Liquid argon TPC signal formation, signal processing and reconstruction techniques

B. Baller

Fermi National Accelerator Laboratory, Batavia, IL, U.S.A.

E-mail: baller@fnal.gov

ABSTRACT: This document describes a reconstruction chain that was developed for the ArgoNeuT and MicroBooNE experiments at Fermilab. These experiments study accelerator neutrino interactions that occur in a Liquid Argon Time Projection Chamber. Reconstructing the properties of particles produced in these interactions benefits from the knowledge of the micro-physics processes that affect the creation and transport of ionization electrons to the readout system. A wire signal deconvolution technique was developed to convert wire signals to a standard form for hit reconstruction, to remove artifacts in the electronics chain and to remove coherent noise. A unique clustering algorithm reconstructs line-like trajectories and vertices in two dimensions which are then matched to create 3D objects. These techniques and algorithms are available to all experiments that use the LArSoft suite of software.

KEYWORDS: Detector modelling and simulations II (electric fields, charge transport, multiplication and induction, pulse formation, electron emission, etc); Noble liquid detectors (scintillation, ionization, double-phase)

ArXiv ePrint: 1703.04024
1 Introduction

The liquid argon time projection chamber (LAr TPC) is ideal for the study of neutrino interactions as it provides mm-scale resolution in a $O(100 \text{ m}^3)$ scale volume. A LAr TPC also produces a prodigious amount of information that must be processed to extract $O(100)$ physics quantities in a neutrino interaction. As an example, the MicroBooNE detector [1] has a volume of $62 \text{ m}^3$. It is instrumented to measure the charge deposited in voxels of size $\approx 10 \text{ mm}^3$ resulting in $6 \times 10^9$ measurements for each beam spill. Images of the data have a photographic quality that has been likened to that of a bubble chamber. This report focuses on the challenge of extracting physics objects from this large data sample.

Significant progress has been made in developing the technology as exemplified by physics results from ICARUS [2] and ArgoNeuT [3] and in the successful operation of test stands such as LAPD [4], LongBo [5], CAPTAIN [6], ArgonTube [7], etc. Results were obtained from semi-automatic reconstruction of complex event topologies or fully-automatic reconstruction of simple events. The large data volumes generated by the current generation of experiments such
as MicroBooNE and LArIAT [8], and the need to reconstruct more complicated events motivates improving reconstruction performance.

Reconstructing isolated particles is routinely done with efficiencies approaching 100%. Minimum ionizing particles deposit $\gtrsim 500$ keV in a wire hit. Efficient reconstruction of tracks $\gtrsim 1$ cm (proton kinetic energy $\gtrsim 20$ MeV) is feasible using techniques that are described here. The difficulty arises in reconstructing tracks that are embedded within a cluster of charge deposited by other tracks. Achieving this capability enables exploring the role of final state interactions in neutrino-argon interactions. Reconstructing tracks in a high-density shower or in the vicinity of a deep inelastic neutrino interaction is a significant challenge.

We begin with a review of a subset of the physics processes that ensue when an ionization event occurs within the detector. Discussion in this section is limited to the processes of creation, loss and transport of ionization electrons that form signals on the TPC wires. Only passing mention is made of the use of a light collection system to provide a trigger. Online processing of wire signals by the readout electronics is discussed only insofar as it motivates the need for offline signal processing. The reader is referred to appendix A that describes a LAr TPC calculator that allows a more precise estimate of the order of magnitude estimates given in this section.

Section 4 describes the deconvolution method that corrects for the response of the readout electronics. This technique was originally developed to overcome an unavoidable artifact of the ArgoNeuT electronics. This method also provides a mechanism for correcting the seemingly complex direct and indirect currents induced on TPC wires by ionization electrons. This processing stage produces a standardized set of wire signals that have a Gaussian-like shape.

A summary description of the standard LArSoft [9] hit finding algorithm is given in section 5. Here we make the observation that an unambiguous connection between a hit and a discrete ionization event cannot be made at this early stage of reconstruction. The TrajCluster reconstruction module described in section 6 is a first step to confront this problem. TrajCluster reconstructs line-like 2D trajectories and creates a new hit collection using local information from those trajectories. TrajCluster includes an effective method for matching 2D trajectories in 3D.

This report describes methods and algorithms that have been developed over the last 8 years. The algorithms were developed in the early days of the ArgoNeuT experiment and are now experiment-agnostic components of LArSoft. More recent users include LArIAT and DUNE [10]. The performance of these methods and algorithms varies with the experiment-specific implementation and the physics processes being studied. The figures in this report highlight general features of real data and some techniques for exploiting those features.

## 2 Signal formation

Electrons liberated by the passage of a particle in a LAr TPC are separated from their parent argon ions by $\approx 2 \mu$m after reaching thermal energies [11]. The ions and electrons form two overlapping columns along the particle trajectory. The columns separate under the influence of the TPC electric field, $E$, that is typically $< 1$ kV/cm. Electron-ion recombination occurs for the next few nanoseconds until the columns are well separated. The fraction of electrons that escape recombination, $R$, can be modeled by a Birks “law” [12, 13] or alternatively by the “Modified Box Model” [14, 15] shown in equations (2.1) and (2.2). Both of these empirical models are based on the columnar theory of
recombination by Jaffe [16] and provide similar performance. The $A_B$, $k_B$, $\alpha$ and $\beta$ parameters of these models are found by a fit to the charge collected from stopping particles. The dependence on $dE/dx$ is shown in figure 2. The $R$ values for these two models differ by less than 10% for $dE/dx < 35$ MeV/cm and approach 25% at 100 MeV/cm.

$$R_{ICARUS} = \frac{A_B}{1 + k_B \cdot (dE/dx)/\mathcal{E}}$$

$$R_{Box} = \frac{1}{\xi} \ln(\alpha + \xi), \text{ where } \xi = \beta(dE/dx)/\mathcal{E}. \tag{2.2}$$

The separation time of the electron and ion columns also depends on the angle of the columns relative to the $\mathcal{E}$ field direction. The two columns for an idealized particle traveling exactly in the $\mathcal{E}$ field direction, $\phi = 0$, would overlap completely for many milliseconds resulting in complete recombination, $R \approx 0$. The expected sin($\phi$) dependence is not included in the above equations because no significant angle dependence has been observed [15]. A likely cause of this discrepancy is that the simple geometric model doesn’t account for the contribution of $\delta$-rays.

![Figure 1](image-url)

**Figure 1.** Left: schematic view of the elements of a LAr TPC in two dimensions. Ionization electrons produced on a track follow the electric field lines (yellow) through the induction planes until they are collected on wires in the collection plane. Right: view of the wire planes as seen by the approaching electrons illustrating that the electron trajectories shown in the left figure are overly simplistic.

Electrons drift to the anode following the electric field lines as shown schematically in figure 1 with a velocity of $\sim 1$ mm/$\mu$s. Figure 2 shows the electric field dependence of the electron drift velocity. A fraction are lost during transport to the anode wires due to attachment on electronegative impurities such as water and oxygen [22]. The charge loss is characterized by the drift electron lifetime, $\tau_e$. The ionization charge remaining after recombination, $Q_o$, is obtained from the collected charge at the wire planes, $Q_c$, using the relation $Q_o = Q_c / \exp(-(t_{\text{arrive}} - t_o)/\tau_e)$ where $t_{\text{arrive}}$ is the time of arrival of the electrons at the anode and $t_o$ is the time of ionization event. The value of $t_o$ may found from a beam timing signal as was done for ArgoNeut, from a photon detection system as is done in MicroBooNE or by selecting cosmic ray muons that pass through both the anode and cathode planes.
Diffusion will increase the spatial extent of the electron cloud governed by the equation
\[ \sigma_D = 2 \sqrt{D_{L/T}(t_{\text{arrive}} - t_0)} \]
where \( D_{L/T} \) is the longitudinal (transverse) diffusion coefficient [18–21]. The diffusion rms in MicroBooNE is 1.5 mm along the drift direction and 2.3 mm transverse to the drift direction at the maximum drift time of 2 ms.

The anode consists of several wire planes with bias voltages set so that ionization electrons travel between the wires of the first set of induction planes inducing a bipolar signal. The charge is collected on the last plane — the collection plane. The wire plane bias voltage settings for transporting 100% of the ionization electrons to the collection plane is a function of the wire diameter, wire spacing and plane spacing [23]. Full transparency, or 100% transmission of electrons to the collection plane, is achieved when the electric field between wire planes increases by \( \gtrsim 40\% \) for each successive plane gap. There is a concomitant increase in the electron velocity of \( \sim 15\% \) in each gap resulting in somewhat narrower wire signals in subsequent gaps.

The wires in each plane are oriented by some angle relative to each other to provide a different view of the ionization event in each plane. The electrons will have a \( \sim 50\% \) longer travel distance in 3D through the wire planes than indicated by the 2D representation shown in figure 1. It frequently happens that a track with \( \phi > 0 \) travels in a direction such that charge is induced on only one wire in one of the views. Track reconstruction is not hindered by this situation if there are more than two views in the TPC however.

Signals formed on anode wires are dominated by the motion of ionization electrons between the wire planes that occur on the time scale of microseconds. The positive leading lobe of the signal induced on the first instrumented induction plane, the U plane shown in figure 1, would be negligible if the Grid plane did not exist because the induced current would occur during the millisecond-scale drift time of electrons in the meter-scale main volume. The negative trailing lobe is sizable since it is created from electrons traveling in the few mm gap between the U plane and the next (V) plane.

The velocity of positive ions is significantly slower than that of electrons, approximately 5 mm/s, resulting in a positive ion buildup, or “space charge”, that can affect the function of long drift TPCs operated in a high rate of background ionization. The electrical circuit is complete when ions reach the cathode plane. This may be a few minutes in a 2 meter drift TPC.
Table 1. Wire configurations for several detectors.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Wire planes and orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArgoNeuT</td>
<td>Grid(0°), Induction(30°), Collection(-30°)</td>
</tr>
<tr>
<td>MicroBooNE</td>
<td>Induction(60°), Induction(-60°), Collection(0°)</td>
</tr>
<tr>
<td>(Proto) DUNE</td>
<td>Grid(0°), Induction(35.7°), Induction(-35.7°), Collection(0°)</td>
</tr>
</tbody>
</table>

The wire plane configurations for several detectors are shown in table 1, where Grid denotes an un-instrumented induction plane. A wire with 0° orientation is vertical. Use of a grid plane restores the positive leading lobe effectively doubling the size of signals on the first instrumented induction plane at the expense of increasing the operating voltage of the wires.

3 Readout system processing

The pioneering ICARUS detector was instrumented with different charge integrating amplifiers on the middle induction plane to produce unipolar signals on all wire plane. Hits on the collection plane were reconstructed by fitting raw wire signals to a double-exponential function with variable rise and fall time.

In contrast, the ArgoNeuT collaboration elected to instrument all planes with the same electronics. Several approaches were considered to process signals on induction plane wires which were found to be poorly represented by analytic functions. A software implementation of a charge integrating amplifier was developed but required a filter to control baseline offsets. Another approach used a library of wire signal shape templates derived from real tracks at various angles in the TPC. This worked well but required significant effort to create and manage the template library for different operating conditions.

Time-domain approaches such as these complicate the identification and suppression of coherent noise. Raw signals in the ArgoNeuT detector had an additional complication due to an impedance mismatch in the readout electronics. The electronics were borrowed from another operating experiment with the condition that they would not be altered. Bench measurements showed a significant baseline undershoot with a time constant of 52 µs. Adaptive baseline correction techniques were studied and found to be inadequate.

The solution described below is unique in that all offline signal processing is done in the frequency domain. ArgoNeuT was the first experiment to employ this approach. It was incorporated into LArSoft and is now used in offline reconstruction by all experiments that use this software suite.

4 Offline signal processing

Detector features and electronics artifacts described in the previous sections can be removed by offline signal processing. The digitized signal is a convolution of the serial effects of signal formation, electron transport through the TPC and processing by the readout chain as shown schematically here: Wire Signal = Ionization ⊗ Recombination ⊗ Diffusion and Attachment ⊗ Field Response ⊗ Noise ⊗ Electronics Response. In this section we describe a Fourier
deconvolution method that removes the most deleterious effects to prepare signals for hit finding. This method also provides a mechanism for removing coherent noise using a Wiener filter \[24\].

We can estimate the frequency range of wire signals by using a Gaussian approximation where the width of the signal in the frequency domain, \( \sigma_f = 1/(2\pi\sigma_t) \), where \( \sigma_t \) is the width in the time domain. The signal in the frequency domain is also a Gaussian having a maximum at \( f = 0 \). An estimate of the high frequency cut-off can be made by observing that the highest frequency component of a wire signal is related to the inverse of the transit time through the wire plane gap. Using an example where \( \sigma_t = 2 \mu s \), a low-pass Wiener filter should have a \( 3\sigma_f \) cut-off at \( \approx 200 \) kHz.

### 4.1 Electronics response

The electronics response of a \( \delta \)-function input is generally obtained from simulations and bench tests. Figure 3 shows the output of the bench test of the ArgoNeuT preamplifier when a \( \delta \)-function signal is injected. The data were taken with the ArgoNeuT DAQ system consisting of 10 bit ADCs sampled at 0.2 \( \mu s \) per tick. The noise rms was \( \approx 1 \) ADC count as can be inferred by the jitter on the long tail. The large baseline shift is due to the impedance mis-match.

![Response of the ArgoNeuT preamplifier to a \( \delta \)-function input. The scatter in the points is due to digitization error of the ADC.](image)

### 4.2 Field response

Determining the field response appears to be a difficult problem. Ionization electrons follow complex 3D trajectories as they pass through the wire planes producing direct and induced signals on nearby wires. Signals are induced on neighboring wires due to inter-wire capacitance. The magnitude and time dependence of these signals is highly dependent on the distance between the electron electron cloud and neighboring wires. Additional complications are that the spatial and temporal distribution of electrons varies with track angle and wire bias voltage settings.

In practice it is sufficient to model the field response with a simple analytic form that can be scaled for different operating conditions, as long as the electronics integration time constant exceeds the intrinsic time constant of electron transport through the wire planes. This method was used in ArgoNeuT and is illustrated in figure 4. The broad features of wire signals simulated by
Figure 4. Top: Garfield simulation of the current on the induction plane in ArgoNeuT (left) and in the collection plane (right) for a track traveling perpendicular to the anode wire with $\phi = 90^\circ$. Solid (dotted) black lines represent the direct (induced) signal. Middle: simplified representation of the Garfield field response used in the Monte Carlo simulation and deconvolution. The values of $N_{\text{ticks}}$, $A_{\text{ind}}$ and $A_{\text{col}}$ were determined by matching simulated and real data. Bottom: real detector wire signals (solid) and simulated wire signals (dashed).

Garfield [25] shown in the top panels are represented by the simple forms shown in the middle panel that are scaled in time and amplitude to match wire signals in the real detector (lower panel). The real detector data shown in the lower panel (solid lines) were obtained by averaging wire signals on 50 wires on a through-going muon with $\phi \approx 90^\circ$. The real muon was reconstructed in 3D and then simulated with the ArgoNeuT Monte Carlo. The dashed lines are the average of 50 wire signals on the simulated muon where the induction (collection) plane, $N_{\text{ticks}}$ was set to the 1.5 (2.5) $\times$ the width obtained from the 2D Garfield simulation. The interpretation of this result is that real 3D
signals can be modeled with a 2D representation if one assumes that the combined effects of the increased path length in 3D and higher drift electron velocity in the inter-plane gaps jointly result in a factor of 1.5 increase in the wire signal width in the first gap and a factor of 2.5 in the second. The field response amplitudes, $A_{\text{Ind}}$ and $A_{\text{Col}}$, were then adjusted and an asymmetry applied to the induction plane response to improve the agreement.

### 4.3 Diffusion and attachment

The effects of diffusion are generally small. Using MicroBooNE as an example, the longitudinal diffusion rms for electrons traveling the full 2.5 m drift is $\sim 1.5$ mm corresponding to an increase in the time spread of 1 tick. Likewise, correcting for electron attachment is a simple multiplicative factor that can be applied in later stages of reconstruction.

### 4.4 Recombination

A recombination correction is required for calorimetric reconstruction but requires knowledge of the path length of the track in space to determine $dQ_o/dx$. It is important to recognize that this correction is unavoidably imprecise since recombination occurs on the physical length scale of microns in the TPC while the correction is applied to the collected charge $Q_o$ with the length scale of the wire spacing. The disparity in the length scales is not important when $dE/dx$ is slowly varying but is significant near the Bragg peak.

Applying a recombination correction to hits reconstructed in electromagnetic showers requires a different but ultimately simpler treatment. The example of a typical low energy photon conversion in which the $e^+e^-$ pair have an opening angle of $\sim 0.01$ radians illustrates this point. The pair initially do not have a dipole field and hence no ionization occurs. After traveling $O(10)$ nm, the dipole separation is about 1 atomic diameter resulting in the onset of two overlapping ionization columns in which recombination occurs between electrons and ions produced by both particles. This situation is somewhat similar to that of a single particle with twice the $dE/dx$ of a minimum ionizing particle. After traveling $O(100) \mu$m, the ionization columns are sufficiently well separated such that recombination occurs preferentially between electrons and ions produced in the same column. For this pair-production example, the charge measured on a wire is largely dominated by the regime in which the particles are well-separated. As a result, it is reasonable to use a recombination factor of $R = 0.64$ for minimum ionizing particle hits that reside in the shower. This approximation clearly fails if the local charge density is very high.

### 4.5 Signal deconvolution

The approach described here is to remove the electronics response and field response, resulting in wire signals that are roughly Gaussian in shape in all wire planes. A fit to a Gaussian distribution should therefore provide a good estimate of the hit position and the ionization charge. Ionization electrons produced on tracks that travel in the electric field direction, $\phi = 0$, will arrive over many microseconds and in this case a single Gaussian fit is inadequate.

The time sampled wire signal read out by the data acquisition system, denoted $W(t)$, is the convolution of the ionization charge approaching the wire plane, $Q(t)$, with the field response, $F(t)$,
and the electronics response $E(t)$.

$$W(t) = Q(t) \ast F(t) \ast E(t)$$  \hspace{1cm} (4.1)

The time-dependent ionization charge can be recovered by deconvolution

$$Q(t) = \mathcal{F}^{-1} \left[ \frac{\Phi(f) \times \mathcal{F}(W(t))}{\mathcal{F}(F) \times \mathcal{F}(E)} \right]$$  \hspace{1cm} (4.2)

where $\mathcal{F}$ is the Fourier Transform and $\Phi(f)$ is a Weiner filter function. Here we have dropped the $t$ dependence of $F$ and $E$ to distinguish the real-time varying wire signal from the real-time invariant response functions. A deconvolution kernel that includes the filter, field response and electronics response is computed for each wire plane. The Wiener filter has the form

$$\Phi(f) = \left[ \frac{S^2(f)}{S^2(f) + N^2(f)} \right]$$  \hspace{1cm} (4.3)

where $S(f)$ is the Power Spectral Density (PSD) of the Signal and $N(f)$ is the PSD of the noise. The filter is constructed by creating histograms of PSD for wires that contain the signal, $S(f)$, of a small angle track, $\phi \approx 90^\circ$, and a sideband histogram that has no detectable signal, $N(f)$. The histograms may then be fit to an analytic function as was done for ArgoNeuT or used directly as done by MicroBooNE. Adjustment of the filter is required to preserve the low frequency components of large angle tracks.

The implementation of these techniques in ArgoNeuT is described in ref. [3] and applied to the calorimetric reconstruction of large angle tracks in ref. [15]. In this work a calibration factor of $< 5\%$ was required to correct for the effects of the offline signal processing chain for tracks with $\phi > 40^\circ$.

These methods are employed by all experiments that use LArSoft however the implementation is different for each experiment. The efficiency of reconstructing hits from wire signals and the quality of those hits is highly dependent on the implementation.

4.6 Discussion

It is important that the filter preserves the most desirable components of the wire signal. The filter may be adjusted to remove lower frequency components resulting in narrow signals in the time domain. This will improve the separation of close tracks but has a detrimental effect on calorimetric reconstruction since signal power is lost in the filter. A minor loss in position resolution will result as well.

The current approach by LArSoft experiments has been to tune the filter for the best calorimetric reconstruction. This approach is well suited to the study of low energy neutrinos produced at Fermilab. The most important particles under study are primary muons and electrons produced by charged current interactions. These have energy less than a few GeV and may be fully or partially contained in the detector. The properties of protons and pions produced in these interactions are also of great interest to better understand neutrino cross sections and the role of final state interactions. These typically have energy less than 0.5 GeV and would be required to be fully contained.
The energy of fully contained particles is best determined using particle identification and range. This requires good calorimetric reconstruction at the expense of poorer position resolution. The momentum of exiting muons can be determined by multiple Coulomb scattering [27, 28] in experiments that lack a muon spectrometer. The contribution of position resolution to these measurements has not been studied in detail.

Suggestions have been made to process wire signals with two different filters; one wide and one narrow. The benefits and complications of this idea have not been explored. There may be some benefit to re-processing wire signals with a narrow filter in the vicinity of neutrino interaction vertices to better identify low energy particles.

5 Hit reconstruction

Hit reconstruction can be approached as the straightforward process of characterizing the position and amount of charge deposited in the detector within a Region of Interest, or ROI, in which a wire signal is above or below a baseline-subtracted threshold. The hit charge, \( dQ \), is the integral of a wire signal above threshold times a calibration constant. The hit time relative to \( t_0 \) can be calculated using the charge weighted mean of the ROI. Information that could be used to separate close tracks or complicated ionization events would be lost using this simple procedure however. The method summarized here from ref. [3] improves the reconstruction of hits; in particular those close to a neutrino interaction.

Each ROI is fitted to a variable number of Gaussian distributions defined by a time, peak amplitude and width. The first step is to estimate the values of these parameters. A set of local peaks is found in the ROI. Each peak is then fitted to a Gaussian distribution with that starting value. An initial fit is performed assuming the noise rms is 1 ADC count. The premise for this assumption is that the wire signals after deconvolution are truly Gaussian in nature and secondly that the deconvolution kernel and filter have been optimized. When these conditions are met, the \( \chi^2/\text{DOF} \) of the fit can be used to distinguish ionization events that are close in time. The \( \chi^2/\text{DOF} \) of the fit is used to decide whether to continue fitting with additional “hidden” Gaussian distributions that do not have a local maximum or to create a “crude hit” that encapsulates the global features of an ionization event in this ROI. The LArSoft hit data product is defined with the expectation that it is Gaussian in nature. A hit is a member of a “multiplet” if it was found in a multi-Gaussian fit.

Figure 5 illustrates these concepts. The top panel is a standard LArSoft event display of wire signals observed in a neutrino interaction in the ArgoNeuT detector in the collection plane. It appears, using visual human-brain pattern recognition and some physics judgment, that three or possibly four charged particles and a photon were produced in the interaction. A different qualitative conclusion results from inspecting the detailed wire signals on four wires near the interaction as shown in the bottom panels. The charge deposited by the two trajectories near Tick 800 on these wires is consistent with the hypothesis that they are individual particles. Interpretation of detailed signals from the large angle trajectory is not as clear. The detailed signals on each ROI has a structure that may be due to dE/dx fluctuations on a single inclined track or the presence of two close small angle tracks.

A second type of ambiguity arises when charge from several tracks overlap and are fit as a single Gaussian hit. This occurs at the primary vertex of every neutrino interaction to some extent.
One negative consequence is that pattern recognition of tracks near the vertex will be incorrect. Another consequence is that the calculation of $dQ/dx$ of short tracks near the primary vertex will be erroneous. Untangling the effects of overlapping tracks should potentially allow the identification of MeV-scale particles produced by final state interactions.

An improvement to the hit reconstruction has recently been made to represent a long wire signal as a series of equally spaced narrow Gaussian hits. A rare event from ArgoNeuT data shown in figure 6 illustrates this feature. A lightly ionizing particle, most likely a beam muon, entered the front of the detector from the left. The muon interacted with an argon atom, producing a charged track that traveled towards the anode plane, $\phi = 180^\circ$, and a neutral particle, most likely a photon, in the forward direction. The ionization from the charged daughter was collected on a single wire in both planes during the 10 cm of travel. The charge deposited on the collection plane wire after deconvolution is shown in blue in the bottom two panels. Hits reconstructed using the traditional hit finder are shown in the middle panel. Hit reconstructed using the improved hit finder shown in the

**Figure 5.** Top: LArSoft event display of wire signals produced in a neutrino interaction. The ADC value is color-coded from blue (low) to red (high). Lower: Detailed shapes of signals on wires in the vicinity of the interaction. Raw (after deconvolution) wire signals are shown in black (blue). Reconstructed hits are shown in orange.
bottom panel are more amenable for pattern recognition and calorimetry. The clustering algorithm described in the next section is designed to take advantage of this improvement.

Figure 6. Top: collection plane and induction plane views of an interaction in the ArgoNeuT detector producing a highly ionizing particle that travels with \( \phi = 180^\circ \), towards the anode plane. The particle travels for 10 cm before it decays or re-interacts. The event display threshold was set close to the noise level to emphasize the small induction plane signal. Middle: hits reconstructed on wire 113 using the standard configuration. Bottom: hits reconstructed on wire 113 using the improved hit reconstruction configuration.

6 Reconstruction techniques — TrajCluster

ArgoNeuT analyses used the LArSoft Hough transform algorithm to reconstruct 2D line-like clusters of hits. Tracks in 3D were then found by matching cluster end points in the two views. This technique is less efficient in a large TPC where particles deviate from a straight trajectory due to the effects of multiple Coulomb scattering. In addition, space charge alters the TPC drift electric field in surface detectors that have a long drift time resulting in 10-cm scale distortions.

The LineCluster algorithm has been used by many LArSoft experiments for several years. LineCluster includes hit merging to mitigate the problem raised in the previous section but is strictly a 2D cluster algorithm. More recently, the Pandora Software Development Kit [29] has been incorporated into LArSoft for use by MicroBooNE and DUNE/ProtoDUNE. Pandora reconstructs a hierarchical set of 3D particles but does not refine hits using information gained from pattern recognition. TrajCluster is a second generation version of LineCluster that reconstructs 2D trajectories and 2D vertices and matches these objects in 3D.
TrajCluster exploits the fact that the energy loss due to ionization on each wire is generally small compared to the kinetic energy of a particle. For example, a muon of 100 MeV kinetic energy has a range of 32 cm in liquid argon which will result in 106 ionization measurements in a plane with a 3 mm wire spacing if the particle travels parallel to the wire plane and perpendicular to the wires. During the first 16 cm of travel, $dE/dx$ increases from 2.3 MeV/cm to 2.6 MeV/cm resulting in a series of 53 hits on adjacent wires which will have similar width and similar amplitude. The trajectory position deviation in this region due to multiple Coulomb scattering is $\sim 0.4$ mm. This is similar to the position resolution obtained from hit reconstruction so all of the hits in this section should be well reconstructed as a straight line. The local state of a trajectory reconstructed in this region is modified as it is extended in other regions to account for the effects of multiple Coulomb scattering, energy loss and other physics processes.

The pattern recognition strategy is tailored to the trajectory under construction and its local environment. Particular attention is given to reconstructing short tracks in the vicinity of neutrino interactions. The strategy differs from other LArSoft modules in several respects. Most of the pattern recognition is done in 2D. Matching of 2D vertices in 3D is done to improve the consistency of 2D trajectories between views. The final matching of trajectories in 3D is a simple process that does not involve pattern recognition. TrajCluster includes 30 algorithms of varying complexity that may be configured for different experiments and analyses. Of relevance for this report is that it produces a refined hit collection.

TrajCluster creates trajectories in a 2D space similar to the LArSoft event display where the wire number increases along the abscissa. A normalized “time” on the ordinate axis is calculated using the TDC tick and drift velocity which are then scaled to the wire spacing. A trajectory is comprised of an arbitrary number of trajectory points along with additional information about the global properties of the trajectory. There is no requirement that a trajectory point lie on a wire coordinate nor is there a requirement that points lie on separate wires. The concept is depicted in figure 7. Points on the trajectory each have a charge and direction vector, $\text{Dir}$, in addition to two positions, “HitPos” and “Pos”. Pos and $\text{Dir}$ are the position and direction found from a trajectory fit. A variable number of hits can be associated with a trajectory point using proximity as the only criterion. The decision to “use” a subset of those hits to define the charge and HitPos of the trajectory point is done in a separate step described below. The local trajectory direction and the Pos position are determined from a linear fit using a variable number of previously added points. HitPos is the charge weighted position of hits that are used in the point. The component of the hit position error transverse to the trajectory is calculated using the error in the time and wire coordinates and the local trajectory angle. The charge, the charge rms, the number of fitted points and the fit $\chi^2$/DOF are important properties of the point that are used to inform tracking decisions. Prior decisions may be easily revised after more points are added to the trajectory and better knowledge of the trajectory is gained.

The seed trajectory is started by selecting two positions in the 2D space to define the position and direction of the first point. The selected positions may be two hits or a hit and 2D vertex. In figure 7, the HitPos position of the first point, TP0, is defined by a single hit on wire 1 and the direction is defined by a hit on wire 2. Open (closed) circles represent points that have (no) used hits. A step is made in the specified direction and search is made for hits on wire 2. A second point, TP1, is added to the trajectory after finding the single hit. A decision was made to use this
hit. The third wire is known to be non-responsive and is ignored. A hit is found on wire 4 and is associated with a new point TP2. The hit charge and width are inconsistent with the previously added hits so it is not used. Three hits found on wire 5 are associated with the trajectory. There are six possibilities for using them; three singlets, two doublets and one triplet. The hit doublet outlined in red is found to be most similar to the the hits in TP0 and TP1 in terms of position and charge. These two hits will later be merged into a single refined hit and written to the output stream. Local and global properties of the trajectory are then updated for use in the next step. Stepping continues until a stopping condition is met. The stopping conditions include encountering a Bragg peak or a 2D vertex.

The properties of the trajectory at the end point are much better known than at the first point. The uncertainty on the angle and the average charge is large for the first points on the seed trajectory. It frequently happens that the first few points should be reconstructed as a separate trajectory, for example when the seed trajectory is created using hits of a daughter particle. TrajCluster performs a reverse propagation procedure to mitigate this problem. An example is given below. After a stopping condition is met, the similarity between the first several points and later points is assessed. If the first points are not similar, they are removed from the trajectory. The trajectory is reversed to step in the opposite direction to re-find the seed hits and possibly re-use them.

An estimate of the momentum is made using multiple Coulomb scattering (MCS) each time a new point is added to the trajectory. It provides a global measure of trajectory point scatter that is used to adjust the pattern recognition criteria. The Gaussian approximation for the MCS scattering
Angle from ref. [30] is:

\[ \theta_0 = \frac{13.6 \, \text{MeV}}{\beta cp} z \sqrt{x/X_o} [1 + 0.038 \ln(x/X_o)]. \]  

(6.1)

where \( \beta c \) is the particle velocity, \( p \) is the momentum and \( x/X_o \) is the distance traveled divided by the radiation length. For this approximation we assume that \( \beta c = 1, z = 1 \), use 14 cm for \( X_o \) and ignore the log term. Figure 8, adapted from ref. [30], illustrates how the rms value of \( \theta_0 \) is estimated. The dotted line in the figure is defined using the positions of the first and last trajectory points. The rms scatter of the intervening points about this line, \( s_{\text{rms}} \), and the trajectory point separation are used to calculate \( \theta_0^{\text{rms}} \). The momentum is then estimated using the equation \( p = 13.6 \times \sqrt{\text{Length}/14}/\theta_0^{\text{rms}} \) (MeV/c), where Length is the separation between the trajectory end points. In later stages of reconstruction the values of \( \theta_0^{\text{rms}} \) of two close trajectories are used to decide whether a 2D vertex should be created between them or if they should be merged into a single trajectory.

\[ x/2 \]
\[ x \]

Trajectory points

\[ s_{\text{plane}} \]
\[ \phi_{\text{plane}} \]

**Figure 8.** Estimating momentum from multiple Coulomb scattering using trajectory points. The red dots overlaid on the figure from ref. [30] represent points on the trajectory. Deviations from the dotted line, \( s_{\text{plane}} \), and the separation between the points is used to calculate \( \theta_0^{\text{rms}} \).

TrajCluster provides the ability to tailor the reconstruction approach for different angle trajectories. To explain this we use the term “small angle” to describe a trajectory of a particle that travels within \( \sim 30^\circ \) of the abscissa in the 2D space defined above. Charge deposited along the trajectory is highly likely to be reconstructed as single narrow hits on each wire. At “large angles”, in the range of \( 30^\circ \)–\( 75^\circ \), wire signals may be reconstructed as single wide hits or as multiplets of narrow hits. Particles that travel at “very large angle”, \( 75^\circ \)–\( 90^\circ \), produce very long wire signals resulting in many hits in a multiplet. As noted above, very large angle trajectories are not necessarily those with \( \phi \) close to zero. An additional consideration is that particles traveling at very large angle are likely to be low momentum and deposit charge on several neighboring wires at a fixed time.

TrajCluster provides the ability to define the angle ranges to account for these effects. Several neighboring wires are searched while stepping for hits to associate with a point if the trajectory is very large angle but only one wire is considered when tracking small angle and large angle trajectories.

This concept is extended further by creating an expectation model for the dependence of hit width on trajectory angle. The model is generated in each wire plane before reconstruction begins using the hit collection passed to the algorithm. The average width of single isolated hits is found
in each plane with the expectation that these will be used in small angle trajectories. The model provides an estimate of the expected width for trajectories of all angles. This information is used in conjunction with hit multiplicity and charge to decide which hits to use in a point.

Reconstruction can be done in several configurable passes through the hit collection. 2D vertices are reconstructed after each pass. On the first pass, a test beam experiment such as LArIAT may elect to reconstruct only long small-angle trajectories. The location of 2D vertices found in this pass can be used to track large angle trajectories. Alternatively the first pass may be configured to preferentially reconstruct through-going muons. This capability is useful in surface detectors such as MicroBooNE and ProtoDUNE which have a large background of cosmic rays. Hits used in these trajectories are removed from consideration on subsequent passes which are then configured to reconstruct increasingly large angle trajectories such as those from low energy neutrino interactions.

Figure 9 illustrates the use of these techniques. The top panel shows a reconstructed trajectory after the first stage of stepping is finished. The trajectory has not been analyzed for quality yet. The seed trajectory was created from a hit on wire 2886 to a hit on wire 2887. This defines the stepping direction to be in the positive direction. Points were added to the trajectory until no more hits were found after wire 3455. It is suggestive by inspecting the top and middle panels that there are two particles that travel almost back-to-back in this view. There is evidence for a small-angle kink in the trajectory near the beginning and one or more Bragg peaks in the charge distribution in this region as shown in the middle panel that suggest a backward going particle in this view. The trajectory point charge is 60 ADC-ticks on the last 540 trajectory points, suggesting that this is the trajectory of a muon that leaves the TPC. Local increases in the charge are likely due to unresolved $\delta$-rays. There is evidence for a resolved $\delta$-ray near wire 3040.

Few points were fitted to the trajectory at the beginning but the trajectory direction was known sufficiently well to cross the first gap of dead wires. The trajectory was tagged as a muon after 109 points were found, triggering a modification of the tracking strategy. The number of points in the trajectory fit, NPtsFit, increased monotonically to 180 indicating that the trajectory has high momentum. The estimated multiple Coulomb scattering momentum exceeded 1 GeV/c. A soft interaction occurred near wire 3106 however the change in the trajectory angle was too small for it to be detected as a kink. Instead the number of fitted points was reduced to 80 to keep $\chi^2$/DOF of the fit below 2. Additional points were added to the fit until a second dead wire gap was encountered. The number of fitted points was held constant even though there was a significant increase in $\chi^2$/DOF. This situation is allowed when reconstructing long muons to improve tracking through large $\delta$-ray showers. While checking the trajectory for quality, large deviations were found in the beginning points. These points were removed and reverse propagation was done. Large deviations were also found at the end points of the trajectory. The end of the trajectory was re-fitted with a lower number of NPtsFit. The bottom panel of figure 9 shows this same event after all algorithms were applied. A 2D vertex is later found between the trajectories.

An alternative sequence of stepping decisions result if hits on wires 3455 and 3454 are chosen for the seed trajectory. In this case the stepping direction is in the negative direction. Points are added to the muon trajectory while stepping from right to left. In this case the trajectory direction is well defined when the kink is encountered at wire 2900, resulting in a stopping condition. No reverse propagation is found to be required for the muon trajectory. The short stopping trajectory between wires 2887 and 2900 is reconstructed separately and a 2D vertex is later found between
Figure 9. Top: LArSoft event display of a trajectory in a wire plane after the first stage of reconstruction. Two particles are mis-reconstructed as a single trajectory. Middle two plots: selected data extracted from a detailed stepping report from the first stage of reconstruction. Trajectory construction proceeded from left to right. NPtsFit is the number of previously added points that were fitted to a line and Chg is the trajectory point charge. When a trajectory point on wire 3106 was added, 180 points on wires 2925 to 3105 were fitted to a line with a \( \chi^2/\text{DOF} \), FitChi, of \( \sim 1 \). Bottom: LArSoft event display of the same event after 2D reconstruction. Two trajectories were reconstructed as a consequence of reverse propagation of the first stage trajectory. A 2D vertex (open star) was created between the two trajectories.
The set of hits that are used in the two trajectories in these two cases is not dependent on the choice of the seed trajectory. The HitPos positions are identical. The trajectory positions, Pos, are well defined but not precisely the same at each point when the stepping direction is reversed however. Improvements will be made to eliminate this small difference if a 2D trajectory LArSoft data product is defined.

A new hit collection is created after reconstruction is complete. Multiple hits that are used in a each trajectory point are merged into a single hit. This process does not trigger the re-reconstruction of hits but instead merges multiple Gaussian distributions from the used hits into a single Gaussian while preserving the total charge. The current state of development is illustrated in figure 10 in which 2D trajectories are converted into a LArSoft cluster data product. Four trajectories were reconstructed near the neutrino interaction vertex.

TrajCluster reconstructs trajectories of varying quality in showering events. Low energy electrons of \( \sim 100 \) MeV energy produce a number of track-like trajectories that can be well reconstructed. The track-like nature of trajectories breaks down at higher energies. The shower in figure 10 shows both a track-like and shower-like features. A long trajectory with high MCS momentum is reconstructed at the start of the shower. The charge at the start of this trajectory can be used to tag it as an electron shower or a photon shower after 3D reconstruction. Short trajectories having low MCS momentum tend to be reconstructed further downstream. Exploiting these features to reconstruct 3D showers is currently under development.

![Figure 10](image.png)

**Figure 10.** LArSoft event display of TrajCluster reconstructed clusters and 3D vertices. Hits in a cluster and connecting lines between hits are drawn with a unique color. Two 2D vertices were reconstructed and matched to create the 3D vertices drawn as small open circles near positions (154, 800) and (194, 1250).

### 7 Performance metric

Another unique feature of TrajCluster is that it includes a metric to assess performance as events are processed. The metric utilizes the LArSoft BackTracker service which returns a list of MC truth particles and the fraction of the deposited energy attributable to each particle in the vicinity of a reconstructed hit. Several performance metrics have been developed within the LArSoft community to measure performance after track reconstruction. All rely on the BackTracker service.
For example, the Pandora group defines track efficiency, completeness and purity metrics in ref. [29]. These quantities are typically presented as a function of particle momentum and angle. These are well-defined metrics but some effort is required to gauge the overall performance.

An important use of performance metrics is to aid in configuring the algorithm for a specific experiment or analysis. There are \( \sim 50 \) parameters for configuring TrajCluster and many are correlated with each other. Exploring the configuration space requires a significant computing and human investment to optimize performance. When MC truth checking is enabled, TrajCluster provides an evaluation of its performance for the given configuration that is updated after each event is processed. This rapid feedback enables a more efficient exploration of the configuration space.

When MC truth checking is enabled, each hit is matched to the MC particle that contributes \( > 50\% \) of the charge to that hit. This requirement is imposed to be consistent with the convention adopted by TrajCluster that a hit is only used in one trajectory. Hits which do not have a dominant charge contribution from a MC particle are ignored. A further requirement is that an MC particle must have at least three matched hits in a plane for it to be used in the metric. This ensures that the metric measures the performance for those particles that are reconstructable.

A reconstructed trajectory may be matched to at most one MC particle — the trajectory that has the highest number of matched hits to that MC particle. The match efficiency, \( E \), is defined to be the number of matched hits in the reconstructed trajectory that are matched to the MC particle divided by the total number of MC truth matched hits for that MC particle. The match purity, \( P \), is defined to be the number of matched hits in the reconstructed trajectory that are matched to the MC particle divided by the total number of hits in the trajectory. The product of efficiency and purity, or \( EP \), is calculated for all reconstructable MC electrons, muons, pions, kaons and protons in all planes separately. A perfectly reconstructed trajectory that is matched to an MC particle will have \( EP = 1 \). A running sum of the combined \( EP \) product for all particles in all planes is printed to the standard output stream to provide real-time feedback. Another feature is the ability to produce a detailed report of reconstruction failures that exceed user-defined thresholds.

The \( EP \) metric is a single number that describes the change in the overall performance of 2D reconstruction for a reference data set when the TrajCluster configuration is varied. One commonly used reference set is a sample of simulated neutrino interactions in the Fermilab Booster Neutrino Beam with average energy of 800 MeV. Interactions with argon produce particles that have kinetic energy in the range of 20–200 MeV. The detector in this reference data set has three wire plane views. The TrajCluster EP performance is 88% for muons and 81% for pions and protons. The overall performance is 86% for these particle types. It should be noted that calorimetric reconstruction requires knowledge of the trajectory in 3D which implies that at least two views must be well reconstructed.

8 Summary

The offline signal processing techniques described here enable a detailed reconstruction of neutrino interactions in a LArTPC. It is evident that the first stage of reconstructing the ionization of particles in the detector is necessarily imperfect. This motivates an iterative approach in which information acquired in clustering is used to refine hits. Existing clustering algorithms, including TrajCluster, provide higher level objects for downstream processing but are not iterative. The development of
a truly iterative reconstruction chain may enable reconstruction of MeV-scale particles from final state interactions.

### A LAr TPC calculator

The LAr TPC calculator is an Excel spreadsheet that is useful for understanding the operation of a LAr TPC. Basic detector parameters such as the wire spacing, wire diameter, drift electric field, etc, are entered resulting in the calculation of measured detector quantities. For example, the wire signal amplitude on the collection plane produced by an idealized track that travels perpendicular to the collection plane wires is calculated. This is the smallest signal of interest for detectors used in neutrino experiments. The calculator utilizes a parameterization of the MicroBooNE and LArIAT cold ASIC preamplifier that was developed at Brookhaven National Laboratory [31], a second stage amplifier and a generic sampling ADC.

Derived quantities such as the drift electron velocity, maximum drift time and signal to noise ratio are displayed in blue shaded cells. Non-scaling properties of liquid argon such as the ionization energy and diffusion coefficients are highlighted by the salmon-colored cells.

The calculator accounts for the electric field dependence of the electron drift velocity. The charge loss due to recombination is calculated using both Birk’s Law and Modified Box Model formulas. The plots shown in figure 2 were produced by the calculator.

### Acknowledgments

Members of the ArgoNeuT collaboration provided helpful comments during the development of these techniques and continue to make improvements. Carl Bromberg and Dan Edmunds from Michigan State University provided guidance during the development of the deconvolution method. I thank Tingjun Yang from Fermilab for his careful reading of section 6.

We gratefully acknowledge the support of Fermilab, the U.S. Department of Energy and the National Science foundation. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.
Table 1. Parameters for MicroBooNE detector configuration.

<table>
<thead>
<tr>
<th>Detector Properties</th>
<th>Value</th>
<th>Units</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max drift distance (L)</td>
<td>2.56</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Drift Field</td>
<td>0.175</td>
<td>kV/cm</td>
<td></td>
</tr>
<tr>
<td>Electron drift velocity</td>
<td>1.3</td>
<td>mm/ns</td>
<td></td>
</tr>
<tr>
<td>Max electron drift time</td>
<td>1.0</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Max drift velocity</td>
<td>6.4</td>
<td>cm/s</td>
<td></td>
</tr>
<tr>
<td>Max electron drift time</td>
<td>8.0</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Gaps 1 &amp; 2 drift for full transparency</td>
<td>0.65, 0.62</td>
<td>kV/cm</td>
<td></td>
</tr>
<tr>
<td>Wire testing</td>
<td>3</td>
<td>mm</td>
<td></td>
</tr>
<tr>
<td>Wire diameter</td>
<td>0.15</td>
<td>mm</td>
<td></td>
</tr>
<tr>
<td>Wire spacing</td>
<td>3</td>
<td>mm</td>
<td></td>
</tr>
<tr>
<td>Wire capacitance in air</td>
<td>16</td>
<td>pF/m</td>
<td>Includes 5% of parasitic capacitance for wire attachment</td>
</tr>
<tr>
<td>Wire length</td>
<td>10</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Wire capacitance in L</td>
<td>293</td>
<td>pF/m</td>
<td></td>
</tr>
<tr>
<td>Omega High voltage</td>
<td>770</td>
<td>kV</td>
<td></td>
</tr>
<tr>
<td>Bias voltage Induction 1</td>
<td>12.2</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>Bias voltage Induction 2</td>
<td>6</td>
<td>V</td>
<td>Assumed grounded</td>
</tr>
<tr>
<td>Bias voltage Collection</td>
<td>391</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>Front end amplifier shaping time</td>
<td>2 microsec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digitizer sampling rate</td>
<td>2</td>
<td>kA</td>
<td></td>
</tr>
<tr>
<td>Sample time</td>
<td>0.33</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Front and amplifier gain</td>
<td>7.8</td>
<td>mV/PE</td>
<td></td>
</tr>
<tr>
<td>ADC noise factor</td>
<td>2</td>
<td>ADC counts/peV</td>
<td></td>
</tr>
<tr>
<td>Electronics Folded Assumption</td>
<td>7.2</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Equivalent CO2 contamination</td>
<td>29</td>
<td>pCi/mL</td>
<td></td>
</tr>
<tr>
<td>Charge loss at maximum drift</td>
<td>17%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of MPS assumption</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (Y)</td>
<td>2.33</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Length (Z)</td>
<td>10.27</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Active Volume</td>
<td>41.7</td>
<td>cm³</td>
<td></td>
</tr>
<tr>
<td>Active Mass</td>
<td>81.1</td>
<td>kg</td>
<td></td>
</tr>
<tr>
<td>Crystal pressure</td>
<td>124</td>
<td>bar</td>
<td></td>
</tr>
<tr>
<td>Crystal pressure</td>
<td>18.6</td>
<td>bar</td>
<td></td>
</tr>
</tbody>
</table>

- **General B/Ar Properties**
  - Unit Temperature: 89.2 K
  - Unit Density: 1.378 g/cm³
  - Effective constant: 1.51
  - Wilson ionization energy: 22.6 eV/pe
  - Multiplicity: 39.5 eV/pe
  - Relative Number of MPS assumption: 2.06
  - Recomposition factor: 0.63
  - Photon recombination factor: 0.49
  - Intrinsic isotropy: 0.0036 cm²/s
  - Diffusion coefficient: 3.1 cm²/s

- **Electronics Signal Noise**
  - After Recombination: 16000 electrons
  - Front and amplifier Signal + Collection: 20 microvols
  - Gain: 42 volts
  - ENC @ 90: 420 electrons
  - ENC @ 900: 1.2 counts/pe
  - ENC @ 9000: 522 electrons
  - Signal to noise ratio: 37
  - Signal to noise ratio @ max: 37
  - Induction 2 signal width estimate: 2.1 microsec
  - Collection signal width estimate: 0.8 microsec
  - Diffusion length: 1.5 mm
  - Longitudinal diffusion longer: 0.8 ADC 80
  - Transverse diffusion: 2.3 mm
  - Resolution: 51200 (photons/cm)

**Figure 1.** LAr TPC calculator configured with MicroBooNE detector parameters.
References


