Modelling detector-specific reconstruction uncertainties in LAr-TPC

Lorenzo Vincenzo D'Auria[1]*, under the supervision of Harry Hausner, Angela Fava

Abstract
The Short-Baseline Neutrino (SBN) program features three Liquid Argon Time Projection Chamber (LAr-TPC) detectors positioned along the Booster Neutrino Beam (BNB) axis: the Short Baseline Neutrino Near Detector, MicroBooNE, and the ICARUS T600. As the largest operational LAr-TPC, ICARUS T600 serves as the far detector, located 600 m from the BNB target. While its primary goal is to record neutrino events, it also detects other ionizing events, including cosmic rays.
This work focuses on analyzing and modeling detector-specific reconstruction uncertainties in LAr-TPC. These inefficiencies, identified during the Pattern Recognition phase handled by the PANDORA algorithm, impact subsequent Particle Fits and Offline Analysis. Specifically, inaccuracies in track reconstruction can lead to significant physical consequences, such as erroneous particle energy estimates and poor Particle Identification (PID), reducing the efficiency of neutrino event characterization. A key issue addressed is split tracks, caused by missing hits or incomplete track stitching by PANDORA. The aim of this internship is to characterize, model, and quantify the impact of split tracks on track reconstruction.

Keywords
ICARUS — SBN — Neutrino — SplitTracks

[1] *Department of Physics and Astronomy, Universiity of Padua, Padua, Italy*
***Author**: lorenzovincenzo.dauria@studenti.unipd.it

Contents

1. Introduction

1.1 Physics Program

Experimental observation of neutrino oscillations have established a picture consistent with the mixing of three neutrino flavors (ν_e, ν_μ, ν_τ) with three mass eigenstates (ν_1, ν_2, ν_3) whose mass differences turn out to be relatively small. However, several experimental anomalies have been reported which could be hinting at the presence of additional neutrino states with larger mass-squared differences participating in the mixing. Mainly two distinct classes of anomalies pointing at additional physics beyond the Standard Model (SM) in the neutrino sector have been reported, namely [1]:

(a.) **Reactor and Gallium Anomaly:** Refers to the deficit of $\bar{\nu}_e$ observed in numerous detectors a few meters away (short baselines) from nuclear reactors compared to the predicted rates. A similar indication comes from radioactive ν_e source in Gallium solar neutrino experiments (e.g., GALLEX and SAGE). This deficit could be explaines through ν_e disappearance due to oscillations at $\Delta m^2 \geq 1eV^2$.

(b.) **LSND/MiniBooNE anomaly**: Refers to the evidence for electron-like excesses in interaction with ν_μ and $\bar{\nu}_\mu$ from particle accelerator in Liquid Scintillator Neutrino Detector (LSND) and MiniBooNE experiments.

The most common interpretation of this collection of data is evidence for the existence of one or more additional, mostly "sterile" neutrino states with masses at or below the few eV range. The minimal model consists of a hierarchical 3+1 neutrino mixing and the new sterile neutrino would mainly be composed of a heavy neutrino ν_4 with mass m_4 such that $\Delta m^2_{41} \approx [0.1 - 10] eV^2$ and $m_1, m_2, m_3 << m_4$ [2].
In this 3+1 minimal extension to the SM, the effective ν_e and ν_μ disappearance probabilities are described by:

$$P^{3+1}_{\nu_\alpha \to \nu_\beta} = \delta_{\alpha\beta} - 4|U_{\alpha 4}|^2(\delta_{\alpha\beta} - |U_{\beta 4}|^2) sin^2\left(\frac{\Delta m^2_{41} L}{4E_\nu}\right)$$

where U_{ij} are elements of the now 4x4 mixing matrix (extension of the 3x3, U_{PMNS}, mixing matrix) and L is the travel distance of the neutrino of energy E_ν.
The Short-Baseline Neutrino (SBN) physics program [3] fits in this framework and it was set up to:

1. Understand the low energy excess found in MiniBooNE by exploiting MicroBooNE;
2. Search for sterile neutrino with Short-Baseline Near Detector (SBND) and ICARUS (both in the appearance and disappearance channels);
3. Further develop the LAr-TPC technology and measure the $\nu - Ar$ cross section in the GeV region for future Long Baseline experiments (e.g., DUNE).

In the following sections, the SBN program will be explored in depth, illustrating the operating principle of the LAr-TPC and paying particular attention to ICARUS.

1.2 The Short Baseline Neutrino program

The Short-Baseline Neutrino (SBN) program [2] includes three Liquid Argon Time Projection Chamber (LAr-TPC) detectors located on-axis in the Booster Neutrino Beam (BNB):

1. The **Short Baseline Near Detector** (SBND, 2024-now) will be located 110 m from the BNB target;
2. The **MicroBooNE** detector (2015-2021), located in the Liquid Argon Test Facility (LArTF) at 470 m, is a 170-ton total mass (89-ton active mass) LAr-TPC contained within a conventional cryostat. It was developed to measure neutrino interactions in argon for multiple reaction channels and to investigate the source of the currently unexplained excess of low-energy electromagnetic events observed by MiniBooNE.
3. The **ICARUS T600** detector (2021-now) serves as the far detector within the SBN program, positioned 600 m from the target along the BNB axis. It contains 760 tons of LAr (476-ton active mass). For a more in-depth description see *section* 1.3.

Figure 1 provides a schematic view of the SBN detectors.

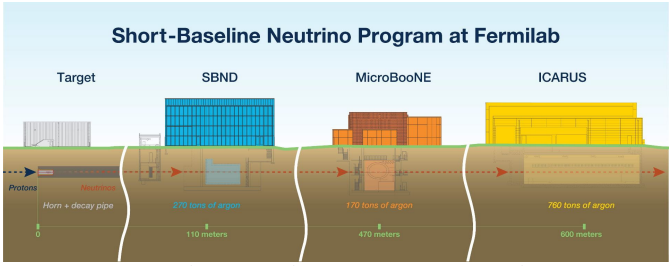

Figure 1. *Short Baseline Neutrino project layout.*

Finally, the BNB is produced by extracting protons from the Booster accelerator at 8 GeV kinetic energy and directing them onto a beryllium target to generate a secondary beam of hadrons (primarily π). These hadrons are focused and allowed to propagate down a 50 m long air-filled tunnel, where the majority decay to produce ν_μ and ν_e; the remaining hadrons are absorbed by a concrete and steel absorber at the tunnel's end.
The physics program of ICARUS is enhanced by the stand-alone studies of neutrino cross-section relevant to the Long Baseline Neutrino Facility program, thanks to the off-axis neutrinos from the NuMI beam. Data collected with the NuMI beam will also be exploited for searches beyond the SM and dark sector analyses. The two beam-lines are schematized along with the rest of the accelerator complex in Figure 2.

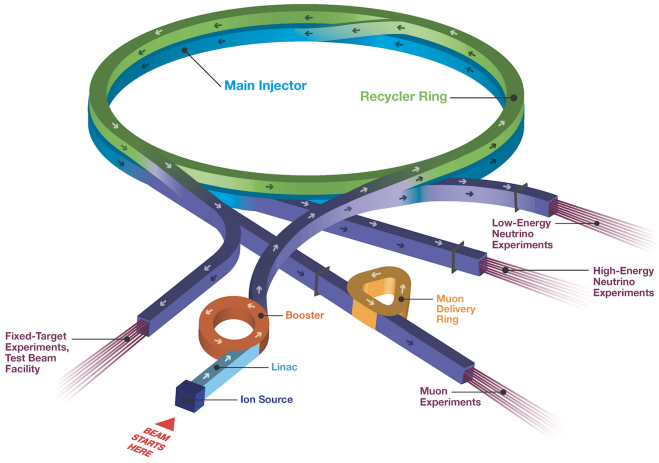

Figure 2. *Schematic of the Fermilab accelerator complex.*

LAr-TPC Working Principle All the SBN detectors are based on LAr-TPC experimental teechnique [4]: this device has the potential to reconstruct tracks and showers with high level of detail and efficiency, as well as to provide a precise measurement of ionization charge necessary for good particle identification based on ionization energy loss. The LAr-TPC principle of operation is schematically illustrated in *figure* 3

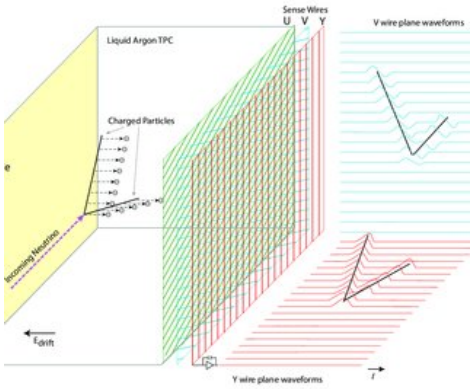

Figure 3. *The LArTPC Principle of Operation.*

Ionization electrons are produced in the medium by ionizing particles associated with an event; due to an electrostatic field maintained in the chamber, they start drifting to the wires, while positive ions travel towards the cathode. Drifting electrons then create a signals on the wires belonging to the induction wire planes "U" and "V". When the finally reach the collection plane "Y" (where they are removed from the detector volume) a signal is also produced on its wire. Then the ionization pattern can be reconstructed in the plane perpendicular to the drift direction by analyzing signals on (U,V,Y); every induction and collection wire is then readout as an individual channel. Reconstruction along the axis of the drift direction can be done using the information contained in the time evolution of the signals on the (U,V,Y) wire planes through the know value of drift velocity.

1.3 The ICARUS T600 Detector

The ultimate goal of ICARUS T600 detector [5] is to record neutrino events, however, in addition to them, the detector records all any kind of ionizing events, also crossing cosmic rays. To be able to discriminate interesting events, the analysis relies on the information provided by the various detector components: TPC, PMT and CRT sub-system. In this section the three subsystems are described in more detail .

Cosmic Ray Tagger (CRT) The CRT system is a subdetector external to the cryostats, designed to identify charged particles coming from outside the beam and passing through or near the active volume of the TPC. It covers an area of $\approx 1100 m^2$ and its divided into three subsystems (top, side and bottom CRT), each covering different regions of the TPC. All these subsystems have a time resolution of the order of *ns*, allowing a filtering of events that rejects those where the primary event trigger was tagged in the subsystem. The CRT is designed to tag $\approx 95\%$ of the cosmic particles entering the TPCs.

Time Projection Chamber (TPC) ICARUS consists of two identical adjacent modules (cryostats, named west and east modules with respect to the BNB beam direction, see *figure* 4) with dimensions of $3.6 \times 3.9 \times 19.6 m^3$ with 760 tons of ultra purified liquid argon. In each module there are two TPCs separated by a central, vertical, common cathode, in which there is an uniform electric field of $\approx 500 V/cm$ going from the cathode to anode, with a maximum drift time of about $1 ms$. As said, each TPC has 3 parallel read-out planes, 3mm apart from each other: the first plane (induction 1) has horizontal parallel wires, while the other two (induction 2 and collection) are oriented $\pm 60^o$ with respect to it. Each of the TPCs consists of 13332 wires: the reconstruction of the image of the charged particle traversing the volume is obtained by combining the coordinates in the wires of each of the planes. For a more in-depth description of event reconstruction see the *section* 2.

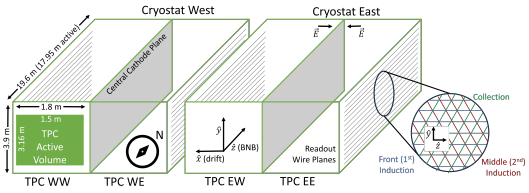

Figure 4. *ICARUS T600 diagram.*

Light Collection System (PMT) Ionization in the LAr TPC is accompanied by a scintillation light emission, which is useful for a possible absolute time measurement of an interaction and triggering. A large set of 360 PMTs (90 for each TPCs) is immersed in the liquid argon, so UV photons from the scintillation provide a signal that allows a measurement of the drift time, and hence of the distance traveled by the drifted electrons. The photomultipliers selected for each TPC were 90 Hamamatsu R5912- MOD of 8 diameter hemispheral glass mounted behind the wire chambers and adapted to operate at cryogenic temperatures on each TPC. The electronic readout system of the PMTs is designed to allow continuous readout, digitization and an independent waveform recording of the signals coming from each one of the PMTs of the light detection system. In *figure* 5 it is possible to appreciate the arrangement of the PMTs inside the ICARUS T300 module

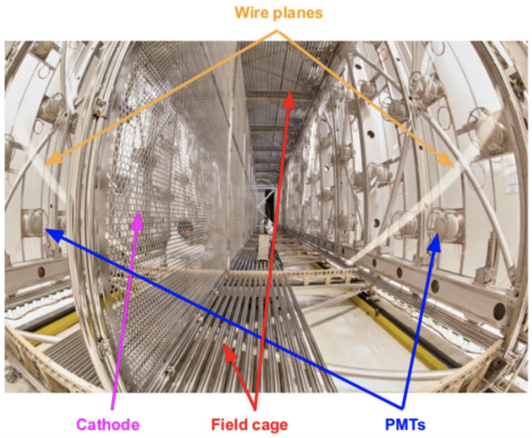

Figure 5. *Internal view of an ICARUS T300 module, during the refurbishing activities at CERN.*

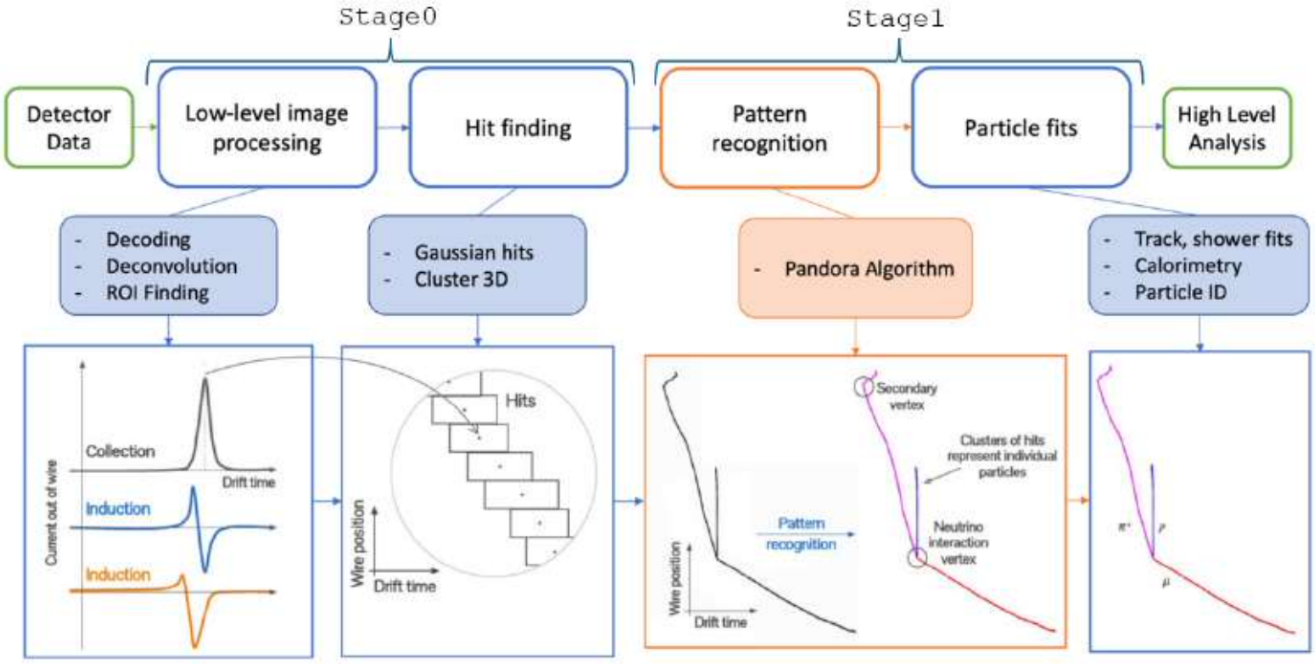

Figure 6. *Illustration of the ICARUS TPC signals processing and reconstruction process.*

2. Event Reconstruction in ICARUS

LAr-TPCs design and acquisition parameters lead to a substantial data rate and volume: to deal with it, several strategies have been introduced in the ICARUS data processing chain. The output data from the detector consists of digitized waveforms from each TPC readout channel, representing the charged induced by the motion of ionization electrons swept by the drift field from the TPC volume. All the collected raw data needs to undergo several software phases before it can be properly analysed; the reconstruction of these events is done in two stages: *Stage 0* and *Stage 1*.

Stage 0 The first stage (*Stage 0*) translates from raw data format to LArSoft[1] format for offline analysis. It also performs signal processing for all three subsystems with the goal of identifying physical signal (*Hits*) to be fed to the poattern recognition and event reconstruction algorithms. *Stage 0* processing for TPC includes:

1. Decoding: Decompresses and unpacks raw data into a format that can be used in later steps. Includes electronic noise filtering to remove TPC coherent noise.
2. Deconvolution: Removes effect of the electrostatic field around the wires and electronic response.
3. Region Of Interest (ROI) finding: Reduces data volume by selection of ROI around candidate signal.
4. Hit Finding: Builds "Hit" objects from the ROIs in the previous step. A Hit represents the identified signal induced by a charged particle on a wire and are the primary input to the pattern recognition algorithms.

Stage 1 The second stage (*Stage 1*) processing is mainly focused on the reconstruction of TPC, CRT and PMT signals (and might include calibrations of each subsystem). *Stage 1* processing for TPC includes:

1. Hit filter: Builds tridimensional space points from combination of 2D Hits across different wire planes.
2. Pattern recognition algorithm (PANDORA): Arranges close hits into clusters that are then used to identify track or shower candidates and event topology information, including cosmic ray identification. Inside each defined interaction, the so-called slices, a hierarchy among all reconstructed objects is build to identify parent-daughter relationships. A more detailed description of PANDORA's reconstruction algorithms is given in the *section* 2.1.
3. Particle fits: Applies detailed algorithms to reconstruct tracks and showers and obtain calorimetric measurements of each particle. It also provides information to allow kinematic reconstruction and to analyze the event.

A schematic view of the event reconstruction algorithm in ICARUS TPCs is shown in *figure* 6.

[1] LArSoft is a toolkit that provides a software infrastructure and algorithms for the simulation, reconstruction and analysis of events in LAr-TPC.

2.1 PANDORA reconstruction algorithms

PANDORA has two main chain algorithms for event reconstruction in neutrino detectors [6]: PANDORACosmic and PANDORANu, targeting the reconstruction of interactions due to cosmic rays and neutrinos respectively.

PANDORACosmic PANDORACosmic clusters 2D hits on each readout plane, splitting them at bifurcations or ambiguities to ensure high purity. Clusters are then merged based on proximity or directional alignment to improve completeness. These refined 2D clusters undergo 3D reconstruction by matching consistent 2D groups across all three planes, with suitability scored and evaluated using a χ^2 test to resolve ambiguities. Unassociated 2D clusters, assumed to be delta ray fragments, are dissolved. The initial position for cosmic muons is assigned at the highest vertical coordinate, with secondary particles linked hierarchically.
Tracks are classified as clear cosmic if they cross the detector boundaries, have hits outside the drift volume, or have incompatible time corrections. Once clear cosmics are removed, the remaining clusters are input to PANDORANu, which processes them into slices to isolate neutrino interactions or cosmic remnants, resulting in neutrino candidates.

PANDORANu The first step in PANDORANu reconstruction involves using track-oriented clustered hits to generate possible 3D vertex candidates. A quality cut evaluates the proximity of each vertex to hits in all plane views, and a scoring algorithm selects the most likely neutrino interaction vertex. The score is based on three factors: how well particles point back to the vertex, asymmetry between downstream and upstream hits, and proximity to the beam origin. The candidate with the highest score is chosen, and 2D clusters crossing the vertex are split, creating new clusters on each side. Next, 3D track reconstruction proceeds similarly to PANDORACosmic, but with additional reconstruction of primary electromagnetic showers. A Boosted Decision Tree (BDT) then classifies each particle as shower-like or track-like. Subsequent steps refine this classification and organize particles into a parent-daughter hierarchy. The final structure starts from the 3D neutrino vertex, adding branches for primary particles and leaves for particles produced by primary parents. The interaction grows until all particles in the slice are associated, resulting in a single neutrino particle with an internal particle hierarchy that represents the flow of the interaction.

2.2 LArSoft event scanner

An important tool for the work lies in the possibility of viewing the events reconstructed by PANDORA using the LArSoft event display: in this way it is possible to directly observe the events, understanding their characteristics and highlighting any anomalies in the reconstruction.

LArSoft Wire Geometry LArSoft assigns logical IDs to each TPC readout channel, organizing the ICARUS detector into two cryostats, each with two TPC sets [7]. Each TPC set contains four readout planes, with channels numbered based on their position relative to the cathode. Channels in each readout plane are ordered first by increasing z-coordinate and then by increasing y-coordinate. The cryostats are labeled as C:0 for module E and C:1 for module W. Each TPC set (labeled S) corresponds to an actual TPC and drift volume, and includes two logical TPCs (each half of the physical TPC, the division of which occurs at the separation of the horizontal wires), labeled T. The TPC sets are further numbered with increasing x-coordinate: TPC EE is C:0 S:0, TPC EW is C:0 S:1, TPC WE is C:1 S:0, and TPC WW is C:1 S:1. Readout planes are groups of channels on the same type of plane within a TPC set, with each readout plane potentially containing one or more logical wire planes. The channel numbering starts from the first TPC set and the readout plane closest to the cathode, with an order based on z and y coordinates. Wireless channels are associated only with readout planes, not wire planes. In total, 55,296 channel IDs are assigned, with 53,248 connected to TPC wires and the rest allocated to wireless, ghost, and virtual channels. LArSoft wire planes are numbered starting from the plane closest to the cathode (P:0).

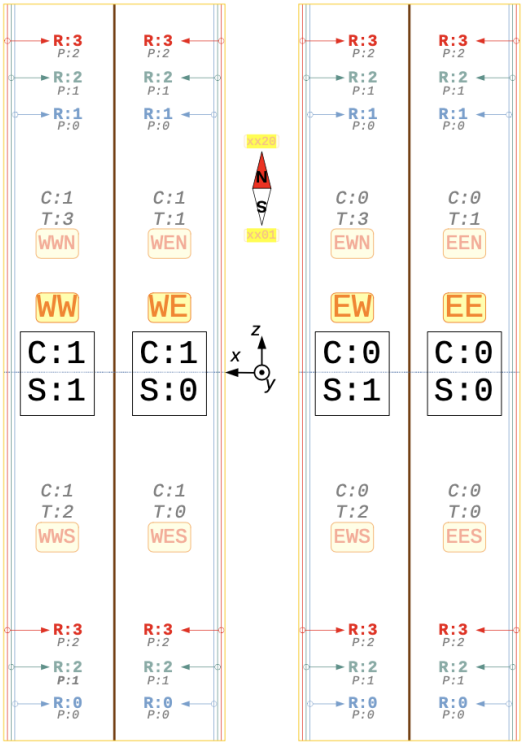

Figure 7. *Illustation of wire plane layout as referenced in [7].*

3. Split Tracks Identification

The purpose of this work is to analyze and model detector-specific reconstruction uncertainties in LAr-TPC. Such reconstruction inefficiencies are identified during the "Pattern Recognition" phase (handled by the PANDORA algorithm already discussed in the previous section) and significantly affect the subsequent steps of Particle Fits and High Level Analysis. In particular, incorrect track reconstruction directly leads to significant physical consequences such as erroneous estimation of particle energies or poor Particle Identification (PID), which together result in a substantial reduction in the efficiency of characterizing events associated with neutrino interactions.
An important class of detector-specific reconstruction uncertainties in LAr-TPC is represented by "Splitting Tracks": these are caused by missing hits that were not accounted for at the reconstruction level or that PANDORA was not able to account for when stitching (joining together the segments belonging to the same particle track) the divided track segments. At the detector level, the most common known reasons for track splitting are [8]:

- *Malfunctioning channels:* as it is known that a set of 32 wires is disconnected in one of the TPCs;
- *Presence of the cathode plane*;
- *Electric field distortions:* produced by the fact that the Induction 1 plane is made by two sets of wires joined in the middle of the plane, with a few cm wide region not equipped with them.

Internship Aim The aim of the internship lies in characterizing and modeling the split tracks and quantifying this effect in track reconstruction. These results were achieved through two main steps:

- The first step consists in writing a plug-in, based on LArSoft's ART framework, whose purpose is to identify the presence of split tracks based on their geometric characteristics (length, gap, and angle between them). The calculation of observables will be discussed in more detail in *section* 3.1, while the determination of the cuts to be made to discriminate the split tracks is discussed in *section* 3.2. Finally, the plug-in's selection capability will be validated in *section* 3.3;
- Once the performance of the plug-in has been validated, the distribution of lost channels and gap lengths will be recorded (*section* 4.1): once the behavior of these distributions is understood, they are implemented in the HARPS tool, with the aim of generating a dataset where hits are removed from the tracks according to the gap distribution (*section* 4.2).

3.1 Plug-in logic and observables calculation

The plug-in written for the identification of split tracks, and used on Stage 1 files, is based on three main steps:

1. Loading of track pairs and calculation of variables;
2. Selection cuts;
3. Storage of results in .root files.

The following paragraphs will describe the content of the various parts of the code in more detail.

Importation and calculation of variables The plug-in, starting from a track identified with an index consistent with that assigned by PANDORA, iterates over all the tracks, constructing a series of possible pairs. For each track in the pair, some basic geometric characteristics are obtained: the logical TPC containing the track, its length, and the position and direction of the start and end points. At this point, the number of possible pairs is significantly reduced through a pre-selection based on track length, requiring that the tracks be longer than $L_{th} = 5cm$: this threshold, currently purely indicative, is set considering the physical hypothesis that a track splitting event is more likely to occur for physically longer tracks. After this pre-selection, the track pairs are significantly reduced, and two fundamental geometric variables between them are calculated: the gap separating them (calculated as the norm of the segment connecting the endpoint of the first track to the start point of the second) and the angle between them. However, the angle calculation is much more complex and requires further explanation.

The angle between the tracks is calculated based on two fundamental cases:

- Case of completely parallel tracks: in this case, the angle θ between the two tracks is obtained as

$$\theta = \pi - \alpha$$

  With α being the angle between the direction vector of the endpoint of the first track and the vector connecting the endpoint of the first track and the start point of the second;

- Case of non-parallel tracks: In this case, the two tracks may not intersect in $\mathscr{R}^3$, so the direction of the last point of the first track is projected forward and the direction of the first point of the second track is projected backward. The midpoint M of the perpendicular line minimizing the distance between them is then found, and the angle θ formed by the end of the first track, M, and the start of the second track is stored. To streamline the discussion, the detailed calculations are reported in *Appendix*. A graphical representation of the process is shown in *Figure* 8.

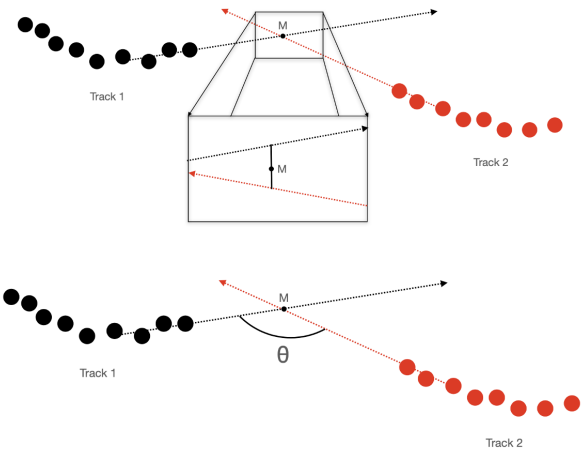

Figure 8. *(Top figure) Example of 3D projection; (Bottom figure) Example of the stored angle.*

A major issue that emerged when applying the methods proposed above is associated with tracks crossing the cathode: these tracks exhibit a distorted profile at the ends (as shown in *Figure* 9), which alters the angle estimation when considering only the extremal points of the tracks. To address this issue, the direction of the first and second tracks is obtained as the average of the last and first $N = 20$ hits, respectively. Although simplistic, this solution provides a satisfactory estimate of the angle between the tracks (as confirmed in the analysis presented in *Section* 3.2).

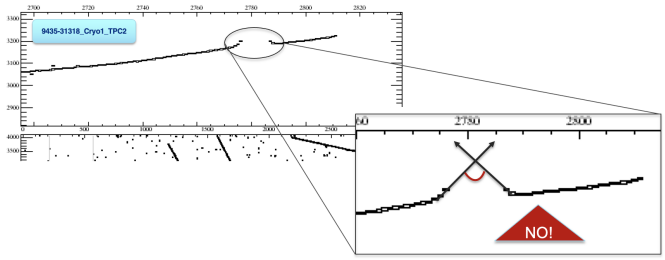

Figure 9. *Example of anomalous behavior in the angle calculation: case of a track crossing the cathode.*

Selection cuts Once the observable reconstruction capabilities have been validated, it is necessary to determine the optimal cut values to apply to the variables in order to discriminate split track pairs from regular track pairs. For this purpose, a run whose split track information was previously obtained from the event display is used, and for a more detailed description of the selection process, refer to *Section* 3.2.

Storage of results The output results from the plug-in lead to the production of a *.root* file consisting of different TTrees depending on the cuts applied: only on the gap, on gap and angle, on gap-angle-length. For each track pair that passes the pre-selection on the length threshold, the observables are also computed and saved, so that comparisons with the data sets after the cuts can be made. For each program execution, information about the tracks (such as the IDs of the tracks identified as a pair) is also printed to the screen, allowing for cross-checks with what is visualized using the event scanner (this function will be used to validate the plug-in's ability to select split track events and will allow for the visual identification of any anomalies, see *Section* 3.3 for more details).

3.2 Determining the selection cuts

This section discusses the results of applying the plug-in, particularly validating the plug-in's ability to satisfactorily reconstruct the observables of interest. The primary goal here is not to evaluate the plug-in's selection capability but rather the consistency of the reconstructed variables' distribution with the reference values. The run under study (hereafter referred to as "Run 9435") has previously been studied using an event scanner, through which information about the presence or absence of split tracks has been obtained, and the relevant variables were calculated. In the following, "reconstructed variables" refers to those calculated by the plug-in, while "true variables" refer to those reported in the archive.

To validate the reconstruction capability, the search was limited to events where the presence of split tracks was certain, and, in addition to the pre-cut on the threshold length, a cut on the gap between two split tracks was applied, selecting only pairs of tracks with a gap smaller than 25*cm*. The results are shown in *Figure* 10.

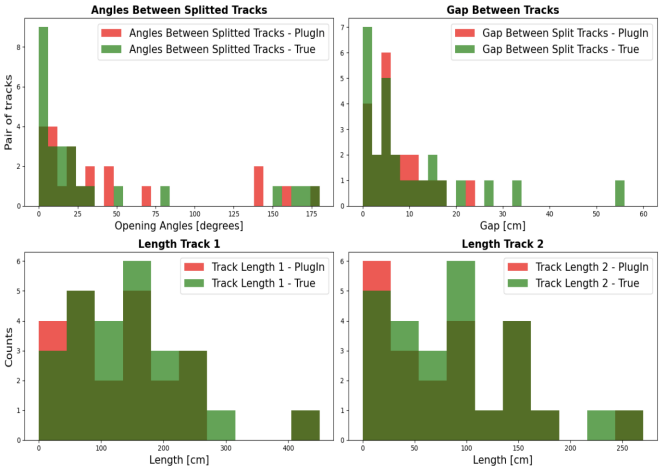

Figure 10. *Comparison between reconstructed and true variables. Light green represents reference variables, red represents reconstructed variables, and dark green shows the overlap of the histograms.*

It can be seen that there is satisfactory agreement both in the selection degree and, especially, in the reconstruction of the variables. Specifically, for the length calculation, the agreement is almost total (fewer than three events where the plug-in does not detect due to the too-tight gap cut).
After confirming the quality of the reconstruction, the distribution of the same variables is analyzed for all track pairs present

in the run (which, let us remember, pass the pre-selection). The trend is shown in *figure* 11.

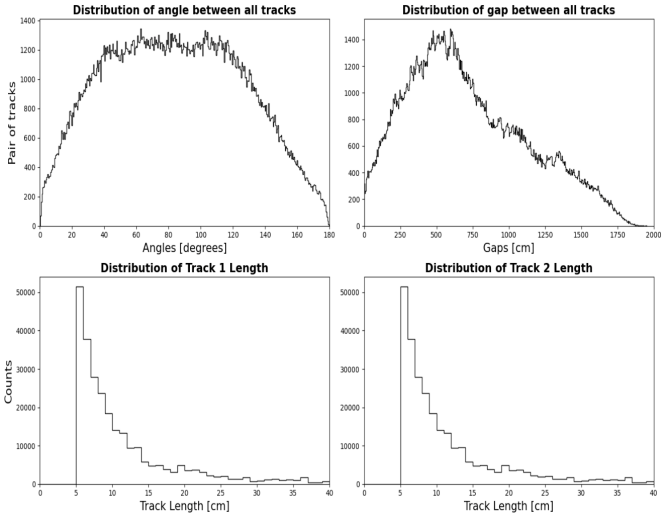

Figure 11. *Distribution of variables for all track pairs passing the pre-selection.*

Based on the event scanner visualization of the split tracks, some recurring geometric characteristics can be observed. In fact, they must be:

- **Parallel and close:** We expect the splitting to occur in a reduced portion of space and likely along a straight trajectory;
- **Long:** Avoiding considering isolated particles as belonging to the same track, while also imposing a considerable track length in relation to the requested gap length.

While the first characteristic is easily understandable, we can convince ourselves of the second by comparing the length distributions for all track pairs and for known split tracks. It is possible to highlight that most of the counts (in the lower plots of *figure* 11) occur for lengths shorter than $20cm$; whereas, in the split track distribution, evident percentages are also present at higher lengths. Therefore, a cut around the length of $L_{cut} = 20cm$ was chosen (which will later be refined to $L_{cut} = 40cm$ by analyzing data samples from other runs), significantly reducing the total number of tracks at the cost of losing a small fraction of split tracks.

3.3 Validating the plug-in capabilities

After validating the plug-in's ability to reconstruct the variables of interest, in order to optimize the cuts to select split track events and validate the selection capability, a dataset comprising eleven different testing runs was used. The choice to use all these runs is motivated by the desire to have sufficient statistics to perform a thorough analysis and validate the reconstruction capability on a larger sample.

At this point, it was first decided to construct the histograms of the variables of interest for the entire dataset to understand the overall trend and compare it with what was obtained in the previous section. *Figure* 12 shows the trend of the variables for the 7xxx runs.

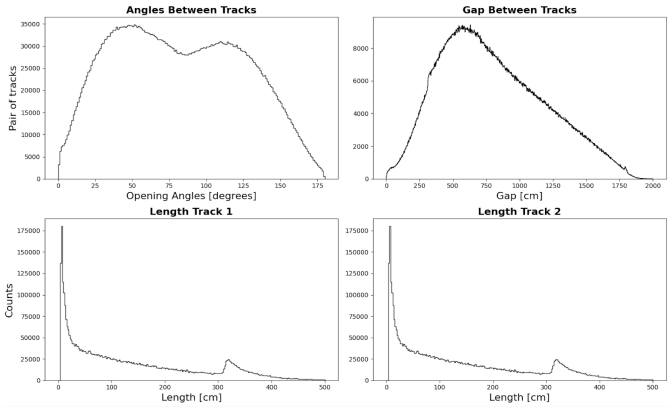

Figure 12. *Distribution of the variables of interest for other testing runs.*

A trend similar to that obtained for run 9435 is identified (although with greater statistics). Starting from an initial number of $N = 4,362,526$ events, the cuts are applied sequentially to understand the reduction in statistics. Specifically:

(a.) With the Gap selection only, the number of tracks that pass this selection is approximately 0.12%. An evident problem related to the spatial distribution of the tracks is identified, identifying all nearby tracks as split tracks leads to the error of evaluating positive match associated with different particles (e.g., tracks 22-29, *Figure* 13). It is evident that it is necessary to limit the observation to only parallel tracks (which most likely refer to the passage of the same particle): from the distribution of the angles, it is possible to highlight the consistency of the angle cut with what was found in run 9435..

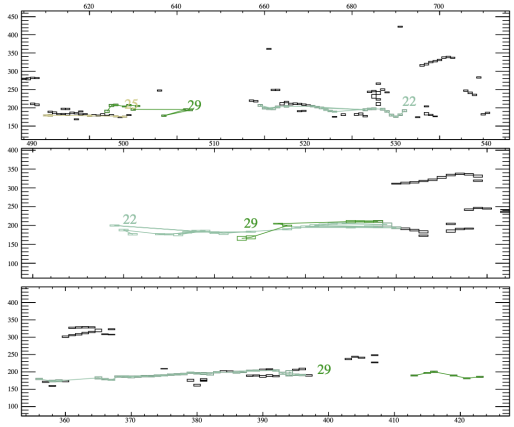

Figure 13. *Example of an event where the Gap-only selection leads to the incorrect identification of split track events.*

(b.) By making a further selection on the angle, the number of tracks that pass the selection is about 0.06%. It is possible to show how the tracks identified as split tracks are now parallel and close to each other, but the selection is still not satisfactory; as can be observed (e.g., tracks 30-34, *Figure* 14), there are many tracks that are too short compared to the gap separating them, leading to an incorrect estimation of their characteristics: this suggests implementing an additional cut on the length (consistent with the one performed for run 9435).

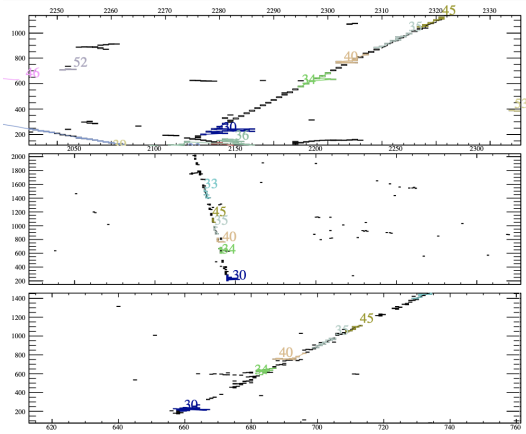

Figure 14. *Example of event where the Gap+Angle selection leads to the incorrect identification of split track events.*

(c.) Following the subsequent selection on the length of the tracks, pairs are identified that are indeed split tracks (e.g., tracks 3-4, *Figure* 15). The number of tracks that pass the selection is about 0.007%, perfectly in agreement with the percentage of split tracks present in run 9435.

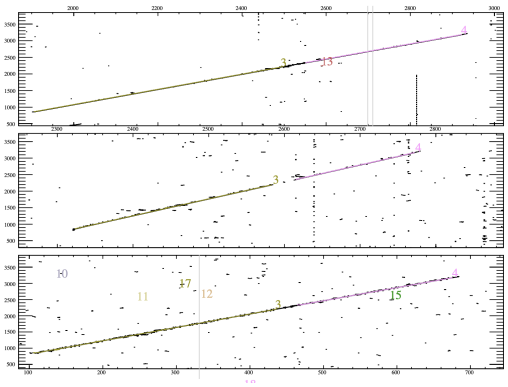

Figure 15. *Example of an event where the identification of a split track phenomenon occurs successfully.*

For an overview, *Figure* 16 shows a comparison of the variable trends at each stage of the selection

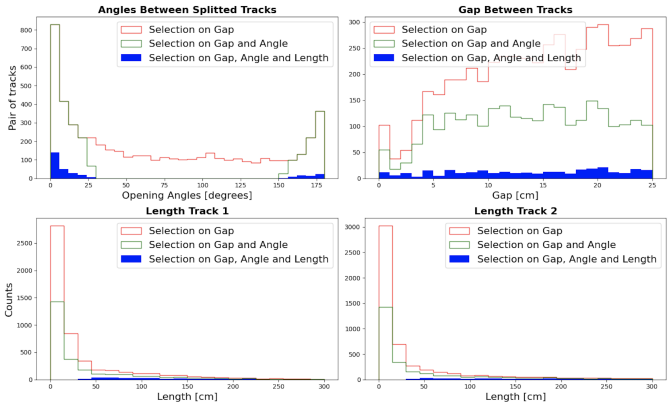

Figure 16. *Distribution of the observables of interest as the applied selection varies.*

Some observations can be made: first of all, by applying a selection only on the gap, the angle distribution flattens significantly in the central region and shows peaks at the extremes, reflecting the actual characteristic of split track events being close and parallel (thus with angles close to 0 or π). As the selection step increases, the statistics are significantly reduced, but the trend from the previous cuts is maintained: particularly in the case of track lengths, the final cut removes a large portion of events with short lengths that were not related to split tracks (as already justified in previous sections). It can thus be concluded that the plug-in has the ability to retrieve the variables of interest and identify split track events with a high degree of accuracy, allowing further analysis in the spatial distribution of split track events (*Section* 4).

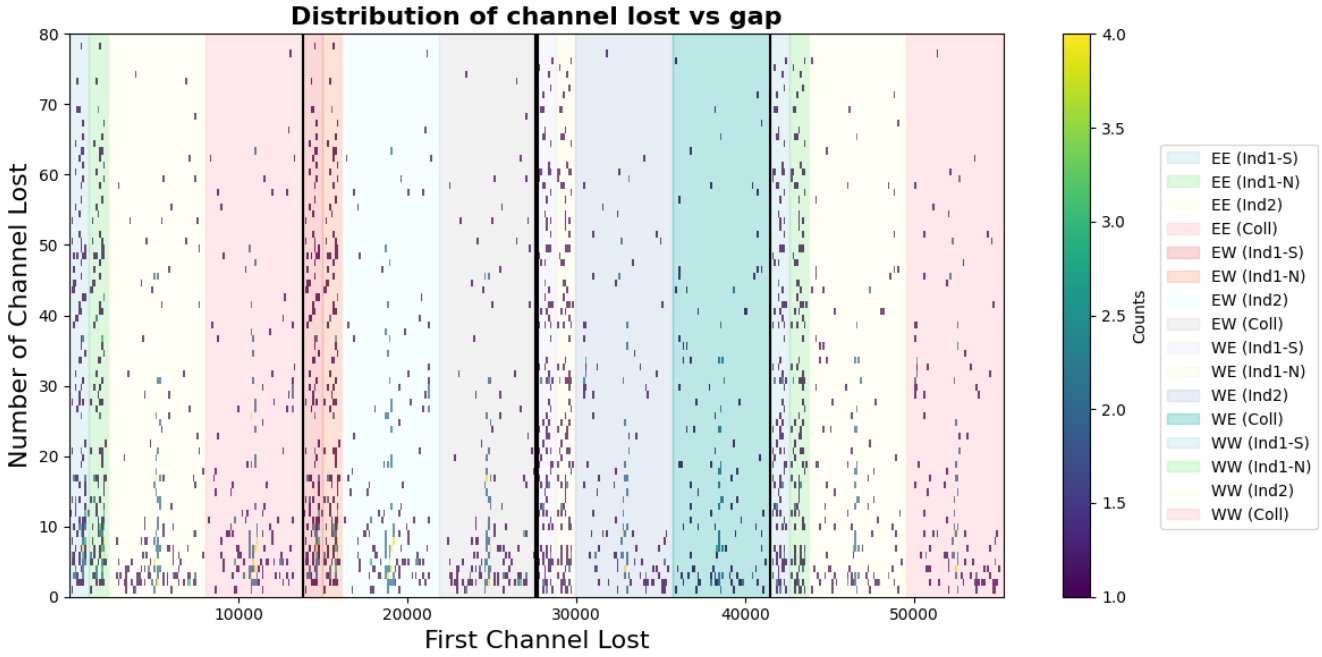

Figure 17. *Distribution of the number of lost channels as a function of the first channel that loses the event. The different colors indicate the different planes according to the correspondence channels-physical planes of the wires implemented in LArSoft.*

4. Spatial distribution of split track events

In this section, the spatial distribution of the split track events, identified with the plug-in written and presented in the previous sections, will be studied. The spatial distribution will be investigated by identifying the missed channels in the pairs of tracks identified as split tracks, and analyzing the distribution across the different wire planes (*Section* 4.1). This study will further serve as the basis for the implementation of missed channels in the HARPS tool (Hit Activity Removal from Particles for Systematics), developed by Bruce Howard and Harry Hausner. The goal of HARPS is to modify the events by removing certain hits, according to the channel lost distribution, and thus generate modified samples to study how reconstruction changes depending on the applied modification. It can work in two ways: drop context (i.e., keeping only a specific selected particle) or keep only context (i.e., selecting only a specific particle to perform systematic studies). In *Section* 4.2, the form of the plug-in, the changes made, and an application example will be presented.

4.1 Missed Channel distribution

The search for missed channels is implemented in the plugin described in the previous sections. Specifically, once a split track phenomenon is identified based on geometric characteristics, the nearest channels to the endpoint of the first track and the starting point of the second track are searched. At this point, two key pieces of information are stored: the first missed channel (i.e. the channel closest to the end of the first track) and the number of missed channels starting from it (i.e. the number of channels between the closest to the start of the second track and the closest to the end of the first). An important note is that the plugin applies a veto before storing these variables: to avoid storing channels from different cryostats, only the channels from the cryostat where there is correspondence with the passage of the track are recorded. Once these two variables are stored, it is possible to produce a two-dimensional plot, shown in *Figure* 17, to verify its trend.

The highlighted zones in the figure represent, according to the wire geometry described earlier and the correspondence with the channels, the three wire planes present in ICARUS: two Induction 1 planes (IND1-S and IND1-N, which are split at $z = 0$), the Induction 2 plane (IND2), and the Collection plane (COLL). The division naturally follows from knowing the correspondence between planes and wires (as described in [7]): knowing the first missed channel, it is possible to construct, for each TPC, four independent sets of Gap data (in terms of the number of missed channels) reflecting the division in the different planes. Then, to increase the statistics, the counts are summed for each TPC and cryostat.

From the distribution presented in *Figure* 17, a uniform distribution is evident for the first missed channel for each plane: this reflects the fact that there is no privileged location where the split track phenomenon occurs. The behavior of the missed channels, however, is more challenging to consider and is shown in *Figure* 18.

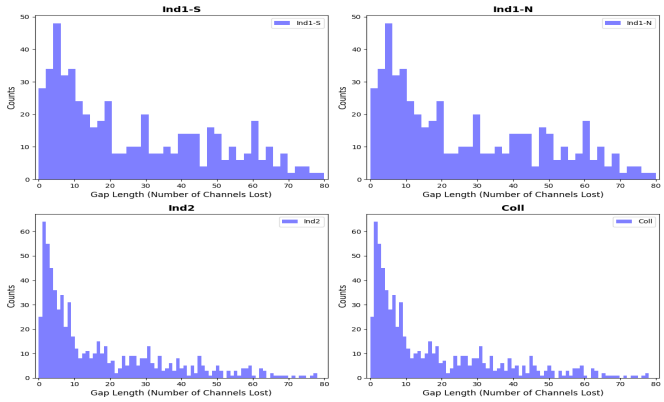

Figure 18. *Distribution of the number of lost channels, the distributions are divided according to the geometric plane of wires.*

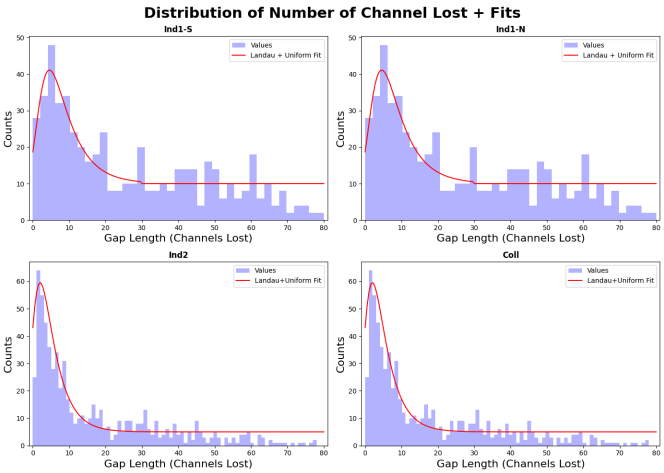

Figure 19. *Landau+Uniform fit on the distribution of the number of lost channels divided by geometric plane of wires.*

It is evident that the trend is significantly different for the IND1 planes compared to the IND2 and COLL planes. The model used to parameterize the distributions must necessarily consist of two parts:

- **Physical:** a component that takes into account the phenomena that cause the split track effect (e.g., a track crossing the cathode). This shape is parameterized as a Landau distribution, centered at $x_1 = 3$ for the IND1 planes and centered at $x_1 = 1$ for IND2 and COLL. As expected, the gaps are small and there is no significant difference in the distribution between the different planes, as the physical phenomenon occurs almost equally on each plane;
- **Reconstruction effect:** a component that accounts for the actual coherent noise experienced by the different planes. The wire planes, during the event reconstruction process, undergo an initial phase where coherent noise is removed; however, this removal could result in the loss of physical signal. For this reason, it is not necessary for the gap to be small, and moreover, this noise tends to occur when a track is aligned with the wires. This component is parameterized by a uniform distribution, which is found to have significantly varying levels between the planes. In particular, the IND1 planes have a much higher noise component than the IND2 and COLL planes. This is predictable considering that the IND1 planes are the outermost planes and are therefore subject to noise from other events much more noticeably than the subsequent planes.

Once the data trend is parameterized, a fit is performed by summing the two distributions. The obtained result is shown in *Figure* 19. At this point, all the necessary data are available to implement the hit removal with HARPS in a simulated sample (implemented in *section* 4.2): we know that each plane has a uniform probability distribution for losing a channel and that starting from it, depending on which plane loses the event, N channels will be lost, following a probability distribution given by the sum of a Landau and a uniform background.

4.2 HARPS Tool

As introduced in the previous section, the HARPS module aims to modify the events by removing hits from the tracks based on known probability distributions. The selection of tracks from which hits are removed is not random: the code requires a .txt file as input, which contains information about the particle IDs for which we want to identify the tracks and remove the hits based on the aforementioned probabilities. In this way, the code allows for the separate study of particles interacting with the detector and the quantification, by varying parameters, of the effect of the split track phenomenon on the reconstruction of neutrino physics events. The plug-in presented in the previous sections, aimed at identifying and characterizing split track phenomena in a sample, can therefore be used in synergy with the HARPS tool: by applying the module to a simulated dataset with Monte Carlo, a dataset with a known abundance of split tracks is generated, and by using the plug-in, it is possible to derive the characteristics of the events and quantify the impact of split tracks on the reconstruction.

The first modification made is related to the management of the .txt input file, where if the file is empty or non-existent, all the tracks in the sample are considered. The second modification, not fully implemented and tested on a sample, concerns the way hits are selected for removal from a given track: the first hit is chosen randomly, and starting from this hit, the next N hits are removed following a Landau distribution with a uniform background. The characteristics of these two distributions were described in the previous section, and the code must take into account that the initial parameters vary significantly depending on the wire plane that sees the track. Additionally, and very importantly, after the random selection of the channel and the subsequent removal of the N following hits, a veto condition is applied to ensure that the final length of the pair of segments is greater than $40cm$, as previously presented. A visual example is shown in *Figure* 20.

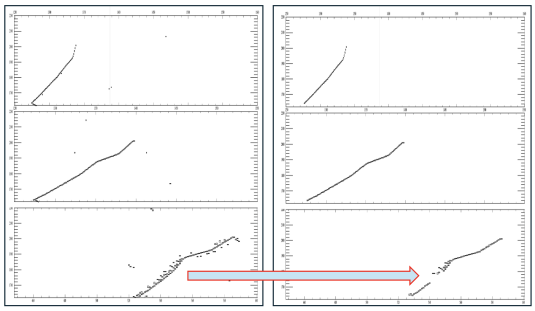

Figure 20. *Example of hits removal using the HARPS tool on an event, displayed with the LArSoft event viewer.*

In this figure, an example of a track recorded on the three wire planes is shown, along with the removal of a cluster of hits on one of the planes. By subsequently applying the plug-in to the dataset following the removal, it is possible, in addition to further validating its selection capabilities, to obtain information regarding the artificially generated Split Track and compare its characteristics with the dataset before the HARPS tool was applied: the ability to select which particle to focus on and vary the probability distribution according to which tracks are removed represents a highly important tool for understanding and quantifying the impact of the Split Track phenomenon on neutrino event reconstruction.

5. Conclusion

In conclusion, the work conducted in this study can be summarized as follows:

- The importance of accurate event reconstruction was understood, and several potential issues were highlighted, with particular focus on the phenomenon of Split Tracks;
- A versatile and optimized plug-in in ART was developed, which can be applied to any dataset to identify Split Track events based on geometric characteristics and store their main observables: angle between tracks, gap, missed channels, and segment lengths;
- The geometric distribution of Split Track events in the detector was studied, identifying the distribution followed by the number of missed channels. Using this, the result was implemented into the HARPS module.

Possible future developments of this work include the implementation of additional methods for selecting split track events (e.g., energy-based selections) and the validation of the plug-in's performance on further runs. Moreover, a more in-depth study of the missed channel distribution, followed by its implementation into the HARPS module, could pave the way for more precise quantifications and direct interventions to address the systematic issue of split tracks in the reconstruction of neutrino events in ICARUS.

Appendix

Calculation of the angle between non-parallel tracks

The angle between the two tracks is obtained through three main steps:

1. **Searching the points of closest approach:** These points are obtained as

$$Closest1 = endPos1 + avgDir1 \cdot t_1$$

$$Closest2 = startPos2 + avgDir2 \cdot t_2$$

where t_1 and t_2 are obtained by minimizing the distance between the two tracks (for the explicit calculations, see [9]).

2. **Searching for midpoint M:** The midpoint is calculated as

$$M = closest1 + \frac{1}{2} \cdot closestSep$$

where $closestSep = closest2 - closest1$ is the vector connecting the points of minimum distance.

3. **Searching for the angle:** The angle is then obtained starting from the dot product of the vectors pointing from the end of the first track to the midpoint and from the midpoint to the beginning of the second track, in formulas:

$$\theta = arccos\left(\frac{startToMid \cdot endToMid}{||startToMid|| \cdot ||endToMid||}\right)$$

Acknowledgments

I would like to extend my profound thanks to the Organizing Committee of the Italian Student Program at Fermilab, especially Giorgio Bellettini, Simone Donati, Emanuela Barzi and Marco Mambelli for granting me the exceptional opportunity to participate. Being selected was a true honor and a defining moment in my academic path.

I am deeply grateful to my supervisor, Harry Hausner, for his continuous guidance and support at all hours of the day. Your invaluable advice has been instrumental in the success of this project, and your expertise has been a constant source of inspiration during challenging moments.

I would also like to express my heartfelt thanks to Angela Fava for her numerous suggestions and for the important discussions that have helped guide me towards achieving the goals of the project. My sincere thanks also go to the ICARUS Collaboration for fostering a collaborative research environment.

References

[1] **P. Abratenko et Al.** ICARUS at the Fermilab Short-Baseline Neutrino Program - Initial Operation. *arXiv:2301.08634v1*, 2023.

[2] **The ICARUS-WA104, LAr1-ND, MicroBooNE Collaborations**. A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam. *arXiv:1503.01520v1*, 2015.

[3] **R. Triozzi**. Studies of the trigger performance of the ICARUS T600 detector at Fermilab. *Fermi National Accelerator Laboratory Internship Report*, 2022.

[4] **M. Diwan et Al.** A novel method for event reconstruction in Liquid Argon Time Projection Chamber. *Journal of Physics: Conference Series 762, 012033*, 2016.

[5] **M. Artero Pons**. Study of the reconstruction of $\nu\mu$CC QE events from the Booster Neutrino Beam with the ICARUS detector. *Philosophiae Doctoral Course Thesis in Physics (Università degli studi di Padova)*, 2022/2023.

[6] **R. A. Jaimes Campos**. Cosmic Ray Tagging in ICARUS: a combined analysis. *Master Thesis in Physics (Alma Mater Studiorum)*, 2023/2024.

[7] **B. Behera et Al.** ICARUS TPC Channel Mapping. *SBN DocDB 25069*, 2022.

[8] **M. Scarnera**. Improving track reconstruction in the ICARUS detector. *Master Thesis in Physics (Alma Mater Studiorum)*, 2023/2024.

[9] **Ronald Goldman**. Intersection of Two Lines in Three Space. *Graphics Gems (p. 304)*, 1995.