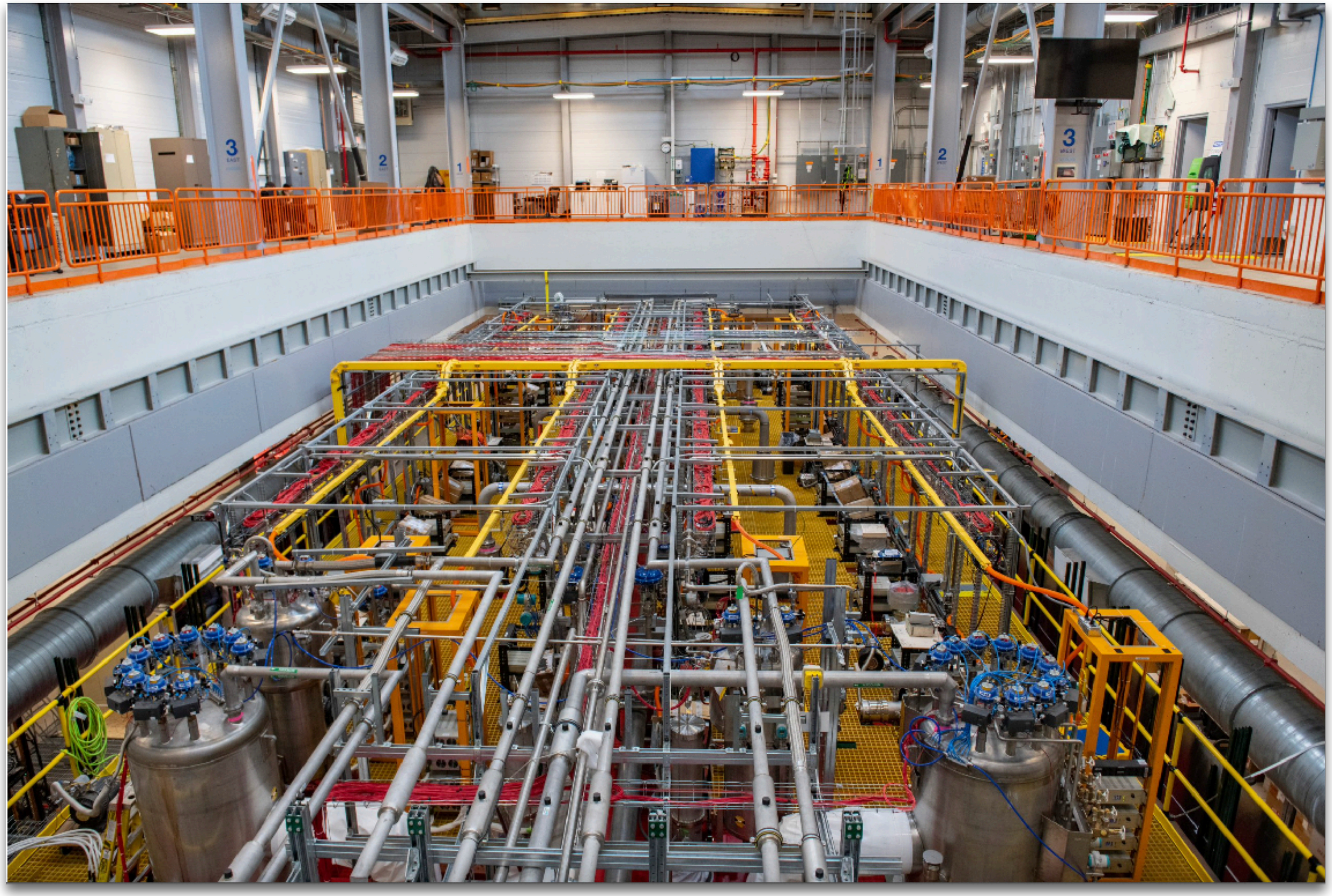


Track reconstruction in liquid Argon TPC experiments

Alice Campani
on behalf of the ICARUS collaboration

Università di Genova

INFN
Istituto Nazionale di Fisica Nucleare

Overview of the content this talk

- Introduction to **LArTPC experiments** and SBN physics program
- General description of **TPC event reconstruction chain** and main steps
- Two *parallel* event reconstruction paths:
  - *Pandora-based* event reconstruction:
    overview of the hierarchy, insights on the main stages
  - *Machine Learning*- (ML) based event reconstruction:
    overview of the full reconstruction chain
- Conclusions and perspectives

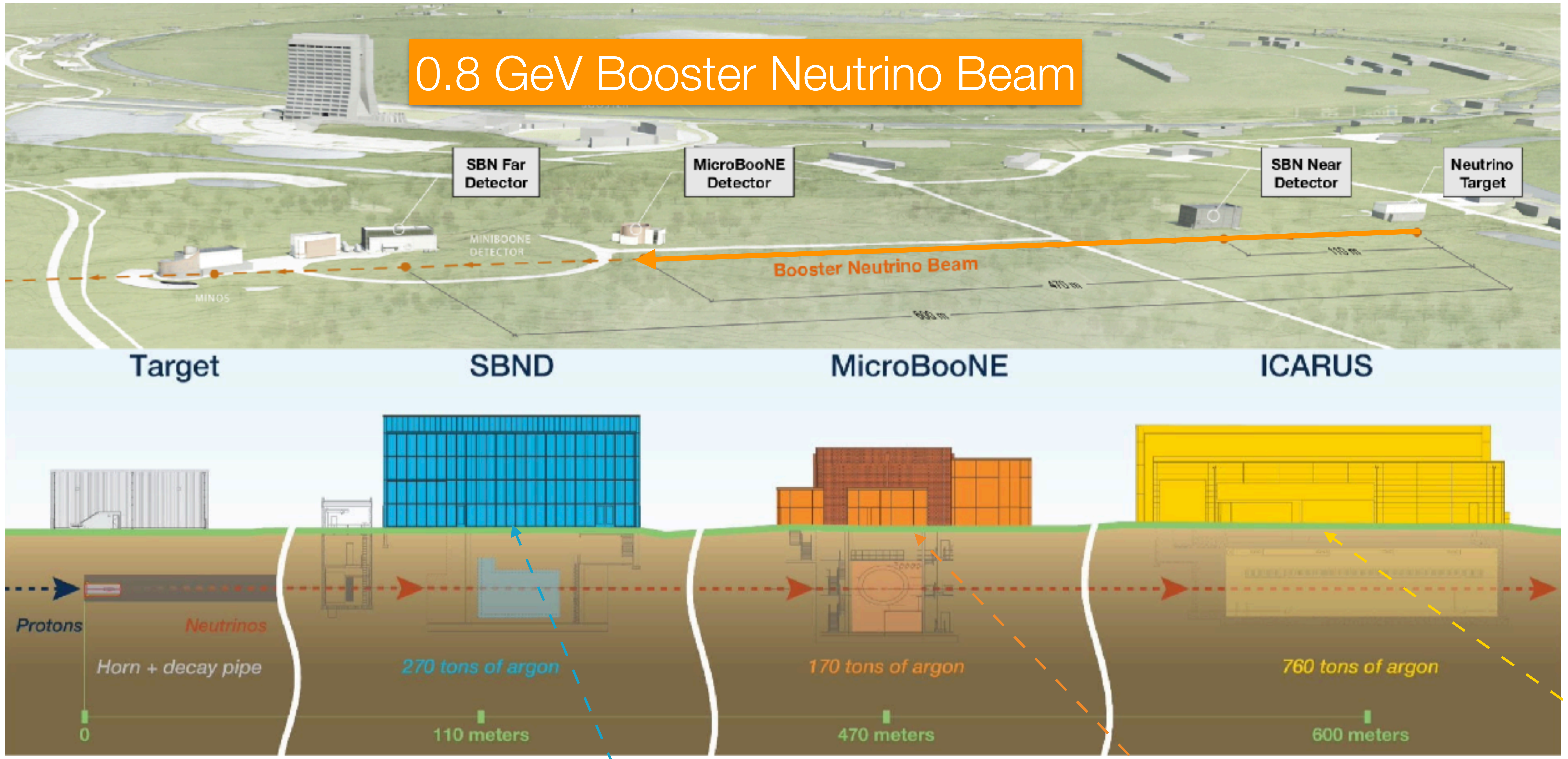
Imaging Cosmic And Rare Underground Signals

The Short Baseline Neutrino (SBN) program

Precision search for **1 eV mass scale sterile ν** to confirm/rule out previous anomalies from past experiments

Sensitive searches for ν_μ **disappearance**, ν_e **appearance**

ICARUS exposed also to **NuMI beam** (6 degrees off axis)



High statistics measurement of **ν-Argon** cross sections for **DUNE**

Search for Beyond Standard Model (**BSM**) physics

Same detector technology to reduce **systematics** and increase sensitivity

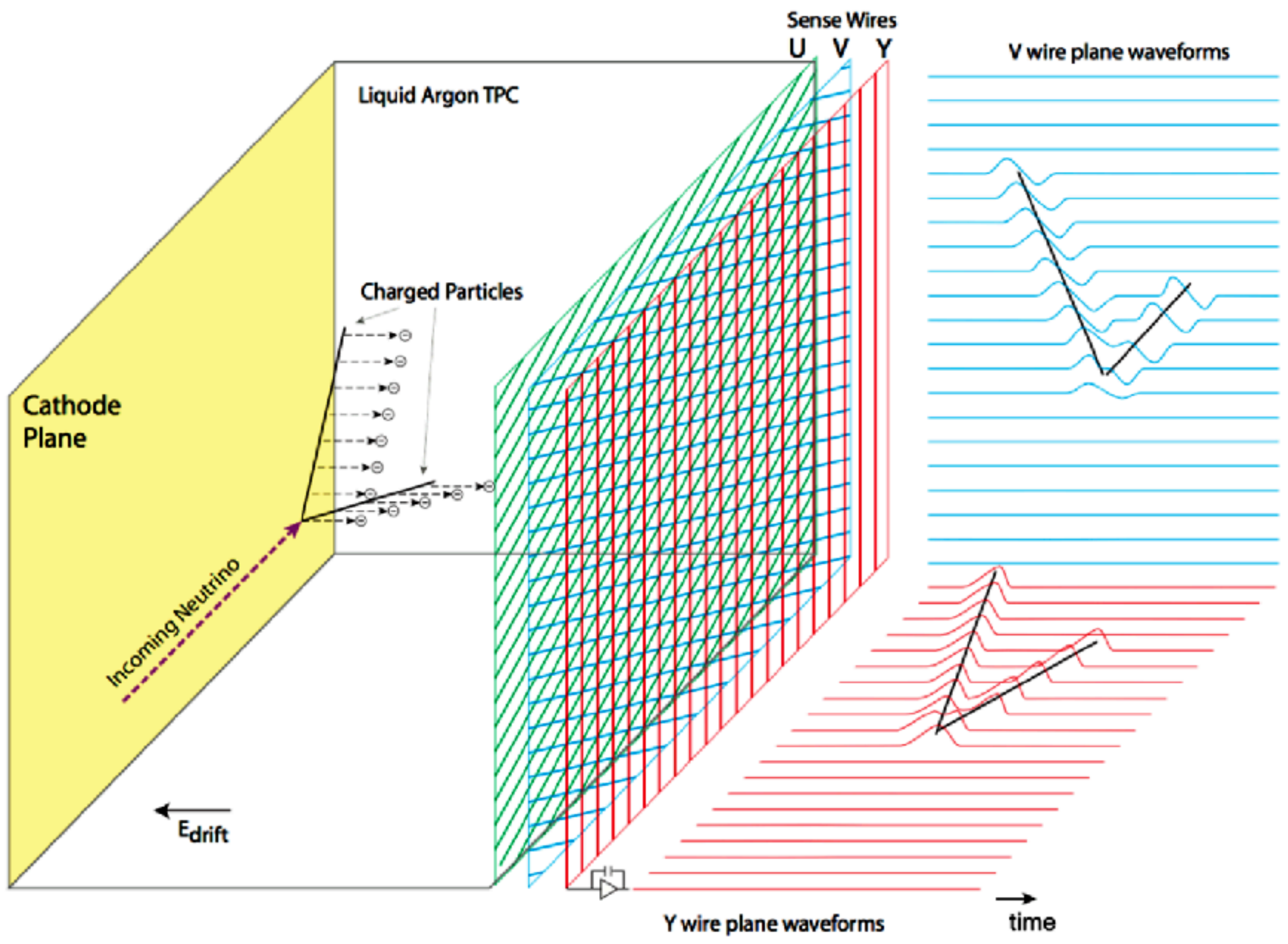
SBND
Near detector

MicroBooNE

ICARUS
Far detector

Liquid Argon Time Projection Chambers (LArTPCs)

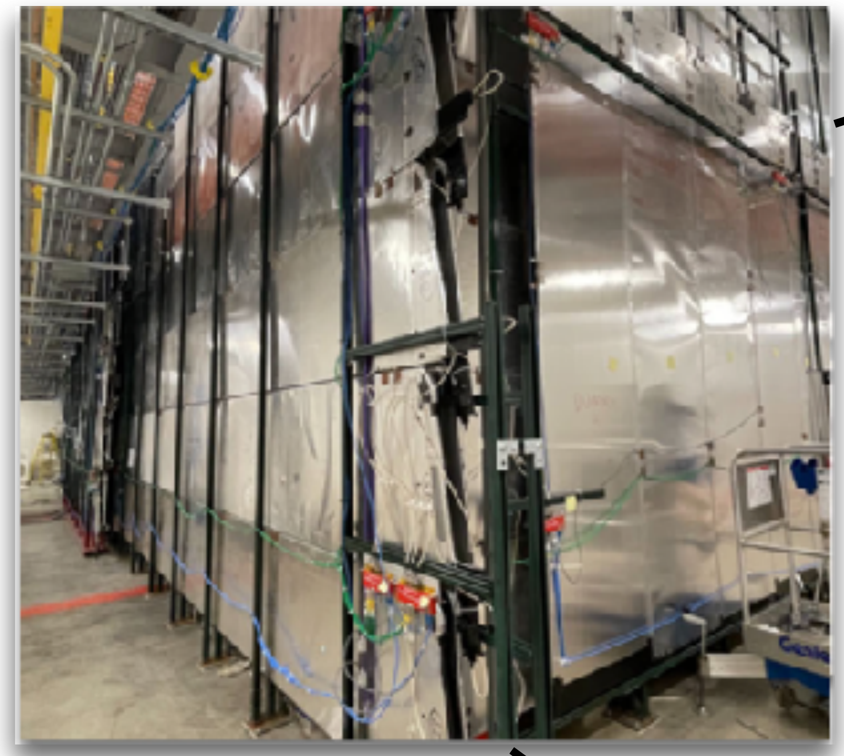
- Proposed by **C. Rubbia** in **1977**, LArTPCs are high granularity, continuously sensitive, self-triggering detectors
  - **Dense medium:** high rate of ν interactions
  - 2/3 **wire planes** (3-5 mm wire pitch) with different orientation to generate **2D views** of particle tracks
  - 3D **imaging** with **mm-scale resolution**
  - **Calorimetric** reconstruction capabilities
  - **Scalable** to large detector volumes $\mathrm{O}(10)\ \mathrm{kton}$



Ideal for ν interaction studies in a wide energy range

Typical LArTPC detector components: ICARUS detector as example

Two **identical cryostats** (3.6 x 3.9 x 19.6 m^3) housing two TPCs each, 760 tons of ultra pure liquid argon for a total **active mass** of **470 ton**

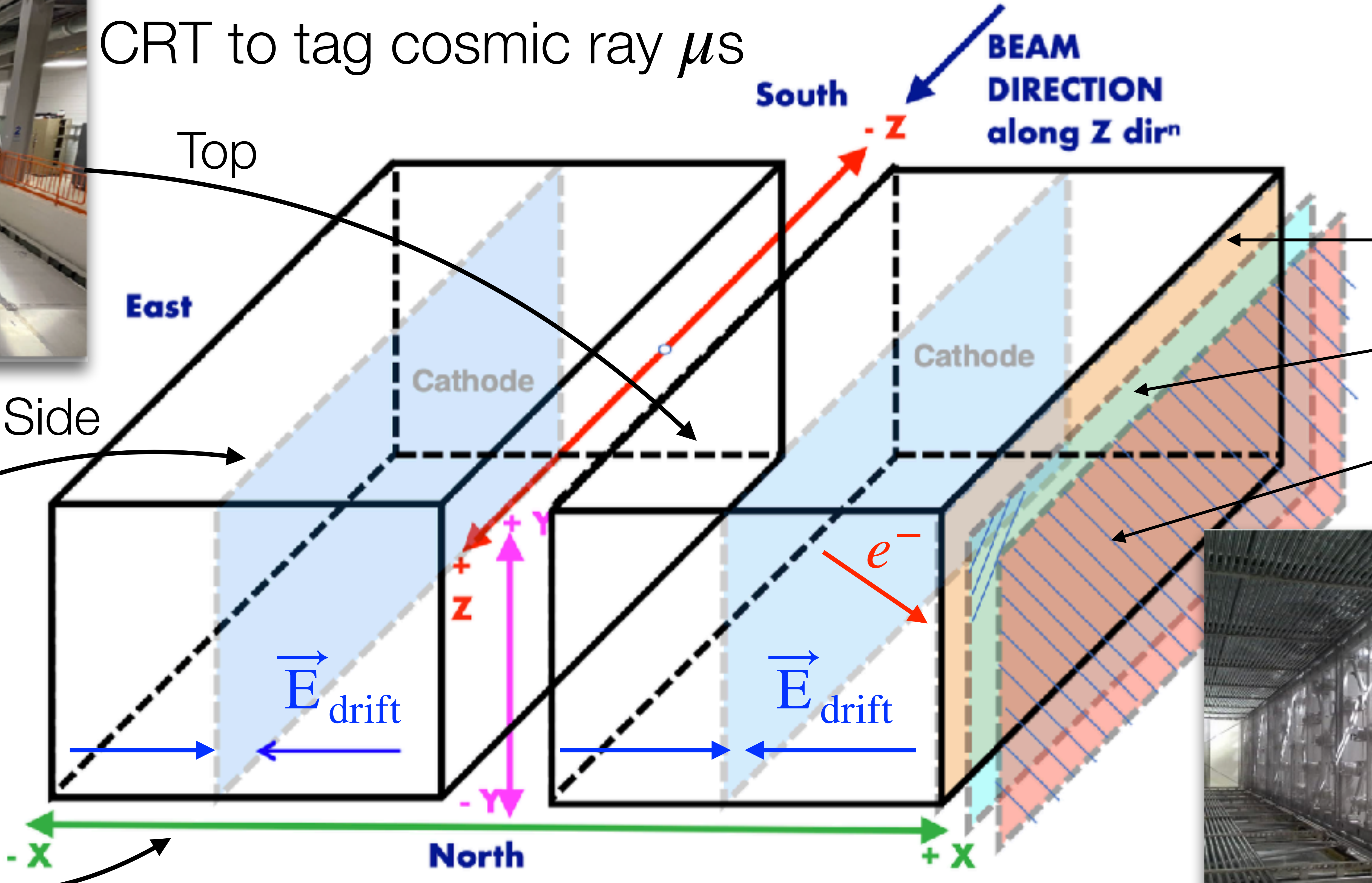


CRT to tag cosmic ray μs

Top

Side

Bottom



$E_{drift} = 500$ V/cm, $t_{drift} \sim 1$ ms

Ionization charge read by **3 wire planes** with different orientation:

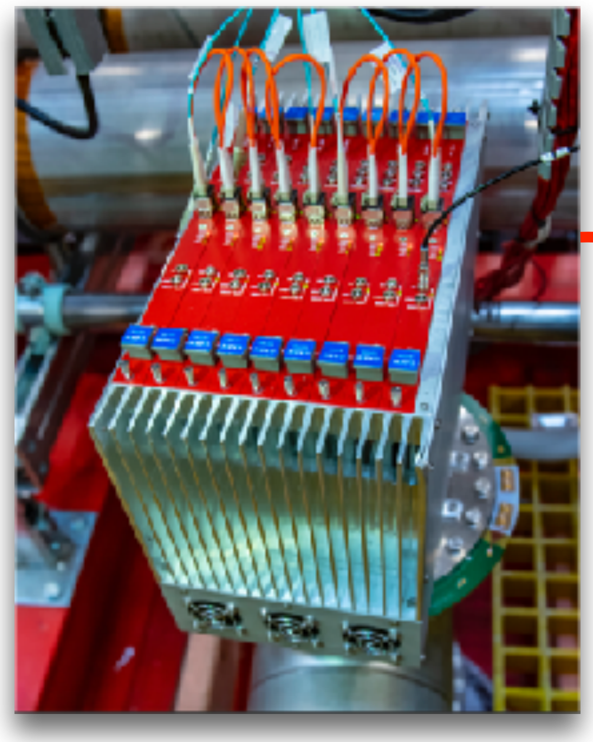
Induction1 (0°)

Induction 2 (+60°)

Collection (-60°)

360 **PMTs** behind the wires to collect scintillation light and trigger events

Event reconstruction in LAr TPCs: ICARUS reconstruction chain



Data

Unpack the data and turn it into a **raw waveform**

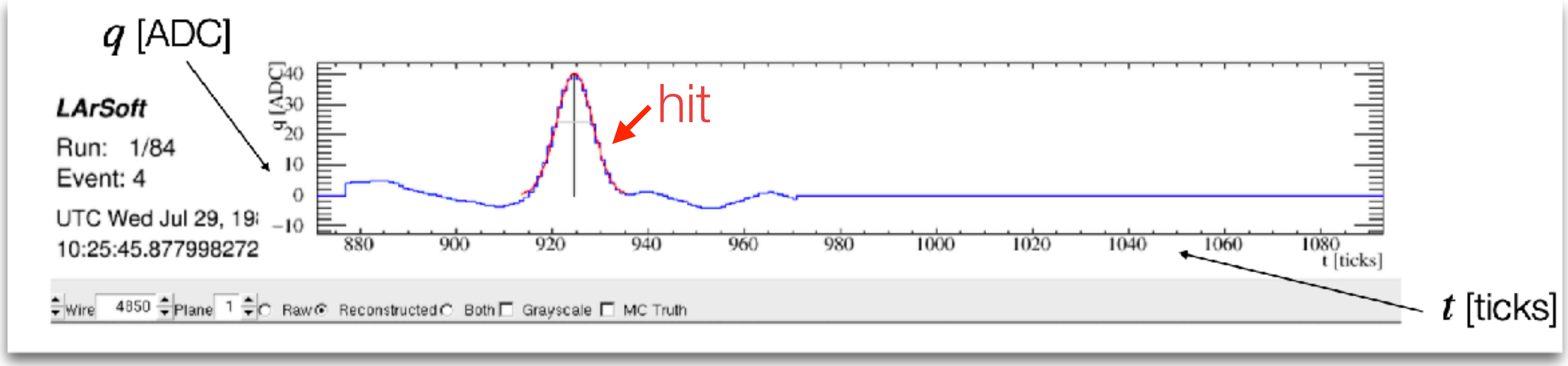
Threshold-based algorithm to identify regions containing *hits*, i.e. segments of waveforms corresponding to signal.

Decoding → **Deconvolution** → **ROI Finder** → **Gauss hits**

- Removal of **coherent noise**
- **Deconvolution** to remove the $\vec{E}$ distortions and electronics shaping effects on wire signals

Fit each signal hit with **Gaussians**: the area is proportional to n_{e^-} drift electrons that generated that.

Example of deconvolved signal (charge vs time) on a single wire plane after ROI finding and Gaussian fit



Event reconstruction in LAr TPCs: ICARUS reconstruction chain

To reduce noise, keep only hits that are **consistent** with 3D points using 2 wire planes combinations.

Gauss hits

Cluster 3D

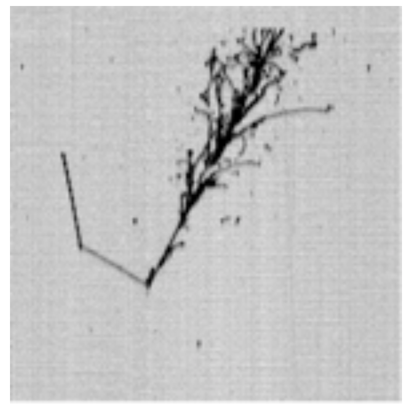
After **Cluster 3D** the reconstruction chain splits in **two parallel paths.**

Pandora-based reconstruction
Downstream reconstruction is based on cluster, *slice* (groups of tracks related to the same particle interaction) and pattern recognition.

Pandora reco

ML-reco

ML-based reconstruction
The full reconstruction chain is based on Machine Learning (ML) techniques



Analysis

Track and shower reconstruction, calorimetry

Signal processing: foreseen change from 1D to 2D deconvolution

- Wire signals are a convolution of **electric field** and **electronics responses**:

$$M(t) = \int_{-\infty}^{+\infty} R(t,t') \cdot S(t')dt$$

Measured signal

Response function

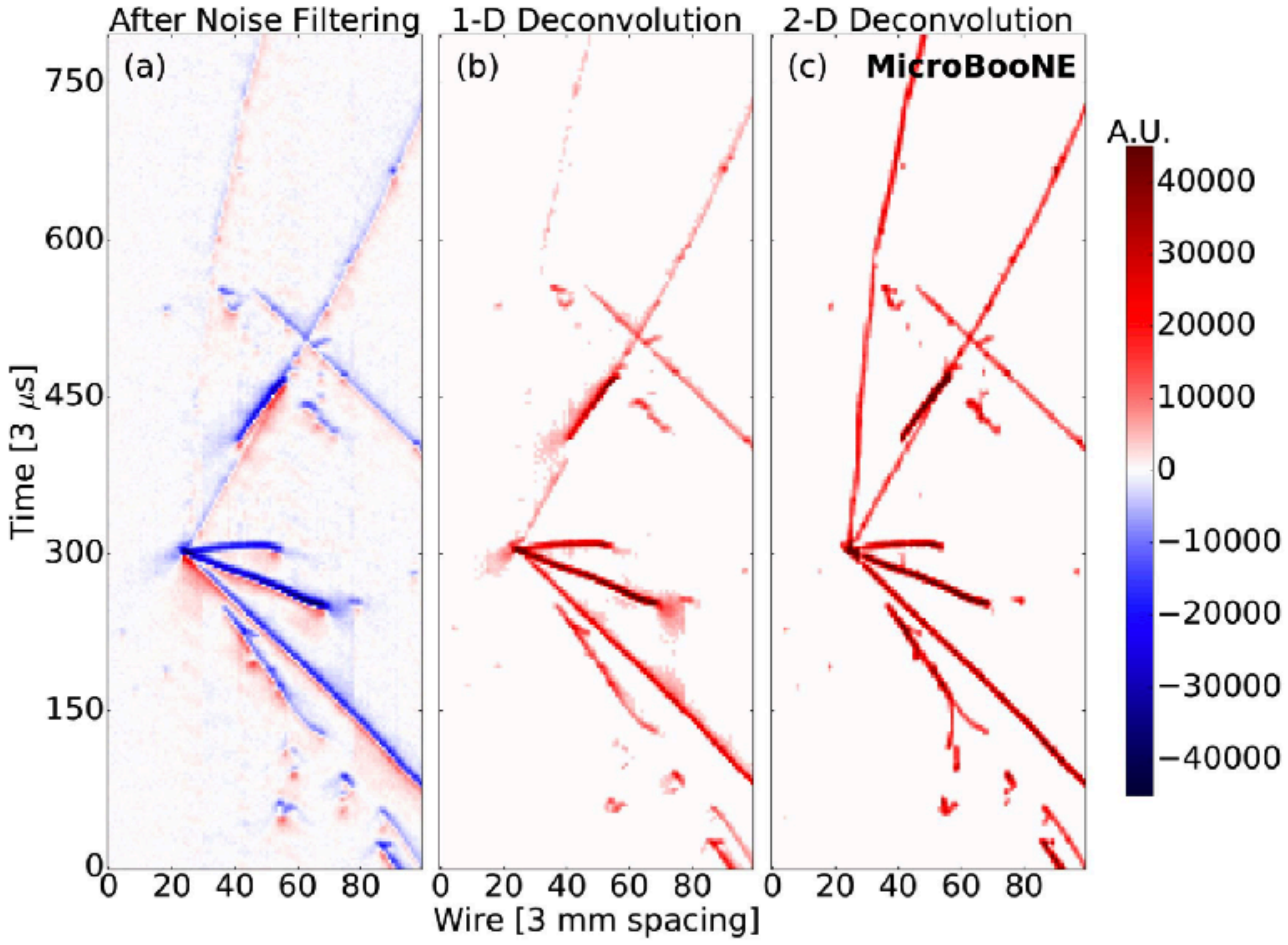
Original wire signal

- Original wire signal extracted with **1D deconvolution** after applying a filter for noise

- **2D deconvolution** to account for induced charge effects, i.e. charge drifting in nearby wire regions

- improvement of the charge resolution

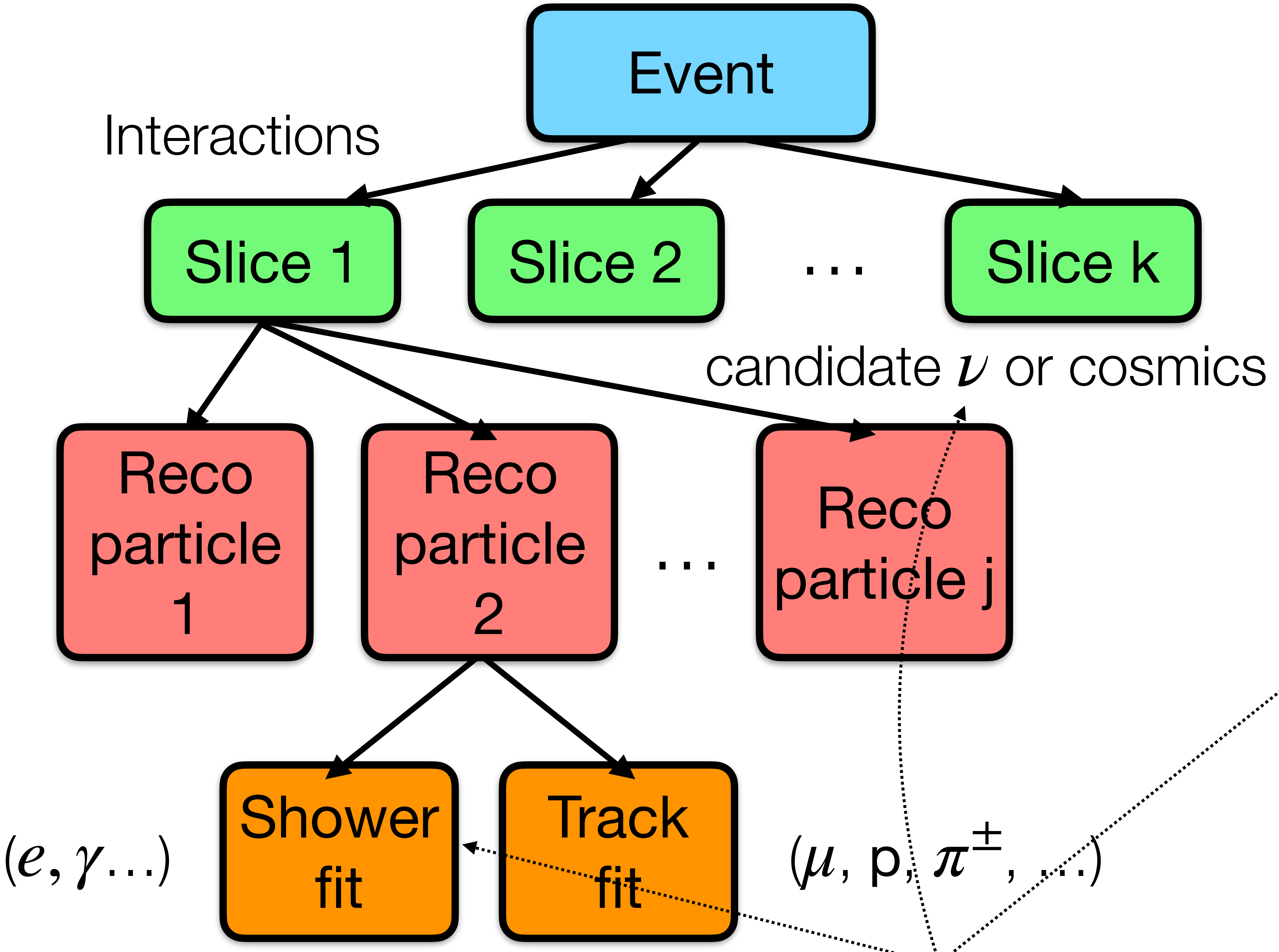
- higher ε on hits reconstruction for specific track classes

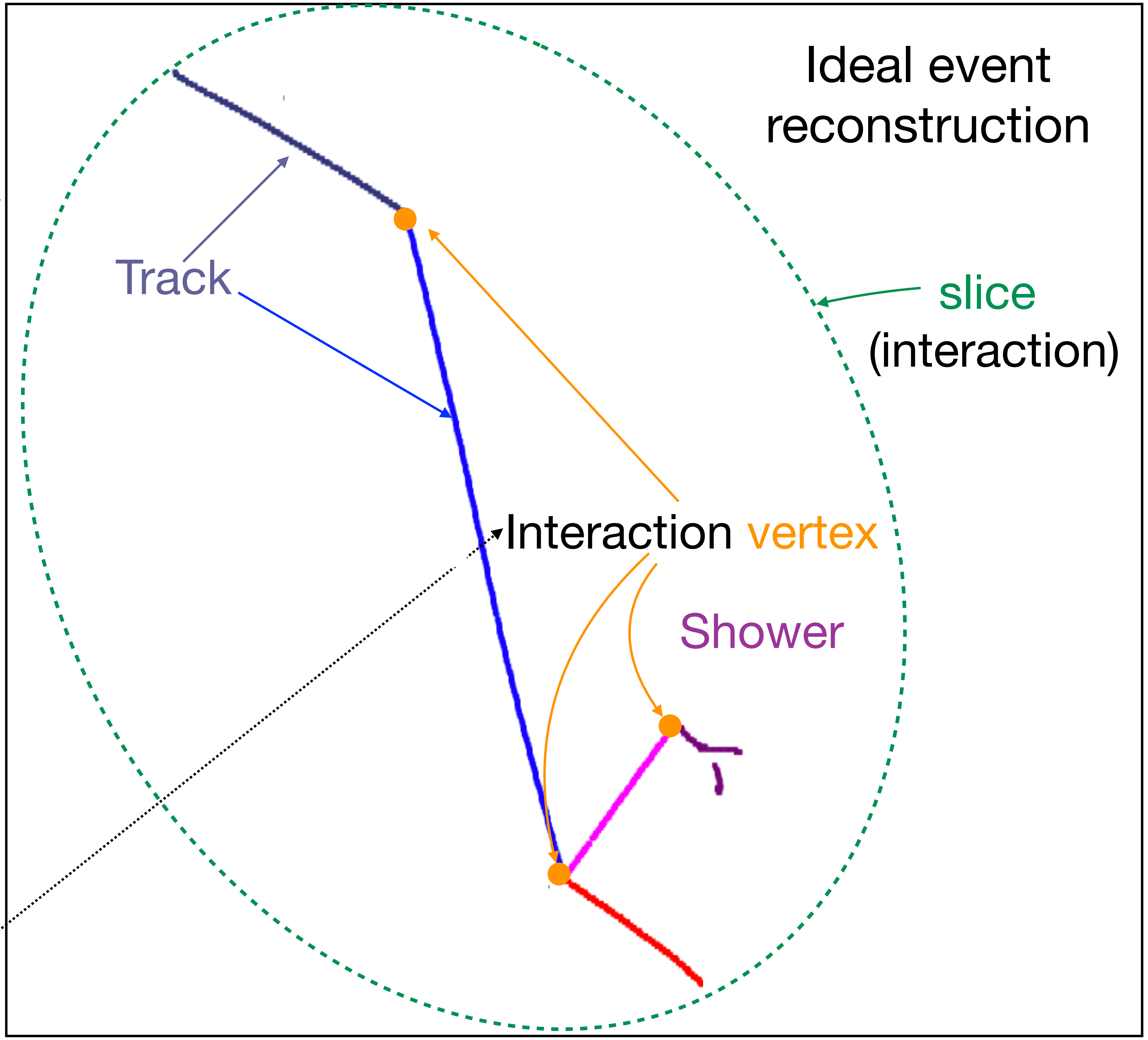


Example of charge recovery from MicroBooNE
Adams, C. et al. 2018 *JINST* 13 P07006 (ArXv link)

Pandora-based event reconstruction

- Multi-algorithm pattern-recognition software
- Goal: reconstruct **interaction hierarchies**





- https://github.com/PandoraPFA

Boosted Decision Trees (**BDTs**), i.e. ML algorithms, employed in three steps of the chain

Boosted Decision Tree (BDT)

We mentioned several places where Pandora uses this algorithm for the reconstruction.

- Idea: Identify a **signal** and a **background** class and a **set of input features** on which you expect there could be a good separation between them.

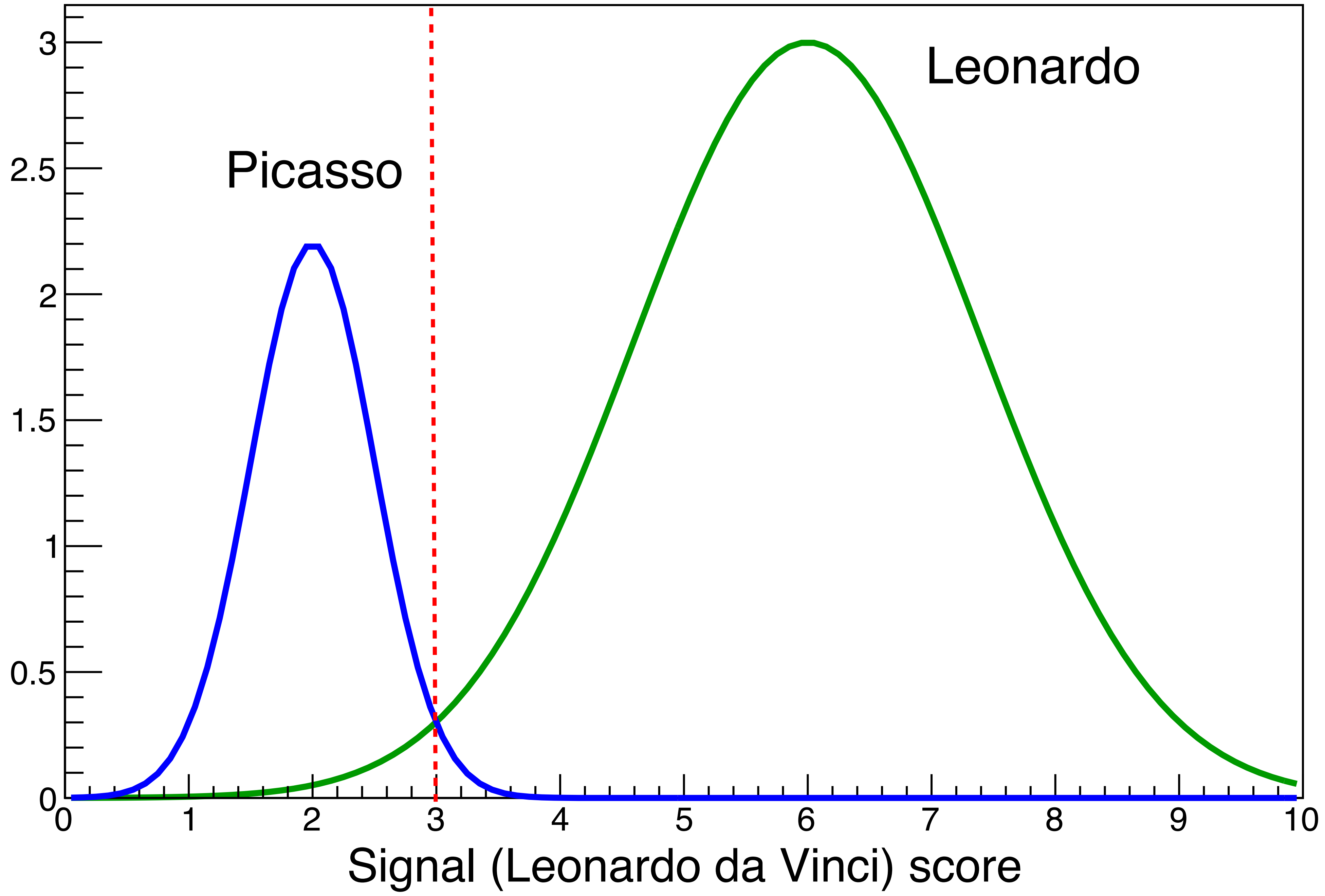
- Method: BDT is first **trained** on a sample where the true class is known and input features are used to have the power to distinguish between signal and background, then for a new sample with unknown class the same set of features is computed to define a **score** that quantifies how “signal-like" the sample is.

- Example:
  Signal: Leonardo da Vinci art work
  Background: Pablo Picasso art work (from the cubism period)
  Sample: a generic painting
  Input parameters: use of colors, light and shadow, presence of geometric shapes

Boosted Decision Tree (BDT)

- Example: Signal: Leonardo da Vinci art work
  Background: Pablo Picasso art work (from the cubism period)
  Sample: a generic painting
  Input parameters: use of colors, light and shadow, geometric shapes, …

Signal



Background

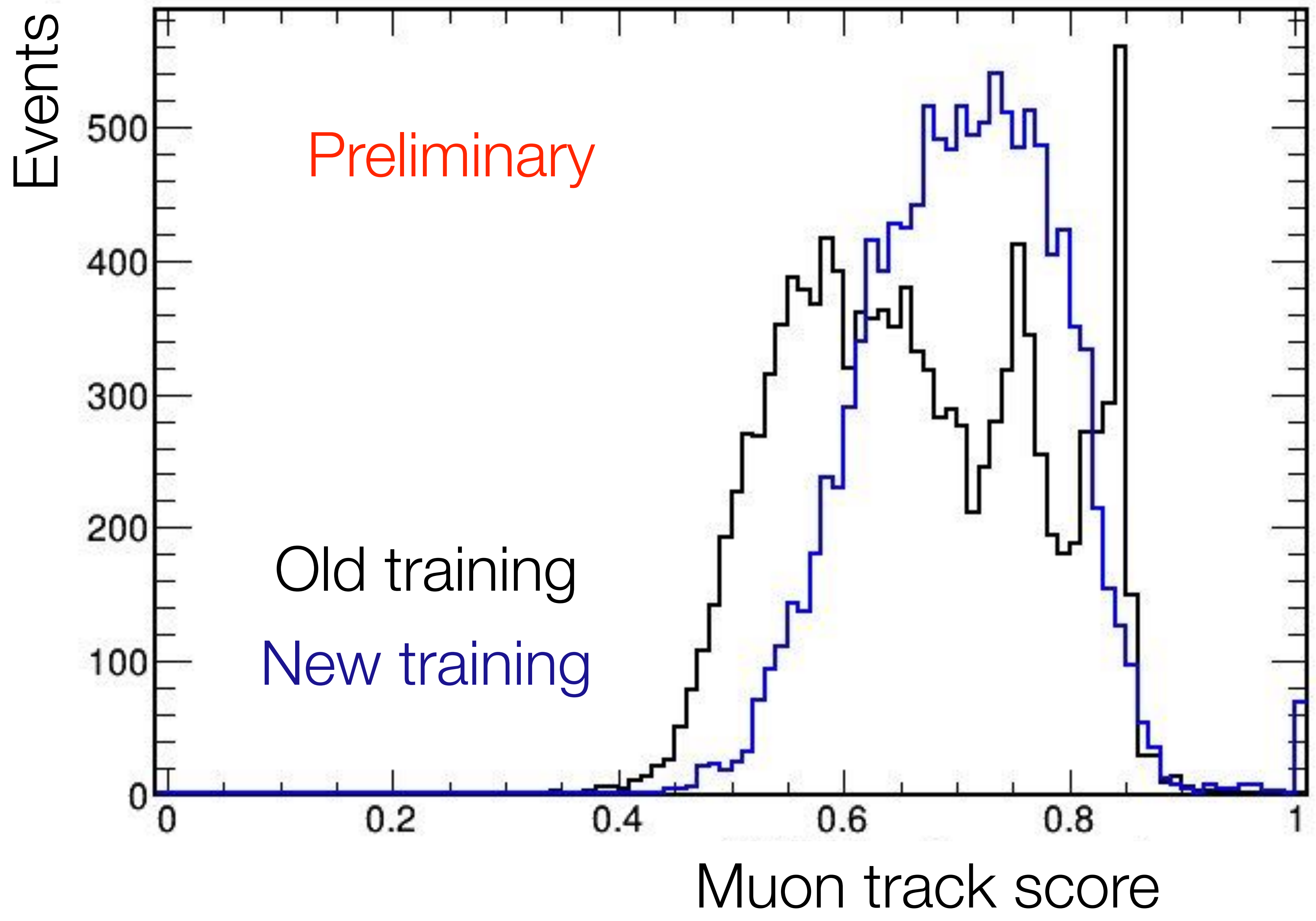
Pandora-based event reconstruction: new BDT training to discriminate tracks and showers

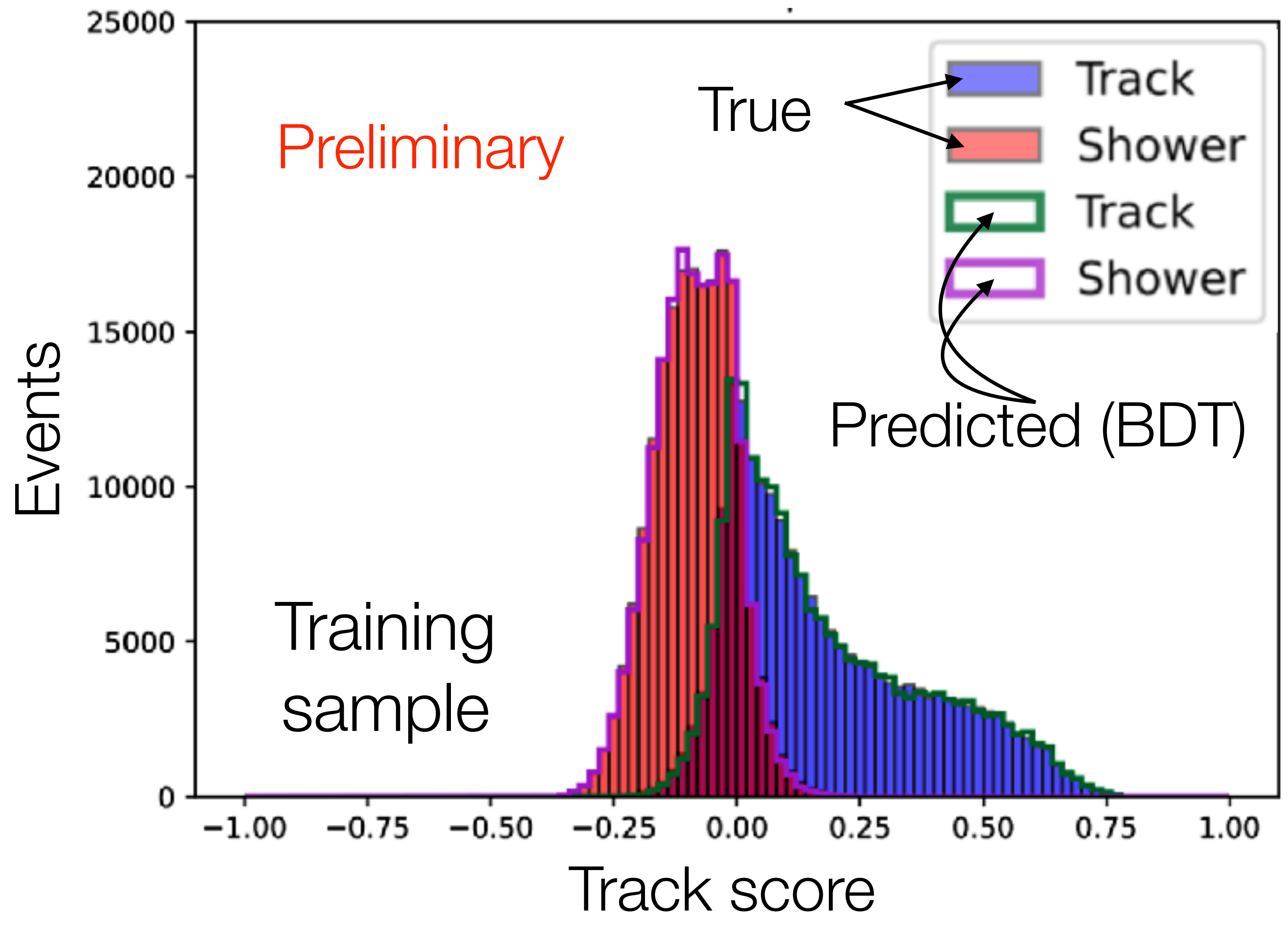
- Training based on 8 **geometrical** variables (5 **calorimetrical**) from the 3D coordinates (charge) of the hits

New training based on BNB ν-only MC

Preliminary

$\varepsilon_{classification}$	Old training	New training	Δ
BNB ν only	72.3%	80.3%	8.0%
NuMI ν only pre-tuning [*]	67.8%	79.9%	12.1%
NuMI ν only tuned [*]	66.7%	79.2%	12.5%

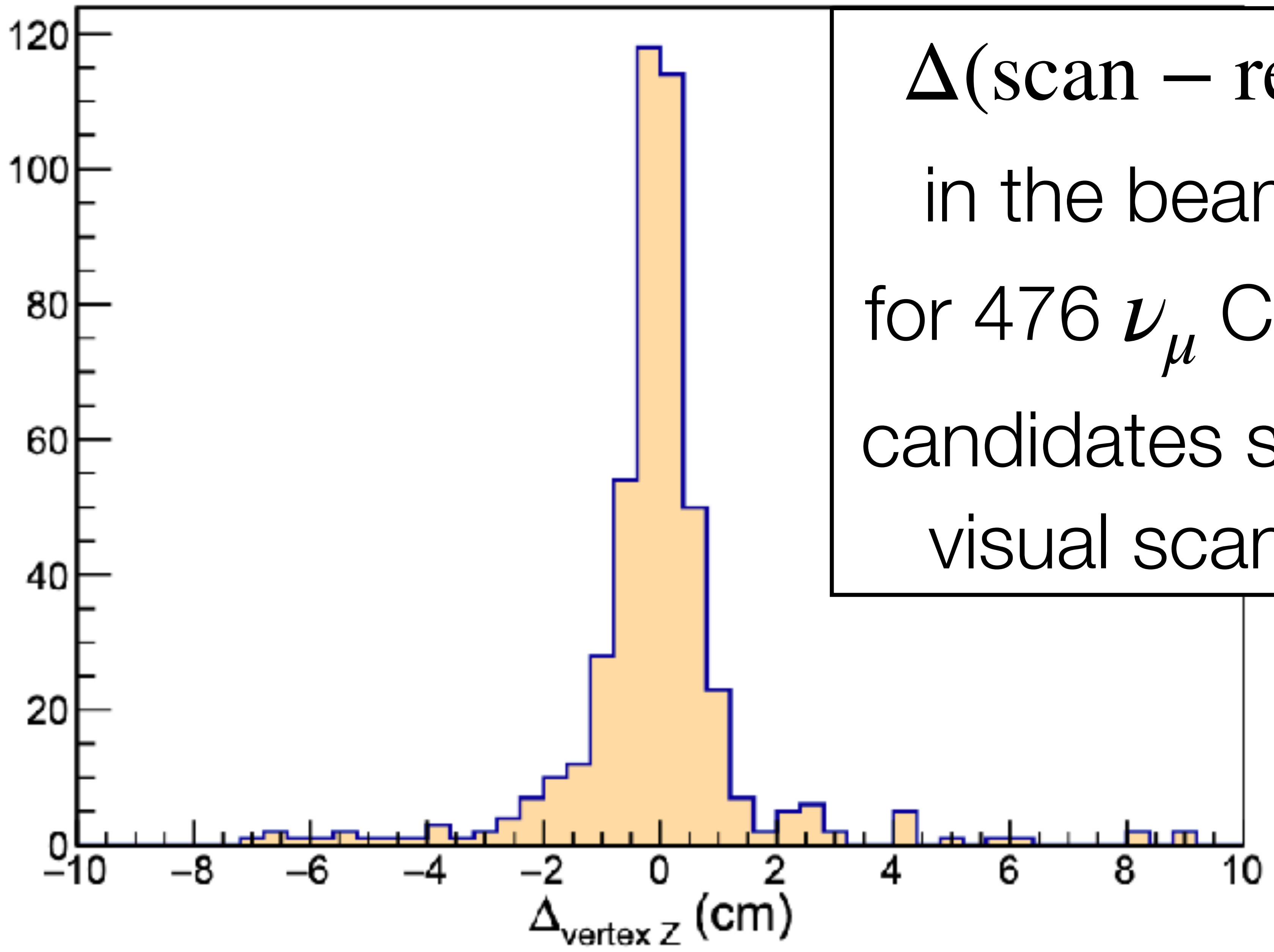




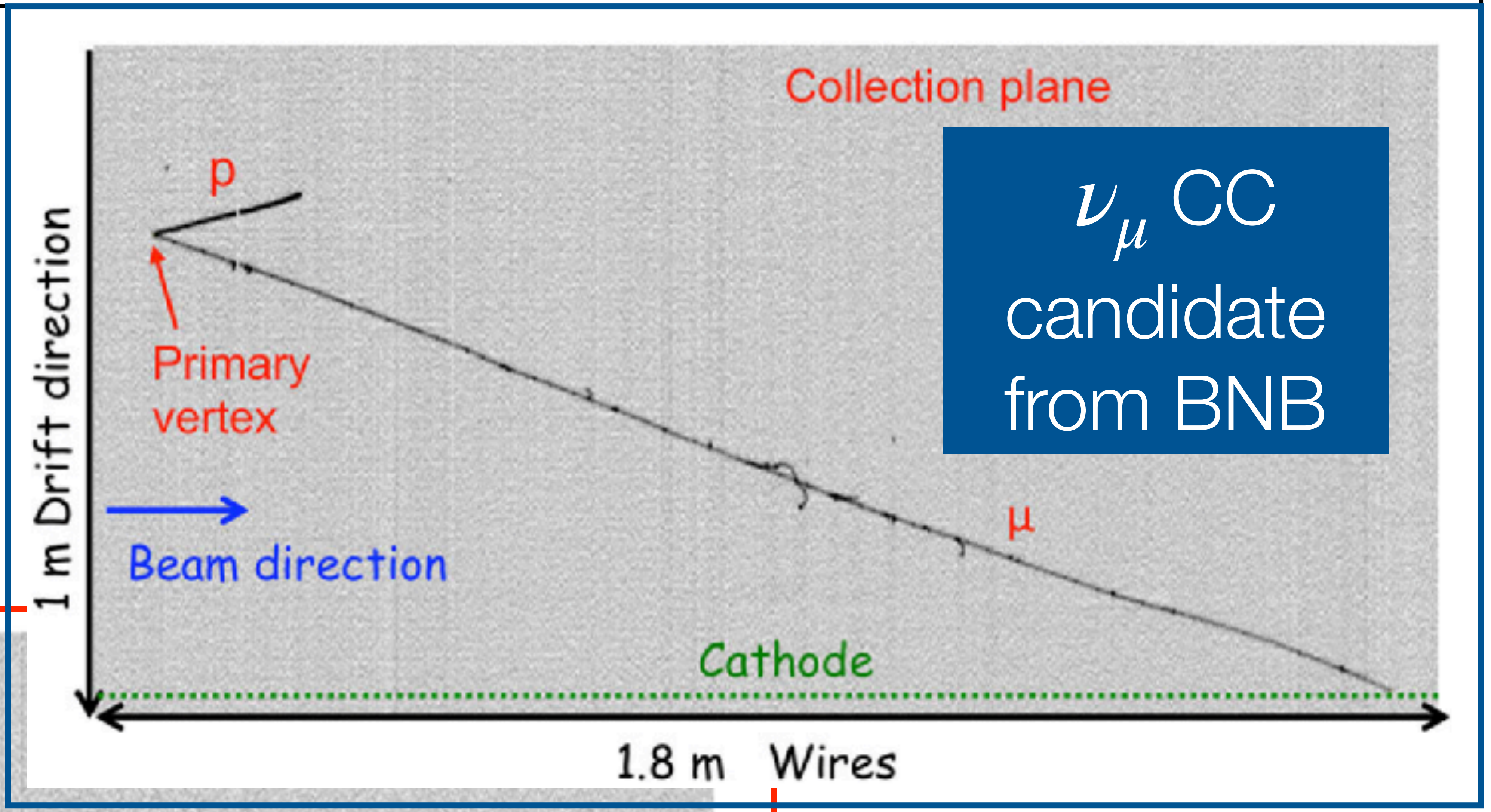
Pandora event reconstruction: visual scanning and data/MonteCarlo comparison to evaluate performance/improvements

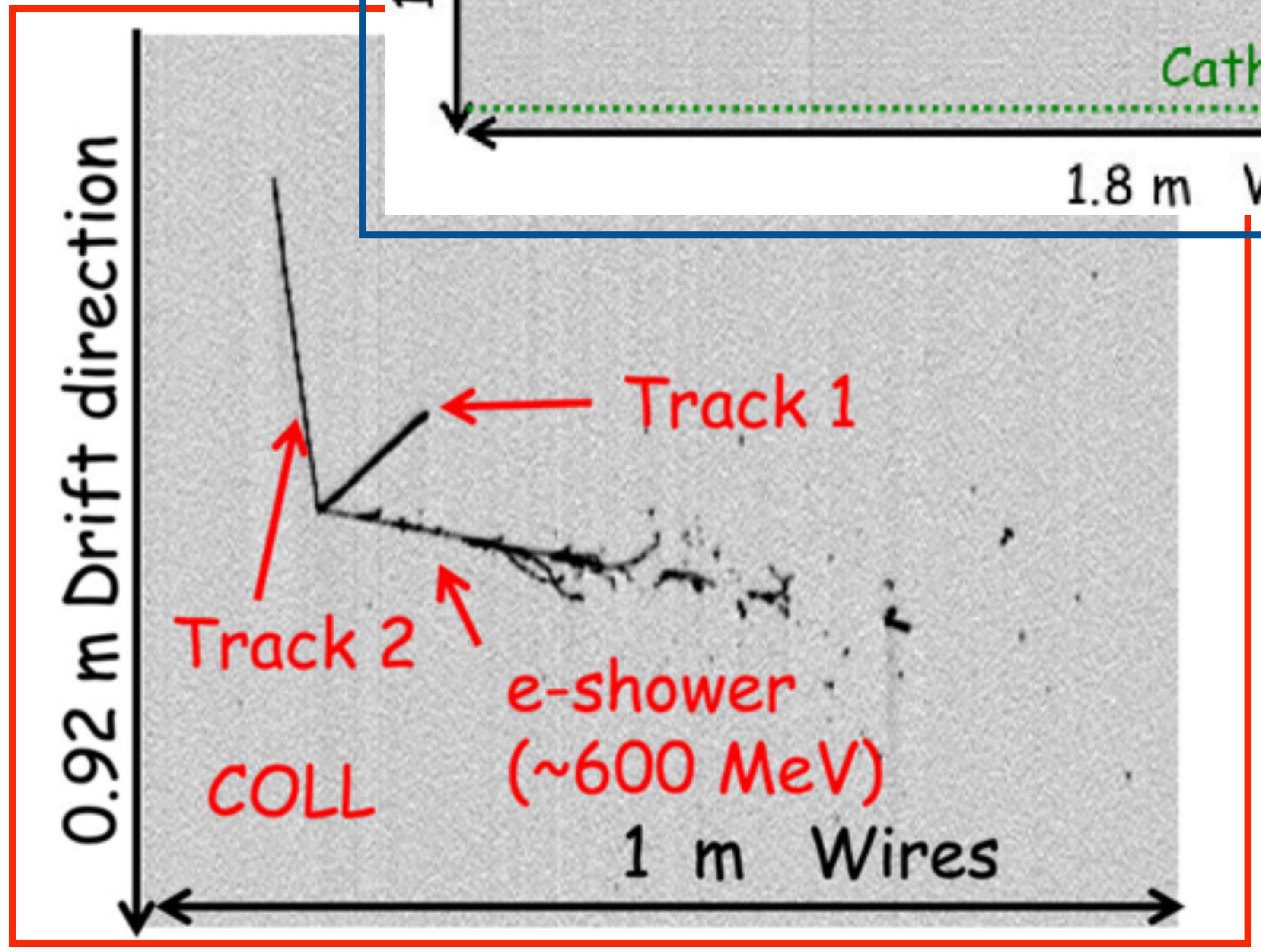
- We employ **visual scan ν events selection** and **Monte Carlo** simulations to identify reconstruction pathologies, explore reconstruction improvements and tune our selection algorithms for analyses

- Most frequent pathology is **track splitting** $\mathcal{O}$(6-7%), followed by **wrong vertex ID** $\mathcal{O}$(4%) and **track/shower ID**
- Validation w/ visual scan based on the 3D position of the vertex V, end point and length of μ track



$\Delta(\mathrm{scan}-\mathrm{reco})\mathrm{vertex}_z$ in the beam direction for 476 ν_μ CC contained candidates selected with visual scan of events



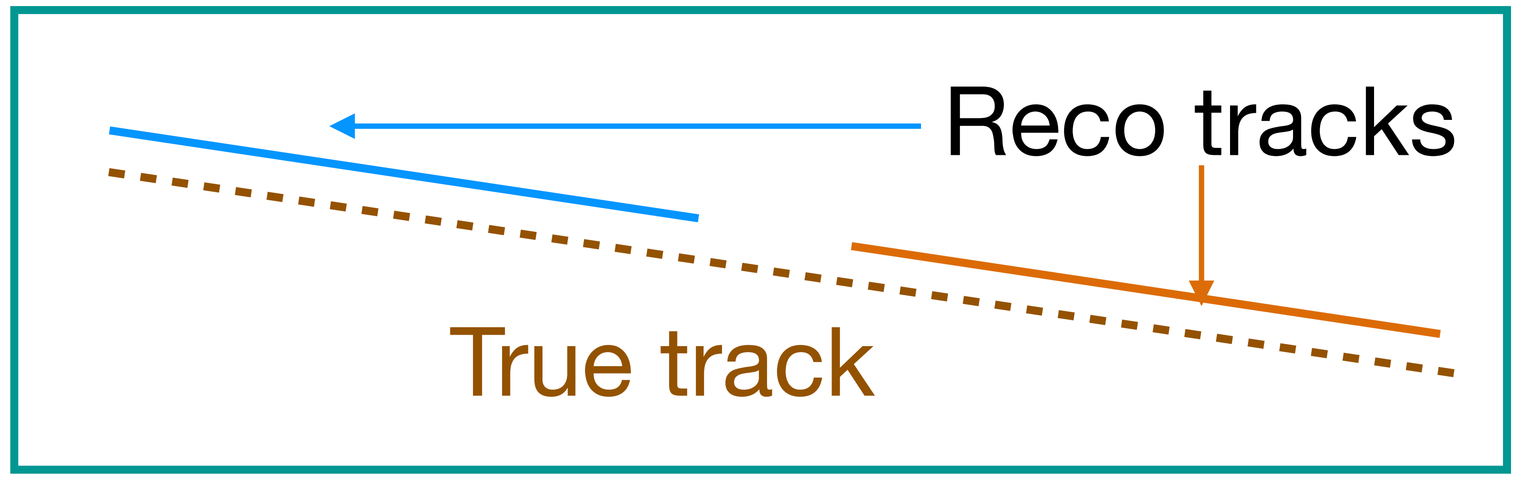


ν_e CC candidate from NuMI

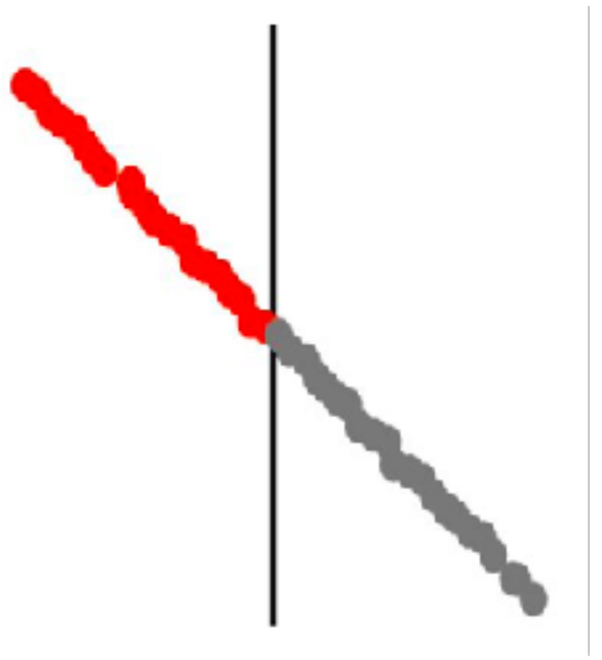
Eur. Phys. J. C 83:467 (2023)

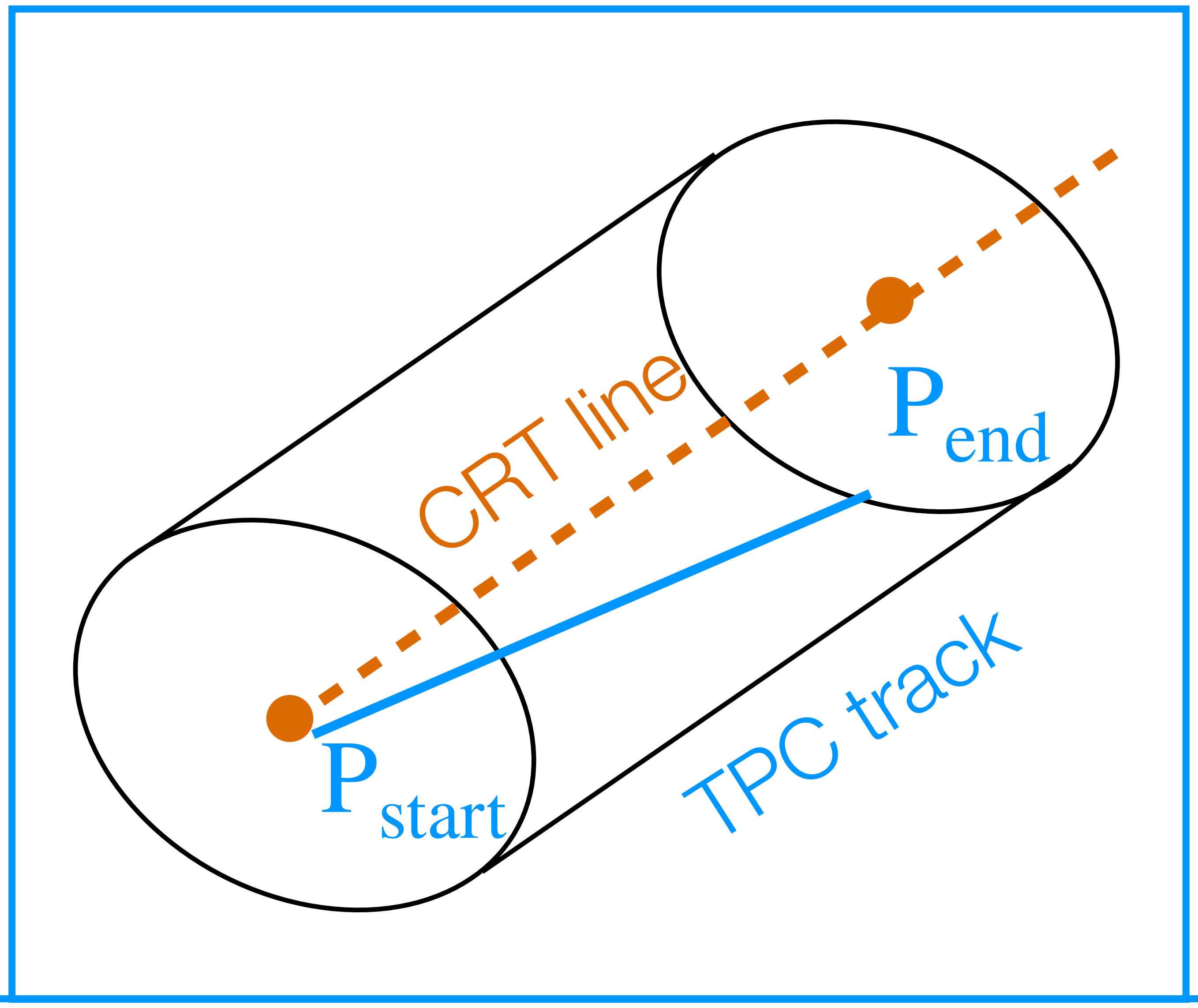
Pandora-based event reconstruction: track splitting

- Several studies to mitigate the problem of **track splitting:**
  e.g. the single track of a μ is reconstructed as $n \geq 1$ segments

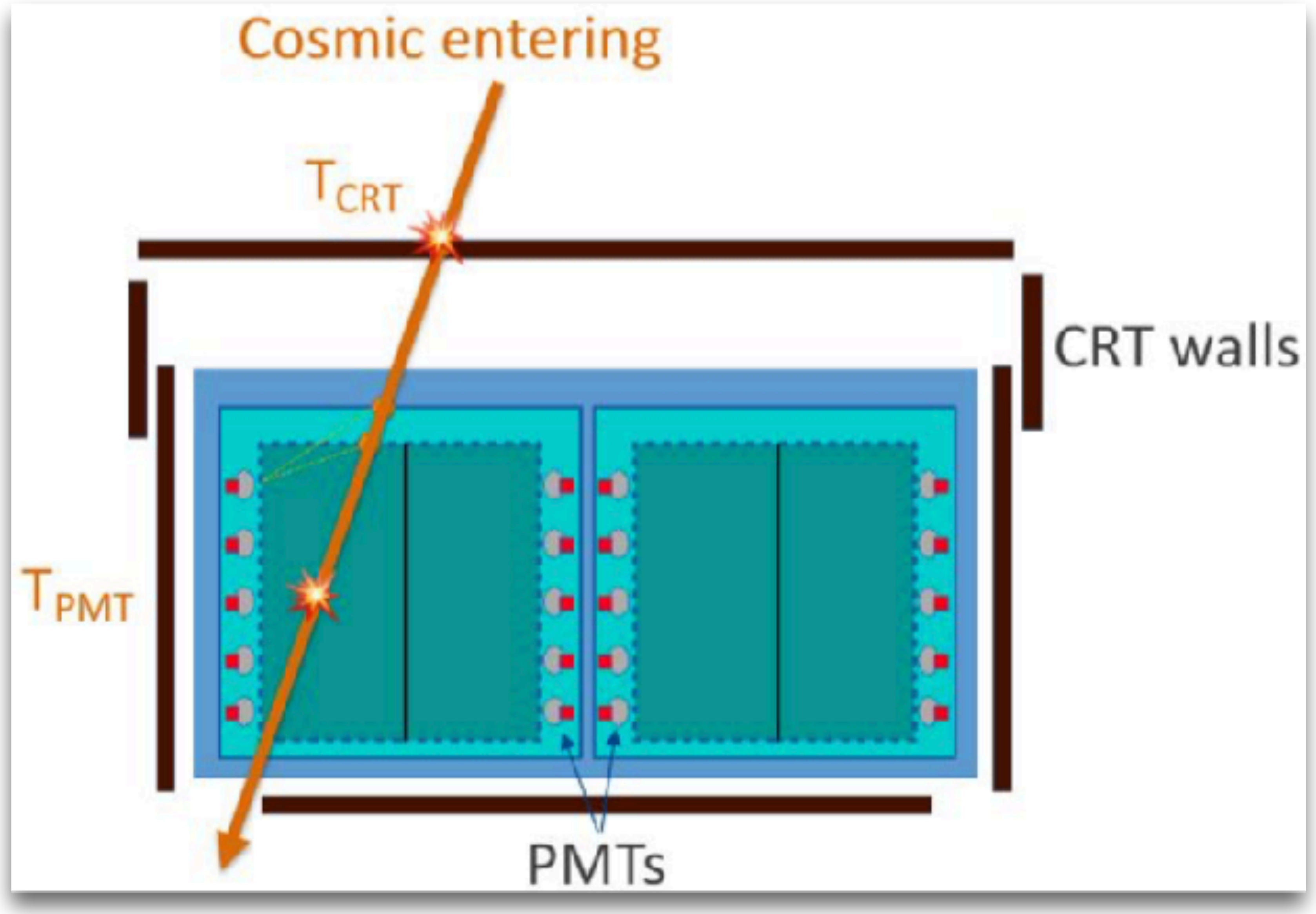


  - Track splitting happening at detector boundaries:
    $z = 0$, at the **cathode**
  - Ongoing study of a **stitching algorithm** to join track pieces post-reconstruction based on **MC**
  - Study of a stitching algorithm on **cosmic μ in data:** TPC tracks are identified after CRT-PMT info
  - Study of the **systematic** induced by track-splitting:





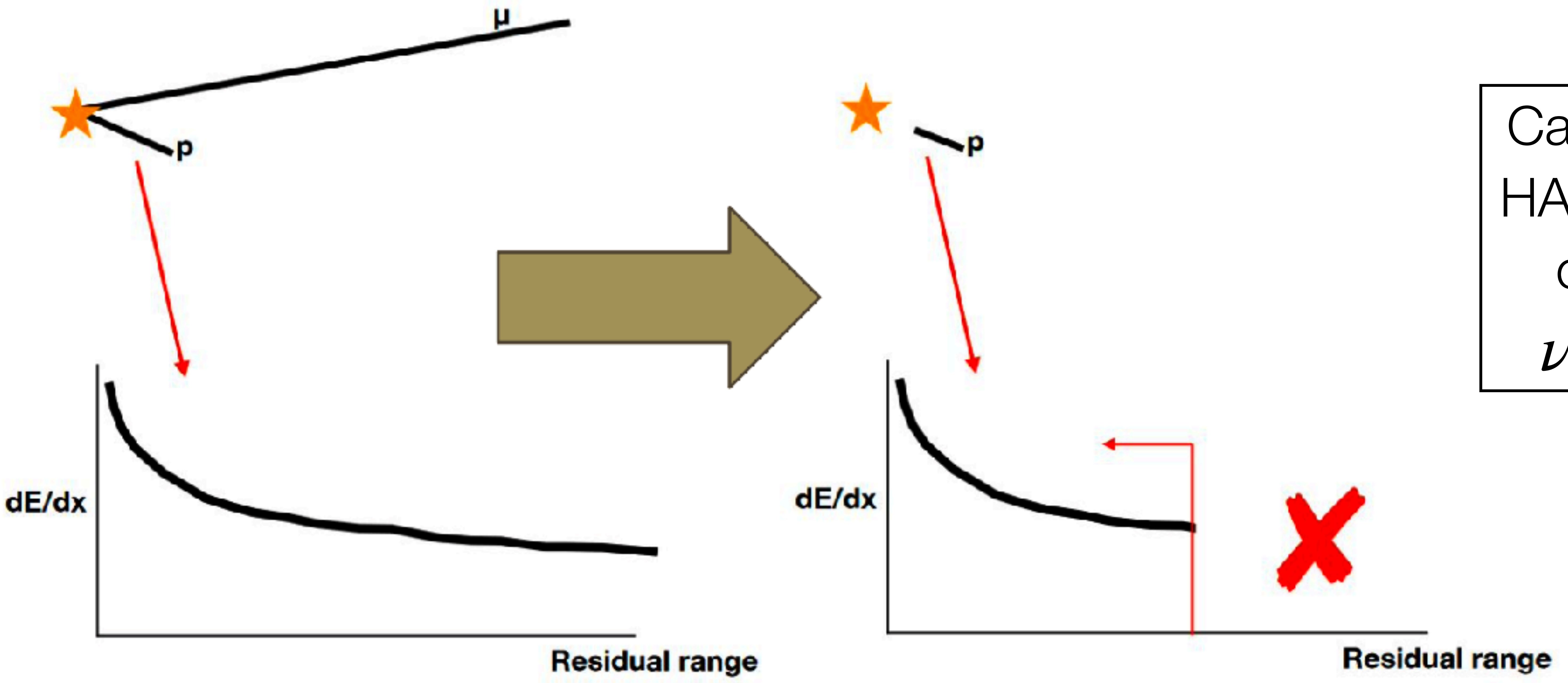
Cartoon of the stitching algorithm



  - Basic idea: break tracks study how reco is affected

Pandora-based event reconstruction: data-driven systematics study

- Goal: understand and account for differences in reconstruction between data and MC
- Foreseen goal: data driven validation of ML algorithms
- *Hit Activity Removal from Particles for Systematics* (**HARPS**): operate on specific particles and reduce their size ↔ similar to starting with a lower energy particle and analyse the impact on reconstructed quantities

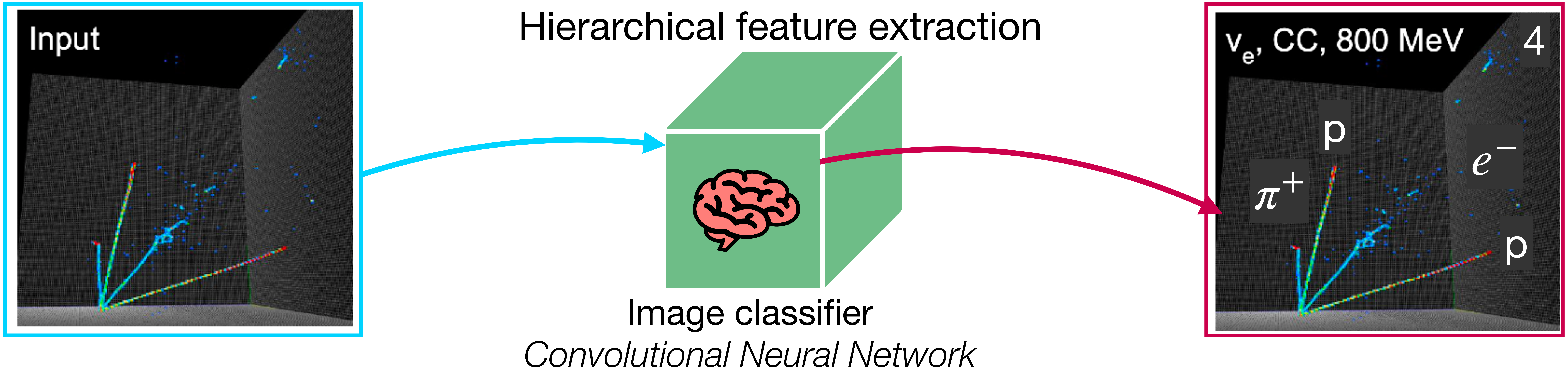


Cartoon of the idea: HARPS on a sample of **protons** from ν + cosmics MC

Pandora-based event reconstruction: summary and next steps

- Strong **interplay** with the needs/results of the ongoing analysis efforts in defining our goals: we are increasing our effort towards evaluating **reconstruction (detector) systematics**
- Several efforts to mitigate the effects of the most relevant reconstruction pathologies at different levels including track splitting, track vs shower misidentification, vertex reconstruction
- Next steps foreseen: continuous **validation** of the reconstruction chain and (re)training of the **ML algorithms** employed in several points of the reconstruction any time relevant changes to signal processing at previous stage are included in the data processing chain

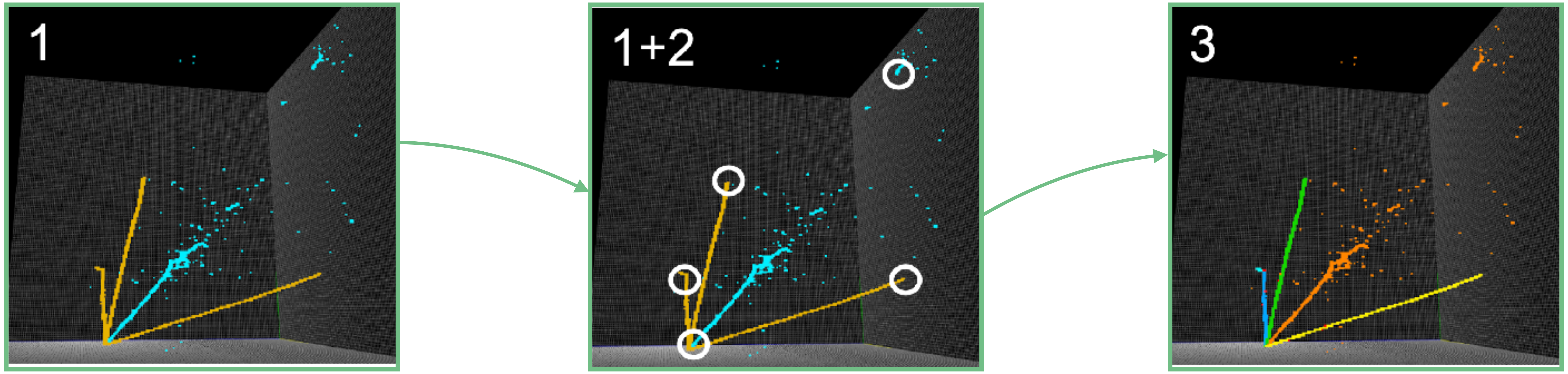
Machine Learning (ML) based LArTPC event reconstruction



1 Separate *voxels* based on the **topology**

2 Find **important points** (vertex V, start/end P)

3 Cluster particles



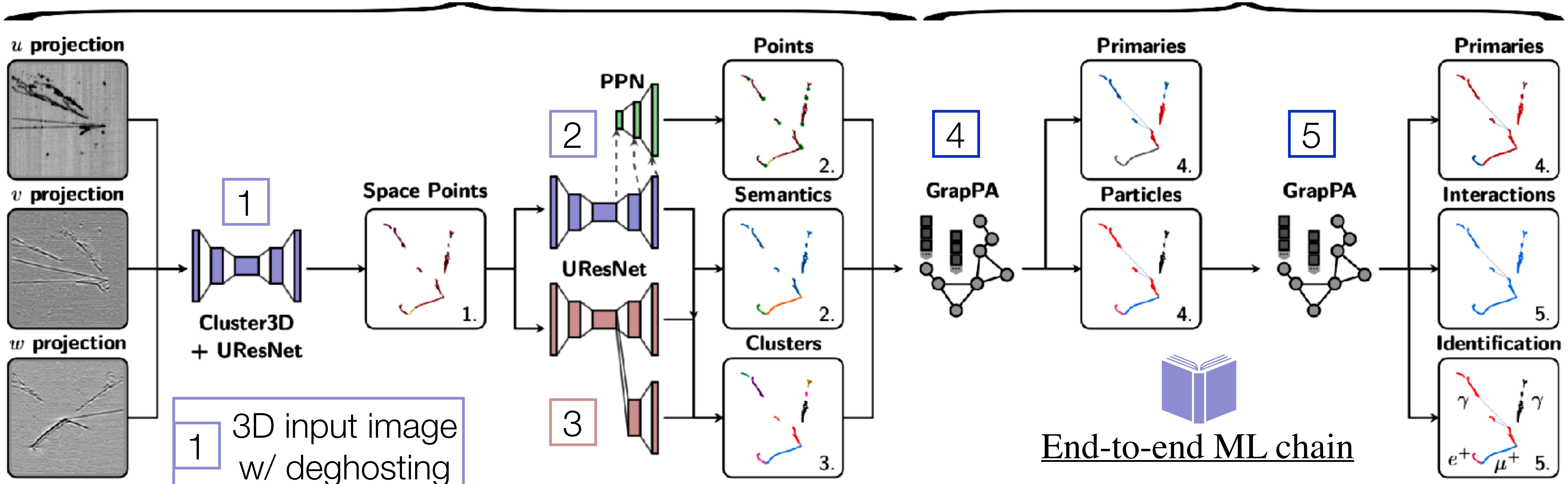
ML-based LArTPC event reconstruction: end to end reconstruction chain

2 Voxels classified in different abstract particle classes + identification of the points of interest

4 Assemble shower objects and identify primary fragments

Convolutional NN

Graph NN



1 3D input image w/ deghosting

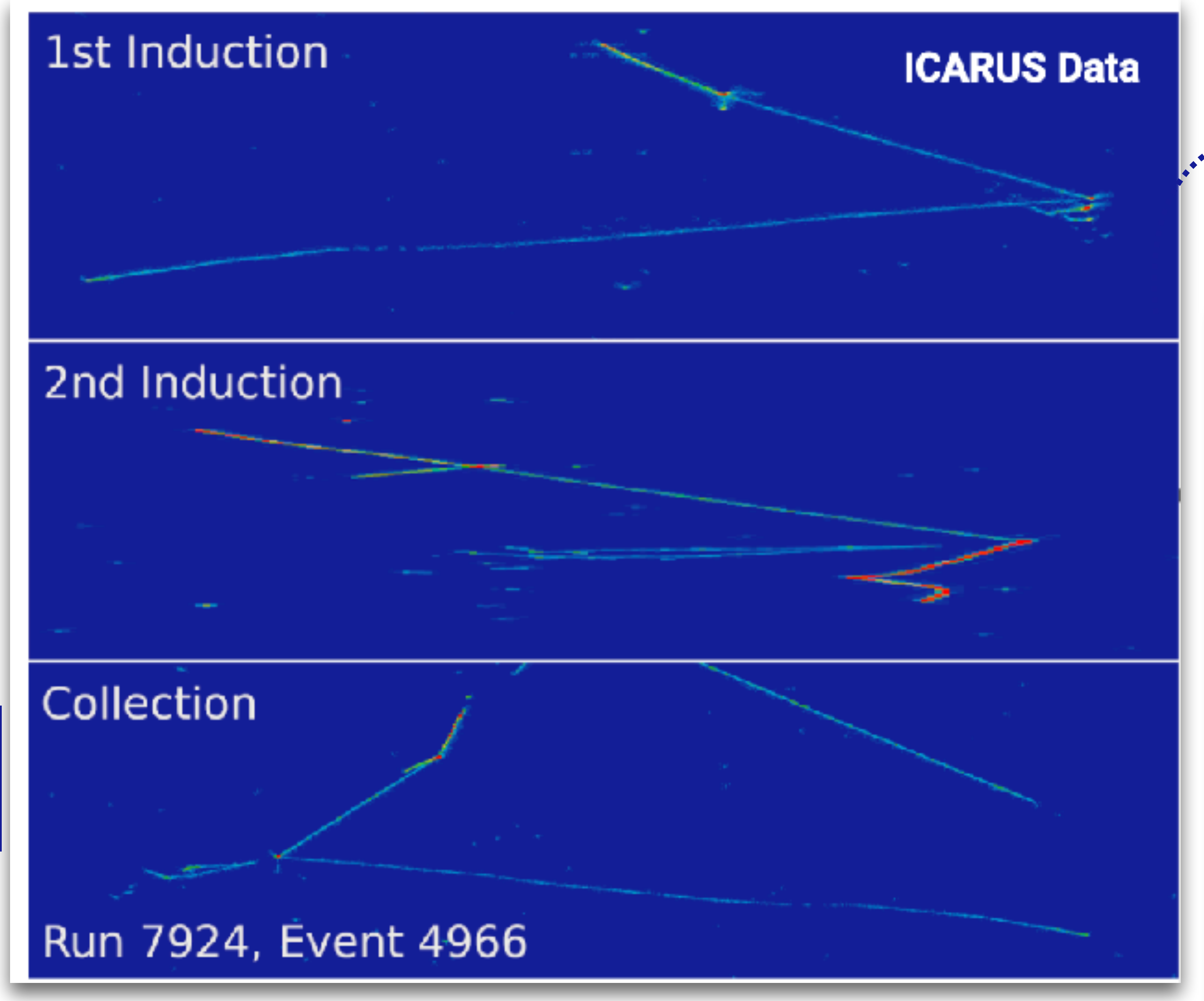
End-to-end ML chain

2D views from wire planes

3 NN to build individual dense particle clusters

5 Particles aggregation into interactions and ID

1 ML-based event reconstruction: hierarchical feature extraction

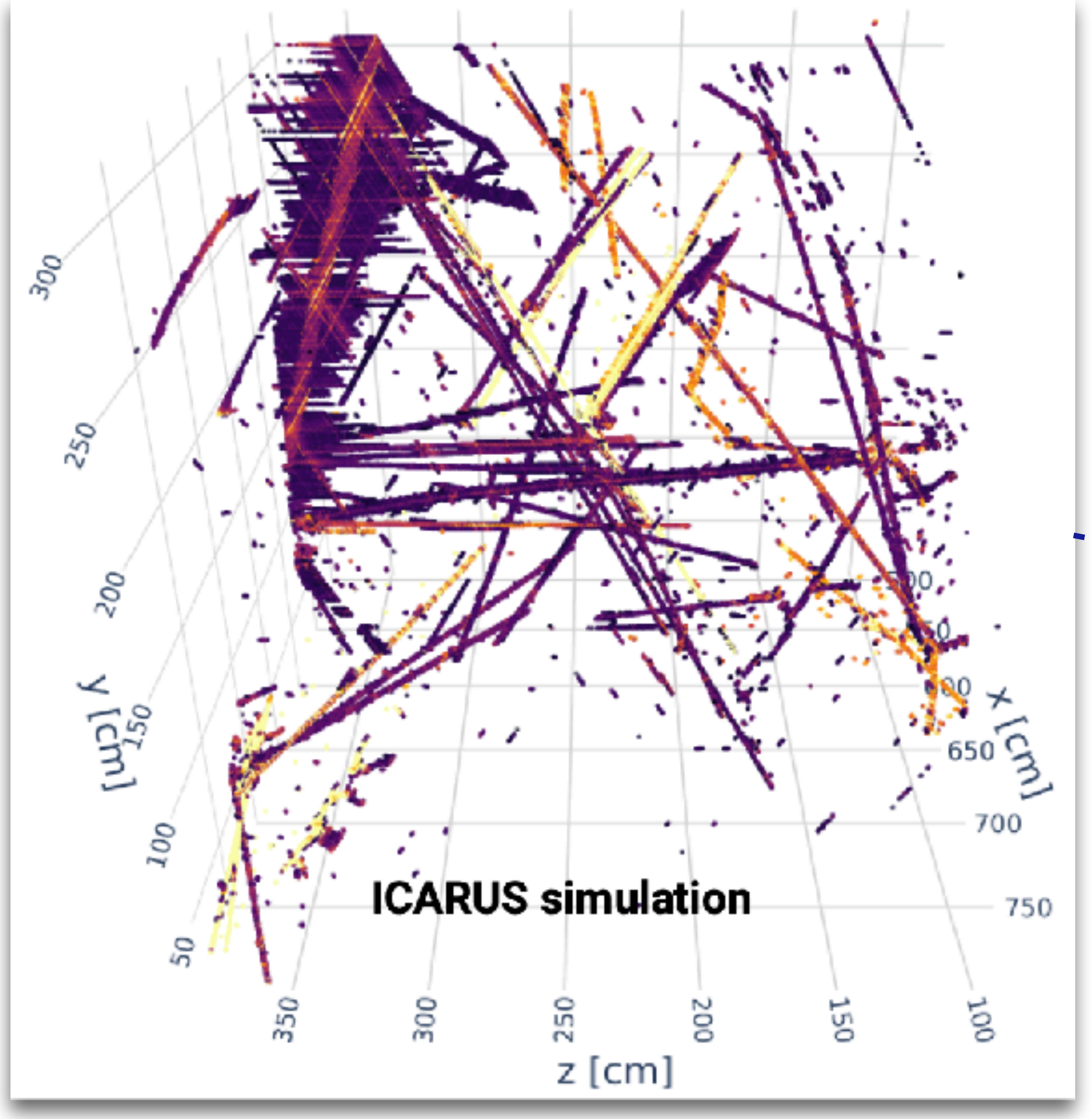


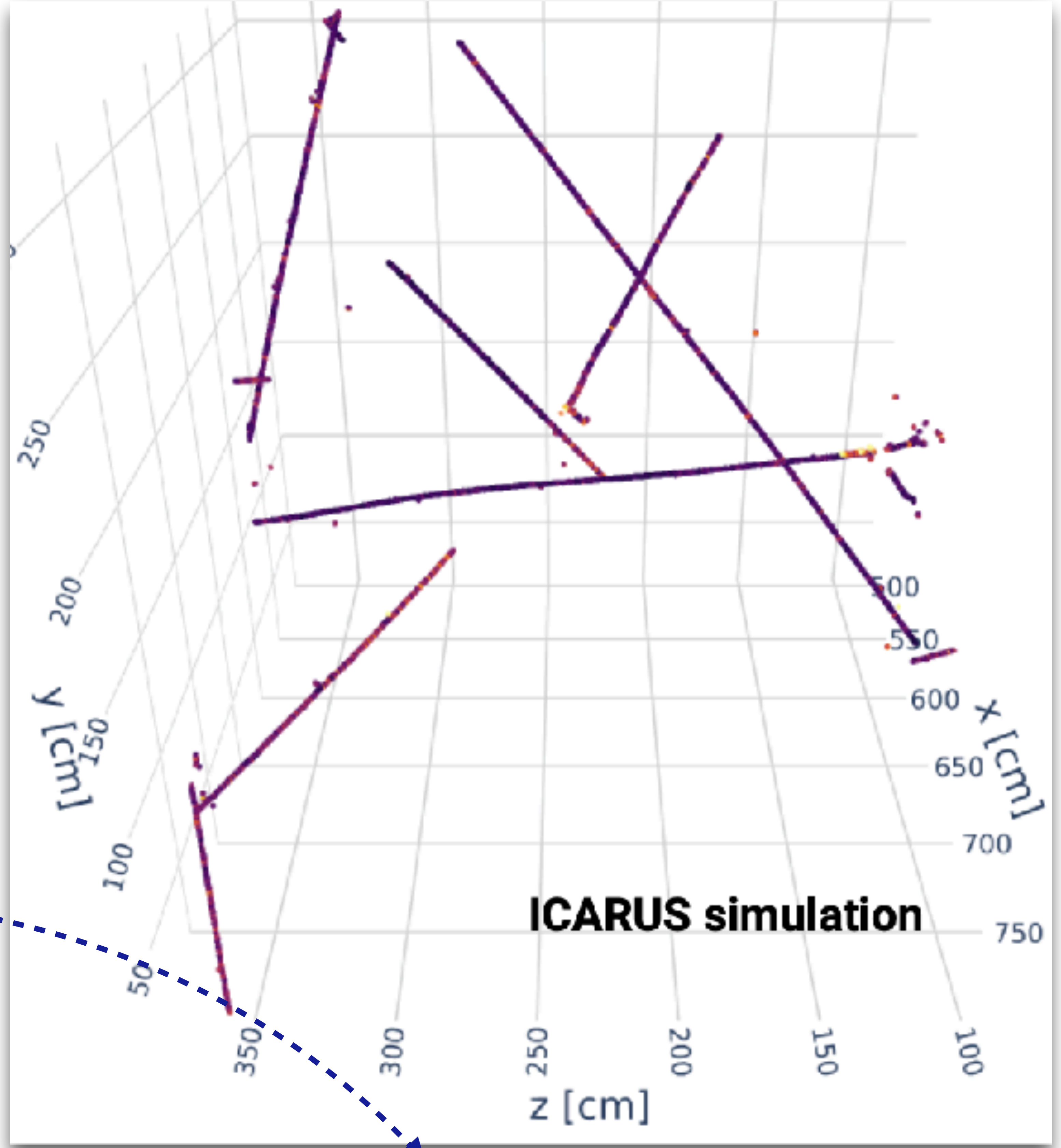
time

wire

Starting point:
3 wire planes ↔ 3 x 2D images

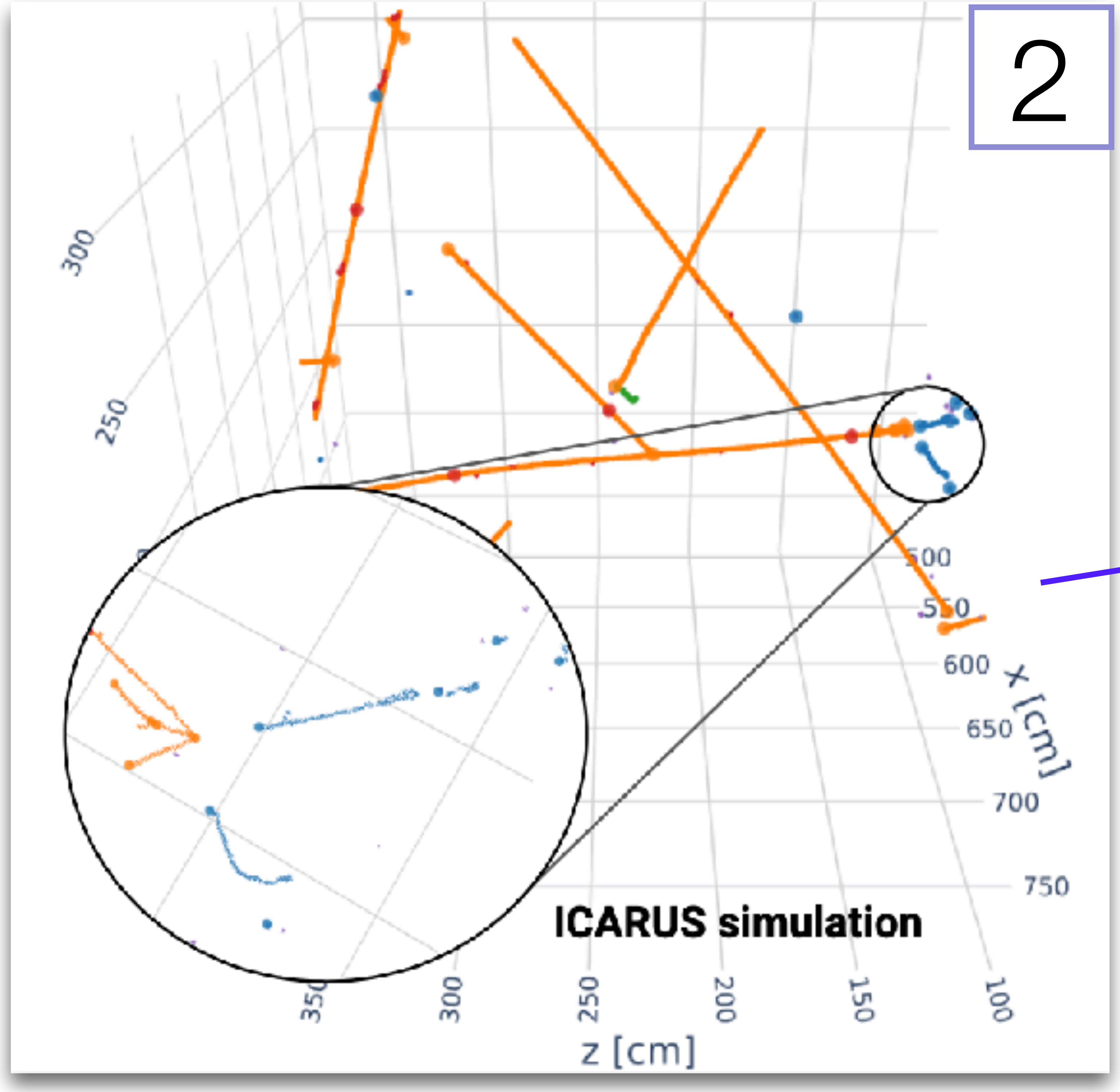
Cluster 3D: make all valid (time-compatible & intersecting) combinations of hits across 2 wire planes





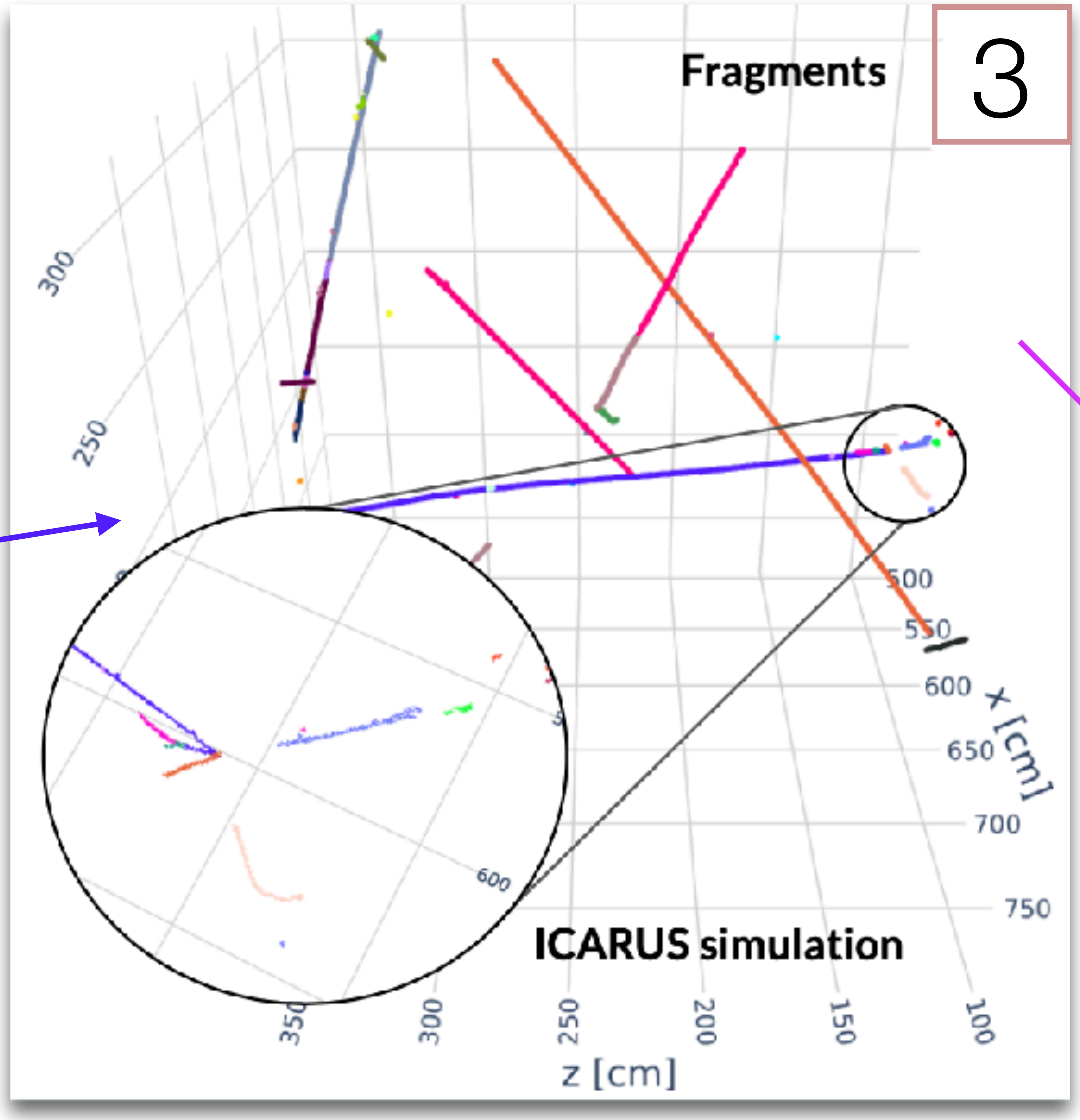
Deghosting: use **U-ResNet** to identify and remove *artifacts* of the reconstruction

ML-based event reconstruction: hierarchical feature extraction



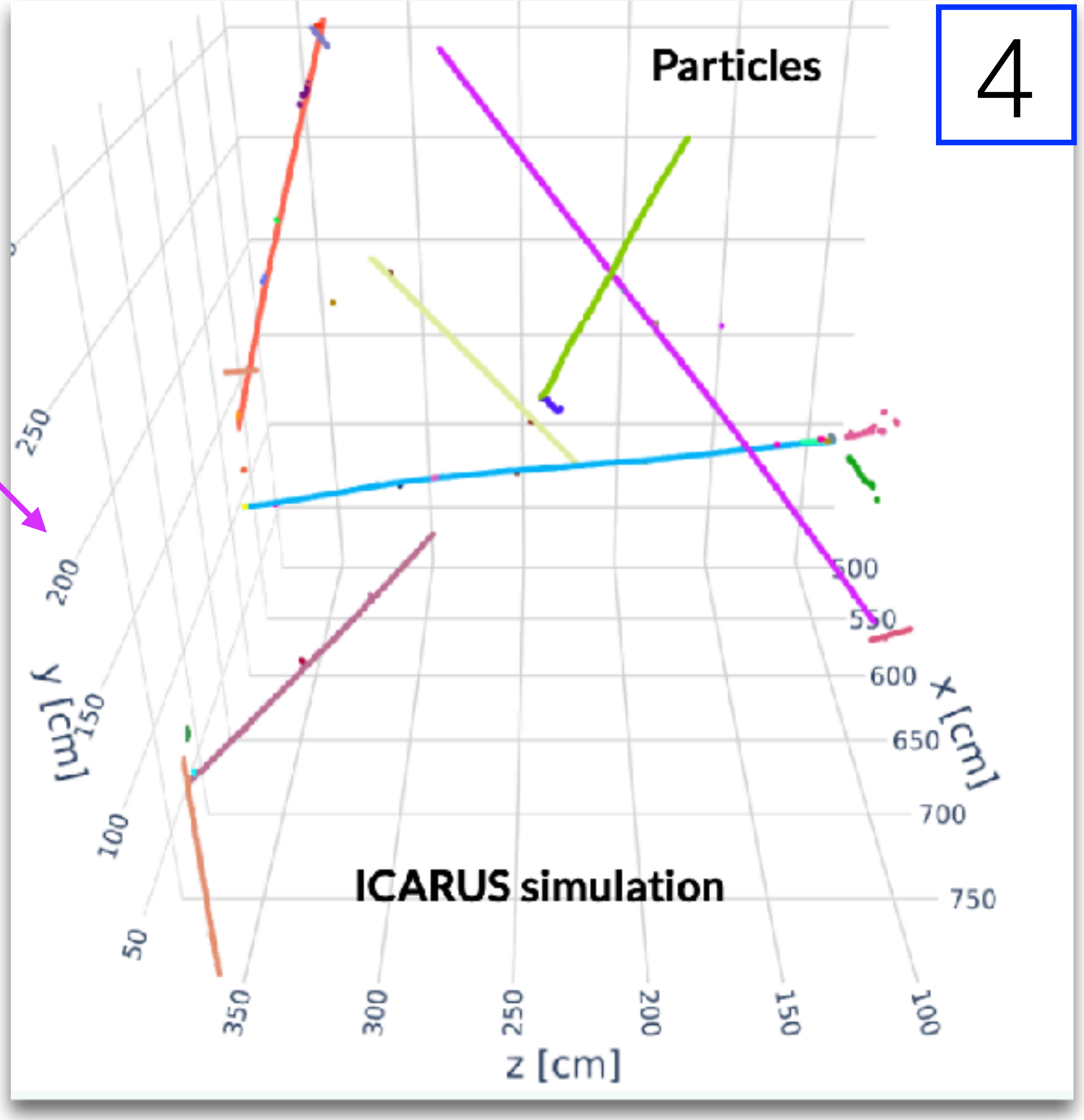
Semantic segmentation & Point of interest (PPN)

Distinguish different particle types based on topological features and identify vertex, start/end points



Particle clustering

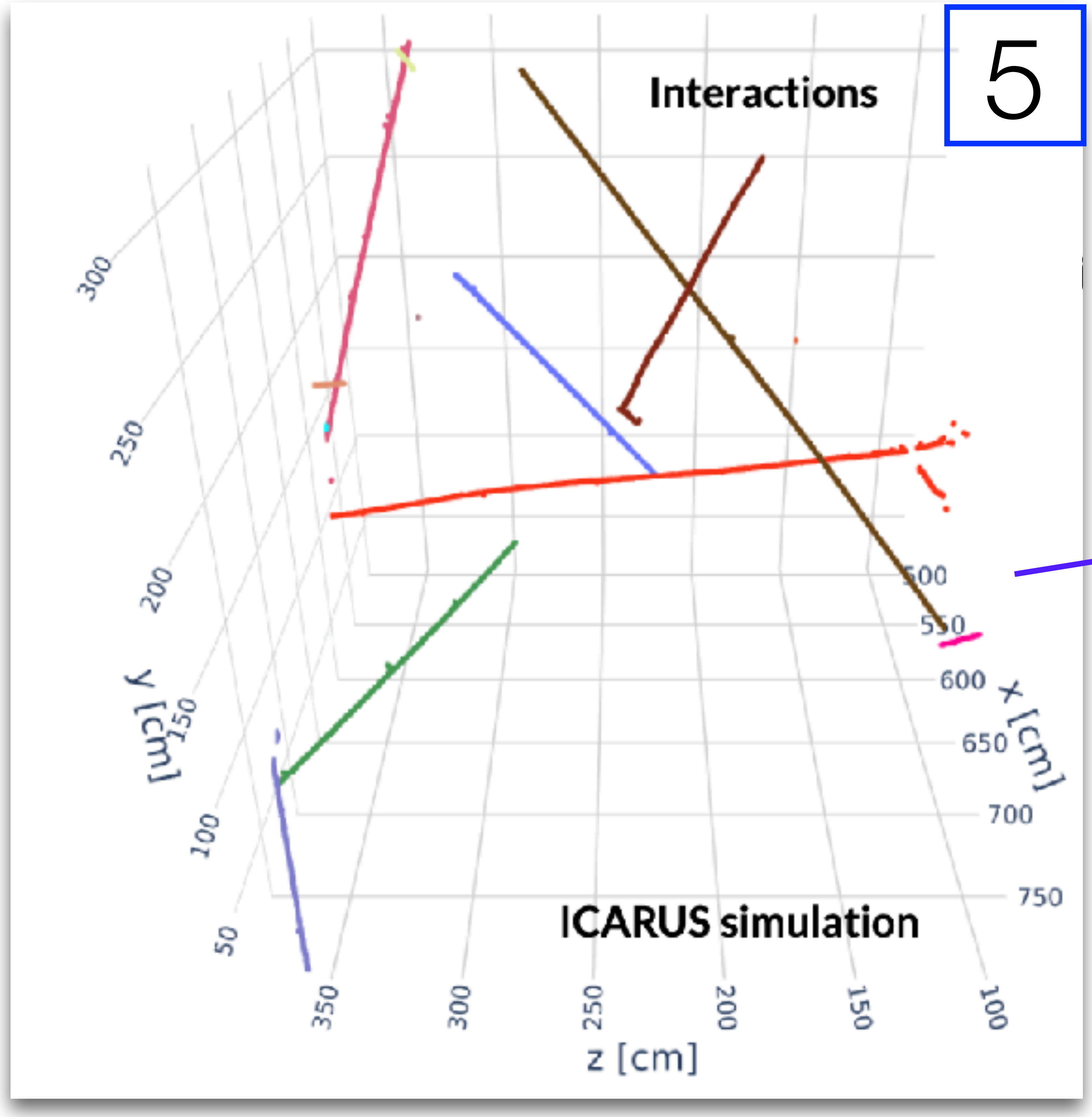
Cluster particle fragments that belong to a common *semantic* class, i.e. break track/shower fragments at PPN

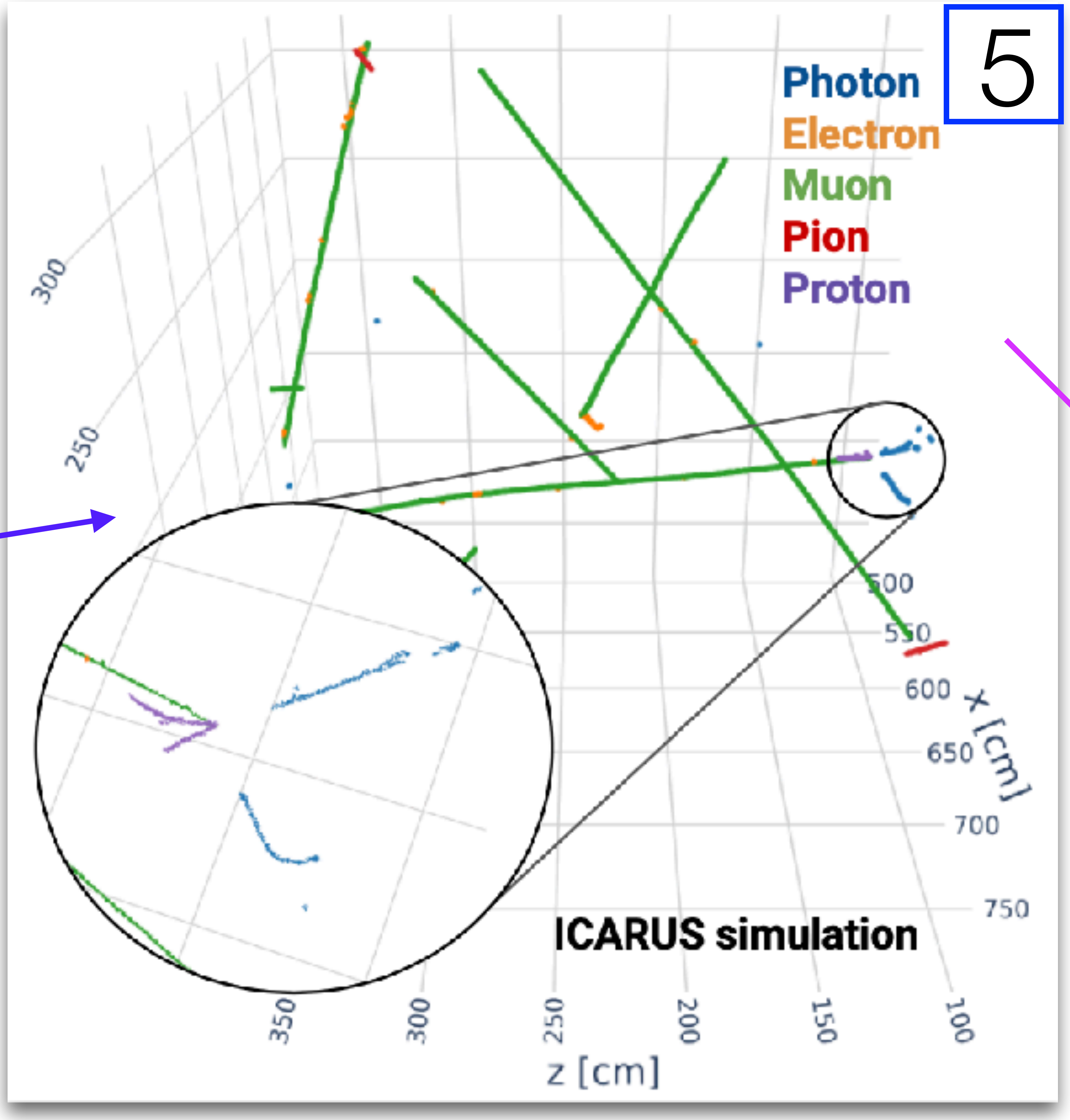


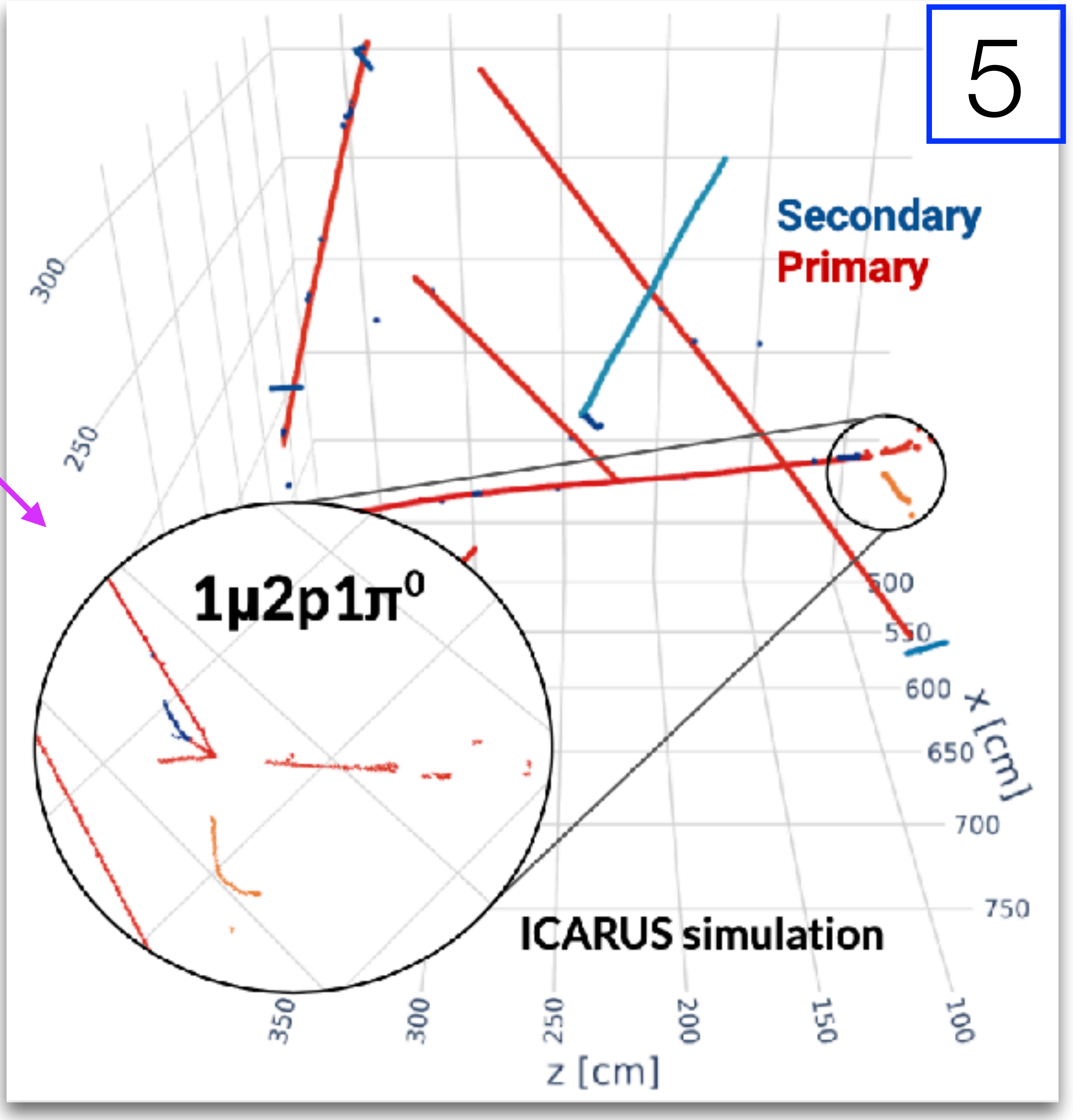
Particle aggregation

Use a Graph Neural Network (GNN) to aggregate fragments and form particles

ML-based event reconstruction: hierarchical feature extraction







Interaction aggregation

Use a Graph Neural Network (GNN) to aggregate particles and form interactions

Particle identification

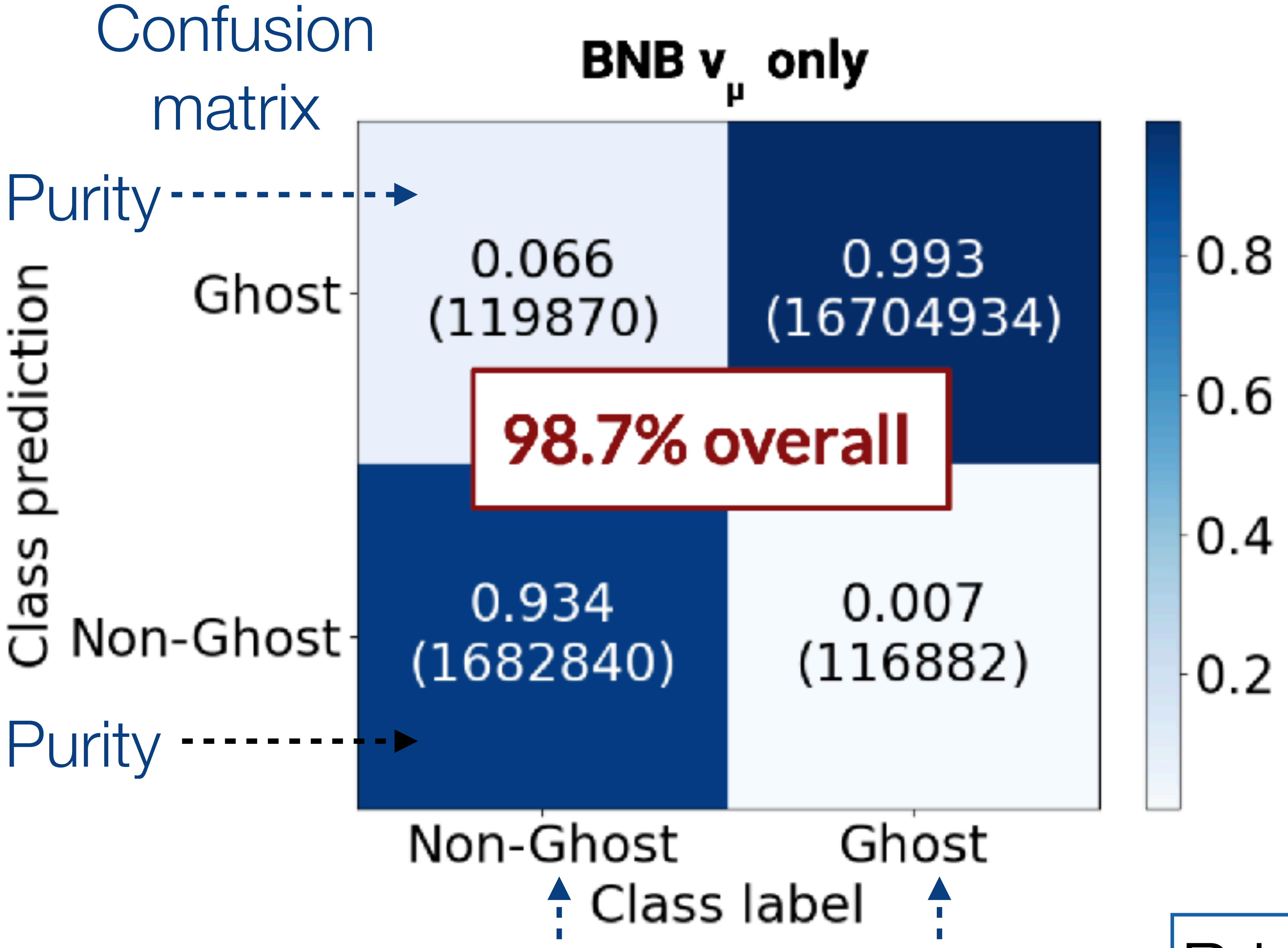
Use GNN to identify particles e, γ, μ, π, p in context

Primary identification

Separate particle(s) which originate from the **vertex**. This is fundamental for analyses.

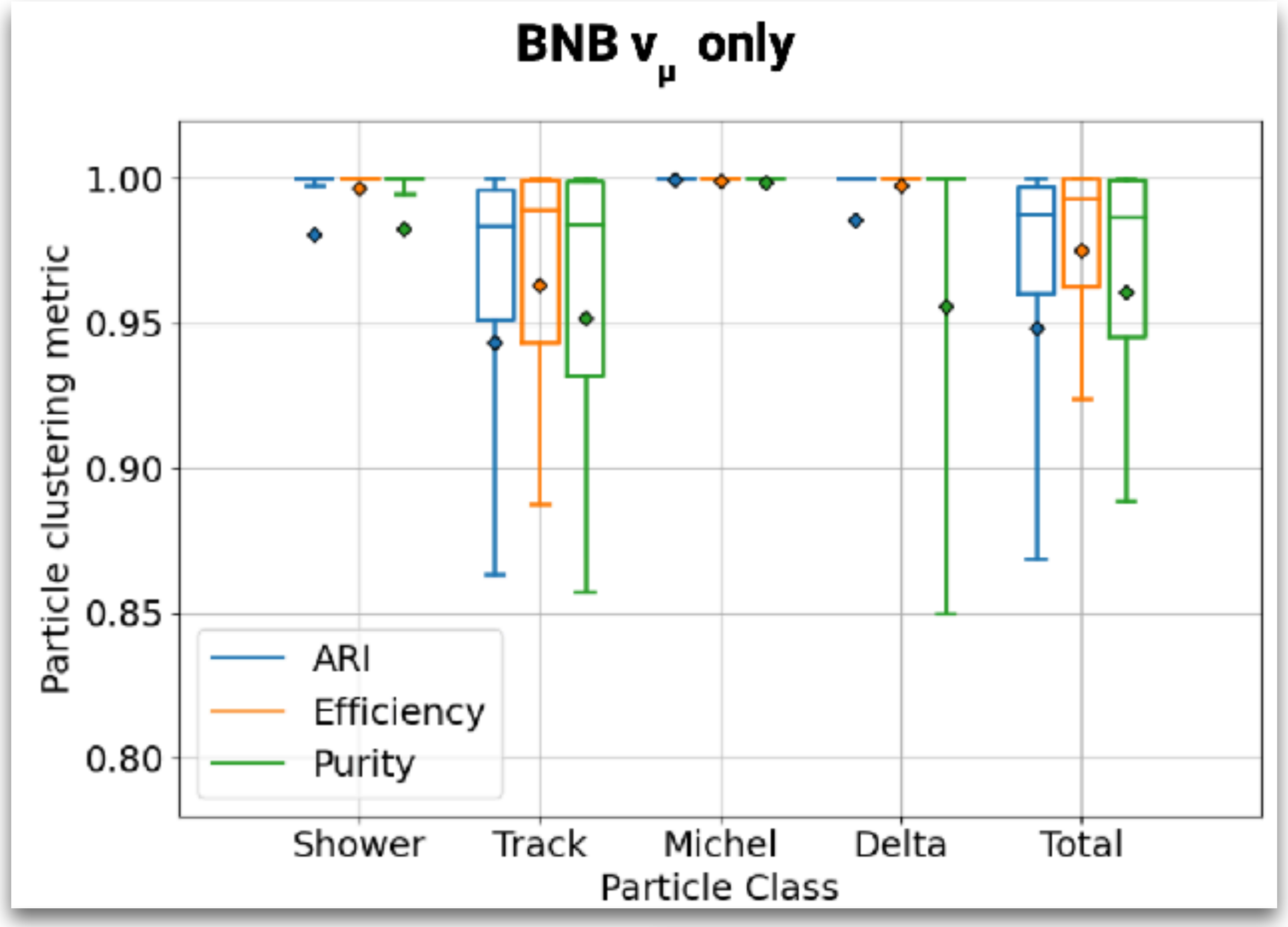
ML-based event reconstruction: performance

Confusion matrix



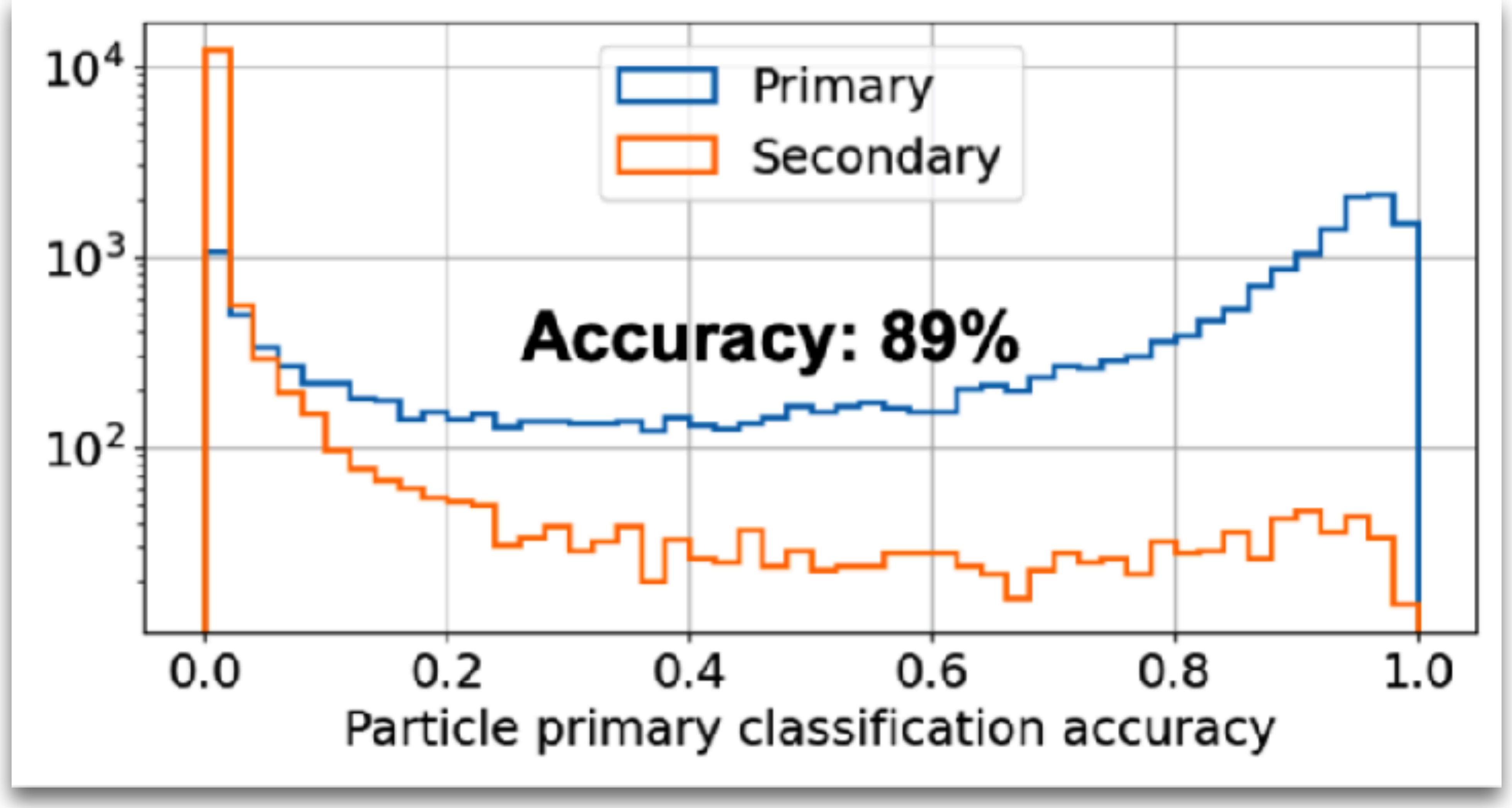
Efficiency Efficiency

Deghosting



Particle clustering

Primary identification



ML-based event reconstruction: current effort and next steps

- Continuous effort to improve the performance of the end-to-end ML-based reconstruction chain as a whole exploiting both MC simulations and visual scanning info

- Several physical analyses underway in ICARUS using ML-based reconstruction:

  - **Beyond Standard Model physics:** Higgs-portal scalar decays, $S \rightarrow ee$, (J.Dyer) see her talk tomorrow!

Thank you for your attention!