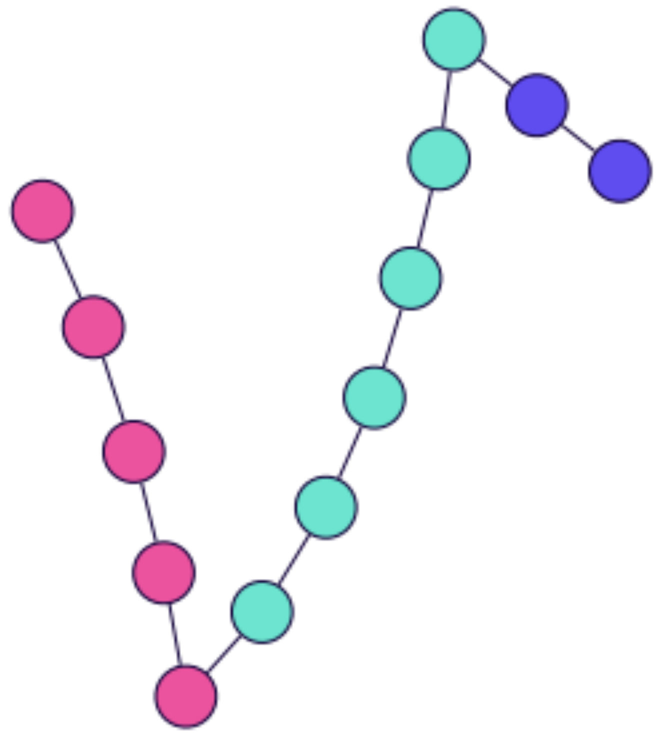
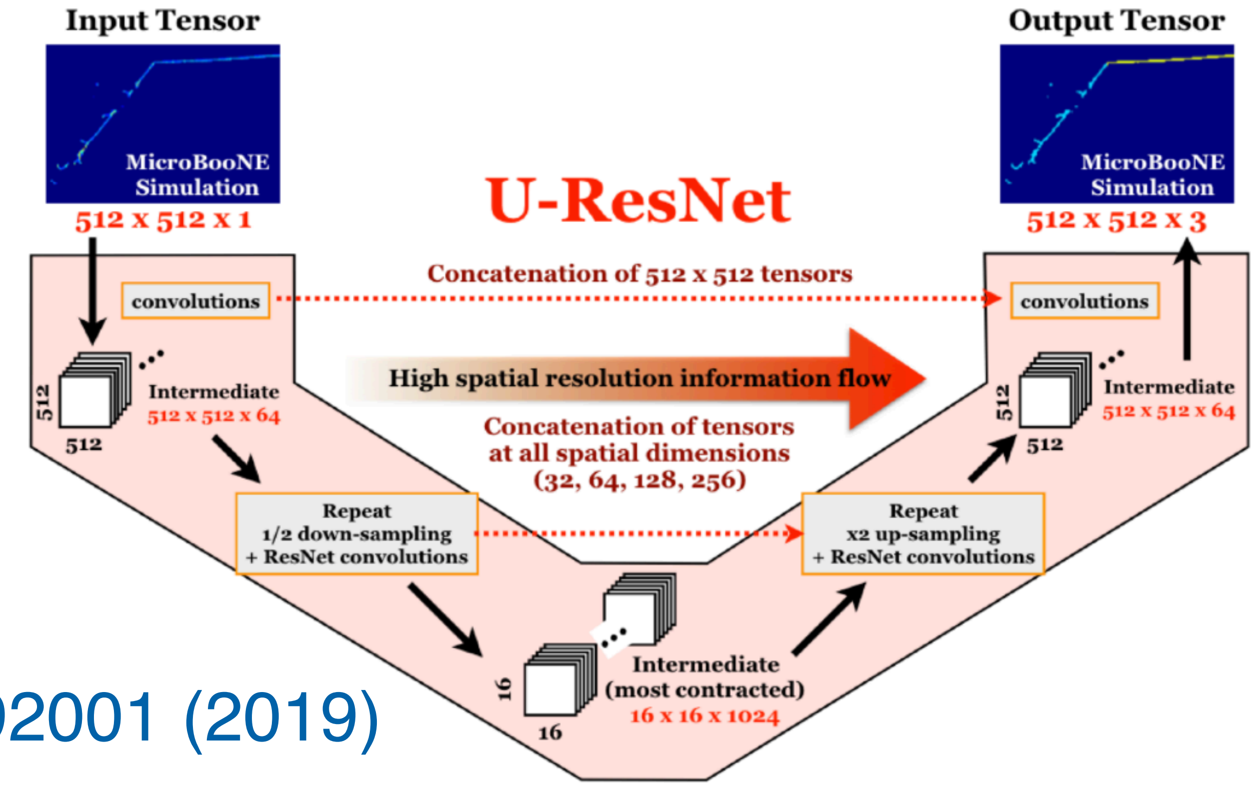
NuGraph

Giuseppe Cerati (FNAL)
Fermilab IF AI Jamboree
Dec. 6, 2024

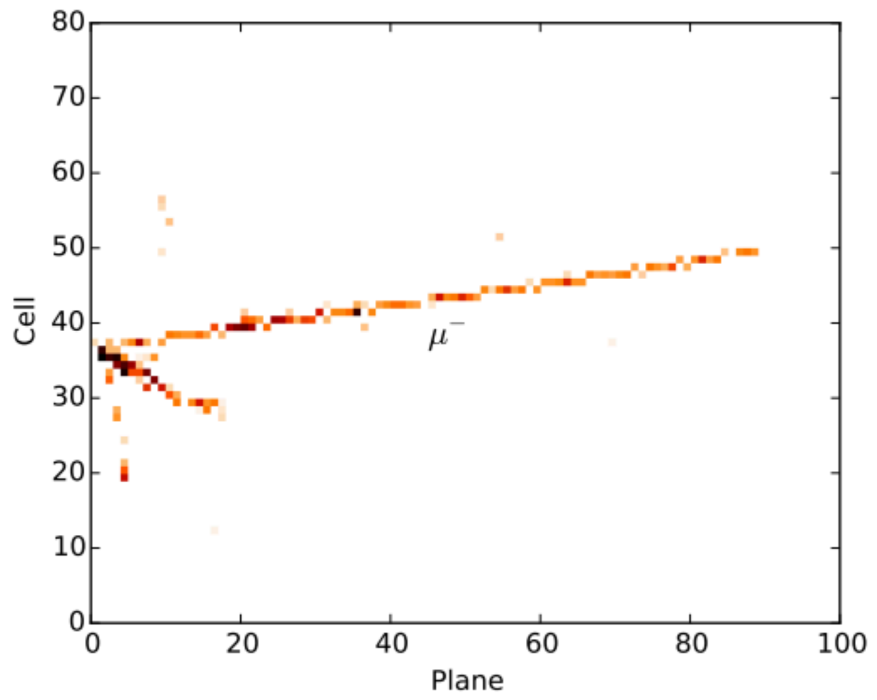


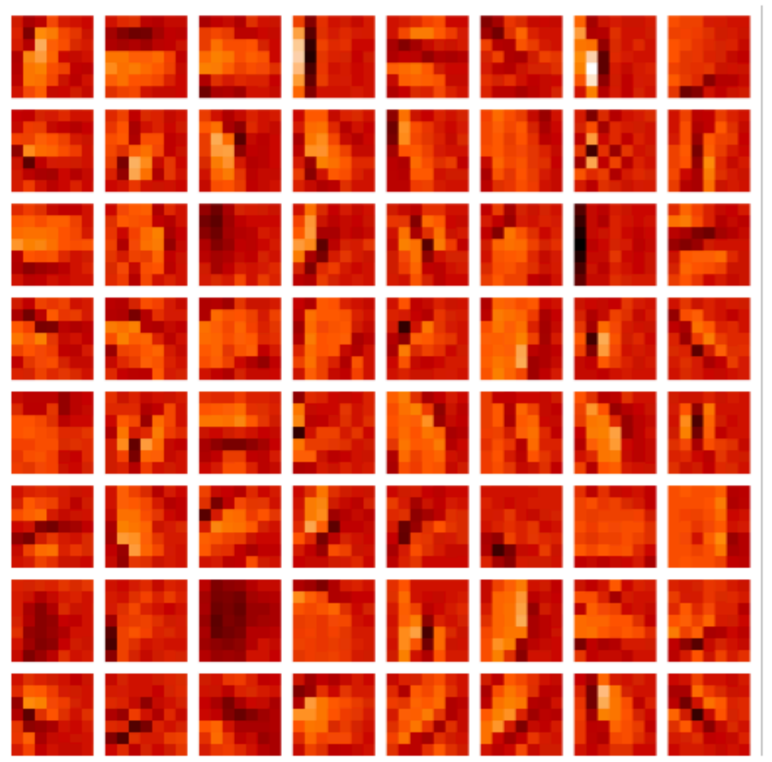
Introduction

- CNN-based networks reshaped event reconstruction for neutrino physics



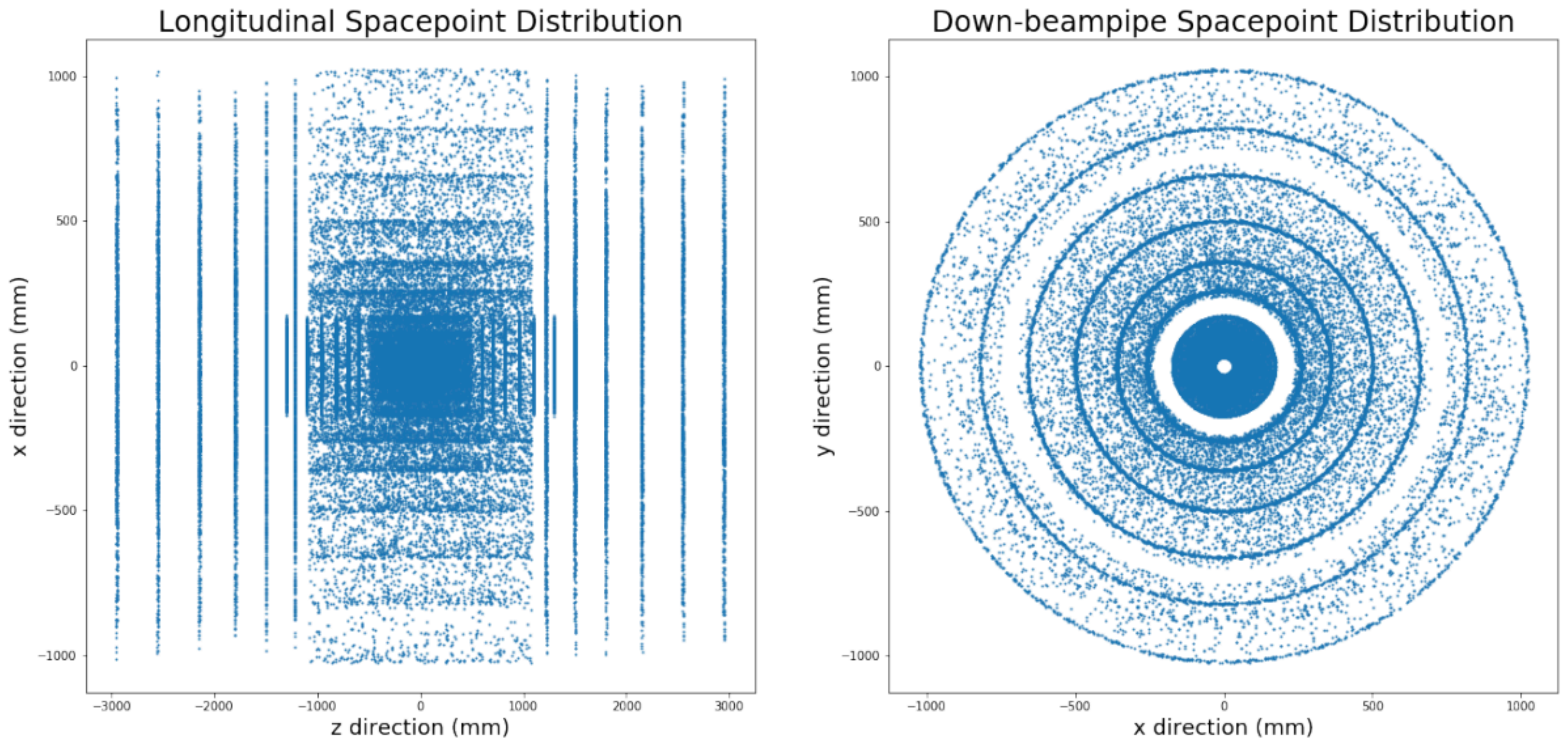
Phys. Rev. D99, 092001 (2019)

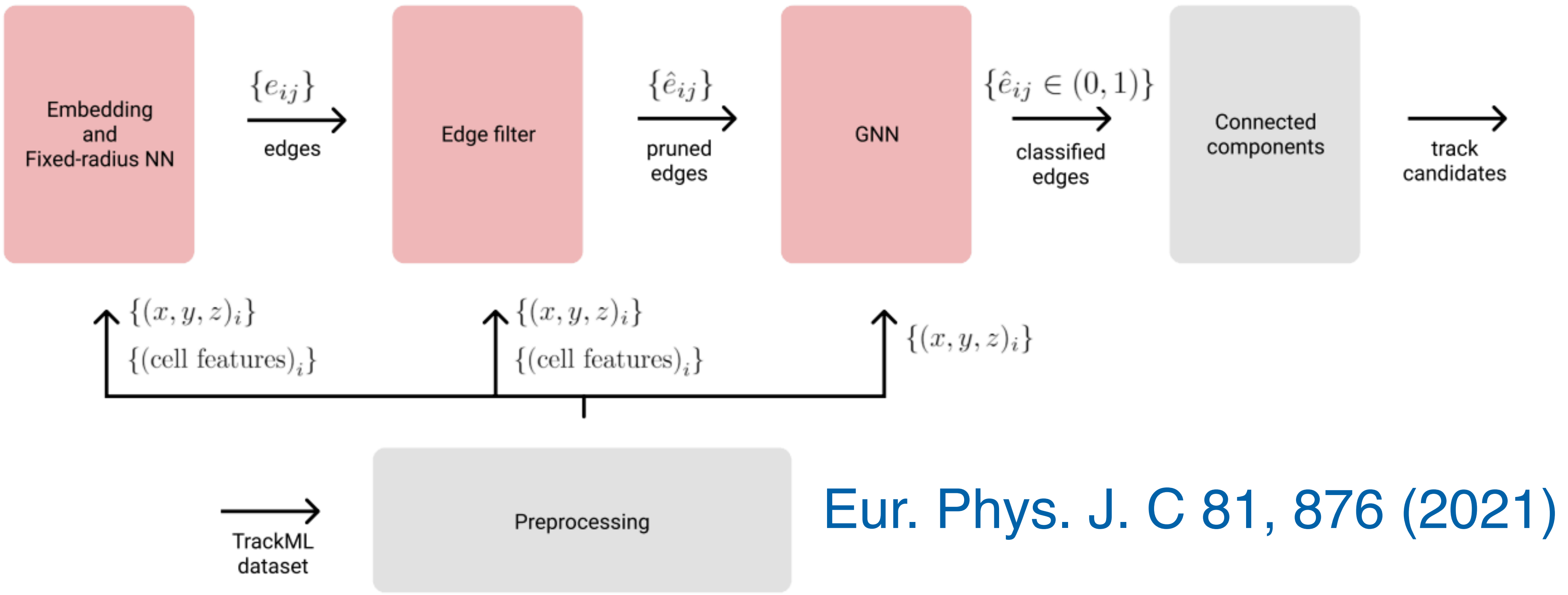




2016 JINST 11 P09001

- GNN proved to be promising for track reconstruction at the LHC
  - naturally sparse
  - no image pre-processing
  - flexible structure





Eur. Phys. J. C 81, 876 (2021)

NuGraph2 Paper Reference [arXiv:2403.11872]

PHYSICAL REVIEW D **110,** 032008 (2024)

Graph neural network for neutrino physics event reconstruction

A. Aurisano and V. Hewes
University of Cincinnati, Cincinnati, Ohio 45221, USA

G. Cerati and J. Kowalkowski
Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA

C. S. Lee and W. Liao
Northwestern University, Evanston, Illinois 60208, USA

D. Grzenda and K. Gumpula
Data Science Institute, University of Chicago, Chicago, Illinois 60637, USA

X. Zhang
Data Science Institute, University of Chicago, Chicago, Illinois 60637, USA
and University of California, Los Angeles, Los Angeles, California 90095, USA

(Received 19 March 2024; accepted 14 June 2024; published 6 August 2024)

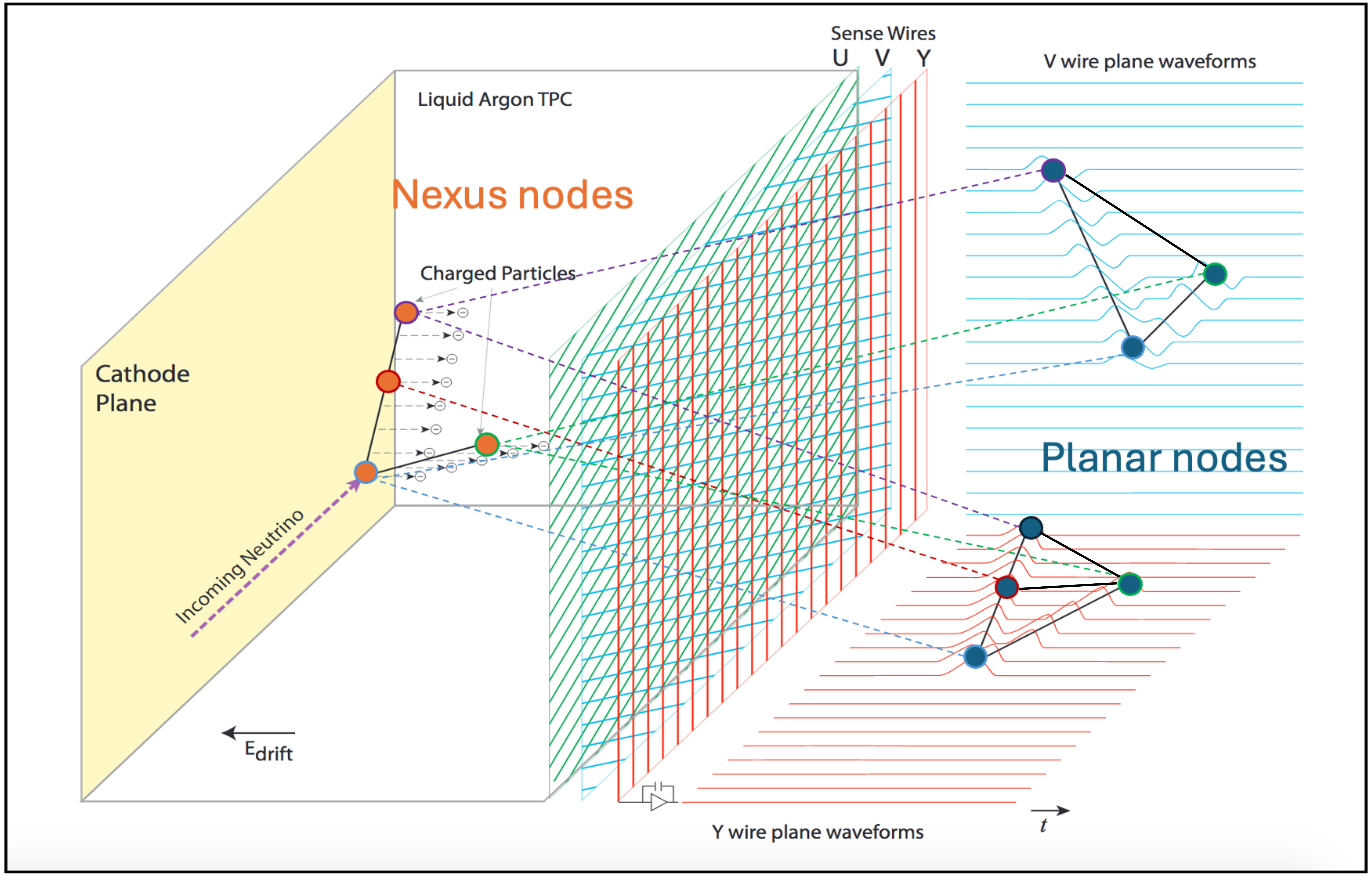
- Preprocessed training data set and trained model available on Zenodo
  - https://zenodo.org/records/12169756

MicroBooNE Open Samples — Overview

- Inspired by **FAIR** principles (findable, accessible, interoperable, reusable data)
- Samples available under "cc-by" license. Template text for acknowledgment is provided.
  - requesting resulting software products to be made available
- Two formats: targeting LArTPC and broader data & computer science communities
  - **art/ROOT** is the same format as used by the collaboration.
    - Files are stored on persistent **dCache** pool area and made accessible with **xrootd**
  - **HDF5** include a reduced subset of the art/ROOT information in a simplified format for usage by non-experts.
    - Files stored on **Zenodo**, providing citable DOI (digital object identifier) & versioning.
- Extensive documentation and tutorials are also made public.
  - Notebooks show how to access the data, demonstrate useful applications, define reference performance metrics

Sample	DOI	HDF5			artroot		
		N events	N files	size	N events	N files	size
Inclusive, NoWire	10.5281/zenodo.8370883	753,467	18	195 GB	1,046,139	24436	6.4 TB
Inclusive, WithWire	10.5281/zenodo.7262009	24,332	18	44 GB	24,332	720	136 GB
Electron neutrino, NoWire	10.5281/zenodo.7261921	89,339	20	31 GB	89,339	2151	761 GB
Electron neutrino, WithWire	10.5281/zenodo.7262140	19,940	20	39 GB	19,940	540	170 GB

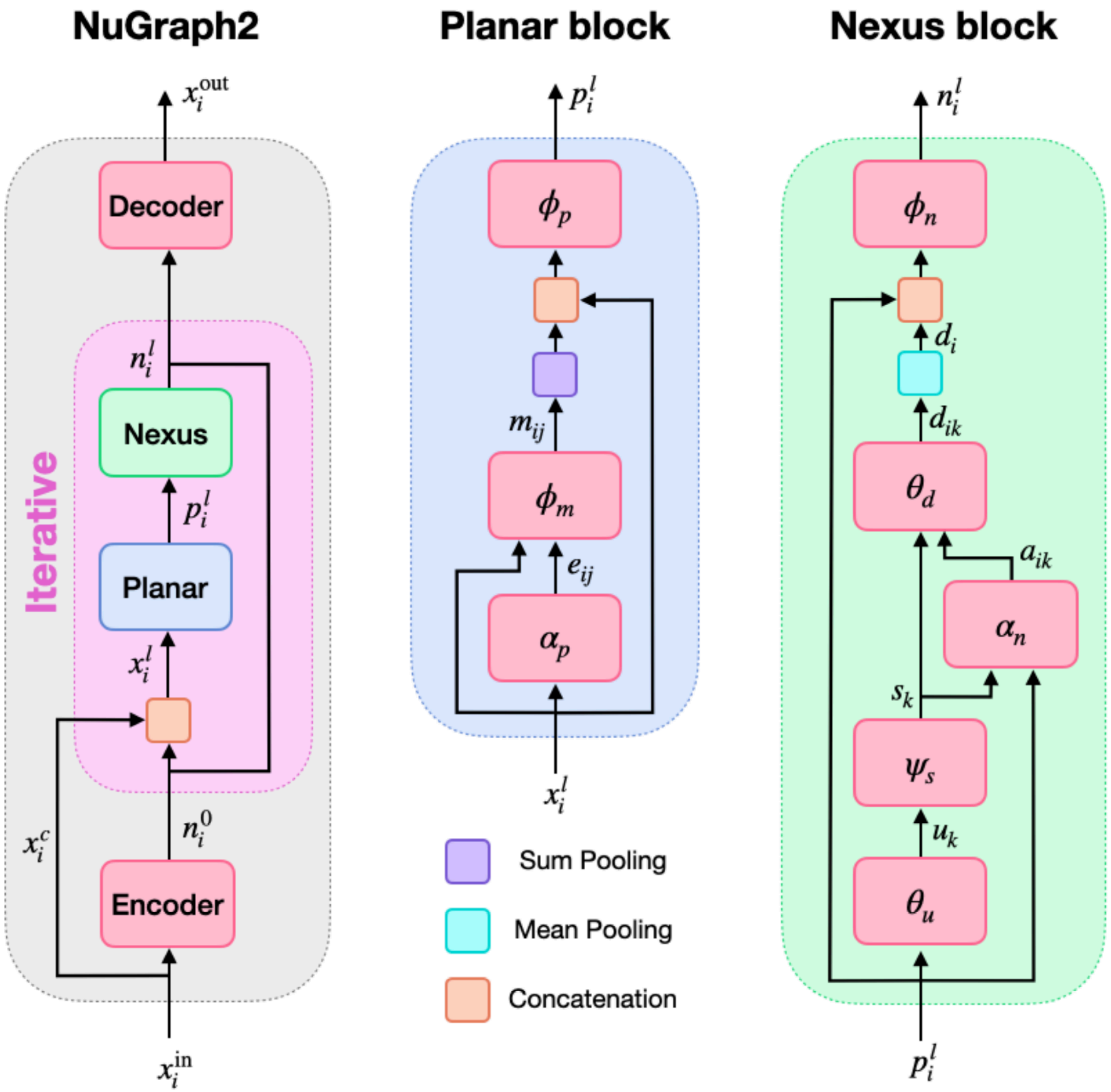
Graph Construction



- Main inputs to the GNN are the Hits
  - hits are Gaussian fits to waveforms
  - features: wire, peak time, integral, RMS
  - currently using Hits associated to the neutrino interaction by Pandora
- Within each plane hits are connected in a graph using Delaunay triangulation
  - fully connected graph
  - both long and short distance edges
  - connect across unresponsive wire regions
- Hit associations to 3D SpacePoints create “nexus” connections across graphs in each plane
  - Currently defined by “Space Point Solver”
  - SPs are not connected among themselves
  - No input features for SPs

Network architecture

- NuGraph2's architecture is an iterative message-passing network.
- Each message-passing iteration consists of two phases:
  - Planar block: pass messages internally in each plane.
  - Nexus block: pass messages up to 3D nexus nodes to share context information.
- Messages are based on a categorical embedding:
  - Each semantic category is provided with a separate set of embedded features, which are convolved independently.
  - Context information is exchanged between different particle types via a categorical cross-attention mechanism.



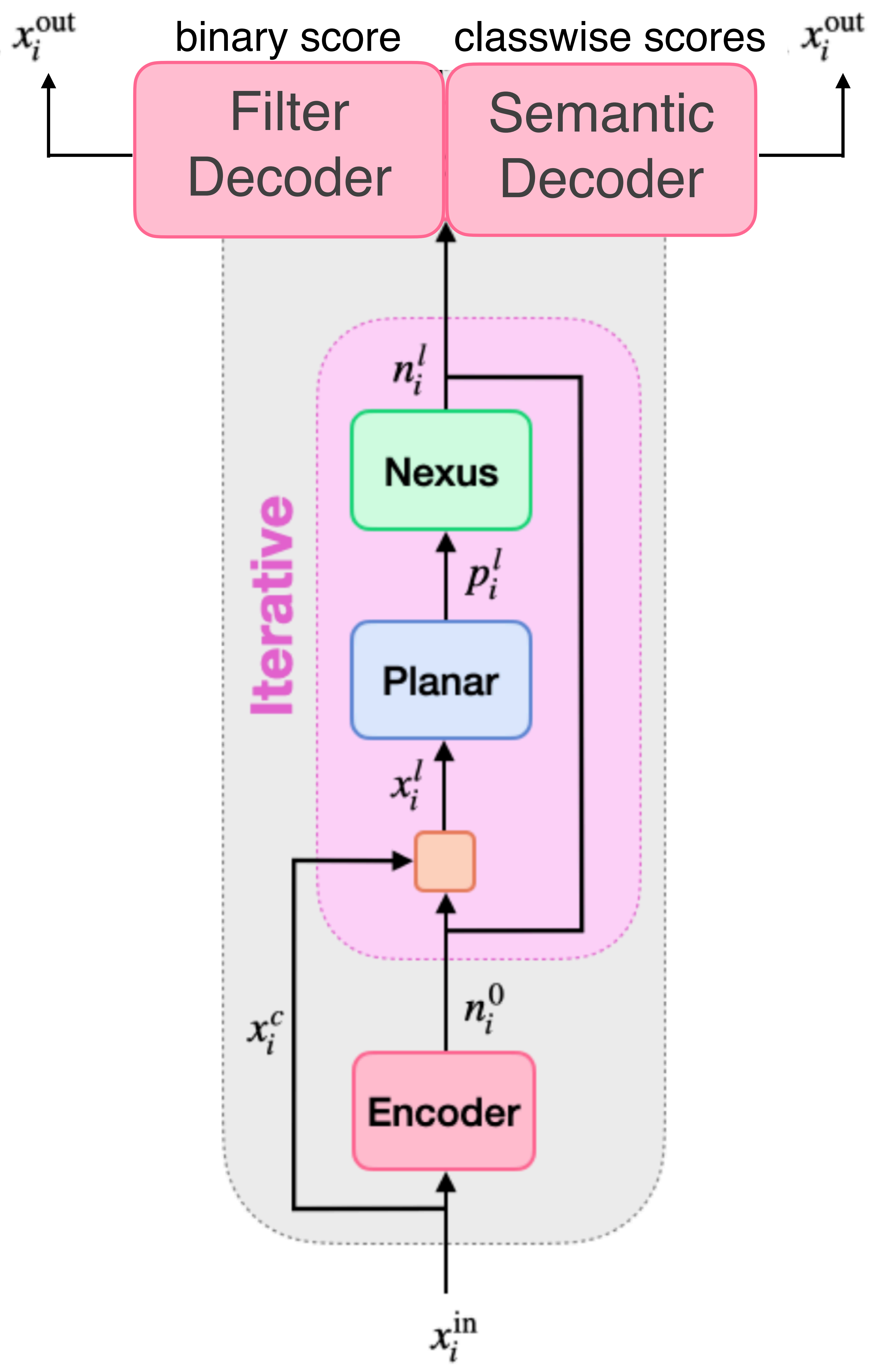
Decoders

- The last step at the end of the message passing network are the decoder steps

- Paper describes two node classifications decoders:
  - Semantic: classify each hit by particle type
  - Filter: separate hits from neutrino interaction from background
  - Output both class-wise scores from the semantic decoder and a binary score from the filter decoder
  - Same learned features are used as input to all decoders
  - Different loss functions weighted based on per-task variance (arXiv:1705.07115)

- Work in progress on more decoders: neutrino flavor, vertex regression, object condensation

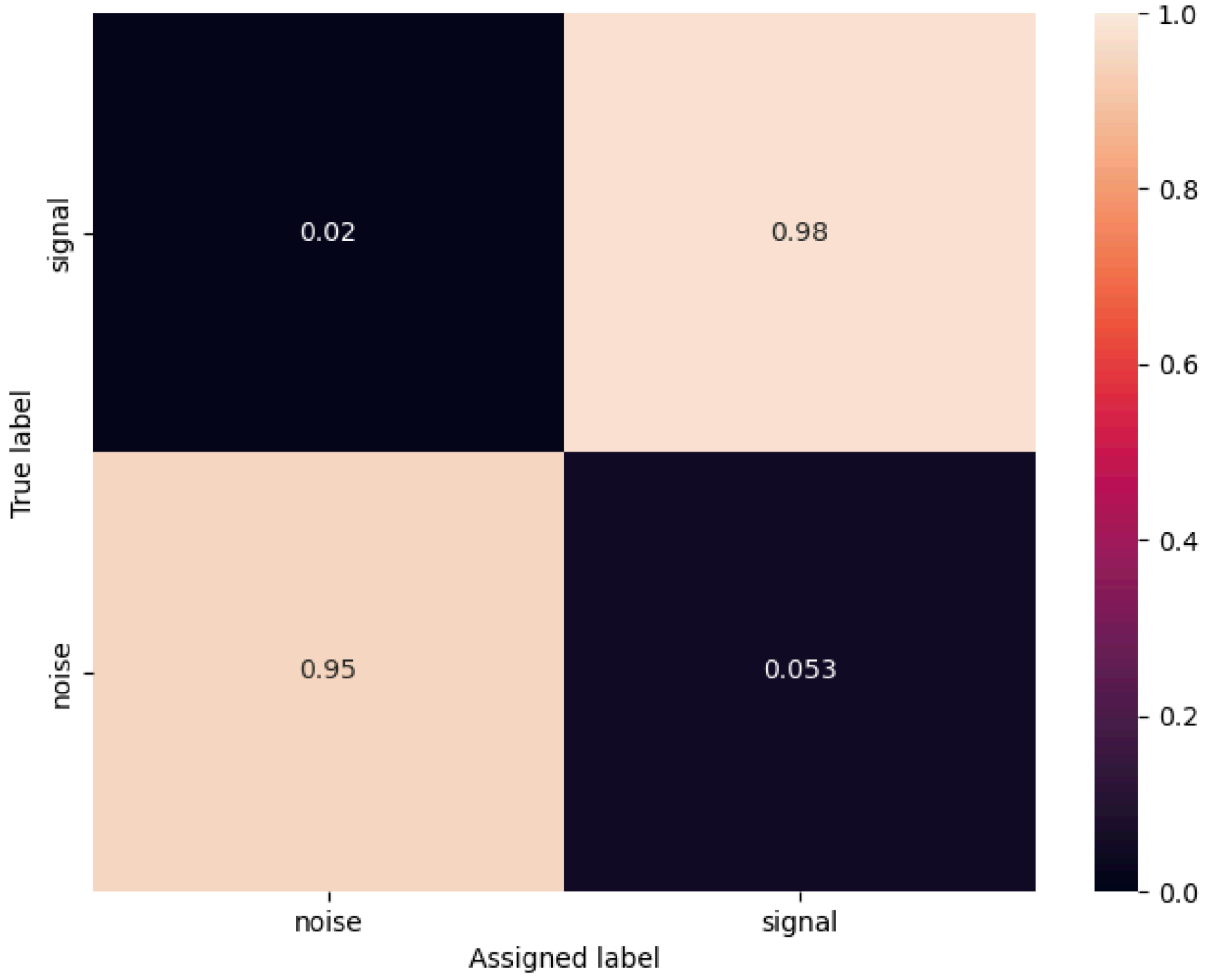
NuGraph2



Performance on Simulation: Filter

- Decoder trained to separate neutrino-induced hits from background (noise or cosmic-induced hits)
  - Pandora slicing tends to prioritize completeness over purity

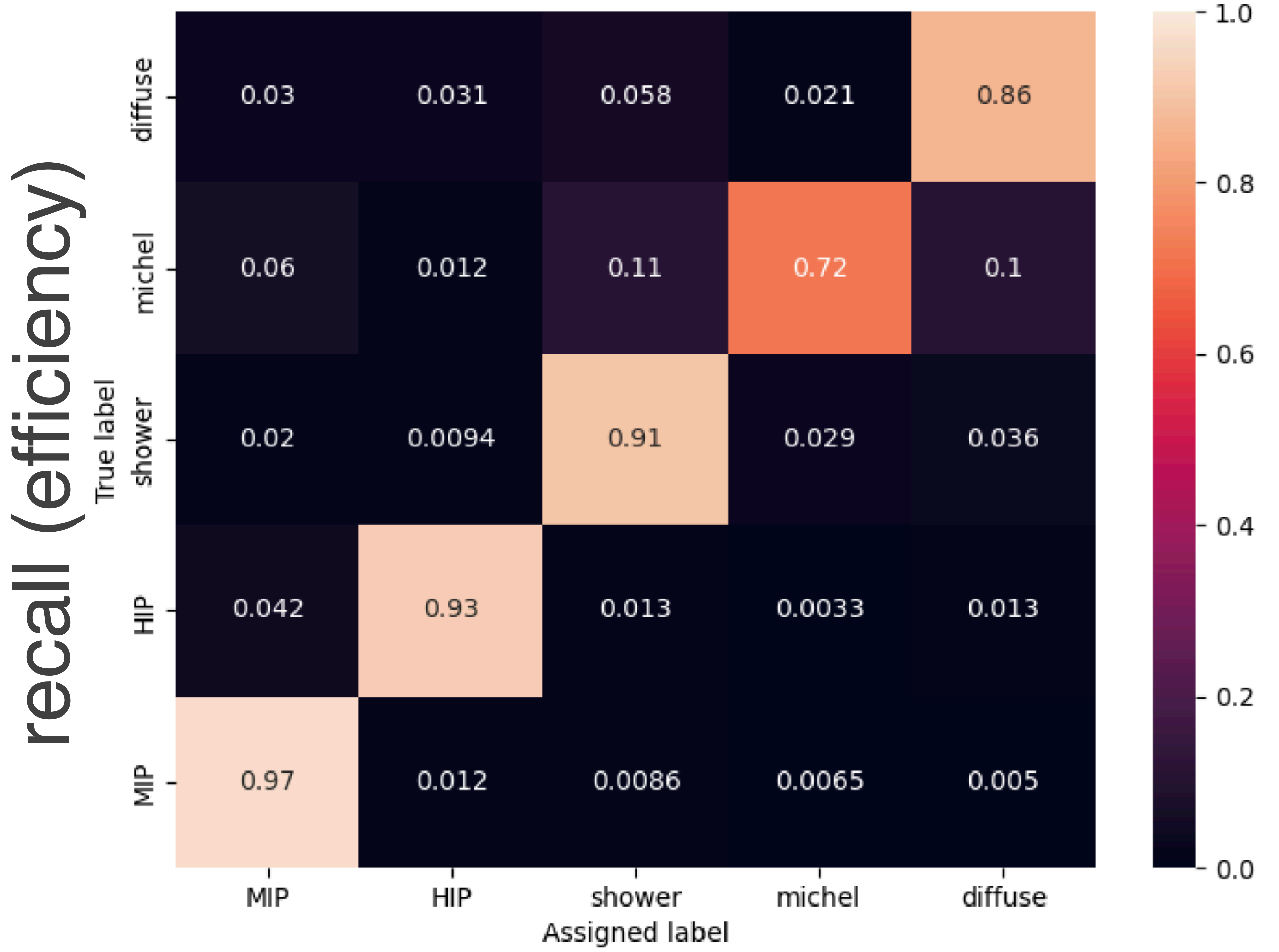
- Performance metrics:
  - recall and precision: ~0.98

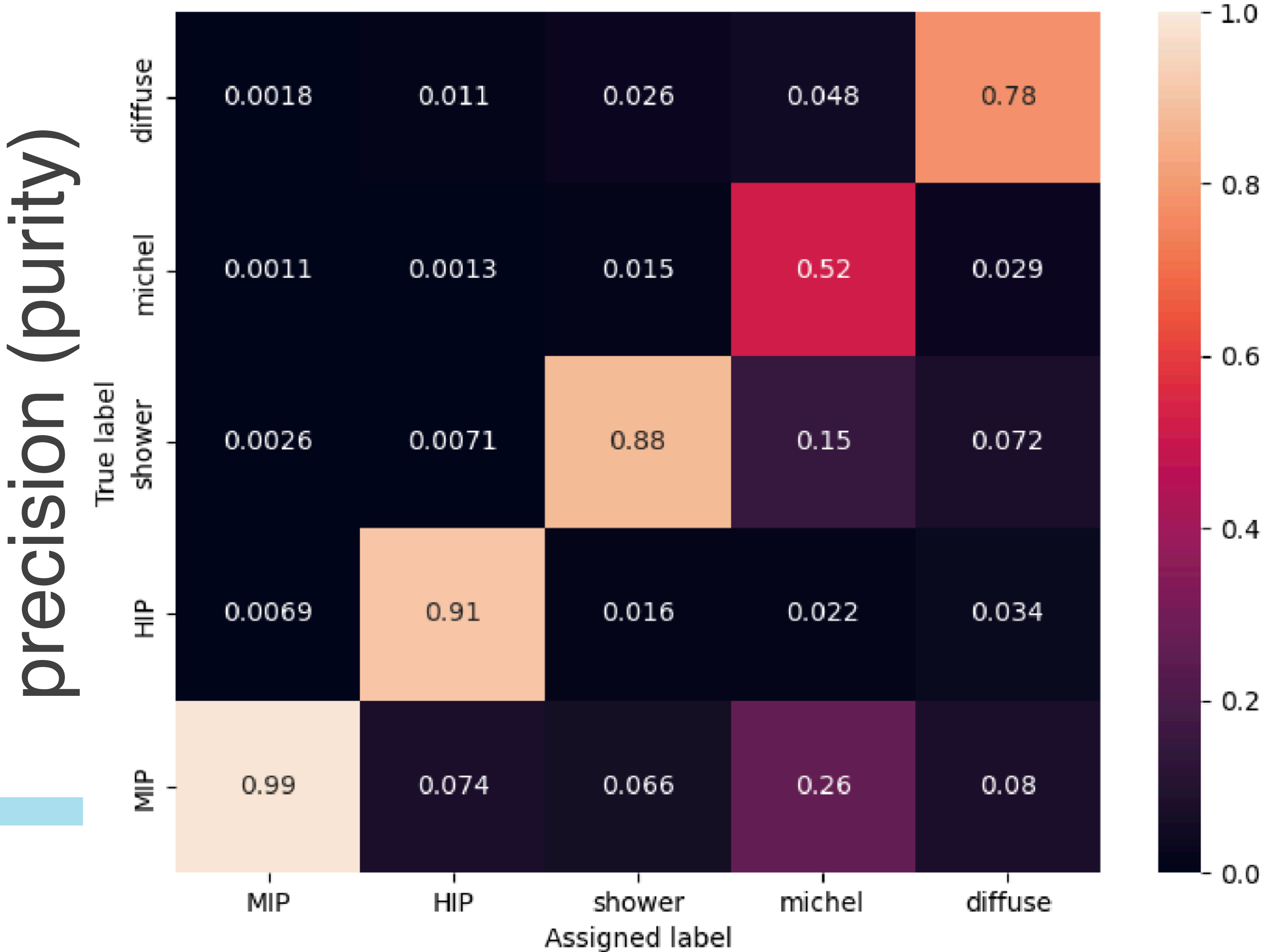


Fermilab

Performance on Simulation: Semantic

- Decoder trained to classify each neutrino-induced hit according to particle type
- Use five semantic categories:
  - MIP: Minimum ionizing particles (muons, charged pions)
  - HIP: Highly ionizing particles (protons)
  - EM showers (primary electrons, photons)
  - Michel electrons
  - Diffuse activity (Compton scatters, neutrons)
- Performance metrics:
  - recall and precision: ~0.95
  - consistency between planes around 98%
    - compared to ~70% without 3D nexus edges



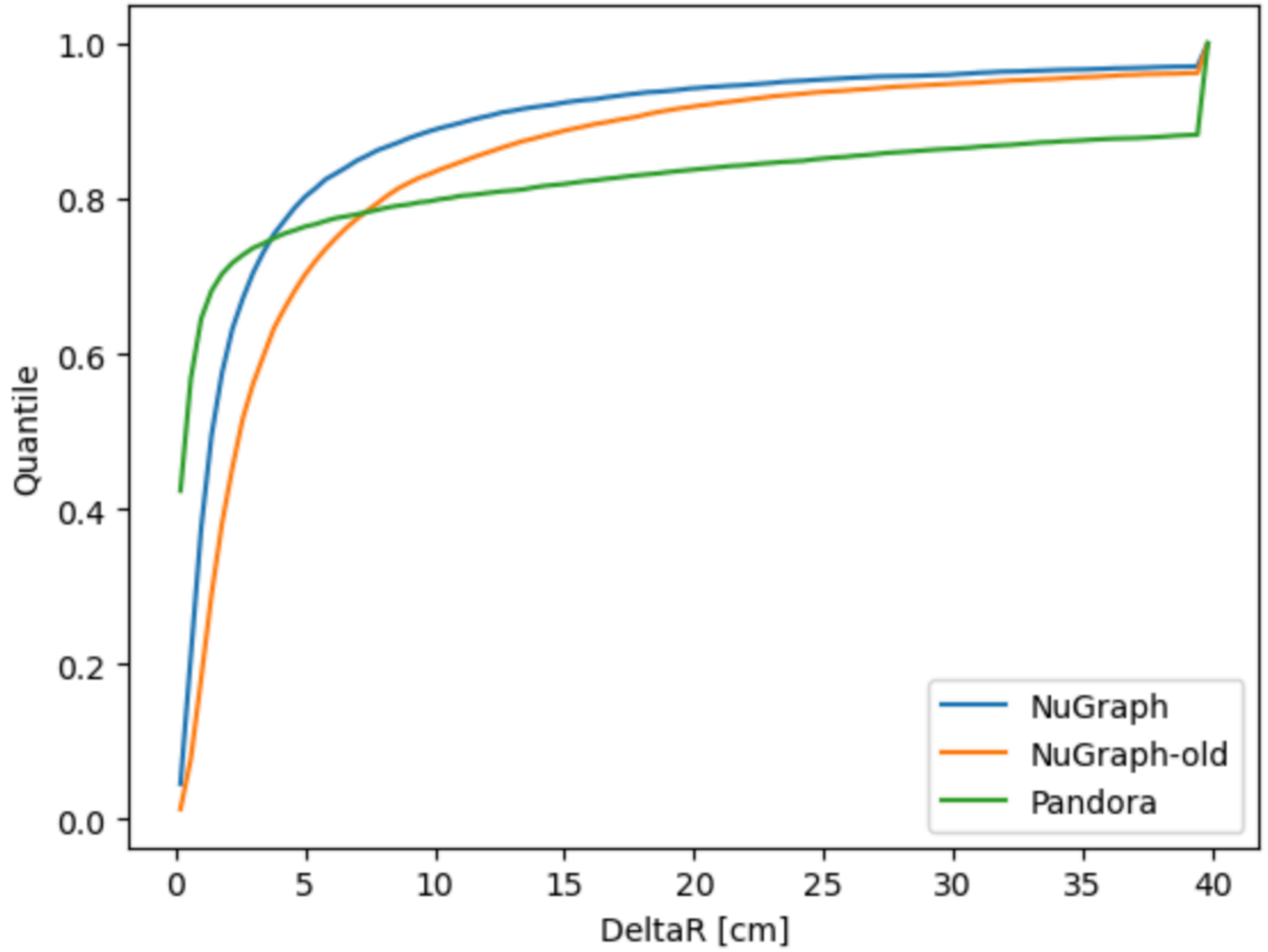


Vertex Decoder

- We also developed a vertex decoder regressing the 3D vertex position
  - Average 3D distance from truth:
    - NG: 6.2 cm
    - Pandora MCC9: 16.9 cm
  - Already better on average, but need to improve position pin-pointing

- Work in progress:
  - Need to consider different approaches wrt pure regression, as e.g. it does not constrain the vertex to hit positions in 2D (for CC interactions)
  - Tested different approaches for aggregating hit information into event-level, now moving to NuGraph3 (see next slide)

MicroBooNE Public Datasets

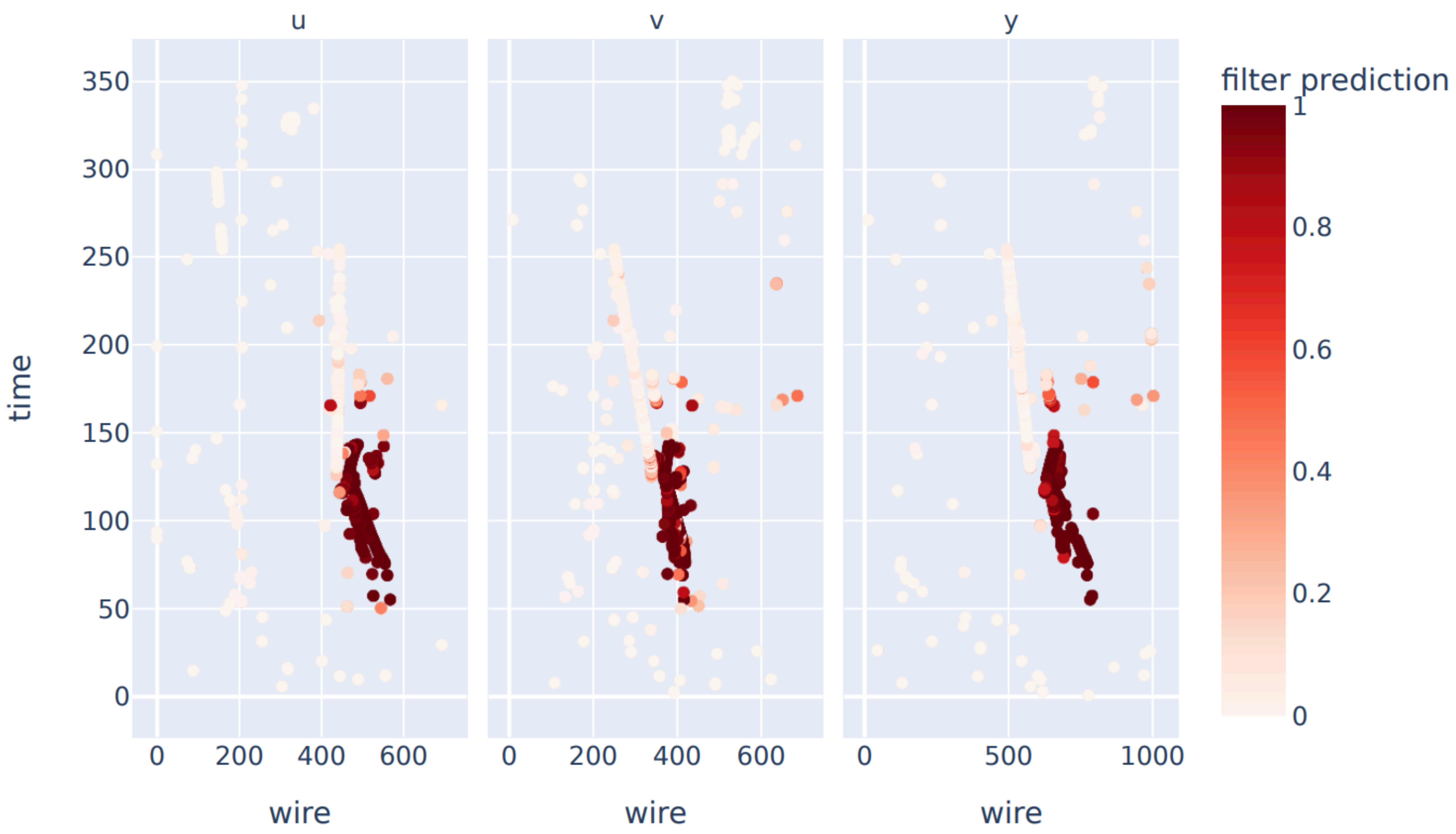


Fermilab

Performance on Simulation: Event Display

- Filter successfully rejects hits that are not from the neutrino interaction, including cosmic tracks that are close to it

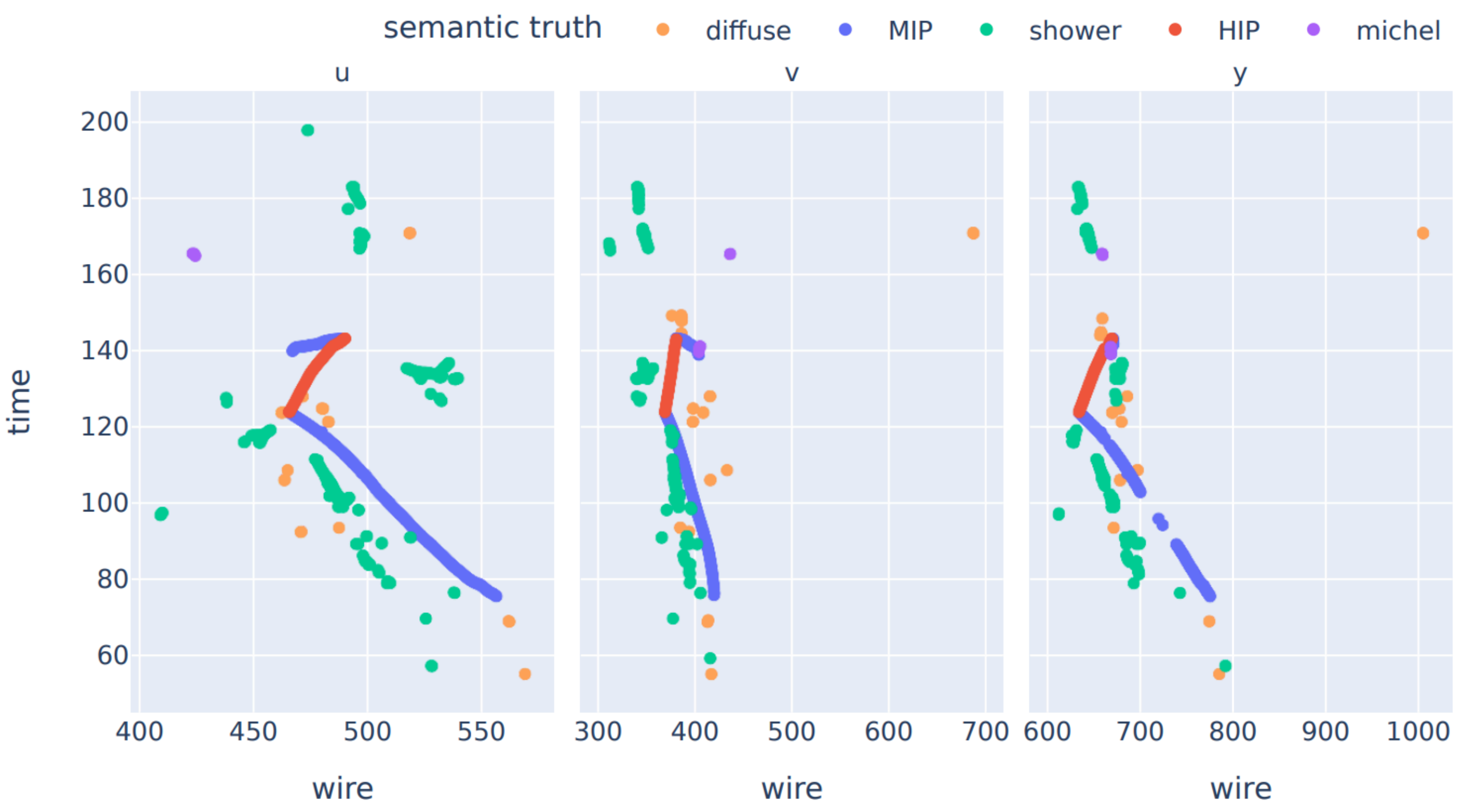
(a) Filter truth



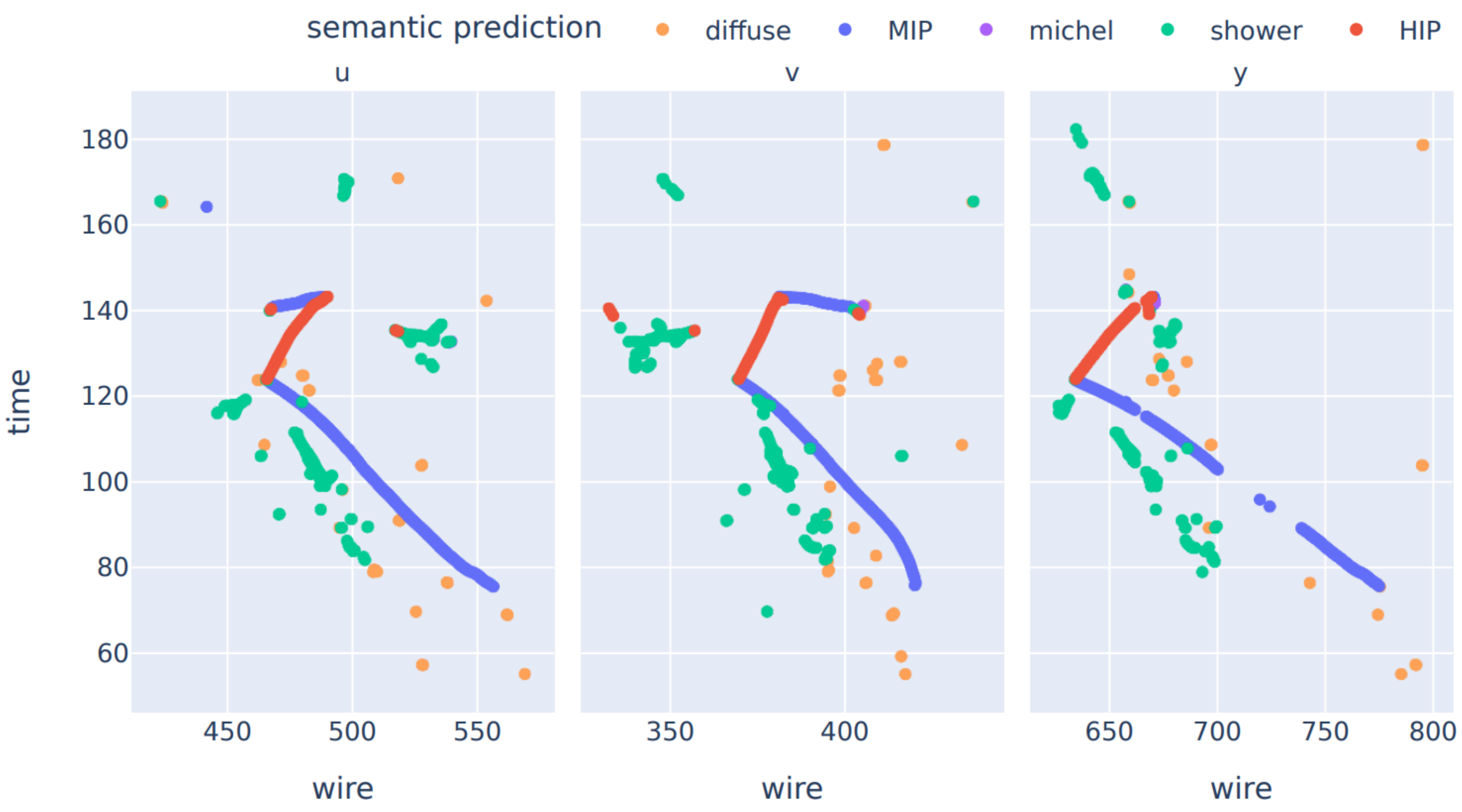
(b) Filter prediction

Performance on Simulation: Event Display

- Semantic classification correctly classifies hits classes both in events with a simple topology and also in higher multiplicity events.



(c) Semantic truth, filtered by truth

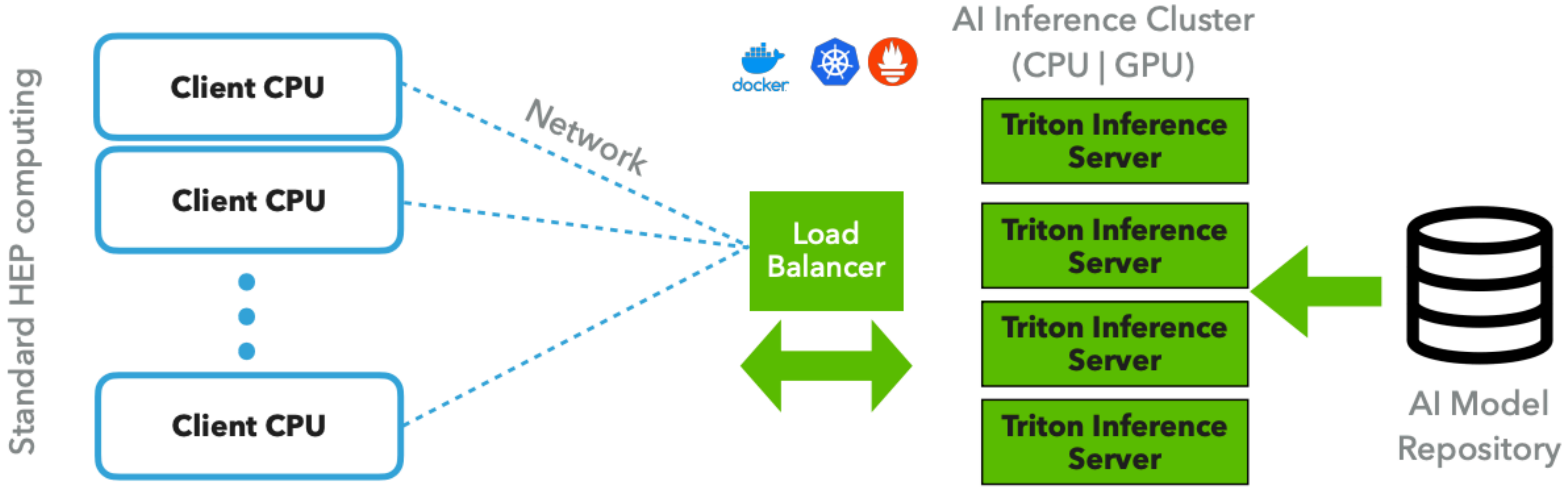


(d) Semantic prediction, filtered by prediction

Integration in LArSoft

- NuGraph2 is integrated in the software framework for LArTPC experiments, LArSoft
- Model compiled with JIT and run using the libtorch C++ library.
  - Integrated a package for Delaunay triangulation as well.
  - Inference results are stored in the Event record for usage in downstream reconstruction and analysis.
- Inference module takes 0.75 s per event event on CPU, including graph construction
- Enables running in production workflows for LArTPC experiments!
- Currently exploring more flexible integration methods based on NVIDIA Triton inference server (NuSonic: arXiv:2009.04509)

TimeTracker printout (sec)	Min	Avg	Max	Median	RMS	nEvts
Full event	0.0450458	3.36097	87.7468	0.237533	12.1151	74
source:RootInput(read)	0.000725606	0.00255304	0.019421	0.00131291	0.00392539	74
reco:nuslhits:NuSliceHitsProducer	0.0411265	0.116099	0.55599	0.0900547	0.0817036	74
reco:sps:SpacePointSolver	0.000110578	2.48479	85.3879	0.000217748	11.6239	74
reco:NuGraph:NuGraphInference	4.7356e-05	0.74844	5.22709	8.83935e-05	1.14997	74
[art]:TriggerResults:TriggerResultInserter	1.4952e-05	2.38511e-05	6.7179e-05	2.1032e-05	9.54137e-06	74
end_path:rootOutput:RootOutput	2.915e-06	4.5257e-06	1.9485e-05	3.9445e-06	2.18303e-06	74
end_path:rootOutput:RootOutput(write)	0.000867838	0.008697	0.0783238	0.00176224	0.0132931	74

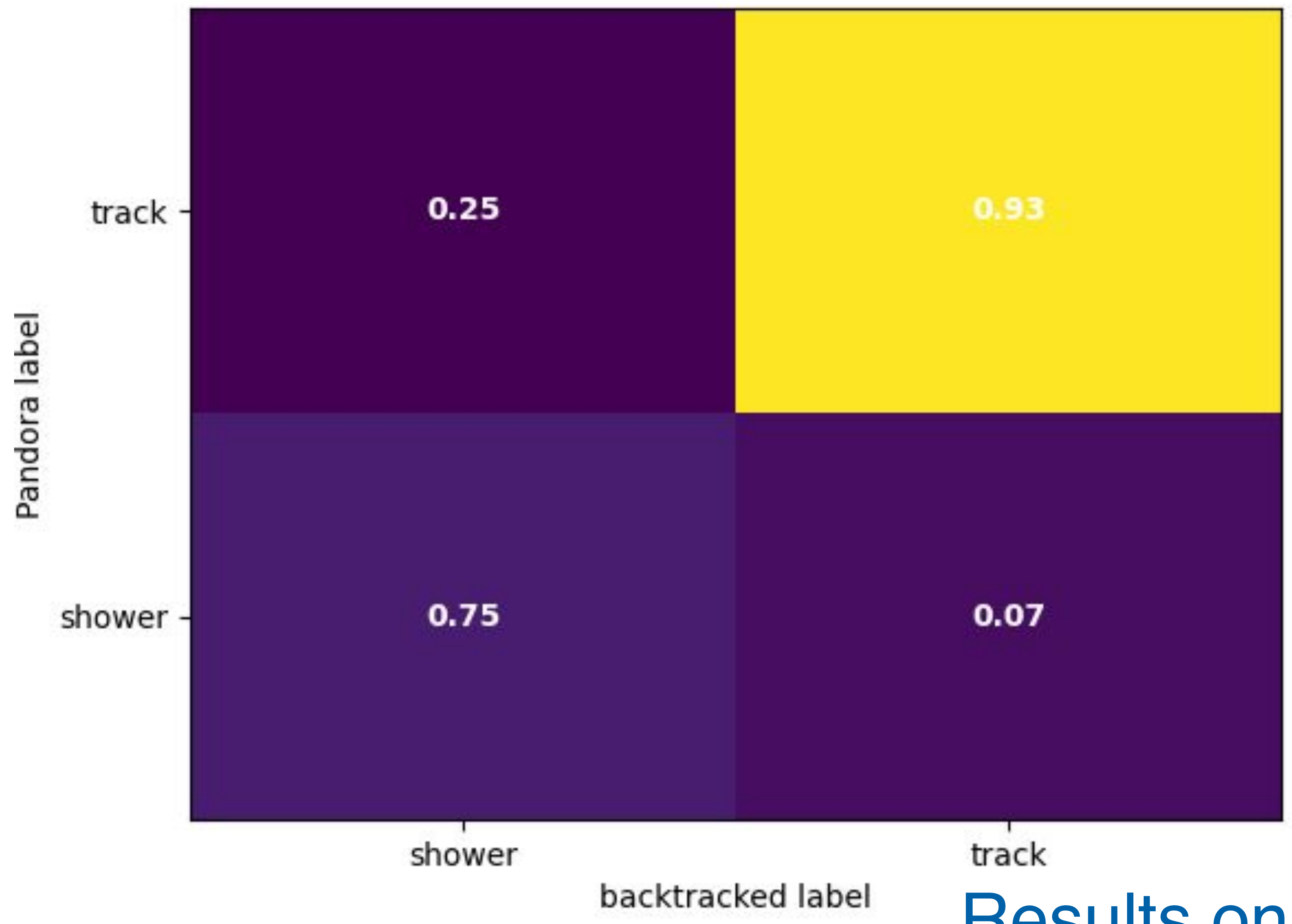


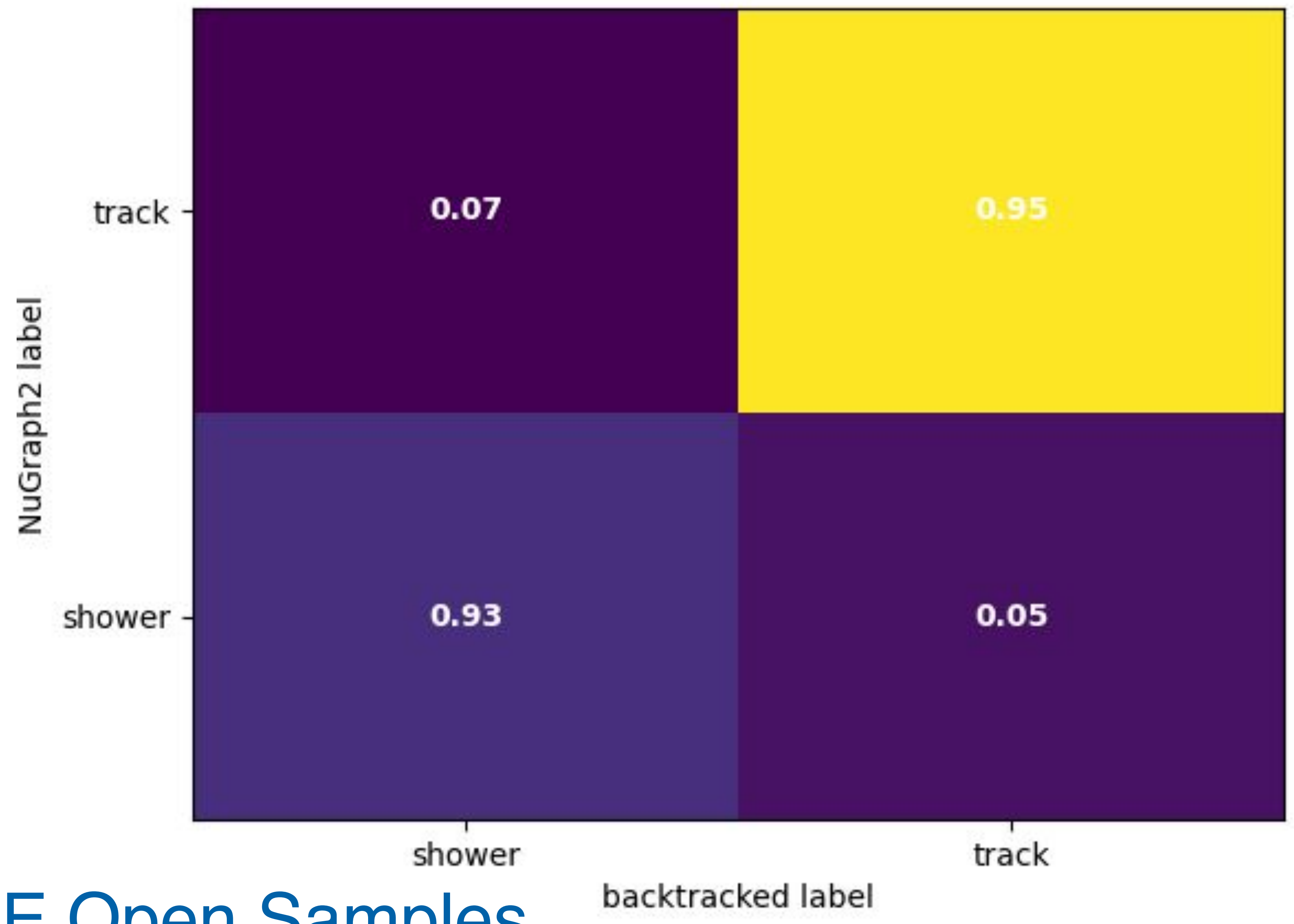
Fermilab

Collaboration with
D. Caratelli and M. Fang

NuGraph2 for Track/Shower Separation

- NuGraph2 "shower" label significantly improves the Pandora classification out of the box
  - ~95% accuracy on track identification
  - ~20% improvement on shower classification





Results on MicroBooNE Open Samples

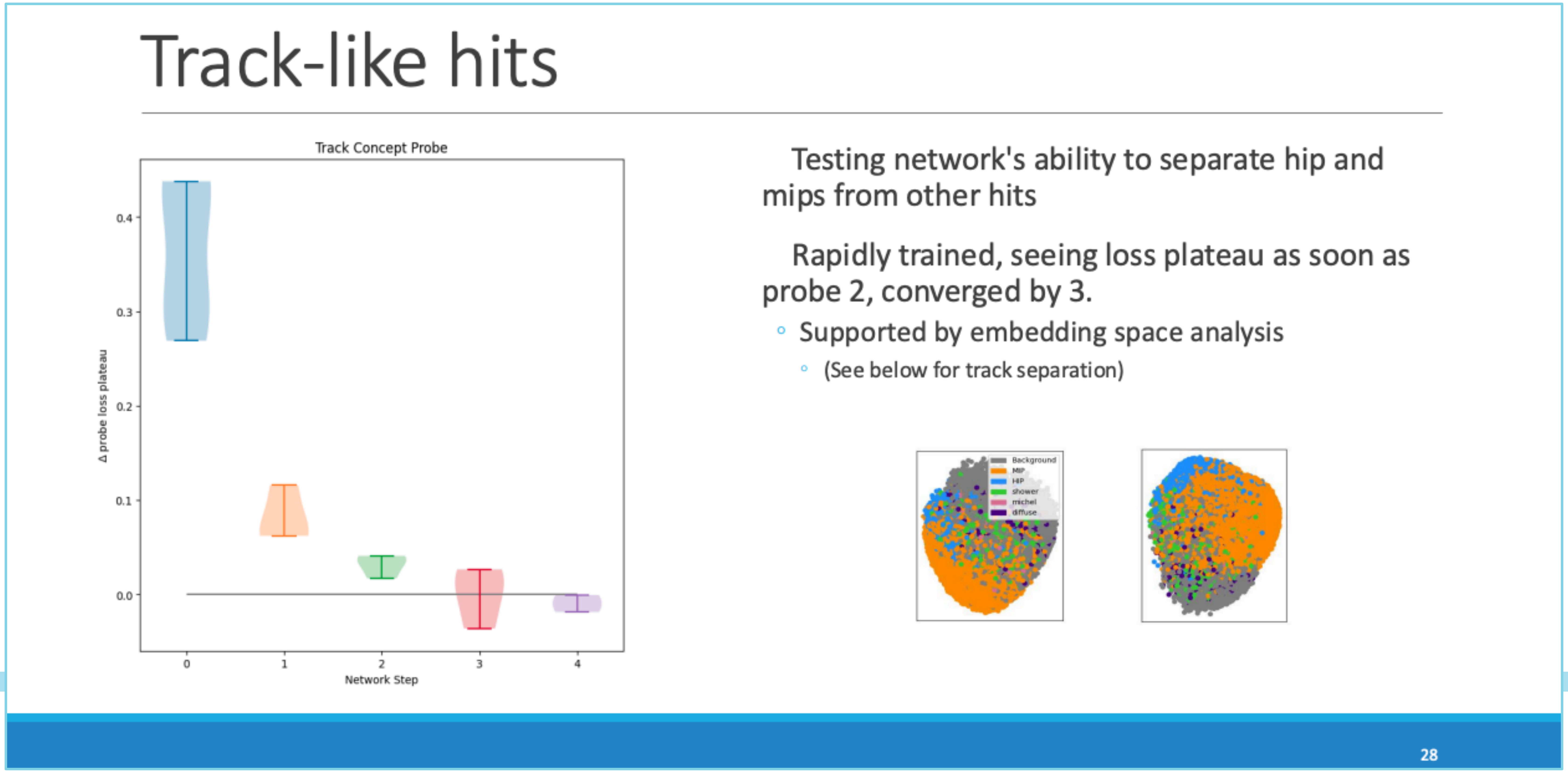
Status of NuGraph applications

- MicroBooNE
  - Ongoing integration in “MCC10” reconstruction workflow
- ICARUS
  - First training completed! Thanks to S. Seo and interns R. Campos and E. Novello
- DUNE
  - Several ongoing developments:
    - Tau neutrino reconstruction
    - Proton decay reconstruction
    - Supernova pointing

Network Explainability

- Explainability: Goal is to “open the black box” to build confidence and drive developments.
  - “Standard” tools for GNN interpretability (e.g. GNNExplainer) struggle with our network

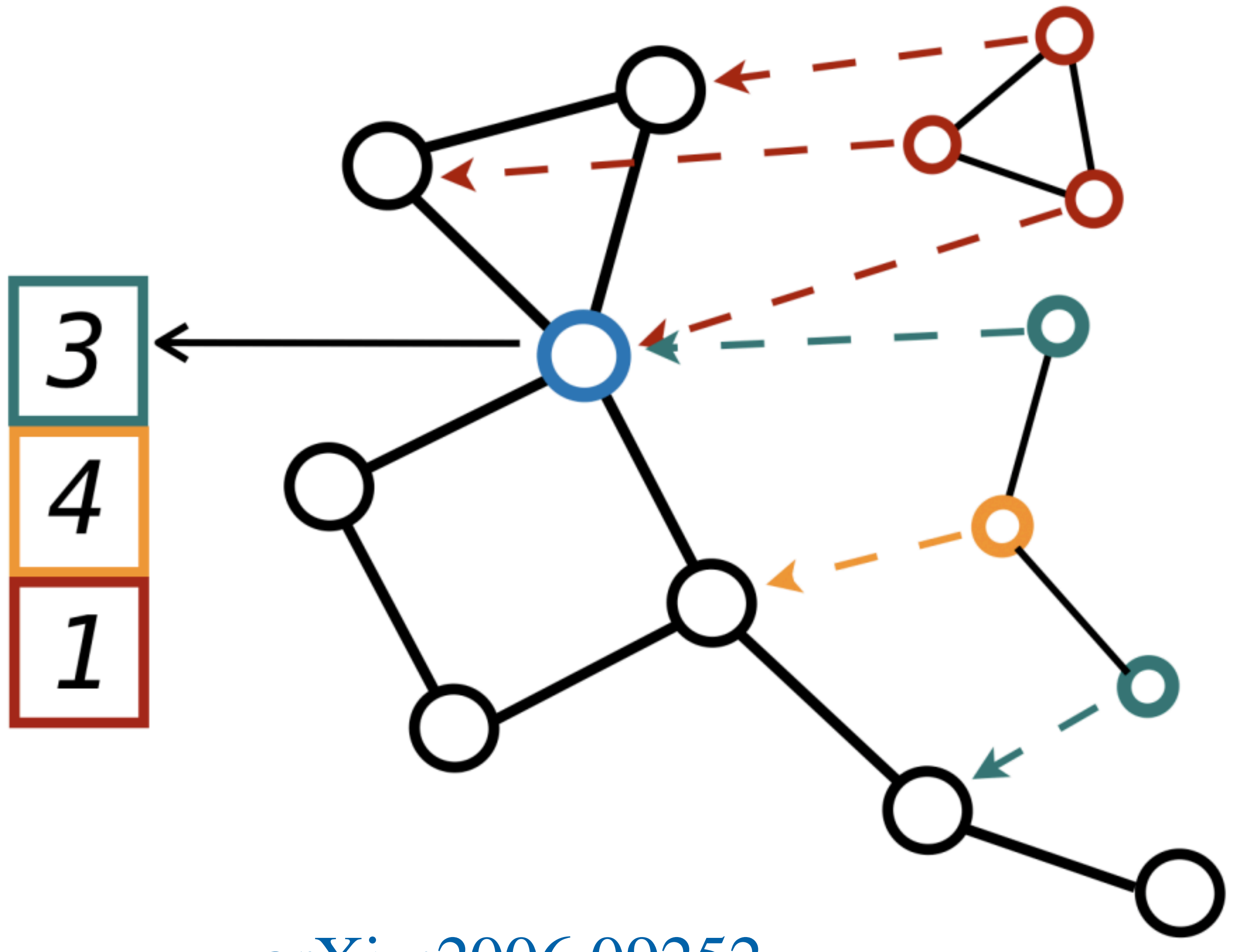
M. Voetberg, https://indico.fnal.gov/event/66124/contributions/301004/attachments/182262/250229/exatrkx%20workshop%20-%20graph%20explanations-1.pdf



Fermilab

Injecting Physics Domain Knowledge: Augmented Features

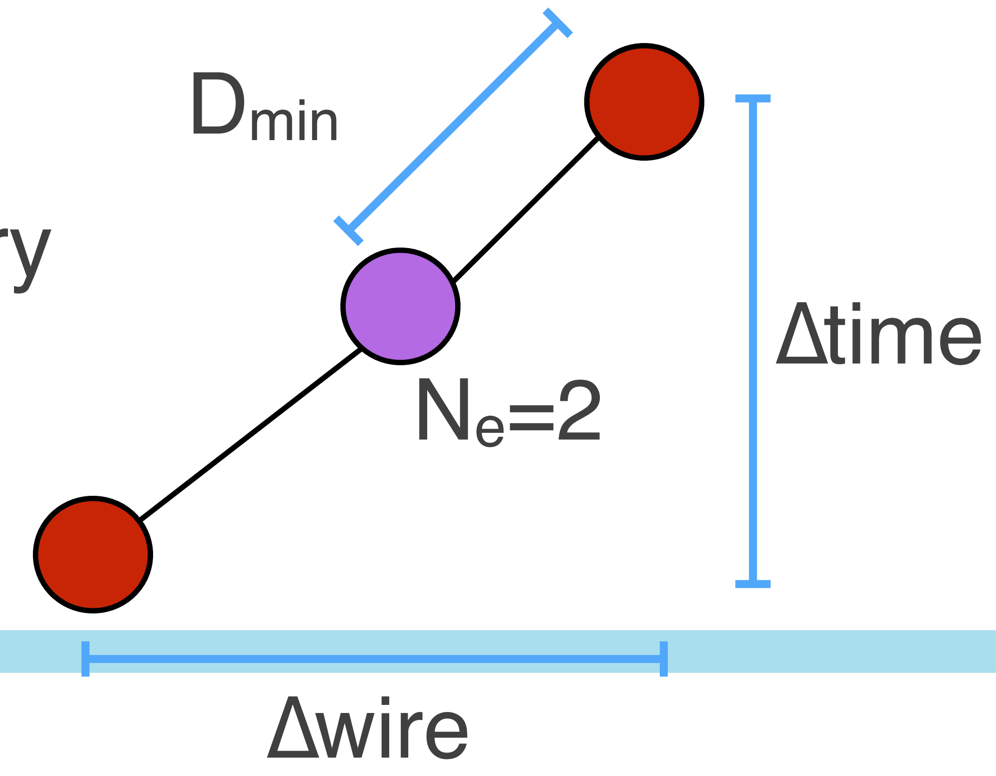
- It turns out that GNNs are not aware of the structural role of nodes
  - They do not learn the graph structure
  - GNNs do not distinguish graphs that are isomorphic according to the Wesfeiler-Lehman test



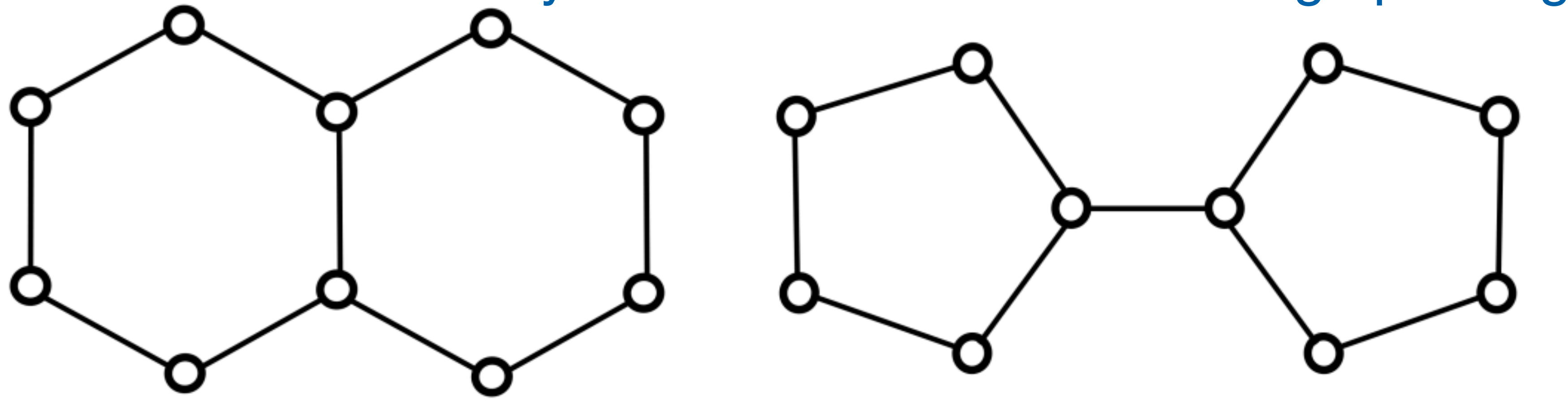
arXiv:2006.09252

- Adding the graph structural information (e.g. triangles, circles) may help with classification
  - This can be implemented by a structure-aware message passing which contains structural information about the nodes

- Add structural and non-local features to nodes improves the network performance across the board:
  - Features: Δtime, Δwire between 2 closest nodes, distance to closest node D_{min}, edge multiplicity N_e
  - ~5% (relative) improvement for the Michel category



Indistinguishable molecules by the WL test and thus message passing NN



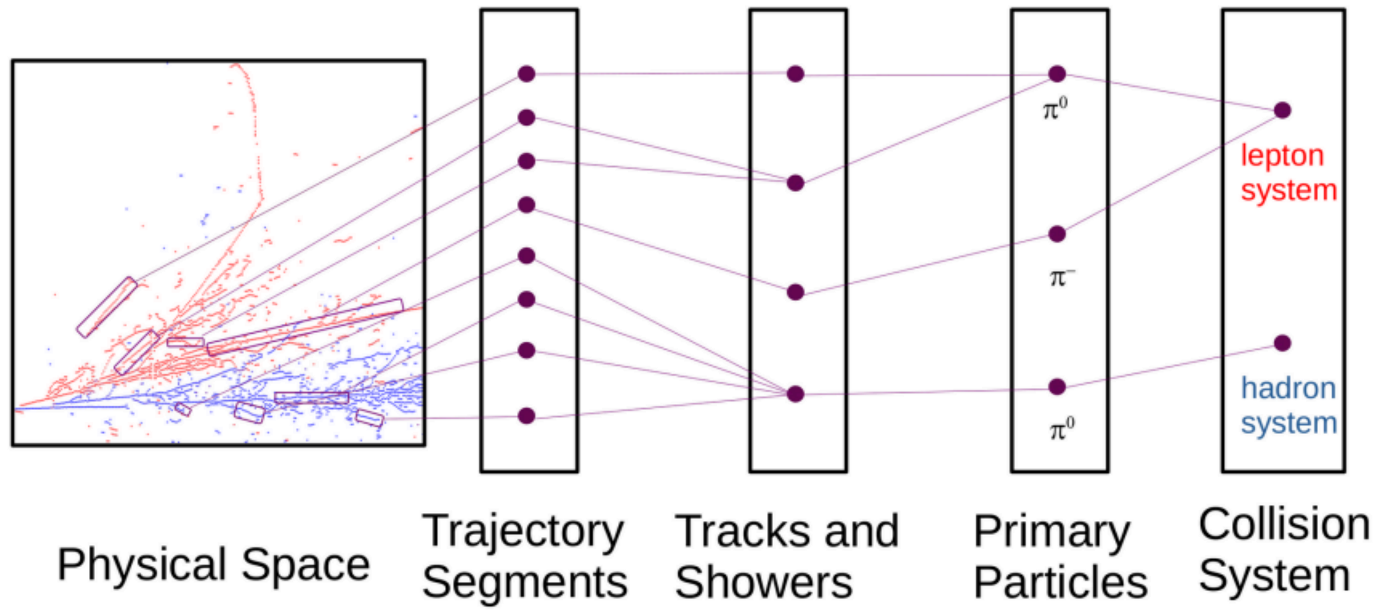
Work by V. Grizzi, H. Meidani (UIUC)

NuGraph3

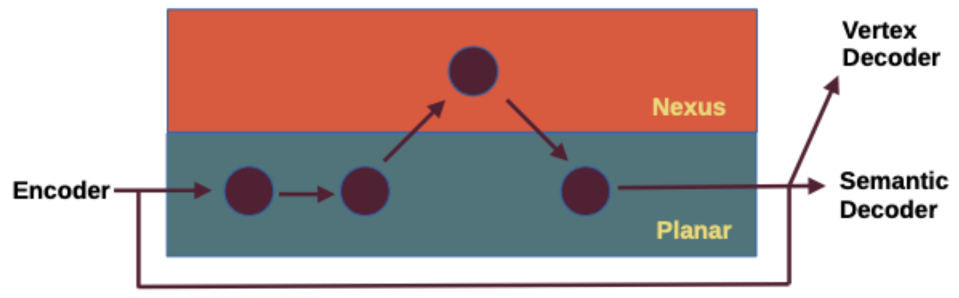
- A. Aurisano @ NPML:
  - https://indico.phys.ethz.ch/event/113/contributions/836/attachments/516/1110/aurisanoNuGraph3NPML.pdf

NuGraph3 Concept

- GNN-based particle flow reconstruction using NuGraph2 as starting point
- Similar to Pandora, consider series of reconstruction stages
- Each stage connects elements from stage before to produce higher level objects
  - Reconstruction chain expressible as a hierarchical graph with each level representing a reconstruction stage
- Avoid lossy serial steps by keeping many plausible reconstruction hypotheses and resolving them simultaneously
  - Expressible through fuzzy membership
    - Nodes on level L-1 can be connected to more than one node on level L
- Hierarchical message passing iteratively improves the particle tree reconstruction by choosing a reconstruction hypotheses using information from all stages simultaneously

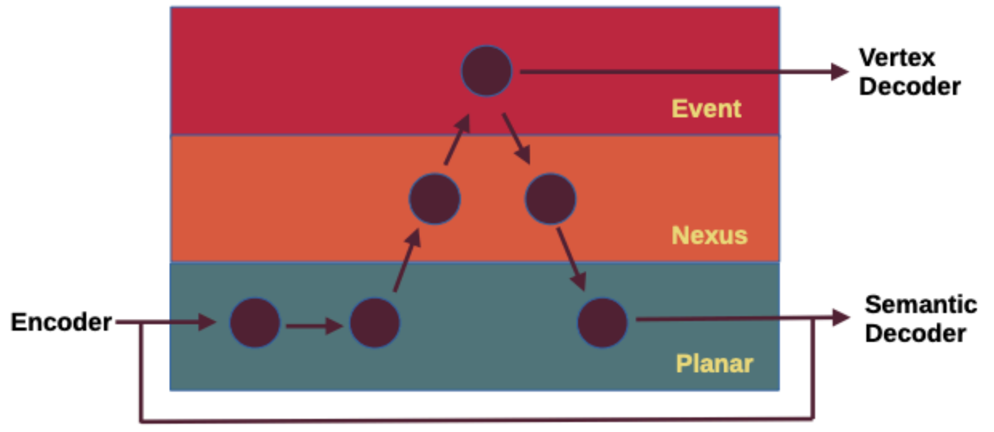


Hierarchical Message Passing



- NuGraph2 consisted of planar and nexus nodes connected in a pseudo-hierarchical fashion
- Nexus nodes primarily provided a way for enforcing consistency between semantic segmentation in each view
- Predicting event-level information was only possible through an aggregation layer (LSTM, transformer, etc)

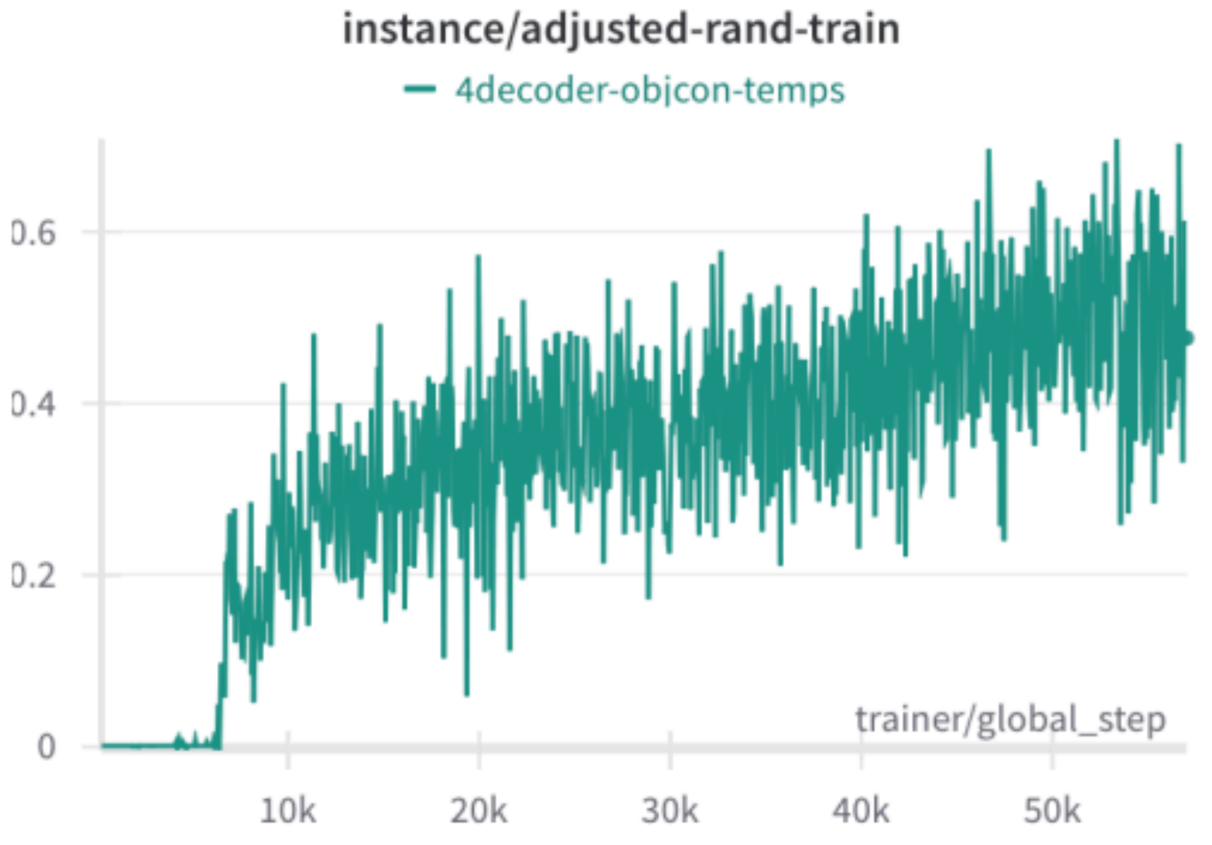
- To test hierarchical message passing, added an event layer with a single node
- Message passing with learned edge weights between nexus nodes and the event node allows for lightweight and smart aggregation



University of CINCINNATI

Clustering

- Utilize **object condensation** to cluster together detector hits into particle instances (2002.03605).
- Materialize object condensation embedding inside model to explicitly generate particle nodes.
- Currently performing this step during **instance decoder forward pass**.
- Naive implementation is not well-optimized, so currently optimizing for **memory overhead**, **speed** and **performance**.
- Ultimately plan to materialize instances inside core message-passing loop, so particle instance nodes can **replace nexus nodes** as the intermediate step in the hierarchy.

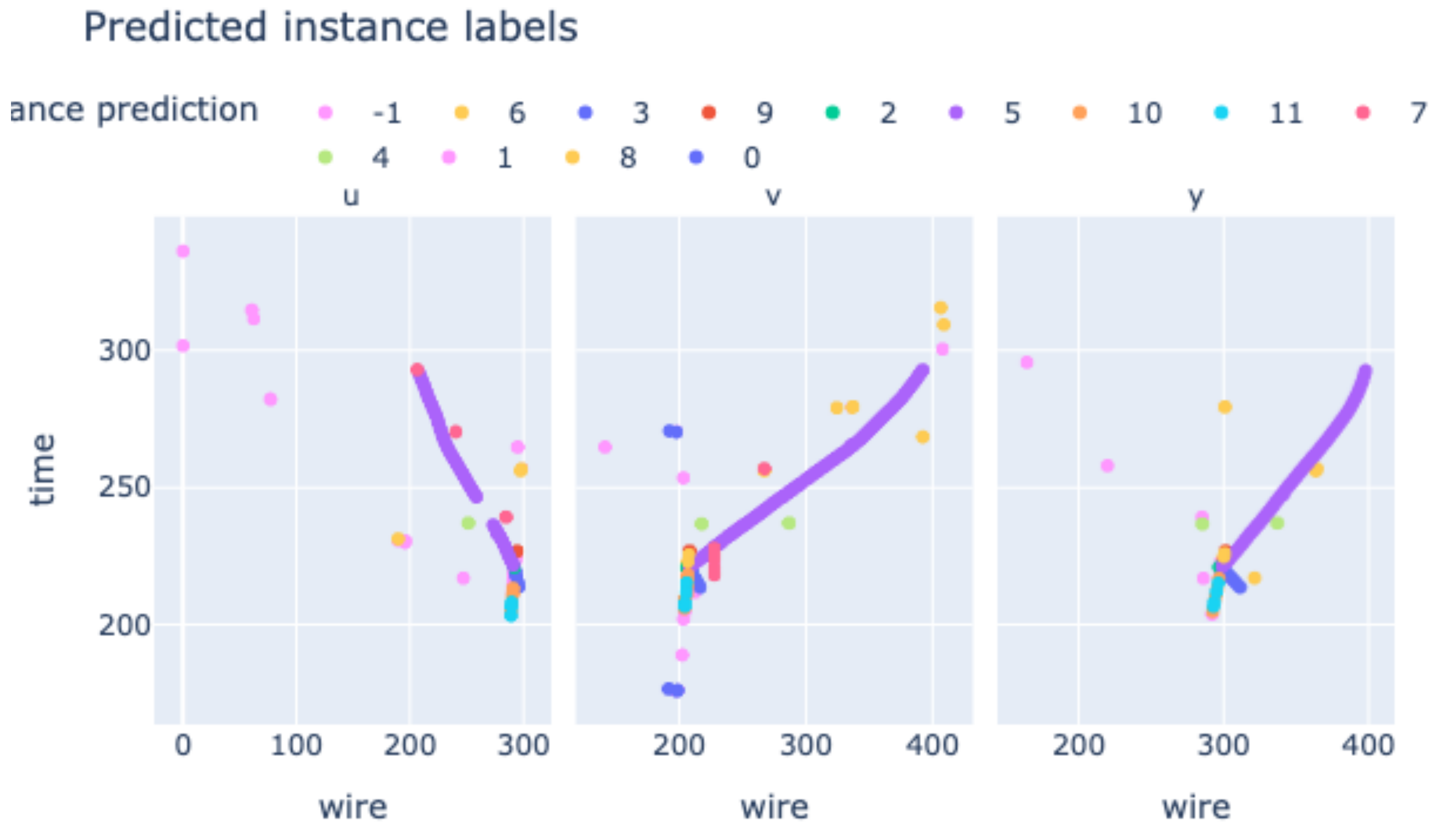


Quantify clustering performance in terms of **Adjusted Rand Index (ARI)**

0 = random clustering
1 = perfect clustering

NuGraph3 – v – 3rd October 2024 11

Recent developments…



Multi-modal network — Adding PMT detector information

Graph Architectures

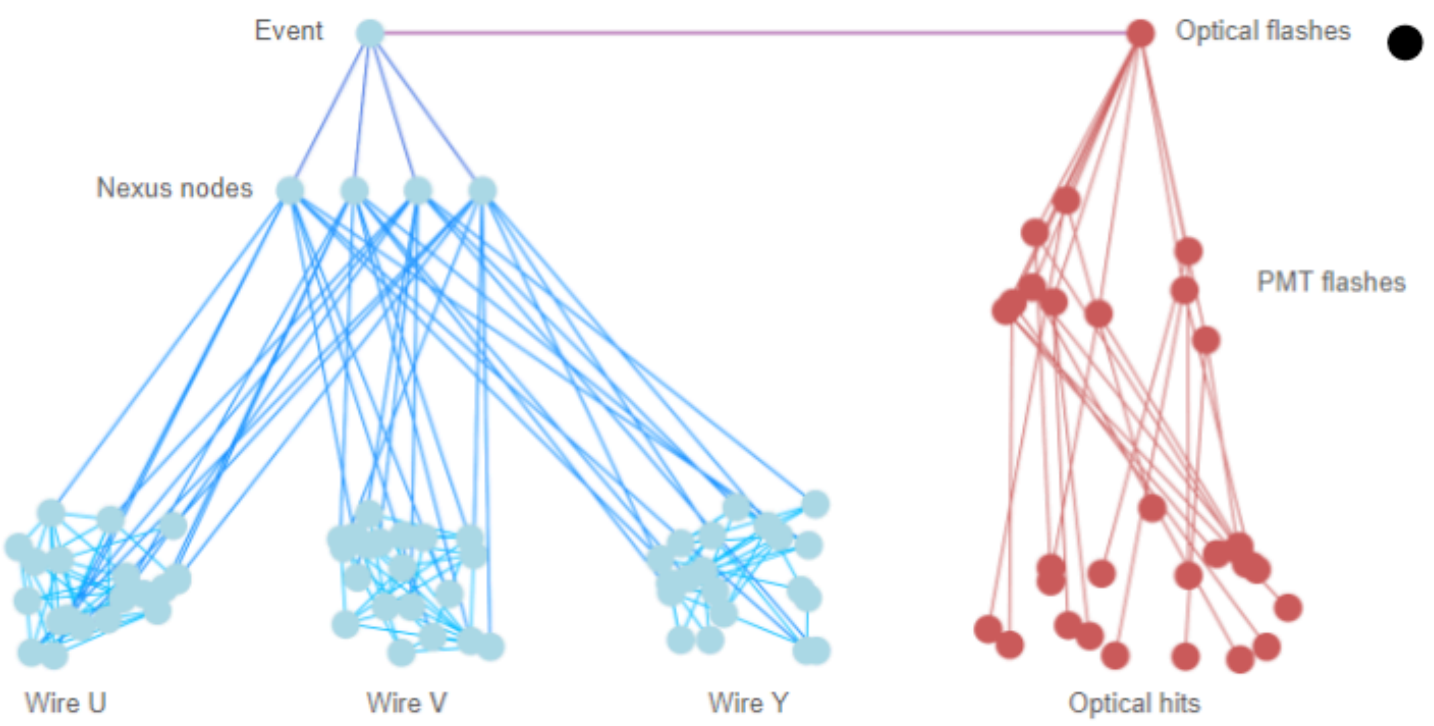


Figure 2. Graph Architecture A

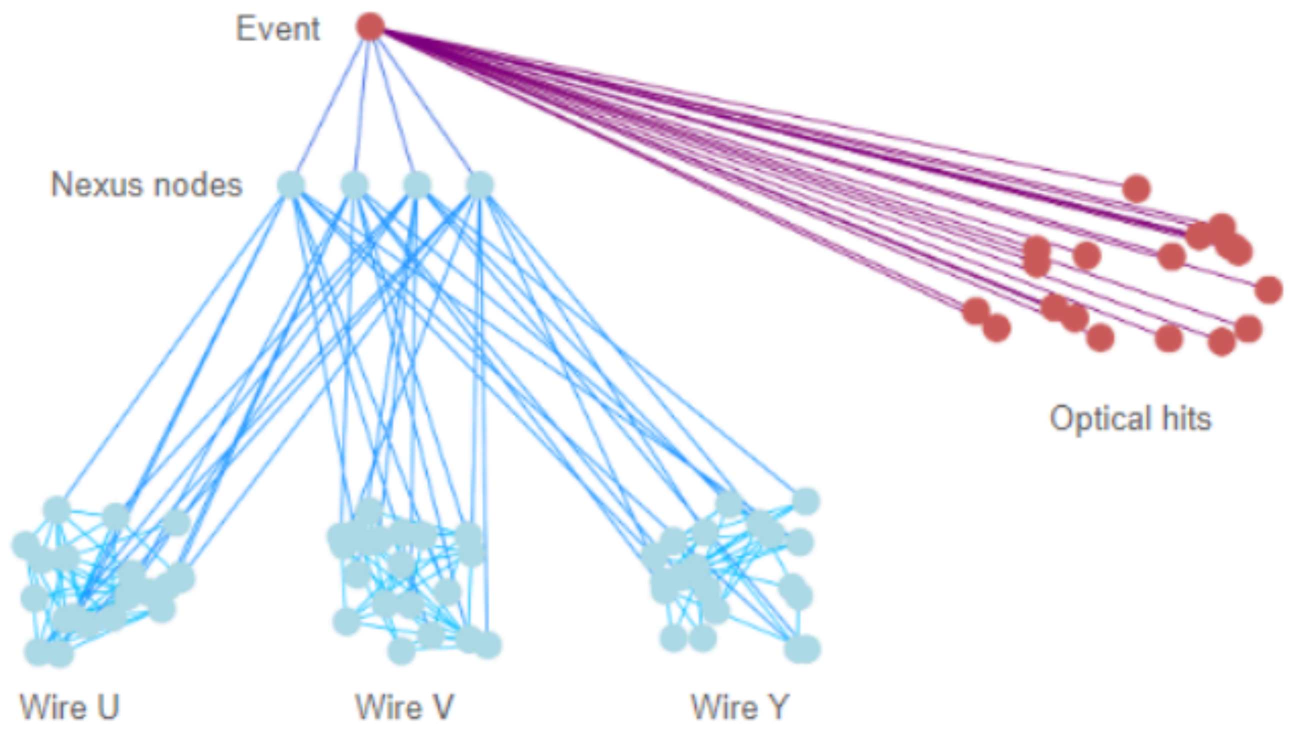


Figure 3. Graph Architecture B

- The blue graph represents the original graph, which consists of connected nodes from Wires U, V, and Y. Each particle hit in the wire is aggregated to a nexus node and finally an event.
- Graph A aggregates the optical hits to one of the 32 PMTs which is further aggregated to one flash representing the interaction.
- Graph B directly connects the hits to the event node.

Results

- After incorporating the optical data, the model was trained successfully and yielded similar results to the original one.
- Although improvement were expected, this opens the door for experimenting with different hyperparameters and graph connections to maximize performance.

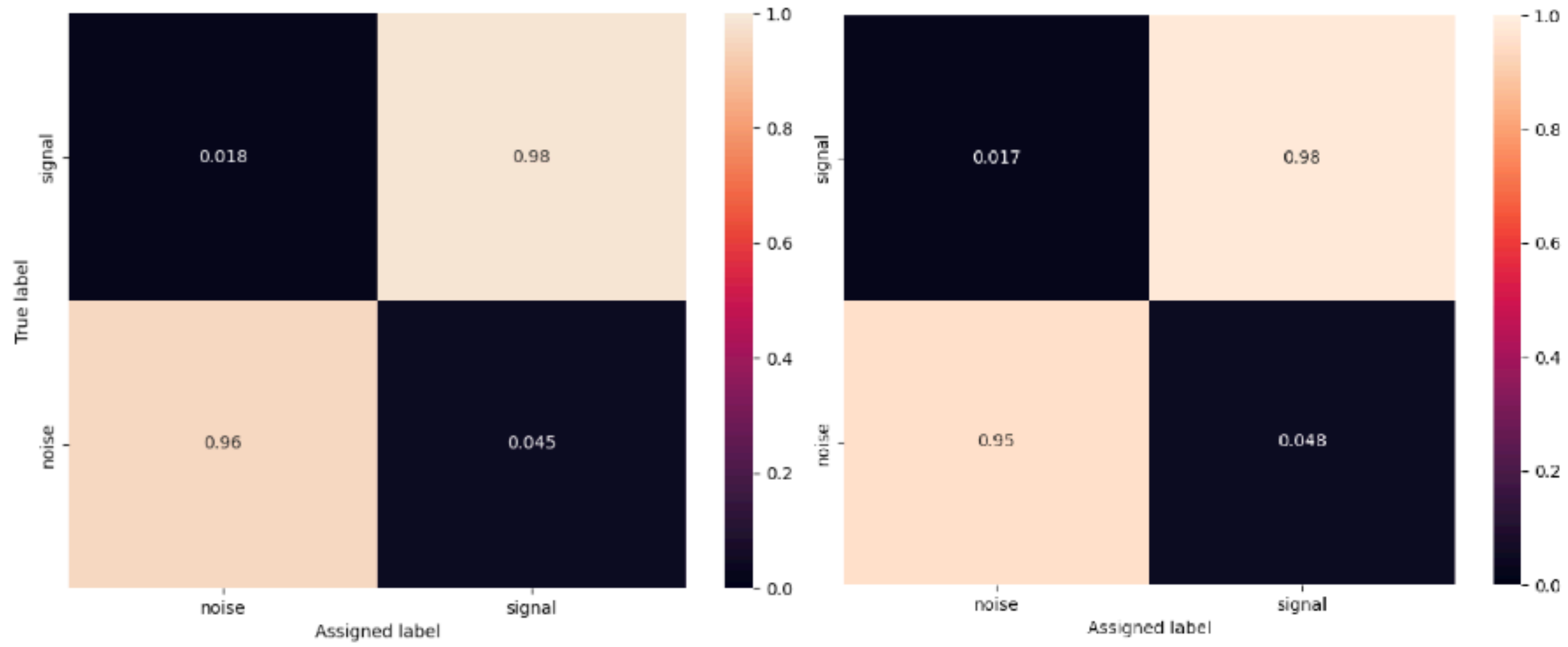
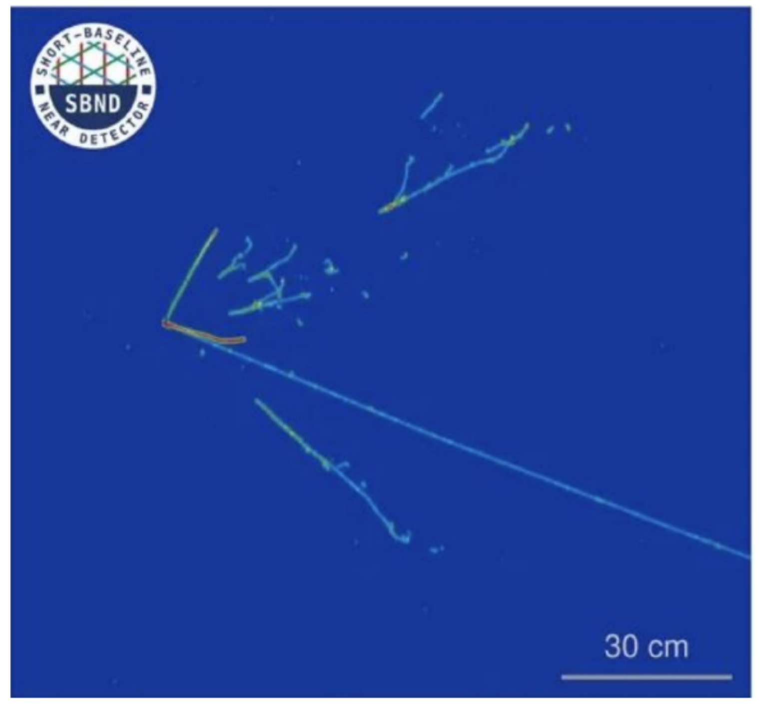


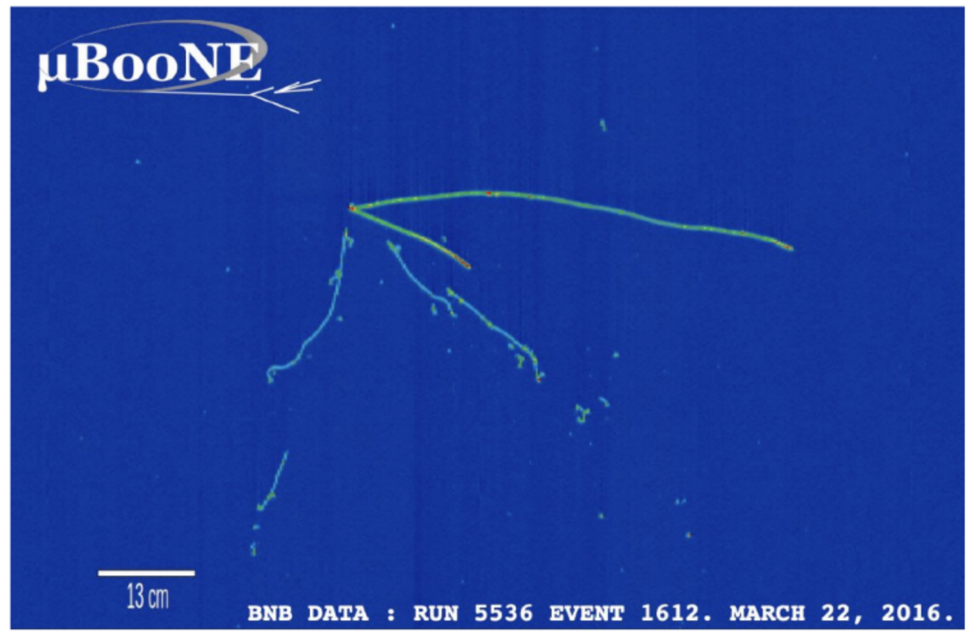
Figure 5. Comparing the Filter Recall (Efficiency). The plot to the right is with the optical data, and the one to the left is without it.

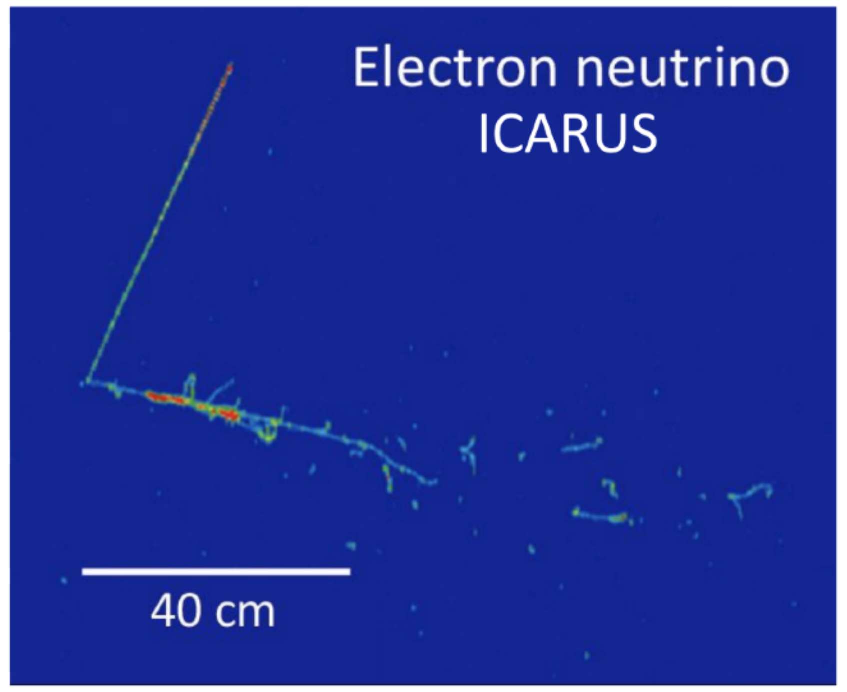
Potential usage for “interaction” decoder, e.g. for DUNE ND.

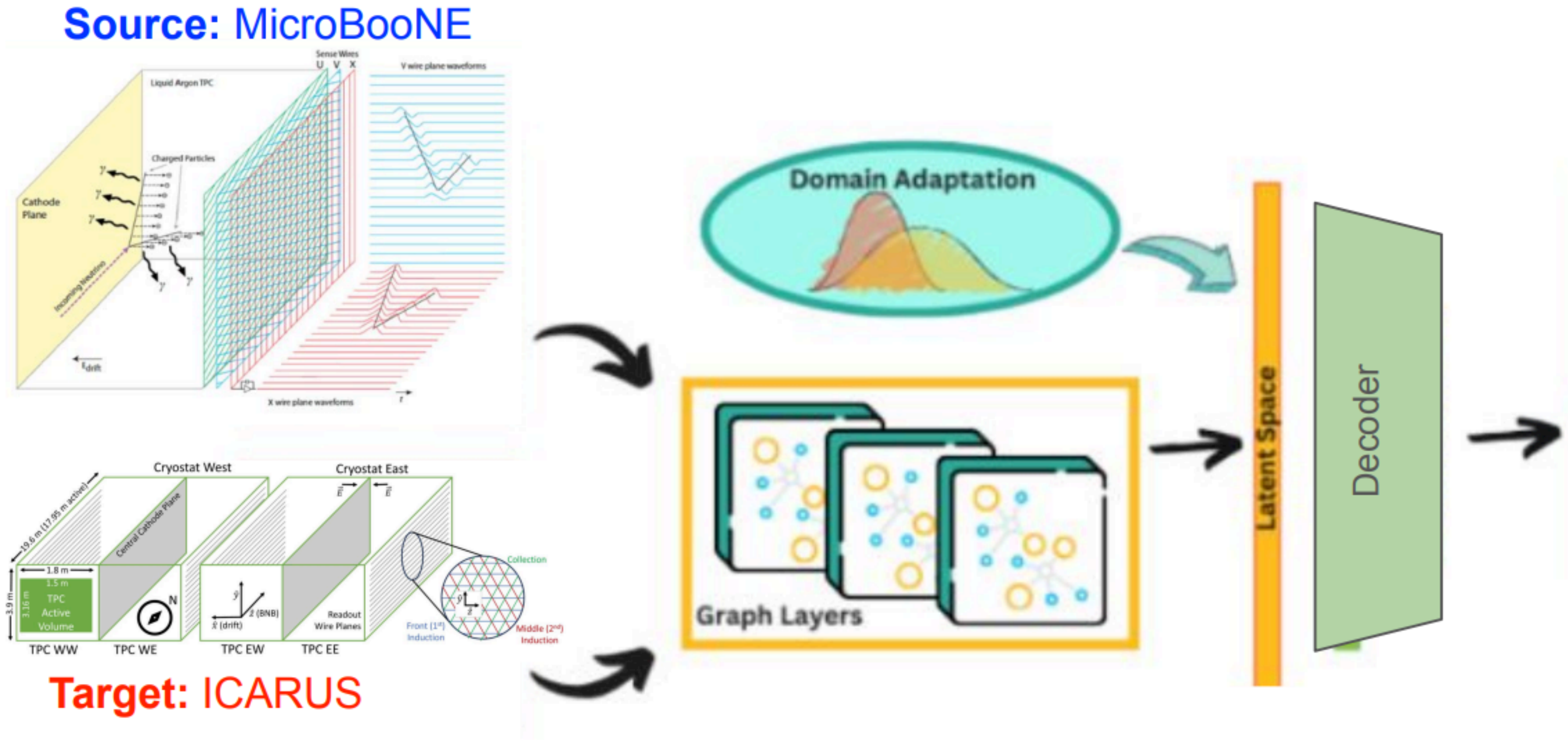
Domain Adaptation

Particles and the overall event looks the same in different detectors.
Image credit: M. Toups

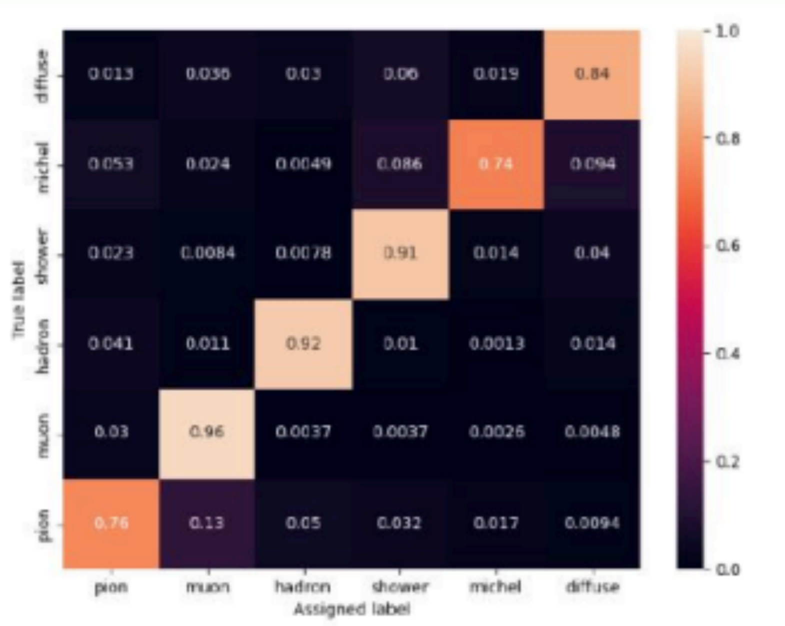


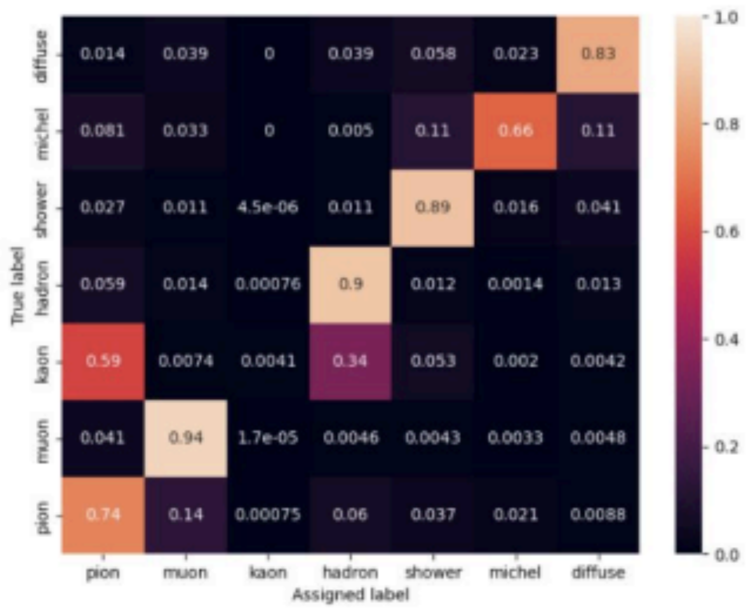






MicroBooNE





ICARUS

- Step one: correct classification of particles in both detectors.
- Future: perform domain adaptation on other levels of the hierarchy and types of tasks.
- Could be useful for combining DUNE near and far detectors?

A. Ciprijanovic, https://indico.fnal.gov/event/66124/contributions/301008/attachments/182275/250251/Exa.TrkX%20meeting%20-%20DA.pdf

Fermilab

NuGraph Social Network