FERMILAB-SLIDES-23-324-CSAID

A Multipurpose Graph Neural Network for Reconstruction in LArTPC Detectors

G. Cerati (FNAL) — with lots of input from V Hewes (UCincinnati)

Connecting the Dots

Oct. 10, 2023

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Introduction

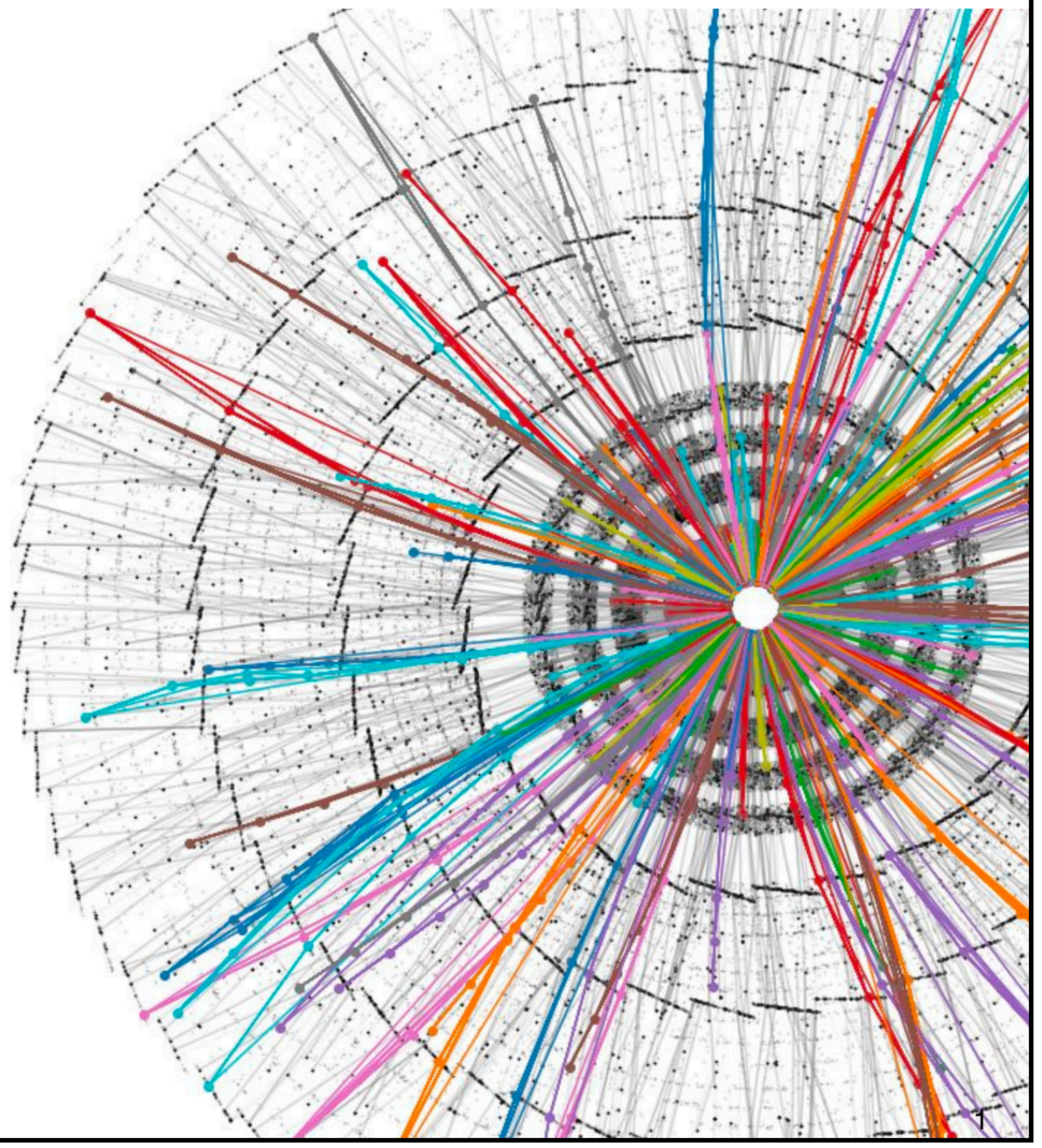
- I am presenting work by the Exa.TrkX collaboration based on the MicroBooNE open samples
  - https://microboone.fnal.gov/documents-publications/public-datasets/
  - we have a paper in preparation, stay tuned!

- This network architecture is developed to have broad applicability, without being tied to any particular detector geometry.
  - This network was initially developed in the context of the DUNE Far Detector geometry for reconstructing high-multiplicity atmospheric and ντ interactions.
  - Also being deployed on non-LArTPC detector technology!
  - See NuML and pynuml packages

credit: V Hewes

Exa.TrkX

- Exa.TrkX is a collaboration developing next-generation Graph Neural Network (GNN) reconstruction for HEP:
  - **Energy Frontier**
    - Expand on HEP.TrkX's prototype GNN for HL-LHC.
    - Incorporate into ATLAS's simulation and validation chain.
  - **Intensity Frontier**
    - Explore viability of HEP.TrkX network for neutrino physics.
    - Develop GNN-based reconstruction for Liquid Argon TPCs.

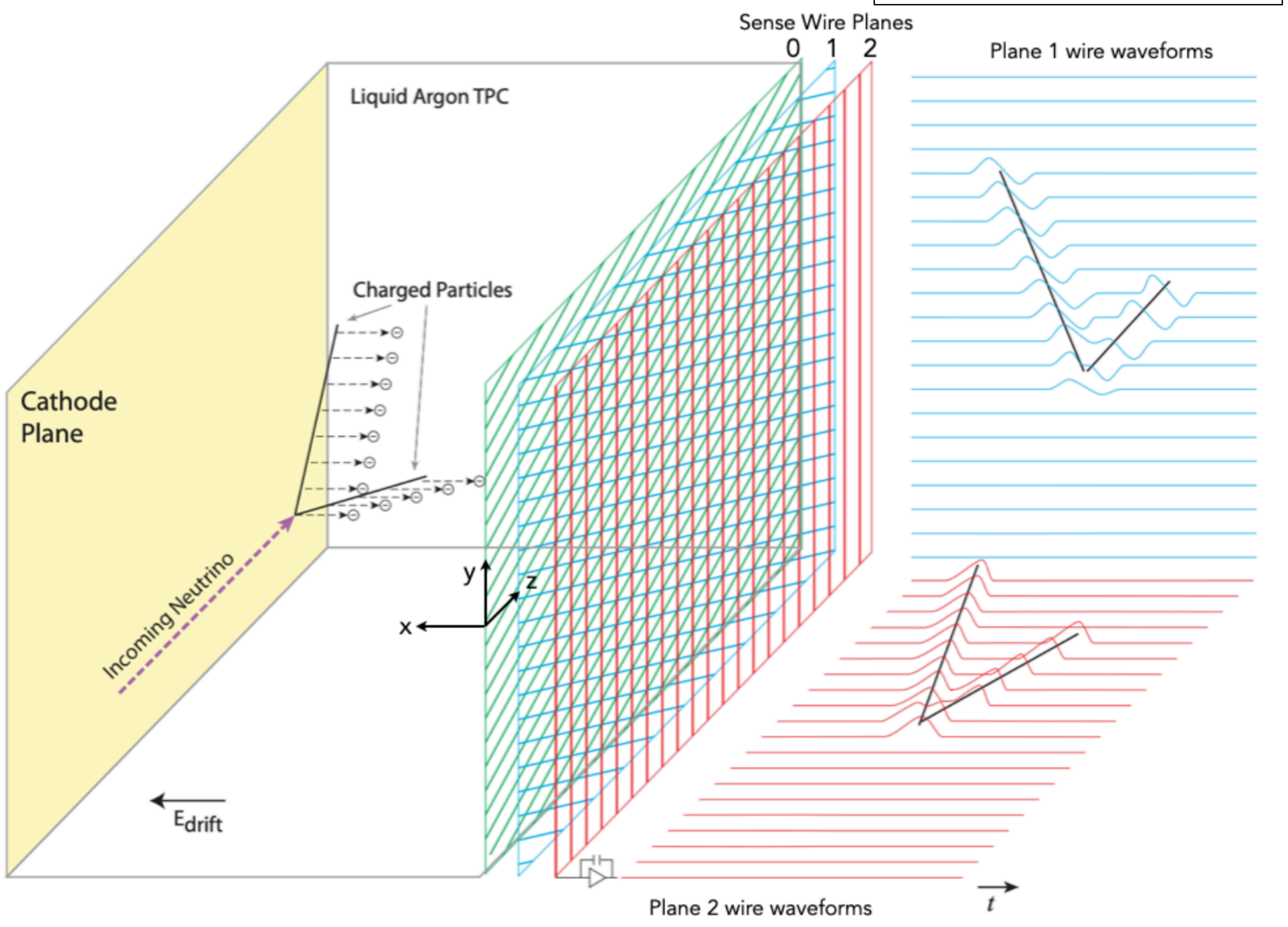


MicroBooNE's Liquid Argon Time Projection Chamber (LArTPC)

- Charged particles produced in neutrino interactions ionize the argon, ionization electrons drift in electric field towards anode planes

- Sense wires detect the incoming charge, producing beautiful detector data images

JINST 12, P02017 (2017)

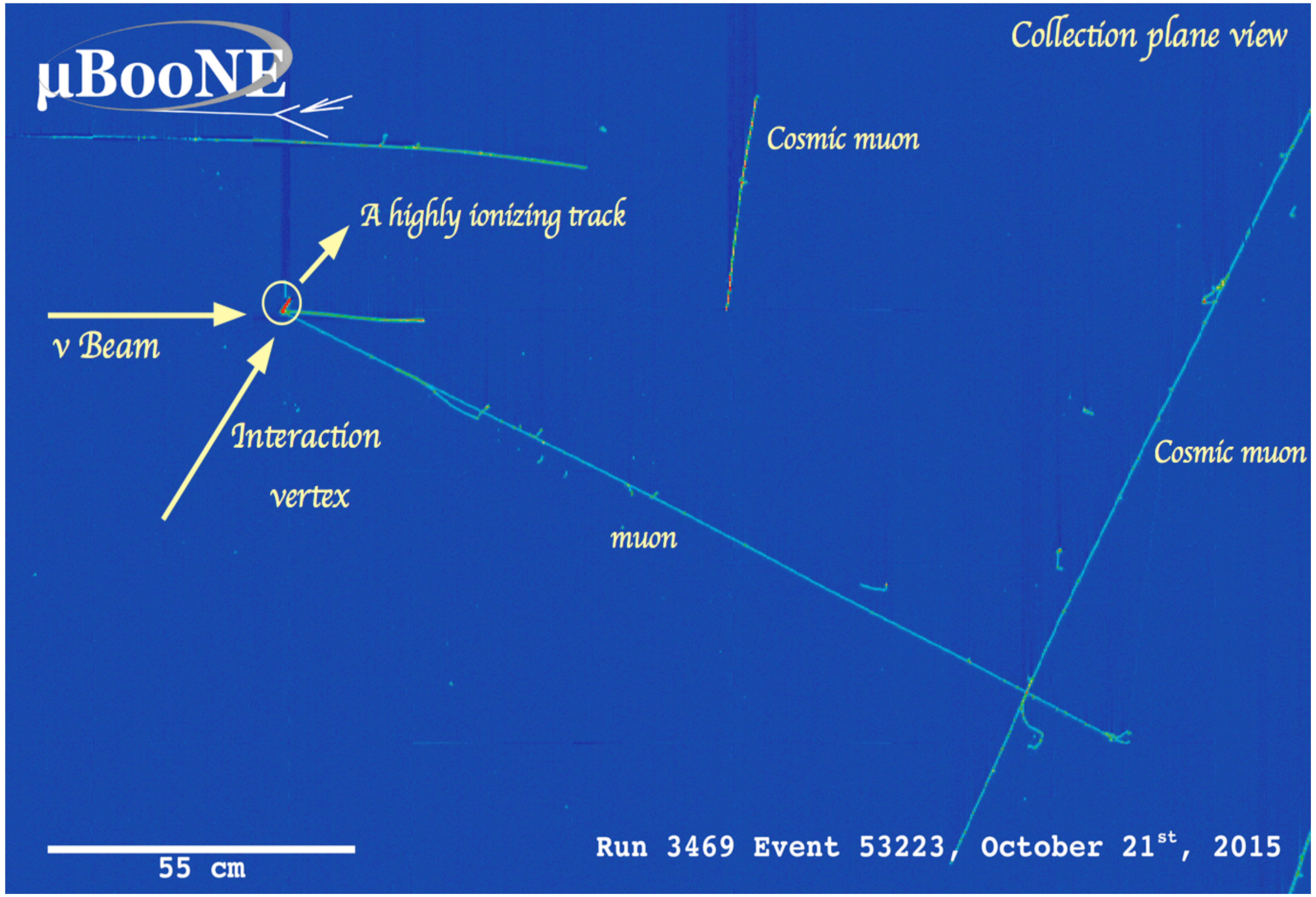


3 planes allow for 3D reco

MicroBooNE's Liquid Argon Time Projection Chamber (LArTPC)

JINST 12, P02017 (2017)

- Charged particles produced in neutrino interactions ionize the argon, ionization electrons drift in electric field towards anode planes
- Sense wires detect the incoming charge, producing beautiful detector data images
- Full detail of neutrino interaction with O(mm) spatial resolution and calorimetric information
- Fast scintillation light detected by Optical system (PMT) for trigger & cosmic rejection



axes: time vs wire - color scale: charge

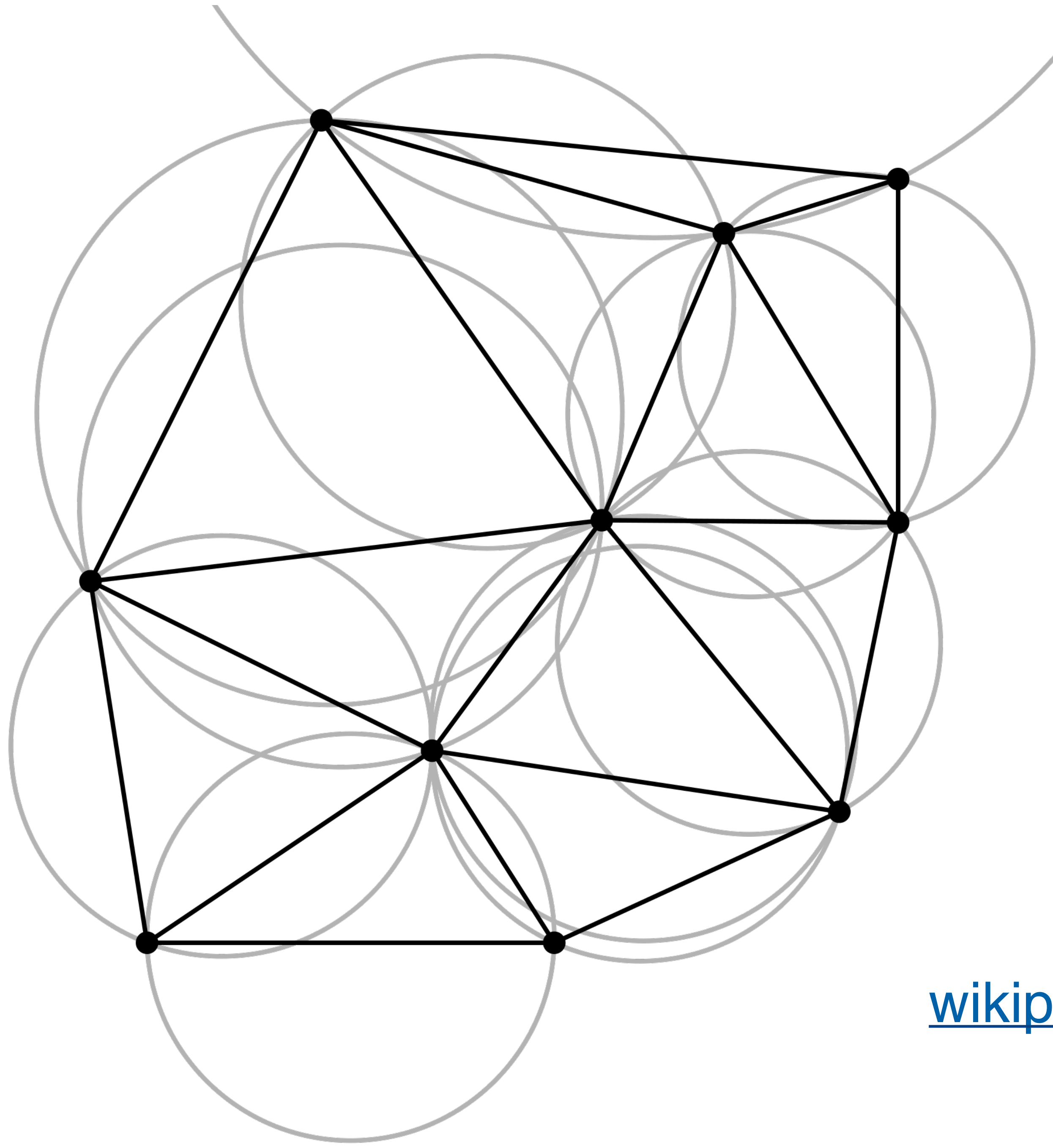
Main idea

- GNNs have been successfully used for tracking application at LHC, can they be used for LArTPC reconstruction?
  - Eur.Phys.J.C 81 (2021) 10, 876 • e-Print: 2103.06995

- Detector hits can be connected in a graph
  - Naturally sparse representation of the data
  - Low-level information, close to native output of the detector
  - Graphs can also connect hits from different planes, thus making the network “3D-aware”

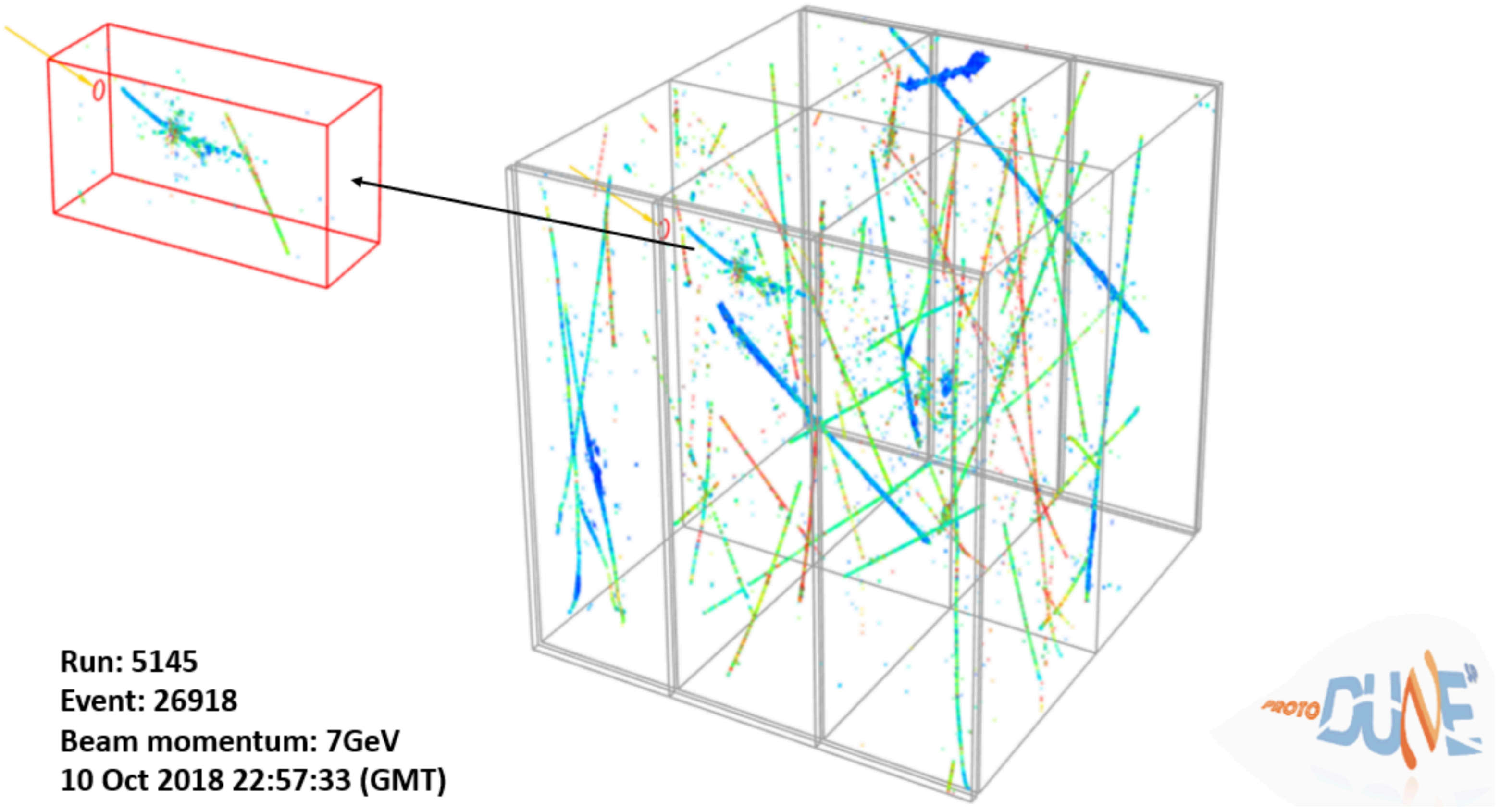
Inputs and Graph formation

- Main inputs to the GNN are the Hits
  - features: wire, peak time, integral, RMS
  - currently using Hits associated to the Pandora neutrino slice
- Within each plane hits are connected in a graph using Delaunay triangulation
  - fully connected graph, both long and short distance edges, able to jump across unresponsive wire regions



wikipedia

- Hit associations to 3D SpacePoints (currently from the SP solver) are used to create “nexus” connections across graphs in each plane
  - SpacePoints are not connected among themselves
  - No input features for SpecePoints



arXiv:2002.03005

NuGraph2 Network Architecture: Overview

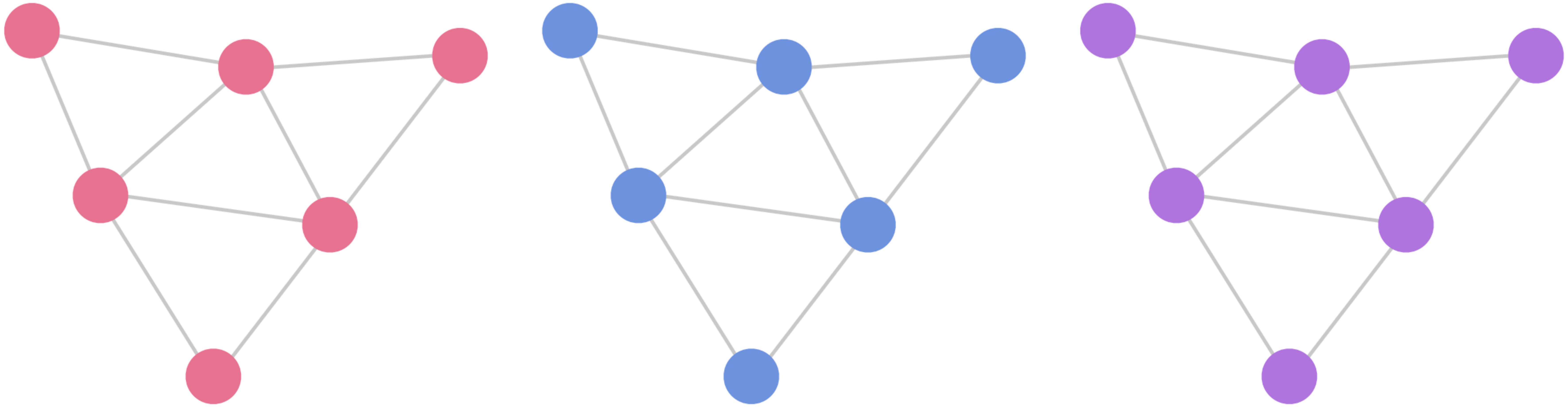
- Initial application for the GNN is semantic hit classification
  - Categories based on the type of particle that produced the hit.

- NuGraph2's core convolution engine is a self-attention message-passing network utilizing a categorical embedding
  - Each particle 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.

- Each message-passing iteration consists of two phases, the planar step and the nexus step:
  - Pass messages internally in each plane.
  - Pass messages up to 3D nexus nodes to share context information.

Message passing iteration through the graph

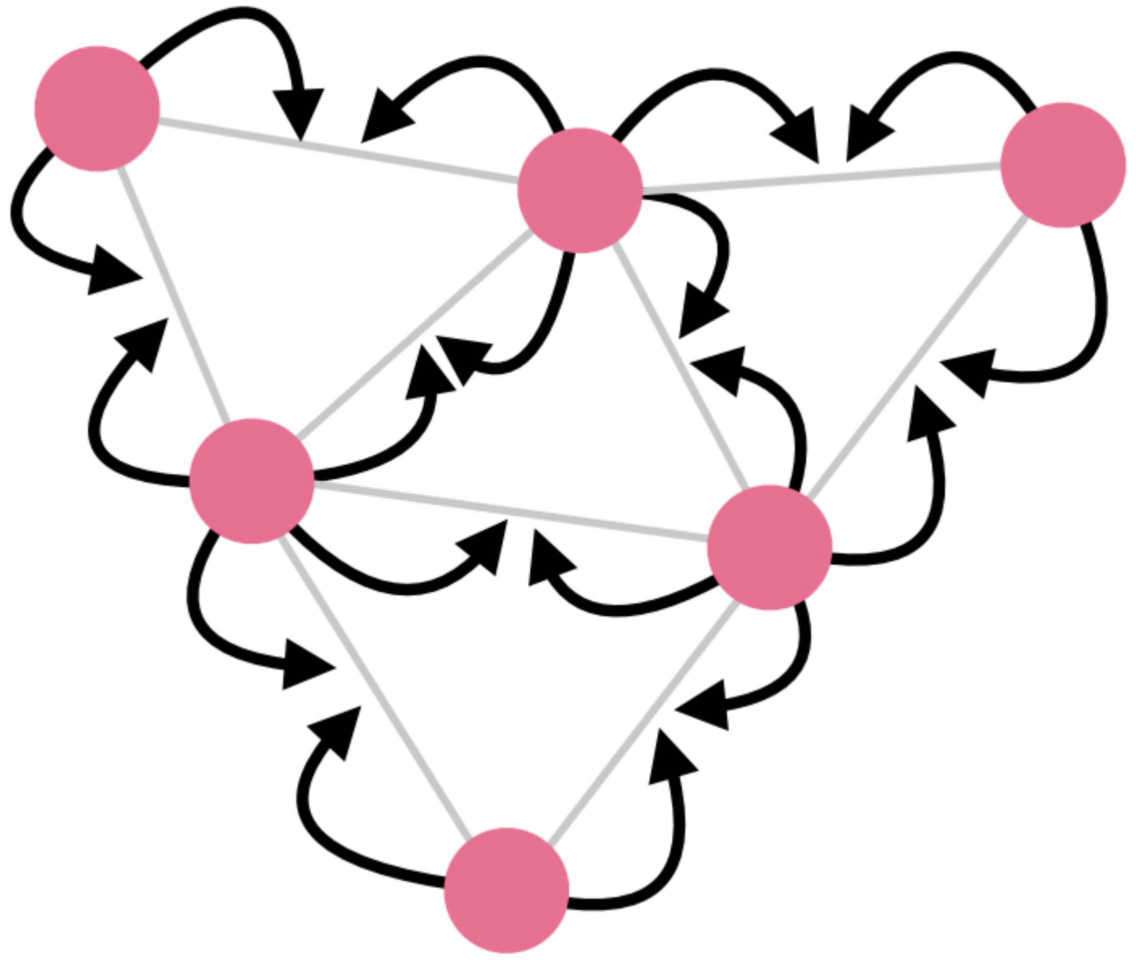
Input graph with node features, in each TPC plane



Fermilab

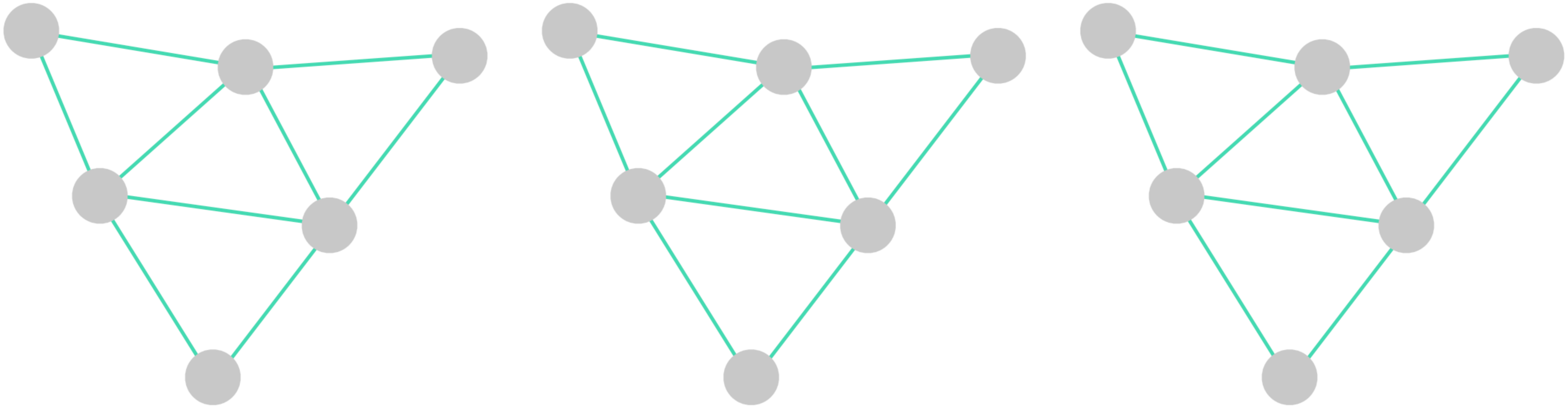
Message passing iteration through the graph

Node features convolved to obtain edge features



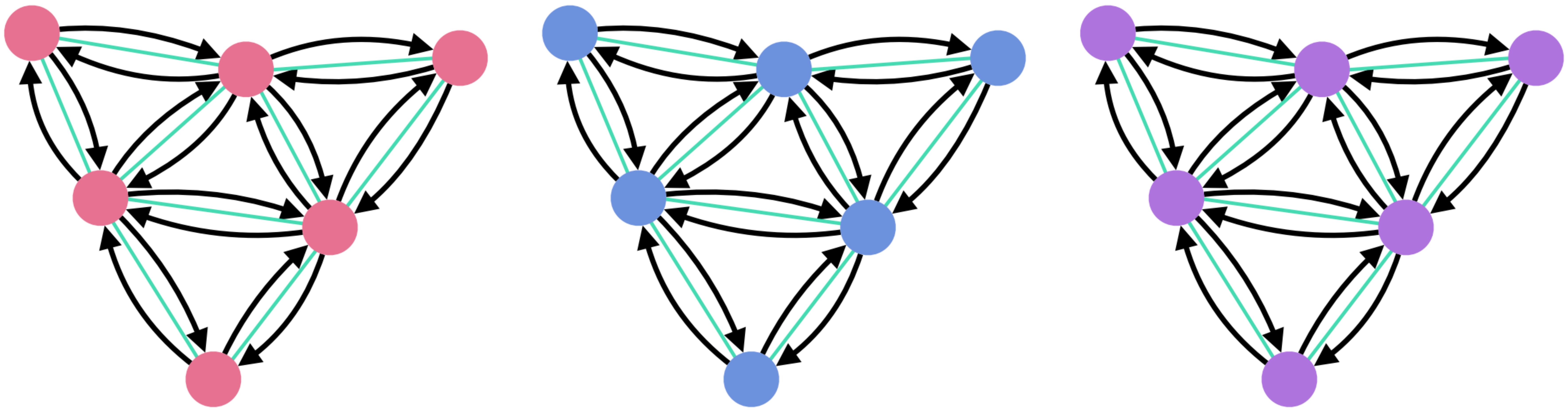
Message passing iteration through the graph

From edge features derive edge weights



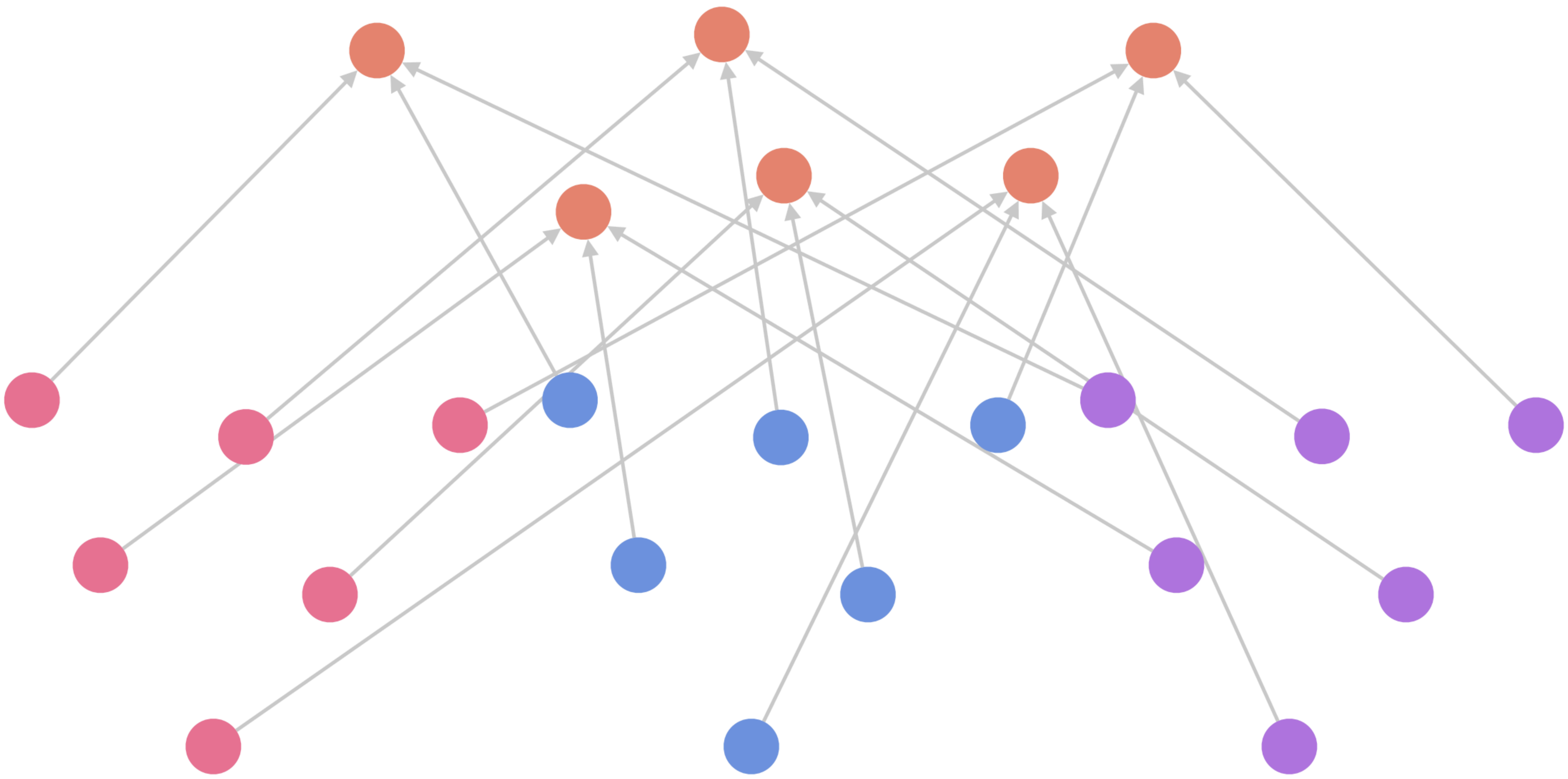
Message passing iteration through the graph

Update node features using edge-weighted features from connected nodes



Message passing iteration through the graph

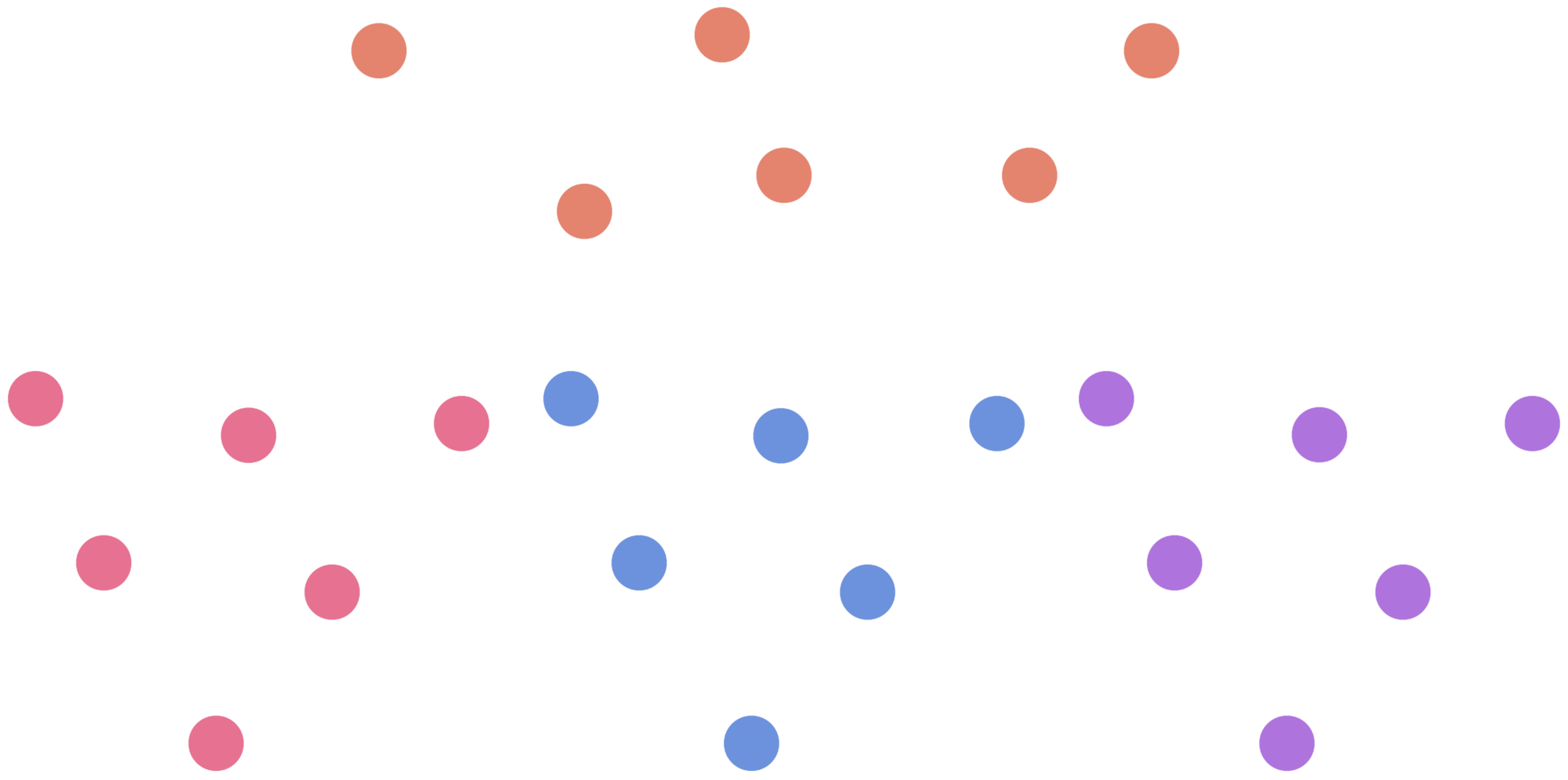
Propagate node features to 3D nexus nodes



Fermilab

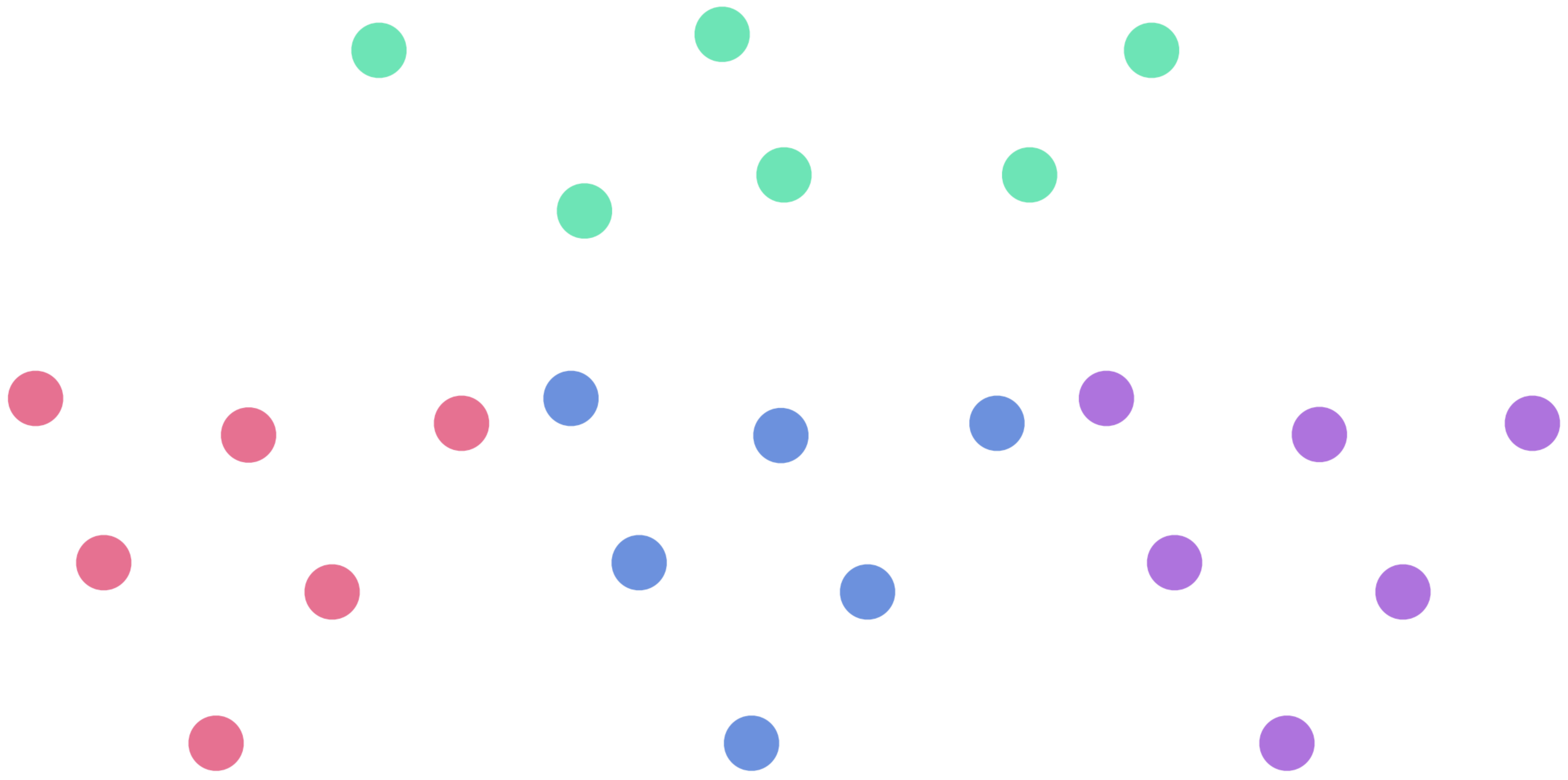
Message passing iteration through the graph

Convolve nexus node features to mix information between detector planes



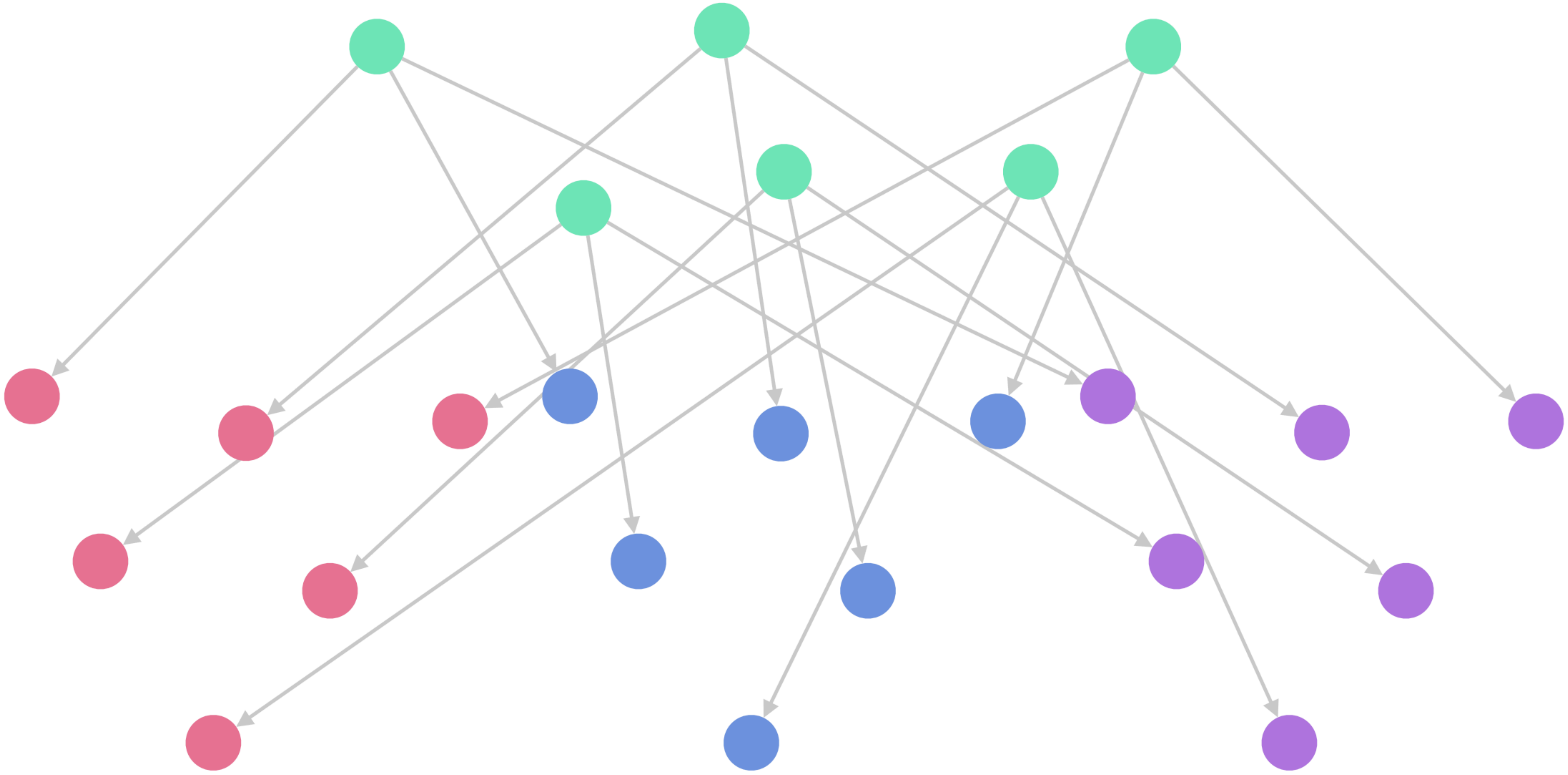
Message passing iteration through the graph

Convolve nexus node features to mix information between detector planes



Message passing iteration through the graph

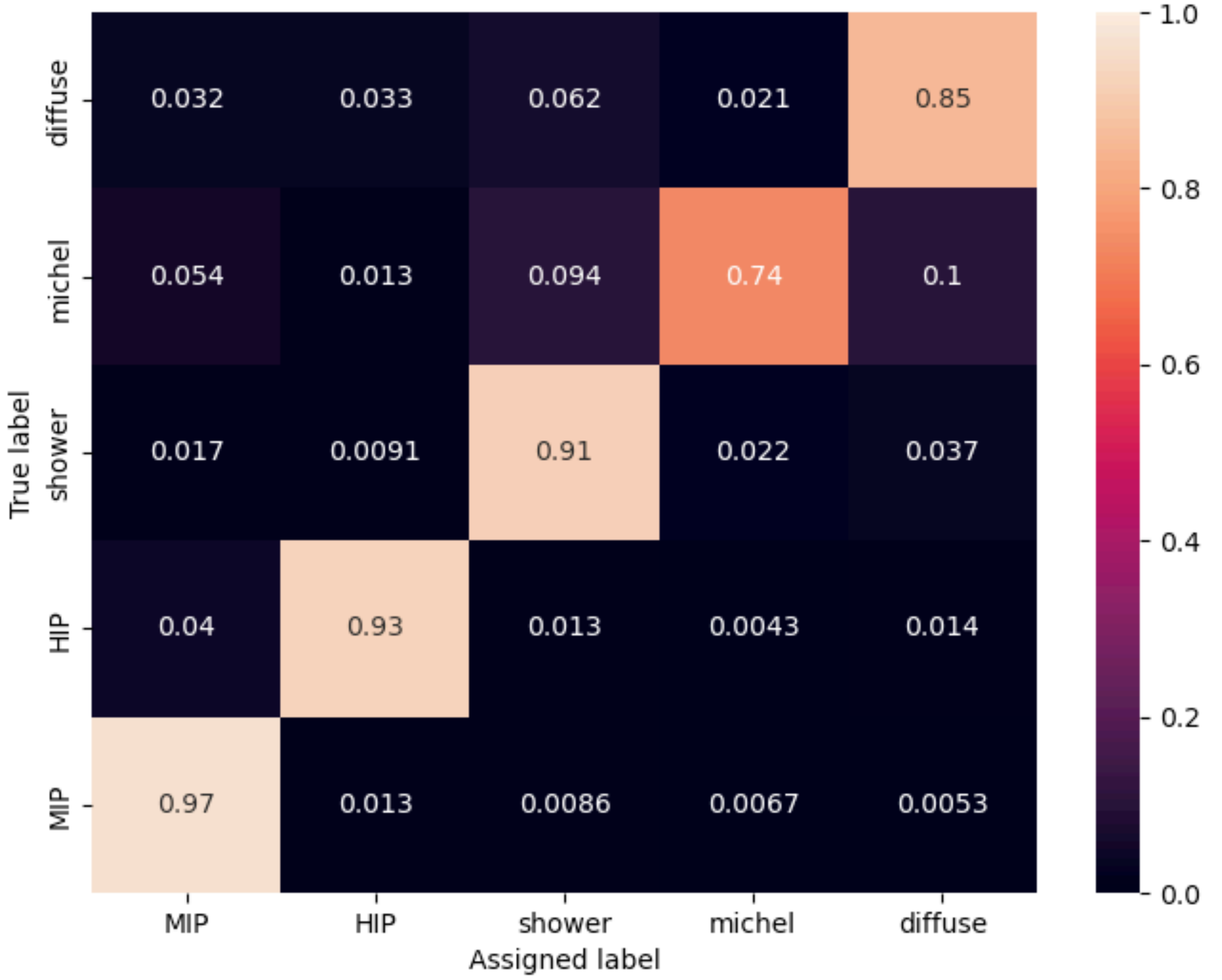
Propagate 3D nexus nodes features back down to 2D planar nodes



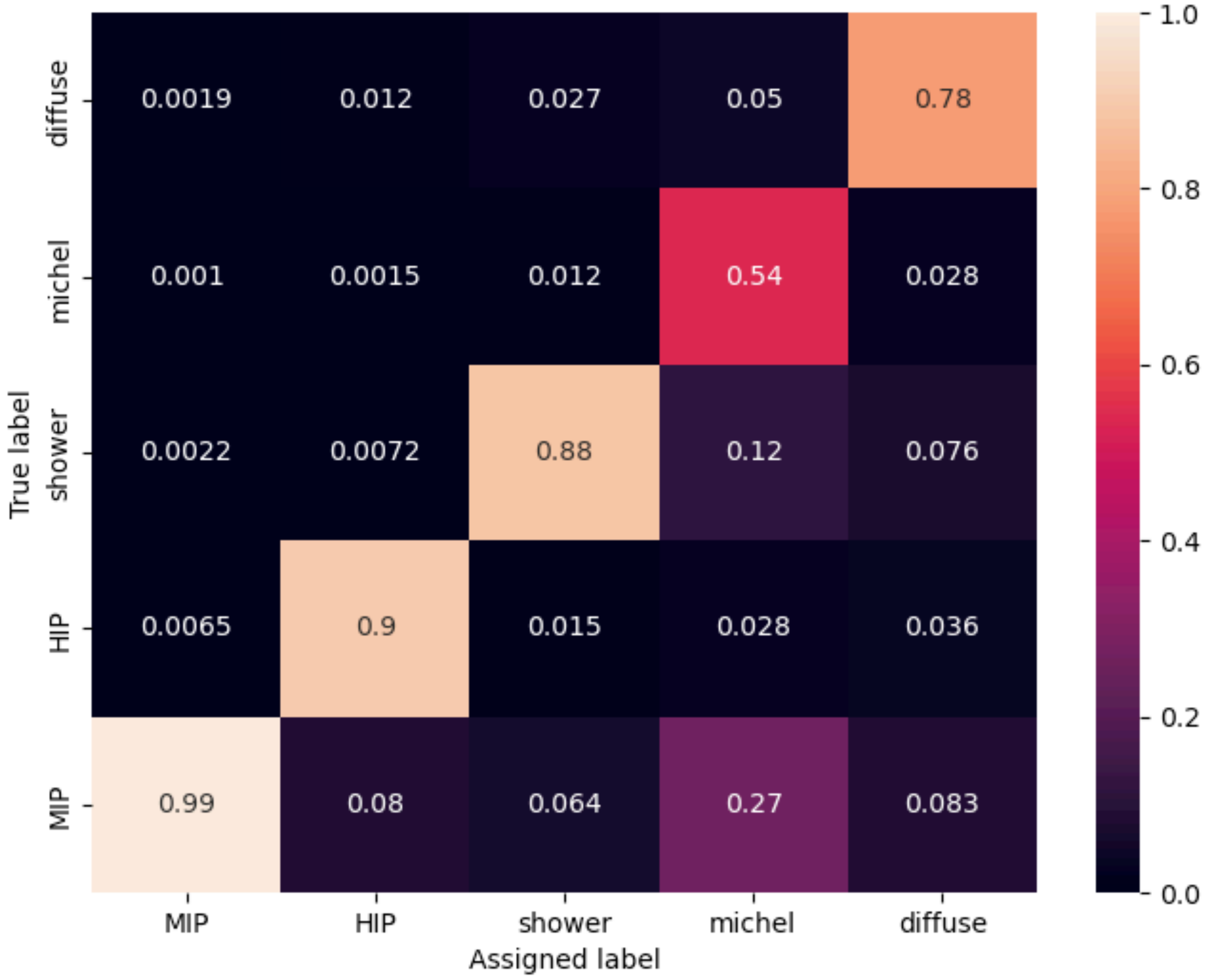
Semantic hit classification

- 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

recall (efficiency)



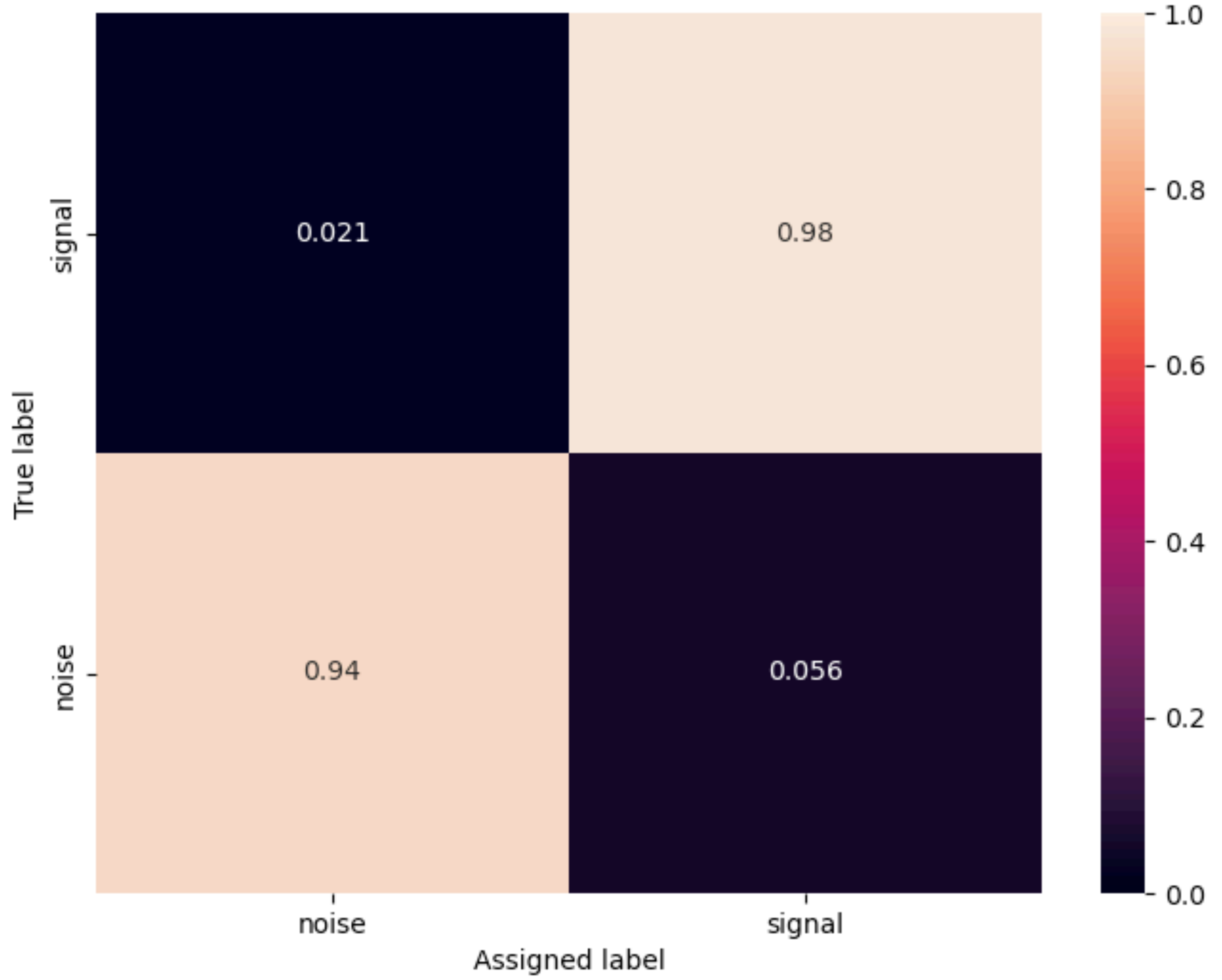
precision (purity)



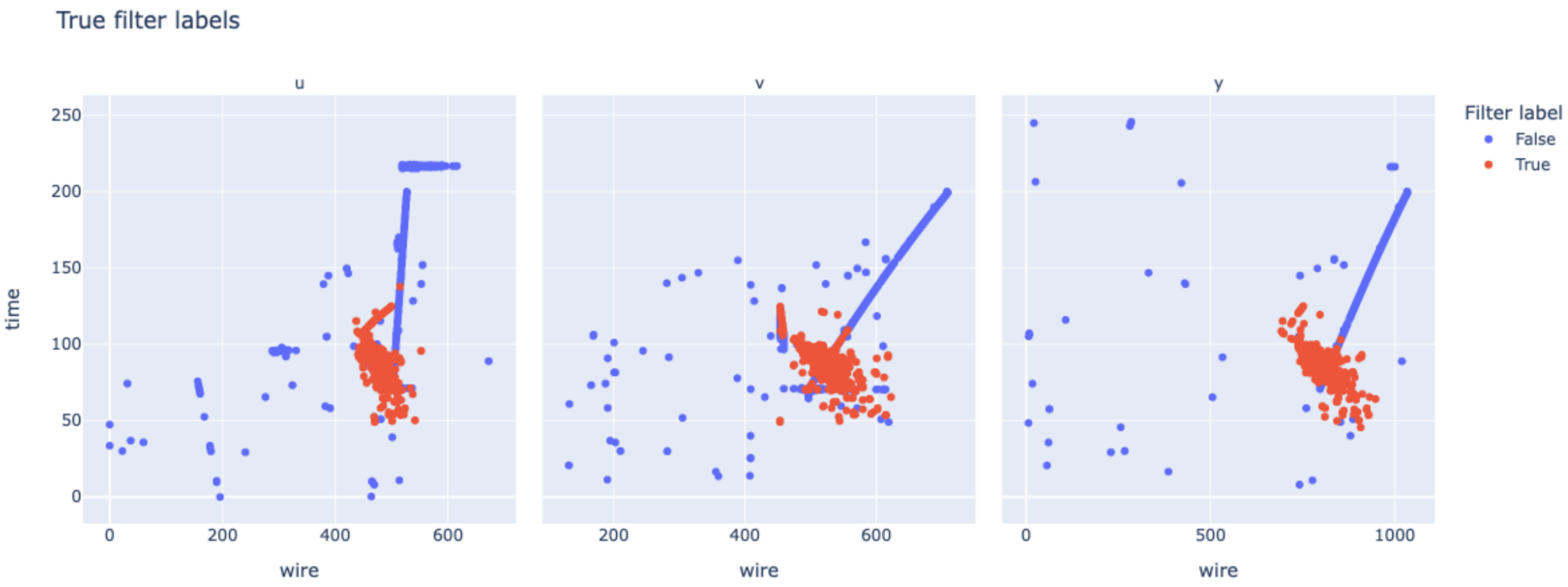
Filter hit classification

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

- Performance metrics:
  - recall and precision: ~0.98

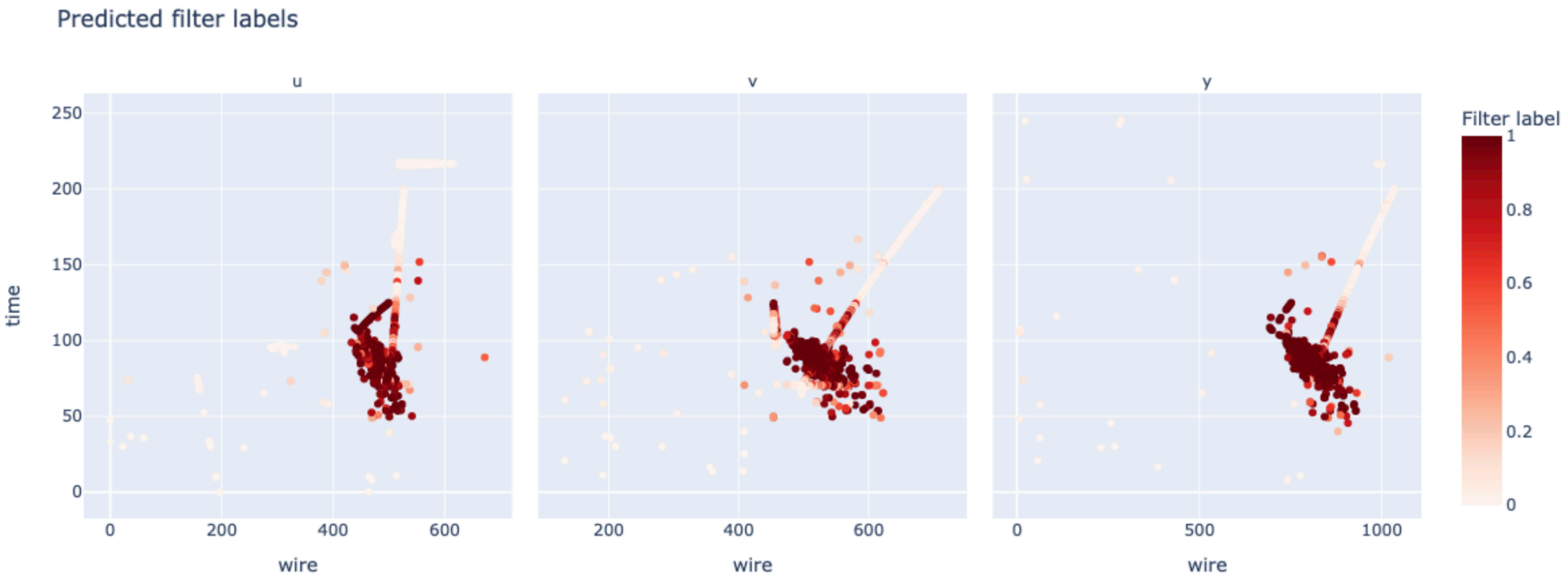


Event display

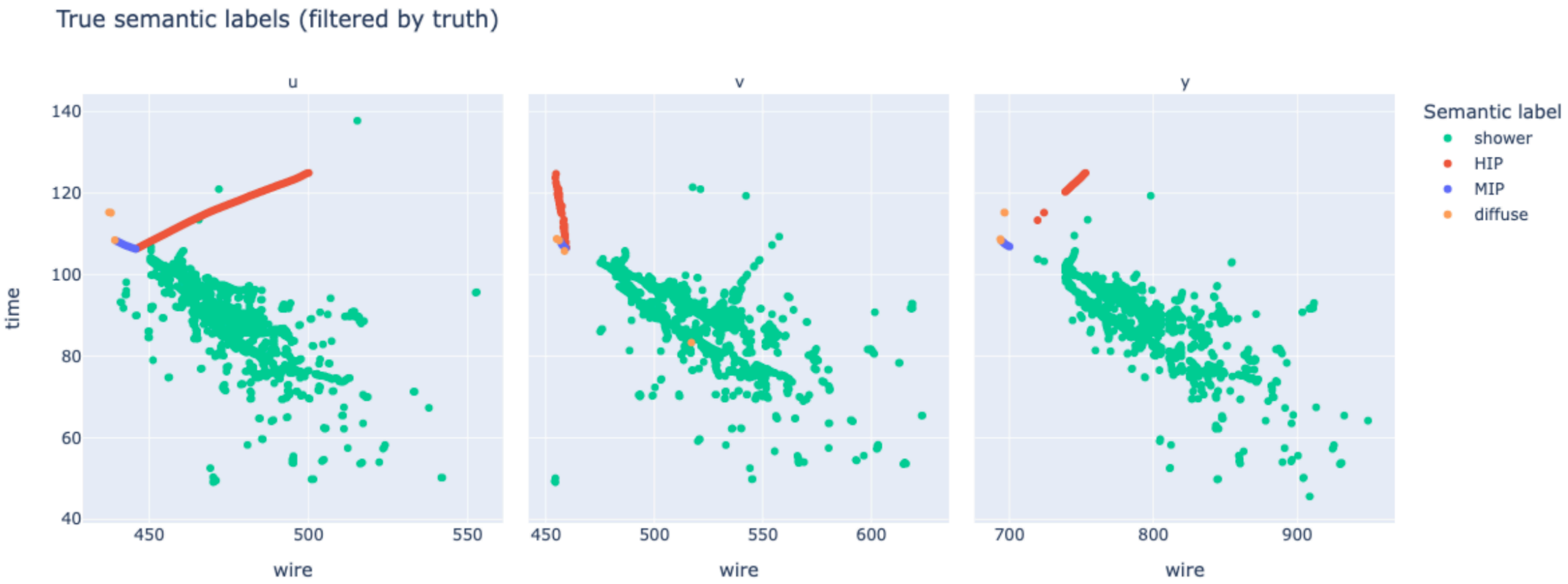


Fermilab

Event display



Event display



Fermilab

Event display

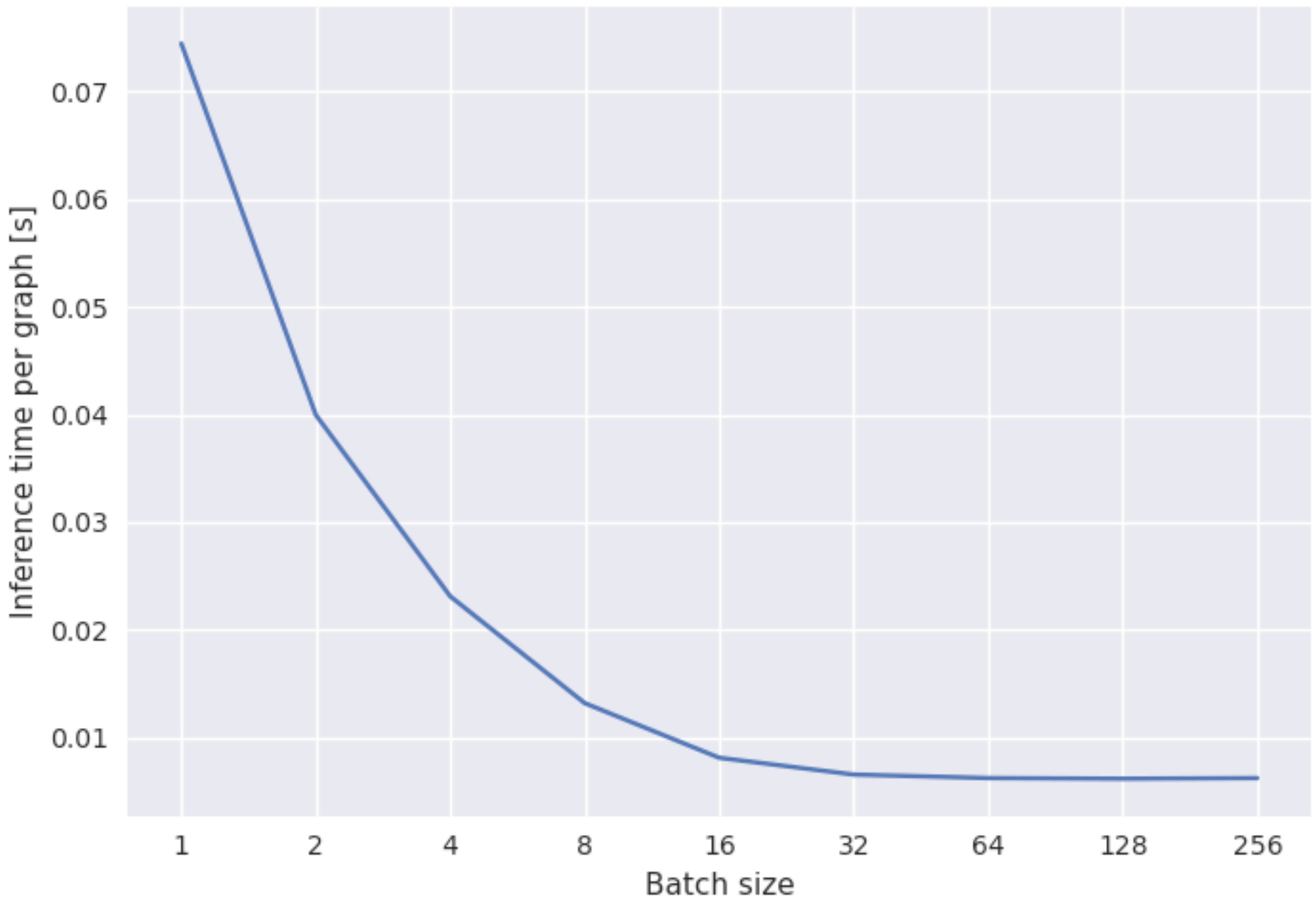
Fermilab

Inference time

- Relatively small network:
  - number of learnable parameters: ~410k
  - max RSS memory on CPU: ~2.5 GB

- Out of the box inference time:
  - 0.12 s/evt on CPU
  - 0.005 s/evt batched on GPU
  - graph construction not included, but also fast

- Implications:
  - can easily run on CPU as part of regular offline processing
  - can run very fast for realtime applications on GPU, or other accelerators



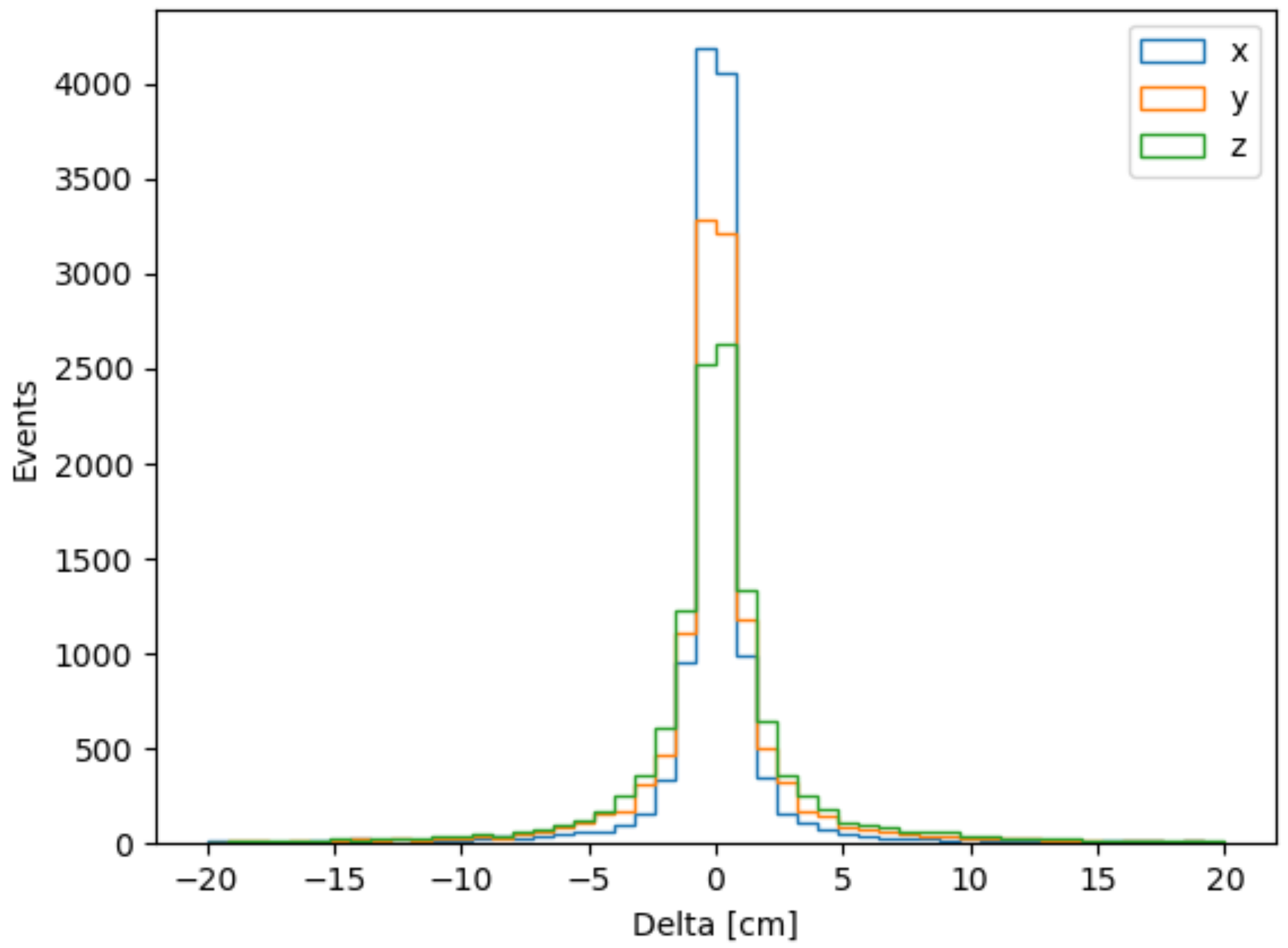
Fermilab

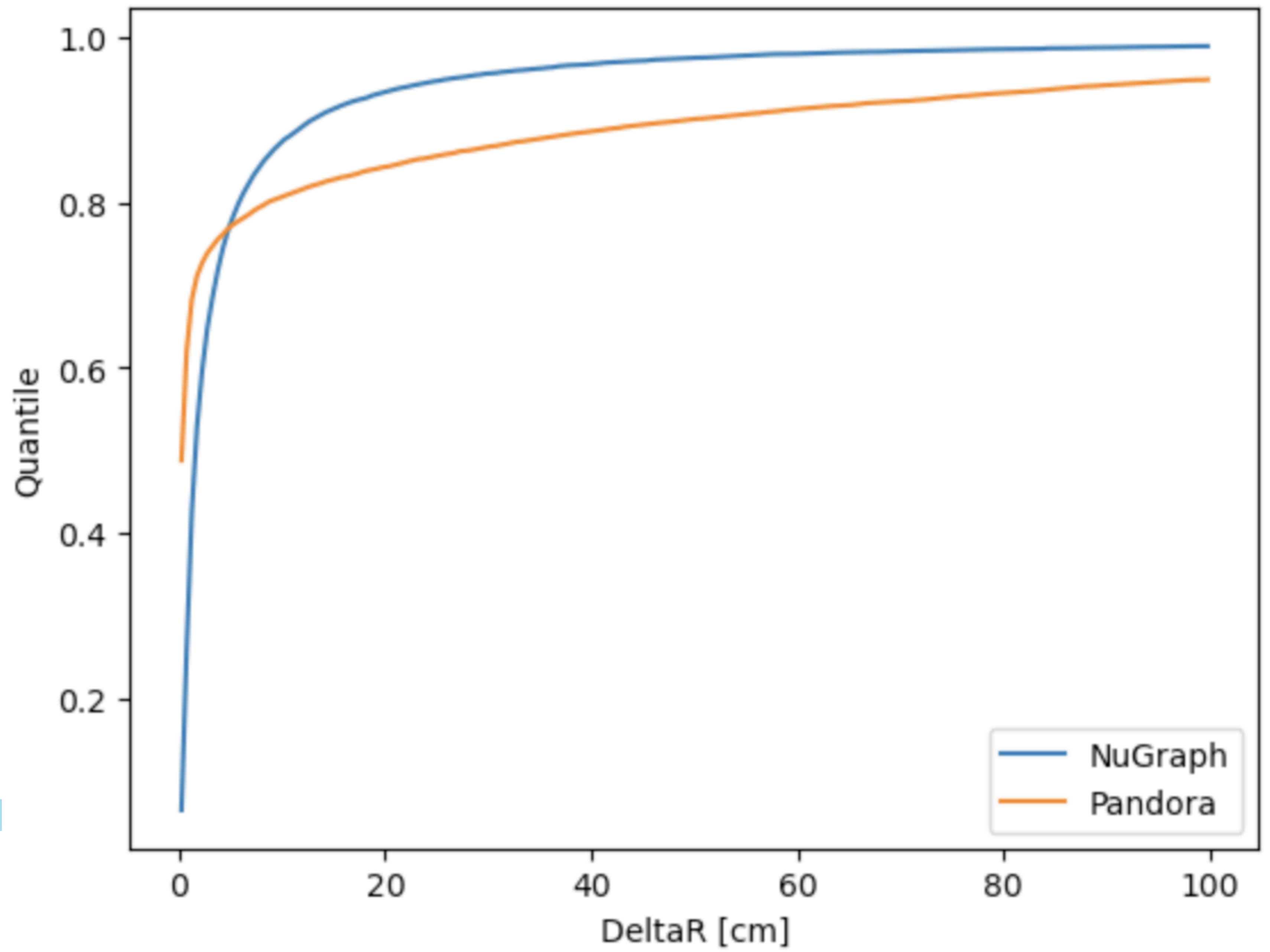
Vertex position classification

- Preliminary work demonstrates that our GNN is able to identify the neutrino interaction vertex position in 3D
  - currently O(cm) level resolution in each coordinate

- Compared to current vertex reconstruction this version shows worse percentile a low ΔR, but better at larger ΔR
  - worse at finding exact point, better at avoiding catastrophic errors

- Issues related to ground truth definition identified and being fixed, expect to achieve better results soon





Summary and next steps

- NuGraph2 is a multi-purpose GNN architecture for reconstructing neutrino interactions in LArTPC
  - Efficiently reject background detector hits.
  - Classify detector hits according to particle type.
  - Lightweight network, allowing fast inference on CPU and GPU.
  - Preliminary results for vertexing are promising.

- Next steps:
  - Immediate plans: paper, and inference in production is high priority!
    - Discussing ways to integrate in LArSoft through NuSonic (also for CPU)
  - Developments: clustering, hierarchical graphs, inclusion of info from other detectors.