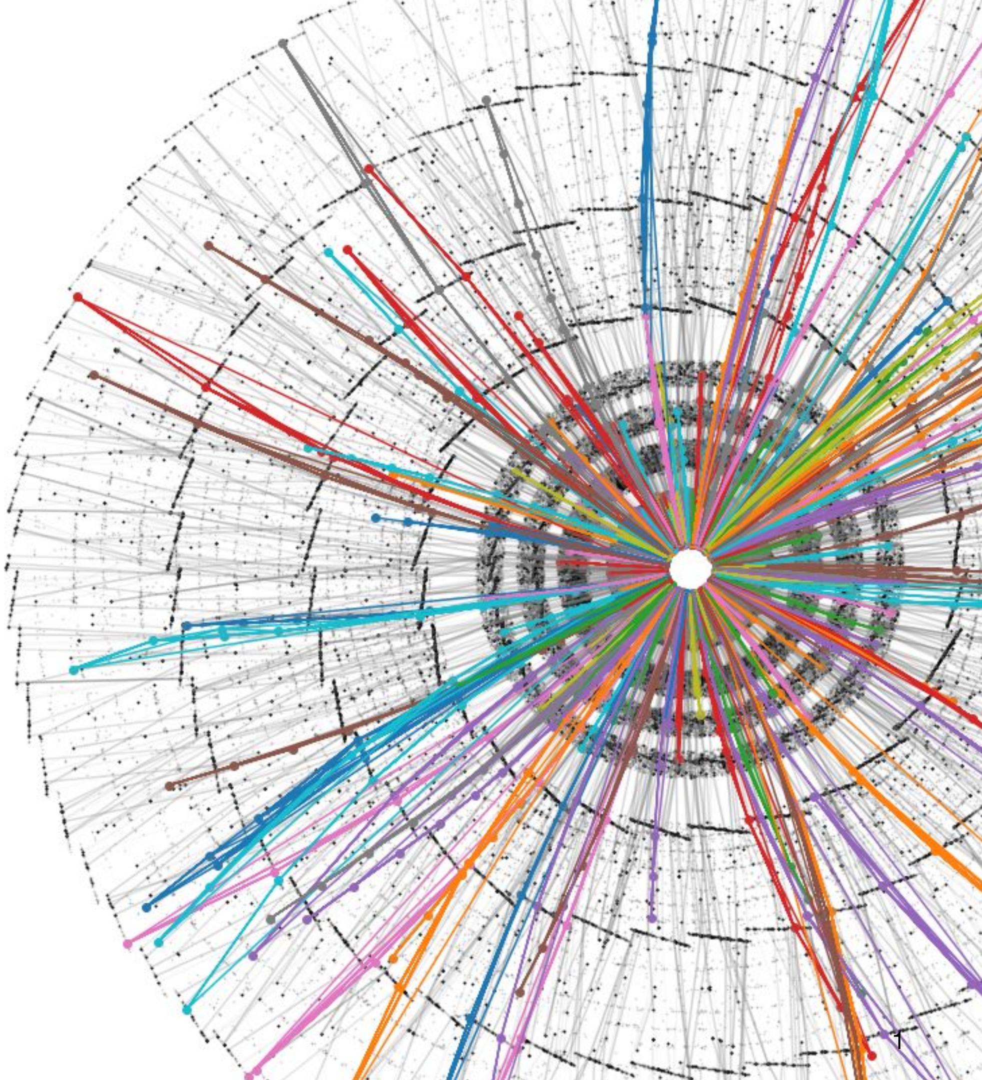
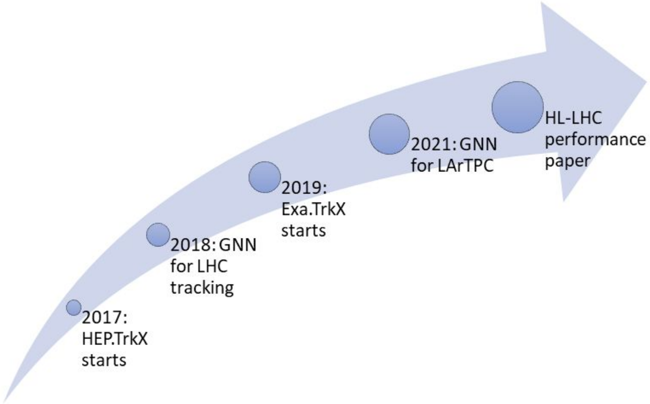
THE EXATRKX PROJECT
PARTICLE PHYSICS WITH GRAPH NEURAL NETWORKS

V Hewes & Daniel Murnane
On behalf of the ExatrkX Collaboration



EXATRKX OVERVIEW

- Approach is to use ML & graphs-based pipeline for particle tracking experiments (GNNs and graph algorithms)
- Graph-based techniques more appropriate for typical Particle Physics data structures than CNNs.
- What is a typical task: point cloud in, build graph, link prediction and/or node regression
- Unique challenges of each frontier has produced a range of innovations specific to each

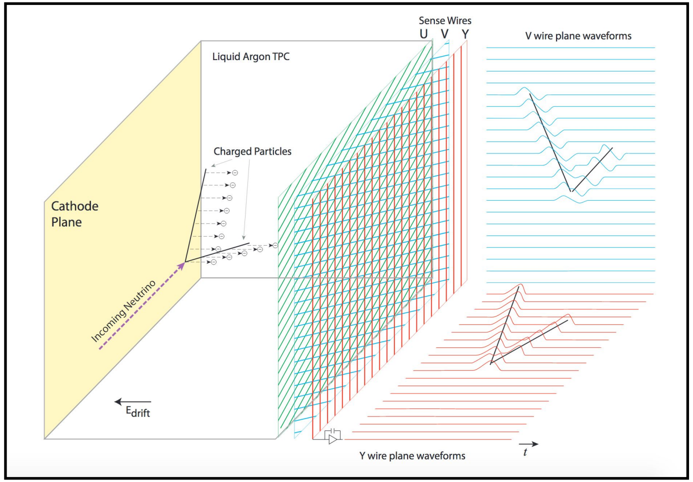


Geometric Deep Learning for the **Intensity Frontier**

TrkX

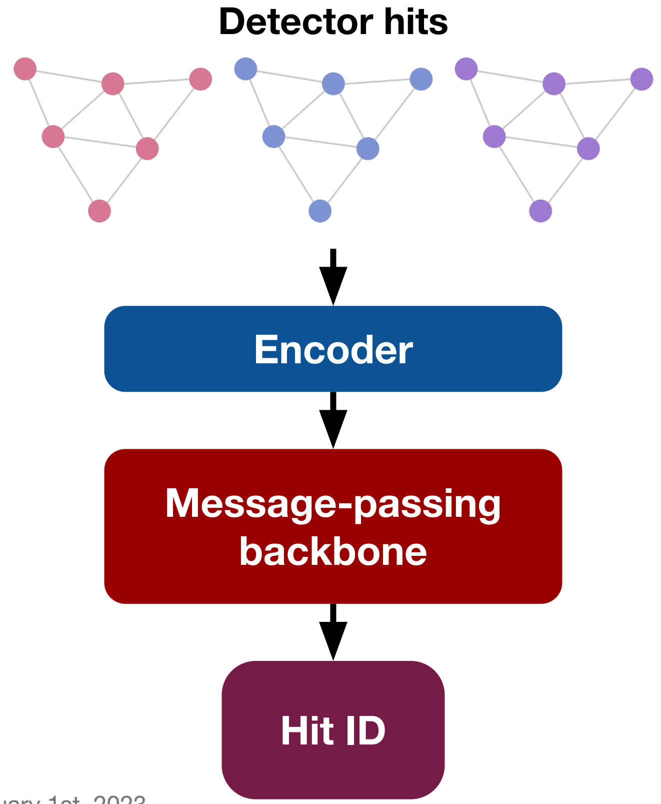
Liquid Argon TPCs

- Liquid Argon Time Projection Chambers (LArTPCs) currently a heavily utilised detector technology in neutrino physics.
  - At FNAL: MicroBooNE, Icarus, SBND.
  - Future: DUNE (70kT LArTPC deep underground, plus near detector).
- Charged particles ionize liquid argon as they travel.
- Ionisation electrons drift due to HV electrode field, and are collected by anode wires.
- Wire spacing ~3mm – **high-resolution detector**.

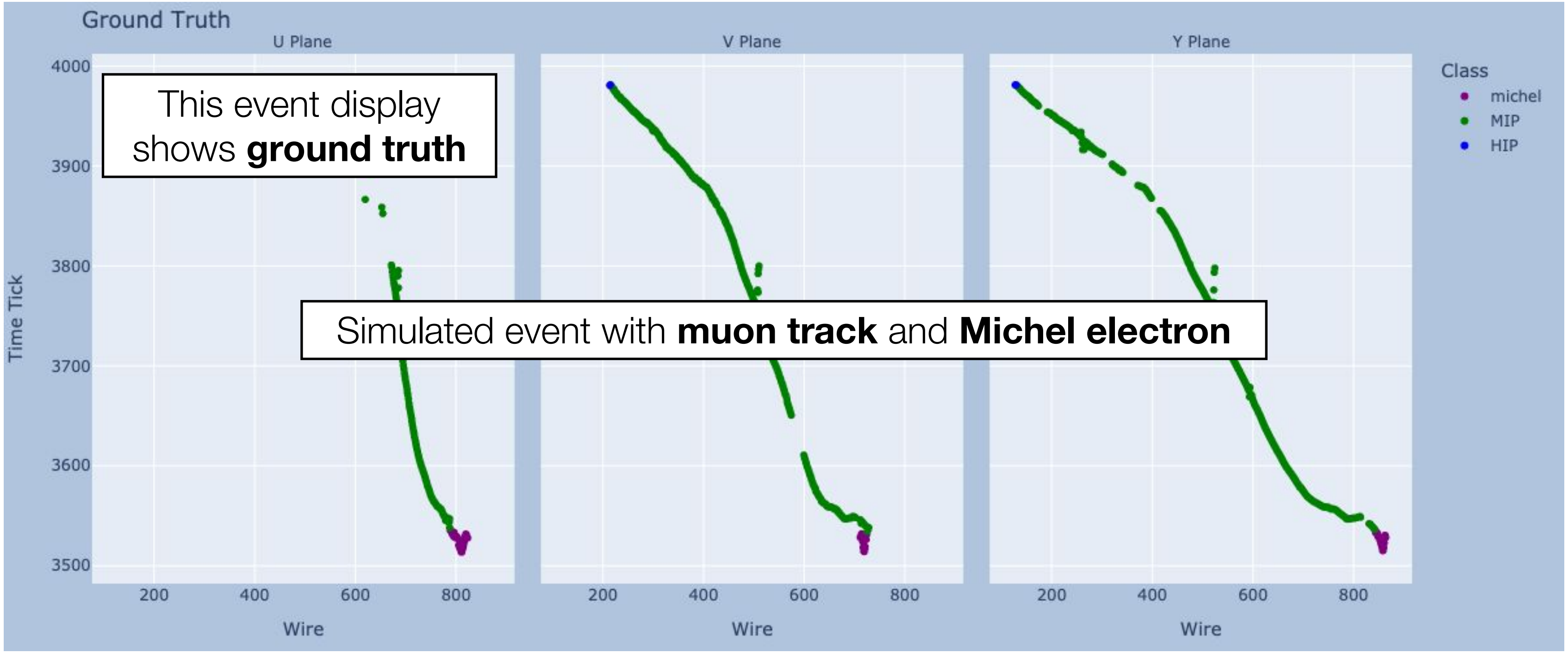


NuGraph2

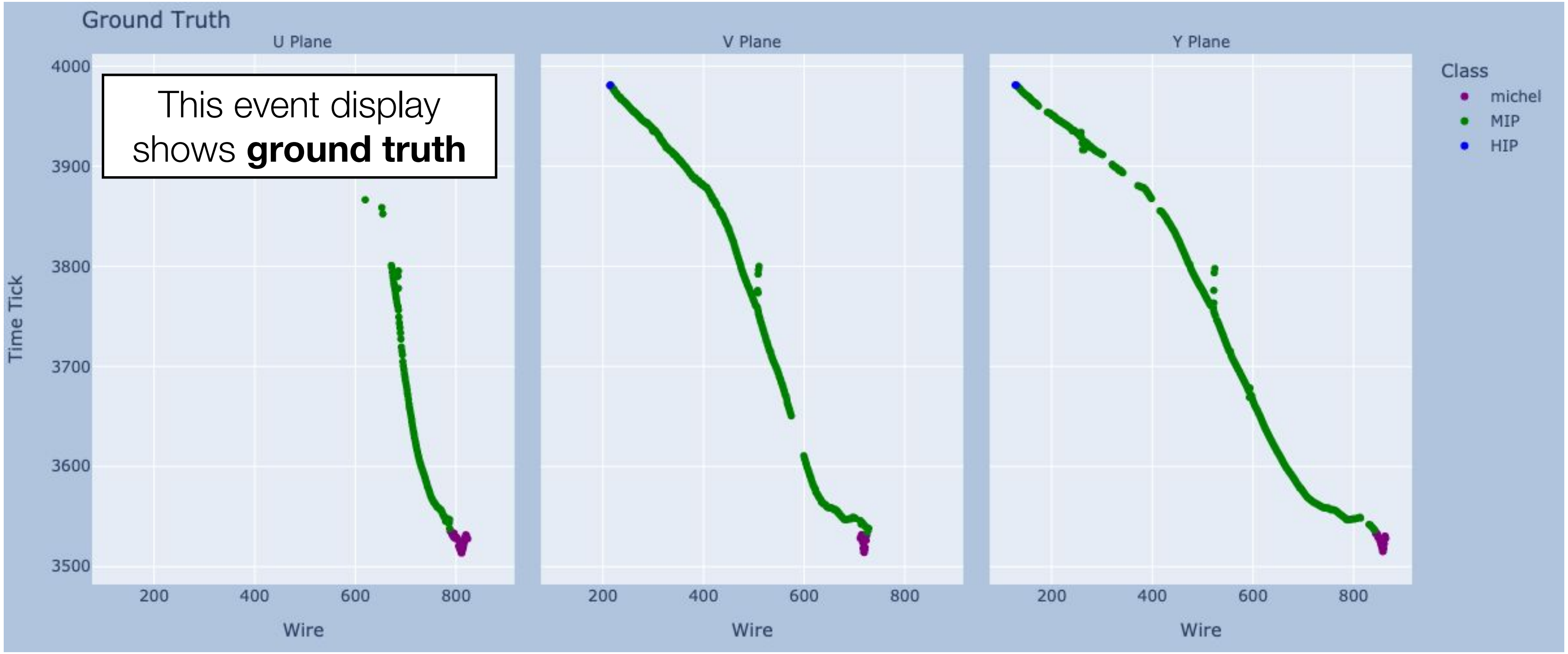
- Developed **NuGraph2**, a message-passing GNN for particle ID.
- **Semantically label detector hits according to particle type**.
- Developed by ExaTrkX in collaboration with the **University of Chicago Data Science Institute**.
- Originally developed for the **DUNE far detector**.
- Current application utilises **MicroBooNE open dataset**.
- Utilise **PyG** for graph convolutions and **PyTorch Lightning** as model backbone.



ν_μ event display

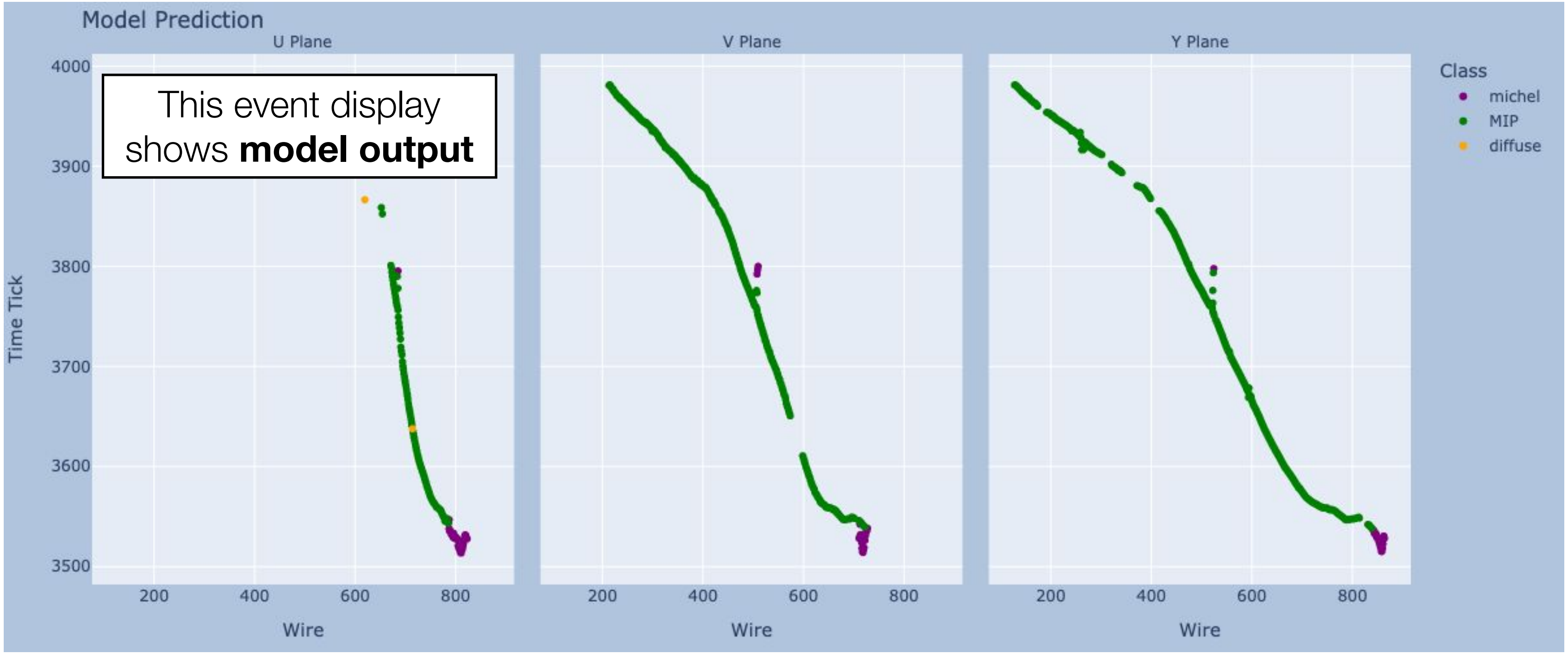


ν_μ event display

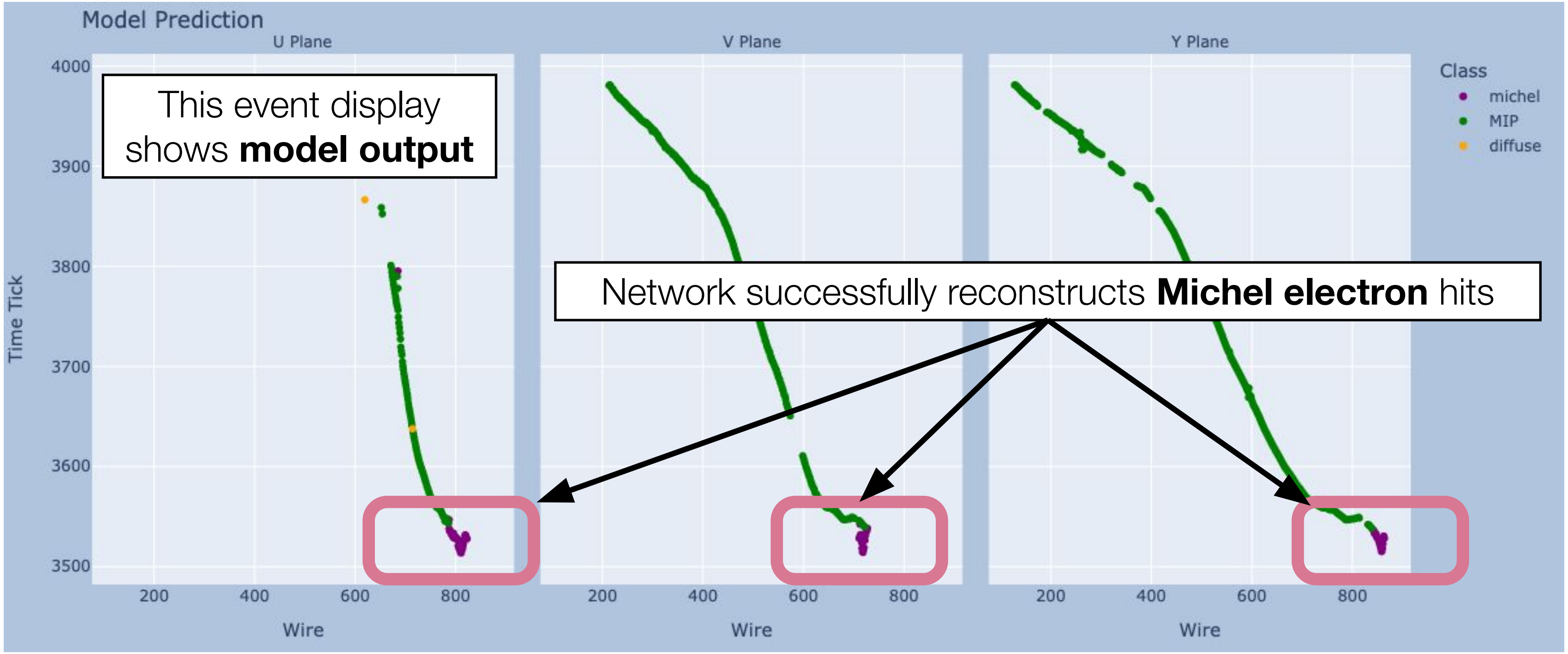


This event display shows **ground truth**

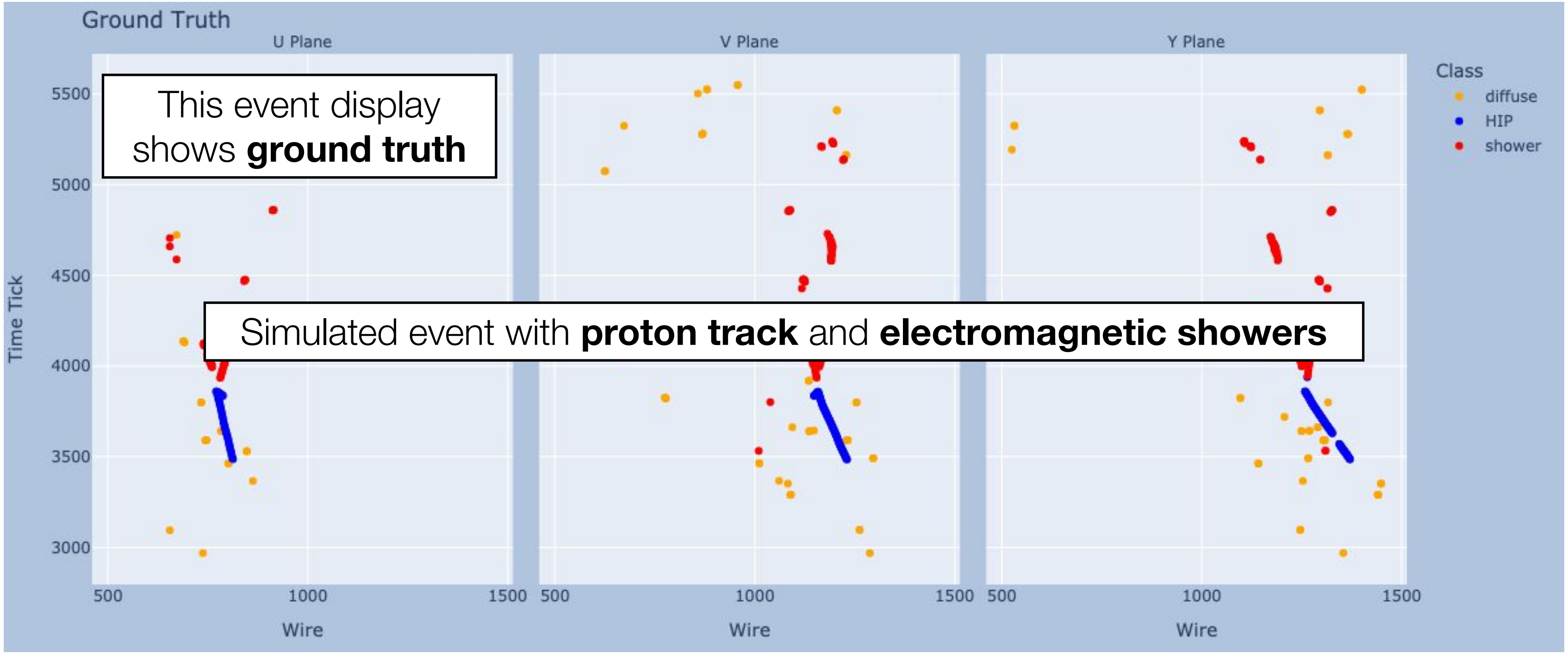
ν_μ event display



ν_μ event display



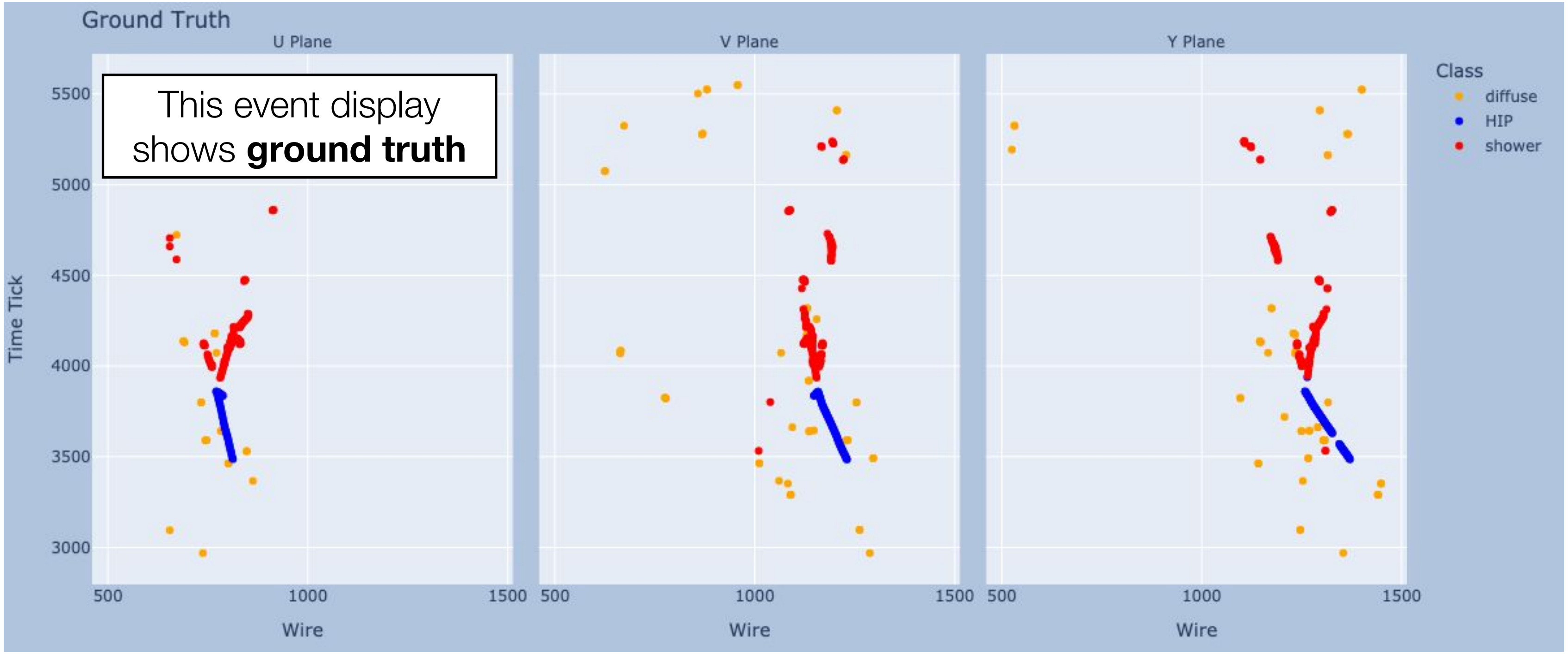
NC π^0 event display



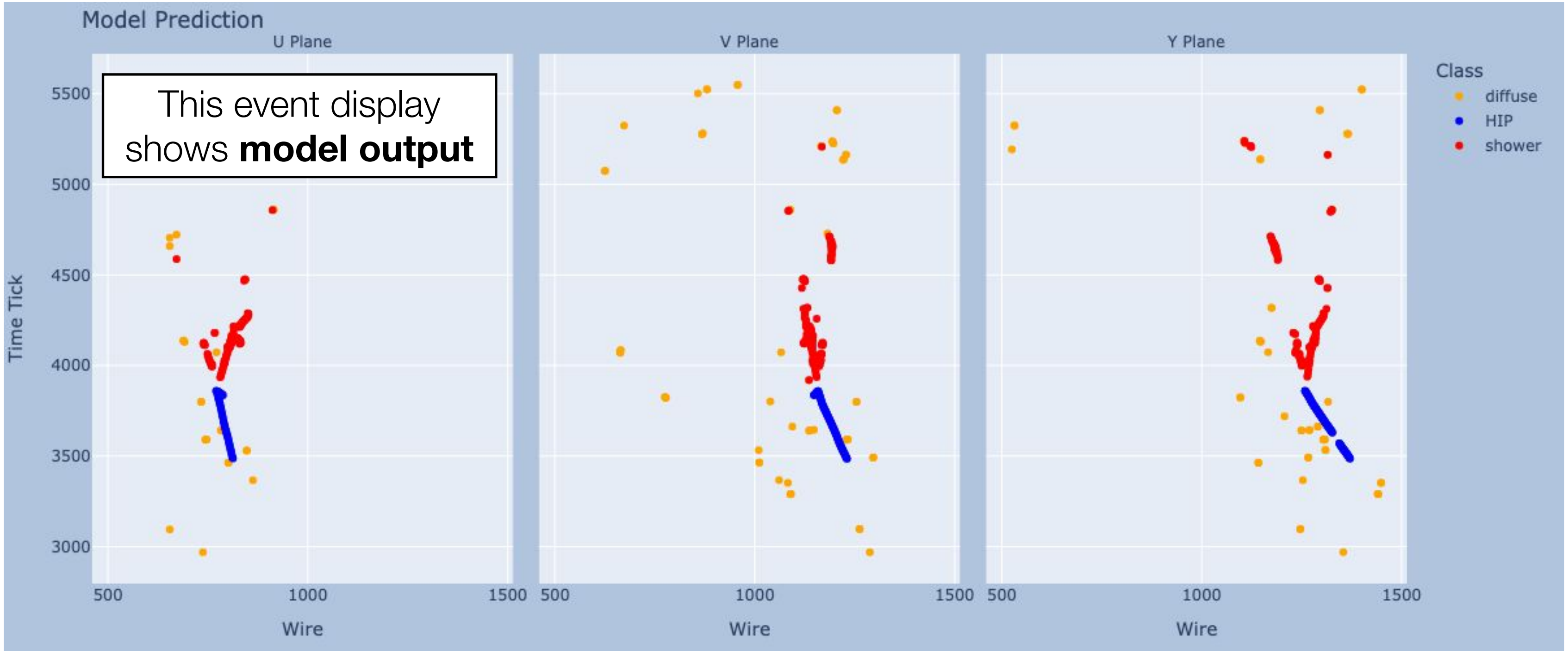
This event display shows **ground truth**

Simulated event with **proton track** and **electromagnetic showers**

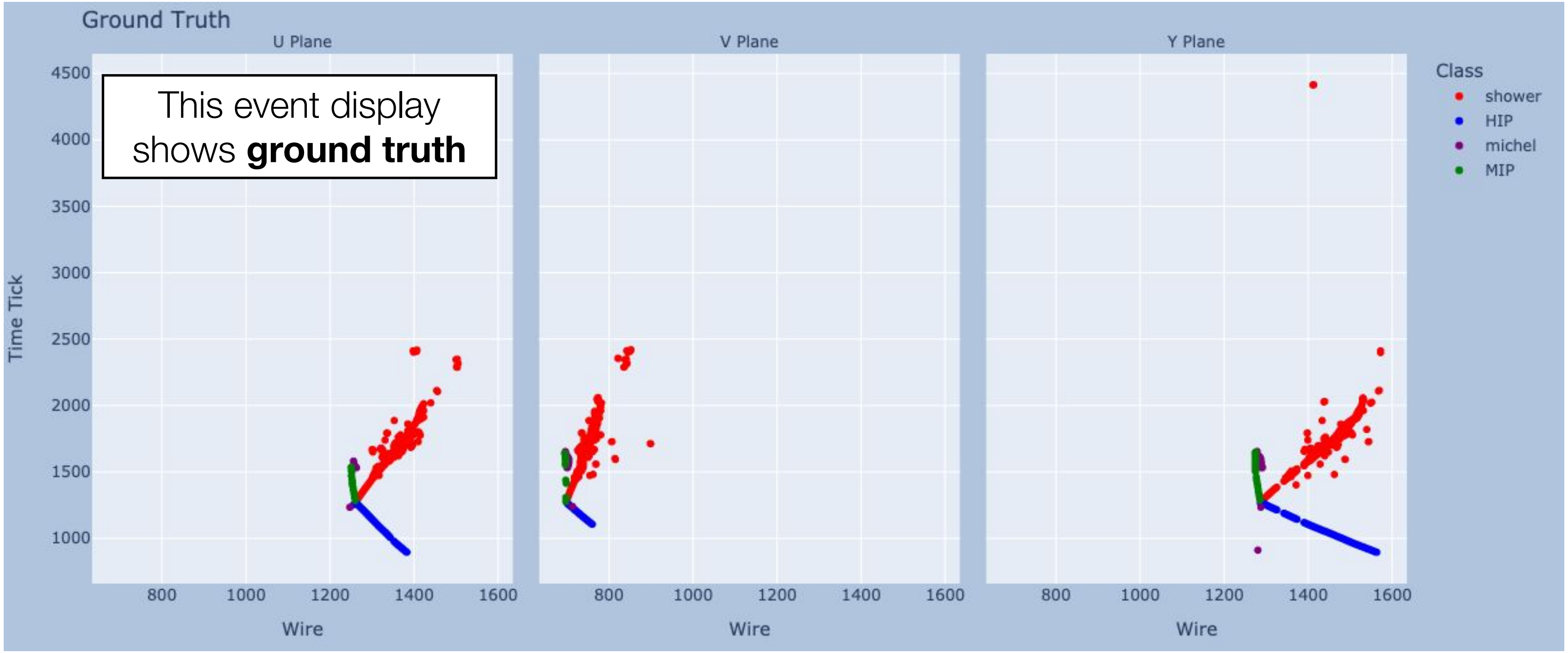
NC π^0 event display



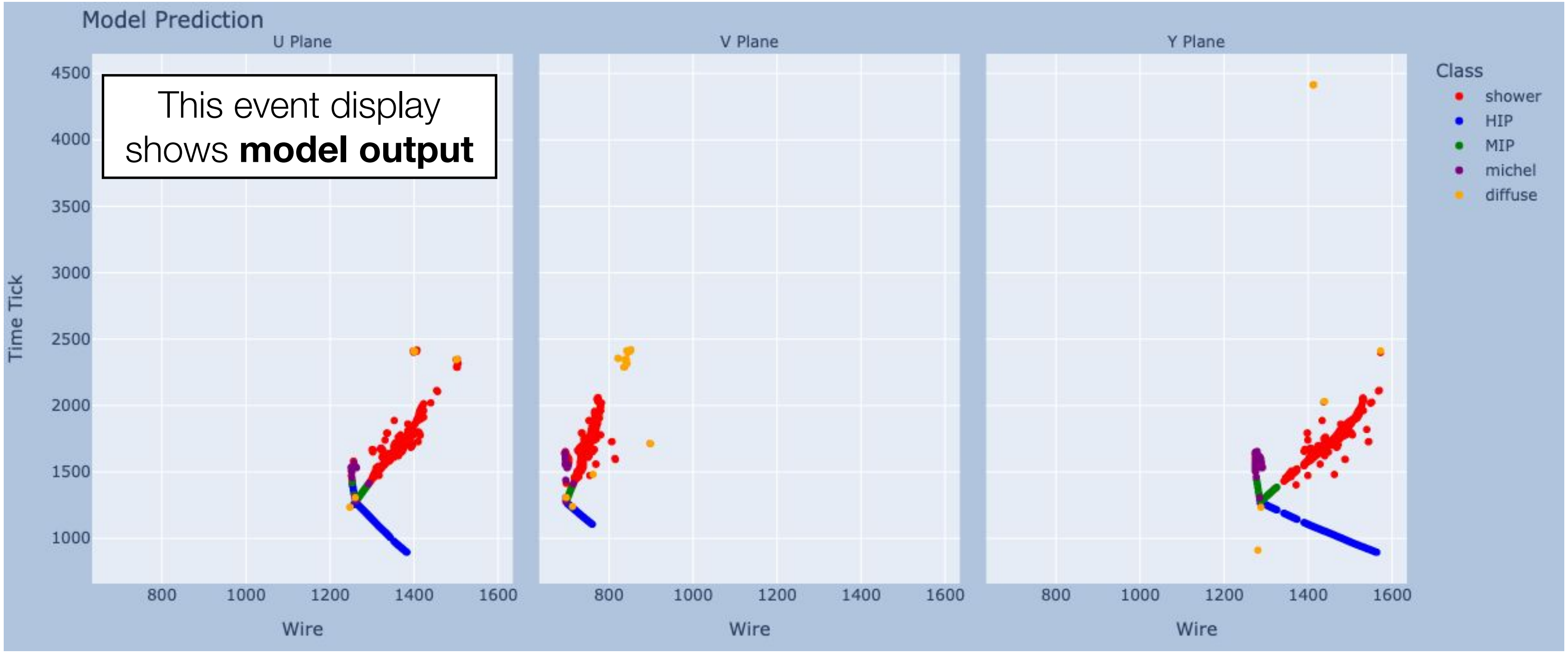
NC π^0 event display



Higher multiplicity event display

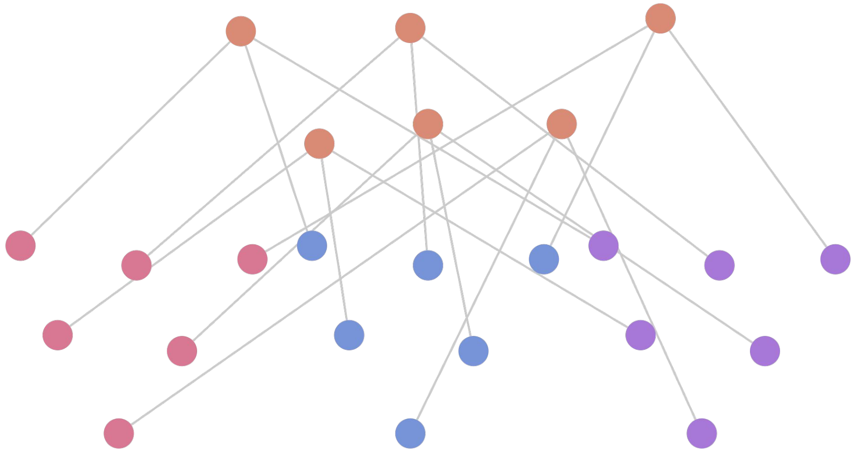


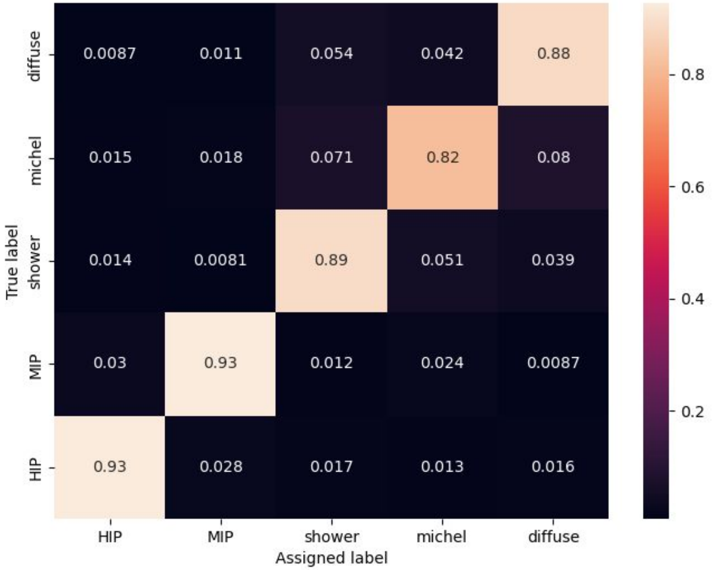
Higher multiplicity event display



TrkX

NuGraph2 performance



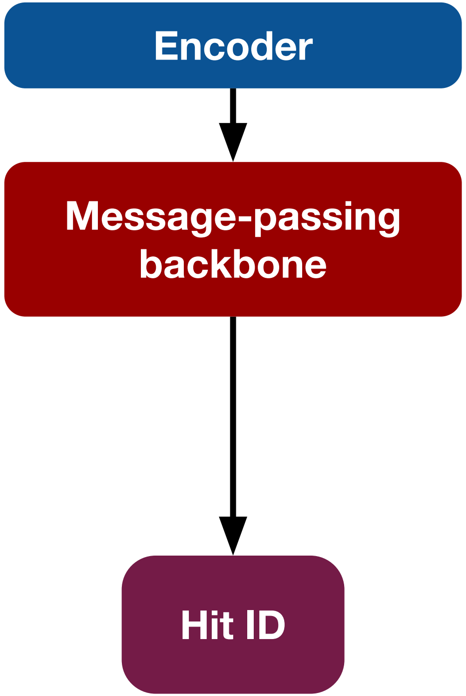


- Current model achieves **94% accuracy** in semantically labeling 2D detector hits.
- With 3D connections, consistency of representations between views is now around **98%**, compared to ~70% without).

- Model memory footprint **~500 kB.**
- Inference on CPU **~0.2 s / event**

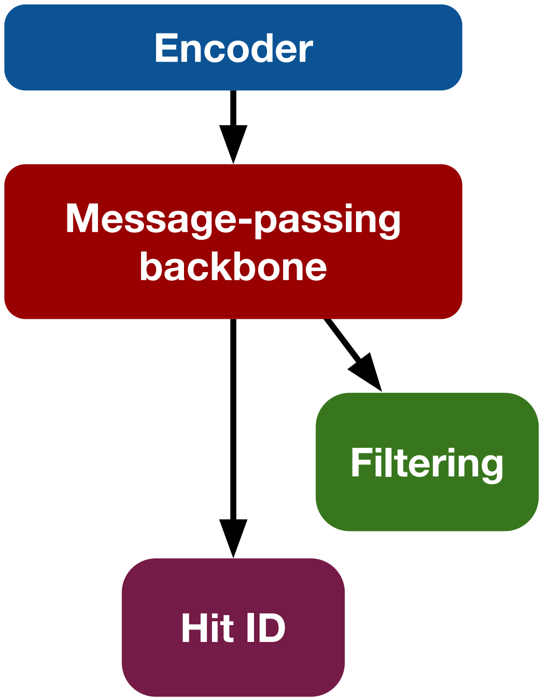
Potential future work: NuGraph3

- Plans for a potential next-generation **NuGraph3** involve leveraging the existing model backbone to train multiple decoders simultaneously:



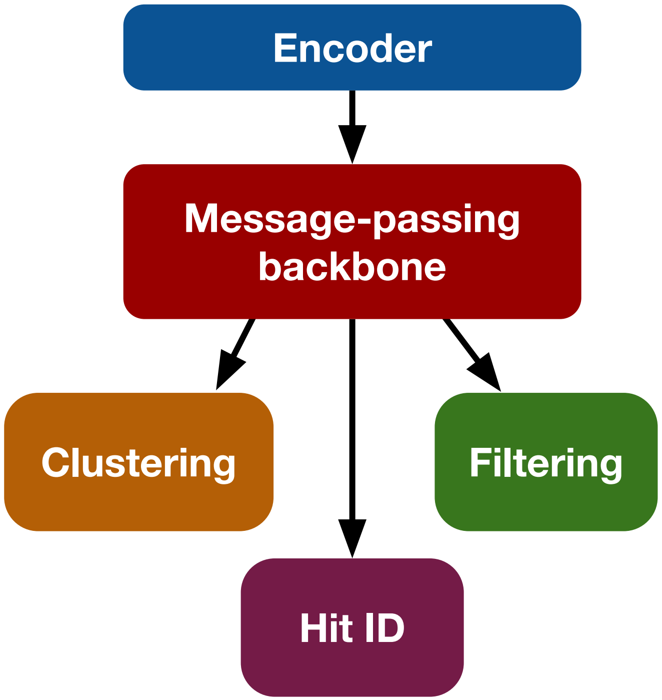
Potential future work: NuGraph3

- Plans for a potential next-generation **NuGraph3** involve leveraging the existing model backbone to train multiple decoders simultaneously:
    - **Binary signal/background classifier** to distinguish signal (ie. beam) hits from background (ie. cosmic) hits.



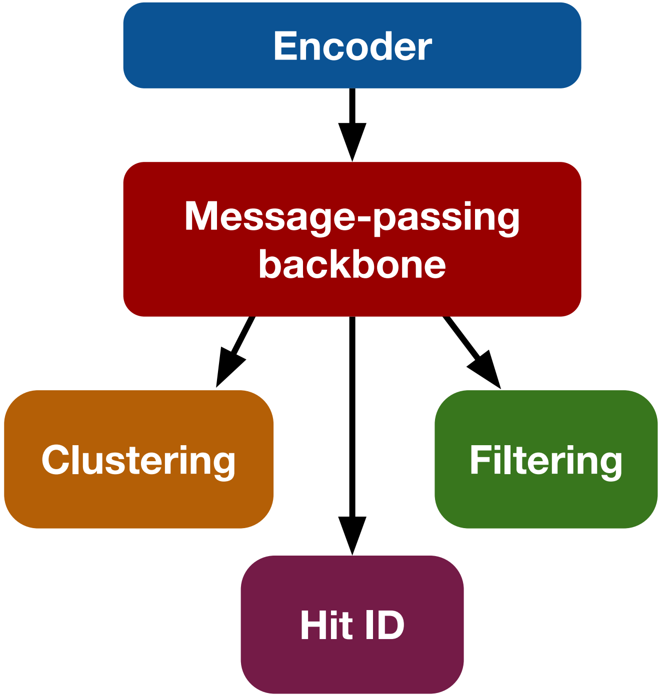
Potential future work: NuGraph3

- Plans for a potential next-generation **NuGraph3** involve leveraging the existing model backbone to train multiple decoders simultaneously:
    - **Binary signal/background classifier** to distinguish signal (ie. beam) hits from background (ie. cosmic) hits.
    - **Instance segmentation** to cluster together hits into particles, which provides **full particle reconstruction**.

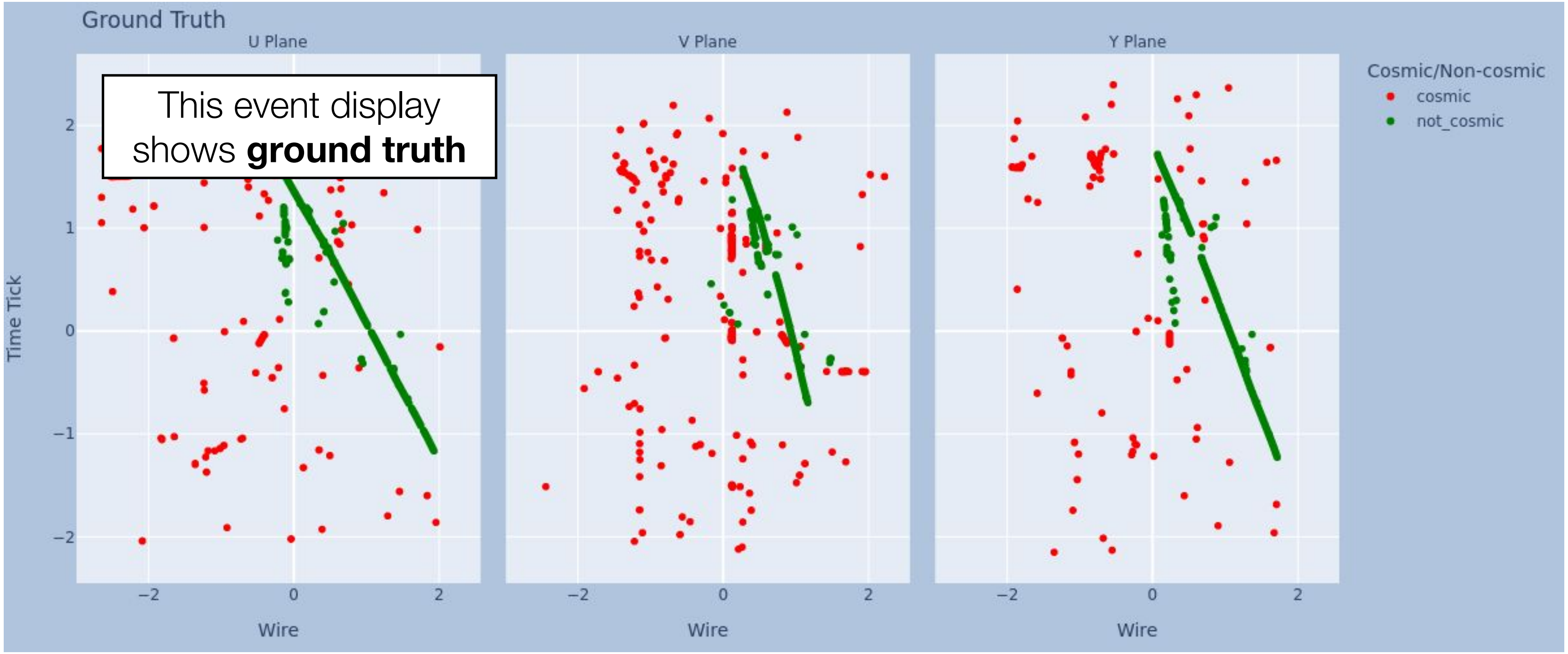


Potential future work: NuGraph3

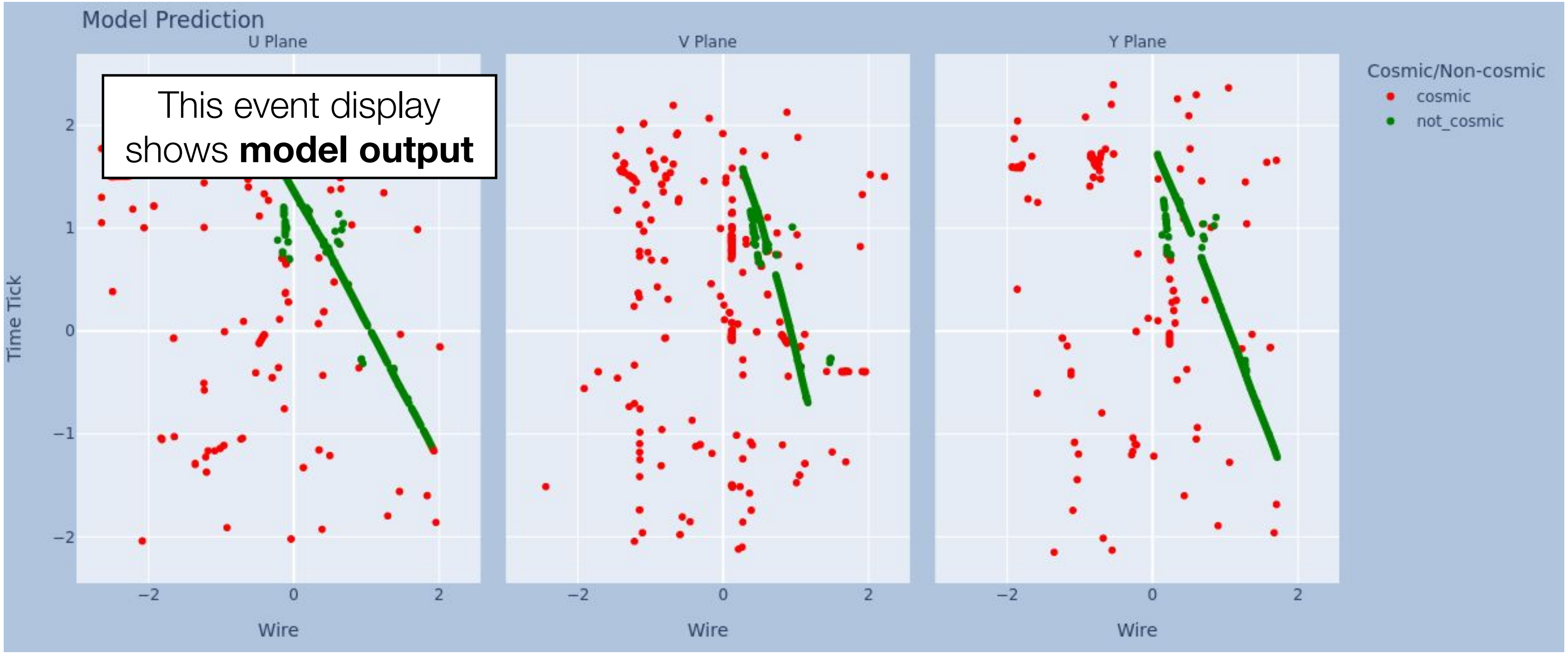
- Plans for a potential next-generation **NuGraph3** involve leveraging the existing model backbone to train multiple decoders simultaneously:
  - **Binary signal/background classifier** to distinguish signal (ie. beam) hits from background (ie. cosmic) hits.
  - **Instance segmentation** to cluster together hits into particles, which provides **full particle reconstruction**.
  - **Hierarchical graph network** for particle flow on reconstructed particles.



Background filtering

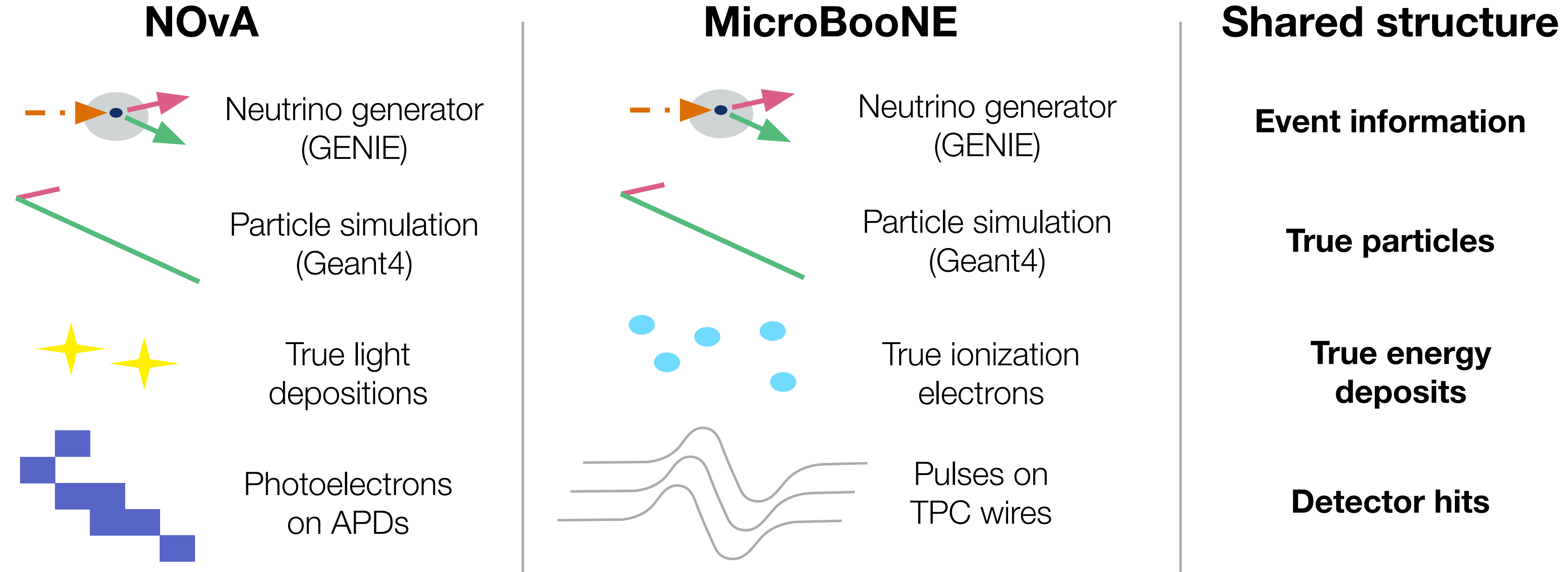


Background filtering



Common abstraction for neutrino experiments

- Although the details of many neutrino physics experiments vary, the majority of them share a common paradigm at a high level.



NuML & PyNuML

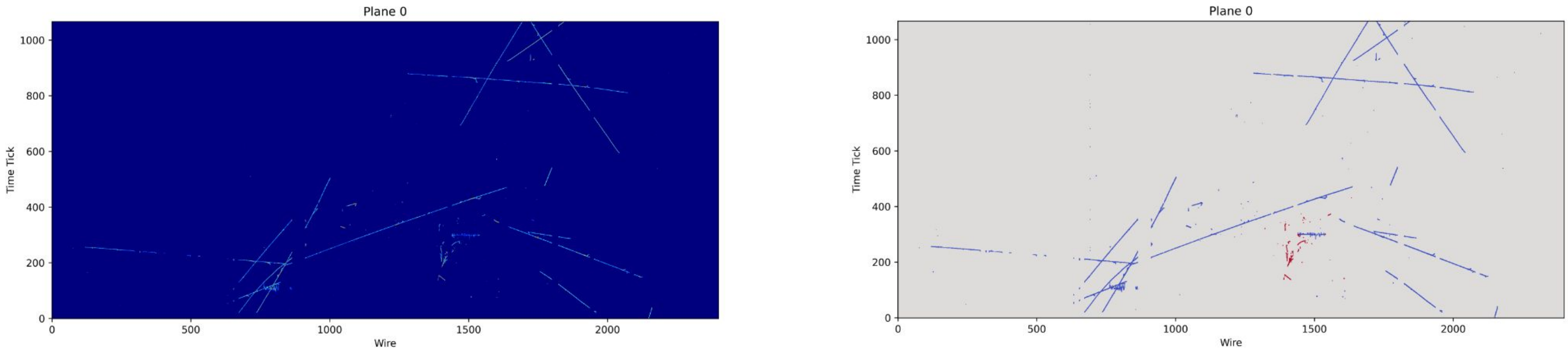
- Beyond the GNN architecture itself, many of the tools developed have **broad utility beyond our specific workflow**.

- The **NuML** package is a toolkit for writing **physics event records** to an **HDF5 file format**.
    - This file format is generic and can be shared across many experiments.

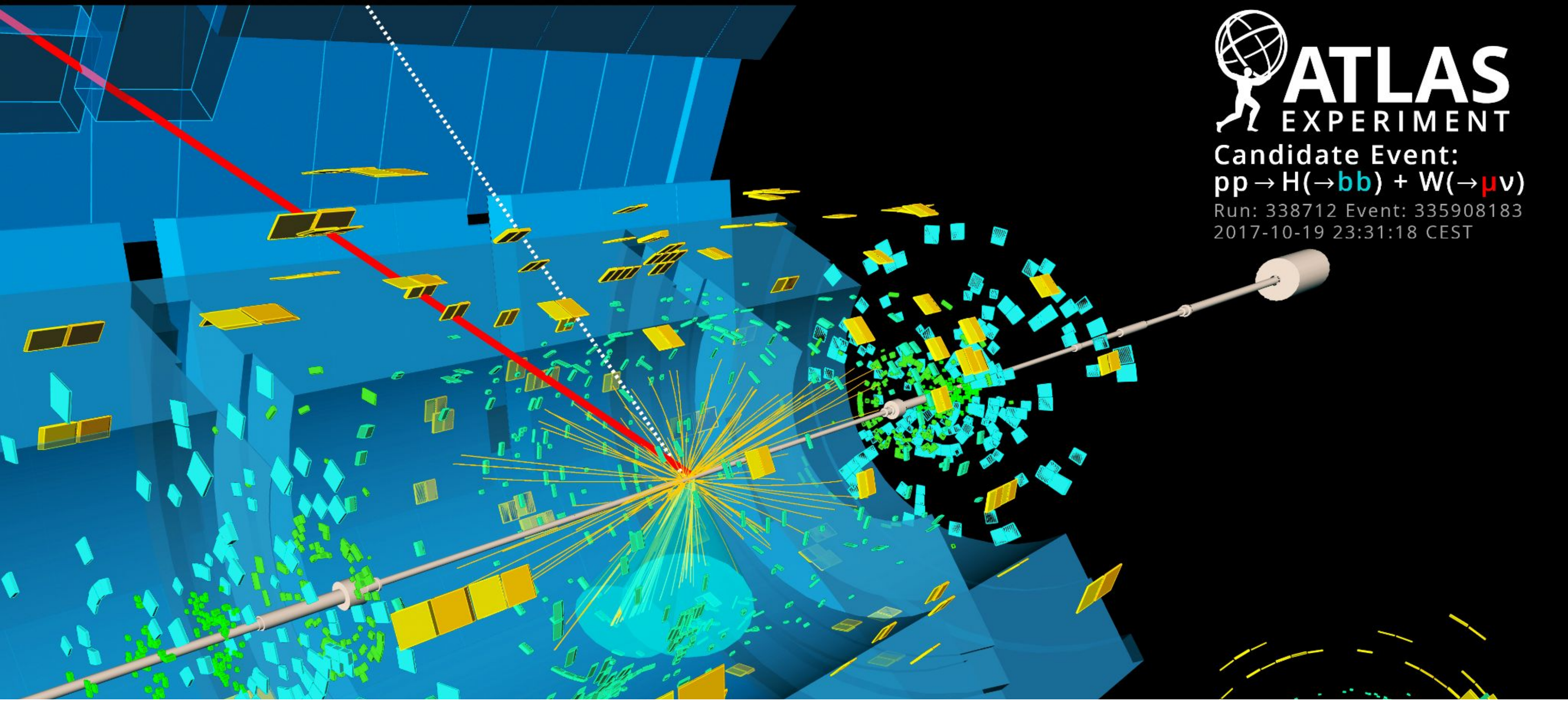
- The **PyNuML** package is designed to provide a **generic**, **accessible**, **efficient** and **flexible** solution for many of the necessary tasks in leveraging ML for particle physics.
    - Define particle ground truth labels for Geant4-simulated particles.
    - Arrange detector hits into ML objects, ie. graphs, CNN pixel maps, etc.
    - Efficiently preprocess ML inputs in parallel in HPC environments using MPI.

- Tools developed in collaboration with **Northwestern University** through **SciDAC's data institute**, **RAPIDS**.

MicroBooNE open data release



- ExaTrkX's NuML pipeline was used as the backbone of the MicroBooNE experiment's recent **open data release** (https://github.com/uboone/OpenSamples).
  - An HDF5 dataset was released containing raw tables of simulated particles, detector hits etc.
  - This data release enables **open development of machine learning algorithms** by making low-level simulation available to a broader machine learning community.
  - ExaTrkX's tools provide a powerful interface for accessing these files.



Geometric Deep Learning for the **Energy Frontier**

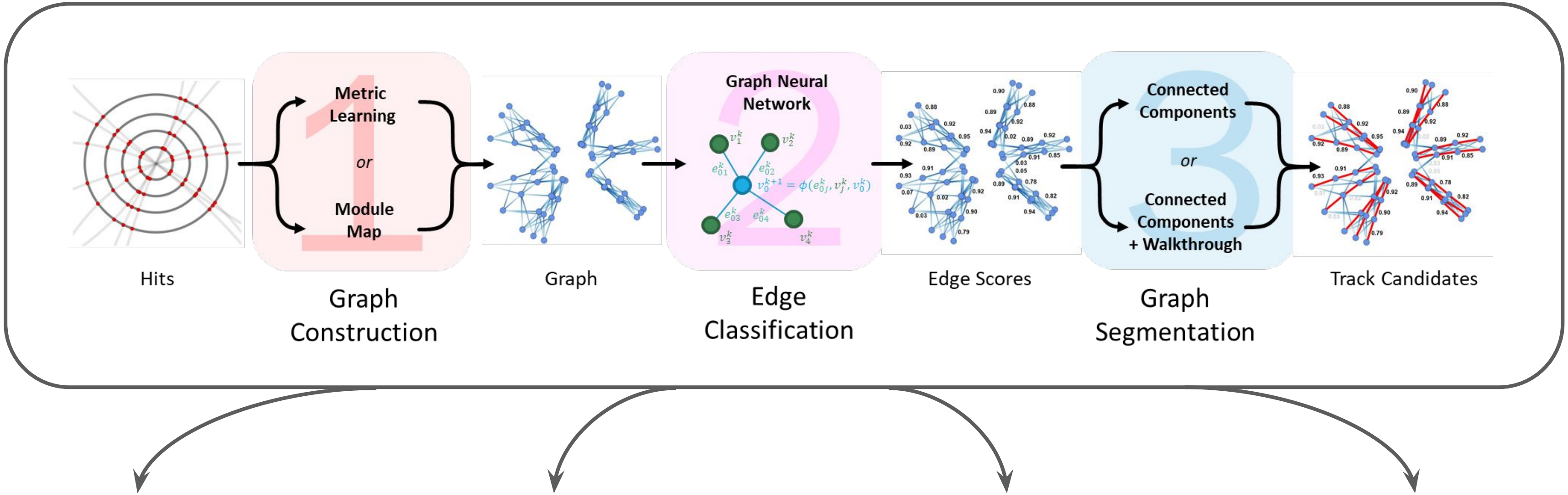
What's been accomplished

Exa.TrkX

HL-LHC GNN Prototype

What's been accomplished

What's next



Physics Optimization

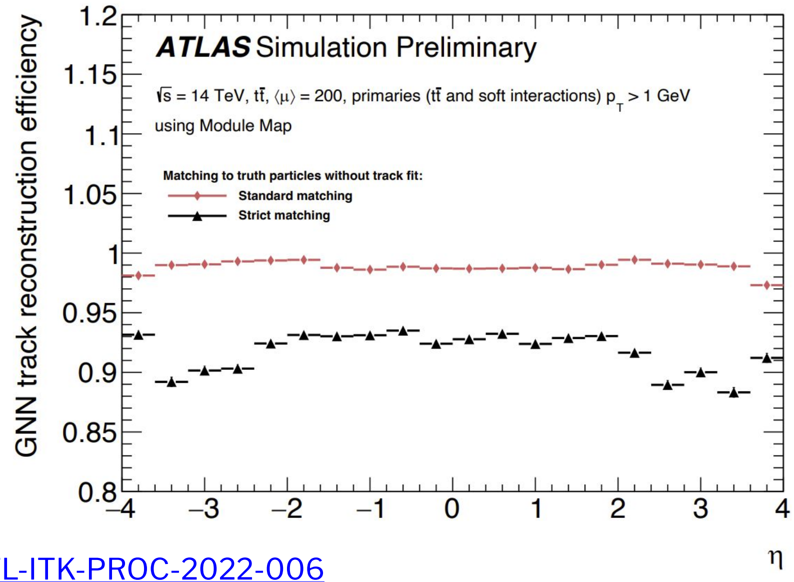
Computational Optimization

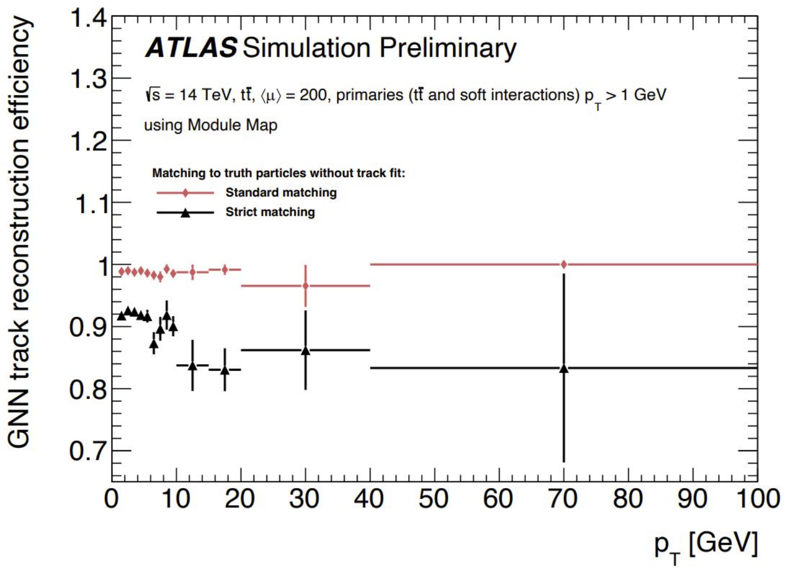
Experiment Integrations

Common Framework

Physics Optimization

What's been accomplished | What's next





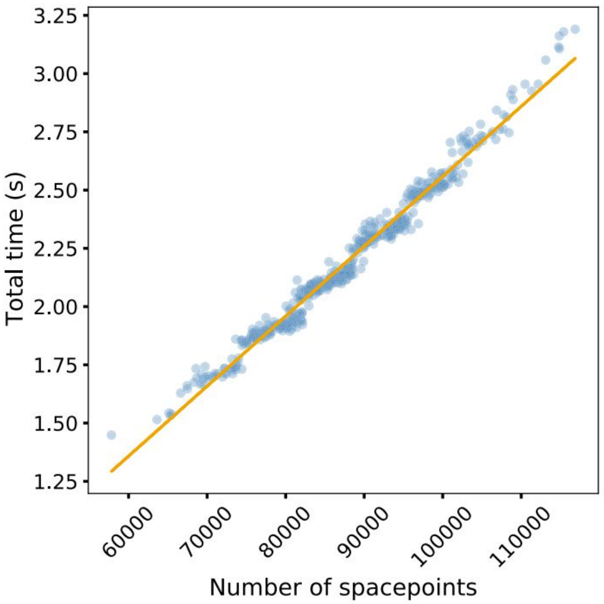
ATL-ITK-PROC-2022-006

Standard matching: single-matched particles with $f_{truth} = 0.5$
Strict matching: double-matched particles with $f_{reco} = 1.0$

- Fake rate is $O(10^{-3})$ using standard truth matching

Computational Optimization

What's been accomplished   What's next



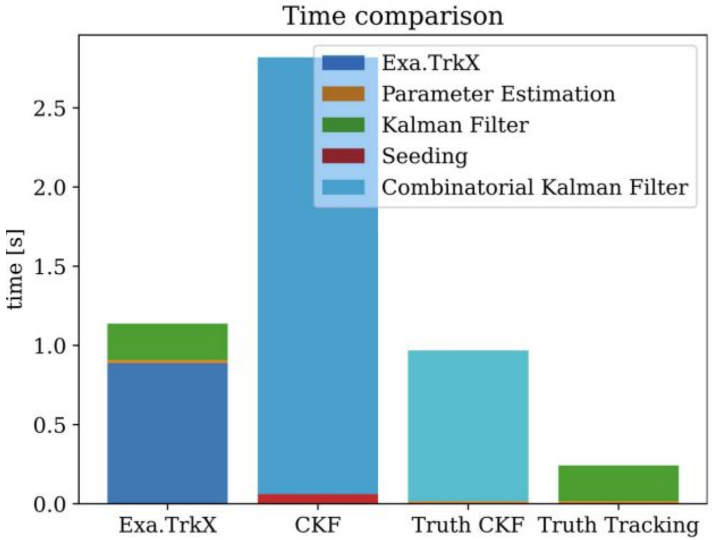
	Baseline	Faiss	cuGraph	AMP	FRNN
Data Loading	0.0022 ± 0.0003	0.0021 ± 0.0003	0.0023 ± 0.0003	0.0022 ± 0.0003	0.0022 ± 0.0003
Embedding	0.02 ± 0.003	0.02 ± 0.003	0.02 ± 0.003	0.0067 ± 0.0007	0.0067 ± 0.0007
Build Edges	12 ± 2.64	0.54 ± 0.07	0.53 ± 0.07	0.53 ± 0.07	0.04 ± 0.01
Filtering	0.7 ± 0.15	0.7 ± 0.15	0.7 ± 0.15	0.37 ± 0.08	0.37 ± 0.08
GNN	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03
Labeling	2.2 ± 0.3	2.1 ± 0.3	0.11 ± 0.01	0.09 ± 0.008	0.09 ± 0.008
Total time	$15 \pm 3.$	3.6 ± 0.6	1.6 ± 0.3	1.2 ± 0.2	0.7 ± 0.1

- Physics is important, but GNNs shine in **scaling behavior**
- When development began, graph-based pipeline started required **15 sec for TrackML**
- Implemented custom Fixed Radius Nearest Neighbor (FRNN) algo., cuGraph Connected Components algo., and Mixed Precision inference
- Now have **sub-second TrackML inference** on 16Gb V100 GPU
- Inference time scales **approximately linearly** across size of event, in TrackML

Experiment Integrations

What's been accomplished
What's next

- Integrated into ACTS (a Common Tracking Software), and available for general usage
    - Collaboration with PhD students in ACTS to performance validation tests
- Integrated into official ATLAS simulation and validation framework
    - Allows for apples-to-apples comparison of timing and physics performance
- Beginning integration with workflows of other HEP communities, such as FPGA-based trigger system

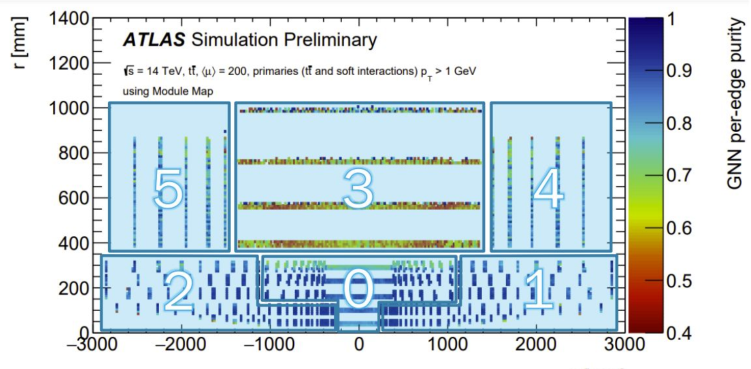


Plot courtesy of Benjamin Huth, *ACTS, University of Regensburg*

Case Study: Heterogeneity
ITk Hardware Heterogeneity

What's been accomplished | What's next

- Different performance across different hardware in ATLAS ITk
  - Have proof-of-concept: Encode different points with **Heterogeneous GNN**
  - Can recover efficiency lost in a homogeneous GNN
- Next step: extend to *any* type of detector input
  - Working on first version of "full-detector cluster-level representation" - pixel clusters, strip spacepoints and calorimeter cells in single graph
  - Can include hierarchical structures (e.g. track-like nodes)



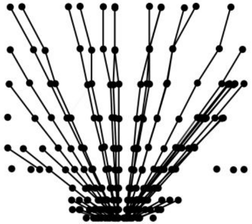
6-region = 6 nodes, 21 edges

Model	F1 Score
HomoGNN 3-feature	0.966
HeteroGNN 3-feature	0.961
HomoGNN 9-feature	0.968
HeteroGNN 9-feature	**0.975**

Summing Up

Geometric Deep Learning for Energy & Intensity Frontiers

HepTrkX



GNN edge classification

ExatrkX Intensity Frontier

- Segmentation and classification
- Object clustering
- Partial heterogeneity
- 2D → 3D mapping
- HDF5 → ML composable interfaces

ExatrkX Energy Frontier

- Large graph memory handling
- Graph segmentation algorithms
- High-throughput nearest neighbors
- Metric Learning
- Multi-stage pipeline software

Future Work

- Full heterogeneity of hardwares
- Cluster-to-particle hierarchical models
- Multi-task and multi-modal generalist ML

Notes for CSAID Roadmap [GC, JBK]

- Seeking additional funding from DOE with the goals of future work described in the previous slides, especially interested in heterogeneous graphs and hierarchical models
  - exploring one year project extension scoped for 1 FTE effort at Fermilab
  - This provides continuity with the exa.trkx collaboration together with LBNL

- Also seeking (seed?) funding to explore explainable and physics-informed AI + visualization techniques to improve GNN understanding, performance, and development process.

- Promising technique in terms of physics and especially computing performance: integration into production workflows and our computing ecosystem is high priority and should be supported by the lab