University of CINCINNATI

Machine Learning in Accelerator-Based Neutrinos

Jeremy Hewes

Fast Machine Learning for Science Workshop

2nd December 2020

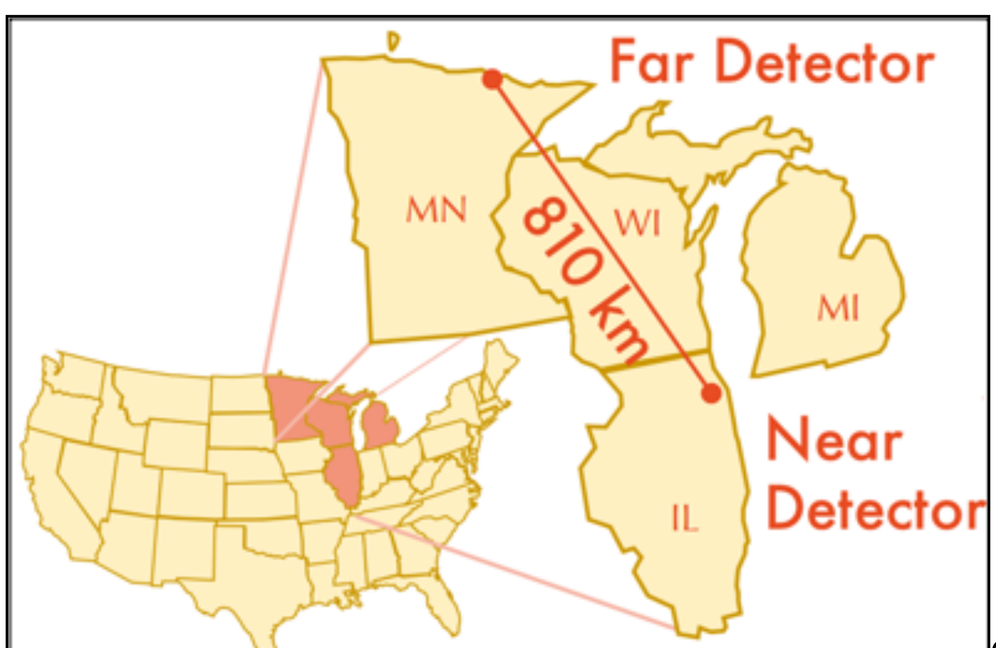
Overview

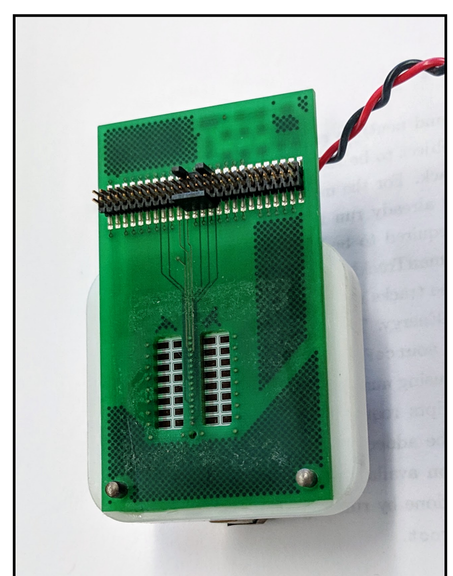
- Overview of ML methods used in accelerator-based neutrino experiments.
  - NOvA
  - DUNE
  - Short-Baseline Neutrino (SBN) Program
- Summarise tools currently in use, as well as new techniques being developed.
  - Much of what this talk covers is not really “Fast ML.”
  - Most ML applications in neutrino physics are offline reconstruction.
  - Hopefully highlight some deficiencies in current workflows.

NOvA

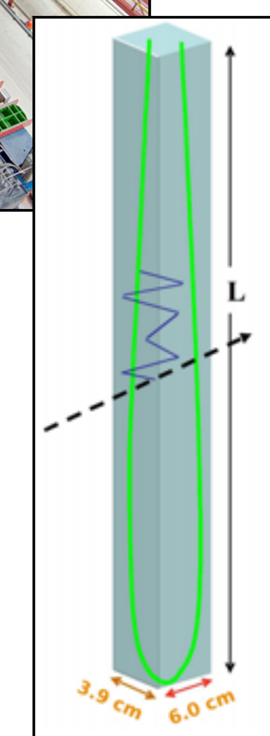
NOvA experiment

- **NOvA** is a long-baseline accelerator experiment based at Fermilab.
- Measures neutrinos from Fermilab's **NuMI beam**.
- Functionally identical near and far detectors.
  - Plastic and liquid scintillator **sampling tracking calorimeter**.
  - **ND:** 1km baseline, FNAL, 300 tons.
  - **FD:** 810km baseline, Ash River, 14 kt, 14 mrad off-axis.

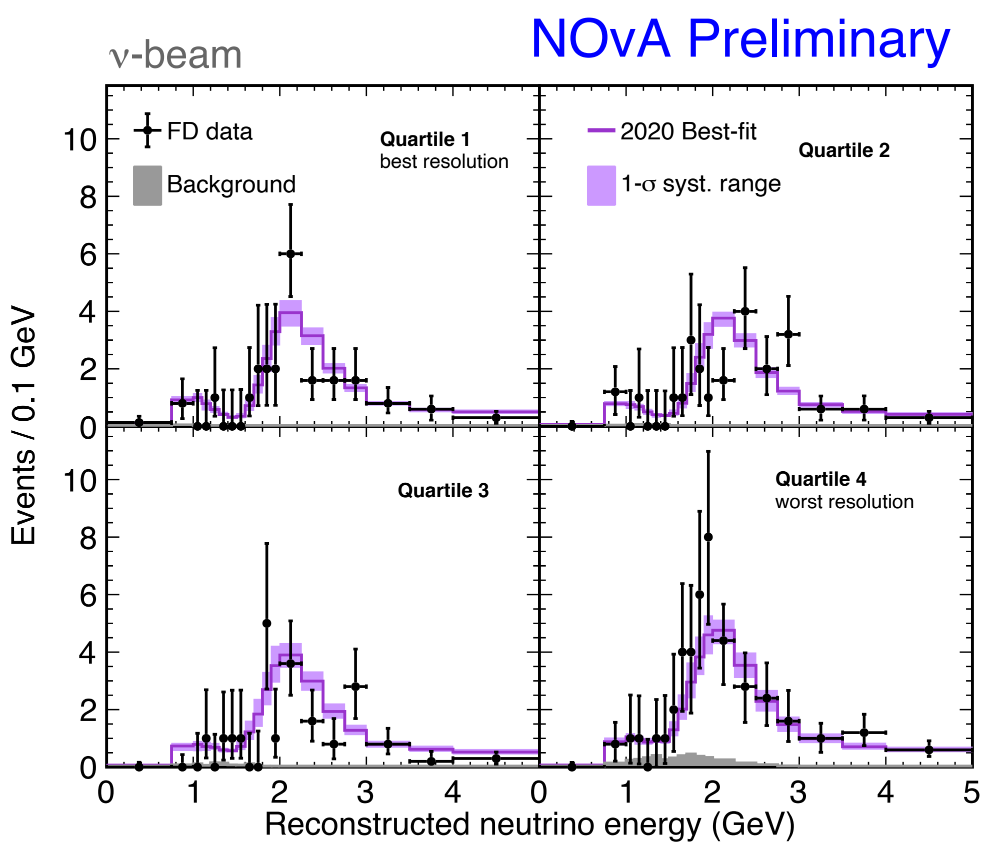


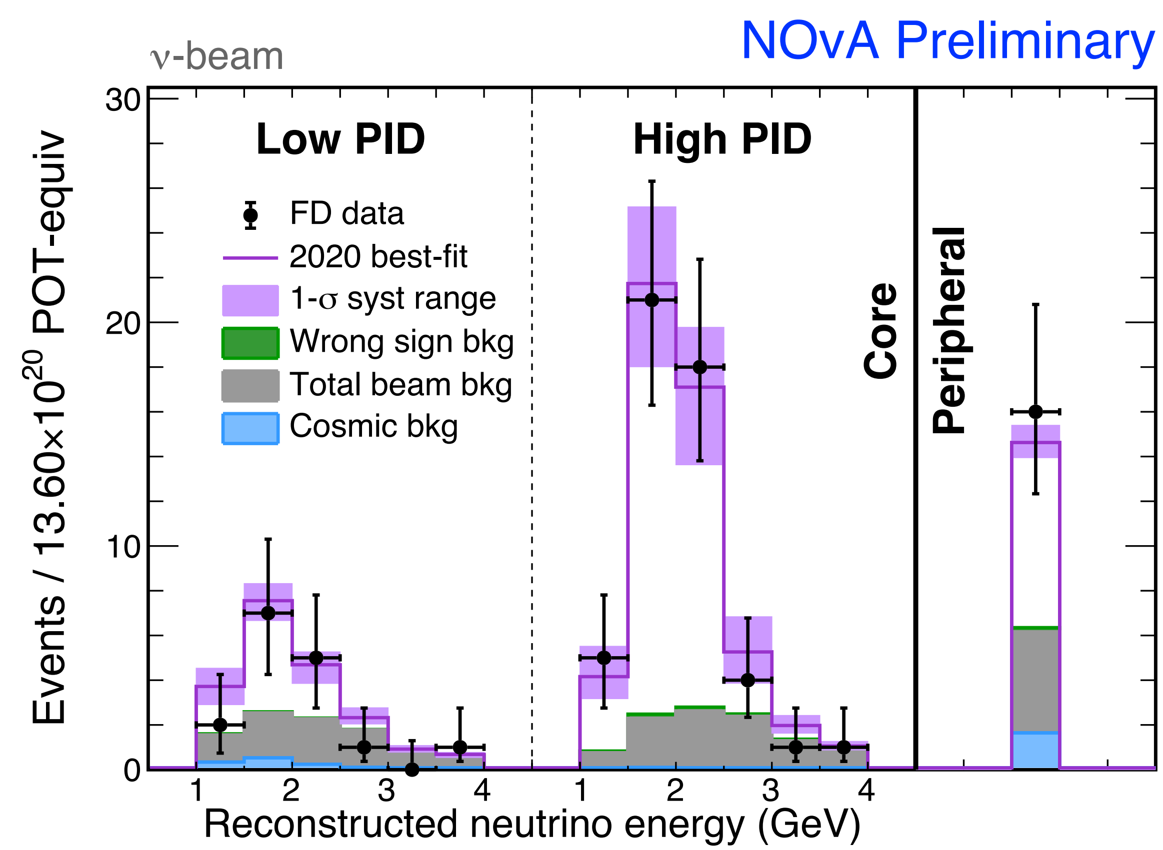


- Charged particles produce light when propagating through scintillator.
- Picked up by wavelength-shifting fibers (right) and amplified by avalanche photodiodes (left).



NOvA physics





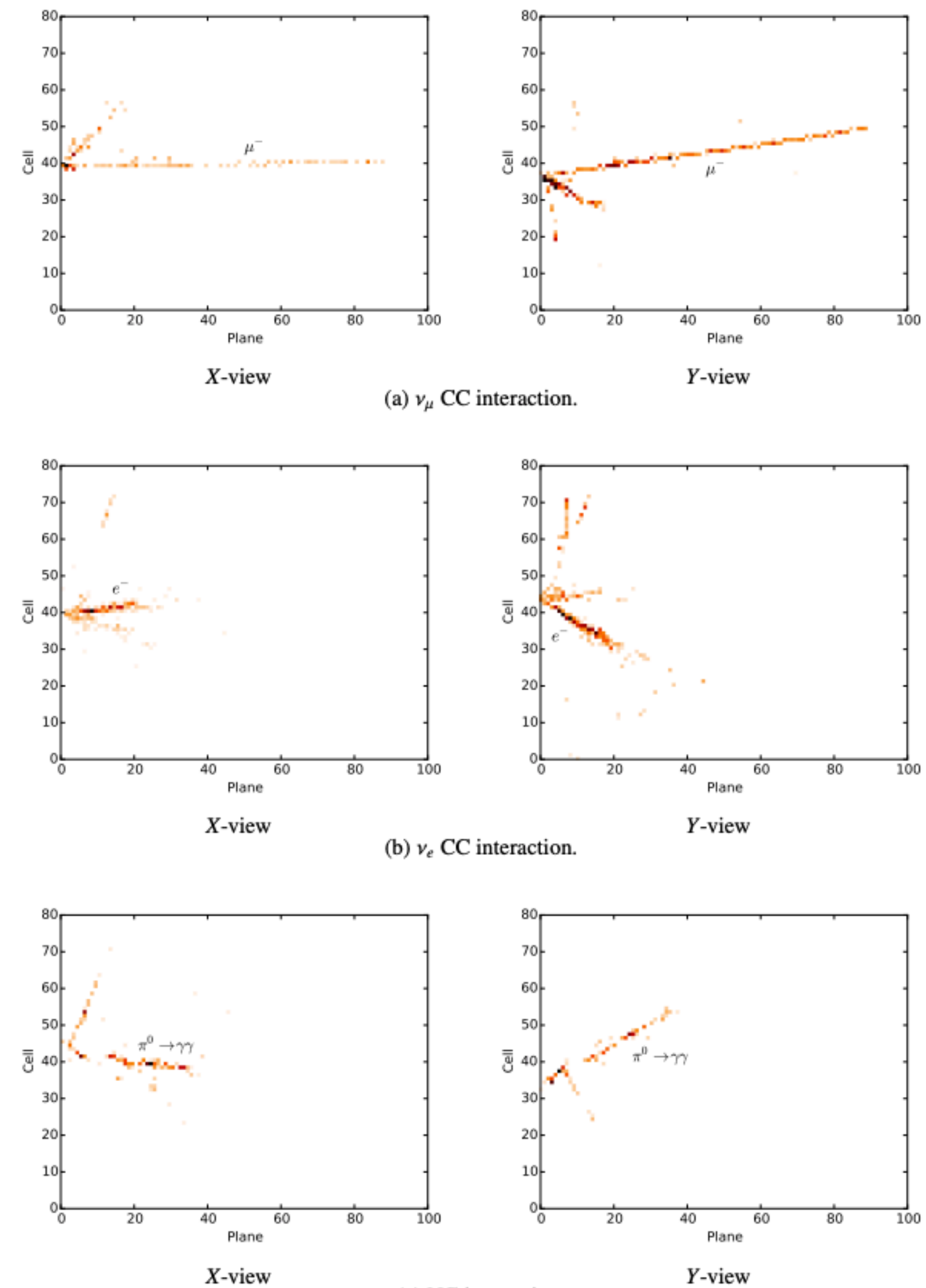
ν_μ **sample**

ν_e **sample**

- NOvA's ability to measure neutrino oscillations relies on an ability to disambiguate different neutrino interaction types.
- Measuring oscillation parameters requires pure charged-current (CC) ν_e and ν_μ samples, and robust rejection of neutral-current (NC) and cosmic events.

NOvA CVN

- NOvA utilises a CNN neutrino interaction classifier called the **Convolutional Visual Network (CVN)**.
- Segmented detector provides **top-down** and **side-on** views of an interaction.
- Lower-level reconstruction identifies neutrino interaction candidates, and selects a region of interest around each interaction.
- Train CNN to disambiguate CC ν_μ & ν_e interactions from NC and cosmic interactions.
- Use a modified MobileNet v2 architecture which operates independently on the two views, then merges them before a final set of layers.



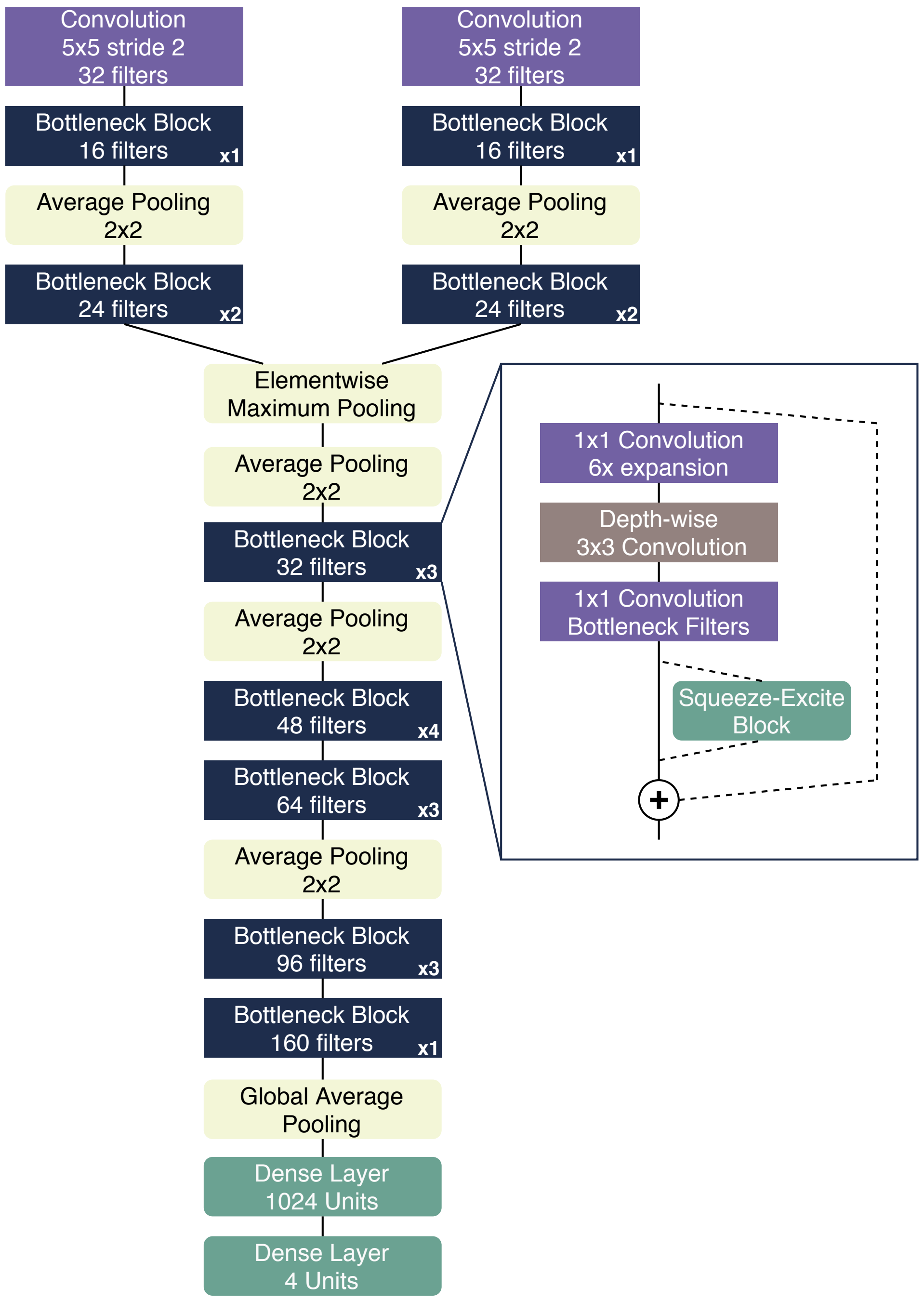
(a) ν_μ CC interaction.

(b) ν_e CC interaction.

(c) NC interaction.

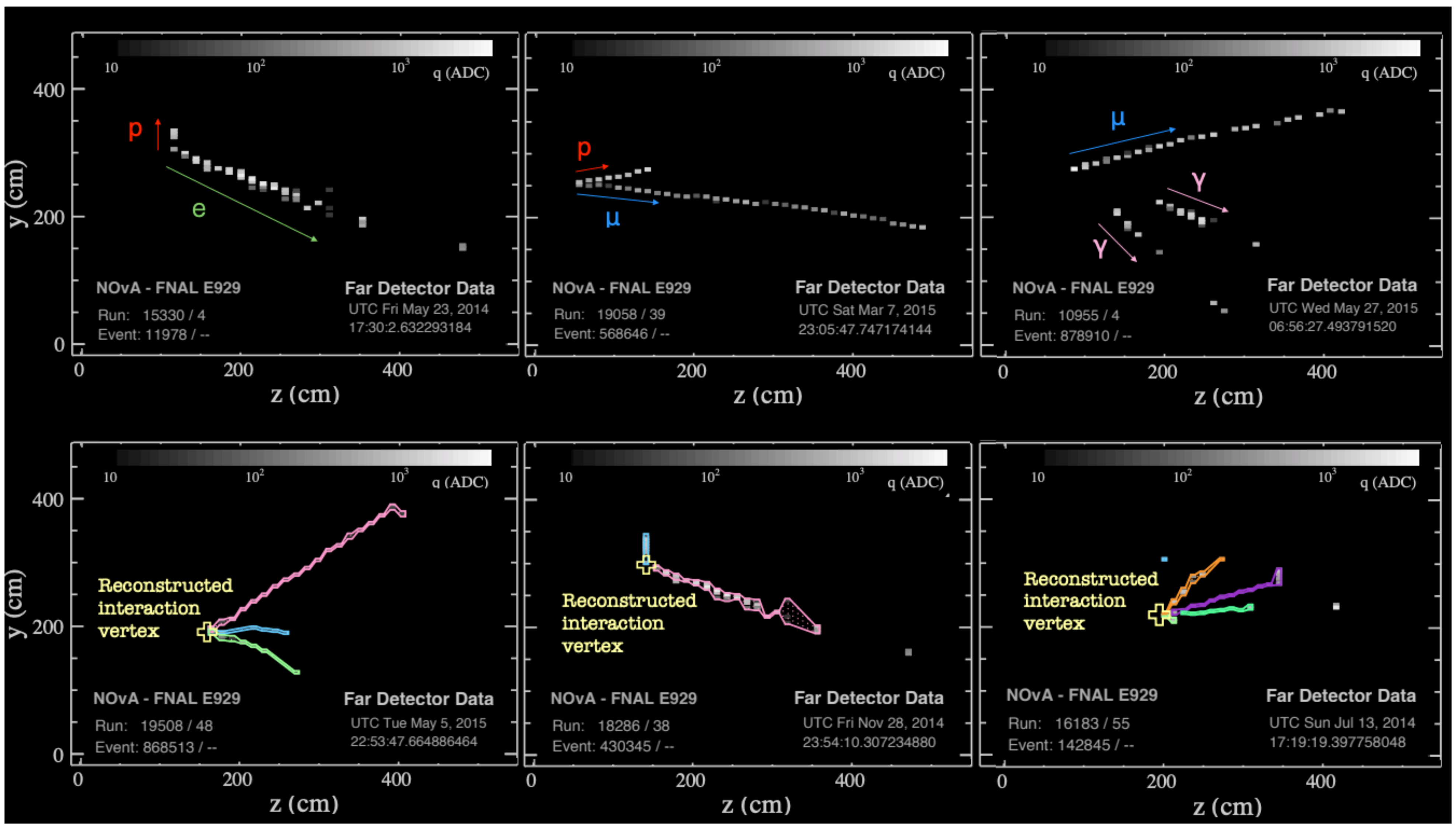
NOvA CVN

- NOvA utilises a CNN neutrino interaction classifier called the **Convolutional Visual Network (CVN)**.
- Segmented detector provides **top-down** and **side-on** views of an interaction.
- Lower-level reconstruction identifies neutrino interaction candidates, and selects a region of interest around each interaction.
- Train CNN to disambiguate CC ν_μ & ν_e interactions from NC and cosmic interactions.
- Use a modified MobileNet v2 architecture which operates independently on the two views, then merges them before a final set of layers.



CVN Particle Classification

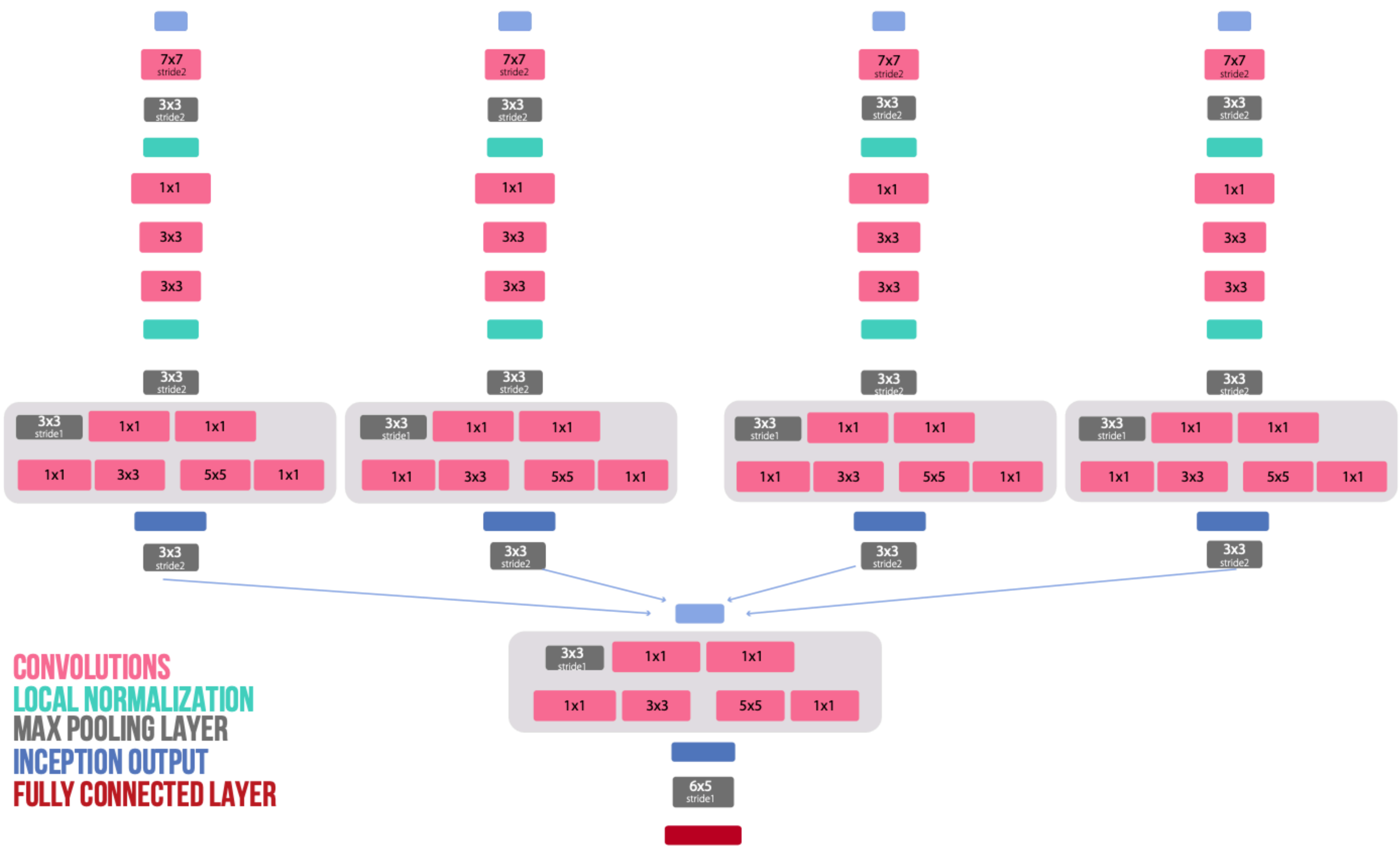
- CVN extension developed which operates on each particle cluster independently.
- Select individual clusters within interaction, and feed them into CVN for particle **type** classification.



arXiv:1906.00713

CVN Particle Classification

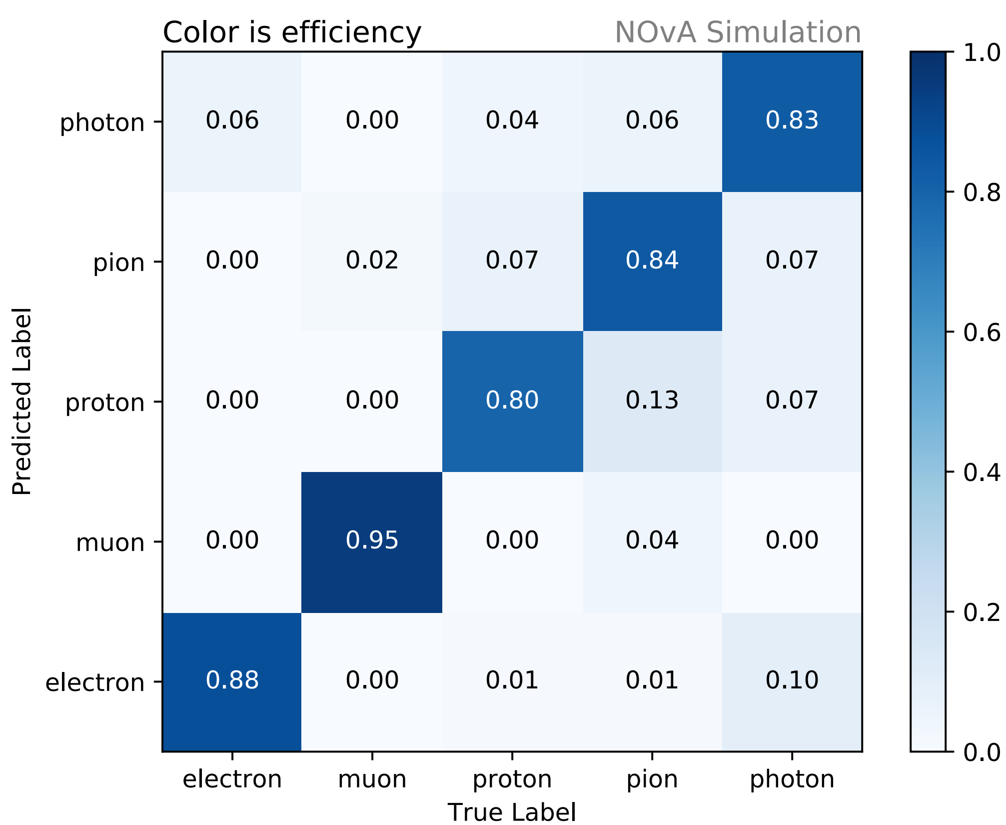
- CVN extension developed which operates on each particle cluster independently.
- Select individual clusters within interaction, and feed them into CVN for particle **type** classification.
- Constructing a four-stack network including the full event for context improves network performance.

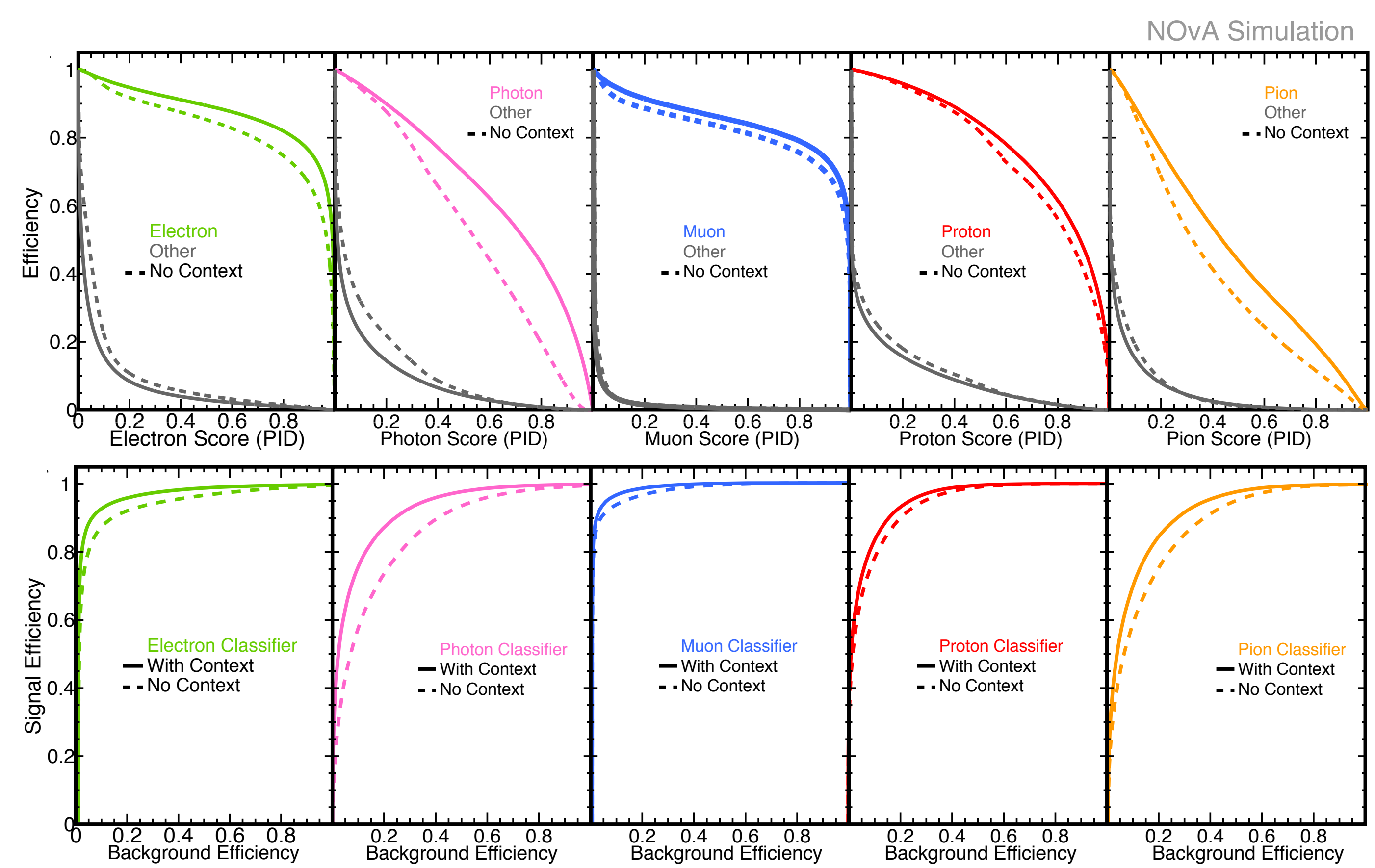


arXiv:1906.00713

CVN Particle Classification

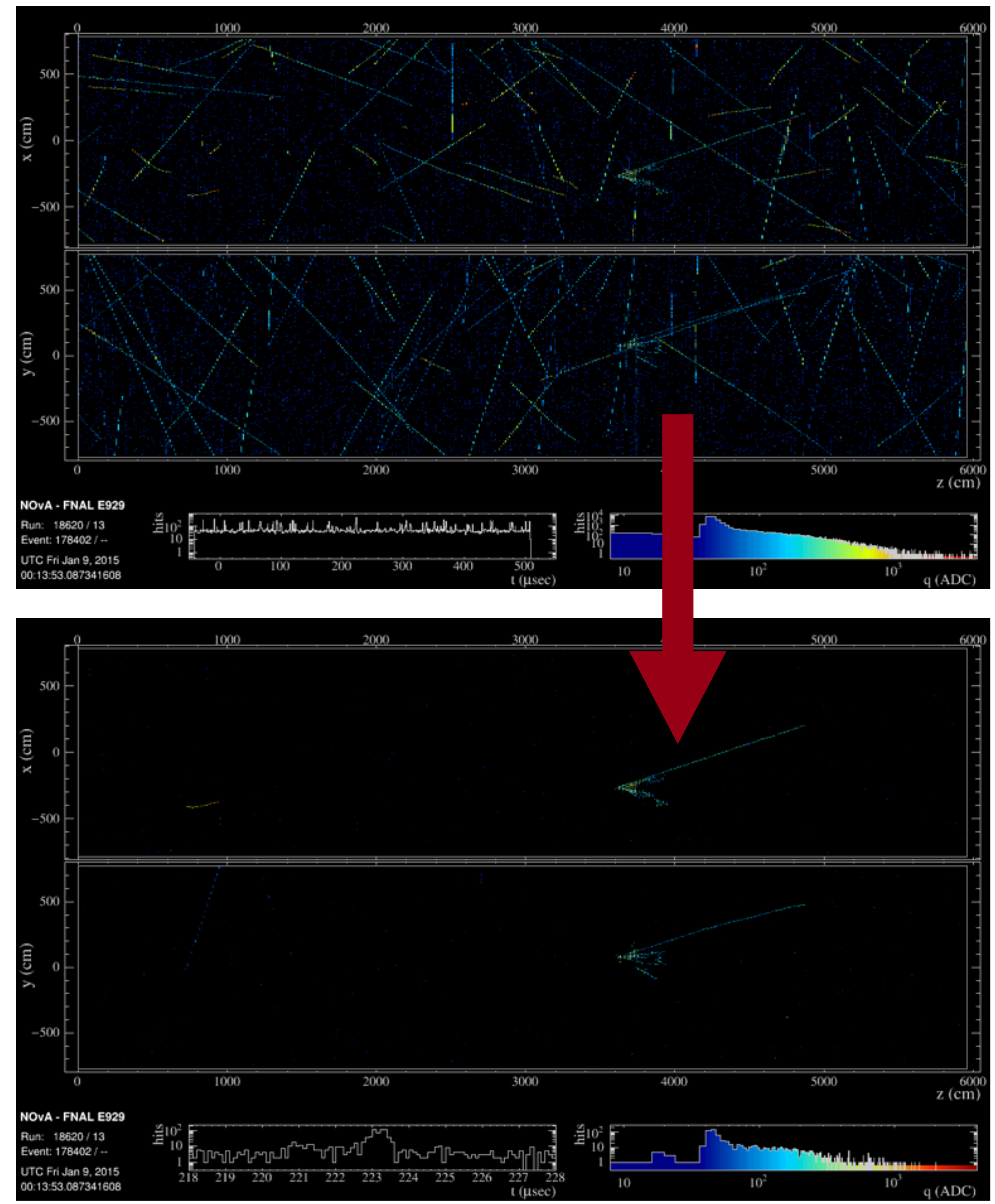
- CVN extension developed which operates on each particle cluster independently.
- Select individual clusters within interaction, and feed them into CVN for particle **type** classification.
- Constructing a four-stack network including the full event for context improves network performance.





Cosmic rejection CNN

- NOvA's far detector is on the surface, so sees an enormous flux of cosmic interactions.
  - Many TBs of cosmic data recorded annually.
- Many 550μs readout windows are fully processed and reconstructed, despite being easy to reject.
- CNN classifier trained to identify and reject readout windows without interesting physics.
  - Weed out easy-to-remove backgrounds without sacrificing any potential signal.
  - Not run online, but in early stages of keep-up data processing to reduce computing cost of downstream reconstruction.



Training and inference in NOvA

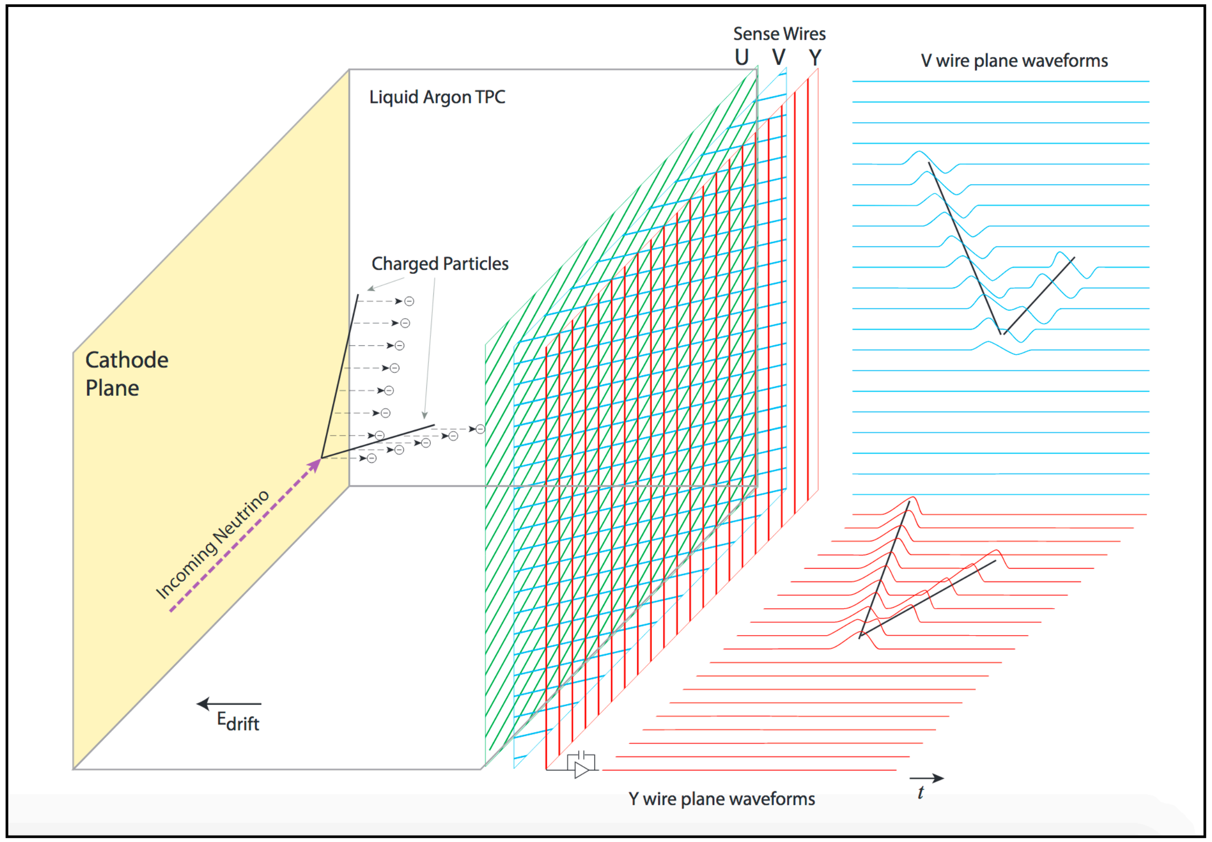
- CNN training typically occurs on GPU clusters such as Fermilab's Wilson Cluster, outside of typical analysis framework.
- NOvA's analysis tools utilise a "Common Analysis Framework (CAF)."
  - Recent developments allow for simple translation of CAF files into HDF5 format, greatly streamlining the training procedure.
- Simulation and reconstruction are performed in the Art framework (heavy-duty, C++).
  - Inference is run event-by-event on CPU in Tensorflow via C++.
  - Logistically the simplest solution, but also extremely slow.
  - More sophisticated pipelines could greatly streamline this process.

Liquid Argon TPCs

Liquid Argon TPCs

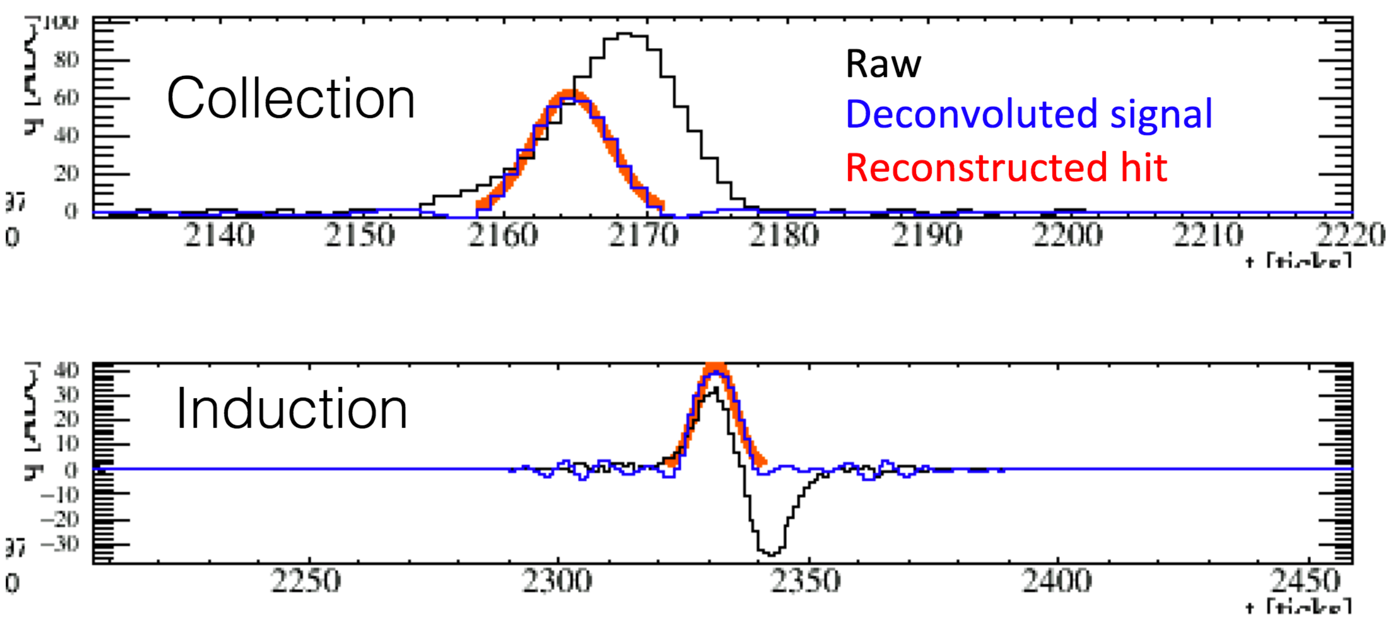
- Liquid Argon Time Projection Chambers (LArTPCs) are currently a very important detector technology for neutrino physics.
  - At FNAL: MicroBooNE, Icarus, SBND.
  - Future: DUNE (70kT LArTPC deep underground, plus near detector).

- 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 – produce **high-resolution images**.

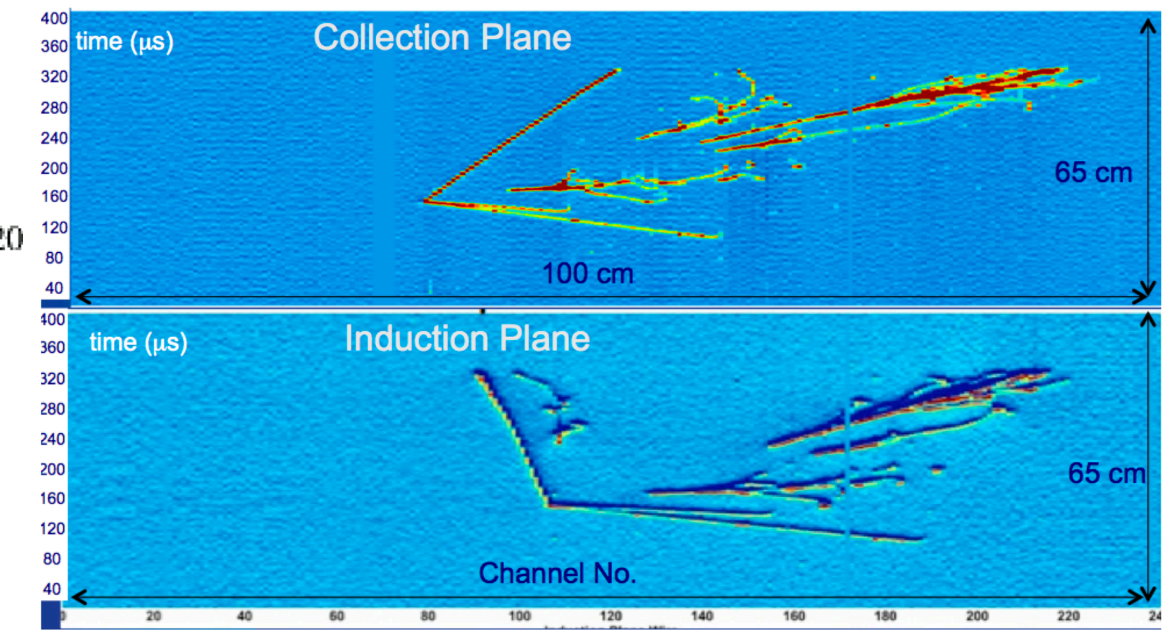


University of CINCINNATI

Standard reconstruction chain



T. Yang (ICHEP 2016)

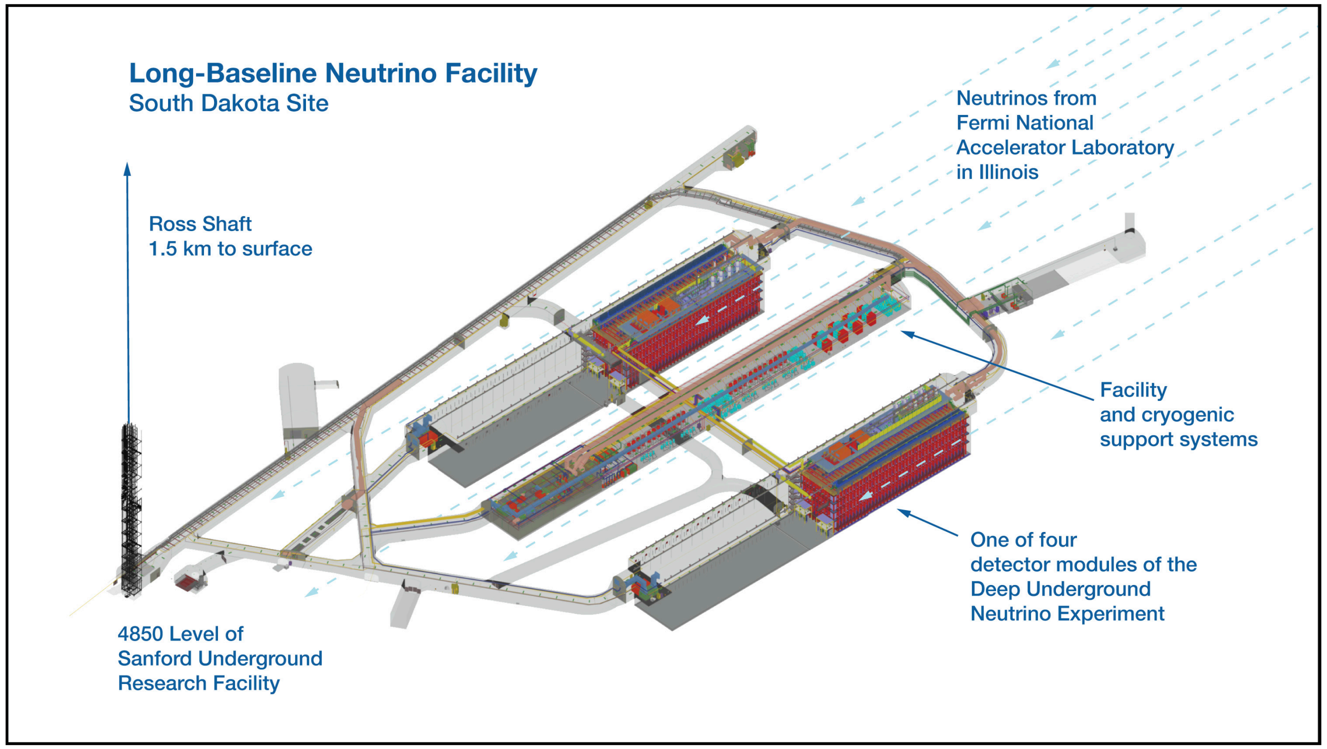


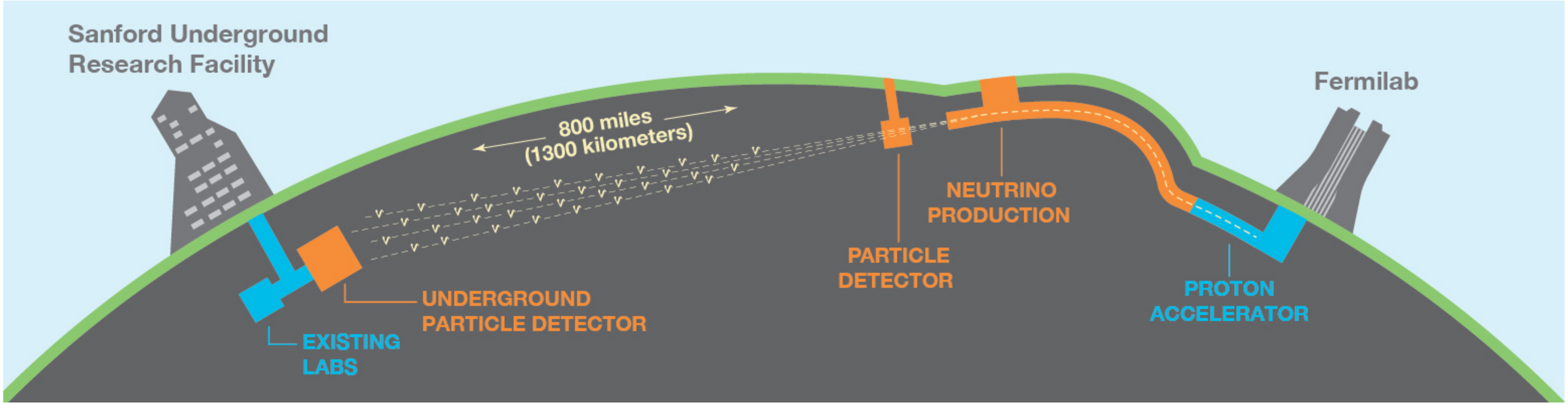
ArgoNeuT data event

- Raw TPC output is wire waveforms.
- Waveforms are then deconvolved and hit-finding is applied to produce Gaussian hits.
- Each wire plane forms a 2D image in the space of wire vs readout time.
- Three wire planes angled at -36°, 0°, 36° provide three 2D representations of the event.
- These 2D representations can be used to construct a 3D representation of the event.

Deep Underground Neutrino Experiment

- **70 kt** LArTPC, **1.5km** underground.
- High exposure in low-background environment.
- **Modular design:**
    - Four large detector modules.
    - Each consists of 200 individual TPCs.
    - Transformations necessary to combine data across multiple modules in 2D.





University of CINCINNATI

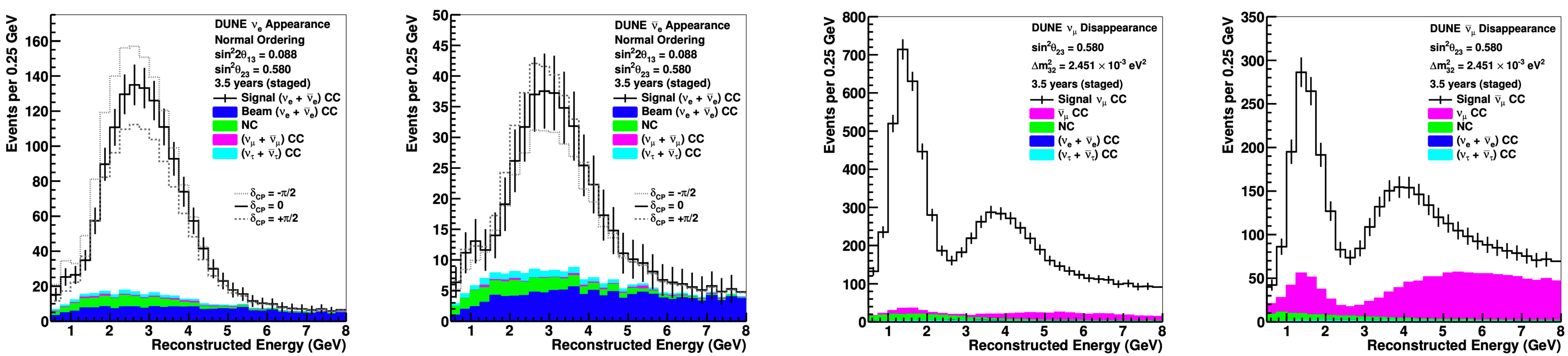
Machine Learning in DUNE

- Liquid Argon Time Projection Chambers (LArTPCs) provide extremely rich and detailed information due to their high resolution.
- Taking full advantage of this information requires sophisticated event reconstruction techniques.
- Machine learning is increasingly being adopted due to its ability to outperform traditional methods:
  - CNNs for **event ID**.
  - 3D CNNs for **pixel** and **instance segmentation**.
  - **Graph neural networks** (GNNs) for both high- and low-level reconstruction.
- DUNE's modular Pandora reconstruction framework beginning to explore ML-based reconstruction modules.

Neutrino oscillations at DUNE

- Neutrino oscillation measurements at DUNE will be performed using similar techniques to those previously discussed for NOvA.
- High-power beam, high-resolution LArTPC detectors and large detector mass will allow neutrinos to be collected with much larger statistics.
- Current sensitivities are produced using simulated fluxes & cross-sections paired with parameterised reconstruction assumptions.
- Current baseline method is CVN trained on fully simulated DUNE neutrino interactions.

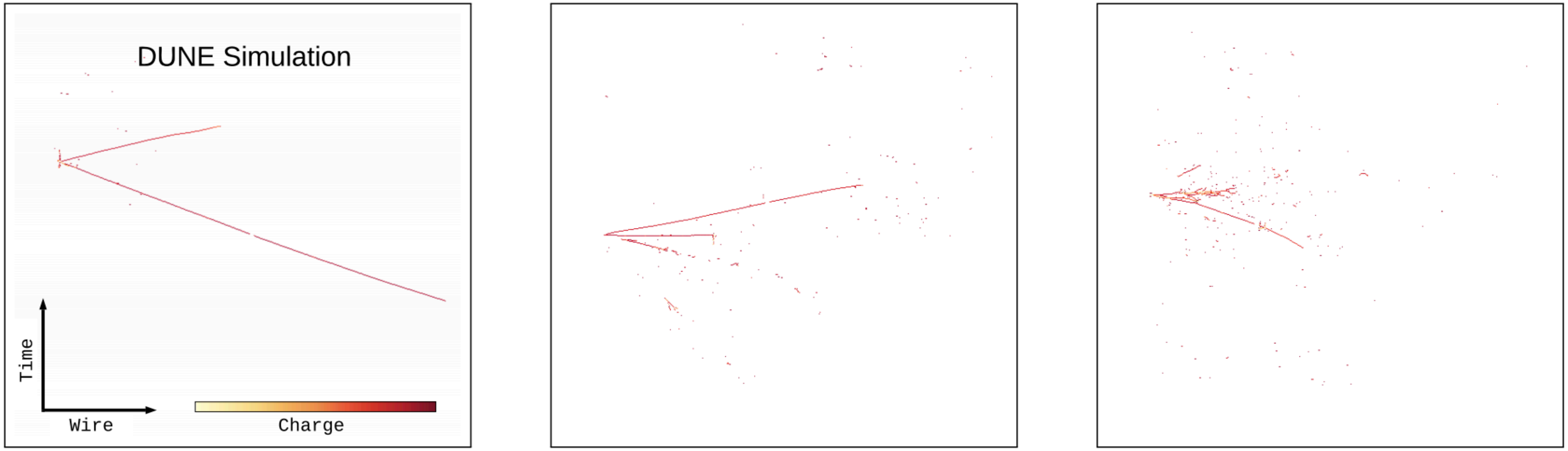
ν_e $\bar{\nu}_e$ ν_μ $\bar{\nu}_\mu$



arXiv:2002.02967

DUNE CVN

- Implementation is the same as NOvA:
  - Perform low-level reconstruction (wire deconvolution, hit finding).
  - Identify a region of interest, and produce pixel map inputs.
  - Stitch together pixel maps across multiple TPC modules.
  - Train to classify neutrino flavour (no cosmics due to DUNE's depth).



(a) 1.6 GeV CC ν_μ. (b) 2.2 GeV NC $1\pi^+$. (c) 2.4 GeV NC $1\pi^0$.

arXiV: 2006.15052

DUNE CVN

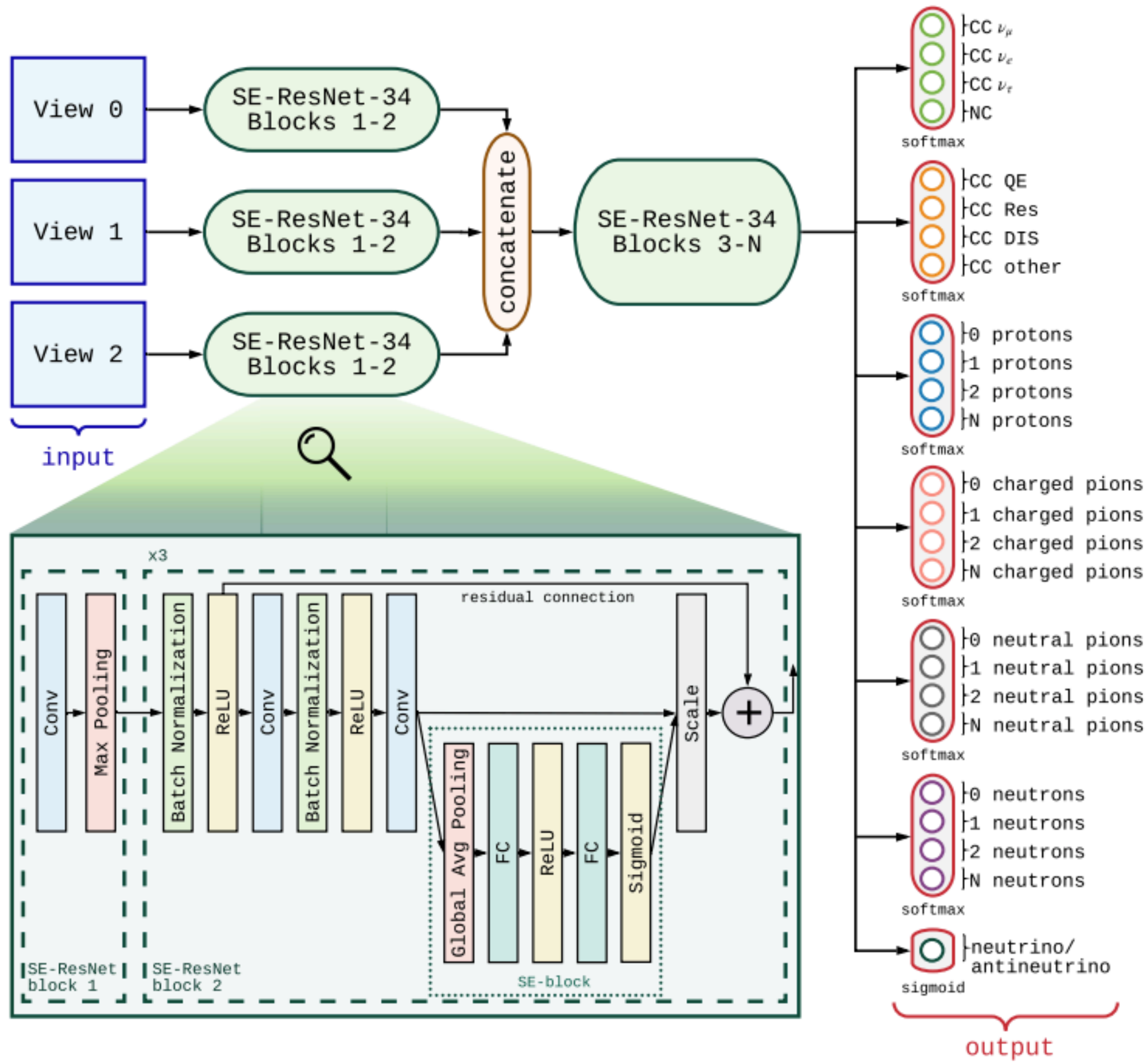
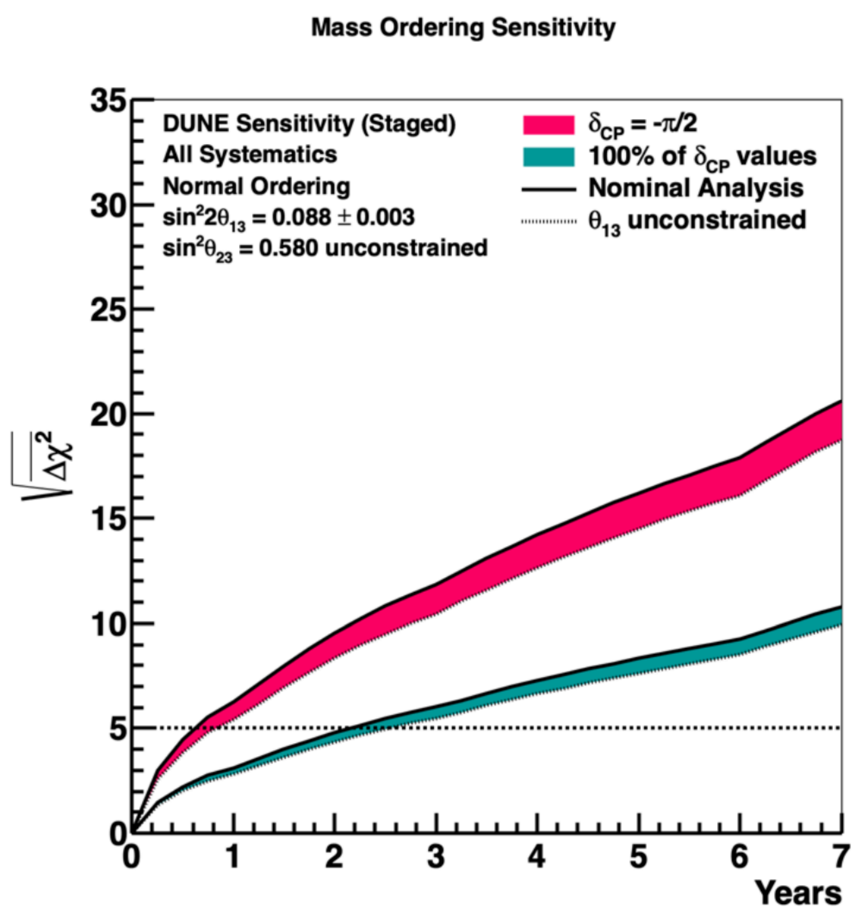
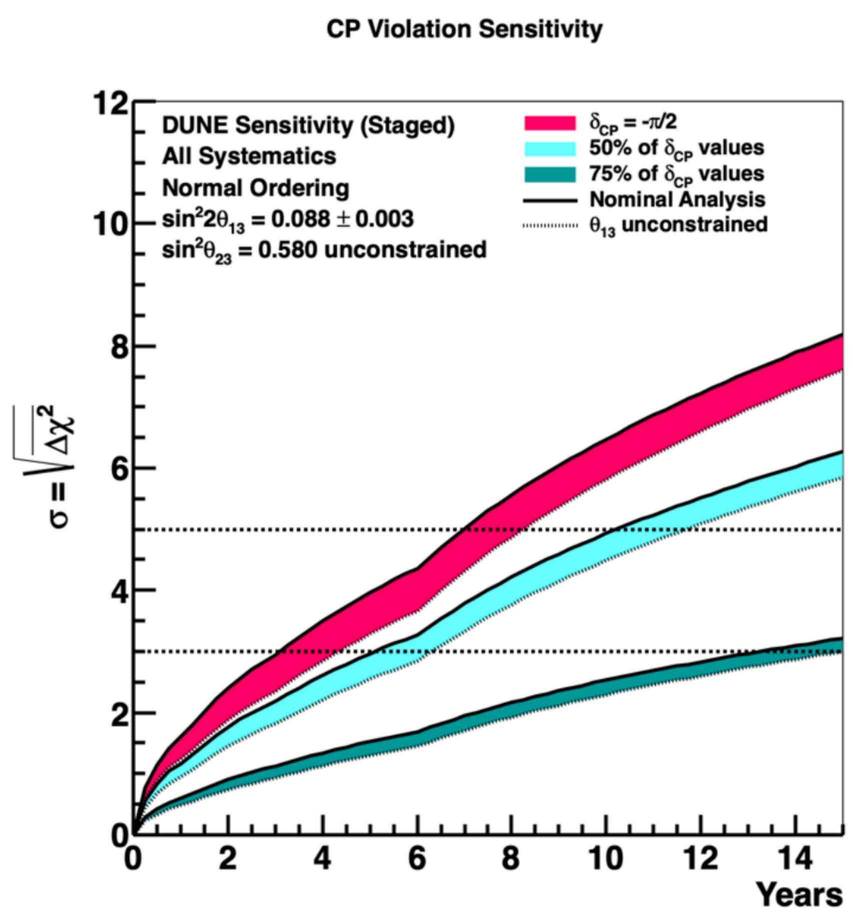


FIG. 4: Simplified diagram of the DUNE CVN architecture.

DUNE sensitivity

arXiv:2002.02967

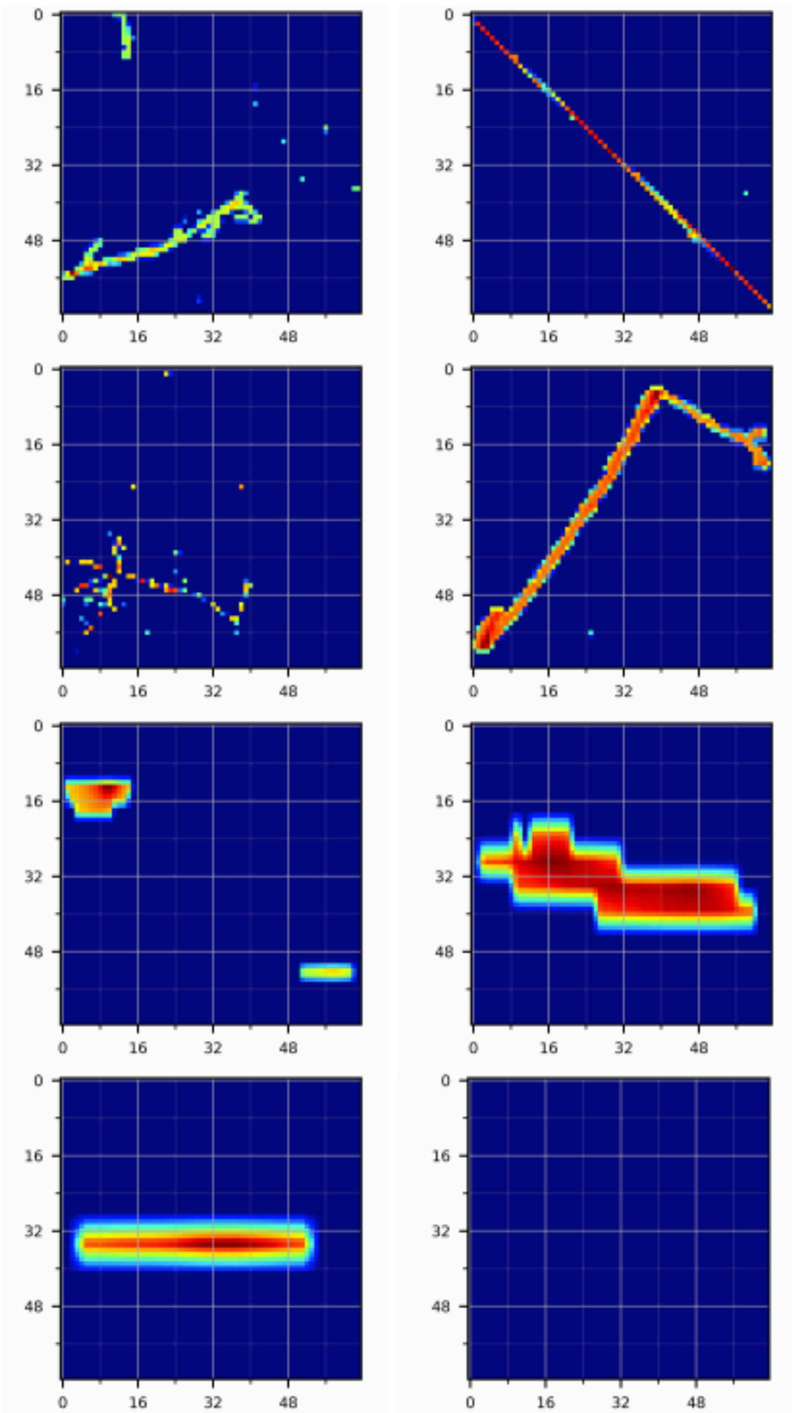




- DUNE's sensitivity to primary physics goals (CP violation, mass ordering) is based on efficiencies and purities achieved by CVN.
    - CVN performance surpasses previous assumptions about event ID capabilities.

University of CINCINNATI

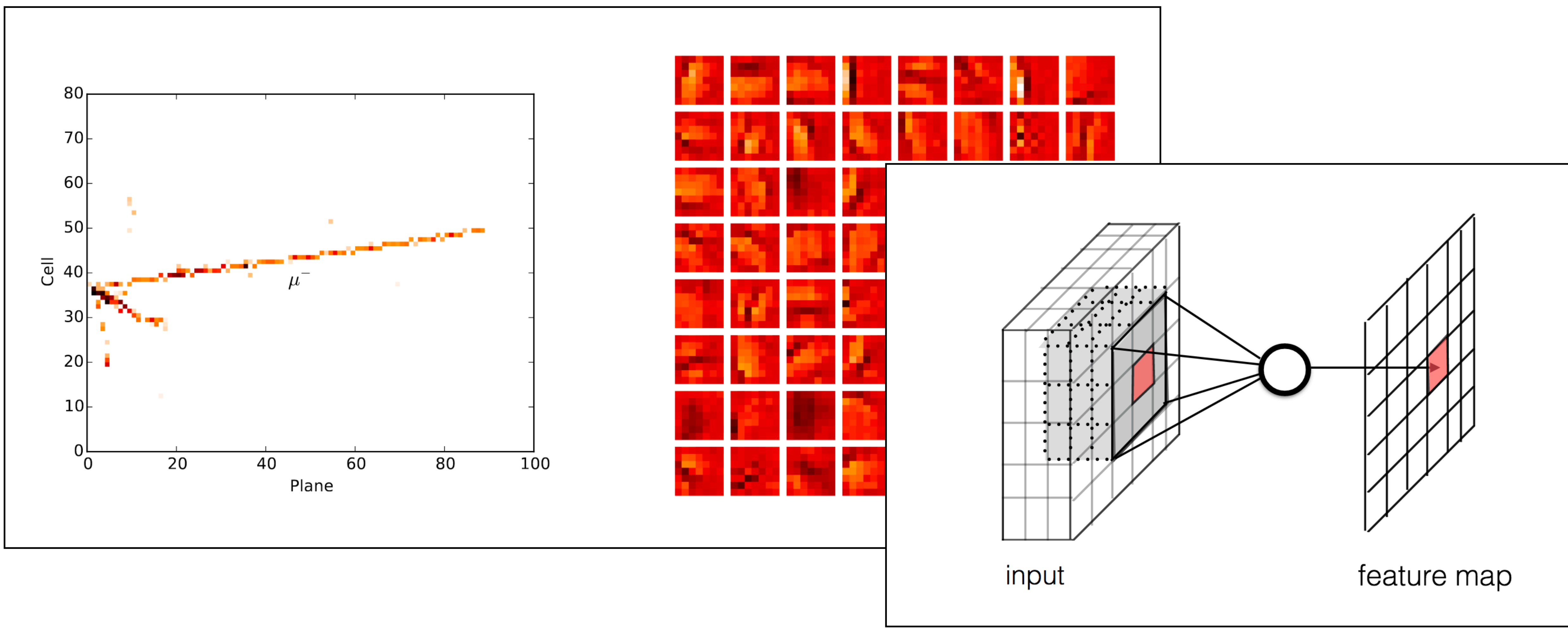
ML triggering at DUNE



- DUNE is a stable large-mass detector in a low background environment.
- Besides neutrino physics, DUNE is well-suited for a range of rare searches:
  - Supernova neutrinos
  - Proton decay
  - $n - \bar{n}$ oscillation
  - …etc
- Studies into using ML for triggering have shown some promise, and FPGA studies show improvements in inference speed (x1.7) and energy efficiency (x2.6) over CPUs (see more here).

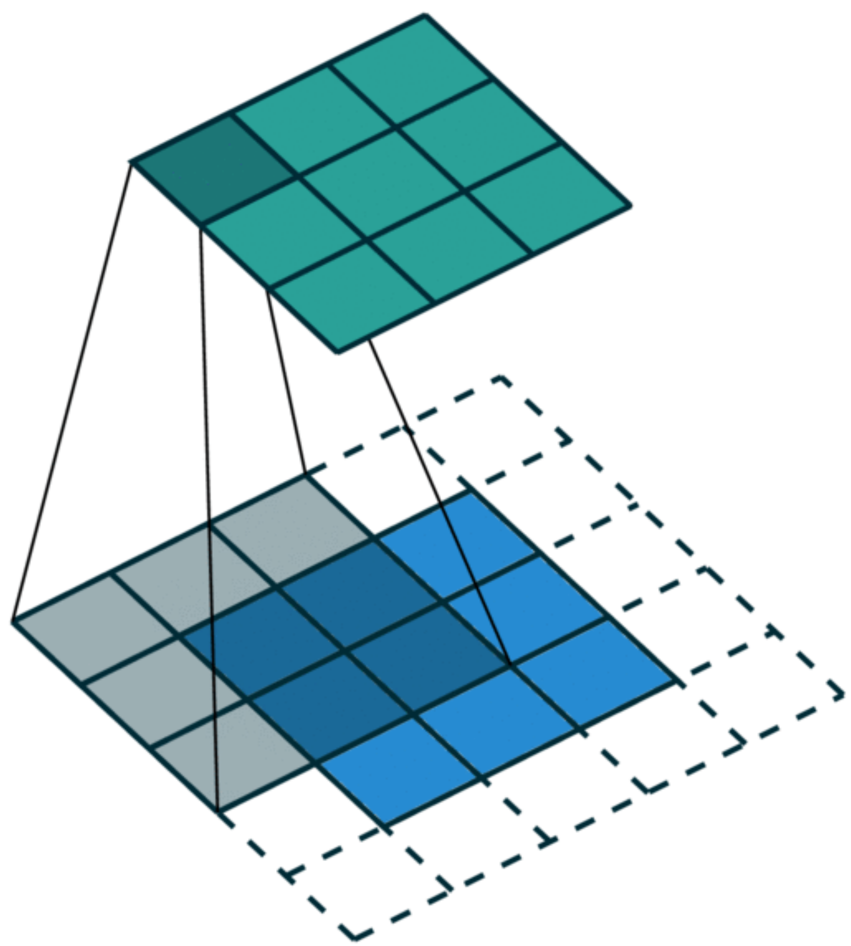
Sparse Convolutions

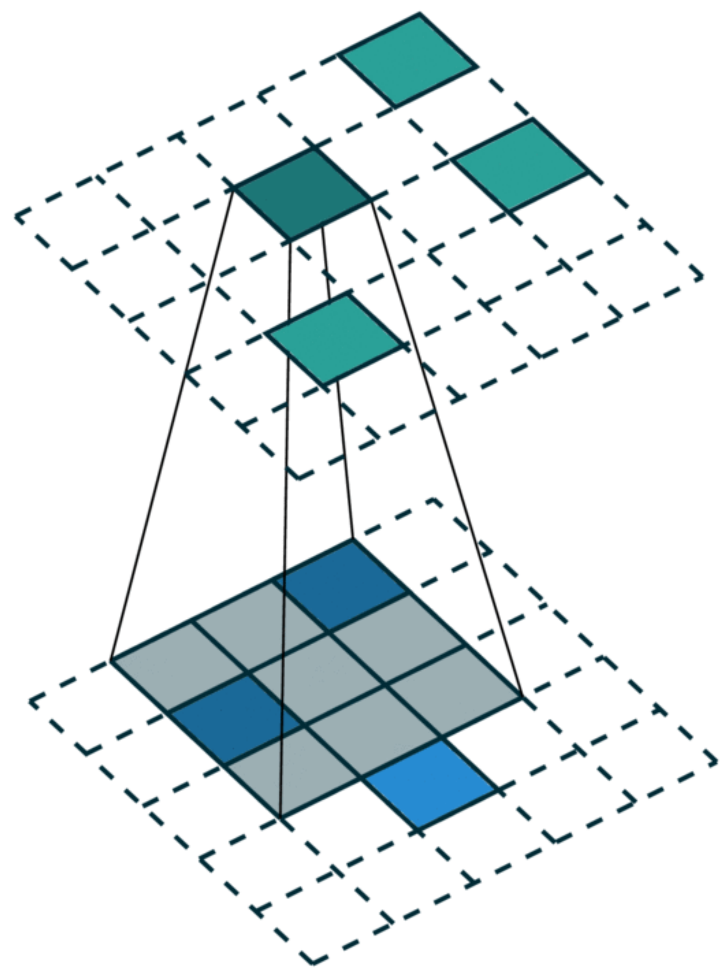
- Many CNN applications in neutrino physics are locally dense but globally sparse.
- Classifying these dense pixel maps involves many computations wasted multiplying your model weights by 0.
- Solution: Use **sparse convolutions** to only operate on interesting regions.



Sparse Convolutions

- Facebook's **SparseConvNet** and later NVIDIA's (formerly Stanford's) **MinkowskiEngine** are PyTorch packages which implement sparse convolutions.
- Only perform convolution if the pixel at the **centre** of the receptive field is non-zero.



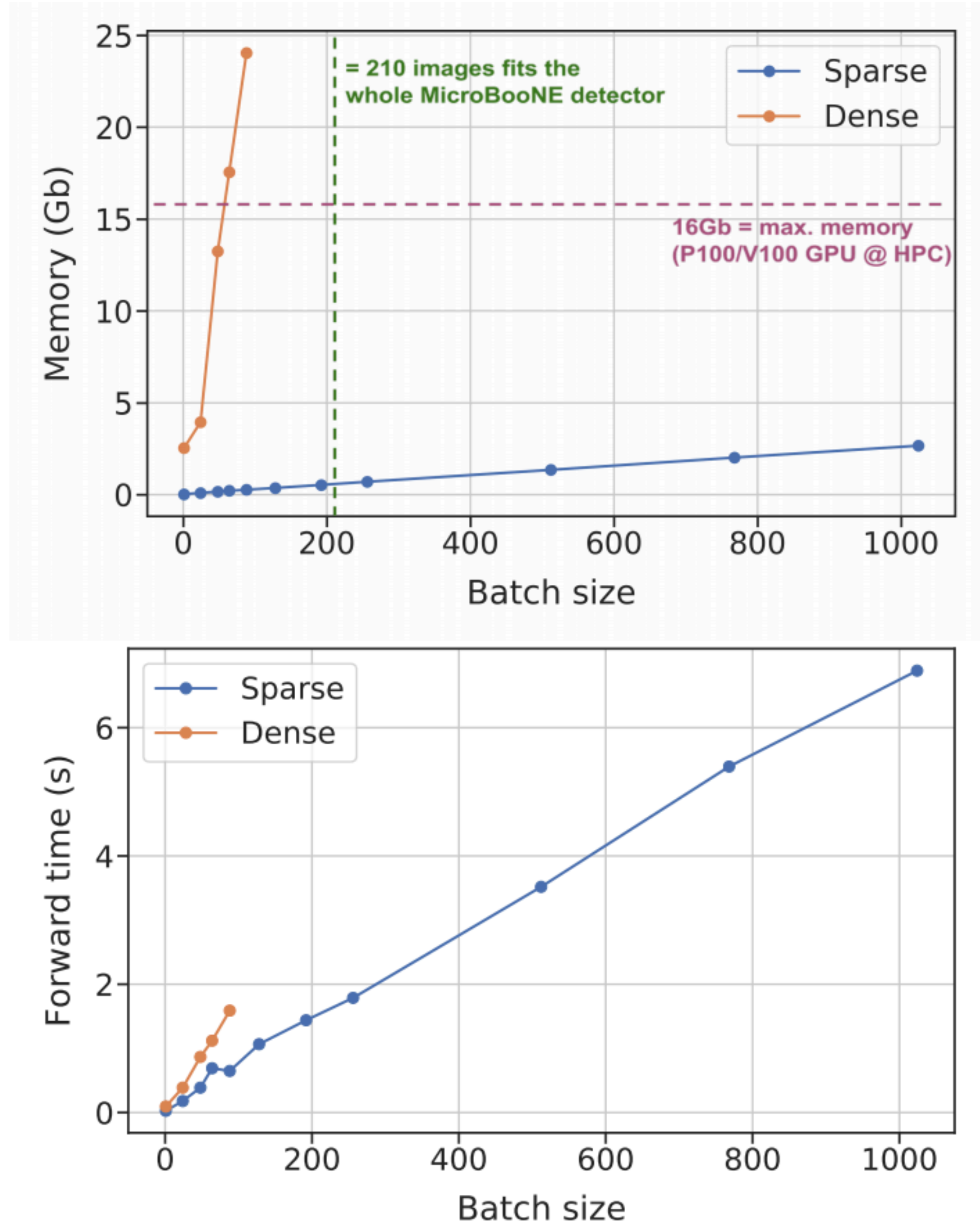


Sparse Convolutions

- First paper investigating sparse CNNs within neutrino physics demonstrated significant improvements in inference time and memory usage for a MicroBooNE-equivalent detector.

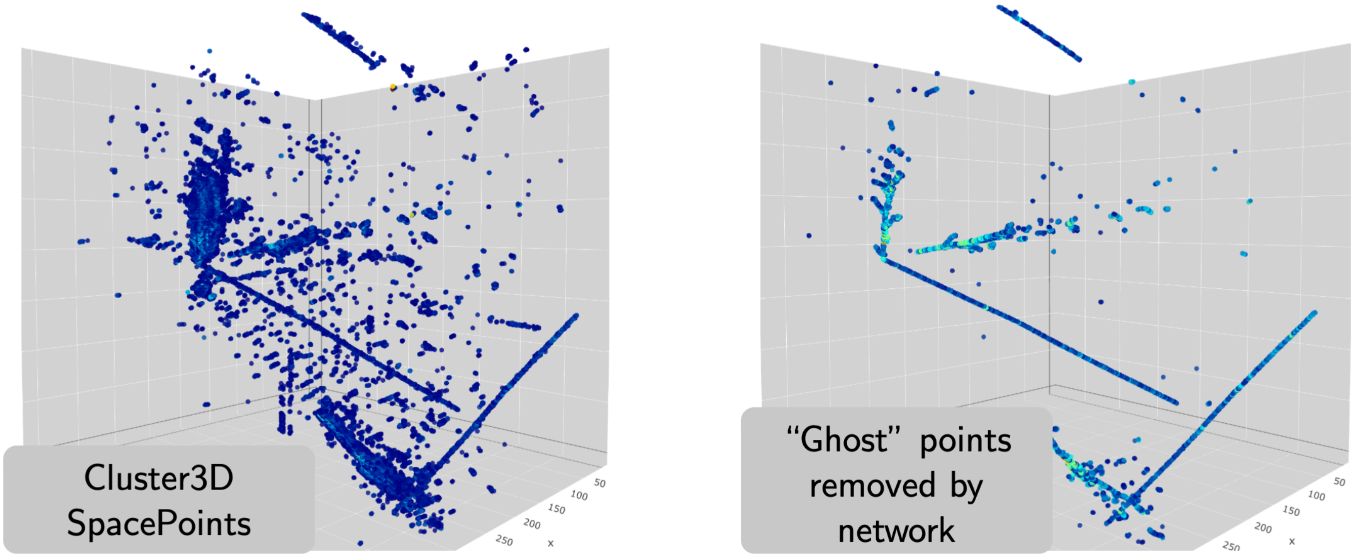
- Sparse convolutions remove the need for ROI-finding in large detectors.
  - Scale of sparse pixel map set by **number of active pixels**.
  - Detector region can be arbitrarily large

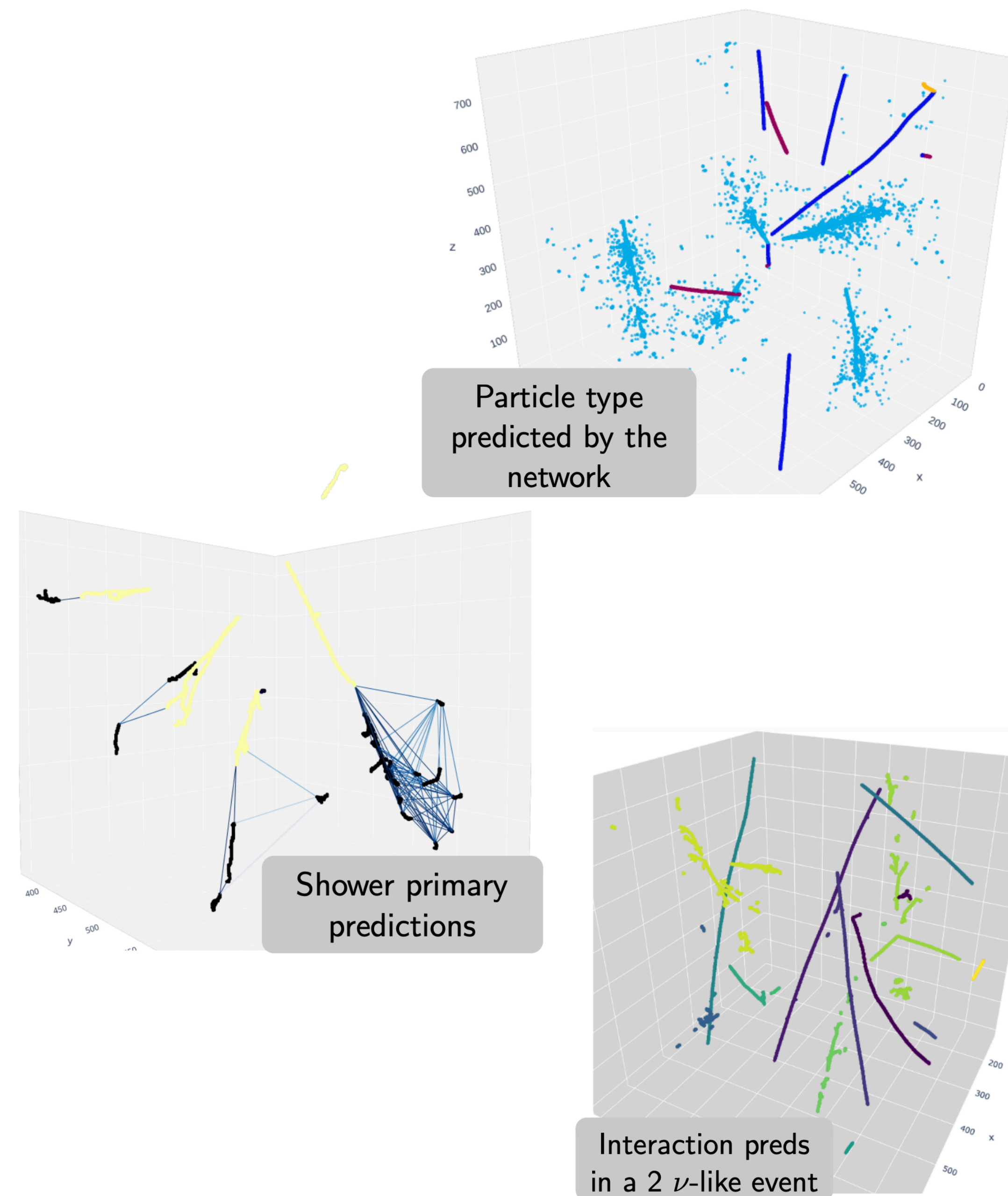
- Sparse CNN approaches are being developed across the SBN and DUNE Near Detector by SLAC, and in NOvA and ProtoDUNE by University of Cincinnati.



arXiv:1903.05663

3D CNN reconstruction





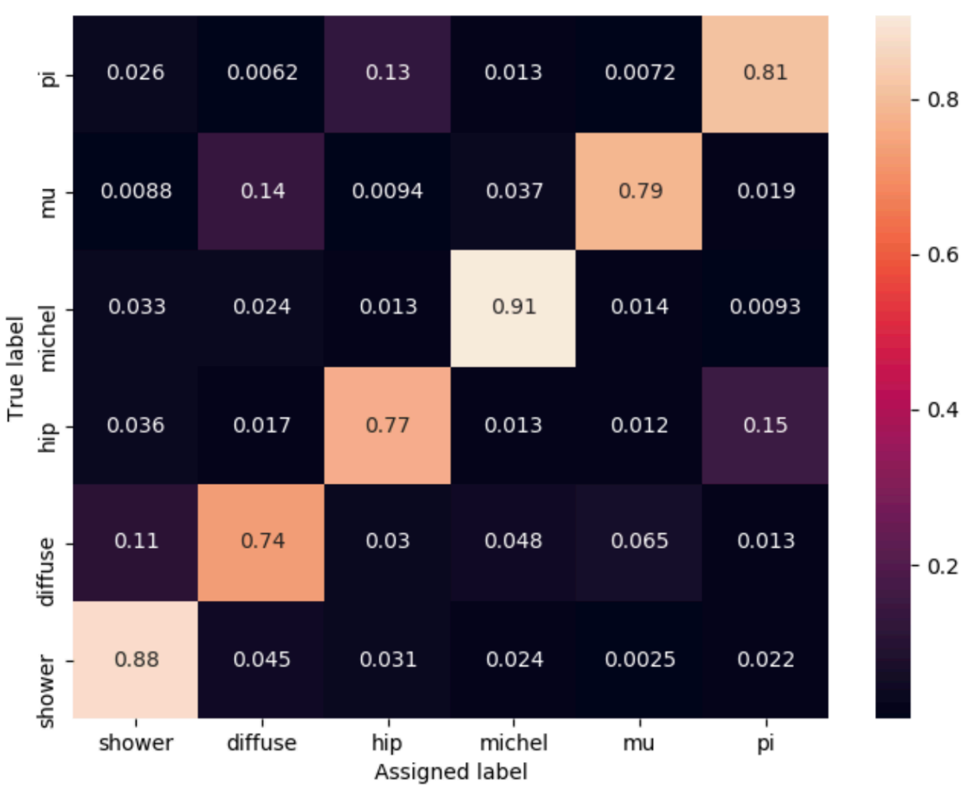
- SLAC LArTPC reconstruction tools utilise 3D sparse CNNs to remove ghost voxels, and classify particle type.
- Instance segmentation to group voxels together into objects.
- Disparate EM shower fragments grouped using a GNN.
- Second GNN layer to group together objects into interactions.

Connecting the Dots 2020
F. Drielsma

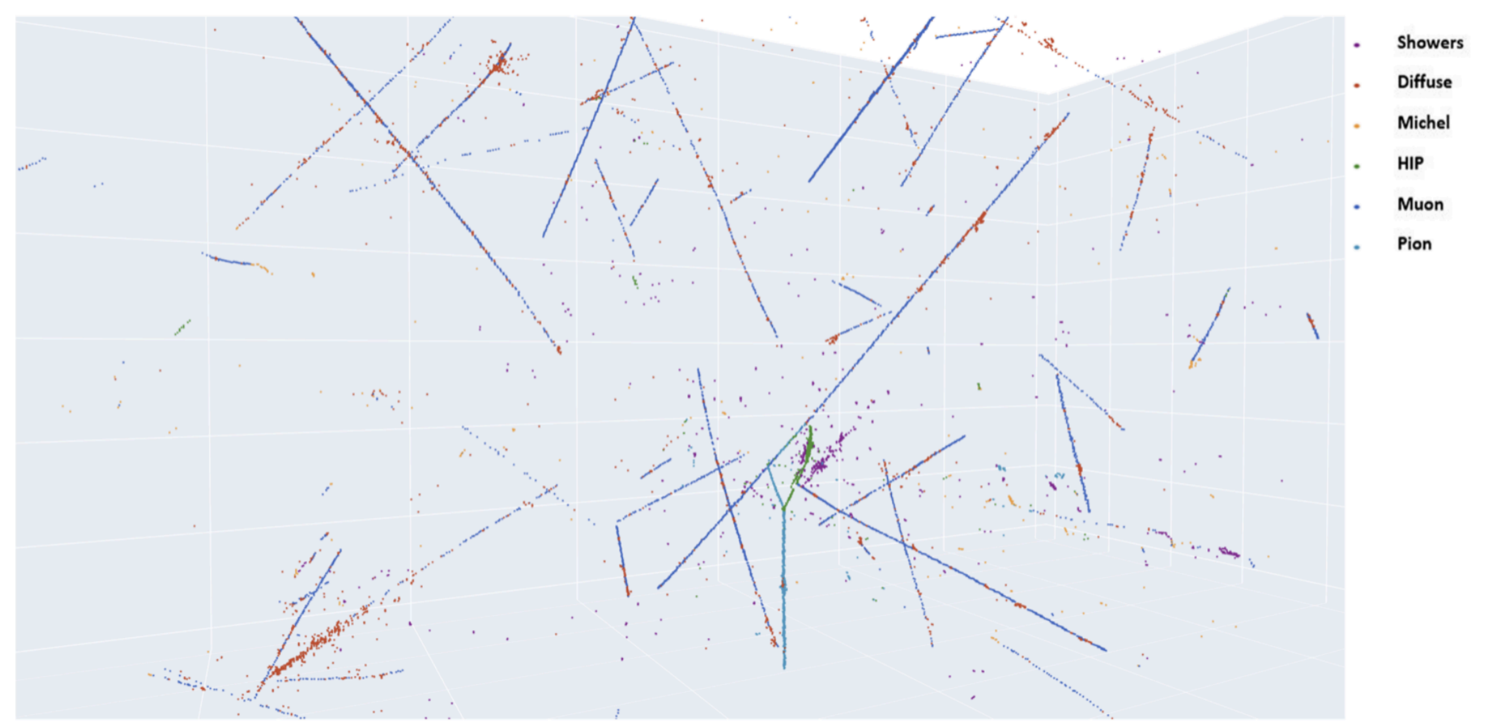
ProtoDUNE Sparse CNN

- Sparse 3D CNN for voxel segmentation also developed for ProtoDUNE.
    - Test beam prototype for DUNE far detector at CERN.

Neutrino 2020
C. Sarasty

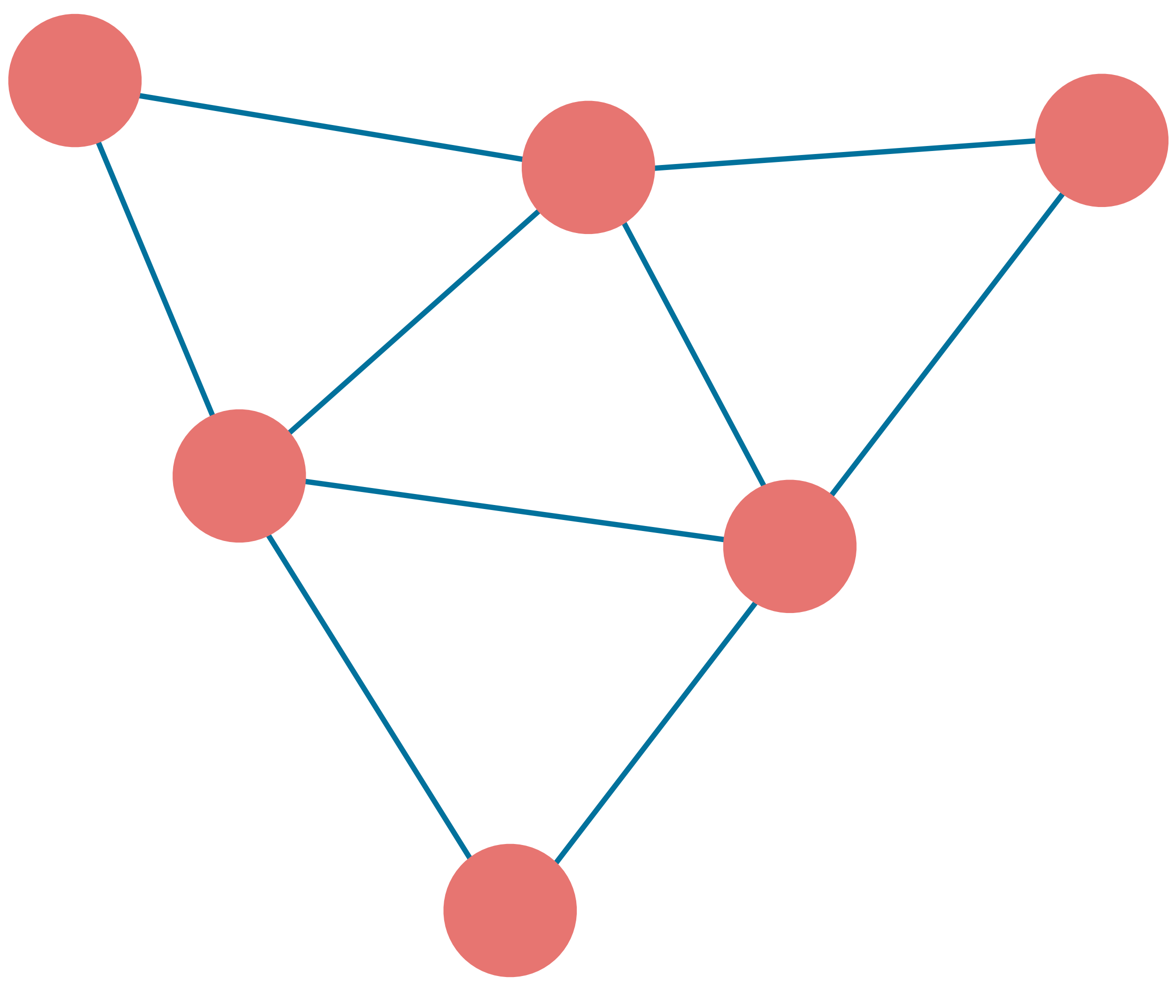


ProtoDUNE-SP Simulation



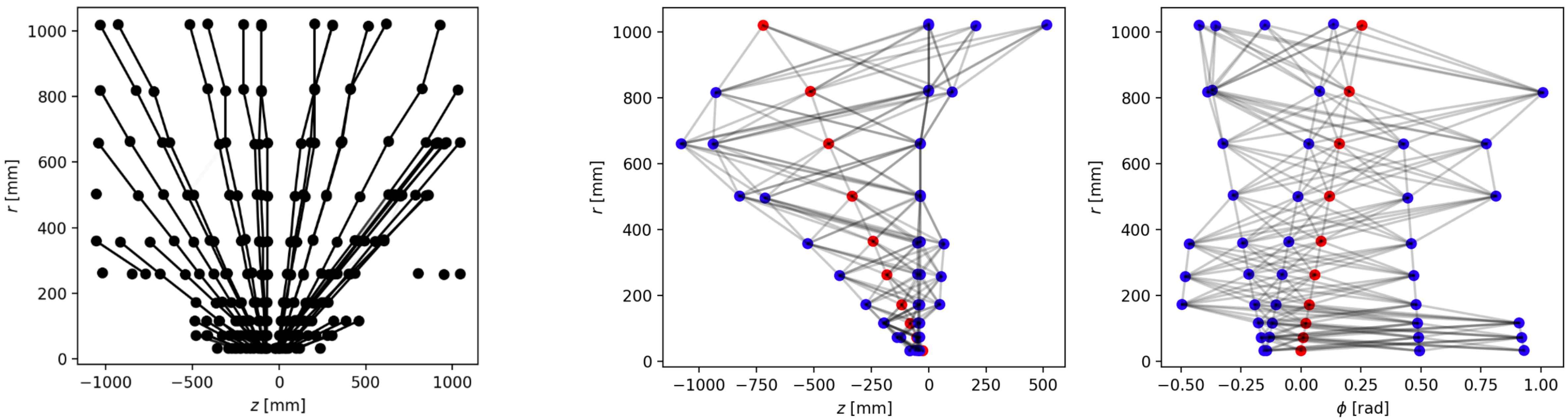
Graph neural networks

- Describe information structure as a **graph** represented by **nodes** and **edges**.



- **Nodes** are generalised as quantised objects with some arbitrary set of **features**.
- **Edges** describe the **relationships** between nodes.
- Perform convolutions on nodes and edges to learn relationships within the graph.
- Output is user-defined:
    - Classify nodes or edges.
    - Classify full graph.
    - Regression outputs.

Exa.TrkX

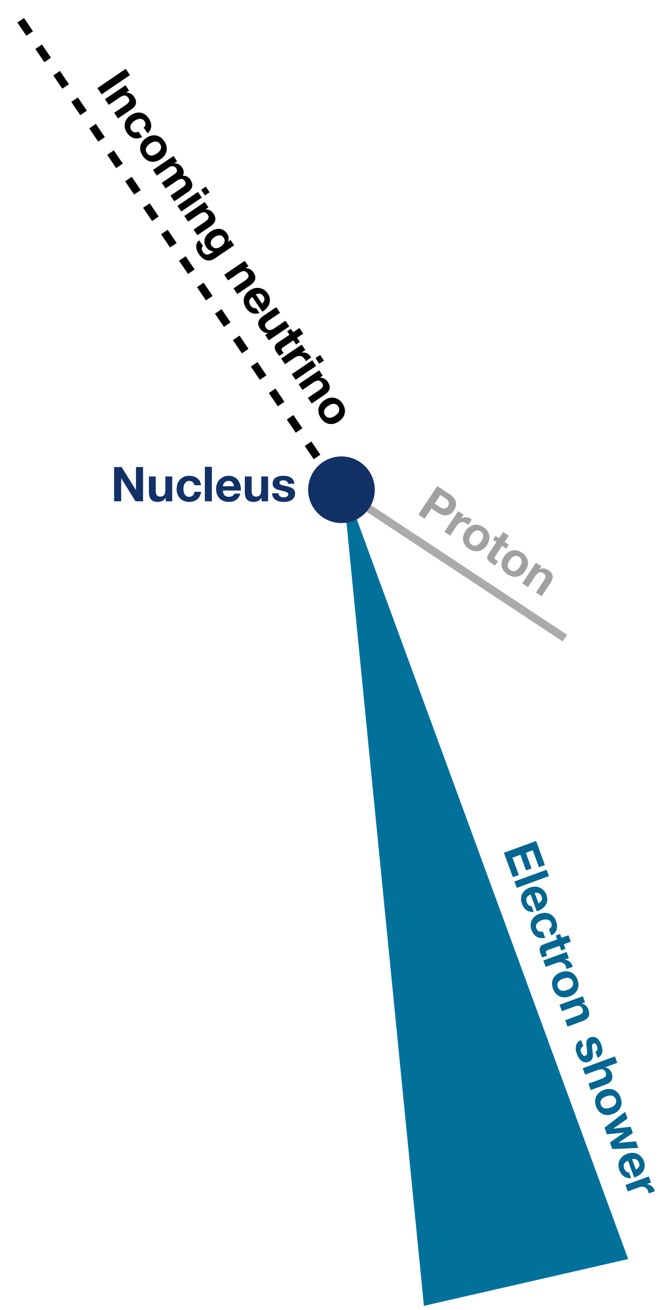


arXiv:1810.06111

- Promising results from the **Exa.TrkX** collaboration using GNN methods for track reconstruction in the HL-LHC.
- Adapting these methods for 2D reconstruction in the DUNE far detector.
  - CNN-based methods perform well, but require transformation of 3D point cloud into a grid.
  - Graph-based techniques can operate on the data in its native structure.

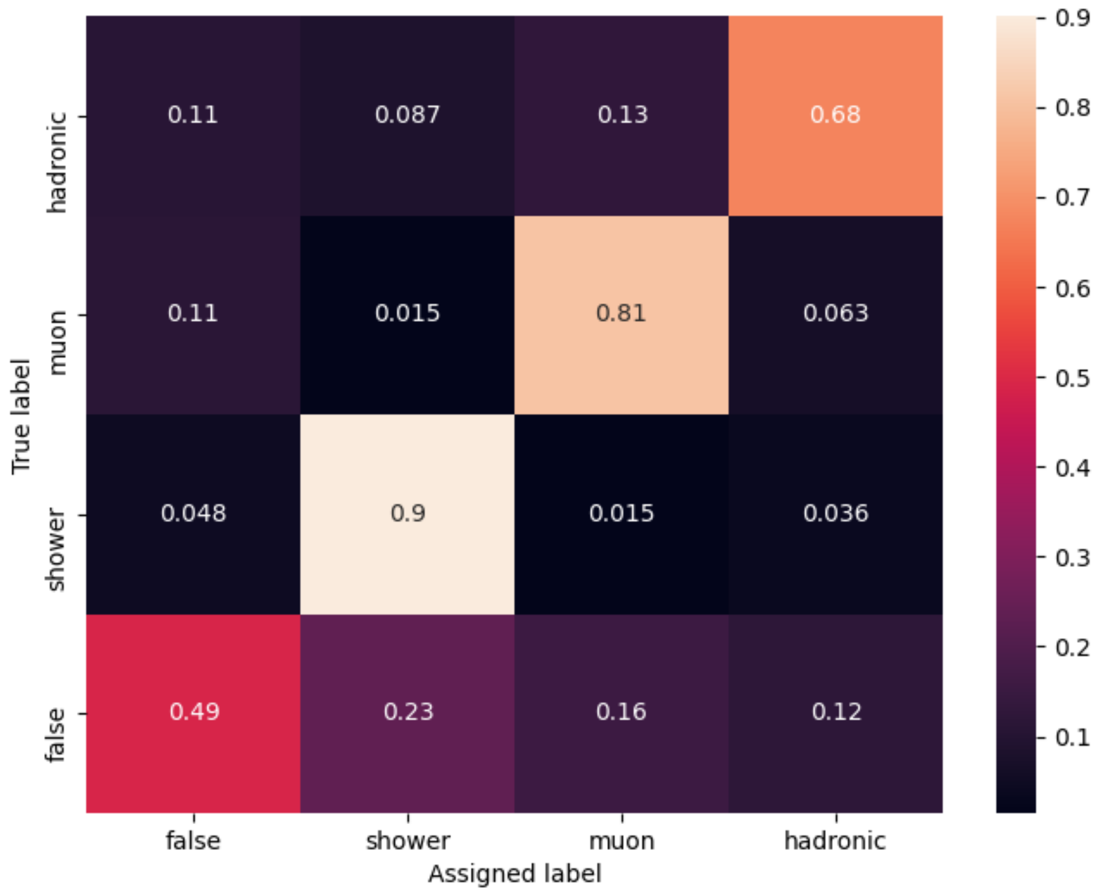
GNNs in DUNE

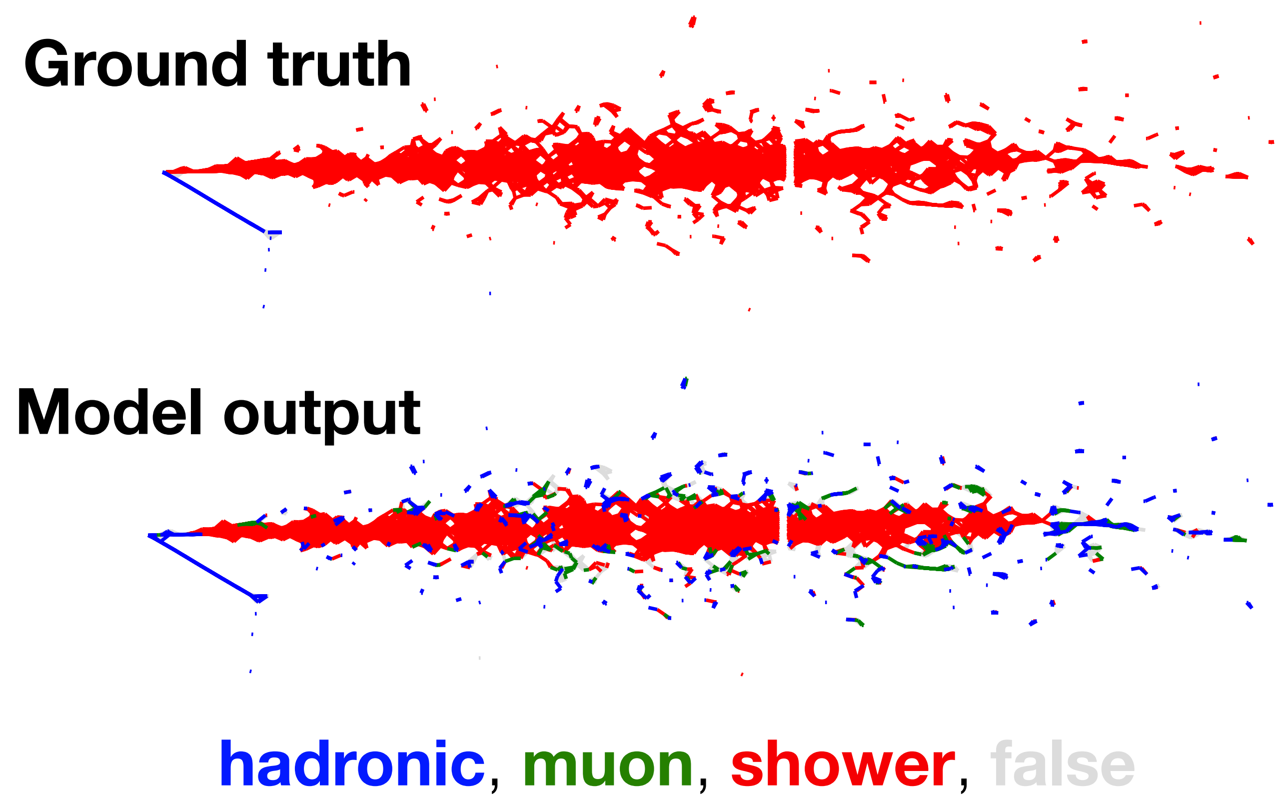
- Developing GNNs for particle reconstruction in 2D using simulated CCQE beam neutrino interactions.
    - Few-GeV energy.
    - Neutrinos travel along beam direction.
    - Typically “clean” interactions – primary lepton (e,μ) and minimal hadronic activity.
- Train multihead attention message-passing network to classify relationships between detector hits.
    - Determine whether the hits were created by the same particle – and if so, whether that particle is an EM shower, μ or hadronic.



GNNs in DUNE

- Current iteration achieves 84% accuracy in classifying graph edges.
- Exploring further graph-based approaches.





Summary

- Accelerator neutrino physics seeing **increasingly widespread ML use**.
  - NOvA, SBN, DUNE, Minerva (not mentioned here!)
- At this point, ML applications in accelerator neutrino physics are **mostly offline**.
  - Reconstruction tools after data has already been collected.
  - Some lower-level ML applications for raw data/triggering are being explored.
- Personal opinion: even in places where techniques are mature, integration between ML and analysis tools could be improved.
  - CPU inference.
  - Training environments separate from sim/reco/analysis ecosystem.
  - Process streamlined in NOvA with HDF5 integration.
  - Active work in FermiGrid, HPC to improve this!

Backup

Neutrino oscillations

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

PMNS matrix

$$P(\nu_\alpha \to \nu_\beta) = \sum_i U^*_{\alpha i} U_{\beta i} U_{\alpha i} U^*_{\beta i} + \sum_{i \neq j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{i \frac{\Delta m^2_{ji} L}{2E}}$$

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} 1 & & \\ & c_{23} & s_{23} \\ & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & & s_{13} e^{i\delta} \\ & 1 & \\ -s_{13} e^{i\delta} & & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & \\ -s_{12} & c_{12} & \\ & & 1 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

$s_{ij} = \sin\theta_{ij}, \; c_{ij} = \cos\theta_{ij}$

University of CINCINNATI

Neutrino oscillations

$$P(\nu_\mu \rightarrow \nu_e) \approx \boxed{\sin^2\theta_{23}} \sin^2 2\theta_{13} \frac{\sin^2 \Delta(1-A)}{(1-A)^2}$$
$$+\alpha \tilde{J} \cos(\Delta \boxed{\pm \delta_{cp}}) \frac{\sin \Delta A}{A} \frac{\sin \Delta(1-A)}{(1-A)}$$
$$+\alpha^2 \boxed{\cos^2\theta_{23}} \sin^2 2\theta_{12} \frac{\sin^2 \Delta A}{A^2}$$

$$\tilde{J} = \cos\theta_{13} \sin 2\theta_{13} \sin 2\theta_{12} \sin 2\theta_{23}$$
$$A = \pm\sqrt{2} G_F n_e E_\nu / \Delta m^2_{13}$$
$$\Delta = \boxed{\Delta m^2_{31}} L_\nu / 4E_\nu$$
$$\alpha = \Delta m^2_{21} / \Delta m^2_{31}$$

- Several parameters in our current neutrino oscillation framework remain unmeasured:
  - Neutrino mass ordering (sign of Δm^2_{32}).
  - CP-violating phase (δ_{cp}).
  - Octant of θ_{23}.