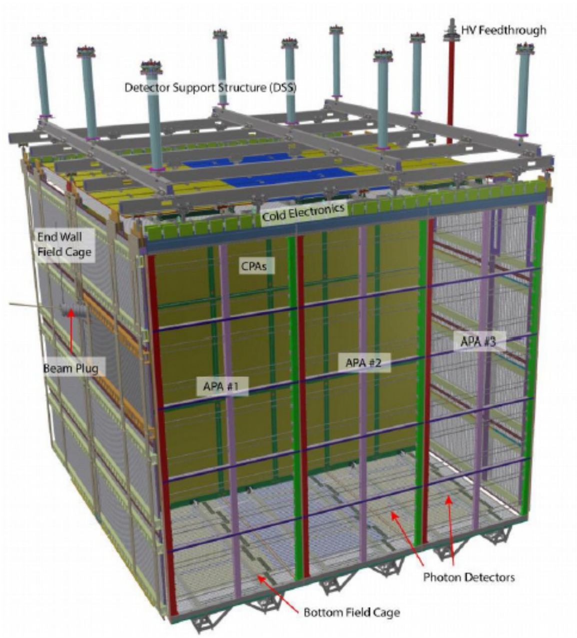
Machine Learning and BSM Searches



Adam Aurisano
University of Cincinnati

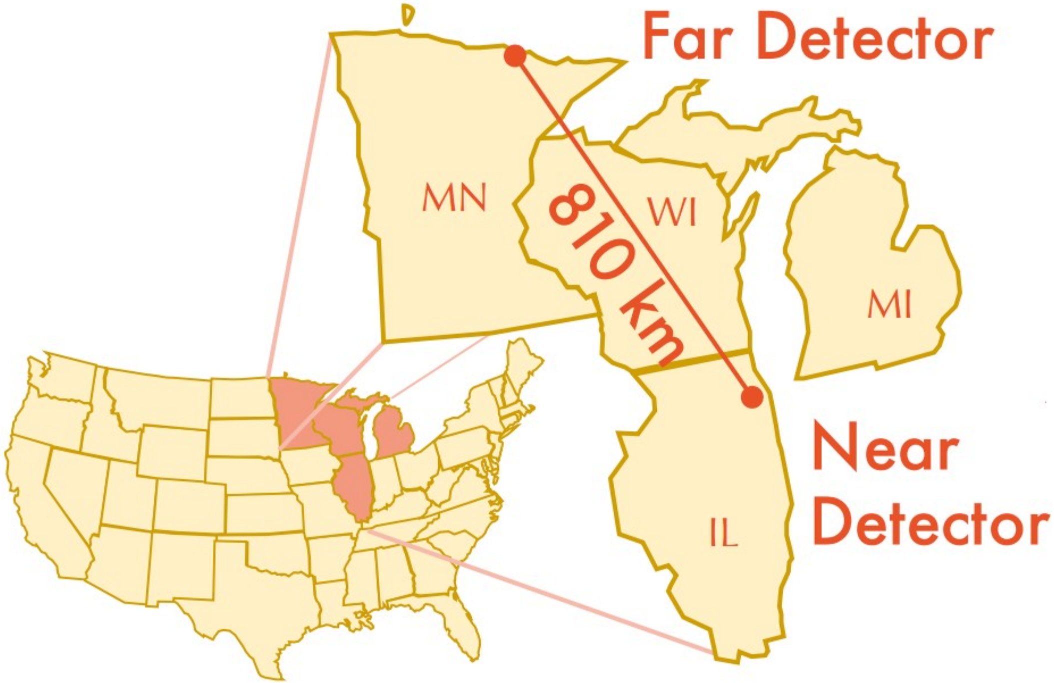
New Opportunities at the Next Generation Neutrino Experiments

Overview

- Machine learning has a long history in HEP experiments
  - Neural networks, BDT, kNN, SVM
- Deep-learning revolution in 2012 has reinvigorated the machine learning field
- Virtually all HEP experiments have realized that new deep-learning techniques have broad applicability to our problems
- I will mostly be talking about NOvA and DUNE, but the techniques I will discuss are increasingly in common usage
- On the cusp of full end-to-end deep-learning-based reconstruction which may open new avenues for BSM searches

NuMI Off-Axis ν_e Appearance Experiment

NOvA is a long-baseline neutrino oscillation experiment located 14 mrad off-axis from the NuMI beam designed to measure:



ν_e appearance

- Mass hierarchy
- CP violation
- θ_{23} octant

ν_μ disappearance

- Improved precision on $|\Delta m^2_{32}|$ and θ_{23}

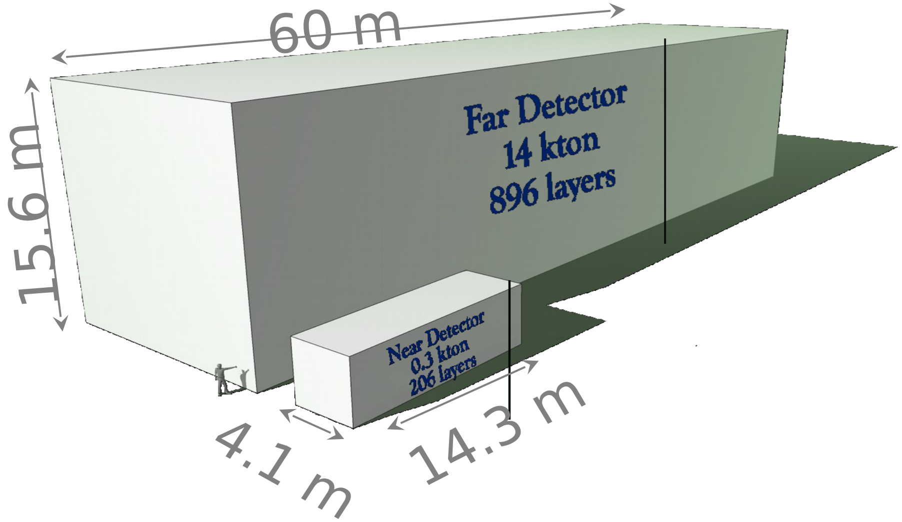
NC disappearance

- Search for sterile neutrinos
- Constrain θ_{34} and θ_{24}

Others

- Short-baseline steriles
- Cross sections
- Supernovae
- Exotics

NOvA Detector Design

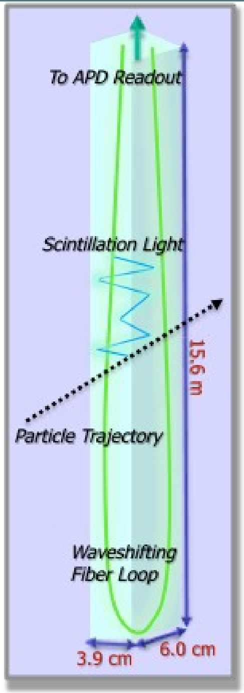


Far detector (FD)

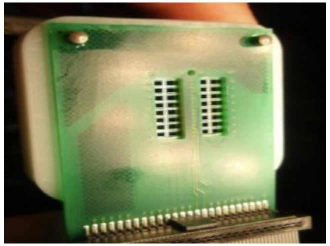
- 14 kton
- 65% active mass
- ~344,000 channels

Near detector (ND)

- 0.3 kton
- Functionally equivalent to FD for systematic uncertainty reduction
  - Faster electronics
  - Muon catcher to contain muons
- ~20,000 channels



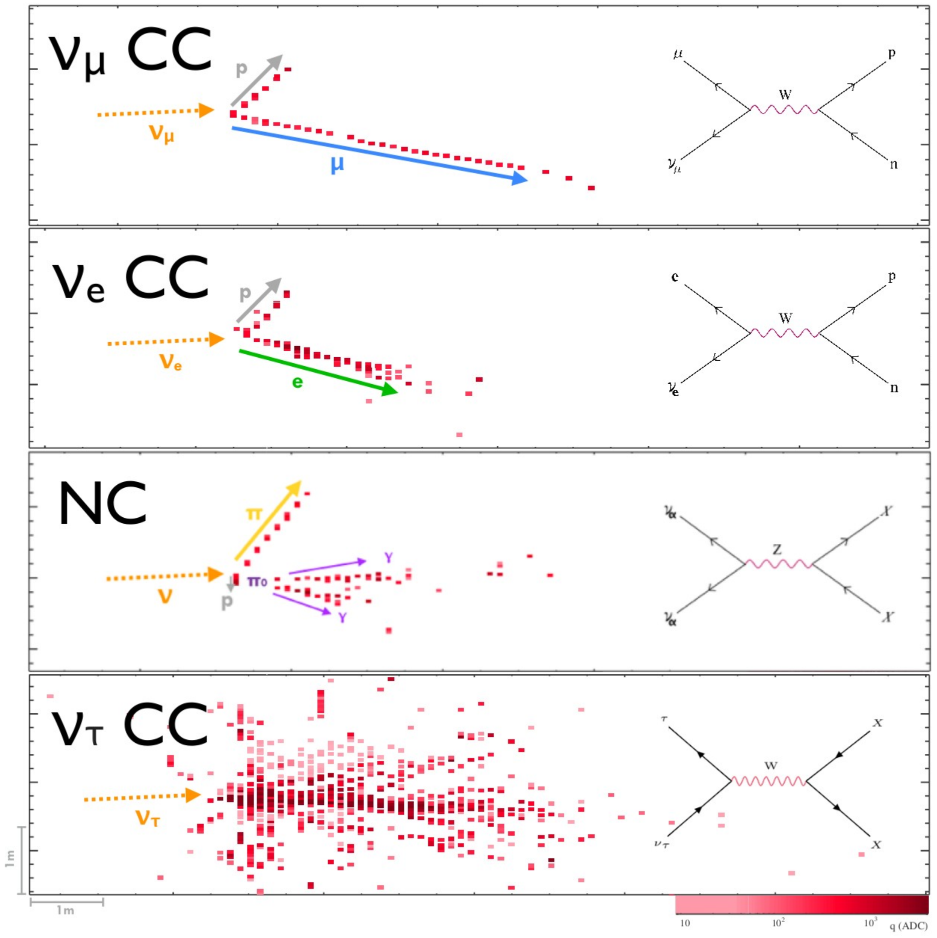
Low Z tracking calorimeter composed of alternating horizontal and vertical planes of liquid scintillator filled cells.



Wavelength shifting fibers carry light out of the cells to APDs.

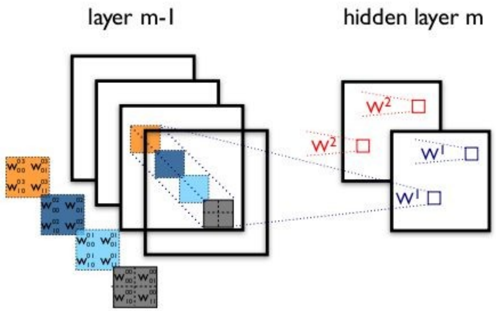
Event Topologies

- Low Z detector materials lead to long tracks and well developed showers
- Key challenges
  - Discriminating between muons and charged pions
    - Both can produce long tracks, but muons are usually longer and interact less with nuclei
  - Discriminating between electrons and photons
    - Electrons start showering immediately, but photons travel a short distance before showering
      - Crucial since neutral pions decay into photons
- In the standard workflow, reconstruct particle content of the event and try to classify the neutrino flavor and interaction type
- Events in NOvA look like images
  - Try using convolutional neural networks for classifying

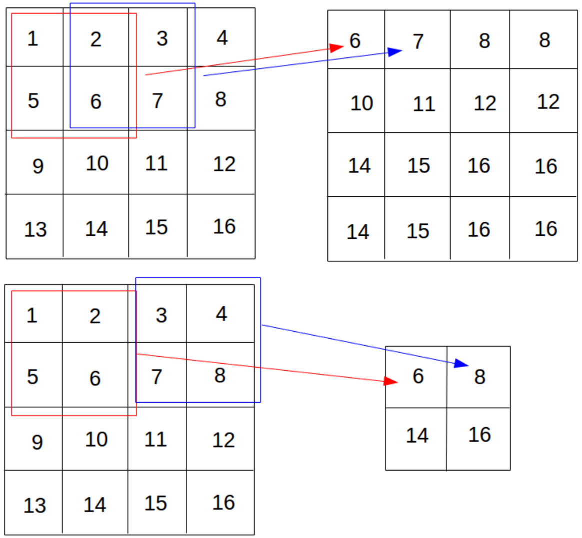


Convolutional Neural Networks

- X and Y views can be interpreted as pictures of a neutrino interaction from the top and the side
- Convolutional neural networks have been highly successful at image recognition tasks
- Two basic type of layers:
  - Convolutional layers – apply discrete convolutions using learned kernels to extract features from the image.
  - Pooling layers – down sample the image and increase translational invariance in the final output.
- Stacked structure of convolutional and pooling layers extract increasingly abstract features from the input raw data encoding both local and global structure.



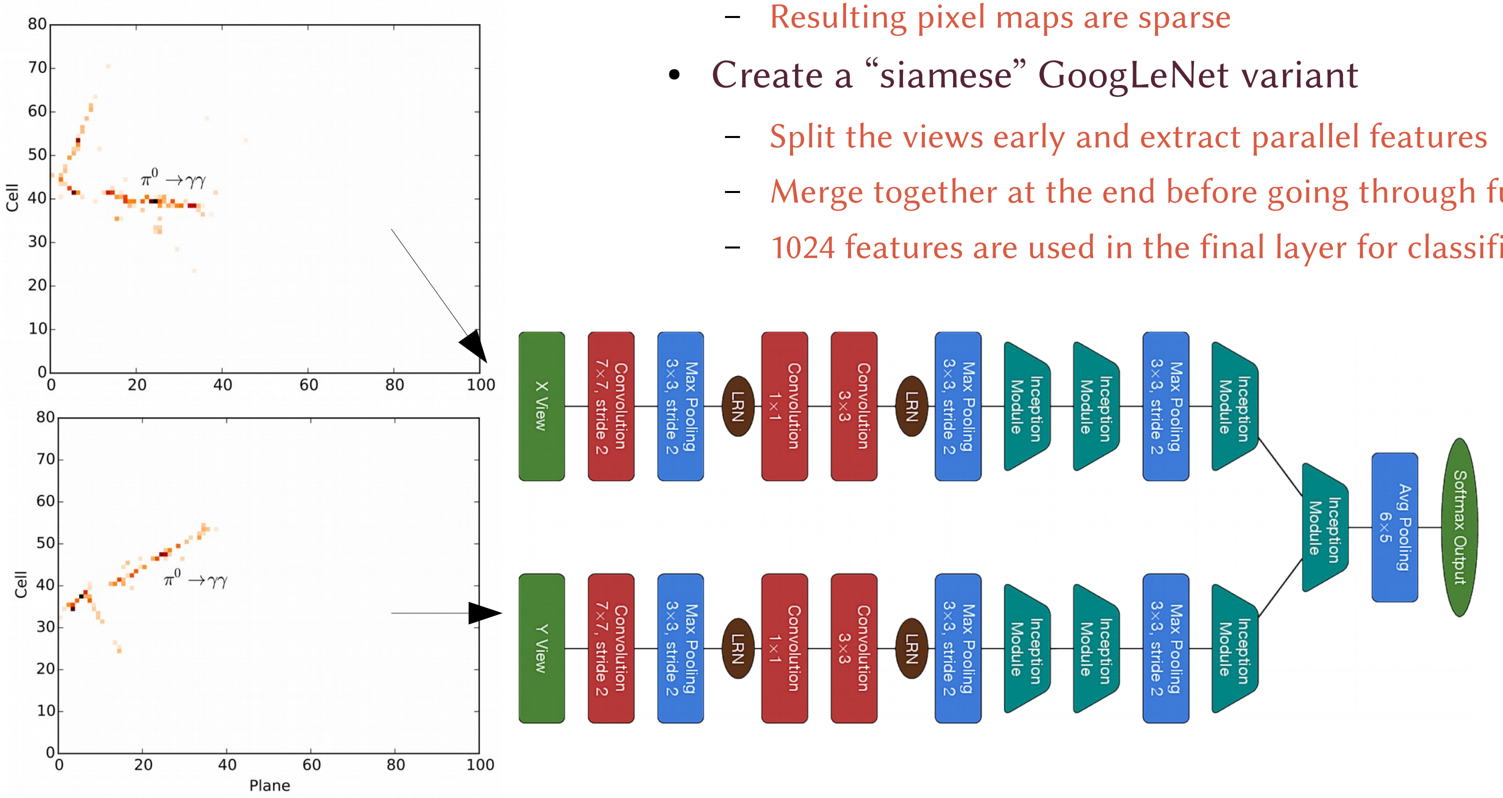
2x2 MaxPool Stride 1



2x2 MaxPool Stride 2

Convolutional Visual Network

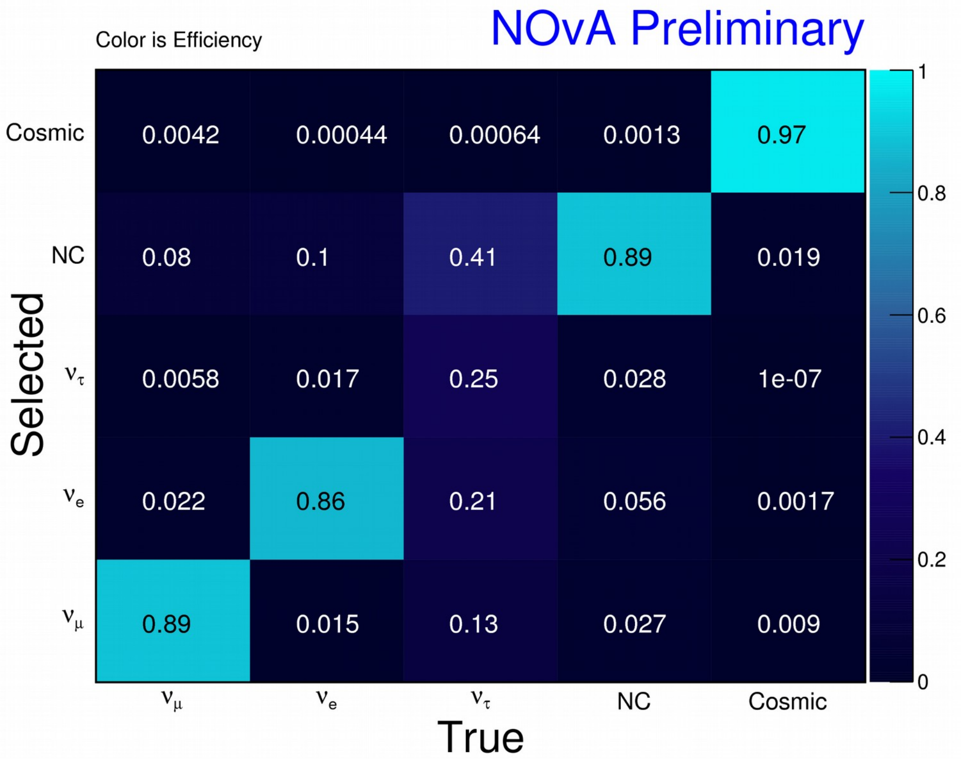
- Detector is composed of alternating horizontal and vertical planes
  - Two views of the event- one from top and one from side
  - Resulting pixel maps are sparse
- Create a “siamese” GoogLeNet variant
  - Split the views early and extract parallel features
  - Merge together at the end before going through fully connected layers
  - 1024 features are used in the final layer for classification.



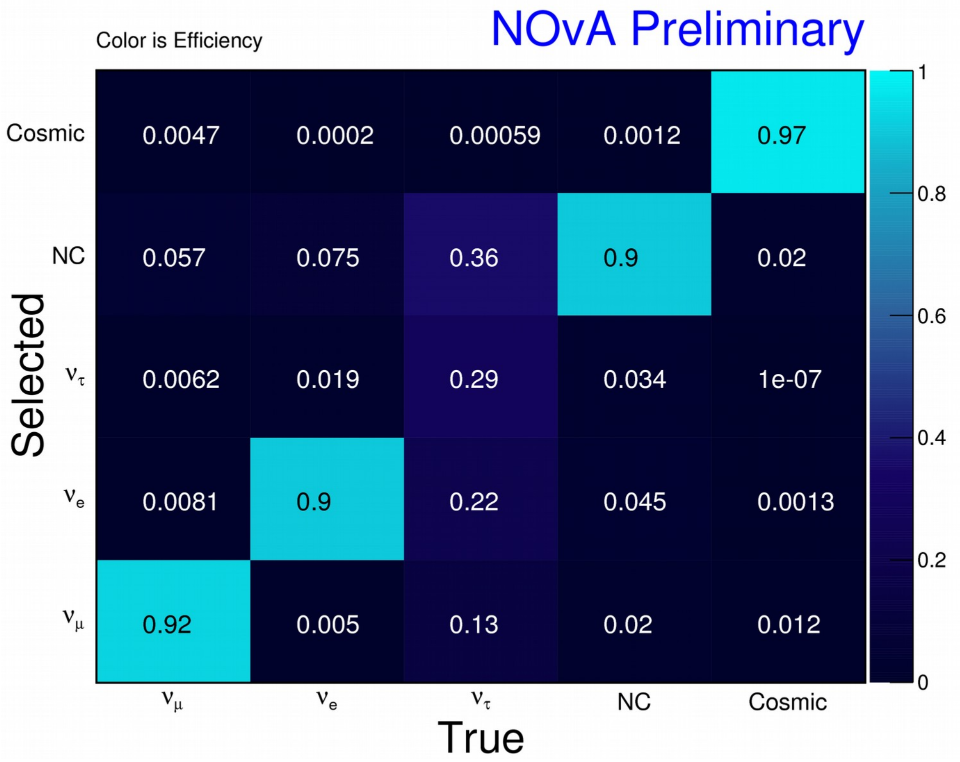
- The architecture is a multi-classifier
  - Neutrino flavor
  - Type of interaction with the nucleus
- In principle, this architecture is a universal neutrino classifier

Training

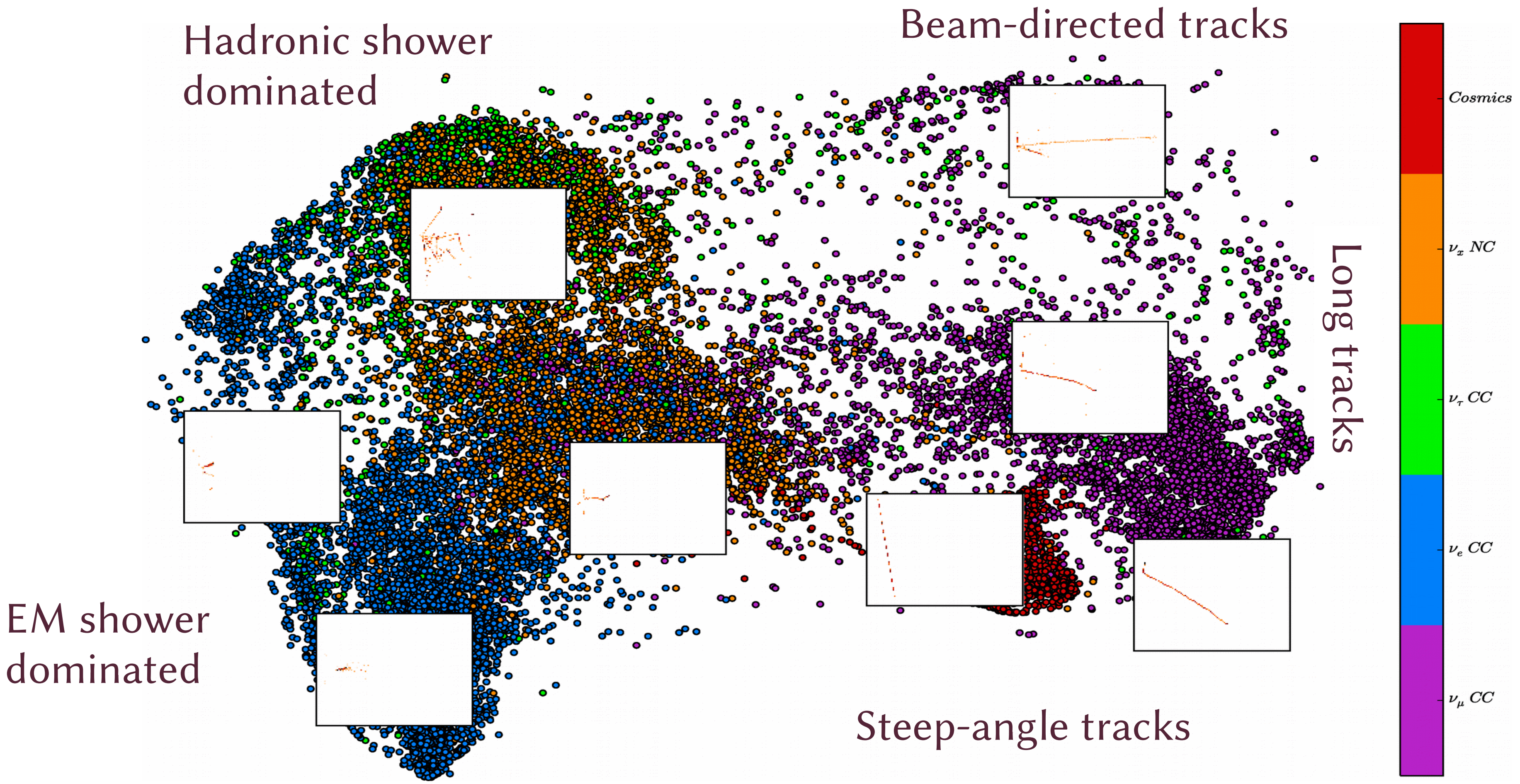
Neutrino mode



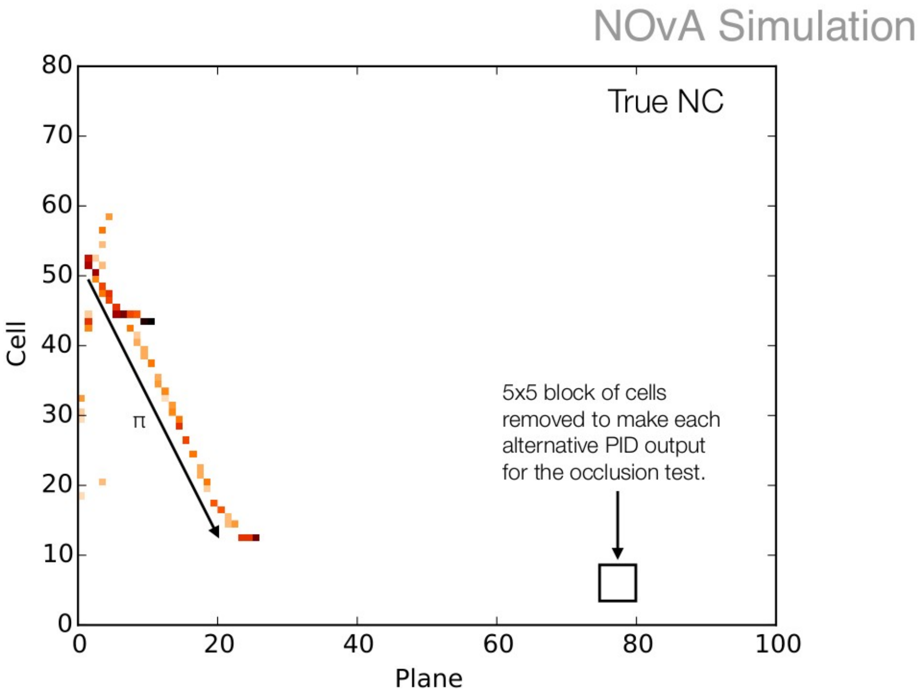
Antineutrino mode

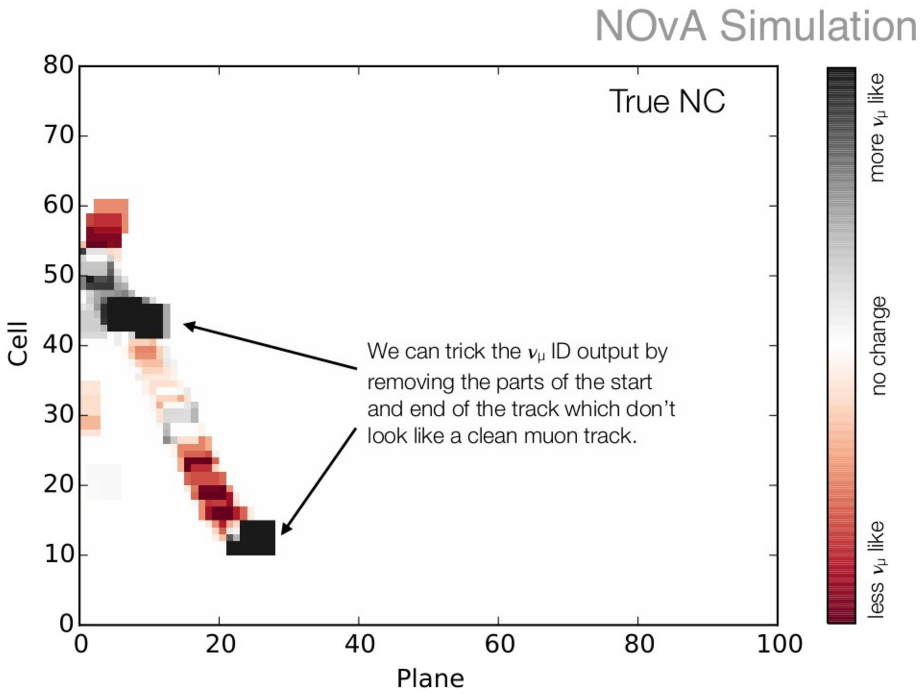


Understanding the Network: Feature Embedding with t-SNE

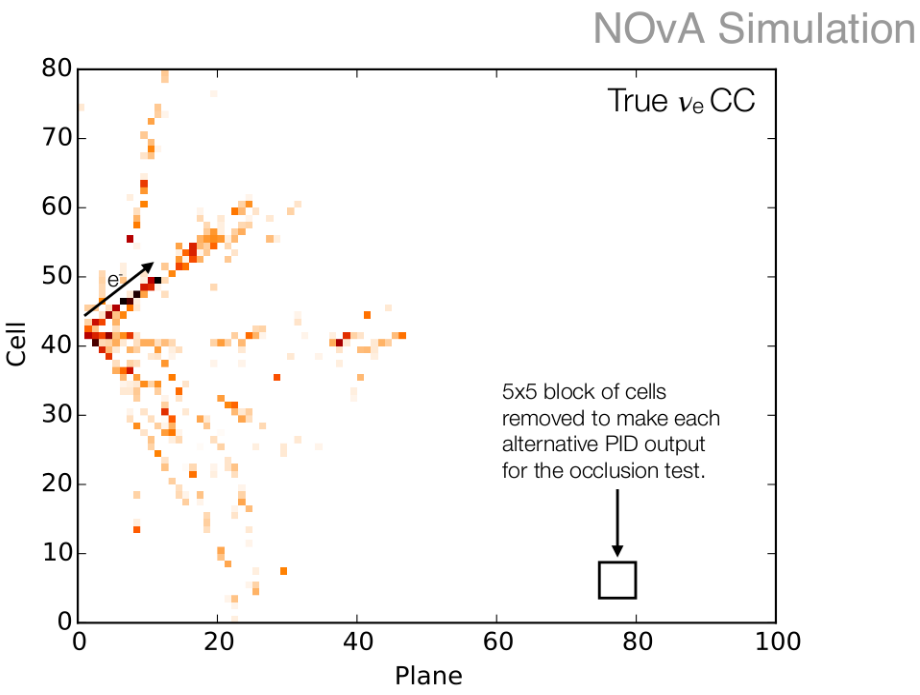


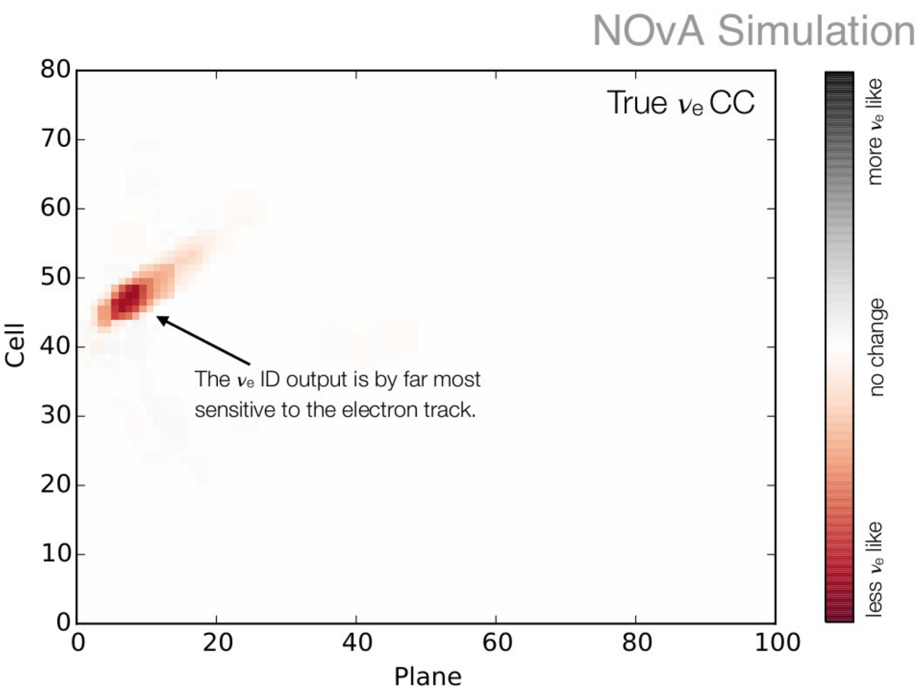
Understanding the Network: Occlusion Tests





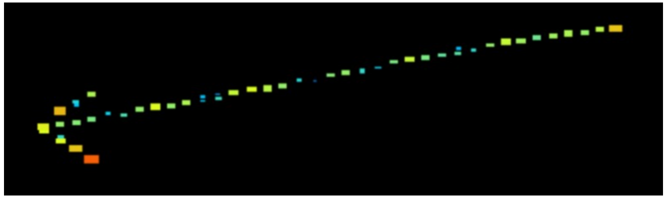
Understanding the Network: Occlusion Tests





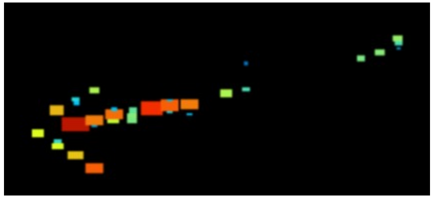
Hybrid Event Testing

- Select likely ν_μ events in the Near Detector

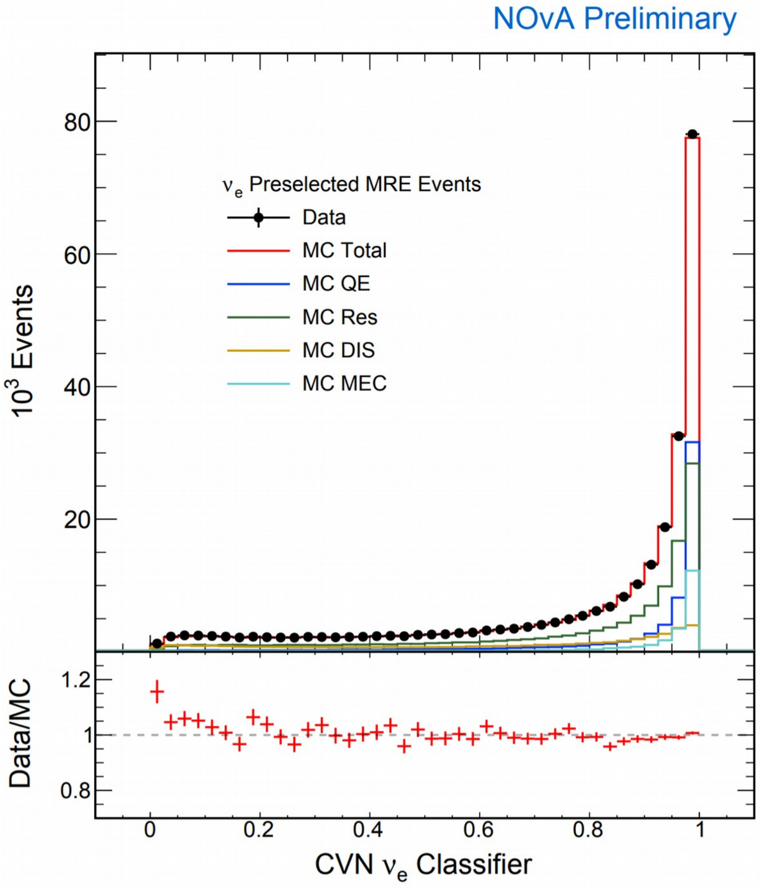


- Remove hits from the identified muon track

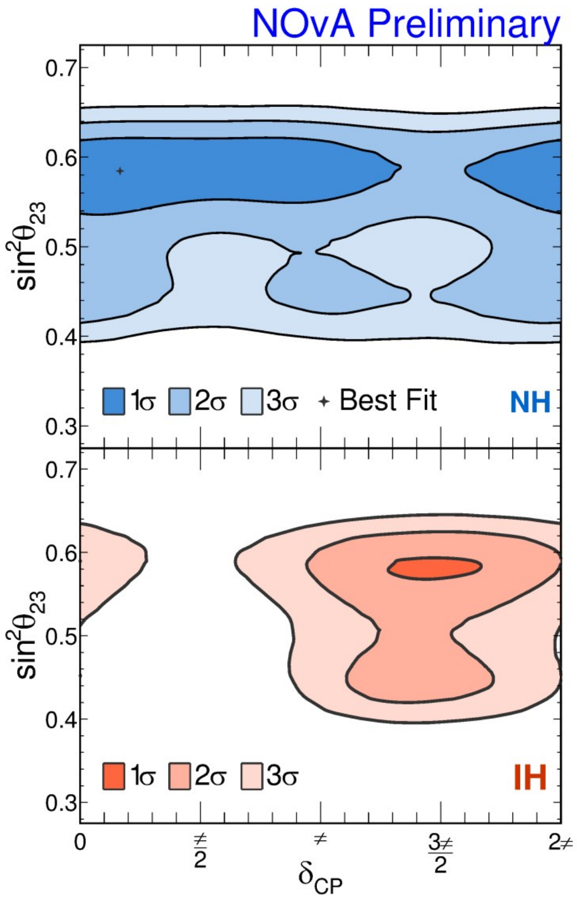
- Insert a simulated electron of the same momentum as the removed muon



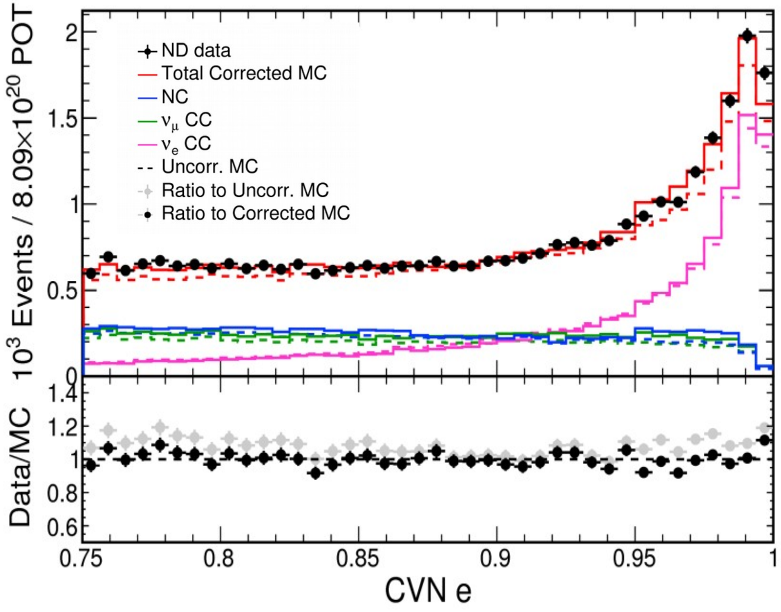
- Data and MC efficiency agree to better than 1%



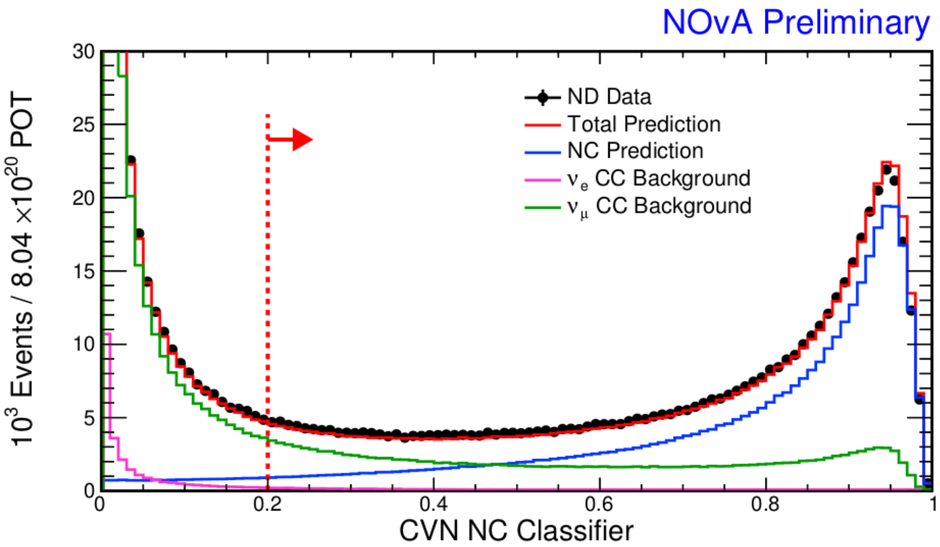
Use Case: ν_e Appearance



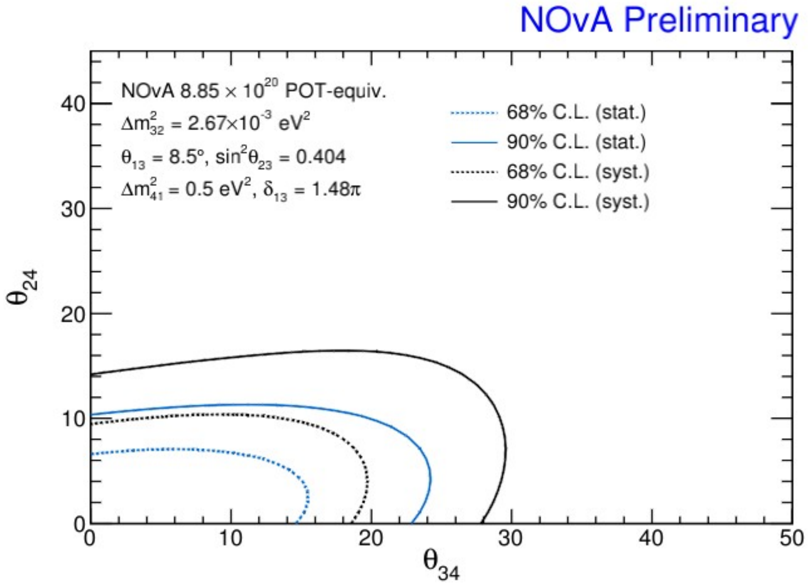
- CVN selected ν_e sample is 73% efficient and 76% pure
  - Improvement of traditional selection techniques equivalent to a 30% increase in detector mass
- Excellent data/MC agreement in the Near Detector



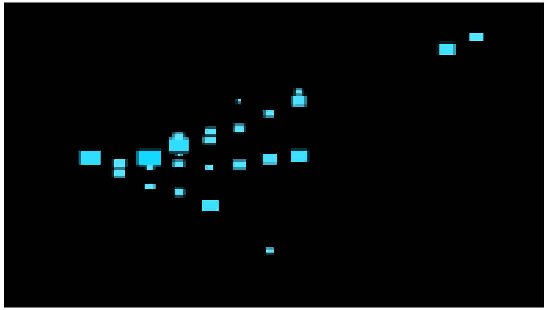
Use Case: Sterile NC Disappearance



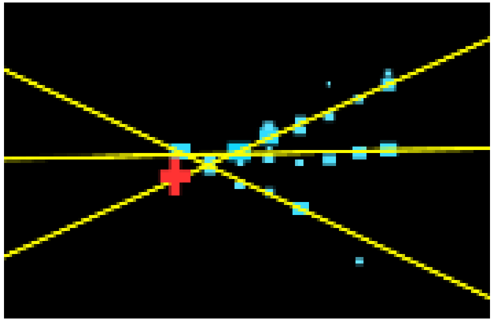
- CVN is in broad use within the experiment
  - Primary selector in standard oscillation analyses
- Is a multi-selector ⟶ provides NC selection for free
- Excellent CC/NC separation
- Good data/MC agreement in the ND
- No evidence of sterile neutrinos



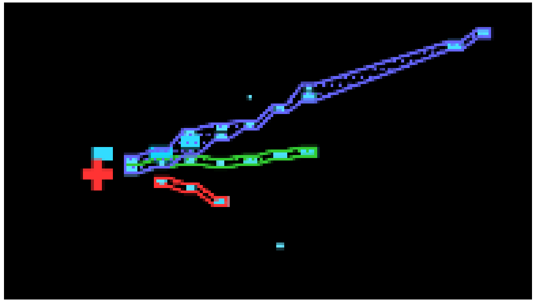
Traditional Reconstruction Methods



Event Separation: Coarse event-level time-space clustering (slicing) using the DBSCAN algorithm.



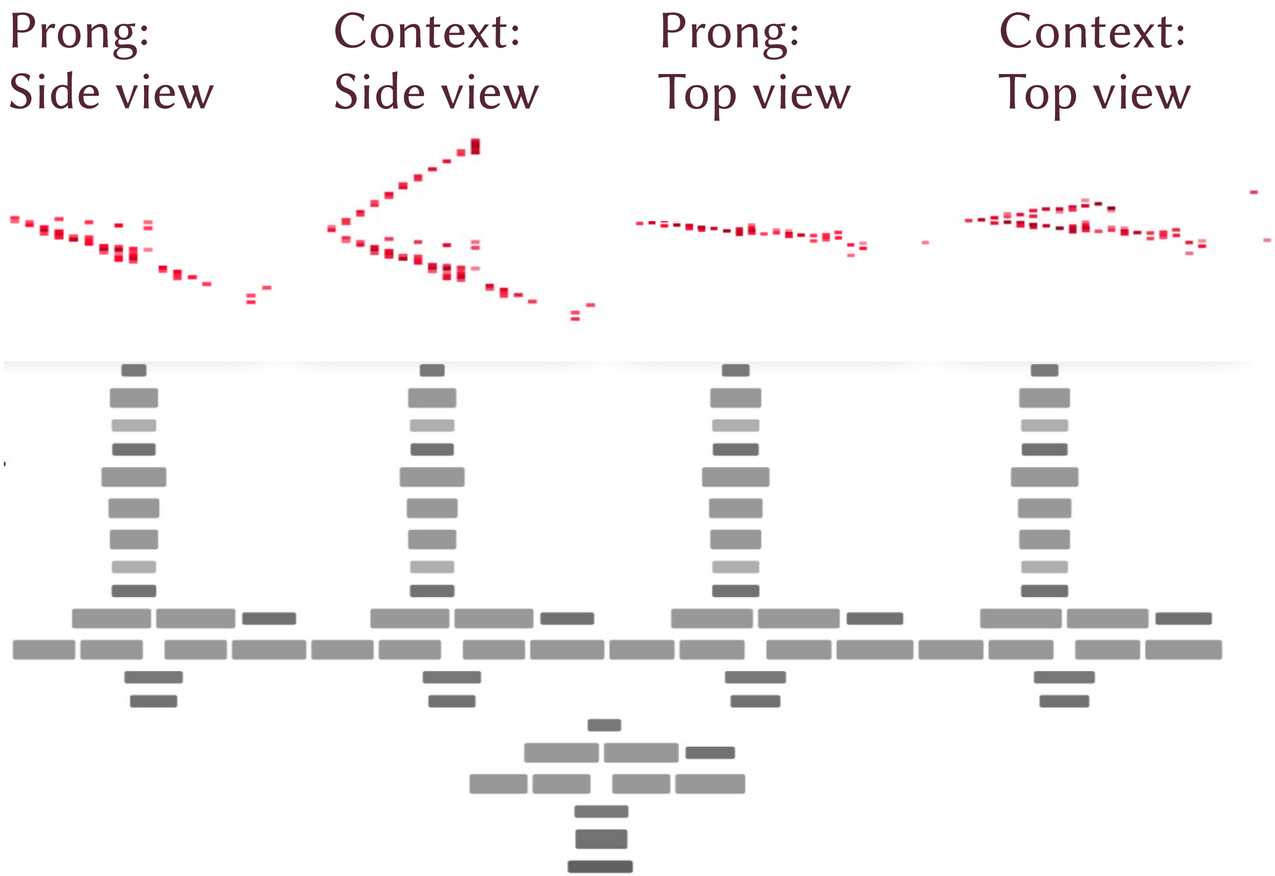
Vertexing: Find lines of energy deposition using the Hough transform. Find the best point line radiate from.



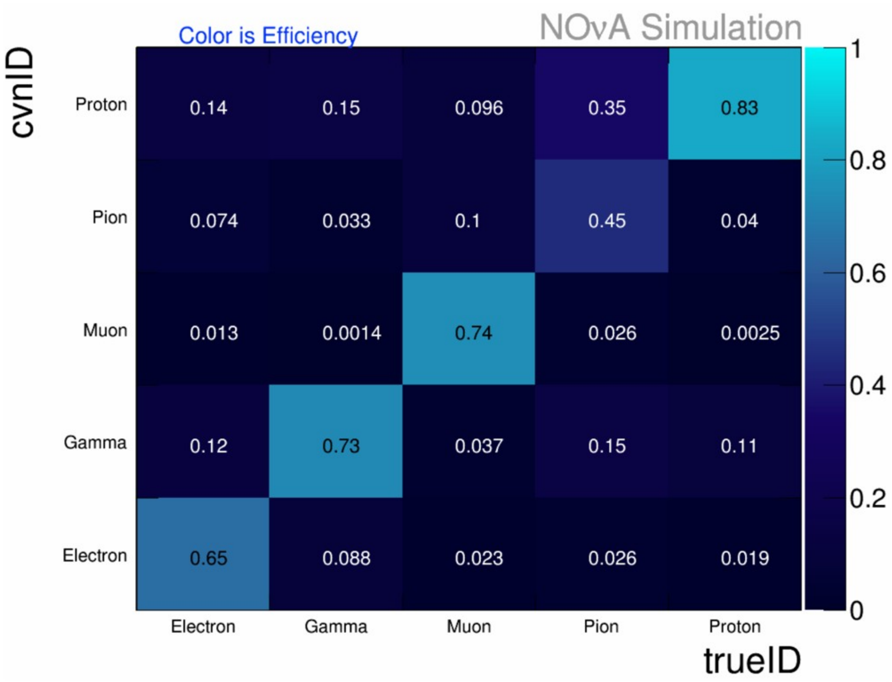
Prong Clustering: Find clusters in angular space around the vertex in each view. Merge views using topology and prong dE/dx.

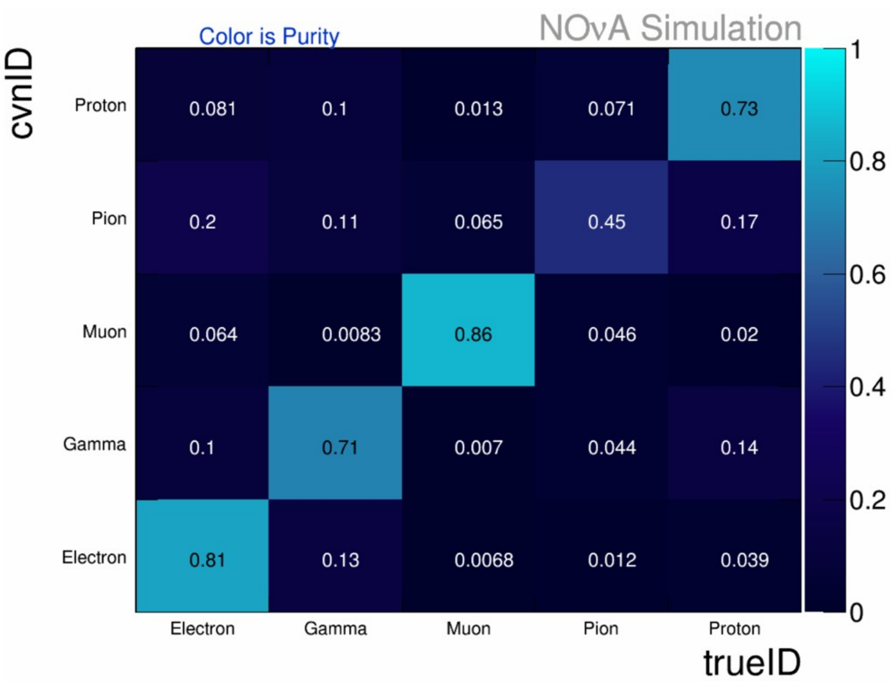
Prong CVN

- Full reconstruction is the dream
  - Still in the future, though instance-aware semantic segmentation is actively being developed
- Attempt to categorize prongs reconstructed using traditional methods to determine the type of particle that created it
- Modify CVN to take 4 views
  - Two prong-only views
  - Two context views
- Depends heavily on the quality of the reconstructed prong
  - This method cannot fix if prong matching between views failed



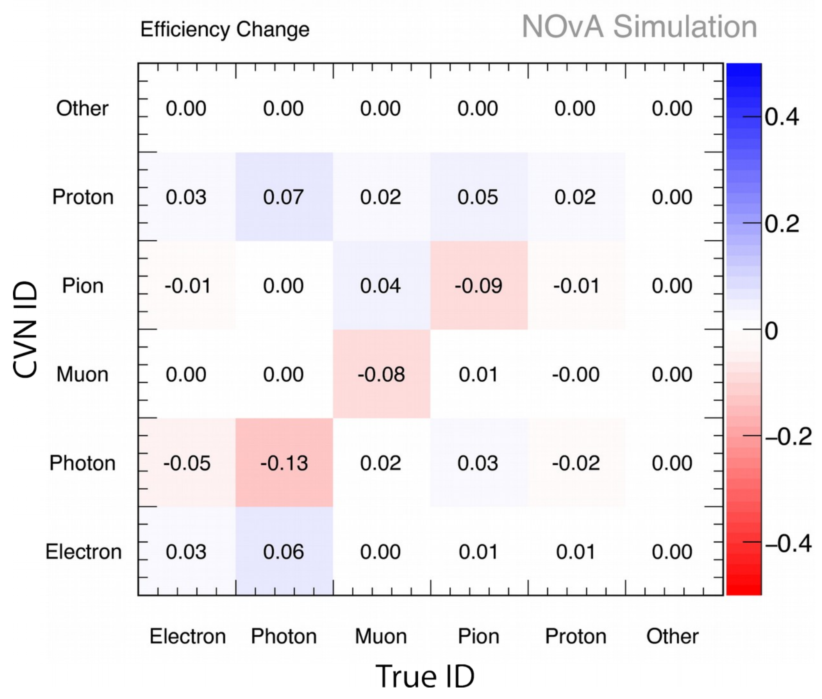
Confusion Matrices



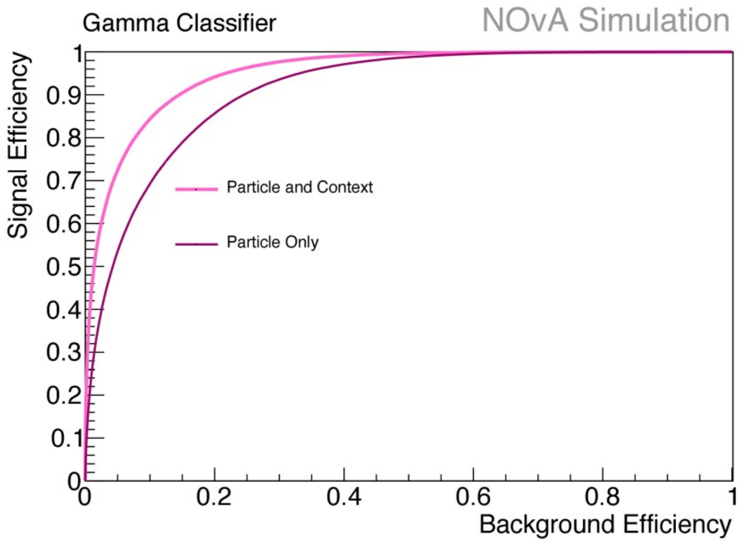


All particles are identified with high efficiency and purity except for pions

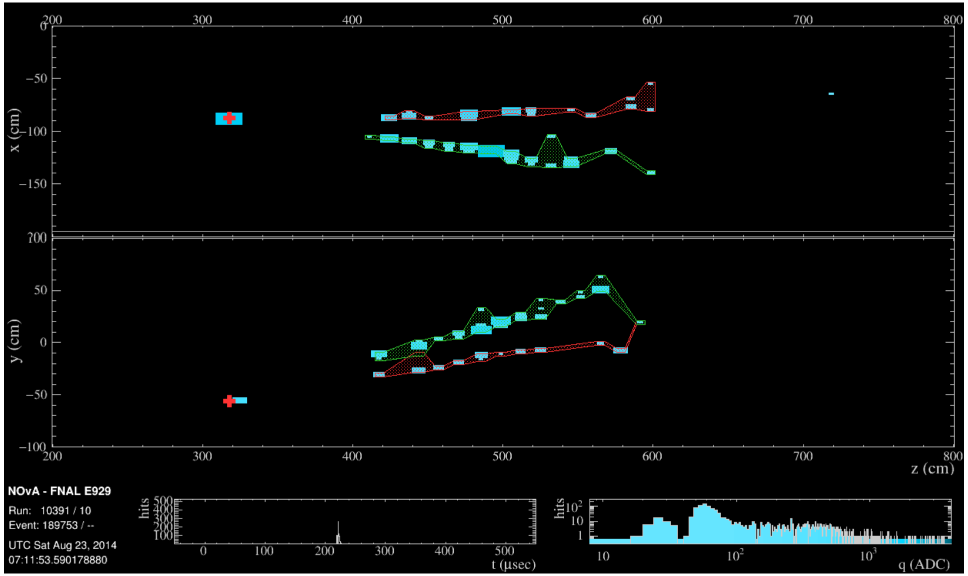
Context Matters



- Adding in full event context introduces dependence on GENIE interaction model
- What happens if context is removed?
- Without context:
  - Muons are more likely to be mis-identified as pions
  - Photons are more likely to be mis-identified as electrons
    - Vertex information is particularly important here

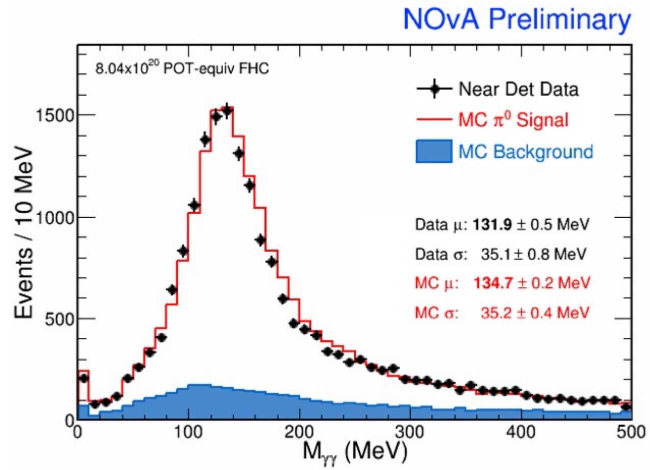


Use Case: π^0 Mass Peak

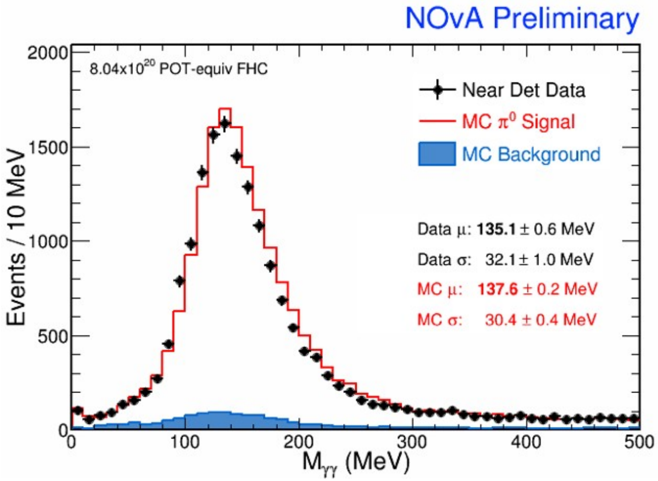


- The π^0 mass peak provides a way to test the ND energy response in data and MC
- Look for two photons emerging from a common vertex
- Prong CVN produces a 12% improvement in selection purity with similar efficiency to traditional methods.

Traditional selection

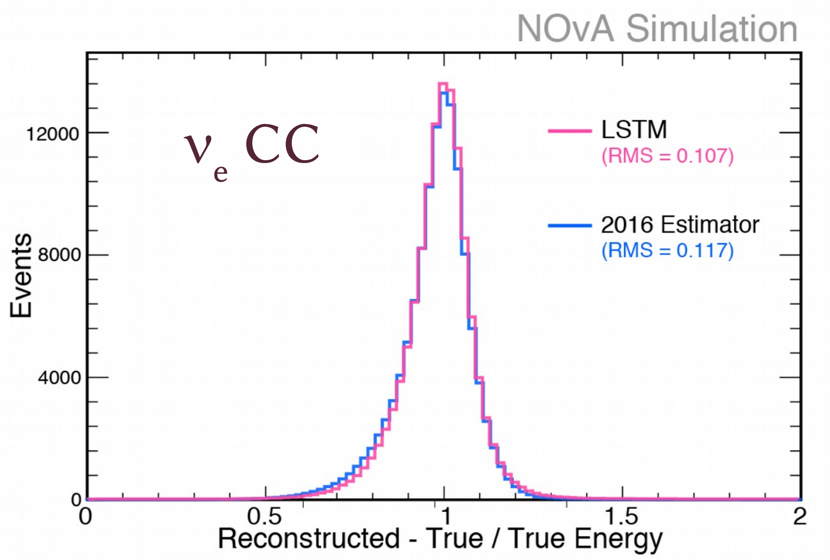


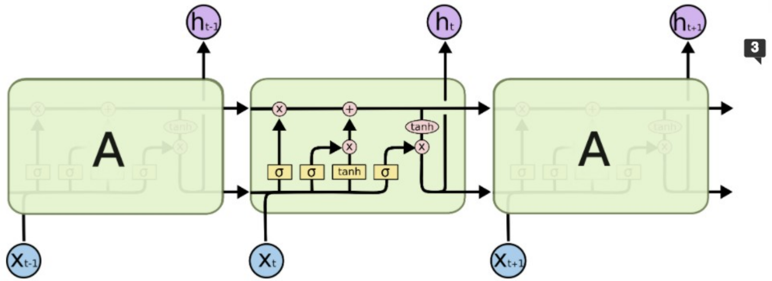
Prong CVN selection



LSTM Energy Estimation

- Recurrent neural networks and LSTMs provide a way to accept sequences as NN input
- Current neutrino energy estimators use coarse-grained energy information
  - Calorimetric energy of full hadronic system is usually used as a single input
- Use an LSTM network to accept variable length list of prongs with particle hypotheses and energy and momentum estimates
- Method builds up a representation of the event that models particle kinematics
- First results show improved energy estimation



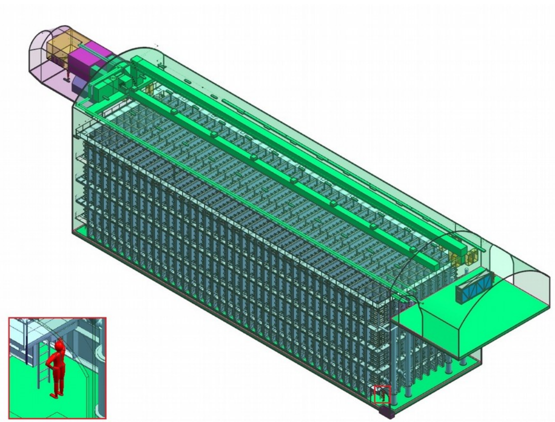


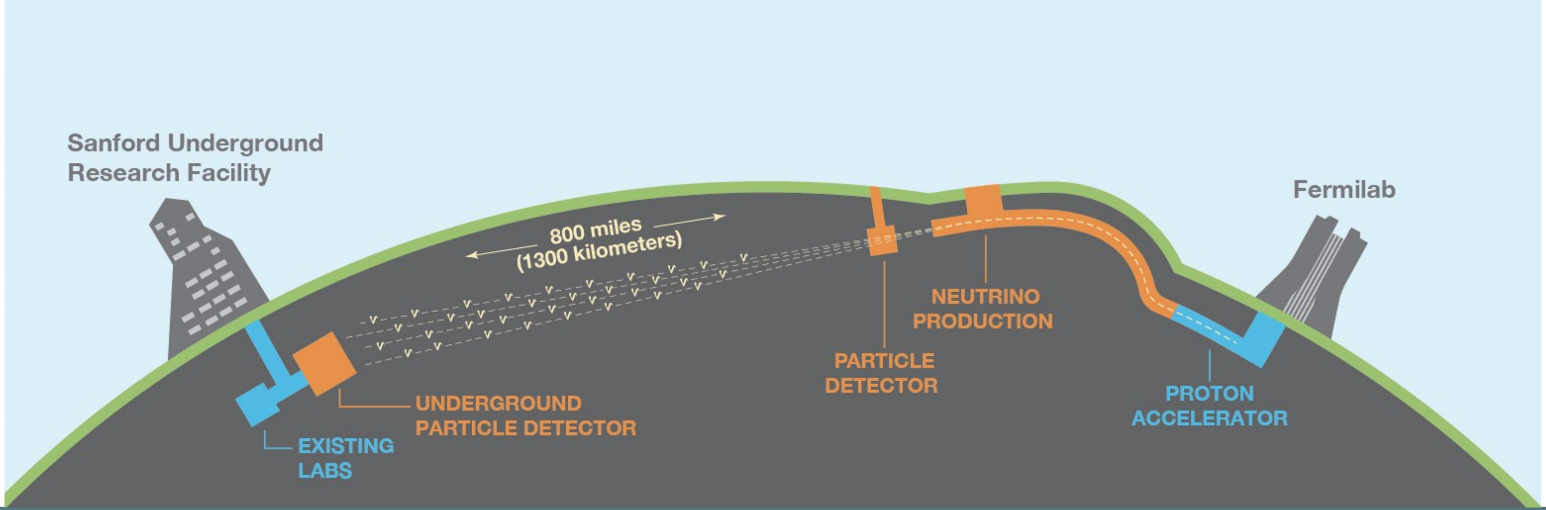
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

DUNE

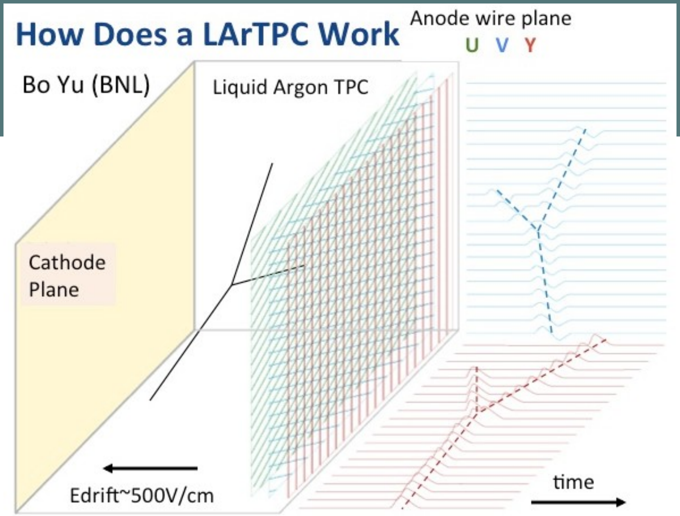
- Future flagship long-baseline neutrino experiment
- Goals:
  - Observe ν_e appearance and ν_μ disappearance in a wideband beam
  - Measure mass hierarchy, CP-violating phase, and atmospheric parameters with high precision
- Ancillary program: sterile neutrinos, ν_τ appearance, solar neutrinos, dark matter, and more

- Near Detector 575 m from target
- Far Detector 1,300 km from target, 1.5 km underground
  - Liquid argon time projection chamber technology for high resolution imaging
  - 4 modules, each with at least 10 kton fiducial mass

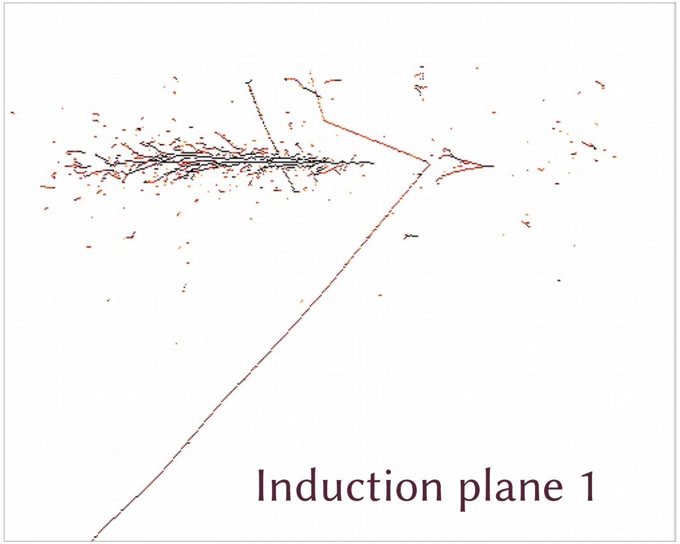


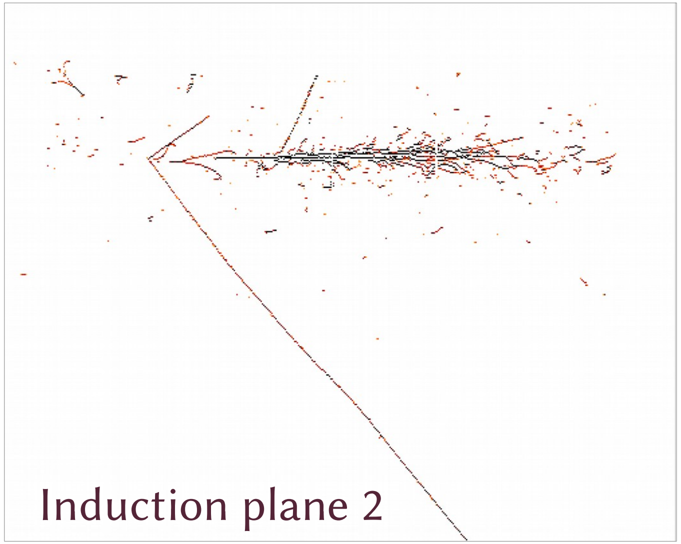


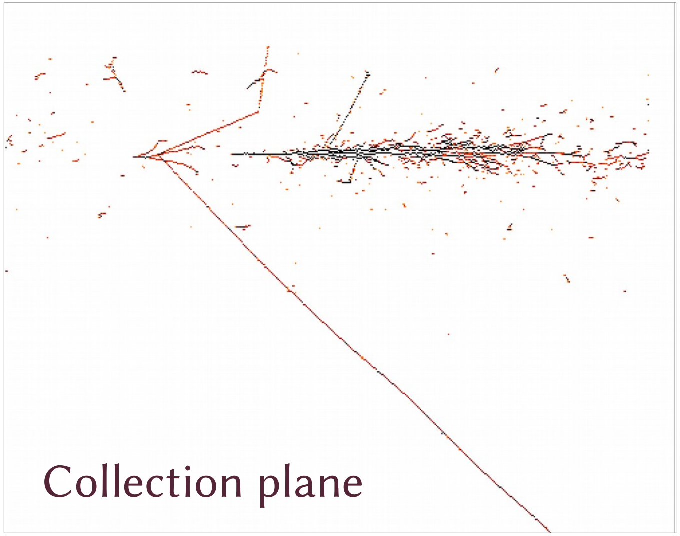
Promise of LArTPC Technology



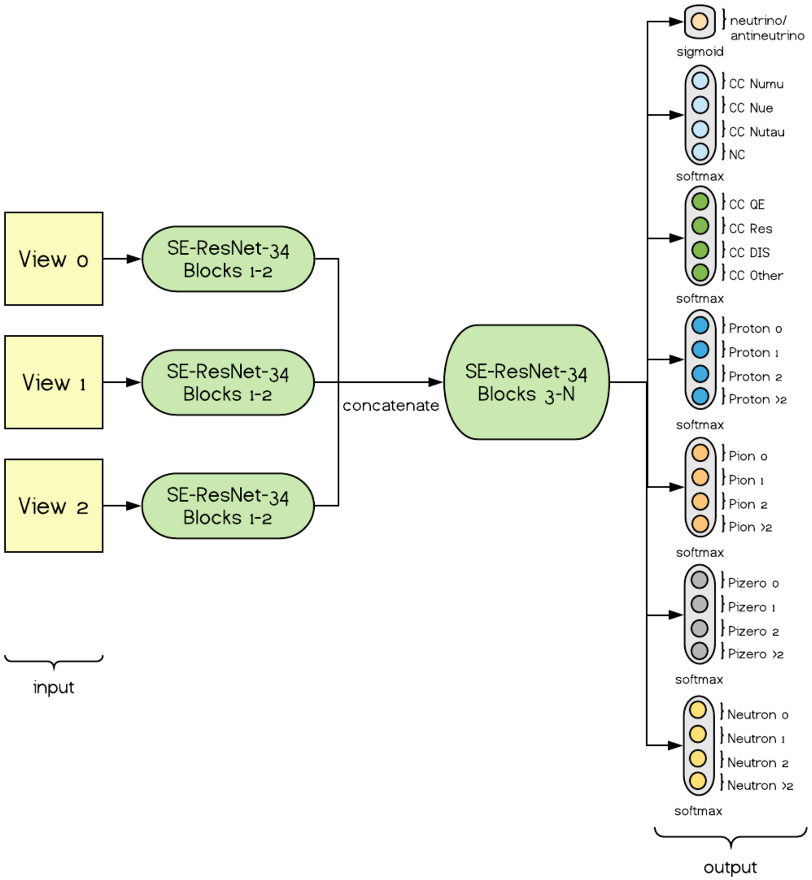
- Wire planes have a 3 mm spacing
- Drift field is 500V/cm
- Maximum drift length is 3.53 m
- Spatial resolution is good enough to distinguish individual electron-positron pairs in an EM shower!
  - Blessing and curse – high resolution images are hard to automatically reconstruct
- For classification, use regions 500 wires long and 1.2 ms wide (in 500 time slices)





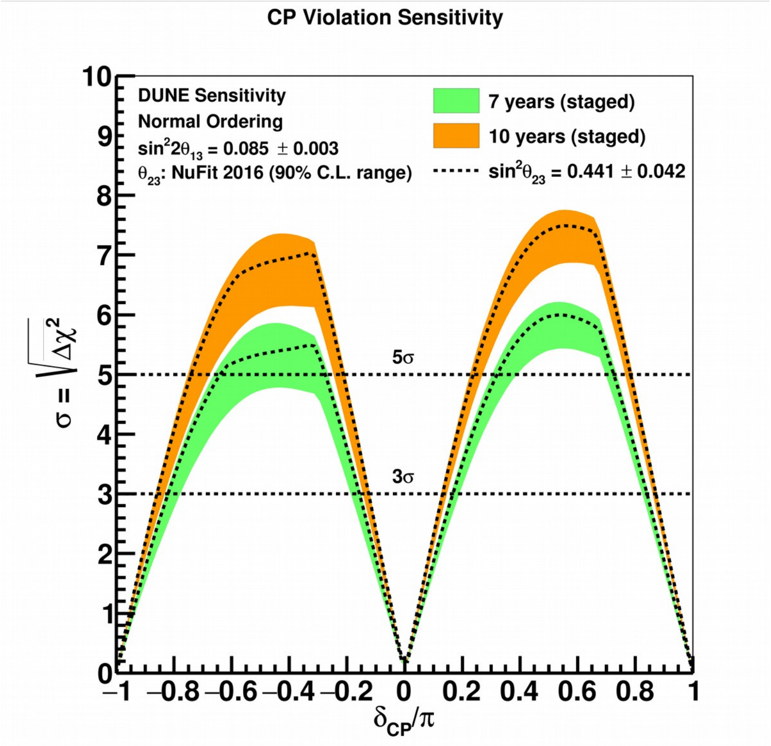


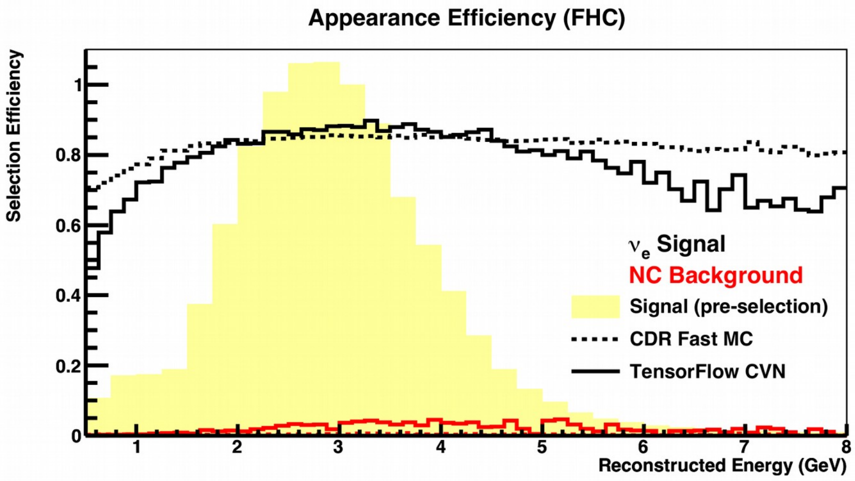
Multi-target Network



- Network is inspired by the NOvA network, but with a few key differences
- Underlying architecture:
  - GoogLeNet → SE-ResNet-34 (arXiv:1709.01507v2)
- Three views separately fed into first two blocks of SE-ResNet-34, then concatenated and fed into remaining blocks
- Multi-target outputs allow for learning many characteristics of the neutrino interaction in one pass
  - Neutrino/Antineutrino
  - Flavor
  - Interaction type
  - Proton multiplicity
  - Charged pion multiplicity
  - Neutral pion multiplicity
  - Neutron multiplicity

Effect on CP-Violation Analysis





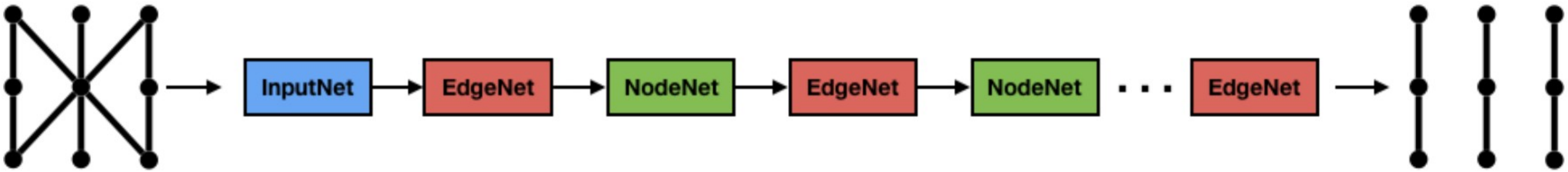
- CDR assumptions of the power of LAr technology implied significant discovery potential over most of the parameter space
- Traditional reconstruction and shallow machine learning methods could not reproduce CDR expectations
- CVN is the first method that exceeds CDR expectations in the most important energy range
  - It does not appear that we have exhausted potential improvements yet

Toward Full Reconstruction

Graph Convolution Network Tracking

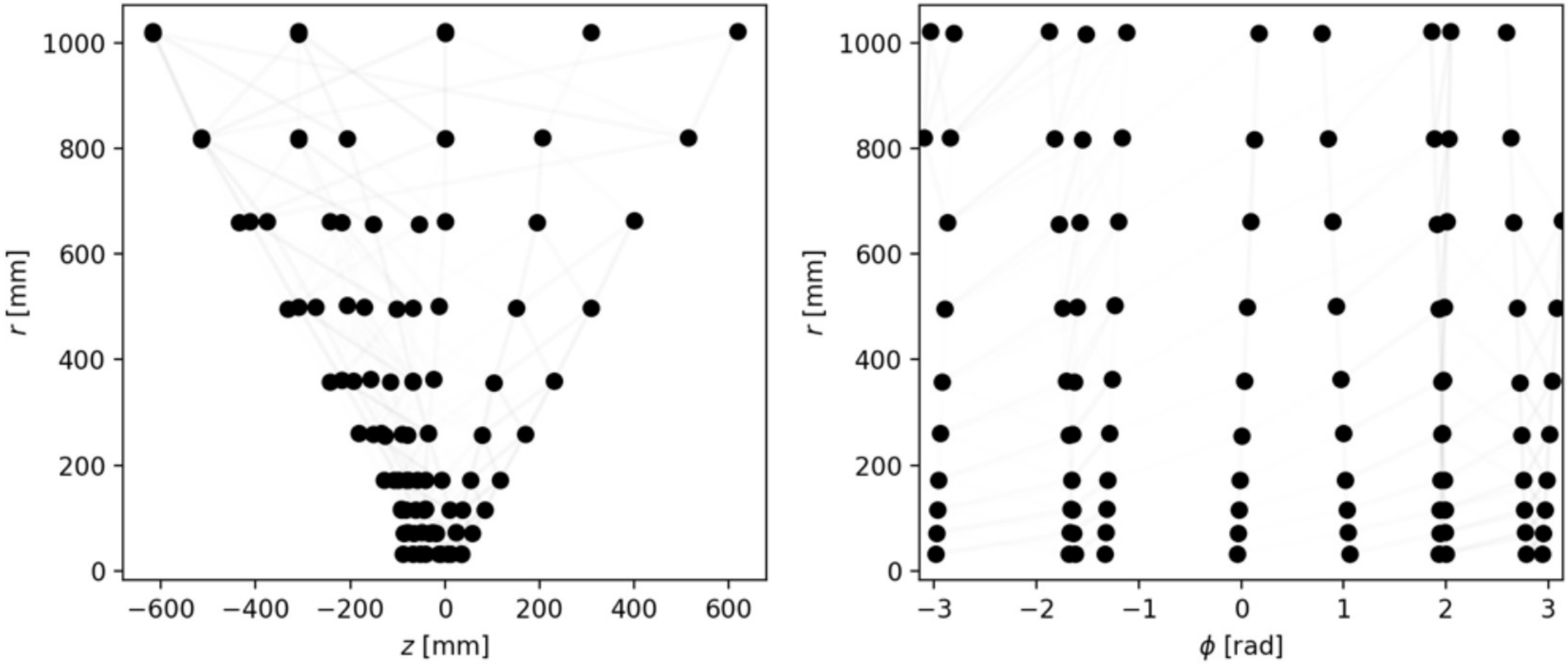
S. Farrell, HEP.TrkX Collab, CtD 2018

- HEP.TrkX collaboration developed novel tracking technique using graph neural networks, part of the Geometric Deep Learning field
- Represent tracker hits as *nodes* in a graph and possible connections between hits as *edges*
  - Need some sort of pre-processor, like a Hough Transform, to generate the initial graph
- Alternate two networks
  - Edge network computes weights for every edge using features of start and end nodes
  - Node network computes new node features using edge weights and current node features



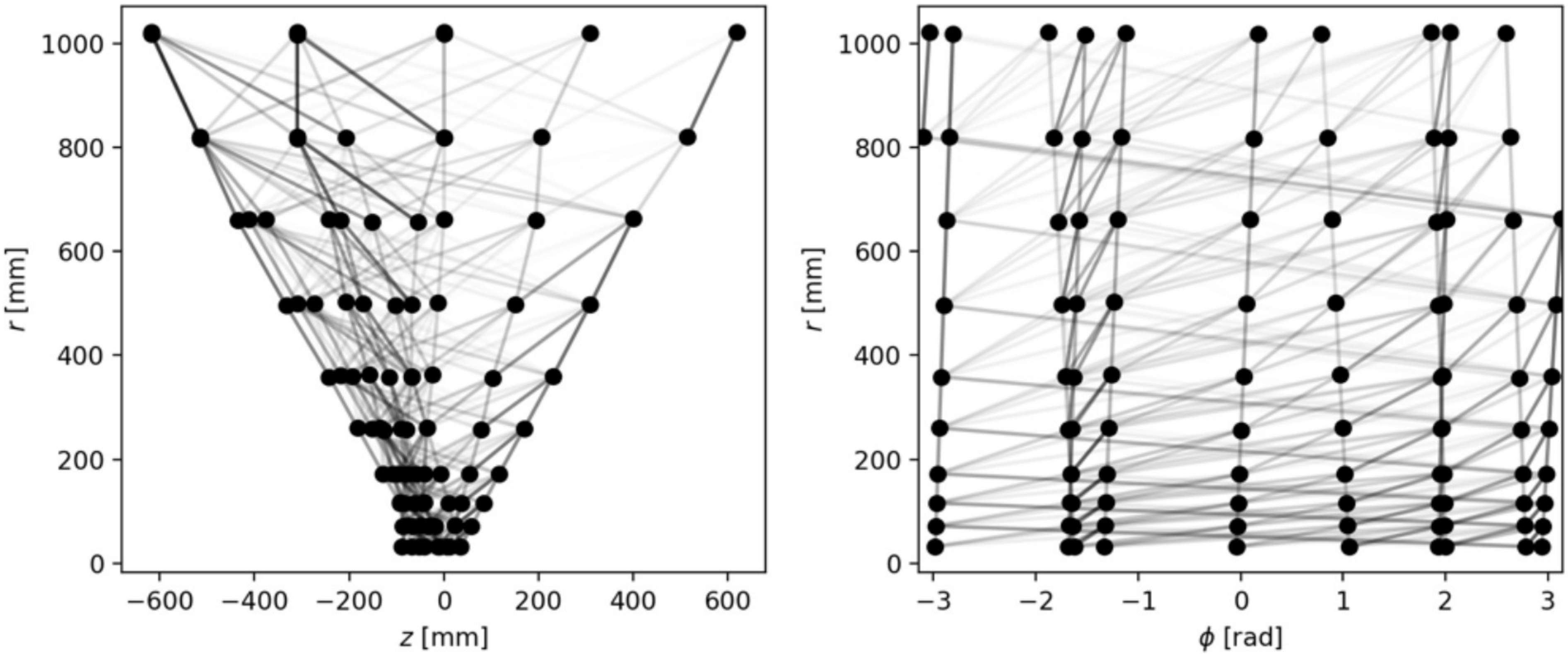
Graph Convolution Network Tracking: Iteration 1

S. Farrell, HEP.TrkX Collab, CtD 2018



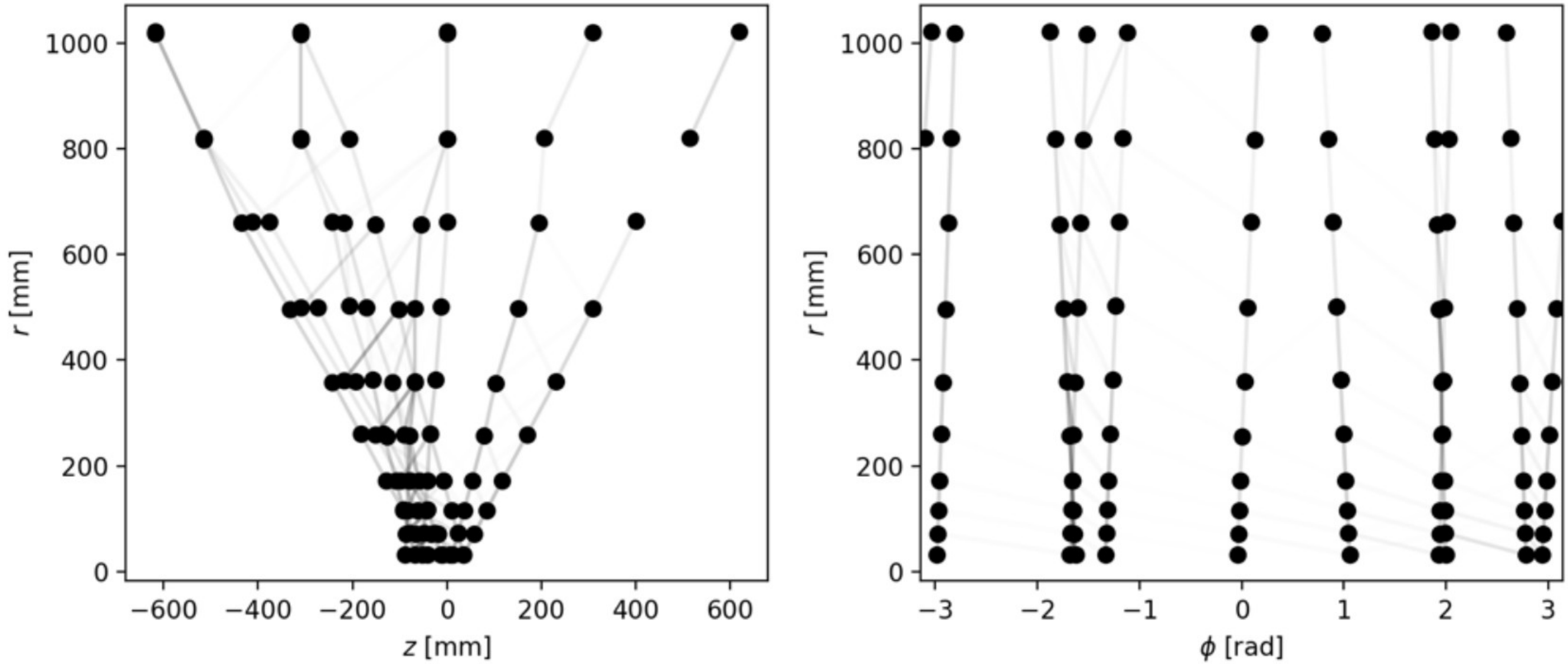
Graph Convolution Network Tracking: Iteration 2

S. Farrell, HEP.TrkX Collab, CtD 2018



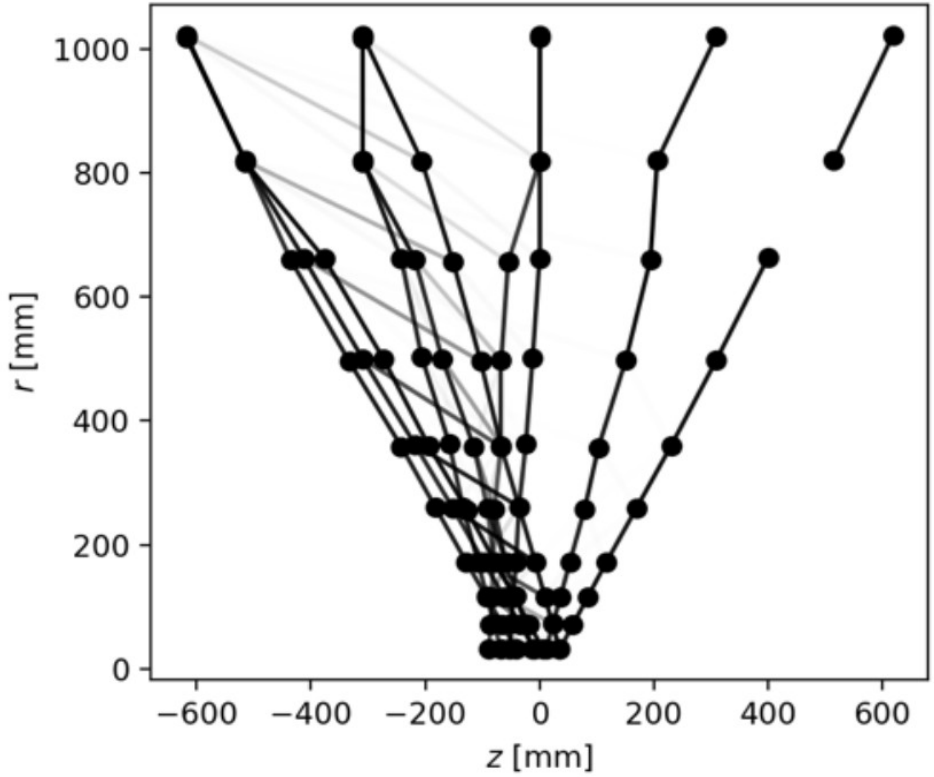
Graph Convolution Network Tracking: Iteration 3

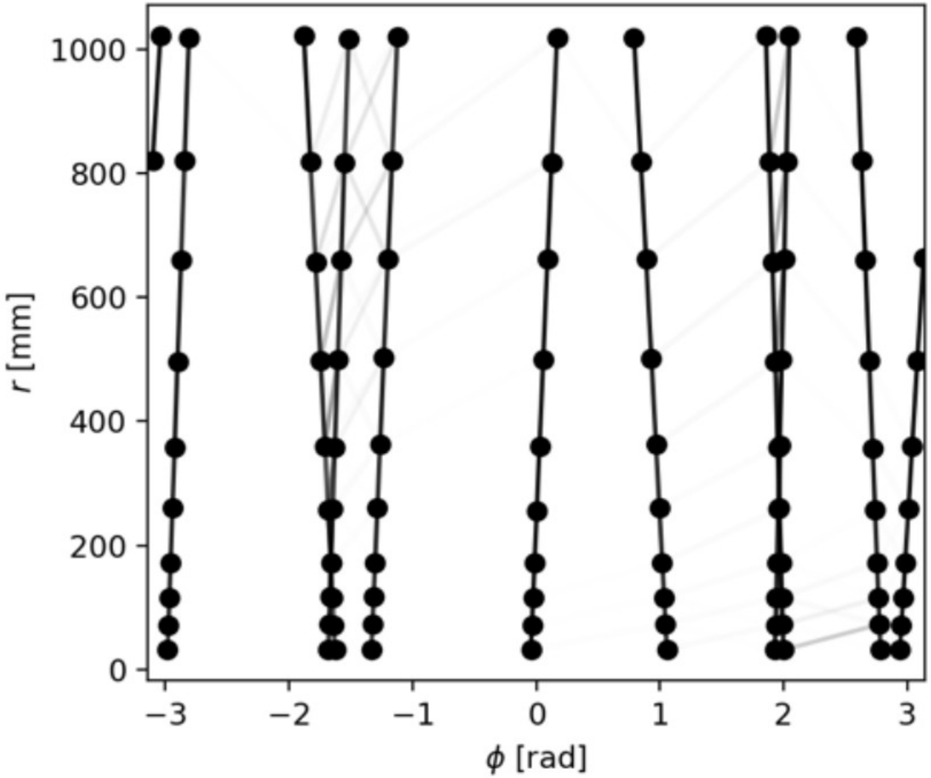
S. Farrell, HEP.TrkX Collab, CtD 2018



Graph Convolution Network Tracking: Iteration 4

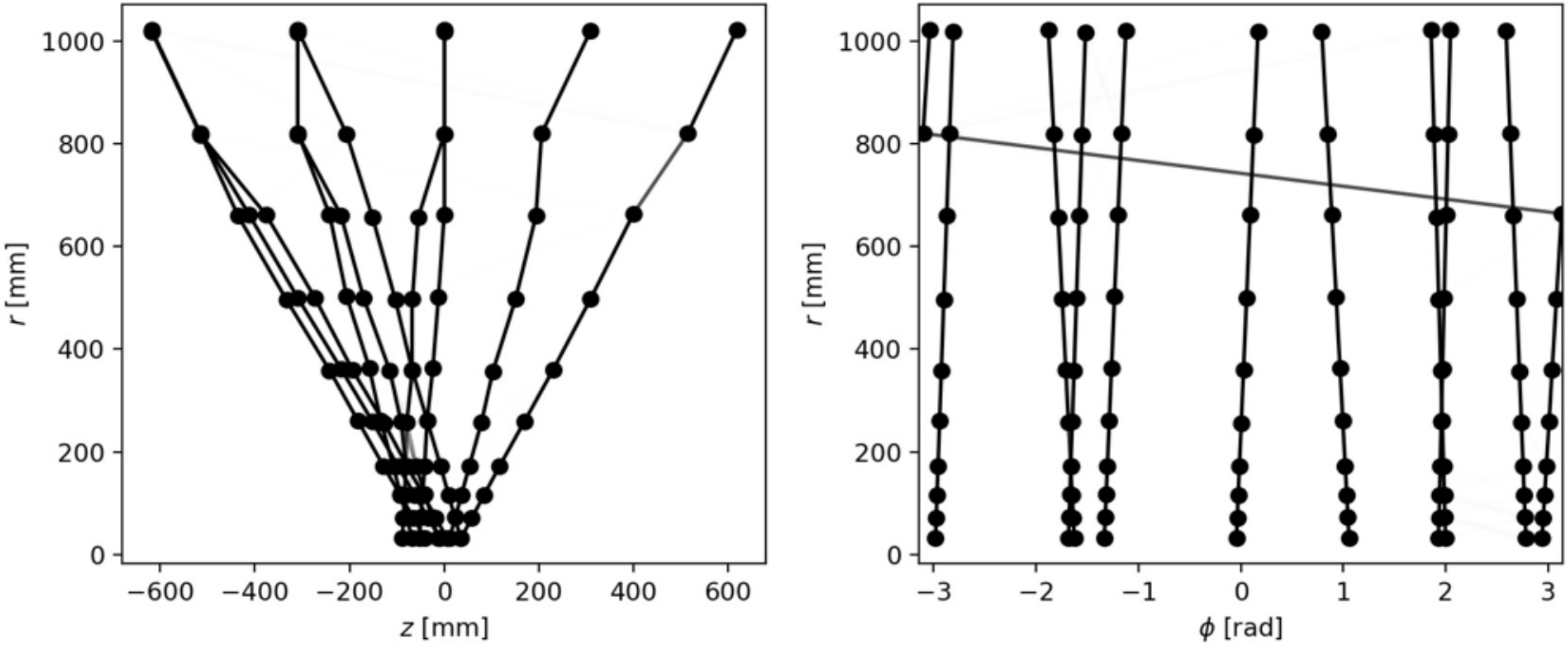
S. Farrell, HEP.TrkX Collab, CtD 2018





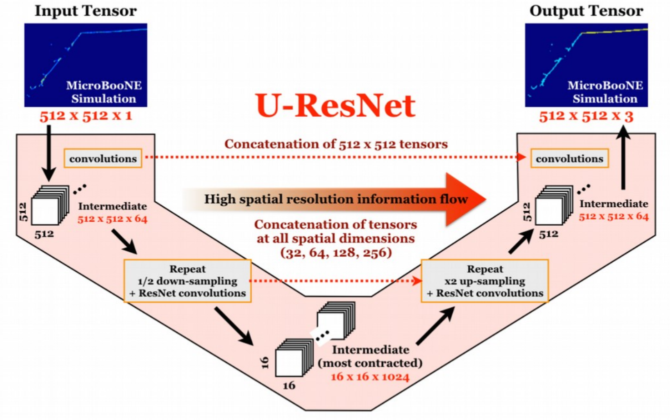
Graph Convolution Network Tracking: Final Iteration

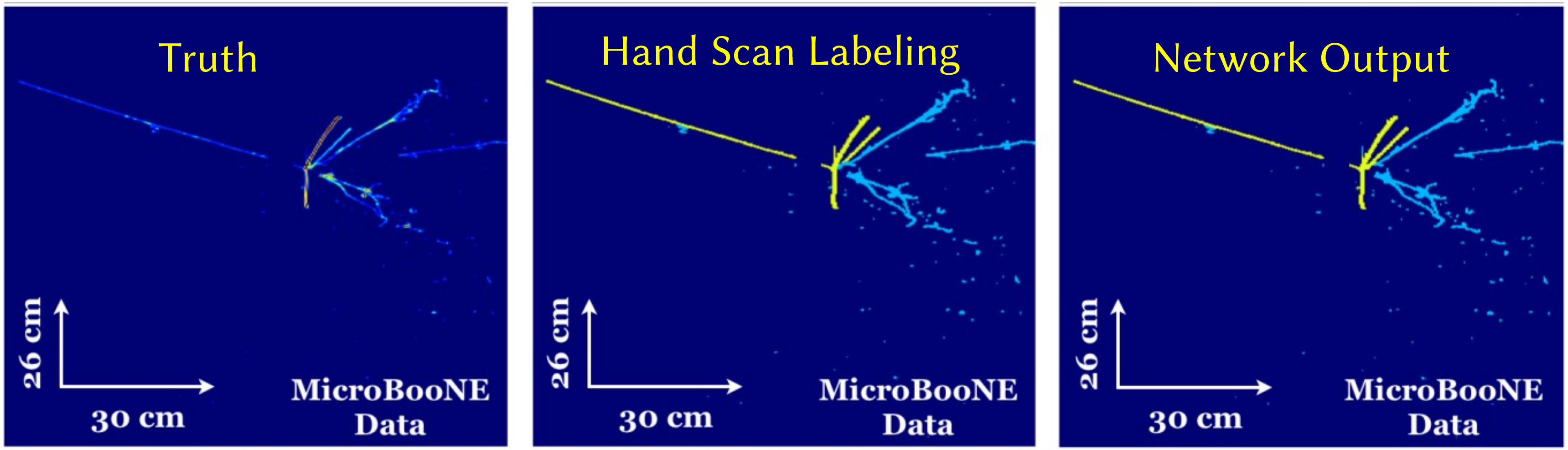
S. Farrell, HEP.TrkX Collab, CtD 2018



Semantic Segmentation

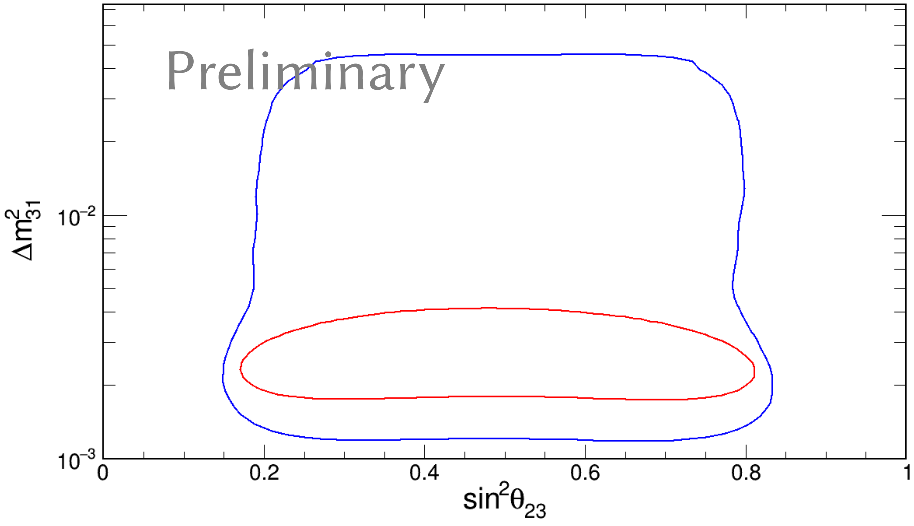
- Based on UNet architecture
- Classifies each pixel as either part of a track or a shower
- Perform classification before higher level reconstruction (classification guides reco)
- Could be used to classify the type of particle creating the hits as well (currently being explored at NOvA)





Why Does This Matter for BSM Searches?

- DUNE will have opportunities to study tau neutrino appearance
  - Low energy beam: 130 ν_τ/year
  - High energy beam: 800 ν_τ/year
  - Atmospherics: 30 ν_τ/year
- High efficiency and purity will be critical
  - For atmospherics, per kton-year
    - 88 ν_e
    - 102 ν_μ
    - 95 NC
    - 1 ν_τ
- Truth studies (J. Conrad, A. de Gouvea, S. Shalgar, and J. Spitz, PRD 82, 093012 (2010)) suggest ~30% signal efficiency and ~0.5% NC efficiency may be possible
  - Traditional techniques and DUNE CVN cannot achieve this yet – may need full deep-learning reconstruction to attain



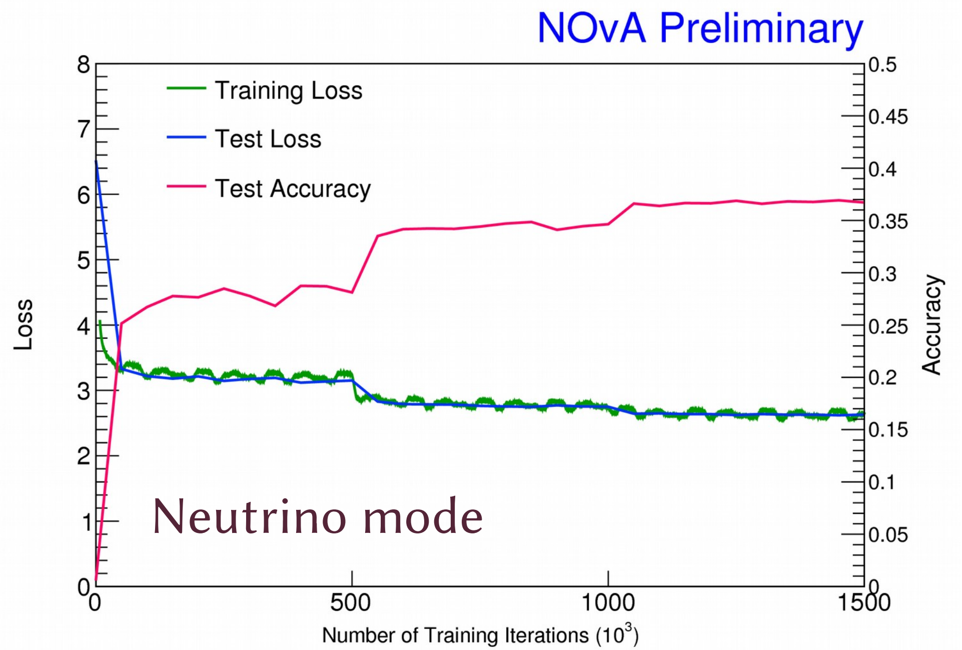
- Atmosopheric ν_τ sensitivity
- Sensitive to reconstruction capabilities
- Blue: IceCube-Deep Core zenith angle and energy resolutions
- Red: ½ IceCube-Deep Core resolutions

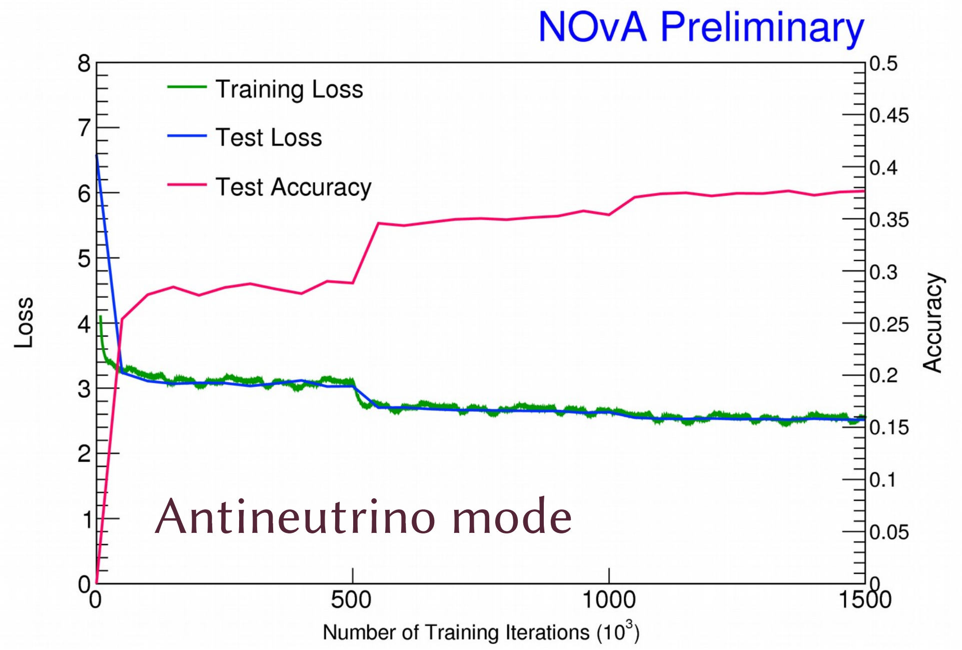
Summary

- Modern deep learning techniques are very powerful and are an excellent fit for neutrino experiments
- Using a modified GoogLeNet architecture, it is possible to create a universal neutrino interaction classifier that
  - Uses minimal reconstruction
  - Significantly increased our ν_e selection efficiency
    - Equivalent to a 30% increase in detector mass
- CVN was central to the ν_e appearance analysis
- In wide use throughout NOvA and is critical for the NC disappearance analysis
- Efforts to develop reconstruction methods using deep learning are underway
  - Prong CVN classifies prongs according to the particle that created them with high efficiency and purity
  - Full reconstruction based on instance aware semantic segmentation is being actively developed
- CNNs were critical for DUNE to achieve design sensitivities with realistic simulation and reconstruction
  - Multi-target outputs provide a fine grained view into neutrino interactions

Thank You!

Training


NOvA Preliminary
Training Loss
Test Loss
Test Accuracy
Loss
Accuracy
Number of Training Iterations (10^3)
Neutrino mode


NOvA Preliminary
Training Loss
Test Loss
Test Accuracy
Loss
Accuracy
Number of Training Iterations (10^3)
Antineutrino mode

CNN Energy Estimation

- Also attempts to estimate energy directly from pixel maps with minimal reconstruction
- A variant of the CVN architecture
  - Adds in extra reconstructed vertex information into the fully connected layer