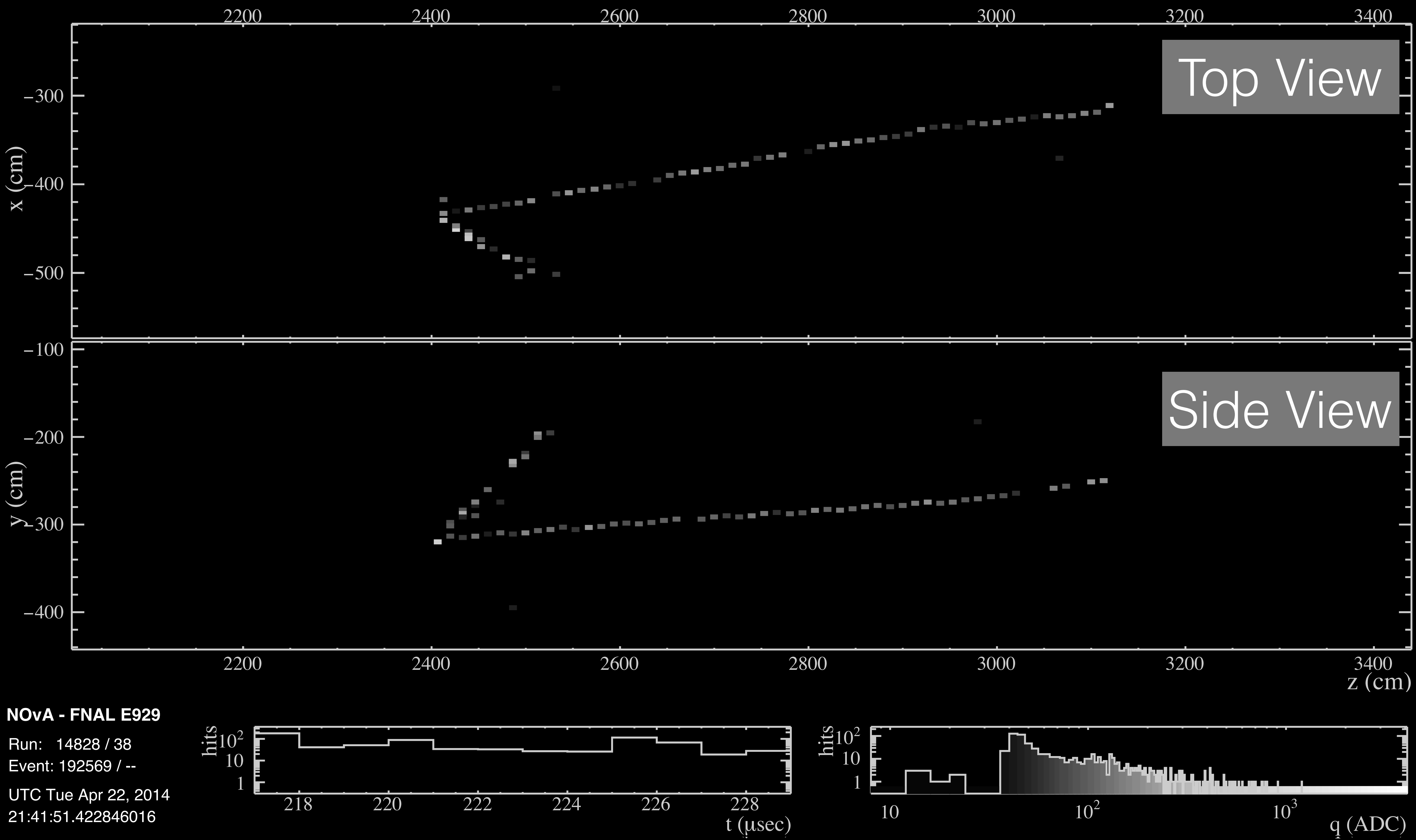
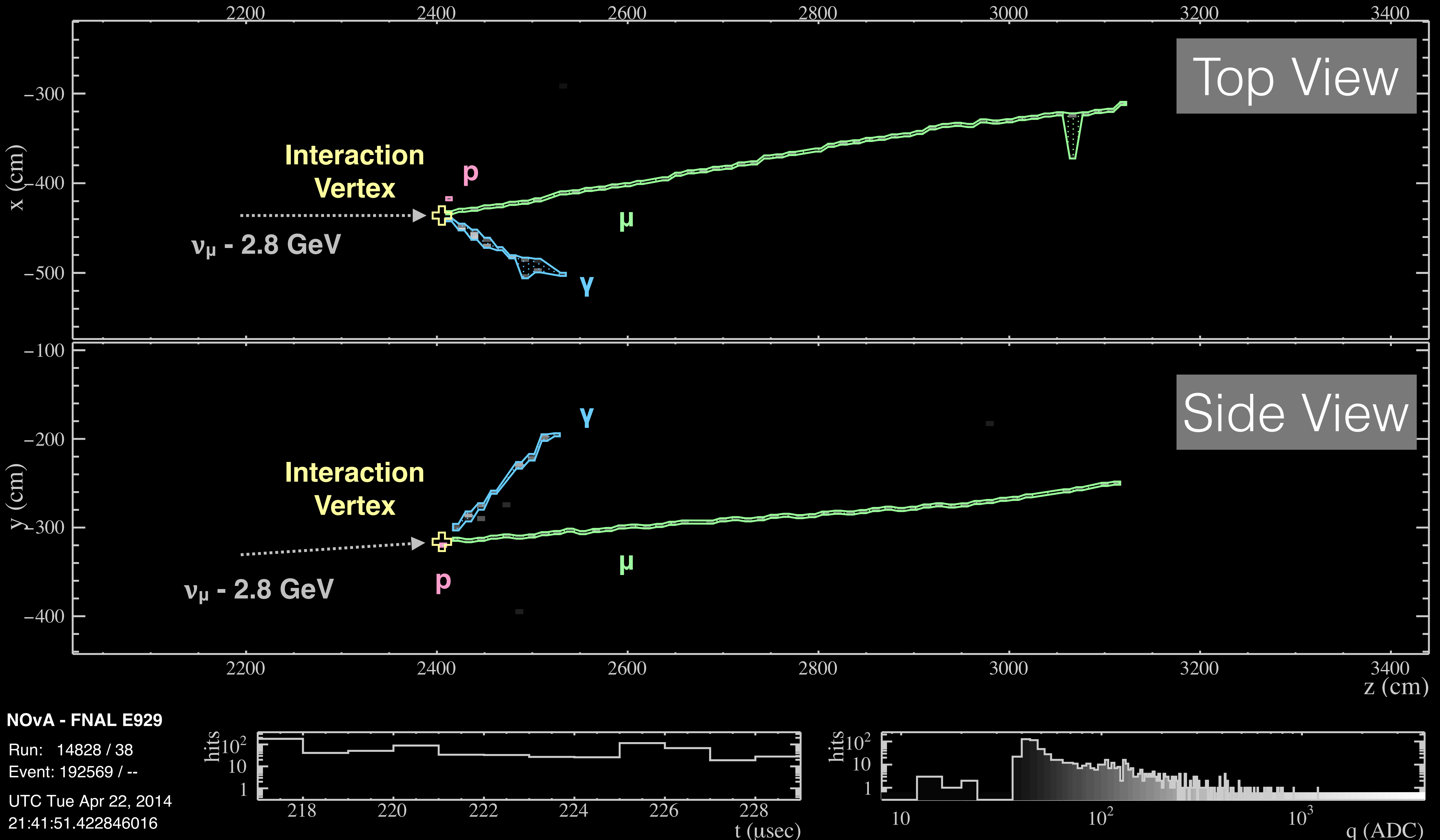
Particle Identification with a Context-Enriched Convolutional Neural Network in the NOvA Experiment

Ryan Murphy | INDIANA UNIVERSITY
On Behalf of the NOvA Collaboration

Typical Event

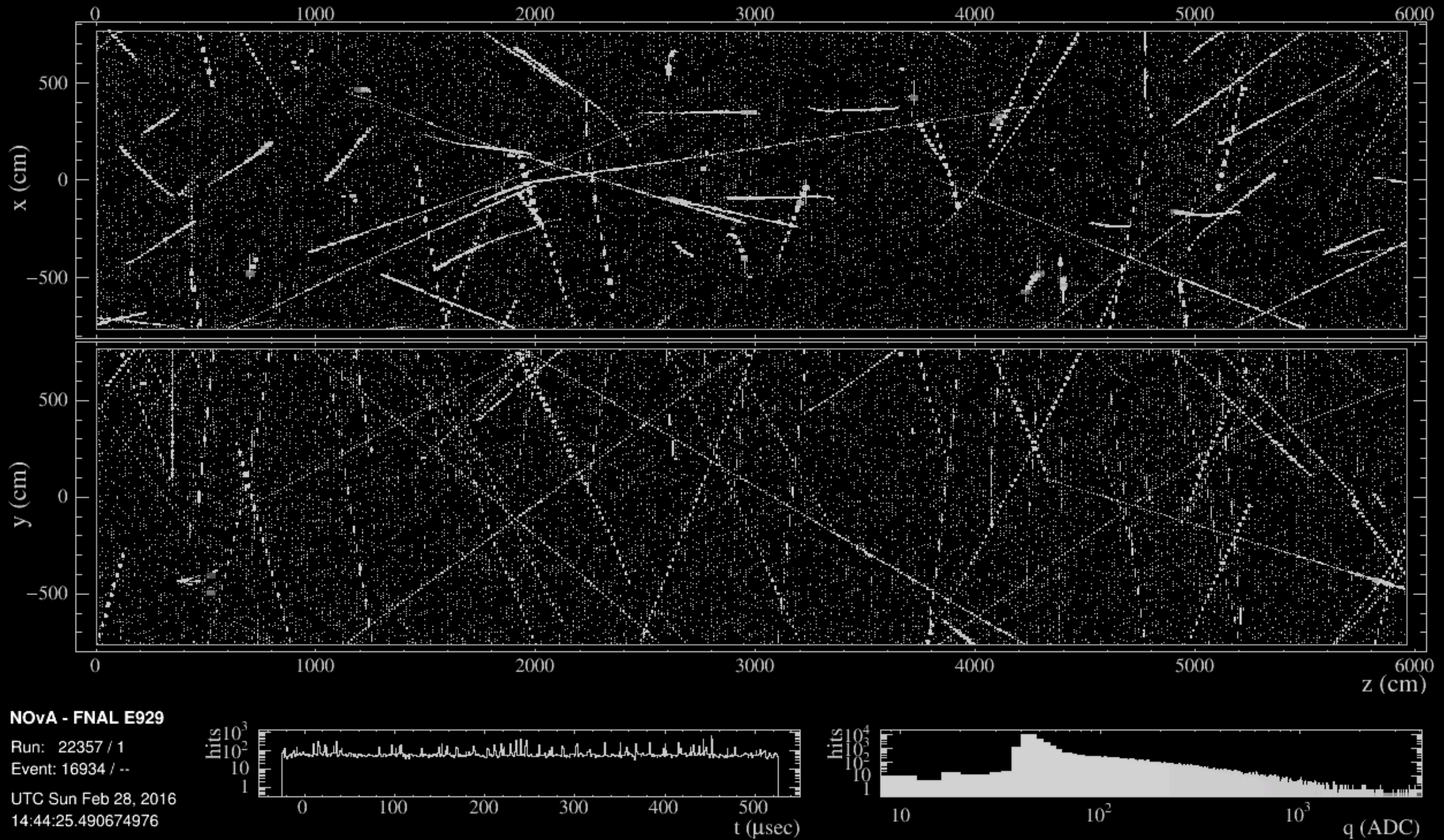


Typical Event



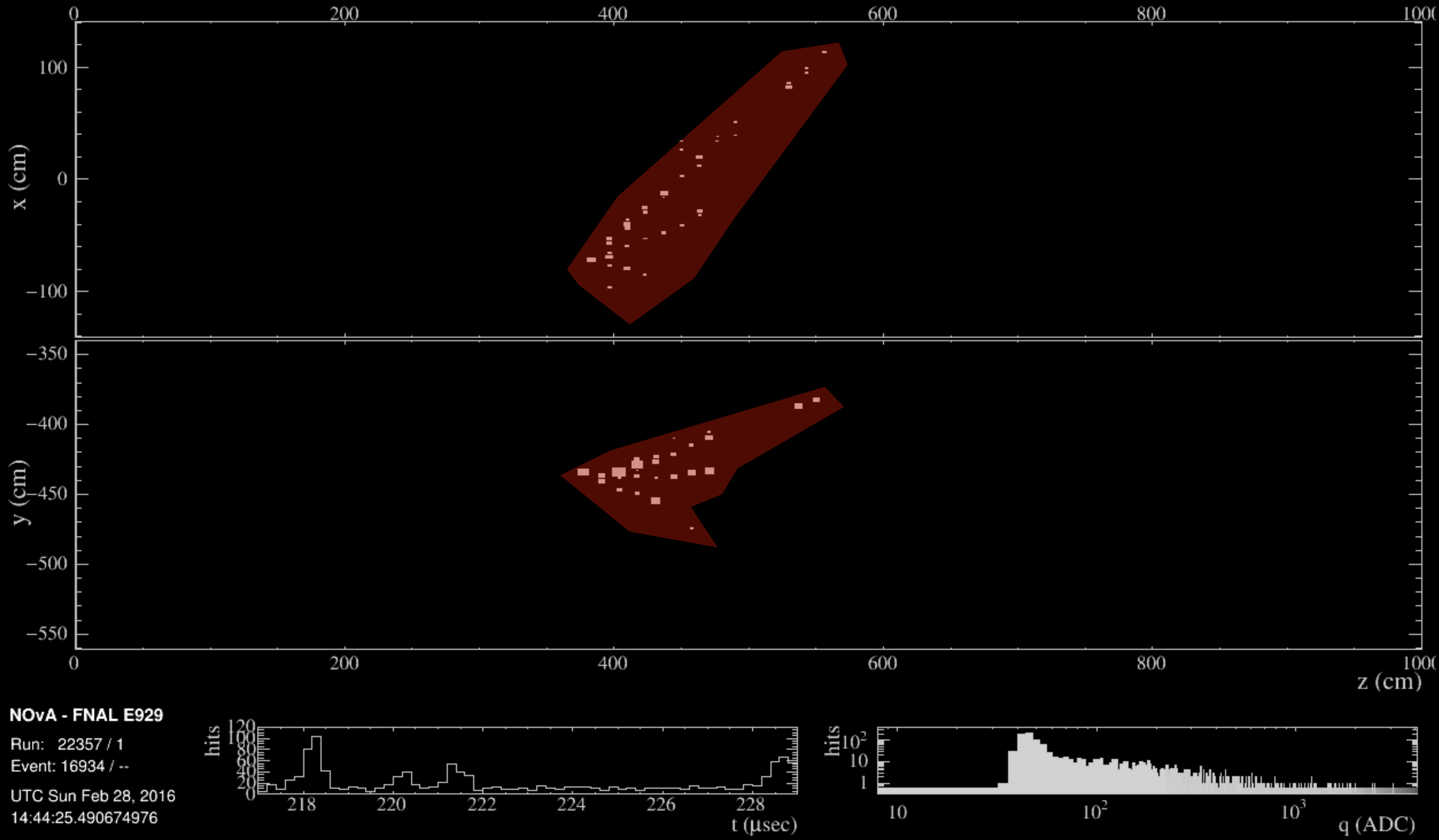
Traditional Reconstruction

550 us of Far Detector data



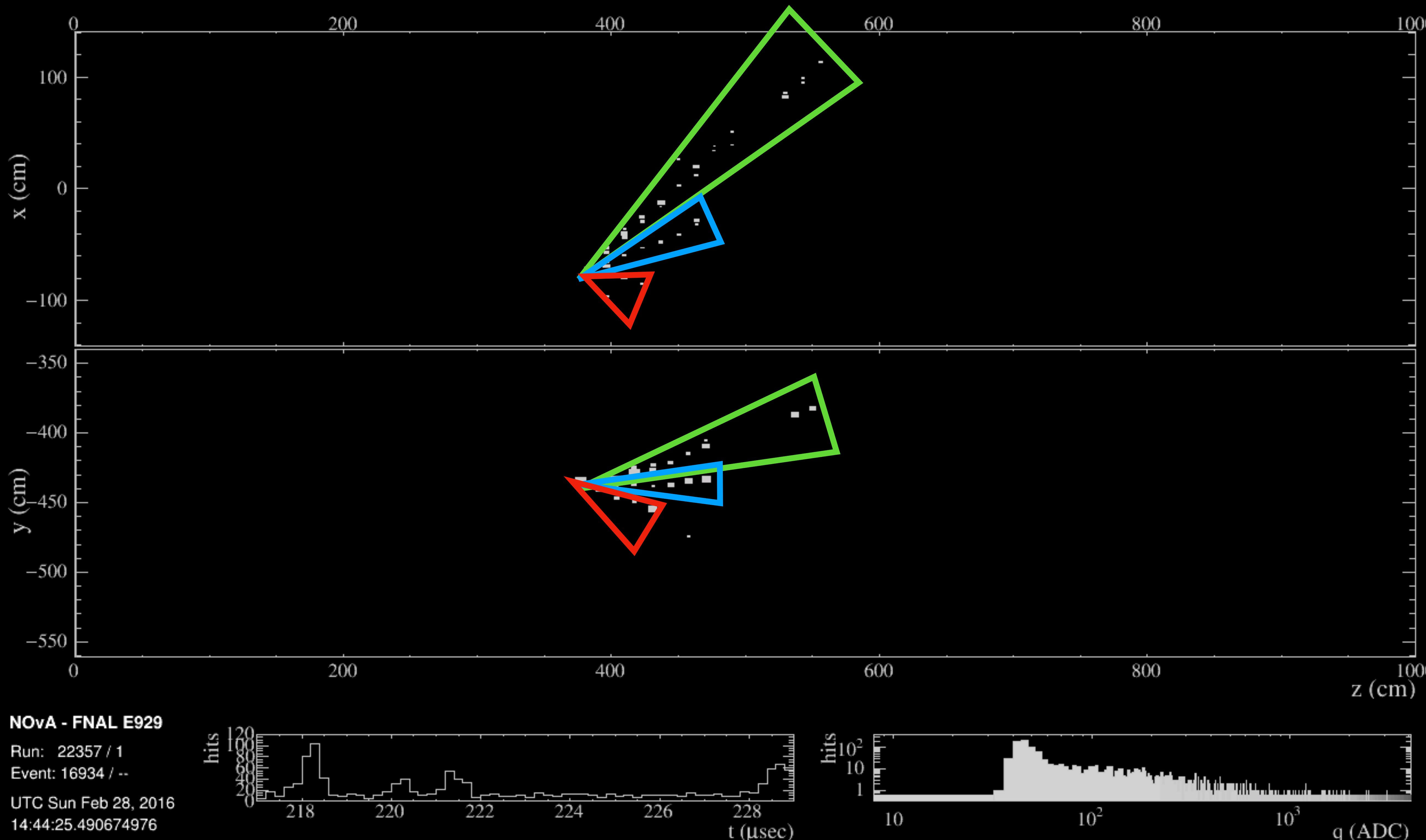
Traditional Reconstruction

Zoom in on beam window, group hits by space and time

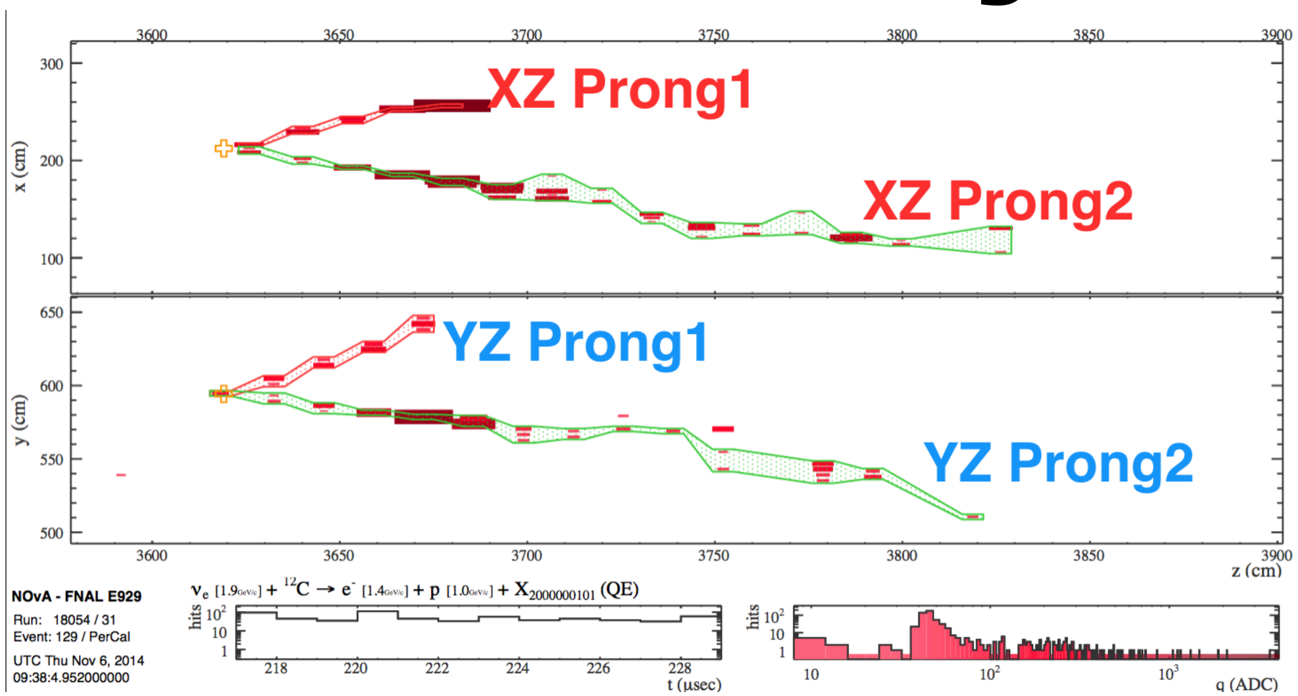


Traditional Reconstruction

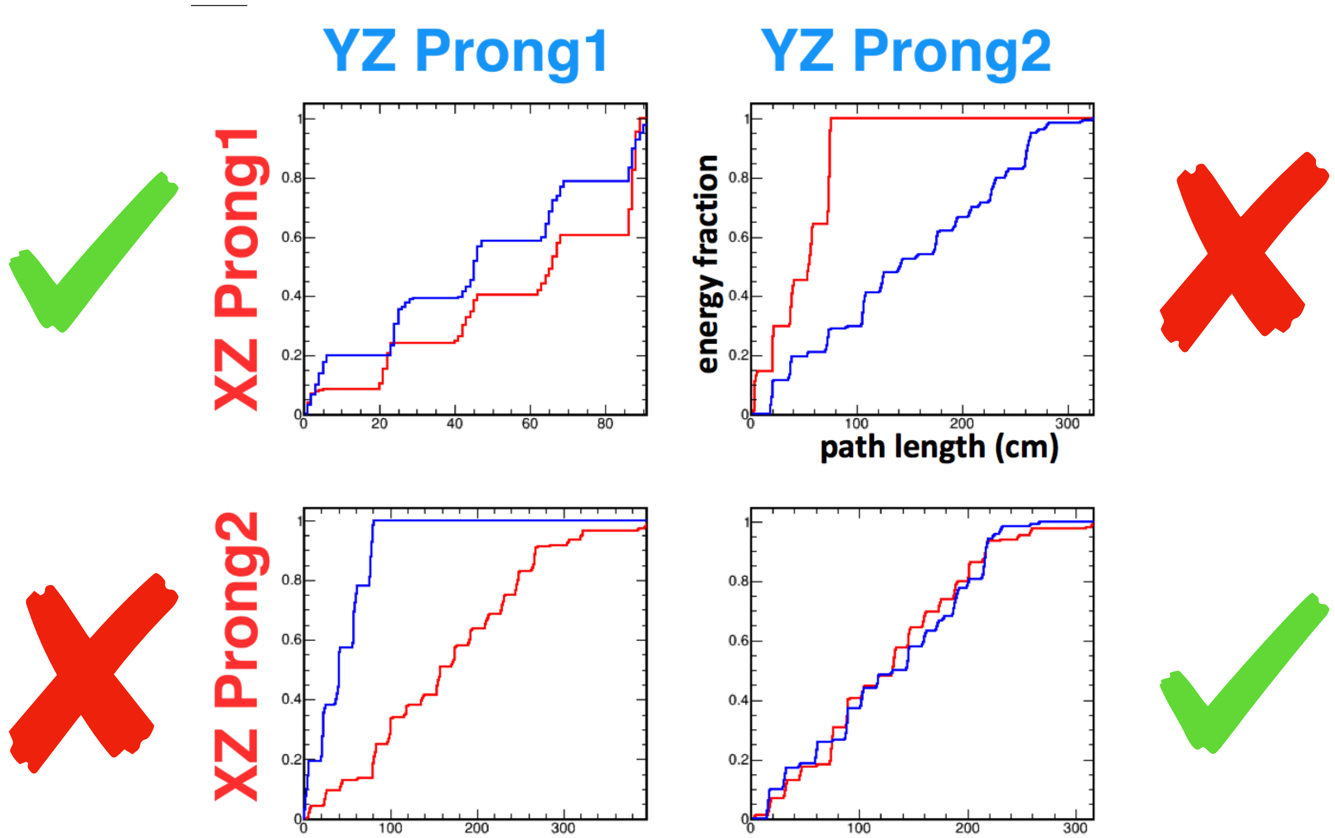
Use Fuzzy K-Means Clustering for individual particles, creating prongs



Fuzzy K Means



- Create **Prongs**:
  -In each of the two detector views, compute distance between hit and center of existing clusters
  -Add hit if overall distance is minimized



- 3D Matching:
  -Look cumulative energy as a function of path length along prong
  -Match is based on Kuiper's Test

Single Particle Identification

- Single particle identification (μ, e, p^+,$\pi^{+/-}$,γ) is important for more in-depth physics, including (but not limited to):
  ⇒ Better, more robust energy reconstruction

  ⇒ Enabling cross-section measurements of exclusive final states

- A convolutional neural network (CNN) could classify the particles

- Current architecture is based on GoogLeNet and uses a 4 tower siamese structure that uses both the prong and event views (the "context")

- Soon to be published in PRD:

Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events

FERMILAB-PUB-19-258-PPD

June 2019

NOvA Publications Page

Particle Signatures

μ

Long and straight, consistent dE/dx

e

Shower, usually associated with hadronic activity

γ

Shower, usually associated with pion, and produced in pairs from π^0

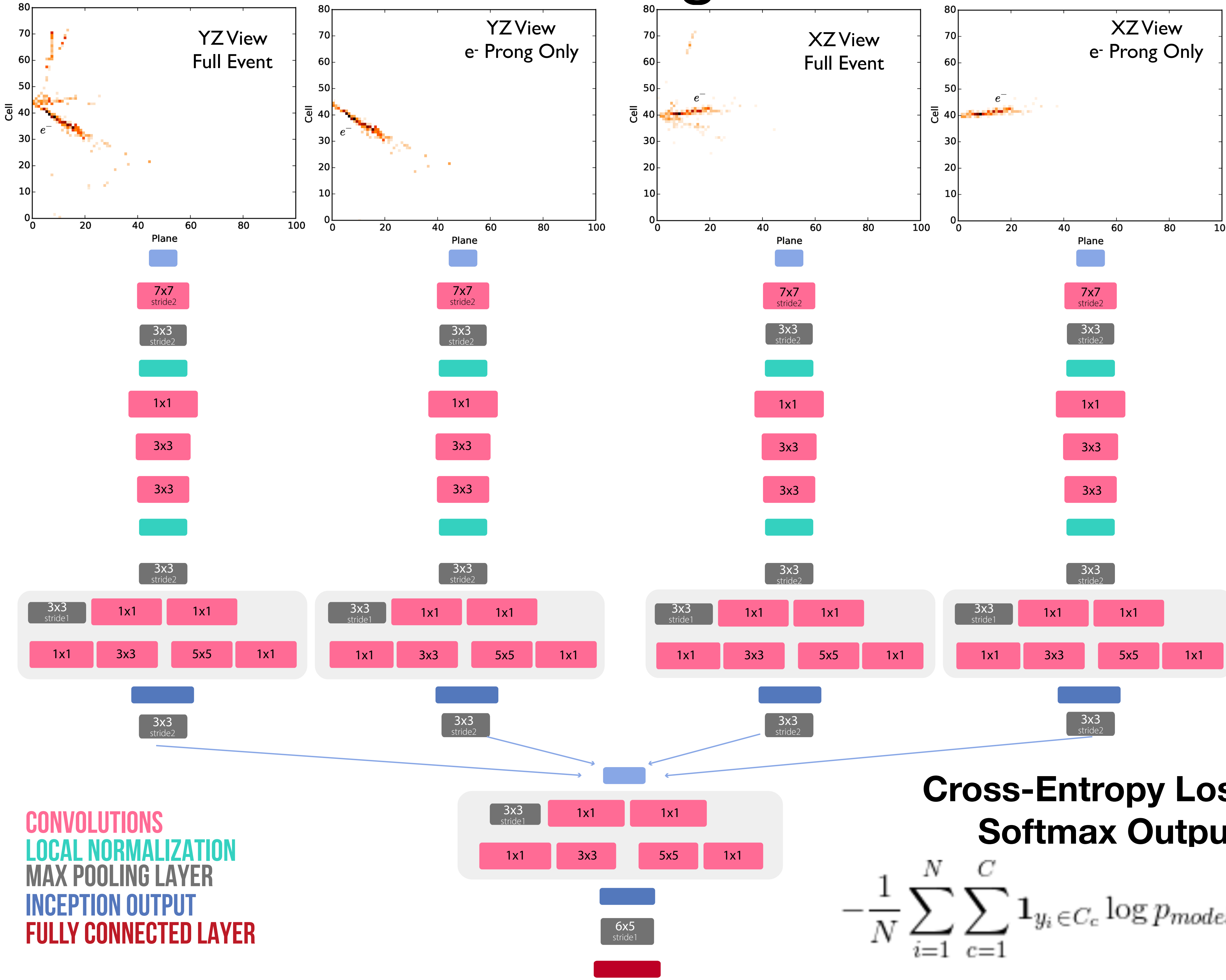
p

Generally short track with large energy deposit at end

$\pi^{+/-}$

Generally short track with consistent dE/dx

Context-Enriched Prong CNN Architecture

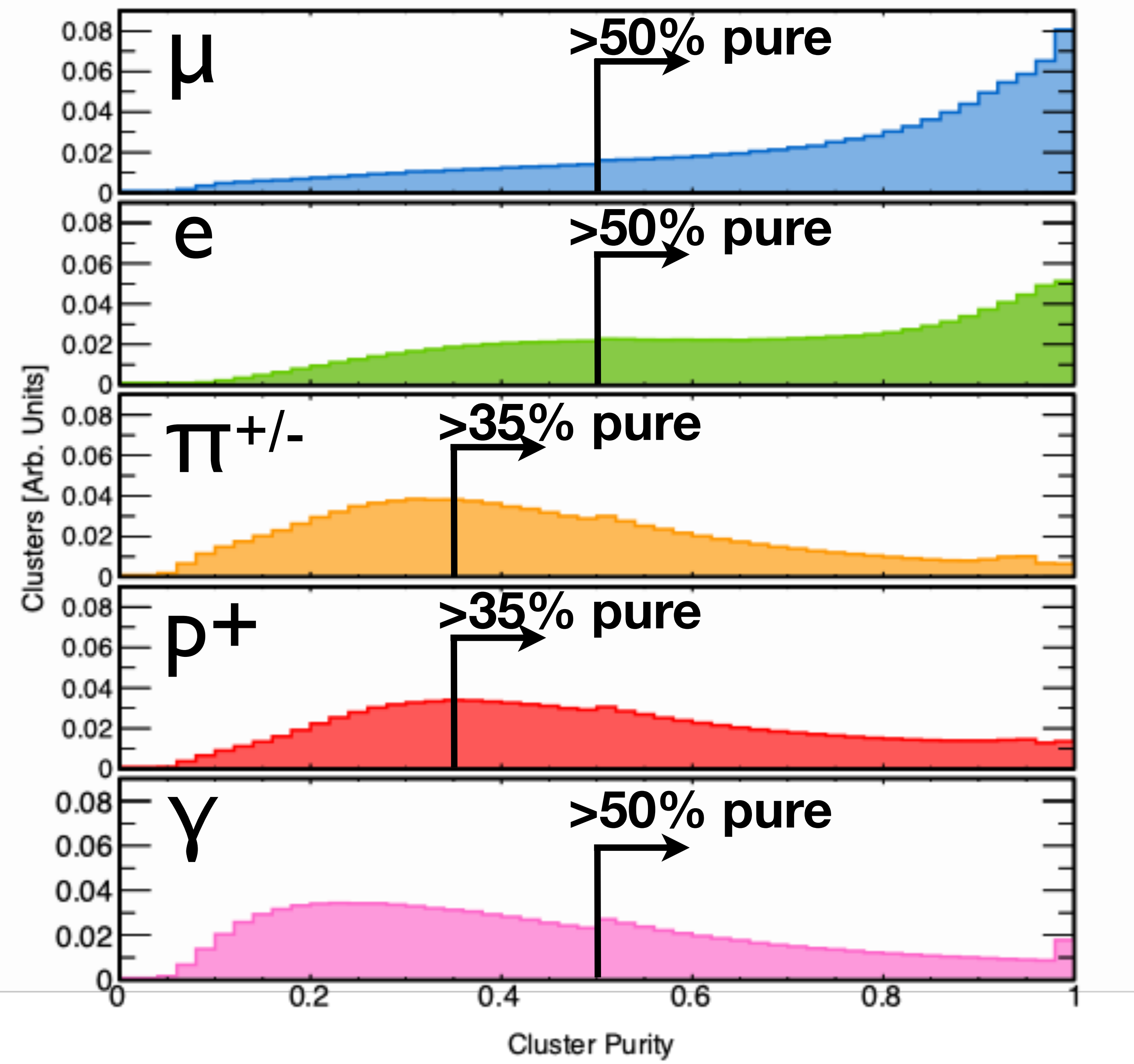


Cross-Entropy Loss w/ Softmax Output:

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\mathbf{1}_{y_i \in C_c}\log p_{model}[y_i \in C_c]$$

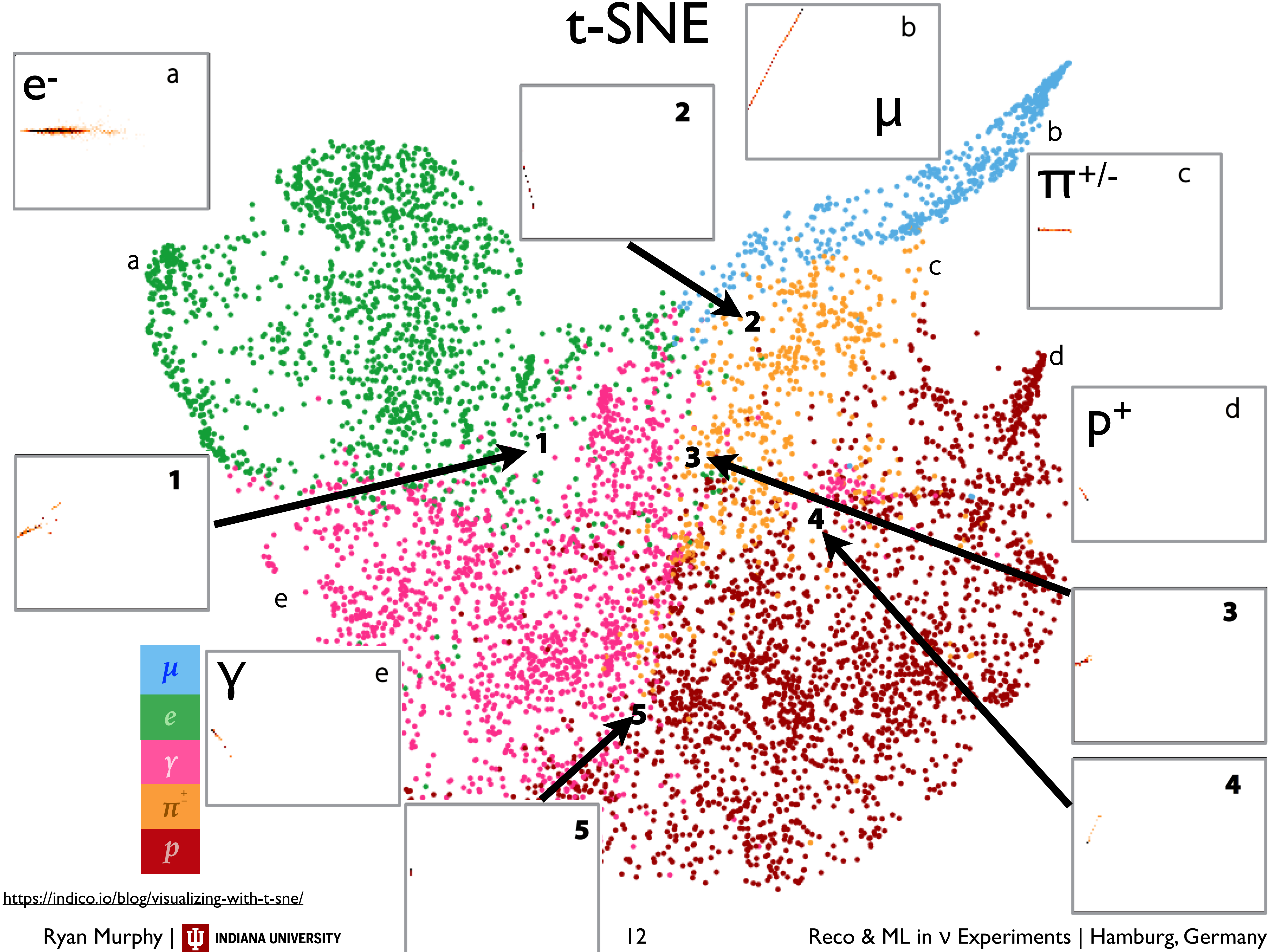
Training Dataset

- Training dataset includes prongs and full interactions
- Total dataset size is 2.95 million events
- Labels are based on which particle deposited the most energy into the prong
- A containment cut is applied to ensure we only train on fully contained events
- 5 meter cut is applied to reduce image size and increase network stability. >95% of all prongs over this length are muons.
- We also applied a selection criteria on the purity of the 3D prong to balance a representative sample of realistic input data with clear identities of the prong



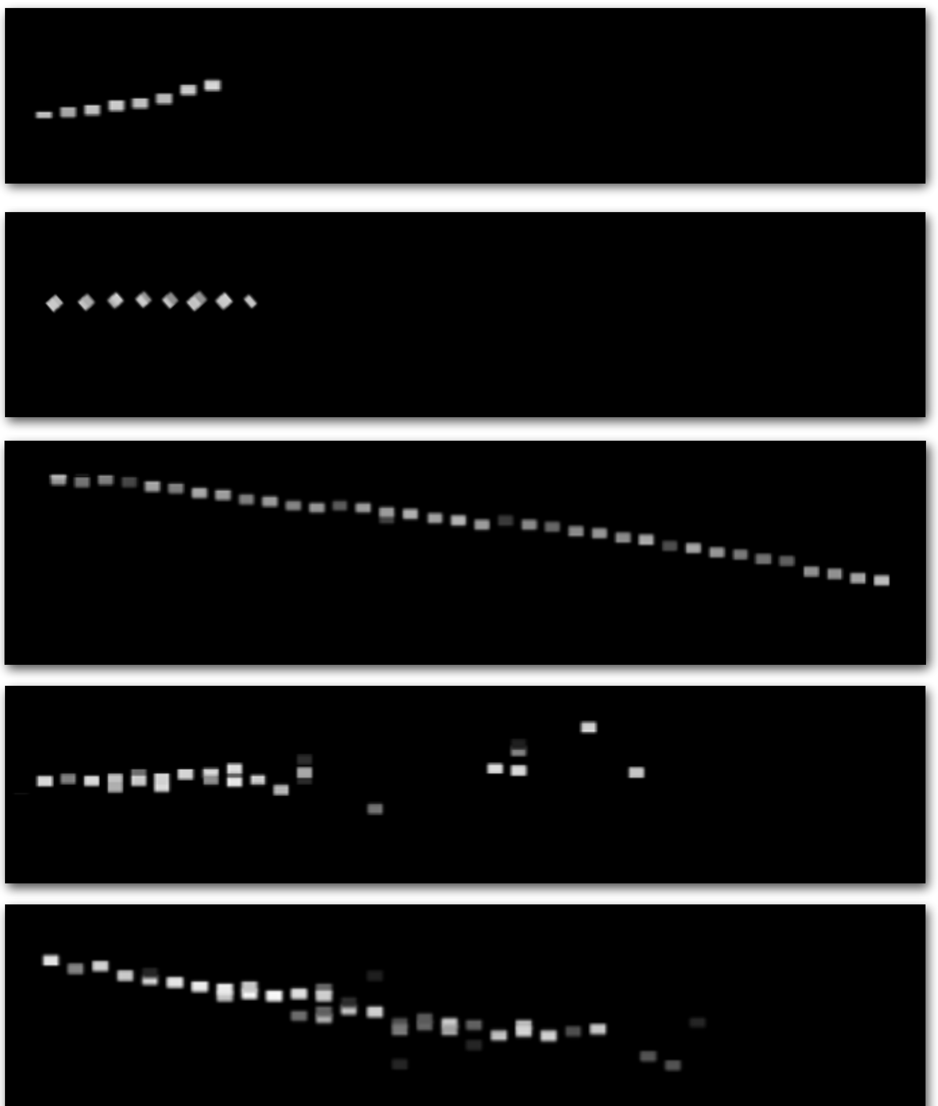
Purity - The fraction of the energy contained in a cluster which comes from the particle it is associated with

t-SNE



https://indico.io/blog/visualizing-with-t-sne/

Overall Performance



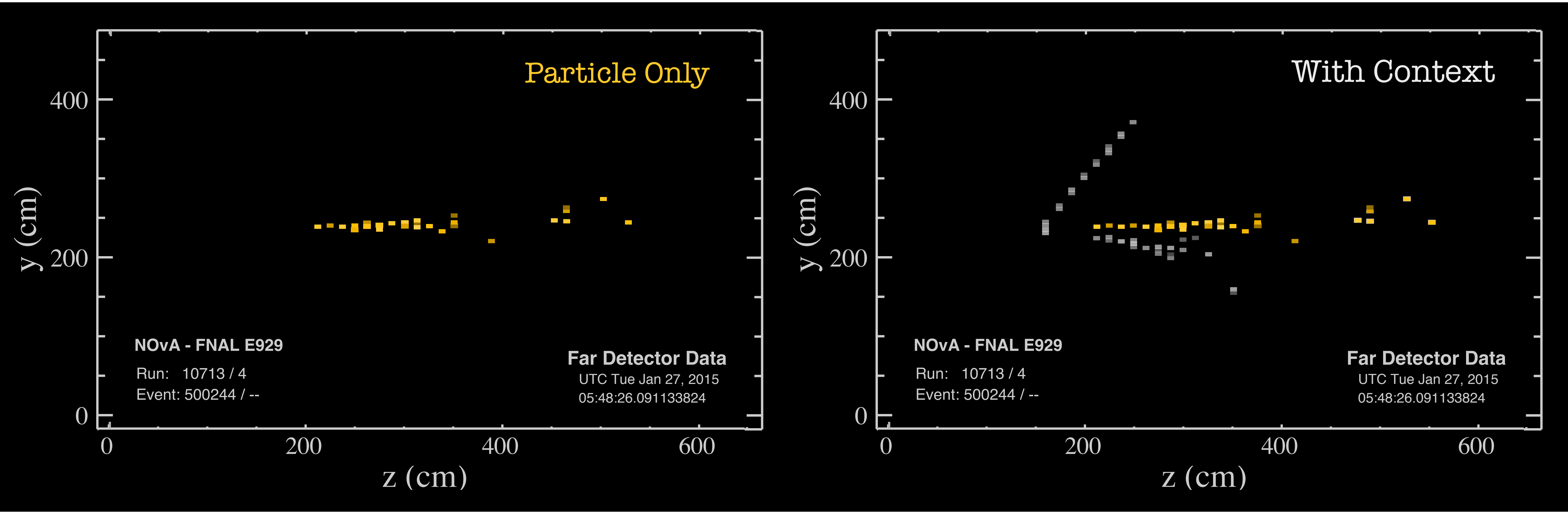
Color is Efficiency

Selected \ True	Electron	Gamma	Muon	Pion	Proton
Proton	0.03	0.12	0.03	0.28	0.89
Pion	0.01	0.03	0.06	0.54	0.04
Muon	0.00	0.00	0.87	0.05	0.00
Gamma	0.06	0.75	0.02	0.11	0.06
Electron	0.90	0.09	0.01	0.02	0.01

The diagonal
shows the efficiency for each category and the off-diagonal
shows how events are misidentified.

Selected = Particle with highest PID score

What does context contribute?



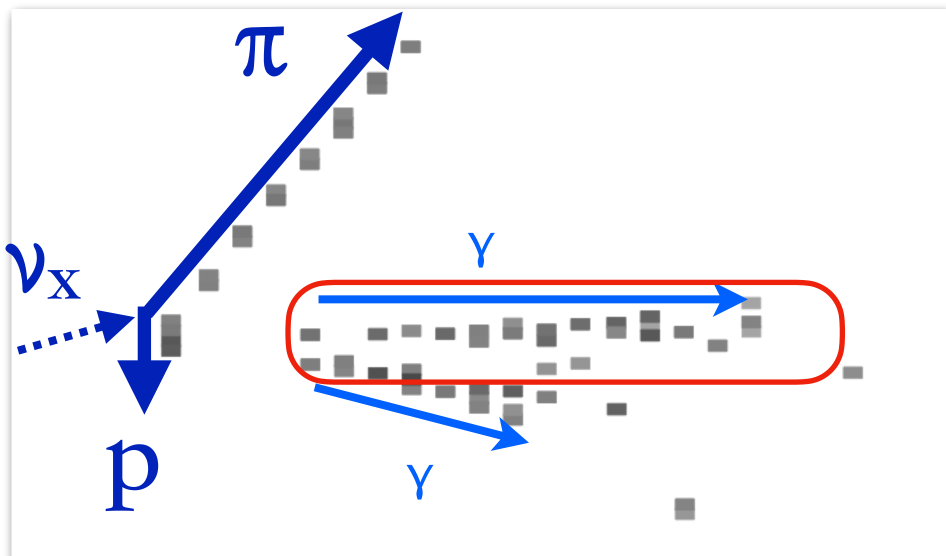
- Our hypothesis is that providing context improves our network's accuracy and purity
- To compare, train a network with prong views as the only input, reducing 4 towers to 2
- Use the exact same dataset, same hyperparameters
- Comparing the models will give us insight into what context improves on

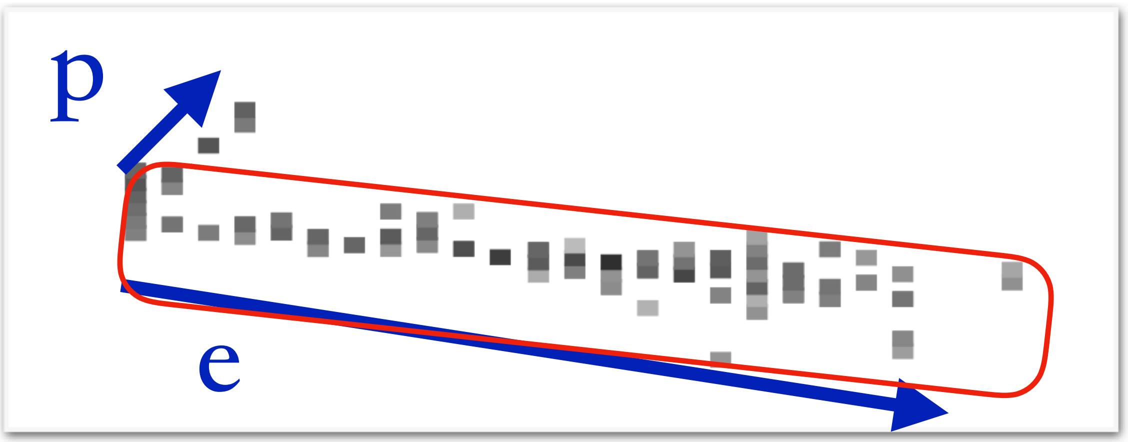
Context provides clarity

Electron or Gamma?

Context provides clarity

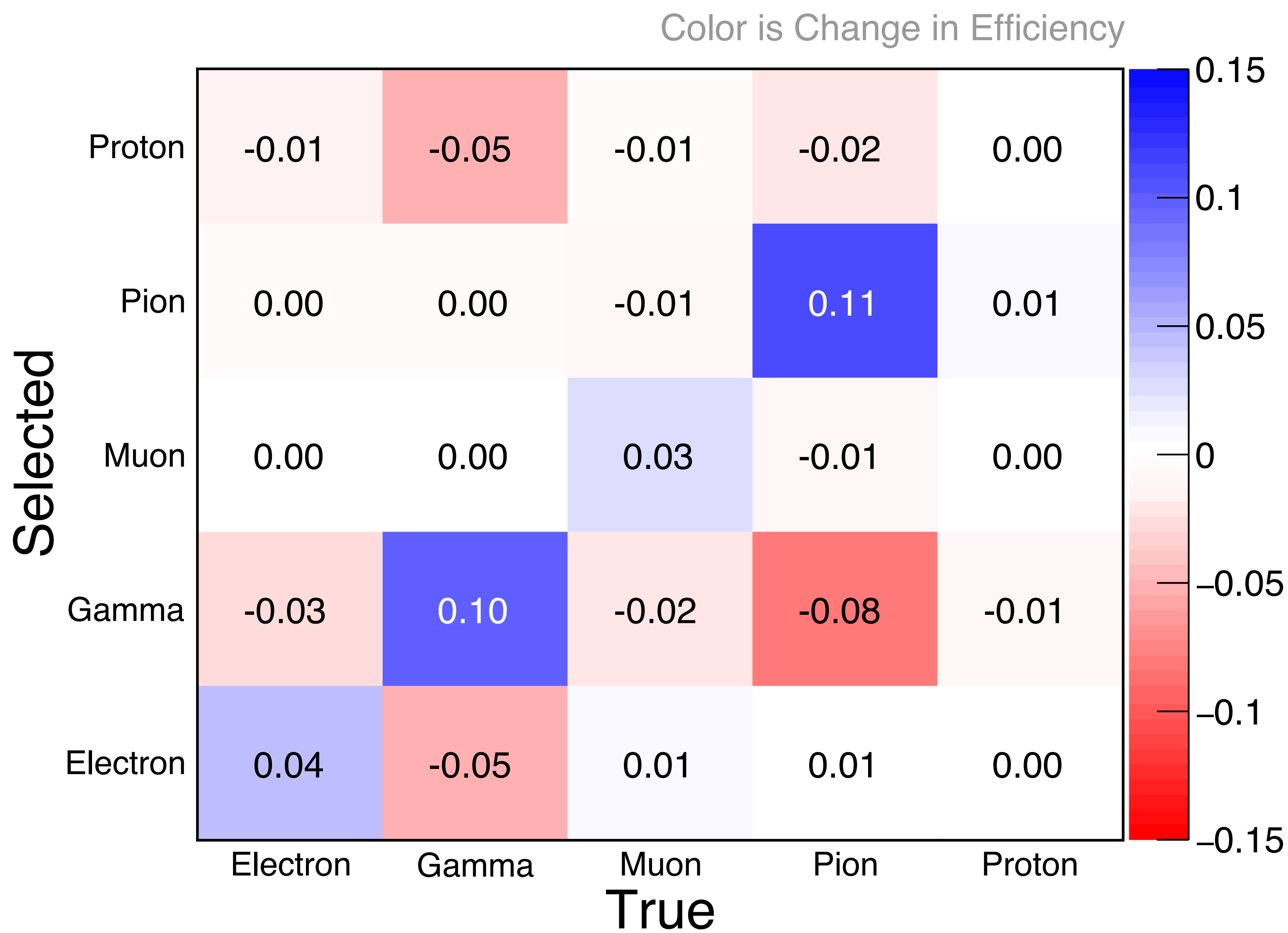
Electron or Gamma?

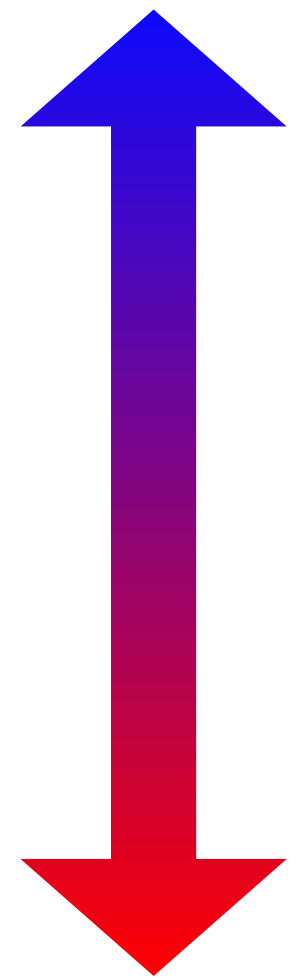




- Gamma is much more easily identifiable in this example with the context of another photon and pion.
- Gap between vertex and photons also helps
- Prongs can also share hits and overlap each other, which may make it indistinguishable without the context information

Context-Enrichment Improvement

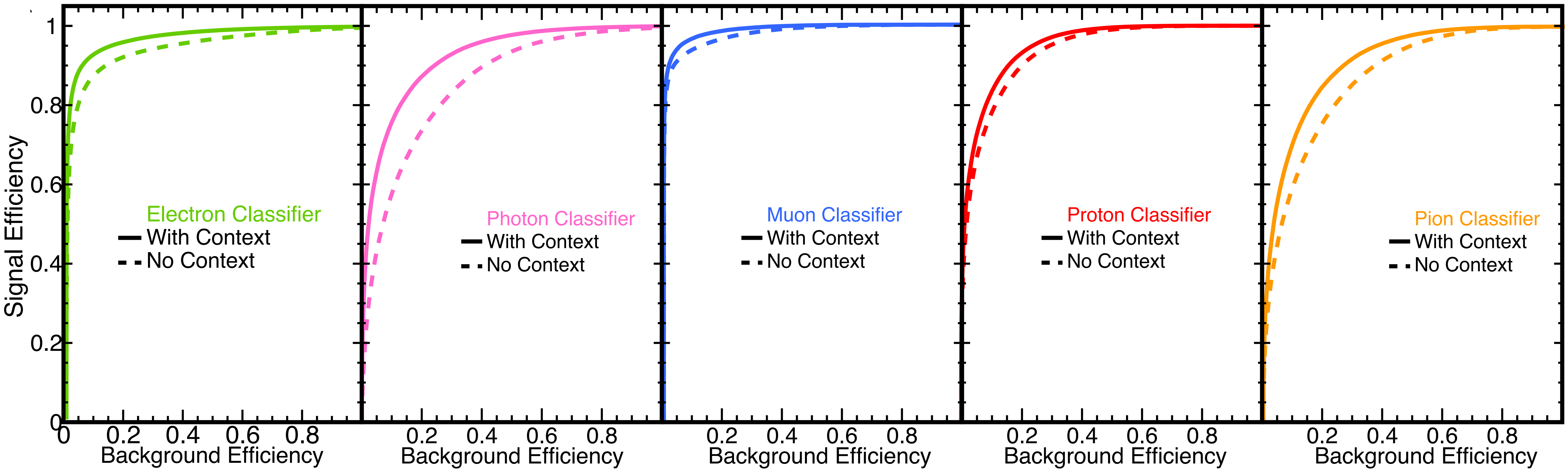




Context-Enriched Minus Particle Only

- Context increases accuracy for almost every label, but especially improves on non-leptonic labels

Context-Enriched Improvement

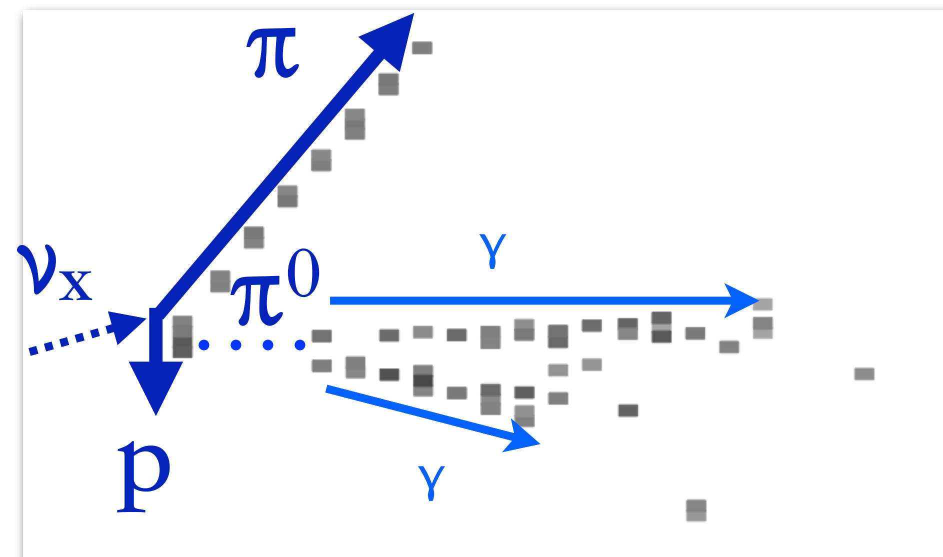


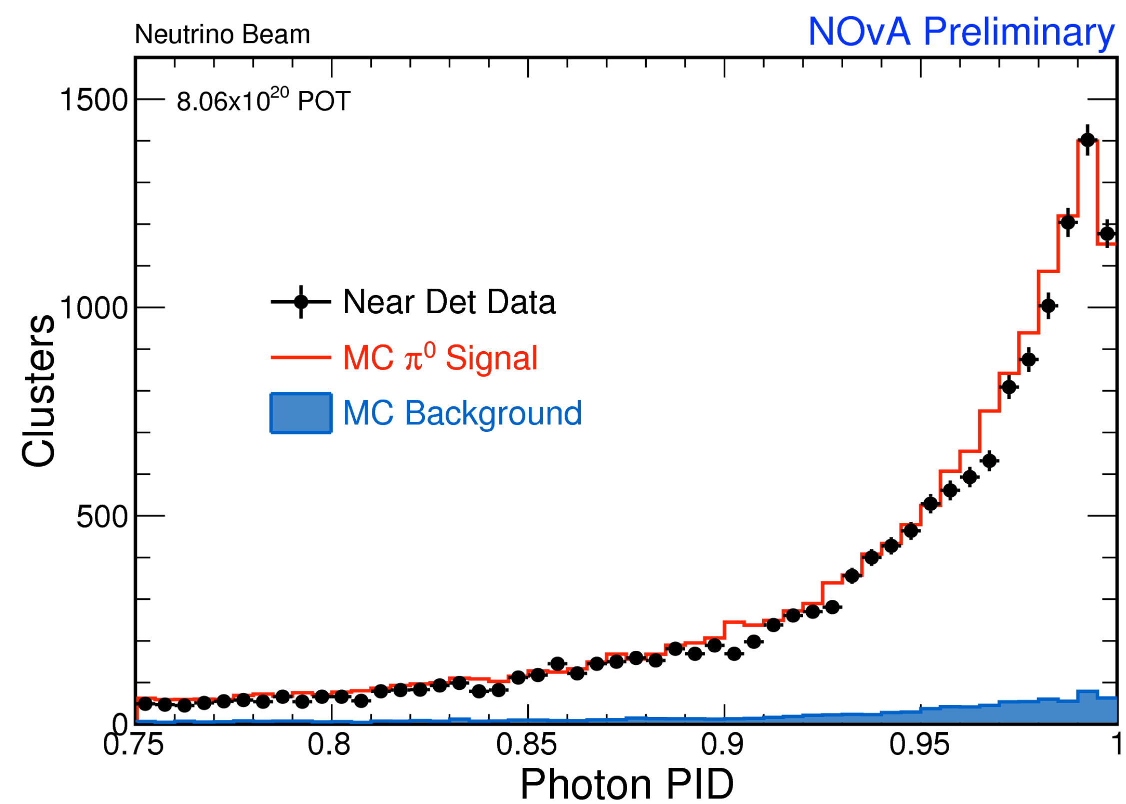
Comparison Metric	Network	Electron	Photon	Muon	Pion	Proton
Background Efficiency for 90% Signal Efficiency	Particle & Context	3.2%	14.5%	1.1%	16.1%	9.4%
	Particle Only	8.0%	24.6%	2.2%	22.4%	12.1%
ROC Integral	Particle & Context	0.983	0.951	0.992	0.944	0.969
	Particle Only	0.967	0.910	0.986	0.920	0.960
Largest Score Selection Efficiency	Particle & Context	90%	75%	87%	54%	89%
	Particle Only	86%	74%	84%	43%	89%
Largest Score Selection Purity	Particle & Context	93%	75%	93%	65%	81%
	Particle Only	90%	64%	92%	60%	77%

- Context increases accuracy and purity for almost every label, but especially improves on non-lepton labels

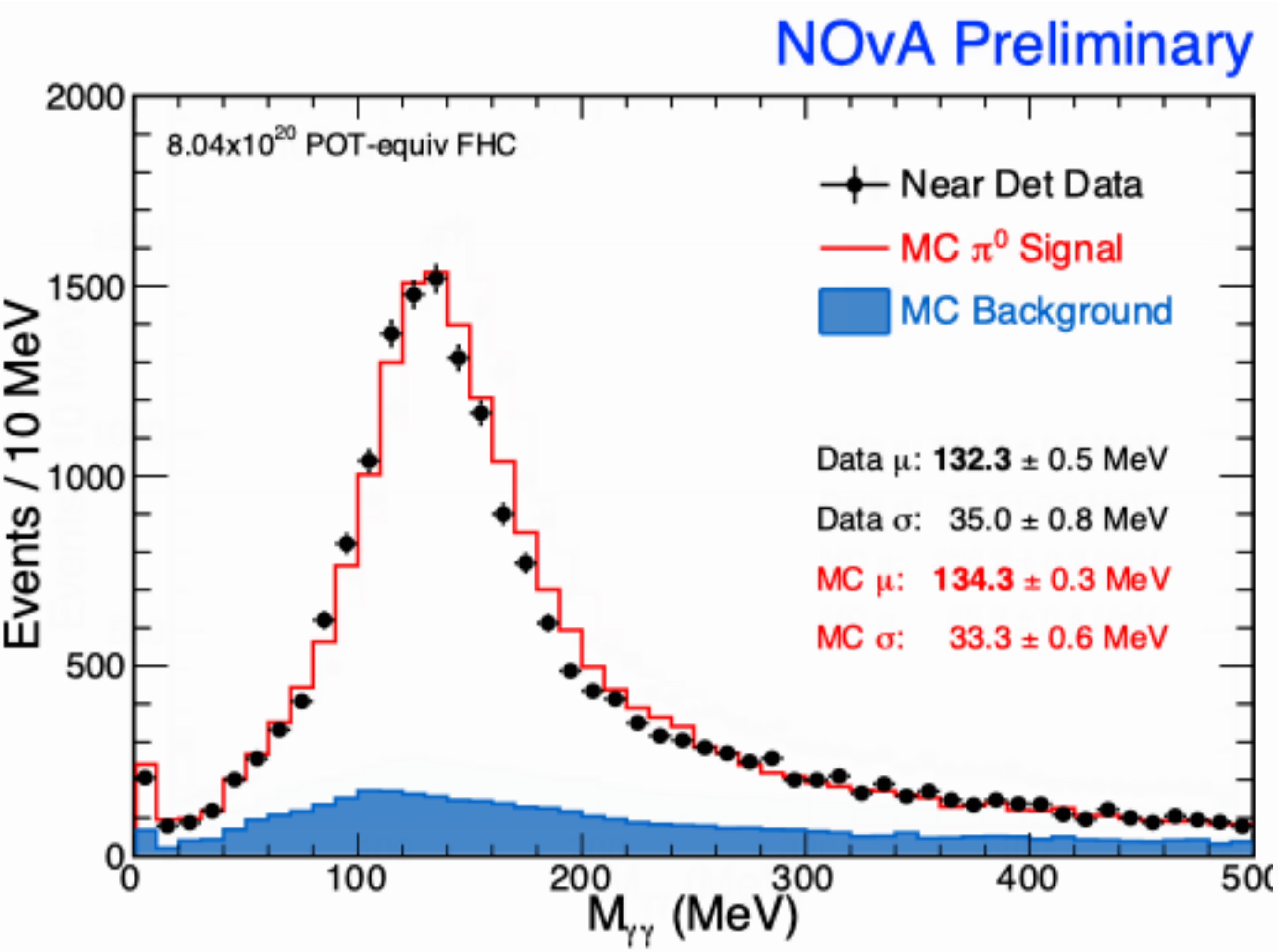
Use case #1: π^0 Mass Peak

- **Data-driven** method to gauge the energy response of our detectors.
- Look for $\pi^0 \rightarrow 2\gamma$
- Can compare old method that uses traditional reconstruction methods and new method that uses Prong CNN (γ pid > 0.75).
- Using Prong CNN lets us decrease backgrounds by 60% at the same efficiency.

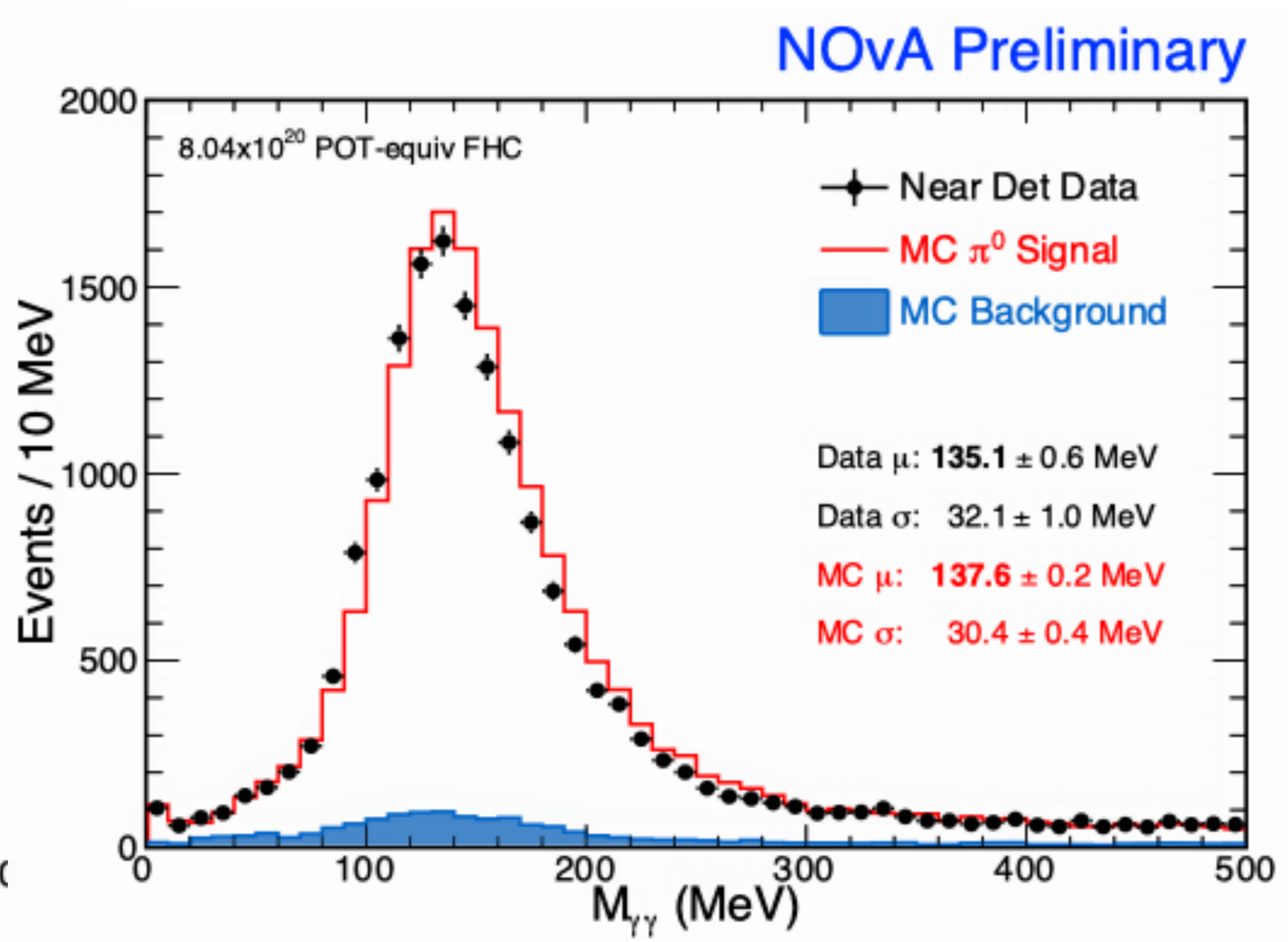




Use case #1: π^0 Mass Peak



Old Method

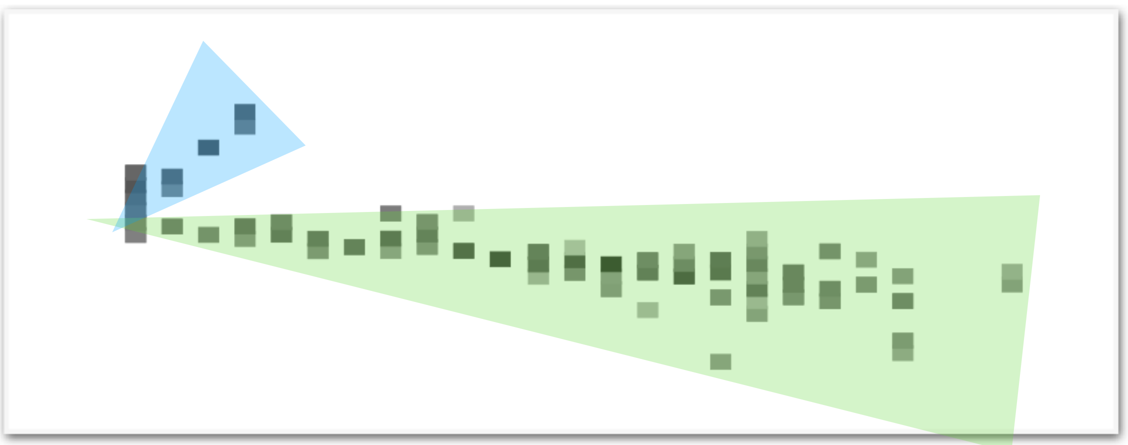


New Method with Prong CVN

π^0 mass ~ 134.96 MeV

Use case #2: ν_e Energy Estimator

- Used in selection of EM-like prongs for energy estimator.



EM Score = Electron ID + Photon ID
Hadronic Score = 1 - EM Score

EM-like prong if:
EM Score > Hadronic Score

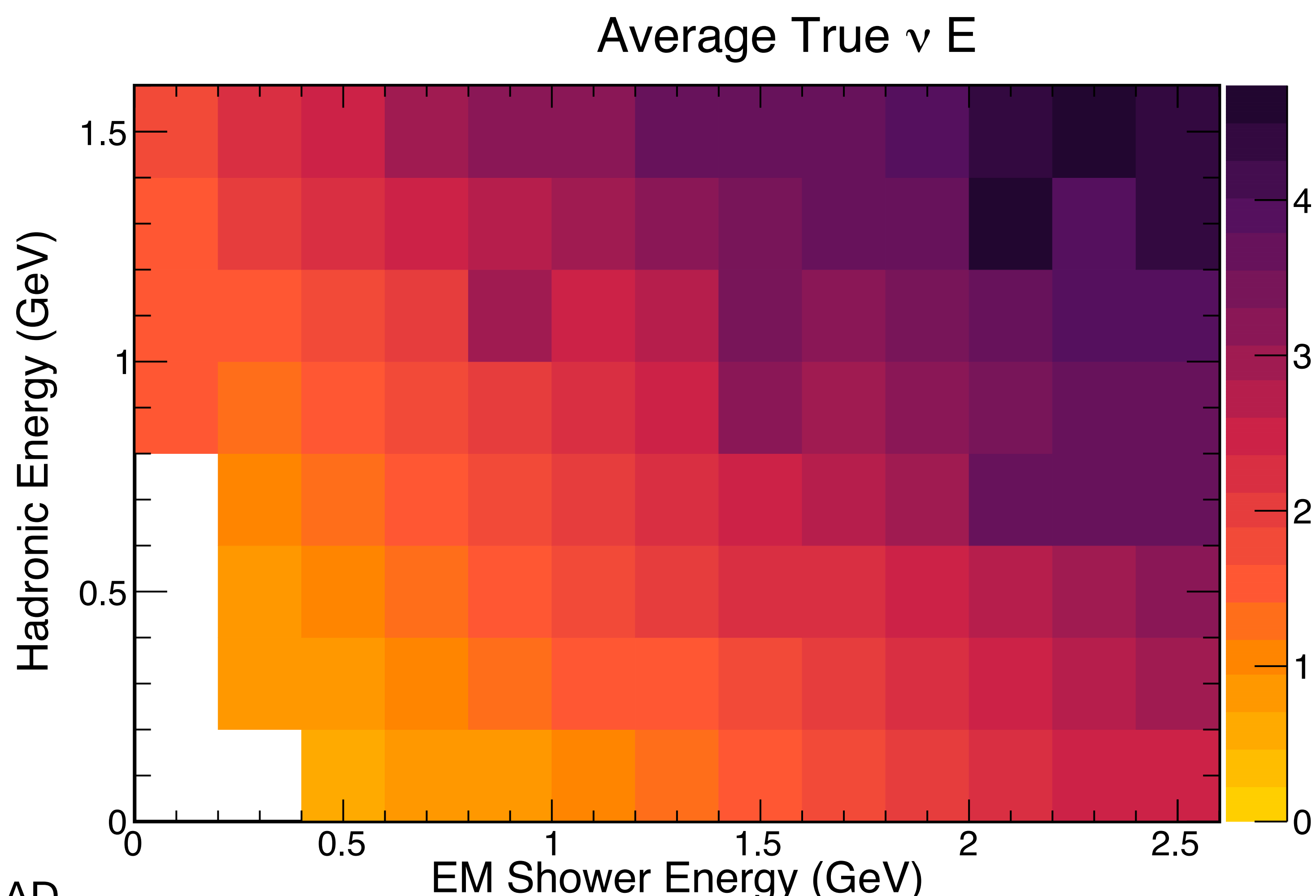
EM Shower Energy:
Sum together energy of all EM-like prongs

Hadronic Shower Energy:
The difference between total event energy and the EM-like prong's total energy
Hadronic Energy = Total Energy - EM-like Energy

Use case #2: ν_e Energy Estimator

- Each bin is filled in by the true average energy
- A quadratic fit is used to estimate the ν_e energy:

$$E_{\nu e} = A*E_{EM} + B*E_{HAD} + C*E^2_{EM} + D*E^2_{HAD}$$



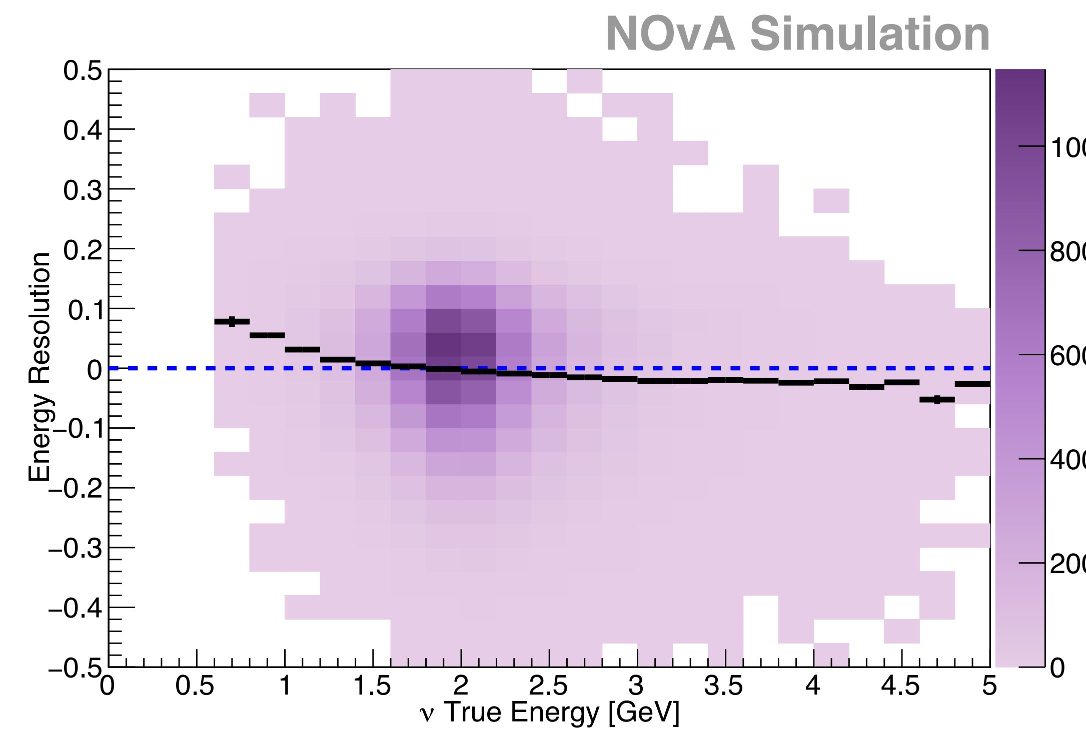
Use case #2: ν_e Energy Estimator

- Each bin is filled in by the true average energy
- A quadratic fit is used to estimate the ν_e energy:

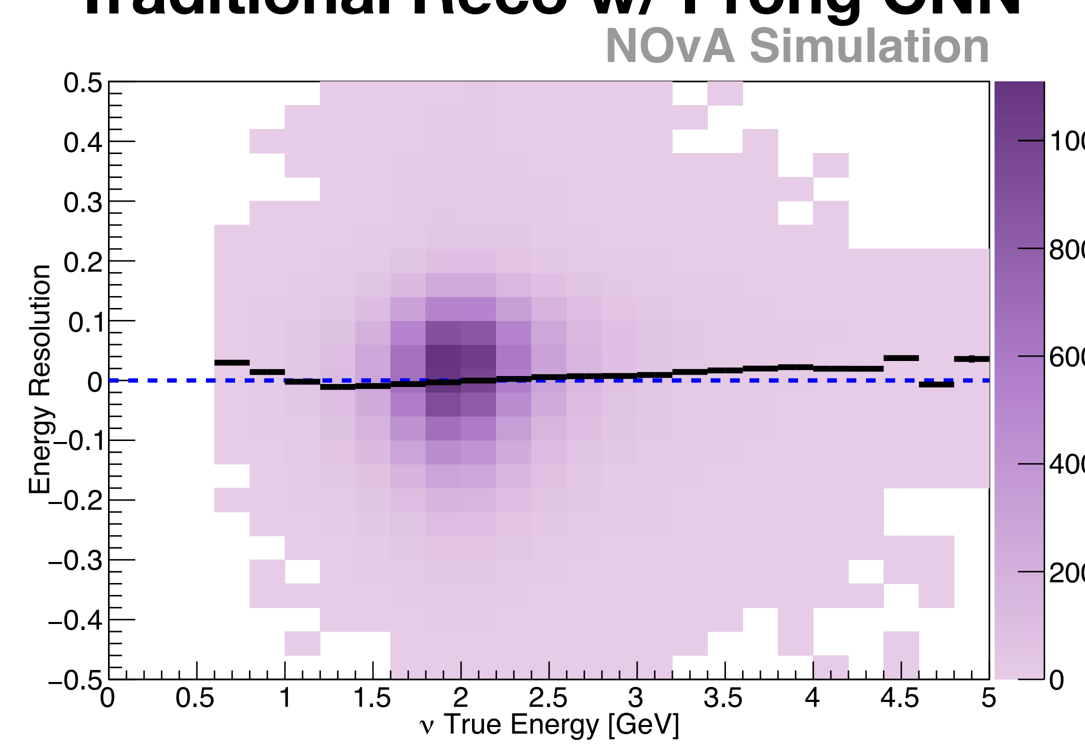
$$E_{\nu e} = A*E_{EM} + B*E_{HAD} + C*E^2_{EM} + D*E^2_{HAD}$$

- Energy resolution is 11%, bias across energy is reduced

Traditional Reco

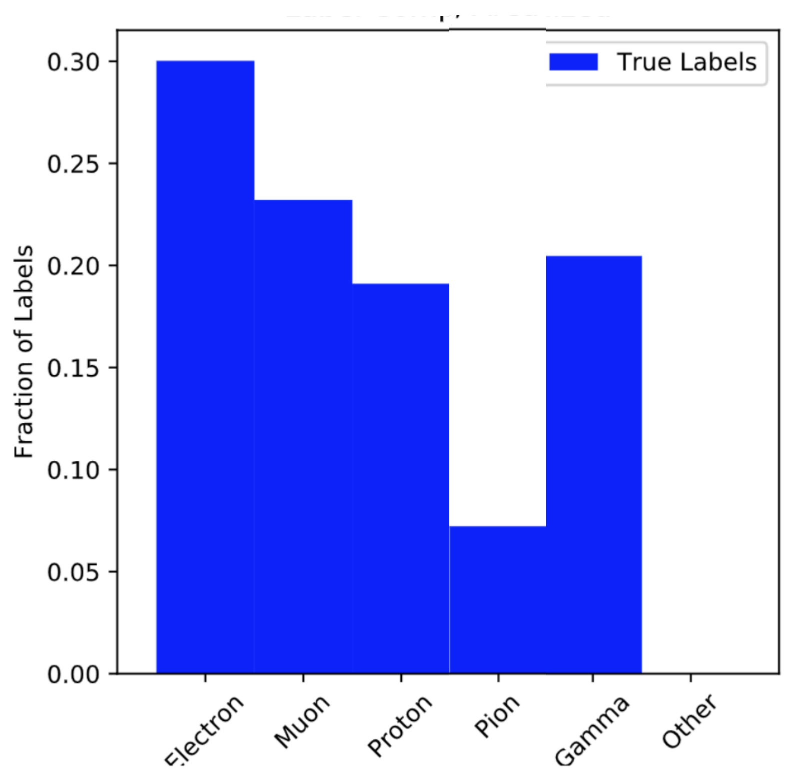


Traditional Reco w/ Prong CNN

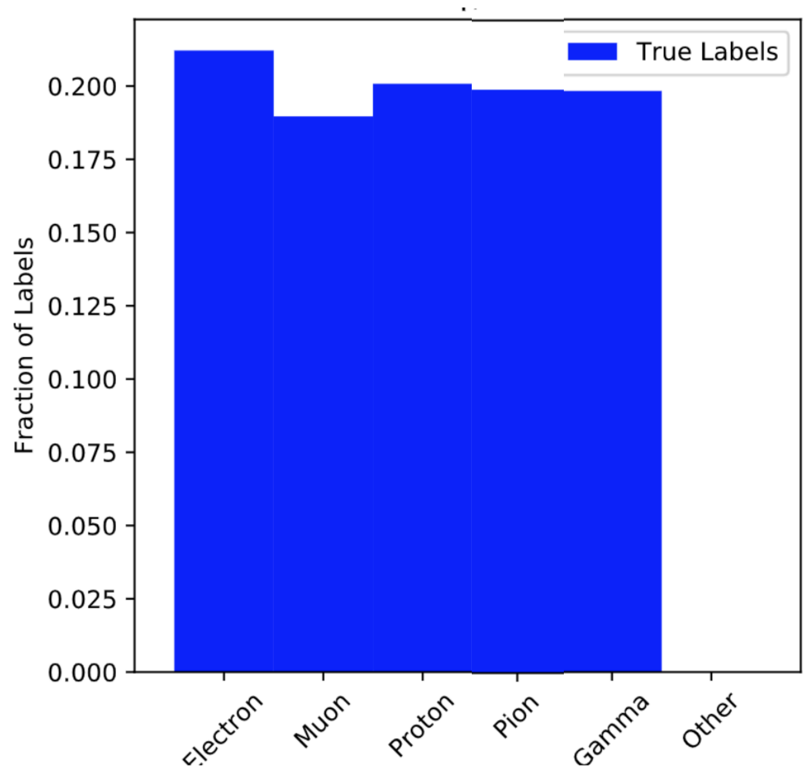


Improvement: Balanced Datasets

Unbalanced Dataset Composition



Balanced Dataset Composition



- Makes our neural networks focus equally on all particle types
- F1 score (harmonic mean of efficiency and purity) increased by ~2%

$$F_1 = 2 * \frac{Efficiency * Purity}{Efficiency + Purity}$$

Summary

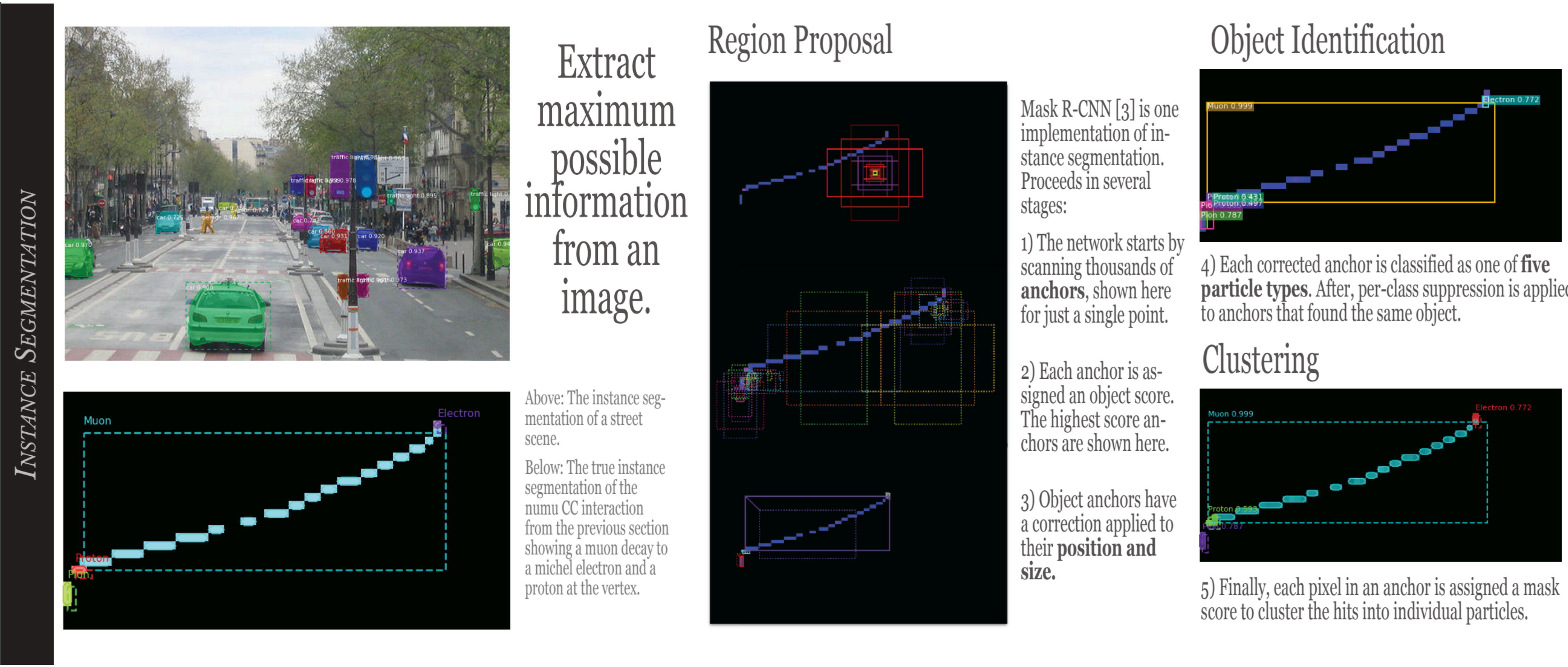
- Context adds significant improvement to our particle identification CNN
- A neural network based particle identifier improves NOvA's physics analysis capabilities through many channels
- NOvA continues to host a rich deep learning program with many more improvements in the pipeline

http://novaexperiment.fnal.gov


Thank you!
VIRGINIA
University of Sussex, UK
Summer 2019

Back ups

Instance segmentation

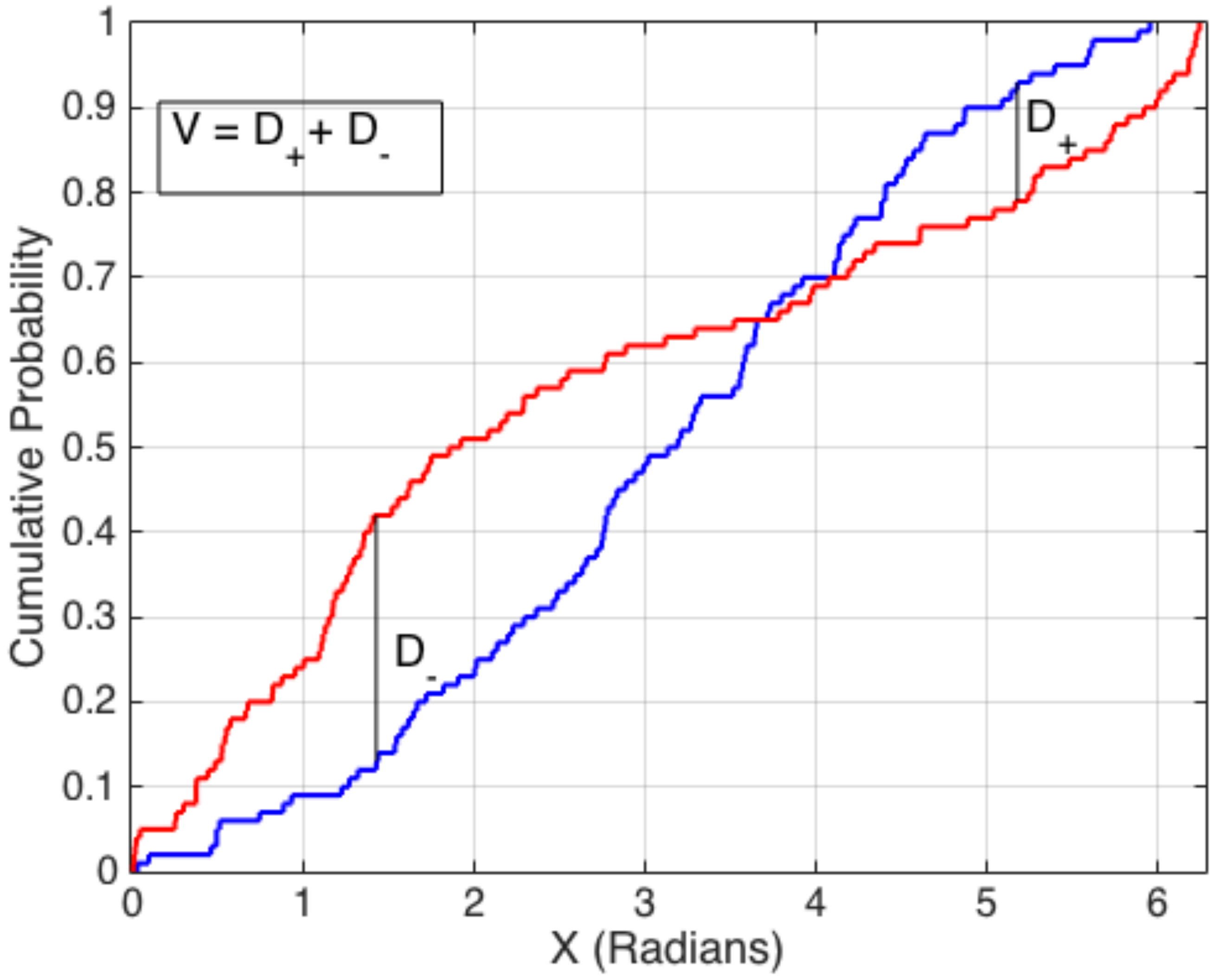


Taken from Micah G.'s

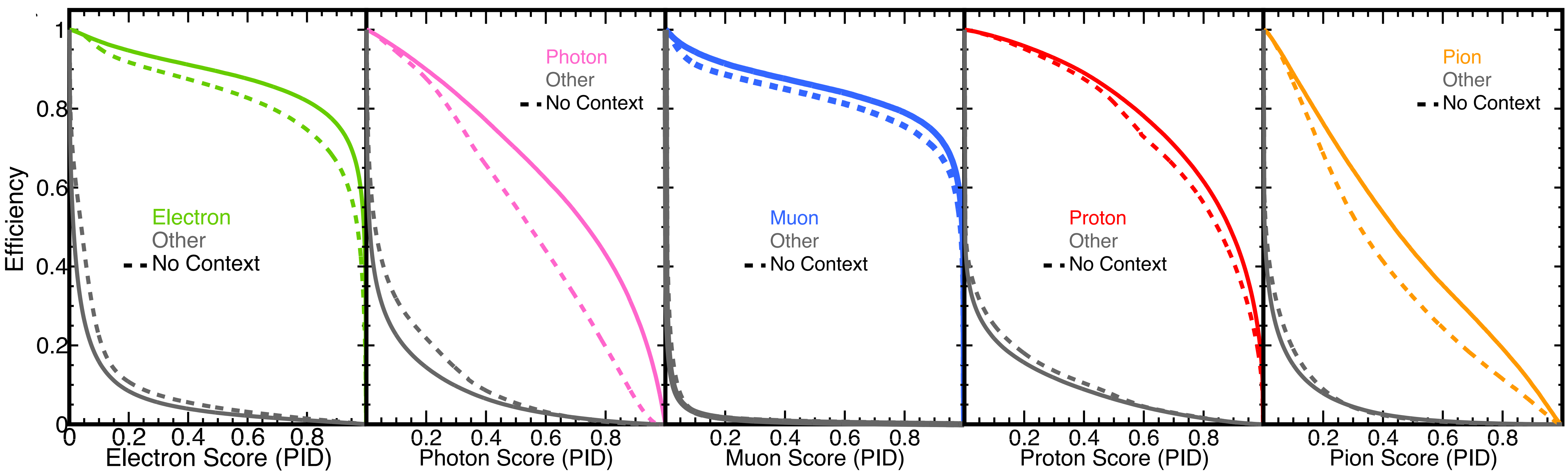
Training Dataset Full Details

- Training dataset includes prongs and full interactions
- Total dataset size is 2.95 million events
- Prongs use a Kuiper test to match XZ and YZ views. Any prong not matched is not used in training
- The spatial and temporal resolution of the detector, along with the inefficiencies of vertex finding and separating overlapping particles effect the quality, completeness, and purity of the prongs
- To make sure there is reasonable data in our training data, we apply a containment cut to ensure we only train on fully contained events
- 5 meter cut is applied to reduce image size and increase network stability. >95% of all prongs over this length are muons. Easily identifiable via traditional reconstruction methods
- We also applied a selection criteria on the purity of the 3D prong to balance a representative sample of realistic input data with clear identities of the prong
  Muons, photons, and electrons are cut at 0.5 and protons and pions are cut at 0.35

Kuiper's Test

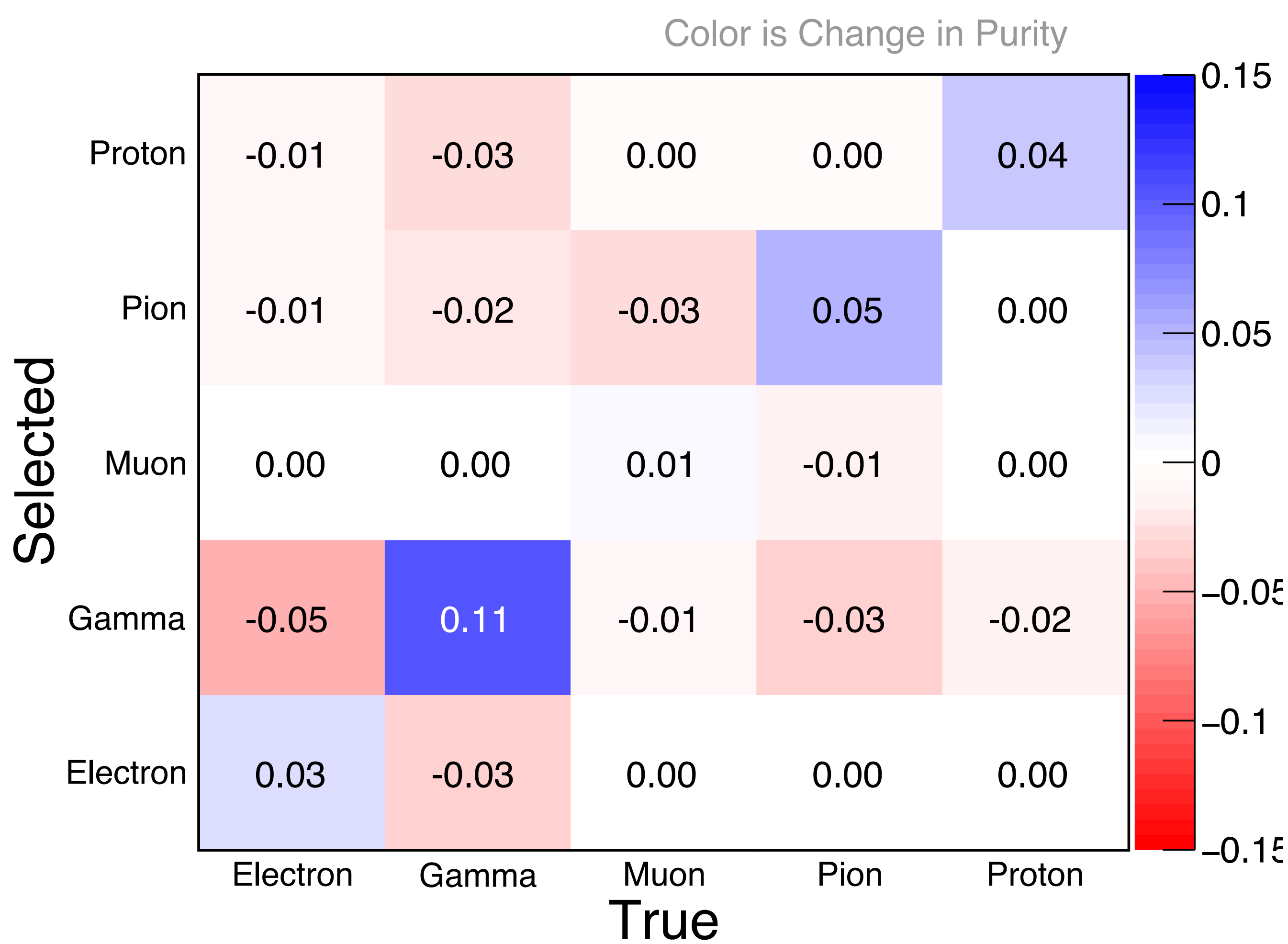


Efficiency vs PID Score

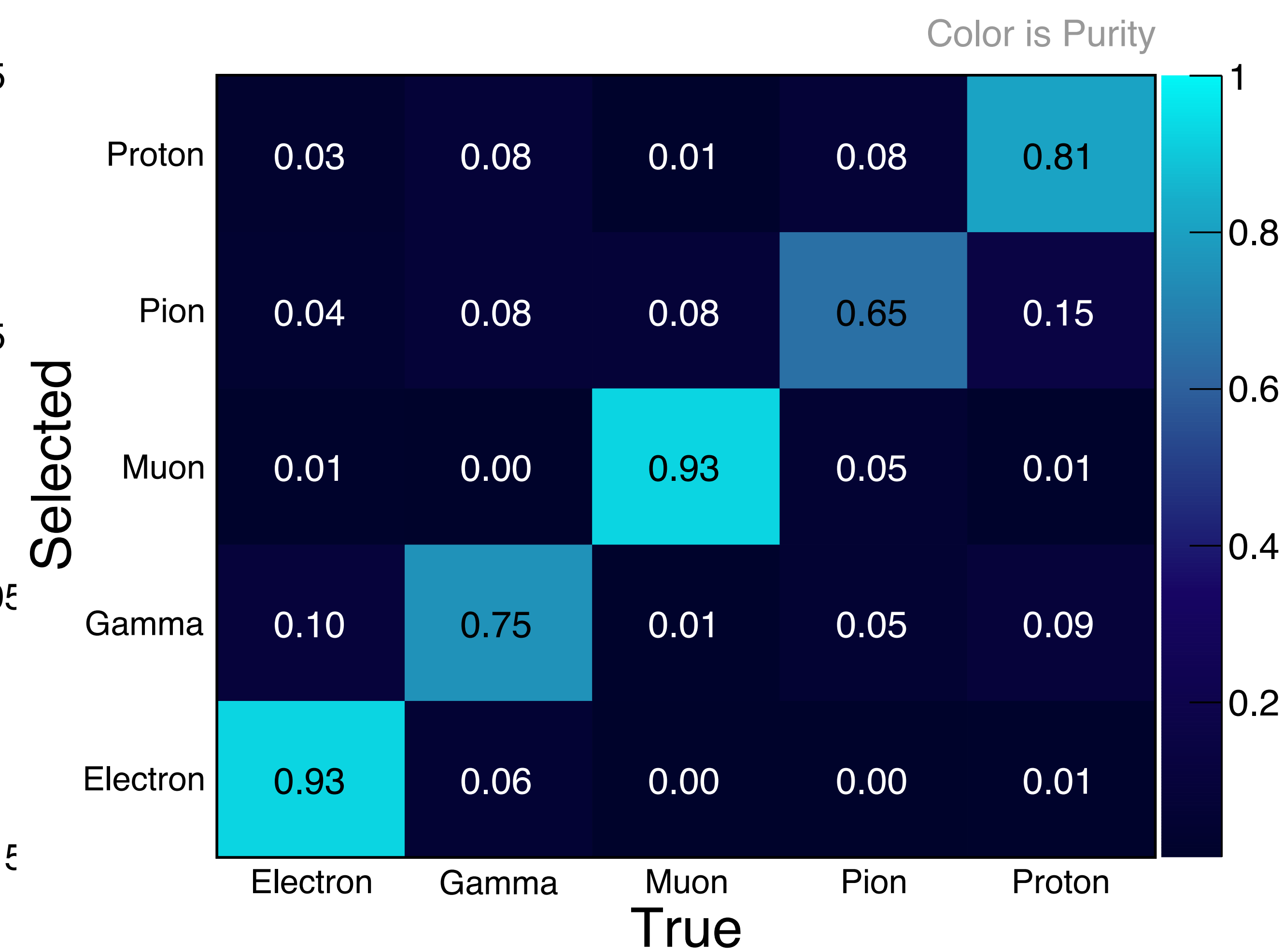


Purity Plots

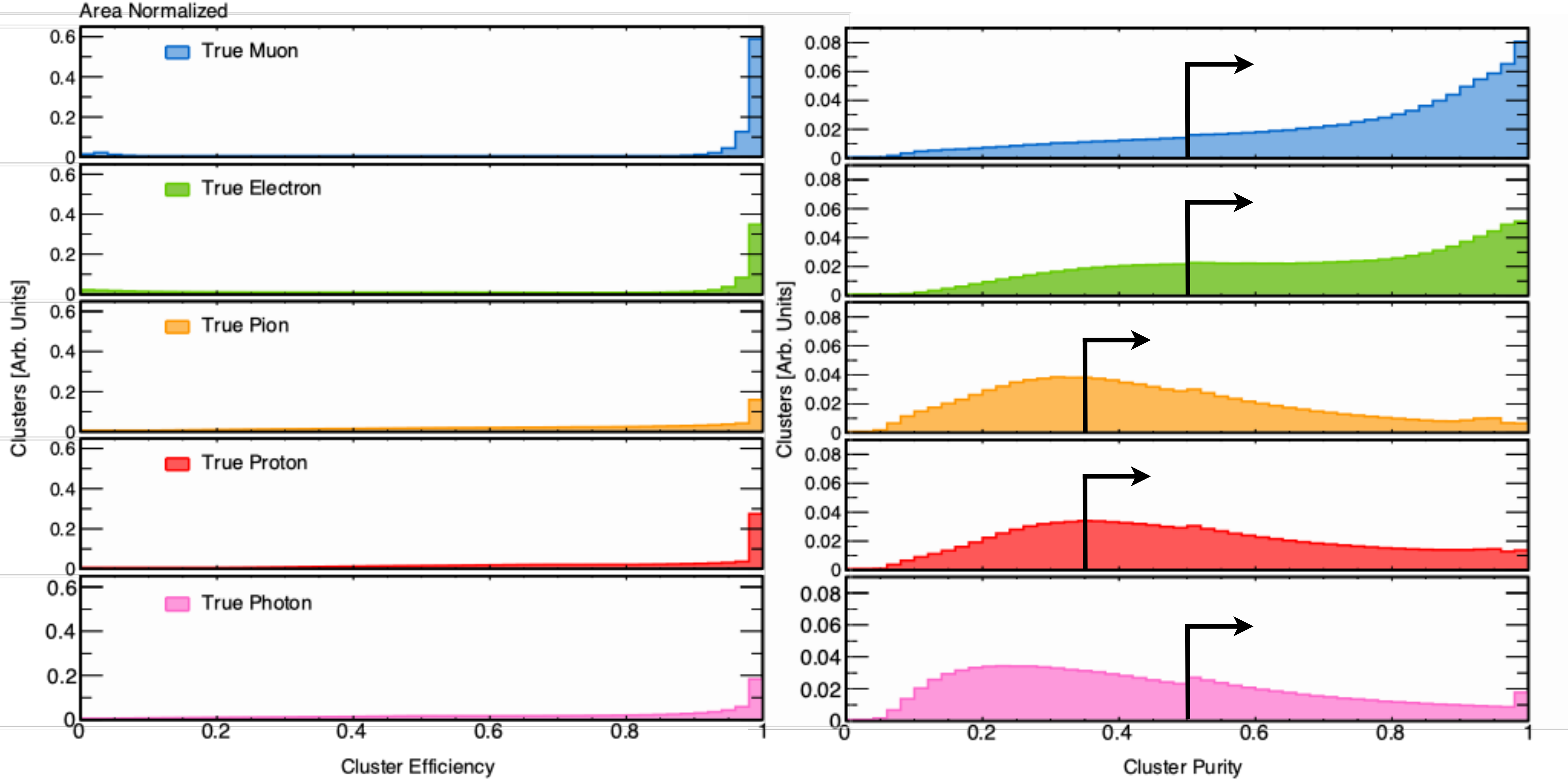
4 view - 2 view



4 View Purity Matrix



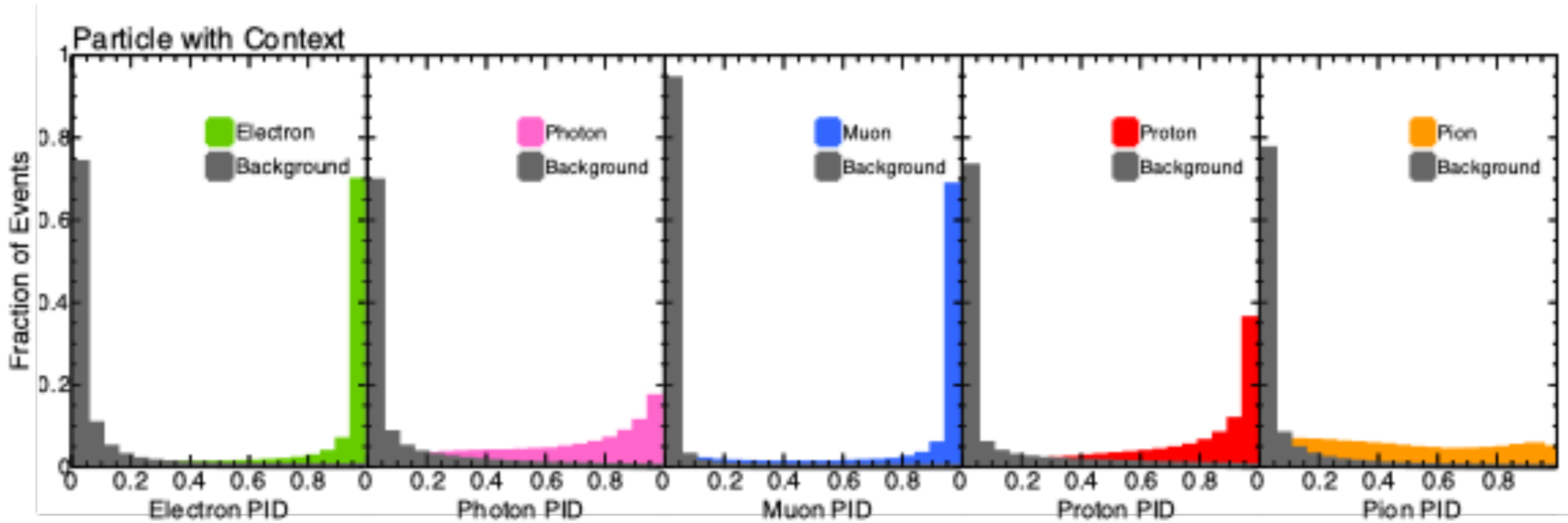
3D Prong Efficiency & Purity



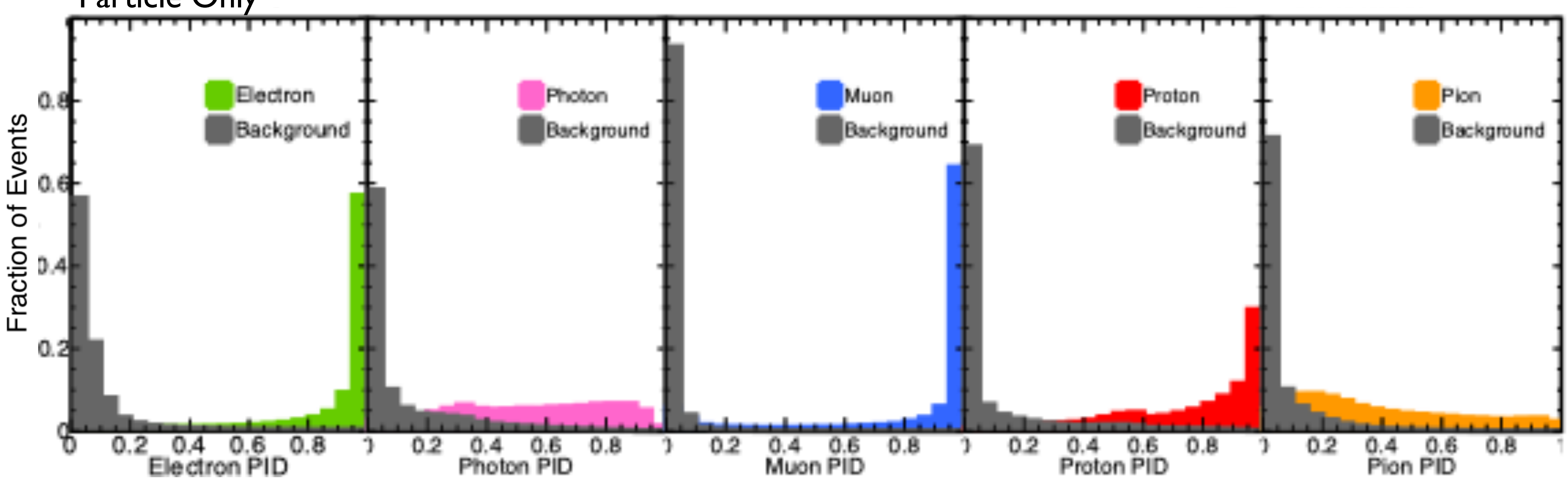
Efficiency - The fraction of energy depositions from the particle associated with a cluster which are contained by the cluster

Purity - The fraction of the energy contained in a cluster which comes from the particle it is associated with

Prong PID Distributions



Particle Only

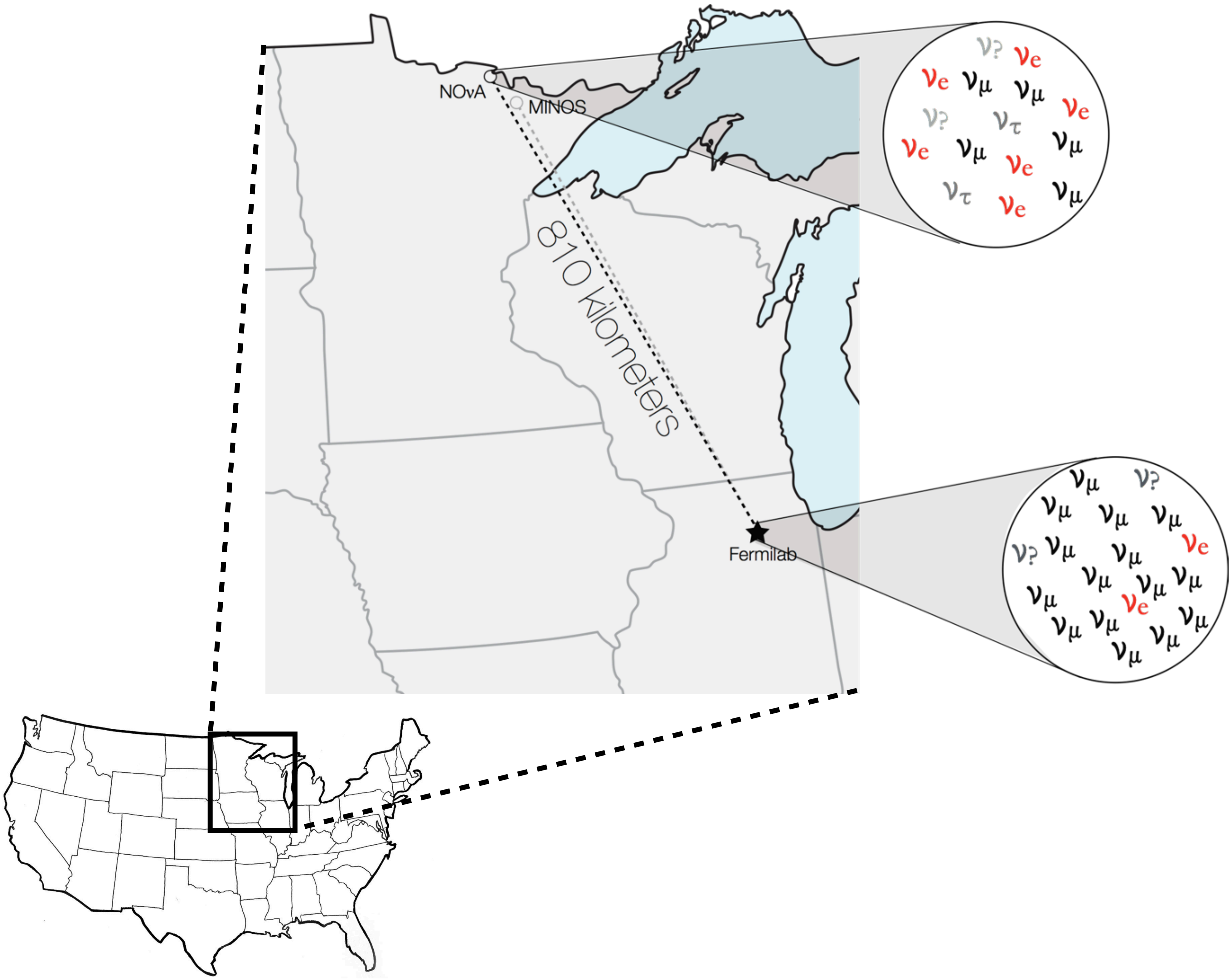


More Comparison Metrics

Comparison Metric	Network	Electron	Photon	Muon	Pion	Proton
Background Efficiency for 90% Signal Efficiency	Particle & Context	3.2%	14.5%	1.1%	16.1%	9.4%
	Particle Only	8.0%	24.6%	2.2%	22.4%	12.1%
ROC Integral	Particle & Context	0.983	0.951	0.992	0.944	0.969
	Particle Only	0.967	0.910	0.986	0.920	0.960
Largest Score Selection Efficiency	Particle & Context	90%	75%	87%	54%	89%
	Particle Only	86%	74%	84%	43%	89%
Largest Score Selection Purity	Particle & Context	93%	75%	93%	65%	81%
	Particle Only	90%	64%	92%	60%	77%

The NOvA Experiment

NuMI **O**ff-axis ν_e **A**ppearance



- NOvA is a long-baseline neutrino oscillation experiment
- Observes neutrinos from NuMI beamline at Fermilab
- Two functionally identical detectors, situated 14 mrad off axis, 810 km apart
- Near Detector is 300 tons, located at FNAL
- Far Detector is 14 ktons, located in Ash River, MN

NOνA Physics Program

Disappearance channel ($\nu_\mu \rightarrow \nu_\mu$ & $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$)

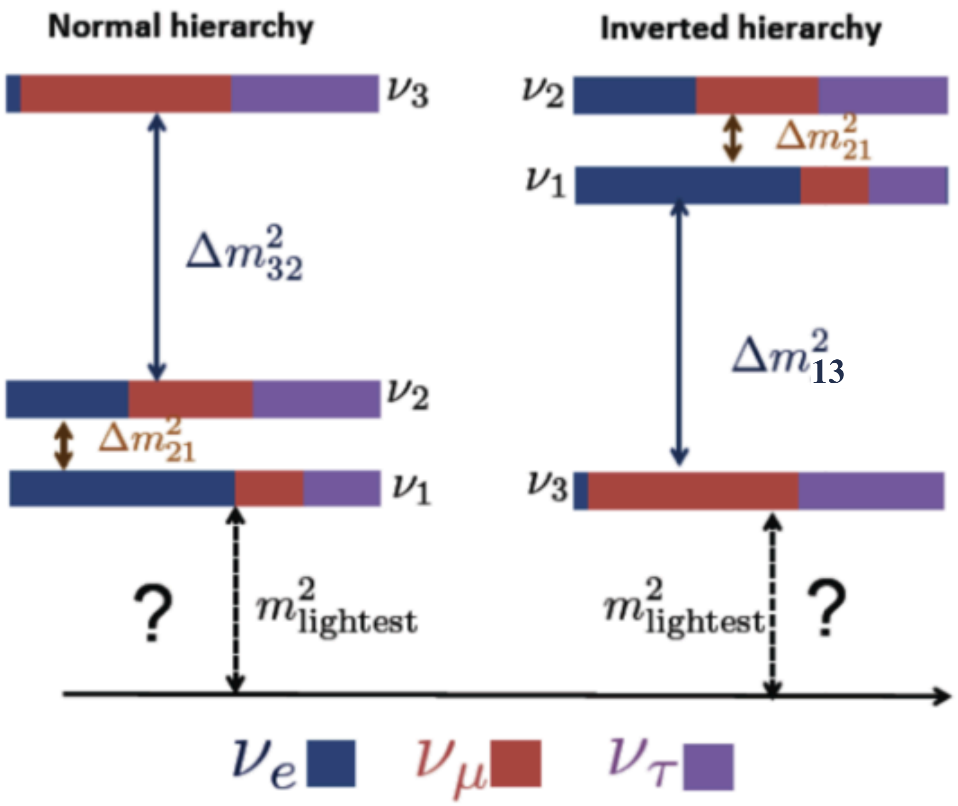
- Measurements of $\sin^2(\theta_{23})$ and Δm^2_{32}

Appearance channel ($\nu_\mu \rightarrow \nu_e$ & $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$)

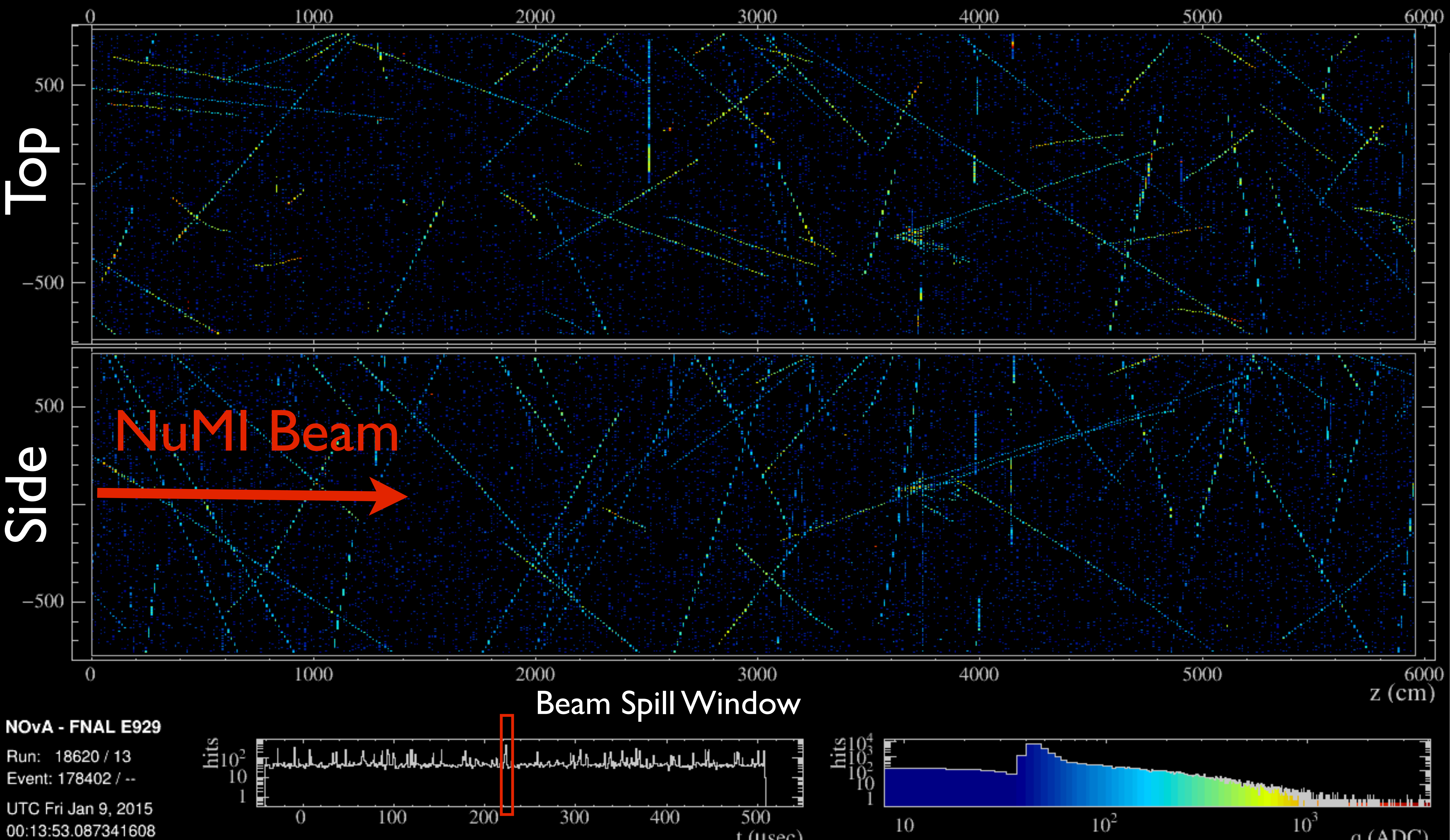
- Determine ν mass hierarchy
- Octant of θ_{23} (> or < 45°)
- Constrain δ_{CP}

Non Oscillation Physics

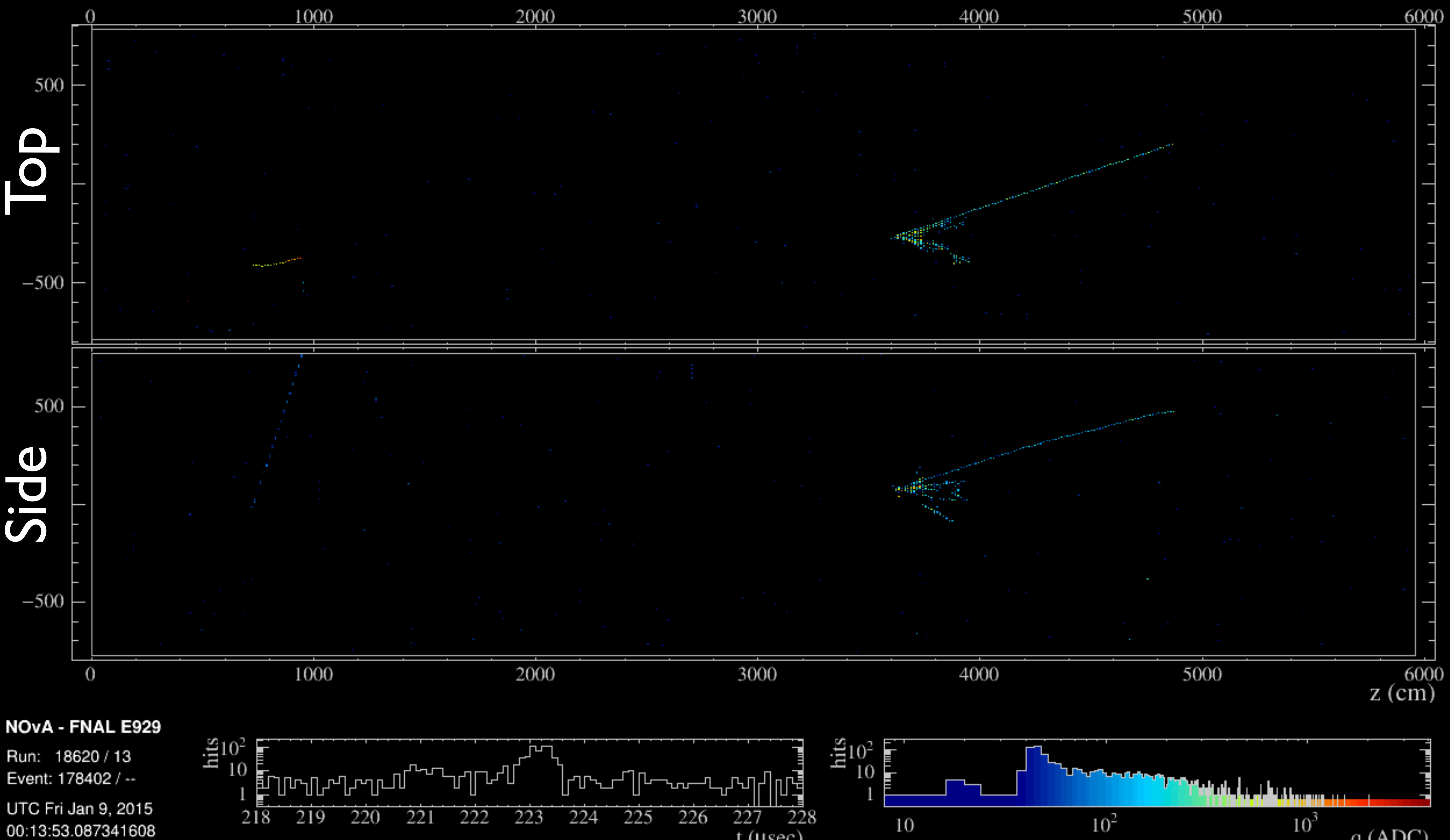
- Cross sections with NOνA ND
- Supernova neutrinos
- Sterile Neutrino Search
- Plus more!



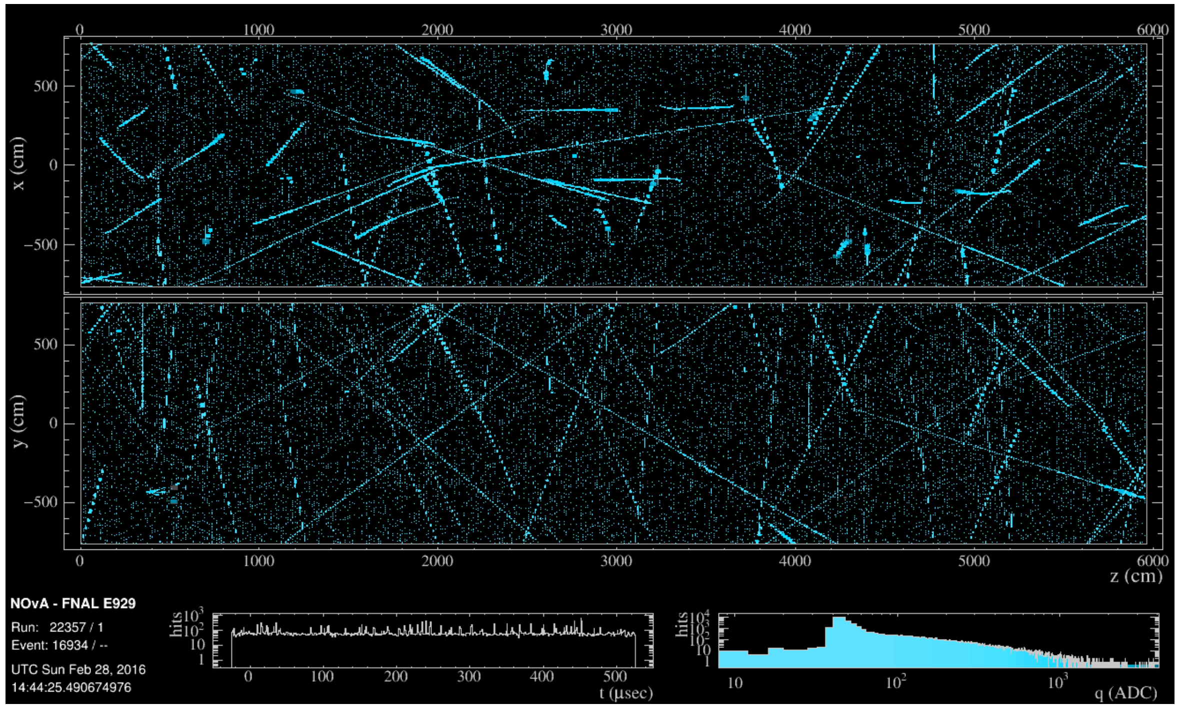
NOvA Event Display



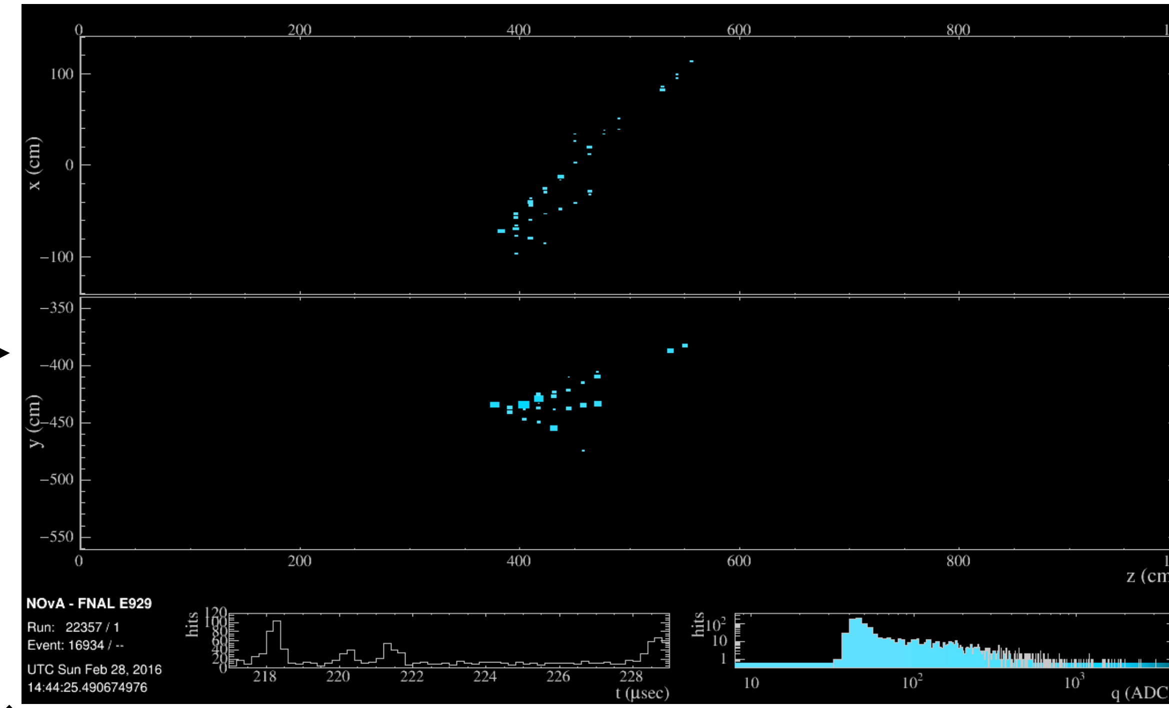
NOvA Event Display



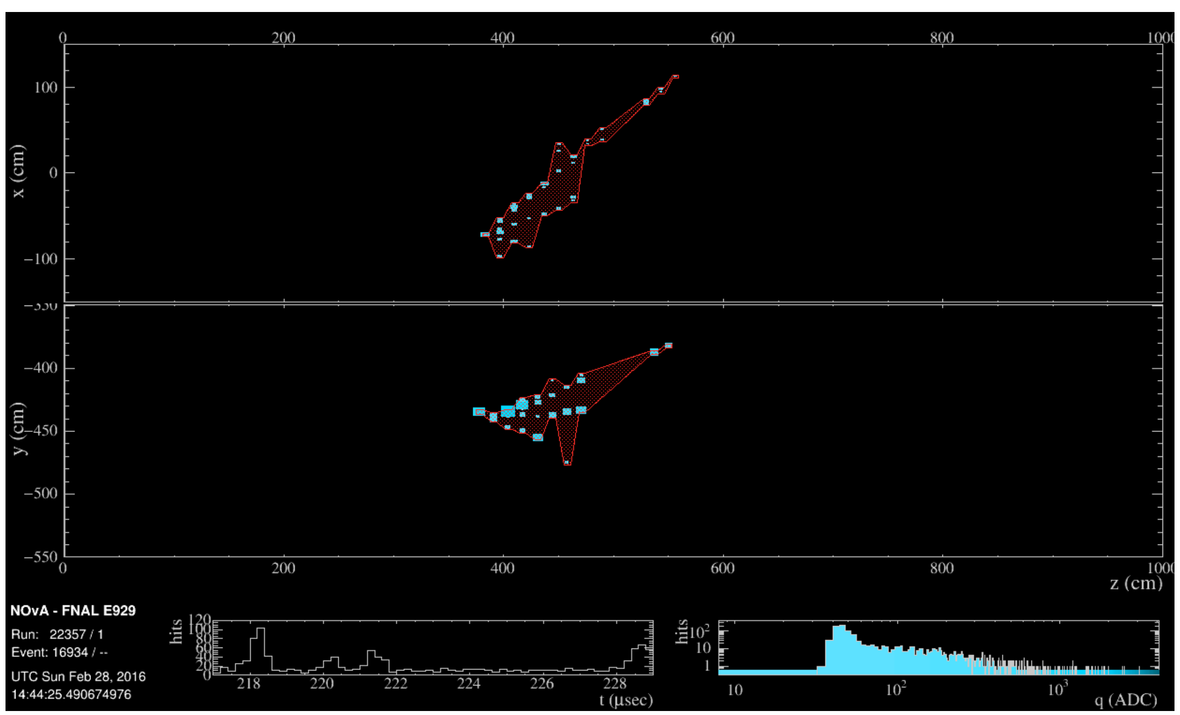
Traditional Reconstruction



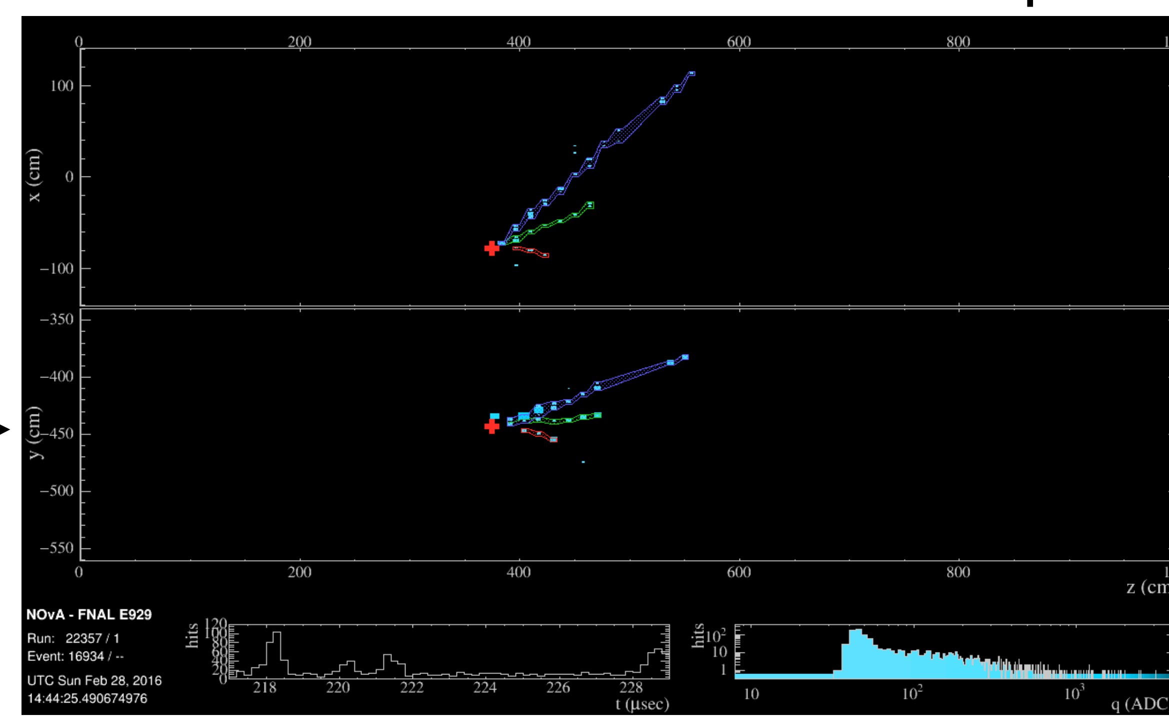
Full 550 us event in Far Detector



Isolate interactions within time and space



Group hits together to form a cluster



Find vertex of interactions and produce prongs for each individual particle

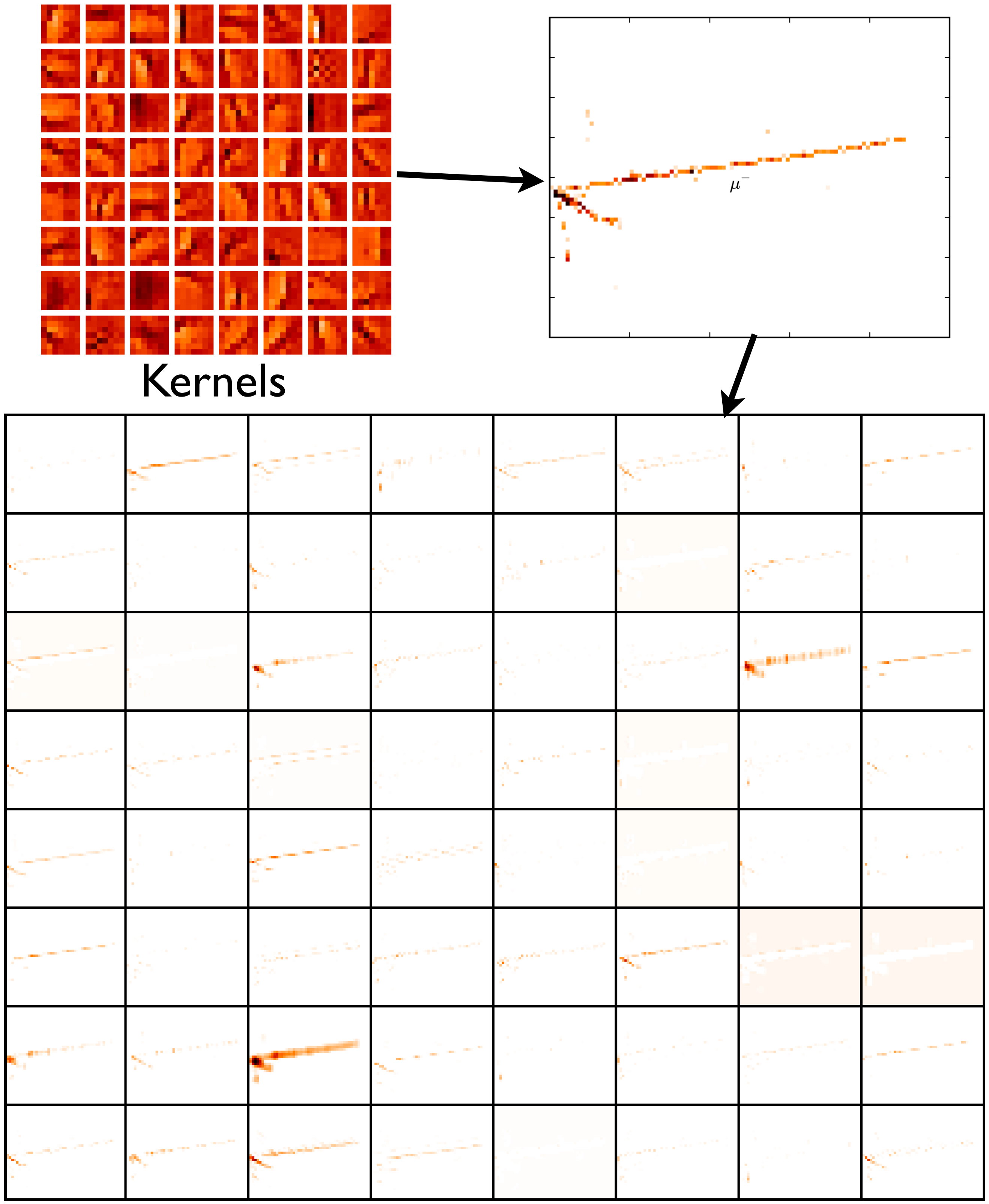
Extracting Features

In the convolutional layers, kernels are used to extract different features and create feature maps

In pooling layers, feature maps are downsampled to help computation time. Also helps access different size features

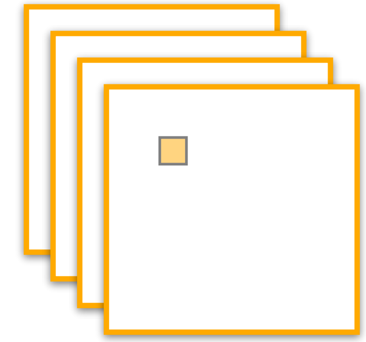
Kernels change through training to produce more useful feature maps

Fully connected layer correlates feature maps to labels. Provides a 0-1 output for each label, roughly probability of each label



Convolved Feature Maps

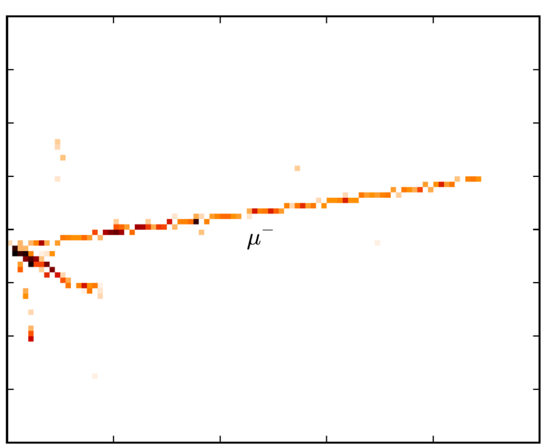
Extracting Features

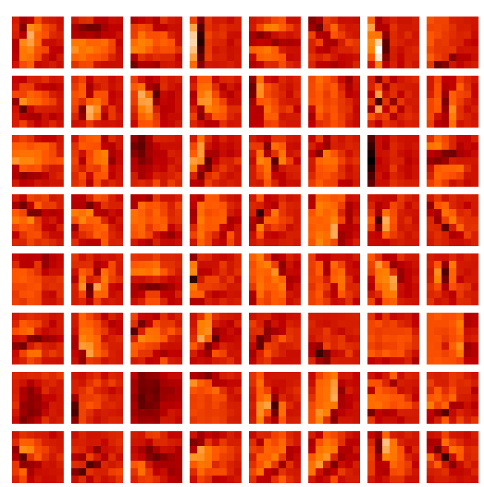


Convolutional Layer

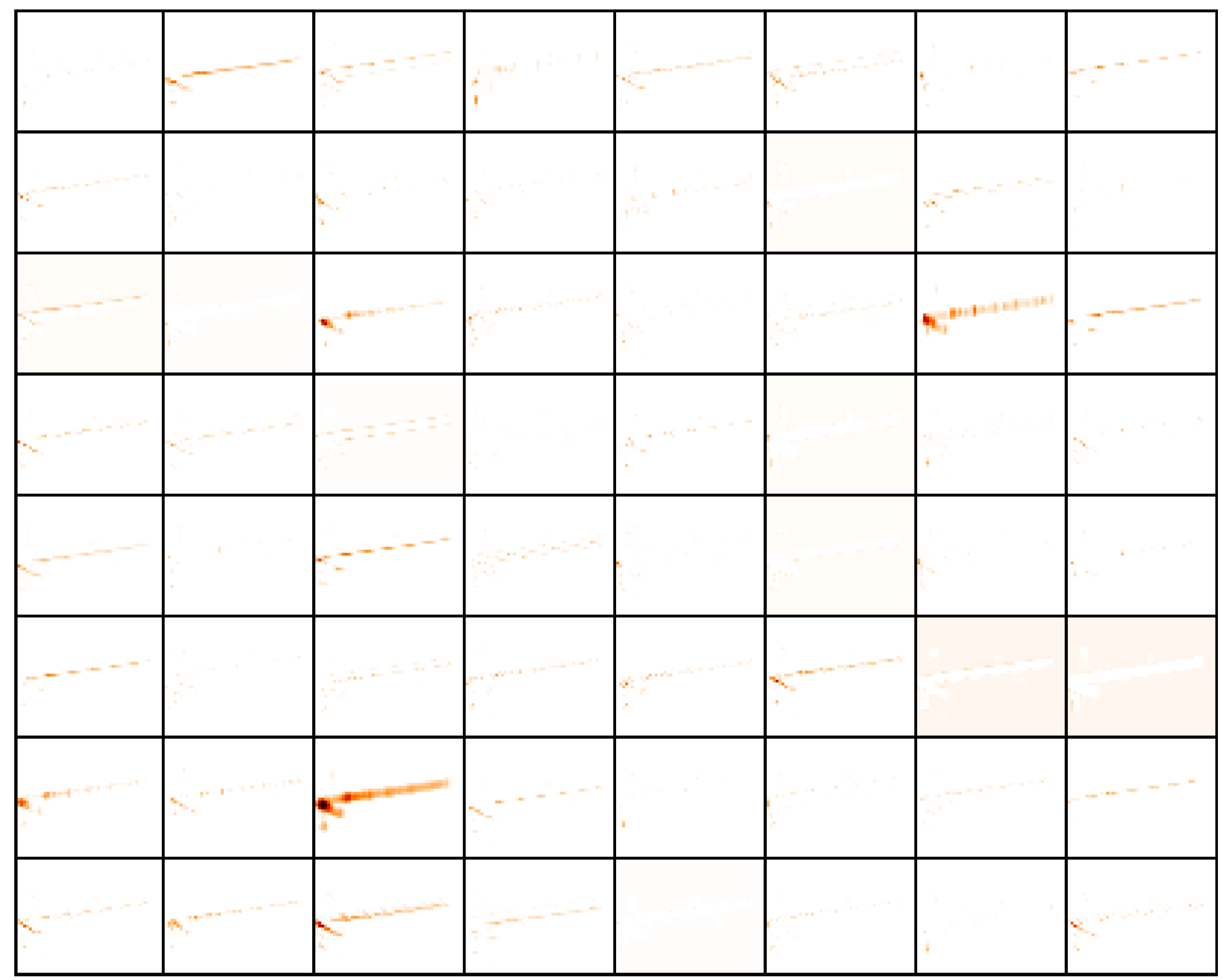
In the convolutional layers, kernels are used to extract different features and create feature maps

The network learns from correlations between different feature maps for each type of event



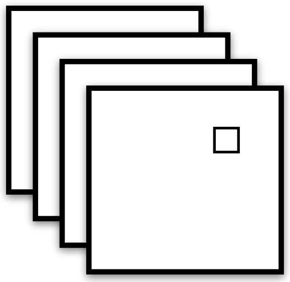


Example of kernels used for convolutions



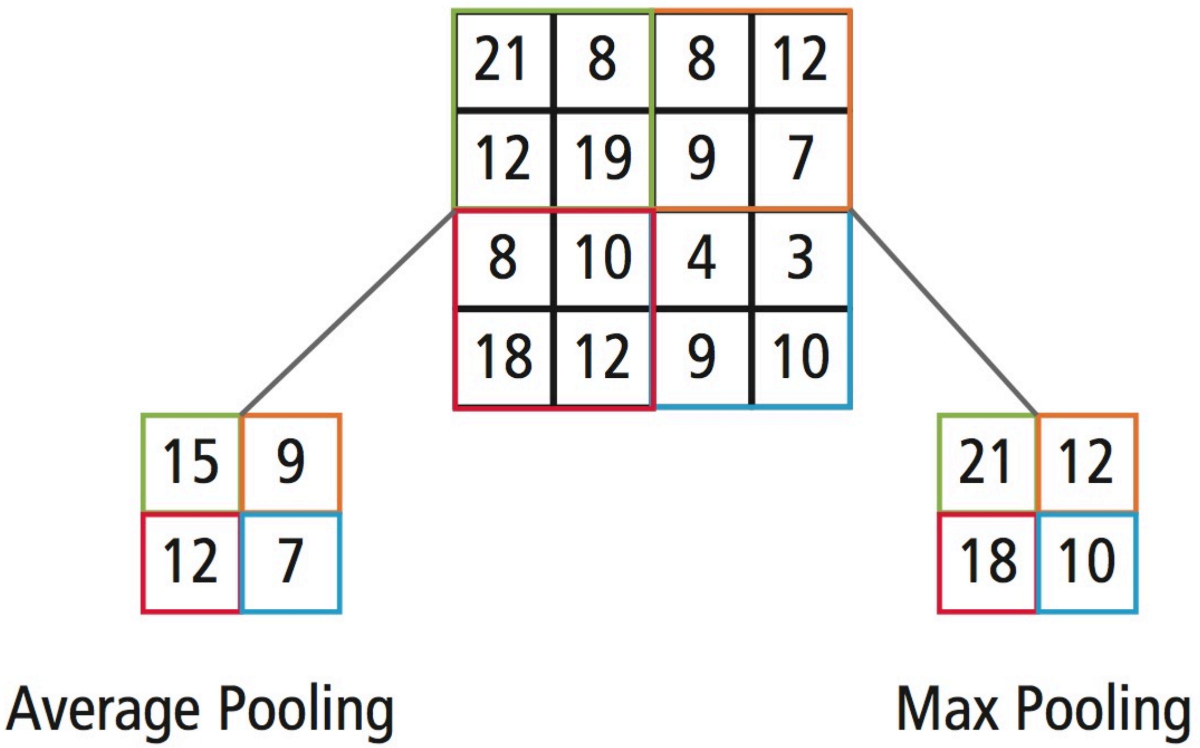
Convolved Feature Maps

Pooling Layer



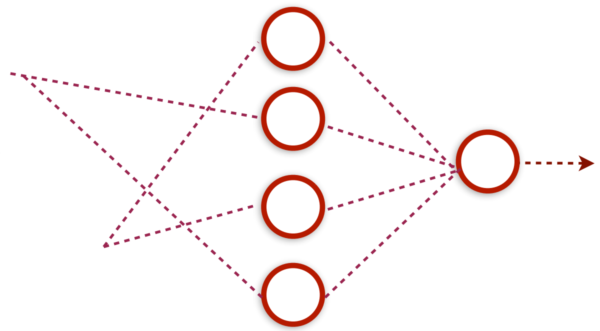
Pooling Layer

These layers downsample feature maps in order to reduce the number of parameters and computation needed. This will help with overtraining.



http://cs231n.github.io/convolutional-networks/#pool

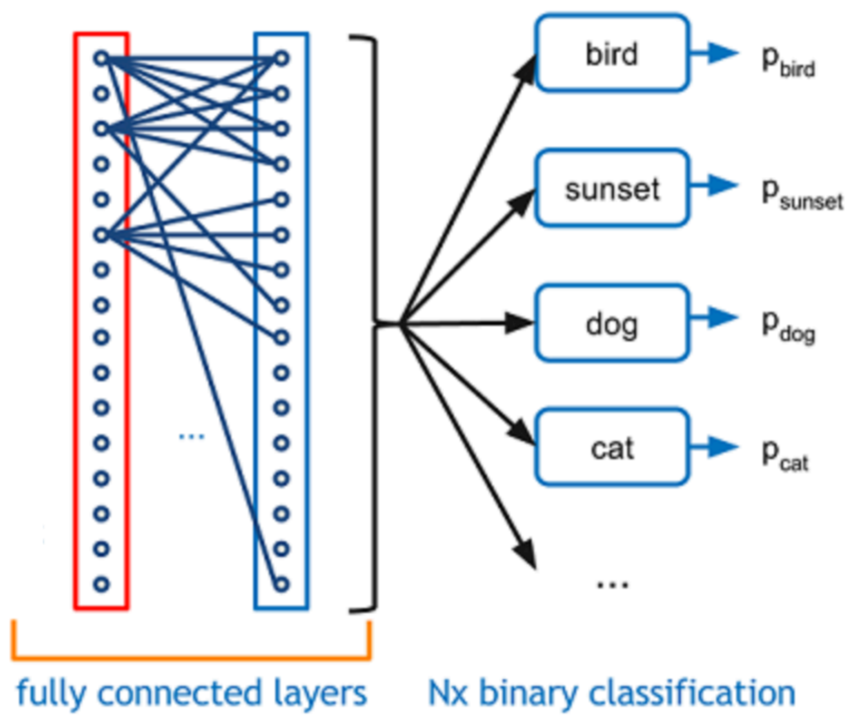
Fully Connected Layer



Fully Connected Layer

- Looks at the feature maps of the previous layer and determines which features correlate to a particular class/label

- Assuming a *softmax* output, the FCL outputs an N length vector, with the length equal to the number of classes/labels you input. Each digit will be between 0 and 1, roughly representing the probability of each class/label.



https://adeshpande3.github.io

Inception Module

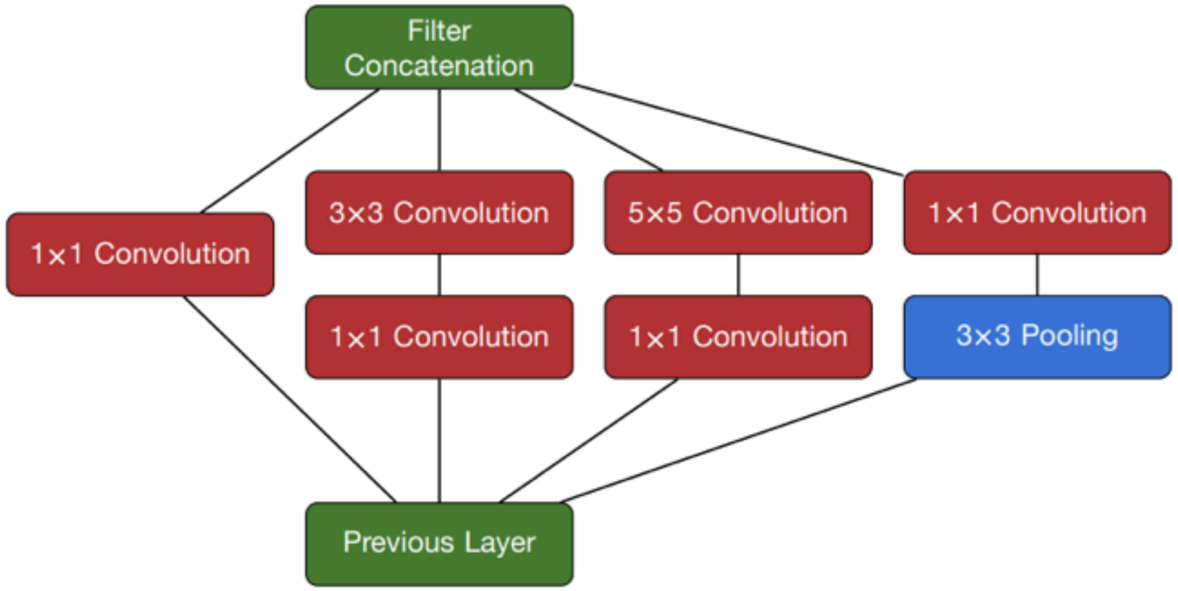
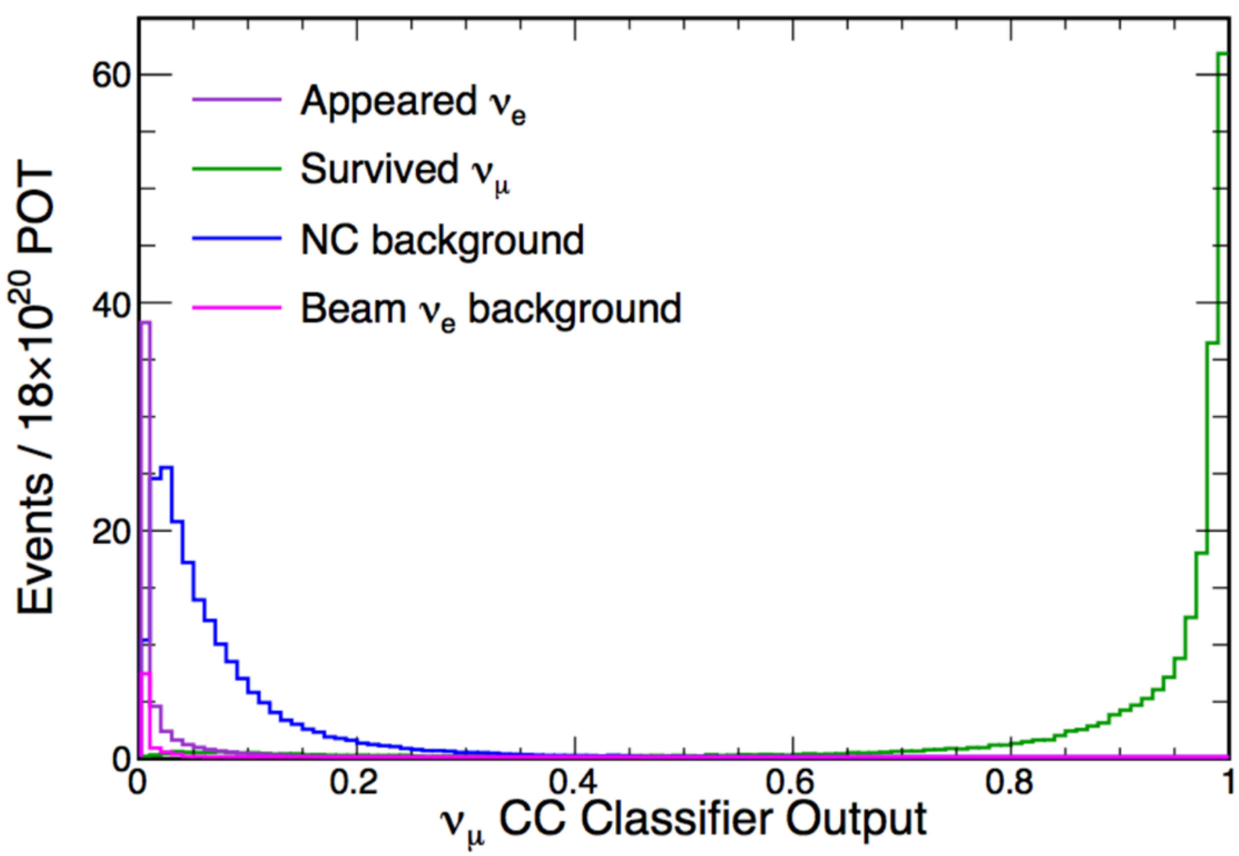


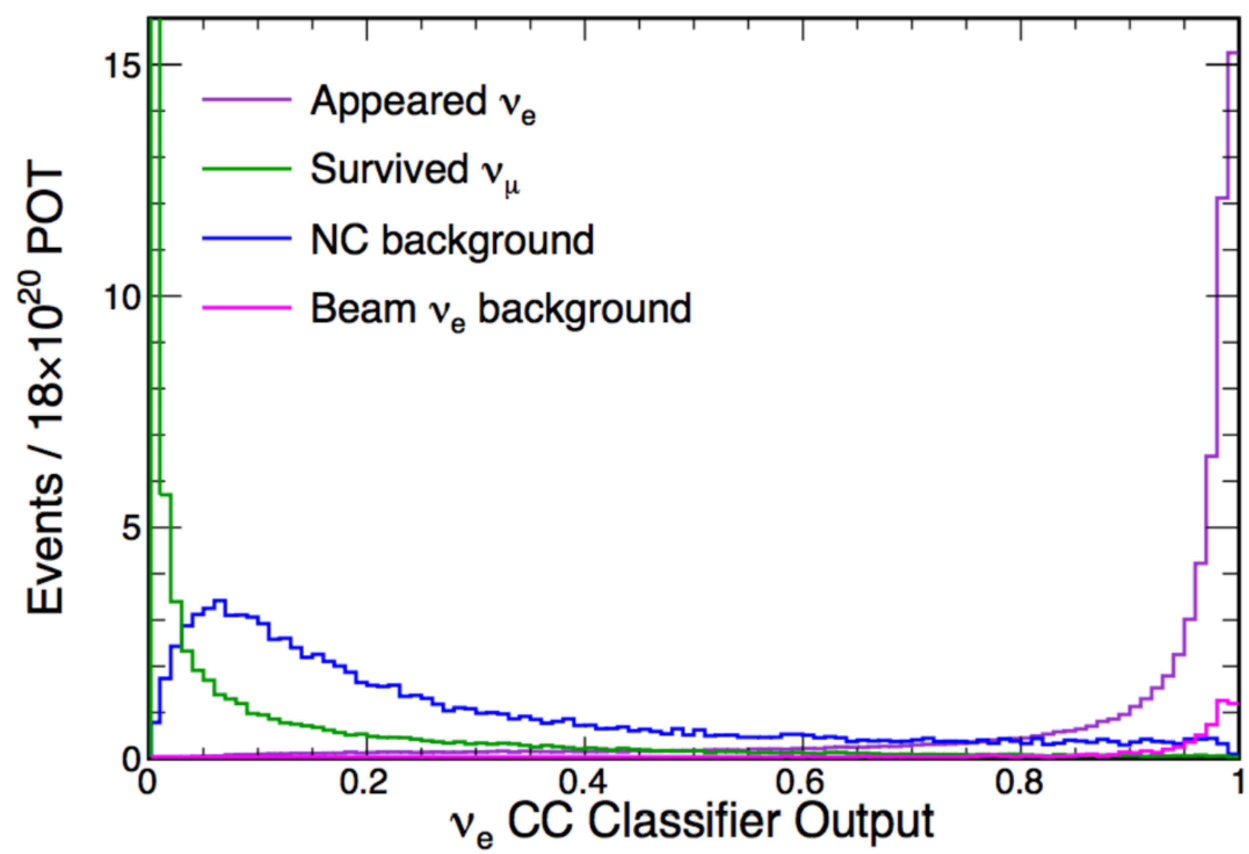
Figure 1. Diagram of the inception module
The inception module distributes filter output from the previous layer to branches, each with filter maps at different scales. NIN architecture is implemented as 1 × 1 convolutions which form linear combinations of the input feature maps to reduce dimensionality through semantic similarity. Separate branches perform 3×3 and 5×5 convolution, as well as 3×3 overlapping pooling. The filtered outputs from each branch are concatenated along the channel dimension before being passed to the next layer.

"A Convolutional Neural Network Neutrino Event Classifier"
A.Aurisano et. al. JINST 11 (2016) no.09, P09001

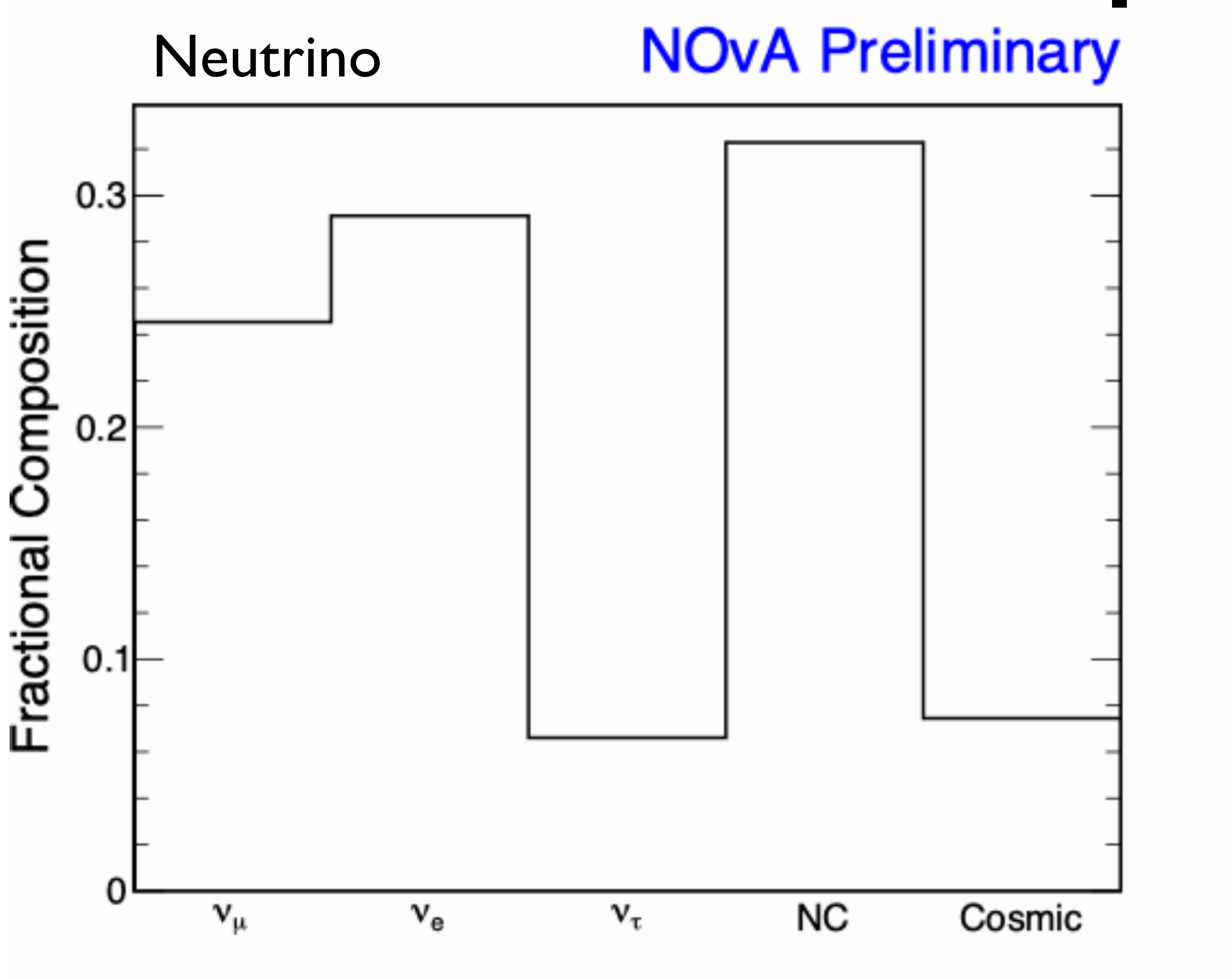
CVN Event Classifier Output

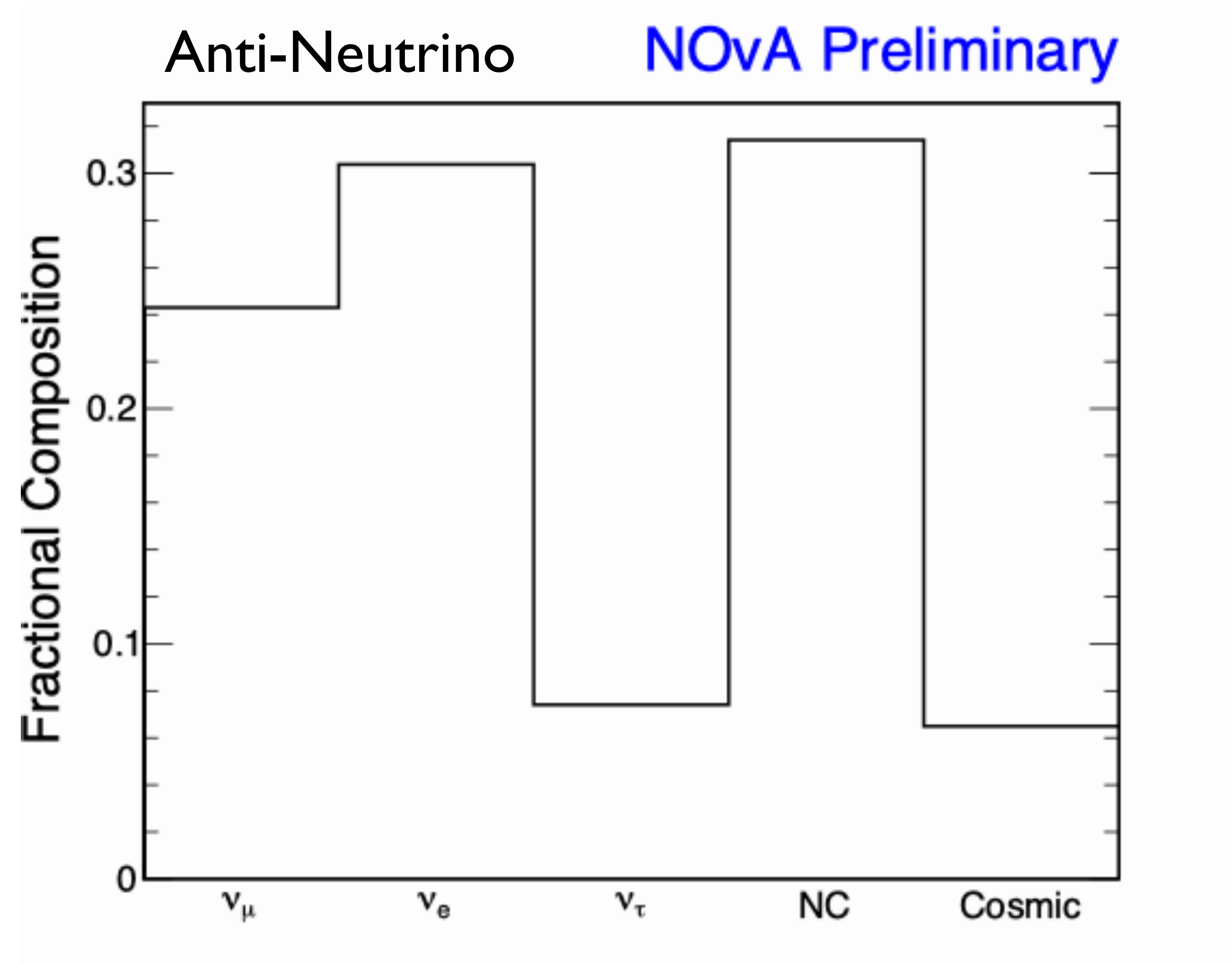
- Output gives a value for each category whose sum is normalized to 1 for all labels





Event CNN Training Sample Compositions

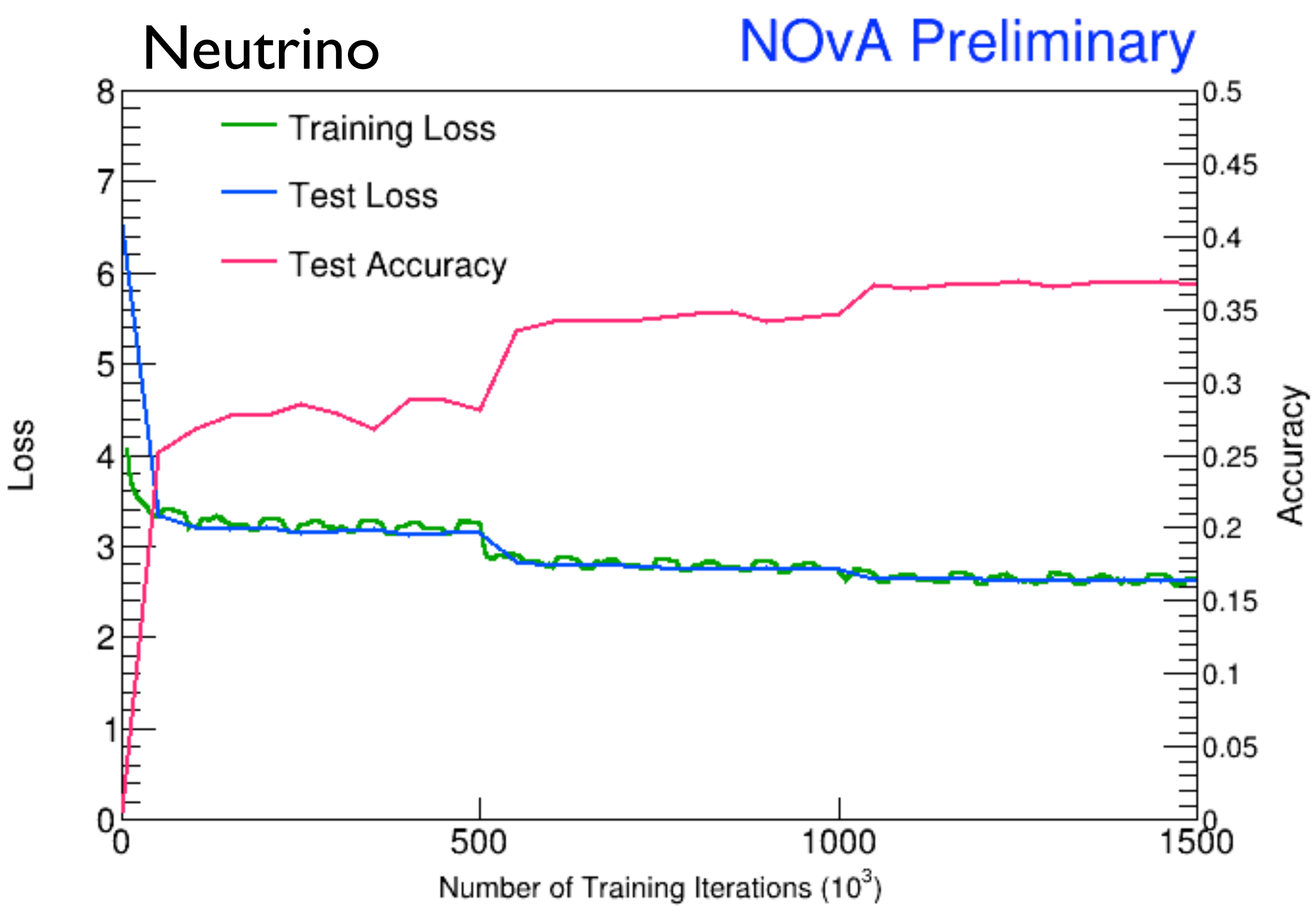




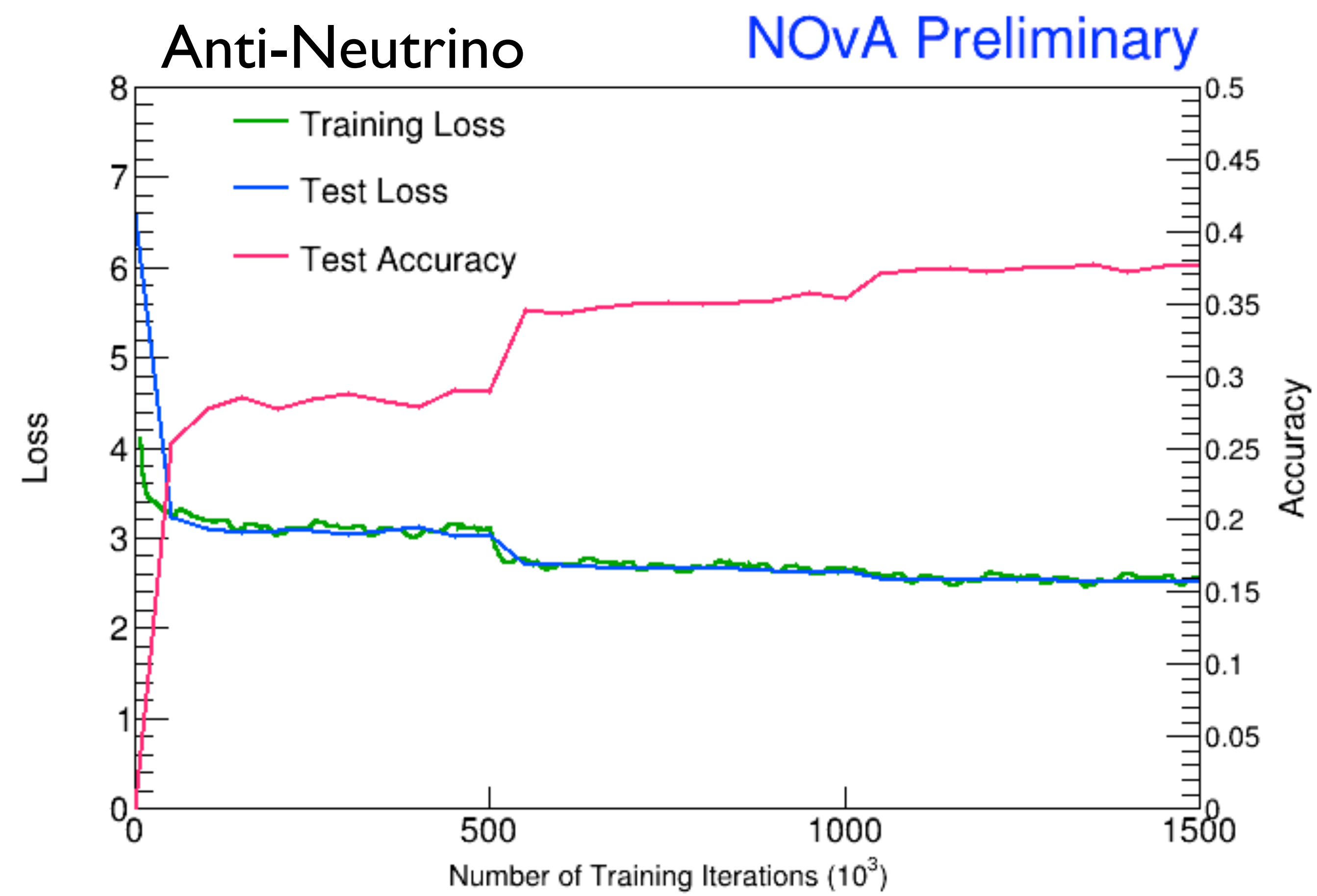
Similar composition of major categories between datasets.
Cosmics are data.

Event CNN Training Evolution

Neutrino

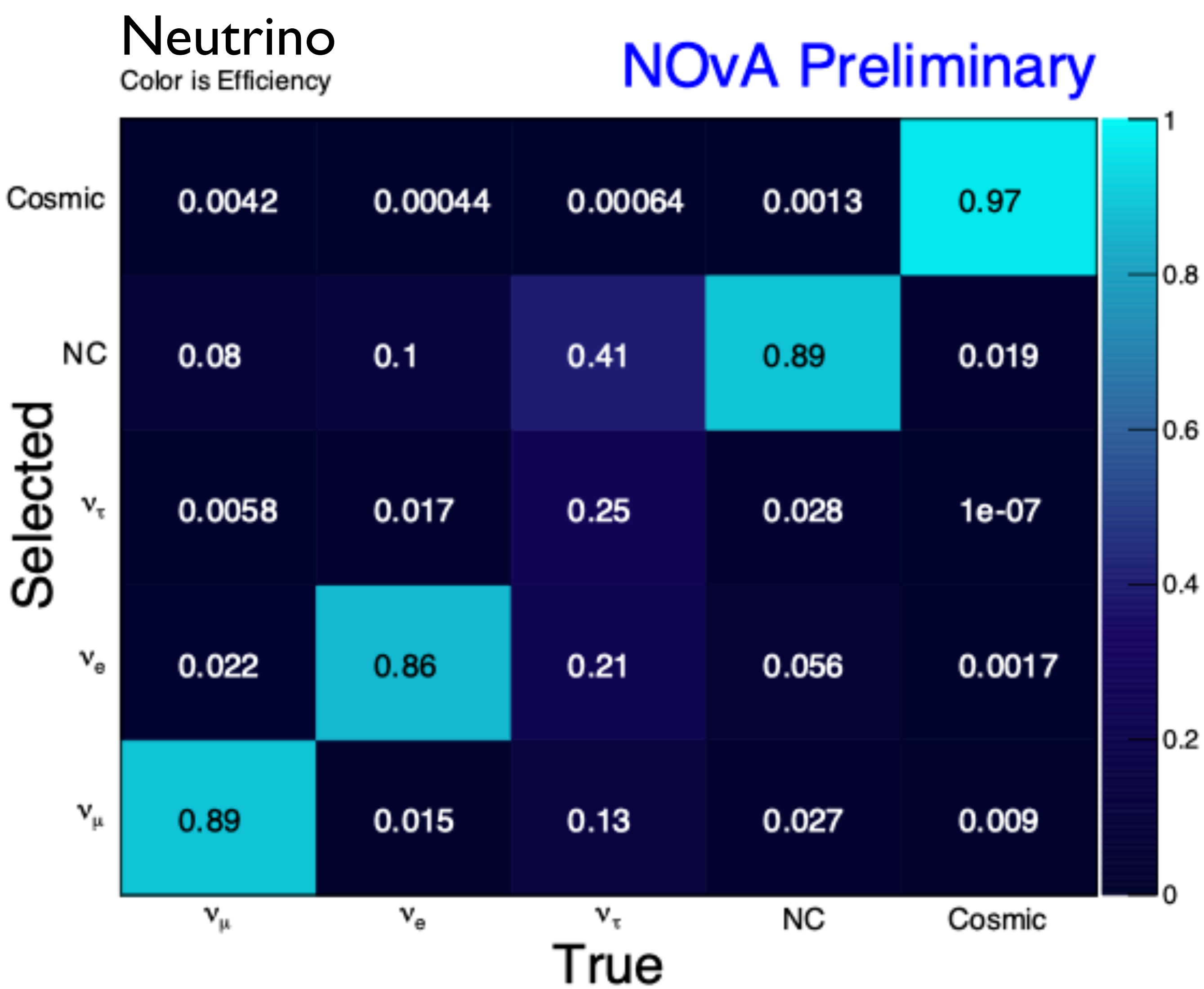


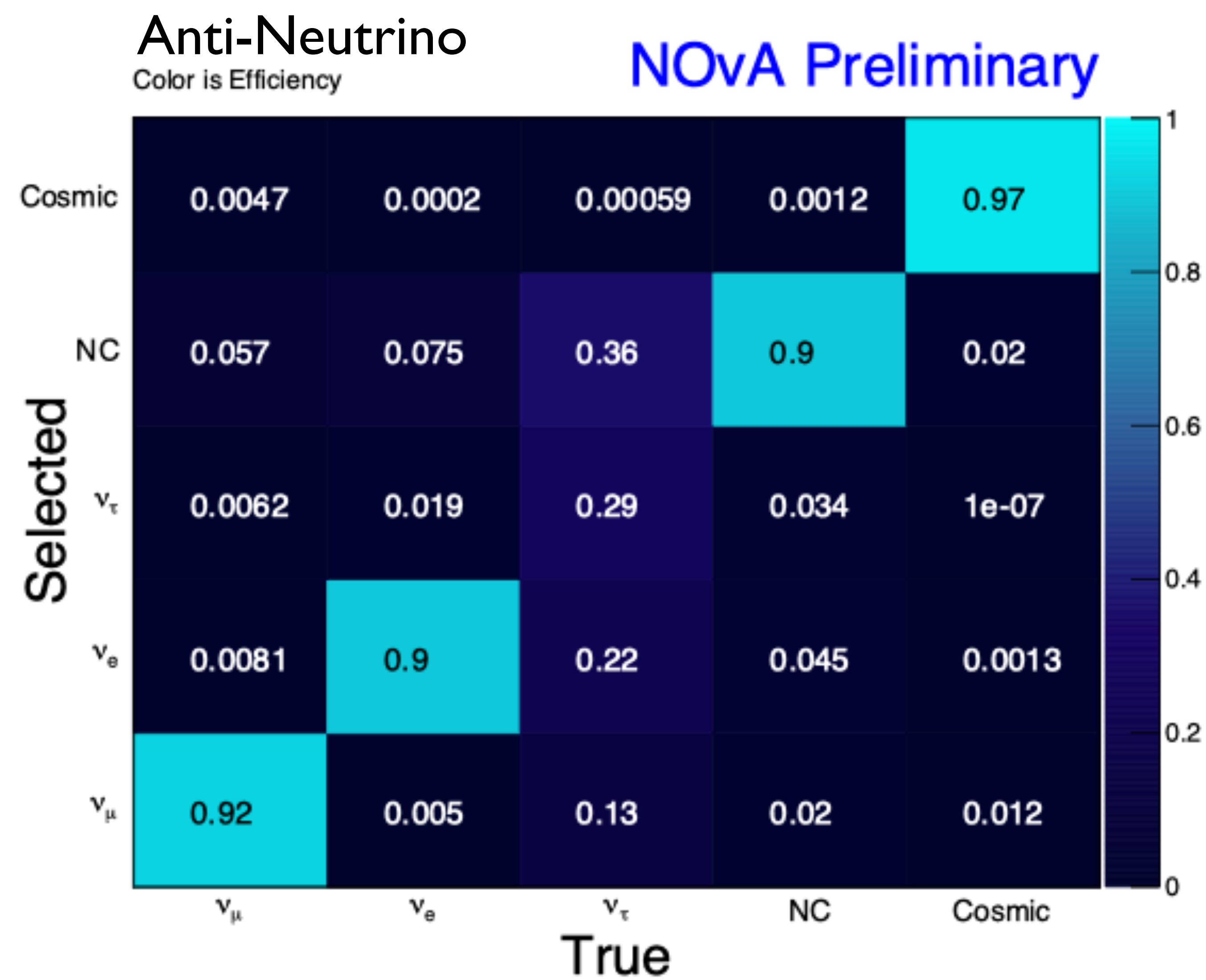
Anti-Neutrino



The red curve is accuracy of the top-1. The blue and green curves are the output of the loss function in the test and training datasets, respectively. The dips at 500k and 1M iterations is where the learning rate of the network becomes smaller. The flatness past 1M iterations shows that the network has found a local minima. The agreement between test and training loss gives a good indication that the network has not overtrained.

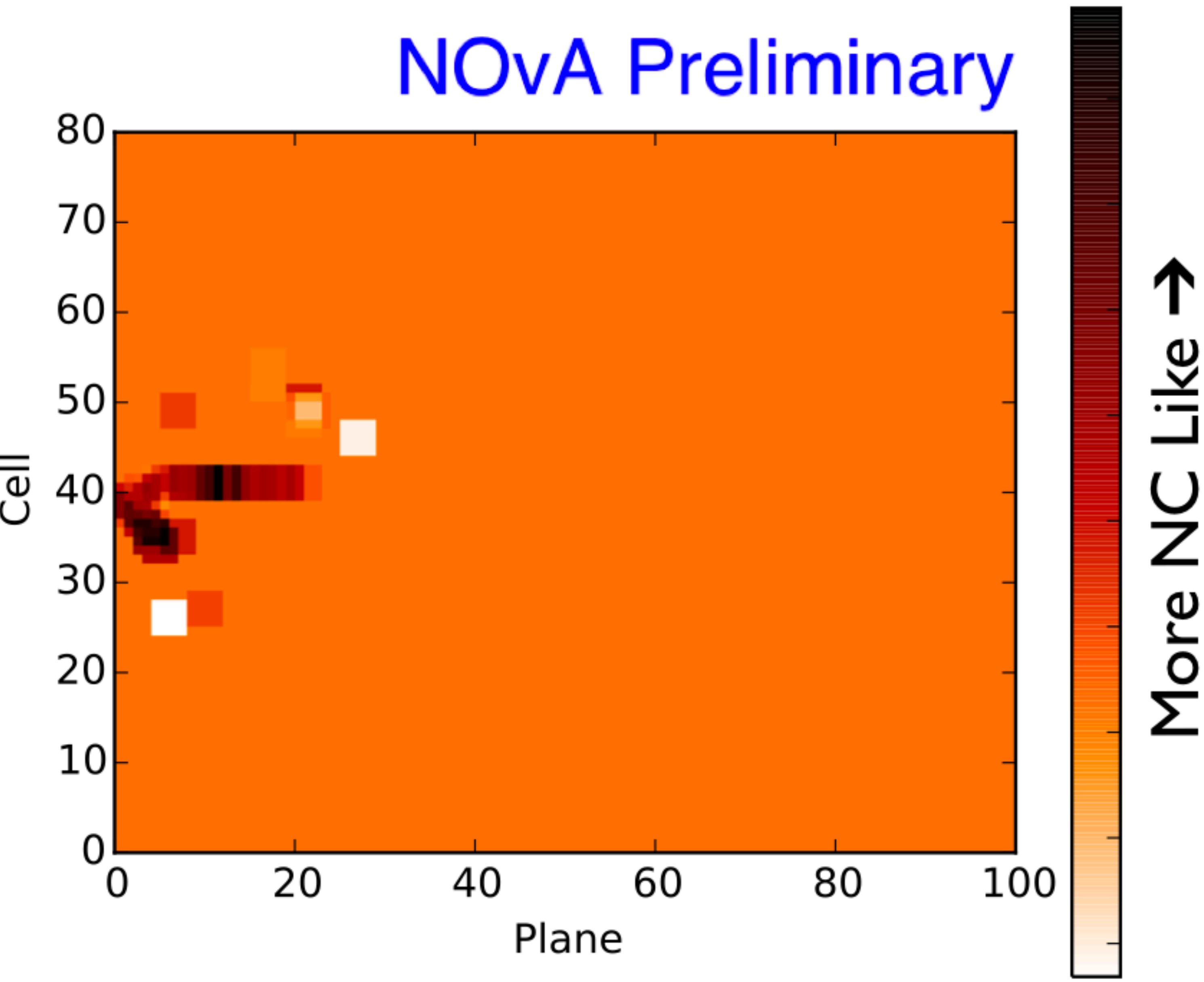
Event CNN Classification Matrices





Events are sorted by the their true category and then selected by whichever CVN output gives highest value. Each column is normalized to 1. Along the diagonal gives the efficiency of each category while the off diagonal gives insight to how the networks misclassify events.

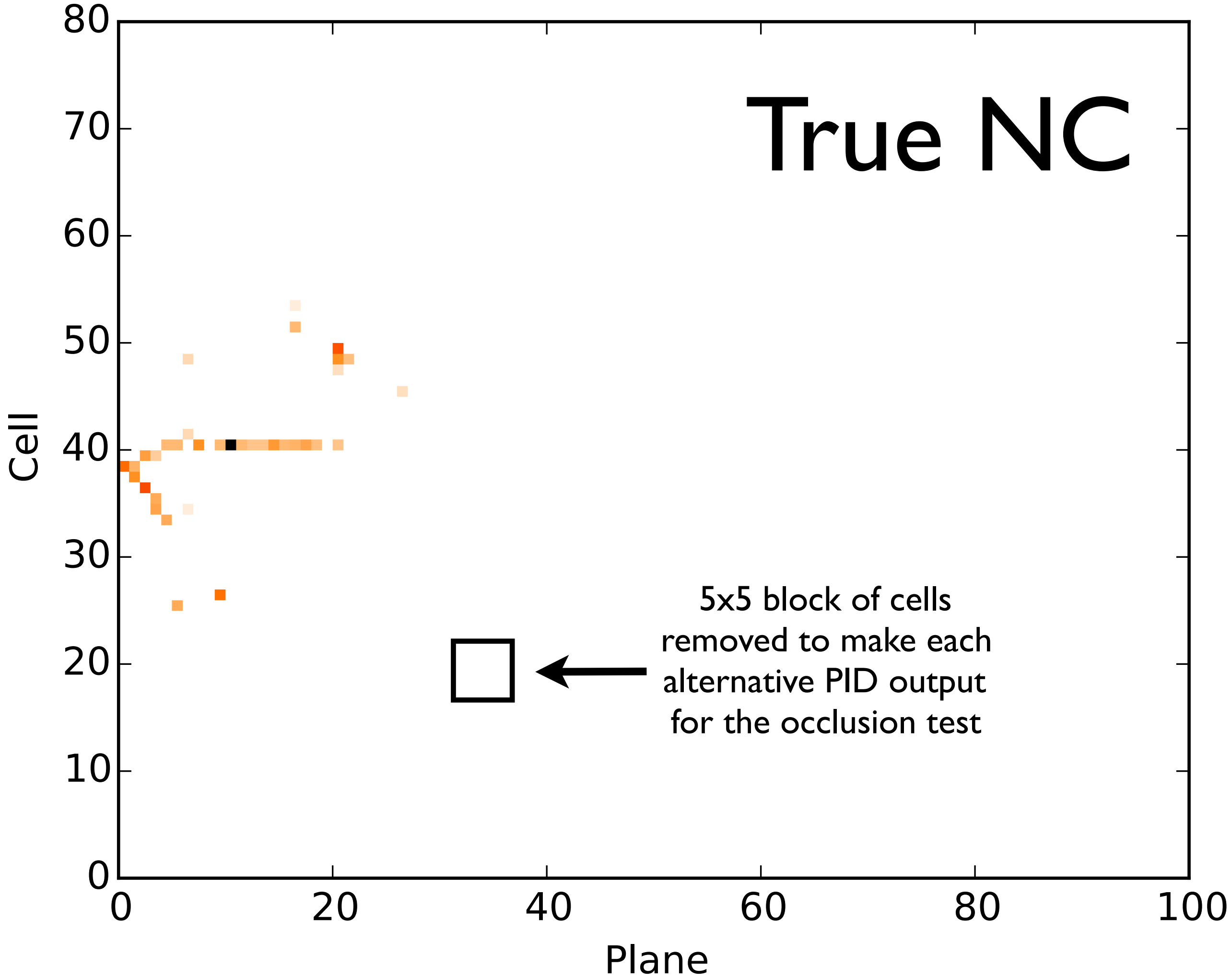
Occlusion Tests



NC PID

NC PID doesn't find tracks to useful in identifying NC. Suggests that the NC PID is more sensitive to the activity outside of the tracks.

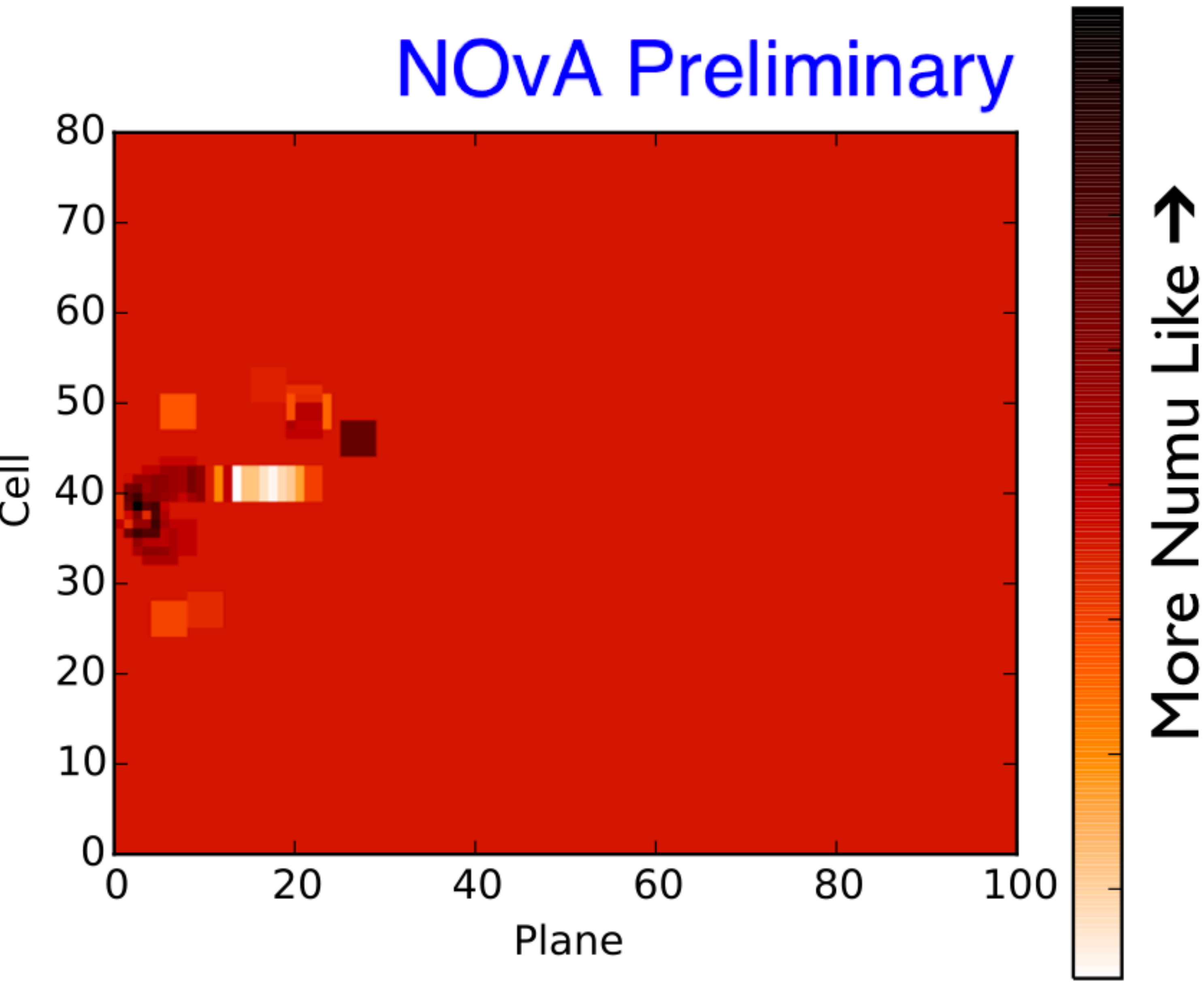
Occlusion Tests



X view

- Offers a way to peer inside what CVN is learning
- Remove 5x5 block of cells from image
- Rerun image through CVN evaluator to get new scores for each PID

Occlusion Tests

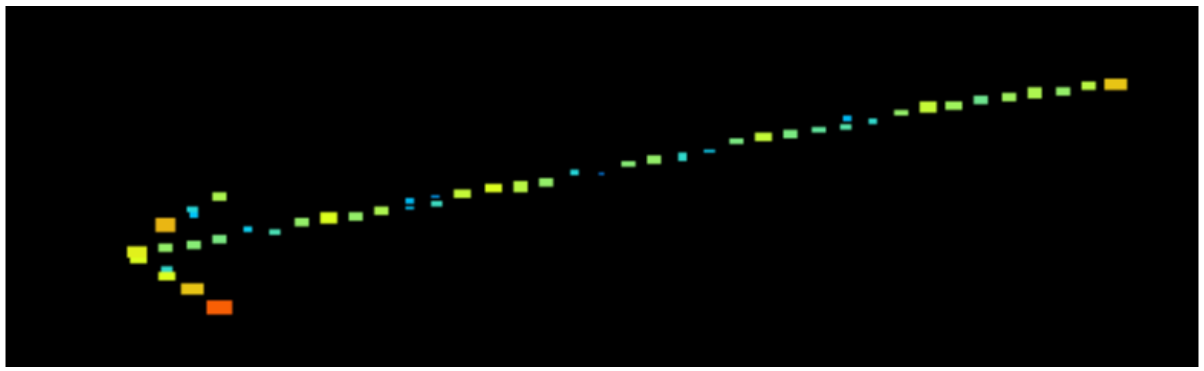


ν_μ PID

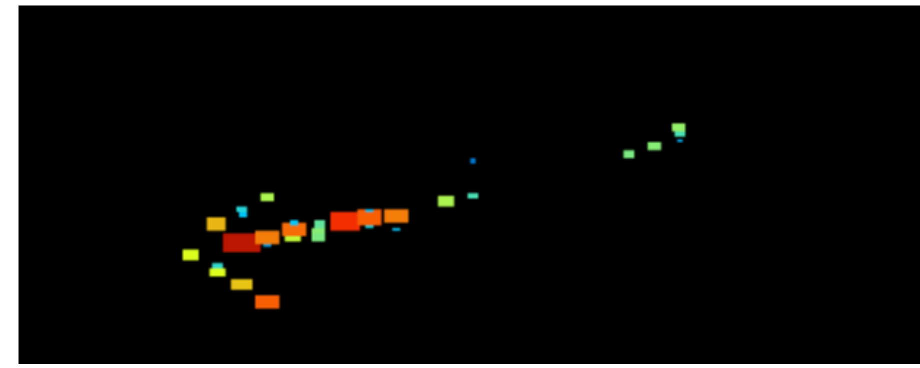
Suggests that ν_μ PID is sensitive to tracks >10 planes. Activity outside of tracks is disfavored.

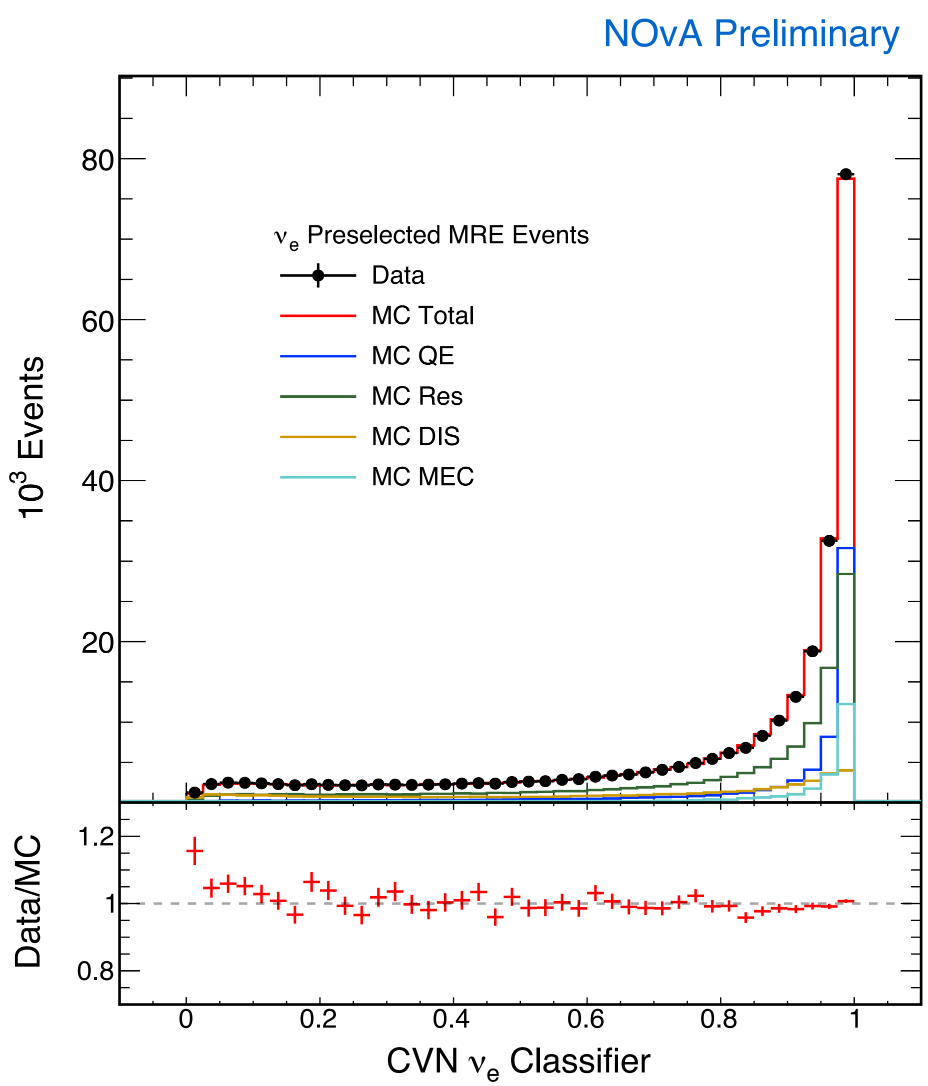
Event CVN on Real Data

Data



Data - muon + MC electron

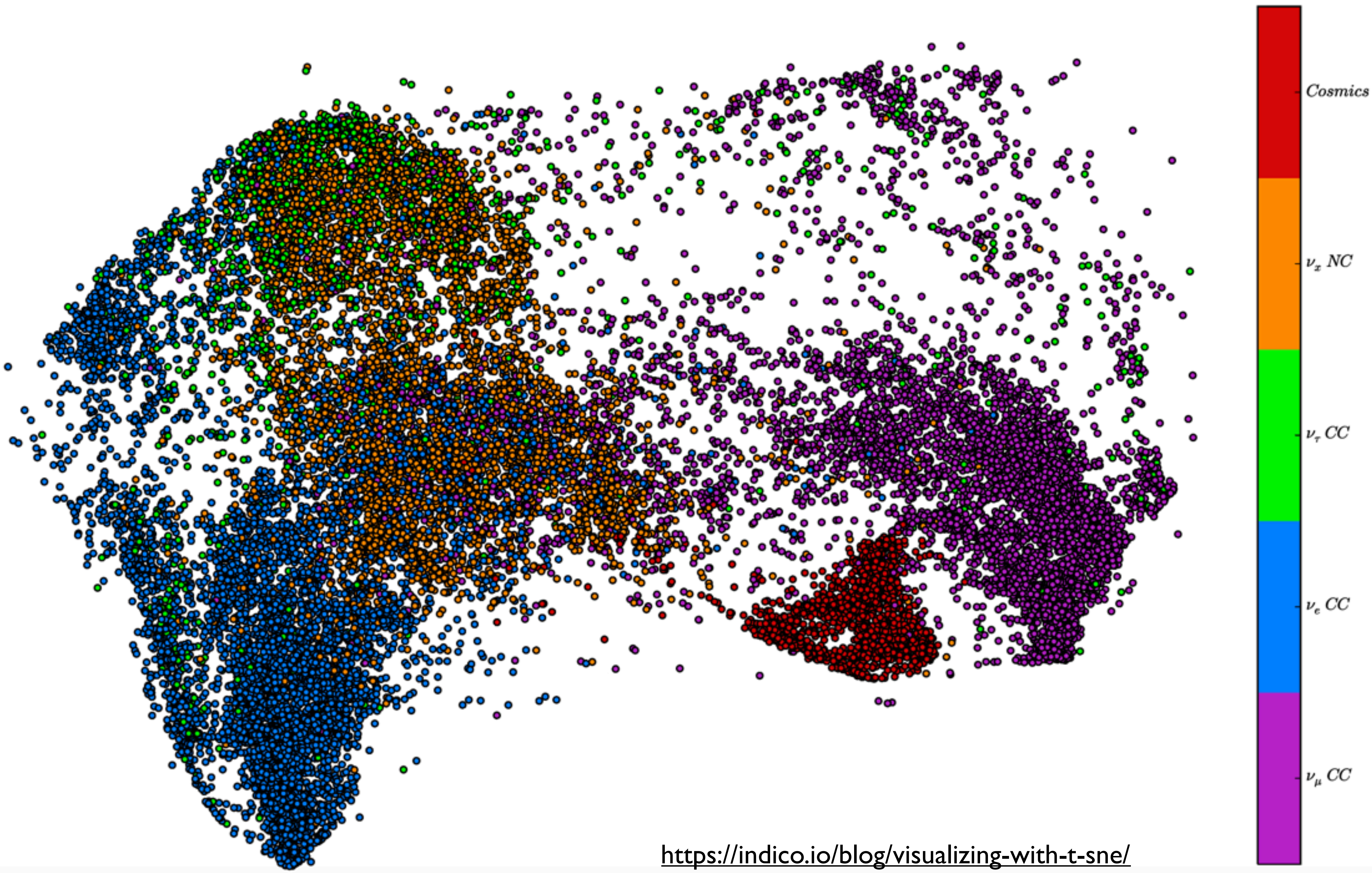




- Select muon neutrino interaction with traditional reco methods
- Remove muon hits and replace with simulated electron
- Less than 0.5% difference in efficiency between Data/MC

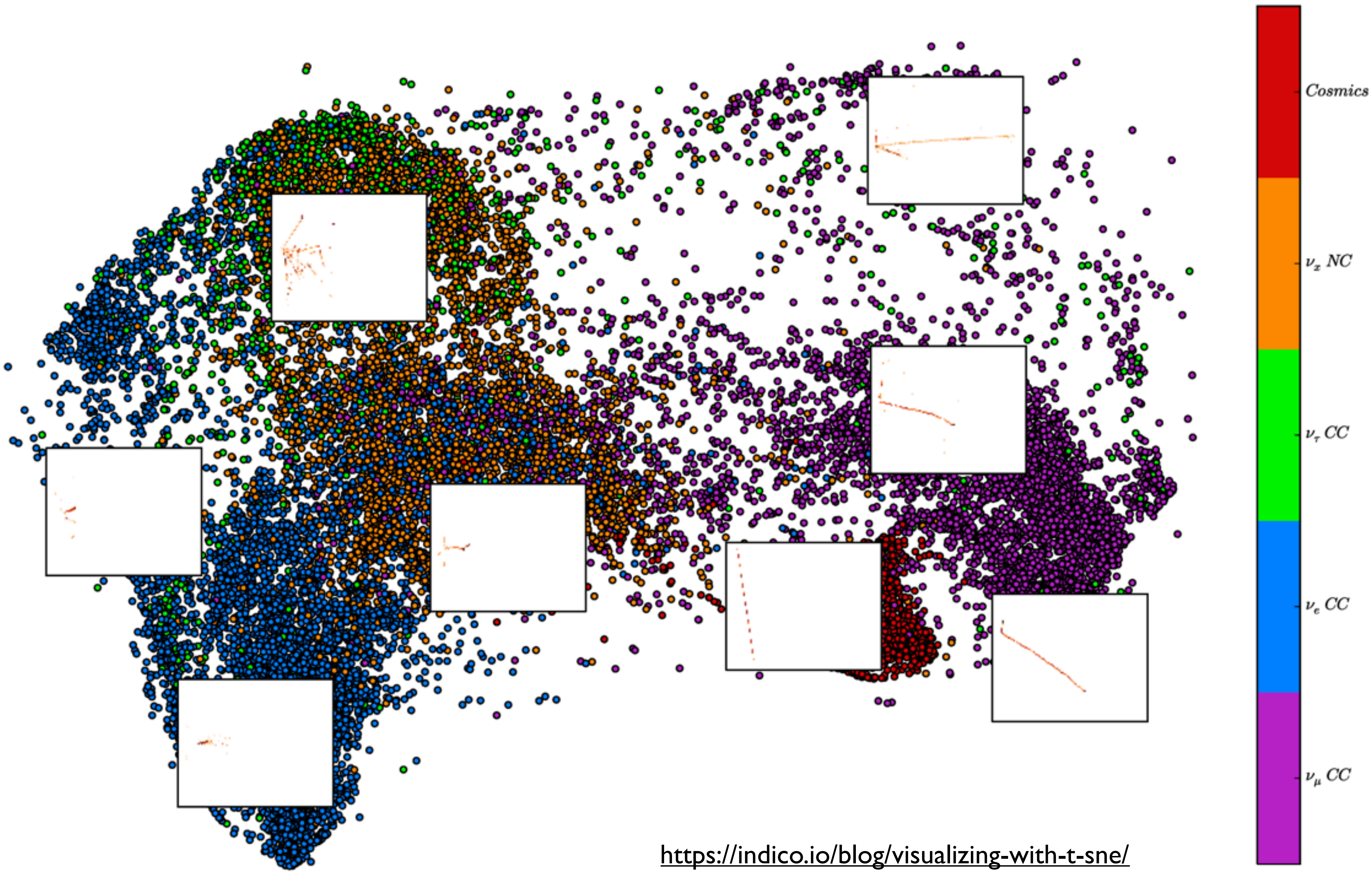
PID	Sample	Preselection	PID	Efficiency	Efficiency diff %
CVN	Data	262884	188809	0.718222	-0.36%
	MC	277320	199895	0.720809	

Event CNN t-SNE



https://indico.io/blog/visualizing-with-t-sne/

Event CNN t-SNE



https://indico.io/blog/visualizing-with-t-sne/