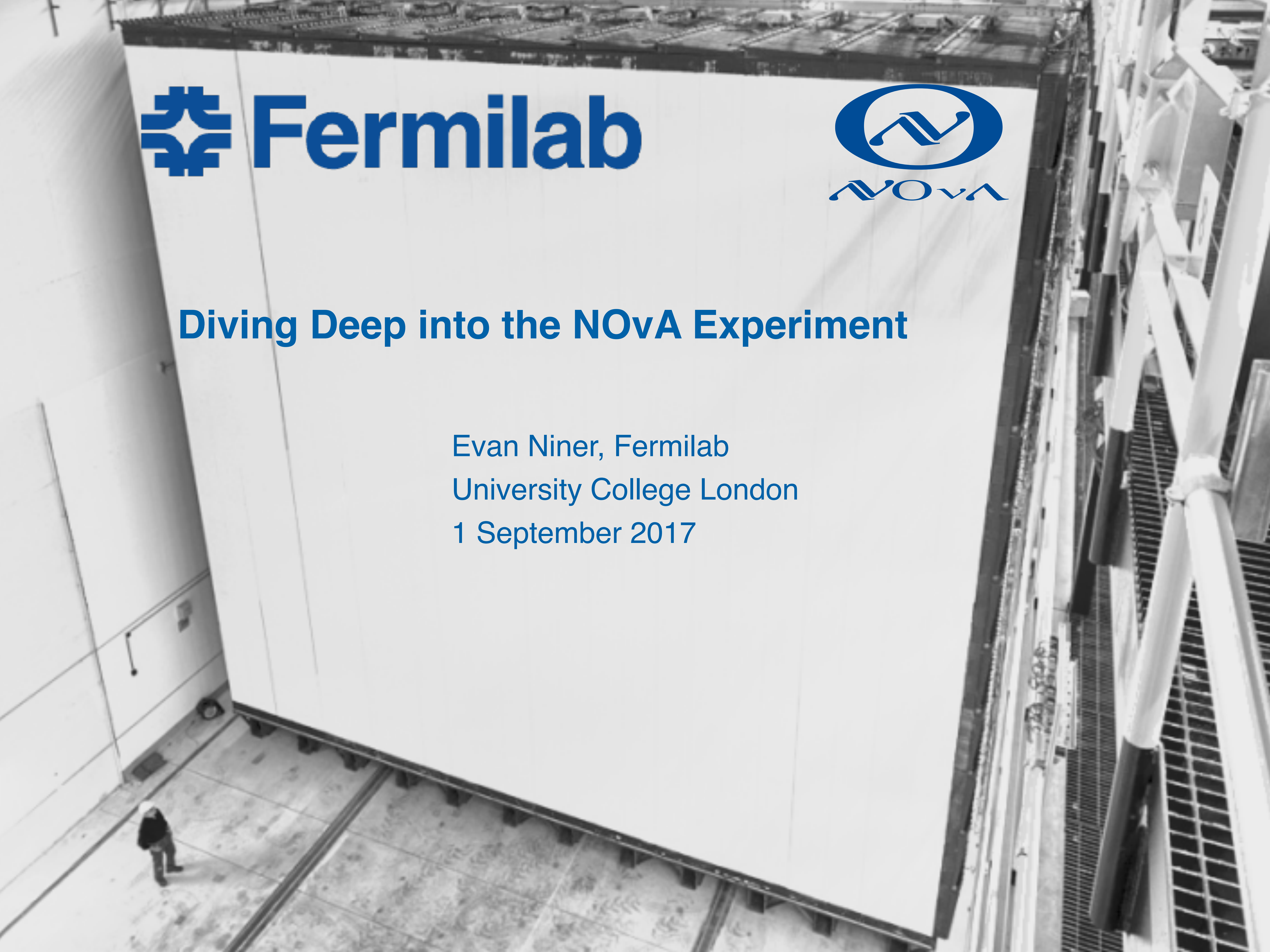


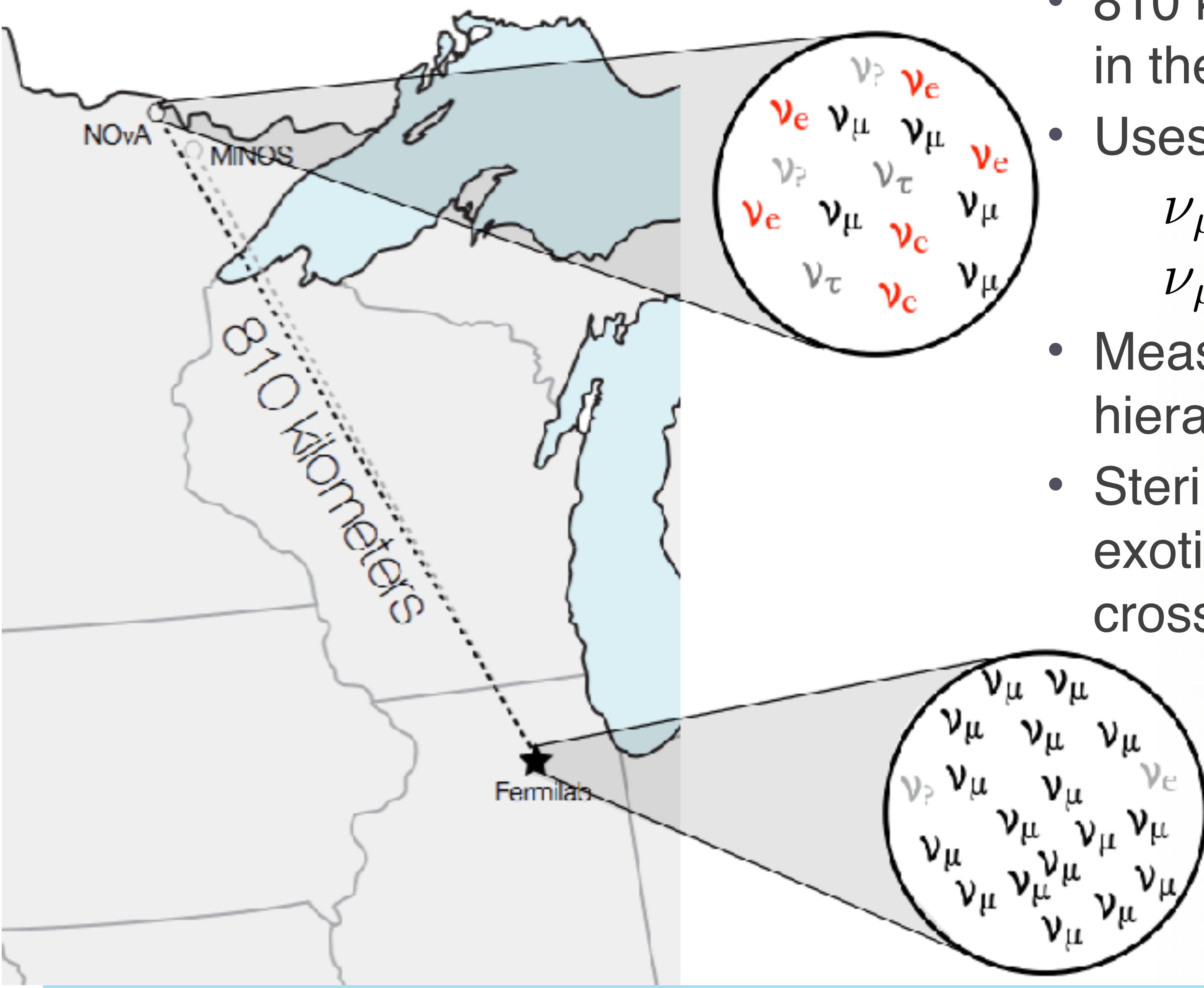
Diving Deep into the NOvA Experiment

Evan Niner, Fermilab

University College London

1 September 2017

The NOvA Experiment

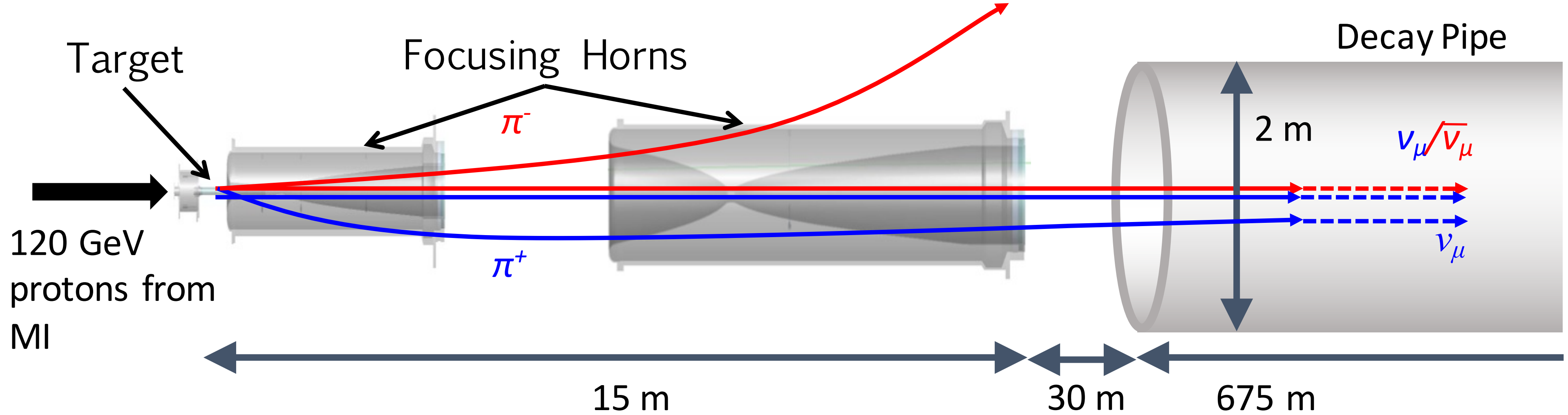


- Observe neutrinos from NuMI neutrino beam line at Fermilab
- Two functionally identical detectors
- 810 km baseline, the longest in the world
- Uses four oscillation channels:

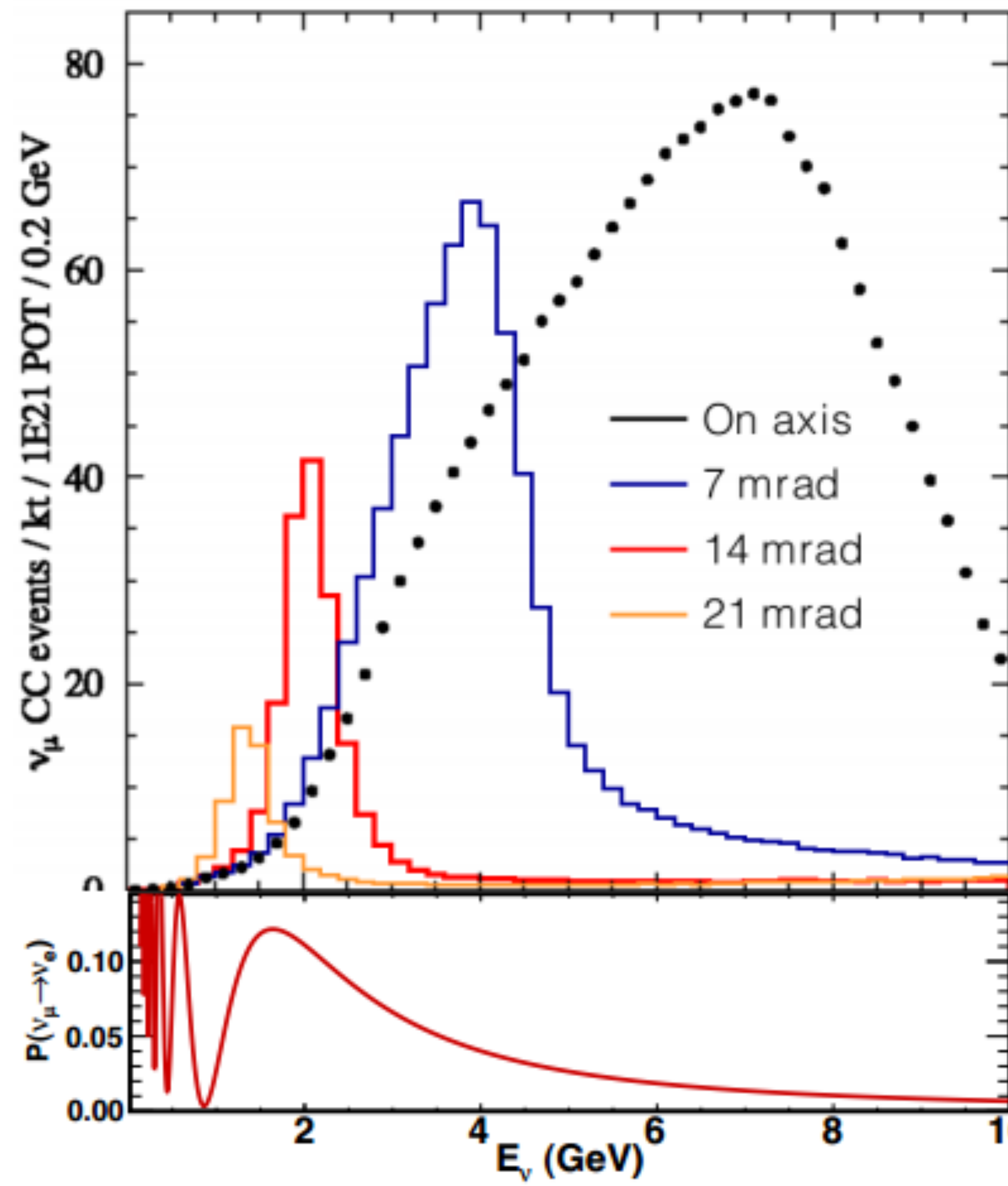
$$\nu_\mu \rightarrow \nu_\mu \qquad \overline{\nu}_\mu \rightarrow \overline{\nu}_\mu$$
$$\nu_\mu \rightarrow \nu_e \qquad \overline{\nu}_\mu \rightarrow \overline{\nu}_e$$

- Measure θ_{13}, θ_{23}, Δm^2_{32}, mass hierarchy, and δ_{cp}
- Sterile neutrino searches, exotic searches, neutrino cross sections

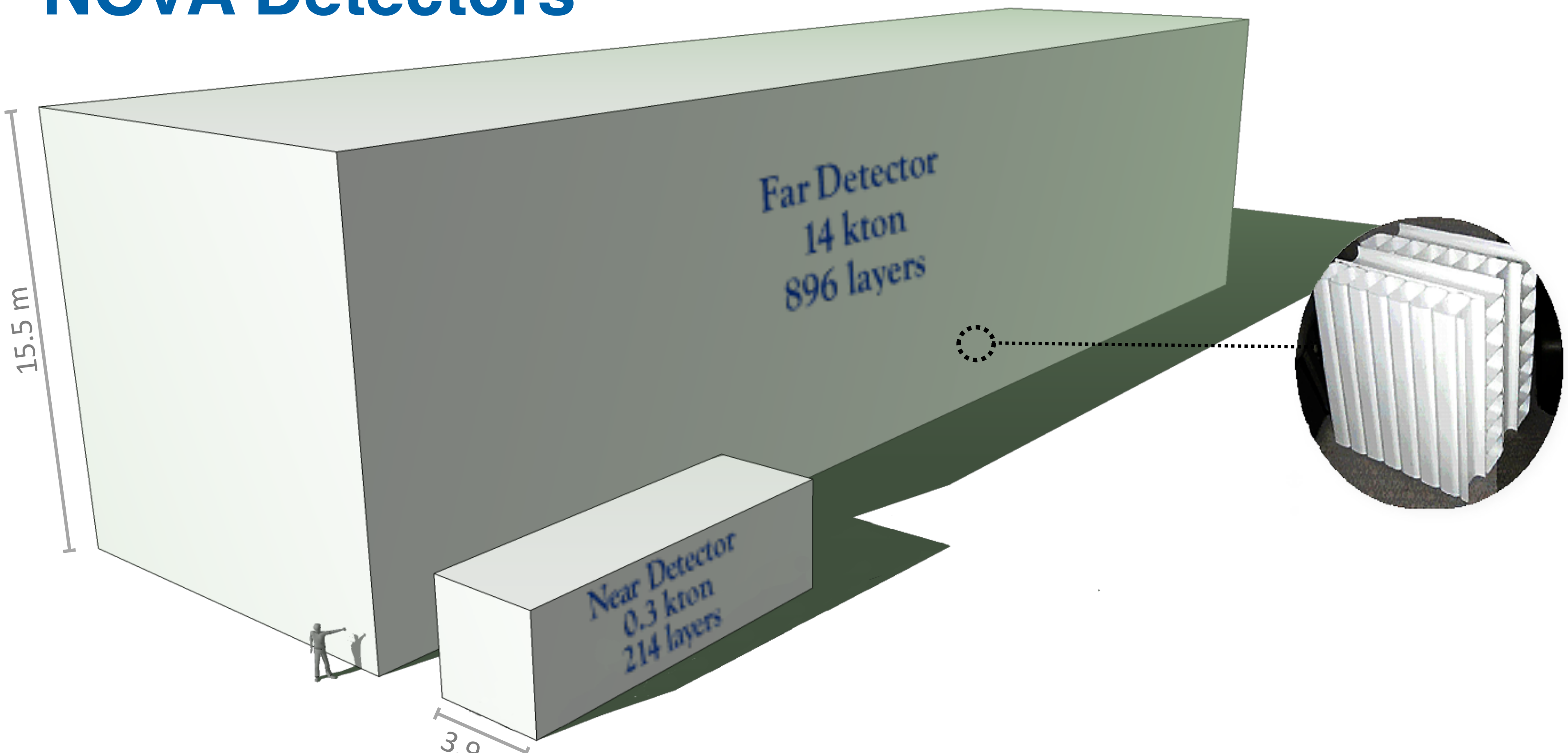
NuMI Beam

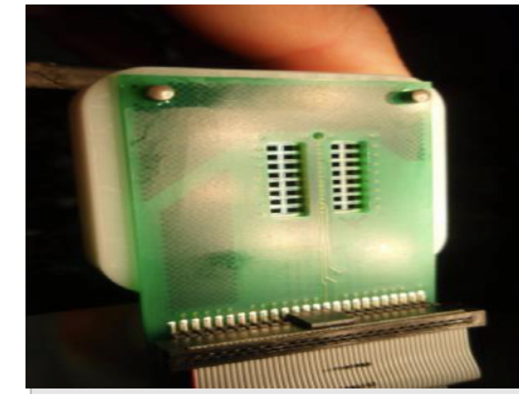


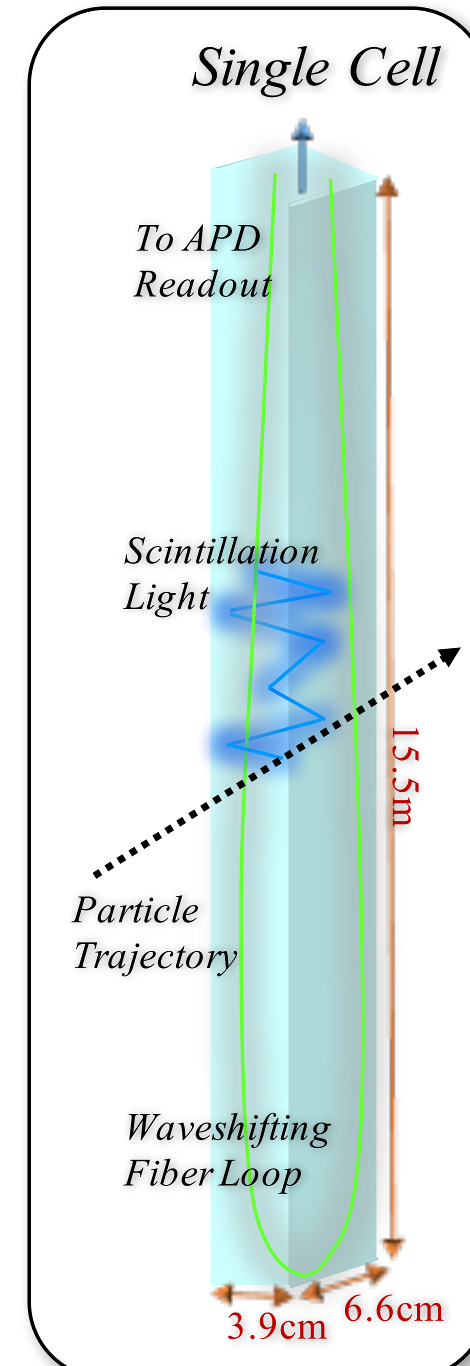
- 120 GeV protons extracted from the Main Injector at Fermilab in 10 μs spills
- Magnetic focusing horns allow selection of charge sign for selecting a neutrino or anti-neutrino beam
- 14.6 milli-radians off-axis, narrow beam around oscillation maximum
- Beam 97.5% ν_μ with 0.7% ν_e and 1.8% wrong-sign contamination



NOvA Detectors







- Two functionally identical detectors
- Extruded plastic cells alternating vertical and horizontal orientation filled with liquid scintillator
- Charged particles passing through cells produce light which is collected.

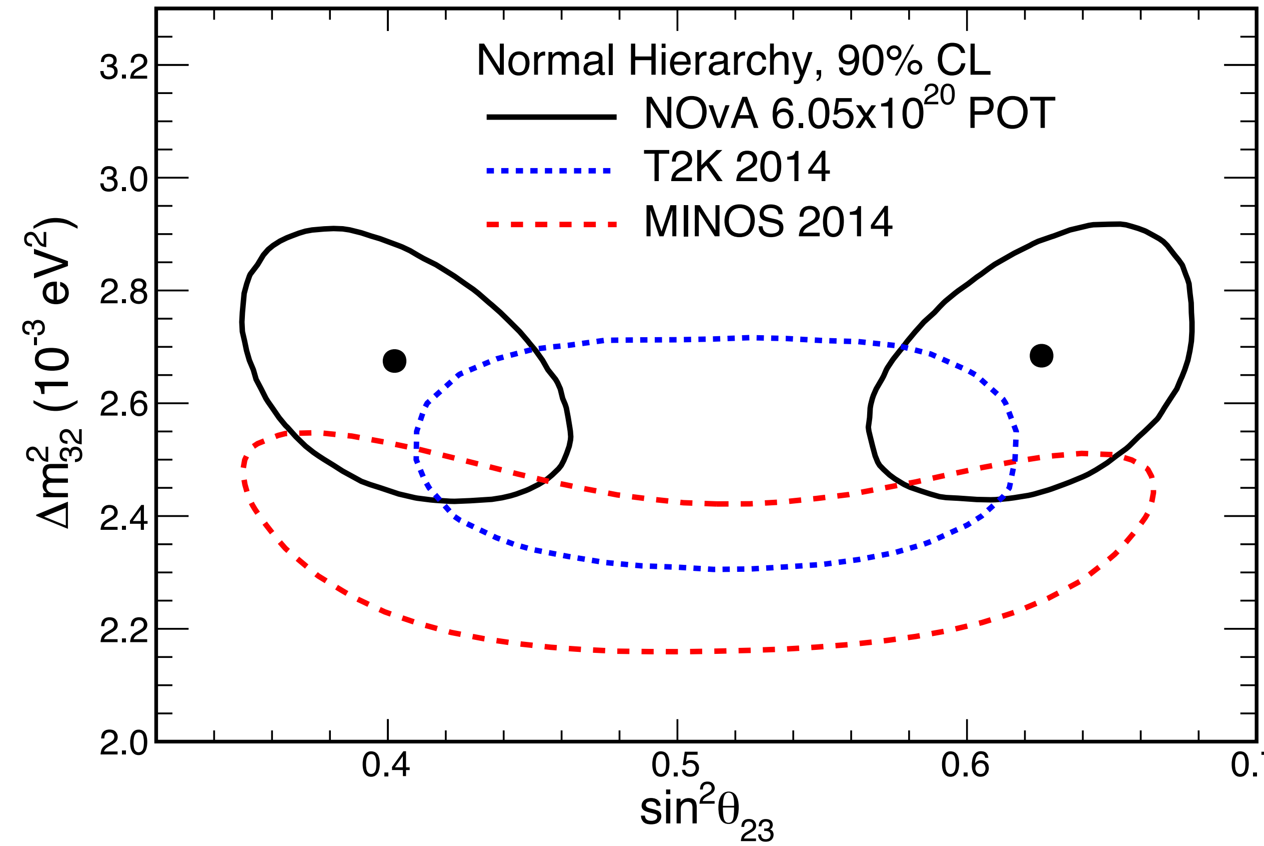
NOvA Results 2016

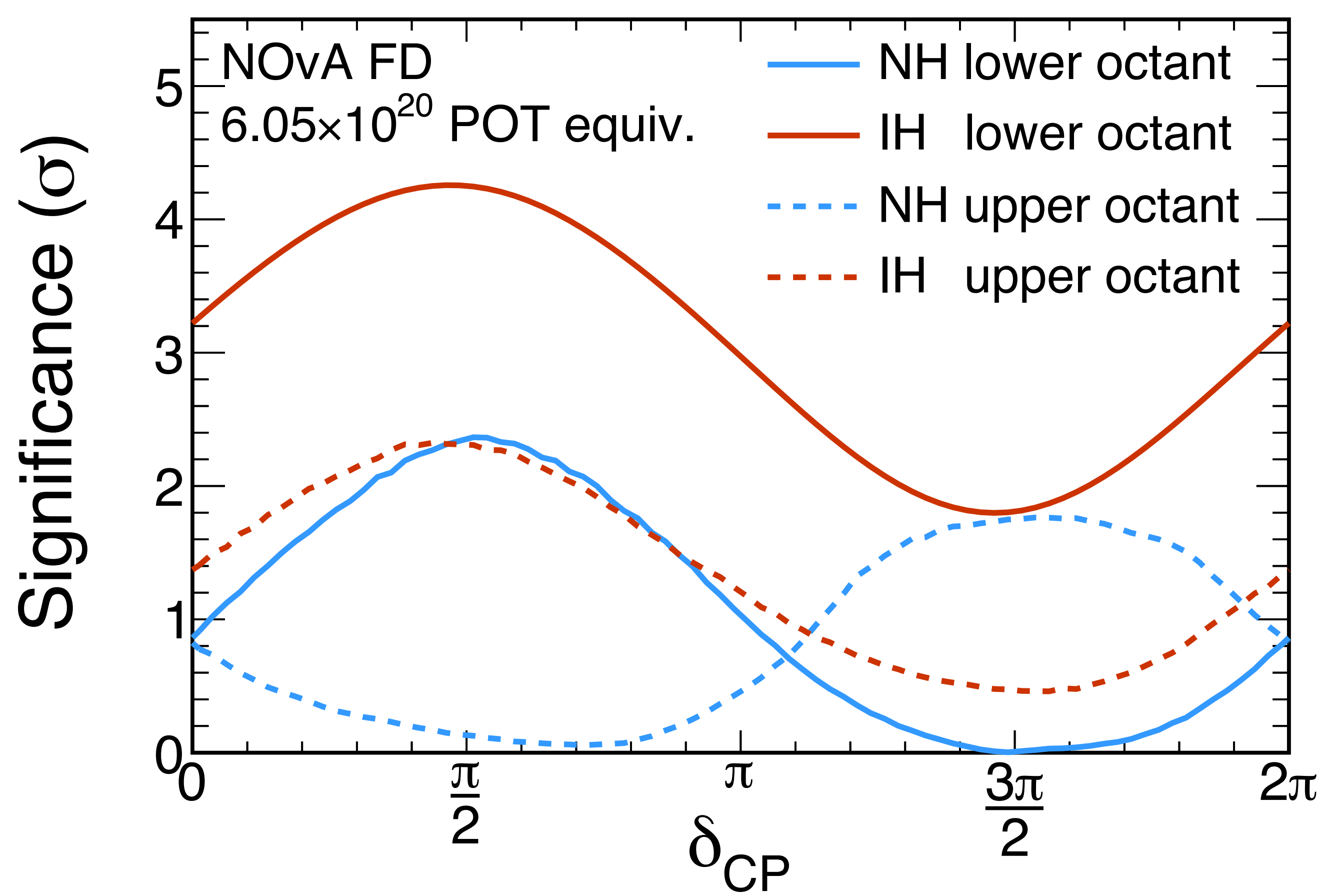
- 6.05e20 POT neutrino data

- **ν_μ —> ν_μ**
  - PRL 118, 151802
  - exclude maximum mixing at 2.5 σ

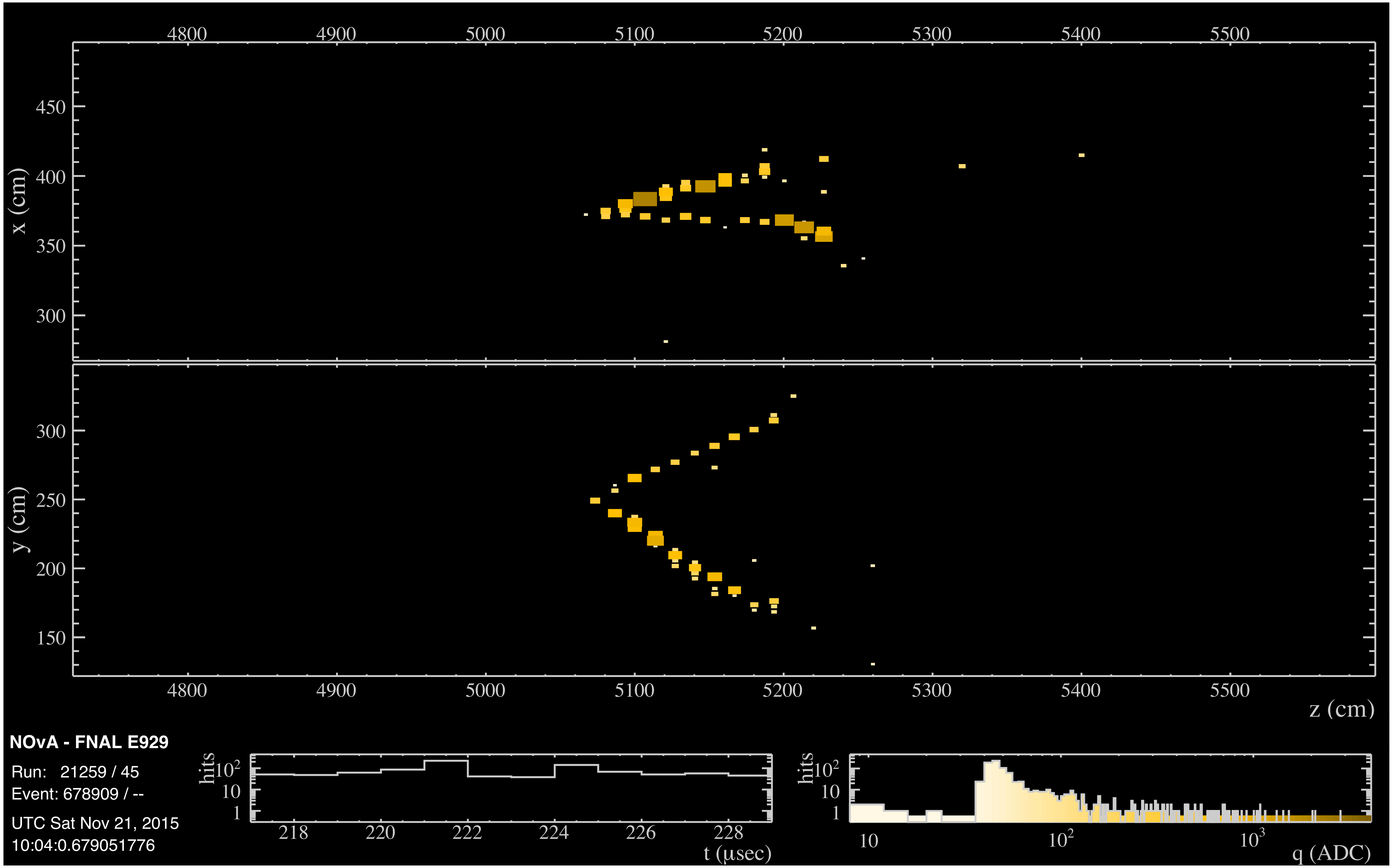
- **ν_μ —> ν_e**
  - PRL 118, 231801
  - Inverted hierarchy, lower octant is excluded at > 93% C.L.
  - first implementation of a convolutional neural network in a HEP result

- **Sterile neutrinos**
  - arXiv:1706.04592
  - No evidence in NC disappearance search



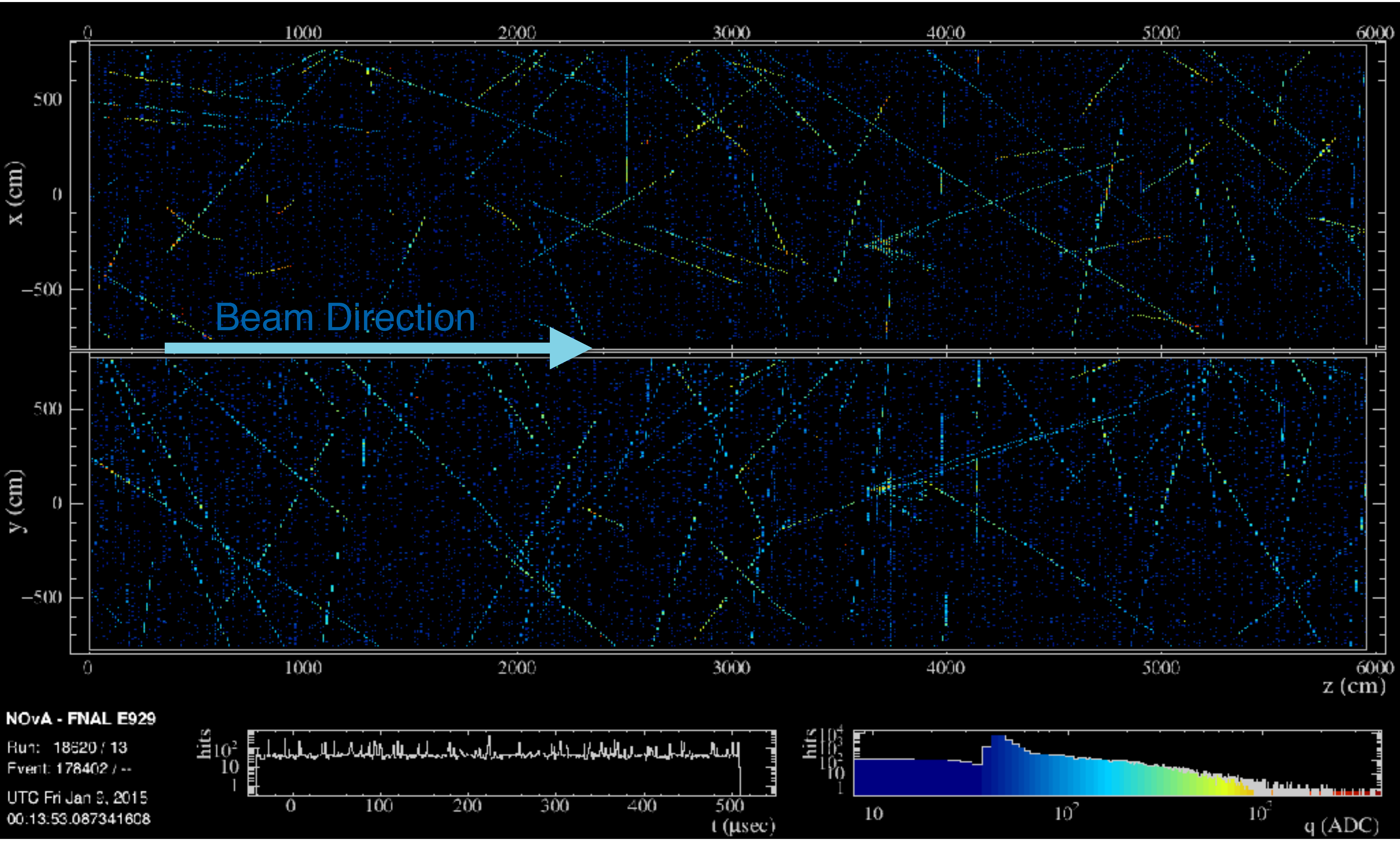


The Challenge



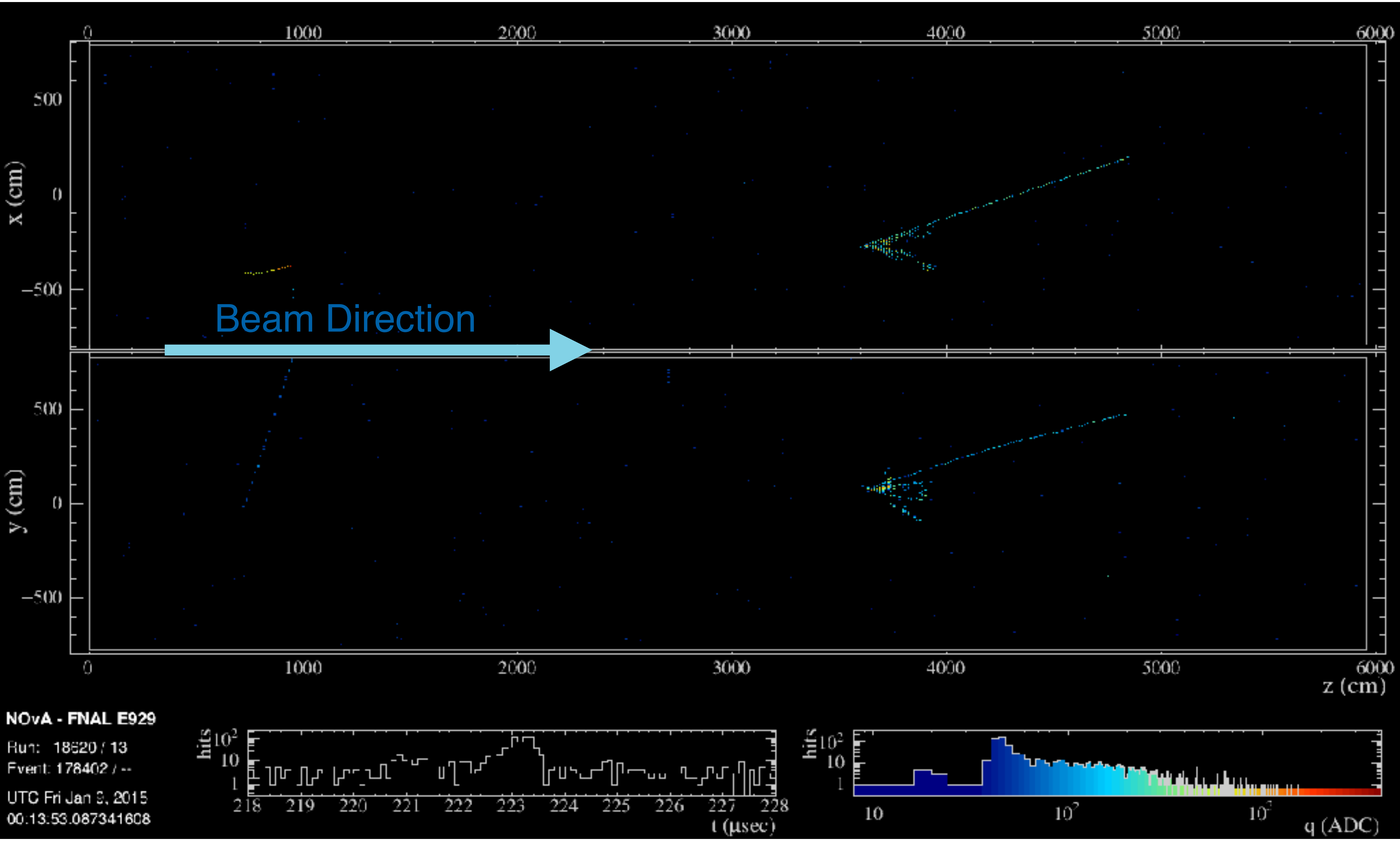
Far Detector 550 μs Readout Window

Cell hits colored by charge deposition



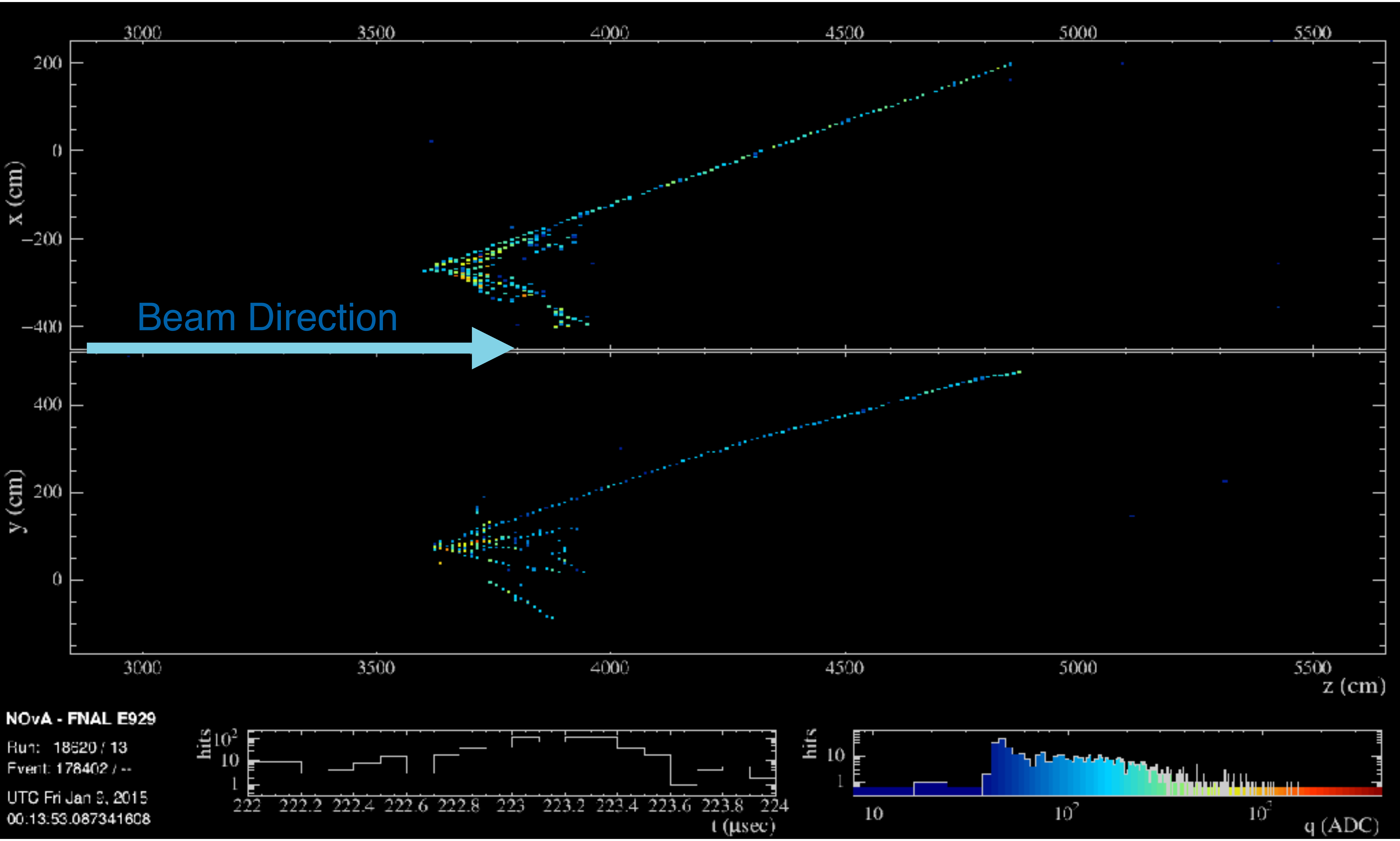
Far Detector 10 μs NuMI Beam Window

Cell hits colored by charge deposition



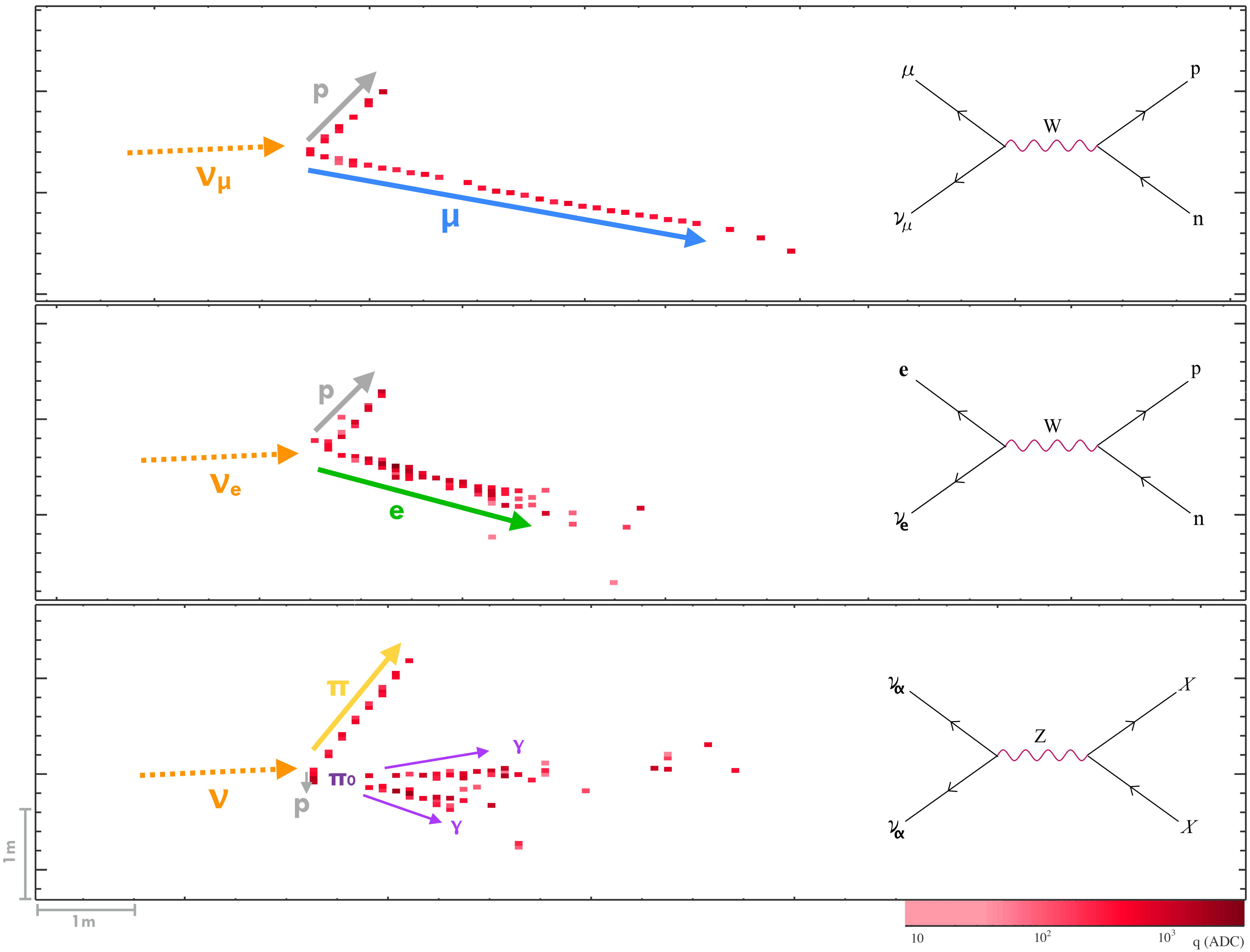
Far Detector Neutrino Interaction

Cell hits colored by charge deposition



Fermilab

Observing Neutrino Interactions



ν_μ

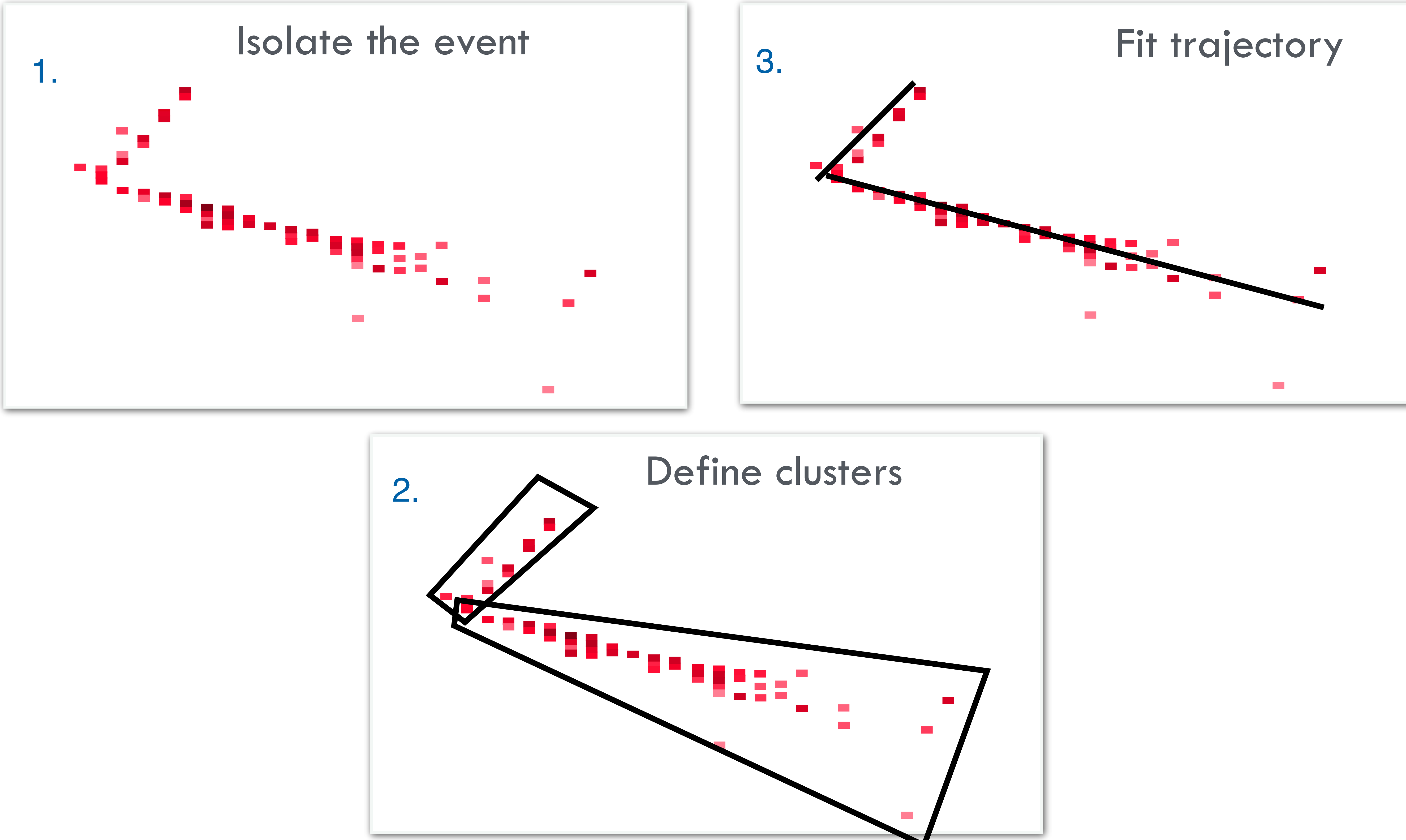
Charged Current

ν_e

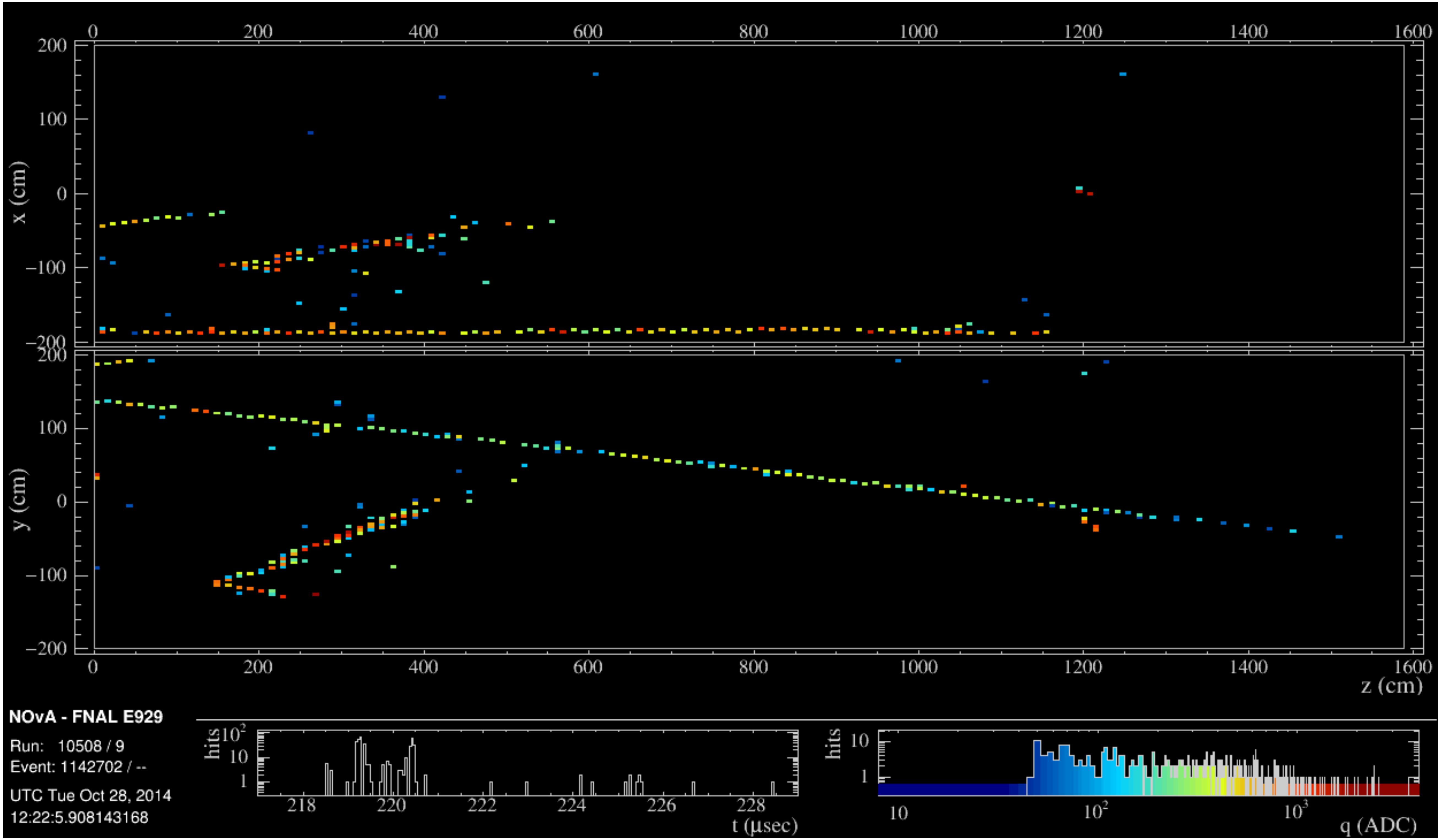
Charged Current

Neutral Current

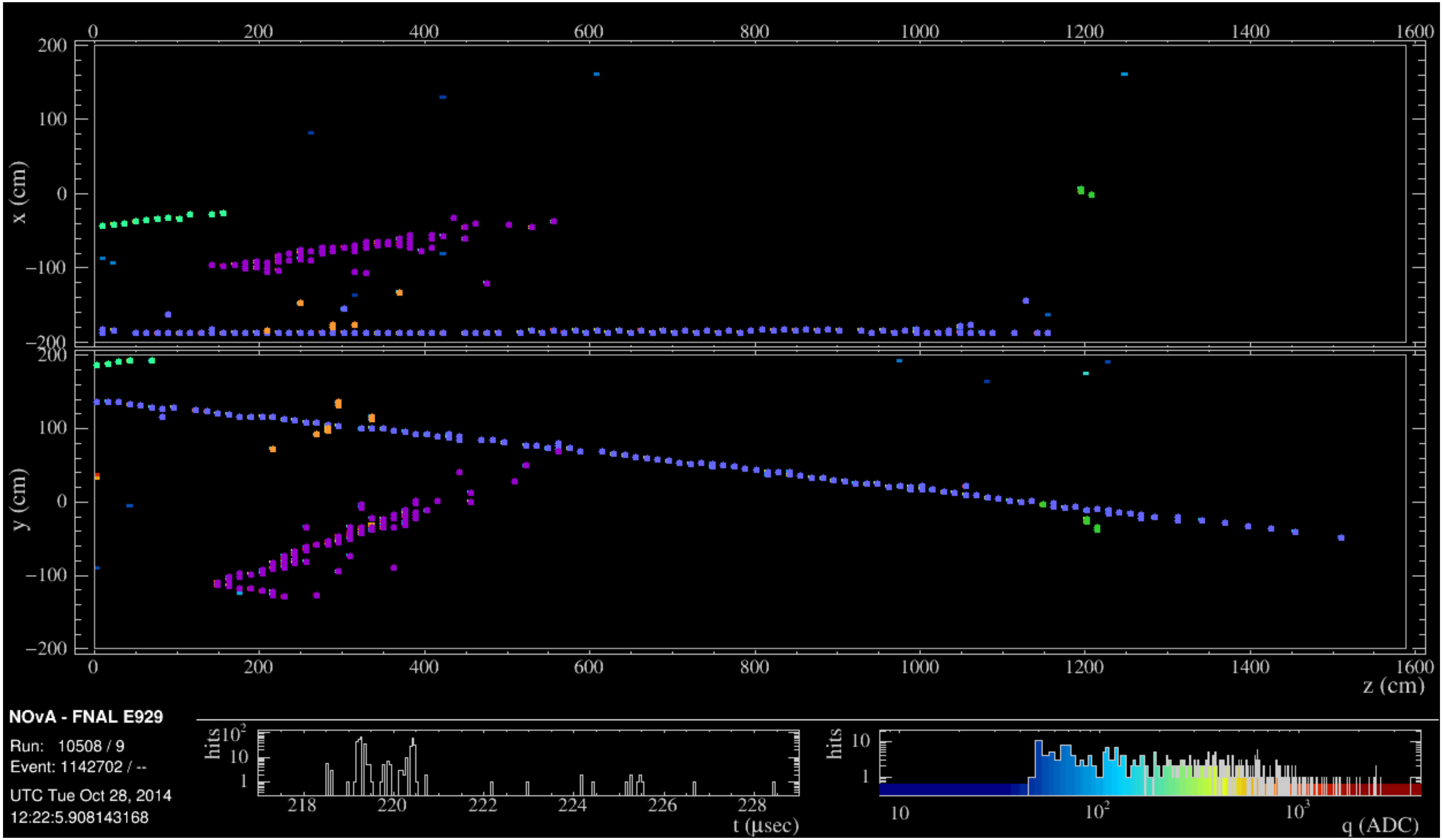
Event Reconstruction in NOvA



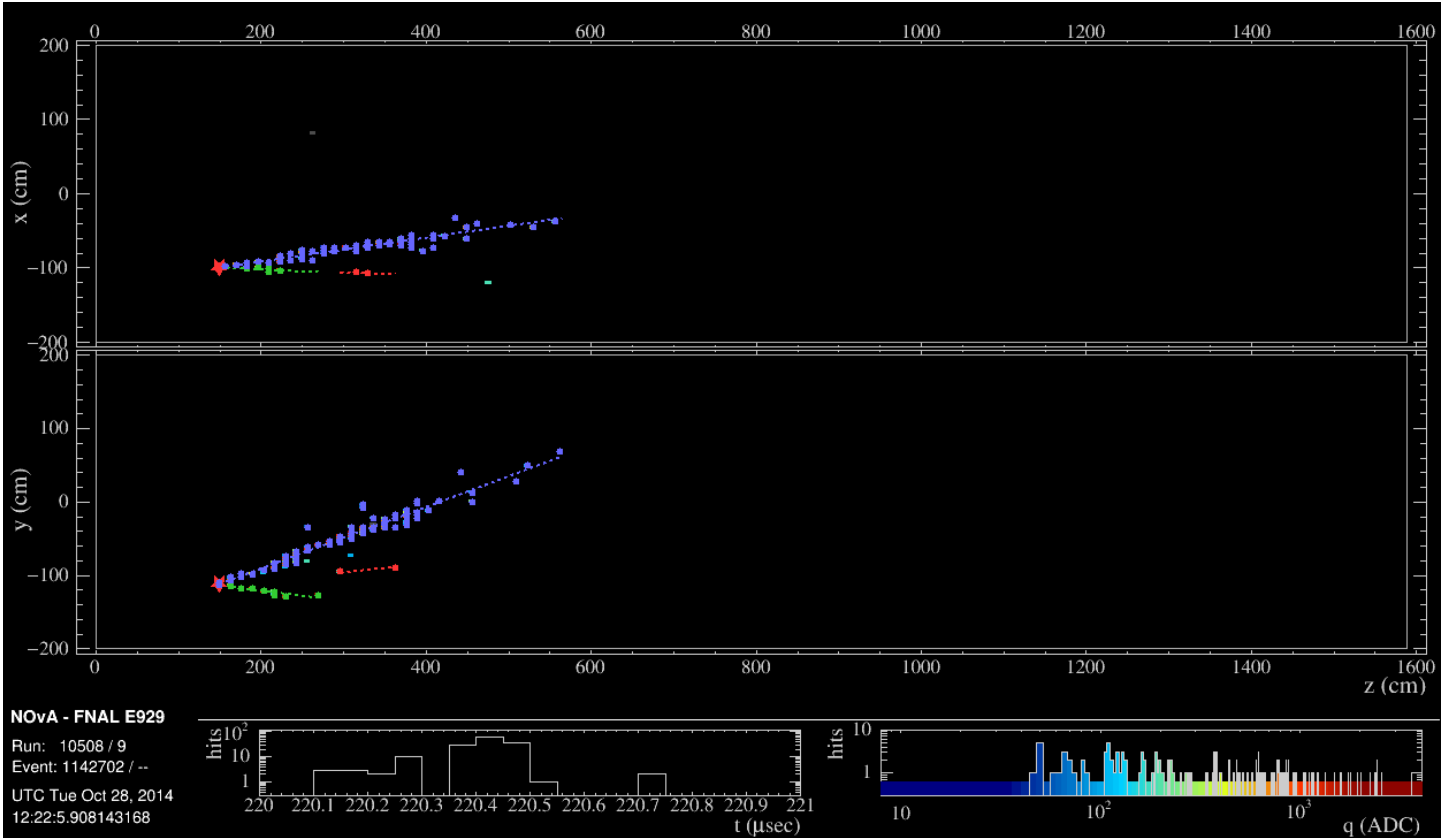
10 µs Near Detector Beam Window



After Space-Time Separation



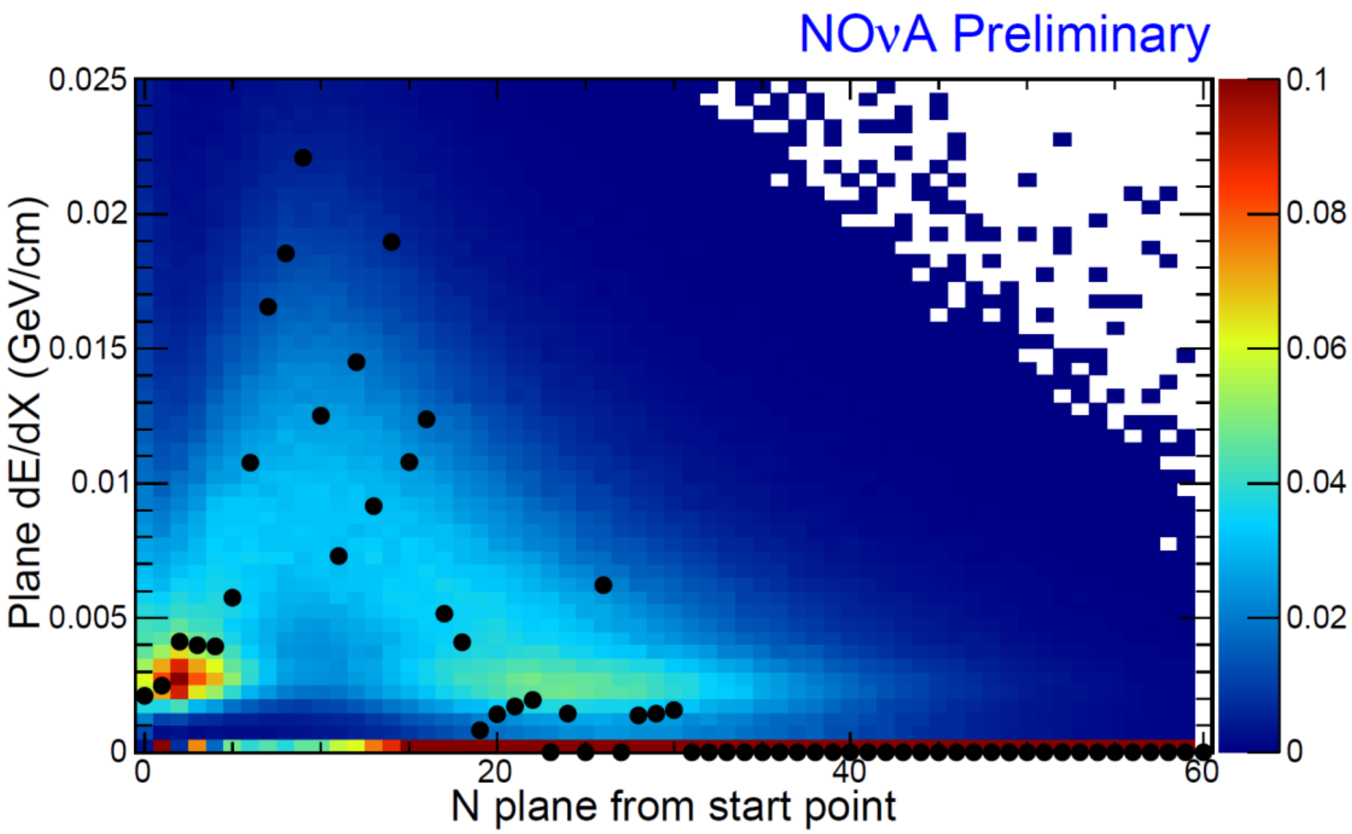
After Vertex and Prong Formation

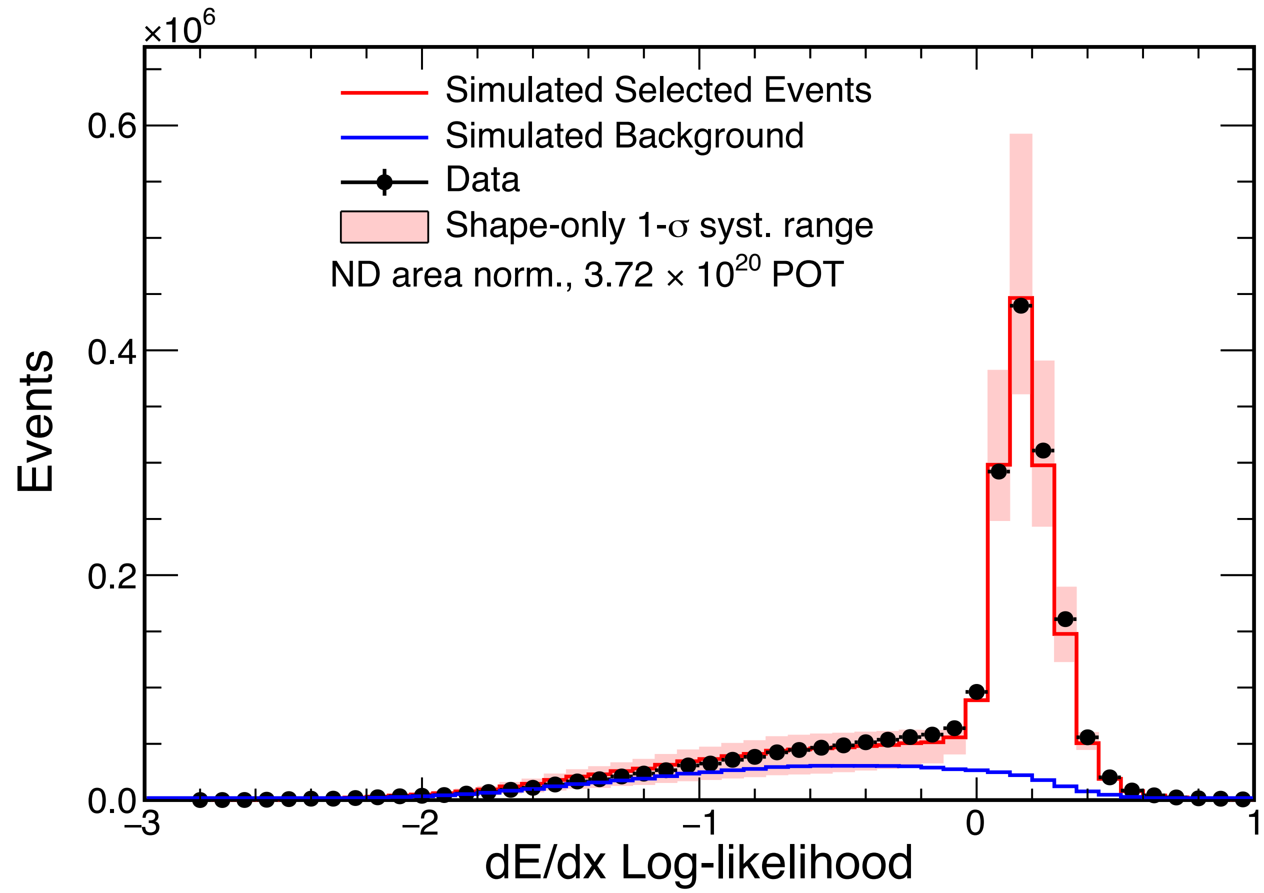


Event Classification before Deep Learning

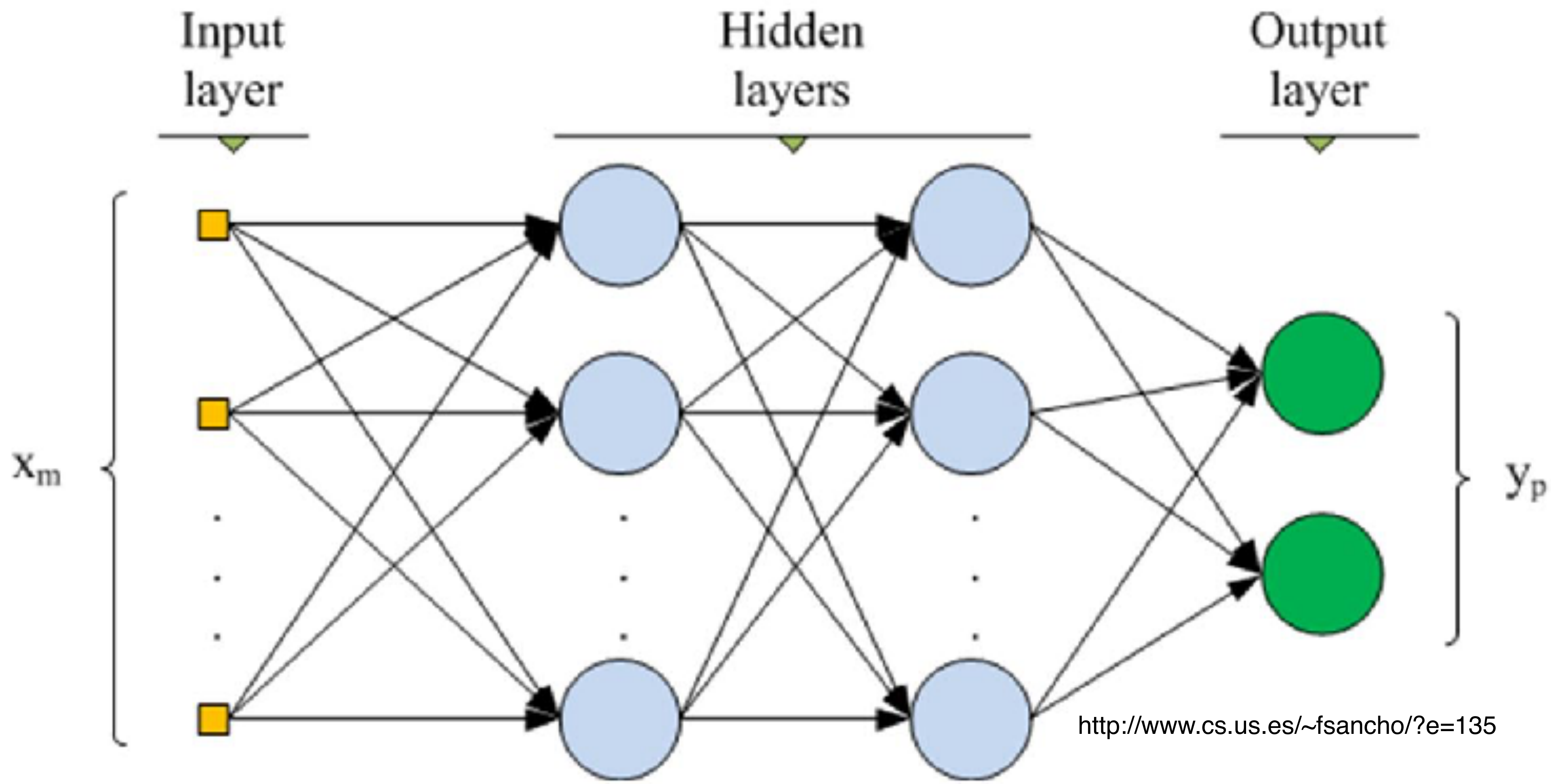
- Likelihood Identifier (LID)
  - Compare longitudinal and transverse dE/dx in leading shower to templates for different particle hypotheses
  - Build neural net from these inputs and reconstructed quantities.
  - Identifies electron neutrinos

- ReMID
  - Build a KNN classifier from four reconstructed quantities related to muons (length, dE/dx, scattering)
  - Identifies muon neutrinos





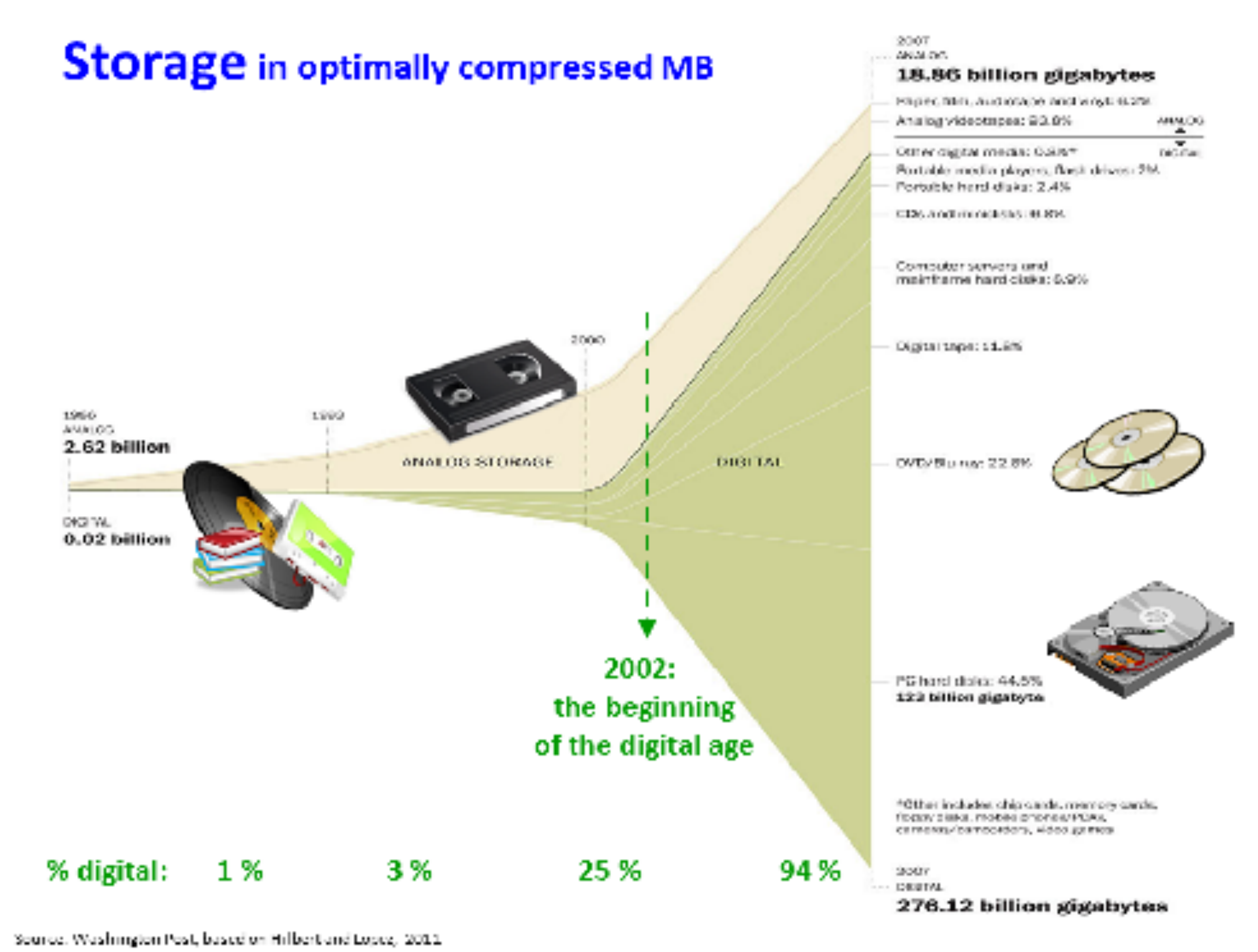
Neural Networks



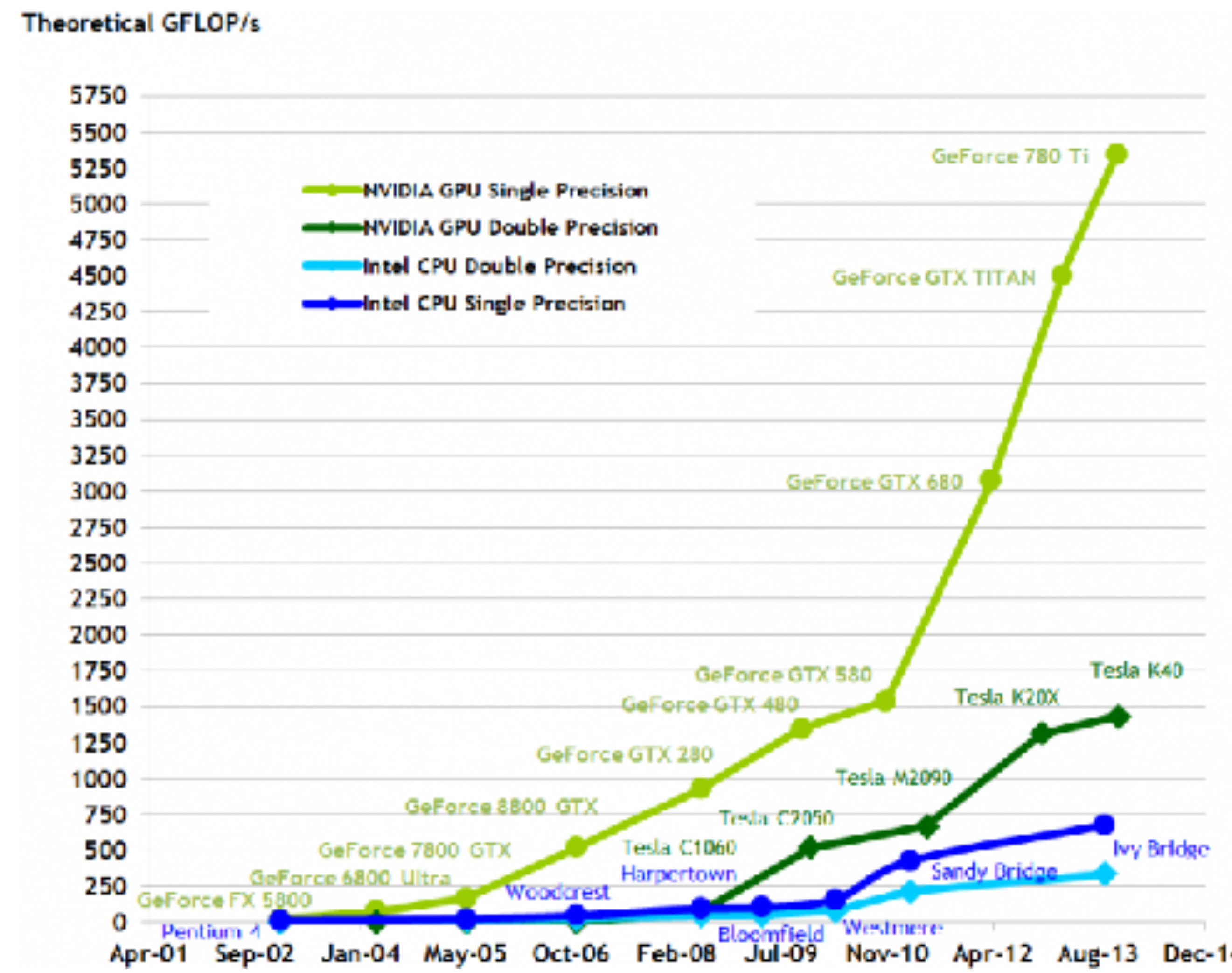
- Neurons with activation and propagation functions, weights between inputs
- Loss function to calculate network performance
- Regularization of weights to avoid overtraining
- Back propagate errors in loss function to nodes, update weights with gradient decent

Enter Deep Learning

Can we improve the networks by making them deeper to extract increasing complex features?



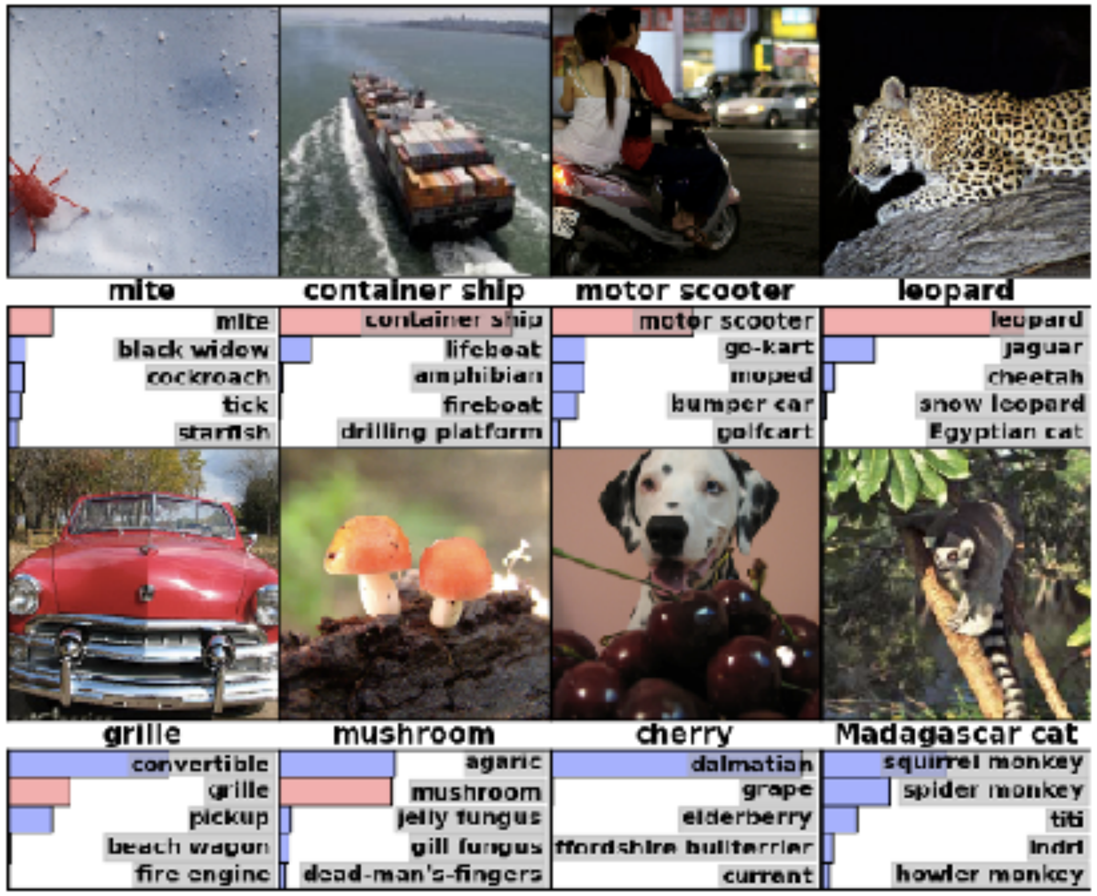
http://www.martinhilbert.net/worldinfocapacityppt-html/



http://www.dual.sphysics.org/index.php/gpu/

Advancement in GPUs and data storage make this possible

Computer Vision

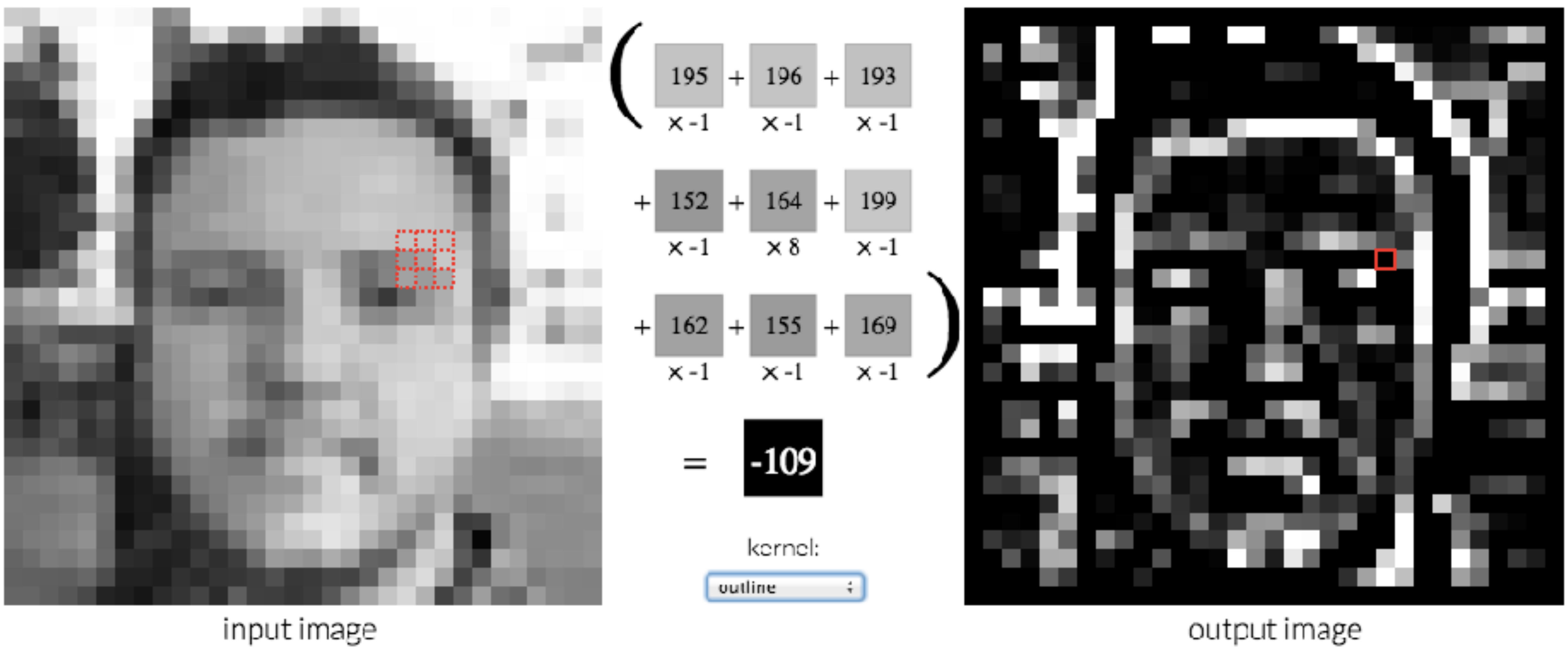


- Deep neural networks for object recognition with pixel inputs to network
- First GPU trained network at 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge) reduced classification error rate from 26.2% to 15.4%
- Now achieving super human performance (<5%) with image net dataset

Convolutional Neural Networks

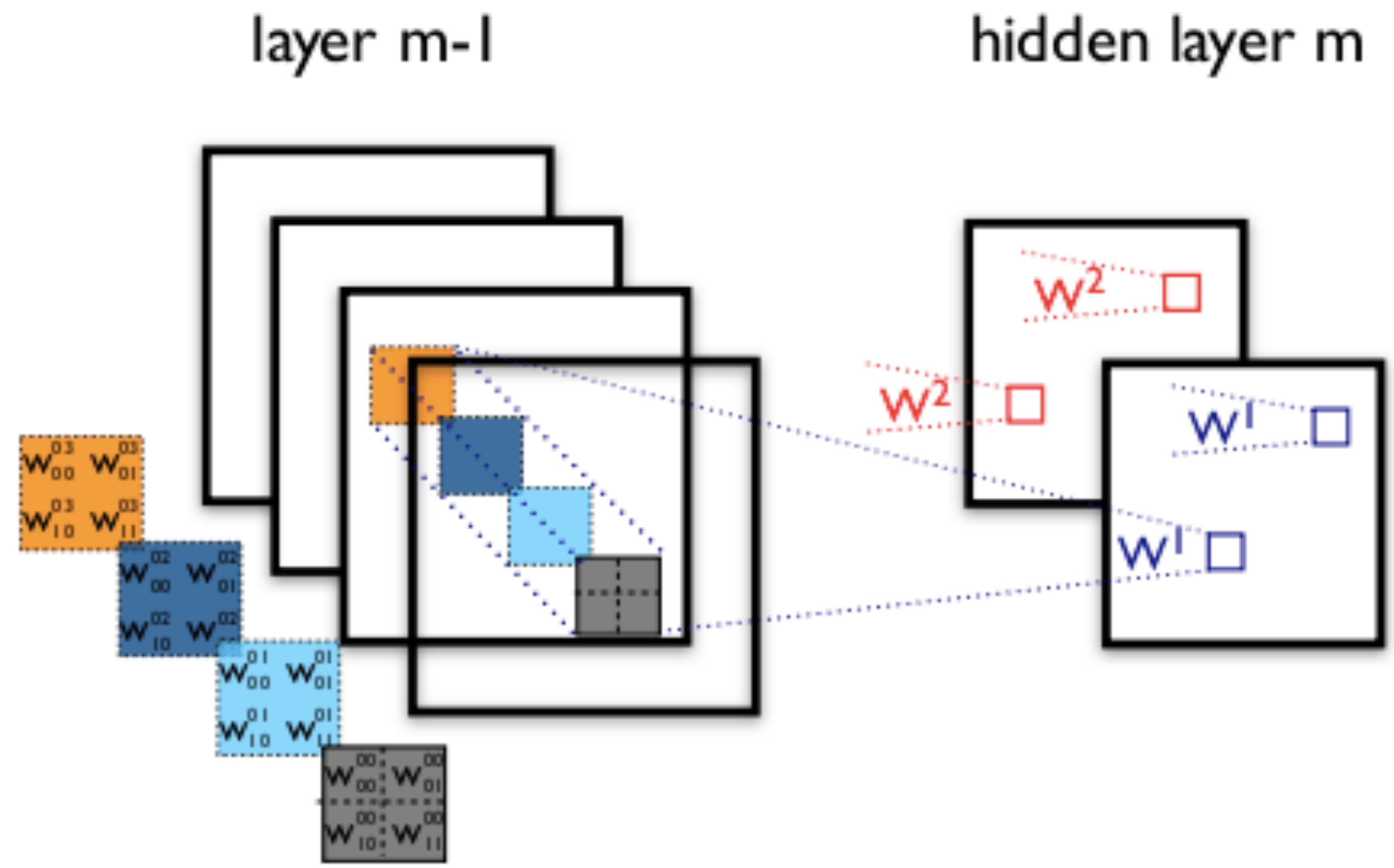
- Instead of training with a weight for each pixel, convolve kernel operations across the image to extract features

- Inspired by the visual cortex

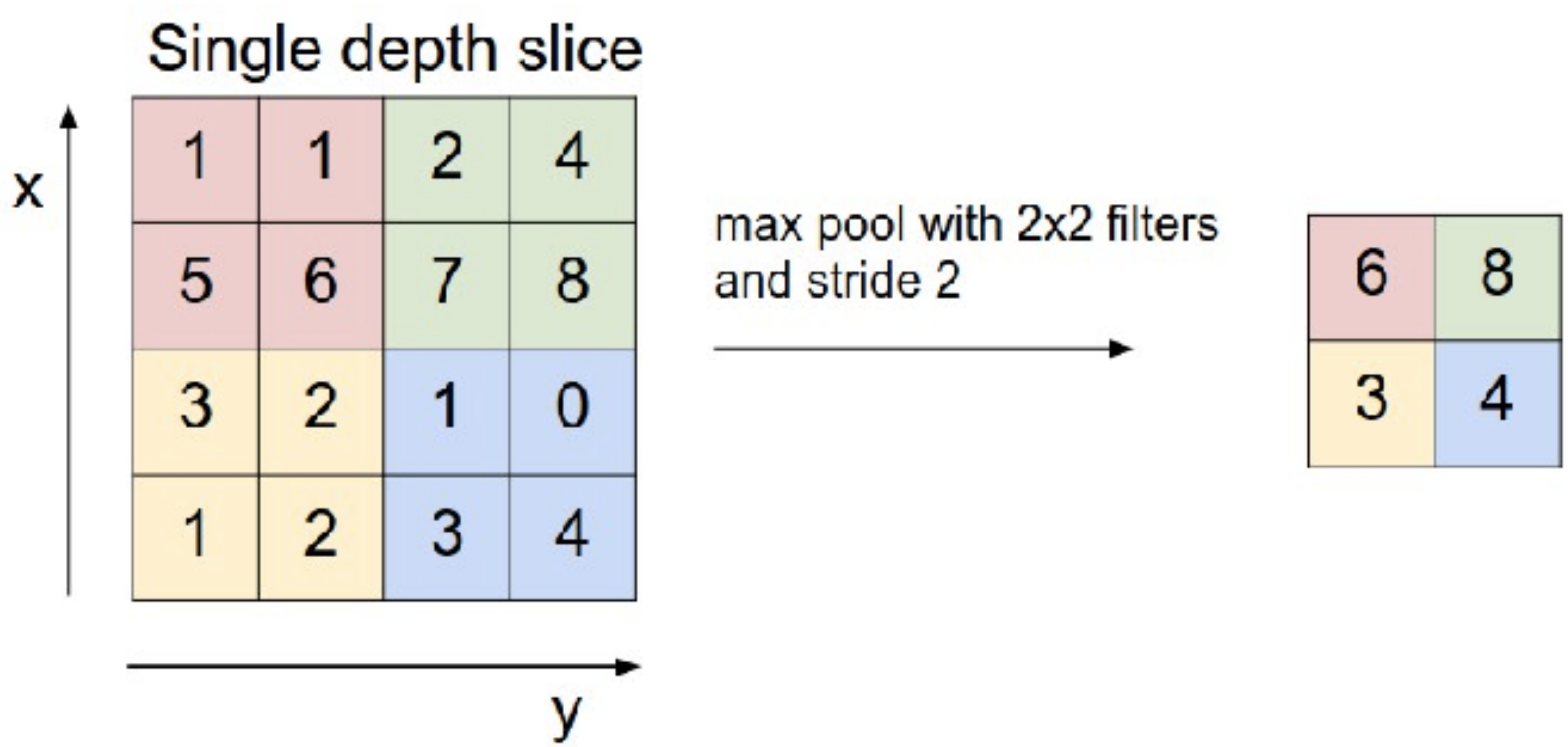


CNN Components

- Convolutional layers train an array of kernels to output feature maps



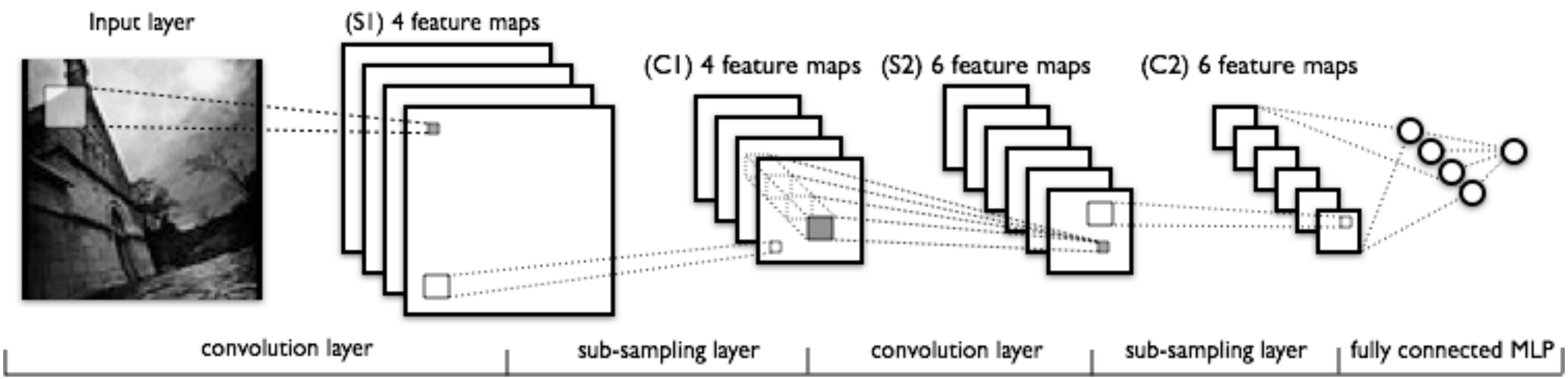
- Pooling layers downsample the feature maps by taking the average or maximum value from image patches



Fermilab

Early CNNs

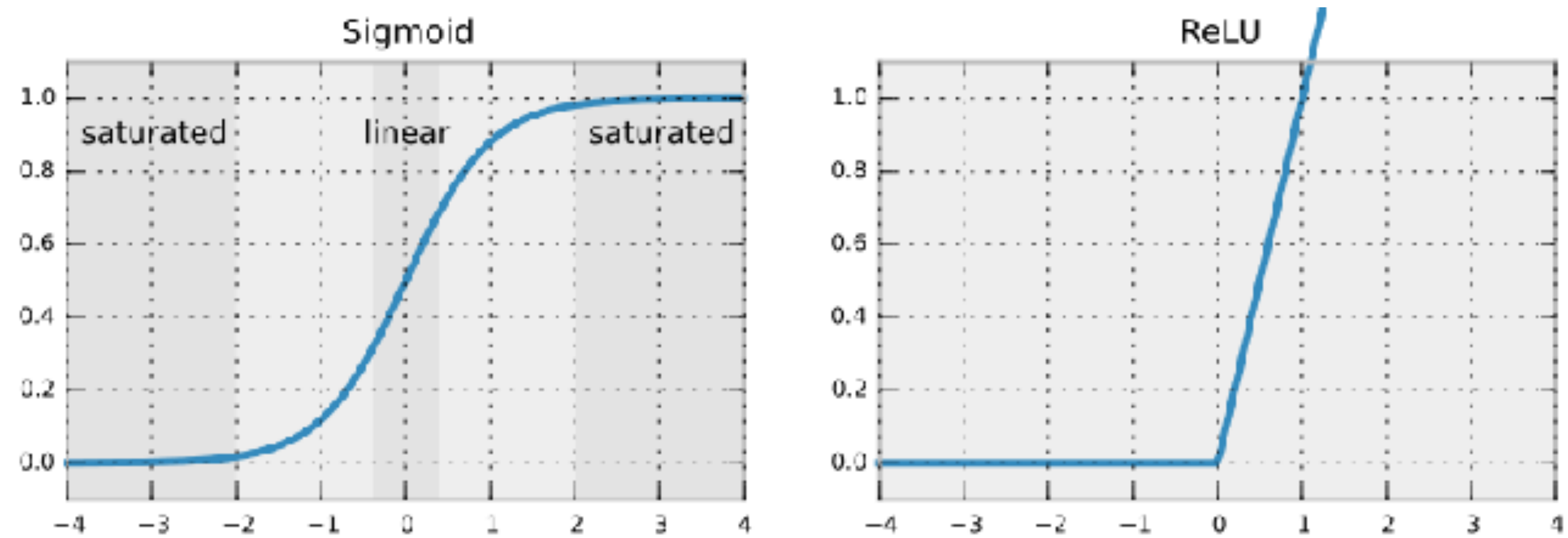
- 1989 “LeNet”

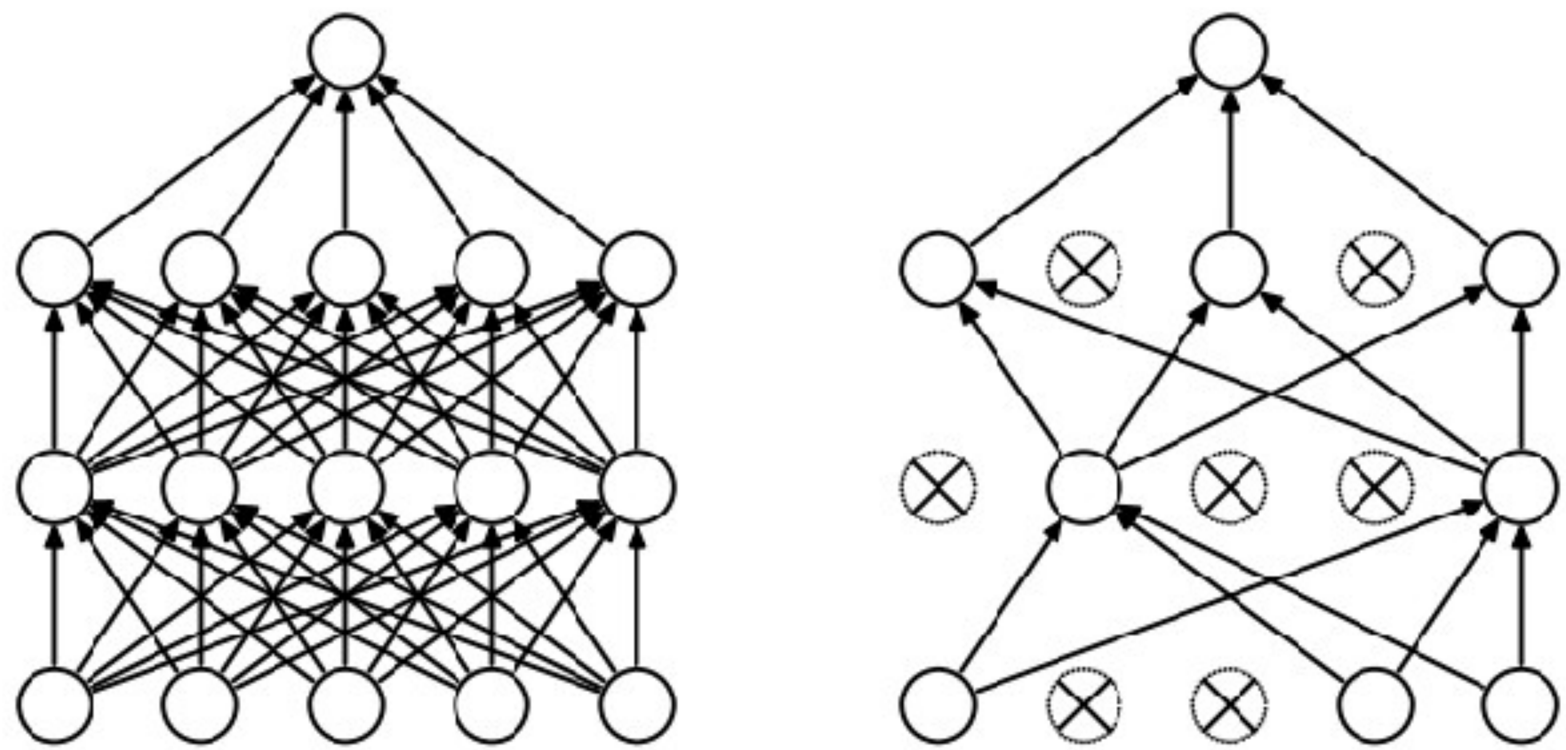


http://deeplearning.net/tutorial/lenet.html http://yann.lecun.com/exdb/lenet/

Training Advancements

- Better activation functions to avoid saturation
- Dropout layers to prevent over training
- Stochastic gradient decent



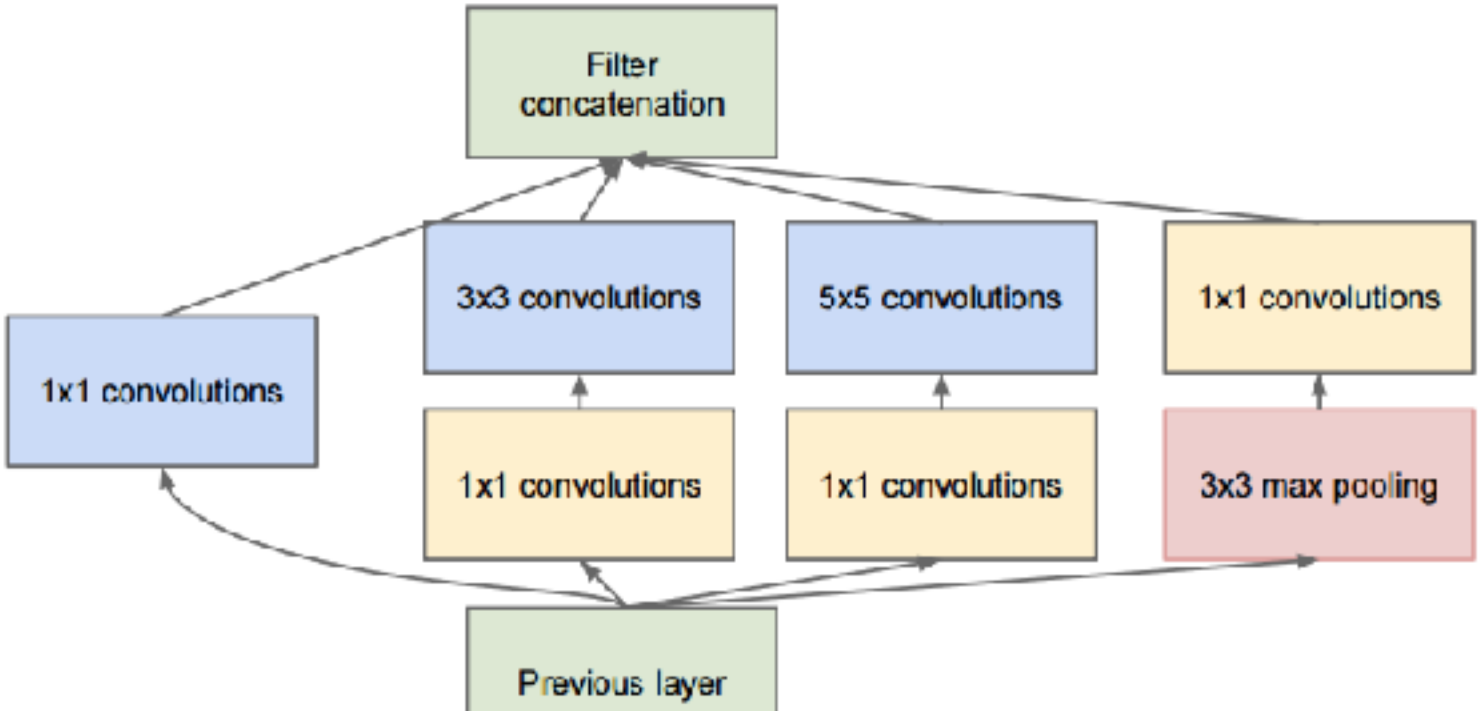


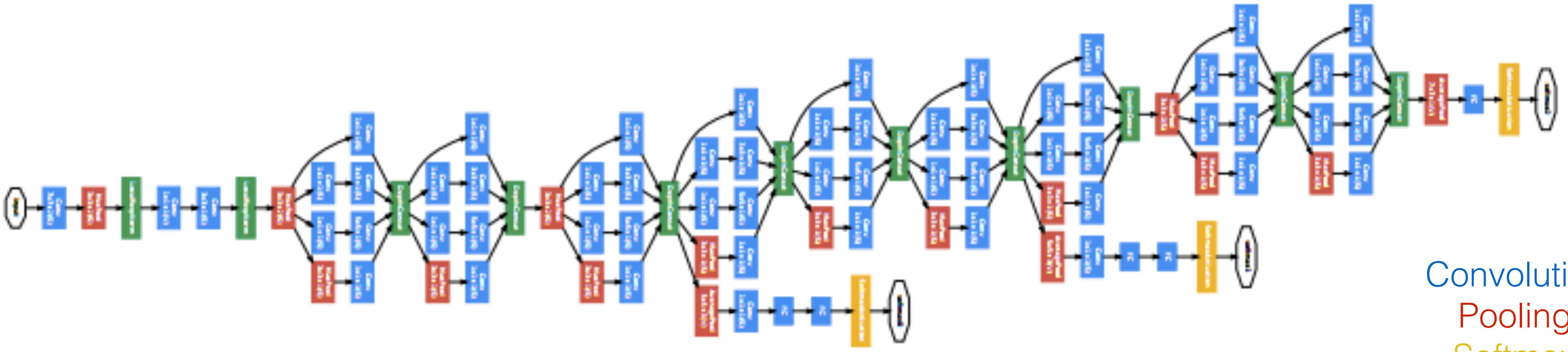
(a) Standard Neural Net

(b) After applying dropout.

Advanced CNNs

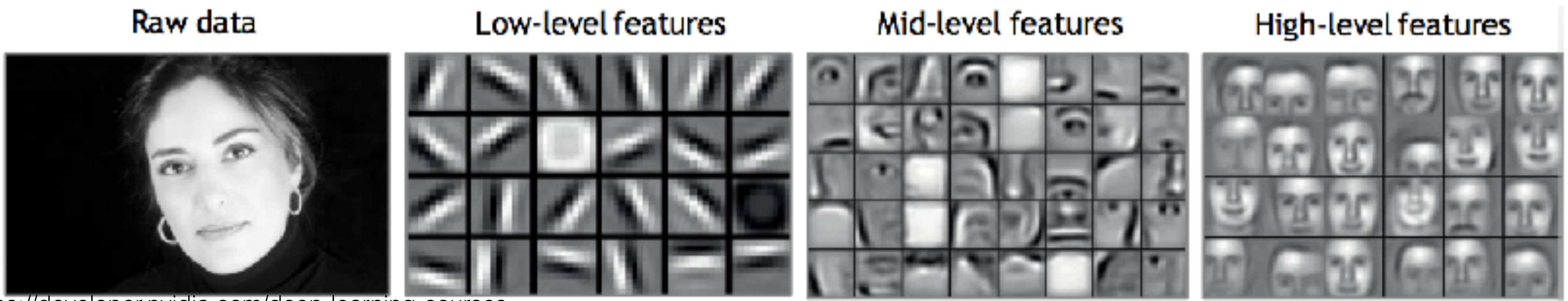
- 2014 GoogLeNet
- C. Szegedy et al., arXiv:1409.4842
- “Network-in-Network”
- Uses kernels of serval sizes
- Number of maps controlled by series of 1x1 convolutions



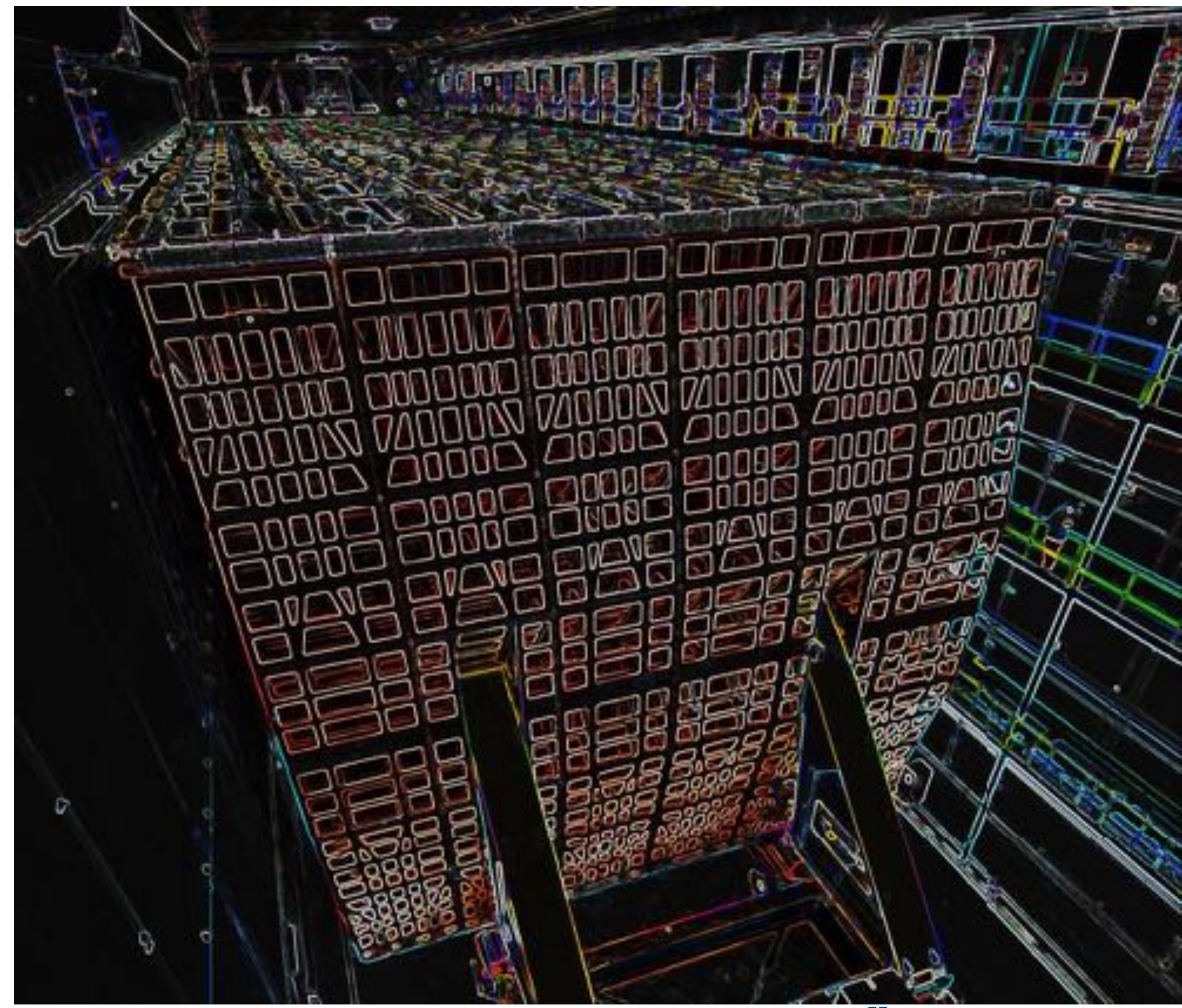


Convolution
Pooling
Softmax
Other

Deep Learning in NOvA

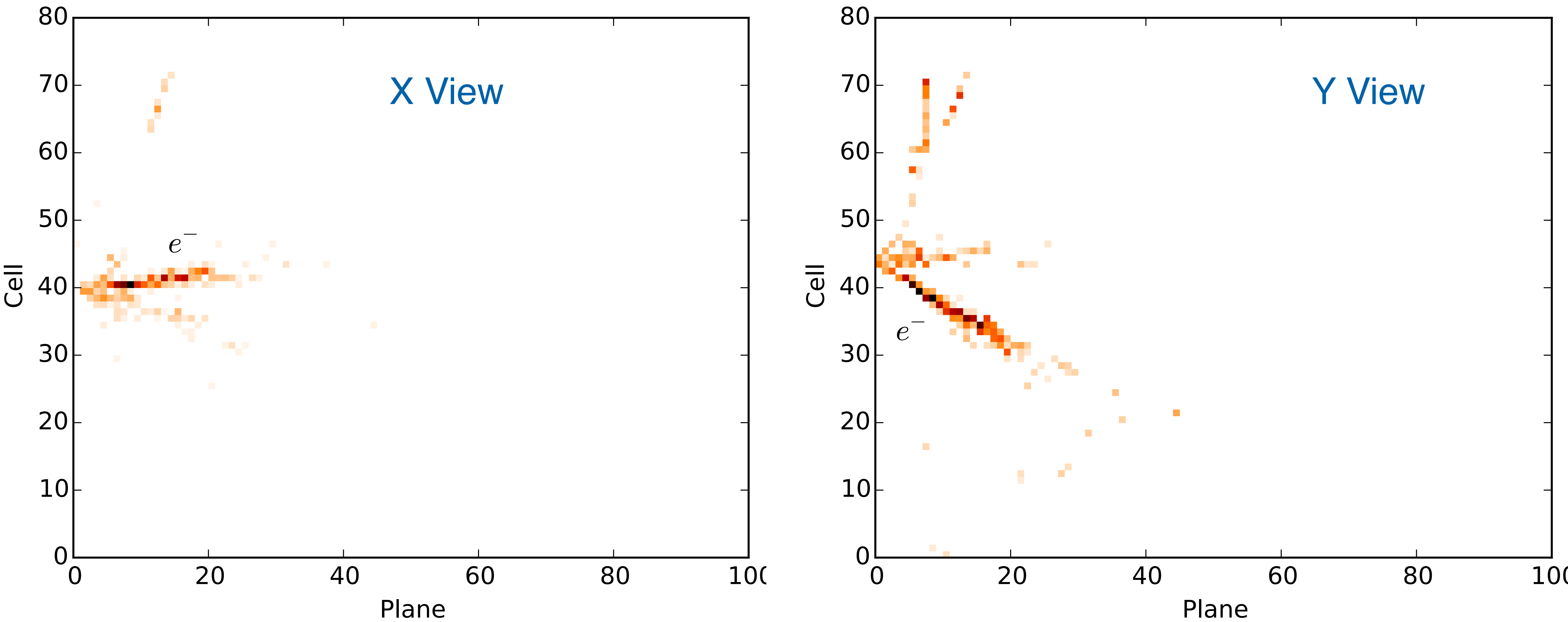


https://developer.nvidia.com/deep-learning-courses



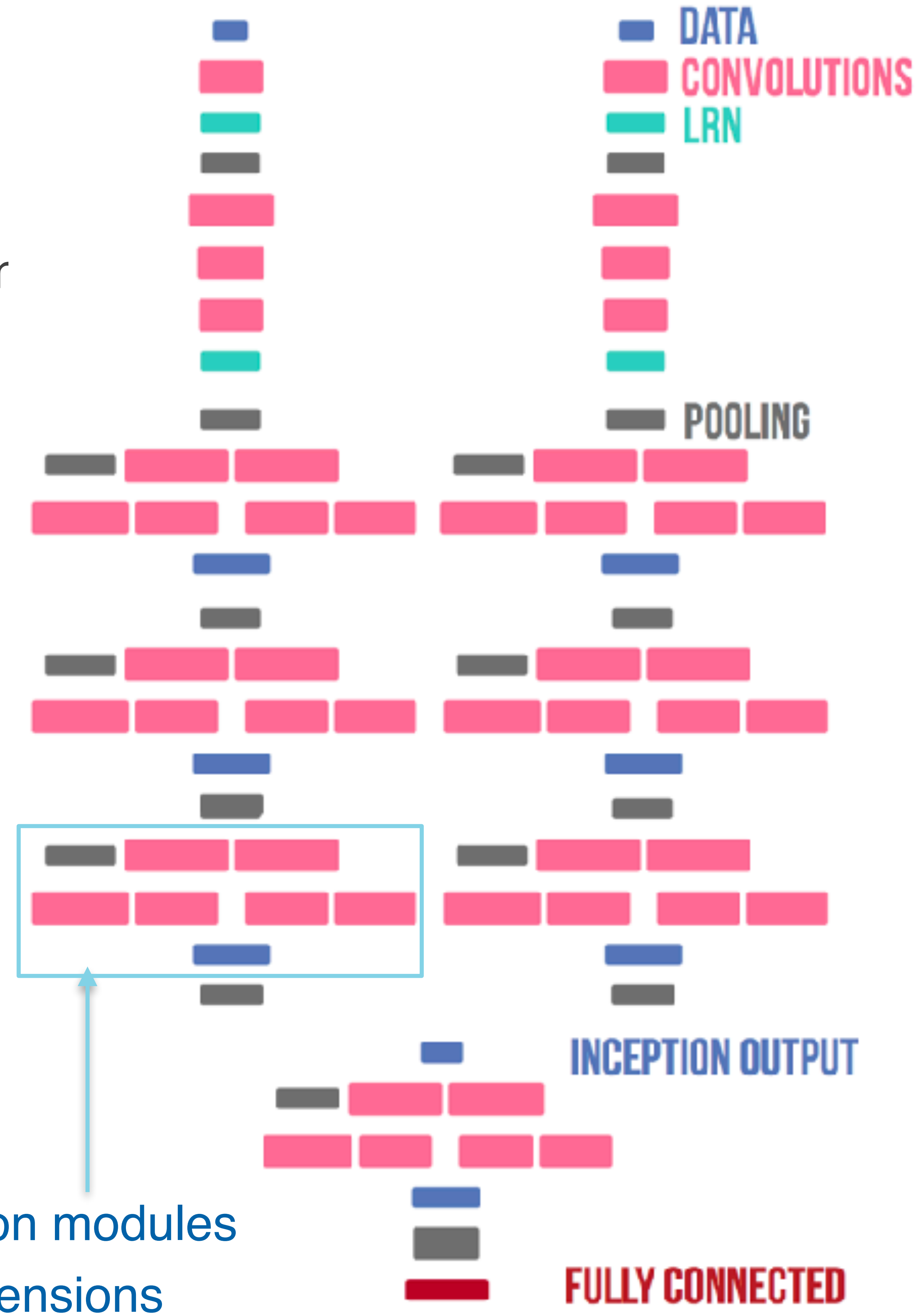
Input Images

- Produce a pair of pixel maps for the X and Y view of each interaction
- Input images are 80 cells by 100 planes
- Sparse images compared to computer vision field



Network Architecture

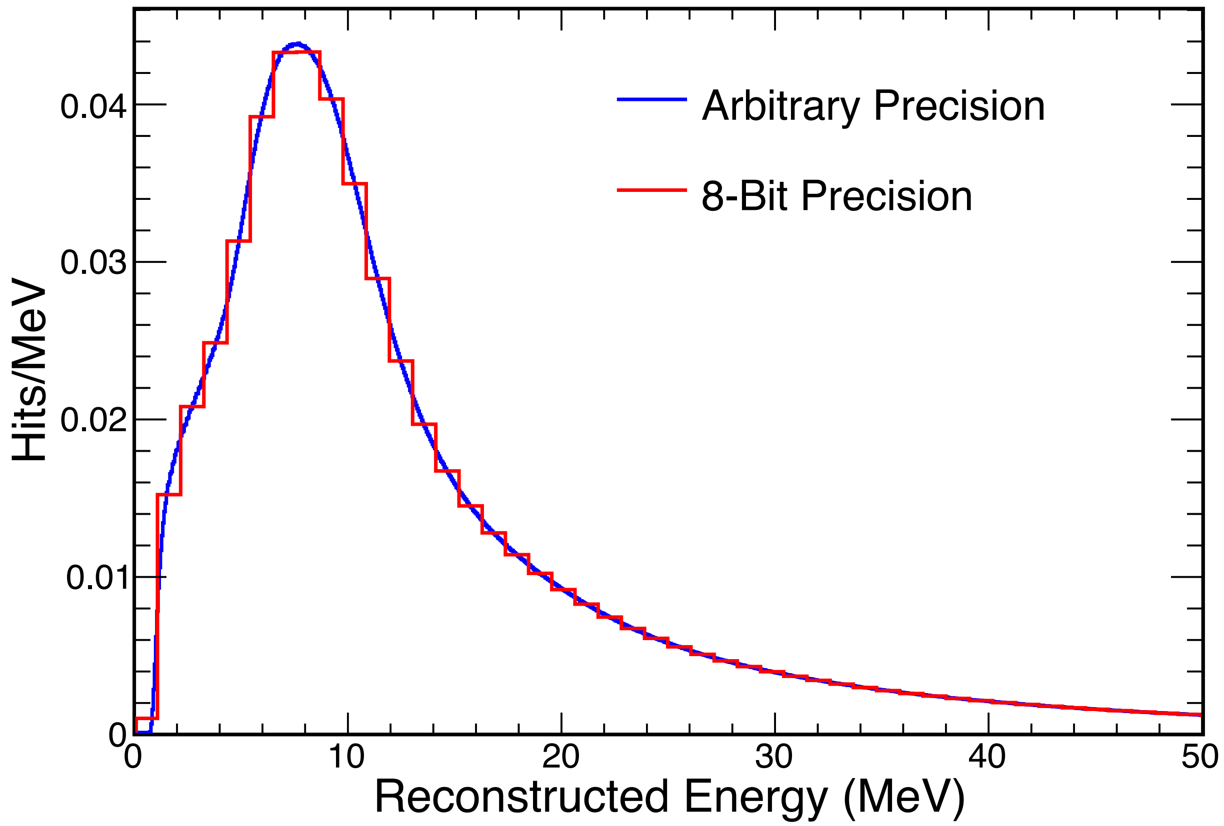
- Architecture adapted from GoogLeNet
  - Each event view processed separately and ther merged
  - Sparse images allow for shallower network
- Convolutional Visual Network (CVN)
- Pixel intensities varied with 1% gaussian noise
- Images randomly flipped along cell axis
- Output classifies neutrino interaction type (ν_μ,ν_τ,ν_e,NC, cosmic)

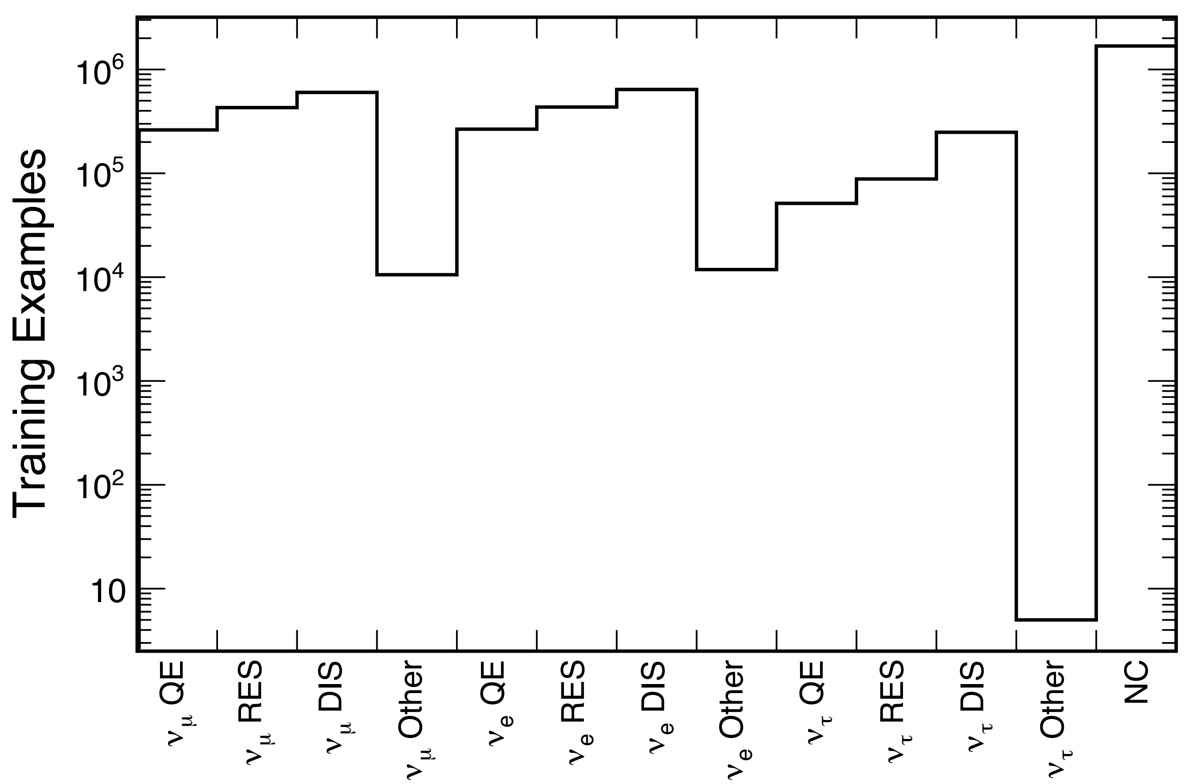


Network-in-network inception modules with kernels of multiple dimensions

Training the Network

- Trained on 4.7 million minimally preselected events, distributed among all neutrino interaction categories (80% training, 20% testing)
- Second training stage added cosmic data events
- Calibrated energy depositions reduced to 8 bit precision to compress files at no loss of information
- Network implemented and trained in the Caffe Framework (Y. Jia et al., arXiv: 1408.5093)
- Use two k40s within Fermilab Wilson Cluster, ~1 week wall time





Training the Network

- Trained on 4.7 million minimally preselected events, distributed among all neutrino interaction categories (80% training, 20% testing)

- Second training stage added cosmic data events

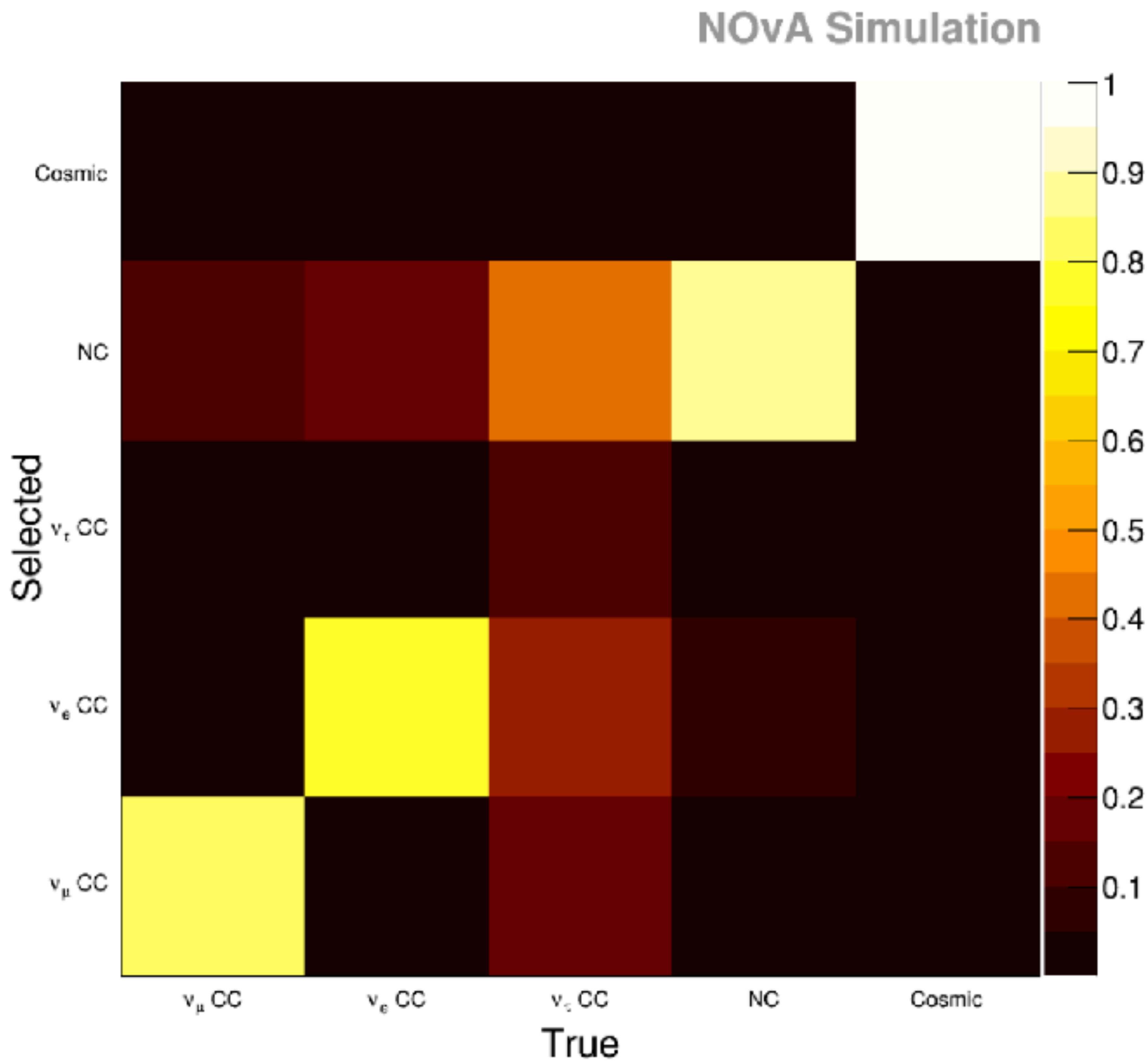
- Calibrated energy depositions reduced to 8 bit precision to compress files at no loss of information

- Network implemented and trained in the Caffe Framework (Y. Jia et al., arXiv: 1408.5093)

- Use two k40s within Fermilab Wilson Cluster, ~1 week wall time

Fermilab

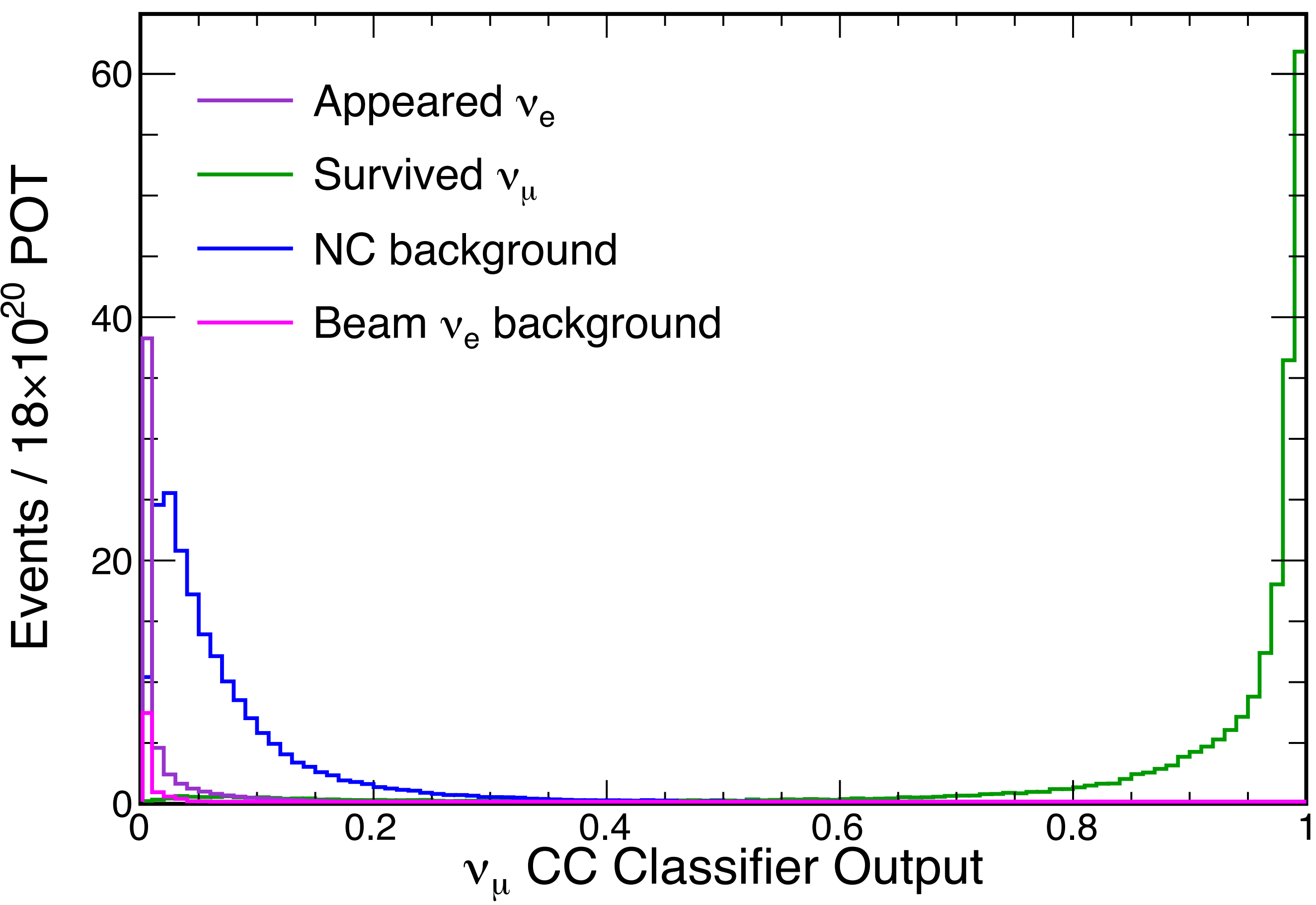
Classification Matrix

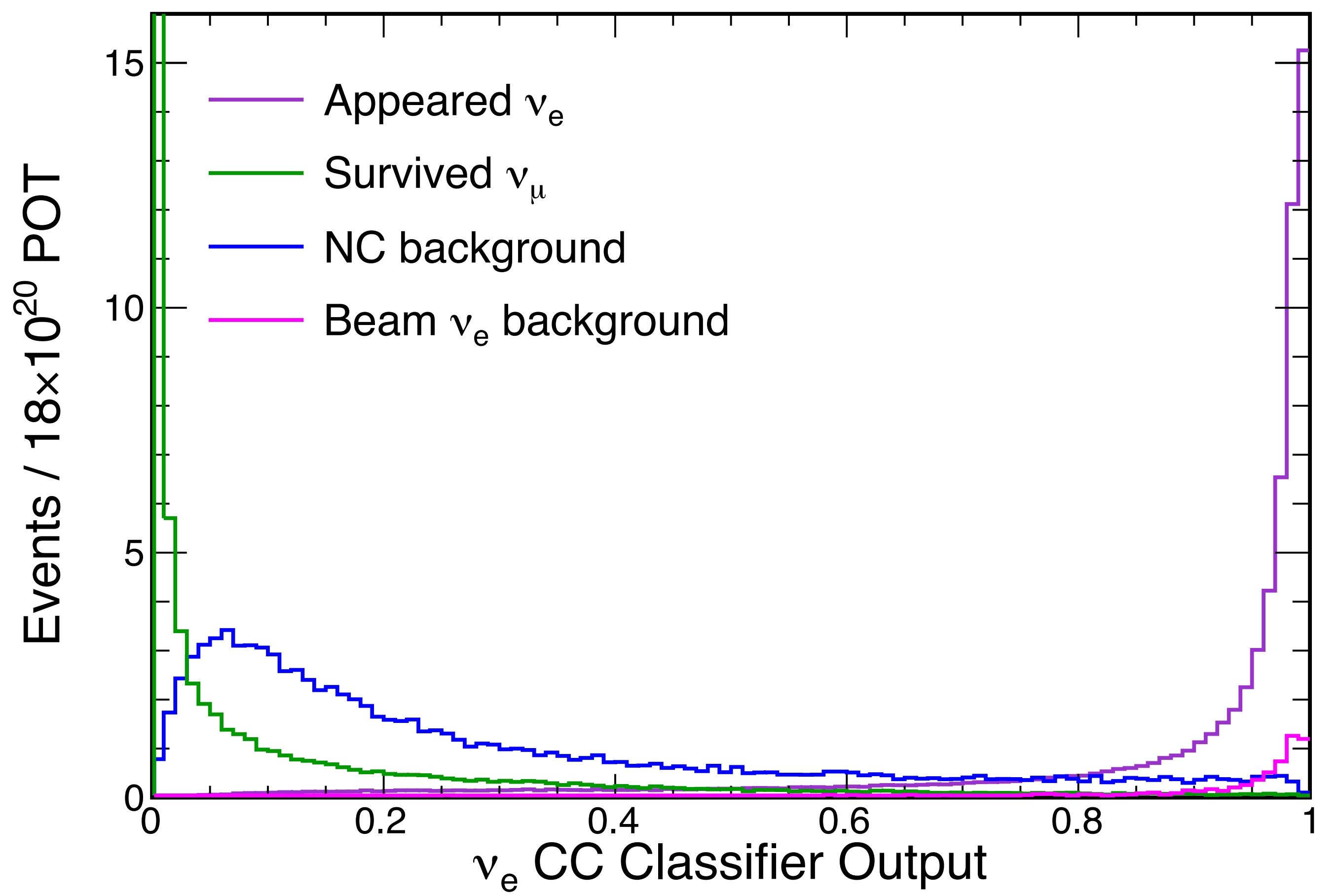


CVN Performance

- ν_μ signal separation is excellent, but ~identical to the existing KNN based selector. Expected since muons are easily identified.

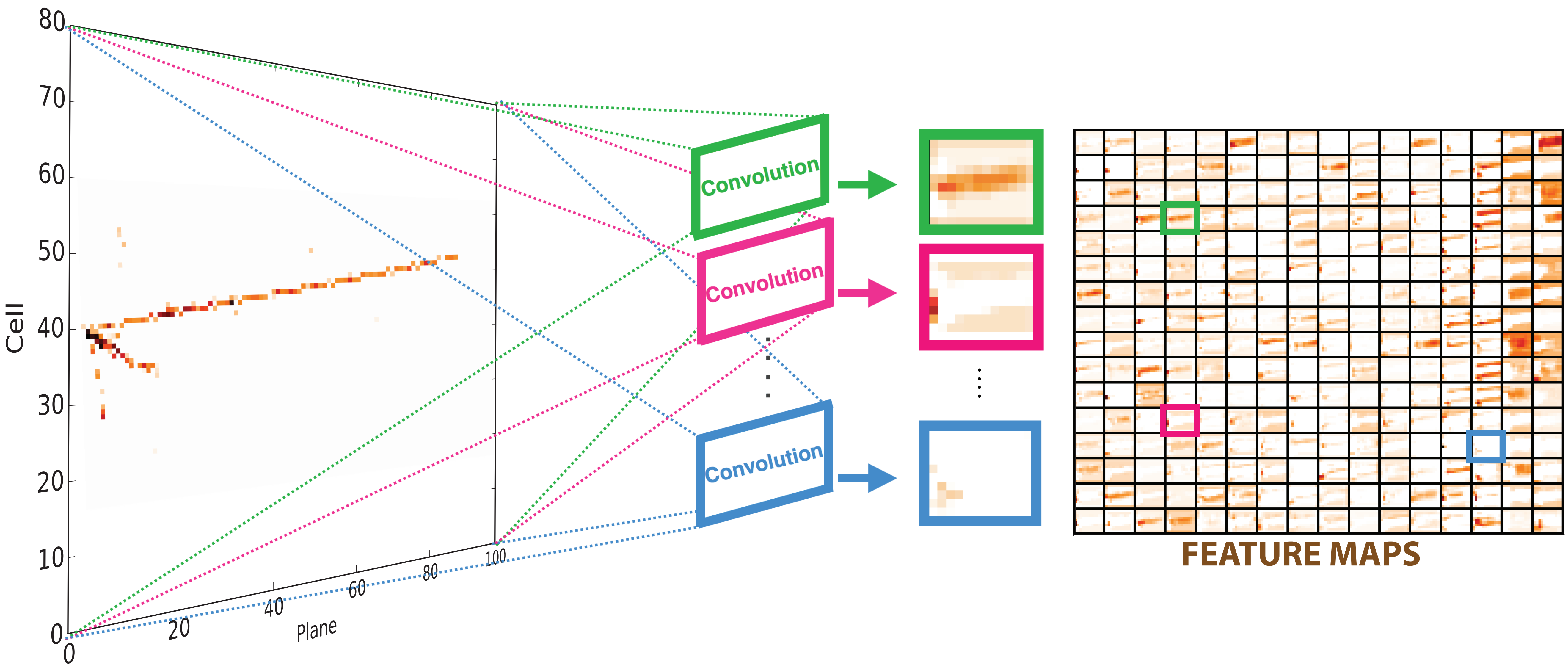
- ν_e selection 73% efficient, 76% pure with the CVN classifier. Performance gain is equivalent to 30% more exposure for the traditional selection techniques.





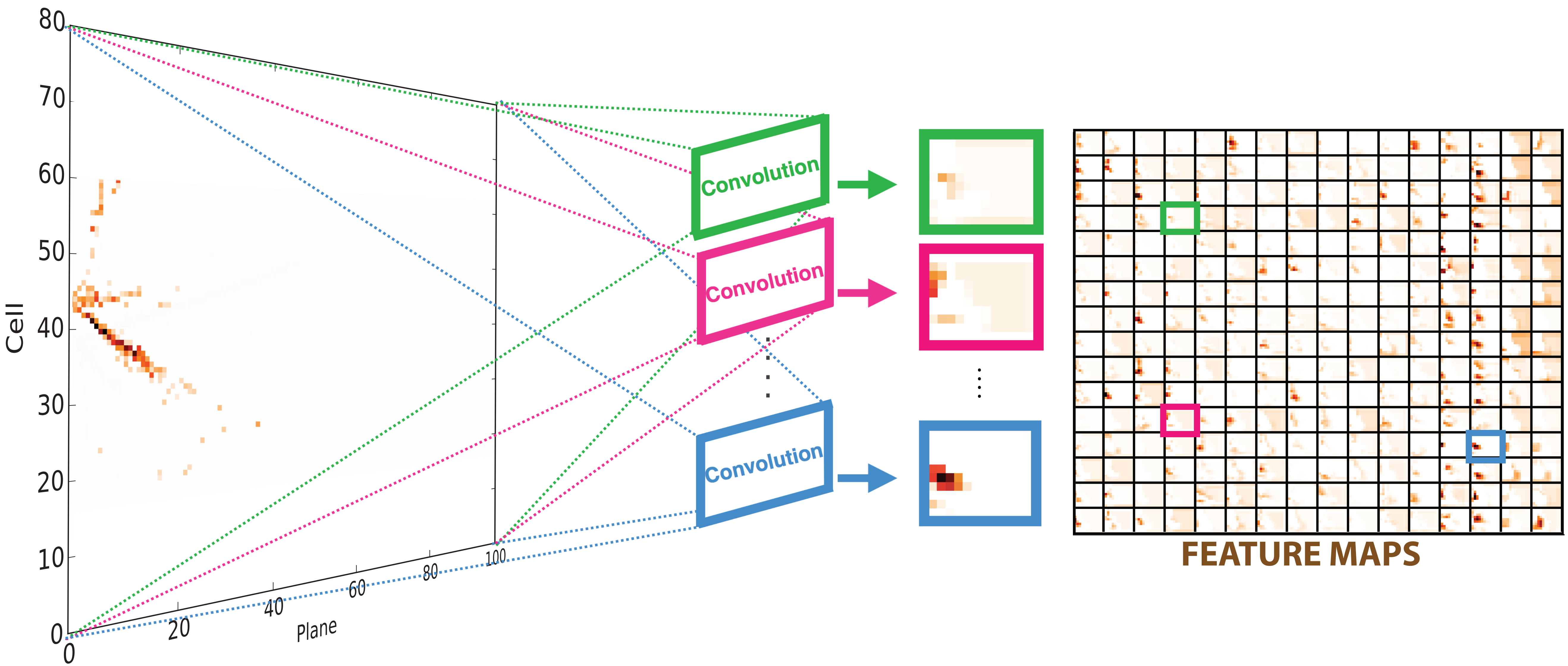
Convolutional Neural Networks

- Showing a muon neutrino interaction and the feature maps extracted from the convolutional kernels after the first inception module.

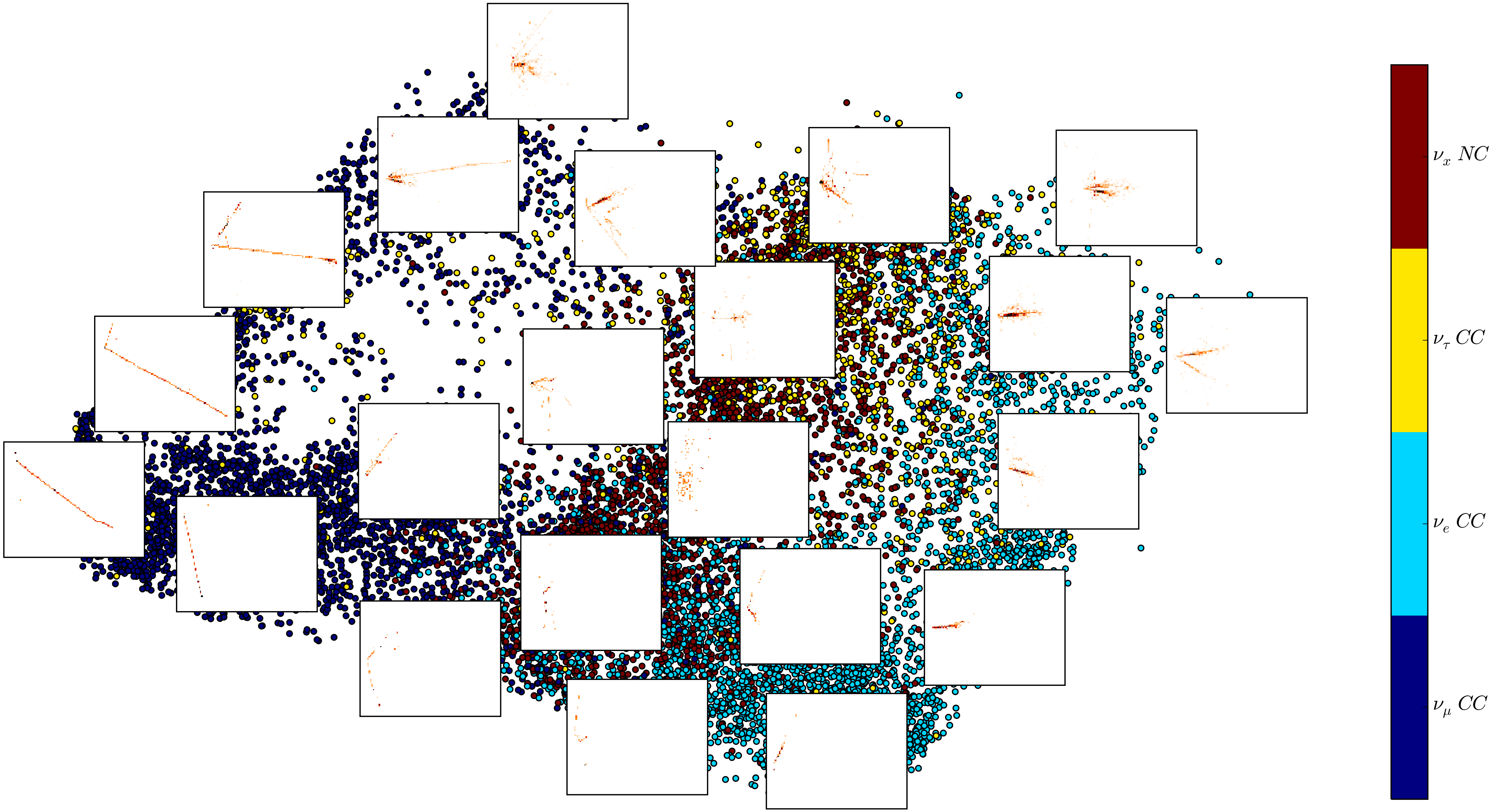


Convolutional Neural Networks

- Showing a electron neutrino interaction and feature maps extracted from the convolutional kernels after the first inception module
- The strong features extracted are the shower as opposed to the muon track

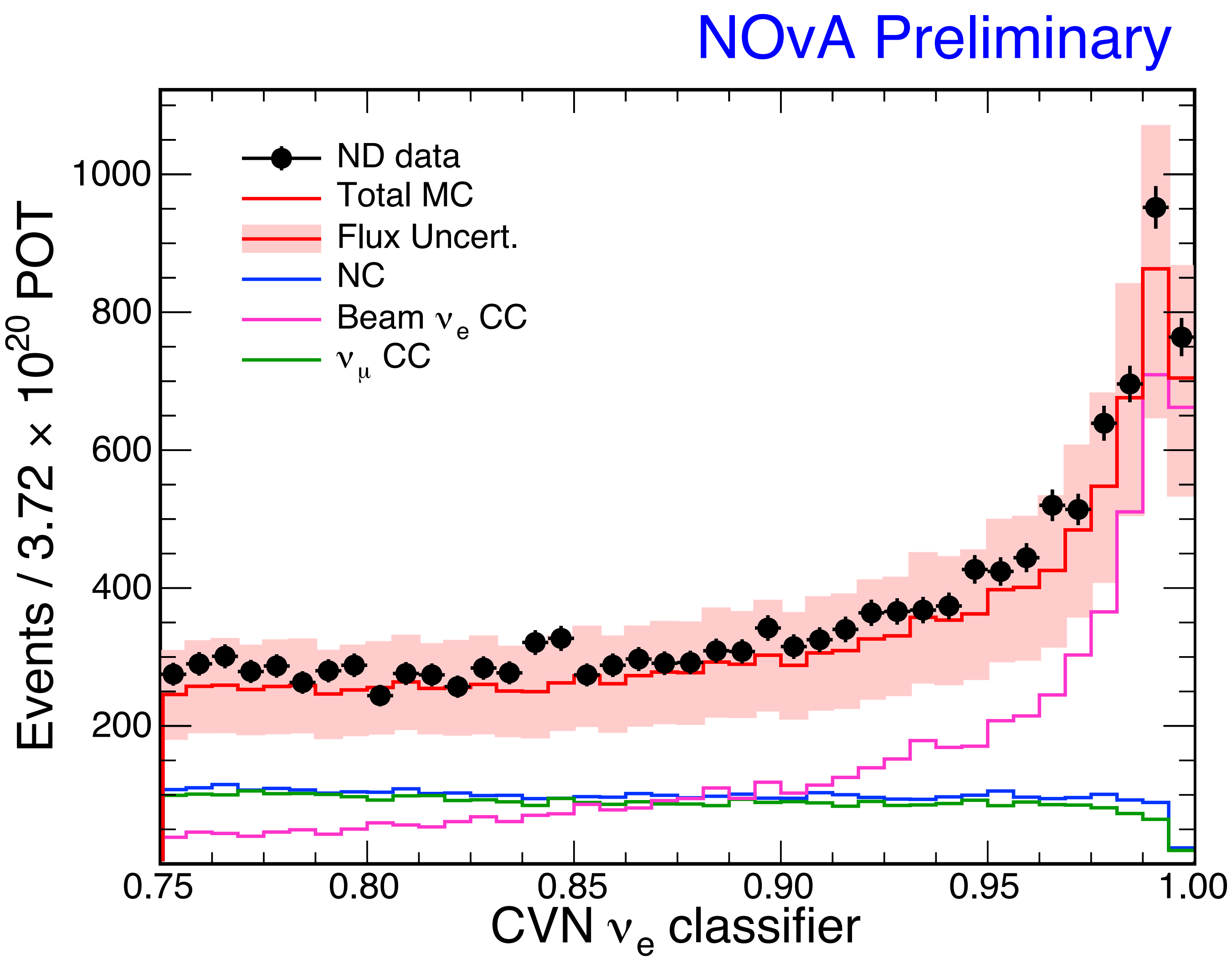


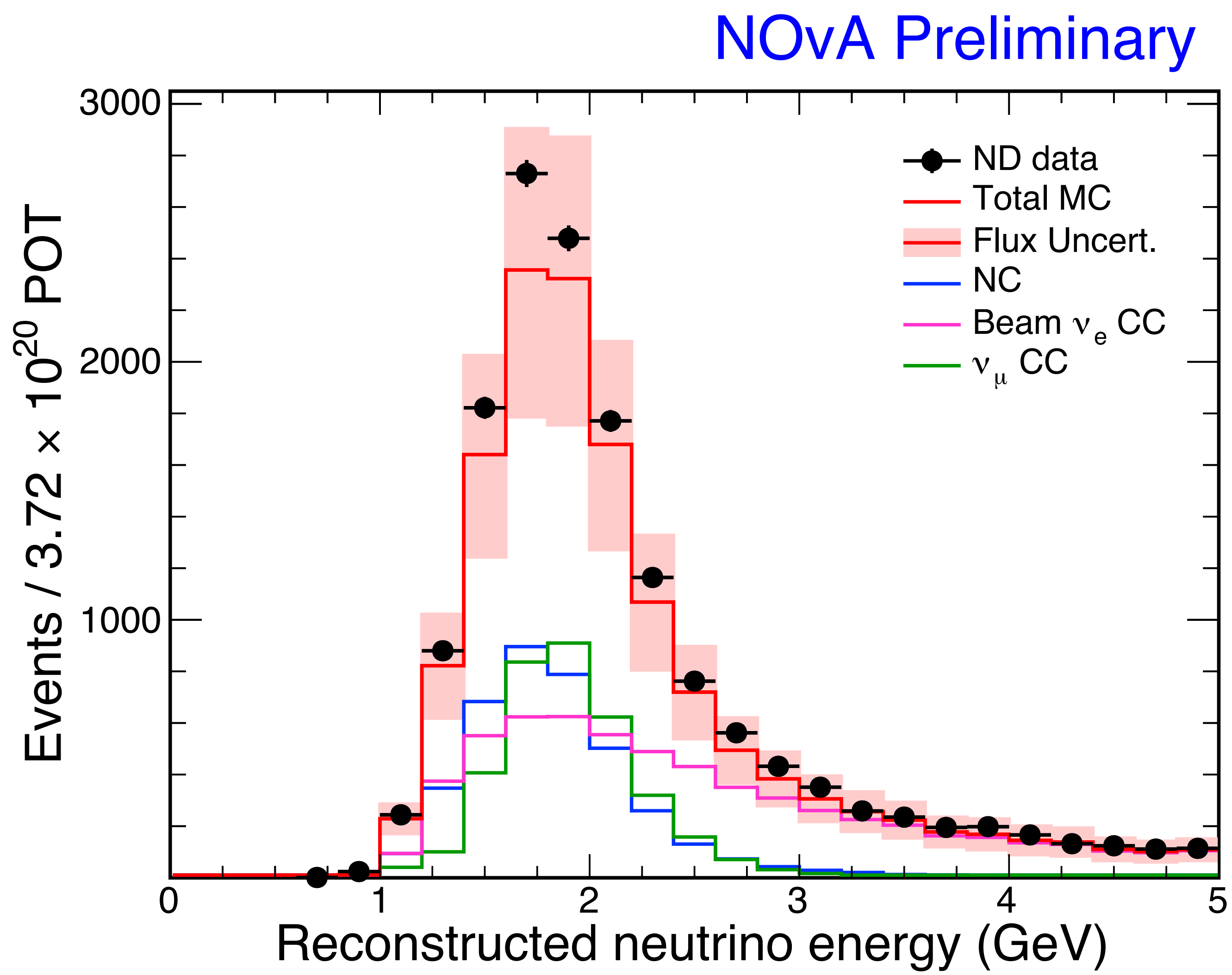
t-SNE representation of CVN classification. Truth labels shown for the training sample.



t-SNE representation of CVN classification. Truth labels shown for the training sample.

Performance Cross-checks on Data



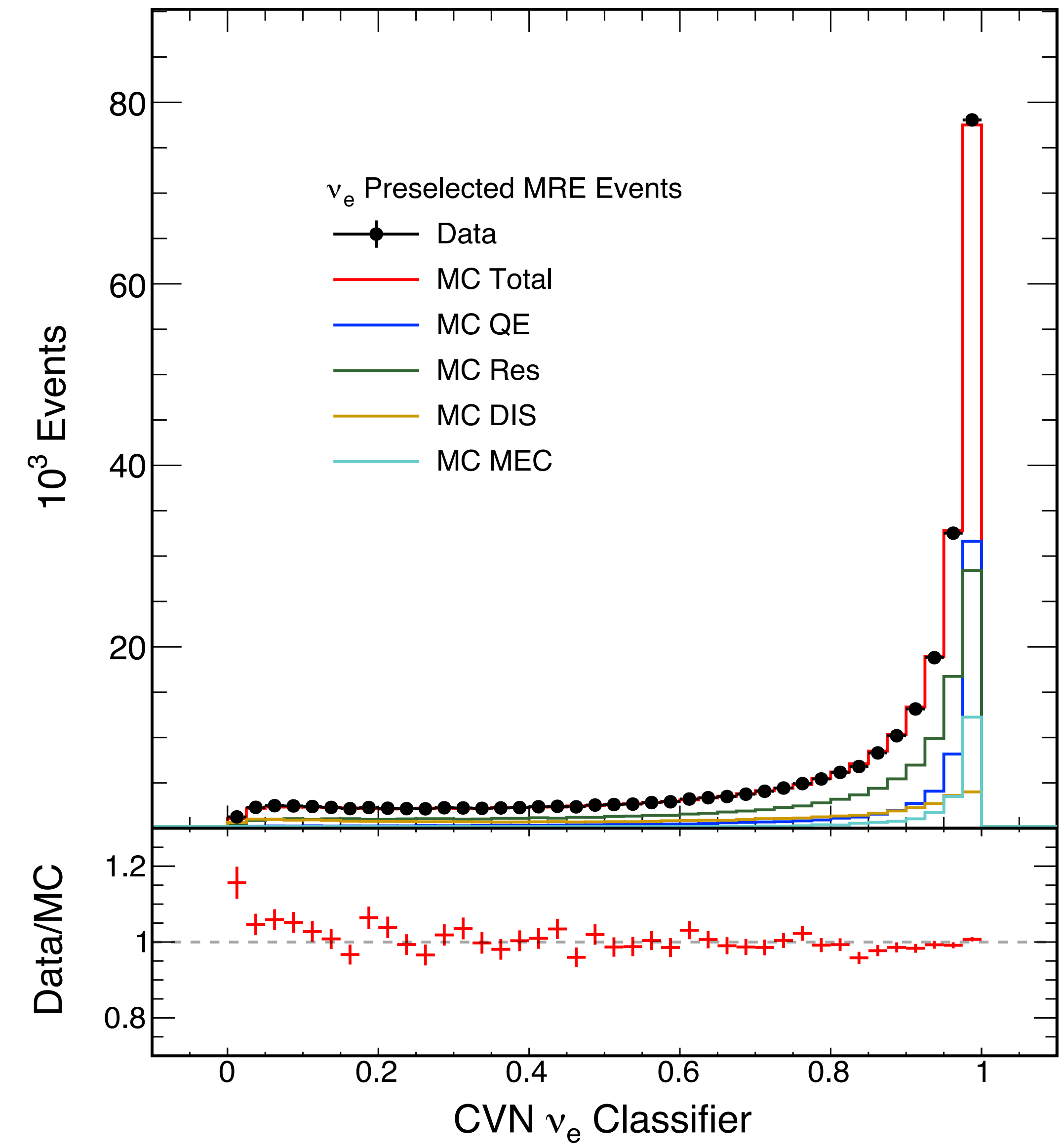


- Excellent data/simulation agreement in the Near Detector with high statistics

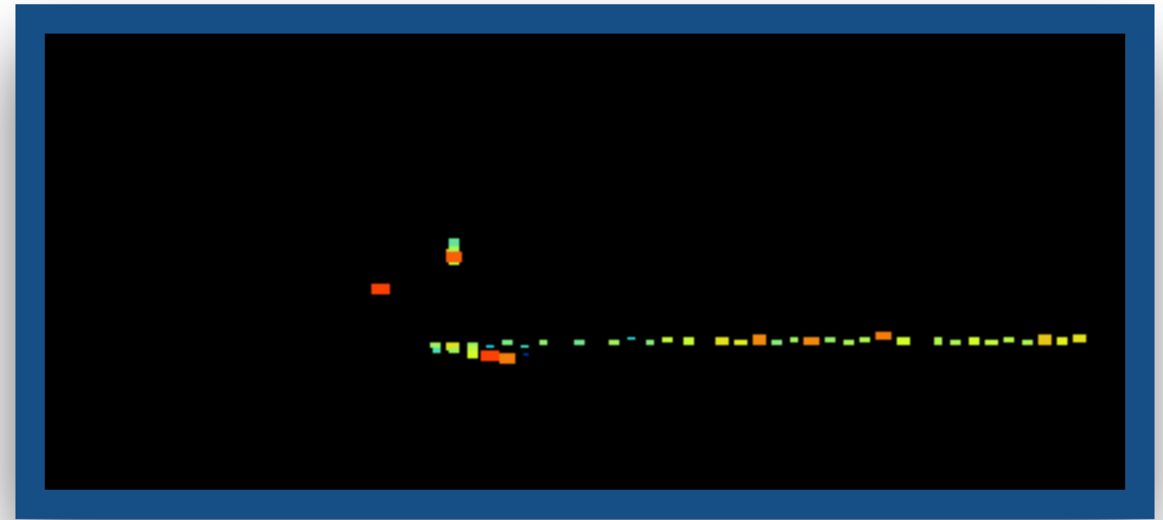
Performance Cross-checks on Data

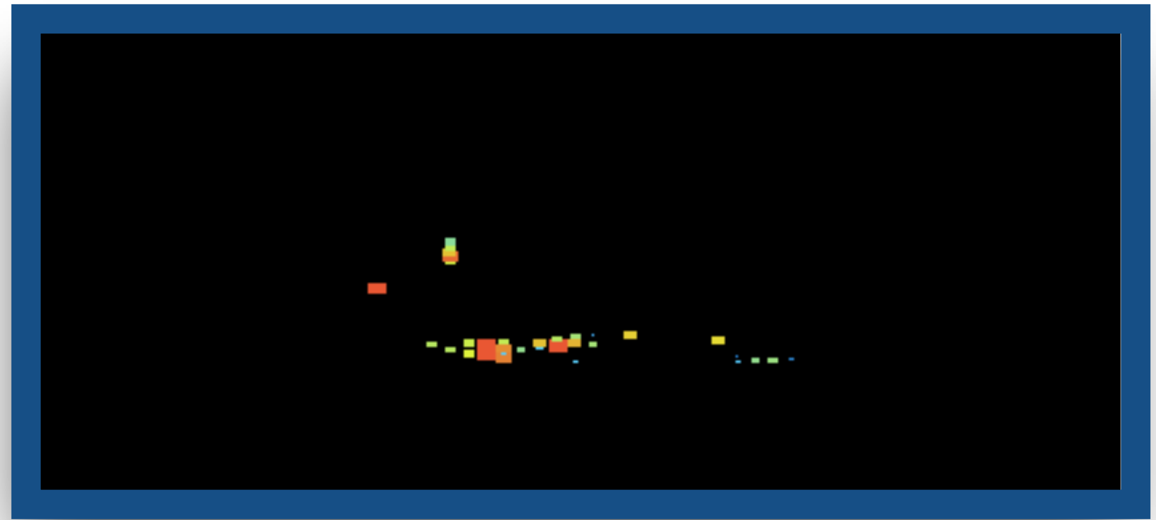
MRE (Muon Removed - Electron):

- Select muons in Near Detector interactions with a traditional classifier.
- Remove the muon hits and replace them with a single simulated electron of matching momentum.
- Data/MC comparisons show less than 1% difference in efficiency.



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

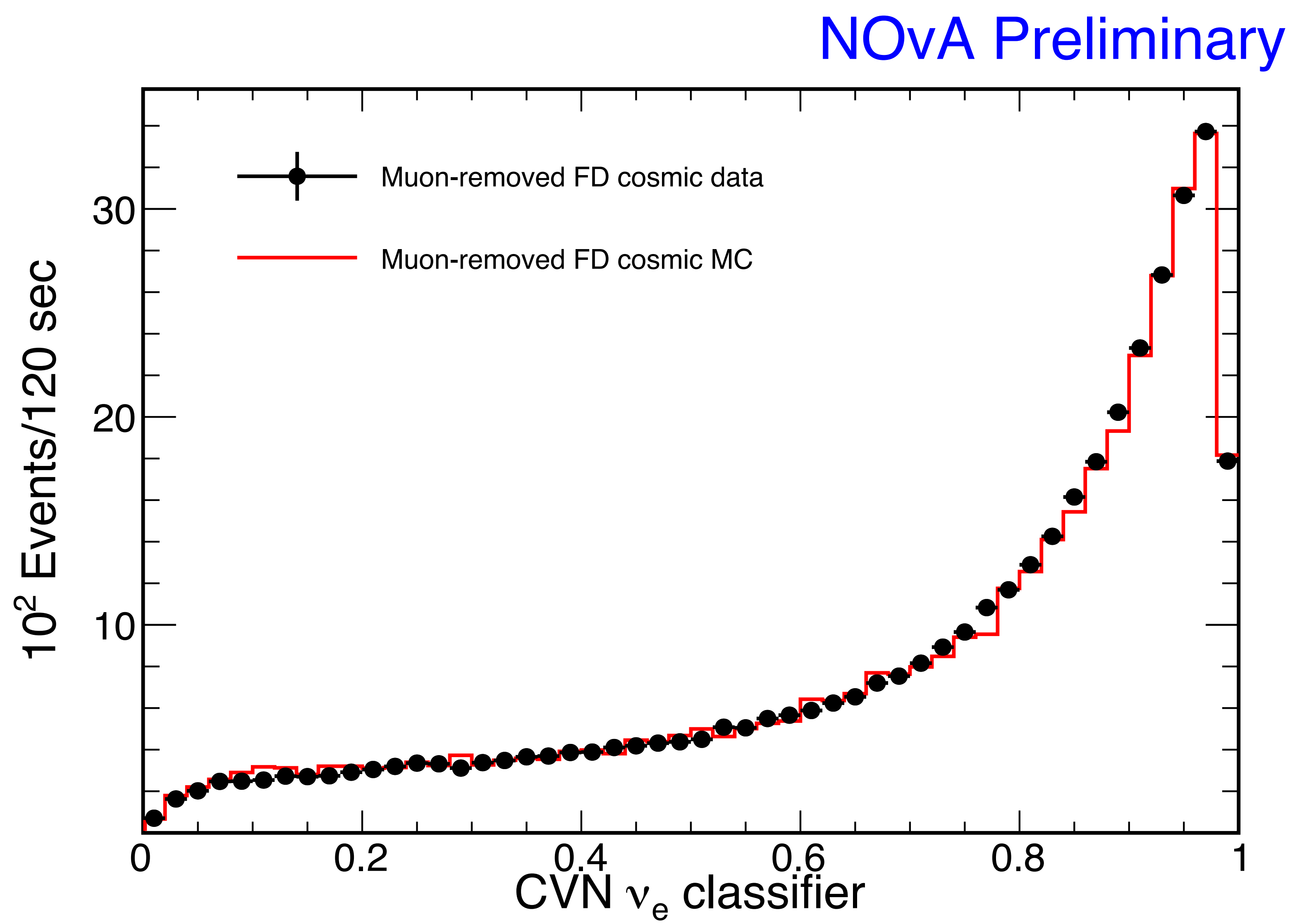




Performance Cross-checks on Data

MRBrem (Muon Removed Bremsstrahlung showers):

- Neutrino events are rare in the Far Detector, multitudes of cosmic ray muon events.
- Select cosmic muon events with an electromagnetic shower from bremsstrahlung radiation.
- Remove the muon hits and apply CVN classification to the remaining electromagnetic shower.
- Data/MC comparisons show very good agreement

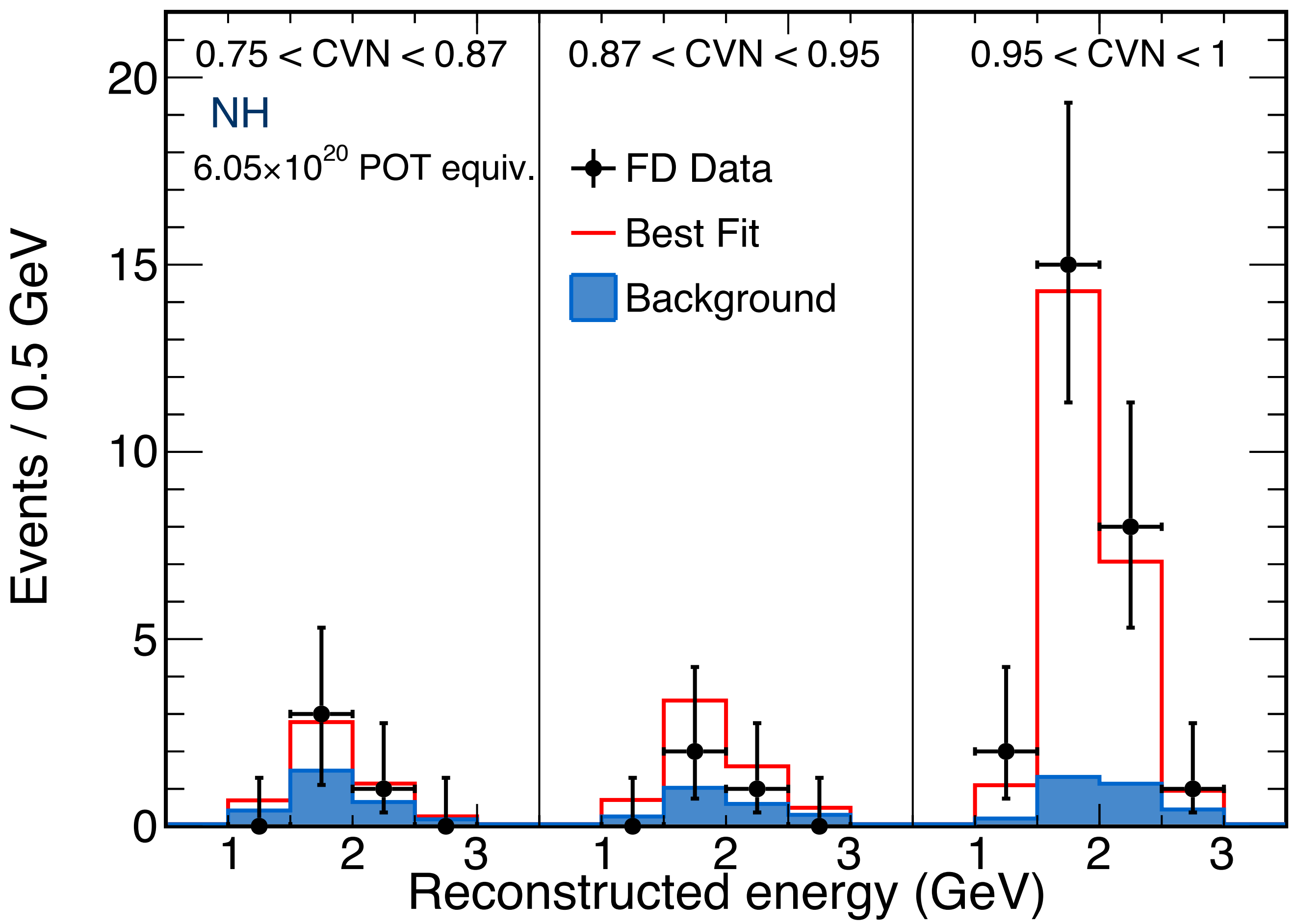


One Network, Many results

Network produces multi-dimensional classification output, normalized to 1.

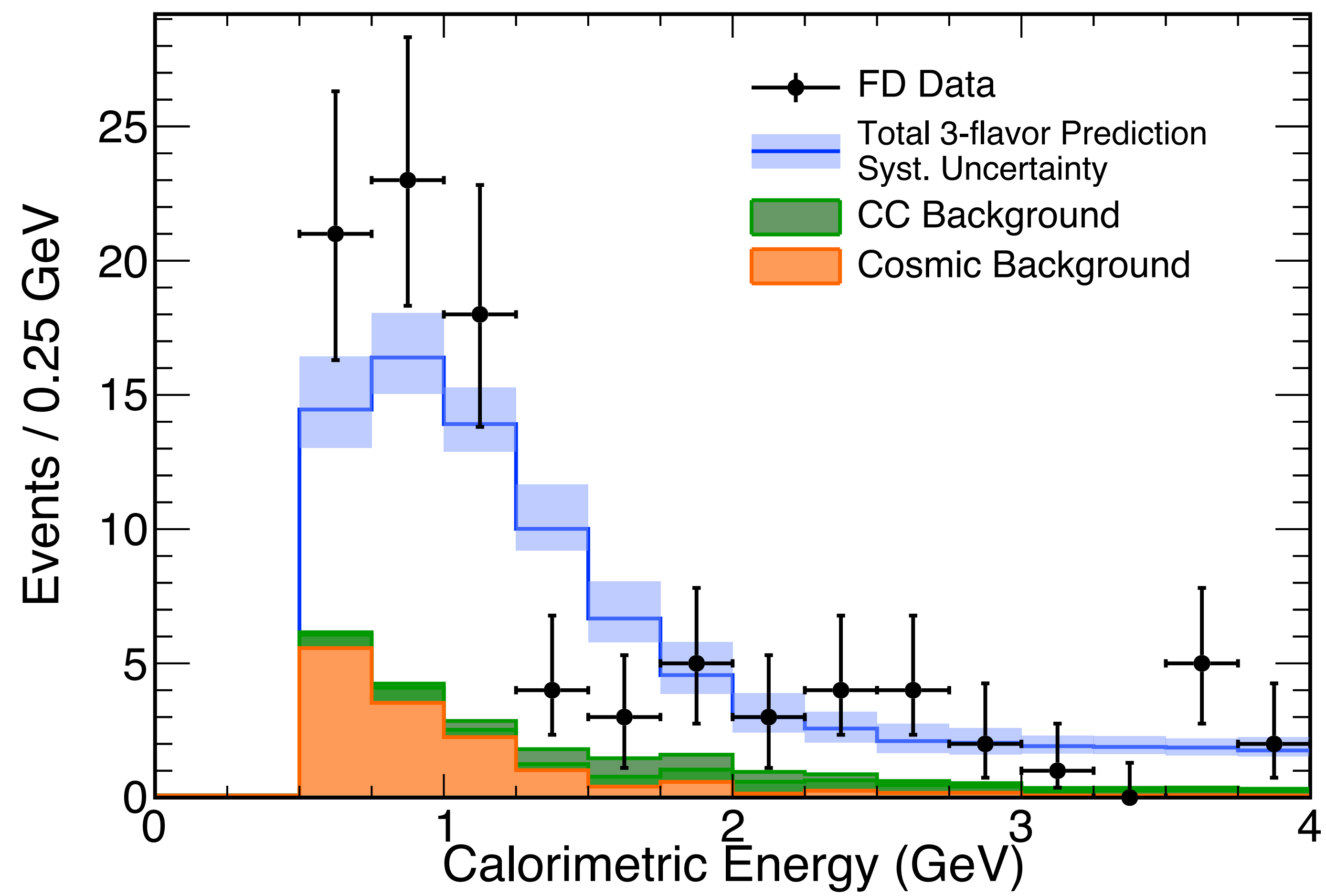
Reduces processing time running one network for many analyzes.

Electron Neutrino Selection



PRL 118, 231801

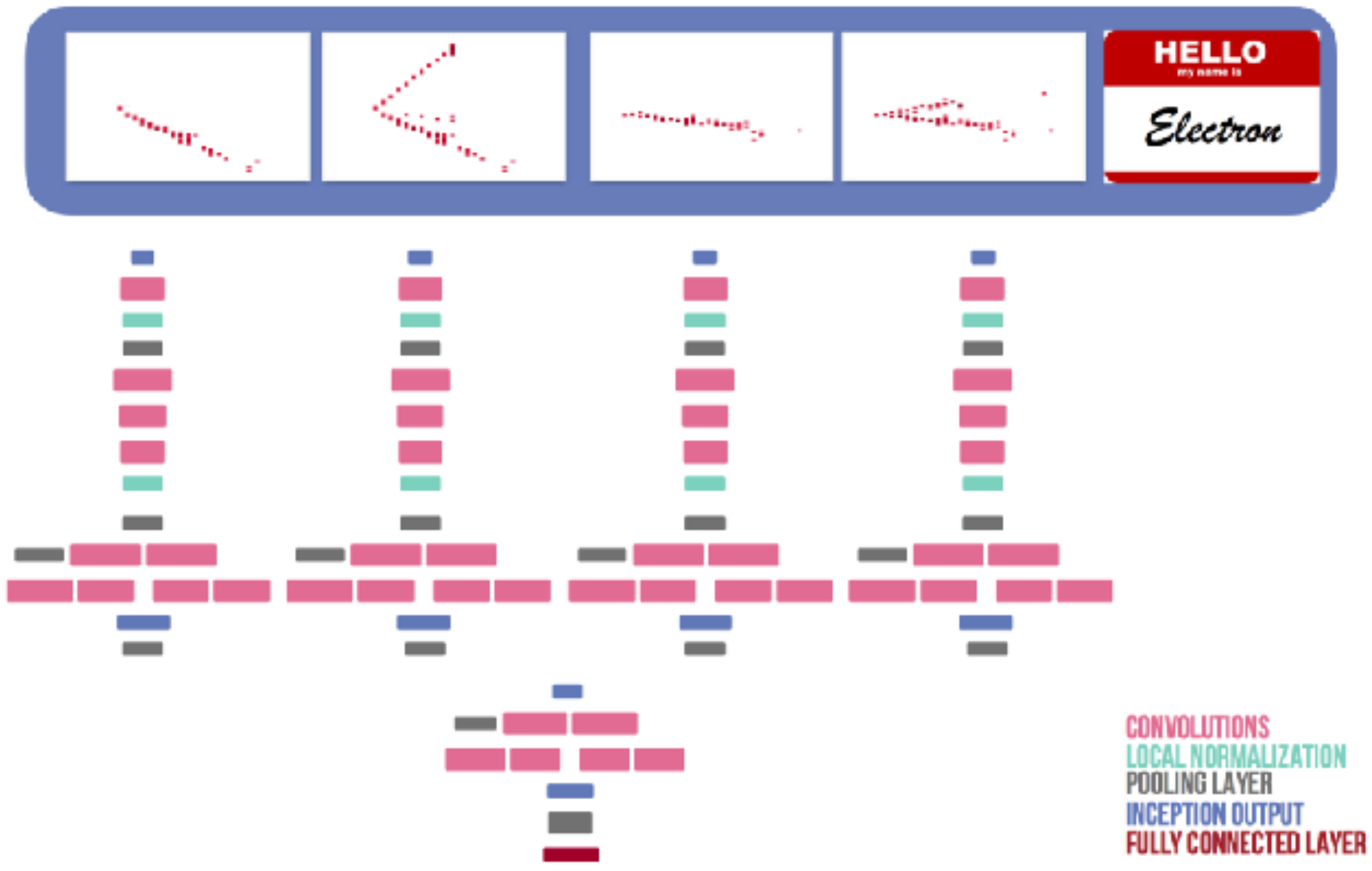
Neutral Current Selection



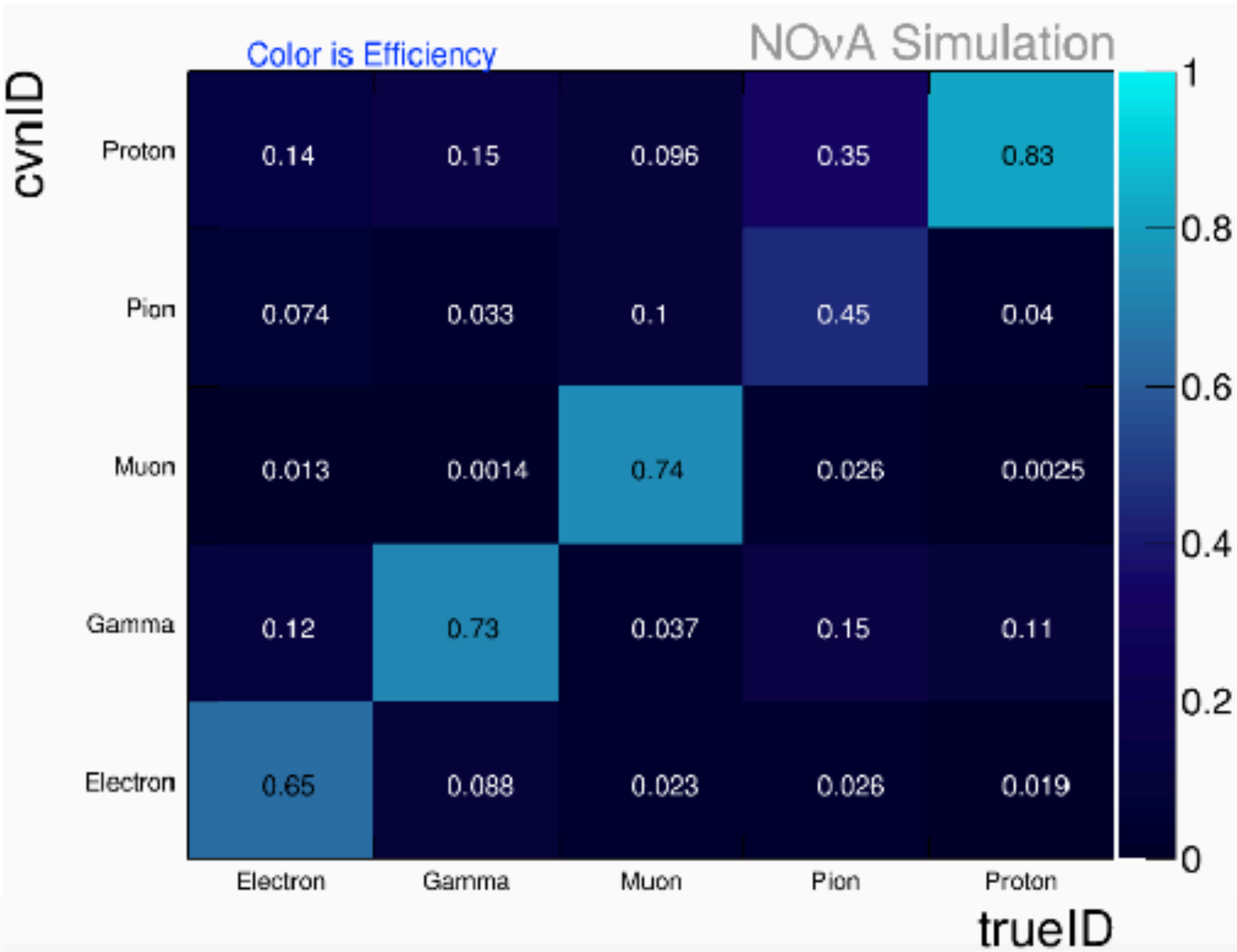
arXiv:1706.04592

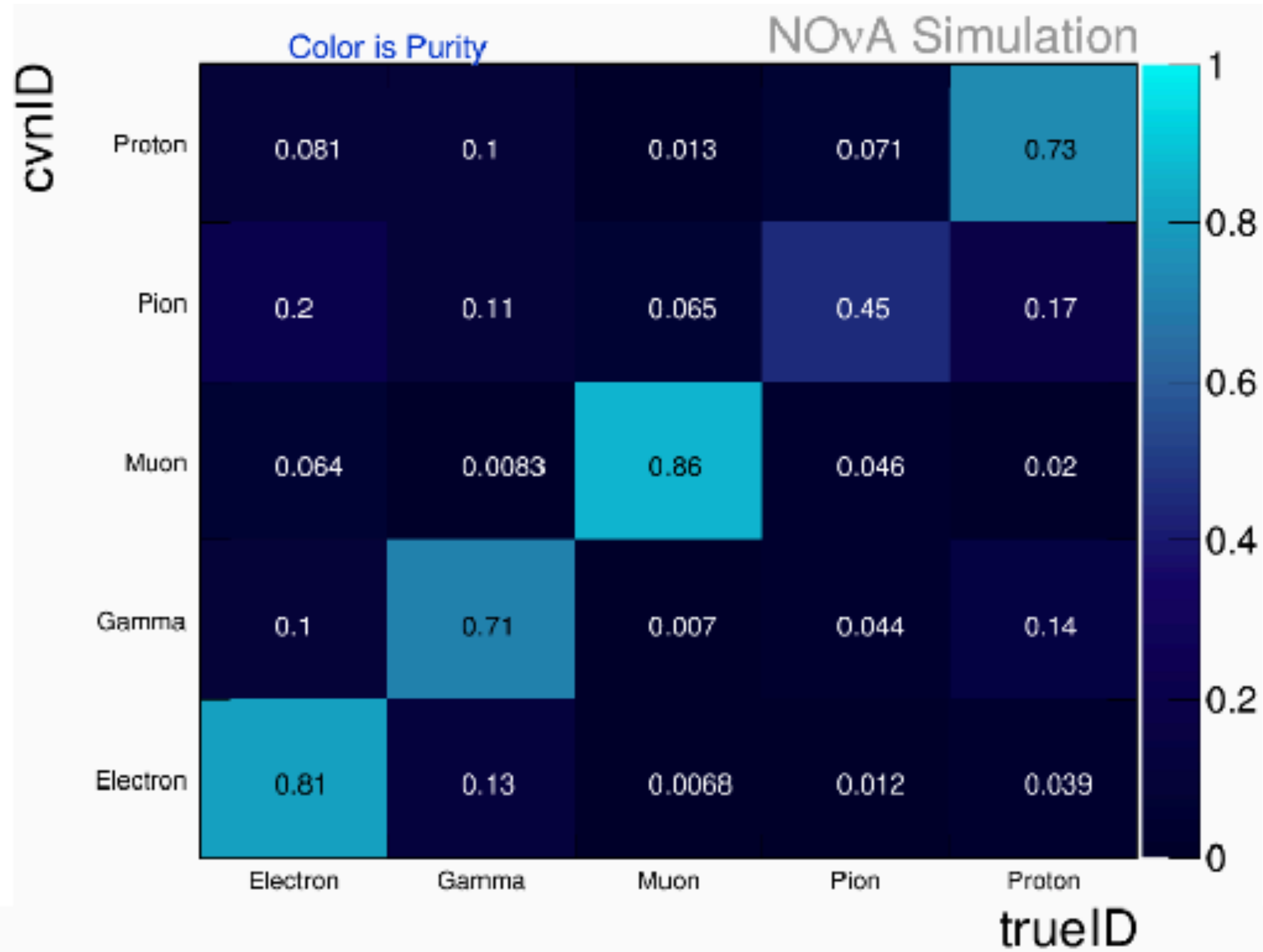
Other applications: Particle Identification

- Instead of classifying the entire event, identify individual particles
- Input pixel maps of the particle and the neutrino interaction
- Couples to reconstruction quality of input tracks, train above a minimum purity



Other applications: Particle Identification

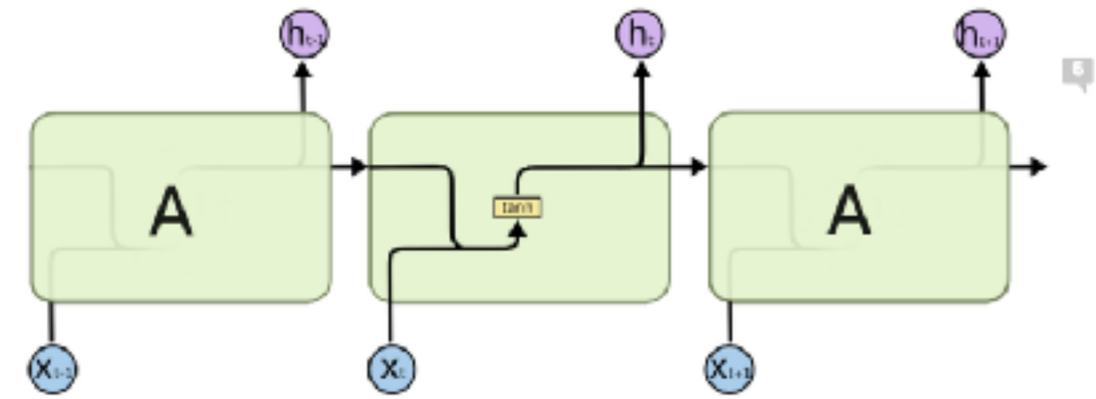




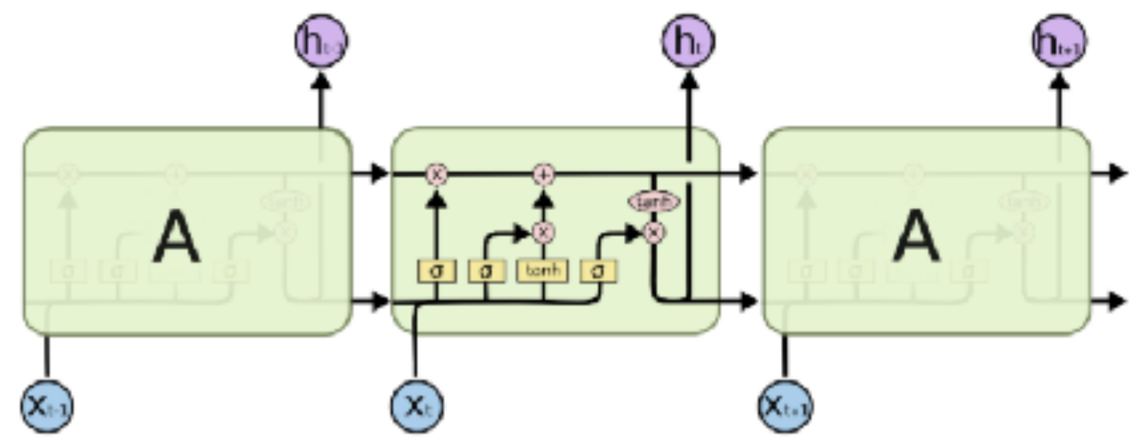
Go one step further in the future, classifying individual image pixels by particle via semantic segmentation, then feed back into the reconstruction.

Other applications: Energy Estimation

- Recurrent Neural Networks are sequential, using the current state of the system and output of last iteration



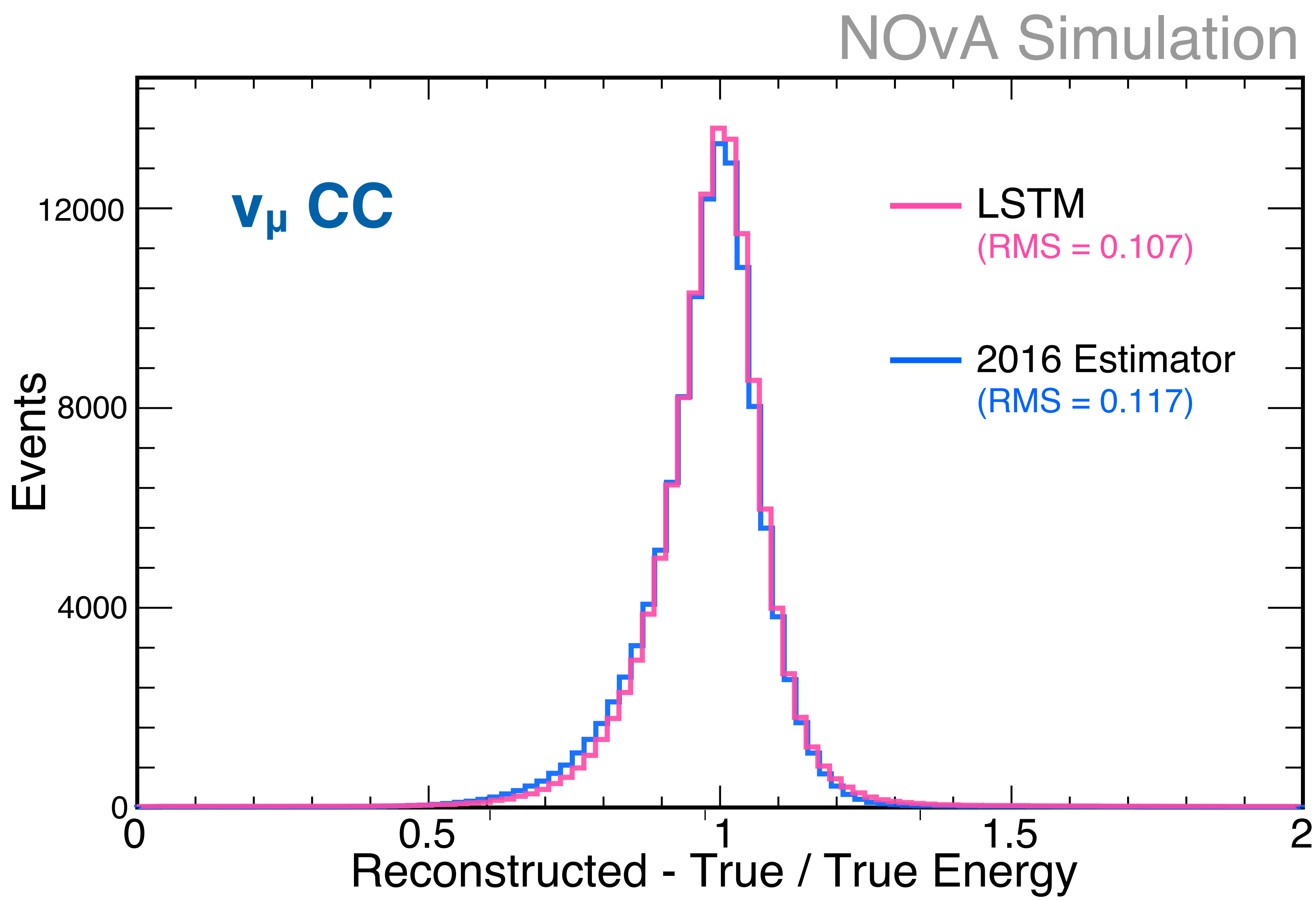
- Add Long-Short Term Memory cell as secondary path to keep longer memory



- Feed reconstructed information from tracks, particle IDs, and energy estimators into network

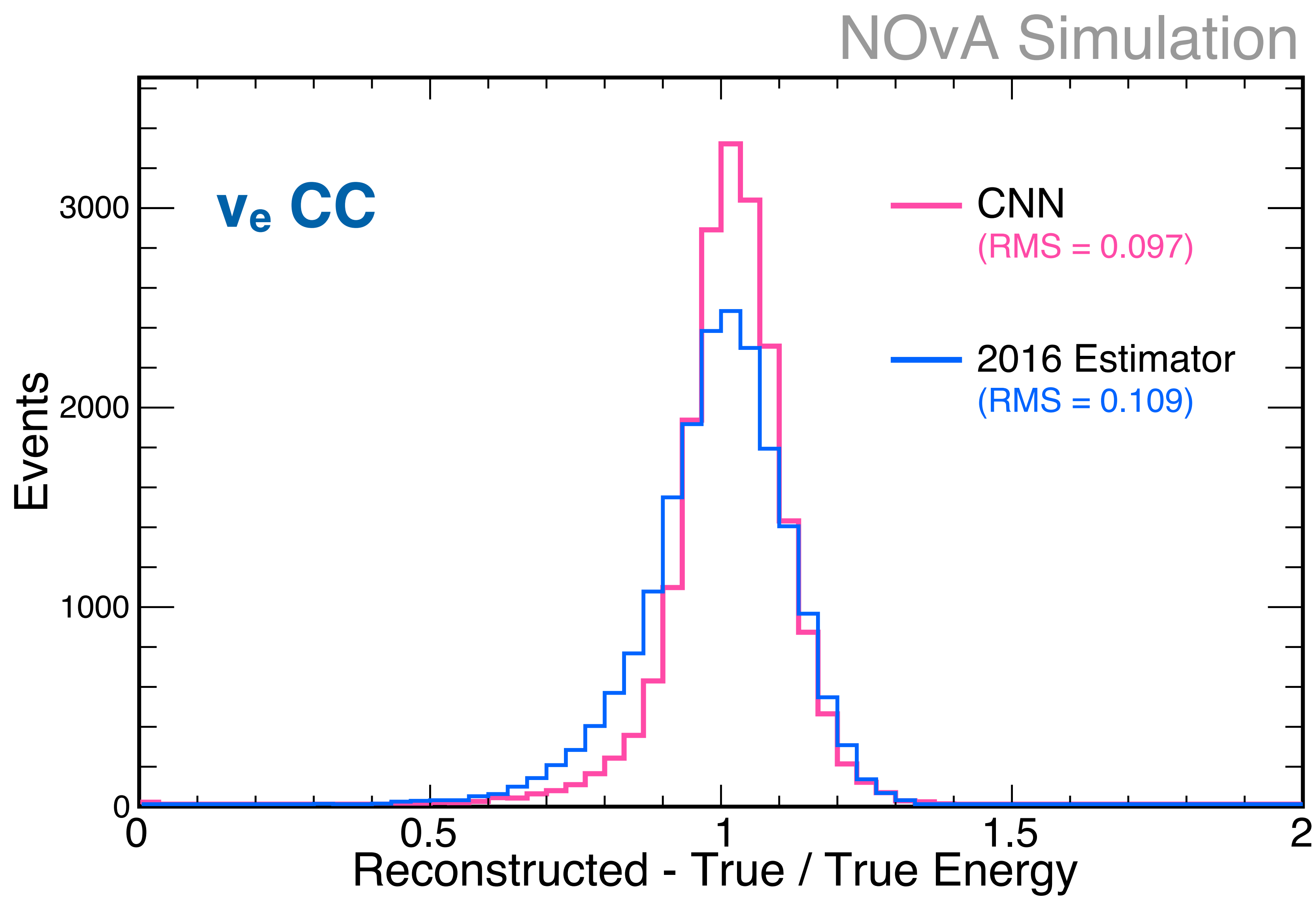
Other applications: Energy Estimation

- Recurrent Neural Networks are sequential, using the current state of the system and output of last iteration
- Add Long-Short Term Memory cell as secondary path to keep longer memory
- Feed reconstructed information from tracks, particle IDs, and energy estimators into network



Other applications: Energy Estimation

- Convolutional Neural network trained for energy estimation instead of a classification task
- Feed vertex reconstruction information into final fully connect layer
- Working toward multi-target networks that do both regression and classification tasks



More to Explore

- Inter-experimental LHC Machine Learning group
  - https://iml.web.cern.ch

- Inter-experimental Machine Learning working group for the intensity and cosmic frontiers
  - http://machinelearning.fnal.gov

- MicroBoone: “Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber” JINST 12 (2017) no.03, P03011

- Vertex Reconstruction in MINEvA
  - ieeexplore.ieee.org/iel7/7958416/7965814/07966131.pdf

- Adaptation of CVN in DUNE as well as semantic segmentation and other networks

- Many applications in colliders, astrophysics and more

Summary

- Advancements in GPUs and the field of computer vision make training deep neural networks with minimal reconstruction feasible
- Many detectors in particle physics have an image like quality that makes them natural candidates for convolutional neural networks
- In NOvA a multi-classification CNN was developed that produced a performance gain equivalent to 30% more exposure, proven robust in Data/ MC studies
- Expanding NOvA DL applications to look at particle identification, energy estimation, vertexing, clustering, and more
- Just the tip of the iceberg! Many more applications of deep networks in HEP
- Thanks!