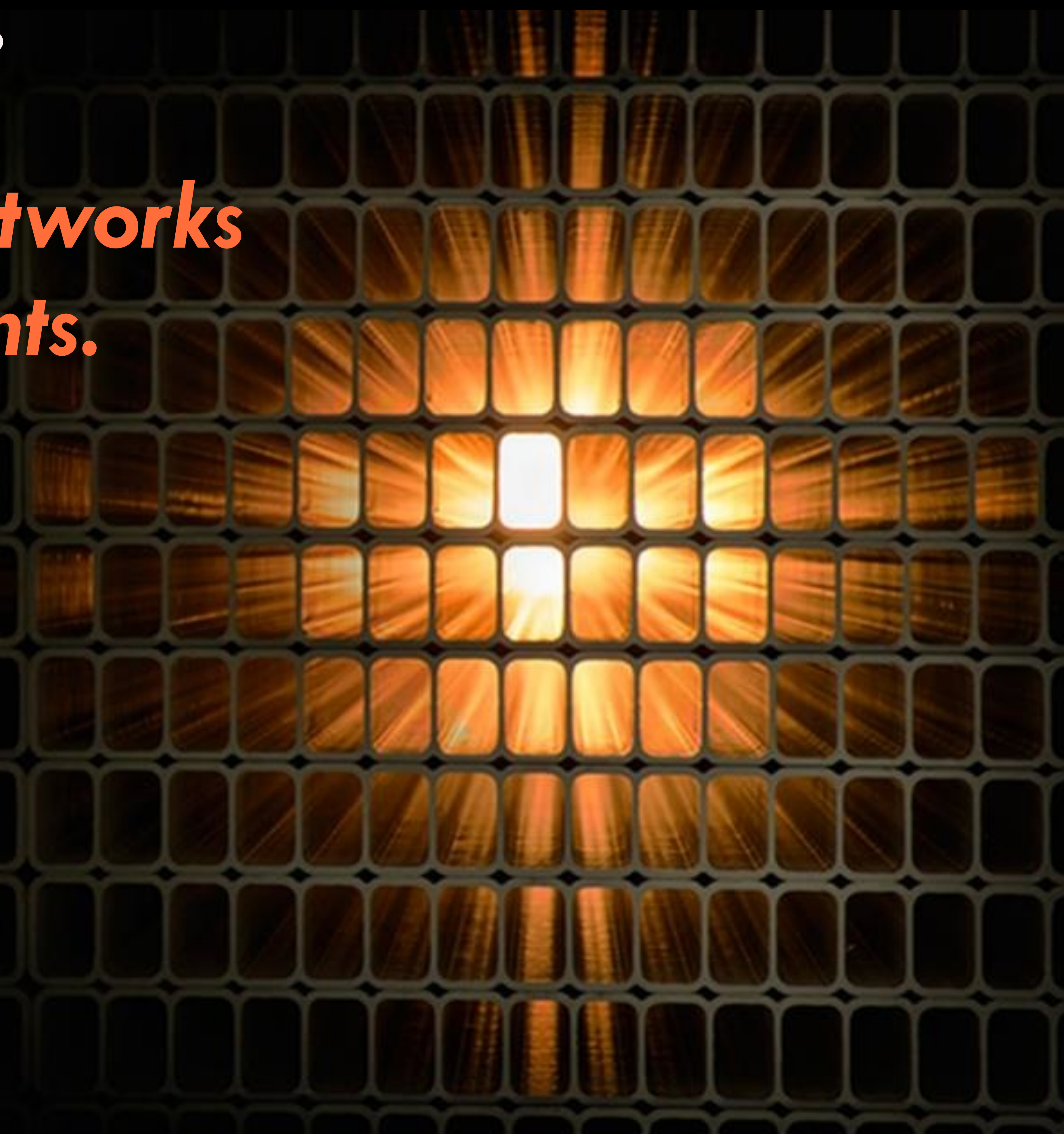
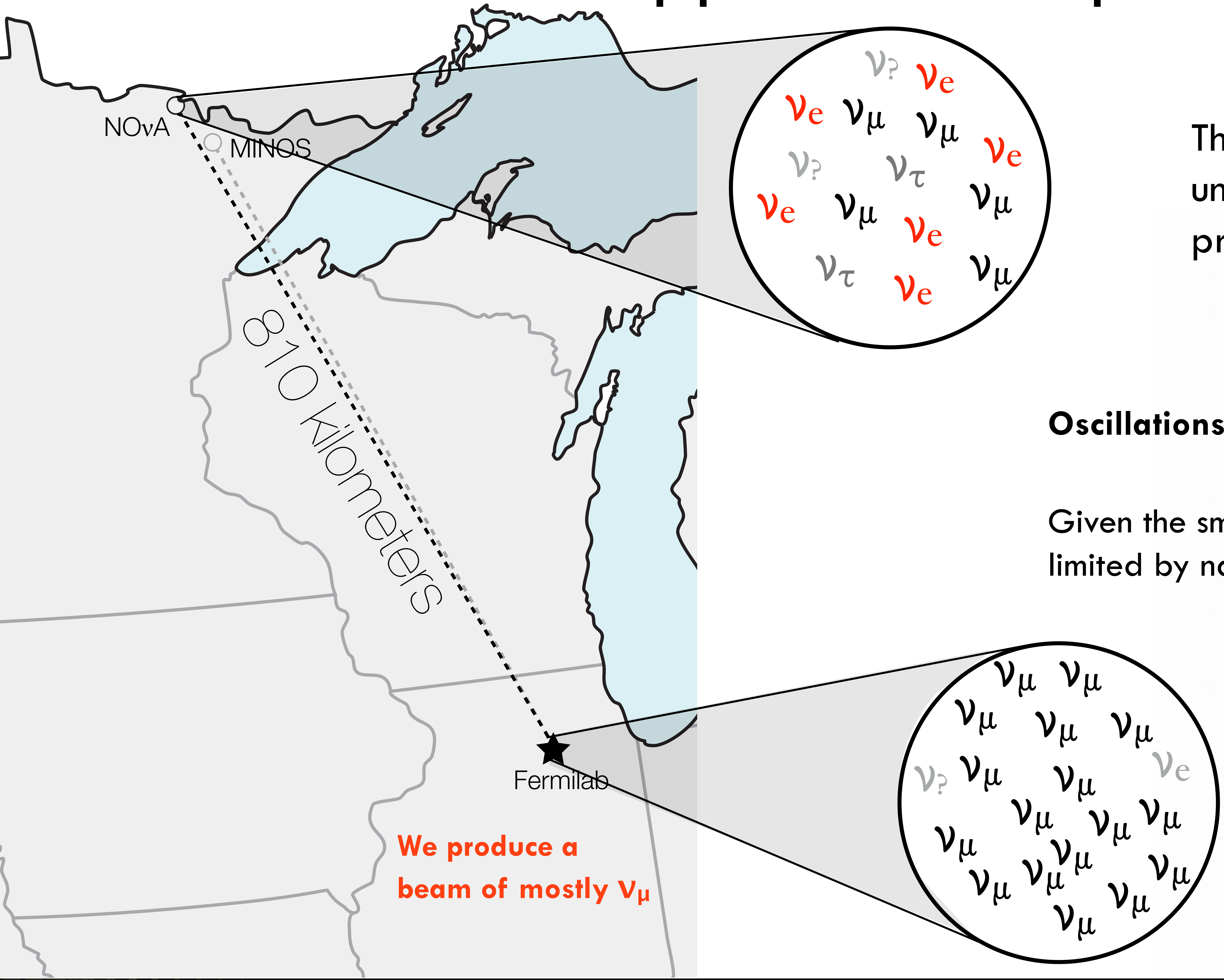
The Convolutional Visual Network for Identification of NOvA Events.

An implementation of Convolutional Neural networks and its applications on neutrino interaction events.



Fernanda Psihas
Indiana University
For the NOvA Collaboration

NuMI Off-axis ν_e Appearance Experiment



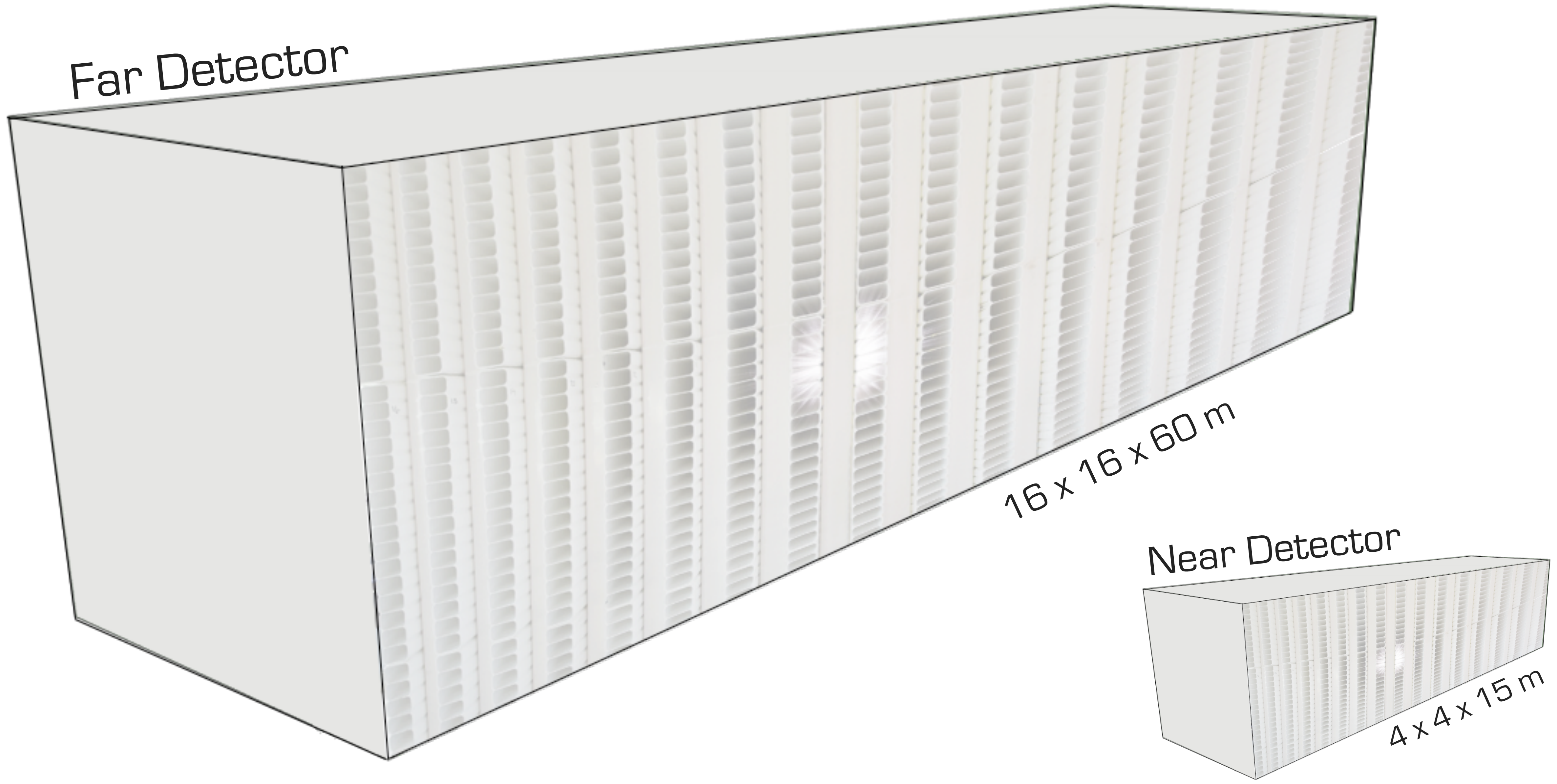
We produce a beam of mostly ν_μ

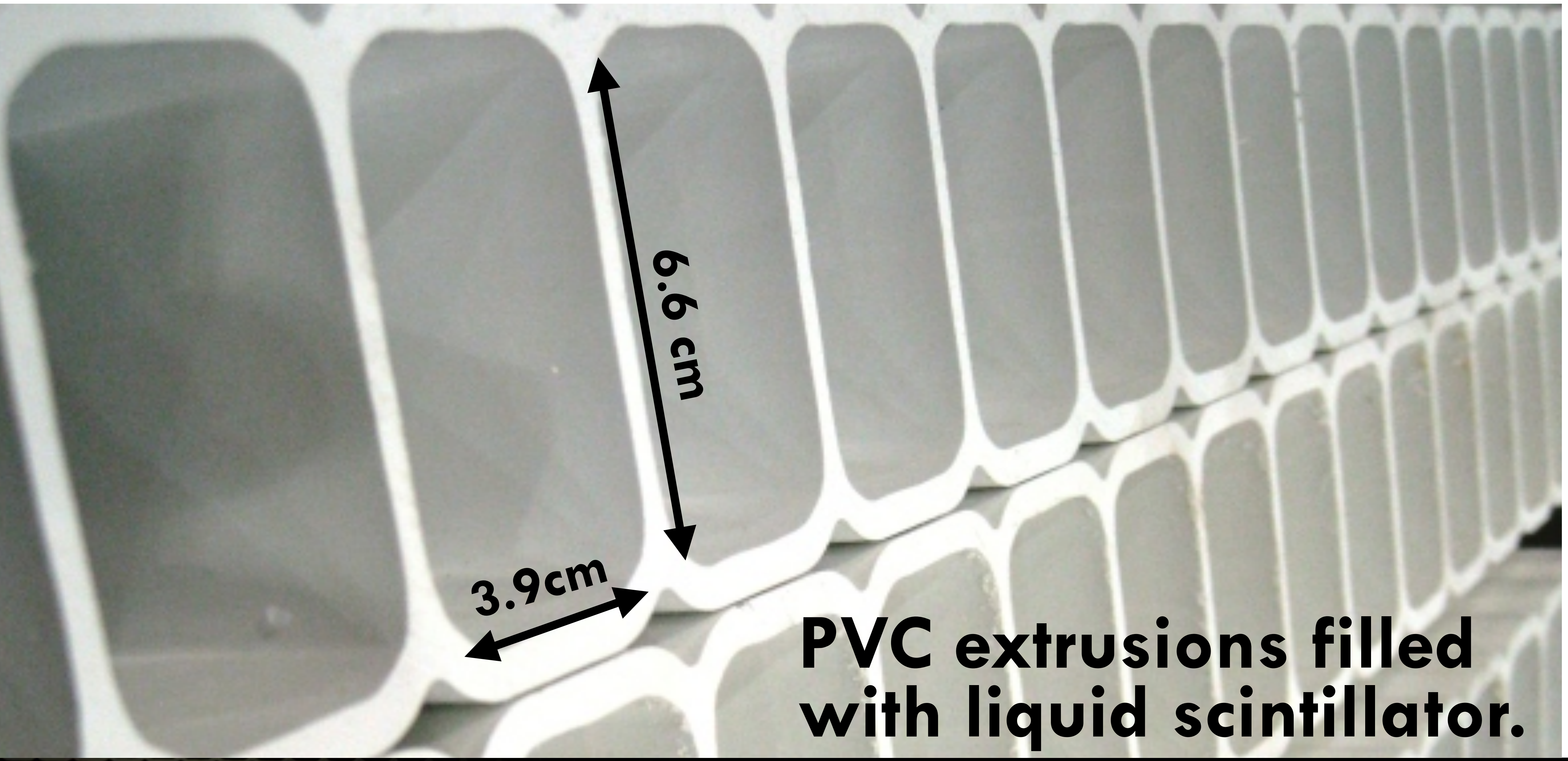
The neutrino flavor eigenstates undergo oscillations as they propagate.

Oscillations measurements rely on flavor identification.

Given the small cross sections of neutrinos, they are statistics limited by nature **i.e. 33 signal events in two years of data.**

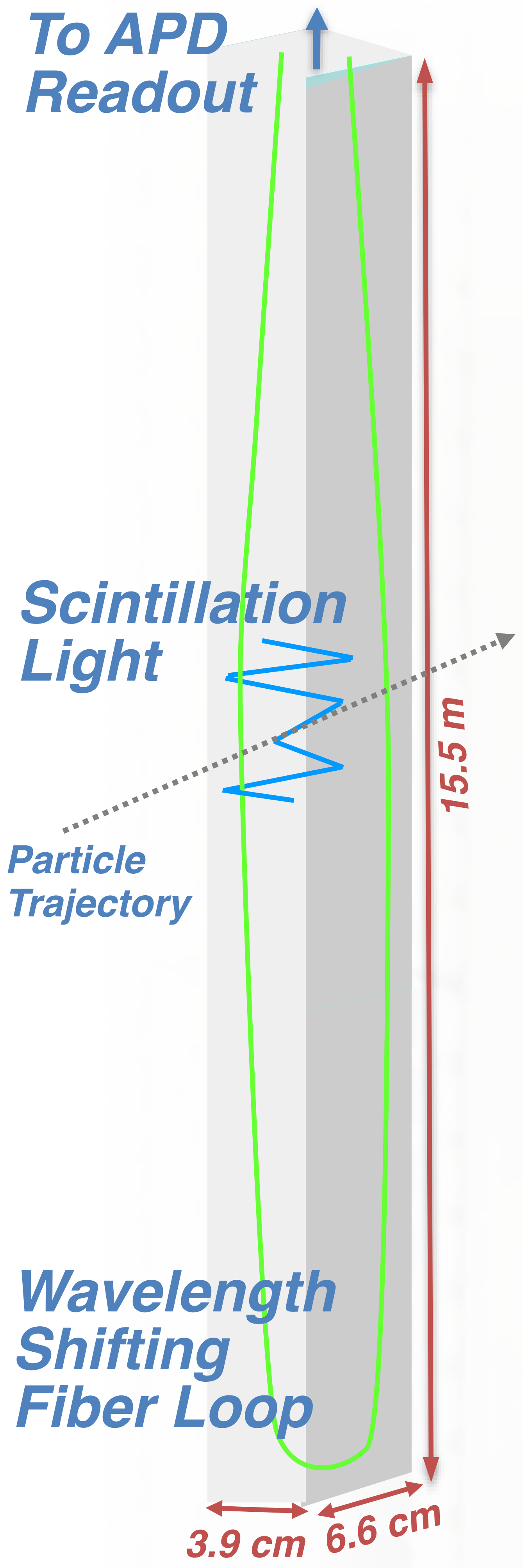
The NOvA Detectors



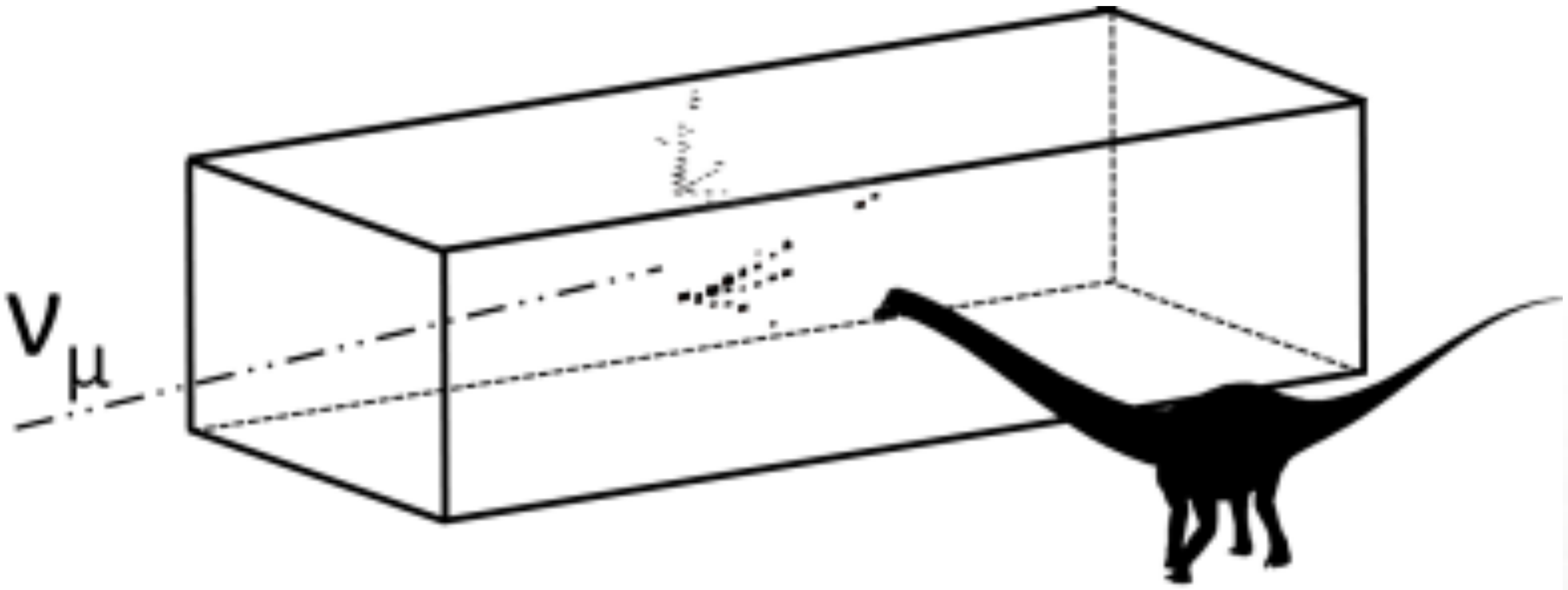


PVC extrusions filled with liquid scintillator.

Charged particles are detected though the scintillation light produced in each cell.

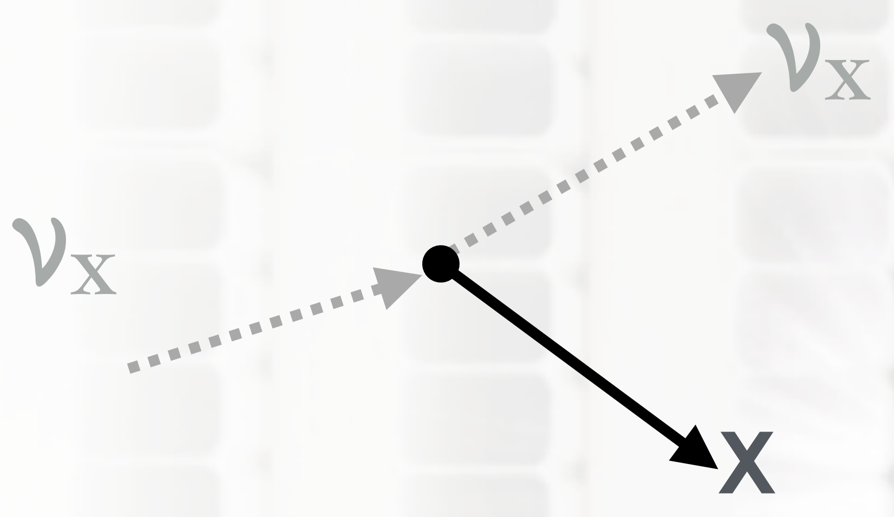


Neutrino Interactions



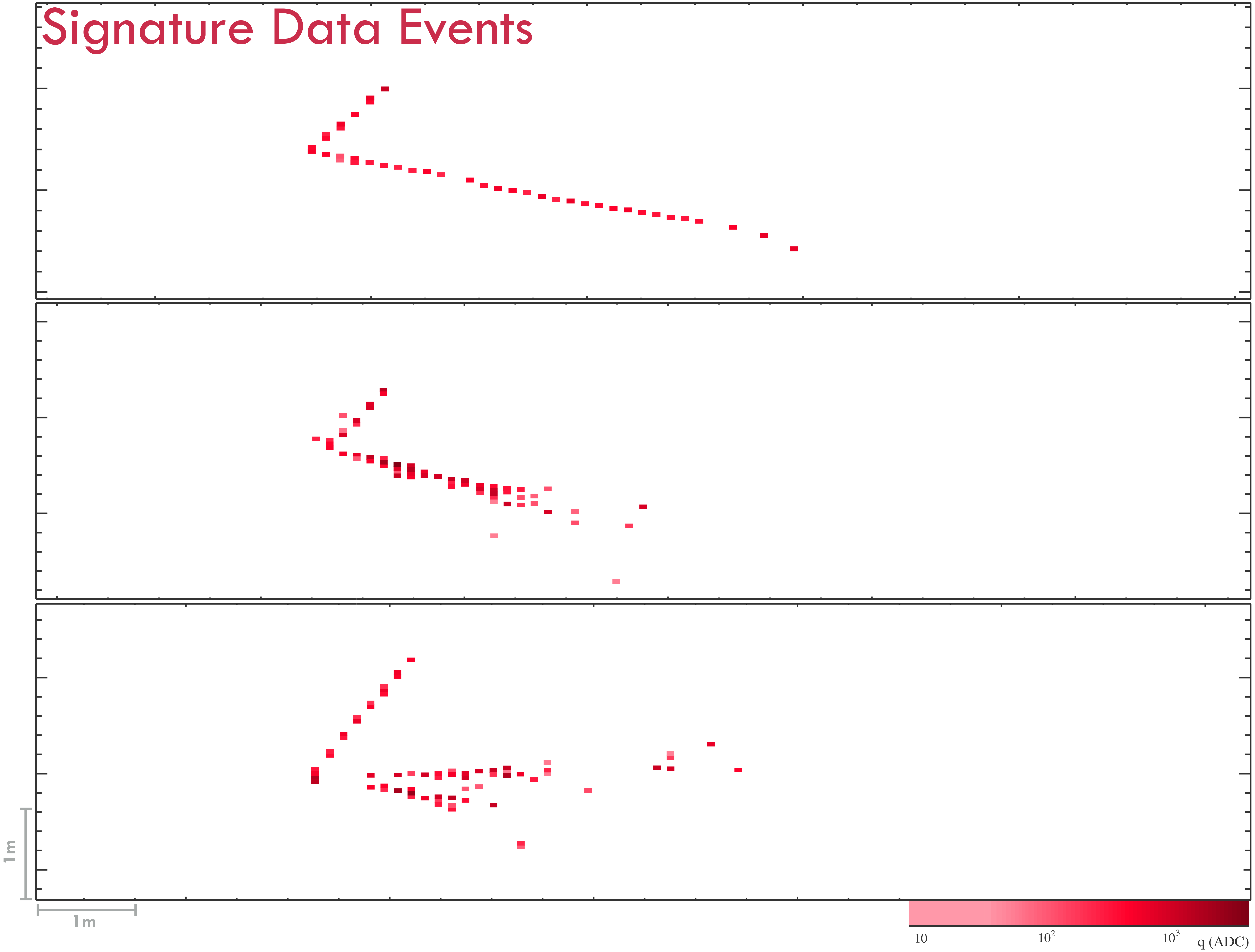
Neutrino interactions are **flavor conserving**, thus, they can be identified from the outgoing particles.

Charged Current Interactions

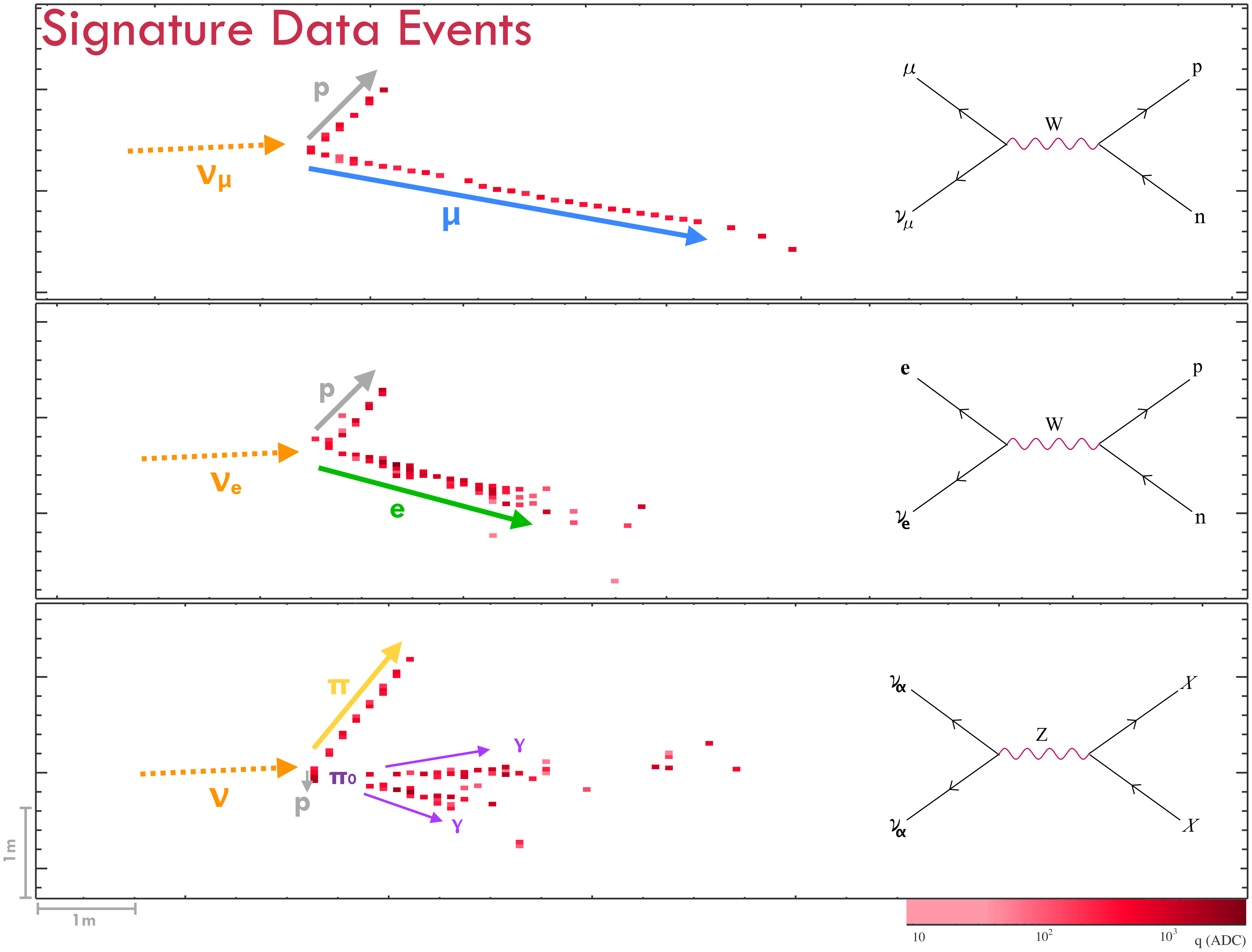


Neutral Current Interactions

Signature Data Events



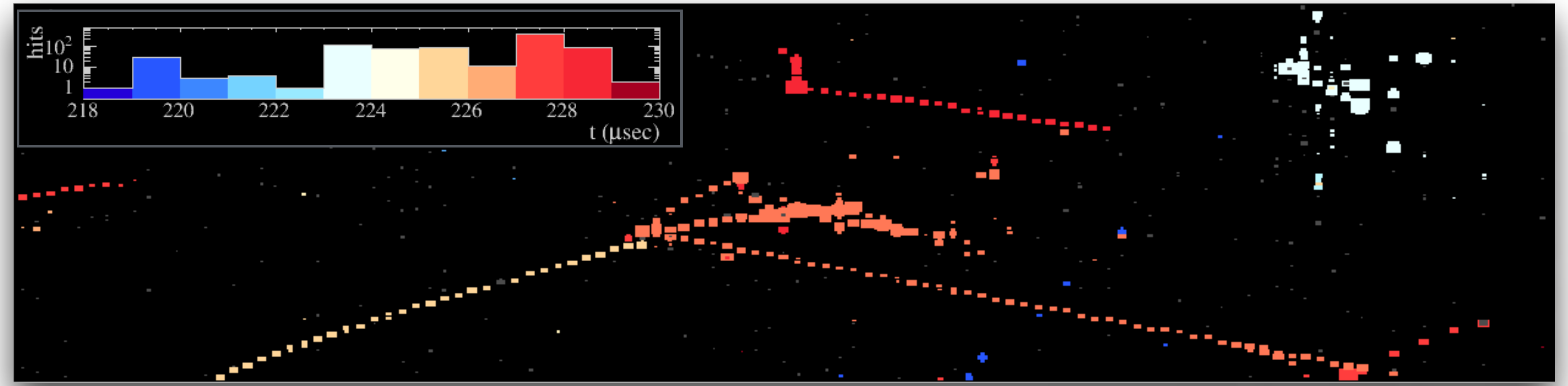
Signature Data Events

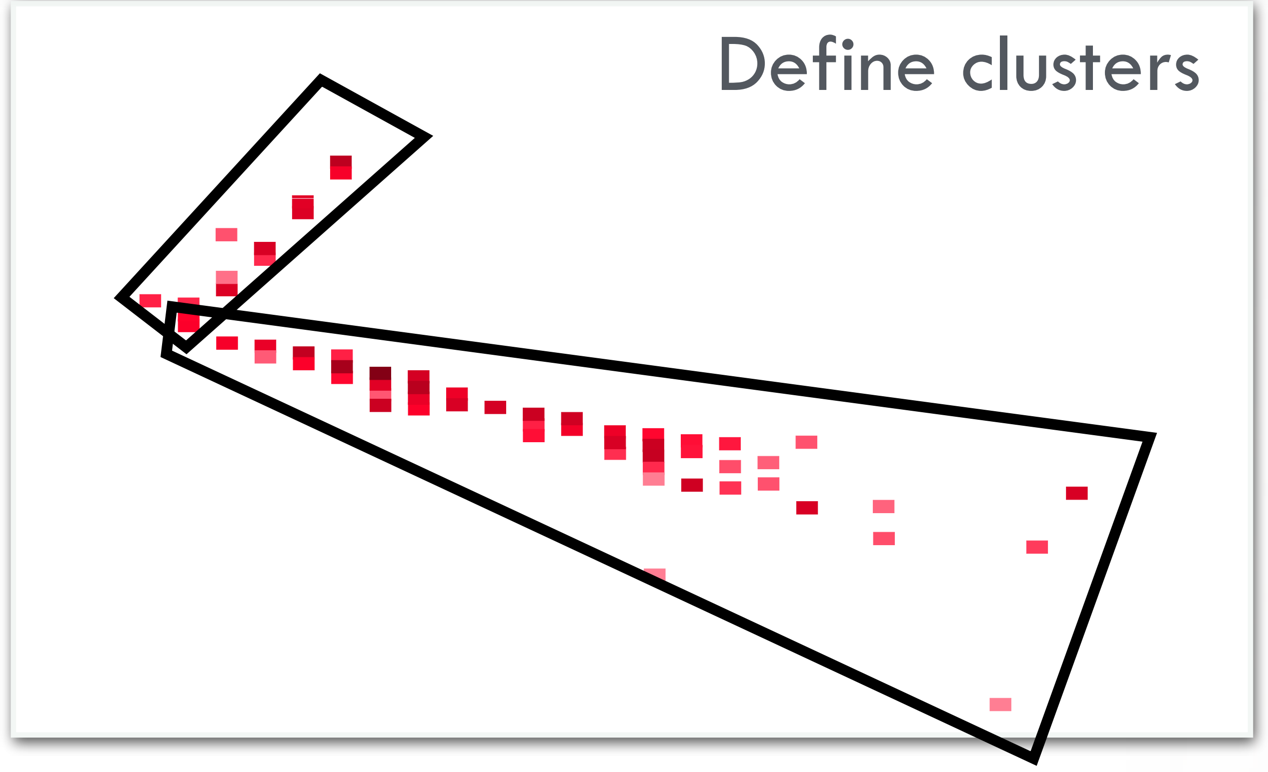


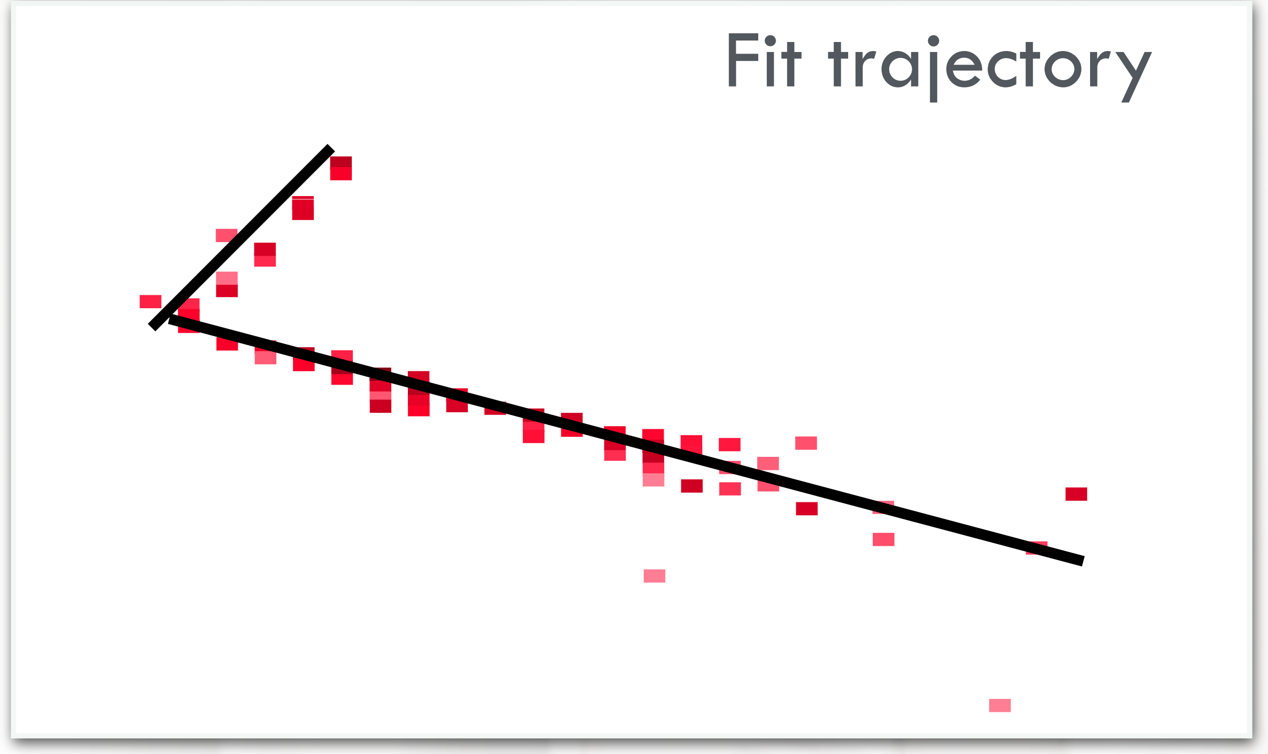
Traditional Reconstruction

Use the topology and magnitude of the energy depositions.

Takes advantage of the granularity and time resolution of our detectors.







We isolate individual interactions using time and space correlation of the hits.

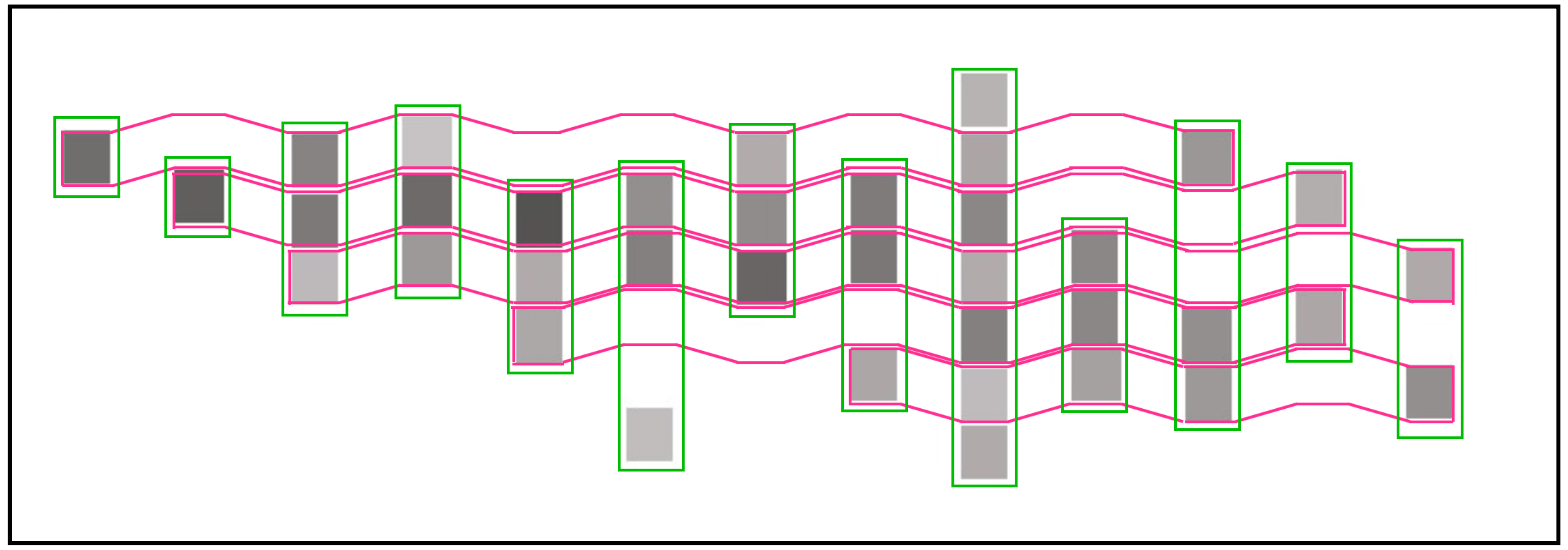
Groups of hits can be clustered as following the path of same particle starting at the interaction point.

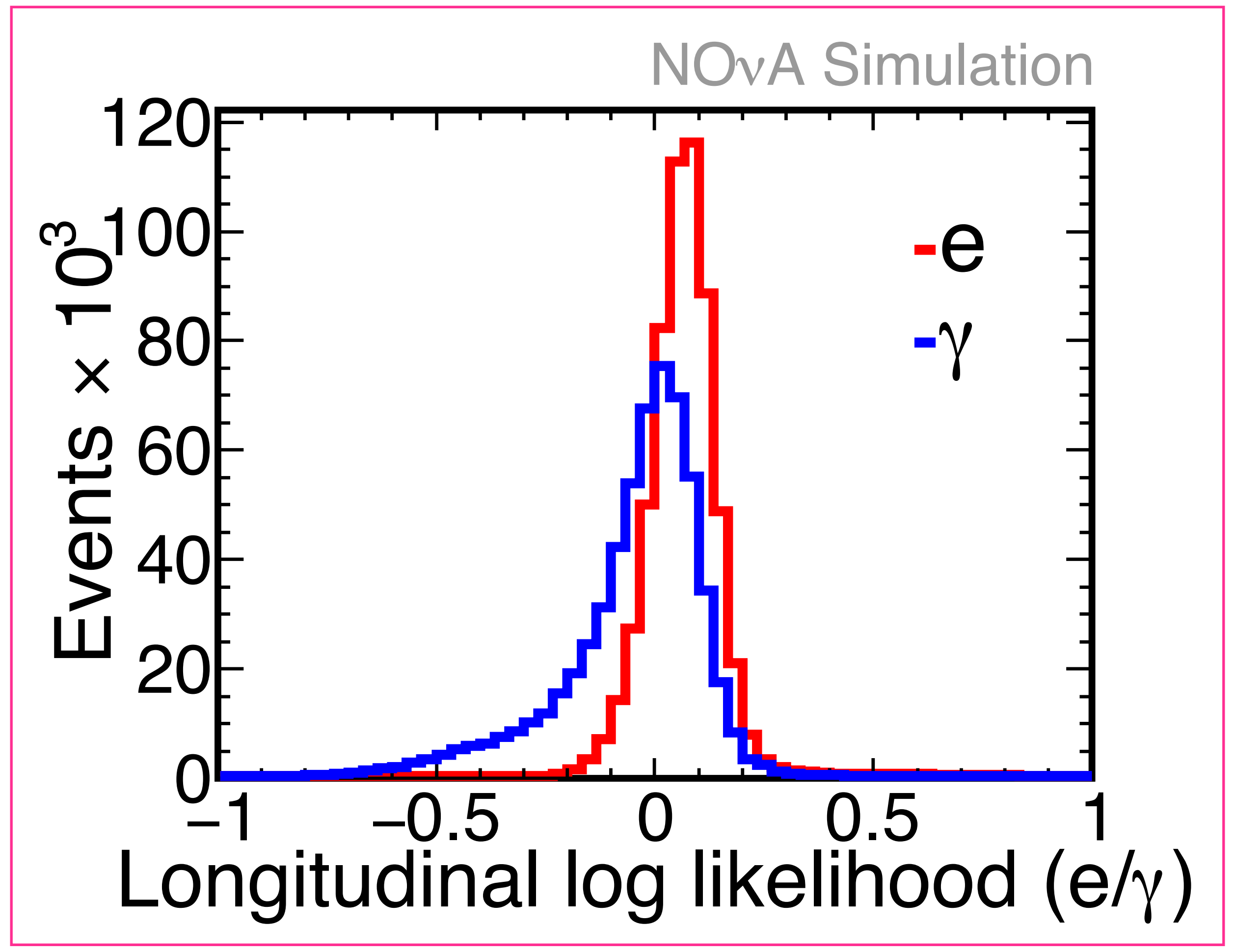
When necessary we can fit an assumed trajectory for each cluster of hits.

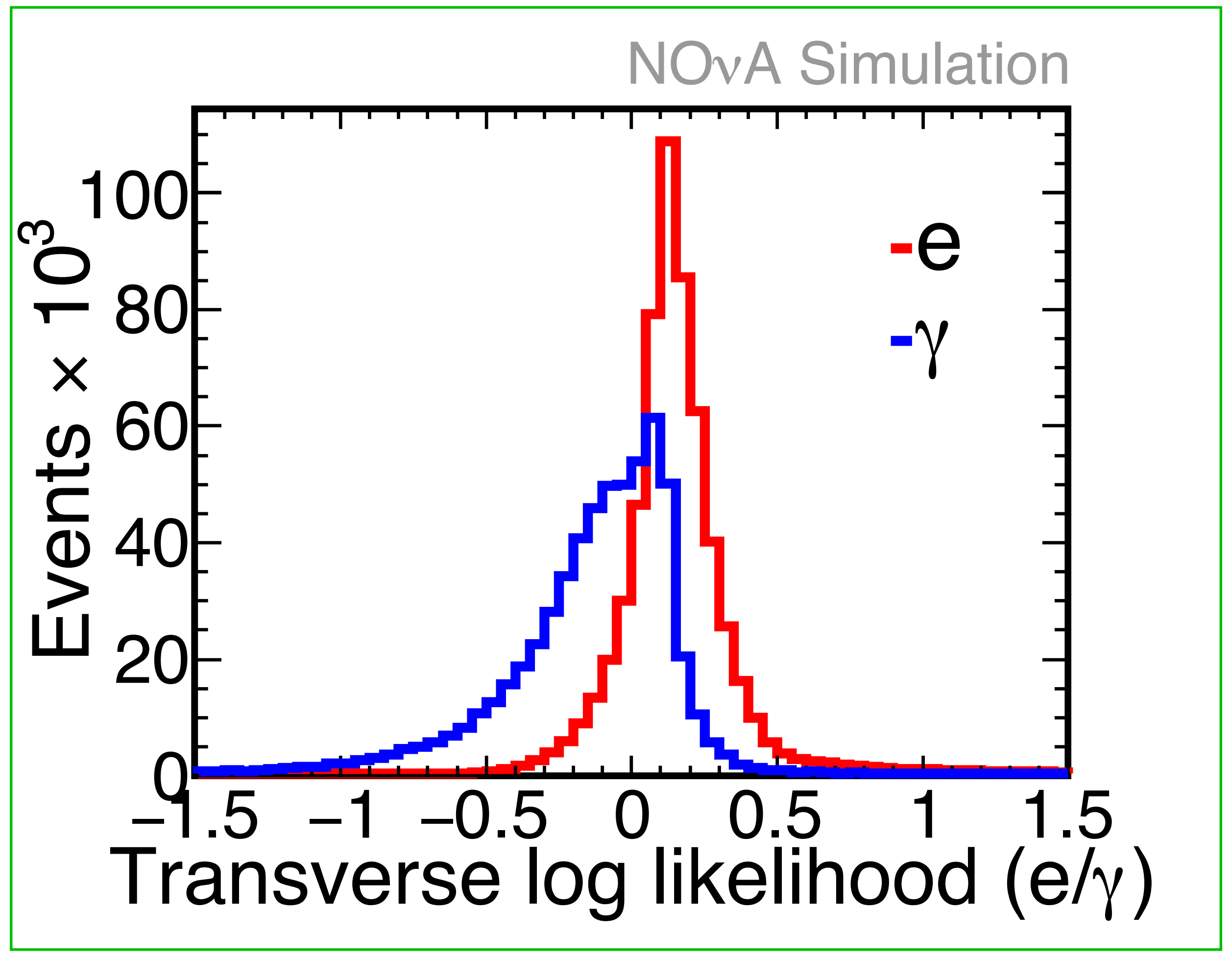
Traditional ID Methods

Mostly focused on identifying the lepton in the event. Extracted features (i.e. track length and scattering for muons, topology of energy depositions for electromagnetic showers)

- ★ **Require Previous reconstruction.**
- ★ **Features are pre-defined, based on MC or test data.**







Example: The Likelihood ID method

- ★ Reconstruct electron shower.
- ★ Find likelihoods from it's dE/dx profiles compared to particle hypotheses.

Likelihoods → *Traditional Neural Network*

ID with Convolutional Neural Networks

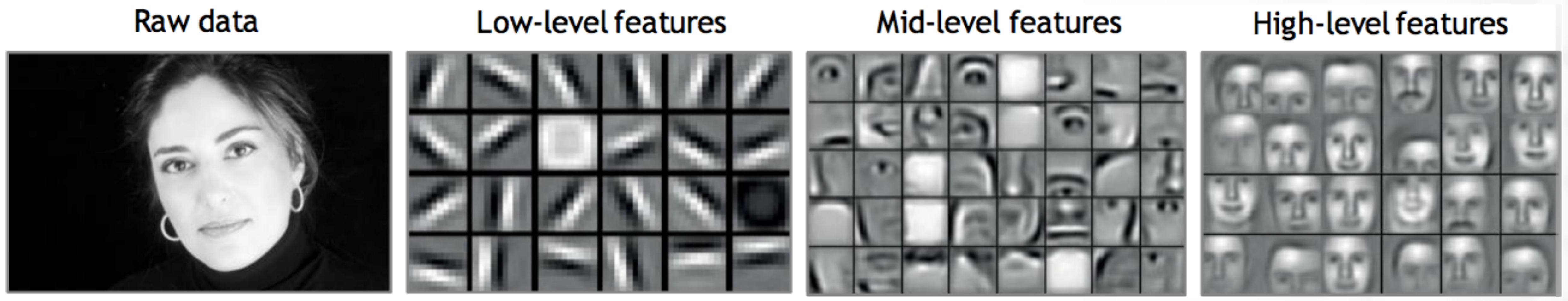
Premise: Rather than select a set of features a priori, let a deep learning network extract features and draw correlations.

In practice: Use “images” of our events to train Convolutional Neural Networks (CNNs) to identify neutrino interactions.

Disentangle the identification from traditional reconstruction.

Allow for features apart from those based in our assumptions of the physics.***

Explore the potential of deep learning beyond event identification.



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

ID with Convolutional Neural Networks

Premise: Rather than select a set of features a priori, let a deep learning network extract features and draw correlations.

In practice: Use “images” of our events to train Convolutional Neural Networks (CNNs) to identify neutrino interactions.

Disentangle the identification from traditional reconstruction.

Allow for features apart from those based in our assumptions of the physics.***

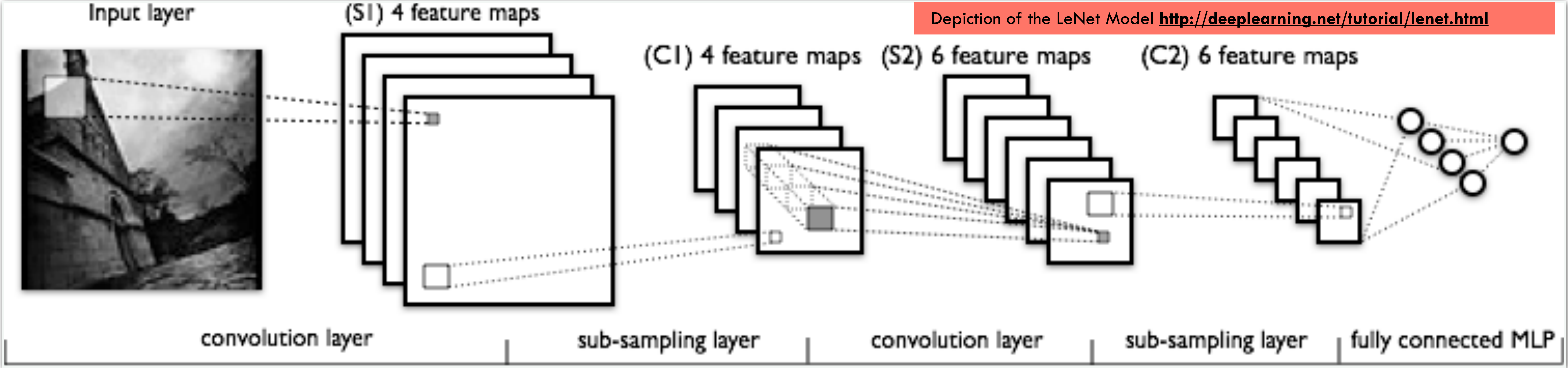
Explore the potential of deep learning beyond event identification.

In absence of test data, these methods **rely on the simulations***

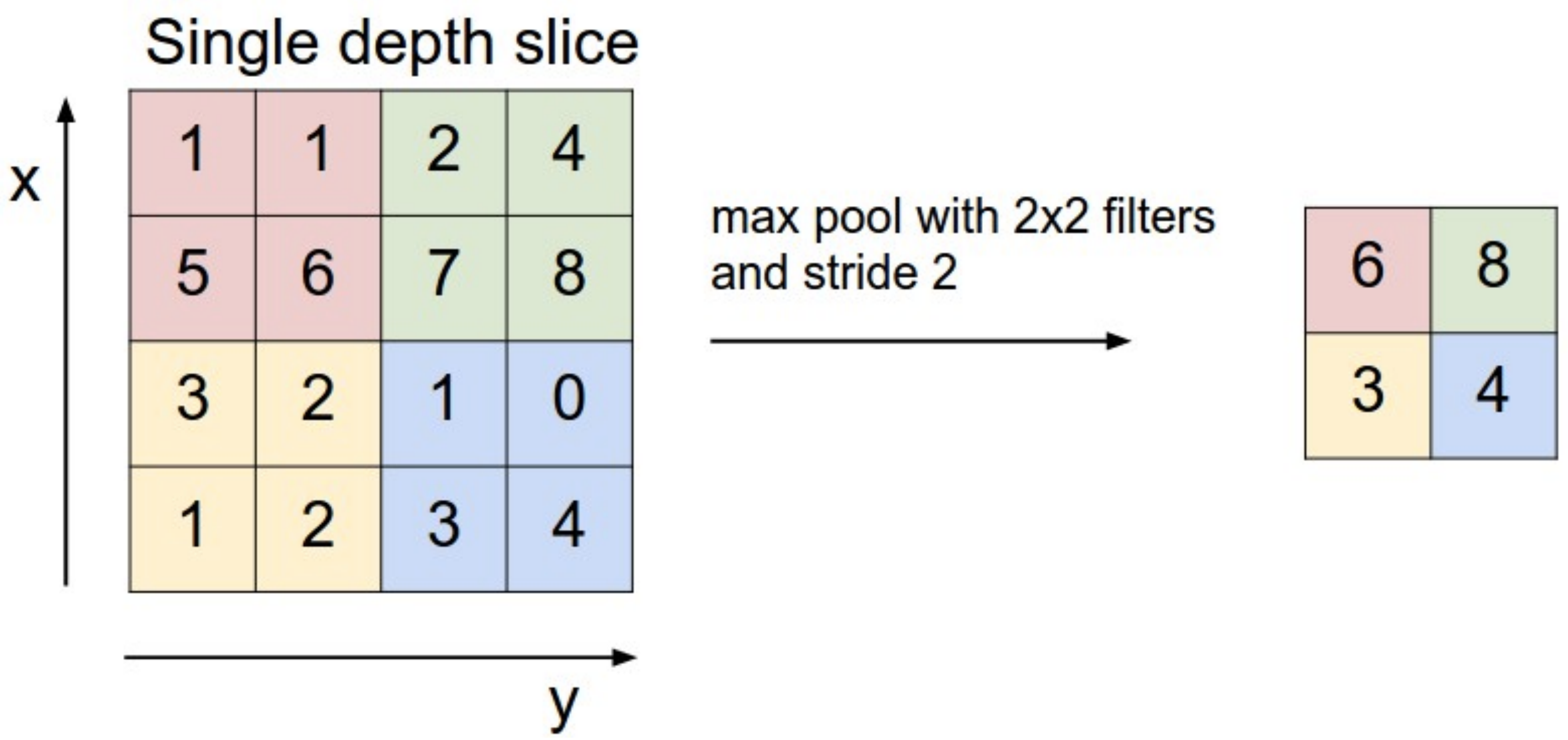
Known features of trained networks like over-training and saturating loss functions.

Network Layers

The simplest form of a CNN includes convolutional layers, max pooling layers and MLP layers.



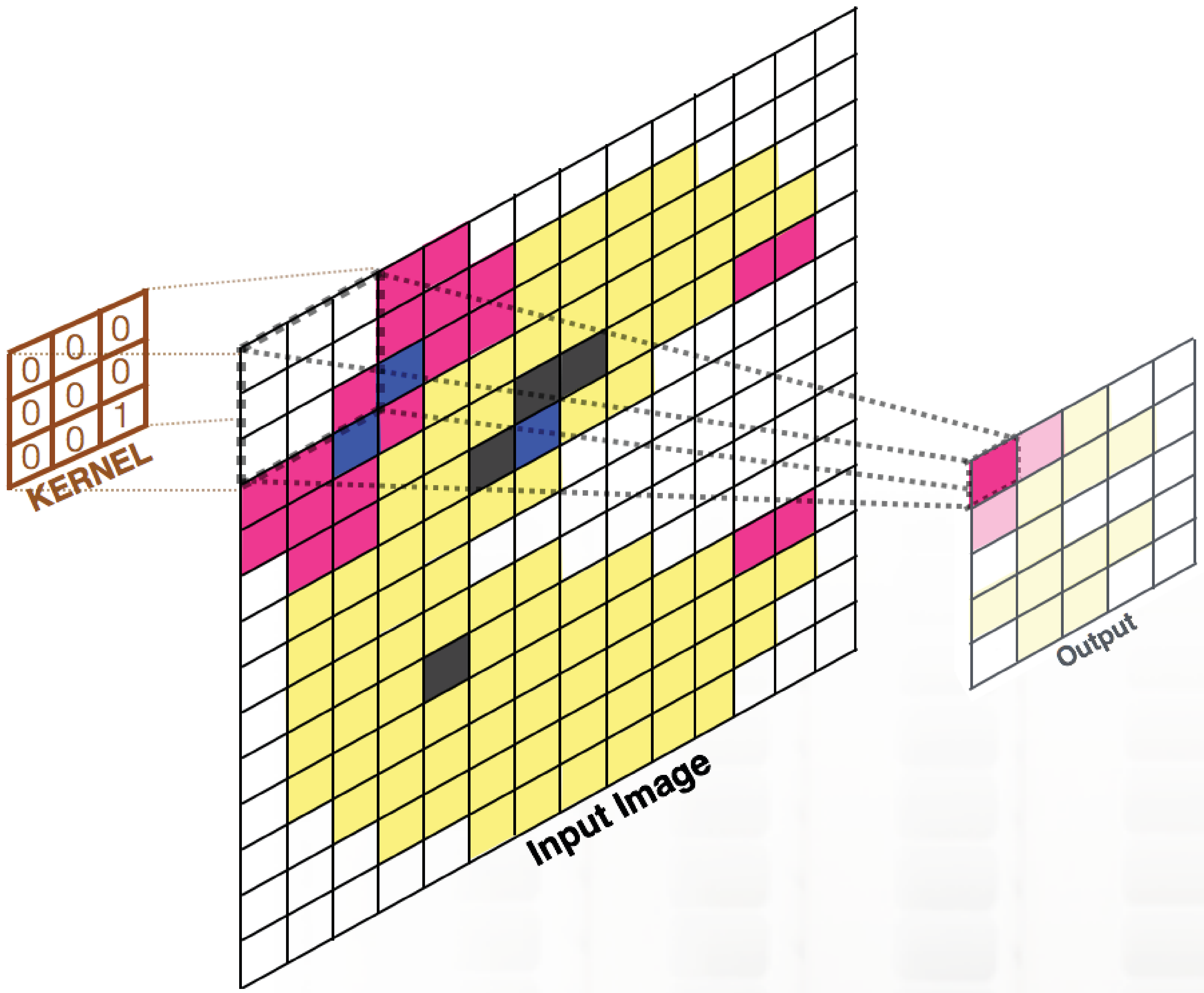
Depiction of the LeNet Model **http://deeplearning.net/tutorial/lenet.html**



Pooling Layers:

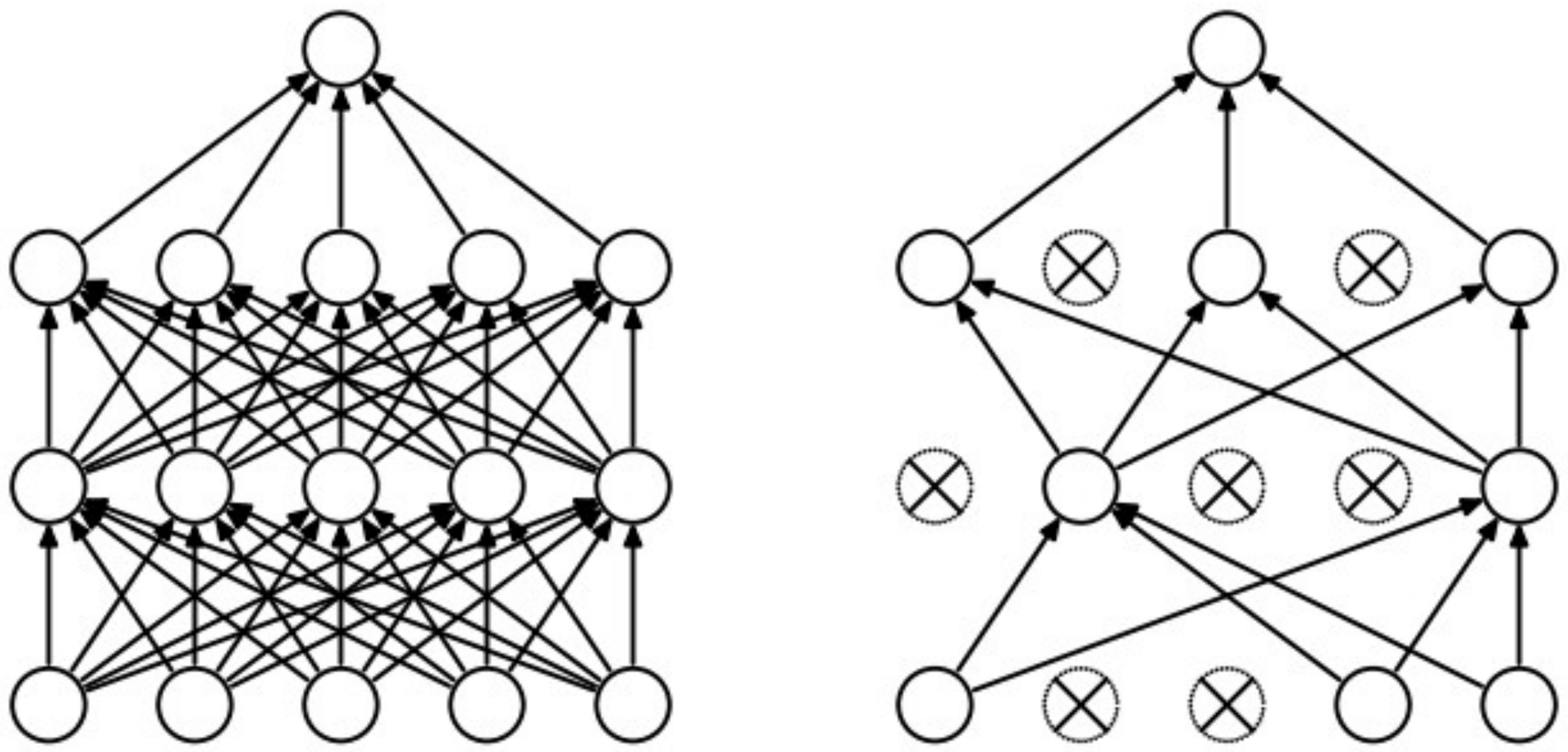
Down-sampling is done by performing operations (average, max, etc) on the feature maps while still preserving the information.

Network Layers



Kernel Renormalization:

Kernels evolve as the training progresses through renormalization. This process uses non saturating functions.



(a) Standard Neural Net

(b) After applying dropout.

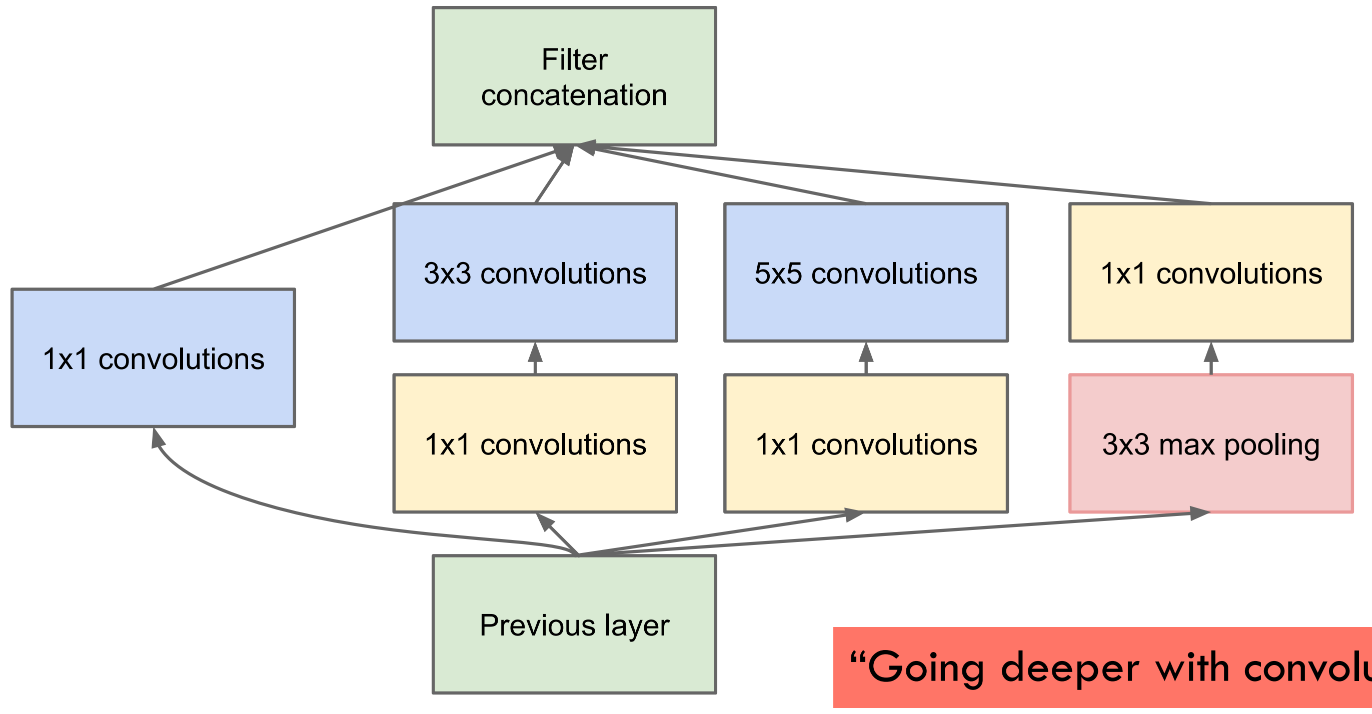
Dropout:

Randomly reset weights, effectively removing whole nodes at each step.

Encourages complex dependence and discourages overtraining

CVN Convolutional Visual Network

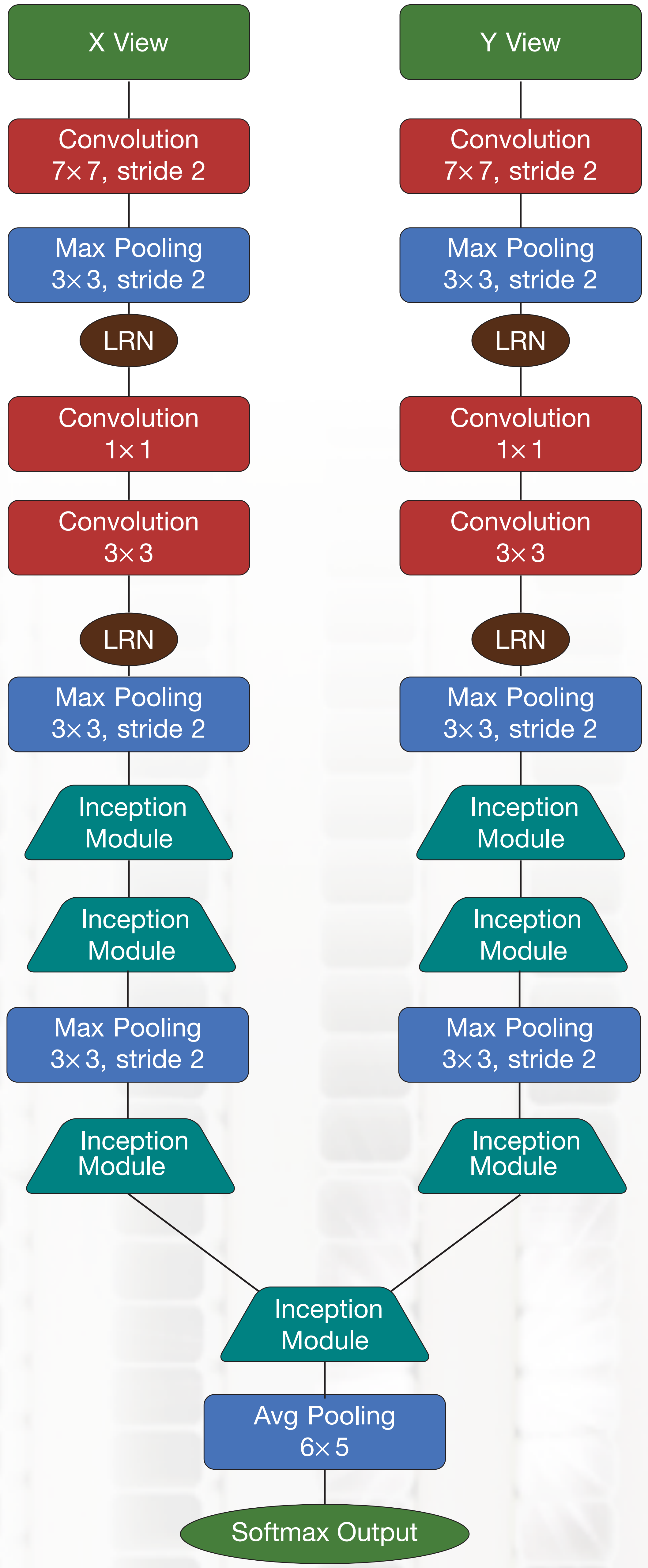
Neutrino Event CVN: Siamese network architecture based on GoogLeNet.



"Going deeper with convolutions" **arXiv:1409.4842**

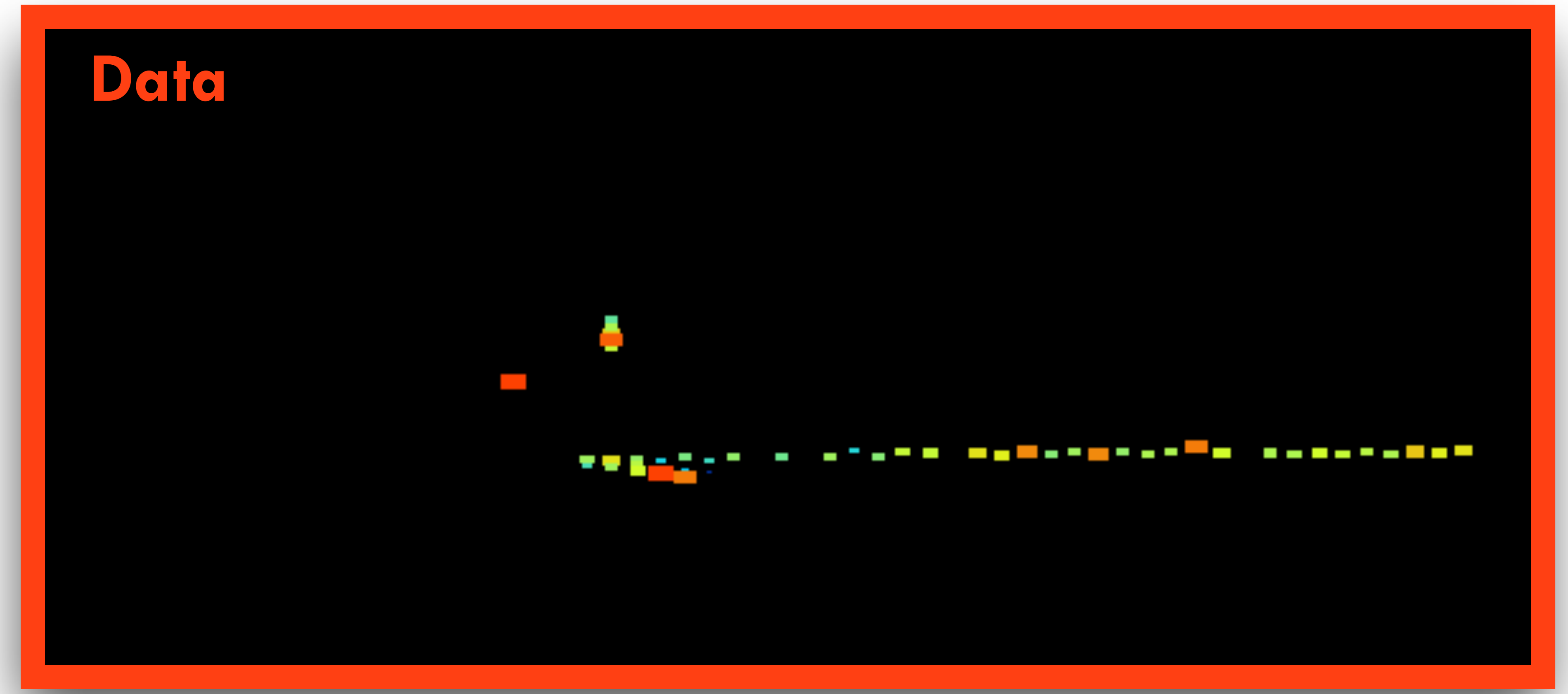
Inception modules:

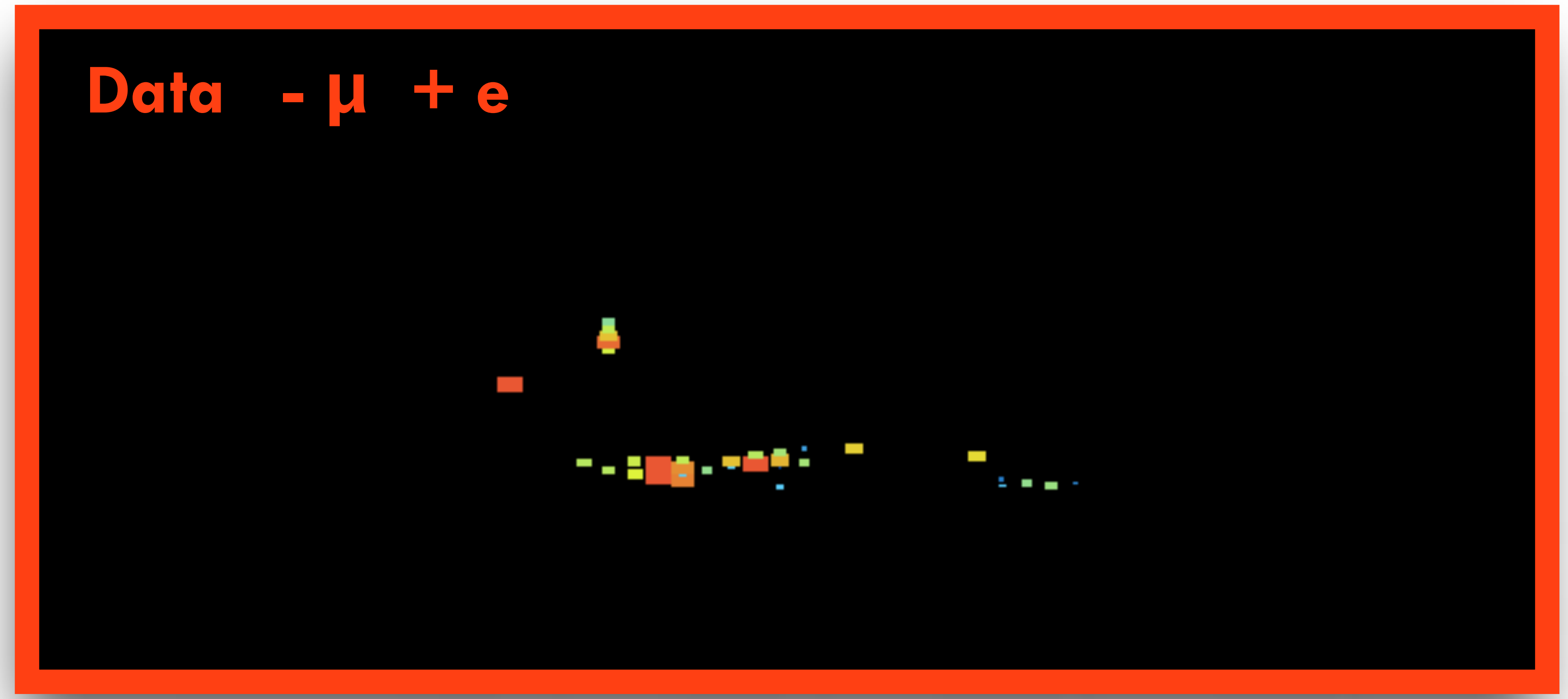
Network in network model with kernels of multiple dimensions

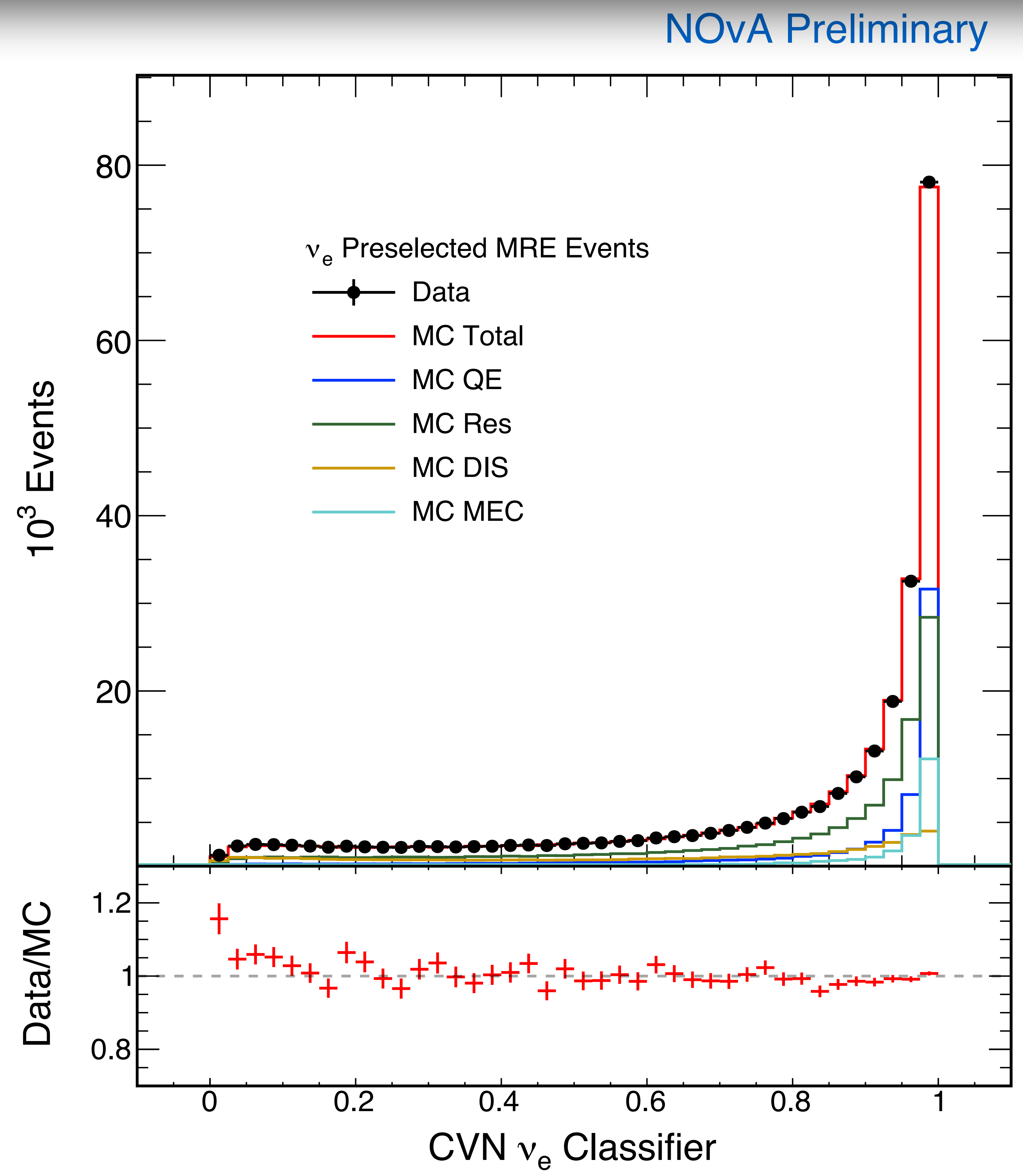


- Inspired by siamese architectures to allow the network to learn from features on each 2D view of the event.
- Using the caffe framework http://caffe.berkeleyvision.org/
- We train on Fermilab's Wilson cluster GPUs **(2 K40s)**
- Trained on 4.7 million simulated events of all neutrino interaction types plus cosmic rays

CVN Performance On Real Data







MRE (Muon Removed - Electron):

Select a muon neutrino interaction with traditional ID methods.

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	

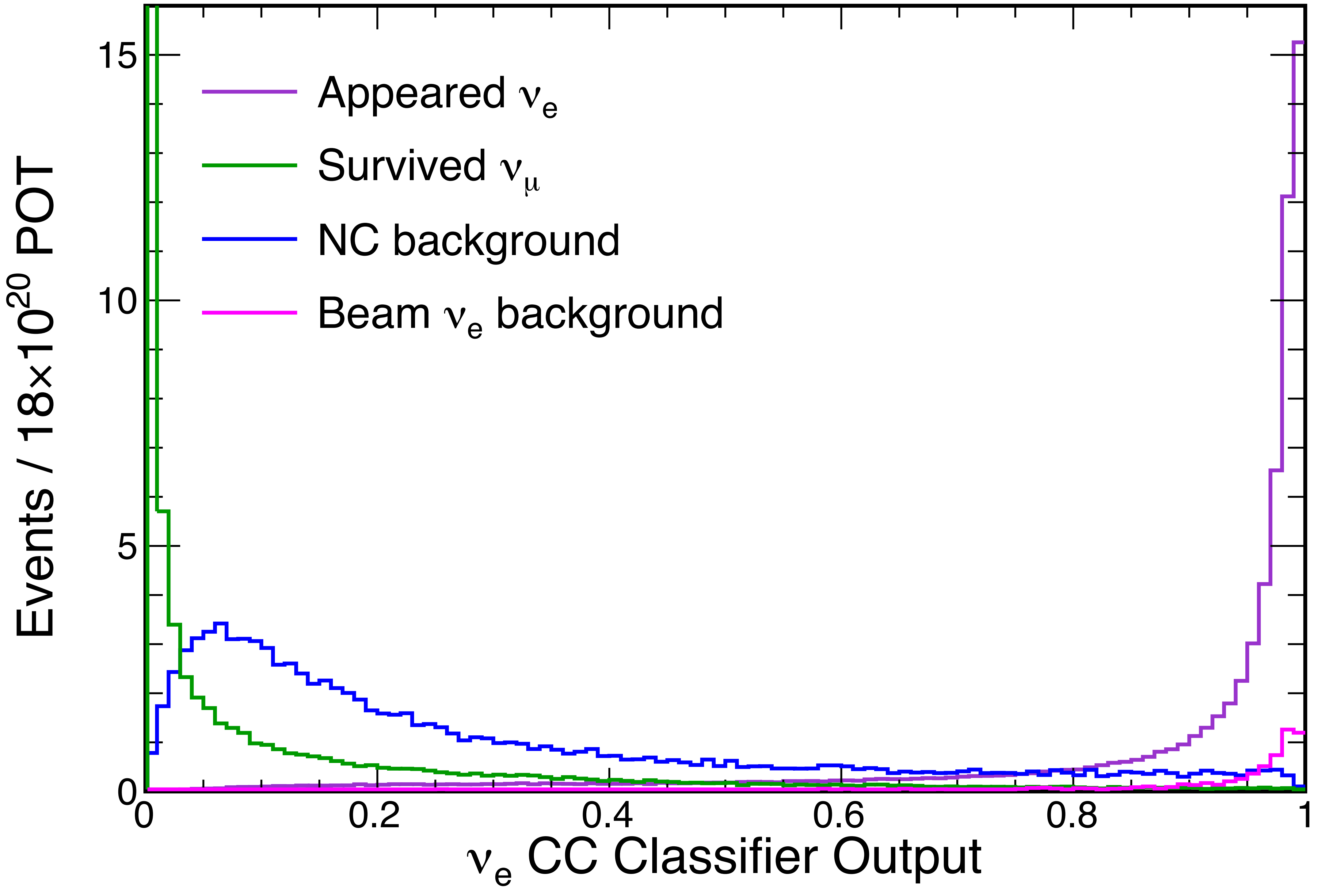
CVN Performance on ν_e

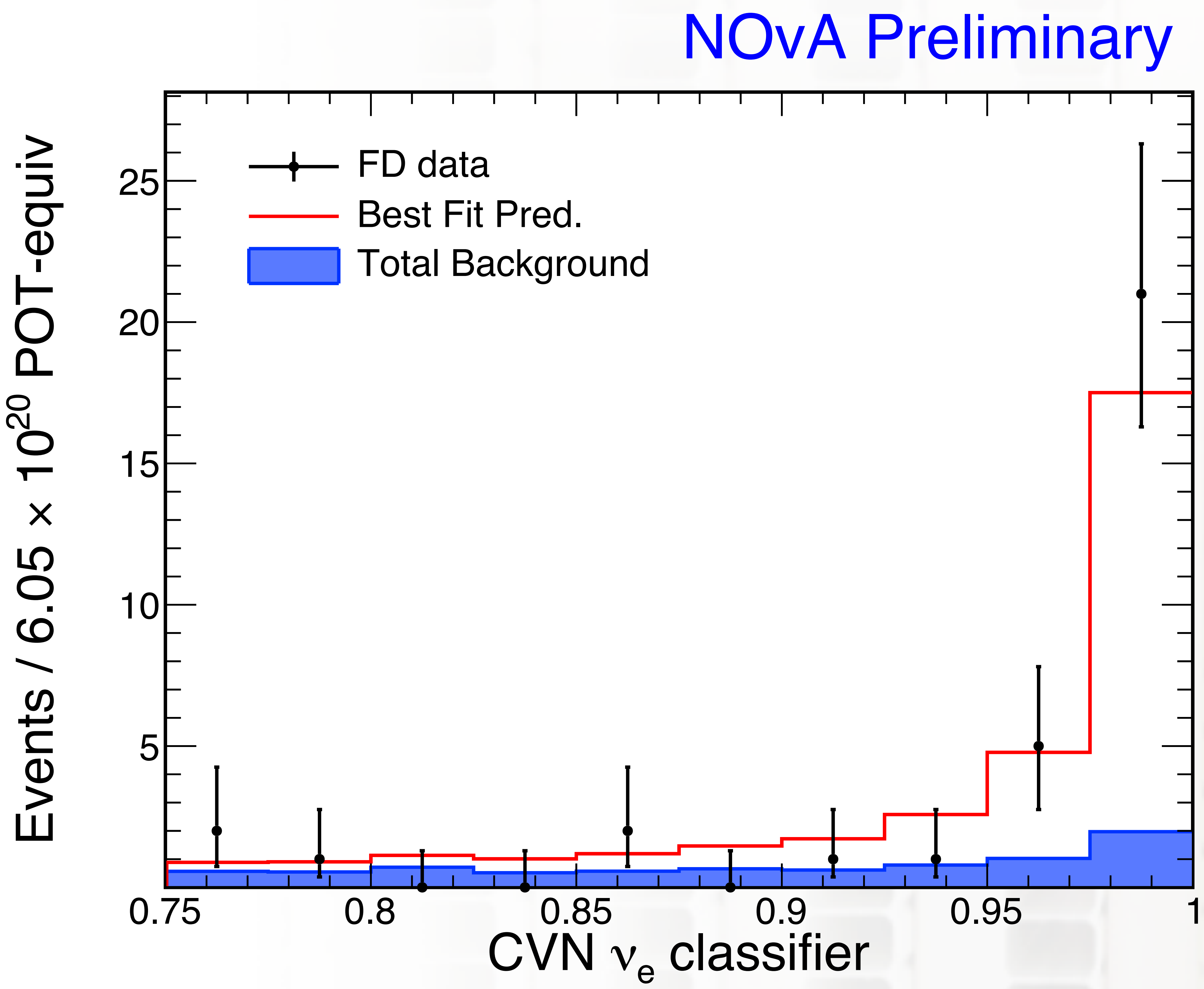
Implemented in NOvA's main analysis for the results shown this summer at Neutrino 2016

Total bkg	NC	Beam ν_e	ν_μCC	ν_τCC	Cosmogenic
8.2	3.7	3.1	0.7	0.1	0.5

33 events selected with estimates background of ~8

76% Purity, 73% Efficiency and an equivalent **increased exposure of 30%**

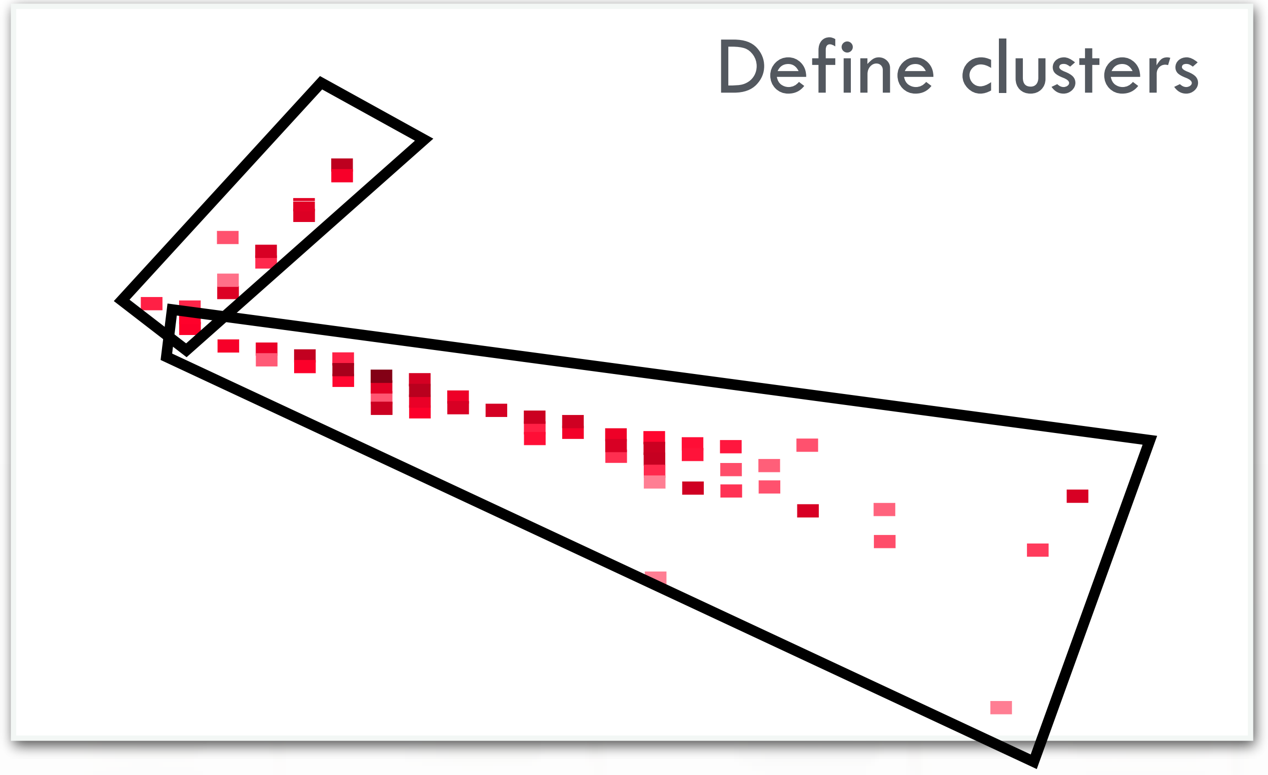


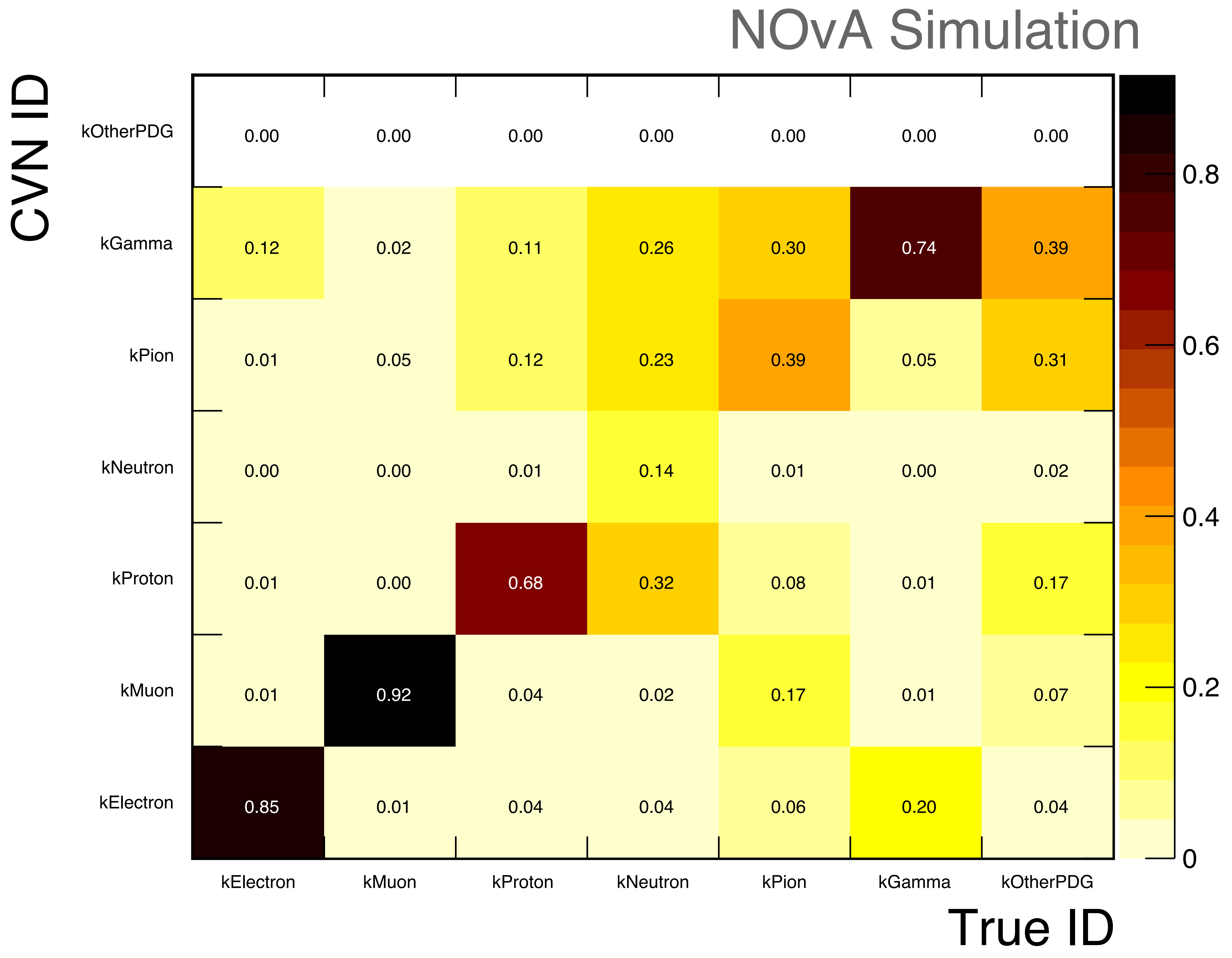


Ongoing Work CVN and Reconstruction

Using the existing reconstruction.

Classify clusters by particle ID





Original CVN network modified to take 4 views (event + prong)

Trained on 50% purity prongs from all events no preselection

Room for improvement in classification and network optimization

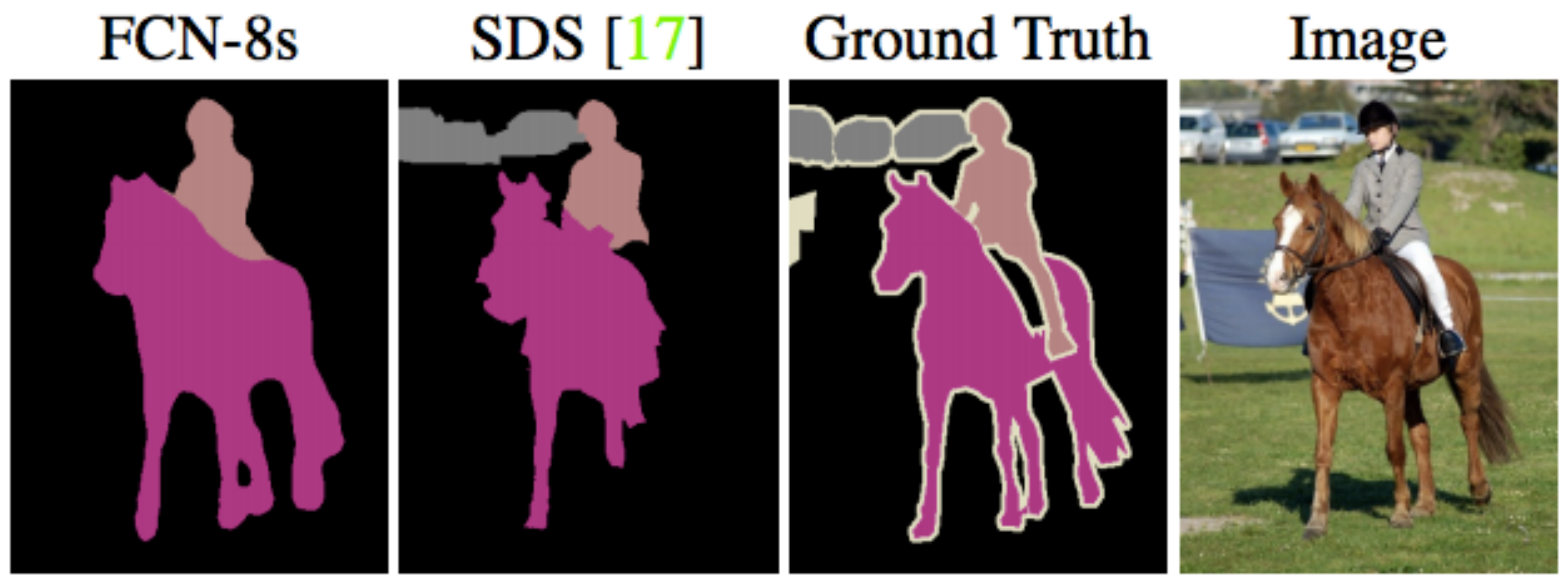
Ongoing Work CVN and Reconstruction

Contributing to reconstruction.

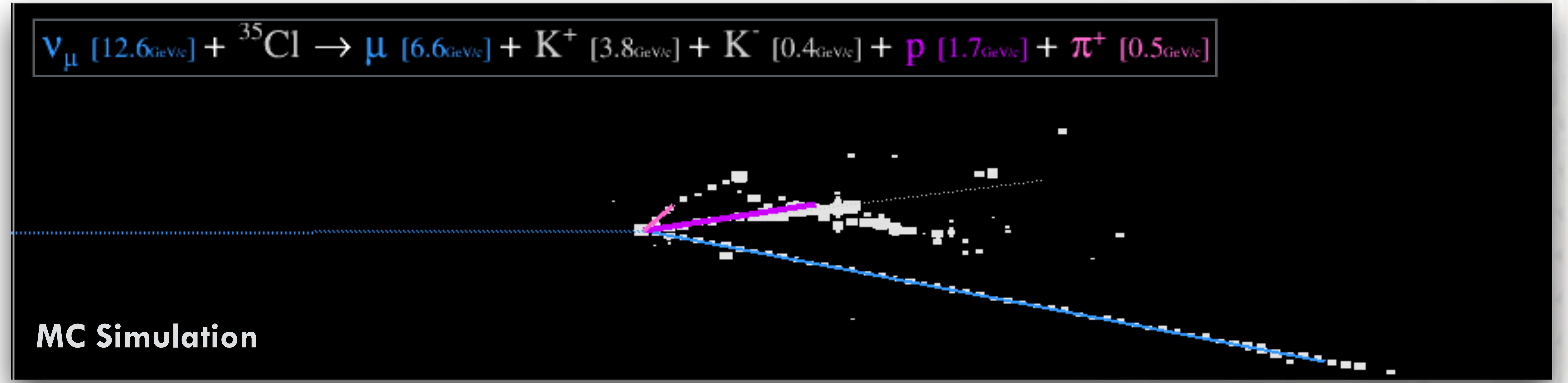
There are CNN implementations in the literature for pixel by pixel classification using semantic segmentation.

In our events that means classify individual hits by the particle which caused them.

Initial studies are ongoing to compare the performance of SS to traditional clustering and the existing particle CVN identifier.



Fully Convolutional Networks for Semantic Segmentation **arXiv:1605.06211**



Summary

Convolutional Visual Network

CVN is our implementation of CNNs for neutrino event classification.

- ⋆It effectively increases out exposure by 30% compared to traditional ID methods.
- ⋆Studies show promise on other analyses, like the muon neutrino disappearance.
- ⋆**Currently being used for multiple physics analysis.**
- ⋆NOvA's nue appearance analysis is the **first implementation of a CNN in a HEP result.**

CVN Paper: *"A Convolutional Neural Network Neutrino Event Classifier"*

A.Aurisano et. al. JINST 11 (2016) no.09, P09001

NOvA's Latest results: Neutrino 2016 *"New Results from NOvA " LINK*

An implementation of **CVN** for cluster/particle classification is in testing stages.

There is ongoing work for **hit classification using semantic segmentation.**

Ongoing studies are learning about the interplay between traditional reconstruction and image classification techniques.

Backup

These are not the slides you're looking for

CVN Convolutional Visual Network

Is NOvA's implementation of Convolutional Neural Networks

Neutrino Event CVN:

Classifier for events in a sampling calorimeter by neutrino interaction type.

For the Electron Neutrino analysis:

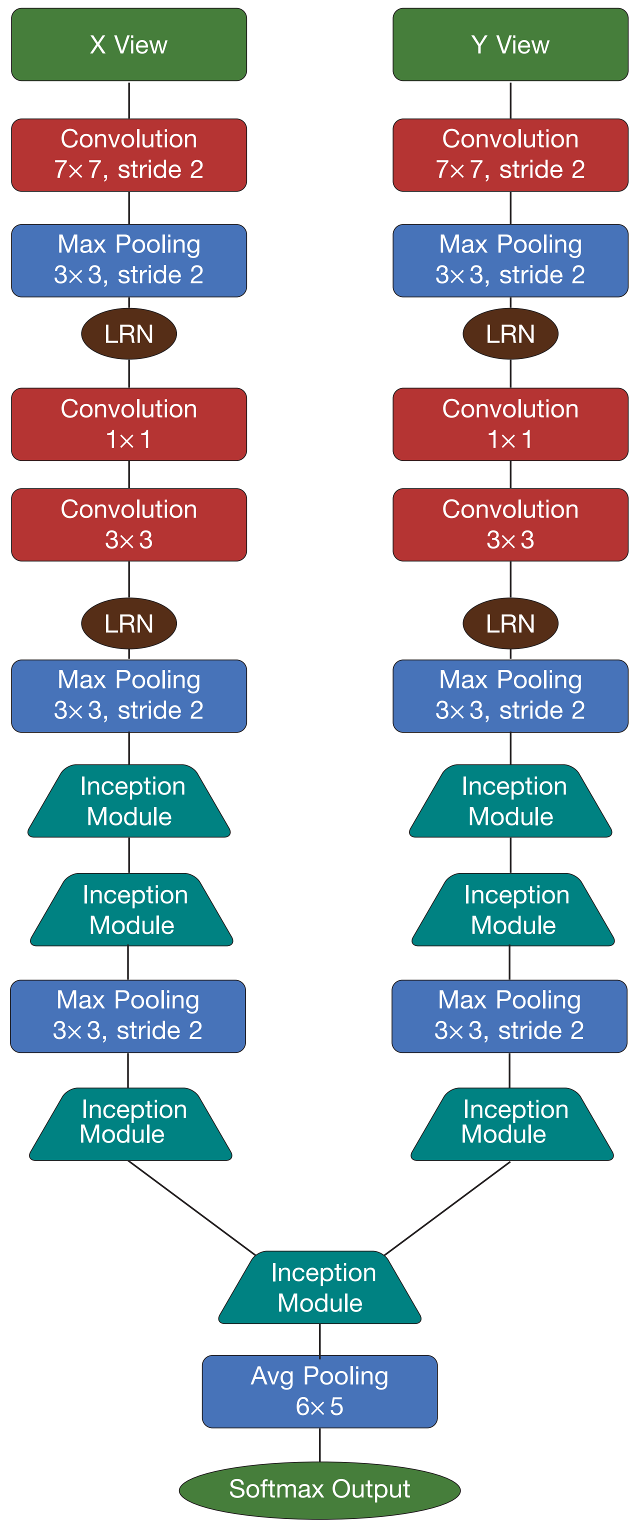
76% purity, 73% efficiency and a 30% equivalent increase in exposure.

NOvA's nue appearance analysis is the **first implementation of a CNN in a HEP result.**

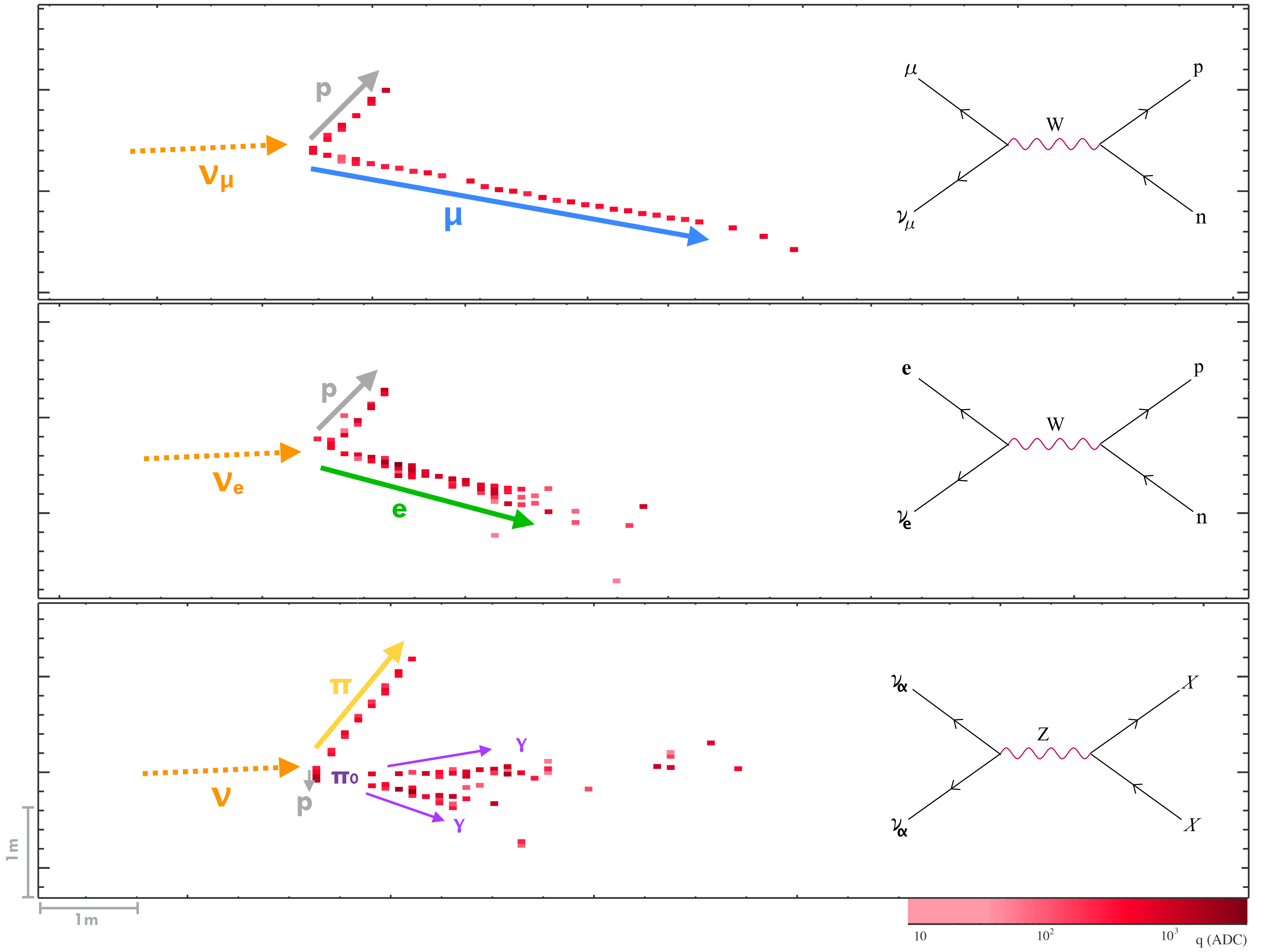
> ***CVN Paper:*** *"A Convolutional Neural Network Neutrino Event Classifier" A.Aurisano et. al. JINST 11 (2016) no.09, P09001*

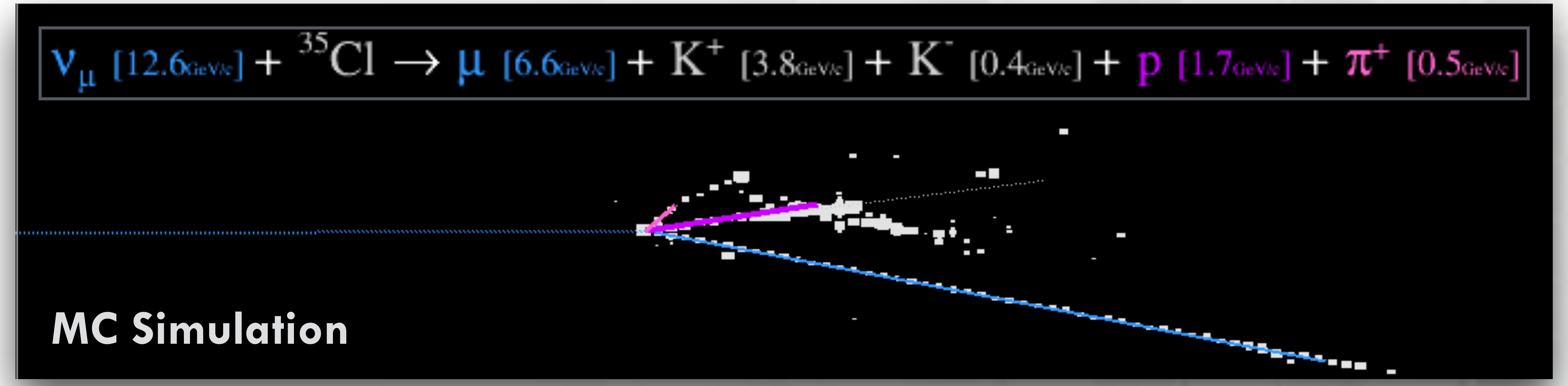
CVN for Reconstruction:

Ongoing studies to identify hit by hit in an event. This type of identification could influence the existing approaches at reconstruction.

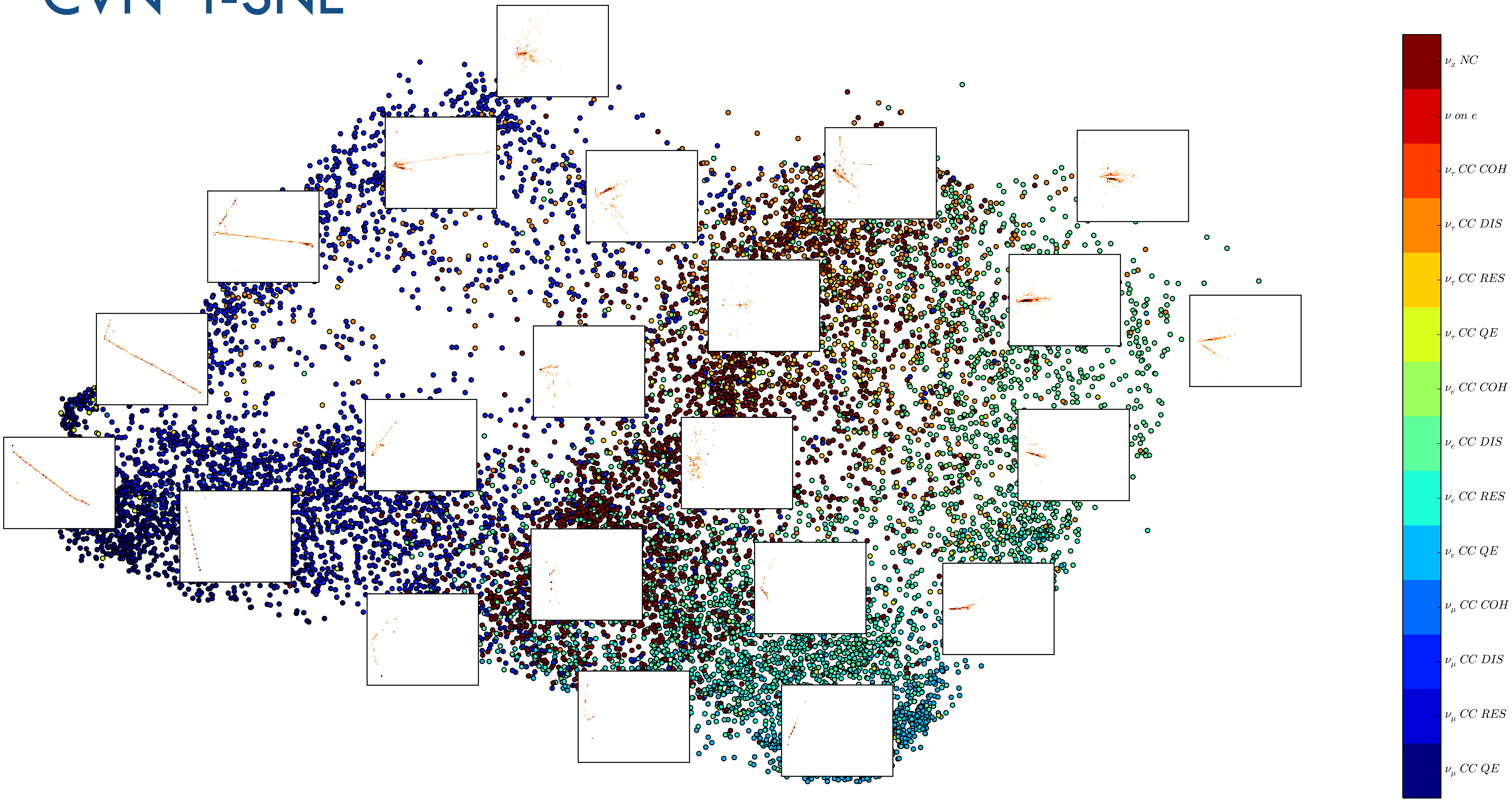


Summary

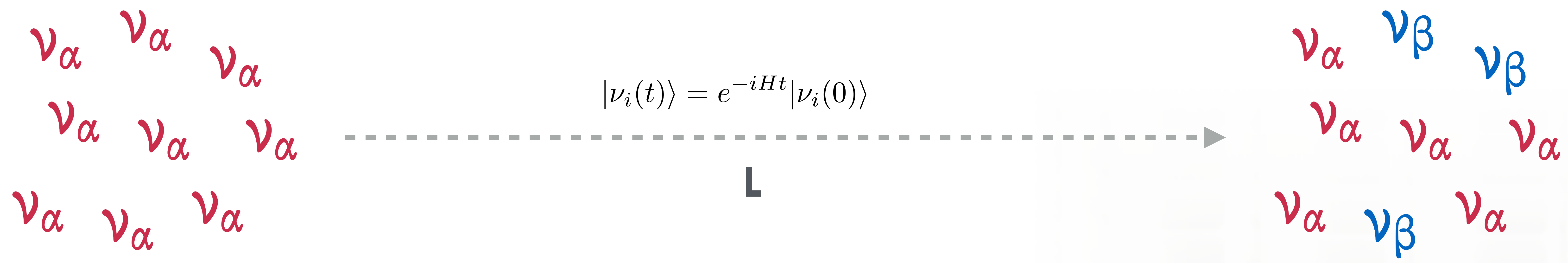




CVN t-SNE

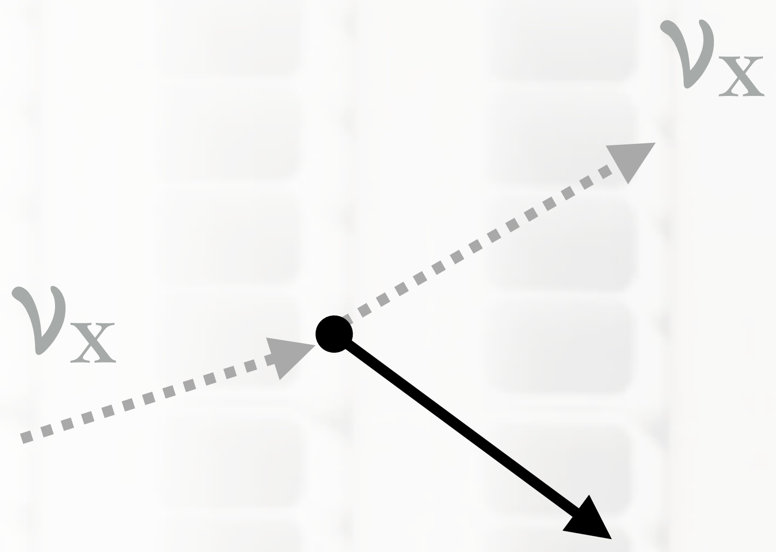

ν_x NC
ν on e
ν_τ CC COH
ν_τ CC DIS
ν_τ CC RES
ν_τ CC QE
ν_e CC COH
ν_e CC DIS
ν_e CC RES
ν_e CC QE
ν_μ CC COH
ν_μ CC DIS
ν_μ CC RES
ν_μ CC QE

Neutrino Interactions



Neutrino interactions are **flavor conserving,** thus, they can be identified from the outgoing particles.

Charged Current Interactions



Neutral Current Interactions

Neutrino Oscillations

The neutrino flavor eigenstate oscillations are described by the PMNS matrix.

PMNS Matrix

$$\begin{vmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{matrix}\right\rangle = \begin{pmatrix} U_{e1}^* & U_{e2}^* & U_{e3}^* \\ U_{\mu 1}^* & U_{\mu 2}^* & U_{\mu 3}^* \\ U_{\tau 1}^* & U_{\tau 2}^* & U_{\tau 3}^* \end{pmatrix} \left|\begin{matrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{matrix}\right\rangle$$

Flavor Eigenstates

Mass Eigenstates

ν_α ν_α ν_α ν_α ν_α ν_α ν_α ν_α ν_α

$|\nu_i(t)\rangle = e^{-iHt}|\nu_i(0)\rangle$

L

ν_α ν_β ν_β ν_α ν_α ν_α ν_α ν_β ν_α

$$P(\nu_\alpha \to \nu_\beta) = \left| \sum_i U_{\beta i} U_{\alpha i}^* e^{-i m_i^2 L/2E} \right|^2$$

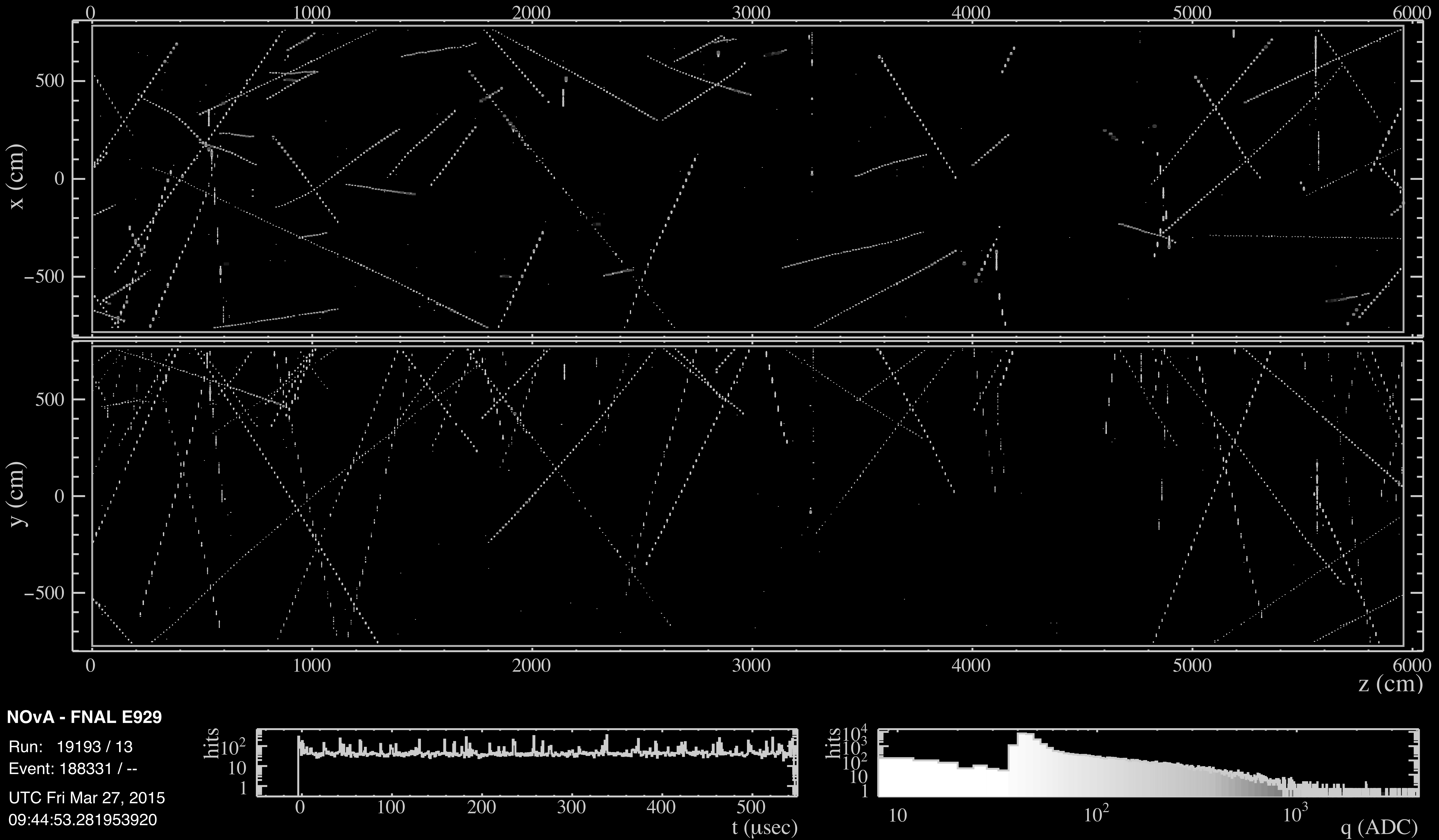
Oscillation probability

The goal of oscillations experiments is to determine the PMNS parameters via oscillation probabilities.

The measurable in these experiments is a count or energy spectrum.

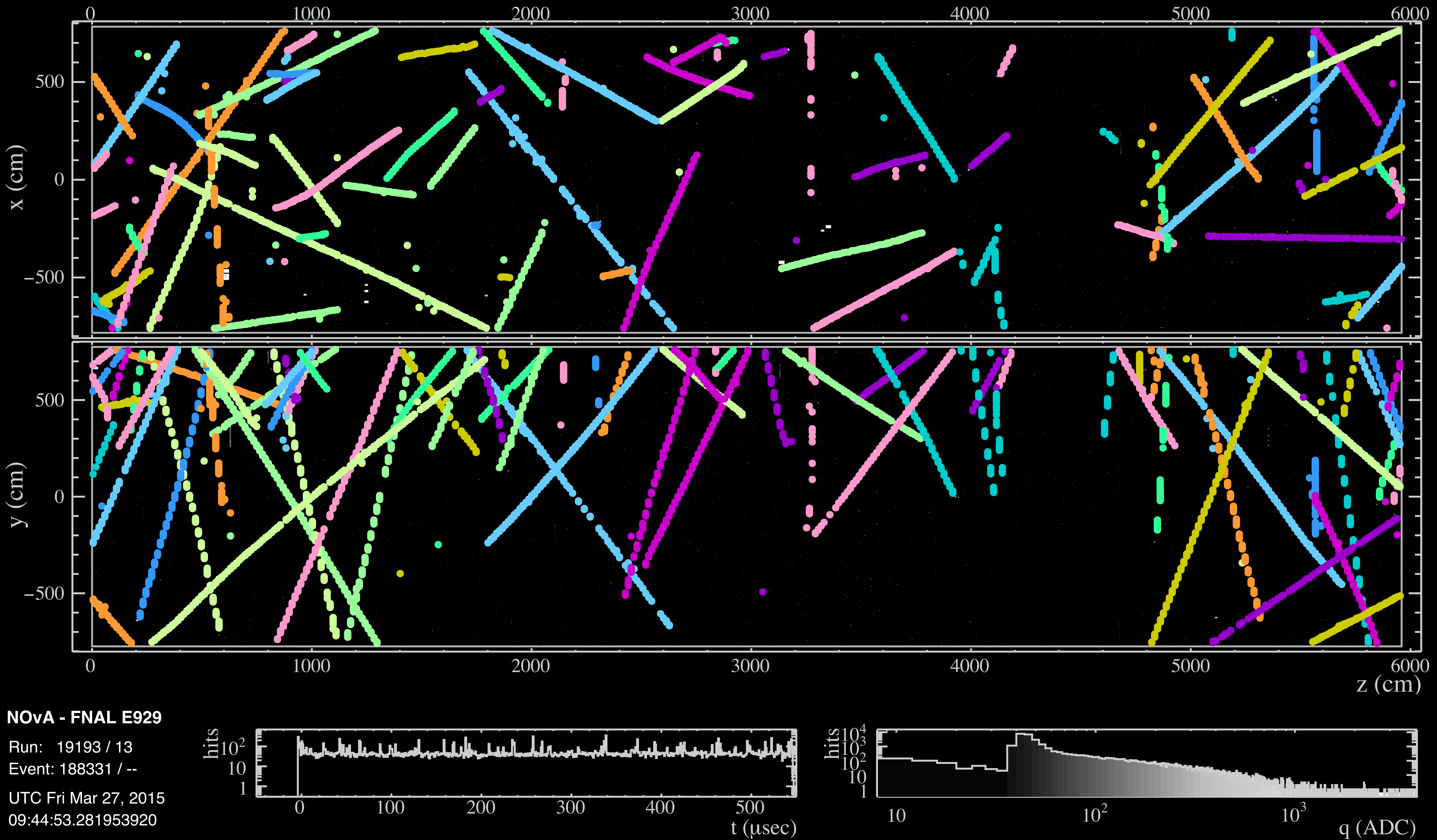
Isolating neutrino interactions

The first step in our reconstruction is dividing an **event** (550 μs of data)

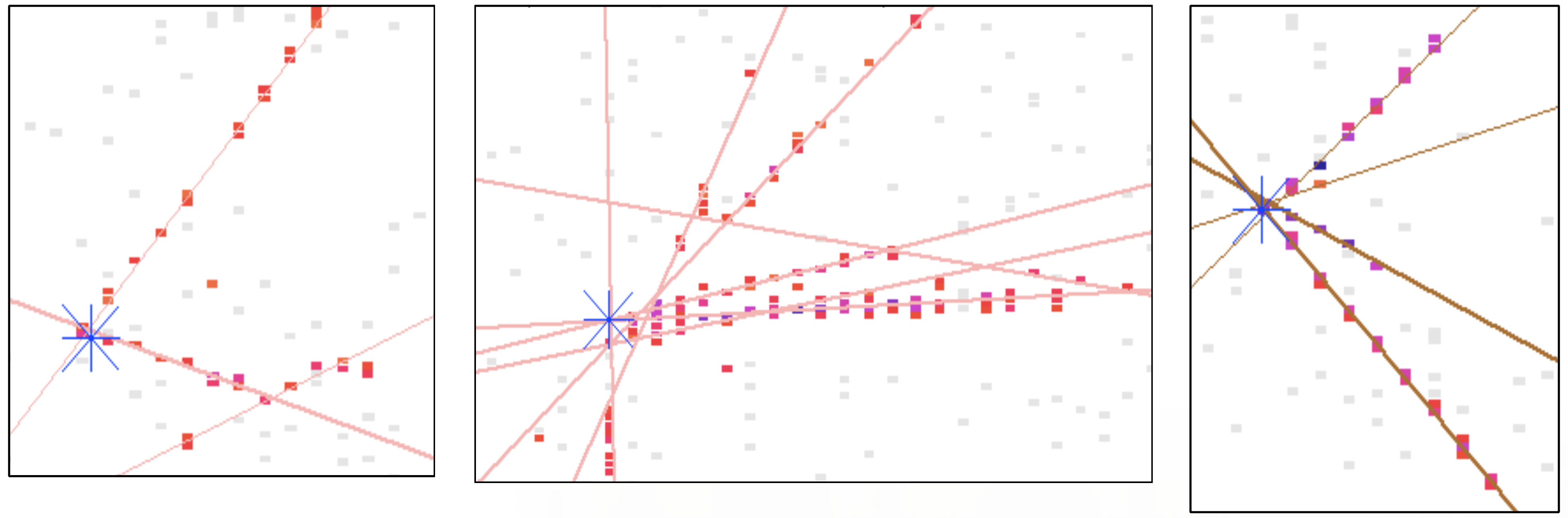


Isolating neutrino interactions

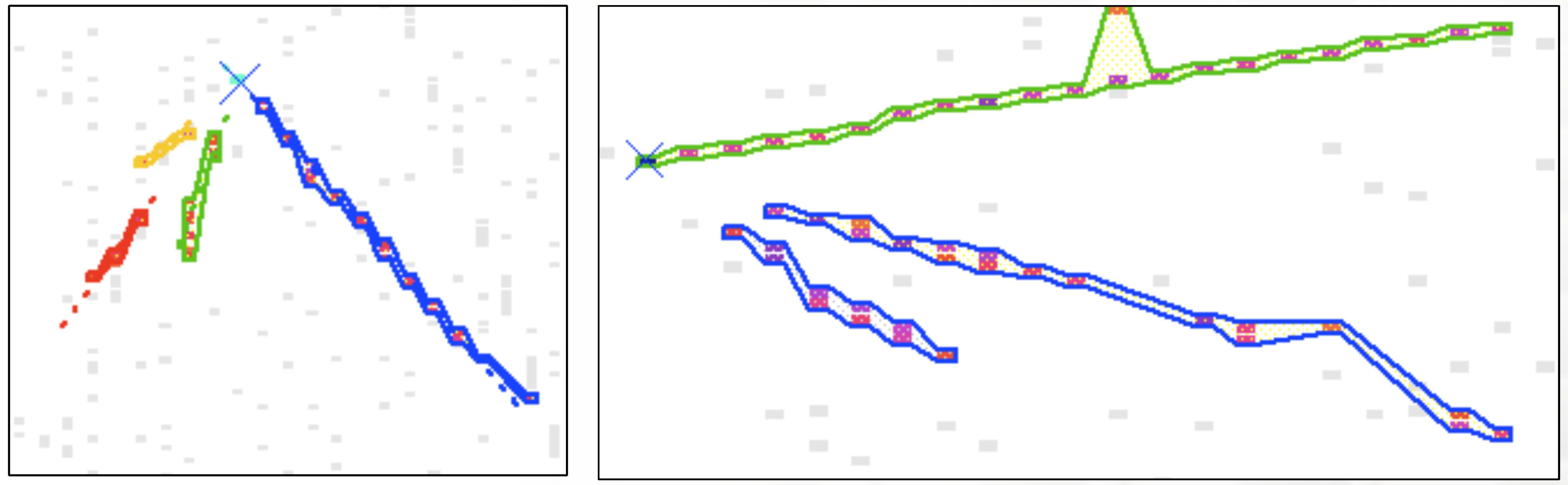
The first step in our reconstruction is dividing an **event** (550 μs of data) into slices (groups of hits with some time and space coincidence)



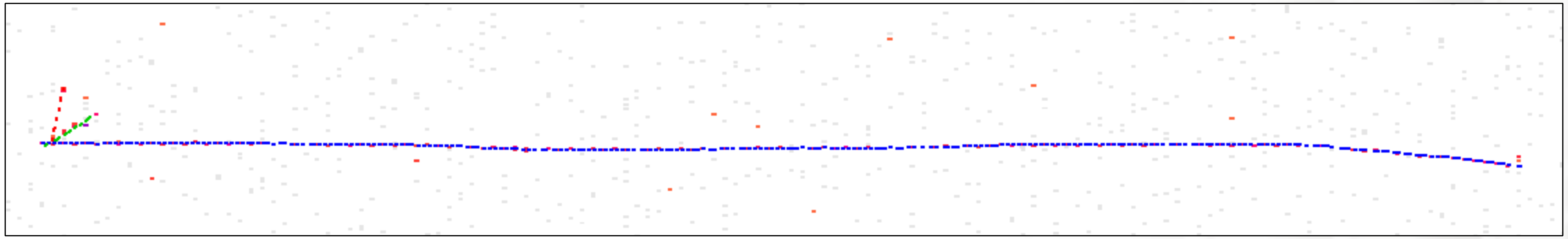
Vertexing: use lines of energy deposition formed with hough transforms to find intersections



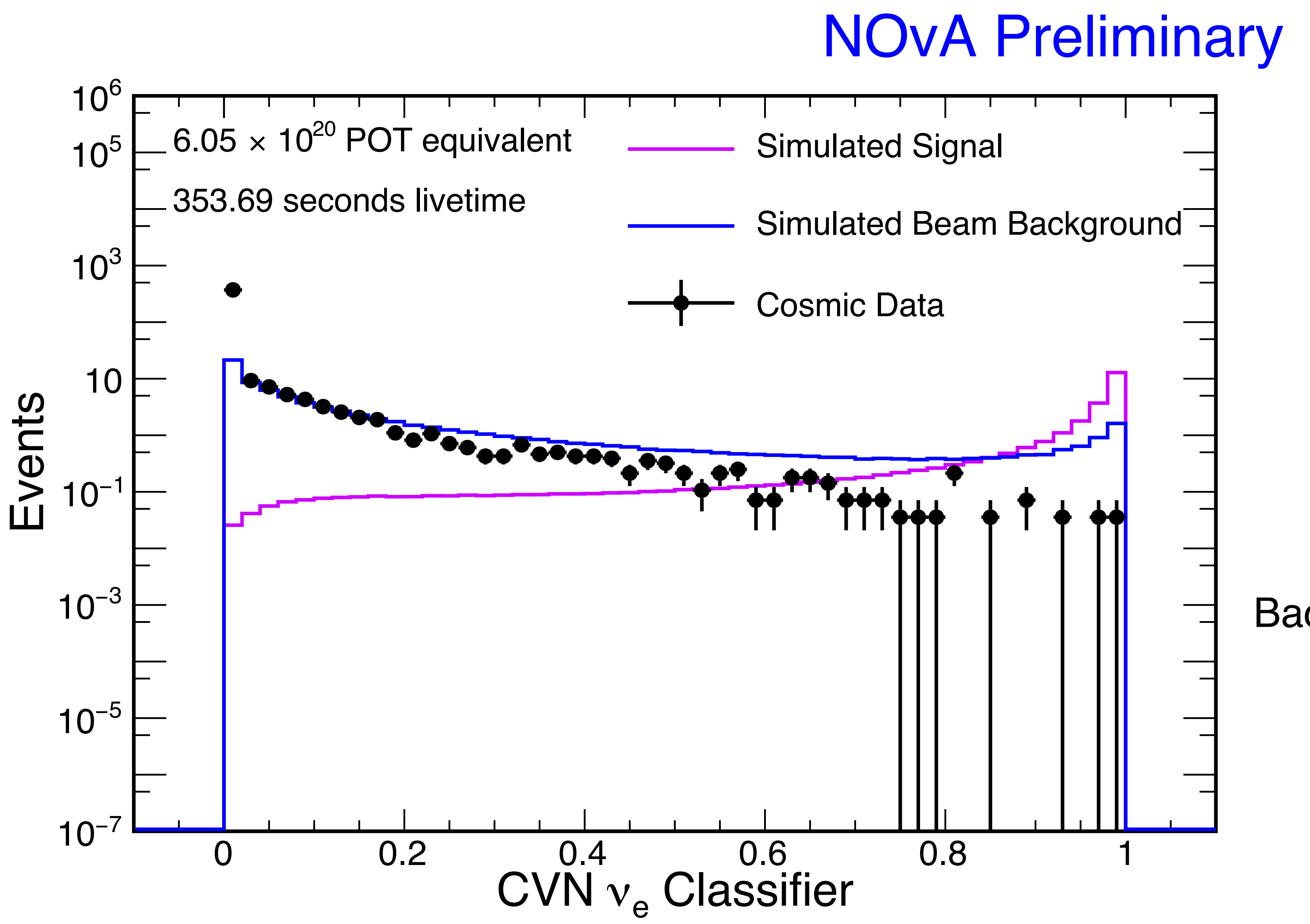
Clustering: find clusters in angular space around the vertex and merge views via topology and prong dE/dx

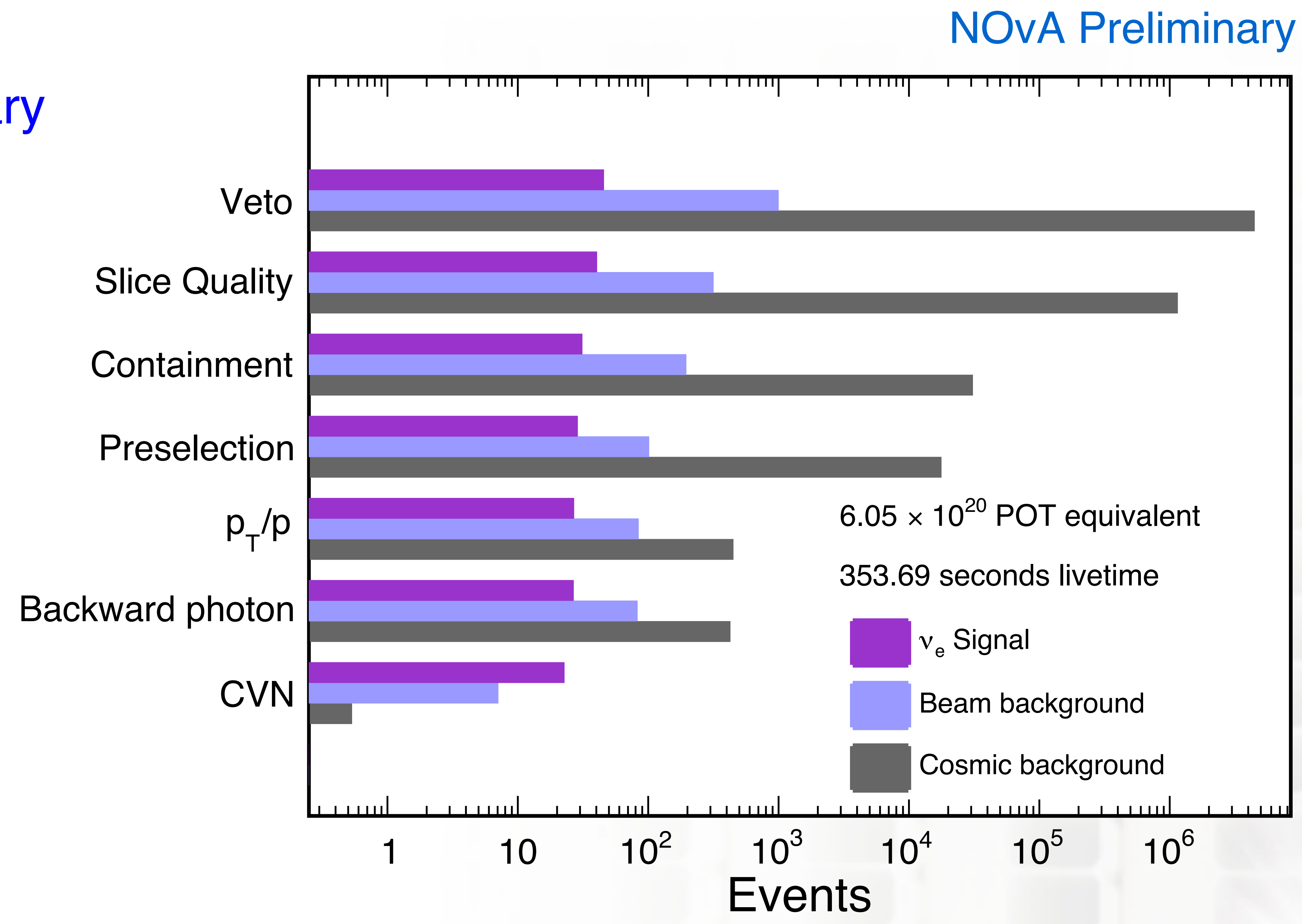


Tracking: Trace particle trajectories using a kalman filter, example below

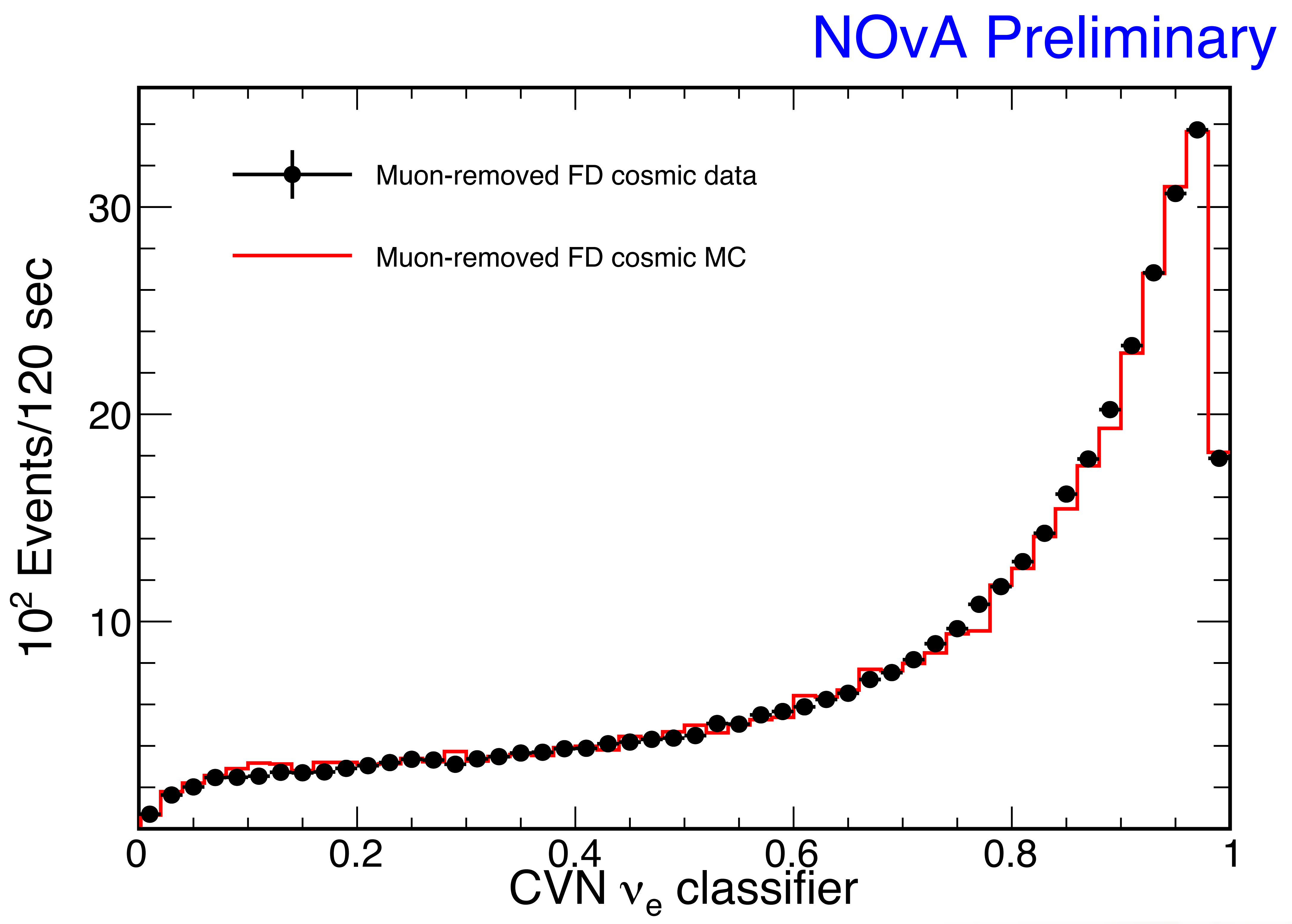


Performance on Cosmic Background

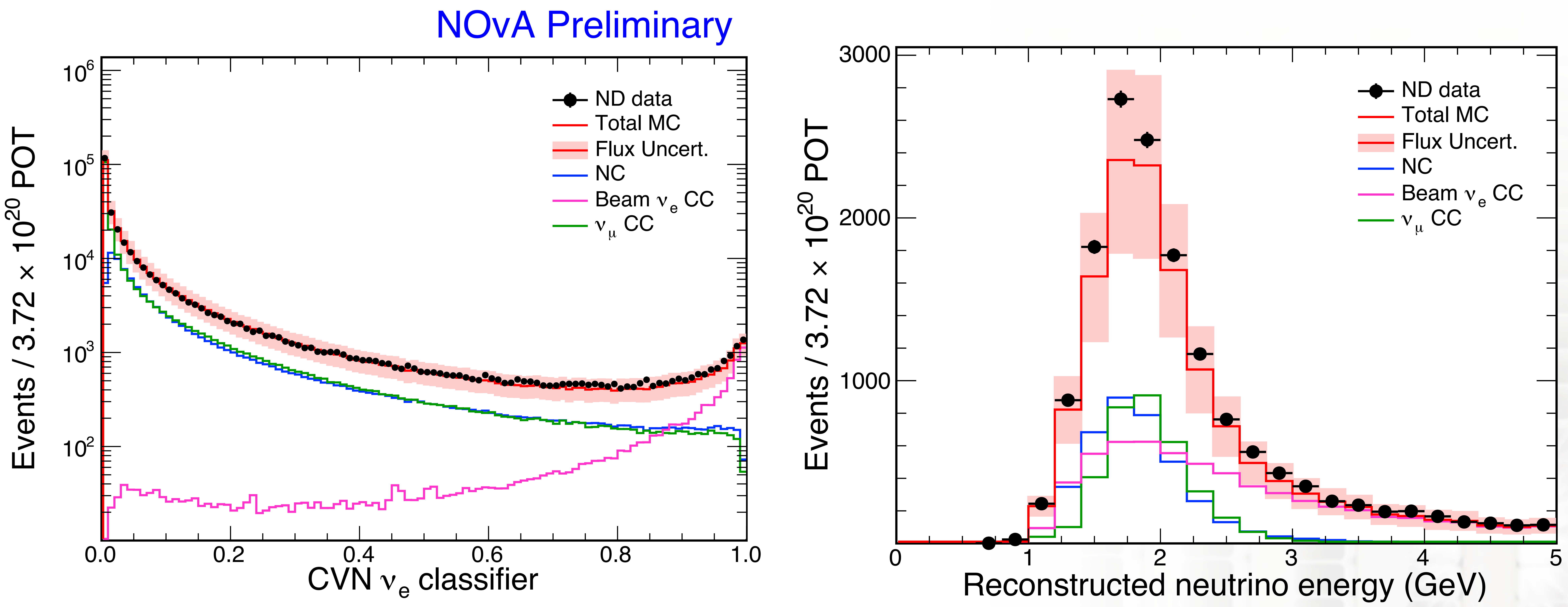




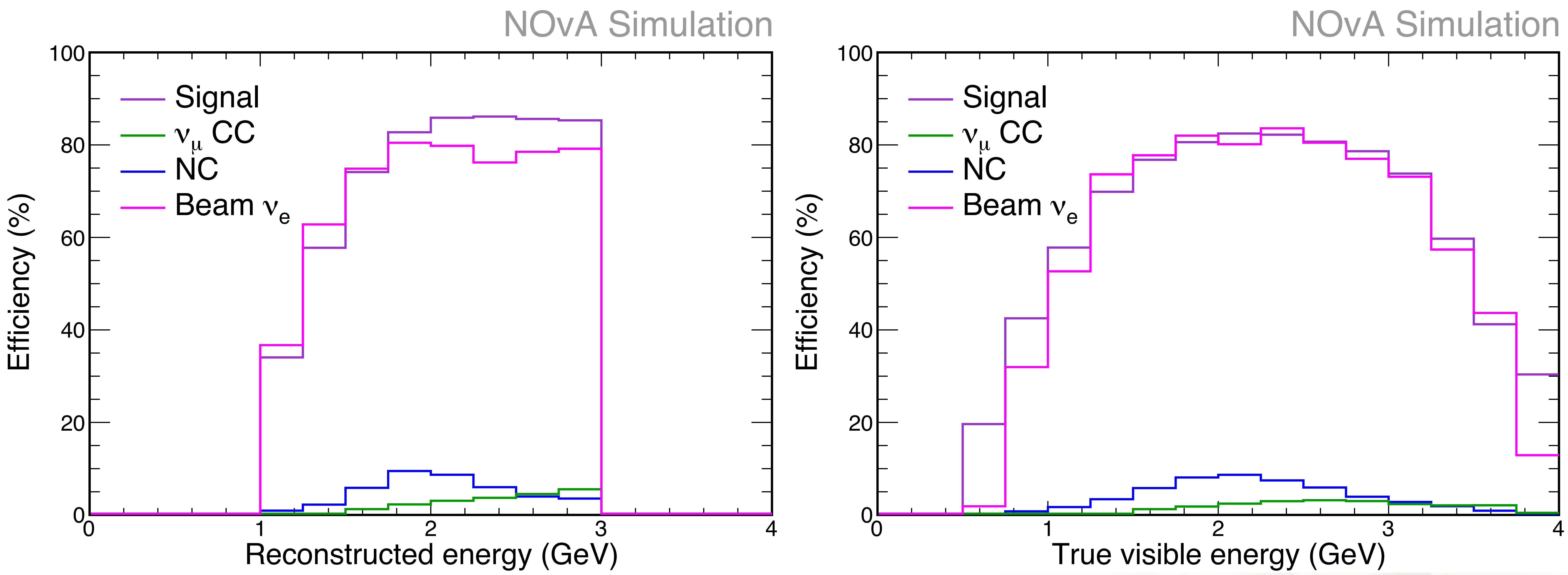
Data Driven Tests - MRBrem



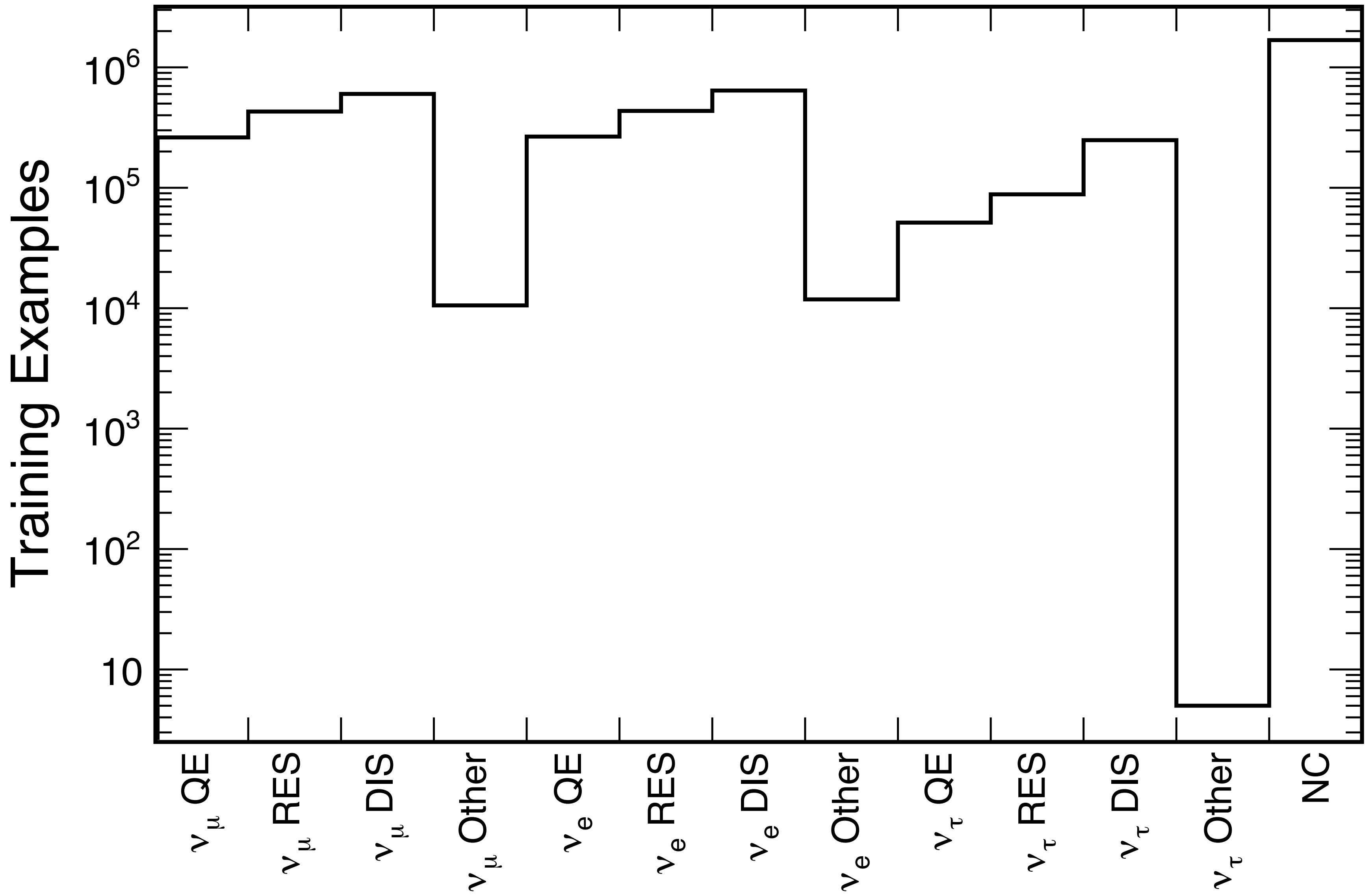
Performance on NearDet Data

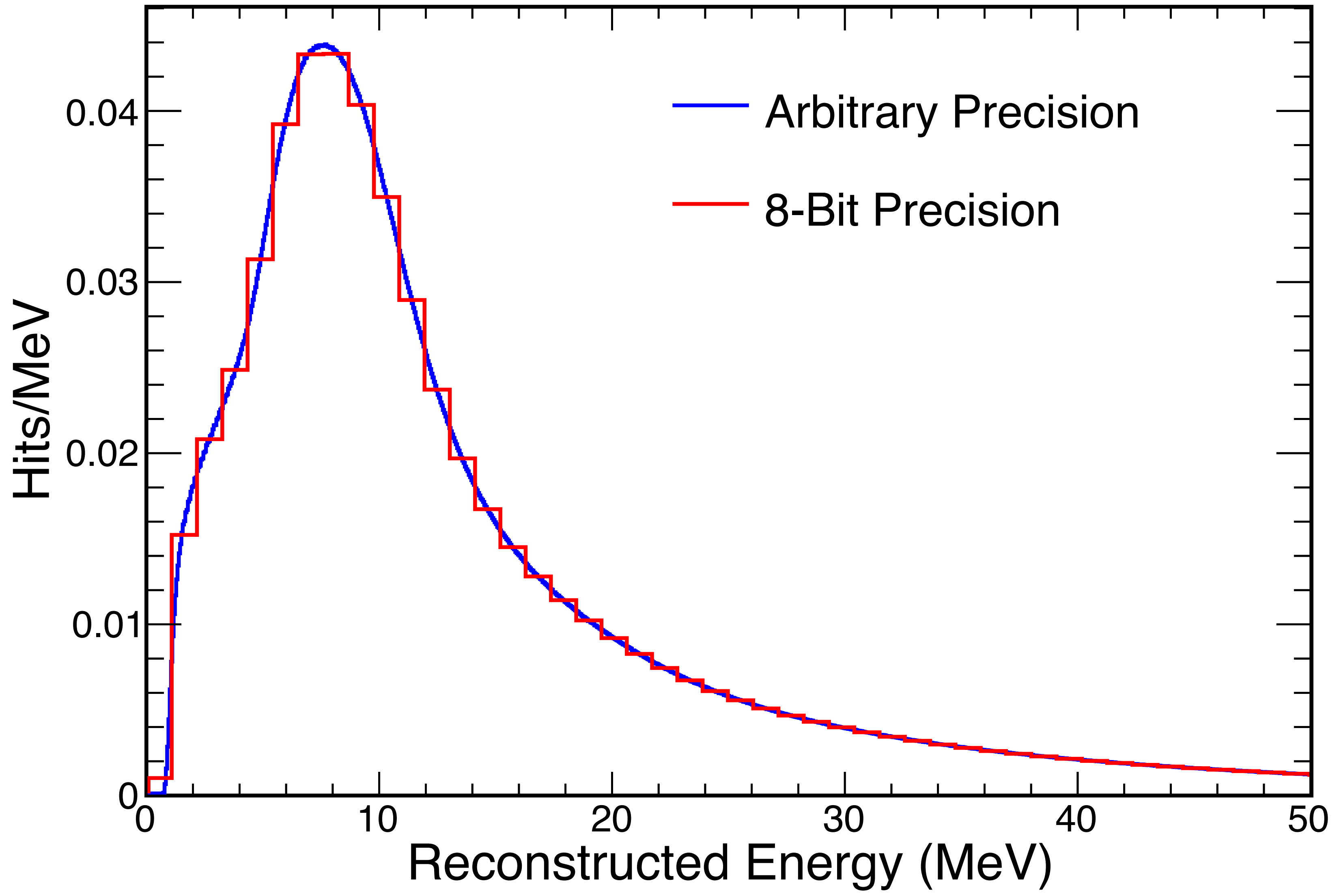


CVN MC Efficiency



CVN Classifier





4.7 million, minimally preselected simulated events, pushed into LevelDB databases: 80% for training and 20% for testing.

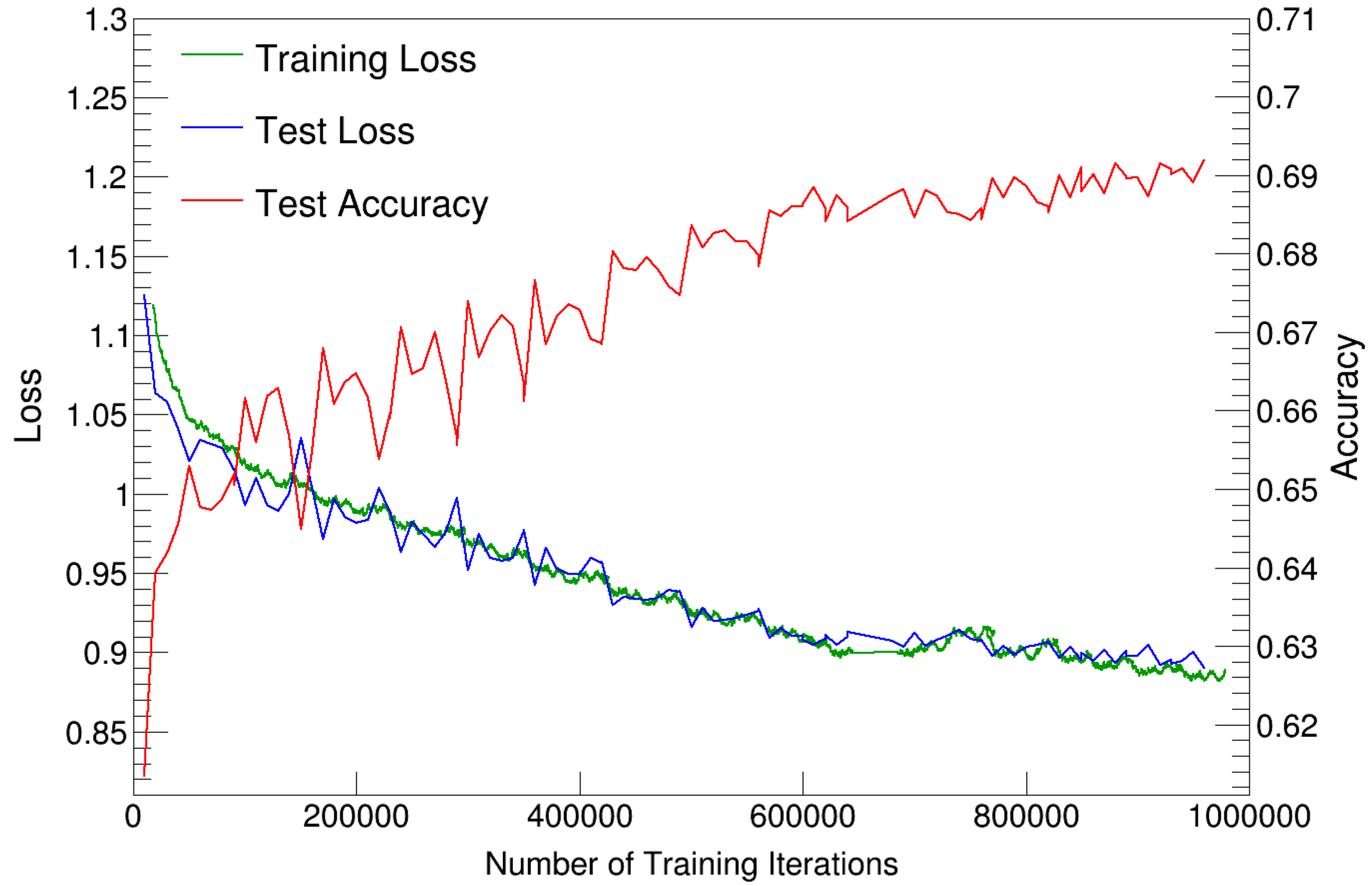
Rescale calibrated energy depositions to go from 0 to 255 and truncate to chars for dramatically reduced file size at no loss of information

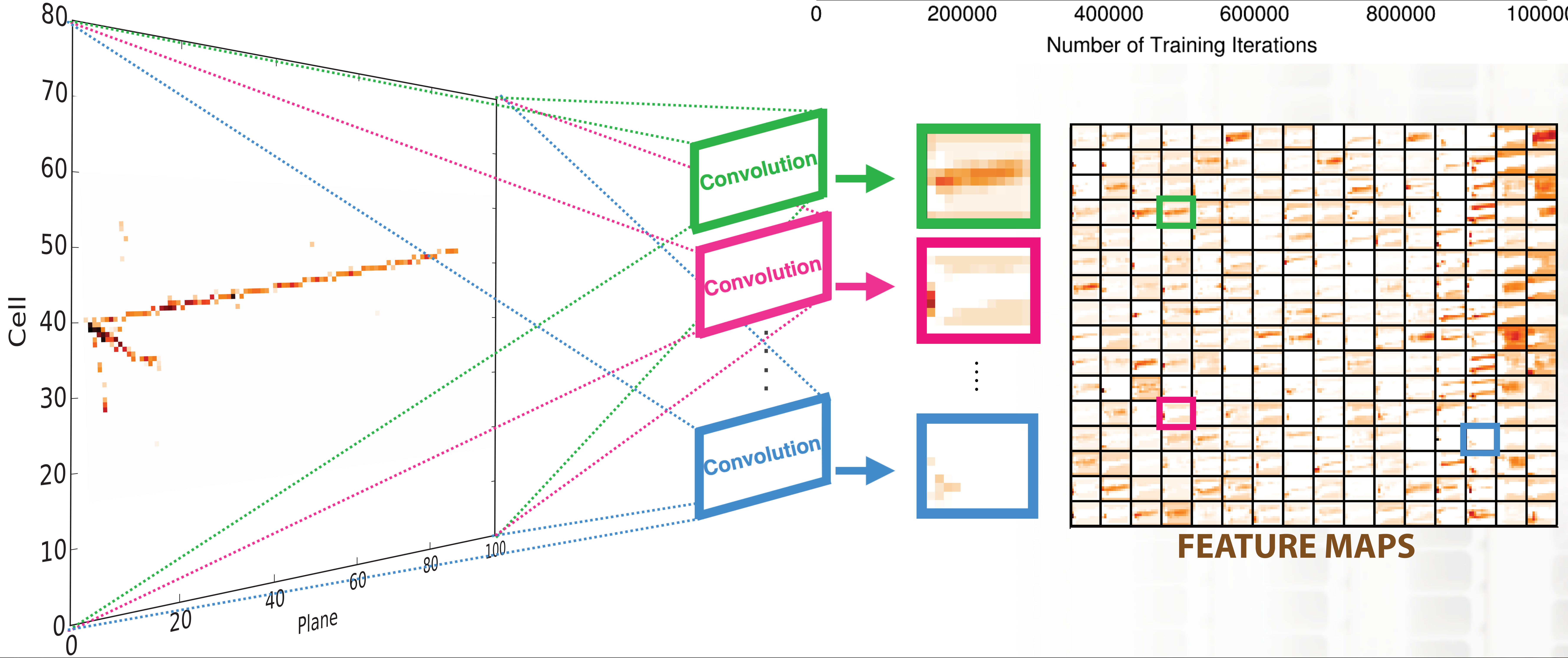
Fine tuned with 5 million cosmic data events taken from an out of beam time minimal bias trigger.

The architecture attempts to categorize events as $\{\nu_\mu, \nu_e, \nu_\tau\} \times$ {QE,RES,DIS}, NC, or Cosmogenic.

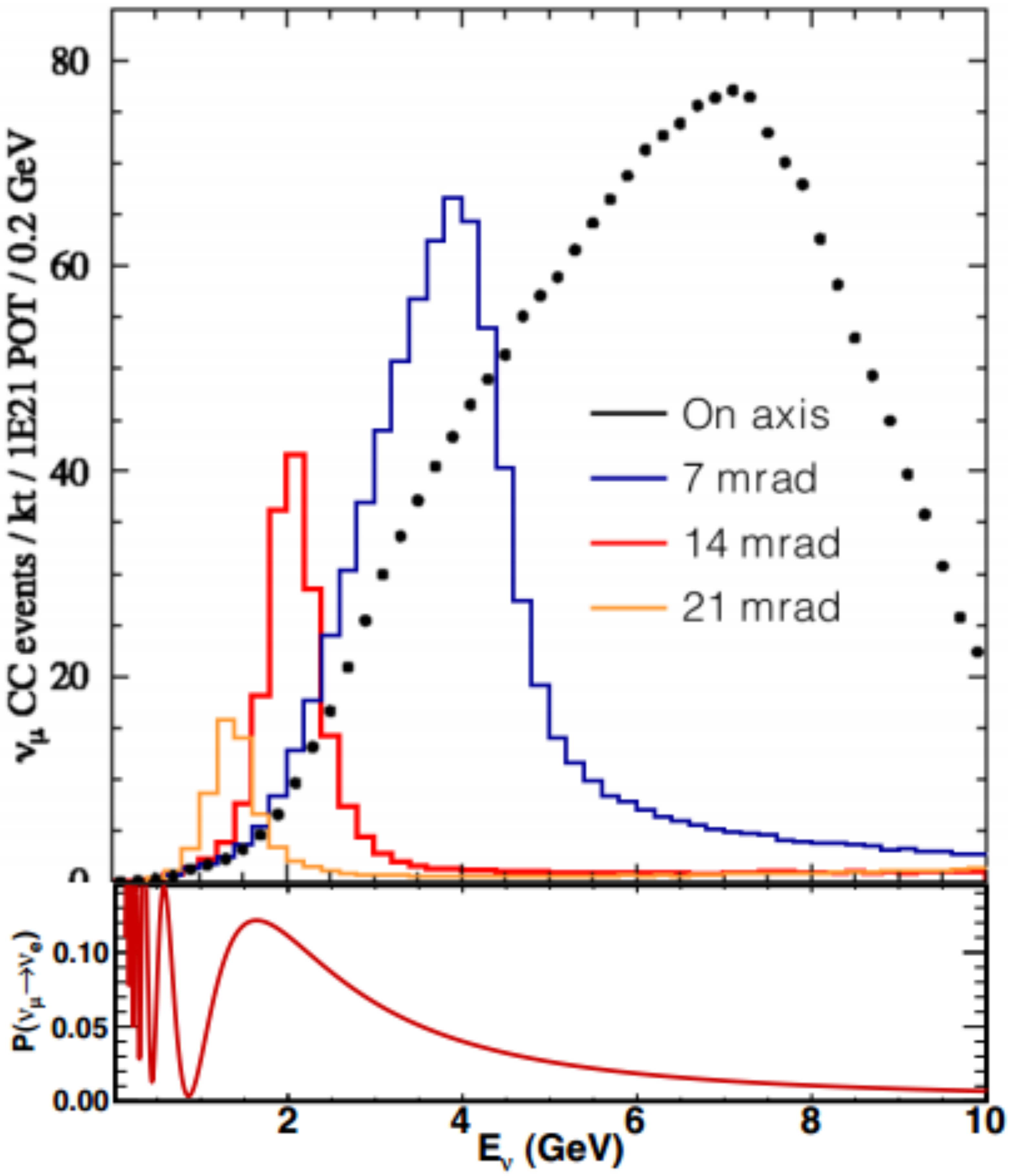
CVN Performance

- Trained on 4.7 million simulated events of all neutrino interaction types plus cosmic rays.
- training sample has minimal preselection

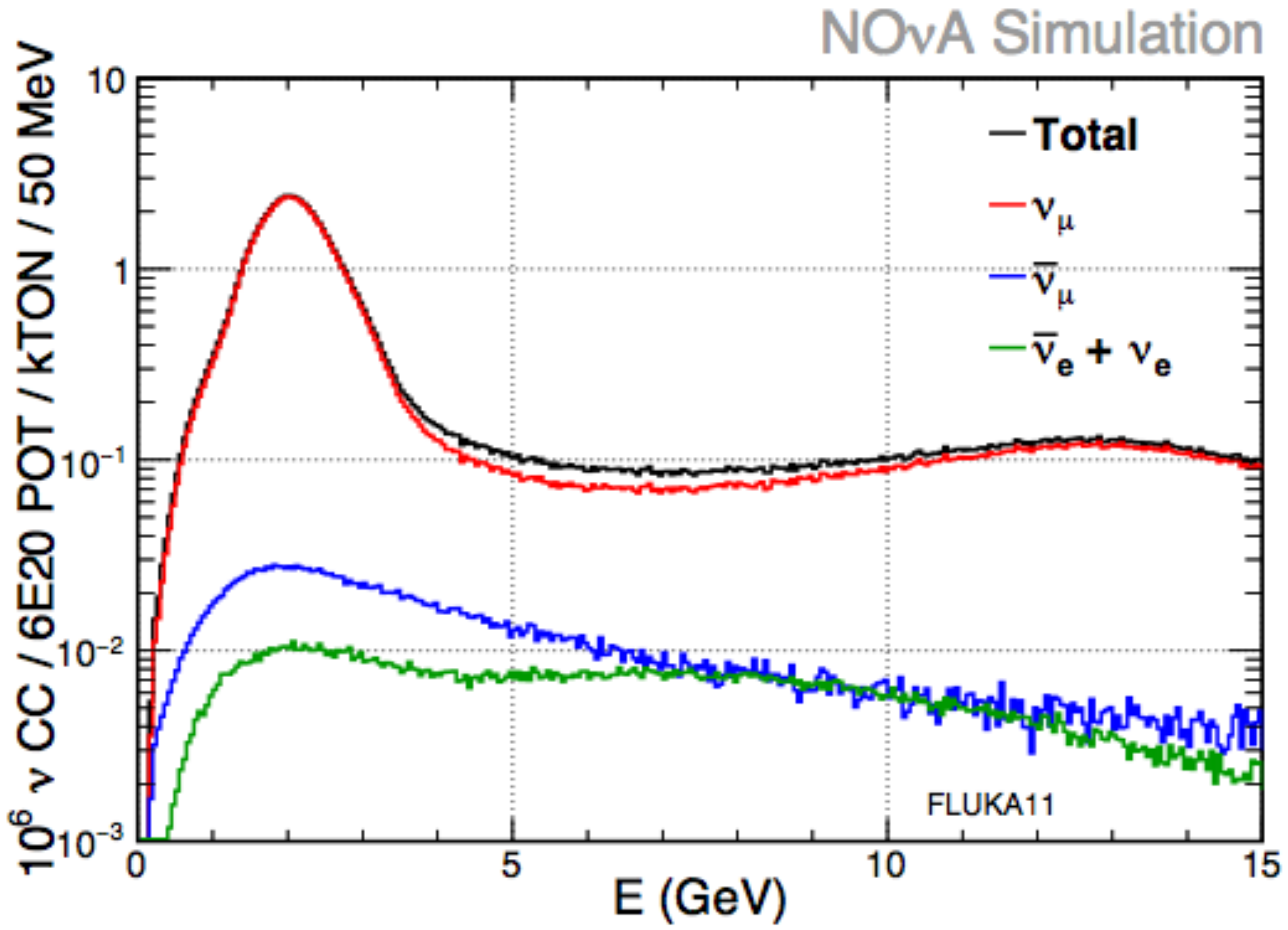




NuMI Beam

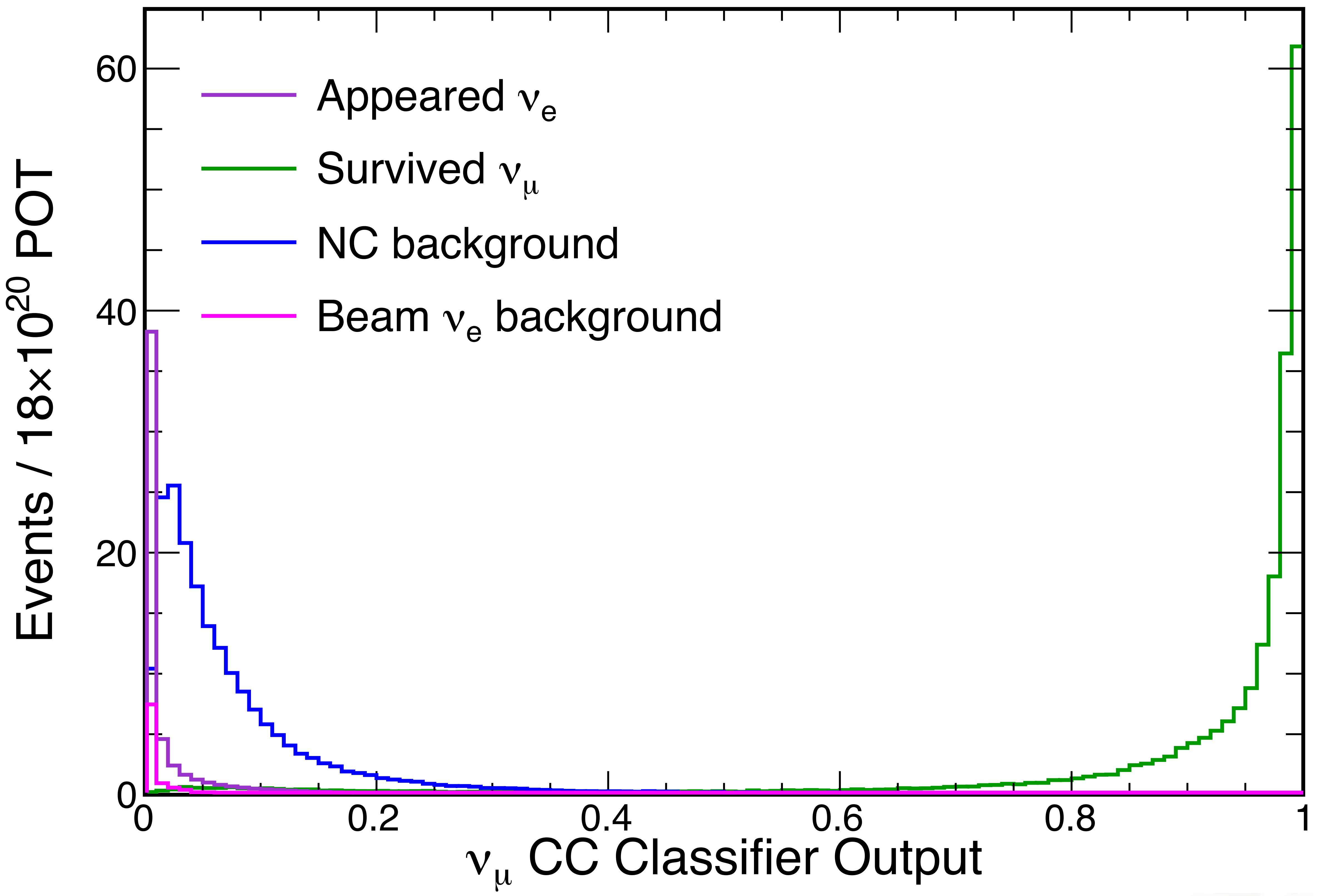


80
60
40
20
0
ν_μ CC events / kt / 1E21 POT / 0.2 GeV
On axis
7 mrad
14 mrad
21 mrad
$P(\nu_\mu \to \nu_e)$
0.10
0.05
0.00
2
4
6
8
10
E_ν (GeV)

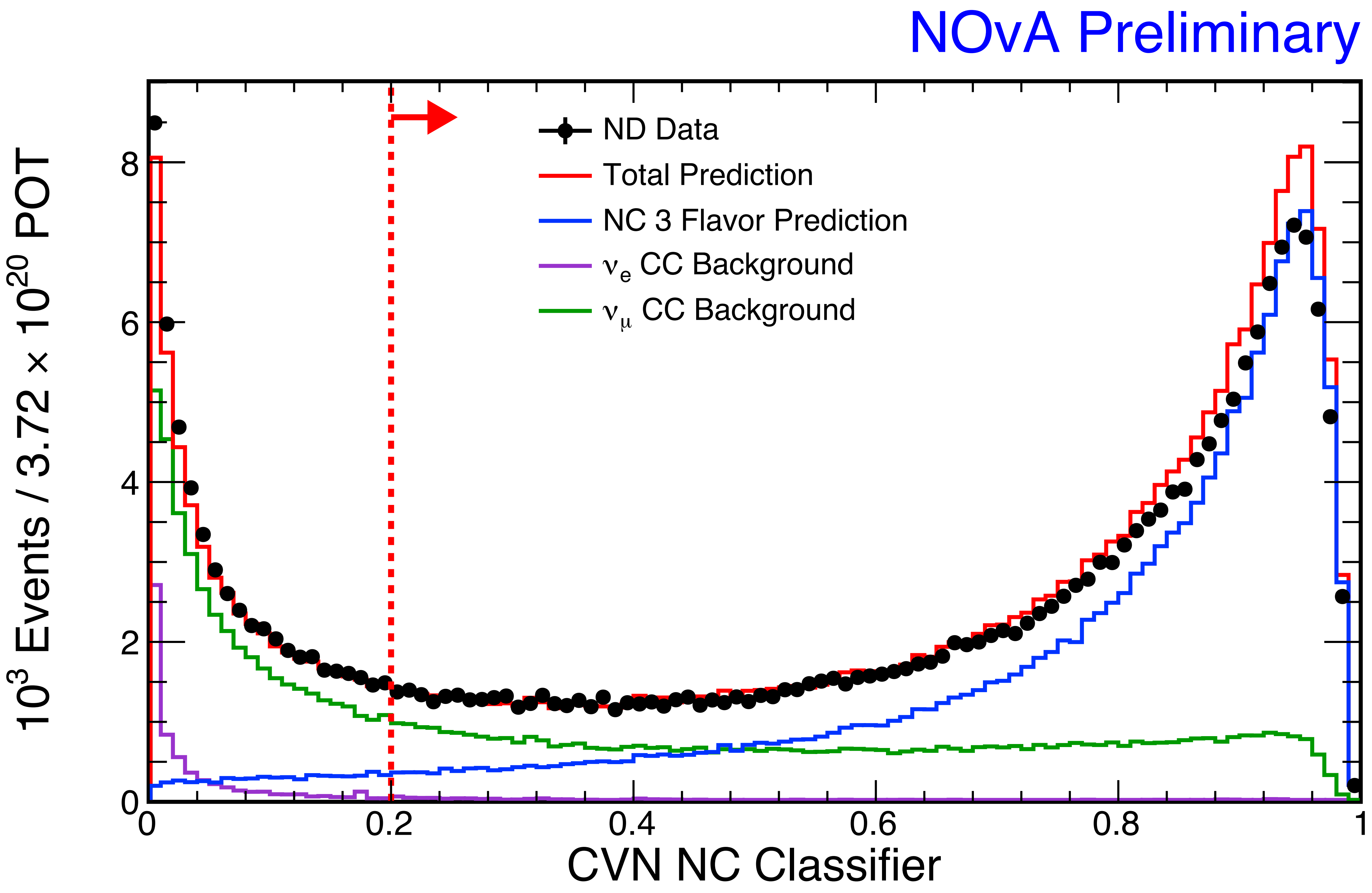


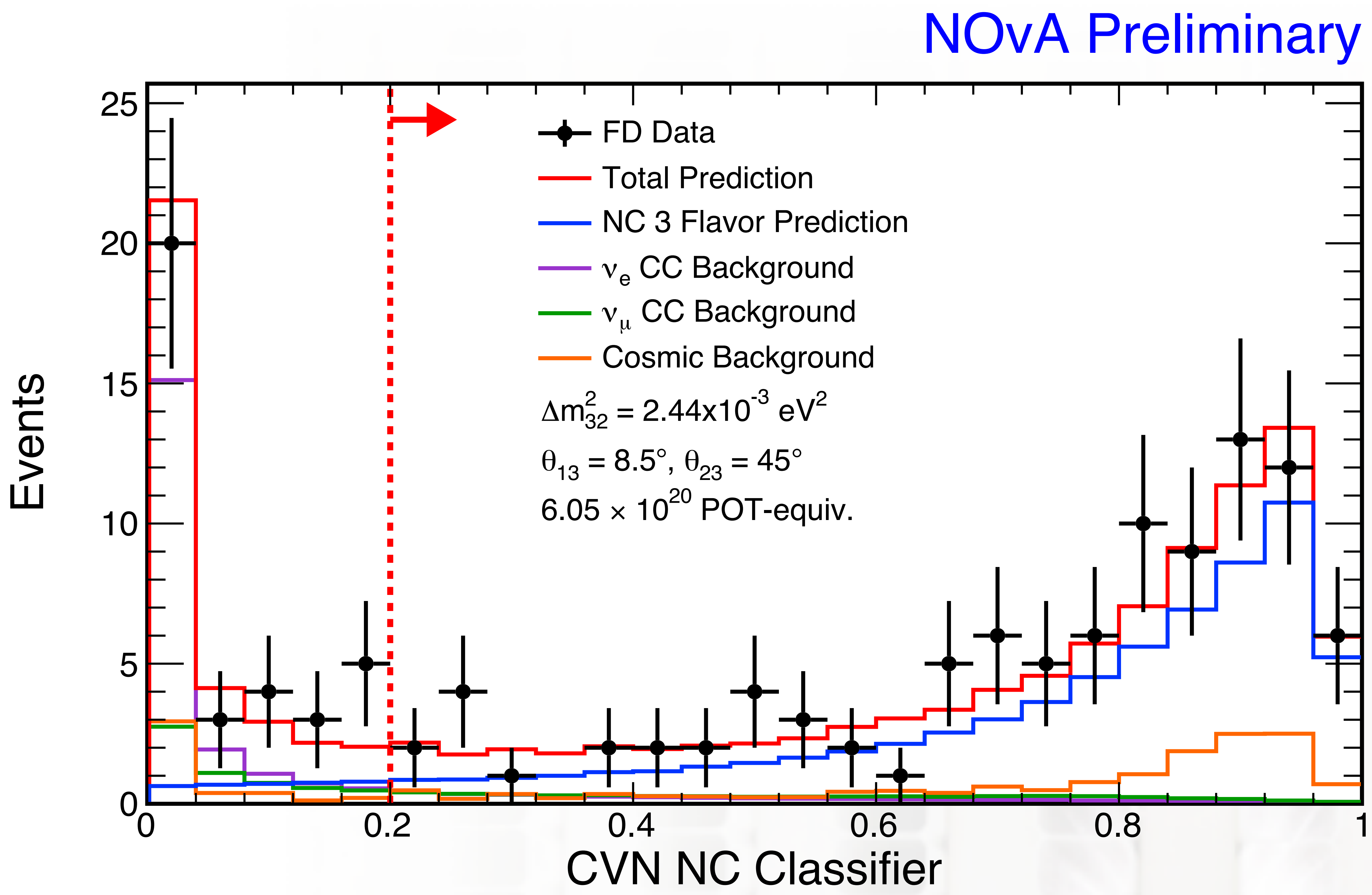
NOvA Simulation
Total
ν_μ
$\bar{\nu}_\mu$
$\bar{\nu}_e + \nu_e$
FLUKA11
10^6 ν CC / 6E20 POT / kTON / 50 MeV
10
1
10^{-1}
10^{-2}
10^{-3}
0
5
10
15
E (GeV)

Muon Neutrino Analysis



Neutral Current Neutrino Analysis

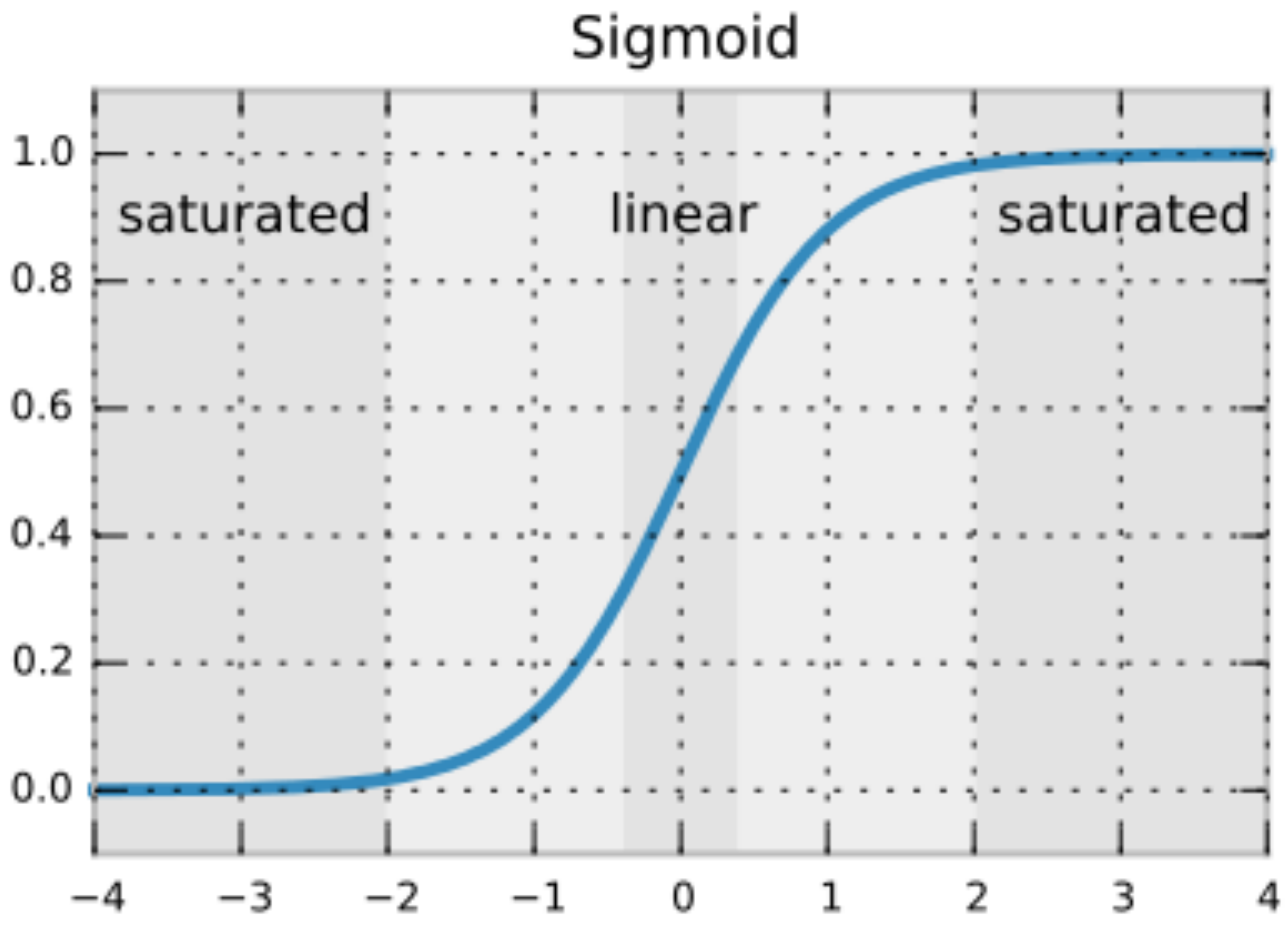


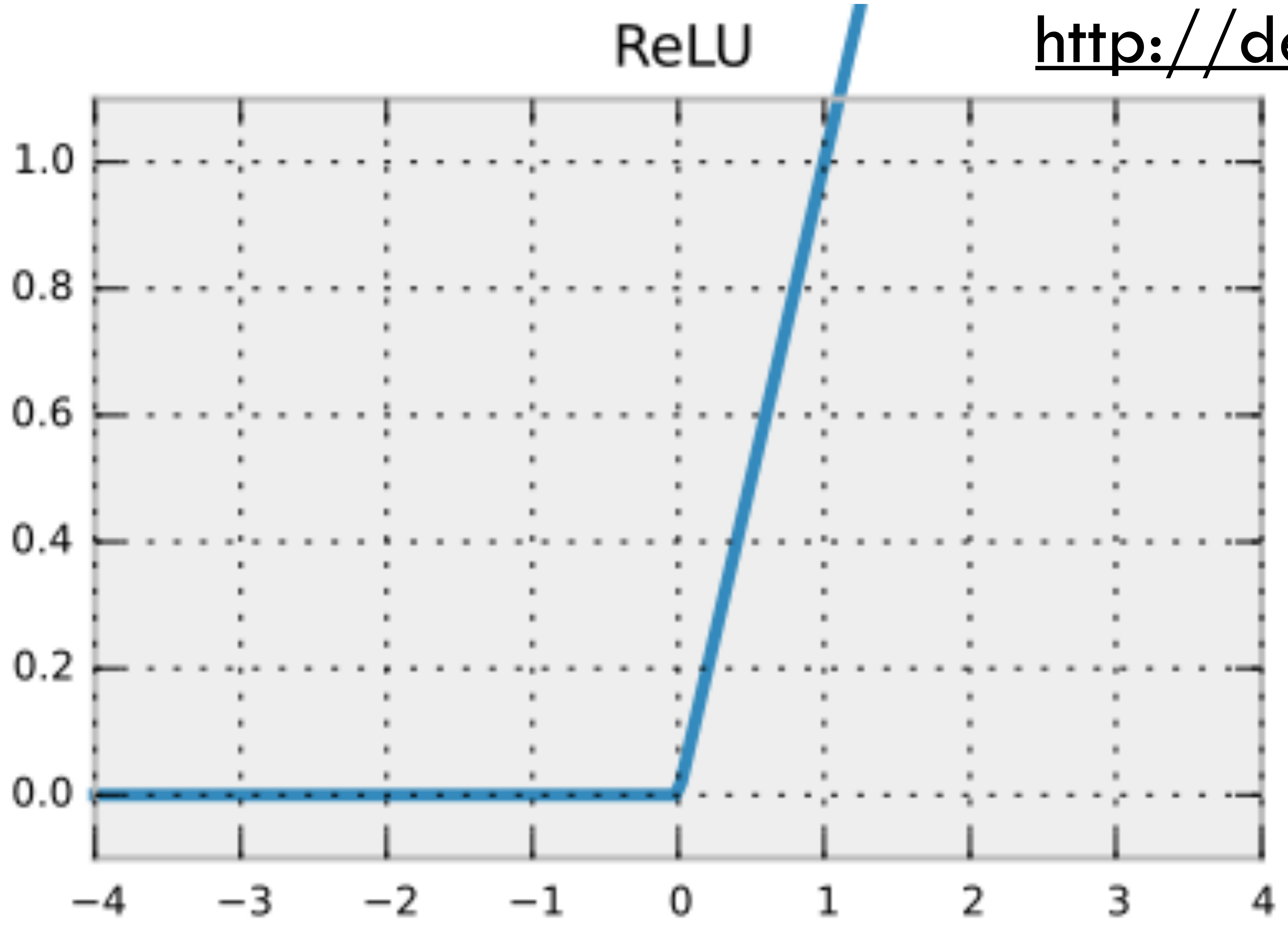


Non Saturating Functions

More effective back propagation due to better weight initialization and saturation functions:

http://deepdish.io/





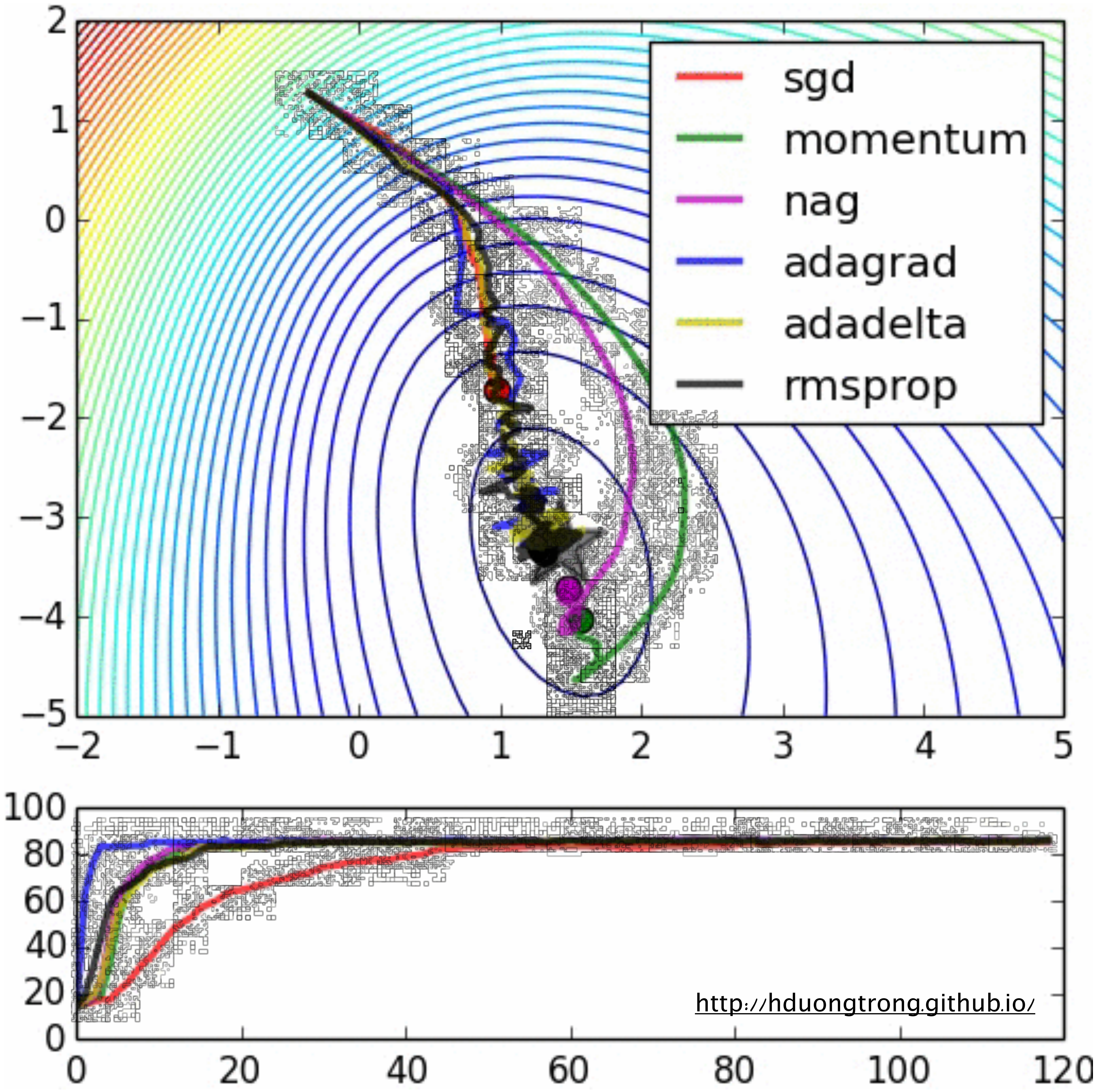
$$\frac{\delta\sigma(x)}{\delta x} = \sigma(x)\left(1 - \sigma(x)\right)$$

$$\frac{ReLU(x)}{\delta x} = \begin{cases} 1 & \text{when } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Sigmoid gradient goes to 0 when x is far from 1. Makes back propagation impossible!

Use ReLU to avoid saturation.

Stochastic Gradient Descent.



In SGD we avoid some of the cost of gradient descent by **evaluating small batches of events one at a time.**

The performance of conventional gradient descent is approximated as the various noisy sub estimates even out, with the stochastic behavior even allowing for jumping out of local minima.