Machine Learning in the NOvA Experiment

Karl Warburton – Iowa State University

For the NOvA Collaboration

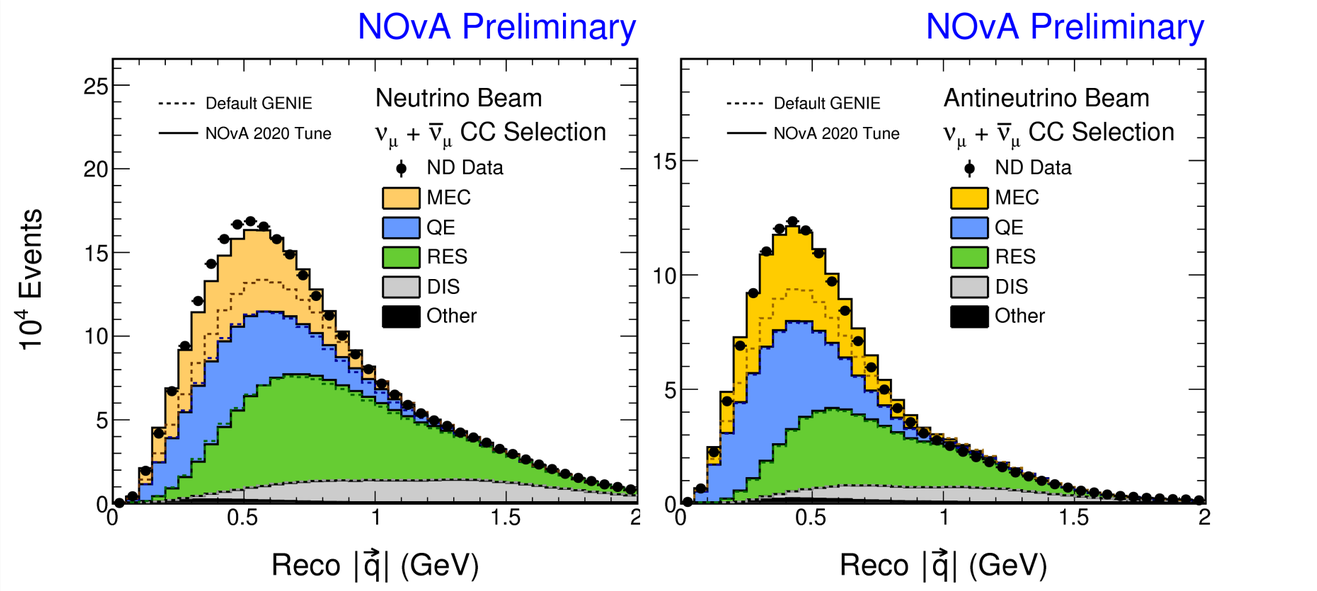
Neutrino Physics Machine Learning Workshop (NPML)

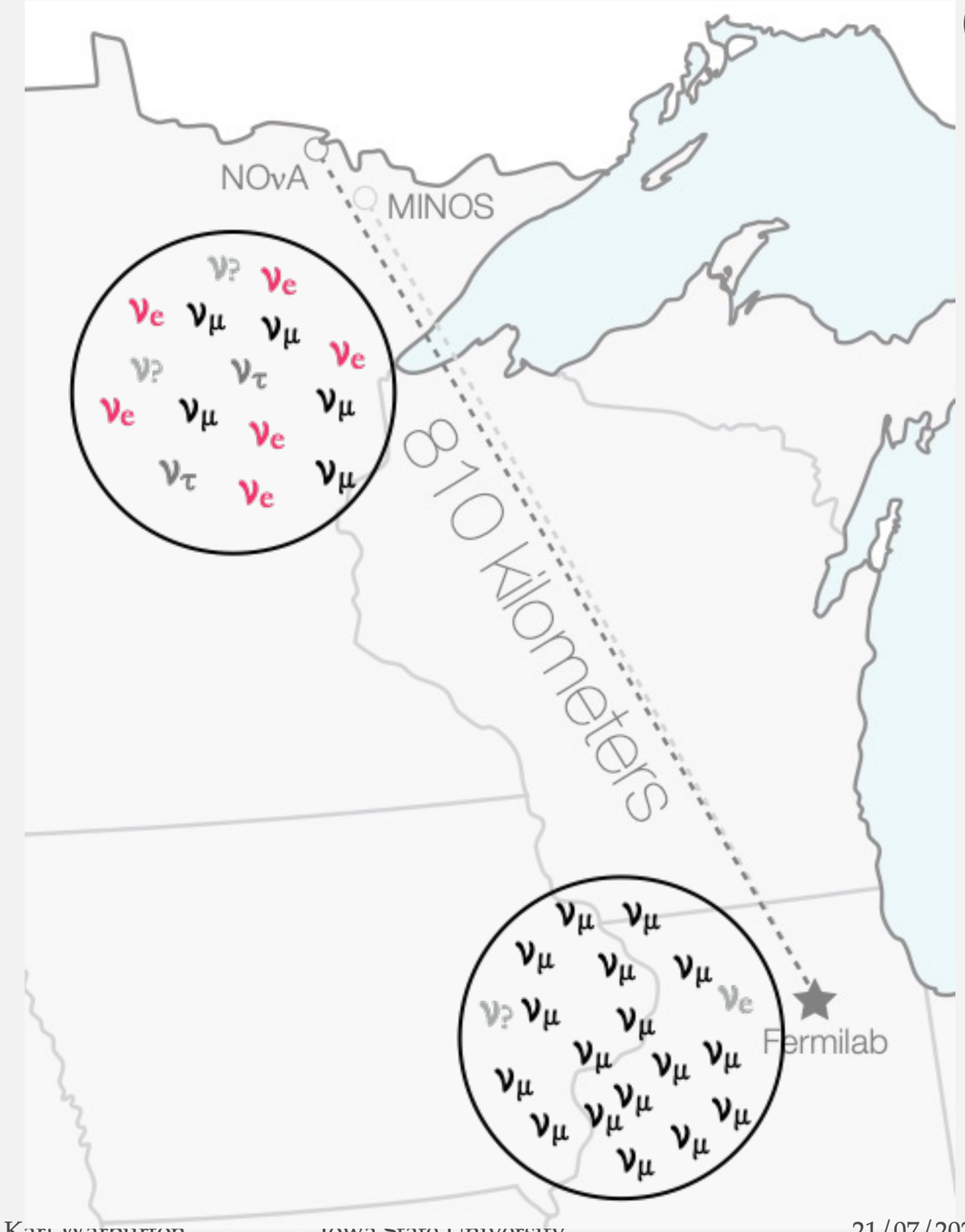
21st July 2020

The NOvA Experiment

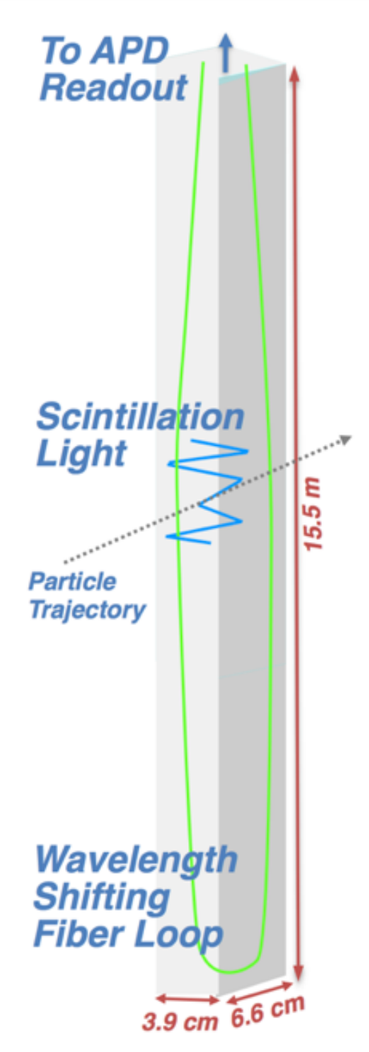
- NOvA is a long-baseline neutrino experiment, which uses the 700 kW NuMI beam of neutrinos from Fermilab.

- Physics goals include neutrino oscillations, neutrino cross-section measurements and exotic physics. Machine learning plays a key role in many of these analyses.

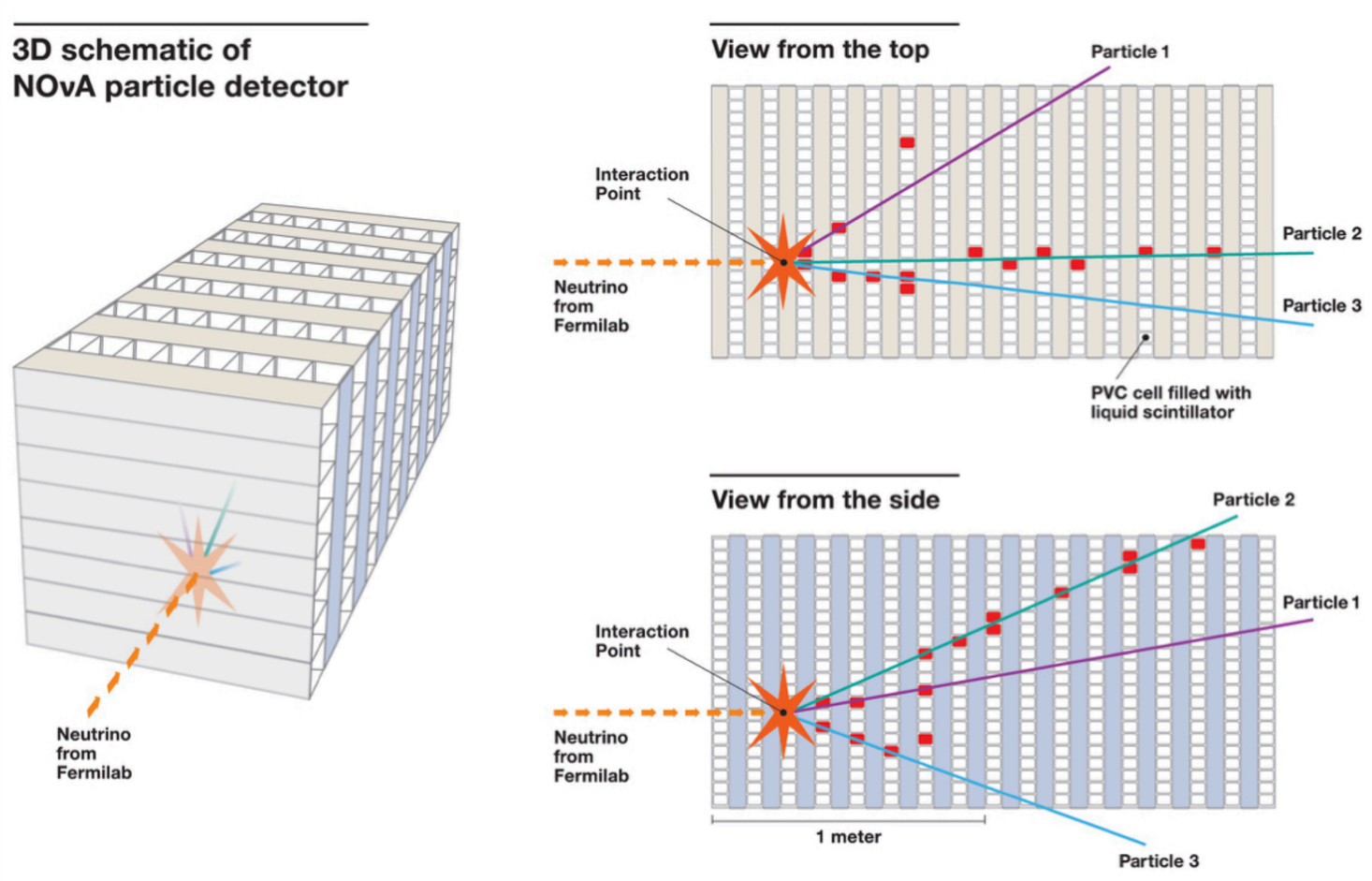


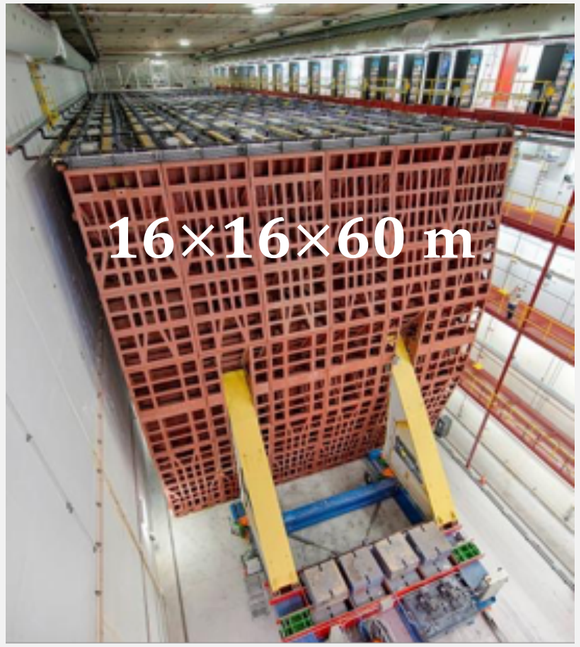


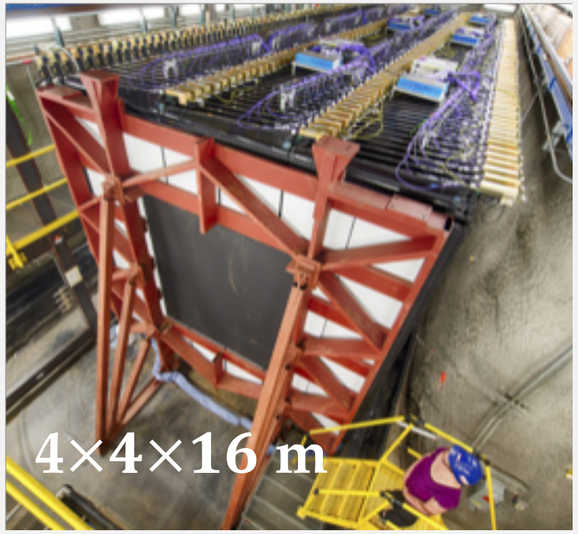
The NOvA Detectors

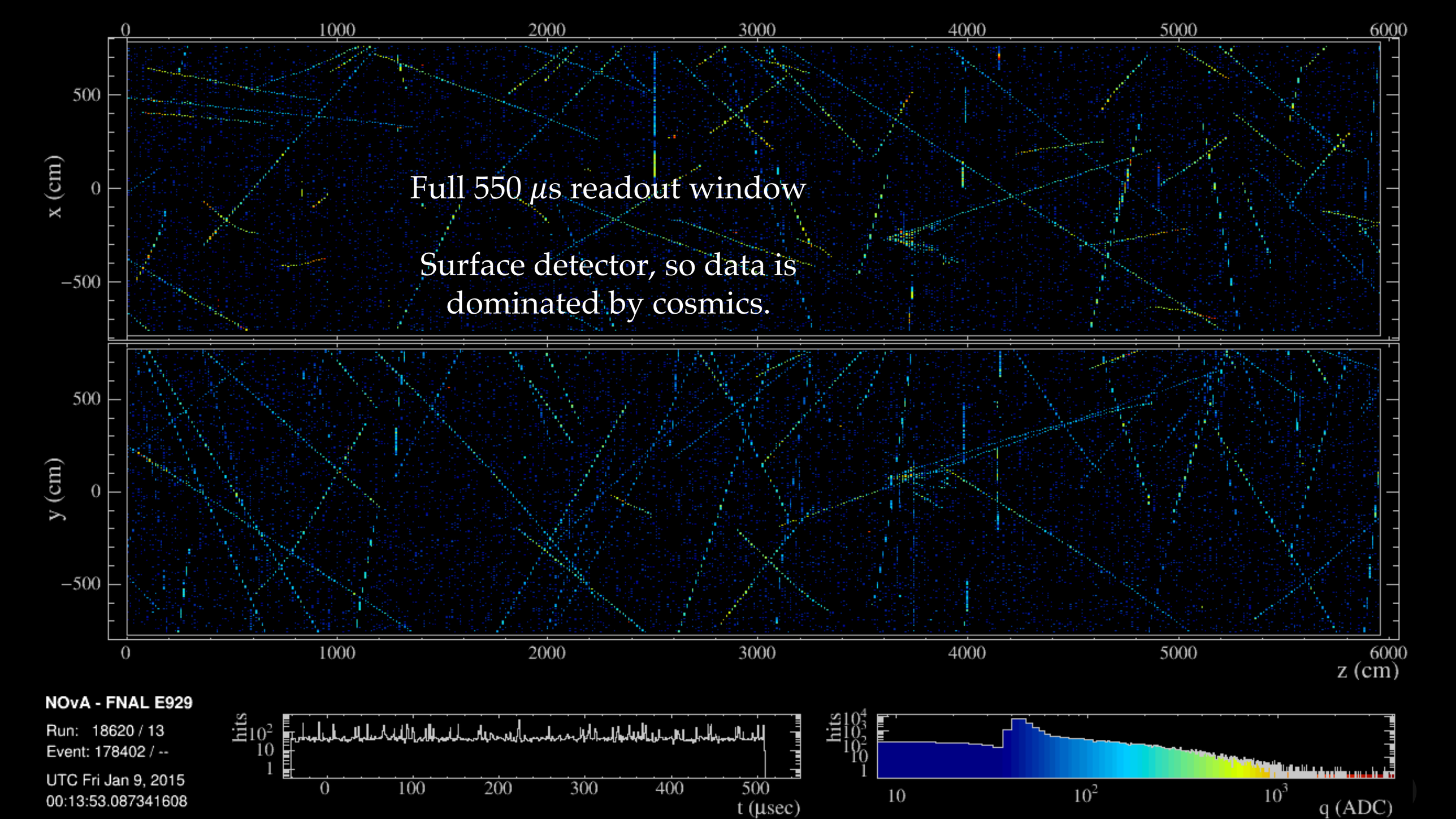


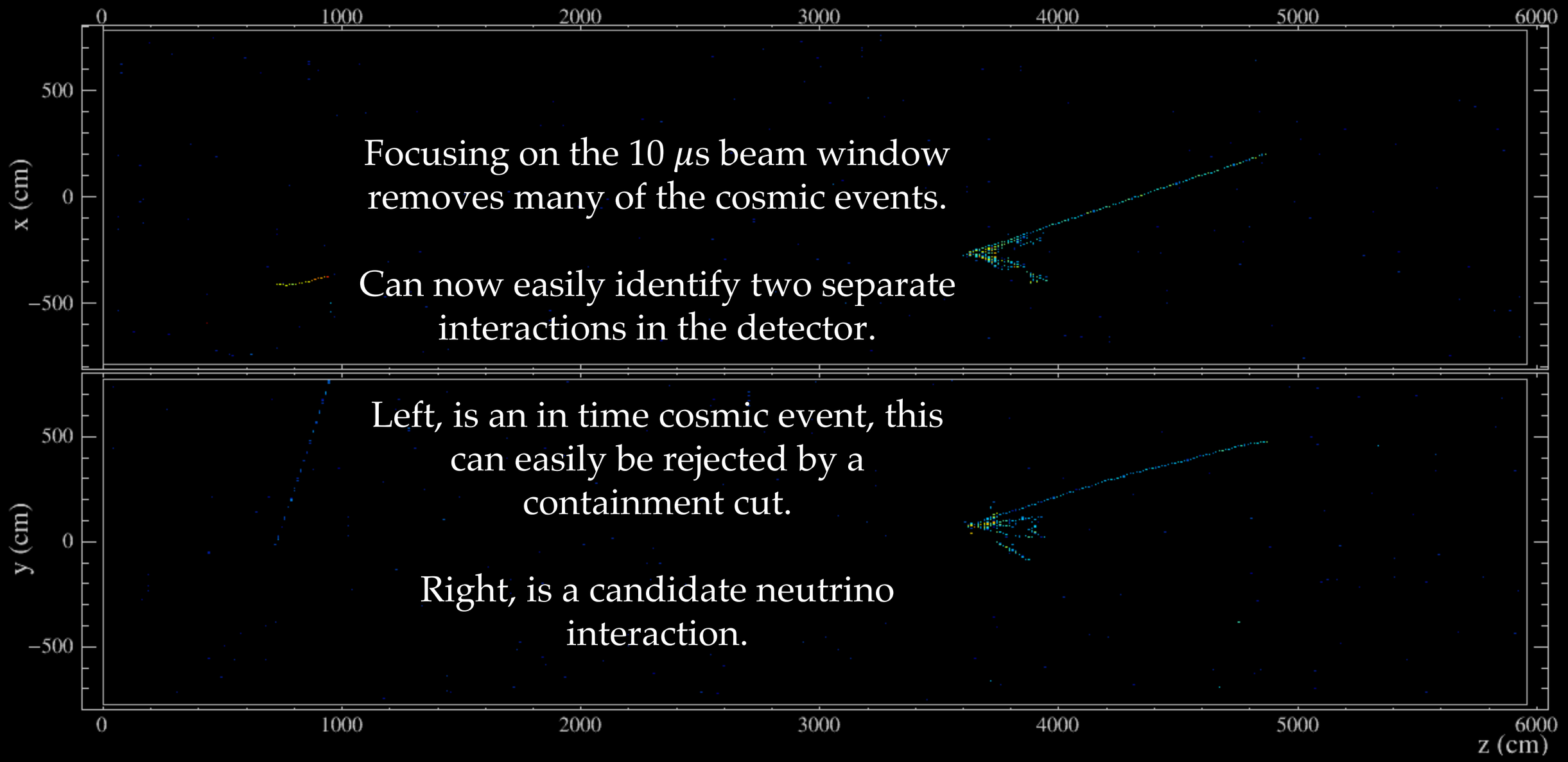
- The detectors are functionally equivalent sampling calorimeters.
- Planes are arranged in alternating directions for 3D reconstruction.
- Low Z material, with a radiation length of roughly 40 cm.








Full 550 μs readout window
Surface detector, so data is dominated by cosmics.
x (cm)
y (cm)
z (cm)
NOvA - FNAL E929
Run: 18620 / 13
Event: 178402 / --
UTC Fri Jan 9, 2015
00:13:53.087341608
hits
t (μsec)
q (ADC)



Focusing on the 10 μs beam window removes many of the cosmic events.

Can now easily identify two separate interactions in the detector.

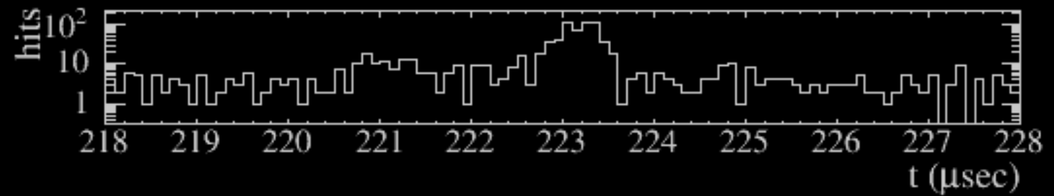
Left, is an in time cosmic event, this can easily be rejected by a containment cut.

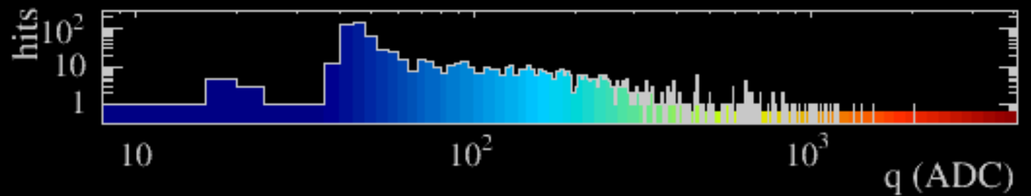
Right, is a candidate neutrino interaction.

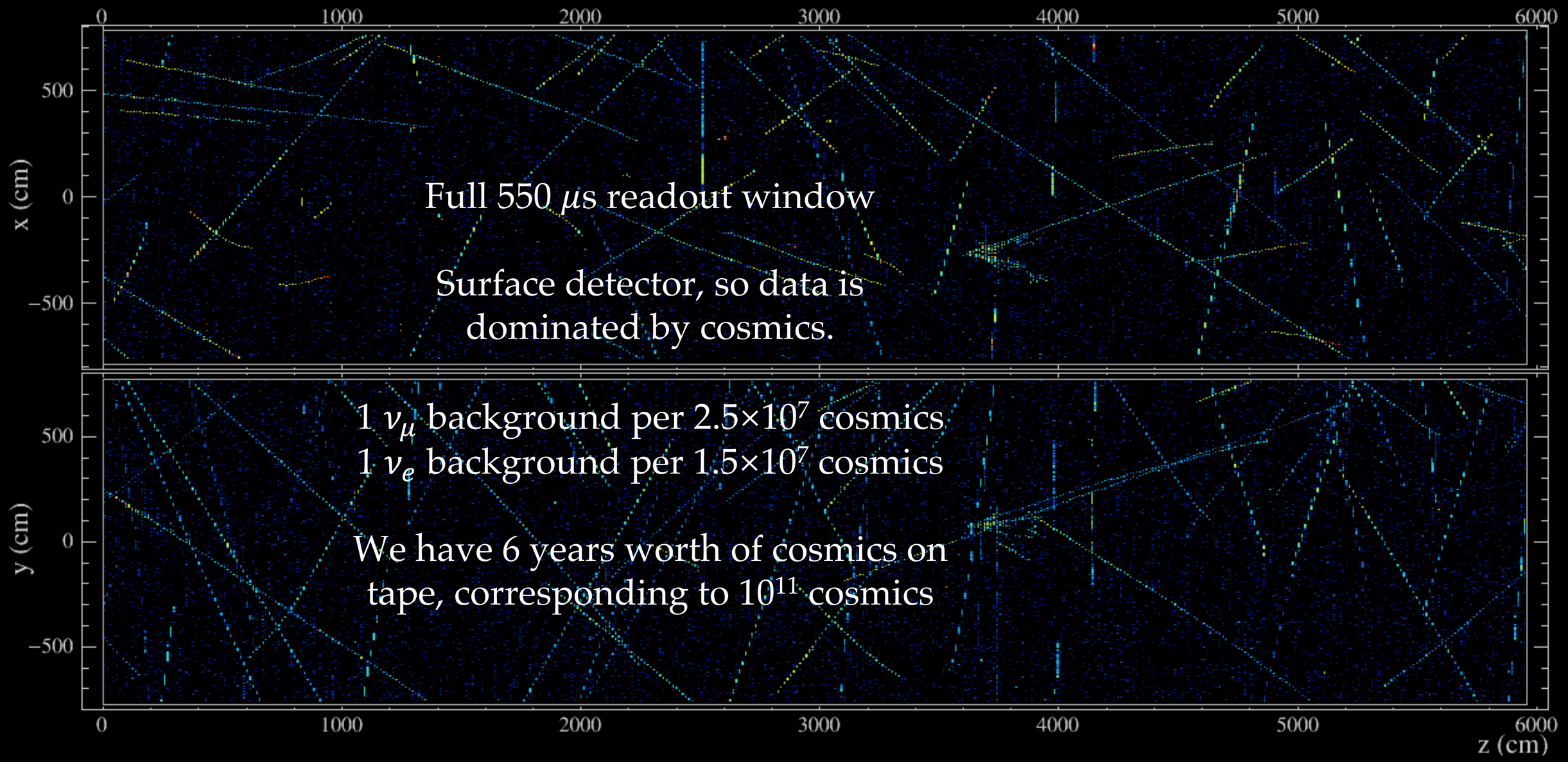
NOvA - FNAL E929

Run: 18620 / 13
Event: 178402 / --

UTC Fri Jan 9, 2015
00:13:53.087341608







Full 550 μs readout window

Surface detector, so data is dominated by cosmics.

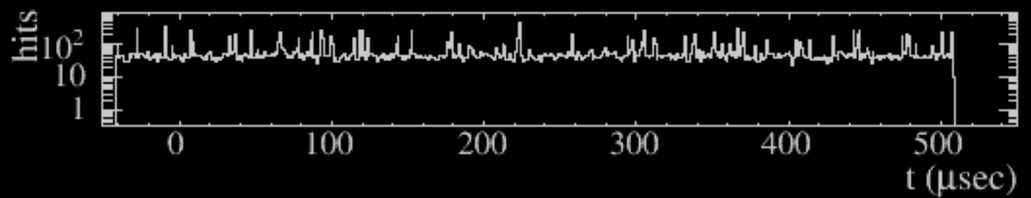
1 ν_μ background per 2.5×10^7 cosmics
1 ν_e background per 1.5×10^7 cosmics

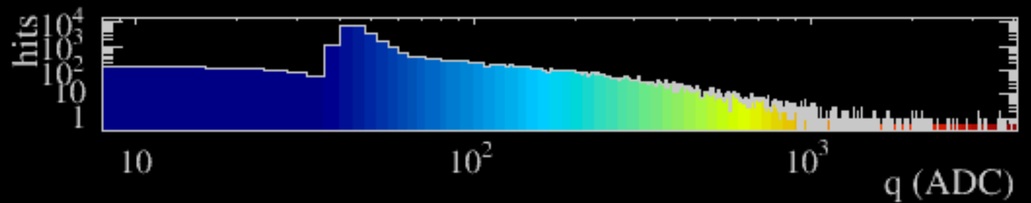
We have 6 years worth of cosmics on tape, corresponding to 10^{11} cosmics

NOvA - FNAL E929

Run: 18620 / 13
Event: 178402 / --

UTC Fri Jan 9, 2015
00:13:53.087341608

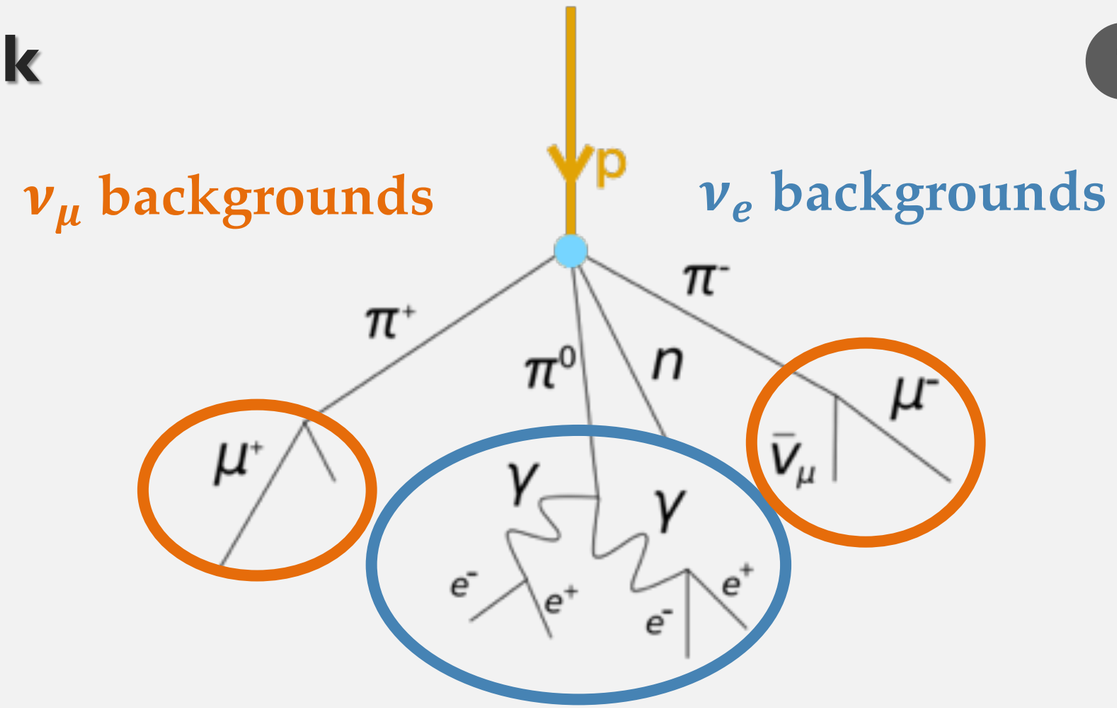




A Cosmic Rejection Neural network

- Fully processing the currently recorded cosmics would take roughly 2 years of continuous running.
  - Run a pre-reconstruction filter to remove the most obvious cosmics with no dependencies on our calibration tunes or clustering algorithms.
  - The network is a "one-pass" network, meaning that it is ran once, and only once.

- ResNet-18 backbone, with a Siamese structure.
  - Trained on 1M $\nu_e, \overline{\nu_e}, \nu_\mu, \overline{\nu_\mu}$ and NC interactions, as well as 5M cosmic events, output is softmax score.
  - The network does not distinguish between neutrino and anti-neutrino mode.



- Leverage NOvA's excellent timing resolution.
  - Split the 550 μs readout window into 18 μs "time slices."
  - The beam window is fully covered in one of these "slices".
  - Determine if there is a neutrino candidate in each "slice."
  - Perform a cut on cosmic score, and only process slices which pass the cut, dropping all other slices.

A Cosmic Rejection Neural network

Data Sample	Traditional Cosmic Rejection	Cosmic Rejection Neural Network
ν_e	93.21	99.71
$\overline{\nu_e}$	92.81	99.82
ν_μ	93.22	99.20
$\overline{\nu_\mu}$	92.82	99.20
ν NC	93.24	97.08
$\bar{\nu}$ NC	92.79	96.82
Cosmic ν	7.80	5.00

- The Neural Network outperforms the traditional cosmic rejection algorithm in all samples.
  - N.B. The CRNN retains cosmics in the same "time slice" as a neutrino candidate, so within the 5% of cosmics which pass the NN, many will still be easy to reject.

A Cosmic Rejection Neural network

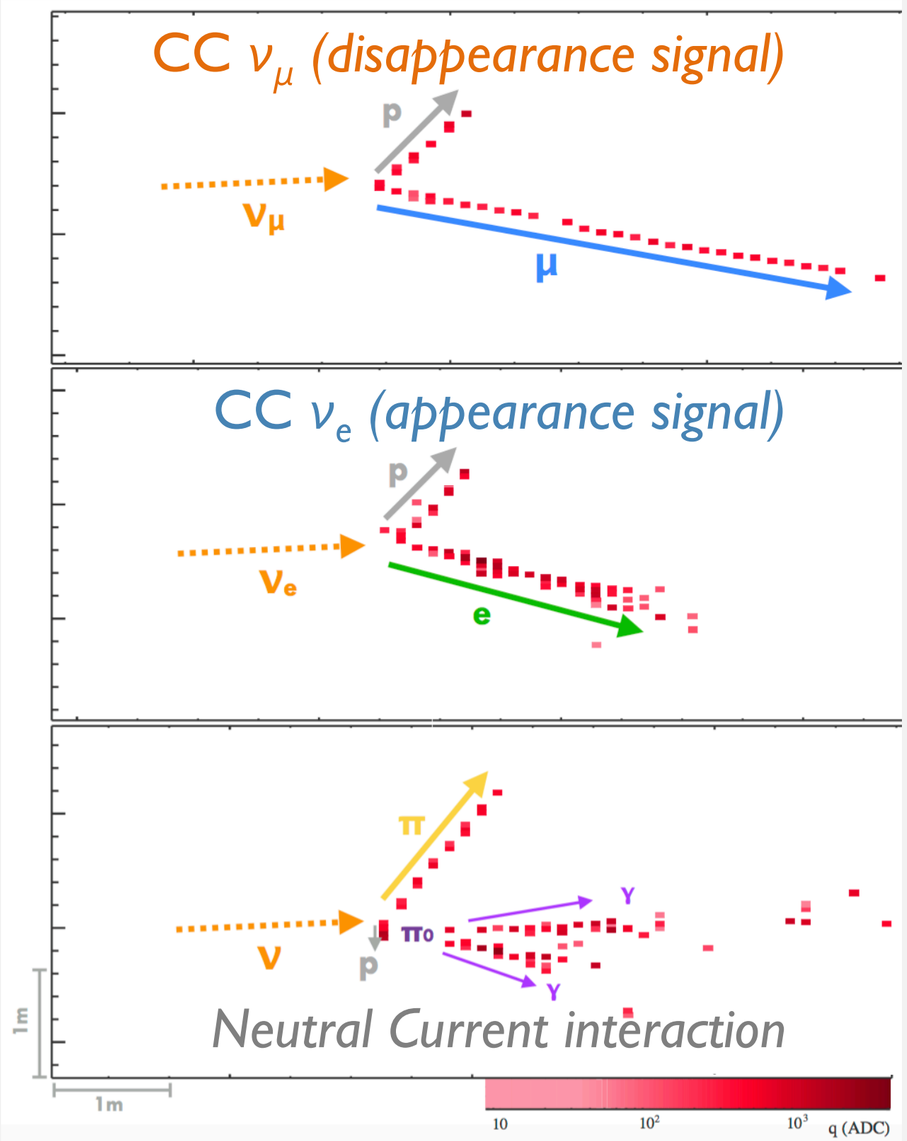
Data Sample	Traditional Cosmic Rejection	Cosmic Rejection Neural Network
ν_e	93.21	99.71
$\bar{\nu}_e$	92.81	99.82
ν_μ	93.22	99.20
$\bar{\nu}_\mu$	92.82	99.20
ν NC	93.24	97.08
$\bar{\nu}$ NC	92.79	96.82
Cosmic ν	7.80	5.00

- The Neural Network outperforms the traditional cosmic rejection algorithm in all samples.
  - N.B. The CRNN retains cosmics in the same "time slice" as a neutrino candidate, so within the 5% of cosmics which pass the NN, many will still be easy to reject.

- Hoped to run inference on GPUs as a separate process, however ran in to numerous problems.
  - Difficulty of even setting up the NOvASoft environment.
  - Difficulty transferring files to/from the GPU.
  - Lack of availability of GPU Open Science Grid (OSG) resources.
- As a result, we ended up performing much of the inference on CPU's, which was both disappointing and much slower.

- Even getting TensorFlow to work reliably on CPUs is very difficult due conflicting dependencies.
  - Further, many OSG nodes are old, and so are unable to run TensorFlow due to old hardware.

Identifying Neutrino Interactions in NOvA



- In order to pursue our physics goals NOvA has to be able to classify neutrino interactions with high efficiency and purity.
  - Machine Learning is one of the tools that we use to do this.
- NOvA was the first HEP experiment to use a CNN in a physics measurement to classify candidate neutrino interactions.
- NOvA is continually working to improve our CNNs, and made many improvements in 2019.

A Convolutional Neural Network Neutrino Event Classifier: https://arxiv.org/abs/1604.01444

Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events: https://arxiv.org/abs/1906.00713

Improving our Training Procedure

Pre-2020 Network Training Workflow

Standard Reco File
- Run Preproc. Job (Month)

LevelDB training files
- Train networks on GPUs (Week)

Trained Network
- Make efficiency tables (Days)

Network with limited val.
- Run reco on dedicated sample of validation files - CPU (Month)

Files with analysis variables
- Run analysis validation scripts (Week)

2020 Network Training Workflow

HDF5 File
- Produced by standard Reco. chain
- Train Network on GPU (Day)

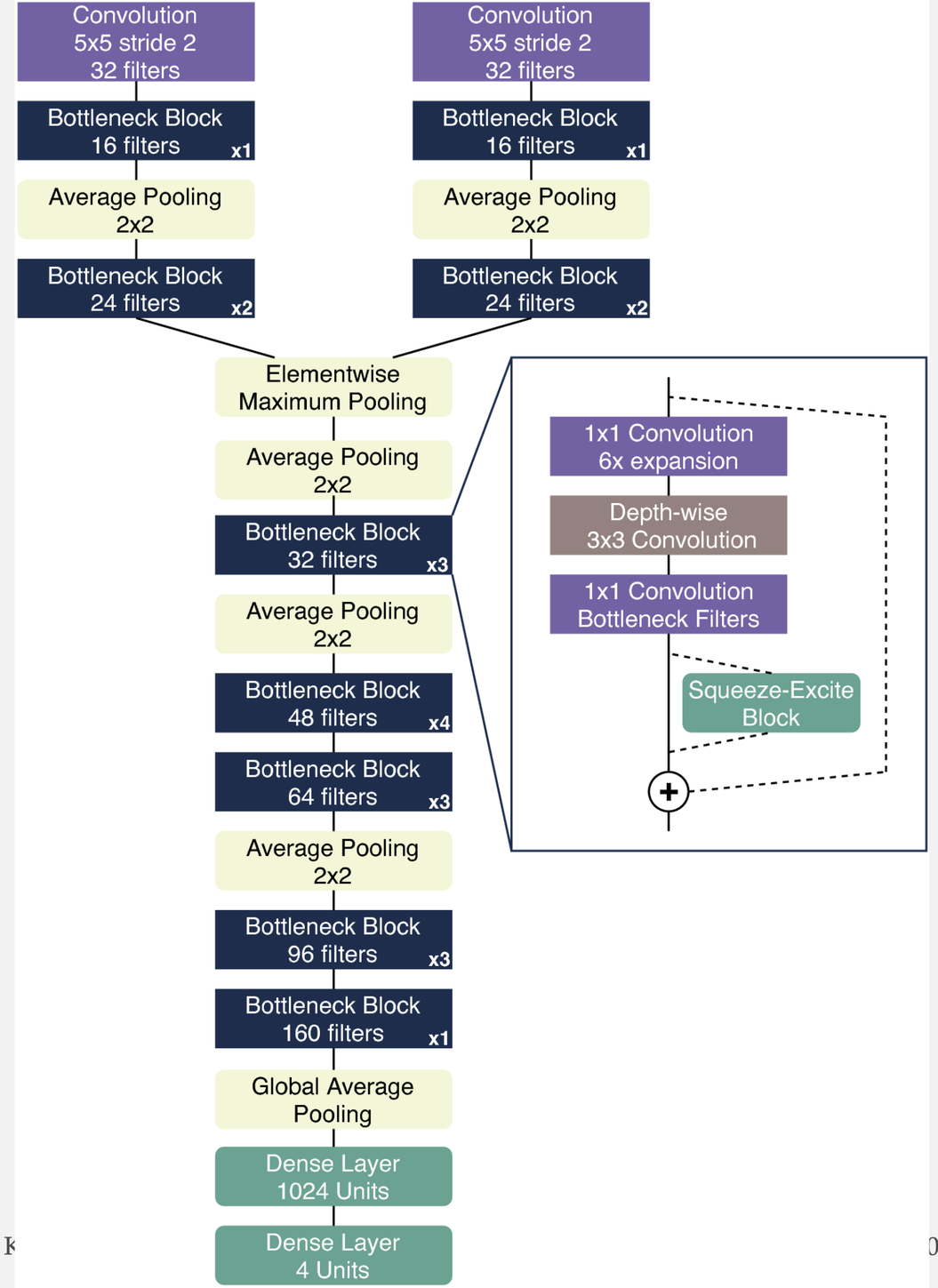
Trained network
- Make efficiency tables (Day)
- Run analysis validation scripts (Day)

Fully trained and validated network
- Complete in the order of a Week.
- All work done on GPUs

- During 2019 we transitioned from training networks using LevelDBs, Root and Caffe to using HDF5, Pandas and Tensorflow.
  - Found that all aspects of training were significantly faster, and that network inference was roughly 7 times faster.
- A process which previously took months, can now be done on the order of a week.
  - This also allows us to train many network variants to explore how systematic uncertainties (eg detector calibration) affect network performance.

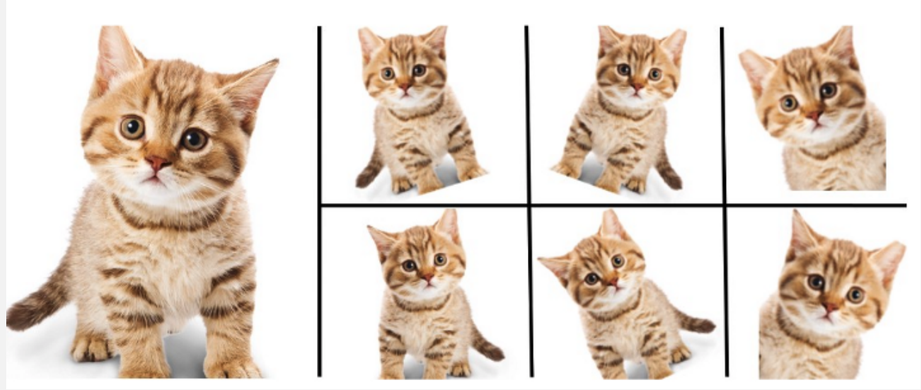
Changing the Network Architecture - Mobile Nets

- Moved away from our custom architecture based off GoogLeNet.
- Considered many new architectures, including Res Nets.
  - Found that Mobile Nets achieved all of the performance gains we could see by using Res Nets, whilst being almost 4 times faster.
- The main component of the architecture is the bottleneck block which expands the input features and performs a depth-wise convolution before compressing the features once again.



Improving our Training Environment

Data Augmentation

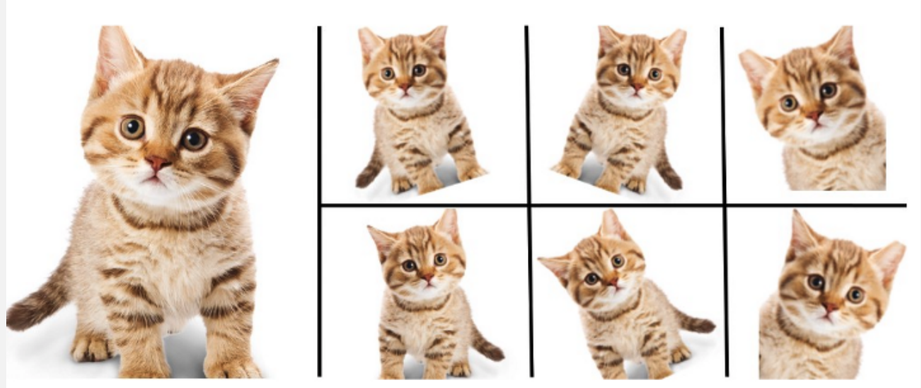


- Uncertainties in Calibration are our leading systematic.
  - Want training to build in resilience to this uncertainty.

- Uniformly scale pixel maps used in training by $\pm 10\%$.
  - We find that this makes the networks more resilient to calibration changes, and more accurate overall.

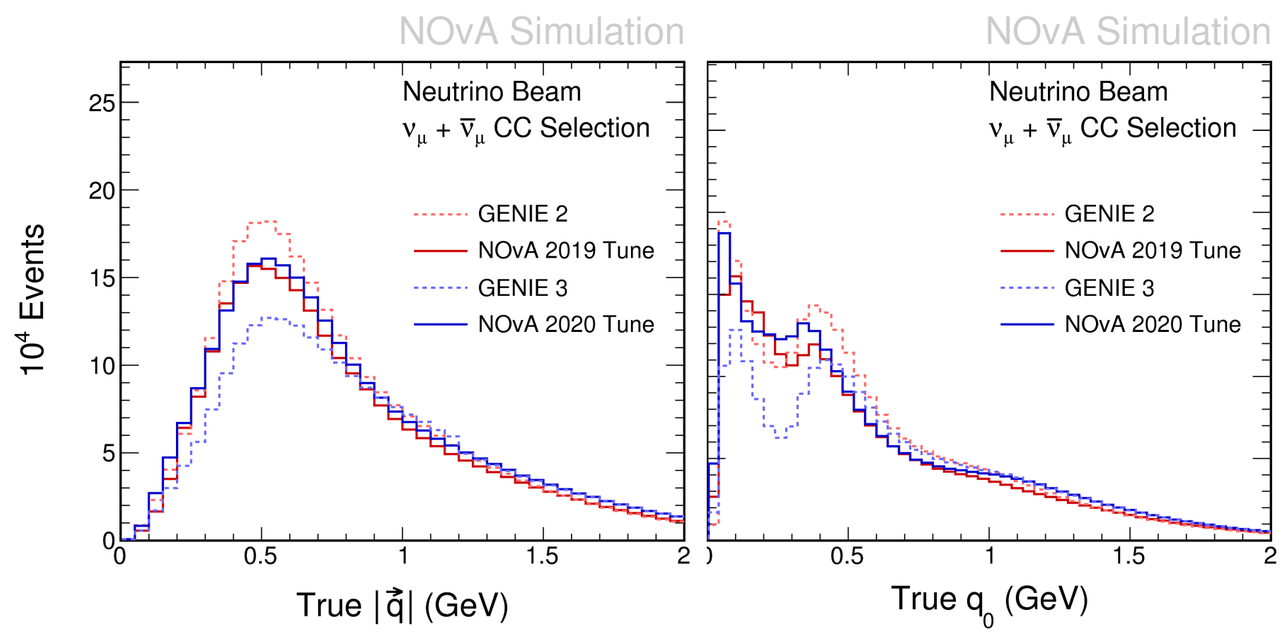
Improving our Training Environment

Data Augmentation



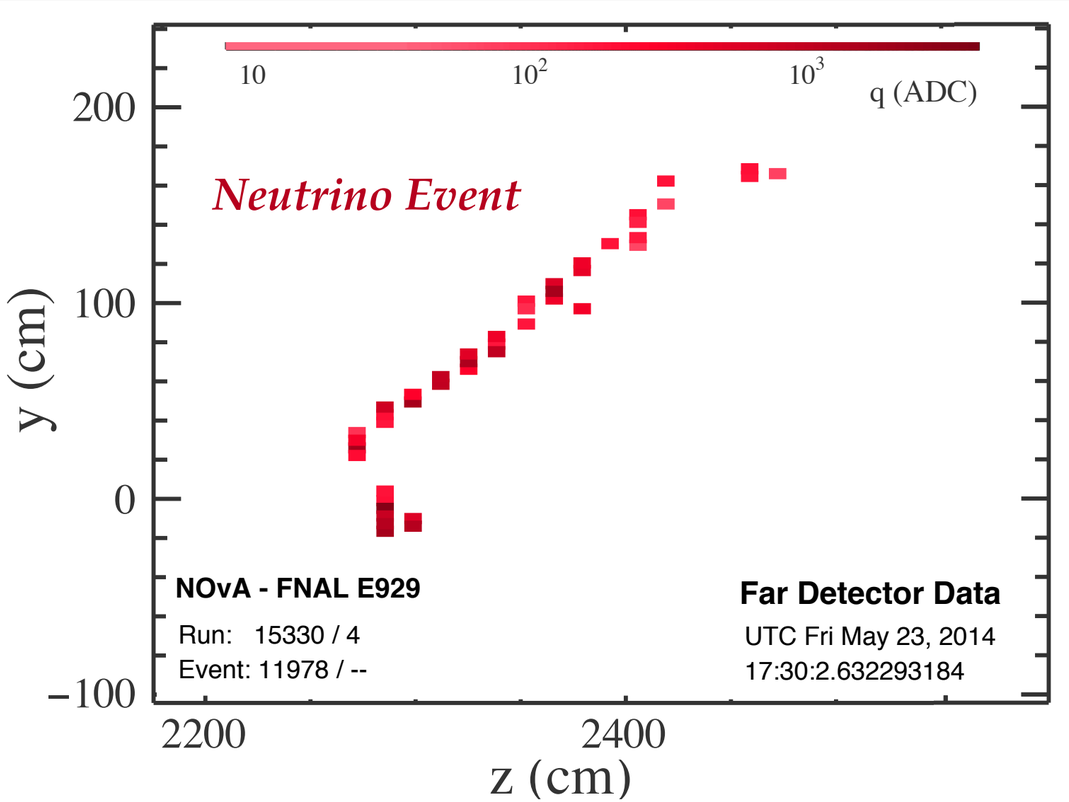
- Uncertainties in Calibration are our leading systematic.
  - Want training to build in resilience to this uncertainty.
- Uniformly scale pixel maps used in training by $\pm 10\%$.
  - We find that this makes the networks more resilient to calibration changes, and more accurate overall.

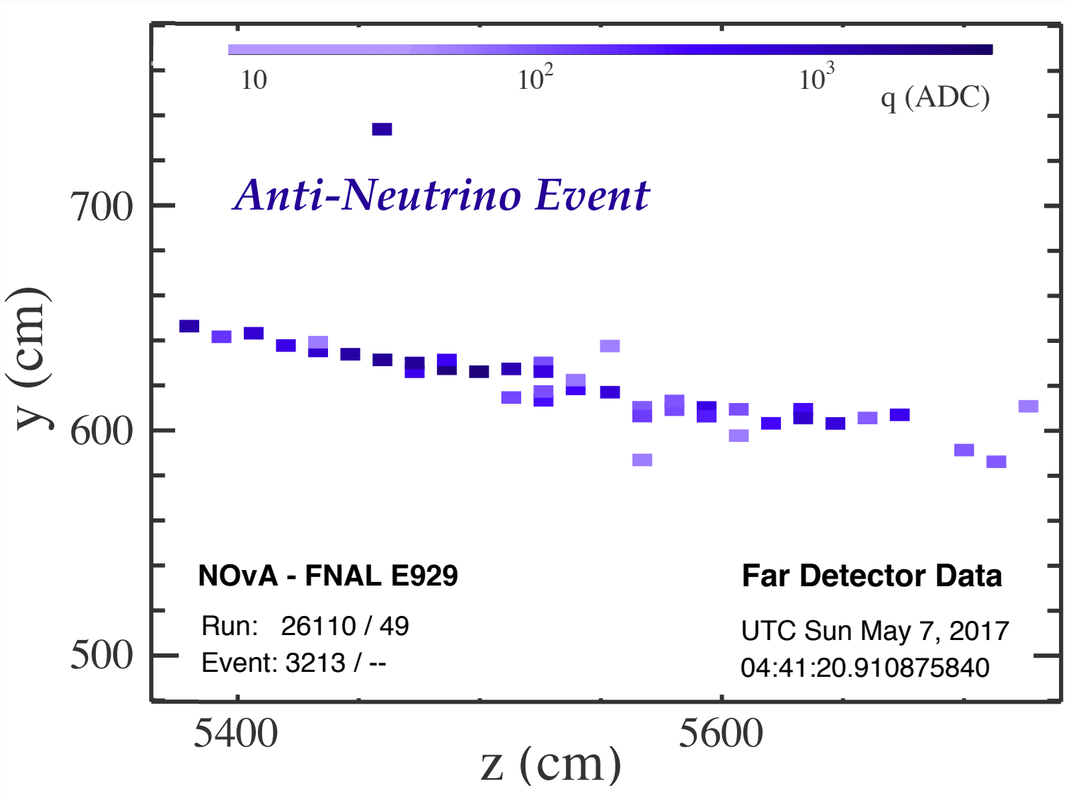
More Realistic Training Sample



- GENIE 3 is a much more accurate depiction of what we see in our data than GENIE 2 was.
  - Therefore, the events which we use to train our networks resemble nature much more accurately than previously.

Separate networks for Neutrino and Anti-Neutrino Modes

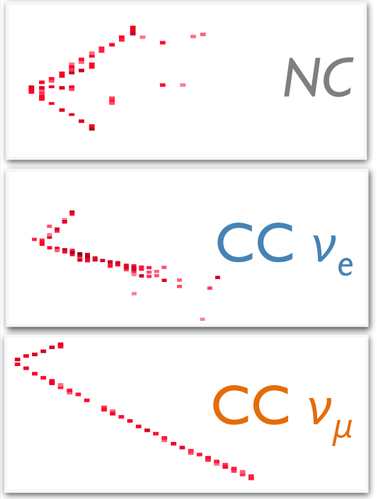




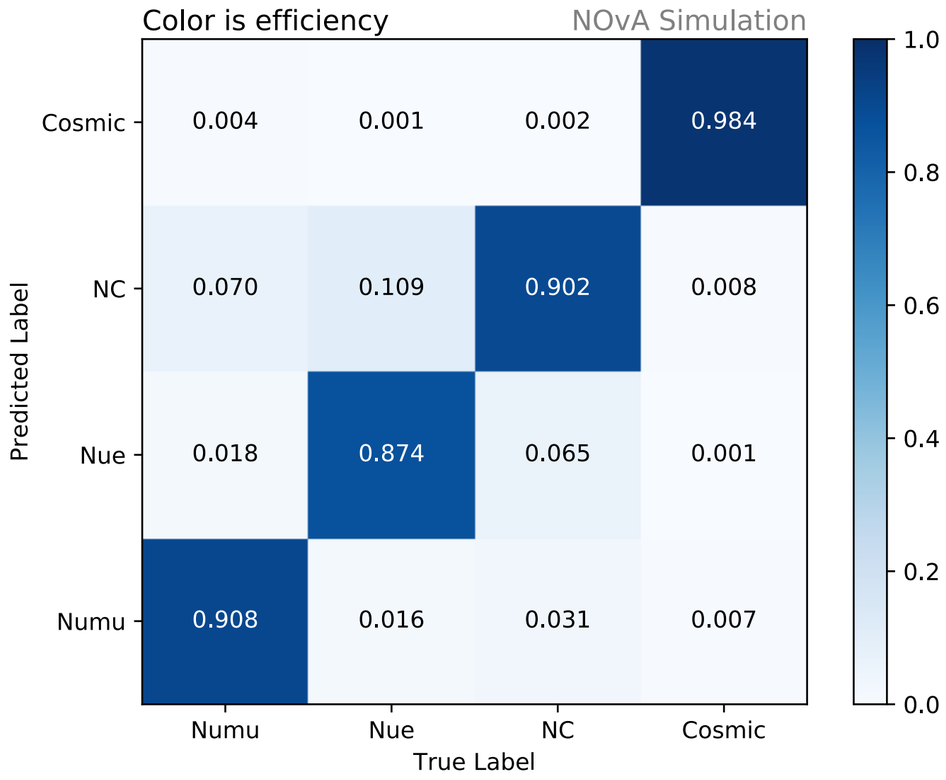
$\bar{\nu}$ **Efficiency Improvement** Training Sample (ID > 0.9)		
$\bar{\nu}_e$ CC Signal	$\bar{\nu}_\mu$ CC Signal	$\bar{\nu}$ NC Signal
14%	6%	10%

The event topologies for neutrinos and anti-neutrinos are different on average, so separate networks are trained for each horn current.

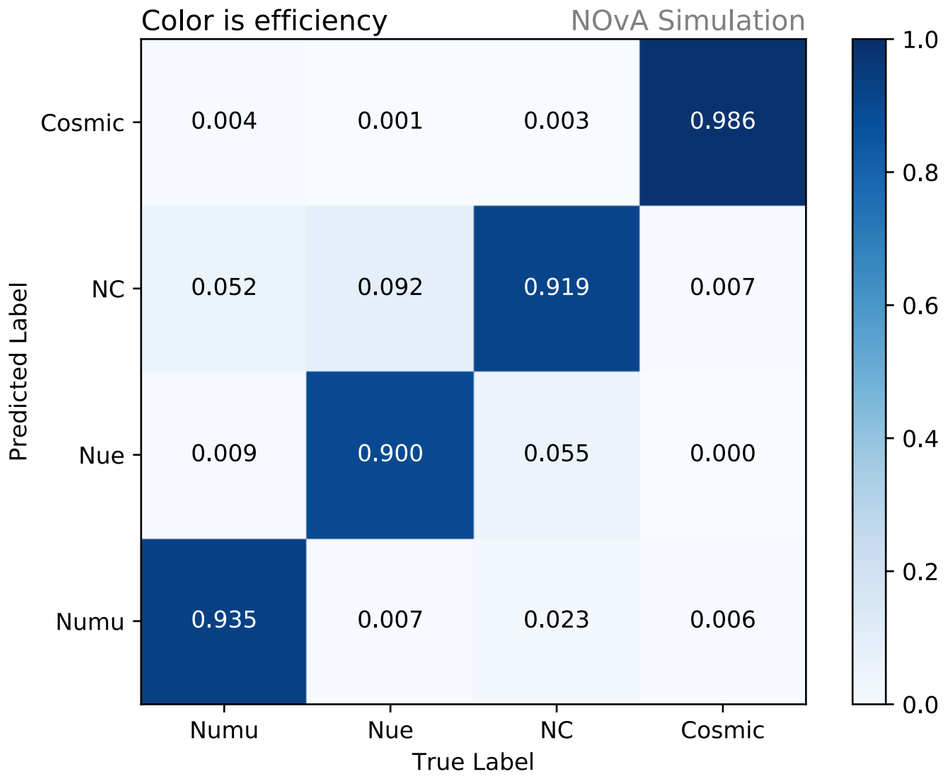
Network Performance - Efficiency



Neutrino Mode



Anti-Neutrino Mode



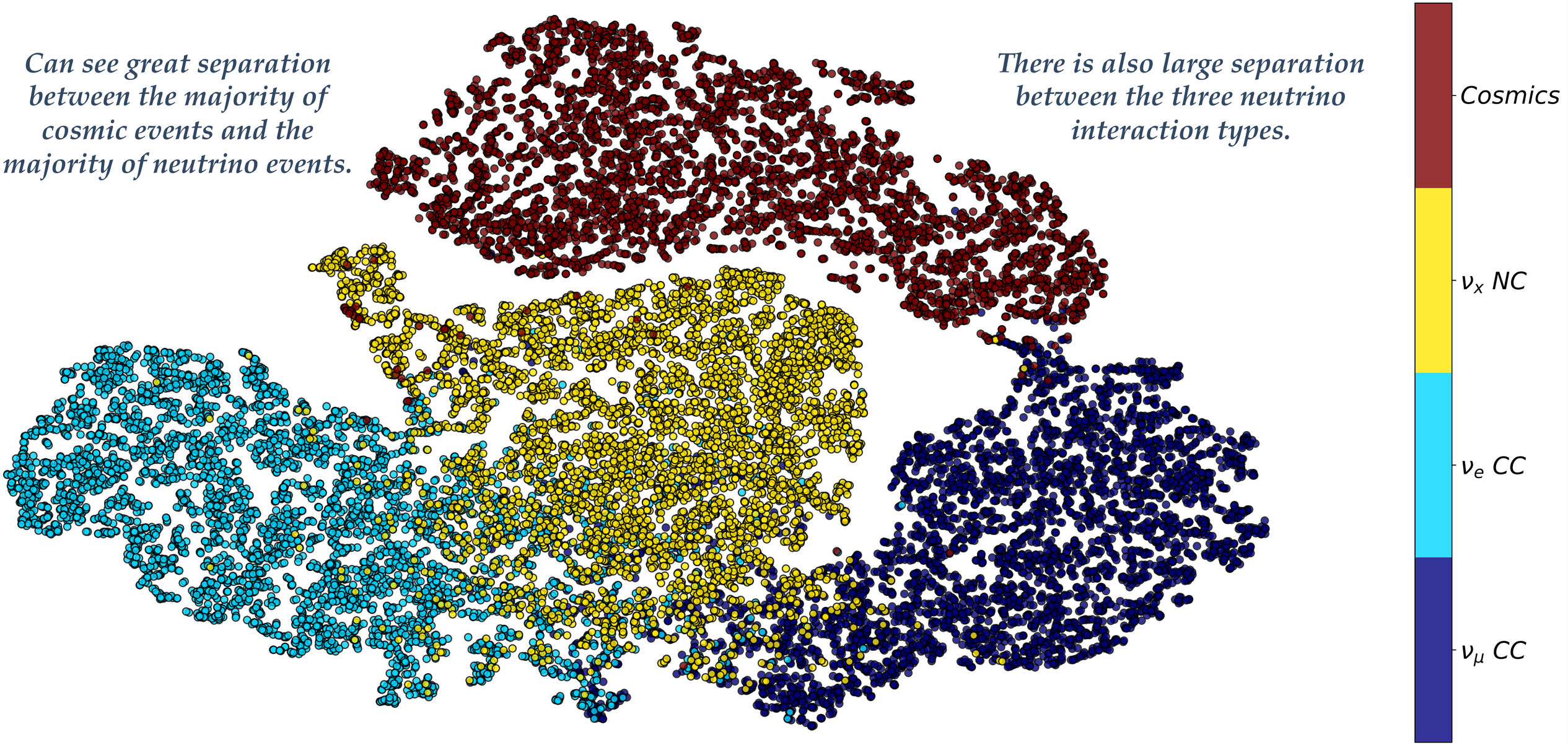
Very similar performance between neutrino and anti-neutrino modes, though the anti-neutrino mode network appears to be more efficient.

See purities of over 99% for cosmics, and over 90% for all neutrino interaction channels for both horn polarities.

t-Distributed Stochastic Neighbour Embedding

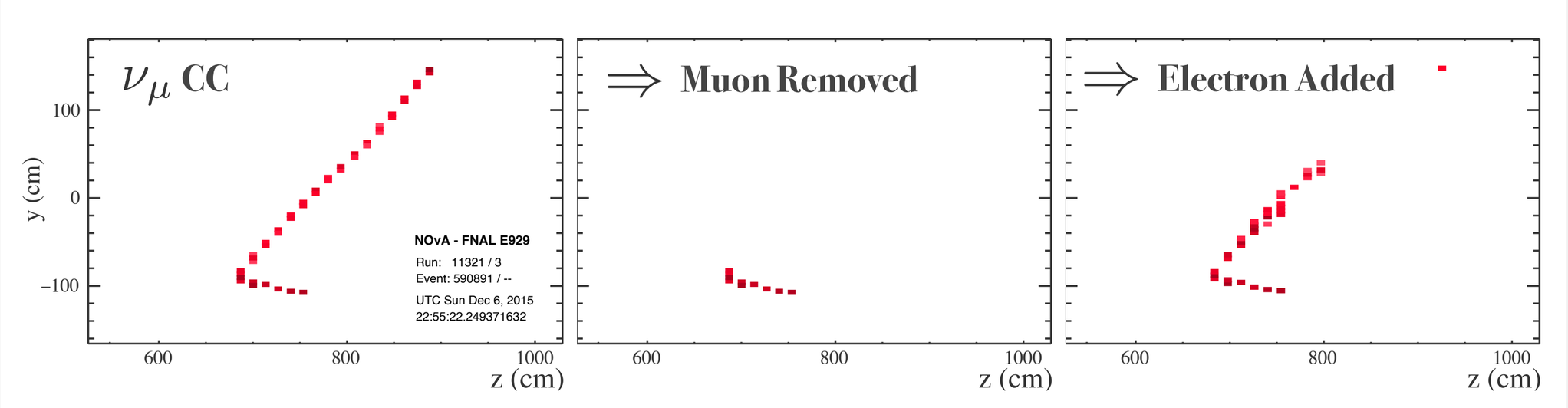
Can see great separation between the majority of cosmic events and the majority of neutrino events.

There is also large separation between the three neutrino interaction types.



Validating Event CVN

Muon Removed, Electron Added in the Near Detector- MRE



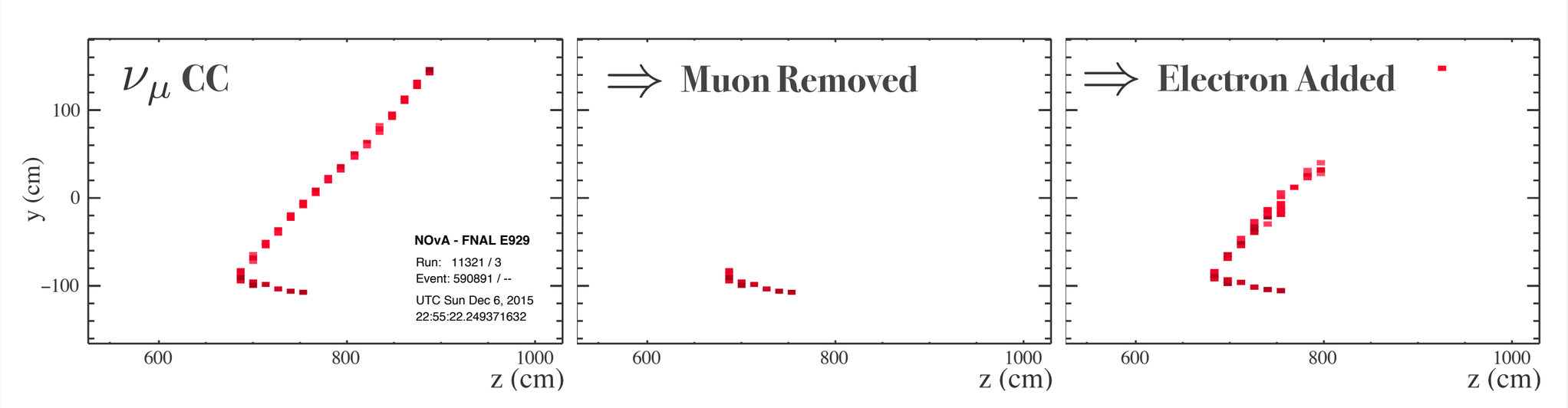
Select a Muon Neutrino interaction from Data/Monte Carlo.

Remove the hits associated with the muon.

Simulate an electron with the same energy as the removed muon.

Validating Event CVN

Muon Removed, Electron Added in the Near Detector- MRE



Select a Muon Neutrino interaction from Data/Monte Carlo.

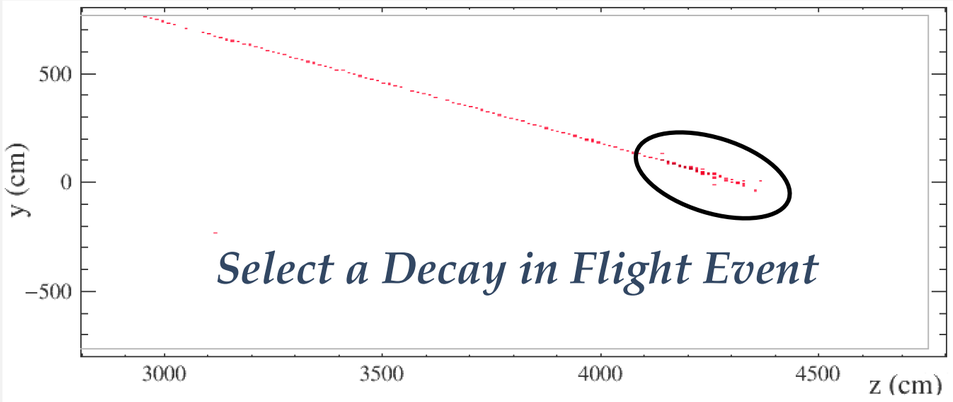
Remove the hits associated with the muon.

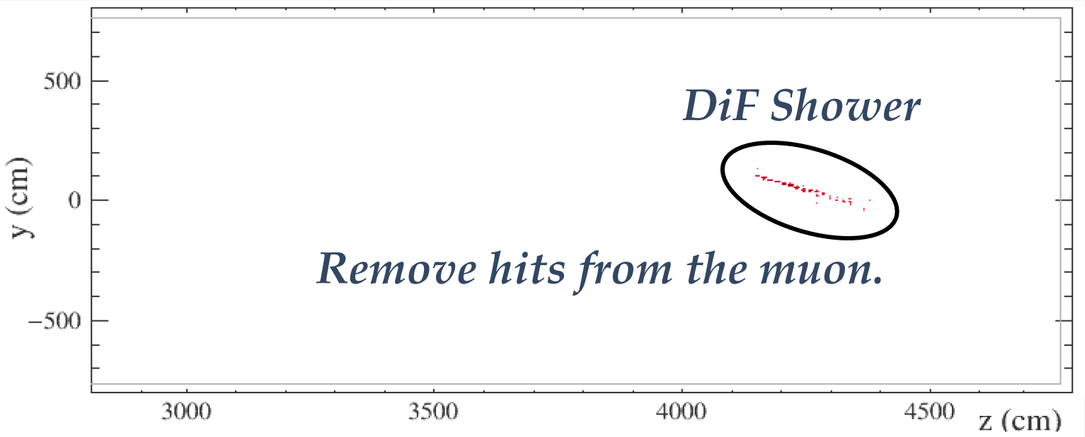
Simulate an electron with the same energy as the removed muon.

		Preselection	Full Selection	Efficiency	Difference
Neutrino beam 9.5×10^{20} POT	Data	709112	564669	0.796	-0.76%
	MC	772566	619908	0.802	
Antineutrino beam 11.8×10^{20} POT	Data	418245	348151	0.832	-0.46%
	MC	475300	397454	0.836	

Validating Event CVN

Muon Removed, Decay In Flight in the Far Detector – MRDiF – New for the 2020 analysis

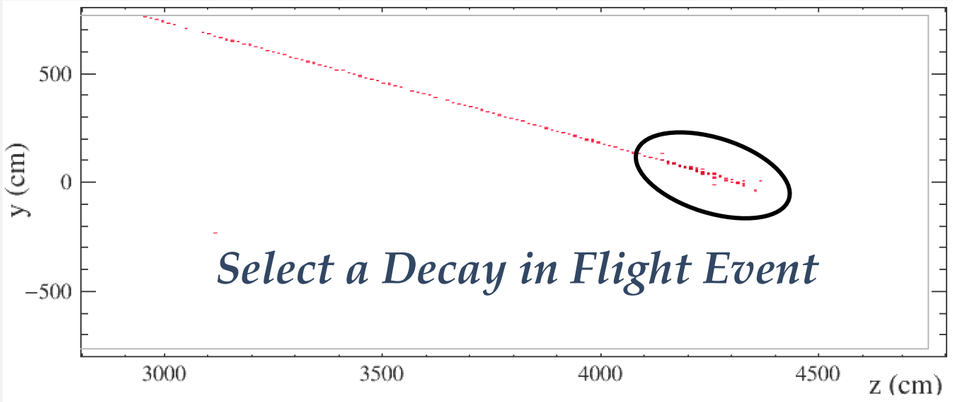


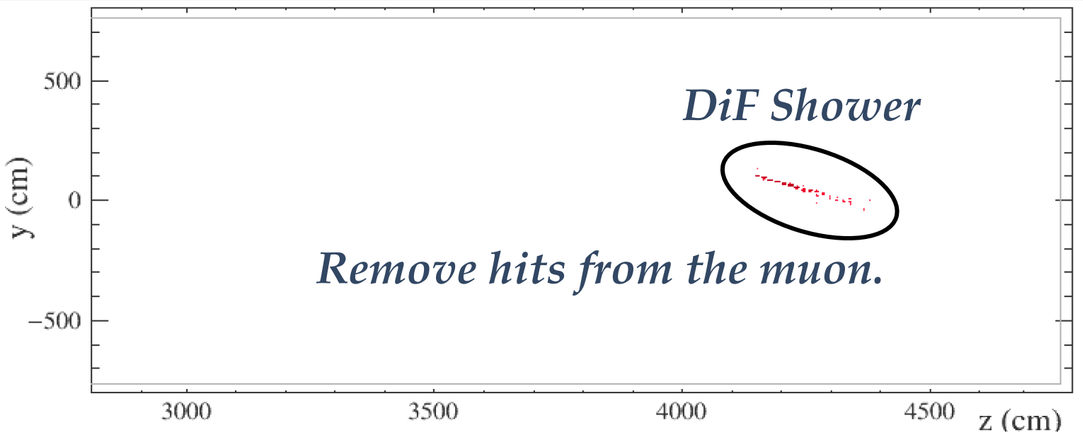


Select an event where the comic muon decays in flight, and remove the muon, leaving only a pure electromagnetic shower.

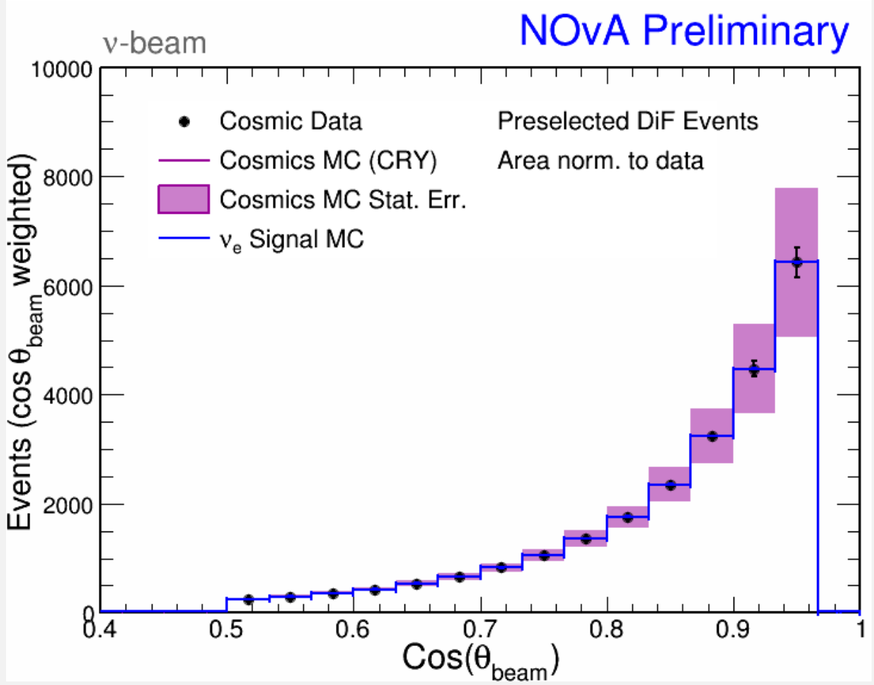
Validating Event CVN

Muon Removed, Decay In Flight in the Far Detector – MRDiF – New for the 2020 analysis



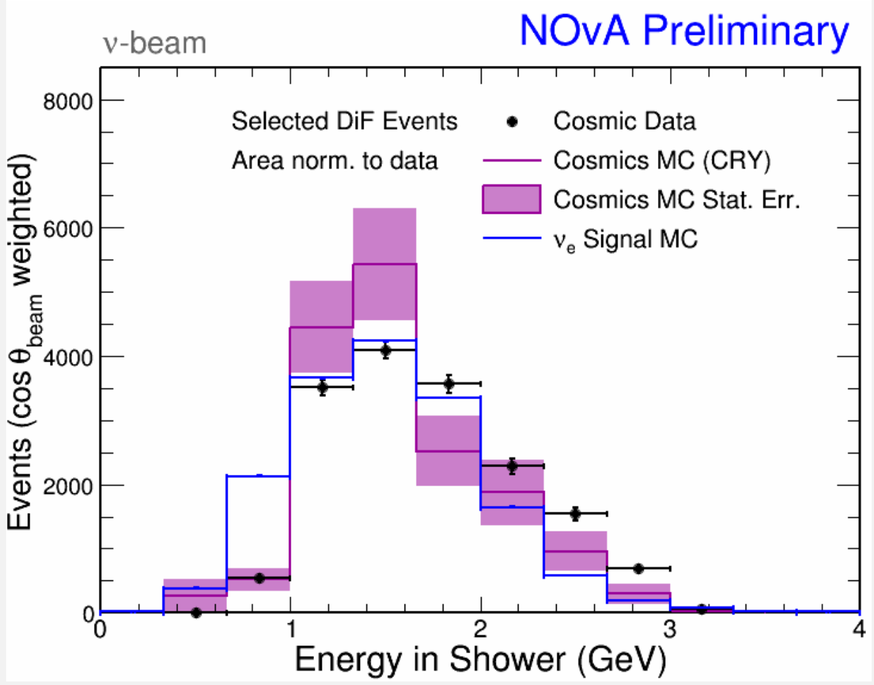


Select an event where the comic muon decays in flight, and remove the muon, leaving only a pure electromagnetic shower.

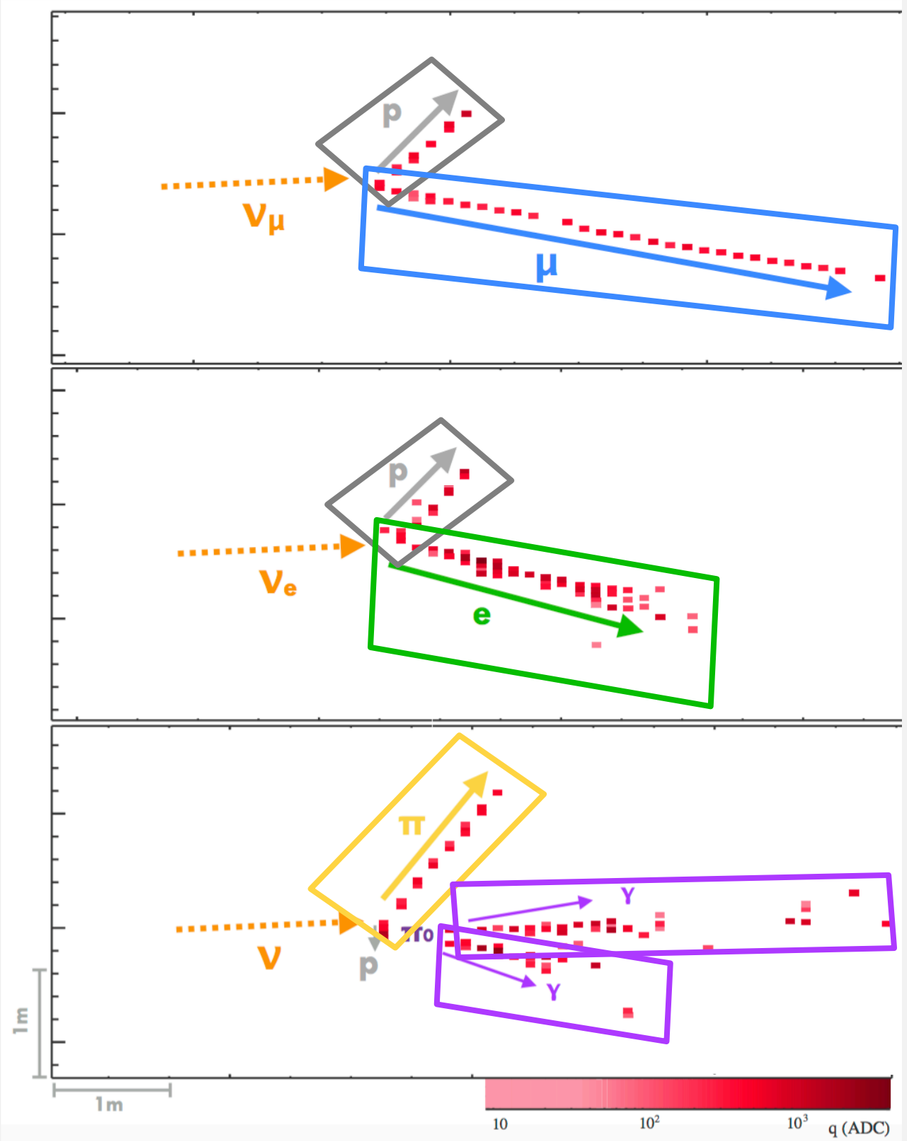


Left: Events are reweighted by their angle wrt to the beam.

Right: The selection efficiencies for MRDiF events in Data and Monte Carlo agree within errors.



Identifying Individual Particles in NOvA

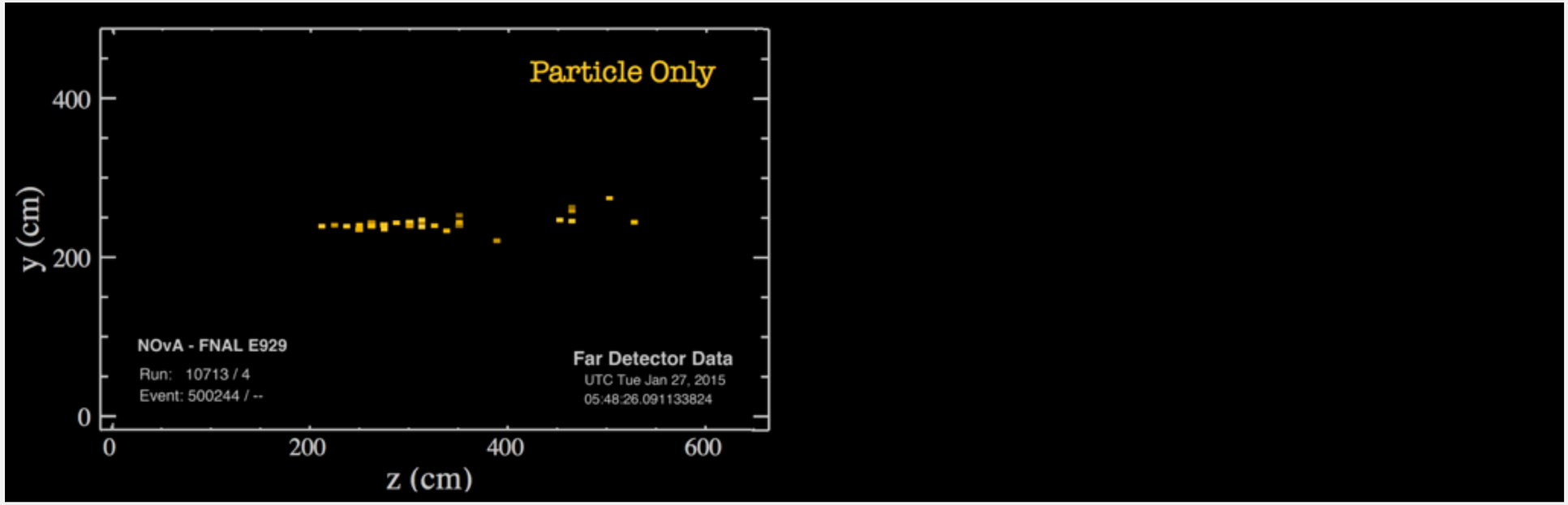


- It is also essential that we are able to classify individual particles with high efficiency and purity.
- We provide hits from individual particle clusters, which are made by geometrical reconstruction.

A Convolutional Neural Network Neutrino Event Classifier: https://arxiv.org/abs/1604.01444

Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events: https://arxiv.org/abs/1906.00713

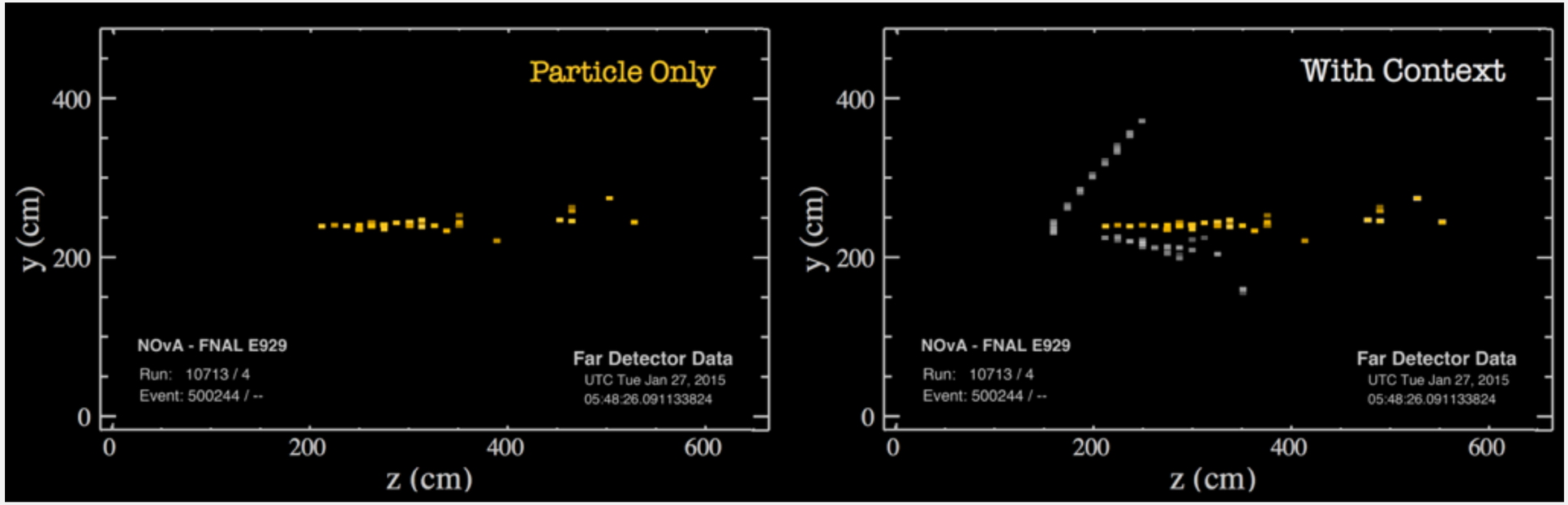
Context-Enriched Particle Identification



Is this particle;

An electron
OR
A Photon?

Context-Enriched Particle Identification



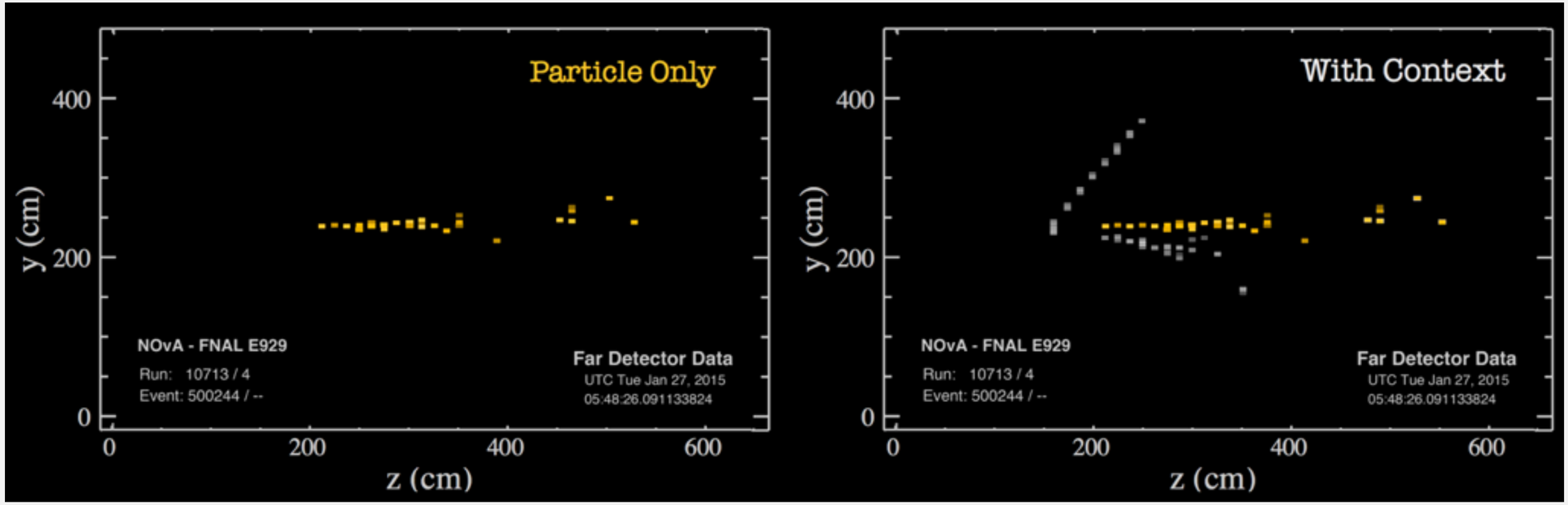
Is this particle;

An ~~electron~~ ✗

OR

A Photon? ✓

Context-Enriched Particle Identification

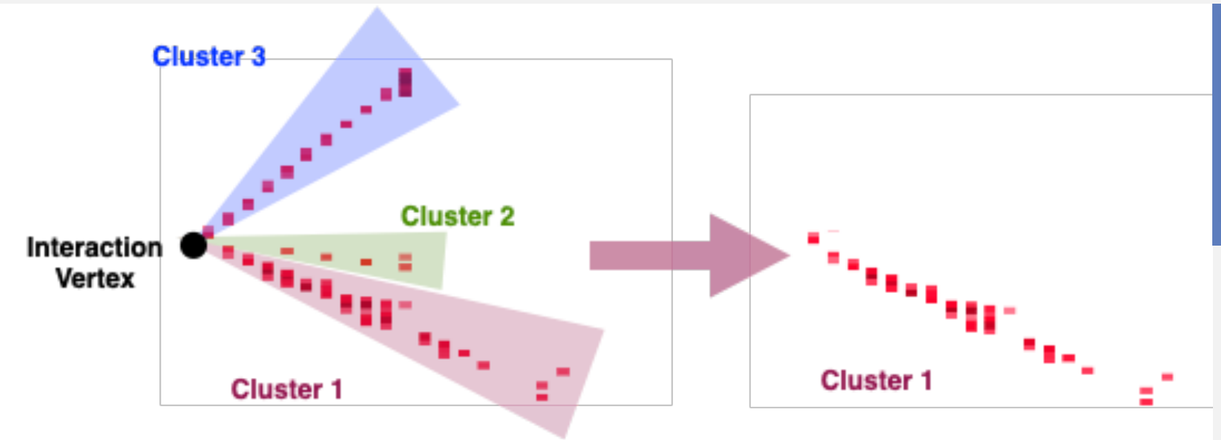


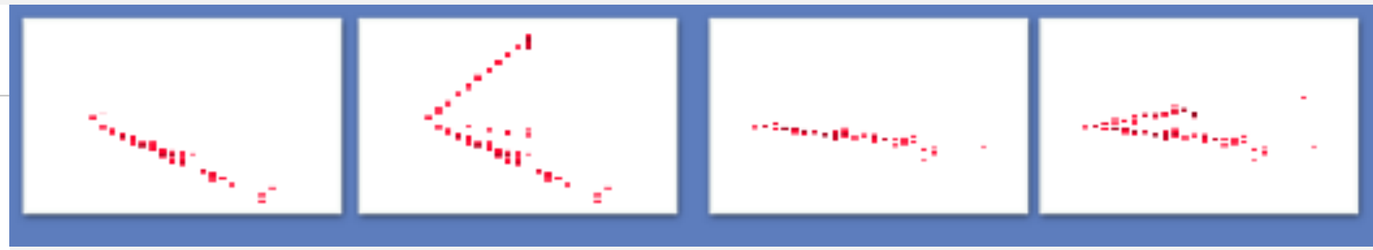
Is this particle;

An ~~electron~~ ✗

OR

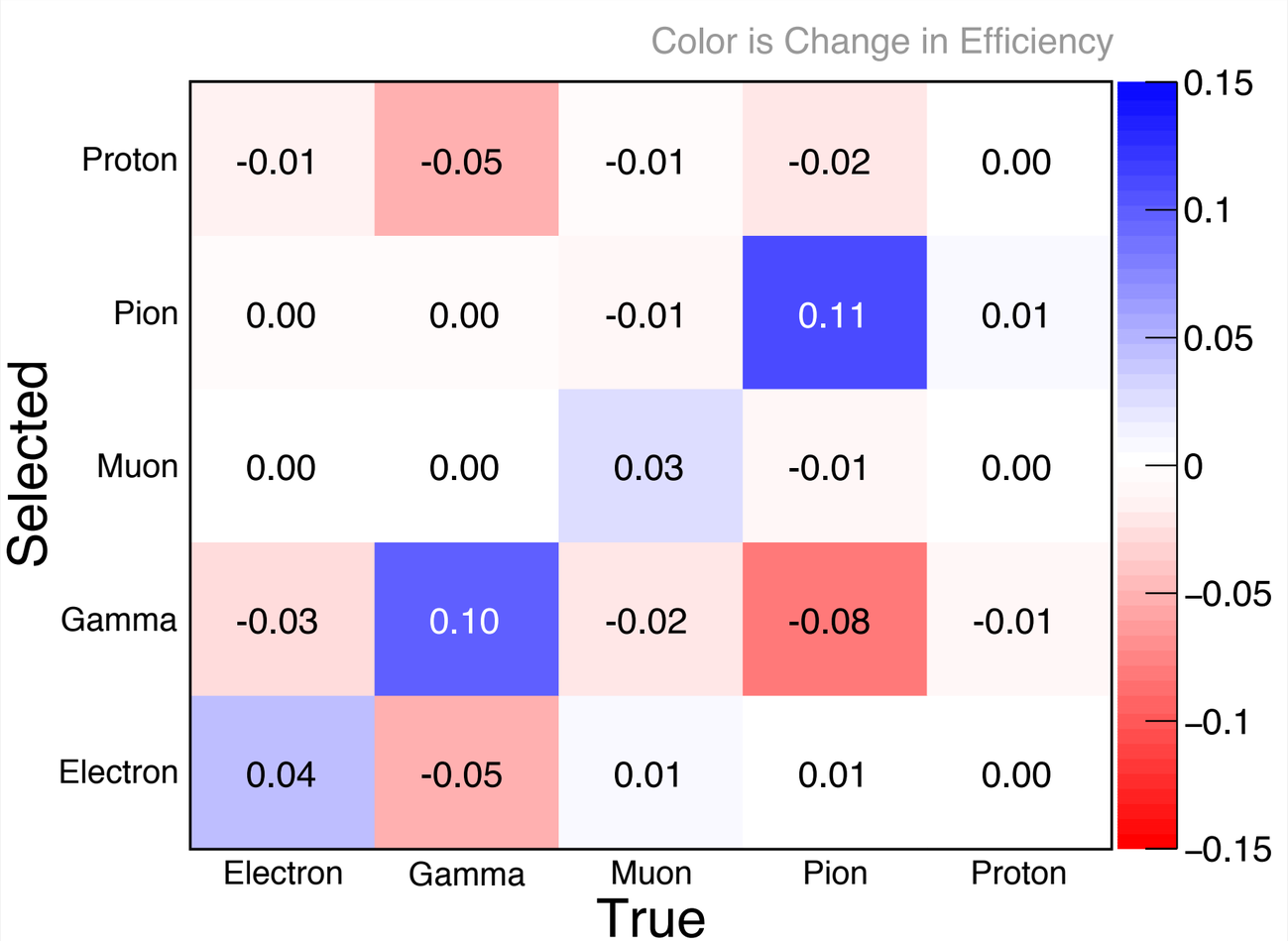
A Photon? ✓

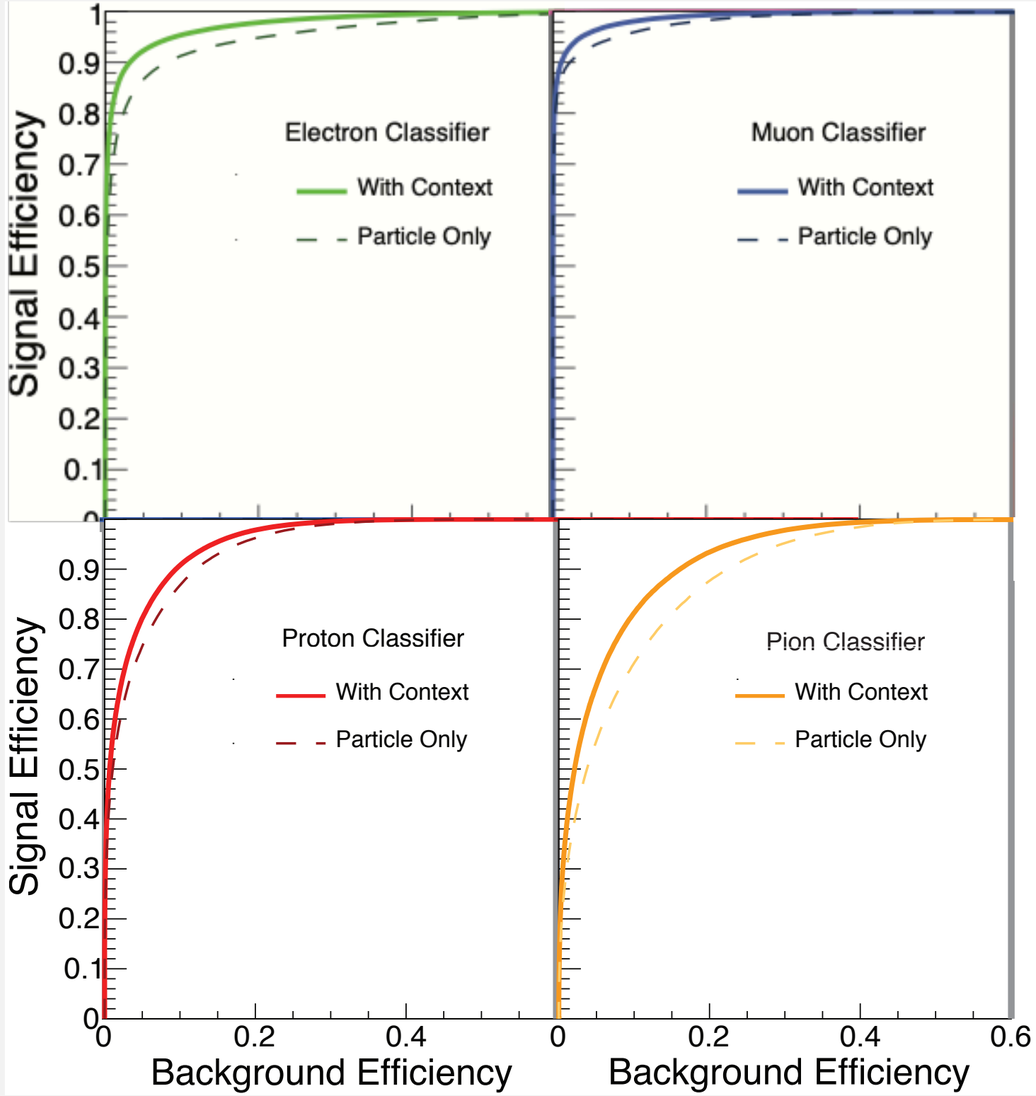




We classify particles using both views of the particle, and both views of the entire event. This provides the network with contextual information regarding the interaction.

The Effect of Providing Contextual Information





Providing contextual information increases network performance by up to 11% in the case of pion identification.

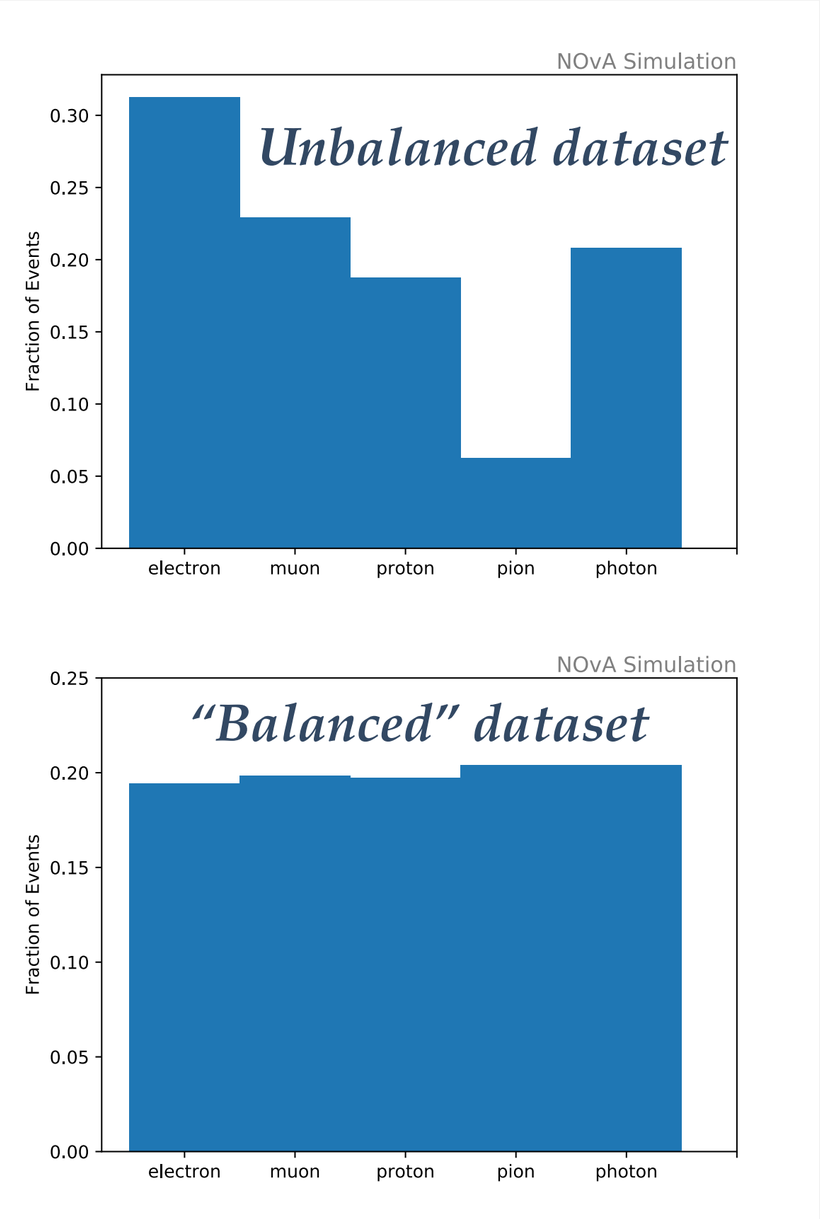
Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events: https://arxiv.org/abs/1906.00713

Changes to the Training Model

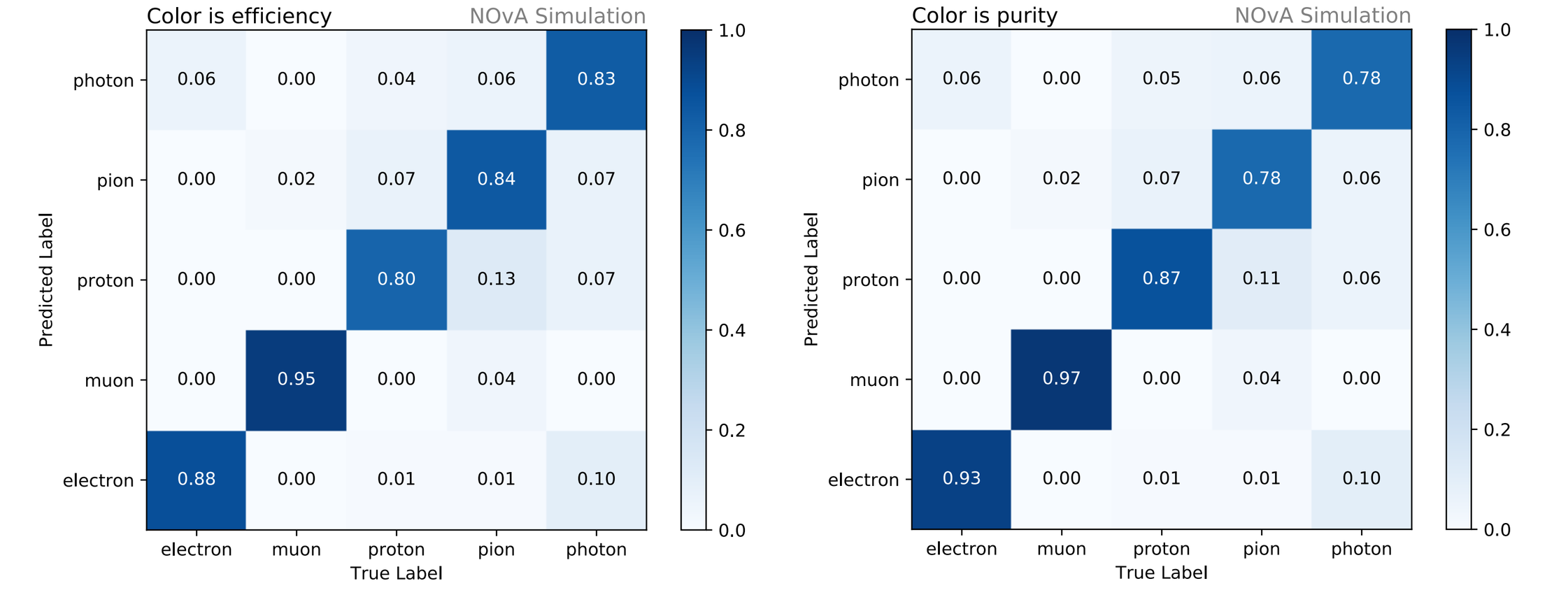
- We also switched to using a Modified Mobile Network architecture for our individual particle identifier.

- Training sample is "balanced" between the 5 main particle types which we see in the detector.
  - Using a balanced dataset significantly improves accuracy.
  - We see a 30% increase in Pion efficiency after moving to the "balanced" dataset.

- In training we require that the particle cluster is contained within the detector in both views, and that the cluster has high hit purity to remove multi-particle clusters.



Performance of the Network

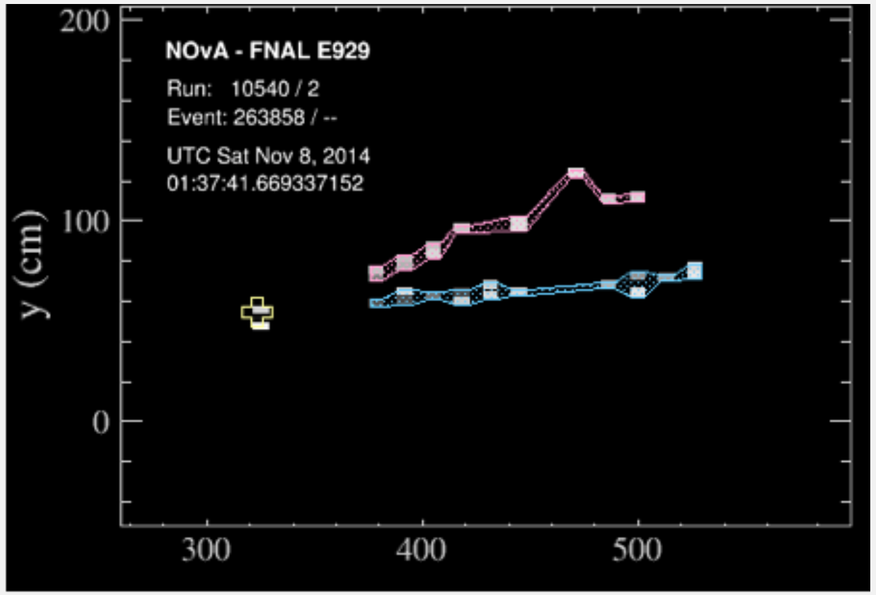


We use a single network for both neutrino and anti-neutrino modes as we see no improvement in performance by having separate networks.

Achieve very high purities for all particle types, particularly for electrons and muons.

Validating Prong CVN

Measuring the π^0 Mass Peak

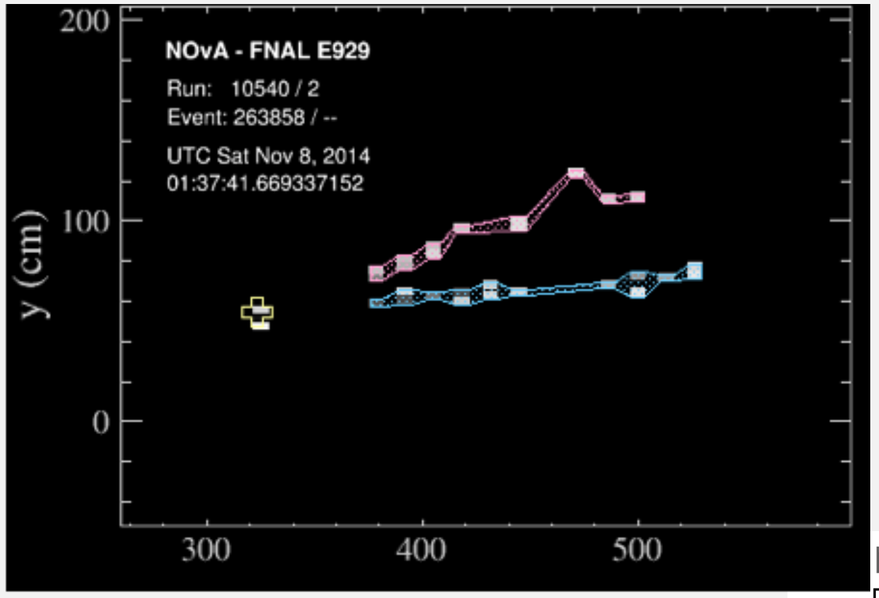


Above: A simulated π^0 decay.

A π^0 decays to a pair of photons with a well-defined invariant mass of 135 MeV.

Validating Prong CVN

Measuring the π^0 Mass Peak

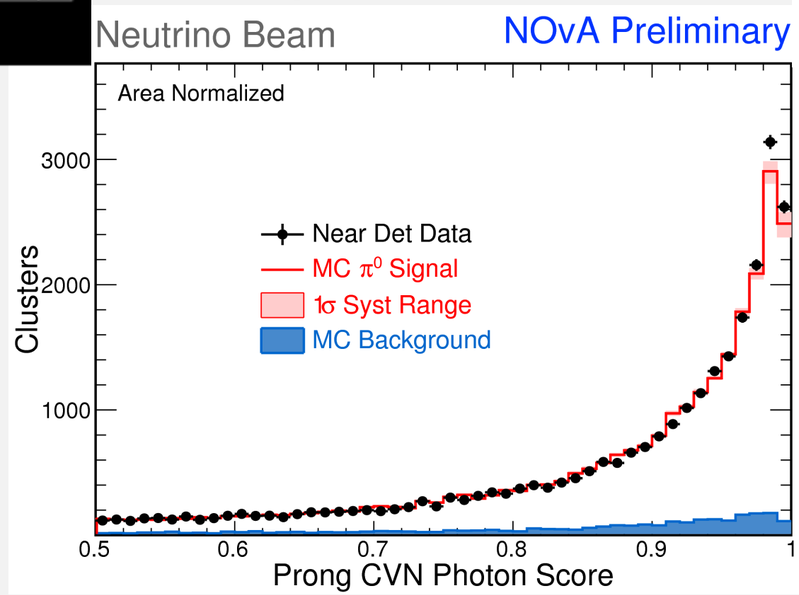


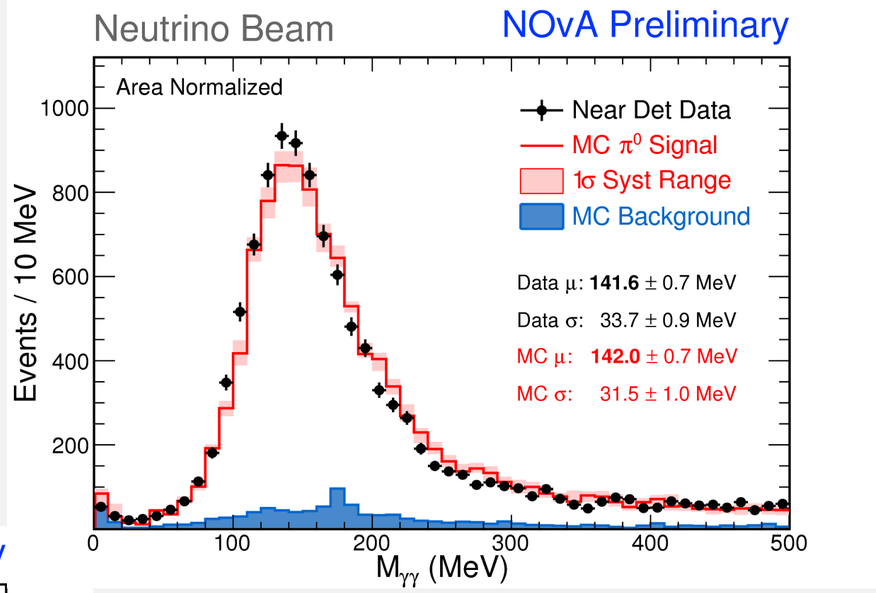
Above: A simulated π^0 decay.

A π^0 decays to a pair of photons with a well-defined invariant mass of 135 MeV.

Below: The Photon score for selected events.

Selected events must have two photons with photon scores of greater than 0.5.



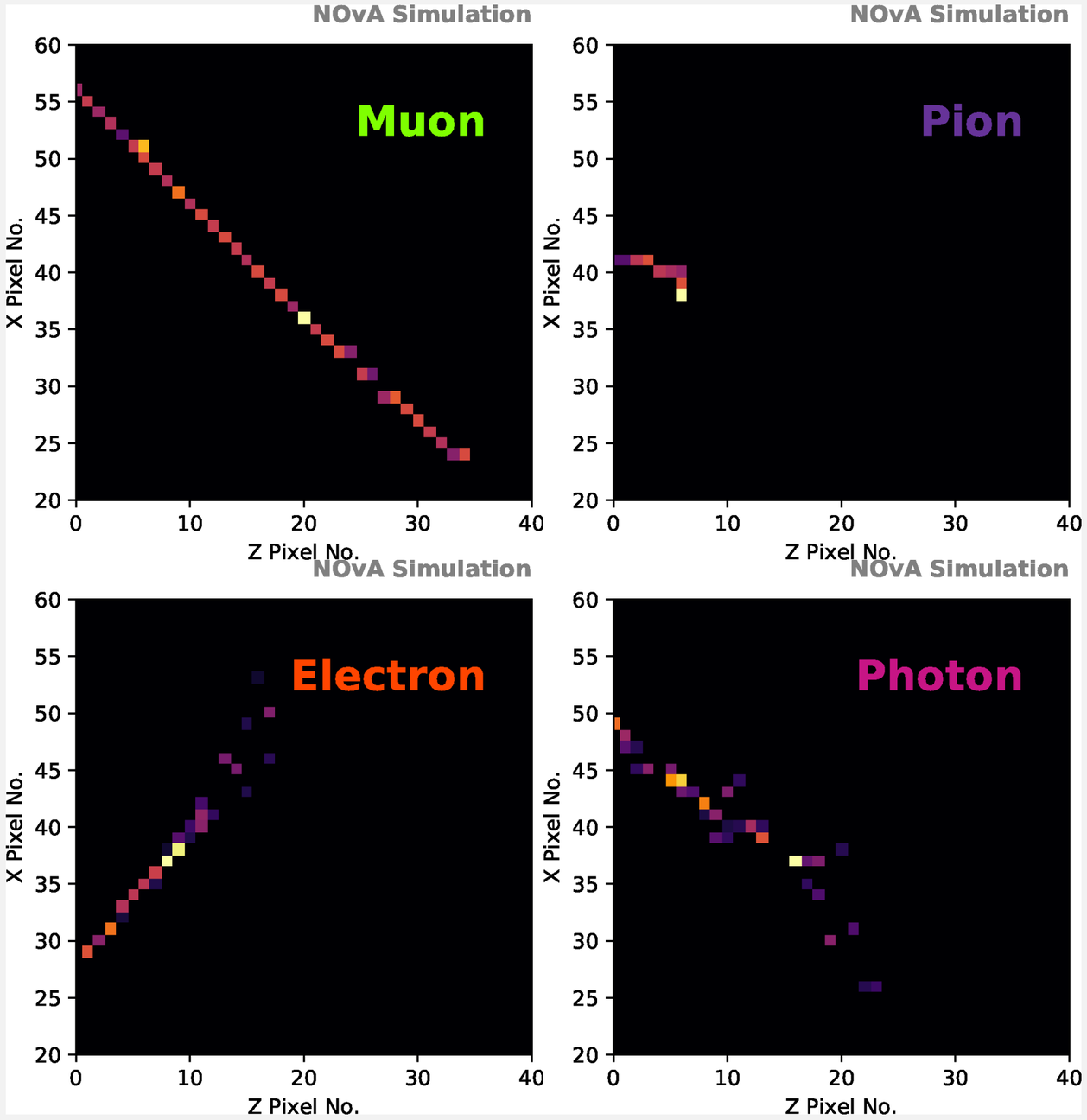


Above: The measured invariant mass of π^0 decays in both Data and Monte Carlo.

The CNN based approach yields a 60% reduction in background, for the same efficiency.

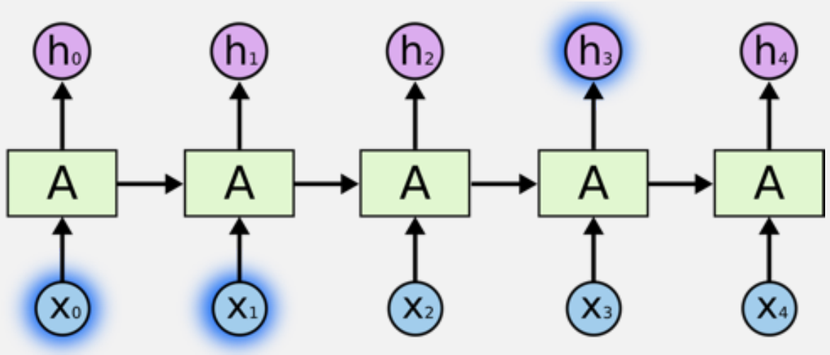
Single Particle Training

- Our context-enriched network is trained on GENIE simulated events.
- No-biases have been observed when training on this sample but out of an abundance of caution a network trained using singularly simulated particles is used in our Near Detector analyses.
  - As such, no contextual information can be used.
- We are also developing a network designed for neutron identification using these samples.



Neutrino Energy Estimation

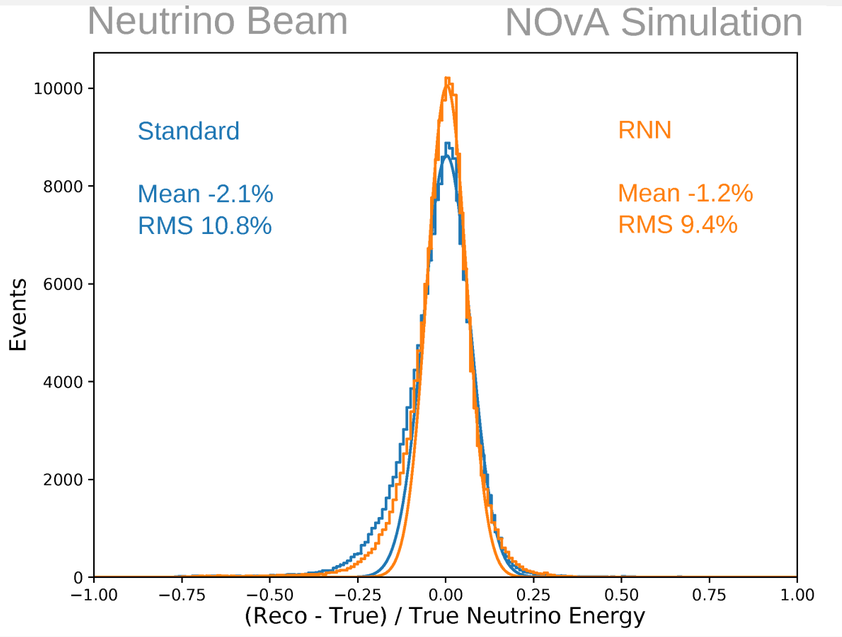
Long Short-Term Memory – Dmitrii Torbunov



Newly developed in 2020.

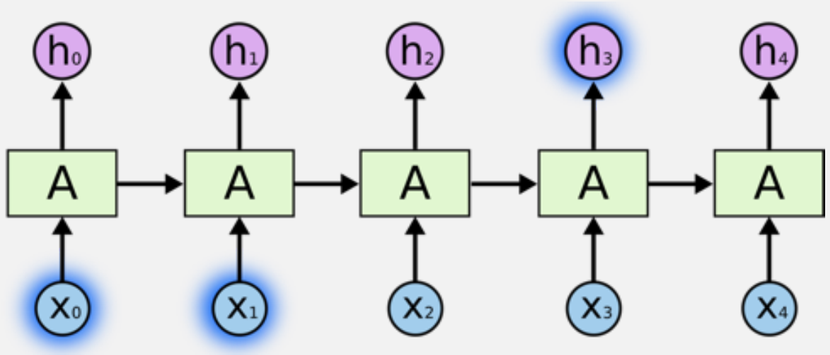
Takes a number of traditional reconstruction quantities as inputs.

Trained using calibration shifts to increase network resilience.



Neutrino Energy Estimation

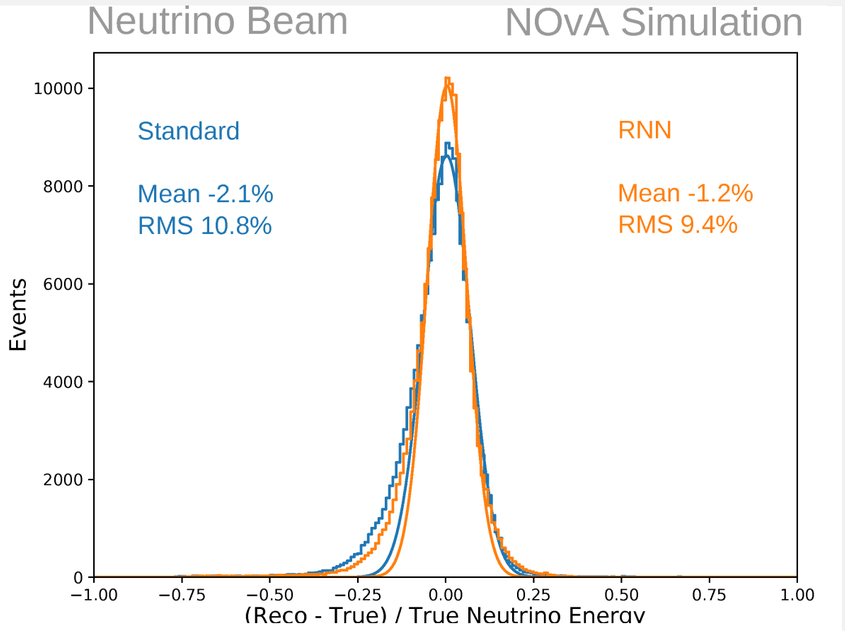
Long Short-Term Memory – Dmitrii Torbunov



Newly developed in 2020.

Takes a number of traditional reconstruction quantities as inputs.

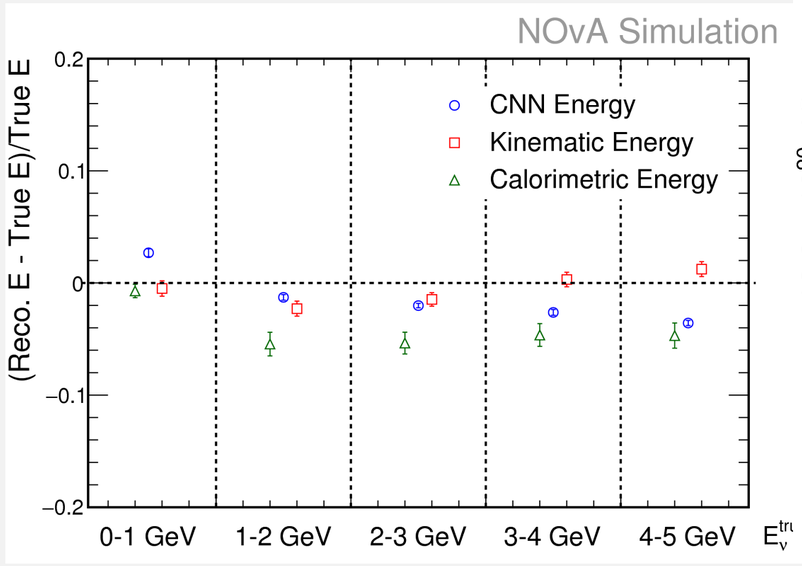
Trained using calibration shifts to increase network resilience.

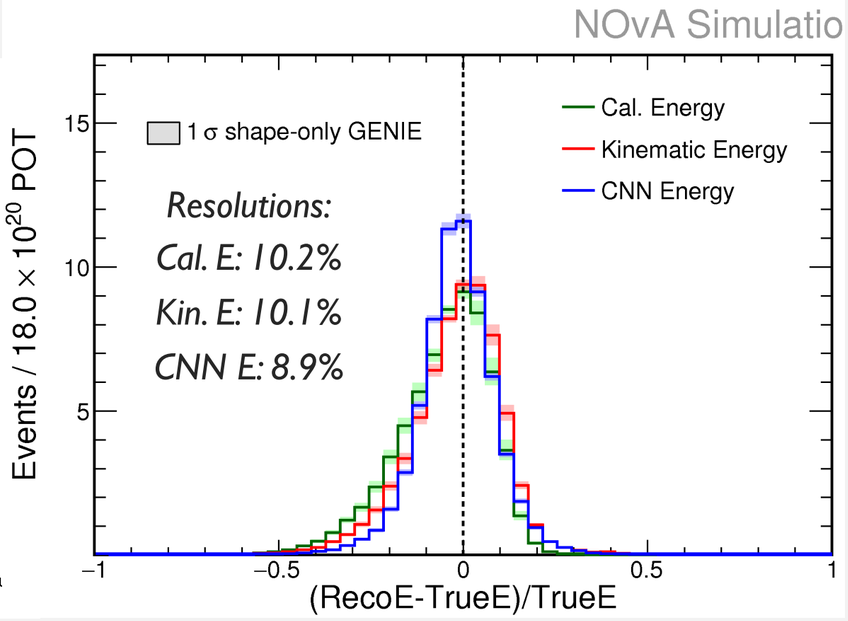


Regression CNNs – Ben Jargowsky

Uses the same pixel maps which Event CVN is trained on.

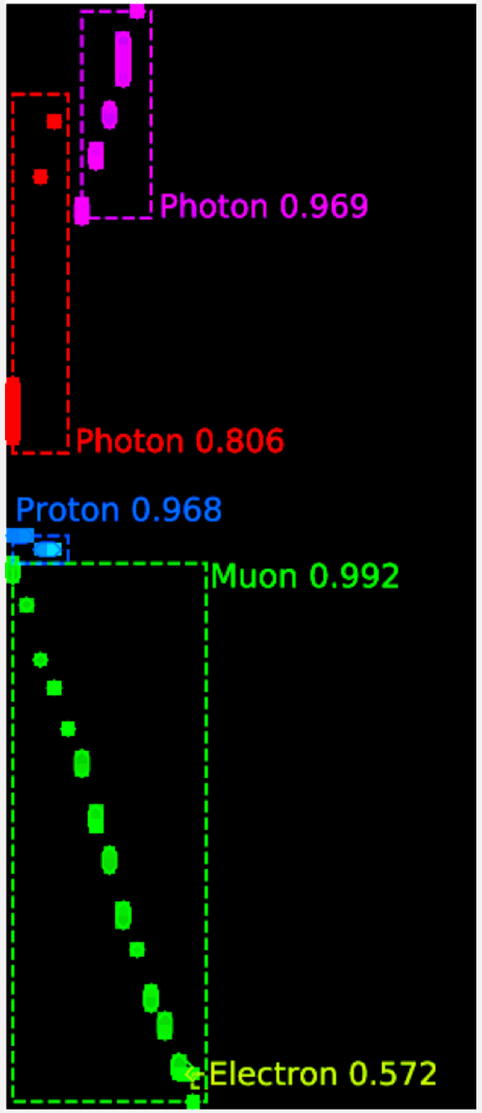
The network is trained on a flattened neutrino energy distribution which yields better control over systematics.





Future Work

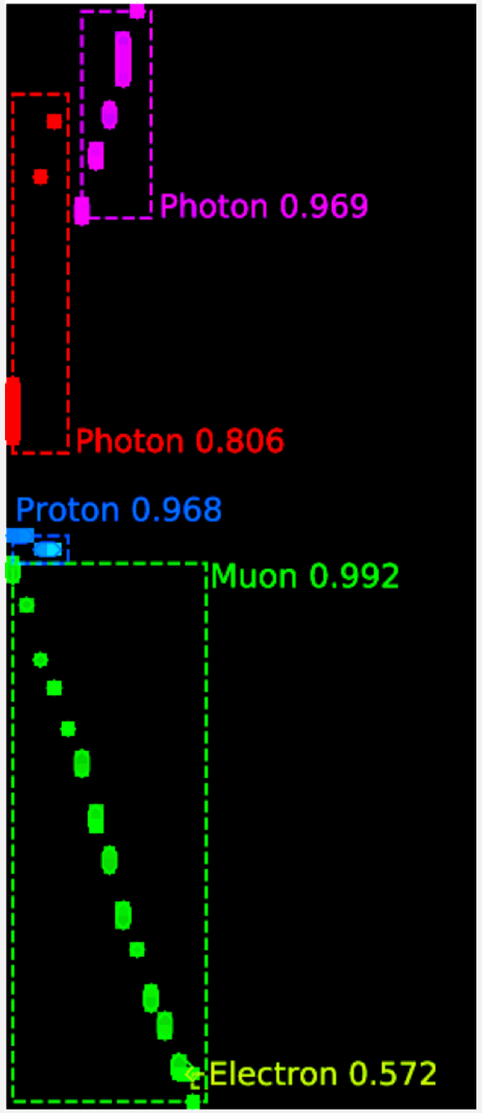
Instance Segmentation – Micah Groh



- Full event reconstruct on a hit-by-hit basis using instance segmentation.
- A bounding box is created around each particle, which are then identified, before individual hits are combined to form clusters.
- Incredibly powerful, but computing limitations threaten viability of running at scale.

Future Work

Instance Segmentation – Micah Groh



- Full event reconstruct on a hit-by-hit basis using instance segmentation.
- A bounding box is created around each particle, which are then identified, before individual hits are combined to form clusters.
- Incredibly powerful, but computing limitations threaten viability of running at scale.

Other Future Plans

There are a number of other projects which we are either pursuing, or will be soon;

- Sparse and Graphical Neural Networks.
- Improving our understanding of systematic effects, focusing on our light level and calibration models.
- Understanding generator biases, by exploring other generators such as NuWro, GIBUU, NEUT.

Summary

- NOvA continues to use and develop Machine Learning tools throughout its analysis program.
  - Some of the tools that we are developing will highlighted in the upcoming talks.

- New for the 2020 analyses;
  - A CNN to filter out cosmic events.
  - A move to Keras/Tensorflow
  - New architectures for our "Event and Prong CVNs."
  - New and improved validation tools.

Can you beat our Event Classification Network?

- NOvA ensures that our ML tools are robust by performing expansive cross-checks, improving our confidence in our analyses. This is something that all experiments should aim to do.

- Though NOvA makes extensive use of ML tools, it is certainly not plain sailing.

NOvA

MAY 2020

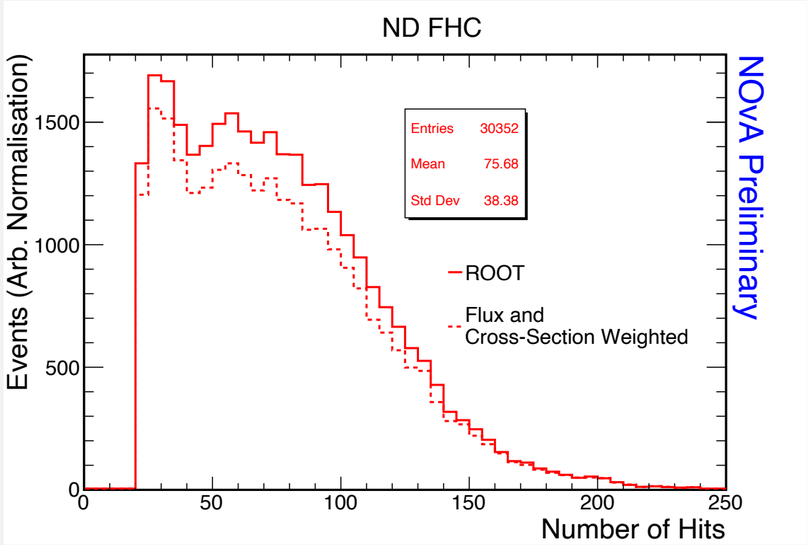
Validating a new analysis framework

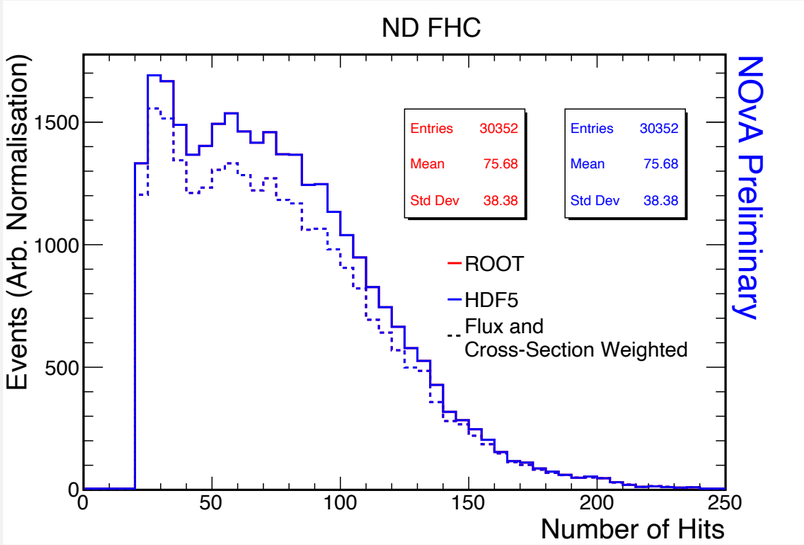
- To fully utilise HDF5s, we had to develop a new Python based analysis framework (*PandAna*). This would complement our ROOT based framework (*CAFAna*).
  - The main difference being switching from row based operations to column based operations (arbitrary sample example shown left).
- PandAna utilises pandas to do much of the heavy lifting.
- Also had to ensure that frameworks were consistent.

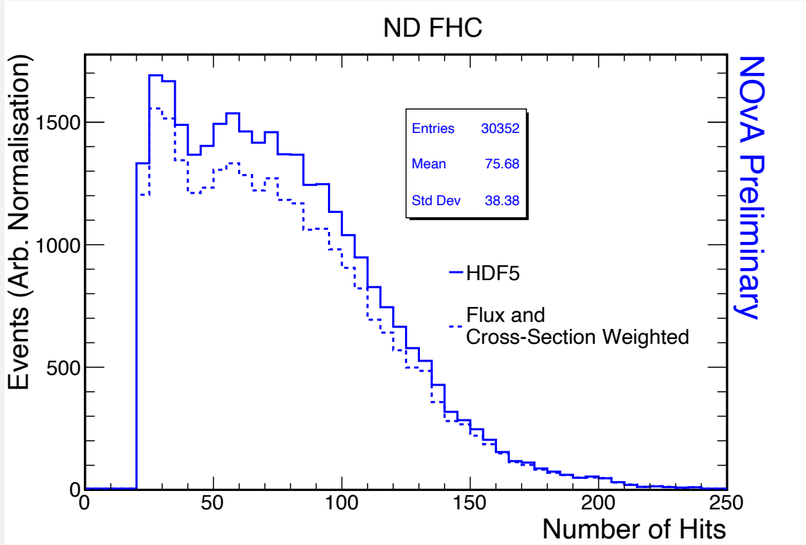
Event	NHit	Max X	Max Y	Max Z
1	75	5.4	5.6	23.9
2	54	3.9	4.8	54.8
3	102	4.5	5.9	12.7
4	63	5.7	2.4	43.6

PANDAna

CAFAna







Detailed MobileNetv2 Architecture for our Event Classification Net.

The input is the hits from the XZ and YZ views of the event.

The initial convolution builds a bank of initial features.

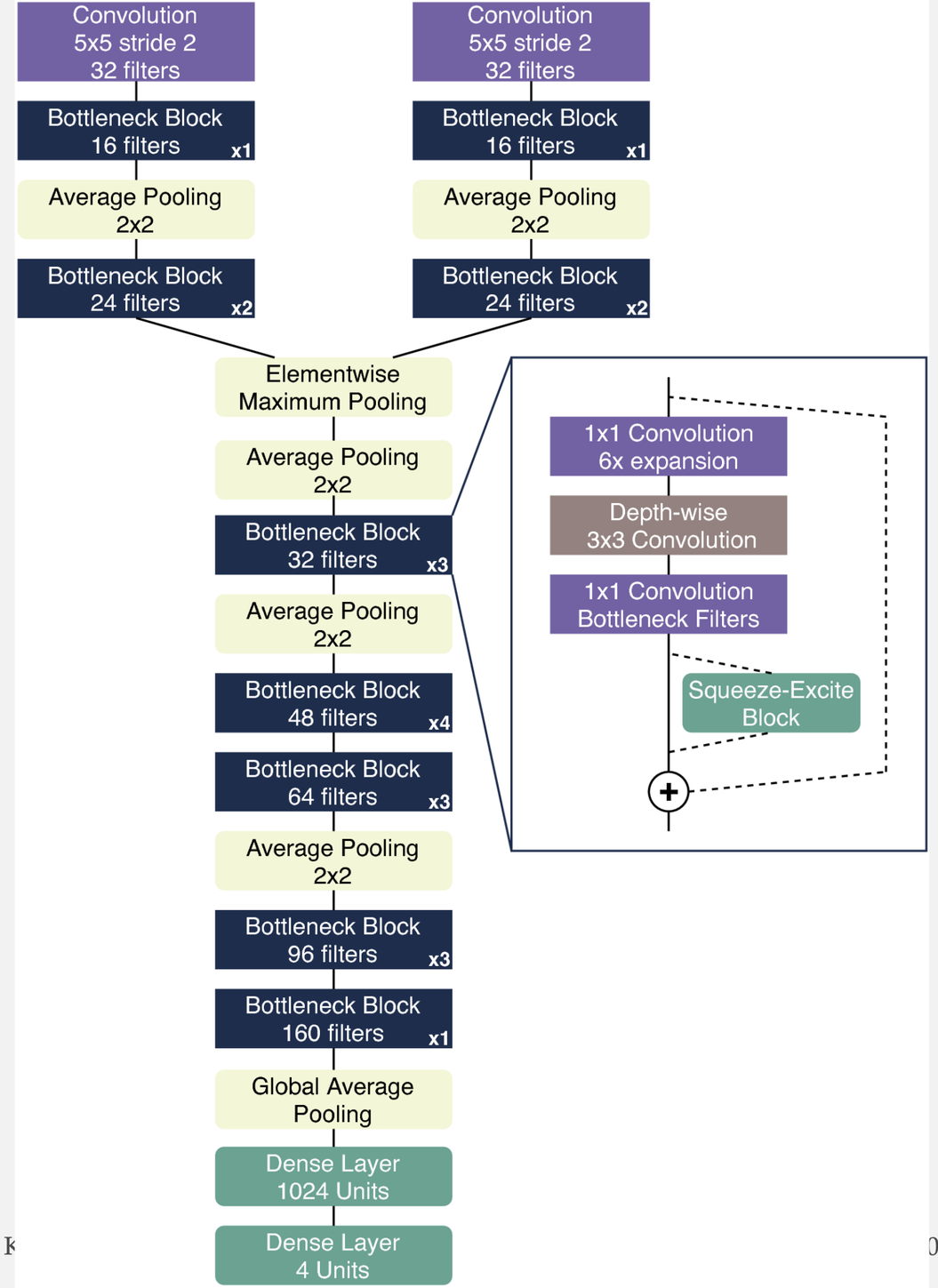
The main component of the architecture is the bottleneck block which expands the input features performs a depth-wise convolution then compresses the features once again.

The residual connection is used when the input and output shapes of the block are the same.

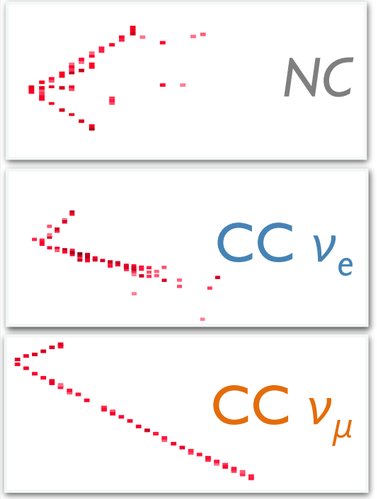
The extra squeeze-excite block is included in the last 7 bottleneck blocks.

Pooling layers are used throughout the model to reduce the feature dimensionality.

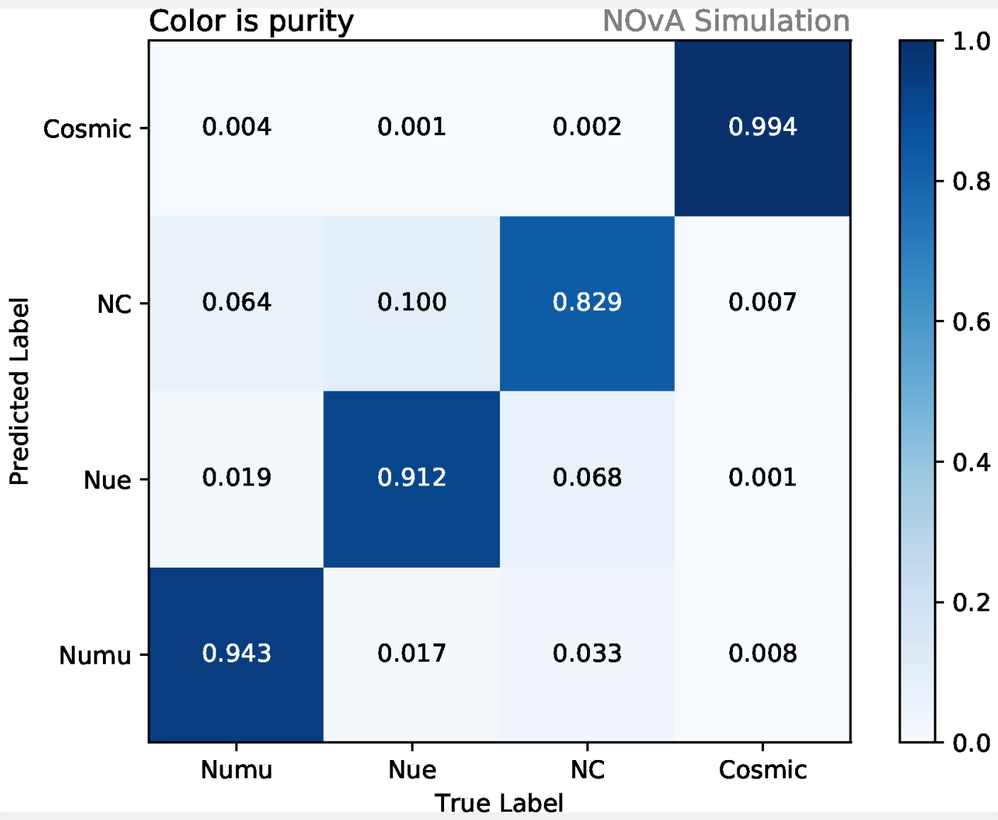
The output of the network is 4 scores used to classify the event as muon neutrino, electron neutrino, neutral current, or cosmic.



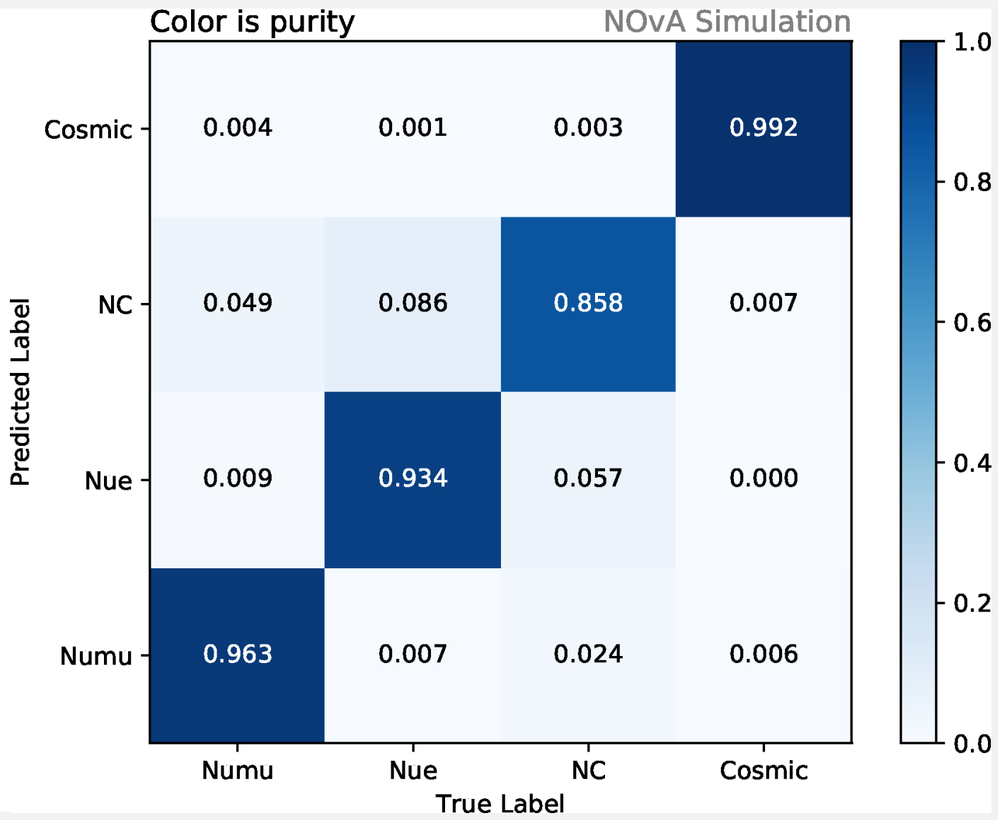
Event Classification Network Performance - Purity



Neutrino Mode



Anti-Neutrino Mode



Very similar performance between neutrino and anti-neutrino modes, though the anti-neutrino mode network appears to be more efficient.

See purities of over 99% for cosmics, and over 90% for all neutrino interaction channels for both horn polarities.

t-Distributed Stochastic Neighbour Embedding