Event Reconstruction with Deep Learning

Amir Farbin

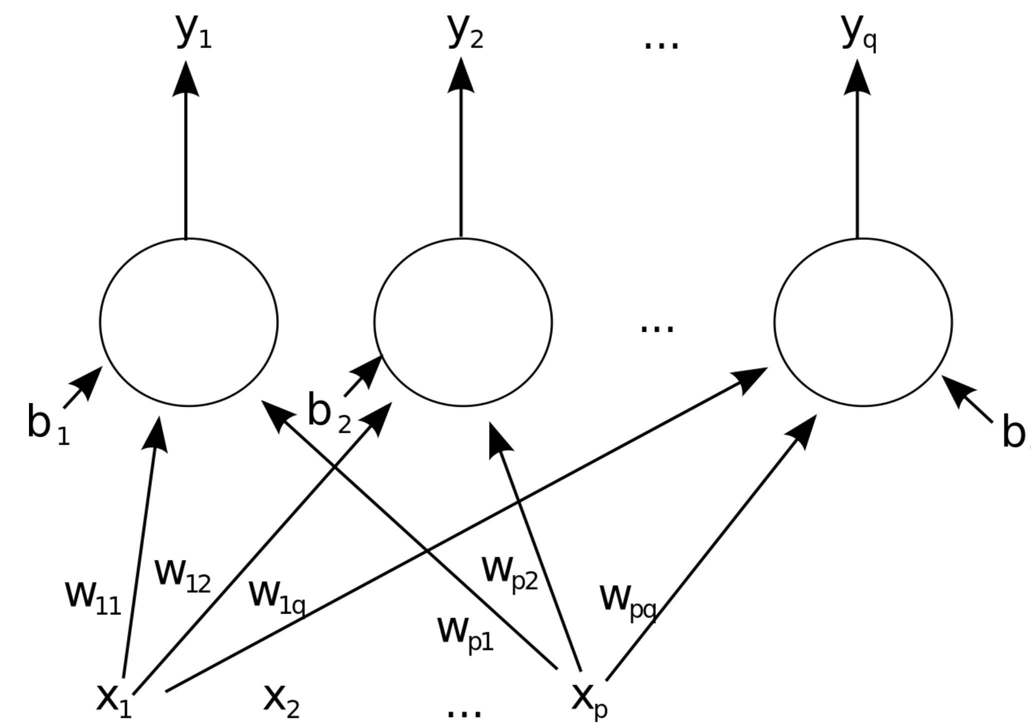
UNIVERSITY OF TEXAS ARLINGTON

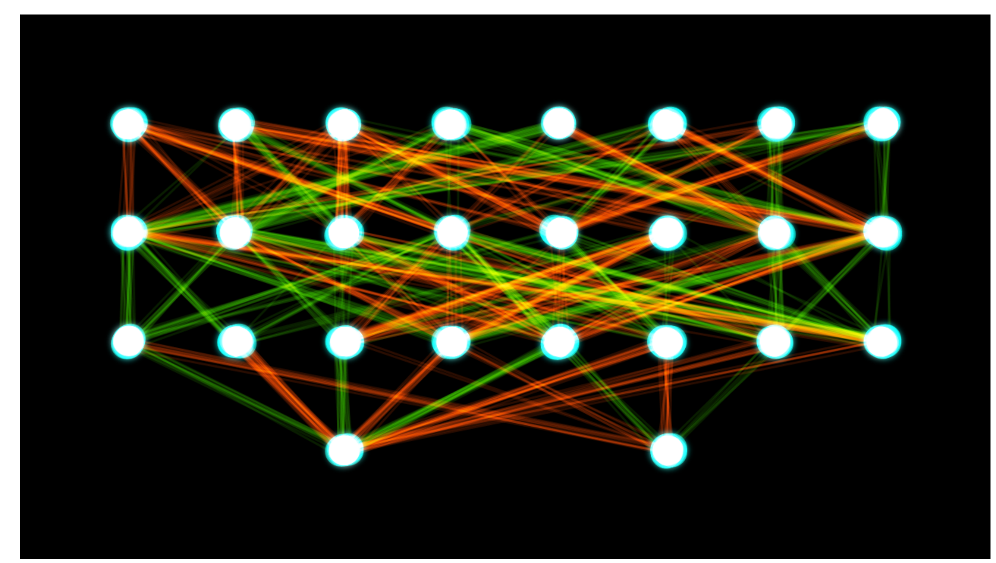
ATLAS EXPERIMENT

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Artificial Neural Networks

- ***Biologically inspired computation***, (first attempts in 1943)
  - ***Probabilistic Inference***: e.g. signal vs background
  - ***Universal Computation Theorem*** (1989)
- Common use in HEP, signal/background or particle ID with ***high-level features derived from raw data*** as input.
- Multi-layer (***Deep***) Neutral Networks:
  - Not a new idea (1965), just impractical to train. ***Vanishing Gradient problem*** (1991)
  - Solutions:
    - New techniques: e.g. better activation or layer-wise training
    - ***More training***: big training datasets and lots of computation … ***big data and GPUs***
  - ***Deep Learning Renaissance***. First DNN in HEP (2014).
  - ***Amazing Feats***: Audio/Image/Video recognition, captioning, and generation. Text (sentiment) analysis. Language Translation. Video game playing agents.
  - ***Rich field***: Variety of architectures, techniques, and applications.

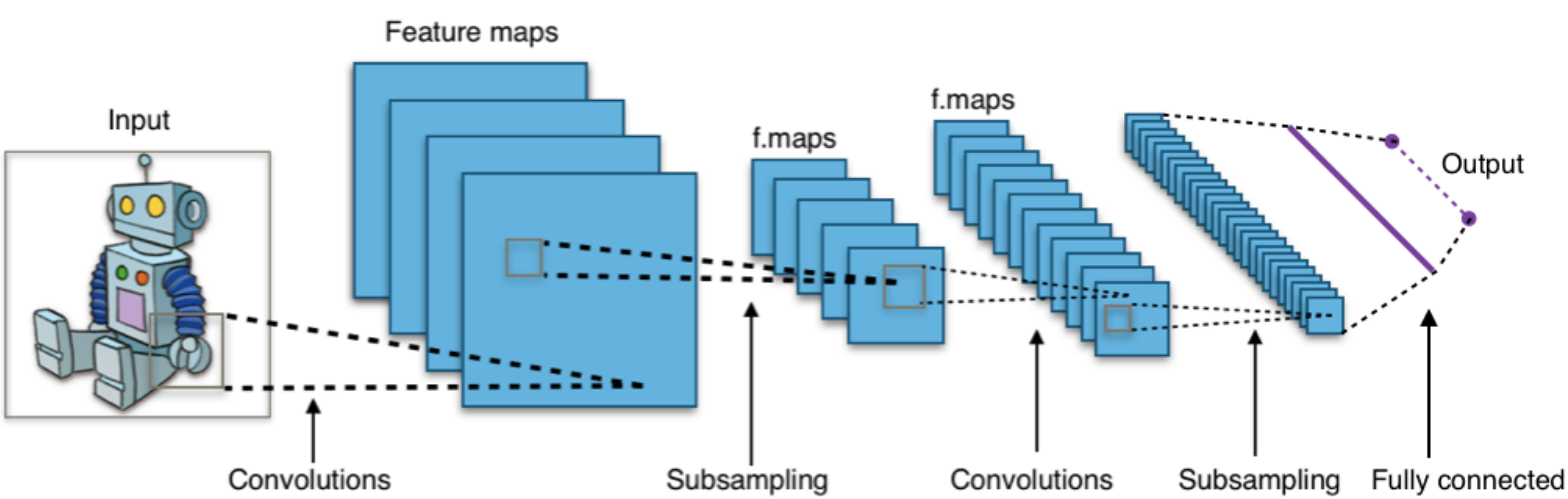


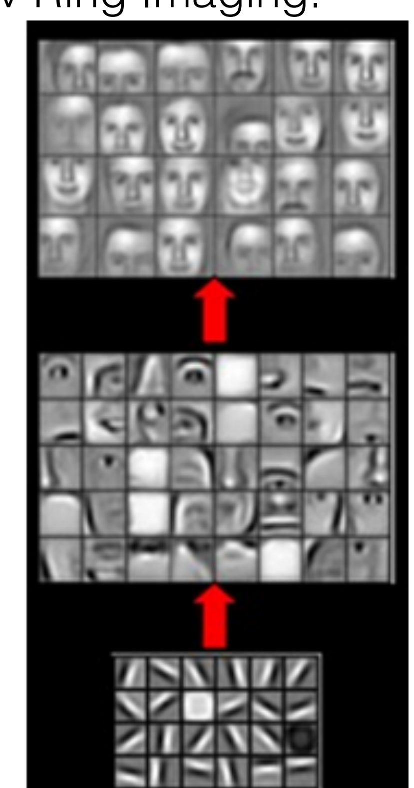


Images from Wikipedia

Feature Learning

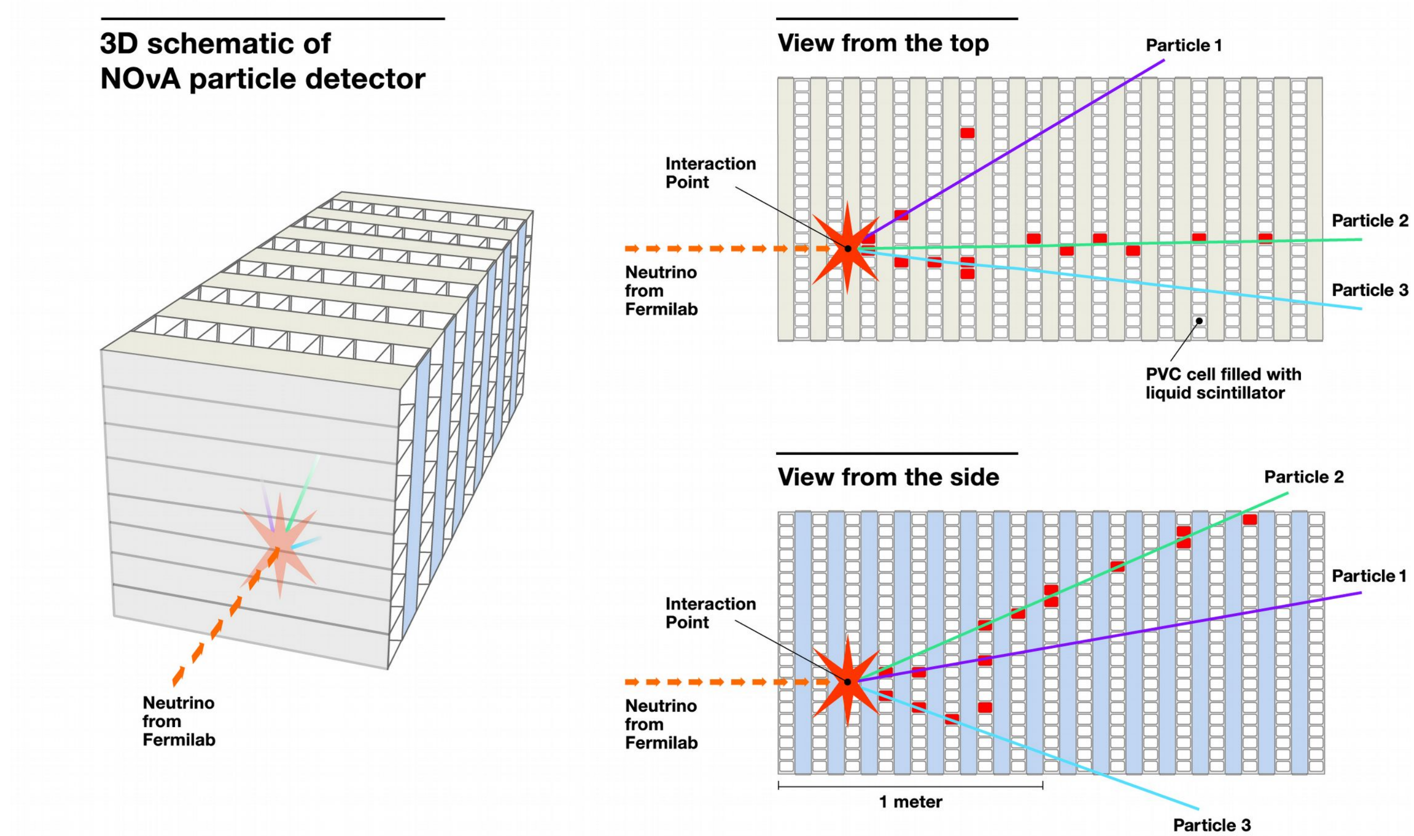
- ***Feature Engineering***: e.g. Event Reconstruction ~ Feature Extraction, Pattern Recognition, Fitting, …
- Deep Neutral Networks can ***Learn Features*** from ***raw data***.
- Example: ***Convolutional Neural Networks*** - Inspired by visual cortex
    - ***Input***: Raw data… for example 1D = Audio, 2D = Images, 3D = Video
    - ***Convolutions*** ~ learned feature detectors
    - ***Feature Maps***
    - ***Pooling*** - dimension reduction / invariance
    - ***Stack***: Deeper layers recognize higher level concepts.
- Over the past few years, CNNs have lead to ***exponential improvement / superhuman performance on Image classification*** challenges. Current best > 150 layers.
- Obvious HEP application: ***"Imaging" Detectors*** such as TPCs, High Granularity Calorimeters, or Cherenkov Ring Imaging.





Neutrino Physics

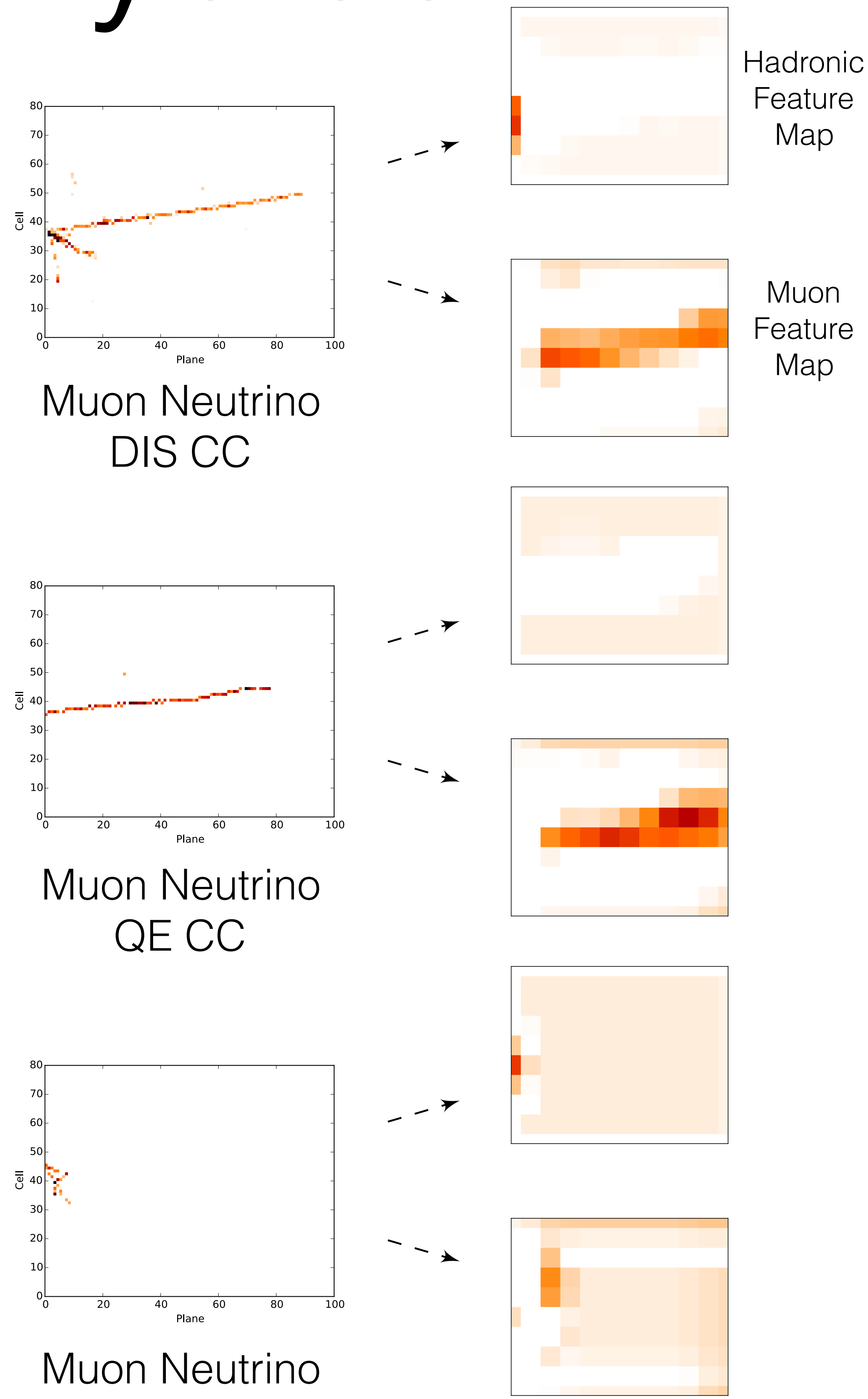
- Core Physics requires just measuring ***neutrino flavor and energy***.
- Generally clean (low multiplicity) and high granularity.
- ***First HEP CNN application: <u>Nova</u>*** using Siamese Inception CNN.



	CVN Selection Value	ν_e sig	Tot bkg	NC	ν_μ CC	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	88.4	509.0	344.8	132.1	32.1	–	14.8%
$s/\sqrt{b}$ opt	0.94	43.4	6.7	2.1	0.4	4.3	49.1%	86.6%
$s/\sqrt{s+b}$ opt	0.72	58.8	18.6	10.3	2.1	6.1	66.4%	76.0%

	CVN Selection Value	ν_μ sig	Tot bkg	NC	Appeared ν_e	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	355.5	1269.8	1099.7	135.7	34.4	–	21.9%
$s/\sqrt{b}$ opt	0.99	61.8	0.1	0.1	0.0	0.0	17.4%	99.9%
$s/\sqrt{s+b}$ opt	0.45	206.8	7.6	6.8	0.7	0.1	58.2%	96.4%

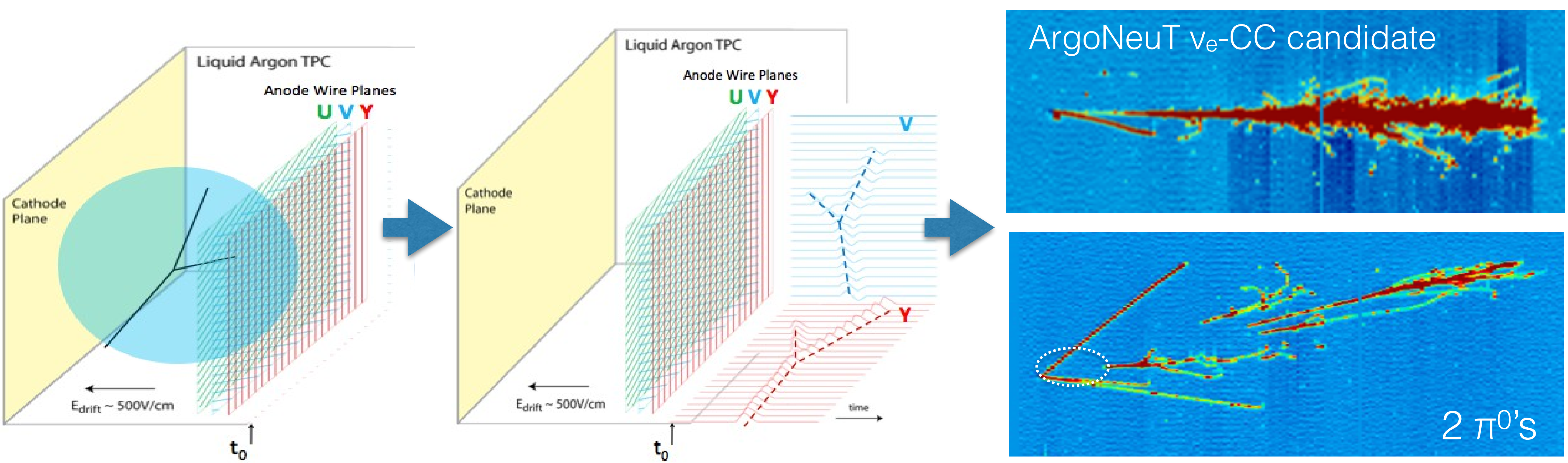
- ***40% Better Electron Efficiency for same background.***



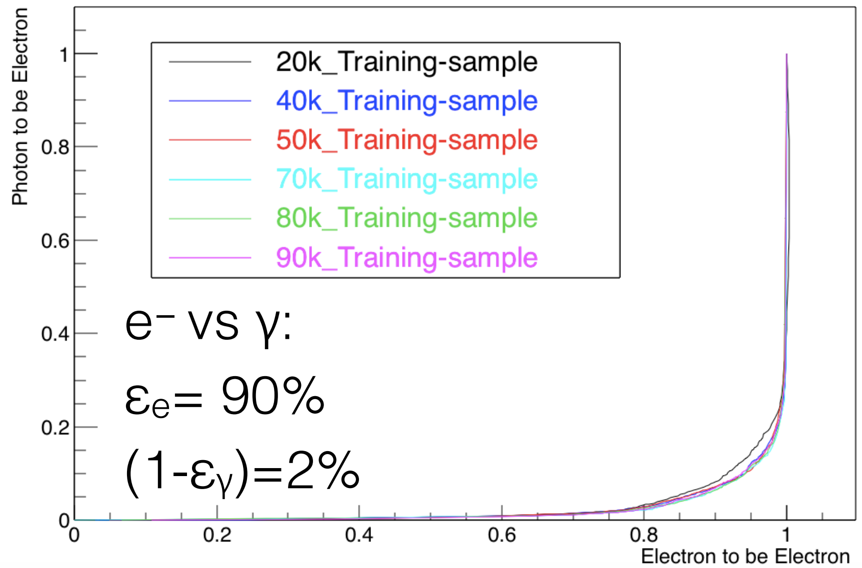
http://arxiv.org/pdf/1604.01444.pdf

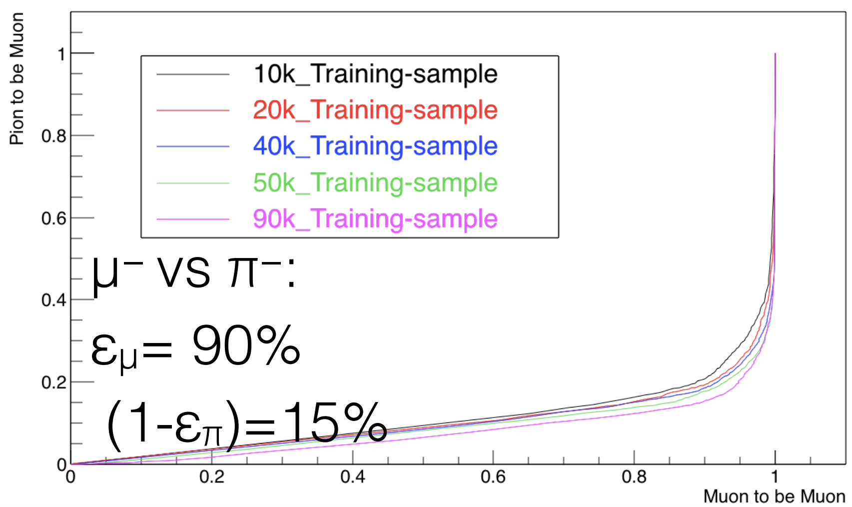
LArTPC

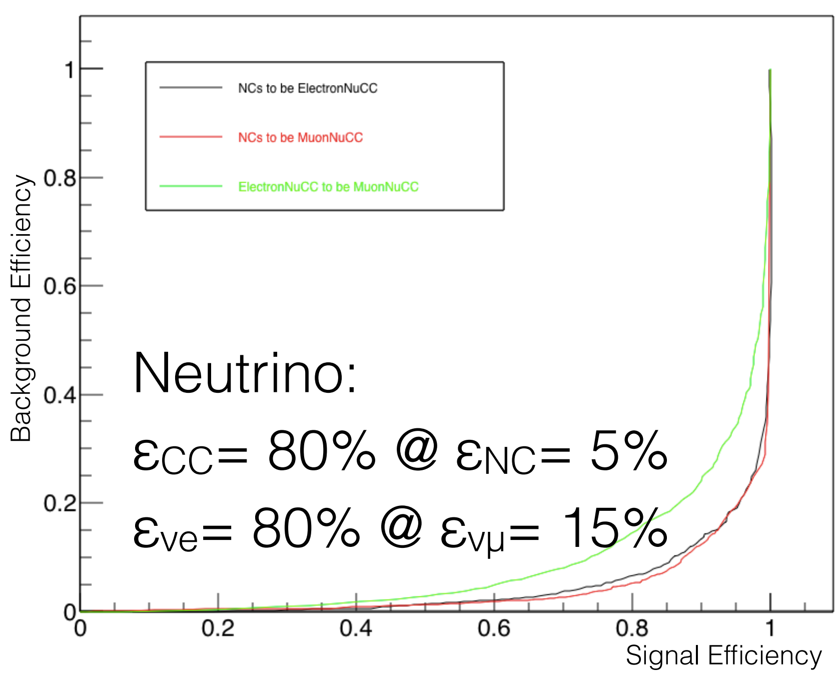
- ***Liquid Argon Time Projection Chambers*** are the chosen technology for ***US's flagship Short and Long Base Line (i.e. LBNF/DUNE) programs***.
    - Tracking, Calorimetry, and Particle ID in same detector.
    - Goal ~80% Neutrino Efficiency. (~ 2x better than current).
- ***Fully automatic reconstruction*** has been challenging and not yet demonstrated with required performance.
    - ICARUS relied on hand scans for pattern recognition. "User assisted reconstruction".
    - Human eye does better than engineered reconstruction.
- ***Ideally suited*** for Convolutional Neutral Networks. Efforts in LArIAT, DUNE, and MicroBooNE.



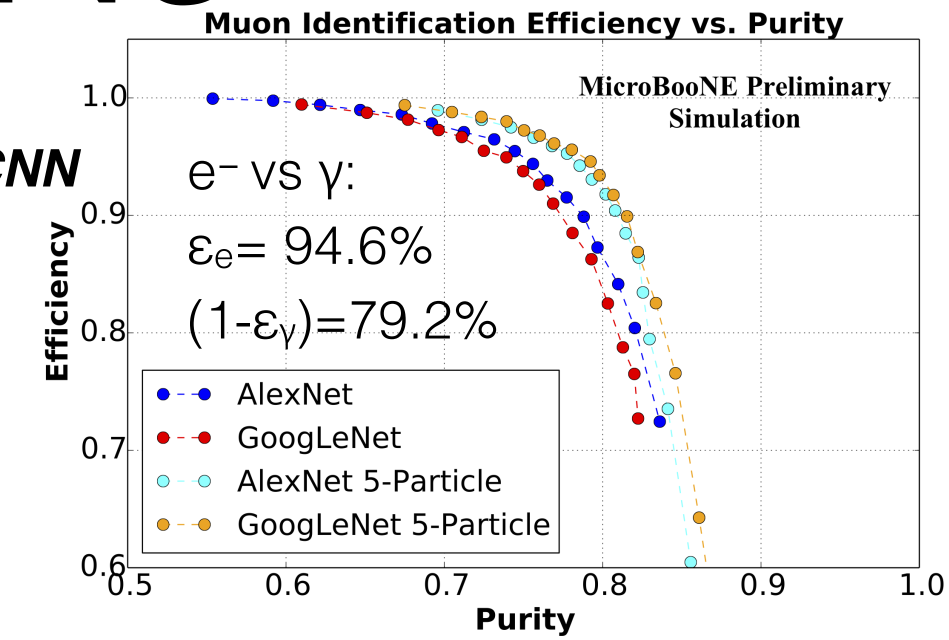
LArTPC CNNs

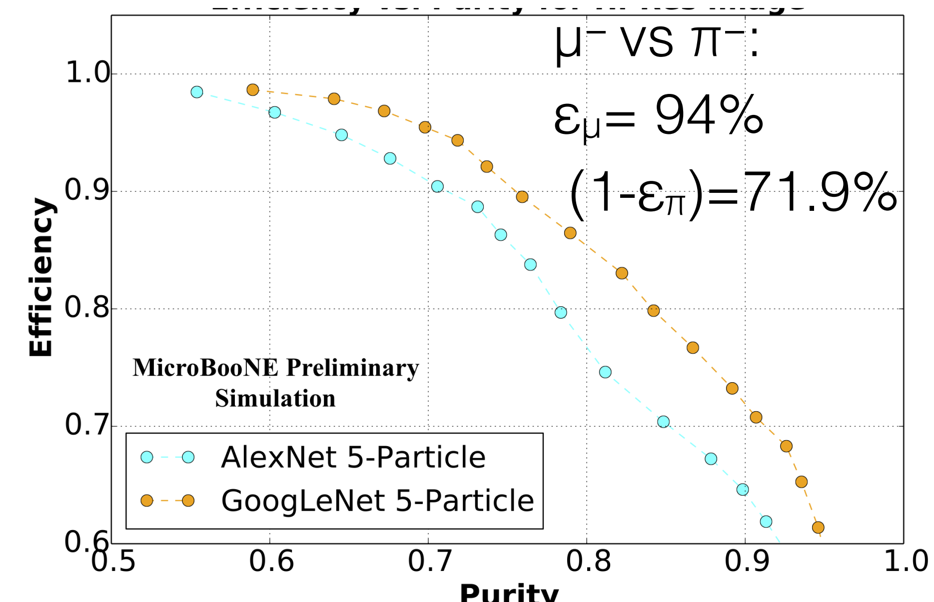


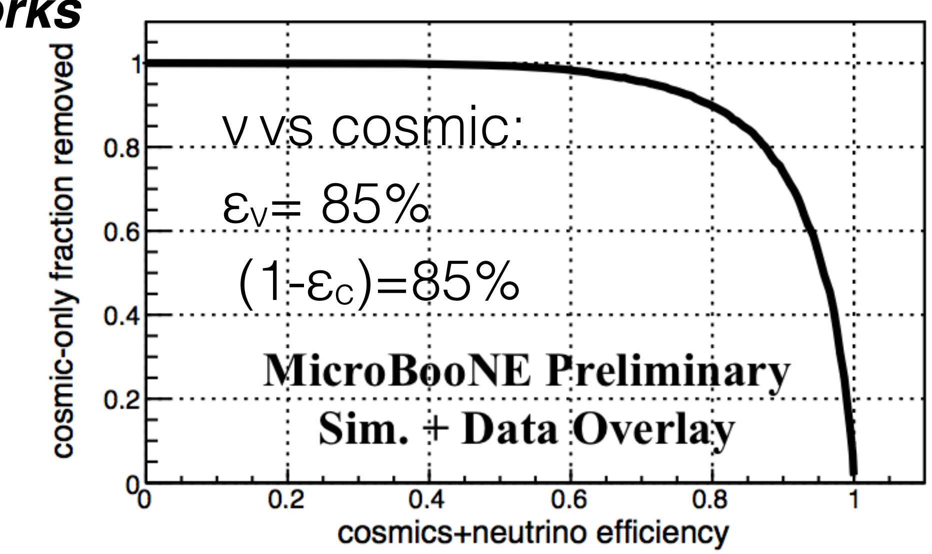




- First Studies Based on ***out-of-box CNN***
- ***MicroBooNE*** Note and ***LArIAT*** (independent by AF):
  - ***Particle ID***: MicroBooNE and LArIAT
  - ***Neutrino ID***: LArIAT
  - ***Cosmic Rejection***: MicroBooNE
  - ***Energy Regression***: LArIAT (in progress)
- Observations:
  - LArTPC easier than image classification. ***Shallower networks sufficient***.
  - Higher resolution is better
  - Larger training sample is better
  - More readout planes are better

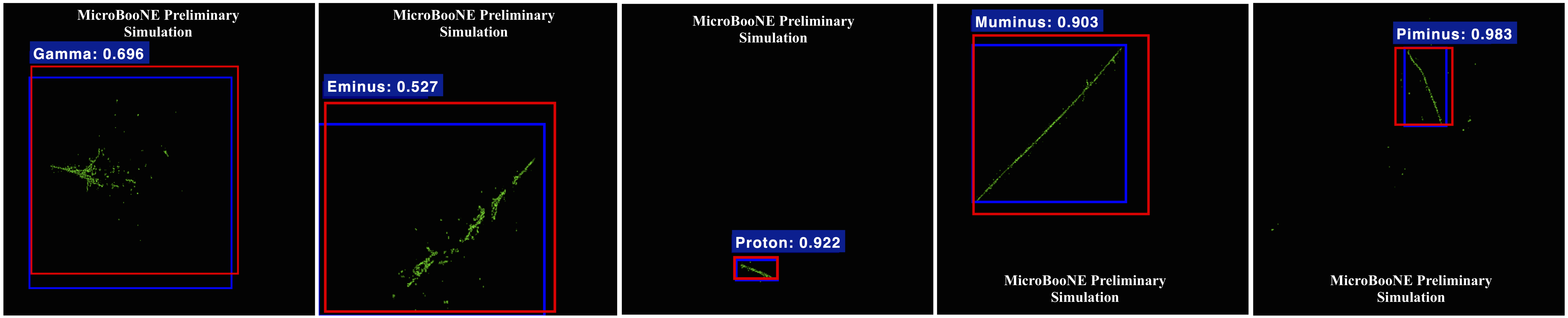


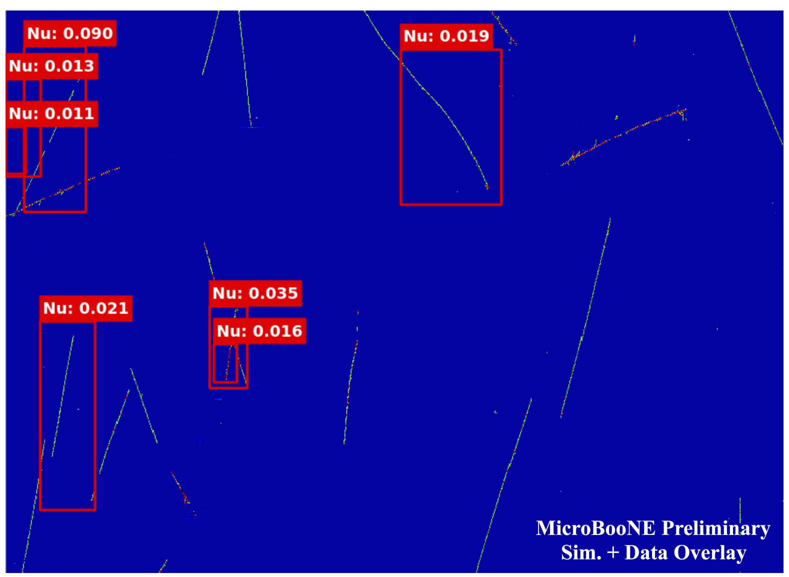




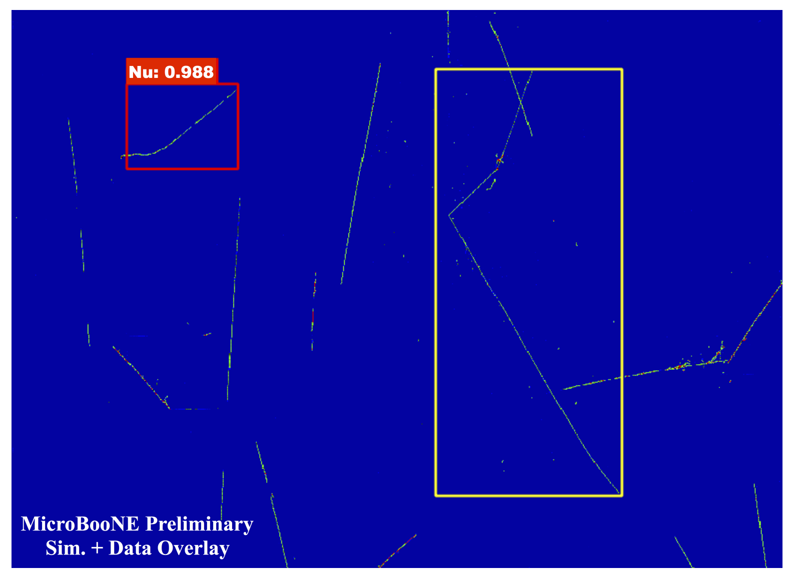
Semantic Segmentation

MicroBooNE

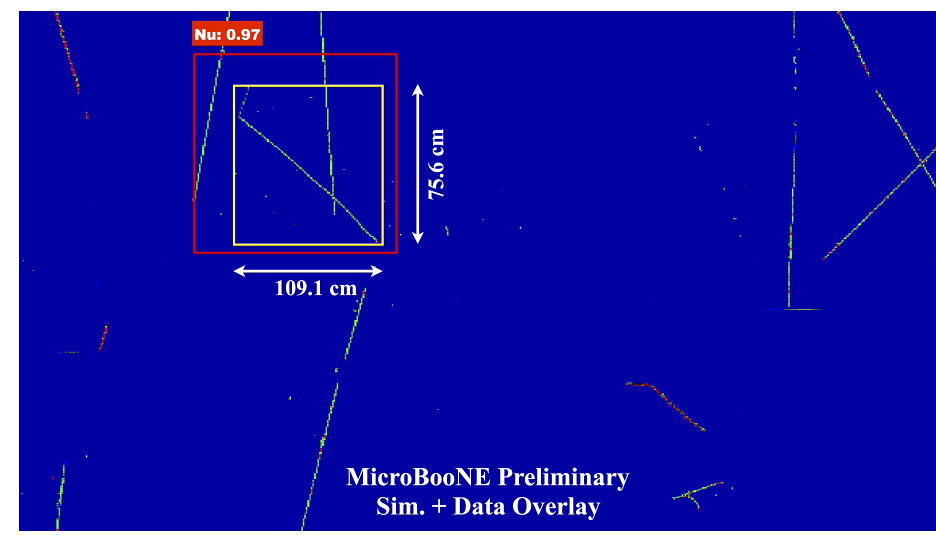




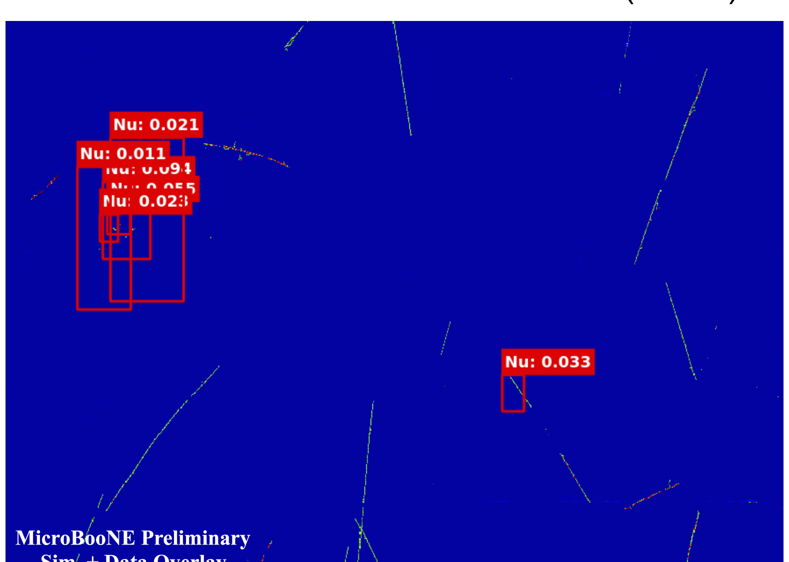
Low score for real Cosmics (data)

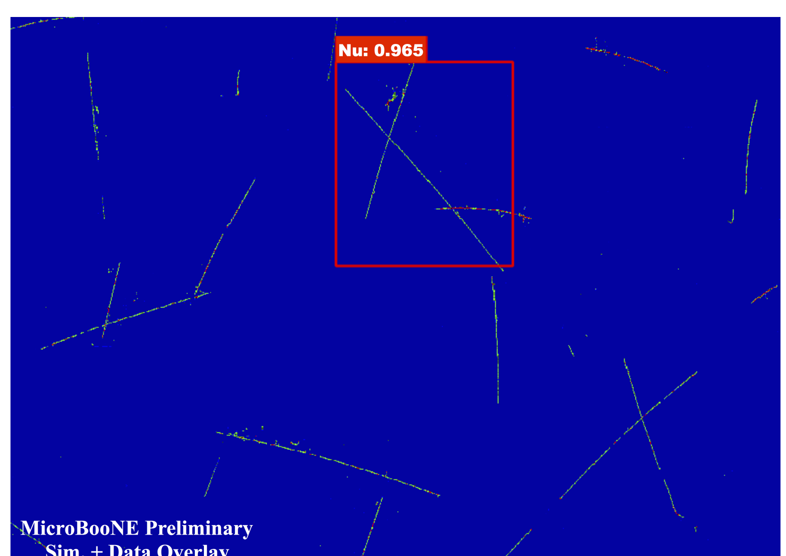


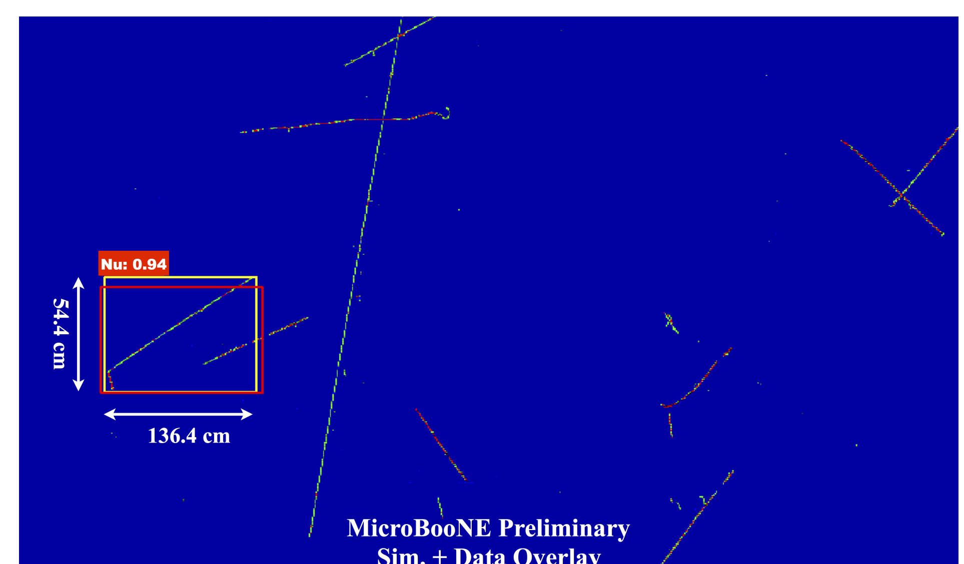
Reasonable Mistakes?



High Score for overlaid neutrinos (sim)

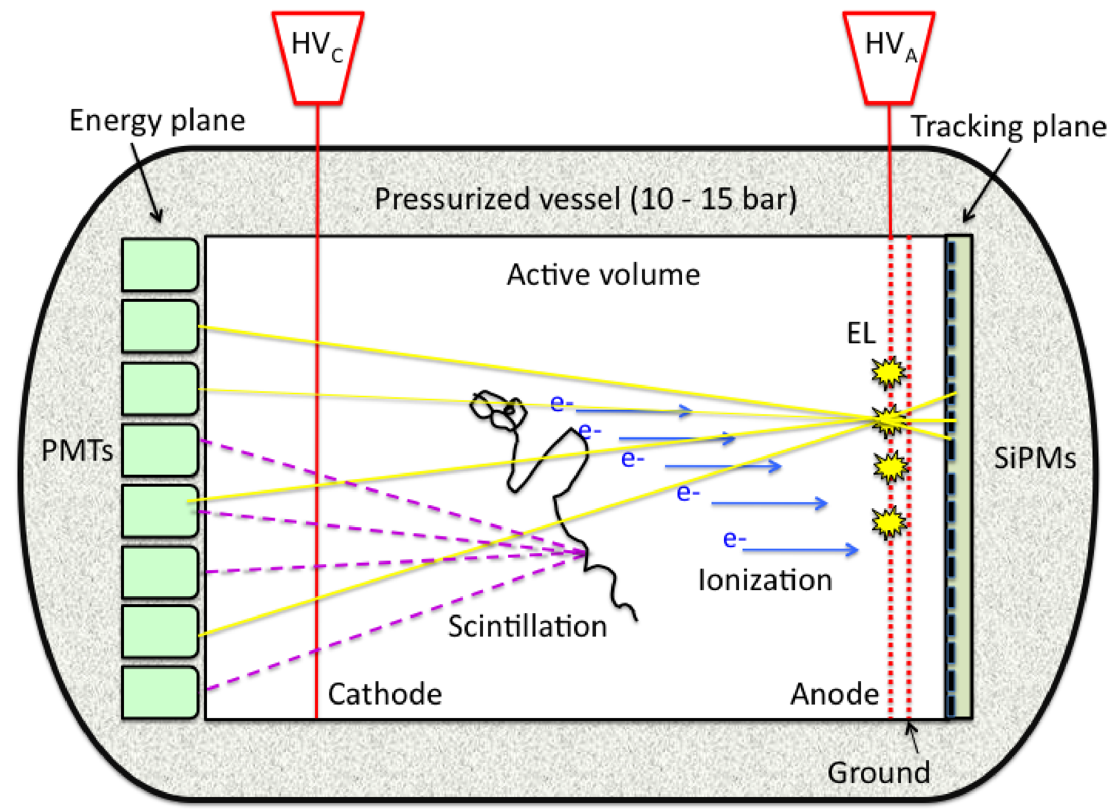


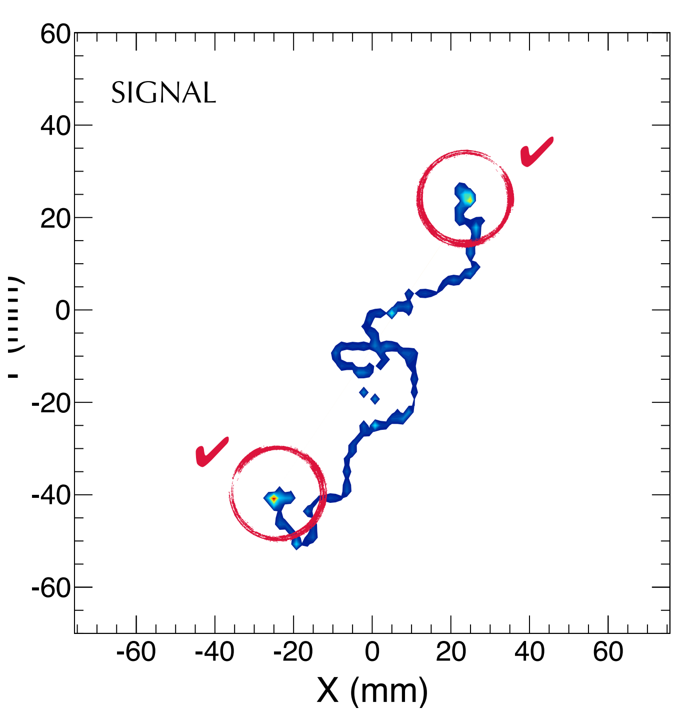


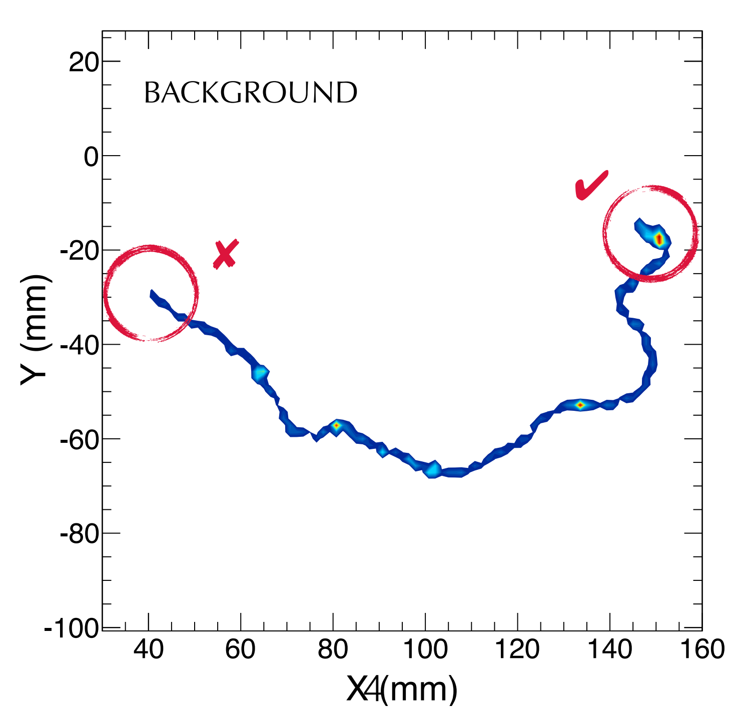


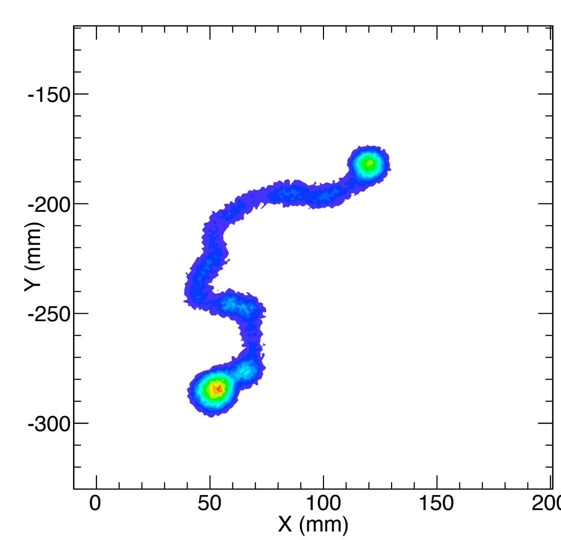
NEXT Experiment

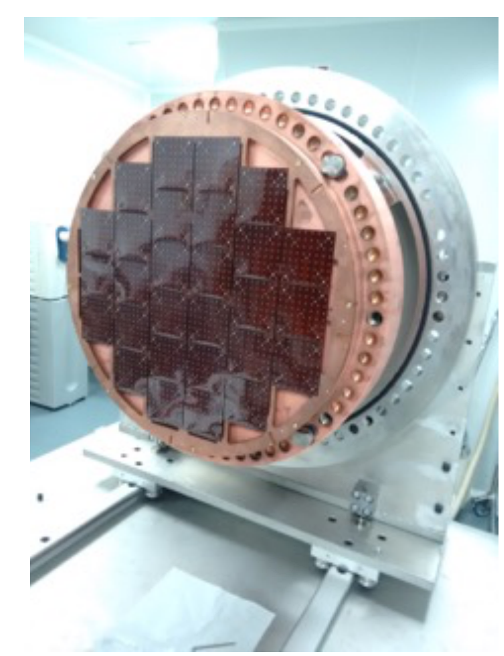
- ***Neutrinoless Double Beta Decay*** using Gas TPC/SiPMs
- Signal: 2 Electrons. Bkg: 1 Electron.
- Hard to distinguish due to ***multiple scattering***.
- ***3D readout***… candidate for 3D Conv Nets.
- Just a handful of signal events will lead to ***noble prize***
    - Can we trust a DNN at this level?







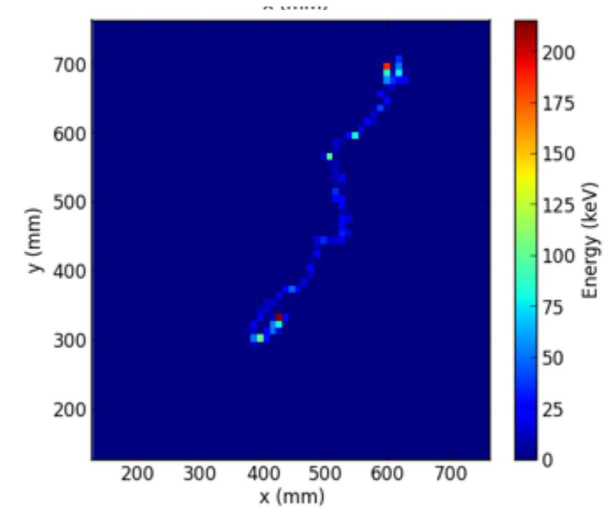


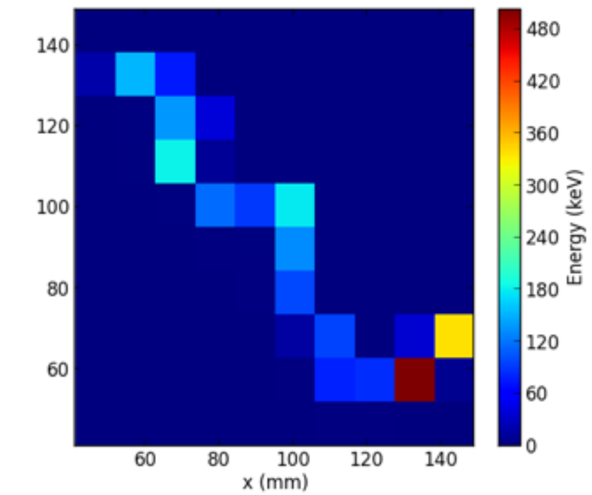


NEXT Detector Optimization

- Idea 1: use DNNs to ***optimize detector***.
    - Simulate data at different resolutions
    - Use DNN to quickly/easily assess best performance for given resolution.

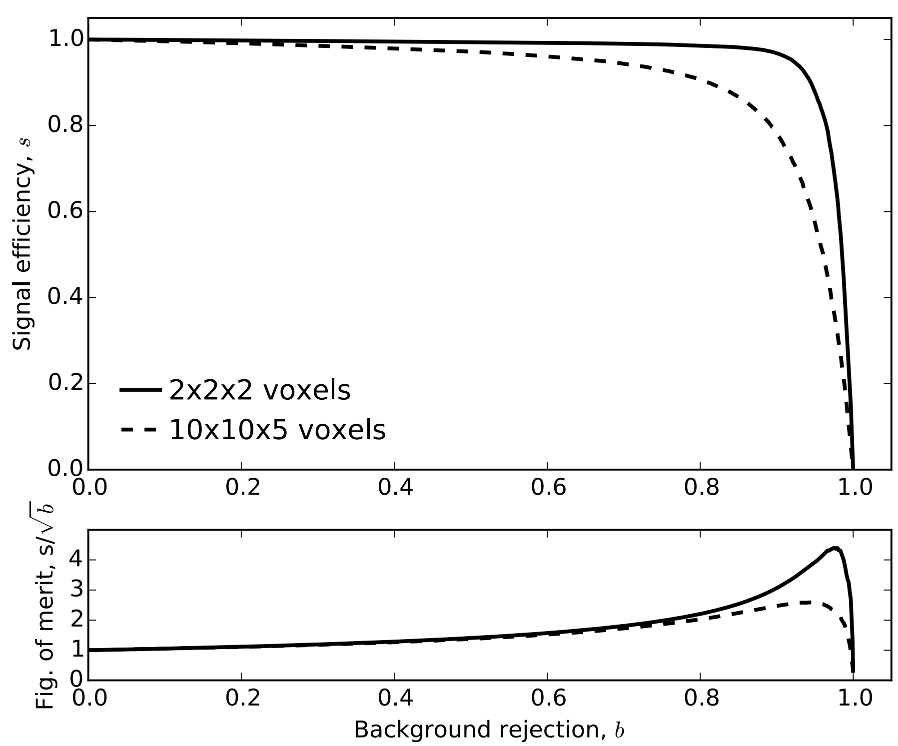
Analysis	Signal eff. (%)	B.G. accepted (%)
DNN analysis (2 x 2 x 2 voxels)	86.2	4.7
Conventional analysis (2 x 2 x 2 voxels)	86.2	7.6
DNN analysis (10 x 10 x 5 voxels)	76.6	9.4
Conventional analysis (10 x 10 x 5 voxels)	76.6	11.0





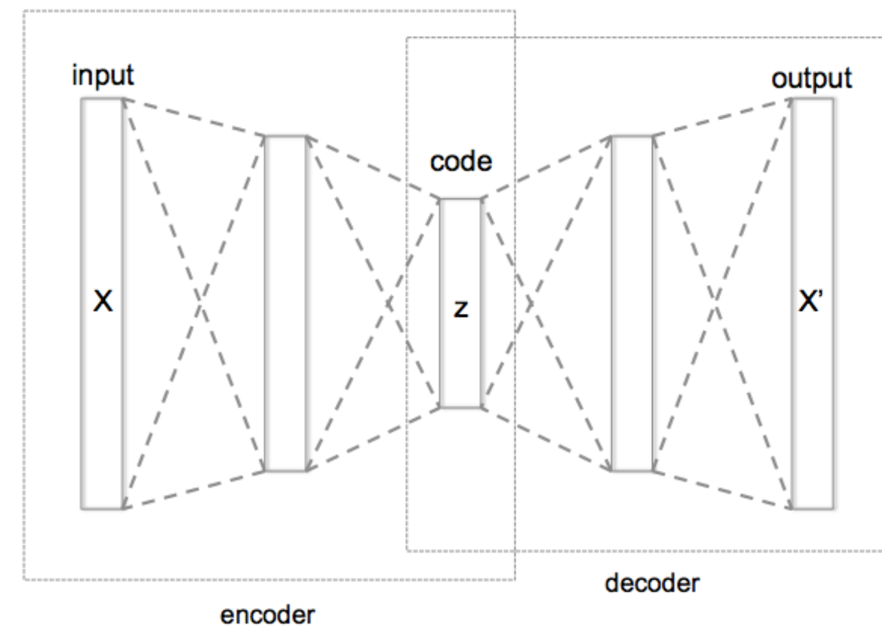
- Idea 2: ***systematically study*** the relative importance of various physics/detector effects.
    - Start with simplified simulation. Use DNN to assess performance.
    - Turn on effects one-by-one.

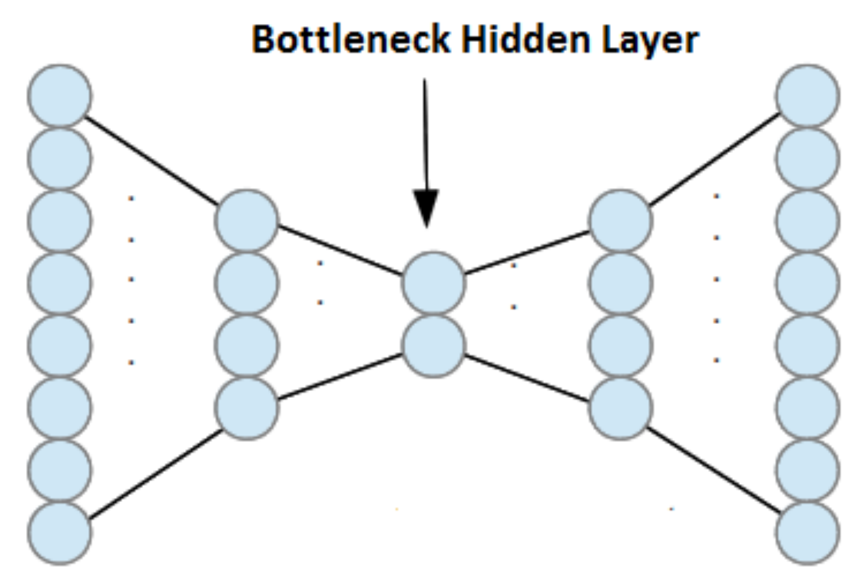
2x2x2 voxels	Run description	Avg. accuracy (%)
	Toy MC, ideal	99.8
	Toy MC, realistic $0\nu\beta\beta$ distribution	98.9
	Xe box GEANT4, no secondaries, no E-fluctuations	98.3
	Xe box GEANT4, no secondaries, no E-fluctuations, no brem.	98.3
	Toy MC, realistic $0\nu\beta\beta$ distribution, double multiple scattering	97.8
	Xe box GEANT4, no secondaries	94.6
	Xe box GEANT4, no E-fluctuations	93.0
	Xe box, no brem.	92.4
	Xe box, all physics	92.1
	NEXT-100 GEANT4	91.6
10x10x5 voxels		
	NEXT-100 GEANT4	84.5



Semi-supervised Learning

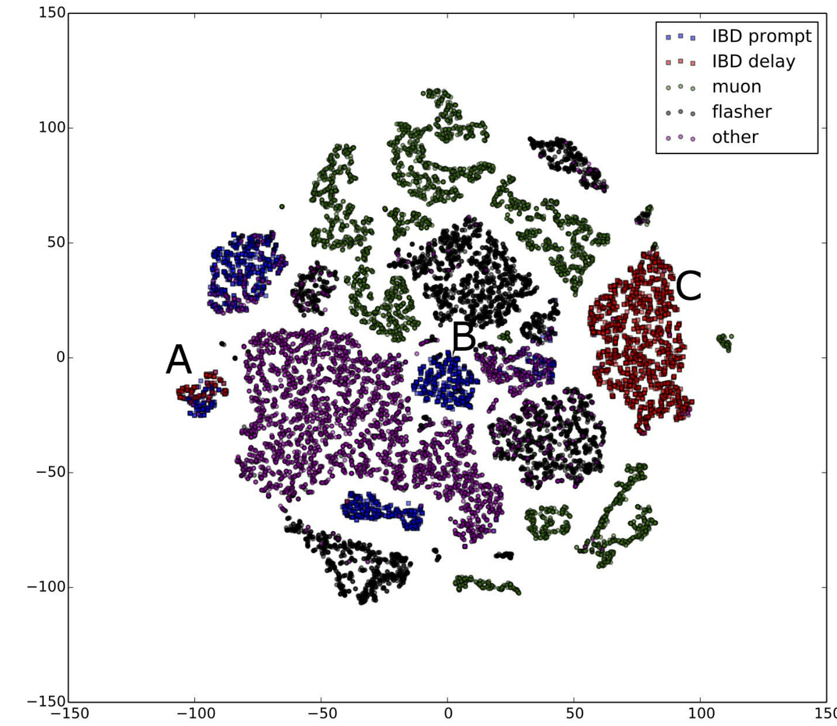
- Basic idea: Train network to ***reproduce the input***.
- Example: ***Auto-encoders***
  - ***De-noising auto-encoders***: add noise to input only.
  - ***Sparse auto-encoders***:
    - ***Sparse latent (code) representation*** can be exploited for ***Compression, Clustering, Similarity testing,*** …
  - ***Anomaly Detection***
    - Reconstruction Error
    - Outliers in latent space
  - ***Transfer Learning***
    - Small labeled training sample?
      - Train auto-encoder on large unlabeled dataset (e.g. data).
      - Train in latent space on small labeled data. (e.g. rare signal MC).
- Not hard to imagine HEP applications.



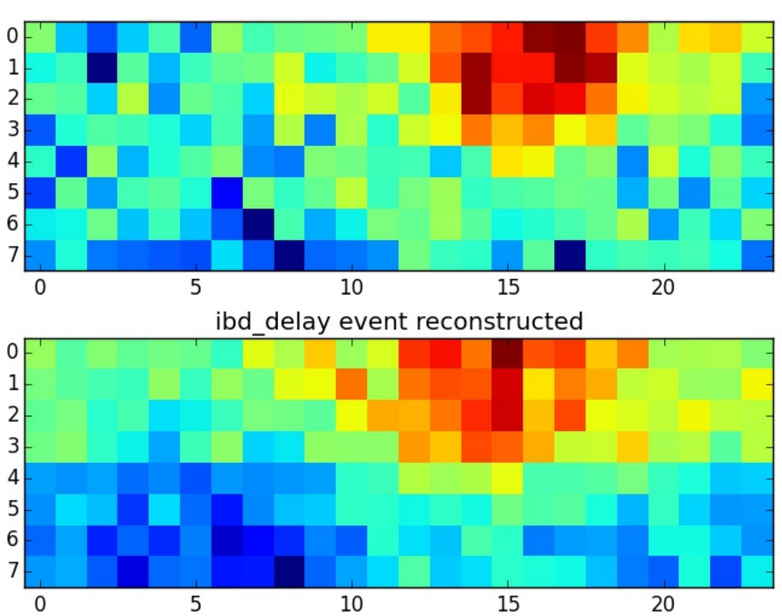


Learning Representations

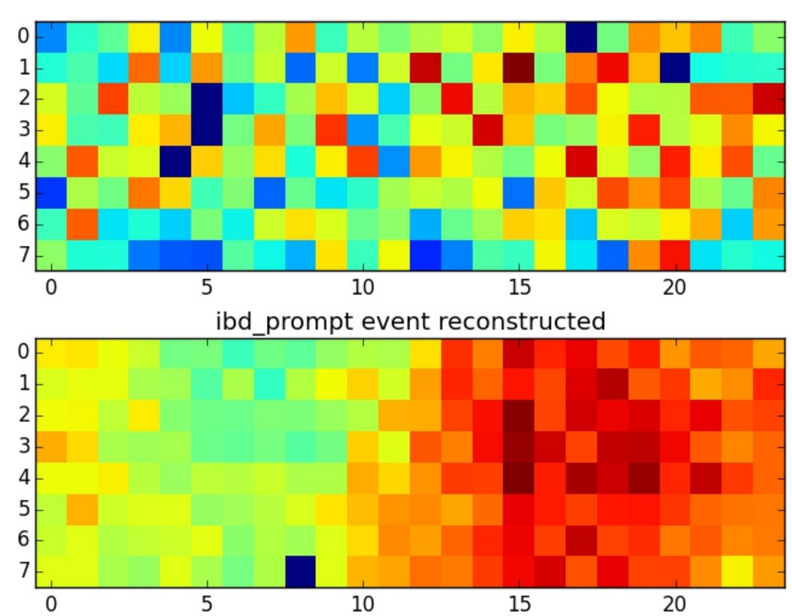
- Example: ***Daya Bay Experiment*** (*Evan Racah, et al*)
- Input: 8 x 24 PMT unrolled cylinder. ***Real Data (no simulation)***
- 2 Studies:
  - ***Supervised CNN Classifier***
    - Labels from standard analysis: Prompt/Delayed Inverse Beta Decay, Muon, Flasher, Other.
  - ***Convolutional Auto-encoder*** (semi-supervised)
    - Clearly separates muon and IBD delay ***without any physics knowledge***.
    - Potentially could have ID'ed problematic data (e.g. flashers) much earlier.



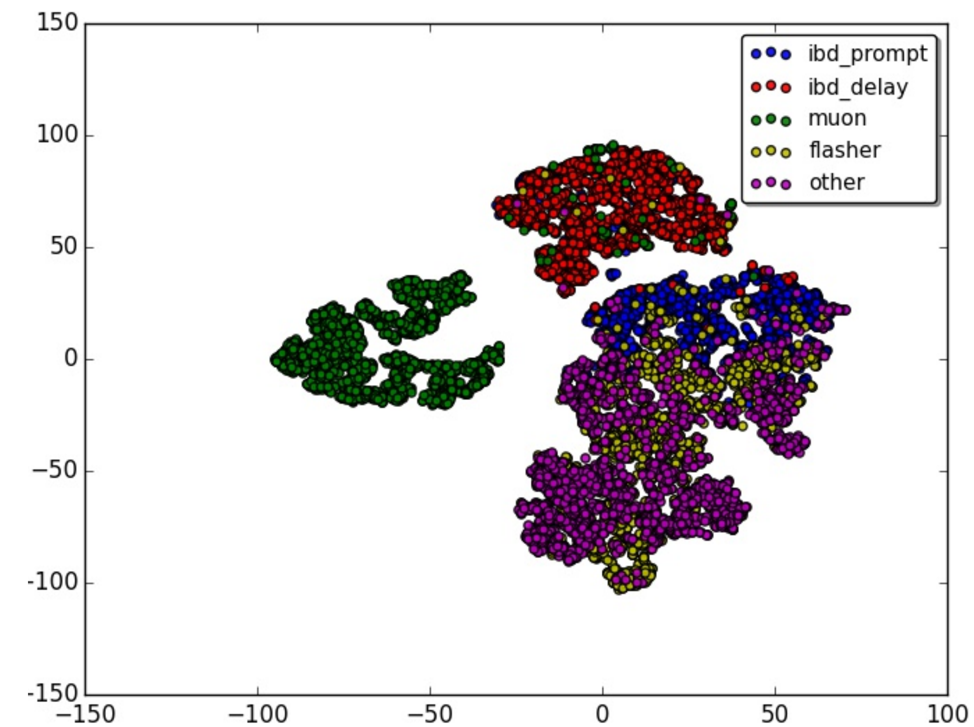
t-SNE reduction of 26-dim representation of the last fully connected layer.



(a) Example of an "IBD delay" event



(b) Example of an "IBD prompt" event



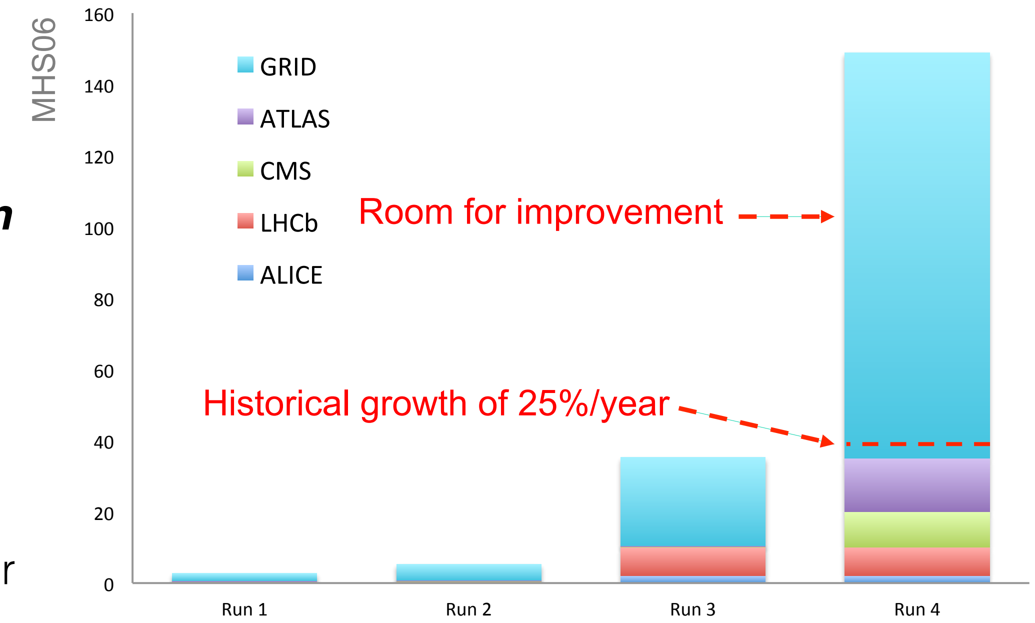
t-SNE reduction of 10 parameter latent representation.

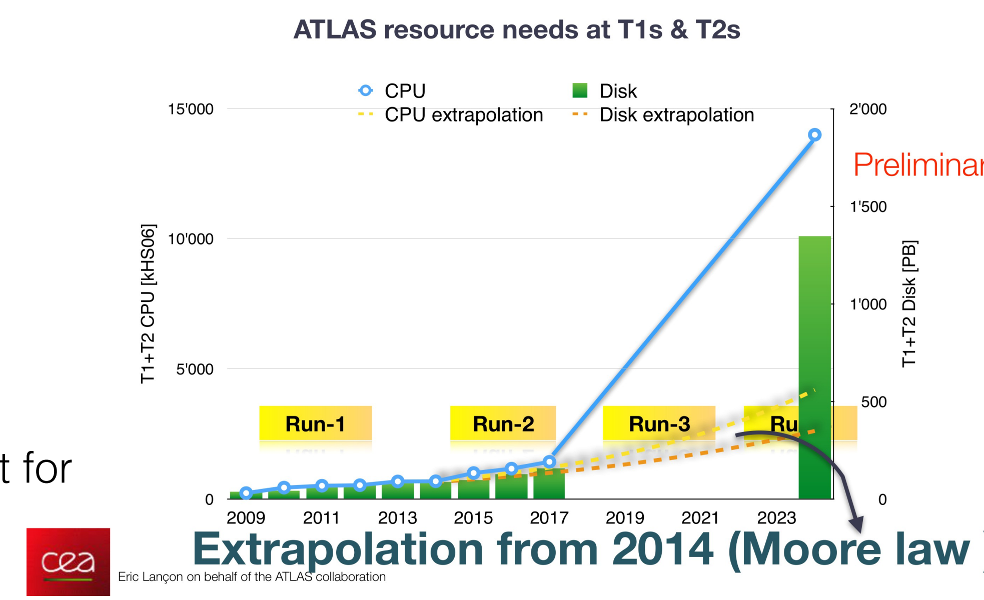
DL on Raw Data

- Recap:
  - ***Improved classification/regression*** with Convolutional NNs.
    - For LArTPC, DNNs may be able to do something we cannot do well algorithmically.
    - Experiments like NEXT may not need to write reconstruction software.
  - ***Detector Optimization and Systematic Studies***.
    - DL provides easily obtainable, consistent, and optimal metrics.
    - Just simulate… no need to tune reconstruction to every possibility.
  - ***Unsupervised techniques***: Identify features in data w/o simulation.
- Other reco applications:
  - Studying DNNs to ***identify unknown features*** that can be exploited in normal reconstruction.
  - Auto-encoders: ***anomaly detection (monitoring), compression, de-noising***, …
  - ***Generative models***: Build model from training data that can generate new data.
    - e.g. Fast showers.
    - Build simulations purely based on data.

HL-LHC

- ***Computing*** is perhaps the biggest challenge for the HL-LHC
    - ***Higher Granularity*** = larger events.
    - ***O(200) proton collision / crossing***.
        - ***High occupancy:*** untenable tracking due to ***tracking pattern recognition combinatorics***.
        - Trigger may require ***Interaction Vertex reconstruction at 40 MHz***.
    - O(100) times = data multi ***exabyte datasets***.
    - ***Moore's law has stalled***: Cost of adding more transistors/silicon area no longer decreasing.
    - Preliminary estimates of ***HL-LHC computing budget many times larger than LHC***.
- ***Solutions***:
    - ***Leverage opportunistic resources and HPC*** (most computation power in highly parallel processors).
    - ***Highly parallel processors*** (e.g. GPUs) are already > 10x CPUs for certain computations.
        - Trend is away from x86 towards ***specialized hardware*** (e.g. GPUs, Mics, FPGAs, Custom DL Chips)
        - Unfortunately parallelization (i.e. Multi-core/GPU) has been extremely difficult for HEP.
    - ***Deep Learning and Neuromorphic processors*** are a promising solution.





Plots from here.

Public Datasets

- ***Neutrino world*** has provided ***ideal playground*** for initial DL-based Reconstruction investigations.
- ***Biggest obstacles*** to DNN research for LHC is ***Data accessibility***.
    - Detector level studies require ***CPU intensive simulations***.
    - DNNs require large training sets with ***full level of detail*** (i.e. not AODs).
    - Experiments have such samples, but they are not easily accessible and ***not public***.
    - ***Difficult to collaborate*** with DL community or other experiments.
- Soon to be ***public datasets***:
    - ***LArTPC*** (Sepideh Shahsavarani, AF): LArIAT detector. 1 M of every particle species (including neutrinos).
    - ***Calorimetry*** (Maurizio Pierini, Jean-Roch Vlimant, Nikita Smirnov, AF): LCD Calorimeter. 1 M electron/pi0 (so far)
        - Many currently ***unused handles*** in LHC experiments… e.g. sampling.
        - Improve ***Identification***, e/gamma and hadronic ***Energy Resolution***.
        - ***Fast Shower Generative Models***
    - ***Tracking ML*** (David Rousseau, Andreas Salzberger, …)
        - HL-LHC like detector/environment.

Why go Deep?

- ***Better Algorithms***
    - DNN-based classification/regression generally **out perform** hand crafted algorithms.
    - In some cases, it may provide a ***solution*** where ***algorithm approach doesn't exist or fails***.
    - ***Unsupervised learning***: make sense of complicated data that we don't understand or expect.
- ***Easier Algorithm Development***: ***Feature Learning*** instead of *Feature Engineering*
    - Reduce time physicists spend writing developing algorithms, ***saving time and cost***.
    - Quickly perform performance ***optimization*** or ***systematic studies***.
- ***Faster Algorithms***
    - After training, DNN inference is often *faster* than sophisticated algorithmic approach.
    - DNN can ***encapsulate expensive computations***, e.g. Matrix Element Method.
    - ***Generative Models*** enable fast simulations.
    - ***Already parallelized*** and optimized for GPUs/HPCs.
    - ***Neuromorphic*** processors.
- ***Deep Learning may be the disruptive enabler of HL-LHC computing.***