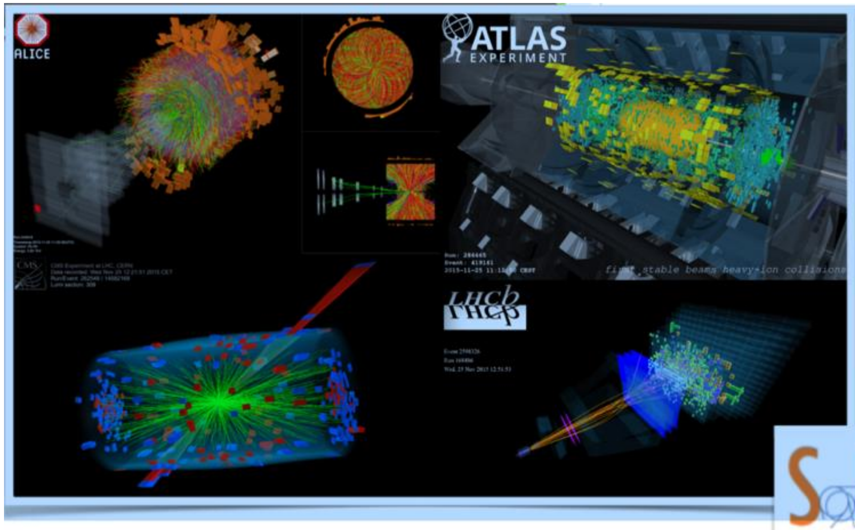


Inter-experimental LHC Machine Learning Working Group Activities

Sergei V. Gleyzer

University of Florida

LHC PHYSICS CENTRE AT CERN

Outline

- Inter-experimental LHC Machine Learning Working Group
- Machine Learning Applications at the LHC
- Community Efforts

IML

Inter-experimental LHC Machine Learning Working Group iml.cern.ch

- Sharing of expertise among LHC (and other HEP) experiments
  - ATLAS, CMS, LHCb, ALICE, Belle-II, neutrino experiments
  - ~450 participants
- Exchange between particle physics and machine learning communities
- Software development and maintenance
- Forum, Training and Education

Working Format

Monthly meetings around machine learning topics relevant to HEP community:

- **Deep Learning**
- **Software and Tools**
- **Hardware Applications**
- **Unsupervised Learning**
- **Anomaly Detection**
- **Multi-class/Multi-objective Learning**
- **Bayesian ML and GANs**
- **Theory Applications**

IML Workshop

First IML Workshop at CERN

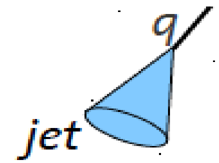
- March 20-22, 2017
  - ~300 participants
  - Industry session
  - Quark/gluon tagging challenge
  - Physics Object Tagging Workshop
  - HEP-ML Community White Paper
- More workshops forthcoming

LHC Applications and Challenges

LHC Applications

Primarily Classification

– **Low Level:**

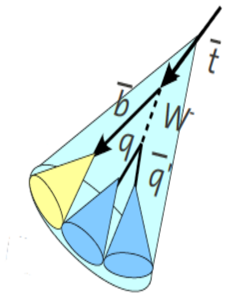


- **Particle identification**

  Photon or a jet?

- **Pattern recognition**

  Tracks, vertices



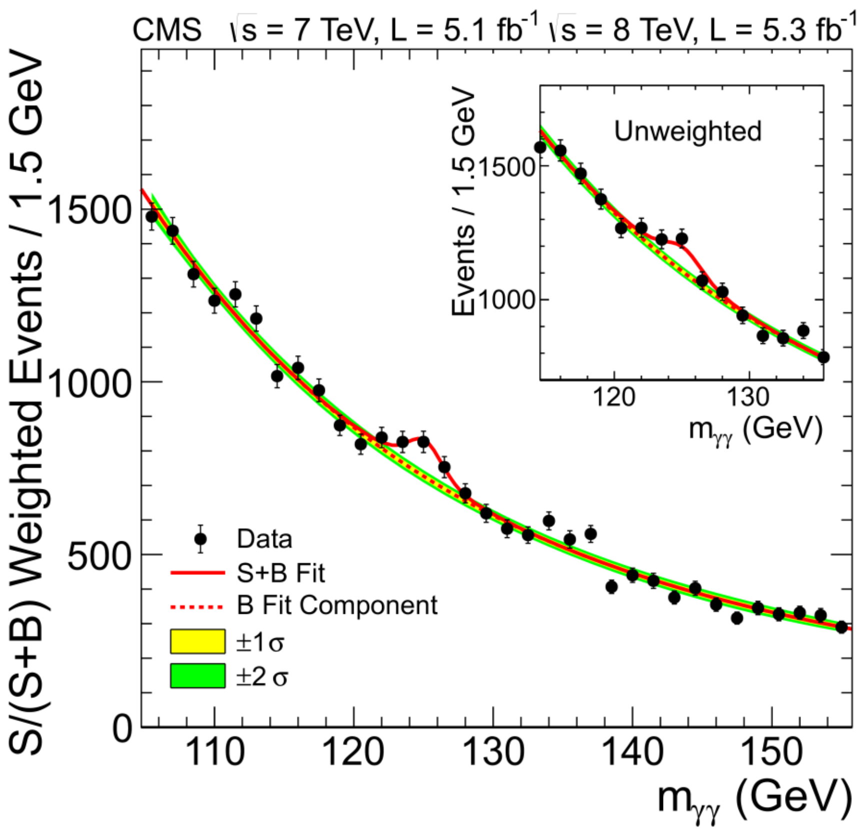
– **High Level:**

- **New Physics searches**

  Higgs/SUSY event or background?

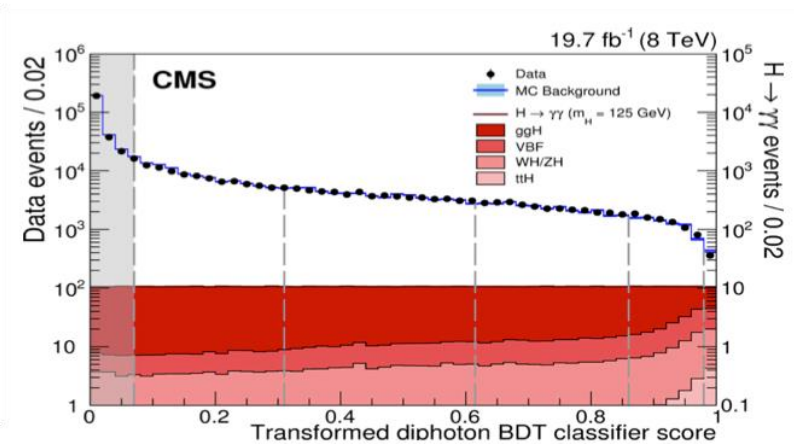
  Jet sub-structure

Higgs Discovery



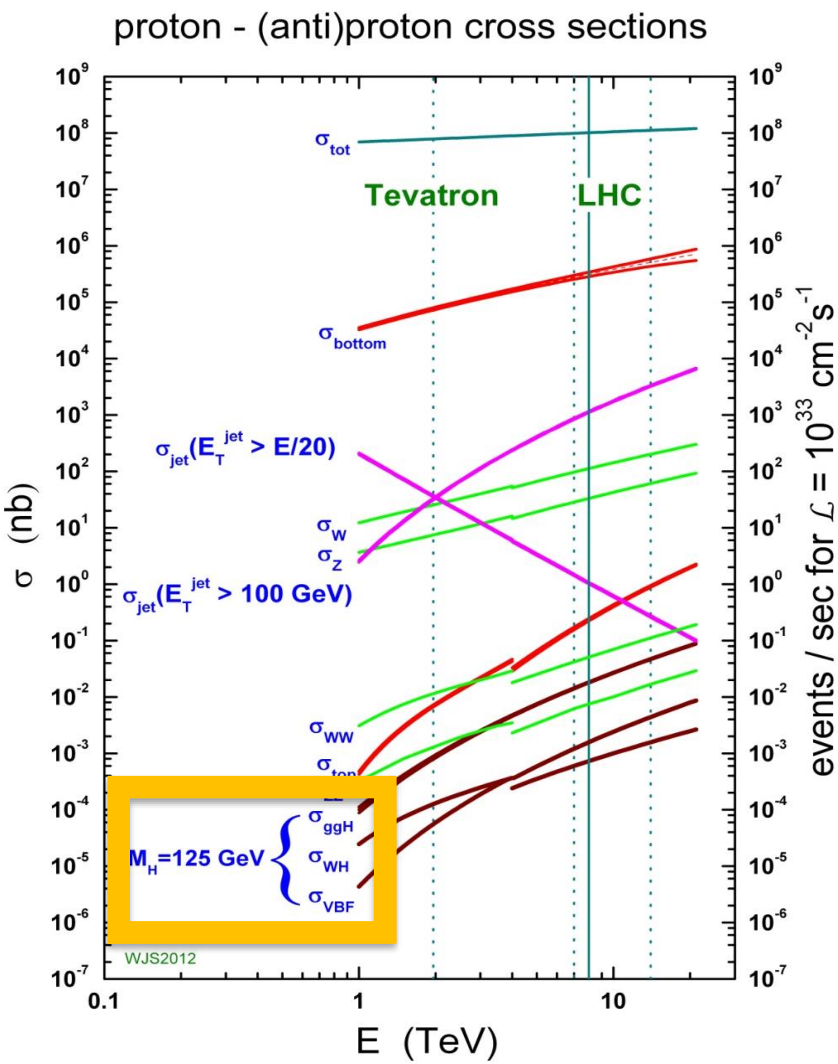
Machine Learning used in Higgs Discovery

- Event selection
- Identification of particles
- Identification of interactions
- Energy regression



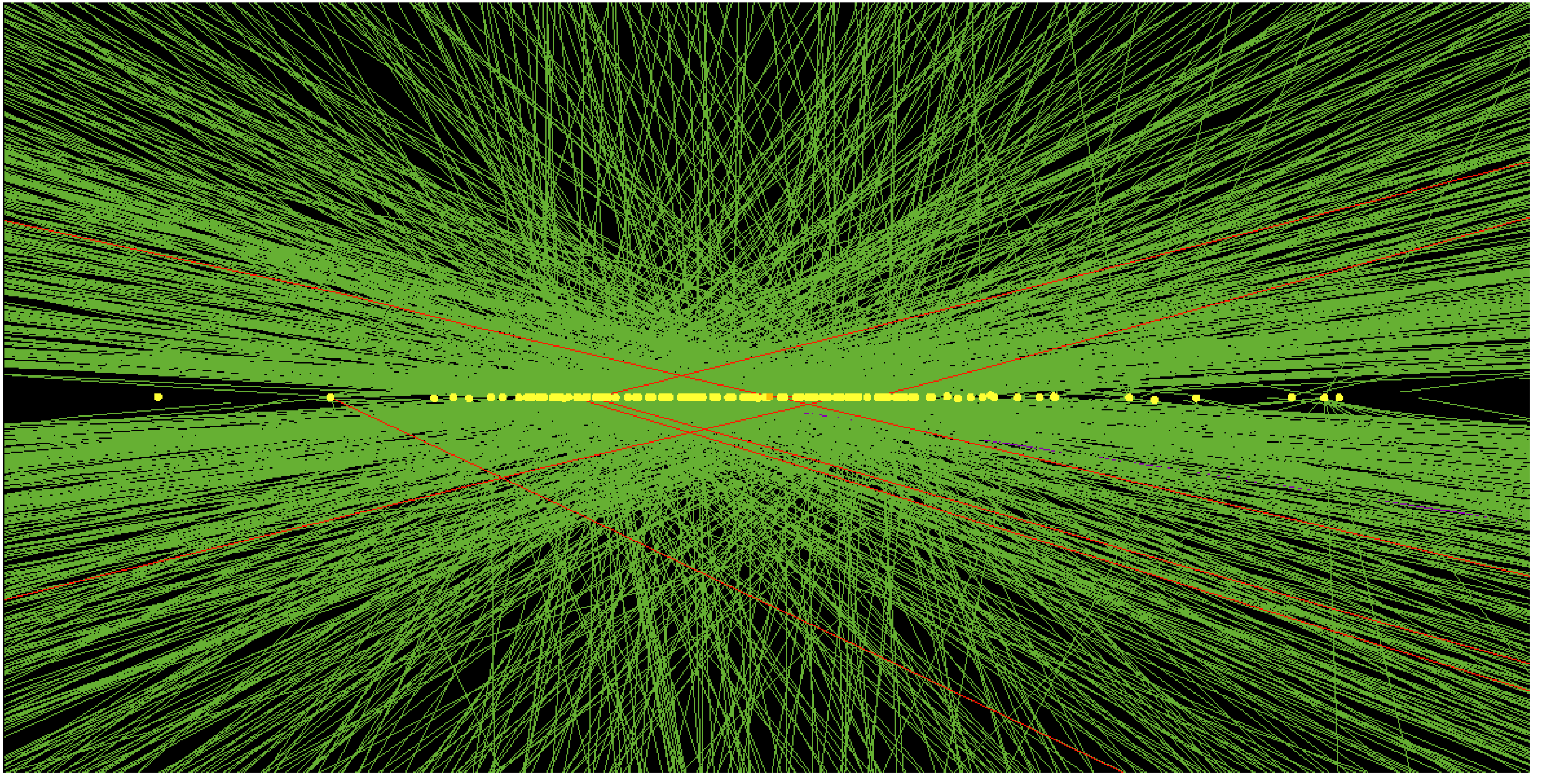
Improvement in analysis from all four areas

Upcoming Challenges

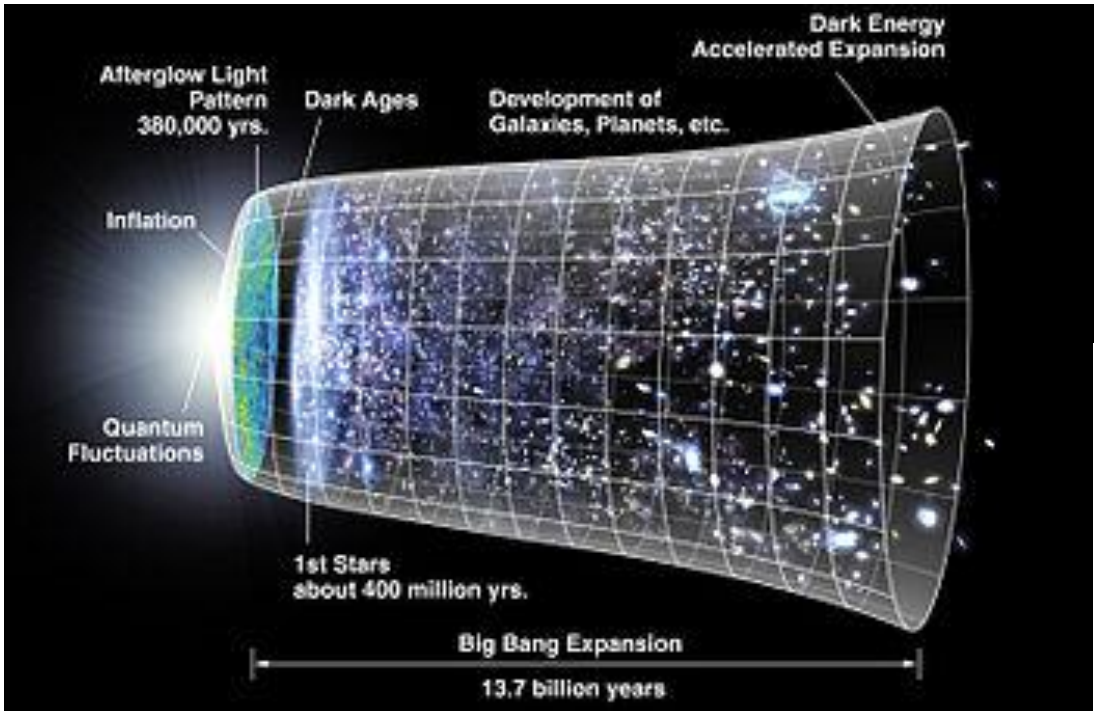


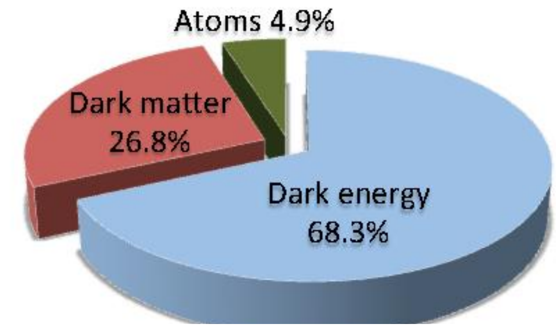
Orders of magnitude between signals and backgrounds

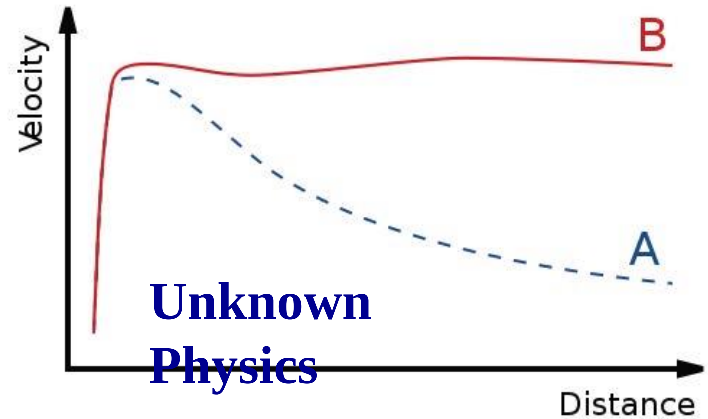
Event Complexity



Upcoming Challenges







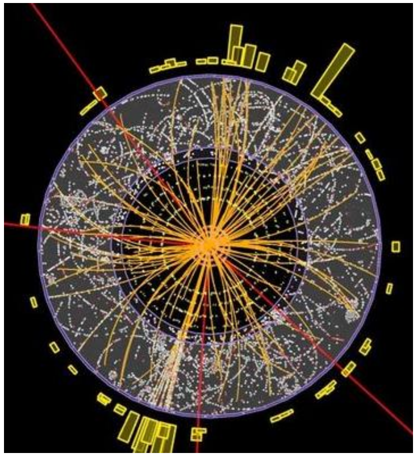
Unknown Physics

Data size:

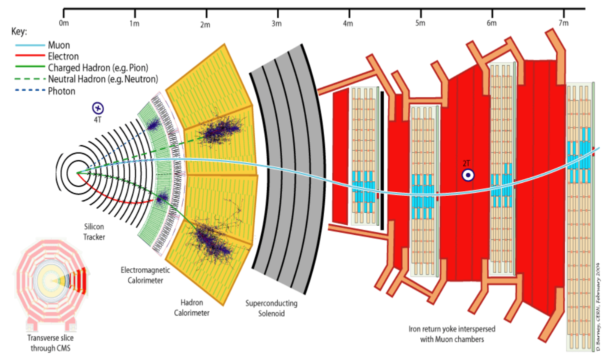
- LHC 15,000,000 Tb 2010 – 2035
- Resources not up as fast as data volume

Topics of interest

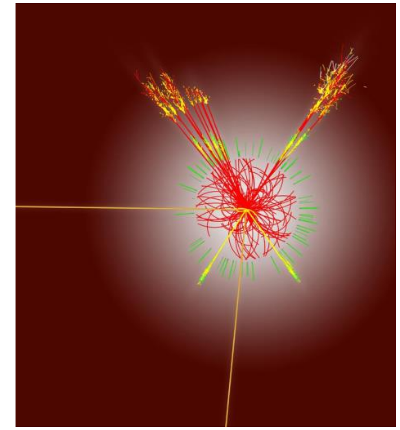
Interesting areas



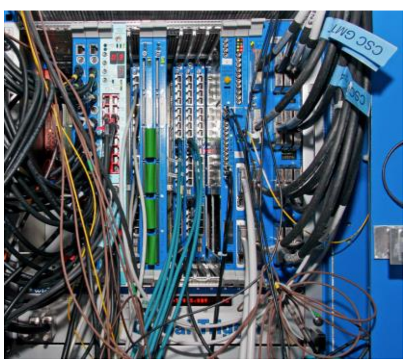
Particle Tracking



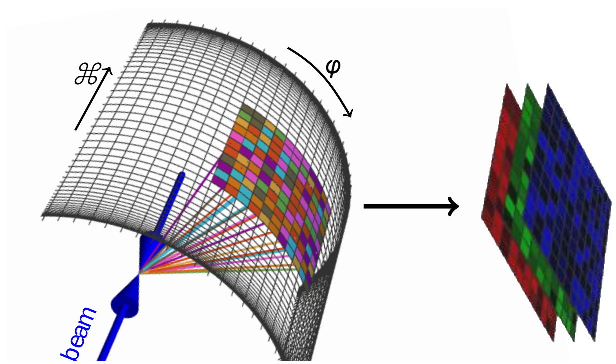
Fast Simulation



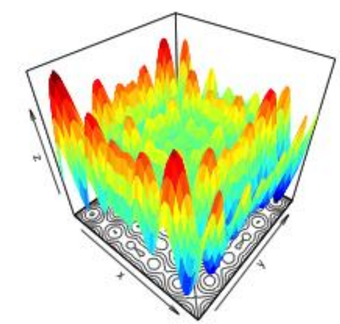
Object Identification



Trigger

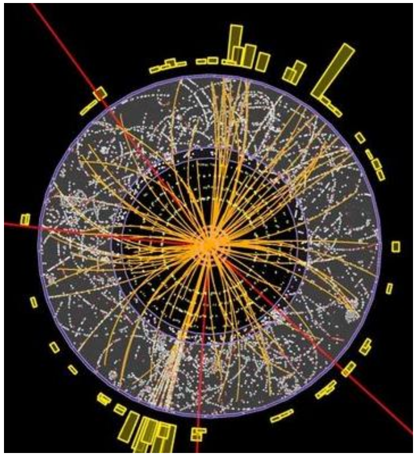


Imaging Calorimetry

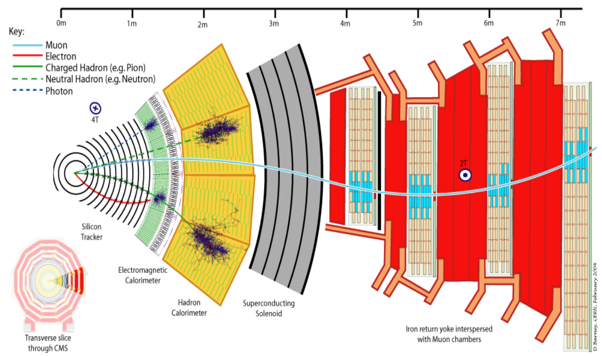


Simulation

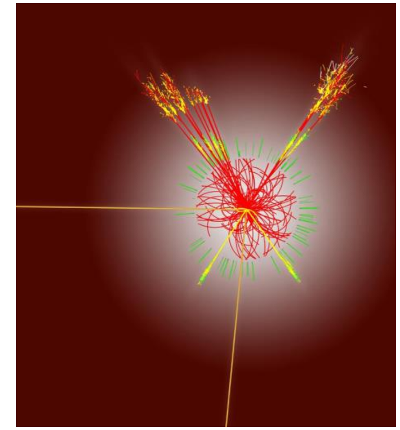
Interesting areas



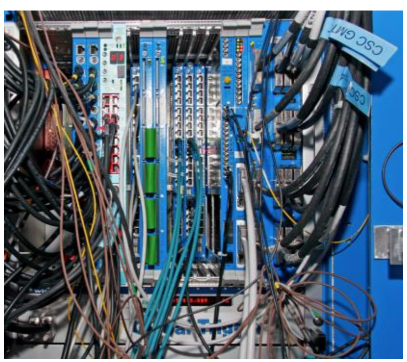
Deep Kalman RNNs



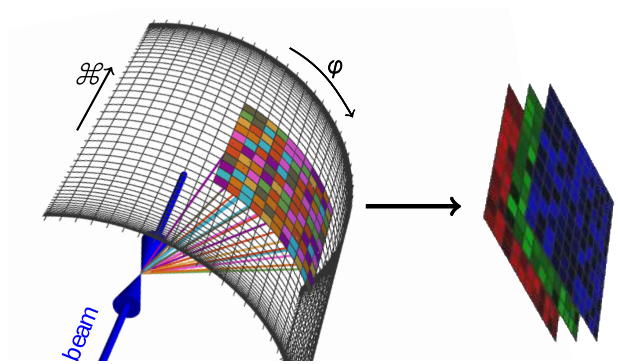
Generative Models, Adversarial Networks



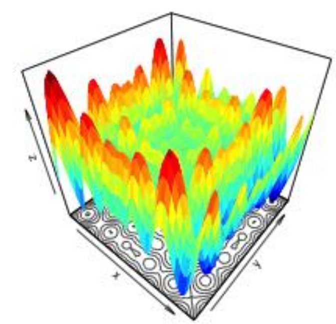
FCN, Recurrent, LSTM NN



Deep ML +FPGA



Convolutional DNN



Multiobjective Regression

Deep Learning

Deep Learning

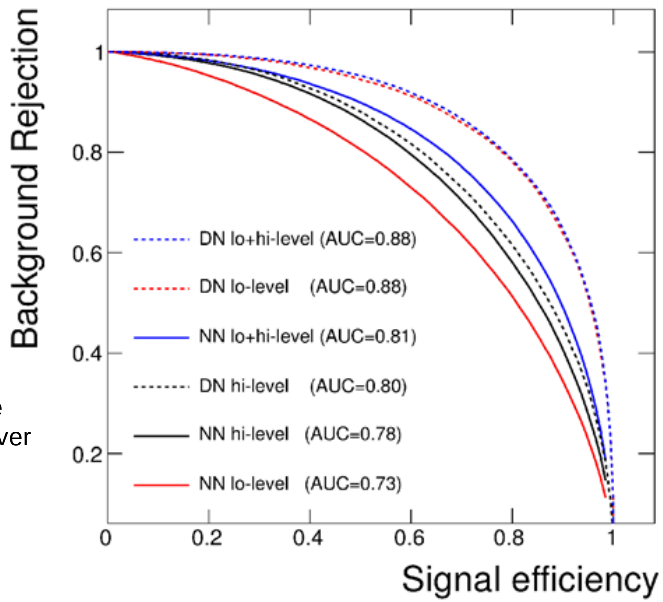
Higgs Boson Example:

P. Baldi, et. al. 2014

Tuning deep neural network architectures.

Hyper parameters	Choices
Depth	2,3,4,5,6 layers
Hidden units per layer	100,200,300,500
Learning rate	0.01, 0.05
Weight decay	0, 0.00001
Pre-training	none, autoencoder
	multi-task autoencoder
Input features	low-level, high-level
	complete set

Best:
- 5 hidden layers
- 300 neurons per layer
- Tanh hidden units, sigmoid output
- No pre-training
- Stochastic gradient descent
- Mini batches of 100
- Exponentially-decreasing learning rate
- Momentum increasing from .5 to .99 over 200 epochs
- Weight decay = 0.00001



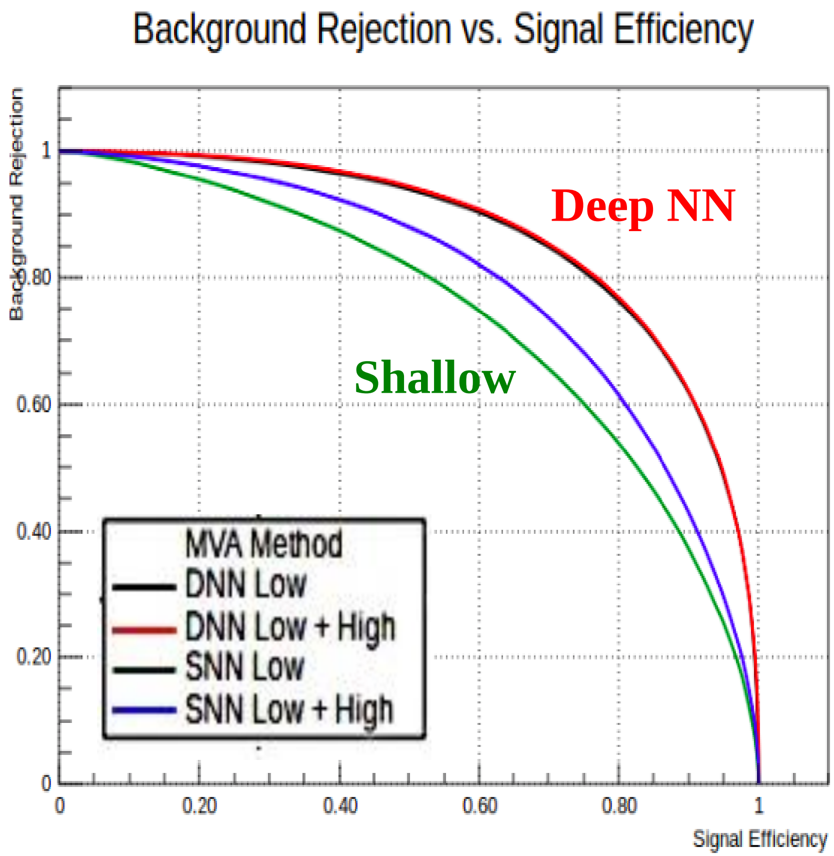
8% improvement

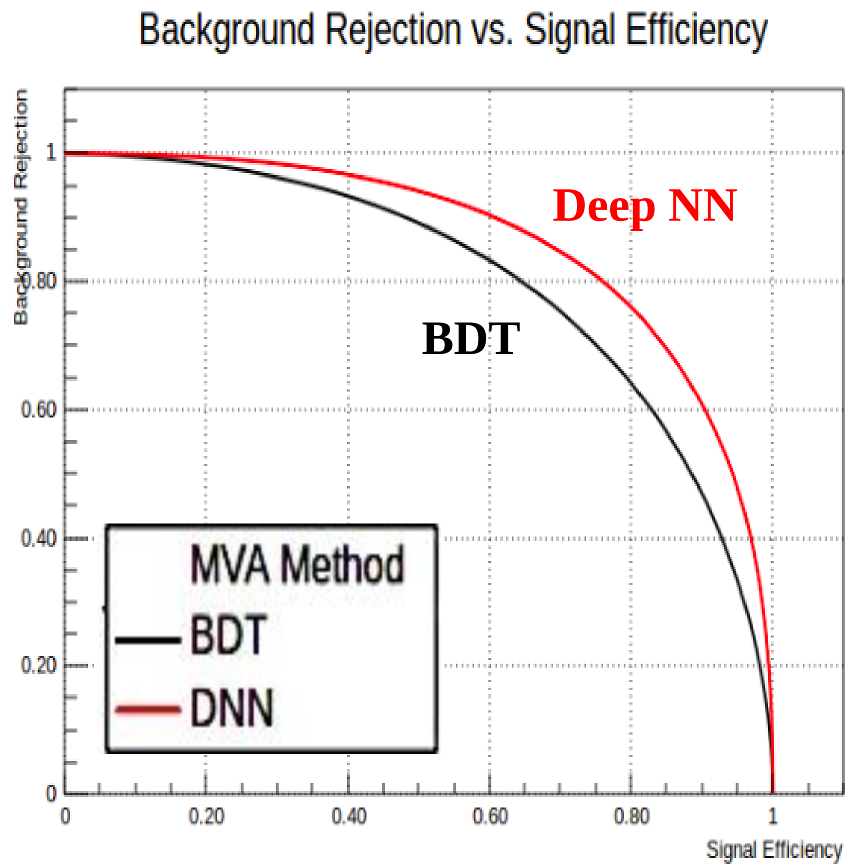
Deep Learning

Papers in HEP:

- Jet images and deep learning: arxiv1511.05190
- Jet substructure and deep learning: http://inspirehep.net/record/1437937/
- Parton shower uncertainties and jet substructure: http://inspirehep.net/record/1485081?ln=en
- Deep learning for ttHhttp://inspirehep.net/record/1491175?ln=en
- Nova http://inspirehep.net/record/1444342
- Daya Bay arxiv1601.07621
- Next: http://inspirehep.net/record/1487439?ln=en
- Microboone: http://inspirehep.net/record/1498561?ln=en

Deep Learning

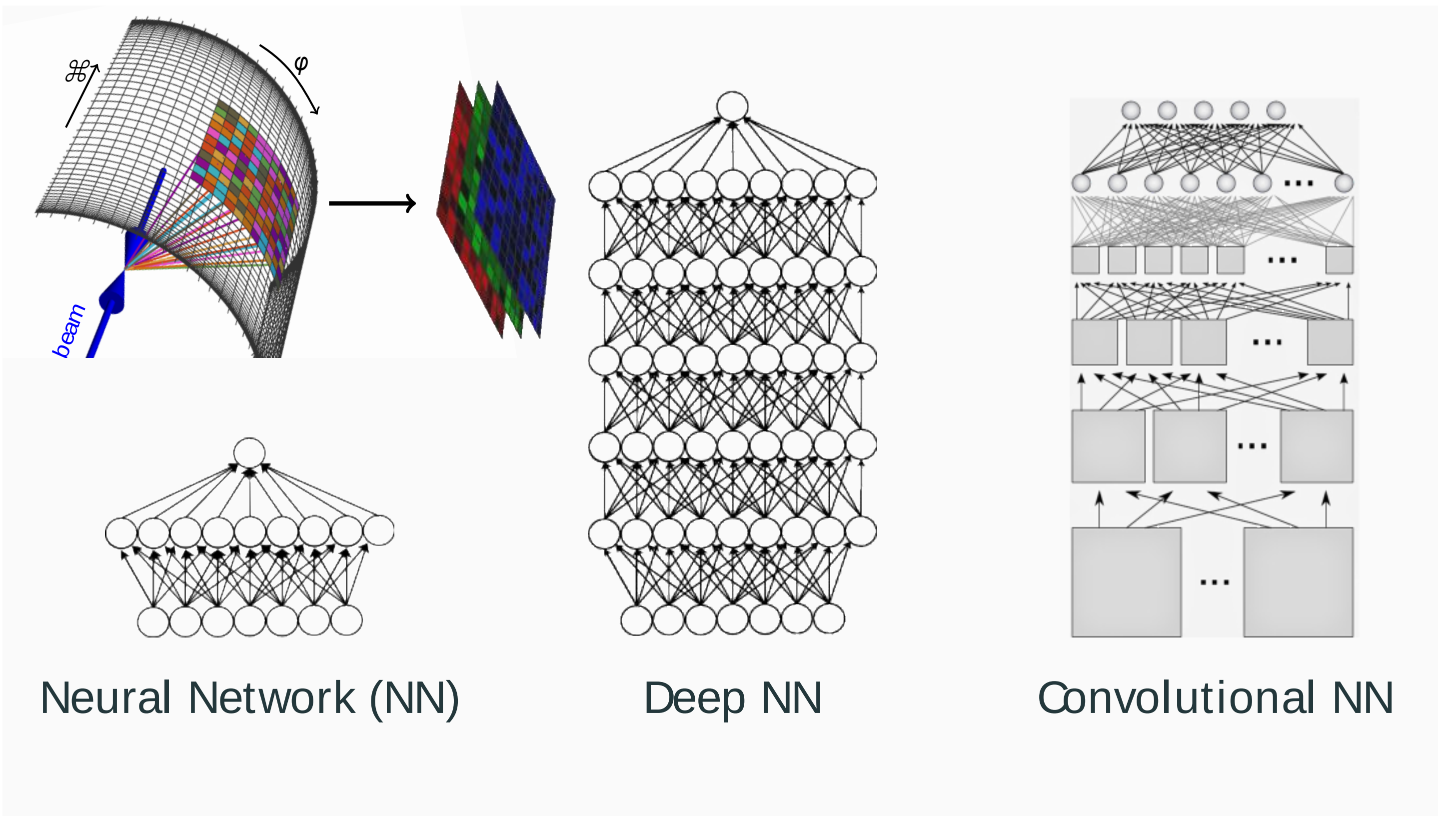




Significant improvements in performance

Deep Learning

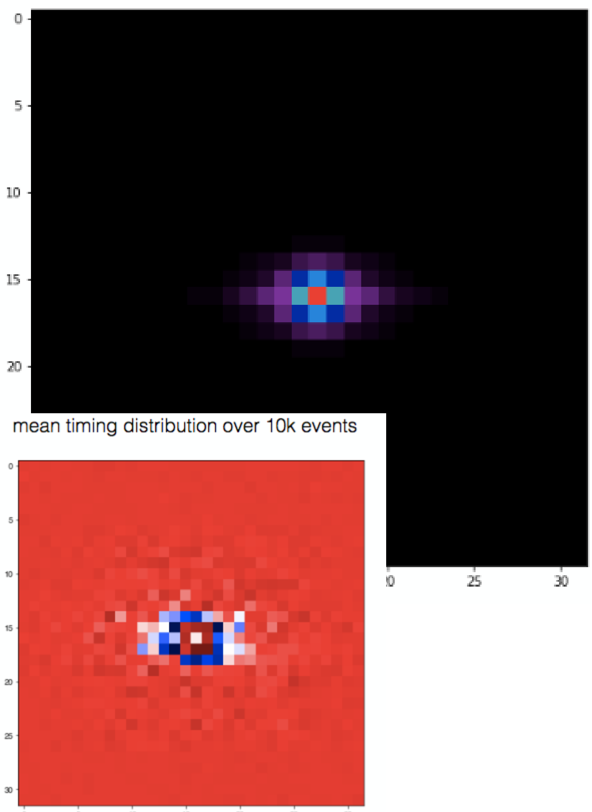
Convolutional Neural Networks



End-to-End Approach

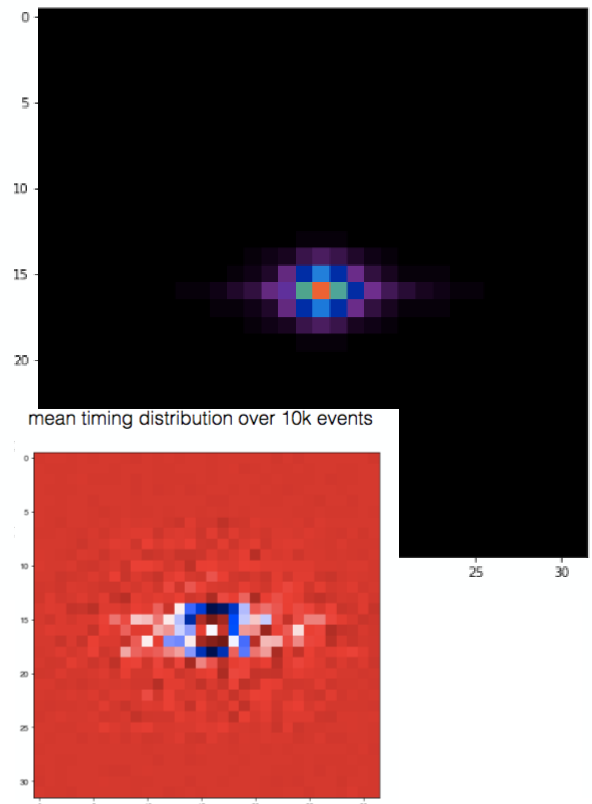
Photon-Induced EM Shower

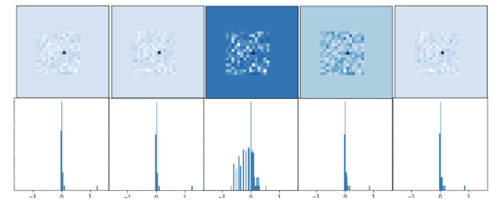
mean energy distribution over 10k events

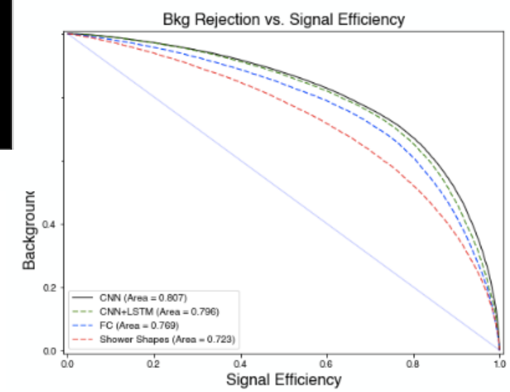


Electron-Induced EM Shower

mean energy distribution over 10k events

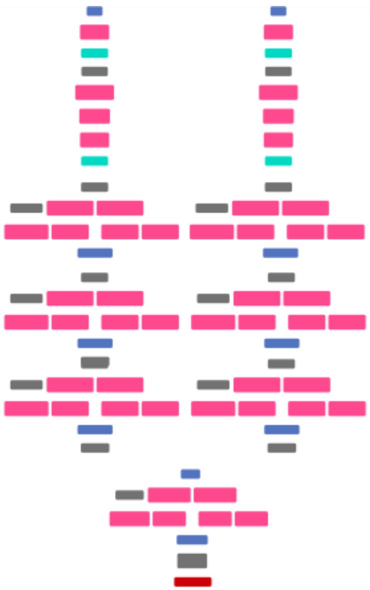


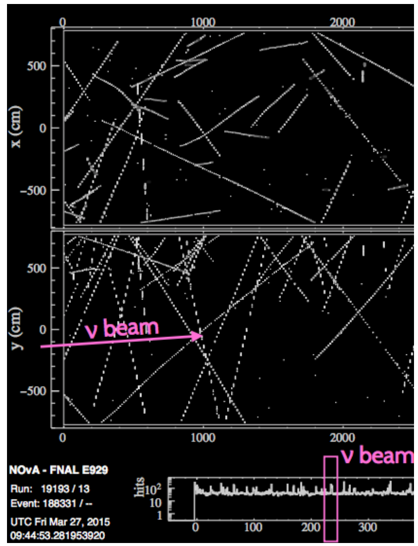


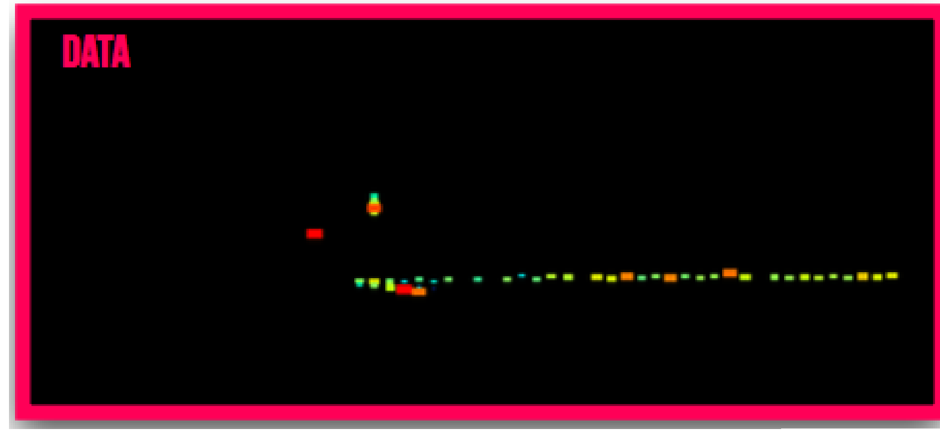


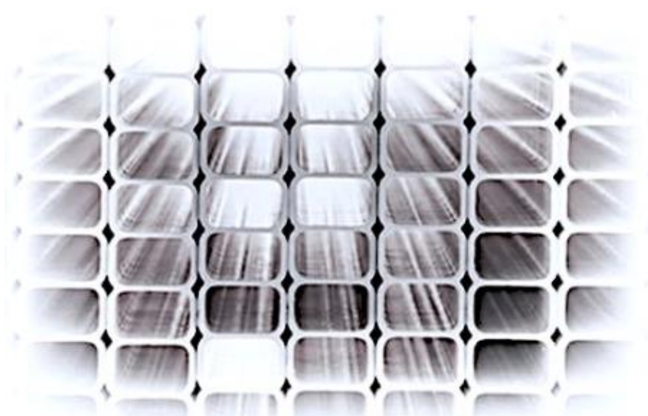
link

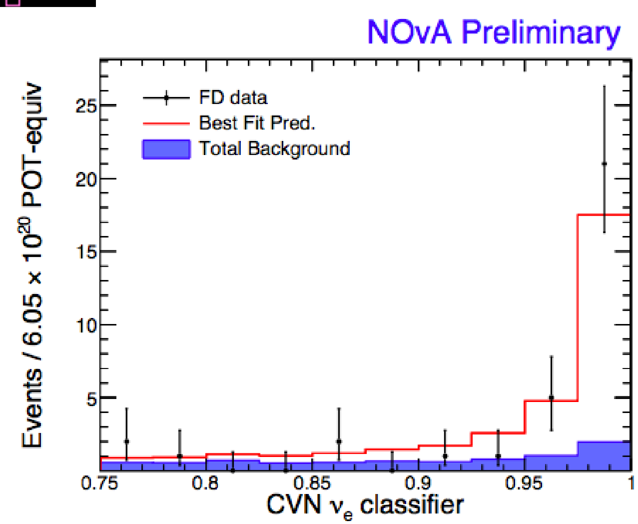
CVN on NOvA

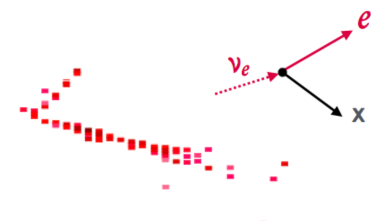












76% Purity

73% Efficiency

An equivalent increased exposure of 30%

link

Beyond Classification

Single-Objective Regression

Train learning model to estimate a single function target or "objective"

- Ex. photon energy/muon momentum

With a machine learning algorithm

- Decision tree, random forest, neural network etc.

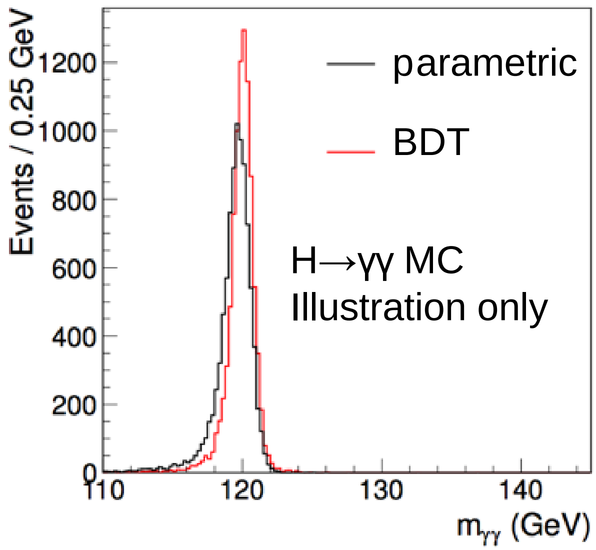
Photon Energy

Single Target Example:

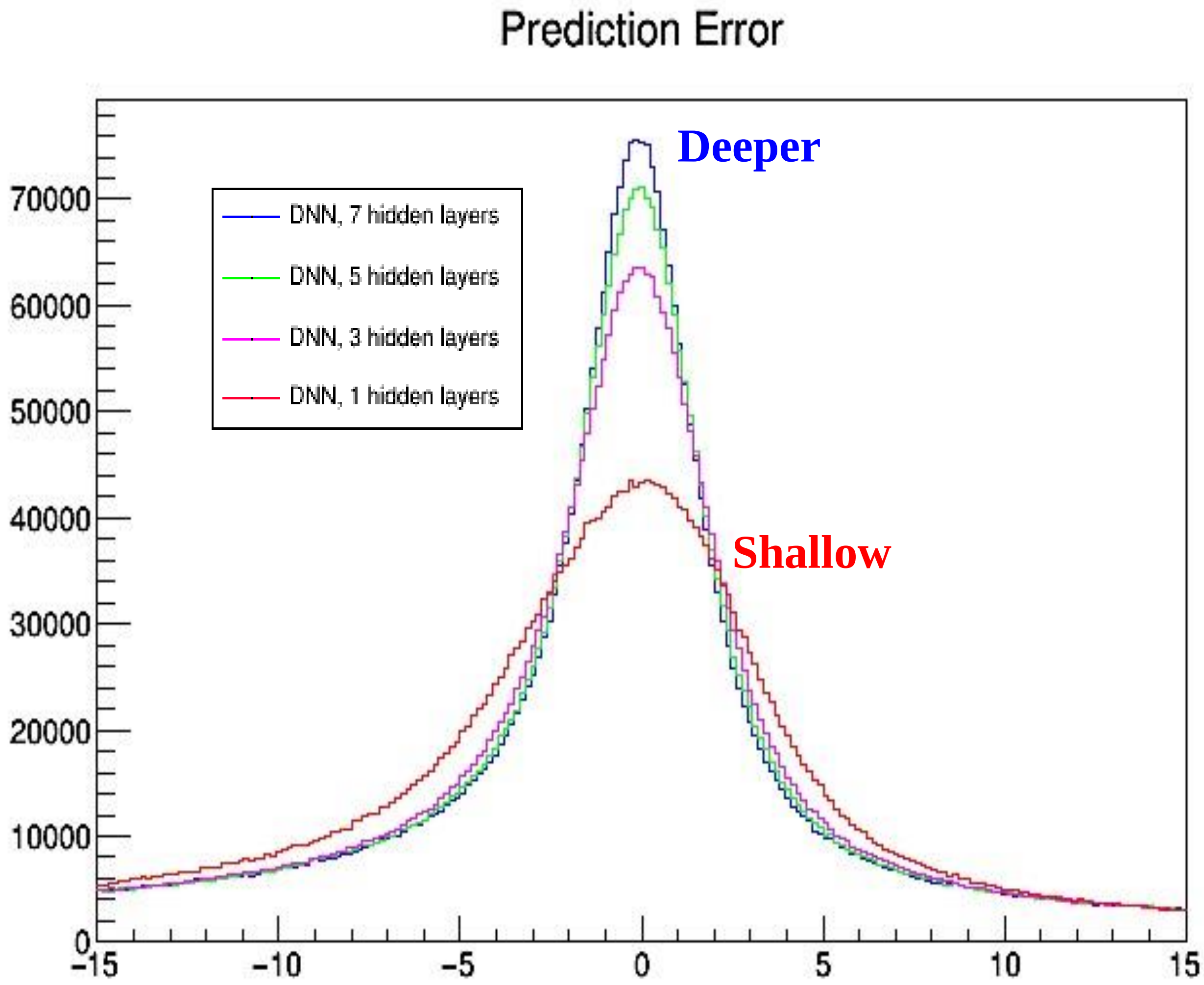
Inputs: shower information, photon coordinates, median event energy

Target Output: $E_{MEASURED}/E_{TRUE}$

~10-30% improvement in resolution



Deep Learning Regression



Multi-Objective Regression

Multi-Objective Regression

Simultaneous estimate of multiple functions or "targets"

- Possibly additionally correlated
  - N single-target models not as optimal lingo: "multi-task" learning
  - and more cumbersome
- Train a single model to simultaneously predict all targets

Applicable Models

Methods:

- Regression decision trees
- Decision rules
- Decision rule ensembles
- Random forest
- Neural networks…

Trade-offs:

- accuracy, model size, interpretability

Multi-objective Example

X input variables {a, b, c, d…}

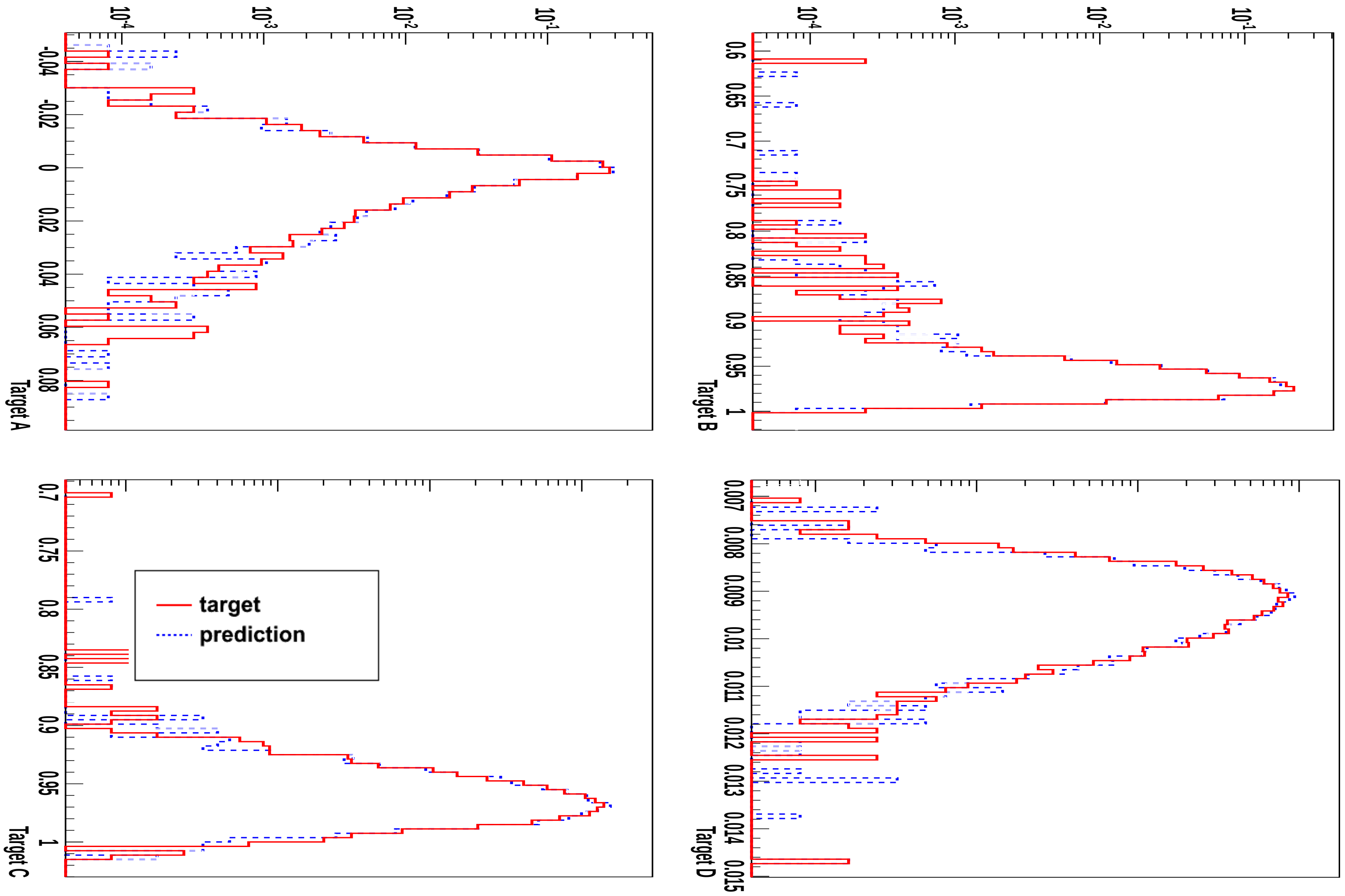
- K of them strongly correlated

Y target outputs to estimate {A, B, C, D…}

- N of them strongly correlated

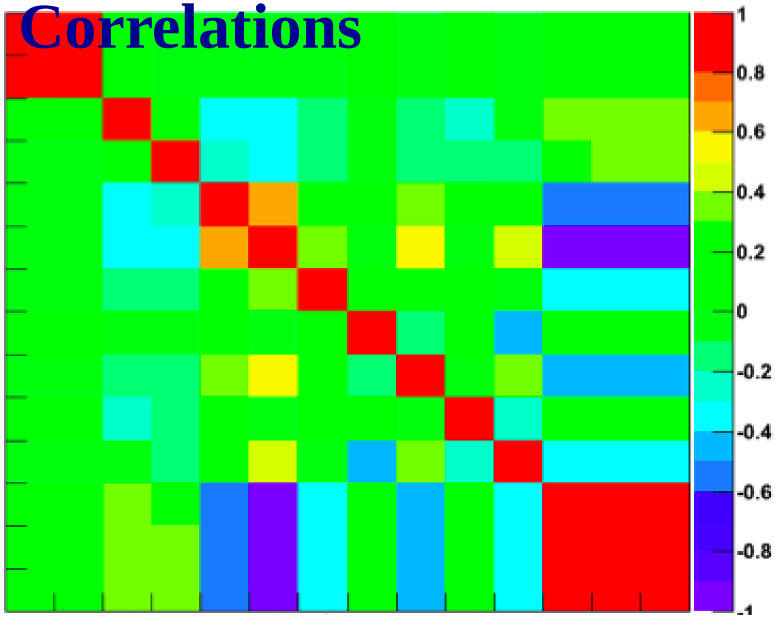
Challenge: build a predictive model to describe simultaneously all the outputs {A,B,C,D…}, provided a corresponding set of inputs.

Illustrative Example

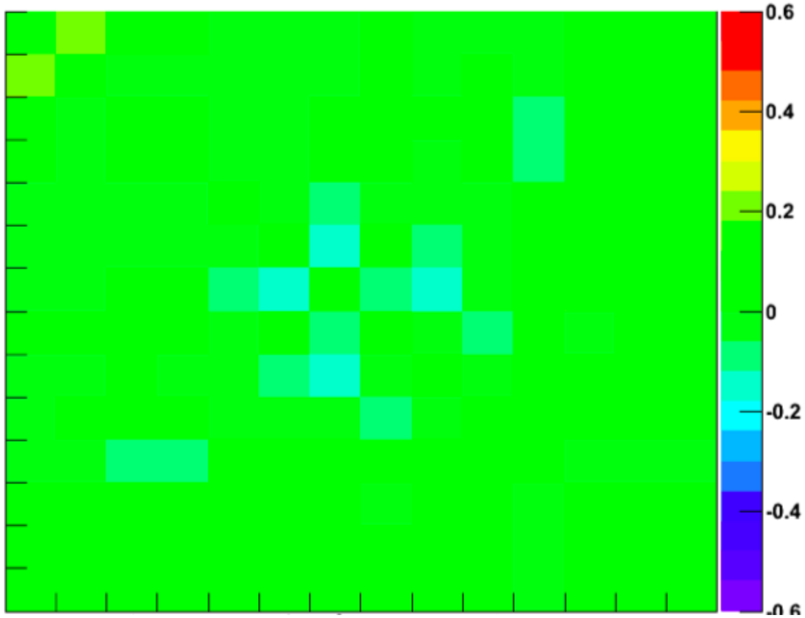


Target Correlations

Target Correlations



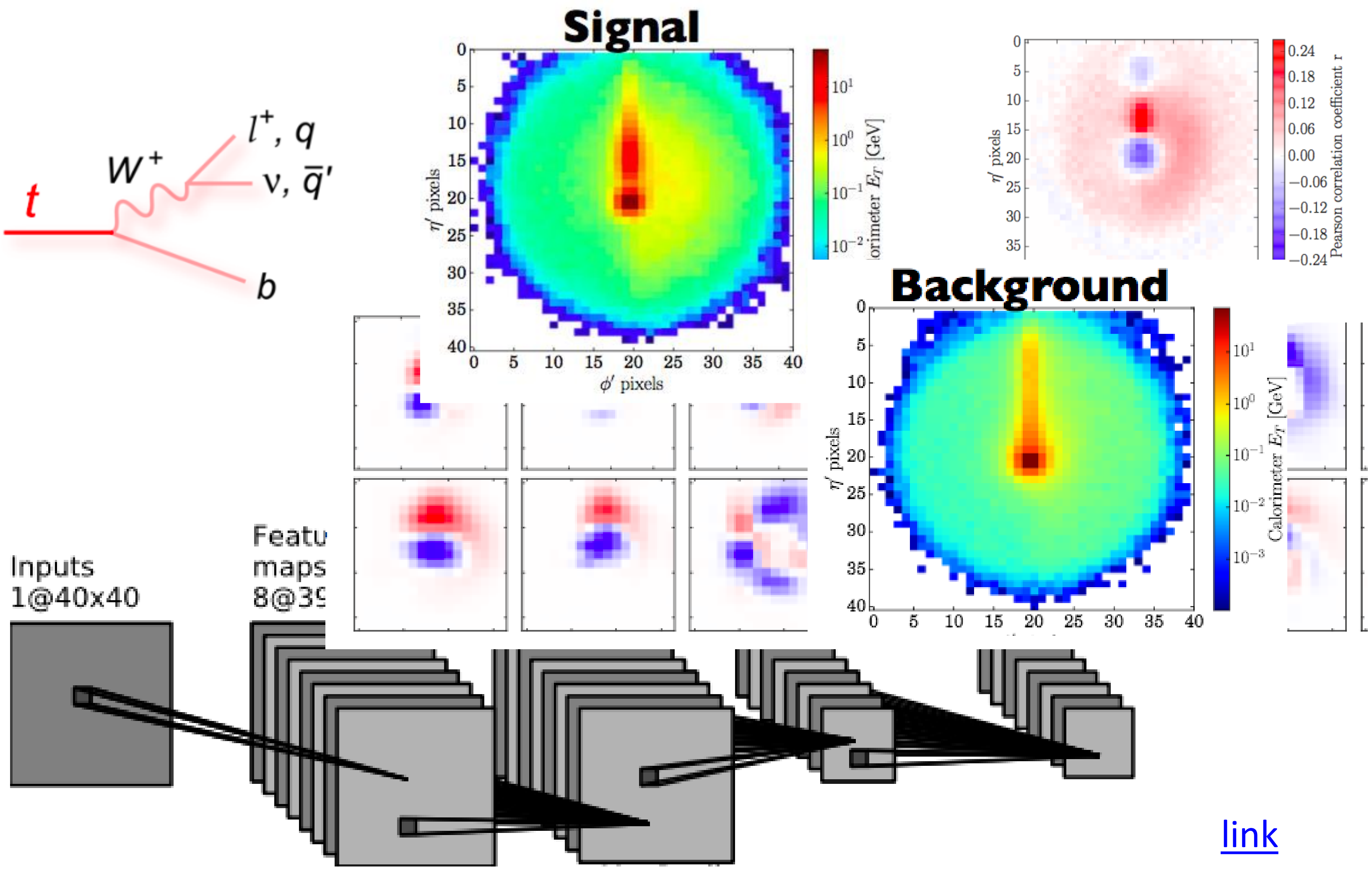
Prediction-Target Difference



Very close to Zero

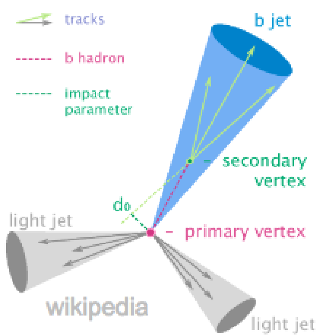
Physics Object Tagging

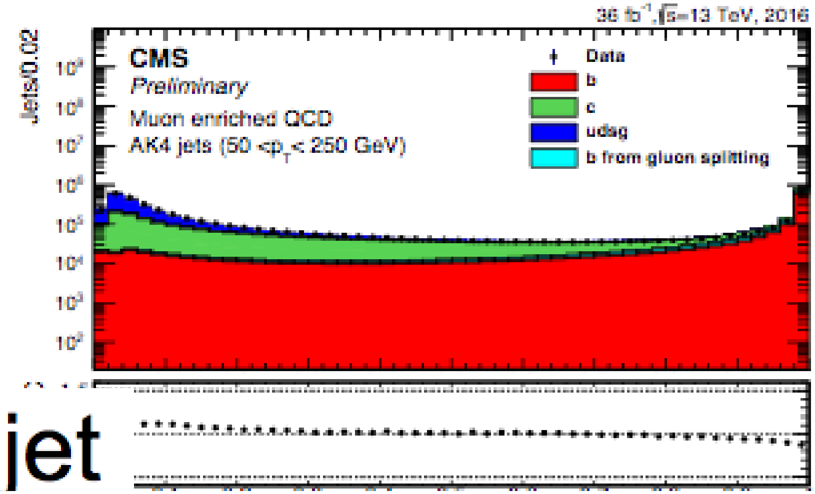
Object Tagging



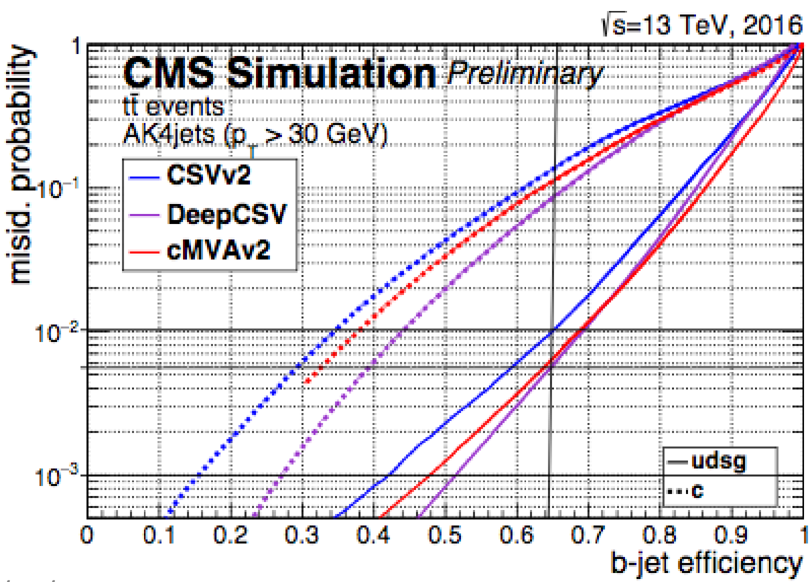
link

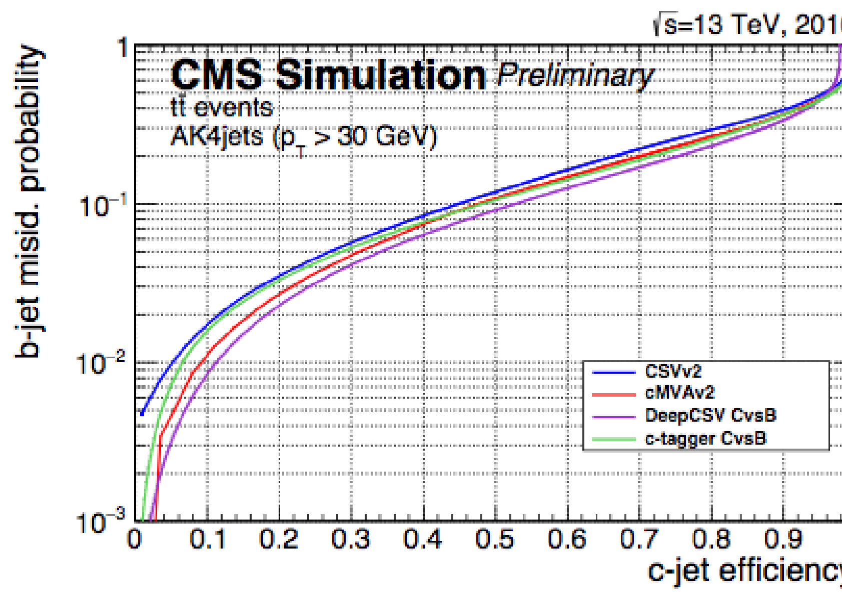
Object Tagging





ROC b-jet vs. light and c-jet



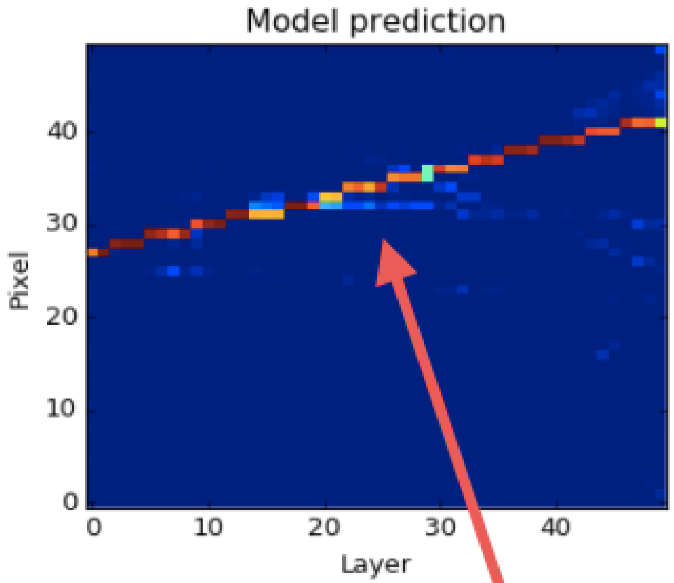


Tracking

HEP.TrackX

Track Extension with LSTM

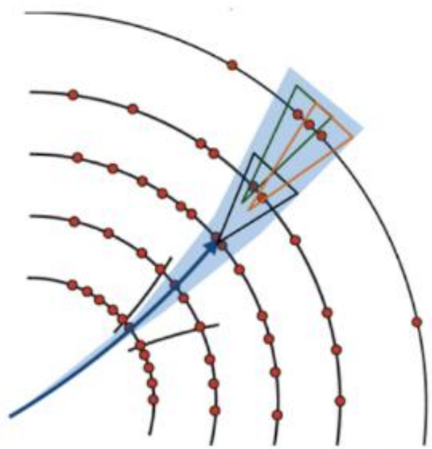
- Repeat for each detector slice to obtain full prediction
- The LSTM memory state propagates relevant information from layer to layer



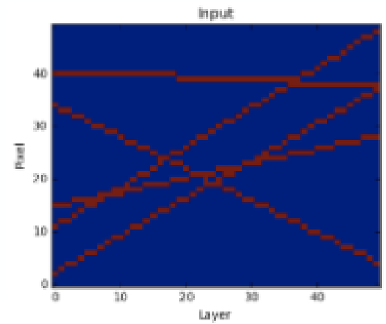
Uncertainty is larger near track intersection points

[link]

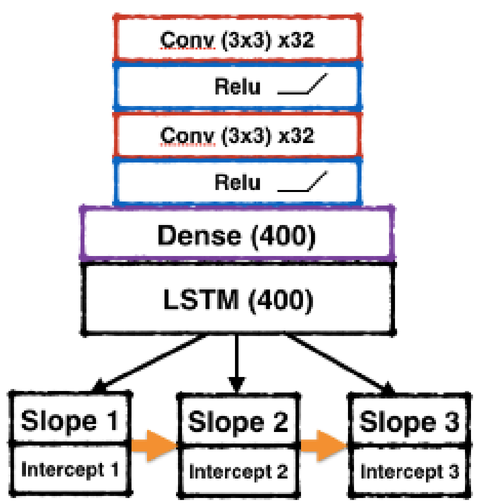
HEP.TrackX

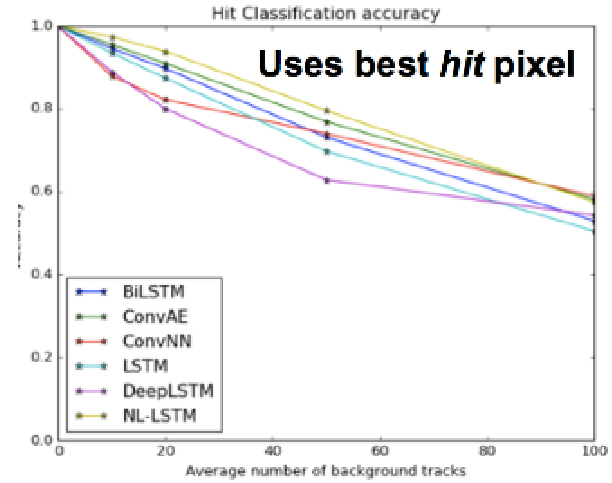


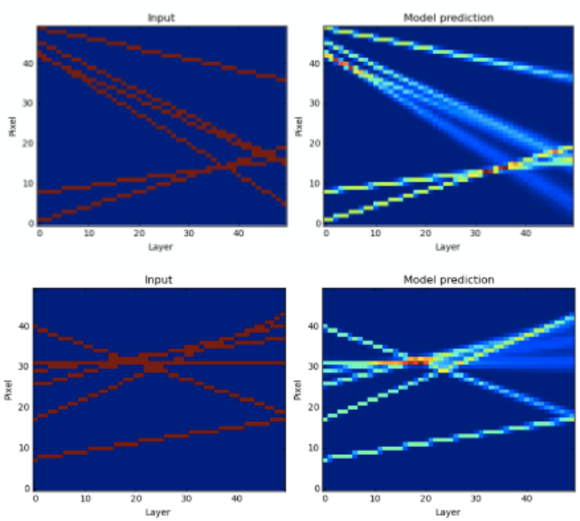
Credit: Andy Salzburger

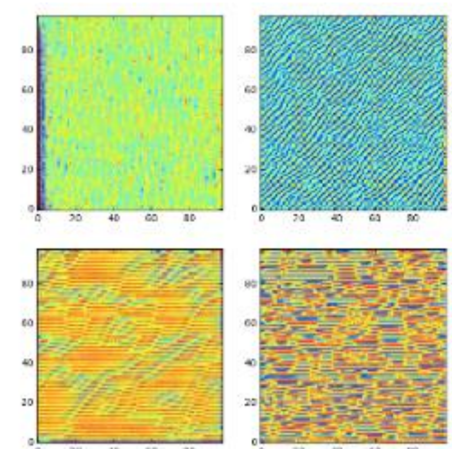


link



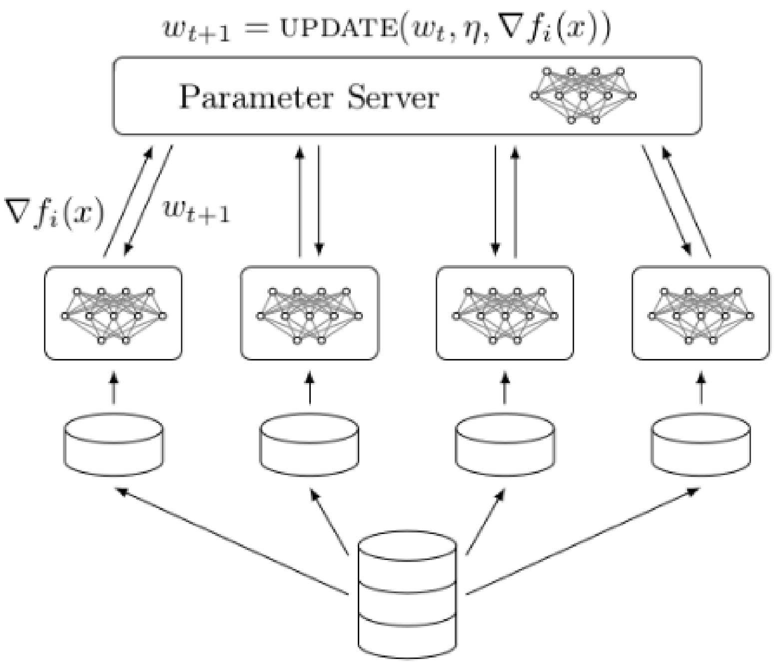


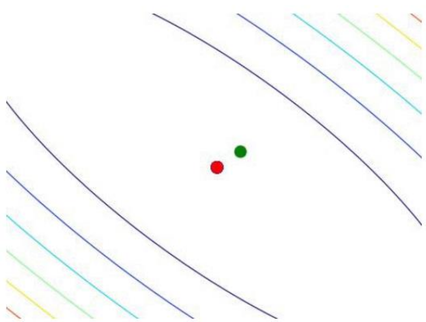


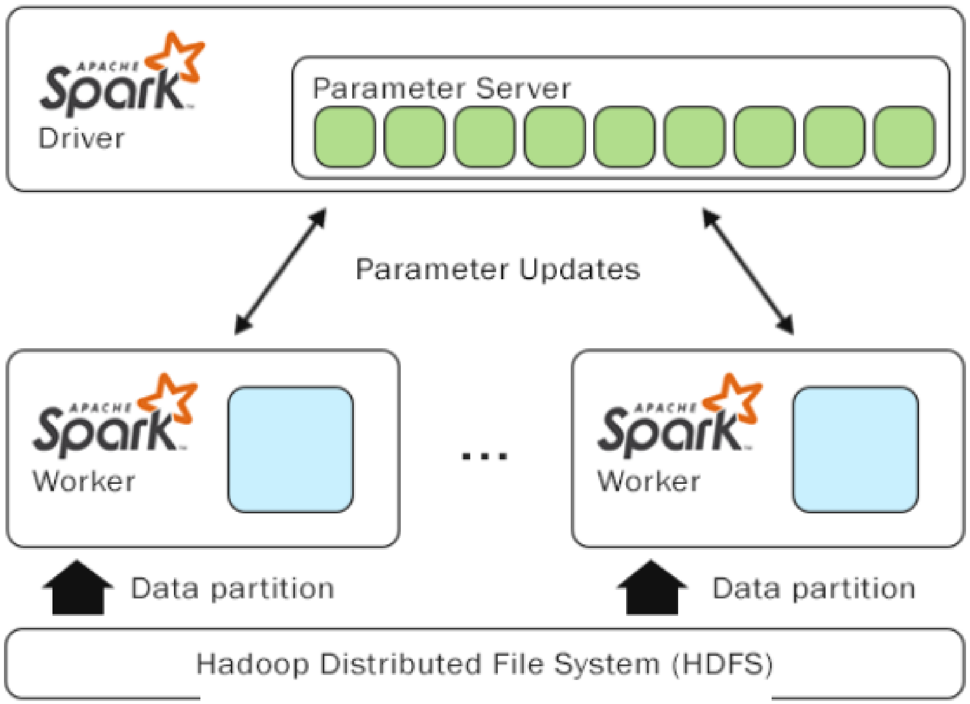


Software and Tools

DNN Apache Spark



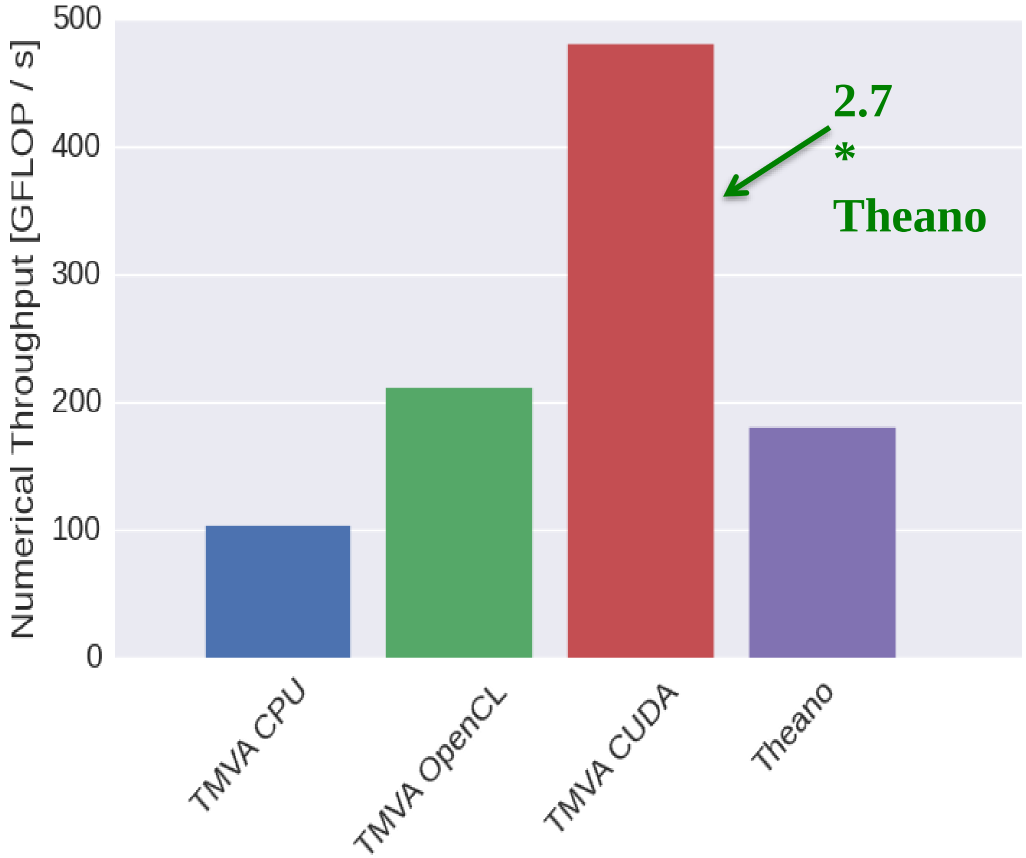




link

Deep Learning

Throughput Comparison



Single precision

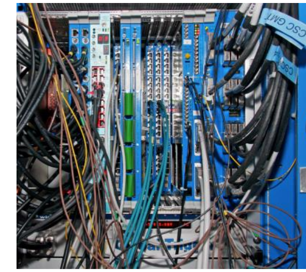
Excellent throughput compared to Theano on same GPU

Hardware Applications

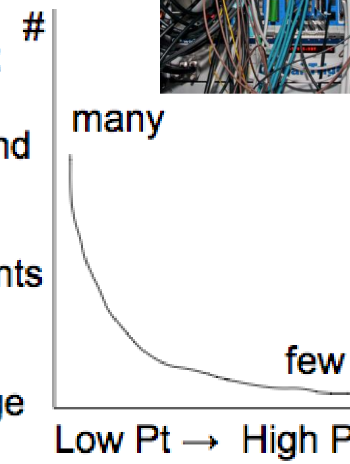
L1 Trigger Implementation

3

- **In a good muon trigger we want to**
  - pass muons whose pt are truly above some threshold
  - reject muons whose pt are truly below some threshold
  - In L1 we operate online, it needs to be fast ~25ns



- **So we train BDTs on a set of simulated muons to predict the Pt**
  - There are way more low pt muons than high pt ones
  - Misinterpreting low pt muons as high ones will increase the rate a lot and you have a terrible trigger (lots of bad events)
  - So we focus on low pt events: (1/pt – 1/pt')^2
    - Using a specific transformation of the target to focus on low pt events – 1/pt → makes low pt a large number
    - Using Least Squares to further focus on low pt events
    - Turned low pt into large numbers and we square differences in large numbers, hence the focus



- **How to implement in hardware?**
- In hardware the inputs must be discrete, so we decide upon a bit scheme for the features
  - 4 bits for feature 1, 8 bits for feature 2, etc
- Run over all possibilities to Pt for every word
- Now we have a map from every possible discretized input → BDT Prediction
- Call this a look up table
- Run this in FPGAs in the L1 Trigger for EMTF

link

Write look up table using BDT

00000000 → BDT → Pt(00000000)
00000001 → BDT → Pt(00000001)
…
…

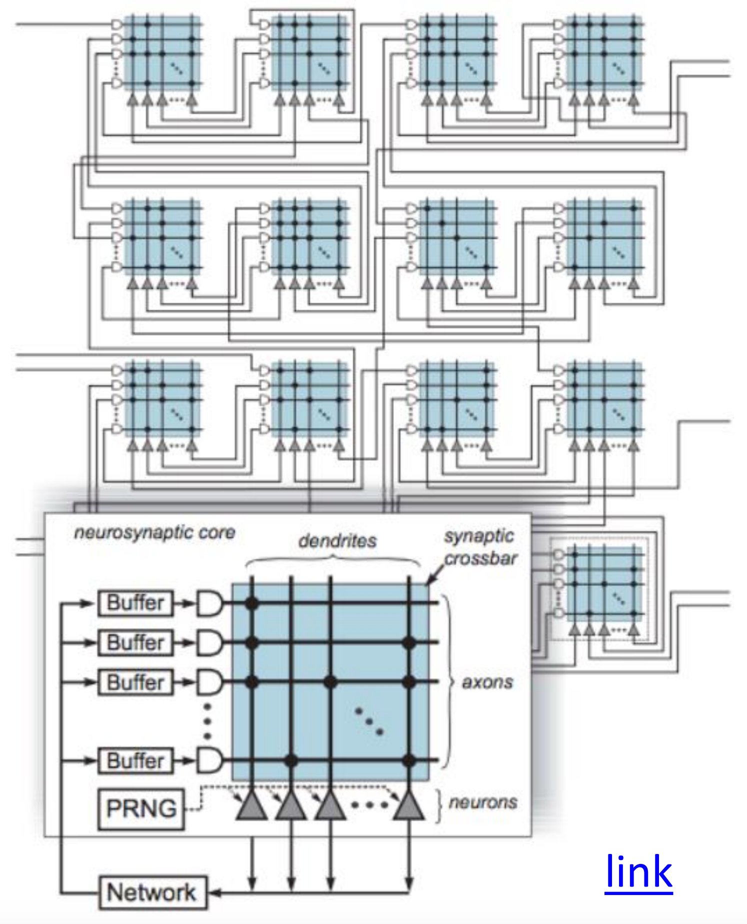
All Possibilities

Look Up Table written, don't need BDT anymore

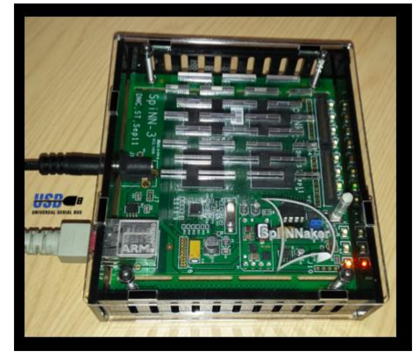
00000000 = Pt(00000000)
00000001 = Pt(00000001)
…

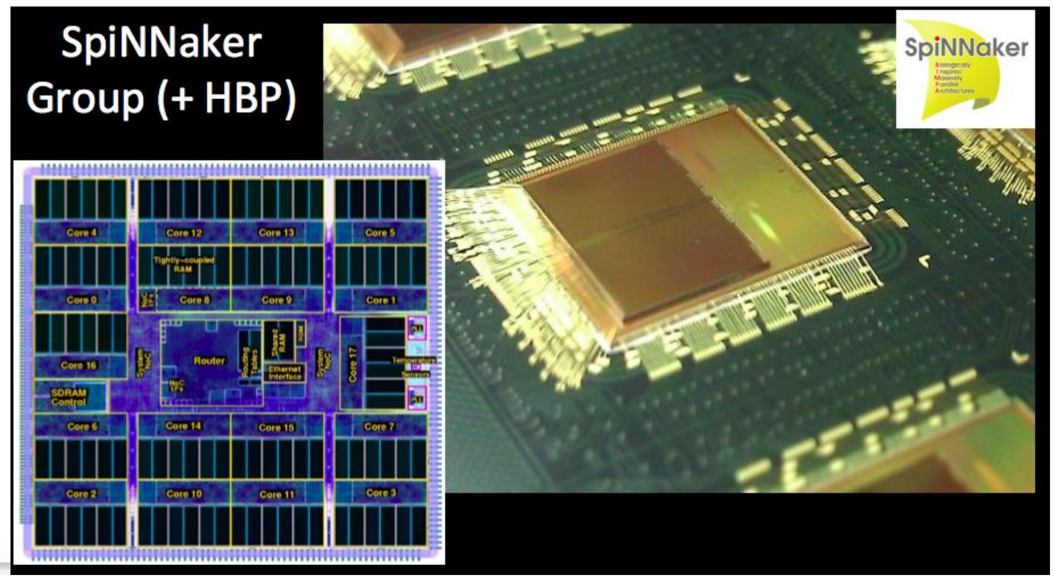
...use this in hardware, it's fast

Neuromorphic



link



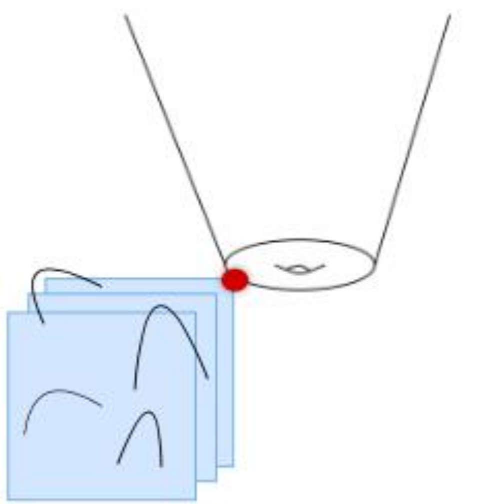


Theory and Phenomenology

The problem (in a nutshell)

Geometry meets field theory:

Toric Calabi-Yau 3-fold

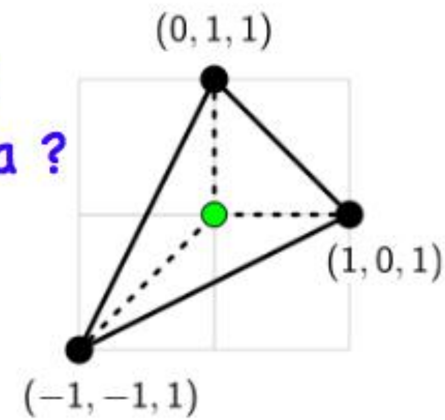


D3-branes

Can be used to calculate the "minimum" volume V_{min} of the base of the geometry (5 dim. subspace)

Q:

Is V_{min} determined by topological data ?



$\mathcal{N}=1$ supersymmetric gauge theory in 4d

Gauge and matter content determined by geometry !

link

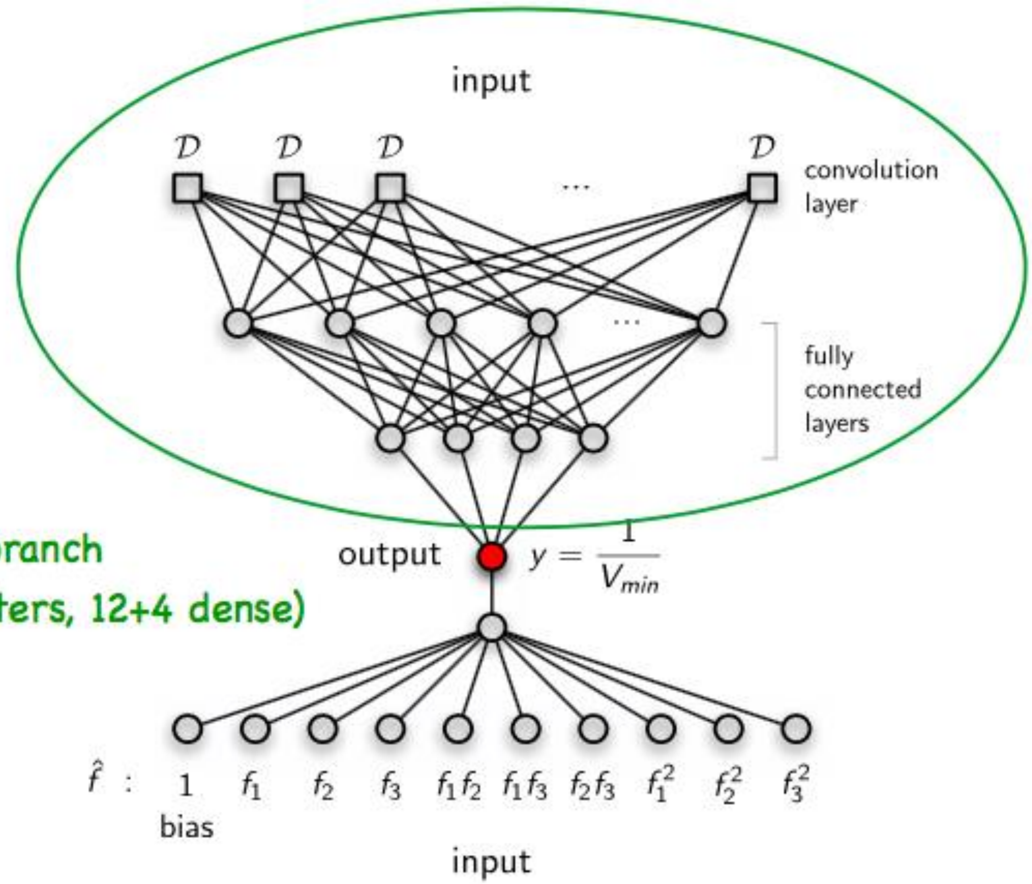
Wide CNN

Second Ansatz:

$$y = F(\hat{f}, \mathcal{D})$$

Idea: Let the ML model learn useful features by itself !

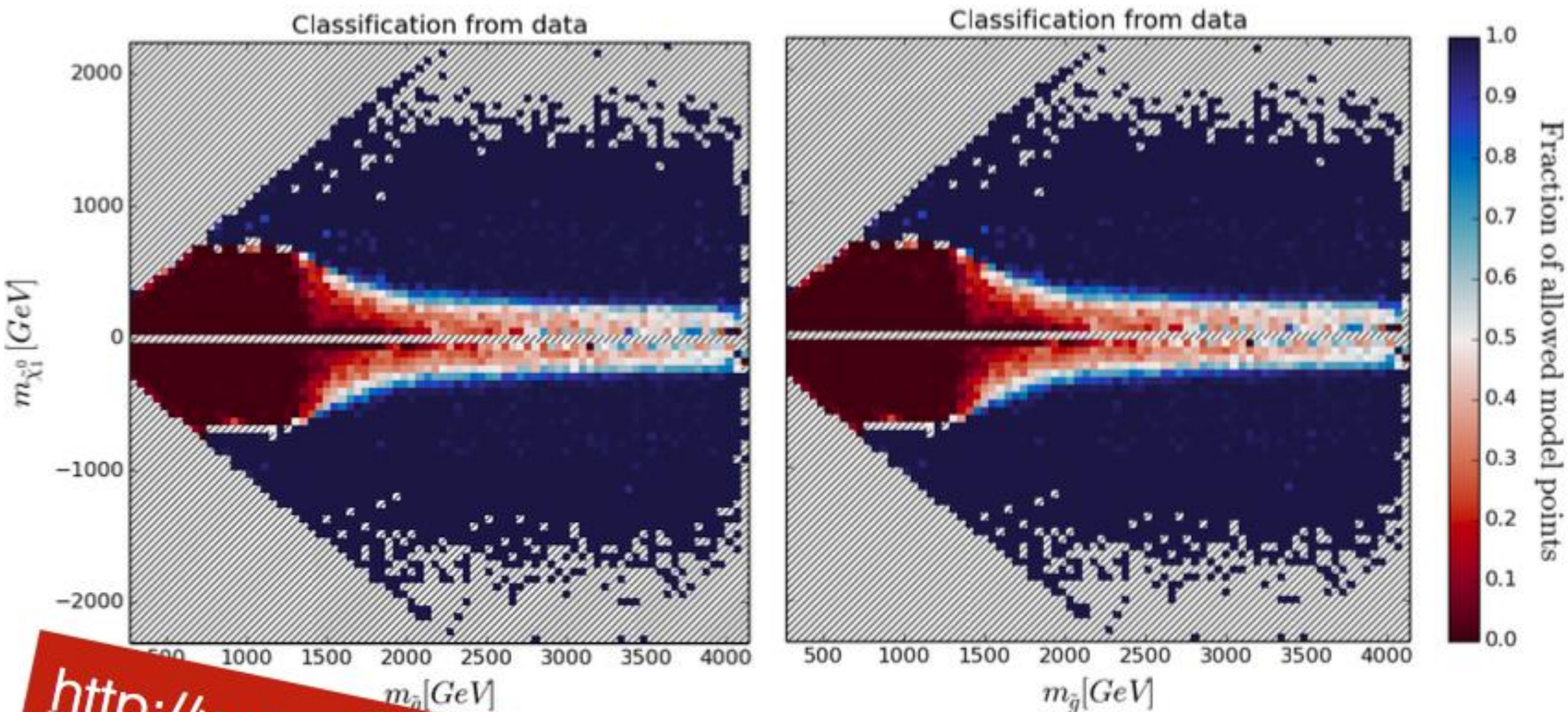
CNN branch
(32 filters, 12+4 dense)



SUSY-AI in the pMSSM (99CL)

99.7% accuracy on 51.6% of total data @ 8TeV

99.7% accuracy on 47.6% of total data @ 13 TeV



http://www.susy-ai.org for online demo!

link

Other areas

- Unsupervised Learning and Anomaly Detection
- Generative adversarial models for fast detector simulation
- Multi-class applications
- Understanding uncertainties associated with decision-making in machine learning applications

HEP Community White Paper in Machine Learning

Community White Paper

HEP Software Foundation

- HSF link
- Community White Paper
  - link to CWP
  - Machine Learning
    - Identification of challenges
    - Roadmap to address them
  - Important to think of these issues now
    - Impact on how we dedicate resources and design our software

Summary

LHC physics and computing challenges will require significant progress:

- **Higher backgrounds and pileup, data volume, unknown new physics**
  - **Machine learning offers a promising direction**
  - **An opportunity to examine new areas of ML applications to HEP**
- **IML an inter-experimental effort to foster collaboration and progress in HEP-ML**