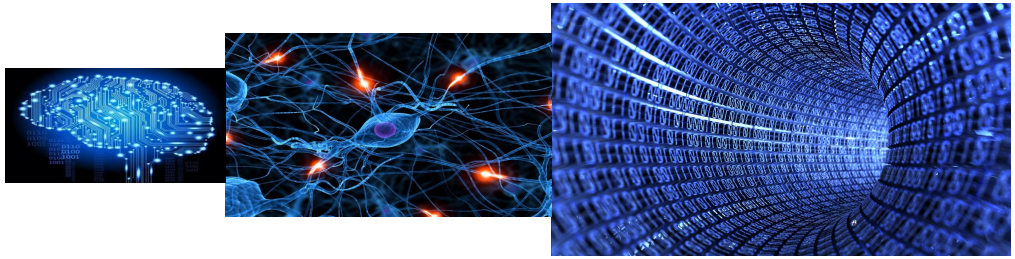
Machine Learning at the LHC:
Status, prospects and challenges

Sergei Gleyzer

University of Florida



September 12, 2018

Today's Outline

- Machine Learning in Practice
- Challenges
- Applications in HEP

Machine Learning

What is Machine Learning?

- Study of algorithms that improve their performance **P** for a given task **T** with more experience **E**

Sample tasks: identifying faces, Higgs bosons

In Computer Science

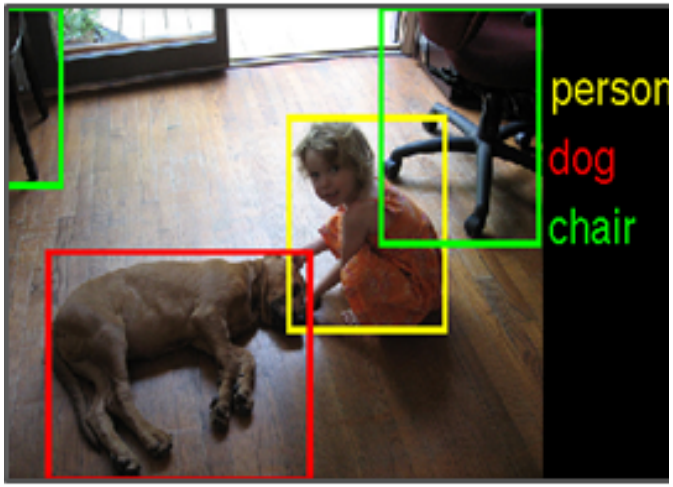
Preferred approach to:

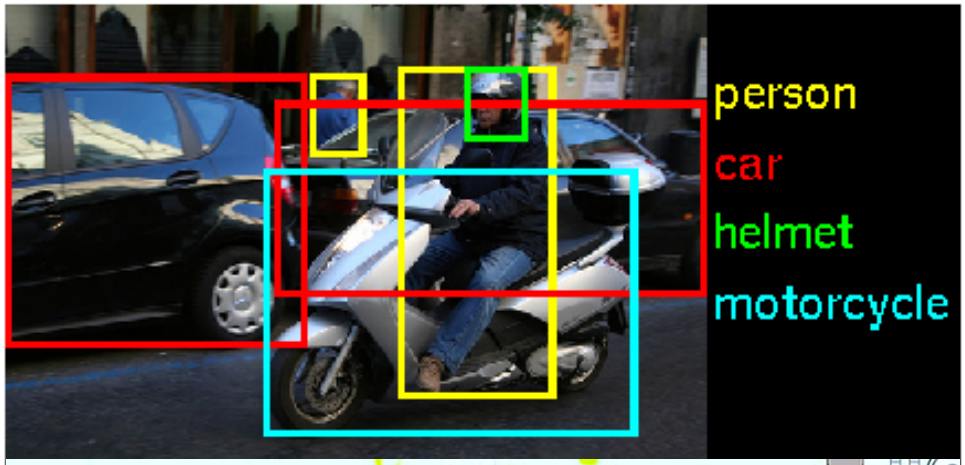
- Speech recognition, natural language processing
- Computer vision, Robot control
- Medical outcomes analysis

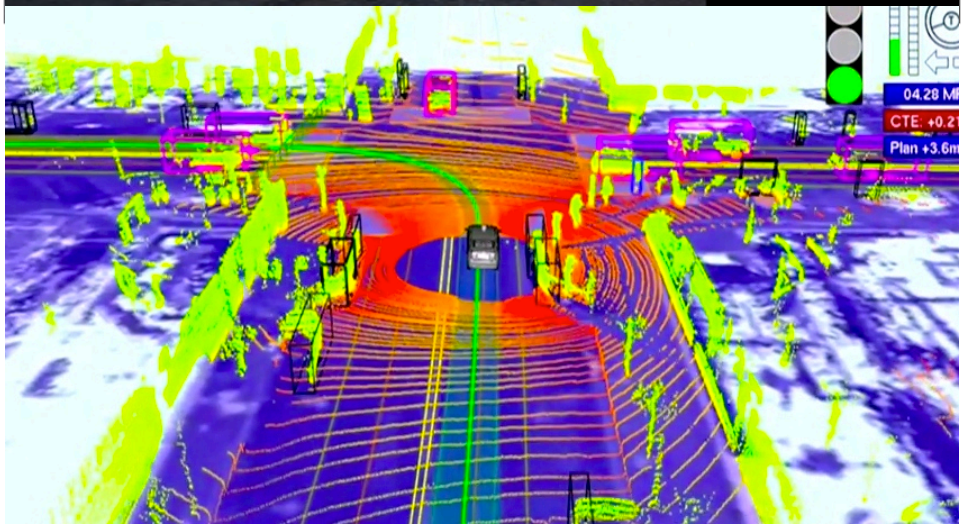
Growing fast:

- Improved algorithms
- Increased data capture
- Software too complex to write by hand

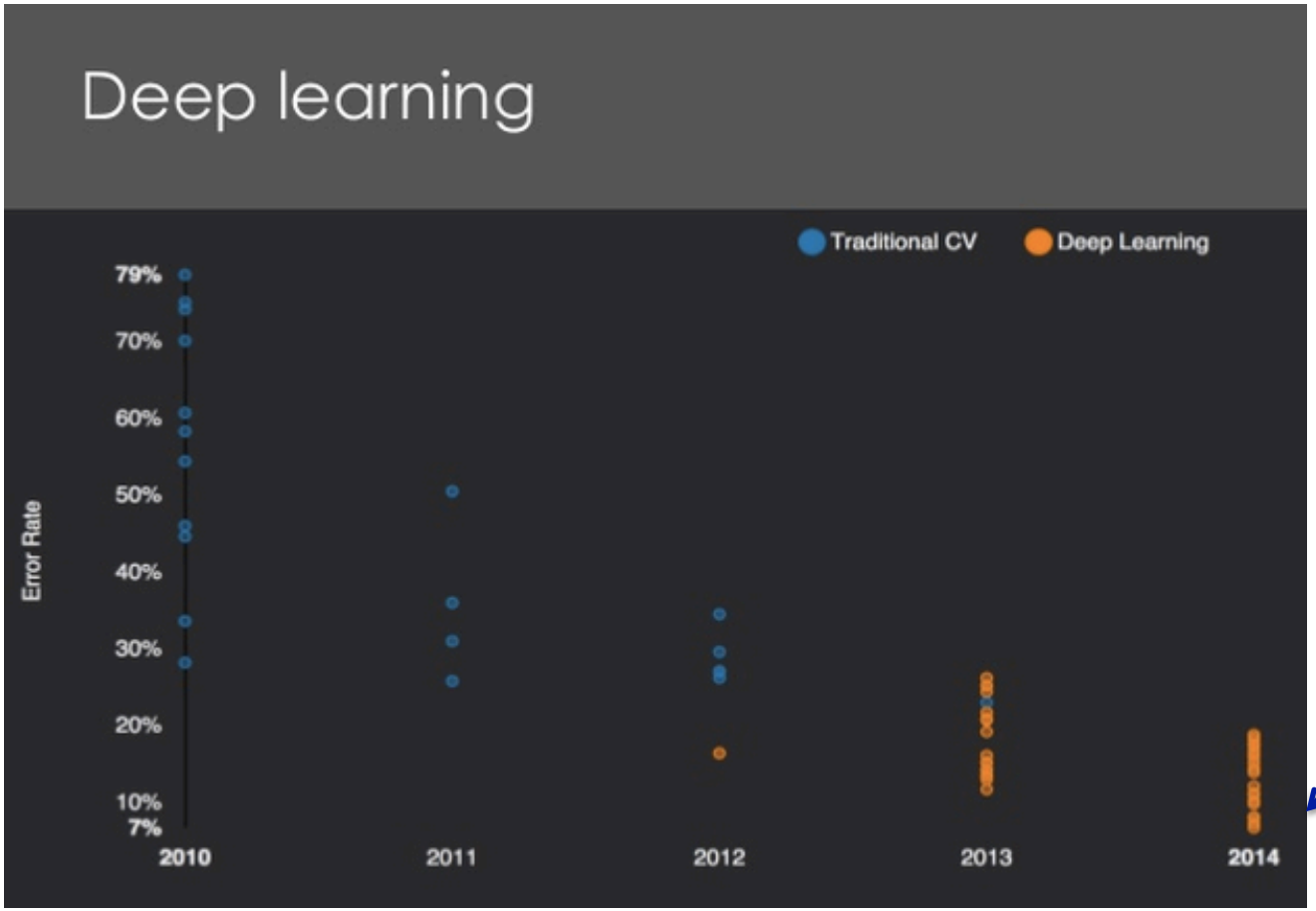
Examples





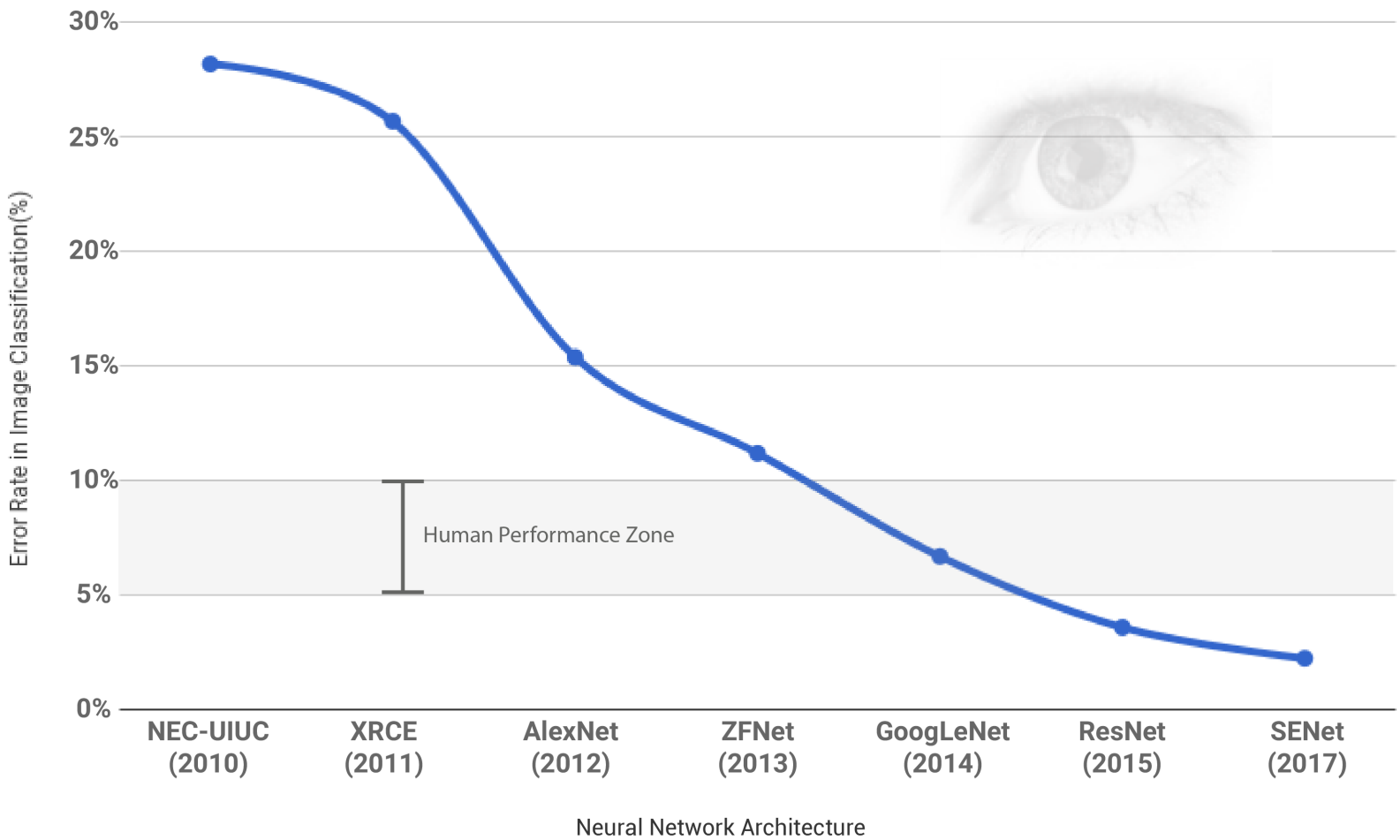


Diving Deeper

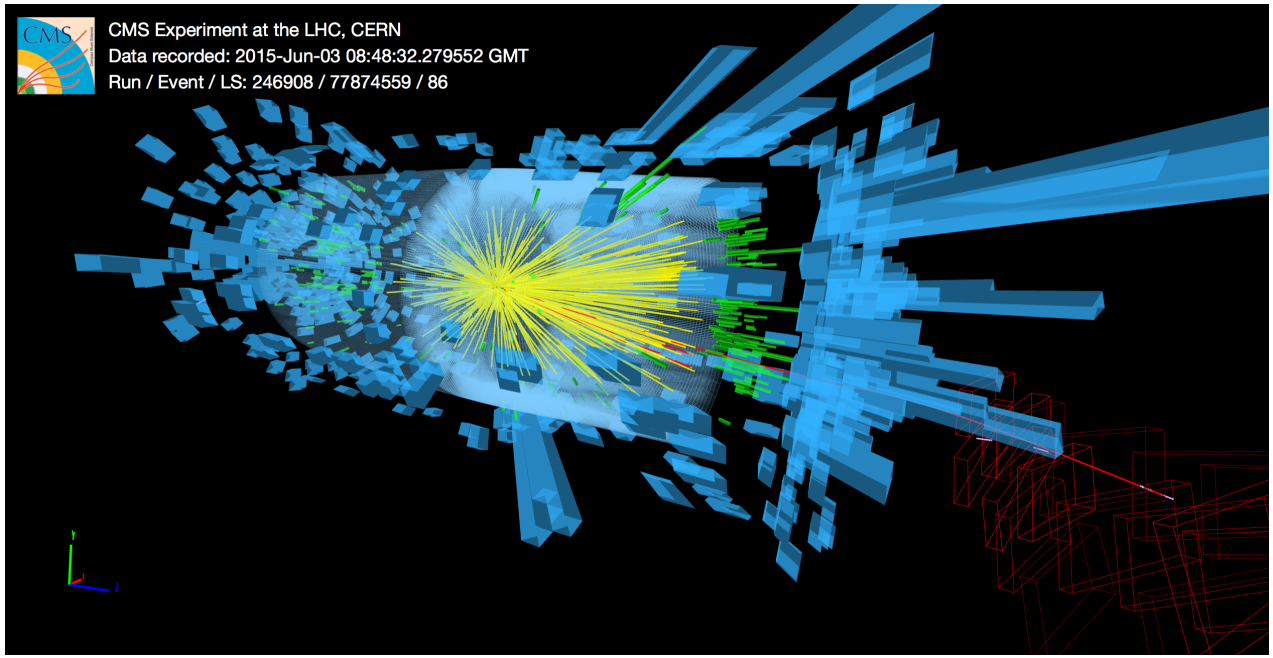


Huge Progress

Diving Deeper



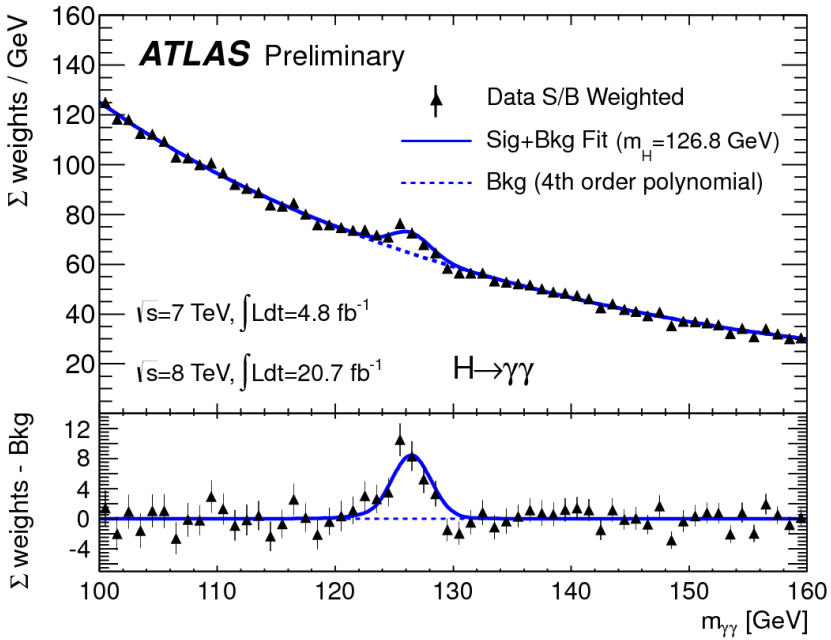
ess

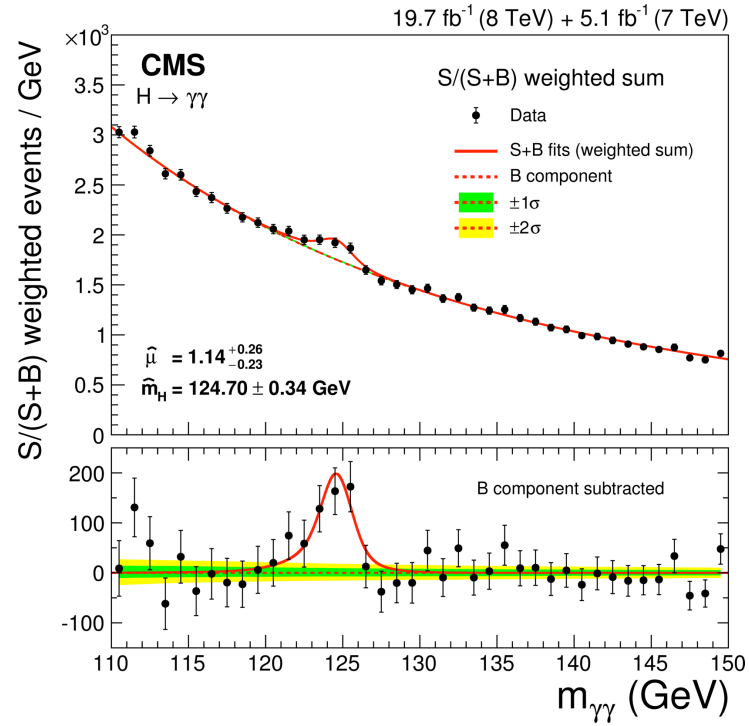


In Particle Physics

Higgs Boson Discovery

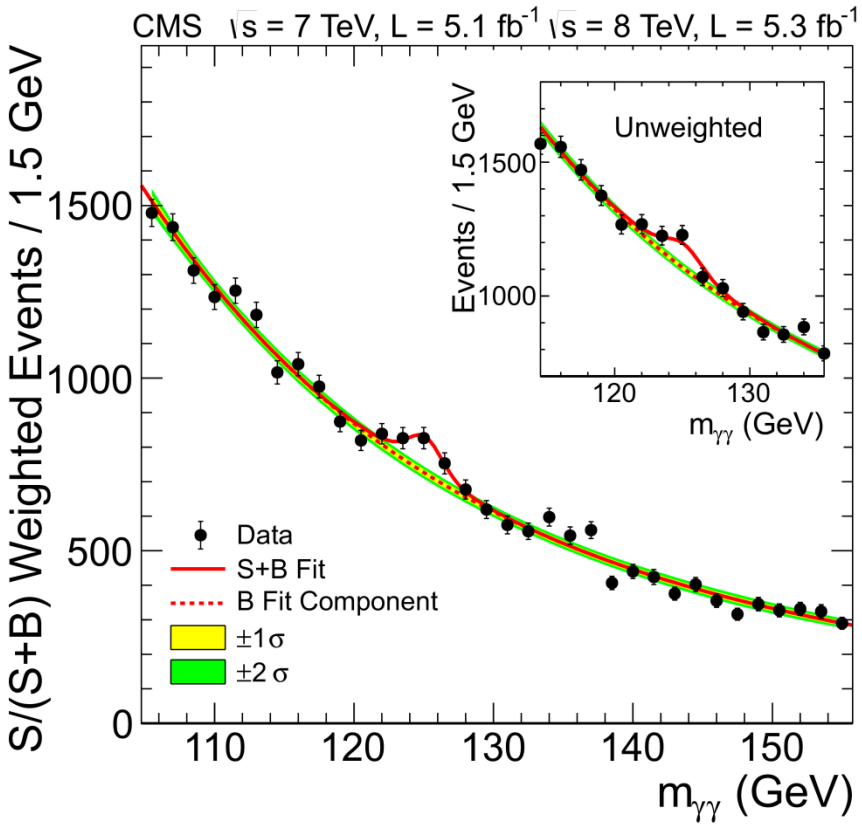
Higgs to di-photons



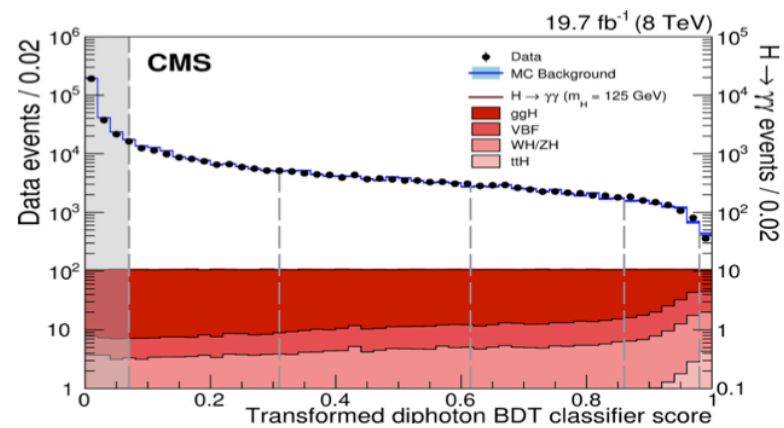


ATLAS **CMS**

In Higgs Discovery



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

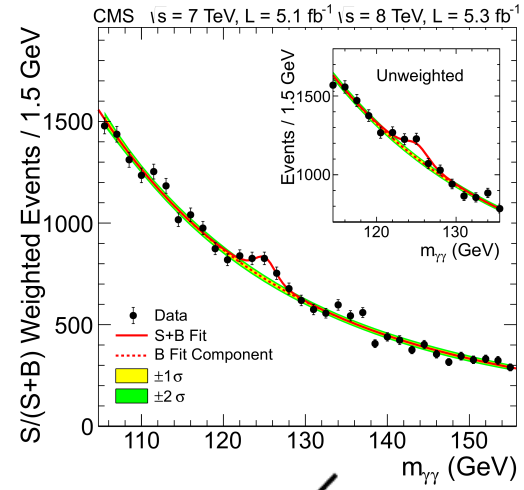


Improvement in analysis from all four areas

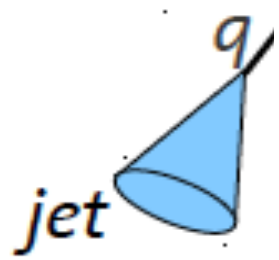
ML in HEP Today

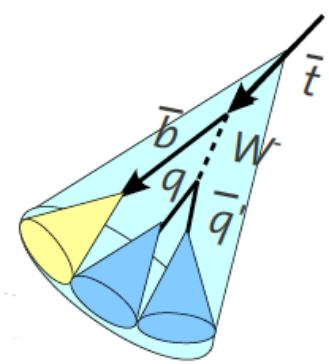
Machine learning already at forefront of what we do:

- Physics object **identification**
- Event type **classification**
- Object properties **regression**

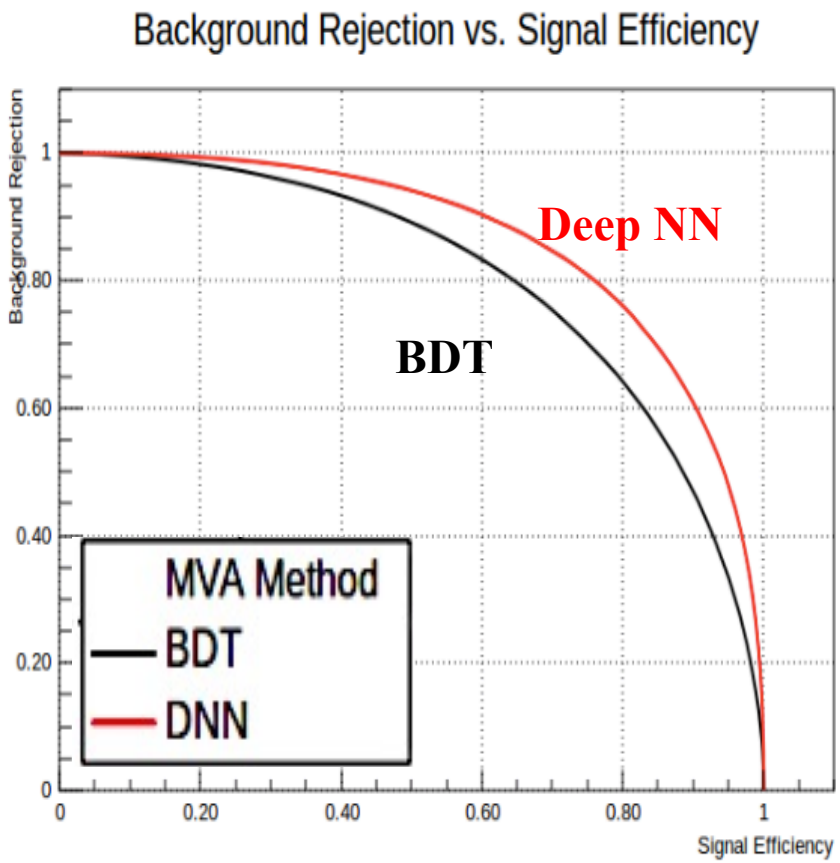


Expanding quickly





Deep Learning



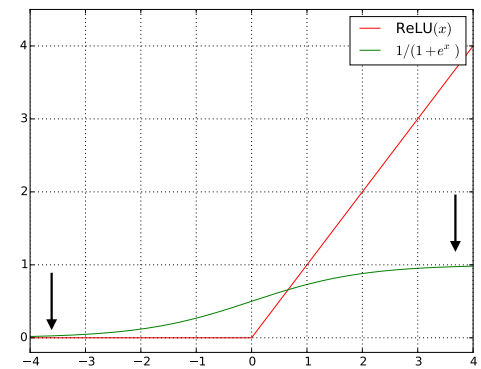
Higher performance compared to previous ML methods

Deep Learning

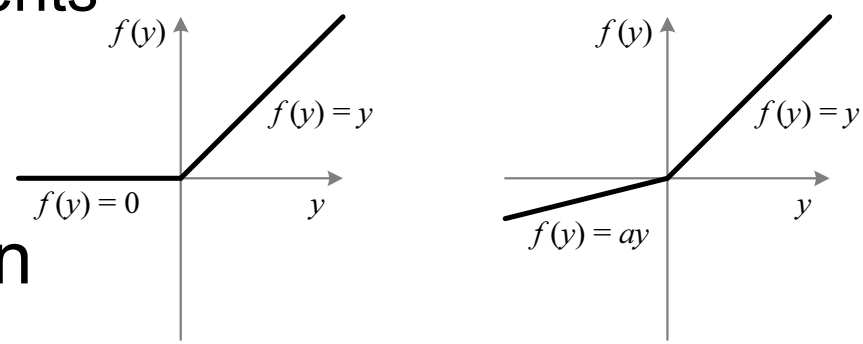
- **Training more complex models**
  - Increased Depth
  - Enlarged Width
  - Specialized Architectures: Convolutional, Recurrent, Graphs
  - Novel activation functions: ReLU
- **Effective strategies avoiding over-fitting**
  - Regularization: L1/2, Data Augmentation, DropOut

ReLU

Rectified Linear Unit (ReLU)



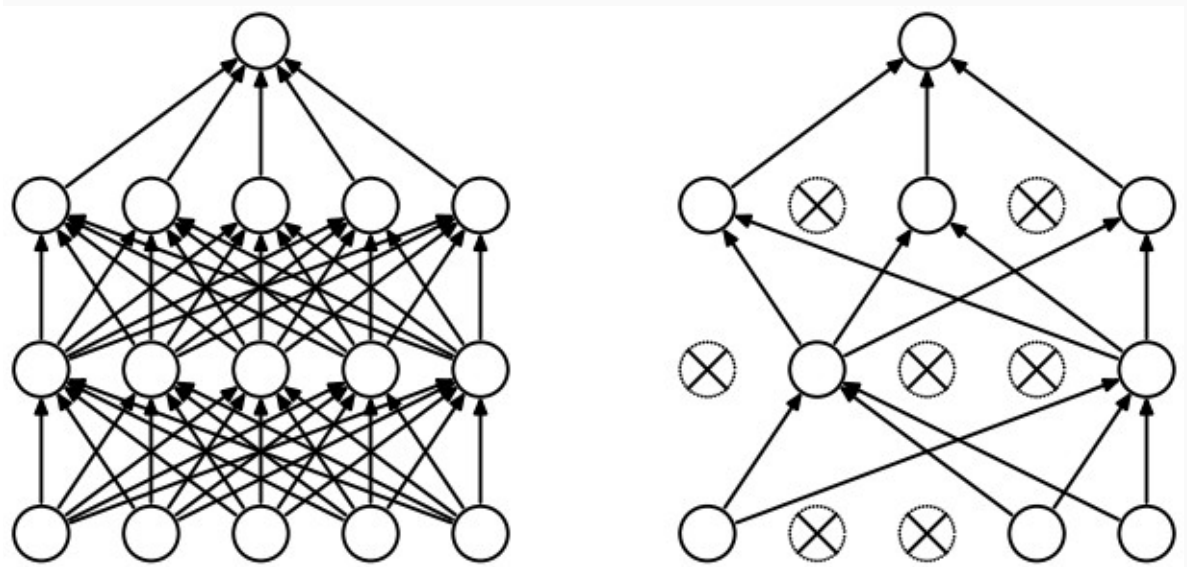
- Rectified neuron
- Faster training convergence
  - Better solutions than sigmoids
    - Vanishing gradients
  - Trained by back-propagation



ReLU and Parametric PReLU

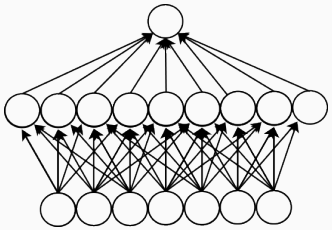
Regularization

- i.e. Drop-Out

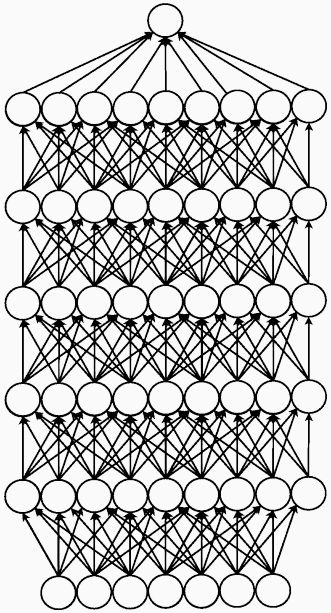


(a) Standard Neural Net (b) After applying dropout.

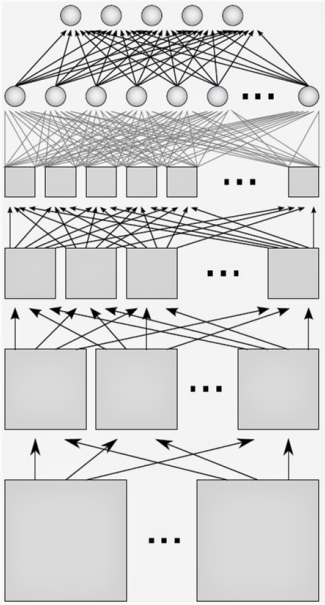
Convolutional



Neural Network (NN)



Deep NN

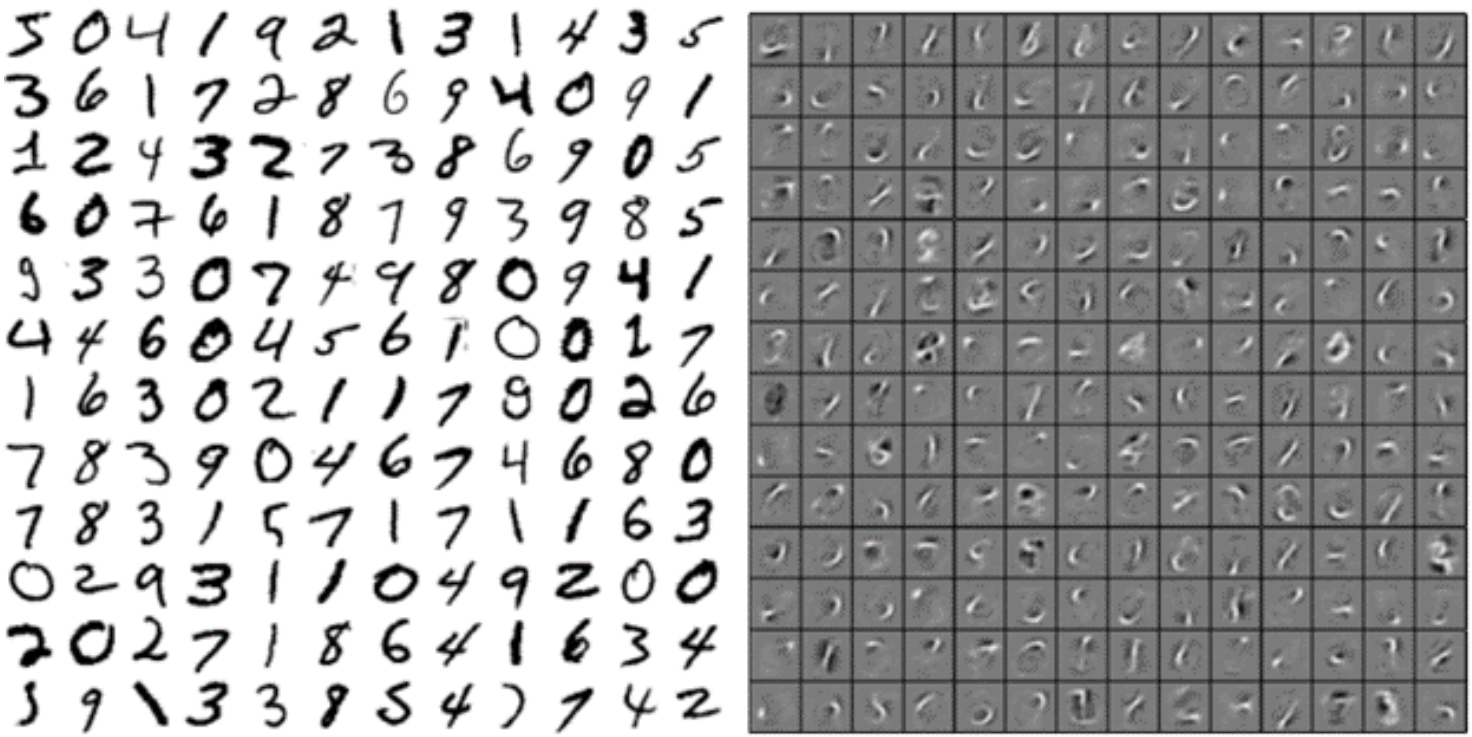


Convolutional NN

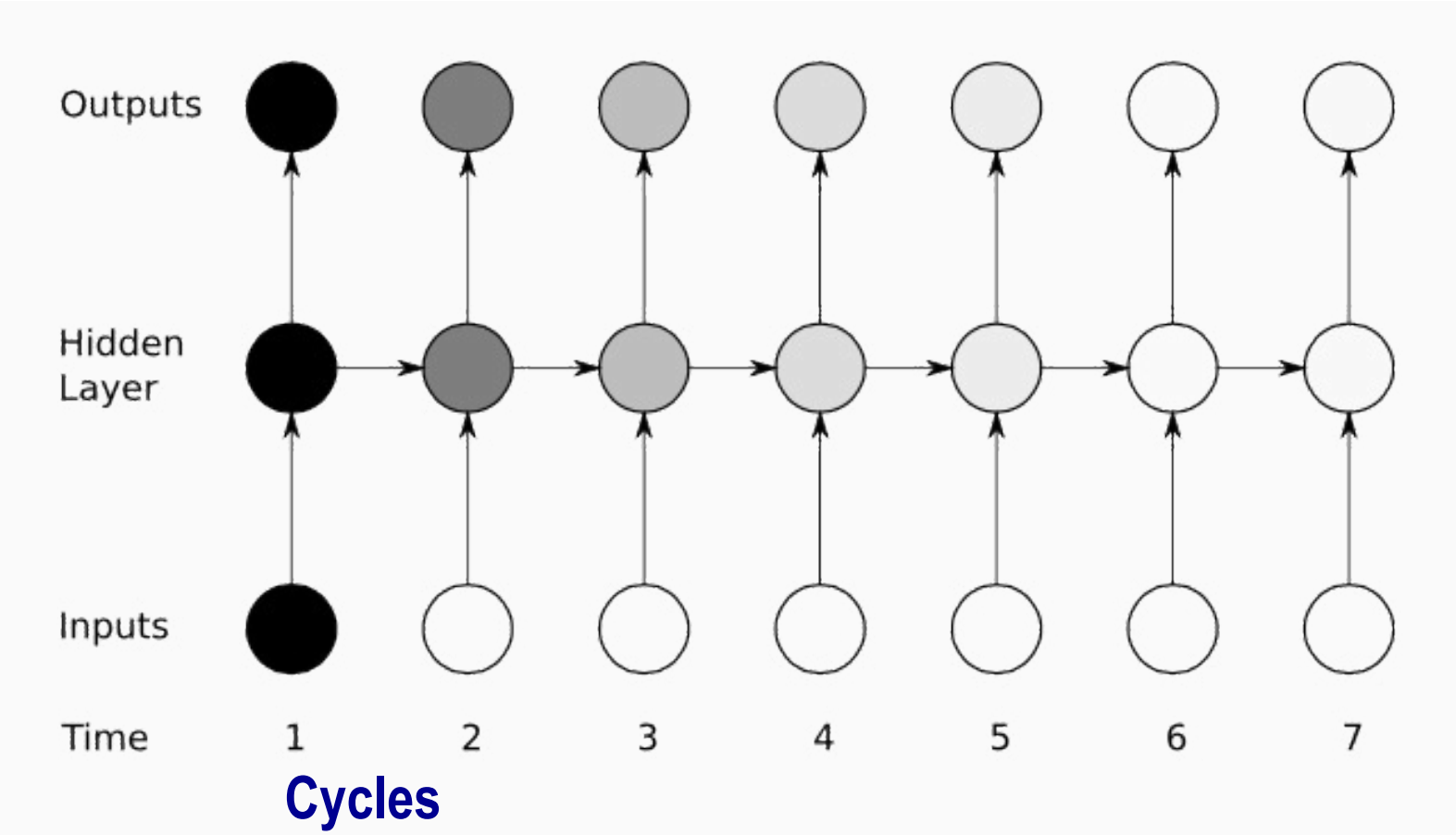
Filters

Convolutional Neural Networks:

Unsupervised Feature Learning



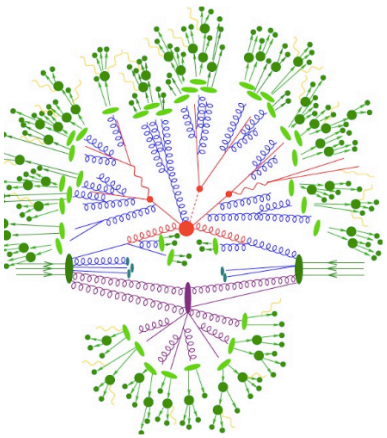
Recurrent NN

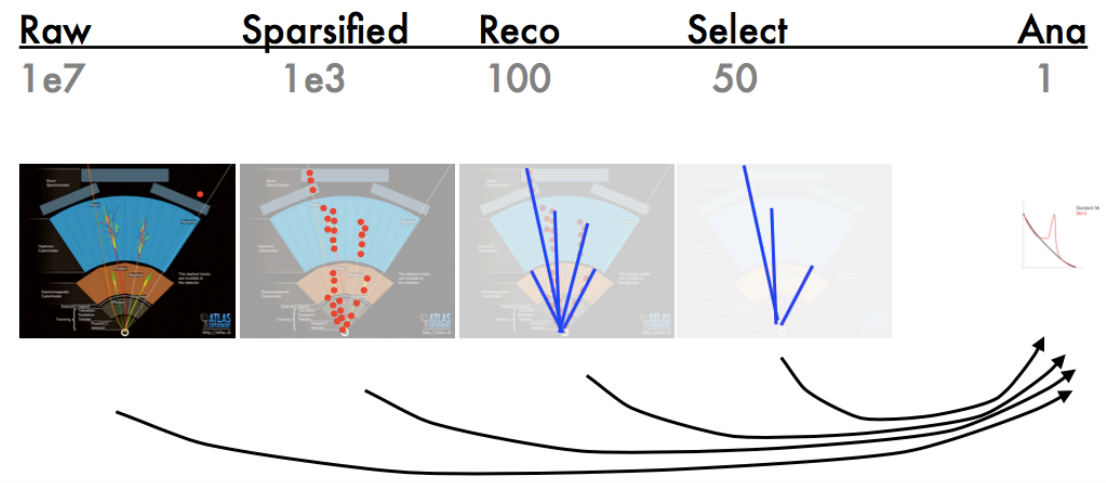


Questions

Can we fully exploit the detectors?

- Raw data, low-level variables

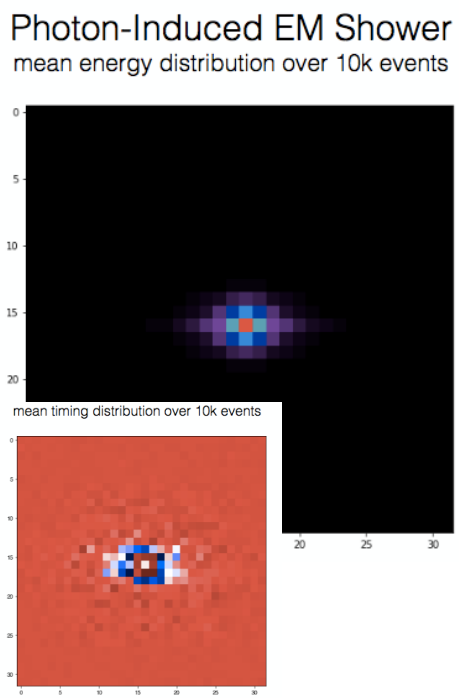


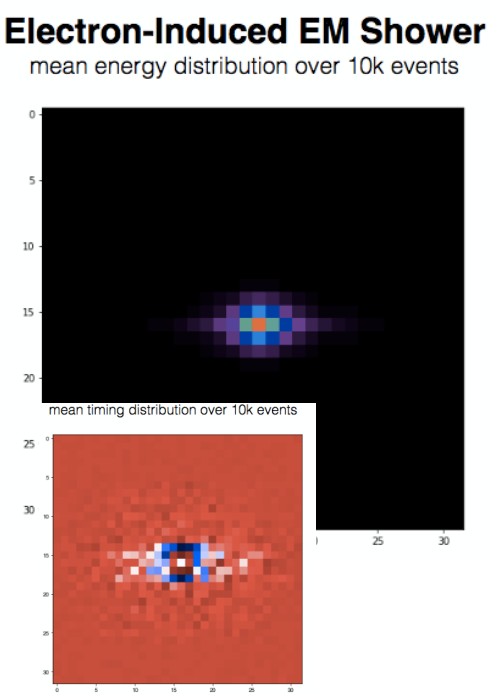


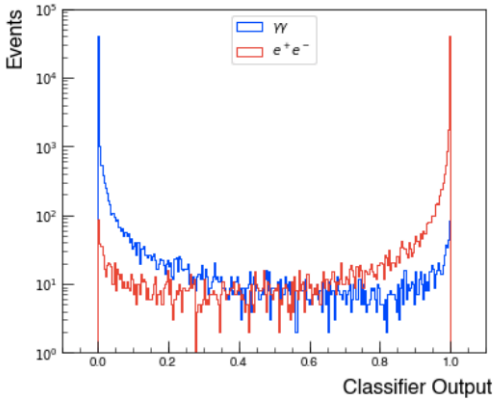
Example

"End-to-end learning" Andrews et al., 2018

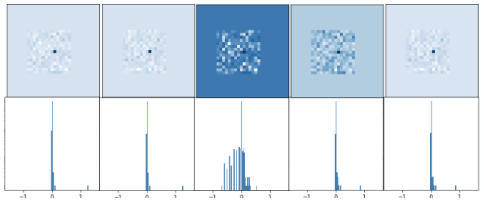
- By-passing traditional reconstruction





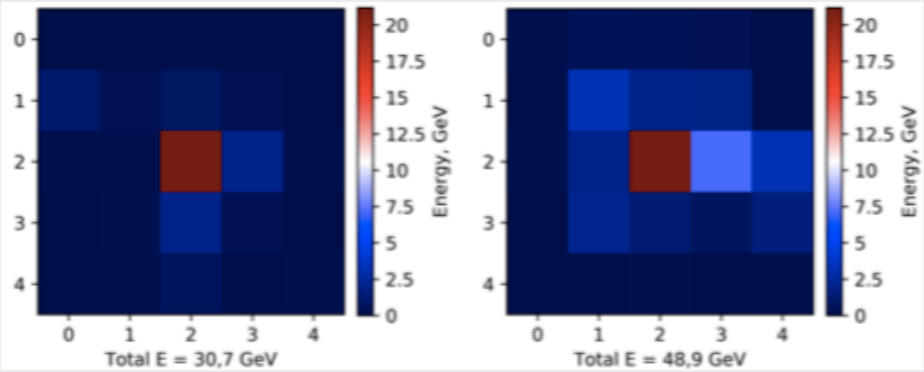


	ResNet-23
Test Set ROC AUC	0.997



Particle ID

Clusters for photon and π^0photon

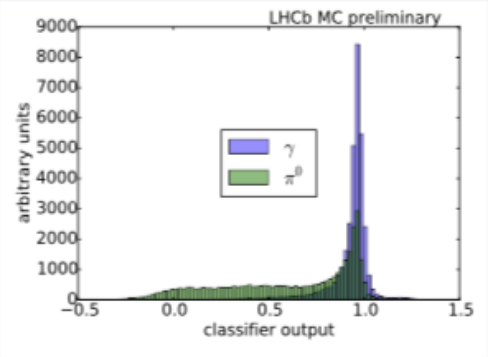


Responses from single photon(left) and merged π^0(right)

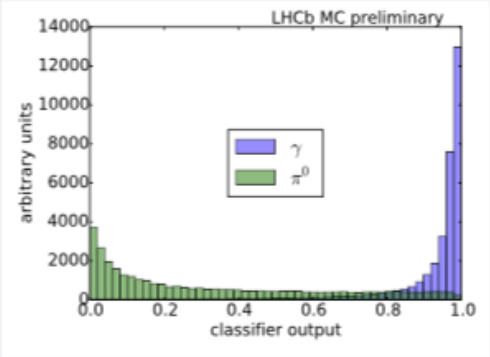
LHCb, CHEP2018

New approach

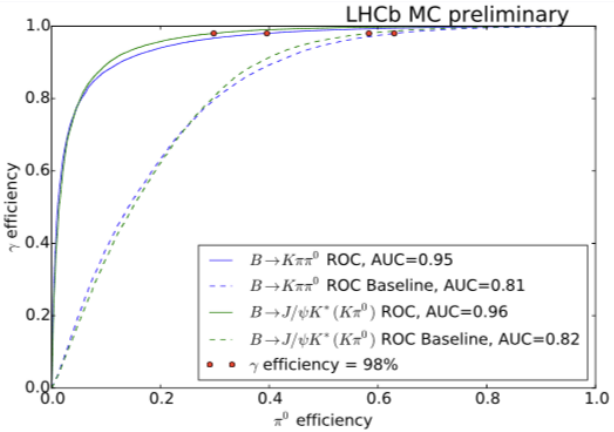
- Consider 5*5 cluster
- Use energy in each cell as a feature
- Use several models and look for the best one



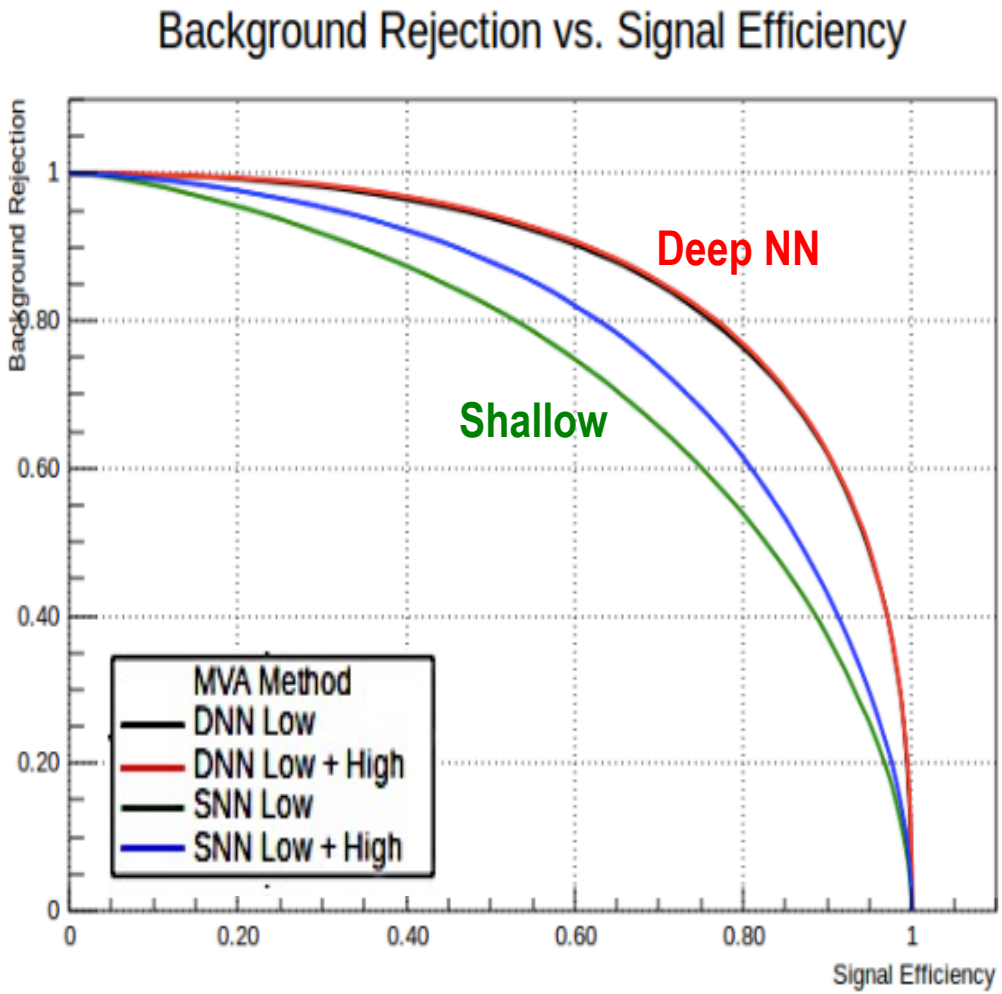
Baseline output

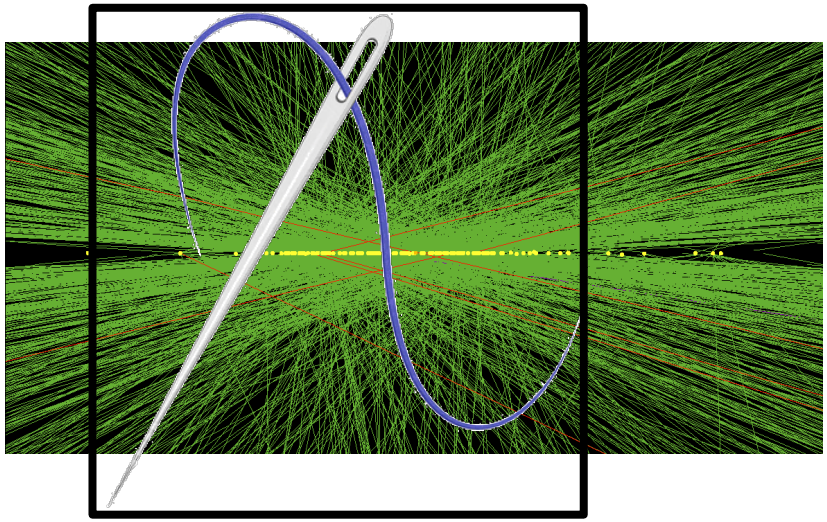


XGBoost approach output



Feature Extraction

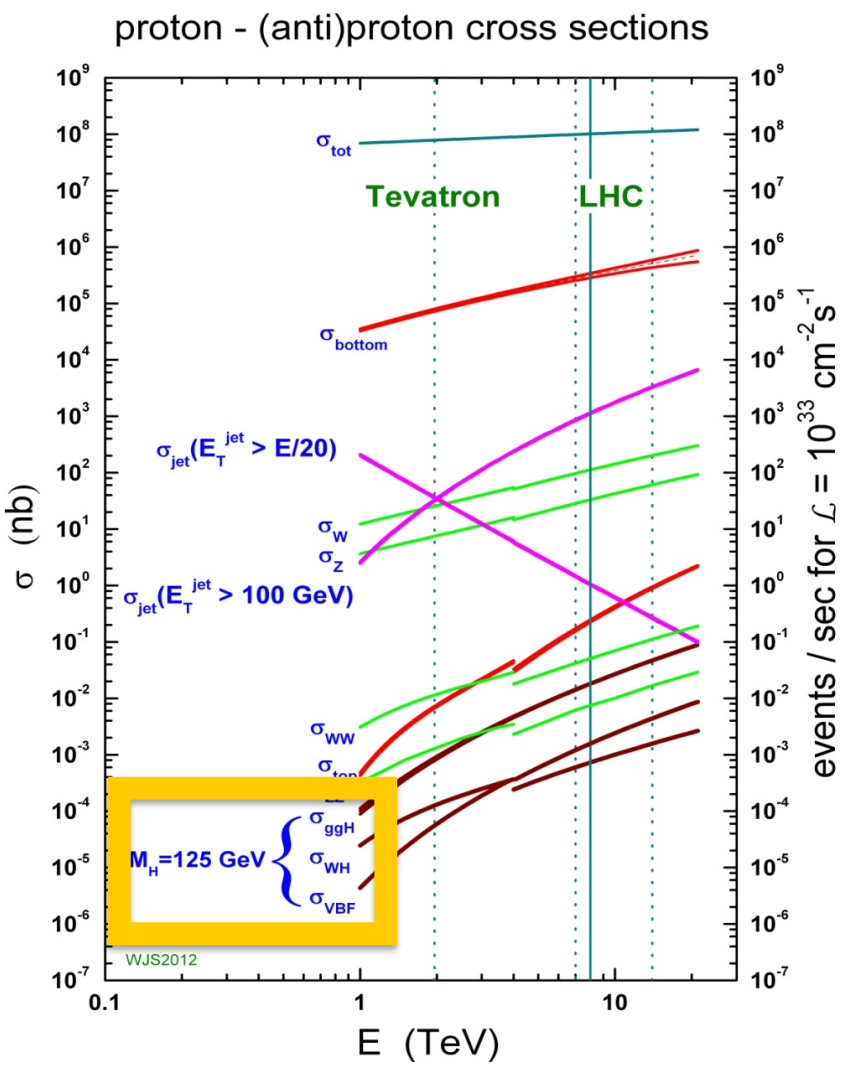




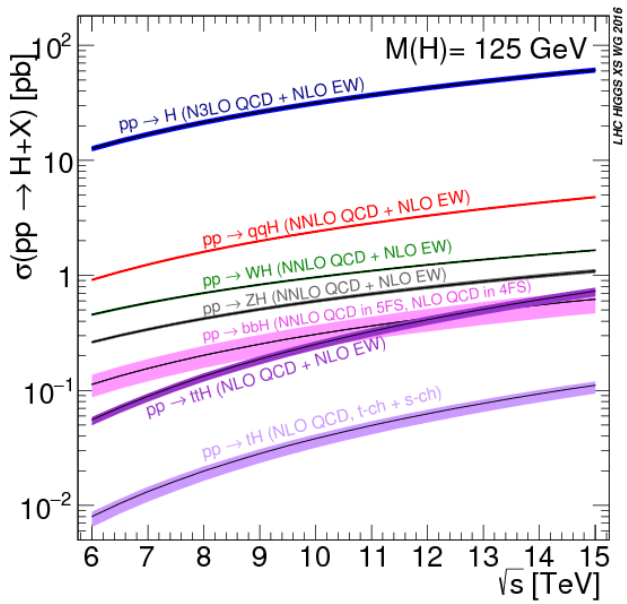
Present and Future Challenges

2

Multi-scale challenge



Orders of magnitude between signals and backgrounds



Observations

New Higgs Boson Observations Reveal Clues on the Nature of Mass

For the first time the scientists have observed the famous Higgs boson, responsible for imparting mass, interacting with the heaviest particle in the universe

By Don Lincoln on June 6, 2018

UF

CMS

SUBSCRIBE

SCIENTIFIC AMERICAN

English Cart 0 Sign In | Register

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

Purch

PHYSICS

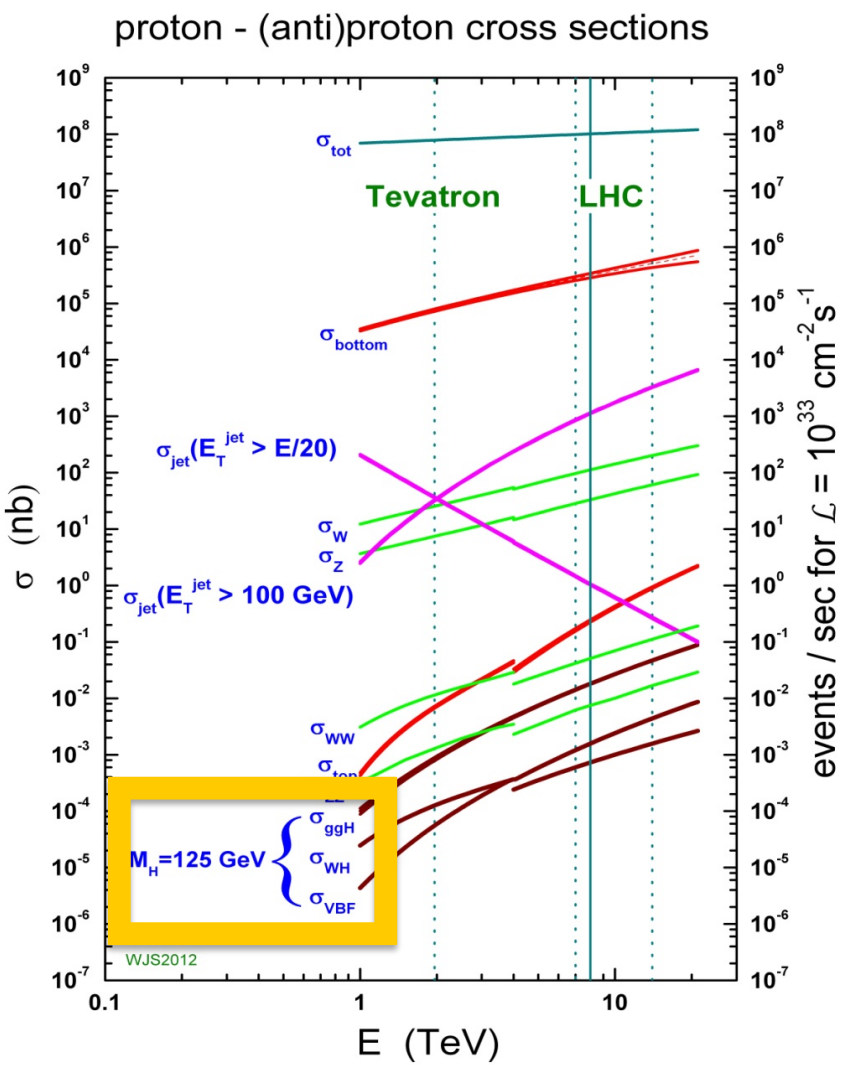
Physicists Observe the Higgs Boson's Elusive Decay

Directly detecting this long-predicted phenomenon further validates the Standard Model of particle physics

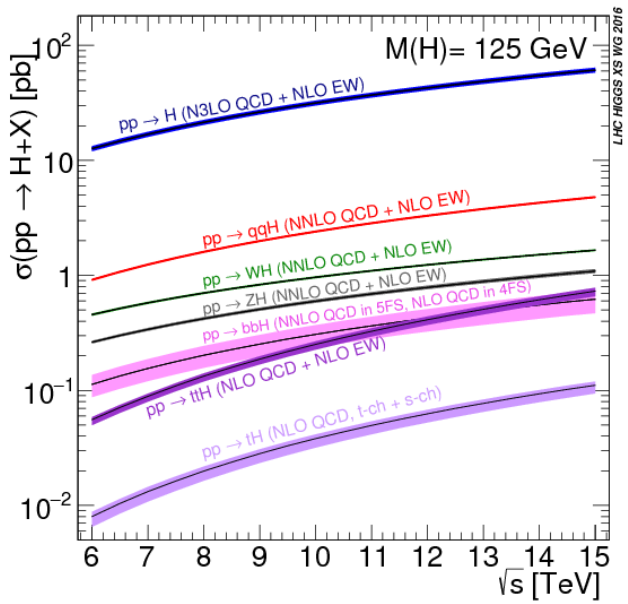
By Chelsea Gohd, SPACE.com on August 29, 2018

E (TeV)

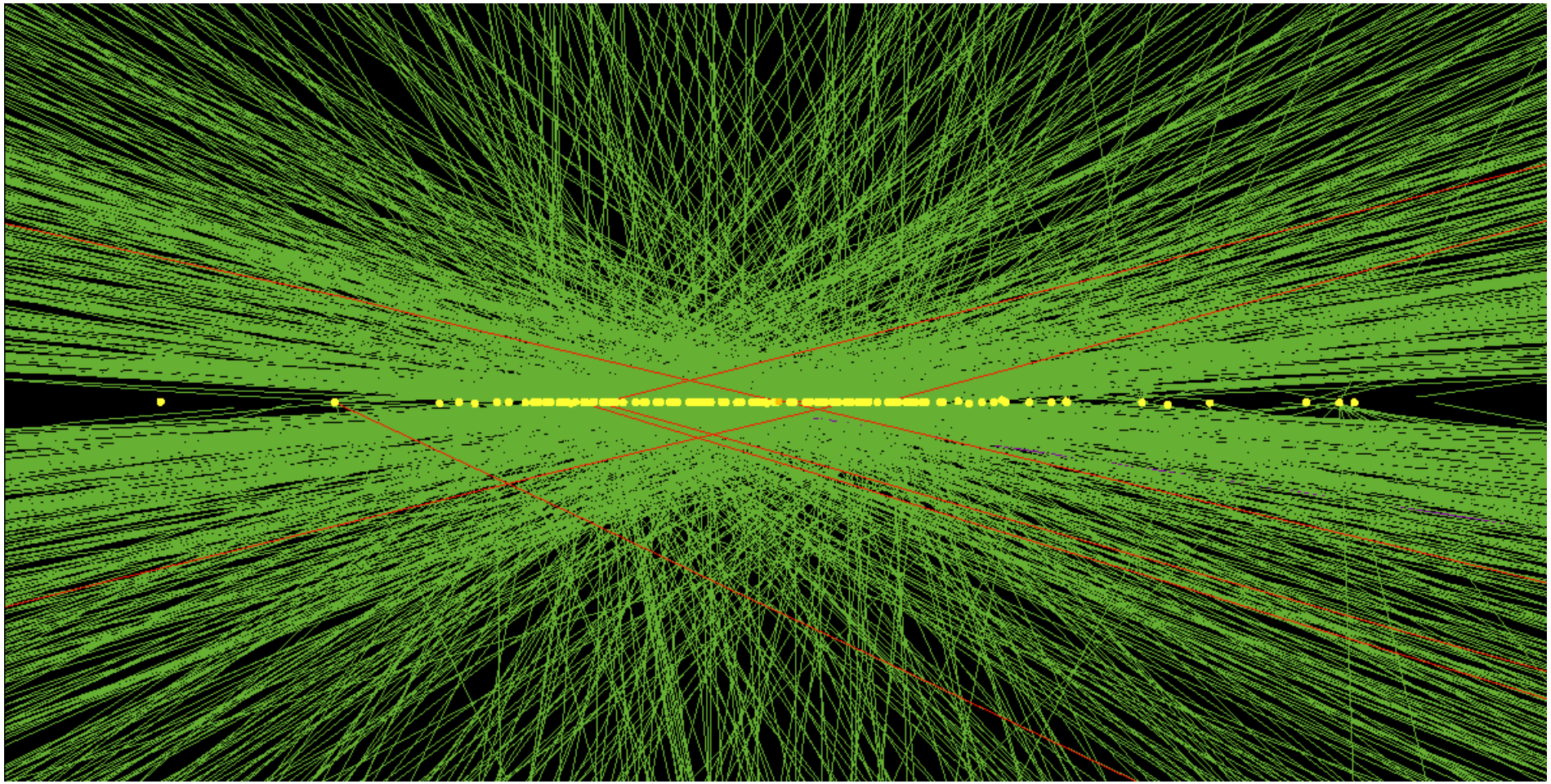
Multi-scale challenge



Orders of magnitude between signals and backgrounds

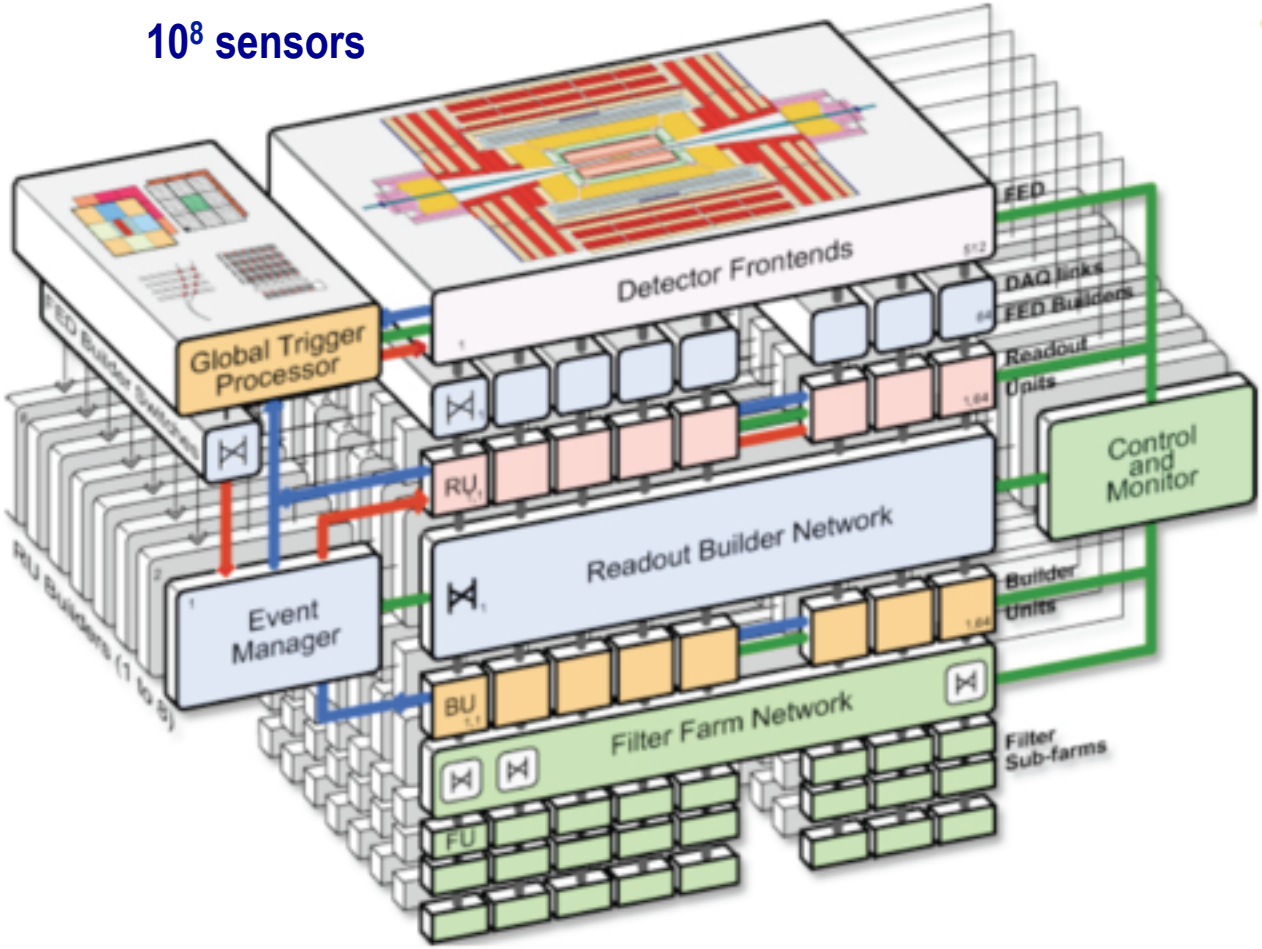


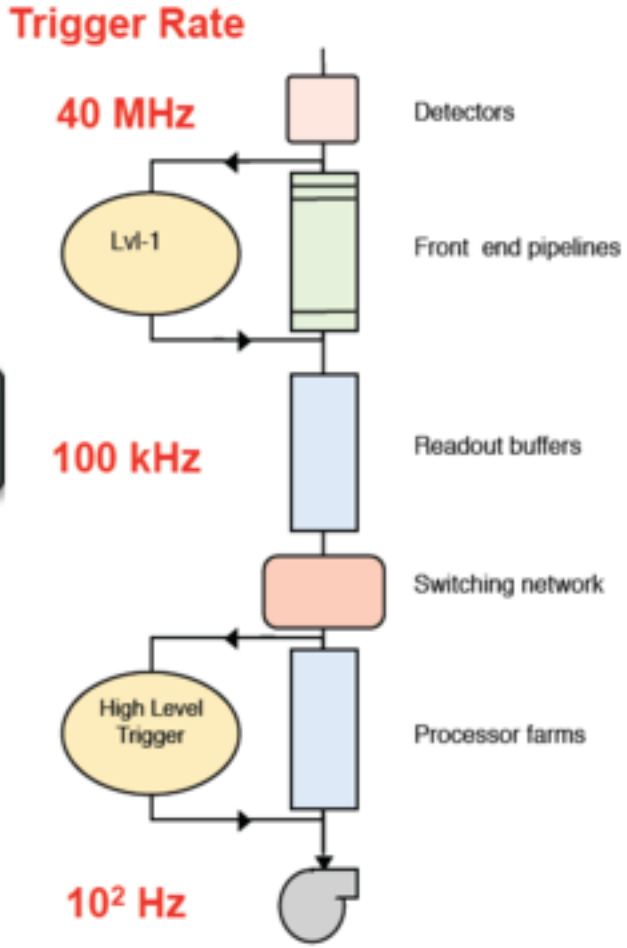
Event Complexity



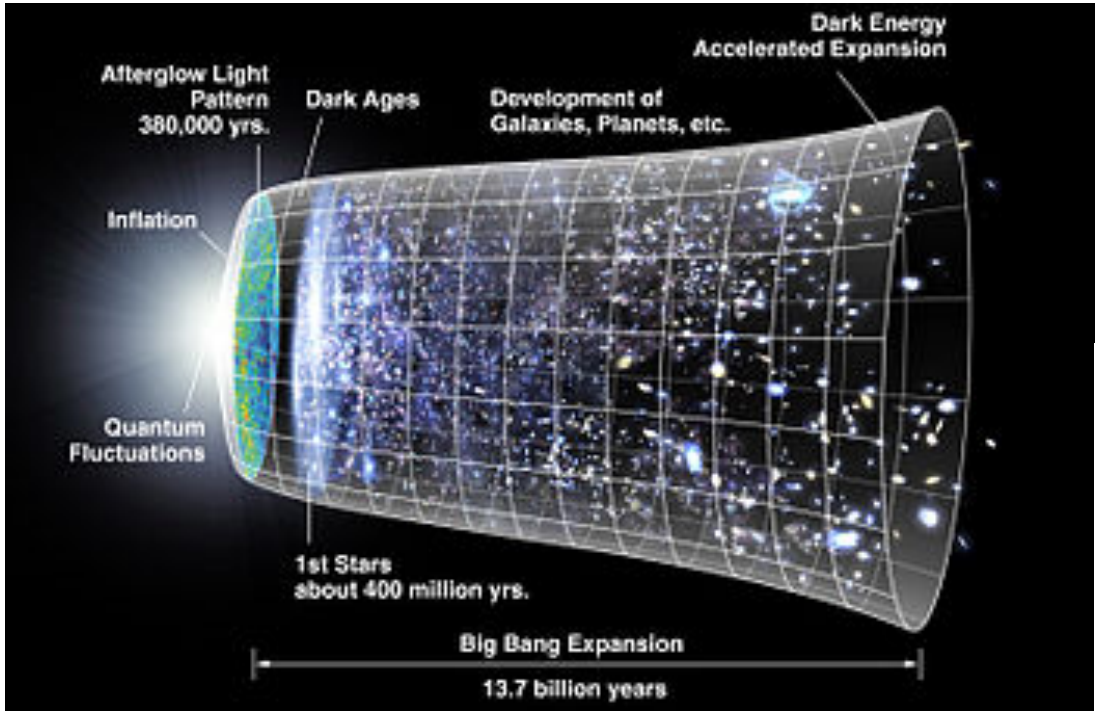
Trigger

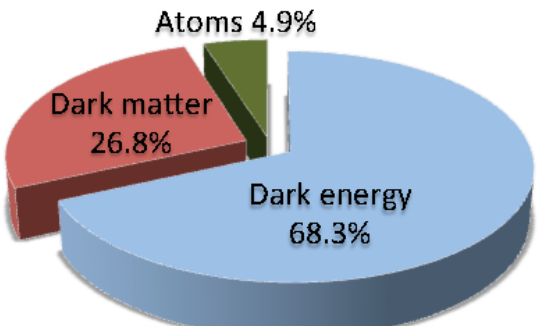
10^8 sensors

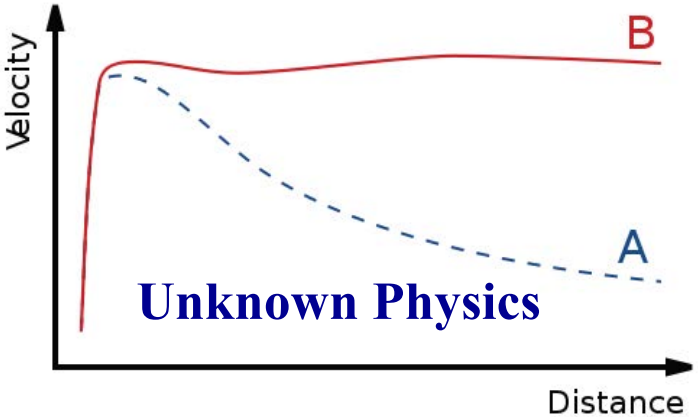




Upcoming Challenges

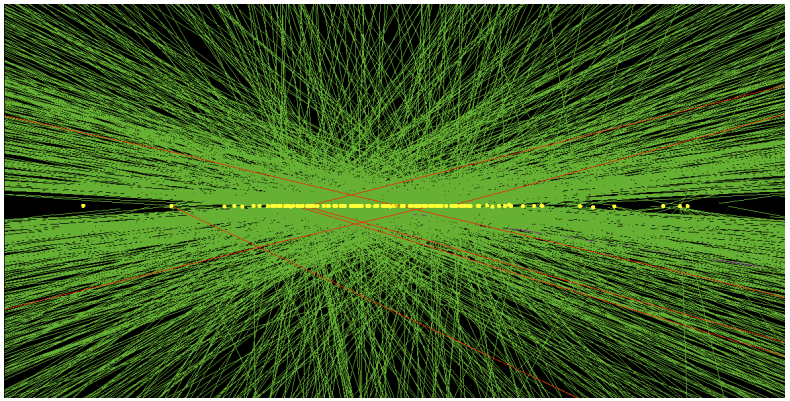






Data size:

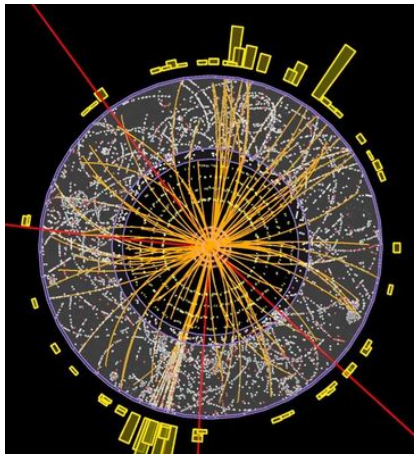
- **LHC 15,000,000 Tb 2010 – 2035**
- **Resources not up as fast as data volume**



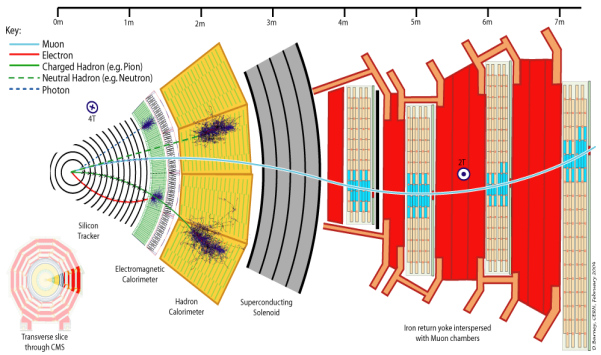
HEP Applications

3

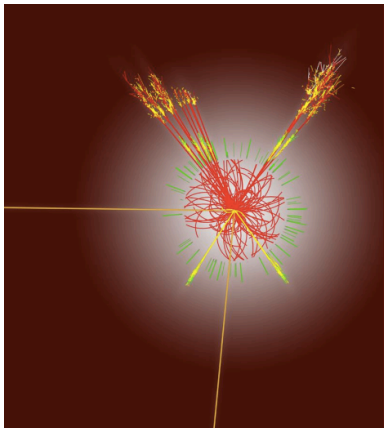
Interesting areas



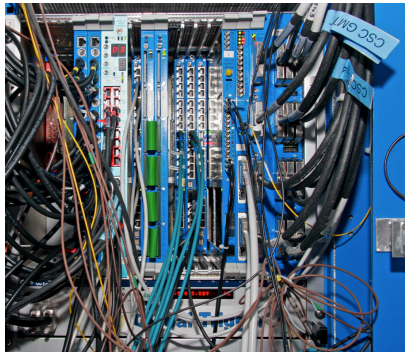
Tracking



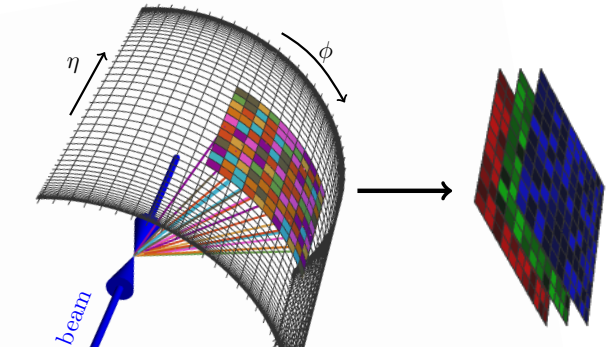
Fast Event Simulation



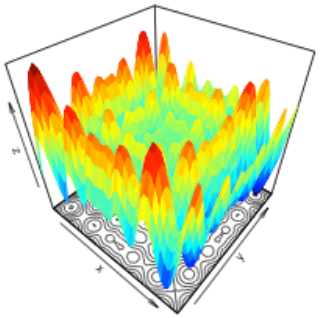
Object Identification



Trigger

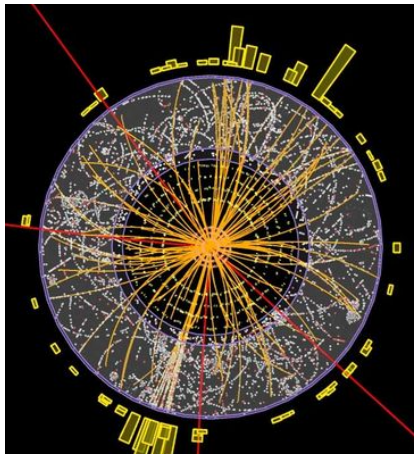


Imaging Techniques

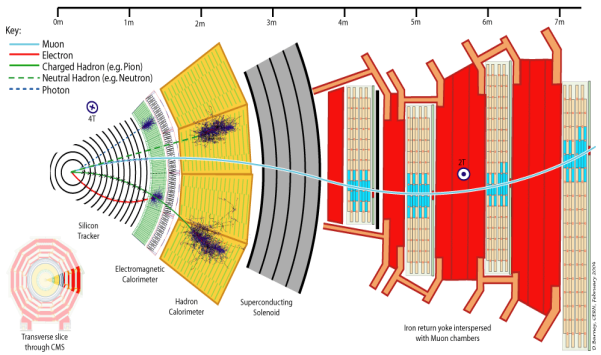


Simulation

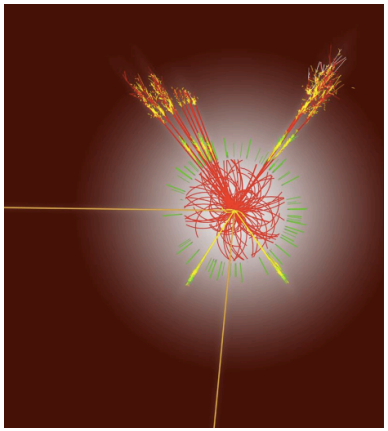
Interesting areas



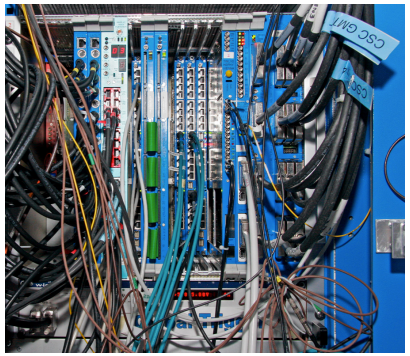
Deep Kalman, LSTMs, GNN



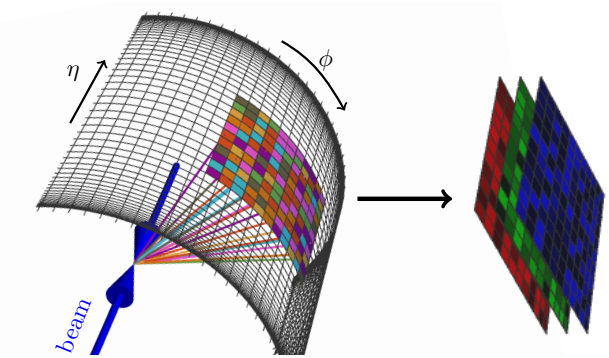
Generative Models, Adversarial Networks



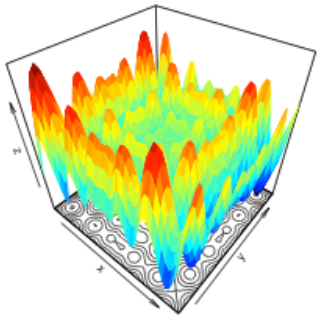
FCN, Recurrent, LSTMs



Deep ML +FPGA



Convolutional NN



Multiobjective Regression

Defining the Problem

If a problem can be expressed as a known problem

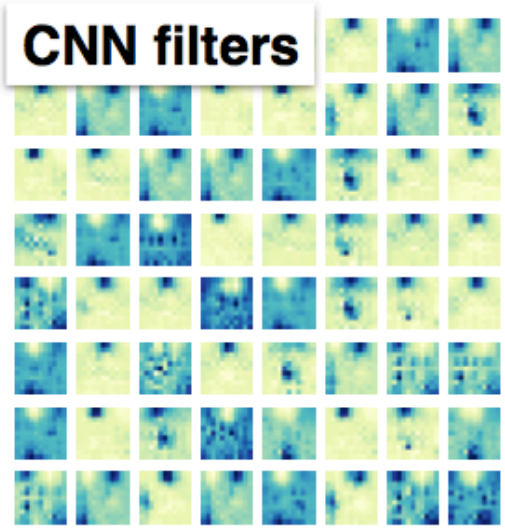
- Apply **existing algorithms**
  - **Example: convolutional neural networks from computer vision**

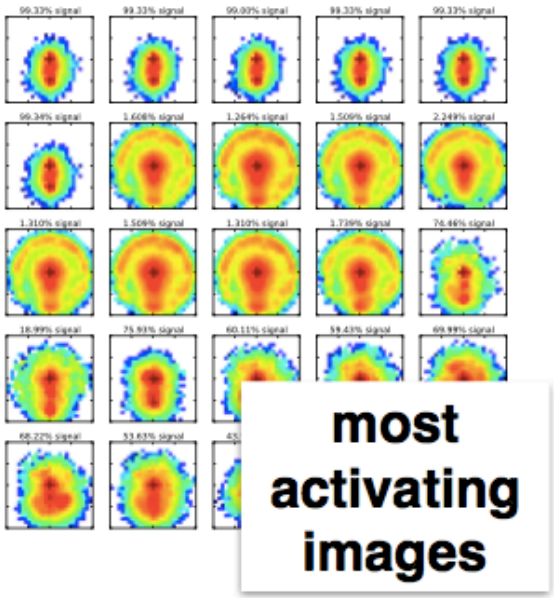
If a problem has not been solved

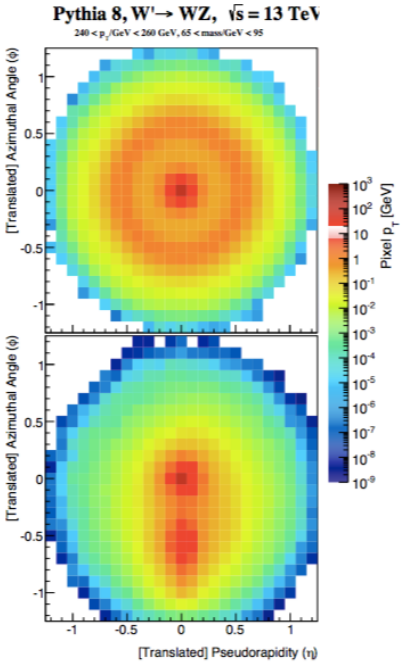
- Push the knowledge boundary forward

Early Examples

Jet images with convolutional nets



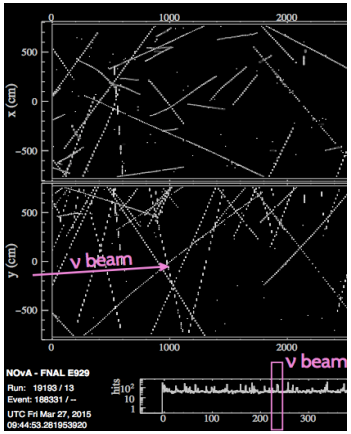


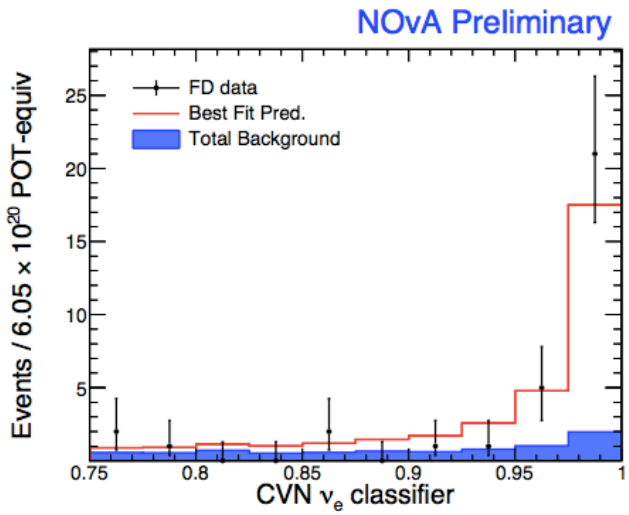


L. de Oliveira et al., 2015

Examples

Neutrinos with convolutional nets



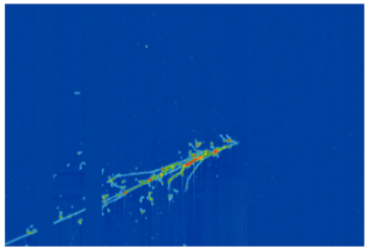


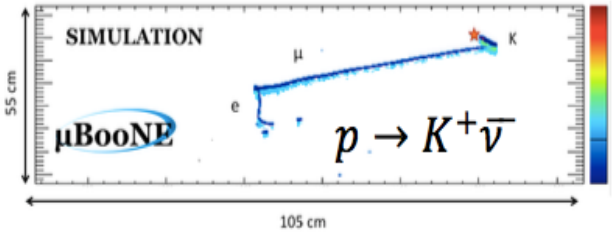
76% Purit

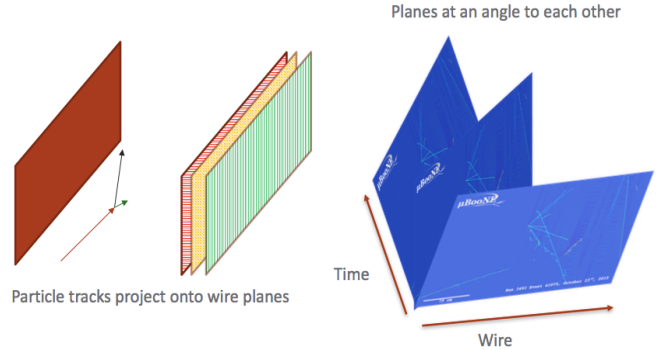
73% Effic

An equivalent increased exposure of 30%

Aurisiano et al. 2016

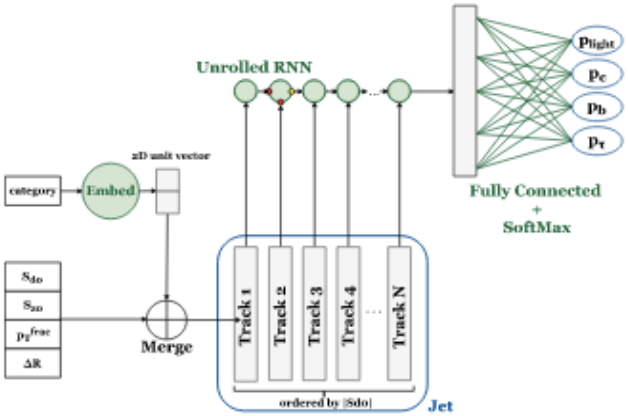


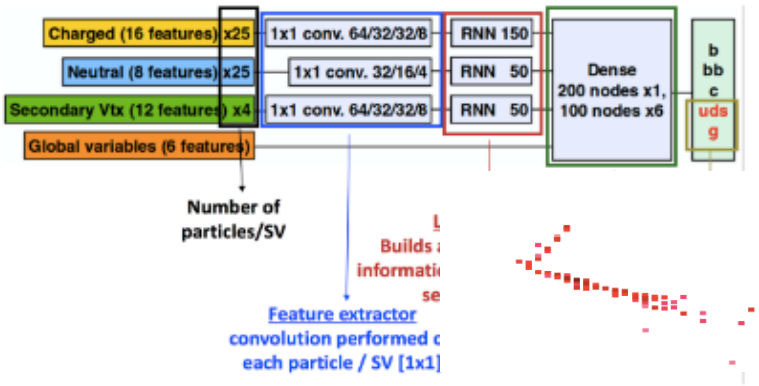


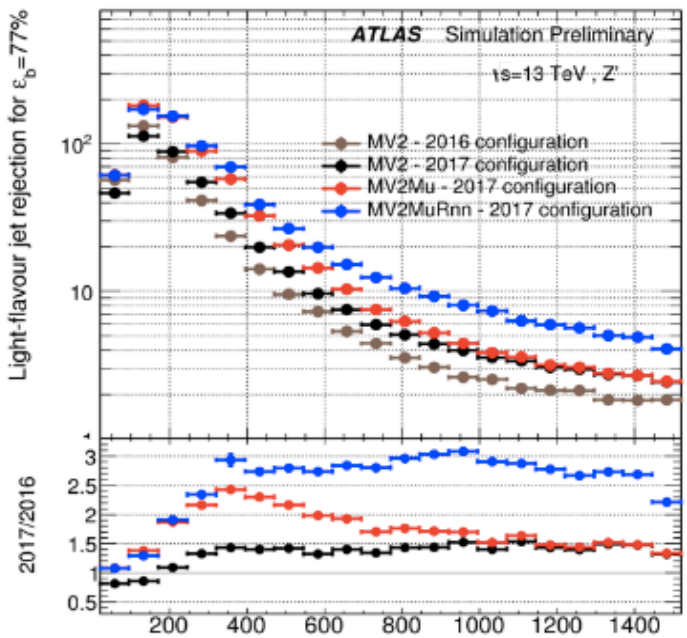


μBooNE

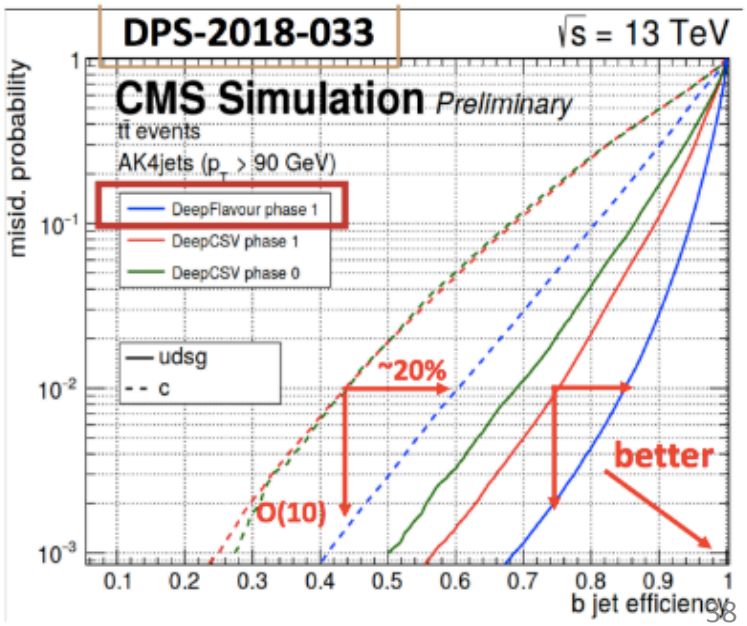
Flavour Tagging





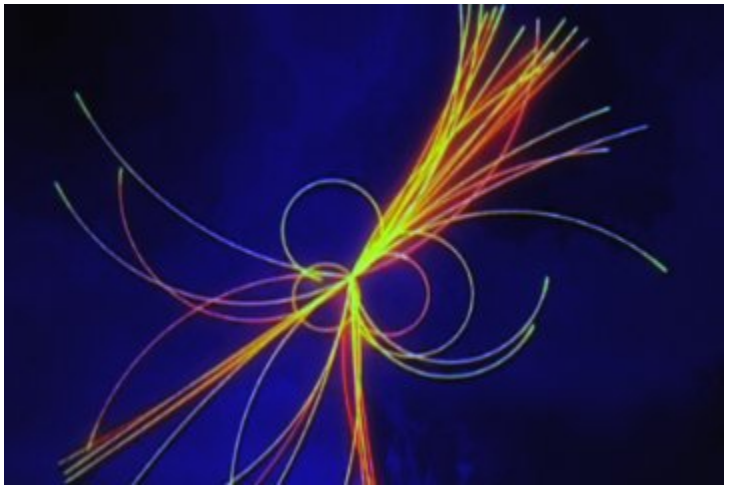


ATL-PHYS-PUB-2017-013



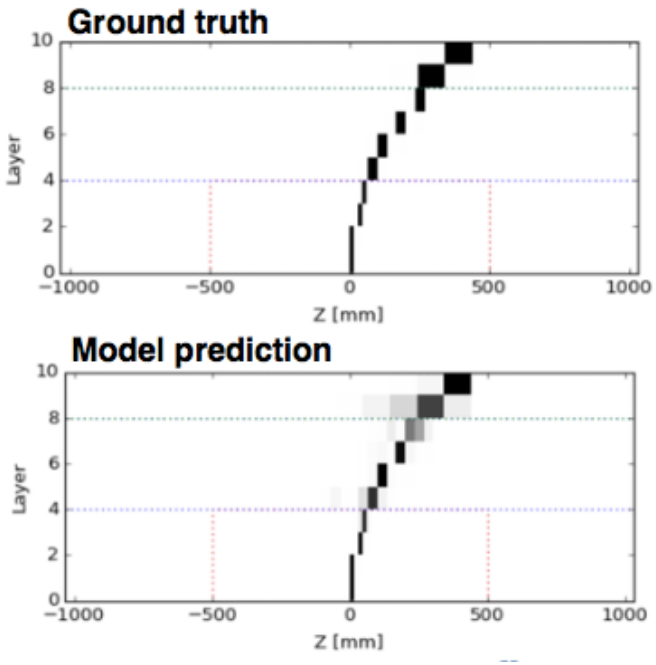
From L. Gouskos, BOOST 2018

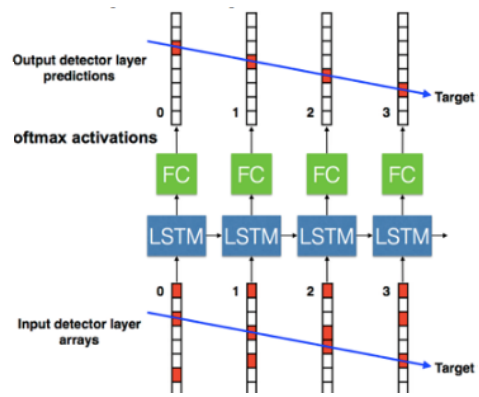
Tracking



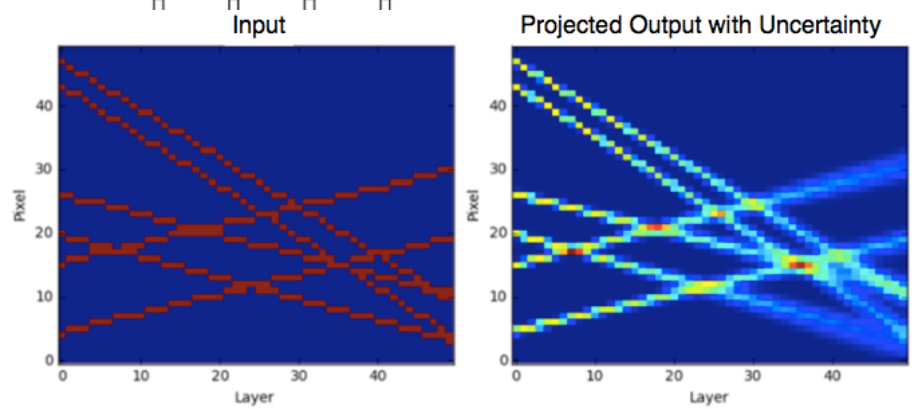
Examples

Tracking with recurrent nets (LSTM)





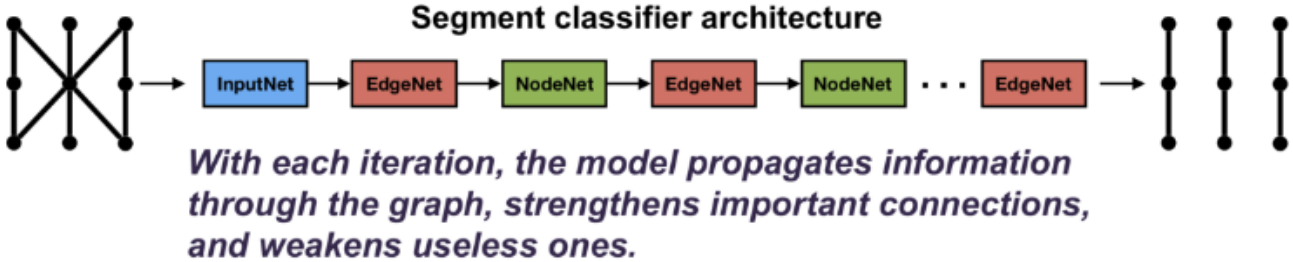
Time dimension (state memory)



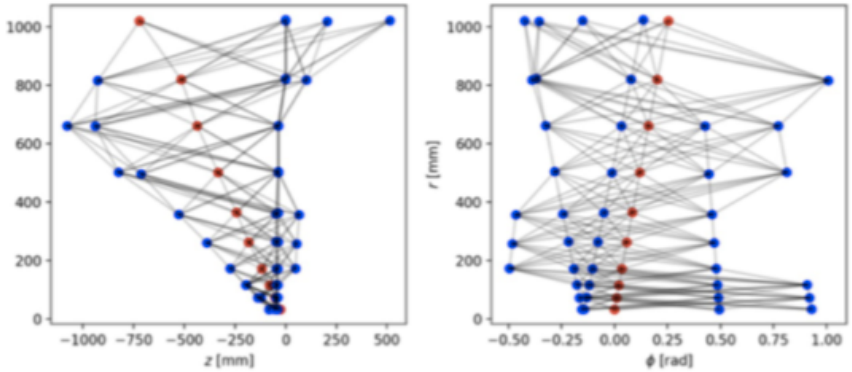
HEP.TrkX, CHEP 2018

Examples

Tracking with Graph Neural Network



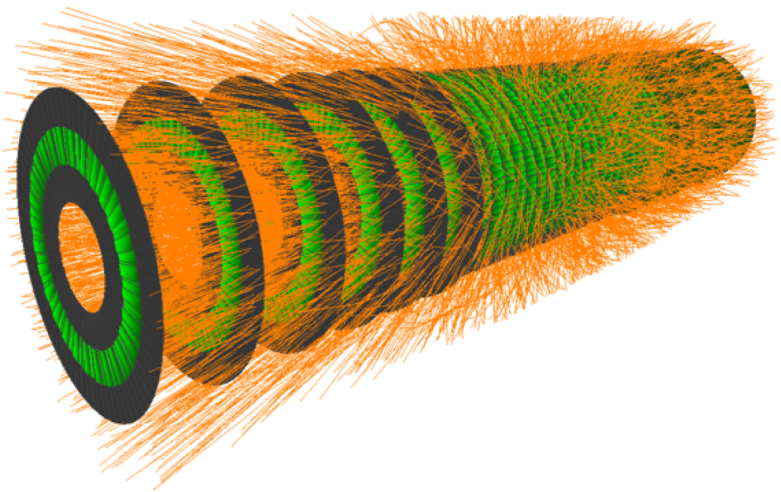
- Unseeded hit-pair classification
- Model predicts the probability that a hit-pair is valid

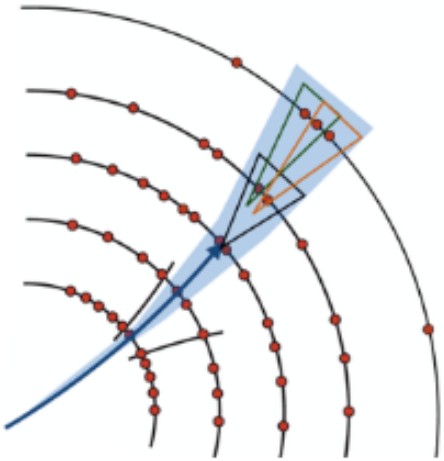


HEP.TrkX, CHEP 2018

TrackML Challenge

https://www.kaggle.com/c/trackml-particle-identification

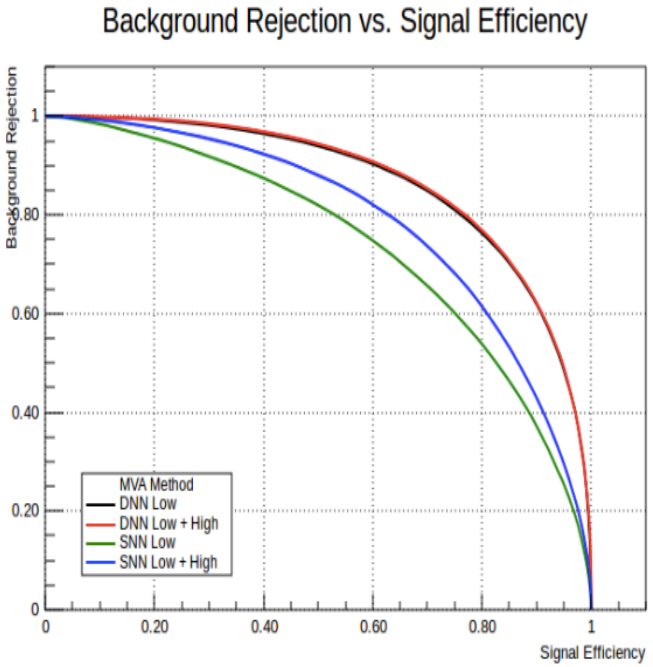




Meaningful Physics

Can we extract features with meaningful physics?

- from low-level variables

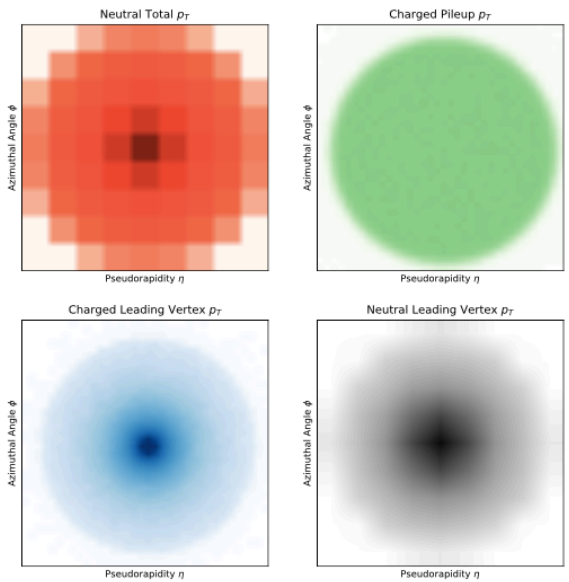


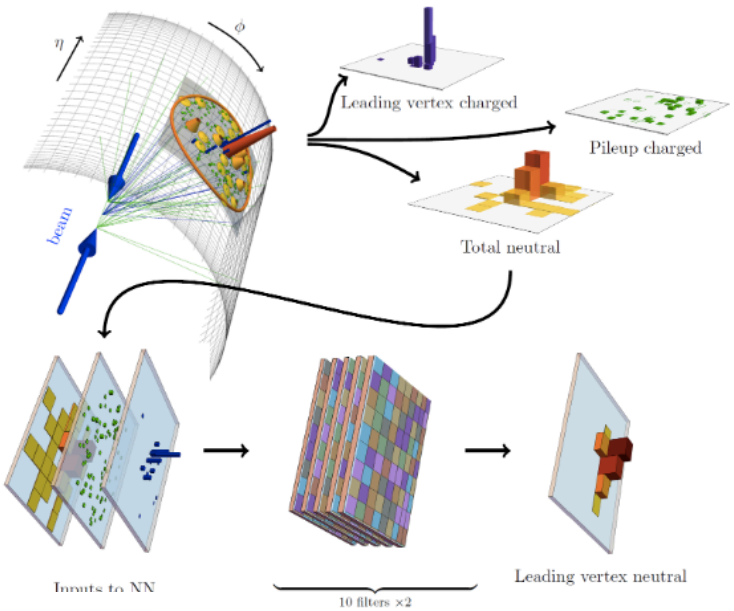
Are we able to understand ML models

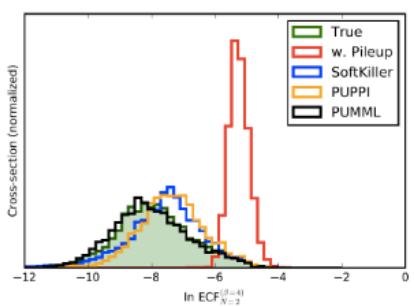
- physics interpretations

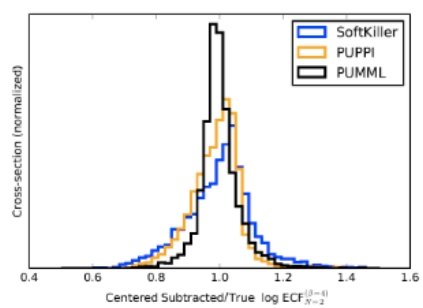
Meaningful Physics

Pile-up removal with CNN





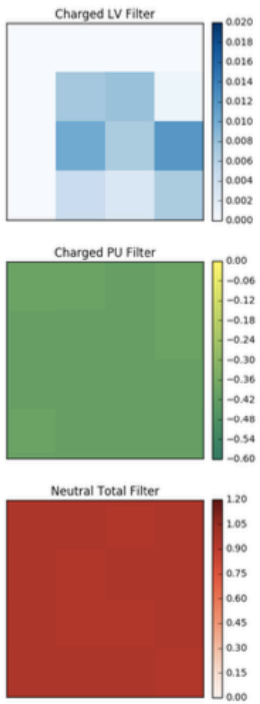




Komiske et al., 2017

Meaningful Physics

Pile-up removal with CNN



What is learned?

- Train a single 4×4 filter and inspect it.
- Pixel-wise: $p_T^{N,LV} \approx p_T^{N,tot} - \frac{1}{2} p_T^{C,PU}$
- This is linear cleansing with $\bar{\gamma}_0 = 2/3$!

$$p_T^{N,LV} = p_T^{N,tot} + (1 - \frac{1}{\bar{\gamma}_0}) p_T^{C,PU}$$

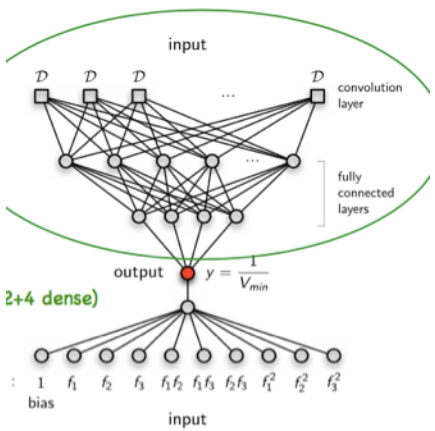
Komiske et al., 2017

Defining the Problem

How to best use domain knowledge we have accumulated?

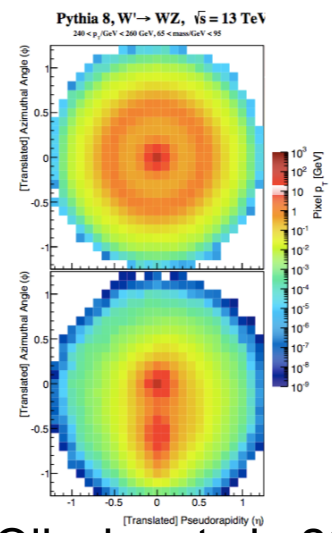
- in designing the algorithms

Strings



Krefl, 2017

Jet Images

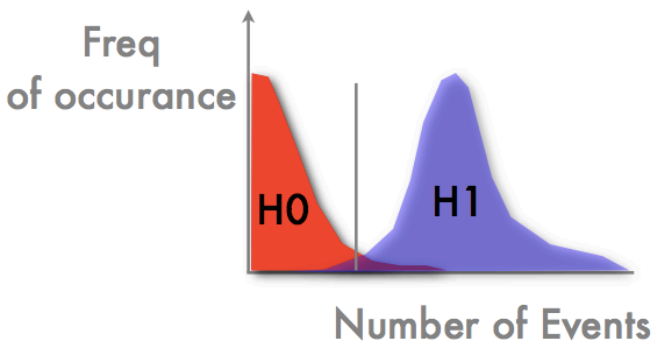


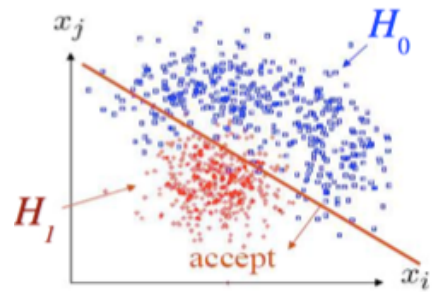
de Oliveira et al., 2015

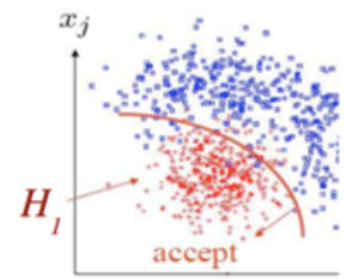
Jet Clustering

Louppe et al., 2017

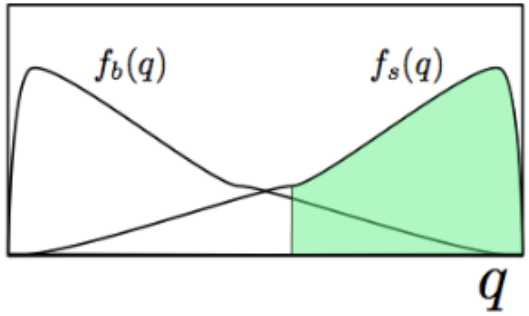
Uncertainties







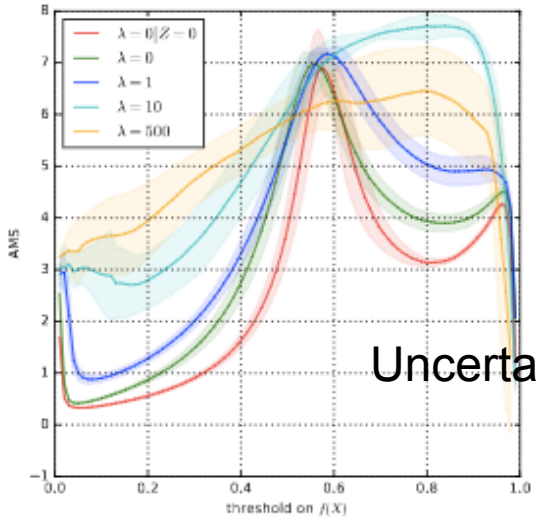
A threshold makes sense.



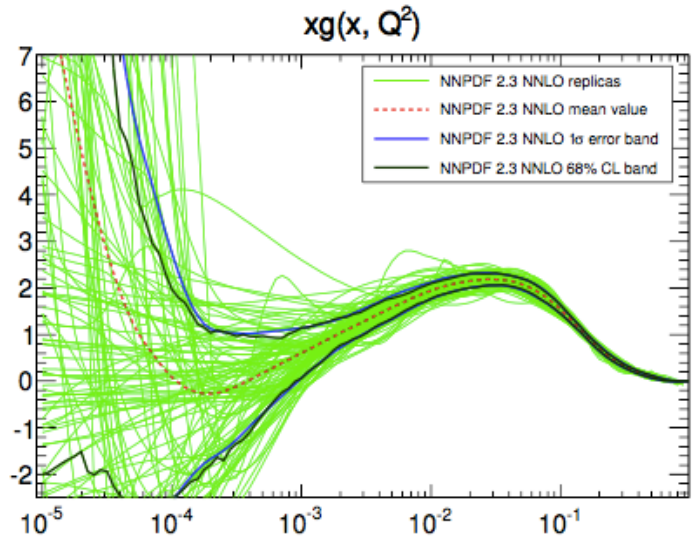
- **Decision making**

Uncertainties

G. Louppe et al., 2016



NNPDF Collaboration

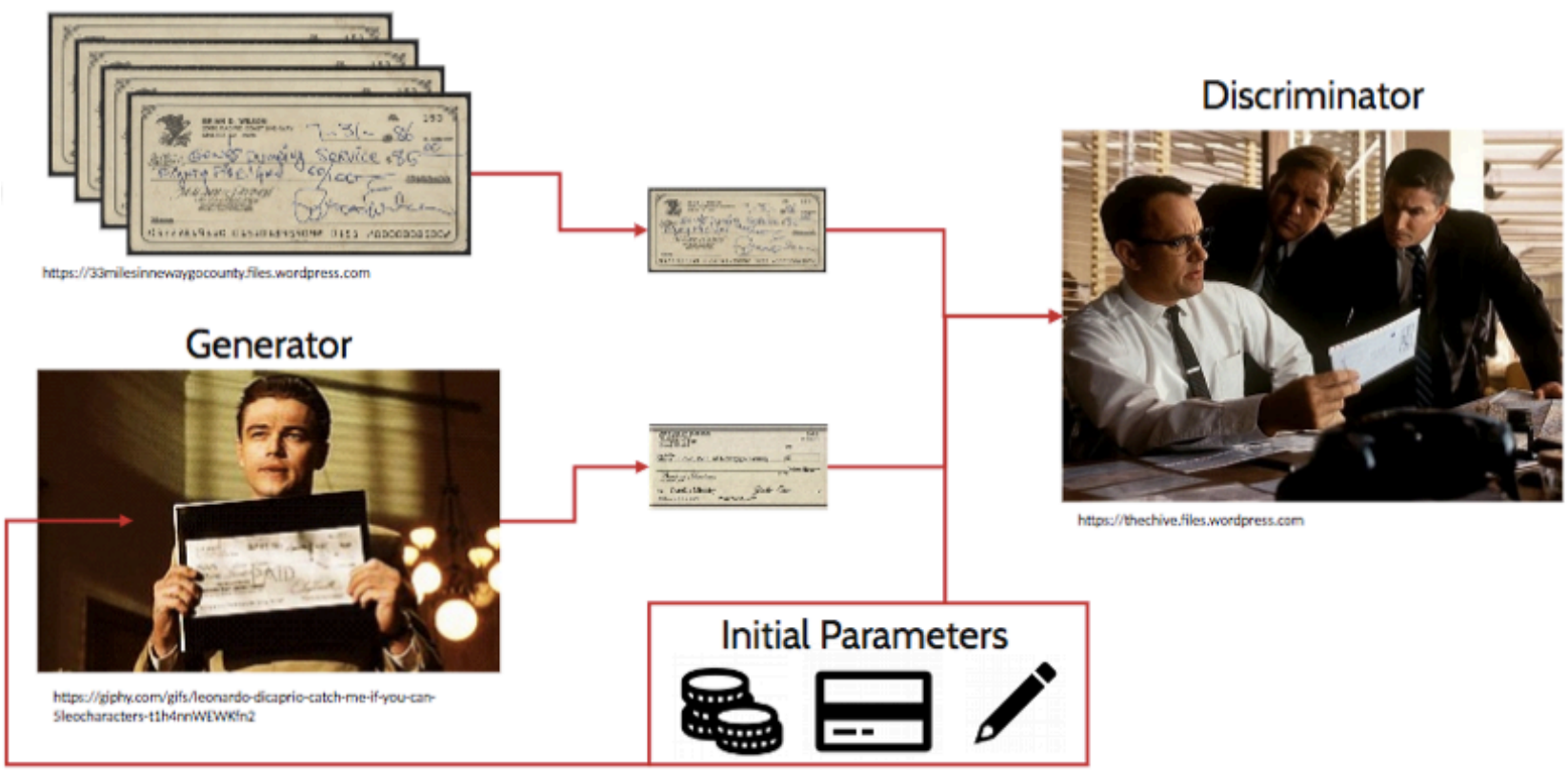


Uncertainties matter

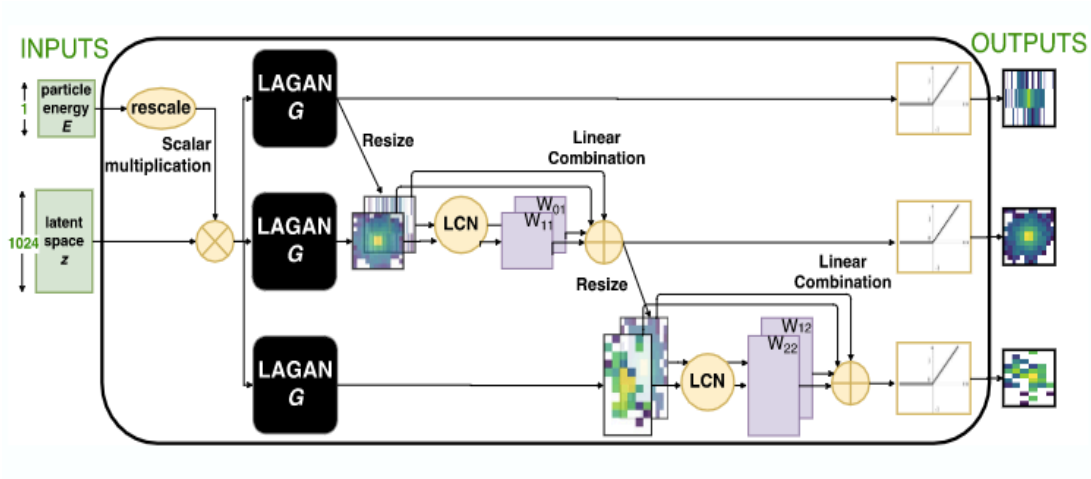
Bayesian connection: Deep neural networks with drop-out approximate variational inference of Bayesian NNs: *Gal and Ghahramani, 2016*

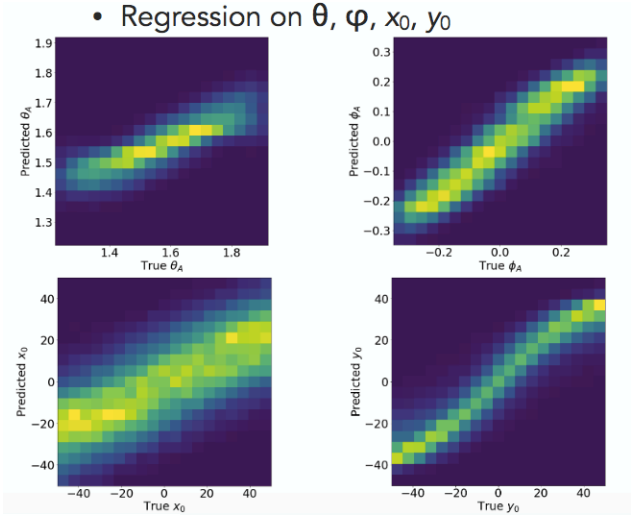
GANs

Generative Adversarial Networks



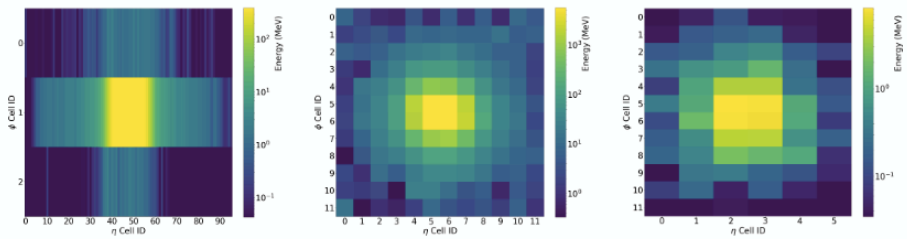
Simulation GANs



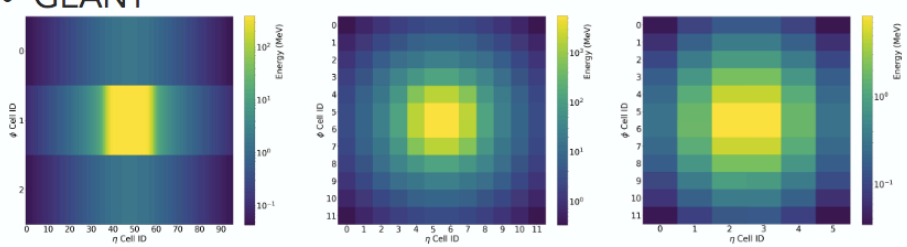


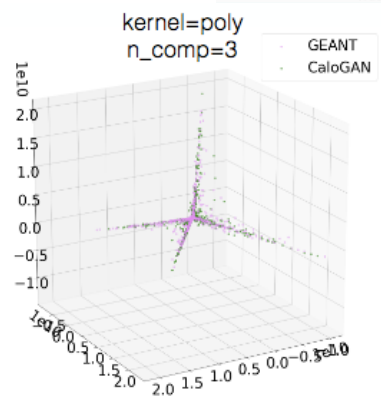
Dataset: 5°; Net: soft sparsity, multiplied E, Conv. attn. and layers

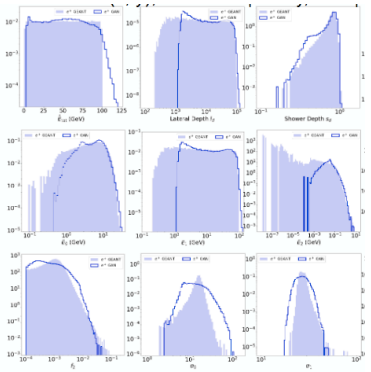
- CaloGAN



- GEANT







L. de Oliveira et al., 2017

Energy Regression

Inputs:

photon coordinates

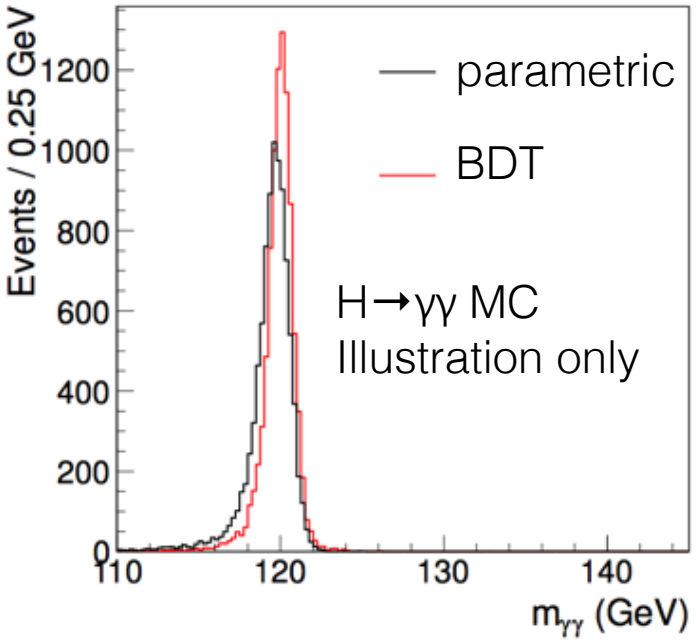
photon shower information

median event energy

Target Output:

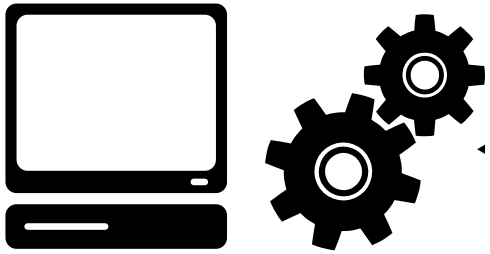
$E_{MEASURED}/E_{TRUE}$

10-30% improvement with shallow ML



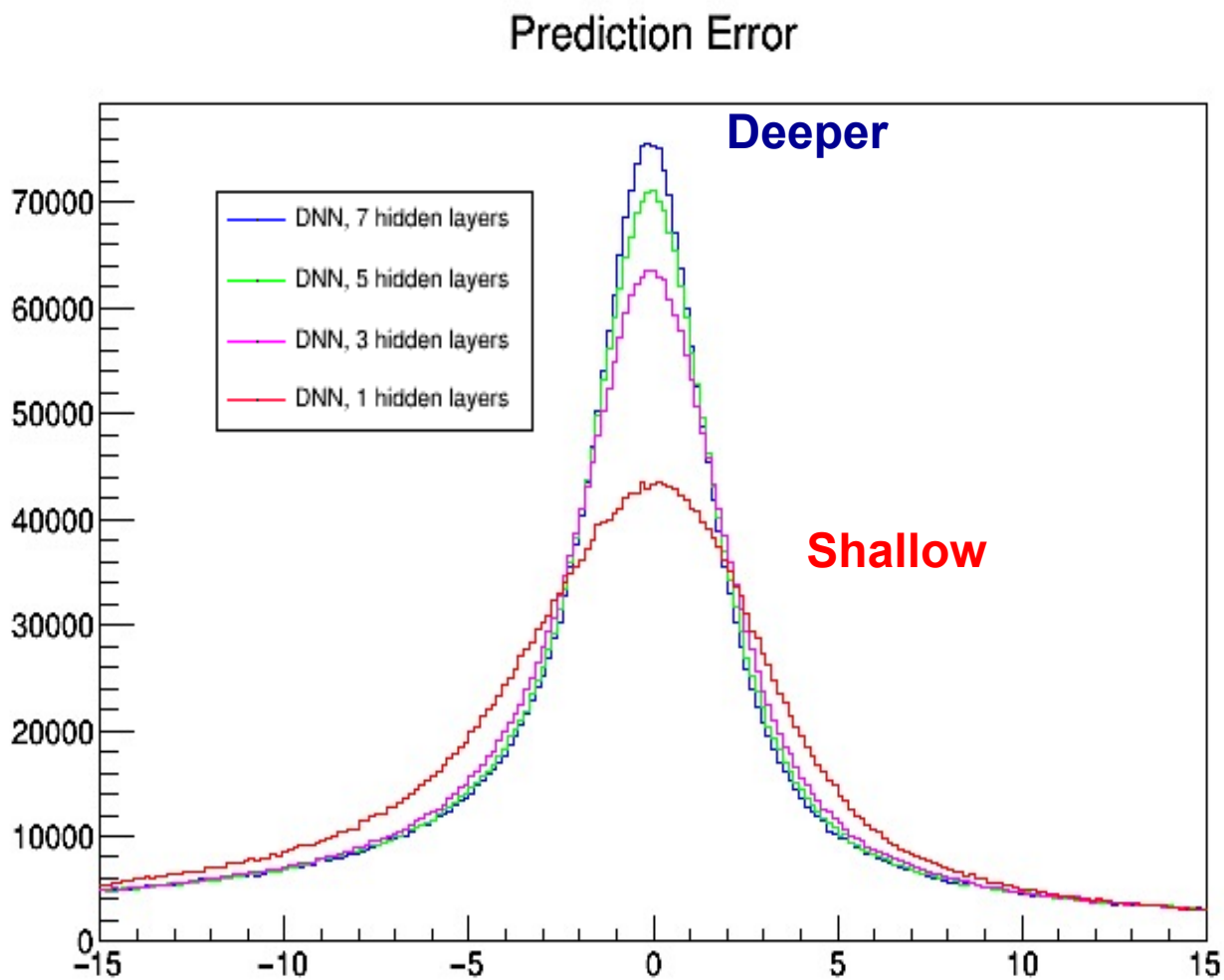
Regression

Modify evaluation in induction algorithm

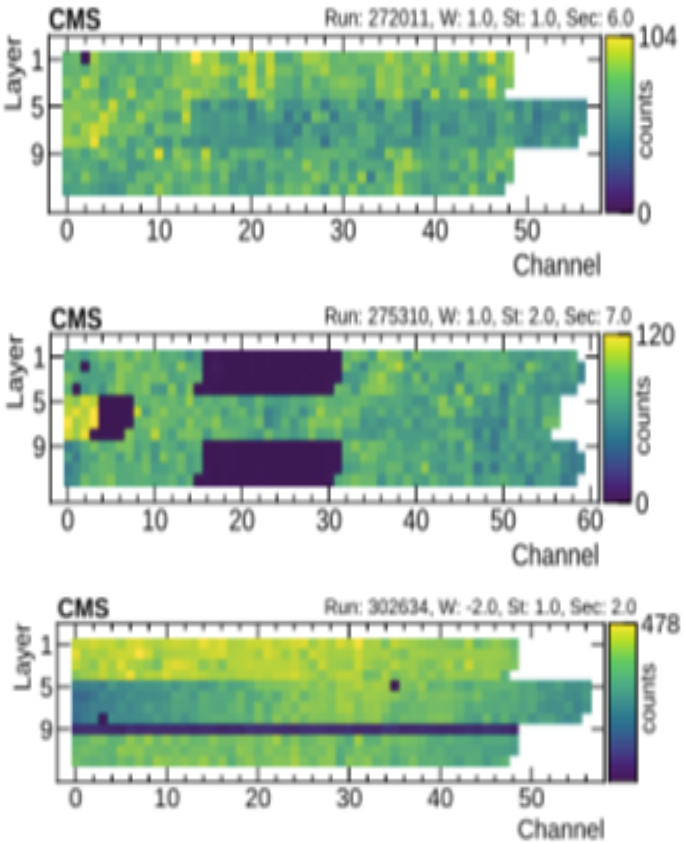


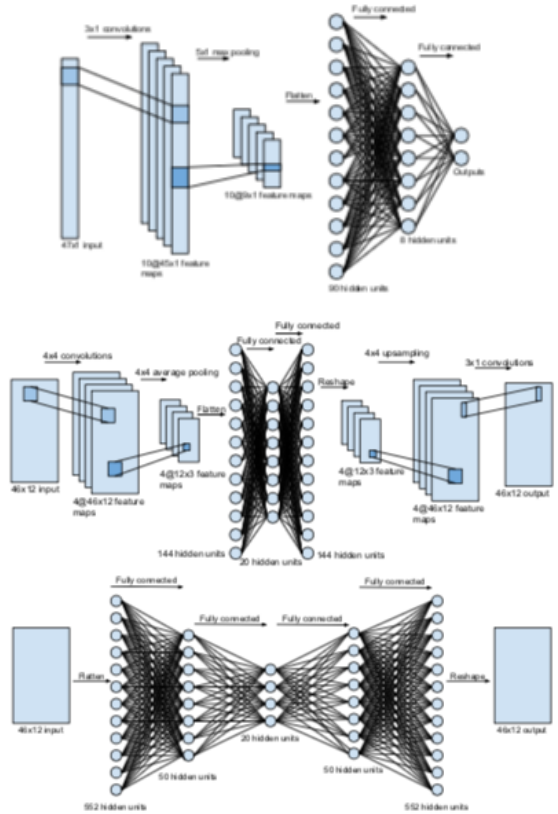
Maximum separation → **Minimal variance**

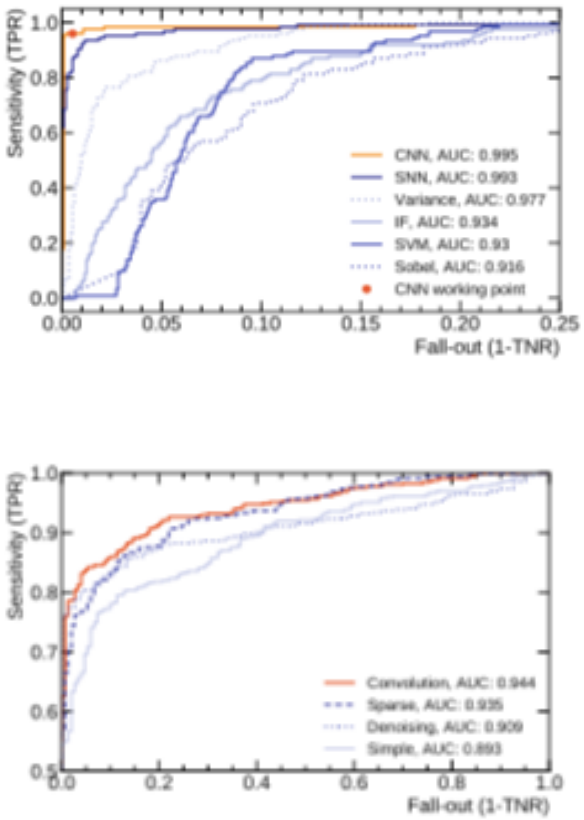
Deep Learning Regression



Online DQM


CMS
Run: 272011, W: 1.0, St: 1.0, Sec: 6.0
Layer
counts
Channel
CMS
Run: 275310, W: 1.0, St: 2.0, Sec: 7.0
Layer
counts
Channel
CMS
Run: 302634, W: -2.0, St: 1.0, Sec: 2.0
Layer
counts
Channel

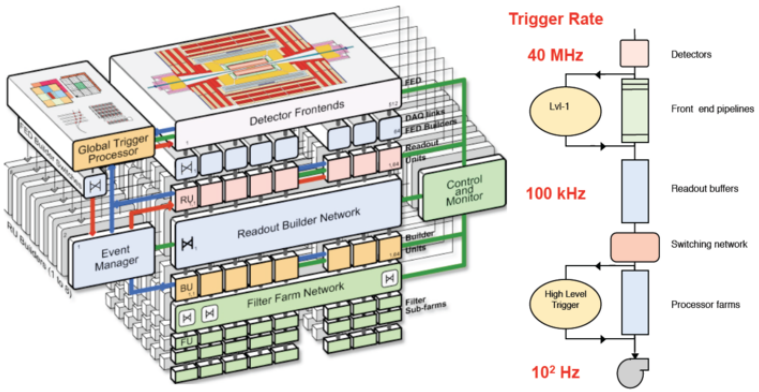



Sensitivity (TPR)
Fall-out (1-TNR)
CNN, AUC: 0.995
SNN, AUC: 0.993
Variance, AUC: 0.977
IF, AUC: 0.934
SVM, AUC: 0.93
Sobel, AUC: 0.916
CNN working point
Sensitivity (TPR)
Fall-out (1-TNR)
Convolution, AUC: 0.944
Sparse, AUC: 0.935
Denoising, AUC: 0.909
Simple, AUC: 0.893

Real Time Application

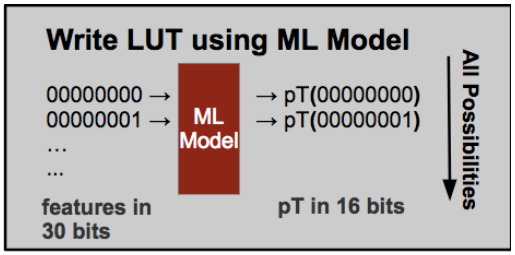
Can we do ML in real-time?

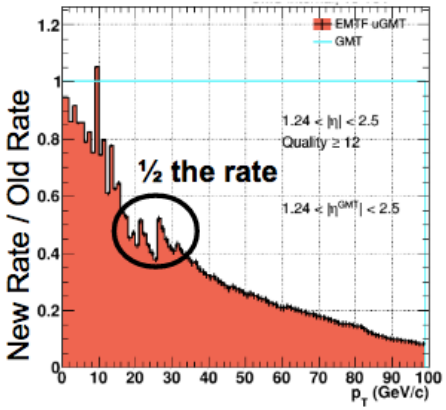
- ML: live video analysis, medical, self-driving cars
- HEP **Trigger Systems** (software and hardware)

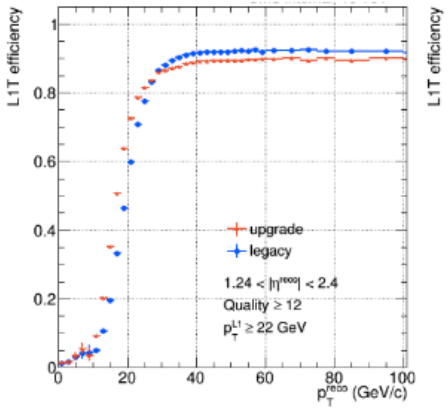


Trigger Applications

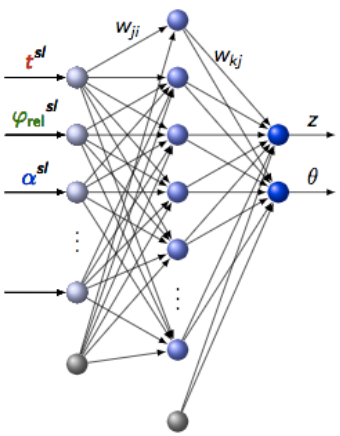
CMS L1

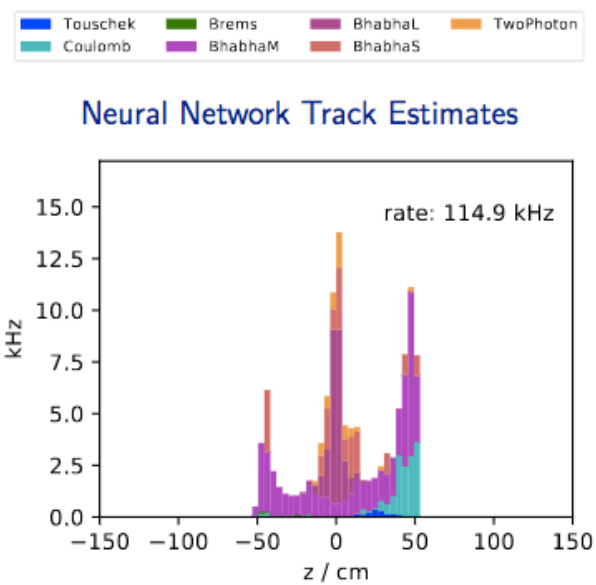


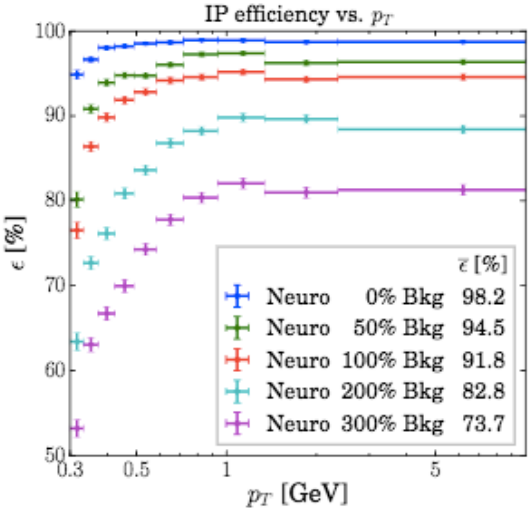




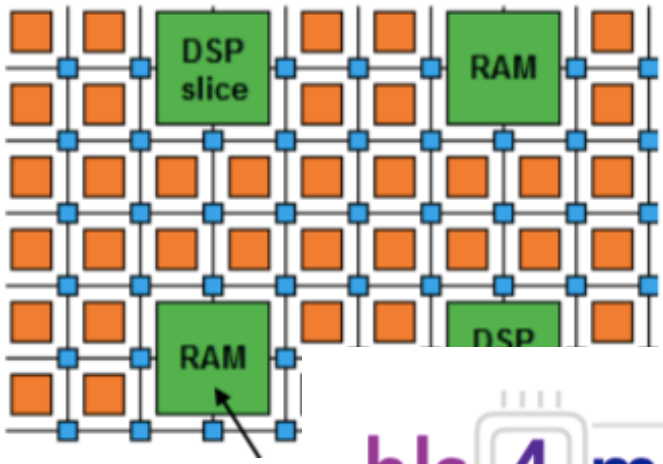
Belle II

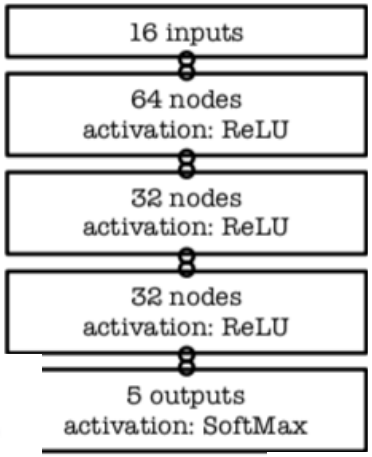


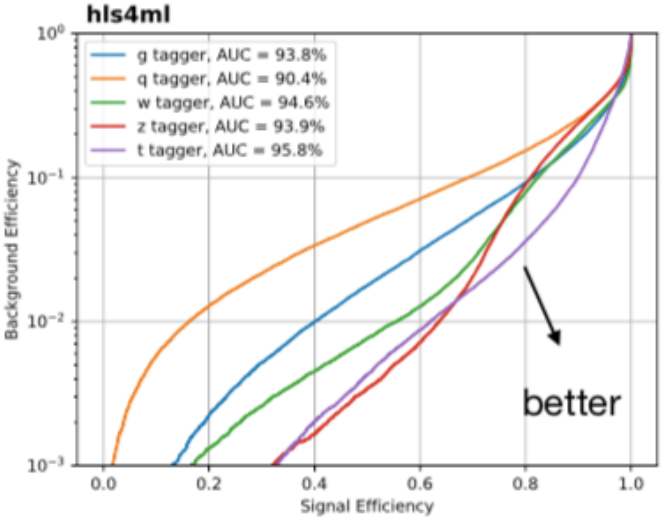


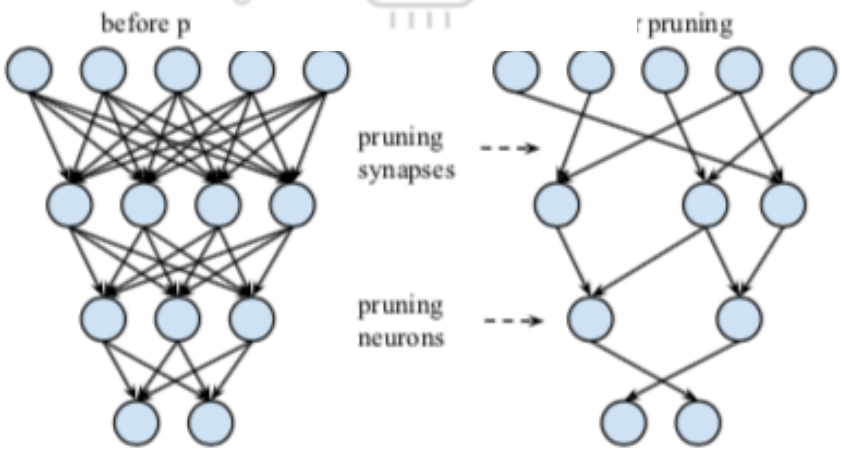


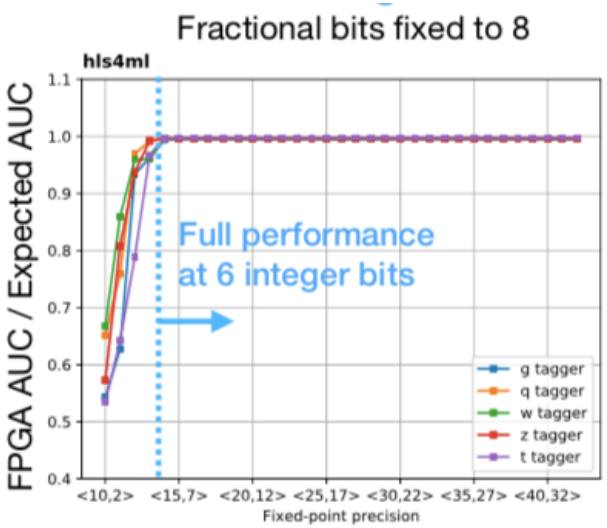
ML on FPGA

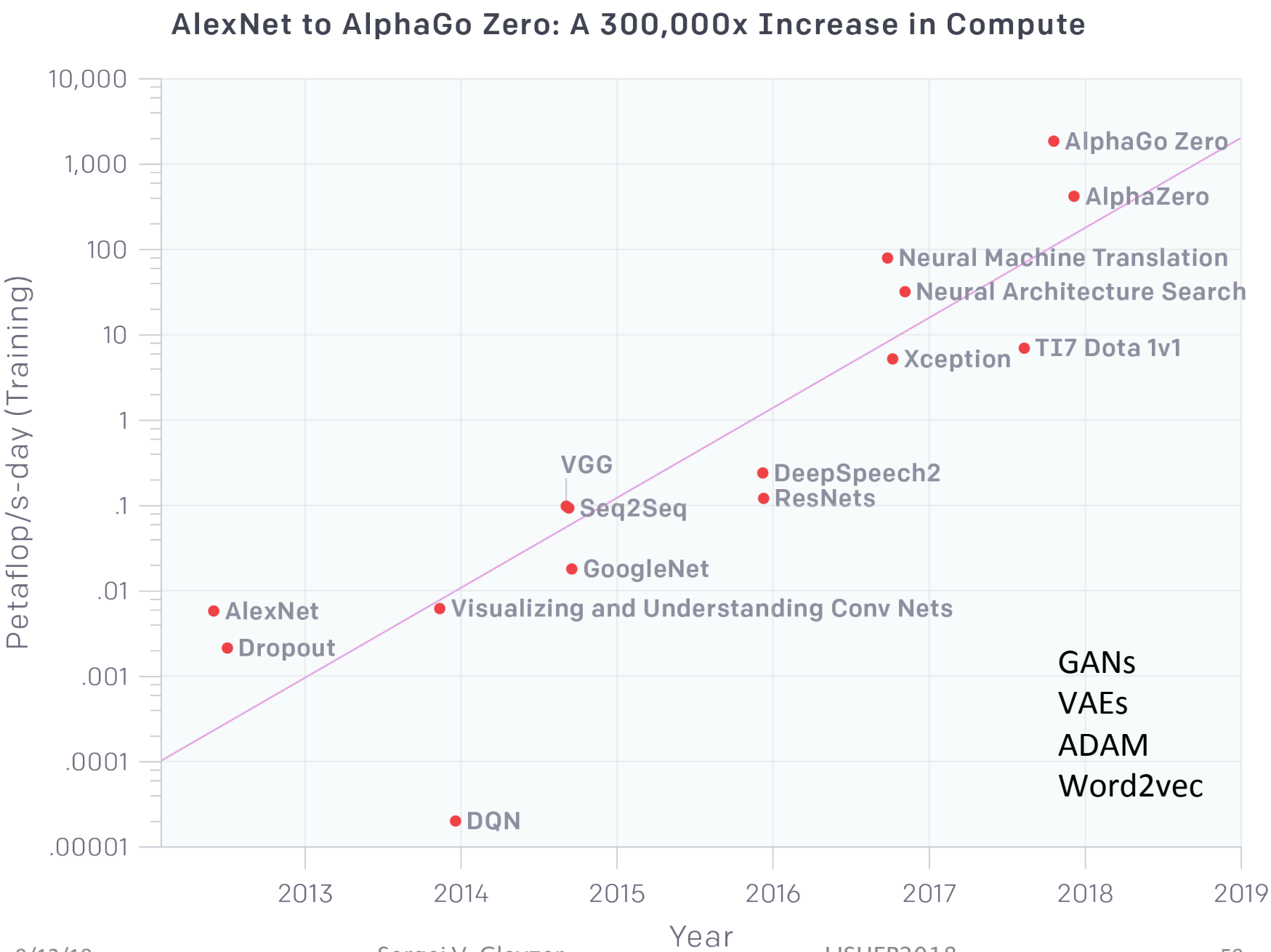

DSP slice
RAM
RAM
DSP


16 inputs
64 nodes
activation: ReLU
32 nodes
activation: ReLU
32 nodes
activation: ReLU
5 outputs
activation: SoftMax


hls4ml
g tagger, AUC = 93.8%
q tagger, AUC = 90.4%
w tagger, AUC = 94.6%
z tagger, AUC = 93.9%
t tagger, AUC = 95.8%
Background Efficiency
Signal Efficiency
better


before p
pruning
pruning synapses
pruning neurons


Fractional bits fixed to 8
hls4ml
FPGA AUC / Expected AUC
Full performance
at 6 integer bits
g tagger
q tagger
w tagger
z tagger
t tagger
<10,2> <15,7> <20,12> <25,17> <30,22> <35,27> <40,32>
Fixed-point precision


AlexNet to AlphaGo Zero: A 300,000x Increase in Compute
Petaflop/s-day (Training)
10,000
1,000
100
10
1
.1
.01
.001
.0001
.00001
2013
2014
2015
2016
2017
2018
2019
Year
AlphaGo Zero
AlphaZero
Neural Machine Translation
Neural Architecture Search
TI7 Dota 1v1
Xception
VGG
Seq2Seq
DeepSpeech2
ResNets
GoogleNet
AlexNet
Dropout
Visualizing and Understanding Conv Nets
DQN
GANs
VAEs
ADAM
Word2vec

HEPML-CWP

Machine Learning in High Energy Physics Community White Paper

July 10, 2018

Abstract: Machine learning is an important applied research area in particle physics, beginning with applications to high-level physics analysis in the 1990s and 2000s, followed by an explosion of applications in particle and event identification and reconstruction in the 2010s. In this document we discuss promising future research and development areas in machine learning in particle physics with a roadmap for their implementation, software and hardware resource requirements, collaborative initiatives with the data science community, academia and industry, and training the particle physics community in data science. The main objective of the document is to connect and motivate these areas of research and development with the physics drivers of the High-Luminosity Large Hadron Collider and future neutrino experiments and identify the resource needs for their implementation. Additionally we identify areas where collaboration with external communities will be of great benefit.

Editors: Sergei Gleyzer[26], Paul Seyfert[11], Steven Schramm[28]

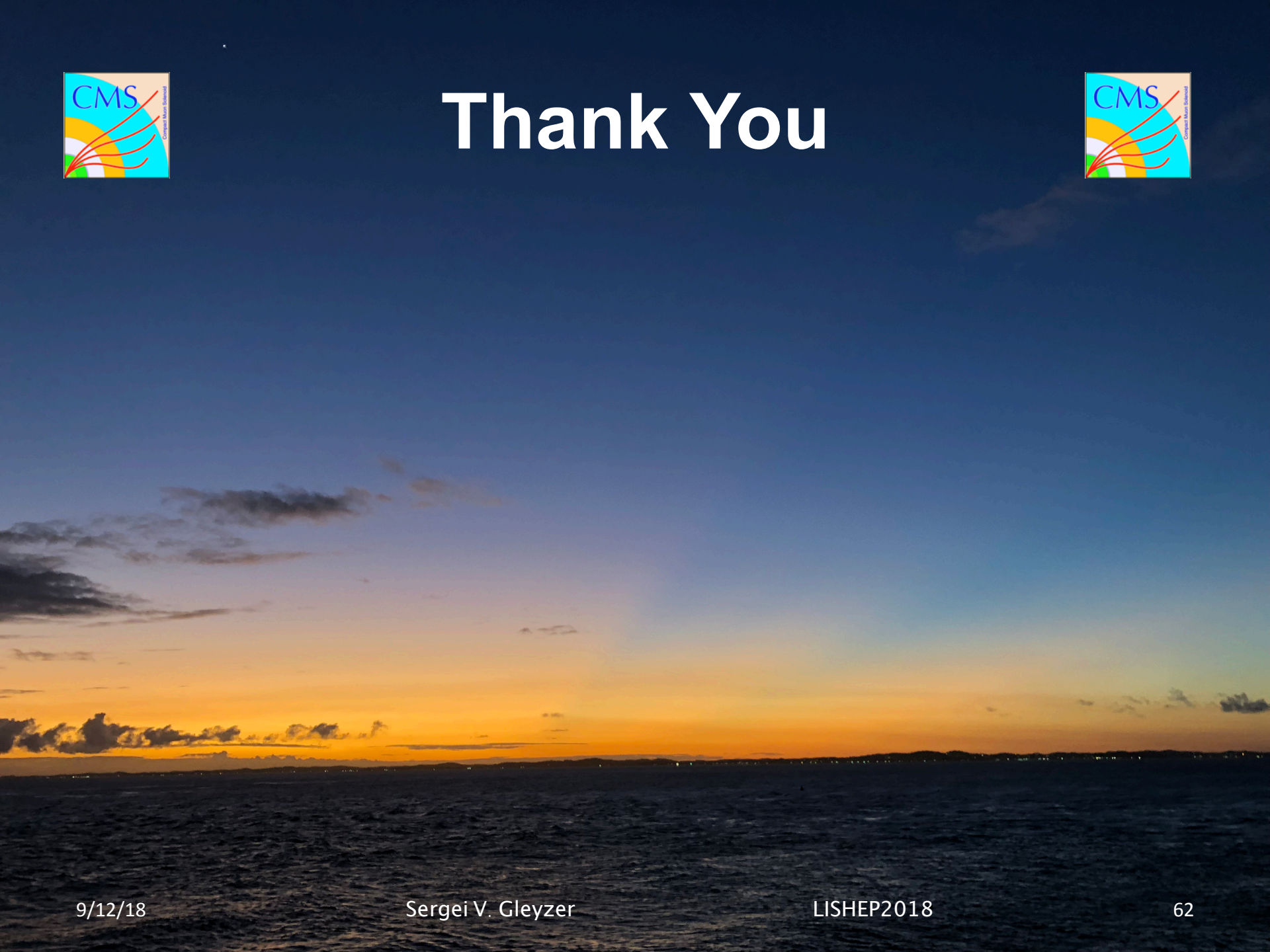
arXiv:1807.02876

Contributors: Kim Albertsson[1], Piero Altoe[2], Dustin Anderson[3], Michael Andrews[4], Juan Pedro Araque Espinosa[5], Adam Aurisano[6], Laurent Basara[7], Adrian Bevan[8], Wahid Bhimji[9], Daniele Bonacorsi[10], Paolo Calafiura[9], Mario Campanelli[8], Louis Capps[2], Federico Carminati[11], Stefano Carrazza[11], Taylor Childers[12], Elias Coniavitis[13], Kyle Cranmer[14], Claire David[15], Douglas Davis[16], Javier Duarte[17], Martin Erdmann[18], Jonas Eschle[19], Amir Farbin[20], Matthew Feickert[21], Nuno Filipe Castro[5], Conor Fitzpatrick[22], Michele Floris[11], Alessandra Forti[23], Jordi Garra-Tico[24], Jochen Gemmler[25], Maria Girone[11], Paul Glaysher[15], Sergei Gleyzer[26], Vladimir Gligorov[27], Tobias Golling[28], Jonas Graw[2], Lindsey Gray[17], Dick Greenwood[29], Thomas Hacker[30], John Harvey[11], Benedikt Hegner[11], Lukas Heinrich[14], Ben Hooberman[31], Johannes Junggeburth[32], Michael Kagan[33], Meghan Kane[34], Konstantin Kanishchev[7], Przemysław Karpiński[11], Zahari Kassabov[35], Gaurav Kaul[36], Dorian Kcira[3], Thomas Keck[25], Alexei Klimentov[37], Jim Kowalkowski[17], Luke Kreczko[38], Alexander Kurepin[39], Rob Kutschke[17], Valentin Kuznetsov[40], Nicolas Köhler[32], Igor Lakomov[11], Kevin Lannon[41], Mario Lassnig[11], Antonio Limosani[42], Gilles Louppe[14], Aashrita Mangu[43], Pere Mato[11], Helge Meinhard[11], Dario Menasce[44], Lorenzo Moneta[11], Seth Moortgat[45], Meenakshi Narain[46], Mark Neubauer[31], Harvey Newman[3], Hans Pabst[36], Michela Paganini[47], Manfred Paulini[4], Gabriel Perdue[17], Uzziel Perez[48], Attilio Picazio[49], Jim Pivarski[50], Harrison Prosper[51], Fernanda Psihas[52], Alexander Radovic[53], Ryan Reece[54], Aurelius Rinkevicius[40], Eduardo Rodrigues[6], Jamal Rorie[55], David Rousseau[56], Aaron Sauers[17], Steven Schramm[28], Ariel Schwartzman[33], Horst Severini[57], Paul Seyfert[11], Filip Siroky[58], Konstantin Skazytkin[39], Mike Sokoloff[6], Graeme Stewart[59], Bob Stienen[60], Ian Stockdale[61], Giles Strong[5], Savannah Thais[47], Karen Tomko[62], Eli Upfal[46], Emanuele Usai[46], Andrey Ustyuzhanin[63], Martin Vala[64], Sofia Vallecorsa[65], Mauro Verzetti[66], Xavier Vilasís-Cardona[67], Jean-Roch Vlimant[3], Ilija Vukotic[68], Sean-Jiun Wang[26], Gordon Watts[69], Michael Williams[70], Wenjing Wu[71], Stefan

Summary

- **Machine learning already at the forefront of high-energy physics applications**
  - Improvements in **particle and event classification**, and **property measurements**
  - **Feature extraction**
  - Estimating and incorporating **uncertainty**
  - Emerging machine learning applications for **fast inference** in hardware

Thank You



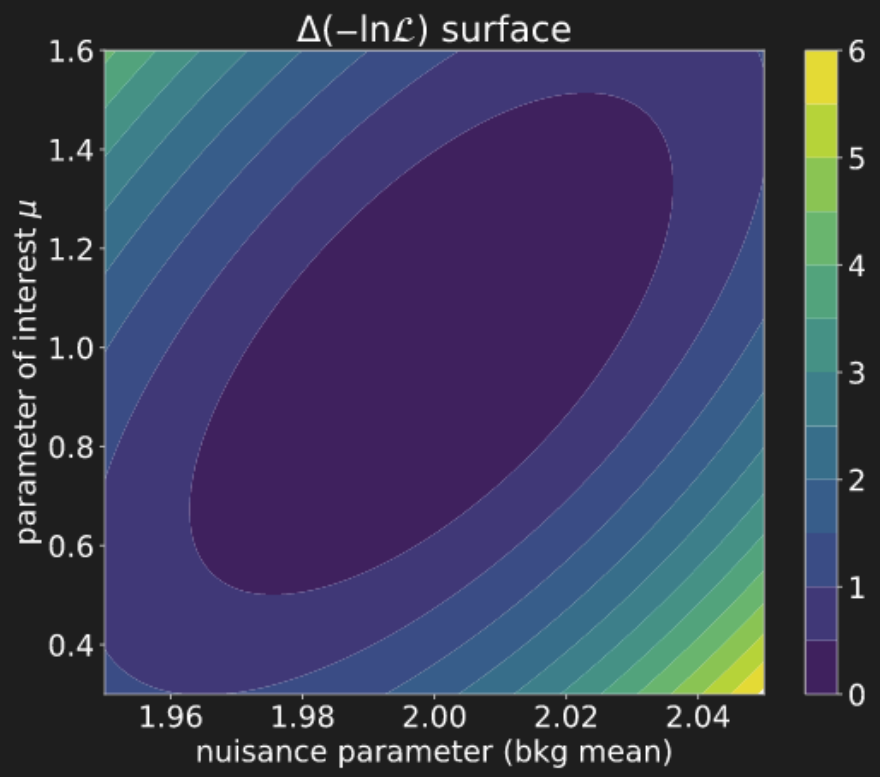
INFERENCE-MOTIVATED LOSS FUNCTION

If we expand the negative log-likelihood around minimum (e.g. Asimov $n_i = \alpha_s \cdot s_i + \alpha_b \cdot b_i$), due to Cramér-Rao bound:

$$\text{covariance} \geq \mathbf{H}^{-1}(-\ln \mathcal{L})$$

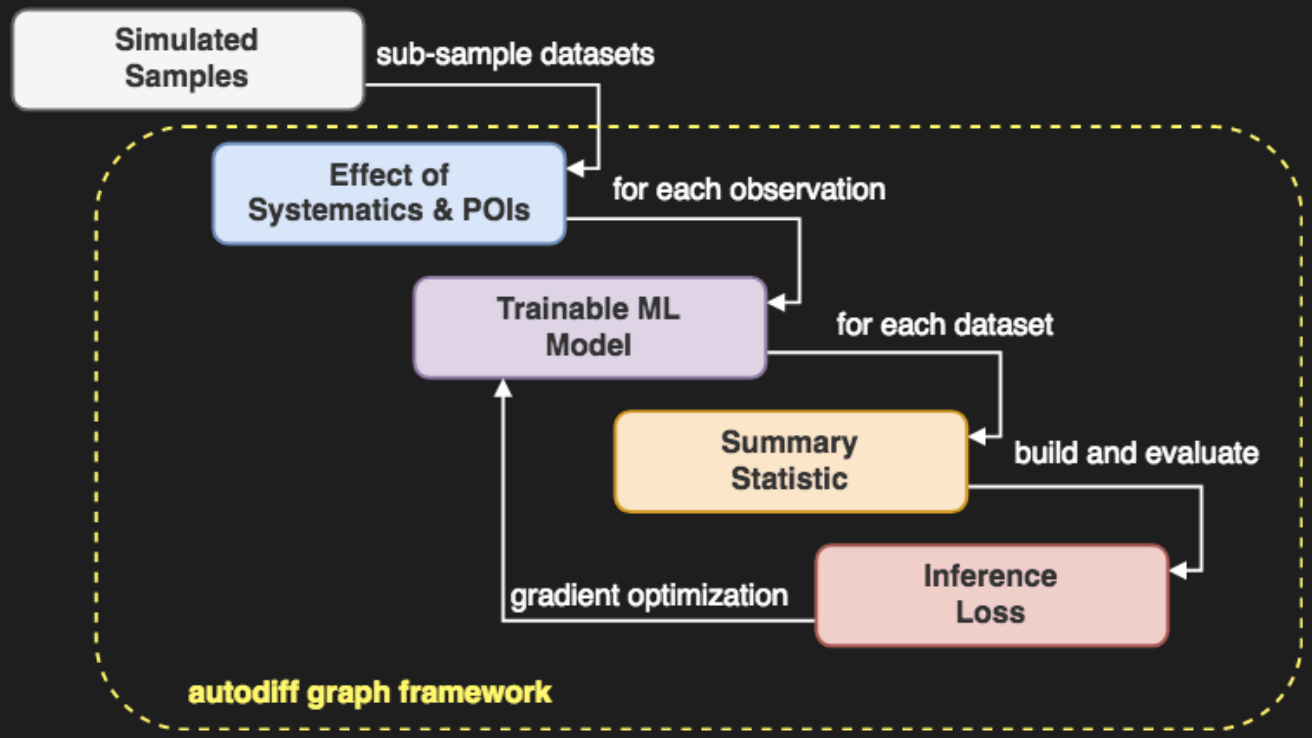
which can be computed via autodiff. Can use as loss function directly the variance bound on the parameters of interest

$$\text{loss} \approx \text{Var}(\mu) \quad \text{(expected)}$$



P. de Castro, 2018

END-TO-END DIFFERENTIABILITY FOR LHC ANALYSES

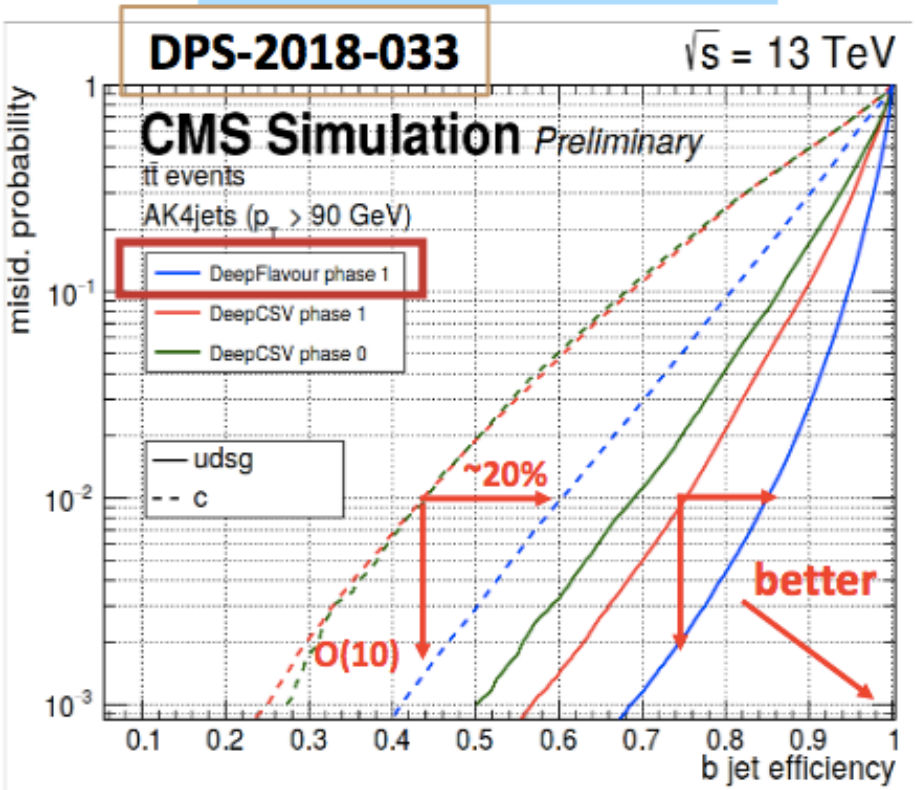


Within this general framework, several approaches are possible, focus here is

DIRECT LEARNING OF SYSTEMATICS-AWARE SUMMARY STATISTICS

Flavor Tagging

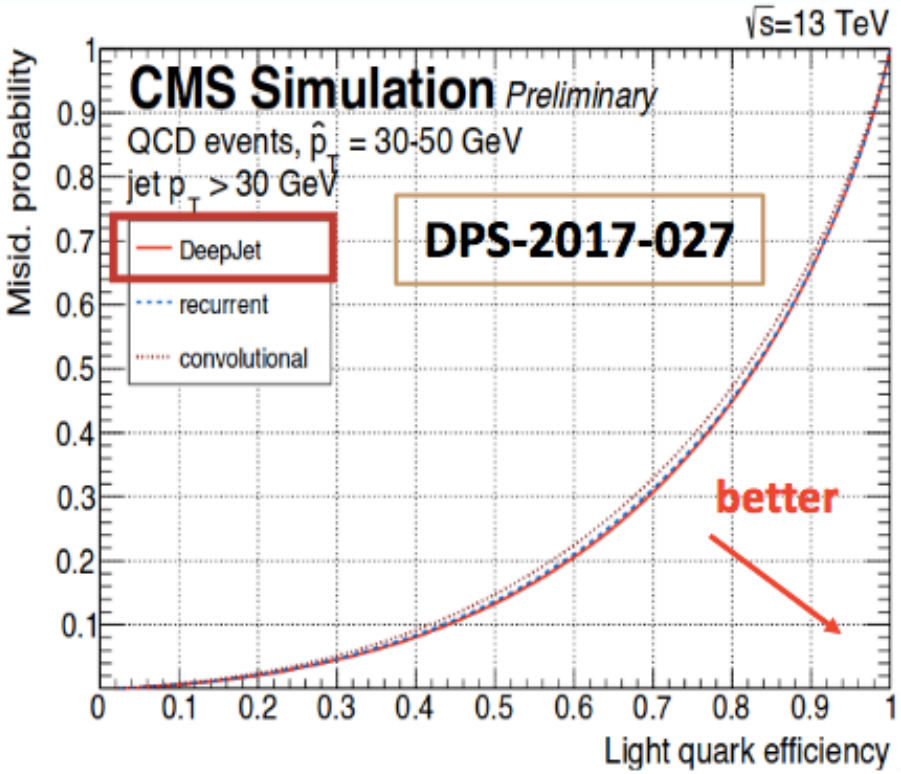
b vs. udsg / b vs. c



→ Significant gain in performance even more significant at higher pT

→ Large part of the performance loss of previous [non particle-based] taggers was due to track preselection

Quark – gluon separation

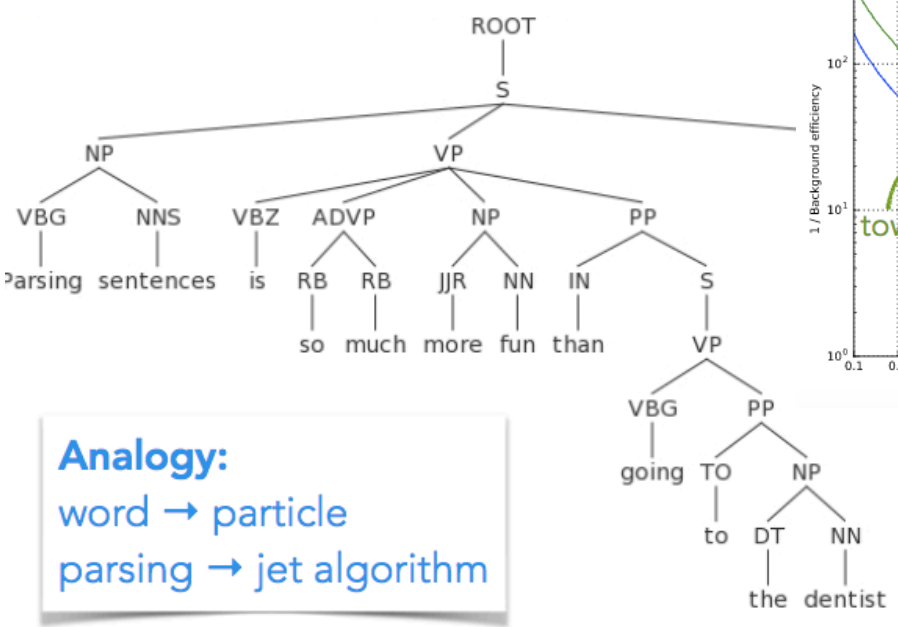


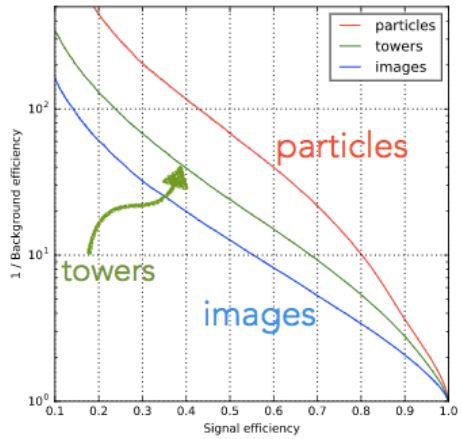
→ Generator level light quarks/gluons that did not split to heavy flavour

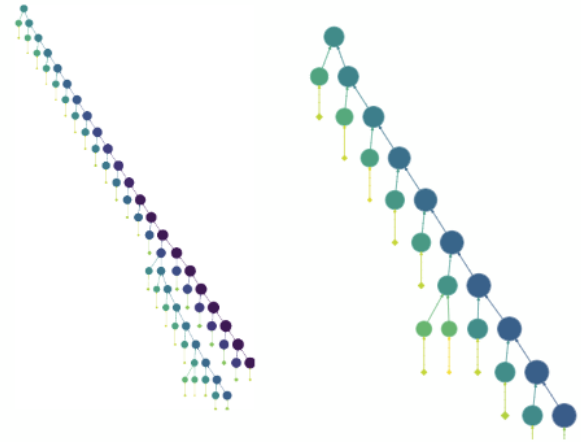
→ Similar performance to simpler & dedicated architectures

Jet Algorithm

Natural Language Processing, RNNs and Jets







T. Cheng, 2017

Analogy:
word → particle
parsing → jet algorithm

G. Louppe et al., 2017

LoLa

Butter, Kasieczka, Plehn, MR '17

Beyond images: LoLa

Why not use the jet constituent 4-vectors directly?
Two ingredients:

1. CoLa* - learns the jet clustering history

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \tilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$$

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ 0 & 1 & & \vdots & C_{2,N+2} & \cdots & C_{2,M} \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix}$$

- Test on-shell conditions

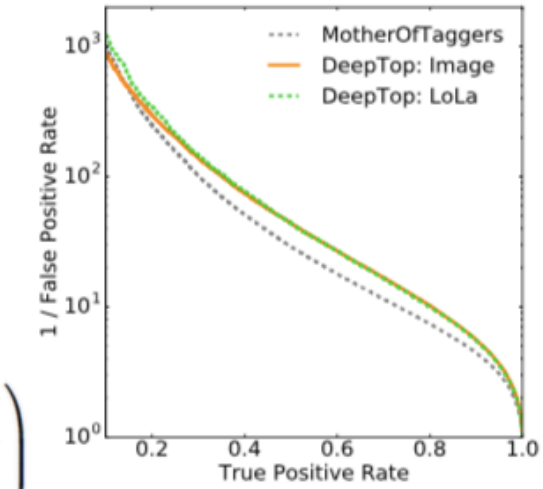
$$\tilde{k}^2_{\mu,1} = (k_{\mu,1} + k_{\mu,2} + k_{\mu,3})^2 = m_t^2$$
$$\tilde{k}^2_{\mu,2} = (k_{\mu,1} + k_{\mu,2})^2 = m_W^2 \,.$$

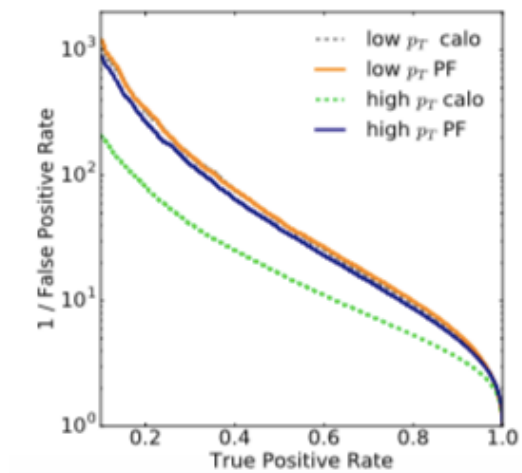
2. LoLa** - learns the kinematics

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w^{(E)}_{jm}\, E(\tilde{k}_m) \\ w^{(d)}_{jm}\, d^2_{jm} \end{pmatrix}$$

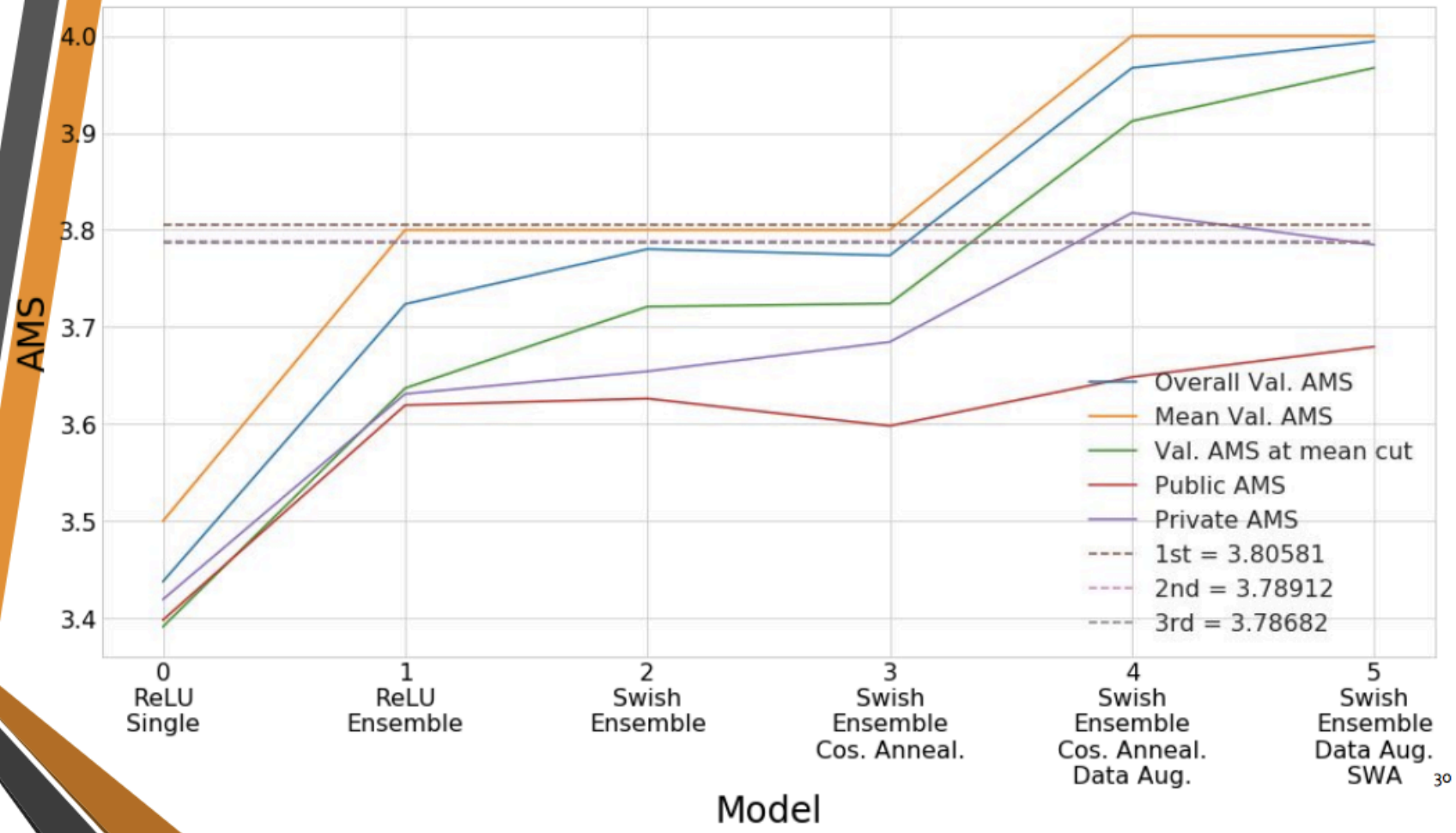
transform 4-vectors into: invariant mass, pT, energy and Minkowski distance
effectively a rotation in observable space

** CoLa = Combination Layer*
*** LoLa = Lorentz Layer*



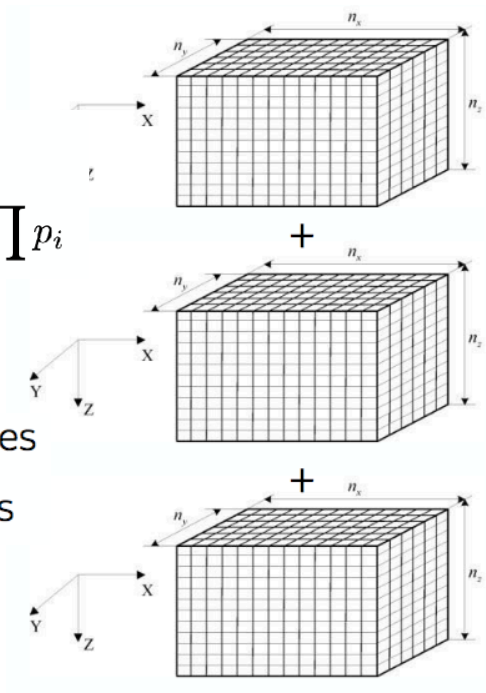


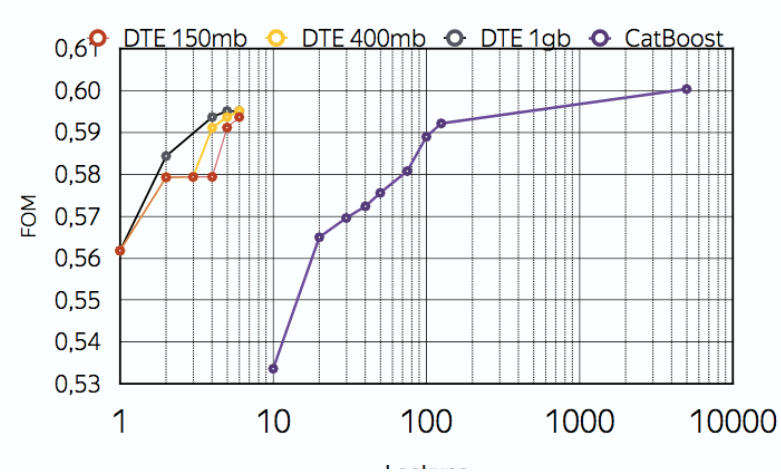
HiggsML Optimization

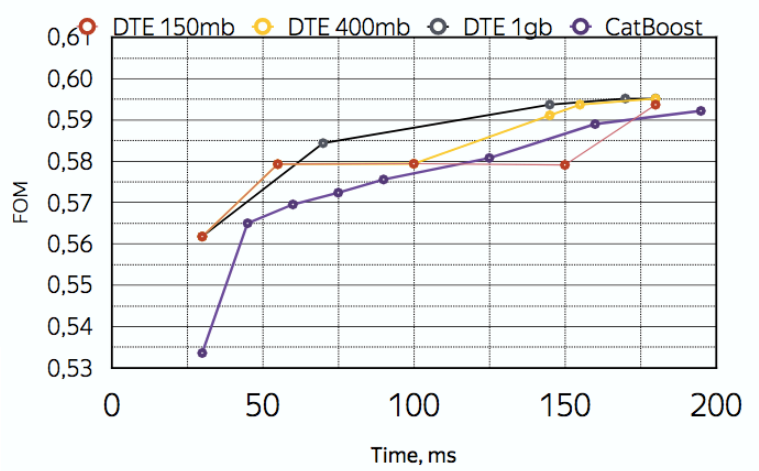


Trigger Applications

- $D = (f_1, ..., f_n)$
  $f_i = \{b_{i1}, ..., b_{ip_i}\}$
  $S(D) \propto \prod |f_i| = \prod p_i$
- CatBoost
- Uses oblivious trees
- Discretize features





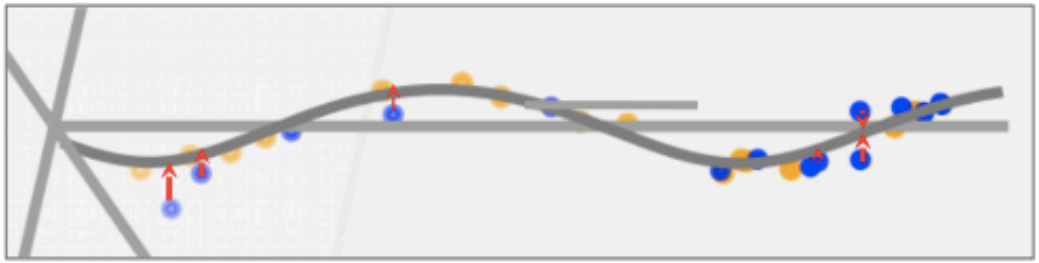


LHCb, ACAT 2017

Results

- Mean Squared Error (MSE) from the original helix as a quality measure
- Evaluation conducted on the separate test-set with ~15000 tracks

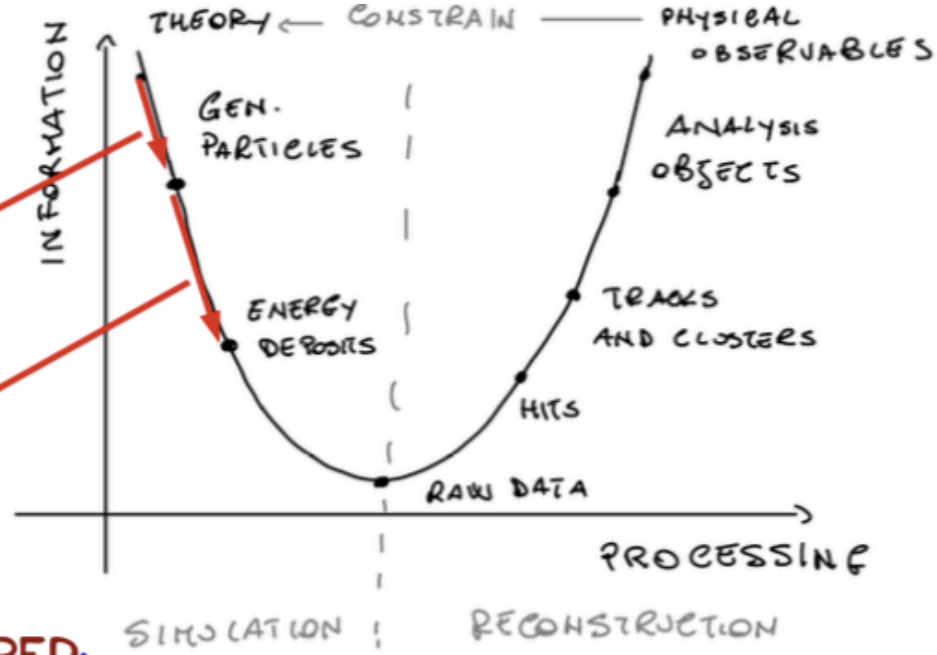
MSE visualisation:
Red - error
Grey- ideal helix
Orange – original clusters
Blue – generated clusters



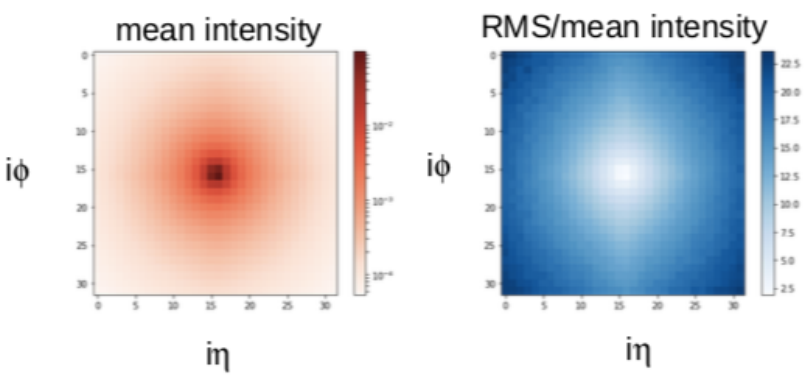
Method	Mean MSE (mm)	Median MSE (mm)	Speed-up
GEANT3	1.20	1.12	1
Random (estimated)	2500	2500	N/A
condLSTM GAN	2093.69	2070.32	100
condLSTM GAN+	221.78	190.17	
condDCGAN	795.08	738.71	25
condDCGAN+	**136.84**	**82.72**	

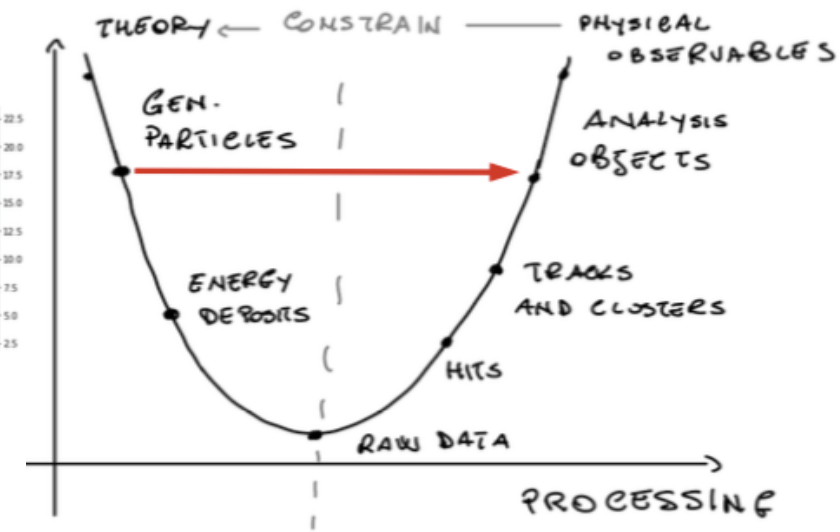
FOCUSES ON LOW LEVEL PART OF THE CHAIN:

- LA-GANS FOR EVENT SIMULATION.
- CALO-GAN AND GEANT-V FOR DETECTOR SIMULATION.



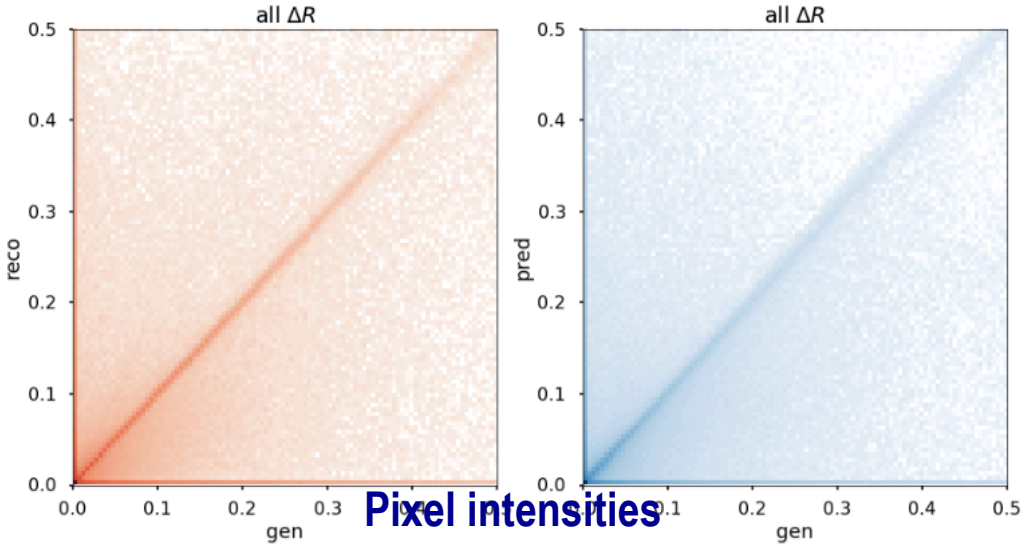
VERY HIGH ACCURACY REQUIRED:



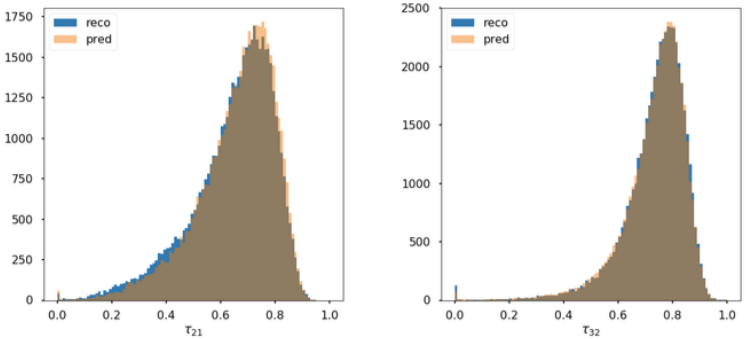


Musella and Pandolfi, 2018

FastSim with GANs



Pixel intensities



Jet substructure