Machine Learning Particle Physics

Sergei Gleyzer

Seminar at Saclay

November 24, 2017

Outline

- **What is Machine Learning**
- **in Particle Physics**
- **in Theory**
- **In Practice**

Machine Learning Basics

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

Machine Learning

General Approach:

Given **training** data $T_D = \{y, \mathbf{x}\} = (y,x)_1 \ldots (y,x)_N$,

function space {f} and a
constraint on these functions

Teach a machine to learn the **mapping** $y = f(x)$

In Computer Science

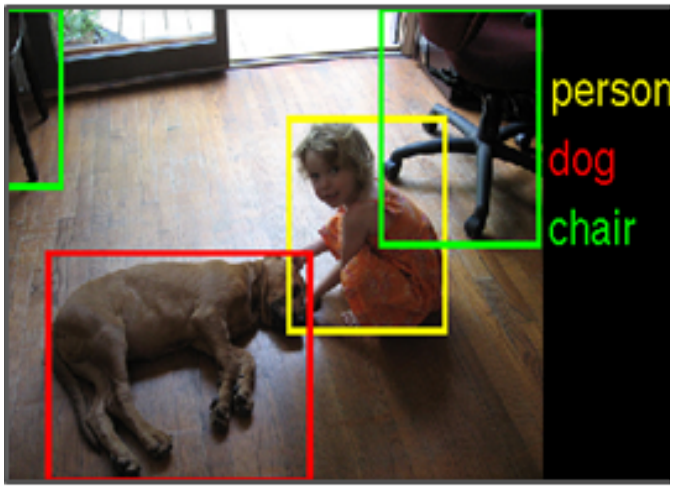
Already the preferred approach to:

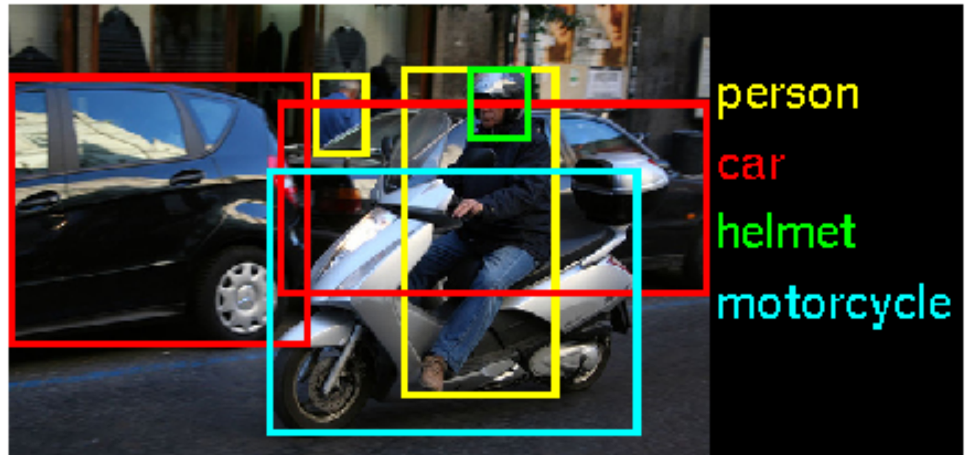
- Speech recognition, natural language processing
- Computer vision, Robot control
- Medical outcomes analysis

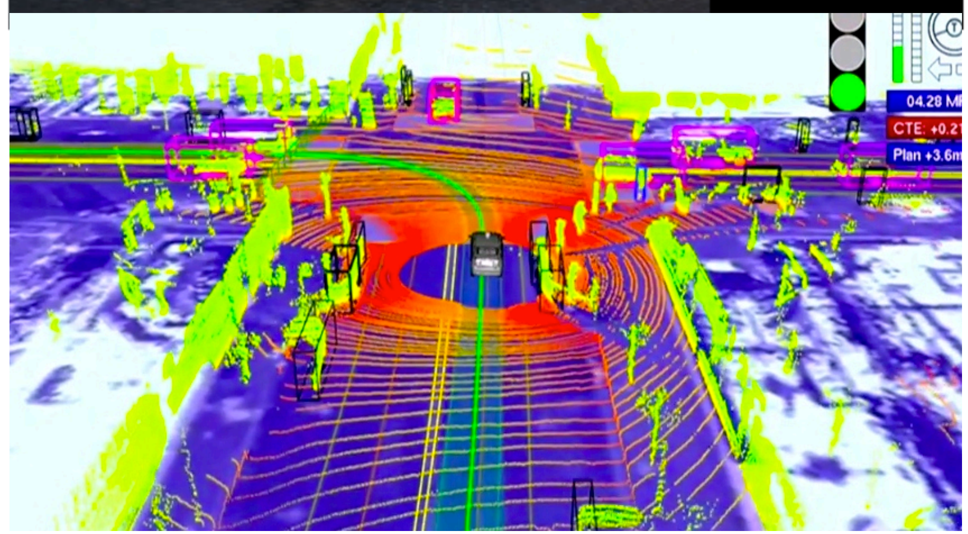
Growing fast

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

Examples







Machine Learning

Choose

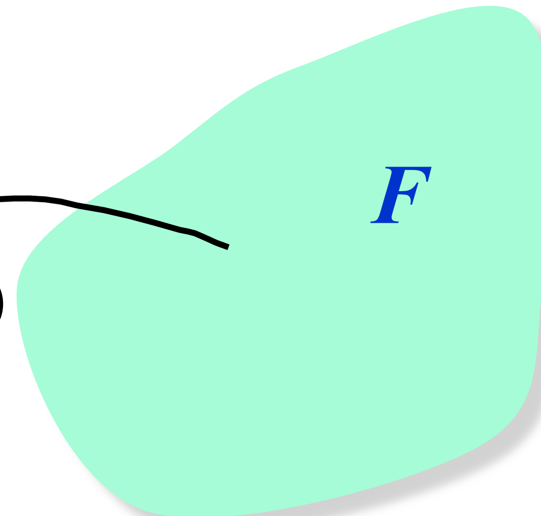
Function space $\boldsymbol{F} = \{ f(x, \boldsymbol{w}) \}$

Constraint $\boldsymbol{C}$

Loss function* $\boldsymbol{L}$

$C(\boldsymbol{w})$

$f(x, \boldsymbol{w}^*)$



Method

Find $f(x)$ by minimizing the empirical risk $R(w)$

$$R[f_w] = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i, w))$$

subject to the constraint $C(\boldsymbol{w})$

*The loss function measures the cost of choosing badly

Machine Learning

Many methods (e.g., neural networks, boosted decision trees, rule-based systems, random forests,…) use the quadratic loss

$$L(y, f(x, w)) = [y - f(x, w)]^2$$

and choose $f(x, w^*)$ by minimizing the ***constrained*** mean square empirical risk

$$R[f_w] = \frac{1}{N}\sum_{i=1}^{N}[y_i - f(x_i, w)]^2 + C(w)$$

History

1950s: First methods invented

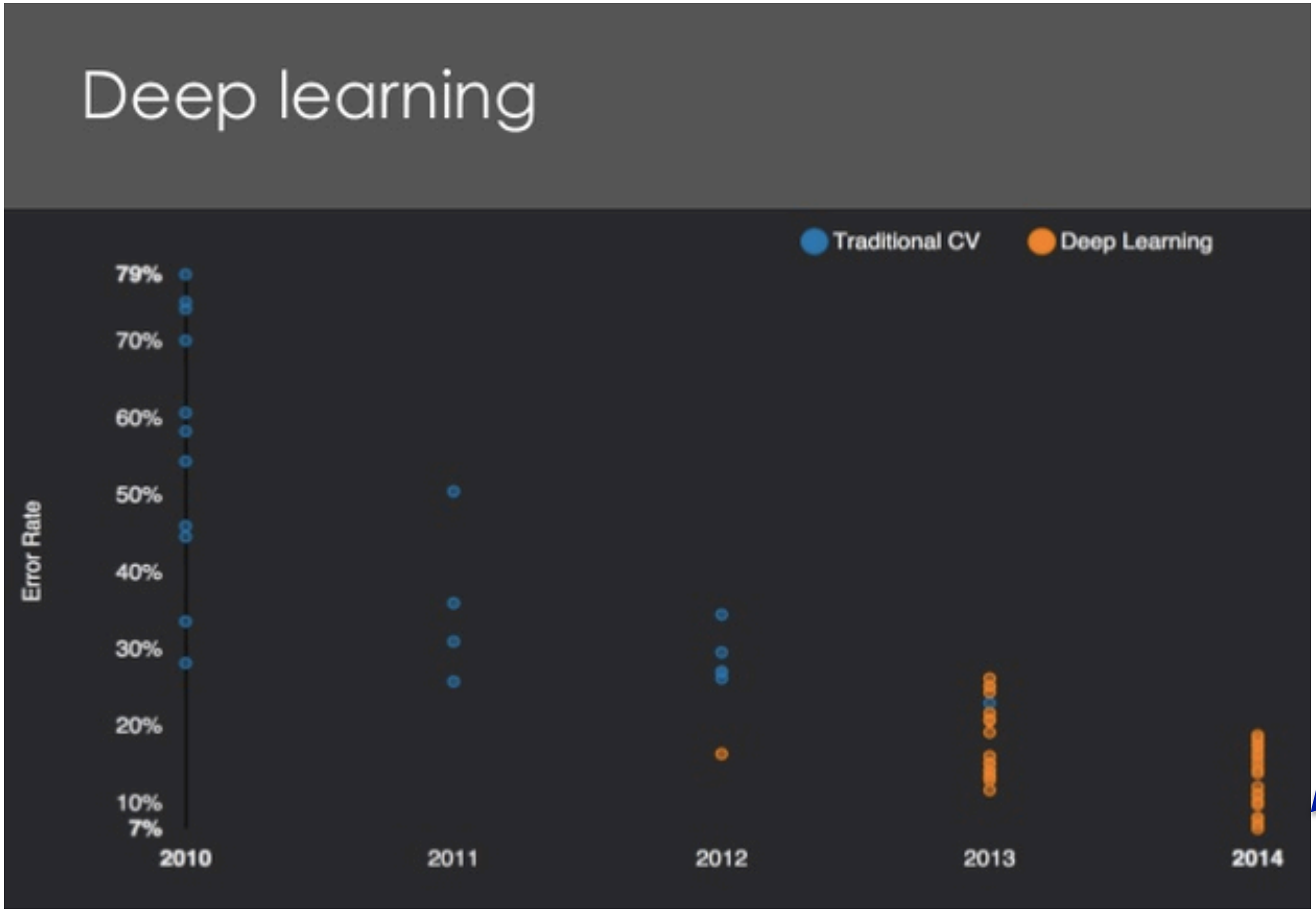
1960-80s: Slow growth, focus on knowledge

1990s: Growth of computing power, new learning methods, data-centric

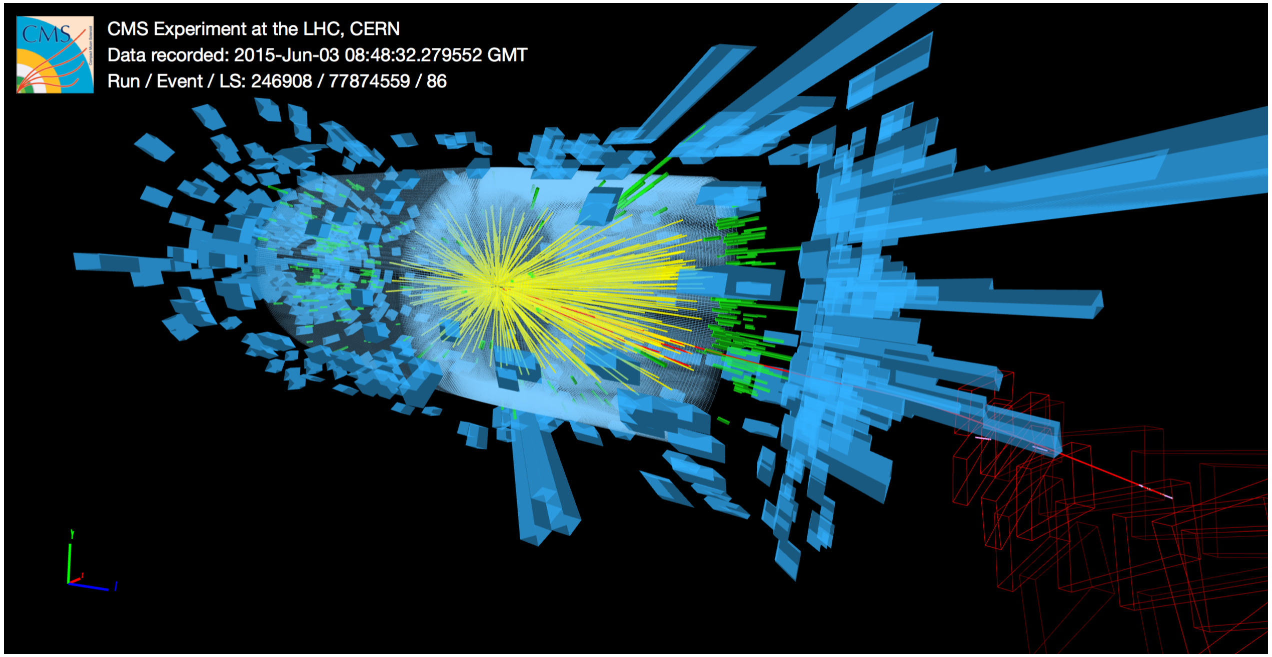
2000-10s: Wider use in research and industry

2010s: Deep learning improvement, specialized hardware

Diving Deeper



Huge Progress

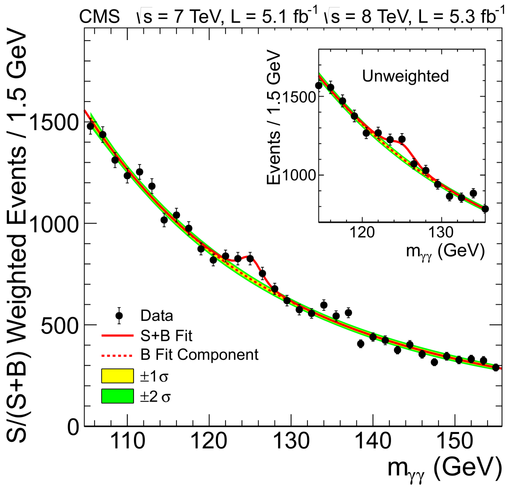


In Particle Physics

ML in HEP Today

Machine learning already at forefront of what we do:

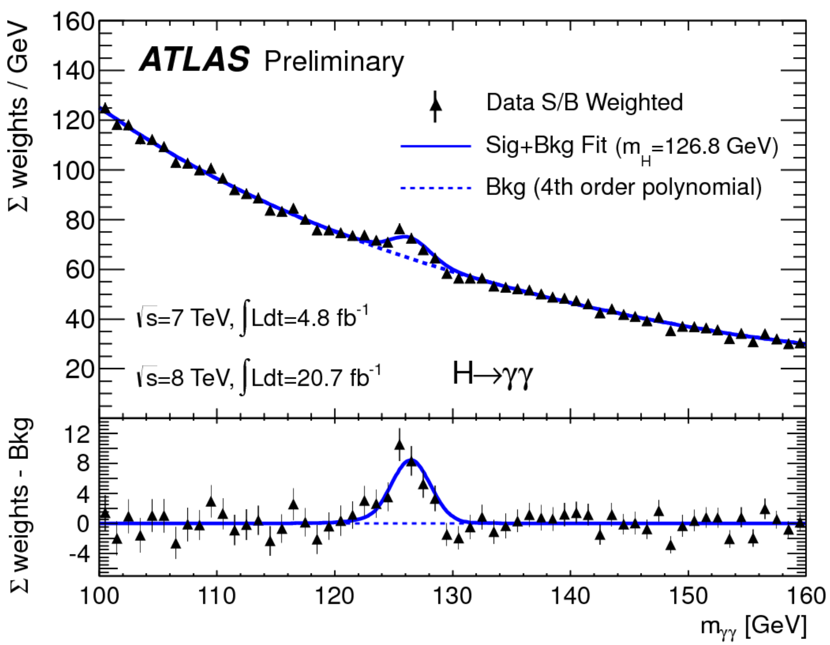
- Physics object **identification**
- Event type **classification**
- Object properties **regression**

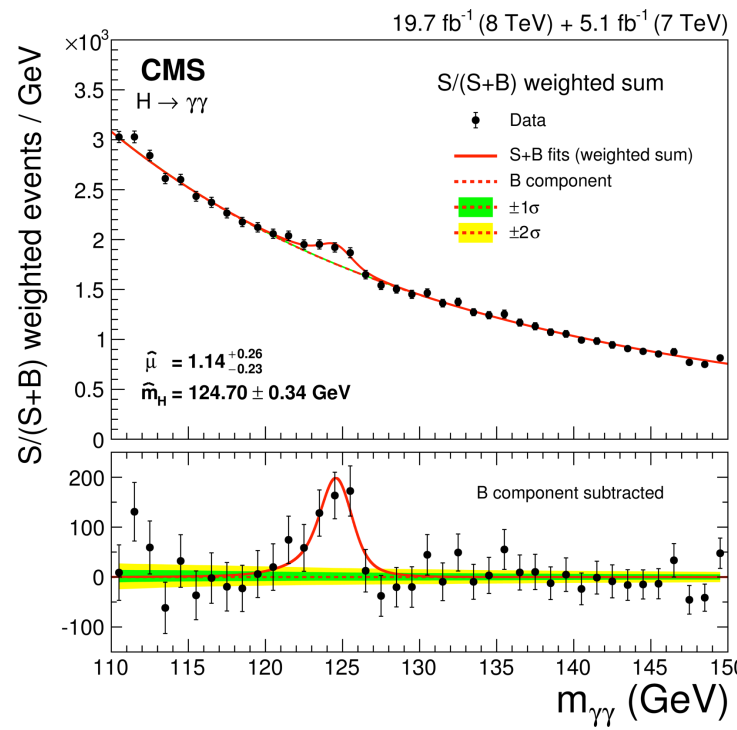


Expanding quickly to more areas

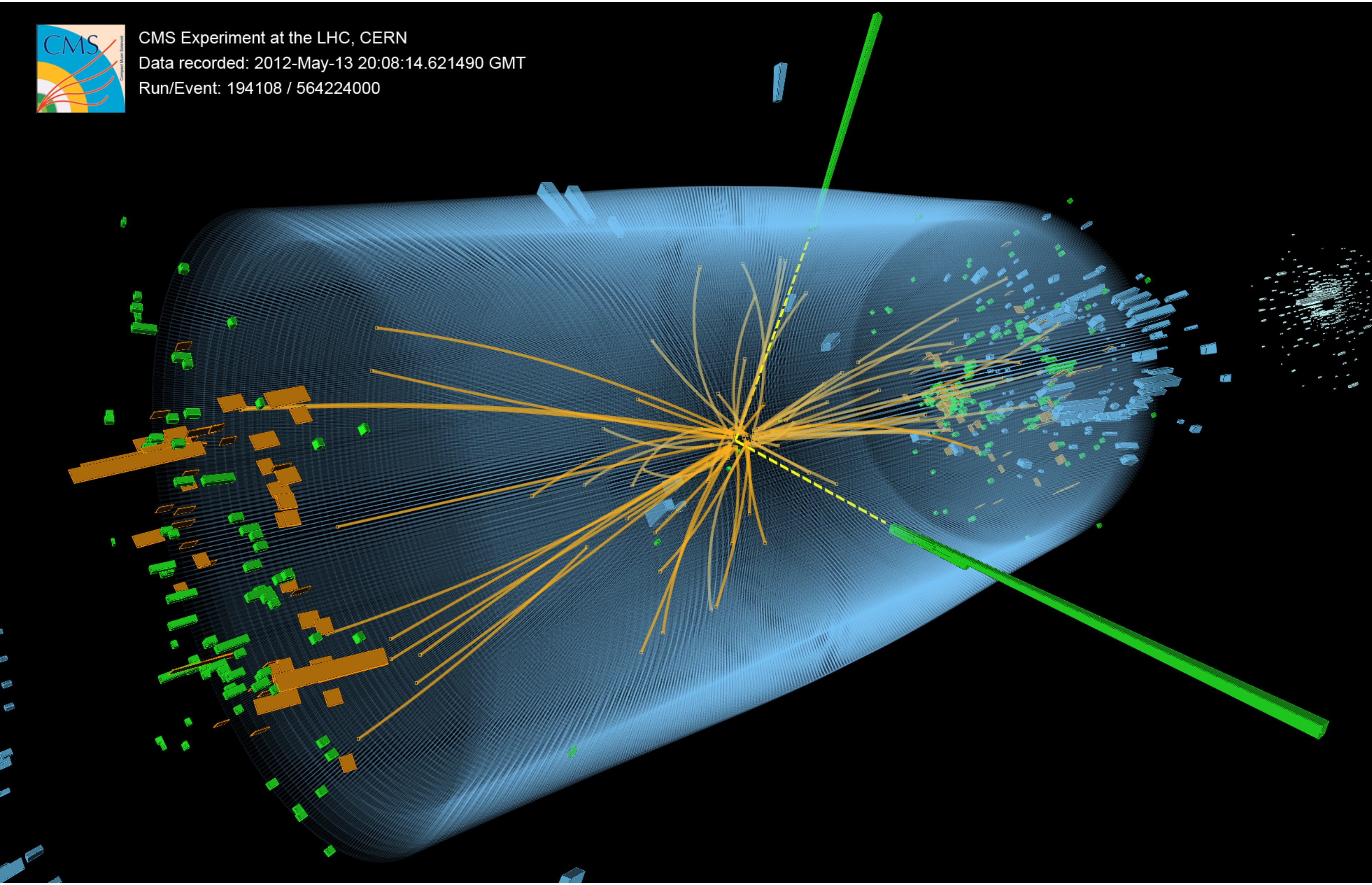
Higgs Boson Discovery

Higgs to di-photons

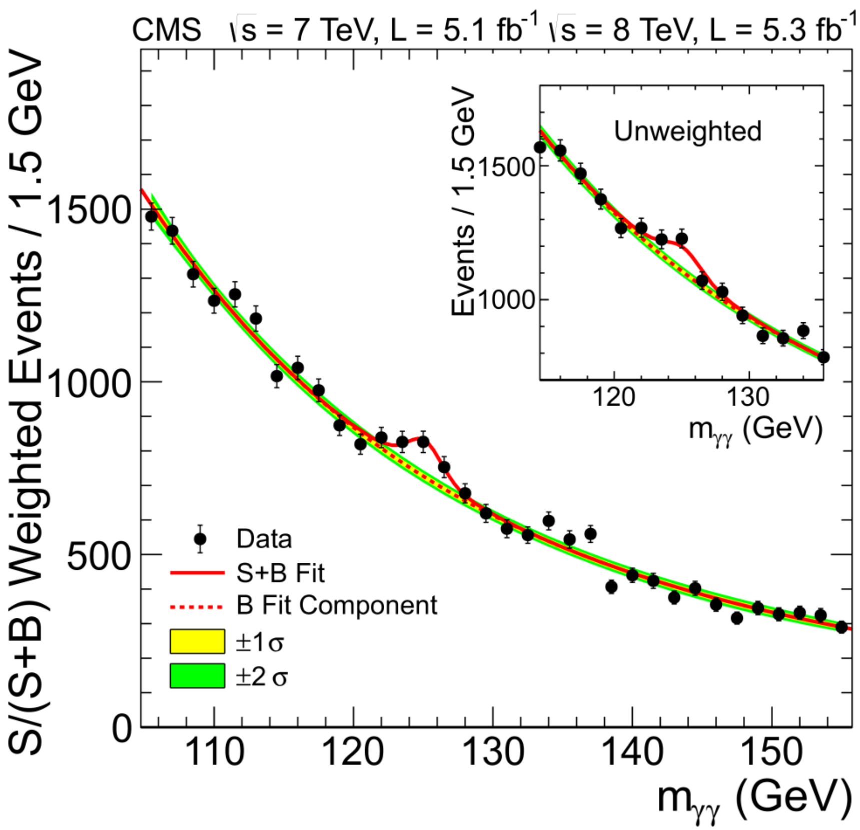




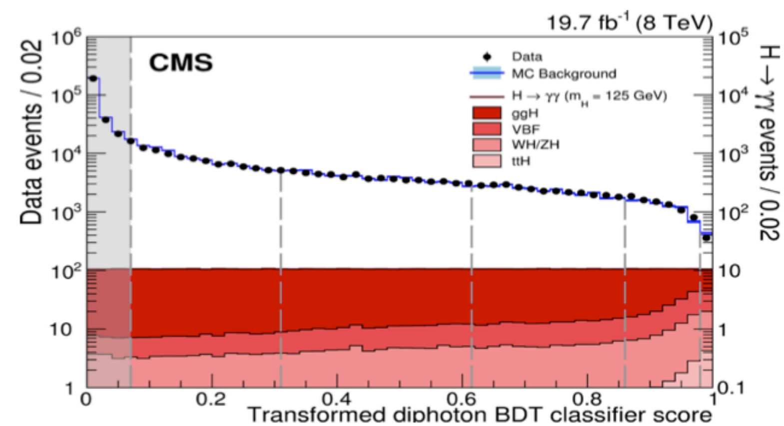
ATLAS **CMS**


CMS Experiment at the LHC, CERN
Data recorded: 2012-May-13 20:08:14.621490 GMT
Run/Event: 194108 / 564224000

in Higgs Discovery



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

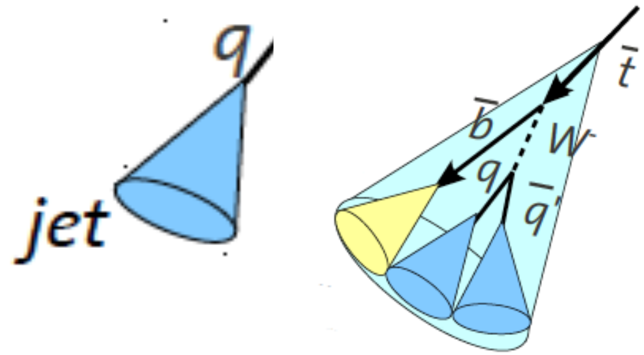


Improvement in analysis from all four areas

Applications

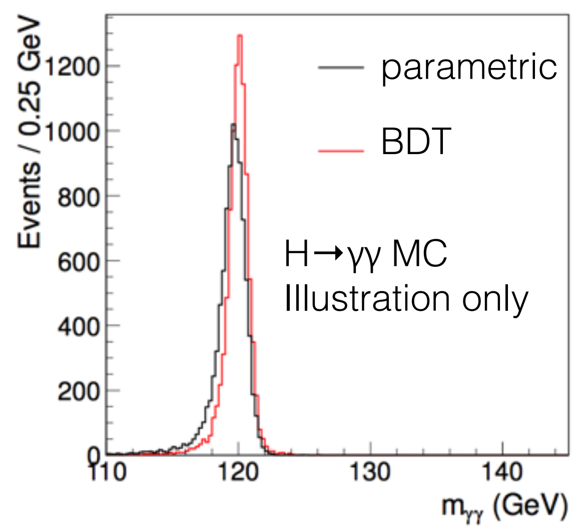
I. Classification

- Particle Identification
- Pattern Recognition (tracks)
- Searches for New Physics
- Data Quality Monitoring



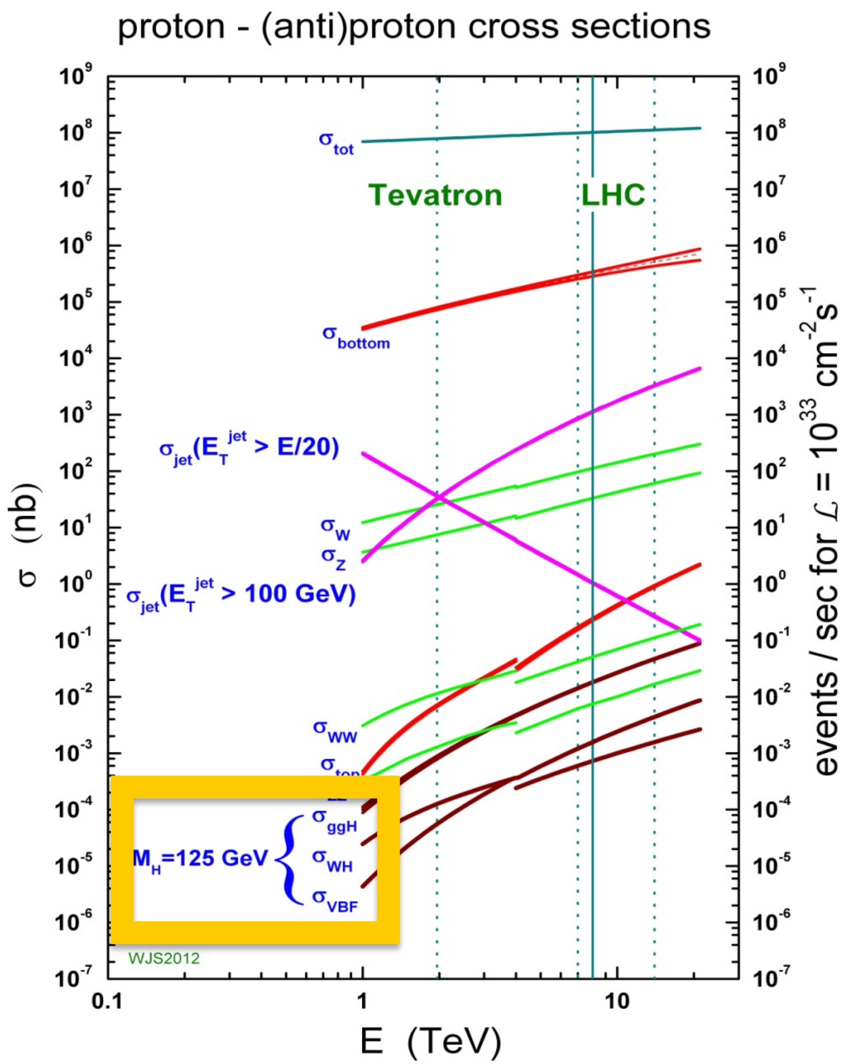
II. Function estimation

- Particle Properties
- Regression



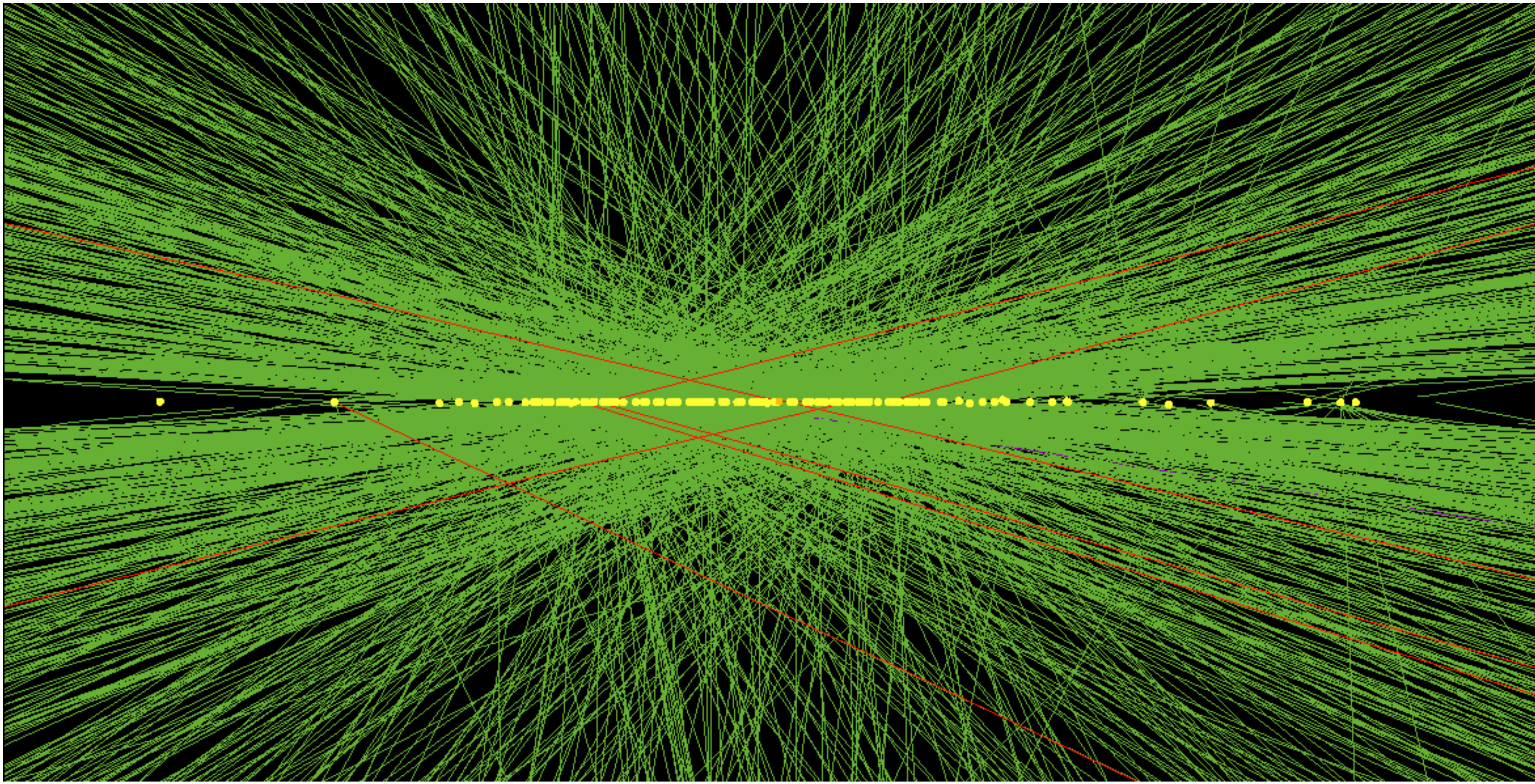
Challenges

Challenges



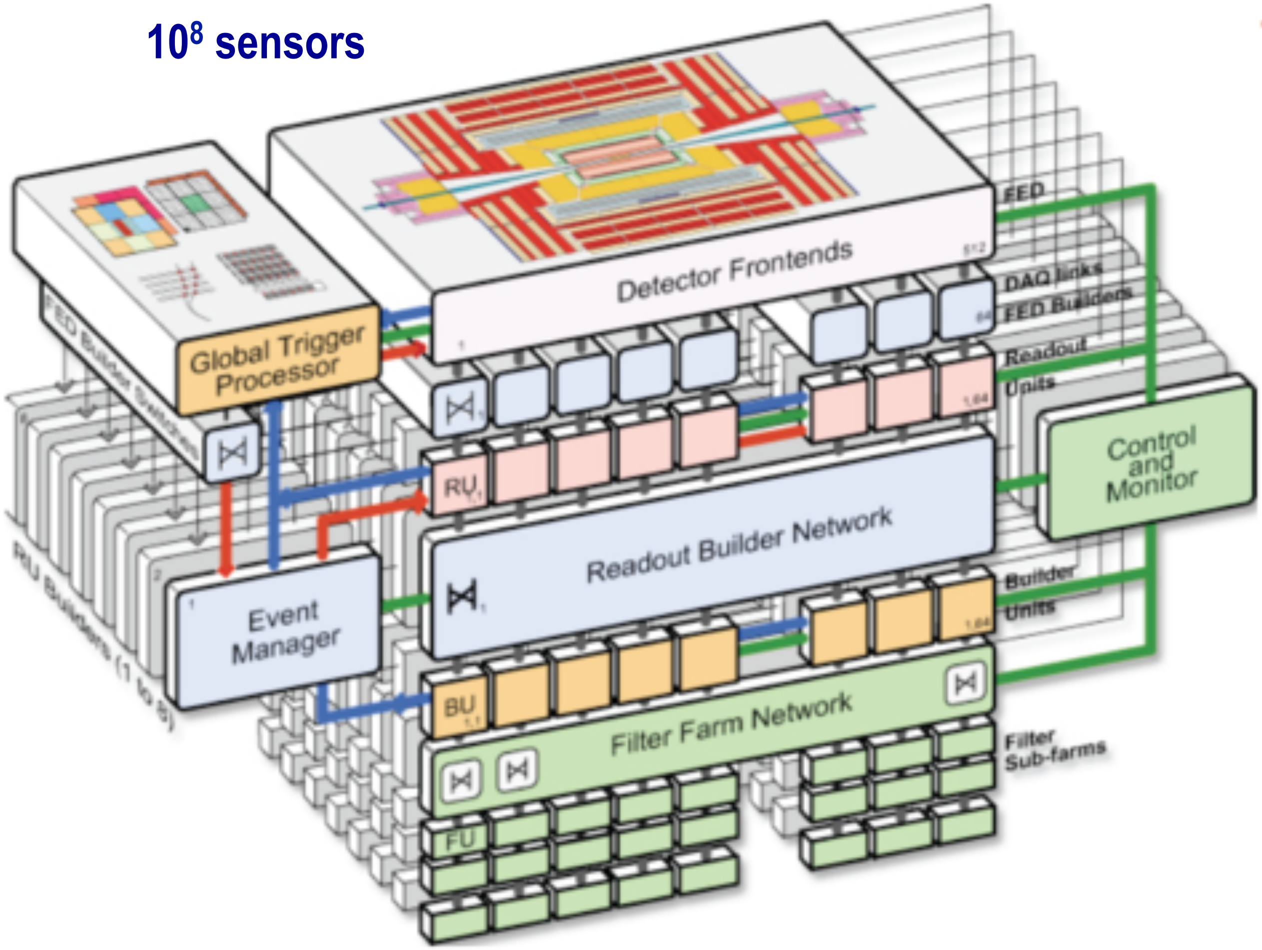
Orders of magnitude between signals and backgrounds

Event Complexity

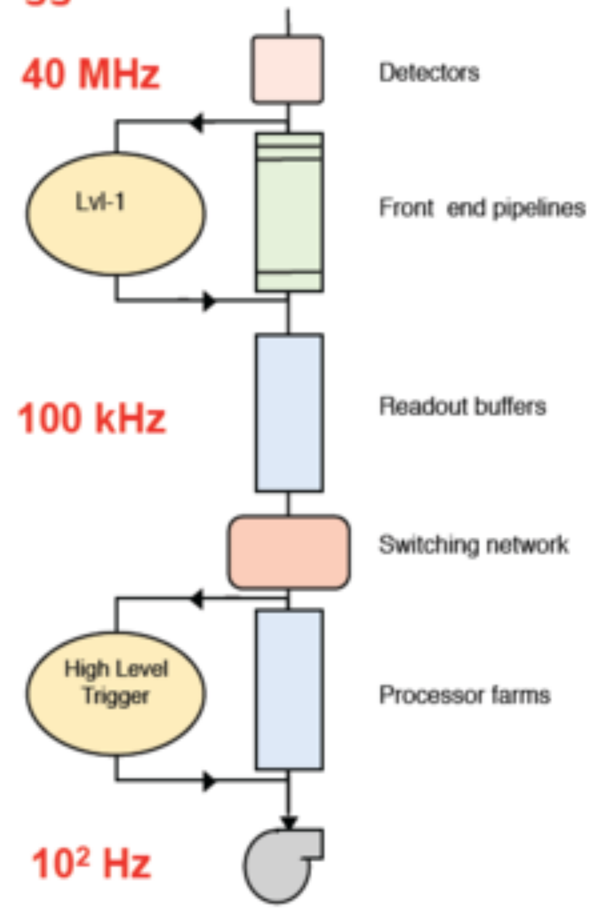


Event Filtering

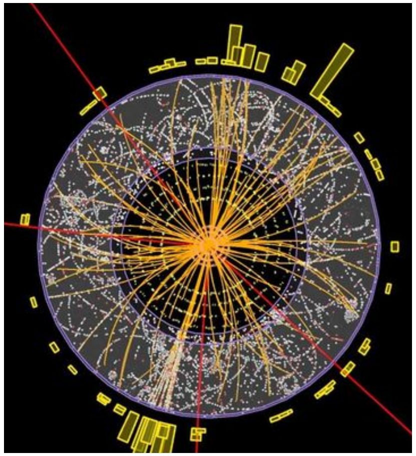
10^8 sensors



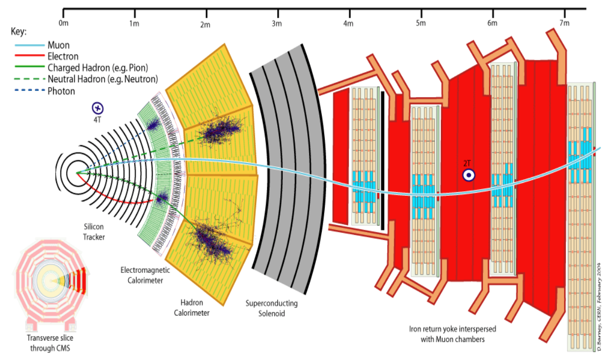
Trigger Rate



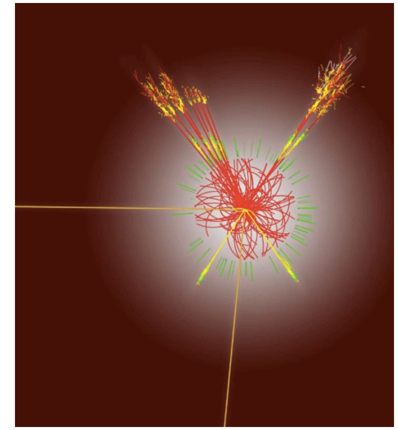
Interesting areas



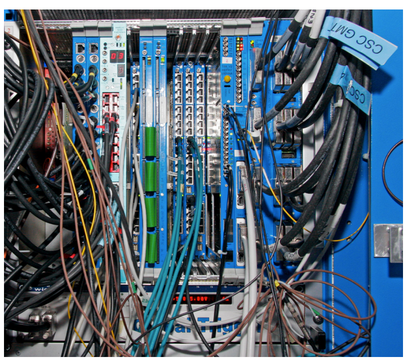
Tracking



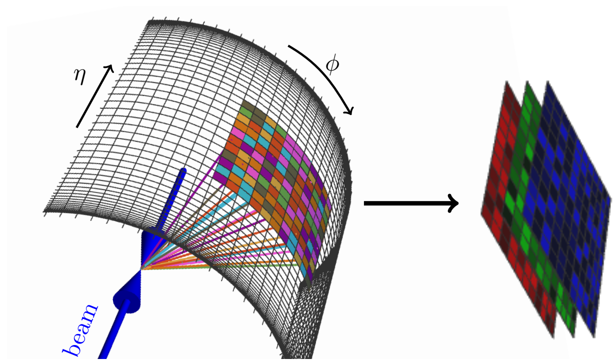
Fast Event Simulation



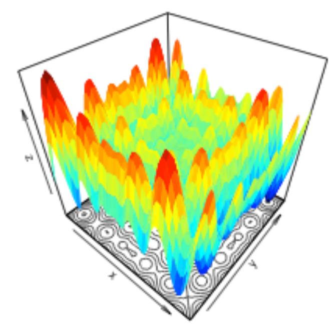
Object Identification



Event Filtering

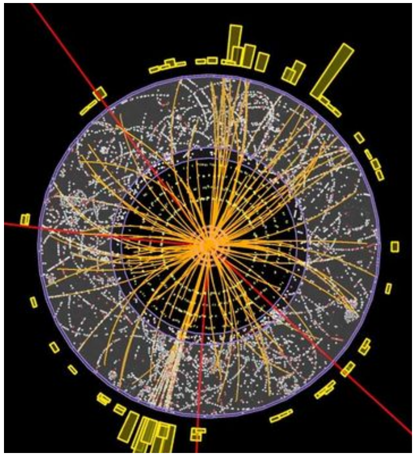


Imaging Techniques

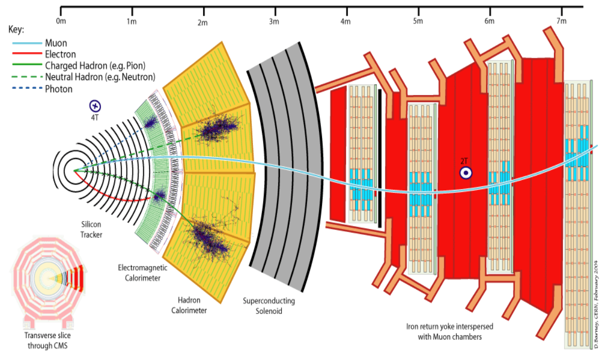


Simulation

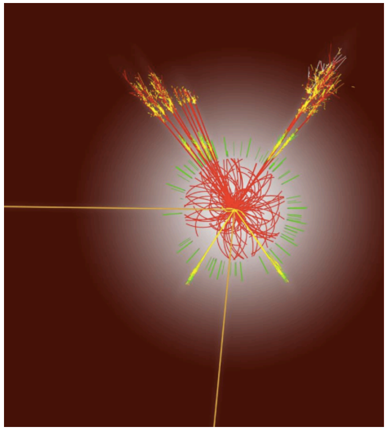
Interesting areas



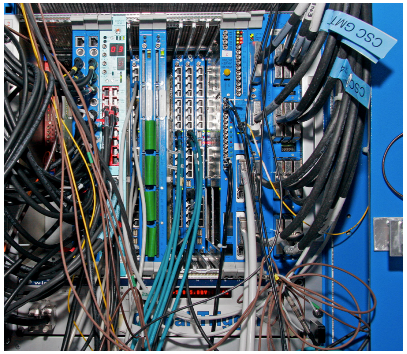
Deep Kalman Recurrent, LSTMs



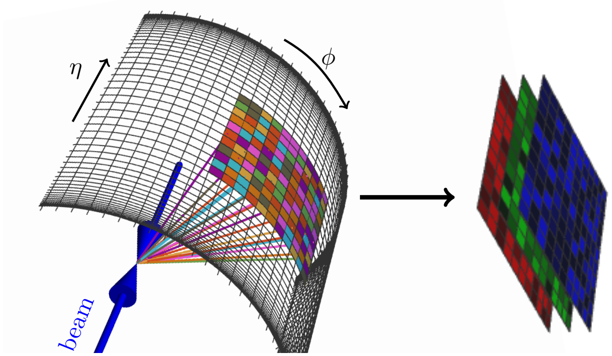
Generative Models, Adversarial Networks



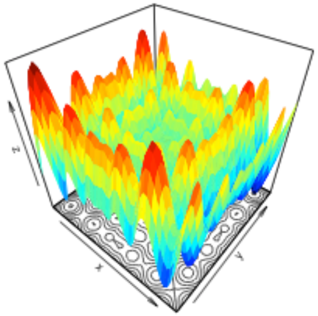
FCN, Recurrent, LSTMs



Deep ML +FPGA



Convolutional DNN

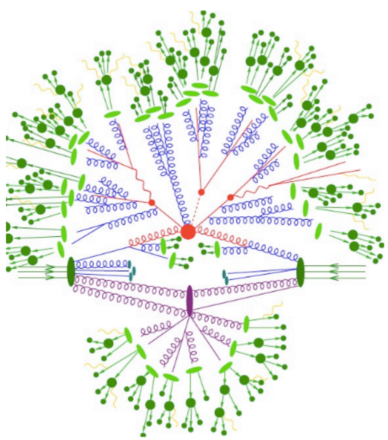


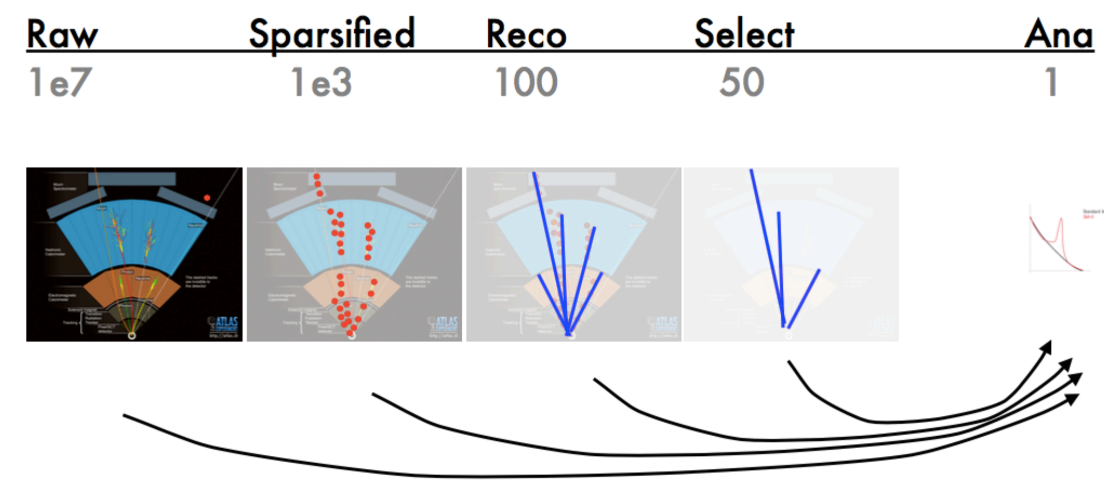
Multiobjective Regression

Questions

Can we fully exploit the detectors?

- Raw data, low-level variables





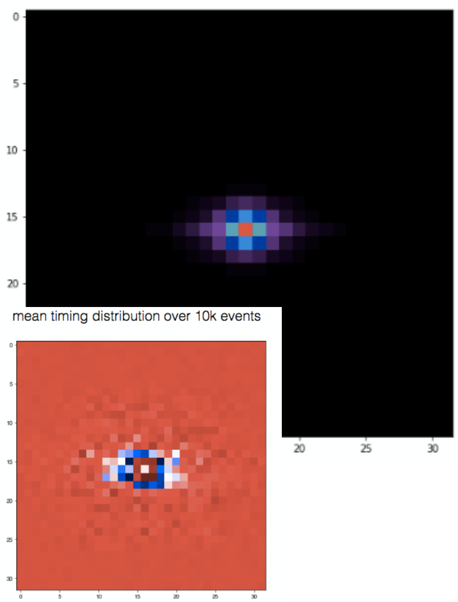
Images: D. Whiteson, K. Cranmer

Example

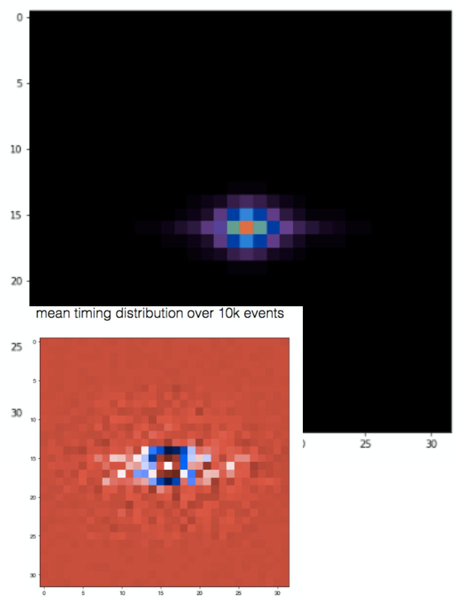
“End-to-end learning”

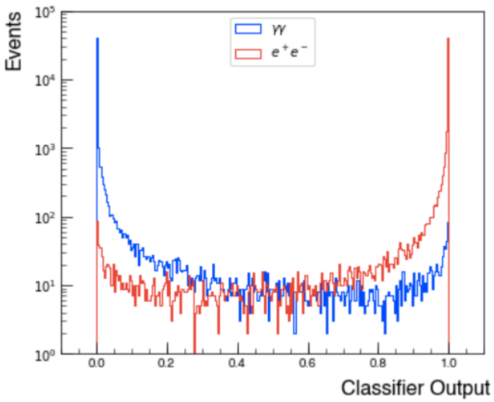
- By-passing traditional reconstruction

Photon-Induced EM Shower
mean energy distribution over 10k events

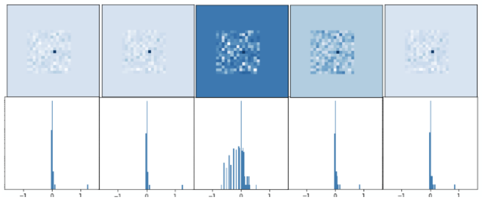


Electron-Induced EM Shower
mean energy distribution over 10k events





	ResNet-23
Test Set ROC AUC	0.997



Defining the Problem

If a problem can be expressed as a known problem

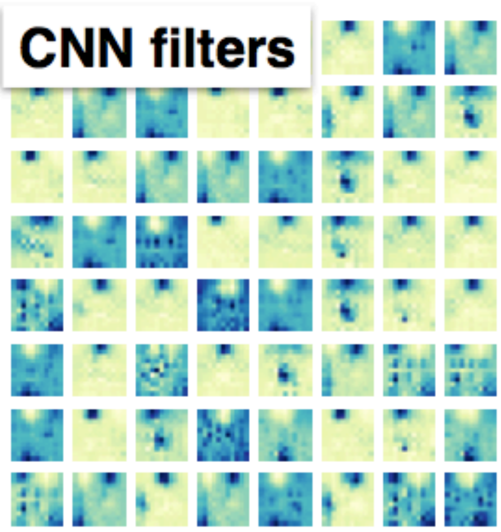
- Apply **existing algorithms**
  - **Example: convolutional neural networks from computer vision**

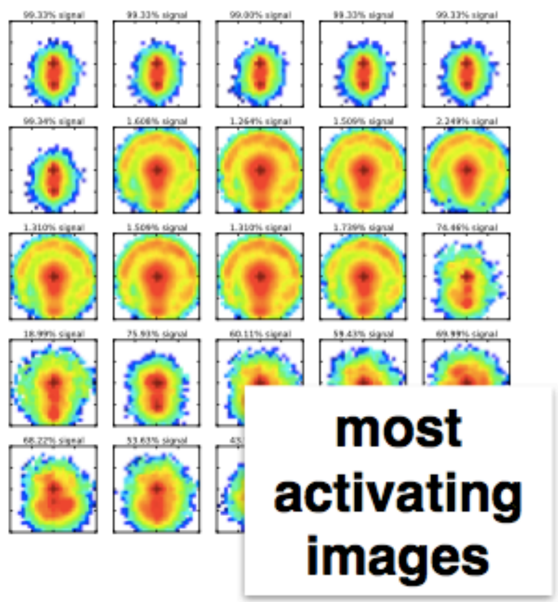
If a problem has not been solved

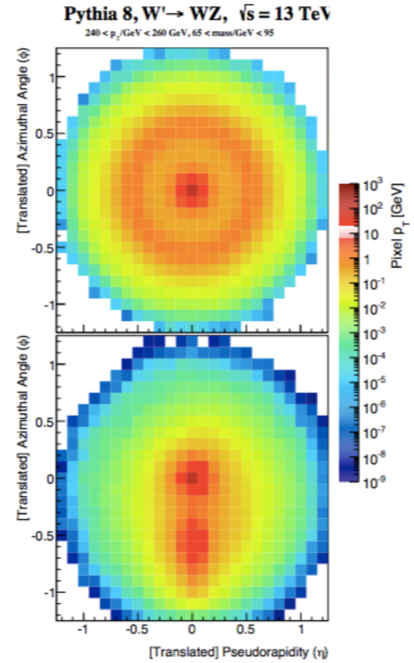
- Push the knowledge boundary forward

Examples

Jet images with convolutional nets



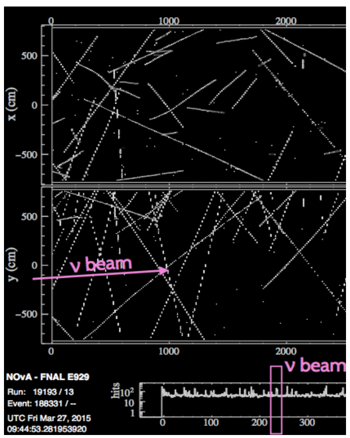


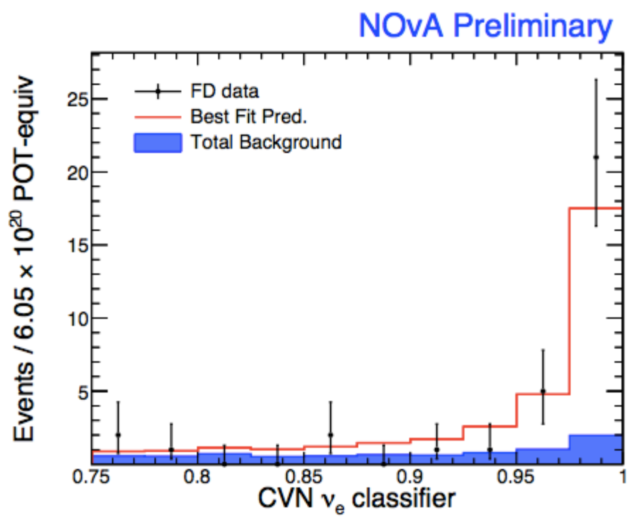


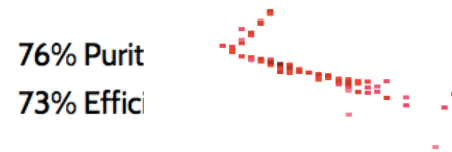
L. de Oliveira et al., 2015

Examples

Neutrinos with convolutional nets

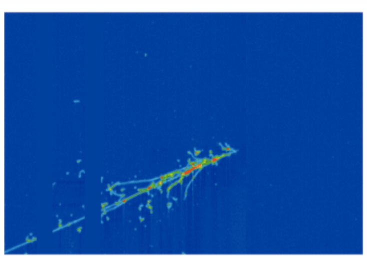


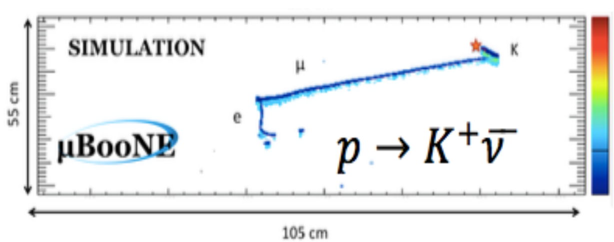


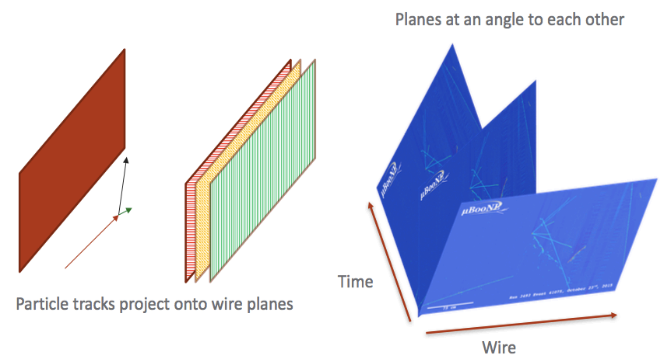


An equivalent increased exposure of 30%

Aurisiano et al. 2016



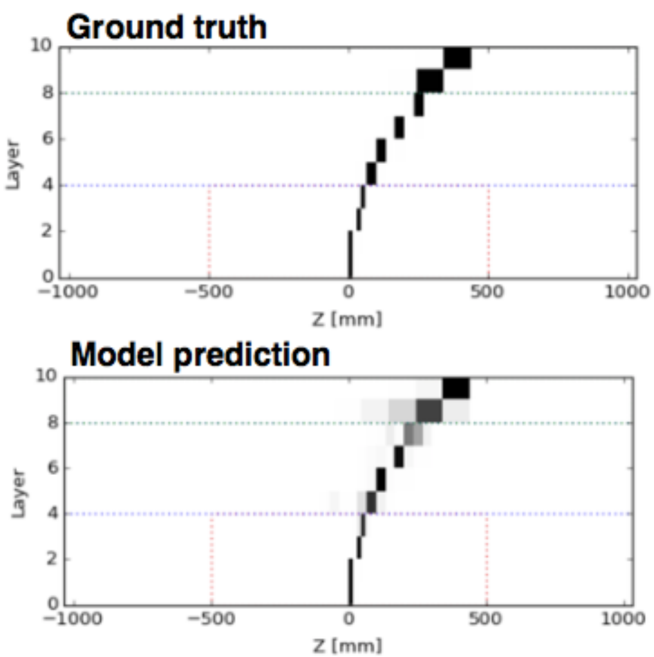


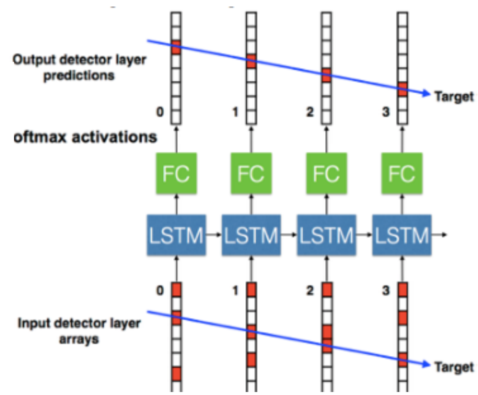


μBooNE

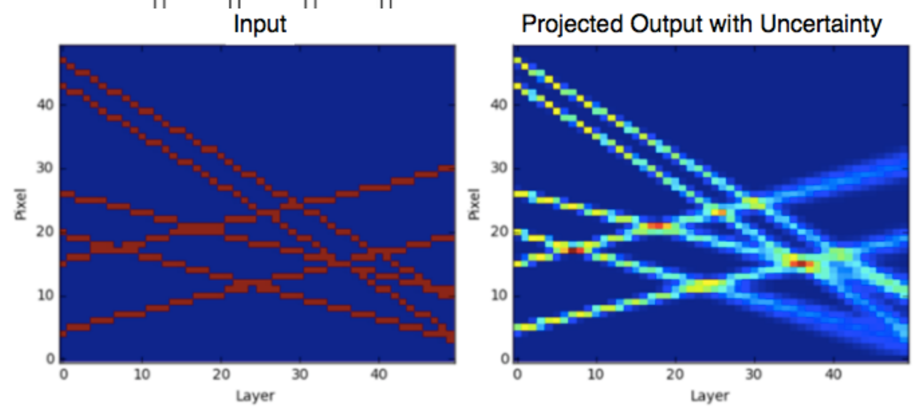
Examples

Tracking with recurrent nets (LSTM)





Time dimension (state memory)

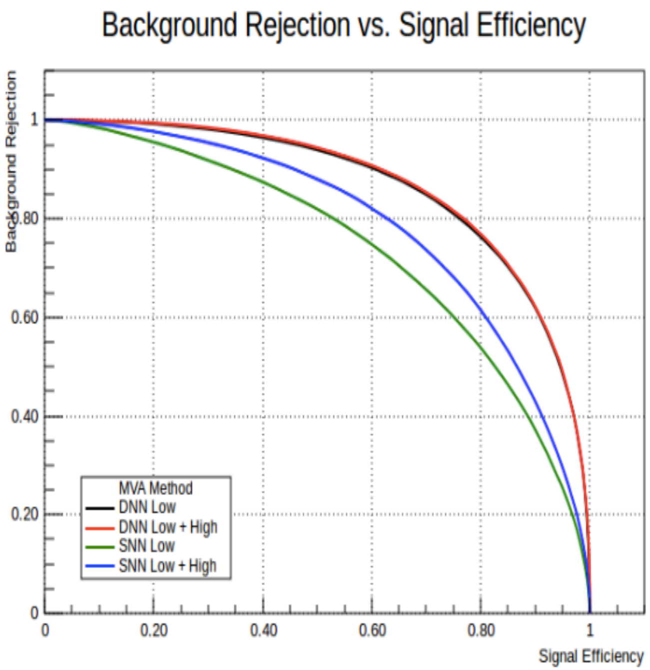


HEPTrkX Project

Meaningful Physics

Can we extract features with meaningful physics?

- from low-level variables

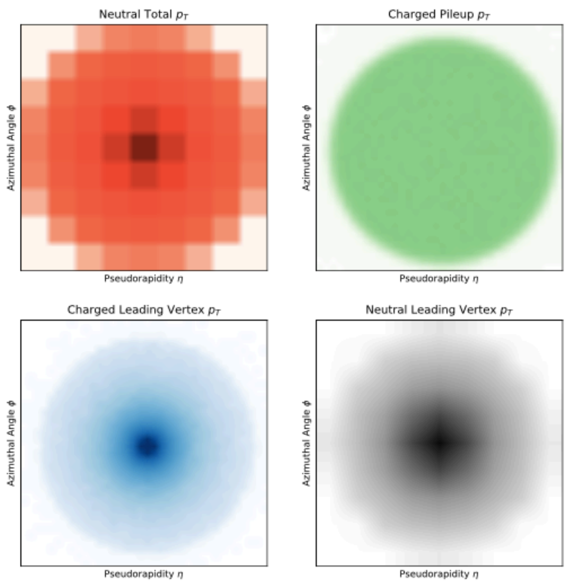


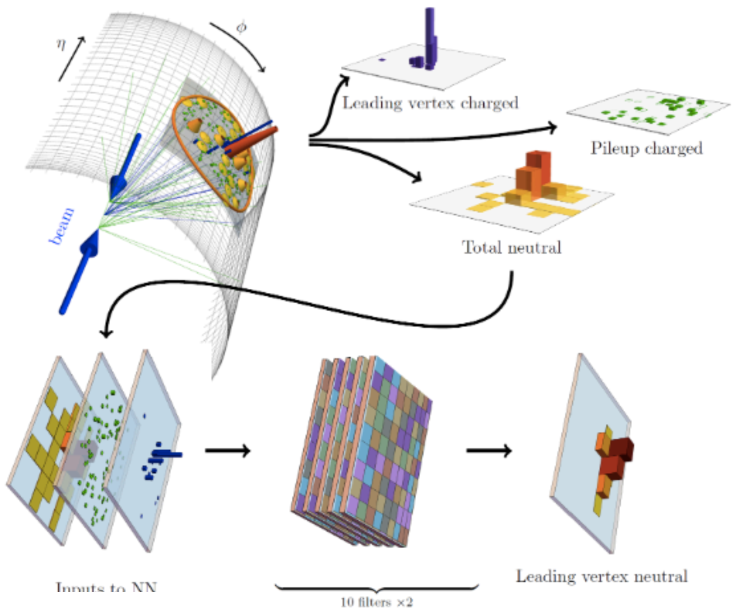
Are we able to understand ML models

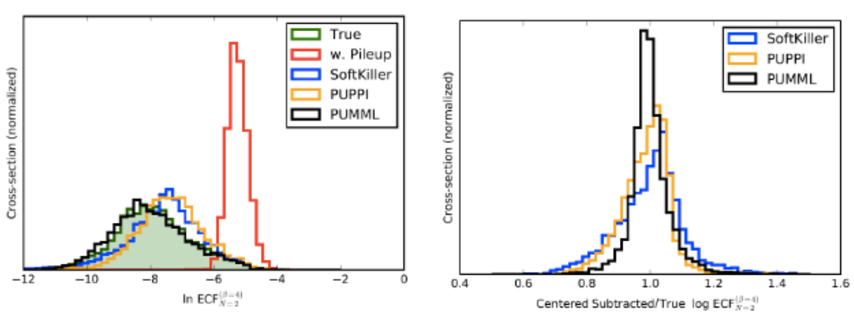
- physics interpretations

Meaningful Physics

Pile-up removal with CNN



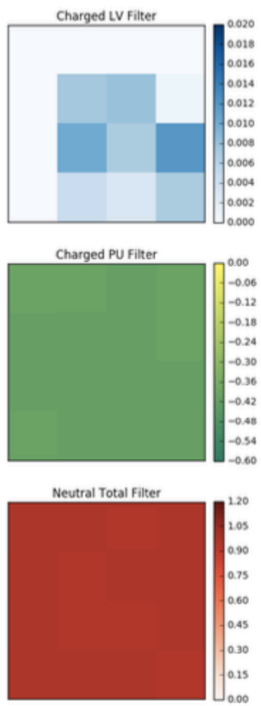




E. Metodiev et al., 2017

Meaningful Physics

Pile-up removal with CNN

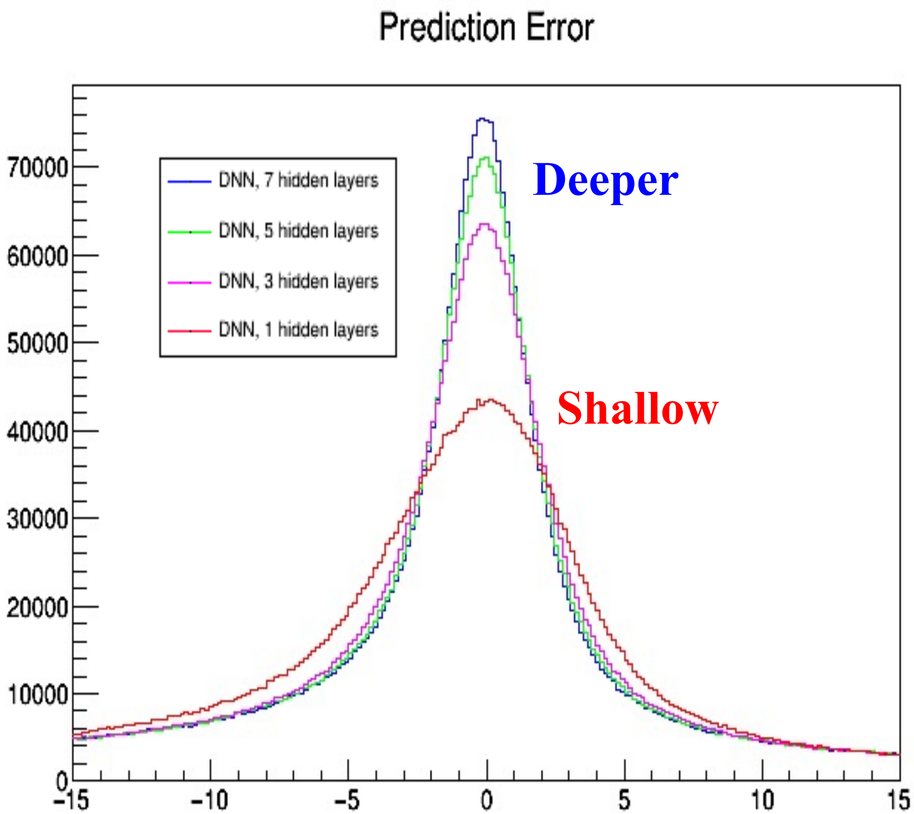


What is learned?

- 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 $\bar{\gamma}_0 = 2/3$!

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

Deep Learning Regression



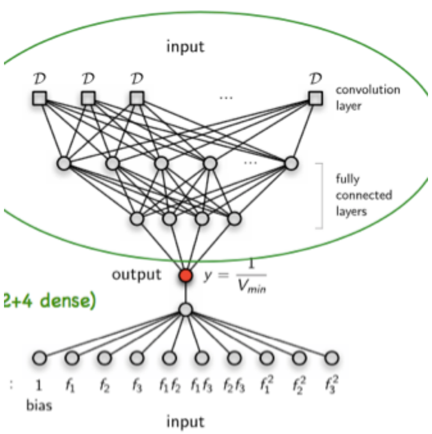
Deep learning improvements apply to regression as well

Defining the Problem

How to best use domain knowledge we have accumulated?

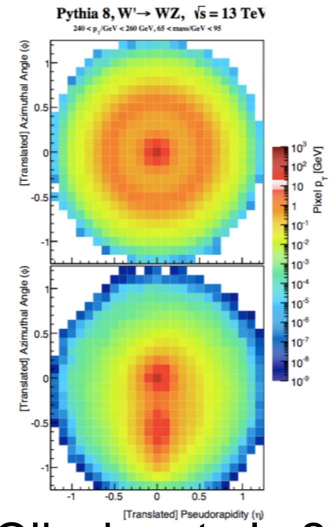
- in designing the algorithms

Strings



Krefl, 2017

Jet Images

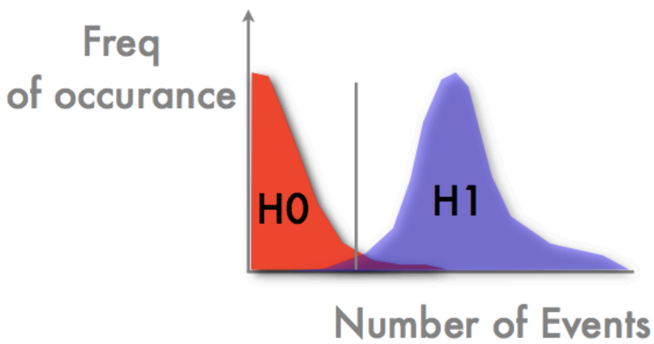


de Oliveira et al., 2015

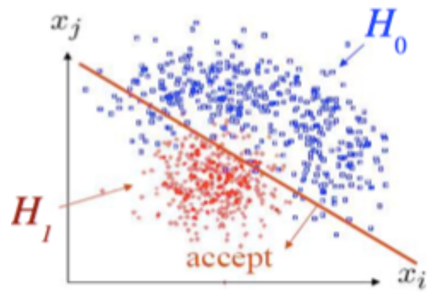
Jet Clustering

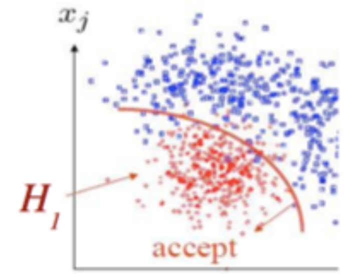
Louppe et al., 2017

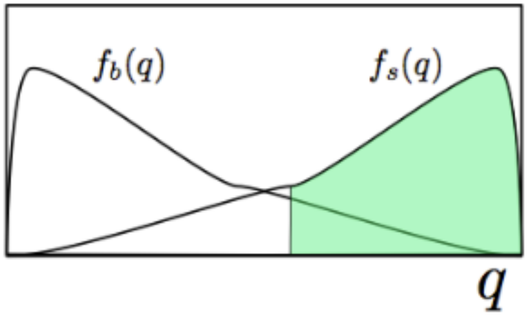
Uncertainties



A threshold makes sense.



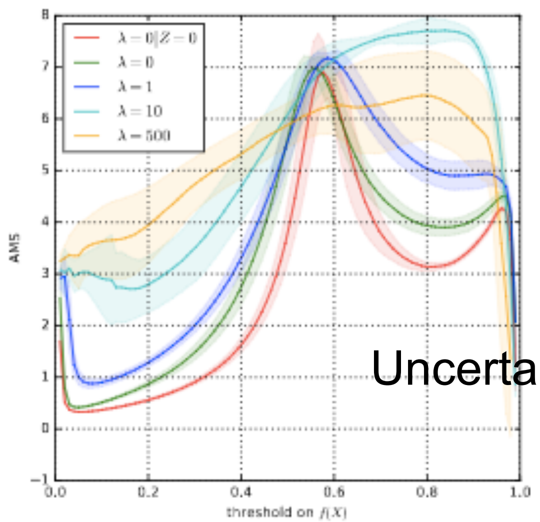




- **Decision making**

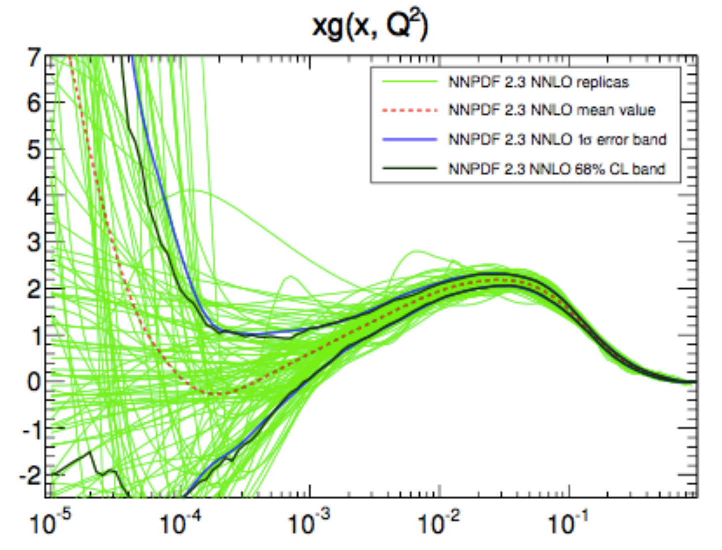
Uncertainties

G. Louppe et al., 2016



Uncertainties matter

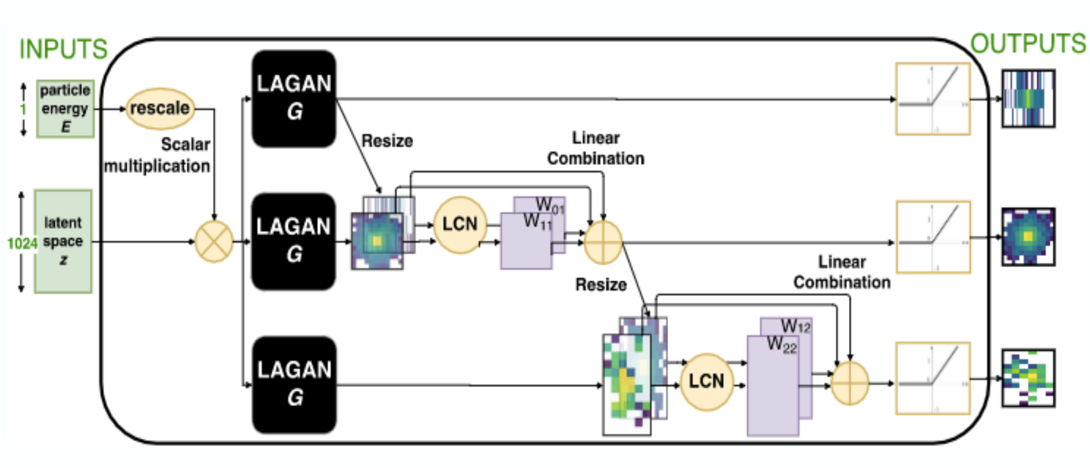
NNPDF Collaboration



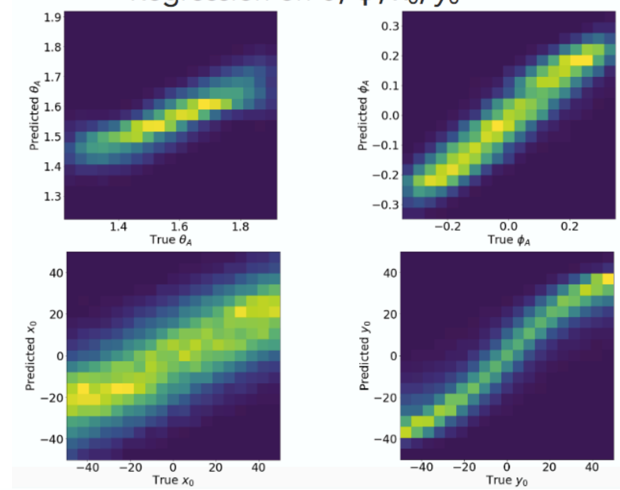
Bayesian connection: Deep neural networks with drop-out approximate variational inference of Bayesian NNs: *Gal and Ghahramani, 2016*

Additional Uses

Simulation GANs

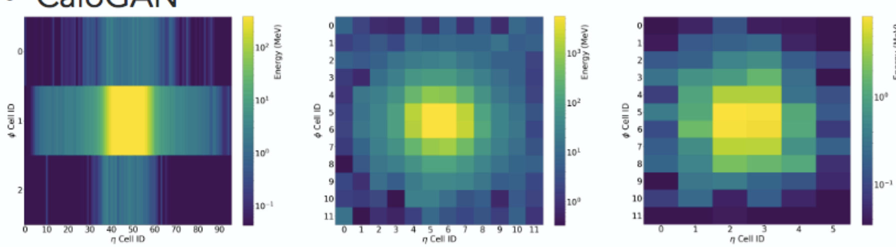


- Regression on θ, φ, x_0, y_0

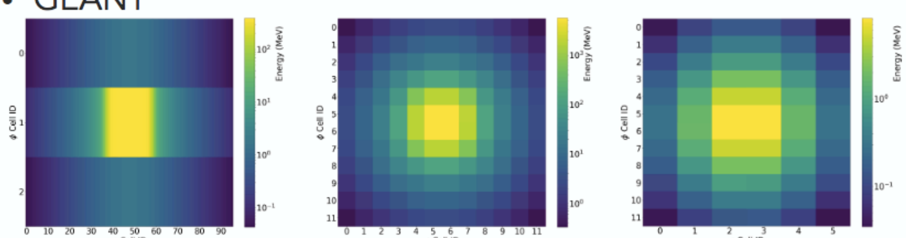


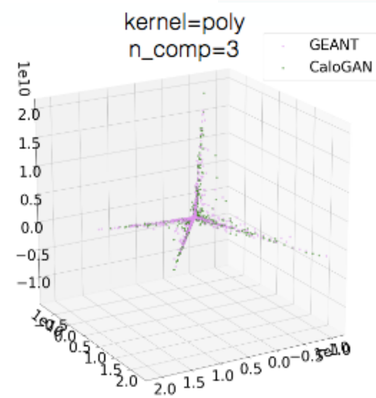
Dataset: 5˚; Net: soft sparsity, multiplied E, Conv. attn. and layers

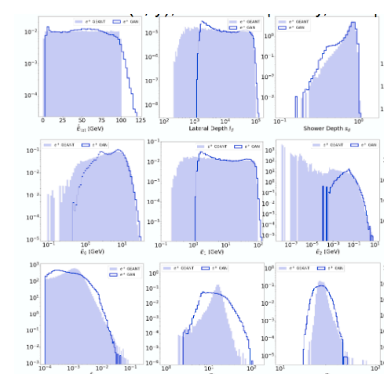
- CaloGAN



- GEANT

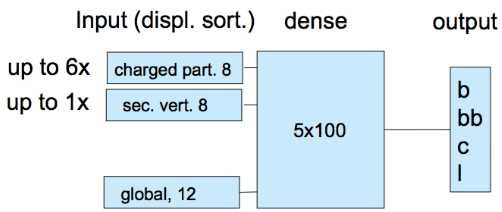


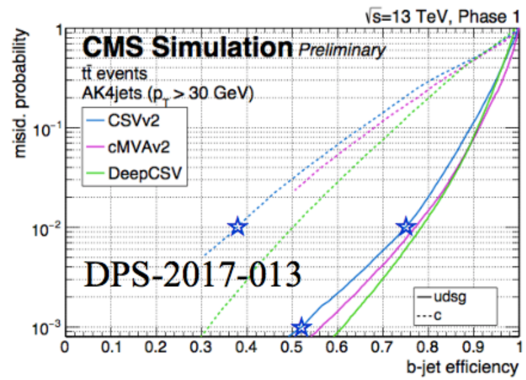


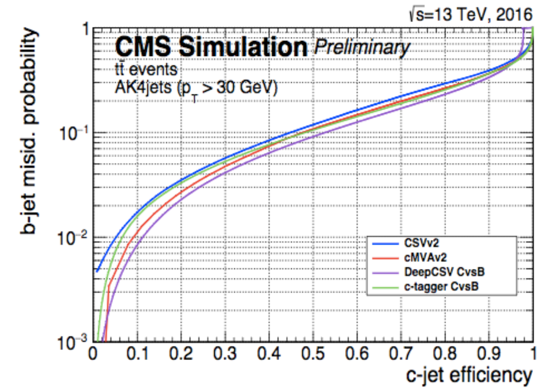


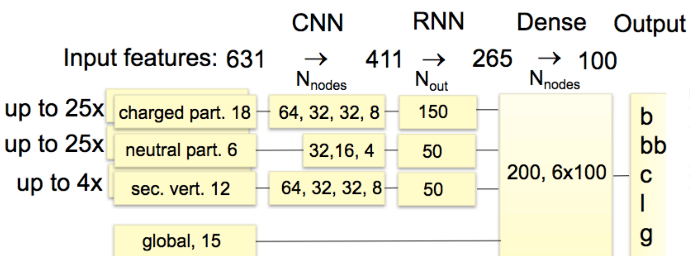
L. de Oliveira et al., 2017

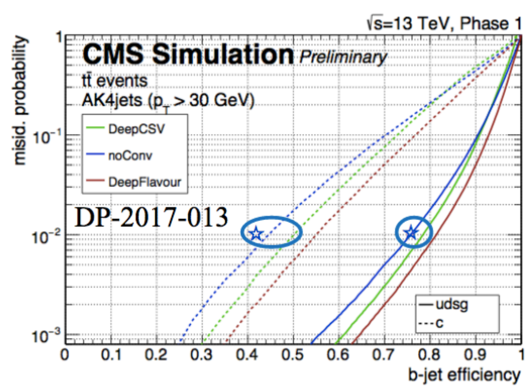
Flavor Tagging

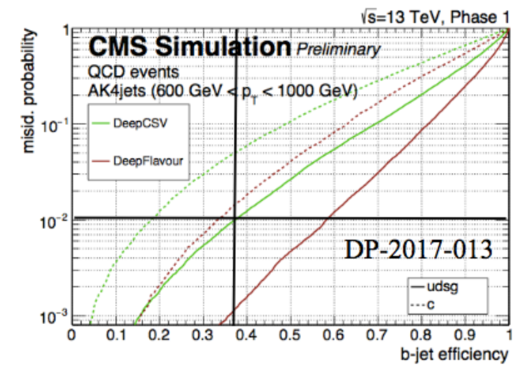


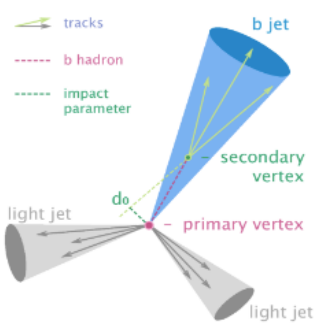








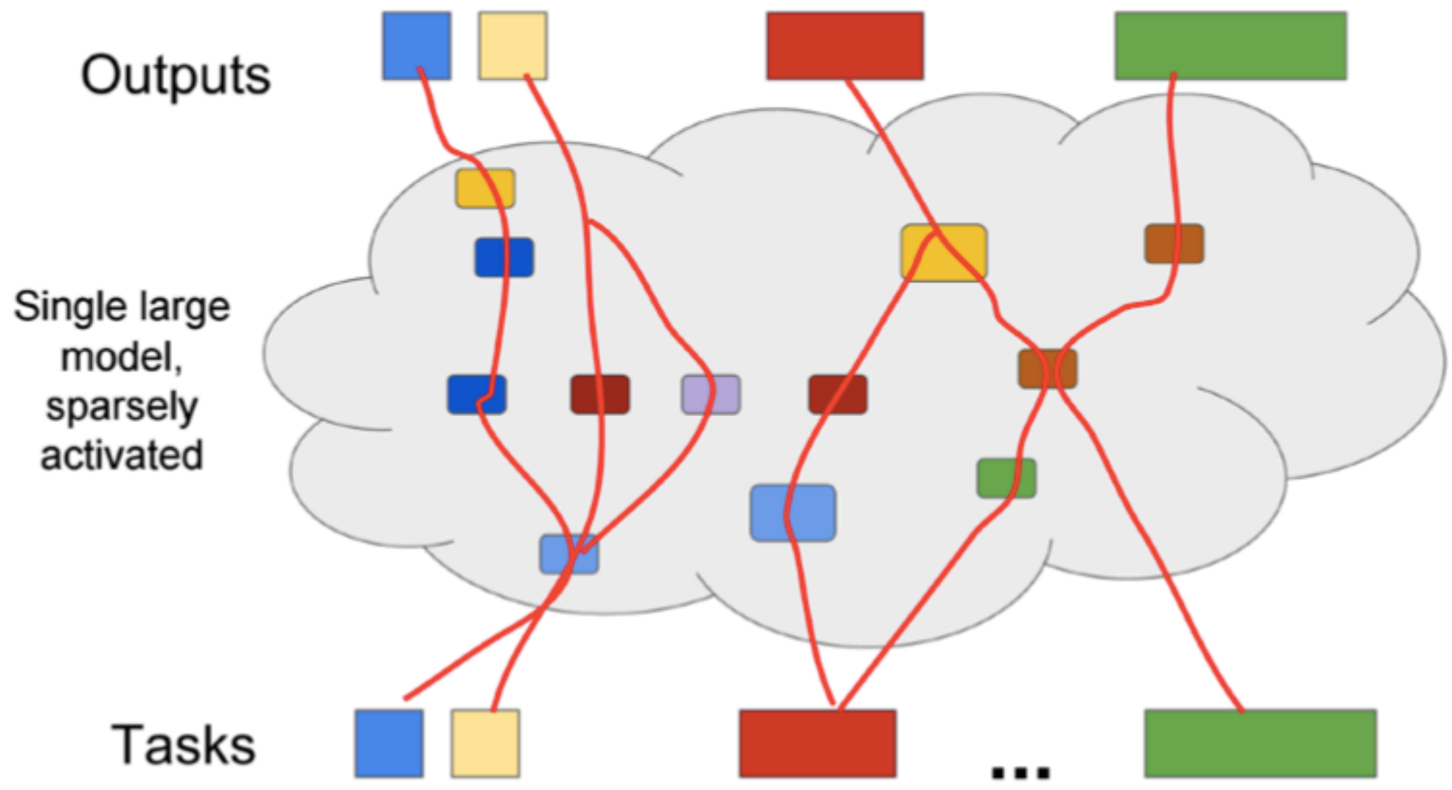




1.7	2.8	Baseline, $p_T^{miss} > 300$, 4b	4.6±0.4	2.5	4.0

e.g. up to ~50% more signal for 15% more bkg

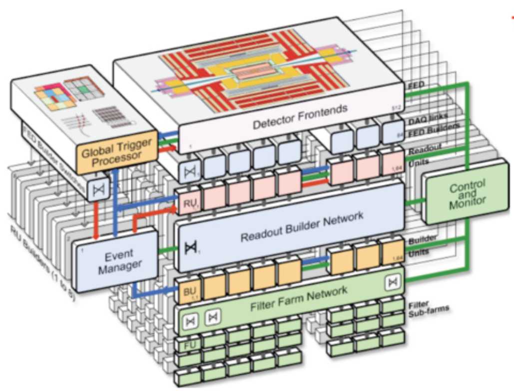
Multi-Task Model

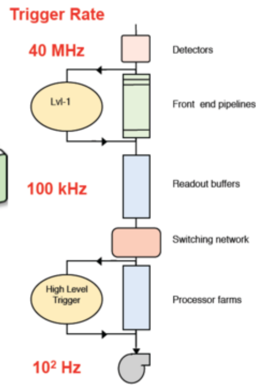


Real Time Application

Can we do ML in real-time?

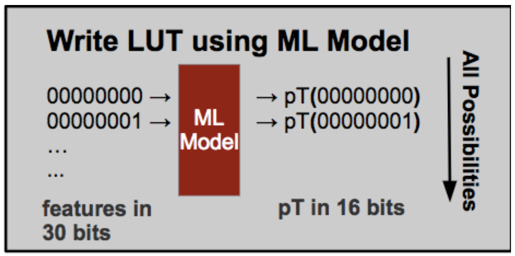
- ML: live video analysis, medical, self-driving cars
- HEP **Trigger Systems** (software and hardware)

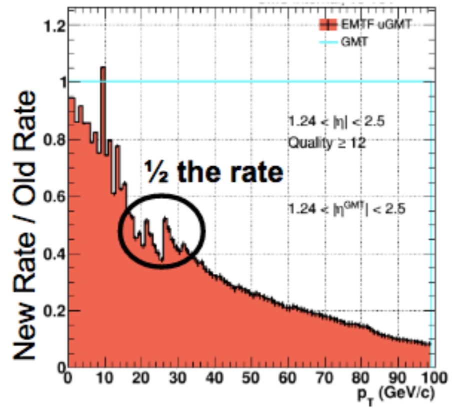


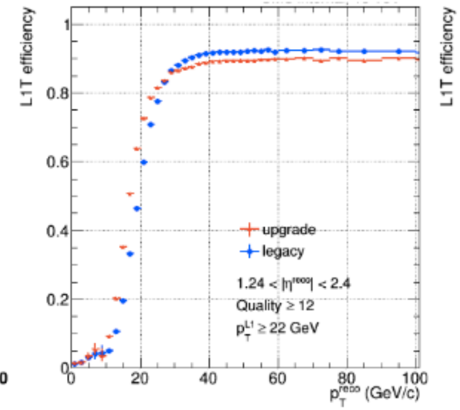


Trigger Applications

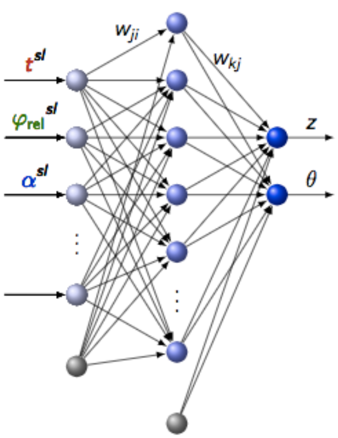
CMS L1

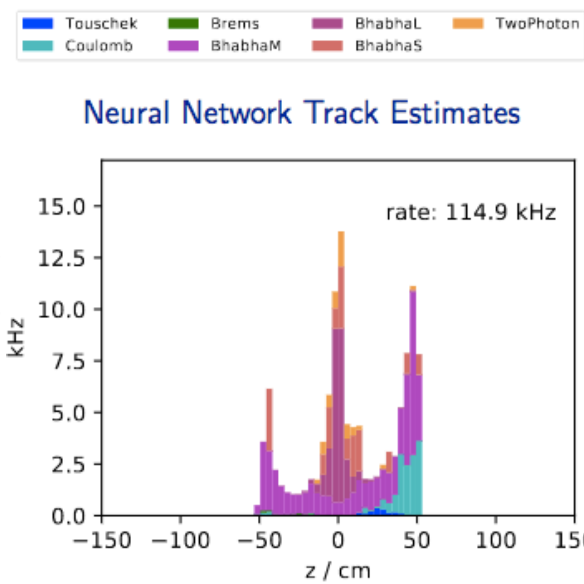


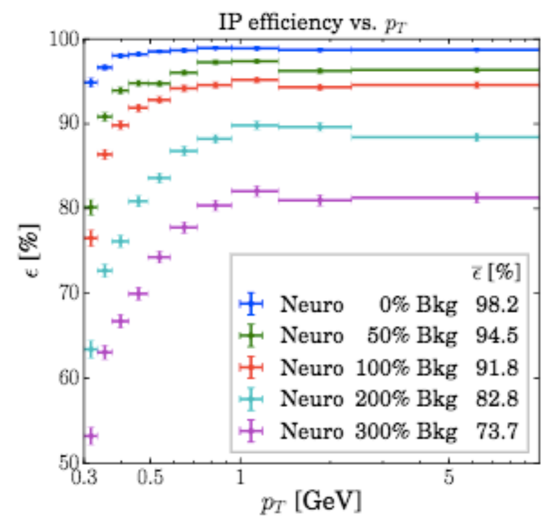




Belle II







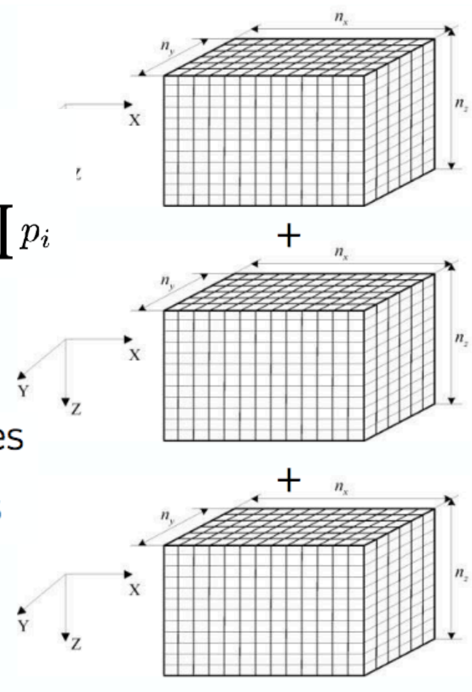
Trigger Applications

LHCb

〉 $D = (f_1, \ldots, f_n)$

$f_i = \{b_{i1}, \ldots, b_{ip_i}\}$

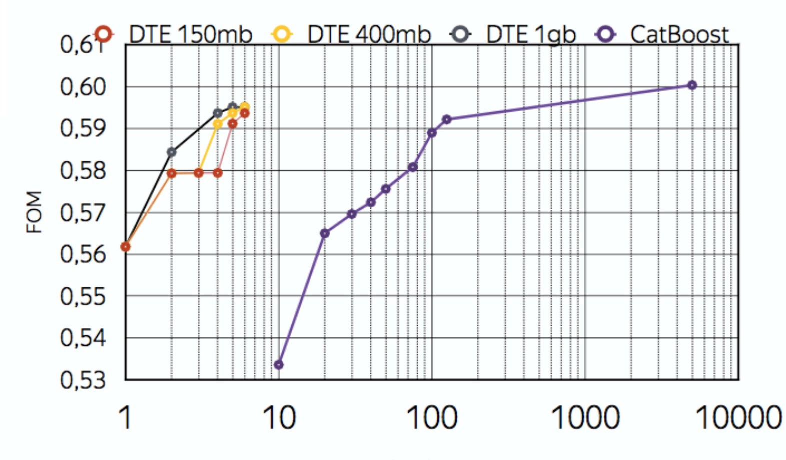
$S(D) \propto \prod |f_i| = \prod p_i$

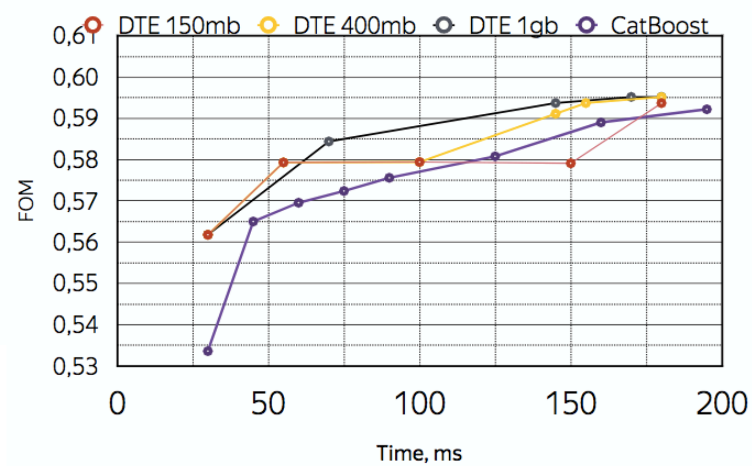


〉 CatBoost

〉 Uses oblivious trees

〉 Discretize features





Other Applications

Unsupervised learning (no labels)

- Anomaly detection, unexpected physics

Generative models

- Simulation and better training

Optimization and tuning

- Bayesian optimization etc.

IML

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

- Exchange between particle physics and machine learning communities
- Sharing of expertise among LHC experiments
- Software development and maintenance
- Forum and Education

Summary

All very exciting directions

- with many challenges to overcome

Opportunity to re-examine how we have done things until now

- from R&D to physics results

Challenges

- Intepretability, scalability and real-time inference