



UF UNIVERSITY of
FLORIDA

Data Science @ LHC Update

**Sergei V. Gleyzer,
Lorenzo Moneta**

**PH-SFT Group Meeting
Dec. 14, 2015**

Outline



- Data Science Workshop
- IML Working Group Update
- EPlanet Follow-up



Data Science @LHC Workshop

9 - 13 November 2015, CERN

DS@LHC

Local Organising Committee

- Xavier Cid (CERN)
- Gilles Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pierini (CERN)
- Jean-Roch Vlimant (Caltech)

Program Committee

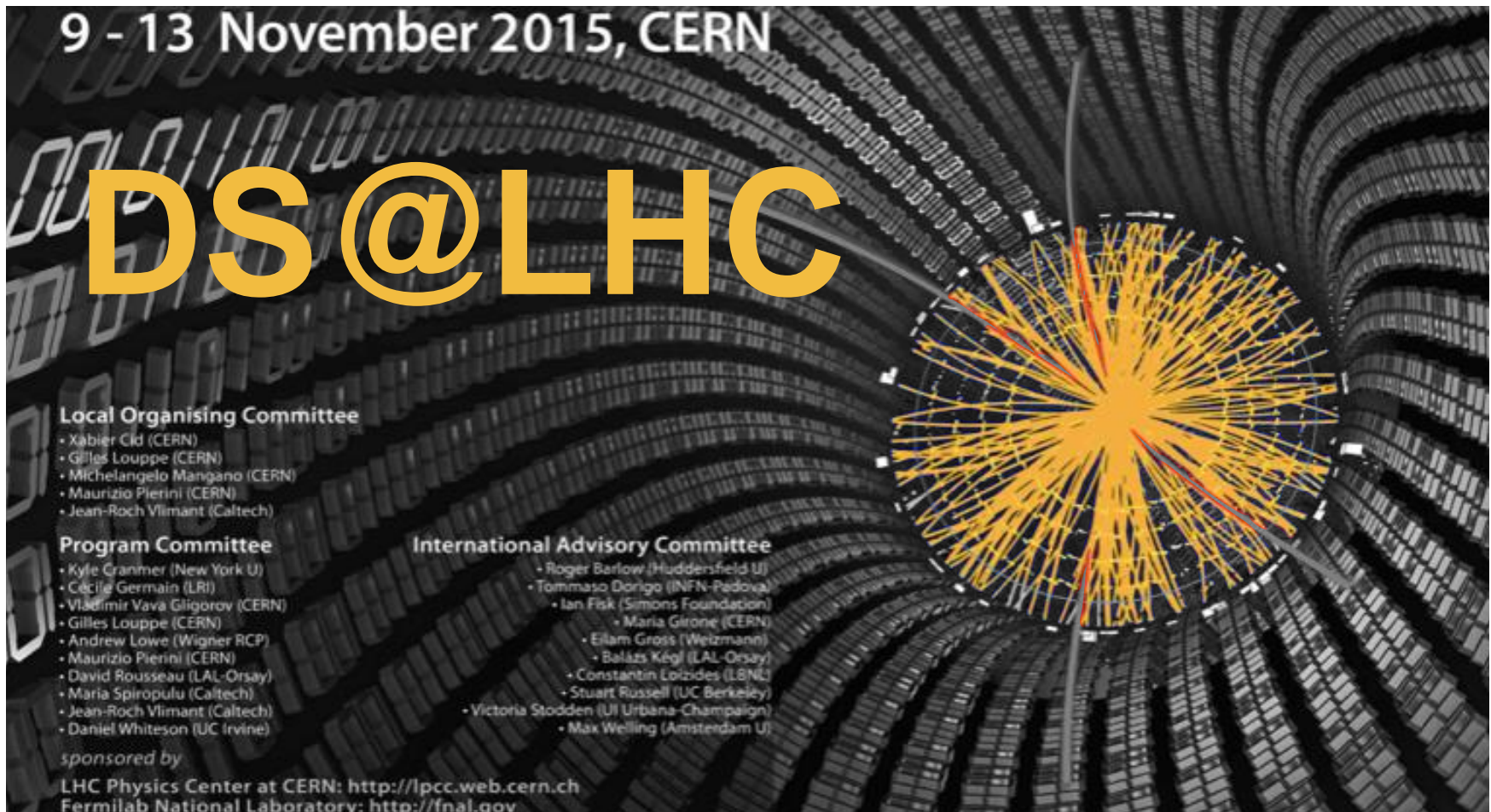
- Kyle Cranmer (New York U)
- Cécile Germain (LRI)
- Vladimir Vava Gligorov (CERN)
- Gilles Louppe (CERN)
- Andrew Lowe (Wigner RCP)
- Maurizio Pierini (CERN)
- David Rousseau (LAL-Orsay)
- Maria Spiropulu (Caltech)
- Jean-Roch Vlimant (Caltech)
- Daniel Whiteson (UC Irvine)

International Advisory Committee

- Roger Barlow (Huddersfield U)
- Tommaso Dorigo (INFN-Padova)
- Ian Fisk (Simons Foundation)
- Maria Gironè (CERN)
- Eilam Gross (Weizmann)
- Balázs Kégl (LAL-Orsay)
- Constantin Loizides (LBNL)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (UI Urbana-Champaign)
- Max Welling (Amsterdam U)

sponsored by

LHC Physics Center at CERN: <http://lpsc.web.cern.ch>
Fermilab National Laboratory: <http://fnal.gov>



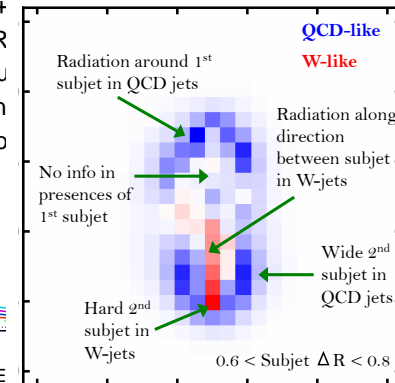
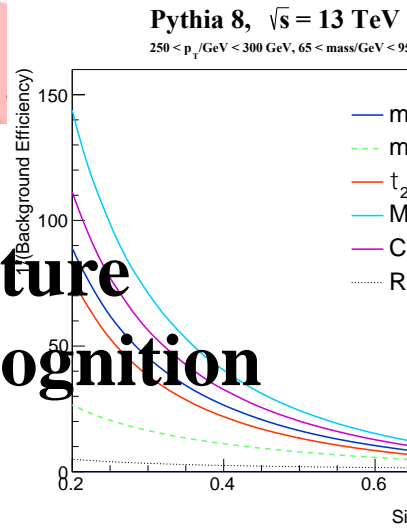
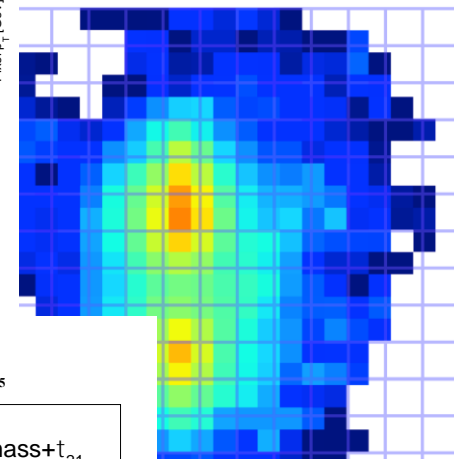
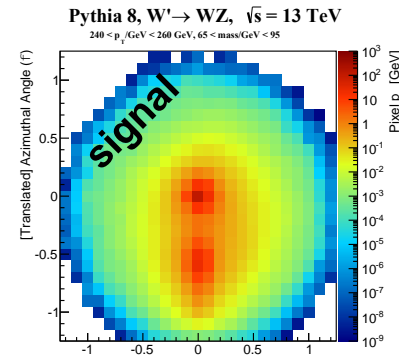
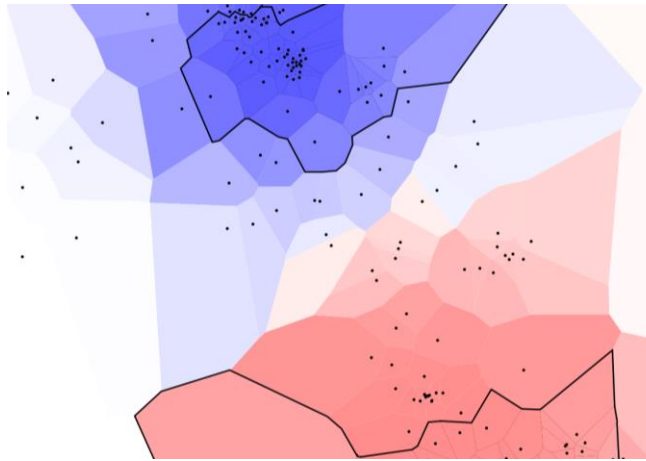


1st Data Science at LHC Workshop:
Nov 9-13 cern.ch/DataScienceLHC2015

- 200+ participants (registration closed at 200)
- Experts in machine learning and HEP
- Presentations from all LHC experiments
- Exciting developments in modern machine learning and their HEP applications
- Practical Tutorials (TMVA, Deep Learning)



We Were Actively
Involved

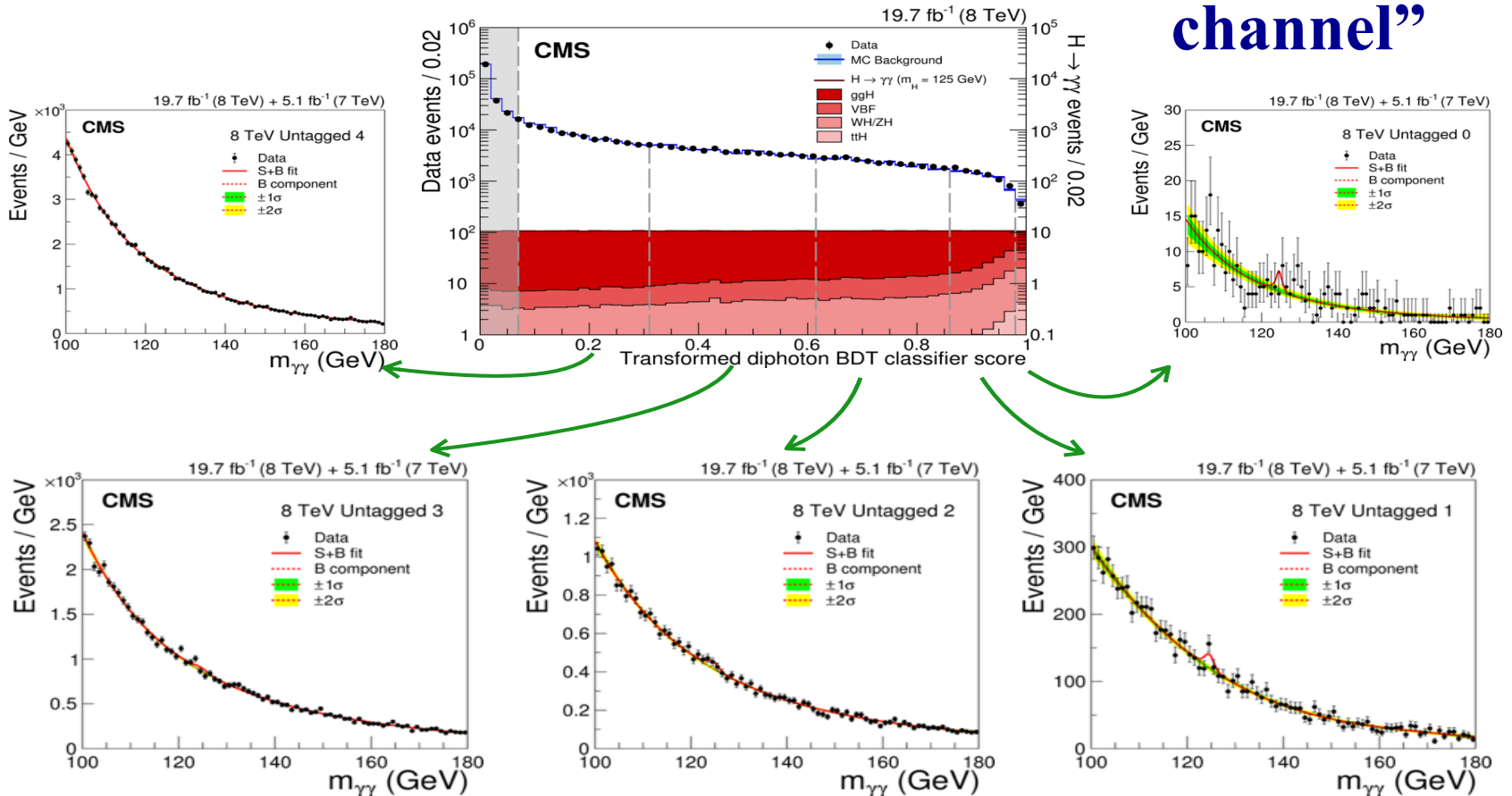


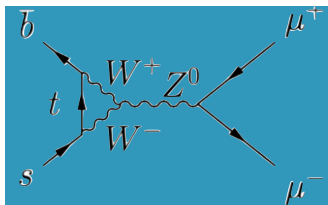
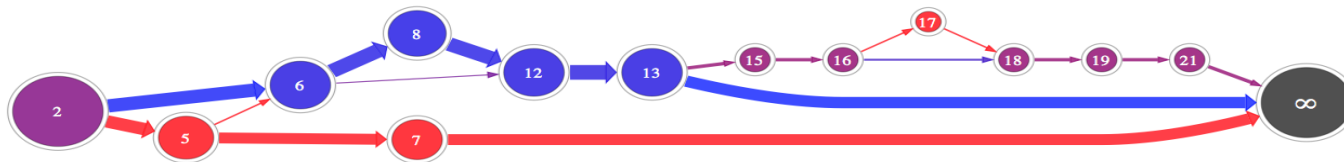
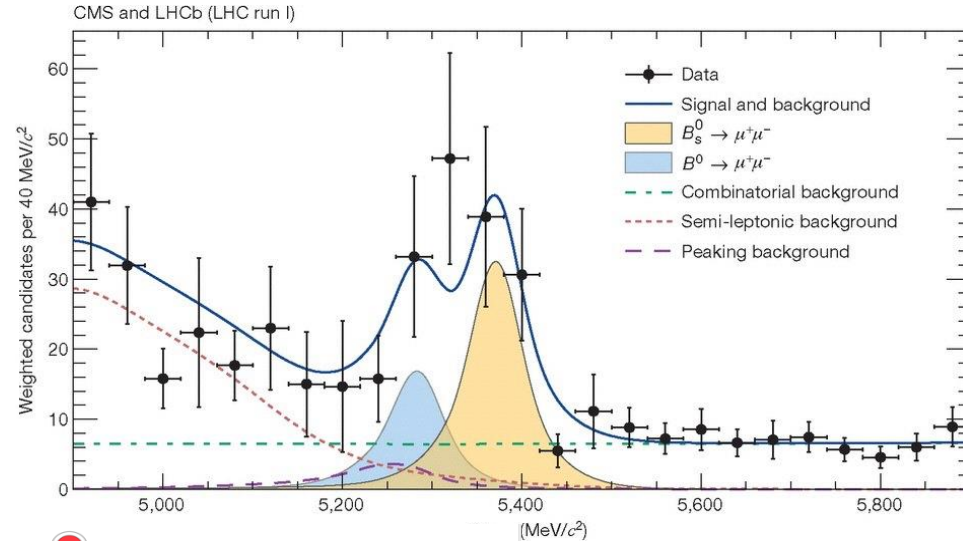
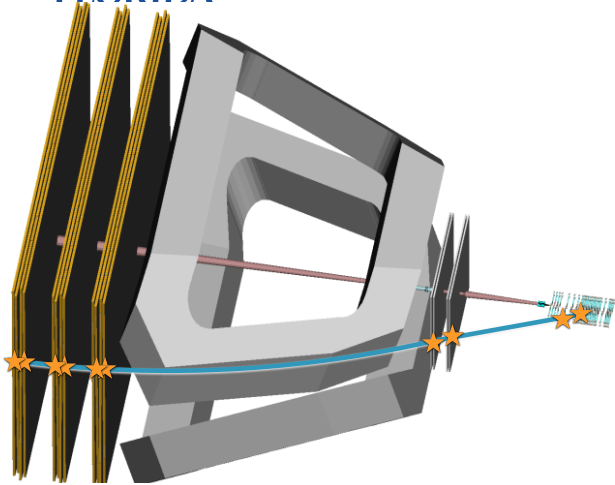
Fuzzy Jets
New ways to sub-structure
Advanced Pattern Recognition
Better Classification
Deep Learning

Mass fit

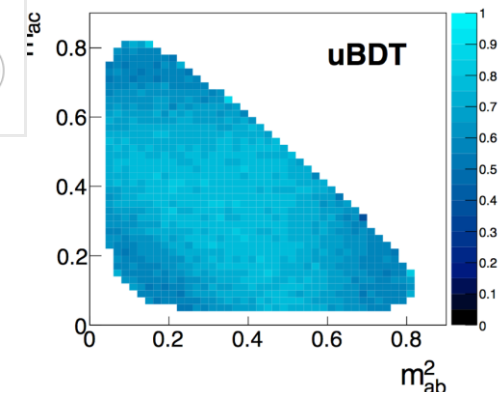
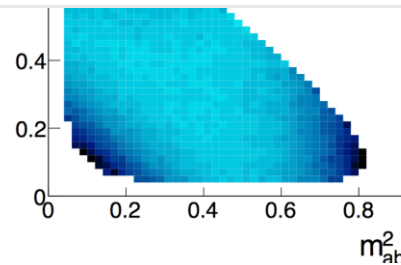
$H \rightarrow \gamma\gamma$

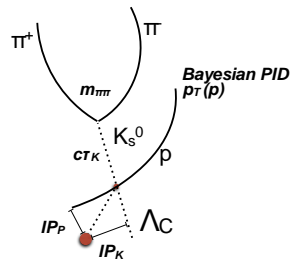
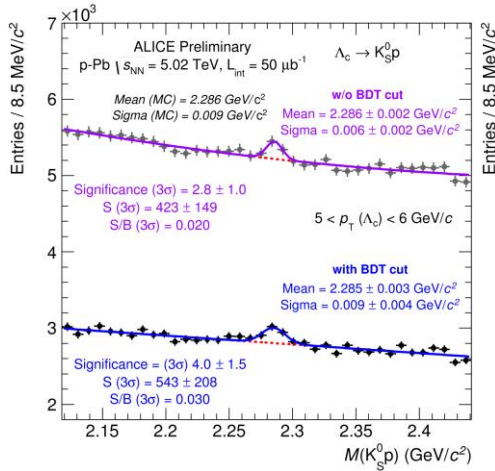
“Golden channel”



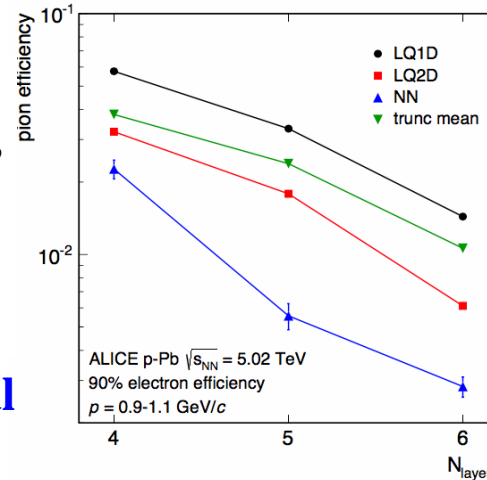


ML Trigger
PID
Event Selection
Dalitz Plots

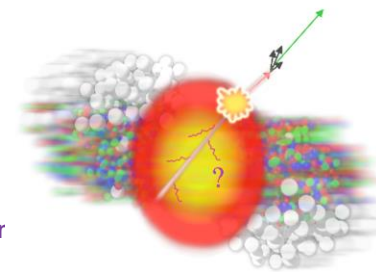
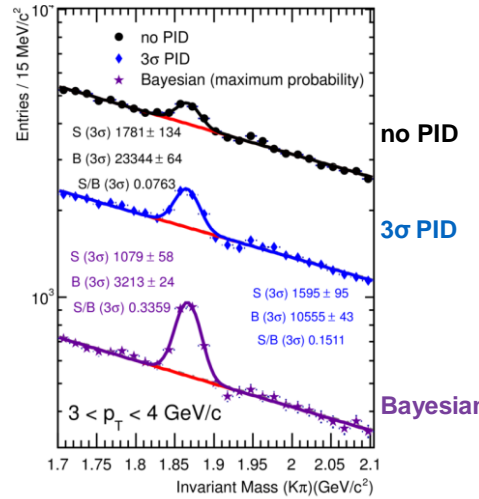
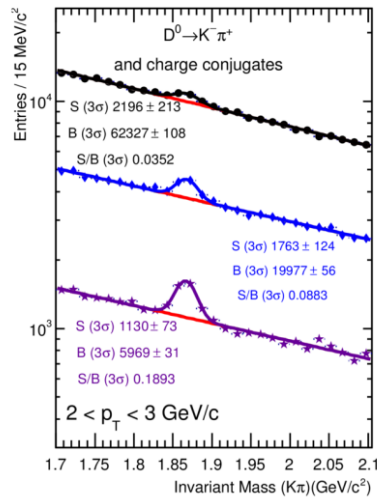
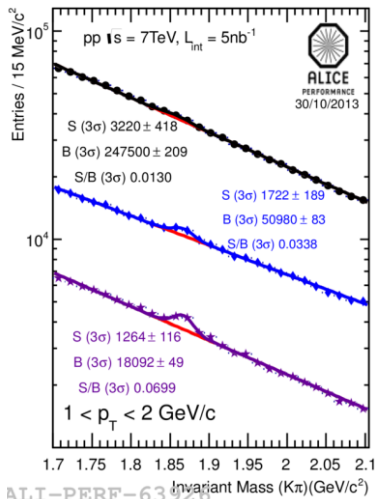
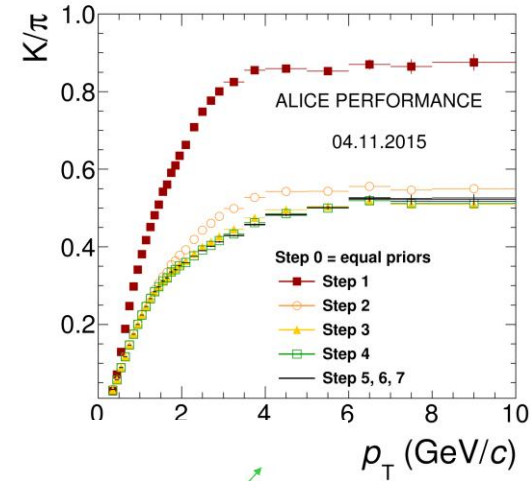




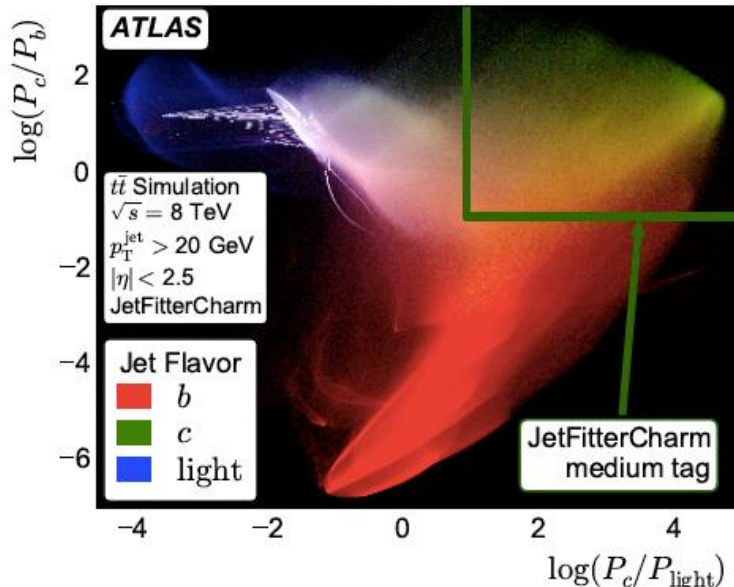
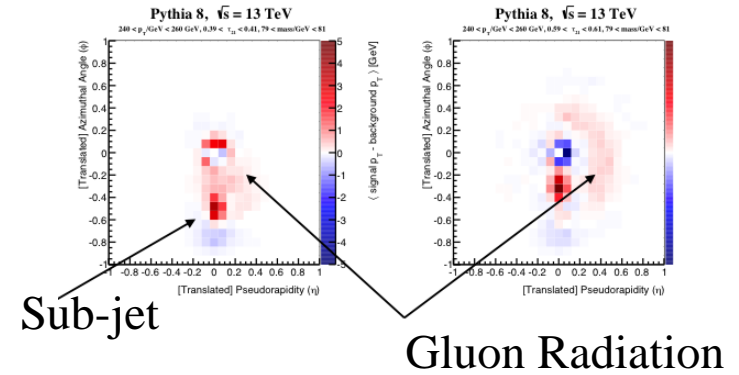
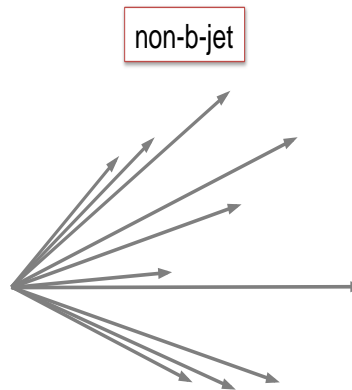
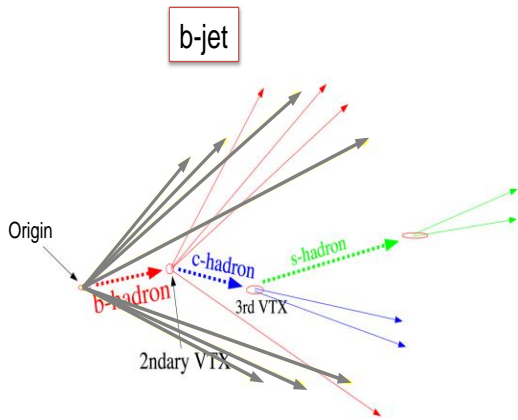
Higher Signal Significance



Better Particle ID

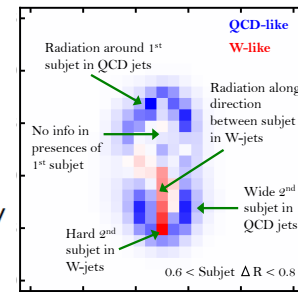


b and c jet identification

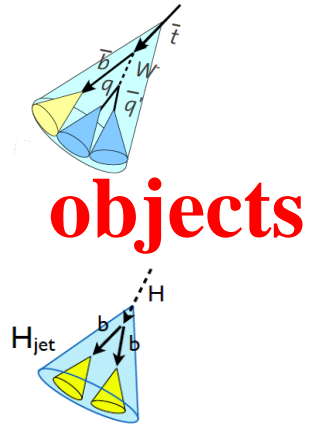


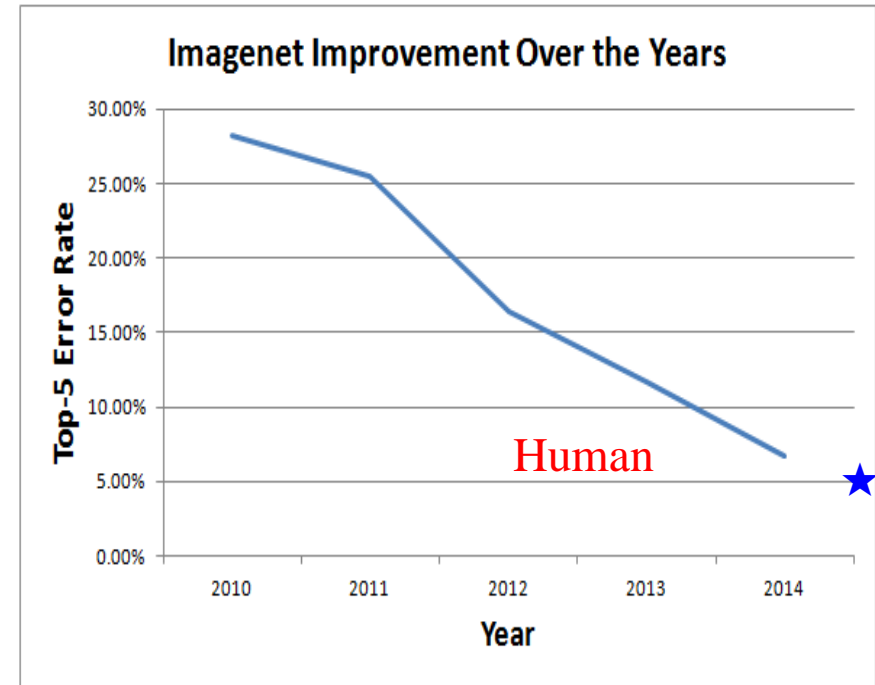
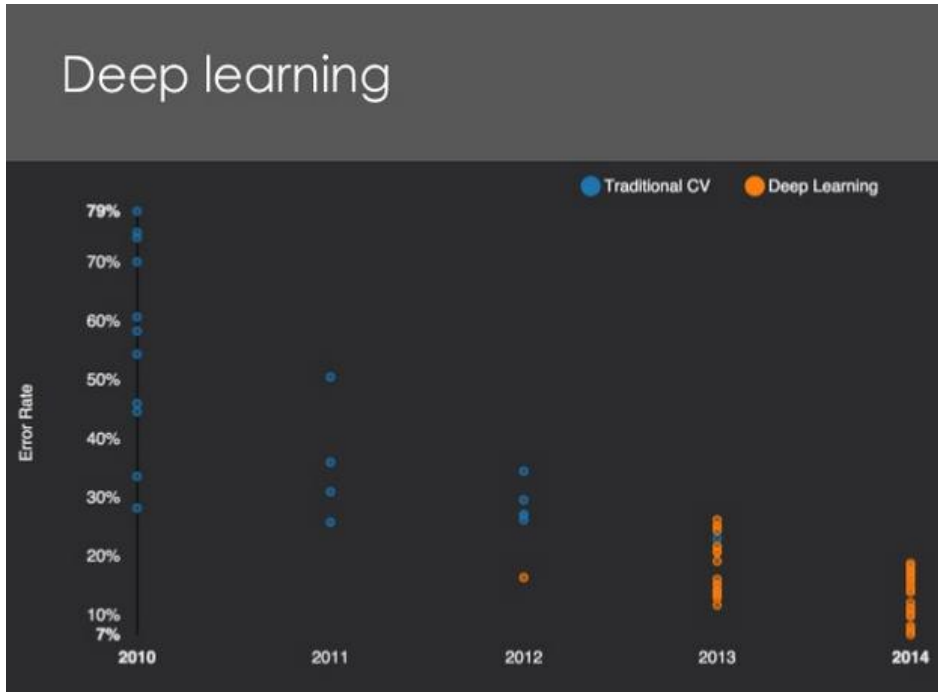
Gleyzer

Boosted



PH-SFT Meeting





Computer Vision (CV) Benchmarks

First super-human result in 2015*

* Google/Microsoft 4.9%

Deep Learning Neural Networks:

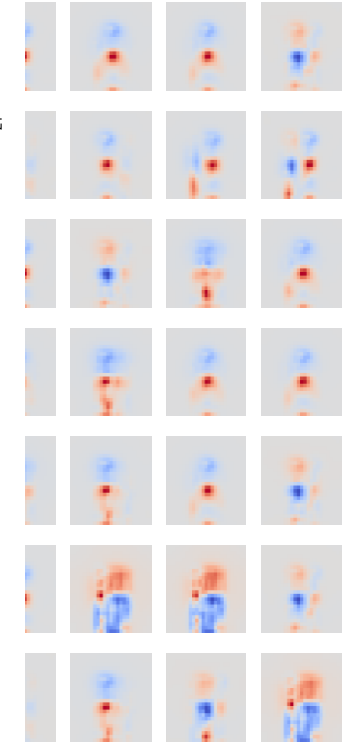
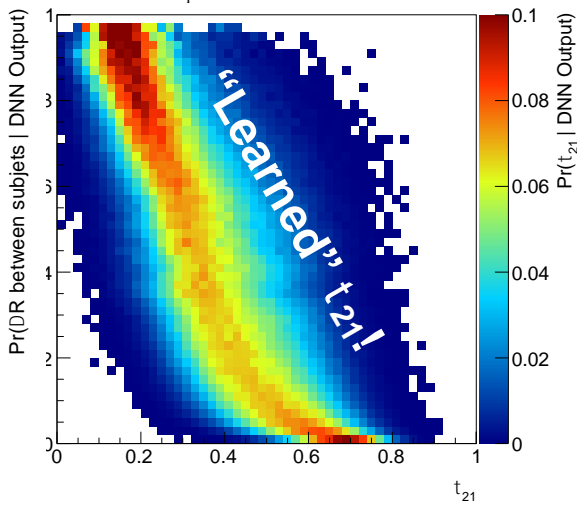
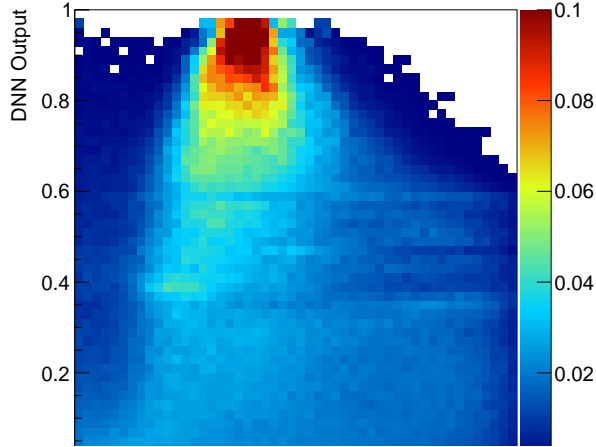
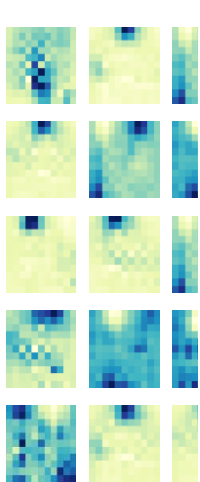
- Tremendous performance improvement
 - Training more complex models
 - **Increased** Depth
 - **Enlarged** Width
 - **Feedback**/Convolution
 - **Novel** activation functions
 - Effective strategies against over-fitting
 - Regularization

DS'15 Deep Learning



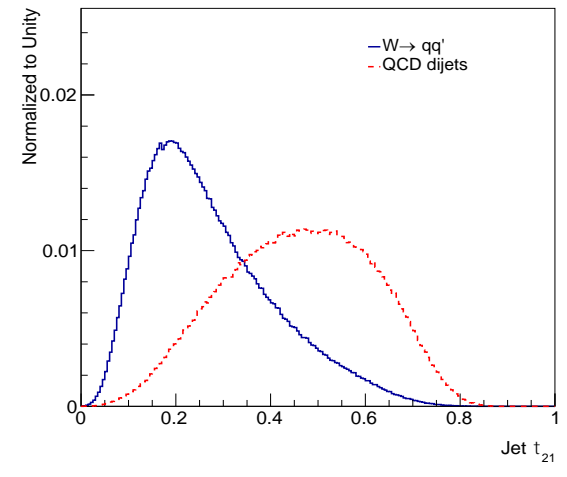
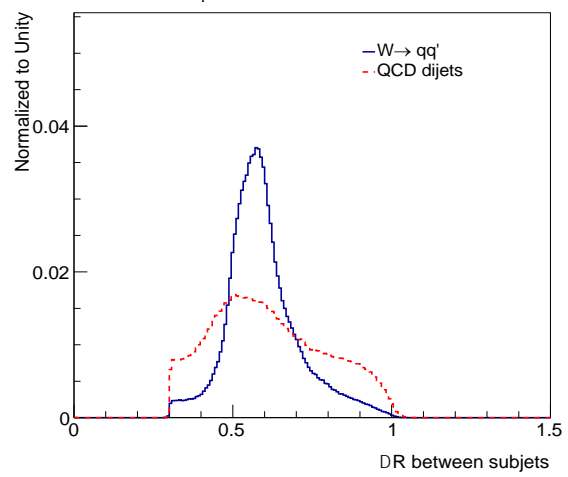
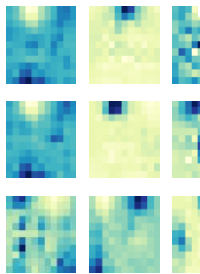
Pythia 8, QCD dijets, $\sqrt{s} = 13$ TeV
250 < p_T /GeV < 300 GeV, 65 < mass/GeV < 95

Pythia 8, QCD dijets, $\sqrt{s} = 13$ TeV
250 < p_T /GeV < 300 GeV, 65 < mass/GeV < 95



Pythia 8, $\sqrt{s} = 13$ TeV
250 < p_T /GeV < 300 GeV, 65 < mass/GeV < 95

250 < p_T /GeV < 300 GeV, 65 < mass/GeV < 95

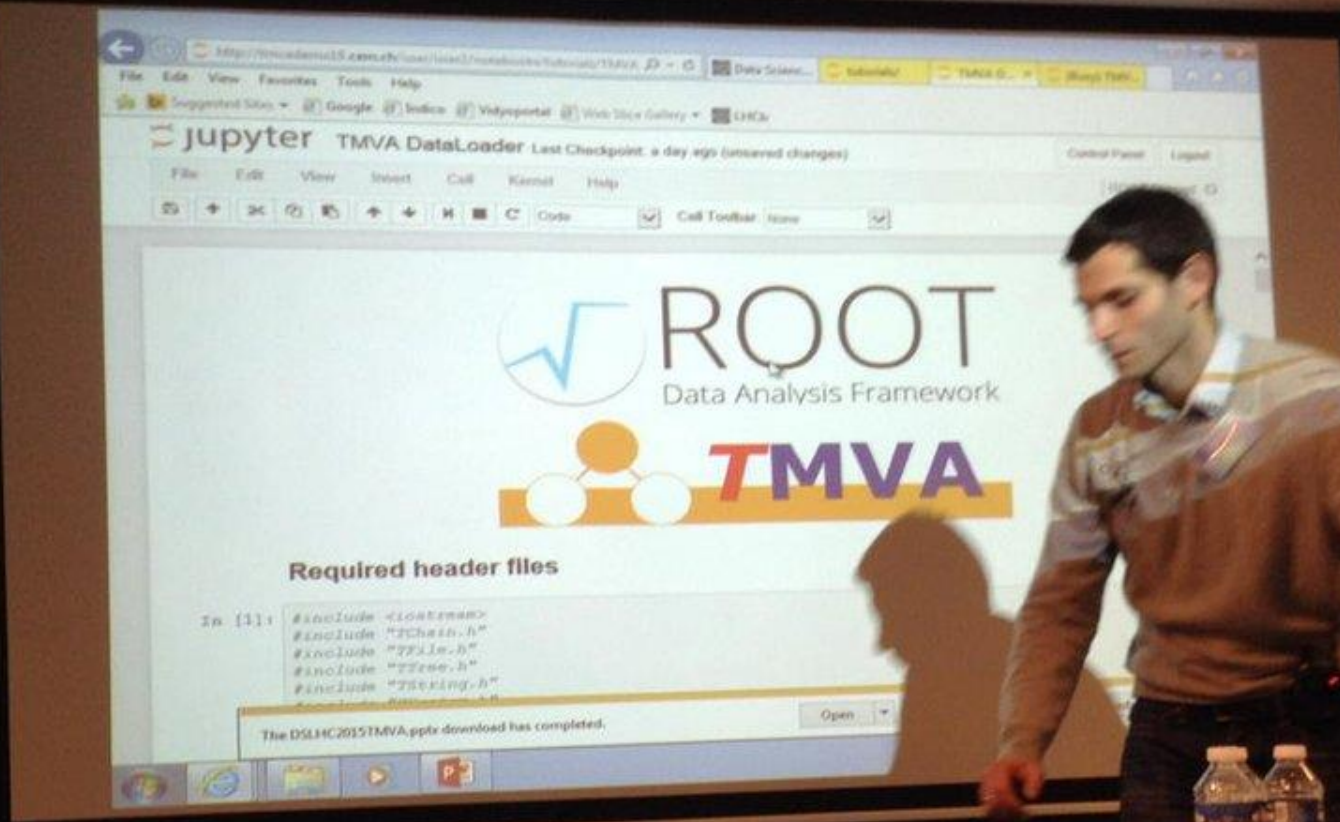


New **tutorials** at DS'15: **Deep Learning**,
TMVA new features, others

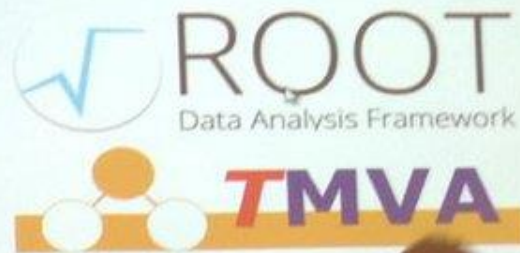
- Tutorials based on **RootBooks** were successful 😊
 - others less so
 - Users got tangled up in setup, dependencies, versions

RootBooks:

- dedicated servers for this tutorial were supplied by OpenStack (1 user/core)
- Very (!) positive feedback from participants



jupyter TMVA DataLoader Last Checkpoint: a day ago (saved changes)



Required header files

```
In [1]: #include <ROOT/RMC>
#include "TChain.h"
#include "TFile.h"
#include "TTree.h"
#include "TBranch.h"
```

The DSLHC2015TMVA.pptx download has completed. Open

From Twitter

- New **TMVA features** were premiered at DS'15
 - with **RootBooks**
 - Data Loader (modularity)
 - Feature Importance
 - Interfaces to R and Python
- Very positive response to new features at the workshop
 - We also expanded ML tutorials for EPlanet trip to Brazil and setup a dedicated server in UERJ
 - Second Rootbook server in Latin America

New **ideas/problems discussed/solved** at
DS'15

- **Deep Learning** applications in HEP
- Handling of **negative event** weights from MC
 - Not all algorithms are suitable (boosting)
- Use of **Matrix Element Methods** (MEM)s as input/combination to Deep learning
- Dark Knowledge
 - Use of simpler ML architectures that learn from trained complex ones
 - runtime, online, hardware, e.t.c.

LHC Machine Learning Working Group (IML) Update

IML: Inter-experimental LHC Machine-Learning working group

- **Exchange of HEP-ML expertise** and experience among LHC experiments
- **ML Forum for LHC-related** development and discussions
- **ML software development and maintenance**
- **Current and future ML R&D in HEP**
- **Exchange between HEP and ML communities**
- **Education** (Tutorials)

Following fruitful/positive discussion with management of the LHC experiments

- IML Working Group will become official Jan. 1, 2016
- Stay tuned for more details
- **Website:** <http://iml.cern.ch>
 - **Latest meeting** Dec. 4
<https://indico.cern.ch/event/463561/>
 - Next meeting in **January**
 - Forum/Mailing-list
 - <https://groups.cern.ch/group/lhc-machinelearning-wg/default.aspx>
 - Please join if you are interested in ML topics



EPlanet Trip Follow-up



EPlanet trip to Brazil (see Lorenzo's talk)

Mini-course in UERJ covered fundamentals of machine learning:

- Classification theory, practice, classic methods (decision trees, neural networks, regression), ensemble methods
- As well as advanced topics:
 - Feature selection, deep learning, multi-objective regression

15:20 - 15:35	<p>Machine Learning Challenges in HEP: HiggsML lessons 15' Speaker: David Rousseau (LAL-Orsay, FR)</p> <p>  tr151104_davidRou...</p>	
15:40 - 15:50	<p>Machine Learning Challenges in HEP: LHCb challenge lessons 10' Speaker: Andrey Ustyuzhanin (Yandex School of Data Analysis (RU))</p> <p>  Ustyuzhanin__FoP....</p>	
15:55 - 16:10	<p>Machine Learning Challenges in HEP: Ideas for new challenges 15' Speakers: Dr. Sergei Gleyzer (University of Florida (US)), Josh Bendavid (California Institute of Technology (US))</p> <p>  IML_Challenges_Se...  mlmemchalleng-De...</p>	
16:10 - 16:30	<p>Machine Learning Challenges in HEP: Upcoming Tracking Challenge 20' Speaker: Andreas Salzburger (CERN)</p> <p>  ML-Tracking-Chall...</p>	
16:35 - 16:45	<p>Machine Learning Challenges and Problem Statements in HEP 10' Speakers: Tim Head (Ecole Polytechnique Federale de Lausanne (CH)), Dr. Gilles Louppe (New York University (US))</p> <p>  tim-gilles-challeng...</p>	
16:50 - 17:05	<p>Introduction to BEAT: a new platform for challenges 15' Speakers: Dr. Sebastien Marcel (Idiap), Dr. André Anjos (Idiap), André dos Anjos (University of Wisconsin)</p> <p>  beat.pdf  BEAT Platform Sign...  BEAT Platform Web...</p>	
17:05 - 17:20	<p>ChaLearn for HEP competitions 15' Speaker: Isabelle Guyon (ClopiNet)</p> <p>  ChaLearn_LHC.pdf  ChaLearn_LHC.ppt</p>	
17:20 - 17:40	<p>Update on Cross-Validation in TMVA 20' Speakers: Thomas James Stevenson (University of London (GB)), Adrian Bevan (University of London (GB))</p> <p></p>	



Follow-up project (IML)

Using **deep learning** in connection with **matrix element method** analysis

- Yet unexplored and high-potential area identified at DS'15
 - Andre Snajder (UERJ faculty)
 - Sandro Fonseca (UERJ faculty) + students/researchers
 - Build upon existing **experience** in UERJ with MEMs and basic ML
 - We will use this project to evaluate deep learning NN in **TMVA** and other tools
 - Identify what needs to improve

- Very **successful** Data Science at LHC workshop in November
- **IML** LHC Machine Learning WG actively expanding adding formal structure
- Very **successful** EPlanet Mini-Course with a special emphasis on Machine Learning
- Started mutually beneficial **follow-up** projects in machine-learning in HEP between Brazilian researchers and us (via IML)