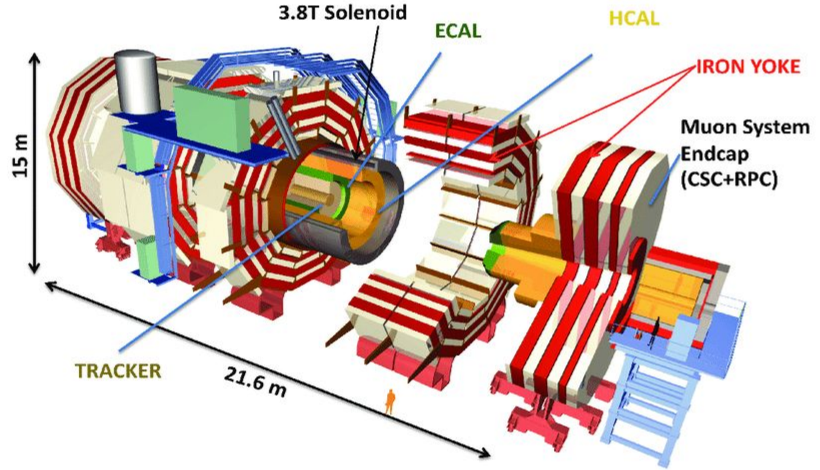
Training and validation of Machine Learning algorithms in the Endcap Muon Track-Finder

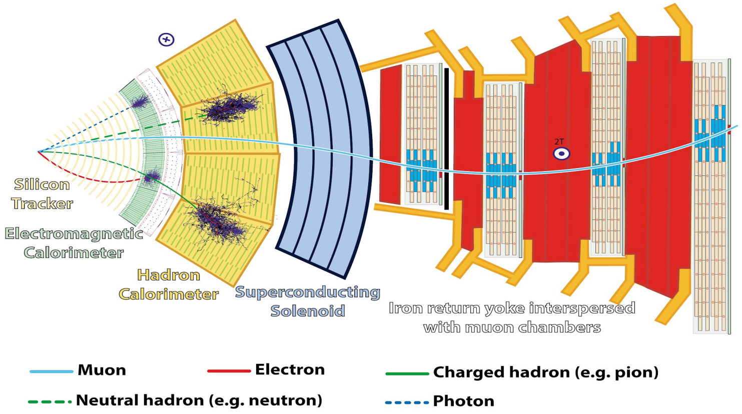
Mentee: Emanuela Riglioni (Xavier University of Louisiana)
Mentors: Dr. Darin Acosta (Rice University), John Rotter (Rice University)

Overview

- CMS Experiment
- Endcap Muon System
- Trigger and EMTF
- Machine Learning and BDTs
- Results

The CMS experiment





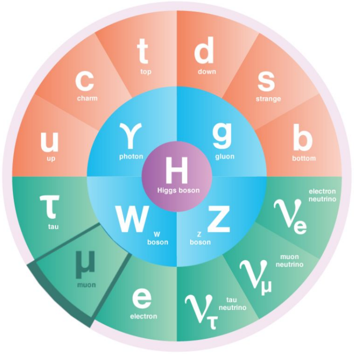
- The Compact Muon Solenoid (CMS) is a detector involved in the data collection and processing of particle collisions generated by the Large Hadron Collider (LHC) at CERN in Switzerland.

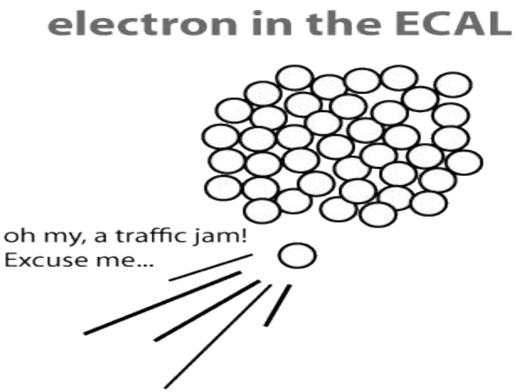
- CMS is composed of several layered sub-detectors each with the ability to interact with different particles.

- The rate of data collected by the detector is too large to be stored. A selection is made on data to keep and analyze.

Muons

- Fundamental particle classified as lepton
- Second generation
- Electric charge of -1 e
- Spin of ½
- 207 times heavier than the electron
- Very penetrating particle
- MIP (Minimally Ionizing Particle)

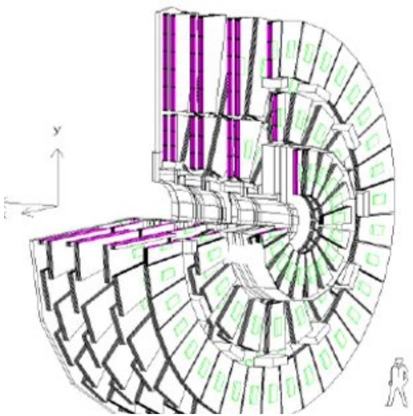


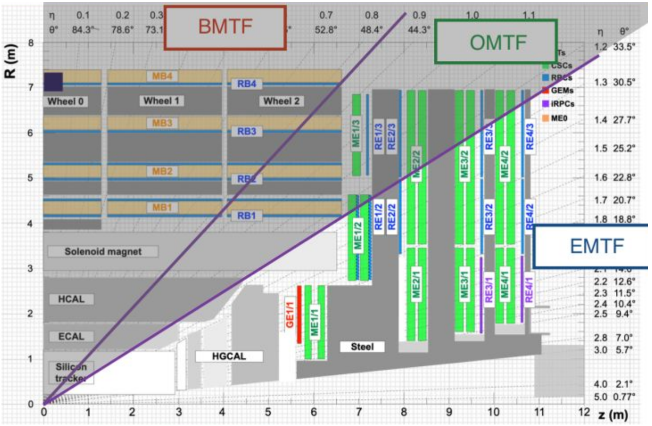


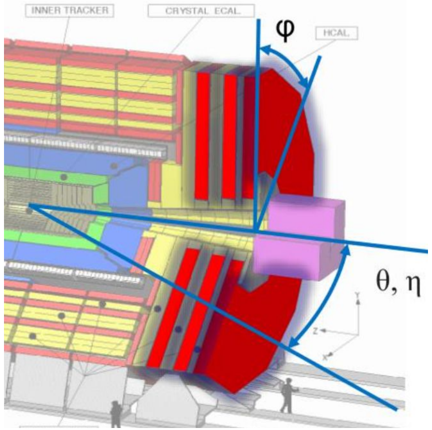
Endcap Muon System

- The Endcap Muon System is made of four different stations containing Cathode Strip Chambers (CSCs) and Resistive Plates Chambers (RPCs).

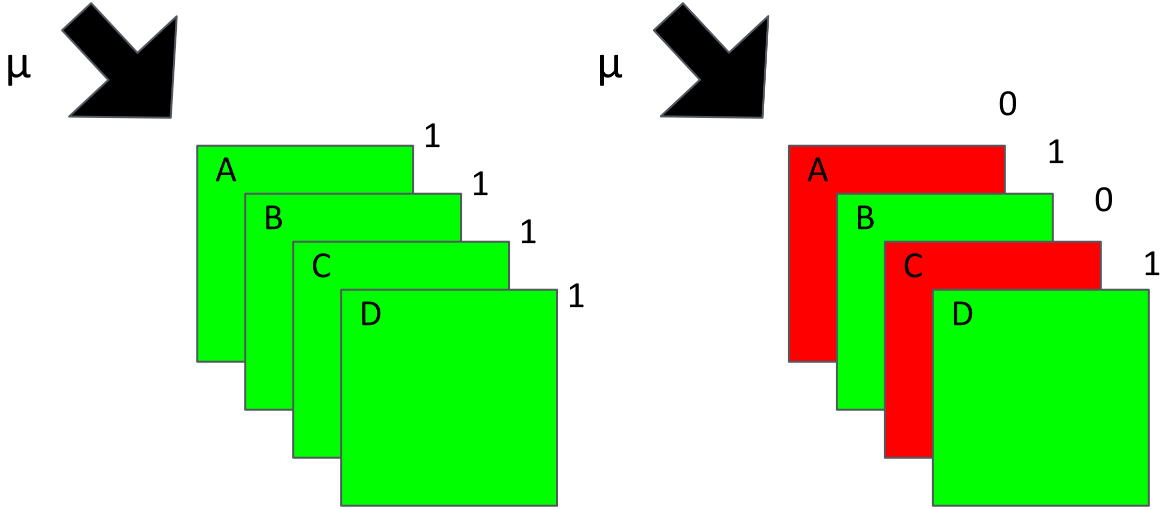
- The system records the coordinates in terms of η and ϕ of the detected particle. These represent the pseudo-rapidity and the azimuthal angle that is sensitive to the deflection of the muon in the magnetic field present.







Mode



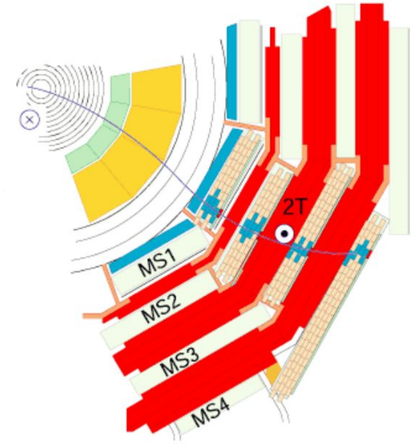
- Depending on which of the four stations is hit by the particles reaching the Endcap, a mode is encoded using a binary system.
  . hit in station B and D → 0 1 0 1 = 5
  . hit in stations A, B, C, D → 1 1 1 1 = 15

Track Reconstruction

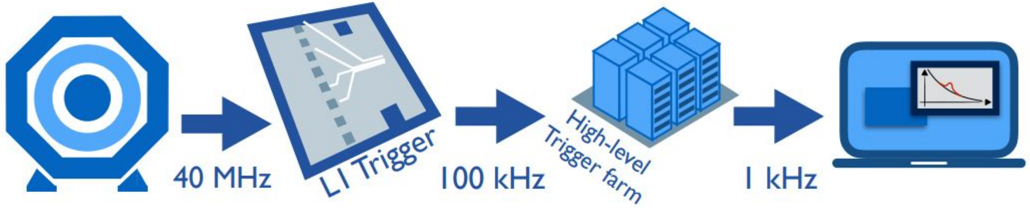
- Information from the mode and the trigger primitives (φ, η, etc.) is used for the reconstruction of the track made by the particle.

- The reconstructed track is used to make predictions on the transverse momentum (Pt) of the particle.

- The curvature of the muons is directly related to their momentum. The higher the momentum, the straighter the trajectory.



Trigger



- A trigger system is used to determine the interesting events to be kept for further analysis. The system includes the Level 1 (L1) and the Higher Level (HLT) Triggers.

- The L1 makes real-time decisions on the collisions at the LHC which produce 40 MHz of data.

- Only one out of 400 collisions is selected by the L1 and sent to the HLT for further investigation, with an output of 100 kHz.

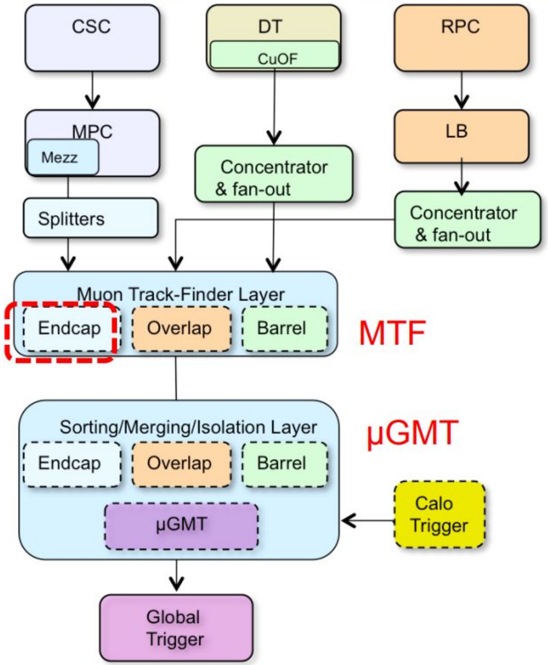
- The selection of muons is based off of specific thresholds in terms of Pt.

Endcap Muon Track Finder

- The Endcap Muon Track Finder (EMTF) is an algorithm that is part of the Level 1 trigger.

- It is used to identify muons and assign them with a momentum based on their deflection in the non-uniform magnetic field of the Endcap.

- Makes use of field-programmable gate arrays (FPGAs) for quicker calculations.

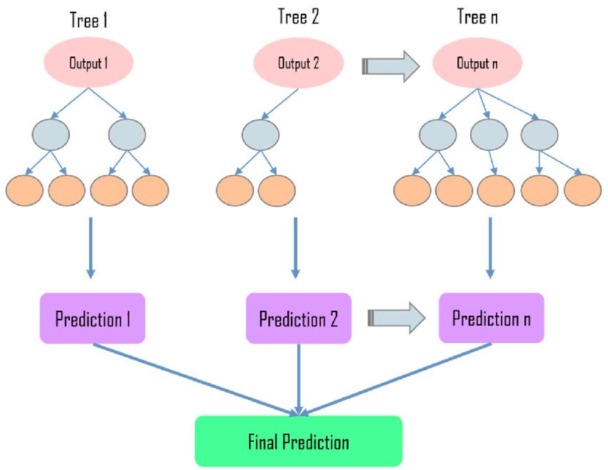


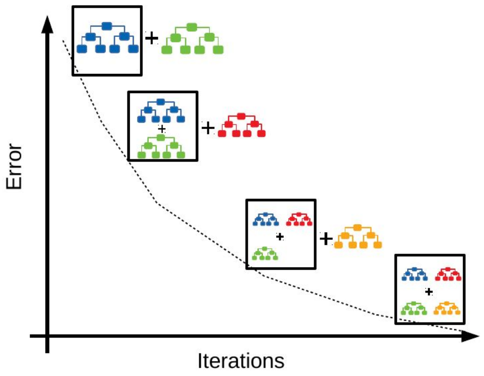
Boosted Decision Trees

- Gradient Boosted Decision Trees Regression is a machine learning technique that is implemented in supervised learning models.

- Individual decision trees, considered to be weak learners, are connected in series to make a strong learner.

- The weak learners are fit so that each new learner fits into the residuals of the previous step with the goal to minimize the chosen loss function.





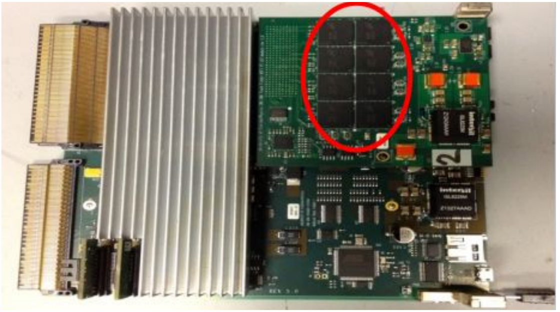
Boosted Decision Trees for EMTF

- Used to regress the momentum of muons from the trigger primitives obtained by the EMTF.

- Initially trained using Monte Carlo simulations using the data format of Run2.

- The variables available for analysis are bit compressed based on their importance for the BDT's prediction, for a total of only 30 bits of information.

Mode		Δφ						Δφ	Δθ						Bend + RPC				F/R				θ	Md	Bits
		1-2	1-3	1-4	2-3	2-4	3-4	sign	1-2	1-3	1-4	2-3	2-4	3-4	1	2	3	4	1	2	3	4			
15	1-2-3-4	7			5		4	2			2				2	1	1	1	1				3	1	30
14	1-2-3	7			5			1		3					2	1	1		1	1			5	3	30
13	1-2-4	7				5		1			3				2	1		1	1	1			5	3	30
11	1-3-4		7				5	1			3				2		1	1	1		1		5	3	30
7	2-3-4				7		5	1					3			2	1	1		1			5	4	30
12	1-2	7							3						3	3			1	1			5	7	30
10	1-3		7							3					3		3		1		1		5	7	30
9	1-4			7							3				3			3	1			1	5	7	30
6	2-3				7							3				3	3			1	1		5	7	30
5	2-4					7							3			3		3		1		1	5	7	30
3	3-4						7							3			3	3			1	1	5	7	30



BDTs Performance

- The performance of the BDTs is evaluated based on their efficiency and resolution.

- Efficiency describes the ability of the algorithm to filter data accurately. Namely, it is the percentage of events that pass for a given true Pt value.

- Resolution is the spread of the prediction about the true value.

- A good performance entails high efficiency above the true Pt threshold, and low efficiency below.

From C++ to Python

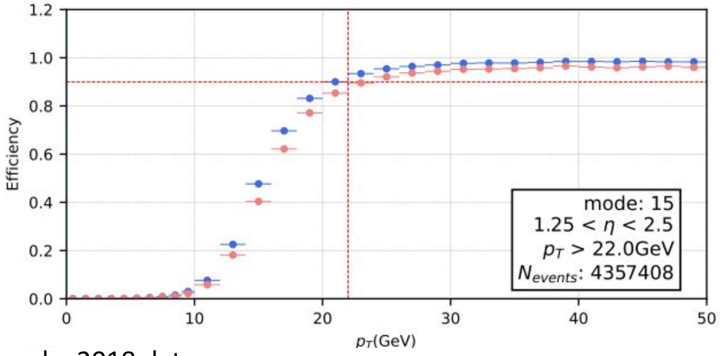
- Initial BDT code was written in C++:
  . slow training
  . outdated module

- Current BDT code written in Python:
  . uses Run2 data format
  . faster
  . validated performance compared to C++ results

- Future BDT code written in Python:
  . altered to utilize improved data format from Run3
  . performance expected to be retained (or improved in the inner ring area)

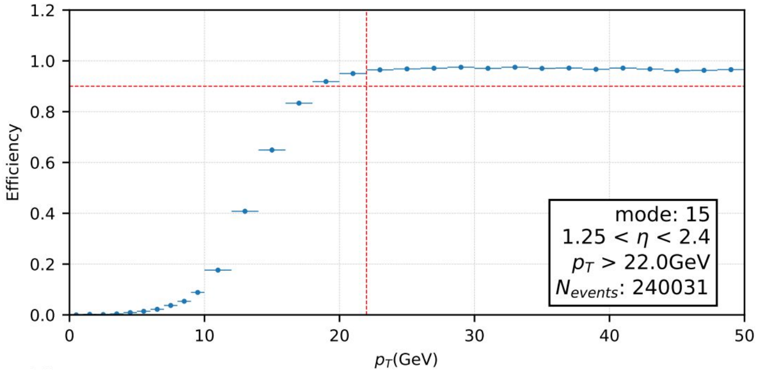
Efficiency vs. Pt

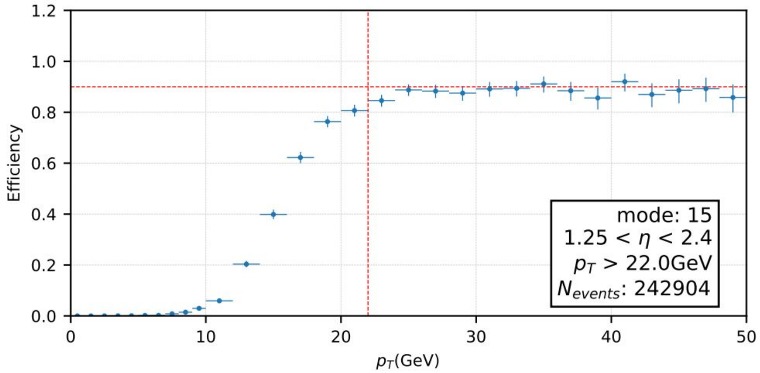
C++



red = 2018 data
blue = 2021 data

Python Run2 data format

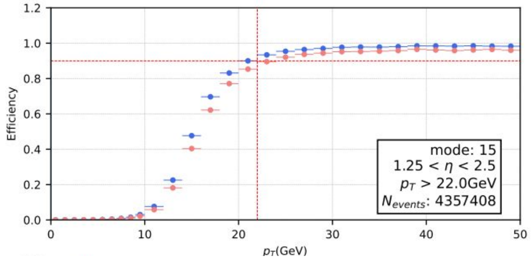




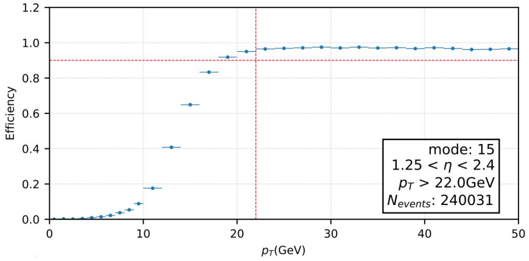
Python Run3 data format

Efficiency Overview

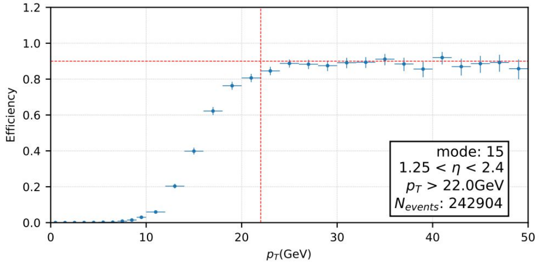
C++



Python Run2 Data Format

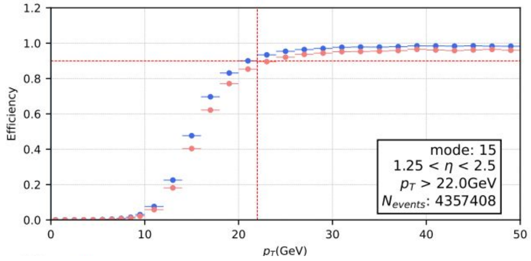


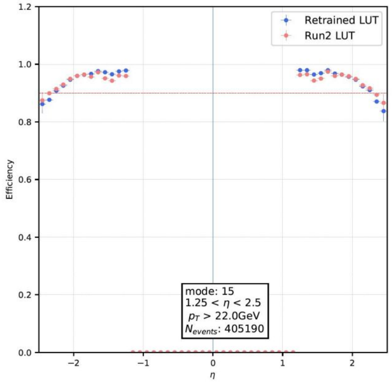
Python Run3 Data Format



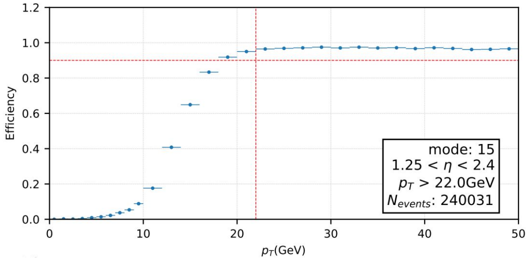
Efficiency Overview

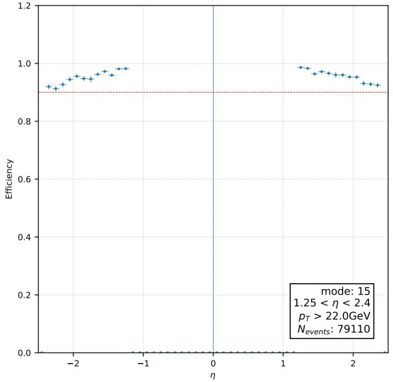
C++



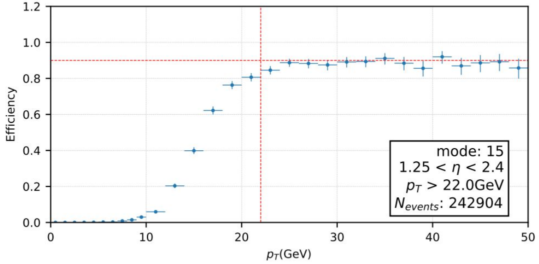


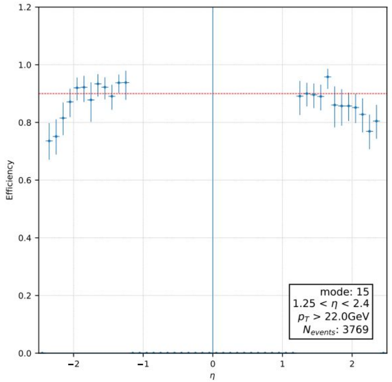
Python Run2 Data Format



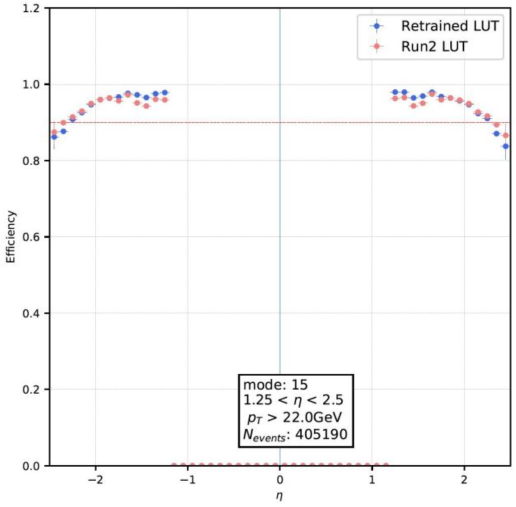


Python Run3 Data Format

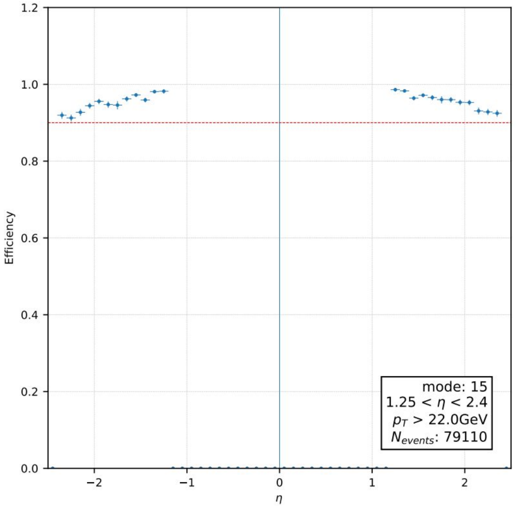




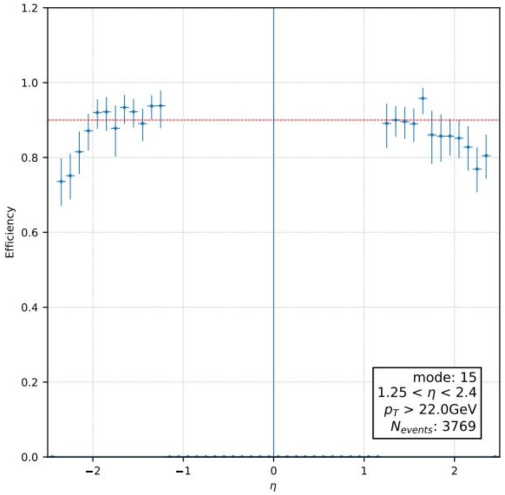
Efficiency vs. Eta



C++

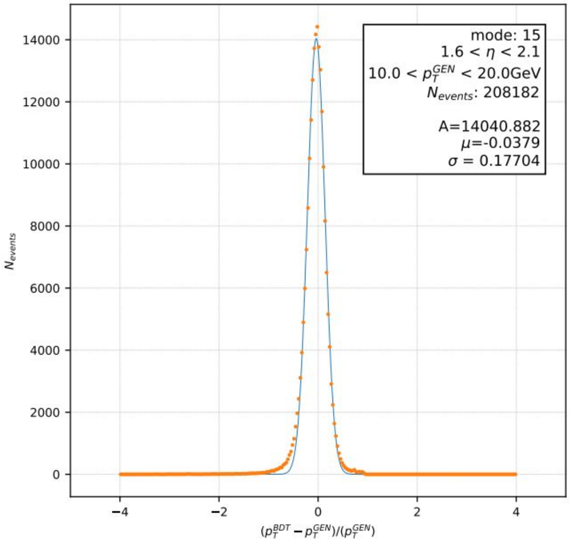


Python Run2 data format

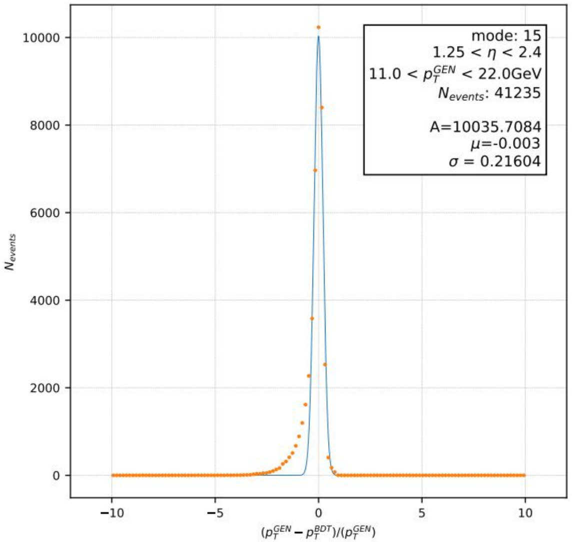


Python Run3 data format

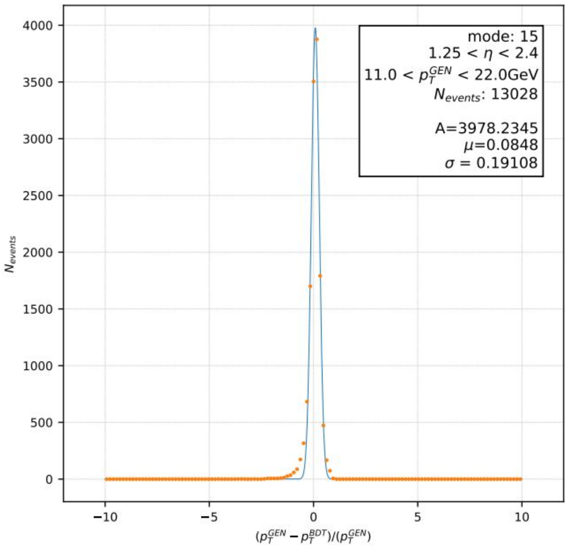
Resolution



C++



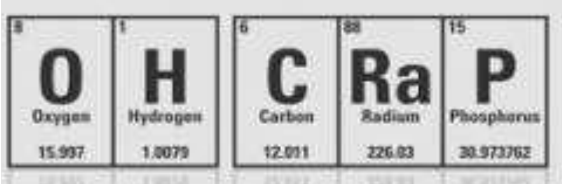
Python Run2 data format



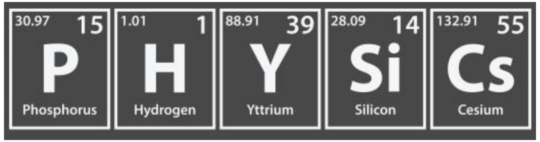
Python Run3 data format

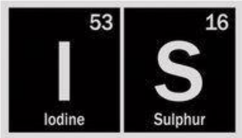
Conclusions

First day at Fermilab:



After 10 weeks of particles Physics:





Acknowledgments

- Dr. Acosta and John for the invaluable professional and personal support
- Particle Owls Group @ Rice, especially Mason, Mark, and Matthew for fostering a stimulating and lively work environment
- Fellow US CMS PURSUE interns Elena, Felipe, Radish and Michael for everything
- Roomies Rachael, Ellie, and Sarah for giving me a home away from home