Mentor Perspective

USCMS PURSUE summer program (2022)

Darin Acosta (Rice University)

Mentee and Mentors

- Summer Student: Andrew Fonseca from the University of Delaware
- Faculty mentor: Darin Acosta, Rice University Physics & Astronomy Dept.
    - I was also the Trigger Coordinator of the CMS experiment during this time
- Graduate student technical mentor joining me : John Rotter

Mentee Research Project

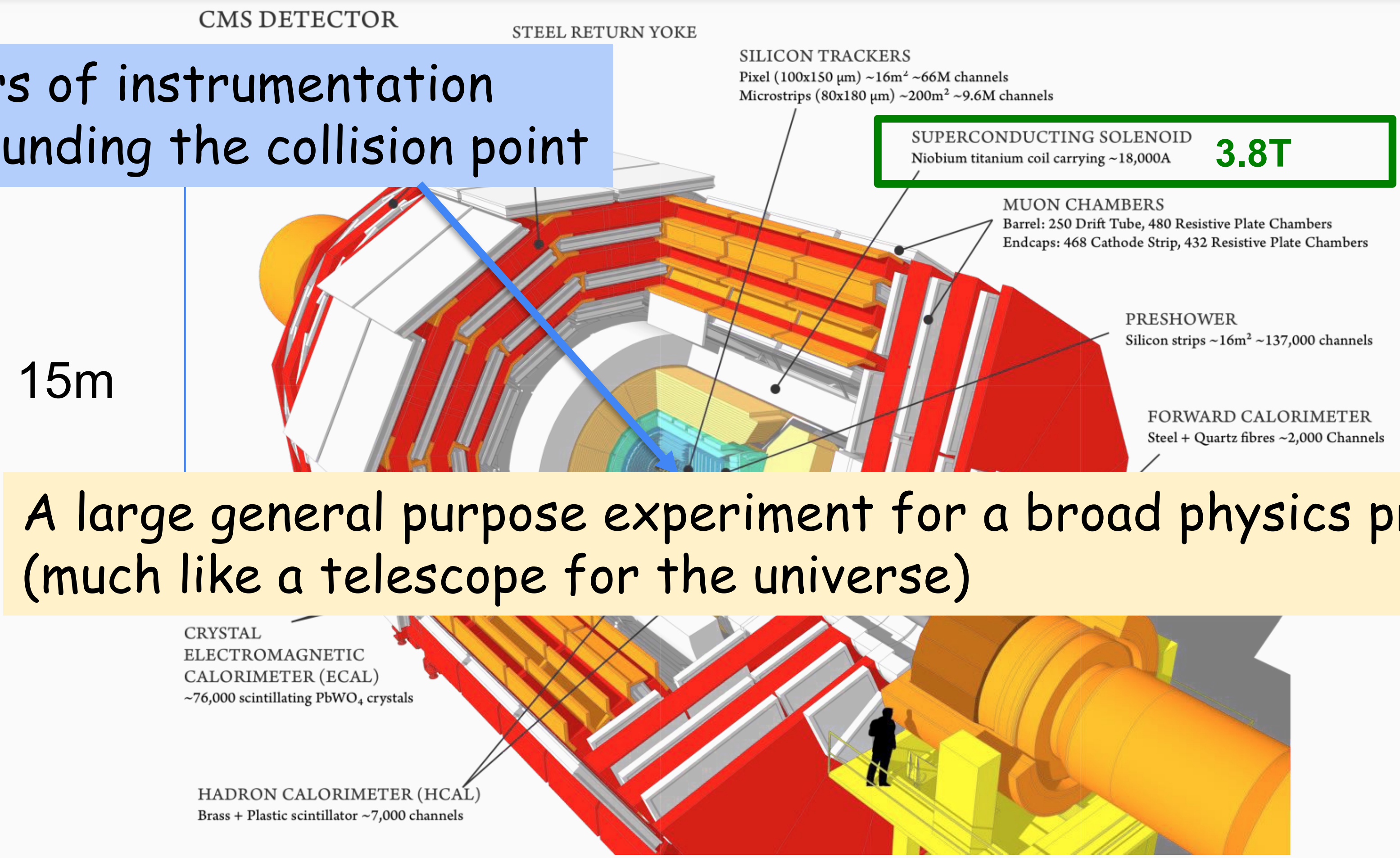
- Title: Trigger systems and filtering muon events
- Become acquainted with a “trigger system” at a collider experiment, and the muon trigger of the CMS experiment in particular
- AI/Machine Learning component: Training of a Boosted Decision Tree:
    - Used to assign momentum of muons from measurements
    - Involves writing/running computer programs to train BDT parameters
    - Recalibrate output to what trigger system needs

But let me try to explain a little bit about what this all means…

- Assessment:
    - The result of these simulations and trainings can be shared, compared, and contrasted through calculations and plots of:
    - Efficiency - How well are you filtering out uninteresting events
    - Rate - How much data are you letting through
    - Resolution - The quality of the events you are letting through

The "Compact" Muon Solenoid (CMS)

Layers of instrumentation surrounding the collision point



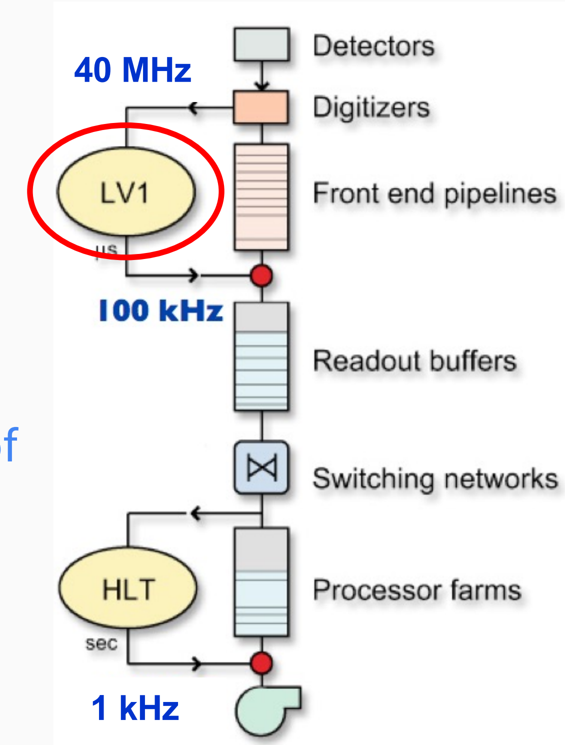
A large general purpose experiment for a broad physics program (much like a telescope for the universe)

Data Acquisition and Trigger System

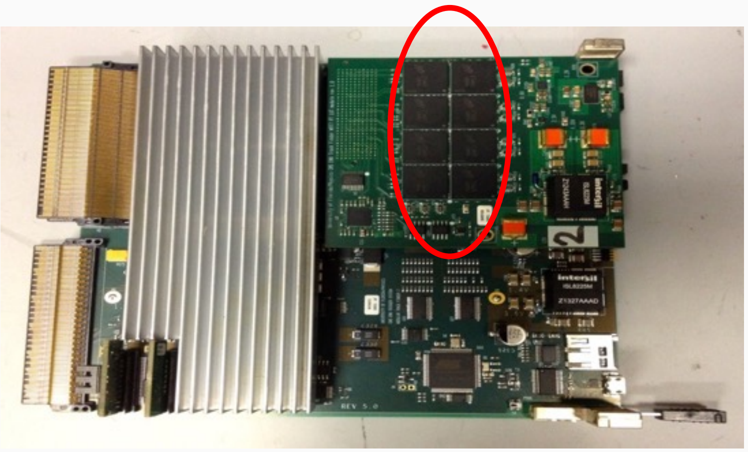
- Not shown in the detector image are the electronics and computers to selectively read out their data
- Why is a real-time data selection system ("trigger") needed?
  - Too much data to continuously stream to disk for storage and/or computer processing reasons
    - The LHC collides proton bunches at 40 MHz
    - An LHC experiment could generate ~100 terabytes per second!
- It is really the *first step of a data analysis*
  - An online data filtering system: select the collisions you want most
  - But irreversible – you can't go back to data you did not record
    - e.g. Throw away 99.998% of all LHC crossings…
      - Keep only ~1000 Hz out of 40 MHz beam crossing rate
    - But, don't throw out the baby (Higgs) with the bath water!

CMS Trigger Architecture

- Two levels:
  - Level-1: custom electronics to reduce the data from a collision rate of 40 MHz to no more than 100 kHz for the detector readout electronics, with only a 4 µs latency (buffer depth)
  - High Level Trigger (HLT): event filter farm comprised of commercial CPUs (and now GPUs also) running software to further reduce event rate to storage to an average of ~1 kHz (for LHC Run 2)
    - Order of 26k CPU cores
-



The Endcap Muon Electronic Trigger System



We designed the muon trigger for the CMS endcaps.
Includes AI/ML algorithms to measure muon momenta.
Working on next-gen algorithms and electronics.

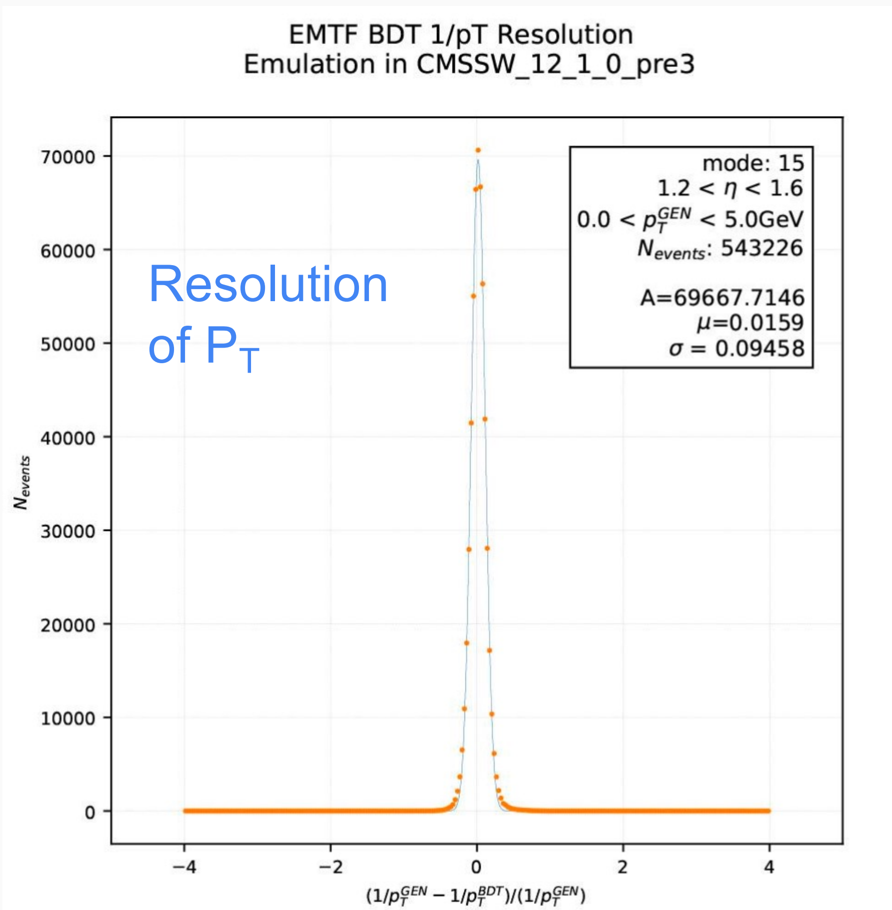
Machine Learning

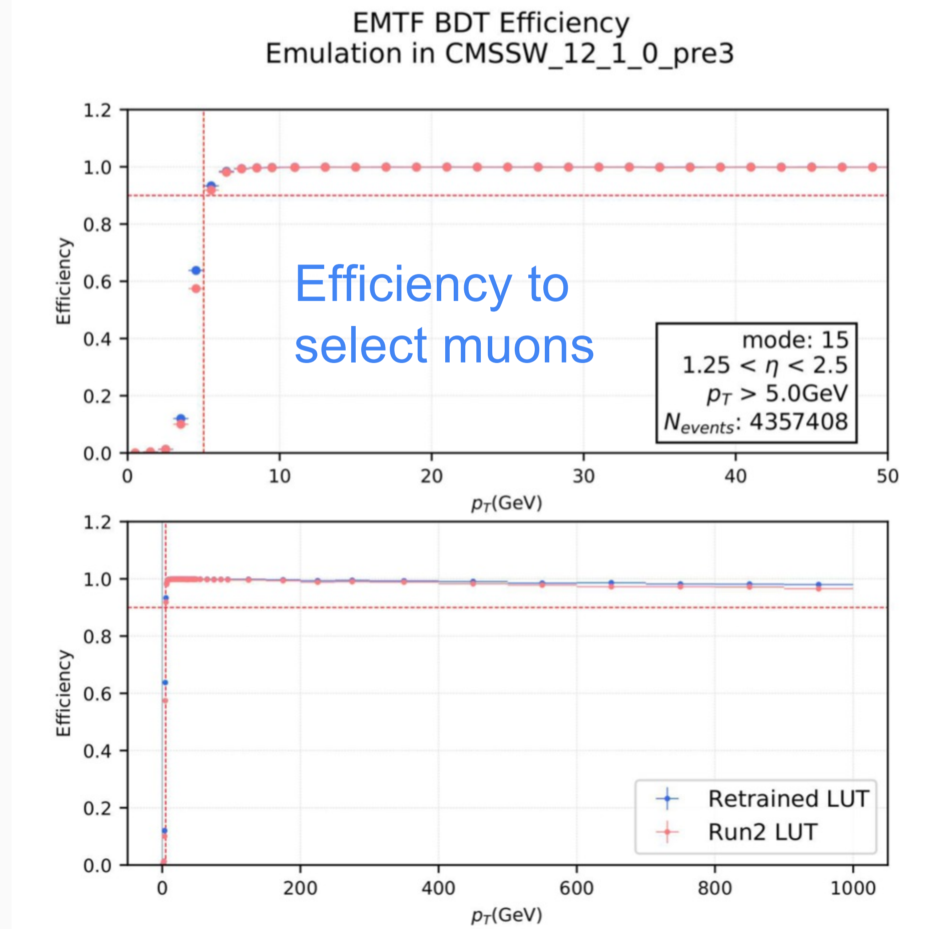
- Boosted Decision Trees are a form of machine learning, like artificial neural networks
  - A decision tree repeatedly splits a dataset into smaller subregions based on features in that dataset
- The algorithm must be trained with a large sample of examples of desired classification
  - Cat vs. not a cat; Higgs boson vs. not a Higgs boson; momentum =10 vs. momentum =100
  - Weights are determined from back propagation and using a specific loss function (penalty)
- The application of trained network is known as inference
  - This is what is stored in the memory of our electronics

Tools used

- Machine learning training used the ‘TMVA’ package in high-energy physics community
- CERN ‘Root’ package also used
- Andrew typically wrote python code, and used matplotlib package for plotting
- Made use of the computing resources offered by the LPC cluster at Fermilab

Some Results from the Research





Overall Experience with Program

- The USCMS workshops and tutorials were very helpful, so it was a smooth process for us to get started on the research project
- Once we got going, met virtually about 3 times per week for extended in-depth discussions about how to conduct the studies (coding, computing, big picture, little picture, etc.)
- Andrew found the machine learning part very interesting and wanted to learn the details
- Grad student mentor also found the mentoring experience very positive
- And I was quite pleased as well! ☺ Useful tools developed!
- In the end we completed the training, the results were loaded into the CMS electronics, and CMS took data using it this year!