Using Unsupervised Machine Learning to Rediscover Standard Model Physics at CMS

Elliott Kauffman
Princeton University
ek8842@princeton.edu

18 December 2024
US LHC Users Association: Lightning Round Talks

Where is new physics?

- **No clear signs of new fundamental particles** since the discovery of the Higgs boson
- CMS has many search attempts, but what if new physics is something we have not thought of?
- Need to explore all possible ways to **maximize LHC discovery potential**

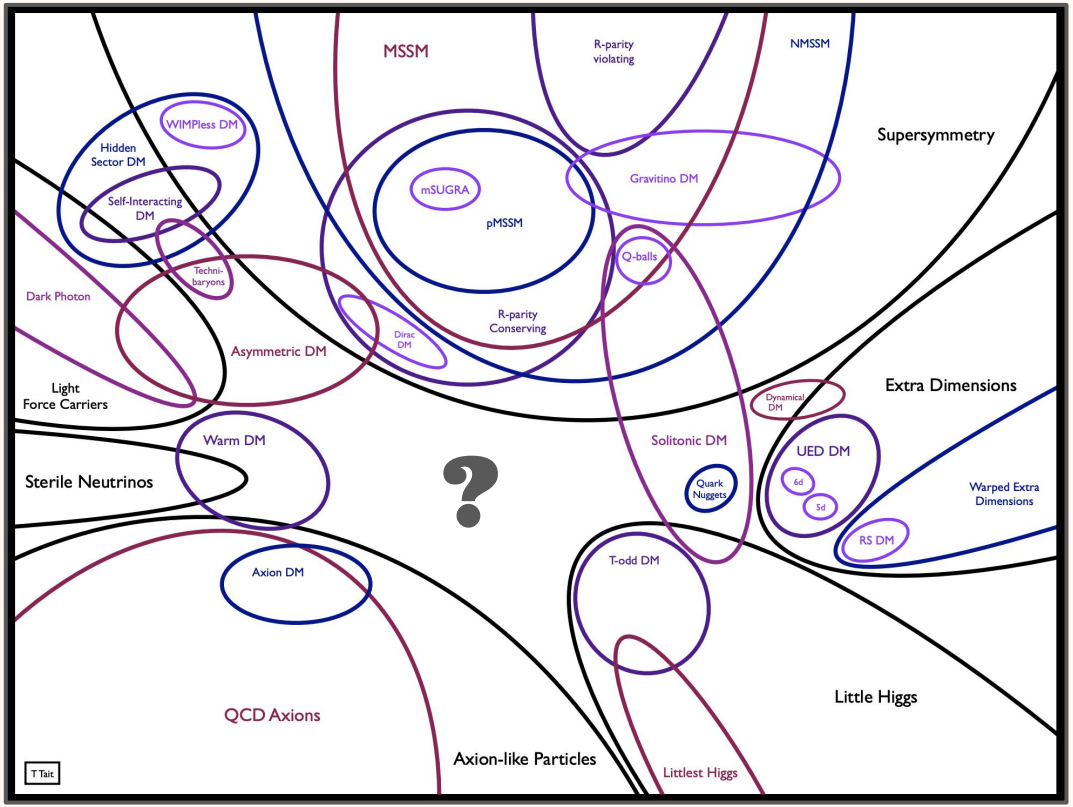
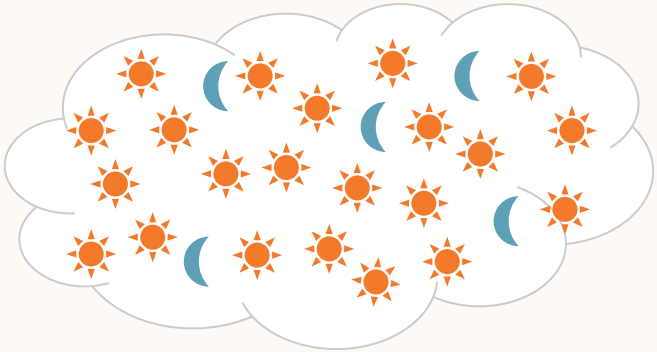


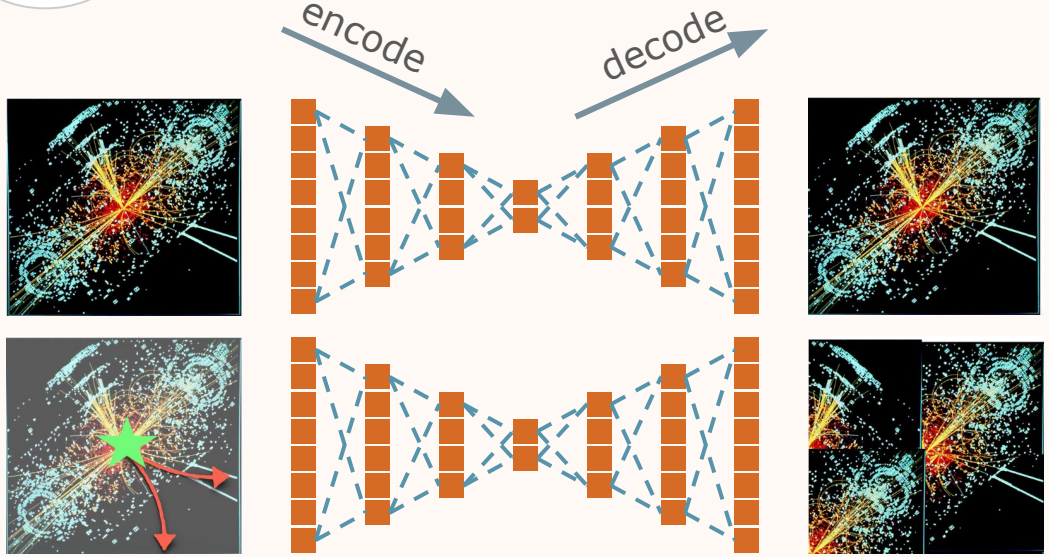
Figure from Timothy Tait

Anomaly detection for LHC Events



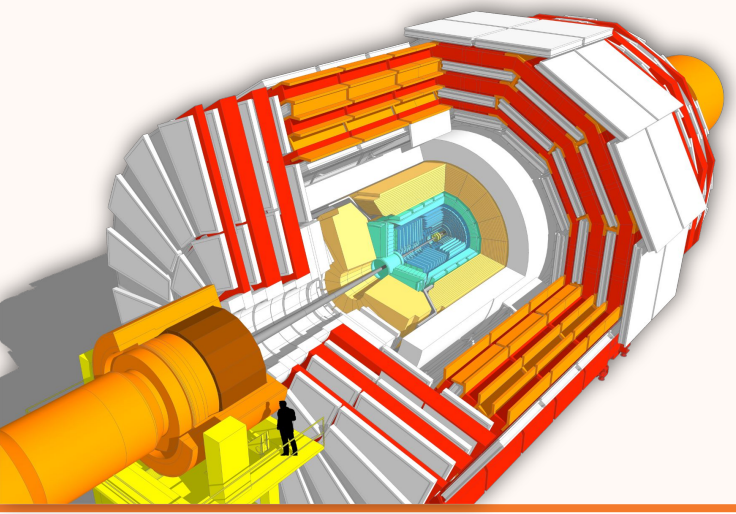
Goal: pick out anomalous events from background majority

Autoencoder method for AD:



CMS Trigger System

algorithms aimed to **pick out interesting physics**



40 MHz

Level-1 Trigger (L1T): 99.75% reduction in 3.8 μs

100 kHz

High Level Trigger (HLT): 1% reduction

1 kHz

for analysis

If the existing trigger system misses new physics, **that data will be lost**

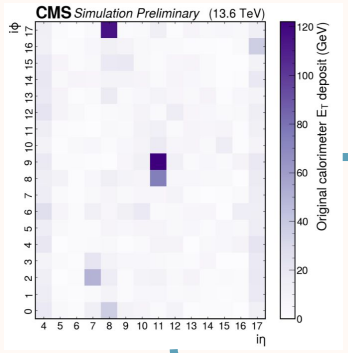
CMS has two anomaly detection-based L1T algorithms: **CICADA** and **AXOL1TL**

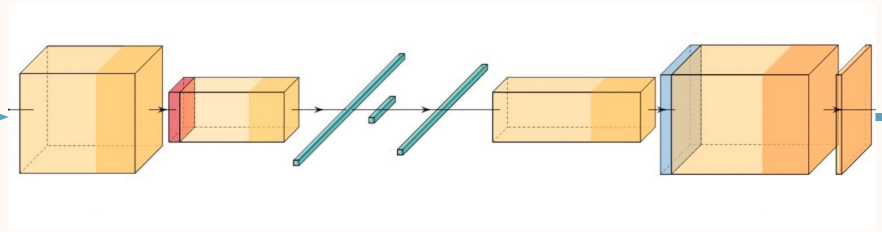
AD in the Level 1 Trigger: CICADA

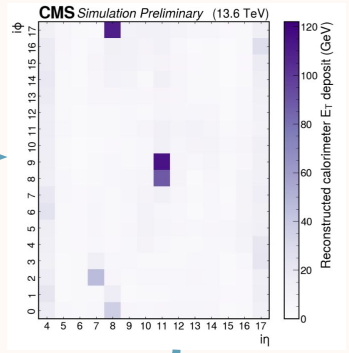
(Calorimeter Image Convolutional Anomaly Detection Algorithm)

Teacher Model:

Input: calorimeter image

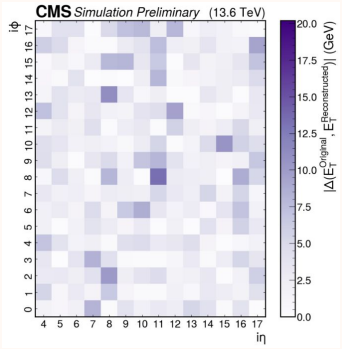






Output: reconstructed image

Reconstruction error



= 14.9

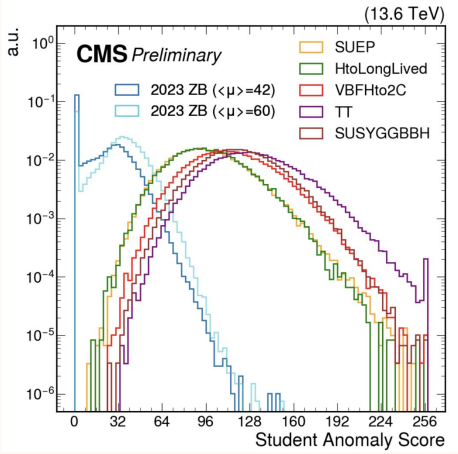
Student model learns anomaly score

Student Model: smaller model, implemented on an FPGA in the L1T

CMS DP-2024/121

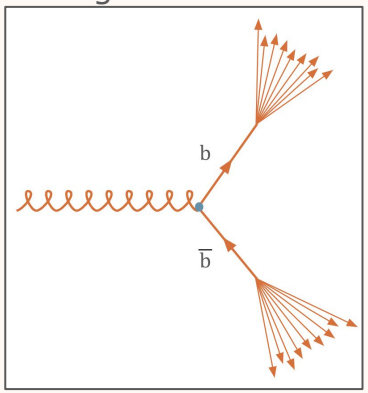
Applying CICADA to Analysis

- Both AD triggers (AXOL1TL and CICADA) are currently taking data.
  - **AXOL1TL**: ~100 fb^{-1} so far
  - **CICADA**: ~5 fb^{-1} so far
  - How to use this data?
- **Goal:** use CICADA to search for new physics
- **Proof of principle:** can we use CICADA to "rediscover" an existing result?



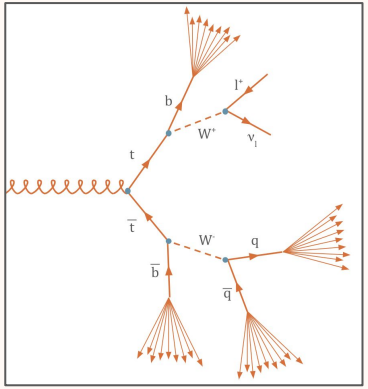
Small overlap between $t\bar{t}$ and background

Background Event



much lower cross section!

$t\bar{t}$ Event

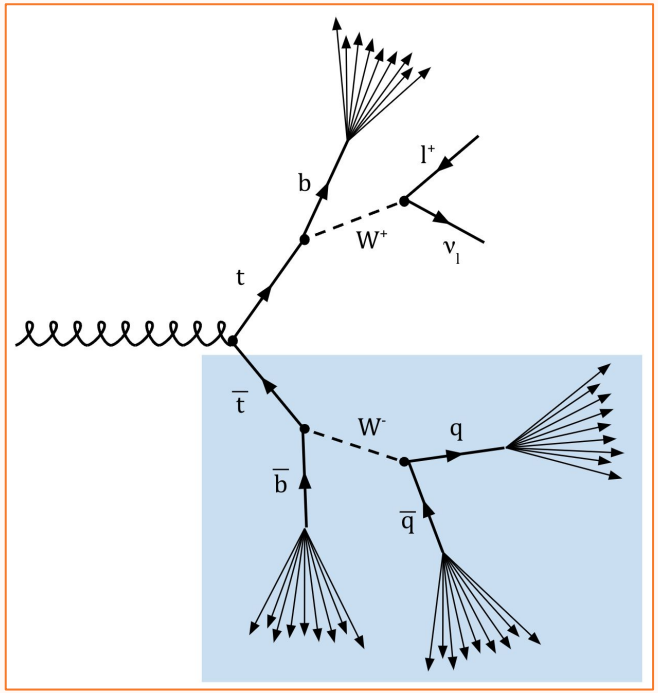


Search for $t\bar{t}$ using anomaly detection/ CICADA

CMS DP-2024/121
CMS DP-2024/059

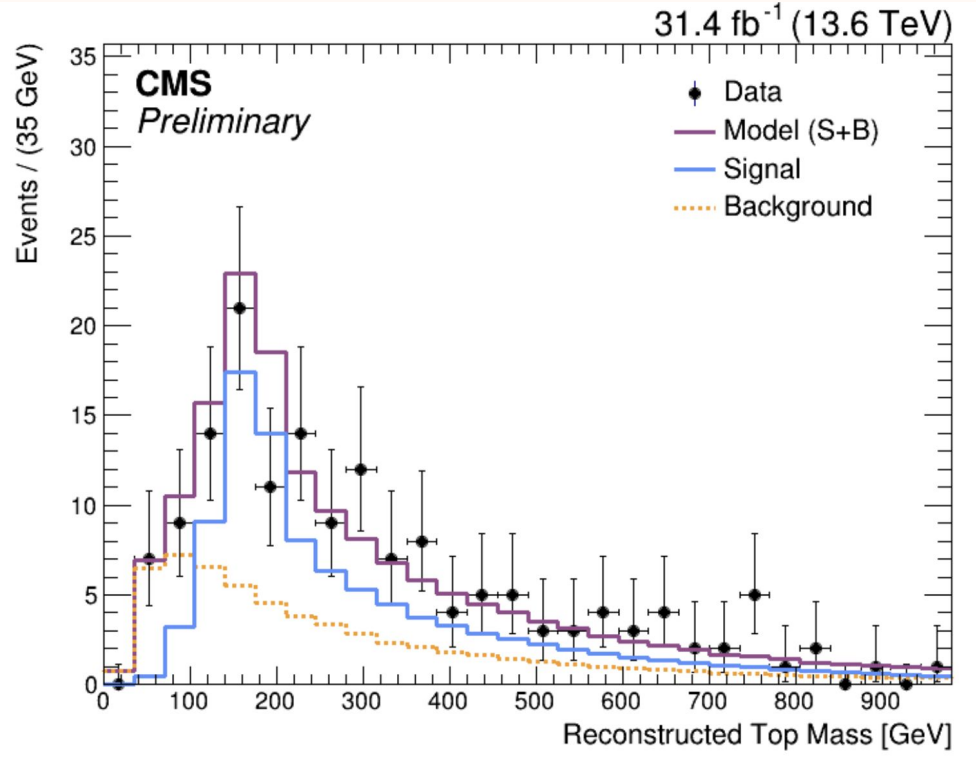
Constructing a search for ttbar using CICADA

- Search strategy developed before CICADA trigger started collecting data - working on limited dataset!
- Require at least **three jets** and at least **one b-tagged jet** (p_T>20 GeV)
- Top mass reconstruction
  - Consider all possible combinations of three jets with at least one b-tagged jet in each event
  - Calculate the **invariant mass of the three jet combination with the highest p_T**
- Use CICADA score to increase signal purity



Top mass reconstruction (m_t=172.76±0.3 GeV)

- Fit signal & background distributions to reconstructed top mass
- **Data:** Zero Bias data with cut on CICADA score > 115
- **Background:** Zero Bias data without cut on CICADA score
- **Signal:** Simulated TTbar data with cut on CICADA score > 115

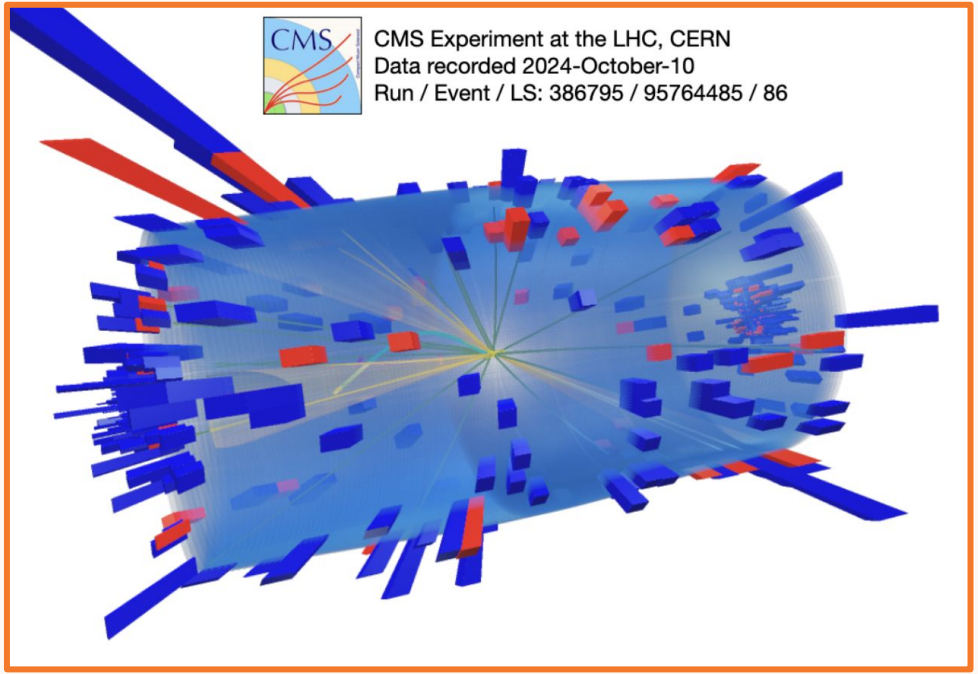


Using anomaly detection on a limited dataset, we observe a peak **consistent with the top mass**

CMS DP-2024/121

Future Directions

- Now we can look at **CICADA-triggered events**, which will offer a larger dataset
- Can look for **other Standard Model resonances** in CICADA-triggered data
  - Z to jets
  - di-electron spectrum
- Can use this information to inform **searches for new physics** using CICADA
  - Bump hunt across di-object spectra
  - Targeted searches for phenomena such as emerging jets
  - Joint search with AXOL1TL, the other CMS L1 AD trigger



Event triggered by CICADA and no other algorithm

CMS DP-2024/121

Thank you!

Questions?

With help from the CICADA team, including Isobel Ojalvo, Andrew Loeliger, Pallabi Das, Kiley Kennedy and Lino Gerlach

References

CMS Collaboration, Model-Independent Real-Time Anomaly Detection at the CMS Level-1 Calorimeter Trigger with CICADA (CMS DP2024/121), Tech. report, CERN, cms-dpg-conveners-l1t@cern.ch, November 2024, Available at https://twiki.cern.ch/twiki/bin/view/CMSPublic/L1TriggerDPGResults. Link to: DP Note

CMS Collaboration, 2024 Data Collected with AXOL1TL Anomaly Detection at the CMS Level-1 Trigger (CMS DP2024/059), Tech. report, CERN, cms-dpg-conveners-l1t@cern.ch, cms-trigger-coordinator@cern.ch, and cms-conveners-ml@cern.ch, July 2024, Available at https://twiki.cern.ch/twiki/bin/view/CMSPublic/L1TriggerDPGResults. Link to: DP Note

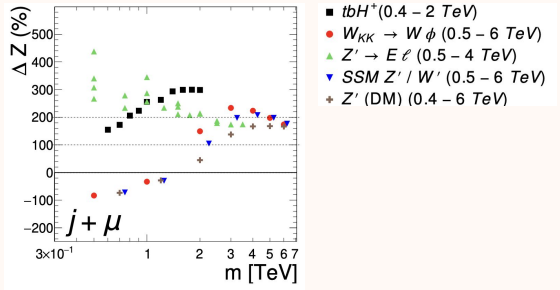
Backup: past AD searches

Anomaly detection on **pre-triggered events**

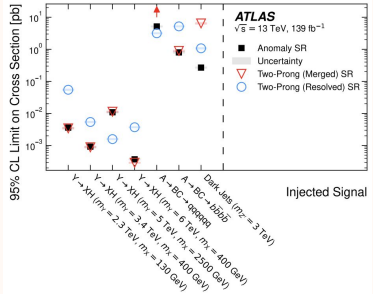
1. ATLAS: Autoencoder AD in di-object resonance search (ATLAS, 2023)
2. ATLAS: Autoencoder AD in search for new resonance decaying to Higgs + new particle X (ATLAS, 2023)
3. CMS: Various AD techniques for anomalous jet substructure in di-jet resonance search (CMS, 2024)

Signal injection tests to evaluate **increase in search sensitivity**:

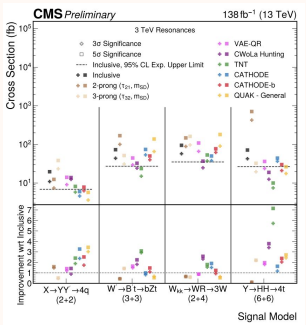
1.



2.

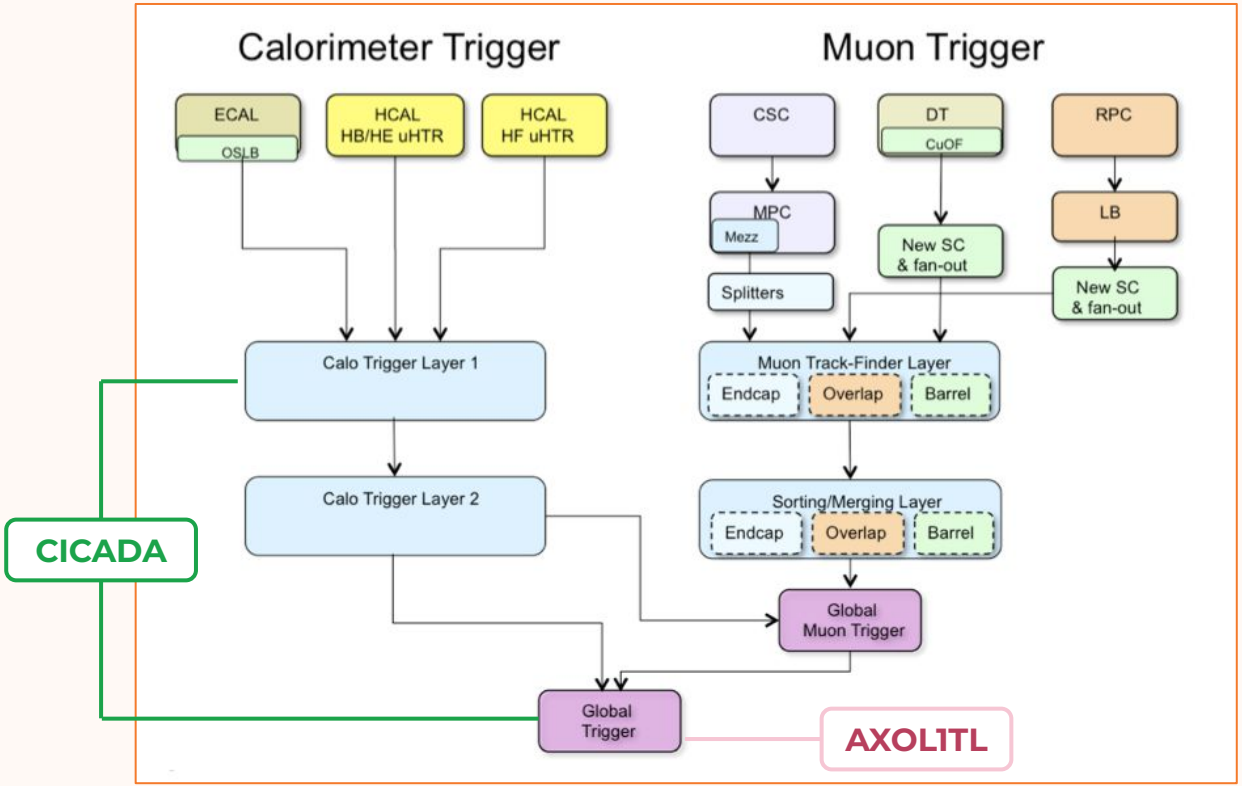


3.



Anomaly detection at the trigger level may further improve search sensitivity

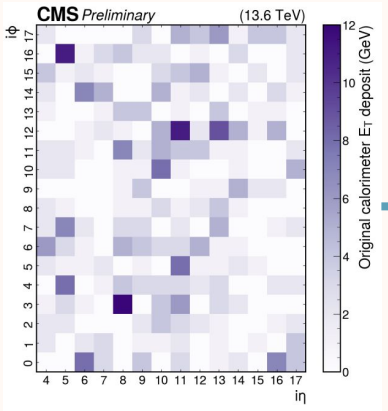
Backup: CMS Level-1 Trigger



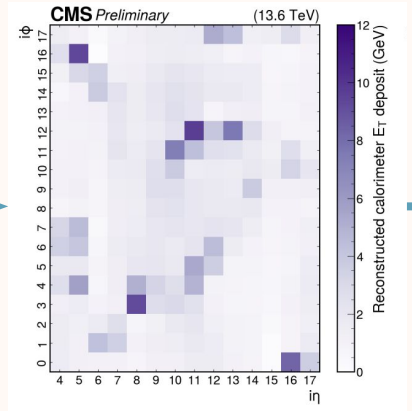
Backup: CICADA Background Event Reconstruction

Teacher Model:

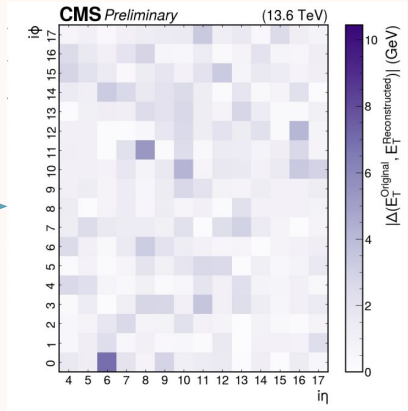
Input: calorimeter image



Autoencoder reconstruction



Reconstruction error

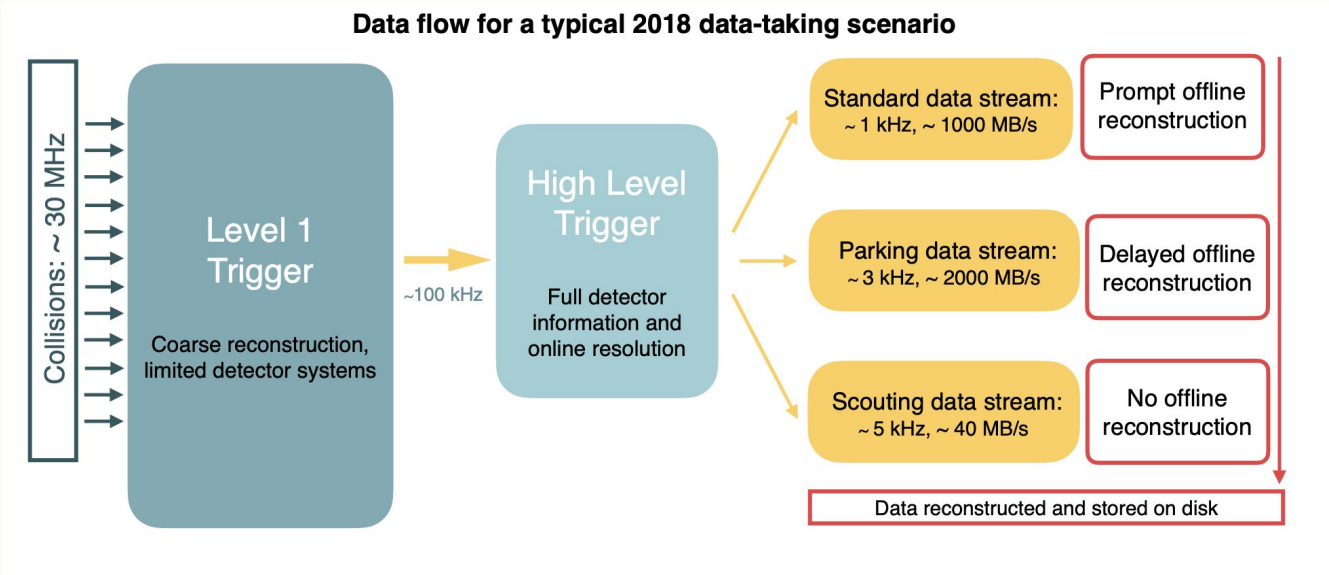


= 2.6

CMS DP-2024/121

Backup: CMS Scouting Data

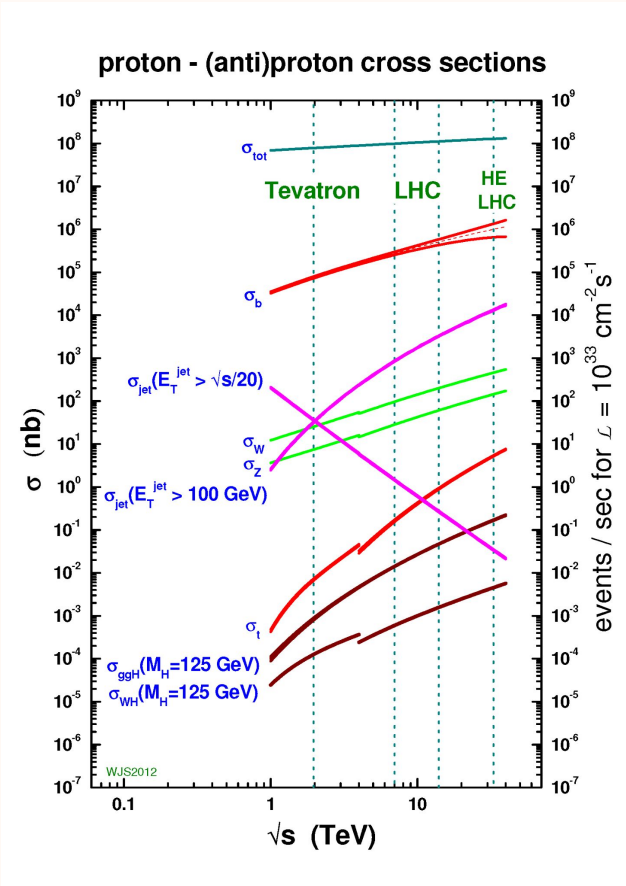
Store less event information in exchange for a higher event rate



The CMS Collaboration, 2024

Backup: Cross-sections

- QCD background: σ_b
- Top quark production: σ_t
- W and Z have higher cross-sections → may also be able to find using CICADA



Backup: more information about top mass reconstruction method

- Slimmed Jets: ak4 PF Jets PUPPI with JEC applied
  - ak4: anti-k_T jet clustering algorithm with $\Delta R=0.4$
  - PUPPI (Pileup Per Particle Identification): removes charged particles originating from different primary vertices
  - JEC: jet energy corrections
- b-tagging algorithm: btagDeepCSV
  - Deep neural network that simultaneously tags c quarks
- Emulated CICADA score method
  - Zero Bias: random trigger in CMS
  - Calculate CICADA score on Zero Bias events
  - Use CICADA score as cut

Backup: AXOL1TL

- **A**nomaly e**X**traction **O**nline **L**evel-**1** **T**rigger **L**ightweight
- Takes in four-vectors of reconstructed particles (muons, e/γ, jets)
- **Variational autoencoder** (VAE):
  - Encodes input into a distribution with mean μ and variance σ instead of a single point
  - Decoder samples from distribution
- Trained on Zero Bias events
- Implementation on FPGA: only use encoding step, **use μ as anomaly score**