# Realtime Anomaly Detection @ L1 Trigger

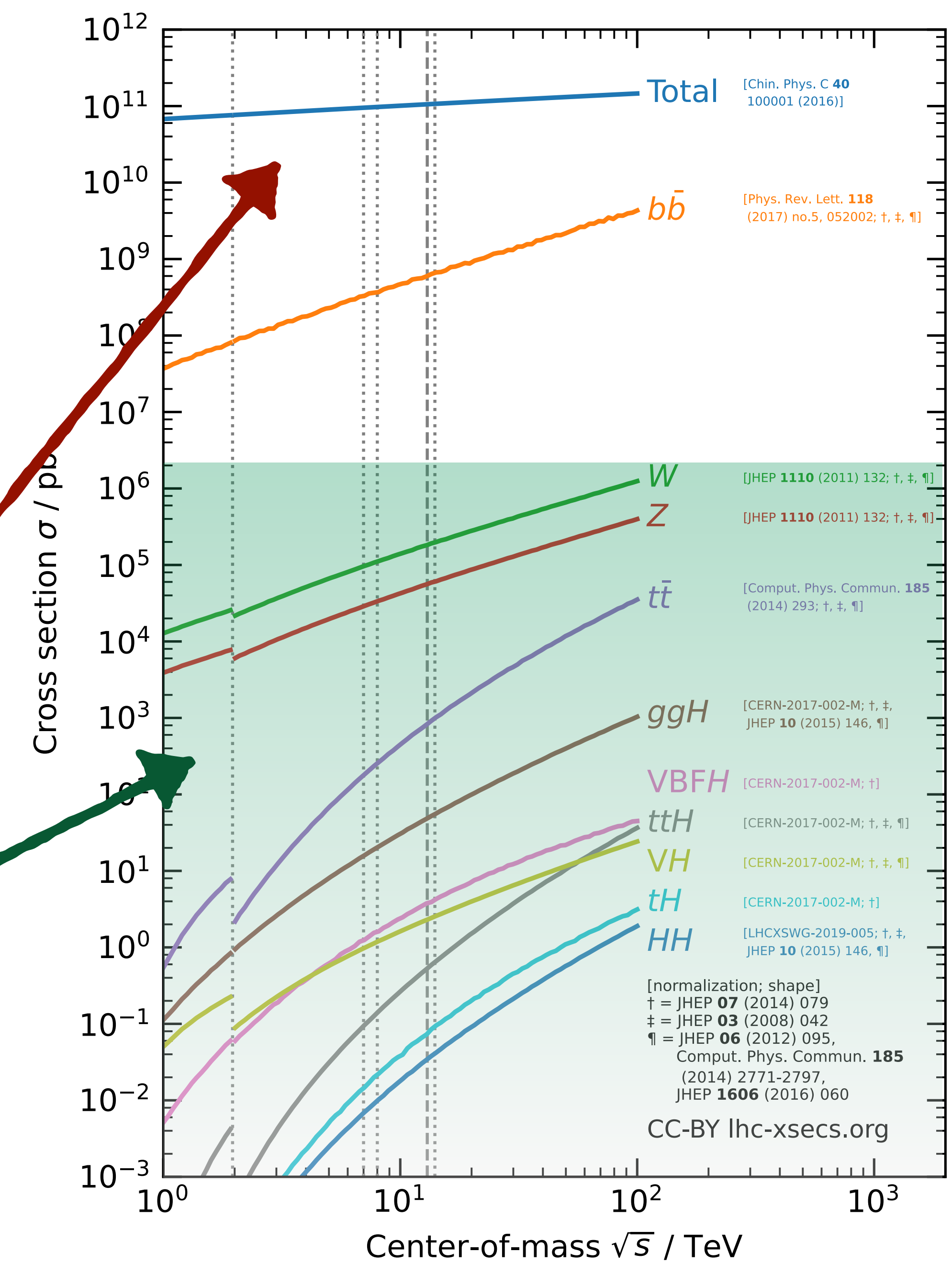
Abhijith Gandrakota
Fermilab

On behalf of CMS collaboration

ICHEP 2024
Prague, Czech Republic

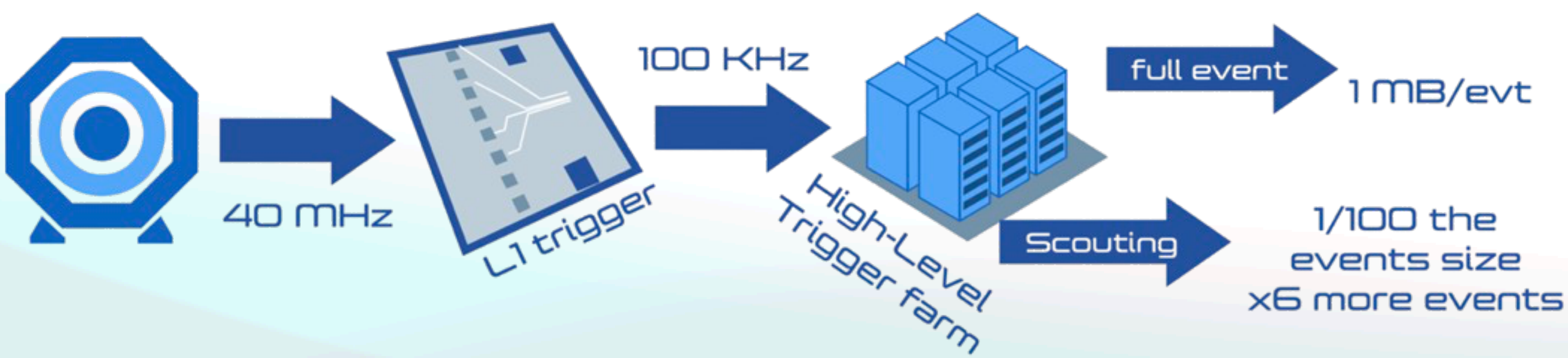
# Introduction

- CMS produces more data than what we can handle
  - Two-level trigger to capture interesting physics
  - Need quick decisions w/ large data rate
    - L1 runs on FPGAs
      (w/ $3.8\mu s$ latency & 110 kHz output rate)




Discard a lot of this

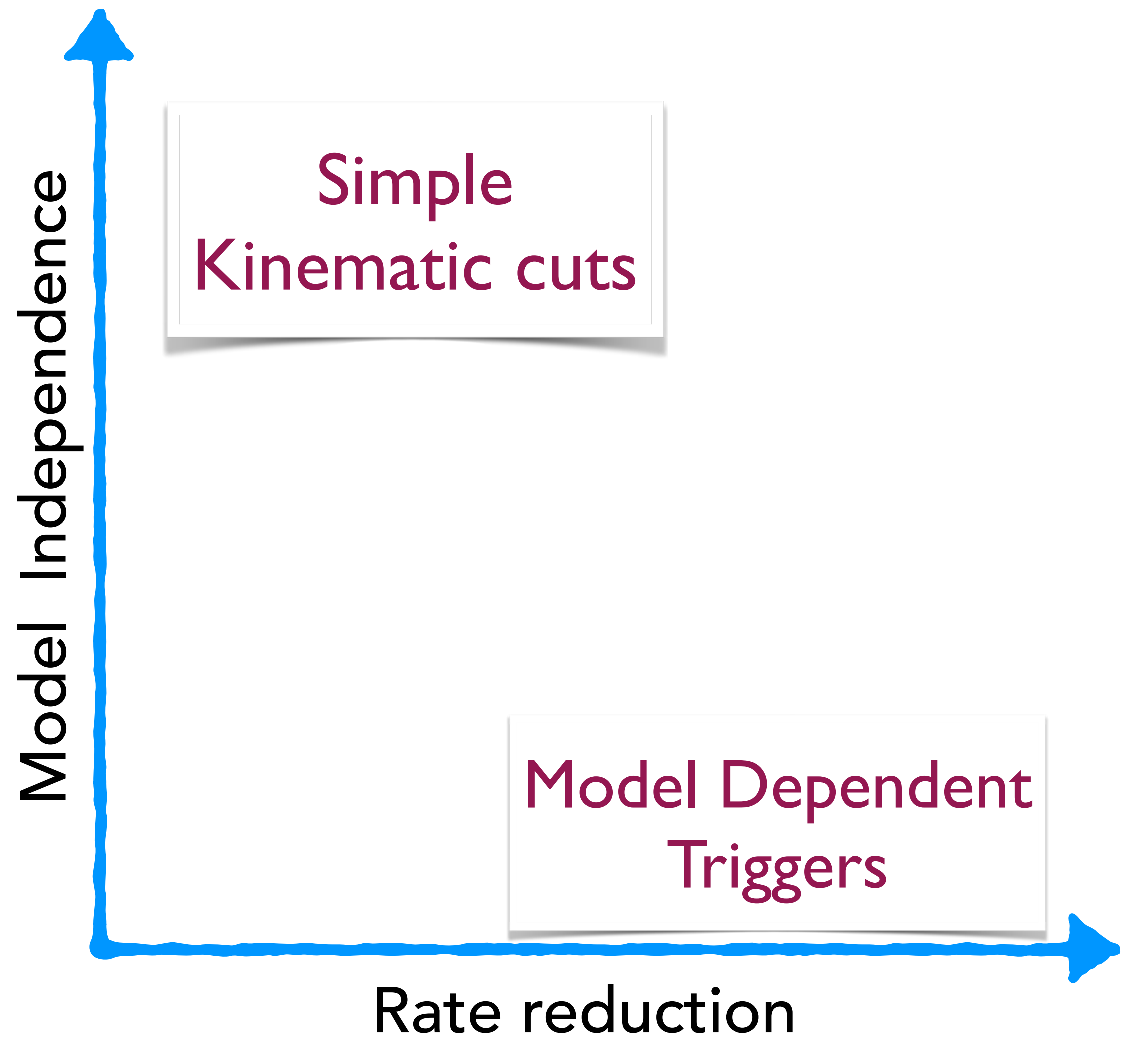
Retain much of this
(Needs to be unbiased)




99.75% gets thrown away in first stage!

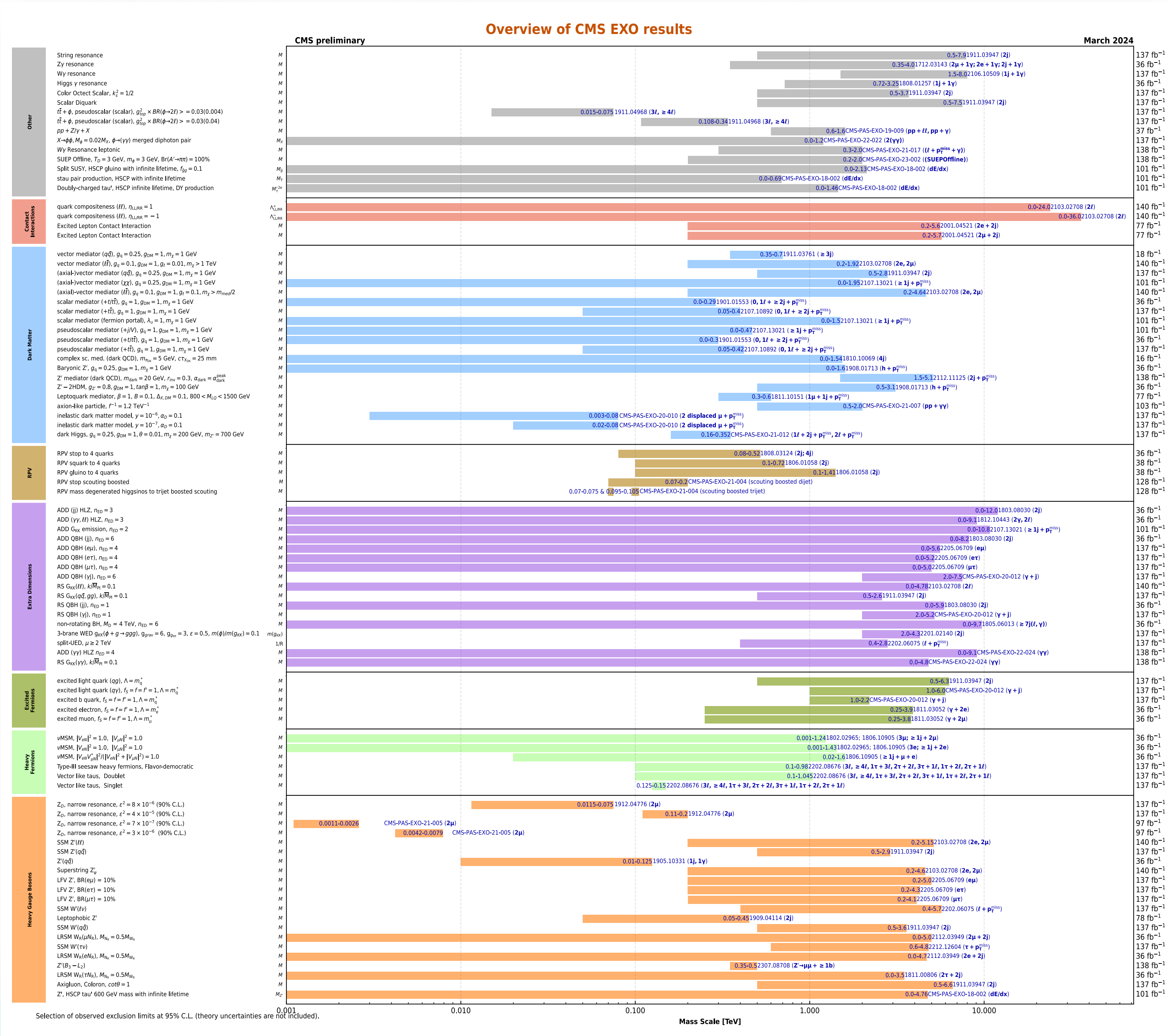
# Introduction

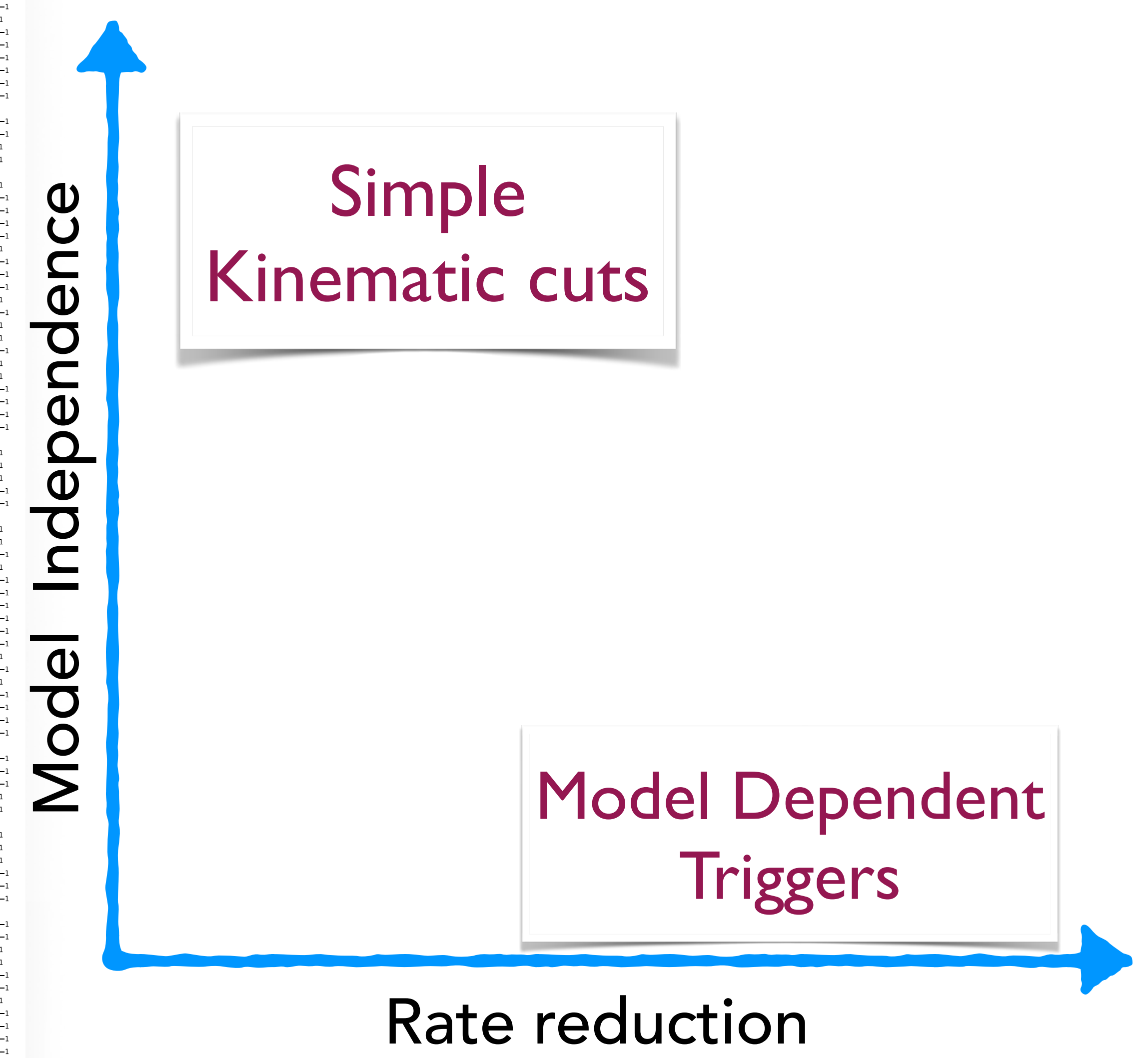
- Objective of a trigger: Manageable rates, Enhance sensitivity to BSM / rare SM
  - Discard typical events, Retain only interesting (*atypical*) events
- Atypicality through kinematic selection
  - Low sensitivity, Model independent
- BSM / Rare SM model signature basis
  - Higher sensitivity, Model dependent



# Introduction

- Objective of a trigger: Manageable rates, Enhance sensitivity to BSM / rare SM

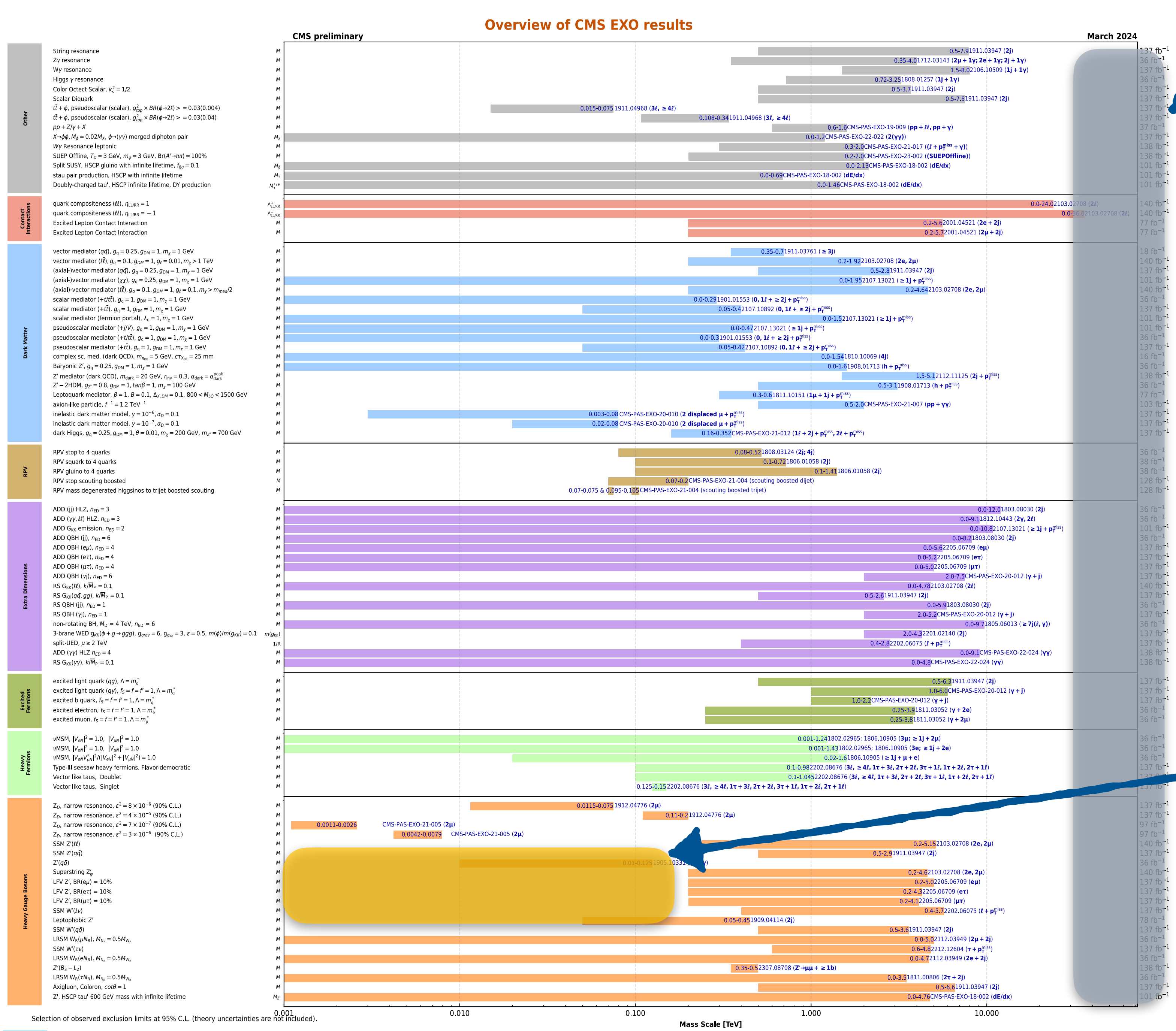


# Introduction

Fermilab

- Objective of a trigger: Manageable rates, Enhance sensitivity to BSM / rare SM




Model Independence

Simple Kinematic cuts

Model Dependent Triggers

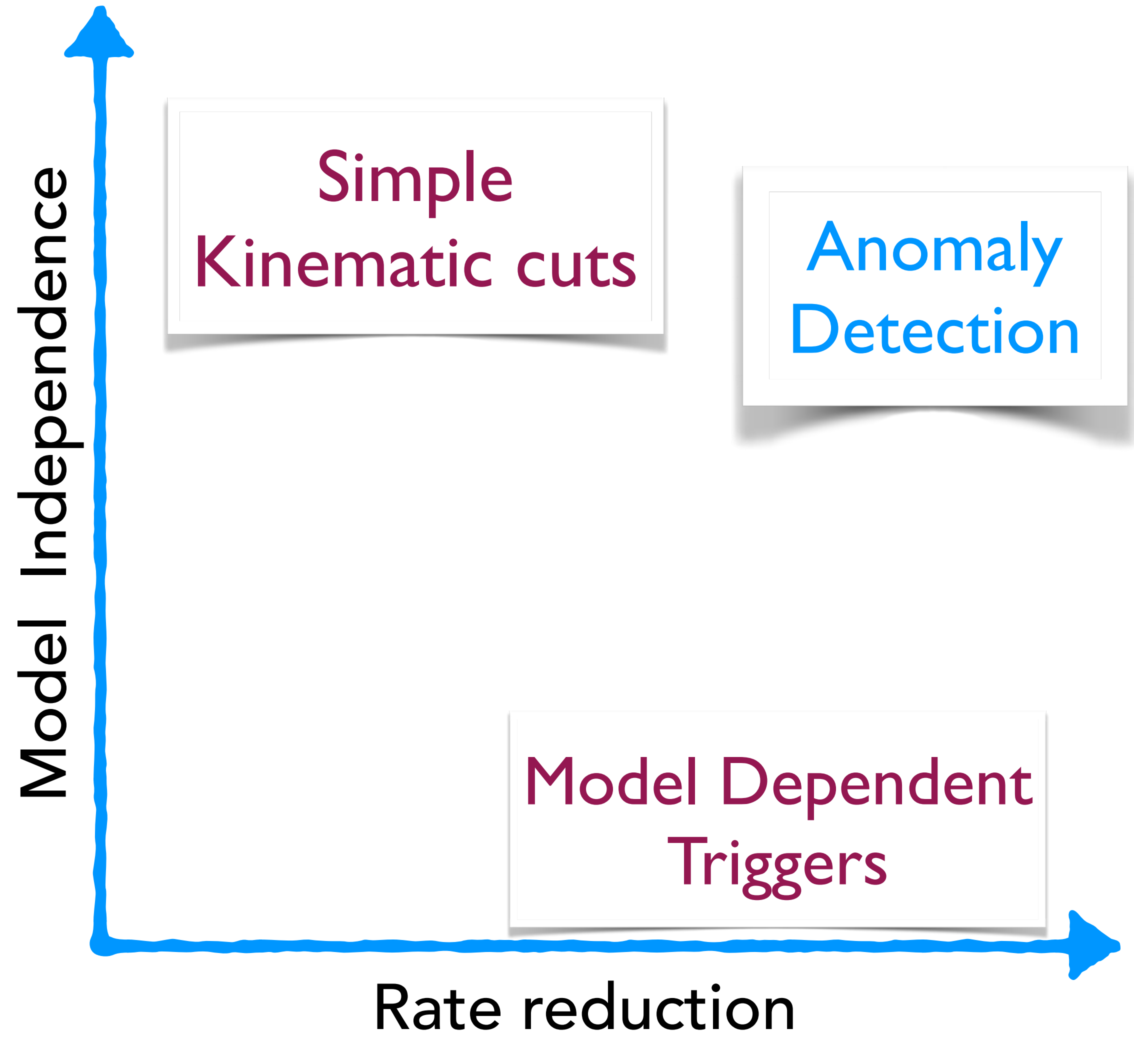
Rate reduction

? ? ?

# Why Anomaly detection ?

- Main objective of a trigger: Manageable rates, Sensitivity to BSM / rare SM
  - Discard typical events, Retain only interesting (*atypical*) events

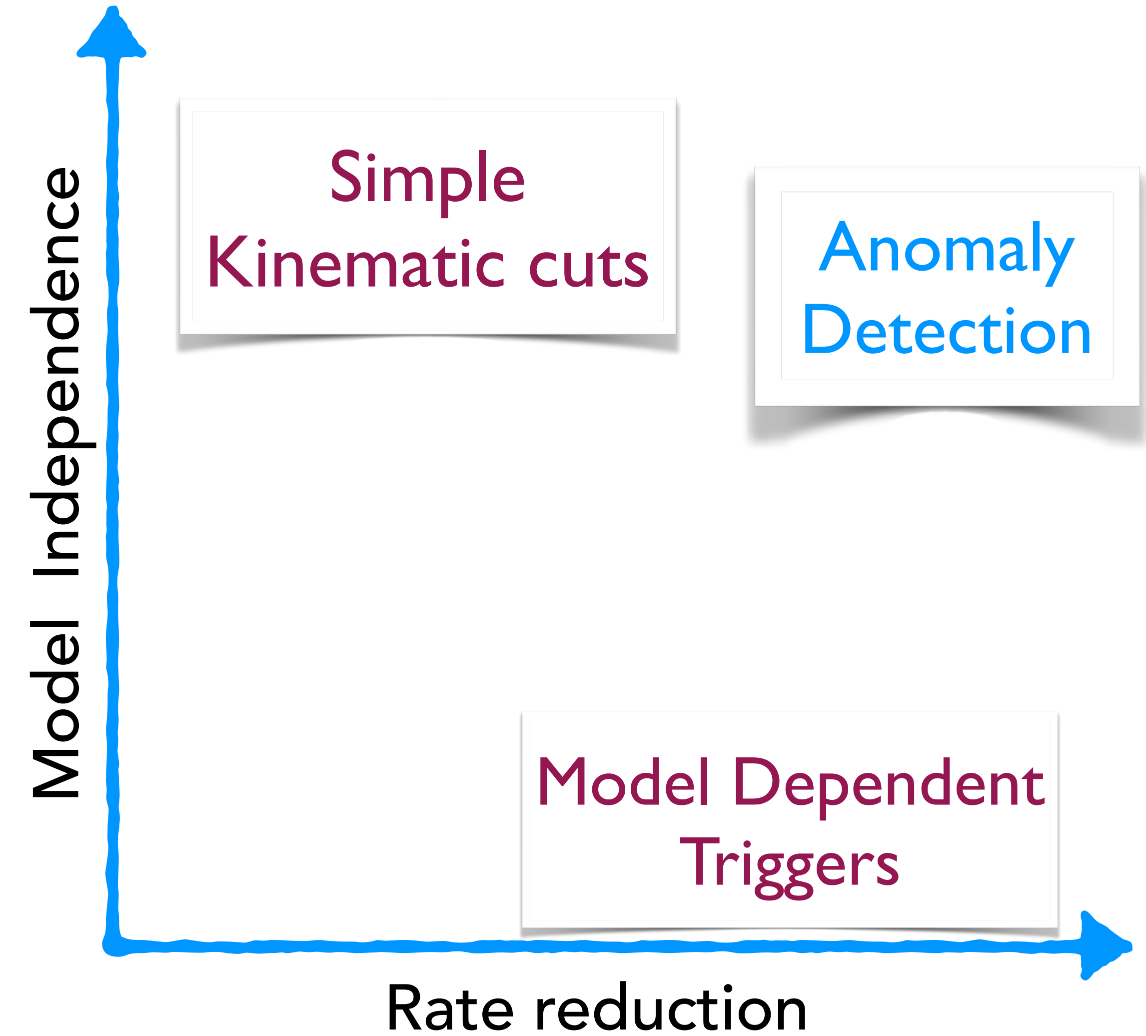
- New Approach: Anomaly Detection
  - High sensitivity, model agnostic
  - With Machine Learning @ L1 trigger
  - Trained directly on data (ZeroBias)
  - Unbiased approach to capture new physics



# Why Anomaly detection ?

- Main objective of a trigger: Manageable rates, Sensitivity to BSM / rare SM
  - Discard typical events, Retain only interesting (*atypical*) events

- New Approach: Anomaly Detection
  - High sensitivity, model agnostic
  - With Machine Learning @ L1 trigger
  - Trained directly on data (ZeroBias)
  - Unbiased approach to capture new physics

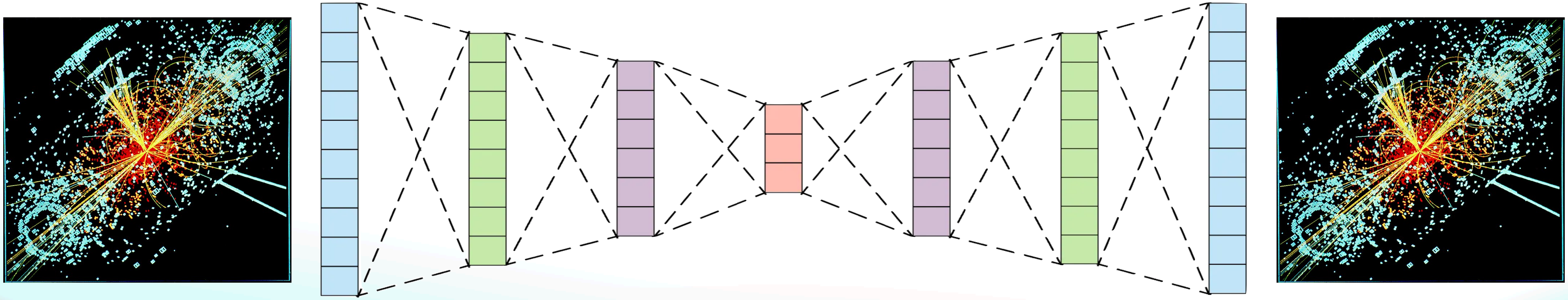



- Two complementary approaches in CMS: AXOL1TL and CICADA
  - Anomaly eXtraction Online Level-1 Trigger aLgorithm
  - Calorimeter Image Convolutional Anomaly Detection Algorithm

Check out Andrew's talk
for all Novel Run-3 Triggers

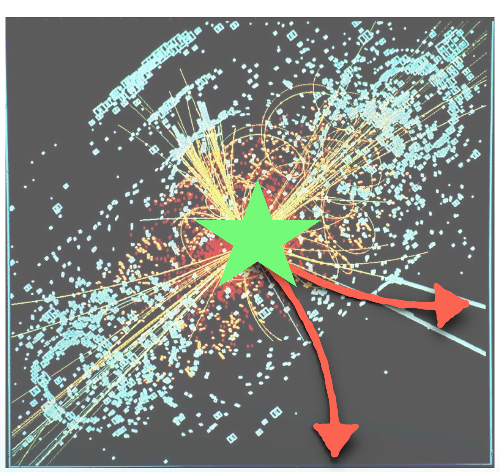
# Autoencoders

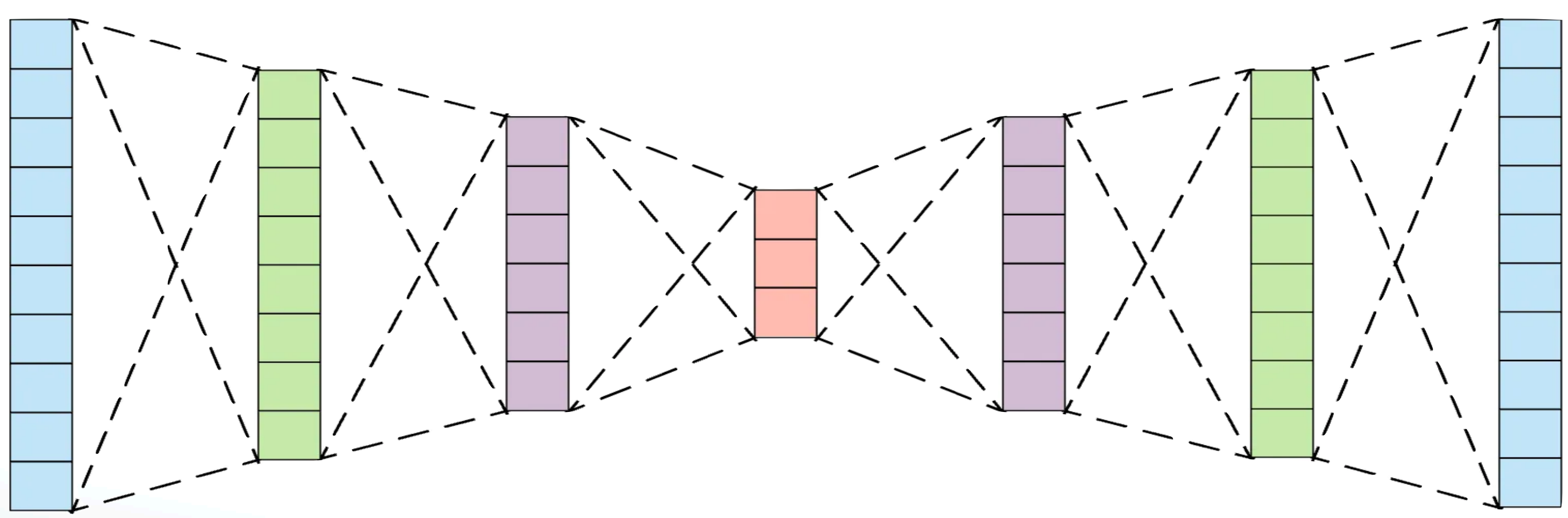
- Unsupervised ML algorithm to learn efficient encodings of backgrounds in data
  - Encodes inputs to latent space, from these decoder tries to recreate input data
  - Trained to have a minimal difference between the input and output
- For a typical event, the output will have a minimal error compared to input

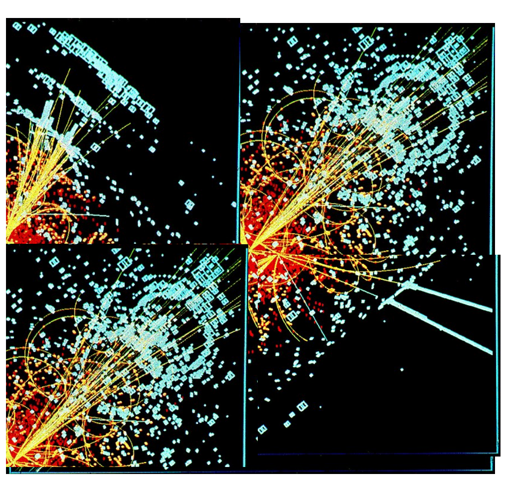


# Autoencoders

- Unsupervised ML algorithm to learn efficient encodings of backgrounds in data
  - Encodes inputs to latent space, from these decoder tries to recreate input data
  - Trained to have a minimal difference between the input and output
- For a typical event, the output will have a minimal error compared to input
  - For a *atypical* event, will result in high reconstruction error!







$$\mathscr{L} = ||x - x'||$$

# How do we train it ?

Fermilab

- Learning *typicality* : By training on Zero Bias dataset

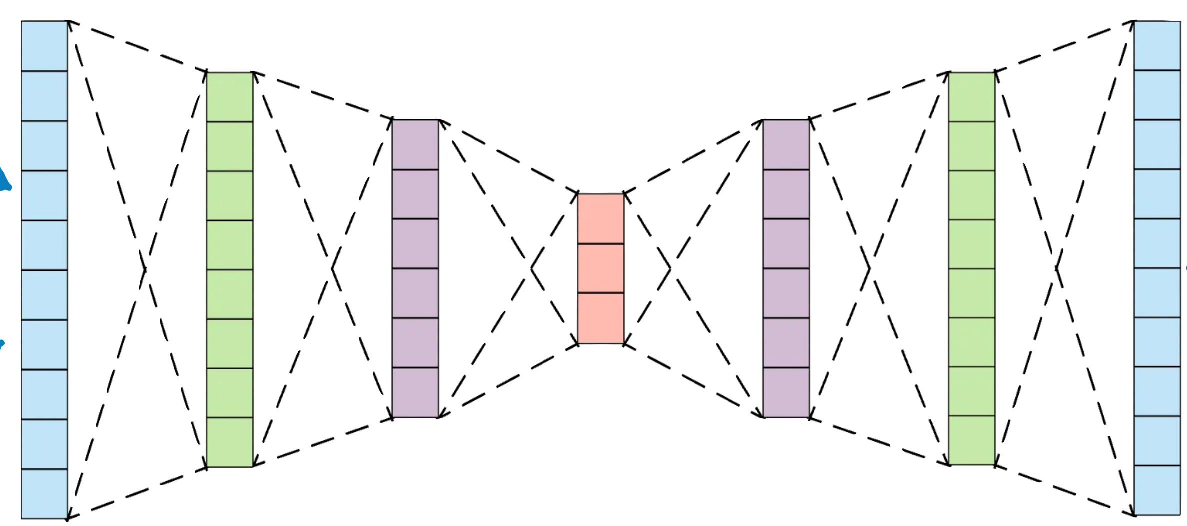
AXOL1TL

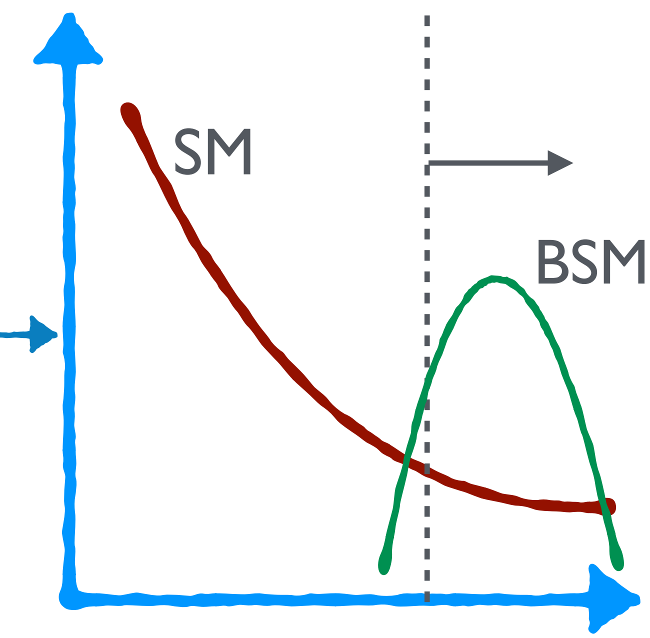
7 million

| | $p_T$ | η | φ |
|---|---|---|---|
| MET | | N/A | |
| 4 e/γ | | | |
| 4 μ | | | |
| 10 jets | | | |

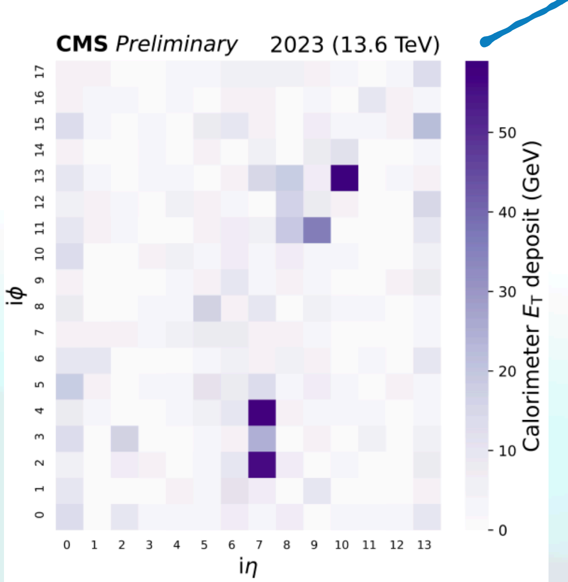
From calorimeter and muon trigger systems:

Objects : Jets(x10) , $e/\gamma$ (x4), $\mu$ (x4), MET
Attributes: $P_T, \eta, \phi$ in raw integer value






CICADA




Low level input :

Calorimeter towers, grouped as calo regions
Basically the energy pixels

# How do we train it ?

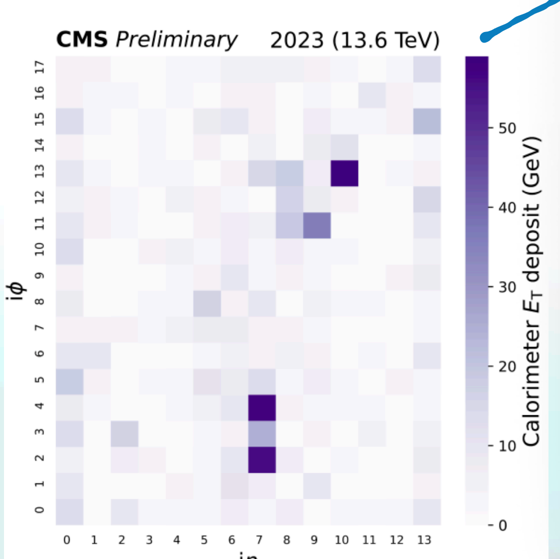
- Learning *typicalit*

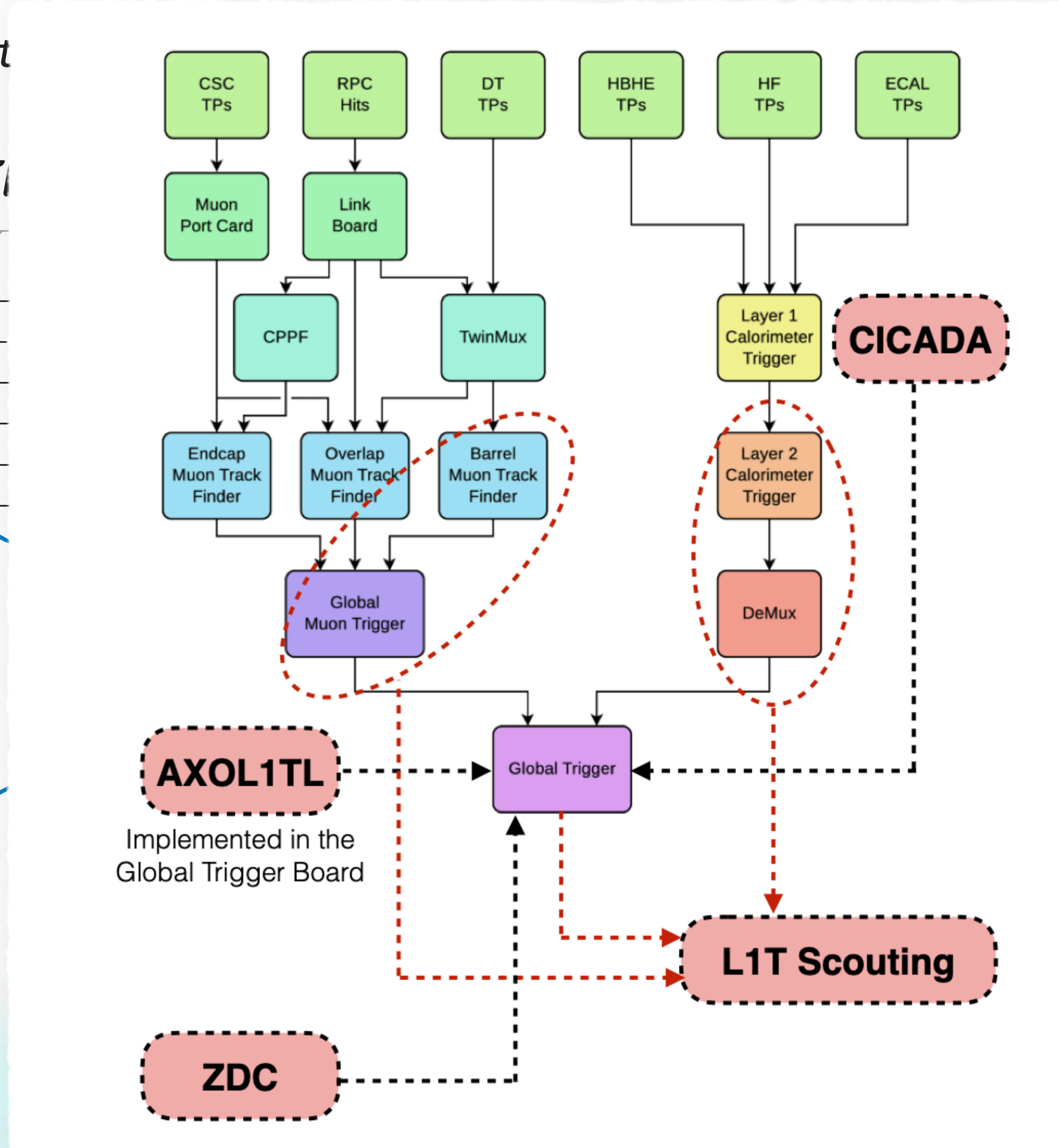
AXOL1TL

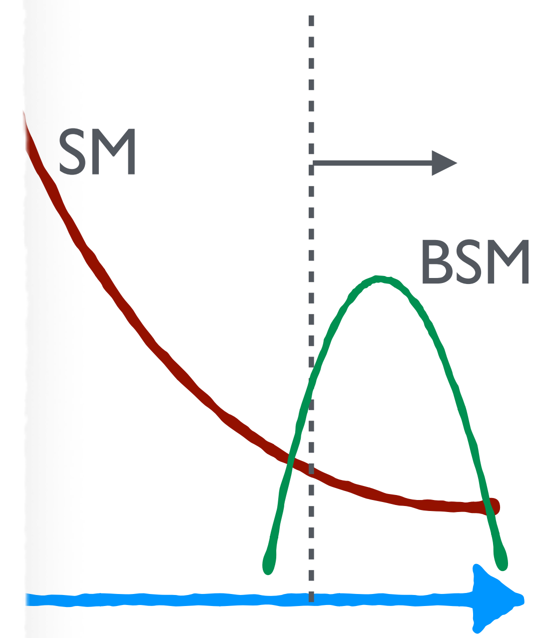
7 million

| | $p_T$ | η | φ |
|---|---|---|---|
| MET | | N/A | |
| 4 e/γ | | | |
| 4 μ | | | |
| 10 jets | | | |

CICADA



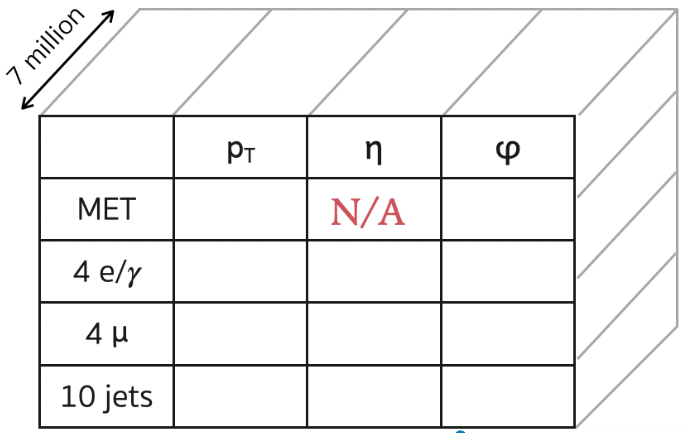



# How do we train it ?

- Learning *typicality* :   By training on Zero Bias dataset

AXOL1TL

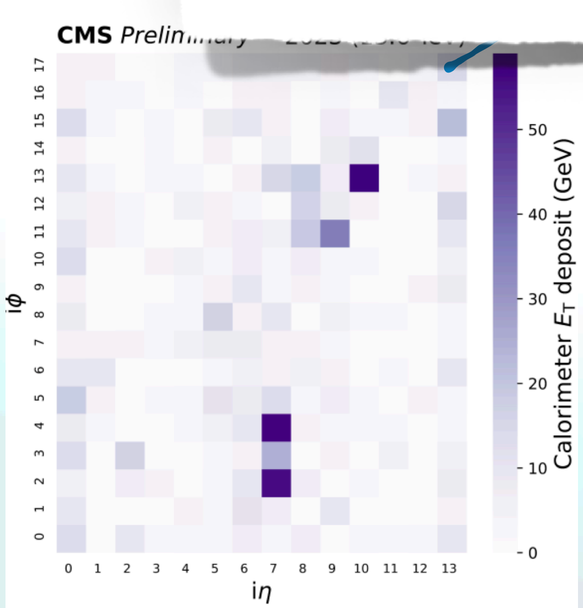



| | $p_T$ | η | φ |
|---|---|---|---|
| MET | | N/A | |
| 4 e/$\gamma$ | | | |
| 4 μ | | | |
| 10 jets | | | |

From Calo layer-2 and GMT:

Objects   : Jets(x10) , $e/\gamma$ (x4), $\mu$ (x4), MET
Attributes: $P_T, \eta, \phi$ in raw hardware value


New tools were developed for ML inference using FPGAs.
NN can be implemented into hardware today!




Low level input :

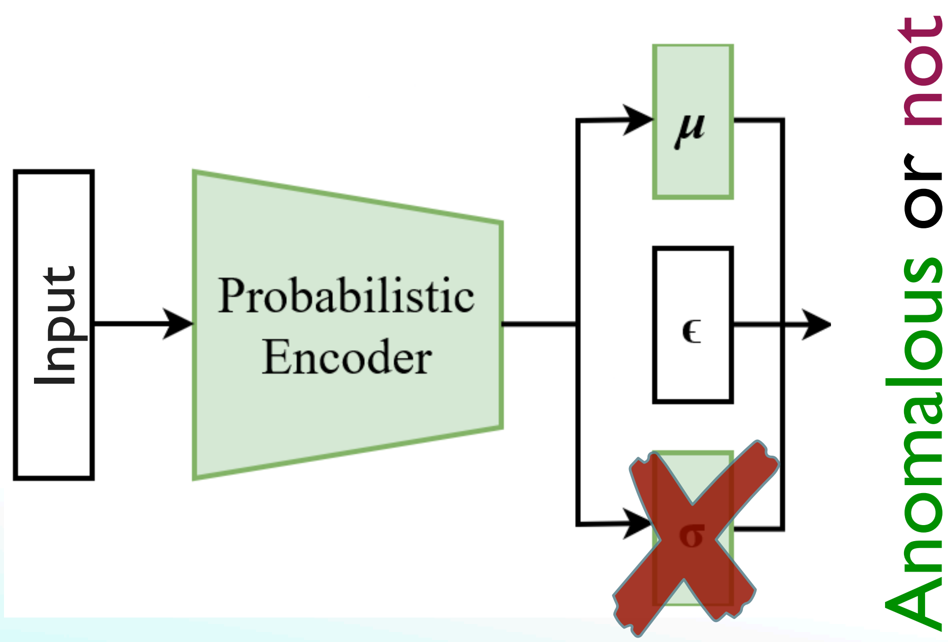
Calorimeter towers, grouped calo regions

# Shrinking the Autoencoders

Fermilab

## AXOL1TL

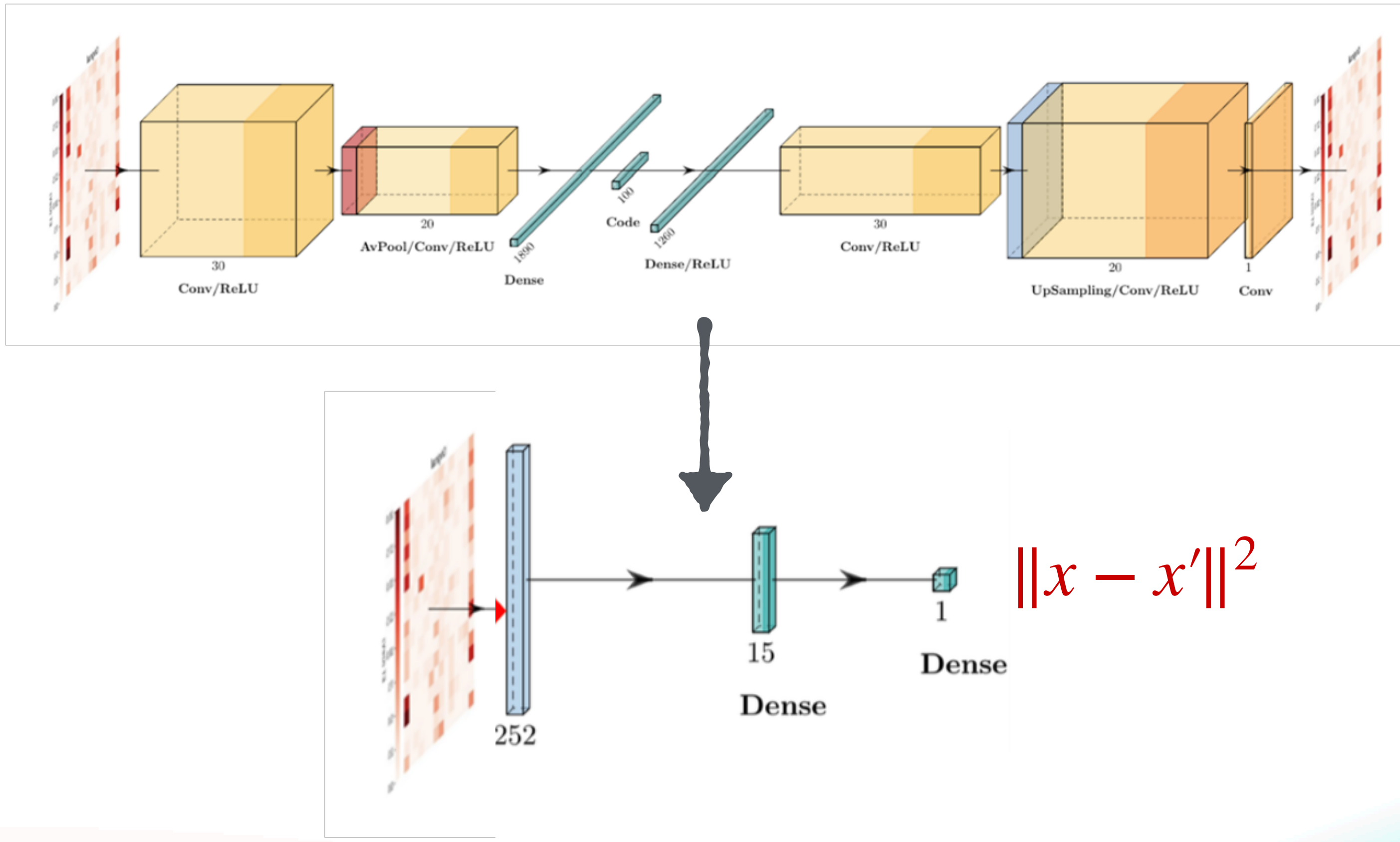
- Uses variational autoencoder
  - In training require the latent space is gaussian distributed $\sim \mathcal{N}(0,1)$
  - For *atypical* events from $\sim \mathcal{N}(0,1)$
- Solution: Deploy only the encoder




Anomaly Metric $= \mu^2$

## CICADA

- Uses convolutional autoencoder
  - Most suited for image like inputs
- Knowledge distillation to compress the autoencoder to a smaller model size

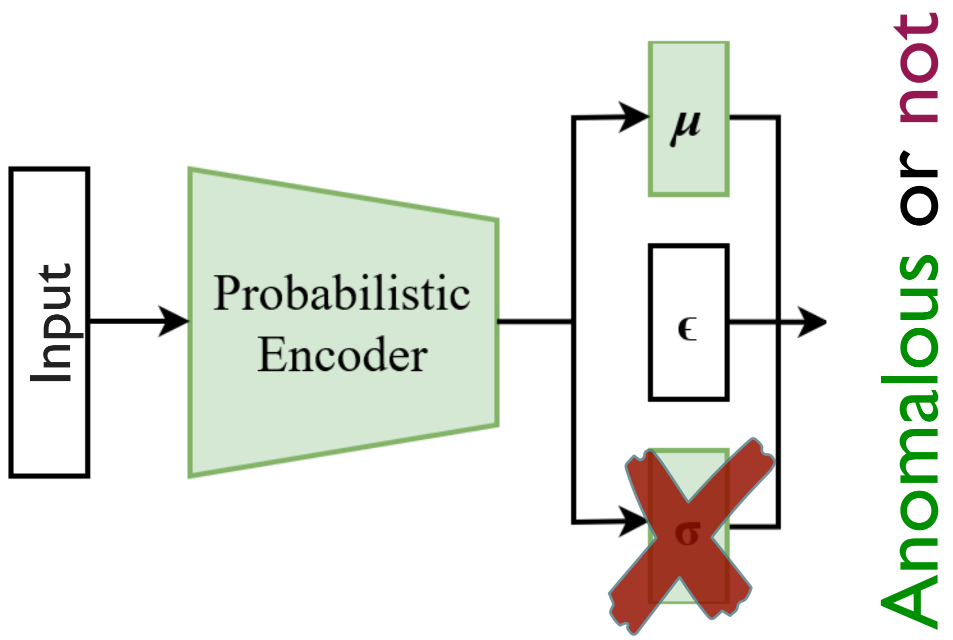



- Compressed model outputs the reconstruction error

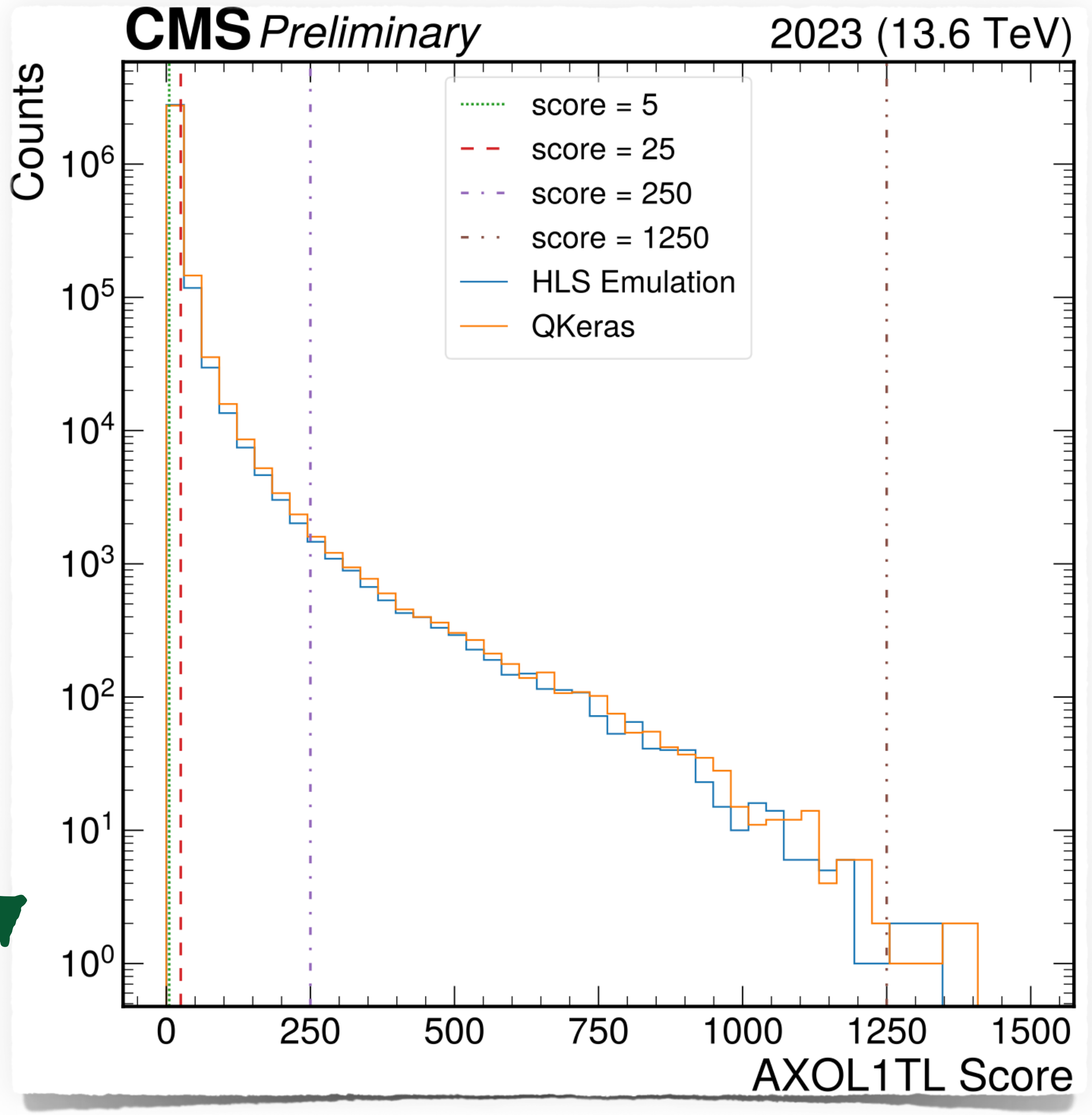
# Shrinking the Autoencoders

AXOL1TL

- Uses variational autoencoder
  - Deploy only the encoder




Anomaly Metric $= \sum \mu^2$

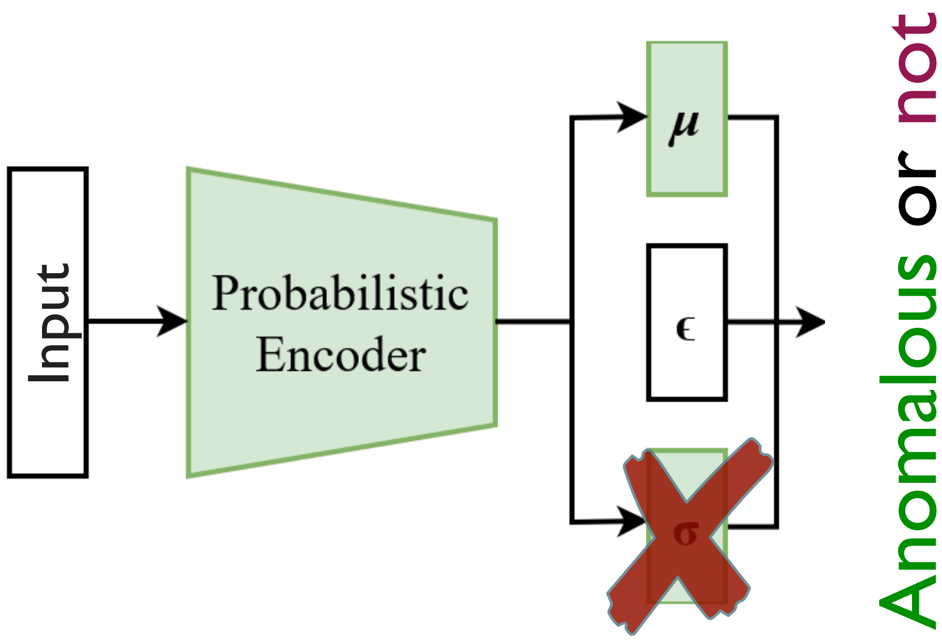



Different anomaly thresholds for targeting different rates

# Shrinking the Autoencoders

AXOL1TL

- Uses variational autoencoder
  - Deploy only the encoder




Anomaly Metric $= \sum \mu^2$

- *Preliminary physics performance*
  - Evaluated on $H \rightarrow aa \rightarrow 4b$ MC

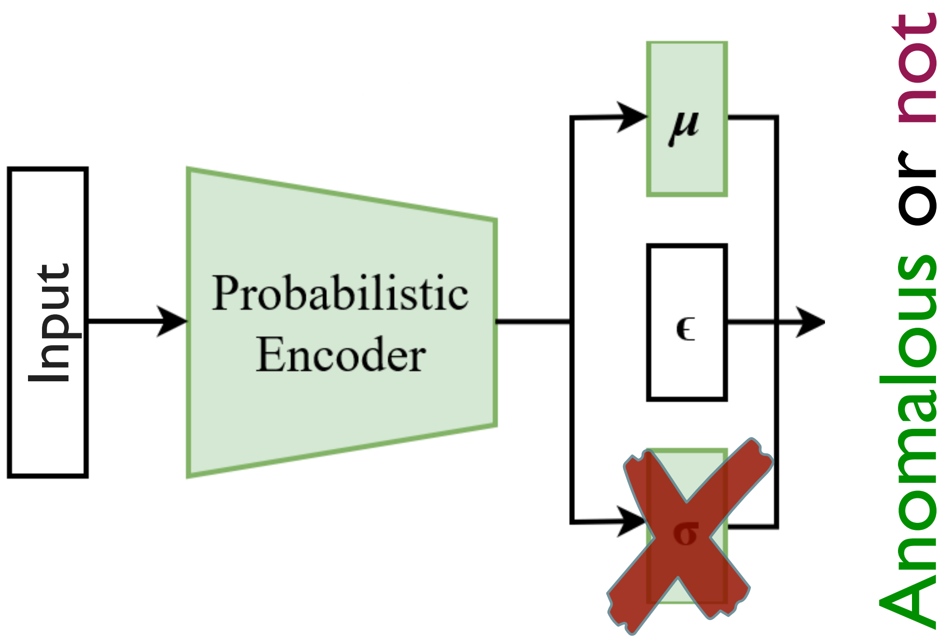
| *AXOL1TL rate* | 1 kHz | 5 kHz |
|---|---|---|
| *Signal Efficiency gain* | *46%* | *100%* |

- Large increase in signal efficiency for a small increase

# Shrinking the Autoencoders

AXOL1TL

- Uses variational autoencoder
  - Deploy only the encoder

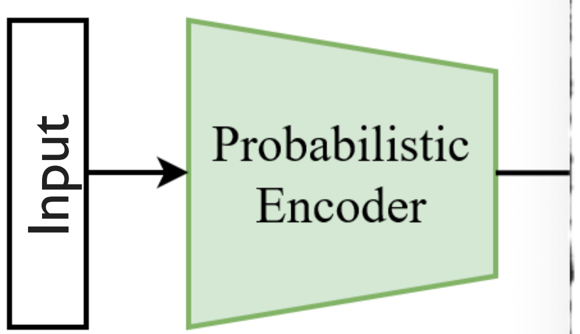



Anomaly Metric $= \sum \mu^2$

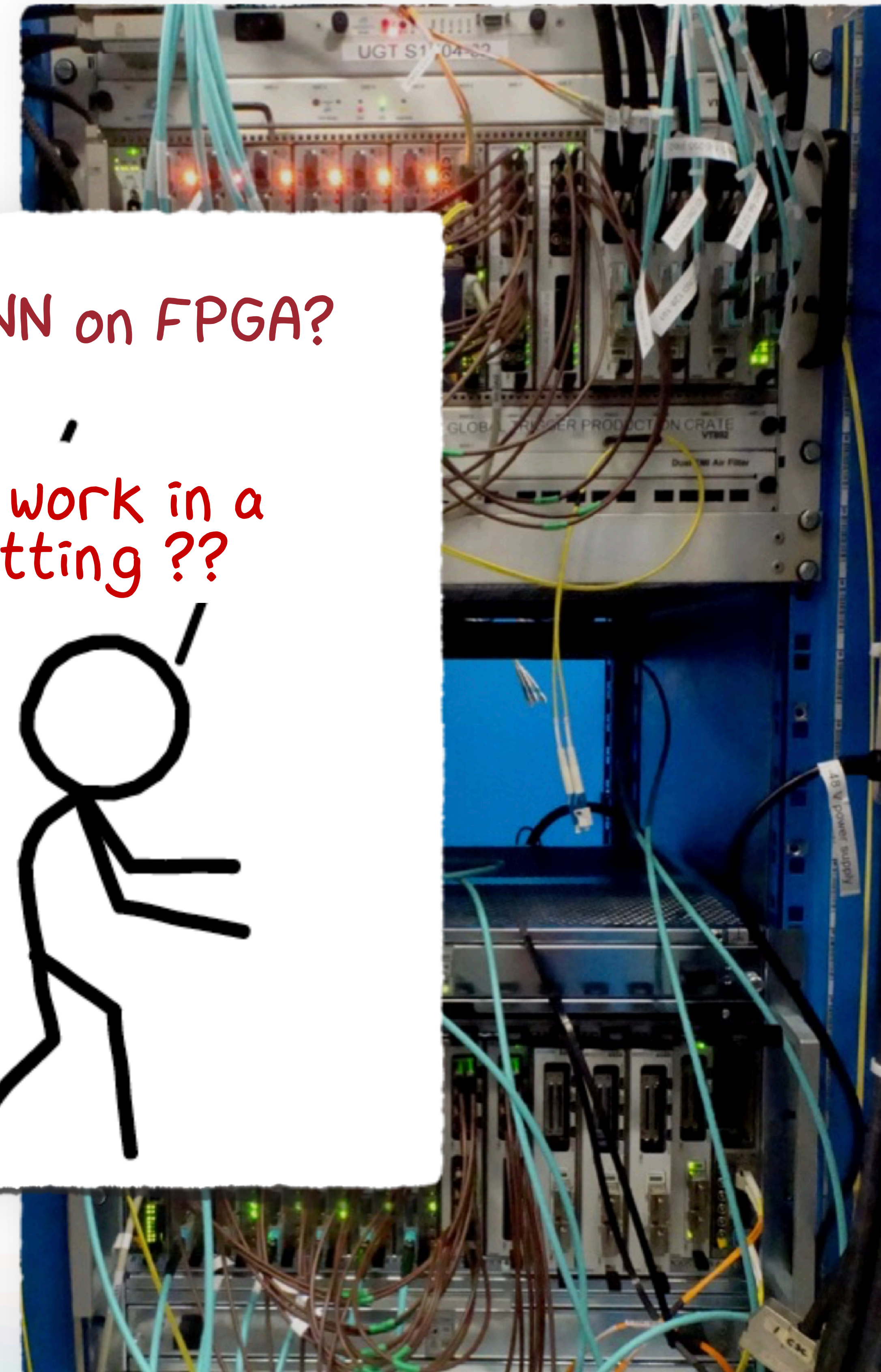
# Shrinking the Autoencoders

AXOL1TL

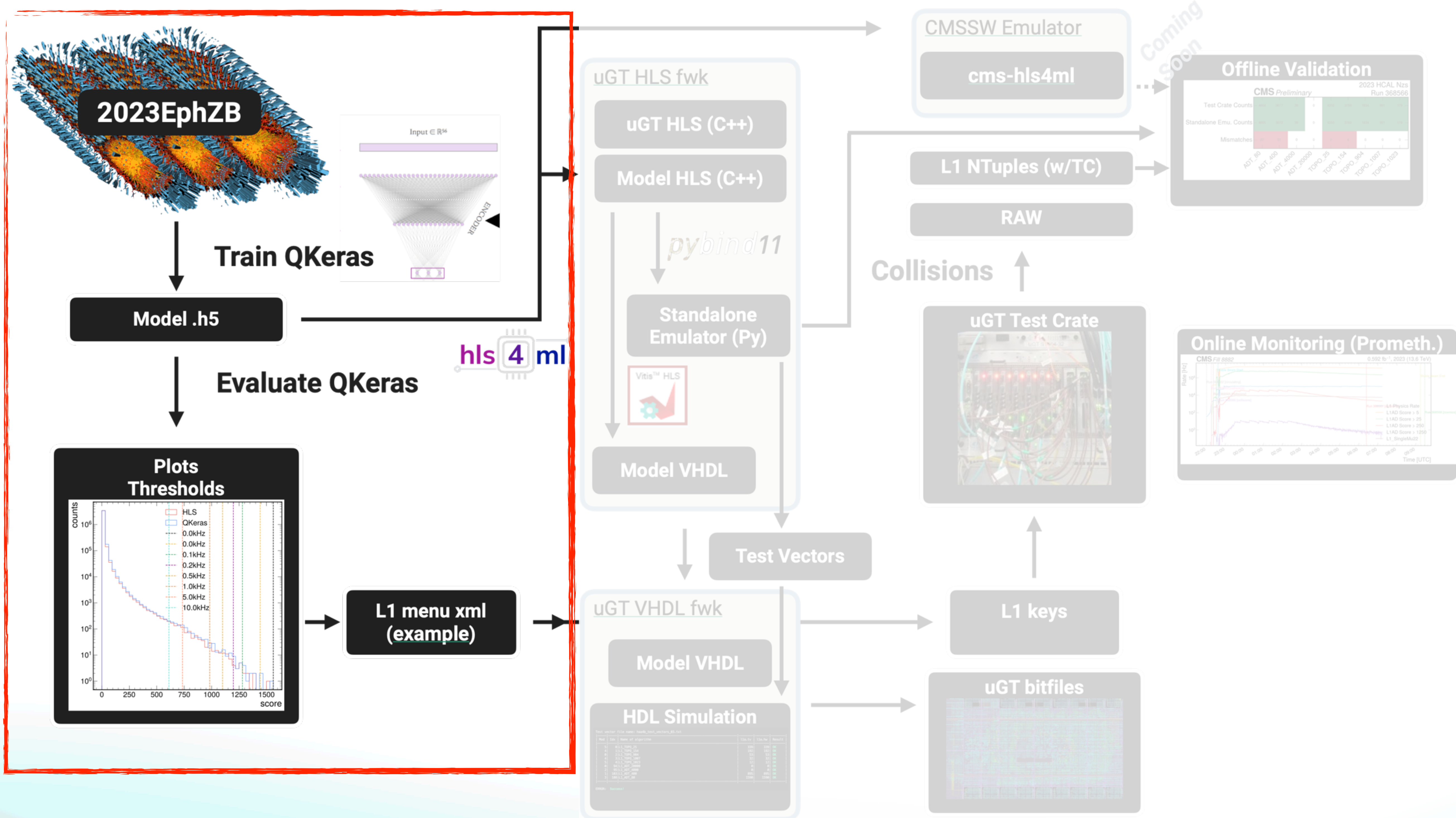
- Uses variational autoencoder
  - Deploy only the e



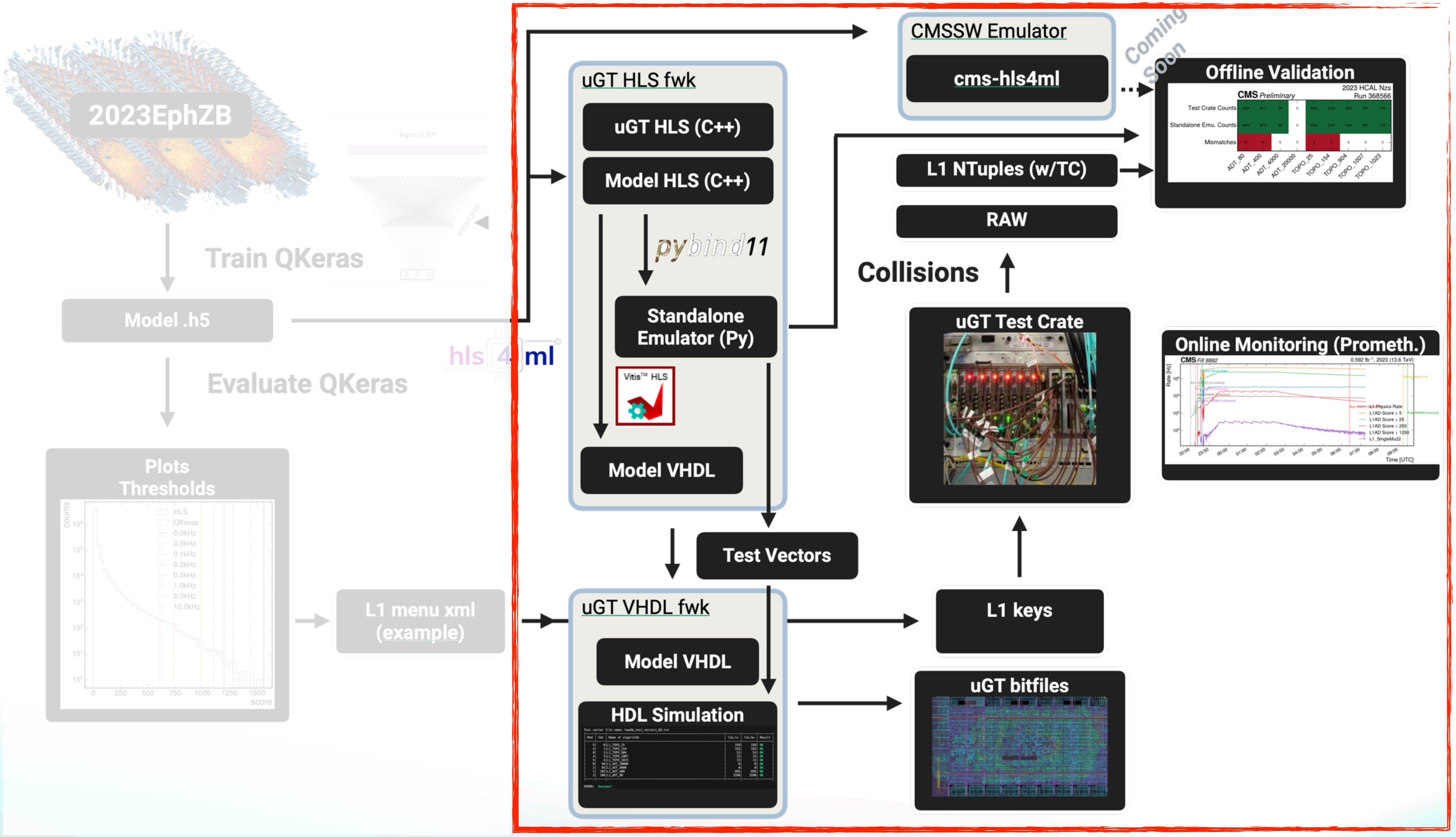

Anomaly Metric

# Algorithm development

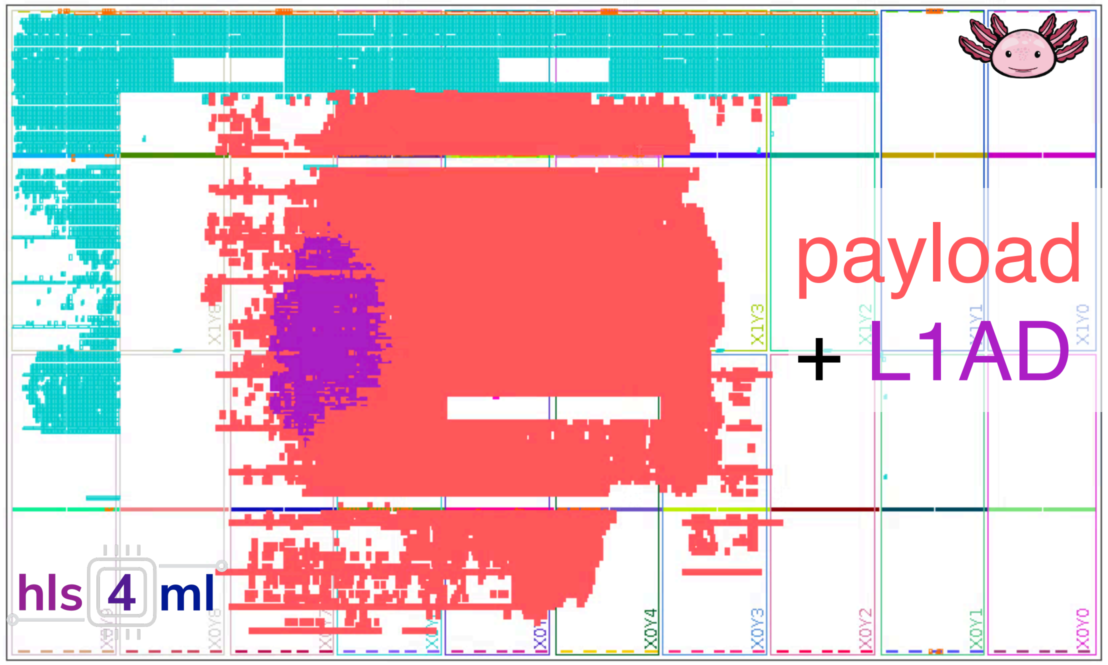


# Technical implementation



# AXOL1TL @ L1 Trigger

- Implementation of NNs @ L1 FPGA
  - Powered by **hls4ml**

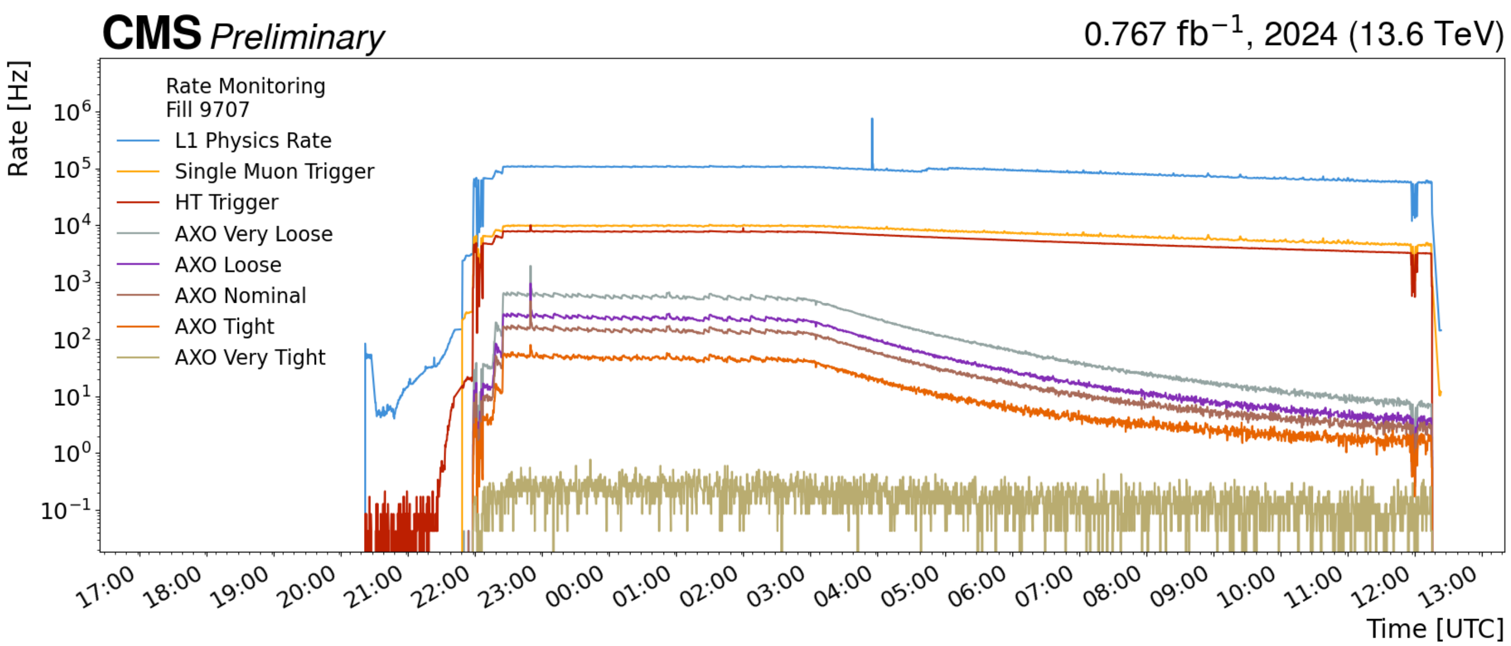



CMS-DP-2023-079

- On Xilinx Virtex-7 XCUP9P uGT boards
  - Needs ultra fast inference
    - Achieved inference time of 50 ns !

| | Latency | LUTs | FFs | DSPs | BRAMs |
|---|---|---|---|---|---|
| AXOL1TL | 2 ticks<br>50 ns | 2.1% | ~0 | 0 | 0 |

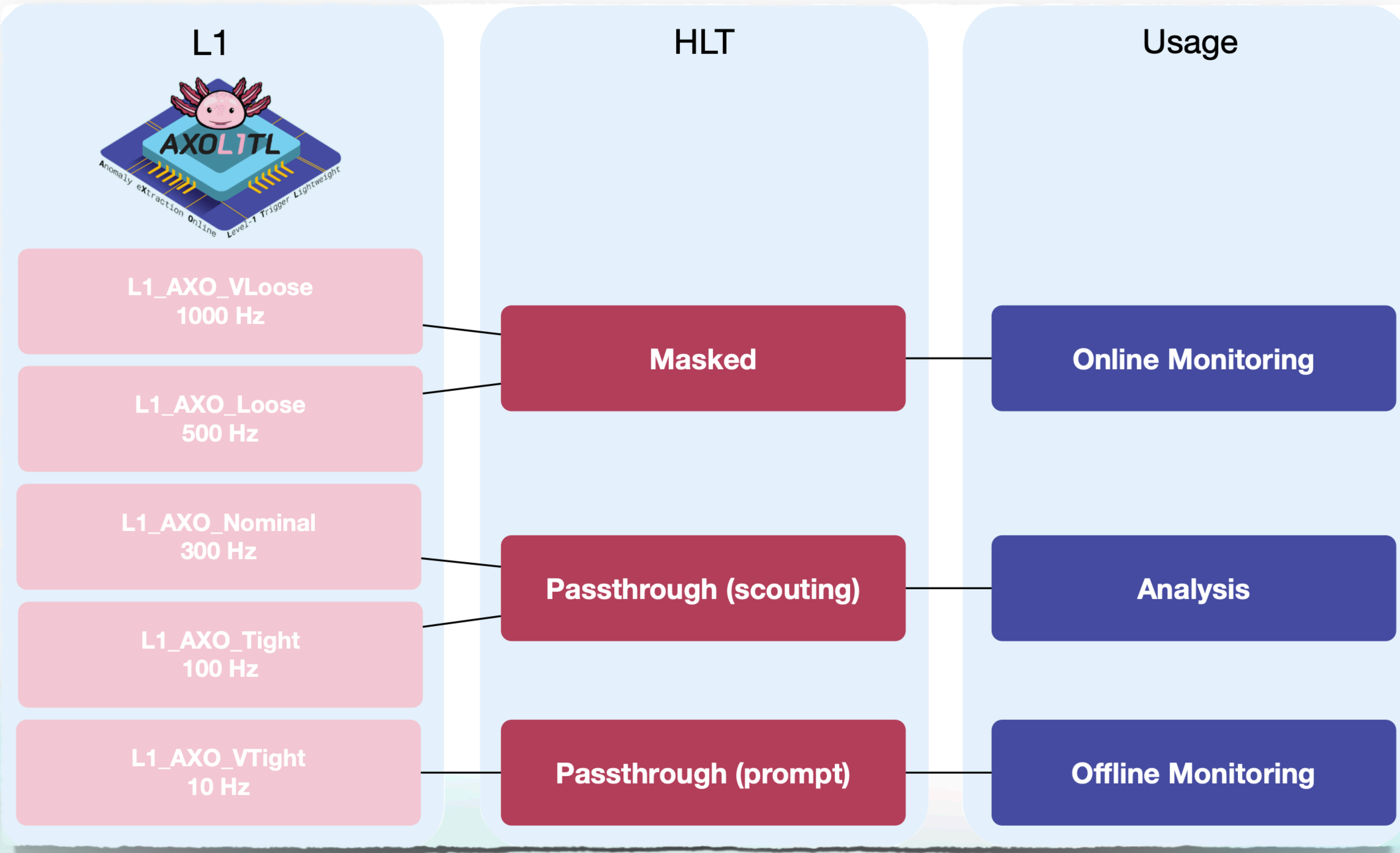
- **AXOL1TL** trigger is successfully integrated
- **AXOL1TL** is collecting data in CMS since last year !
  - Data collection started in 2024




CMS-DP-2024-059

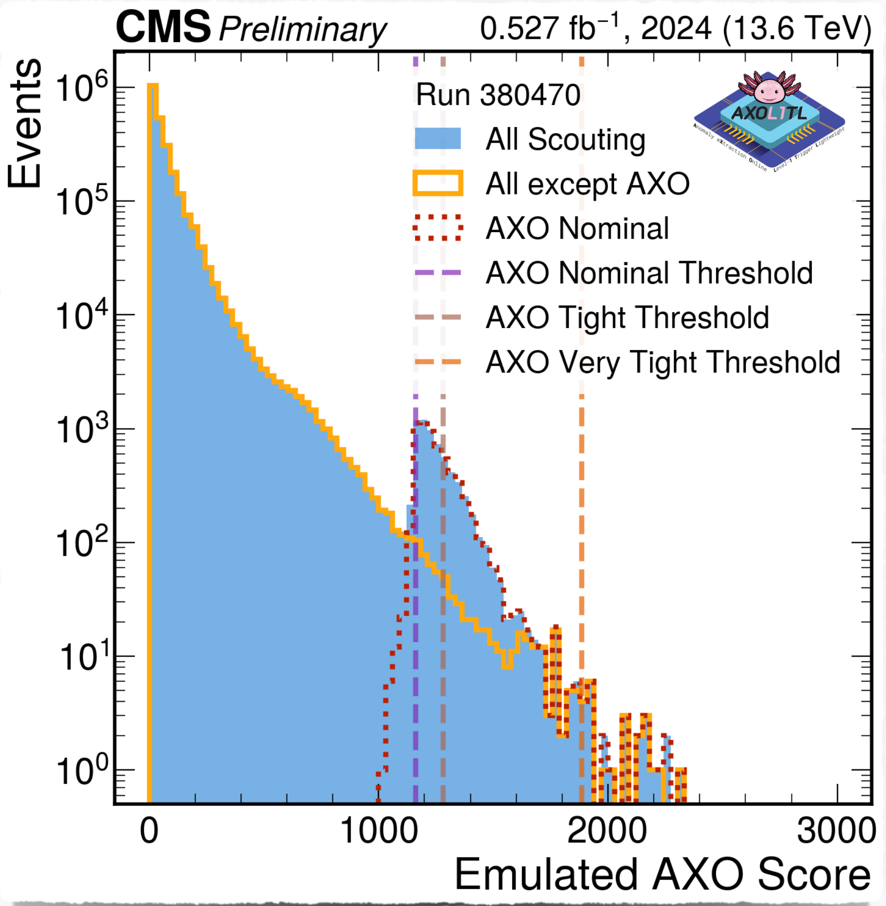
# Triggering strategy

- Triggers currently in place for AXOL1TL, will be shared with CICADA soon!
  - Collects high statistics samples of anomalous events in Run-3 !

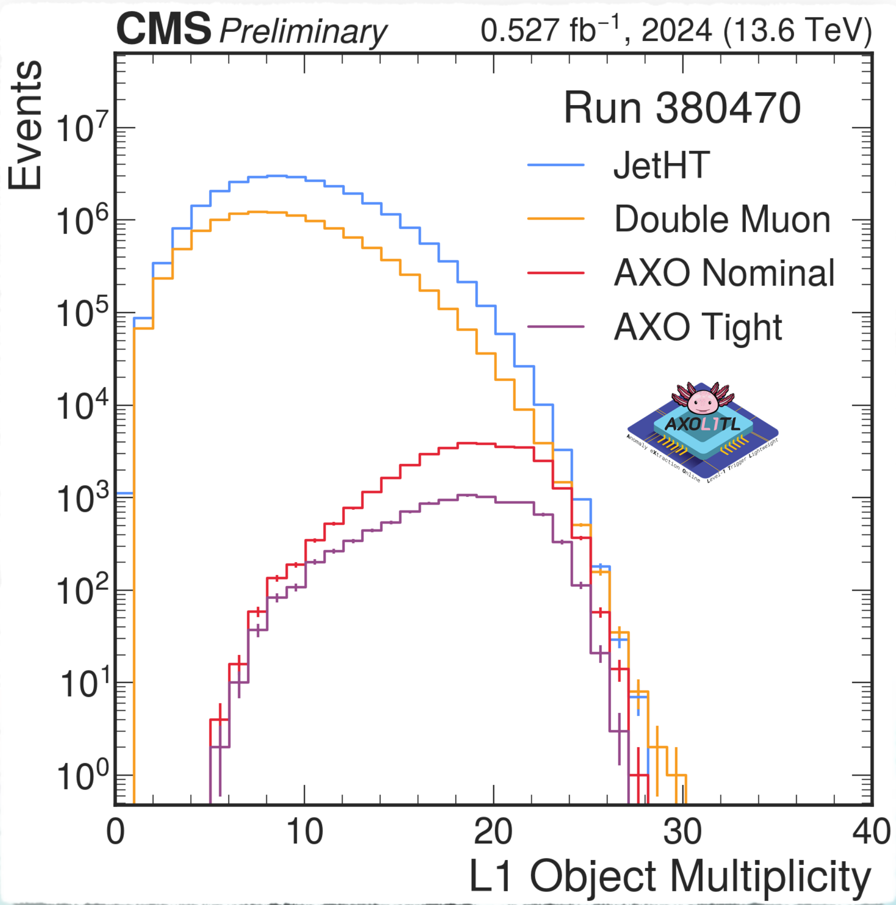


# Preliminary Insights

- *Purity of the Trigger:*
  - Many events triggered by AXOL1TL are unique
  - Mostly orthogonal to the current L1 Trigger menu



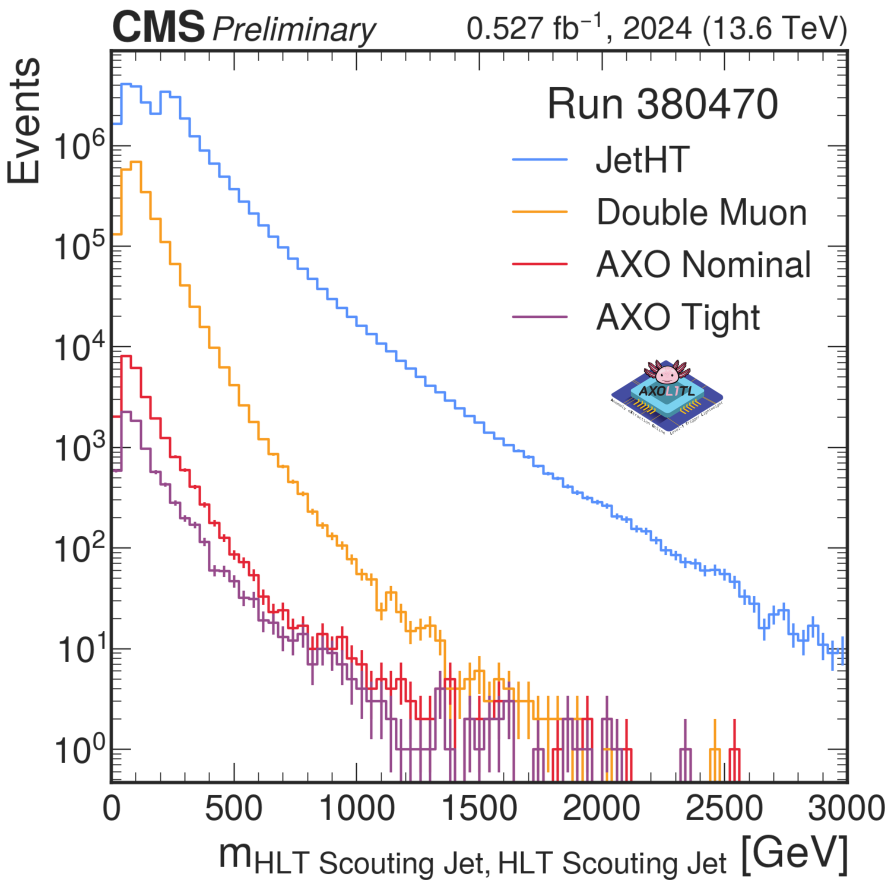

- *Features of events selected:*
  - Preference of events with high object multiplicity
  - Key in identifying certain BSM signatures (eg: SUEPs)

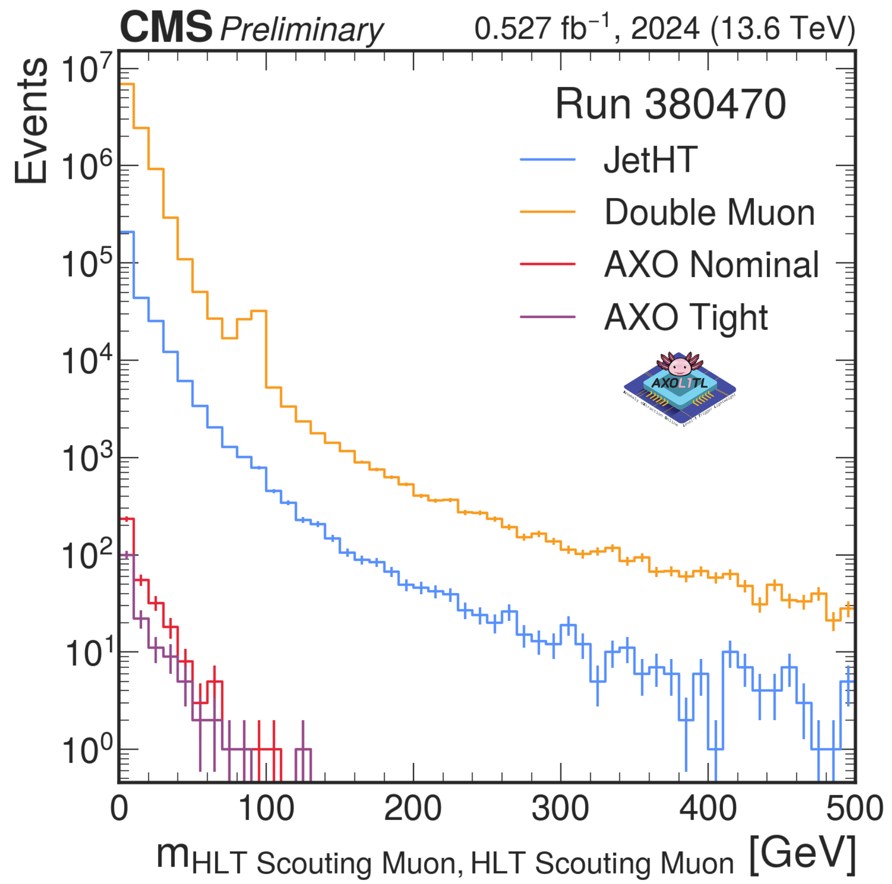


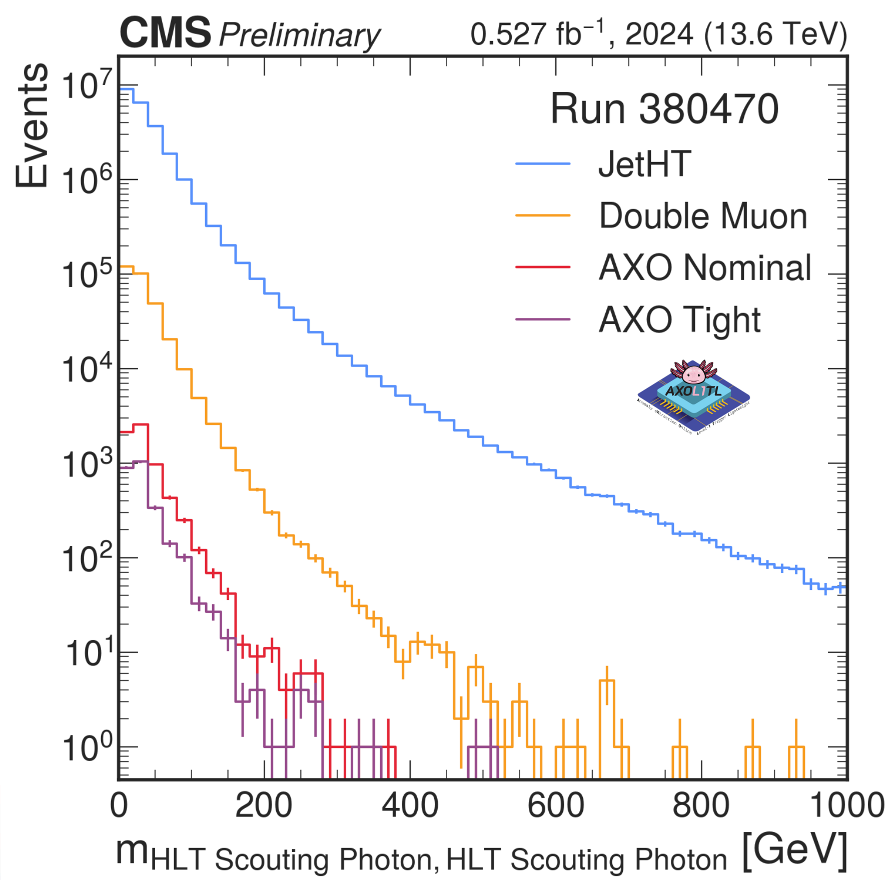

CMS-DP-2024-059

# Preliminary Insights

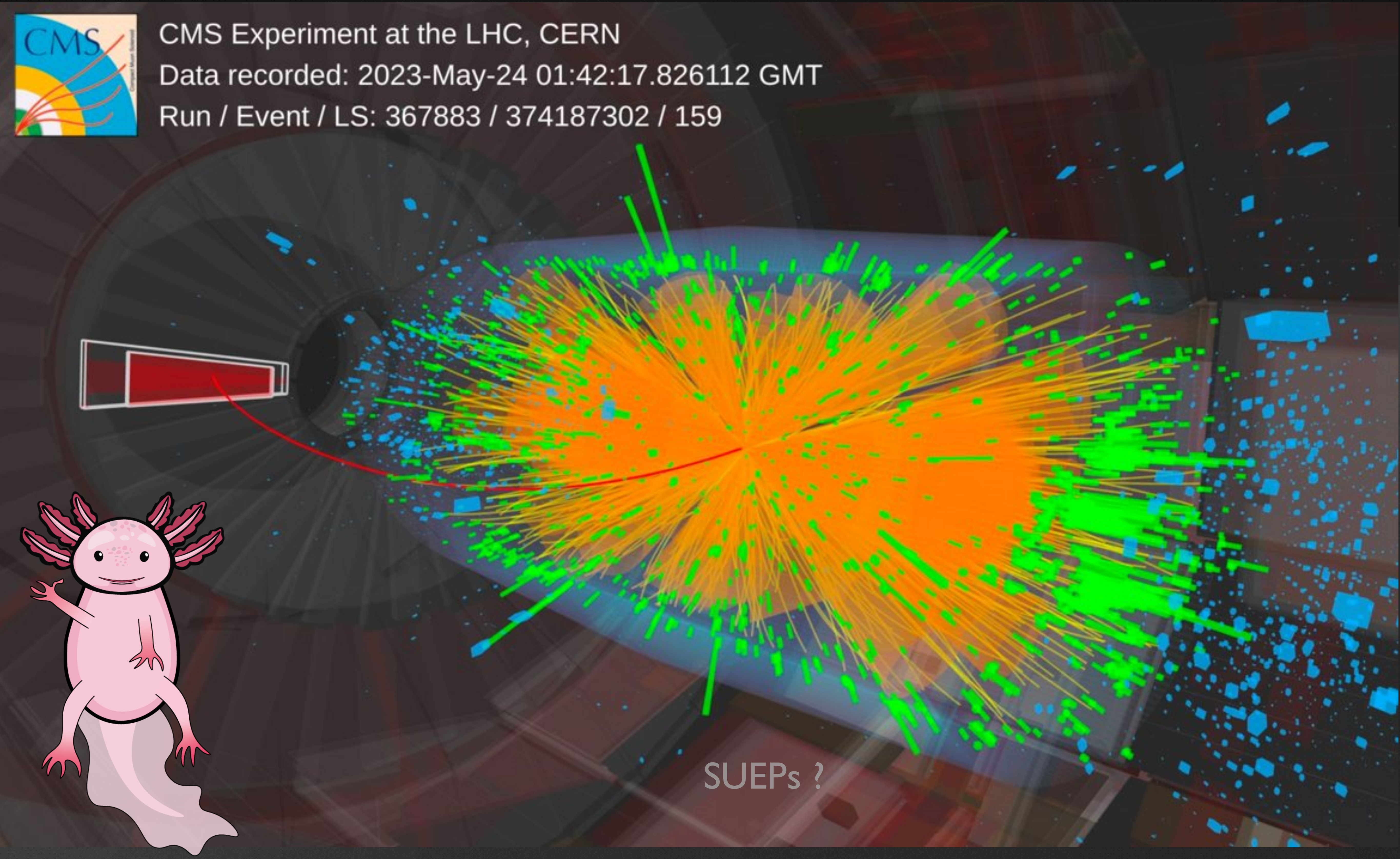
- Preliminary analysis performed on fraction of the dataset
  - *Distribution of search variables*
    - Invariant mass distributions of the objects recorded with Data Scouting
    - No sculpting and nominal mass distributions: Ideal for BSM (/ SM) searches



# How does anomalous event look like ?




CMS-DP-2023-079

# Summary

- Anomaly detection great at improving efficiency to all new physics, even at low rates
  - Cuts the rate down, but stays sensitive to BSM / Rare SM physics
  - Will significantly expand the reach of current L1 Menu
- Two complimentary approaches AXOL1TL and CICADA
  - Enabled by huge support from L1 Trigger community!
  - AXOL1TL is running on CMS Trigger and collecting data
- Demonstrated how to implement AD in a highly resource constrained environment
  - Rich insights for developing triggers for HL-LHC

***CMS taking bold steps and entering a new triggering paradigm***


Thank you !