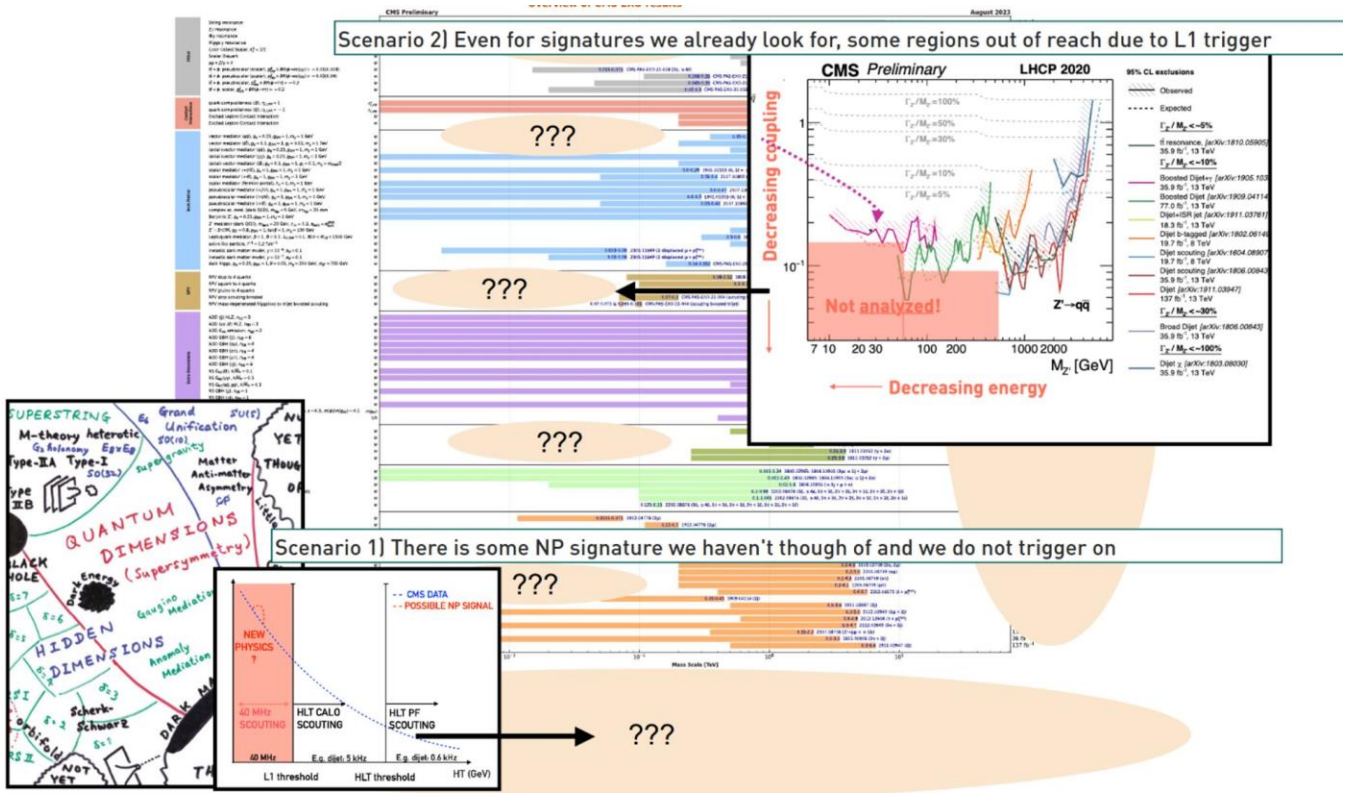
Anomaly Detection in the CMS Global Trigger Test Crate for Run 3

Chang Sun

on behalf of the CMS Collaboration

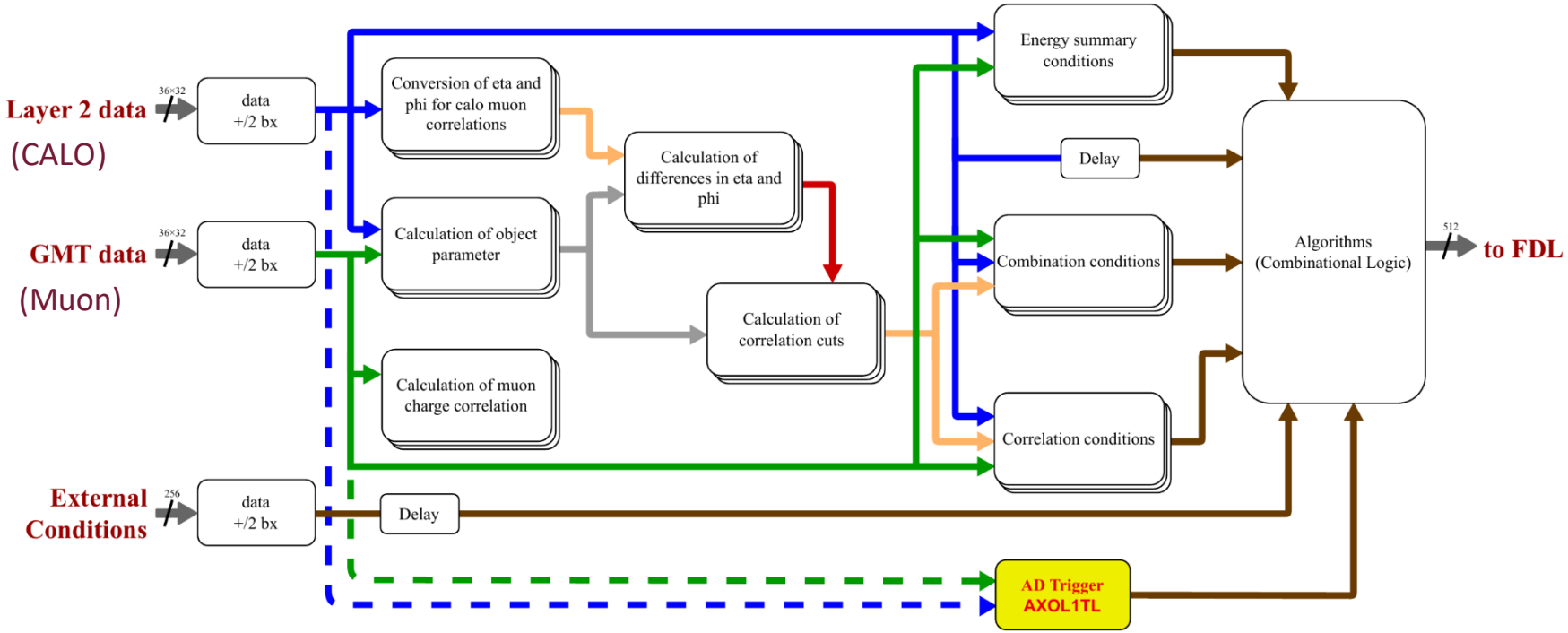
Why Anomaly Detection?

- At the CMS experiment, we need to discard ~99.8% of data in real time within micro seconds.
  - See Thea's talk this morning
- We have signals that are hard to trigger on with the current trigger system.
- The signature for new physics is unknown.



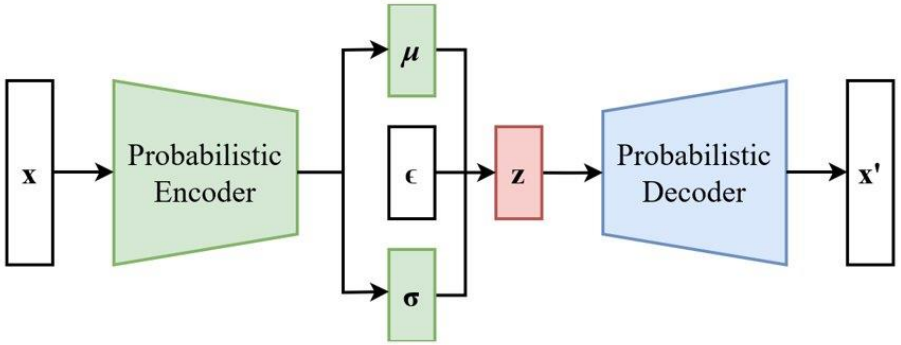
Our Solution

- We propose a new ML-based L1 global trigger at the CMS experiment for new physics detection, AXOL1TL



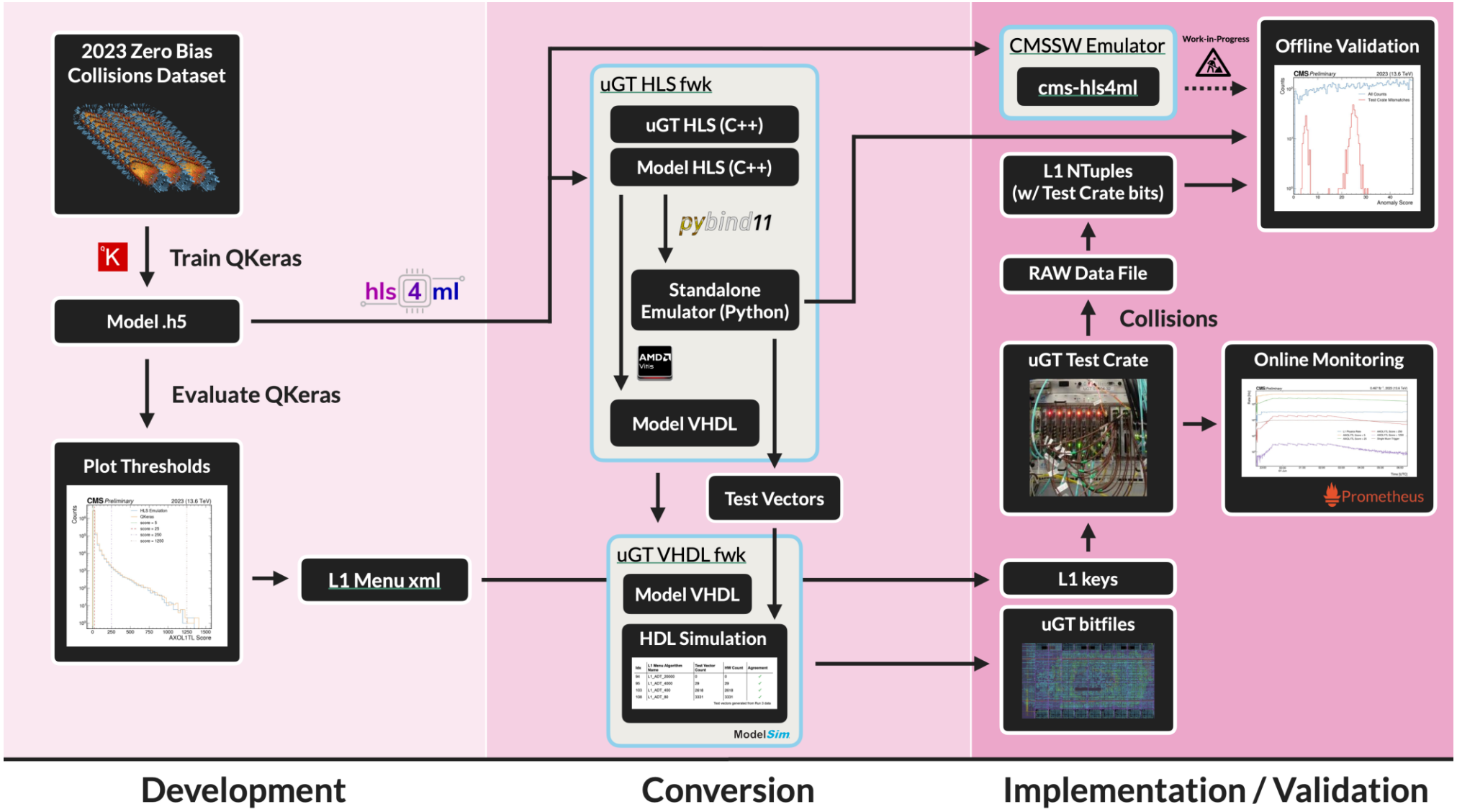
Schematic the CMS Global Trigger

Our Solution

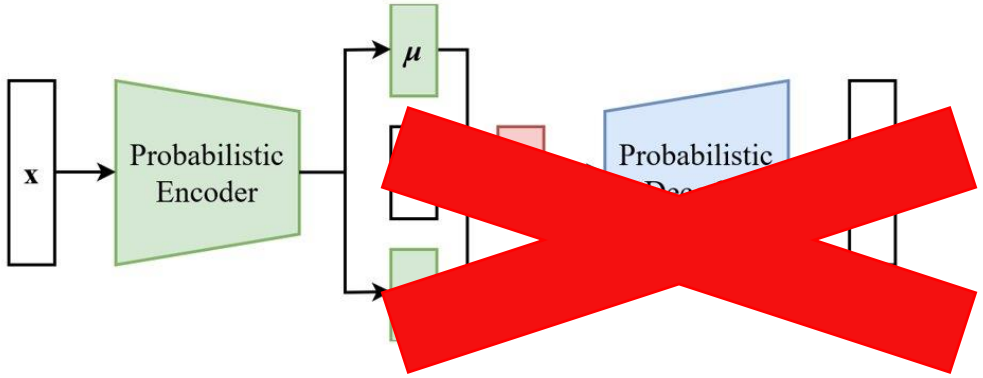


- Use a Variational Autoencoder (VAE) based trigger algorithm, AXOL1TL, to select anomalous events in real-time
  - Sensitive to a wide variety of signals
  - VAE is a common method for anomaly detection
- VAE trained on ZeroBias data to detect data outliers
  - ZeroBias data: minimally biased, unfiltered data from the detector
- Firmware generation with hls4ml+vivado toolchain
- Standalone HLS emulation of Anomaly Detection output for validating performance

Workflow


2023 Zero Bias Collisions Dataset
Train QKeras
Model .h5
Evaluate QKeras
Plot Thresholds
L1 Menu xml
hls 4 ml
uGT HLS fwk
uGT HLS (C++)
Model HLS (C++)
pybind11
Standalone Emulator (Python)
AMD Vitis
Model VHDL
Test Vectors
uGT VHDL fwk
Model VHDL
HDL Simulation
ModelSim
CMSSW Emulator
cms-hls4ml
Work-in-Progress
Offline Validation
L1 NTuples (w/ Test Crate bits)
RAW Data File
Collisions
uGT Test Crate
Online Monitoring
Prometheus
L1 keys
uGT bitfiles
Development
Conversion
Implementation / Validation

Anomaly metric



- The loss for a VAE is

$$\text{Loss} = (1-\beta)||x-\hat{x}||^2 + \beta \cdot \frac{1}{2}(\mu^2 + \sigma^2 - 1 - \log\sigma^2)$$

Reconstruction term KL regularization term

- Observe the KL term is a good anomaly metric, skip the decoder:

$$2 \cdot \text{KL} = \mu^2 + \sigma^2 - 1 - \log\sigma^2$$

- As the σ term is insignificant compare to the μ term, skip the σ term:

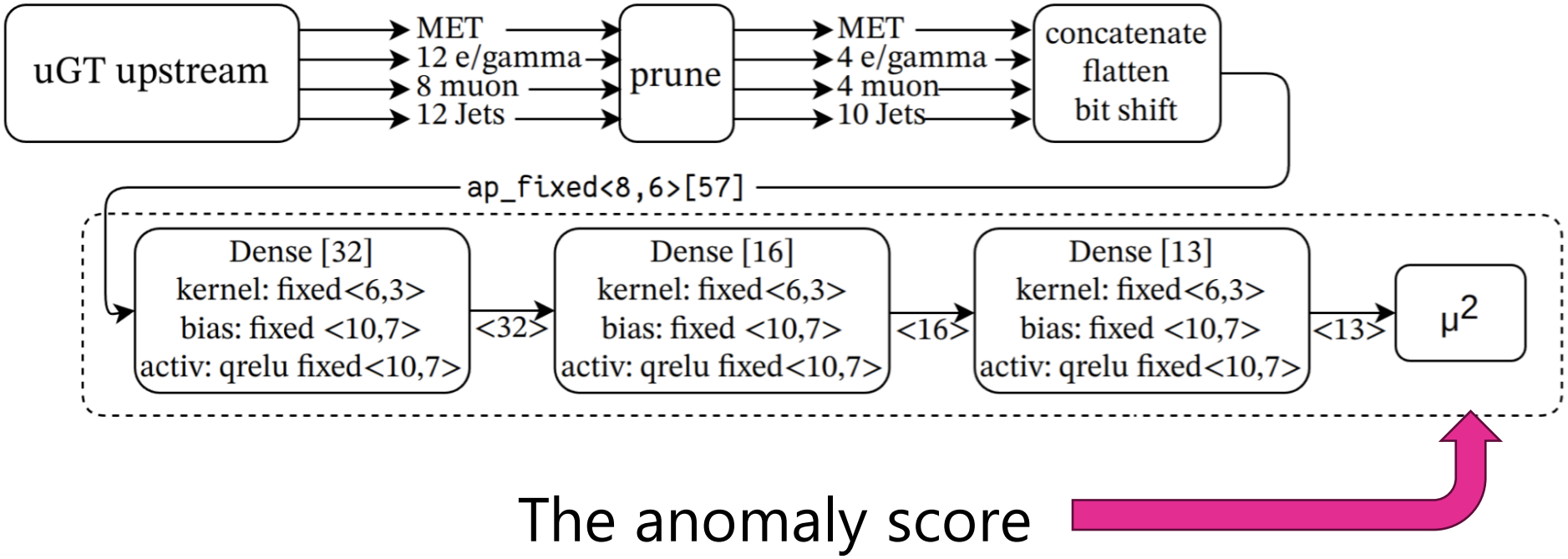
$$2 \cdot \text{KL} = \mu^2 + \cancel{\sigma^2 - 1 - \log\sigma^2}$$

- These modifications reduce resource and latency of the network on FPGAs
  - Remove the decoder
  - Skip the slow exponential/logarithm operation

Dataset

- The model is trained with ZeroBias data collected by the CMS Experiment during 2023 with √s=13.6 TeV
  - Using the first 4 e/gamma, 4 muons, 10 jets, and missing ET objects as inputs
  - Each object has pt, eta, phi in hardware integers
  - 10.5 million events are used
    - ~50% for training
    - other ~50% for setting anomaly thresholds based on the score distribution
- Preprocessed by scaling with bit-shift and clipping extreme values.
- Signal samples are Monte-Carlo generated.

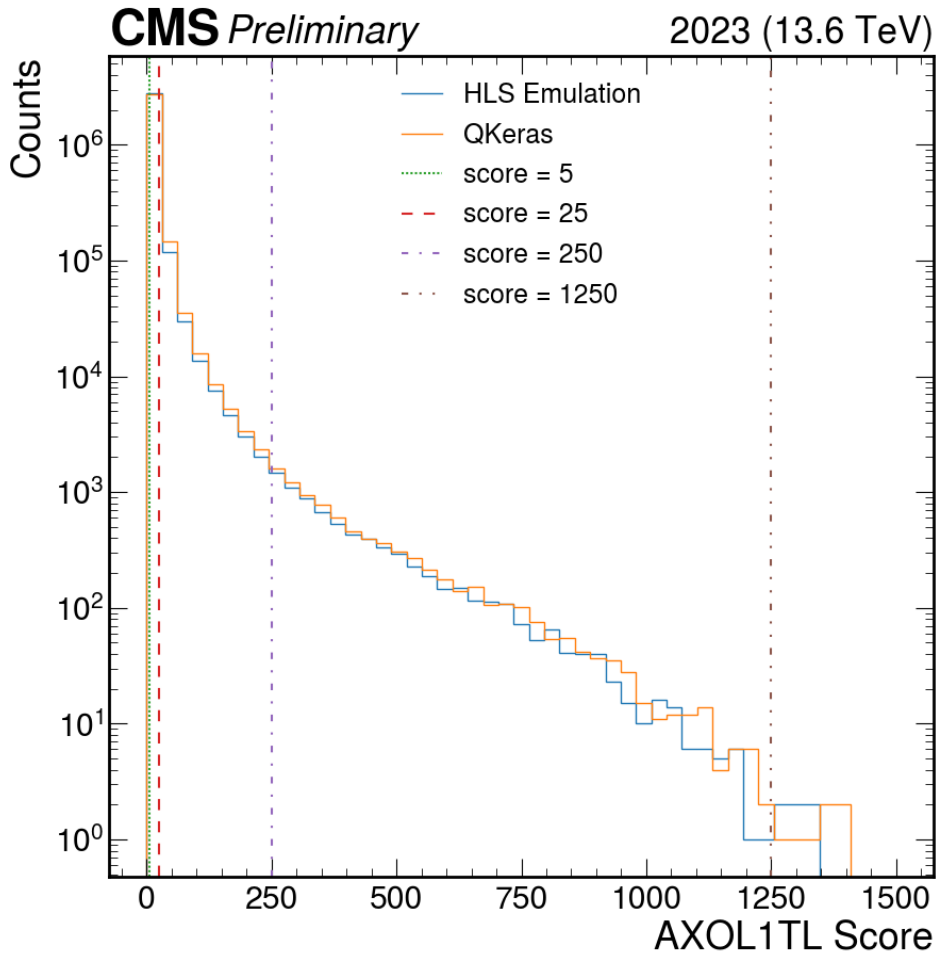
Architecture



The anomaly score

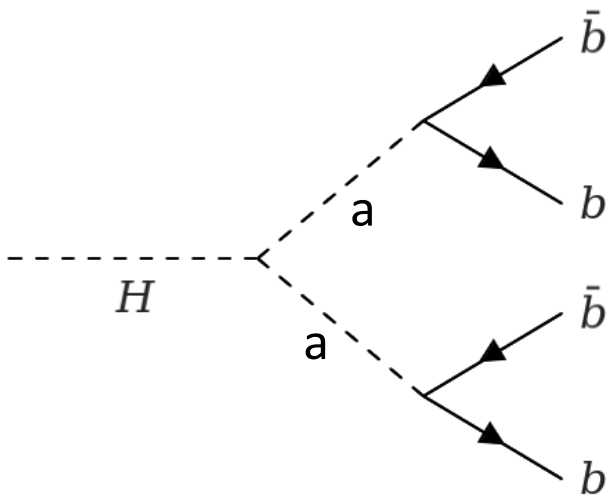
Anomaly Score and Thresholds

- Anomaly score distributions for 2023 ZeroBias events.
- Event scores for the QKeras model and standalone HLS emulator are consistent.
- Dotted lines represent the score thresholds implemented in the µGT trigger test crate.



Physics Performance

- We observed significant performance improvement on various SM and BSM signals by adding AXOL1TL to the 2023 trigger menu
  - $\text{Improvement} = \frac{\text{L1 Efficiency w/ AXOL1TL@freq}}{\text{L1 Efficiency w/o AXOL1TL}} - 1$
- Shown in the table is the performance improvement for a Higgs decaying to 2 (pseudo-)scalars to bottom quarks.
  - H->aa[15 GeV]->4b

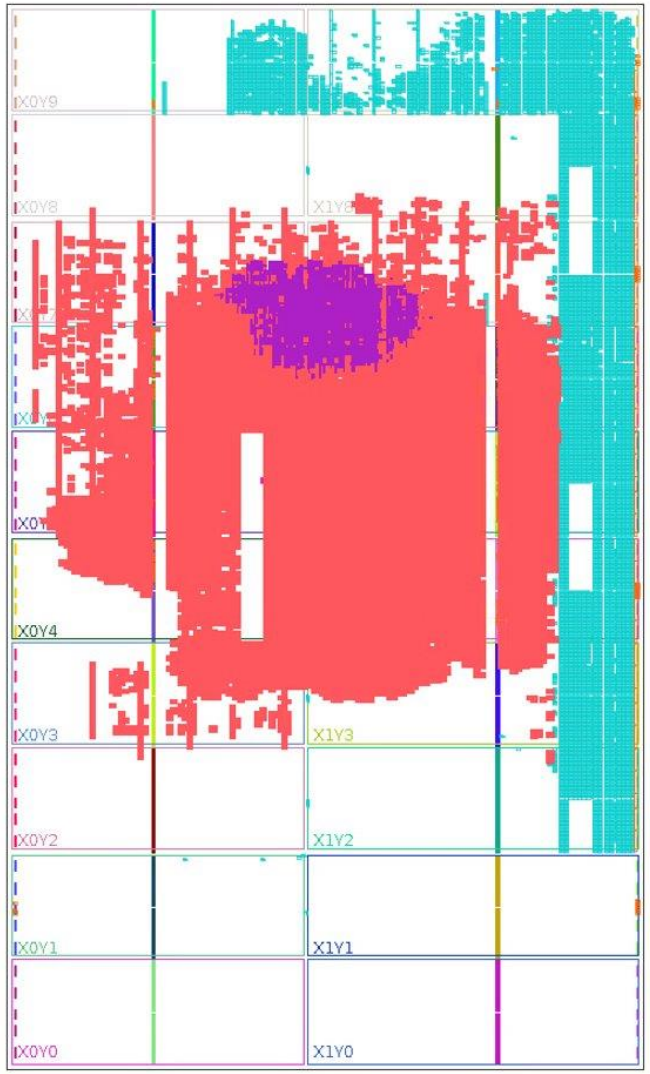


AXOL1TL Rate	1 kHz	5 kHz	10 kHz
Signal Efficiency Gain	46%	100%	133%

Firmware Validation

- Xilinx Virtex-7 XCVU9P is used
  - 1728000 LUTs, 3456000 FFs, 12288 DSPs,
- No timing violation
- Resource and latency requirement met
  - 2 ticks@40MHz w/ initialization interval=1
  - LUT usage of O(1%)

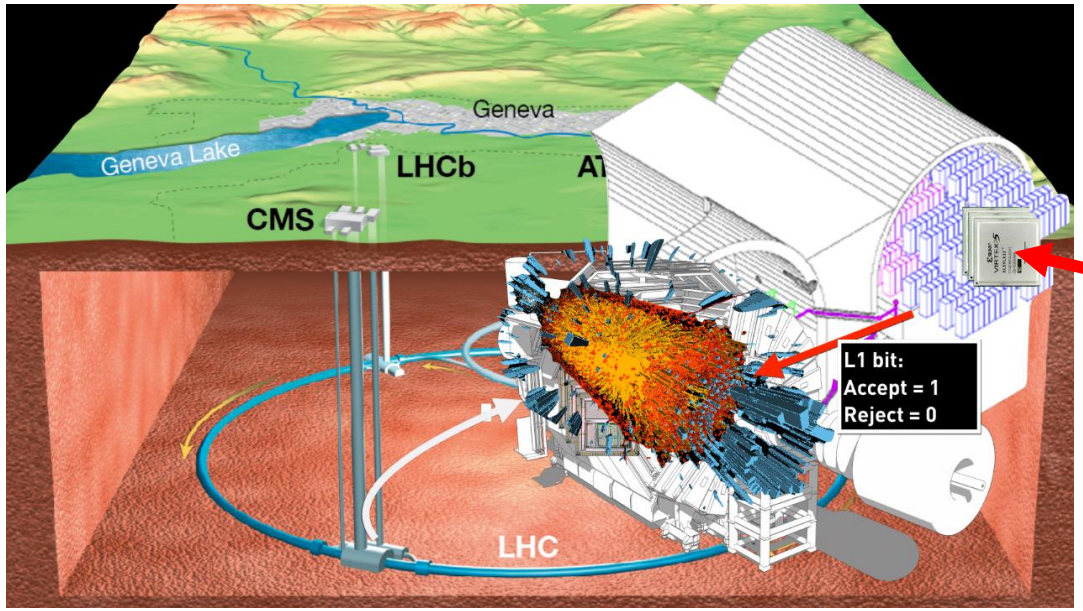
	Latency	LUTs	FFs	DSPs	BRAMs
AXOL1TL	2 ticks 50 ns	2.1%	~0	0	0

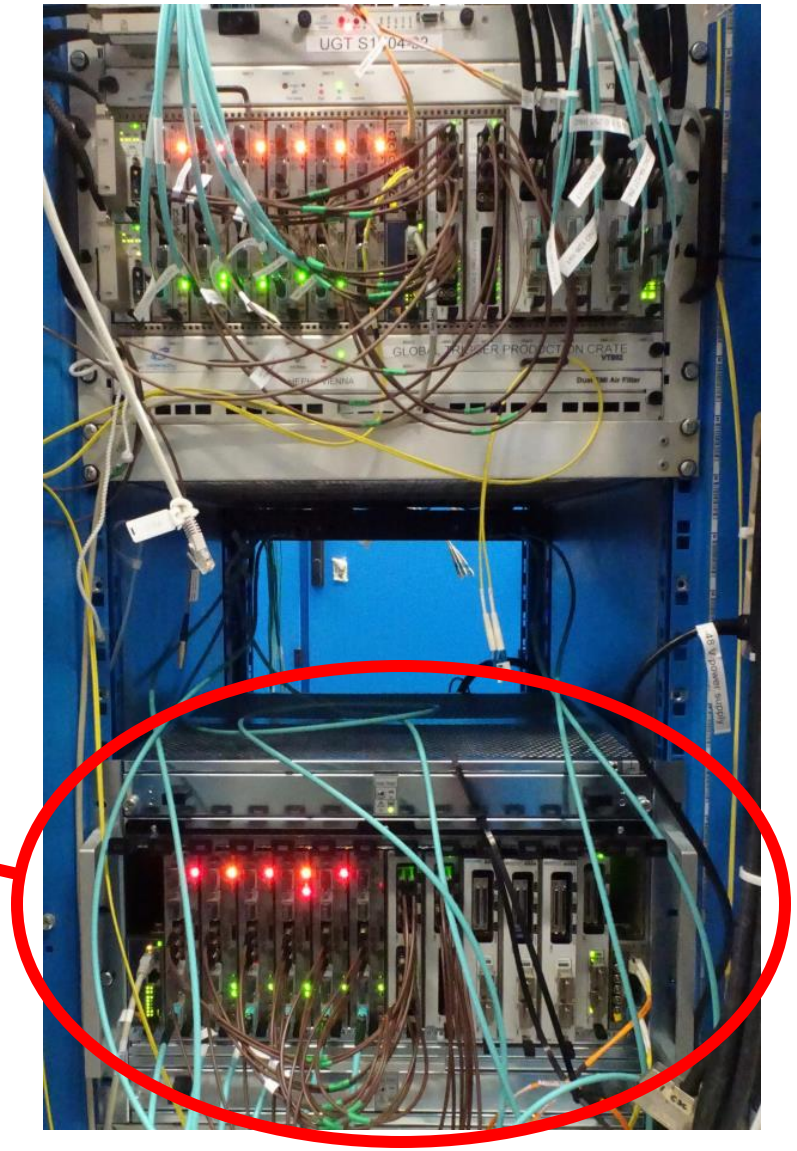


hls 4 ml

Test Crate Monitoring

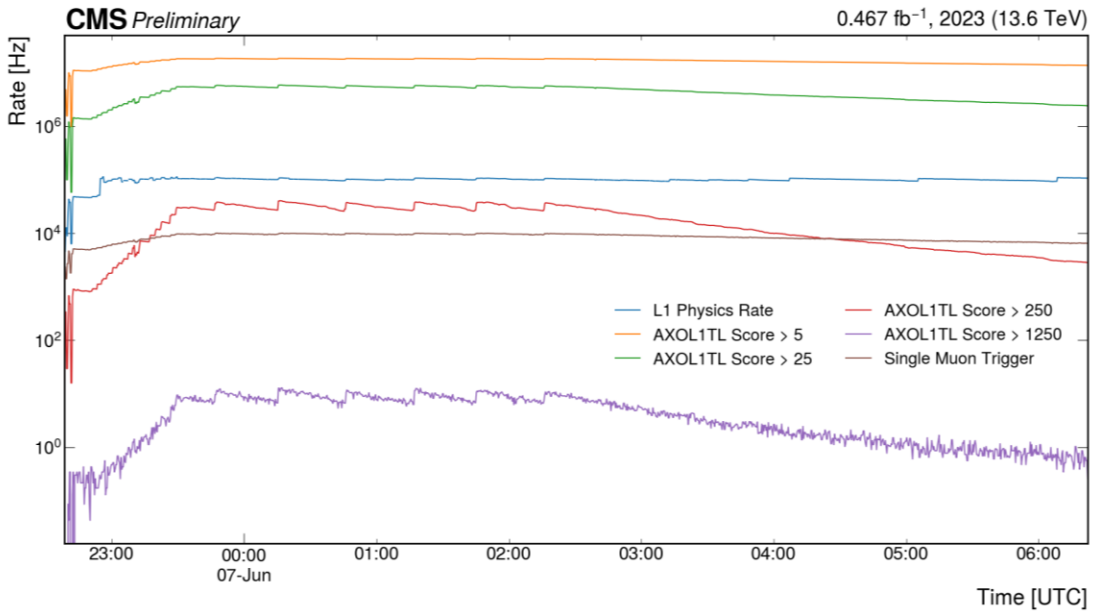
- An earlier version of AXOL1TL is installed in the test crate
  - Trained on 2018 ZeroBias data





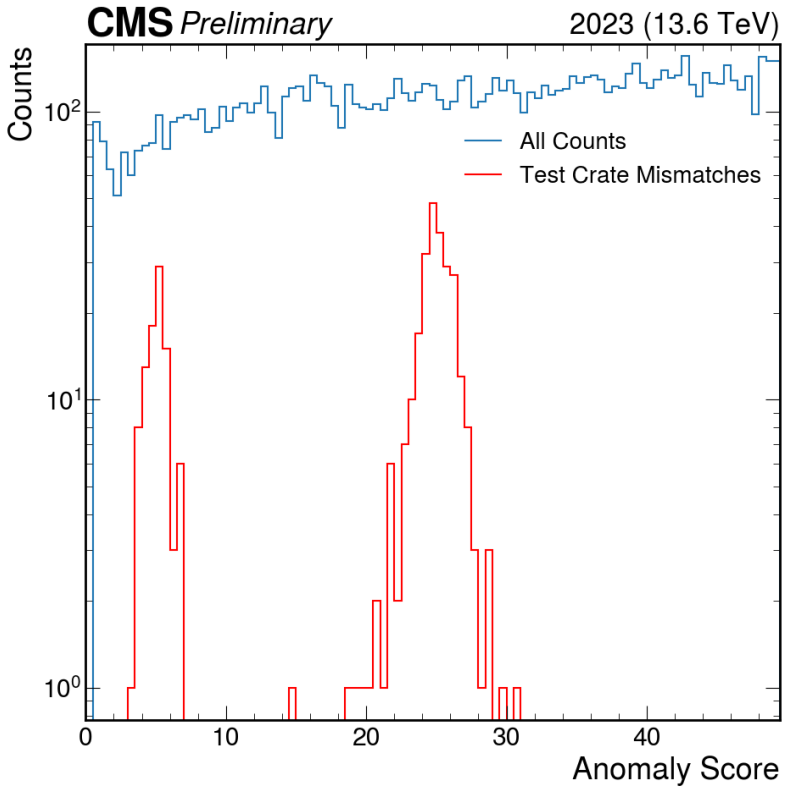
Test Crate Monitoring

- An earlier version of AXOL1TL is installed in the test crate
  - Trained on 2018 ZeroBias data
- The rates shown are
  - AXOL1TL with 4 different thresholds, selected for testing purpose only. **They are not meant to be model realistic trigger rates.**
  - Single muon trigger
  - Total L1 physics rate
- Consistent performance observed through the partial fill cycle.
  - Initial rise-up corresponds to the beginning of a LHC fill
  - sawtooth pattern corresponds to luminosity leveling



Test Crate Validation

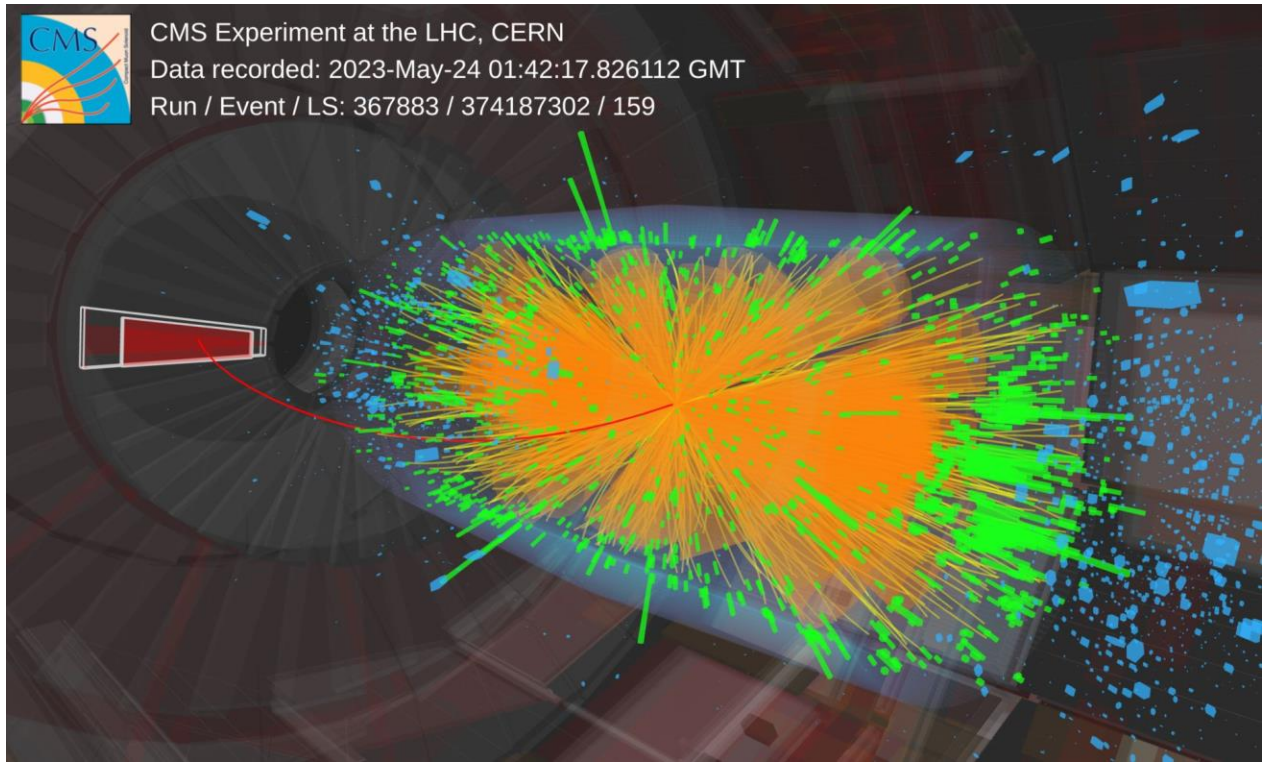
L1 Menu Algorithm Name	Test Crate Count	Standalone Emulator Count	Mismatches
L1_ADT_20000	1	1	0
L1_ADT_4000	742	741	19
L1_ADT_400	21236	21229	253
L1_ADT_80	25468	25481	93



- Test Crate Count
  - from configured μGT test crate.
- Standalone Emulator Count
  - From offline inference with reconstructed L1 objects
- Minimal mismatches present (≤ 1%)
- Events are from promptly reconstructed 2023 ZeroBias data where trigger bits are recorded
- On the figure, red bins are mismatched events. As mismatches are concentrated around decision boundaries, rounding/precision issues are suggested.

Event Display

- Shown is the event with highest anomaly score that is not triggered by the normal L1 menu.
  - Ephemeral Zero Bias 2023 Run 367883.
- This is a very busy event with respect to the given the run condition



- At L1, this event has 12 jets (of which 11 has $E_T > 20$ GeV), and 1 muon of 3 GeV.
- Offline construction identifies 7 jets (PUPPI algorithm) with $p_T > 15$ GeV, and 1 muon.
- Has 75 reconstructed vertices
  - Given the pile-up profile of the data taken in Run 2 and Run 3, this is highly unlikely.

Conclusion

- We demonstrated that our VAE-based model AXOL1TL is performant in anomaly detection as an L1 physics trigger.
- Our model meets the timing and resource requirements imposed by the L1 trigger system. A working copy of the model is installed in the test crate at the CMS Experiment.
- Further integration with the L1 trigger system is underway.

Backup: Test Crate Firmware Validation

- Number of trigger issued with four thresholds
  - (1250, 250, 25, 5)
- Test Vector Count
  - By standalone emulator
- HW Count
  - By standard global trigger firmware simulation workflow with ModelSim

Idx	L1 Menu Algorithm Name	Standalone Emulator Count	HW Count	Agreement
94	L1_ADT_20000	0	0	✓
95	L1_ADT_4000	29	29	✓
103	L1_ADT_400	2618	2618	✓
108	L1_ADT_80	3331	3331	✓

Test vectors generated from Run 3 data

Backup: References

- CMS Collaboration. "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter". CERN-LHCC-2012-015, CMS-TDR-10 (2012). https://cds.cern.ch/record/1481837.
- M. Jeitler, et al. "The level-1 global trigger for the CMS experiment at LHC". JINST 2, P01006 (2007). https://doi.org/10.1088/1748-0221/2/01/P01006.
- E. Govorkova, et al. "Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider". Nat. Mach Intell. 4, 154 (2022). https://doi.org/10.1038/s42256-022-00441-3.
- FastML Team. hls4ml (Version v0.7.1) [Computer software]. https://doi.org/10.5281/zenodo.1201549.
- J. Duarte, et al. "Fast inference of deep neural networks in FPGAs for particle physics". JINST 13, P07027 (2018). https://doi.org/10.1088/1748-0221/13/07/P07027.
- Xilinx Virtex-7 FPGA. https://www.xilinx.com/products/silicon-devices/fpga/virtex-7.html.
- L1 Menu Repository. https://github.com/herbberg/l1menus.