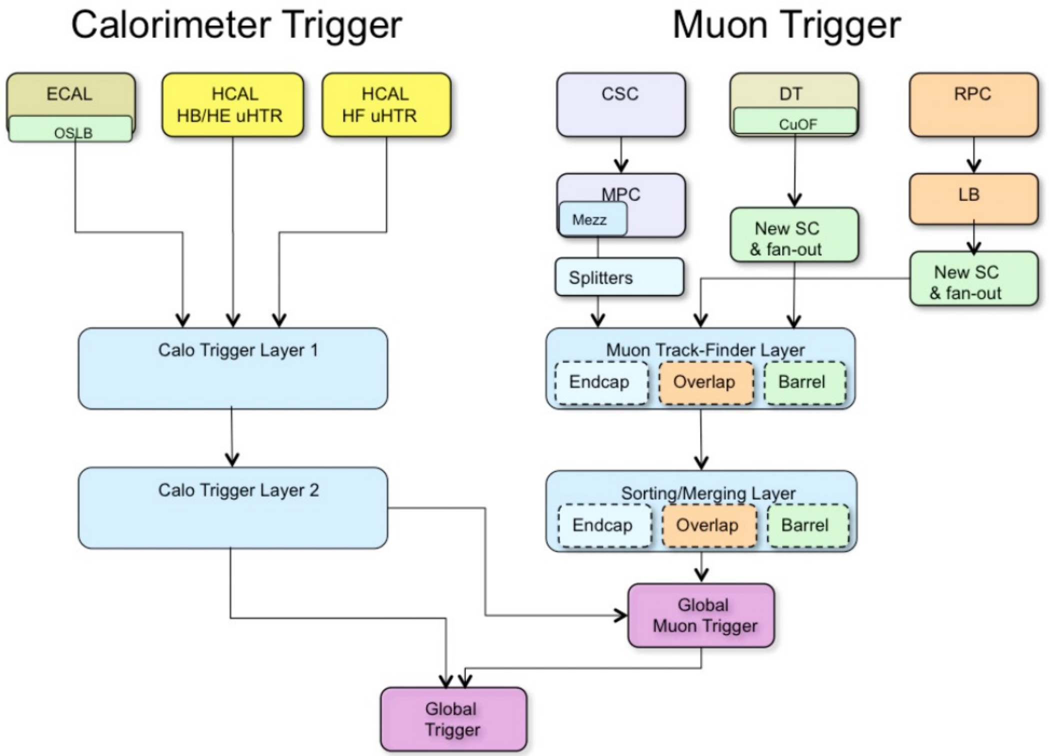
Topical Workshop on Electronics for Particle Physics 2023

Anomaly Detection at the CMS Level-1 Trigger

Noah Zipper on behalf of the CMS Collaboration

The CMS Level-1 Trigger

- CMS produces more data than we can handle
  - Terabytes per second from front-end electronics
  - Need to reduce by >99%
  - **Trigger's challenge is to keep interesting physics**
- Real-time decisions for what to keep
  - Built on Field Programmable Gate Array (FPGA) hardware chain
  - Collisions every 25 nanoseconds mean microsecond latency constraints
- Stability is crucial
  - Errors lead to trigger "dead time" ⇒ lost data
- Experimentation is encouraged
  - Phase-1 flexibility allowed early adoption of new trigger ideas
  - 6 Global Trigger production boards + **6 for testing**



Why Anomaly Detection?

Problem:

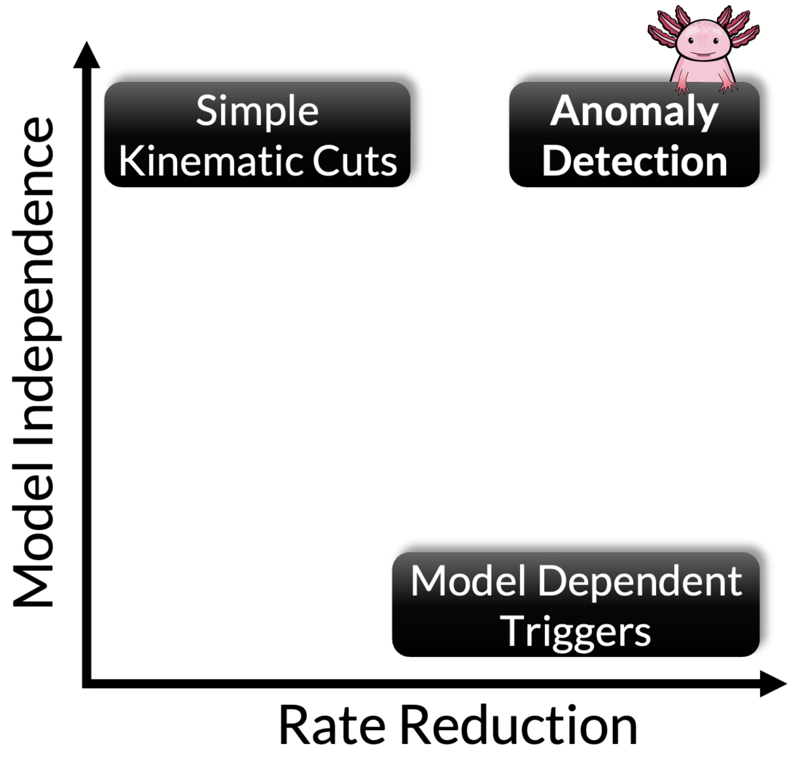
Traditional trigger strategies rely on a priori knowledge of signal or generic kinematic selections.

What if we miss new physics because we don't have the right trigger?

Solution:

Triggering on "anomalousness" offers an answer that is both

1. Signal agnostic - Applicable to signatures that we have not had the foresight or person-power to target specifically
2. Highly sensitive – Can boost signal efficiency to signatures limited by L1 trigger bandwidth

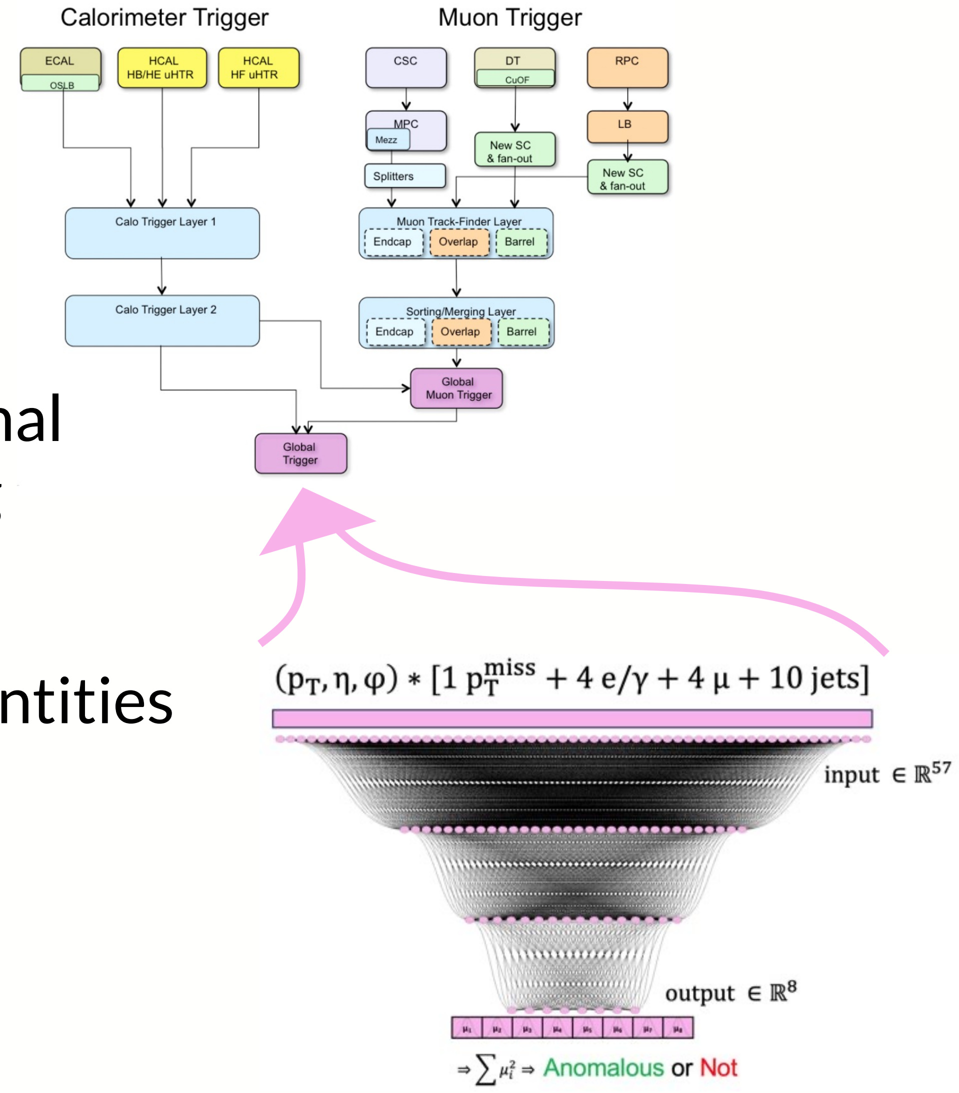


What is AXOL1TL? Anomaly eXtraction Online Level-1 Trigger aLgorithm

- Variational autoencoder (VAE) trained on real unbiased data to detect outliers

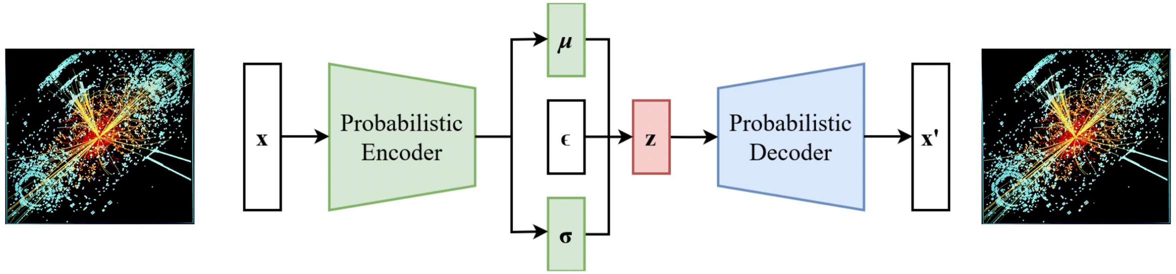
- Information bottleneck created by small-dimensional latent space enforces efficient encoding ⇒ learning

- Calculated from standard Global Trigger (μGT) quantities
  - (pT, η, ϕ) hardware integer inputs from: 1 p_T^{miss}, 4 e/γ, 4 μ, and 10 jets



Model Design

Level-1 Trigger constraints informed design

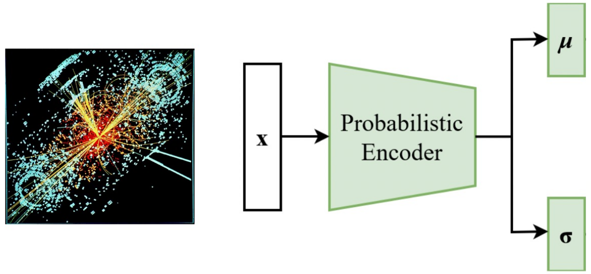


$$\text{Loss} = (1-\beta)||x-\hat{x}||^2 + \beta\frac{1}{2}(\mu^2+\sigma^2-1\ -\log\sigma^2)$$

Reconstruction term Full regularization term

Model Design

Level-1 Trigger constraints informed design



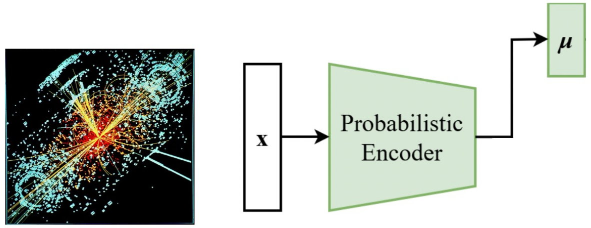
$$\text{Loss} = \cancel{(1-\beta)\|x-\hat{x}\|^2} + \beta\,\frac{1}{2}\left(\mu^2+\sigma^2-1\;-\log\sigma^2\right)$$

Reconstruction term | Full regularization term

- Remove decoder network
  - Significant latency & resource savings, minimal performance degradation

Model Design

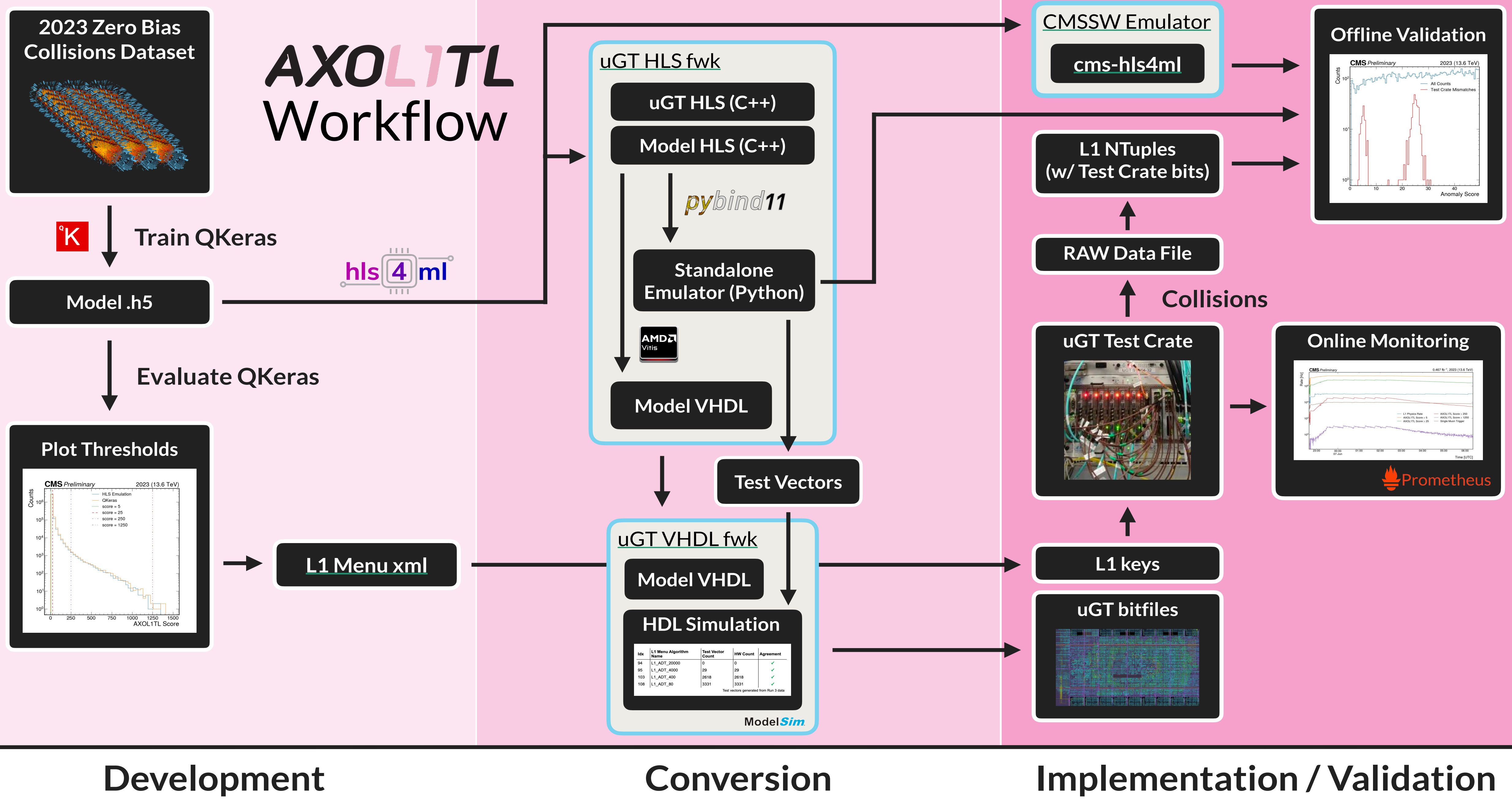
Level-1 Trigger constraints informed design



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

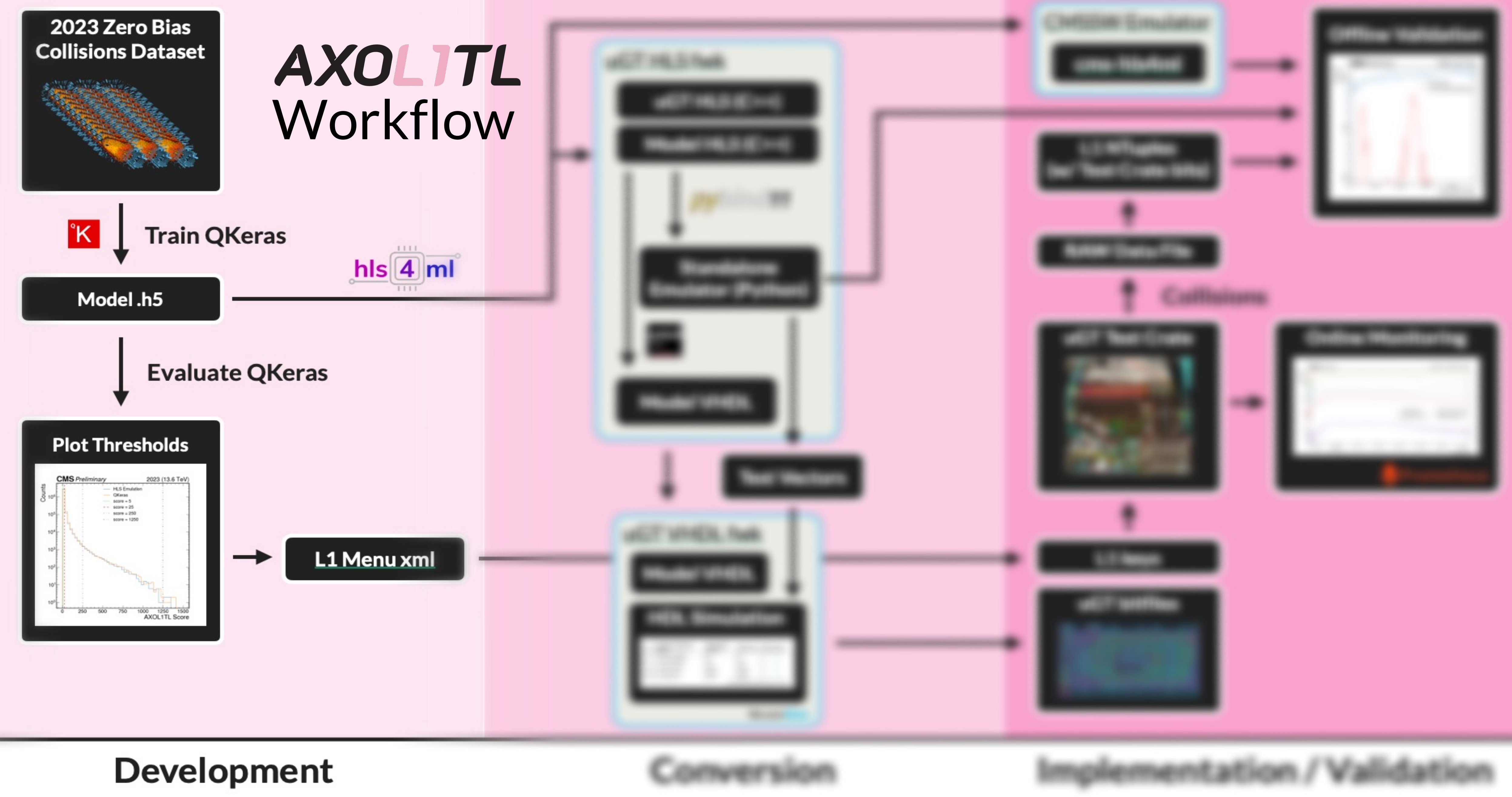
Reconstruction term — Full regularization term

- Remove decoder network
  - Significant latency & resource savings, minimal performance degradation
- Remove latent σ term from loss calculation
  - Saves even more on timing, negligible performance degradation



AXOL1TL
Workflow
2023 Zero Bias Collisions Dataset
Train QKeras
Model .h5
hls4ml
Evaluate QKeras
Plot Thresholds
L1 Menu xml
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
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

AXOL1TL Workflow

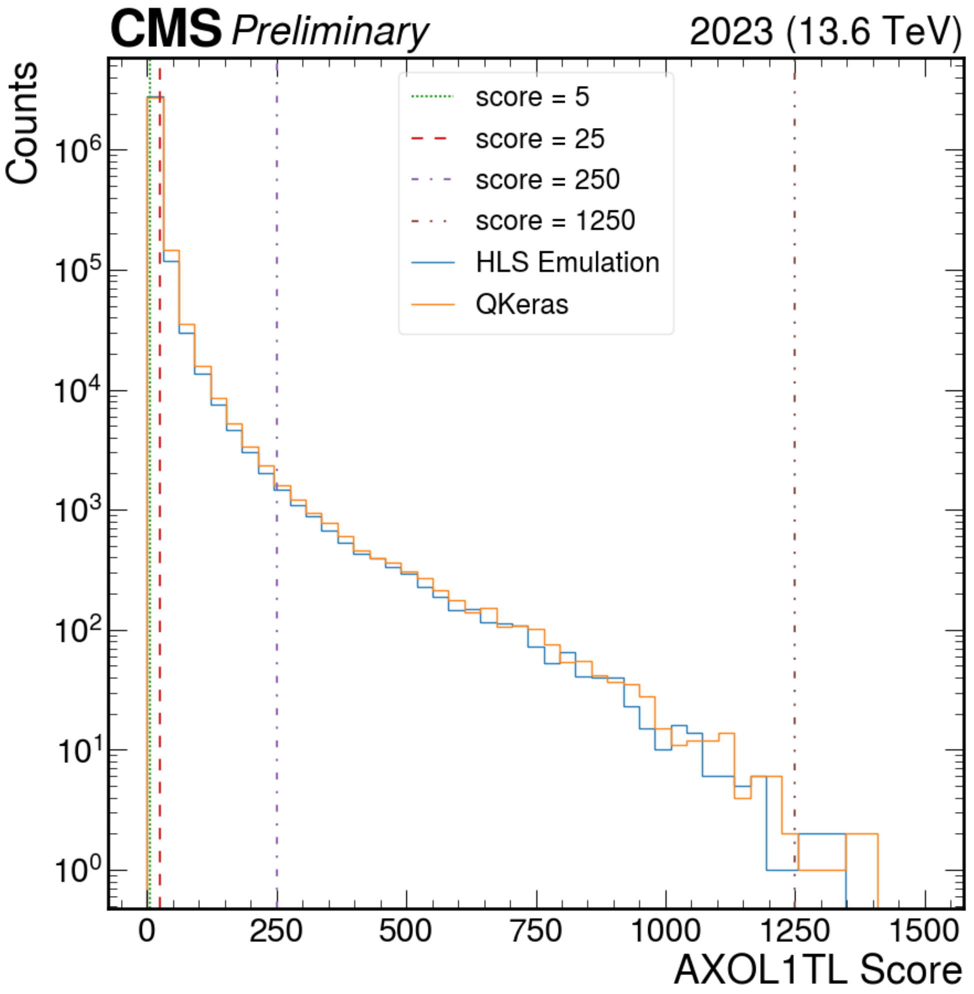


Model Performance

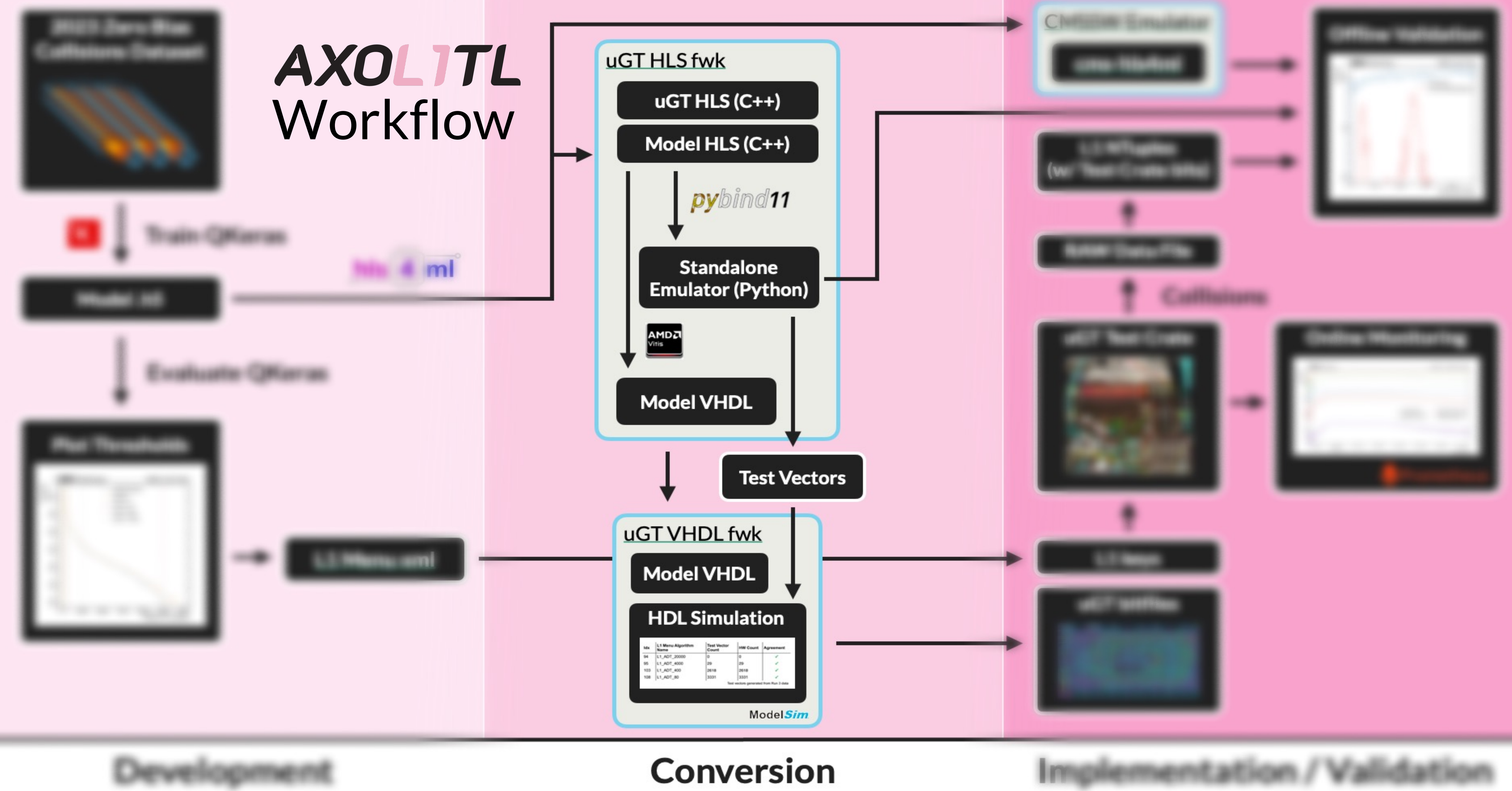
- AXOL1TL is trained with unbiased data collected by the CMS Experiment during 2023 with $\sqrt{s}$=13.6 TeV
  - 10.5 million events used – 50% for training, 50% for setting thresholds
- Dotted lines represent the score thresholds implemented in the Global Trigger Test Crate
- Significant performance improvement on various SM and BSM signals by adding AXOL1TL to the 2023 trigger menu
  - Signal samples are Monte-Carlo generated
  - Table shows performance improvement for a Higgs decaying to 2 (pseudo-) scalars to bottom quarks

h->a(bb)a(bb)

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

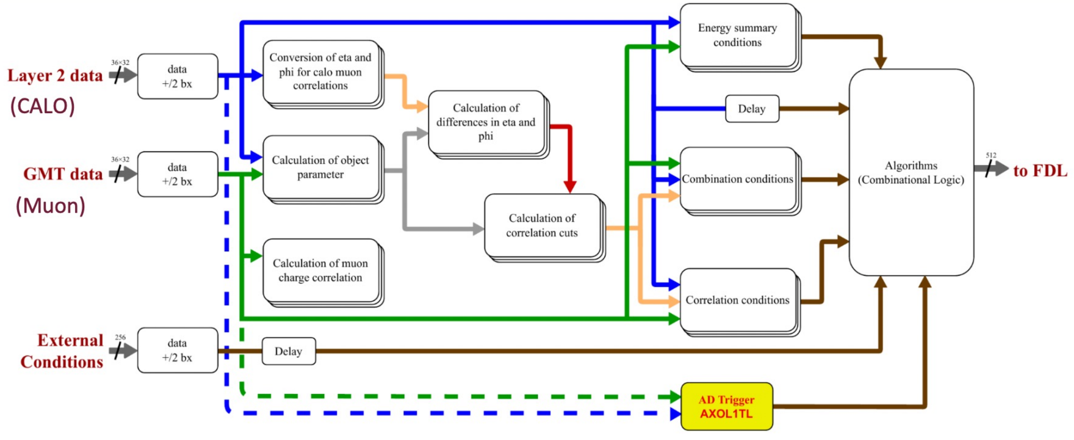


AXOL1TL Workflow



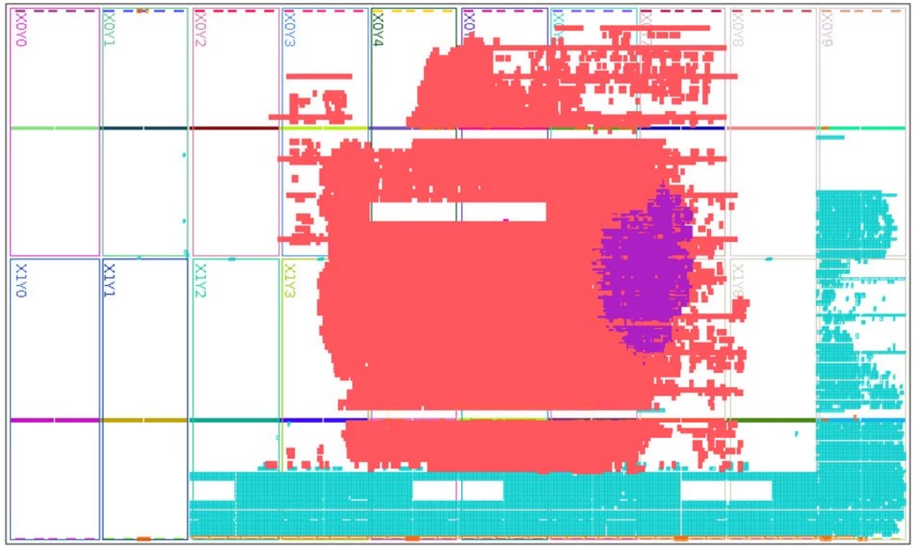
Firmware Development

- Anomaly detection algorithms integrated into Global Trigger firmware structure
  - Anomaly score calculated in concert with other global trigger quantities & output via same links
- High-Level Synthesis (HLS) implementation of Global Trigger firmware synthesizes hardware code (VHDL) for FPGAs
  - CERN Gitlab repository for HLS dependencies
  - Generate bitfiles for for MP7 boards
- AD firmware performance:
  - Fits 2 clock cycles @ 40 MHz latency requirement
  - Resources usage small
- hls4ml simulation of Virtex-7 FPGA chip on MP7 µGT board shows MP7 firmware payload , MP7 infrastructure, and the AXOL1TL network



Schematic the CMS Global Trigger

Virtex-7 FPGA Layout



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

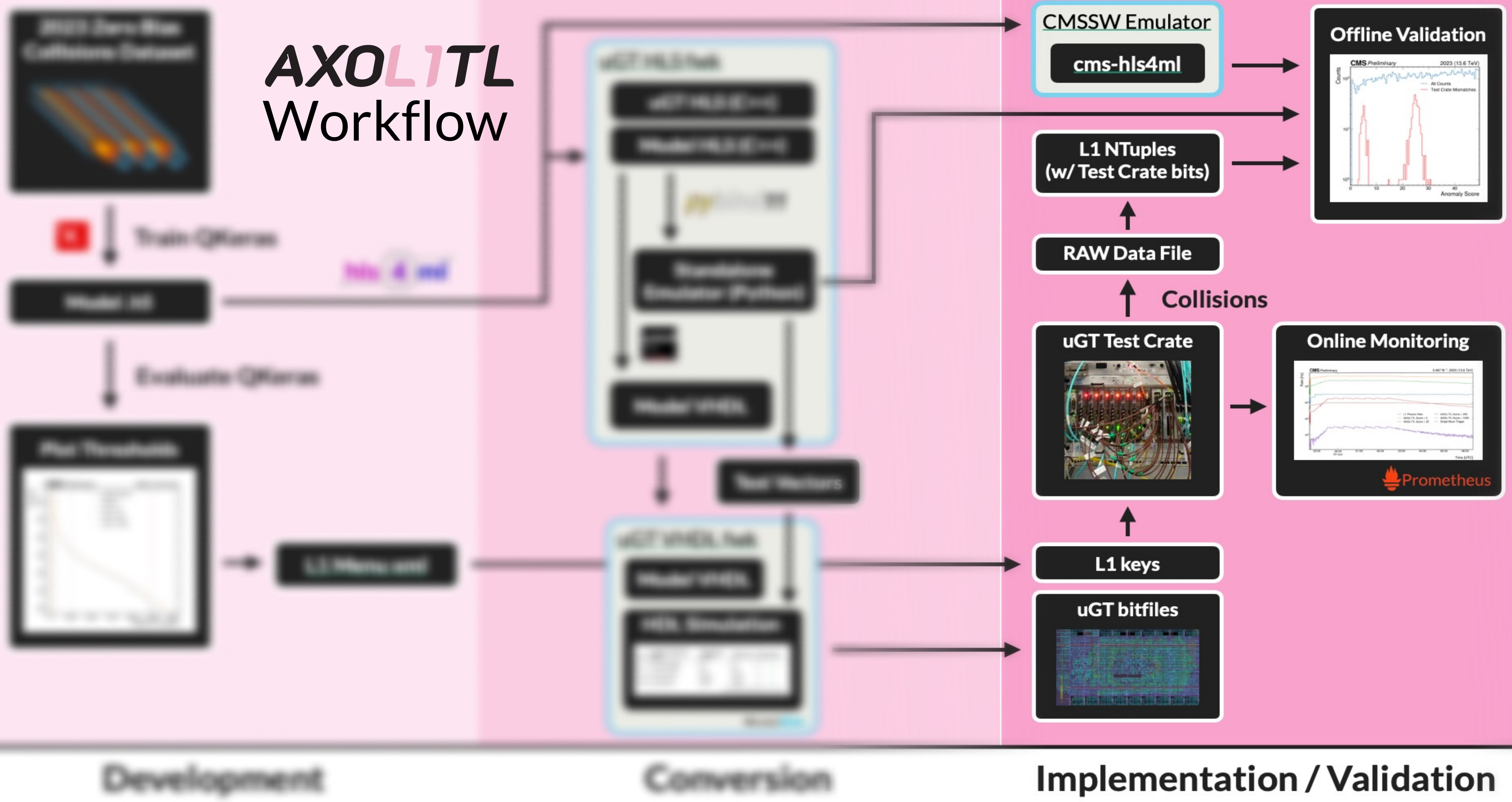
Level-1 Menu Validation

- New Level-1 Trigger Menu is built from μGT anomaly score output link and defined thresholds
- Test vector files are generated with data and MC
  - Trigger objects & detector conditions formatted as bitstring inputs
  - Reference decision (pass or fail) for AXOL1TL made with HLS emulator
- Standard Level-1 ModelSim environment used for validating menu builds
  - Reads μGT VHDL code and simulates decisions from test vector inputs
- Perfect trigger decision bit agreement

Idx	L1 Menu Algorithm Name	Test Vector 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 368566

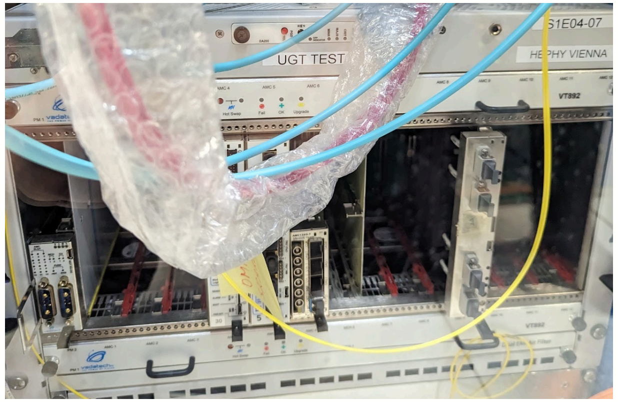
AXOL1TL Workflow

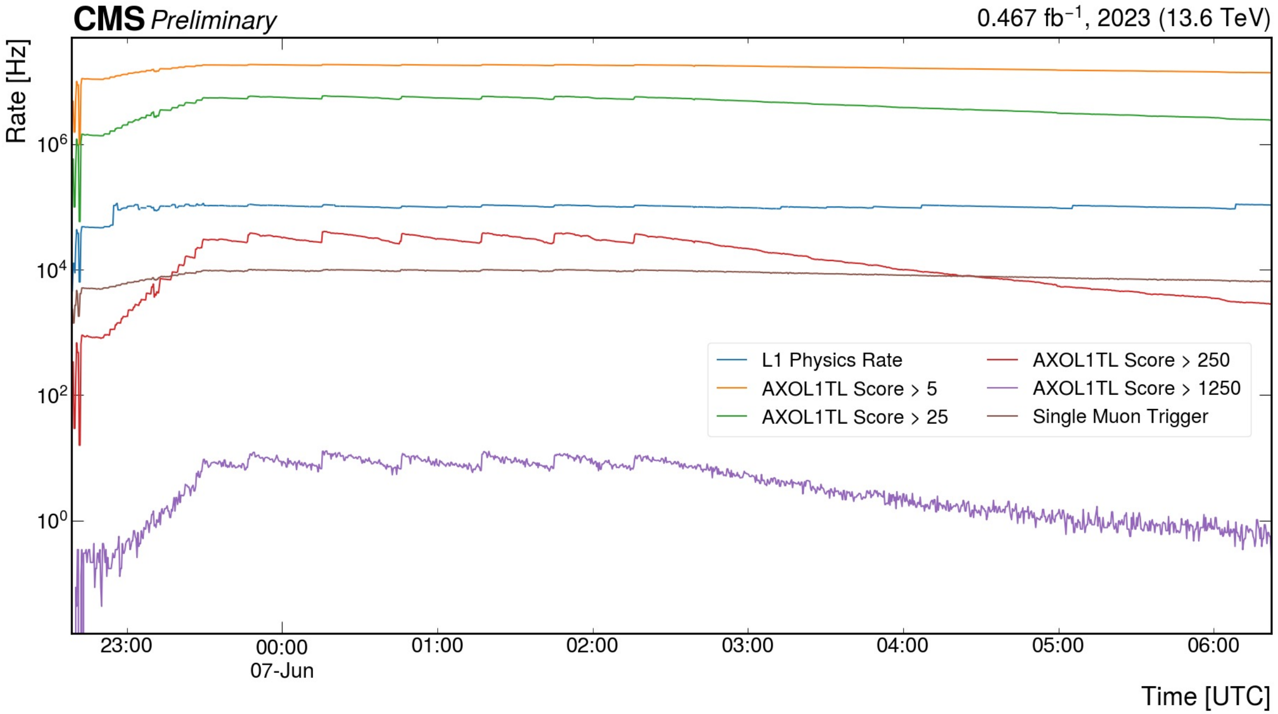


Test Crate Implementation

- CMS Global Trigger Test Crate
  - Identical copy of μGT board, used as backup & testing
  - Bit readout connected to data acquisition system but not configured to select events
- Prometheus monitoring tool answers real-time queries of trigger metrics
  - Used to monitor AXOL1TL rates during 2023 physics data-taking
- Test Crate model is trained on 2018 data with 4 score thresholds used to test rate boundaries
  - Used for firmware testing, not realistic proposal for trigger paths
- Consistent trigger performance shown for fill cycle
  - Single muon trigger (p_T > 22 GeV) shown for reference
  - Dips in rate due to LHC ramp-up and luminosity-levelling scheme

Global Trigger Test Crate sitting underground at CERN Point 5, next to CMS Detector

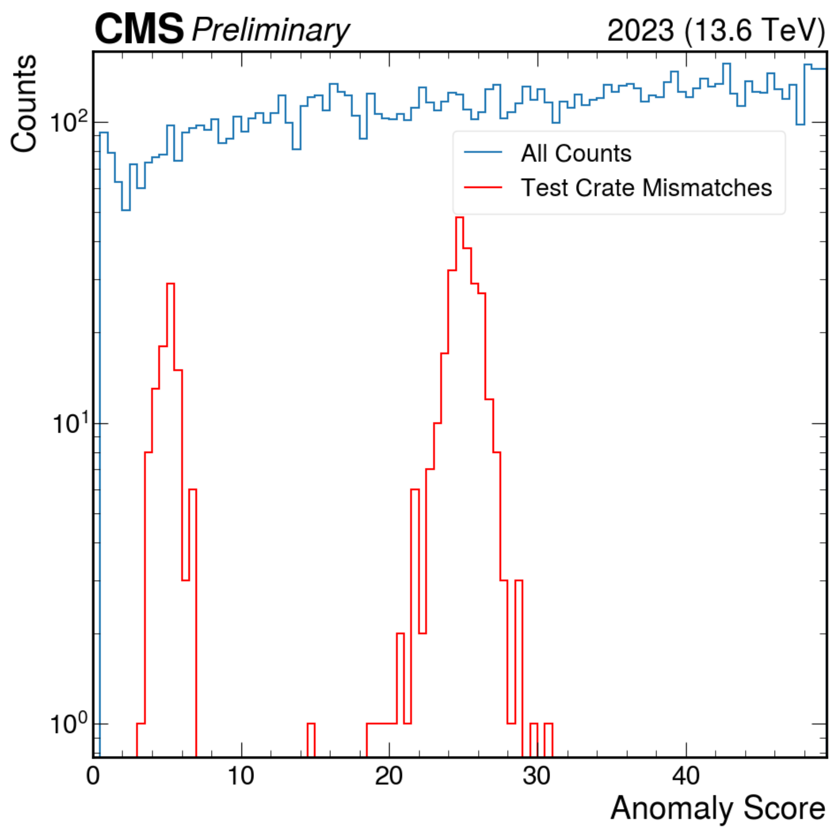




Test Crate Validation

- For certain runs, Test Crate decisions are recorded in 2023 data files
  - Use these bits to validate emulation and show rate agreement

- Minimal (~1%) mismatches between trigger hardware and emulation
  - Mismatches clustered near decision boundaries, most likely due to rounding issue



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

Summary

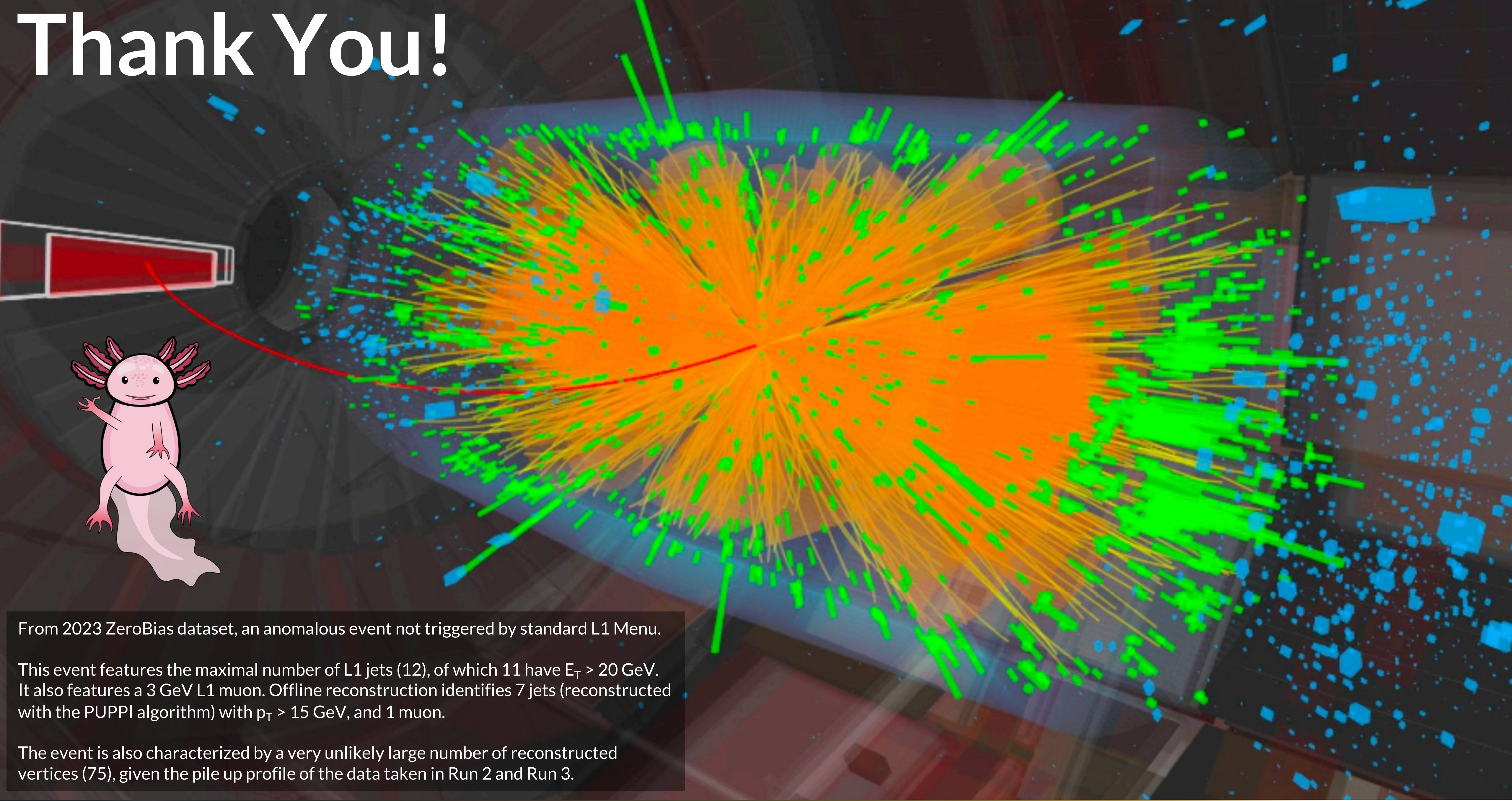
So far, we have shown

- A signal-agnostic trigger model sensitive to interesting physics
- A firmware implementation successfully integrated into the CMS Level-1 Trigger architecture
- A hardware trigger active on the CMS Global Trigger Test Crate that performed consistently during 2023 collisions
- Validation for all steps using HLS emulation

What's left to do

- Implement plans for downstream trigger logic
- Pending approval, integrate into production trigger and begin taking data!

Thank You!



From 2023 ZeroBias dataset, an anomalous event not triggered by standard L1 Menu.

This event features the maximal number of L1 jets (12), of which 11 have E_T > 20 GeV. It also features a 3 GeV L1 muon. Offline reconstruction identifies 7 jets (reconstructed with the PUPPI algorithm) with p_T > 15 GeV, and 1 muon.

The event is also characterized by a very unlikely large number of reconstructed vertices (75), given the pile up profile of the data taken in Run 2 and Run 3.

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.