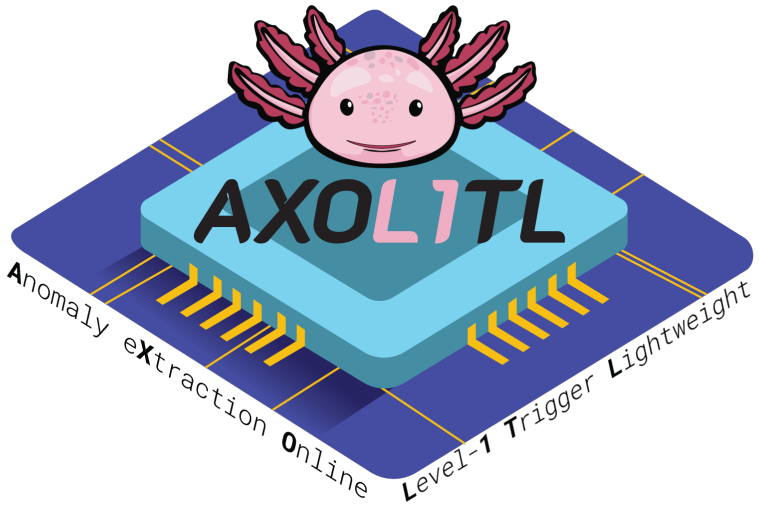


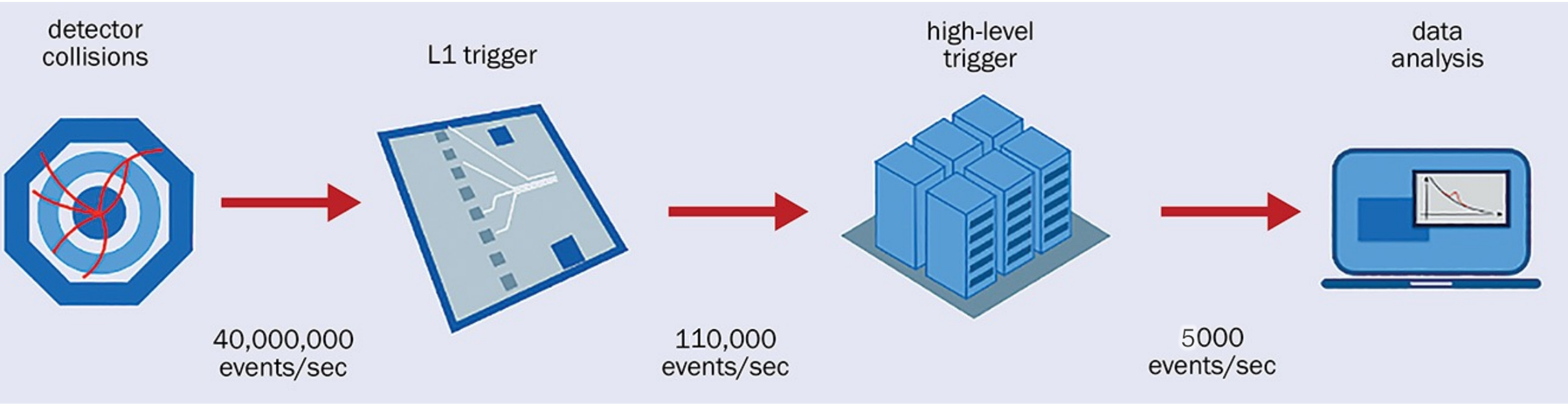
Realtime Anomaly Detection with the CMS Level-1 Global Trigger Test Crate

Jannicke Pearkes, for the CMS Collaboration

ML4Jets, Hamburg

07.11.2023

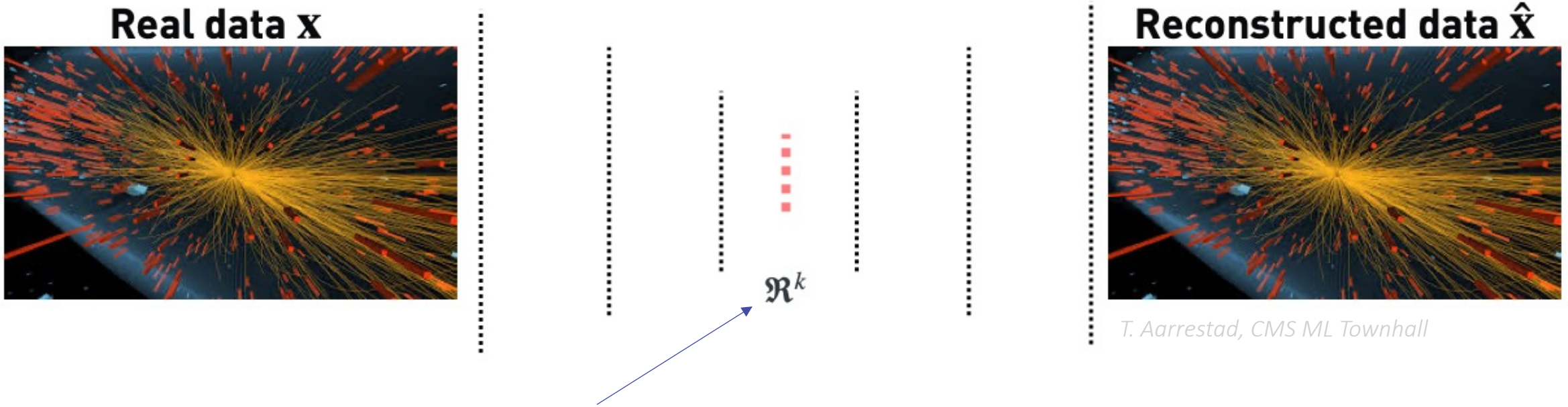
Data Collection in CMS



L1 trigger rejects
99.75% of LHC events

"What if we are missing new physics because we did not design the right trigger?"

Anomaly detection with autoencoders



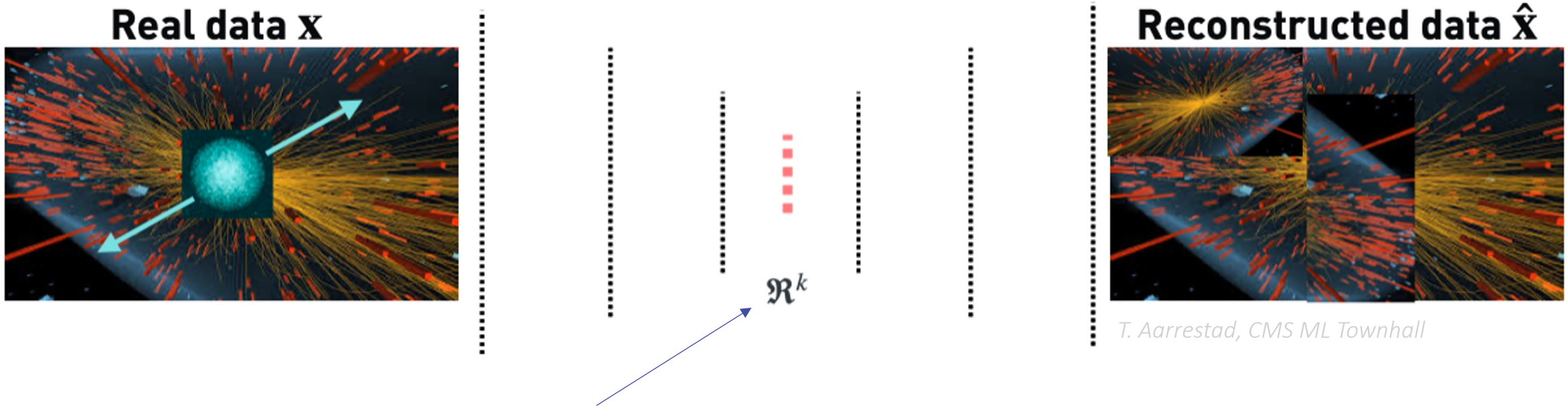
T. Aarrestad, CMS ML Townhall

Train on ZeroBias LHC data

Bottleneck: autoencoder learns to compress high dimensional inputs into low dimensional latent space

$x - \hat{x}$ represents degree of abnormality

Anomaly detection with autoencoders



T. Aarrestad, CMS ML Townhall

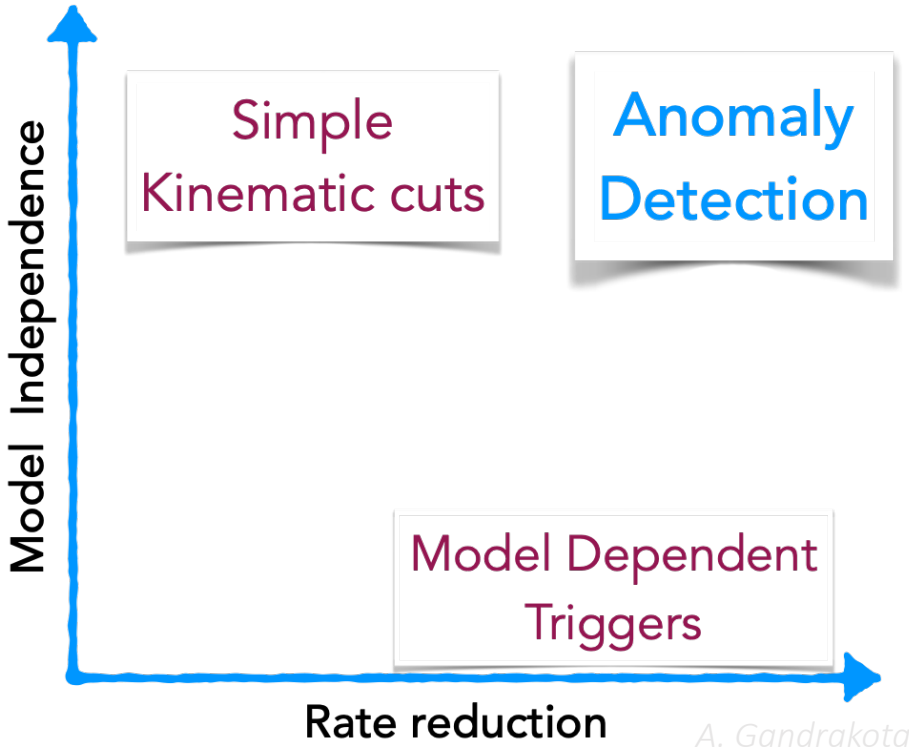
Train on ZeroBias LHC data

Bottleneck: autoencoder learns to compress high dimensional inputs into low dimensional latent space

$x - \hat{x}$ represents degree of abnormality

Why anomaly detection in the trigger?

- *Signal agnostic* - Applicable to signatures that we have not had the foresight or person-power to design specific triggers for
- *High sensitivity* – Can improve signal efficiency to signatures that are limited by L1 trigger bandwidth

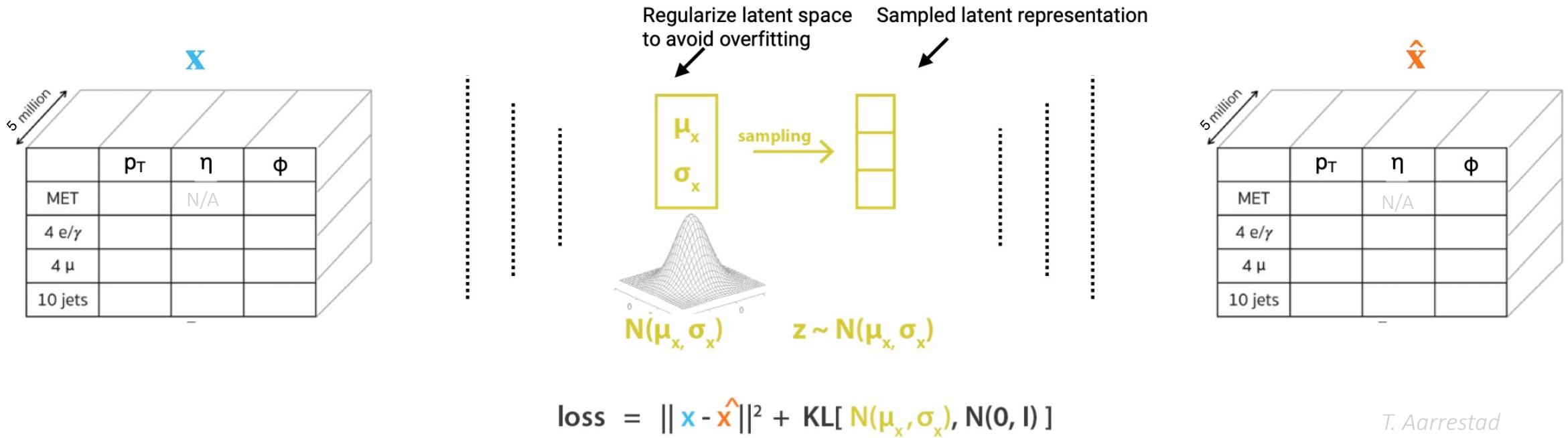


AXOL1TL Design

AXOL1TL is a variational autoencoder:

- Encodes input as a distribution over the latent space
- Additional loss term regularizes latent space to be Gaussian

Uses L1 trigger objects as inputs: (p_T, η, ϕ) of MET, up to 4 electron/photons, 4 muons, and 10 jets

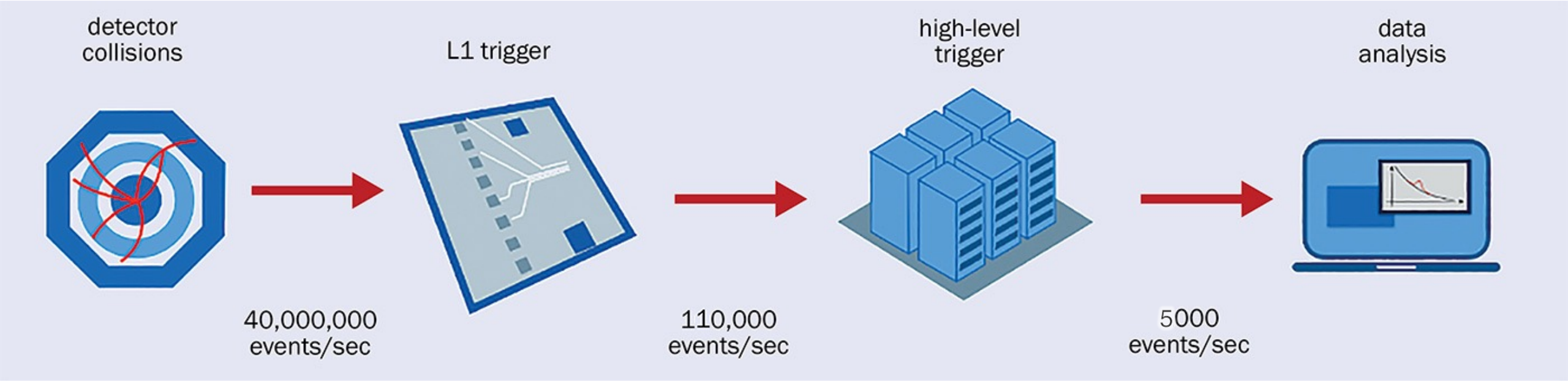


$$\text{loss} = ||x - \hat{x}||^2 + KL[N(\mu_x, \sigma_x), N(0, I)]$$

T. Aarrestad

Train on data collected by CMS in 2023 at √s=13.6 TeV, 10.5 million events 50% for training, 50% testing

Constraints for Implementing AXOL1TL in L1 Trigger

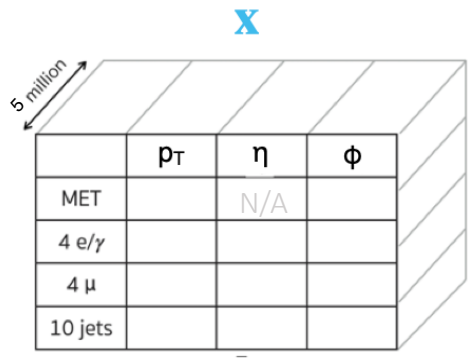


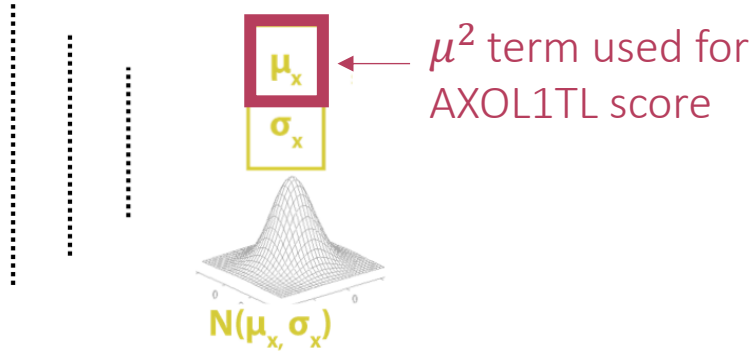
Global trigger constraints:

- Latency of 50 ns
- Low resource utilization on FPGA

AXOL1TL Design Choices

Only deploy encoder half of the network, compute degree of abnormality from latent space directly.
Halves the network size and latency!





T. Aarrestad, CMS ML Townhall

Small, fully connected network architecture (encoder: 32,16,8 nodes per layer)

Quantization aware training with QKeras

hls4ml to translate network into VHDL design for FPGA

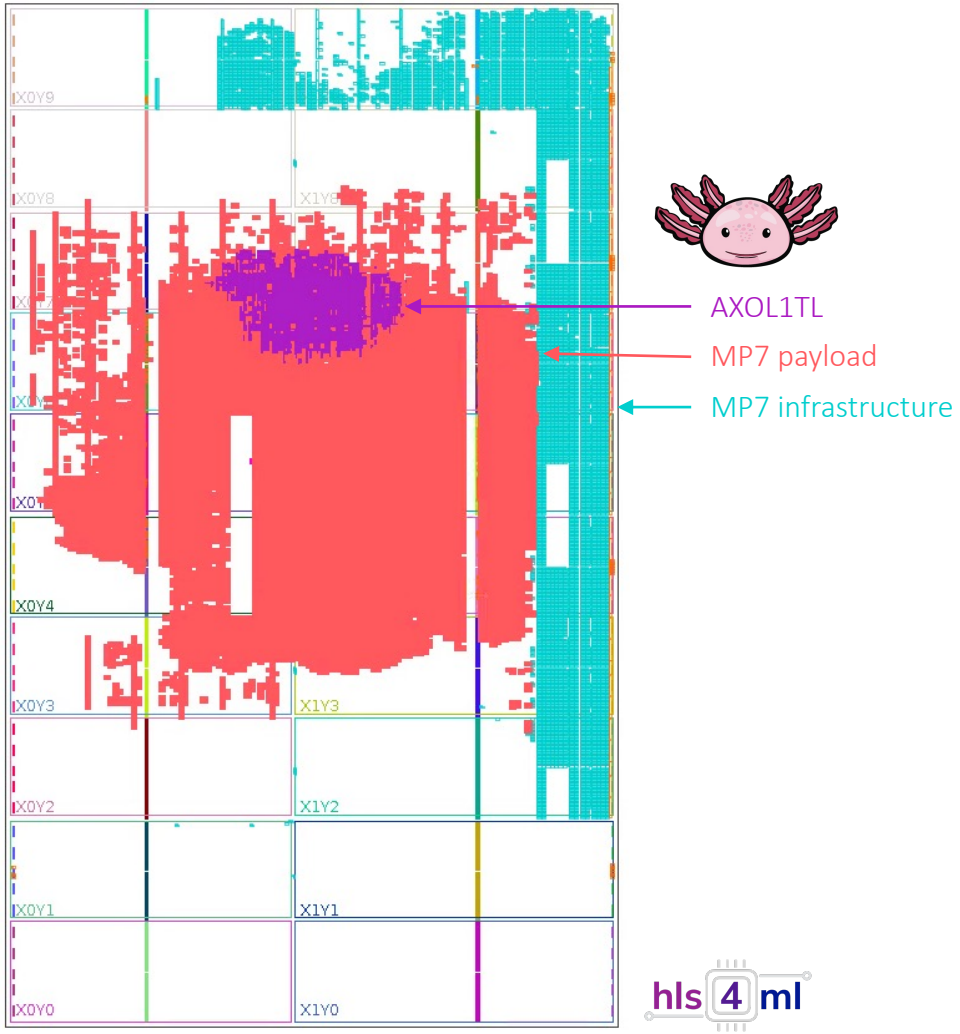
AXOL1TL Implementation

Implemented on Xilinx Virtex-7 XCVU9P FPGA:

50ns latency and resource requirements met

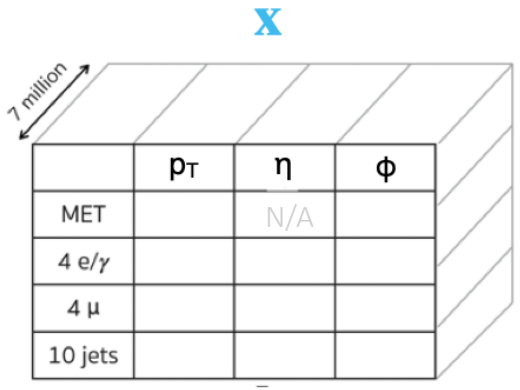
Resource utilization of Virtex-7 FPGA chip on Imperial College MP7 µGT board

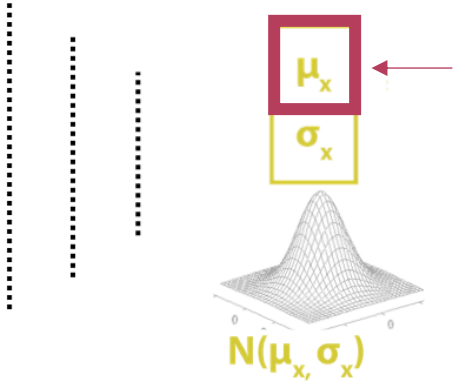
	Latency	LUTs	FFs	DSPs	BRAMs
AXOL1TL	2 ticks 50 ns	2.1%	~0	0	0



AXOL1TL Thresholds

CMS-DP-2023-079

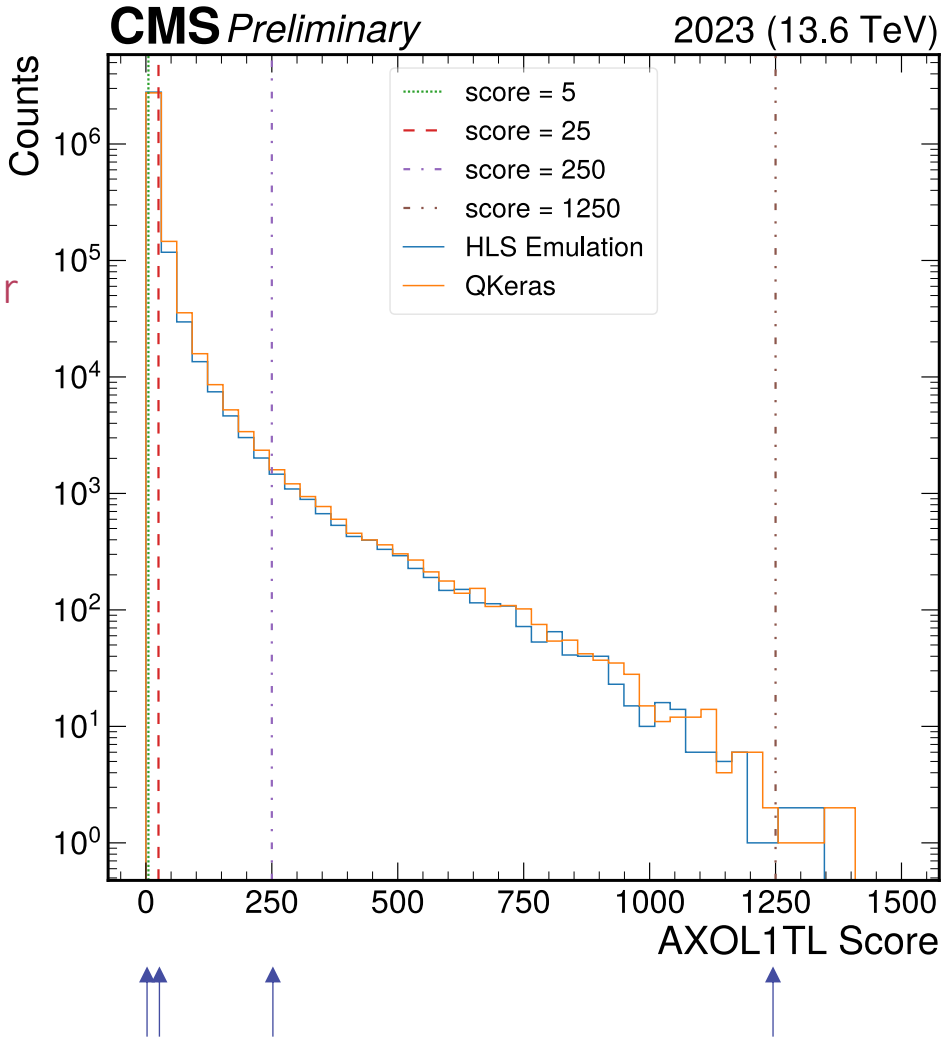




μ^2 term used for AXOL1TL score

$$\text{loss} = ||x - \hat{x}||^2 + \text{KL}[\,N(\mu_x, \sigma_x), N(0, I)\,]$$

T. Aarrestad, CMS ML Townhall

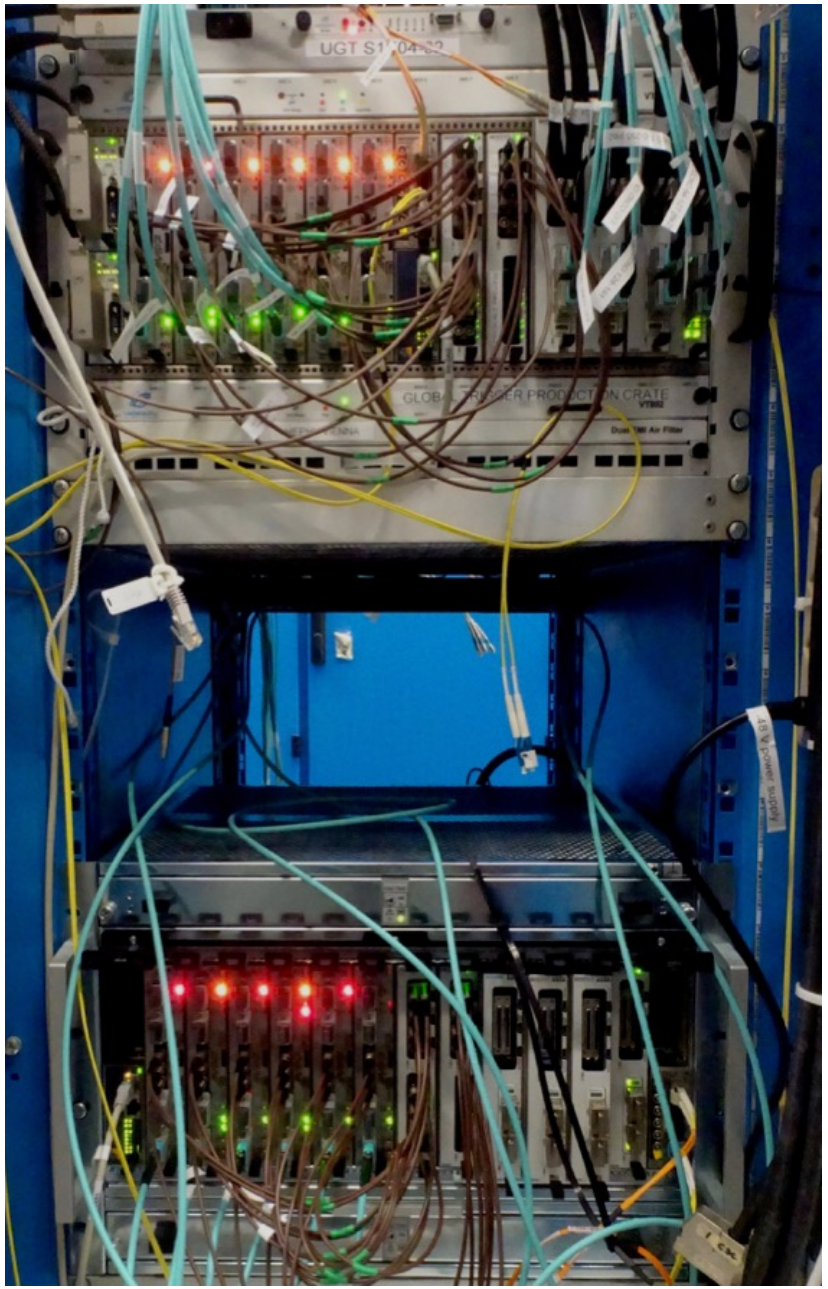


Different anomaly score thresholds can be used to target different trigger rates

AXOL1TL Status

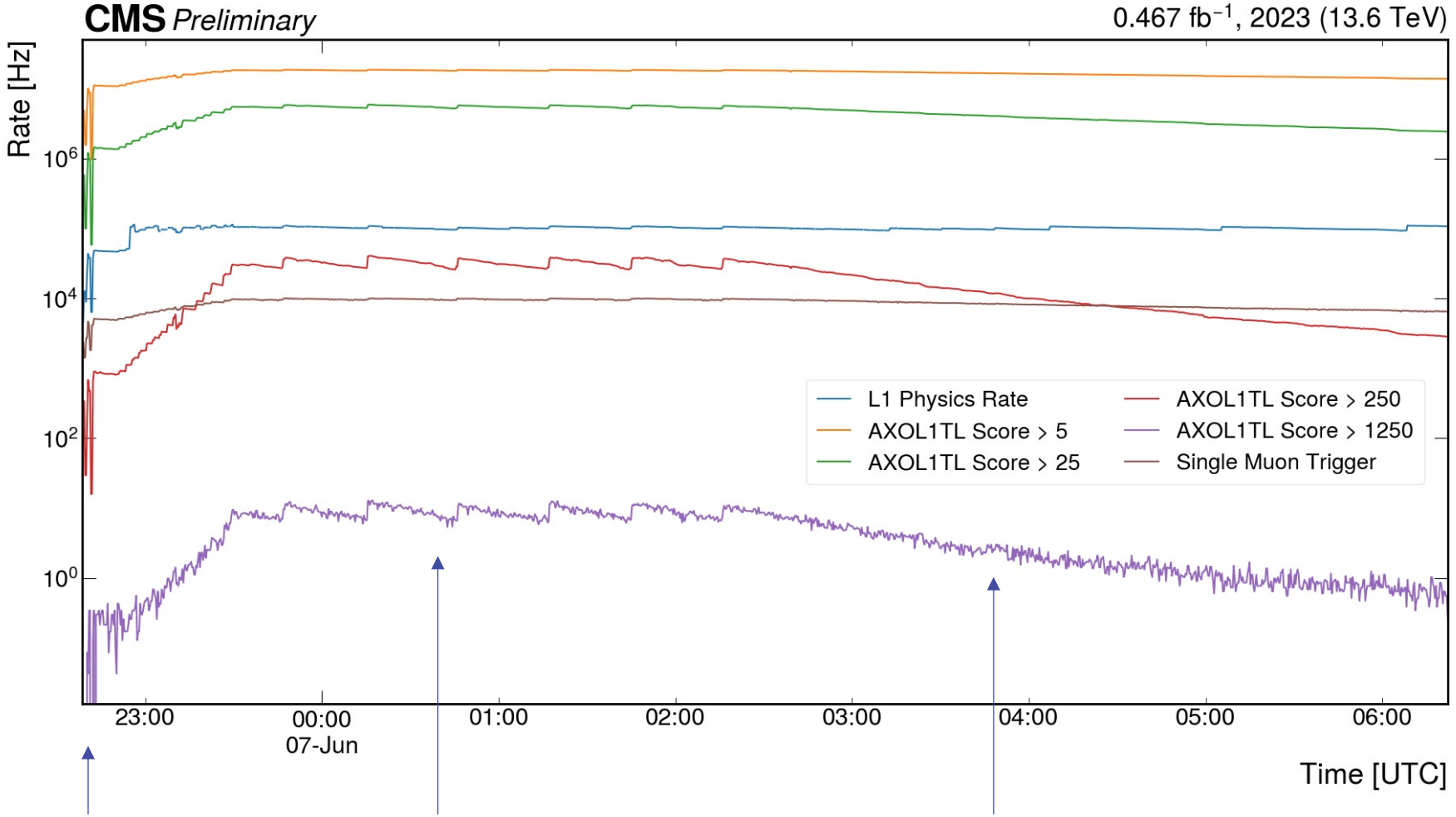
Running in the CMS L1 Global Trigger test crate:

- Identical to uGT production crate
- Used for global trigger algorithm R&D using same input data as production crate, but not used to collect physics data
- Added to CMS L1T system in Run 3
- Used for monitoring rates and comparing with other L1 triggers
- Preliminary model trained on 2018 ZeroBias data deployed in test crate in 2023



Test Crate Monitoring

CMS-DP-2023-079



start of fill

luminosity levelling

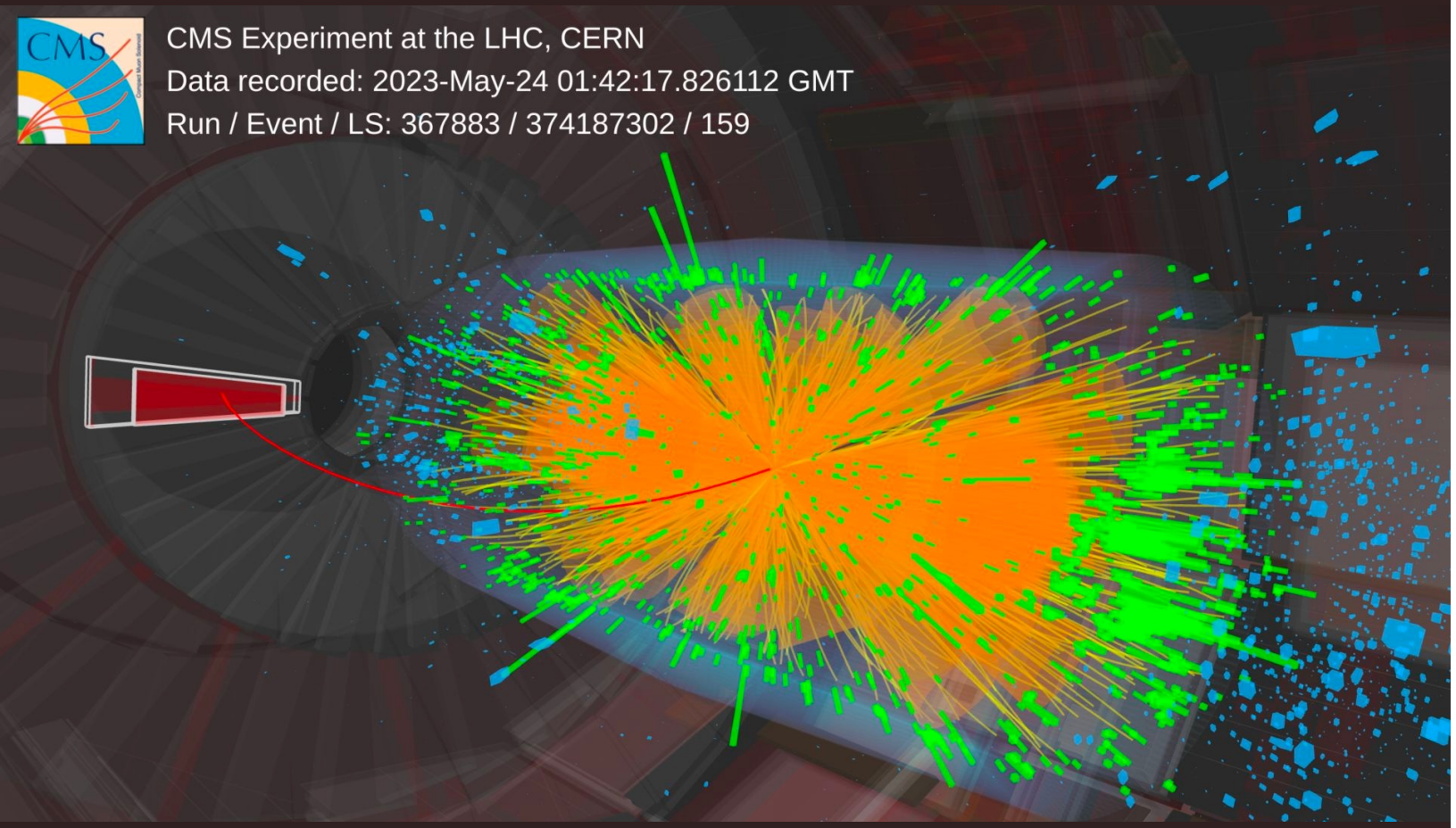
luminosity decreases at end of fill

AXOL1TL rates are stable relative to other L1 triggers.

AXOL1TL rate closely follows the changes in instantaneous luminosity during the run.

Thresholds for monitoring do not correspond to realistic level 1 rates.

A highly anomalous event:



CMS-DP-2023-079

An anomalous event with the highest AXOL1TL score that was not also triggered by the Level 1 menu in 2023, Run 367883, Zero Bias data.

At Level 1:
- 12 jets (11 with $E_T > 20$ GeV)
- 1 muon with 3 GeV

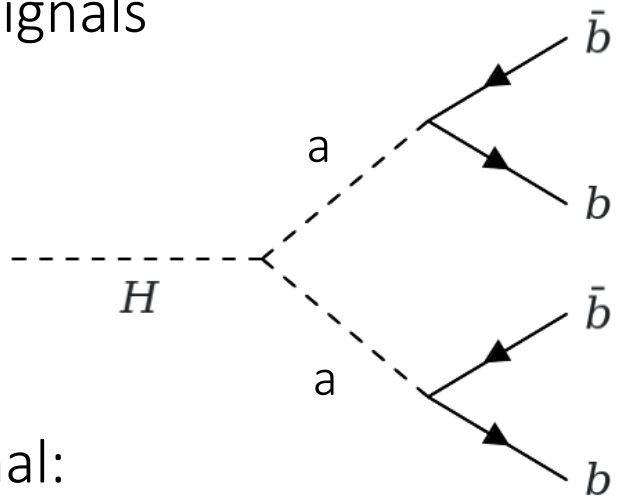
Offline reconstruction:
- 7 PUPPI jets $p_T > 15$ GeV
- 1 muon
- 75 reconstructed vertices

Busy event given the pile up profile of the Run 2-2018 training data and data collected in Run 3-2023. See back-up for PU profiles.

Physics Performance

Significant performance improvement on various SM and BSM signals by adding AXOL1TL to the 2023 trigger menu.

$$\text{Improvement} = \frac{\text{L1 Efficiency w/ AXOL1TL@freq}}{\text{L1 Efficiency w/o AXOL1TL}} - 1$$



Example performance improvement for H->aa[15 GeV]->4b signal:

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

CMS-DP-2023-079

Large increase in signal efficiency for a small increase in rate of 1-10 kHz relative to total L1 rate of ~110 kHz.

Conclusions

- AXOL1TL is an anomaly detection-based level 1 trigger for CMS
- Our model meets the timing and resource requirements imposed by the CMS L1 trigger system
- An initial version of the model is implemented in the test crate at the CMS Experiment
- Further integration with the L1 trigger system is underway!

Thank you!

References

CMS Collaboration. ”Anomaly Detection in the CMS Global Trigger Test Crate for Run 3”. CMS-DP-2023-079 , CERN-CMS-DP-2023-079 (2023). https://cds.cern.ch/record/2876546.

N. Zipper, “Testing a Neural Network for Anomaly Detection in the CMS Global Trigger test crate during Run 3”, TWEPP 2023, https://indico.cern.ch/event/1255624/contributions/5444028/

C. Sun, “Realtime Anomaly Detection in the CMS Experiment Global Trigger Test Crate”, FastML 2023, https://indico.cern.ch/event/1283970/contributions/5554350/

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.

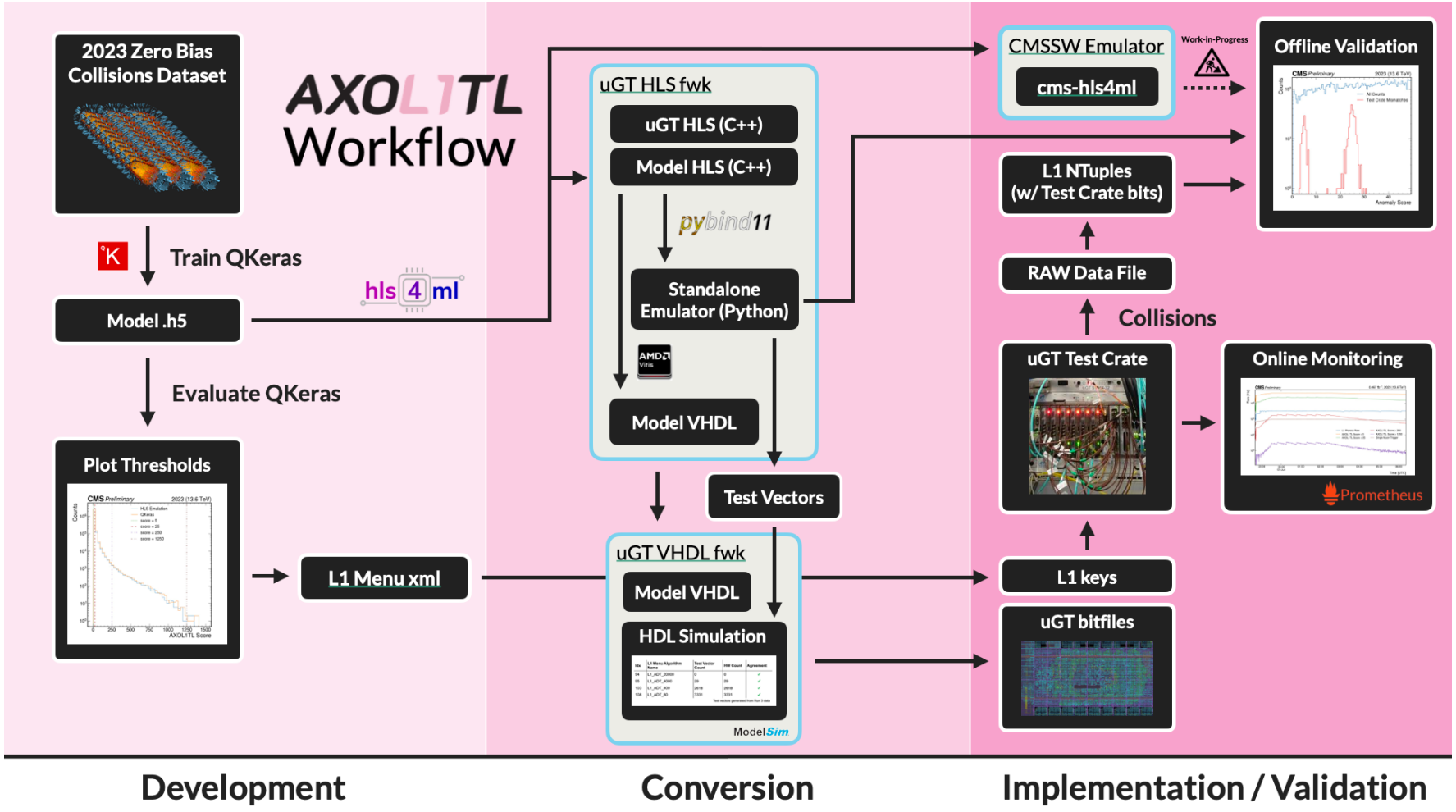
FastML Team. hls4ml (Version v0.7.1) [Computer software]. https://doi.org/10.5281/zenodo.1201549.

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.

Axol1tl Pipeline



Firmware Validation

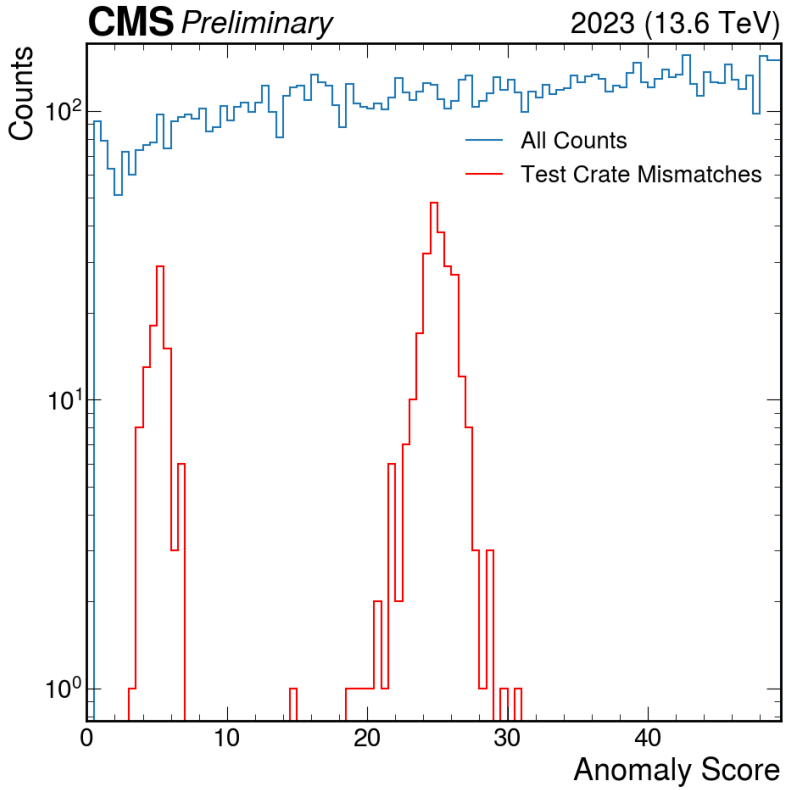
Test Crate firmware validation. The table shows trigger bits for the L1 menu including 4 anomaly detection thresholds: scores >1250, >250, >25, and >5 from top to bottom. Test vector column is generated from inference results of standalone emulator and HW count comes from standard global trigger firmware simulation workflow using ModelSim. Perfect bit agreement is observed.

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 3 data

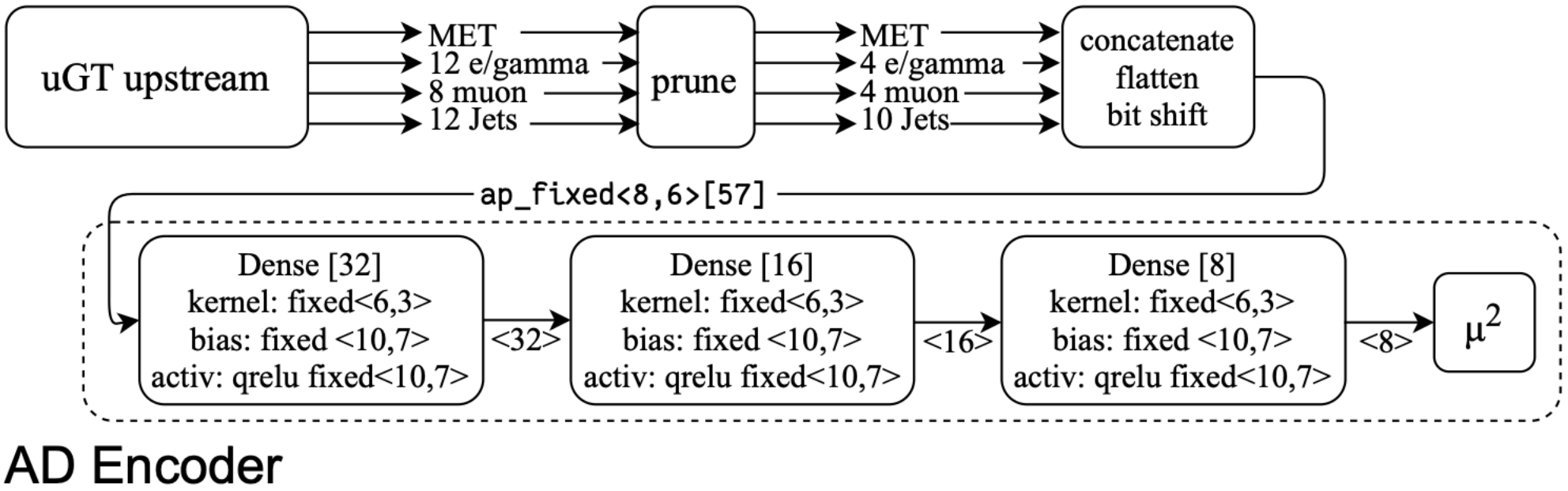
Test Crate Validation

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

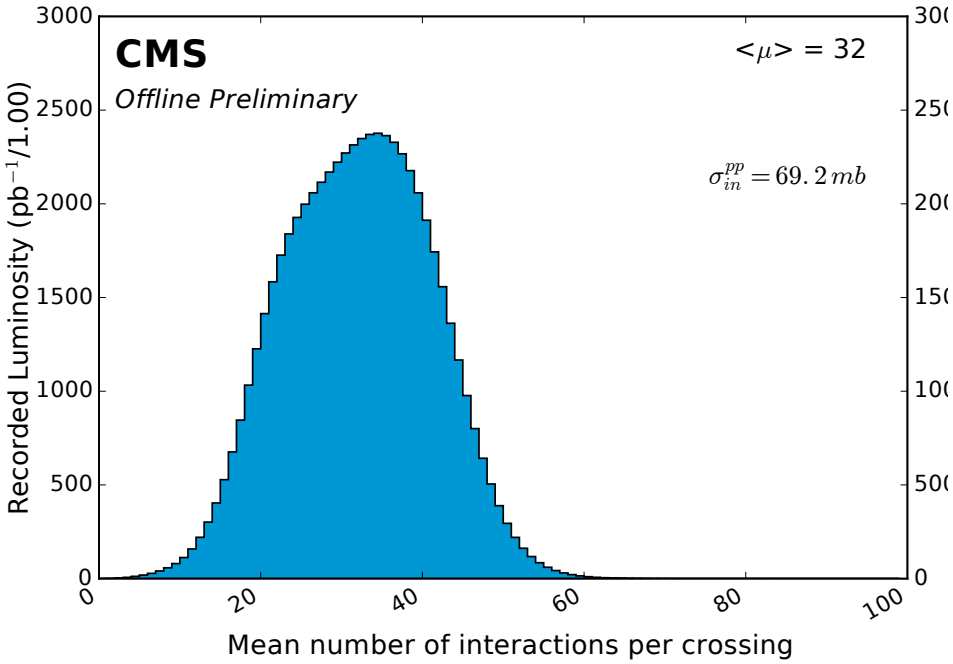


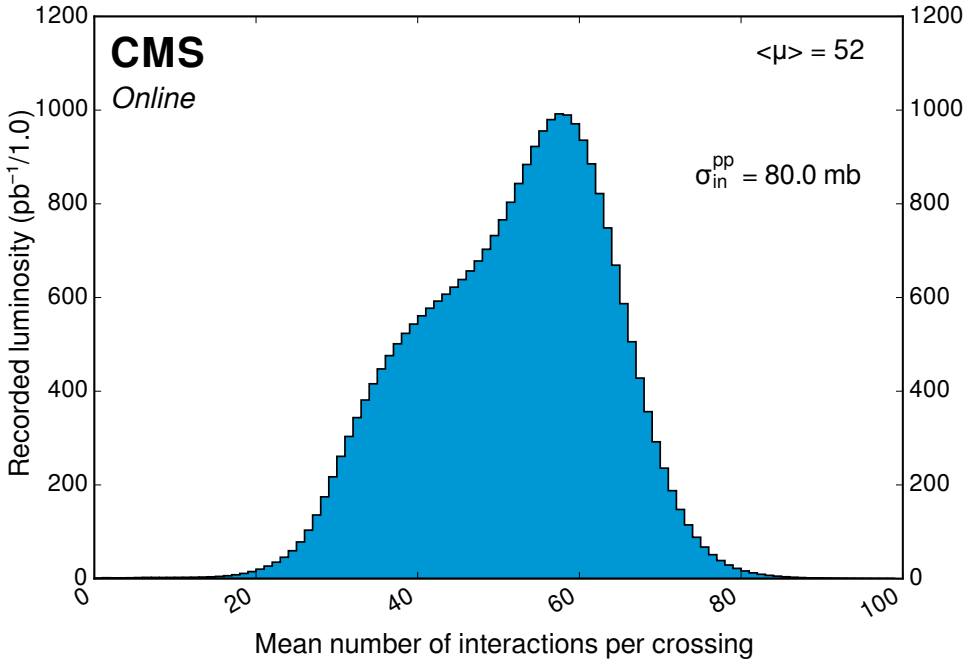
Anomaly Detection hardware vs. emulation trigger mismatches. Events from promptly reconstructed 2023 Ephemeral ZeroBias data where hardware bits are recorded from configured μGT test crate. In table (left), Test Crate Count shows events triggered in hardware and read out into data and Standalone Emulator Count is evaluated via offline inference with L1 objects. Anomaly score distribution of all events (right): red segments represent mismatches between hardware and emulation. Clustering near decision boundaries implies issue is due to precision/rounding problem. Minimal mismatches in hardware vs. emulation (≤ 1%) observed.

AD Encoder Implementation



Pile-up profiles in 2018 and 2023





2018 PU profile

2023 PU profile

From: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults