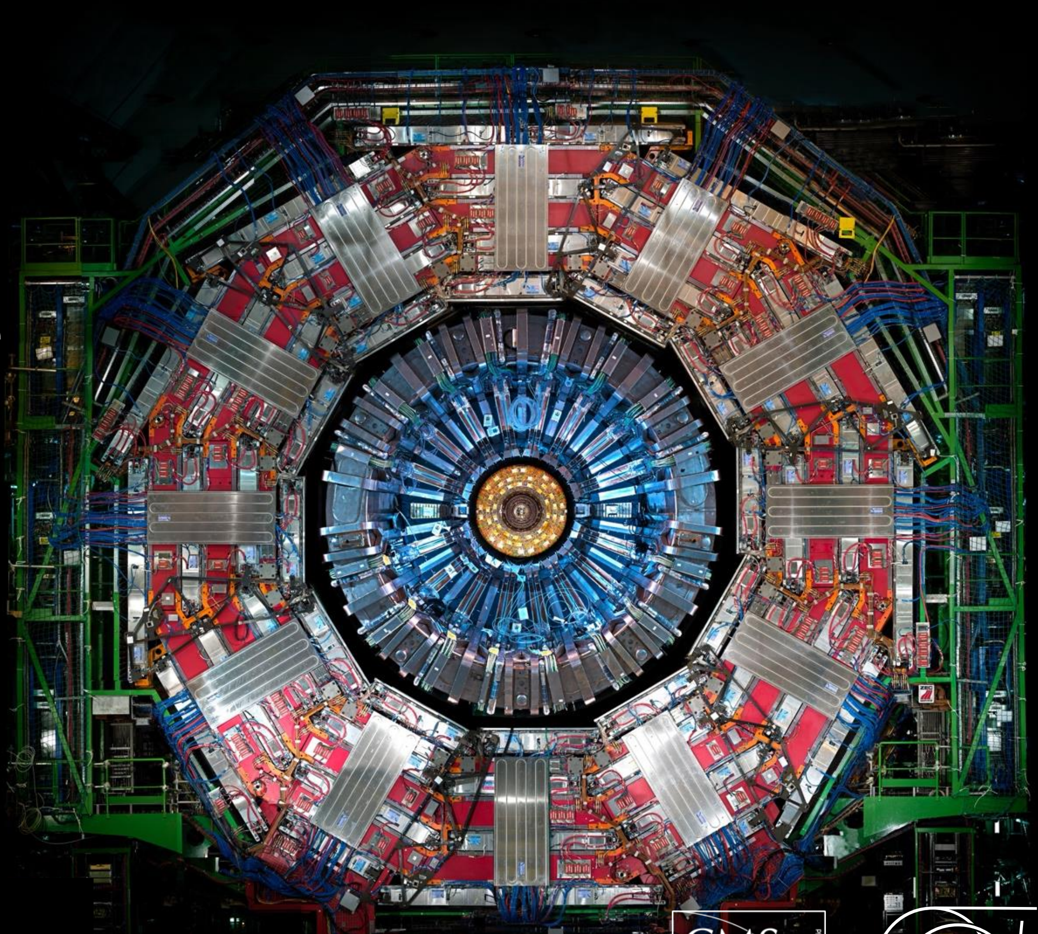
CMS Experiences with Anomaly Detection in the L1 Trigger



Jannicke Pearkes

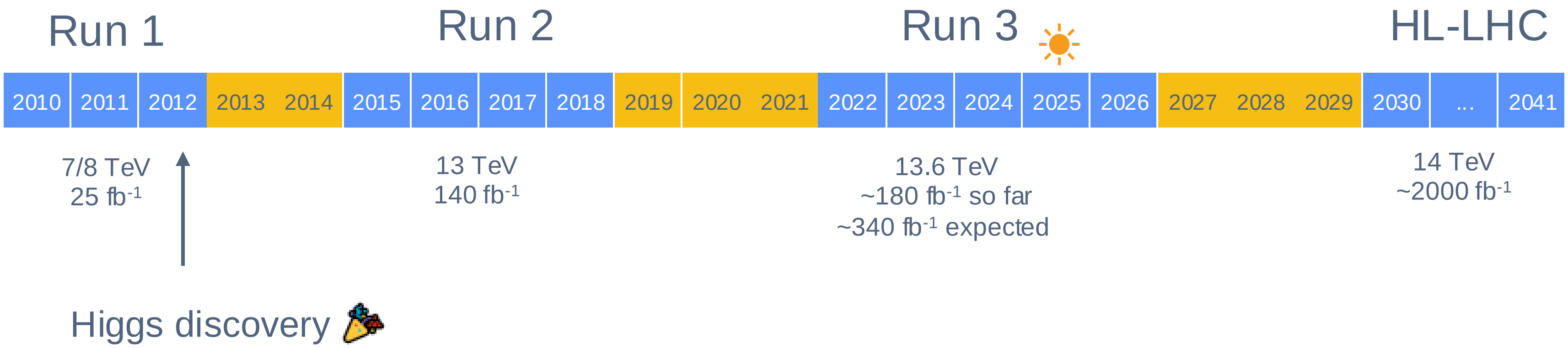
CERN Openlab Technical Workshop

March 4th, 2025

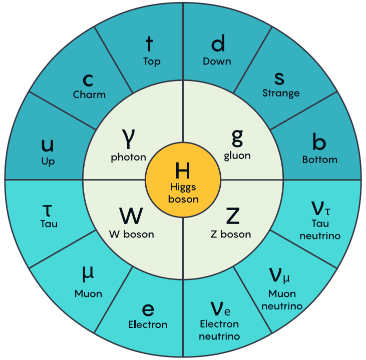
University of Colorado Boulder

NexTGen
Next Generation Triggers

LHC Timeline



LHC Timeline



More than 1,300 research papers from CMS alone

Run 1 | Run 2 | Run 3 ☀ | HL-LHC

2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 | ... | 2041

Run 1: 7/8 TeV, 25 fb^{-1}

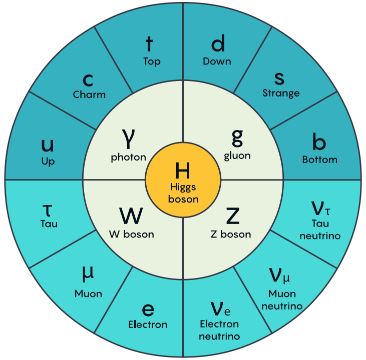
Run 2: 13 TeV, 140 fb^{-1}

Run 3: 13.6 TeV, ~180 fb^{-1} so far, ~340 fb^{-1} expected

HL-LHC: 14 TeV, ~2000 fb^{-1}

Higgs discovery 🎉

LHC Timeline



More than 1,300 research papers from [CMS](CMS) alone

Run 1					Run 2							Run 3								HL-LHC		
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	...	2041

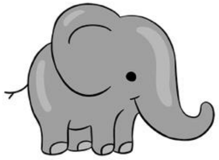
Run 1: 7/8 TeV, 25 fb^{-1}

Run 2: 13 TeV, 140 fb^{-1}

Run 3: 13.6 TeV, ~180 fb^{-1} so far, ~340 fb^{-1} expected

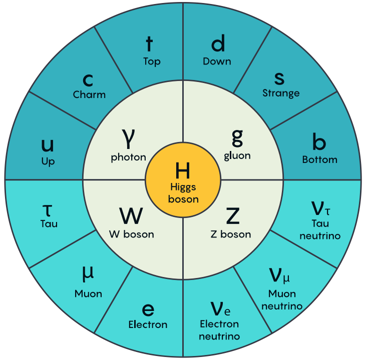
HL-LHC: 14 TeV, ~2000 fb^{-1}

Higgs discovery



No sign of beyond the standard model physics so far

LHC Timeline



More than 1,300 research papers from CMS alone

Run 1					Run 2							Run 3								HL-LHC		
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	...	2041

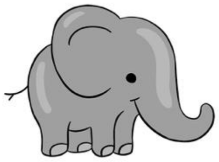
Run 1: 7/8 TeV, 25 fb^{-1}

Run 2: 13 TeV, 140 fb^{-1}

Run 3: 13.6 TeV, ~180 fb^{-1} so far, ~340 fb^{-1} expected

HL-LHC: 14 TeV, ~2000 fb^{-1}

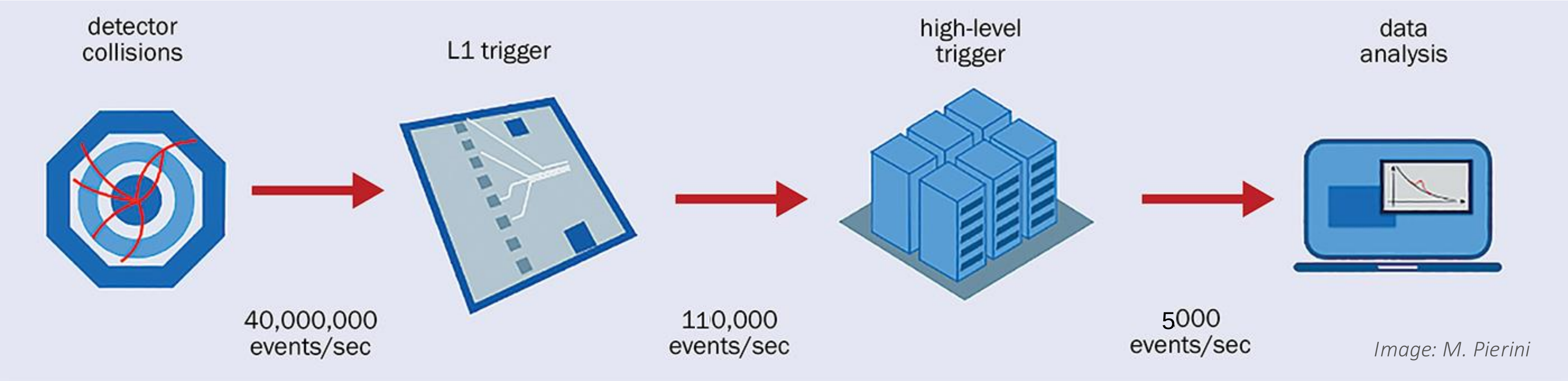
Higgs discovery



No sign of beyond the standard model physics so far

New technologies are opening doors to search for new physics more creatively than ever before.

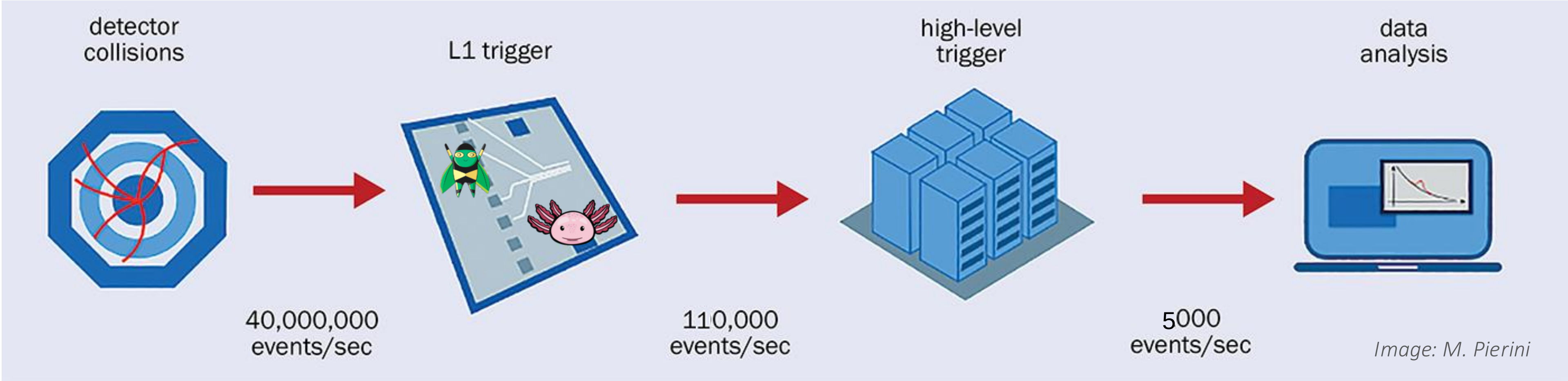
CMS Trigger Overview



L1 trigger filters of 99.75% of collision events

If we don't identify interesting events in trigger, we lose them forever!

Motivation for Anomaly Detection at L1



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

AXOL1TL and CICADA anomaly detection algorithms use machine learning to be more signal agnostic

Anomaly Detection with Autoencoders

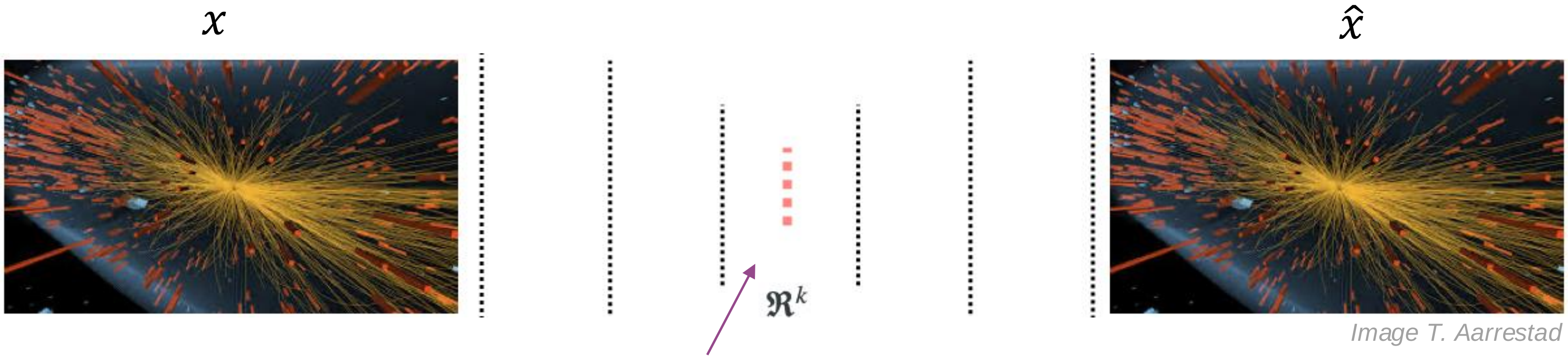


Image T. Aarrestad

Train on randomly sampled data

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

Unsupervised learning: $x - \hat{x}$ represents degree of abnormality

Anomaly Detection with Autoencoders

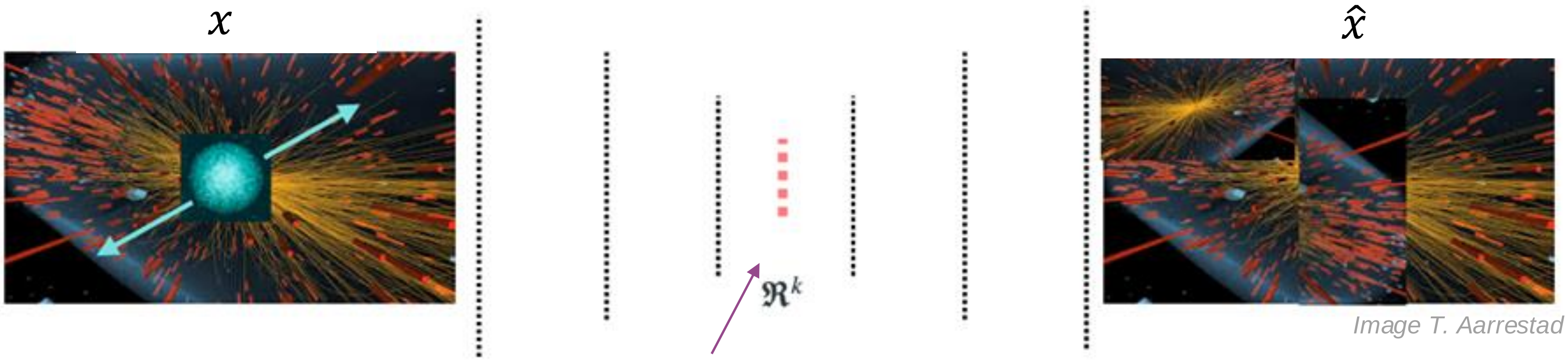


Image T. Aarrestad

Train on randomly sampled data

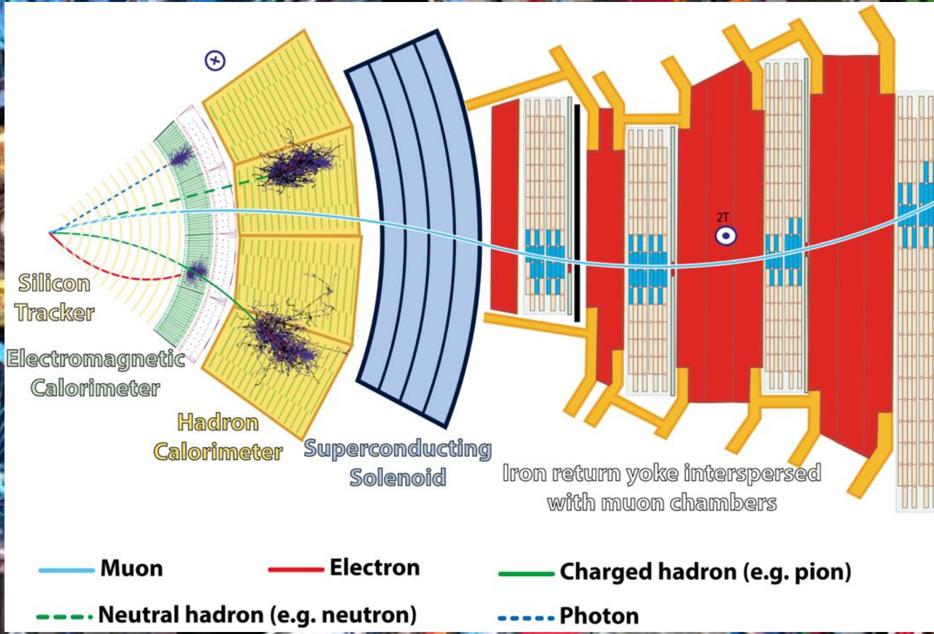
Bottleneck: autoencoder learns to compress high dimensional inputs into low dimensional latent space

Unsupervised learning: $x - \hat{x}$ represents degree of abnormality

CMS Experiment

Like a giant, 14,000-ton, camera

Different detector components sensitive to different types of particles

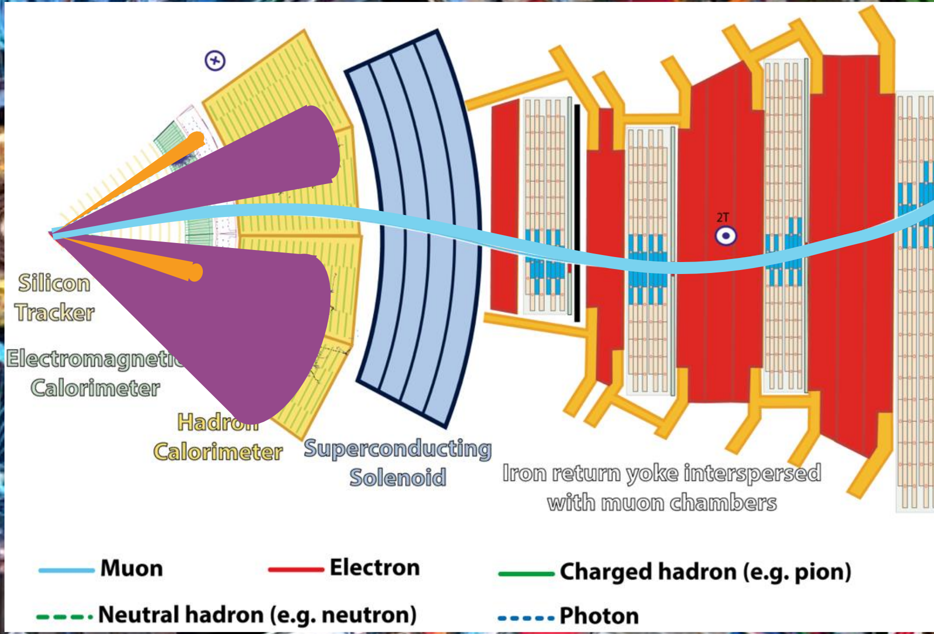


CMS Experiment

Like a giant, 14,000-ton, camera

Different detector components sensitive to different types of particles

AXOL1TL takes in reconstructed L1 trigger objects: jets, muons, electron/photons, missing energy

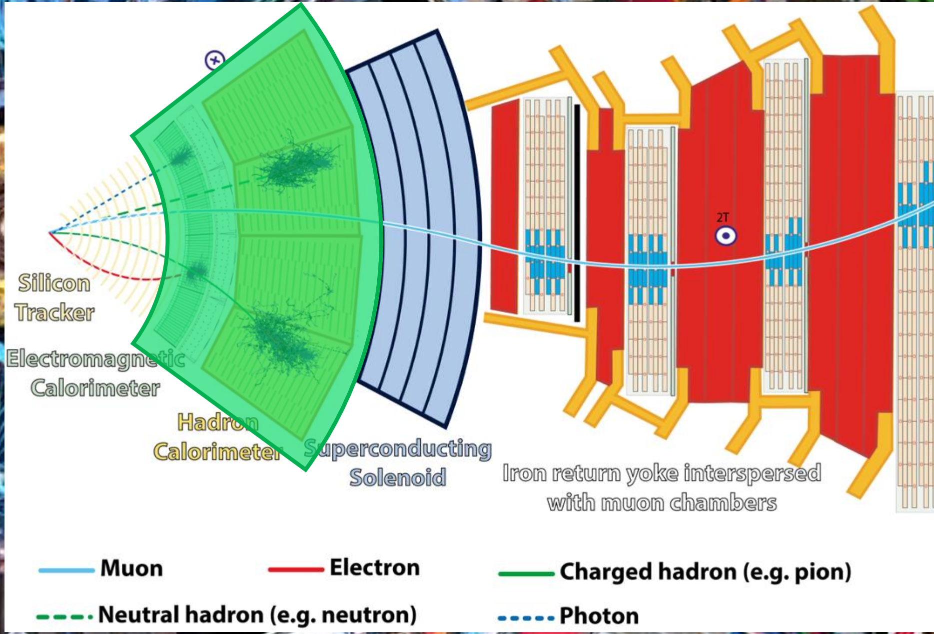


CMS Experiment

Like a giant, 14,000-ton, camera

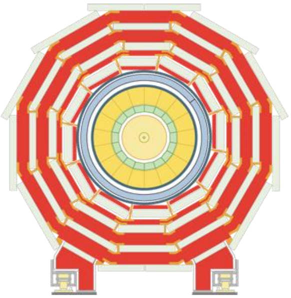
Different detector components sensitive to different types of particles

AXOL1TL takes in reconstructed L1 trigger objects: jets, muons, electron/photons, missing energy



CICADA takes in L1 calorimeter towers

Anomaly Detection in the L1 Trigger



Muon Trigger

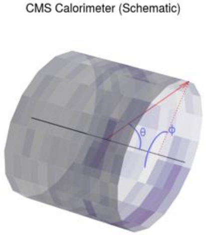
Calorimeter Trigger

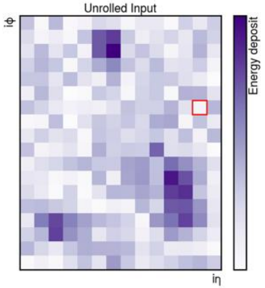
AXOL1TL: L1 trigger object inputs

MET - (p_T, ϕ)
10 jets - (p_T, η, ϕ)
4 muons - (p_T, η, ϕ)
4 electrons / photons - (p_T, η, ϕ)

Global Trigger

L1 Accept / Reject





CICADA image by L. Gerlach

CICADA: L1 calorimeter tower inputs

252 E_T deposits corresponding to 14x18 towers in $\eta \times \phi$

Both algorithms must be lightweight enough to fit within the existing L1 trigger system.

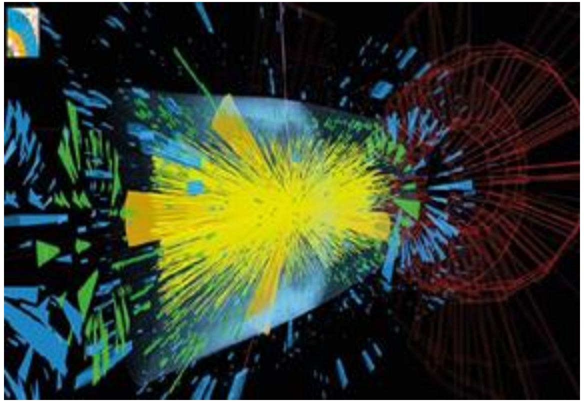
Challenge: Ultrafast Inference

Image B. Ramhorst

~1-3 seconds

~50ms

~100 ns



AXOL1TL and CICADA have to be ultrafast

Deployed Neural Networks on FPGAs

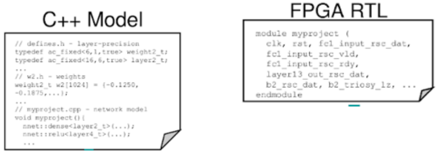
TensorFlow
PyTorch
ONNX
scikit learn
XGBoost

hls4ml

Quantized Model → hls4ml (NN) & Conifer (BDT) → C++/HLS project → FPGA

Quantization and pruning:
QKeras, AutoQ (Keras)
Brevitas (Pytorch)

Conifer

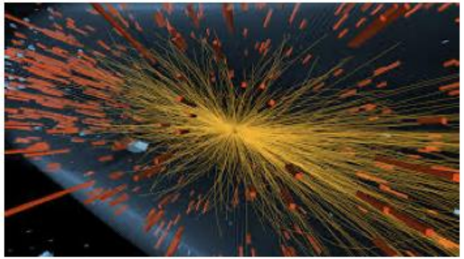


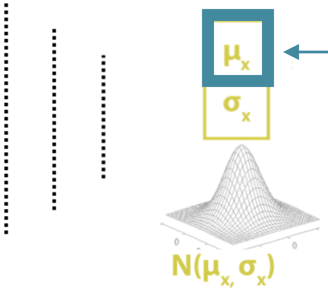
AXOL1TL and CICADA use hls4ml. hls4ml (2018) & conifer (2021) are open source projects primarily developed by LHC community to convert neural networks and boosted decision trees into HLS.

AXOL1TL Implementation

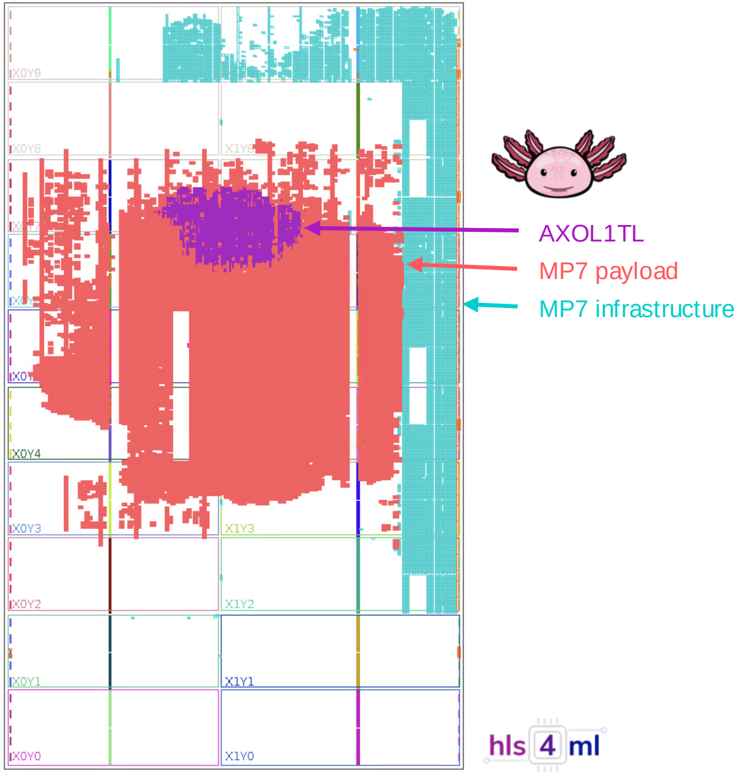
CMS-DP-2023-079

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





← μ^2 term used for anomaly score



Implemented on Xilinx Virtex-7 FPGA

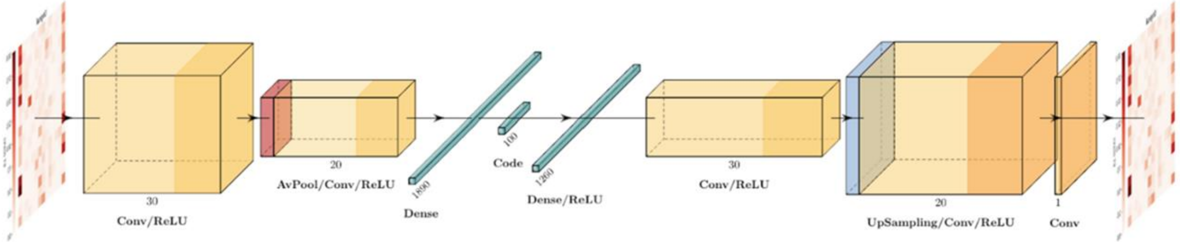
50 ns latency and resource requirements met

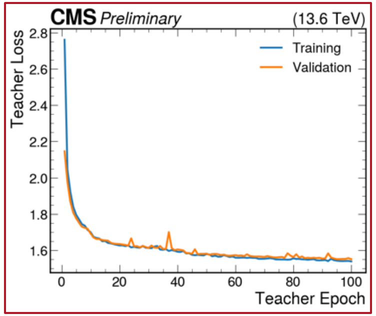
CICADA Implementation

[V. Sharma PPC 2024](#)
[CMS-DP-2024-121](#)

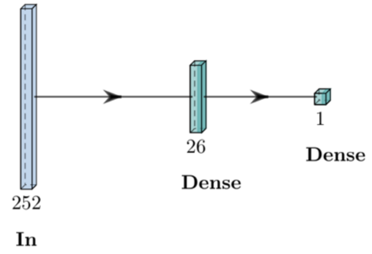
Knowledge distillation: student learns from teacher model

The teacher model

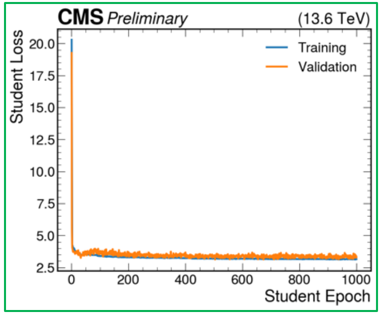




The student model



Inference latency ~ 100 ns
on Virtex-7 FPGA



Status of Anomaly Detection Data-taking

$\int \mathcal{L}\, dt = 47\, fb^{-1}$

$\int \mathcal{L}\, dt = 55\, fb^{-1}$

May 2024

Start of data taking with AXOL1TL

Aug 2024

AXOL1TL model update

Oct 2024

CICADA starts taking data

April 2025

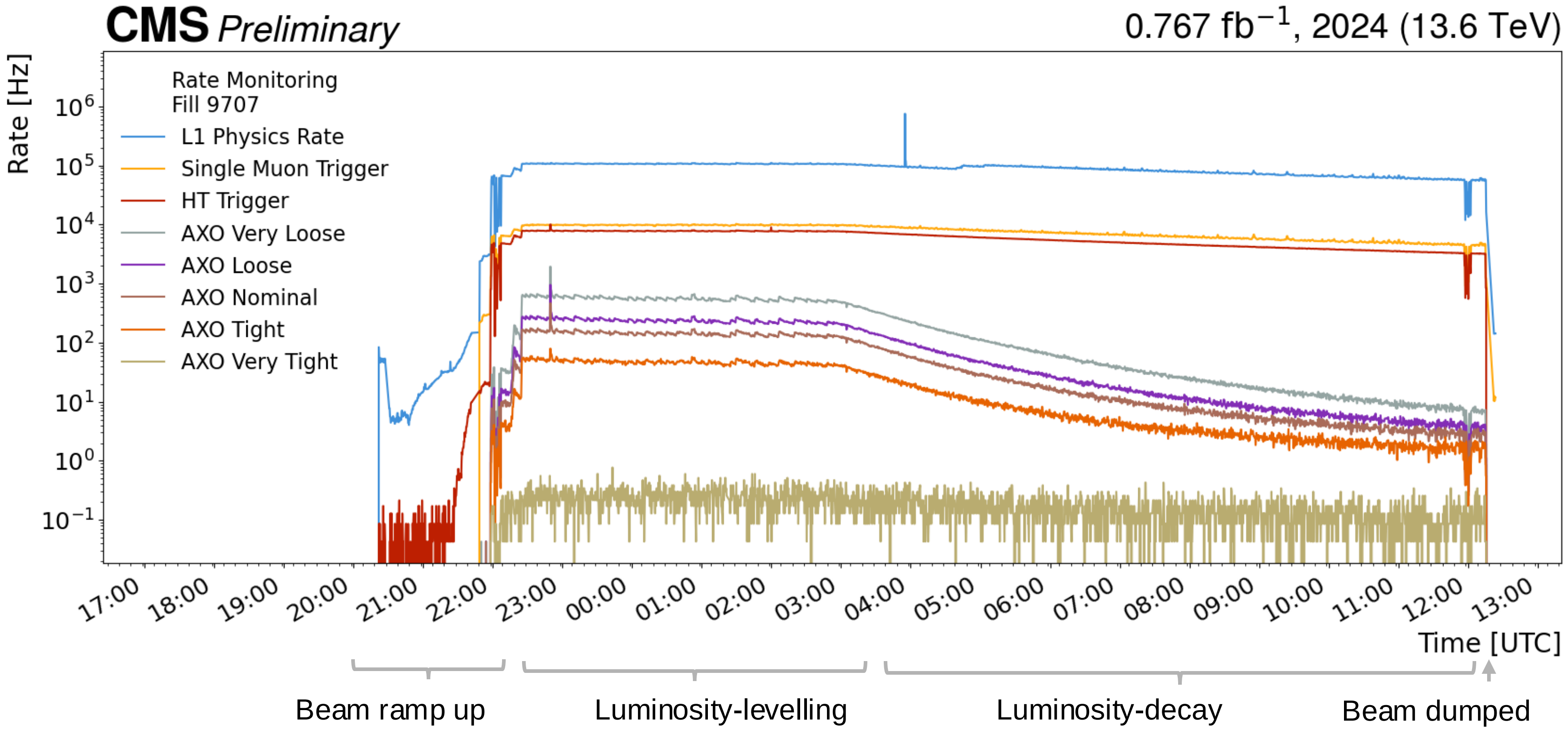
AXOL1TL & CICADA model updates

2025?

ATLAS starts taking data with Anomaly Detection Trigger?

AXOL1TL Rate Stability

CMS-DP-2024-059



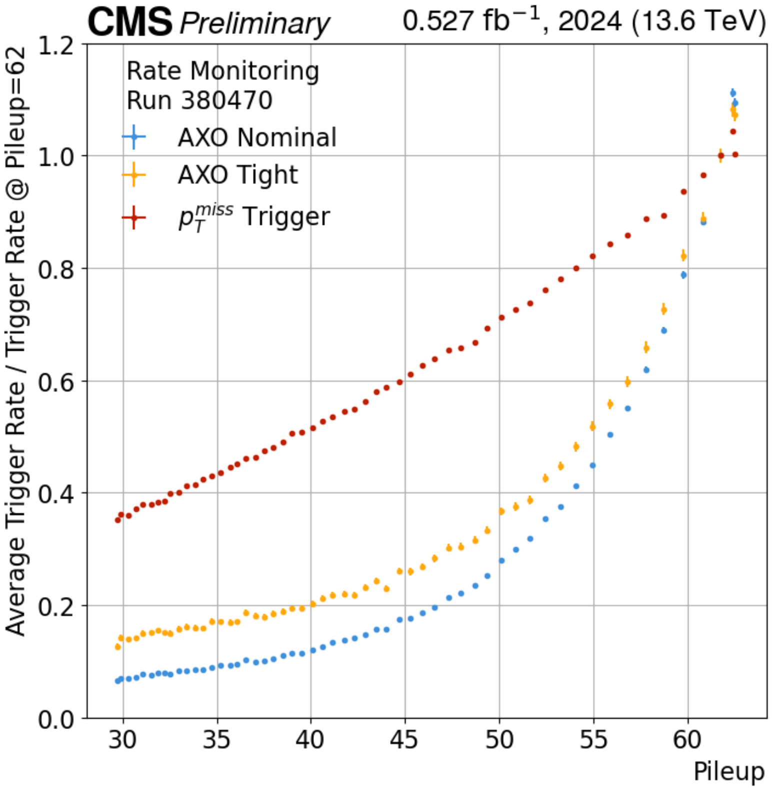
Challenge: Pile-up Dependance

CMS-DP-2024-059

Events with high pileup contain more “information” (more jets, more calorimeter cells) are inherently harder to encode. This contributes to higher rates at high pileup.

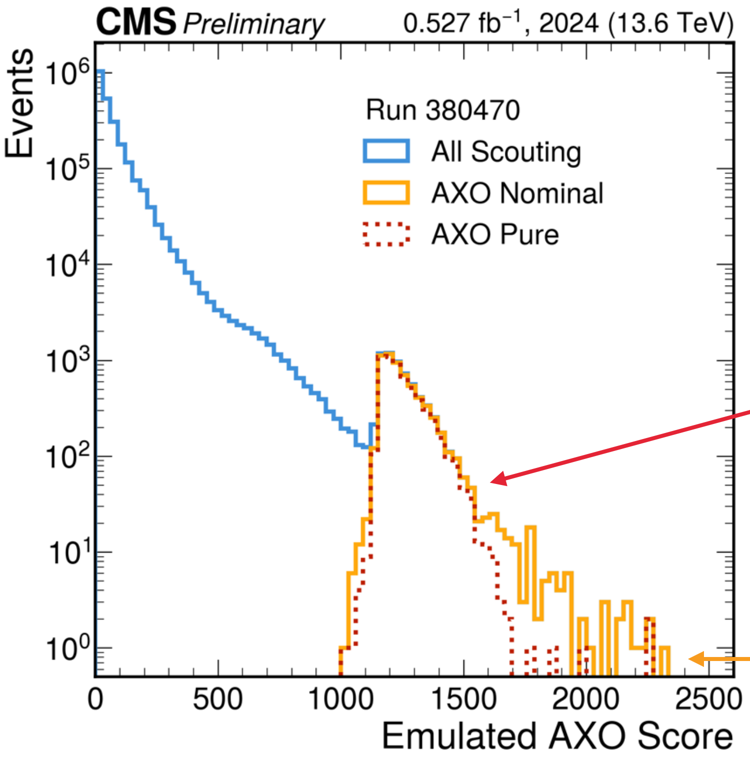
Care must be taken during model development and deployment:

- more robust training procedures
- back-up paths & “emergency off” columns
- conservative rate estimates



AXOL1TL triggered events

CMS-DP-2024-059



Large fraction of unique events recorded that would otherwise be rejected

High anomaly score events, also triggered by existing L1 trigger

Challenge: Data / MC agreement

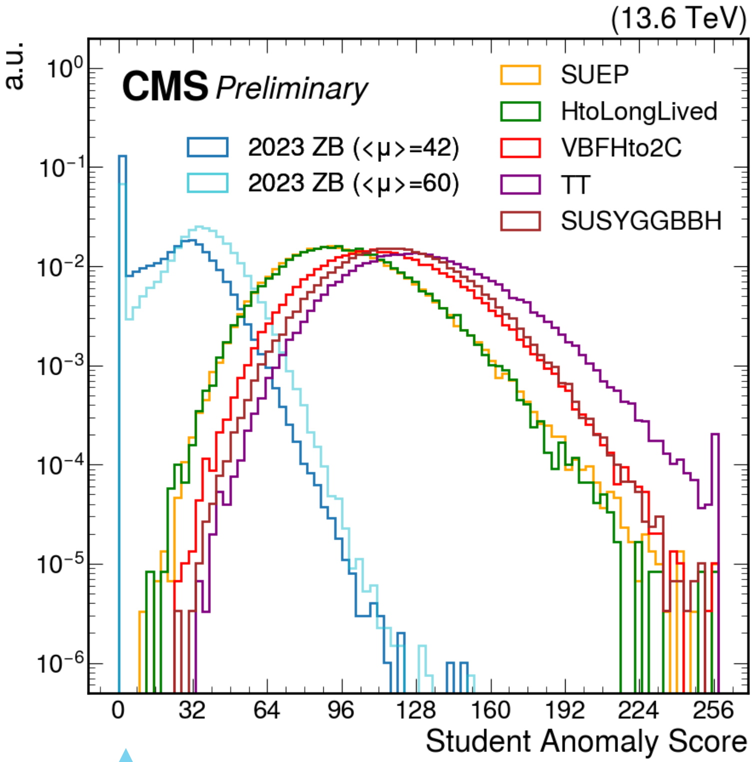
[CMS-DP-2023-079]

"But how much do you trust your Monte Carlo?"

Opposite problem of most ML models deployed at the LHC which are trained on MC then deployed on data.

We train on data and evaluate on MC.

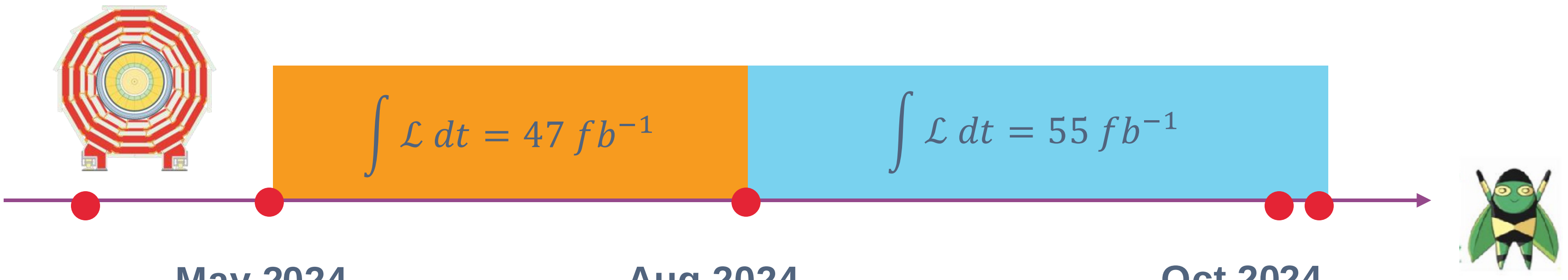
Developing processes for studying trigger efficiencies in data and evaluating on standard candles.



Zero Bias data events have low anomaly score

Signal Monte Carlo has high anomaly score

Challenge: Tracking model updates - MLOps



May 2024

Start of data taking with AXOL1TL

Aug 2024

AXOL1TL **model update**

Oct 2024

CICADA starts taking data

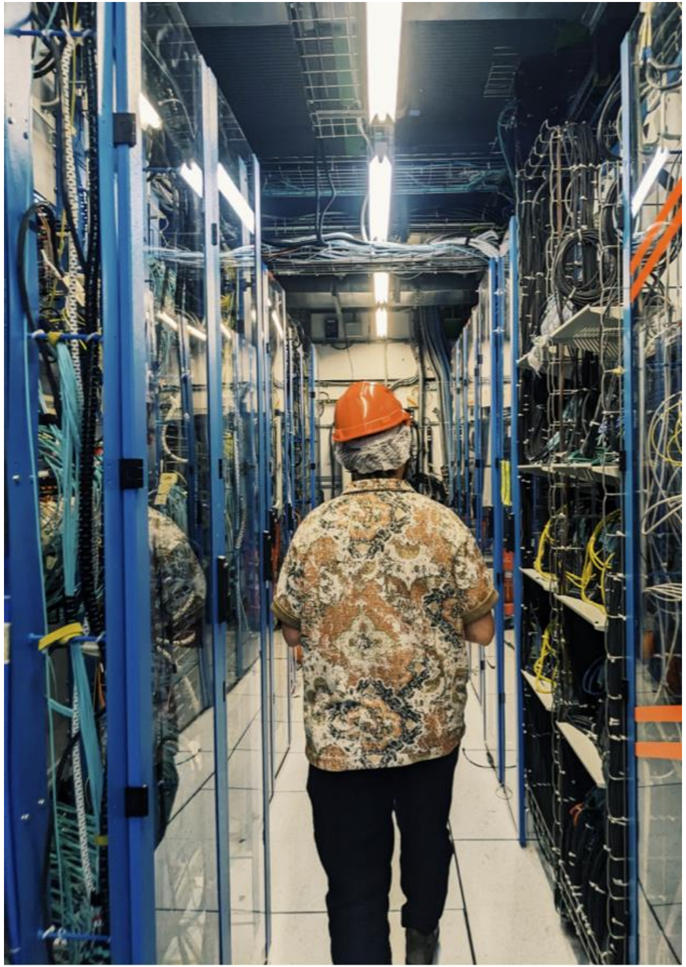
April 2025

AXOL1TL & CICADA **model updates**

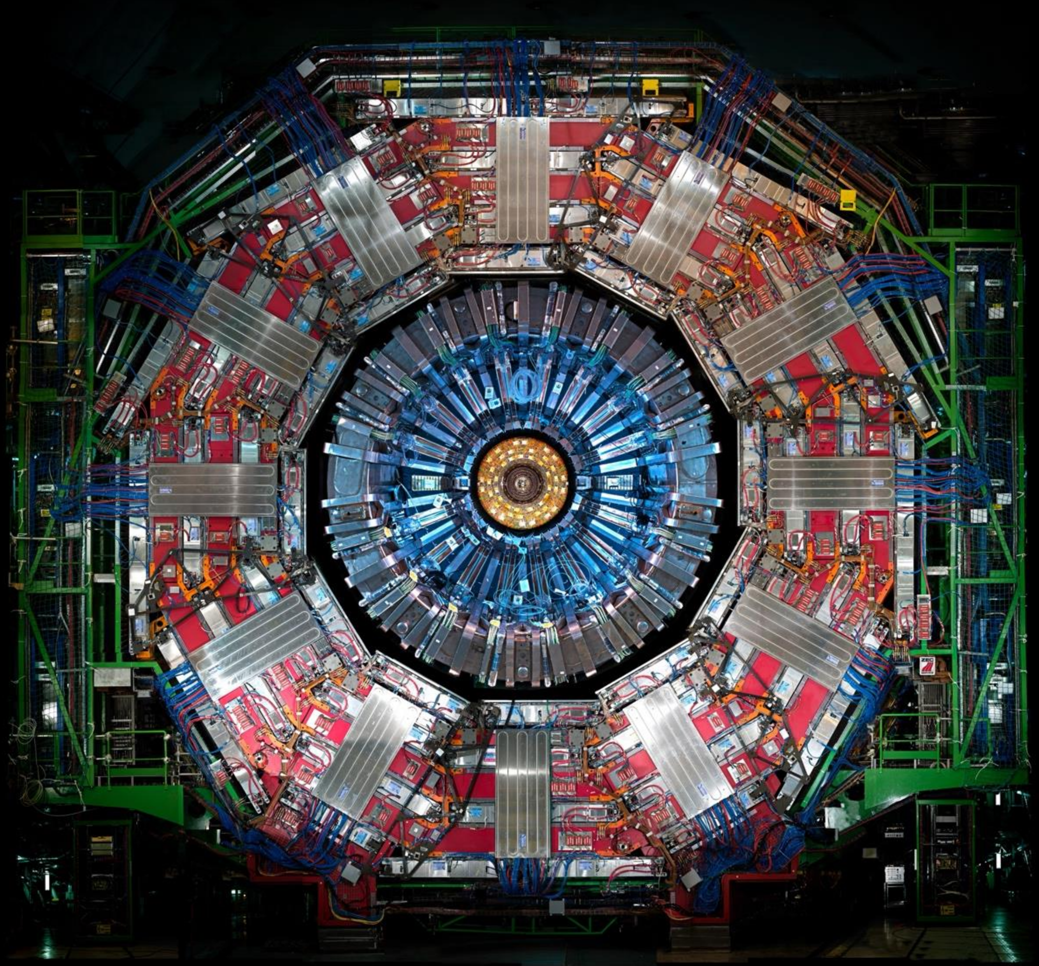
- Want to be able to retrain for new detector conditions and update models often
- For analysis, it is essential to store, track and be able to re-emulate all deployed models
- Experiences **now** will be invaluable at HL-LHC where L1 models are expected to contribute **25 billion inferences / second in CMS**

Summary

- CMS has been taking data with AXOL1TL and CICADA since 2024
- The are the first ML-based anomaly detection triggers deployed at the LHC
- Have had to overcome numerous challenges during development and deployment
- Currently analyzing data collected in 2024 and developing improved algorithms for the future
- This is just the very beginning!



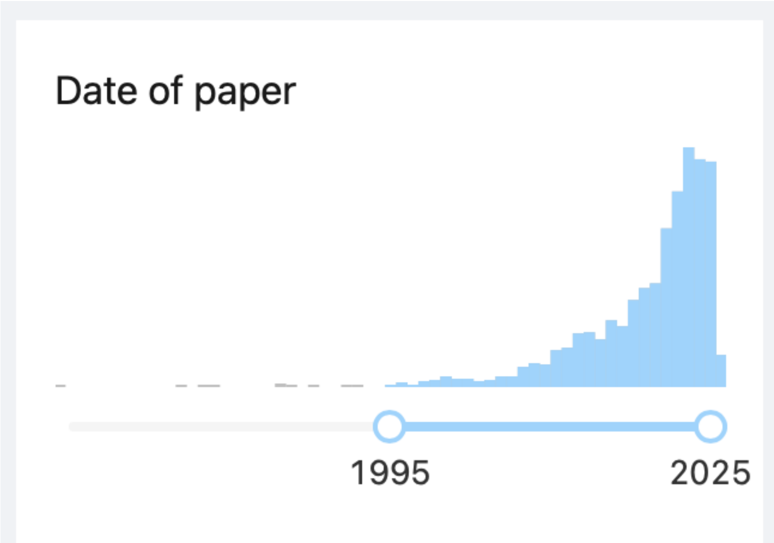
Thank you!



Anomaly Detection

>1000 papers on “Anomaly Detection” in INSPIRE HEP search

Majority posted in the past 3 years

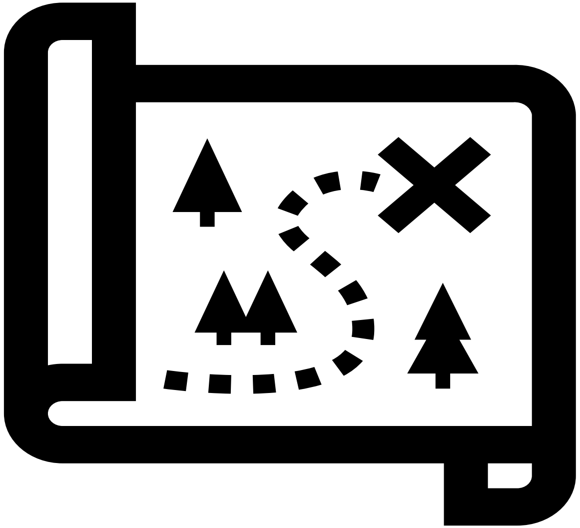


Applications:

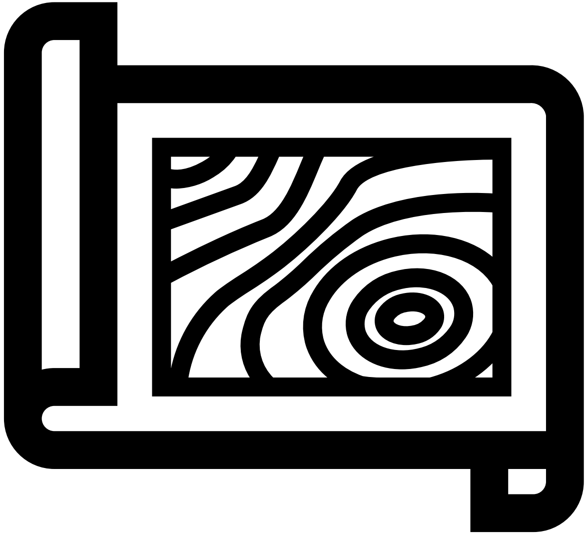
- New physics searches
- Data quality monitoring
- Data compression
- Triggering!

Searching for new physics

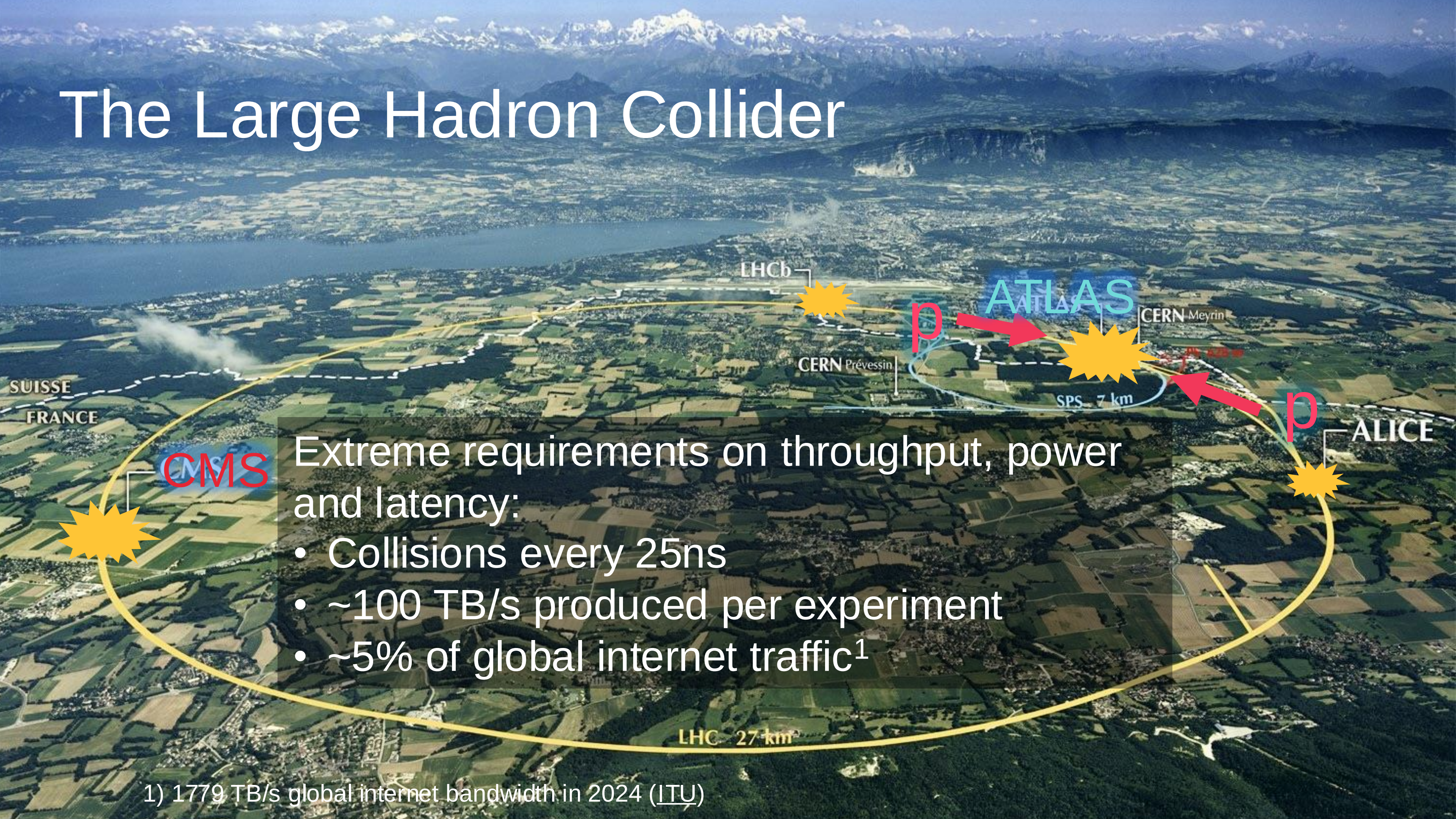
Less like this



And more like this



The Large Hadron Collider

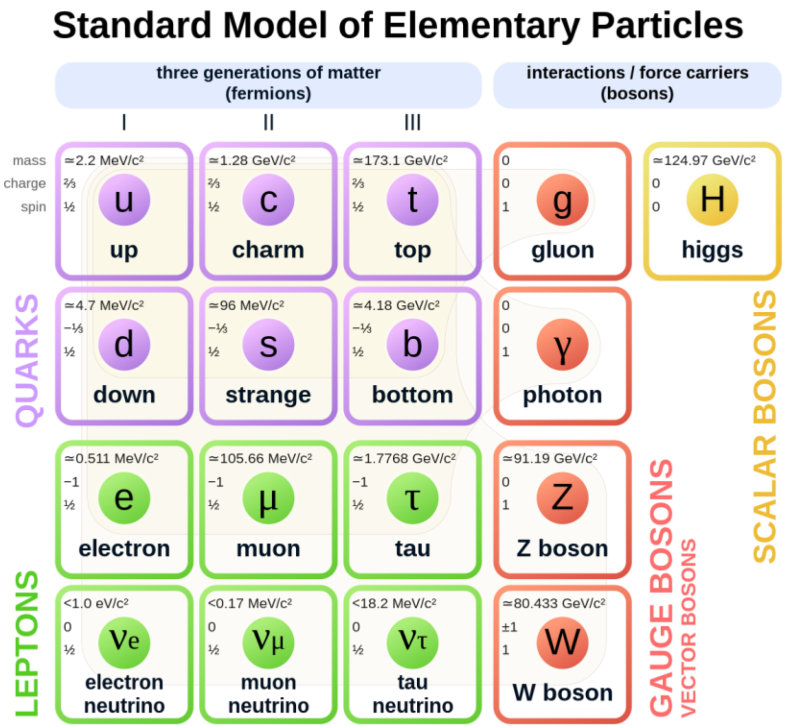


Extreme requirements on throughput, power and latency:

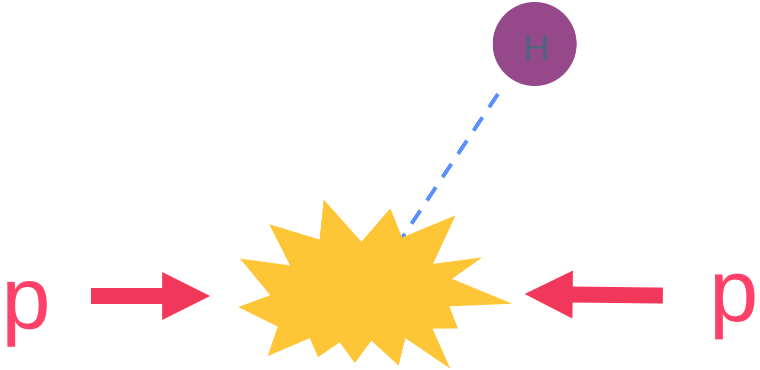
- Collisions every 25ns
- ~100 TB/s produced per experiment
- ~5% of global internet traffic[1]

1) 1779 TB/s global internet bandwidth in 2024 (ITU)

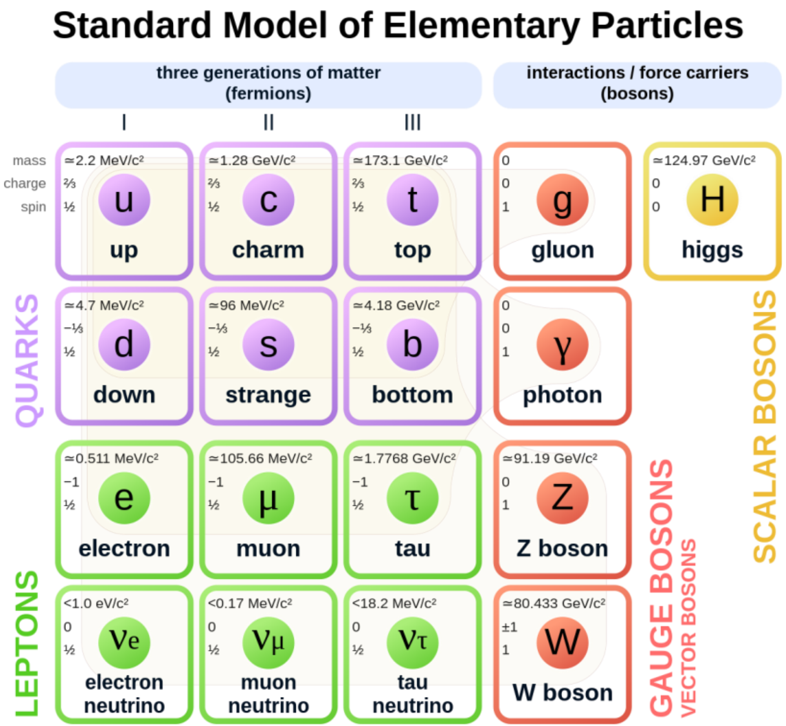
What we study at the LHC



Dark matter?

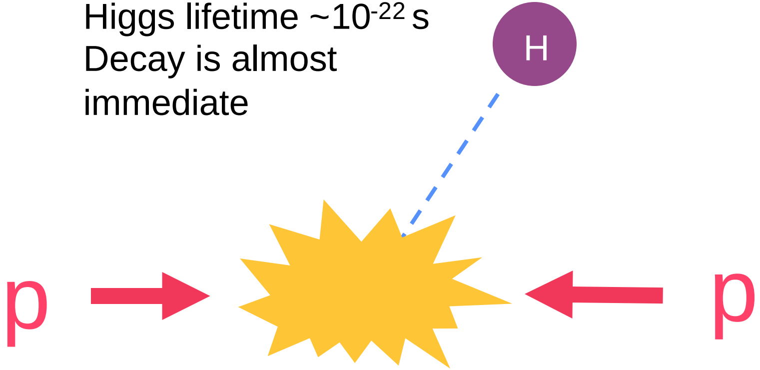


What we study at the LHC

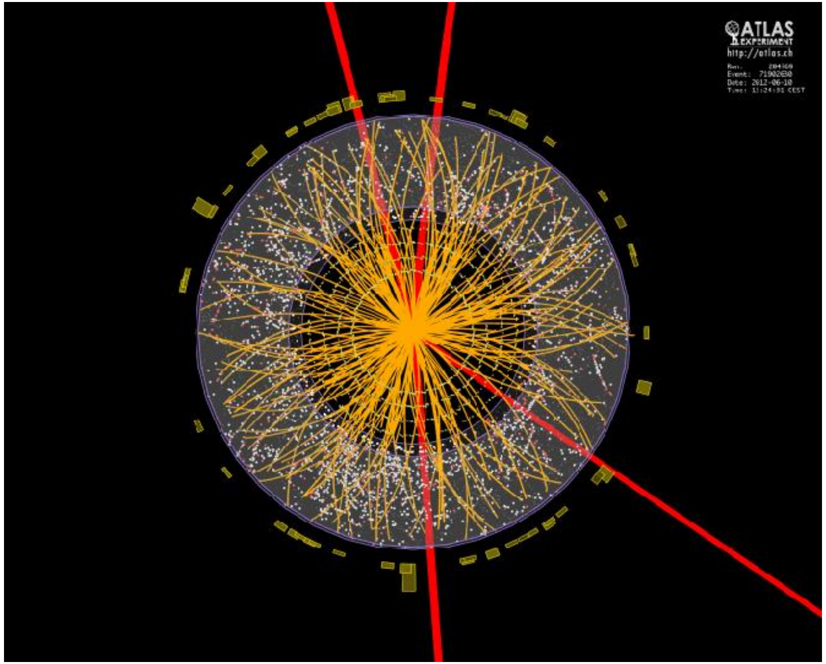


Dark matter?

Higgs lifetime ~10^{-22} s
Decay is almost immediate

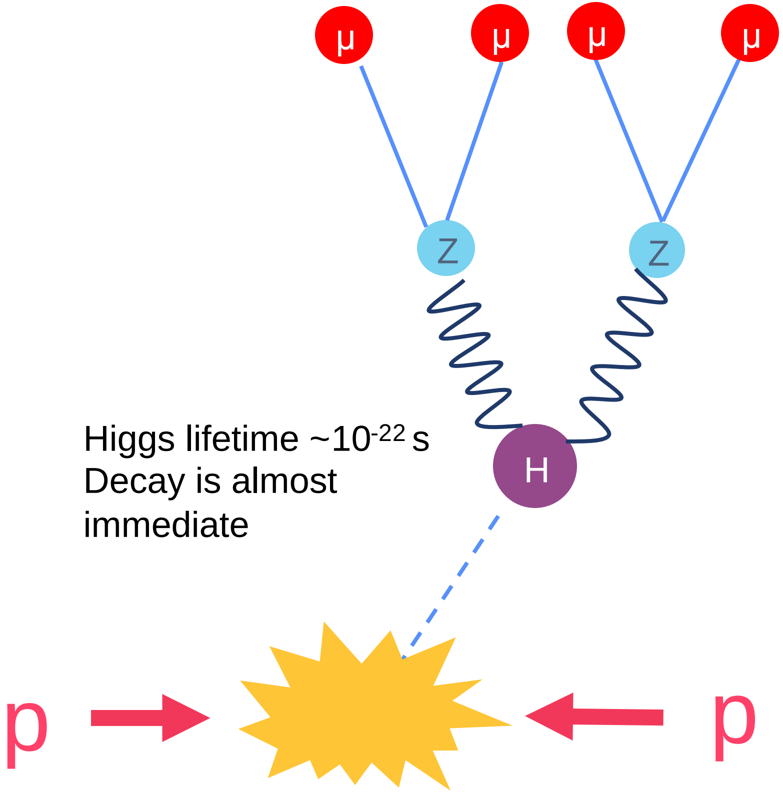


What we study at the LHC



- Detector takes a snapshot of collisions
- We work backwards to infer what happened

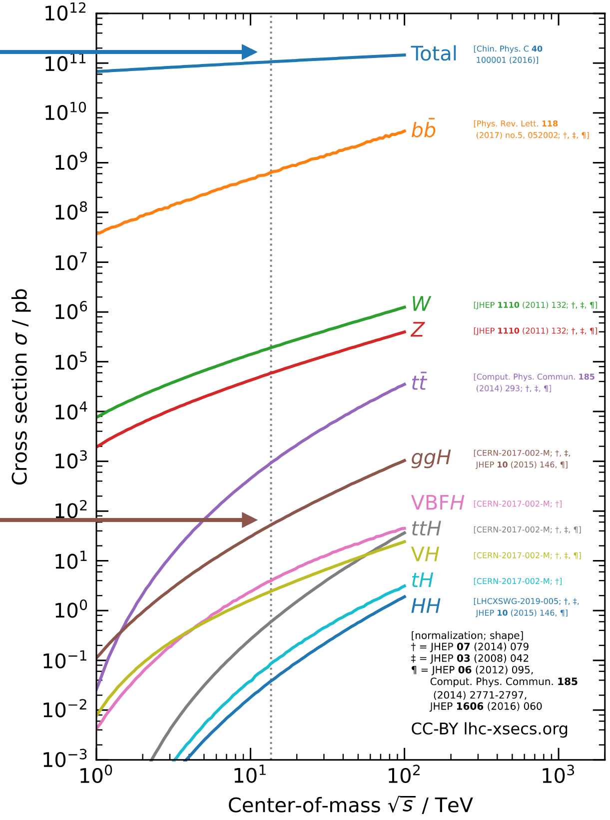
Higgs lifetime ~10^{-22} s
Decay is almost immediate



Why are our data rates so high?

“Probability” of
producing a collision

~1 Higgs boson per
billion collisions



A lot of the physics we are interested in is *very rare.*

AXOL1TL and CICADA

AXOL1TL

<u>AXOL1TL L1 trigger objects are inputs:</u>

MET - (p_T, ϕ)

10 jets - (p_T, η, ϕ)

4 muons - (p_T, η, ϕ)

4 electrons / photons - (p_T, η, ϕ)

Potentially more sensitive to processes with muons in the final state.

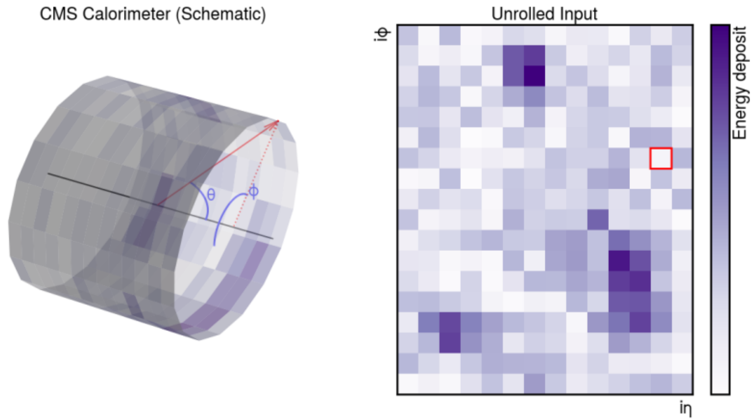
Very similar HLT strategy for both triggers.

CICADA

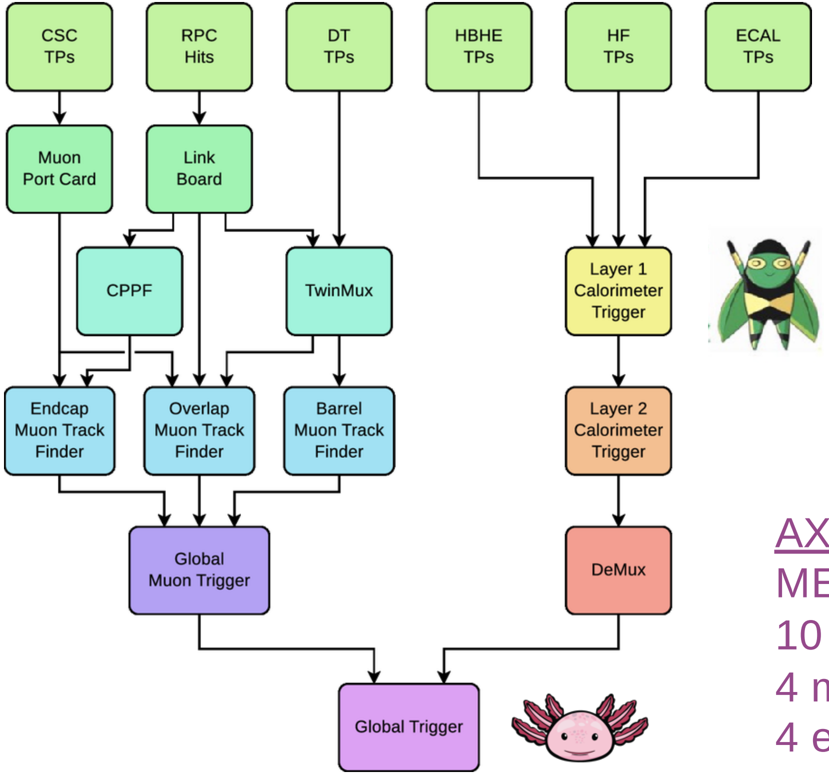
<u>CICADA L1 calorimeter towers are inputs:</u>

252 E_T deposits corresponding to 14x18 towers in ηxϕ

Potentially more sensitive to processes with interesting jet substructure.



AXOL1TL & CICADA in the L1 Trigger



CICADA L1 calorimeter towers are inputs:
252 E_T deposits corresponding to 14x18 towers in ηxϕ

AXOL1TL L1 trigger objects are inputs:
MET - (p_T, ϕ)
10 jets - (p_T, η, ϕ)
4 muons - (p_T, η, ϕ)
4 electrons / photons - (p_T, η, ϕ)

Data Streams

L1T

AXOL1TL

5 thresholds:
L1_AXO_VLoose (1 kHz)
L1_AXO_Loose (500 Hz)
L1_AXO_Nominal (300 Hz)
L1_AXO_Tight (100 Hz)
L1_AXO_VTight (10 Hz)

HLT

HLT Scouting passthrough

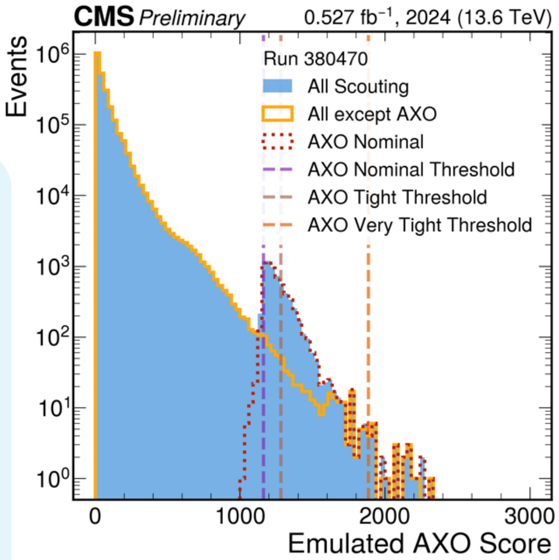
HLT Scouting Datset

DST_PFScouting_AXONominal (300 Hz)
DST_PFScouting_AXOTight (100 Hz)
DST_PFScouting_AXOVTight (10 Hz)

Full reconstruction passthrough

JetMET dataset

AXO_VTight (10 Hz)



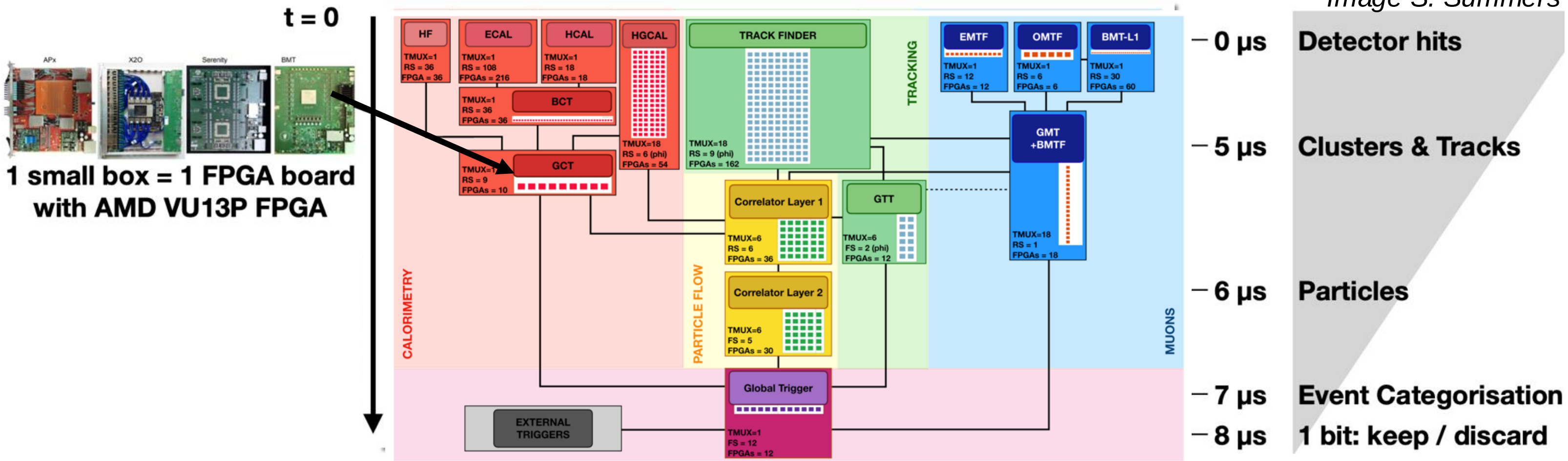
Rates shown are NNv3 target rates

Passthrough trigger: Anything accepted at L1 is also accepted by the HLT

HL-LHC Upgrades to the CMS L1 Trigger

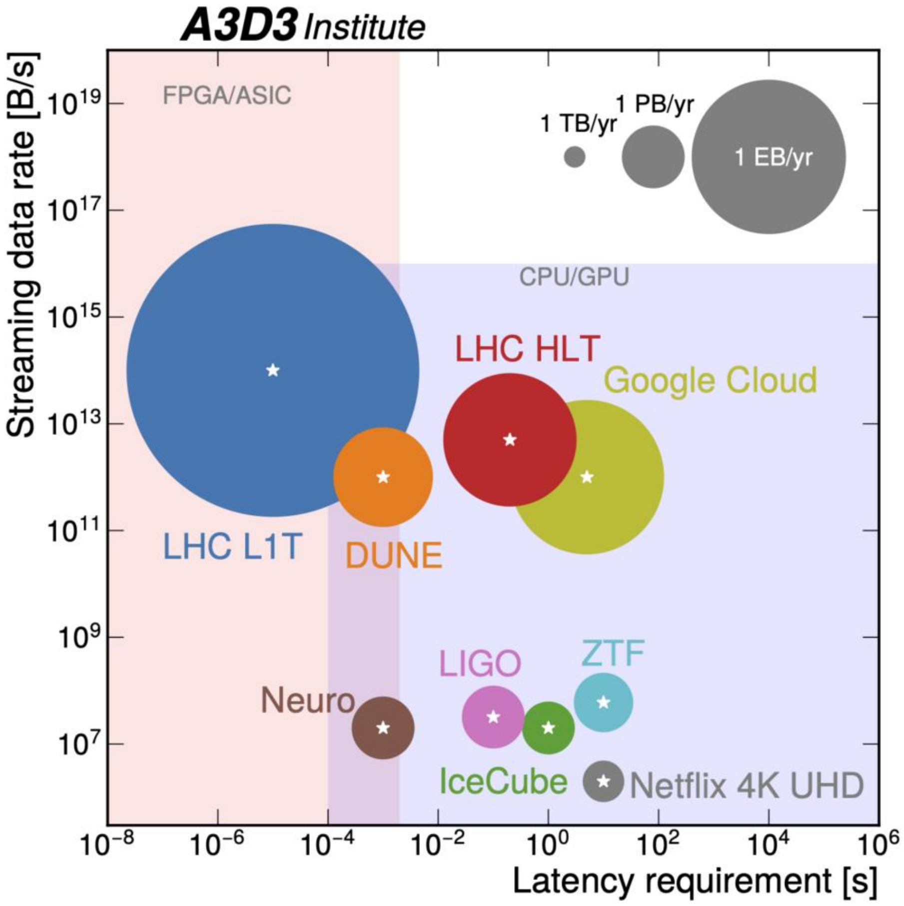
- Machine Learning heavily incorporated into upgraded L1 trigger design
- Anticipate **25 billion inferences/s** from ML models

Image S. Summers



- **MLOps challenge**:
  Tracking and monitoring deployed models, and adapting to changing environments

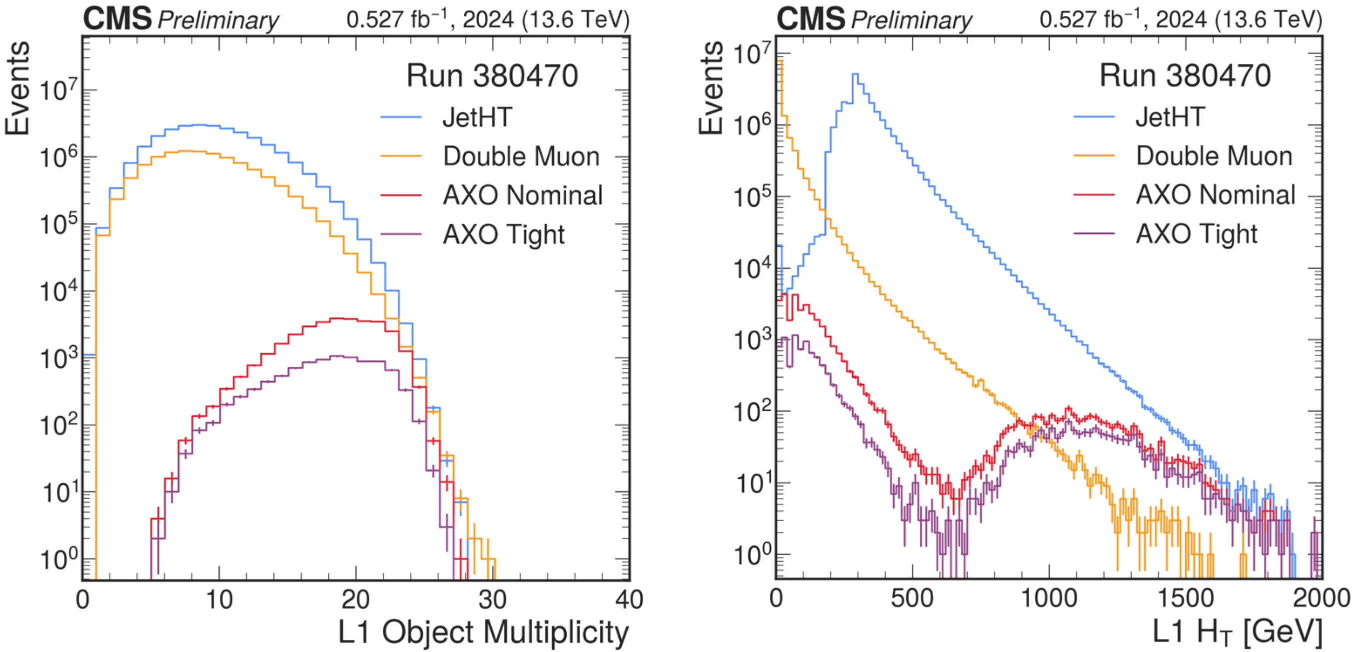
LHC Rates & Latency Reqs.



What does AXOL1TL trigger on?

CMS-DP-2024-059

High object multiplicity and total transverse momentum.



Peak at AXO triggered data: □

- Smoothly falling mass distributions shown here in small fraction (<1%) of the 2024 data
- More plots in our DP Note: https://twiki.cern.ch/twiki/bin/view/CMSPublic/AXOL1TL2024