Anomaly Detection in CMS

12th Large Hadron Collider Physics Conference

Amandeep Kaur (Purdue University)
on behalf of CMS Collaboration

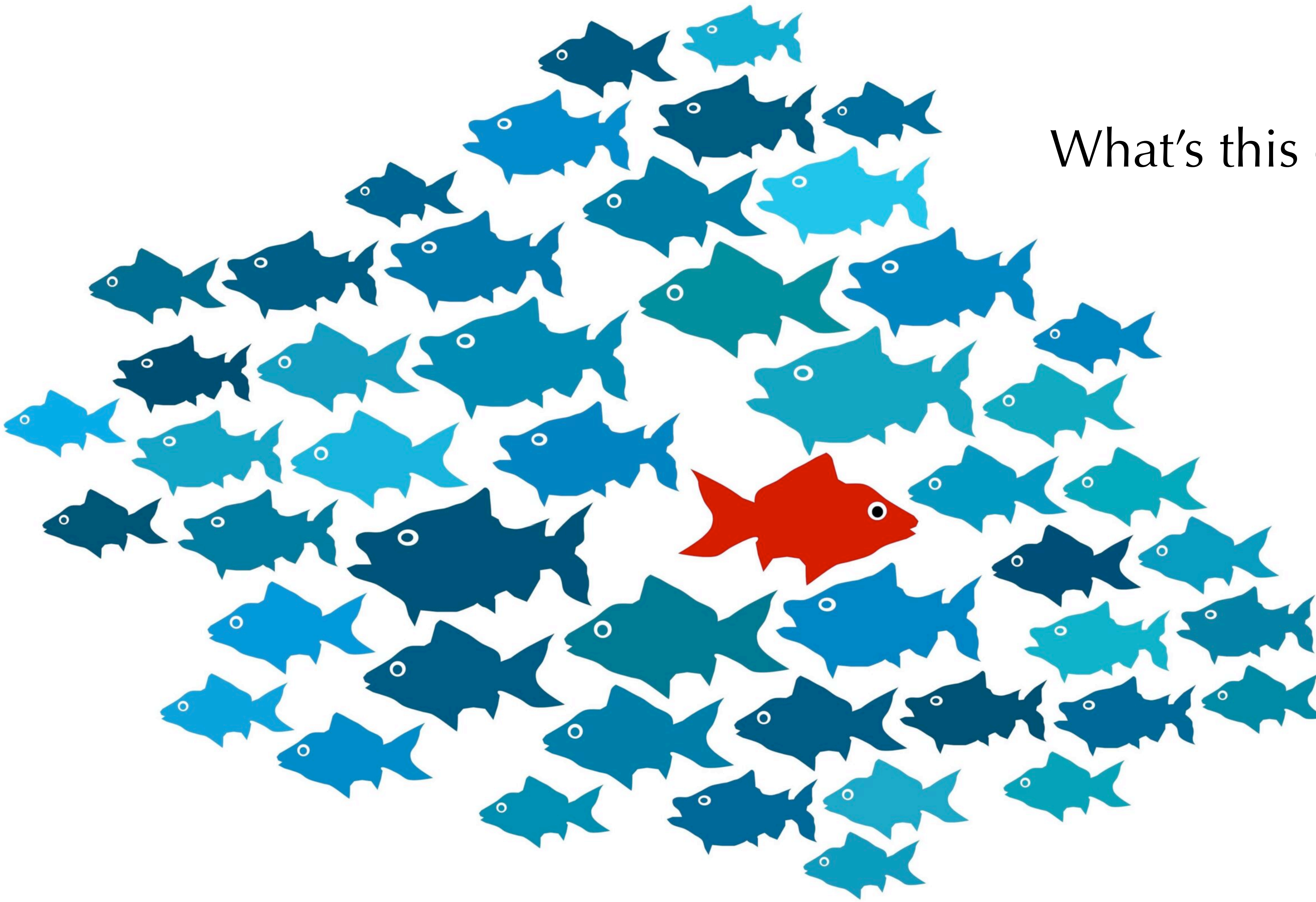
June 5, 2024

LHCP Boston 2024

CMS

PURDUE UNIVERSITY

CMS & anomaly detection

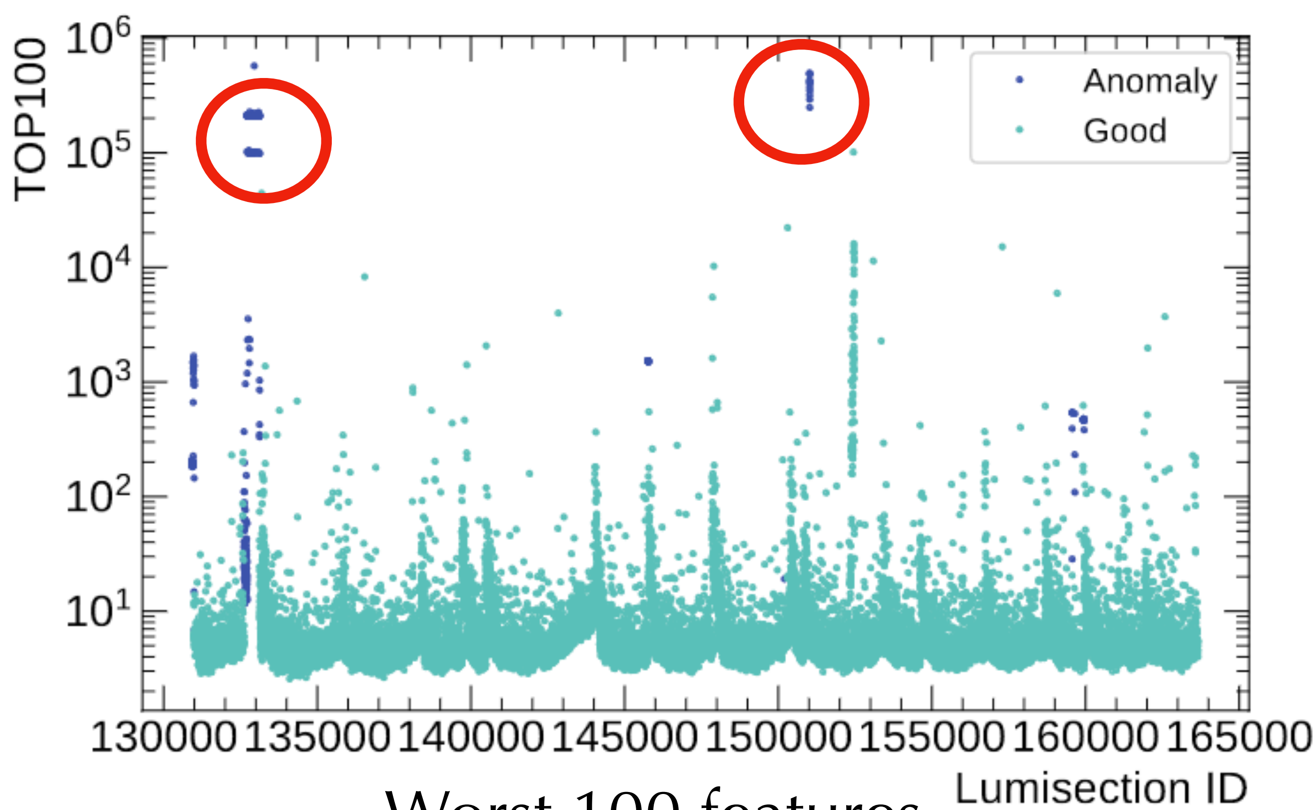


What's this exotic fish?

CMS & anomaly detection

- Anomaly detection methods are in practice since 2016 for data quality monitoring and data certification @ CMS
- Proven to be additional quality indicator in the current certification framework
- Filters the work for human experts

Credits: doi.org/10.1051/epjconf/201921406008

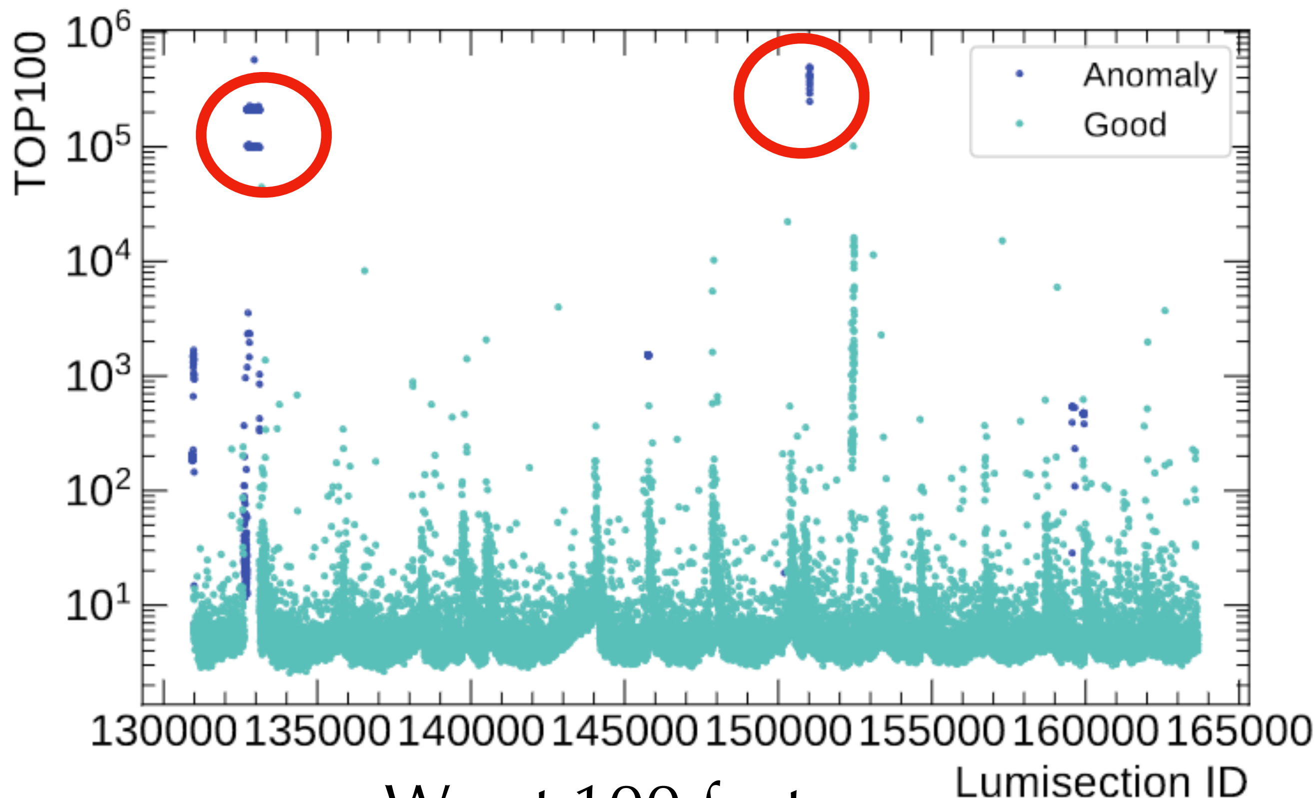


Worst 100 features

CMS & anomaly detection

- Anomaly detection methods are in practice since 2016 for data quality monitoring and data certification @ CMS
- Proven to be additional quality indicator in the current certification framework
- Filters the work for human experts

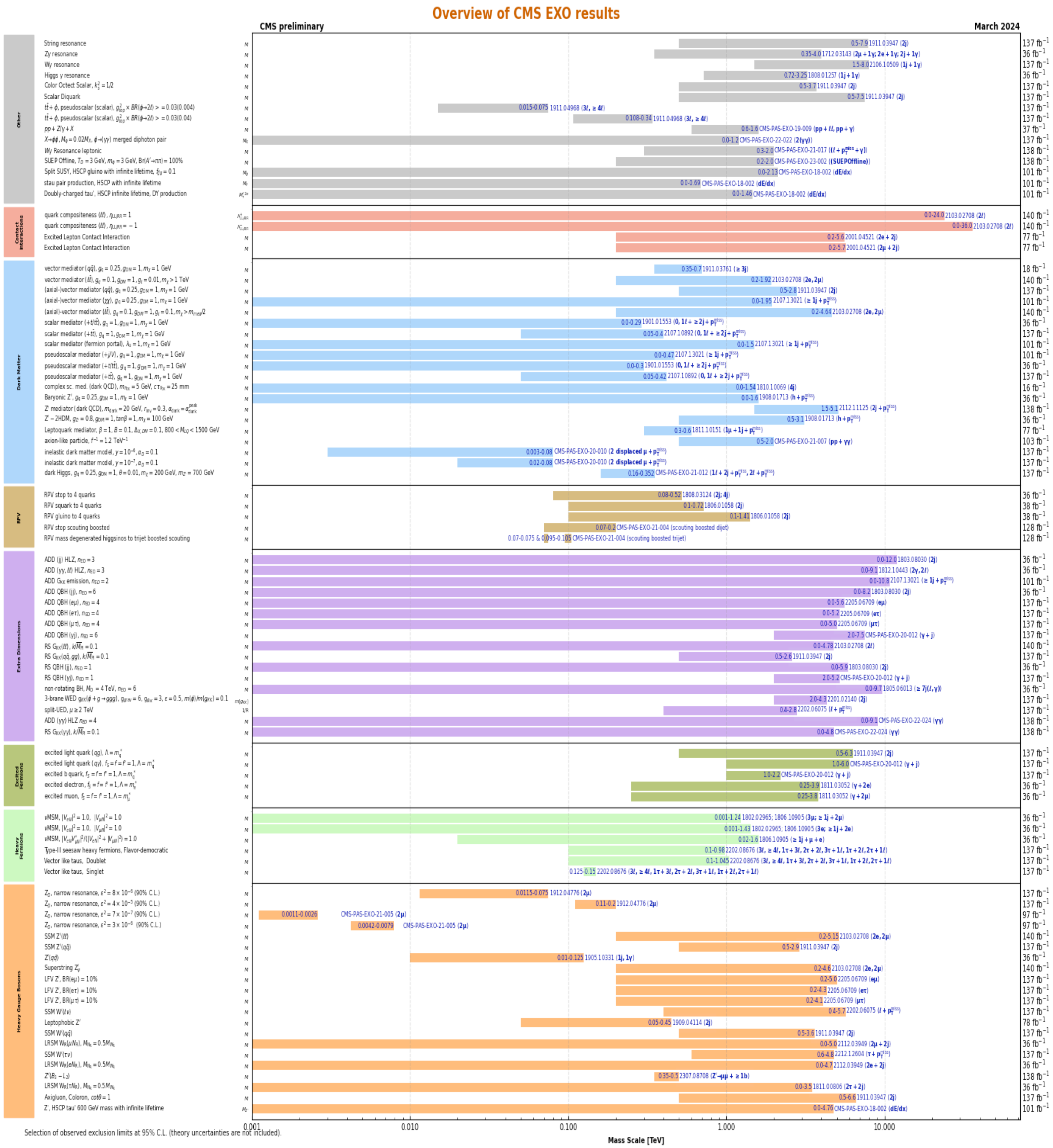
Credits: doi.org/10.1051/epjconf/201921406008



Worst 100 features

Incorporated such techniques in physics analyses to increase discovery potential and look for new exotic signals

New physics search @ CMS

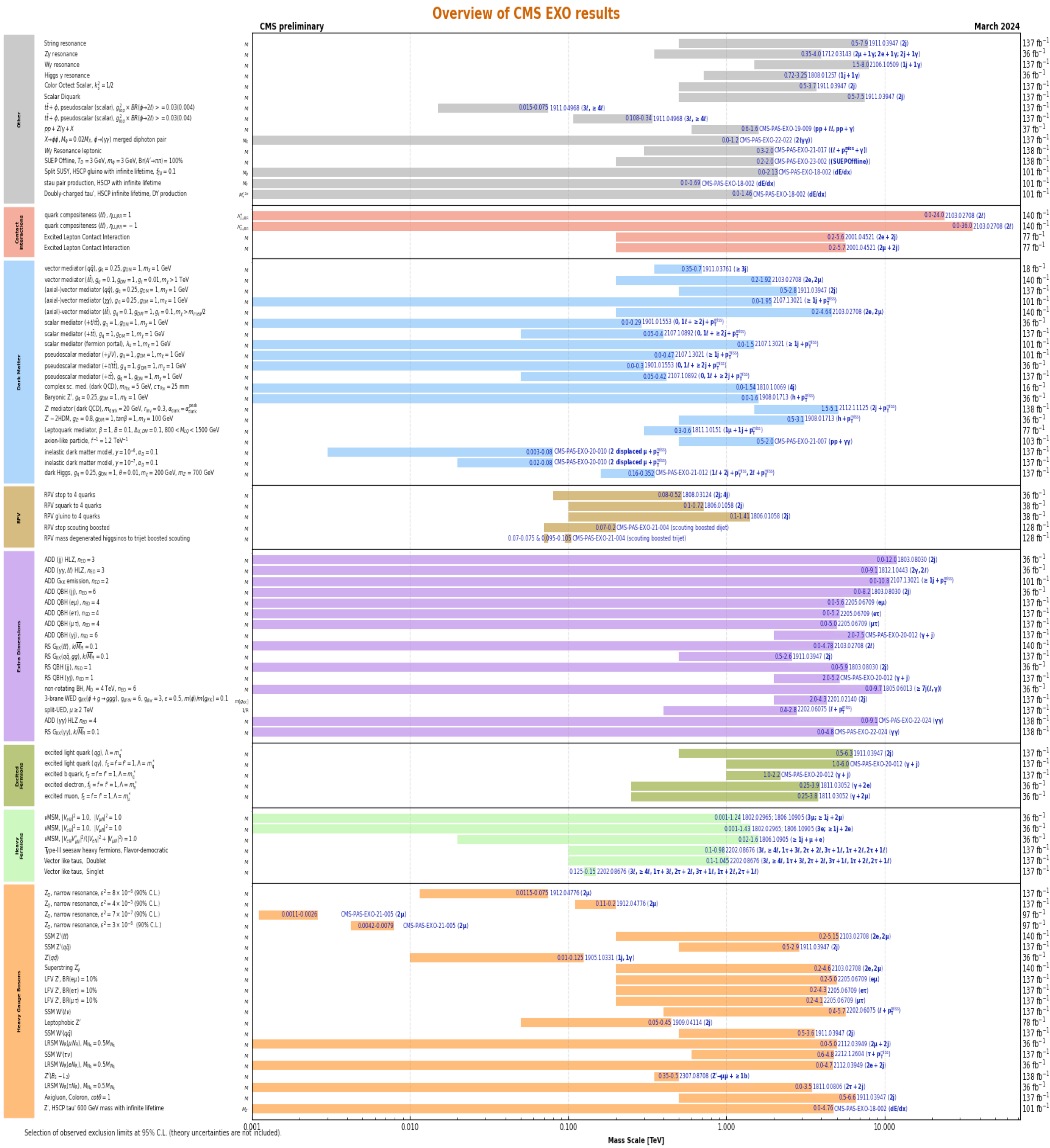


CMS EXO'24 Summary Plot

- O(1000) CMS searches targeting different physics models
- Most searches target a specific signal phase space
- Performed signal model independent search, but are background model dependent

Can we perform signal and background model independent search?

New physics search @ CMS



CMS EXO'24 Summary Plot

- O(1000) CMS searches targeting different physics models
- Most searches target a specific signal phase space
- Performed signal model independent search, but are background model dependent

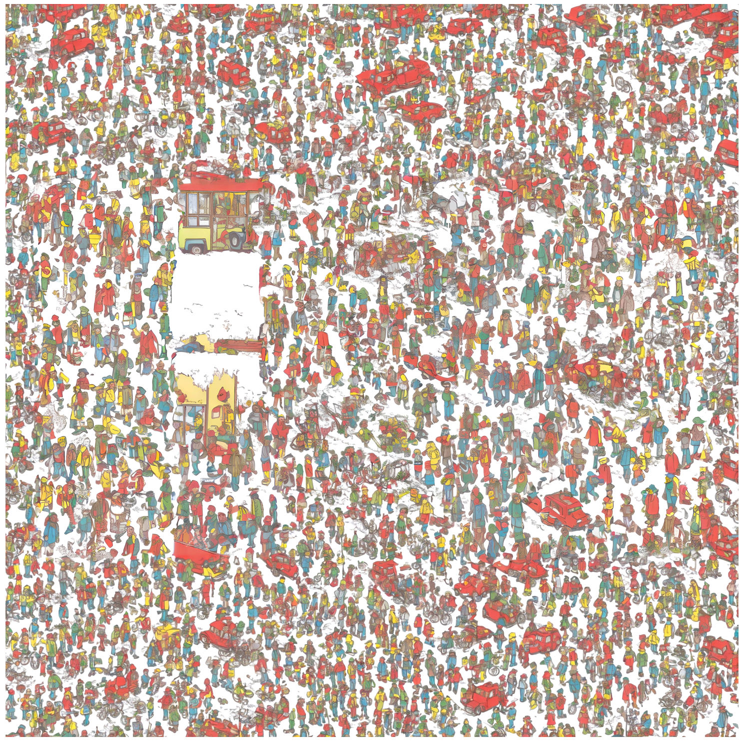
Can we perform signal and background model independent search?

Yes!!

Designing the search

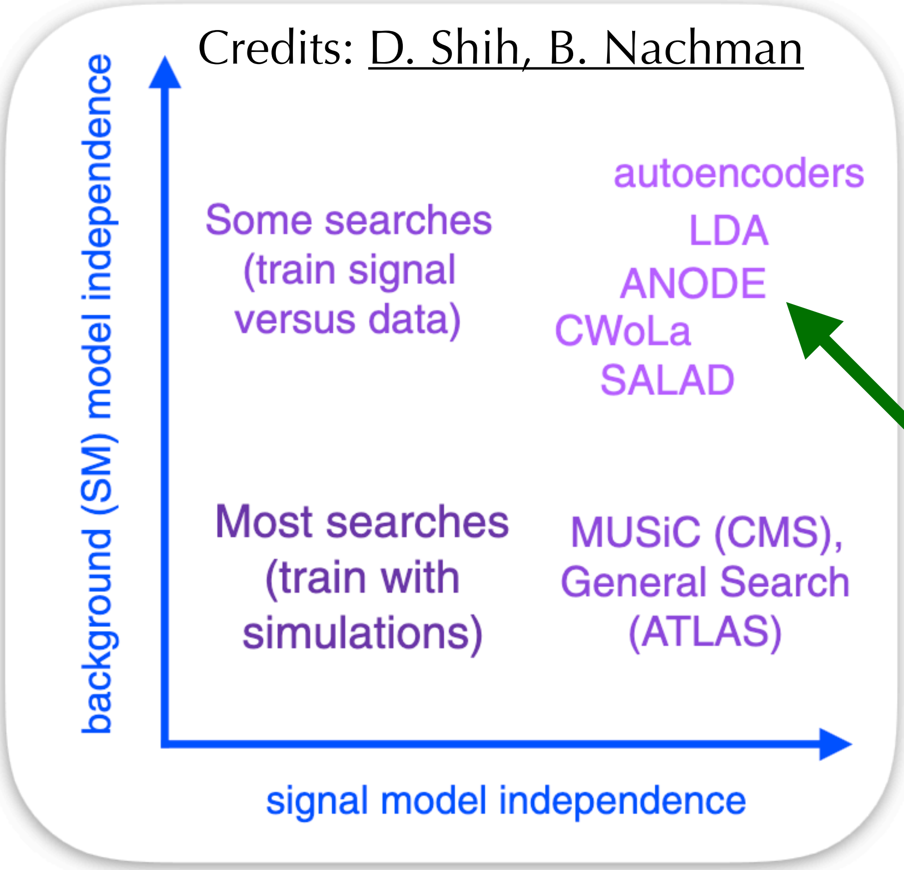
- Best approach, take data and start eliminating the known i.e. standard model
- Use machine learning to learn from data itself how the standard model looks like → complete signal & background model independent search
- **Search for any anomaly w.r.t standard model**

Courtesy: Where's Waldo!

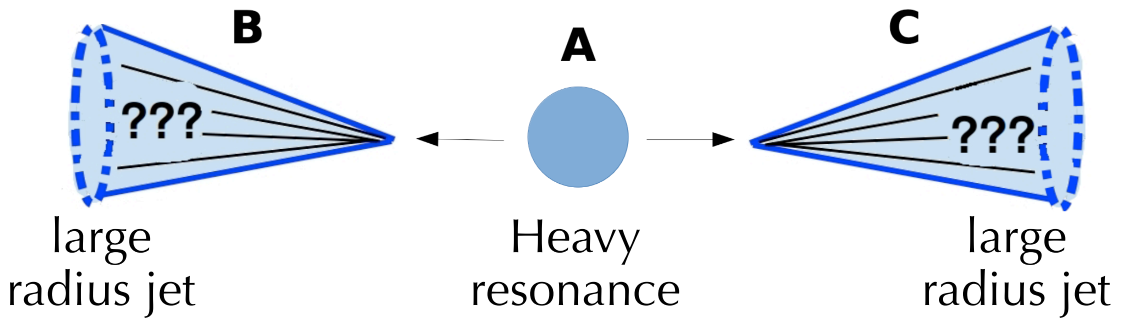


Fun Fact : It's "Where's Waldo", except we don't know what it looks like?

Credits: D. Shih, B. Nachman

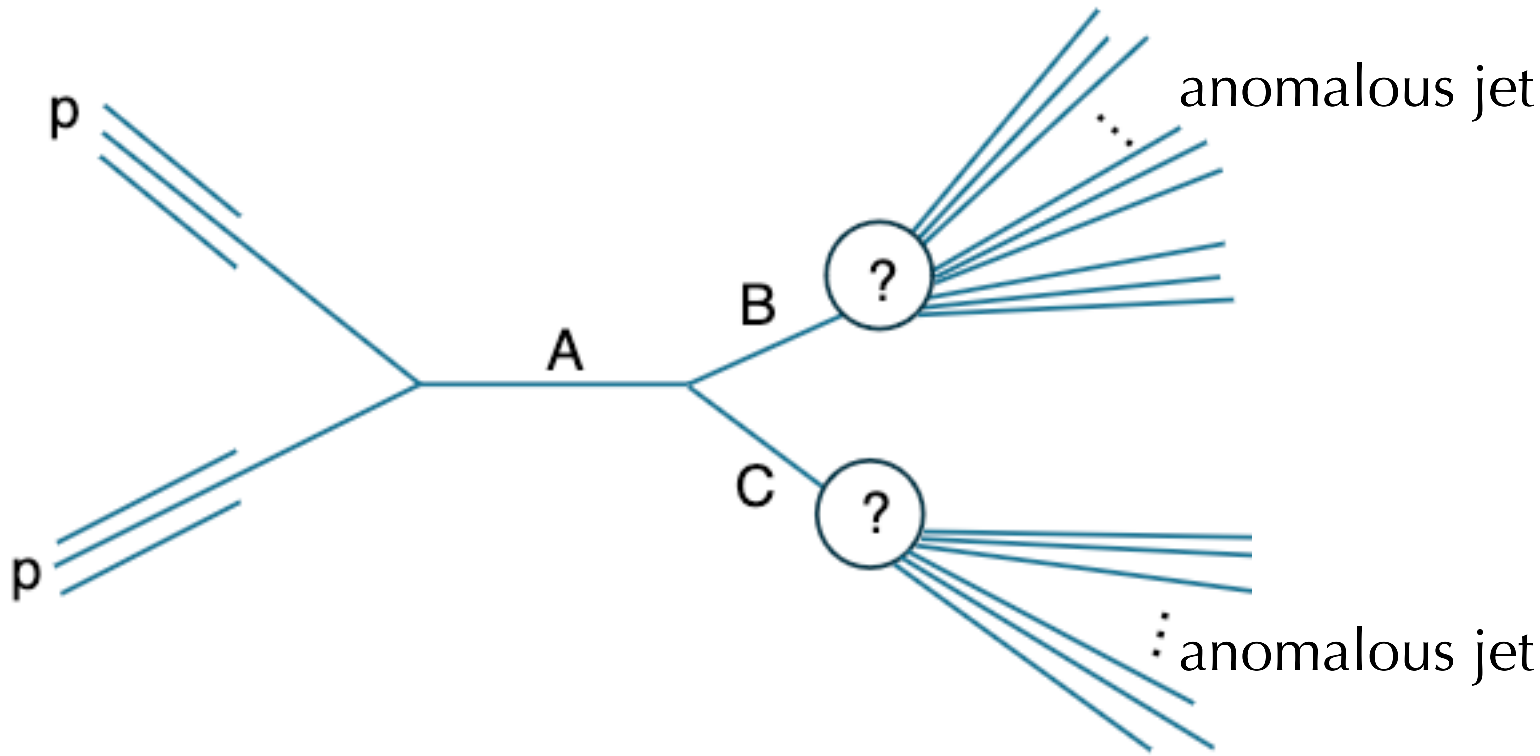


Physics case @CMS dijet anomaly search



CMS-PAS-EXO-22-026

Dijet resonance anomaly search



How to identify these anomalous jets?

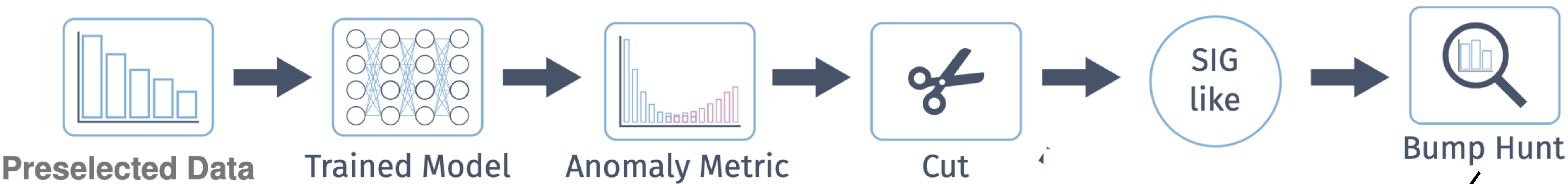
Different assumptions → Complementary approaches

- Used five AI methods based on three different training algorithms

Learn regular jets and look for outliers	Train and Classify signal vs. background on data	Encode a prior of new physics, look for similar
Variational Autoencoder (VAE)	Classification w/o Labels (CWoLa) Tag N′ Train (TNT) CATHODE	Quasi Anomalous Knowledge (QUAK)

Increasing model dependence →

Dijet resonance anomaly search



2 large radius jets

five different anomaly detection methods to ensure robust coverage

Look for an excess over smoothly falling SM background

Dijet resonance anomaly search

Input features to each methods

Increasing model dependence

VAE ← p_x, p_y, p_z of each particle

CWoLa ← $m_{SD}, \tau_{21}, \tau_{32}, \tau_{43}, \eta_{const}$, leptonic energy fraction, sub-jets B tag score

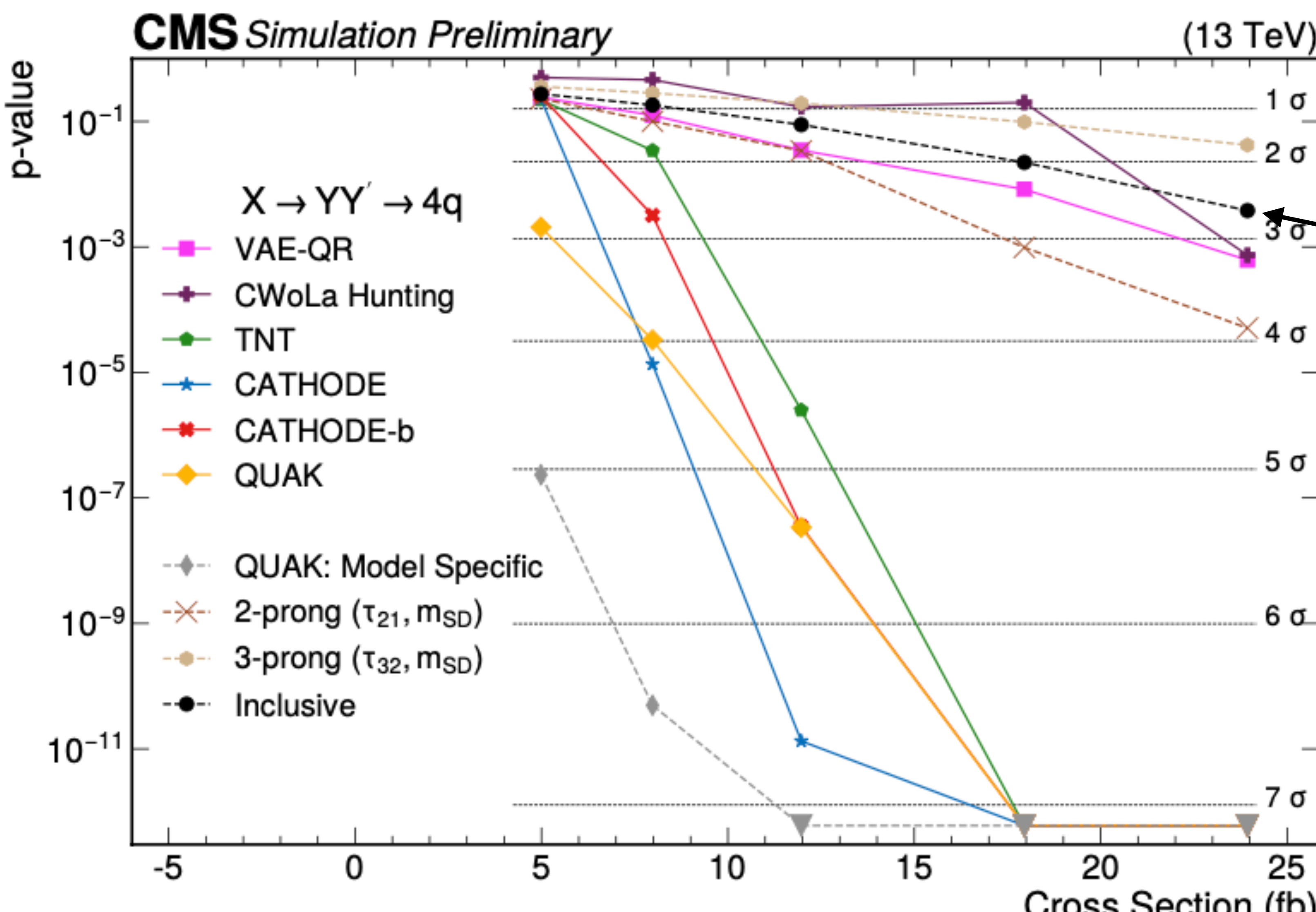
TNT ← $m_{SD}, \tau_{21}, \tau_{32}, \tau_{43}, \eta_{const}$, leptonic energy fraction, sub-jets B tag score

CATHODE ← $m_{SD}^{j1}, m_{SD}^{j1} - m_{SD}^{j2}, \tau_{41}^{j1}, \tau_{41}^{j2}$, B tag score j1

QUAK ← $\rho = m_{SD}/p_T\,, \tau_{21}, \tau_{32}, \tau_{43}, \eta_{const}\,, \sqrt{\tau_{21}}/\tau_1$ jet B tag score

Dijet resonance anomaly search

- Compared the performance of anomaly detection methods with several standard ones
- Used mock data demonstrating different backgrounds and tested sensitivity of new methods



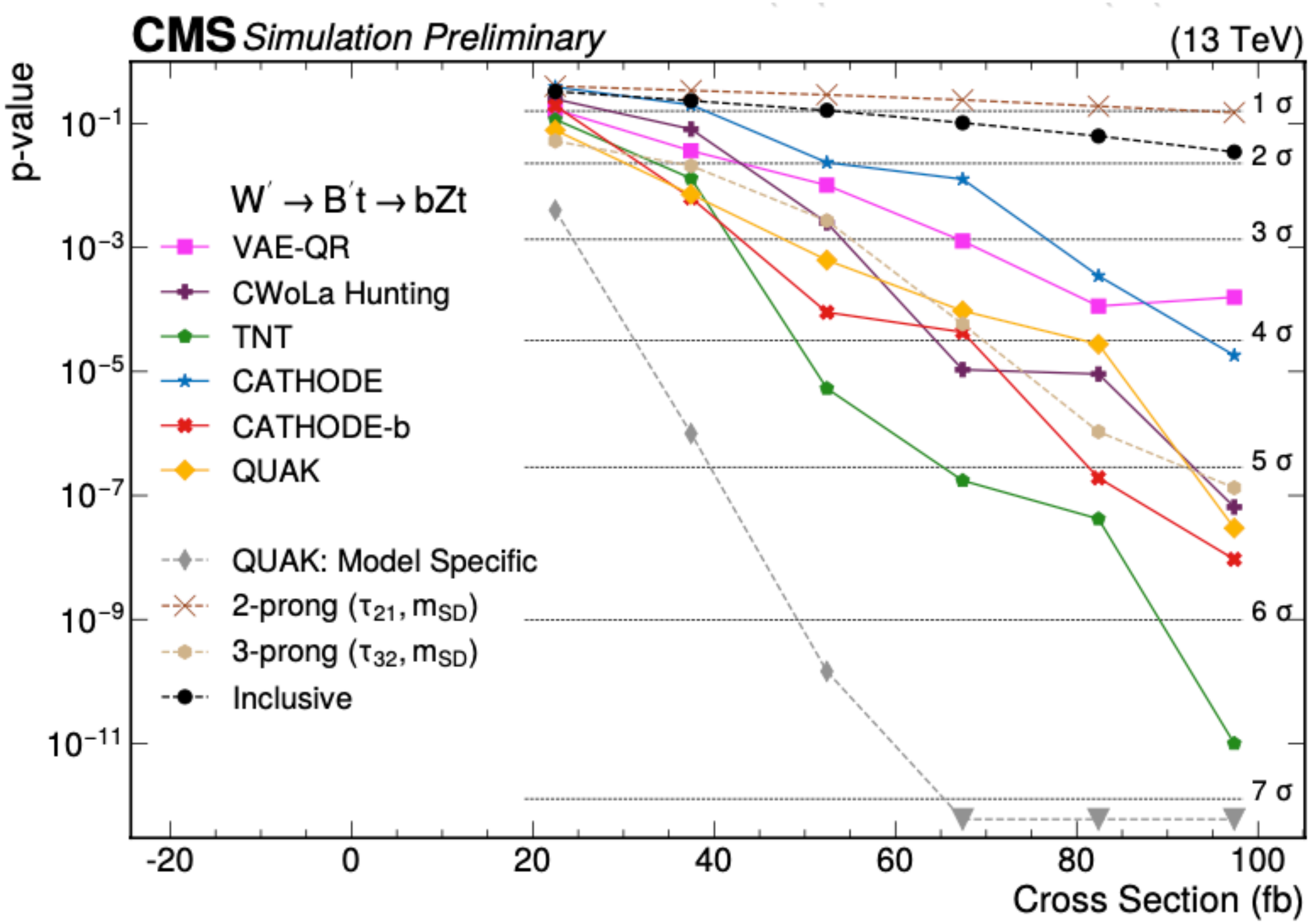
signal scenario : 2-pronged
$X \to YY' \to 4q$

Inclusive with no substructure information does not reach discovery-level

All anomaly detection methods demonstrate increased sensitivity above inclusive search

Dijet resonance anomaly search

- Similar observation with **3-pronged signal scenario** ($W' \to B't \to bZt$)
- For both signal scenarios, the inclusive search does not reach discovery-level significance

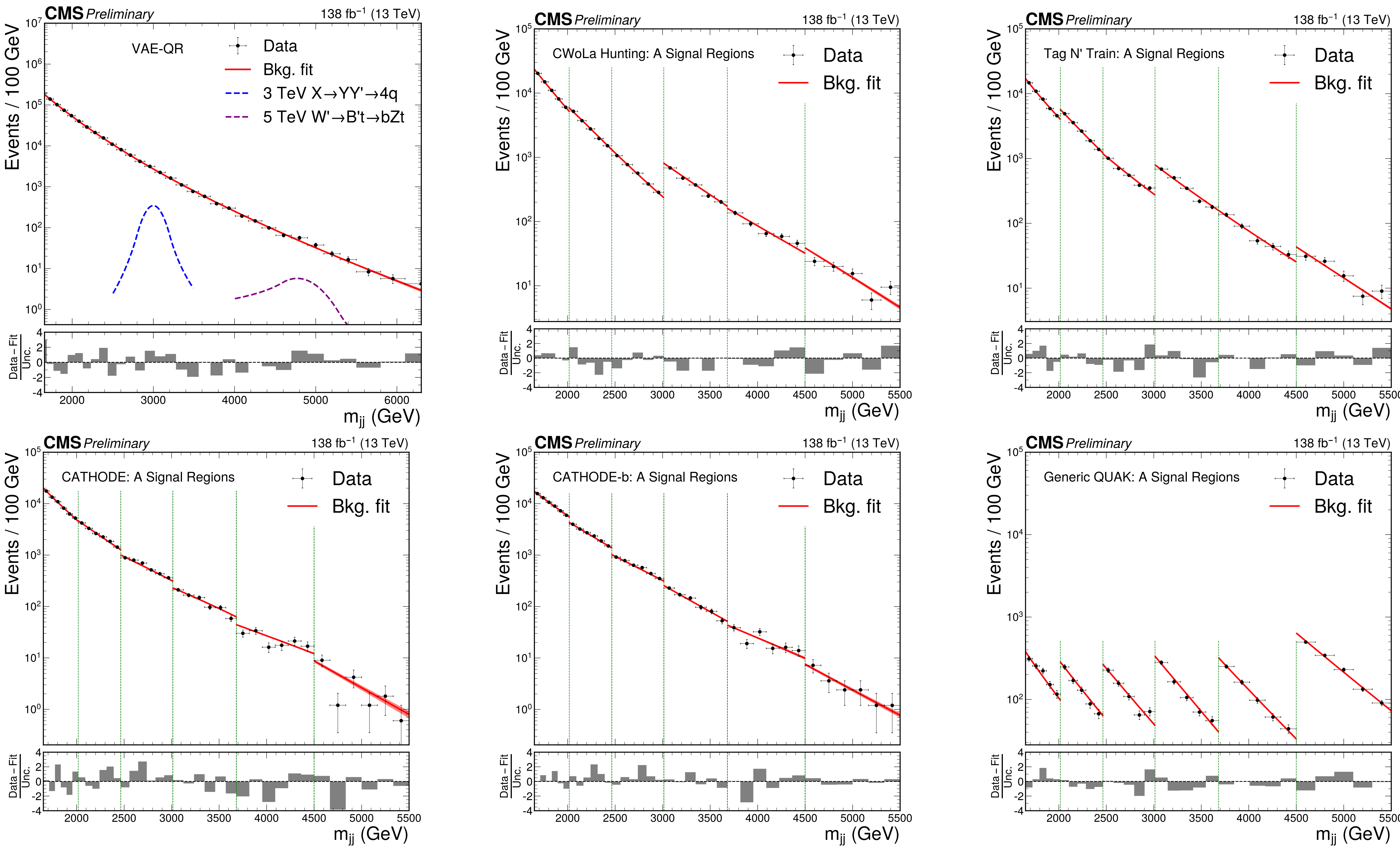


Performance of different anomaly methods vary for the two signal scenarios and no method is optimal for both

All anomaly detection methods demonstrate increased sensitivity above inclusive search

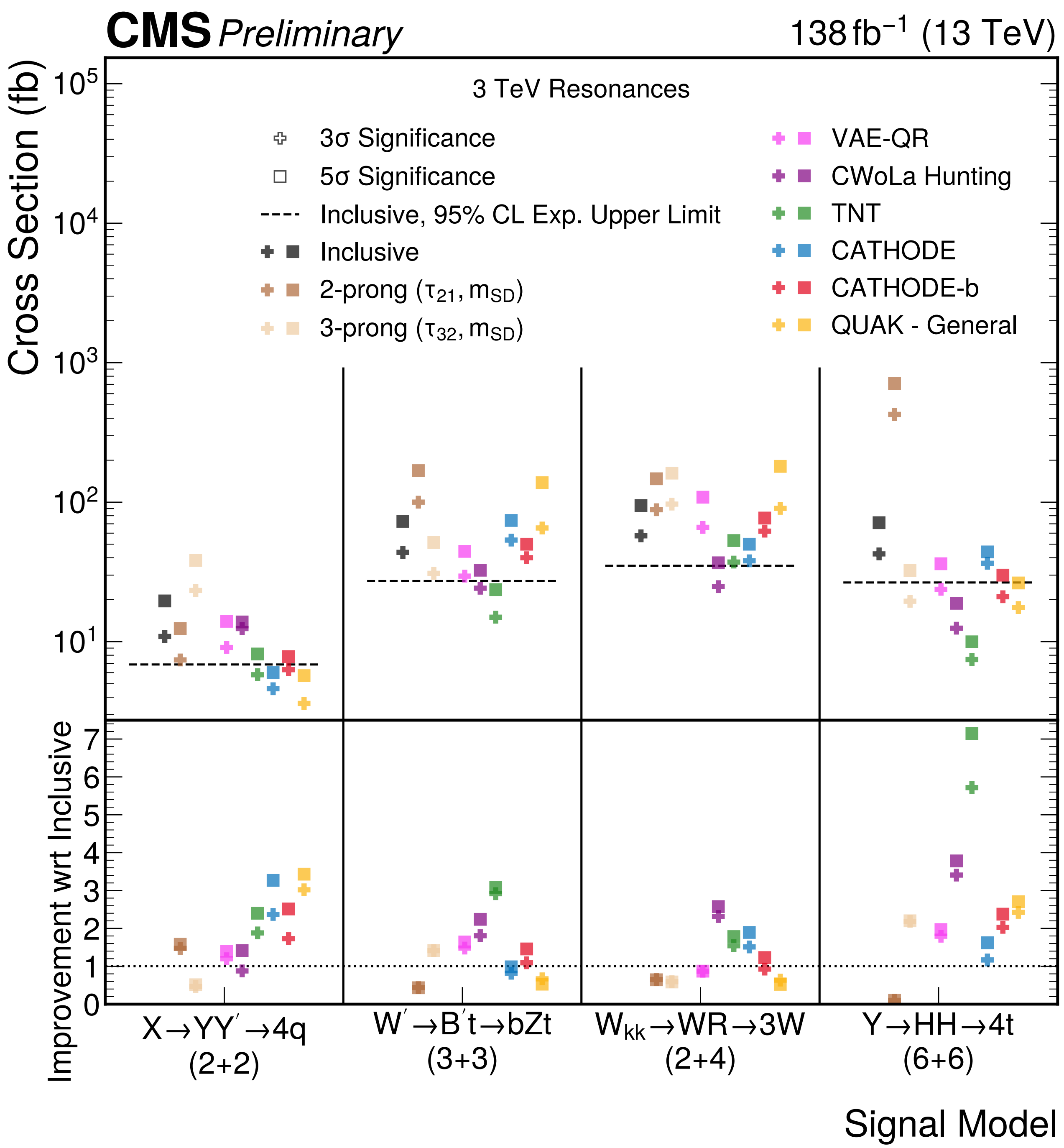
Dijet resonance anomaly search

- Performed bump hunt with all five methods on dijet invariant mass spectrum; **no excess observed**



Dijet resonance anomaly search

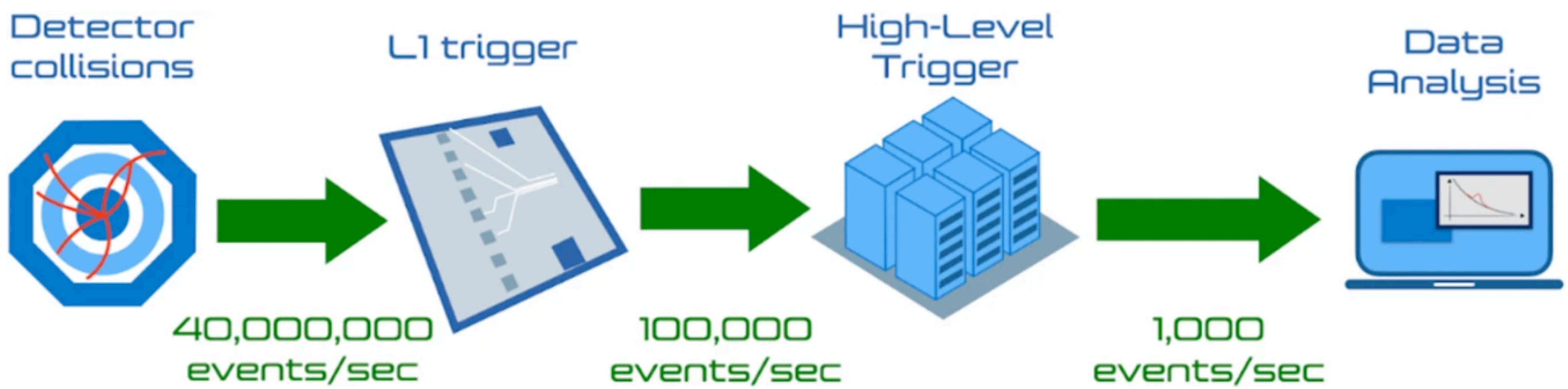
Discovery Sensitivity



- Determined which cross-sections would led to discovery ($3\sigma/5\sigma$)
- Observed **improvement with anomaly detection methods by a factor of 7 (3.3)** as compared to inclusive (cut-based) method
- For every signal, at least one anomaly detection method reaches 5σ level

Anomaly detection in real-time @ CMS

Courtesy: doi.org/10.1038/s41597-022-01187-8



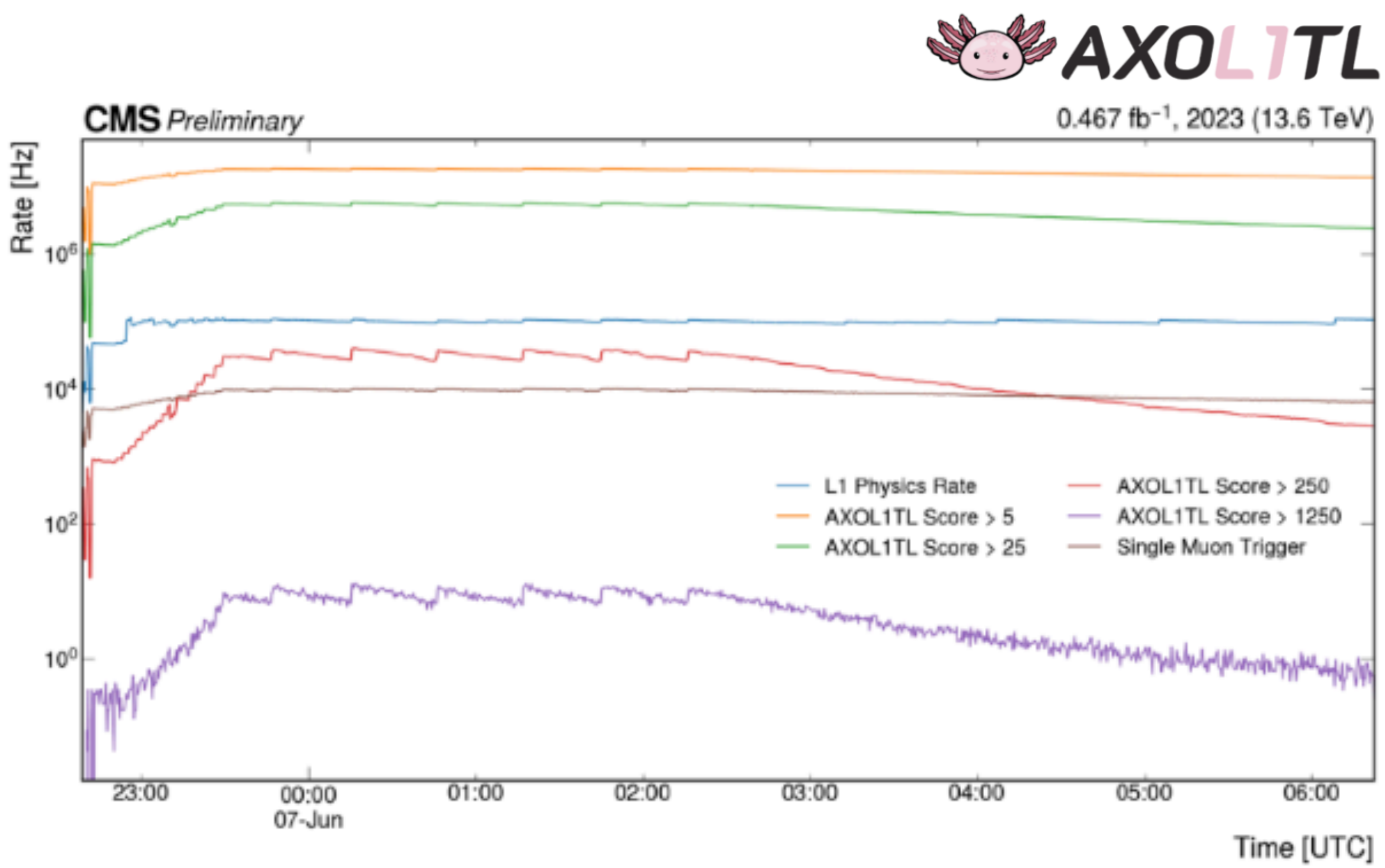
real-time data filter

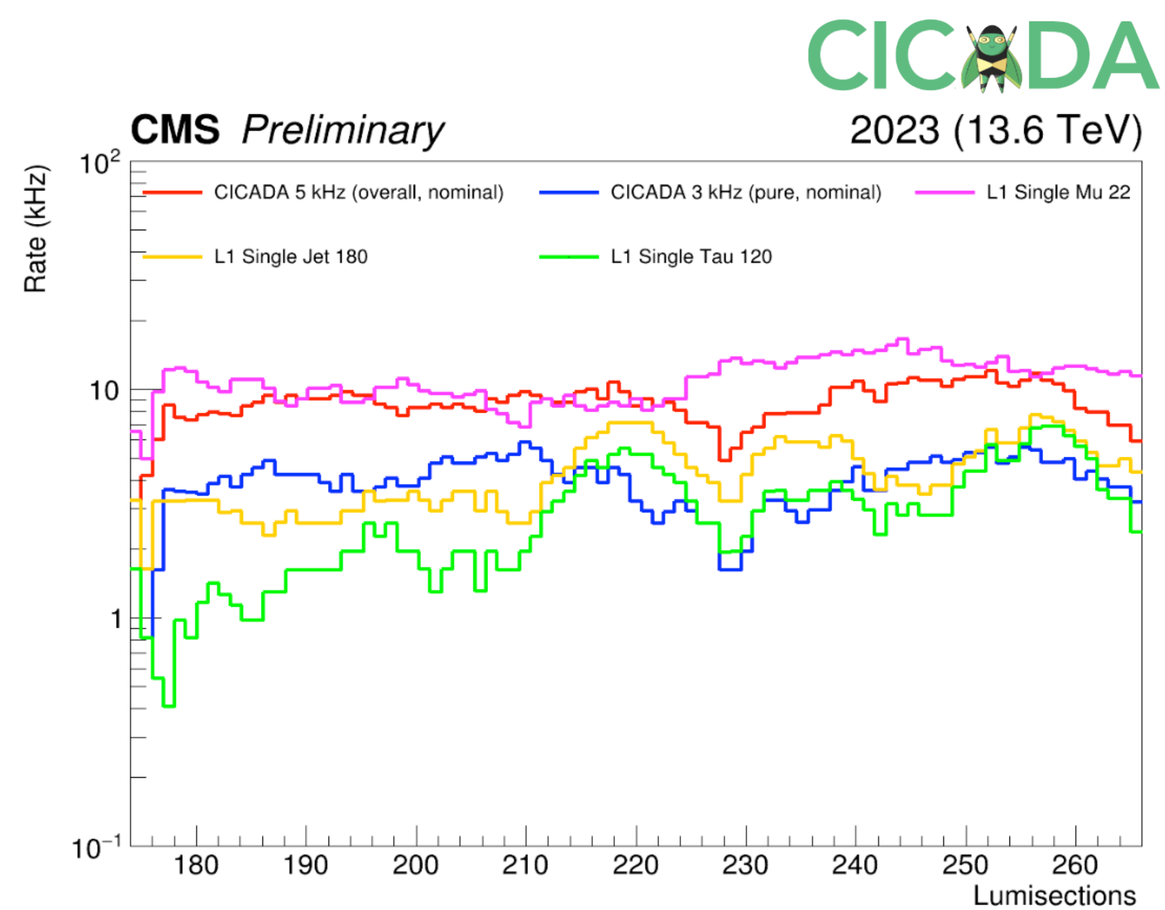
- L1 trigger quickly decides if an event should be recorded or not, based on muon-system and calorimeter information.
- Are some BSM/rare SM events also lost?
- New development: Using anomaly detection @L1 trigger
  - unbiased approach → potentially enhanced sensitivity to SM or BSM physics that evades other triggers
  - Trained directly on data
  - runs in hundreds of nanoseconds on a single FPGA

Anomaly detection in real-time @ CMS

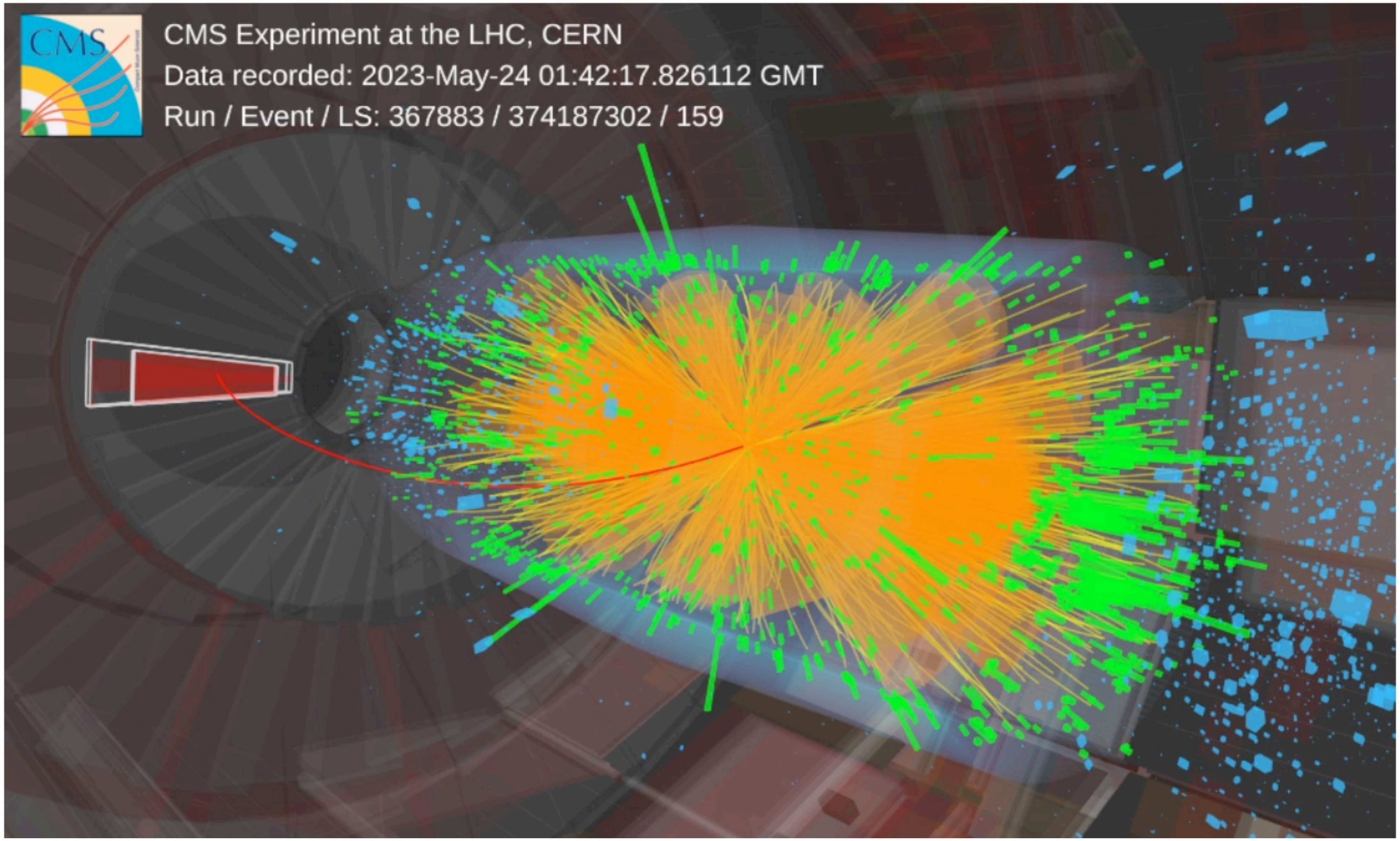
- Rate stability tests were performed during 2023 data-taking; rate stability observed throughout the fill cycle
- More sensitive to final states with high object multiplicity
- Currently recording data with the AXOL1TL seeds (+ corresponding HLT paths), at rates up to a few hundred Hz
- Further developments on the algorithms are ongoing

CMS-DP-2023-086
CMS-DP-2023-079





Anomaly detection in real-time @ CMS



- Anomalous event captured with highest anomaly score ; not selected with normal L1T menu
- Offline reconstruction identified 7 jets and 1 muon in an event
- Excited to look at the anomalous events dataset collected by these triggers

CMS-DP-2023-079

Conclusions

- Interesting results from Run 2 data; but we see **no significant excess** in any of the signal phase space
- Need to **develop new strategies** to search wider phase space but with **increased sensitivity to new physics**
- CMS has used machine-learning based **anomaly detection methodology** which is fully based on data; making it fully **signal and background independent**
- A full **dijet resonance search** is performed with various anomaly detection methods and have set the **most stringent limits on the cross section of various signal scenarios**
- Besides this, CMS developed **anomaly detection algorithm at the L1 trigger level** to enable fully unbiased analysis of rare/anomalous events

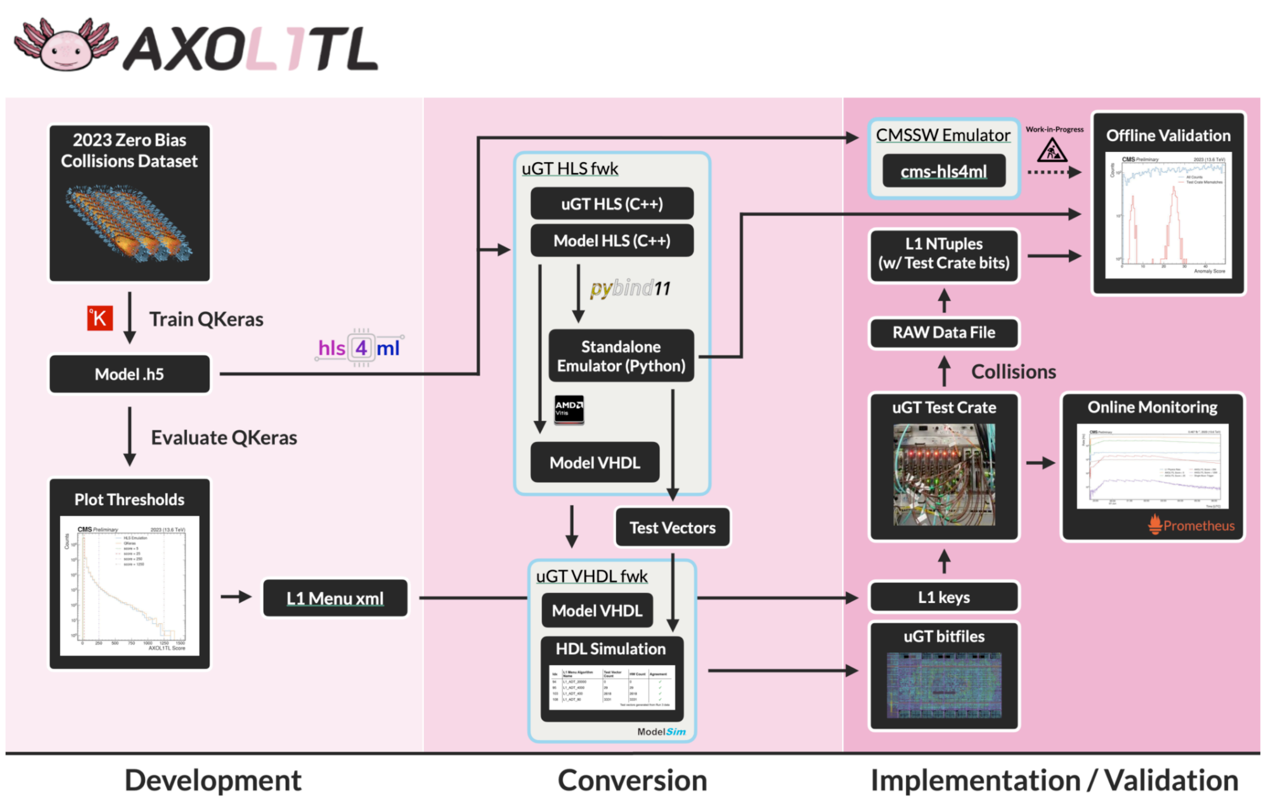
Thank You

Backup

Anomaly detection in real-time @ CMS

- Developed two complementary algorithms for **anomalous event selection at L1 trigger** level
  - AXOL1TL : uses L1 global trigger input (10 jets, 4 electrons, 4 muons and 1 MET)
  - CICADA : uses raw calorimeter energy deposits as inputs
- Uses autoencoders to pick out topologies different from majority of events seen at the LHC

CMS-DP-2023-086
CMS-DP-2023-079



- Firmware generation with hls4ml package to interface with frontend hardware
- runs in hundreds of nanoseconds on a single FPGA