Gefördert durch:

Bundesministerium
für Forschung, Technologie
und Raumfahrt

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Recent Results from Anomaly Detection Searches on CMS

Tore von Schwartz for the CMS Collbaboration

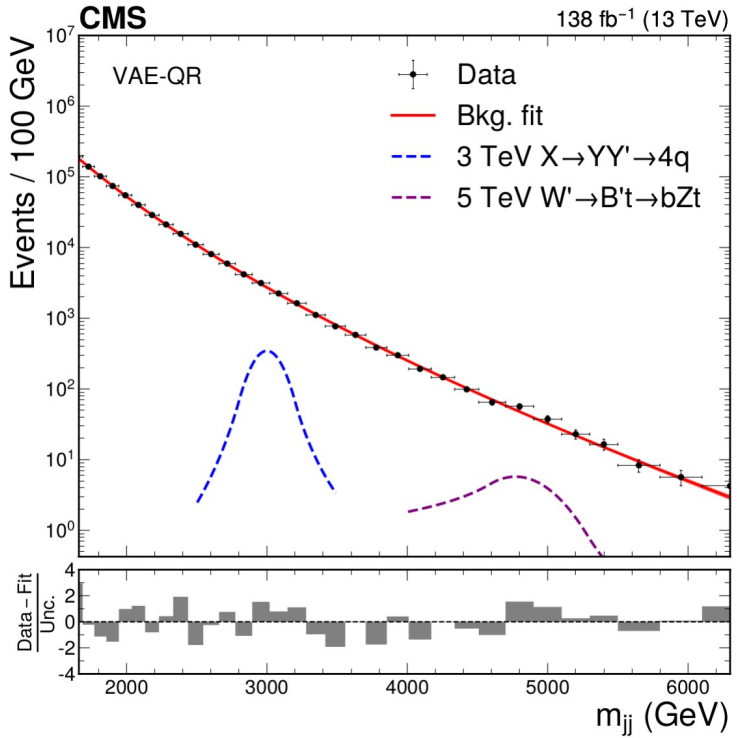
tore.von.schwartz@cern.ch

17.09.26

ML4Jets 2026

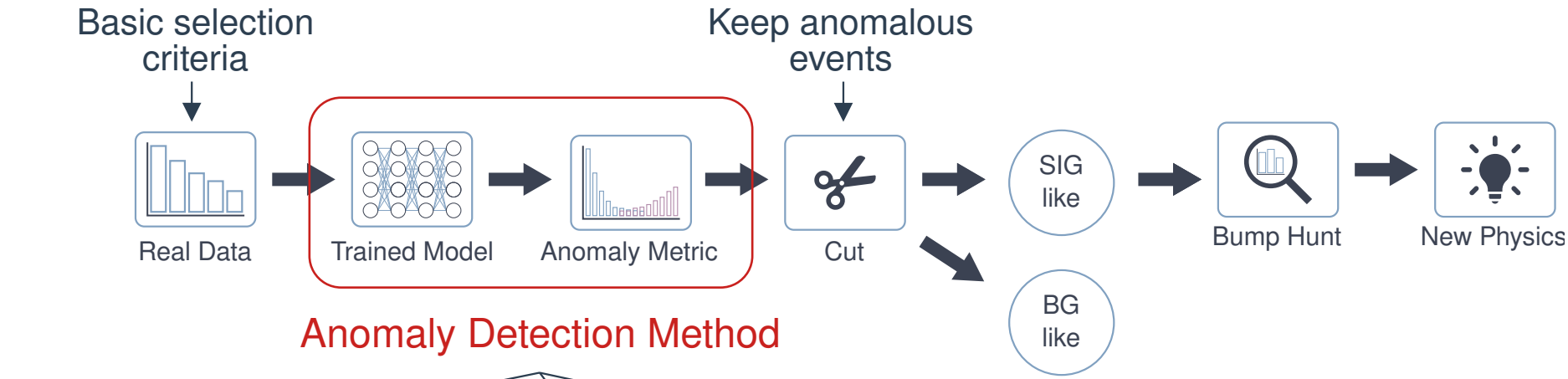
Introduction

- New physics may appear in signatures that traditional searches miss
- Goal is to find rare events without a signal model
- Search for new particles that may appear as invariant-mass peak
  [2412.03747, 2512.20395, 2509.13635]



[2412.03747]

Analysis Strategy



Unsupervised
VAE-QR [2412.03747]

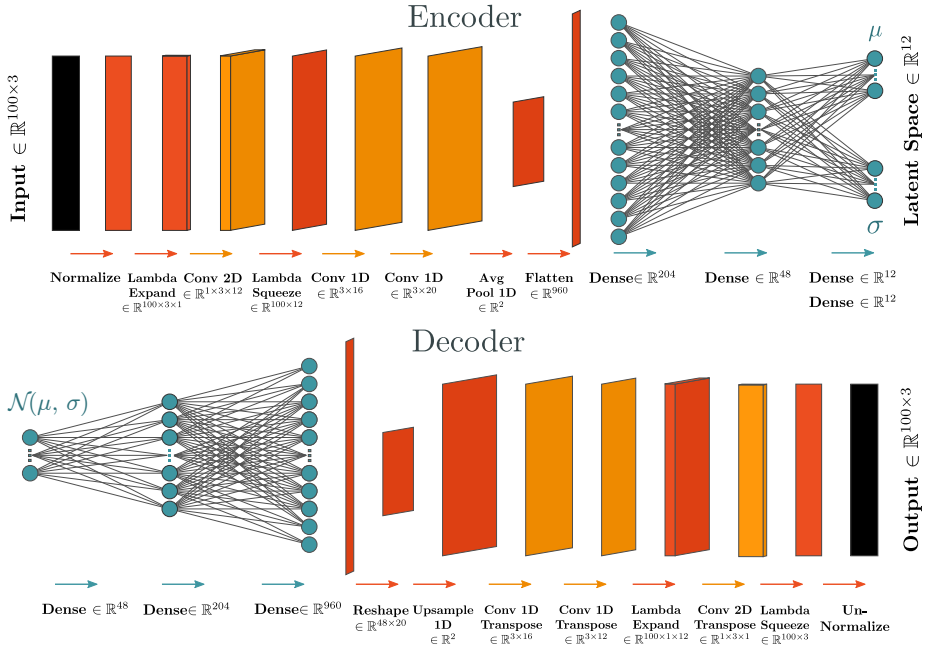
Semi-supervised
QUAK [2011.03550]

Weakly supervised
CwoLA [1902.02634]
TNT [2002.12376]
CATHODE [2109.00546]

Unsupervised Anomaly Detection

VAE-QR (Variational Autoencoder - Quantile Regression)

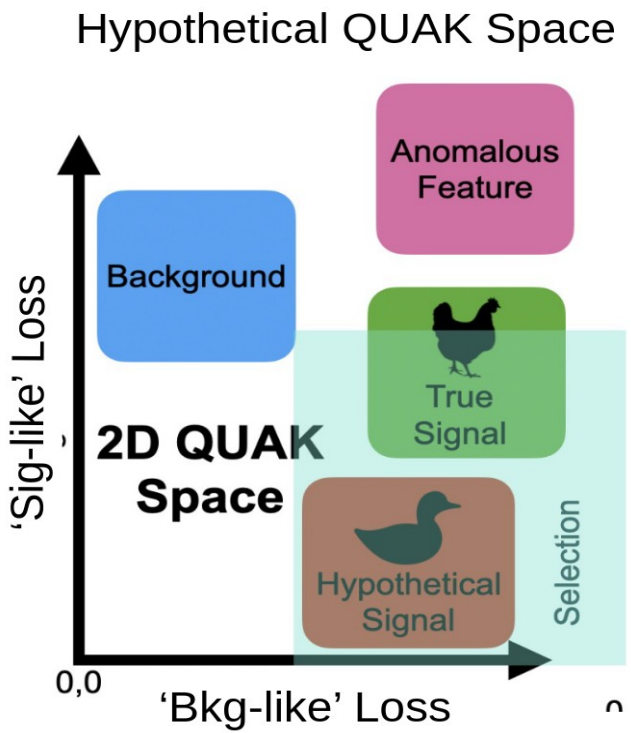
- Encodes 100 particles per jet to 12 dimensions, then decodes
- Trained in background-dominated sideband
- Learns to reconstruct background, not signal
- Per event score: Worst reconstructrion loss of the two jets



Semi-supervised Anomaly Detection

QUAK (QUasi Anomalous Knowledge)

- Hybrid approach, uses a prior from simulated signal events
- Train separate models on simulated signal and background
- Assign signal and background loss scores to events
- Selected events are signal- but not background-like



[2011.03550]

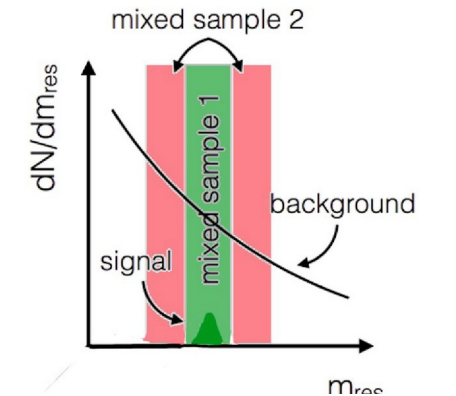
Weakly supervised Anomaly Detection

- Assume signal is narrow resonance and choose a mass window that is defined as SR (signal region)
- Define the SB (side bands)
- Train a classifier between data and background-like events

$$\frac{p_{\text{data}}(\vec{x})}{p_{\text{bkg}}(\vec{x})} = \frac{(1-\epsilon)p_{\text{bkg}}(\vec{x}) + \epsilon p_{\text{sig}}(\vec{x})}{p_{\text{bkg}}(\vec{x})} = 1 - \epsilon + \epsilon \frac{p_{\text{sig}}(\vec{x})}{p_{\text{bkg}}(\vec{x})}$$

data vs background

signal vs background



[1708.02949]

Weakly supervised Anomaly Detection

CWoLA

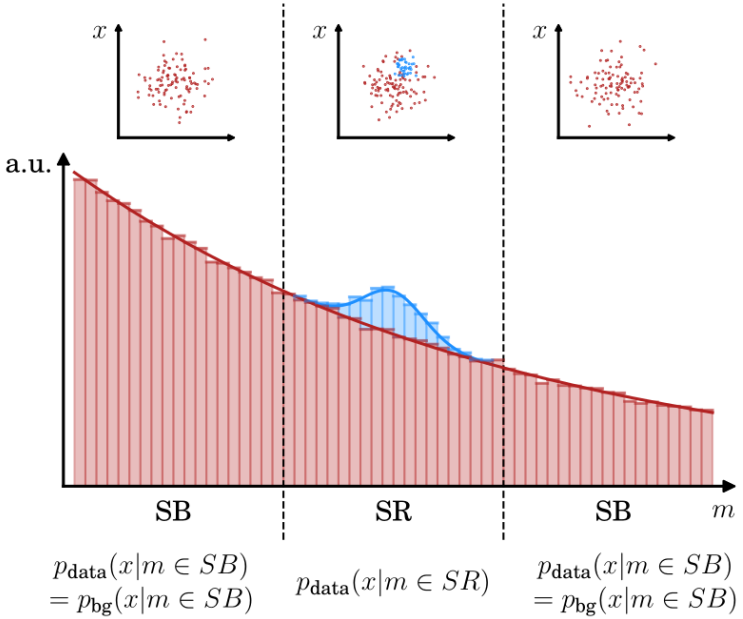
- Background from SB, directly into classifier training

TNT (Tag N' Train)

- Use Autoencoder to tag events with two anomalous jets
- Classify between accepted jets, rejected jets and SB jets

CATHODE

- Density estimator trained on SB data can sample background-like data in SR
- Classification between SR data and sampled SR background

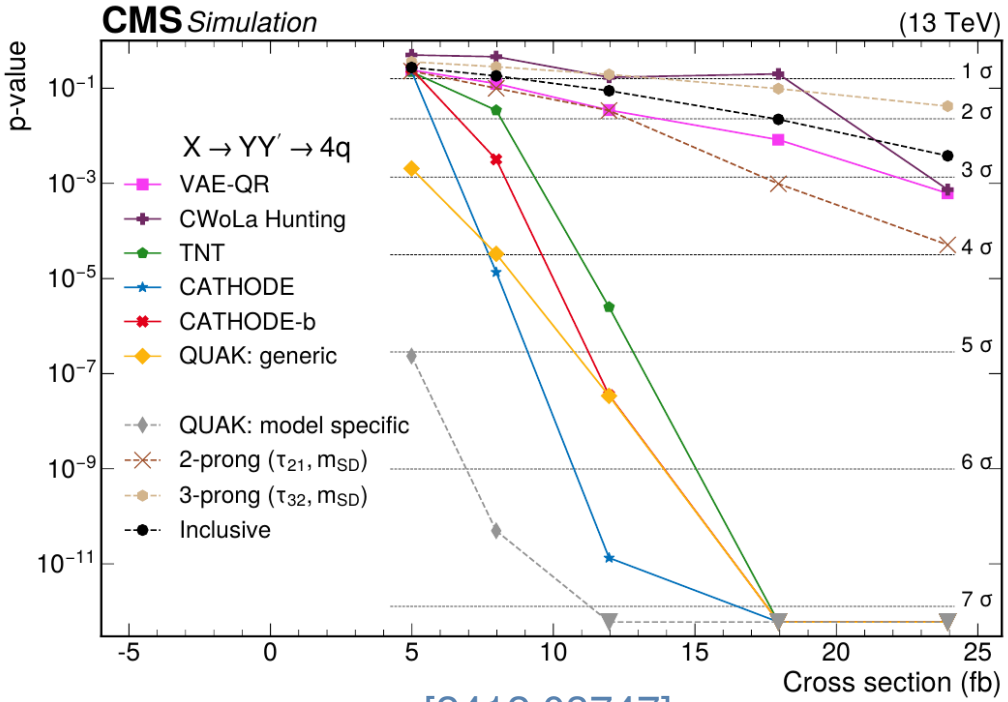


[2109.00546]

Performance on benchmark signal

- Pick one benchmark model:

  $X \rightarrow YY' \rightarrow 4q$
- Check significance for varying cross sections
- Four methods with similar performance
- Detection with fewer events compared to traditional, cut-based methods



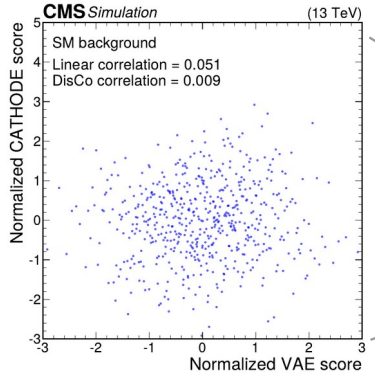
[2412.03747]

Complementarity

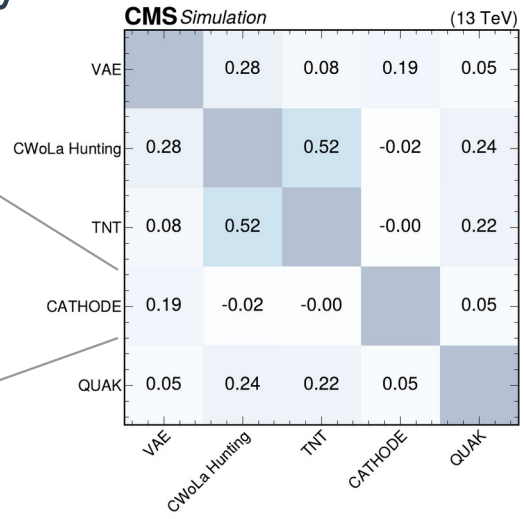
- Study correlations between anomaly scores of different methods
- Do different methods flag the same events as background/signal?
- Pearson correlation matrices show small correlations

  → signals are found in different ways

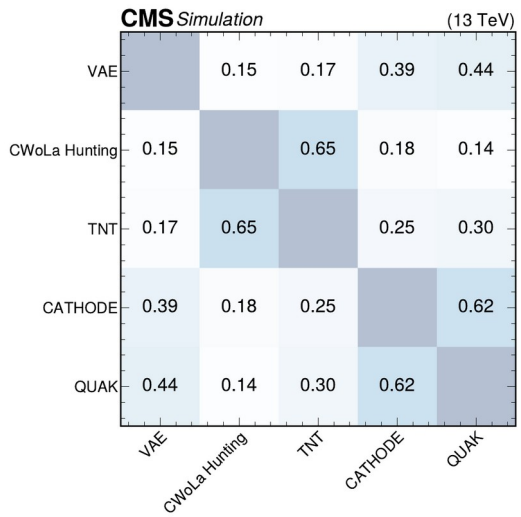
X(3000) → YY' → qq qq
[2512.20395]



background

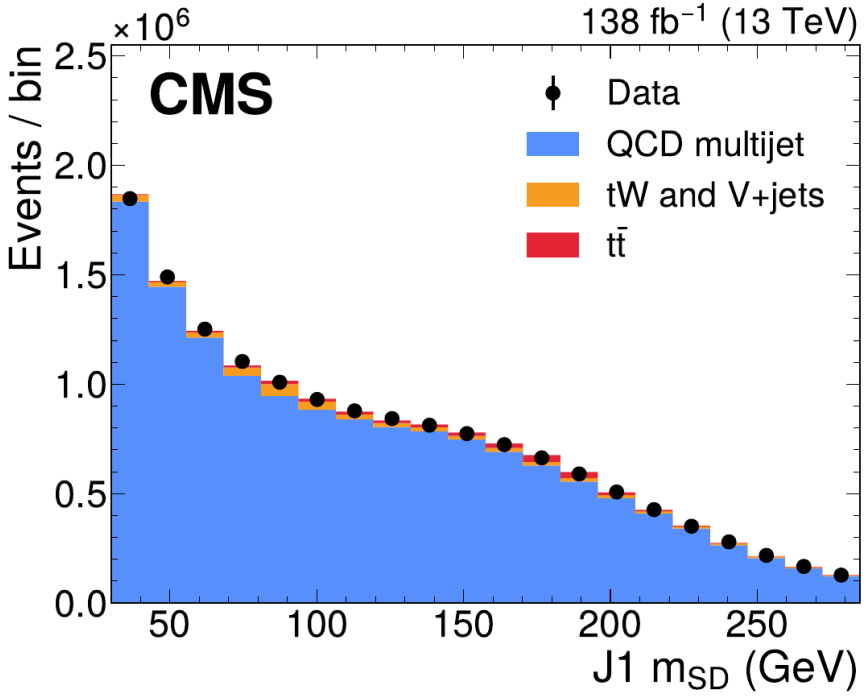


signal



Finding the top quark

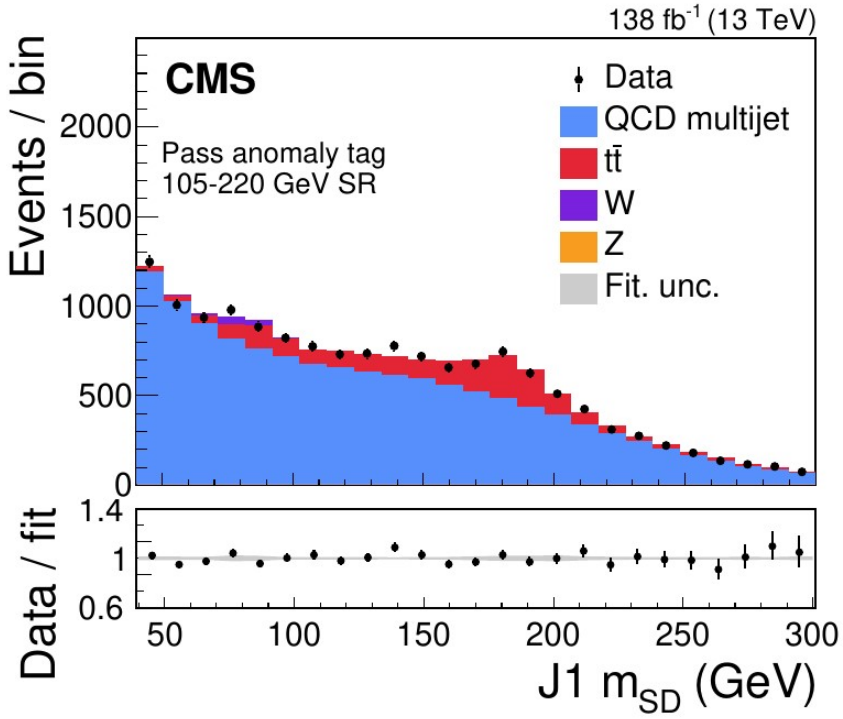
- Run weakly-supervised anomaly detection in data
- Jet mass as the resonant variable
- Setup adapted for pair-produced boosted resonances
- Initially, QCD dominates



[2512.20395]

Finding the top quark: After Anomaly Detection

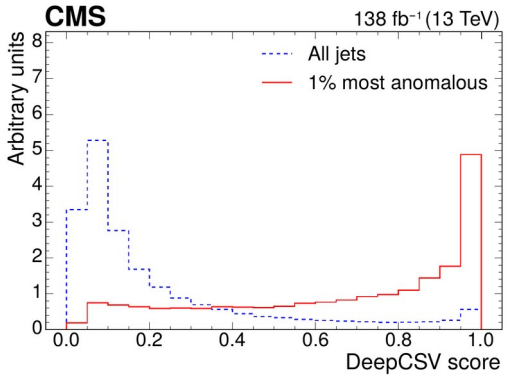
- The top-quark shows up as peak in jet mass distribution
- Reminder:
  - Trained on data
  - Evaluated on data
  - No MC involved



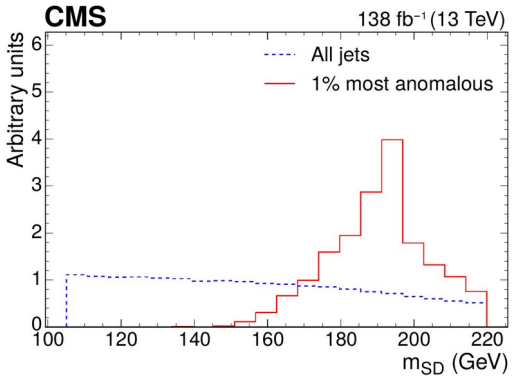
[2512.20395]

Finding the top quark: Excess Interpretation

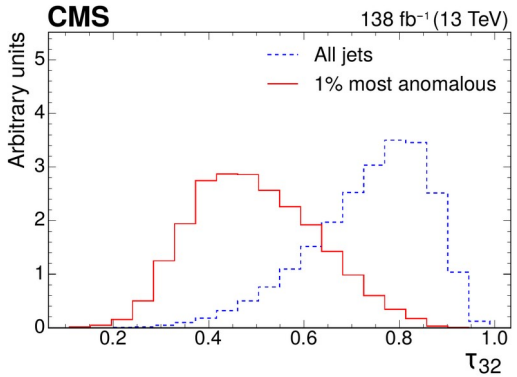
- In case of excess more characterization of the signal is needed
- Looking at properties of most anomalous events, we can spot the top properties



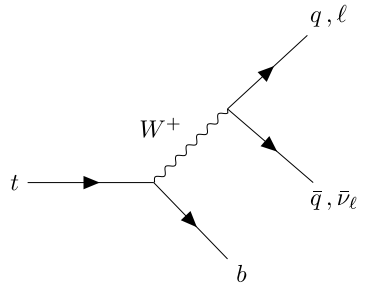
Decay products contain b-jets



Signal resonance around 190 GeV



Signal decay in 3-prong structure



X → HY, where Y is anomalous jet

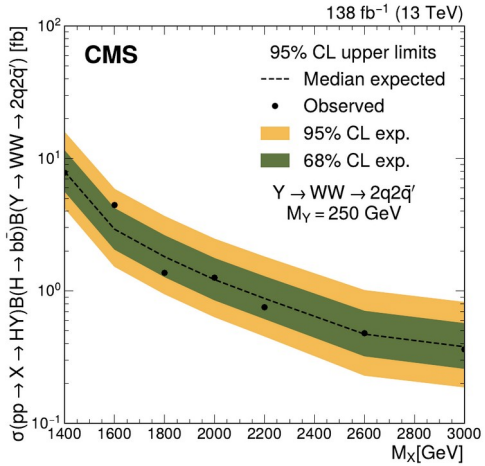
- Covers many models

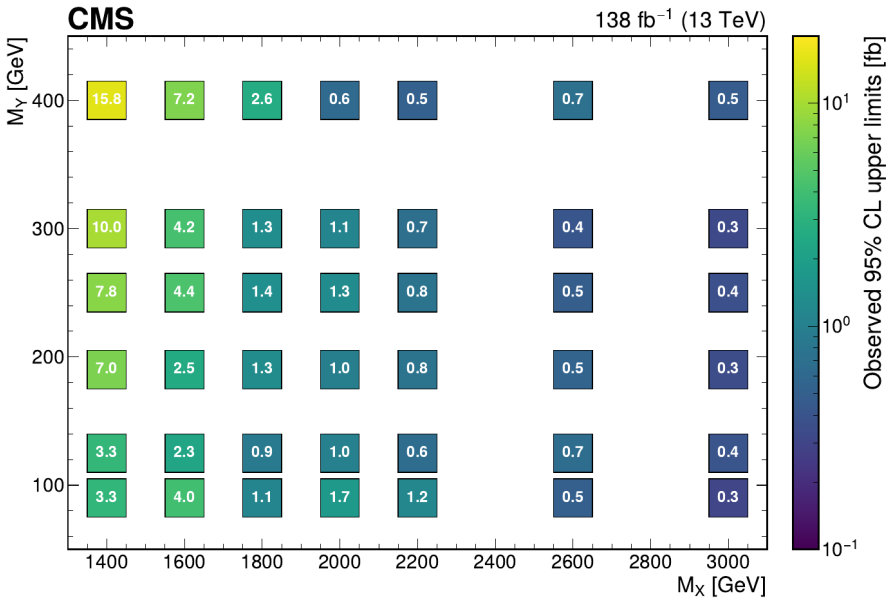
  X → H(bb)Y(4q) X → H(bb)Y(ud)

  X → H(bb)Y(tt(6q)) T' → H(bb)t(bqq)

- Search for a bump in a 2D (m_X, m_Y) plane

T' → Ht, t → bqq
[2509.13635]





X → HY, Y → WW → 4q
[2509.13635]

Summary

- ML methods used for dijet bump-hunting in CMS [2412.03747]
  - Weak supervision: CWoLA, TNT, CATHODE
  - Quasi-anomalous knowledge (QUAK)
  - Variational autoencoder with quantile regression (VAE-QR)
- Study and comparison of the methods [2512.20395]
  - Performance on benchmark signal
  - “Rediscover“ the top quark
- Search for anomalous Jet-Higgs resonances [2509.13635]
- Studies for extension to LLPs and multi-jets ongoing