SPONSORED BY THE
Federal Ministry of Education and Research

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Latest developments in CATHODE

Chitrakshee Yede, Gregor Kasieczka, Louis Moureaux, Tore von Schwartz
(chitrakshee.yede@cern.ch)

AD4HEP Workshop, Columbia University, Nevis laboratories

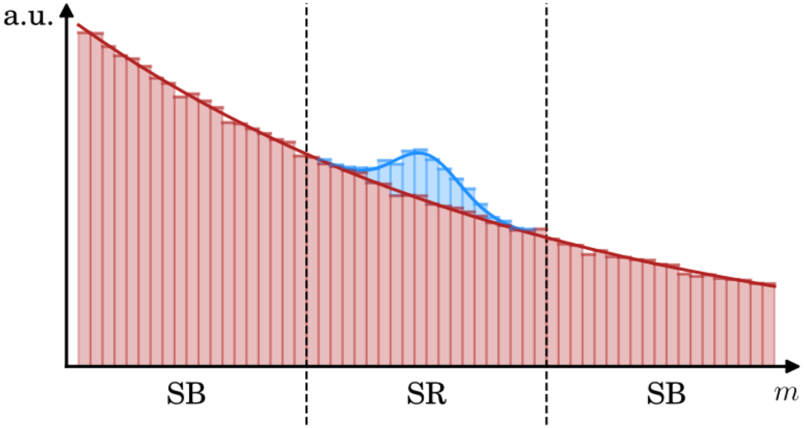
17th June 2025

What is Cathode?

It is a weakly supervised anomaly detection method

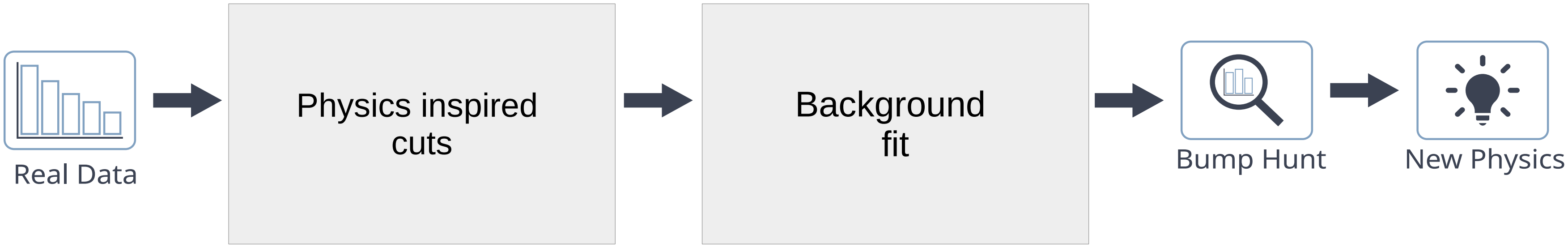
Search for new physics

- Many dedicated searches
- No new physics found
- Broader approach needed (Use data-driven, model-agnostic searches)



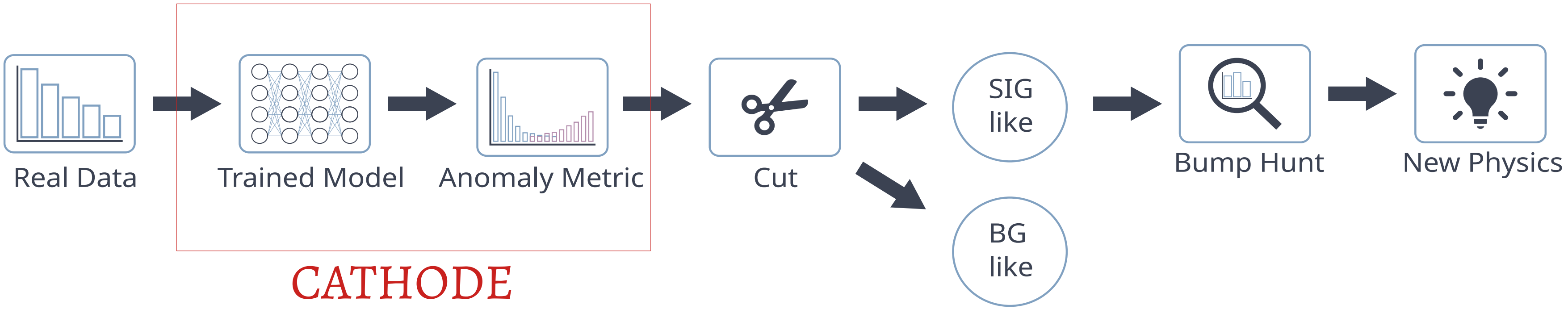
Bump Hunt

- Resonance in feature ***m***
- Smooth background, localized signal



Anomaly detection

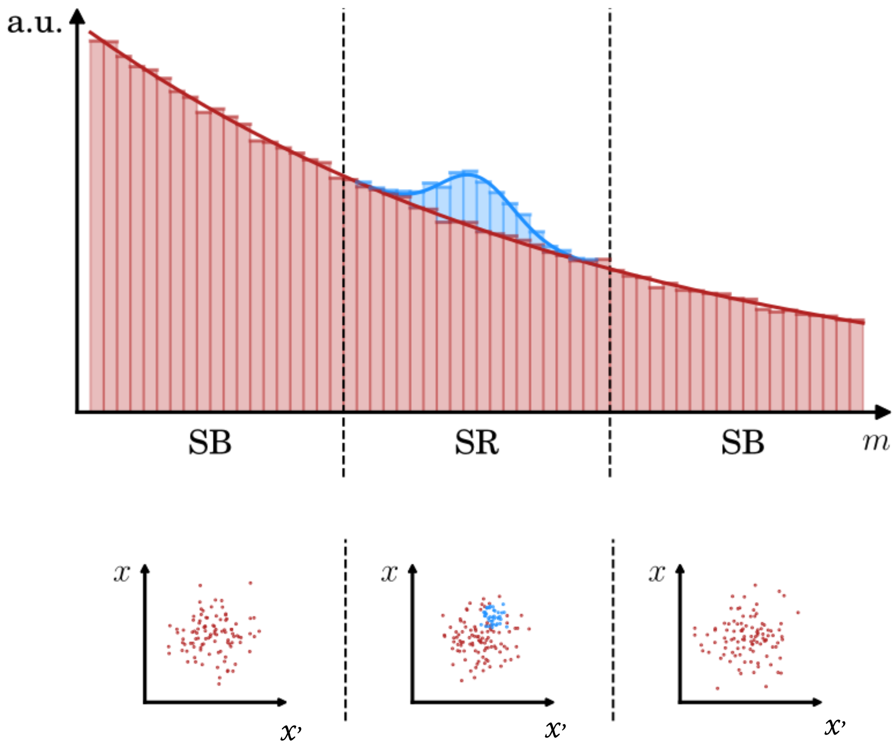
- Data-driven, model-independent approach
- Train an ML model to assign anomaly scores to events.
- Apply cut to these scores to select the most anomalous events to reduce background, keep signal



(2109.00546, Hallin et al)

Anomalous Resonances

- Features:
  - resonant feature, $\boldsymbol{m}$
  - auxiliary features : $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3} \ldots$
- Train the anomaly detector using auxiliary features
- Utilizing x gives further separation in signal-background



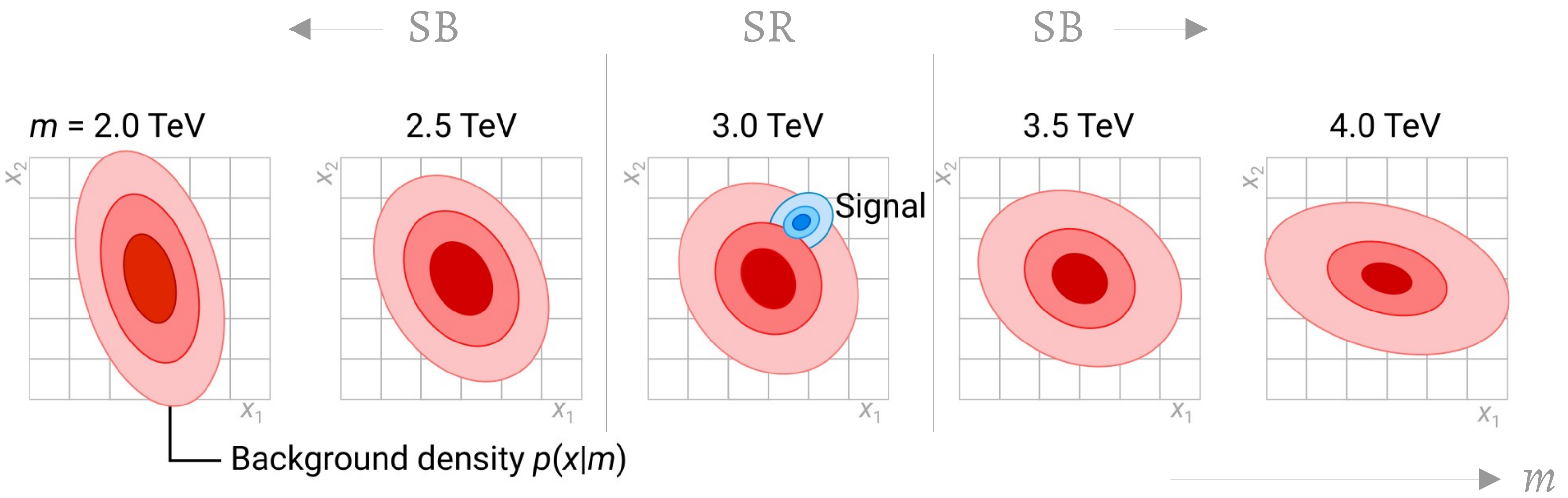
CATHODE

Classifying Anomalies THrough Outer Density Estimation

- Start with data, choose resonant feature m and auxiliary features $x_1, x_2, x_3\ldots$
- Split the mass spectrum into SR (signal region) and the SB (side bands)

1. Density Estimation (Normalizing Flows)

- Train a generative model only on SB to learn the density of x conditioned on m
- Sample the background in the SR



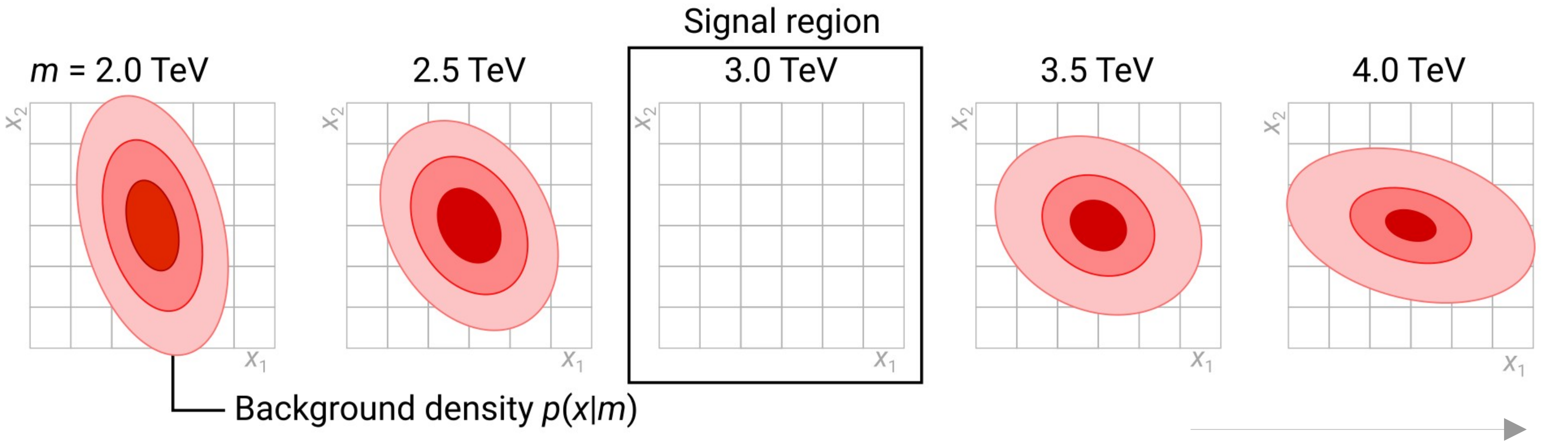
CATHODE

Classifying Anomalies THrough Outer Density Estimation

- Start with data, choose resonant feature m and auxiliary features x_1, x_2, x_3...
- Split the mass spectrum into SR (signal region) and the SB (side bands)

1. Density Estimation (Normalizing Flows)

- Train a generative model only on SB to learn the density of x conditioned on m
- Sample the background in the SR



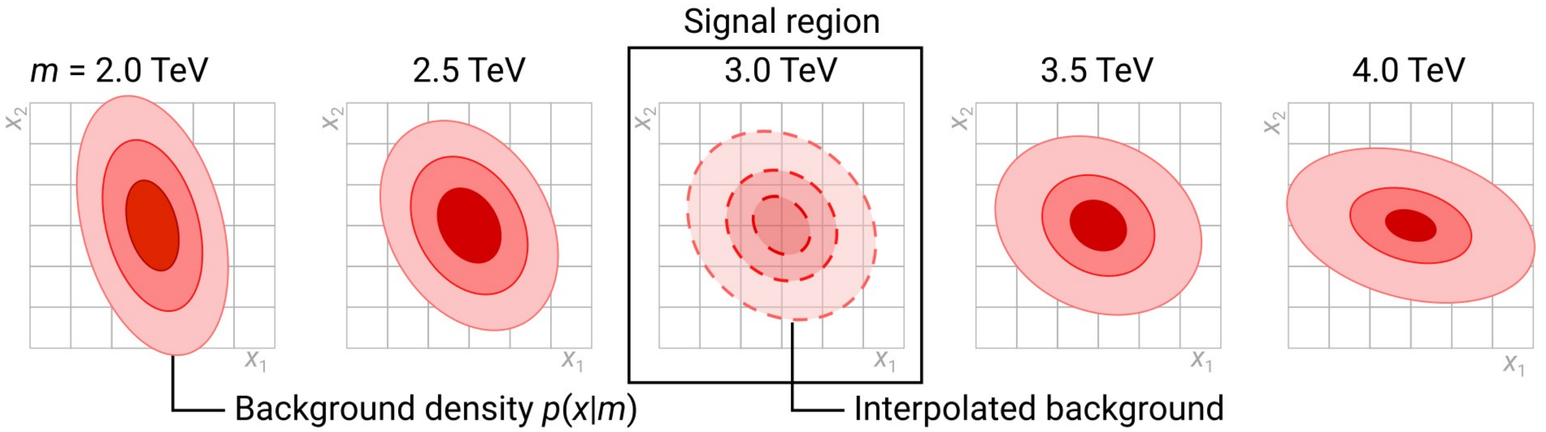
CATHODE

Classifying Anomalies THrough Outer Density Estimation

- Start with data, choose resonant feature m and auxiliary features $x_1, x_2, x_3\ldots$
- Split the mass spectrum into SR (signal region) and the SB (side bands)

1. Density Estimation (Normalizing Flows)

- Train a generative model only on SB to learn the density of x conditioned on m
- Sample the background in the SR

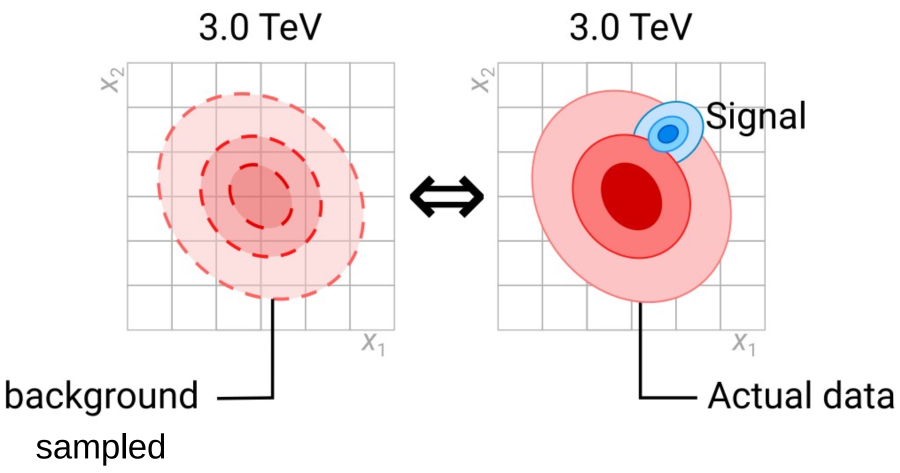


CATHODE

Classifying Anomalies THrough Outer Density Estimation

2. Classification (Weak Supervision)

- Train a classifier in SR between the actual data and the sampled background
- Put a cut on the score to select the most anomalous events



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

data vs background

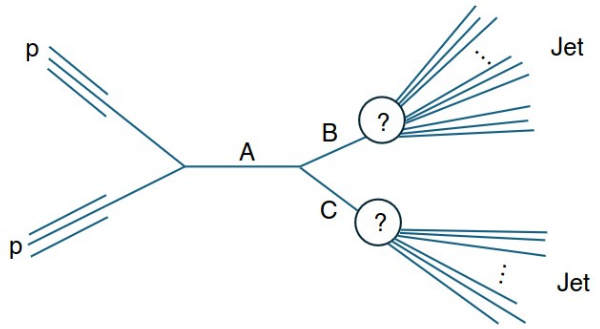
signal vs background

Application and latest developments

- Cathode applied in CMS analysis for dijet anomaly detection
- Using Features: M_{JJ}, M_{J1}, ΔM_{JJ}, $\tau_{21}J_1$, $\tau_{21}J_2$

CMS analysis (2412.03747)

ATLAS analysis (2502.09770)

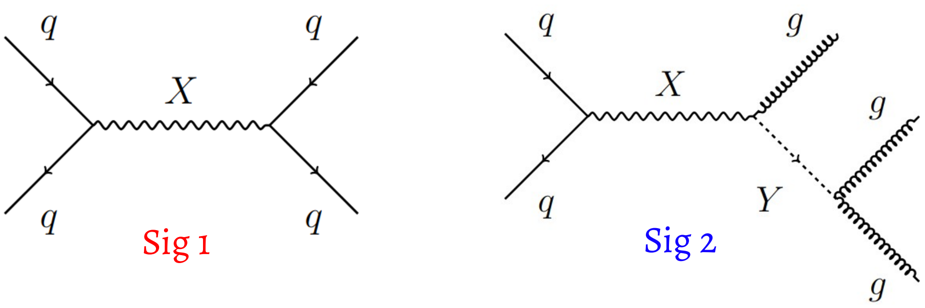


Now Exploring → New Signal topologies and new feature sets

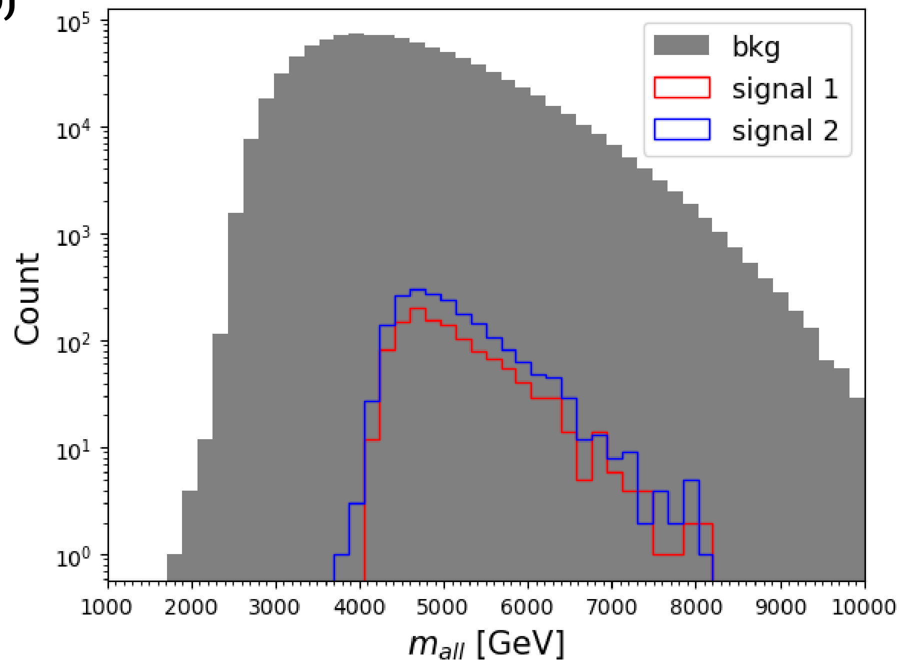
1. Generalizing the signal topologies
2. Weak supervision using event level features (2504.13249)

(1) Generalizing the signal topologies

- Use complicated signal → **LHCO BB3 Dataset (2101.08320)** with $m_X = 4.2$ TeV



- Two different decay modes, dijet and trijet final states
- How to define the resonant feature in this case?



M_{all} – mass of all the particles

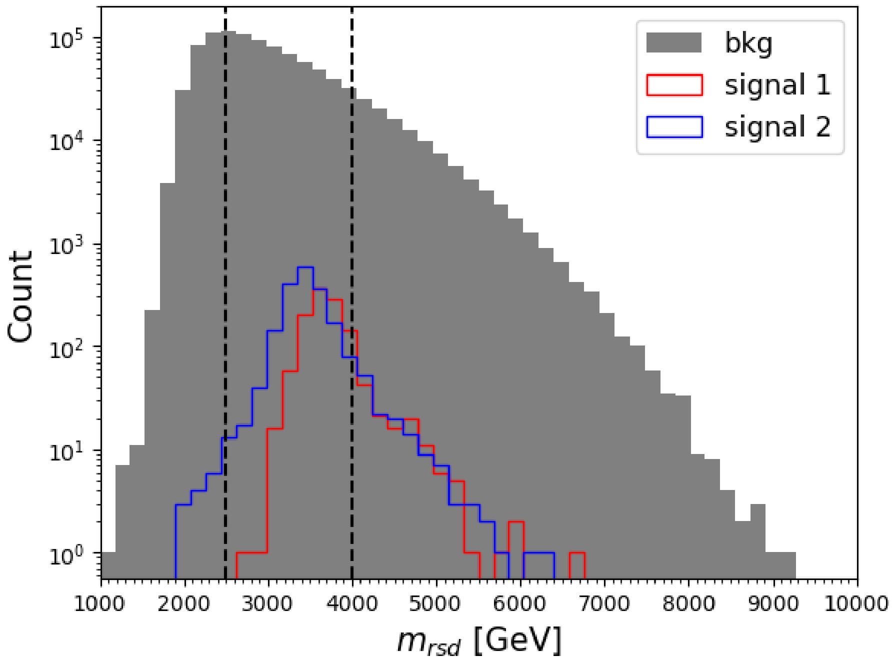
(1) Generalizing the signal topologies

- Apply Recursive Soft Drop (1804.03657),

  $M_{all} \rightarrow M_{RSD}$

- Results in more narrow peak that allows definition of signal region

- Check if Anomaly detection works!

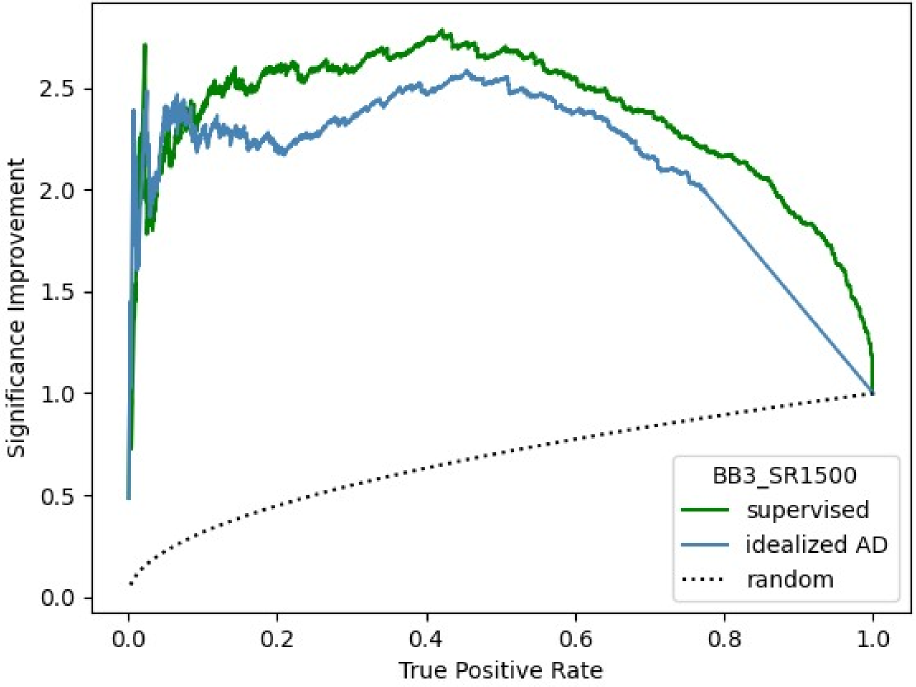


M_{all} – mass of all the particles

(1) Working point for BB3

- Results shown for
  - Supervised classifier
  - Idealized anomaly detector
    (data Vs. perfect bkg)

- Input features used: M_{RSD}, H_{T}, τ_1, τ_2, τ_3, τ_4
- Next step: Check if Cathode classifier works



In collaboration with David Shih*, Sung Hak Lim**, Luigi Favaro***

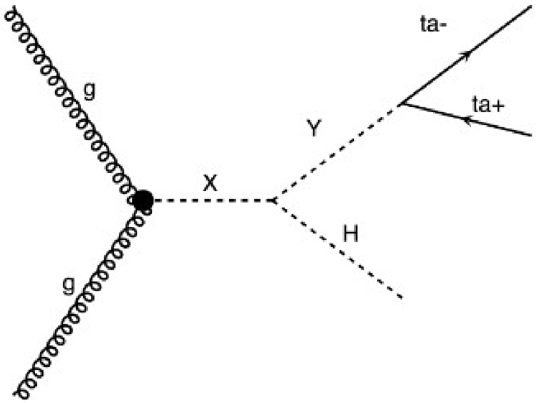
*Rutgers University, US **IBS,CTPU,South Korea ***Université catholique de Louvain

(2) Weak supervision using event level features

(2504.13249, L. Brennan et al)

- Prior studies show that low level features for weak supervision requires large signal to work
- Used summary variables + more event level observables like non τ jet p_T , MET

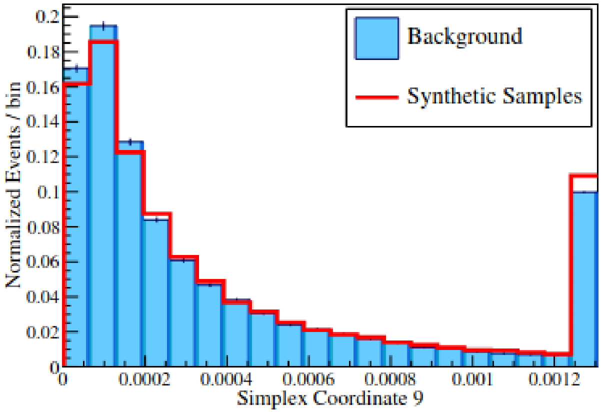
Input Feature Selection : first ten simplex coordinates, MET, non-τ jet pT sum, and di-τ ΔR

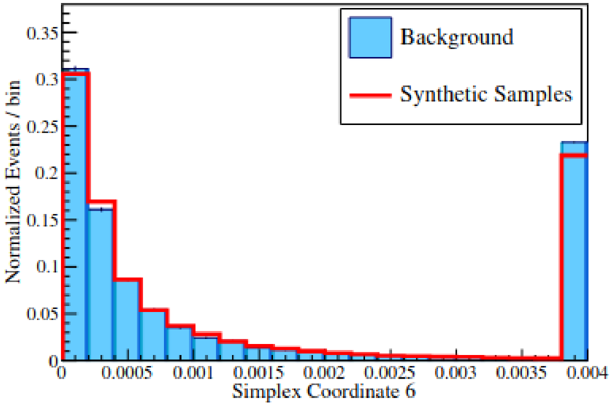


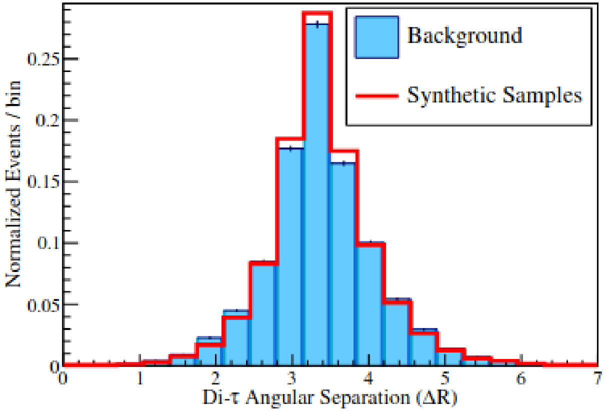
Signal topology: di-τ + X

(2) Weak supervision using event level features

- CATHODE applied
- Density Estimation done: Bkg Vs. Sampled Bkg
- Weakly supervised classification
- Cut on anomaly score

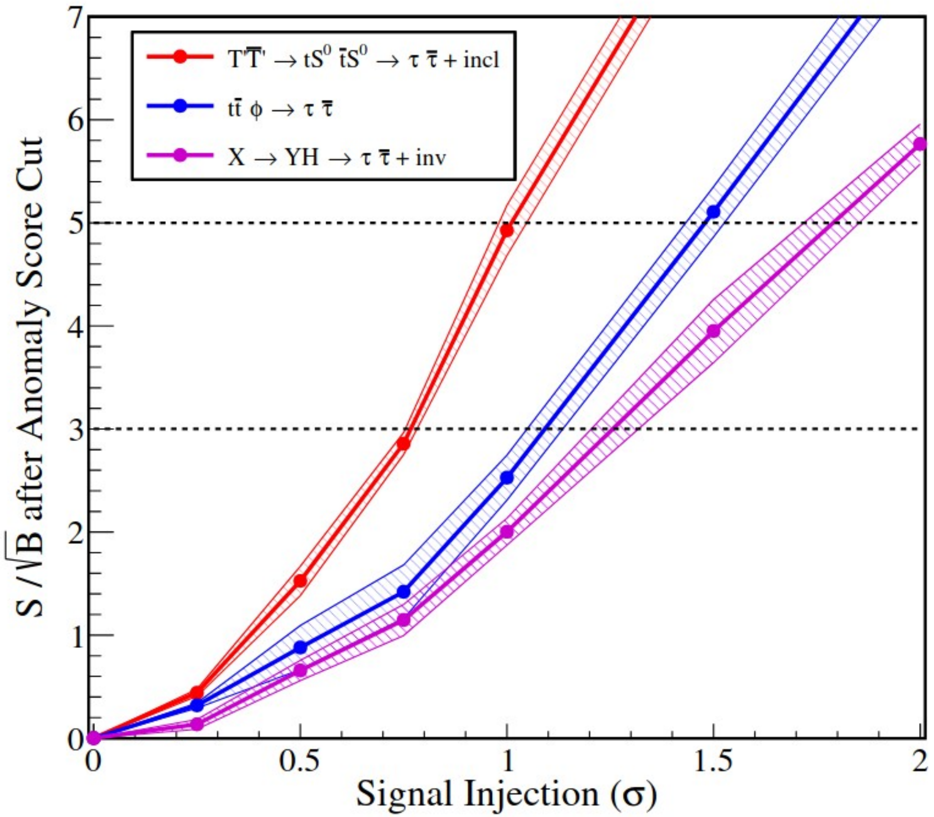






(2) Weak supervision using event level features

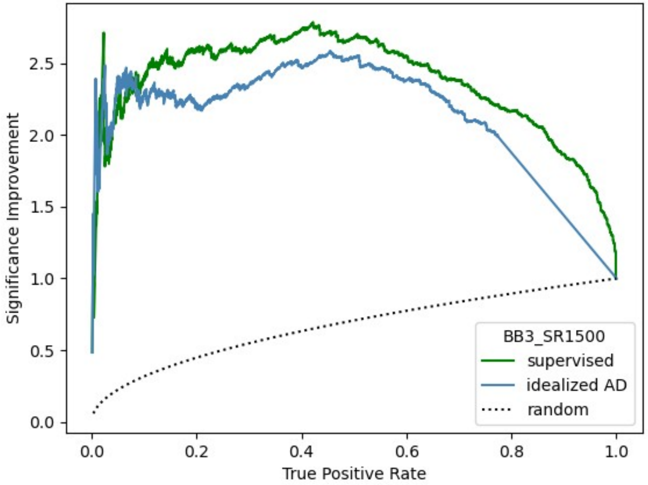
- Results shown for 3 signal (diτ+X) scenarios
- Cathode results in enhancement in signal sensitivity
- Signal significance above discovery level for $\geq 2\sigma$ signal injection

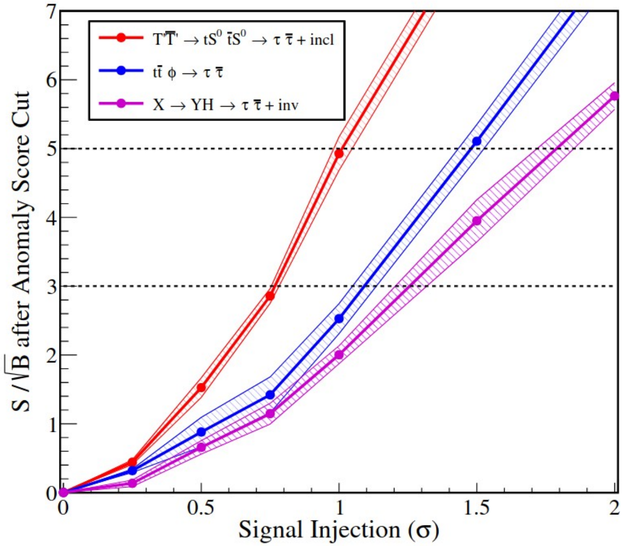


Summary

- Cathode – weakly supervised anomaly detection (2109.00546)
- Used in CMS analysis (2412.03747)
- Improvements:
    1) Exploring new signal topologies
    2) Event level features for weak supervision

(2504.13249, L. Brennan et al)





Bacakup – Phase Space Coordinates

- References for the summary variables used-
  1) The Phase space distance between collider events (2405.16698)
  2) Covariantizing Phase Space (2008.06508)
- Represents collider events using geometric phase space → provides a coordinate system that is both global and intrinsic to the event's kinematics