Model-agnostic search for dijet resonances with the CMS detector

Results on CDS

Louis Moureaux on behalf of the CMS Collaboration
cms-pag-conveners-exo@cern.ch

AISSAI workshop in Clermont-Ferrand, 06.03.2024

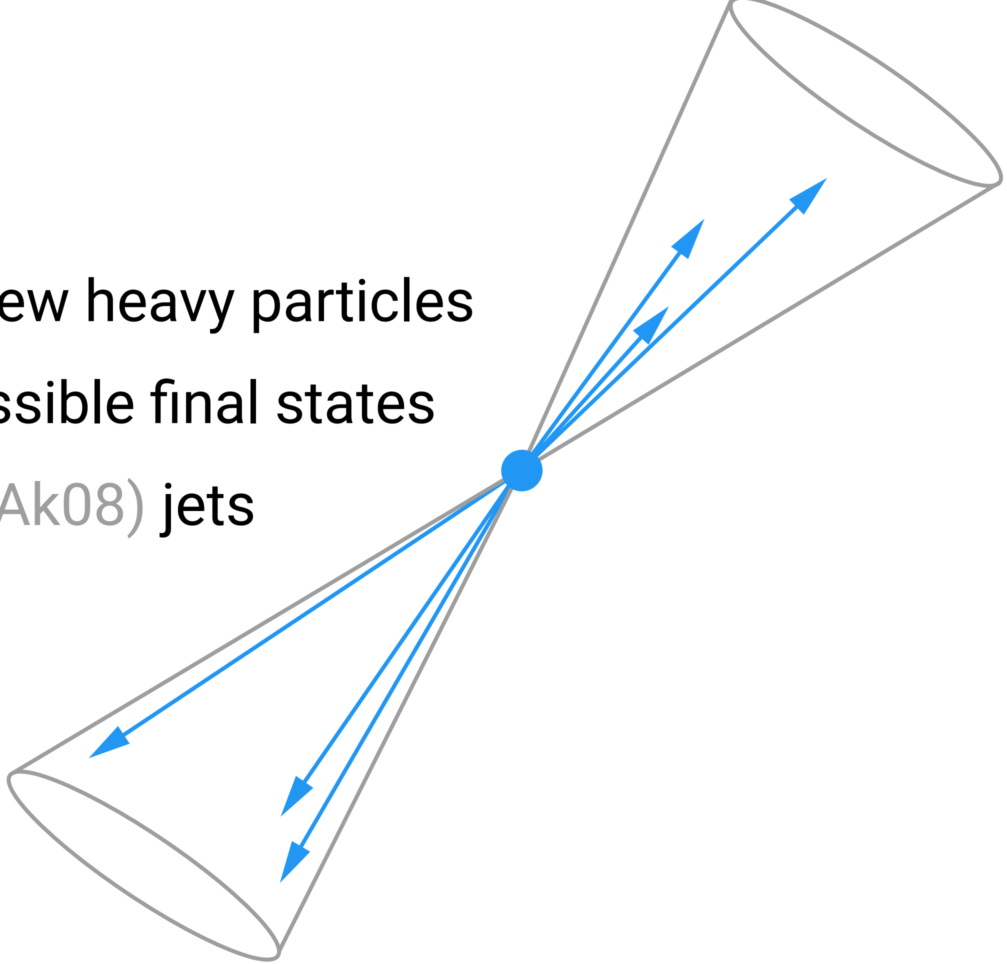
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

SPONSORED BY THE
Federal Ministry of Education and Research

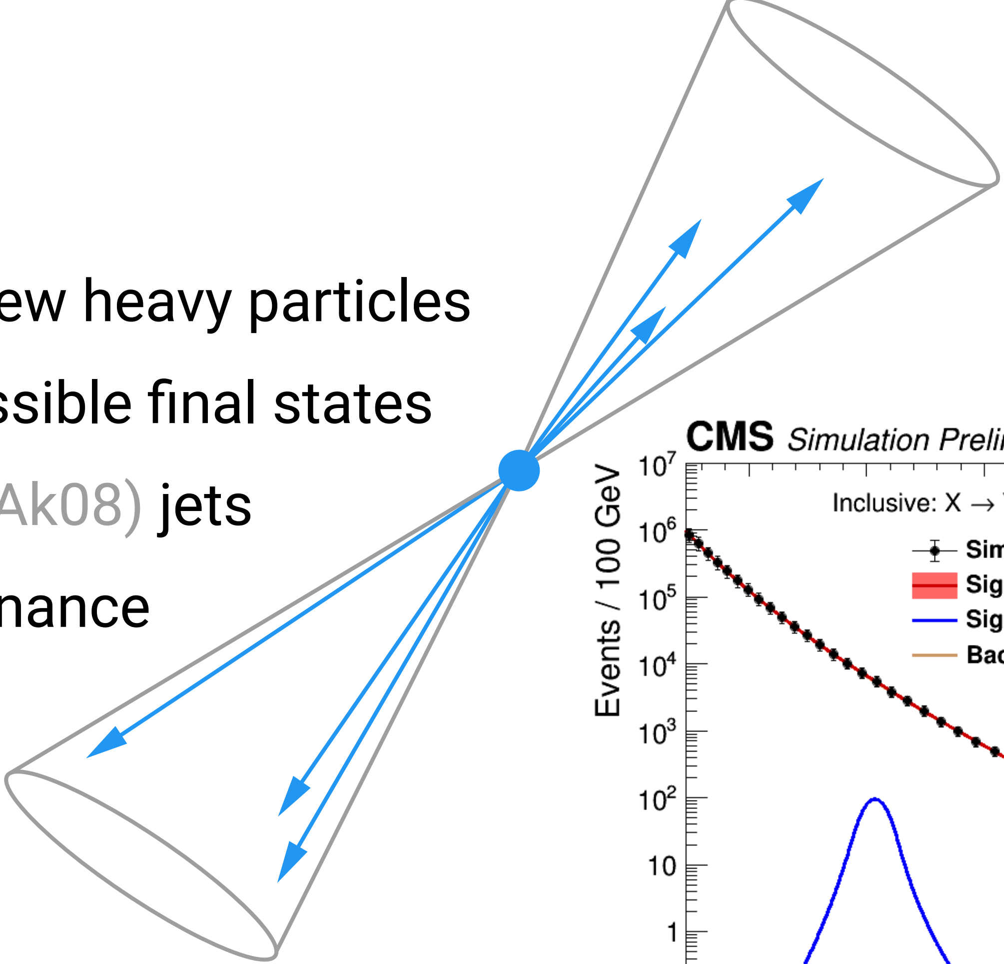
Motivation

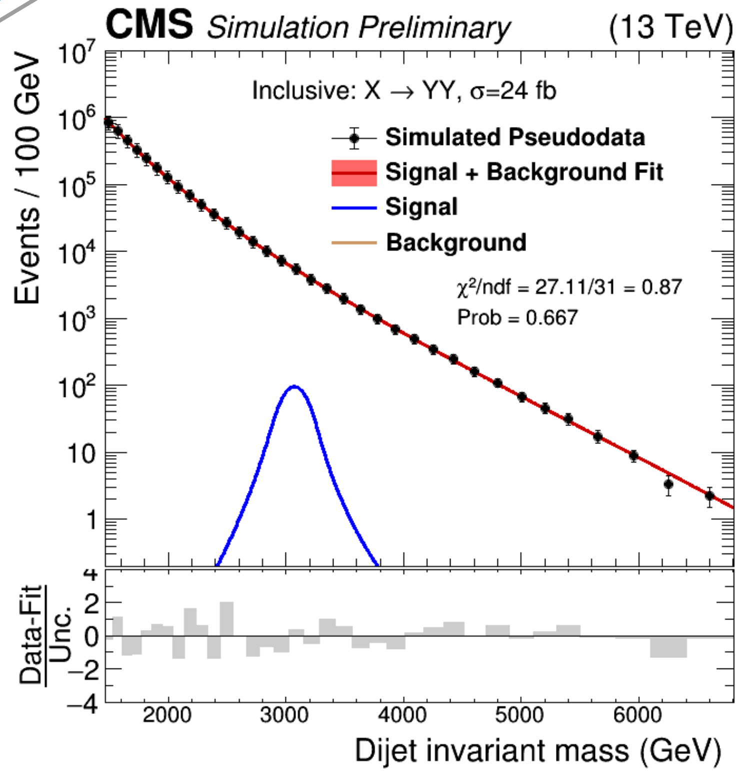
- CMS is looking for new heavy particles
- Large number of possible final states
- Looking at pairs of (Ak08) jets



Motivation

- CMS is looking for new heavy particles
- Large number of possible final states
- Looking at pairs of (Ak08) jets
- Searching for a resonance

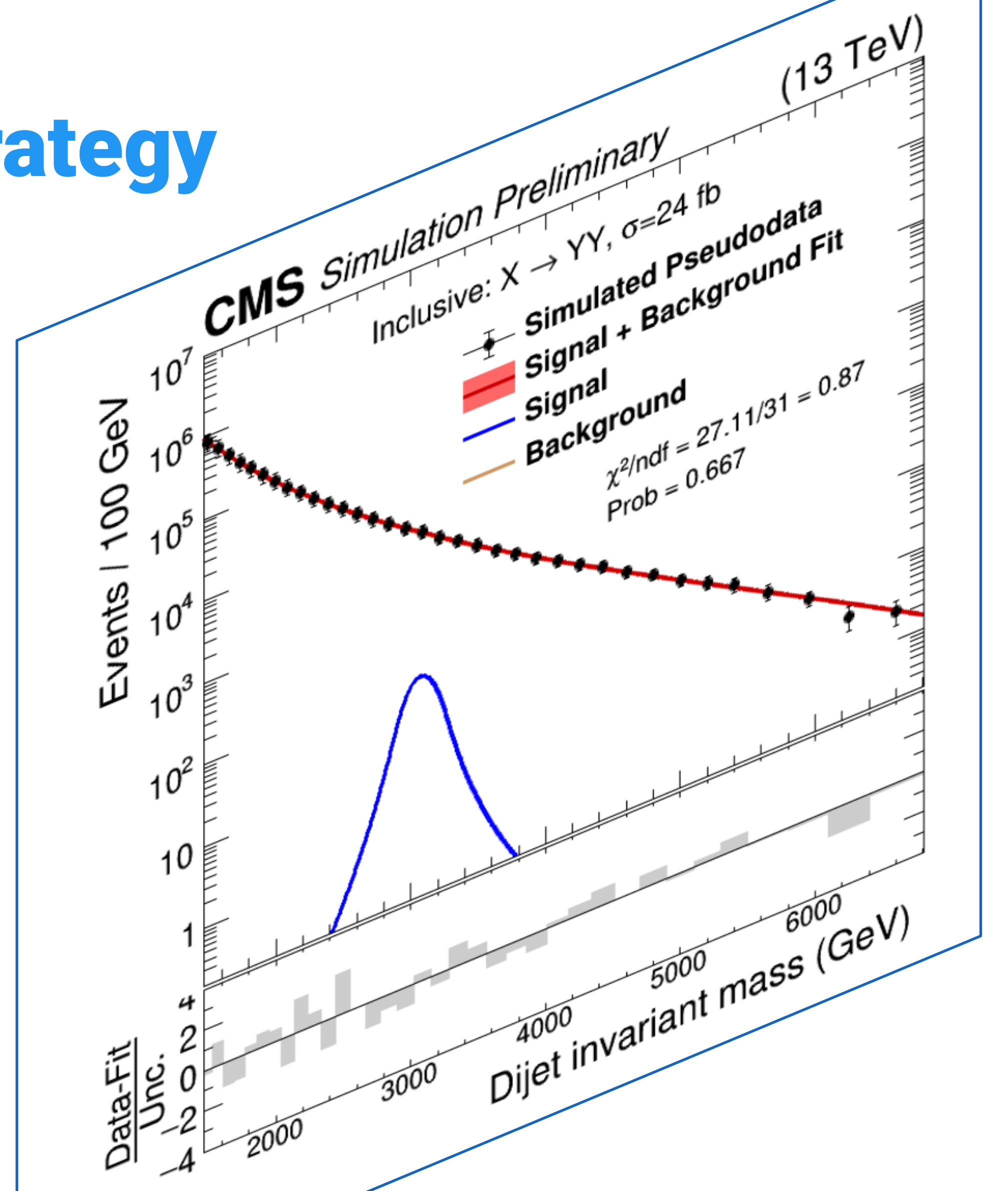




Examples

Z' → T'T' → tZtZ $m_{Z'}$ = [2,3,5] TeV $m_{T'}$ = [400] GeV	
	Y → HH → tttt m_Y = [2,3,5] TeV m_H = [400] GeV

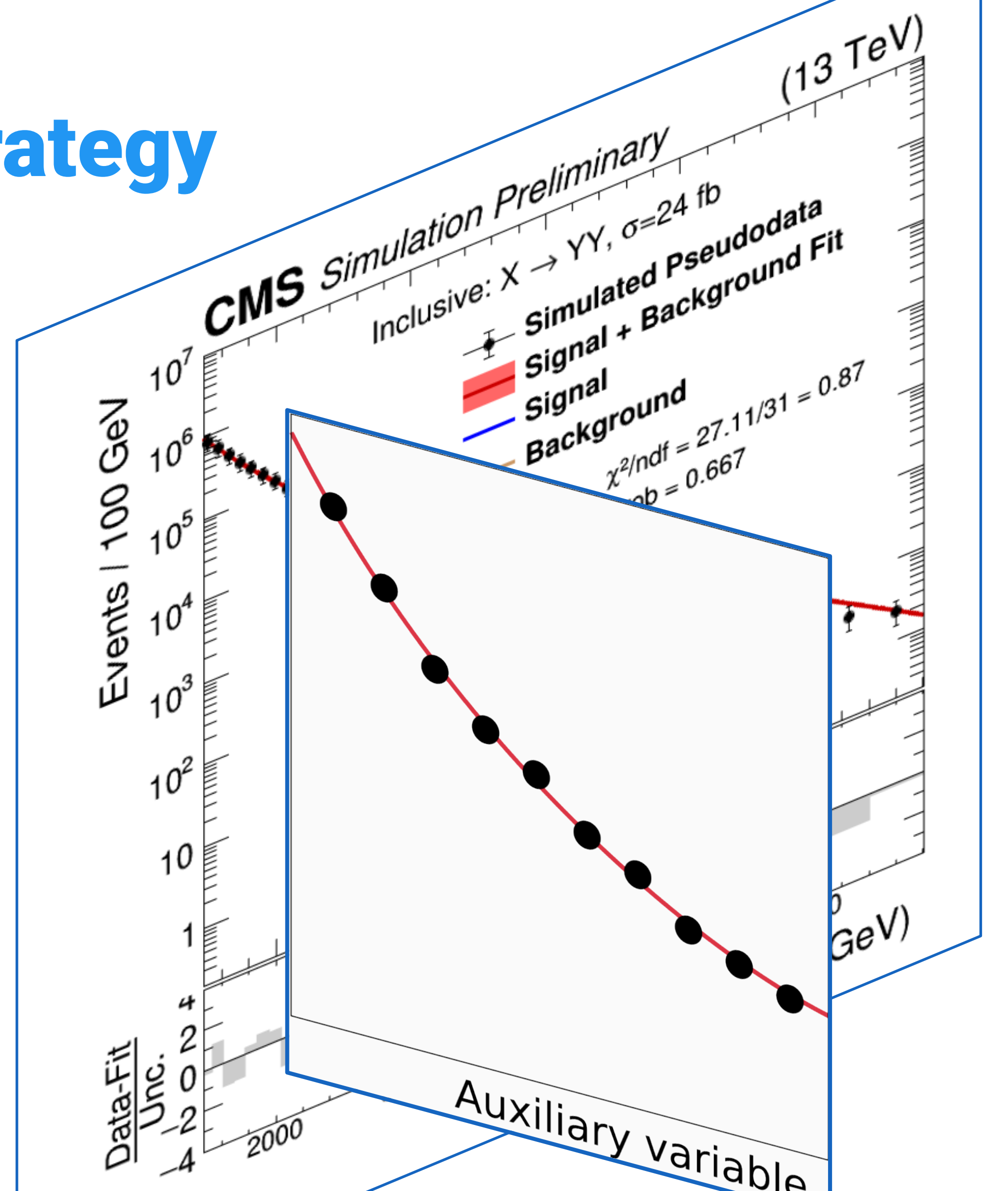
Strategy



We assume a signal would differ from the background in some auxiliary variable

(= in jet substructure)

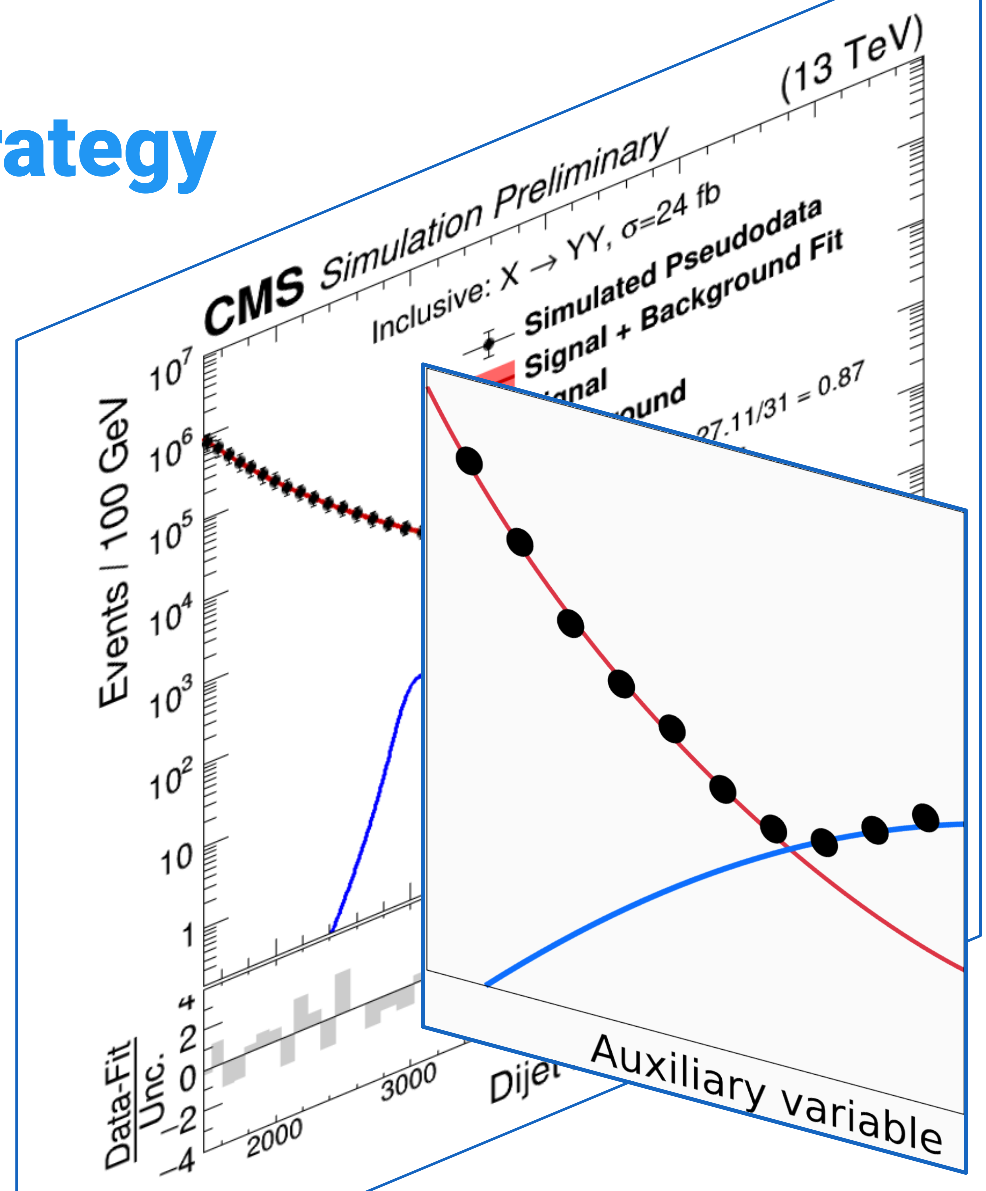
Strategy



We assume a signal would differ from the background in some auxiliary variable

(= in jet substructure)

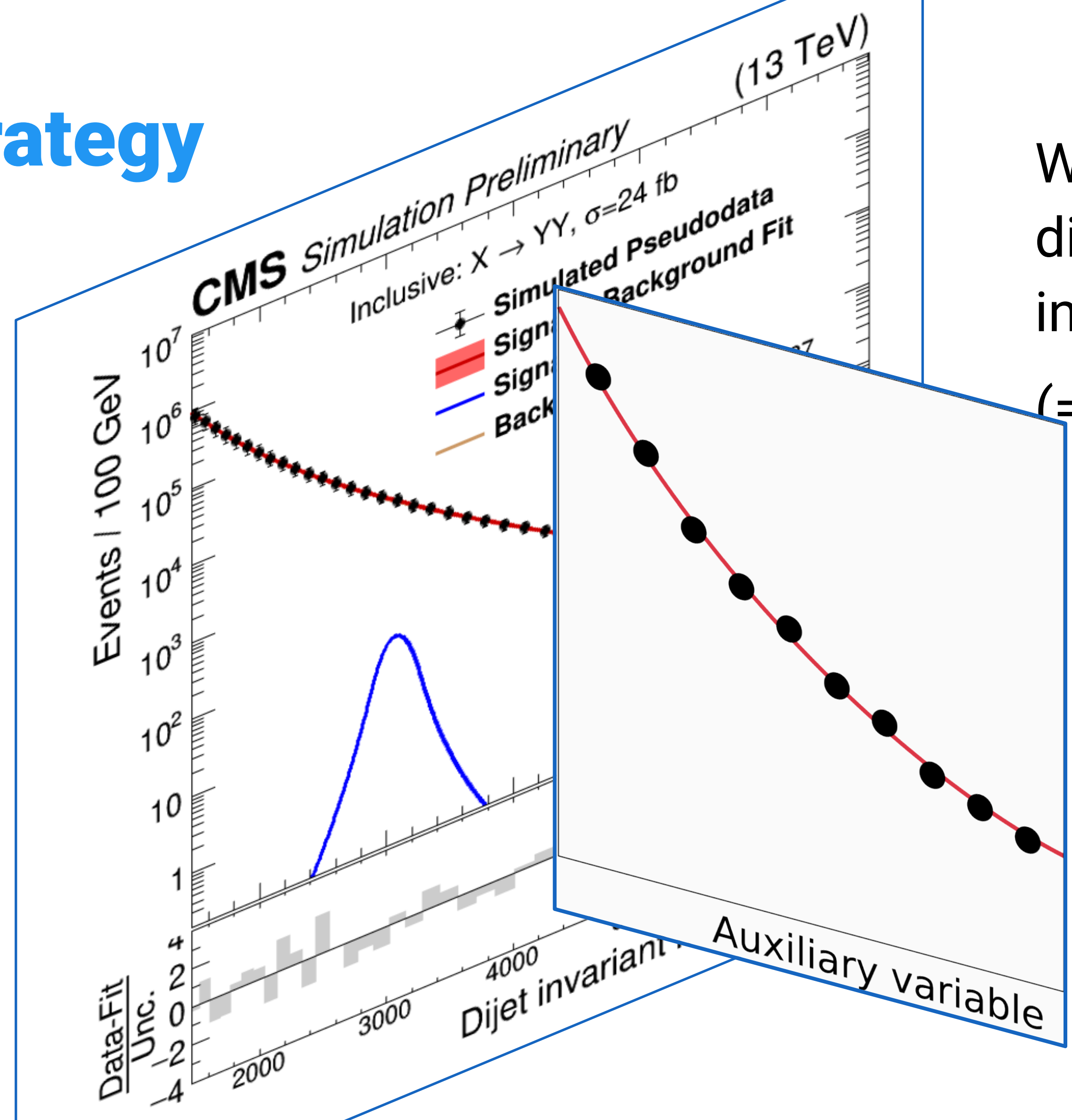
Strategy



We assume a signal would differ from the background in some auxiliary variable

(= in jet substructure)

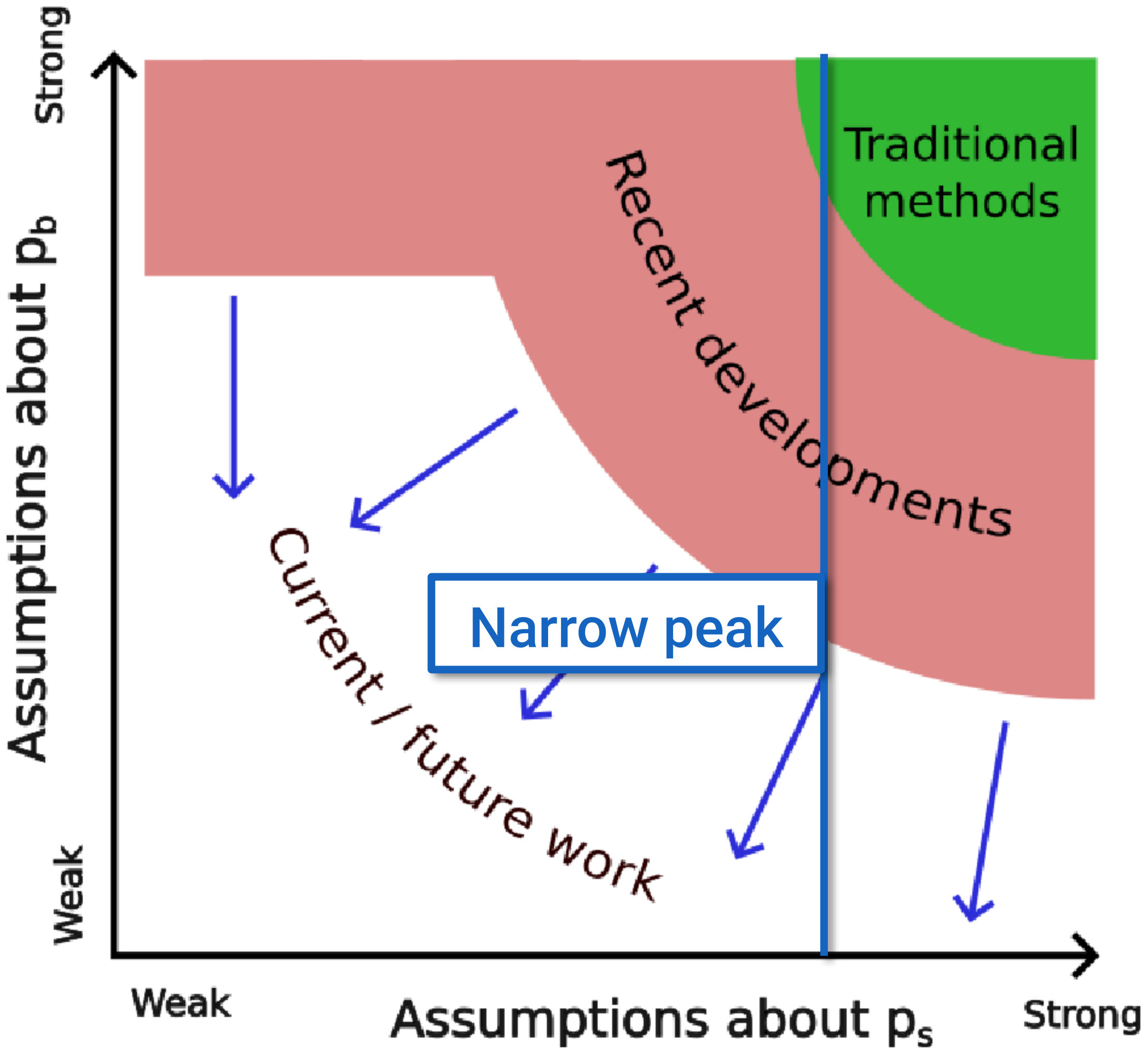
Strategy



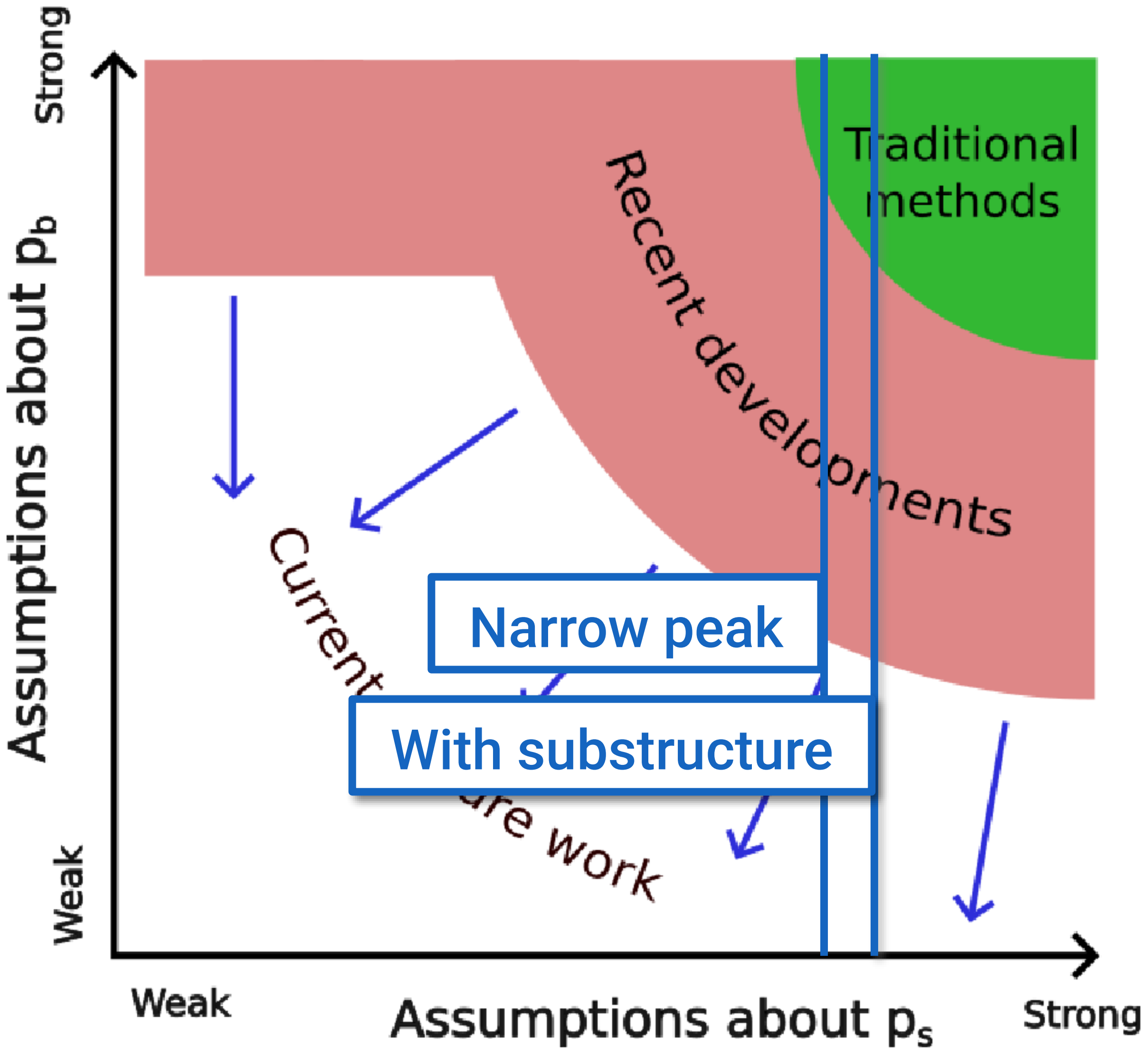
We assume a signal would differ from the background in some auxiliary variable

(= in jet substructure)

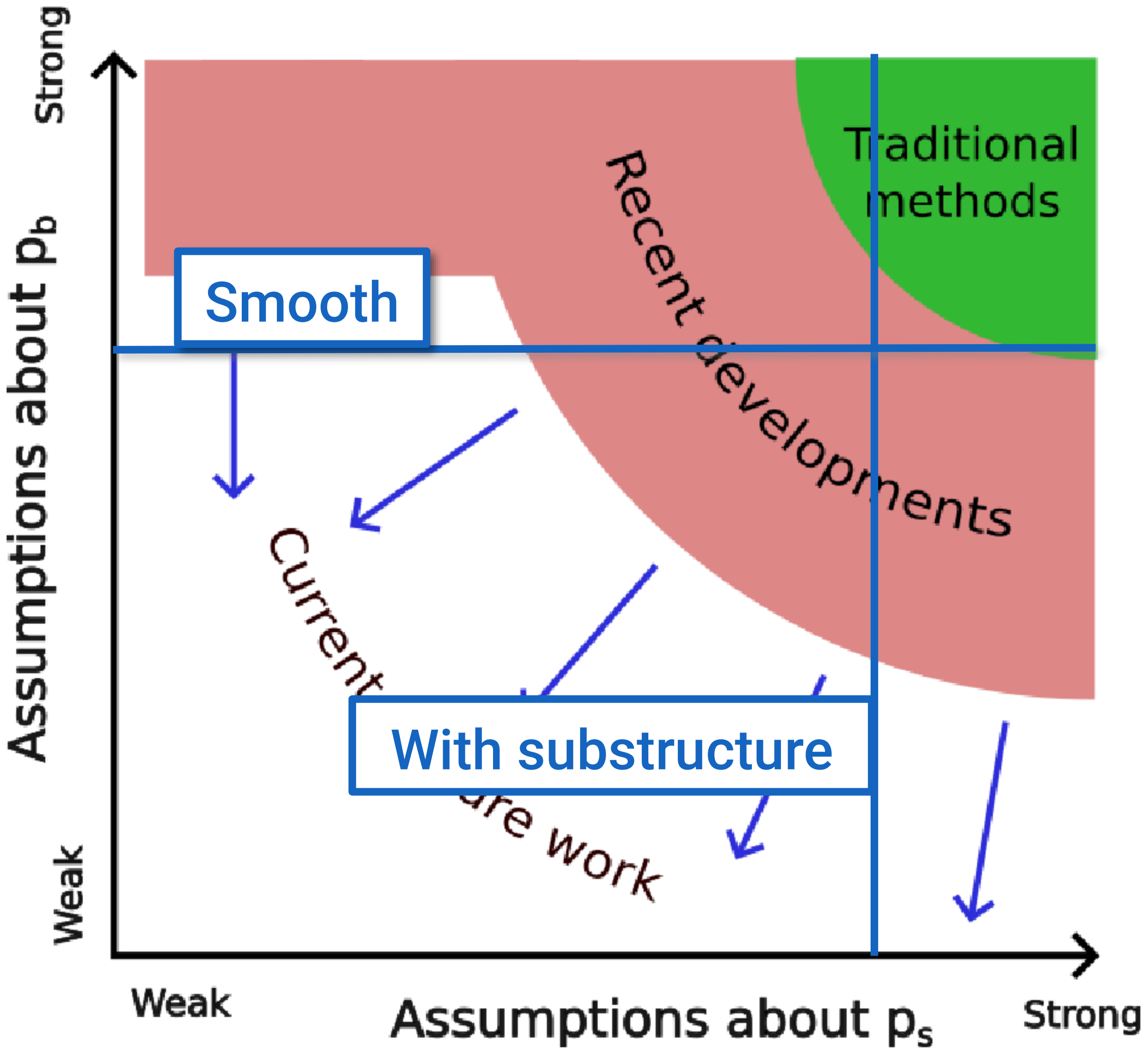
Assumptions



Assumptions



Assumptions



Methods

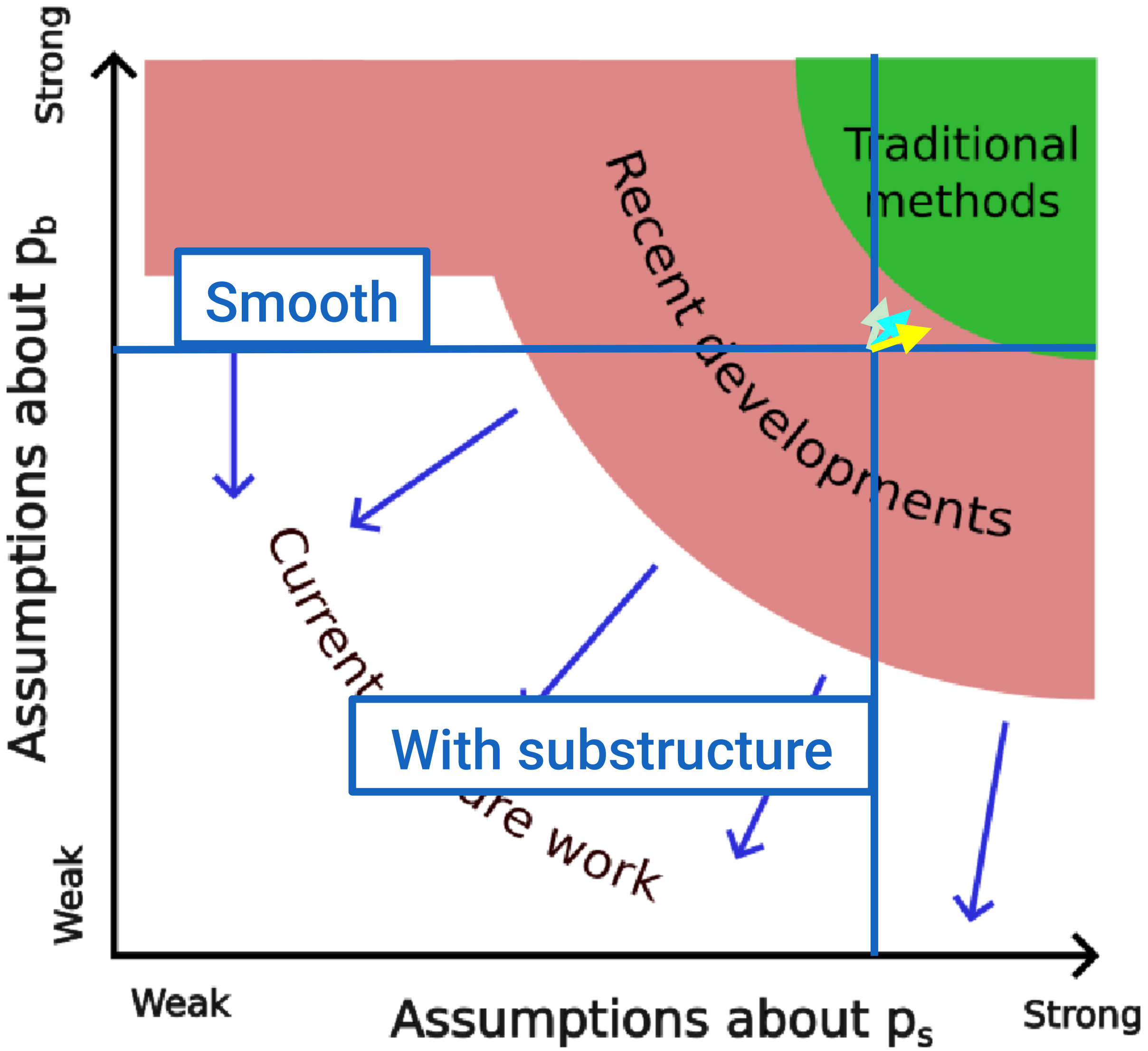
5 techniques starting from different assumptions:

- Signal is in the tails ➔ **VAE** = Variational Autoencoder
- Narrow signal peak ➔ **CATHODE, CWoLa** Hunting
  & *two* anomalous jets ➔ **TNT** = Tag'N'Train

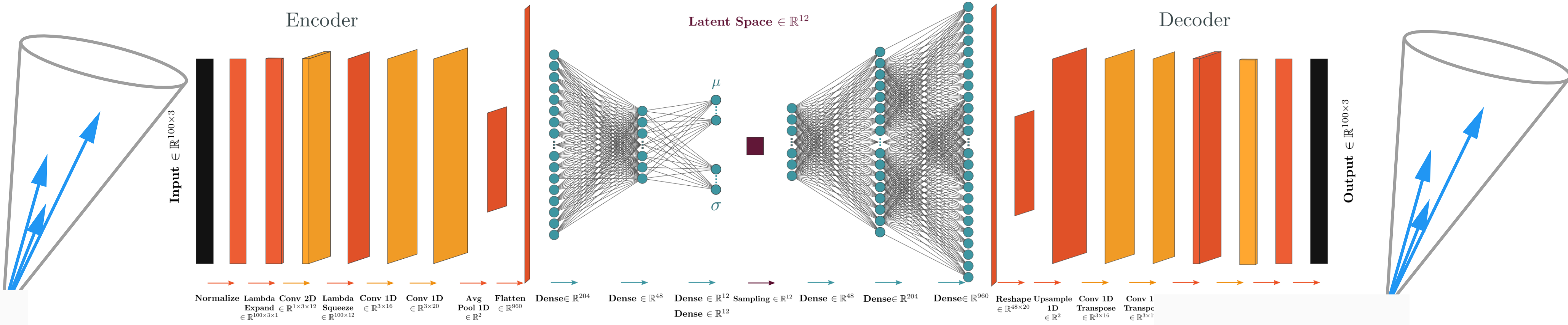
  Weak supervision

- Take bias from theory ➔ **QUAK** = Quasi-Anomalous Knowledge

Assumptions



Methods: Variational Autoencoder



- Encodes up to 100 PF* constituents per jet
- Trained with jets from a QCD-dominated sideband ($\Delta\eta$ > 1.4)
- Final score: lowest reconstruction loss of the two jets
- Background sculpting controlled with quantile regression

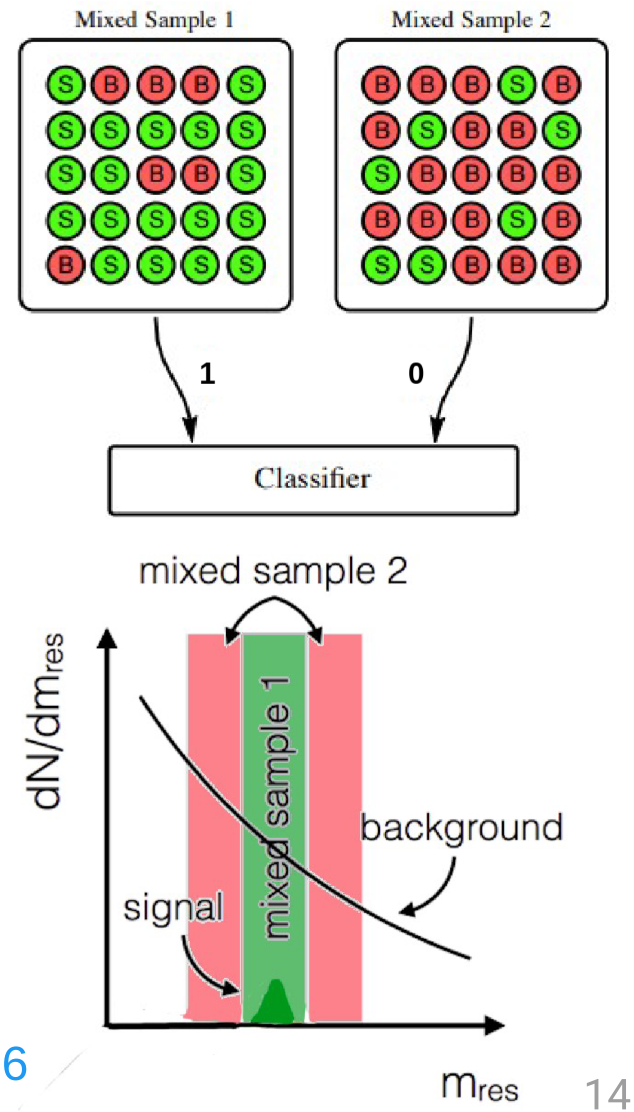
* Particle Flow

Methods: **Weak Supervision**

Train a classifier between data and a background-like sample

- **CWoLa:** background taken from sidebands
- **CATHODE:** background interpolated from sidebands
- **Tag N' Train:** autoencoder preselection, targets events with **two** anomalous jets

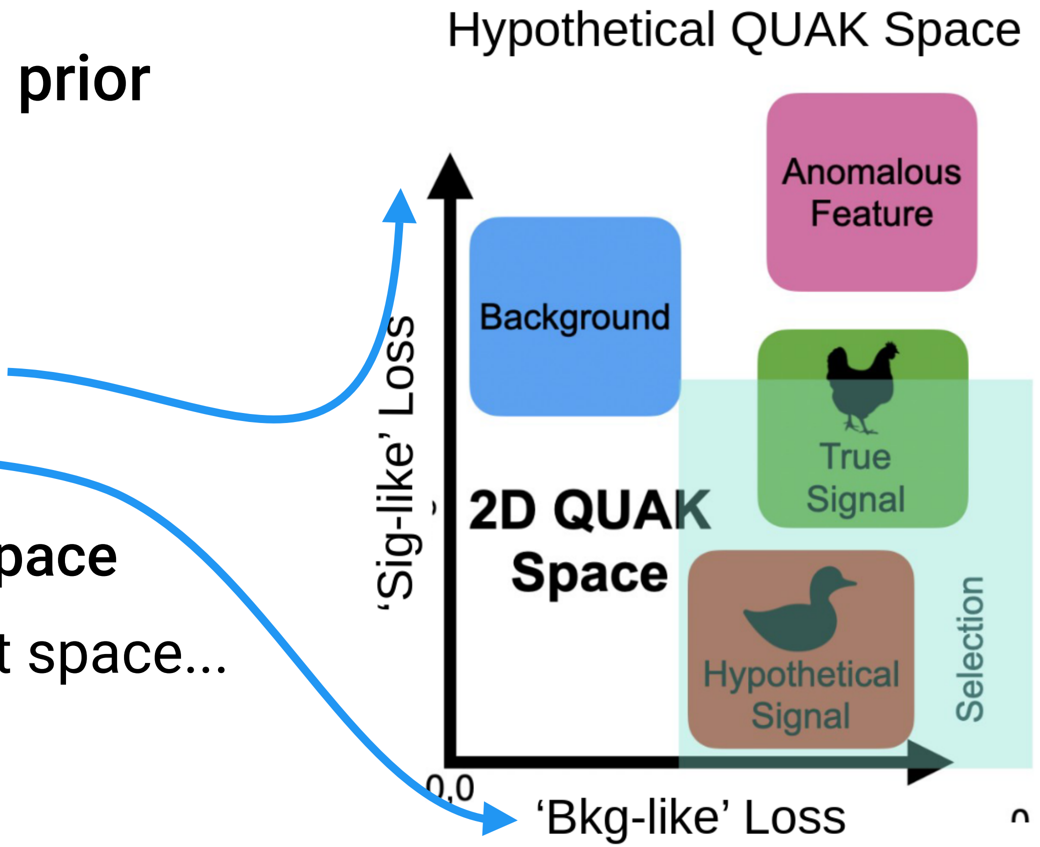
Fewer features for CATHODE than CWoLa/TNT



CWoLa: 1902.02634 / CATHODE: 2109.00546 / TNT: 2002.12376

Methods: **QUAK**

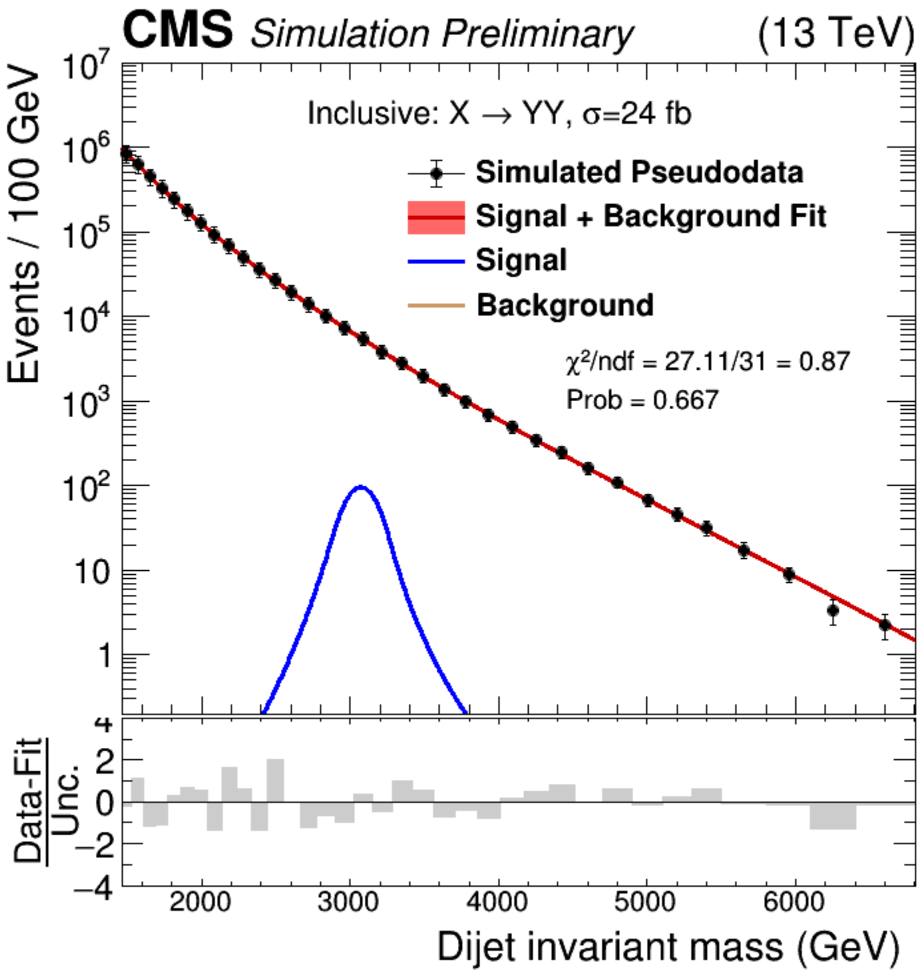
- Hybrid approach, encoding a **prior** on signal-like features
- Train two normalizing flows:
  - On a mixture of signal MCs
  - On background MC
- The losses define a **2D QUAK space**
- The signal is somewhere in that space...



2011.03550

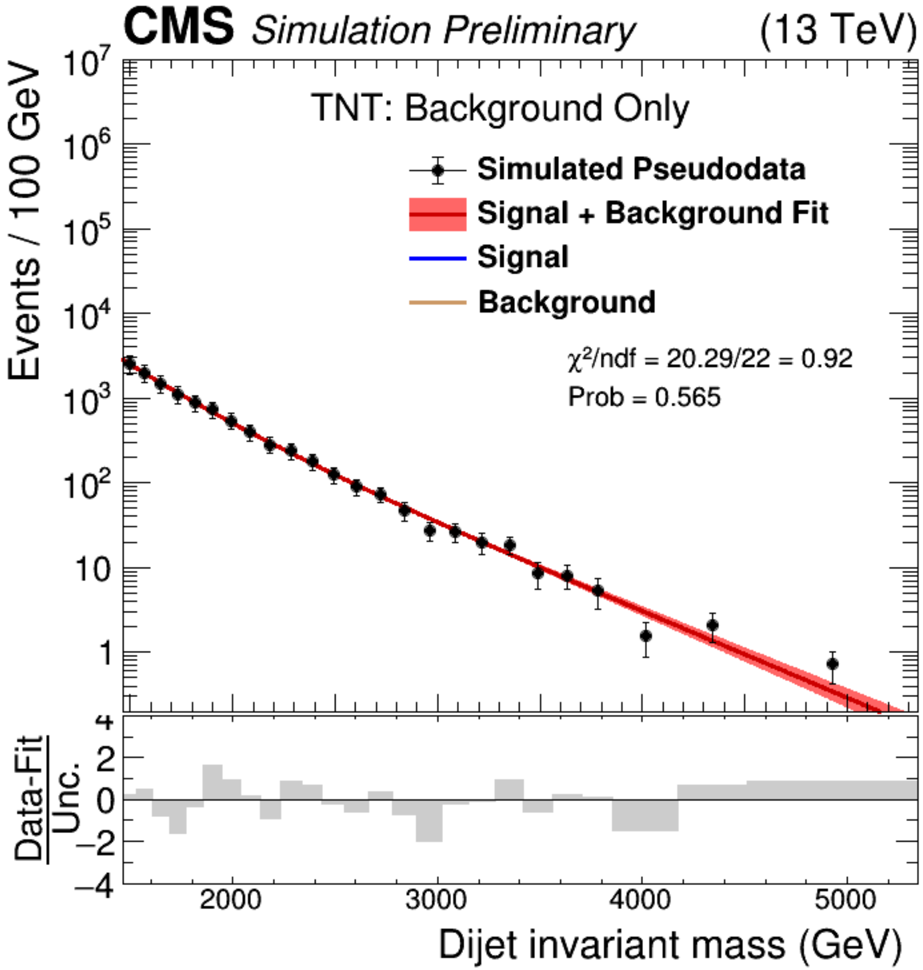
After Tagging

- Choose a working point



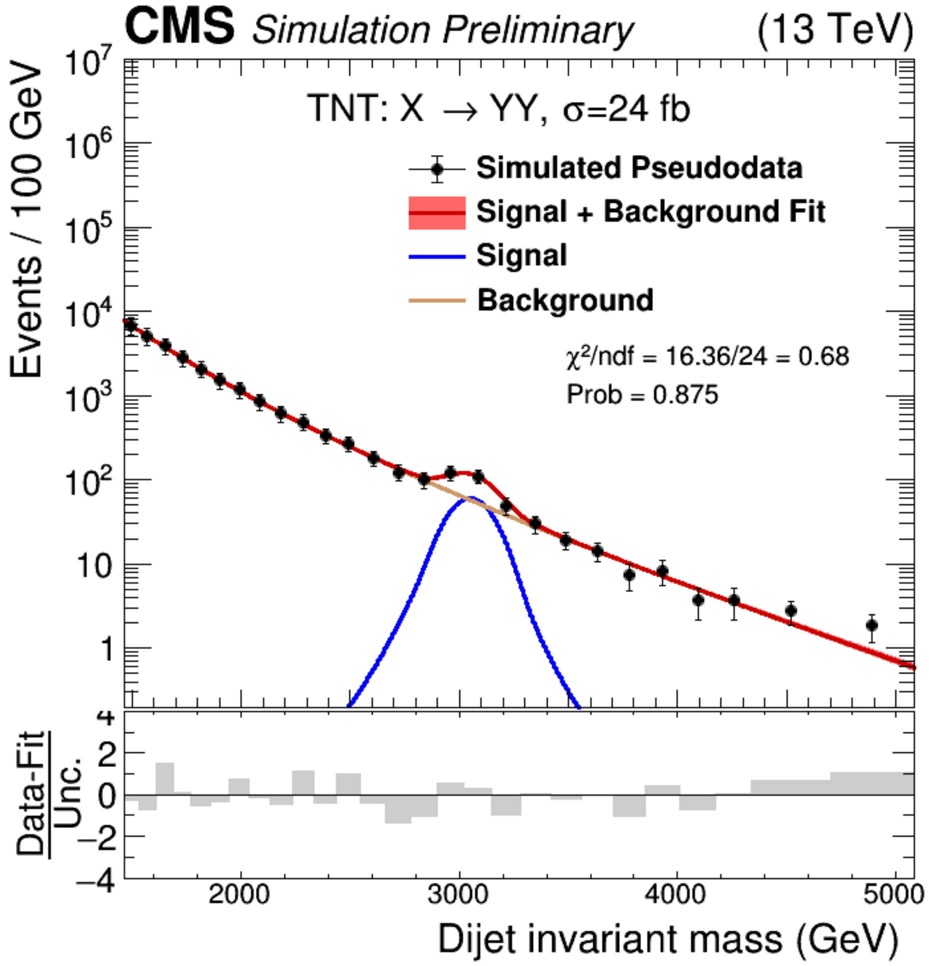
After Tagging

- Choose a working point
- Select events
- Look at m_{jj} spectrum
- Fit with analytic functions



After Tagging

- Choose a working point
- Select events
- Look at m_{jj} spectrum
- Fit with analytic functions
- Find a bump (maybe)
- Derive a *p*-value

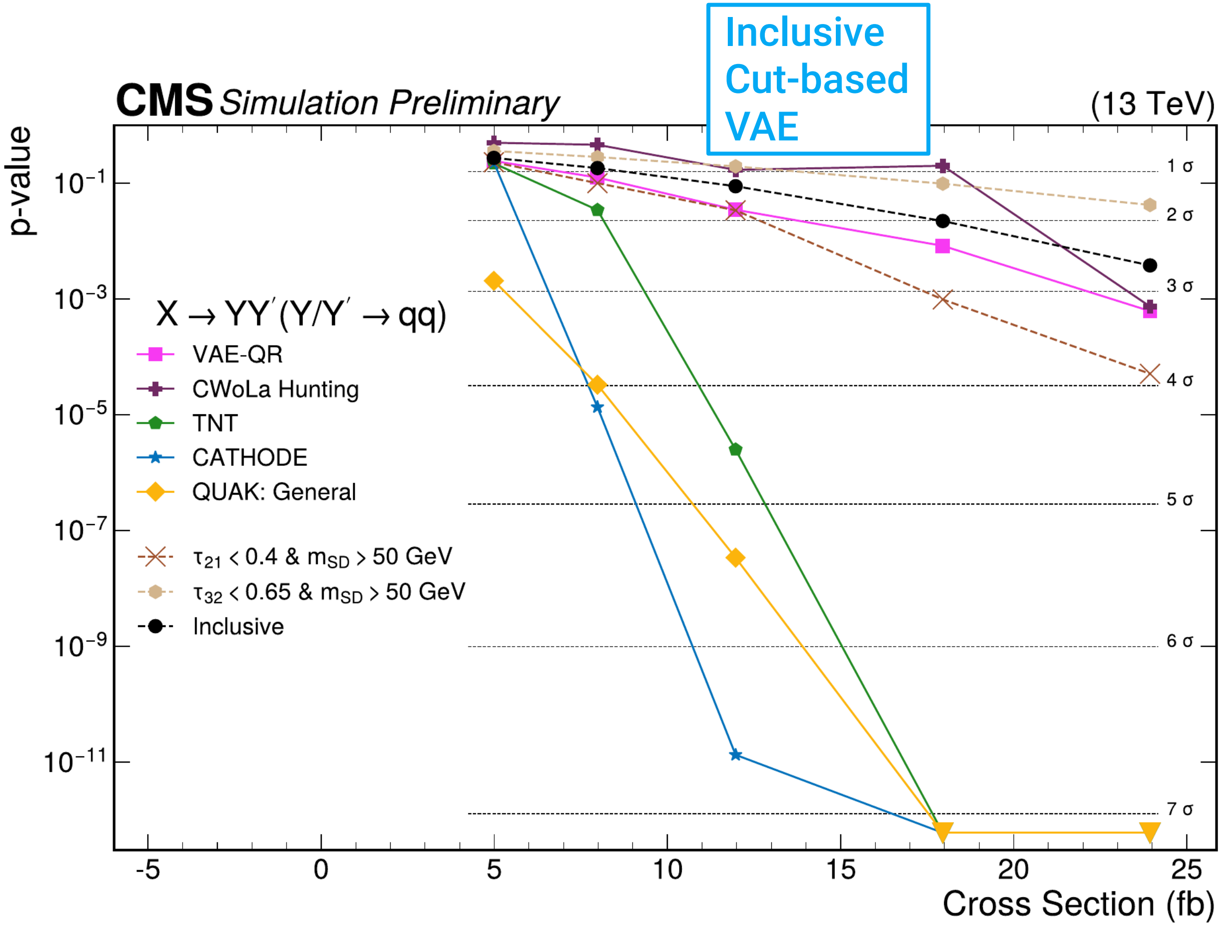


Performance: 2 + 2

Testing performance on:

X(3000) → YY' → qq qq

Two W-like jets

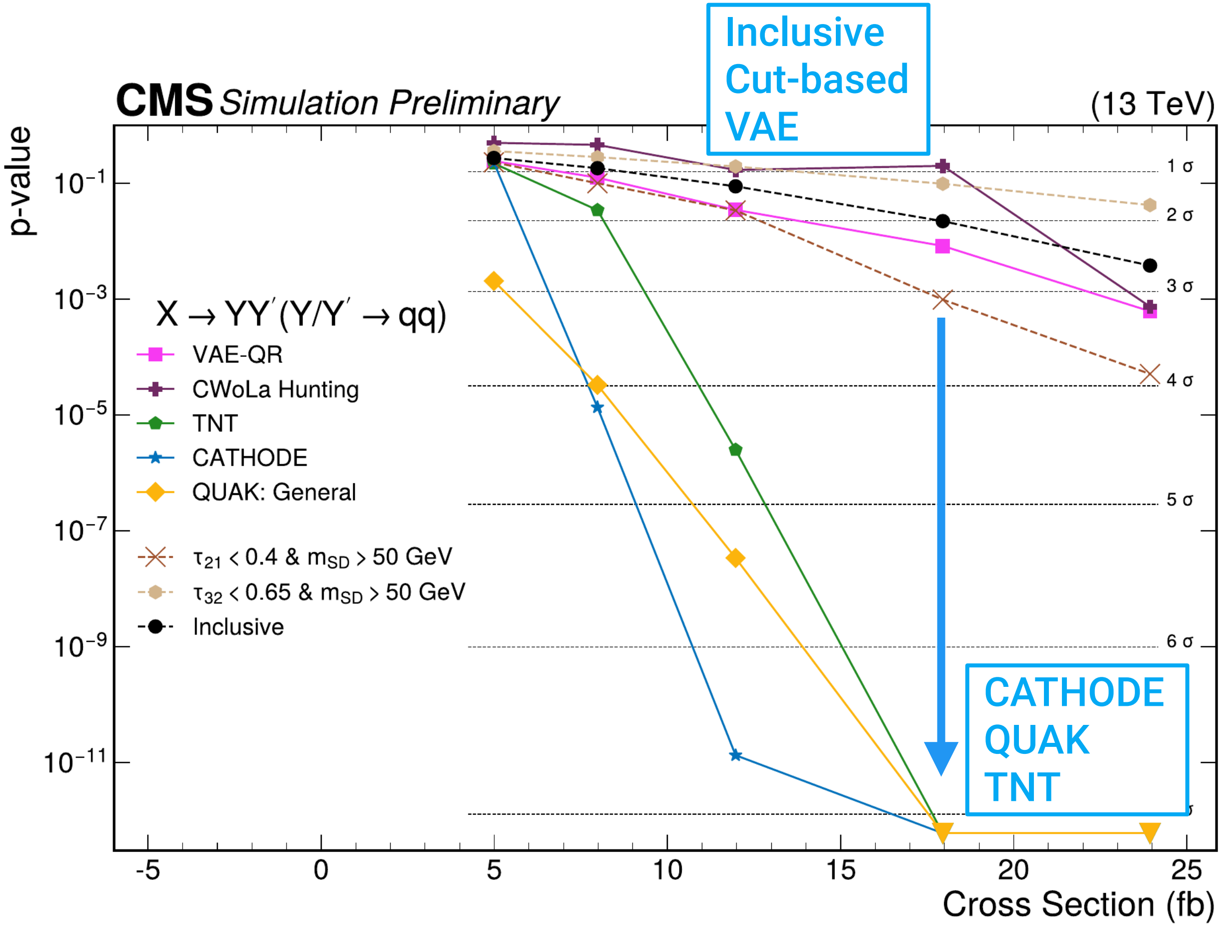


Performance: 2 + 2

Testing performance on:

X(3000) → YY’ → qq qq

Two W-like jets



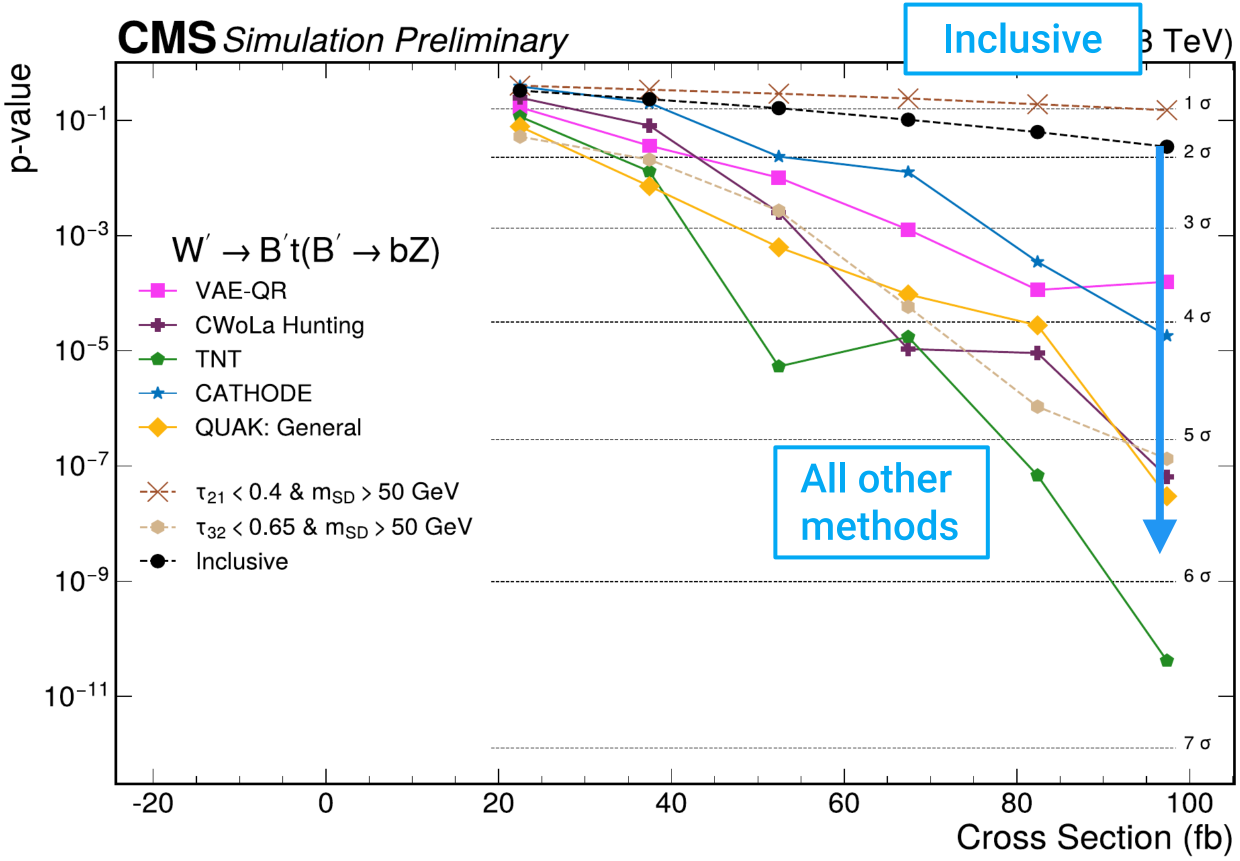
Performance: 3 + 3

Testing performance on:

W' → B't → qqq qqq

3 + 3

Two top-like jets

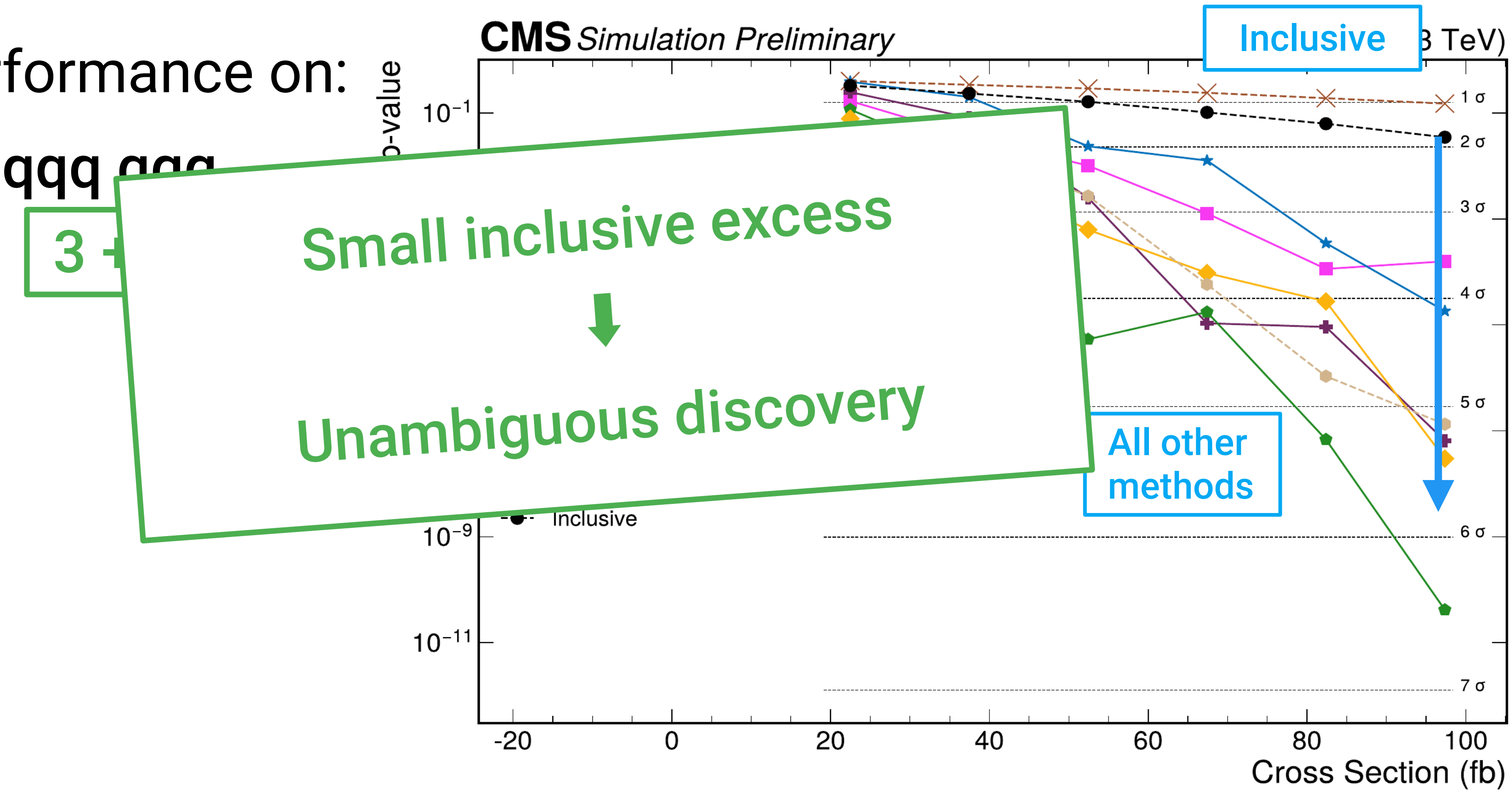


Performance: 3 + 3

Testing performance on:

W’ → B’t → qqq qqq

3 +



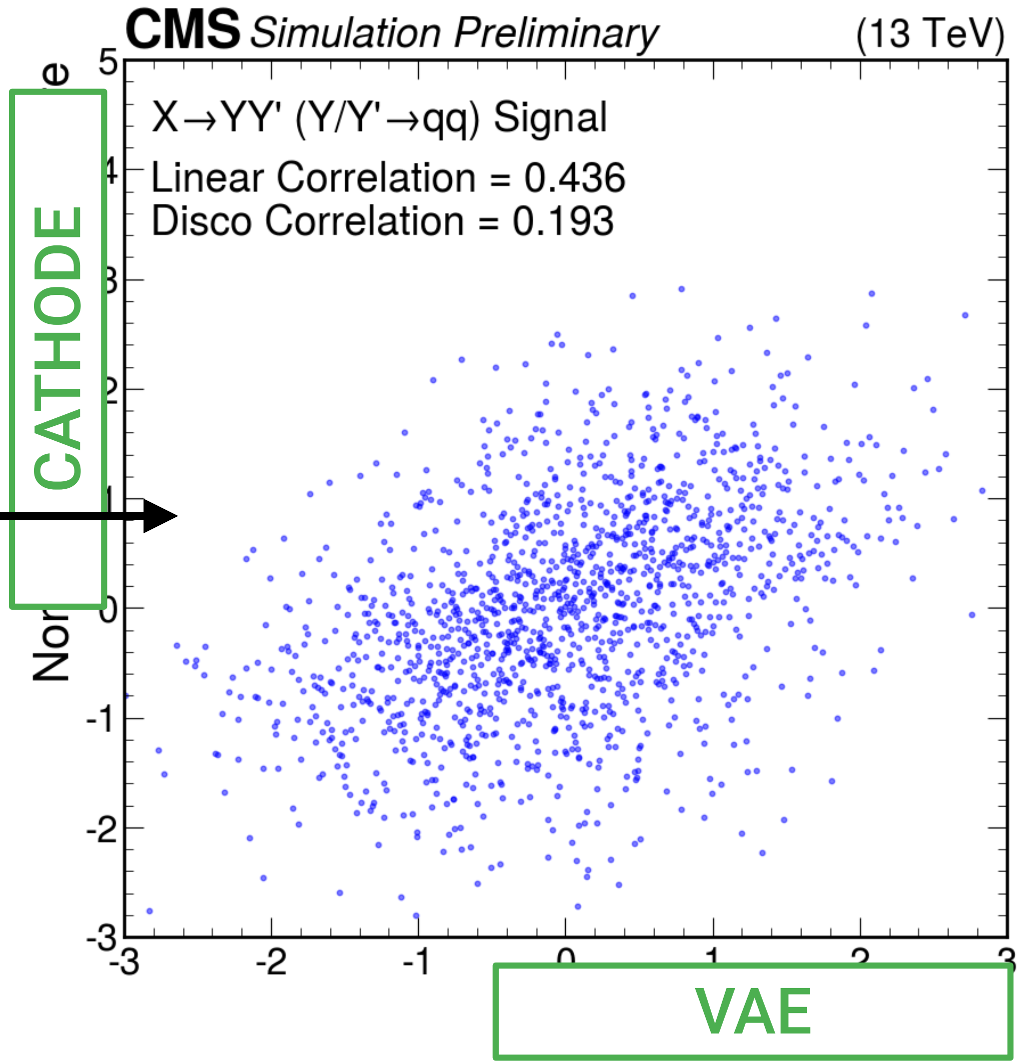
Complementarity

Do all methods find
the same events?

Complementarity

Do all methods find the same events?

Check correlation between scores



Two W-like jets

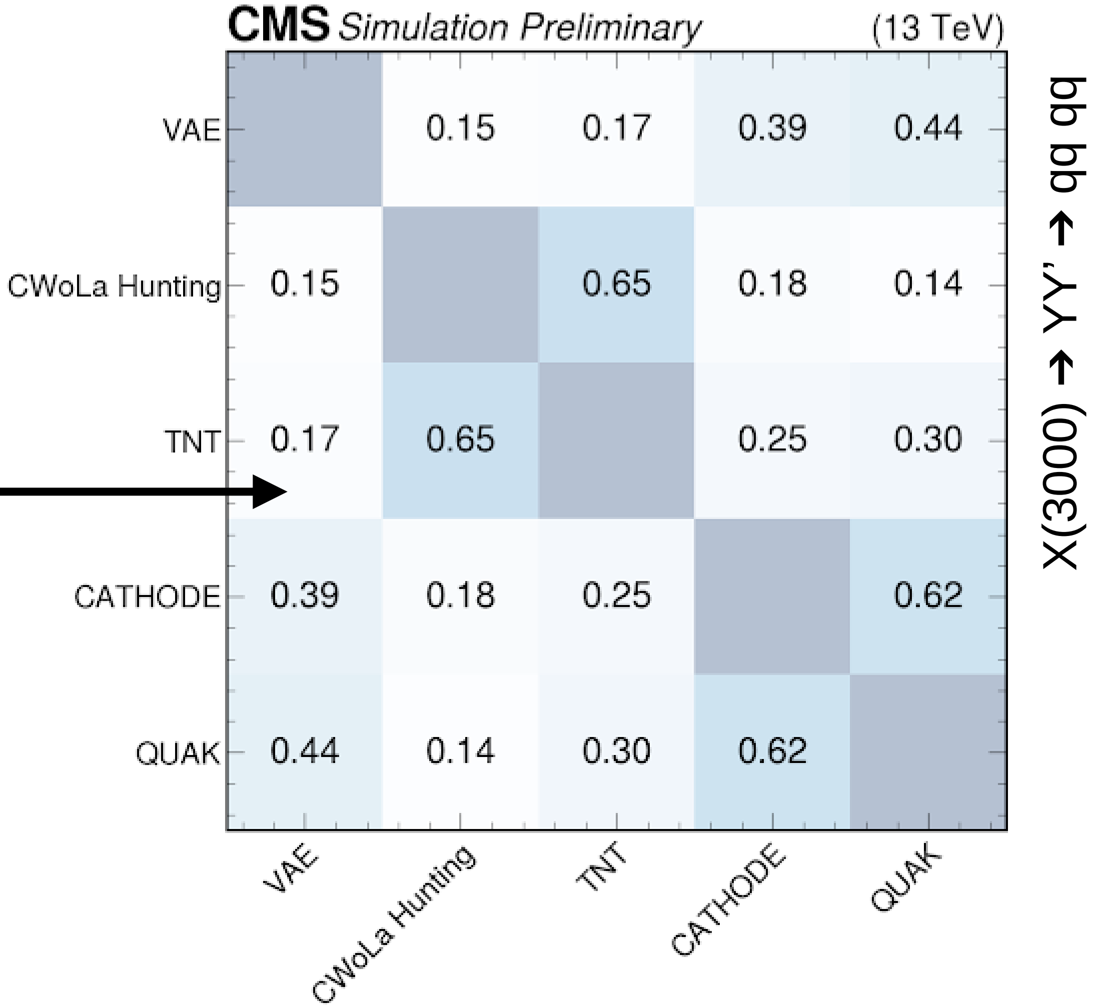
Signal:

X(3000) → YY' → qq qq

Complementarity

Do all methods find the same events?

Check correlation between scores



Small correlations

➡ Complementarity

Summary

- CMS joining the anomaly detection party
- Looking for dijet resonances with **5** methods:

  VAE, CWoLa, CATHODE, Tag N' Train, QUAK
- Promising performance in simulation
- No method to rule them all

Results on CDS [link]

Finalizing analysis in data: stay tuned

CMS

Compact Muon Solenoid

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Federal Ministry of Education and Research

CERN

KIT
Karlsruhe Institute of Technology

Fermilab

ETH zürich

JOHNS HOPKINS
UNIVERSITY

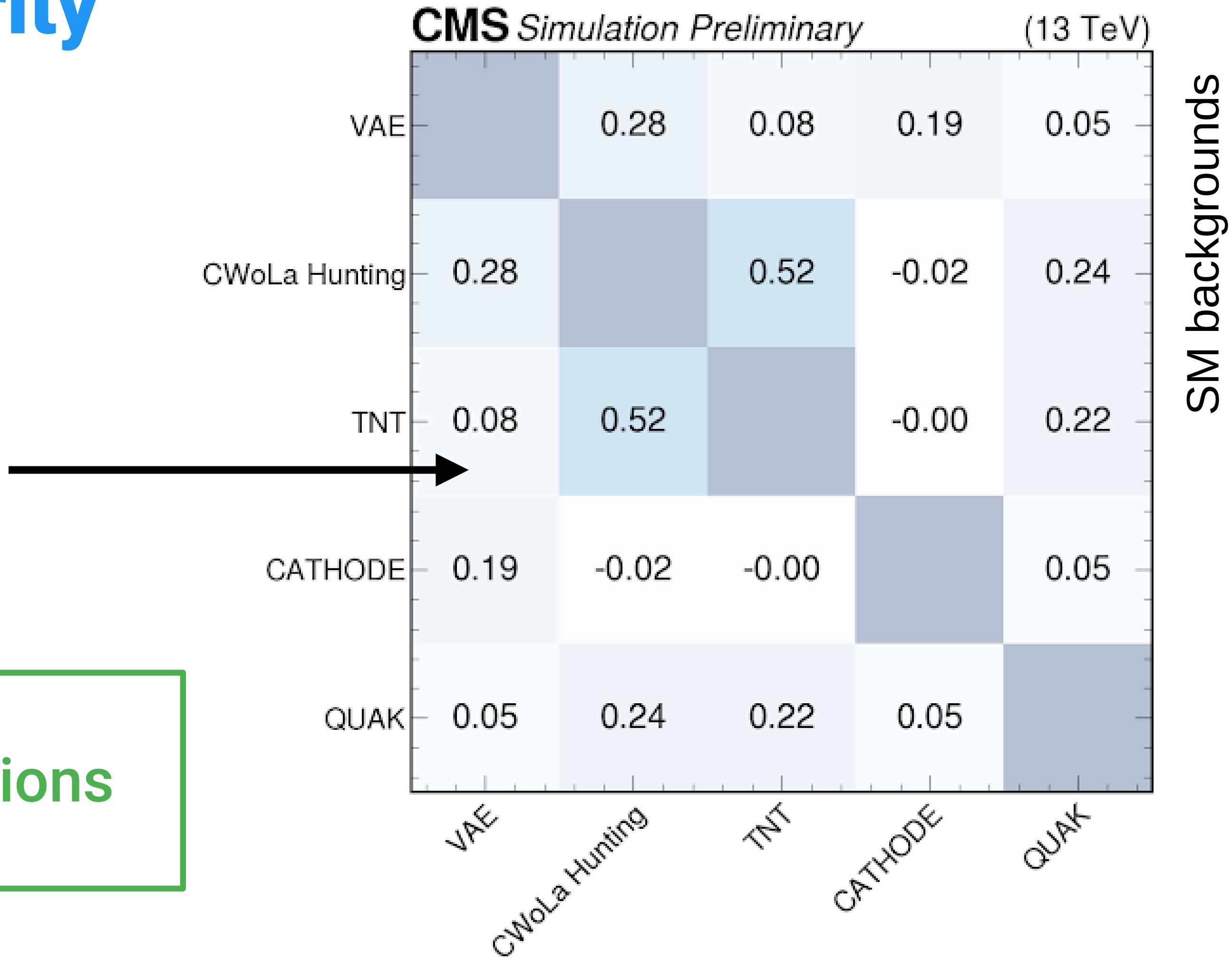
Input Features

- **VAE:** p_T, η, ϕ of leading 100 particle flow constituents (per jet)
- **CWoLa, TNT:** m_{SD}, τ_{21}, τ_{32}, τ_{43}, n_{PF}, LSF_3, b-tagging score (per jet)
- **CATHODE:** m_{SD1}, $m_{SD1} - m_{SD2}$, $\tau_{41,1}$, $\tau_{41,2}$ (per event)
- **QUAK:** m_{SD}, τ_{21}, τ_{32}, τ_{43}, $\sqrt{\tau_{21}}/\tau_1$, M/p_T (for each jet, per event)

Complementarity

Do all methods find the same events?

Check correlation between scores (background)



Small bg correlations