Treasure Hunting without a Map:
First anomaly detection results from CMS

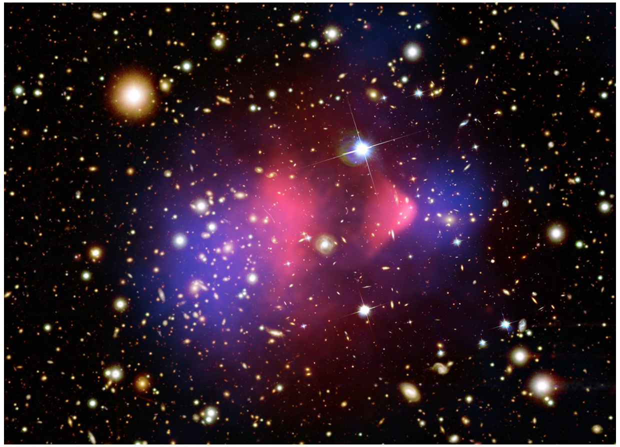
Oz Amram
May 24th, 2024
Fermilab Wine & Cheese

Fermilab

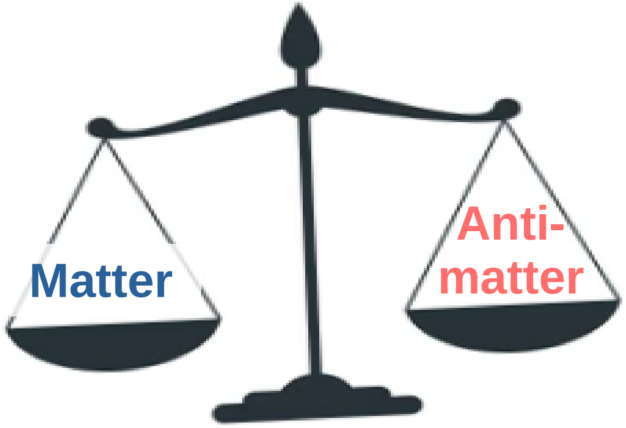
CMS

Lots of Questions

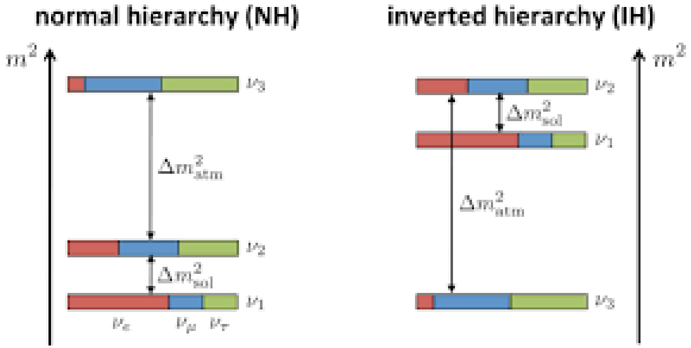
Dark Matter?



Baryogenesis?



Neutrino Mass?



Hierarchy Problem?

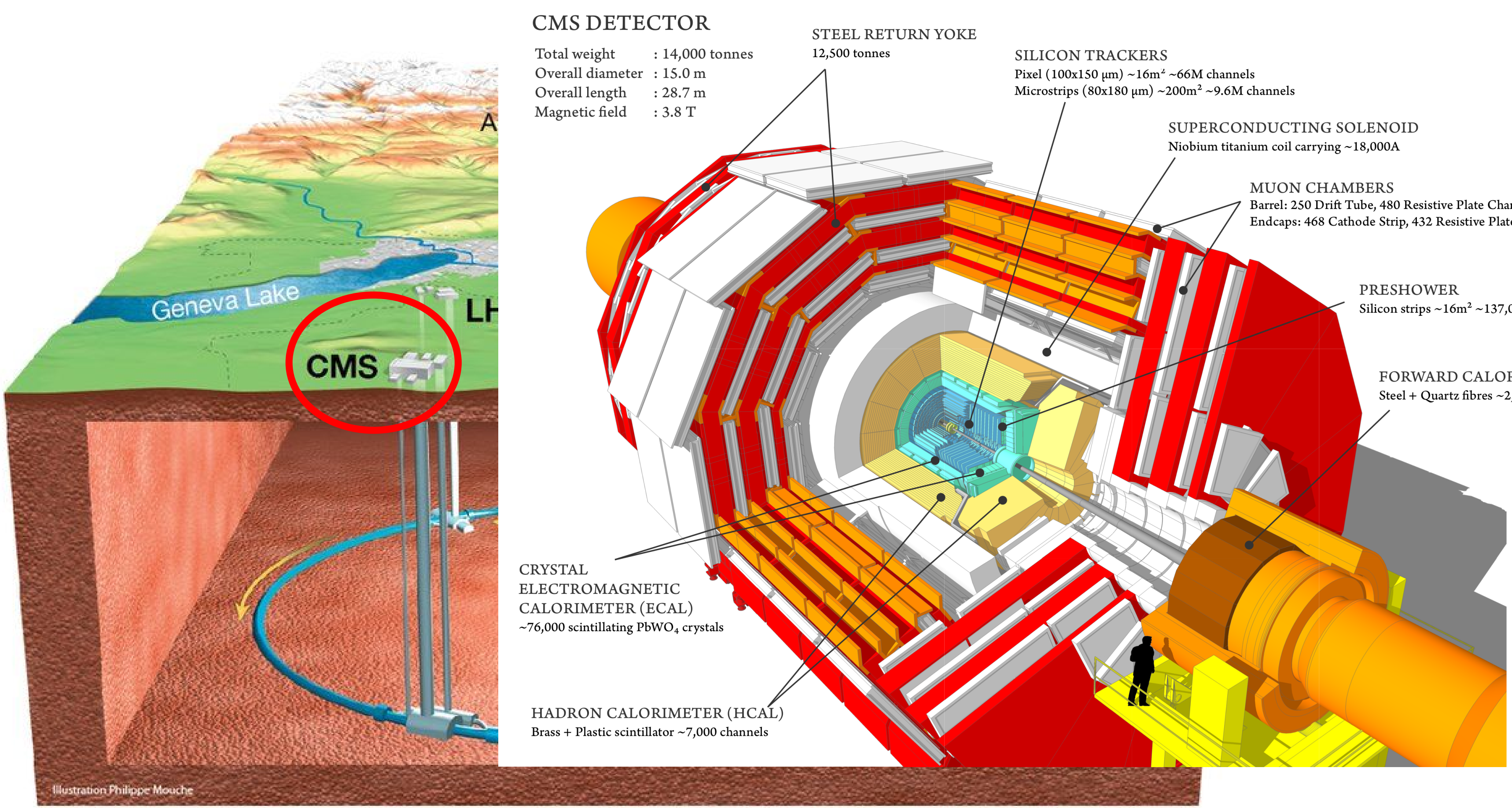
Grand Unification?

And many more...

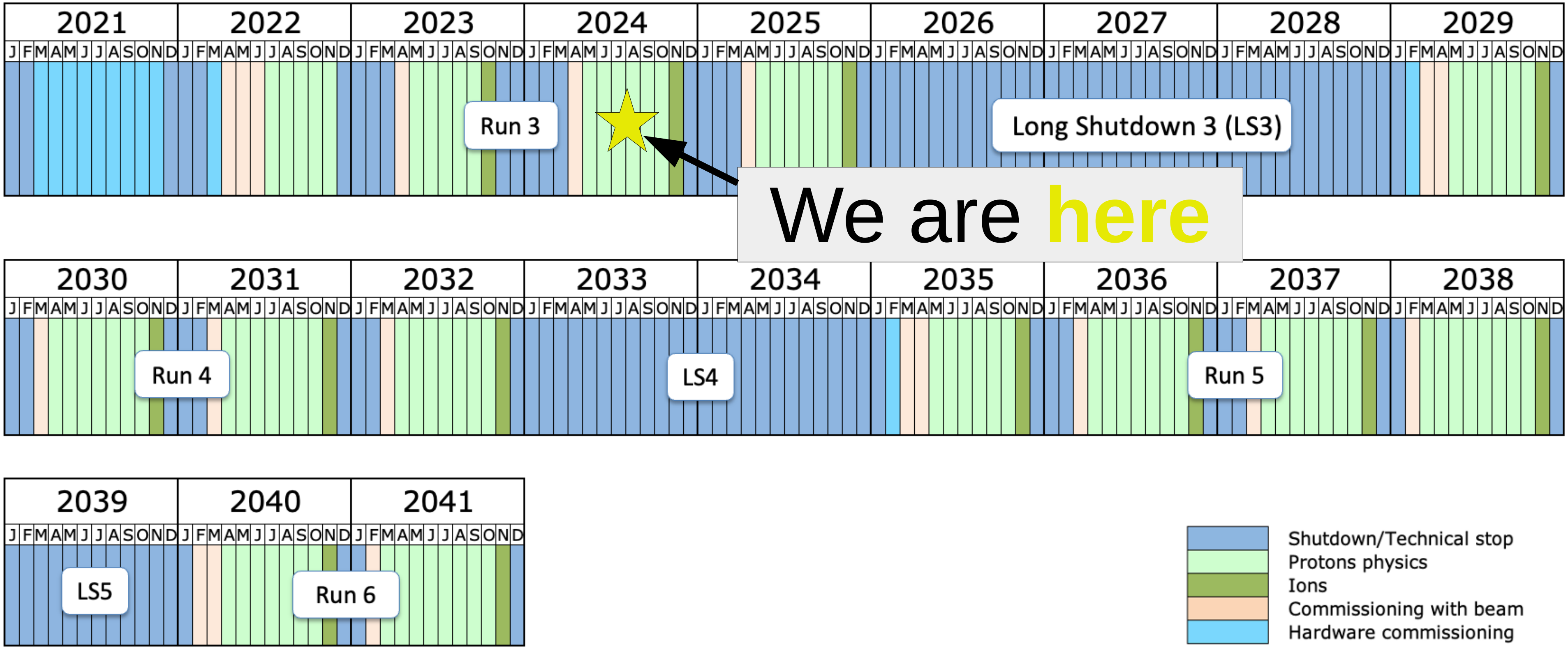
g-2?

Flavor Anomalies?

LHC & CMS

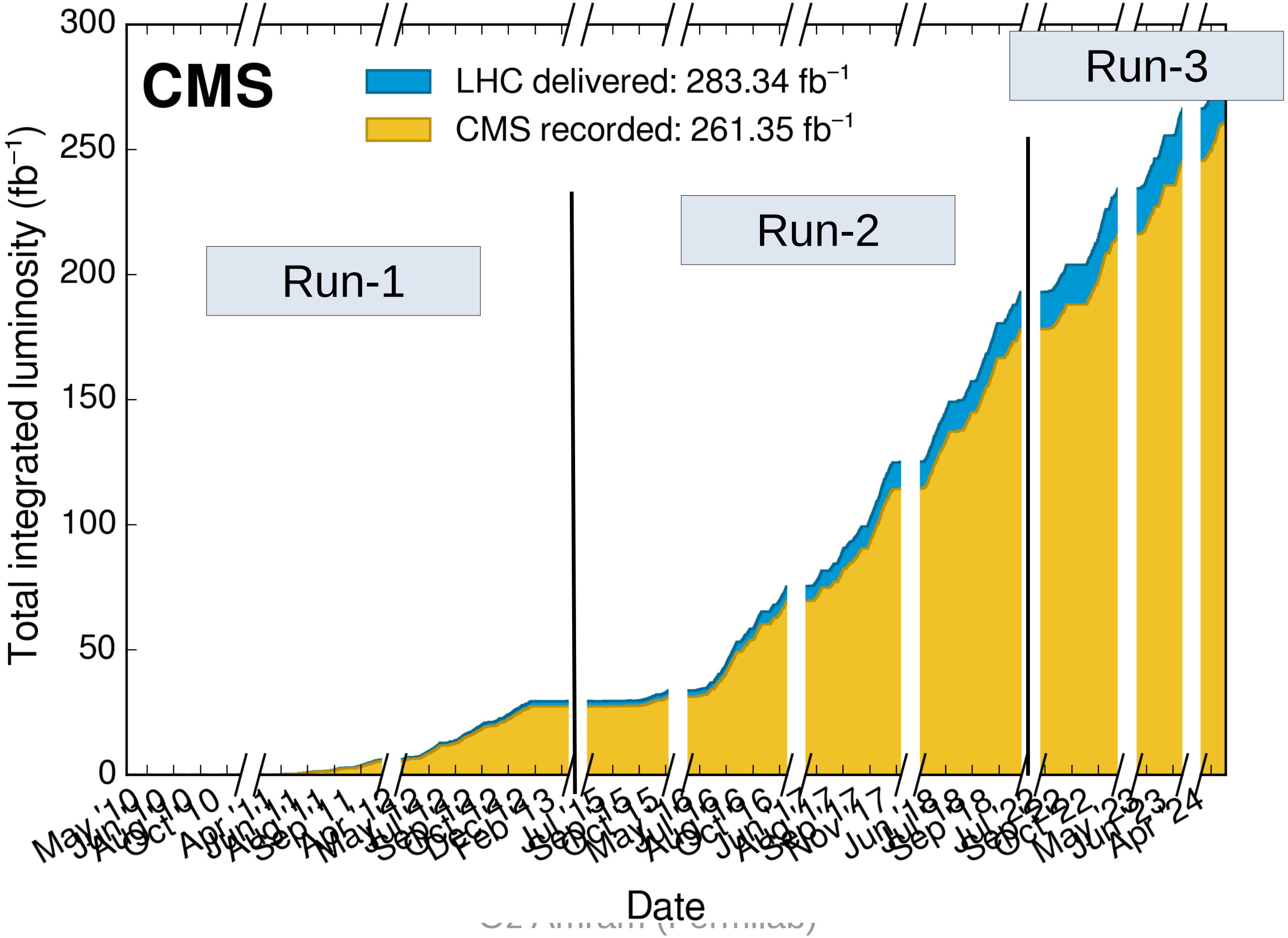

Geneva Lake
CMS
Illustration Philippe Mouche
CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO4 crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

LHC Schedule



Last update: April 2023

CMS Data



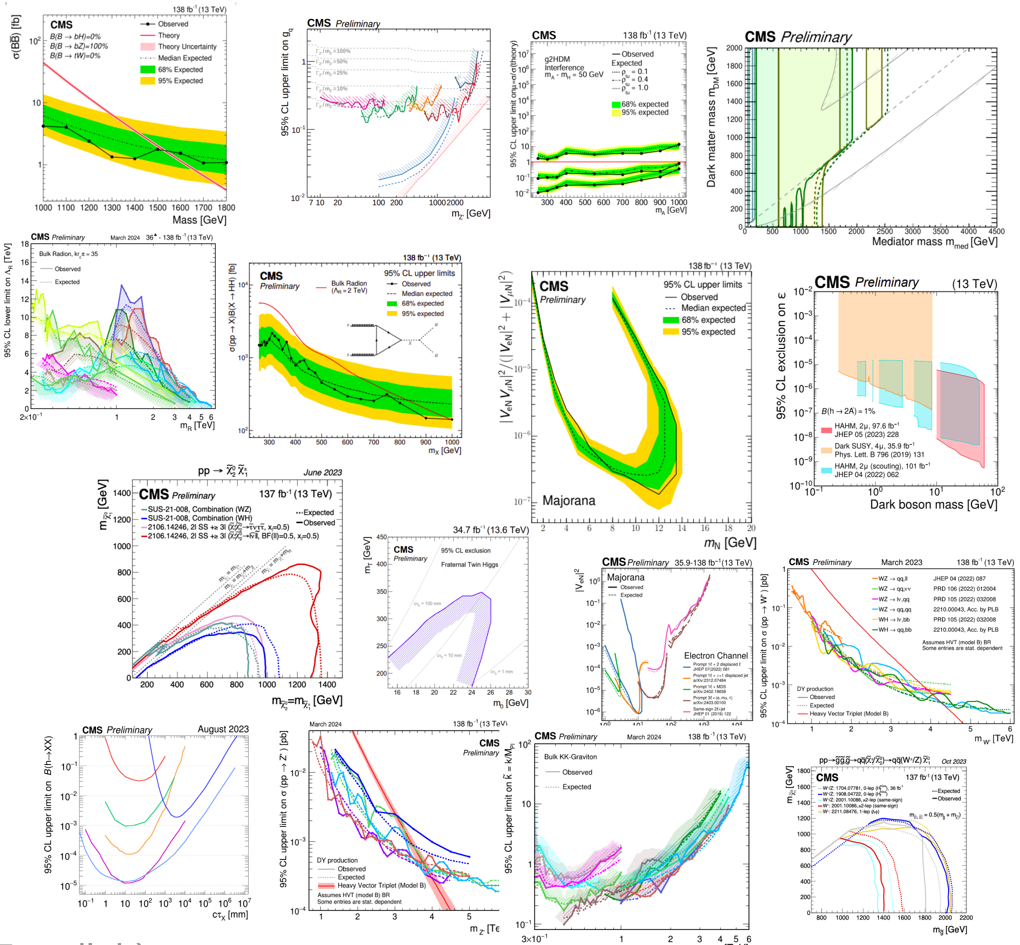
CMS
LHC delivered: 283.34 fb^{-1}
CMS recorded: 261.35 fb^{-1}
Run-1
Run-2
Run-3
Total integrated luminosity (fb^{-1})
Date

LHC Answers?

DM graphic from T. Tait

LHC Answers?

None so far...

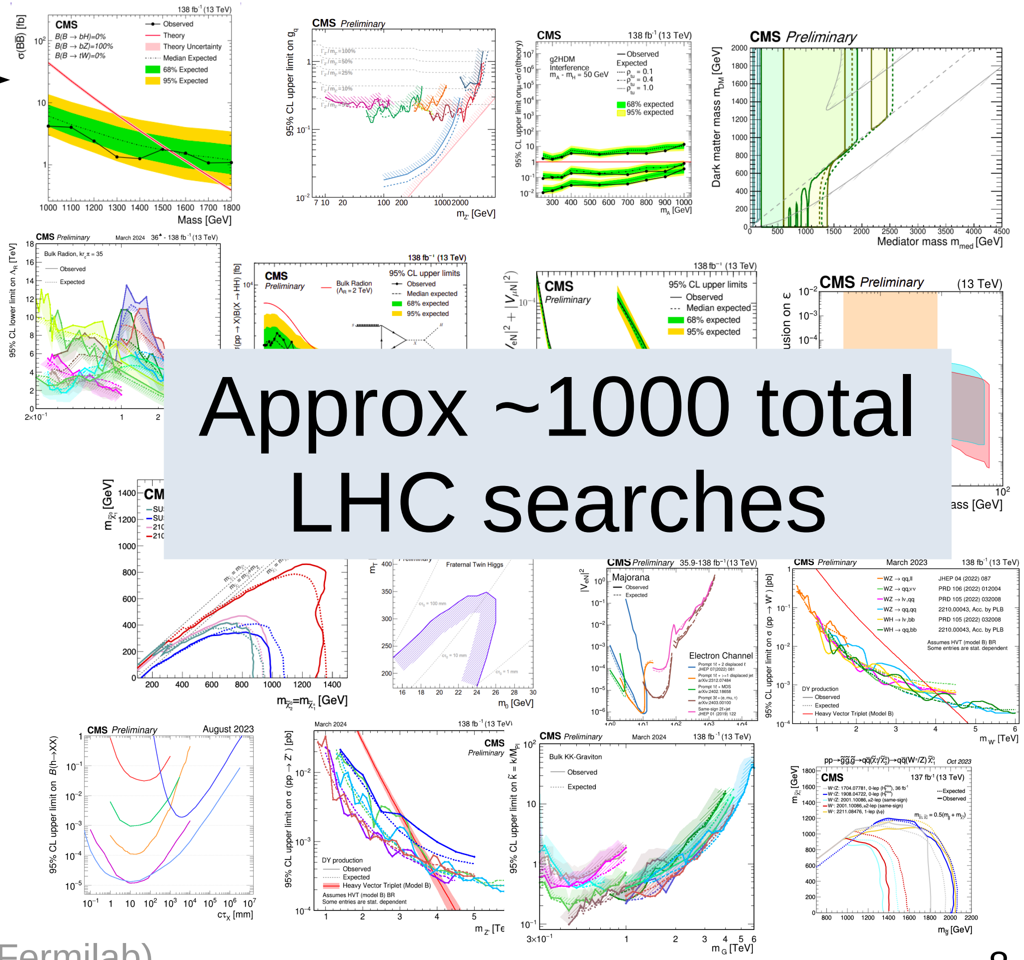


DM graphic from T. Tait

LHC Answers?

Each is ~3 person-years of work

Huge amount of effort!

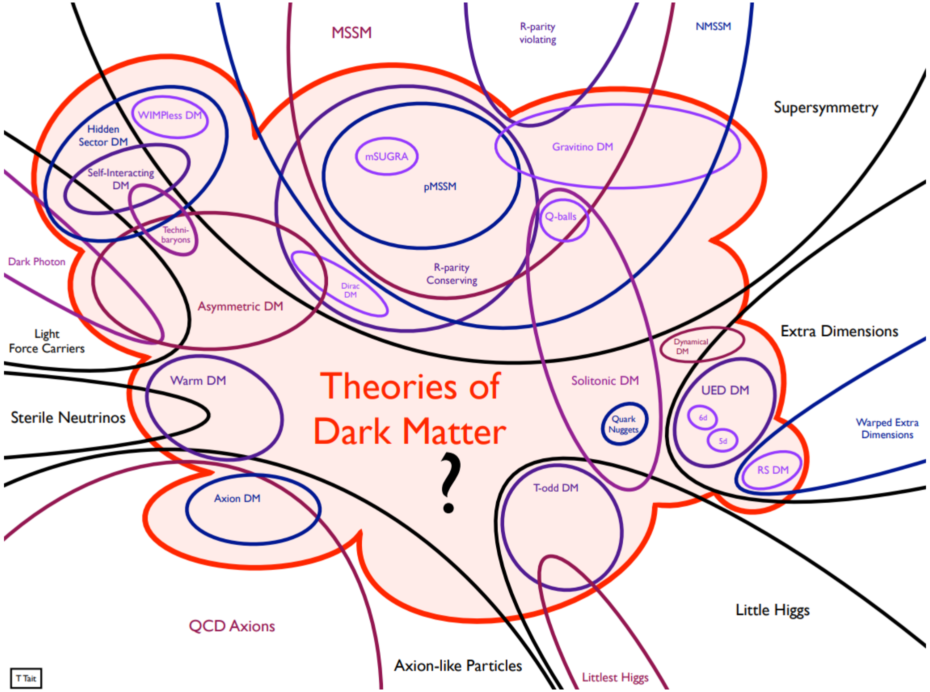


Approx ~1000 total LHC searches

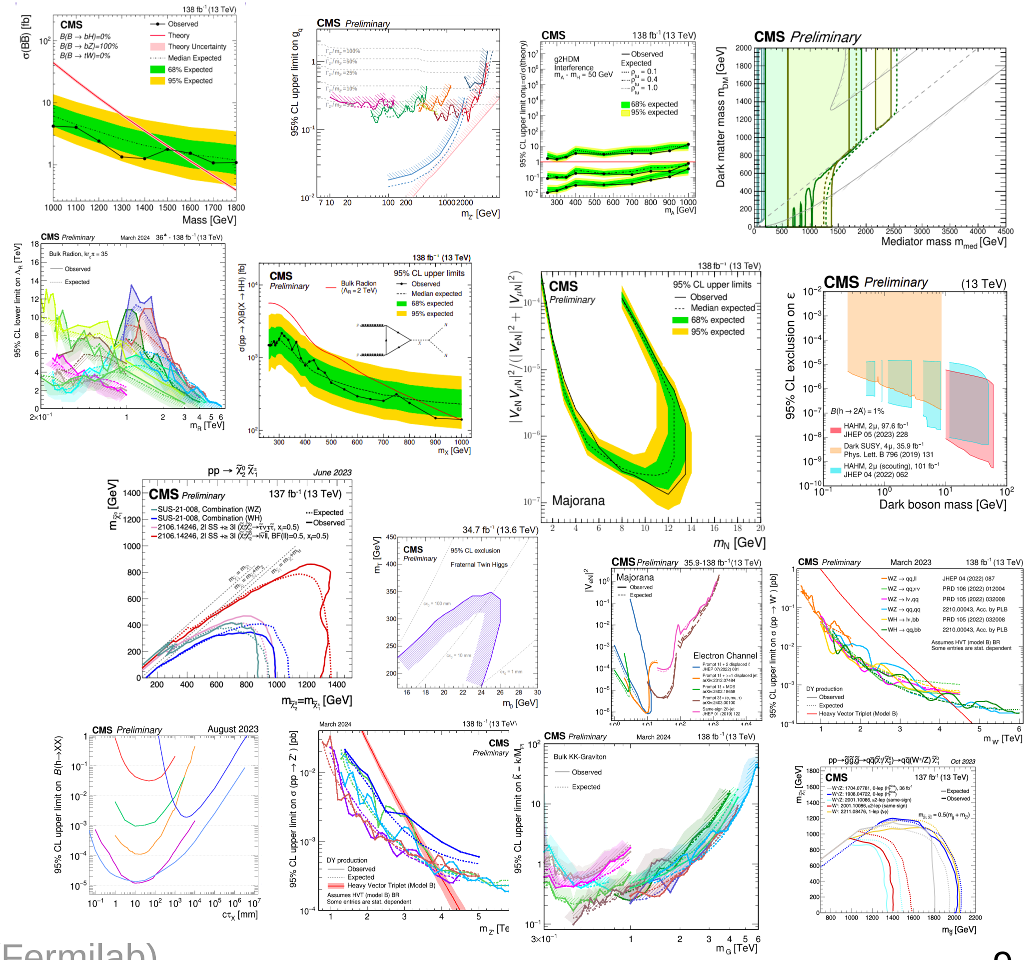
DM graphic from T. Tait

LHC Answers?

But...



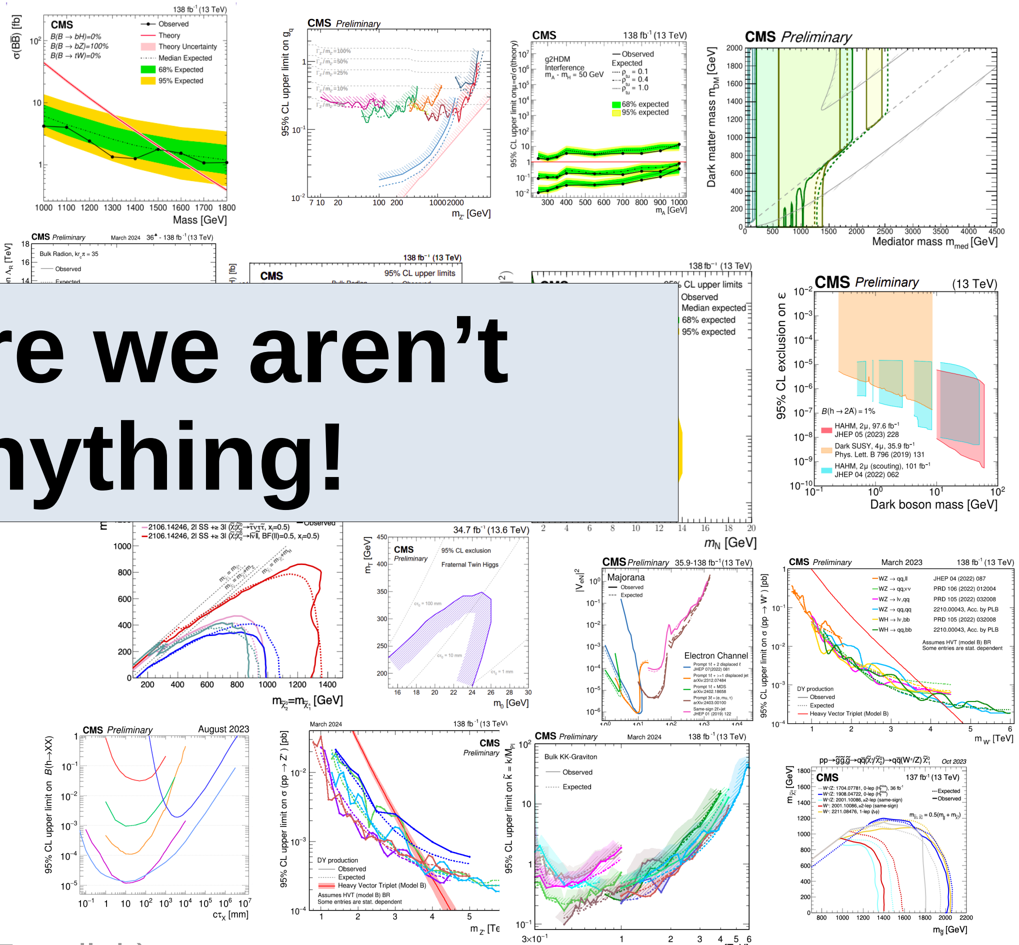
>



DM graphic from T. Tait

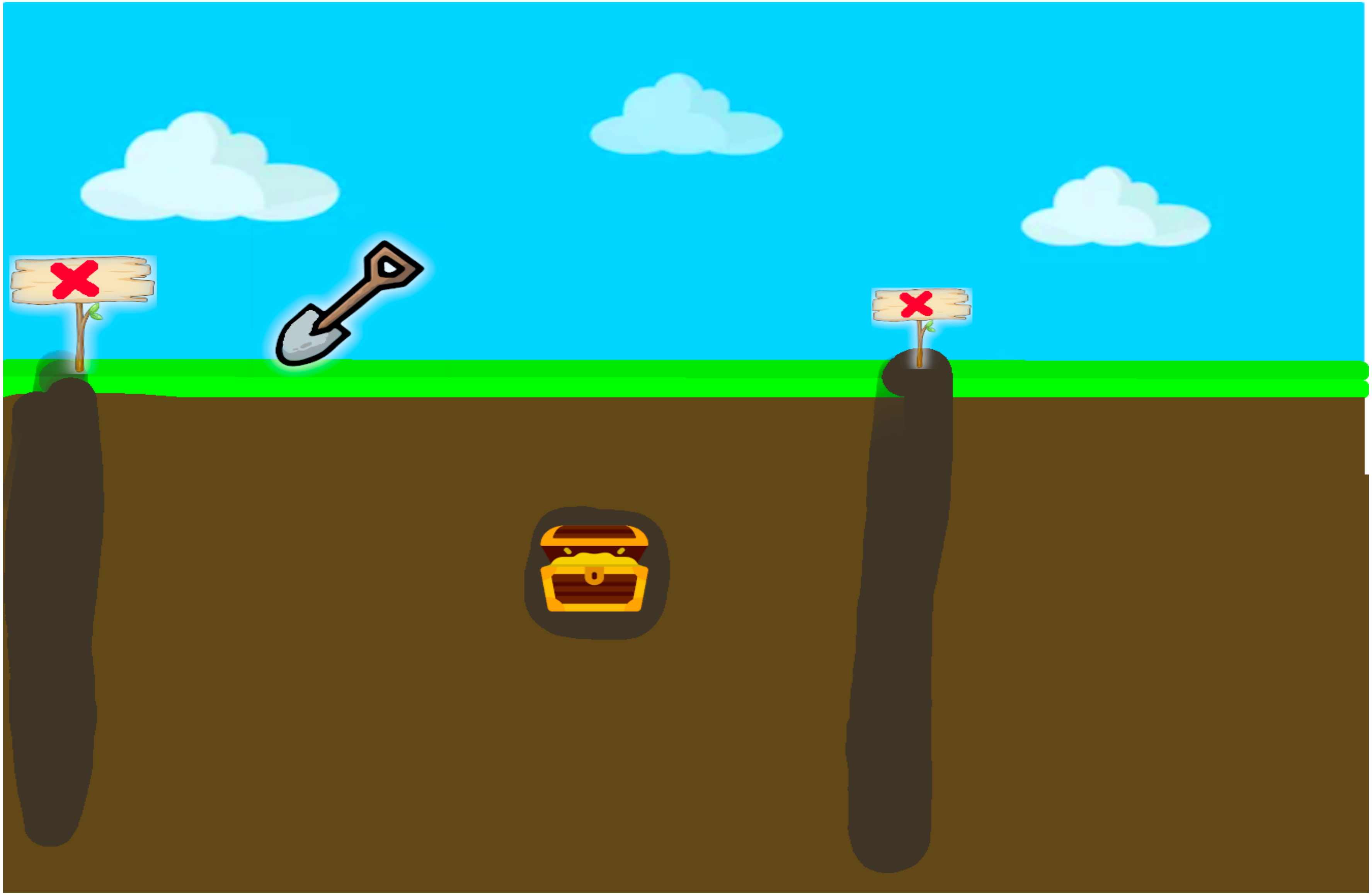
LHC Answers?

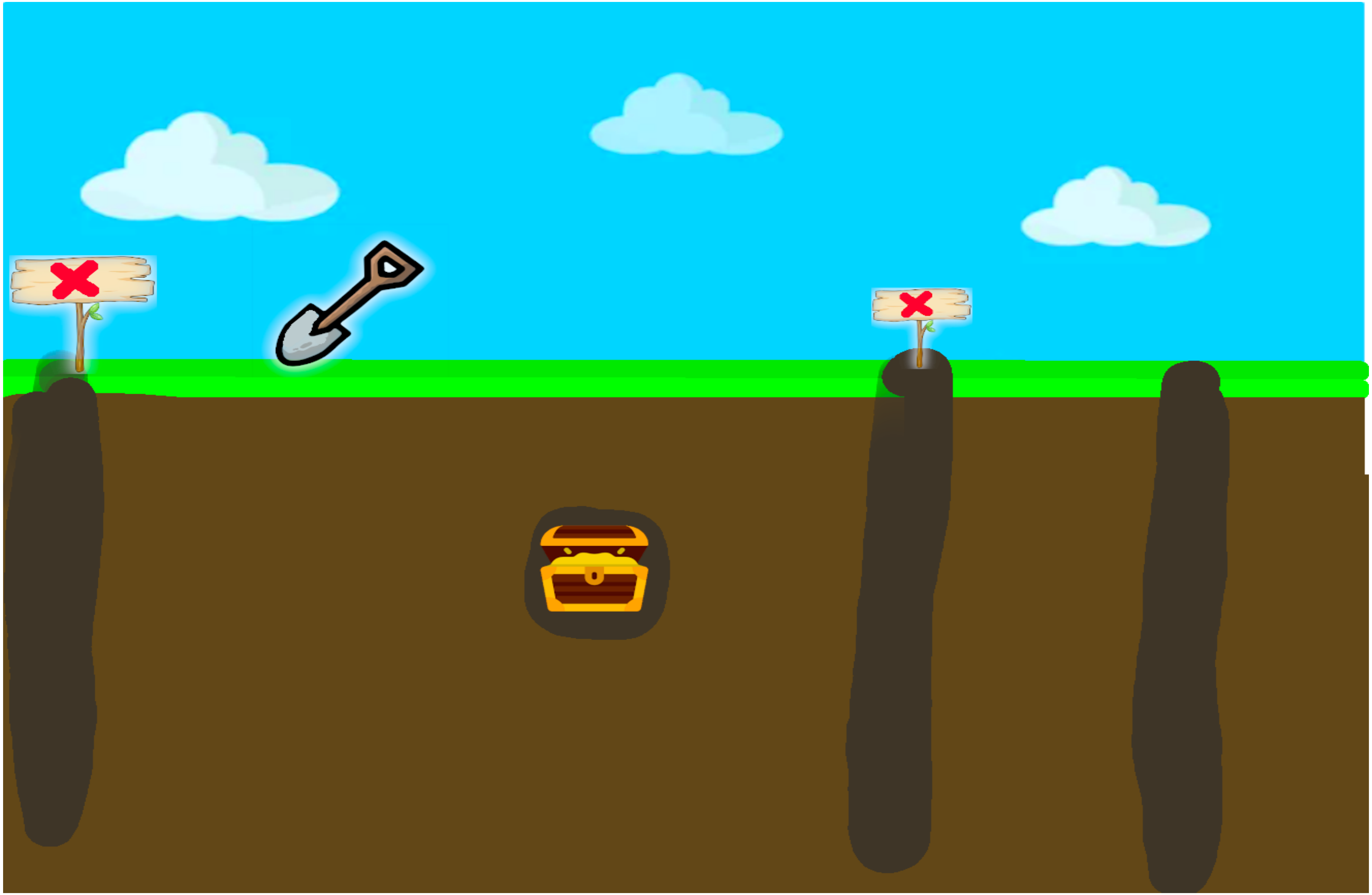
But...

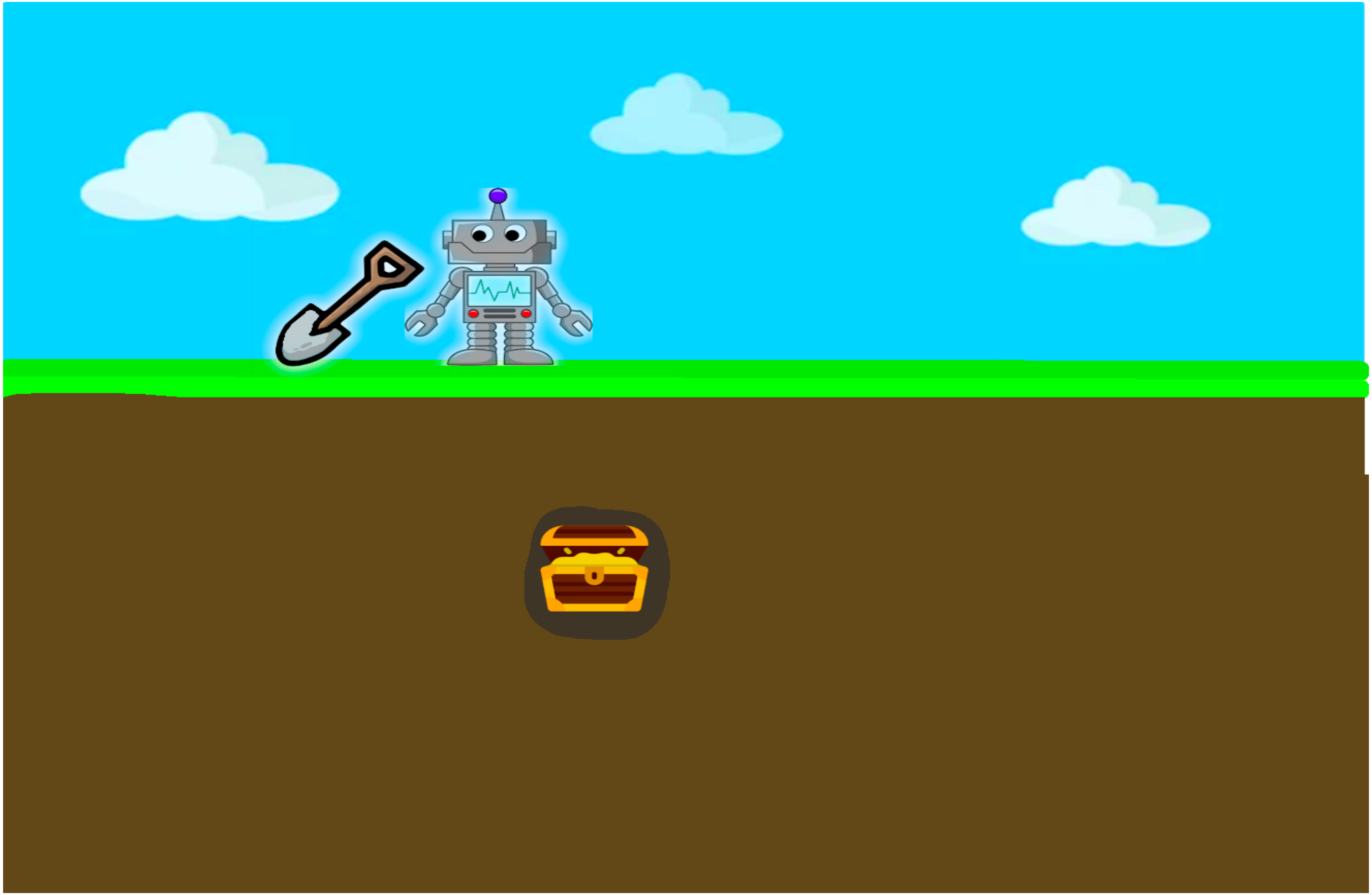


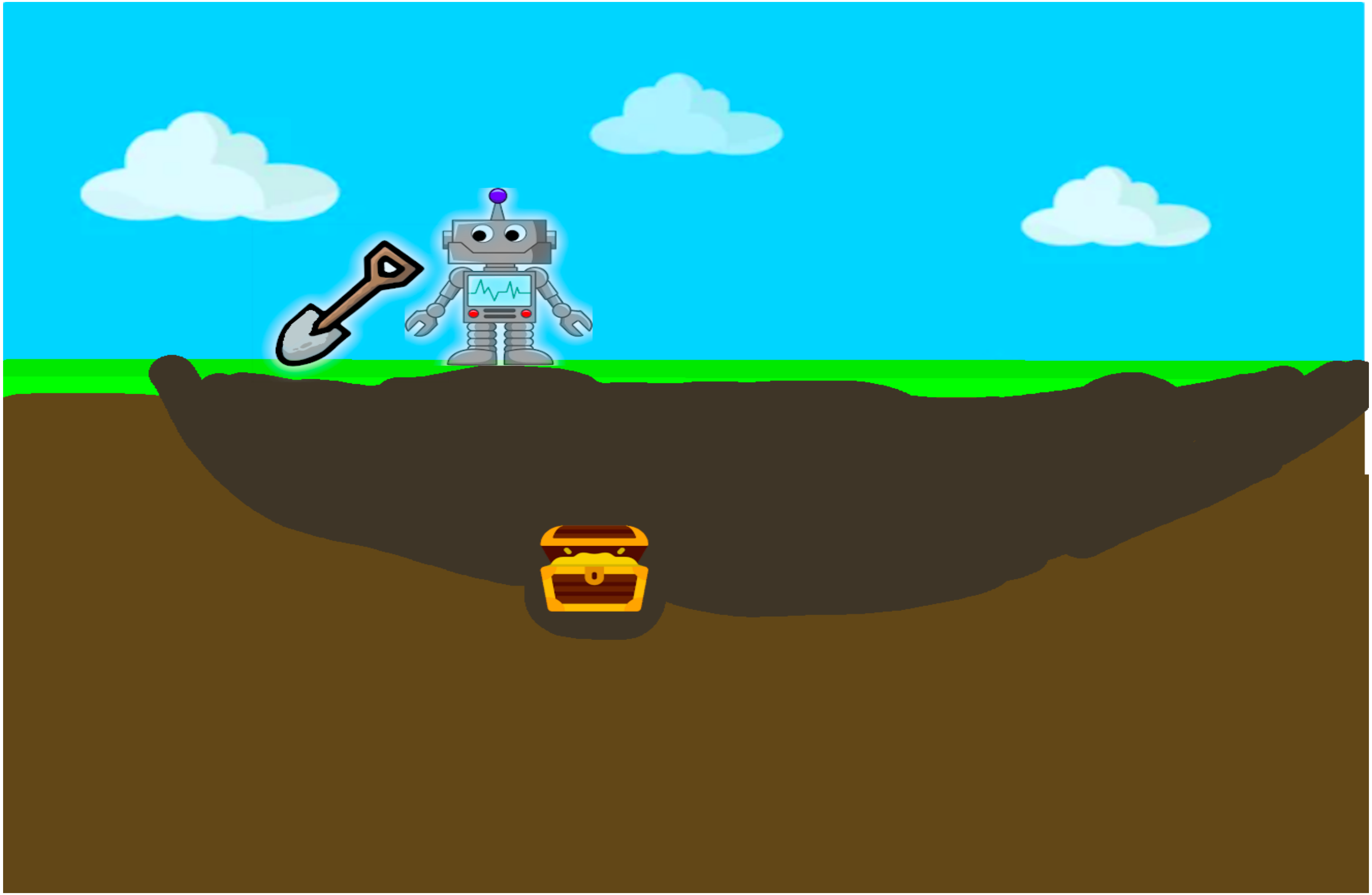
Lets make sure we aren’t missing anything!

DM graphic from T. Tait









History

VOLUME 86, NUMBER 17 PHYSICAL REVIEW LETTERS 23 APRIL 2001

Quasi-Model-Independent Search for New High p_T Physics at D0

We apply a quasi-model-independent strategy ("Sleuth") to search for new high p_T physics in $\approx$100 pb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV collected by the D0 experiment during 1992–1996 at the Fermilab Tevatron. We systematically analyze many exclusive final states and demonstrate sensitivity to a variety of models predicting new phenomena at the electroweak scale. No evidence of new high p_T physics is observed.

PHYSICAL REVIEW D **78,** 012002 (2008)

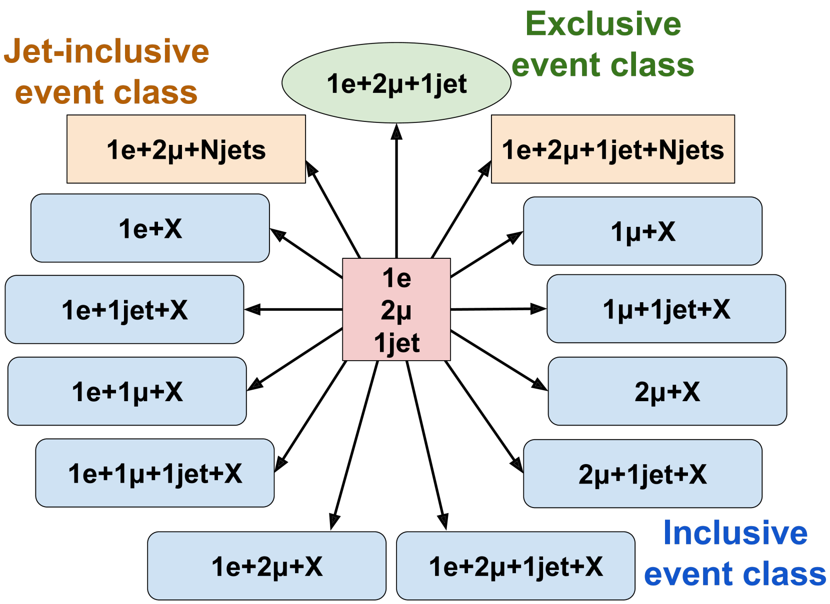
Model-independent and quasi-model-independent search for new physics at CDF

"Vista"

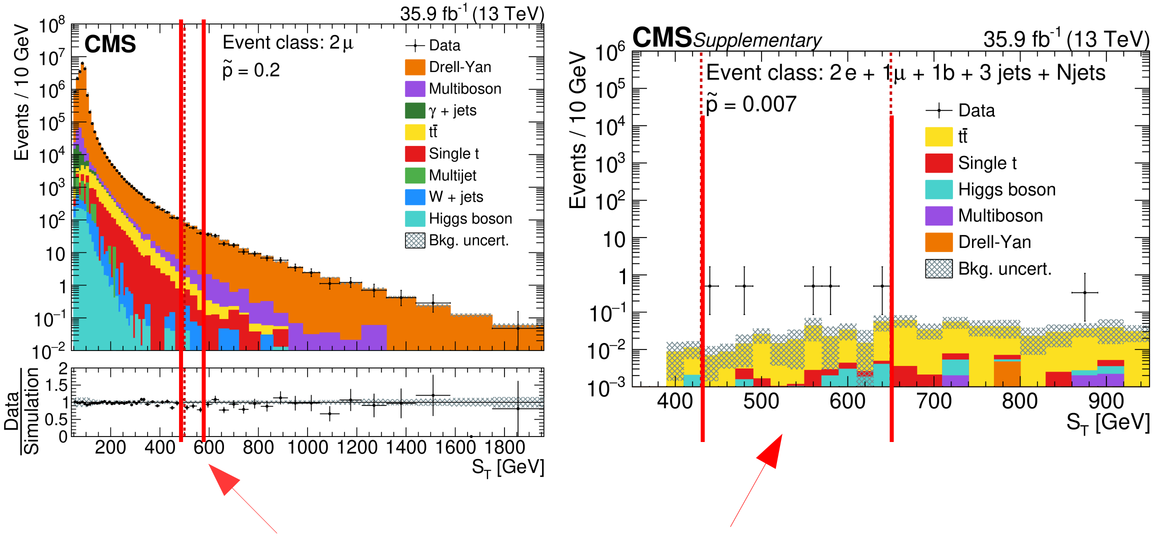
Classic Strategy

Using CMS MUSiC Search as an example

Categorize



Data-MC Comparison



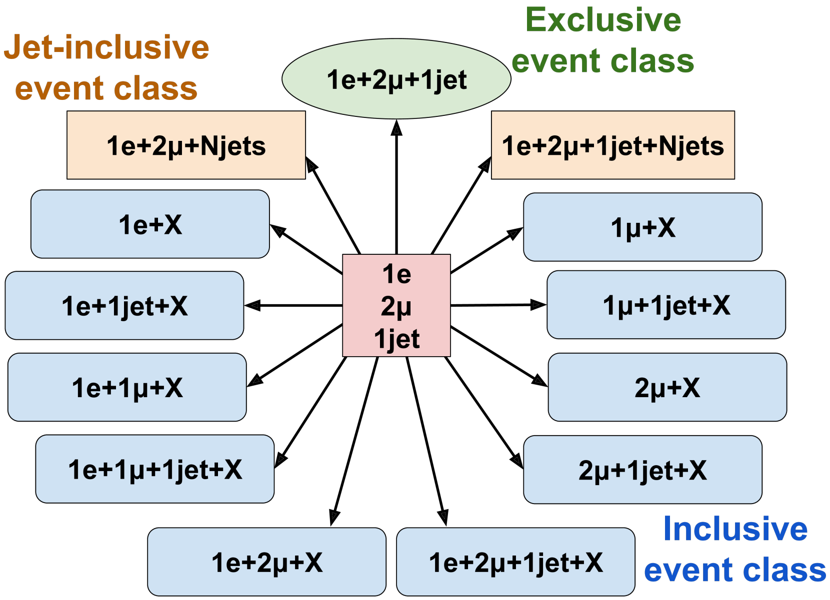
Find Largest Local Deviations

~1.5k event classes

Classic Strategy

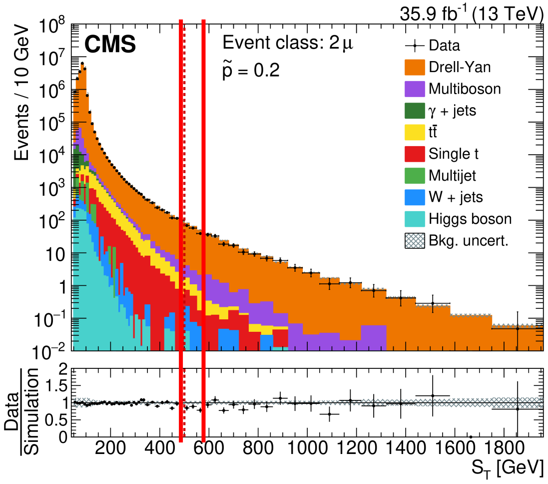
Using CMS MUSiC Search as an example

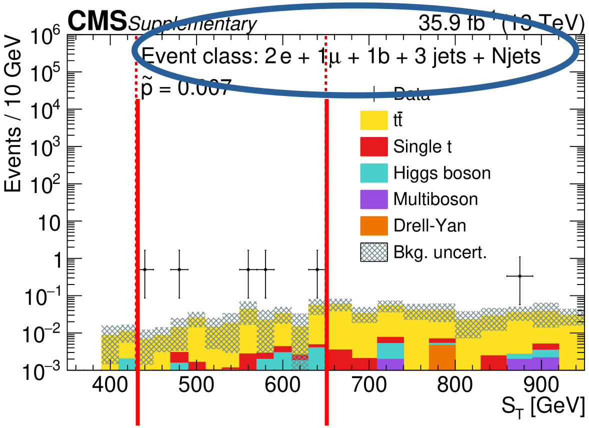
Categorize

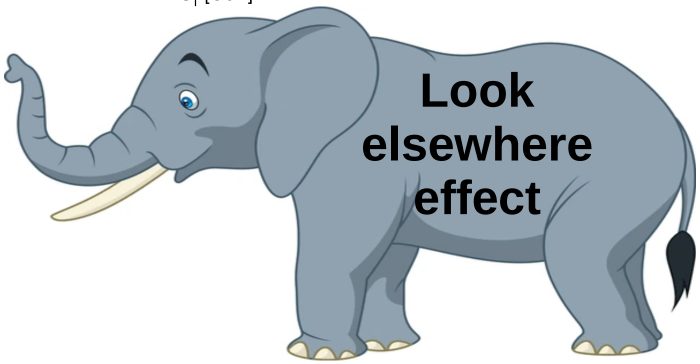


~1.5k event classes

Data-MC Comparison







Modern ‘Anomaly Detection’

arXiv: 2101.08320

- Focus on a single topology at a time
- Entirely **data-driven**
- Novel ML methods to reduce bkg

Modern ‘Anomaly Detection’

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

- Focus on a single topology at a time

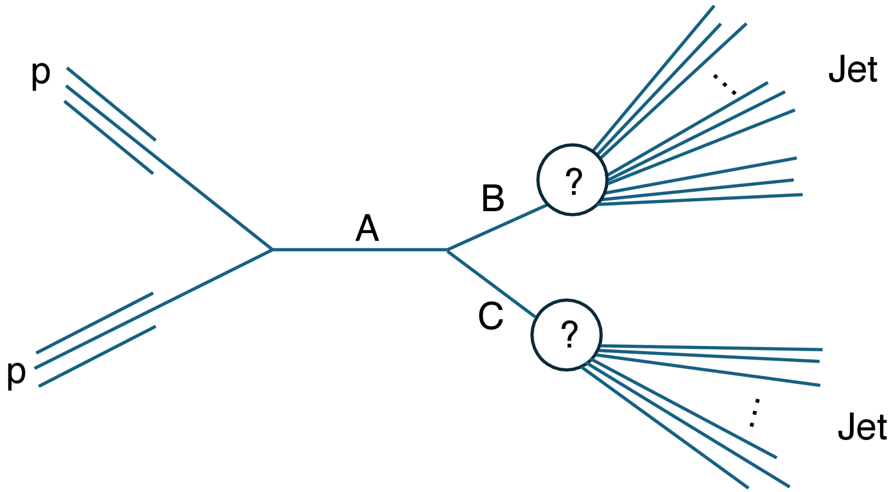
The Philosophy

“No free lunch” → Drop full model independence

But “**discounts for buying in bulk”!**
→ Cover a large model space in an efficient way

arXiv: 2101.08320

Dijet Resonance Anomaly Search

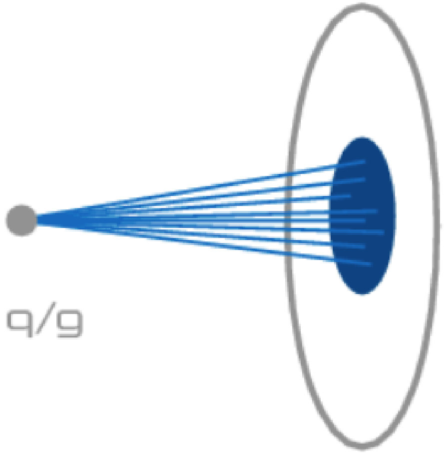


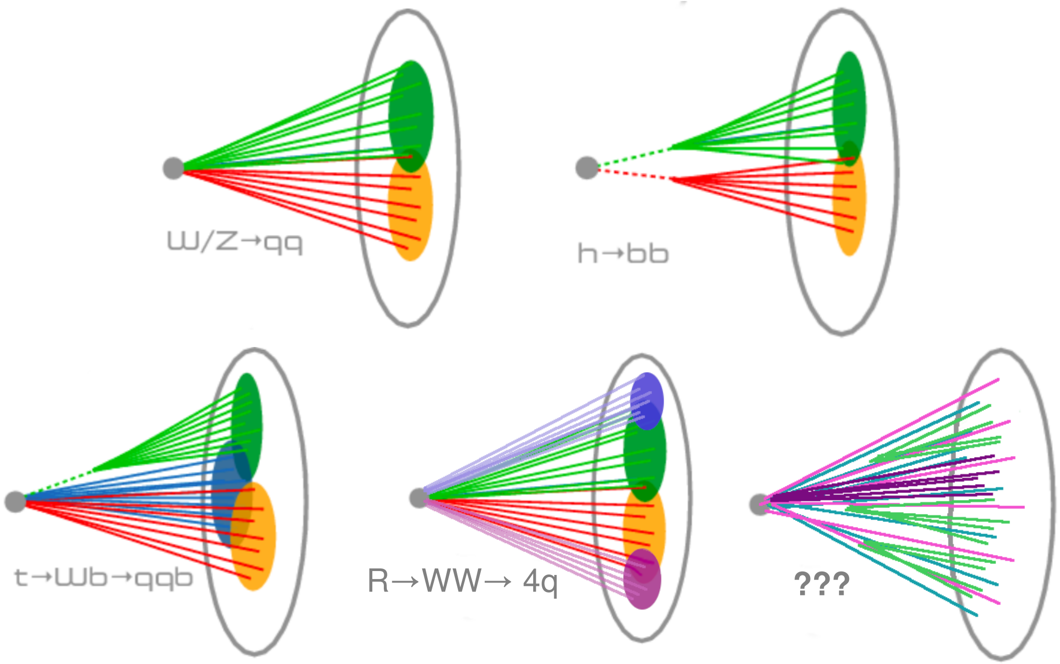
All material from
CMS-EXO-22-026

- **A → BC** topology
  - Heavy resonance (A) → daughters B and C
  - B & C are boosted → contained in a large radius jet
- Look for B & C jets with ‘anomalous’ substructure

Graphics source

Jet Substructure





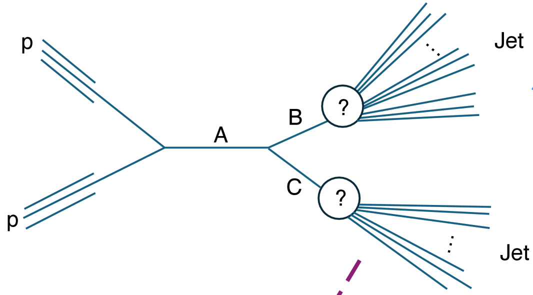
Typical jet

- One central axis (prong)
- From primary vertex
- ...

Anomalous jets

- Multiple prongs
- Displaced vertices
- ???

Signal Models



Picked a set of unexplored models to **evaluate performance**

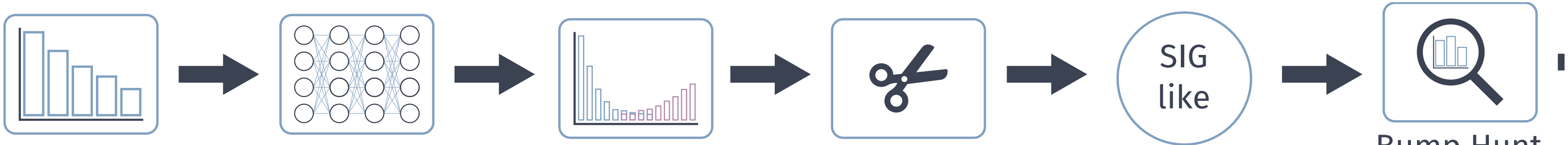
B Jet substructure

C Jet substructure

	1 prong	2 prong	3 prong	4 prong	5 prong	6 prong
1 prong		**$Q^* \rightarrow qW$** m_{Q^*} = [2,3,5] TeV m_W = [25,80,170,400] GeV				
2 prong		**$X \rightarrow YY'$** m_X = [2,3,5] TeV m_Y = [25,80,170,400] GeV $m_{Y'}$ = [25,80,170,400] GeV		**$W_{KK} \rightarrow RW \rightarrow WWW$** m_{WKK} = [2,3,5] TeV m_R = [170,400] GeV		
3 prong			**$W' \rightarrow tB'$** $m_{W'}$ = [2,3,5] TeV $m_{B'}$ = [25,80,170,400] GeV			
4 prong				**$X \rightarrow YH \rightarrow WWWW$** m_X = [2,3,5] TeV m_Y = [170,400] GeV m_H = [170,400] GeV		
5 prong					**$Z' \rightarrow T'T' \rightarrow tZtZ$** $m_{Z'}$ = [2,3,5] TeV $m_{T'}$ = [400] GeV	
6 prong						**$Y \rightarrow HH \rightarrow tttt$** m_Y = [2,3,5] TeV m_H = [400] GeV

Expect sensitivity to many additional kinds of signals!

Analysis Overview

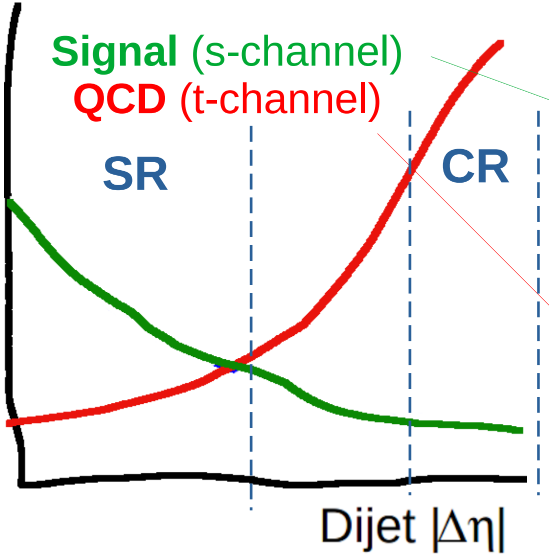


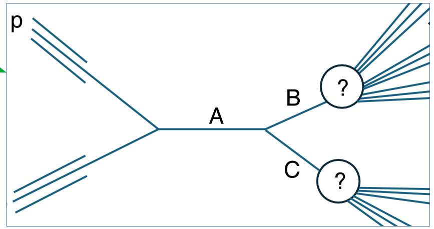
Preselected Data → Trained Model → Anomaly Metric → Cut → SIG like → Bump Hunt

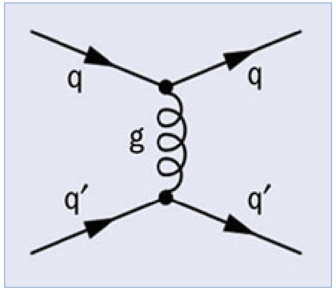
2 large radius (AK8) jets

Signal Region
$|\Delta\eta| < 1.3$

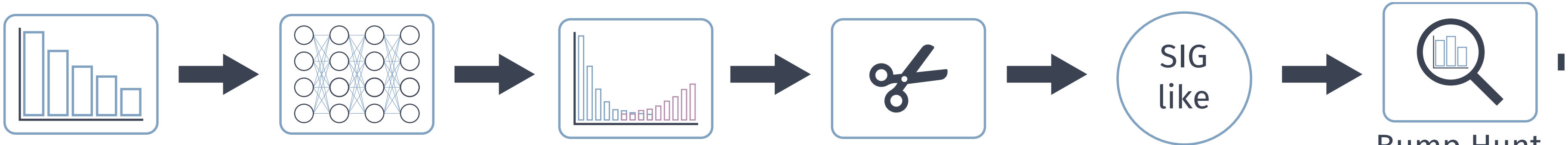
Control region
$2.0 < |\Delta\eta| < 2.5$ *







Analysis Overview



Preselected Data → Trained Model → Anomaly Metric → Cut → SIG like → Bump Hunt

2 large radius (AK8) jets

Signal Region
$|\Delta\eta| < 1.3$

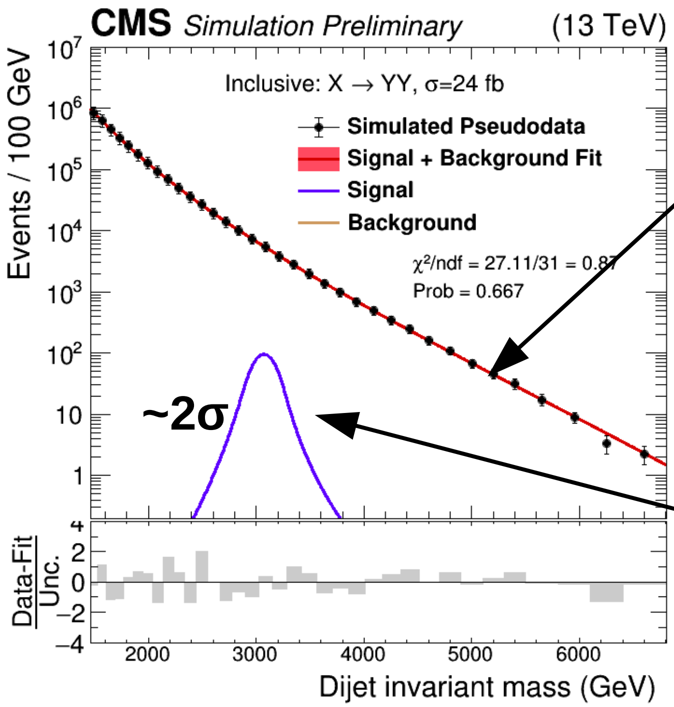
Control region
$2.0 < |\Delta\eta| < 2.5$ *

Bkg: Standard Dijet Parameterization
Signal: Double Crystal Ball

Bumps?

Limits on signal models

The Bump Hunt

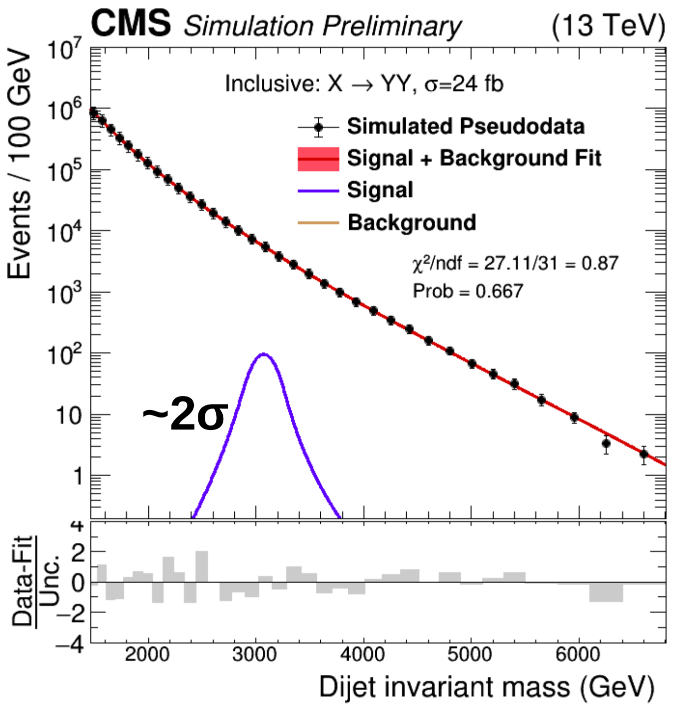


Background fit with standard analytic functions

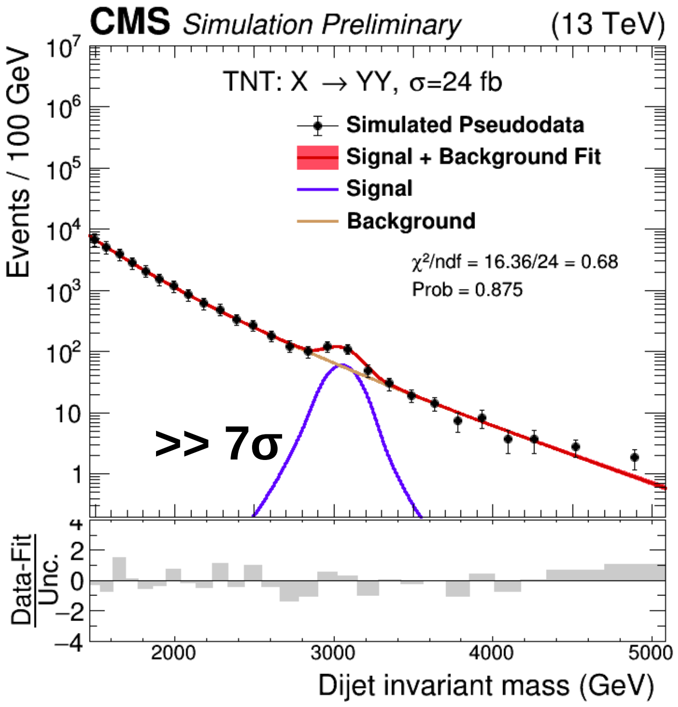
Double Crystal Ball signal shape

Without any substructure cuts →
Signal swamped by QCD background…

The Bump Hunt



Cut on anomaly score



Anomaly detection finds hidden resonance!

Analysis Overview

2 large radius (AK8) jets

The fun part!

5 different approaches

Signal Region
$|\Delta\eta| < 1.3$

Sideband region
$2.0 < |\Delta\eta| < 2.5$ *

Bkg: Standard Dijet Parameterization
Signal: Double Crystal Ball

Bumps?

Limits on signal models (tricky)

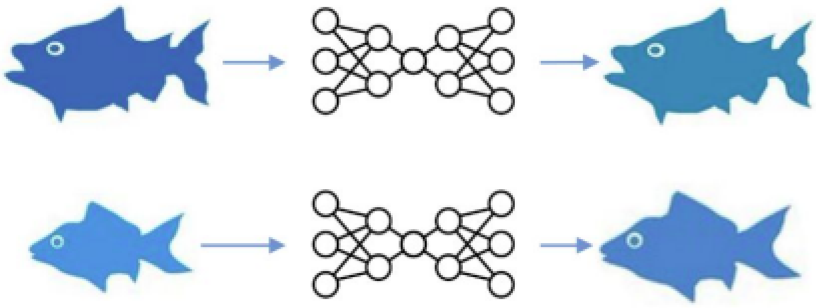
How to identify anomalous jets?

Learn QCD jets → look for outliers

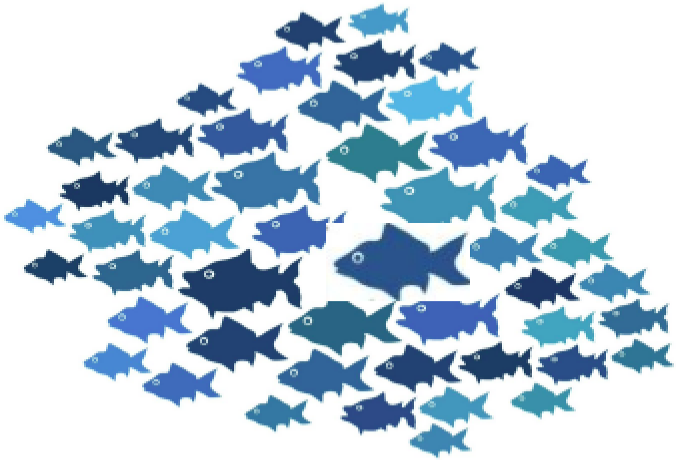
Increasing Model Dependence

Looking for Outliers

Train 'Autoencoder'



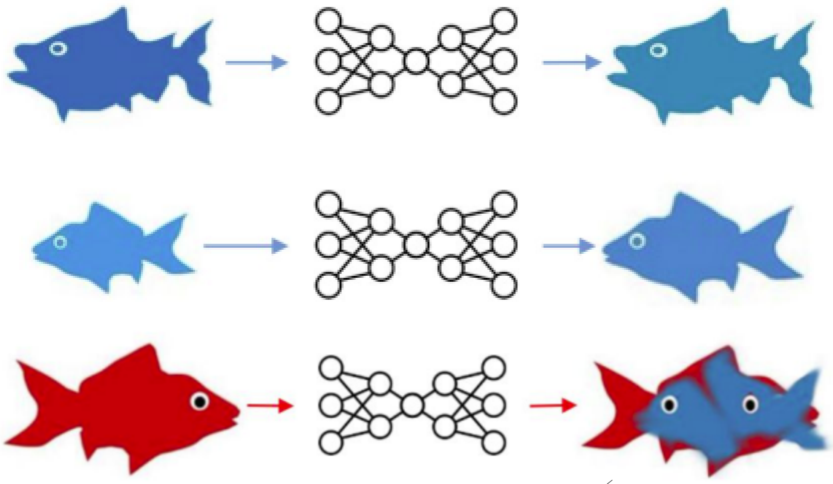
Training Sample from data sideband



Illustrations: J Gonski, A Kahn

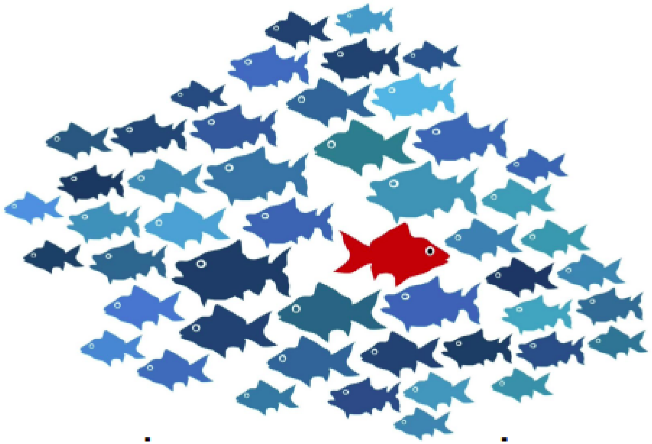
Looking for Outliers

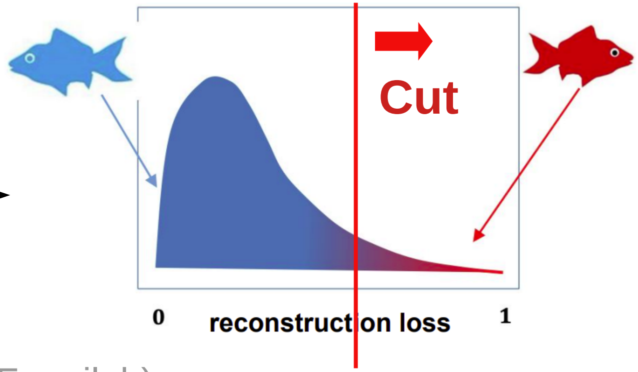
Apply Autoencoder



Take difference

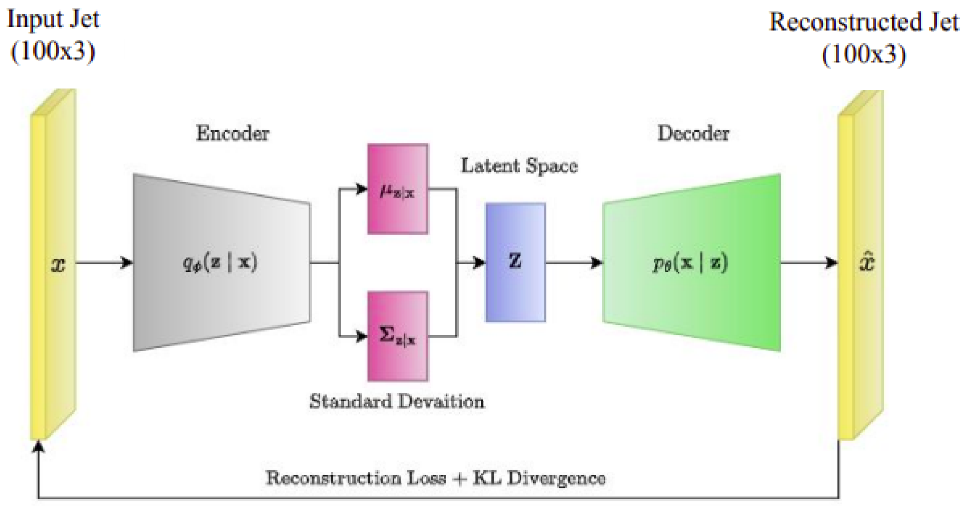
Data from signal region





Illustrations: J Gonski, A Kahn

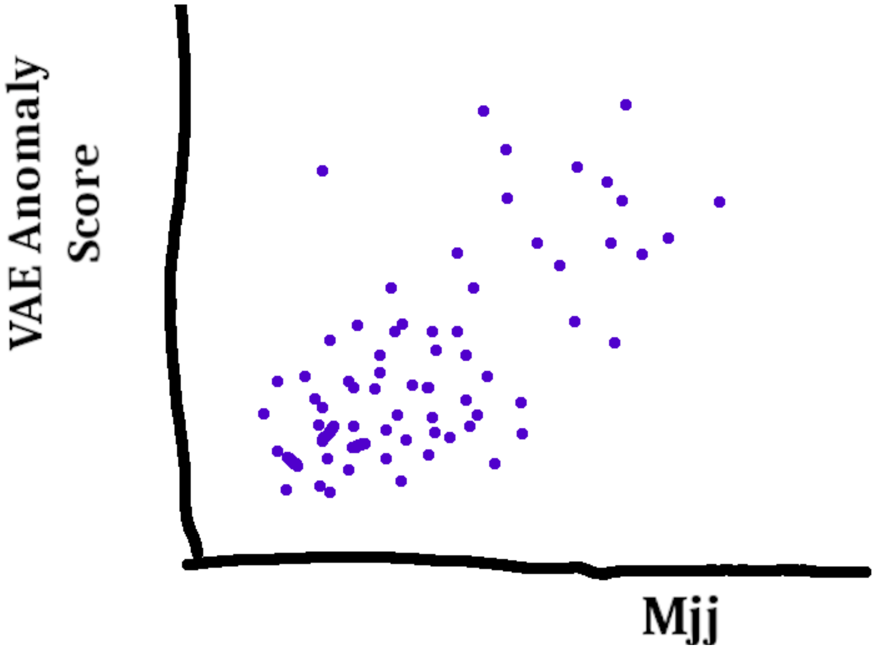
Variational Autoencoder (VAE)



Latent space forced to be Gaussian thru additional term in loss

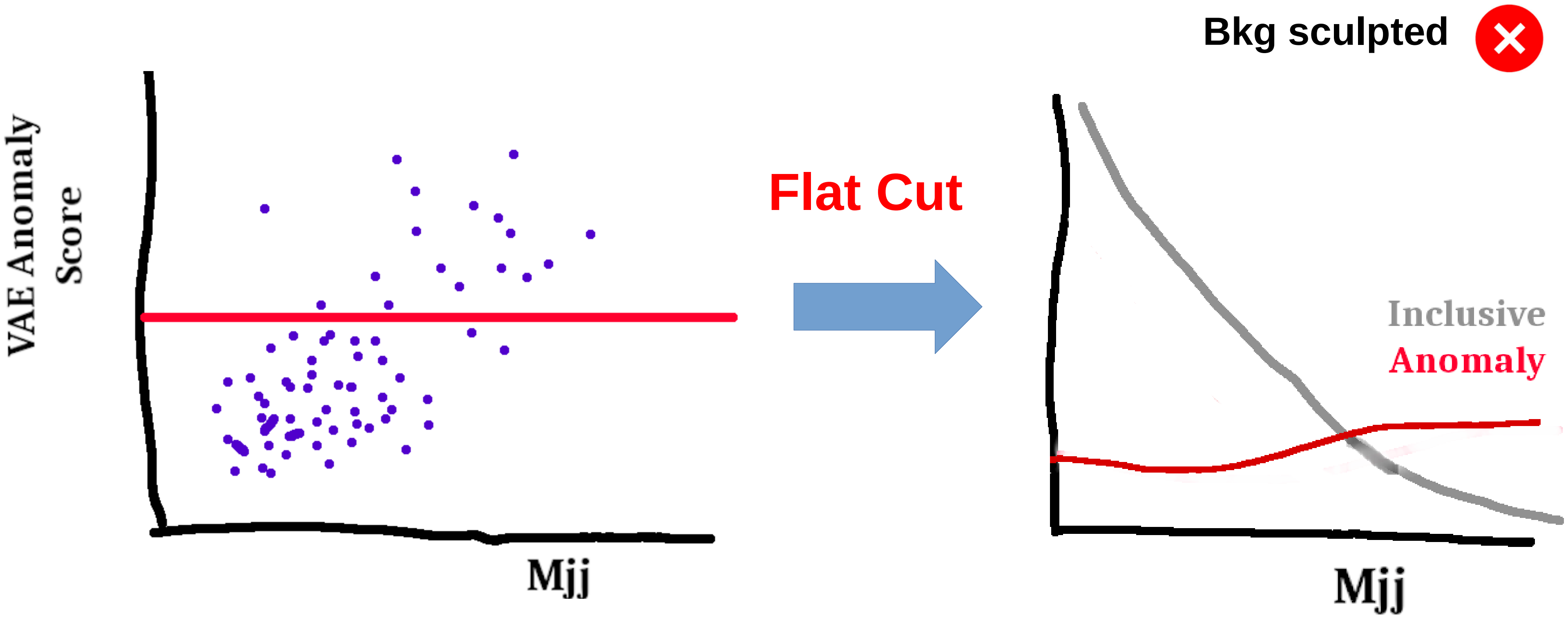
- Jet represented by up to 100 highest p_T constituents (p_x, p_y, p_z)
- 100x3 matrix compressed to latent space of size 12
- Trained on jets from $|\Delta\eta|$ sideband
  - Sampled to match SR kin.

Decorrelate with Mjj



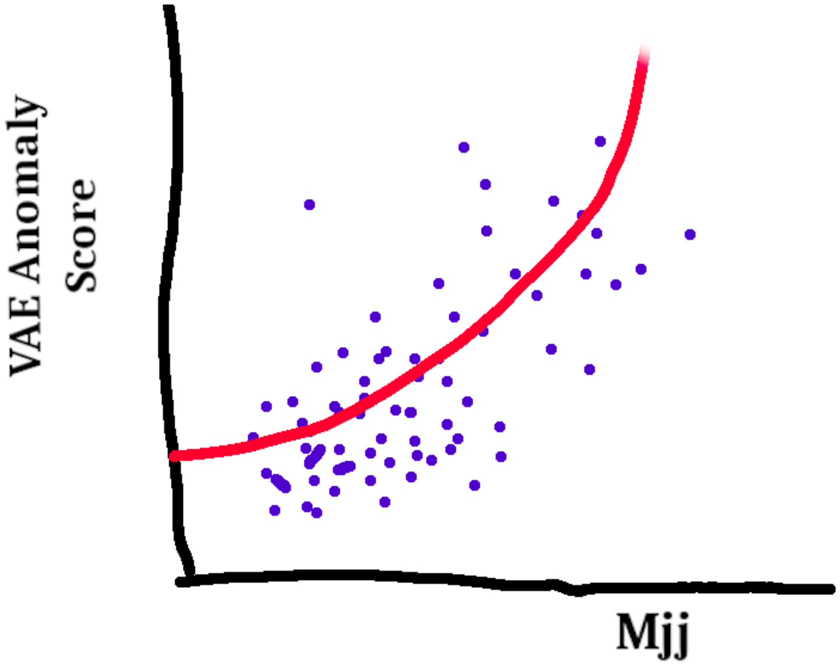
High Mjj events are rarer → higher anomaly score

Decorrelate with Mjj



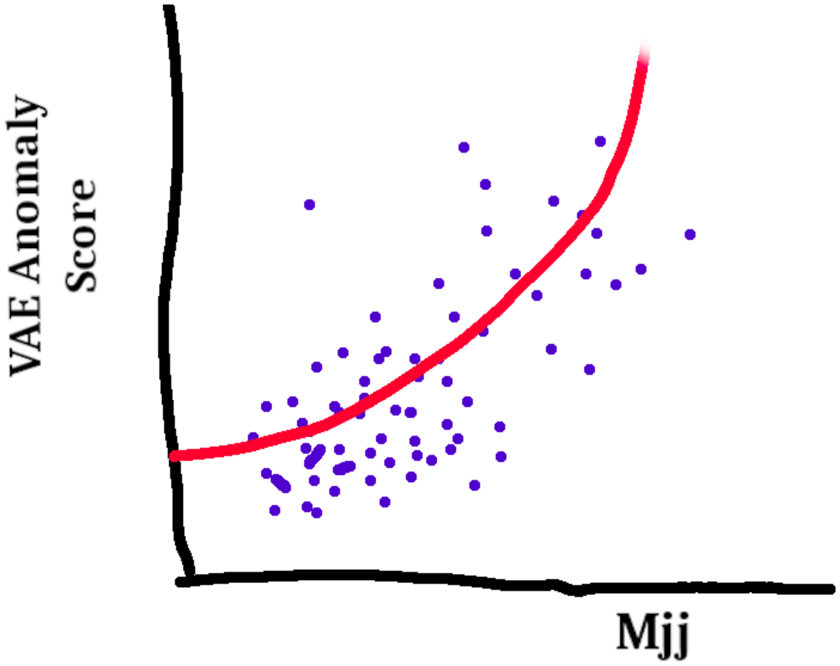
Decorrelate with Mjj

'Quantile Regression' (QR)

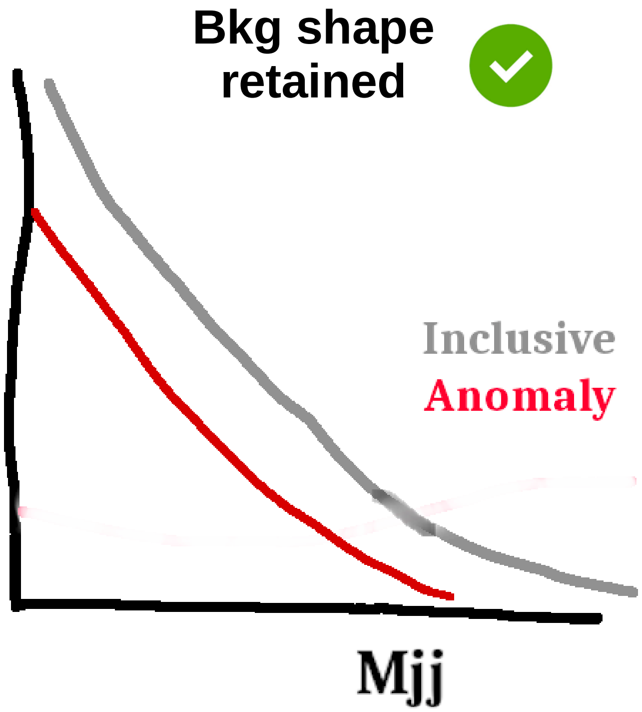


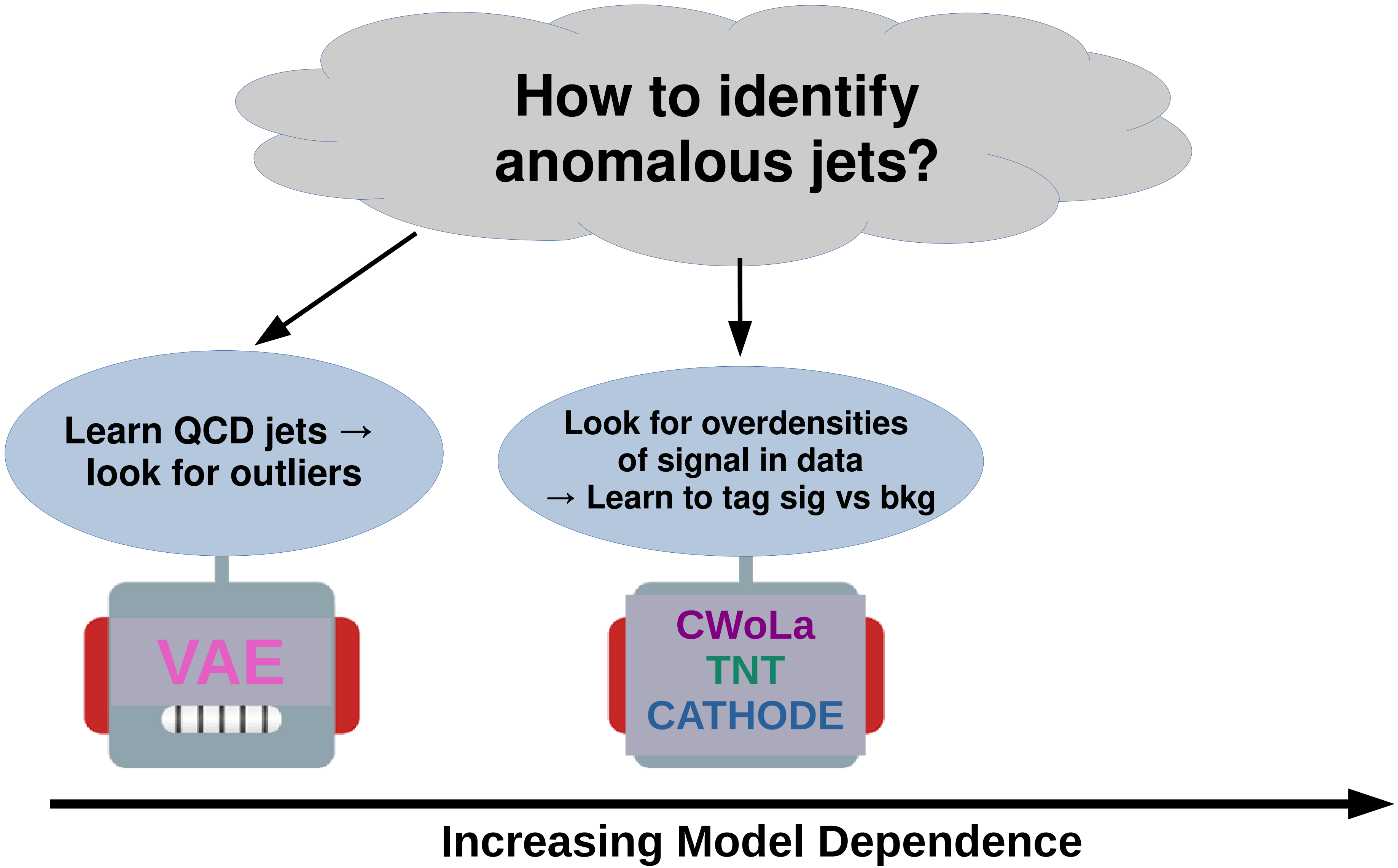
Adjust cut to have a constant efficiency vs Mjj

Decorrelate with Mjj



Cut with Flat Eff.

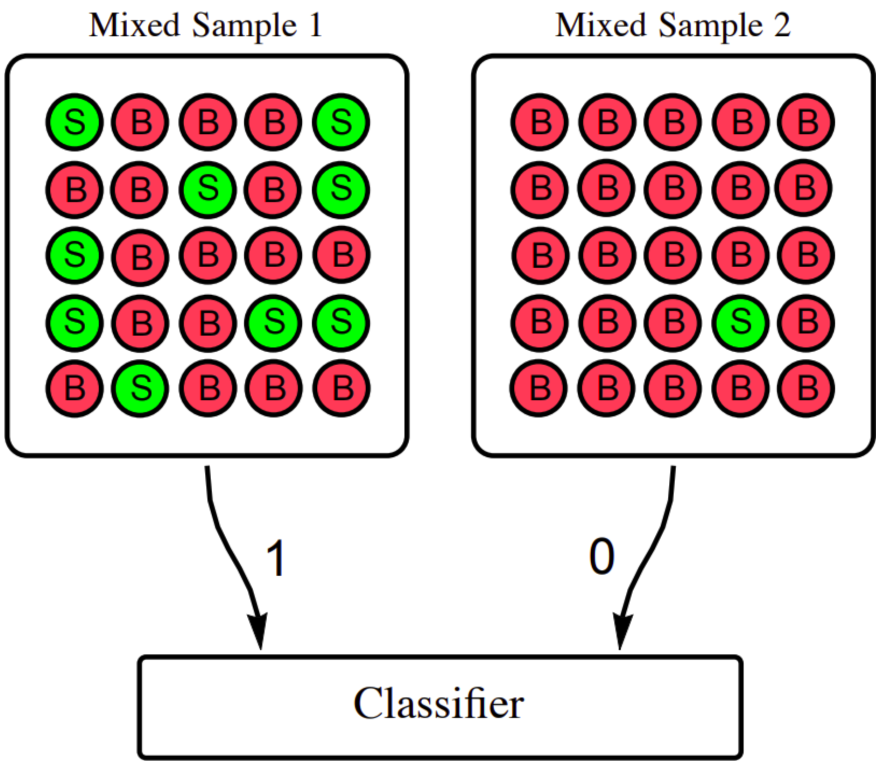



How to identify anomalous jets?
Learn QCD jets → look for outliers
Look for overdensities of signal in data → Learn to tag sig vs bkg
VAE
CWoLa
TNT
CATHODE
Increasing Model Dependence

Weak Supervision

Aka ‘Classification Without Labels’ (CWoLa)

Train on two mixed samples



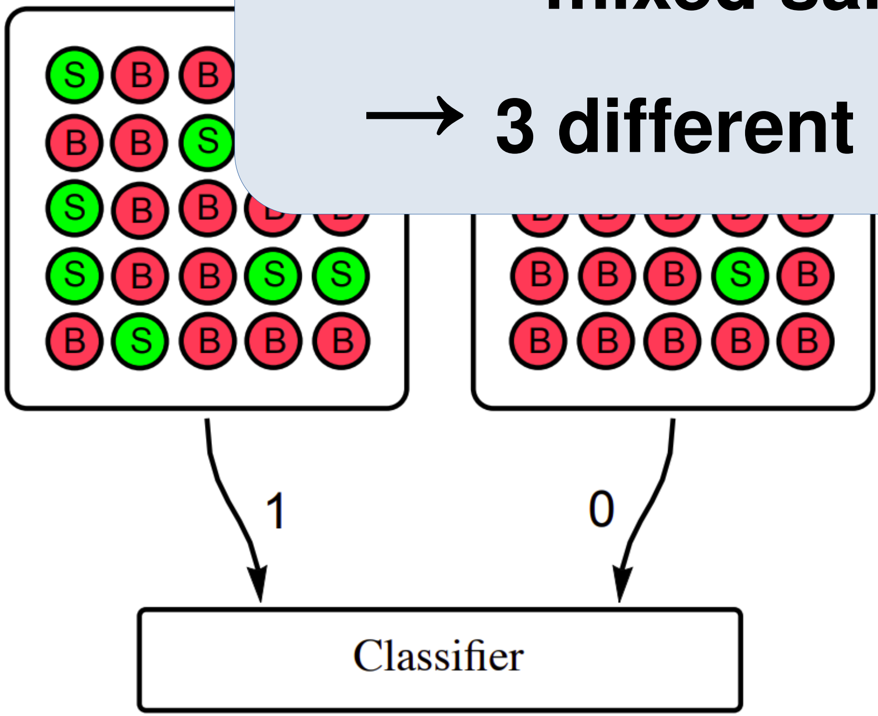
- Train a classifier between **signal-rich** and **background-rich** mixed samples
  - → Learns to tag signal vs. bkg
- Performance changes with amount of signal in training data
  - – No signal → learn random noise
  - – Lots of signal → approach ‘supervised’ (optimal) classifier

Weak Supervision

Aka ‘Classification Without Labels’ (CWoLa)

Trai

Mixed Sa



How can we construct these mixed samples in data?

→ 3 different methods employed

veen

kground-

al vs. bkg

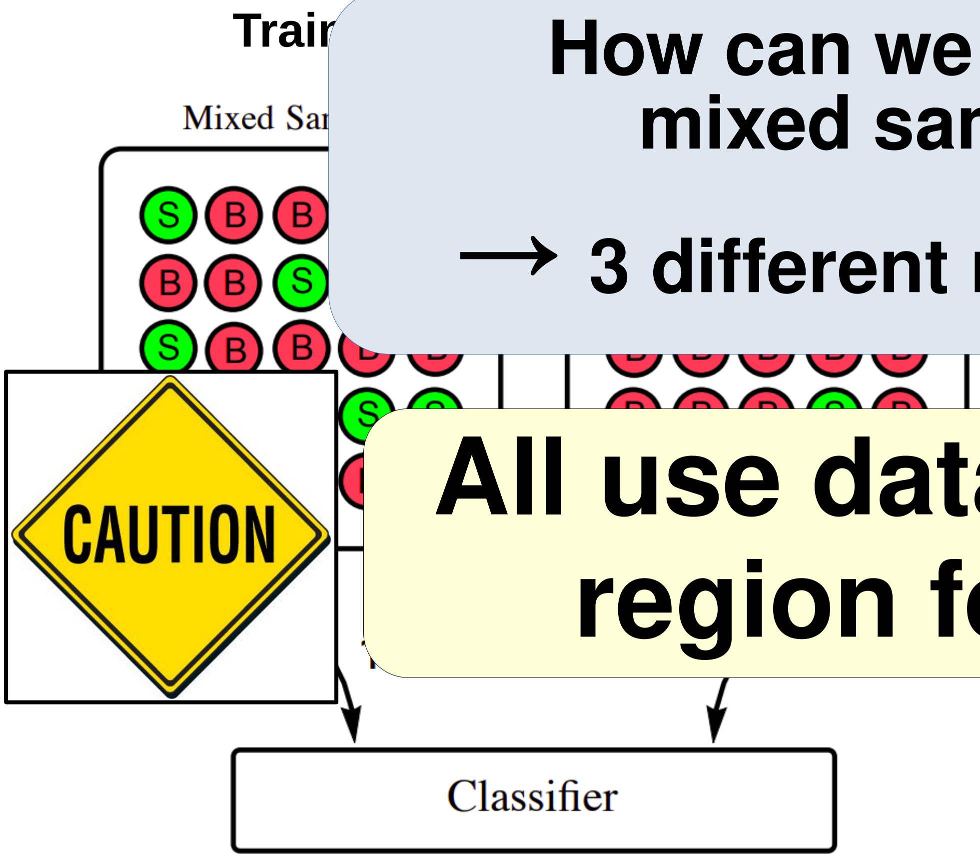
es with amount of signal in training data

- No signal → learn random noise
- Lots of signal → approach ‘supervised’ (optimal) classifier

[Metodiev, Nachman, Thaler, 1708.02949]

Weak Supervision

Aka 'Classification Without Labels' (CWoLa)



How can we construct these mixed samples in data?

→ 3 different methods employed

All use data from signal region for training!

veen
kground-

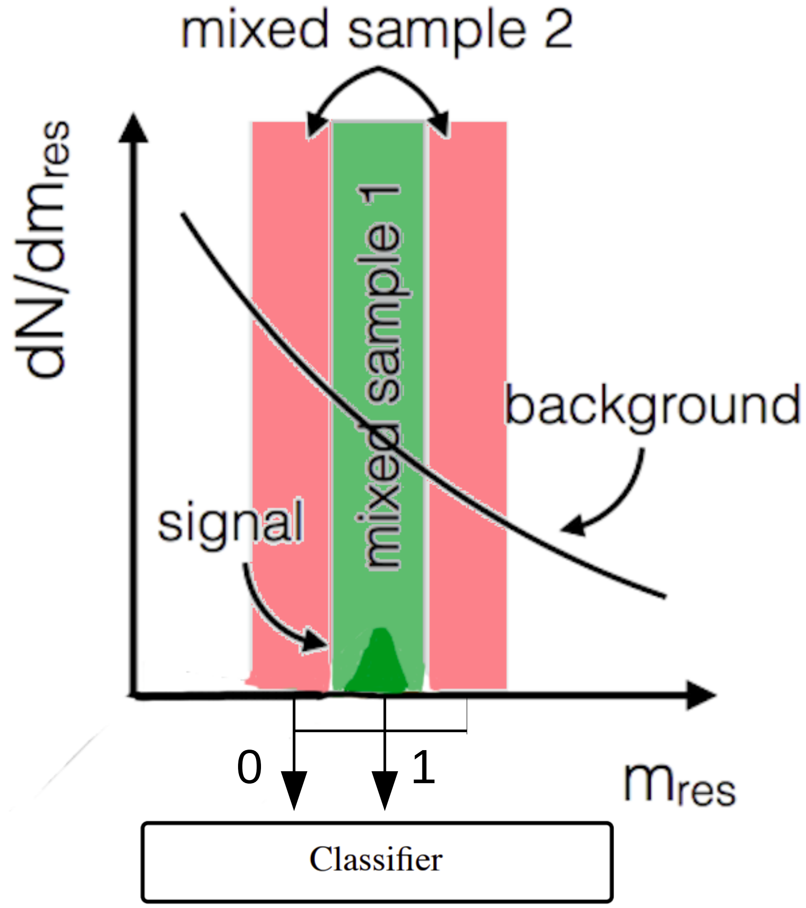
al vs. bkg

es with
amount of signal in training

andom

– Lots of signal → approach 'supervised' (optimal) classifier

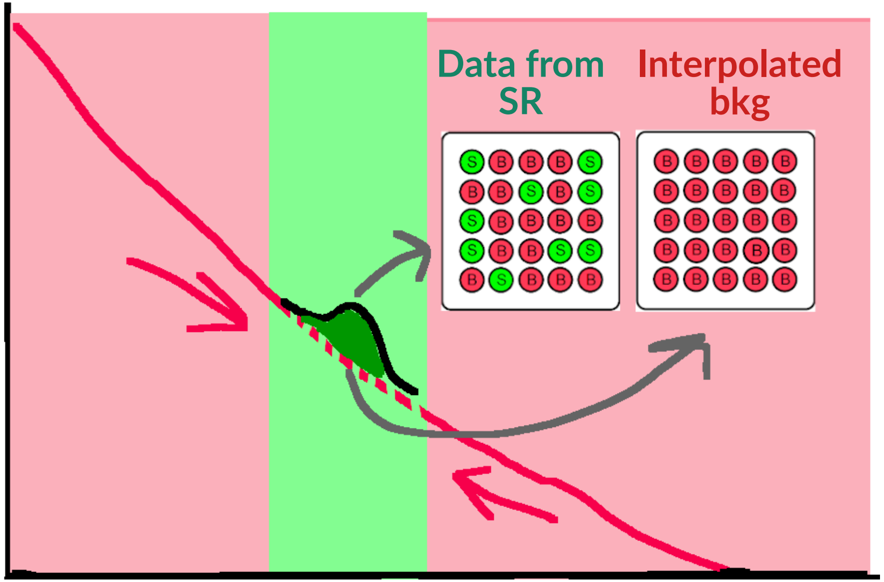
CWoLa Hunting



- Assume signal is a **narrow** resonance
- **Guess** a mass window where it lives
  - Train signal window vs. narrow sidebands using weak supervision
- **Repeat procedure**, scanning over different mass windows
  - (2x6 windows used)
- Need to be careful about correlations with Mjj

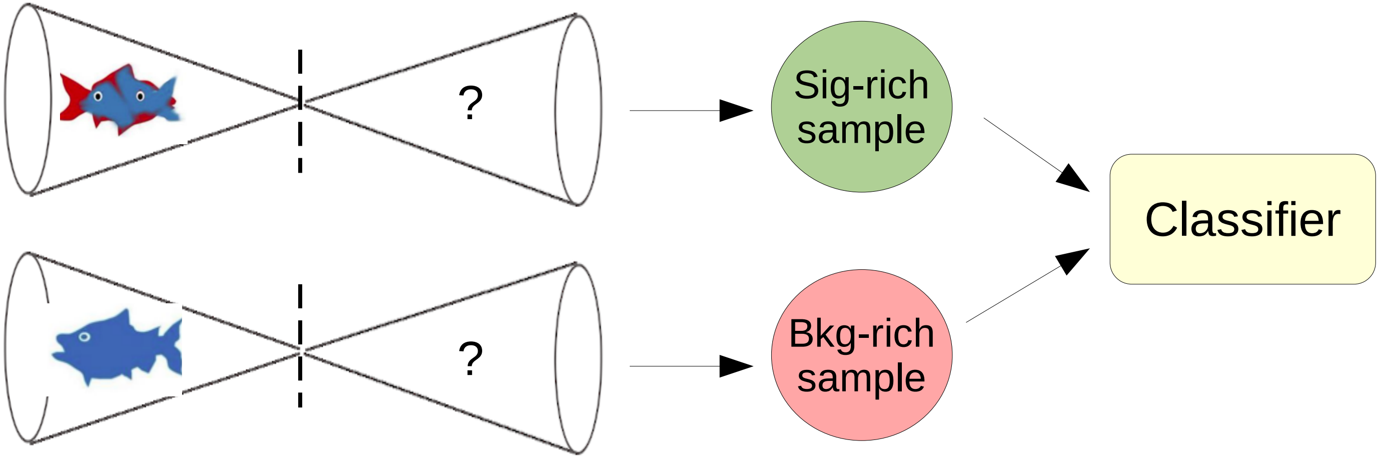
[Collins, Howe, Nachman 1902.02634]

CATHODE
Interpolates bkg events into SR to construct sample



[Hallin et al 2109.00546]

Tag N' Train
purifies samples by first tagging with AE



[**OA** & Suarez 2002.12376]

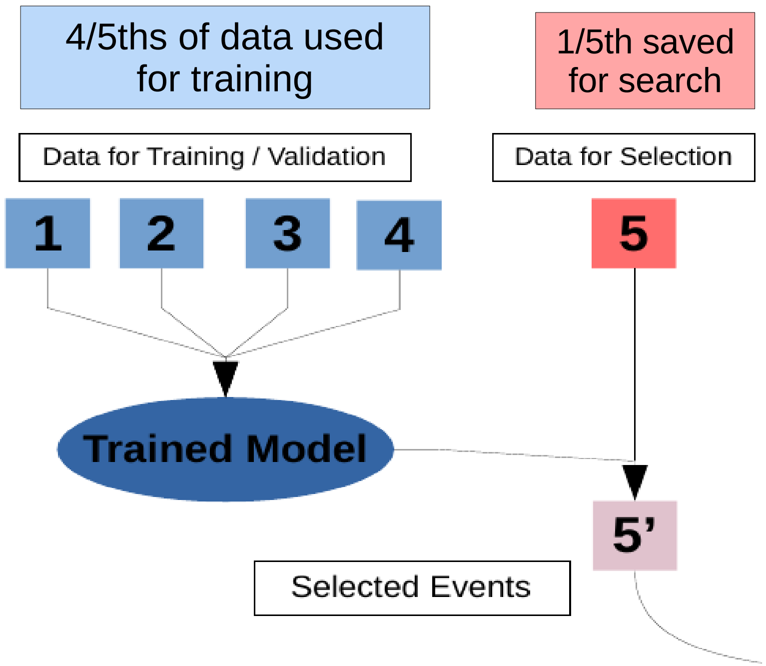
Cross Validation

Weakly supervised methods

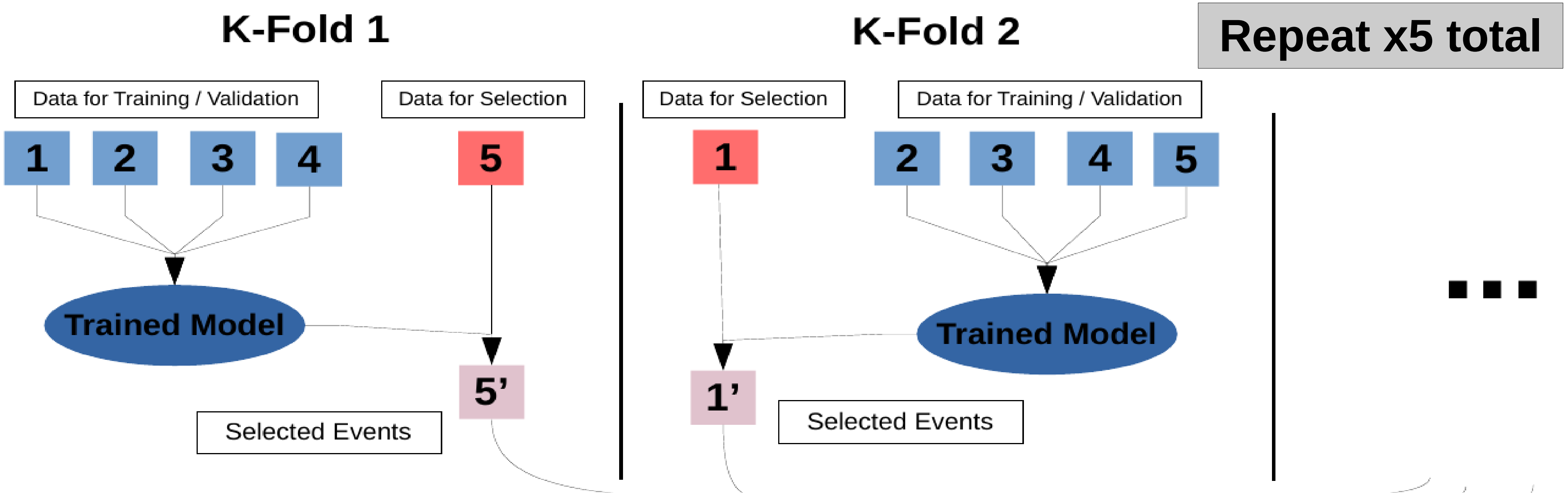
train on events from signal region

→ Ensure **no network reuses an event for both training and evaluation** to prevent overfitting issues

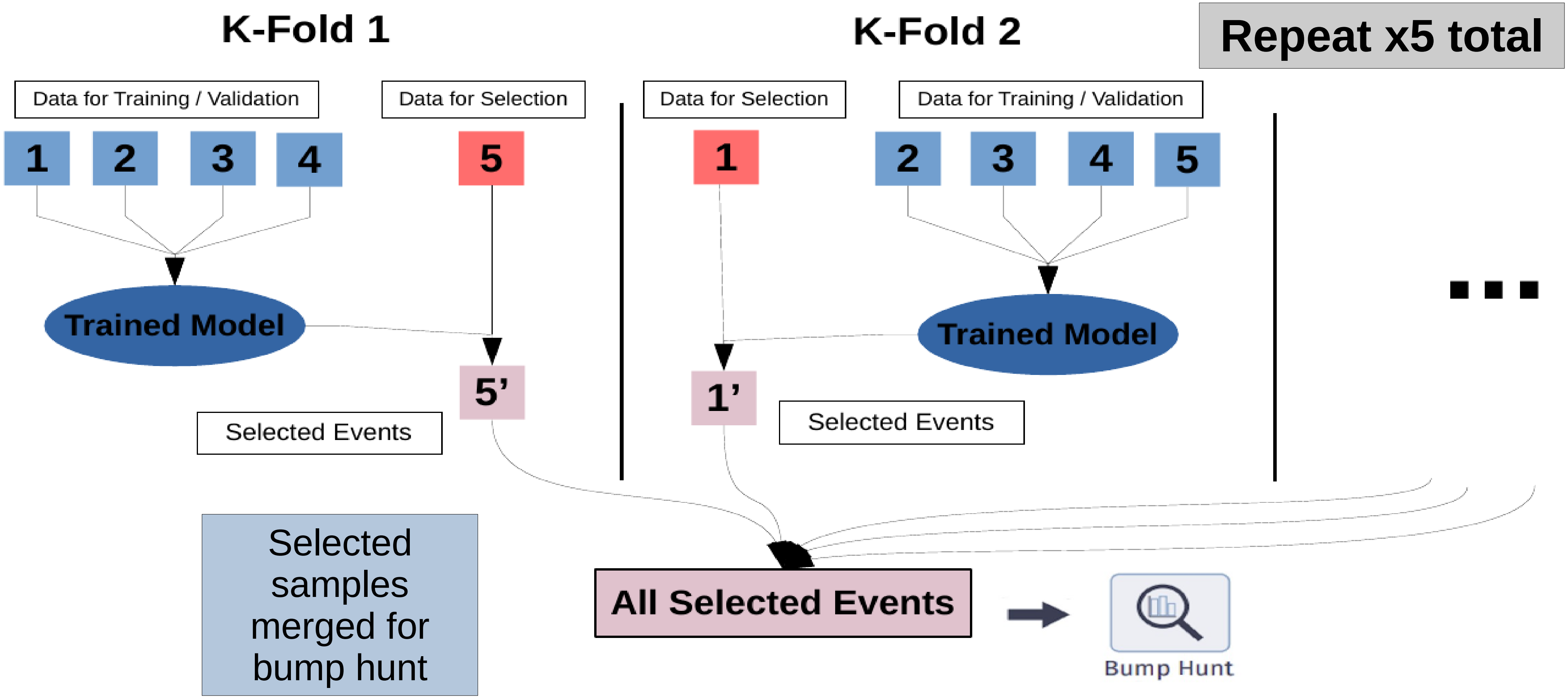
Cross Validation



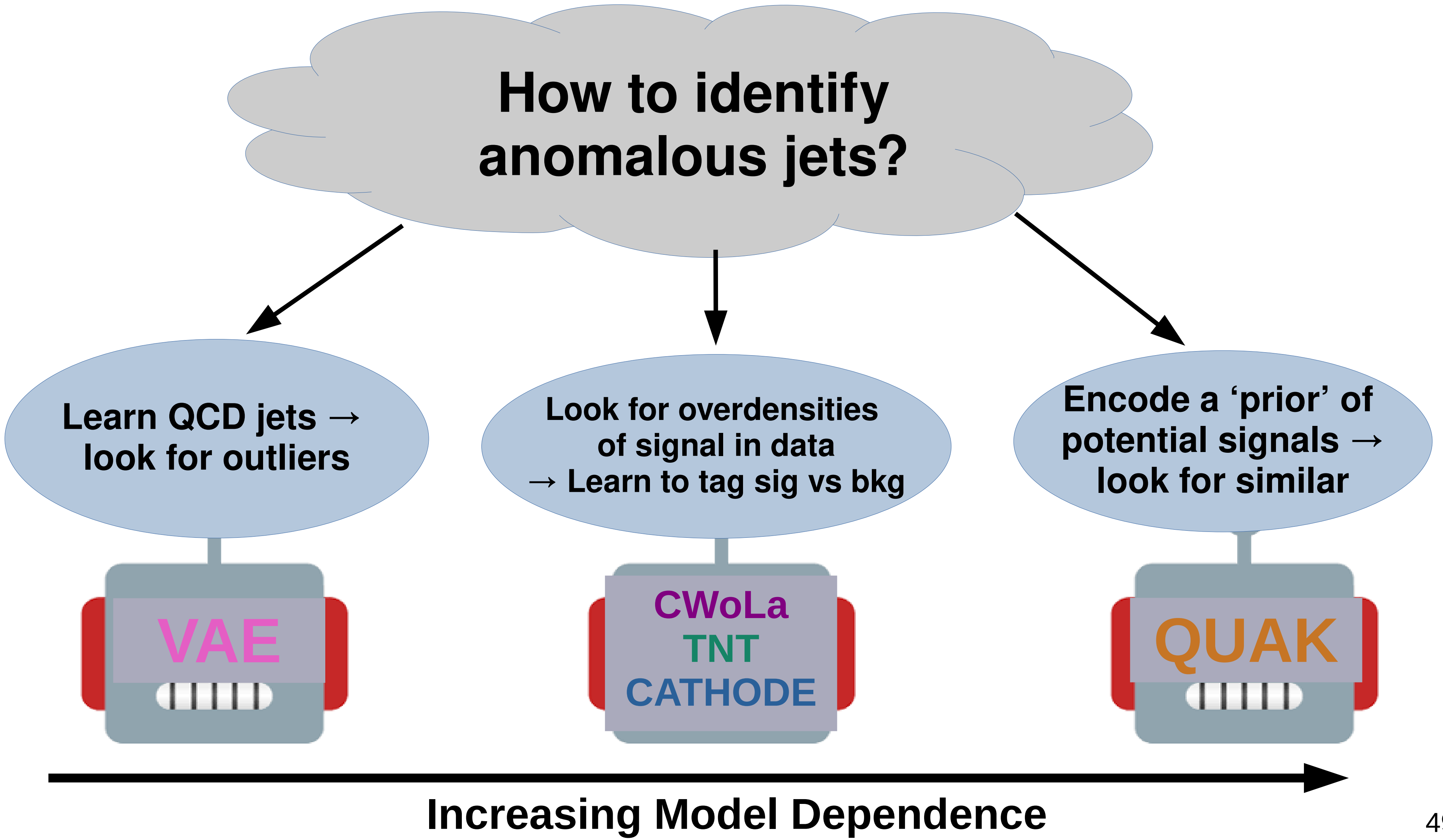
Cross Validation



Cross Validation

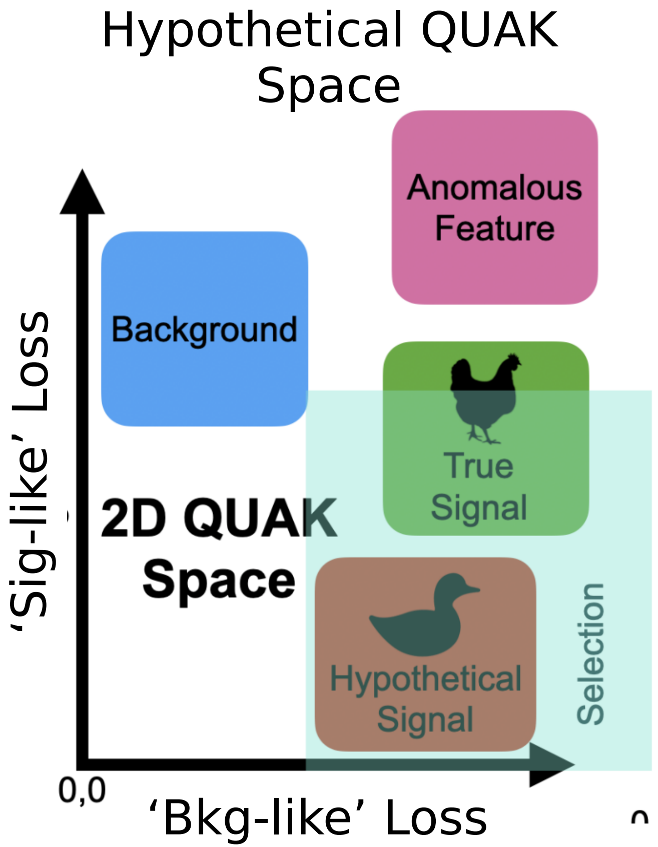


*Weakly supervised methods use additional layer of cross val for stability (see backup)


How to identify anomalous jets?
Learn QCD jets → look for outliers
Look for overdensities of signal in data → Learn to tag sig vs bkg
Encode a 'prior' of potential signals → look for similar
VAE
CWoLa
TNT
CATHODE
QUAK
Increasing Model Dependence

Quasi Anomalous Knowledge (QUAK)

- **Hybrid approach** between fully model-indep. and standard search
- **Encode a prior** on what a potential signal may look like
  - Use an AE trained on a variety of different signal MC's
- Construct 'QUAK space':
  - Loss of signal AE vs bkg AE
- Select events with low sig loss and high bkg loss



[Park et al 2011.03550]

Input Features

Low-level features | Hand-picked high-level features

VAE

Jet Constituents
p_x, p_y, p_z

CWoLa Hunting

Jet mass
τ_{21}
τ_{32}
τ_{43}
N_{const}

Leptonic energy frac.

Sub-jets b-tag score

TNT

Same as CWoLa Hunting

CATHODE

Jet masses
τ_{41}'s

CATHODE-b

+ Subjet b-tag scores

QUAK

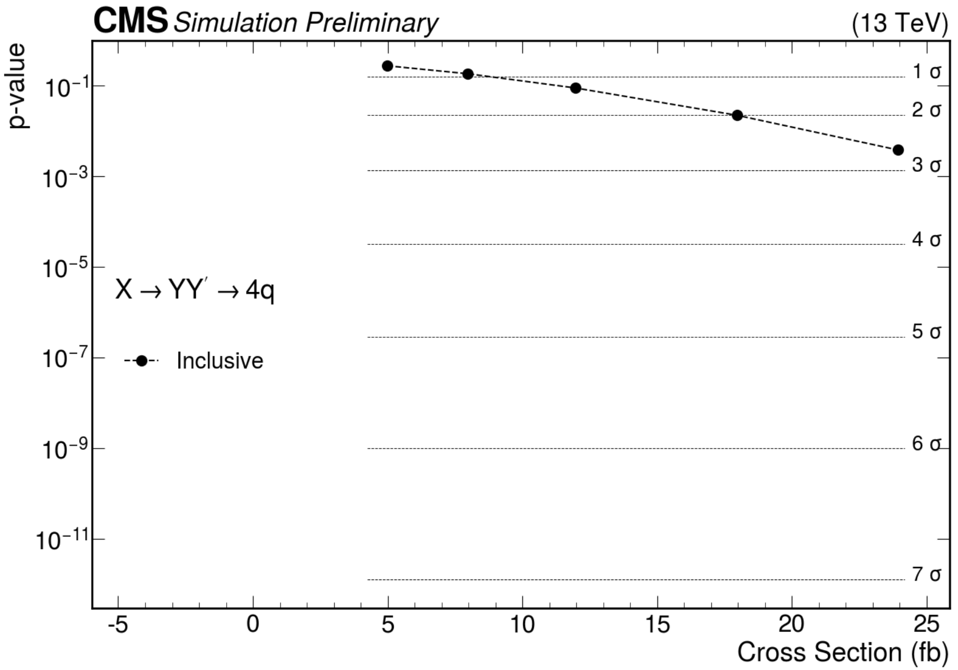
ρ = jet mass / p_T
τ_{21}'s
τ_{32}'s
τ_{43}'s
N_{const}'s

$\sqrt{\tau_{21}}/\tau_1$

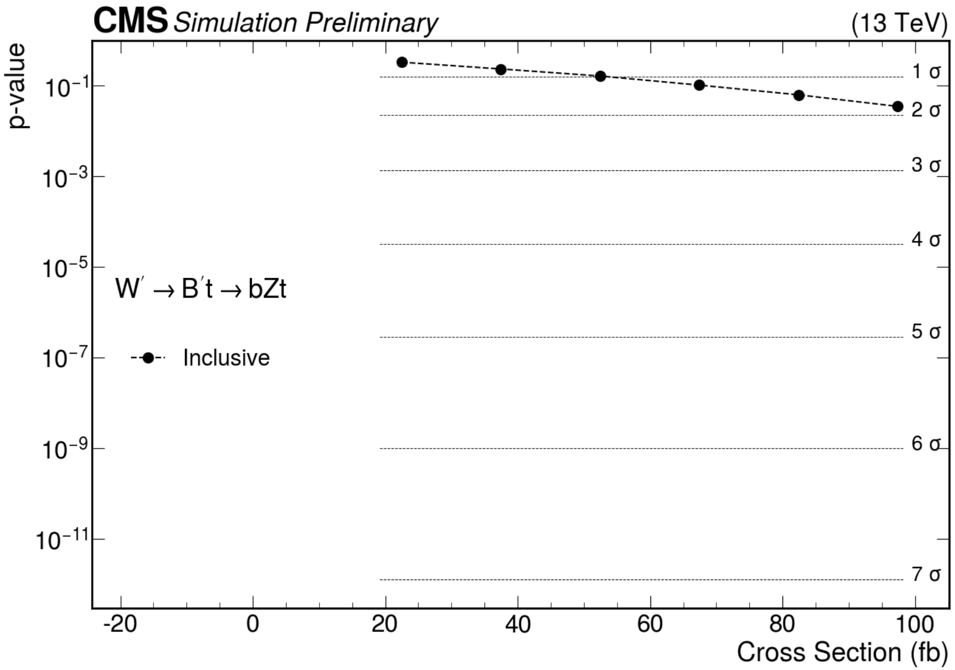
Sub-jets b-tag scores

Sensitivity

2 Pronged Signal



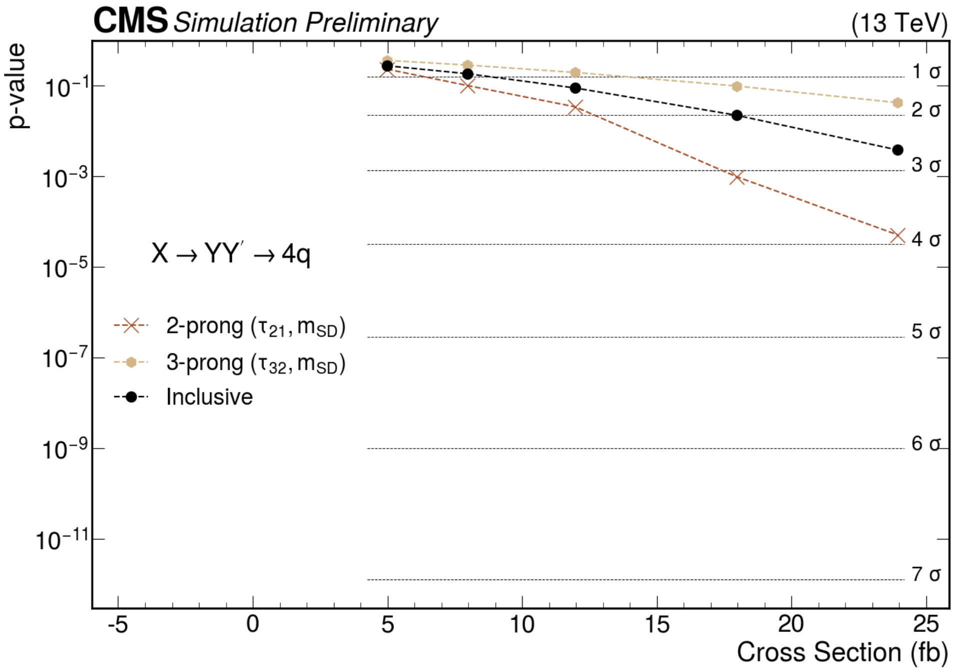
3 Pronged Signal



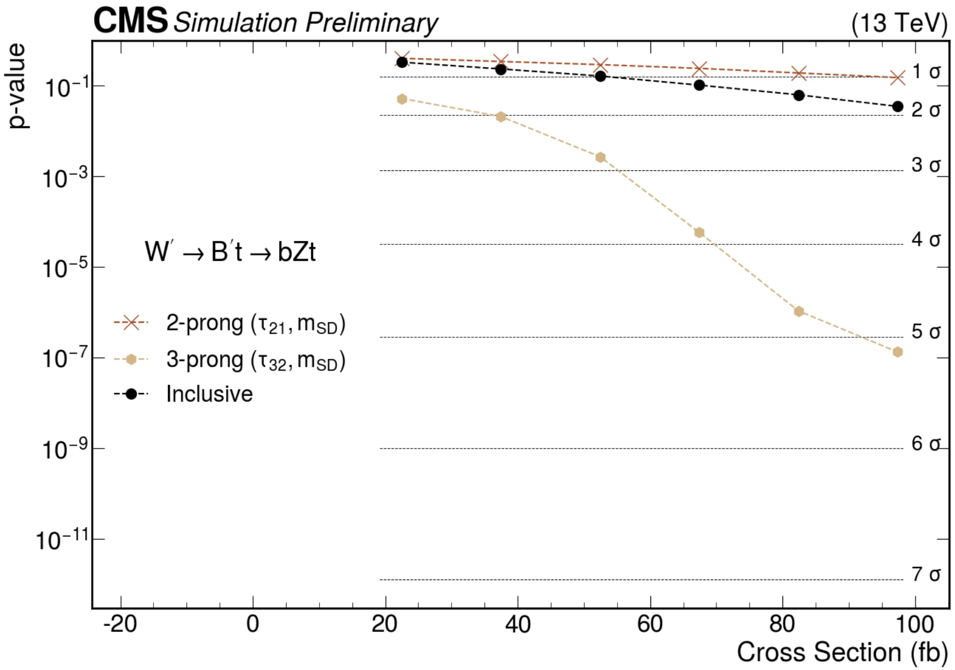
Inclusive analysis (no substructure cuts) sees only "hints"

Sensitivity

2 Pronged Signal



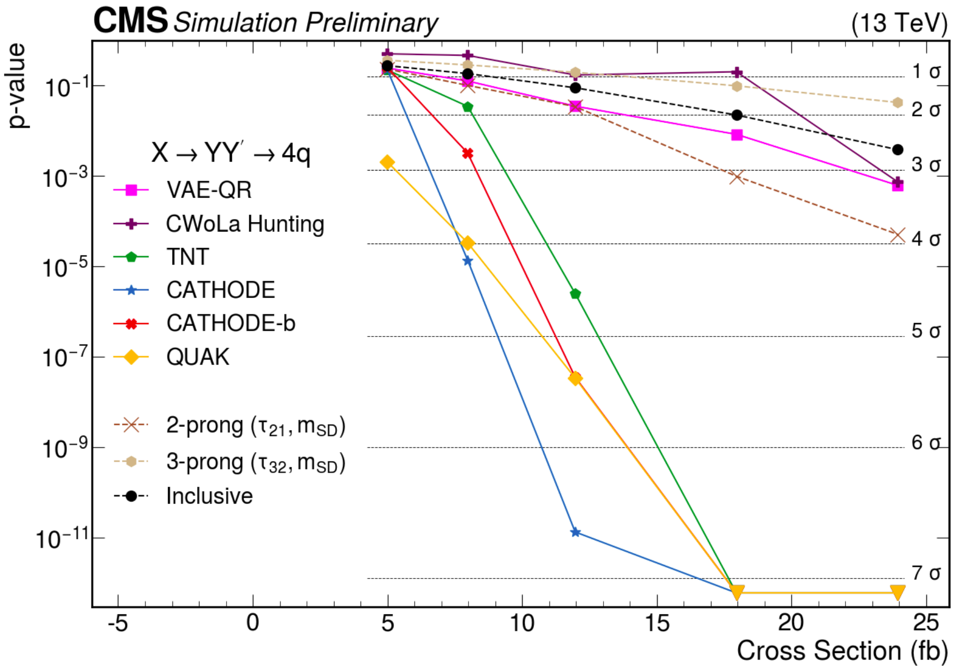
3 Pronged Signal



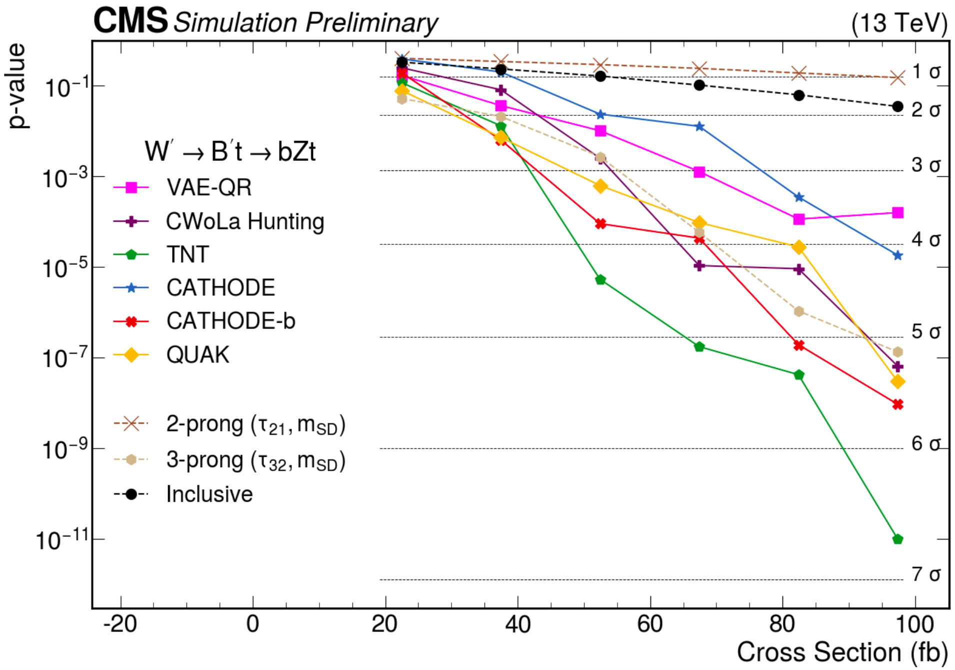
Traditional substructure cuts enhance sensitivity for a specific model, but not others

Sensitivity

2 Pronged Signal



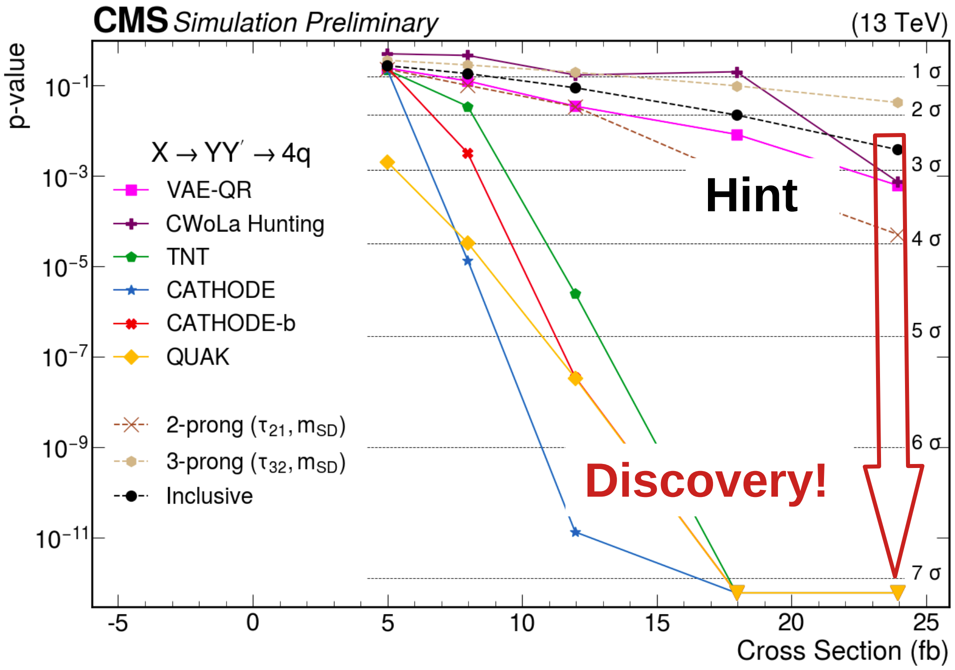
3 Pronged Signal



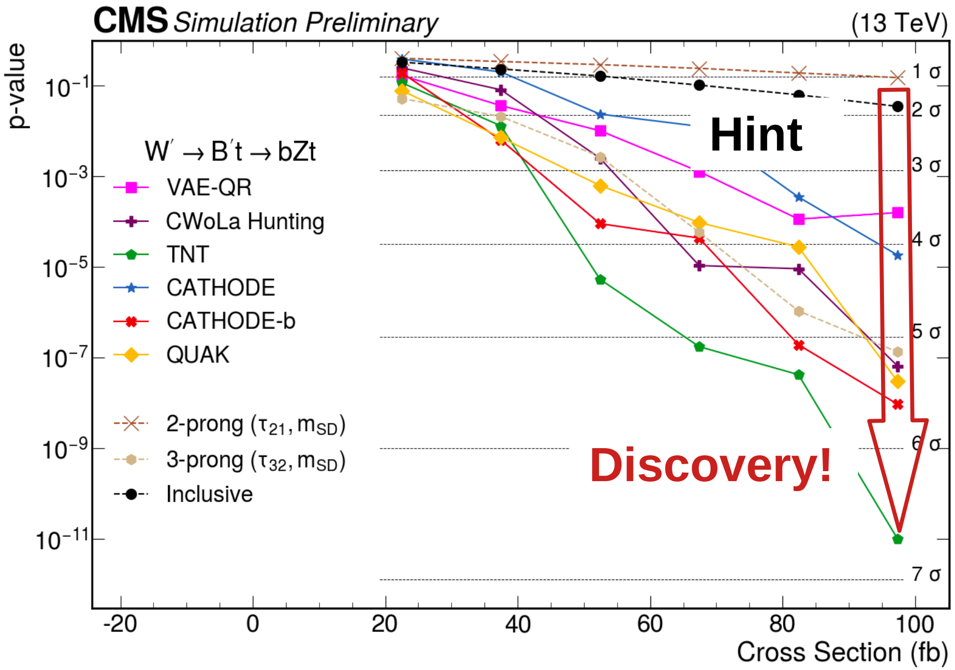
Anomaly detection enhances sensitivity for many models at once!

Sensitivity

2 Pronged Signal



3 Pronged Signal

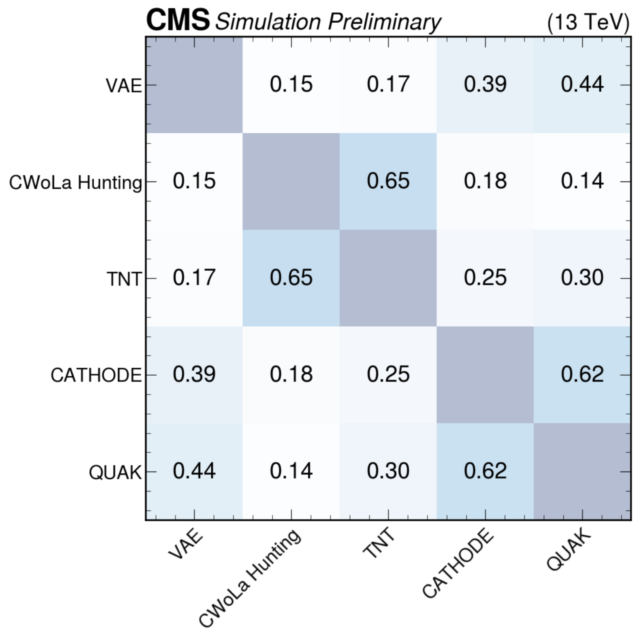


Anomaly detection enhances sensitivity for many models at once!

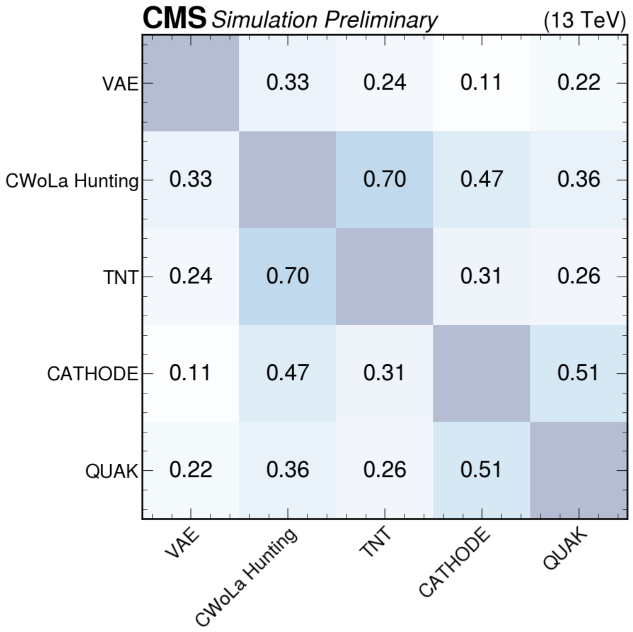
Complementary

CMS-NOTE -2023-013

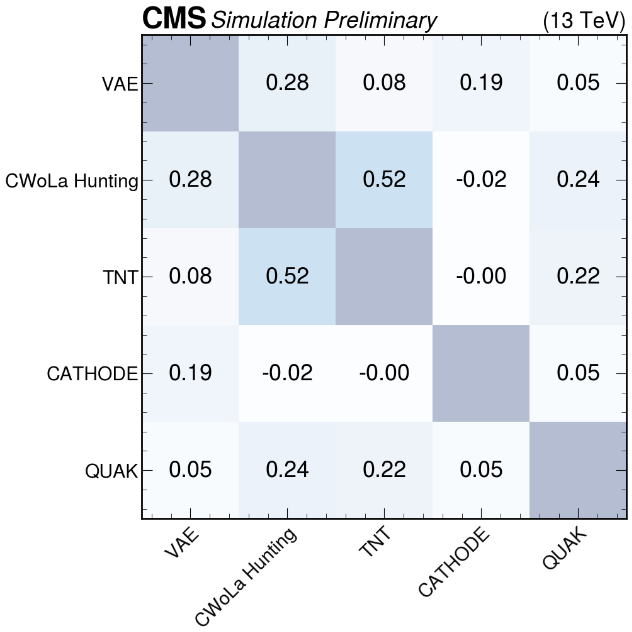
X→ YY → qq qq



W’→ B’t → bqq bqq



QCD Bkg.



- Compute correlation coefficients between different anomaly scores
- Complementary approaches lead to relatively low correlations!

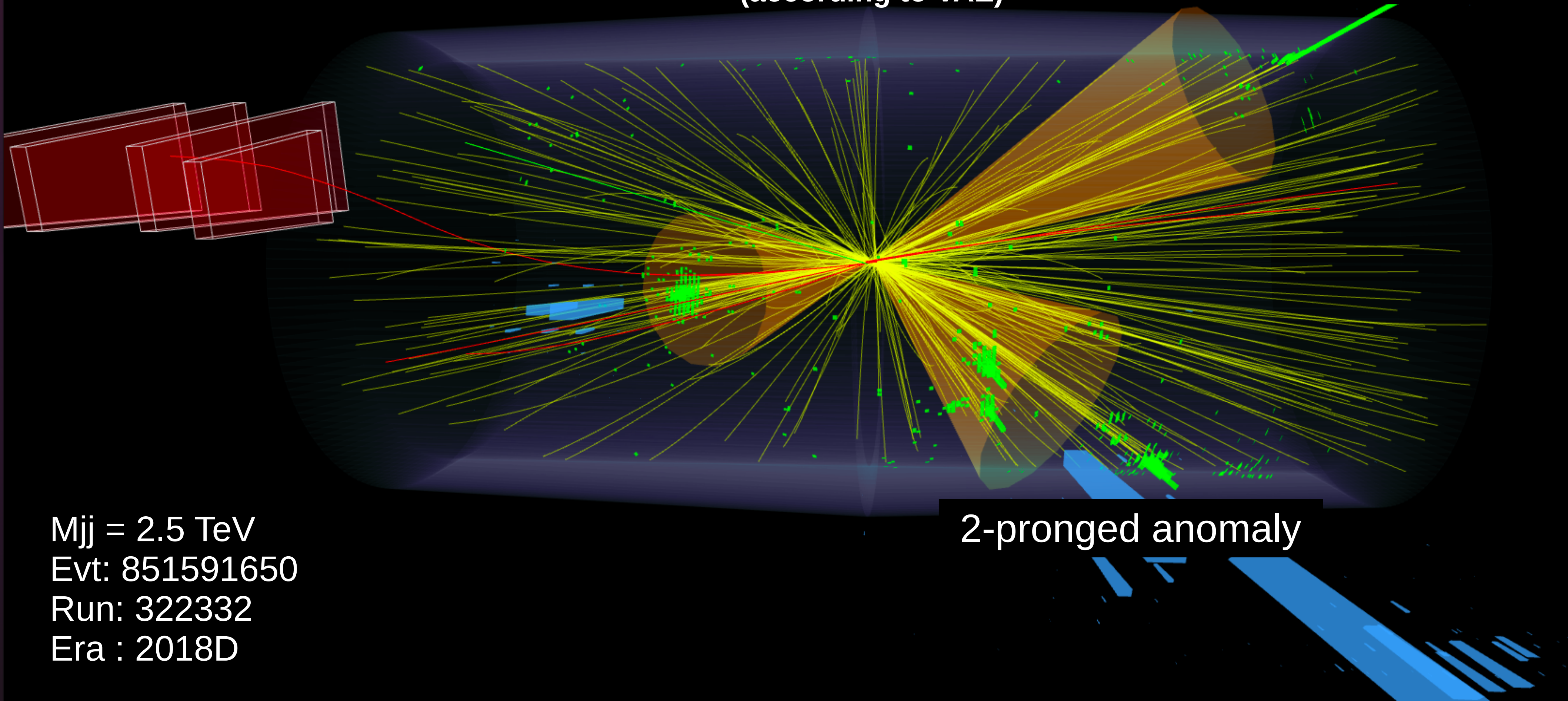
Steps to Unblinding

- ✔ No method creates artificial excesses in MC
- ✔ Can successfully find anomalies in MC
- ✔ Can characterize anomalies if found
- ✔ Apply to data $|\Delta\eta|$ sideband $\rightarrow$ no excesses

Time to apply to unblind!

One of our most anomalous events!

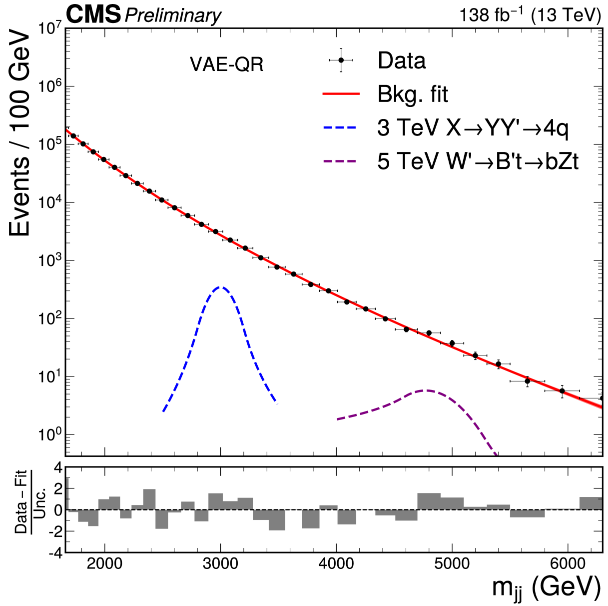
(according to VAE)

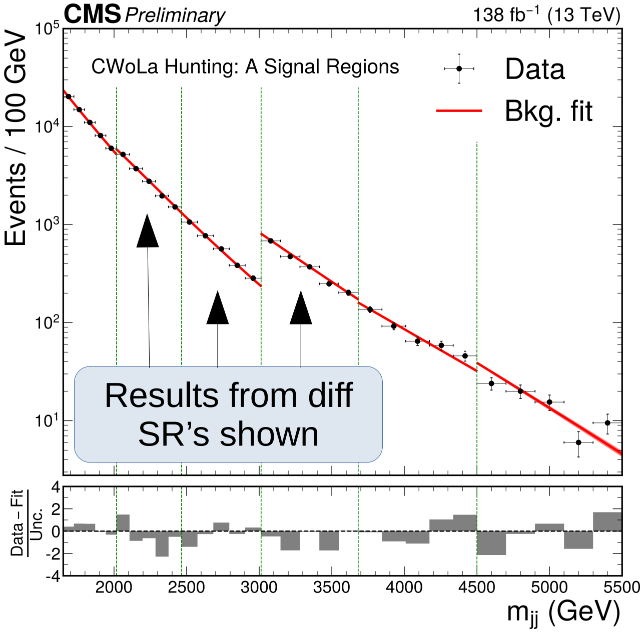


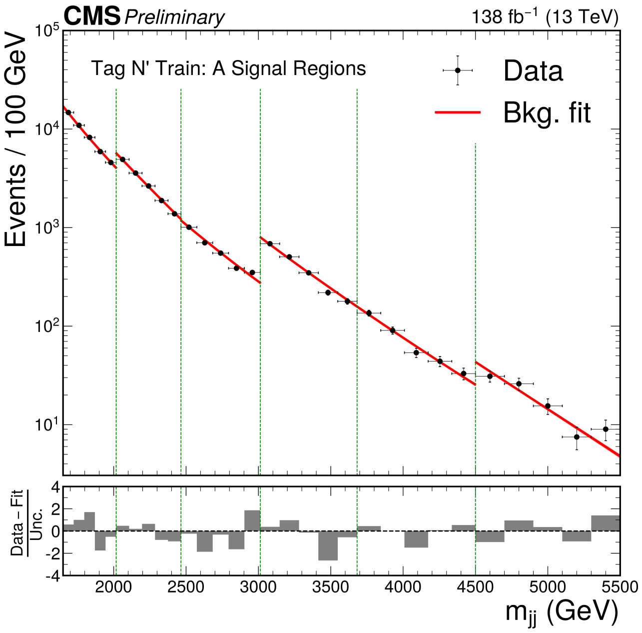
Mjj = 2.5 TeV
Evt: 851591650
Run: 322332
Era : 2018D

Search Results

No significant excesses from any method



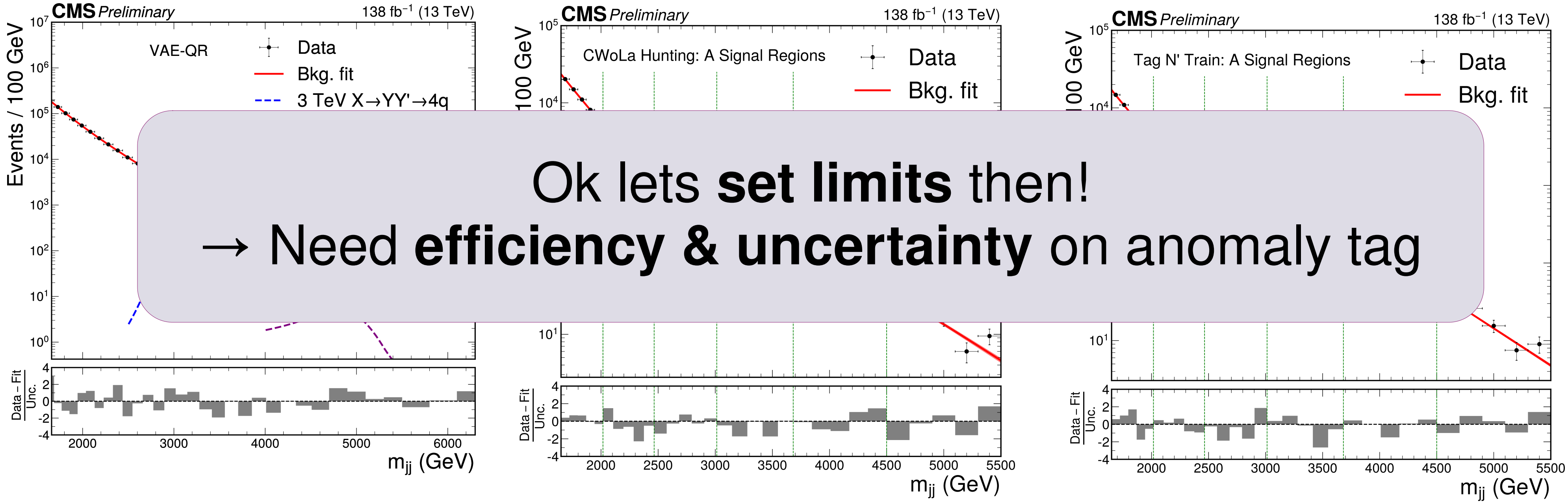




QUAK & CATHODE
results similar

Search Results

No significant excesses from any method



Ok lets **set limits** then!

→ Need **efficiency & uncertainty** on anomaly tag

QUAK & CATHODE
results similar

Efficiency & Uncertainties

To set a limit on a **specific signal model** proceed as usual

- Signal MC + anomaly detector $\rightarrow$ efficiency

Efficiency & Uncertainties

To set a limit on a **specific signal model** proceed as usual

- Signal MC + anomaly detector → efficiency
- One **complication** for weakly supervised methods : **signal eff depends on signal xsec**!
  - Novel methods to calibrate this (requires training lots & lots of NN's), see backup

Efficiency & Uncertainties

What about uncertainties?

Efficiency & Uncertainties

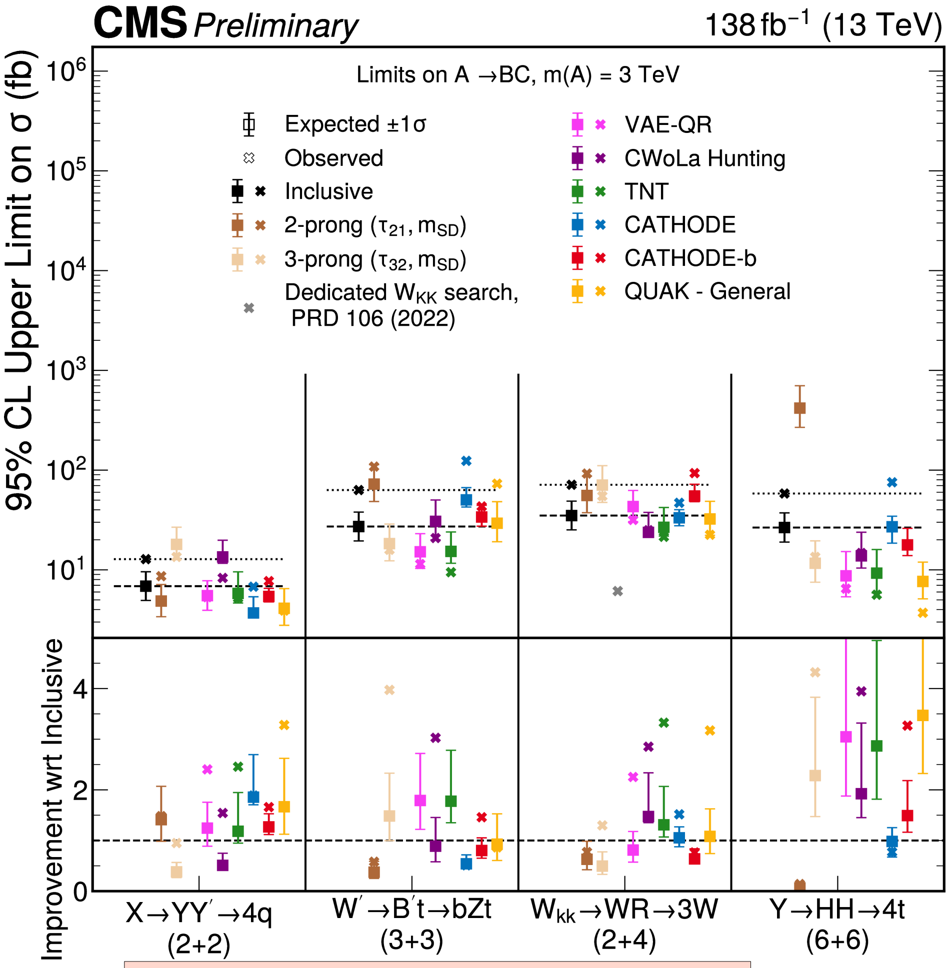
What about uncertainties?

- Anomaly cut just like any other multivariate cut → no ‘special’ uncertainties
- Largest uncertainty is from MC **modeling of jet substructure**
- Developed **new** data-driven correction + uncertainty for modeling high prong jets!
  - Per-prong substructure correction using Lund Jet Plane
  - CMS DP-2023/046

Limits

- Compute limits on benchmark from all **anomaly methods** on variety of signal models
  - Compare against **inclusive** & traditional **model-specific** approaches
  - First-ever limits on most of these models!

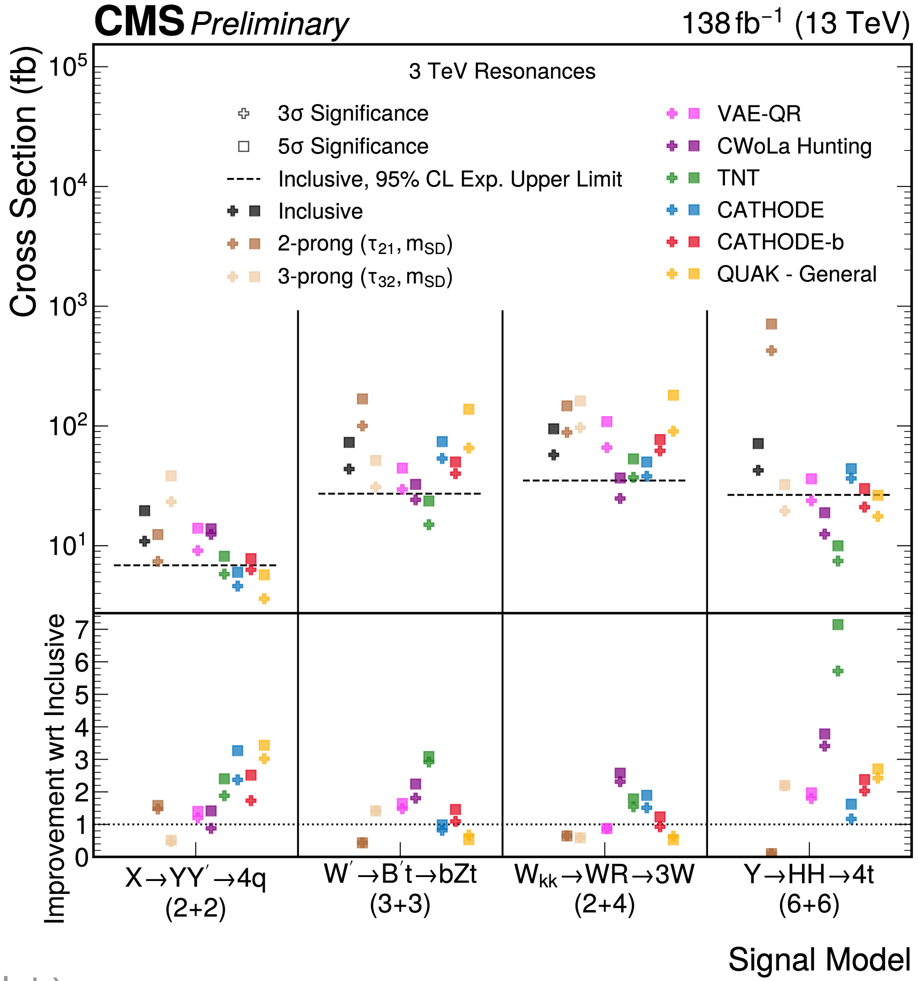
- Anomaly detection improves limits by ~2-3x!
  - Does not reach sensitivity of **dedicated search**



Very different sig. models!

Discovery Sensitivity

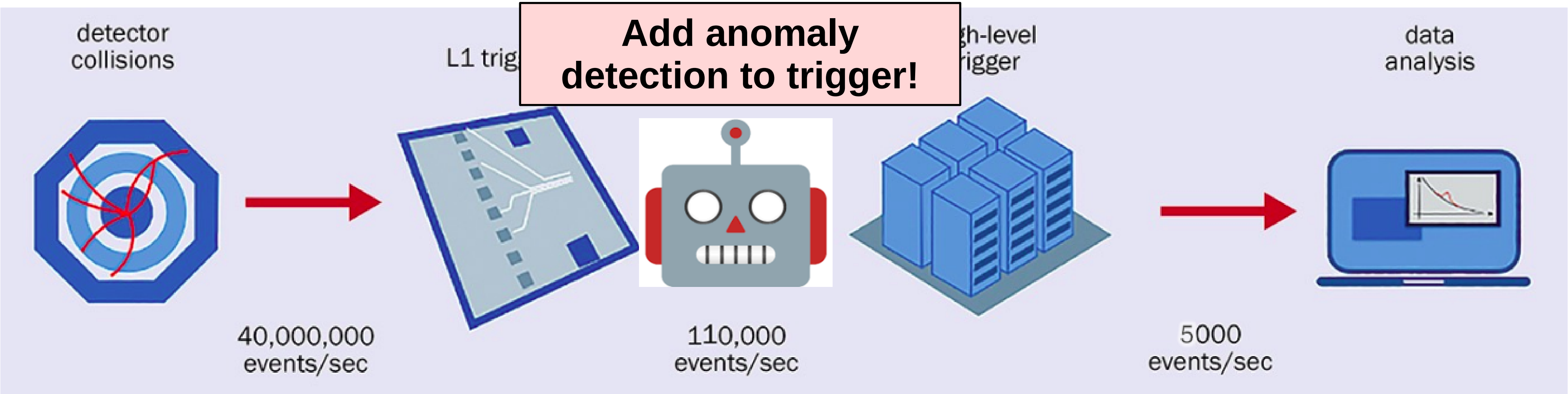
- ‘**Discovery focused**’ performance metric
- “What cross section do I need to get an expected 3σ/5σ excess?”
- **Anomaly methods** improve sensitivity by **~3-7x**!



Whats Next?

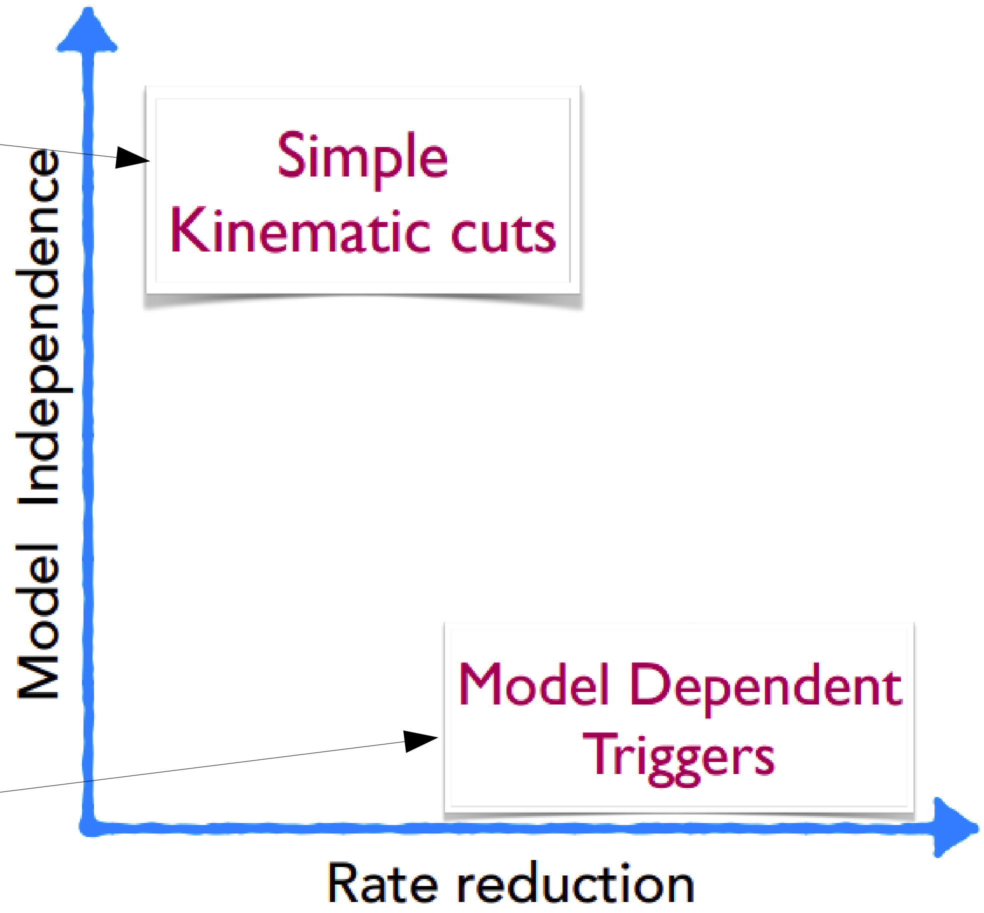
Trigger

- Lots more to explore
- But could still be bottlenecked by the trigger!



L1 Trigger Strategies

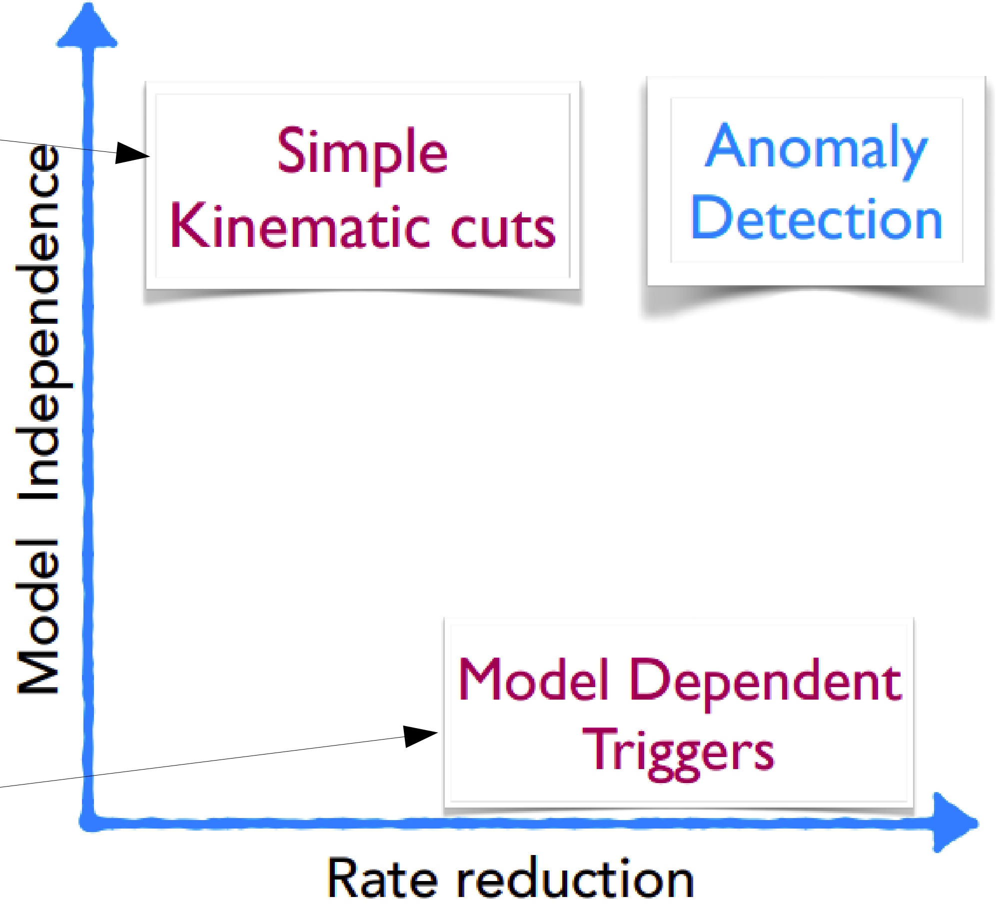
Eg.
Lepton’s with a given p_T
HT, MET, etc.



Eg.
LLP Triggers
VBF Triggers
Etc.

L1 Trigger Strategies

Eg.
Lepton's with a given p_T
HT, MET, etc.



Best of both ?

Eg.
LLP Triggers
VBF Triggers
Etc.

Anomaly Detection at L1

- CMS has developed **two** an anomaly detection triggers
- Based on autoencoder's trained on zero bias data
- Many 'tricks' used to fit onto FPGA and operate at 40 MHz!!

Global Trigger

AXOL1TL

Calorimeter

CICADA

hls 4 ml

QKeras

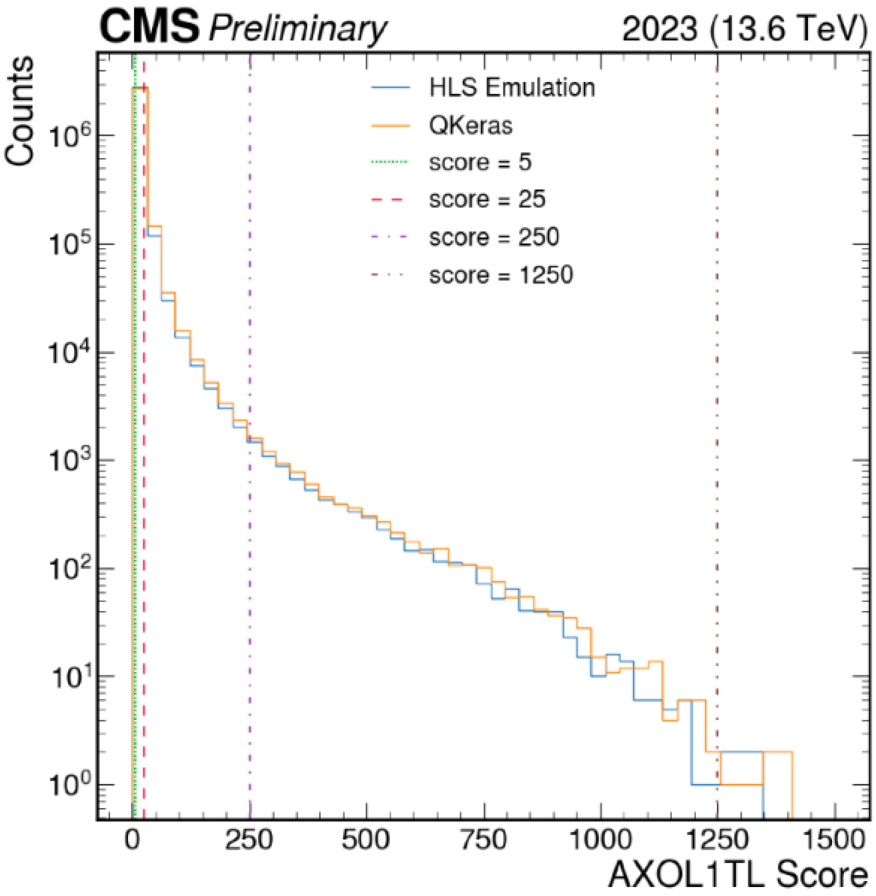
AXOL1TL led by FNAL postdoc Abhijith Gandrakota

AXOL1TL CMS-DP-2023-079
CICADA CMS-DP-2023-086

Anomaly Detection at L1

AXOL1TL

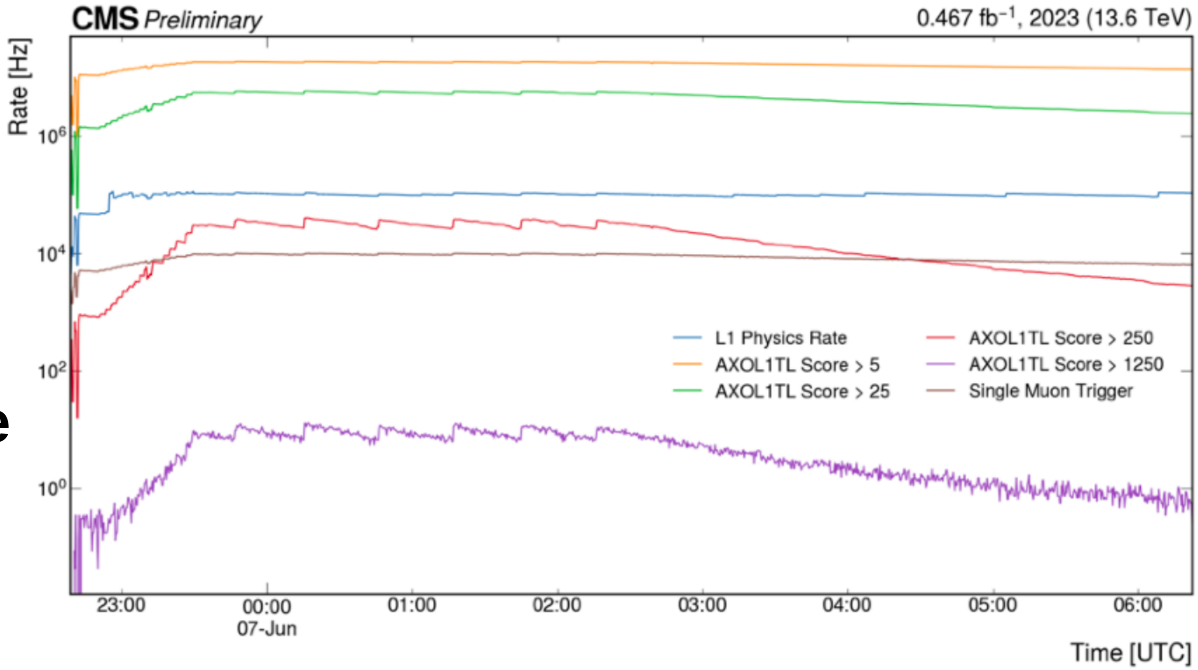
Thresholds on anomaly score chosen to achieve desired rate



CMS-DP-2023-079

In Action!

AXOL1TL was deployed in CMS trigger test crate during 2023 →
rates found to be stable

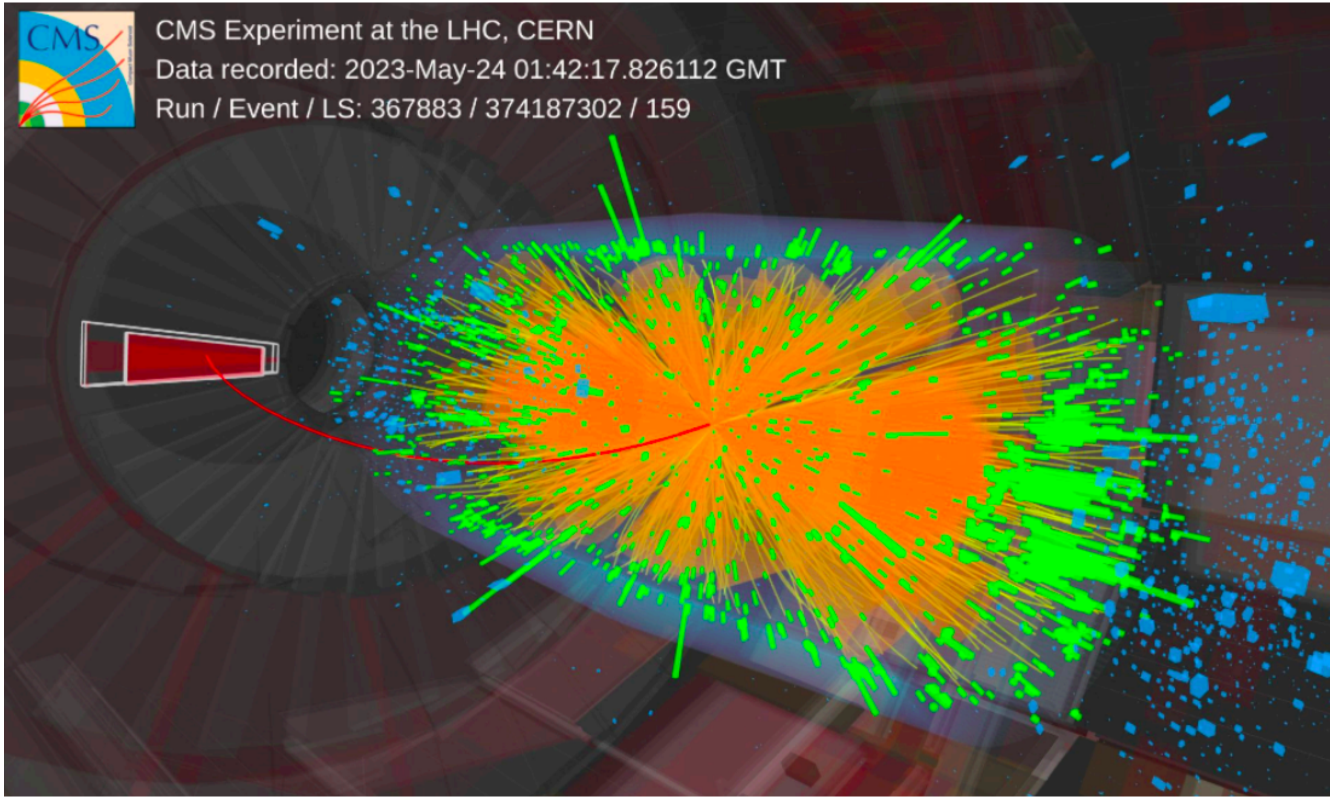


Deployed for real data taking in 2024 !

A L1 Anomalous Event

2023 event triggered only by **AXOL1TL**

Very busy, 11 jets + 1 muon



Future Questions

- No universally ‘best’ anomaly detection method ?
  - Can we combine approaches? Better methods?
- What **other generic topologies** can we apply anomaly detection to?
- How to analyze events from an **anomaly detection trigger**?
- How can we be sure our algorithms are **robust, safe and interpretable**?

Conclusions

- **First** usage of **anomaly detection** in CMS
  - Dijet resonance search with anomalous substructure
- Demonstrated sensitivity to **broad range of signals**
- New anomaly detection **trigger** coming online for 2024
- Many new directions to explore!

Excited to keep digging!

Backup

Dijet Fit

- After anomaly event selection, all methods use a common fitting/statistical framework
- **Bump hunt** performed on Mjj spectrum with 4 GeV bin size
  - Goodness of fit (χ^2) computed & plots shown on larger 'dijet binning' (bin size ~resolution)
- Background distribution modeled with standard 'dijet function', with 2, 3 or 4 params
  - For each fit, optimal number of params chosen with Fisher's F-test
- Signal shape is a double Crystal-Ball taken from fits to MC
  - For search use X→YY shape (relatively generic), interpolated to masses every 100 GeV
  - For limits use specific signal MC

Dijet Fn's

$$\frac{dN}{dm_{jj}} = \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}, \qquad \frac{dN}{dm_{jj}} = \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3\times\log(m_{jj}/\sqrt{s})}}, \text{and}$$

$$\frac{dN}{dm_{jj}} = \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3\times\log(m_{jj}/\sqrt{s})+P_4\times\log(m_{jj}/\sqrt{s})^2}}$$

Anomaly Cut

VAE

Single event selection

Search : 10% most anomalous events

Limits: 3 categories Top 1%, 1-5%, 5-10%

CWoLa Hunting

Selection changes for each SR (12 total)

Cut based on sideband eff.

Varies from 1% (low Mjj SR's) 5% (high Mjj SR's)

TNT

Same as CWoLa Hunting

CATHODE

Selection changes for each SR (12 total)

Cut based on Signal region eff.

1% for all SR's

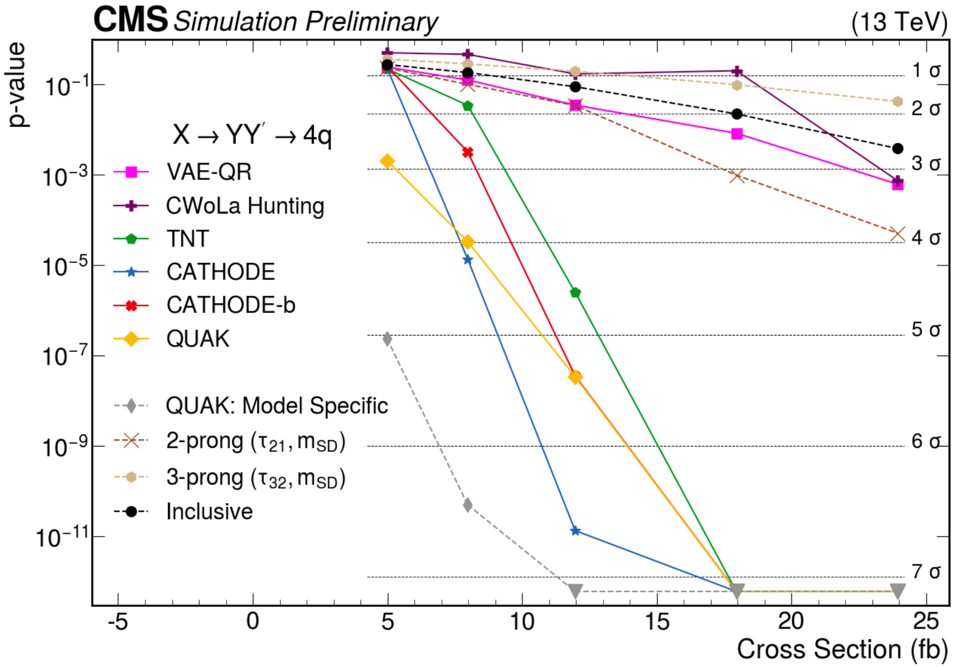
QUAK

Selection changes for each mass Hypothesis (100 GeV scan)

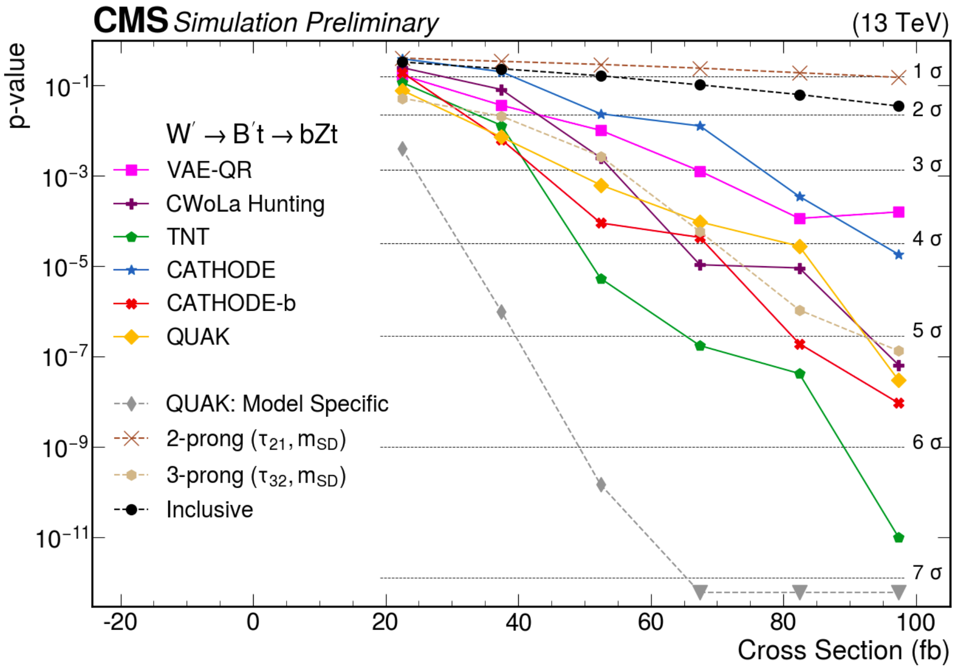
Iteratively select least populated QUAK space bins from SB's until reach specified # of events in SR

Sensitivity

2 Pronged Signal



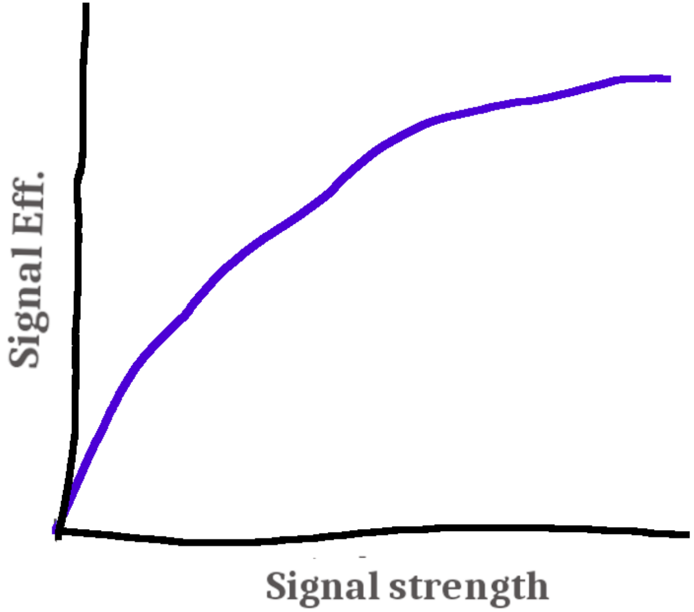
3 Pronged Signal



Changing Efficiency

For weakly supervised methods,

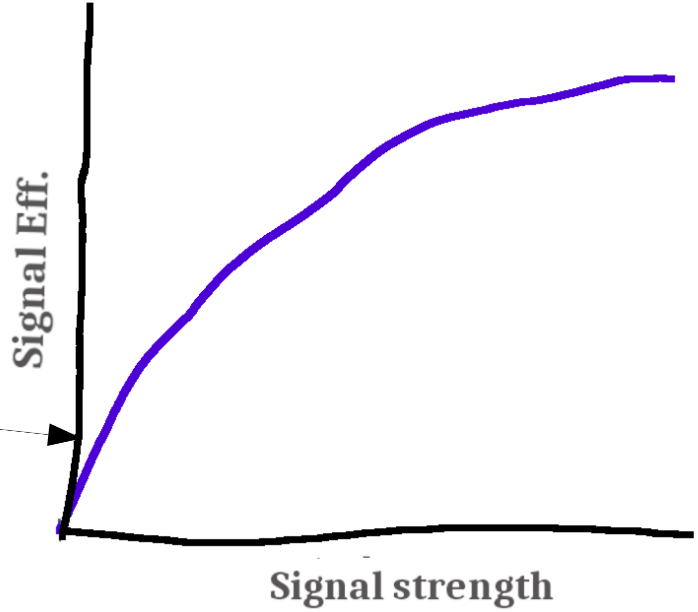
signal efficiency depends on signal cross section



Changing Efficiency

For weakly supervised methods,

signal efficiency depends on signal cross section



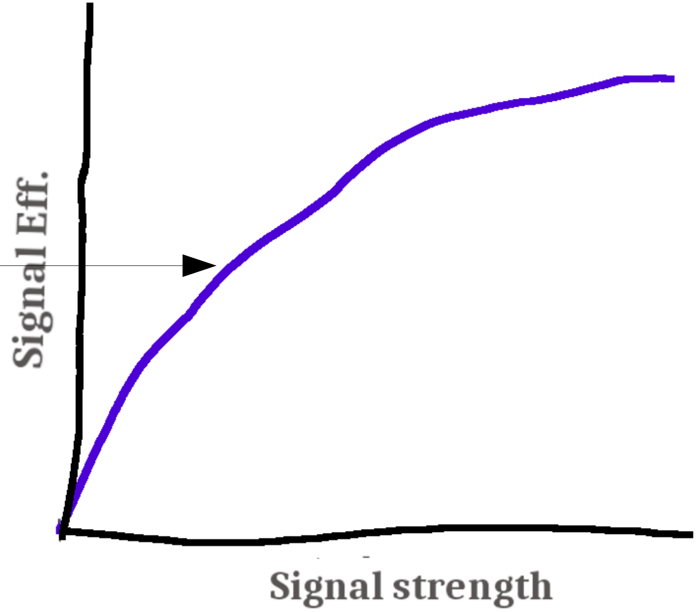
Low signal → classifier learns random noise → low eff.

Changing Efficiency

For weakly supervised methods,

signal efficiency depends on signal cross section

Performance improving with increasing signal

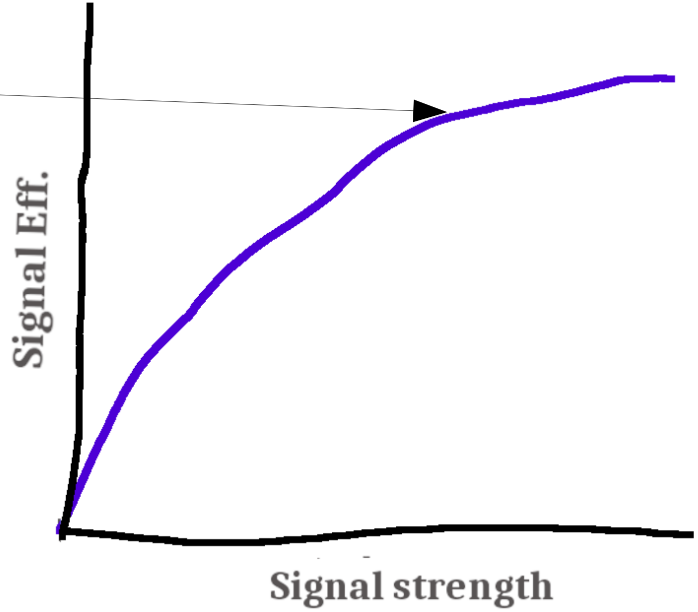


Changing Efficiency

For weakly supervised methods,

signal efficiency depends on signal cross section

Performance plateau's when reaches 'supervised' classifier

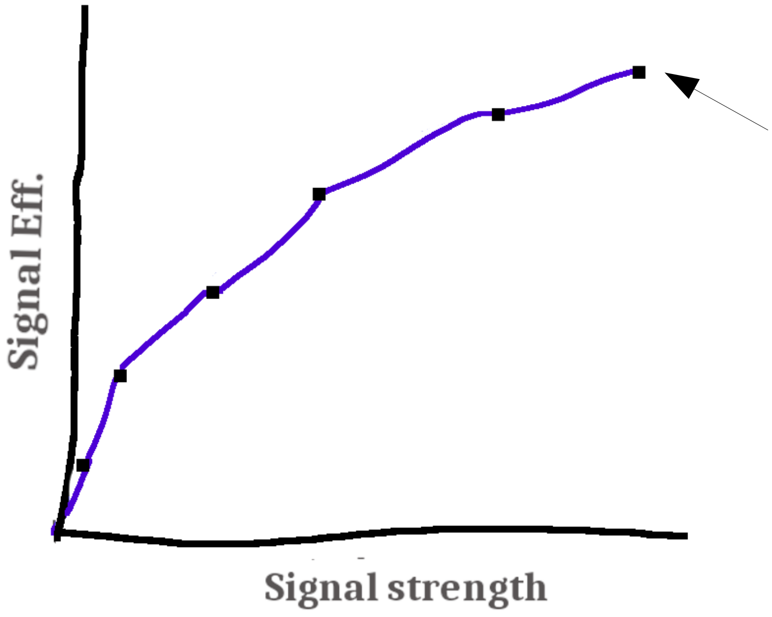


Changing Efficiency

For weakly supervised methods,
signal efficiency depends on signal cross section

Don't know the shape of the eff. curve a priori!

Calibrate by injecting signal at varying cross sections & check eff.



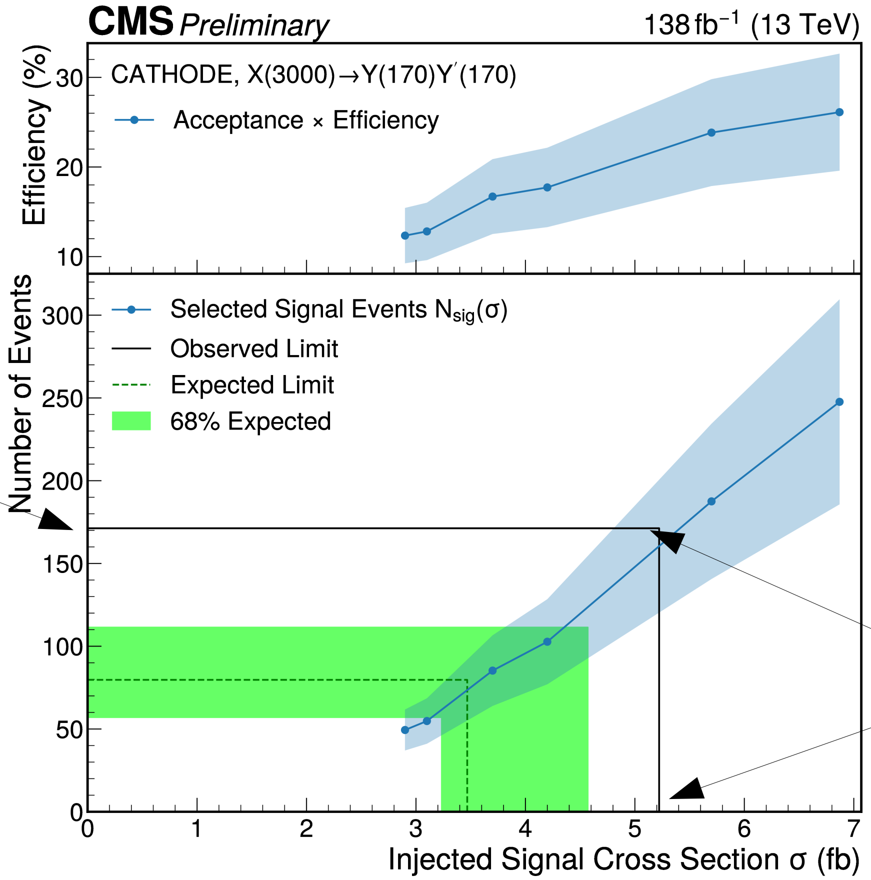
Evaluating each point requires training ~50 NN's!

Expensive for large scans of signal parameters...

Some systematics also require retraining
(see backup)

Limits with Changing Eff.

Limit on # of signal events in SR from fit (N_{exc})



Find $N_{sig}(\sigma)= L*\sigma*\varepsilon(\sigma)$ that matches N_{exc} → σ is limit

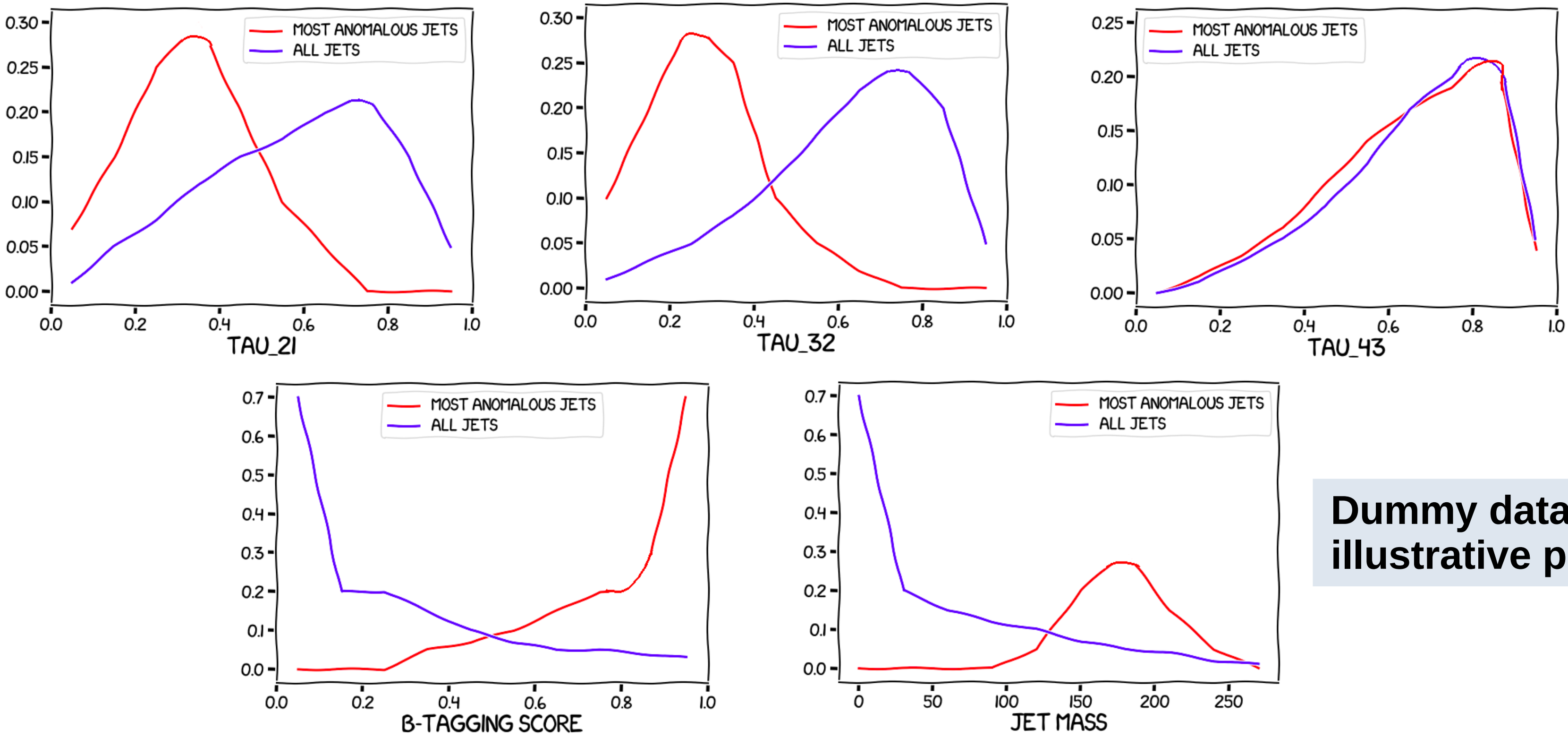
Weak Supervision Mass Scans

- Weakly supervised methods **assume a signal window** for training procedure
- Need to **scan** this window to cover full mass range
  - Repeat training procedure for each window
- Two sets of mass bins, **A** & **B**
  - **B** shifted half a bin width over wrt to **A**
  - Each with 6 signal regions
  - Require a sideband on either side of every signal region
- 12 total signal regions, different event selection for each one!

Bin Name	Range (GeV)	Signal Masses (GeV)
A0	1350-1650	-
A1	1650-2017	1800, 1900
A2	2017-2465	2200, 2300
A3	2465-3013	2600, 2700, 2800
A4	3013-3682	3200, 3300, 3400, 3500
A5	3682-4500	3900, 4100, 4200, 4300
A6	4500-5500	4800, 4900, 5000, 5100, 5200
A7	5500-8000	-
B0	1492-1824	-
B1	1824-2230	2000, 2100
B2	2230-2725	2400, 2500
B3	2725-3331	2900, 3000, 3100
B4	3331-4071	3600, 3700, 3800
B5	4071-4975	4400, 4500, 4600, 4700
B6	4975-6081	5300, 5400, 5500, 5600, 5700, 5800
B7	6081-8000	-

Understanding Anomalies

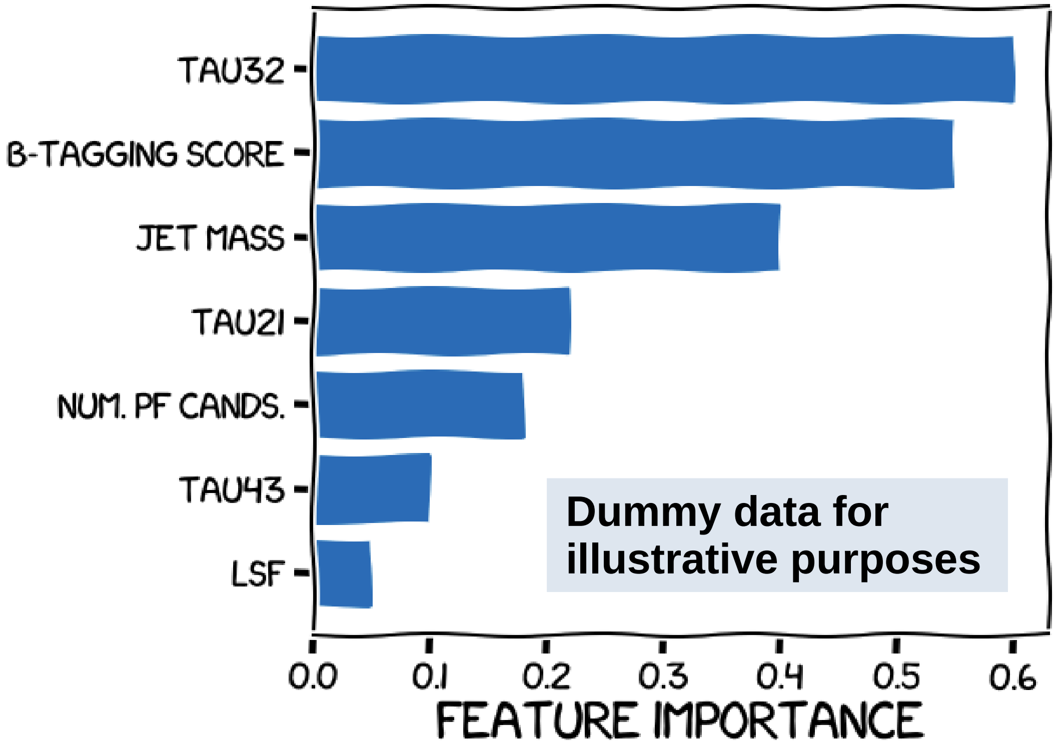
Compare features of **anomalous jets** to **regular ones**



Dummy data for illustrative purposes

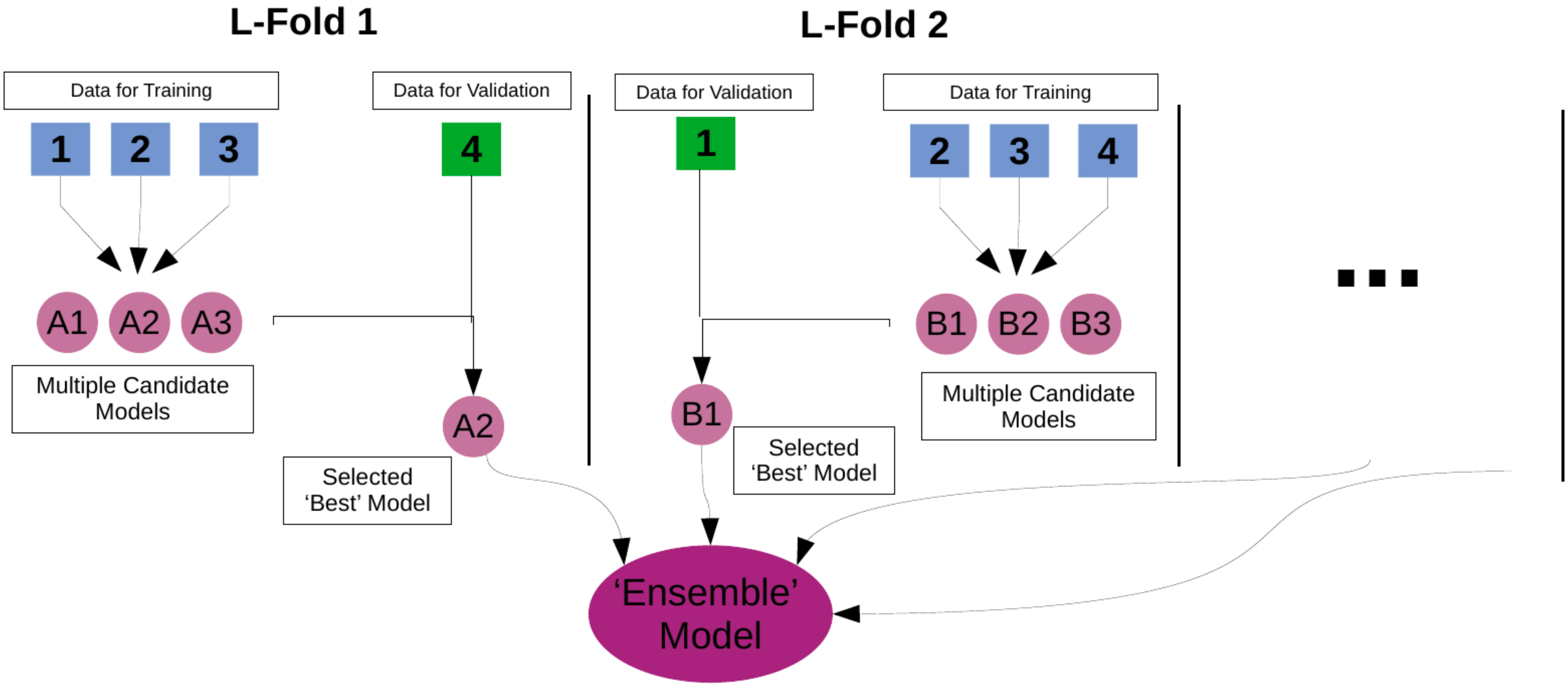
Understanding Anomalies

'Ask' the network!

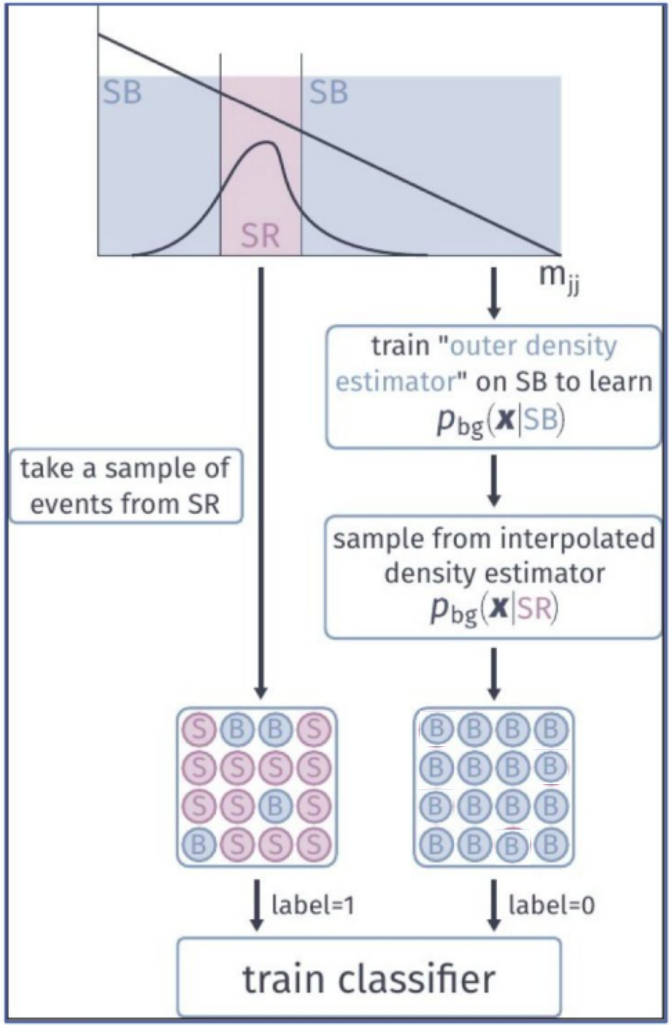


How much does my anomaly score change if I randomly perturb each feature?

Cross Val Part 2



CATHODE



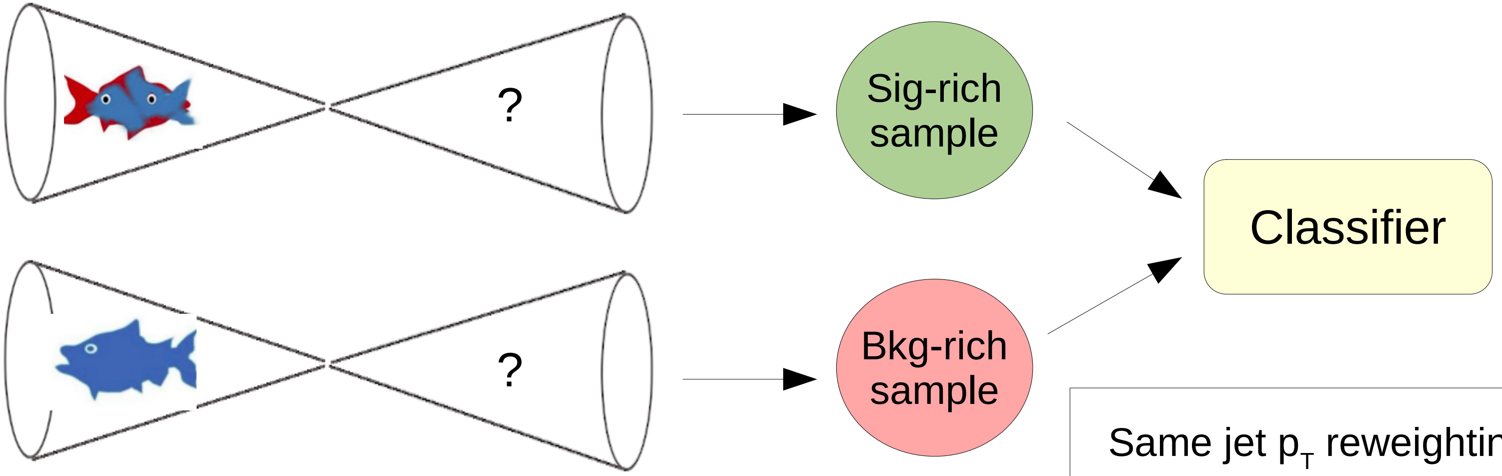
- Learn full multi-dim density $P_{bkg}(x | Mjj)$ from sidebands & **interpolate** into SR
  - 'Normalizing Flow'
- Draw samples to construct **bkg-rich sample**
- Weak supervision btwn data in SR and interpolated bkg samples

[Hallin et al 2109.00546]

Tag N' Train (TNT)

- Similar to CWoLa Hunting, but additional assumption that for signal **both jets are anomalous**
- Enhance purity of mixed samples by first tagging one jet each SR event with an autoencoder

AE is a CNN based on jet images

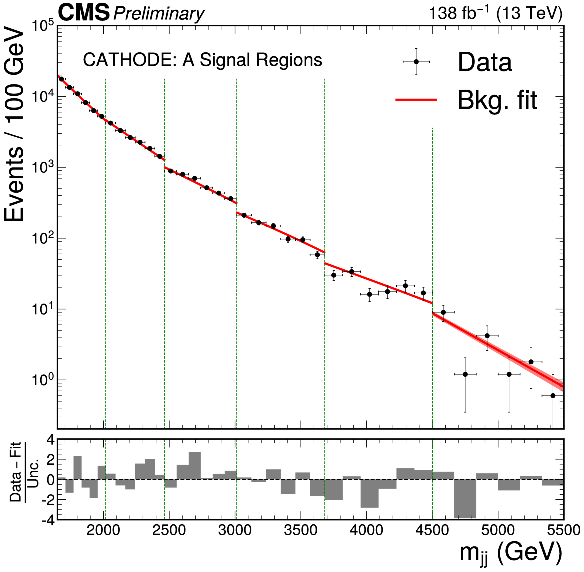


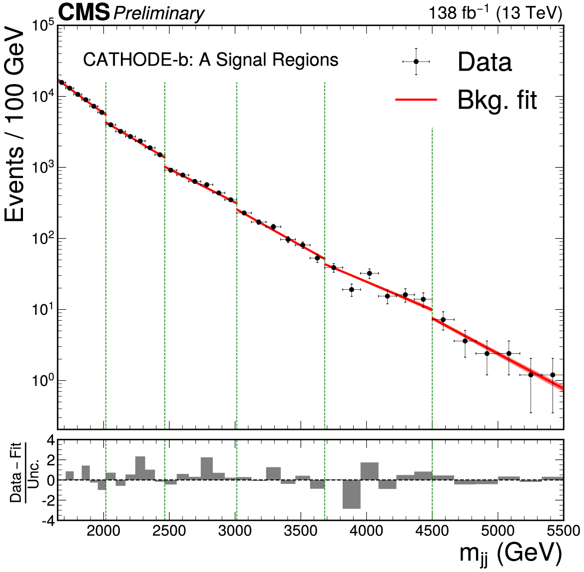
(full algo in backup)

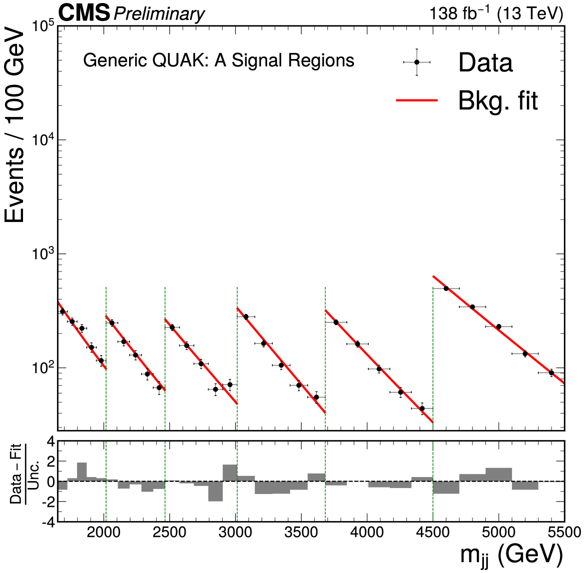
Same jet p_T reweighting method to ensure no correlation with Mjj

[**OA** & Suarez 2002.12376]

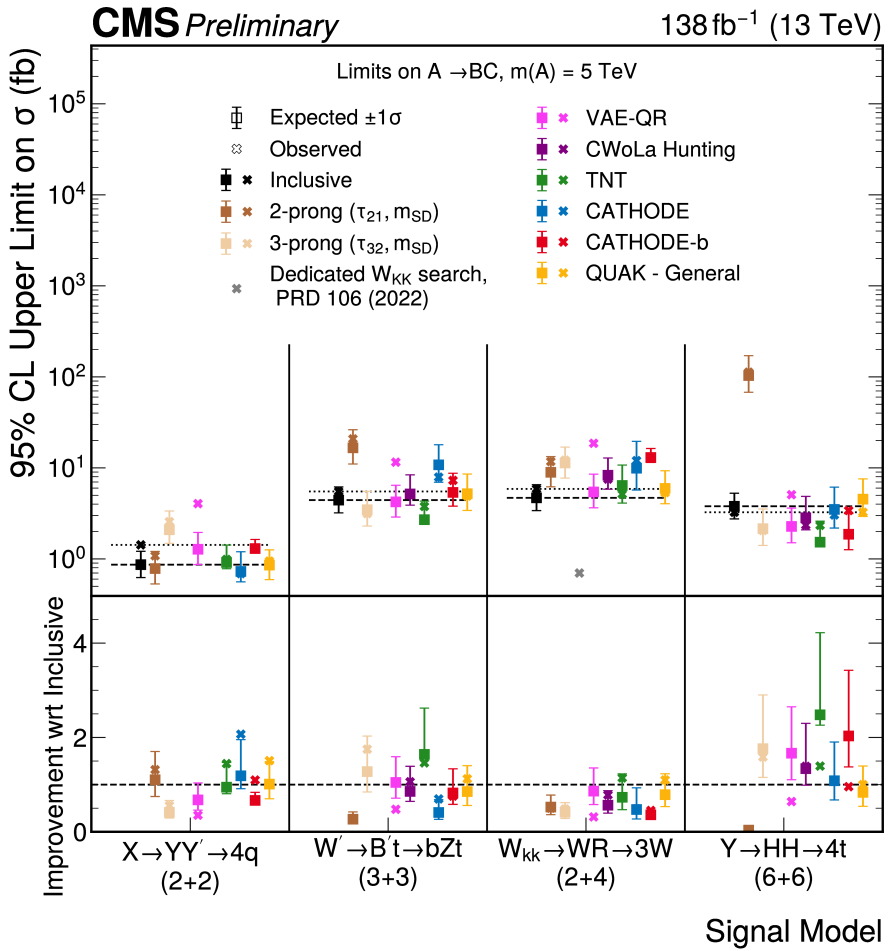
CATHODE & QUAK







5 TeV Limits



Oz Amram (Fermilab)

More Limits (3 TeV)

Signal Model (3 TeV)	Daughter Masses (GeV)	Method	Exp. (Obs.) Limit (fb)	Improvement wrt Inclusive
$Q* \to qW'$	25	*CWoLa Hunting*	61.1 (30.1)	0.3
$Q* \to qW'$	80	*CATHODE*	50.0 (95.2)	0.4
$Q* \to qW'$	170	*VAE-QR*	52.5 (37.5)	0.4
$Q* \to qW'$	400	*CWoLa Hunting*	45.8 (24.3)	0.5
$X \to YY' \to 4q$	25/25	*CATHODE*	8.0 (9.9)	0.9
$X \to YY' \to 4q$	25/80	*CATHODE*	7.6 (13.2)	0.9
$X \to YY' \to 4q$	25/170	*CATHODE*	10.3 (18.4)	0.7
$X \to YY' \to 4q$	25/400	*VAE-QR*	13.6 (12.5)	0.6
$X \to YY' \to 4q$	80/80	*CATHODE*	4.2 (8.0)	1.6
$X \to YY' \to 4q$	80/170	*CATHODE*	5.7 (11.4)	1.2
$X \to YY' \to 4q$	80/400	*CATHODE*	6.0 (7.3)	1.2
$X \to YY' \to 4q$	170/170	*CATHODE*	3.7 (6.8)	1.9
$X \to YY' \to 4q$	170/400	*VAE-QR*	4.4 (4.0)	1.7
$X \to YY' \to 4q$	400/400	*VAE-QR*	2.1 (1.9)	4.2
$W' \to B't \to bZt$	25	*TNT*	25.2 (17.4)	1.5
$W' \to B't \to bZt$	80	*TNT*	22.3 (14.6)	1.5
$W' \to B't \to bZt$	170	*TNT*	12.2 (7.3)	2.1
$W' \to B't \to bZt$	400	*VAE-QR*	15.2 (11.4)	1.8
$W_{KK} \to RW \to 3W$	170	*TNT*	25.1 (20.1)	1.4
$W_{KK} \to RW \to 3W$	400	*CWoLa Hunting*	23.8 (25.0)	1.5
$Z' \to T'T' \to tZtZ$	400	*QUAK*	28.3 (13.9)	2.7
$Y \to HH \to 4t$	400	*QUAK*	7.7 (3.7)	3.5

More Limits (5 TeV)

Signal Model (5 TeV)	Daughter Masses (GeV)	Method	Exp. (Obs.) Limit (fb)	Improvement wrt Inclusive
$Q* \to qW'$	25	*QUAK*	3.5 (3.1)	0.7
$Q* \to qW'$	80	*QUAK*	3.2 (2.8)	0.8
$Q* \to qW'$	170	*QUAK*	3.3 (3.6)	0.8
$Q* \to qW'$	400	*QUAK*	3.9 (9.9)	0.7
$X \to YY' \to 4q$	25/25	*QUAK*	1.7 (1.6)	0.5
$X \to YY' \to 4q$	25/80	*QUAK*	1.3 (1.3)	0.7
$X \to YY' \to 4q$	25/170	*QUAK*	1.1 (1.1)	0.8
$X \to YY' \to 4q$	25/400	*VAE-QR*	1.0 (3.4)	0.9
$X \to YY' \to 4q$	80/80	*TNT*	1.1 (1.2)	0.8
$X \to YY' \to 4q$	80/170	*QUAK*	0.9 (1.0)	0.9
$X \to YY' \to 4q$	80/400	*VAE-QR*	0.9 (3.0)	0.9
$X \to YY' \to 4q$	170/170	*CATHODE*	0.7 (0.7)	1.2
$X \to YY' \to 4q$	170/400	*VAE-QR*	0.7 (2.3)	1.2
$X \to YY' \to 4q$	400/400	*VAE-QR*	0.4 (1.1)	2.3
$W' \to B't \to bZt$	25	*TNT*	4.4 (6.2)	1.3
$W' \to B't \to bZt$	80	*TNT*	3.9 (5.7)	1.4
$W' \to B't \to bZt$	170	*TNT*	2.8 (3.5)	1.6
$W' \to B't \to bZt$	400	*TNT*	2.7 (3.8)	1.6
$W_{KK} \to RW \to 3W$	170	*TNT*	6.1 (7.2)	0.8
$W_{KK} \to RW \to 3W$	400	*VAE-QR*	5.4 (18.6)	0.9
$Y \to HH \to 4t$	400	*TNT*	1.5 (2.3)	2.5

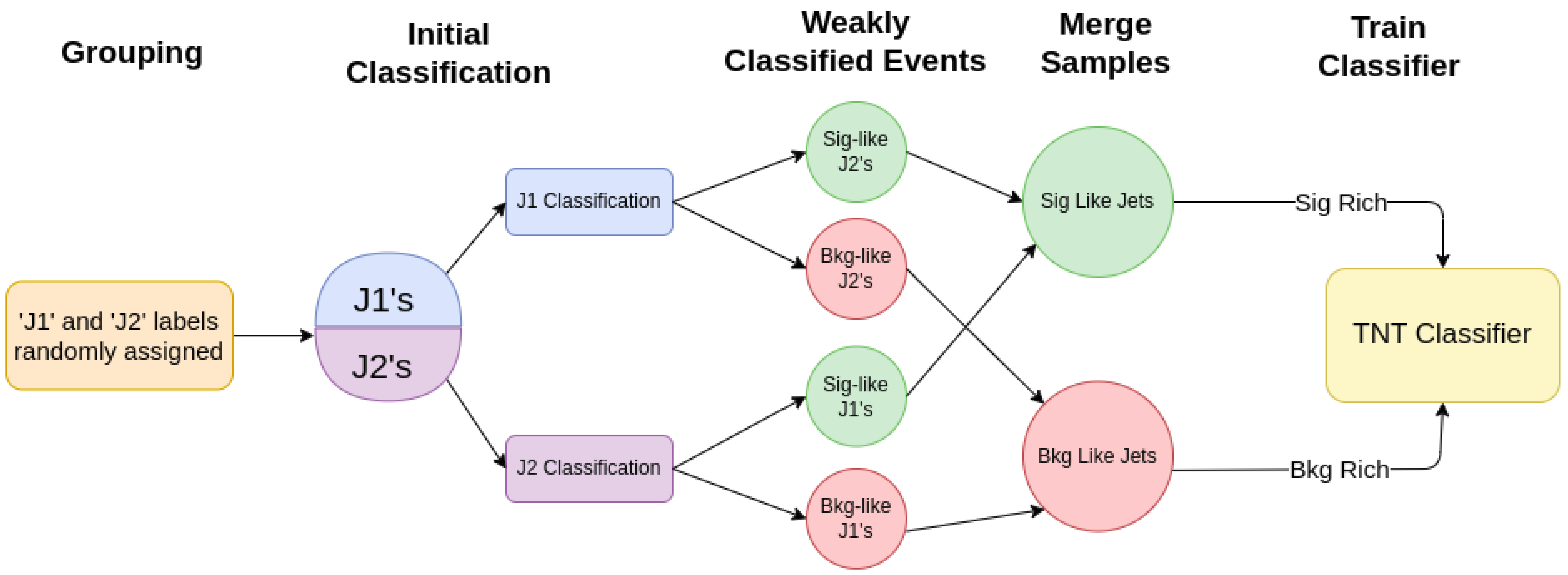
VAE Technical Details

- Latent space size of 12
- Training uses Chamfer loss + Kullback-Leibler divergence of between latent space & Gaussian
- Cross validation with 20 folds used for Quantile Regression
  - Average QR fit of other 19 folds used when selecting events on 20th
- QR fits use dense NN with 5 layers and 30 nodes per layer, output smoothed with 3^{rd} order polynomial
- **Three** categories used in limit setting
  - Cat1: Most anomalous 1% (>99%)
  - Cat2: Next most anomalous 4% (95-99%)
  - Cat3: Next most anomalous 5% (90-95%)
- In model-indep search, use single category, >90%

TNT Autoencoder

- AE used by TNT algorithm
- Based on ‘image’ representation of jets, follow approach of (1803.00107)
  - 32x32 pixels, covering η/φ from -0.6 to 0.6 around center of jet
  - Normalize sum of pixels to be 1 → less p_T dependence
- Trained with same k-folding as weakly supervised methods
- Separate AE trained for each SR using corresponding sidebands

TNT Diagram



Grouping
Initial Classification
Weakly Classified Events
Merge Samples
Train Classifier
'J1' and 'J2' labels randomly assigned
J1's
J2's
J1 Classification
J2 Classification
Sig-like J2's
Bkg-like J2's
Sig-like J1's
Bkg-like J1's
Sig Like Jets
Bkg Like Jets
Sig Rich
Bkg Rich
TNT Classifier

QUAK Signal Prior

- Train 6 separate AE’s trained on different signal samples
  - Grouped by daughter masses
- Signal AE’s
  - **M80-80**: XYY2000_Y80_Yp80
  - **M80-170**: Wkk2000_R170, Wkk3000_R170,Wp2000_B80_T170, Wp3000_B80_T170, XYY2000_Y80_Yp170
  - **M80-400:** Wkk2000_R400, Wkk3000_R400, XYY2000_Y400_Yp80, XYY2000_Y80_Yp400, XYY3000_Y80_Yp400
  - **M170-170**: Wp2000_B170_T170, Wp3000_B170_T170, XYY3000_Y170_Yp170
  - **M170-400**: Wp2000_B400_T170, Wp3000_B400_T170, XYY2000_Y170_Yp400, XYY2000_Y400_Yp170, XYY3000_Y400_Yp170
  - **M400-400:** YHH2000_H400, YHH3000_H400, ZTT2000_Tp400, ZTT3000_Tp400
- Normalize each score so mean 0, std dev 1
- Combine 6 scores into single ‘signal-like’ score use L5 signed norm

$$L_p^{\mathrm{sgn}} = \mathrm{sgn}\left(\Sigma_p\right) |\Sigma_p|^{1/p}, \quad \text{where}\, \Sigma_p = \sum_{i=1}^{k} \mathrm{sgn}(\ell_i)|\ell_i|^p.$$

Signal Systematic Uncertainties

- Analysis fully data driven, only systematics come from signal modeling
- Sys. unc. affects signal distribution → changes what NN learns in training → change in tagging efficiency
  - Need to do a dedicated injection + training to evaluate effect
- Once 'optimal' σ_{inj} found, do injections with different systematics variations and compute change in signal efficiency
  - Don't use event weights in NN training but rather a weighted random of sampling of signal events to produce sample to inject
  - Note that train *and* evaluate efficiency on shifted sample
- Once have an uncertainty on signal efficiency, rerun fit with it as a nuisance to get final limit

Systematic Filtering

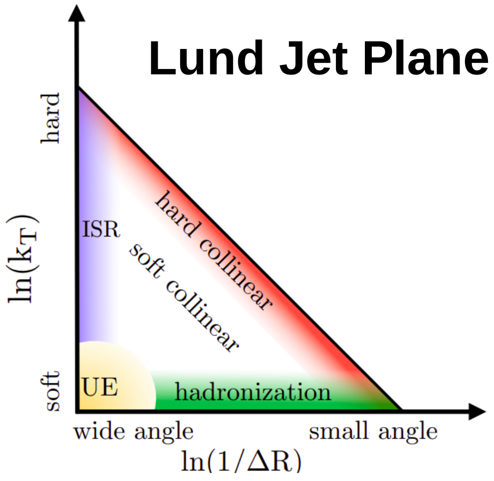
- Before re-training see which systematics realistically could have an effect on the training
  - Make histograms of training features + Mjj and see which systematics have deviations
- Every histogram 10 bins, spanning 1%-99% of distribution (cutting out outliers)
- Compare deviations of systematics as compared to **statistical uncertainty of signal injection divided by 5**
  - Very conservative threshold
- For systematics that have deviations larger than this threshold → retrain
- For others evaluate change in efficiency with nominal (fixed) classifier on shifted samples

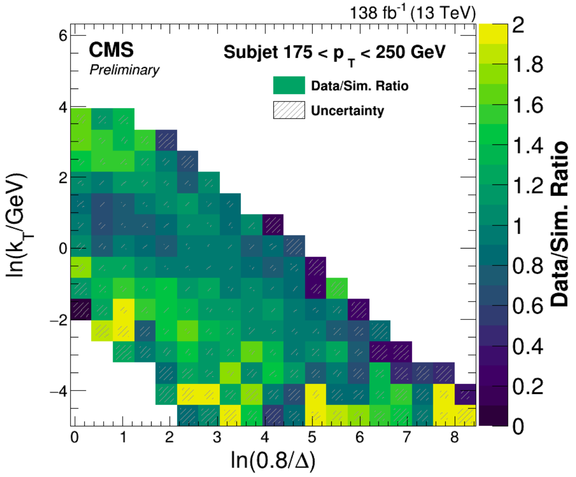
Jet Substructure Modeling

CMS DP-2023/046

Paper for this method on the way!

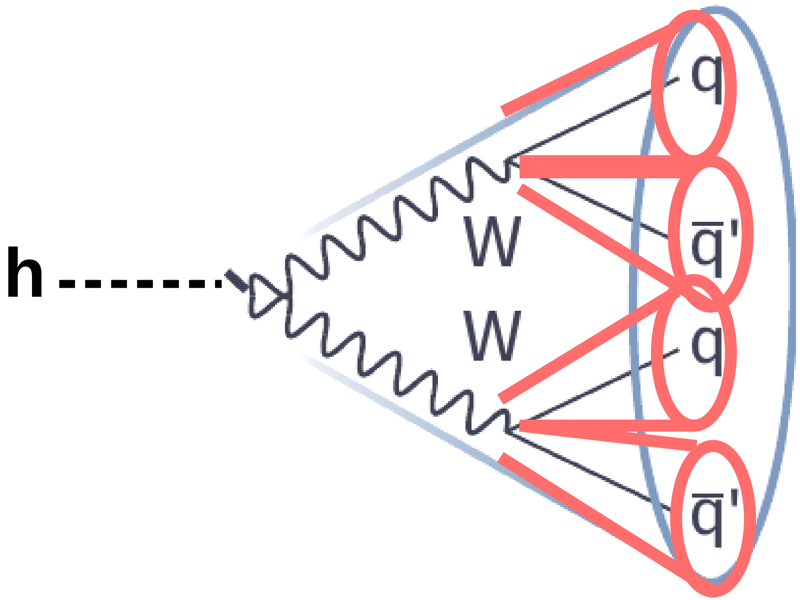
- Dominant uncertainty on anomaly tag from **jet substructure modeling** of signals
- Exotic signals → no SM proxies in data to calibrate
- Developed **new** data-driven method to calibrate **high prong jets**!
  - Recluster prongs into separate subjets
  - Correct modeling of each subjet using Lund Jet Plane
- Gives a correction to modeling & uncertainty





Multi-Prong Calibration Technique

CMS DP-2023/046

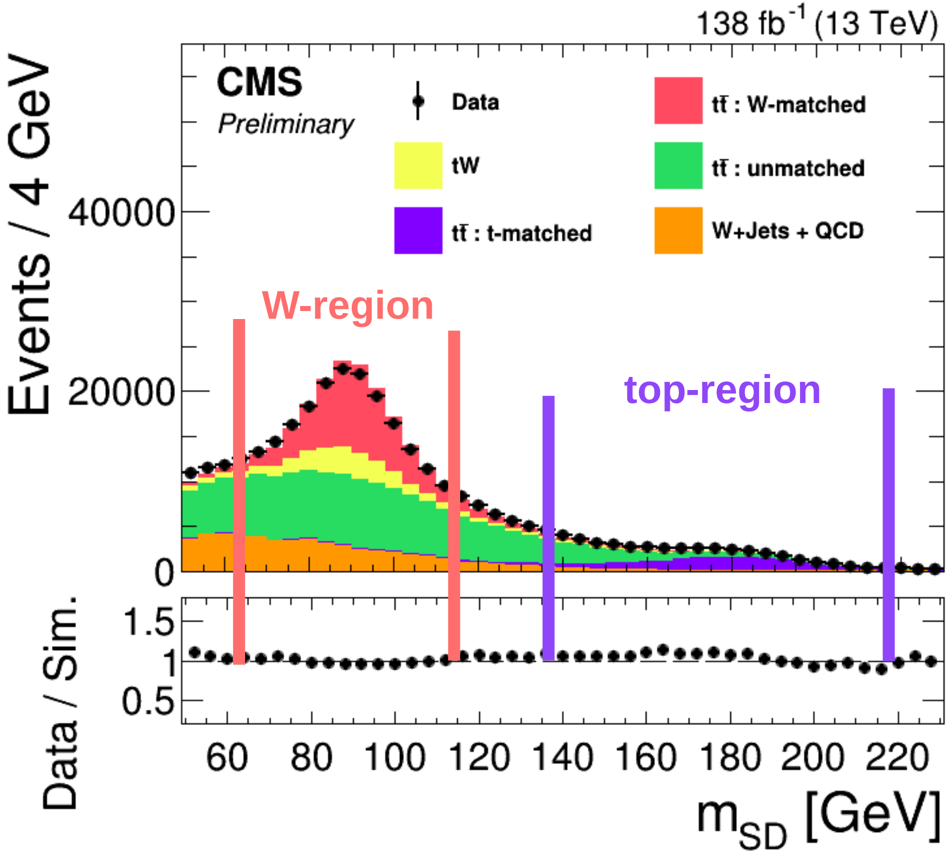


Key assumption : Each prong originates from a SM quark

- **Recluster** AK8 jet so each prong in a **separate subjet**
- Data-driven correction for each **subjet** using the **Lund Jet Plane**
- Correction is 'per-prong' so can extrapolate to higher-prong jets!

Semi-lep. $t\bar{t}$

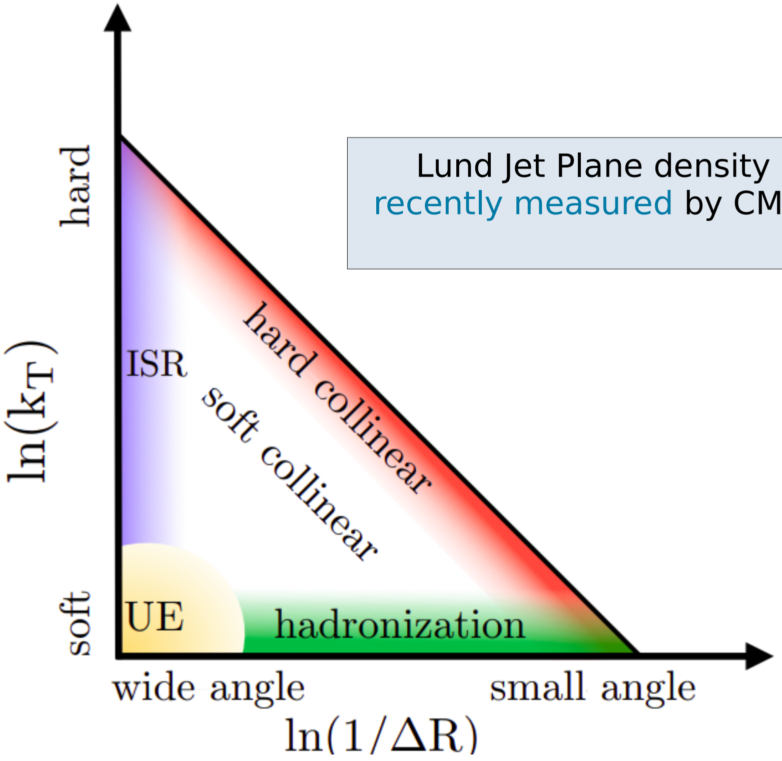
- Extract & test data-driven subjet correction using semi-leptonic $t\bar{t}$ events
- Derive data/sim. ratio of **Lund Jet Plane** from **boosted W's**
- Test calibration on **W's** and **top's**



The Lund Jet Plane (LJP)

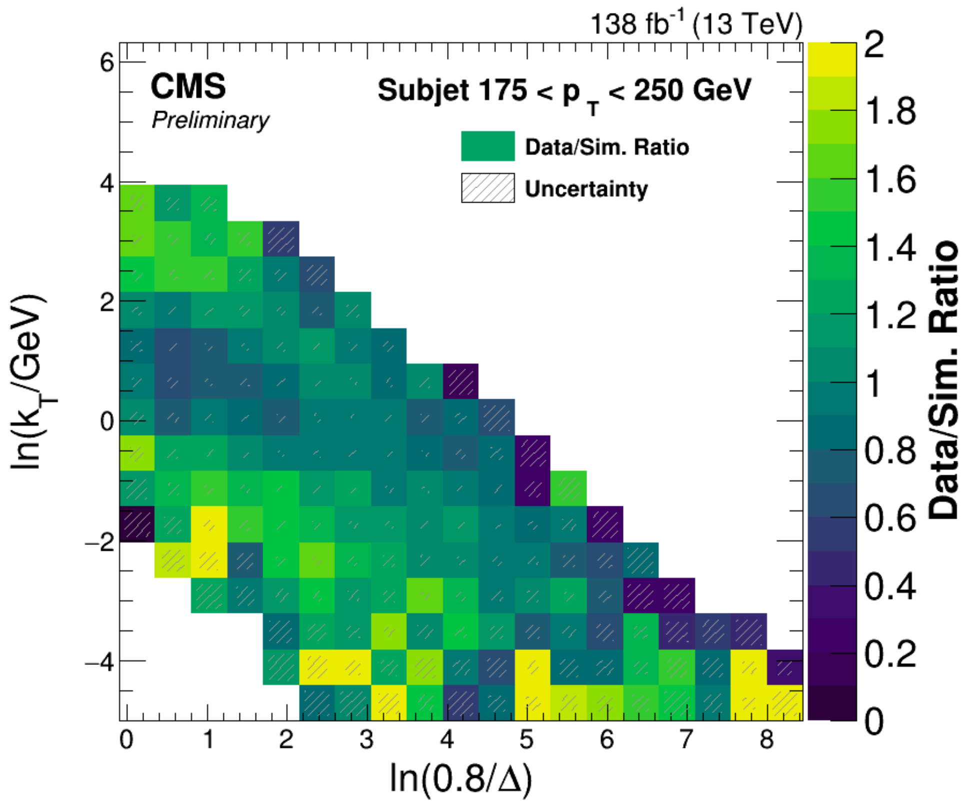
1807.04758

- A 2D representation of the density of splittings inside the jet
- To construct our **subjet Lund Jet Plane**
  - **Recluster** AK8 jet into #prongs using **exclusive kt** algorithm
  - Recluster each subjet using **Cambridge/Aachen** to get splitting history
  - **Fill** points based on **splittings** along hardest branch



Lund Jet Plane density
recently measured by CMS

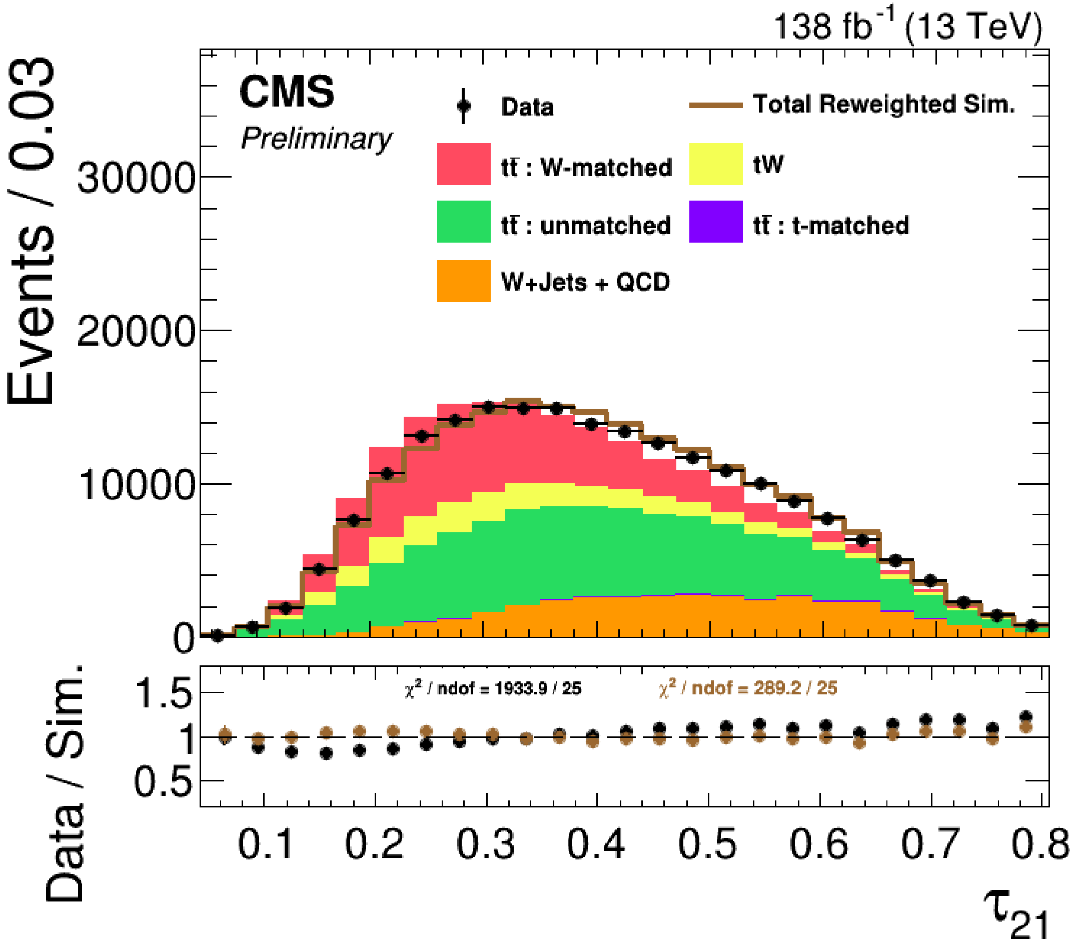
Derivation of LJP Correction



- Recluster AK8 jets from **W-region** into 2 subjets
- Construct **LJP**'s of data and sim. → take ratio
  - Done in 6 bins of subjet p_T
- Use this ratio to **correct simulated jets**
- For each prong, reweight based on the multiplication of the **LJP** ratio of prong's splittings

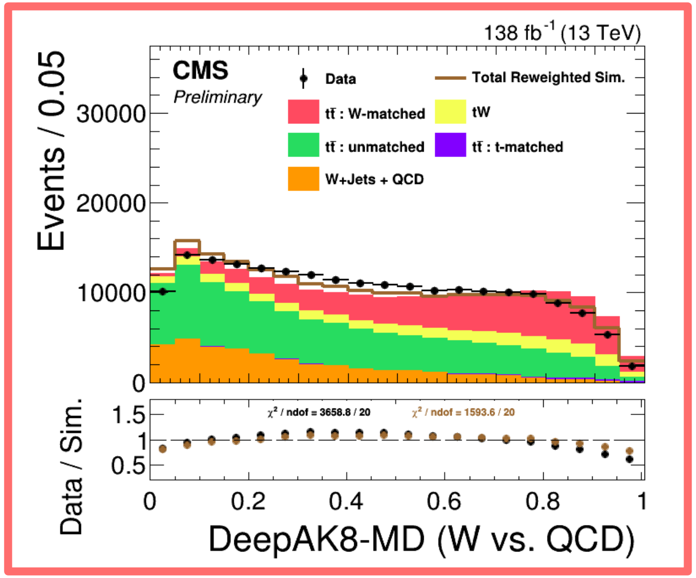
LJP Application to W Jets

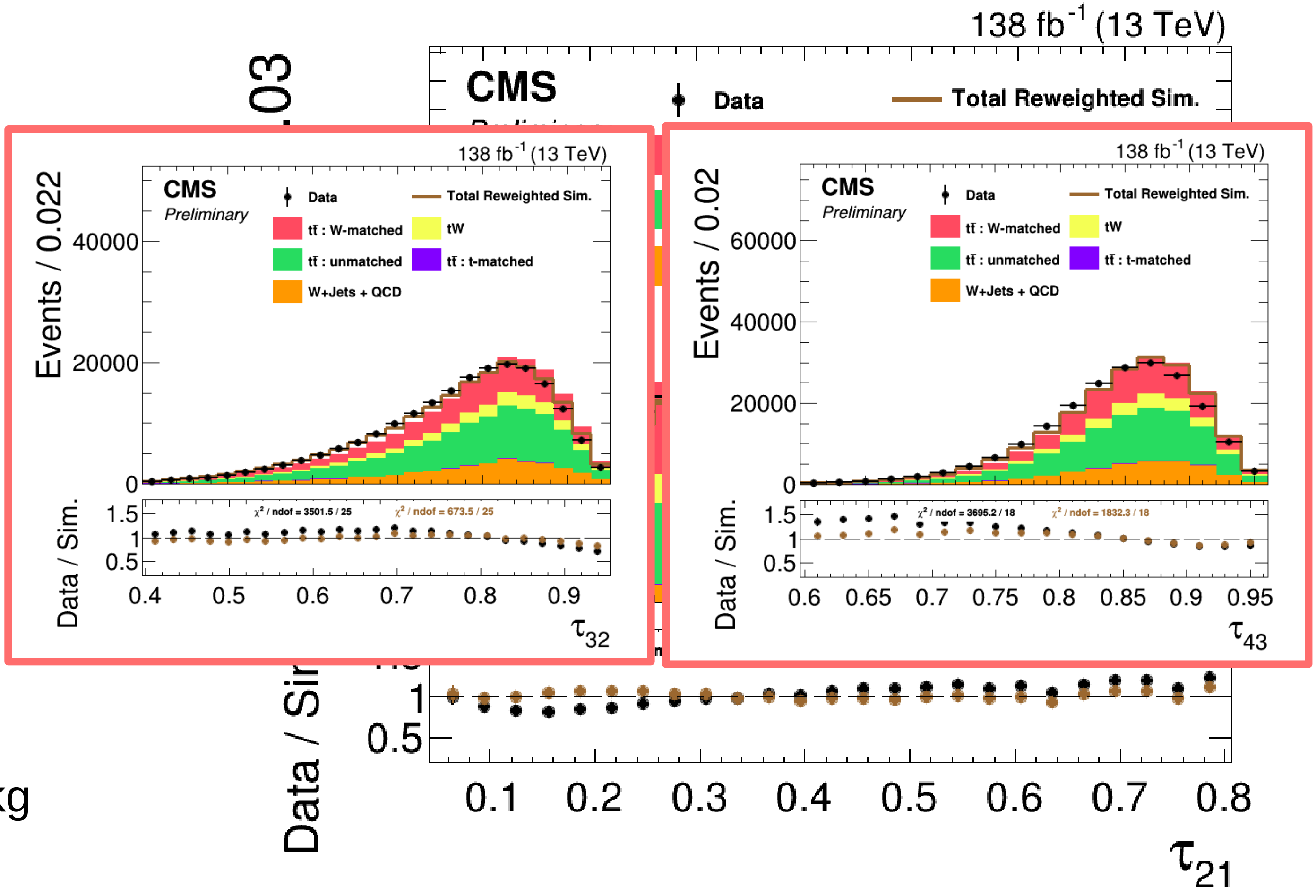
Application of **correction** to **W jets** significantly improves data/sim. agreement!



NB: Non-perfect closure b/c bkg processes are not corrected

LJP Application to W Jets

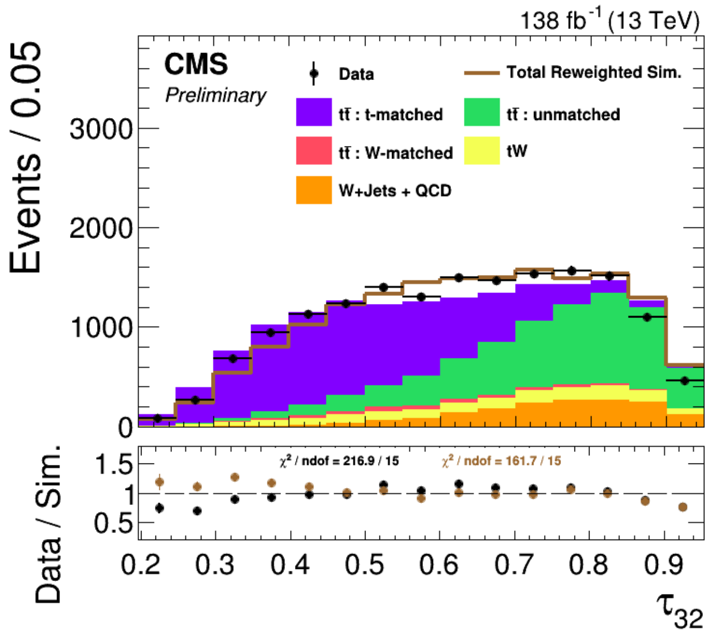


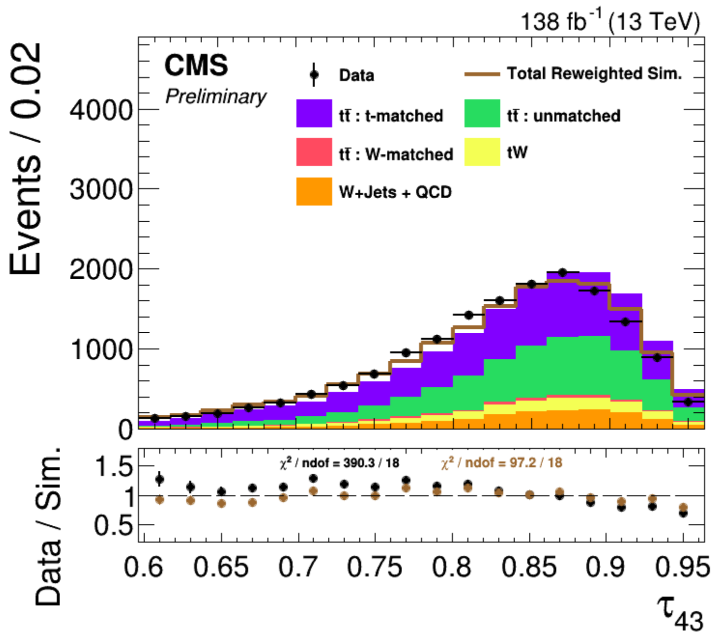


NB: Non-perfect closure b/c bkg processes are not corrected

LJP Application to Top Jets

- Recluster **top jets** into 3 subjets
- Apply data/sim **LJP** correction derived from **W**'s

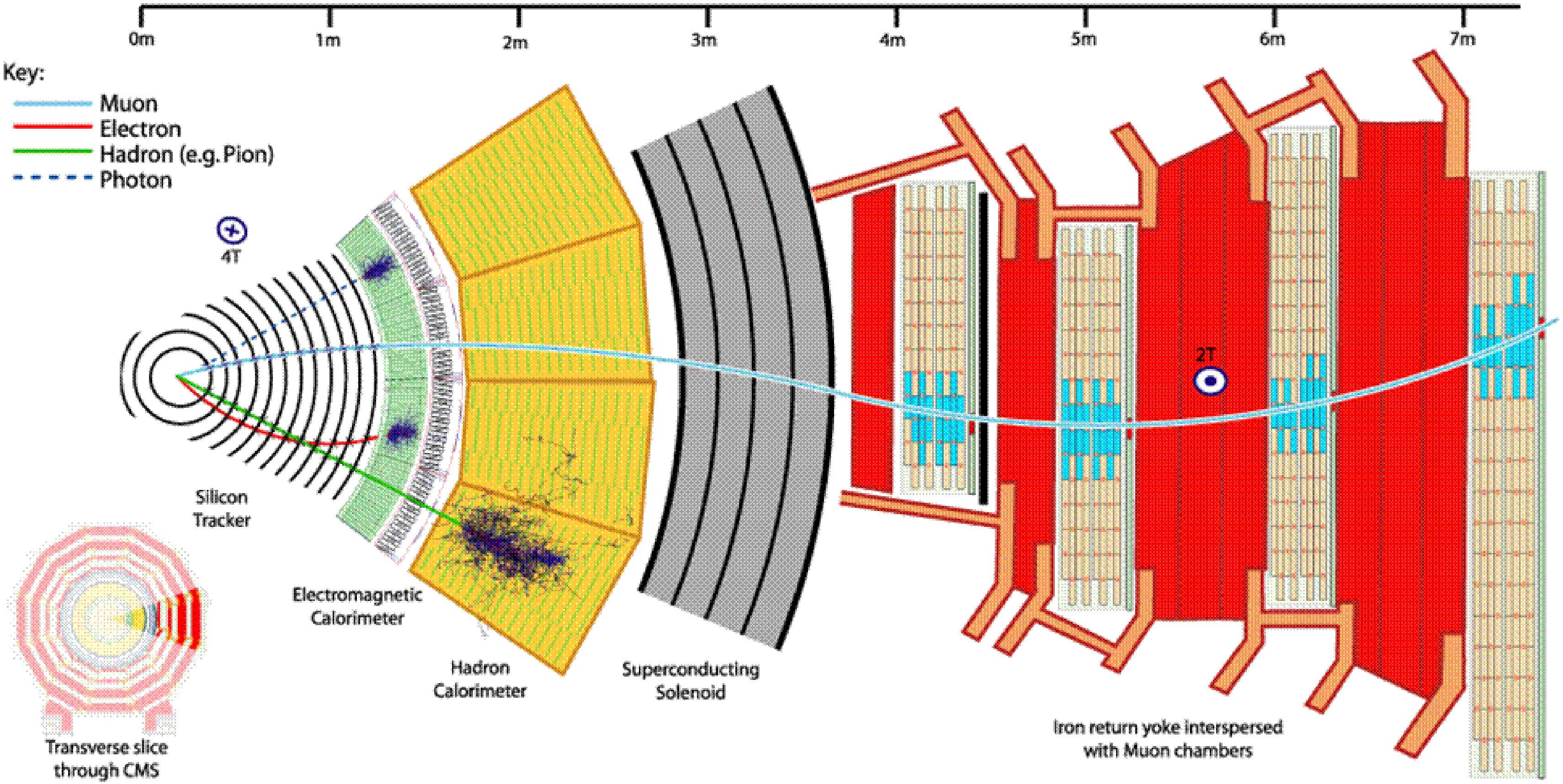




NB: Non-perfect closure b/c bkg processes are not corrected

Correction significantly improves agreement!

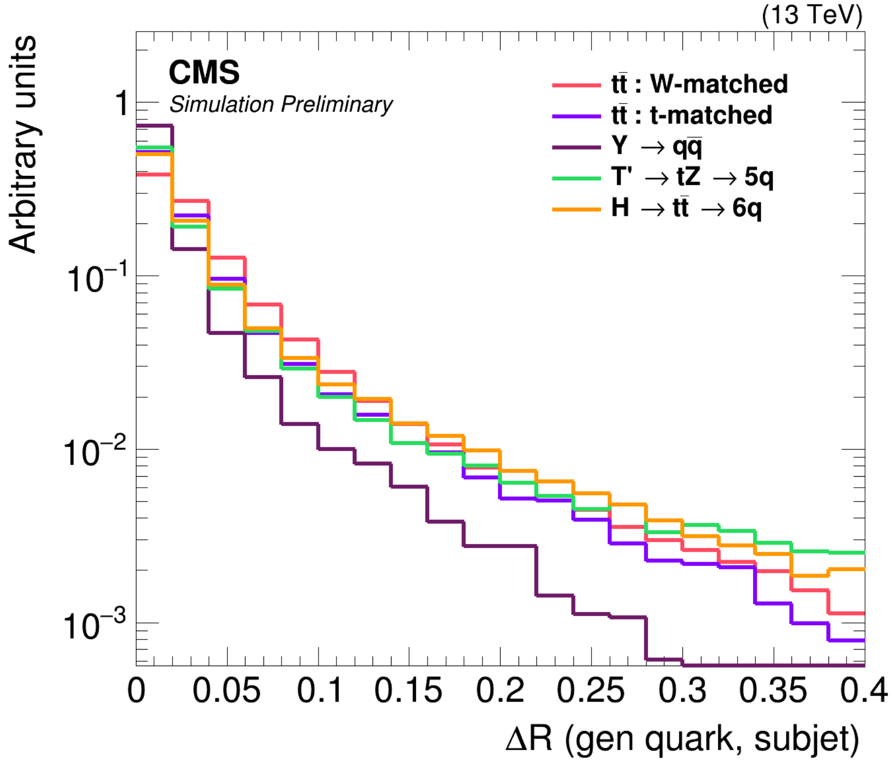
CMS Reconstruction



LJP Uncertainties

- **Stat.** and **sys.** on extraction of data/sim. **LJP** ratio
- **Matching uncertainty** on how well the reclustered subjets correspond to the quarks from the hard process
  - Largest unc., grows with # of prongs
  - ~5% for 2-prong → 50% for 6-prong
- Minor uncertainties:
  - Extrapolation of correction in **subjet p_T**
  - Differences in showering of **bottom quarks** and light quarks

Reclustering works even for **6 prong** jets!

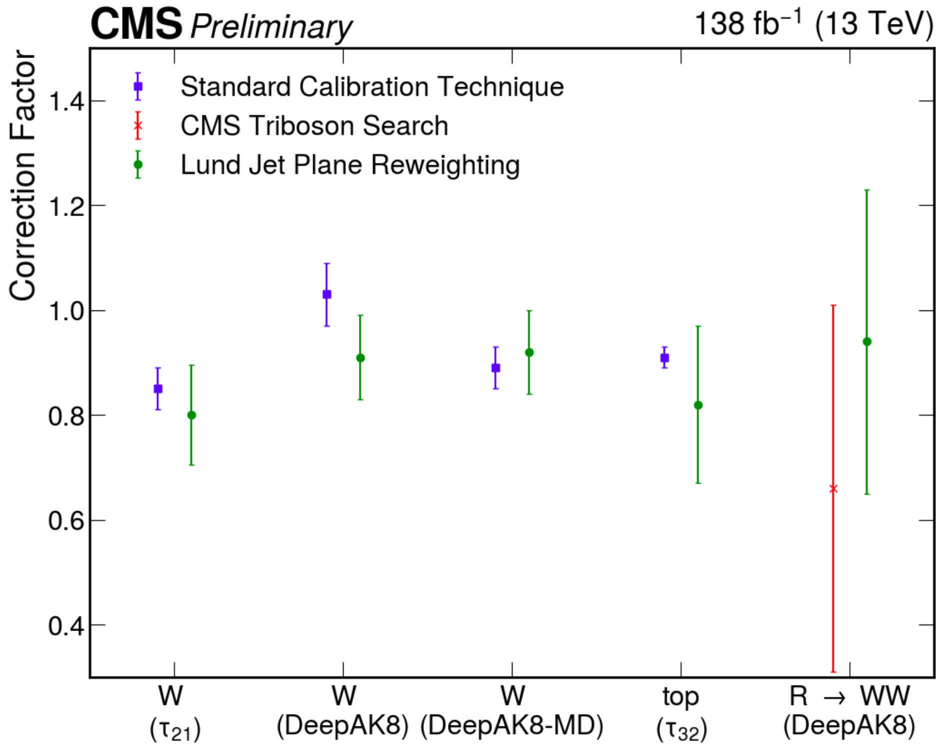


LP Correction Factor Comparison

- Use method to calibrate tagging efficiency
- Compare correction factor (ε_{data} / ε_{sim}) from **std technique** and **LJP reweighting**

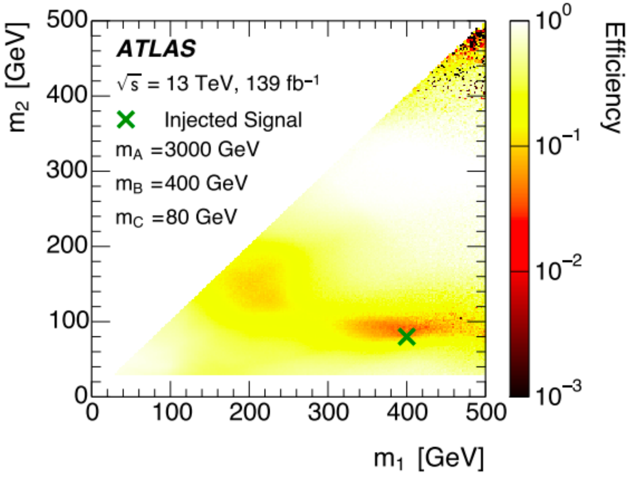
  → Good agreement

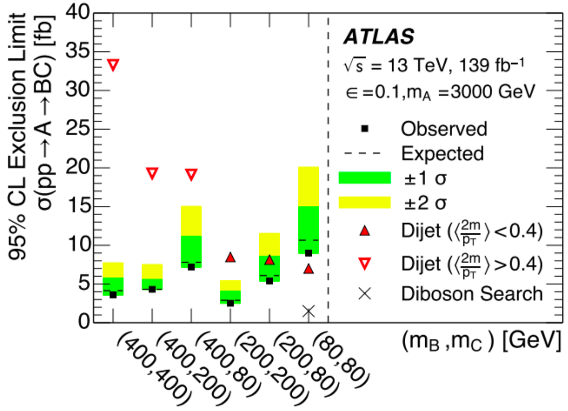
- **LJP** has larger uncertainties b/c more general method
  - BUT enables calibration of high prong jets!



ATLAS CWoLa Hunting

- First anomaly detection analysis
- CWoLa Hunting (per-event instead of per-jet)
- 2D feature space (jet masses)
- Designed limit setting procedure for weak supervision



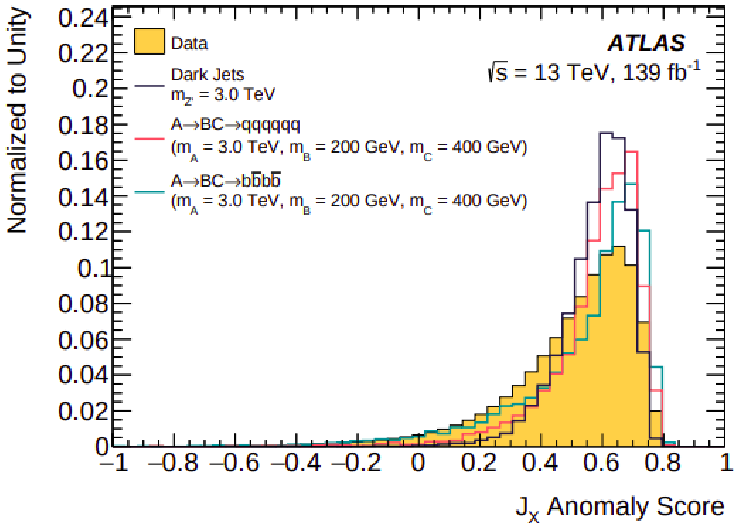


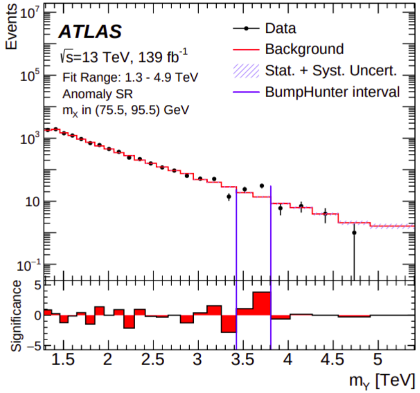
Compared to previous narrow dijet searches

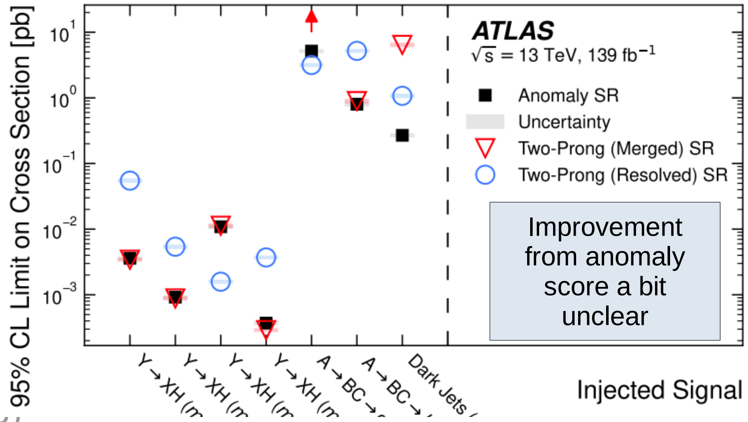
2005.02983

ATLAS Y→ H+X

- Resonance decaying to Higgs + anomalous jet
- Autoencoder-like network used to tag X jet
  - (first unsupervised search)





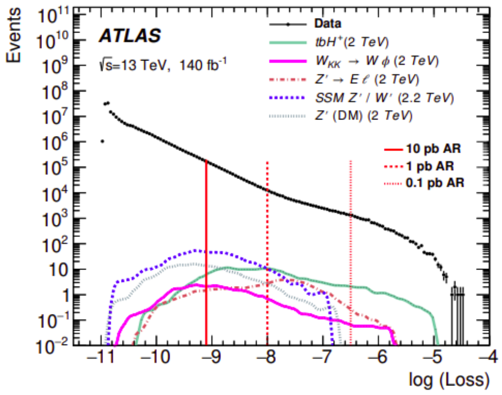


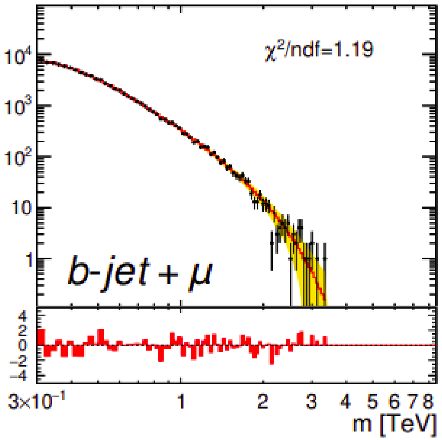
Improvement from anomaly score a bit unclear

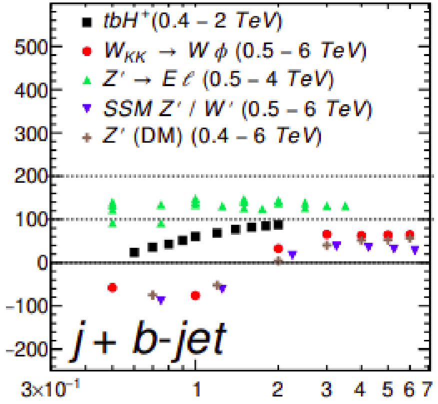
2306.03637

ATLAS Lepton + jet resonances

- Look for resonances with at least 1 lepton
- ‘Event level’ anomaly detection
  - Autoencoder trained on uses ‘rapidity mass matrix’







Nice demonstration anomaly cut improves sensitivity

2307.01612