Anomaly detection for new physics searches in dijet events at CMS

CMS-EXO-22-026

Mar 20, 2024

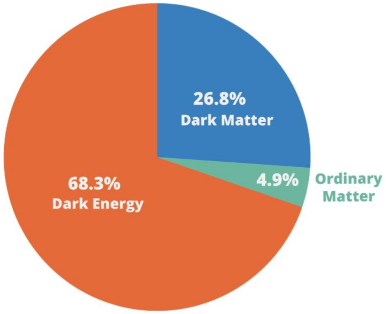
Benedikt Maier, Imperial College London

Many open questions in cosmology and particle physics

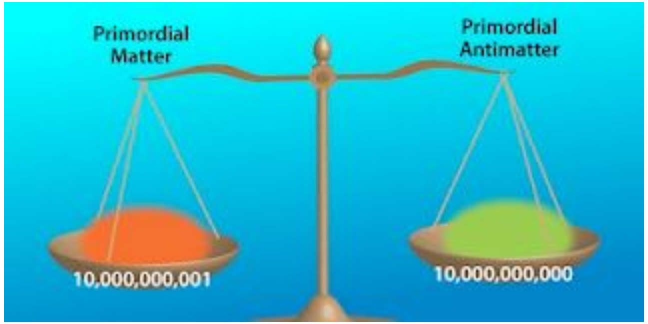
Experiment-driven:

Dark Matter & Dark Energy

Matter-antimatter asymmetry



michaelgstrauss.com



Theory-driven:

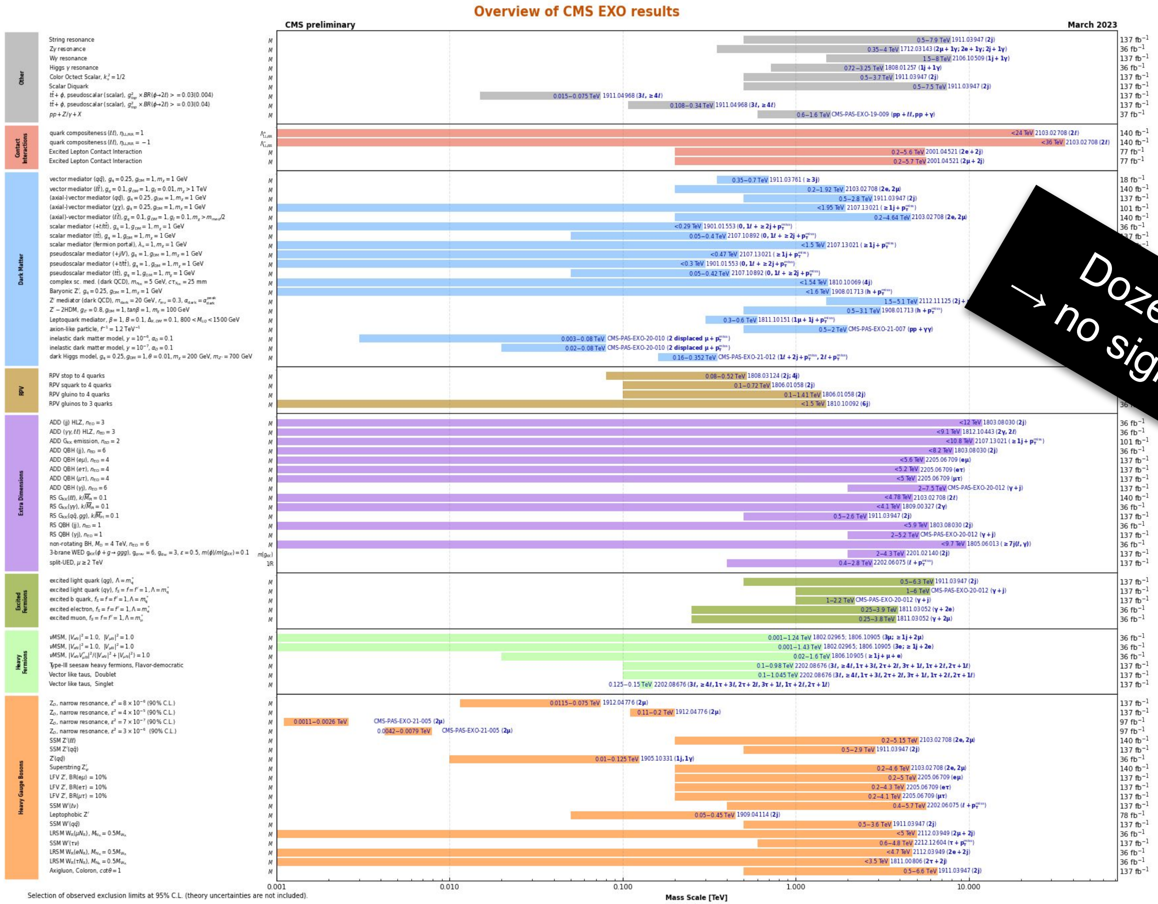
Hierarchy problems (weakness of gravity, fine tuning at level 10^{16})

Why 3 generations of fermions?

→ New physics within LHC reach?

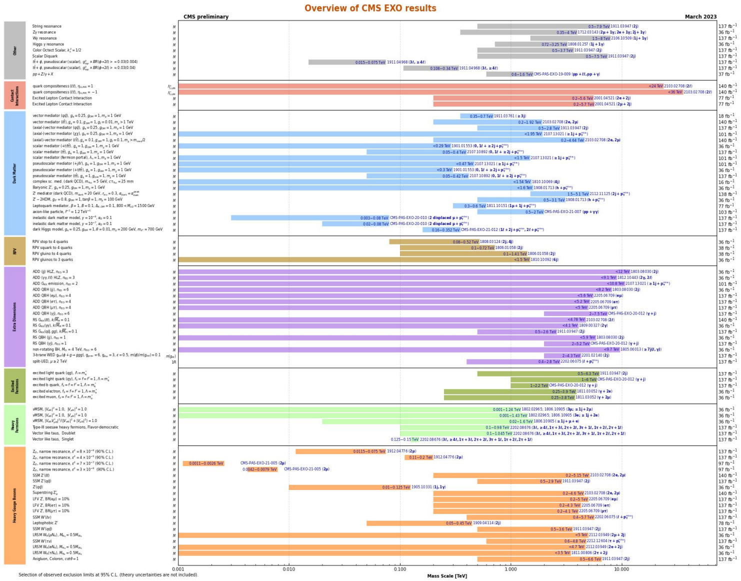
quantumdiaries.org

CMS established rich search program for new physics



Dozens of searches → no sign of new physics

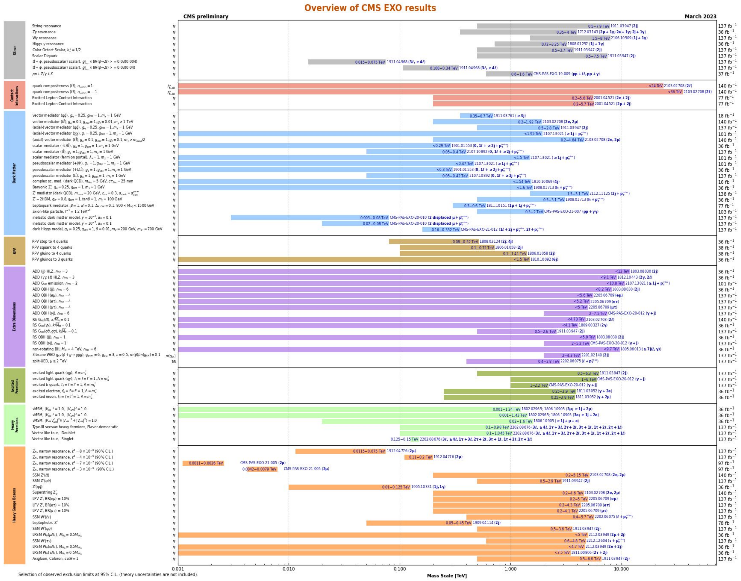
Are we searching for the wrong signatures?



Maybe looking in the wrong spots or for the wrong models?

→ Need **safeguard** against missing signs of new physics

Are we searching for the wrong signatures?



Maybe looking in the wrong spots or for the wrong models?

→ Need **safeguard** against missing signs of new physics

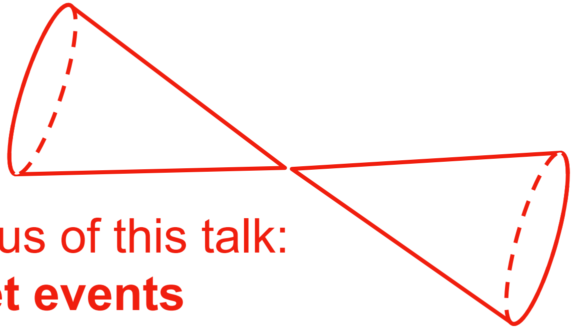
Re-formulate the question

“Does this event look like BSM theory XYZ?”

↓

“Does this event look like the Standard Model?”

→ Anomaly detection



Focus of this talk:
Dijet events

Outline

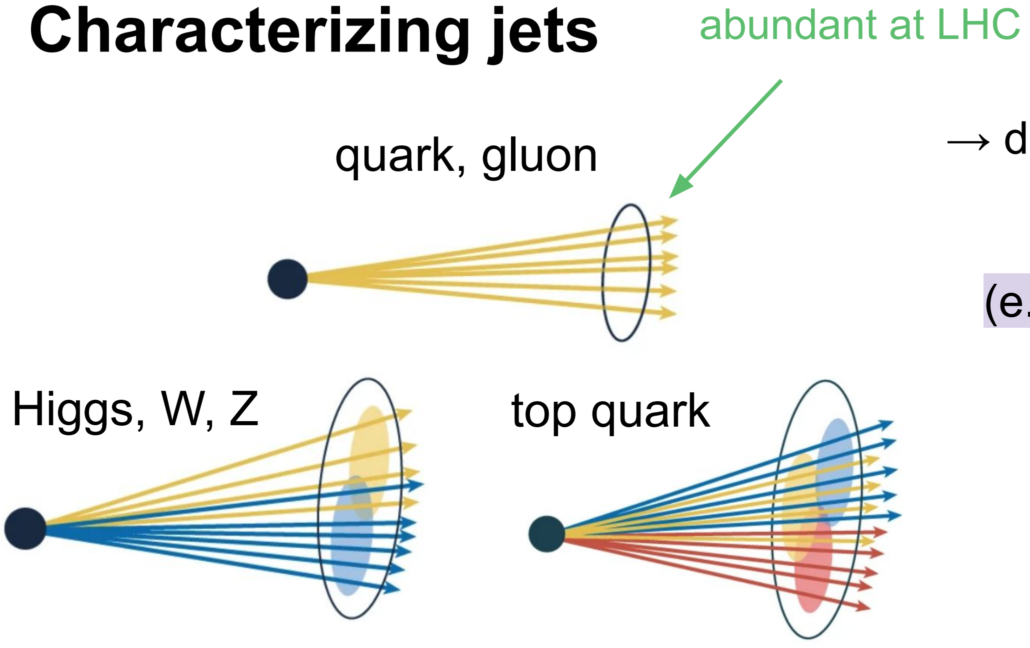
1. What are jets at CMS

2. Anomaly hunting

3. Expected performance

4. Actual look into data

Characterizing jets

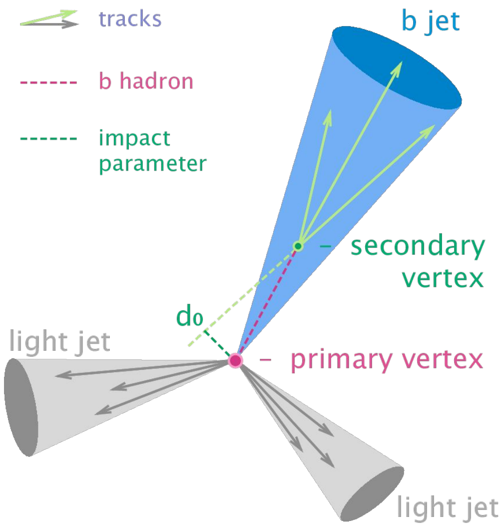


Different particles → different substructure

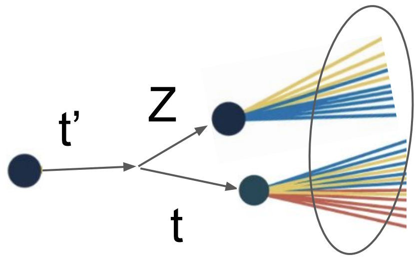
→ Exploit substructure
(e.g., N-subjettiness τ_{21}, grooming algorithms)

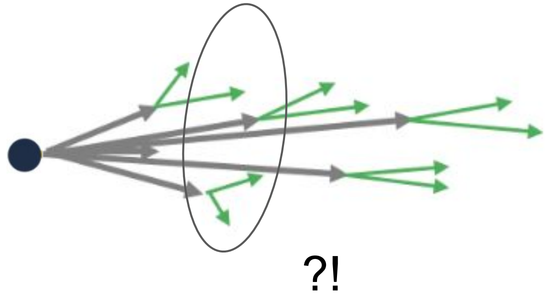
B hadrons long-lived
→ displaced "secondary" vertex within jet

→ Exploit flavor content
(e.g., multivariate B tagging algorithms)



Anomalous jets





- **No assumption** on how exotic jets look like, but they could have:
  - long decay chains of SM resonances
  - completely exotic particles & weird radiation patterns

- Want our algorithm as model-independent as possible

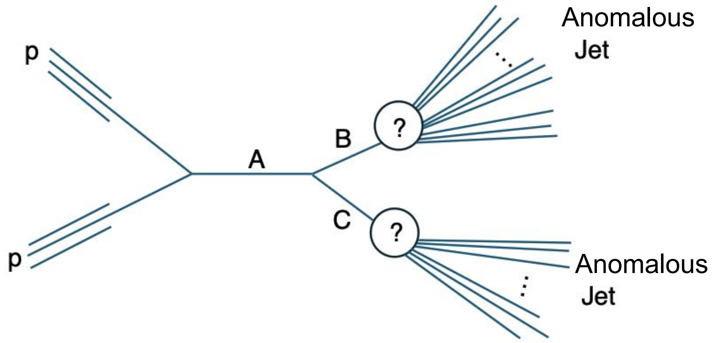
Model independence:

1.) Not assuming specific signal model
2.) Not relying on imperfect background model (→ QCD Monte Carlo) either

→ **Train directly on data!**

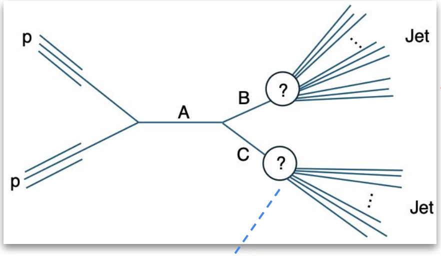
Searching in dijet topology

- Target narrow resonance A → BC, B & C decay hadronically
- Goal is to be sensitive to broad range of possible A,B,C
- Assumption: $m_A >> m_B, m_C$
  - Results in highly boosted B, C
  - Decay products contained in large-radius jets
  - → Anomalous large-radius jet on either side



- Select events with at least two AK8 jets
  - $p_T > 300$ GeV, $|\eta| < 2.5$
- Dijet invariant mass $m_{jj} > 1455$ GeV
  - Dictated by trigger turn-on
- $\Delta\eta_{jj} < 1.3$ (to target *s*-channel resonance)

Developing and testing with a suite of BSM signals



Jet B substructure

Jet C substructure

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

Expect sensitivity to many additional kinds of signals!

Search probes unexplored signals

Outline

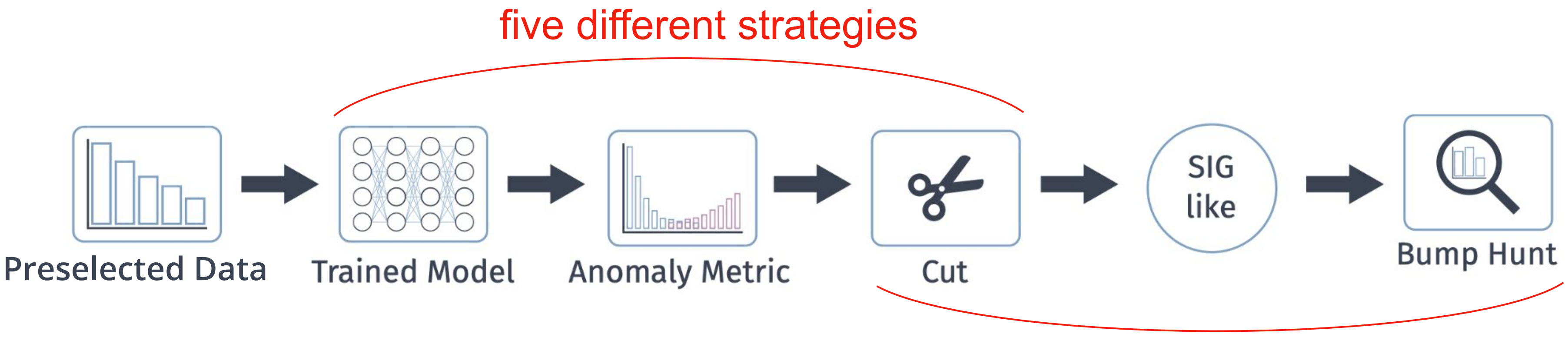
1. Characterizing jets at CMS

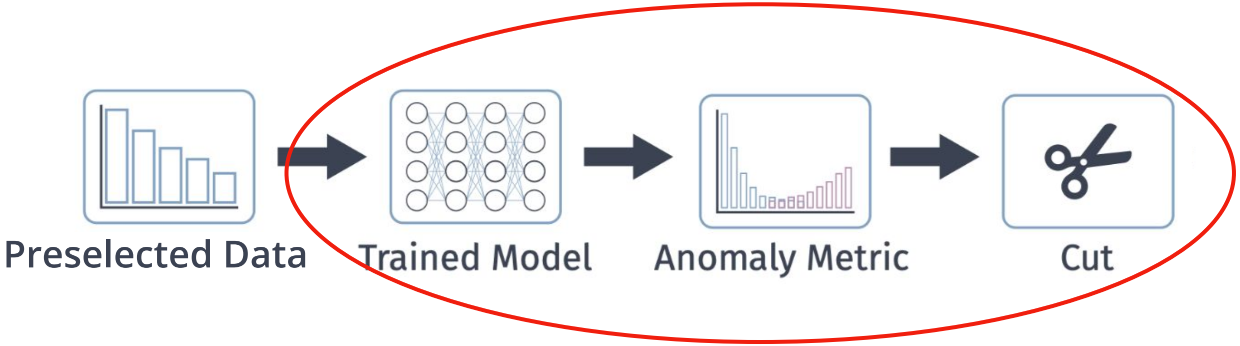
2. Anomaly hunting

3. Expected performance

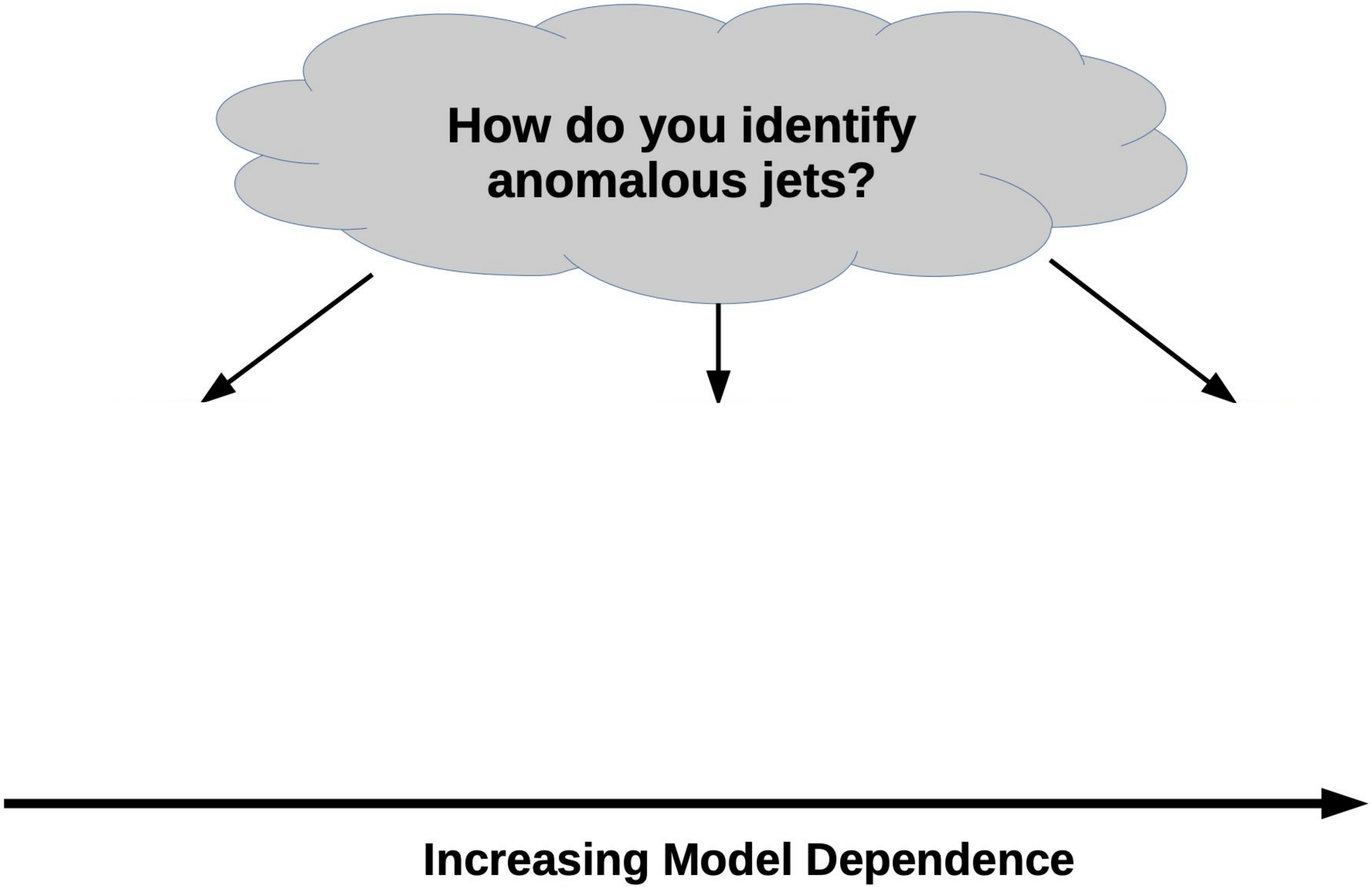
4. Actual look into data

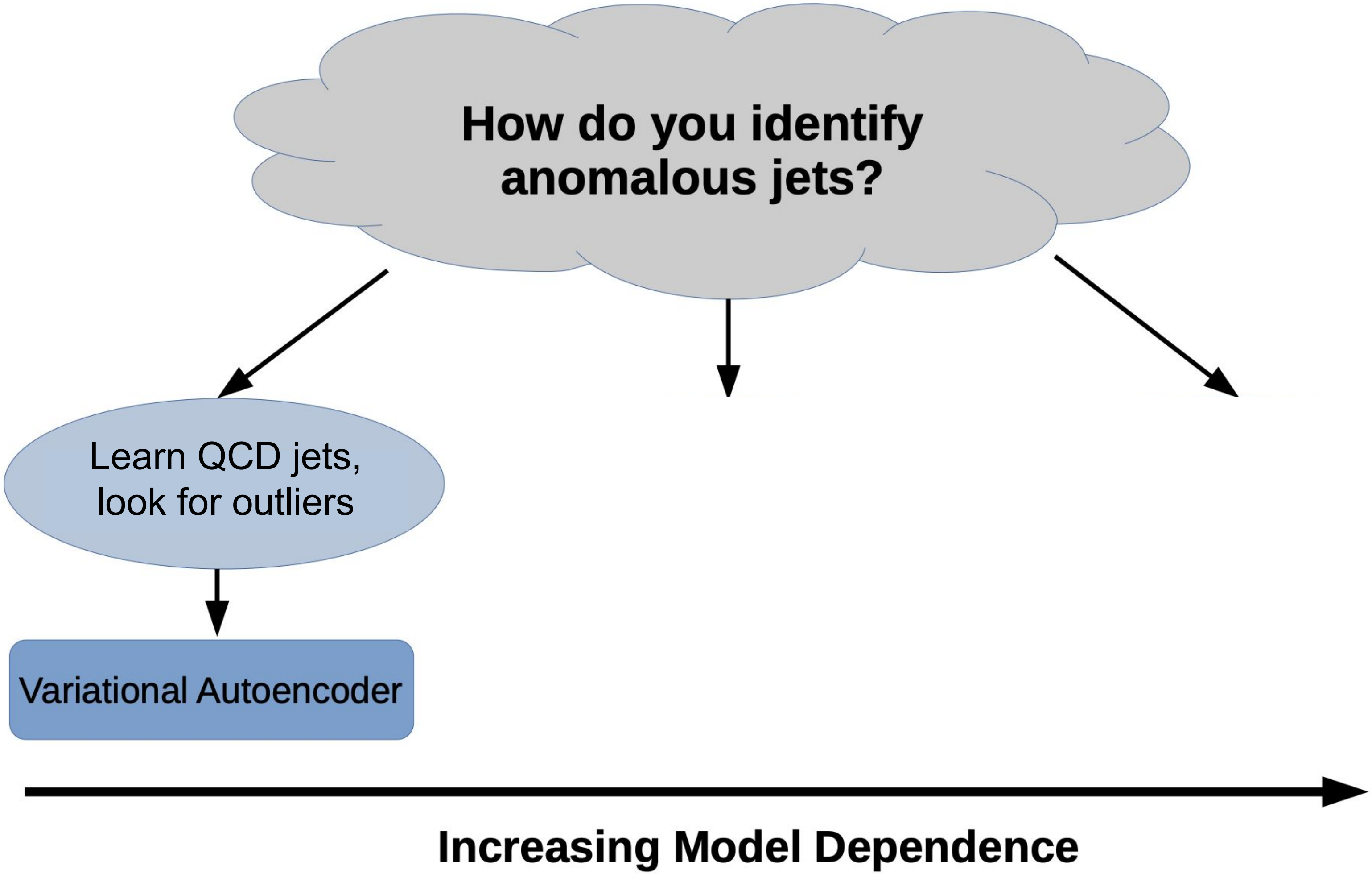
Analysis Strategy



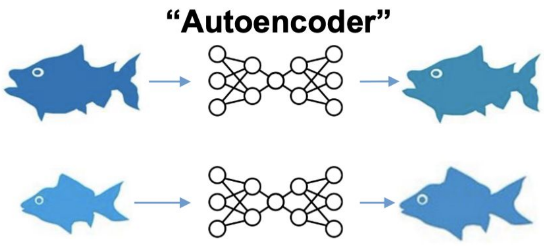


Preselected Data
Trained Model
Anomaly Metric
Cut

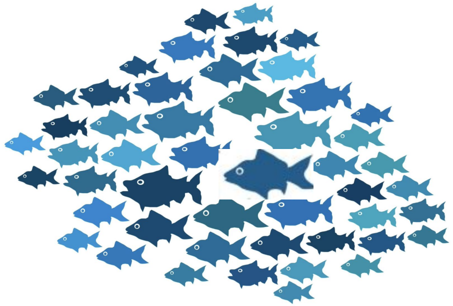

How do you identify anomalous jets?
Increasing Model Dependence


How do you identify anomalous jets?
Learn QCD jets, look for outliers
Variational Autoencoder
Increasing Model Dependence

Autoencoder Fundamentals



training examples

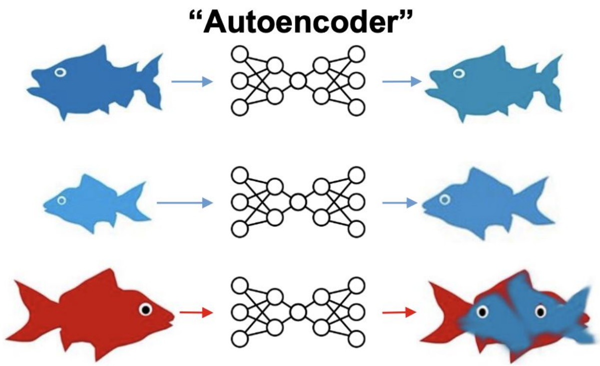


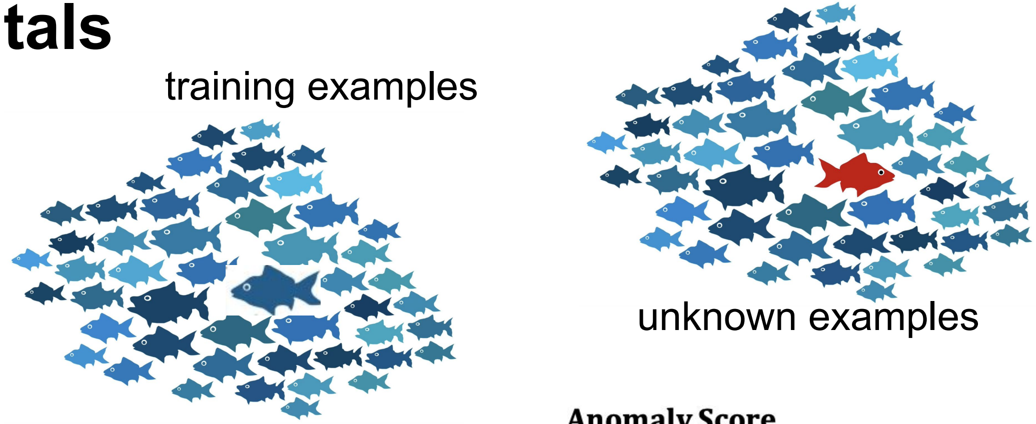
- Train on non-anomalous examples, learns to reconstruct input

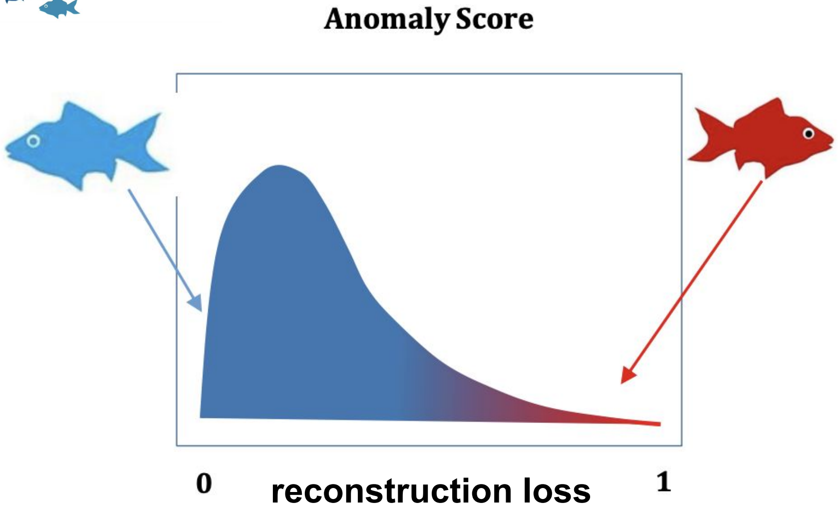
- Force information through a bottleneck
  - Focus on core features of normal examples

Illustrations: J Gonski, A Kahn

Autoencoder Fundamentals







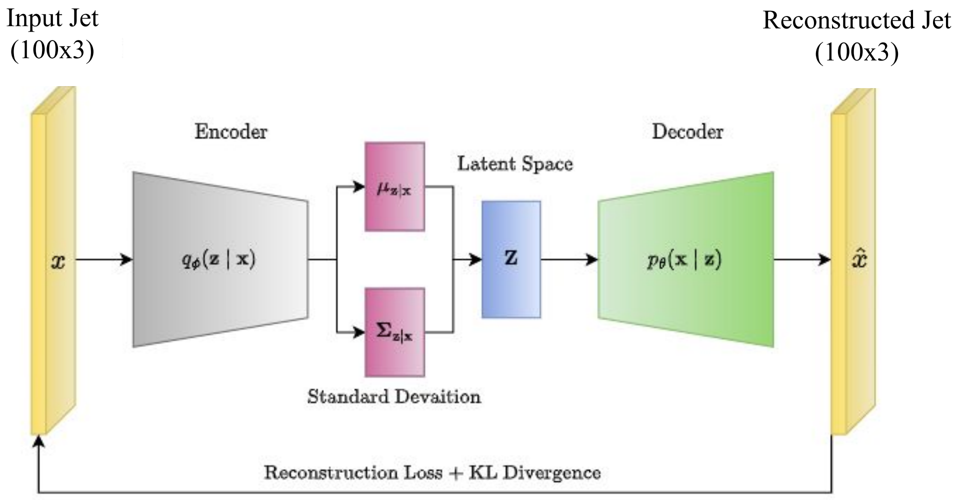
- Train on non-anomalous examples, learns to reconstruct input

- Force information through a bottleneck
  - Focus on core features of normal examples

→ Fails at reconstructing exotic examples

Illustrations: J Gonski, A Kahn

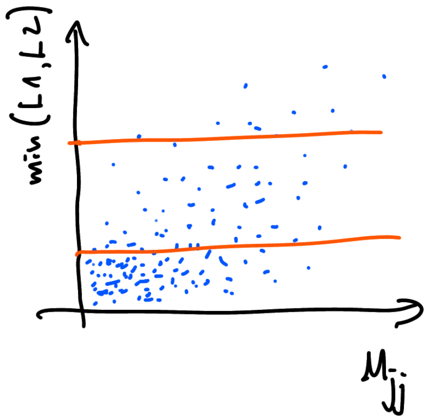
Variational Autoencoder (VAE)



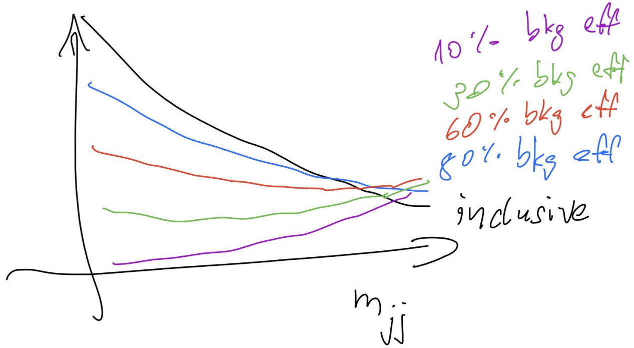
- VAE encouraging (through additional loss term) latent space to be Gaussian

- Trained on jets from signal-depleted $\Delta\eta_{jj}$ sideband
  - $\Delta\eta_{jj}$ > 1.4
- Jet represented by 100 highest-p_T constituents: p_x, p_y, p_z
- Constituents sorted by clustering sequence
  - More expressive in terms of substructure
- 100 x 3 matrix processed with 2D and 1D convolutional layers, bottleneck dimension = 12
- Anomaly metric: min(loss j1, loss j2)

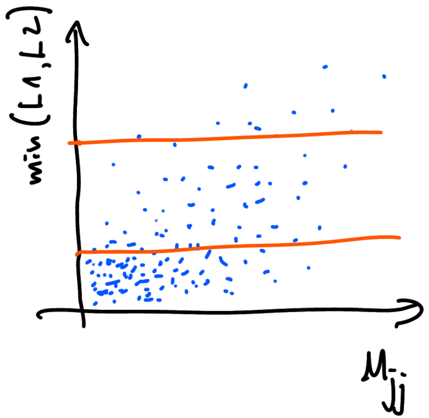
Variational Autoencoder: Decorrelation from dijet mass



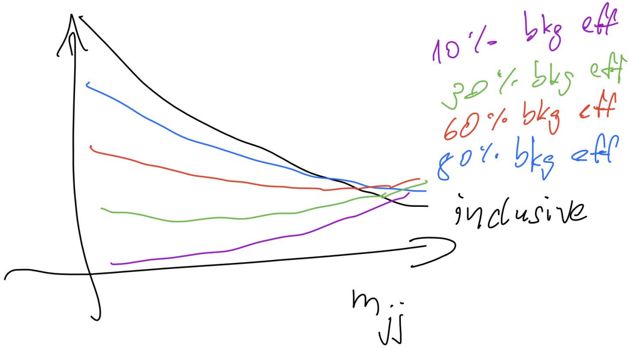
trivial cuts on min(L1,L2)
result in background sculpting

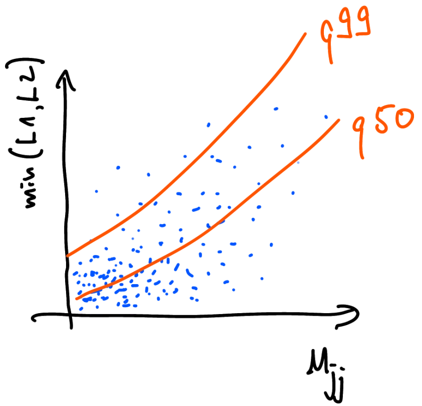


Variational Autoencoder: Decorrelation from dijet mass



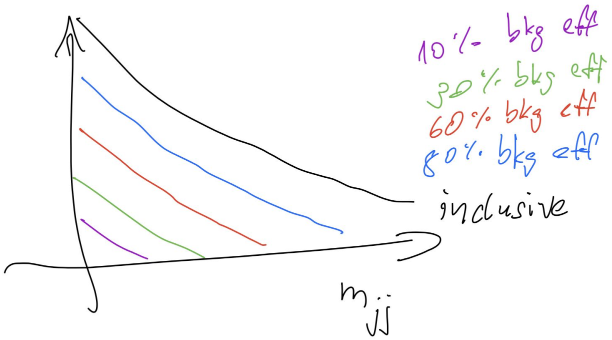
trivial cuts on min(L1,L2)
result in background sculpting

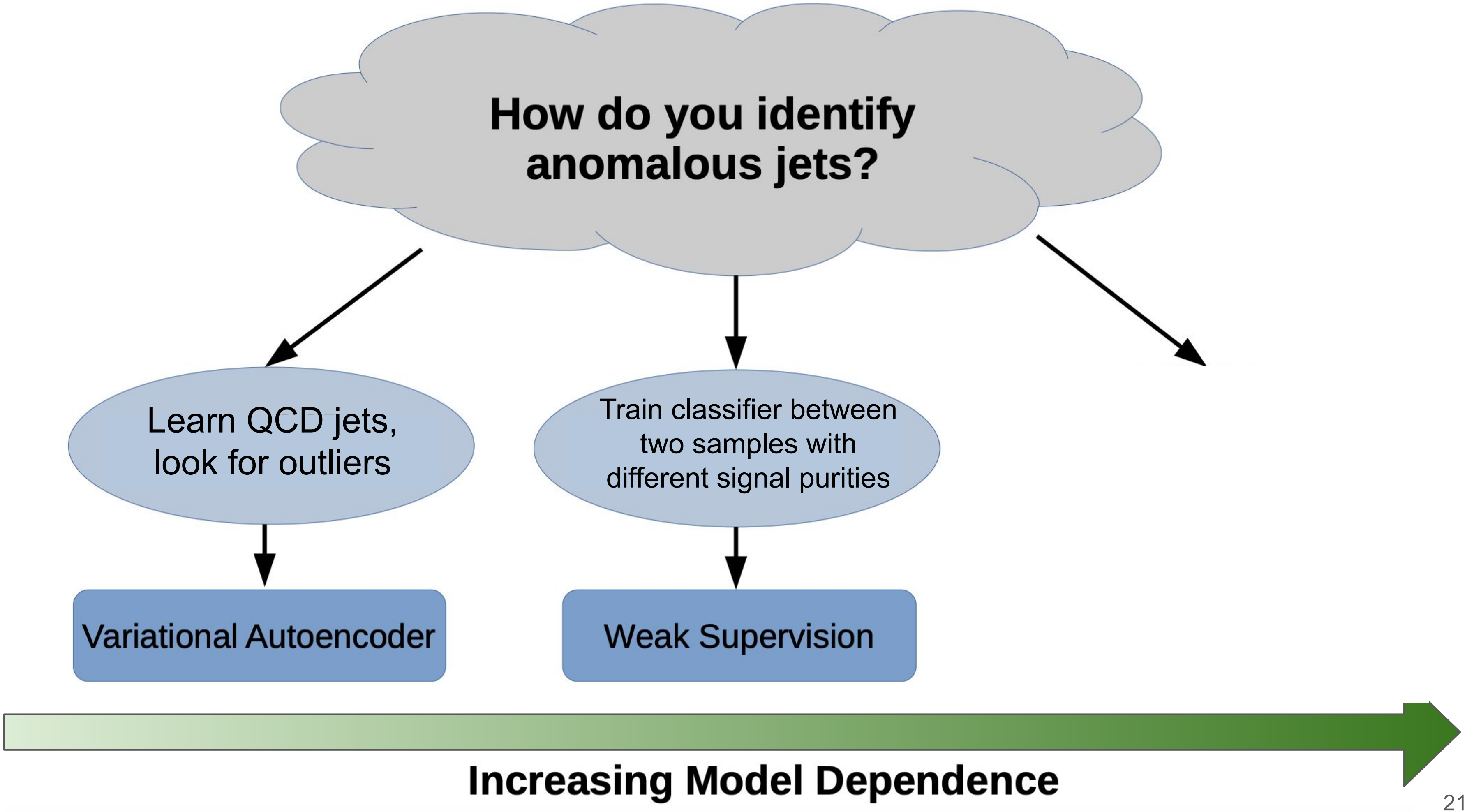




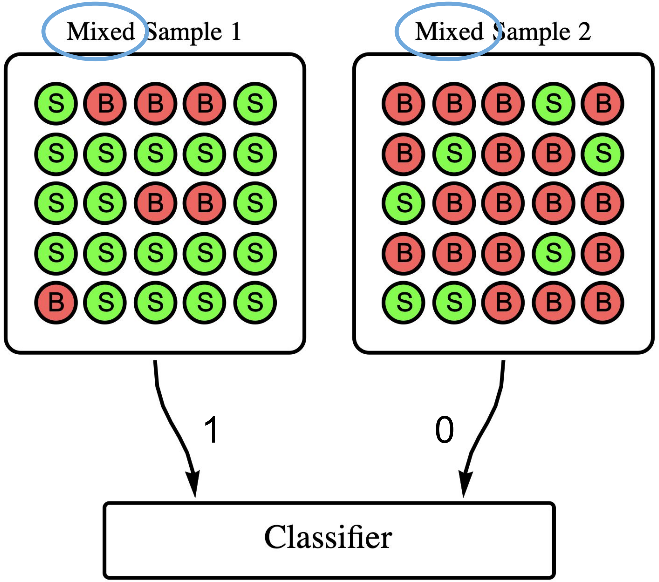
train NN to regress quantiles
corresponding to fixed efficiencies

leaves background unsculpted




How do you identify anomalous jets?
Learn QCD jets, look for outliers
Train classifier between two samples with different signal purities
Variational Autoencoder
Weak Supervision
Increasing Model Dependence

Weak Supervision Fundamentals



JHEP 10 (2017) 174

- Two mixed samples of signal and background events **with different purities**

- Train classifier on the two samples
  - Learns to distinguish signal vs background
  - Bias from also learning different background shapes if they are different

- Higher signal fraction → better classifier performance

3 weakly supervised methods presented here
→ Differ in how they construct the mixed samples

Weak Supervision #1: CWoLa

JHEP 10 (2017) 174

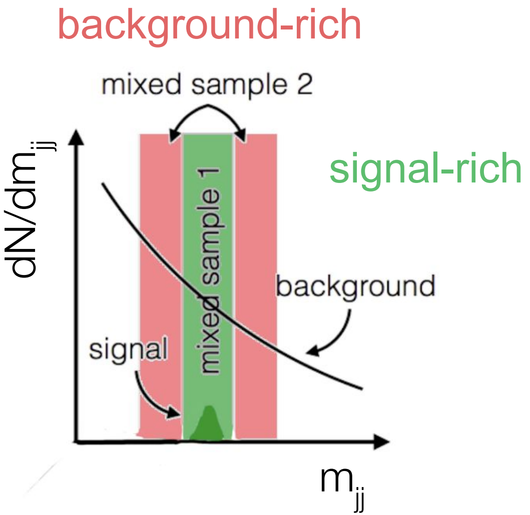
- **Assume** signal is a narrow resonance
  → narrow peak, **choose** a mass window accordingly
  - Signal-rich sample = events from window in m_{jj}
  - Background-rich sample = events from sidebands

- Train a NN classifier on jets from signal-rich sample vs. background-rich sample

- Must ensure NN classifier does not learn m_{jj}

→ Train separate classifiers for heavier and lighter jet*
  - Allows to reweight jets in p_T so that distributions are identical between two samples
  - Avoids learning m_{jj} through jet p_T

→ Event anomaly score: max(score j1, score j2)



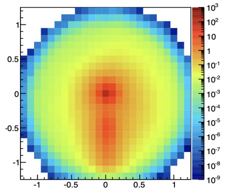
* modification w.r.t. original paper

Weak Supervision #2: Tag N’ Train (TNT)

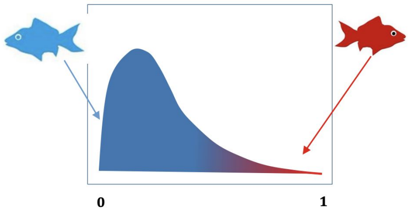
JHEP 01 (2021) 153

- Similar to CWoLa, but **enhance** signal purity of Mixed Sample 1 **with autoencoder**
  - Autoencoder: 2D CNN trained on jet images from jets in sideband

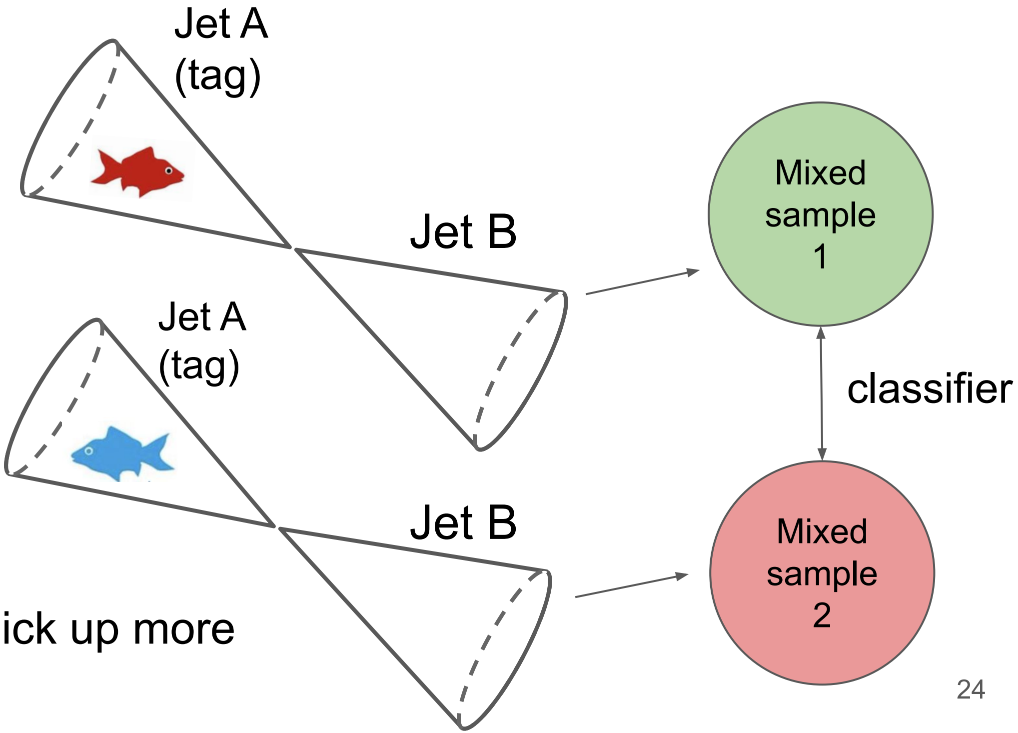
- Assumes both jets are anomalous



Sideband jet images + autoencoder



Apply AE to signal region jets → reconstruction loss



- Better starting point let’s classifier pick up more easily on signal jets

Weak Supervision #3: CATHODE

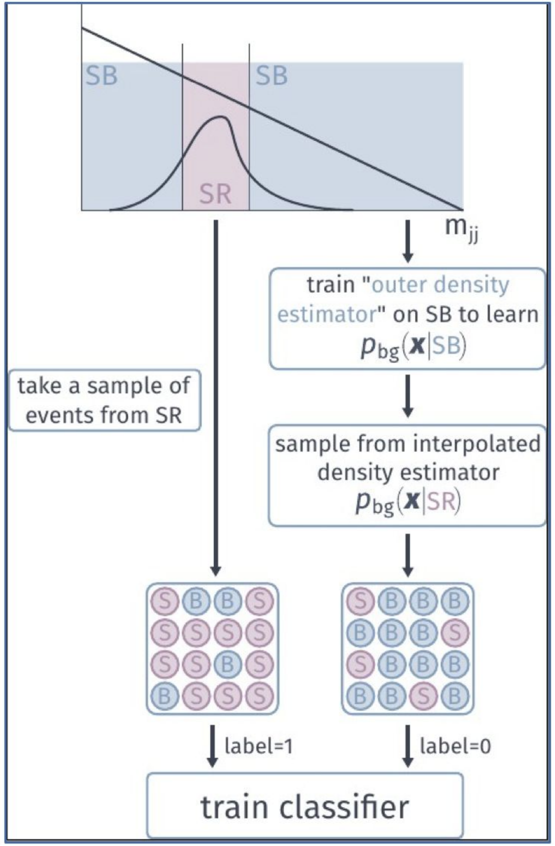
Phys.Rev.D 106 (2022) 5, 055006

- Instead of using sidebands for Mixed Sample 2, **interpolate** their distributions into signal region
    - More robust to feature correlations with m_{jj}

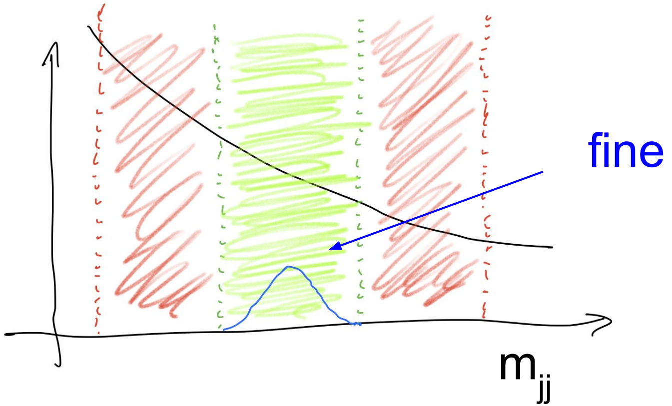
- Train normalizing flows to learn **density** $p_{data}(x \mid m_{jj})$ of features x in sidebands, conditioned on m_{jj}
    - Results in bijective, invertible map $f(x;m_{jj})$

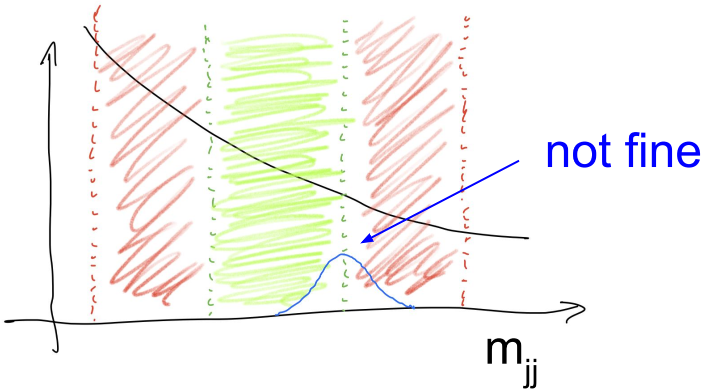
- **Interpolate** to get density in the signal window

- Draw samples in signal window to construct a **synthetic** background sample in SR → Mixed Sample 2
    - Proceed as usual (train classifier Mixed Sample 1 vs. Mixed Sample 2)



Weak Supervision: Training Quirks



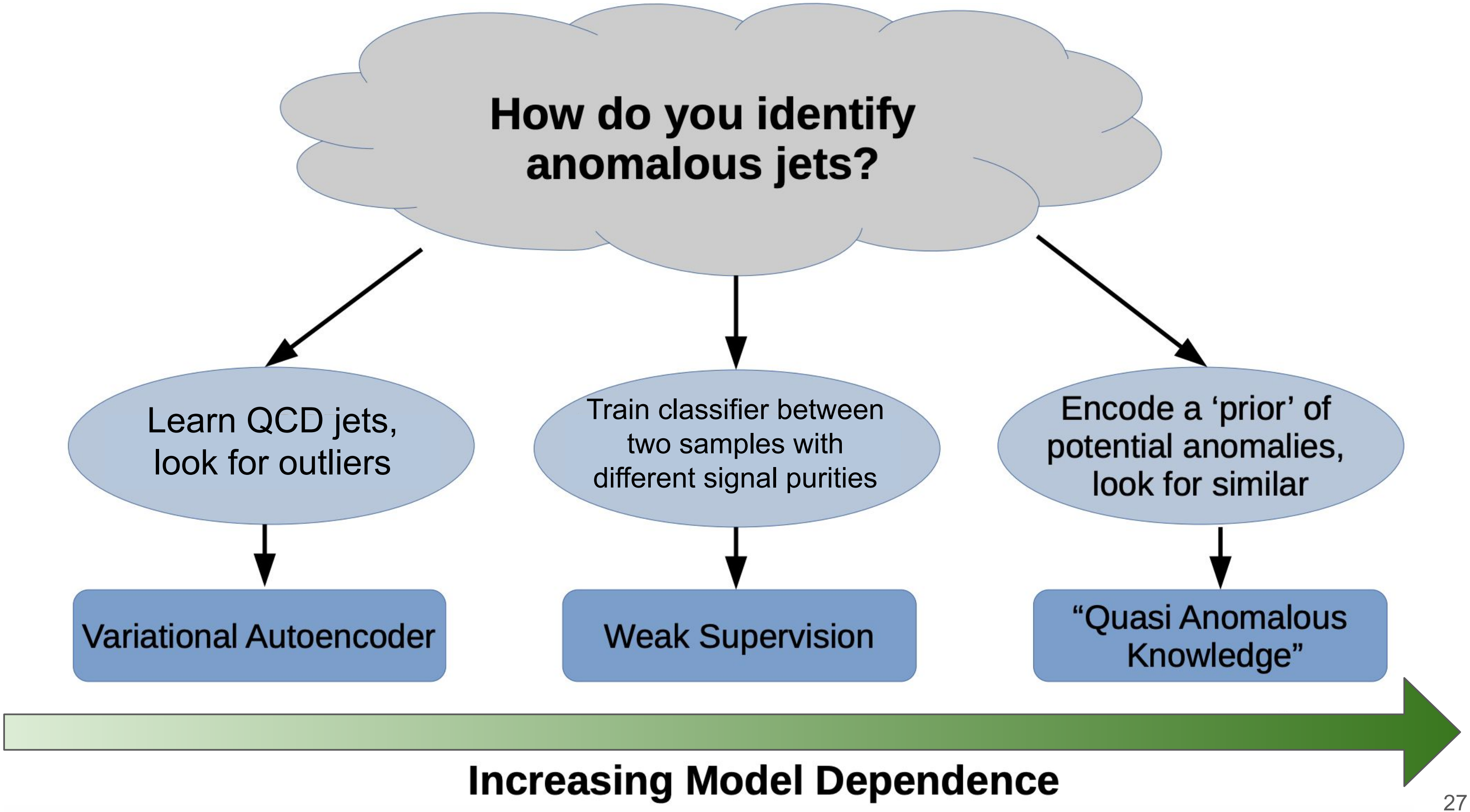


- Weakly supervised methods assume signal window for training
  - Fine if signal well within window, not fine if at the edge
  - Need to **slide window** and **repeat trainings** to cover full mass range

- Define two sets of bins, A and B

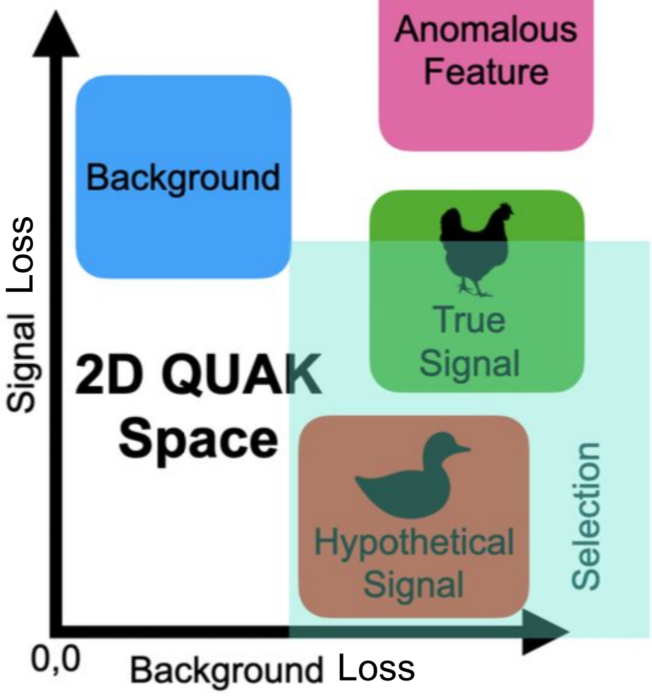
- Set B is shifted by half window w.r.t. Set A

- In total 12 signal regions, different trainings and event selection for each one


How do you identify anomalous jets?
Learn QCD jets, look for outliers
Train classifier between two samples with different signal purities
Encode a 'prior' of potential anomalies, look for similar
Variational Autoencoder
Weak Supervision
"Quasi Anomalous Knowledge"
Increasing Model Dependence

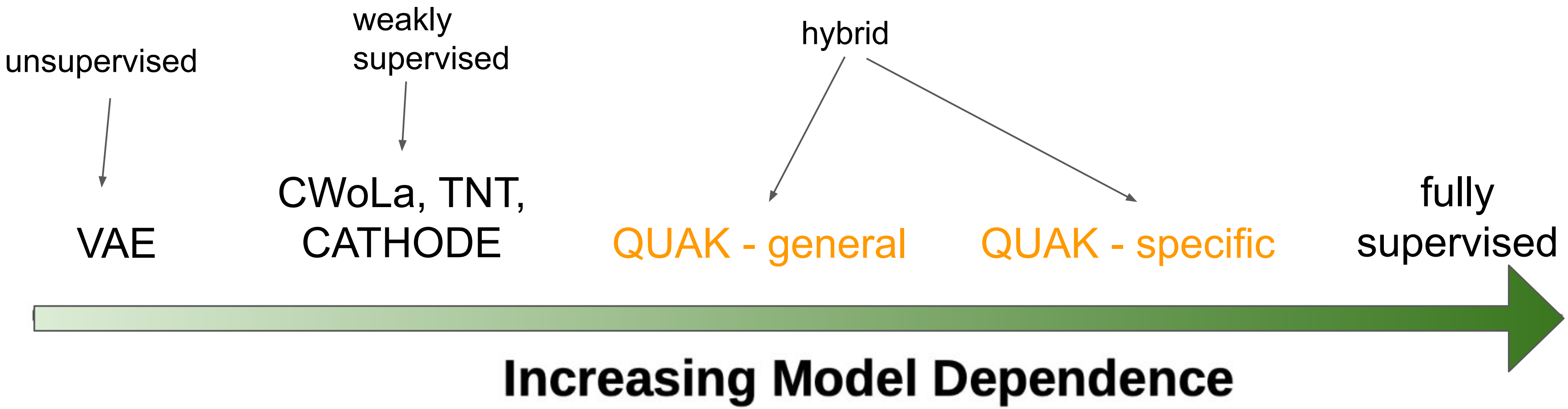
Quasi Anomalous Knowledge (QUAK)

JHEP 21 (2020) 030



- **Hybrid** approach between model-independent and standard search
- Idea: **encode prior knowledge** of how a signal could look
- Train density estimator (normalizing flow) on colorful mix of simulated signals
- Train additional normalizing flow on background **simulation**
- Construct 2D space, select events with high background loss and low signal loss

QUAK: general vs. specific



- Which signals to use for encoding prior knowledge?
  - general - mixture of several signal models
  - specific - using only model to be probed

- Can use QUAK to “interpolate” towards fully supervised approach

Input features

Nice complementarity

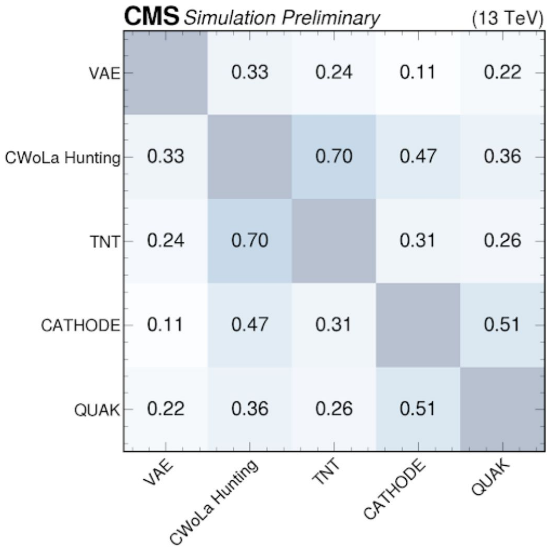
VAE	CWoLa	TNT	CATHODE	QUAK
Constituents p_x , p_y , p_z	m_{SD} τ_{21} τ_{32} τ_{43} n_{const} leptonic energy fraction sub-jets B tag score	same as CWoLa	m_{SD}^{j1} m_{SD}^{j1} - m_{SD}^{j2} τ_{41}^{j1} τ_{41}^{j2} --- + B tag scores of j1, j2 **CATHODE-b**	$\varrho = m_{SD}/p_T$ τ_{21} τ_{32} τ_{43} n_{const} $\sqrt{\tau_{21}} / \tau_1$ jet B tag score

targets individual jets (VAE, CWoLa, TNT)

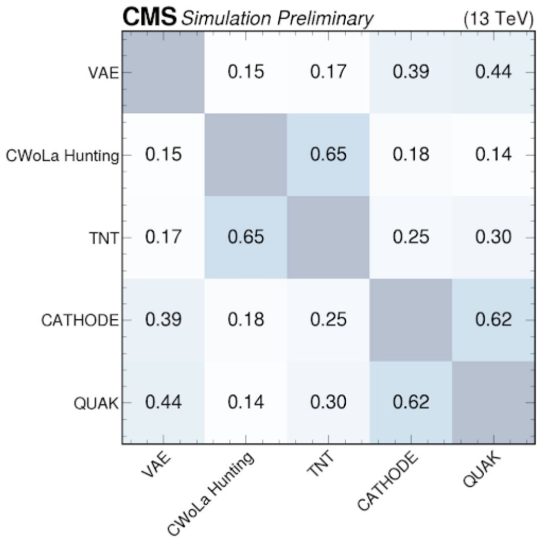
targets events (CATHODE, QUAK)

Checking the correlations

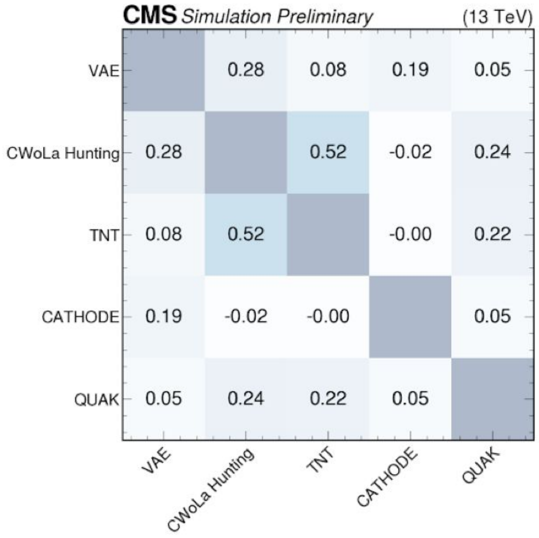
$W' \rightarrow B't \rightarrow qqq\ qqq$



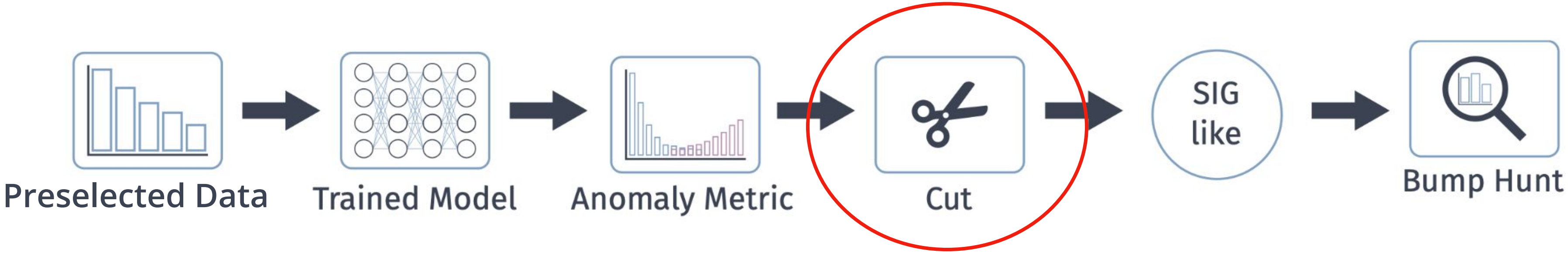
$X \rightarrow YY' \rightarrow qq\ qq$



QCD background



- Complementary architectures and input features reflected in low Pearson correlation among anomaly scores
- TNT and CWoLa most similar (in approach and thus in score)


Preselected Data
Trained Model
Anomaly Metric
Cut
SIG
like
Bump Hunt

Cut definition

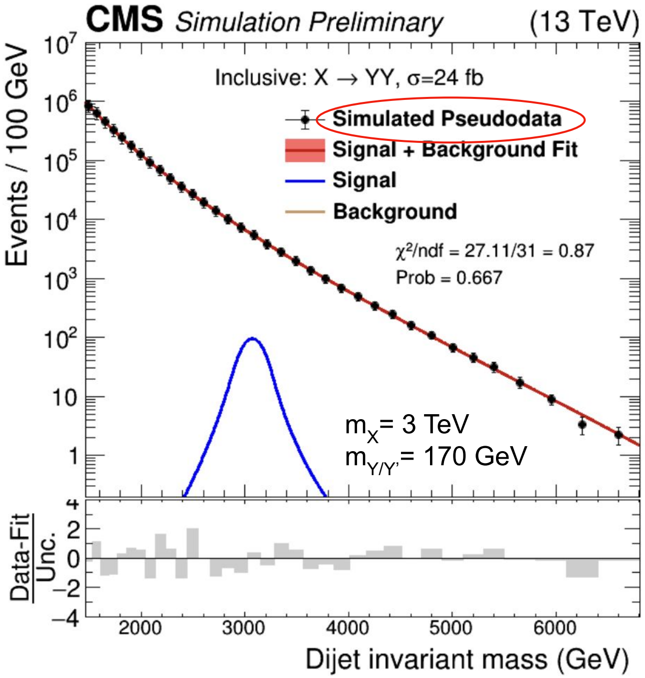
VAE	CWoLa	TNT	CATHODE	QUAK
Single event selection for all masses **10%** most anomalous events	Selection changes for each SR Cut determined on sidebands Chosen such that eff == **1%** for low m_{jj}, eff == **5%** for high m_{jj}	same as CWoLa	Selection changes for each SR Cut determined in SR Chosen such that eff == **1%**	Selection changes for each mass hypothesis Iteratively select least-populated bins in sideband QUAK space, until decent population in SR

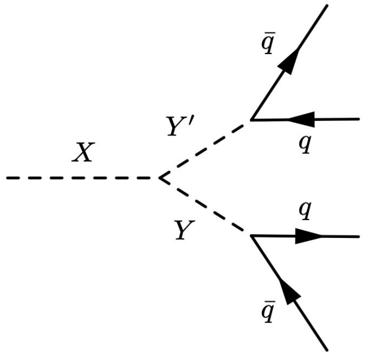
Outline

1. Characterizing jets at CMS
2. Anomaly hunting
3. Expected performance
4. Actual look into data


Preselected Data
Trained Model
Anomaly Metric
Cut
SIG like
Bump Hunt

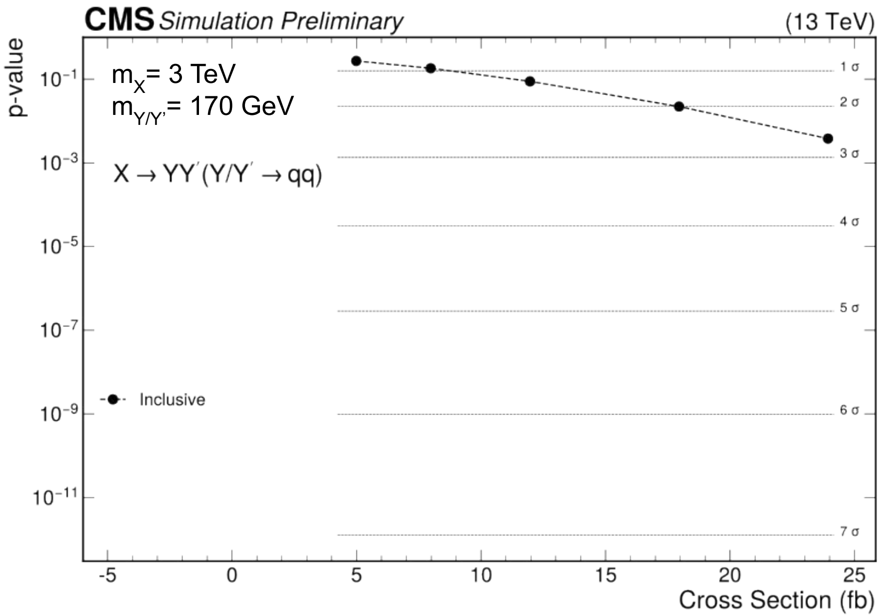
Let's inject some signal



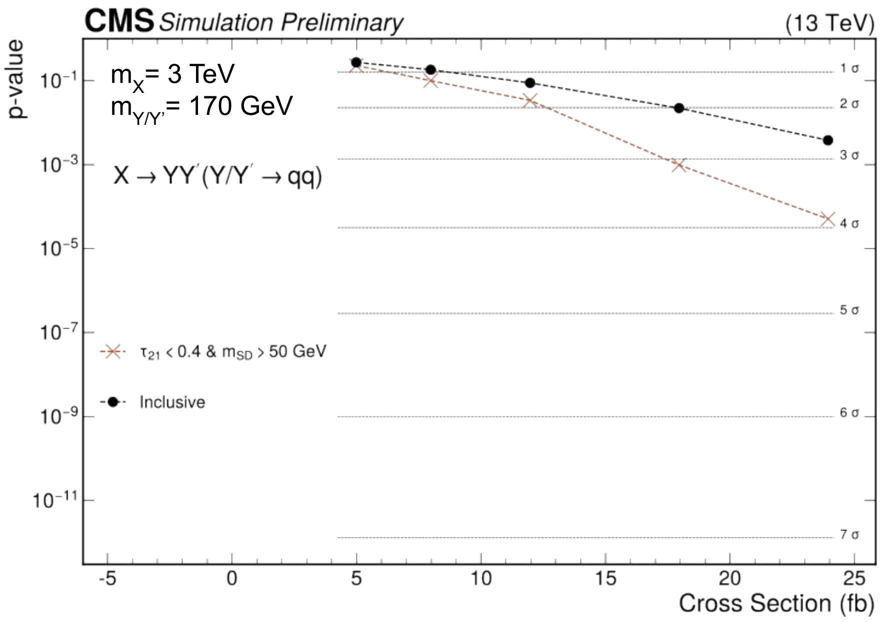


- Injected X→YY' signal with 24 fb cross-section
- Quantifying performance on **simulated** mock dataset worth ~30 fb^{-1}
- Model background and signal with analytic functions
- Signal not visible by eye on top of background

Significance of excess: inclusive

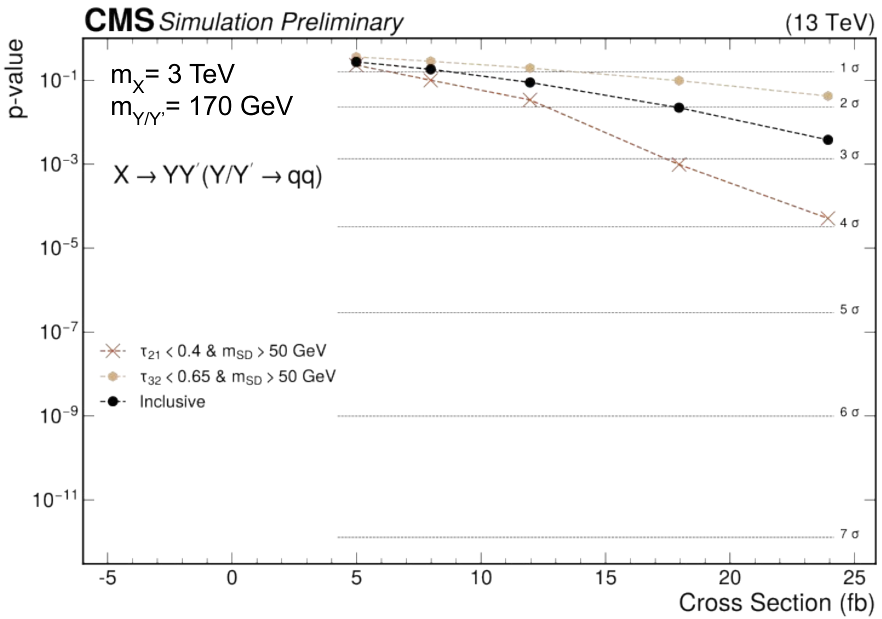


Significance of excess: decent traditional cut



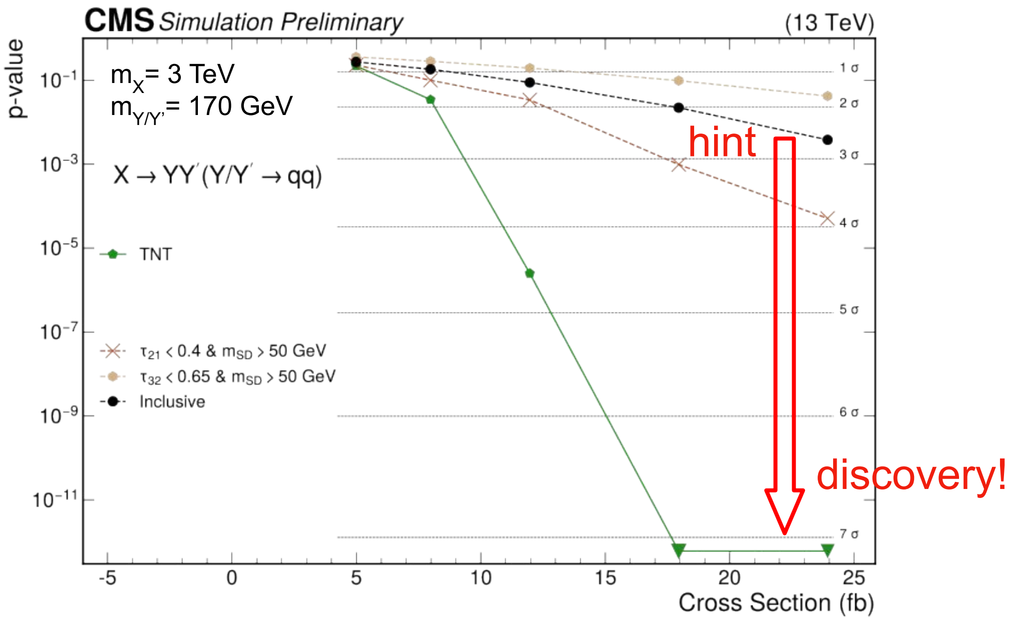
- Cut on N-subjettiness τ_{21} to enrich in 2-prong jets
- Sizeable improvement over inclusive search
- τ_{21} used in many searches

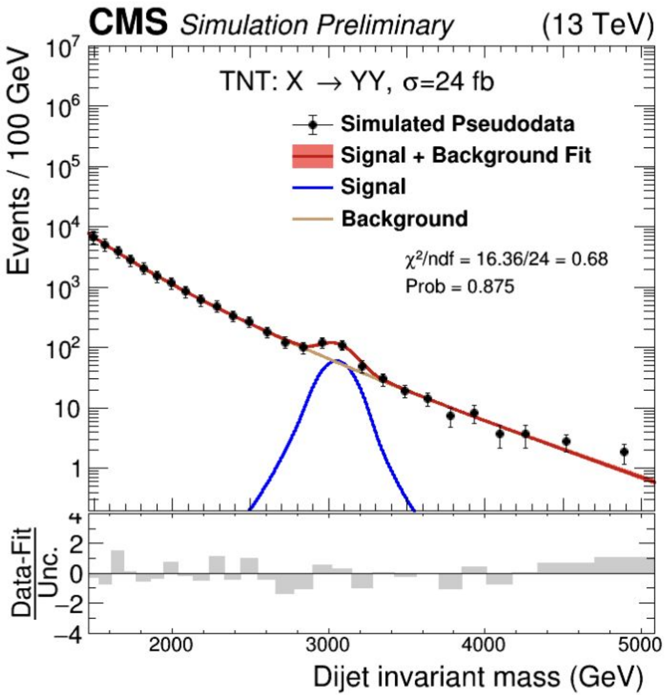
Significance of excess: decent traditional cut



- Cut on N-subjettiness τ_{21} to enrich in 2-prong jets
- Sizeable improvement over inclusive search
- τ_{21} used in many searches
- Cut on N-subjettiness τ_{32} to enrich in 3-prong jets gives worse sensitivity than inclusive search

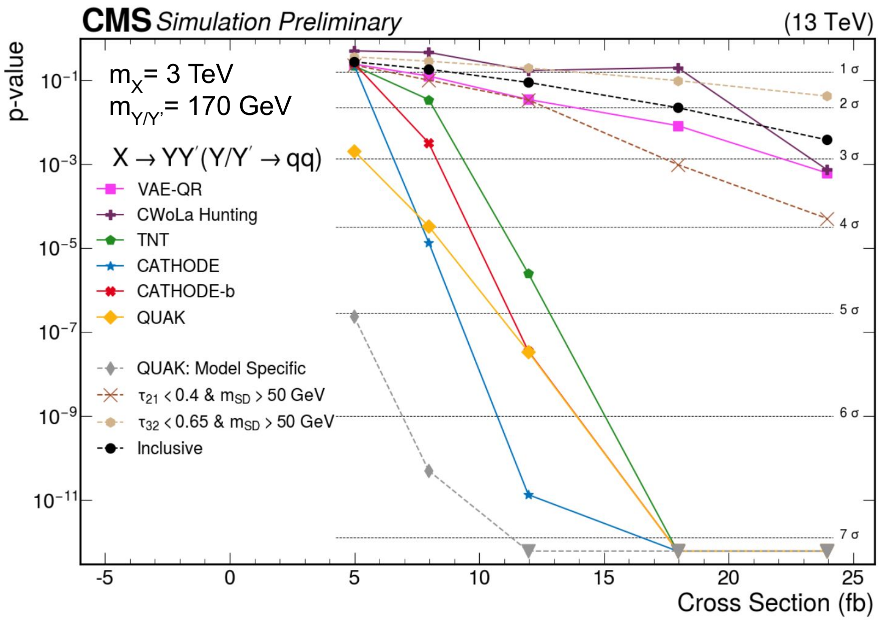
Significance of an excess: TNT





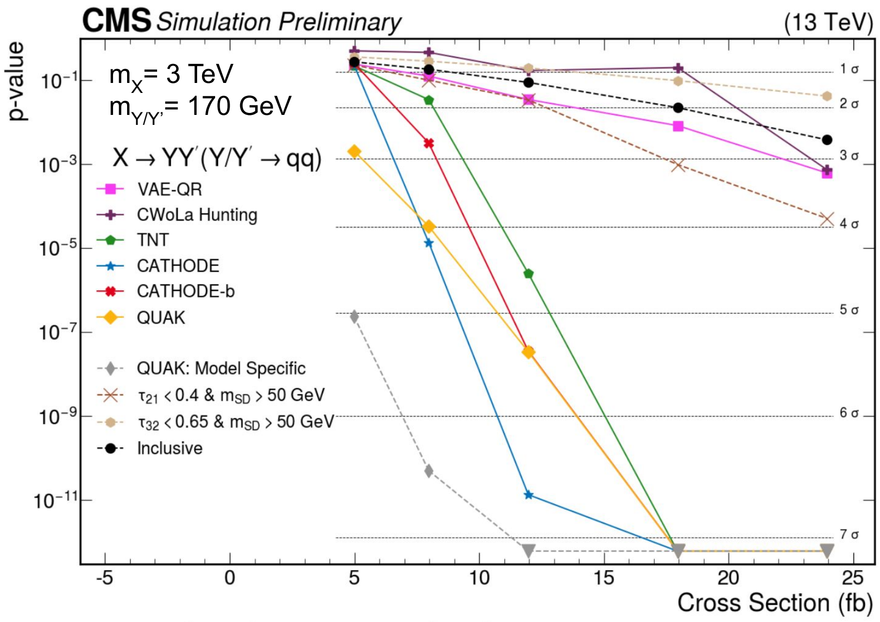
- Prominent peak after cut on TNT anomaly metric
- At 18 fb: Improving from 3 sigma (for τ_{21}) to 7 sigma!

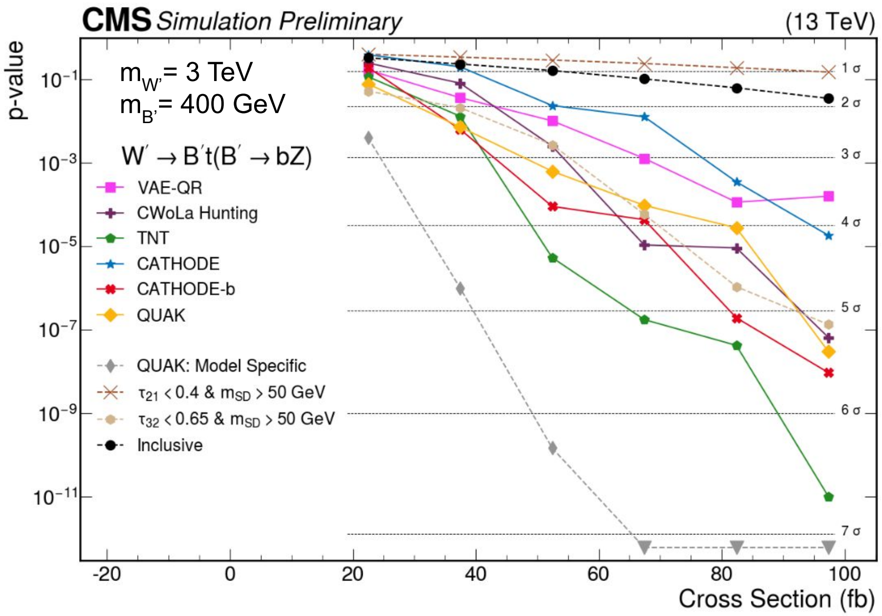
The full picture: a couple of observations



- Model specific QUAK performs best
  - used knowledge about the exact signal
- CATHODE-b worse than CATHODE
  - B tag score in normalizing flows has detrimental effect on signal without b quarks (acts as noise)
- VAE and CWoLa little or no improvement w.r.t. inclusive search

The full picture: a couple of observations



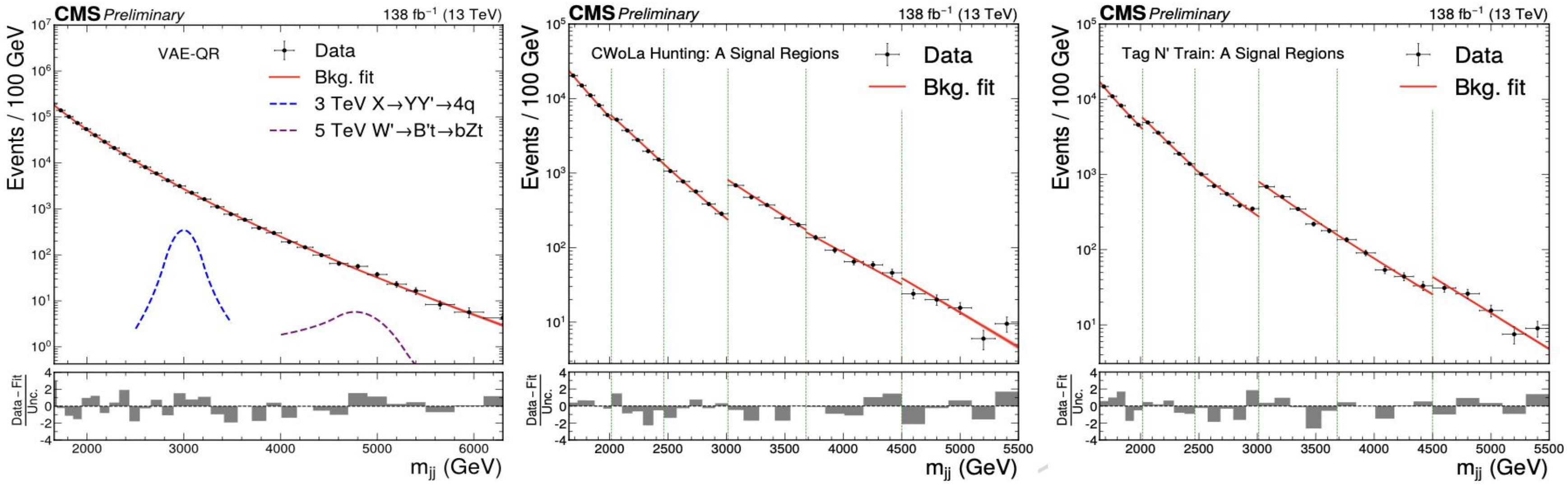


- CATHODE-b better than CATHODE for signals with b quarks (W’ → B’(bZ)t → qqq qqq)
- VAE and CWoLa performing well on rich, broad jets

Outline

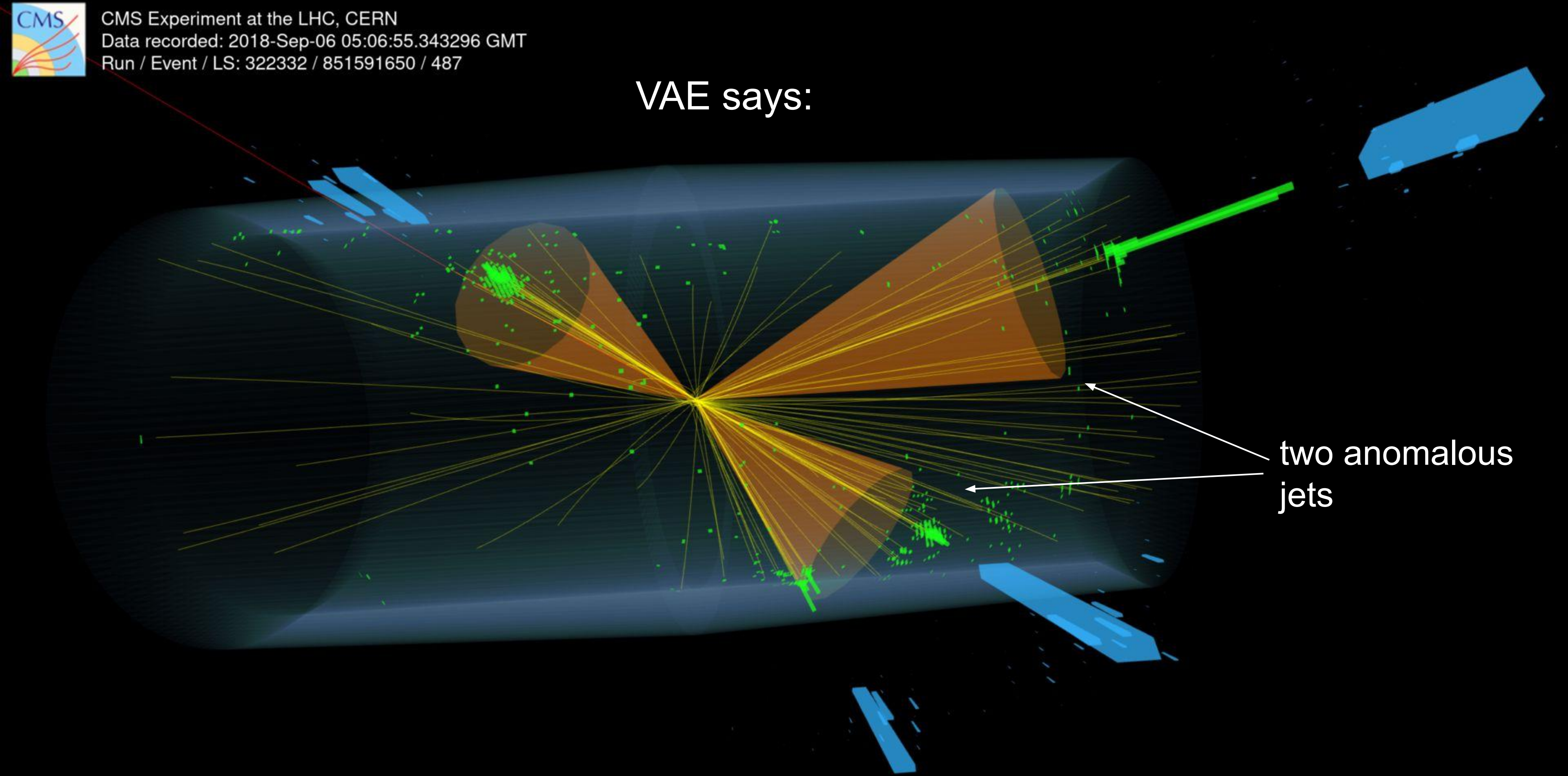
1. What are jets at CMS
2. Anomaly hunting
3. Expected performance
4. Actual look into data

Data spectra - no excess

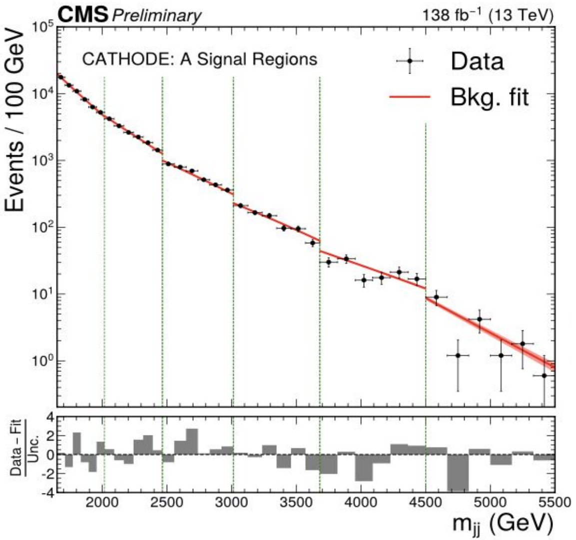


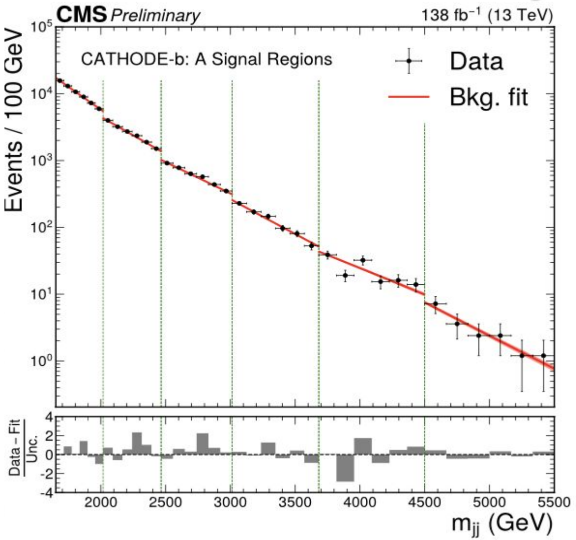
- Reminder: for VAE, only 1 anomaly cut, totally independent of probed mass
- Six different A regions for weakly supervised models (B regions in Backup)
- No significant excess

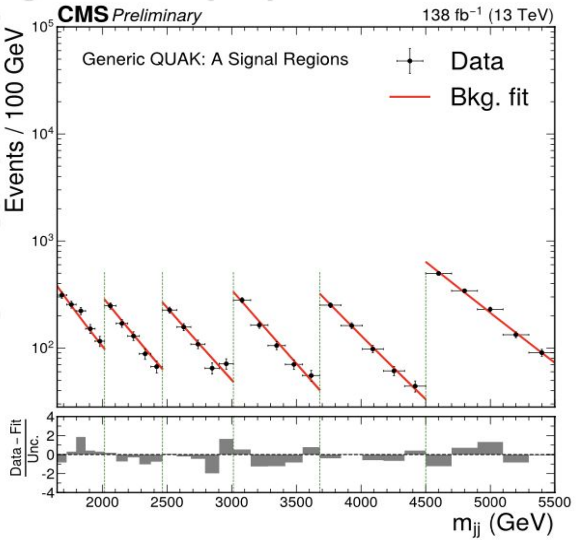
VAE says:


CMS
CMS Experiment at the LHC, CERN
Data recorded: 2018-Sep-06 05:06:55.343296 GMT
Run / Event / LS: 322332 / 851591650 / 487
two anomalous
jets

Data spectra - no excess



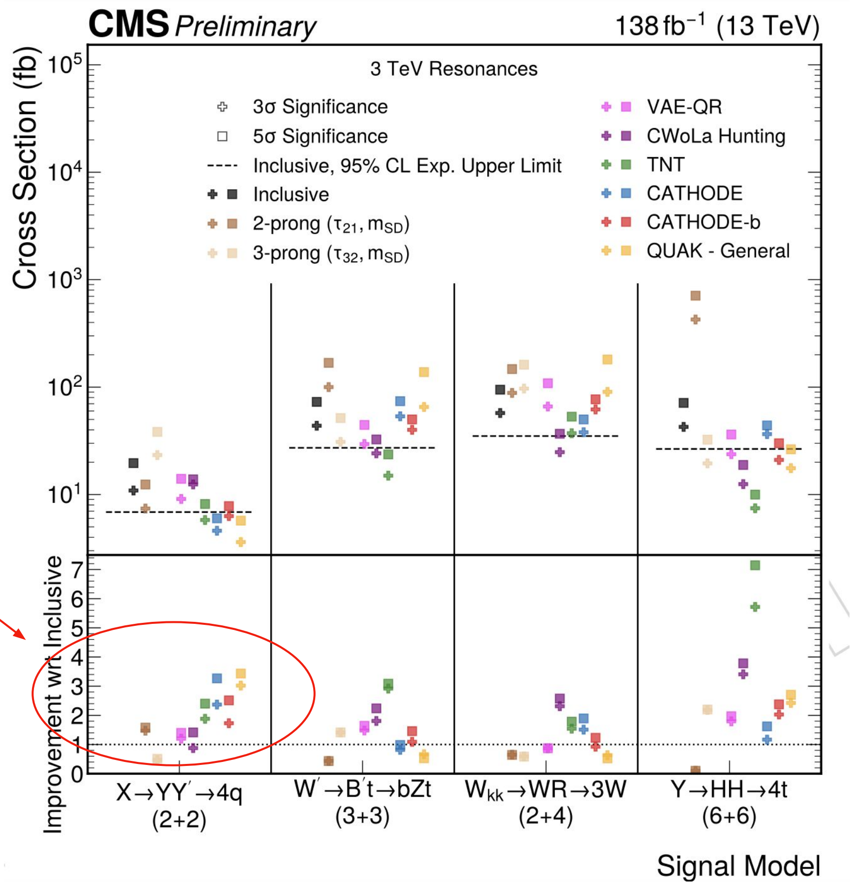




- No significant excess for remaining methods either

Discovery sensitivity

- No excess observed → Which injected cross section would have lead to 3σ / 5σ?
- All methods almost always better than inclusive / traditional search strategy
- For every benchmark, at least one method could claim discovery where inclusive strategy can only set upper limits



m(Y) = 170 GeV, m(B',R,H) = 400 GeV

Model-dependent searches: setting limits

- Statistical inference gives upper limit on number of signal events still allowed in data, N_{UL}

  → Which cross section does this correspond to?

$$N_{\mathbf{sig}} = \sigma \times \mathcal{L} \times A \times \epsilon$$

  → Solve for σ. Done.

Model-dependent searches: setting limits

- Statistical inference gives upper limit on number of signal events still allowed in data, N_{UL}

  → Which cross section does this correspond to?

$$N_{\mathrm{sig}} = \sigma \times \mathcal{L} \times A \times \epsilon$$

  → Solve for σ. Done.

Not the full story!

Limit setting for weakly supervised methods

For weakly supervised methods, efficiency **depends** on number of signal events in data!

- Much signal present → classifier learns well how to pick up on it → high selection efficiency

- Low signal present → classifier cannot learn it properly → low selection efficiency

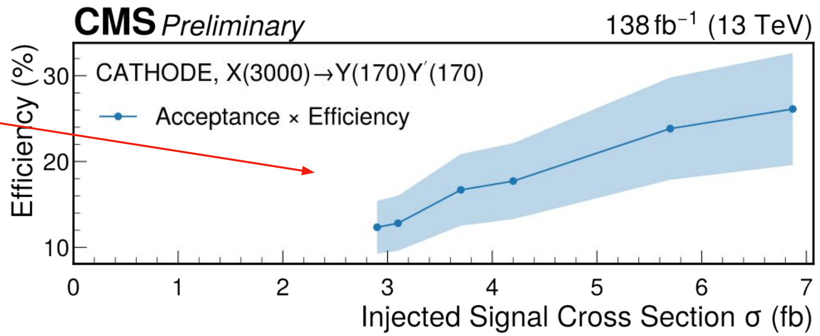
not constant

$$N_{\mathrm{sig}}(\sigma) = \sigma \times \mathcal{L} \times A \times \epsilon(\sigma)$$

- Need to find σ such that $N_{\mathrm{sig}} = N_{\mathrm{UL}}$

Limit setting for weakly supervised methods

$$N_{\text{sig}}(\sigma) = \sigma \times \mathcal{L} \times A \times \epsilon(\sigma)$$



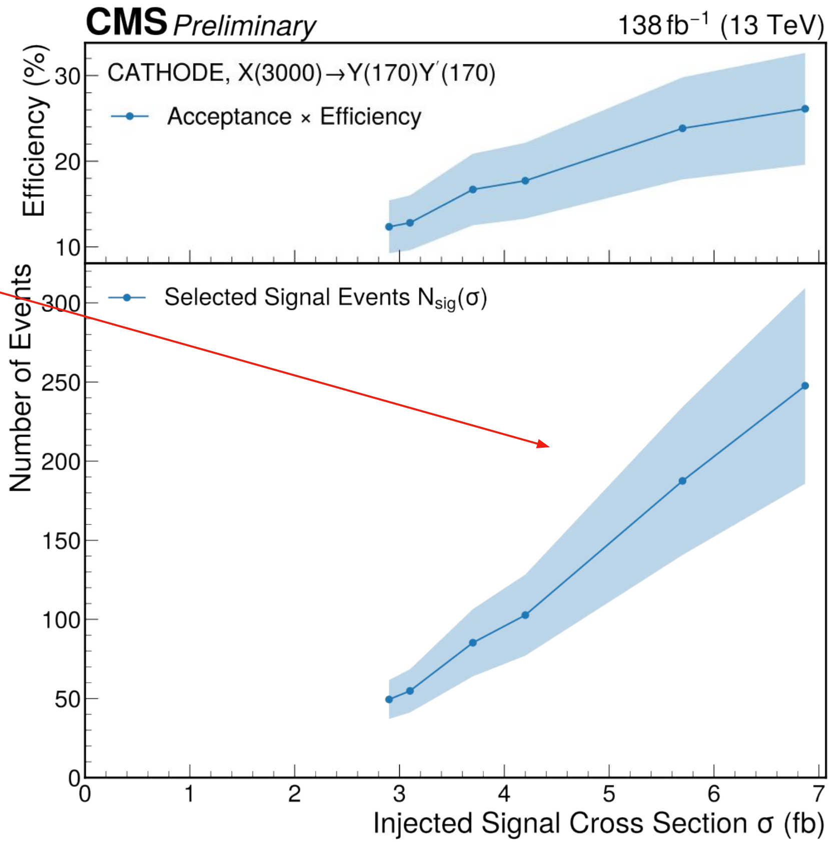
- Inject different # signal, retrain algorithms, measure efficiencies
  - Shaded band: syst. + stat. error

Limit setting for weakly supervised methods

$$N_{\mathrm{sig}}(\sigma) = \sigma \times \mathcal{L} \times A \times \epsilon(\sigma)$$

- Inject different # signal, retrain algorithms, measure efficiencies
  - Shaded band: syst. + stat. error

- Gives number of selected signal events

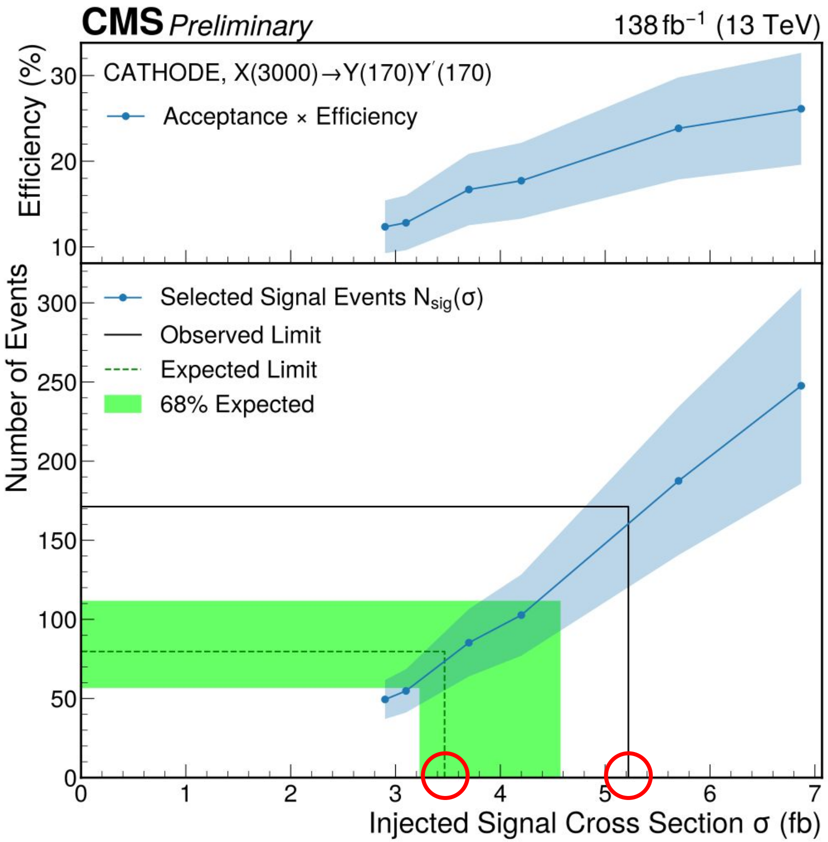


Limit setting for weakly supervised methods

$$N_{\text{sig}}(\sigma) = \sigma \times \mathcal{L} \times A \times \epsilon(\sigma)$$

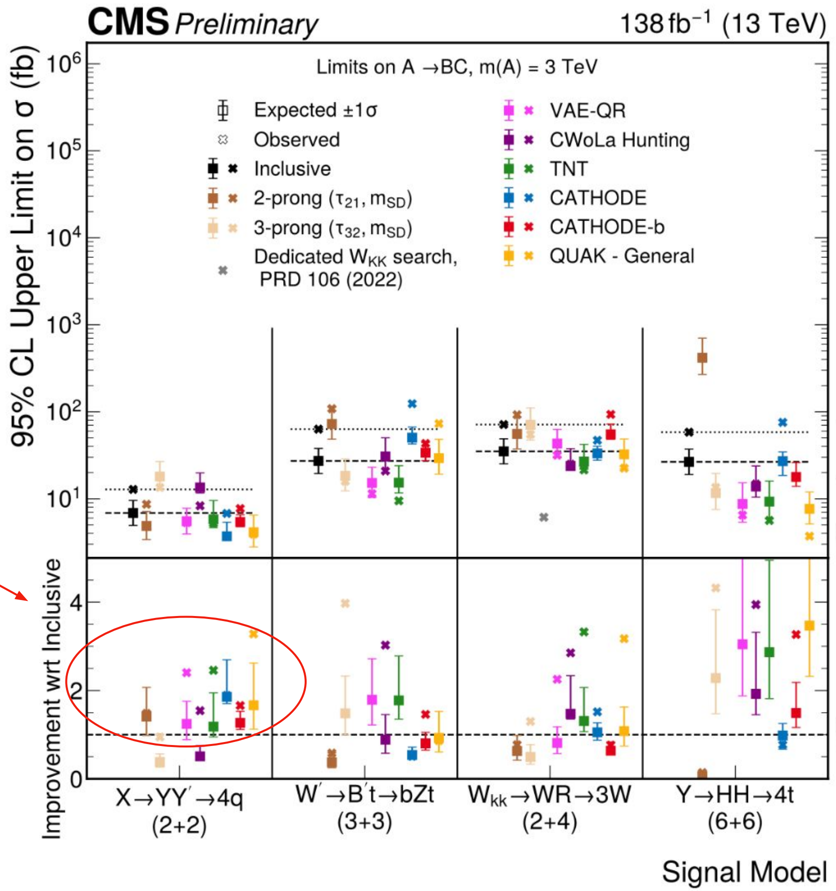
- Inject different # signal, retrain algorithms, measure efficiencies
  - Shaded band: syst. + stat. error
- Gives number of selected signal events
- Find intersection with obs. / exp. event limits

→ Obs. & exp. limits on cross section!



Final limits

- Several 3 TeV resonance scenarios
  - 5 TeV in Backup
- Large improvement over inclusive strategy
- Dominant uncertainty from substructure modelling for signal
- Dedicated W_{KK} search beats all anomaly detection methods (expected)



m(Y) = 170 GeV, m(B',R,H) = 400 GeV

Summary

- For the first time: search with five different & complementary anomaly detection algorithms
- Cast a wide net to catch potential new heavy resonances A decaying into B, C → dijet final state
- Large sensitivity improvements over inclusive search or searches with generic traditional cuts
- No excess observed
- Check out CMS-EXO-22-026 & CMS-NOTE-2023-013
- Stay tuned for more

ChatGPT

I've created a "Where's Waldo?" style image set at a high energy physics conference for you. Look closely among the scientists and researchers, amidst the posters and presentations, to find Waldo. Enjoy your search in this academically rich scene!

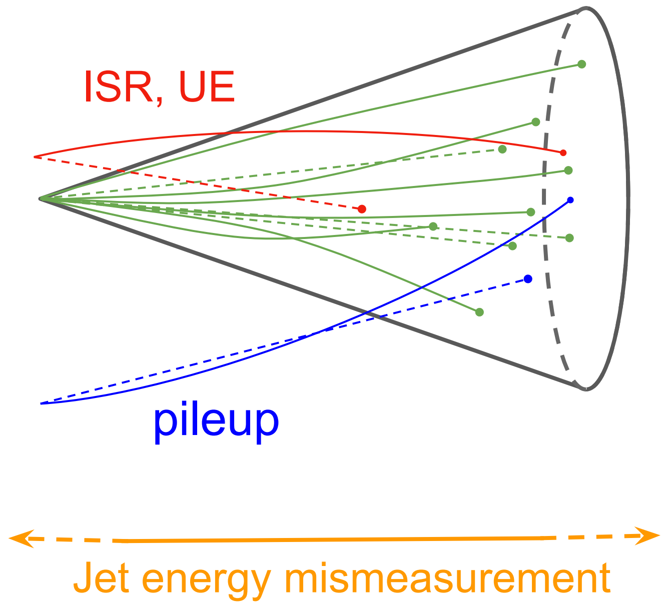
BACKUP

Want a starting point as clean as possible

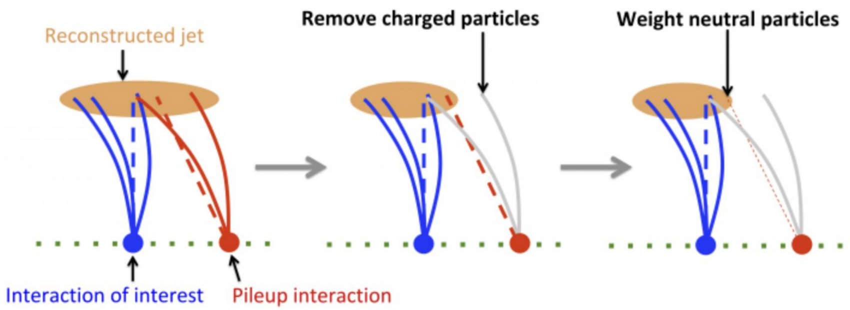
- Jets inevitably not a perfect representation of the resonance decay
- Not everything in jet comes from particle at origin
  - Initial-state radiation, underlying event, pileup
- Measurement inaccuracies for constituents

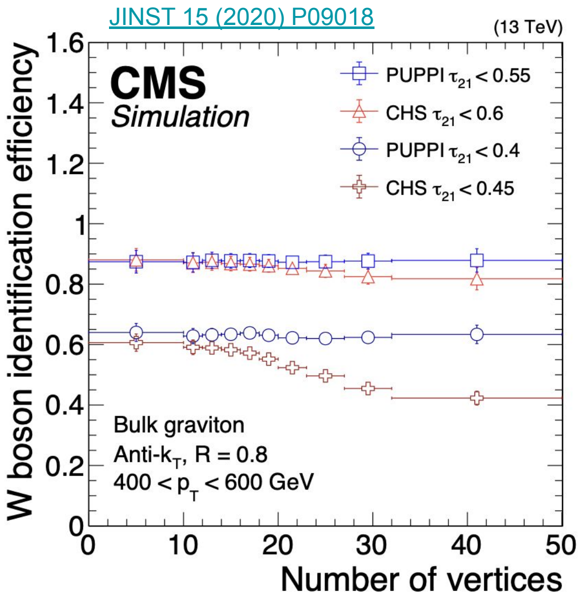
Can give appearance of anomalous radiation patterns

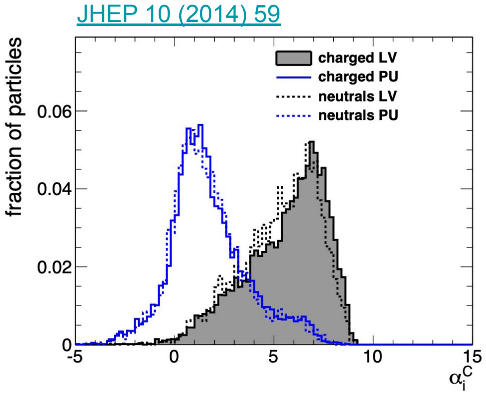
Can reduce expressiveness of truly exotic features, pronginess, etc …



Removing particles from pileup

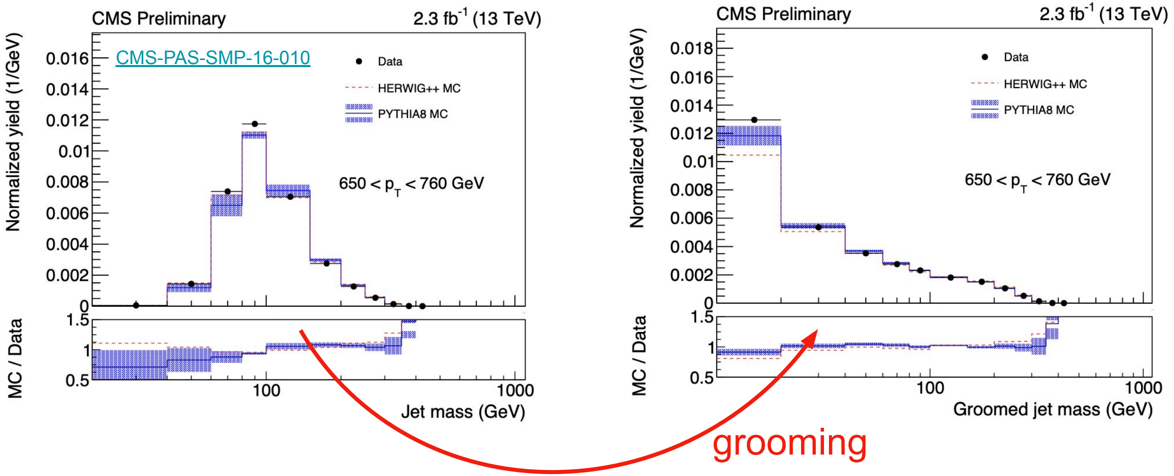


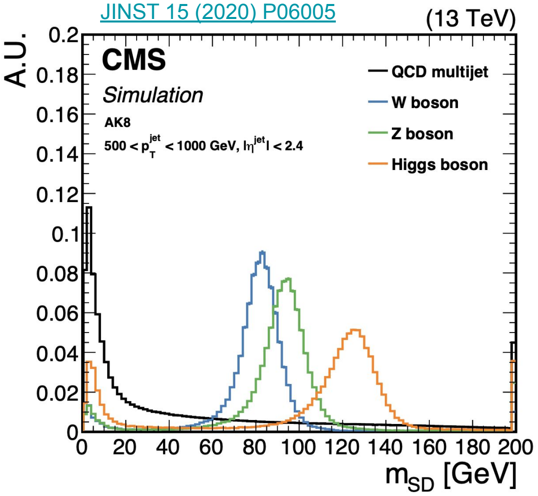




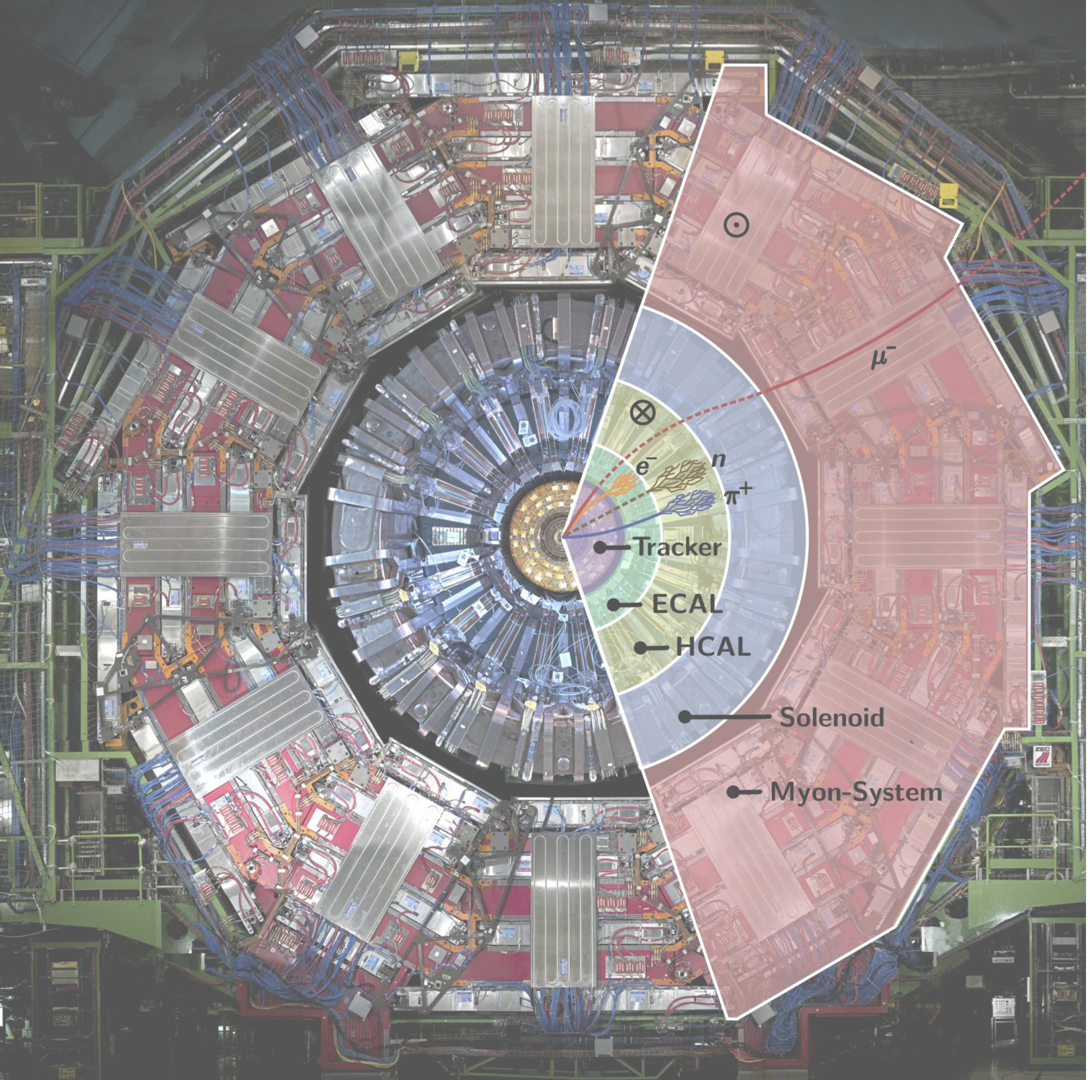
- Need to resolve products from up to 50 simultaneous collisions
  - Tracks: good pointing resolution
  - Neutrals: poor pointing resolution
- PUPPI algorithm discards neutral pileup particles by extrapolating from charged to neutrals
- Large improvement in jet resolution + substructure

Grooming the jet with soft-drop algorithm





- Soft-drop algorithm to remove soft and wide-angle radiation
- Iteratively undo clustering and remove sub-jets that fail: $$\frac{\min(p_{T1}, p_{T2})}{p_{T1}+p_{T2}} > \overset{0.1\,=}{z_{\text{cut}}}\left(\frac{\Delta R_{12}}{R_0}\right)^{\overset{0\,=}{\beta}}$$
- Moves QCD Sudakov peak out of the way → soft-dropped jet mass m_{SD} much more expressive



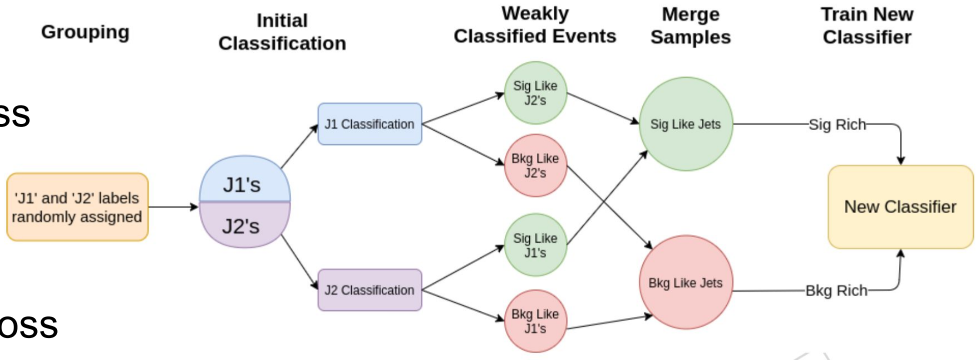
CMS uses **Particle Flow** reconstruction:

- Aimed at reconstructing each individual particle
- "Follow" the path of a particle through the detector
- Match deposits between subdetectors
- For each particle combine subdetector information for best E/momentum measurement

Weak Supervision #2: Tag N' Train (TNT)

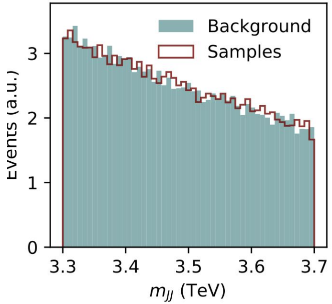
- Mixed Sample 1:
  - In m_{jj} window + in top 20% of AE loss
- Mixed Sample 2:
  - In m_{jj} sideband OR
    In m_{jj} window & bottom 40% of AE loss
- Train Mixed Sample 1 vs Mixed Sample 2

→ Anomaly score: score j1 * score j2



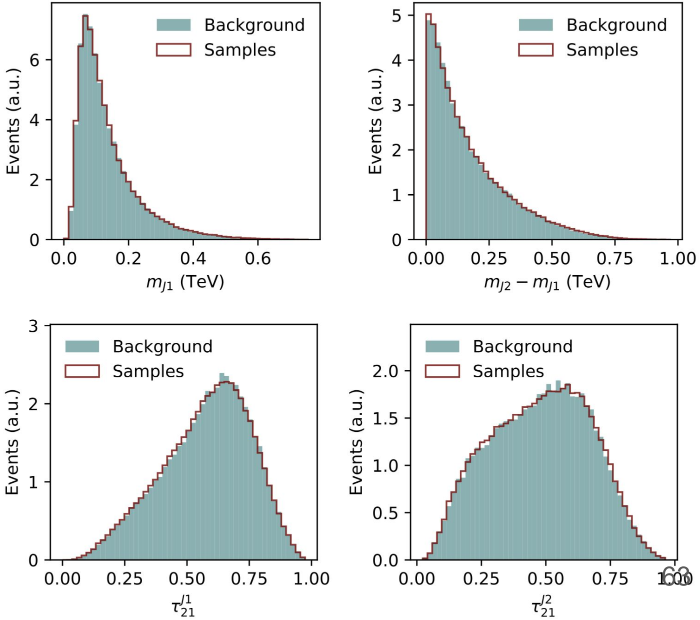
Weak Supervision #3: CATHODE

m_{jj} from KDE fit



x from $f^{-1}(z;m_{jj})$

Phys.Rev.D 106 (2022) 5, 055006



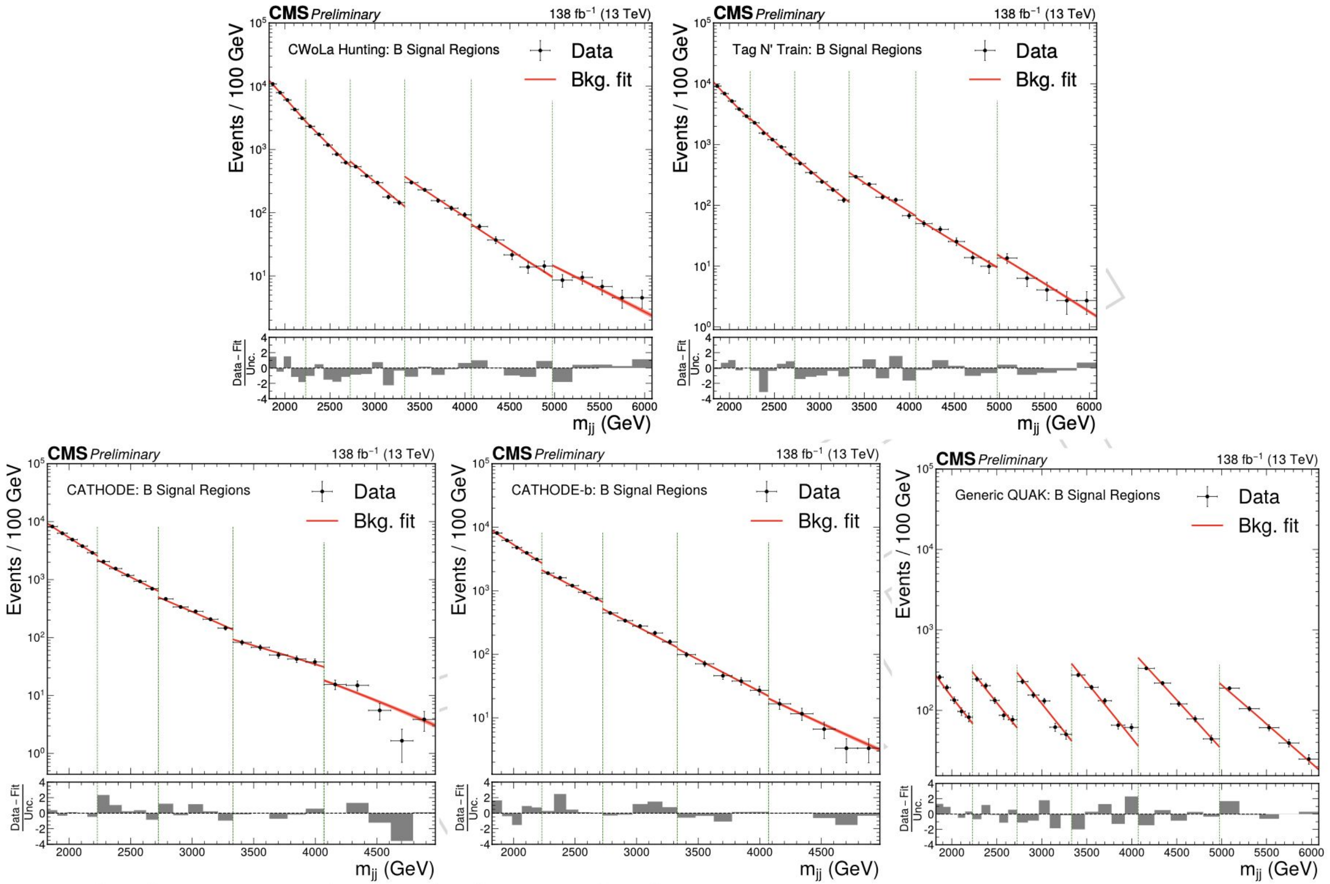
QUAK Signal Prior

- Train 6 separate normalizing flows (six-layer autoregressive rational quadratic splines) on different signal samples
  - Grouped by daughter masses (M80-M80), (M80-M170), (M80-M400), (M170-M170), (M170-M400), (M400-M400)

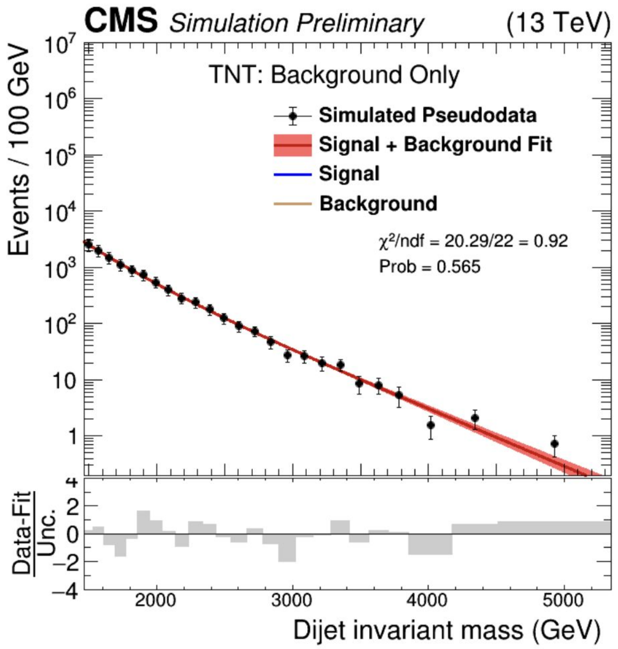
- Normalize each score so mean 0, std 1

- Combine 6 scores into single ‘sig-like’ score using L5 signed norm $(|s1|^5+|s2|^5+...)^{(1/5)}$

Signal B Regions



What happens if no signal present?



- No sculpting
- No artificial excesses
- Checked for all methods
- Validated also on data in $\Delta\eta_{jj}$ sideband ($2.0 < \Delta\eta_{jj} < 2.5$)

Signal extraction - bump hunt on dijet mass spectrum

- After anomaly selection, all methods share common statistical framework

- Bump hunt performed on m_{jj} spectrum with 4 GeV bin size to emulate an unbinned fit (for plots, coarser binning is used)

- Background distribution modeled with standard dijet function

$$\frac{dN}{dm_{jj}} = \frac{P_0(1-x)^{P_1}}{(x)^{P_2+P_3\log(x)+P_4\log^2(x)}}$$

- Starting with P3, P4 = 0, but can be added if found they improve fit quality (Fisher's F-test)

- For signal use a double Crystal ball (from fits to MC); generic shape
- Fit quality: compute chi2/ndf → p value > 0.05

Fit bias study

- Test for bias in functional form
- Generate toys according to alt. fit functions, fit & check bias in signal strength
- Perform for different signal regions / masses
- No significant bias seen

Global p value for weakly supervised methods

- Approximate each SR as fully indep. search → trial factor of 12 for whole scan
- Within each signal region, use traditional methods (toys) to compute effective trial factor based on mass points scanned
- Global pval = (local pval) * (SR trial factor) * 12

Effect of Signal in Data on Limit Setting

- Presence of the **same signal already in the data** prior to injection for limit setting would lead to biased estimate of signal efficiency → limits could undercover
- Tested in two MC studies
  - One using the CATHODE method on a 5 TeV Y→HH signal (low stats. regime)
  - One using the TNT method on a 3 TeV X→YY signal
- For each study, construct 100 mock datasets containing some signal
  - Run limit setting procedure on each dataset (assuming no signal)
  - Compare distribution of limits to true xsec to check coverage
- Excellent coverage observed for signal strengths giving up to 2sigma
- TNT sees some undercoverage (85%) for very large (3.6σ) signal
- In data no excesses larger than ~2.5σ → these studies demonstrate we will have coverage

Mismatched signals?

- If signal already in the data (A) that is different from the one we are setting a limit on (B), that will affect results
- During injections, training will be performed with both A & B signal events
- Found that efficiency when training on A+B is the same or less than the efficiency when training on just A/just B
- Same or lower eff. → same or more-conservative limits

Protocol if we see something

- Consult with detector experts to look at most anomalous events and exclude detector noise, etc.

- Look at features for most anomalous jets vs. standard jets

- Should be clear indication which feature(s) are triggering excess

- Can inform more targeted search with traditional methods / features

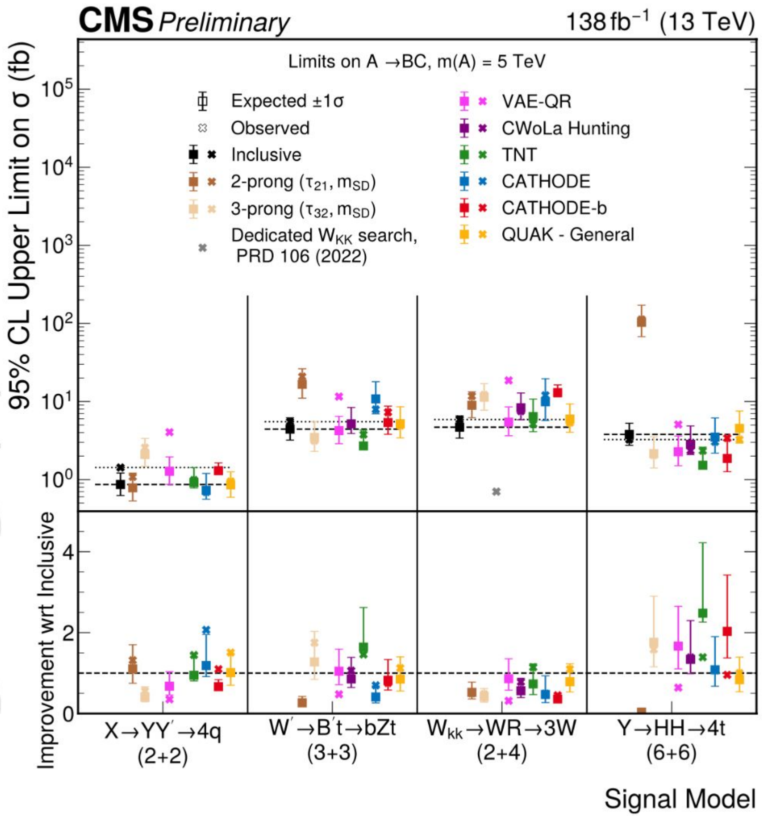
3 TeV limits

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

5 TeV limits

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

5 TeV limits



CMS Preliminary
138 fb^{-1} (13 TeV)
Limits on A →BC, m(A) = 5 TeV
Expected ±1σ
Observed
Inclusive
2-prong (τ_{21}, m_{SD})
3-prong (τ_{32}, m_{SD})
Dedicated W_{KK} search, PRD 106 (2022)
VAE-QR
CWoLa Hunting
TNT
CATHODE
CATHODE-b
QUAK - General
95% CL Upper Limit on σ (fb)
Improvement wrt Inclusive
X→YY′→4q (2+2)
W′→B′t→bZt (3+3)
W_{kk}→WR→3W (2+4)
Y→HH→4t (6+6)
Signal Model

Systematic Uncertainties

- Effects on signal efficiency considered for:
  - Substructure modeling
  - Pileup
  - PDF uncertainties
  - B tagging
  - Jet energy scale & resolution
  - Renormalization / factorization scales

Substructure Modeling

- No correction factors for signals with >3 prongs can be derived with SM proxy
- Derive per-prong correction via Lund Plane Reweighting ([CMS-DP-2023-046](CMS-DP-2023-046))
  - Derived on boosted W's, validated on boosted tops
  - Applicable to signals with any number of prongs
- Recluster large-radius jet with exclusive kt algorithm, recluster subjects with CA algorithm to get splittings,
- Sort splittings into Lund plane, divide Data/MC for per-splitting correction
- Per splitting → per subjet → per jet → per event