# Unsupervised Machine Learning-based Anomaly Detection in an ATLAS Dijet Resonance Search

[ATLAS-CONF-2022-045]

Julia Gonski

*2 November 2022*

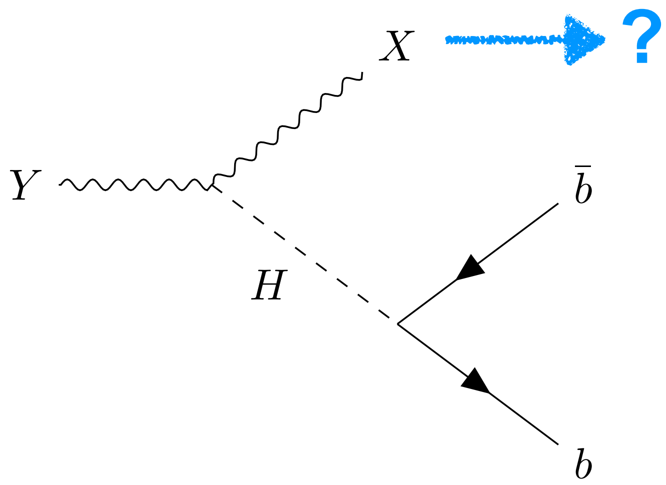
ML4Jets @ Rutgers

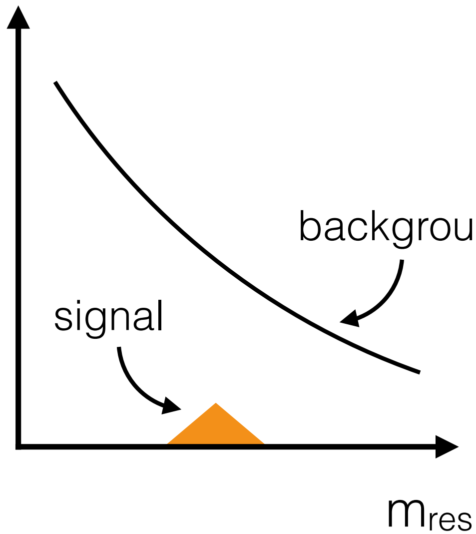
Nevis Laboratories
Columbia University

ATLAS Experiment

# Analysis Overview

- Search for new heavy resonance Y (~TeV) decaying to an SM Higgs ( $\rightarrow b\bar{b}$) and a new particle X (~100s GeV)
    - X and H are highly boosted: reconstruct as large-R jets (R=1.0)
- Machine learning highlights
    1. Unsupervised learning for signal model-agnostic X tagging
    2. Neural net-based tagging of boosted H→$b\bar{b}$ topology
    3. DNN-based reweighting procedure for data-driven background estimation
- Previous search in Y→XH signature using 36 fb-1 [1709.06783]: no significant excesses
    - Assumed X decays to $q\bar{q}$: generated this final state as our primary signal grid for supervised tagging benchmark



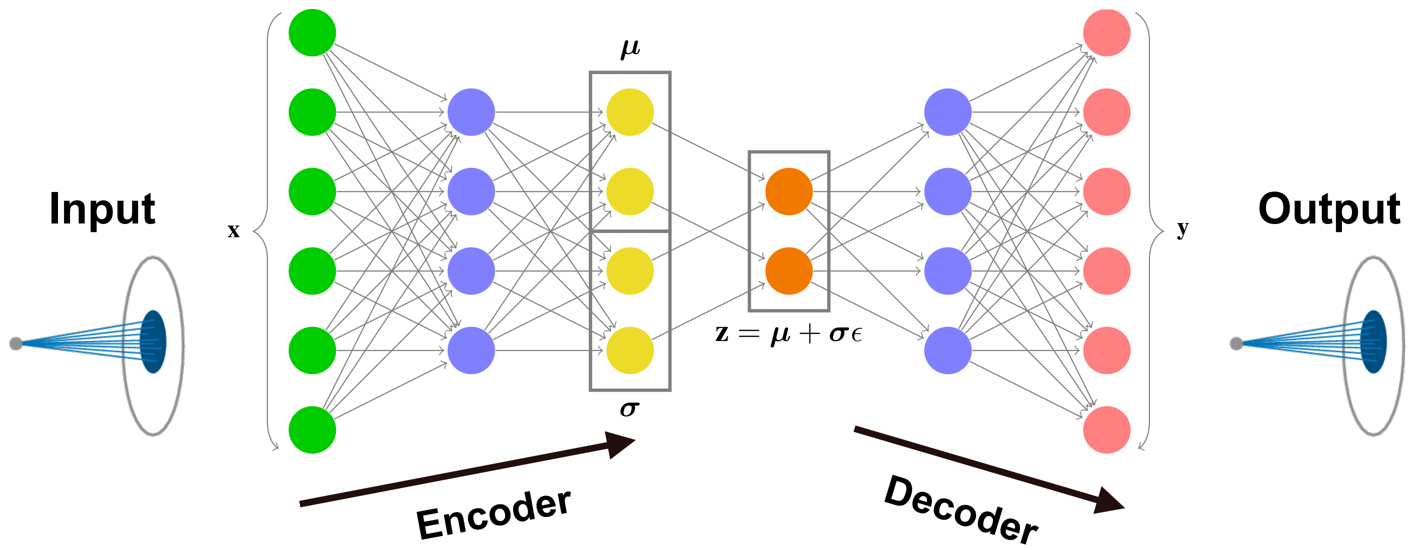

J. Collins, K. Howe, BPN
PRL 121 (2018) 241803

J. Collins, K. Howe, BPN
PRD 99 (2019) 014038

# Autoencoders for Jets

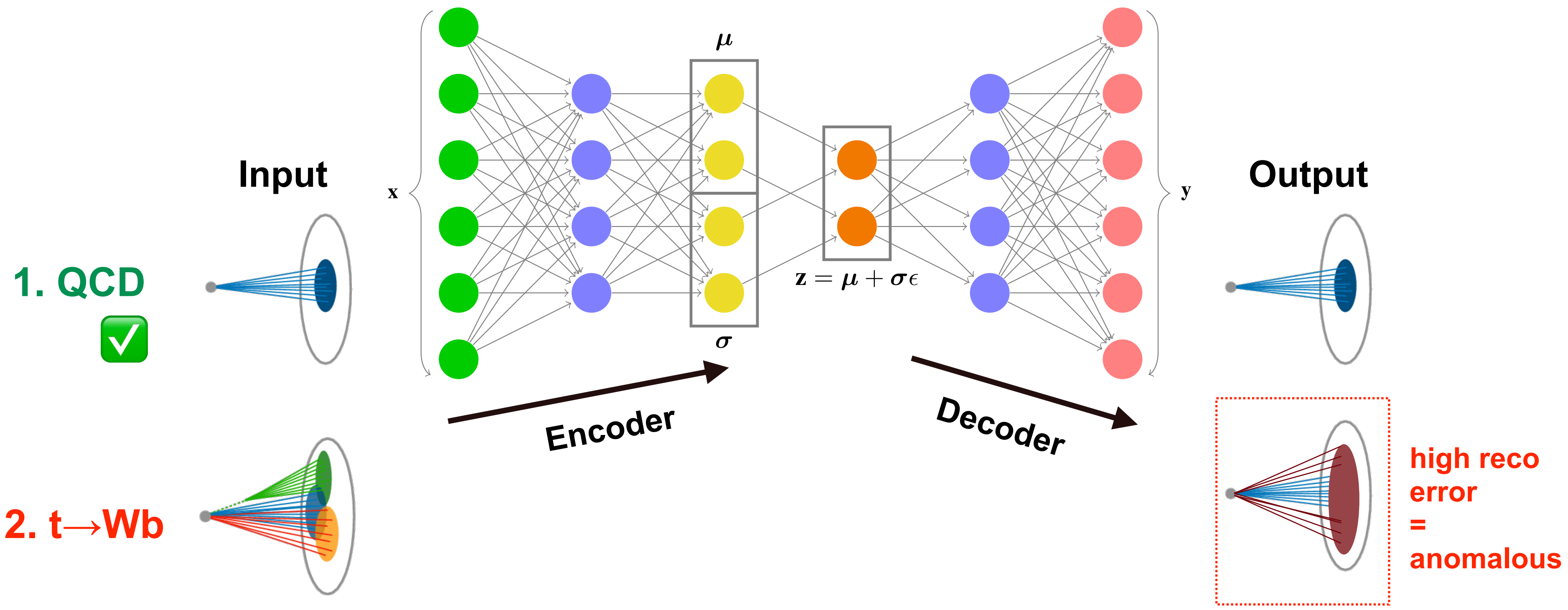
- **Autoencoder:** generative model that *encodes* input in lower-dimensional latent space, *decodes* from latent space, and checks reconstruction error
  - Train over data (mostly QCD)

1. QCD ✅



# Autoencoders for Jets

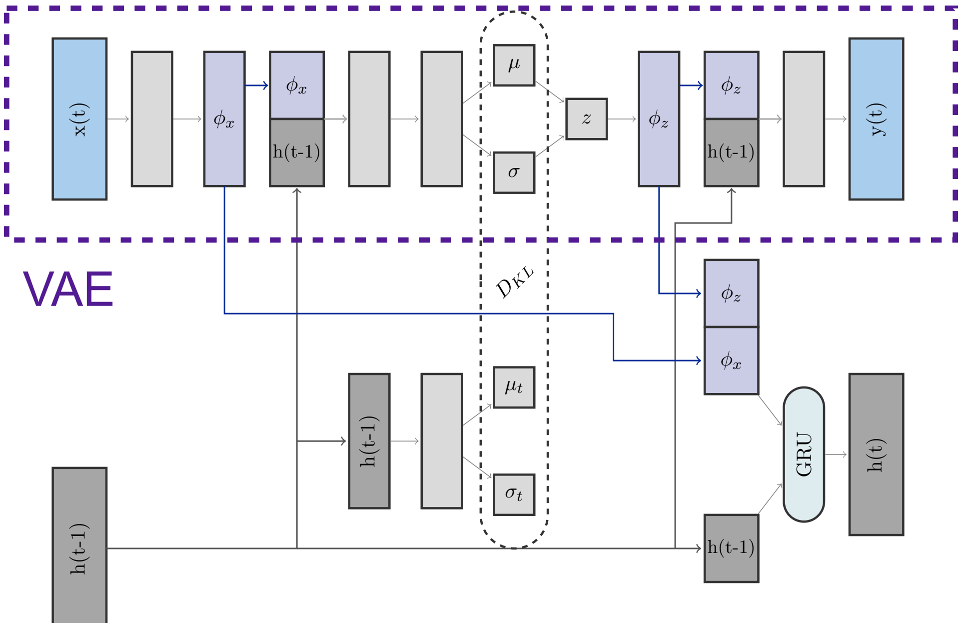
- **Autoencoder:** generative model that *encodes* input in lower-dimensional latent space, *decodes* from latent space, and checks reconstruction error
  - Train over data (mostly QCD)
  - Model jets by their constituent 4-vectors: *jet substructure* is an anomalous feature
  - Order constituents by clustering step: **sequence information is relevant!**



# Variational Recurrent Neural Network

- **Variational RNN:** recurrent neural network (RNN) that updates a VAE latent space at each time step; accommodates variable-length input sequences
- Define *anomaly score* per jet as a function of the KL divergence loss term: $AS = 1 - e^{-D_{KL}}$

➡ VRNN paper published using LHC Olympics dataset [2105.09274]




**VRNN Cell**

**Loss**

$$\mathcal{L}(t) = |\mathbf{y}(t) - \mathbf{x}(t)|^2 + \lambda D_{KL}(z||z_t)$$

Mean-squared reconstruction error

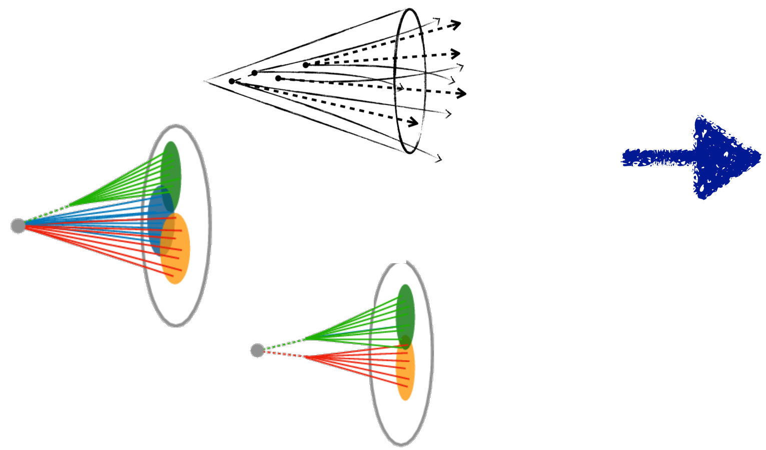
Kullback-Leibler Divergence

# VRNN Jet Tagging in Y→XH

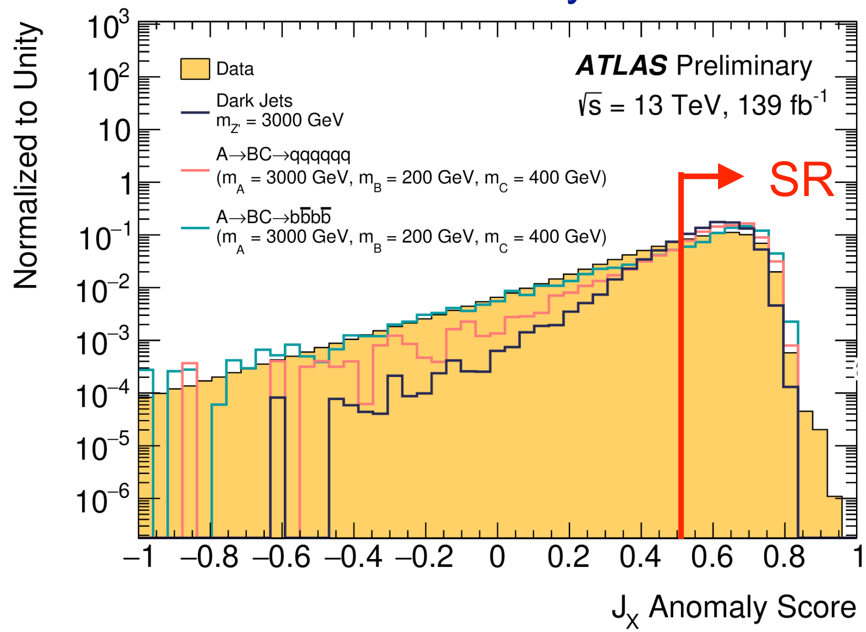
- Train over full Run 2 dataset of large-R jets (R=1.0) with $p_T$ > 1.2 TeV
  - Up to 20 constituents ordered by kt splitting + D2, $\tau$32, Split12, Split23
- Evaluate over four substructure hypotheses to verify degree of model dependence
  - 2 prong, 3 prong, heavy flavor ($b\bar{b}$), and dark jets (Pythia Hidden Valley Model A)

➡ Use a flat cut of AS > 0.5 as SR definition for broad sensitivity enhancement: competitive with D2 on 2-prong signals and ~10x better for dark jets

Possible Signal Jet Models



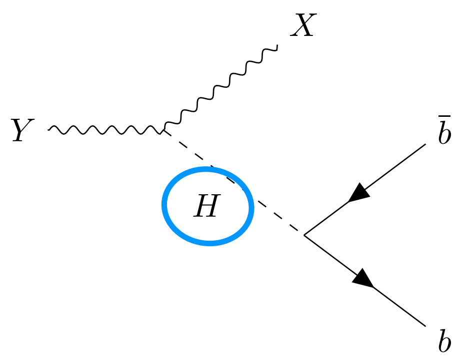
Jet Anomaly Score



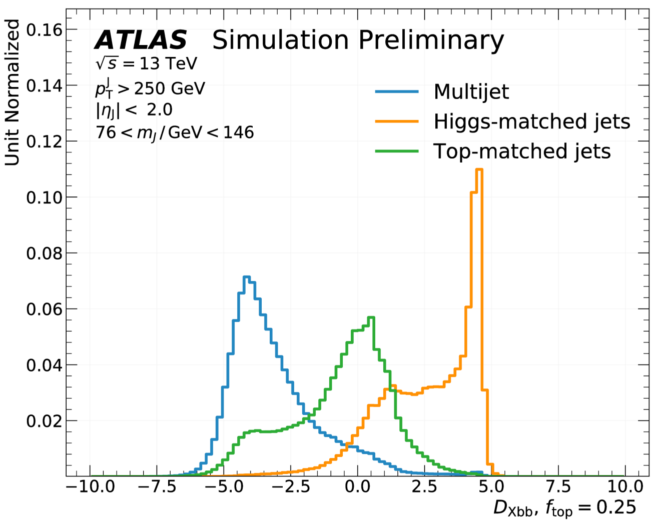
# Neural Net H→$b\bar{b}$ Tagging

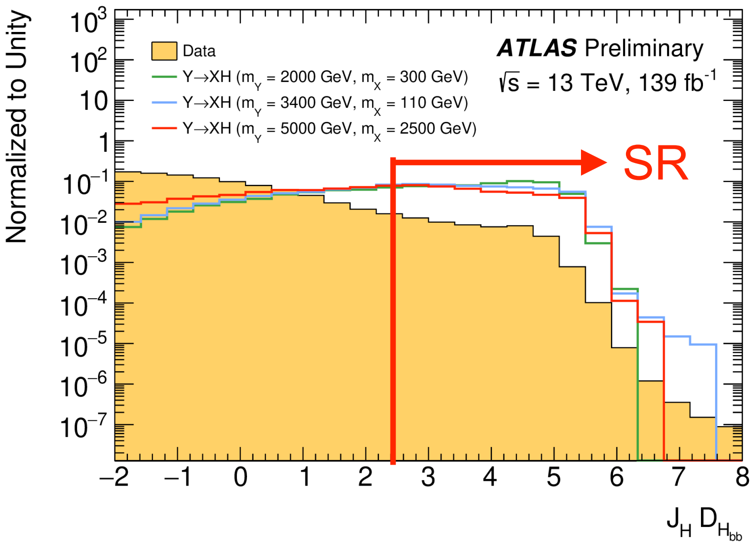
- First use of ATLAS neural net-based double b-tag algorithm to select Higgs vs. dijet or top backgrounds [ATL-PHYS-PUB-2020-019]
  - Train over large-R jet $p_T/\eta$ and up to 3 subjet b-tagging scores
  - Outputs: three class probabilities → discriminant $D_{Hbb}$

➡ Tag Higgs boson using 60% WP and ftop=0.25 as per central FTag recommendation




$$D_{Xbb} = \ln \frac{p_H}{f_{\text{top}} \cdot p_t + (1 - f_{\text{top}}) \cdot p_{\text{QCD}}}$$

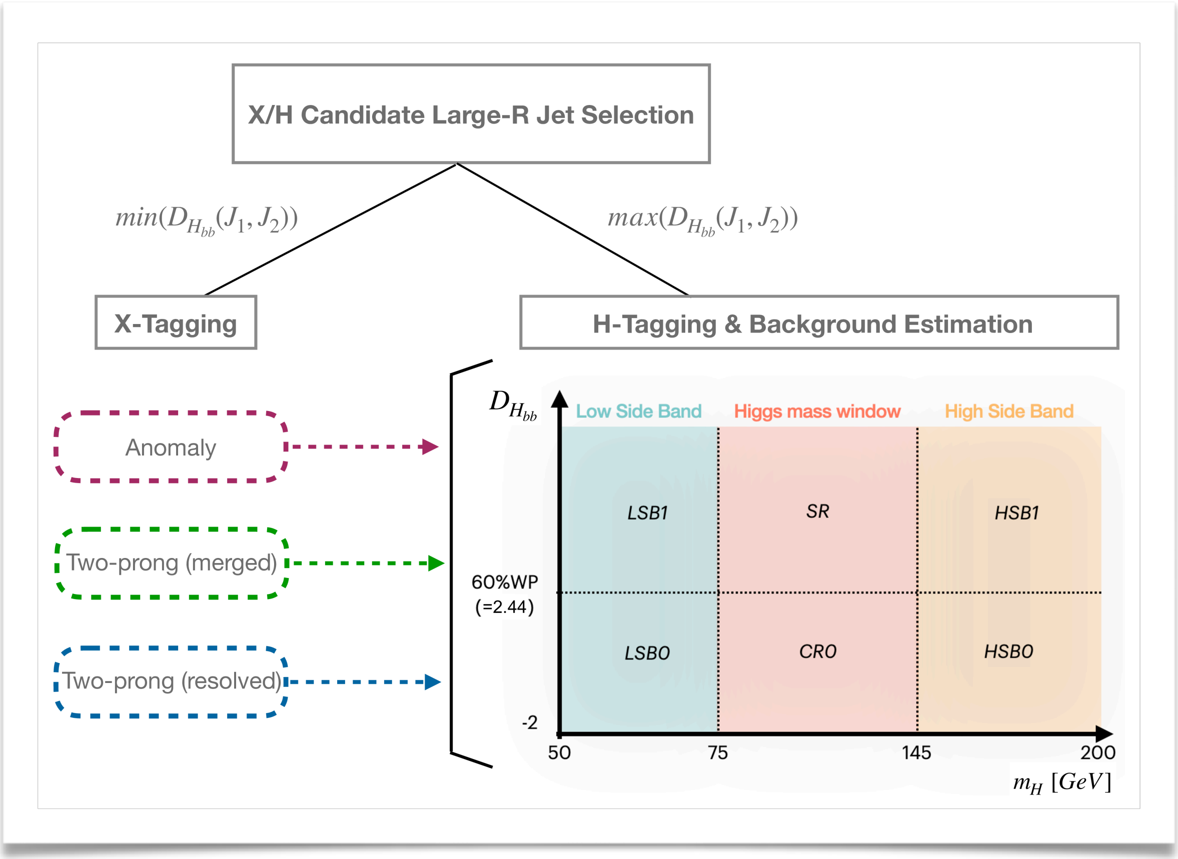


# Analysis Flow

1. **Large-R jet trigger**: $J_1(p_T) > 500$ GeV and $m_{JJ} > 1.3$ TeV
2. **Ambiguity resolution**: jet with highest $D_{Hbb}$ score is Higgs candidate
3. **X-tagging**: AS of X candidate > 0.5
   - Background estimation + SR in single bin of anomaly score
   - Separate 2-prong regions with D2

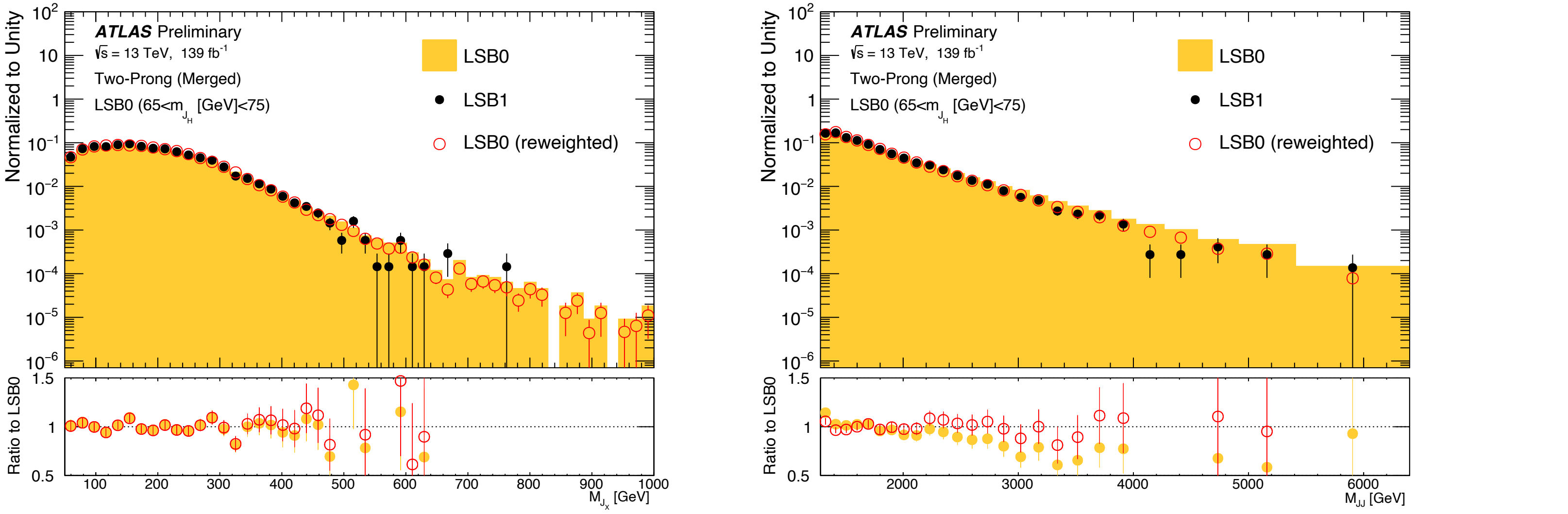
- **SR selection**:
  - Higgs tagging: $D_{Hbb}$ of H candidate > 2.44
  - $75 < m_H < 145$ GeV
- **Background estimation**: reweighted untagged high sideband (HSB0→HSB1)
- **Validation**: low sideband (LSB)



# Background Estimation

- Fully data-driven background estimation (~97% multijet processes)
- Derived from data template in high Higgs mass sideband that fails H tagger score, reweighted to shape in H-tagged region
- Build **DNN** to provide a reweight for each event
  - Train inclusively in X-tagging over variables associated to the Higgs large-R jet (4 vector, 4-vectors of leading & subleading track jets associated to Higgs, # tracks)
  - Minimized on log-likelihood ratio of tagged to untagged regions

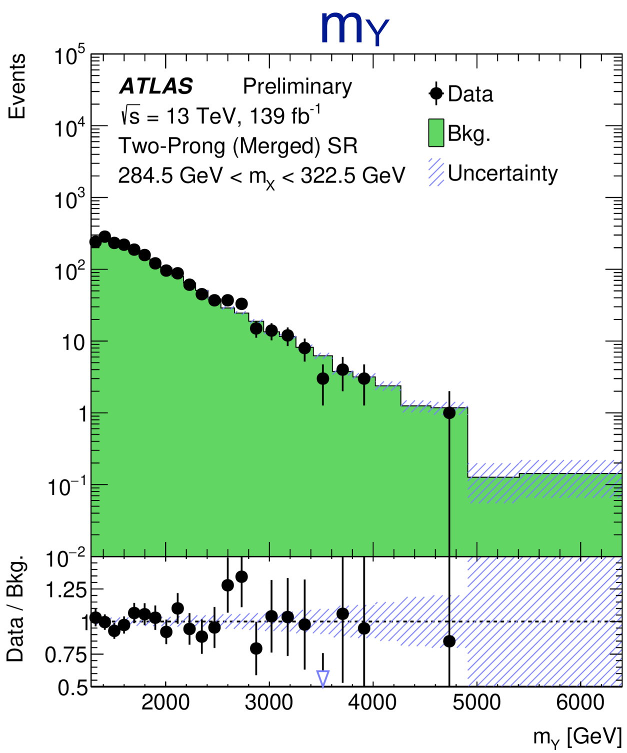
## LSB Validation

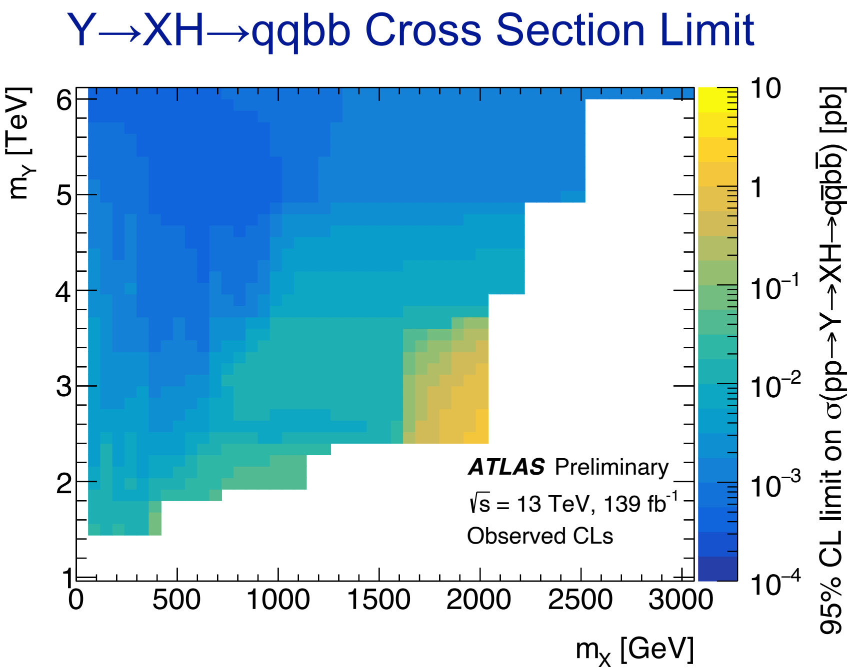



$m(J_X)$ $m_{JJ}$

# Results

- Use BumpHunter for signal model-independence and fit $m_Y$ in overlapping categories of $m_X$
- No significant deviations in anomaly region across $m_X$ bins
  ➡**No interpretation in anomaly region (no signal systematics); limits provided from 2-prong SRs**
- Interpret in nominal X→qq: upper limit at 95% CL on Y→XH→qqbb cross section across signal grid
  - Sensitive up to 6 TeV in resonance mass



# Discussion Topics & Future Outlook

- How to interpret results from unsupervised taggers?
  - We plan to provide cross section limits on signals injected in anomaly region, without signal systematics, only for the purpose of comparison to a supervised approach in this specific analysis context
  - Is it a well-defined task to use anomaly regions to compare sensitivity for a given model to a dedicated supervised search?

- How to provide appropriate analysis preservation?
  - Can upload model files & post-training weights as HEP Data
  - Depending on the tagger, this is still not sufficient to reinterpret with other models
    - Presence of signal model in training dataset could have unspecified consequences on tagger performance

# Conclusions

- Demonstrated an application of the VRNN architecture to classify anomalous jets via data-driven unsupervised learning

- Integrated jet-level anomaly score into ATLAS search
  - Broadened sensitivity to a variety of new physics jet topologies

➡ **First application of unsupervised machine learning to an ATLAS analysis & many others in the works!**

# Backup

# Background Systematic Uncertainties

- All background systematics are determined inclusively in $m_x$, and then applied to each exclusive $m_x$ bin

1. DNN Source Systematic
    - Difference in resulting mJJ distribution due to the choice of training region
    - O(1-10)% effect across mJJ
2. DNN Bootstrap Systematic
    - Statistical error from neural network performance determined via the bootstrap procedure
    - O(1)% effect across mJJ
3. Non-Closure Systematic
    - Determined in the LSB as the difference between reweighted LSB0 and LSB1 data, with smoothing
    - Characterizes additional mis-estimation of data in the VR after determining weighting parameters from the HSB
    - Negligible for low mJJ, O(10)% effect in the tails

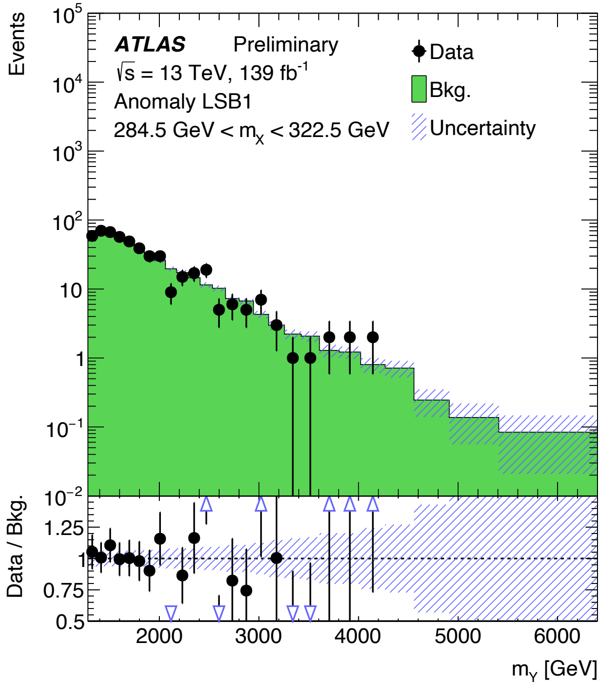
# Signal Systematic Uncertainties

- Flat luminosity uncertainty of 1.7% (as measured with LUCID)
- Jet uncertainties implemented with standard variations from jet/$E_T^{miss}$ CP group
  - Included for both large-R (merged and resolved) and small-R (resolved only) jets
  - Rtrk Baseline, Modeling, Tracking, TotalStat, Closure uncertainties
  - JER Mass and $p_T$ variations
- PDF variation uncertainties
  - ISR/FSR included as flat 3% uncertainty
- XbbSF uncertainties

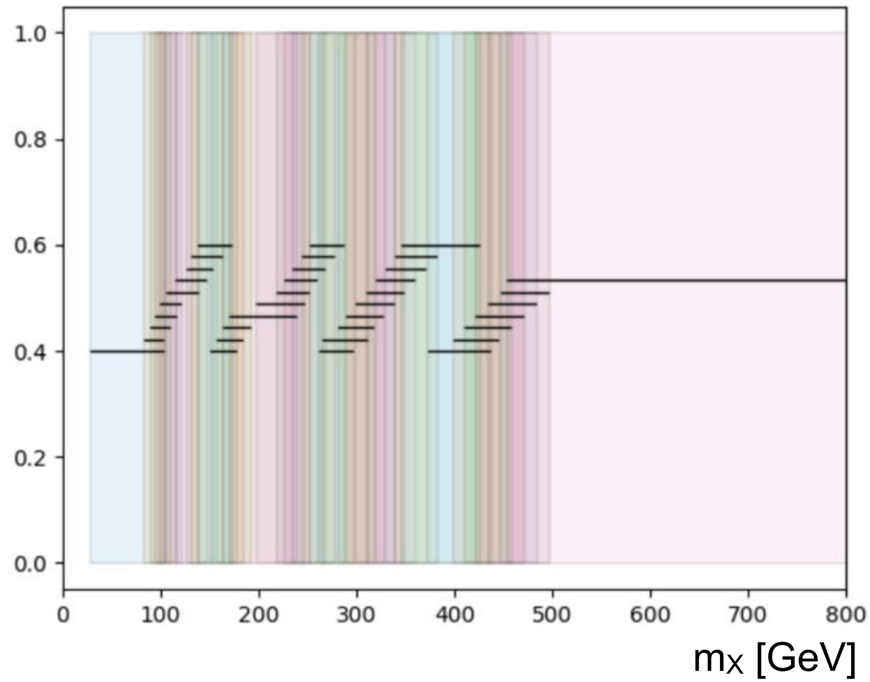
# Statistical Analysis

- Fit $m_Y$ across overlapping categories of $m_X$
  - Bins chosen based on signal mass resolution
- Use BumpHunter as signal model-independent "excess finder" [1101.0390]
  - No significant (p-val < 0.01) excess across $m_X$ bins in the LSB VR

## Post-fit LSB

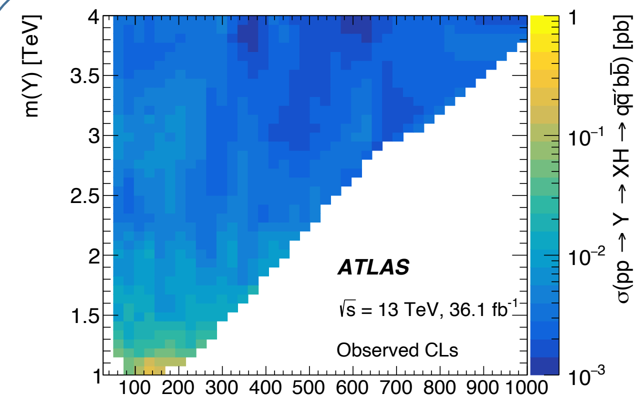



## $m_X$ Windows



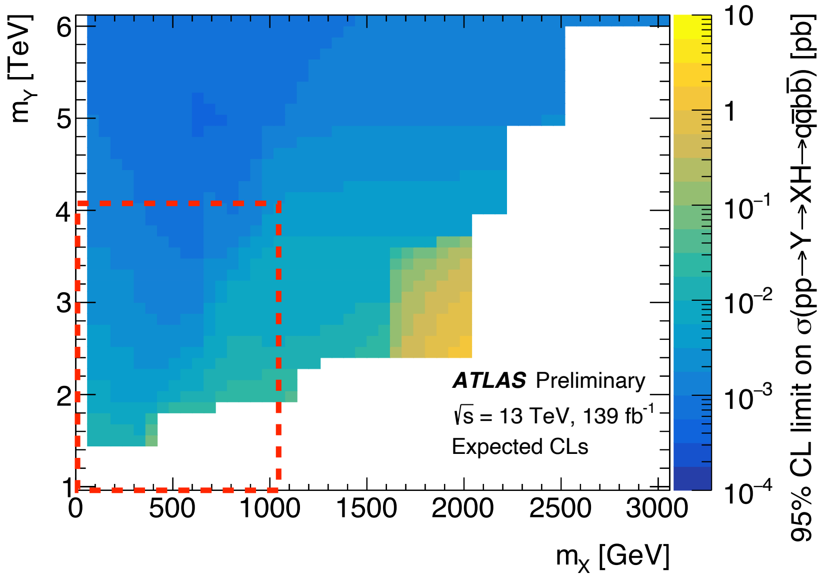
# 2-Prong Interpretation

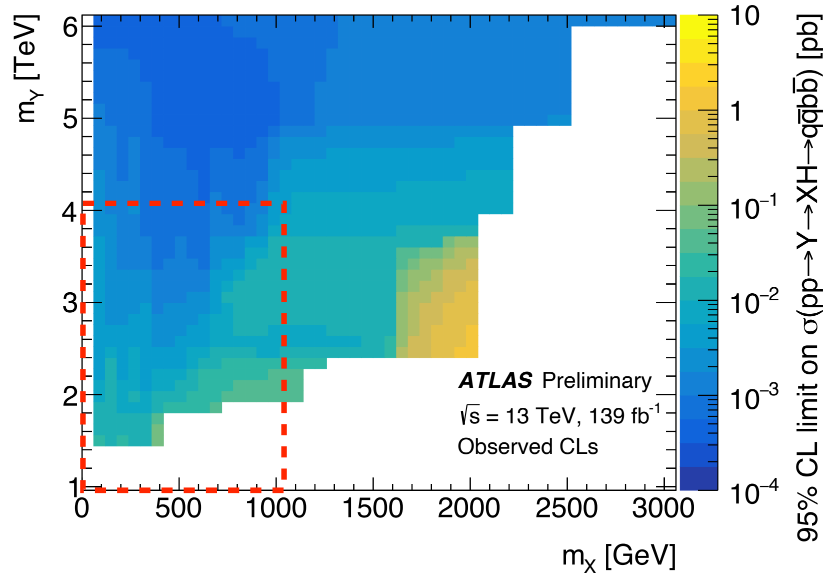
- Results for Y→XH→qqbb given as upper limit at 95% CL on production cross section across signal grid
- With respect to previous result:
  - Improved limit in highly boosted regime by ~10x
  - Increased upper resonance mass sensitivity from 4 to 6 TeV
  - Added coverage of unexplored resolved X decay phase space




36.1 fb-1 exclusion result

= previous analysis coverage

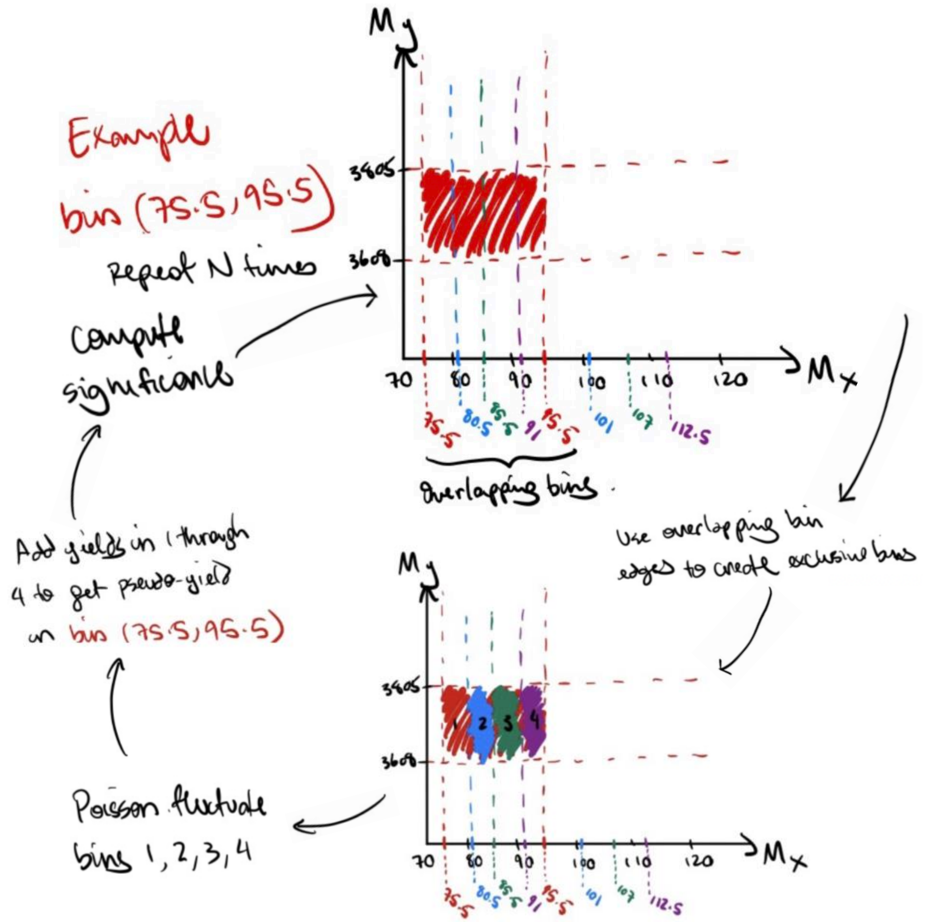


# Global Significance Calculation

- Goal: Determine an overall global significance – difficulty arises from overlapping $m_x$ windows

- 27 Exclusive $m_Y$ bins and 57 overlapping $m_X$ bins = 1539 analysis bins
    - Make exclusive $m_X$ bins from each overlapping window
    - Draw N events in each exclusive $m_X$ bin where N is drawn from a Poisson distribution whose mean is the expected background yield in the exclusive bin
    - Sum the yields from each exclusive bin to arrive at the yield for each analysis bin
    - The p-value is determined for each analysis bin, and the maximum significance recorded
    - This process is repeated N times, where N is the total number of events in the SR
- The global p-value is the fraction of toys with maximum significance greater than that observed in data



# Comparing to VV Searches

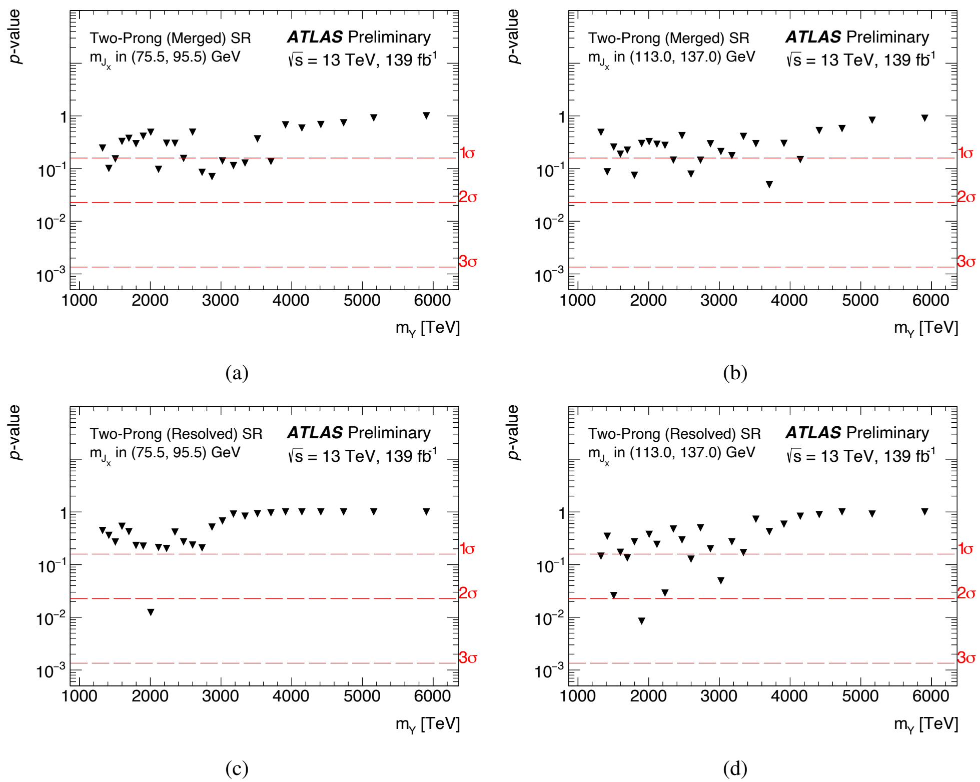
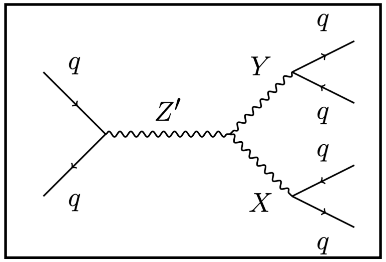



Figure 10: The $p$-value per $m_Y$ bin for both two-prong SRs, calculated using only statistical uncertainties. Two $m_X$ bins are shown, [75.5, 95.5] and [113.0, 137.0] GeV, which corresponds to a window containing the $W/Z$ and Higgs boson mass respectively. Events are thus split into merged $W/Z$ window (a), merged Higgs window (b), resolved $W/Z$ window (c), and resolved Higgs window (d). The background is determined by a background-only fit to the data with all statistical and systematic uncertainties included. In both $m_X$ windows, the $p$-value approximates a constant value of 0.5 for the high $Y$ mass region of the resolved SR, as this region of phase space is far more likely to produce a highly boosted $J_X$ that falls in the merged SR selection.

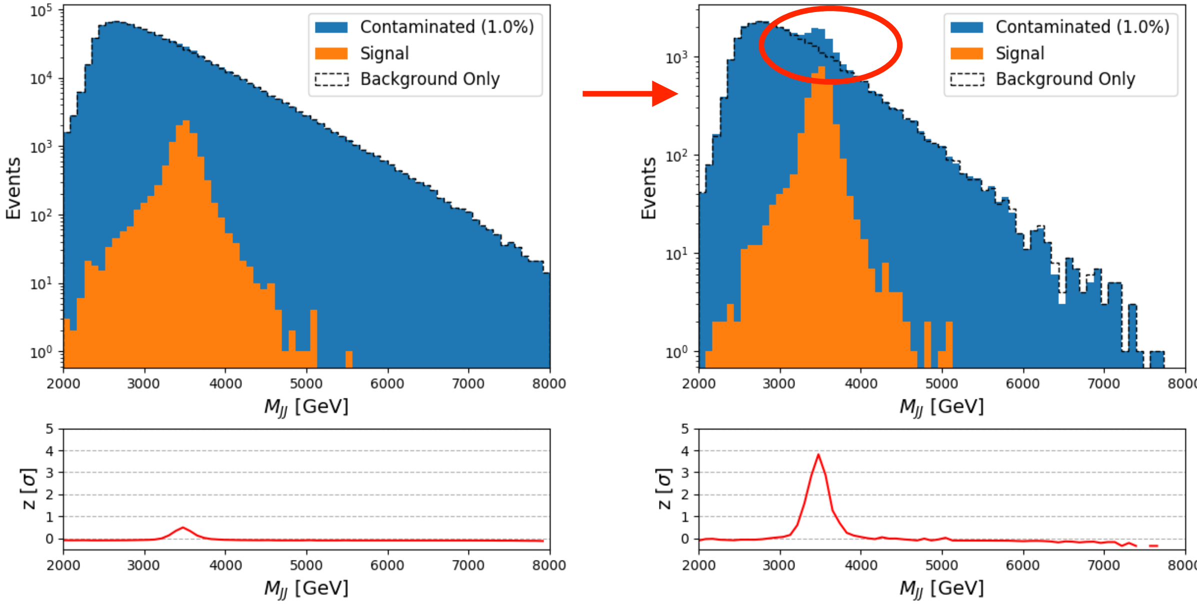
# VRNN in the LHC Olympics

- Developed in simulation via the LHC Olympics community anomaly detection challenge [2101.08320]
- LHC Olympics dataset: Pythia generated + Delphes detector simulation (no pileup)
- **Signal:** 3.5 TeV Z' → 500 GeV X + 100 GeV Y
  - Two substructure hypotheses: 2-pronged and 3-pronged X/Y decays
- Reconstruction = two large-radius (R=1.0) jet, leading $p_T > 1.2$ TeV

➡VRNN paper published using this dataset [2105.09274]



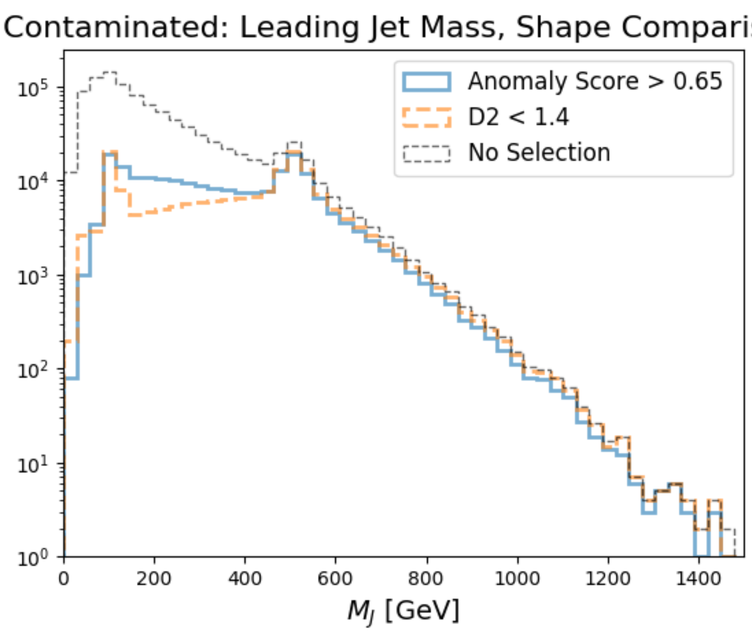

No Selection

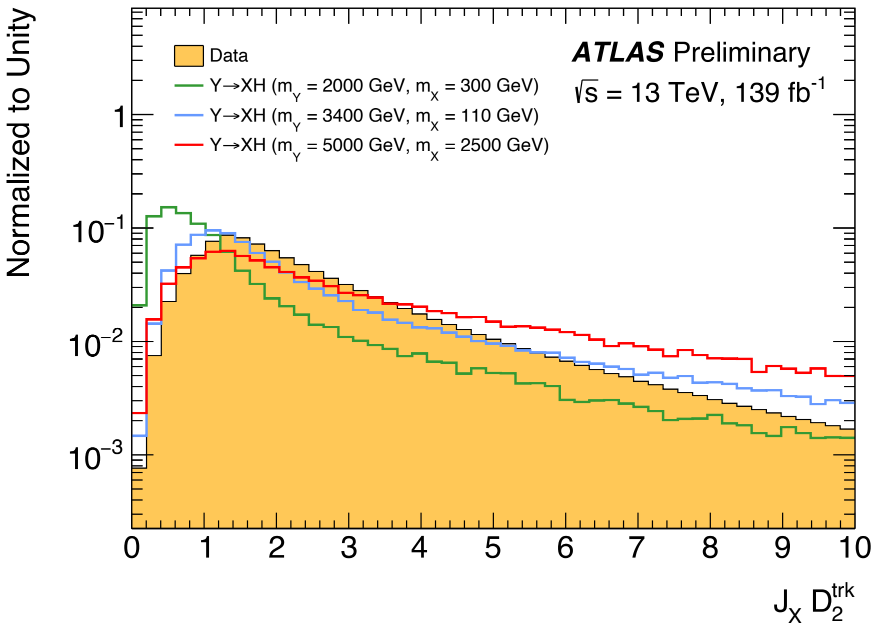



Cut on VRNN anomaly score

# Comparison to D2

- Dataset = 2-prong % contaminated
  - Selections: D2 < 1.4 / AS > 0.65 (equivalent background rejection)
  - AS creates less mass sculpting than substructure variables
- In Y→XH→qqbb, cut on $D2_{trk}$ < 1.2 (merged) or > 1.2 (resolved)



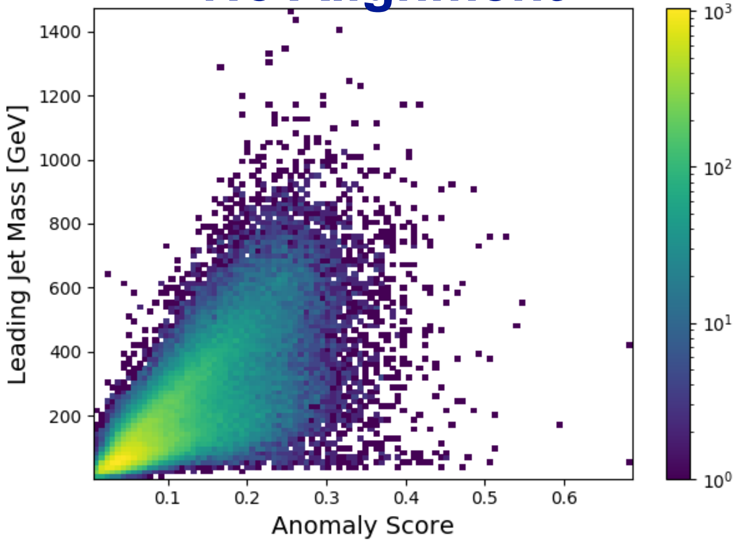
# Alignment

- **Goal**: remove mass and pT information from input jets to avoid tagging on kinematics alone
- **Procedure:**
  1. Rescale each jet to the same mass
  2. Boost each jet to the same energy
  3. Rotate each jet to the same η/Φ orientation
- **Result:** anomaly score far less correlated with mass in background jets

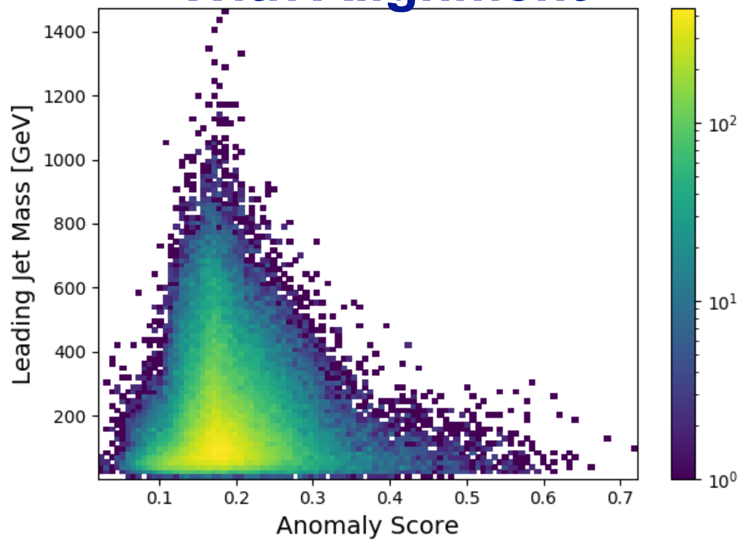
**Start**
Boost jet in $z$ direction until $\eta_{Jet} = 0$
Rotate jet about $z$ axis until $\phi_{Jet} = 0$
Rescale jet four-vector such that $m_{Jet}$ =0.25 GeV
Boost jet along its axis until $E_{Jet}$ =1 GeV
Rotate jet about $x$ axis until hardest constituent has $\eta_1 = 0, \phi_1 > 0$
**if** *Any constituents have* $\Delta R > 1$[a] **then**
  Remove all constituents with $\Delta R > 1$
  Rebuild jet with remaining constituents
  Repeat from start
**else**
  continue
**end**
**if** *Number of constituents* $> 20$ **then**
  Keep up-to the first 20 constituents, ordered in $p_T$
  Rebuild jet with remaining constituents
  Repeat from start
**else**
  continue
**end**
Reflect constituents about $\phi$ axis such that the second hardest constituent has $\eta_2 > 0$

[a] $\Delta R$ is computed as $\sqrt{\eta^2 + \phi^2}$ for each constituent, where $\eta$ and $\phi$ are measured relative to the x axis.

## No Alignment




## With Alignment



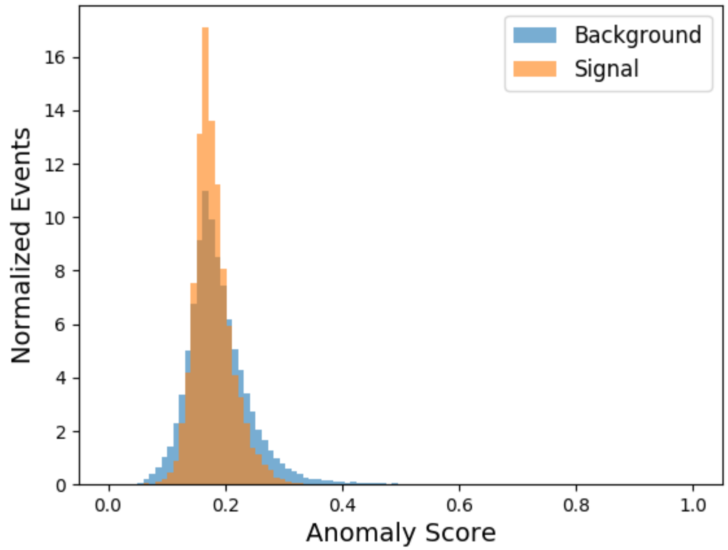
# Sequence Ordering

- In a recurrent architecture, apt sequence modeling of jets (eg. order of constituents) can highlight importance sequence features & boost performance
- Select **$k_t$-distance** ordering to highlight substructure: $n^{th}$ constituent has highest $k_t$-distance relative to previous, starting with highest pT constituent

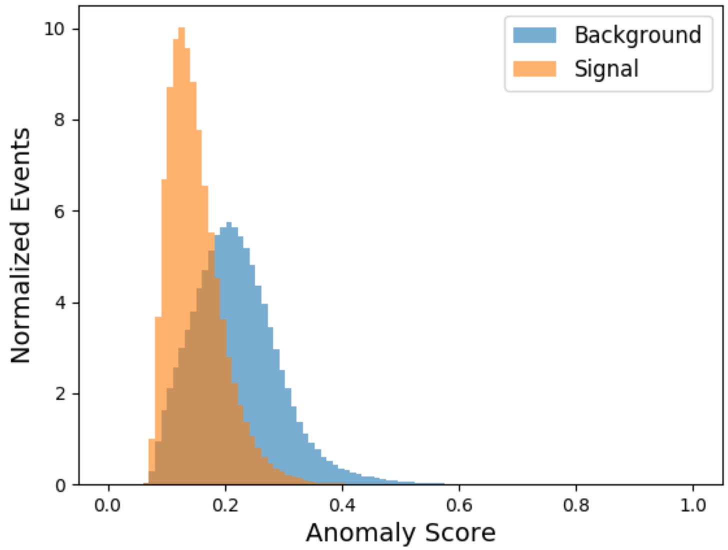
$$c_n = max(p_{Tn} \Delta R_{n,n-1})$$

- **Result:** better separation of two-prong signal from diffuse QCD background than pT-sorting

**pT-sorted**




**$k_t$-sorted**



# Analysis Application

- Compute anomaly score for each jet
  - Higher KL divergence = higher loss = lower anomaly score
  - ➡ Transform such that higher AS corresponds to more anomalous jets

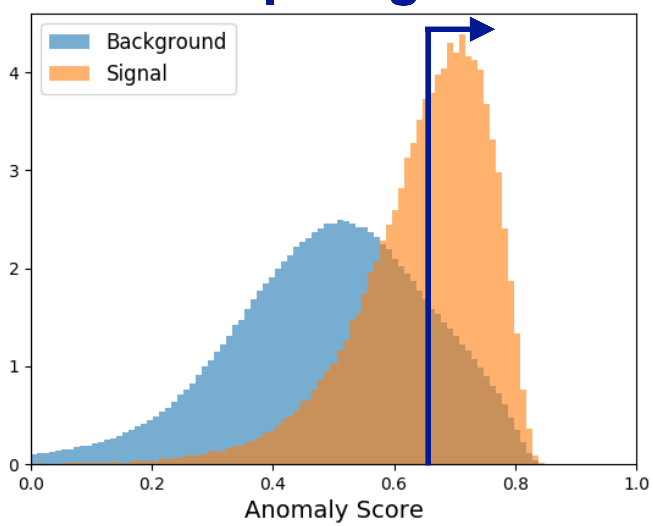
**Jet Anomaly Score**

$$\rho = 1 - e^{-\overline{D_{KL}}}$$

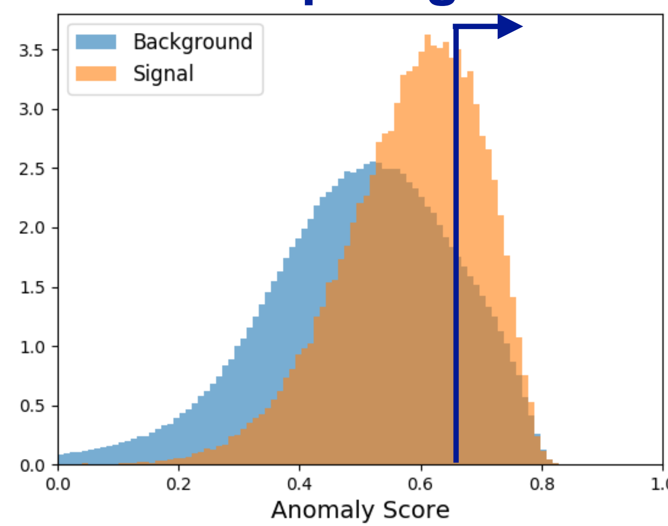
$$\rho' = 1 - \left(\frac{\rho}{2\overline{\rho}}\right)$$

- **Analysis strategy:** cut-and-count on $\rho' > 0.65$ as sole signal region selection & test signal significance in bins of mJJ

**2-prong**




**3-prong**



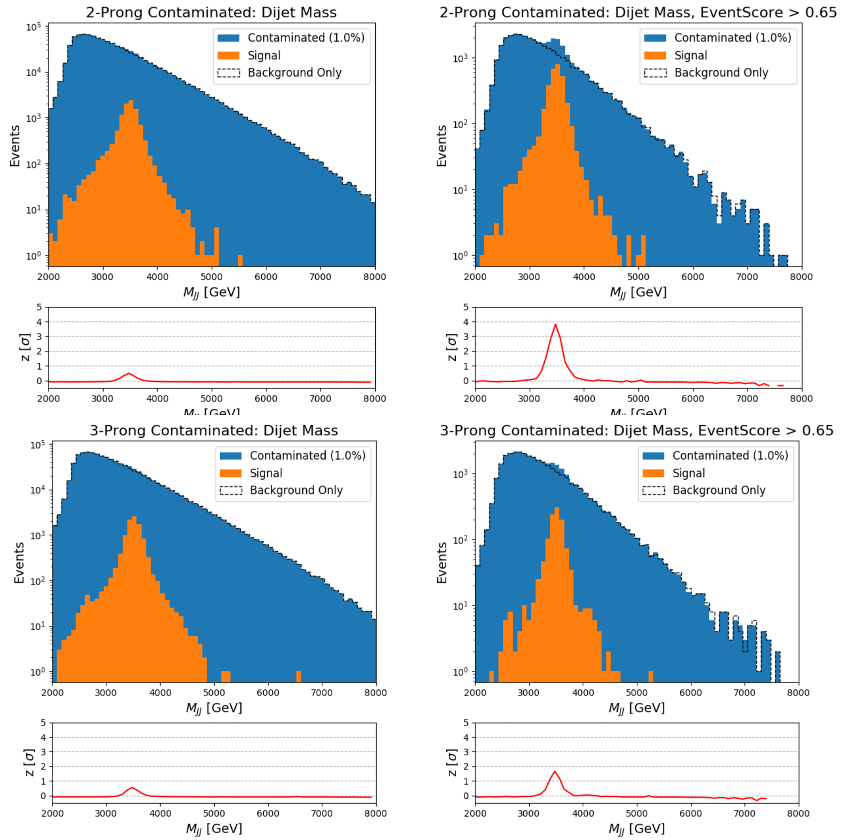
# Results: 2 and 3-Prong Signal

- Perform bump hunt on $m_{JJ}$ with selection on *Event Score* = max of two leading jet Anomaly Scores
- Dataset = background + 1% signal contamination

➡ Enhance a **0.5σ two-prong signal excess to 4.0σ** solely from an Event Score cut at 0.65

➡ Enhance a **0.5σ three-prong excess to 1.5σ** using the same score

**No cut** **Event Score > 0.65**



# Performance vs. Contamination

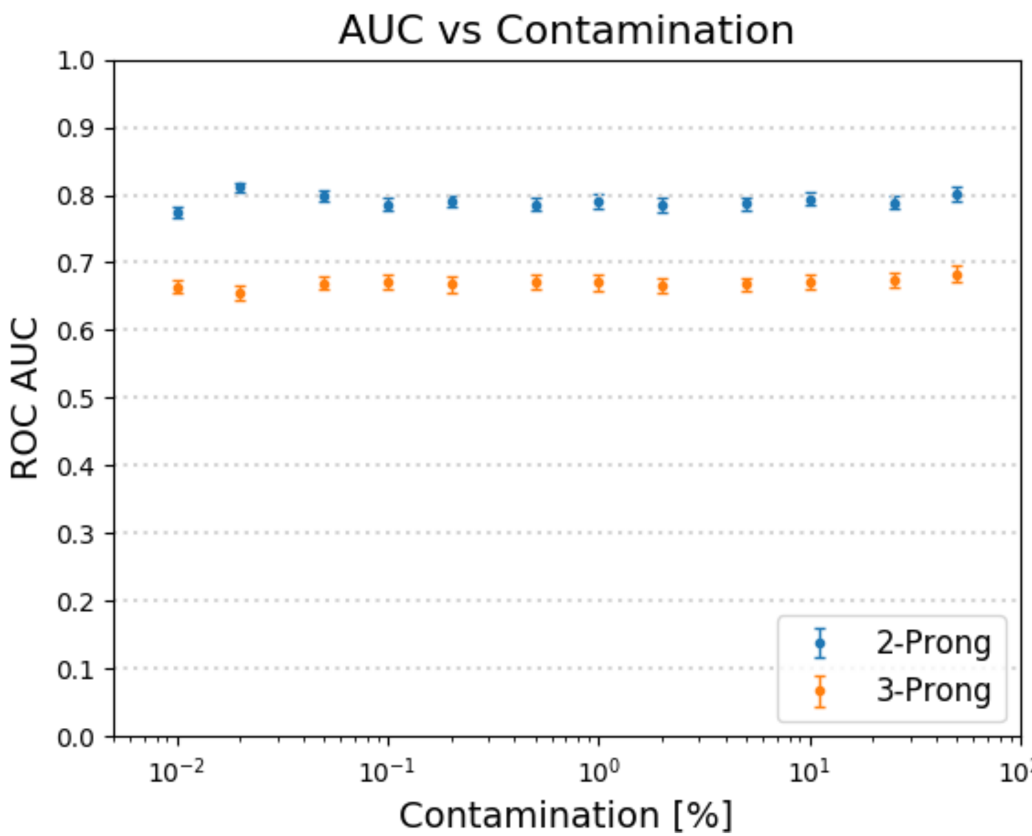



**Figure 11**. ROC AUC vs. percent signal contamination in training datasets. The performance of the Anomaly Score is consistent across a wide range of contamination levels.

# Performance vs. Training Time

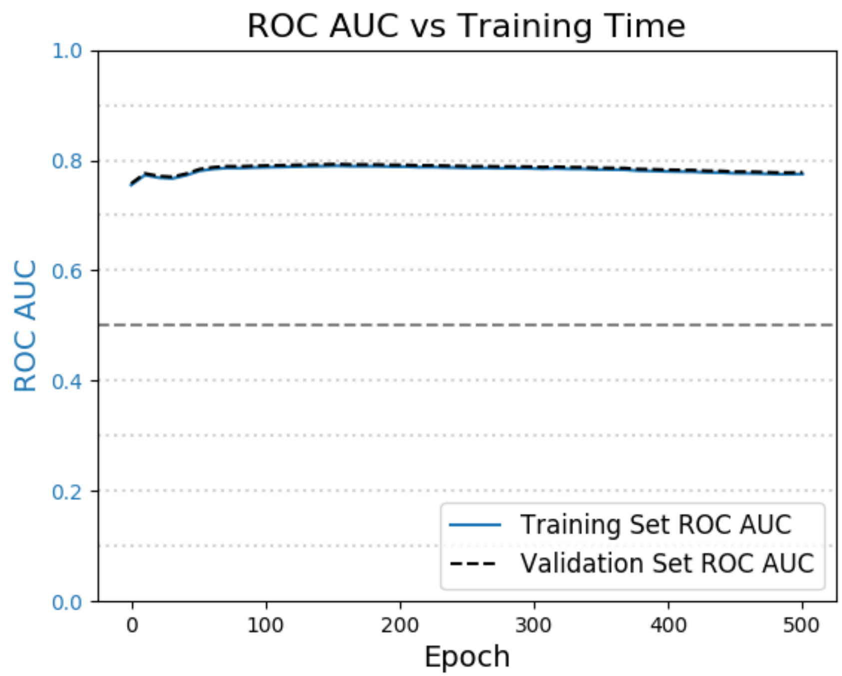



**Figure 6**. Area Under the Curve (ROC AUC) vs. training time in epochs on a 1% signal-contaminated dataset. The VRNN reaches an optimal performance quickly, and retains this performance over a long training period. The difference in performance between the training and validation sets is a result of the former containing elements of signal.