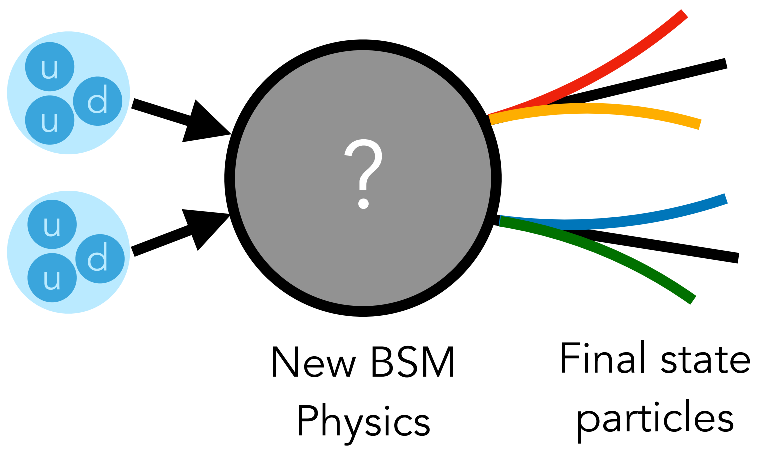
Compare & Contrast: Anomaly Detection in ATLAS

Gabriel Matos
Columbia University
June 16, 2025 - **AD4HEP 2025 Workshop**

Motivation for Model Generic Searches

- Beyond the Standard Model (BSM) searches in colliders follow a general recipe

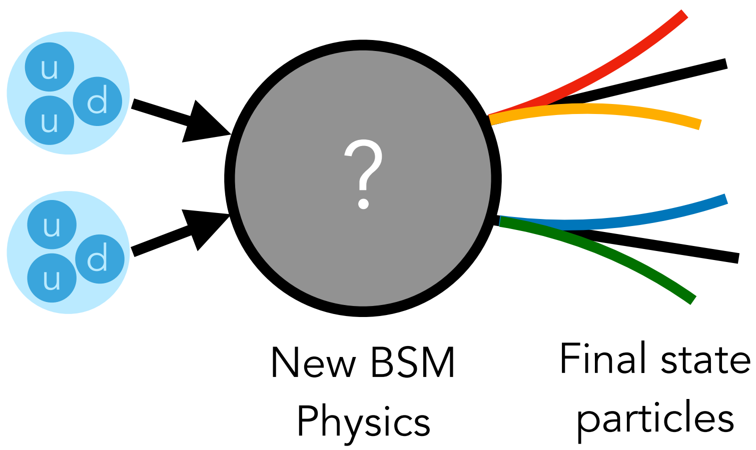


1. Pick a model for some final state signature
2. Determine relevant parameters for the model
3. Design selections on your observables to enhance the signal over background

ATLAS summary plots

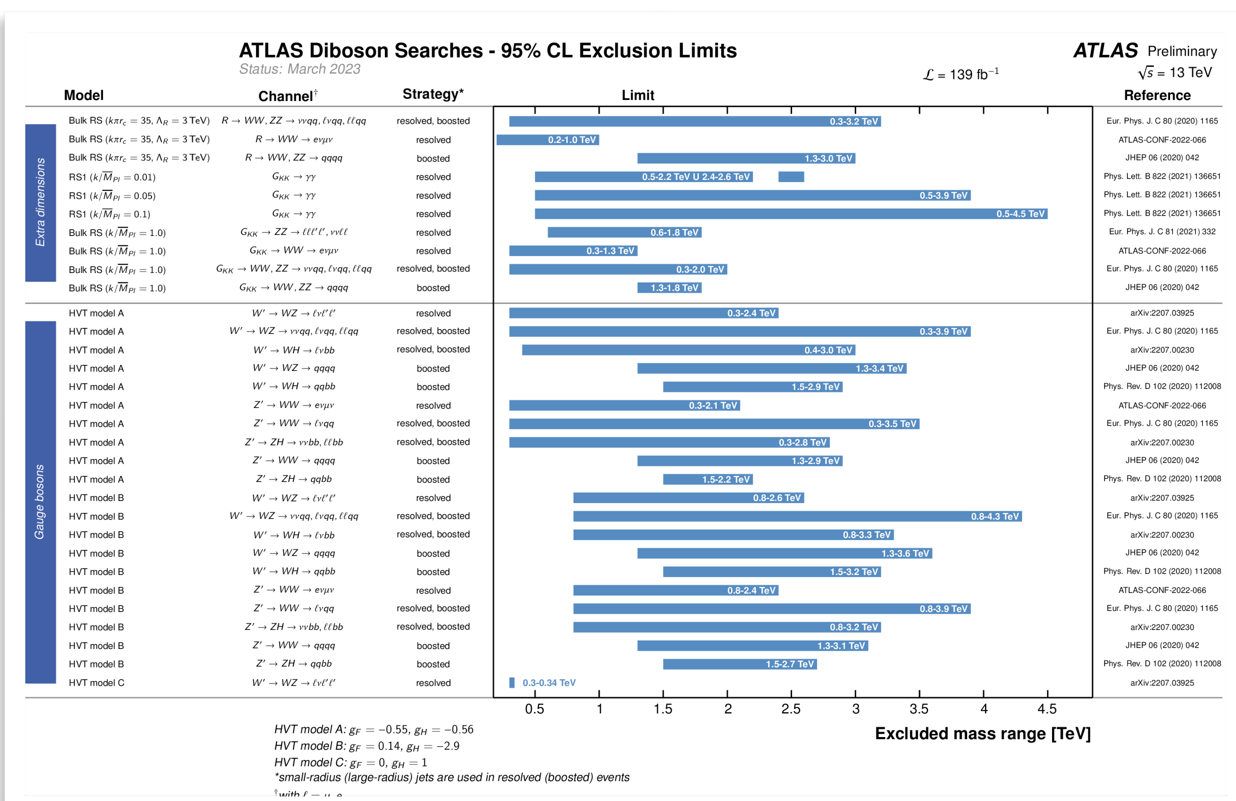
Motivation for Model Generic Searches

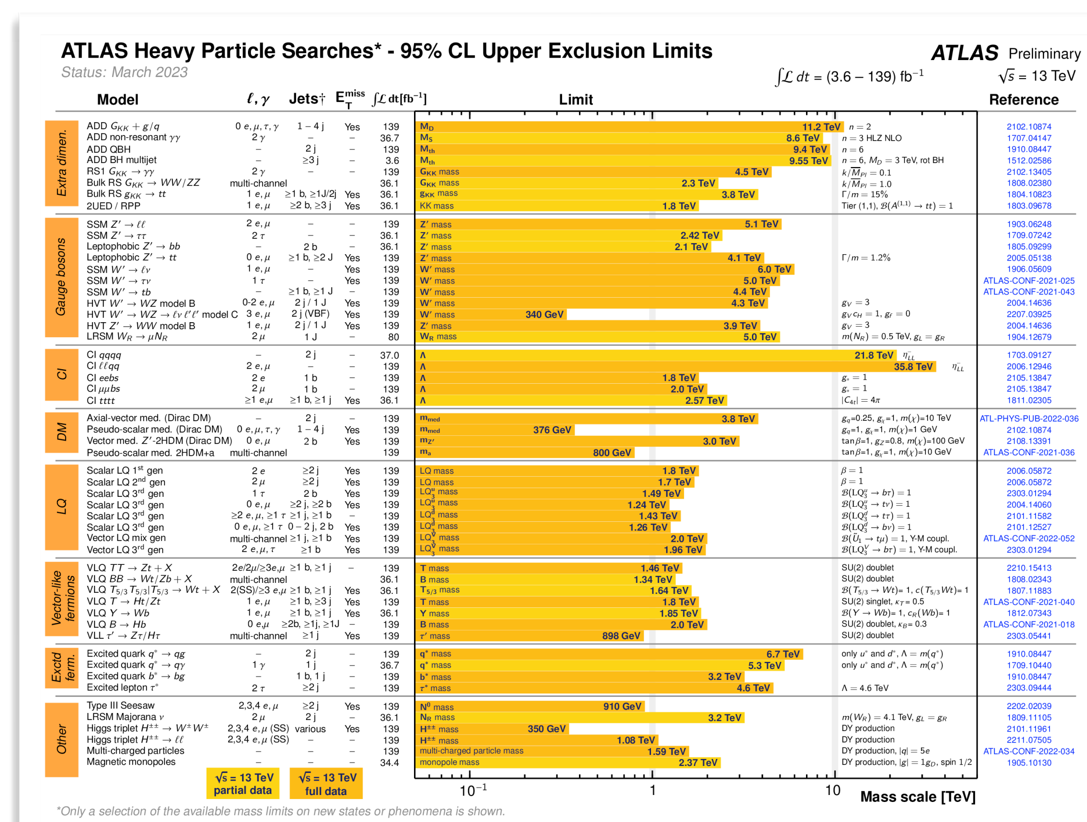
- Beyond the Standard Model (BSM) searches in colliders follow a general recipe



1. Pick a model for some final state signature
2. Determine relevant parameters for the model
3. Design selections on your observables to enhance the signal over background

Problem: way too many models/signatures to develop independent searches for!

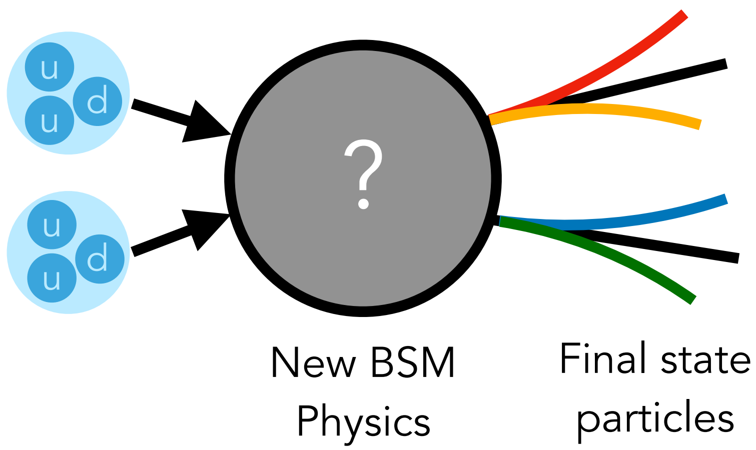




ATLAS summary plots

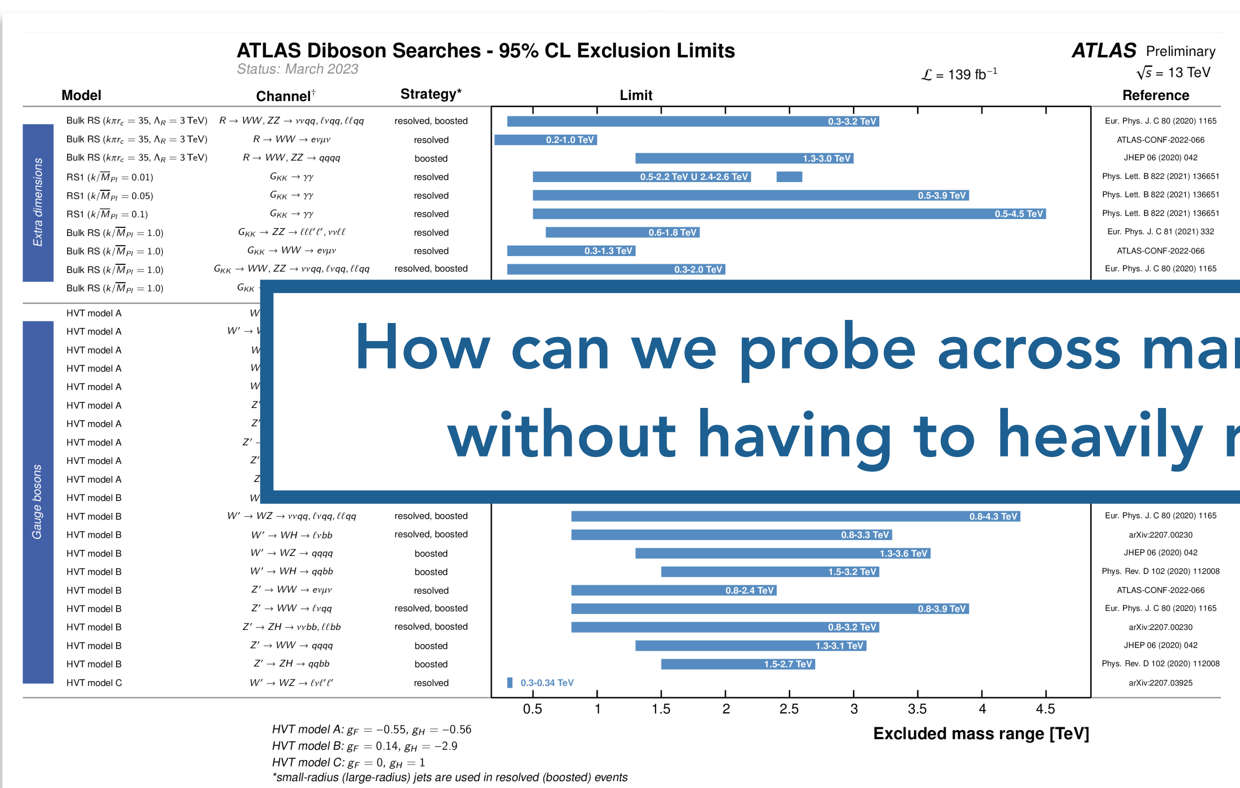
Motivation for Model Generic Searches

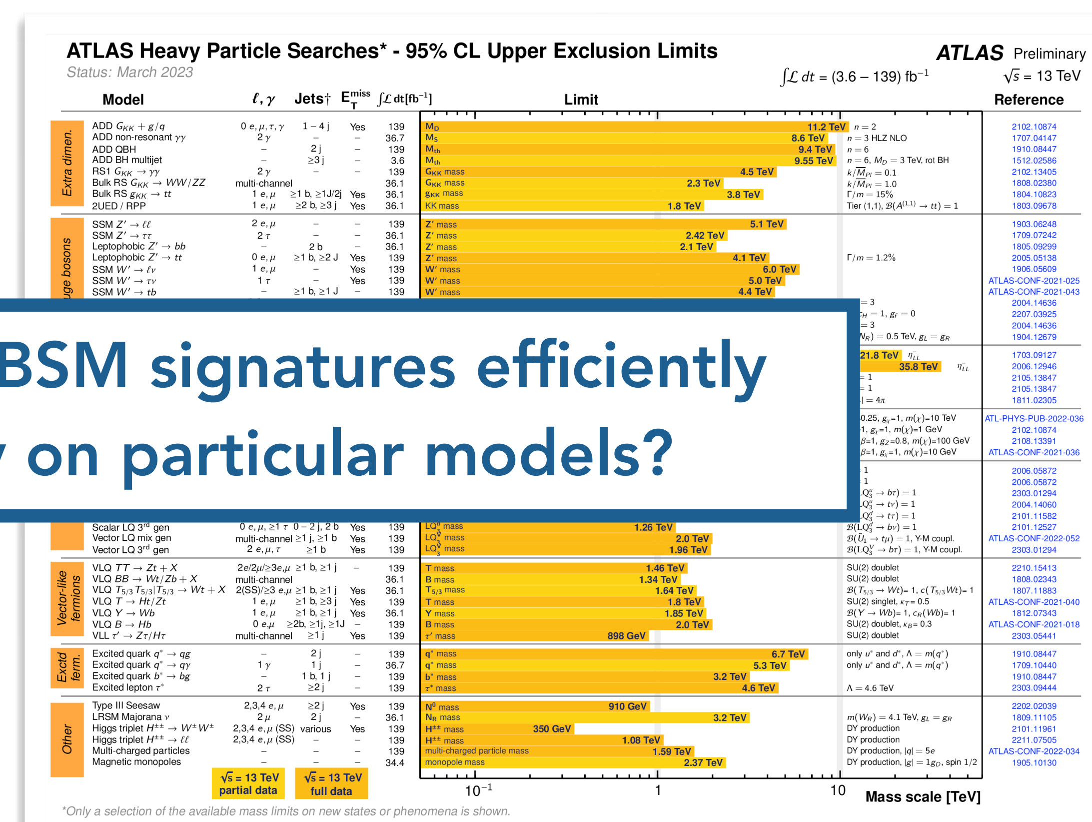
- Beyond the Standard Model (BSM) searches in colliders follow a general recipe



1. Pick a model for some final state signature
2. Determine relevant parameters for the model
3. Design selections on your observables to enhance the signal over background

Problem: way too many models/signatures to develop independent searches for!

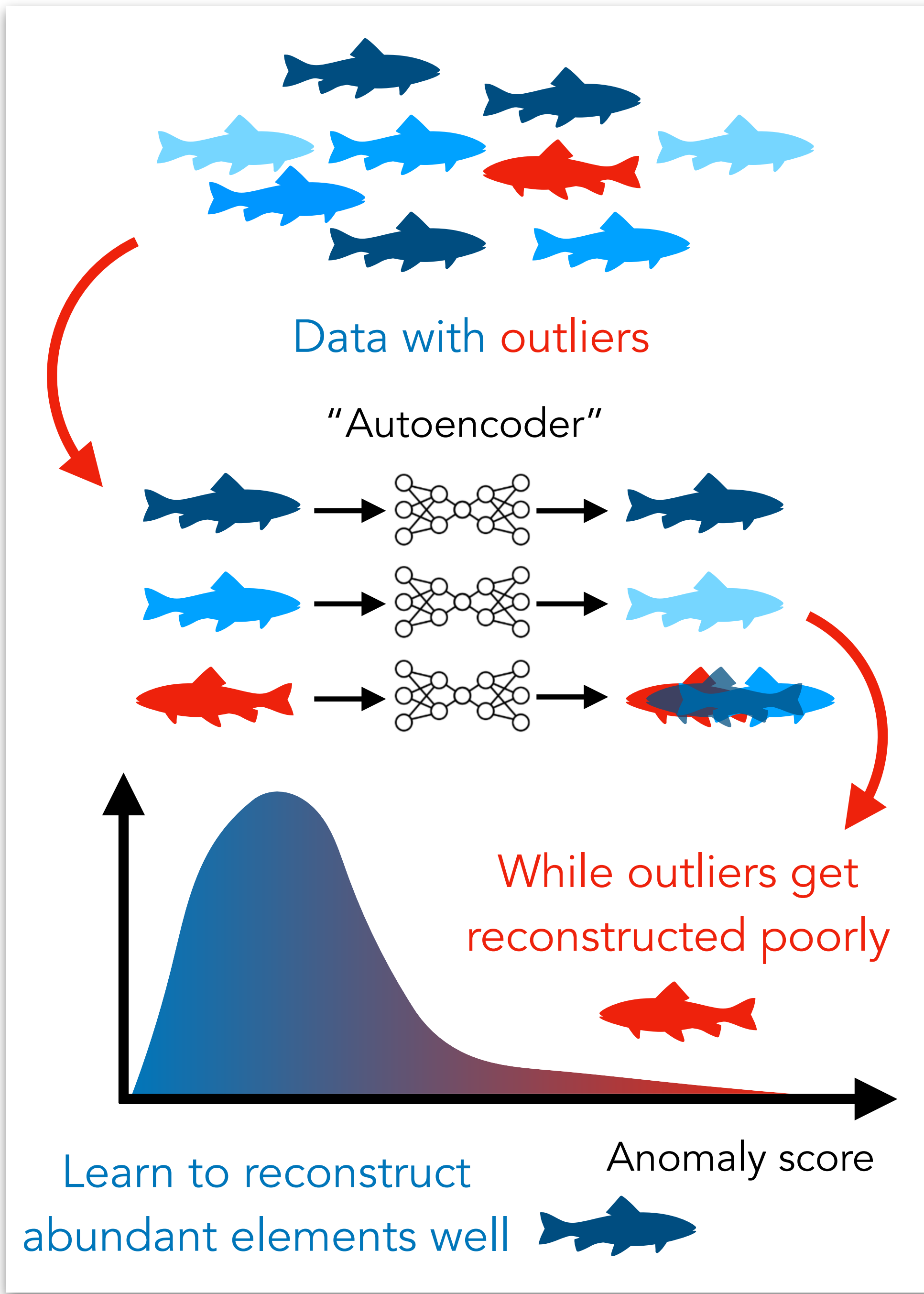


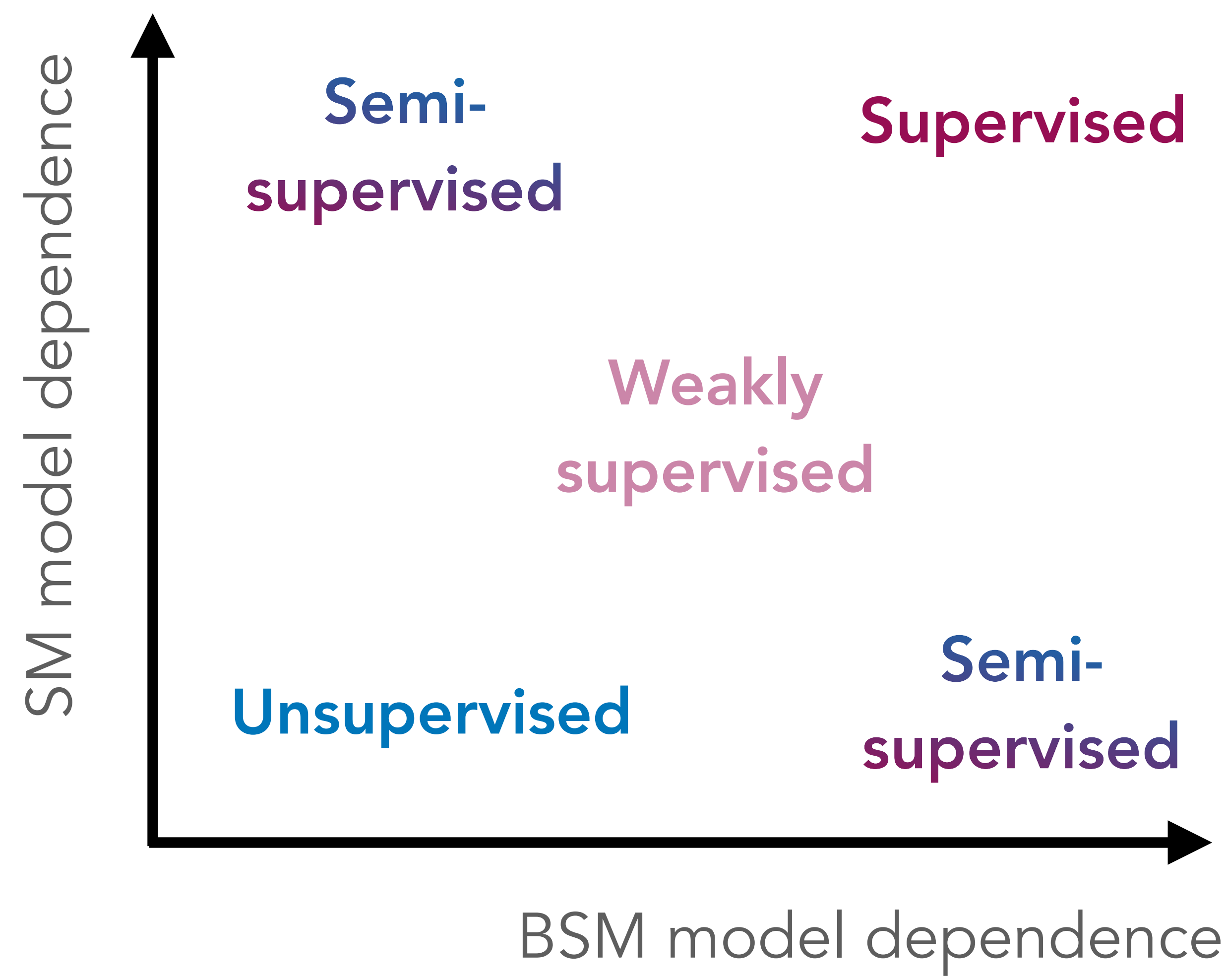


How can we probe across many BSM signatures efficiently without having to heavily rely on particular models?

Anomaly Detection (1)

- Machine learning-based **anomaly detection** (AD) can be used to train models to detect anomalous features in a dataset inconsistent w/ a background-only model





arXiv:2112.03769

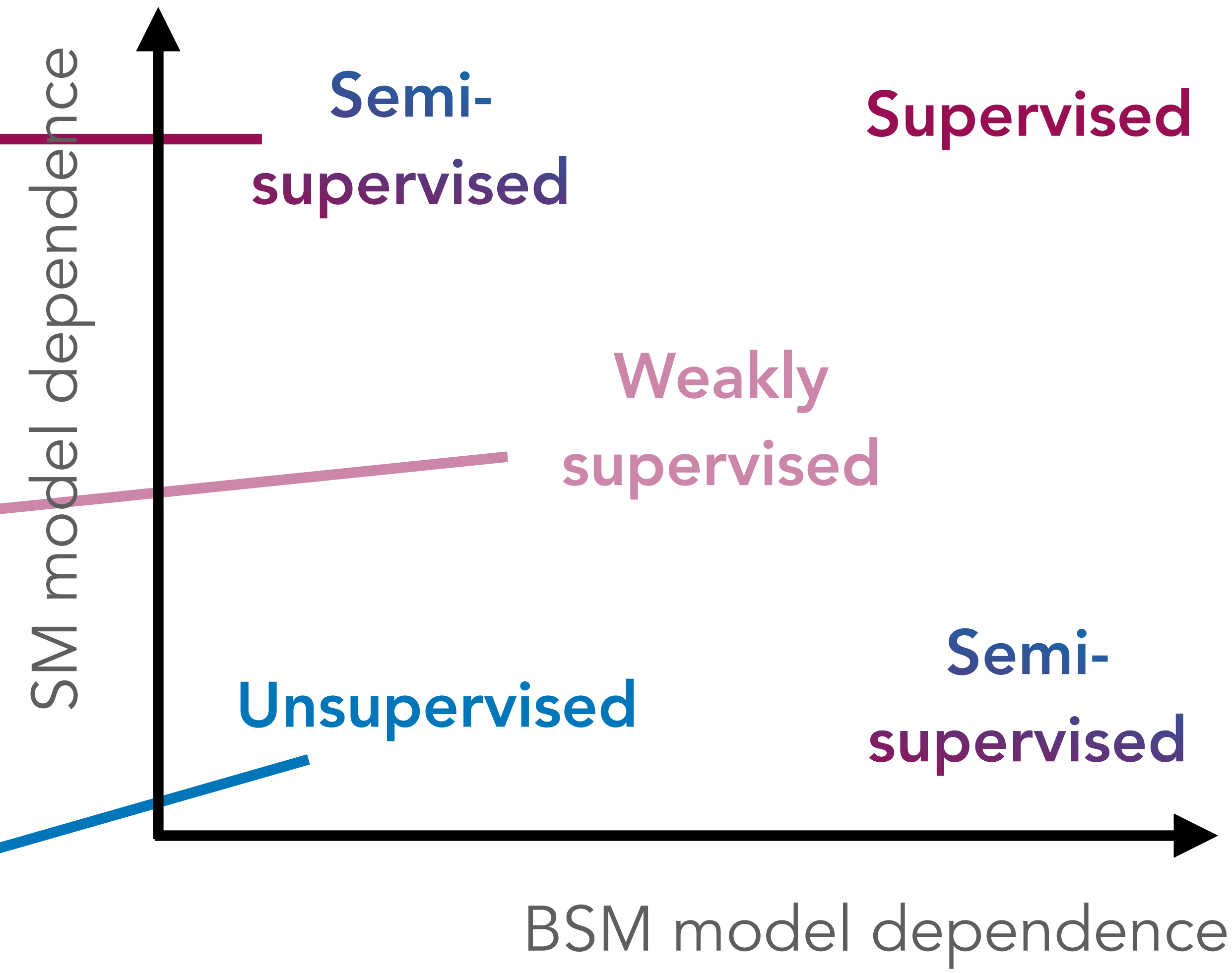
Anomaly Detection (2)

- Machine learning-based **anomaly detection** (AD) can be used to train models to detect anomalous features in a dataset inconsistent w/ a background-only model

Semi-Supervised
Detect outliers by relying on some signal priors

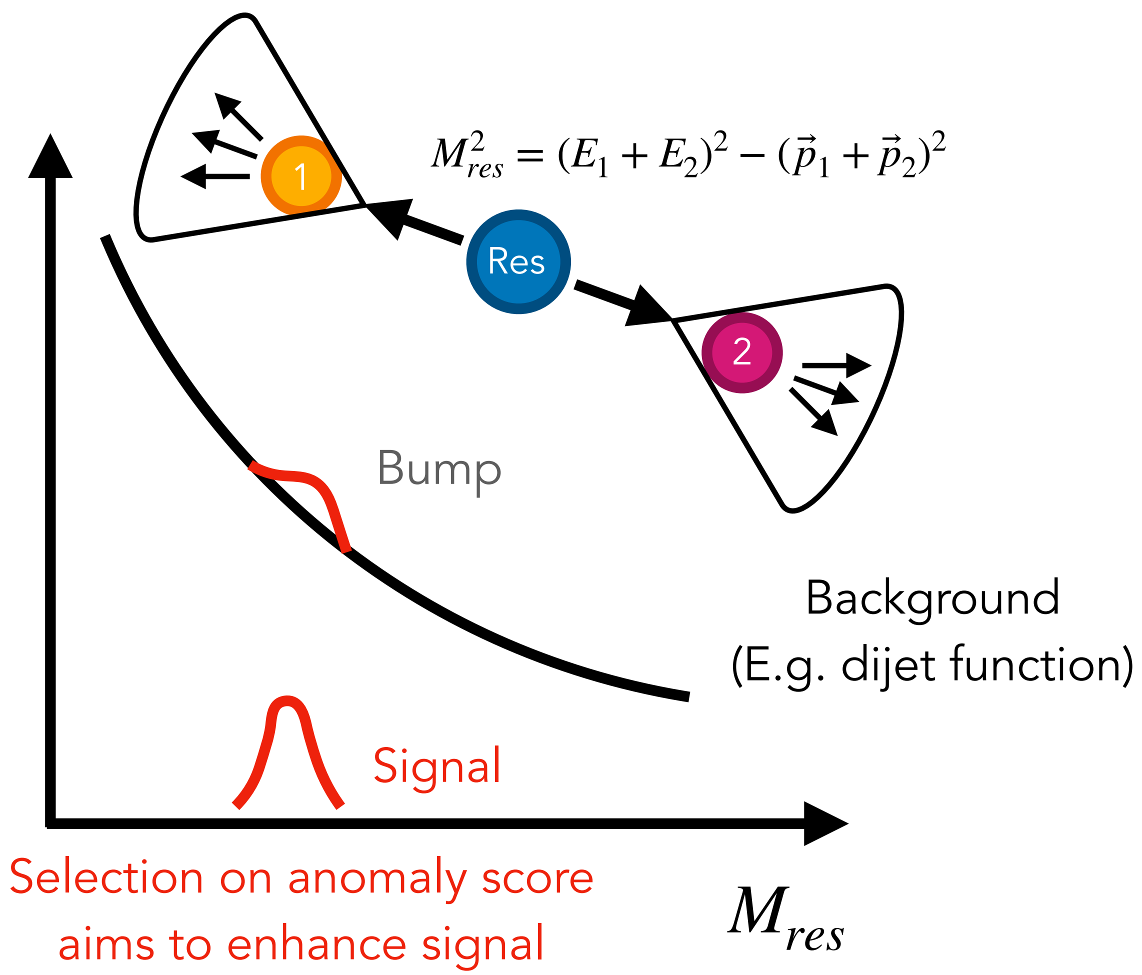
Weakly-Supervised
Detect outliers with a combination of labeled and unlabeled data

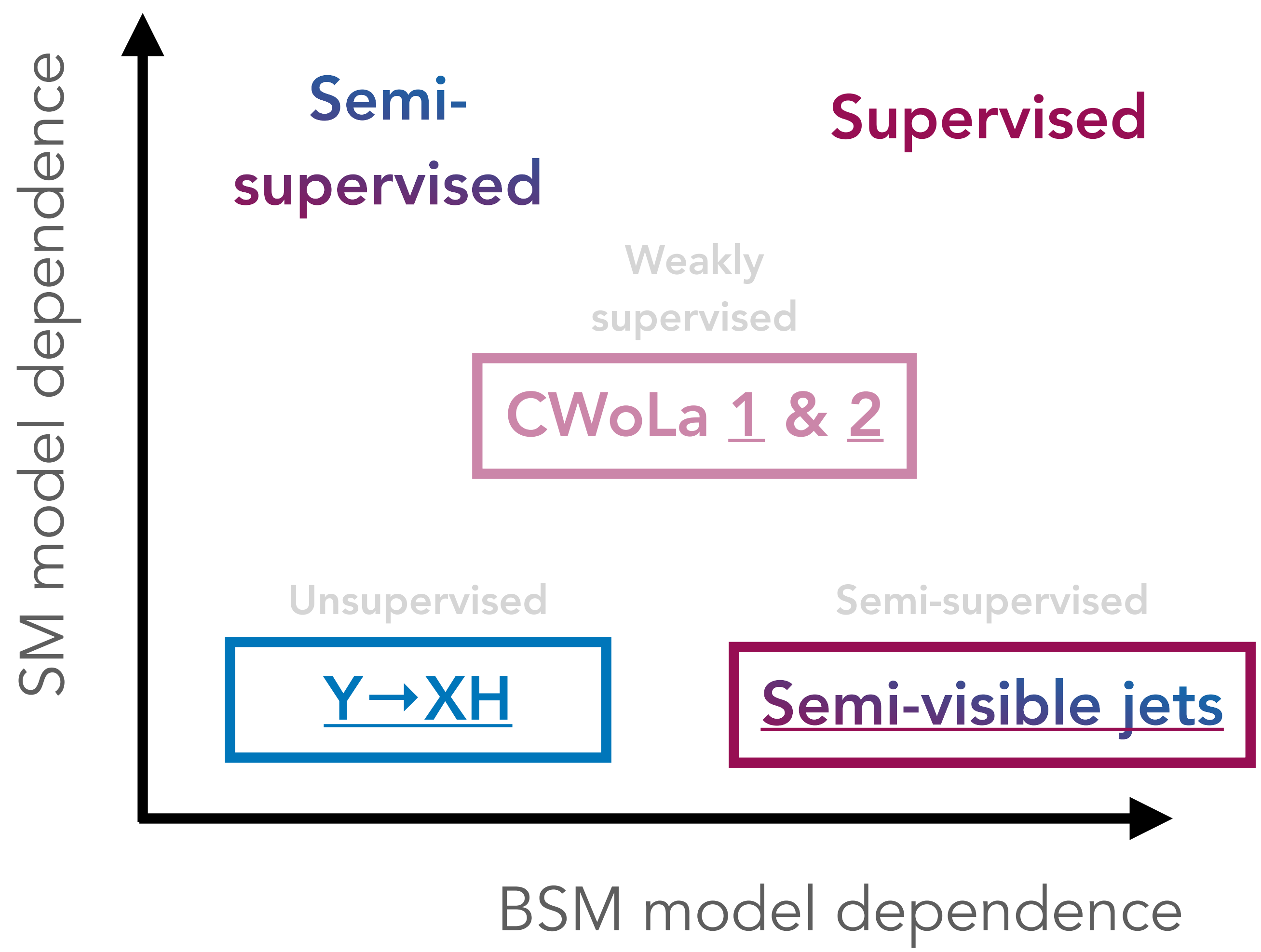
Unsupervised
Detect outliers by training directly on data



Anomaly Detection in ATLAS

- So far, all AD results in ATLAS have been searches for new resonances on a 2-body invariant mass spectrum, so called "bump hunts"

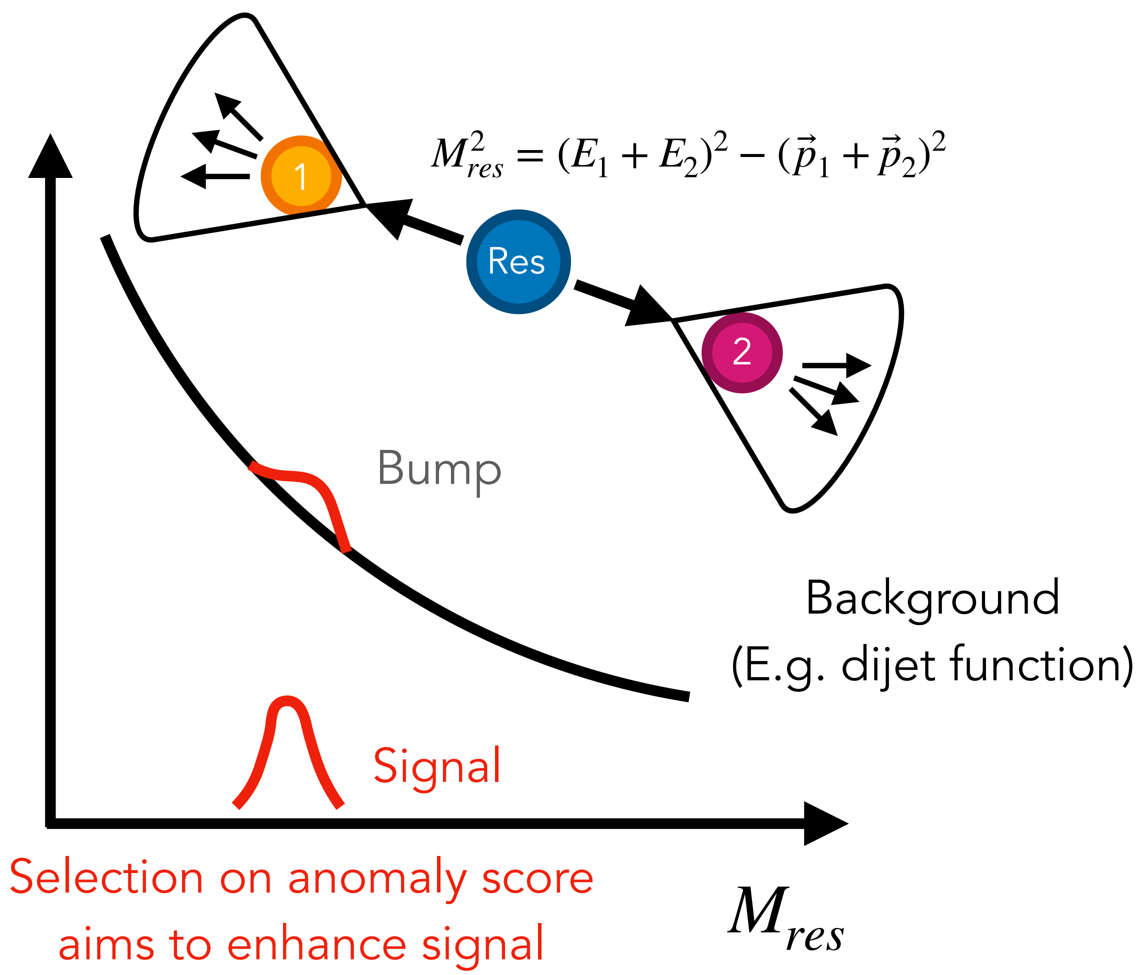


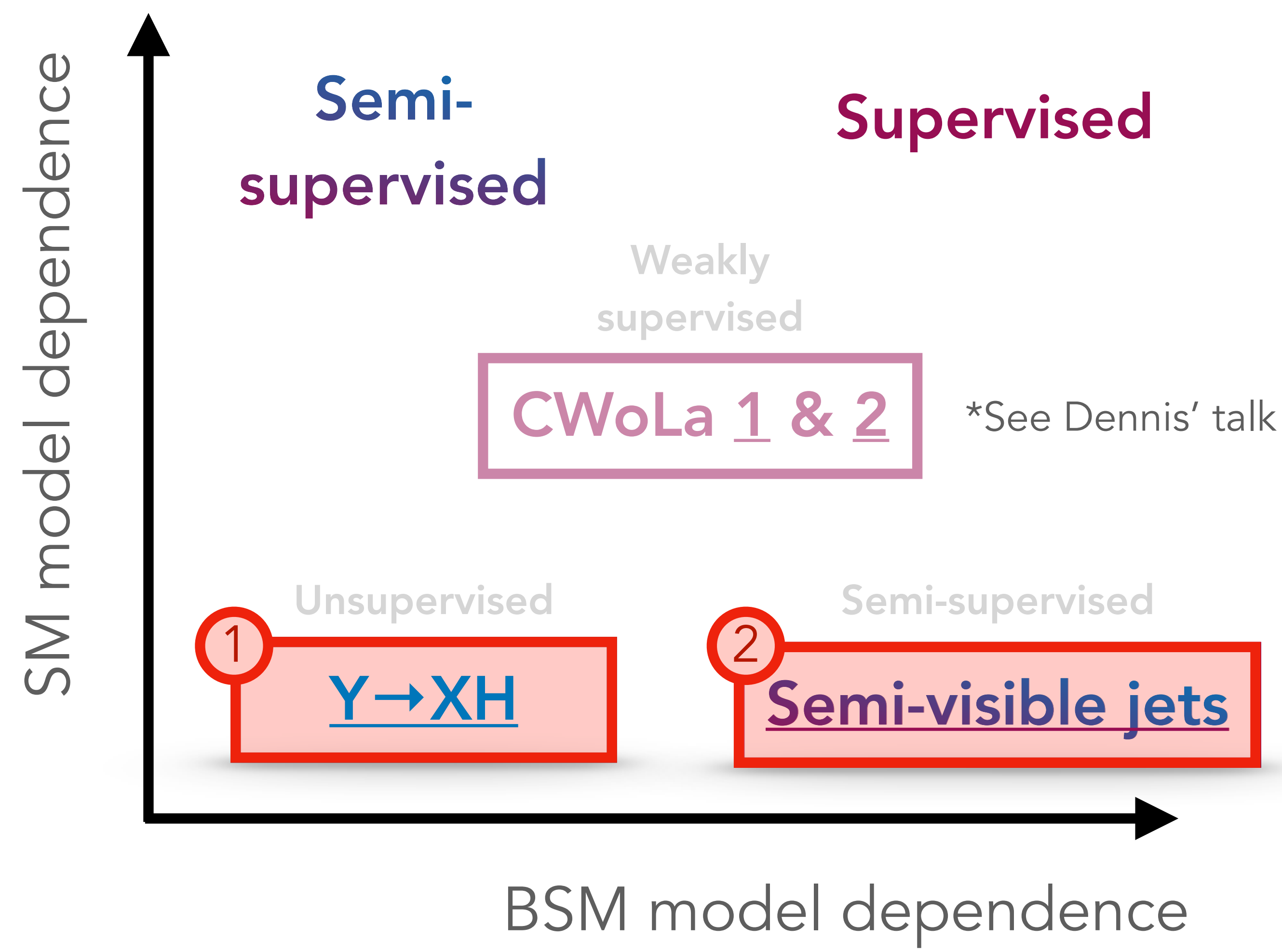


- We'll compare & contrast how these searches tackled key questions in collider-based AD, including input modeling, level of supervision, and result reporting

Anomaly Detection in ATLAS

- So far, all AD results in ATLAS have been searches for new resonances on a 2-body invariant mass spectrum, so called "bump hunts"

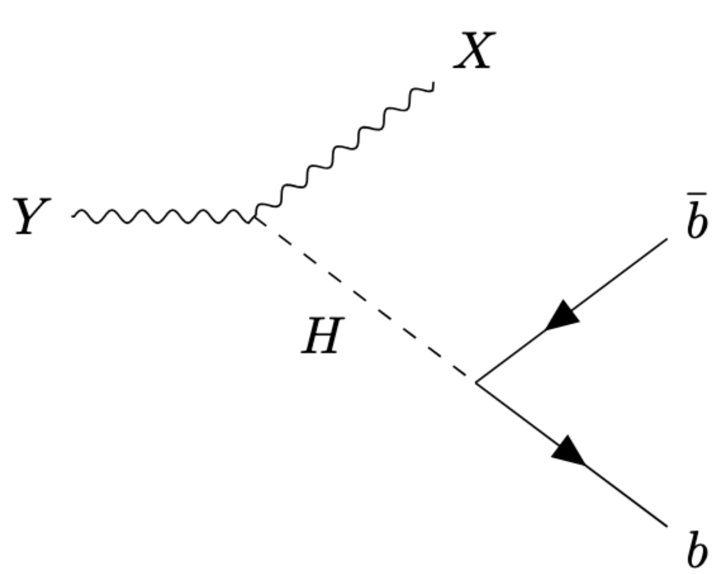


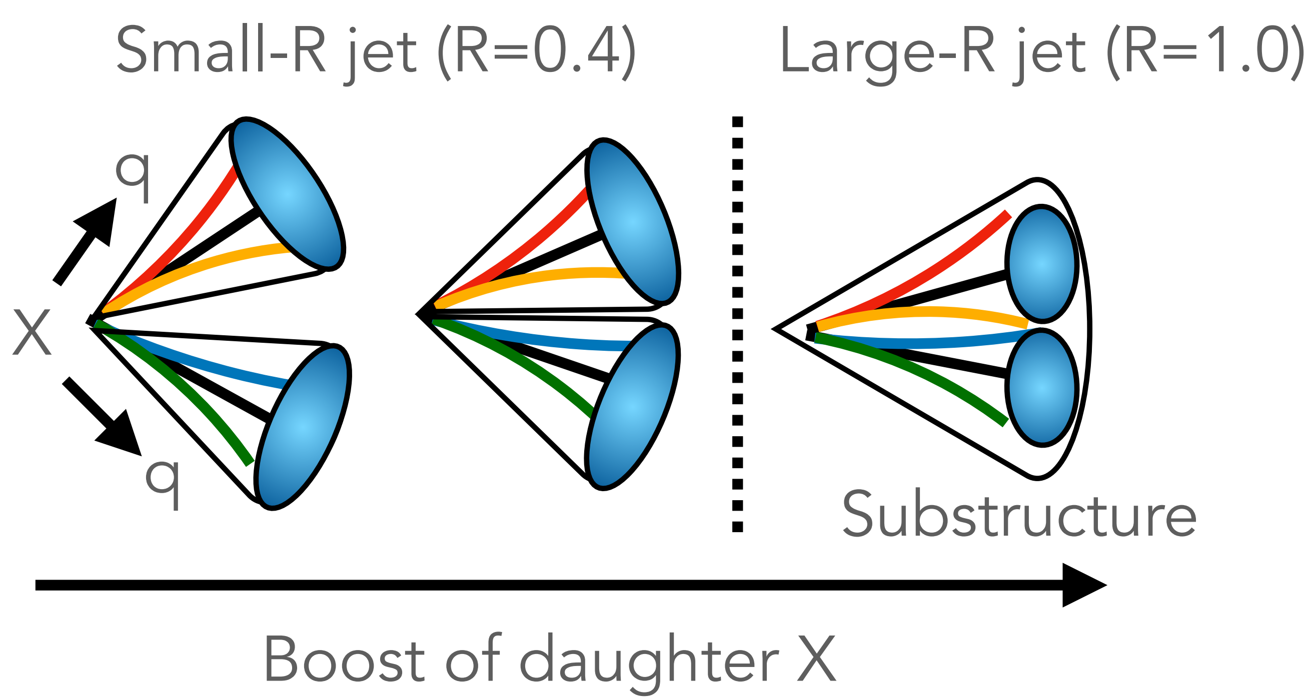


- We'll compare & contrast how these searches tackled key questions in collider-based AD, including input modeling, level of supervision, and result reporting

Anomalous Jet Substructure in Y→XH

- The Higgs boson coupling to mass motivates the search of new TeV scale particles produced in association with a Higgs
- Hadronic final states in these searches often lead to **jets** that have **substructure** due to the boosting of the daughter particles (reconstructed as large-R jets)
- Y→XH is a fully hadronic search for a new resonance Y (~TeV) decaying into a SM Higgs and a new particle X (~100s GeV) displaying **anomalous jet substructure**

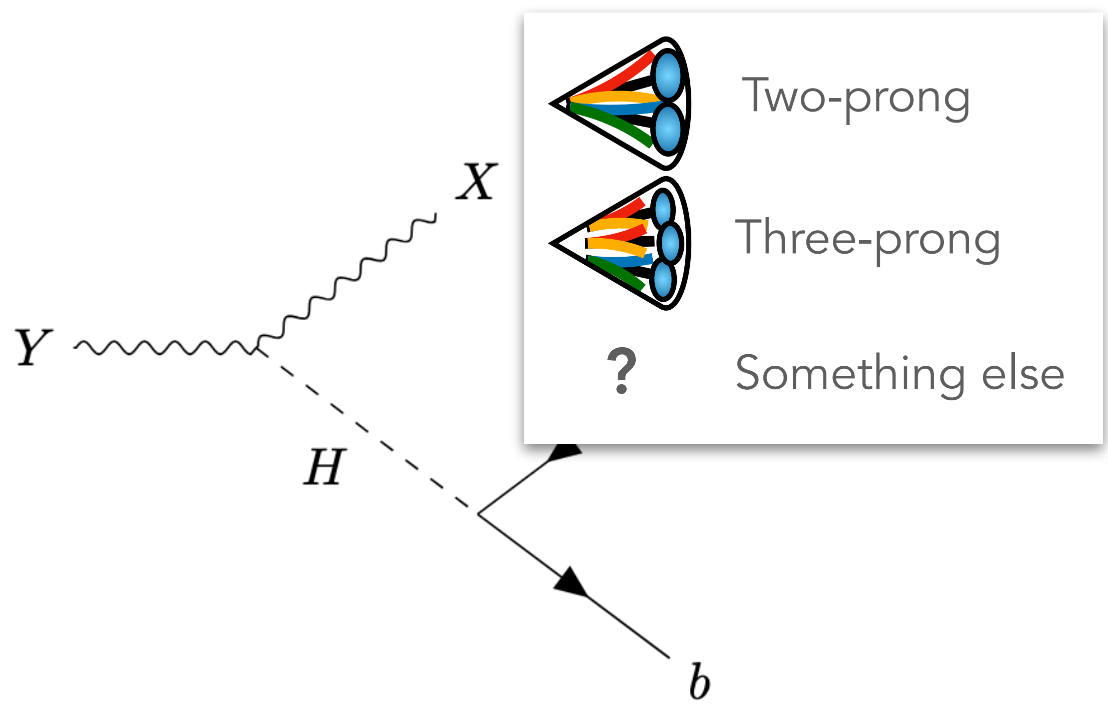


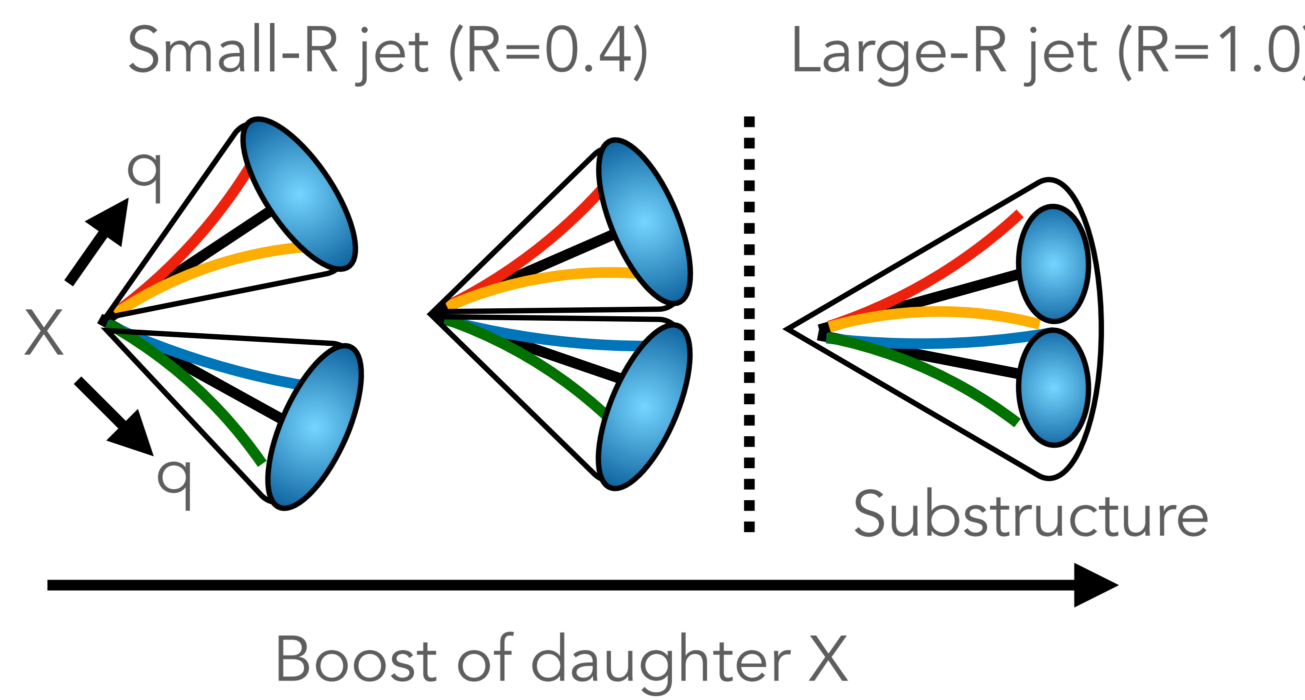


- Two-prong X→$q\bar{q}$ assumption used for exclusion result, signal grid production, & to define model-dependent benchmark → **anomaly & exclusion regions**

Anomalous Jet Substructure in Y→XH

- The Higgs boson coupling to mass motivates the search of new TeV scale particles produced in association with a Higgs
- Hadronic final states in these searches often lead to **jets** that have **substructure** due to the boosting of the daughter particles (reconstructed as large-R jets)
- Y→XH is a fully hadronic search for a new resonance Y (~TeV) decaying into a SM Higgs and a new particle X (~100s GeV) displaying **anomalous jet substructure**

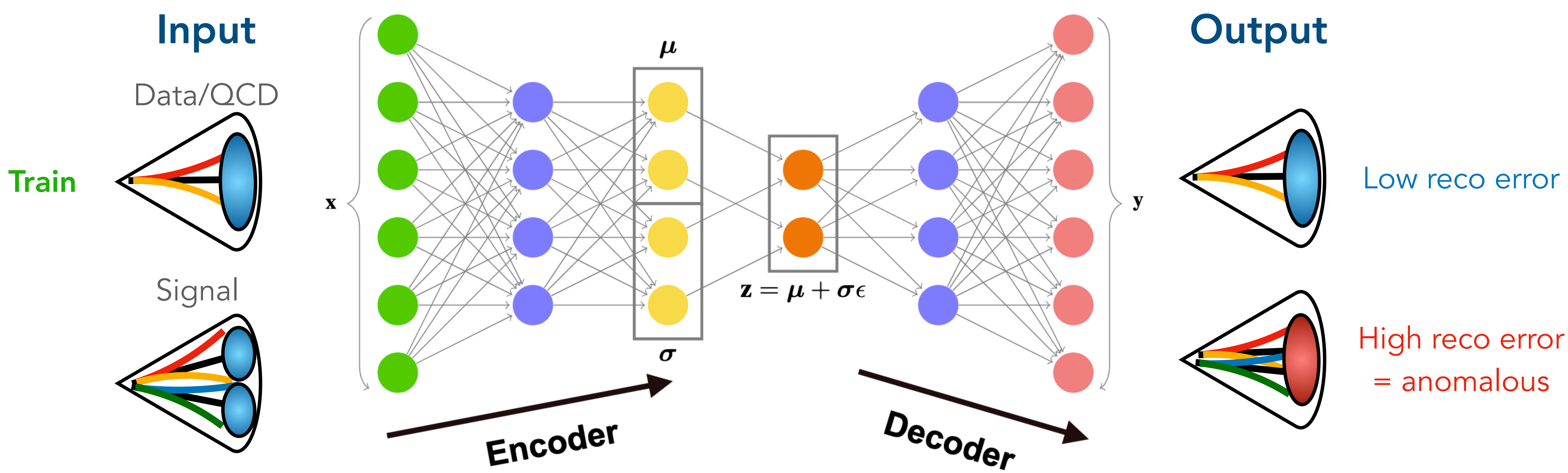




- Two-prong X→$q\bar{q}$ assumption used for exclusion result, signal grid production, & to define model-dependent benchmark → **anomaly & exclusion regions**

Challenge: Autoencoder for Jet Substructure

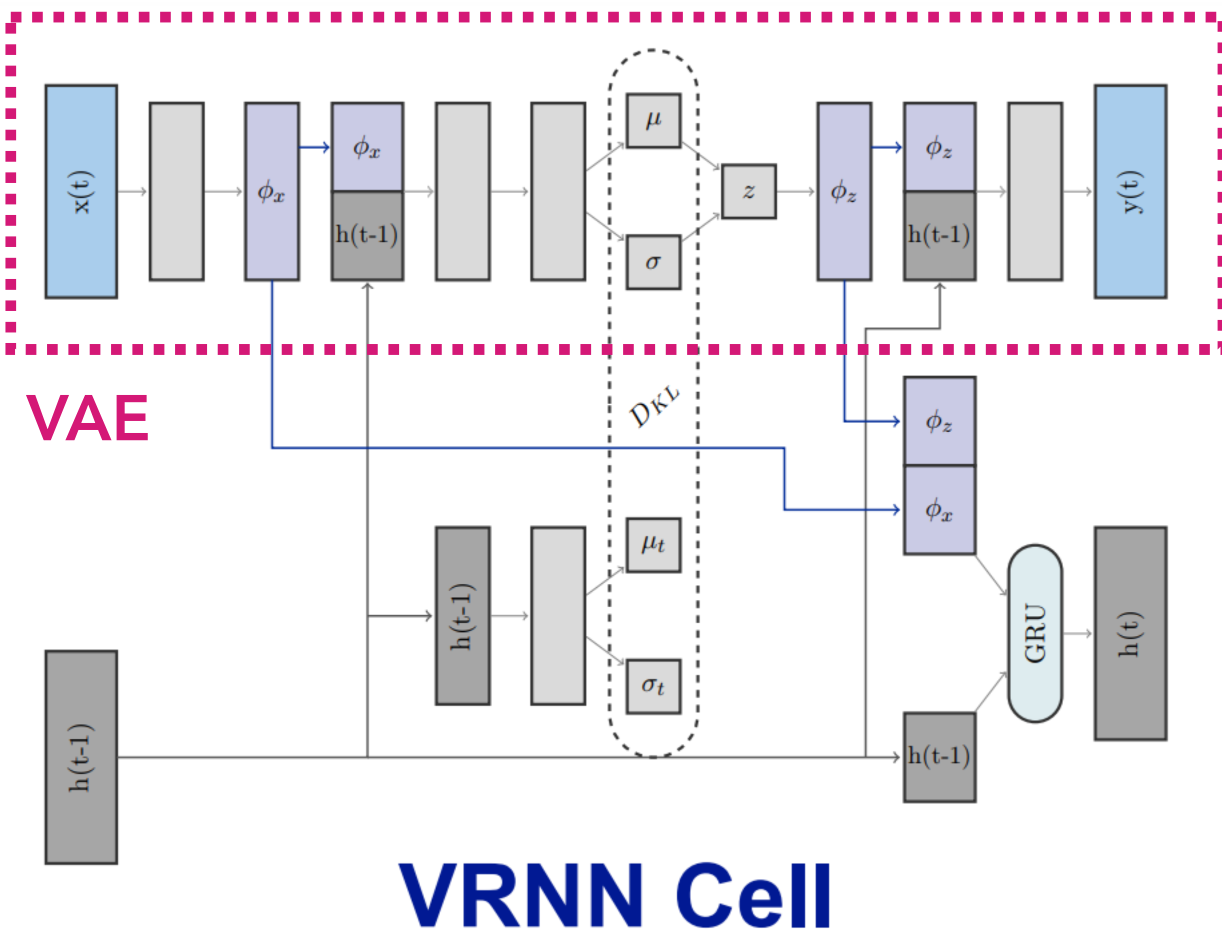
- **Idea:** we can model jets by their constituent 4-vectors and feed them to an AE
  - Vast abundance of "featureless" QCD lets one train over data → **unsupervised**
  - Jets with substructure get flagged with a high anomaly score



- **Challenge:** # of constituents varies per jet & AE requires a fixed length input
  - Need a way to accommodate variable number of inputs

Variational Recurrent Neural Network

- We can solve this problem with a recurrent neural network (RNN)
- **Variational RNN:** recurrent architecture that updates a variational AE latent space at each time step; accommodates variable-length input sequences
- Define **anomaly score (AS)** per jet as a function of the KL divergence loss term



Loss

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

Mean-squared reconstruction error

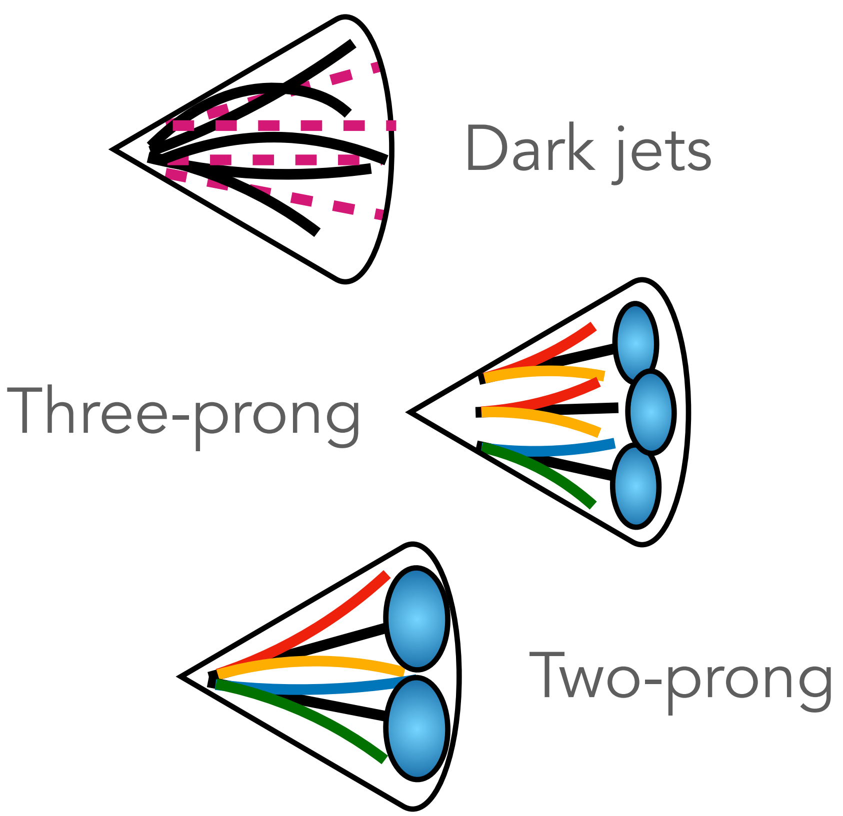
Kullback-Leibler Divergence

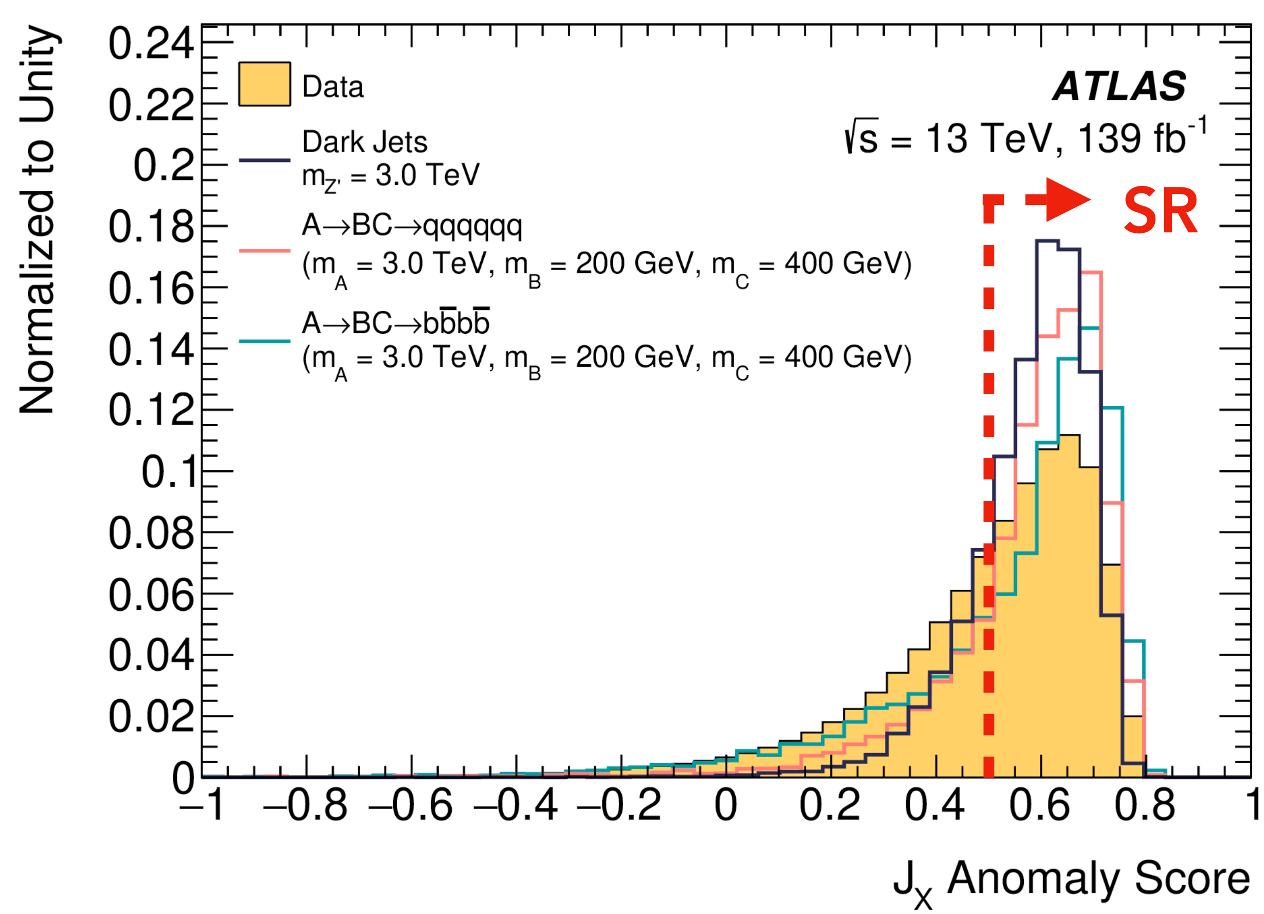
Anomaly score

$$\mathrm{AS} = 1 - e^{-D_{KL}}$$

VRNN Jet Tagging in Y→XH

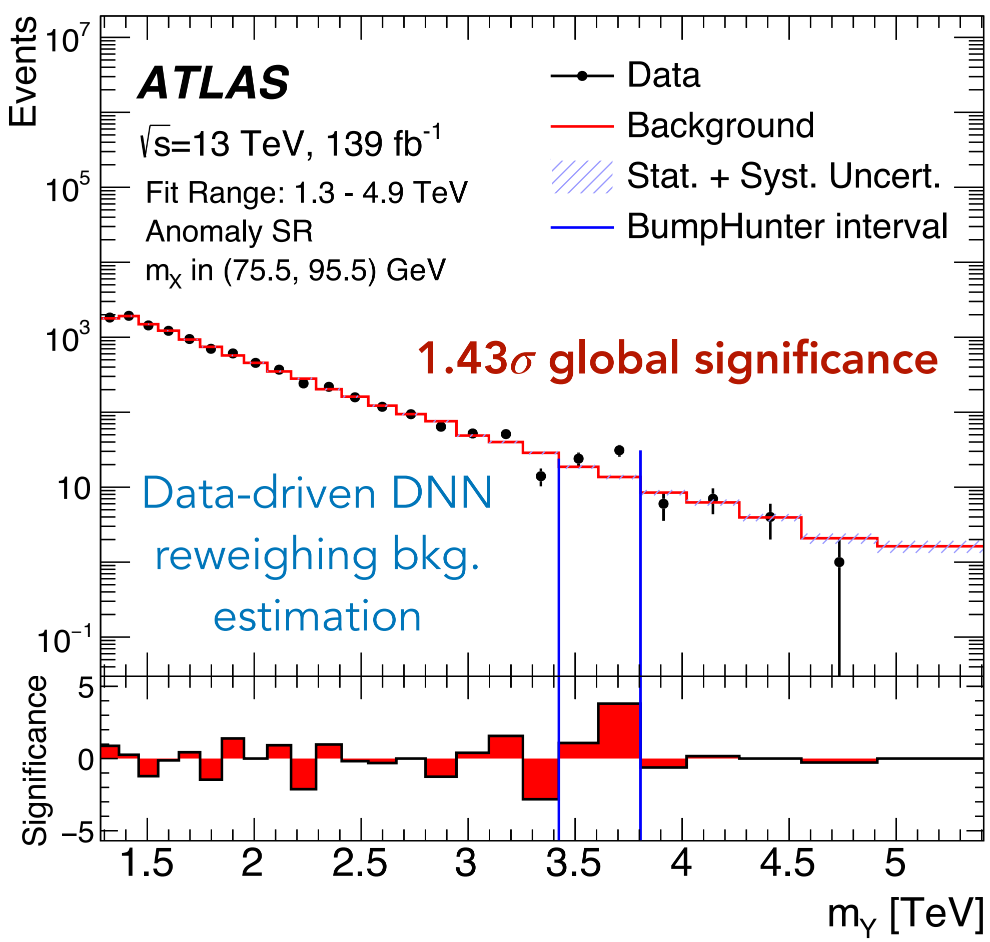
- Train over full ATLAS Run 2 dataset of large-R jets with p_T > 1.2 TeV
  - Up to 20 constituents ordered by k_T splitting & D_2, τ_{32}, Split_{12}, Split_{23}
- Evaluate over four substructure hypotheses to assess model independence
  - 2-prong, 3-prong, heavy flavor ($b\bar{b}$), and dark jets (Pythia Hidden Valley)
- **Employ AS > 0.5 SR definition for sensitivity to a broad range of signatures**

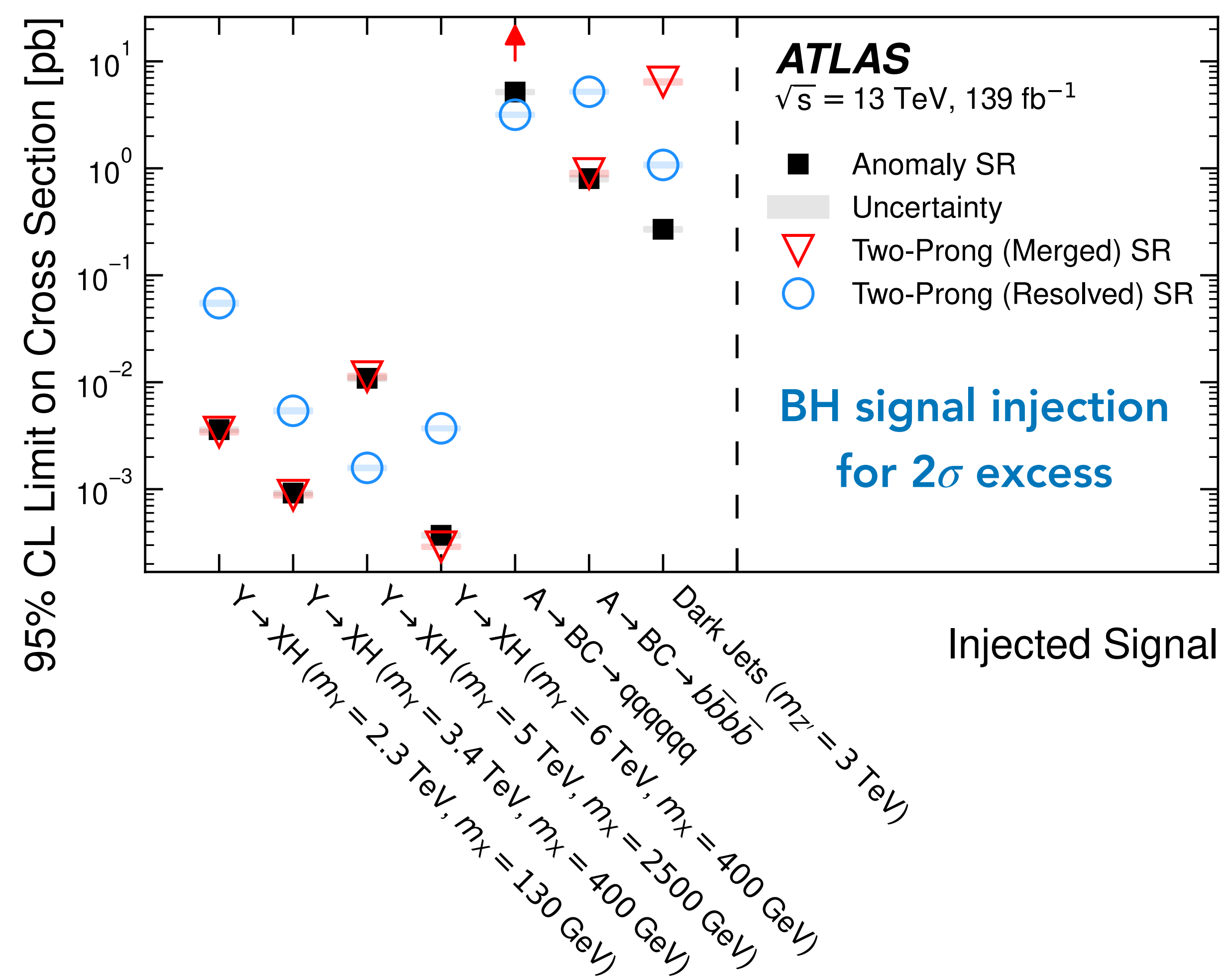




Results & Outlook

- Scan M_Y, M_X parameter space & quote most significant excess w.r.t. the expected background via the BumpHunter (BH) algorithm since no signal assumption
- Compare with X→$q\bar{q}$ model-dependent region to assess sensitivity breadth
  - **AS selection competitive on two-prong signals & x10 better for dark jets**



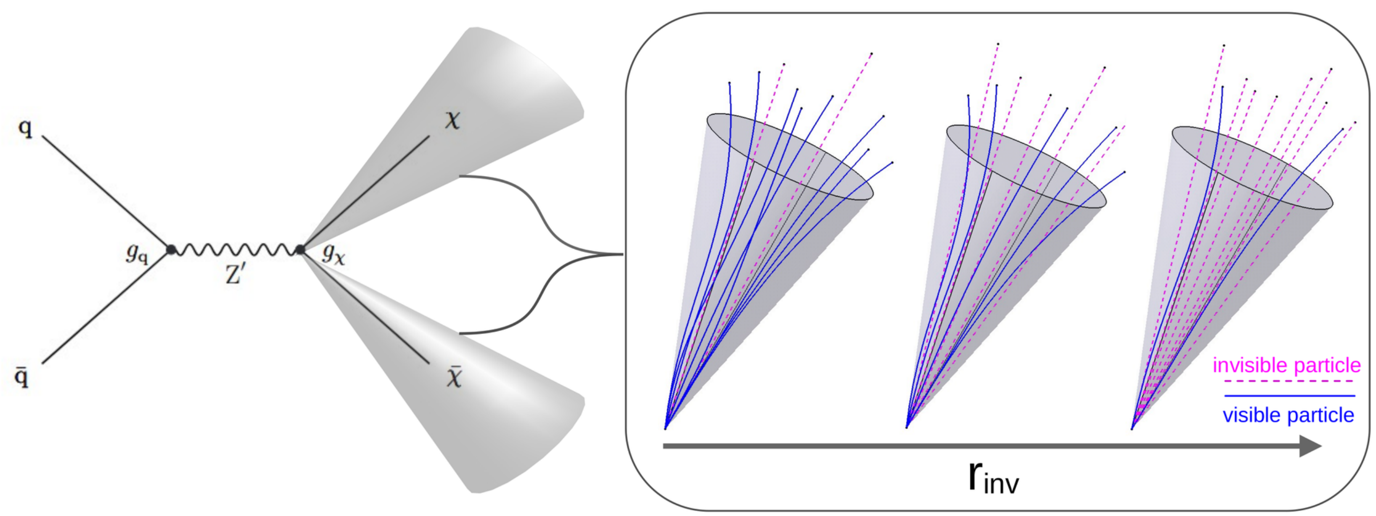


First fully unsupervised result in ATLAS!

Semi-Visible Jets Search

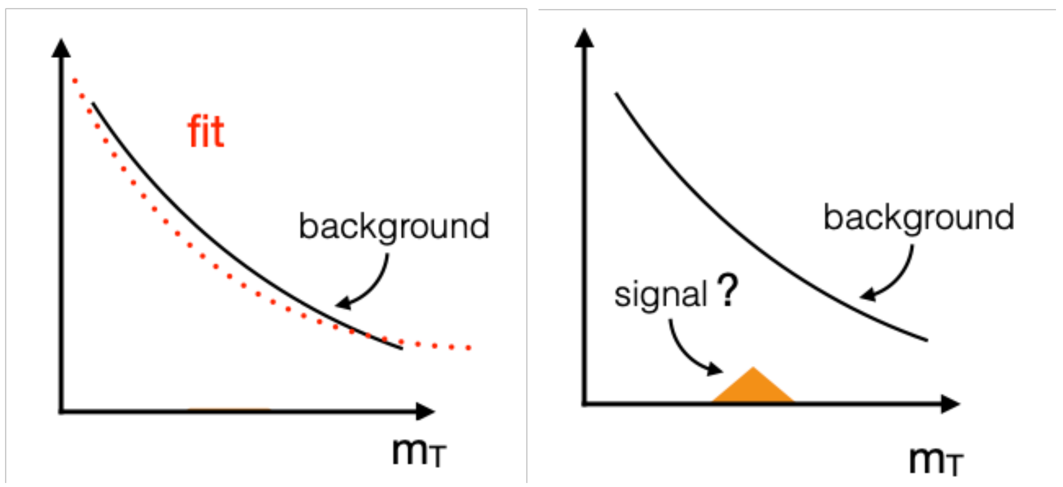
- **Semi-visible jets** (SVJs) arise from Hidden Valley models describing a dark sector that is strongly interacting, allowing for dark matter (DM) particles to hadronize
  - These lead to complex signatures in which invisible dark hadrons partially decay back to visible SM particles and produce jets that are **semi-visible**
- Subtle shower differences between dark and SM QCD motivates the use of **low-level track variables** to spot key differences between signal and bkg. correlations
  - Like Y→XH, define model-dependent **exclusion region** for limits and model-independent **anomaly region** for generalizability beyond SVJs

χ dark quark



Heavy Z' mediator

$$r_{inv} = \frac{\#\ \text{dark hadrons}}{\#\ \text{total hadrons}}$$

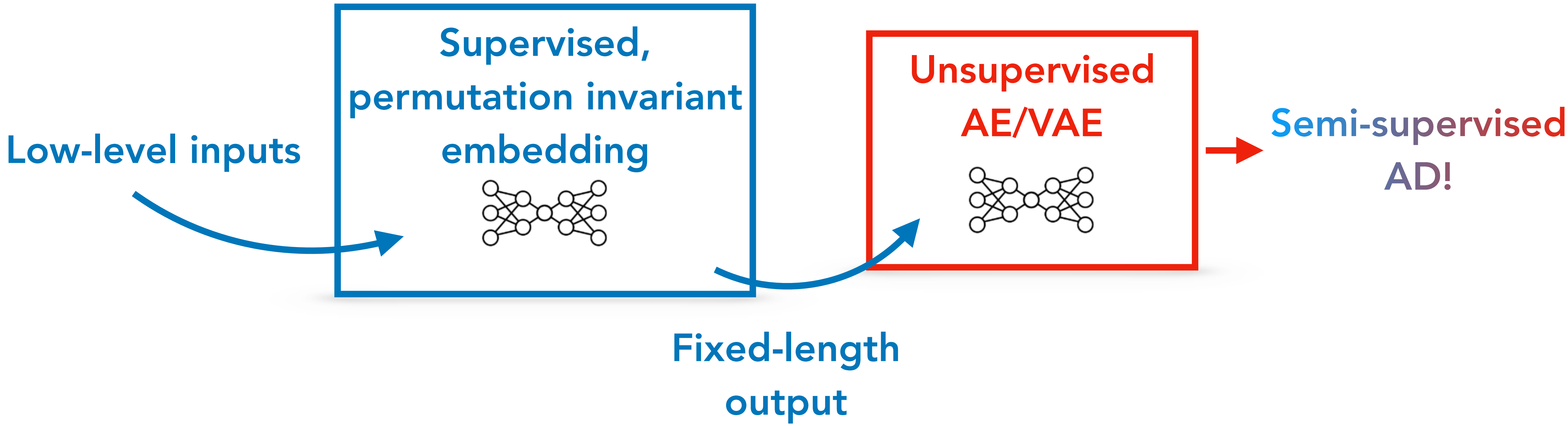


Resonant production in m_T

$$m_T^2 = \left[E_{T,JJ} + E_T^{\text{miss}}\right]^2 - \left[\vec{p}_{T,JJ} + \vec{p}_T^{\,\text{miss}}\right]^2$$

Challenge: Permutation Invariance

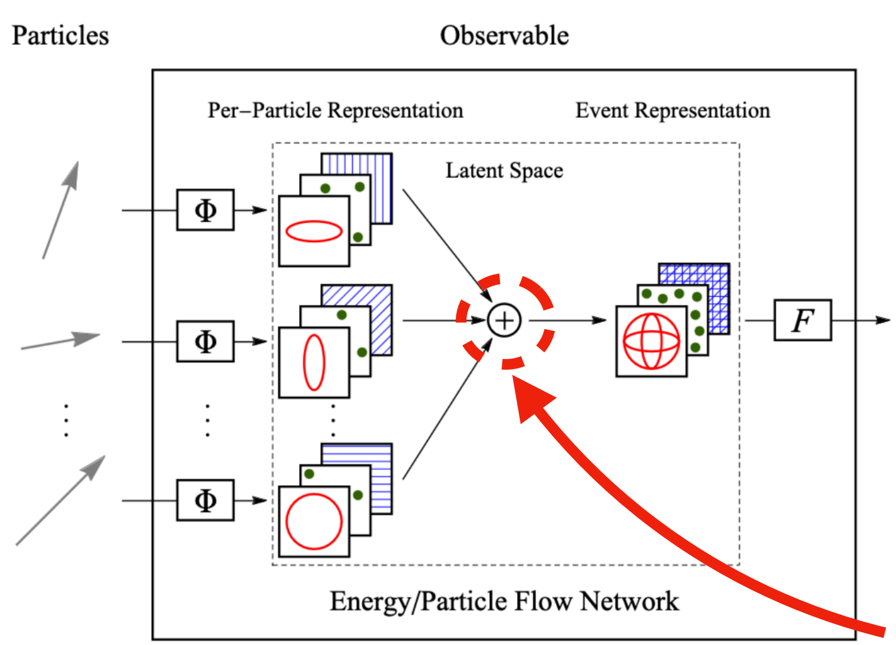
- One pitfall of the VRNN approach is that the ordering of constituents matters
- Collider data is best described as a *set* of objects (e.g. particle tracks, calo clusters, etc.) that are not only variable in length but also **permutation invariant**
    - Artificial manipulations such as zero-padding or imposing an ordering scheme can impact our ability to fully exploit low-level information from our detectors
- **Challenge:** employ AD on low-level objects conserving permutation invariance
    - We can use a supervised classifier to create a **smart embedding** that is fixed-length, permutation invariant, and can be fed to an AE/VAE for AD



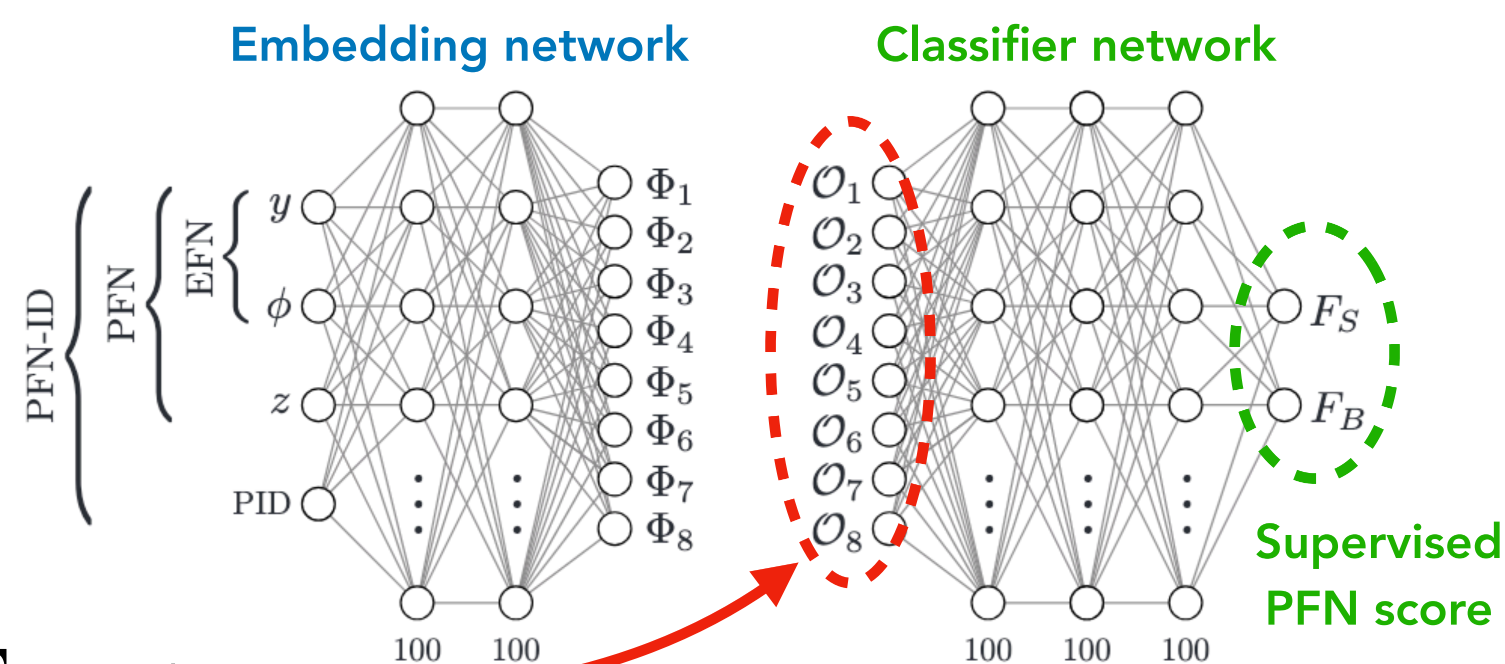
Particle Flow Network

arXiv:1703.06114
arXiv:1810.05165

- One way to achieve this embedding is with a **P**article **F**low **N**etwork (PFN)
- The PFN is a **supervised** classifier based on the Deep Sets framework
  - The network takes in an arbitrary number of particle features that are encoded into a latent space, per-particle representation by a set of learned functions Φ_a
  - These per-particle representations are combined into event level observables $\mathcal{O}_a$ that are **inherently permutation invariant** by **summing** over input particles



$$\mathcal{O}_a = \sum_i \Phi_a(y_i, \phi_i, z_i, \mathrm{PID}_i)$$

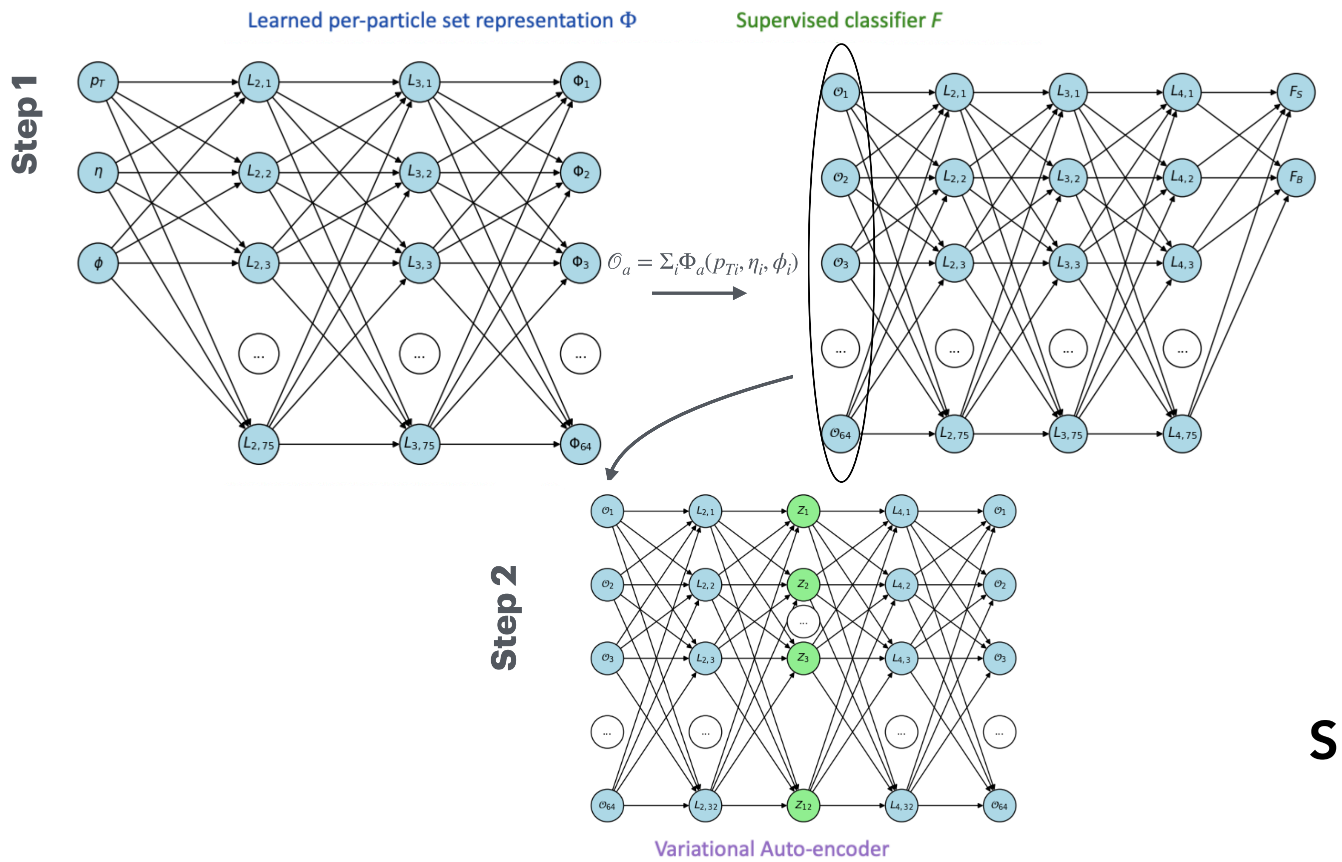


- Classifier F used to define model-dependent, supervised exclusion region

arXiv:2408.17409

Anomaly Detection on Particle Flow Latent Space

- The fixed-length PFN embedding $\mathcal{O}_a$ is derived by training the supervised classifier to distinguish SVJs from QCD jets → smart embedding with SVJ prior
- We pass data through this embedding and use it to train a **unsupervised** VAE & define a novel architecture **ANTELOPE** performing AD on the PFN's latent space



Anomaly score (AS) determined from VAE loss

$$\mathcal{L} = \text{MSE} + \text{KLD}$$

$$\text{AS} = \left(1 + e^{-\log_{10}(\mathcal{L})}\right)^{-1}$$

Supervised prior + **Unsupervised VAE** = **Semi-supervised AD**

- Allows performing AD in low-level detector objects in a permutation invariant way

ANTELOPE in SVJ Search

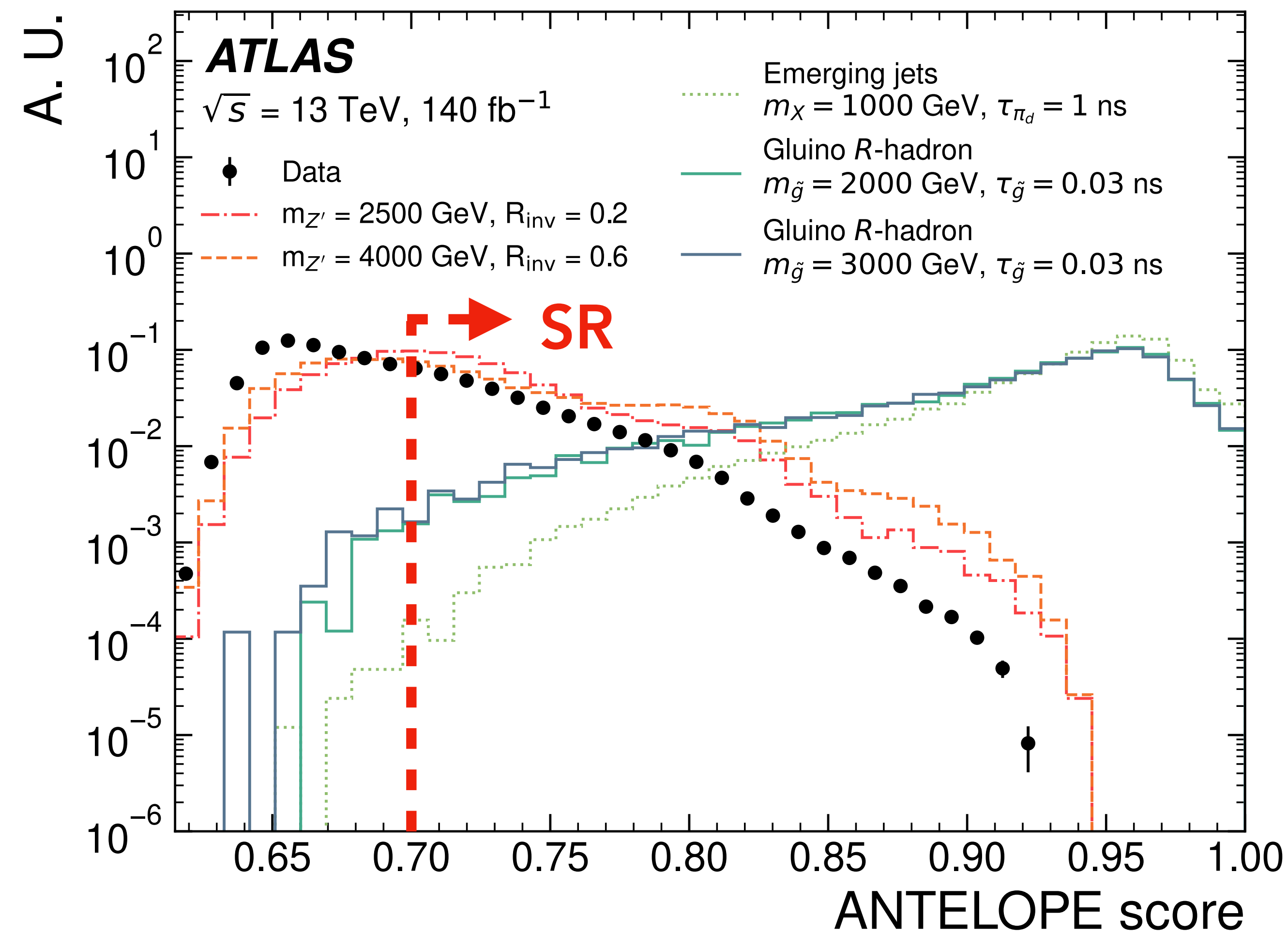
- PFN and ANTELOPE trained over the 80 hardest ghost-associated track 4-vectors (p_T, η, ϕ, E) and impact parameters (d_o, z_o) in the two leading jets of the event
  - PFN is trained on QCD MC & SVJ signals, ANTELOPE only on data
- Tested over signatures high in MET & displacement to assess model independence

- **ANTELOPE score > 0.7 used to define anomaly region**
- Provides sensitivity *beyond* SVJs

> ▸ **Emerging jets (EJs)**
> ▸ **Gluino production → *R*-hadron + LSP**

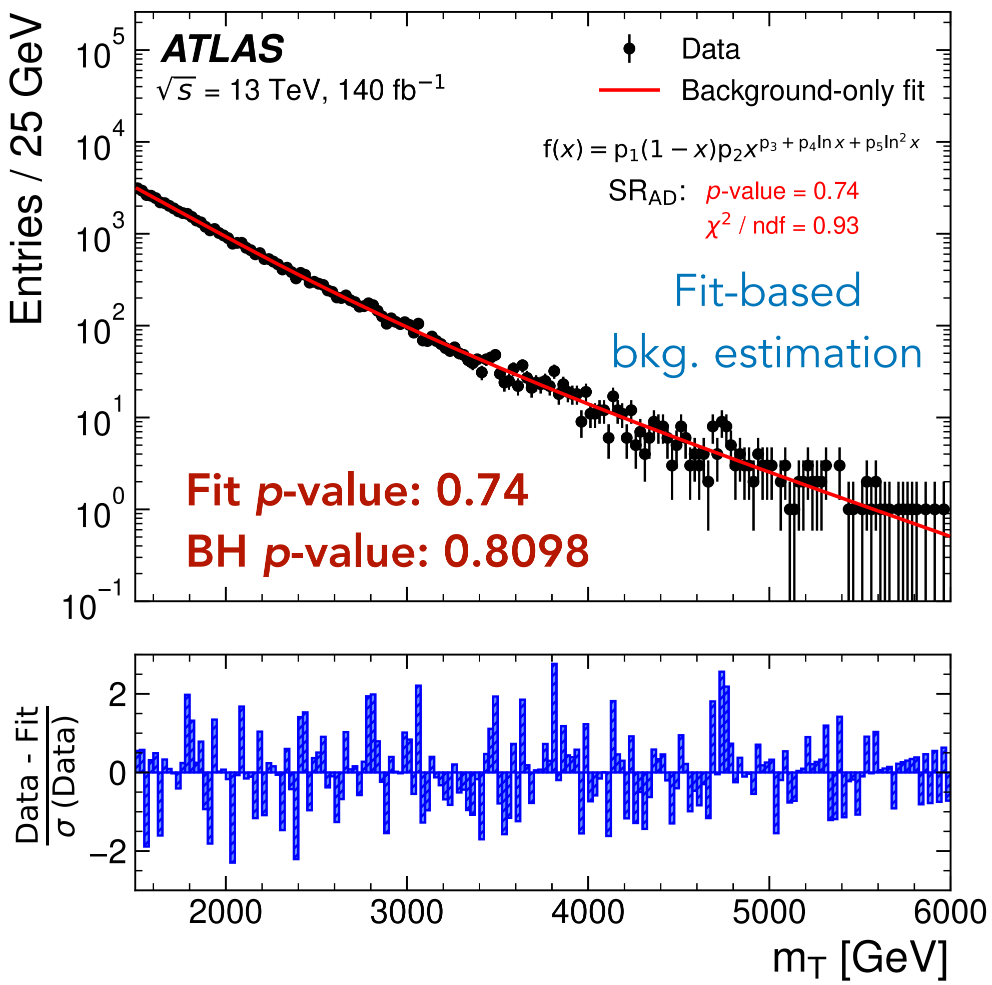
High MET, high displacement signatures

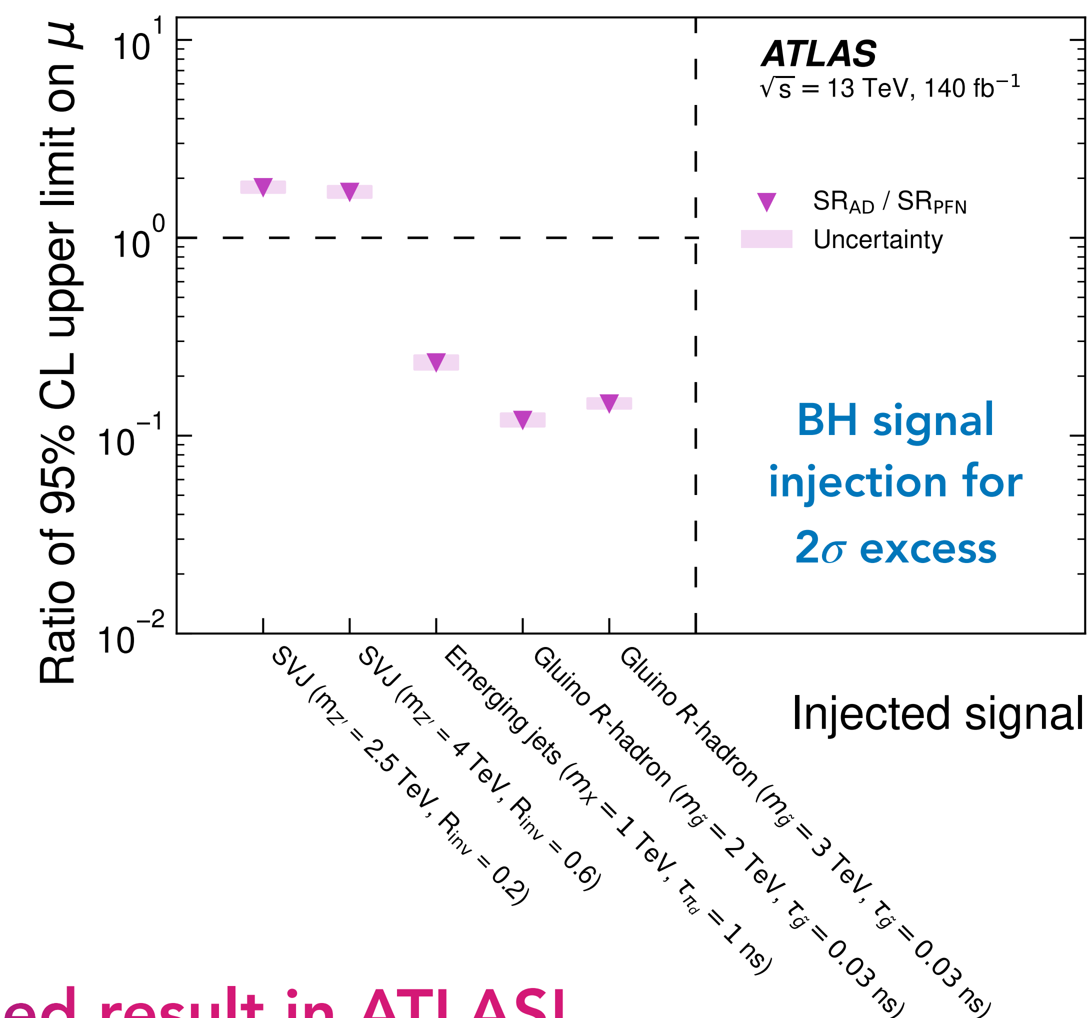
→ high ANTELOPE score



Results & Outlook

- Perform fit over m_T spectrum in anomaly region and quote p-value
  - Use as background template for BH and quote largest deviation
- Compare with model-dependent PFN region to assess model independence
  - **Broader sensitivity & order of magnitude better for EJs, Gluino R-hadron**





First semi-supervised result in ATLAS!

Compare & Contrast

	Y→XH	SVJ
Challenge	Variable-length inputs	Variable length inputs & permutation invariance
Input modeling	Jet constituents	Particle tracks (after PFN embedding)
Architecture	VRNN	ANTELOPE
Trained on	Data	Data (PFN on QCD & SVJs)
Background estimate	DNN reweighing	Polynomial fit
Result	BH *p*-value	Fit & BH *p*-values
Supervision	Unsupervised	Semi-supervised
Areas of improvement?	VRNN requires ordering	ANTELOPE relies on signal prior*

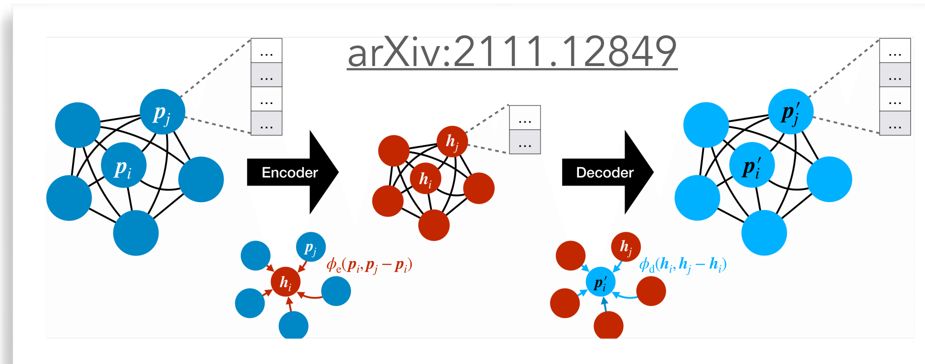
*Maybe an advantage instead of disadvantage

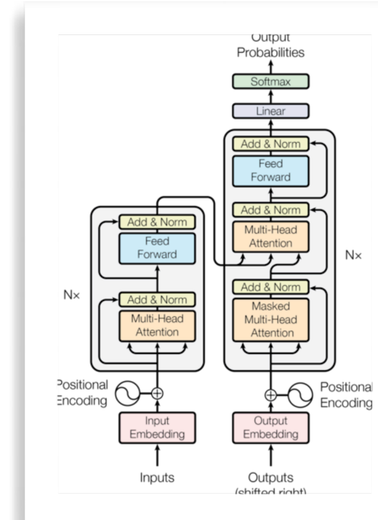
What Does the Future Hold?

[arXiv:1706.03762](#)

- **Sophisticated networks**
  - Adapting graphs, transformers, normalizing flows for less-than-supervised applications
  - See tomorrow's session

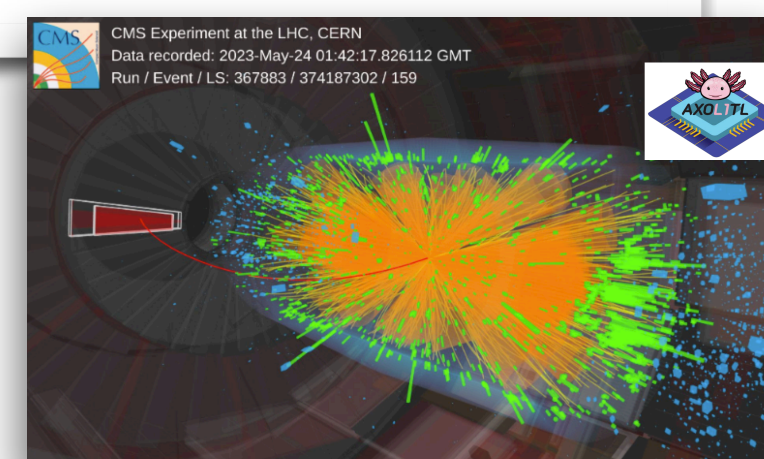
arXiv:2111.12849





CICADA

- **Trigger-level AD on FPGAs**
  - AXOL1TL & CICADA on the CMS side
  - Similar efforts ongoing in ATLAS L1+HLT



- **Can all have significant impact on future AD analysis results**

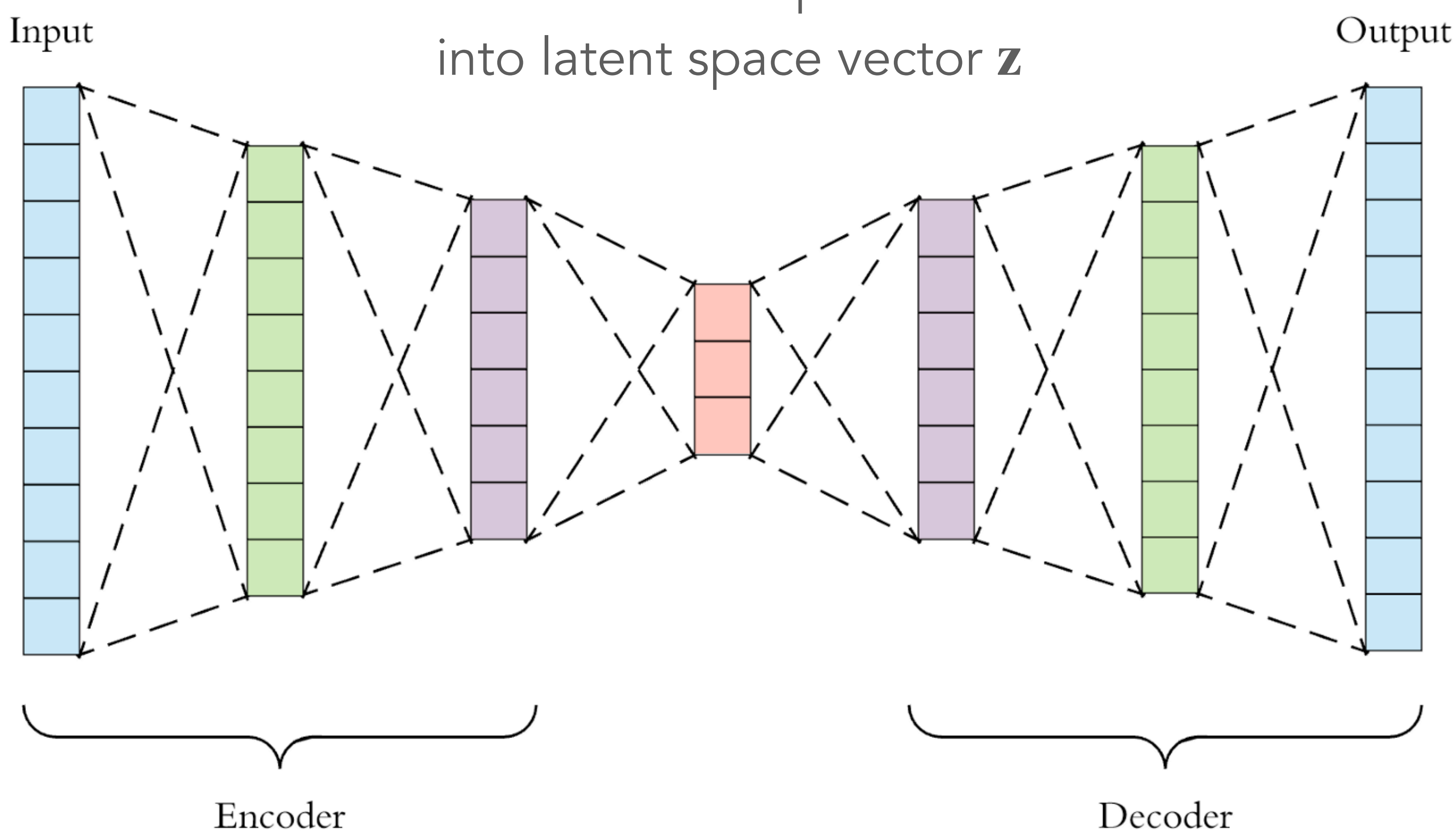
Thank you!

Backup

Overview: Autoencoder (AE)

Features compressed
into latent space vector $\mathbf{z}$

Input
features $\mathbf{x}$



Reco’d $\mathbf{y}$

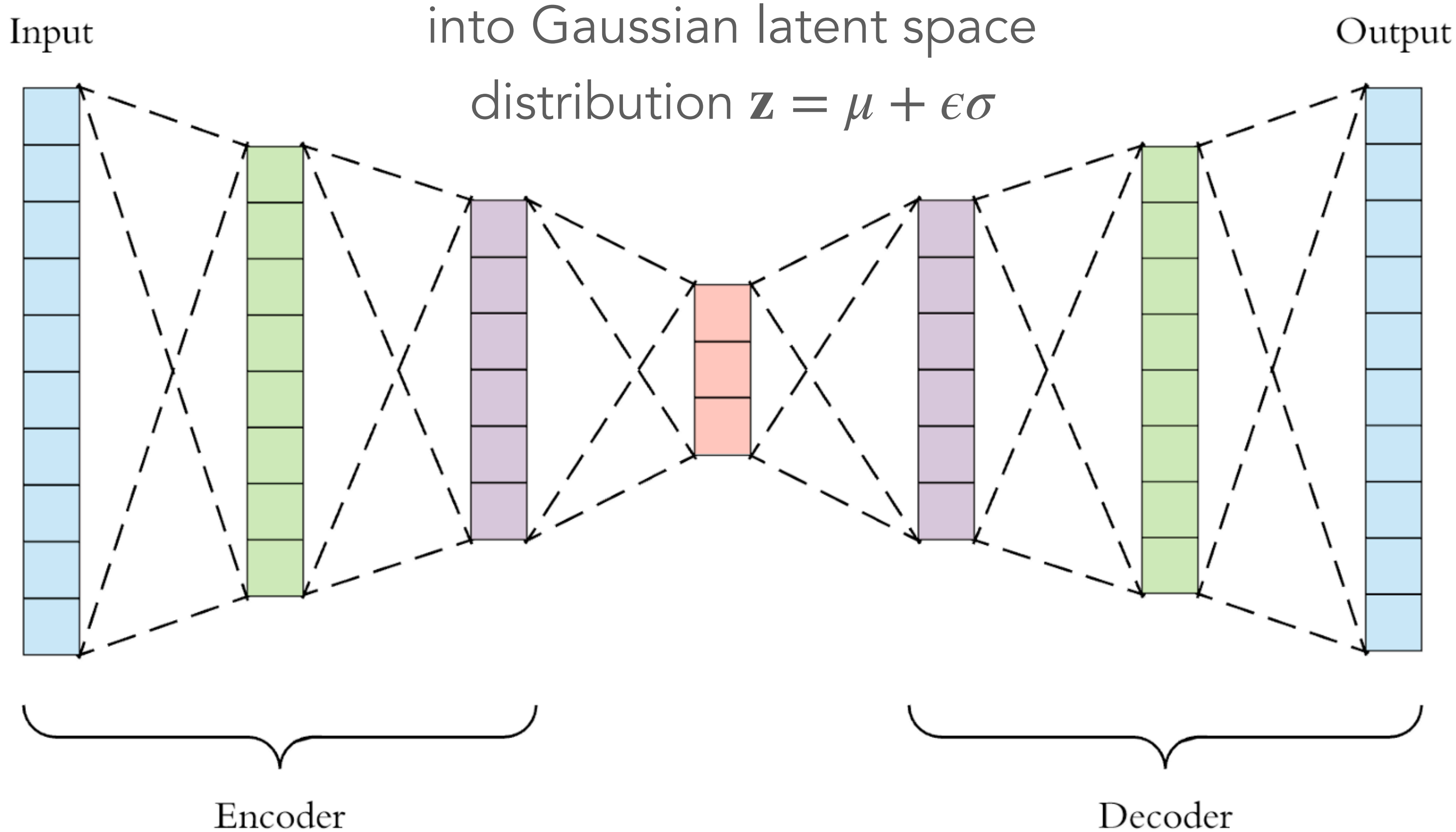
Loss = Reconstruction Error

$$\mathscr{L} = |\mathbf{x} - \mathbf{y}|^2$$

Overview: Variational AE (VAE)

Features compressed into Gaussian latent space distribution $\mathbf{z} = \mu + \epsilon\sigma$

Input features $\mathbf{x}$



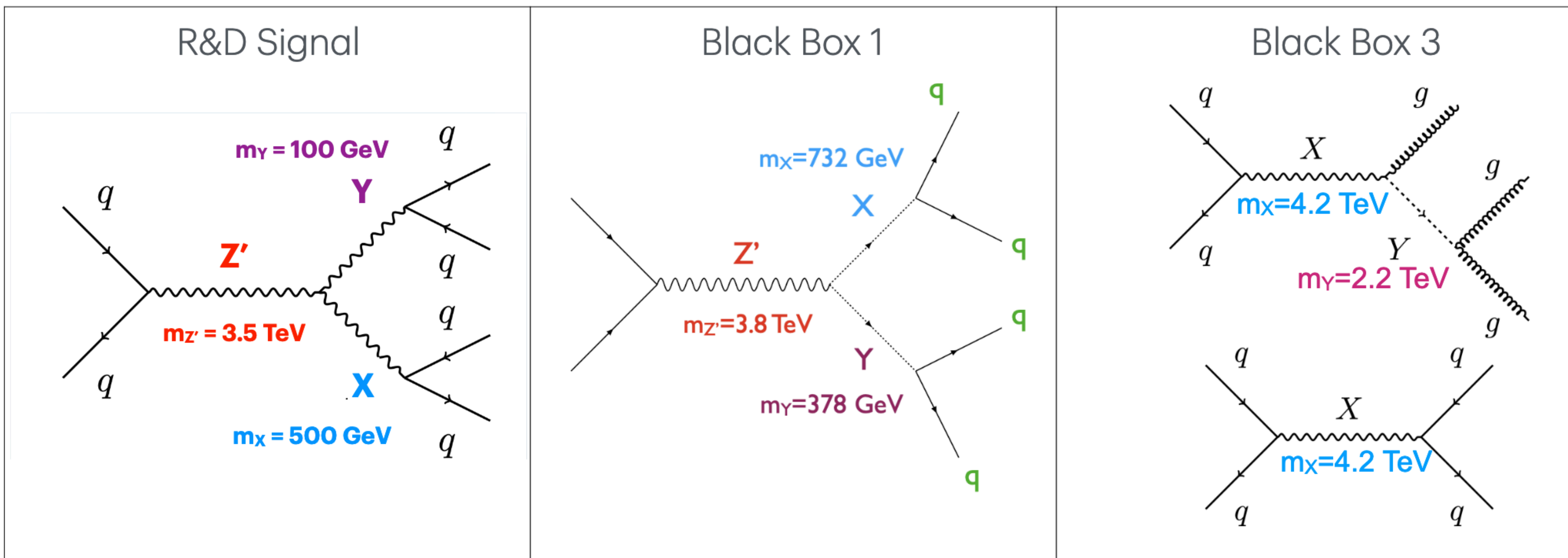
Reco'd $\mathbf{y}$

Loss = Reconstruction Error
$\mathscr{L} = |\mathbf{x} - \mathbf{y}|^2 + \lambda D_{KL}$

$\epsilon \sim \mathcal{N}(0,1)$ used for sampling latent space

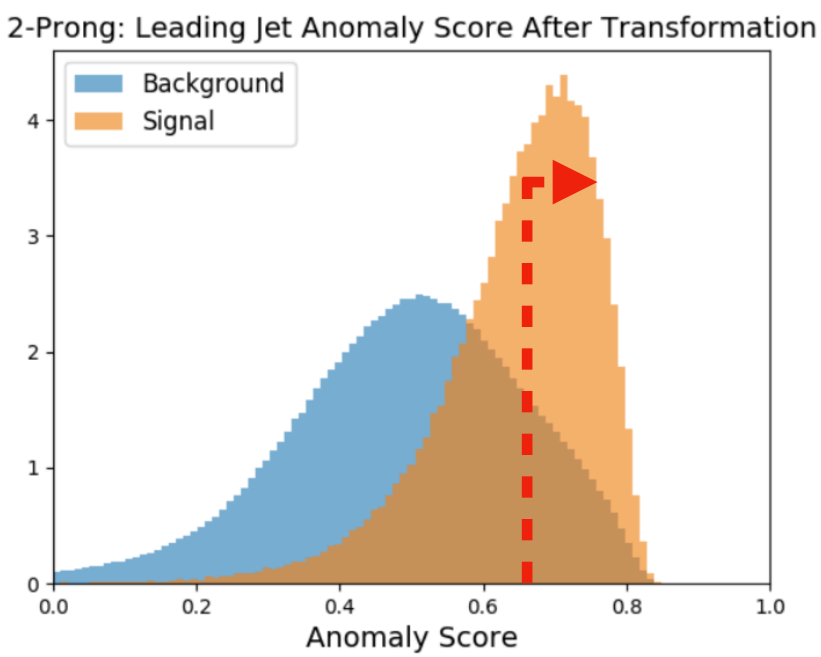
LHC Olympics Dataset

- Both the VRNN and ANTELOPE were developed with the LHC Olympics dataset
- The dataset consists of 3 R&D and 3 black box samples
    - Each event described as set of up to 700 (massless) particle 4-vectors (p_T, η, ϕ)
    - **R&D**: QCD multijet, 2-prong, and 3-prong
    - **Black boxes:** BB1 2-prong BB2 QCD multijet (No signal) BB3 Resonance→Dijet/Trijet



VRNN in the LHC Olympics

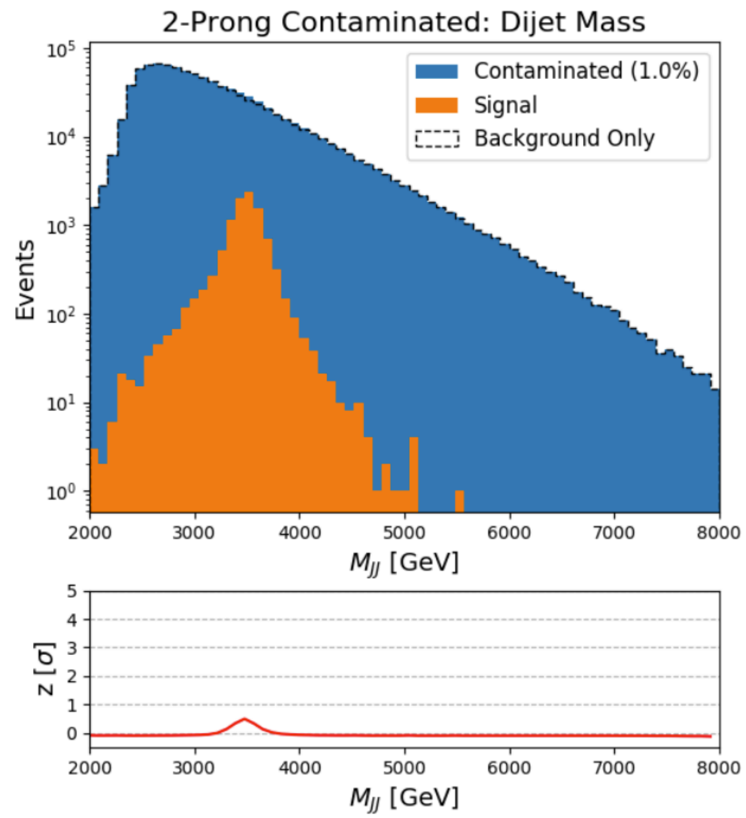
- LHCO events reco'd into two large-radius (R=1.0) jets with leading p_T > 1.2 TeV
- VRNN trained on QCD-only to derive anomaly score
  - Performance assessed on two & three-pronged samples



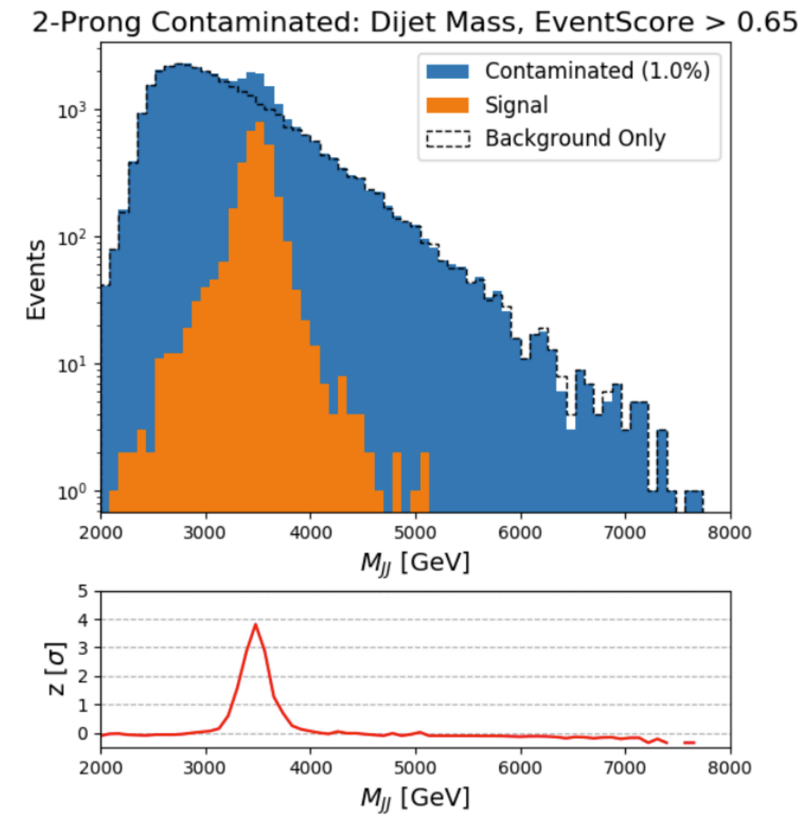
Anomaly score

$$AS = 1 - e^{-D_{KL}}$$

Before Selection



After Selection

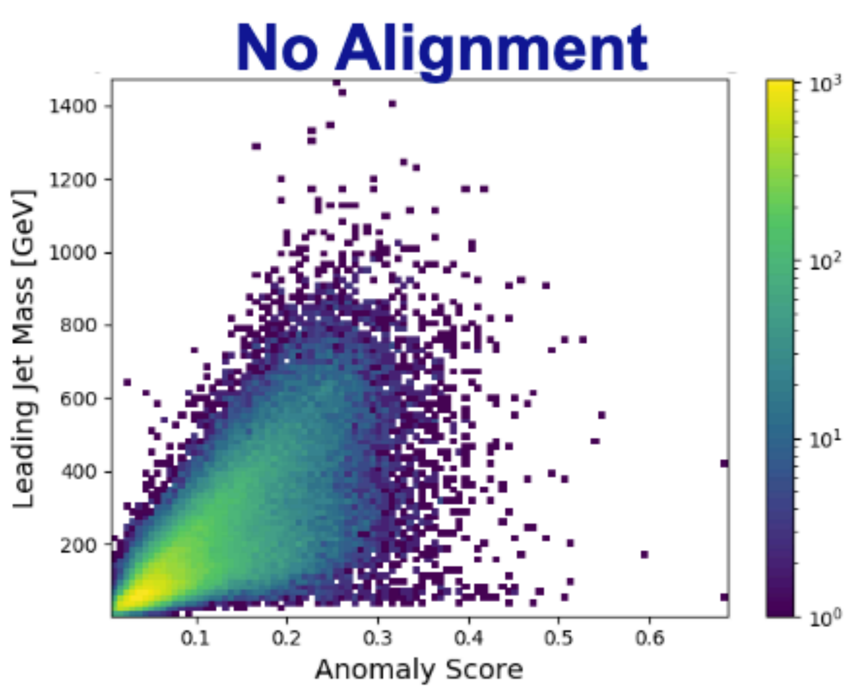


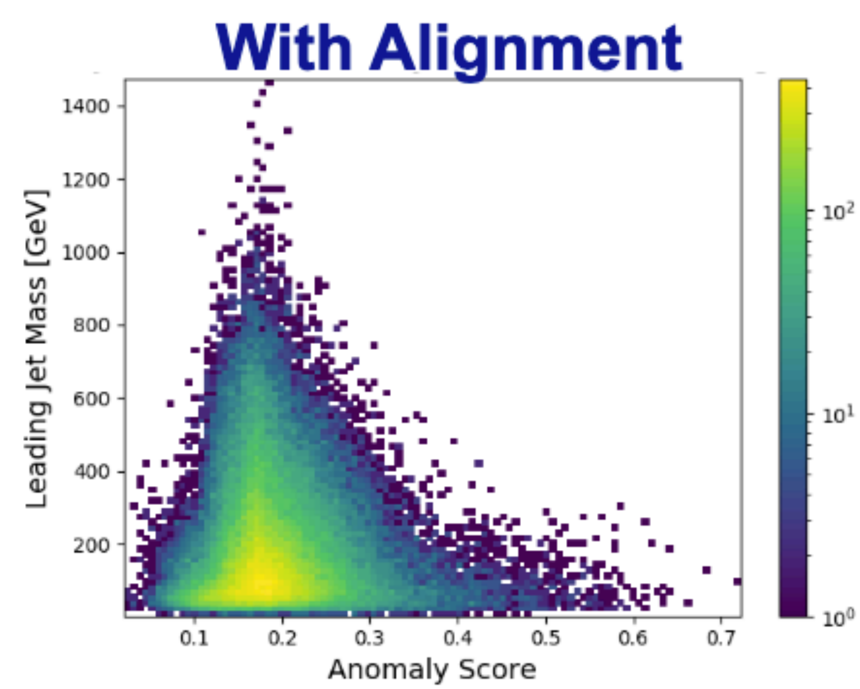
VRNN Preprocessing: Alignment

- Alignment procedure done to remove mass and p_T 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

Algorithm 1: Jet Alignment

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 ϕ axis such that the second hardest constituent has $\eta_2 > 0$



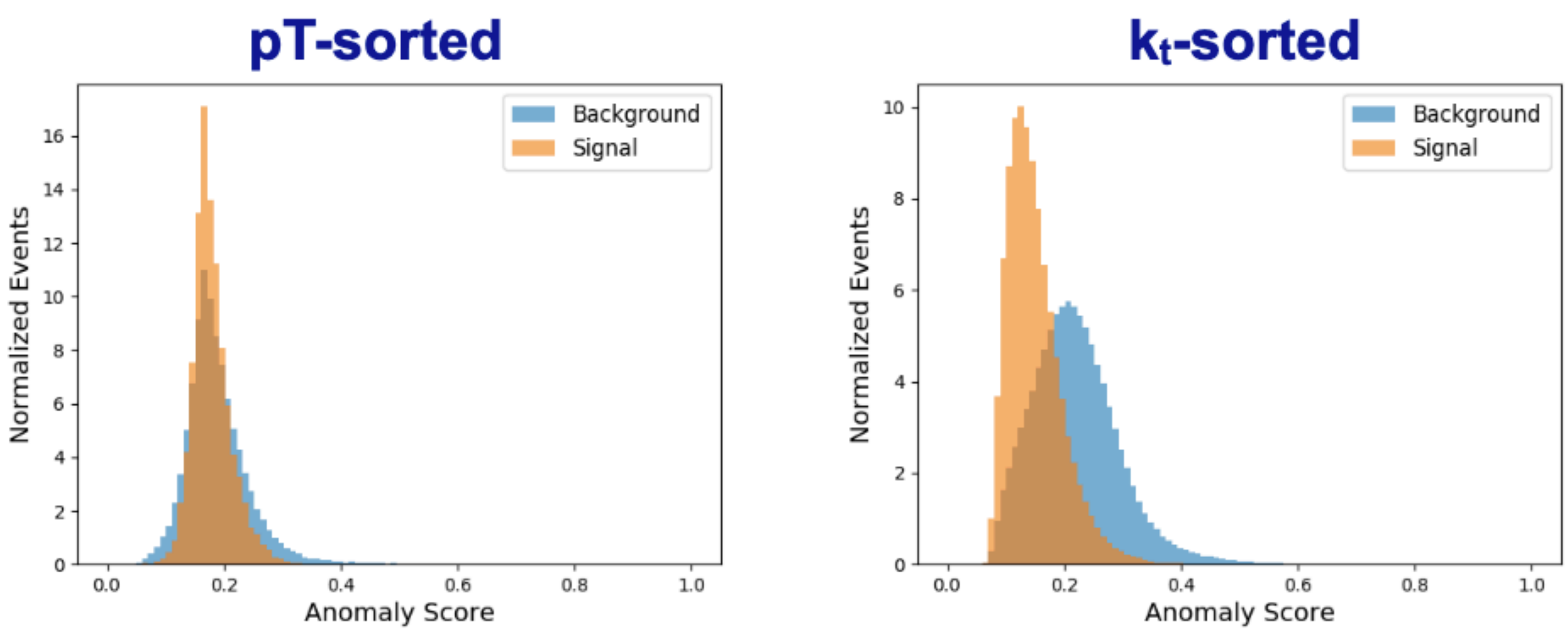


VRNN Preprocessing: Ordering

- Selecting an appropriate ordering scheme in recurrent neural networks can highlight important sequence features & boost performance
- Select k_T -distance ordering to highlight substructure: nth constituent has highest k_T -distance relative to previous, starting with highest p_T constituent

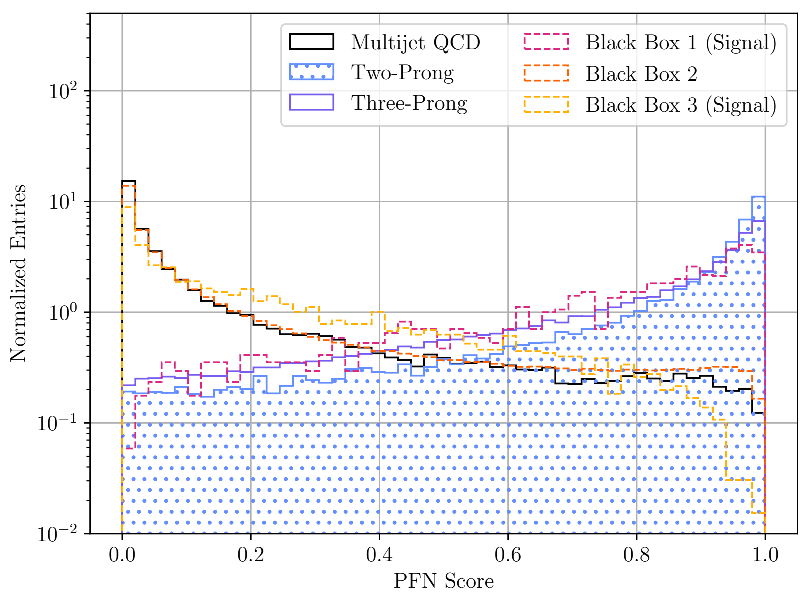
$$c_n = \max(p_{T,n} \times \Delta R_{n,n-1})$$

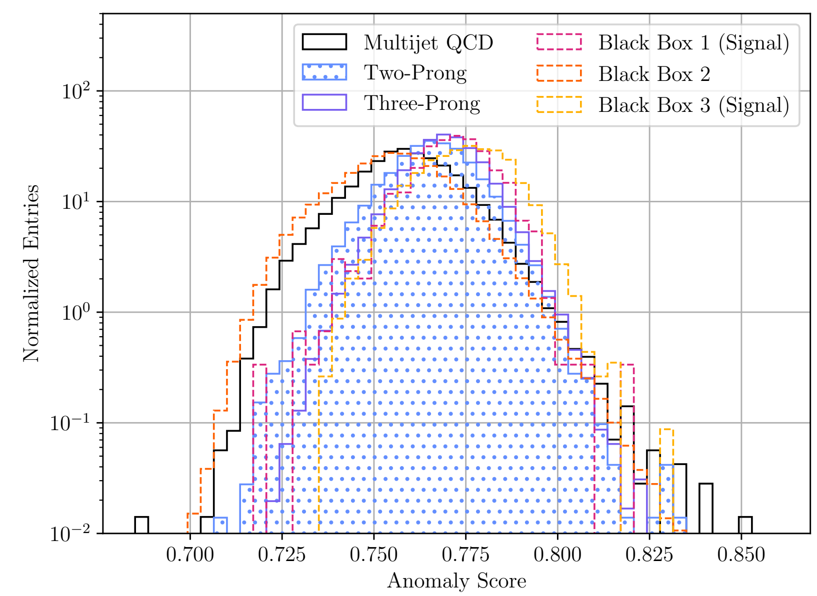
- **Result:** better sep. of two-prong signal from QCD background than p_T sorting



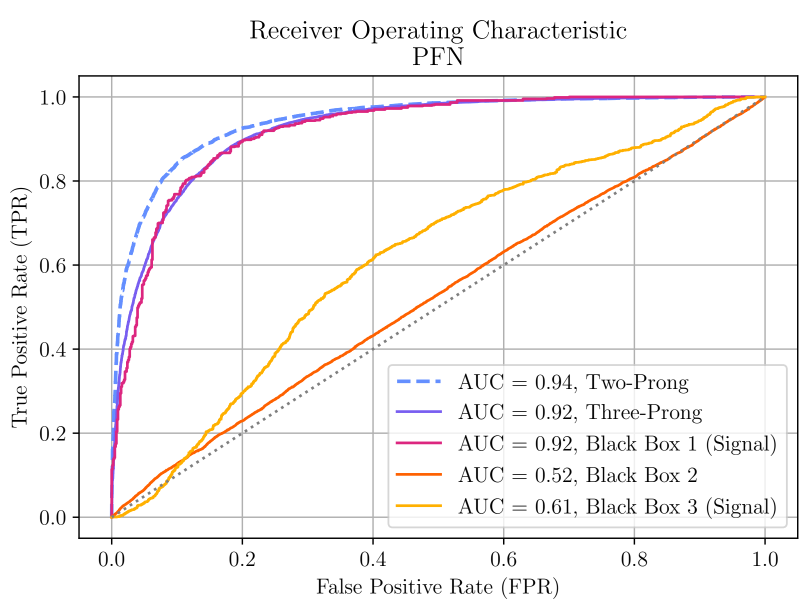
ANTELOPE in the LHC Olympics

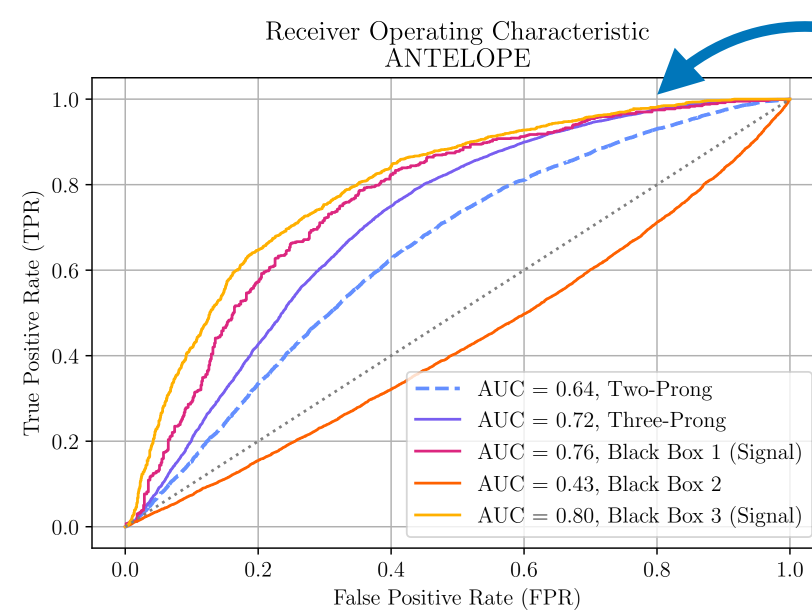
- We use the two-prong and QCD LHCO events to create the PFN embedding and then an orthogonal slice of QCD to train ANTELOPE





PFN sensitivity still best for 2/3-prong



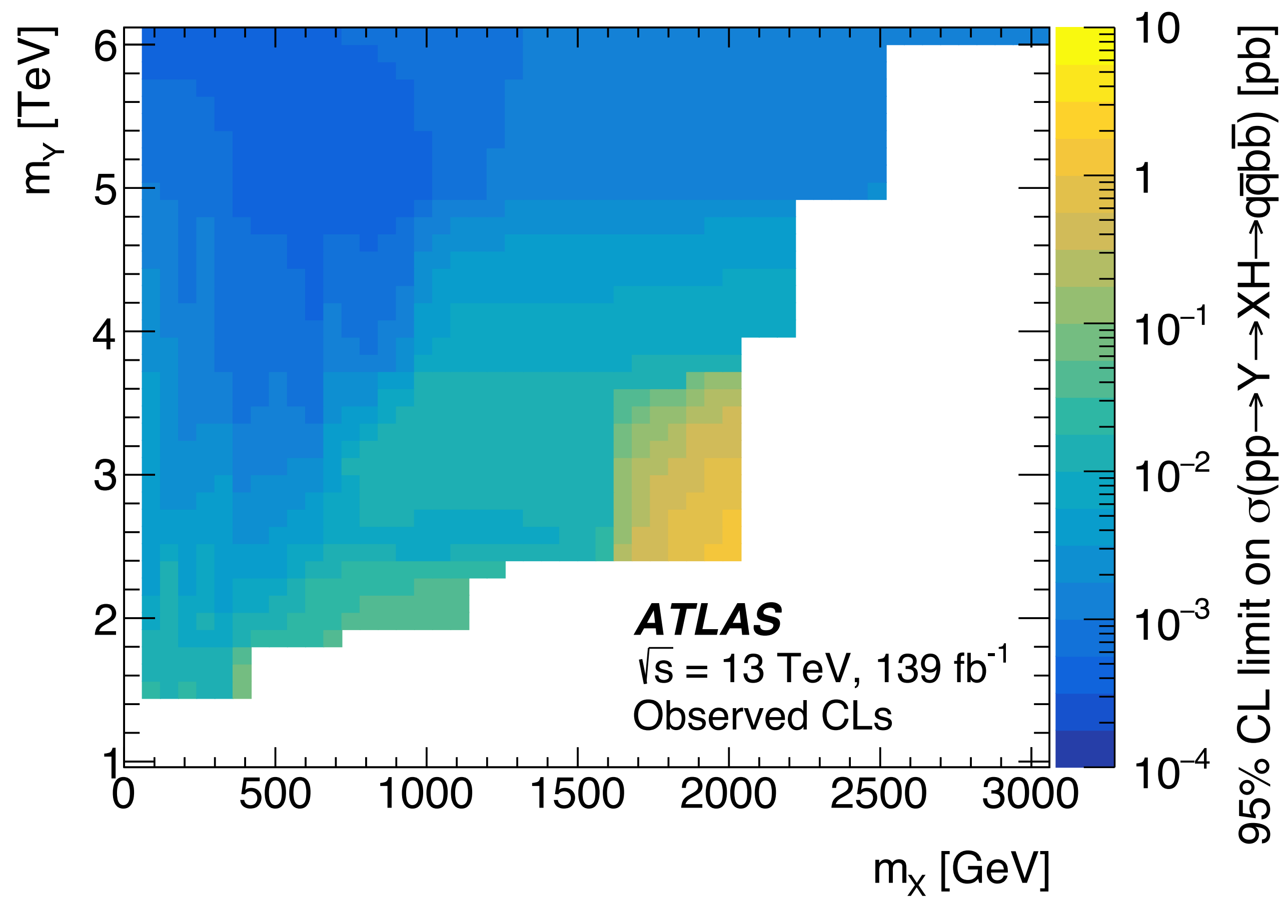


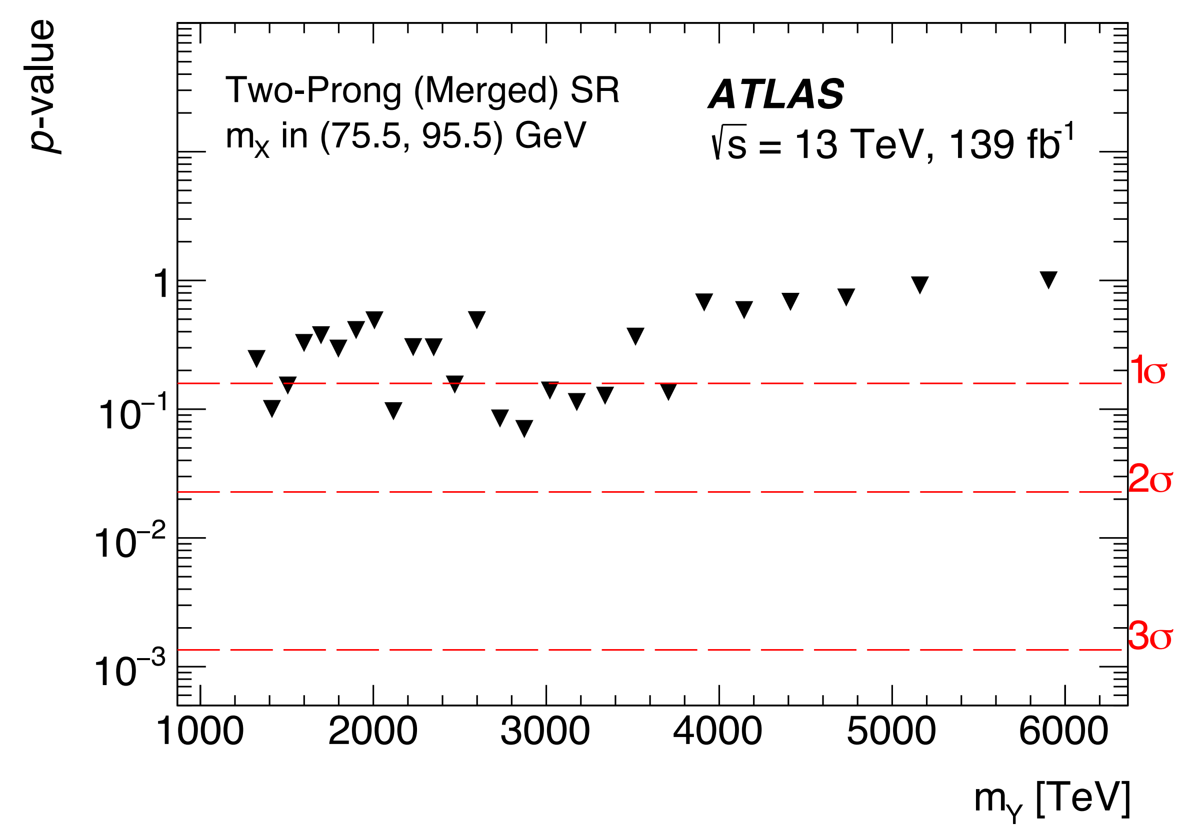
ANTELOPE generalizes for all signals & shows best sensitivity for BB3

In LHCO, no method was sensitive to BB3!

Y→XH: Exclusion Region Results

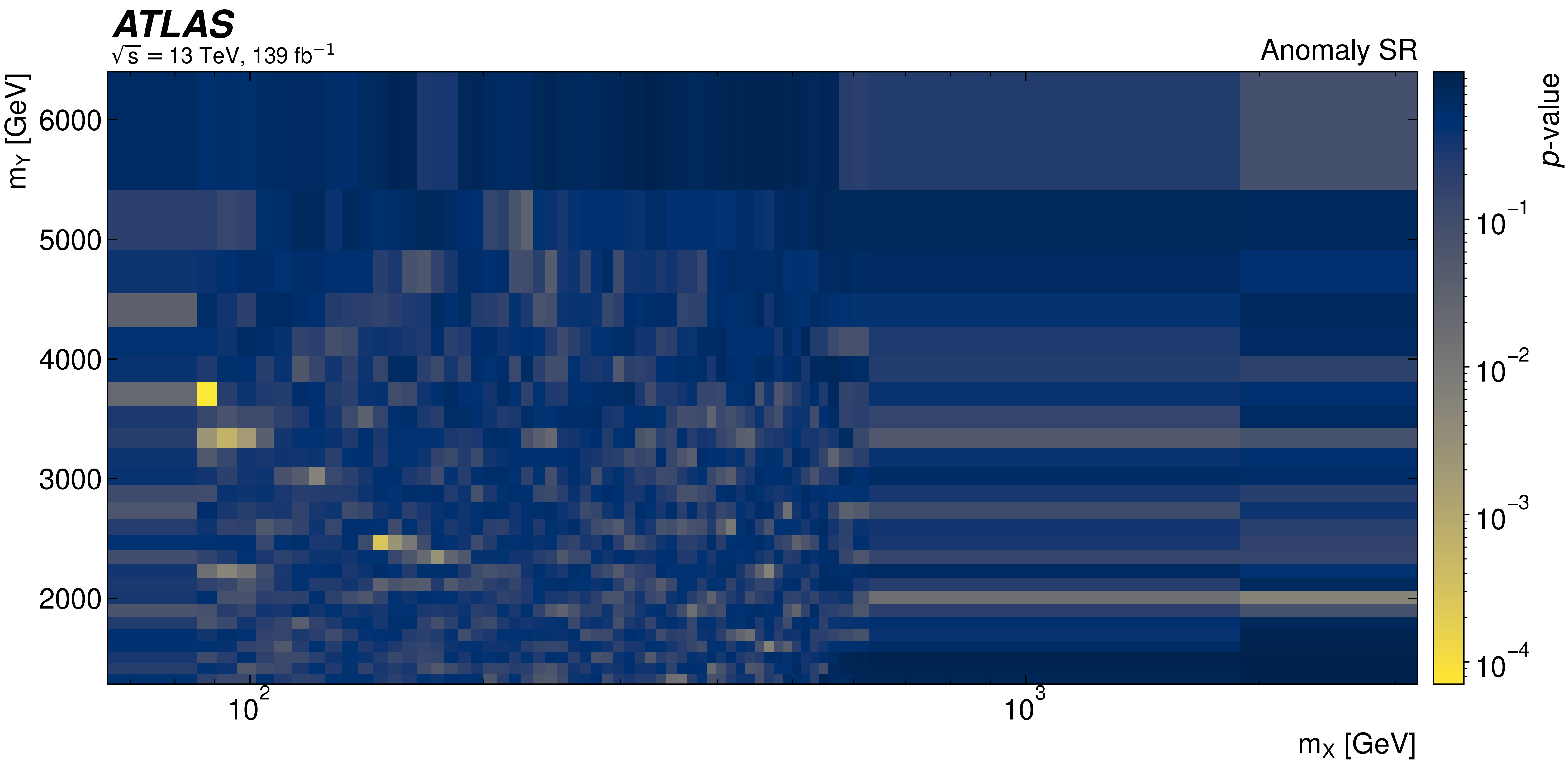
- Exclusion region optimized to select X→$q\bar{q}$ events using D_2 energy correlator
  - Perform fits across M_Y spectrum in optimized bins of M_X
  - Limits are derived on the production cross-section of Y→XH→$q\bar{q}b\bar{b}$





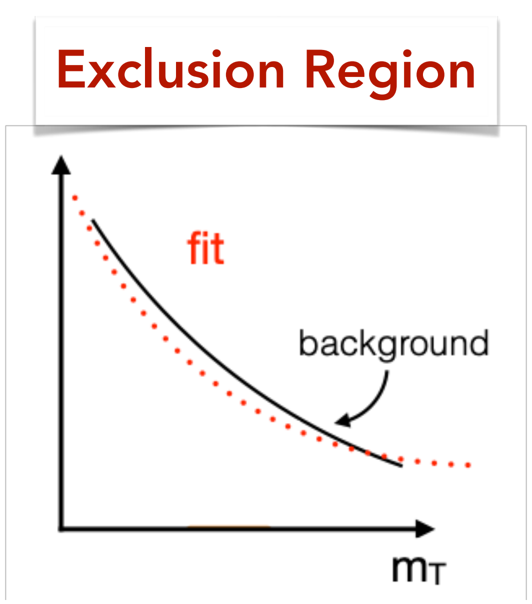
Y→XH: 2D BumpHunter Results

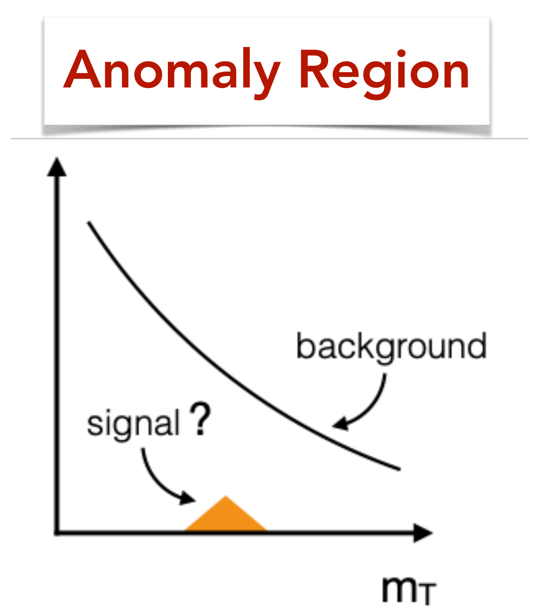
- **Global significance of 1.43σ when accounting for look-elsewhere effect**

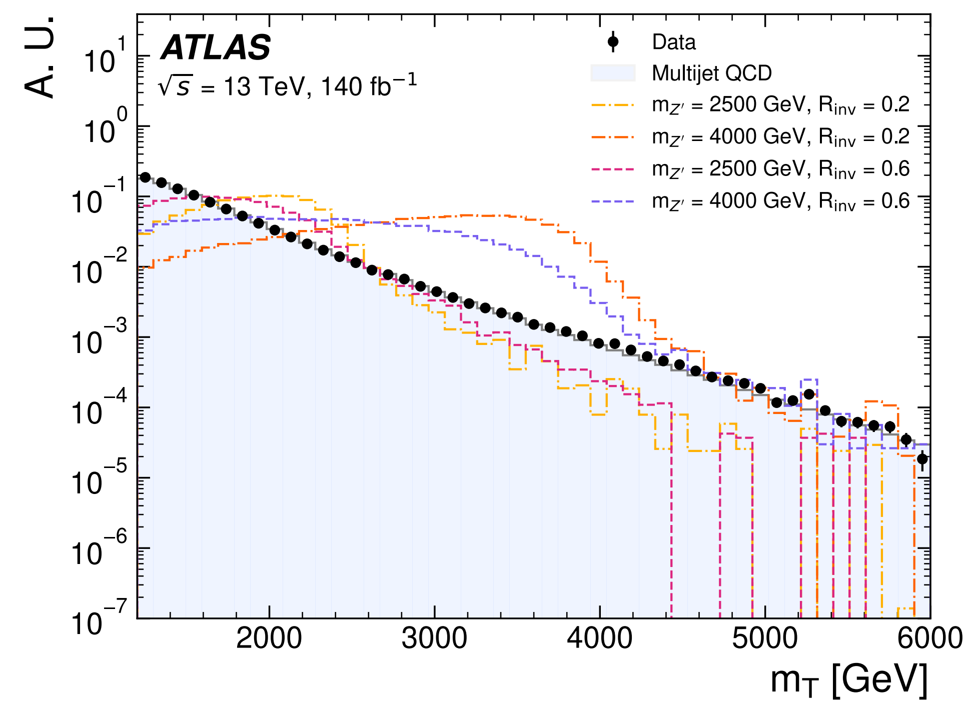


SVJ: Analysis Strategy (1)

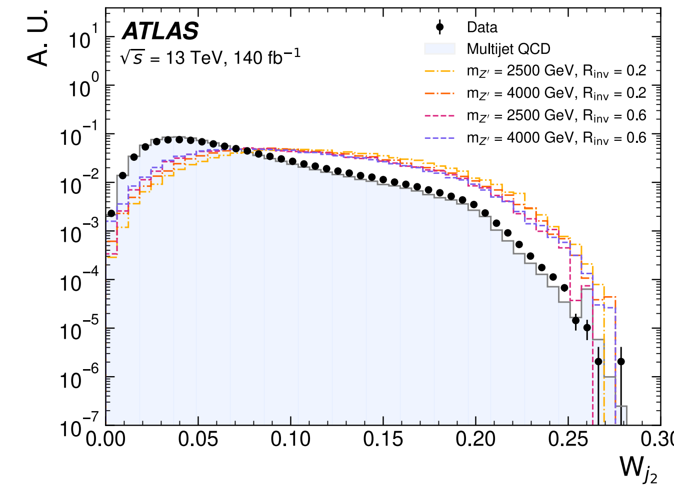
- Subtle shower differences between dark and SM QCD motivates the use of low-level track variables in the selection of SVJs and **two different ML strategies**
  - Exclusion region: Particle Flow Network (PFN) → *supervised ML*
    - Uses a functional form fit of m_T to define the background shape & set limits on signal cross section for SVJ signal model by leveraging track-level inputs
  - Anomaly detection region: ANTELOPE → *semi-supervised ML*
    - Uses the functional form fit of m_T to define the background shape & perform a bump hunt for any excesses (no SVJ model input)



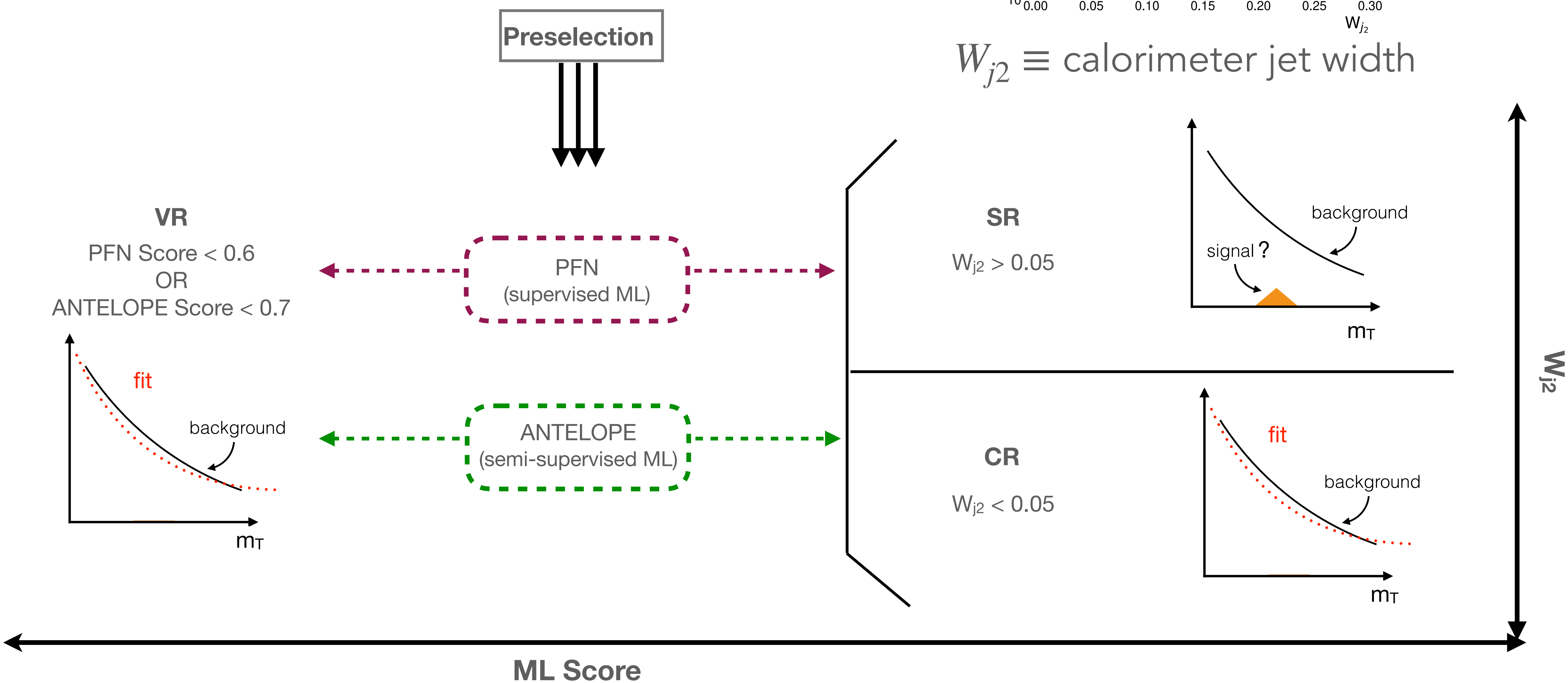




SVJ: Analysis Strategy (2)



$W_{j2} \equiv$ calorimeter jet width



SVJ: Exclusion Region Results

- Exclusion region defined from supervised classifier score of the PFN
  - SR defined from PFN score > 0.6 with W_{j2} > 0.05 selection
  - Limits set on the SVJ production cross-section at 95% CL