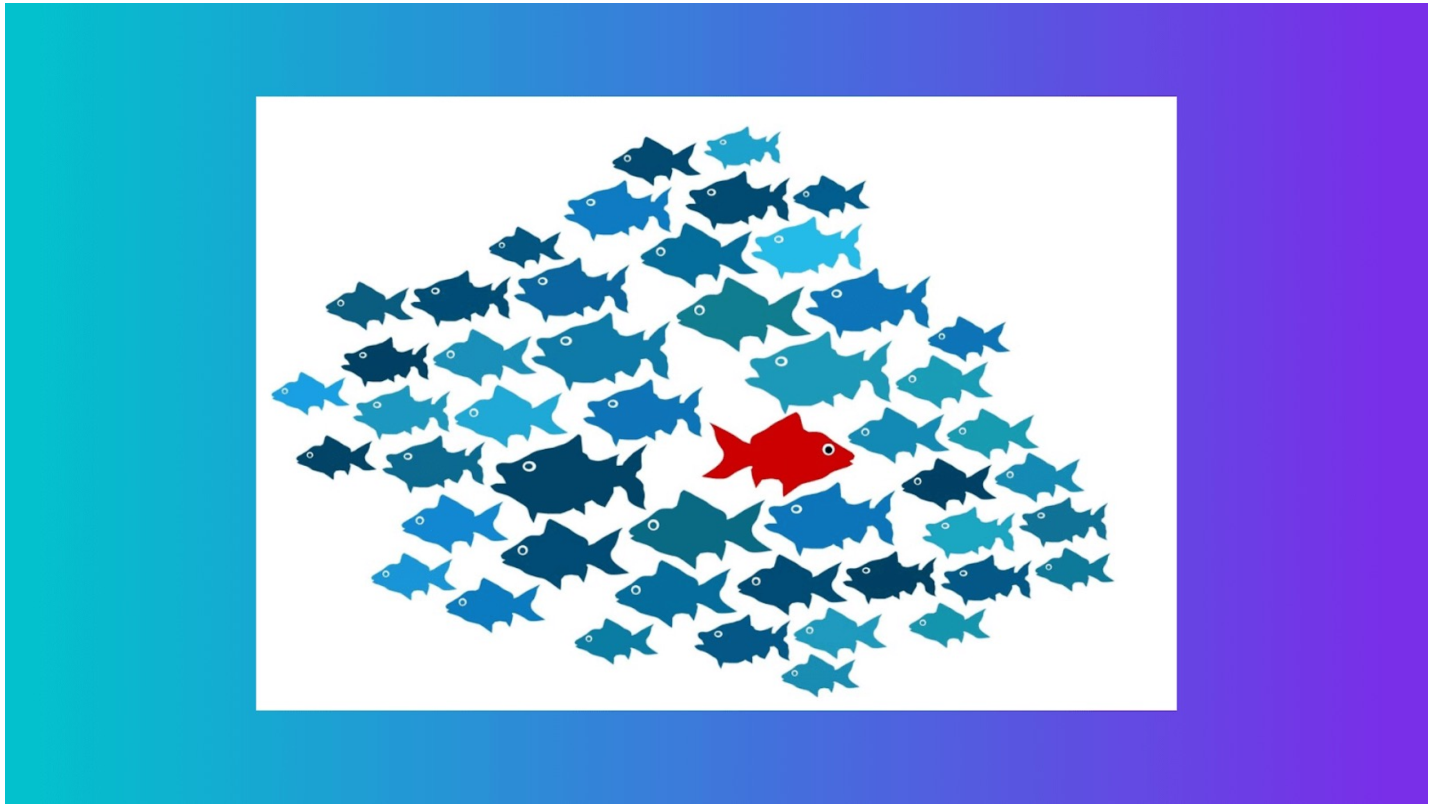
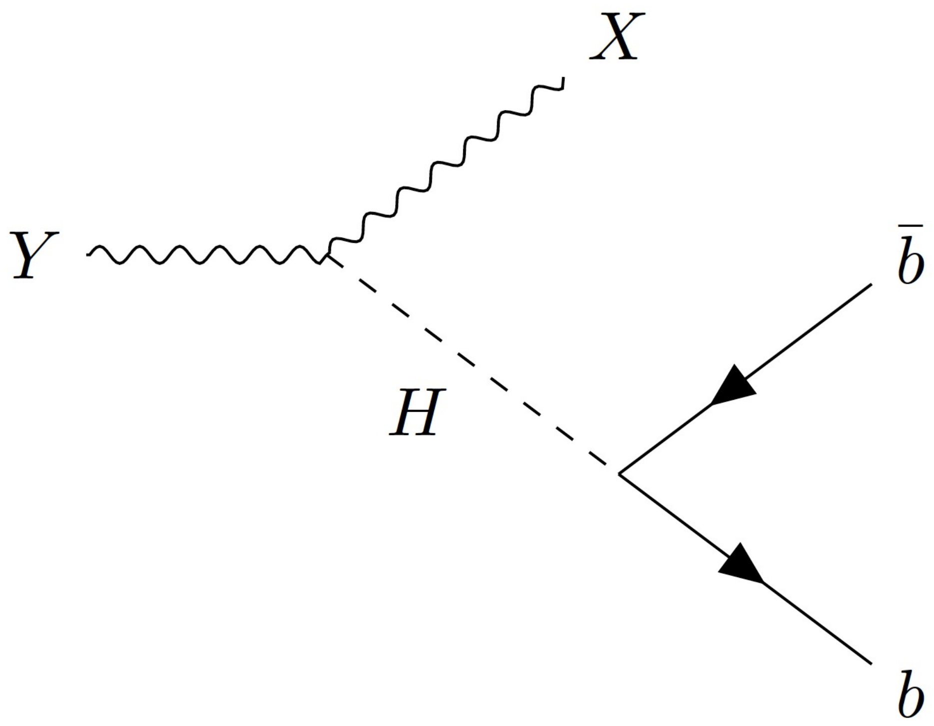
Anomaly detection search for new **resonances decaying** into a **Higgs** boson and a generic **new particle X** in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector



Outline

1. Theoretical Motivation
2. Analysis Regions
3. Anomaly detection with VRNNs
4. Background Estimation
5. Statistical Analysis and Results
6. Summary

Theoretical Motivation



Search for Boson *Y*, $m_Y \sim \mathcal{O}(1\text{ TeV})$ decay to:

- SM Higgs Boson H, with $\bar{b}b$ final state
- heavy boson *X*, $m_X \sim \mathcal{O}(10\text{ GeV-1 TeV})$ with hadronic final state

Why interesting?

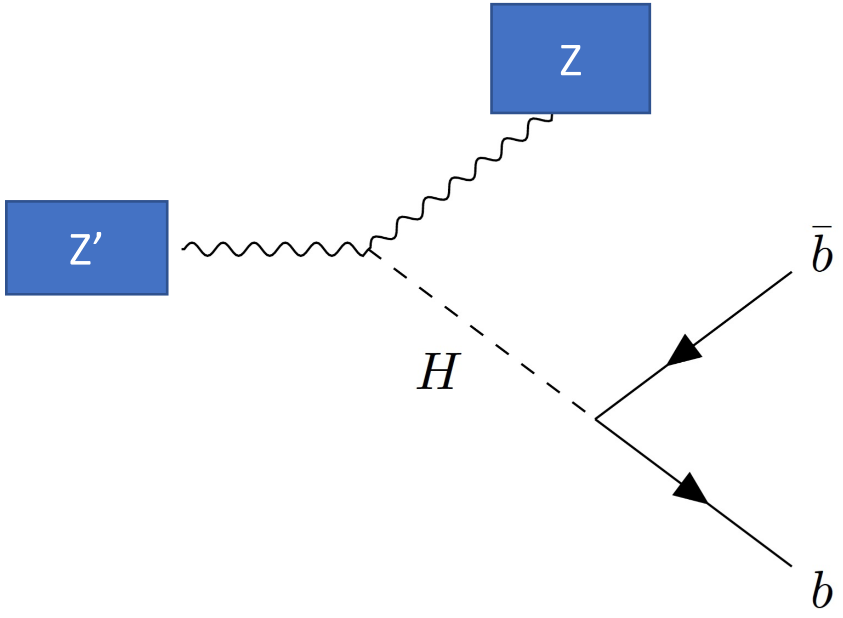
- SM needs extension!
- Many extensions propose new particles, which interact with SM bosons, like Higgs

Extended Gauge Sectors

- Can we unify all forces into one fundamental force?

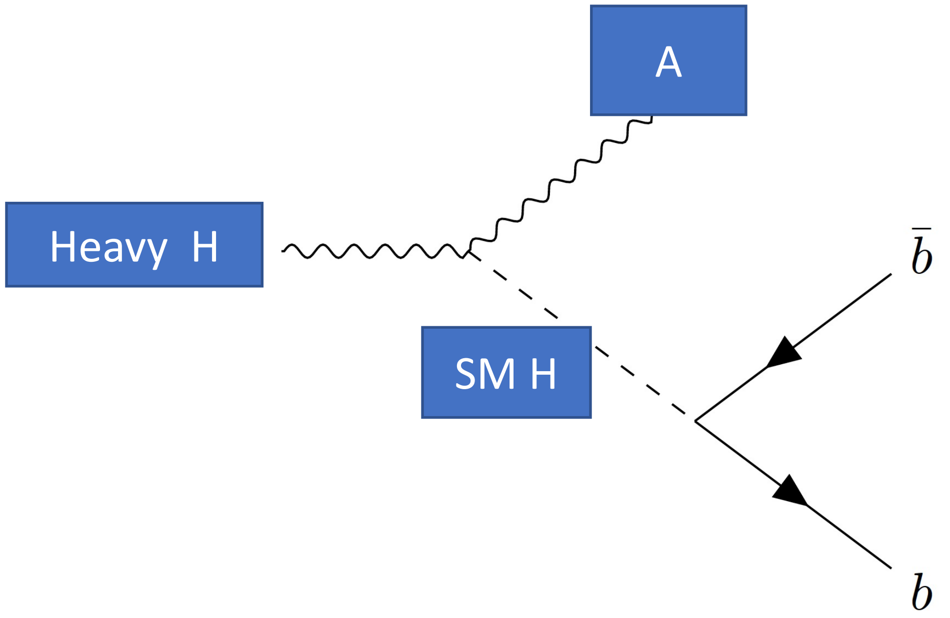
⇒ Grand Unified Theory (GUT)

- SU(5), E6, SO(10)...
- Ex.: $SO(10) \supset SU(3) \otimes SU(2) \otimes \boldsymbol{U(1)} \otimes U(1) \supset SM$
- additional heavy gauge boson **Z'**
- Interacts with SM bosons and Higgs

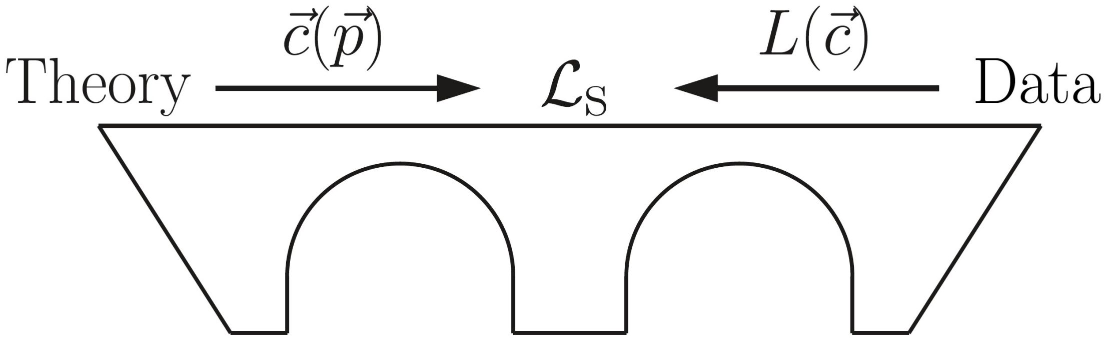


Two Higgs Doublet Models

- Can there be more than one Higgs?
- Motivation in Supersymmetry, Baryogenesis,...
- Simplest extension: Two Higgs Doublet model
- 5 physical d.o.f. after SSB
- multiple "Higgs bosons":
  - h (SM)
  - Neutral Scalar H (heavier)
  - Pseudoscalar A
  - Charged scalar $H^{\pm}$



Bridge Model: Heavy Vector Triplet

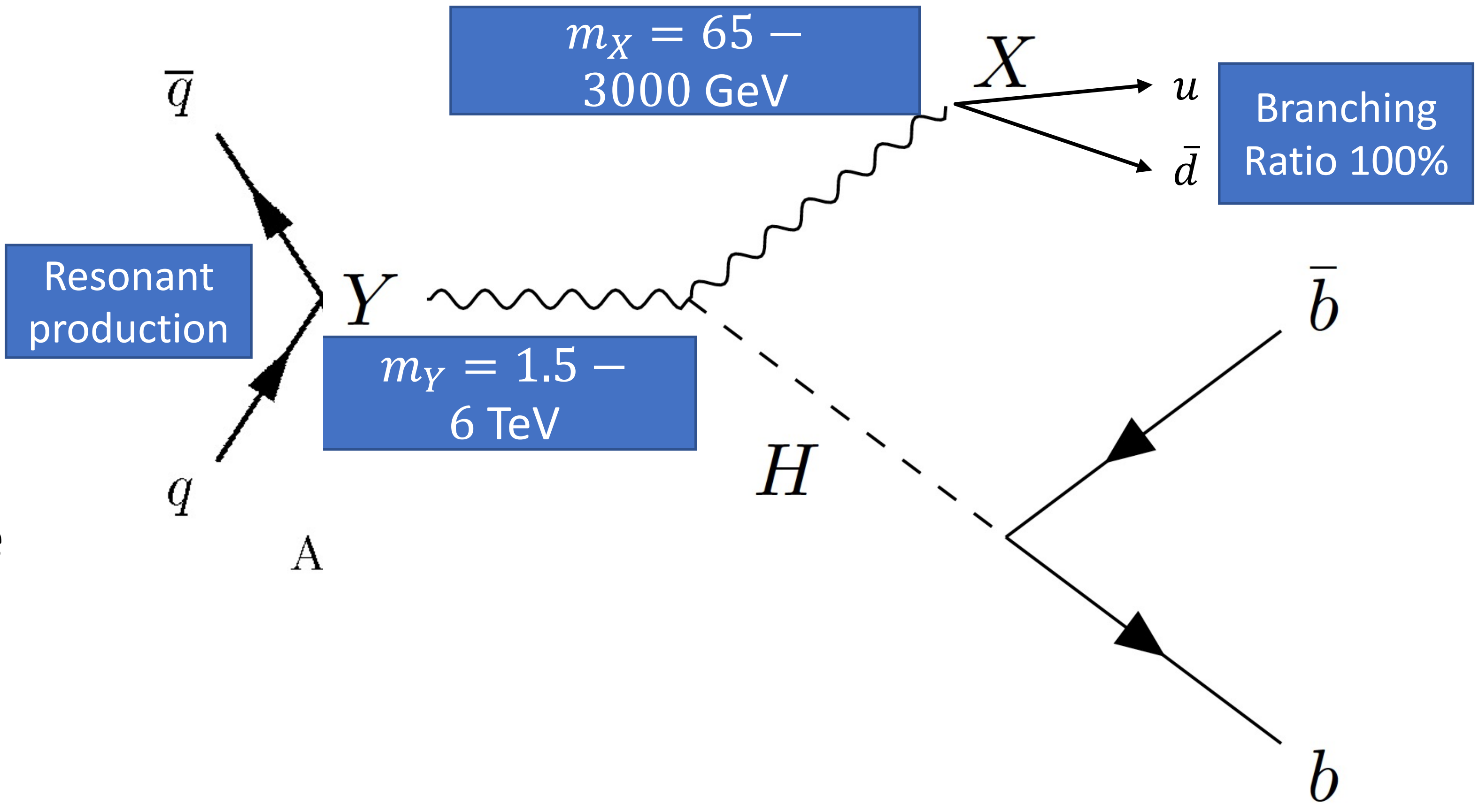


Simplified model used here:

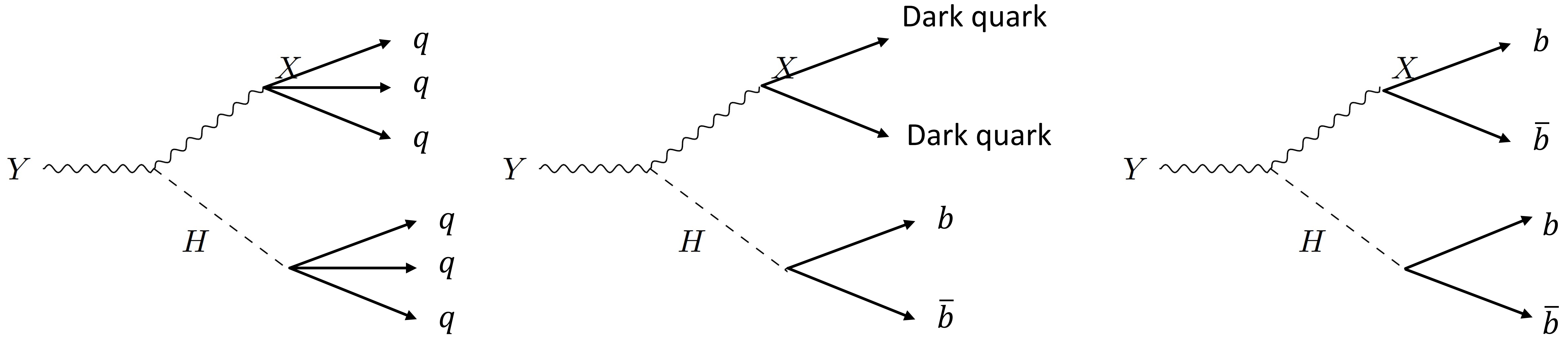
- SM + **3** massive vector bosons: $\boldsymbol{V^{\pm}, V_0}$
- Couple to SM fermions like W,Z bosons
- Couple to W,Z bosons as well

Monte Carlo Simulations

- Simulate the signal
- Necessary for:
  - signal+background fits
  - Assess model independence
- Output is same as data
- Using the HVT model
- Include pileup and up-to date calculations of PDFs



Alternative Signatures



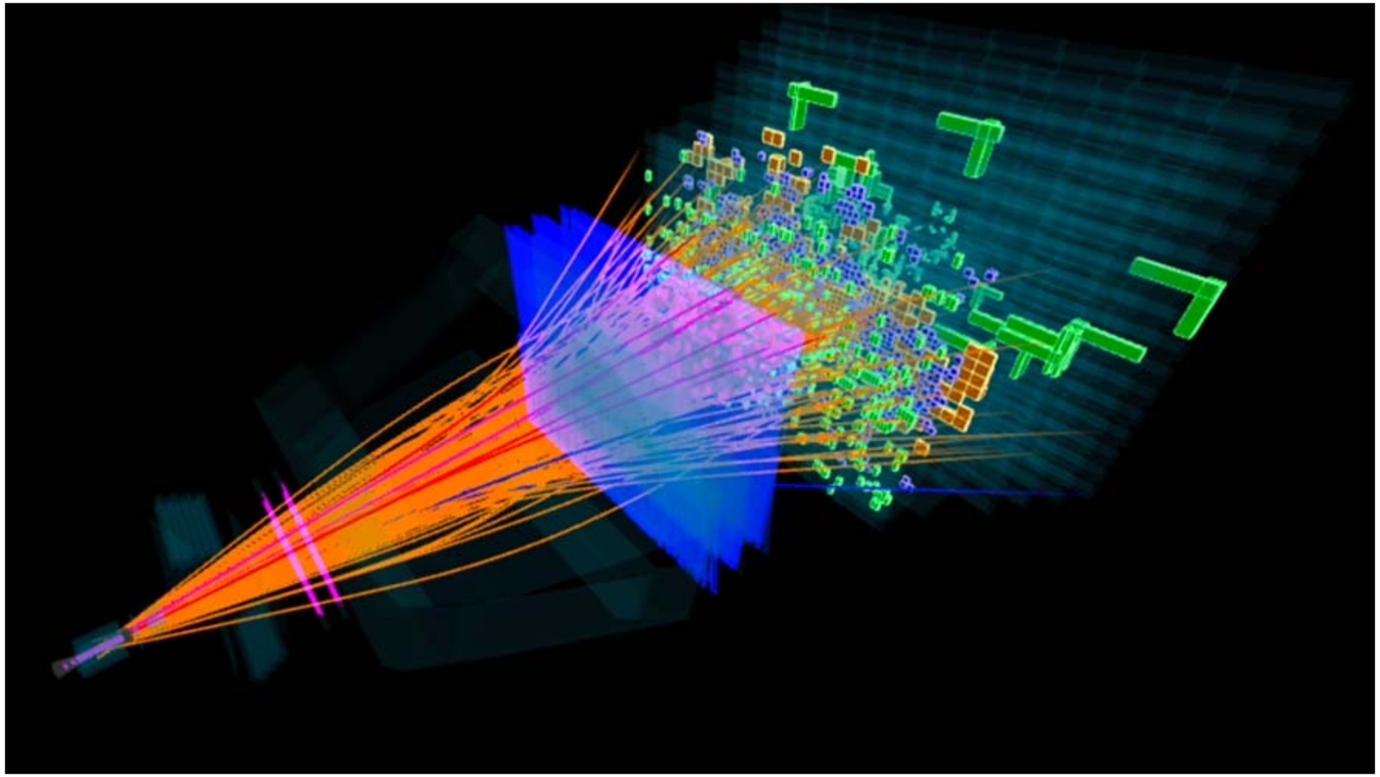
Test Anomaly detection method for
model independence!

- Before starting to analyse our data, we still need to obtain it in the right form!

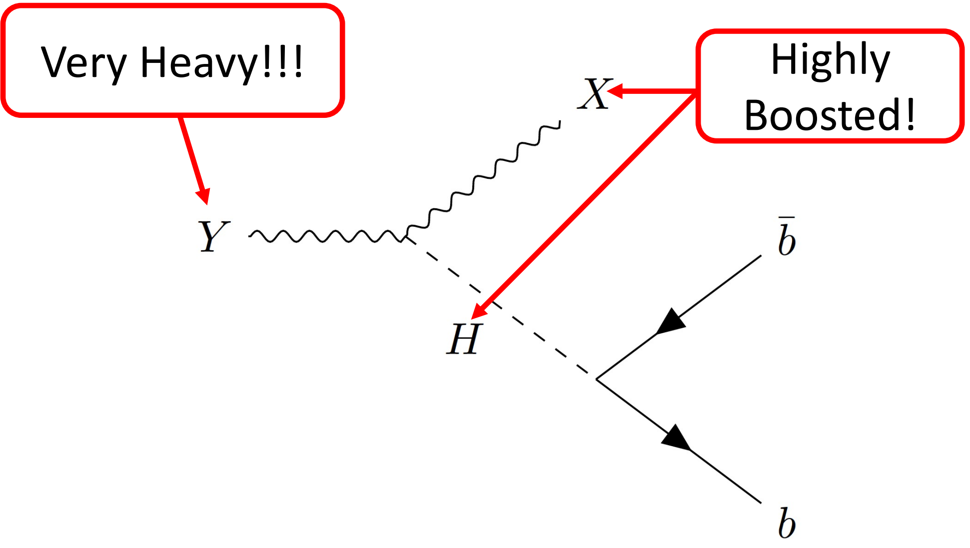
- How do we disentangle them from the massive amounts of data at our disposal without leaving out objects of interest?

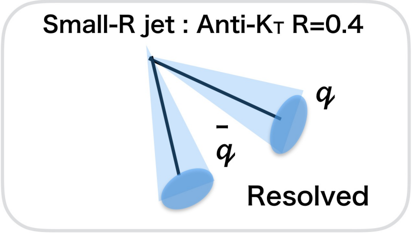
- What kind of experimental signature do we expect?

- What are the objects we are looking for?

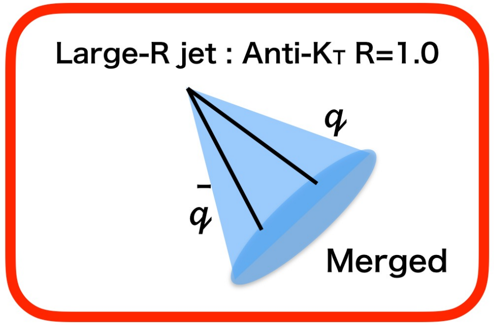


Experimental Signature





From jet constituents
$p_T > 20$ GeV
$|\eta| < 4.7$



$p_T > 200$ GeV
$|\eta| < 2.0$
Trimmed

<u>Qualitatively:</u>

- Heavy Y decays to high energy X,H
- Collimated decay products
- Reconstructed as large R jets!
- substructure analysed to distinguish from background
- Leptons are not used in the analysis

<u>Quantitatively</u>:

- Trigger: presence of a large R jet
- Keep if $p_T > 500$ GeV and $m_{JJ} > 1.3$ TeV
- 2 leading large R jets kept if $m_{Ji} > 50$ GeV
- Small R-jets constructed from constituents (later)

Open Problems:

1. Which jet corresponds to a Higgs and which to an X?

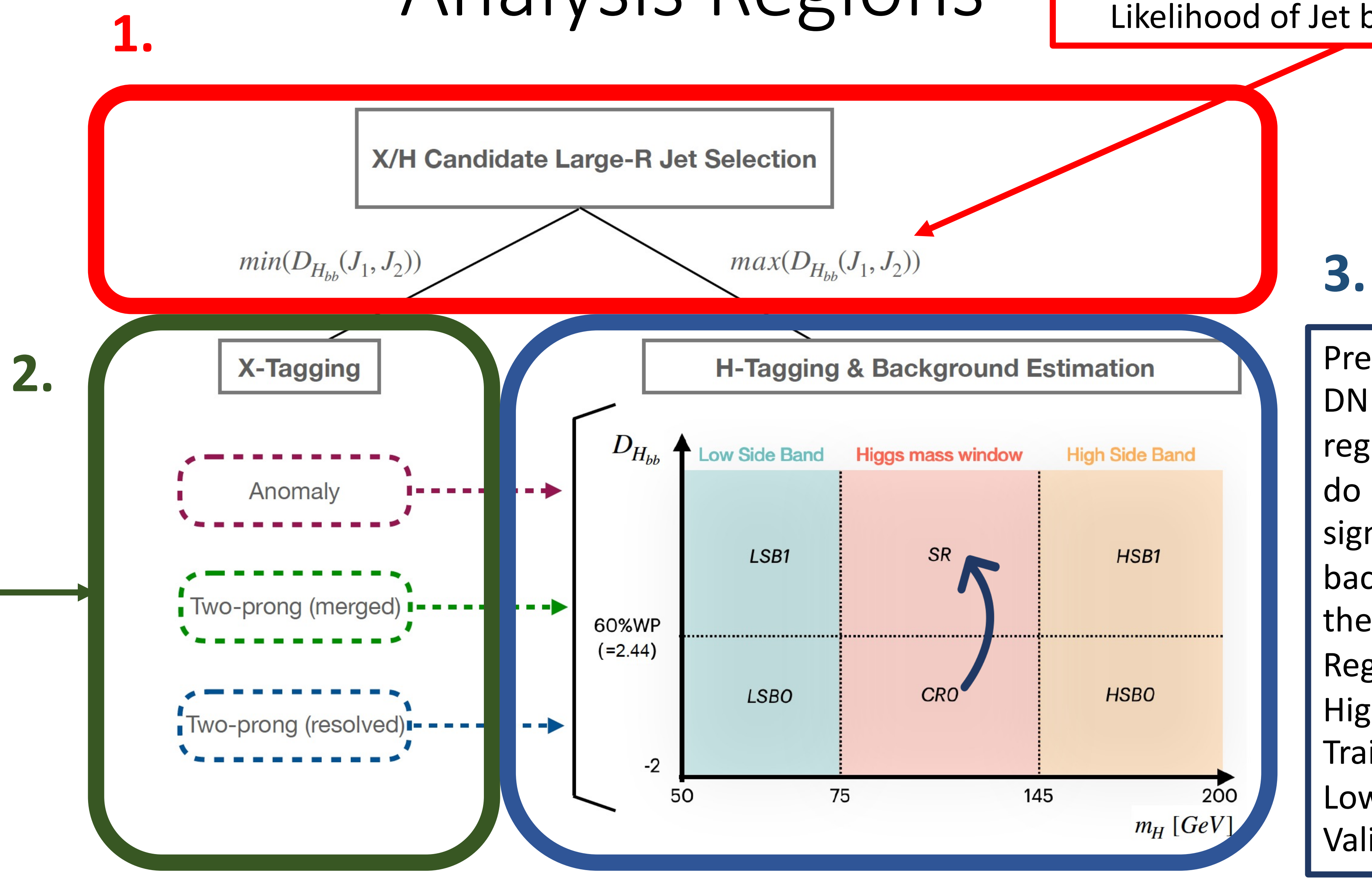
2. What is the X mass?

➔ We need to cover **all possible kinematics**!

3. How do we estimate our background in a **model independent** way?

Analysis Regions

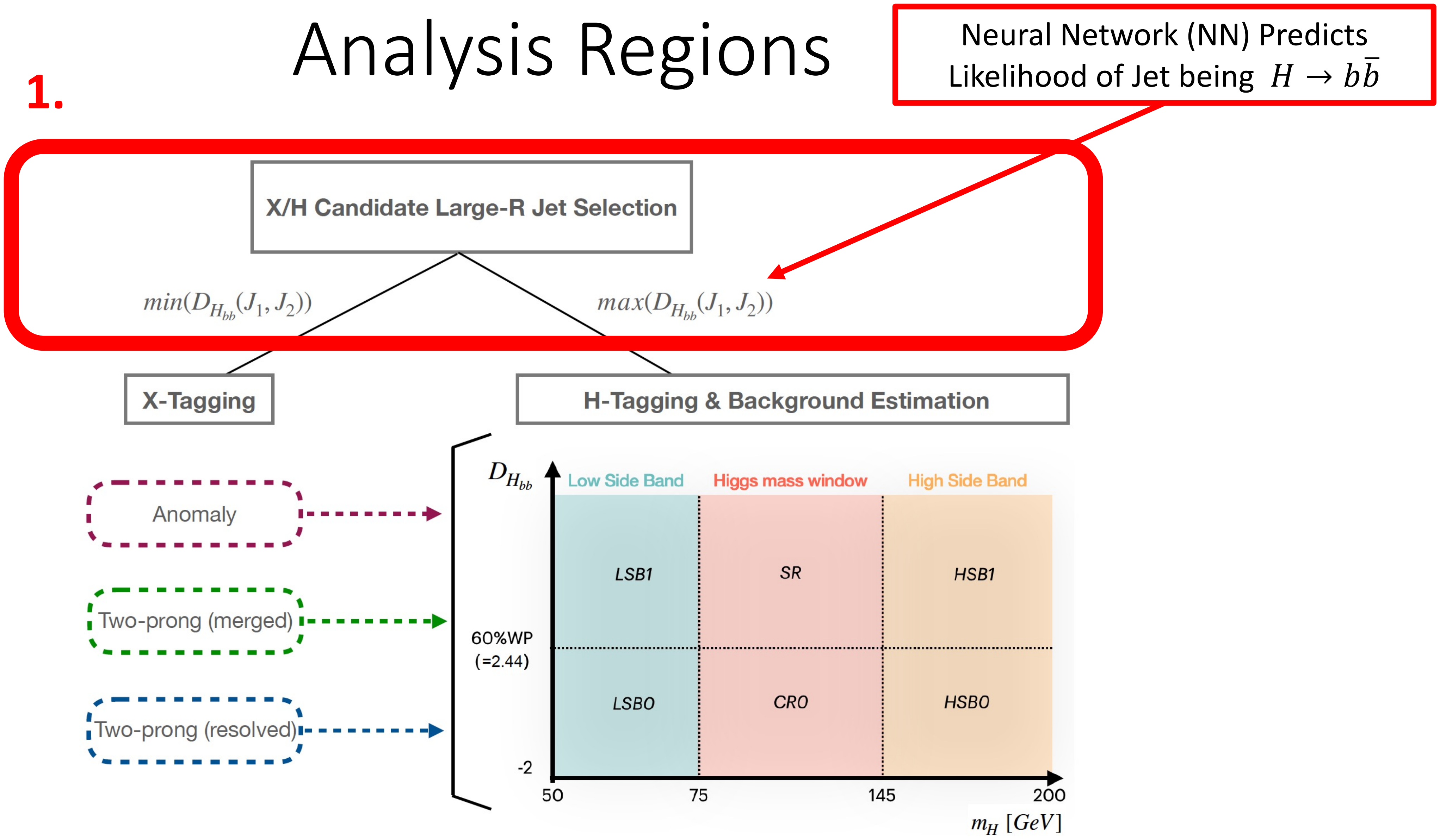
Neural Network (NN) Predicts Likelihood of Jet being $H \rightarrow b\bar{b}$



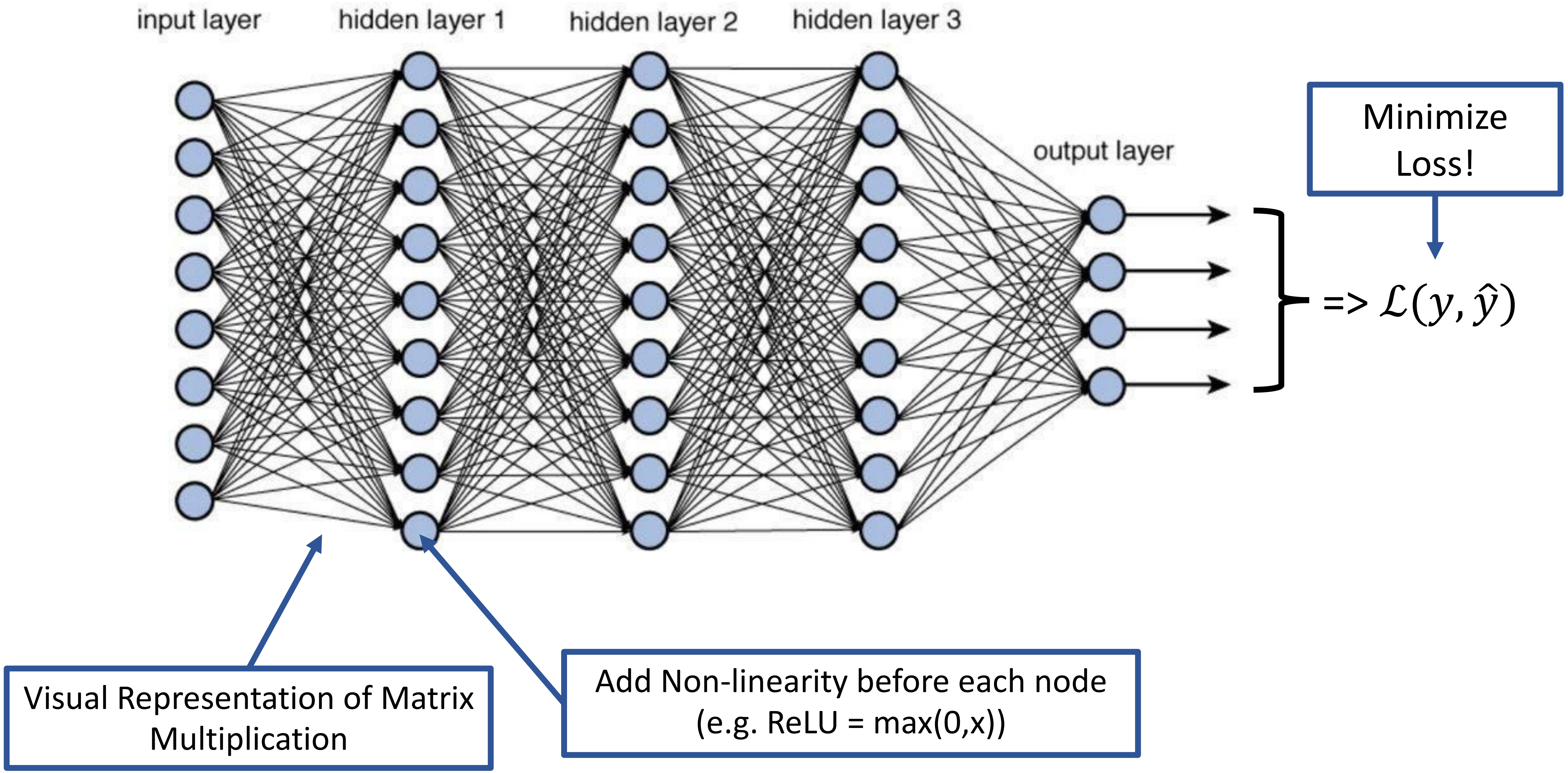
Cover all kinematics depending on X mass

Predict with a DNN from lower region where we do not expect a signal the background in the Signal Region!
High Side Band: Training
Low Side Band: Validation

Analysis Regions



Neural Networks Basics



X/H-Jet Candidate selection

Input:

- Large R jet variables (pT, eta) for jet with two or three subjets
- Output of high level single b taggers DL1r for each jet above a threshold with variable radius :

  -> p_c , p_b , $p_{lightjet}$

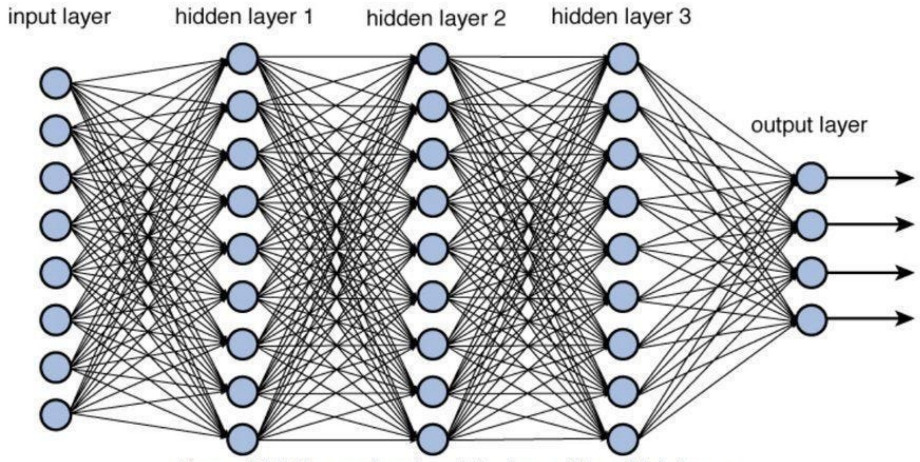
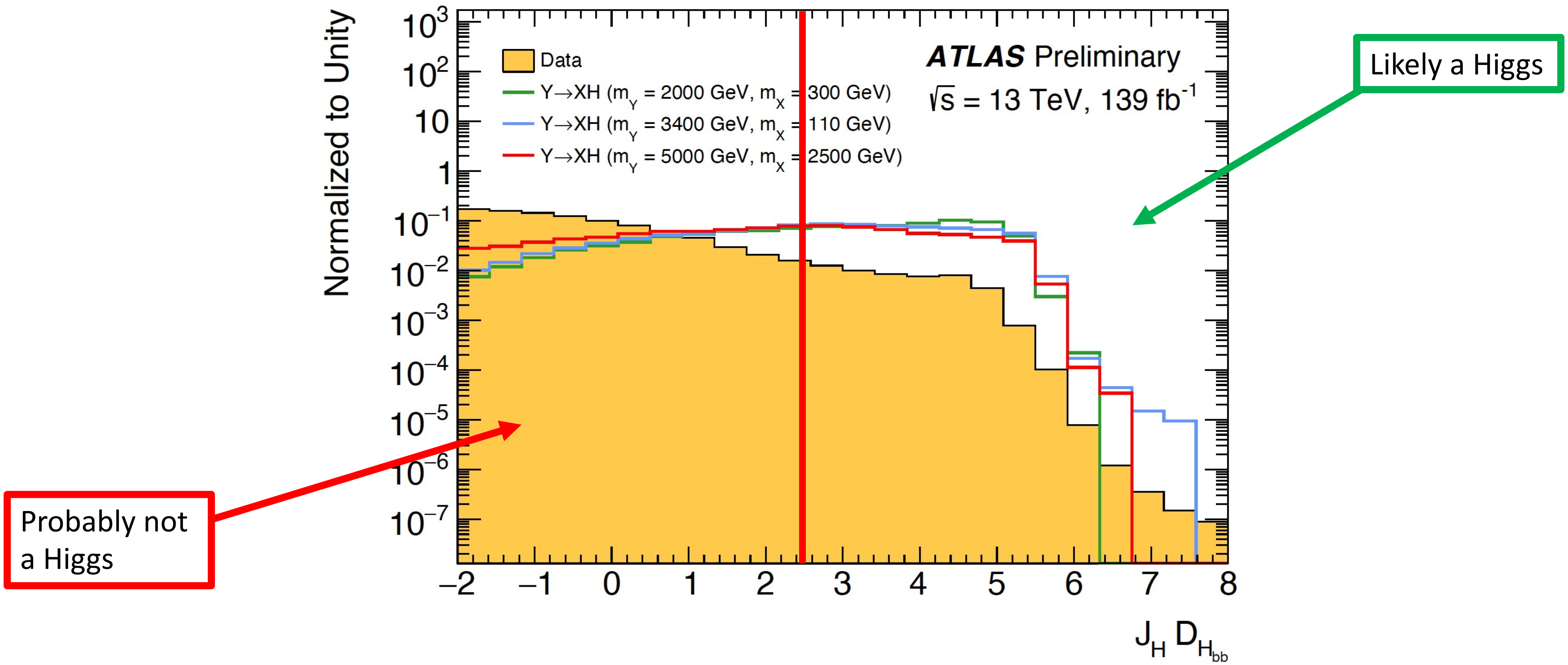


Figure 12.2 Deep network architecture with multiple layers.

$$\begin{pmatrix} p_{Higgs} \\ p_{top} \\ p_{multijet} \end{pmatrix}$$

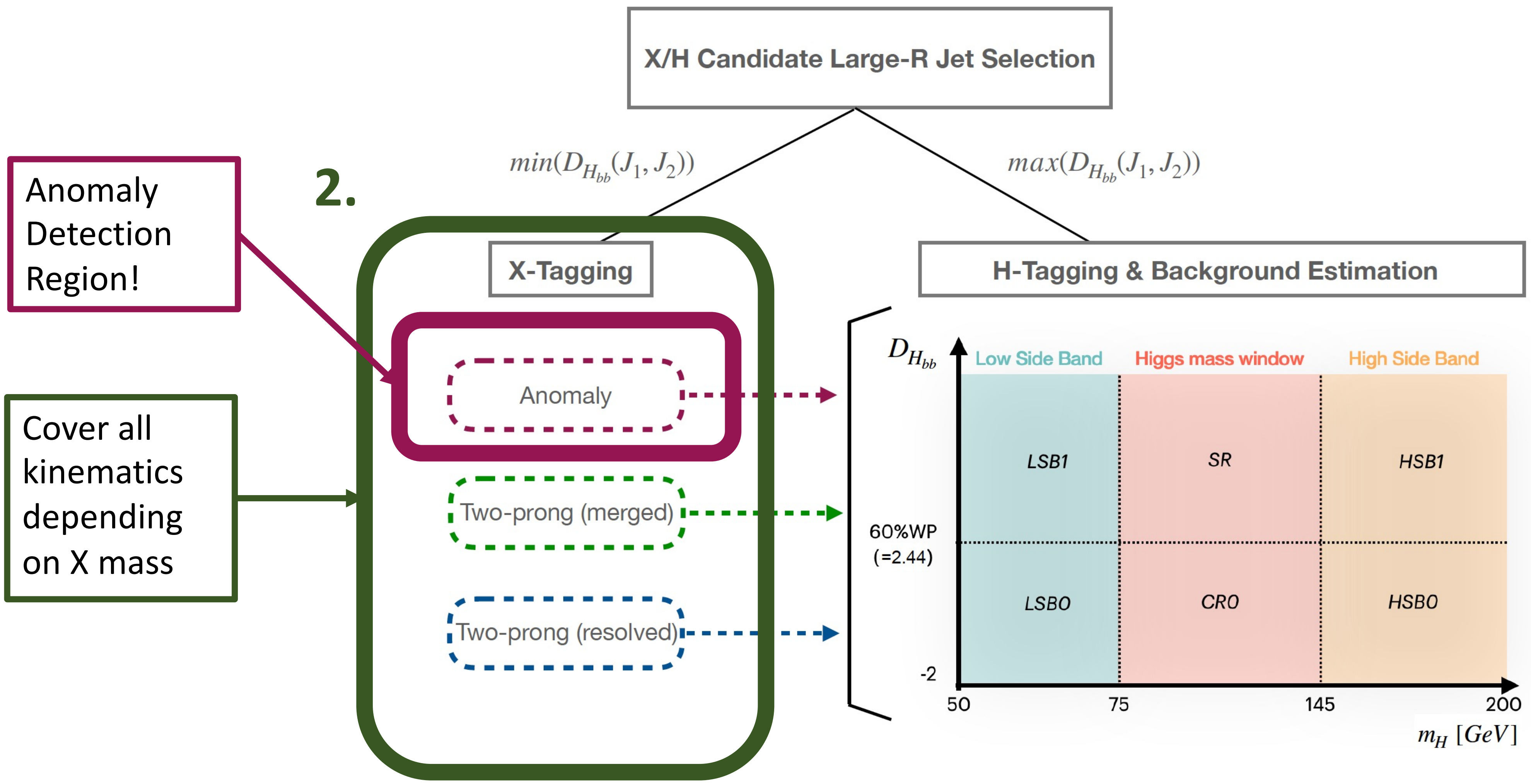
$$D_{H_{bb}} = \ln \frac{p_{\text{Higgs}}}{f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{\text{top}}) \cdot p_{\text{multijet}}}$$

“Logarithmic difference in the probabilities of the jet being a Higgs”

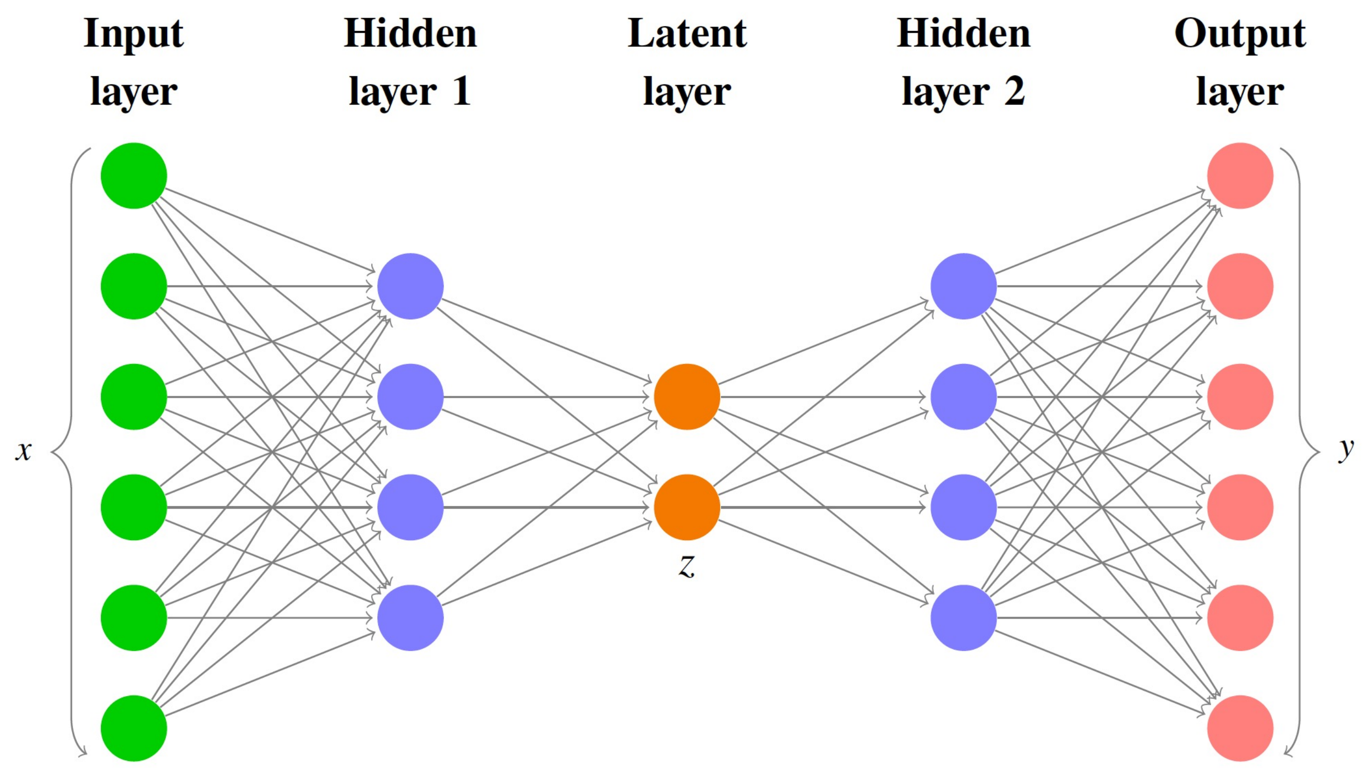


$D_{H_{bb}} = 2.44$
60% Working Point

Analysis Regions



Anomaly Detection: Autoencoder



Idea:

- Encoder: **Reduces** input vector to latent vector z (Extract Features)
- Decoder: **Reconstructs** input x from z, i.e. **y = f(z) ≈ x**
- Train with LHC data -> **bad** reconstruction for **unknown** signals

⇒ **Anomalies in tails of $\mathcal{L}$ distribution!**

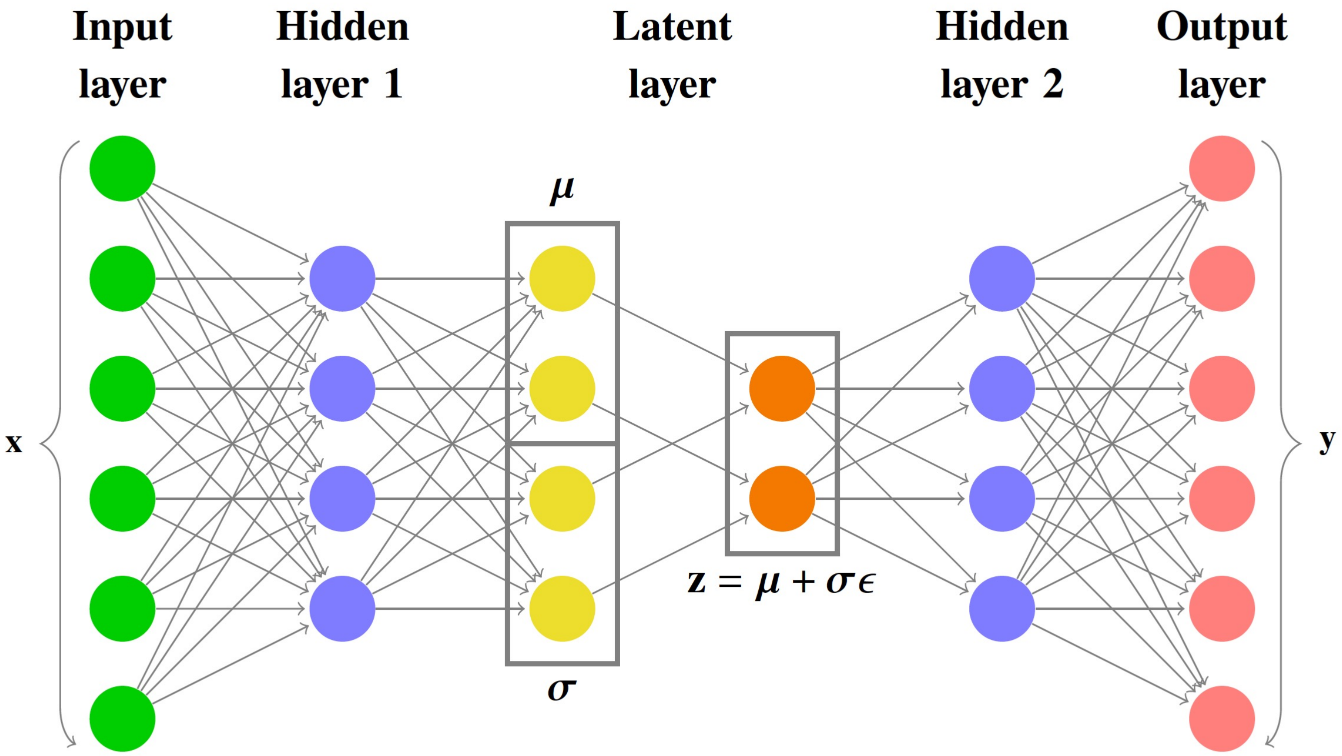
Loss Function:

$$\mathcal{L}(x, y) = |x - y|^2$$

Positive:

- Unsupervised learning

Variational Autoencoder



Idea:

- **x is generated randomly** from some underlying **distribution p(z)** -> latent layer **approximates** this distribution with **q(z,x)**

Loss Function contains **Kullback-Leibler Divergence**:

$$\mathcal{L}(x, y) = |x - y|^2 + D_{KL}(\, q(z|x) \,||\, p(z) \,)$$

Expectation Value of Log difference of PDFS

Anomaly Score:

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

Positive:

- Unsupervised learning

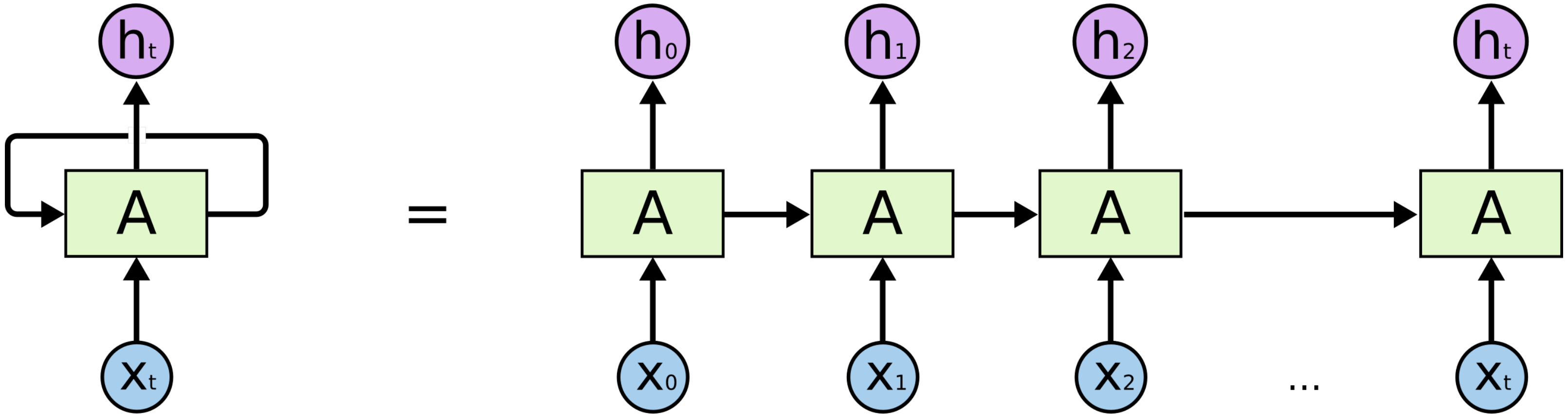
Drawback:

- **Fixed length input data**

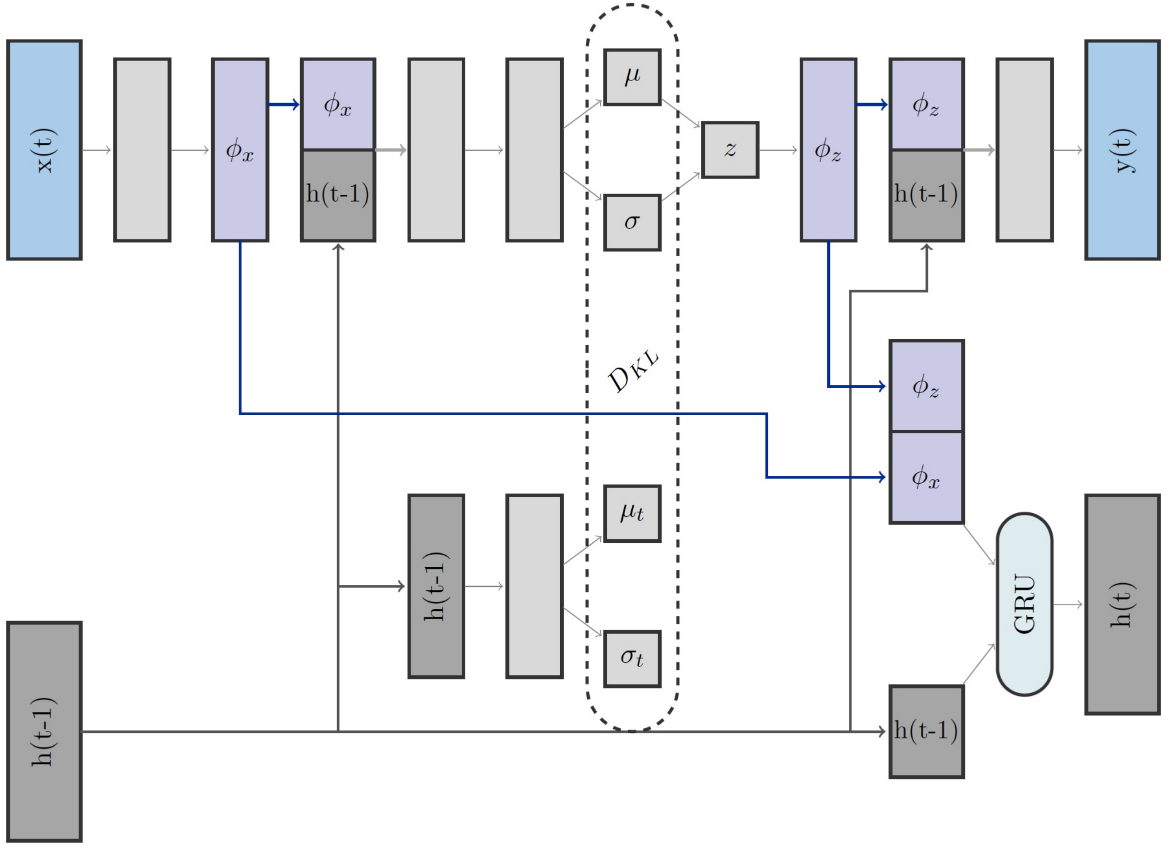
Recurrent Neural Networks (RNNs)

Idea:

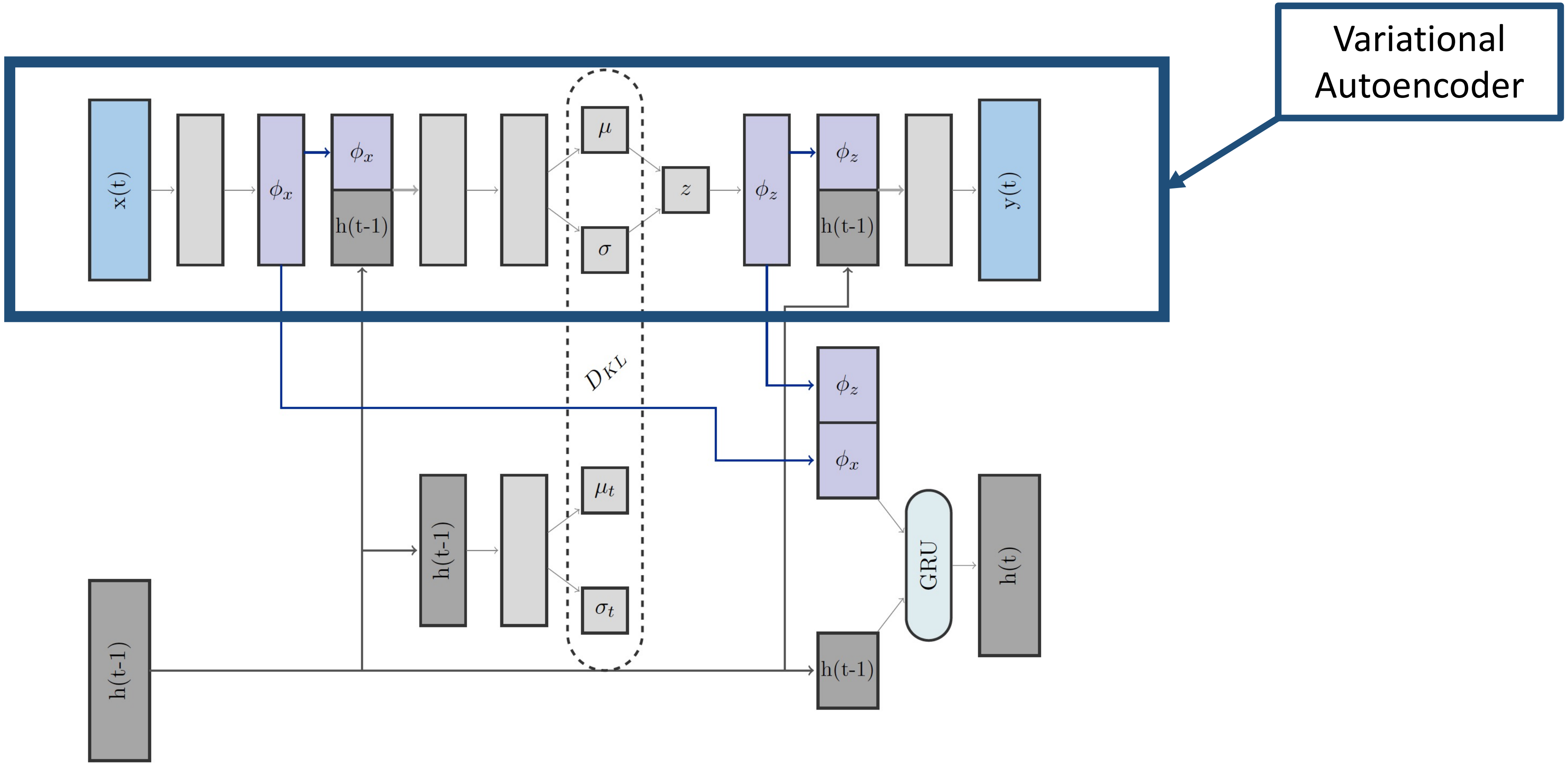
- Input is a **variable length sequence of fixed length objects** (e.g. jet constituents)
- Each step: hidden state calculated that passes on information from all previous time steps!



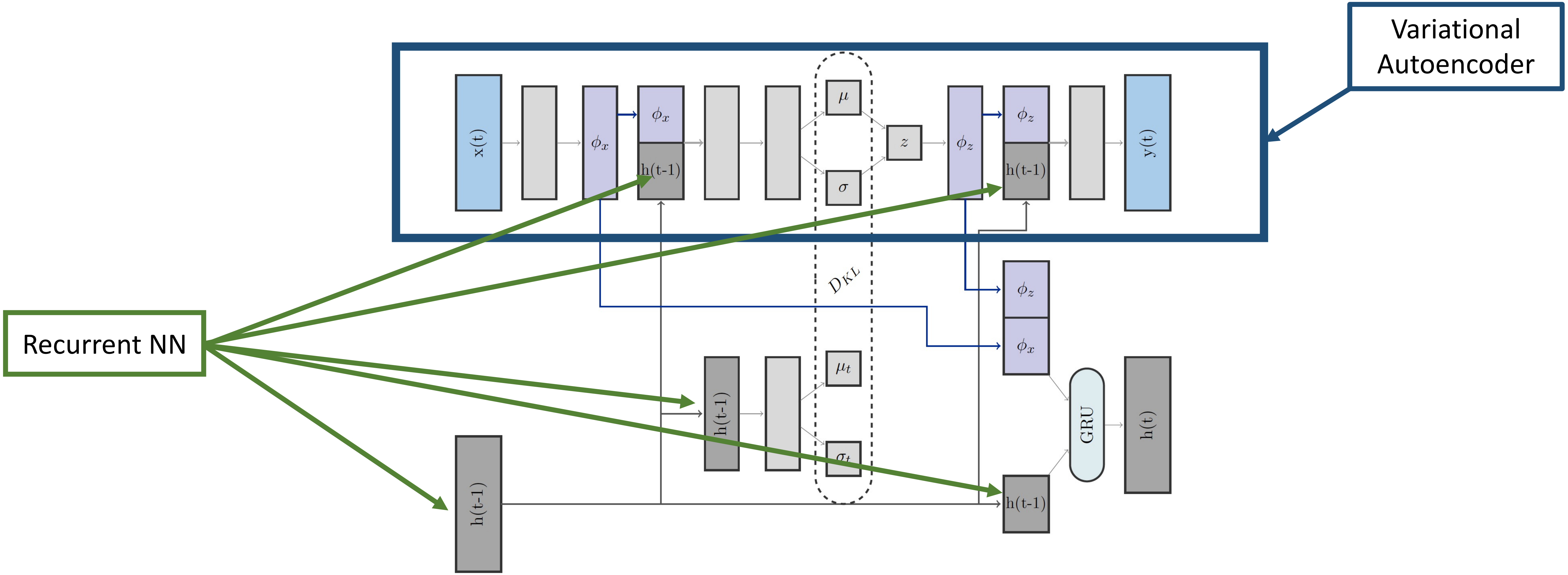
Variational Recurrent Neural Networks (VRNNs)



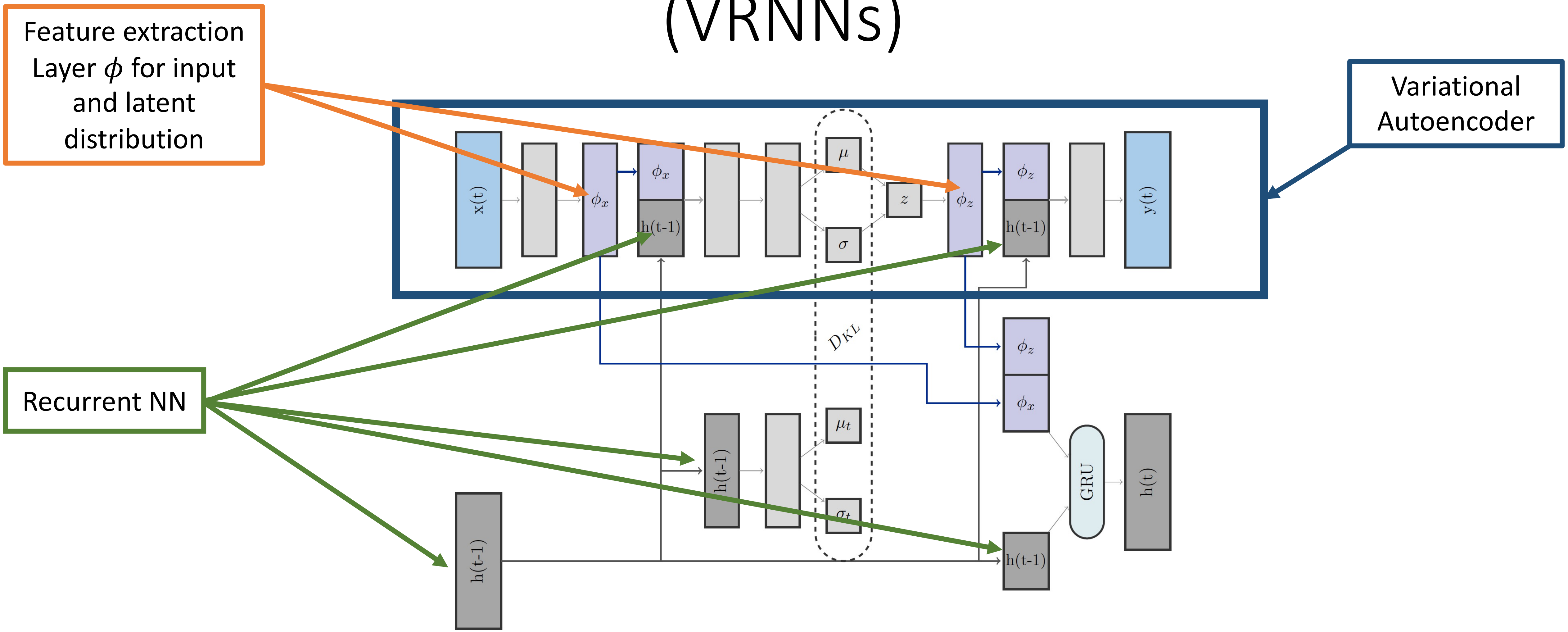
Variational Recurrent Neural Networks (VRNNs)



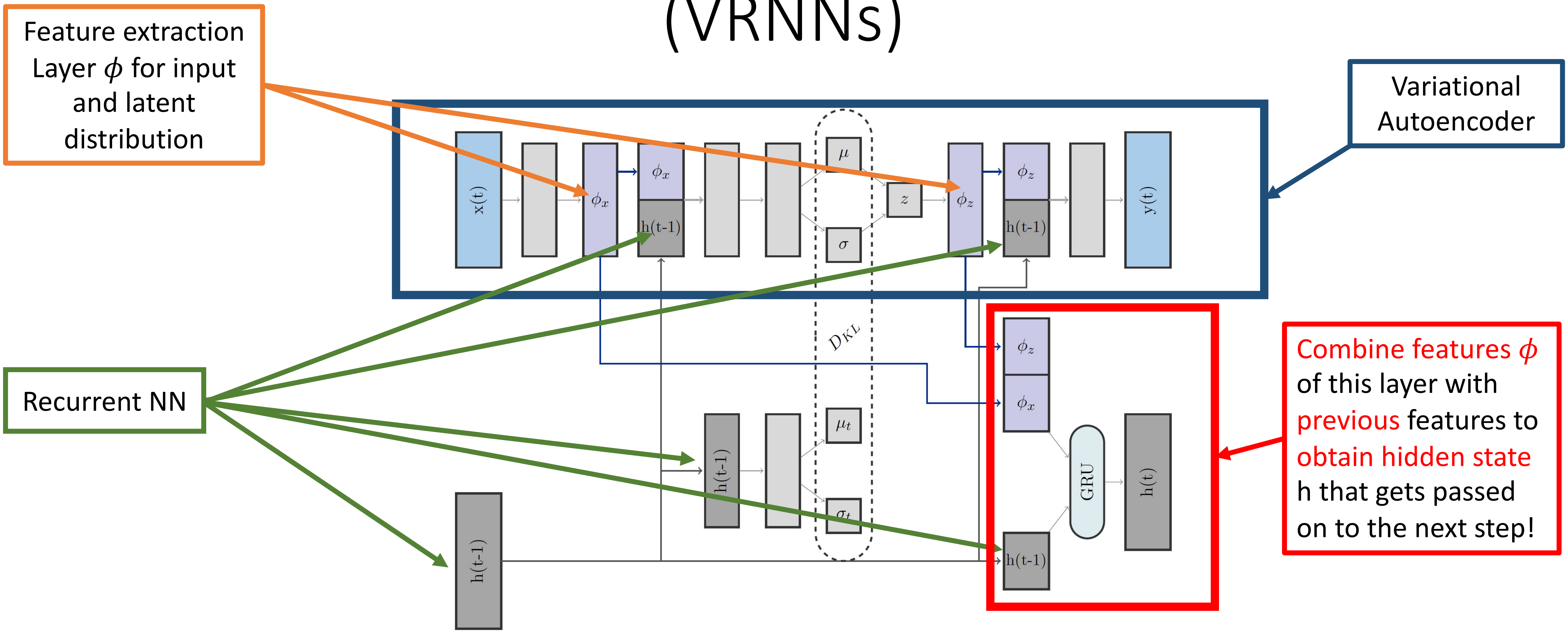
Variational Recurrent Neural Networks (VRNNs)



Variational Recurrent Neural Networks (VRNNs)



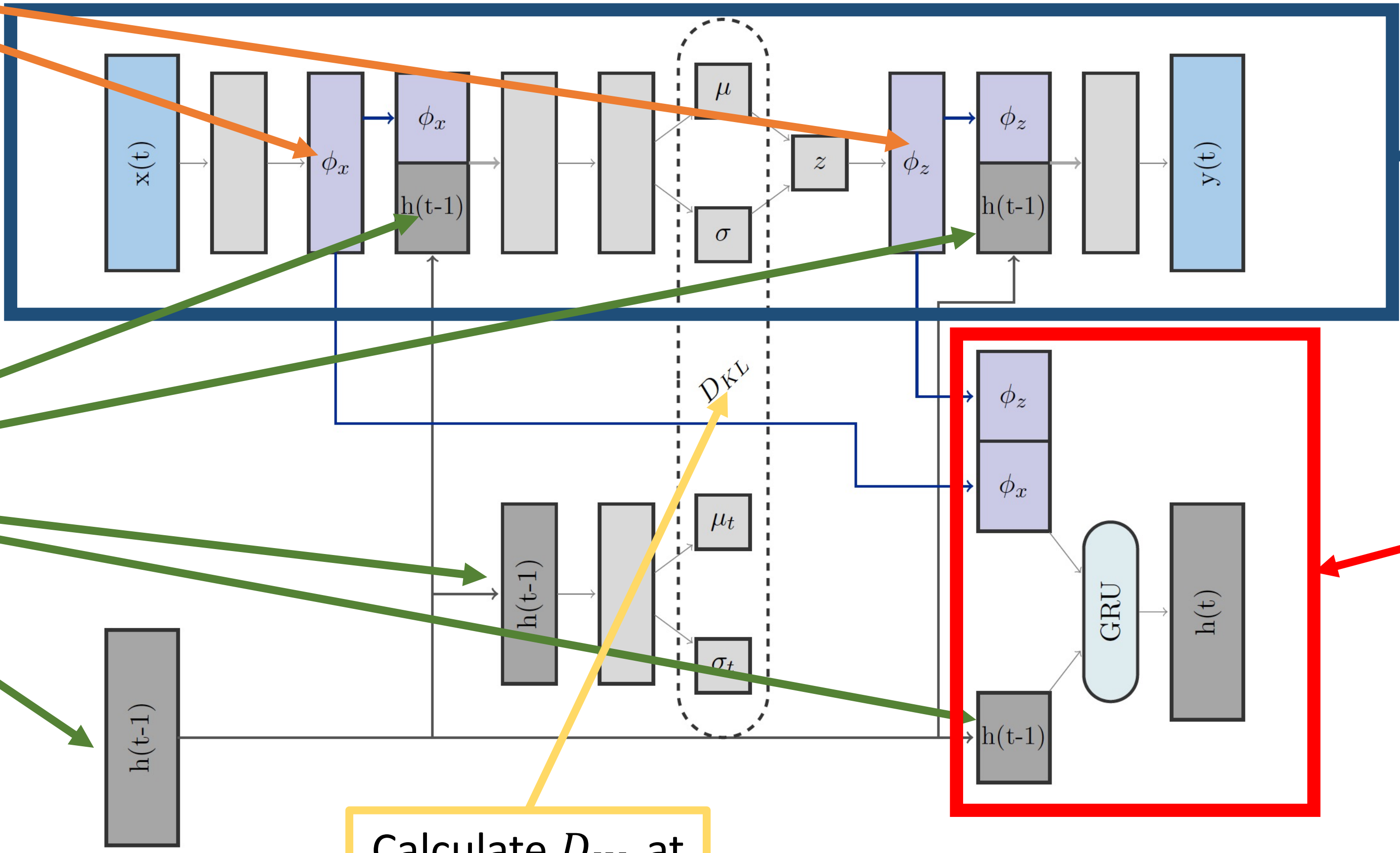
Variational Recurrent Neural Networks (VRNNs)



Variational Recurrent Neural Networks (VRNNs)

Feature extraction Layer ϕ for input and latent distribution

Variational Autoencoder



Recurrent NN

Combine features ϕ of this layer with previous features to obtain hidden state h that gets passed on to the next step!

Calculate D_{KL} at each step

Variational Recurrent Neural Networks (VRNNs)

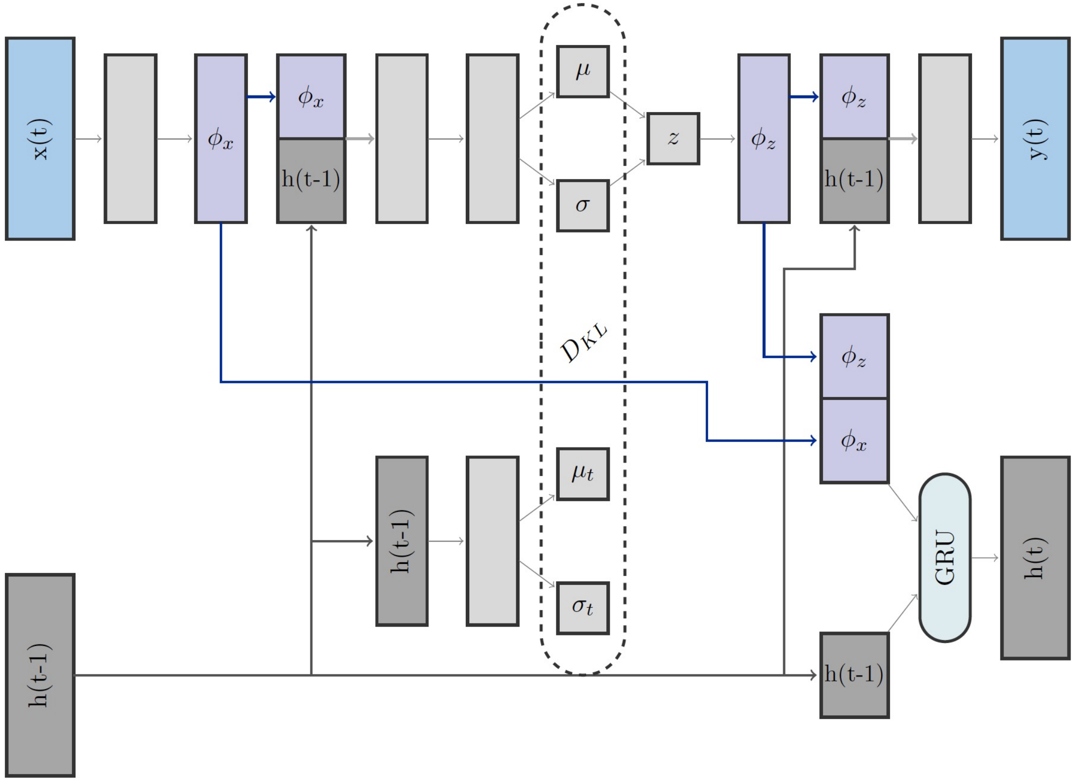
Input:

- Sequence of up to twenty constituent 4-vectors per jet
- Ordered by energy
- 4 high-level variables:
  1. D_2 -> Energy Correlation
  2. τ_{32} -> N-subjettines ratio
  3. d_{12}, d_{23} -> 3-prong sensitivity

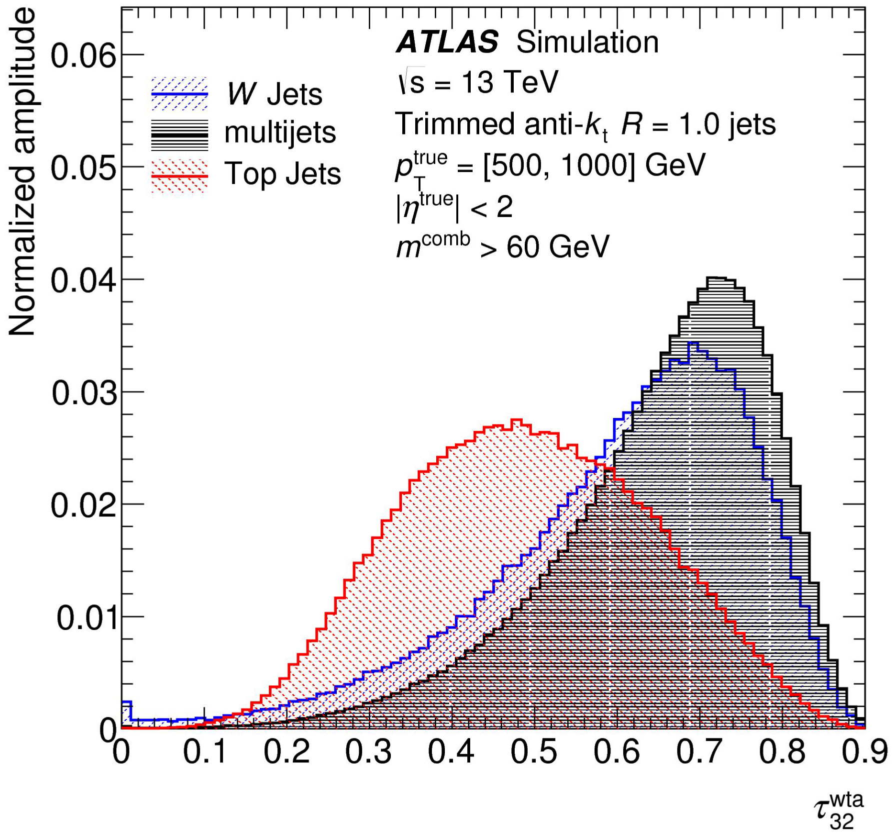
Output:

- Anomaly Score J:

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



N - Subjettiness Ratio



- N – Subjettiness: How compatible is this jet with a N-prong substructure
- Ratio τ_{ij}: preference of i over j
- Small ratio ⇒ high compatibility with i subjets

Variational Recurrent Neural Networks (VRNNs)

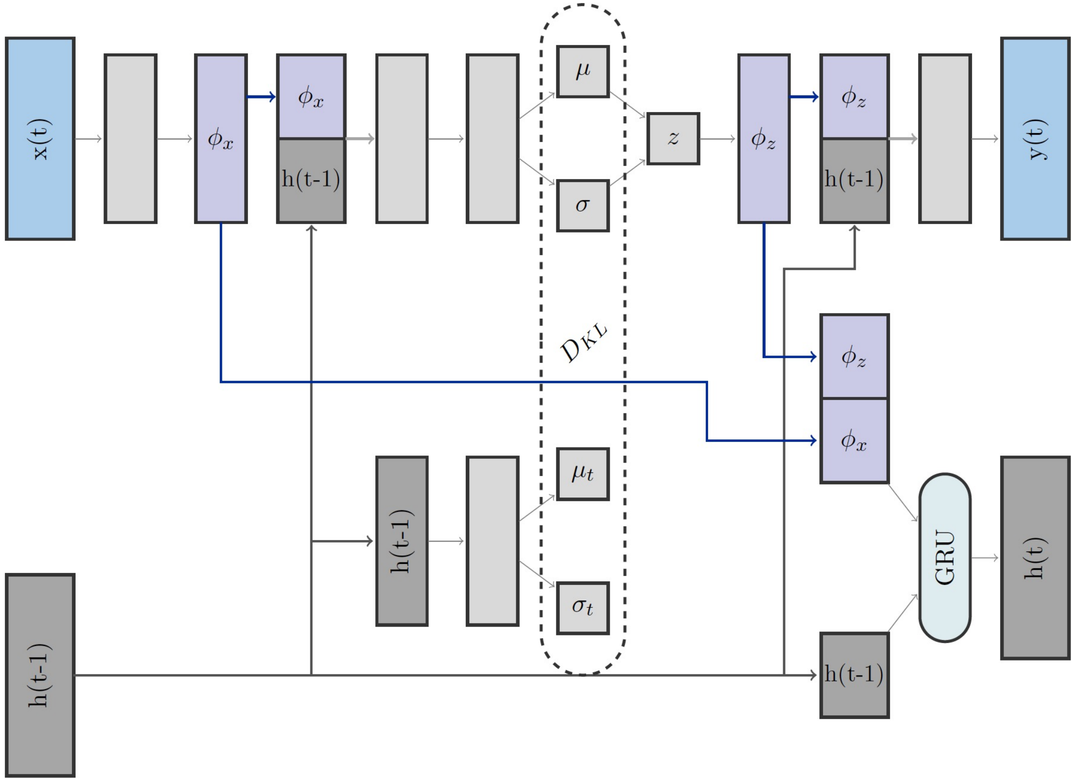
Input:

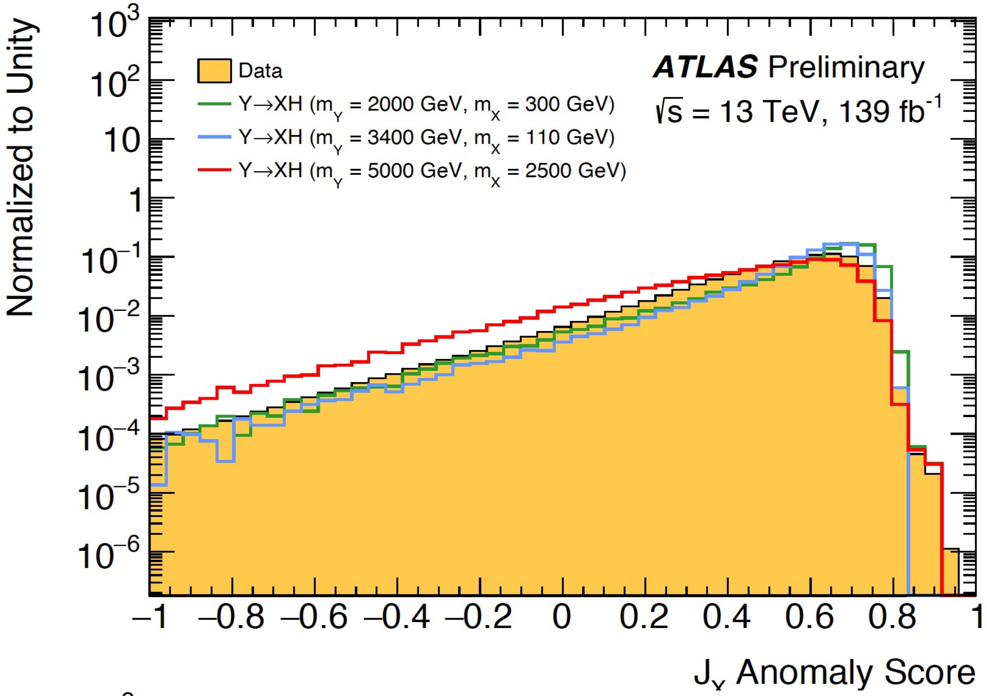
- Sequence of up to twenty constituent 4-vectors per jet
- Ordered by energy
- 4 high-level variables:
    1. D_2 -> Energy Correlation
    2. τ_{32} -> N-subjettines ratio
    3. d_{12}, d_{23} -> 3-prong sensitivity

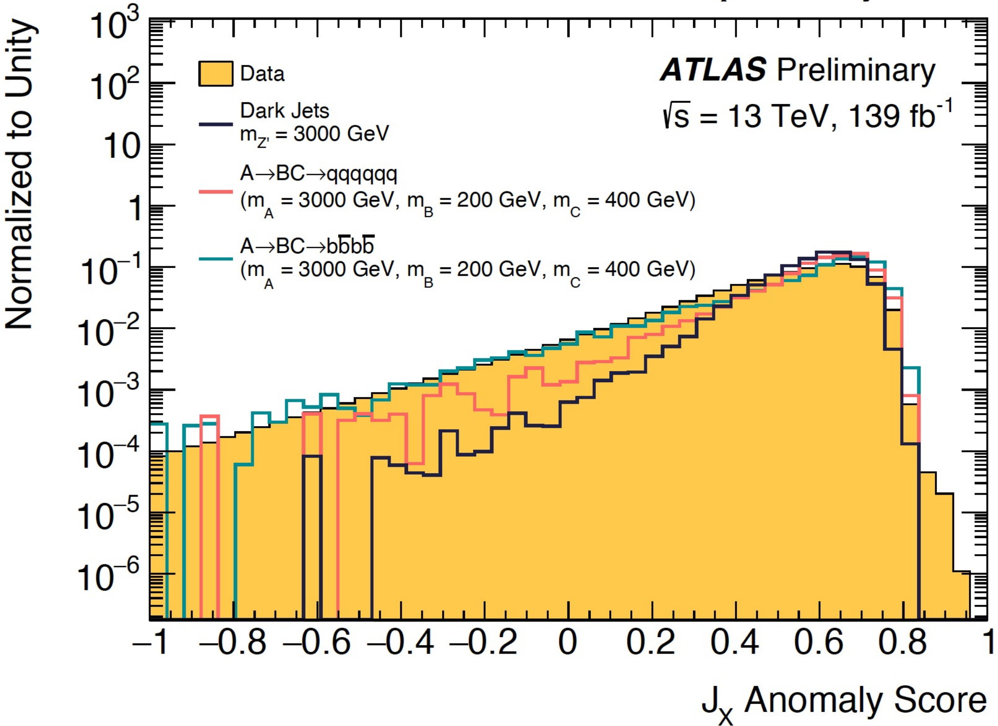
Output:

- Anomaly Score J:

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







Top:

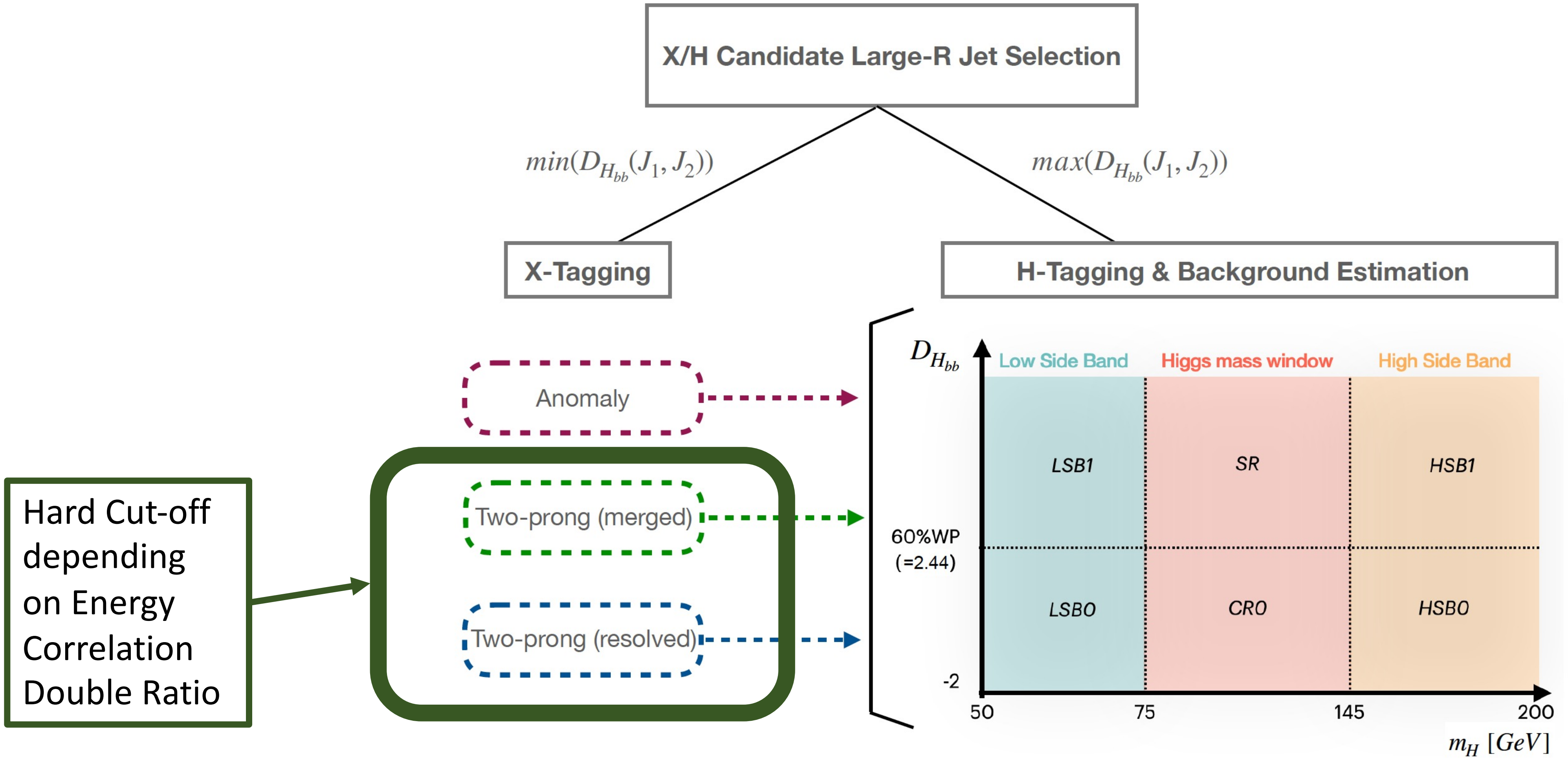
- Distribution of J for data and MC Simulation for different X and Y masses
- Especially **sensitive** for large mass differences
  ➔ **highly boosted regime!**
- Especially for the red line (not boosted) lot of points with small anomaly score
  ➔ Two prong region!

Bottom:

- Same distribution for alternative Jet signatures
- Even then the distribution peaks for high J

⇒ Highly Model Independent!

Analysis Regions



Two Prong Region

Resolved Two Prong Region:

- We want to cover all possible X masses
- What happens if $m_X \lesssim m_Y$?
- X decay products will no longer be boosted!
- Reconstruction as large-R-jet fails ➔ inaccurate results!
- Reconstruct constituents as 2 small-R-jets + add some extra filtering steps

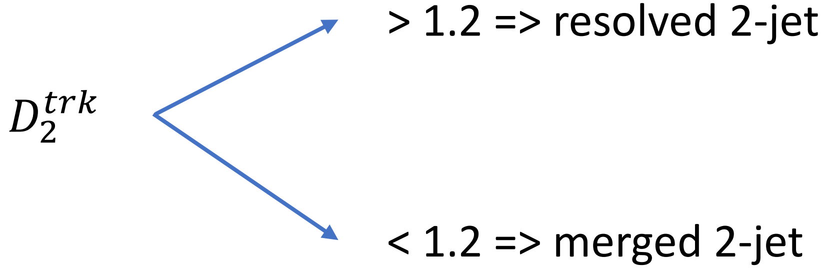
Merged Two Prong Region:

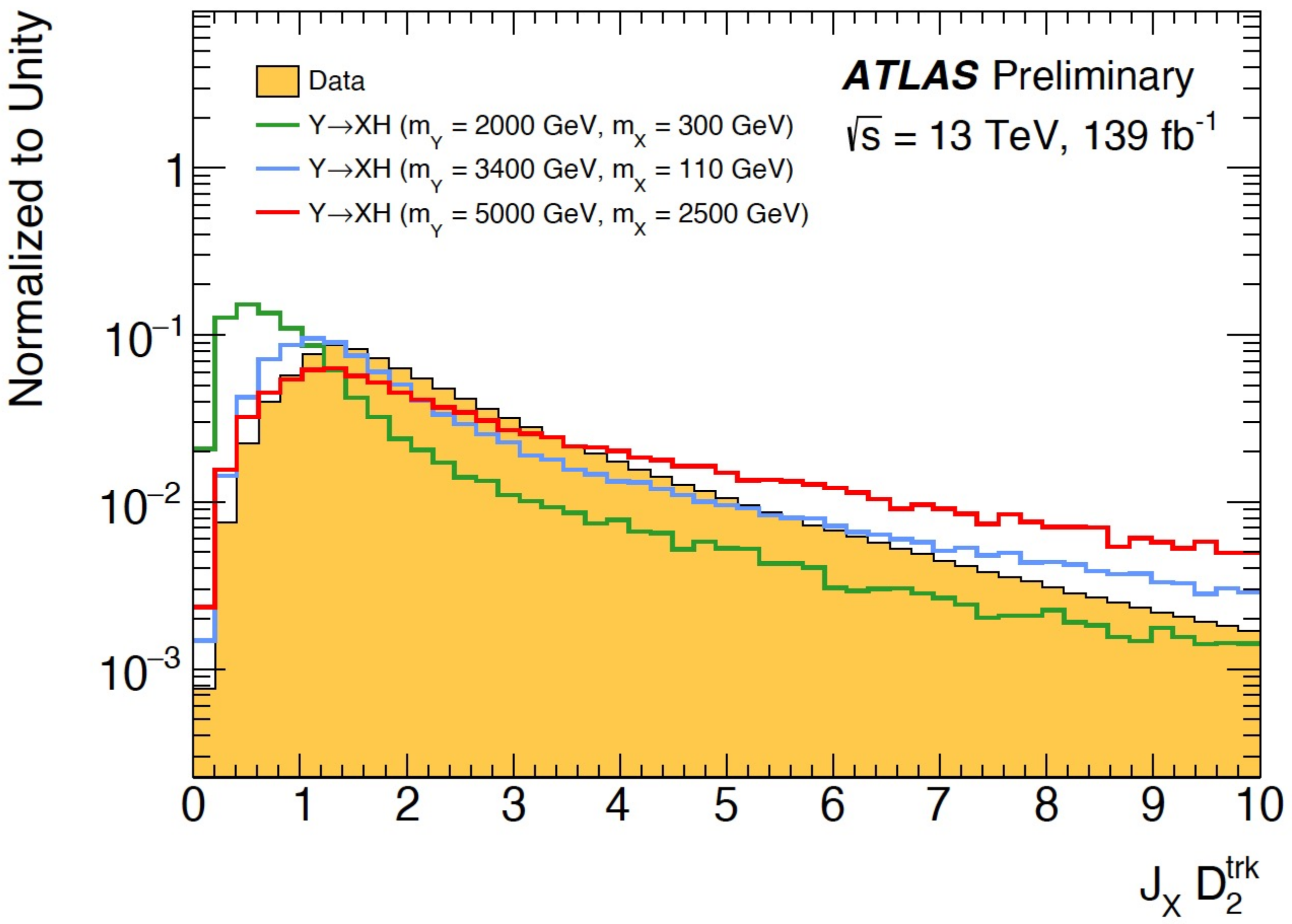
- Covers similar kinematic region as anomaly signal region
- **BUT** it is not model independent!
- Reason is to also test how well Anomaly detection performs compared to dedicated searches

Energy Correlation double ratio D_2

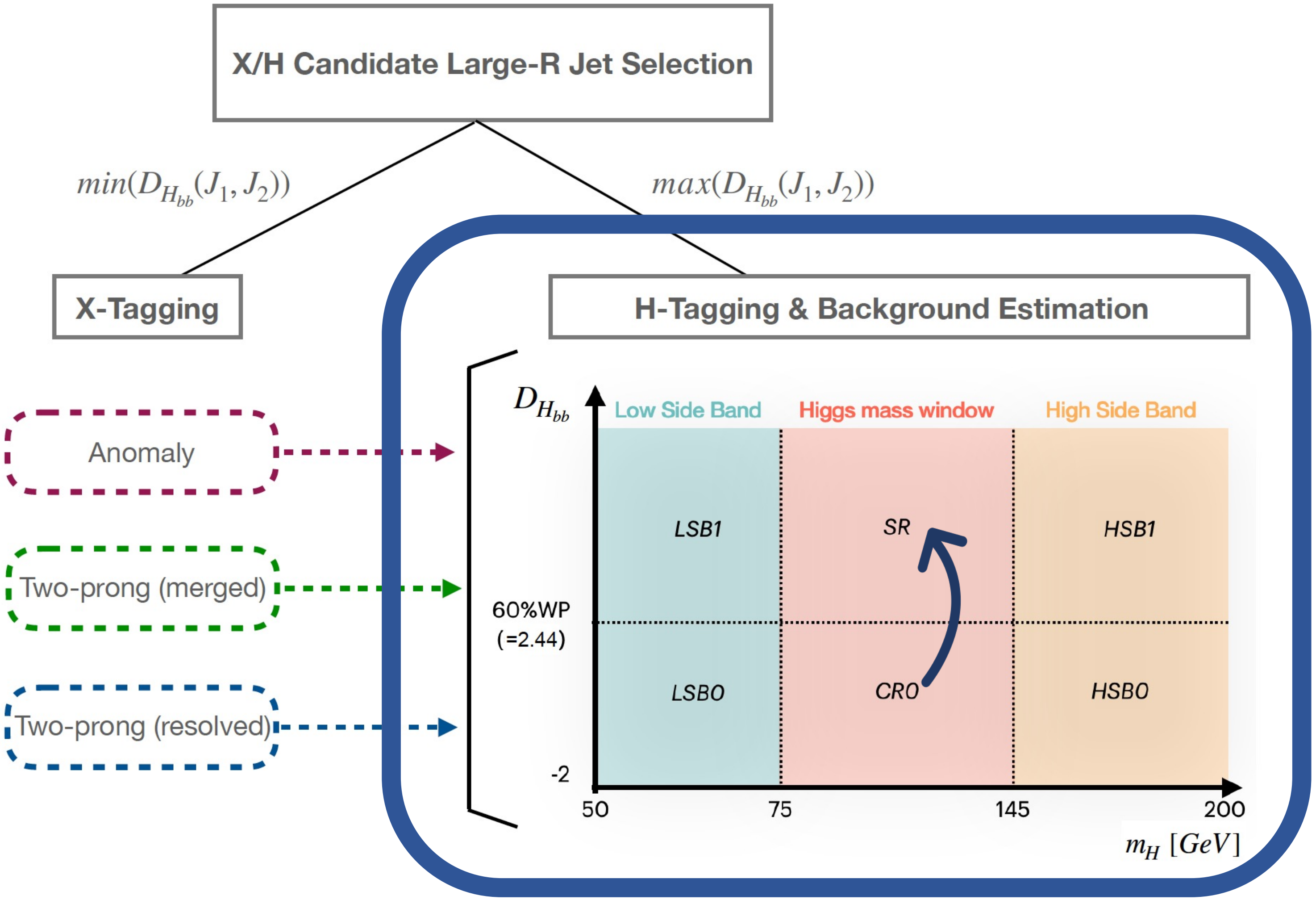
D_2:

- Similar to N-subjettiness
- 2 -> sensitive to two-prong substructure
- Large ⇒ two or more jets
- Small ⇒ less than two jets
- D_2^{trk}: only use jet constituents





Analysis Regions



3.

Predict with a DNN **from lower region** where we do not expect a signal the **background in the Signal Region!**

High Side Band: Training

Low Side Band: Validation

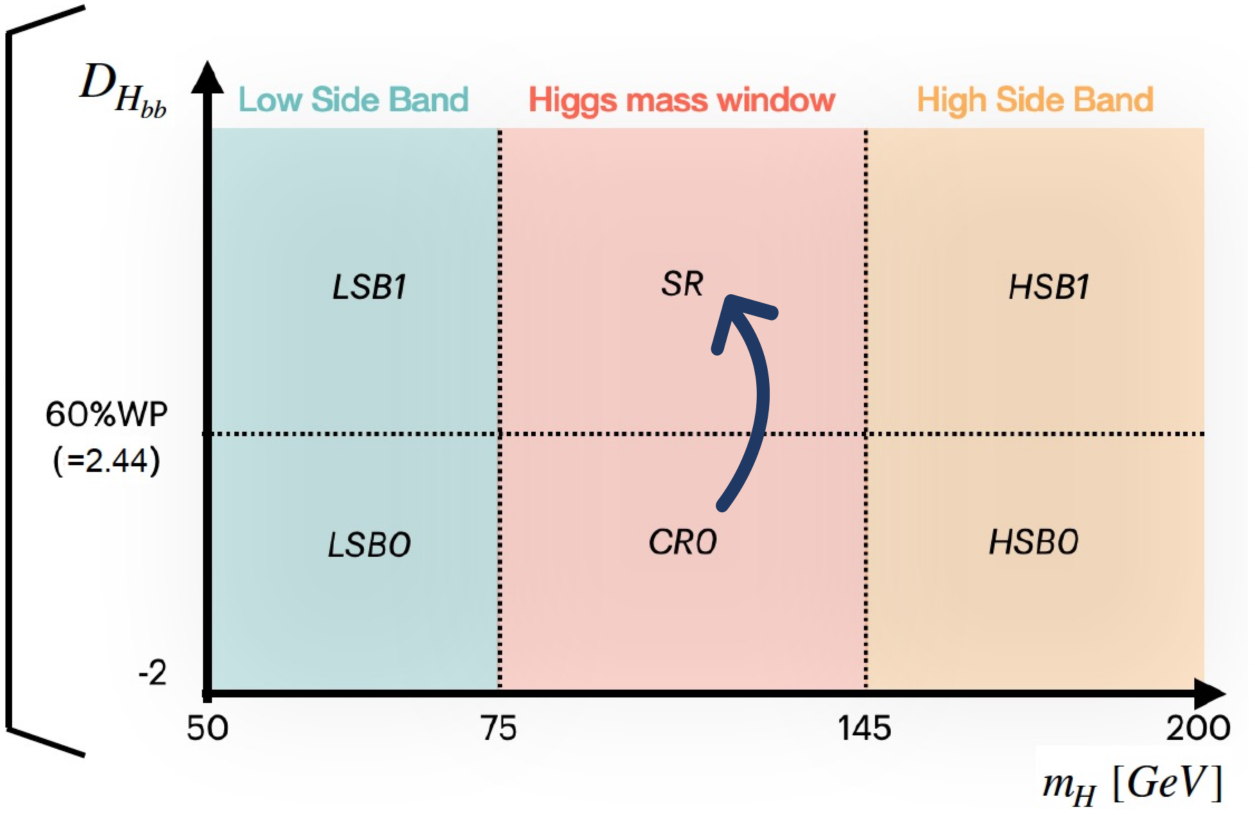
Background Estimation

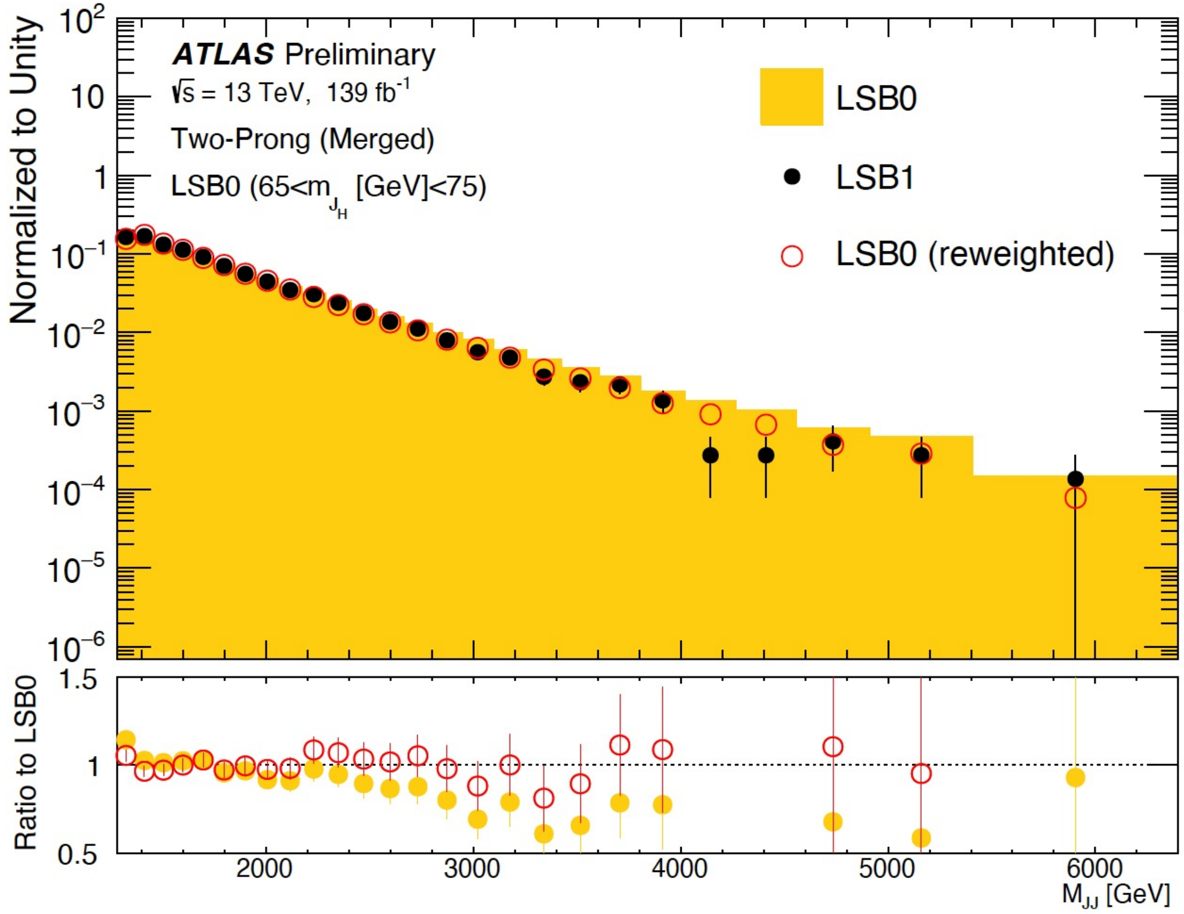
Idea:

- Divide Higgs into 3 mass windows: Low side Band (LSB), Higgs Mass Window (HMW), High Side Band (HSB)
- Split each mass window at $D_{H_{bb}} = 2.44$

⇒ 60% probability of being a Higgs!

- 0 ⇒ no Higgs, only background
- 1 ⇒ contains Higgs
- Using data from Control Region 0, DNN can predict the expected background in the signal region!
- Train with HSB data
- Validate with LSB data
- Normalization of the Background is allowed to float and is used as a fit parameter





Input:

- Unordered set of variables associated to each Jet
- Basically: **shape of the histogram**

Output:

- **Event-level weights** to obtain HSB1 PDF from HSB0 PDF
- Basically: In which bin would a similar event be in the HSB1 region
- Predict the histogram in HSB1

Key Assumption:

- Weights are independent of mass window

⇒ Validate in LSB region!

Systematic Uncertainties

Background:

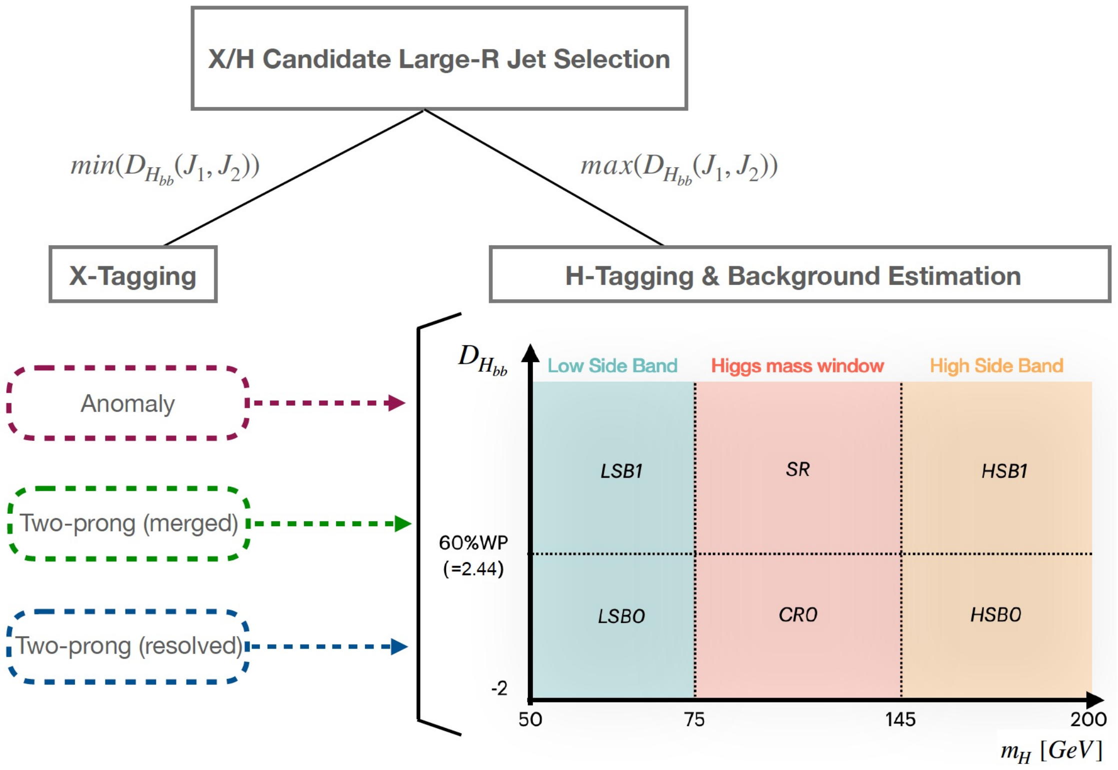
- Arbitrary training window $\sim\mathcal{O}(1-10\%)$
- Finite Statistics and Random Weight Initialization $\sim\mathcal{O}(1\%)$
- Approximation that weights are Mass independent
  - ➔ Take from LSB comparison of data and background
  - ➔ negligible for small m_{JJ}, $\sim\mathcal{O}(10\%)$ in m_{JJ} tail

Signal:

- Luminosity Uncertainty $\sim\mathcal{O}(1.7\%)$
- Theoretical Uncertainty in model $\sim\mathcal{O}(3\%)$
- Instrumental Systematics: jet scale and resolution uncertainty $\sim\mathcal{O}(8\%)$

Analysis Regions: Summary

Parameter	Preselection requirements				
m_{JJ} [GeV]	> 1300				
$p_T(J_1)$ [GeV]	> 500				
m_J [GeV]	$m_{J_1} > 50 \parallel m_{J_2} > 50$				
$D_{H_{bb}}$	> -2				
	Signal regions				
	Merged		Resolved		Anomaly
m_H [GeV]	(75, 145)				
$D_{H_{bb}}$	> 2.44				
D_2^{trk}	< 1.2		> 1.2		-
$\lvert\Delta y_{j_1,j_2}\rvert$	-		< 2.5		-
p_T^{bal}	-		< 0.8		-
Anomaly Score	-		-		> 0.5
	Background estimation regions				
	CR0	HSB0	HSB1	LSB0	LSB1
m_H [GeV]	(75, 145)	(145, 200)		(65, 75)	
$D_{H_{bb}}$	< 2.44	< 2.44	> 2.44	< 2.44	> 2.44



ANALYSIS AND RESULTS

Hypothesis Test

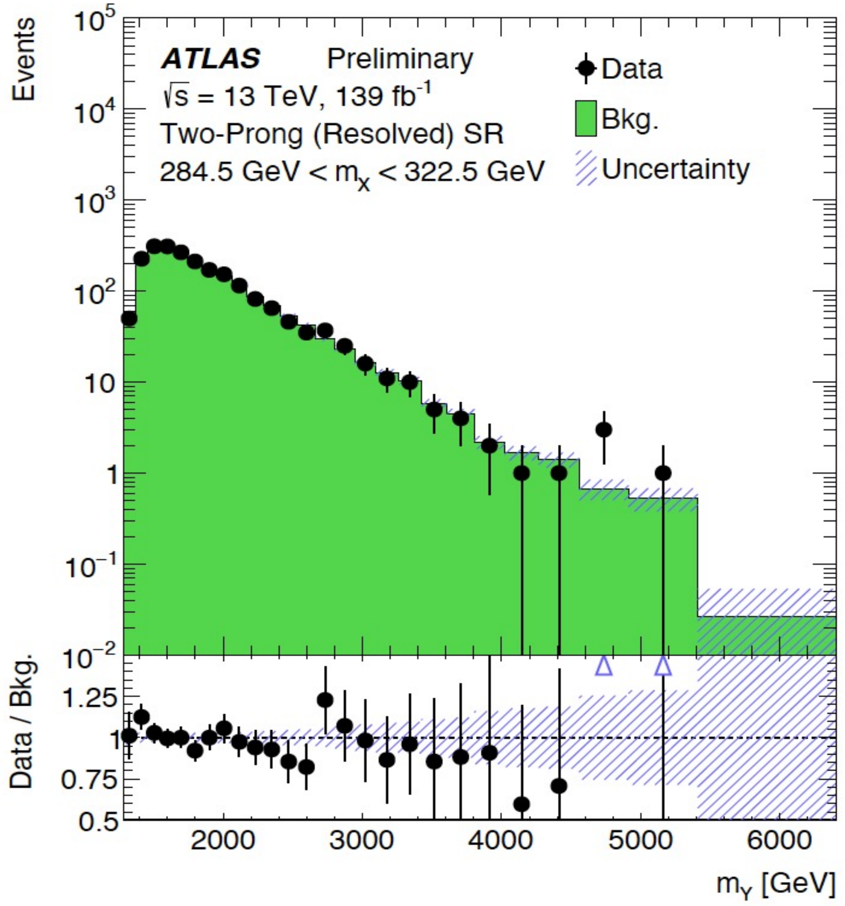
- Hypothesis Testing for bg-only an bg+signal hypotheses
- Observable to be fit: **m_{JJ}-distribution**
- Systematic Uncertainties incorporated as **nuisance parameters** in the fit
- X mass is not fixed, so where do we expect an excess?

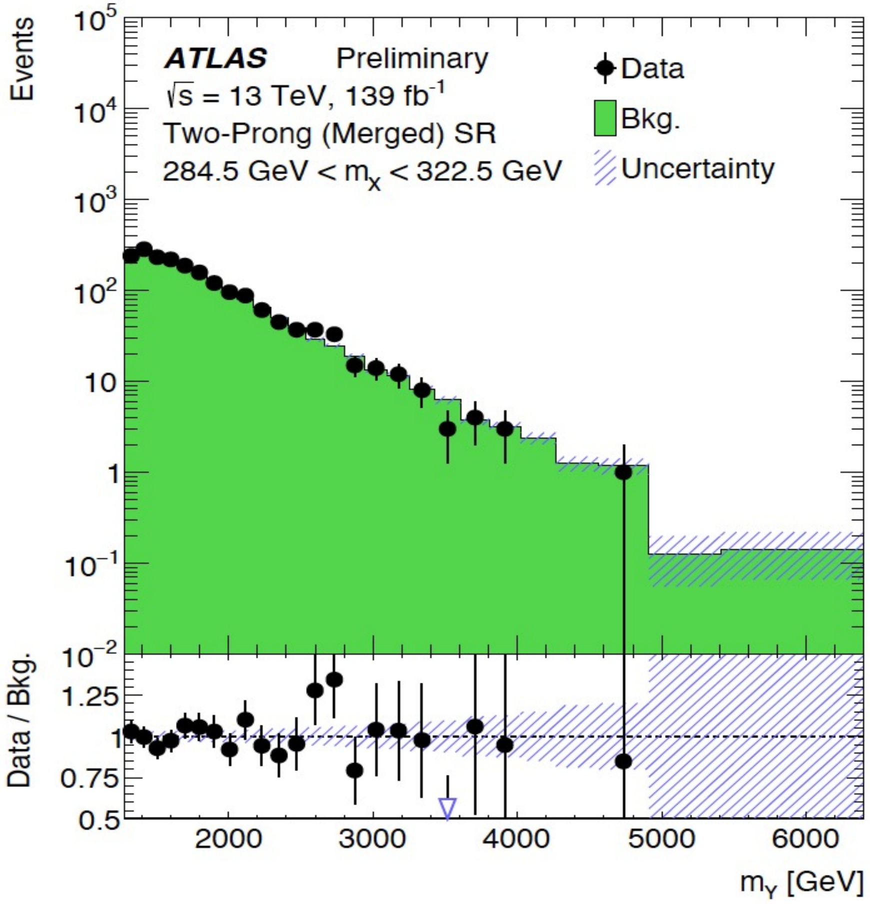
⇒ Analysis repeated in **overlapping m_X-bins** and for all **3 Signal Regions**

- Normalization of Background approximation is allowed to float

➔ only look at the shape!

Two Prong Signal Region





No Significant deviation again!

Anomaly Signal Region

BumpHunter:

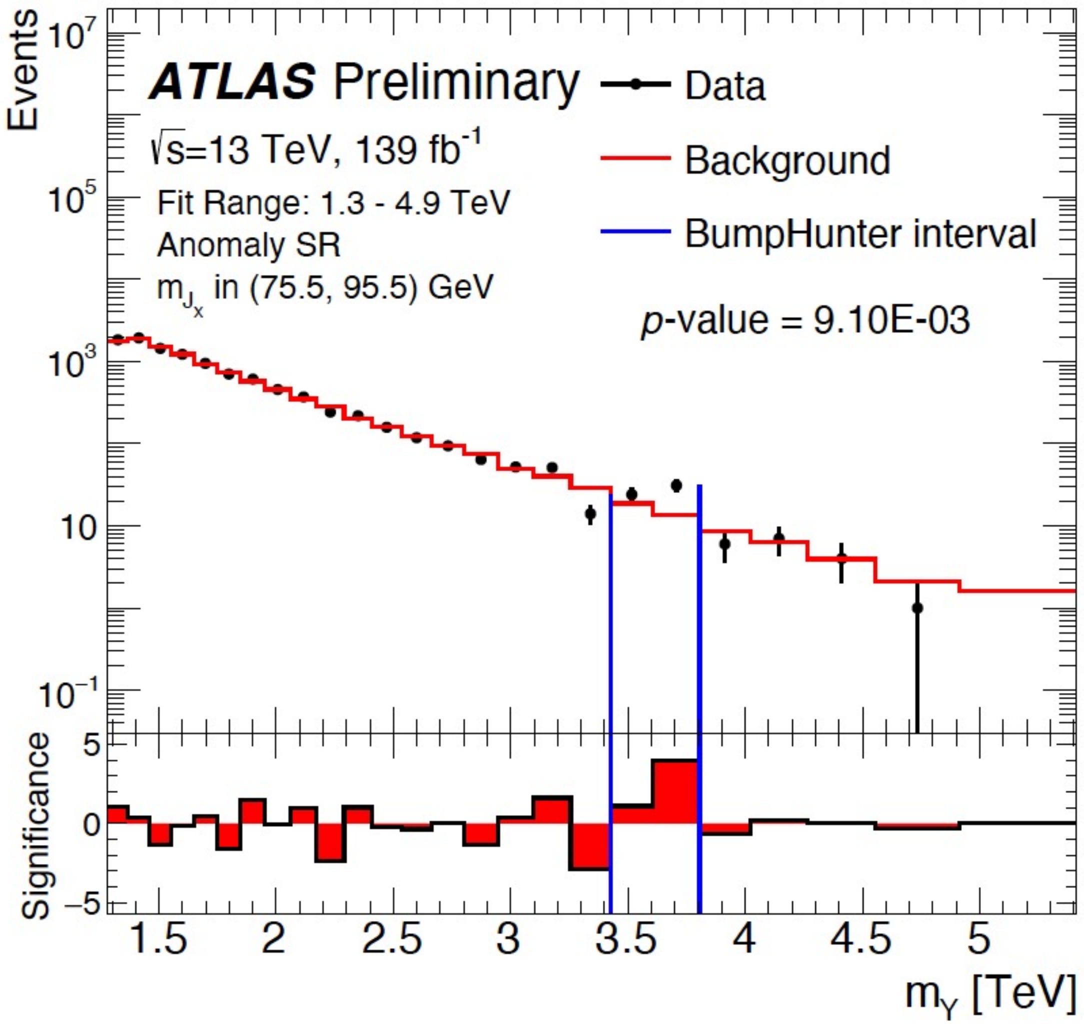
- Hypothesis Hypertest

Returns:

  - most significant bump in data
  - m_X mass window
  - m_Y distribution + fit
  - local p-value taking into account the "trials factor"

Results:

- Background shows good fit with data
- Largest excess: m_X in $[75.5, 95.5]$ GeV
- **Local** p-value = $9.1 \cdot 10^{-3}$
- Corresponds to global significance of $\mathbf{1.47\sigma}$
- But: **Substructure Incompatible with signal**



Constraints

- Signal + Background fit to data

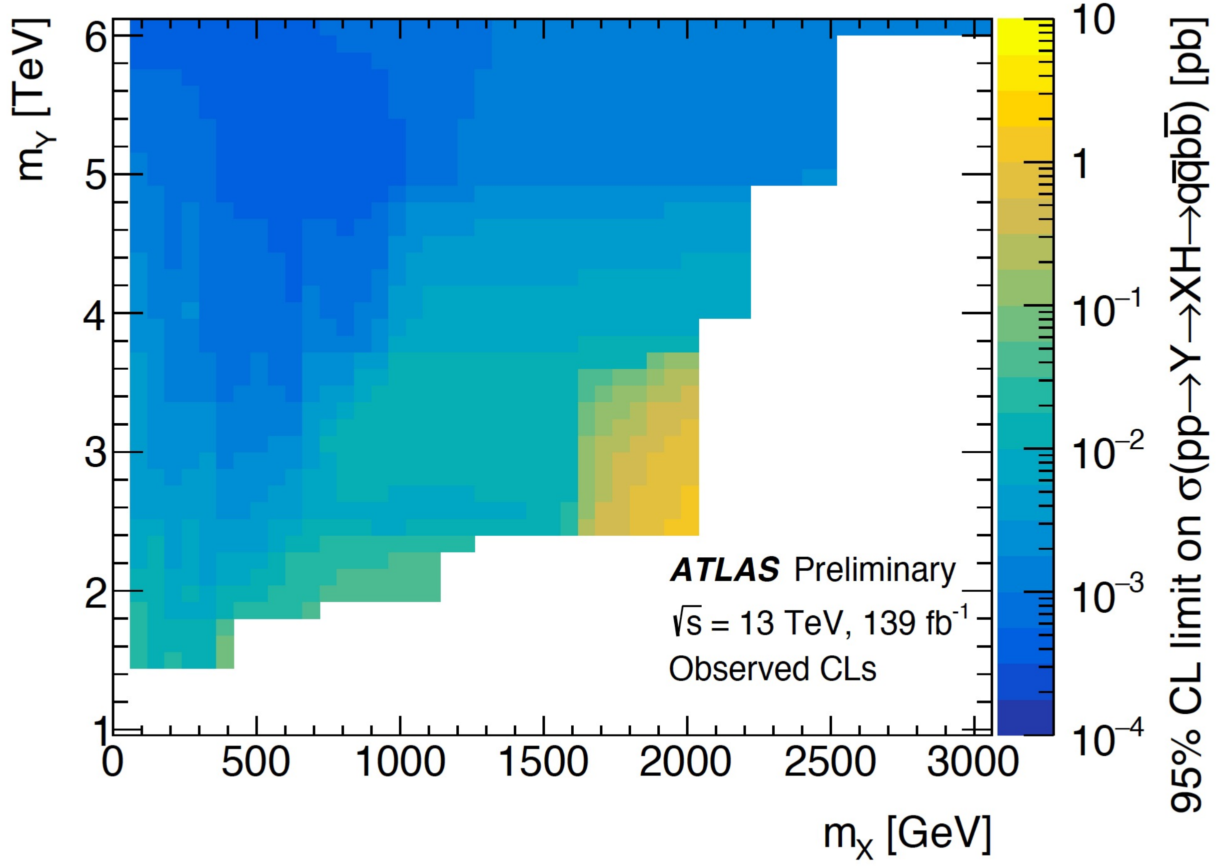
➔ Only two Prong, Anomaly region is supposed to be model independent!

- Assume Heavy Vector Triplet Model
- Find **95% Confidence Level Limits on $\boldsymbol{\sigma}$**

➔ If σ had been higher or equal than this, we should have seen it with 95% likelihood

- **Most stringent** in highly boosted regime for $m_Y = 5$ TeV and $m_X = 600$ GeV:

$$\boldsymbol{\sigma = 0.342}\ \textbf{fb}$$



Sensitivity of Anomaly Detection

- Assess Sensitivity of **model Independent** Anomaly SR to **dedicated search** in two prong region (2PR)
- Compare constraints obtained from signal + background fit for all signals (including alternative signals!)

Standard Signal:

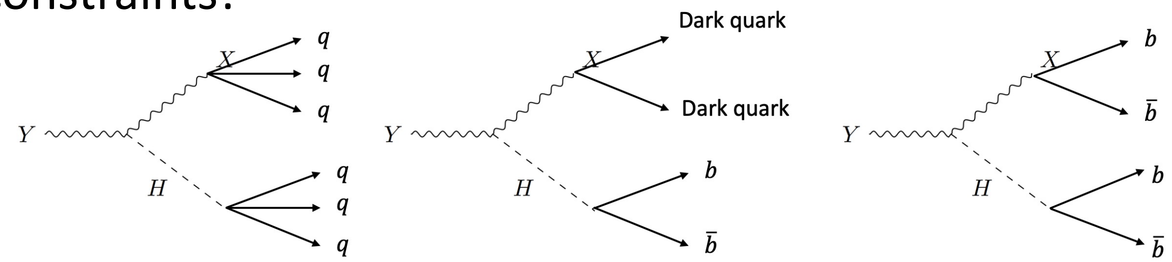
- Anomaly Region is sensitive to highly boosted regime

➔ The upper limit is similar as for Merged 2PR

- Combined merged + resolved 2PR is more sensitive in rest of Parameter Space

Alternative Topologies:

- Anomaly Detection Significantly improves the constraints!
- **20x improvement for Dark Jets!**



Summary

- Search for heavy Y decaying to new particle X and SM Higgs with hadronic final states reconstructed as boosted large R jets
- Anomaly detection with a Variational Recurrent Neural Network

→ **1st application of fully unsupervised ML to ATLAS search!**

- DNNs also used for $H \to b\bar{b}$ tagging and data driven background estimation
- No significant deviations found → Largest excess in anomaly SR of 1.47σ
- But: **Substructure Incompatible with Signal**
- Most stringent in highly boosted regime for $m_Y = 5$ TeV and $m_X = 600$ GeV:

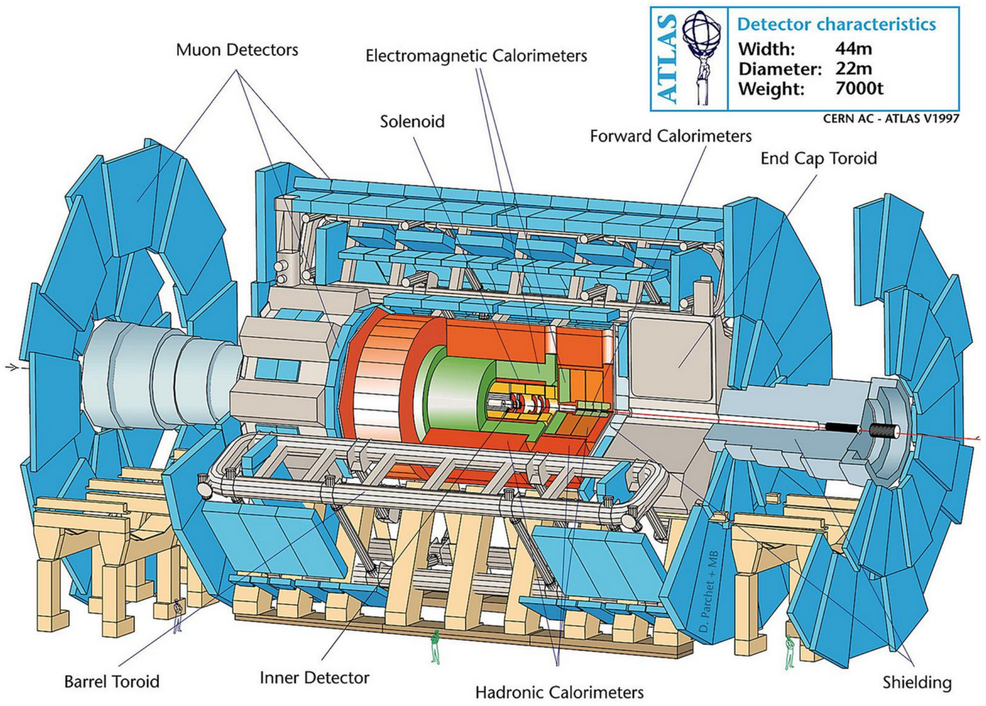
$$\sigma = 0.342 \text{ fb}$$

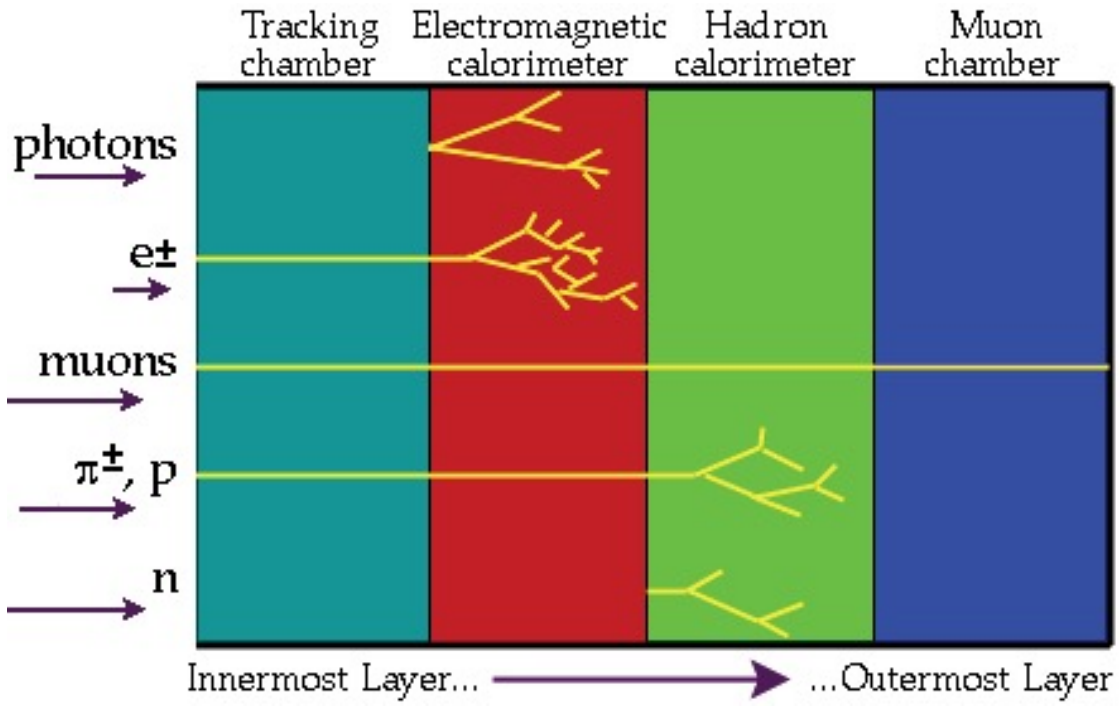
- Dedicated search is more sensitive for the exact signal, **Anomaly Detection outperforms** it for all other alternative signatures!

Thank you for your attention!
Questions?

Backup Slides

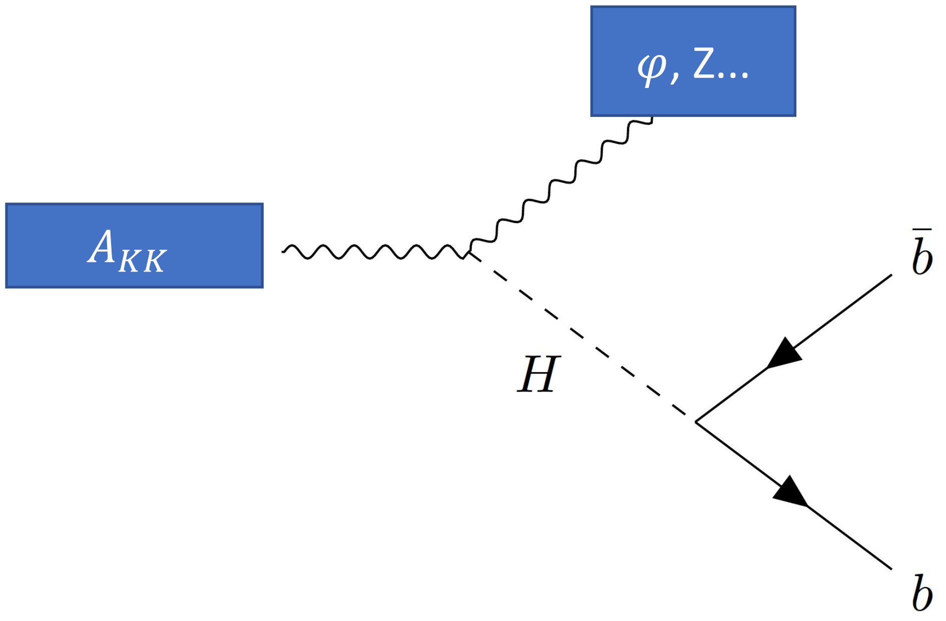
The ATLAS Detector



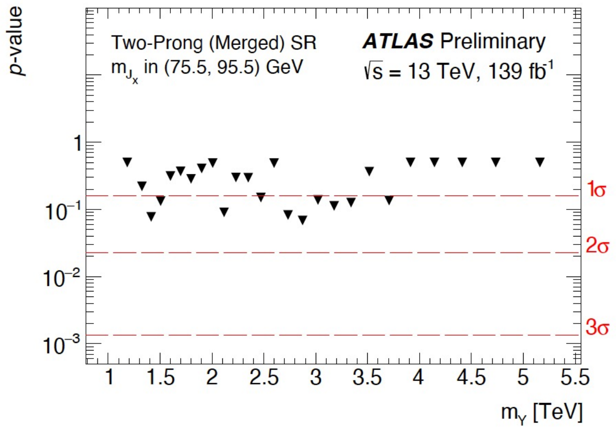


Universal Extra Dimensions

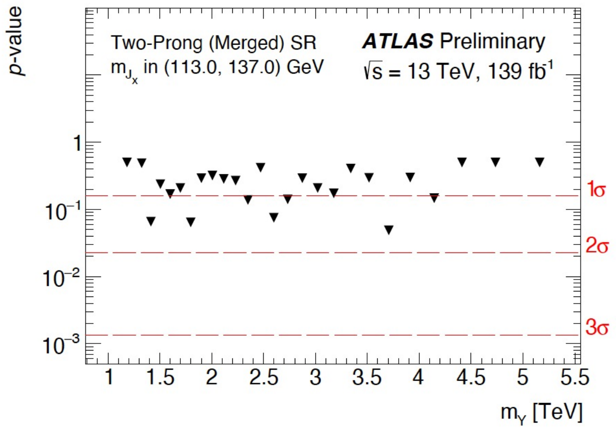
- (3+n+1)-D bulk, (3+1)-D brane (us)
- Compact extra dimensions $\sim\mathcal{O}(R)$:
  - ⇒ p quantized $p^2 \sim {}^{1}/_{R^2}$
  - ⇒ In brane we see this as a tower of states with masses $m_n \sim {}^{n}/_{R}$
  - ⇒ Many new particles!
  - ⇒ Many possibilities for such a decay!
- Lightest KK Particle -> Dark matter?



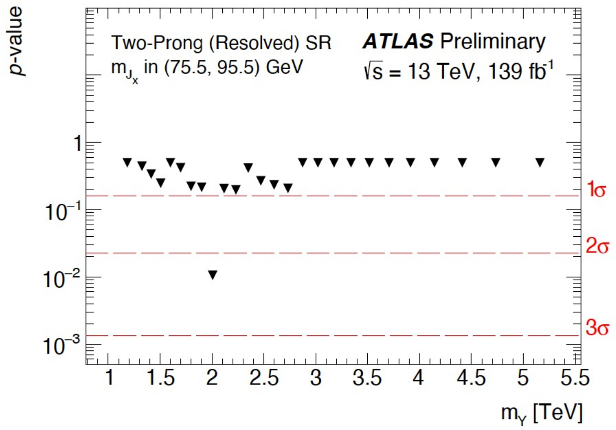
Bump Hunter for Two Prong region



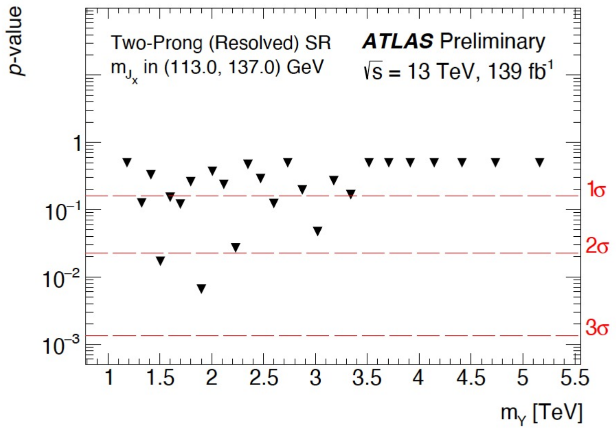
(a)



(b)



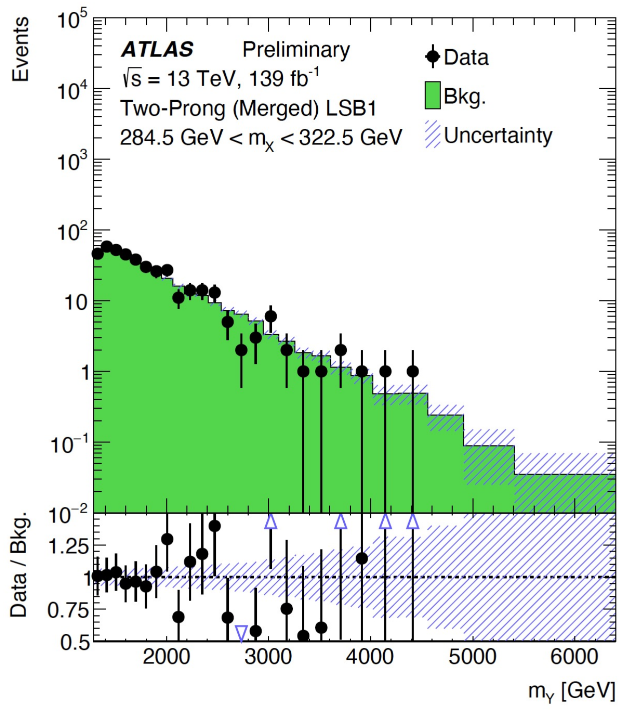
(c)

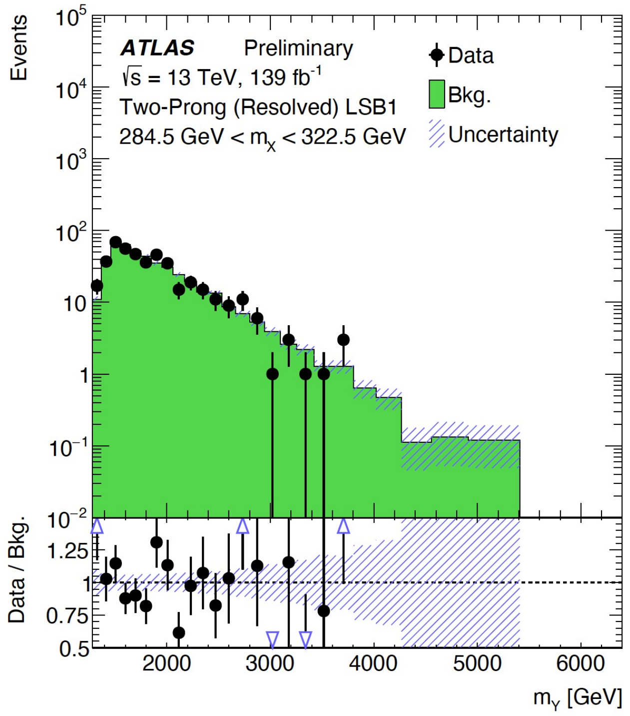


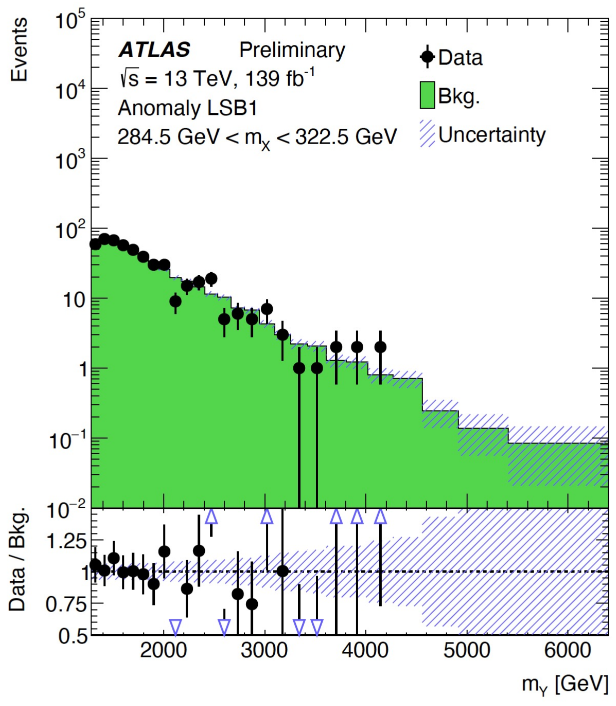
(d)

Sanity Check

In Low Side Band region we do not expect a signal







No Significant deviation! We can start with the results!