Searches for new physics using unsupervised machine learning

B. Wynne
On behalf of the ATLAS collaboration
University of Edinburgh

31/07/24

Introduction

The last ~decade of BSM physics searches at ATLAS have mostly been motivated by specific theoretical models, using analysis selections optimised for a corresponding final state

Lack of discoveries for popular theories, and a broad choice of models, lead to desire for model independence

e.g. Yoran and Jingjing's talks

- Reinterpretation of existing results
- Presentation of generic limits in terms of cross-section x efficiency
- Simplified, generic models for summary plots

Inclusive searches for generic final states (e.g. di-jets) suffer from high backgrounds

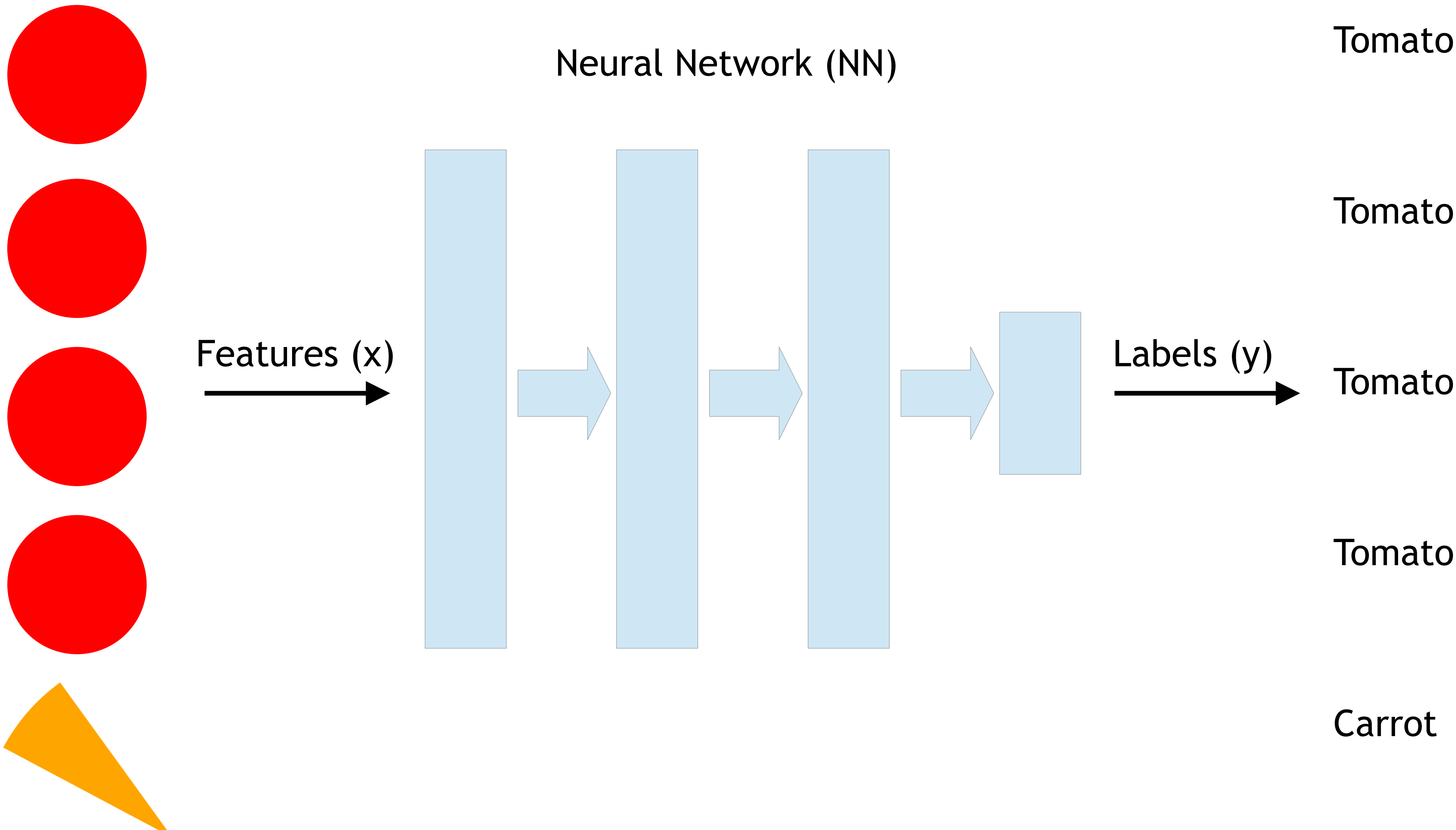
Anomaly Detection (AD) is an alternative approach, that aims to make analysis selections sensitive to a variety of models while still suppressing background

Typically using Machine Learning (ML) methods, events are assessed for their similarity to expected events

Greater model-independence still comes from unsupervised techniques, where the ML system is trained on data rather than simulation

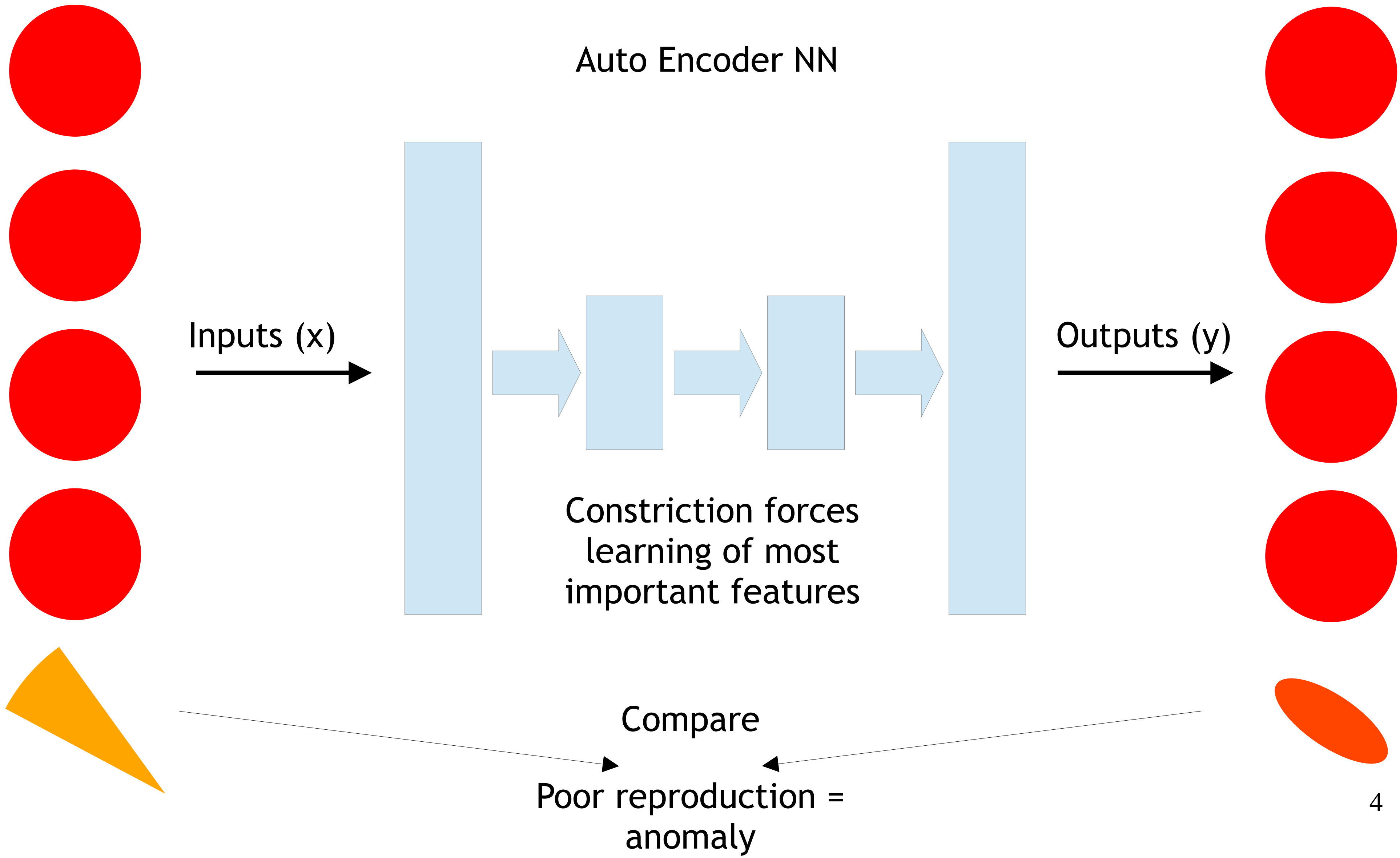
Supervised classifier – “classic ML”

All events have a set of features (x) and a label (y). The model is trained to produce the label y most appropriate for a given x



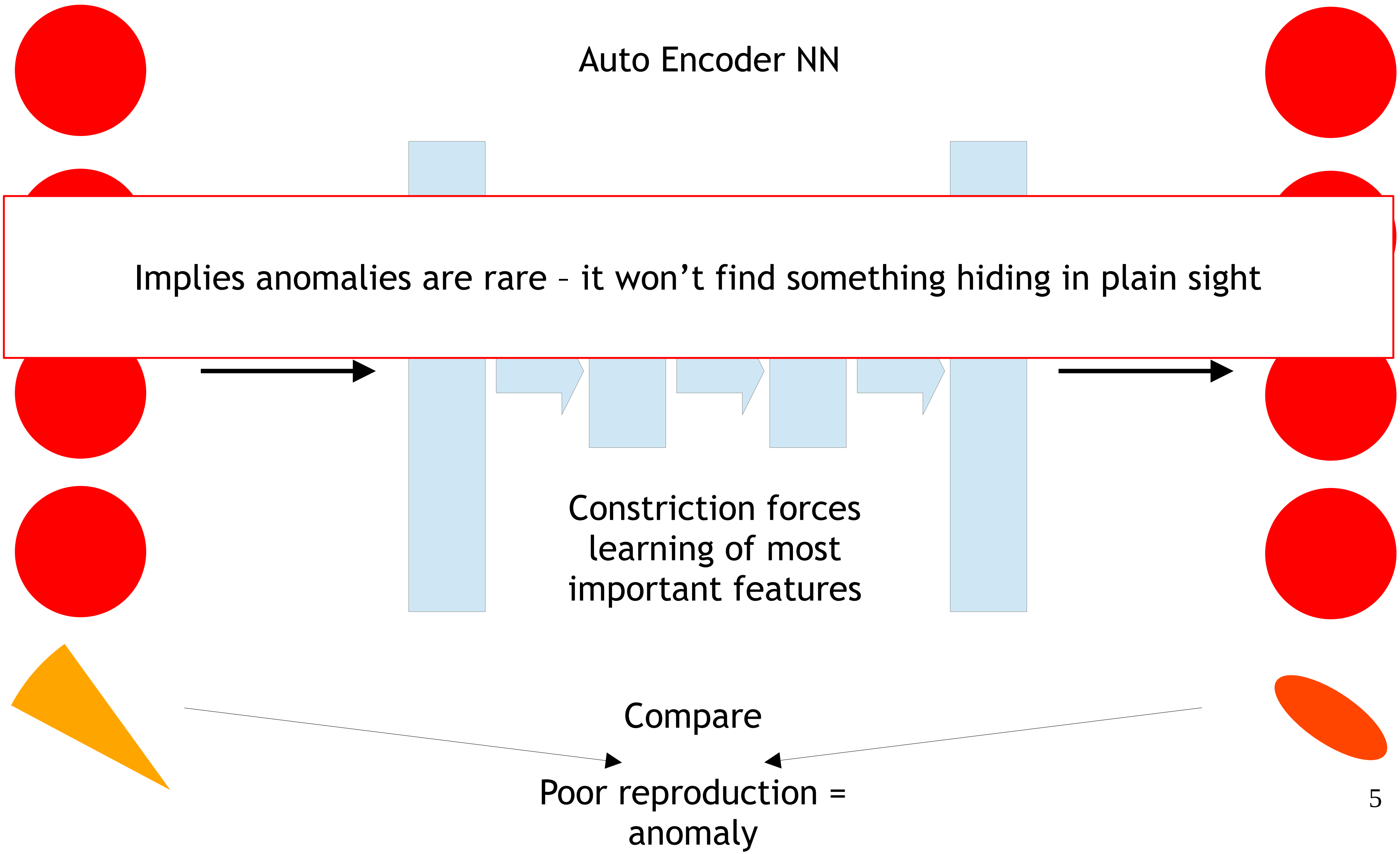
Unsupervised anomaly detection

The NN is simply trained to reproduce its input. Unfamiliar inputs will be reproduced less well, so comparing outputs to inputs gives anomaly score



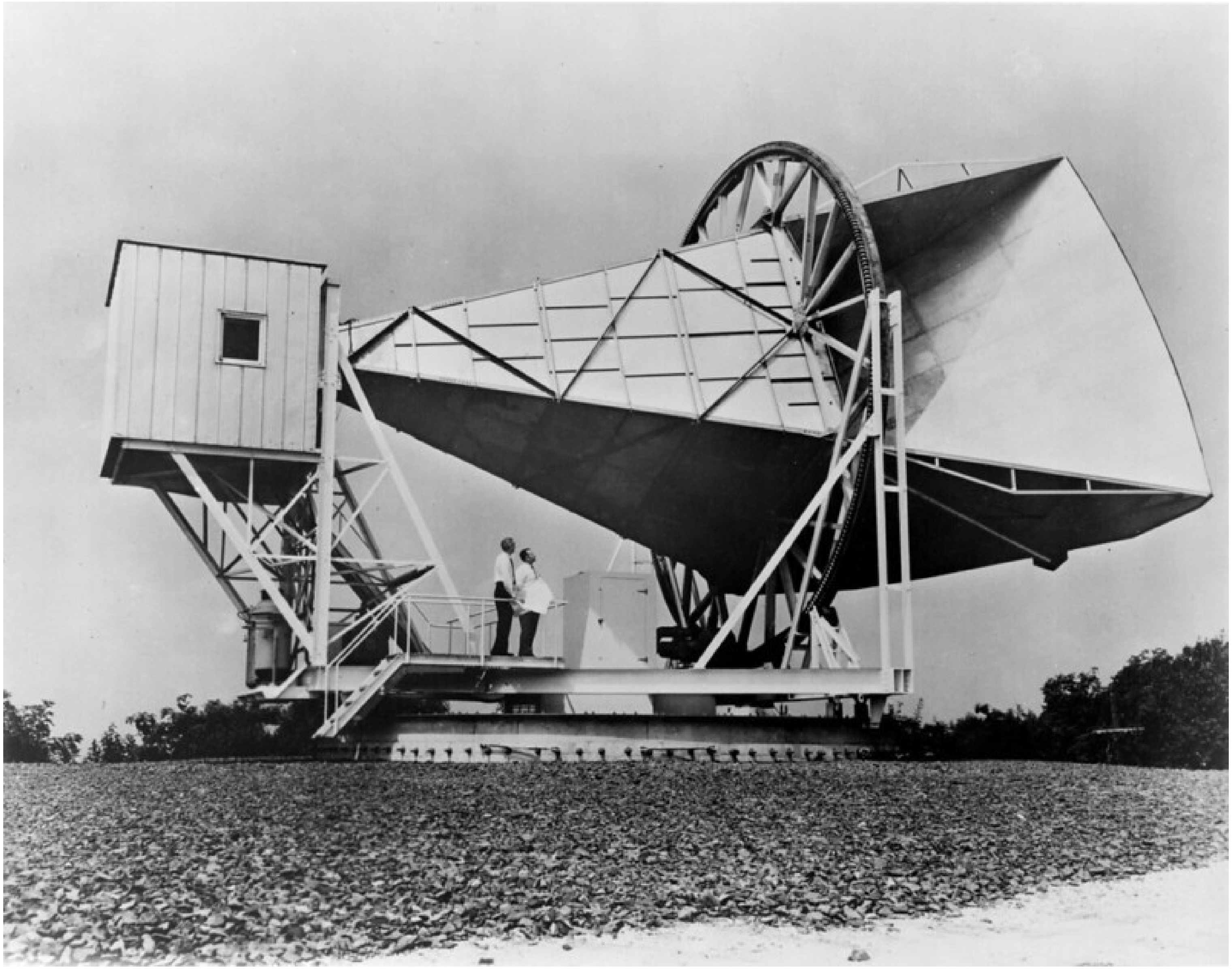
Unsupervised anomaly detection

The NN is simply trained to reproduce its input. Unfamiliar inputs will be reproduced less well, so comparing outputs to inputs gives anomaly score



Anomaly detection

Some of the most important results in physics have come from noticing anomalies...



Anomalous event tagging

Encoding features

Particle physics events aren't naturally amenable to ML input formats, since we typically have "awkward arrays" – multidimensional data of variable lengths – rather than fixed-size flat arrays

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

$$
\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j_1}) & m(j_1, j_2) & \ldots m(j_1, j_N) & m(j_1, \mu_1) & m(j_1, \mu_2) & \ldots m(j_1, \mu_N) \\
h_L(j_2) & h(j_1, j_2) & \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_2}) & \ldots m(j_2, j_N) & m(j_2, \mu_1) & m(j_2, \mu_2) & \ldots m(j_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1, j_N) & \ldots & \ldots \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_N}) & m(j_N, \mu_1) & m(j_N, \mu_2) & \ldots m(j_N, \mu_N) \\
h_L(\mu_1) & h(\mu_1, j_1) & h(\mu_1, j_2) & \ldots h(\mu_1, j_N) & \mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{1}}) & m(\mu_1, \mu_2) & m(\mu_1, \mu_N) \\
h_L(\mu_2) & h(\mu_2, j_1) & h(\mu_1, j_2) & \ldots h(\mu_2, j_N) & h(\mu_1, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{2}}) & m(\mu_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N, j_1) & h(\mu_N, j_2) & \ldots h(\mu_N, j_N) & h(\mu_N, \mu_1) & h(\mu_N, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{N}})
\end{pmatrix}
$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

Choose particle flavours to include (e.g. jets, muons) and max multiplicity N

Jets 1→N (e_T ordered)

$$
\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j}_1) & m(j_1, j_2) & \ldots m(j_1, j_N) & m(j_1, \mu_1) & m(j_1, \mu_2) & \ldots m(j_1, \mu_N) \\
h_L(j_2) & h(j_1, j_2) & \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j}_2) & \ldots m(j_2, j_N) & m(j_2, \mu_1) & m(j_2, \mu_2) & \ldots m(j_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1, j_N) & \ldots & \ldots \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j}_N) & m(j_N, \mu_1) & m(j_N, \mu_2) & \ldots m(j_N, \mu_N) \\
h_L(\mu_1) & h(\mu_1, j_1) & h(\mu_1, j_2) & \ldots h(\mu_1, j_N) & \mathbf{e}_{\mathbf{T}}(\mu_1) & m(\mu_1, \mu_2) & m(\mu_1, \mu_N) \\
h_L(\mu_2) & h(\mu_2, j_1) & h(\mu_1, j_2) & \ldots h(\mu_2, j_N) & h(\mu_1, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_2) & m(\mu_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N, j_1) & h(\mu_N, j_2) & \ldots h(\mu_N, j_N) & h(\mu_N, \mu_1) & h(\mu_N, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_N)
\end{pmatrix}
$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

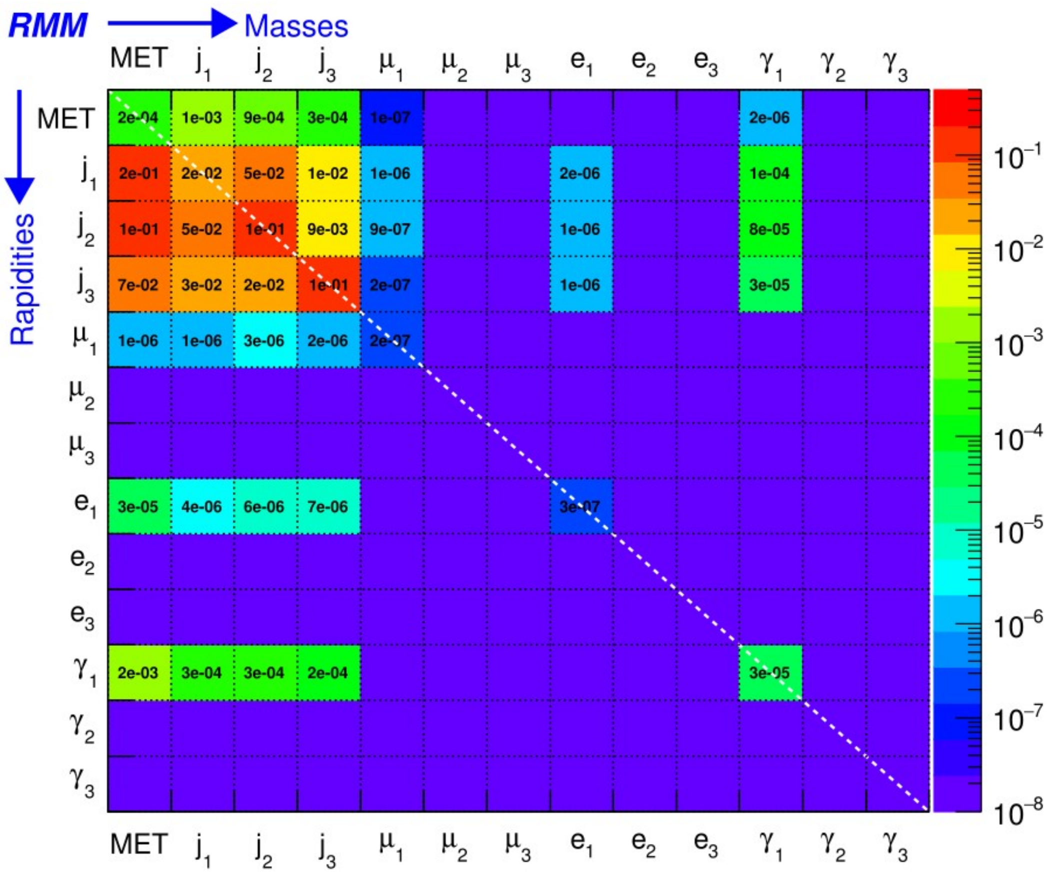
- Elements above the diagonal are masses or mass combinations
- Elements below the diagonal are rapidities or rapidity differences
- Details in backup

$$
\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j}_1) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j}_2) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j}_{\mathbf{N}}) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & \mathbf{e}_{\mathbf{T}}(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_2) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{N}})
\end{pmatrix}
$$

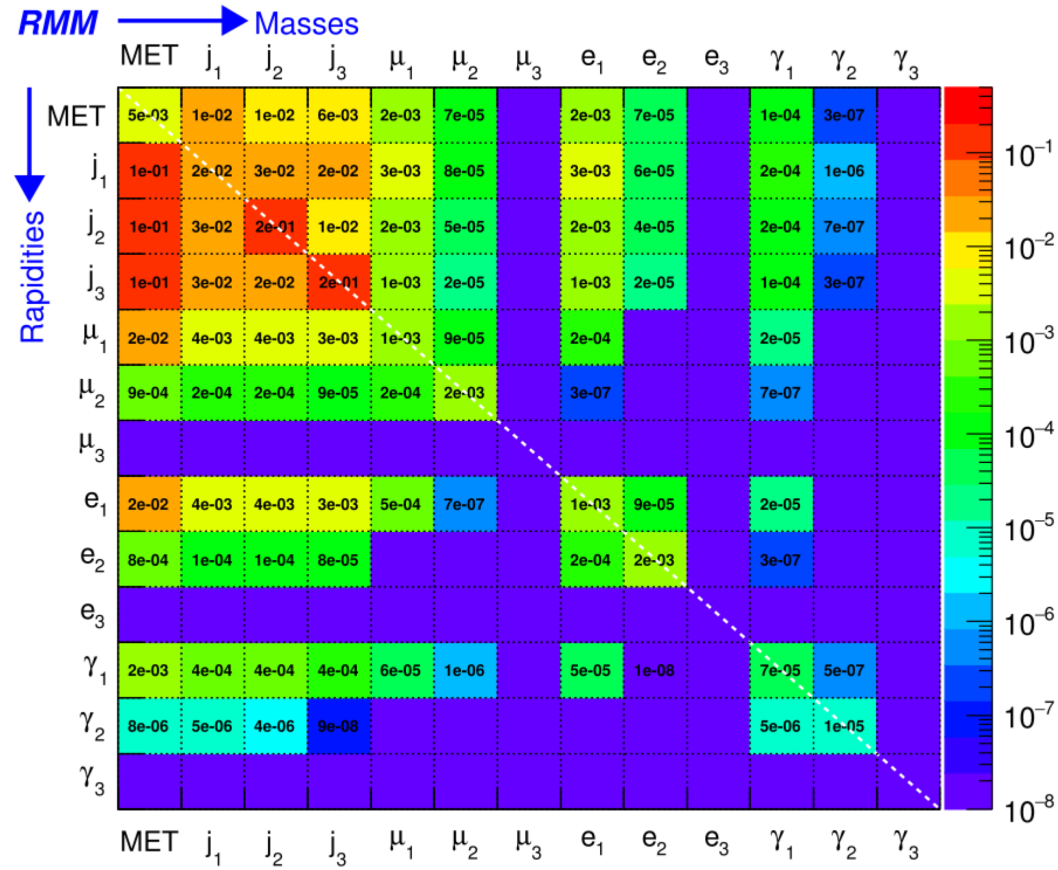
Nucl.Instrum.Meth.A 931 (2019) 92-99

RMM signatures

Different processes can give rise to distinctive patterns in the RMM, although these are strongly linked to particle multiplicity and corresponding sparsity of the matrix



Multi-jets QCD



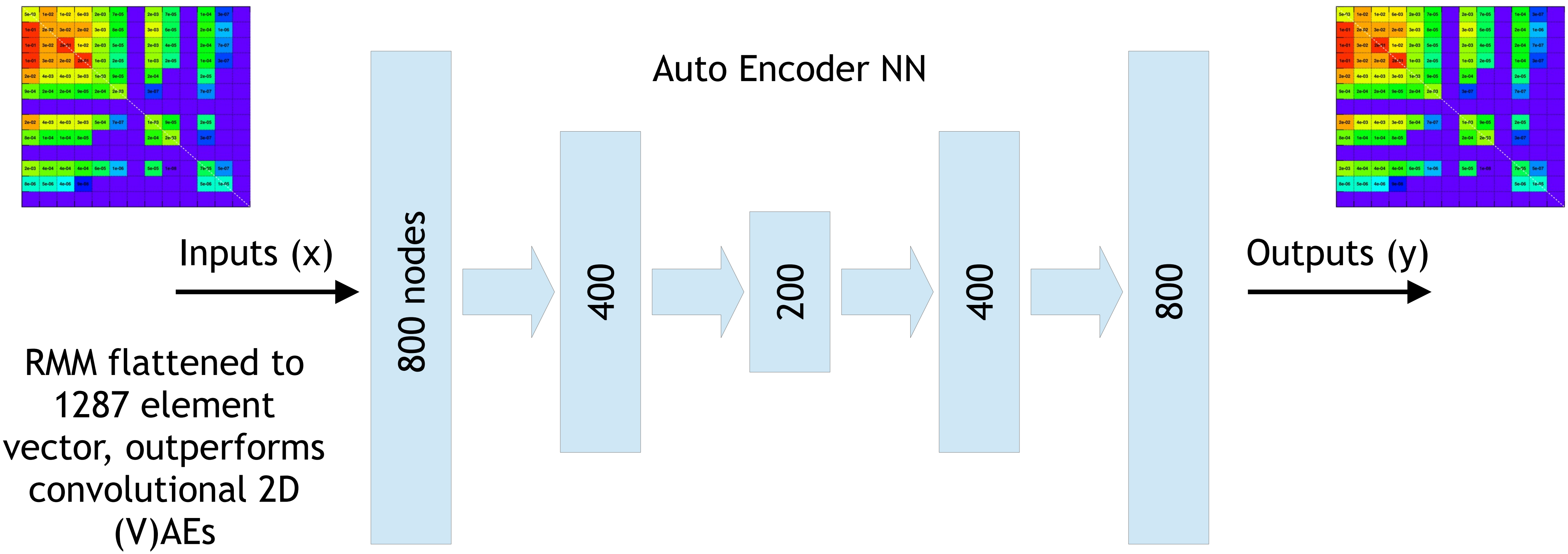
Top ($t\bar{t}$) production

Nucl.Instrum.Meth.A 931 (2019) 92-99

RMM for anomaly detection (2024)

Use an RMM containing 10 light jets, 10 b-jets, 5 electrons, 5 muons and 5 photons to perform a multi-channel anomaly search (trigger on single lepton)

Remove invariant mass observables for search to avoid biasing these spectra



Anomaly score is log(L), where L is the training loss L=|**y**-**x**|^2

Phys.Rev.Lett. 132 (2024) 8, 081801

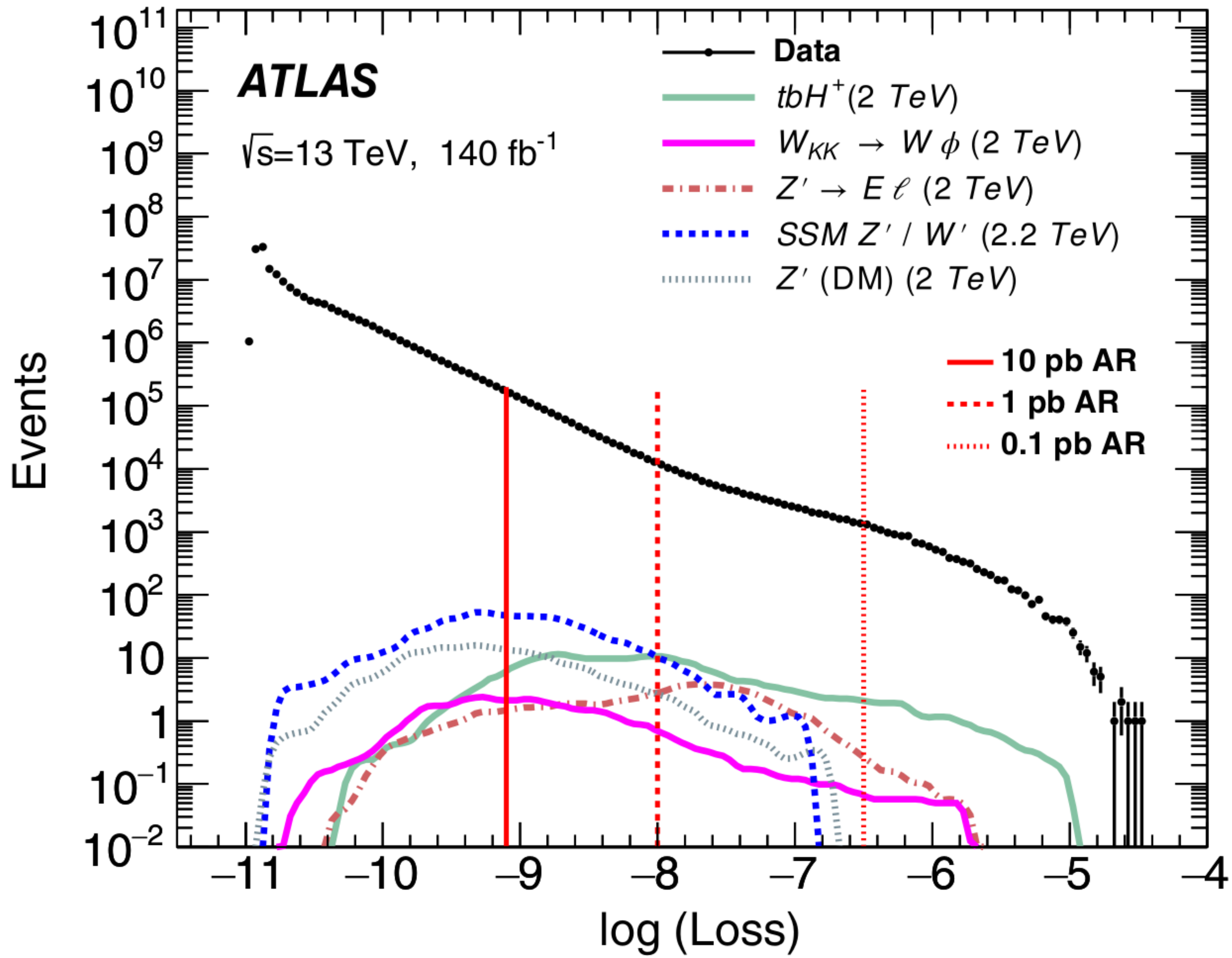
RMM for anomaly detection (2024)

Plot the anomaly score distribution for Run2 dataset, alongside example BSM signals

- One or more leptons with p_T>60 GeV
- One or more jets with p_T>30 GeV

AE trained on 1% of full Run2 dataset (140 fb^{-1})

Define three anomaly score thresholds corresponding to 10, 1 and 0,1 pb cross-sections



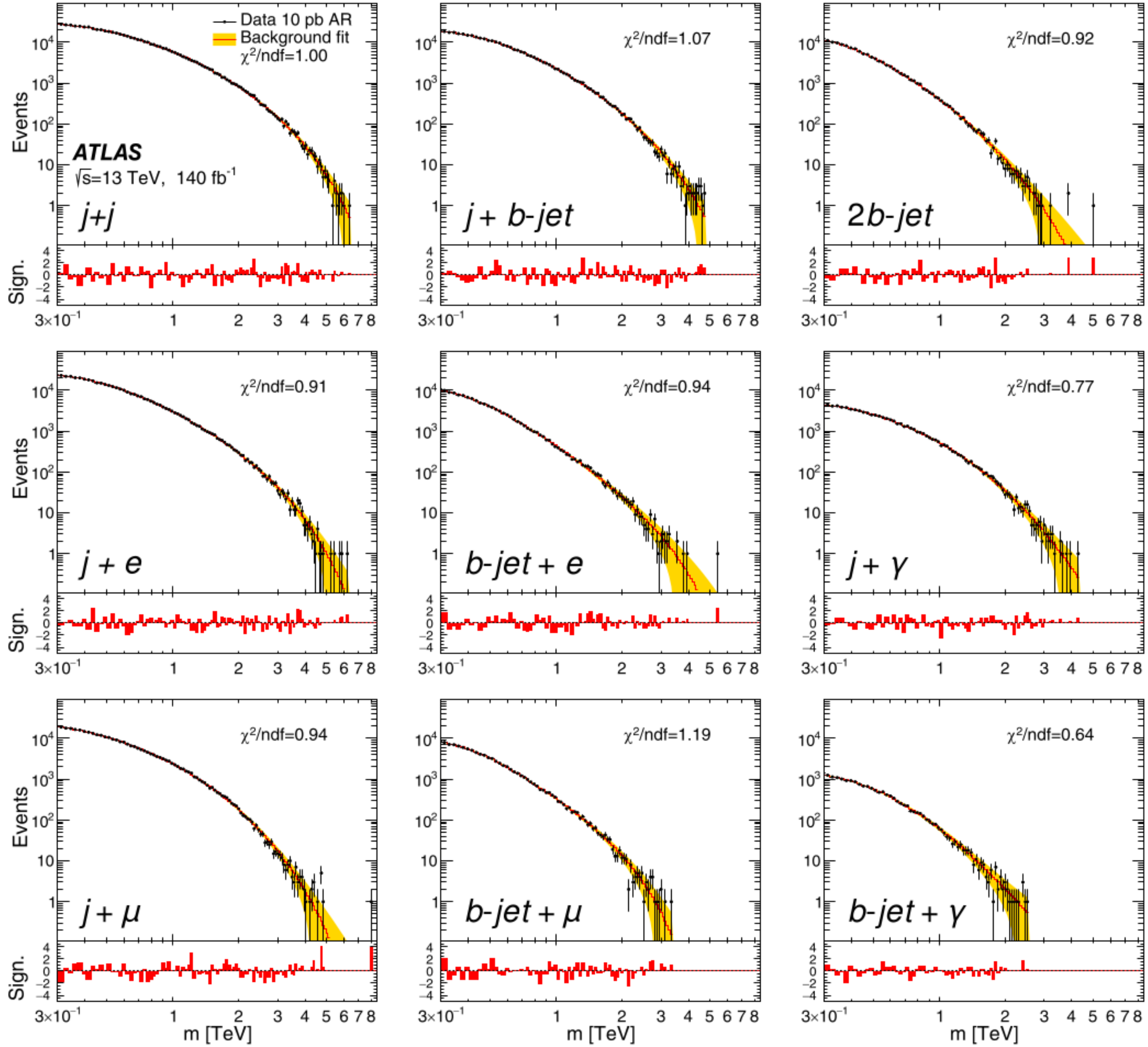
Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

Plot the invariant masses that were excluded from the RMM earlier

Bump-hunt search for local excesses above a smooth, analytical function for the background

These plots are for the 10pb anomaly region: more stringent selections were found to be statistically-limited



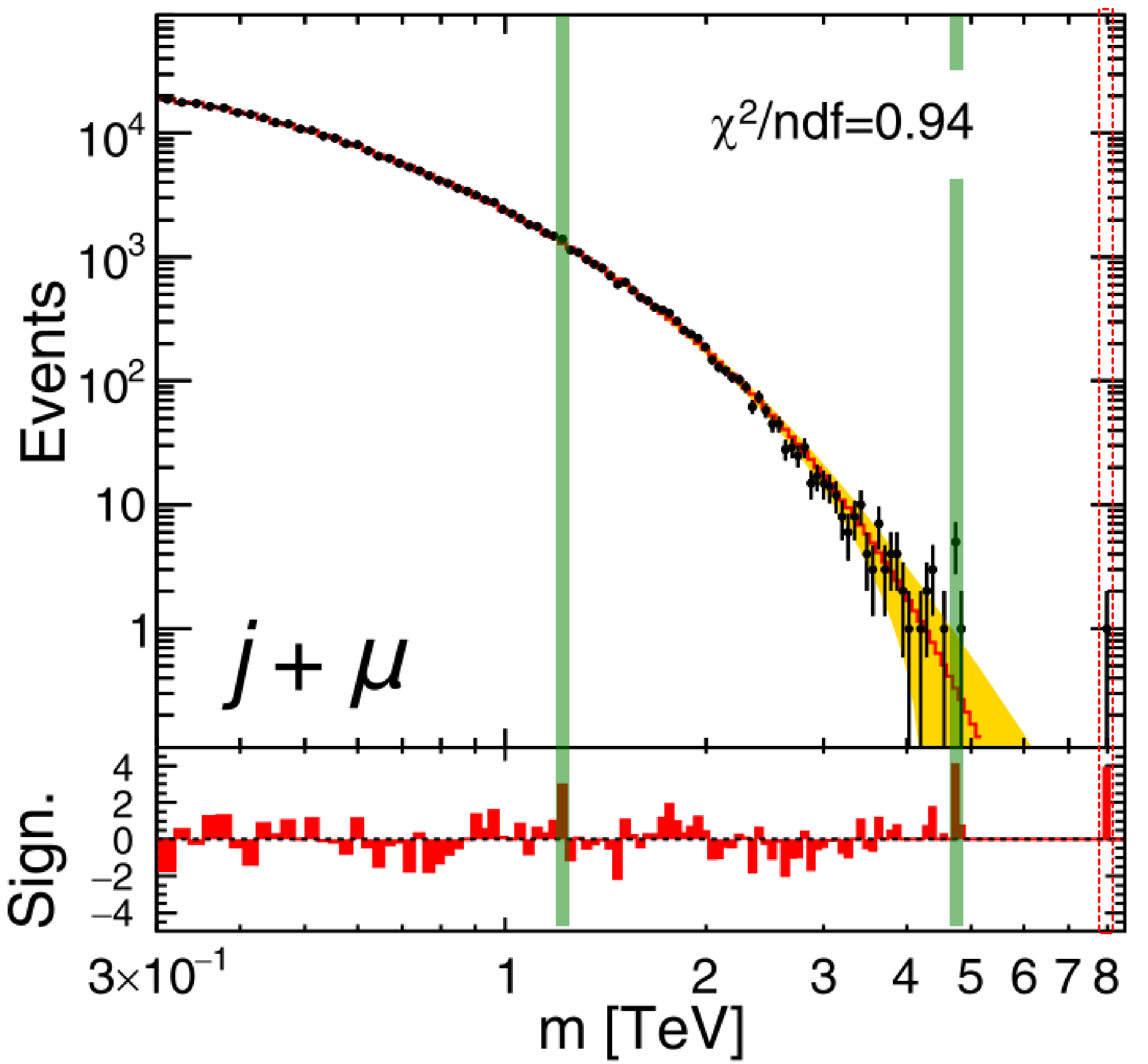
Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

Largest excesses in the m(j+μ) distribution at 1.2 and 4.8 TeV

Assuming a zero-width decay at these masses, local significances of 2.8 and 2.9σ

Event at m(j+μ)=8 TeV is removed by more stringent muon ID



Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

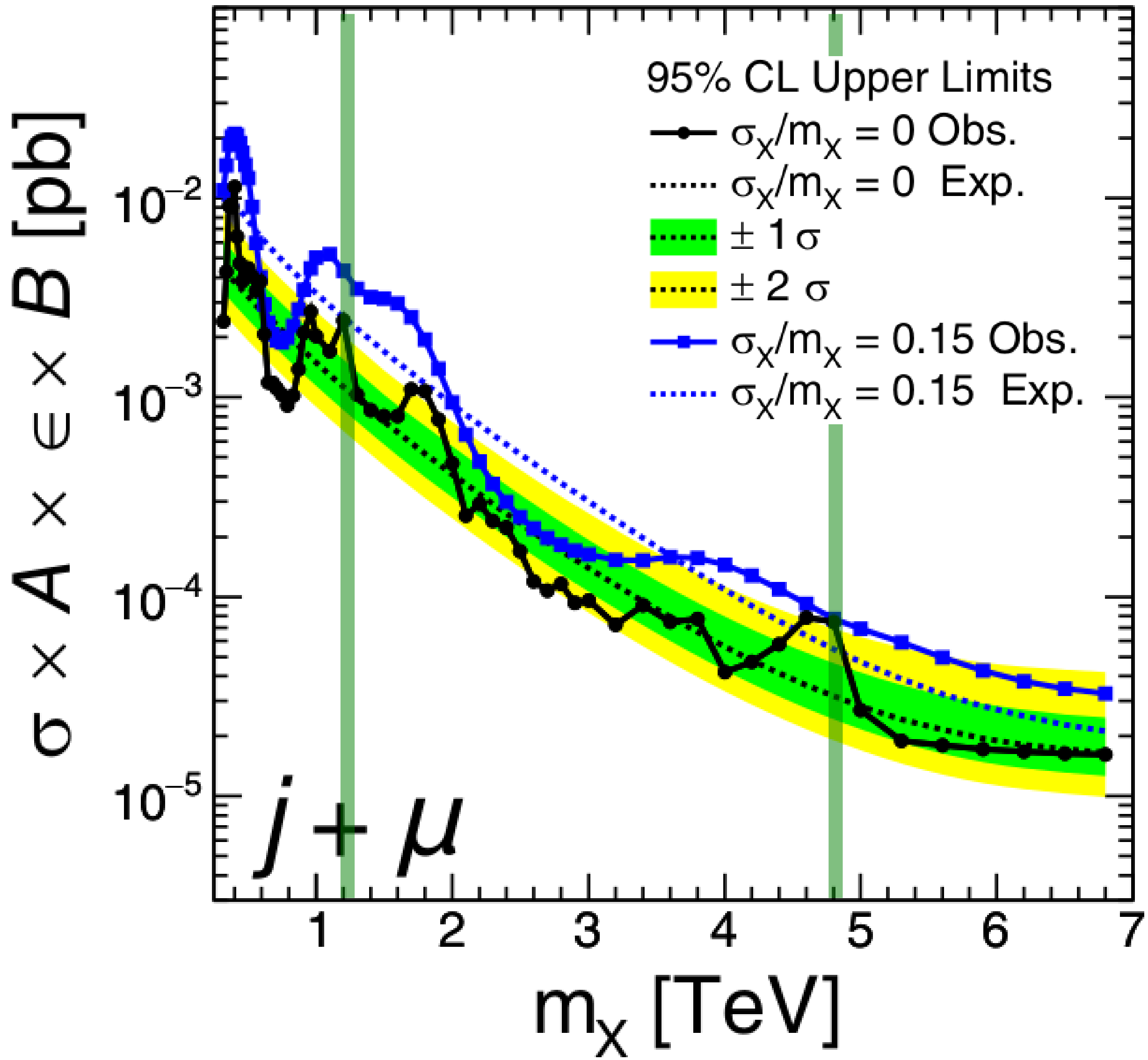
Largest excesses in the m(j+μ) distribution at 1.2 and 4.8 TeV

Assuming a zero-width decay at these masses, local significances of 2.8 and 2.9σ

Calculated assuming a zero-width decay

Results for a 15% width also shown (less significant)

Other channels in backup



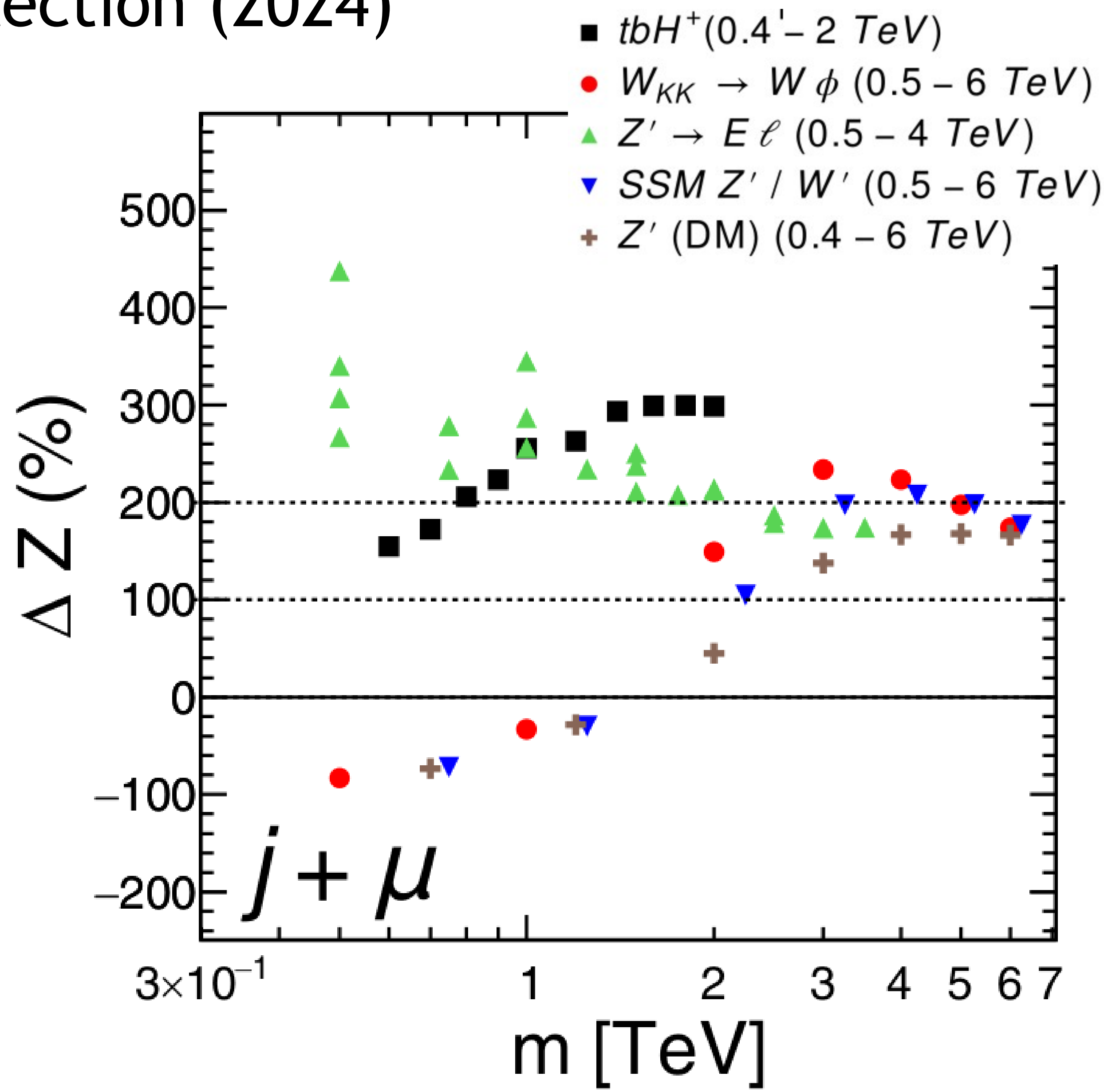
Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

Sensitivity for the different example signals are shown for the 10pb anomaly score selection, as percentage improvement (ΔZ) over no anomaly score cut

Lower masses and more SM-like final states give worse performance, as you might expect

Other channels in backup



Phys.Rev.Lett. 132 (2024) 8, 081801

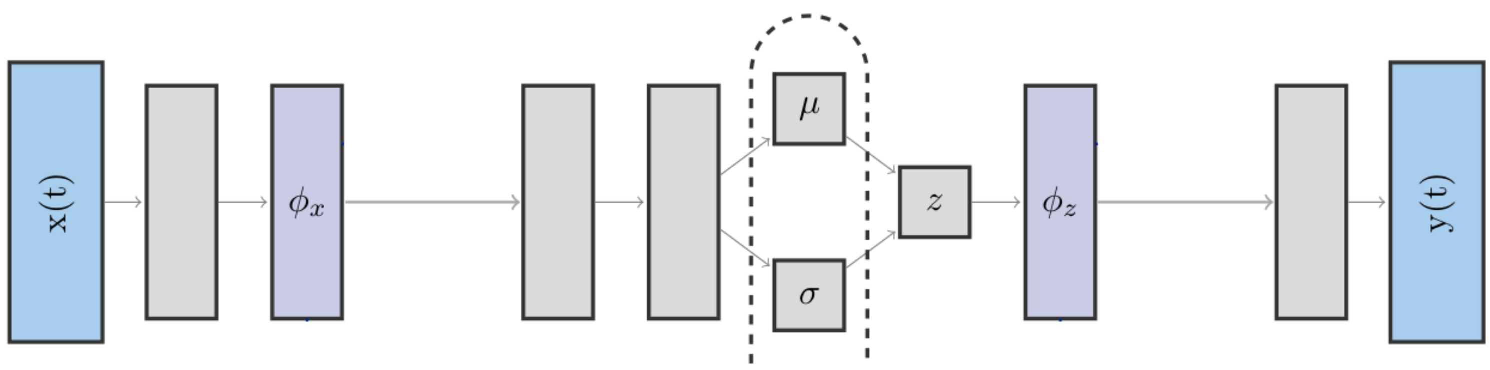
Anomalous jet tagging

Encoding features – variable size

Fixed-size RMMs must strike a balance between truncating event data, and sparsity. Even so, the loss term in training has a dependency on matrix content, and thus particle multiplicity

Recurrent Neural Networks (RNN) are designed to process variable-length data vectors as sequences, with an internal state affected by earlier steps in the sequence

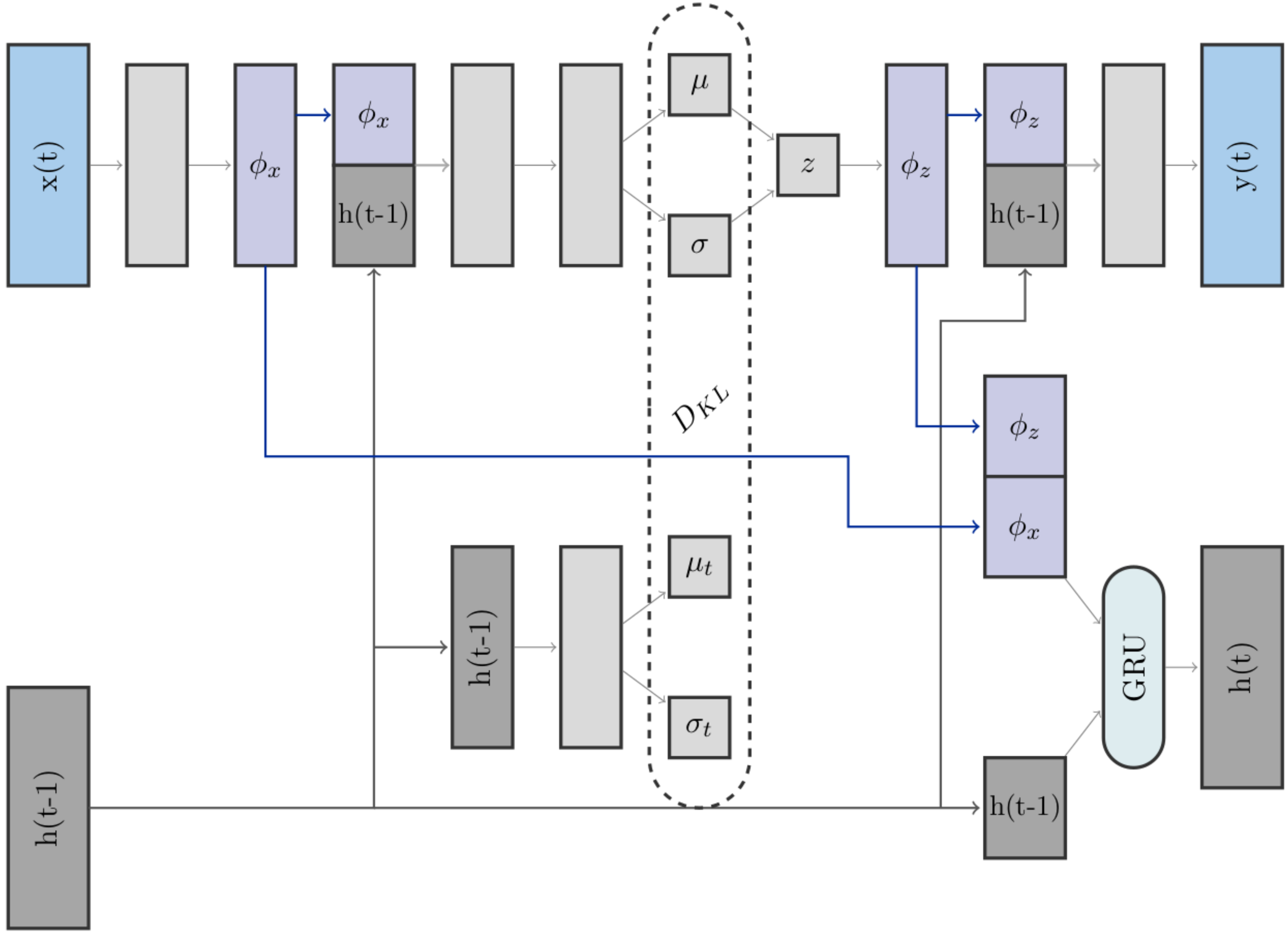
We can use this approach to modify a standard (Variational) Auto Encoder...



JINST 16 (2021) 08, P08012

Encoding features – variable size

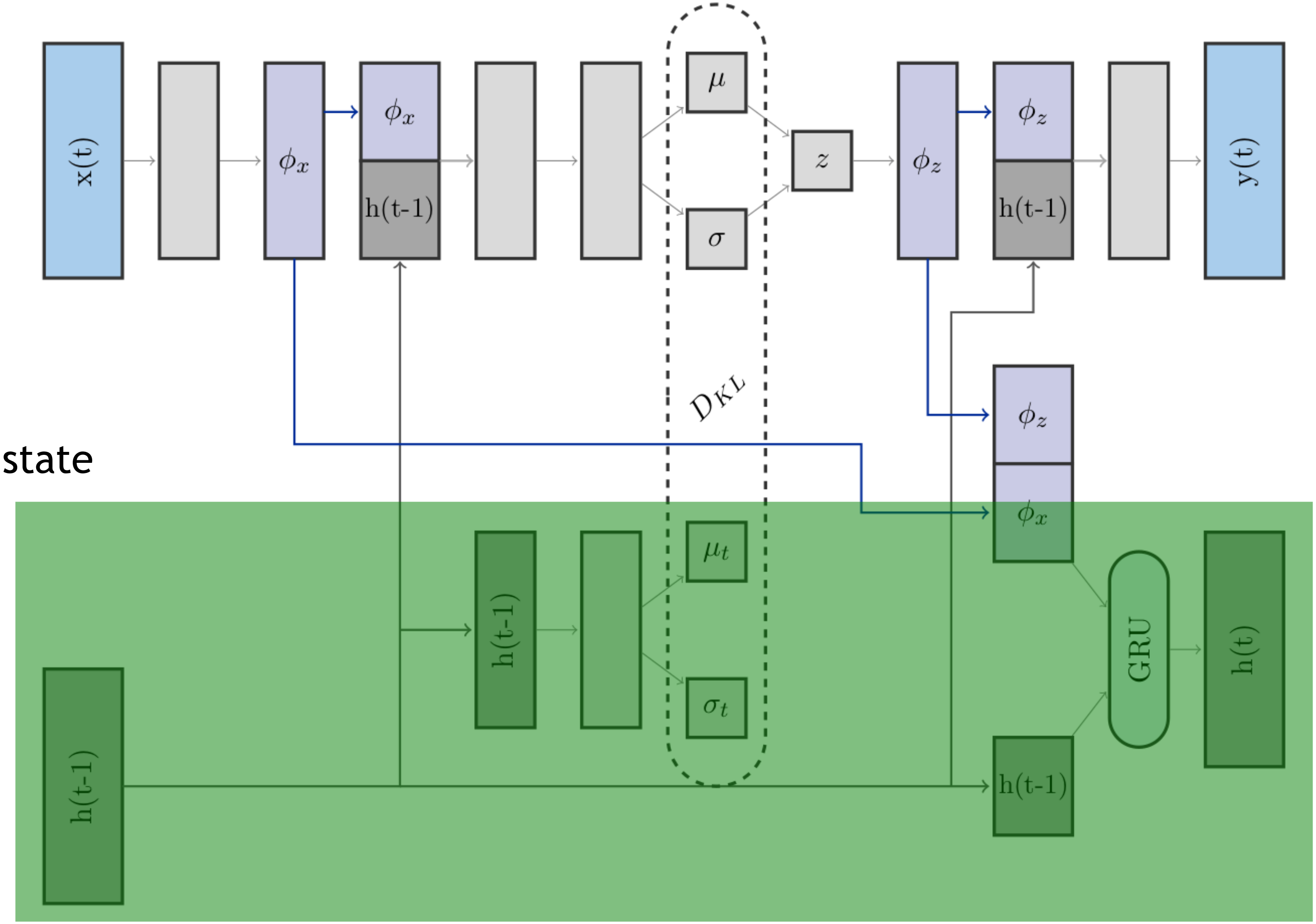
… adding an RNN structure around it, creating a Variational Recurrent Neural Network (VRNN)



JINST 16 (2021) 08, P08012

Encoding features – variable size

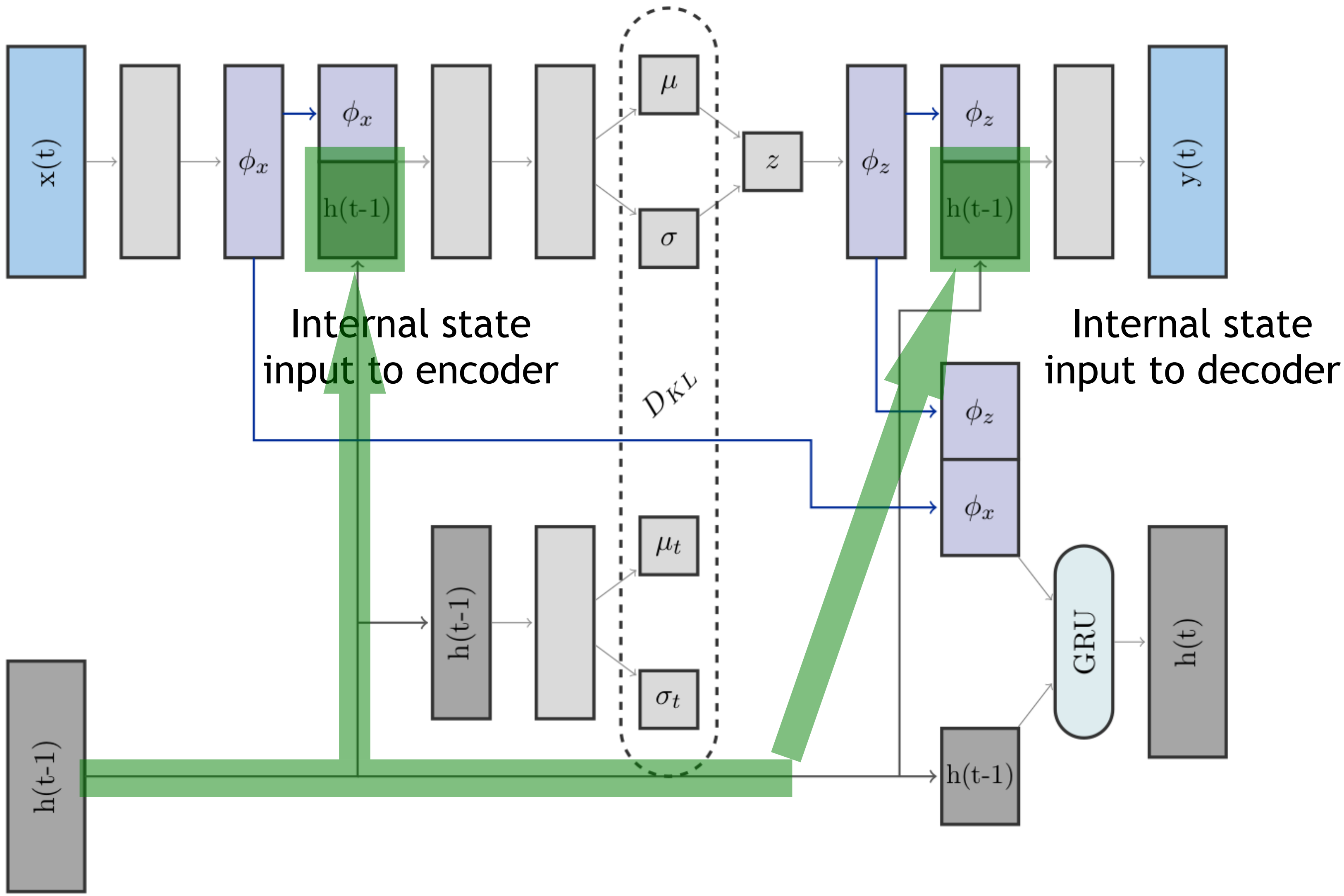
… adding an RNN structure around it, creating a Variational Recurrent Neural Network (VRNN)



JINST 16 (2021) 08, P08012

Encoding features – variable size

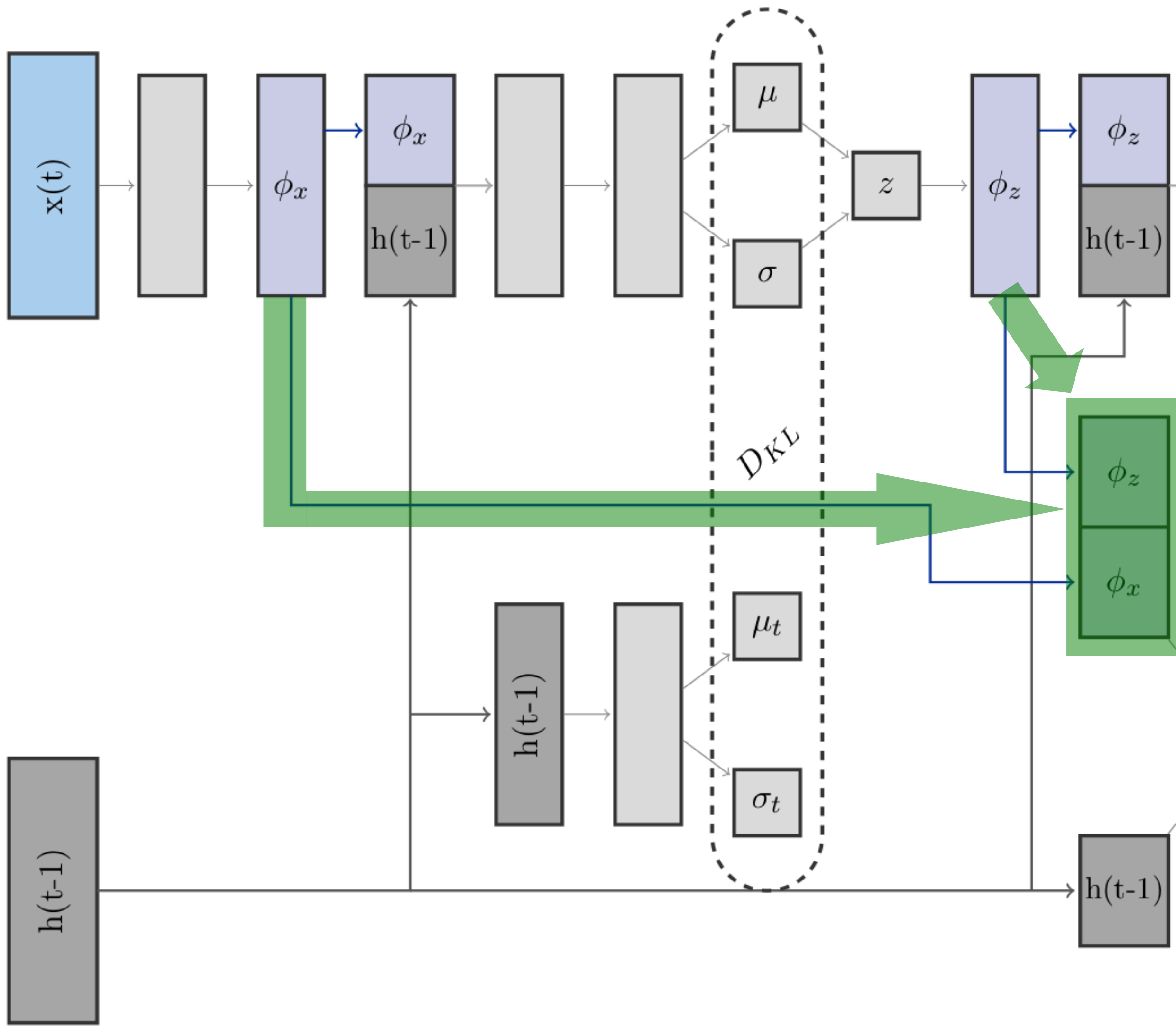
… adding an RNN structure around it, creating a Variational Recurrent Neural Network (VRNN)



JINST 16 (2021) 08, P08012

Encoding features – variable size

… adding an RNN structure around it, creating a Variational Recurrent Neural Network (VRNN)



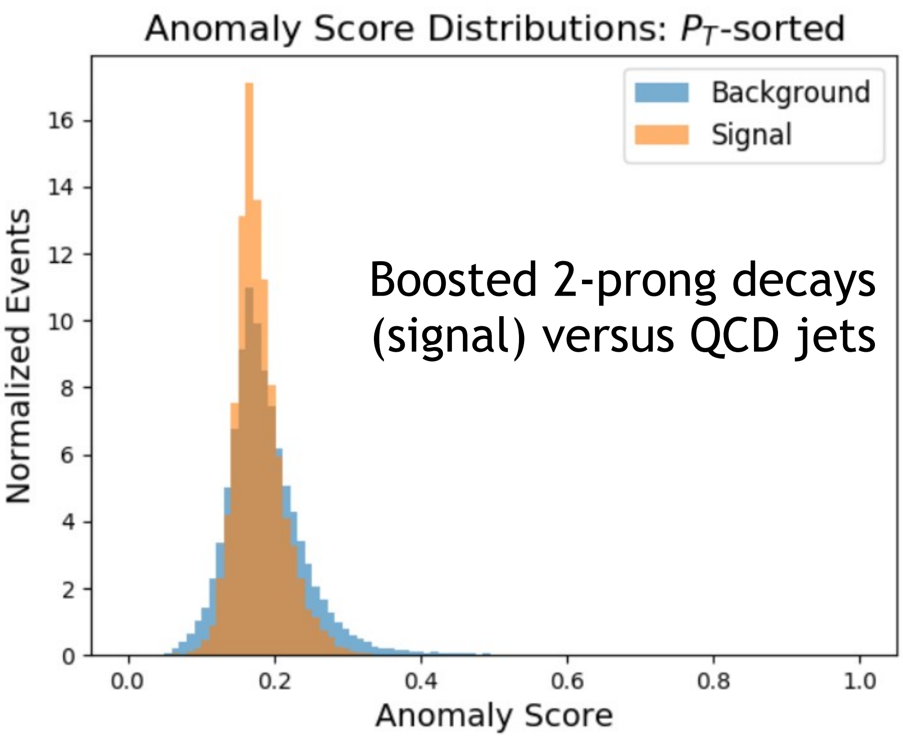
Internal state updated from encoder and decoder

JINST 16 (2021) 08, P08012

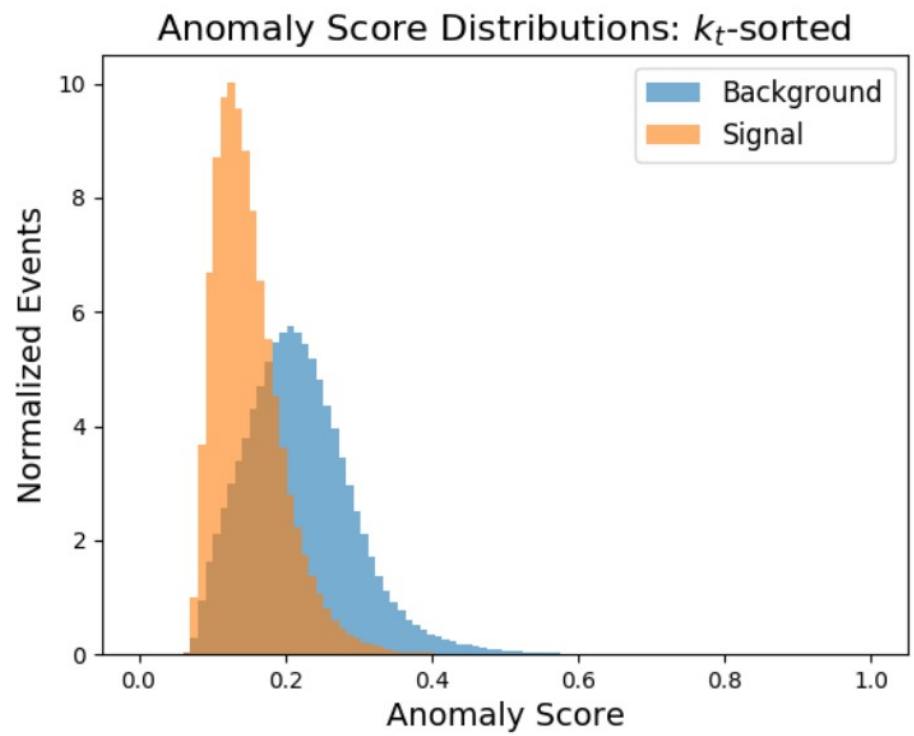
Encoding features – jets

The VRNN is able to perform anomaly detection on sequential data of unknown length, in this case the 20 highest p_T constituents of a jet as 4-vectors

Unlike a standard Auto Encoder, the ordering of the input is meaningful and has an effect on the performance



Boosted 2-prong decays (signal) versus QCD jets

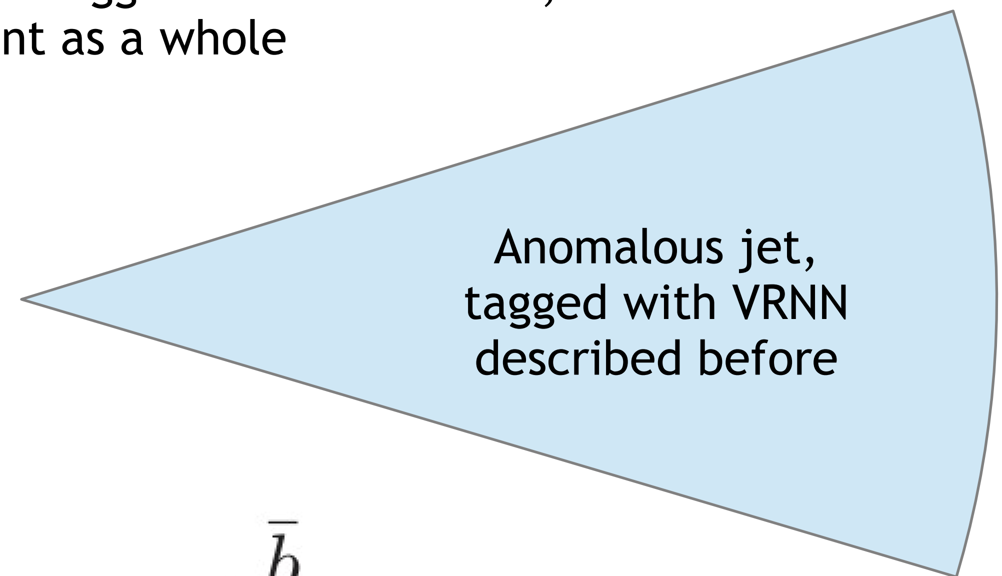


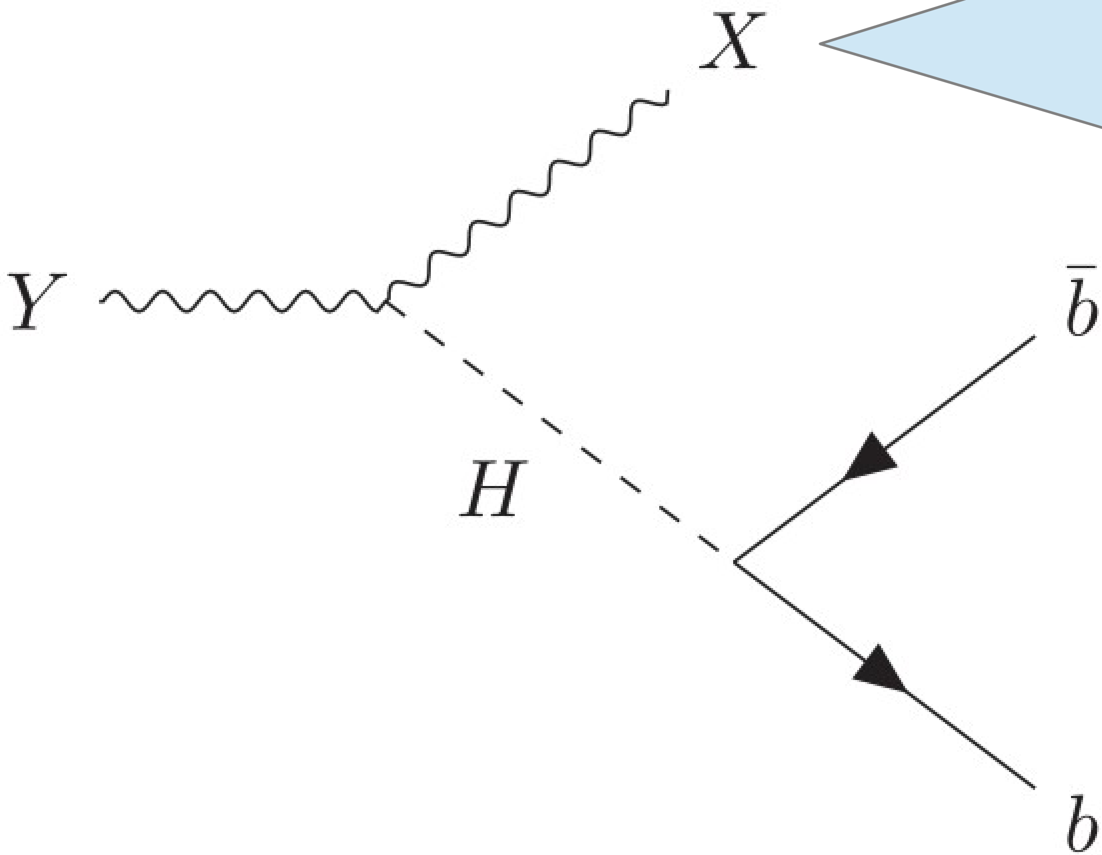
Identify subjets by ordering to maximise k_t-distance between sequence steps
Jets are scaled to avoid mass-dependence (see backup)

JINST 16 (2021) 08, P08012

Anomalous jet tagging analysis

Using the VRNN, anomalous jets can be tagged within an event, rather than assigning an anomaly score to the event as a whole



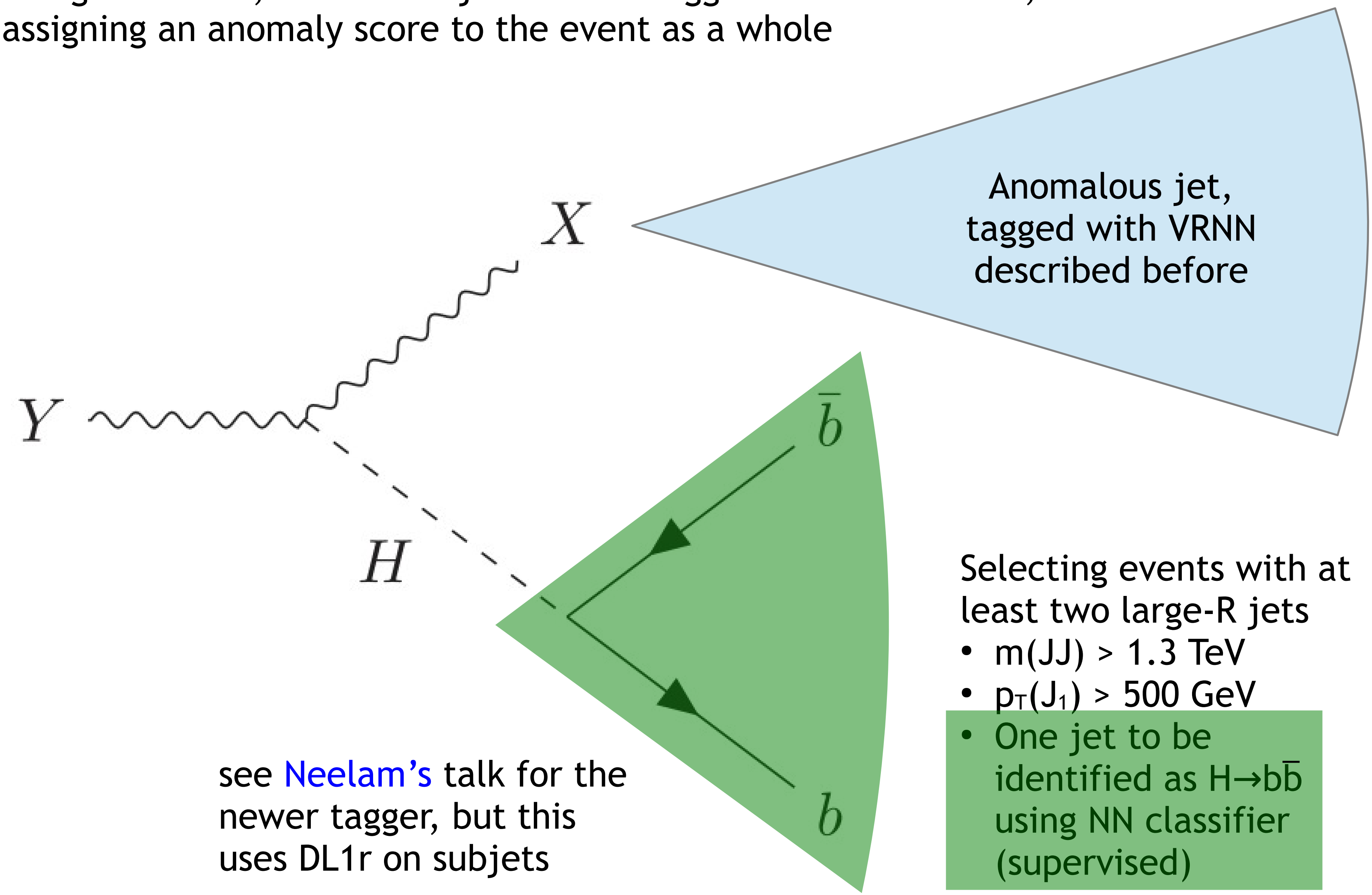


Selecting events with at least two large-R jets
- m(JJ) > 1.3 TeV
- $p_T(J_1)$ > 500 GeV
- One jet to be identified as $H \to b\bar{b}$ using NN classifier (supervised)

Phys.Rev.D 108 (2023) 052009

Anomalous jet tagging analysis

Using the VRNN, anomalous jets can be tagged within an event, rather than assigning an anomaly score to the event as a whole



Selecting events with at least two large-R jets

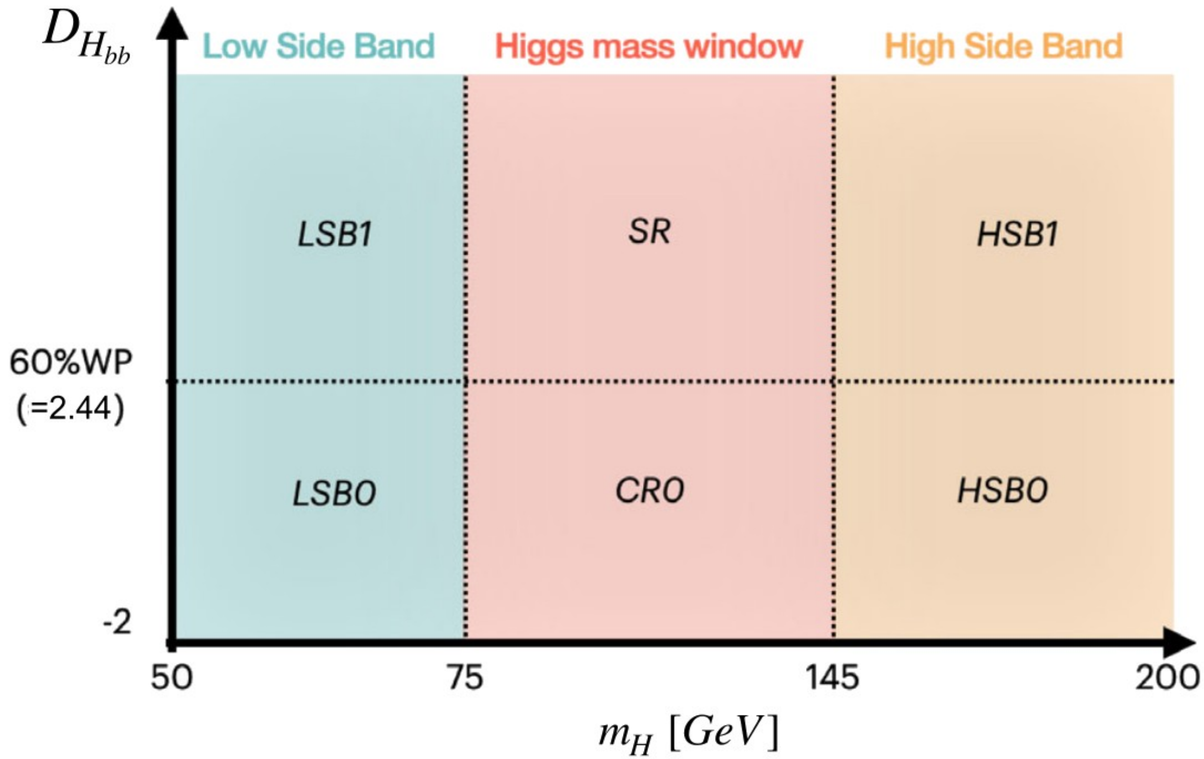
- m(JJ) > 1.3 TeV
- $p_T(J_1)$ > 500 GeV
- One jet to be identified as $H \to b\bar{b}$ using NN classifier (supervised)

see Neelam's talk for the newer tagger, but this uses DL1r on subjets

Phys.Rev.D 108 (2023) 052009

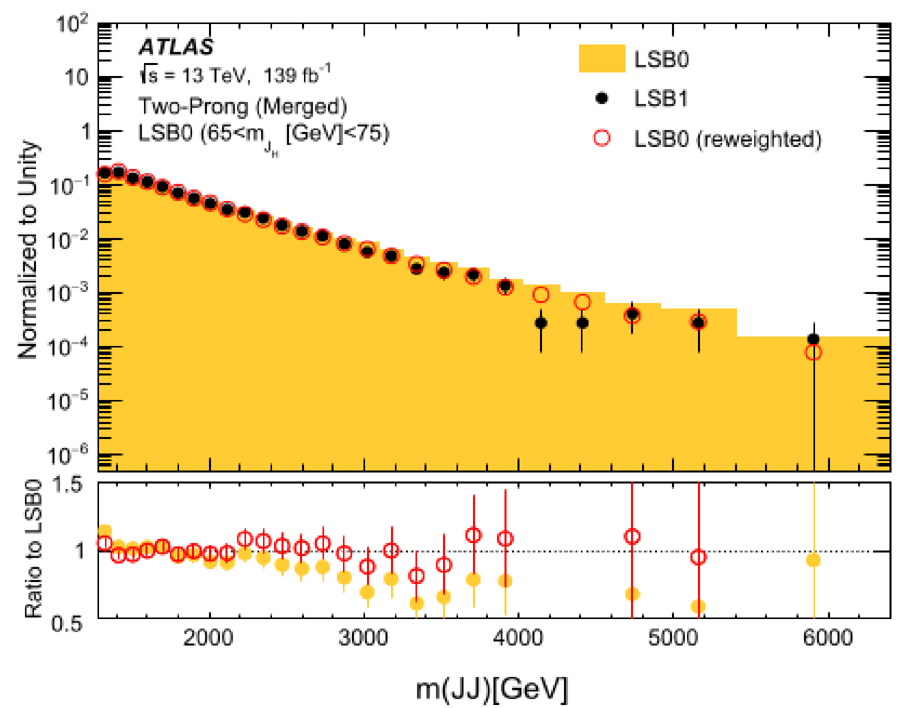
Analysis regions

Two variables are used to define the signal region: the mass of the $H \to b\bar{b}$ candidate jet, and the NN classifier score for that jet



Background modelling:

- Use NN (3 layers, 20 nodes) to predict event-level weights that will correct HSB0 data to reproduce HSB1
- Validate by comparing LSB0 (corrected) with LSB1
- Apply to CR0

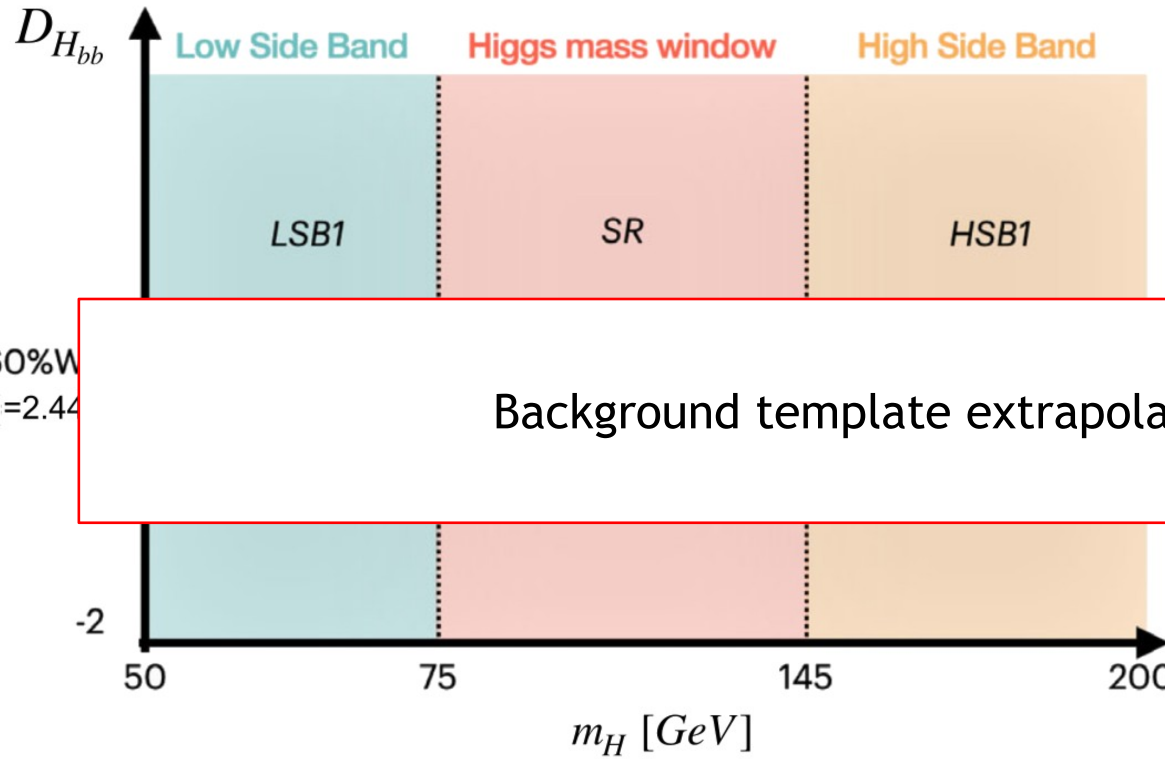


Event weighting uses properties of the $H \to b\bar{b}$ candidate jet (e.g. p_T, η) and those of its sub-jets

Phys.Rev.D 108 (2023) 052009

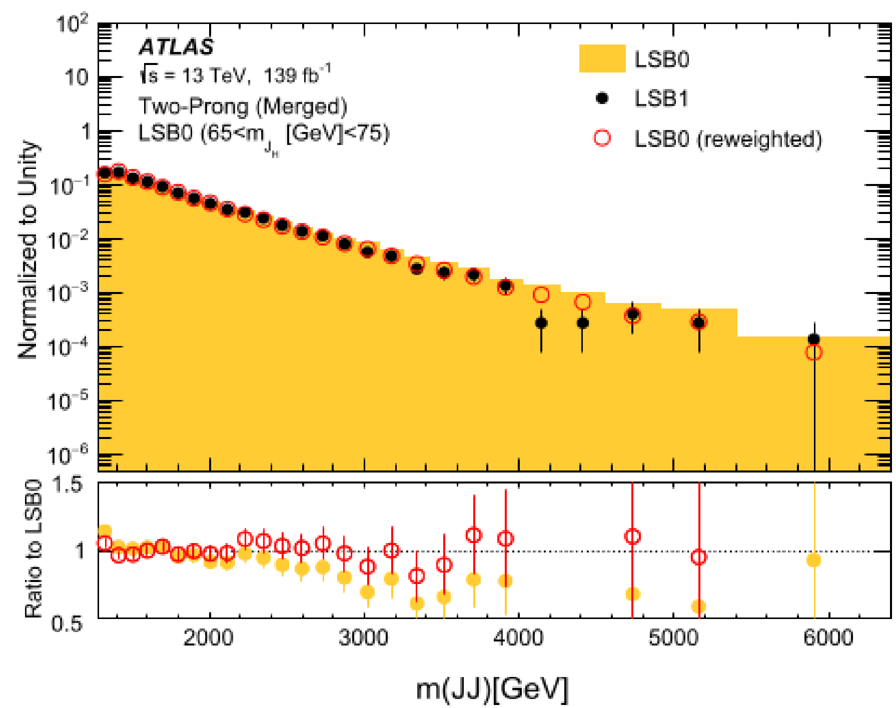
Analysis regions

Two variables are used to define the signal region: the mass of the H→b$\bar{b}$ candidate jet, and the NN classifier score for that jet



Background modelling:
- Use NN (3 layers, 20 nodes) to predict event-level weights that will correct HSB0 data to reproduce HSB1

Background template extrapolated from sidebands

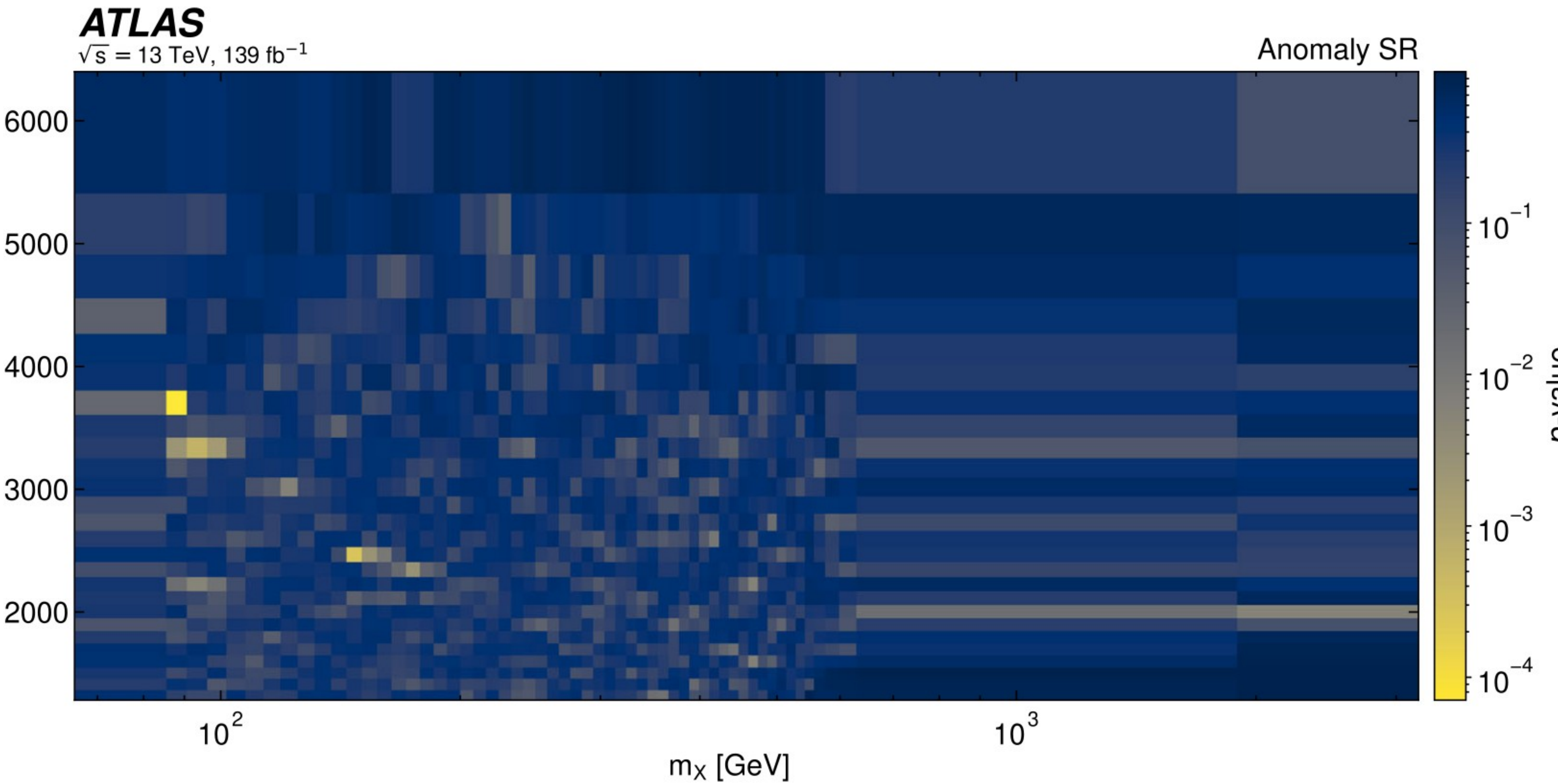


Event weighting uses properties of the H→b$\bar{b}$ candidate jet (e.g. p_T, η) and those of its sub-jets

Phys.Rev.D 108 (2023) 052009

Results

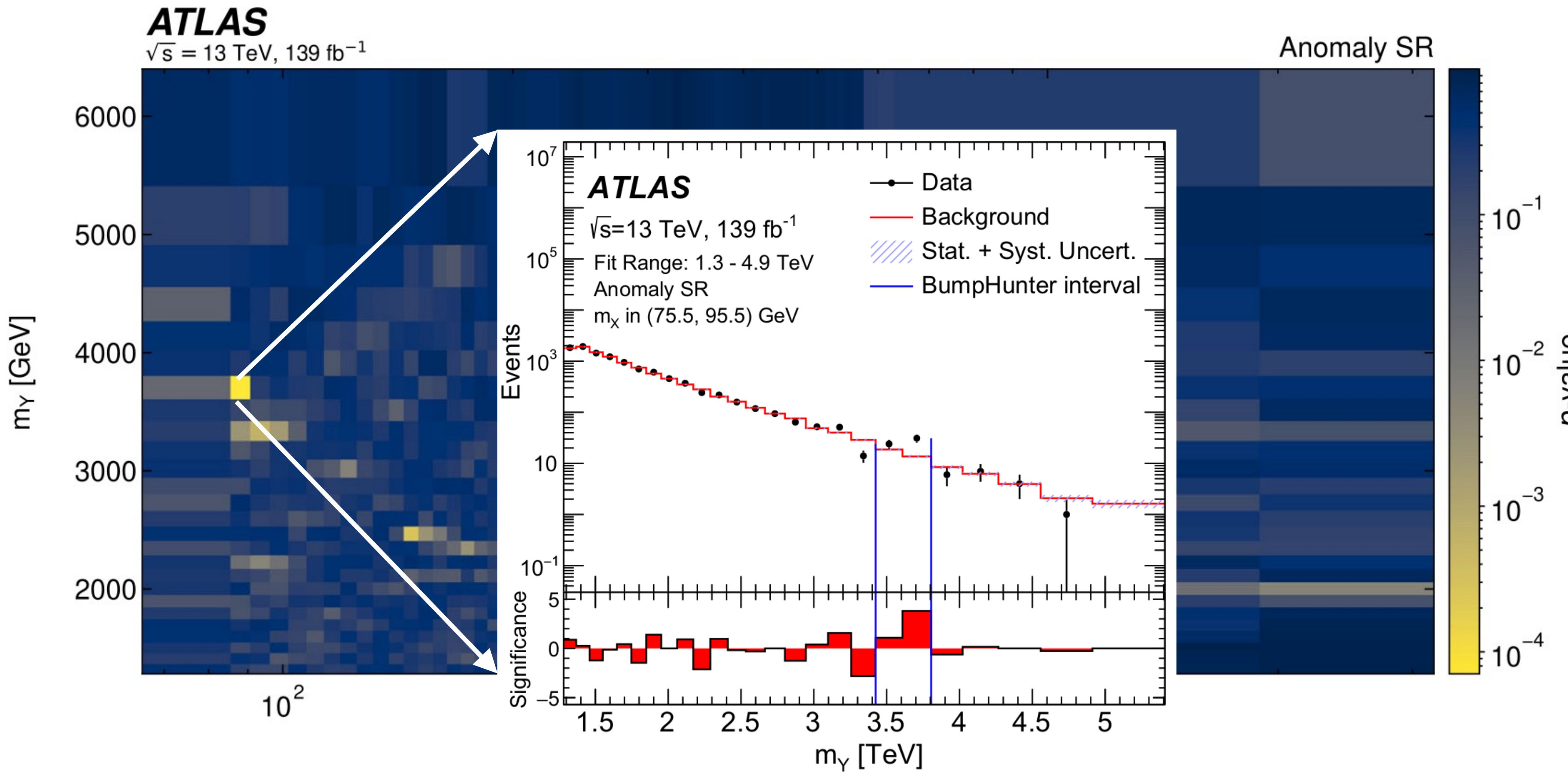
Large space of potential m(Y) and m(X) values searched, looking for excesses above the predicted background



Phys.Rev.D 108 (2023) 052009

Results

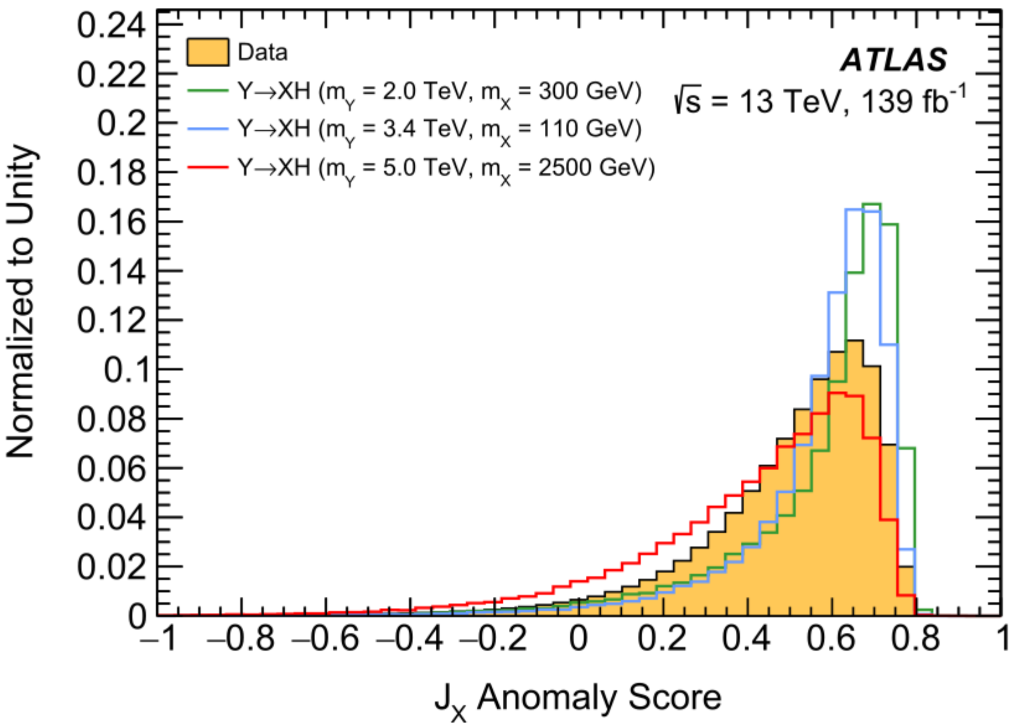
Largest single excess has a global significance of 1.4σ once the trials factor from all these bins is accounted for

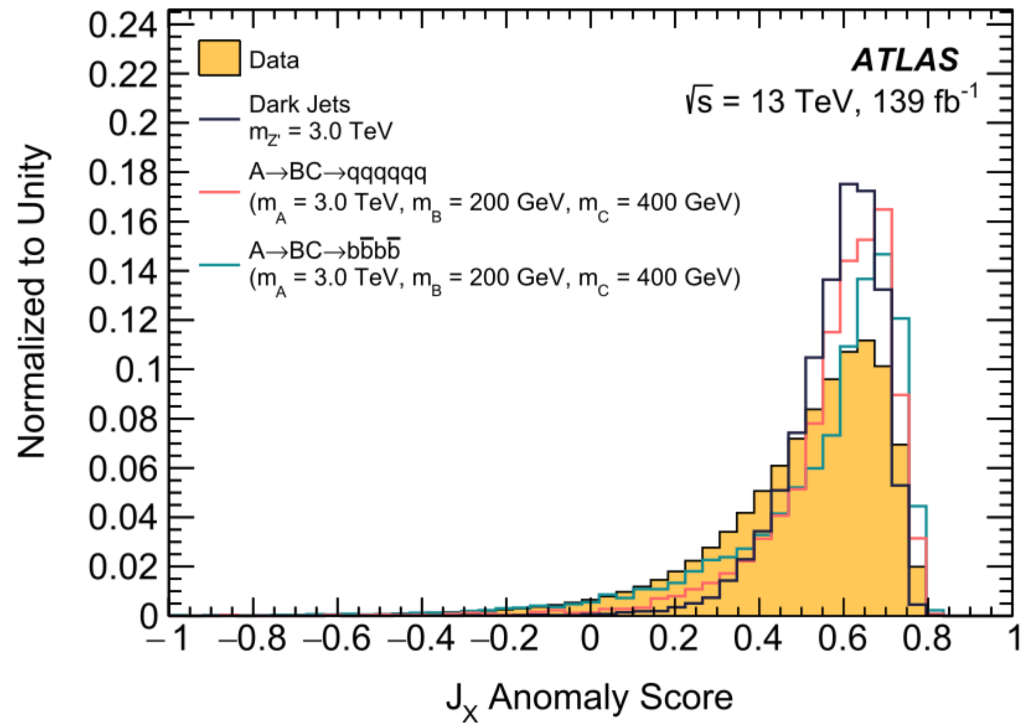


Phys.Rev.D 108 (2023) 052009

Anomalous jet tagging effectiveness

Comparing the anomaly scores for data and a variety of potential signal models, we see modest discriminating power: the chosen working point of score > 0.5 gives approximately 25% improved S:B ratio compared to an inclusive selection

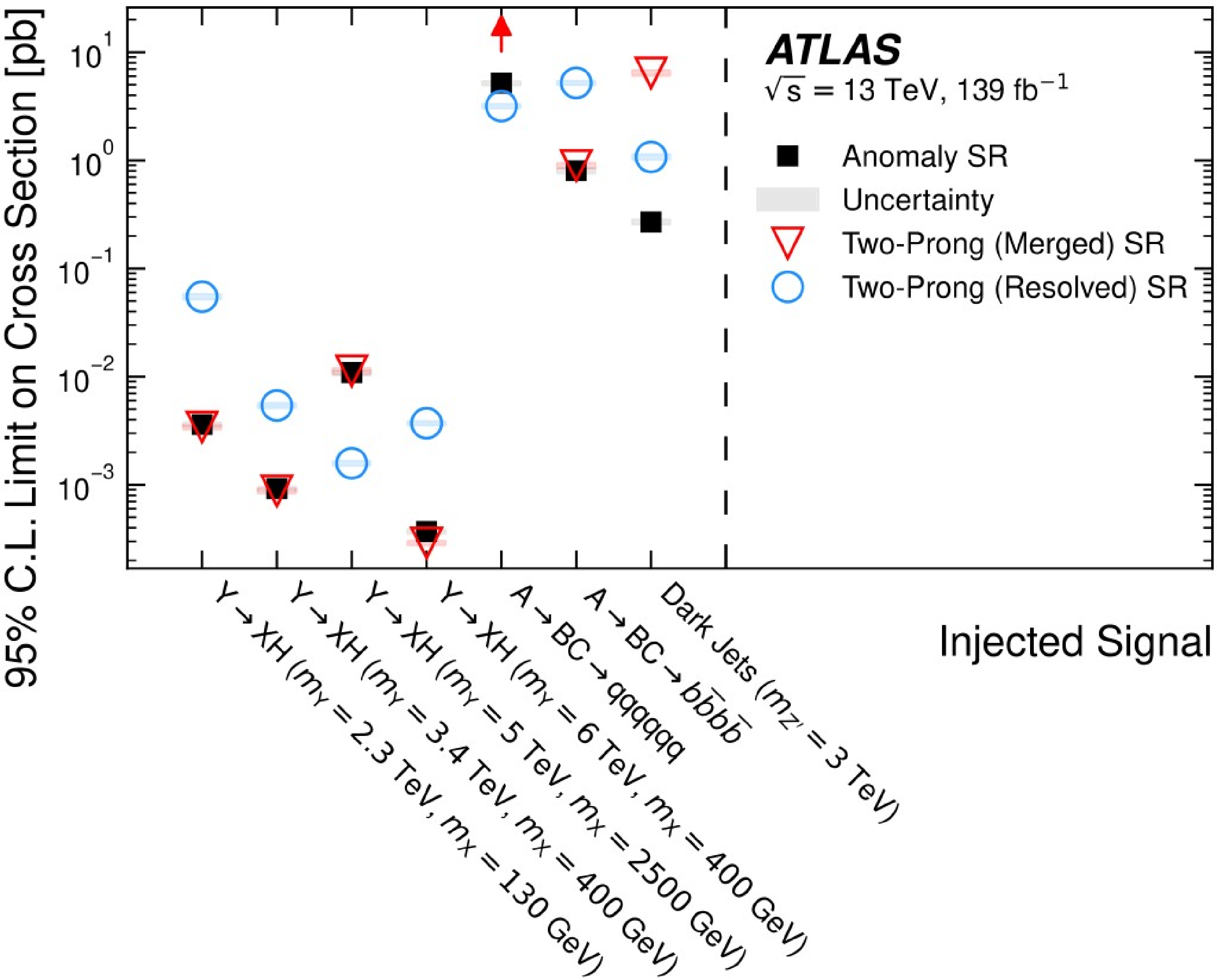




By way of a comparison, two other versions of the analysis were performed, replacing the anomalous jet tag with either a large-R jet with $D_2 < 1.2$, or two small-R resolved jets

Phys.Rev.D 108 (2023) 052009

Anomalous jet tagging effectiveness

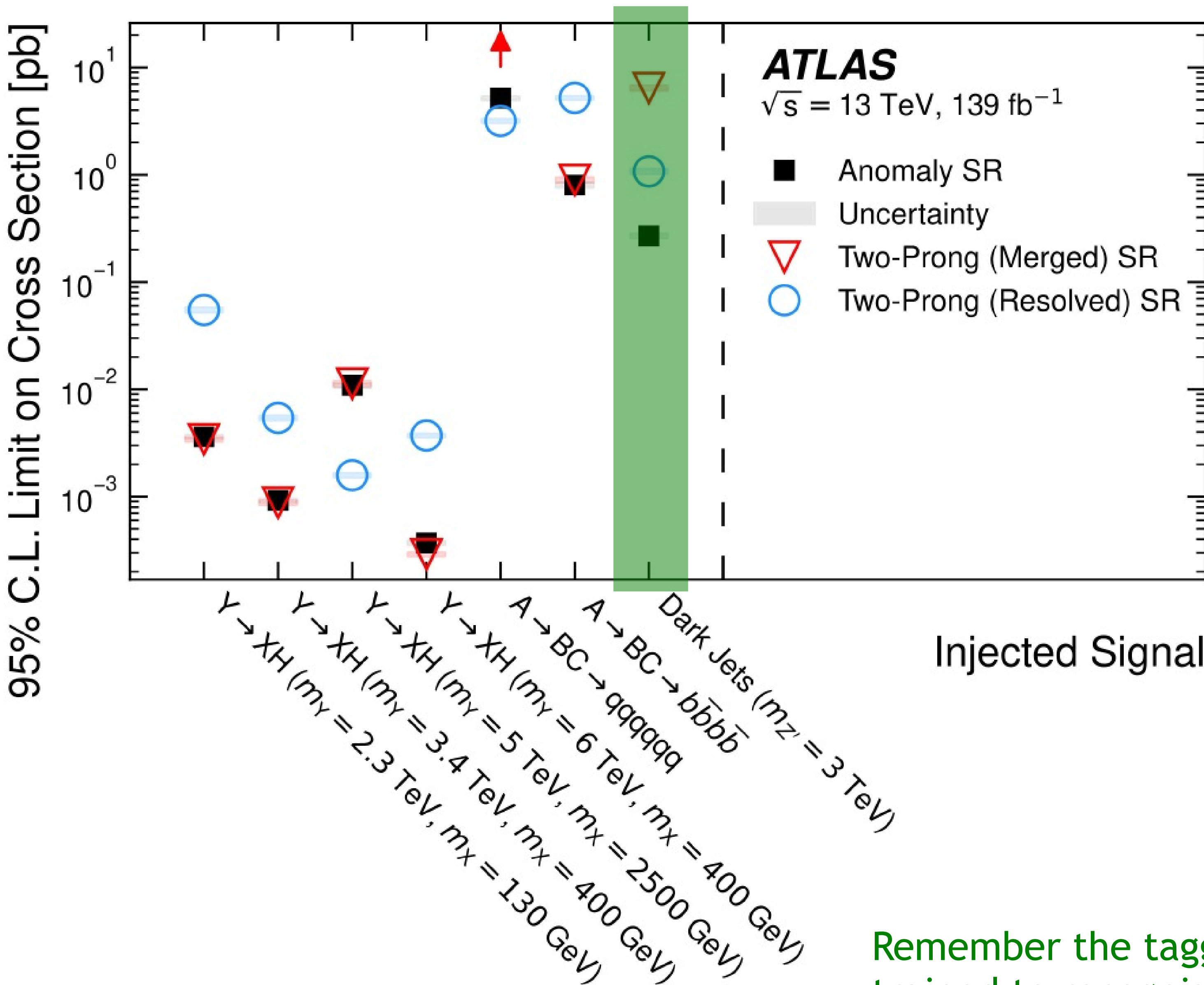


For the different signal models shown, the anomaly-tagging approach is approximately equivalent to finding substructured large jets $D_2 < 1.2$

Note the Dark Jets signal though – the anomaly detection method is significantly better (order of magnitude) for this unusual final state, while remaining competitive in other channels

Phys.Rev.D 108 (2023) 052009

Anomalous jet tagging effectiveness



For the different signal models shown, the anomaly-tagging approach is approximately equivalent to finding substructured large jets $D_2 < 1.2$

Note the Dark Jets signal though – the anomaly detection method is significantly better (order of magnitude) for this unusual final state, while remaining competitive in other channels

Remember the tagger has never been trained to recognise any of these models

Phys.Rev.D 108 (2023) 052009

Summary

Anomaly detection is a promising technique, allowing for identification of unusual signatures without targeting a specific new physics model

Improved sensitivity over inclusive searches in the same channels

Unsupervised training avoids modelling uncertainties in backgrounds, particularly for important LHC processes like QCD-multijets

Full event-level information can be summarised in fixed-size structures like the RMM, allowing searches across many channels, with anomaly scoring provided by an Auto Encoder NN

Anomalous, substructured jets can be identified using a VRNN, taking an ordered sequence of jet constituents as its input

- Comparable performance to existing substructure variables
- Additional sensitivity, e.g. semi-visible signatures from Dark Jets

No significant excesses discovered, but these techniques show plenty of room for improvement (e.g. only 25% sensitivity improvement for VRNN so far)

Looking forward to more Run-3 data, and HL-LHC!

BACKUP

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

Choose particle flavours to include (e.g. jets, muons) and max multiplicity N

$$
\begin{array}{c}
\hspace{22em}\text{Muons } 1\rightarrow N \ \ (e_T \text{ ordered}) \\
\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e_T}(\mathbf{j_1}) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta\mathbf{e_T}(\mathbf{j_2}) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta\mathbf{e_T}(\mathbf{j_N}) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & \mathbf{e_T}(\boldsymbol{\mu_1}) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta\mathbf{e_T}(\boldsymbol{\mu_2}) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta\mathbf{e_T}(\boldsymbol{\mu_N})
\end{pmatrix}
\end{array}
$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

Diagonal components are (missing) transverse energies, or differences:

$$\delta e_T(i_n) = \frac{E_T(i_{n-1}) - E_T(i_n)}{E_T(i_{n-1}) + E_T(i_n)}, \quad n = 2, \dots, N,$$

$$\begin{pmatrix}
\mathbf{e_T^{miss}} & m_T(j_1) & m_T(j_2) & \dots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \dots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e_T(j_1)} & m(j_1,j_2) & \dots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \dots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \boldsymbol{\delta}\mathbf{e_T(j_2)} & \dots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \dots m(j_2,\mu_N) \\
\dots & \dots & \dots & \dots, & \dots & \dots & \dots \\
h_L(j_N) & h(j_1,j_N) & \dots & \dots \boldsymbol{\delta}\mathbf{e_T(j_N)} & m(j_N,\mu_1) & m(j_N,\mu_2) & \dots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \dots h(\mu_1,j_N) & \mathbf{e_T}(\boldsymbol{\mu_1}) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \dots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \boldsymbol{\delta}\mathbf{e_T}(\boldsymbol{\mu_2}) & m(\mu_2,\mu_N) \\
\dots & \dots & \dots & \dots & \dots & \dots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \dots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \boldsymbol{\delta}\mathbf{e_T}(\boldsymbol{\mu_N})
\end{pmatrix}$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

First row characterises decays for one observed particle and one invisible

$$M_T = \sqrt{(E_T + E_T^{\mathrm{miss}})^2 - (\boldsymbol{p}_T + \boldsymbol{E}_T^{\mathrm{miss}})^2}.$$

$$\begin{pmatrix}
\mathbf{e}_\mathbf{T}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_\mathbf{T}(\mathbf{j_1}) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta\mathbf{e}_\mathbf{T}(\mathbf{j_2}) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta\mathbf{e}_\mathbf{T}(\mathbf{j_N}) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & \mathbf{e}_\mathbf{T}(\mu_\mathbf{1}) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta\mathbf{e}_\mathbf{T}(\mu_\mathbf{2}) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta\mathbf{e}_\mathbf{T}(\mu_\mathbf{N})
\end{pmatrix}$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

Other elements above the diagonal are two-particle invariant masses

$$
\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j_1}) & m(j_1, j_2) & \ldots m(j_1, j_N) & m(j_1, \mu_1) & m(j_1, \mu_2) & \ldots m(j_1, \mu_N) \\
h_L(j_2) & h(j_1, j_2) & \delta \mathbf{e}_{\mathbf{T}}(\mathbf{j_2}) & \ldots m(j_2, j_N) & m(j_2, \mu_1) & m(j_2, \mu_2) & \ldots m(j_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1, j_N) & \ldots & \ldots \delta \mathbf{e}_{\mathbf{T}}(\mathbf{j_N}) & m(j_N, \mu_1) & m(j_N, \mu_2) & \ldots m(j_N, \mu_N) \\
h_L(\mu_1) & h(\mu_1, j_1) & h(\mu_1, j_2) & \ldots h(\mu_1, j_N) & \mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{1}}) & m(\mu_1, \mu_2) & m(\mu_1, \mu_N) \\
h_L(\mu_2) & h(\mu_2, j_1) & h(\mu_1, j_2) & \ldots h(\mu_2, j_N) & h(\mu_1, \mu_2) & \delta \mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{2}}) & m(\mu_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N, j_1) & h(\mu_N, j_2) & \ldots h(\mu_N, j_N) & h(\mu_N, \mu_1) & h(\mu_N, \mu_2) & \delta \mathbf{e}_{\mathbf{T}}(\mu_{\mathbf{N}})
\end{pmatrix}
$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

First column is a function of particle rapidities, scaling them to range 0-1

$$h_L(i_n) \;=\; C(\cosh(y) - 1),$$

$$\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j_1}) & m(j_1, j_2) & \ldots m(j_1, j_N) & m(j_1, \mu_1) & m(j_1, \mu_2) & \ldots m(j_1, \mu_N) \\
h_L(j_2) & h(j_1, j_2) & \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_2}) & \ldots m(j_2, j_N) & m(j_2, \mu_1) & m(j_2, \mu_2) & \ldots m(j_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1, j_N) & \ldots & \ldots \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_N}) & m(j_N, \mu_1) & m(j_N, \mu_2) & \ldots m(j_N, \mu_N) \\
h_L(\mu_1) & h(\mu_1, j_1) & h(\mu_1, j_2) & \ldots h(\mu_1, j_N) & \mathbf{e}_{\mathbf{T}}(\boldsymbol{\mu_1}) & m(\mu_1, \mu_2) & m(\mu_1, \mu_N) \\
h_L(\mu_2) & h(\mu_2, j_1) & h(\mu_1, j_2) & \ldots h(\mu_2, j_N) & h(\mu_1, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\boldsymbol{\mu_2}) & m(\mu_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N, j_1) & h(\mu_N, j_2) & \ldots h(\mu_N, j_N) & h(\mu_N, \mu_1) & h(\mu_N, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\boldsymbol{\mu_N})
\end{pmatrix}$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – fixed size

The Rapidity Mass Matrix (RMM) is a structure that encodes particle properties and two-particle correlations into a fixed-size matrix suitable for ML

Other elements below the diagonal are functions of two-particle Δ-rapidity

$$h(i_n, j_k) \;=\; C(\cosh(\Delta y/2) \;-\; 1)$$

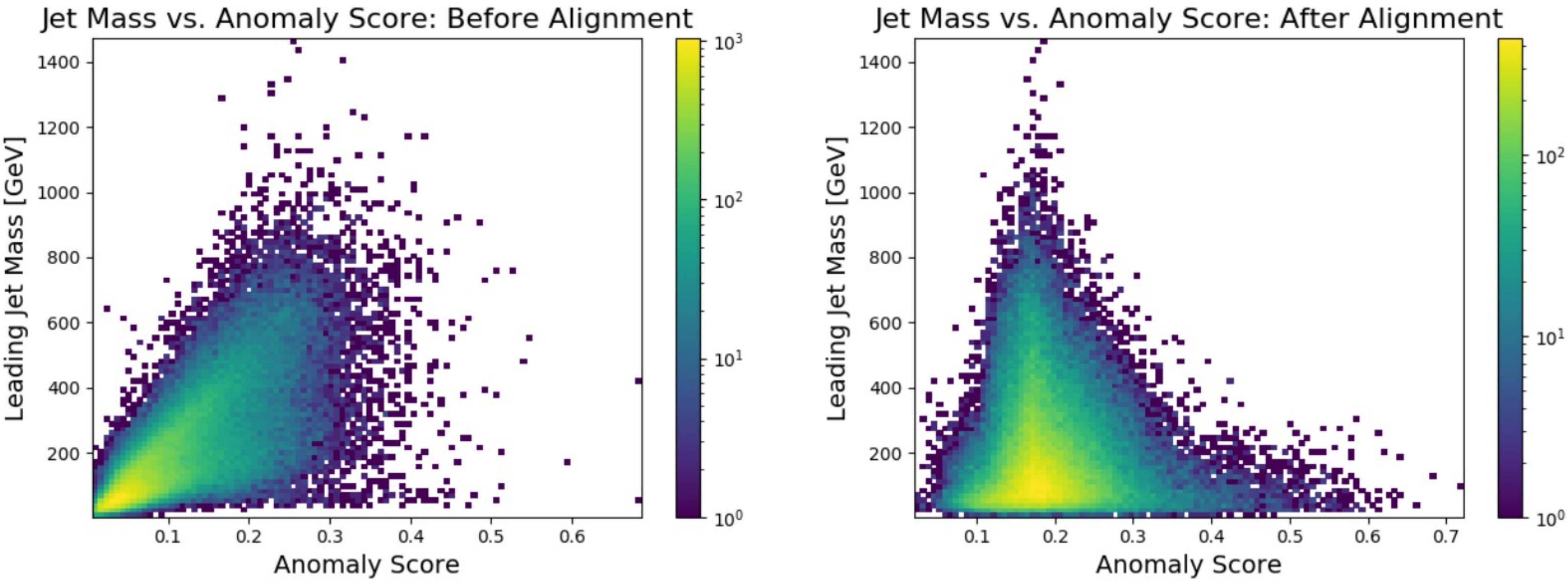
$$\begin{pmatrix}
\mathbf{e}_{\mathbf{T}}^{\mathbf{miss}} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & \mathbf{e}_{\mathbf{T}}(\mathbf{j_1}) & m(j_1, j_2) & \ldots m(j_1, j_N) & m(j_1, \mu_1) & m(j_1, \mu_2) & \ldots m(j_1, \mu_N) \\
h_L(j_2) & h(j_1, j_2) & \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_2}) & \ldots m(j_2, j_N) & m(j_2, \mu_1) & m(j_2, \mu_2) & \ldots m(j_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1, j_N) & \ldots & \ldots \delta\mathbf{e}_{\mathbf{T}}(\mathbf{j_N}) & m(j_N, \mu_1) & m(j_N, \mu_2) & \ldots m(j_N, \mu_N) \\
h_L(\mu_1) & h(\mu_1, j_1) & h(\mu_1, j_2) & \ldots h(\mu_1, j_N) & \mathbf{e}_{\mathbf{T}}(\mu_1) & m(\mu_1, \mu_2) & m(\mu_1, \mu_N) \\
h_L(\mu_2) & h(\mu_2, j_1) & h(\mu_1, j_2) & \ldots h(\mu_2, j_N) & h(\mu_1, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_2) & m(\mu_2, \mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N, j_1) & h(\mu_N, j_2) & \ldots h(\mu_N, j_N) & h(\mu_N, \mu_1) & h(\mu_N, \mu_2) & \delta\mathbf{e}_{\mathbf{T}}(\mu_N)
\end{pmatrix}$$

Nucl.Instrum.Meth.A 931 (2019) 92-99

Encoding features – jets

To avoid simple correlation of anomaly score with jet mass, jets are pre-processed

- Rescale each jet to the same mass
- Boost each jet to the same energy
- Rotate each jet to the same orientation in η, φ



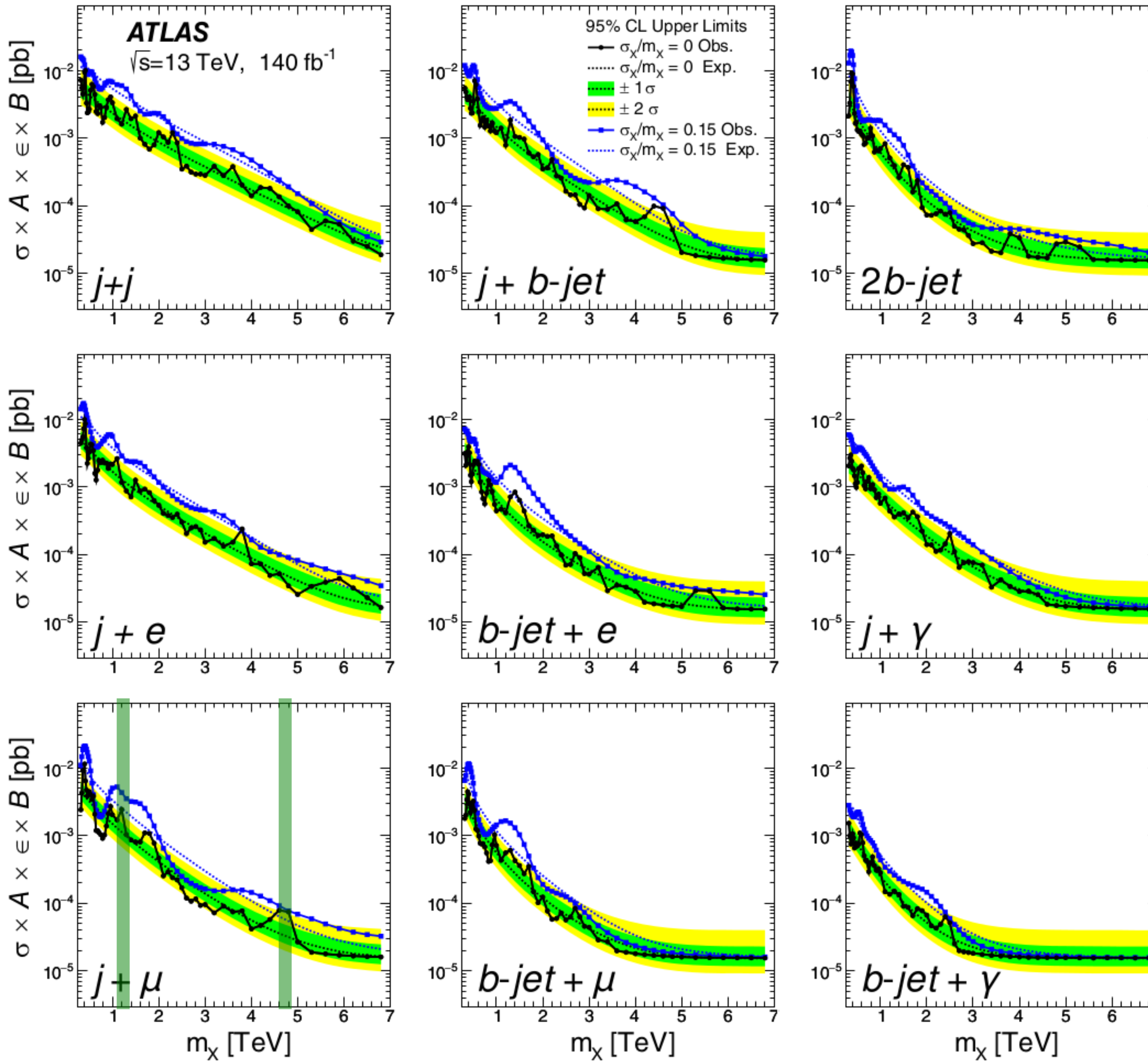
JINST 16 (2021) 08, P08012

RMM for anomaly detection (2024)

Largest excesses in the m(j+μ) distribution at 1.2 and 4.8 TeV

Assuming a zero-width decay at these masses, local significances of 2.8 and 2.9σ

Event at m(j+μ)=8 TeV is removed by more stringent muon ID



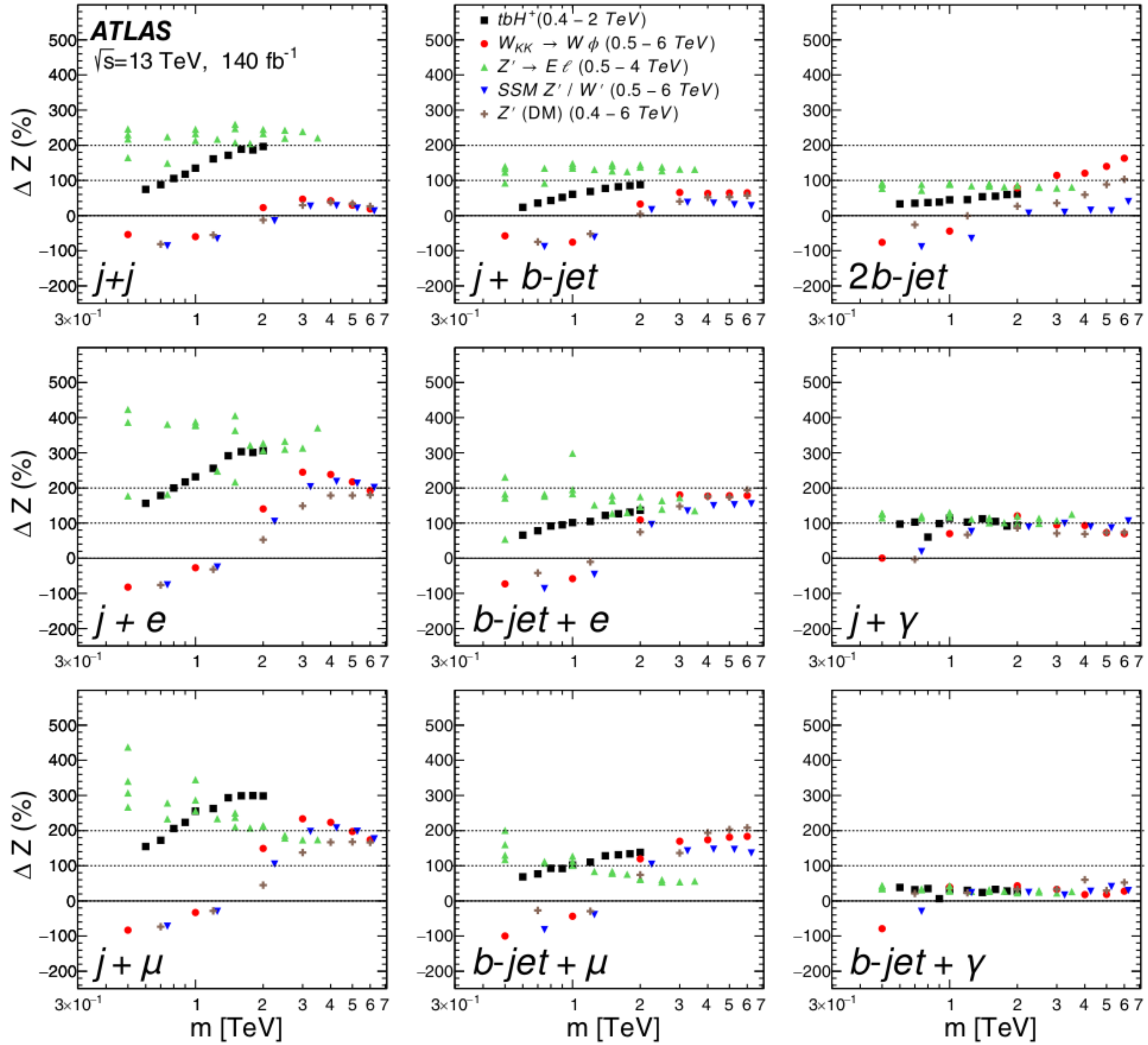
Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

Improvements in sensitivity for the different example signals are shown for the 10pb anomaly score selection, compared to no anomaly score cut

Calculated by counting events within 1σ of BSM particle mass

Lower masses and more SM-like final states give worse performance, as you might expect



Phys.Rev.Lett. 132 (2024) 8, 081801

RMM for anomaly detection (2024)

(1) charged Higgs boson production in association with a top quark, tbH^+ with $H^+ \rightarrow t\bar{b}$, with the mass of H^+ between 0.4 TeV and 2 TeV with a varying step size. All top decays are enabled;
(2) a Kaluza-Klein gauge boson, W_{KK}, with the SM W boson and a radion ϕ that decays into two gluons, with the mass of W_{KK} between 0.5 TeV and 6 TeV, and the mass difference of W_{KK} and the radion being 0.25 TeV;
(3) a model with a Z0 boson and composite SU(2)L fermion doublets that breaks lepton-flavor universality ("composite lepton"), Z0 → El, with the Z0 boson mass of 0.5–4 TeV and various mass hypotheses for the composite lepton E decaying to Zl with $Z \rightarrow q\bar{q}$ [50];
(4) the sequential standard model (SSM) W' → WZ' → $l\nu q\bar{q}$, with the mass of the W' boson ranging from 0.7 TeV to 6.2 TeV with a varying step size and the mass difference of the W' and Z' bosons being 0.25 TeV;
(5) a simplified dark-matter (DM) model Z' → $q\bar{q}$, with an axial-vector mediator Z' boson whose mass ranges between 0.5 TeV and 6 TeV with a varying step size, where one of the quarks radiates a W boson which decays into lν

- ■ $tbH^+ (0.4 - 2\ TeV)$
- ● $W_{KK} \rightarrow W\phi\ (0.5 - 6\ TeV)$
- ▲ $Z' \rightarrow E\ell\ (0.5 - 4\ TeV)$
- ▼ $SSM\ Z'/W'\ (0.5 - 6\ TeV)$
- ✚ Z' (DM) $(0.4 - 6\ TeV)$

Phys.Rev.Lett. 132 (2024) 8, 081801