ATLAS EXPERIMENT

INFN Istituto Nazionale di Fisica Nucleare

ML4Jets
Vienna

14 - 18 September 2026

Unsupervised Machine Learning for Jet-Based Anomaly Detection in ATLAS

Francesco Cirotto, on behalf of the ATLAS Collaboration

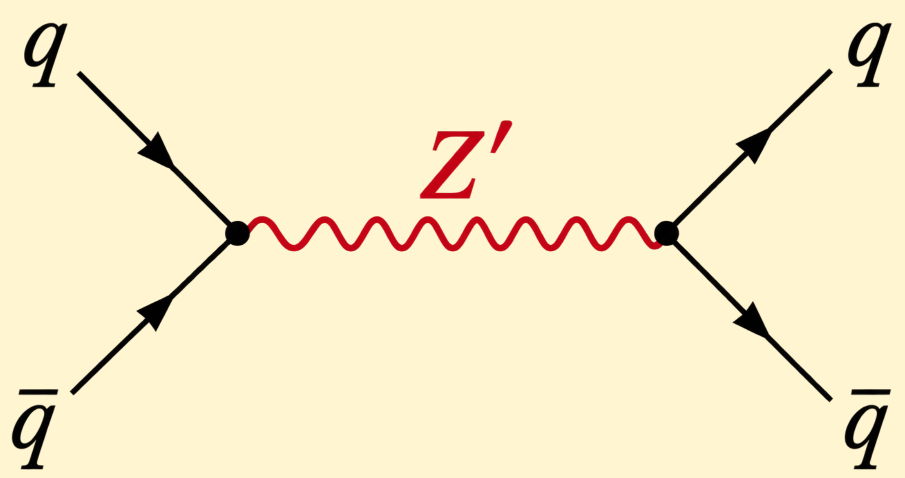
Università degli studi di Napoli Federico II & INFN

Why Anomaly Detection?

A complementary way to search for new physics with fewer assumptions on the signal

Model-dependent approach

Does the data contain **this signal**?



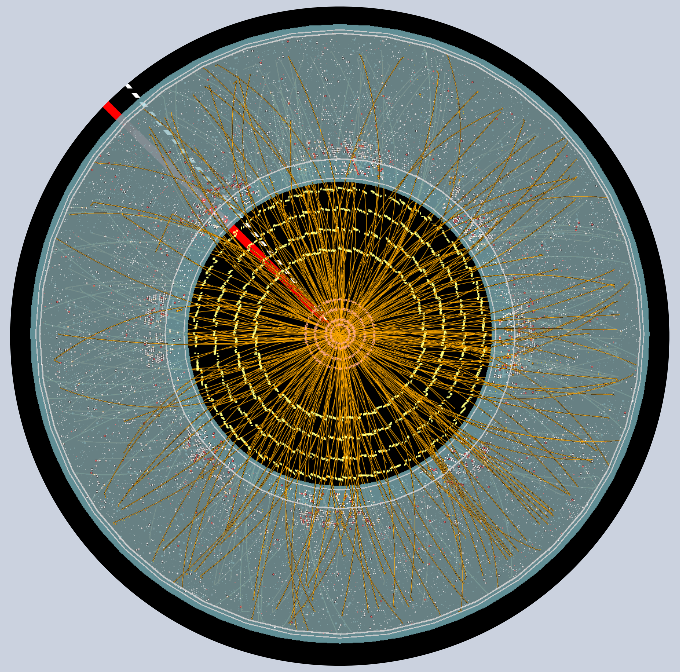
Signal hypothesis
(Model, masses, couplings…)

Signal-optimized analysis
(Cuts, variables, dedicated classifiers)

Statistical inference
(Discovery or limits)

Model-independent approach

Does the data contain **something unusual**?



Background-dominated data
(Minimal assumptions on signal)

Learn what is typical
(Unsupervised learning)

Search for deviations
(Anomaly score, statistical test)

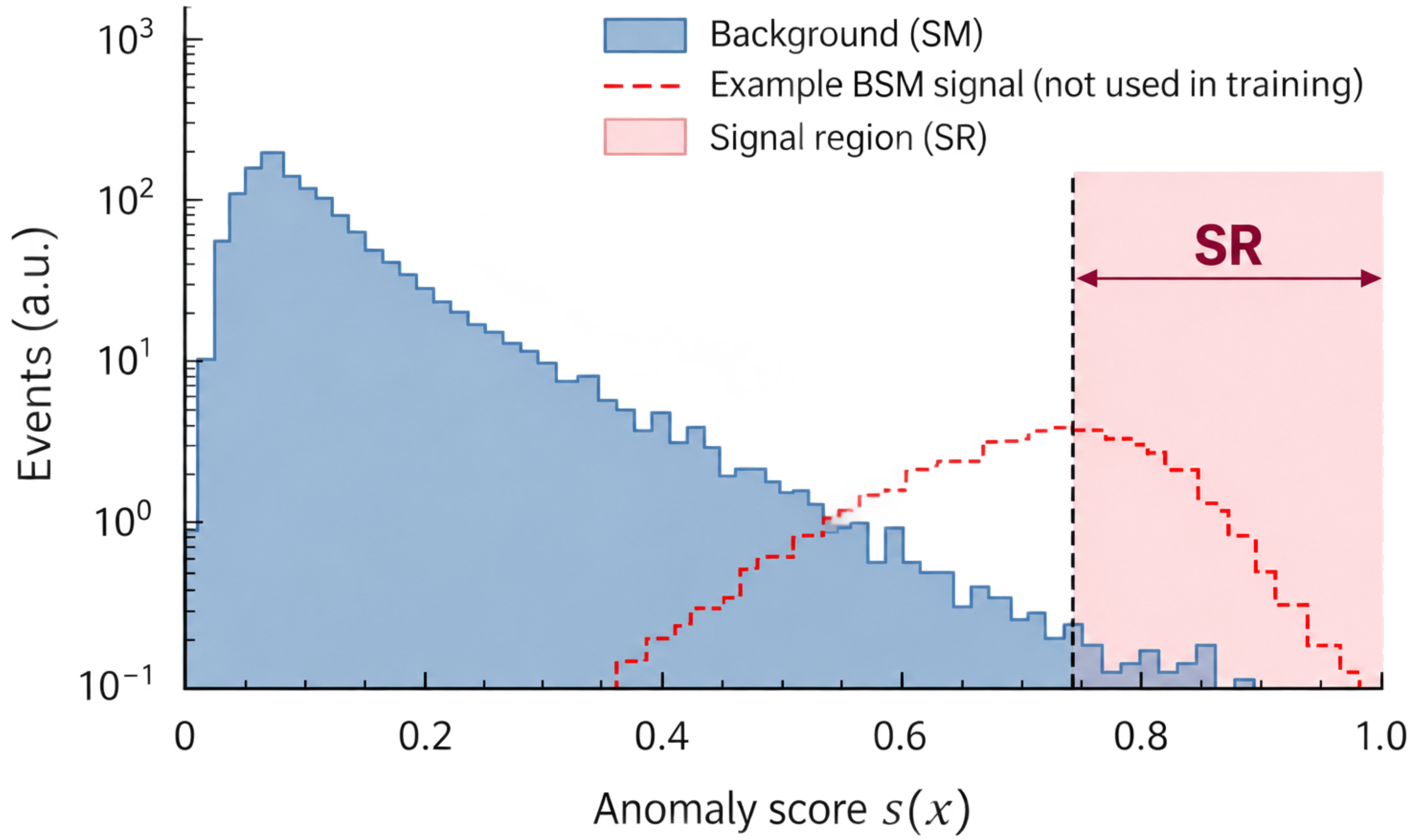
Unsupervised Anomaly Detection

Learning the background to identify deviations from the expected

Train ML models to encode the properties of SM physics and recognize unusual events

Typical events (low score)

Unusual events (high score)



Anomalous $\neq$ New Physics

Detector effects, rare SM processes or mismodelling can also look anomalous

Unsupervised $\neq$ Assumption-free

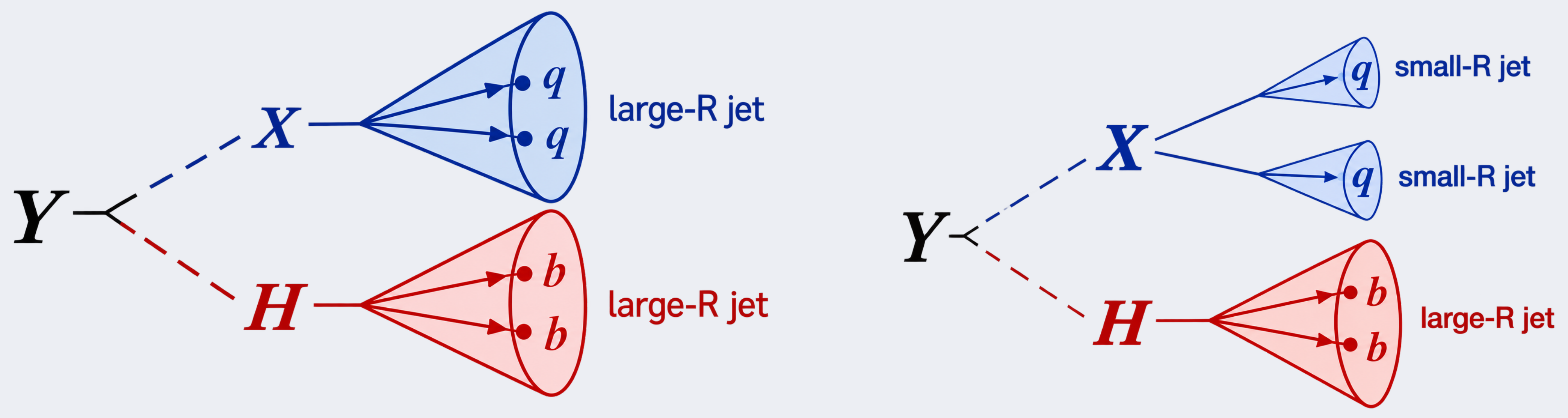
The definition of anomaly depends on the input representation, preprocessing, architecture and training sample

Anomalous regions must be followed by background estimation and statistical inference

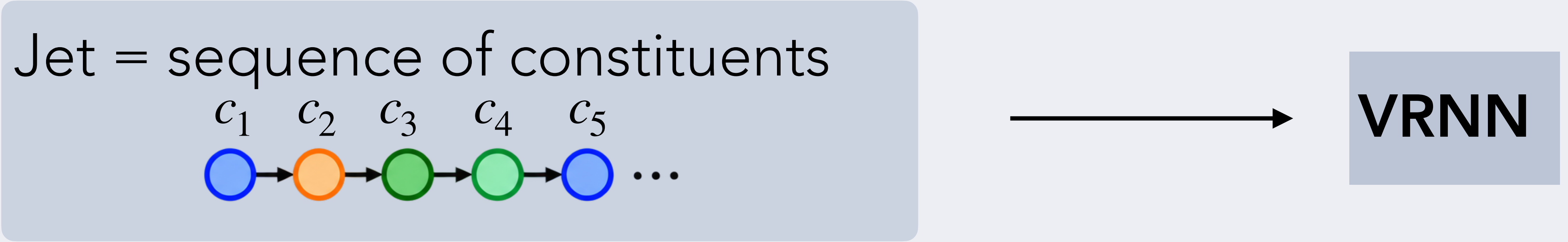
ANOMALY DETECTION IN ATLAS $Y \to XH$ SEARCH

Learning anomalies from sequences of jet constituents

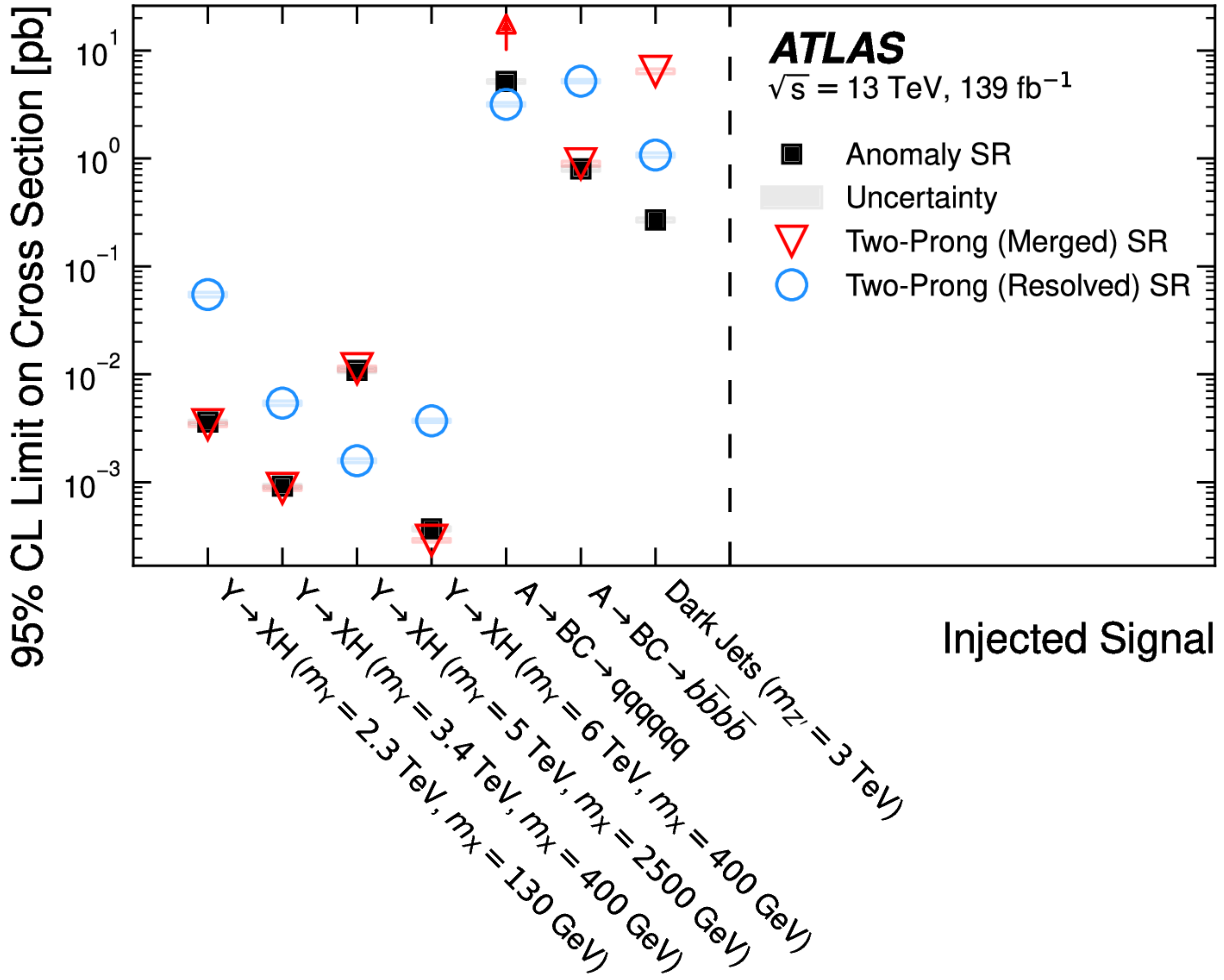
- Search for a heavy-mass resonance Y decaying in a Higgs boson ($H \to b\bar{b}$) and a new particle X in the fully hadronic channel
- First ATLAS publication with a fully unsupervised AD approach based on a Variational Recurrent Neural Network (VRNN)
- Anomaly Score computed from VRNN output to select X boson



Fully hadronic final state in boosted/resolved regime
X jet as anomaly candidate



Trained on data-only over constituents calo cells, modeled as a sequence of four-vectors of jets with p_T> 1.2 TeV

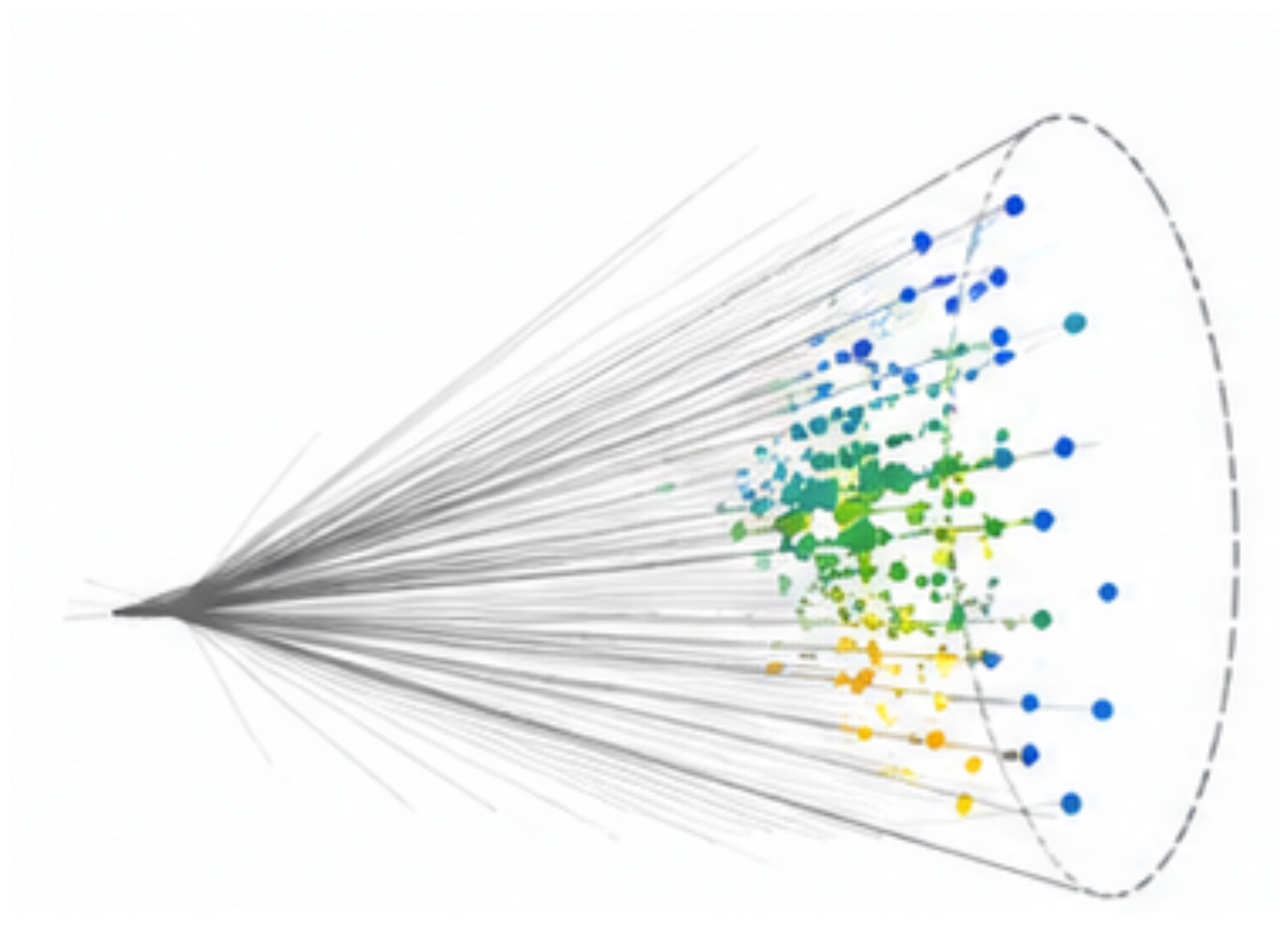


Competitive on targeted signatures

Phys. Rev. D 108 (2023) 052009

WHY CONSTITUENT-LEVEL INFORMATION?

Unlocking the Full Potential of Detector Data



A jet is a complex, multi-particle object

High-level observables

e.g. $m_J, p_T, \tau_{21}, \ldots$

- Compact and low-dimensional
- Physics-motivated and interpretable

Powerful for known signatures, but may miss unexpected topologies that do not manifest in standard observables

Low-level observables

e.g. jet constituents, hits, tracks

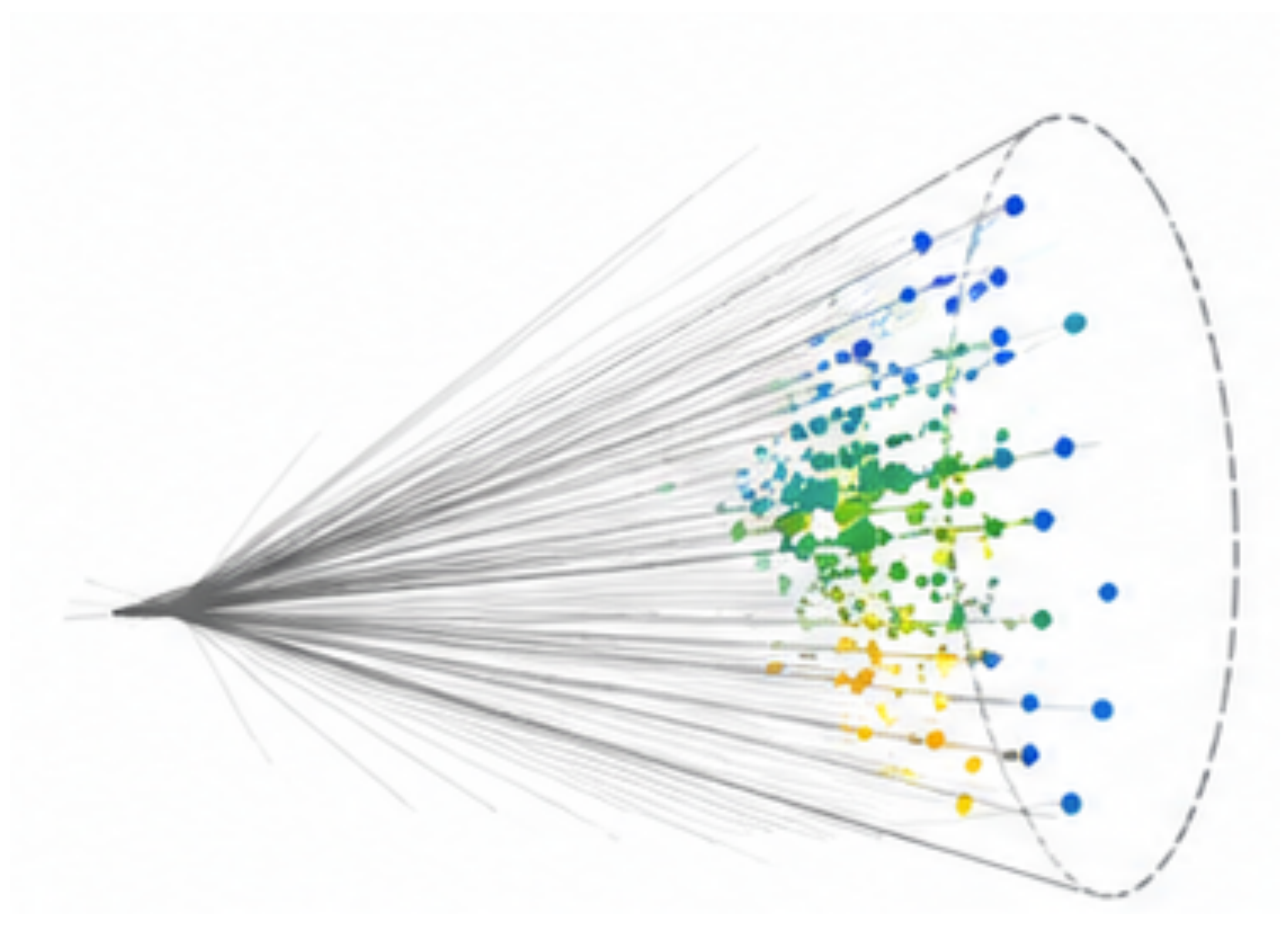
- preserve the raw structure of the event
- Allow ML to learn directly from detector-level patterns

Retains maximal information from the detector

In this talk: jet constituents as the starting point for anomaly detection

HOW SHOULD WE REPRESENT A JET?

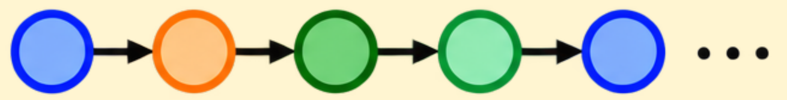
Same constituents, different organizations



A jet is a complex, multi-particle object

Sequence representation

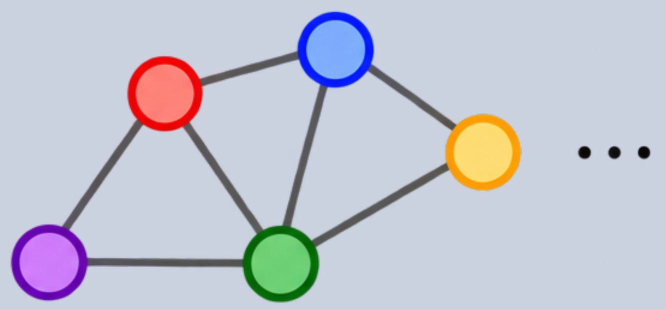
Ordered list of constituents



- Ordering is explicit
- Sequential processing
- Used in the $Y \rightarrow XH$ search

Graph representation

Constituents as nodes with geometrical relations



- Explicitly exploits local structures
- Graph representation: nodes (entities) and edges (connections)

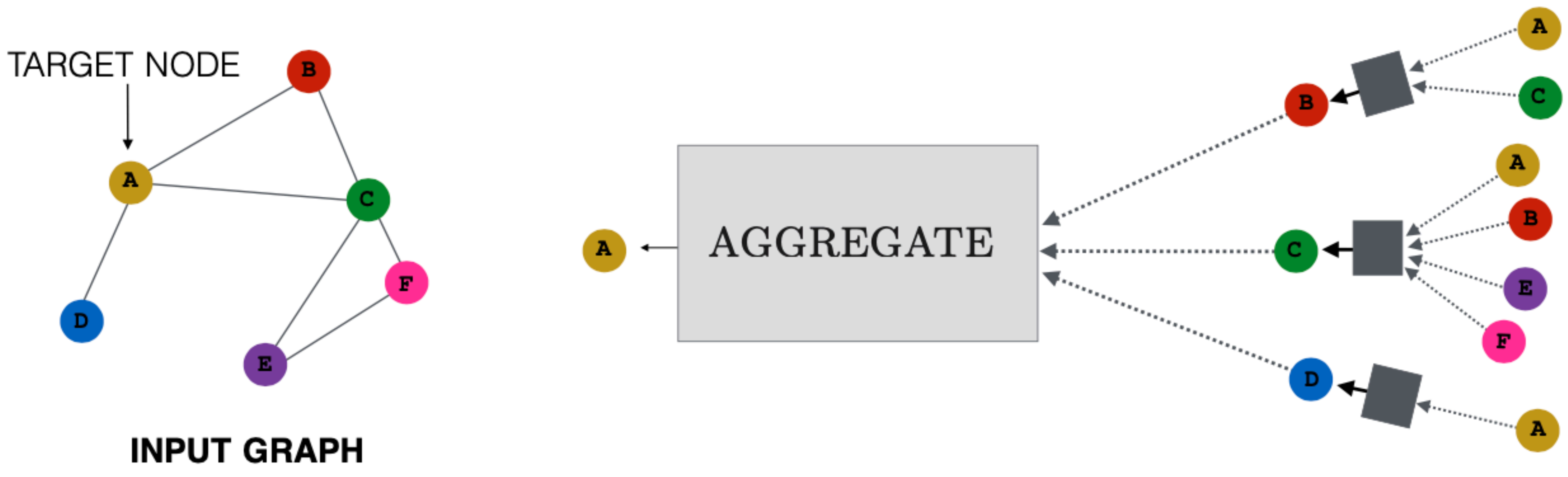
Graph Anomaly Detection for New Physics Searches

Extract Information: message passing

- GNN transforms the structure of the jet into a learned embedding that captures its relevant substructure

- Message passing: vector messages are exchanged between connected nodes and processed by neural networks

- At each layer, each node aggregates information from its local neighbourhood
  - ↪At the end each node embedding contains more and more information from further reaches of the graph.
  - ↪Node embeddings encode both constituent features and graph structure



GRAPHS ARE NEW JETS

Features

What is a node?

↪Jet constituents

↪Fraction of jet pT, eta, phi

What is an edge?

↪Weight message from neighboring nodes

↪Using ΔR "distance" between jets

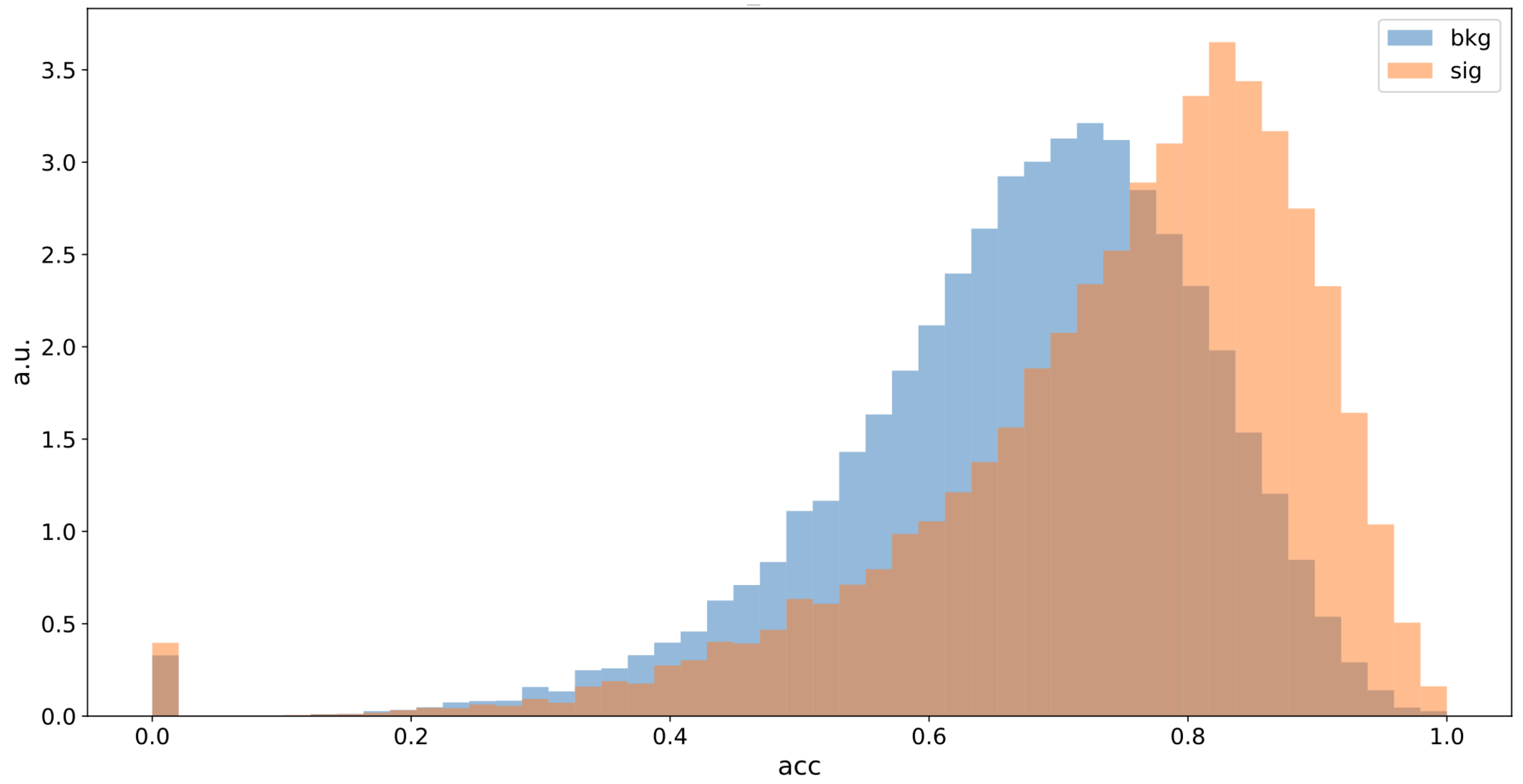
How are they connected?

↪No self-loops, ΔRcut = 0.6

Transformation/Preprocessing

Data augmented for mass decorrelation

- Transform constituent four-vectors
- Prevents the network for trivially learning jet mass
- Focus on substructure



Average Clustering Coefficient: Measure of the degree to which nodes tend to cluster together

Transformation

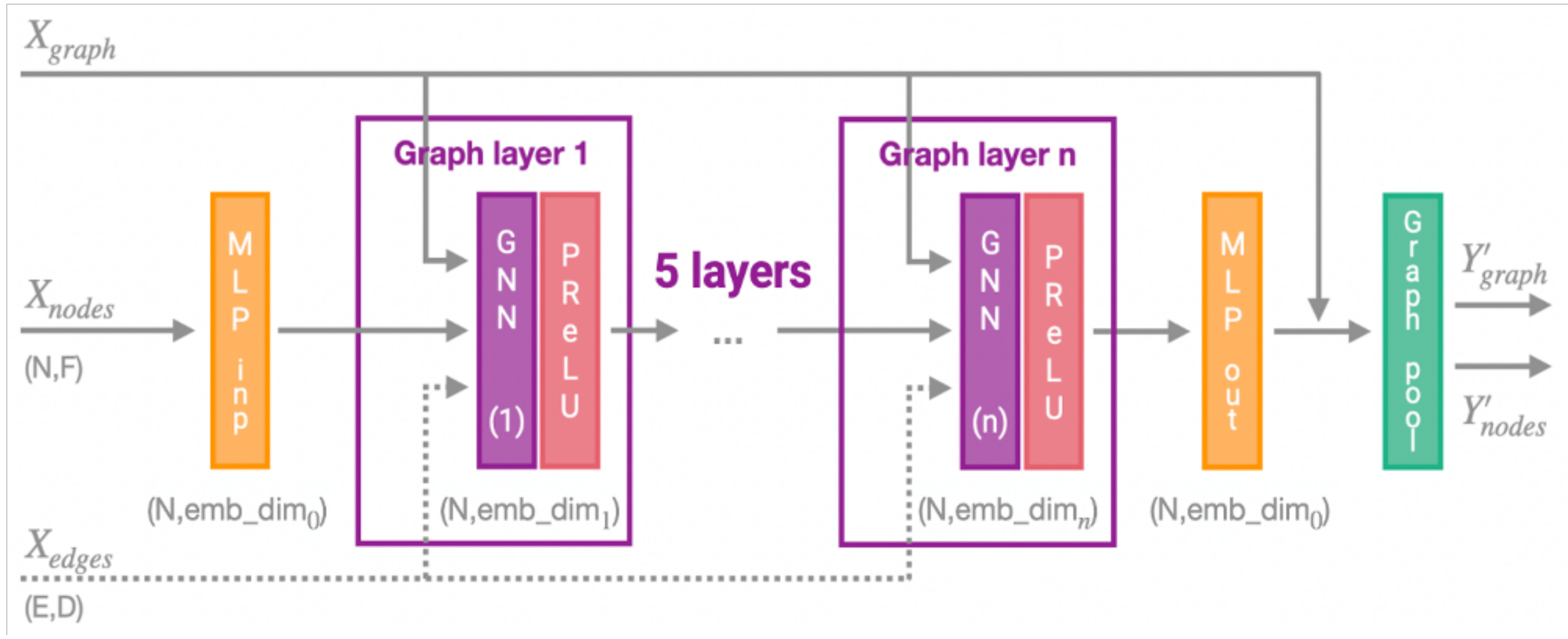
LEARNING ANOMALIES FROM JET GRAPHS

From local structure to a learned representation

EGAT

Edge Graph Attention Network: GNN with attention and edge-feature updates, learning which constituent interactions are most relevant

- Local message passing between connected constituents
- Geometrical information encoded in the edge weight: $e^{-\Delta R}$
- Attention uses both node and edge features
- Node and edge representations are updated layer by layer



N: # of nodes
F: # of node features
E: # of edges
D: # of edge features

EGAT maps the constituent graph into a learned jet representation

UNSUPERVISED ANOMALY LEARNING

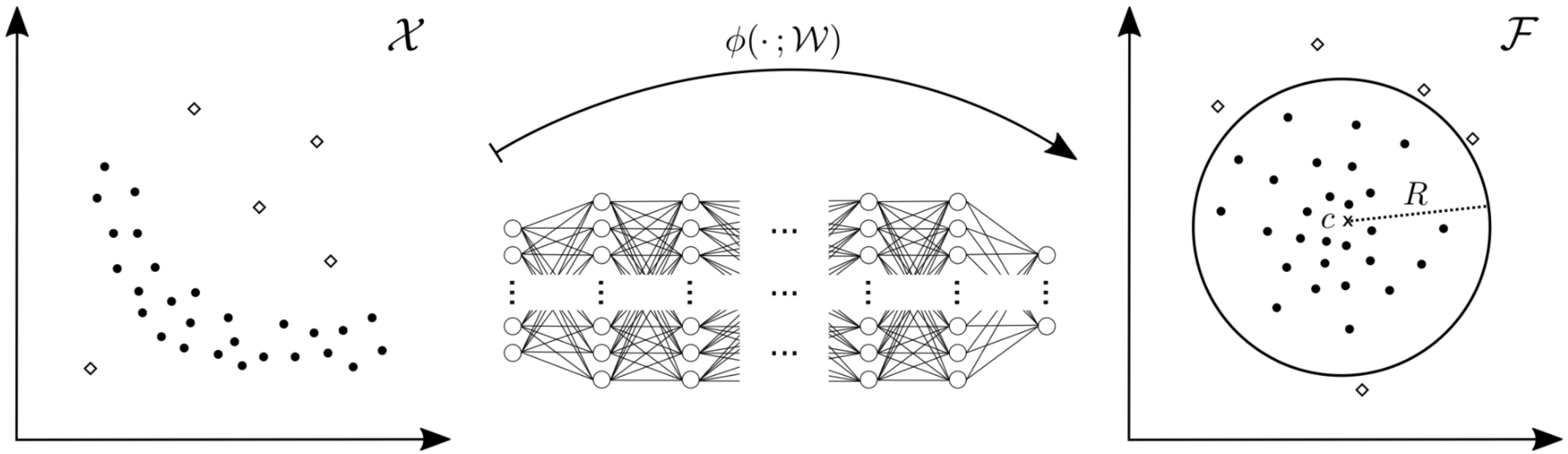
Learning the background, not the signal

Unsupervised training on background-dominated data

- No signal labels or signal model enter the training
- Deep SVDD learns a neural network transformation
- Attempts to map most of the data network representations into a hypersphere
- Mappings of normal examples fall within, whereas mappings of anomalies fall outside the hypersphere

Anomaly Score

- Distance of the learned jet representation from c



$$s(\boldsymbol{x}) = \|\phi(\boldsymbol{x};\mathcal{W}^*) - \mathbf{c}\|^2$$

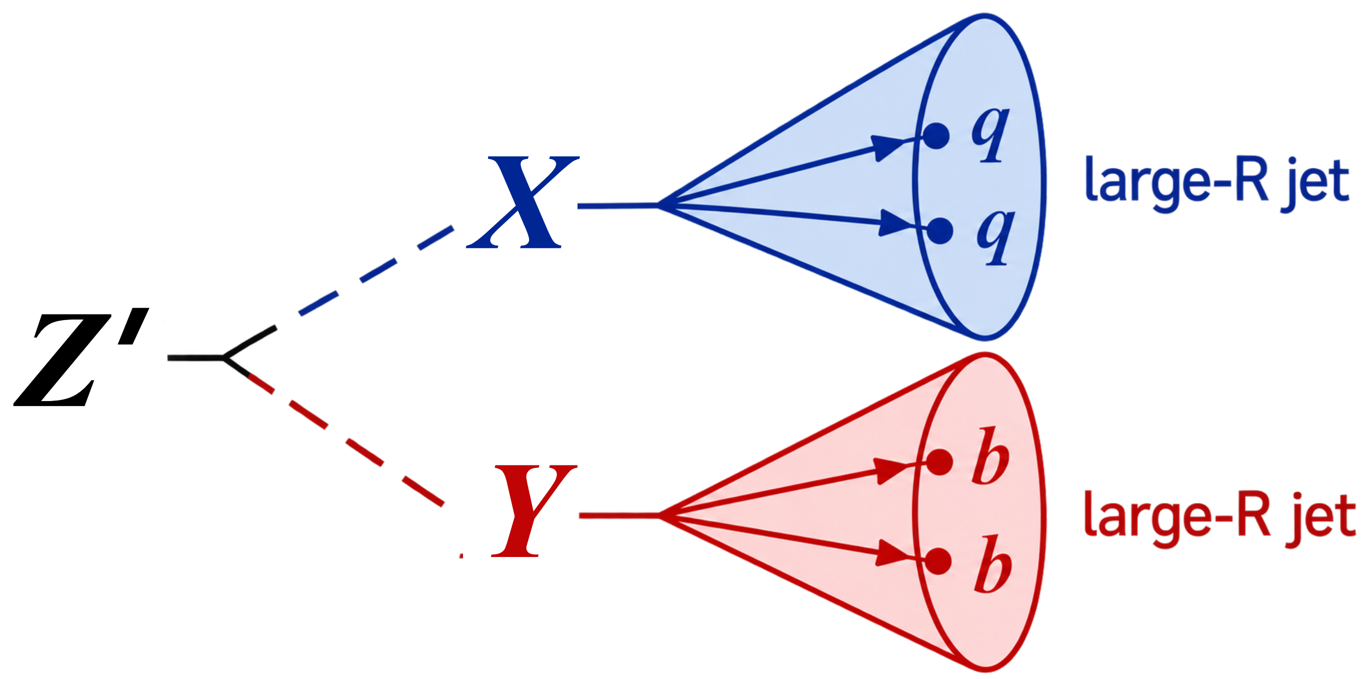
Deep SVDD

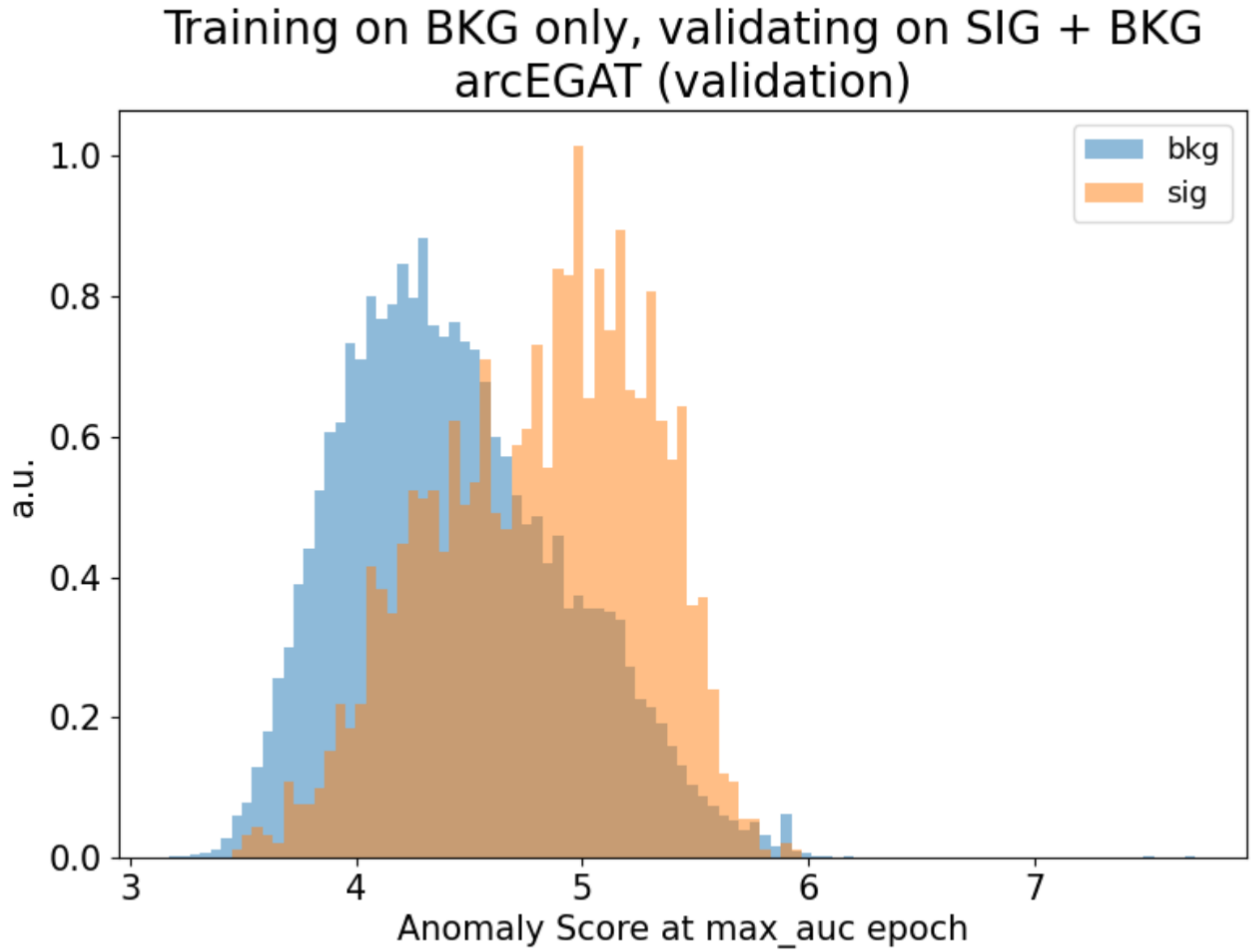
R&D RESULTS ON OPEN DATA

The LHC Olympics Dataset

LHC 2020 Olympics Dataset: a public dataset

- Background: QCD dijets
- Signal: $Z' \rightarrow XY \rightarrow qqqq$, $m_{Z'} = 3.5\,TeV$, $m_X = 500\,GeV$, $m_Y = 100\,GeV$
- Background only-training





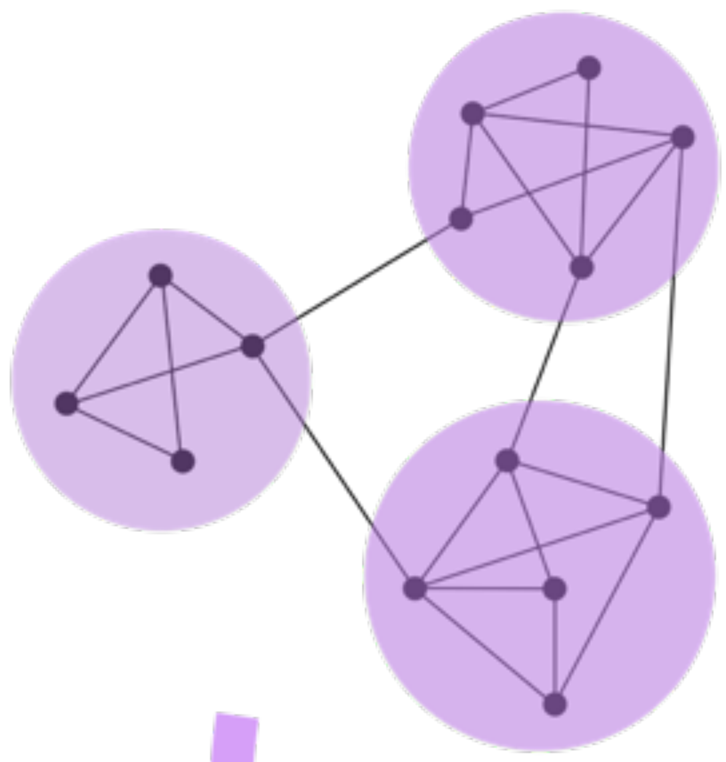
[LHC Olympics Dataset]

THE SEARCH IN ATLAS

From benchmarks to a physics search

Event Selection

- Fully hadronic final state with two boosted large-R jets
- High mass resonance search in ATLAS with data collected during LHC Run 3 period (2022-2026)
- 2 large-R jets selected per event
- Use jet constituents as input for the anomaly tagger

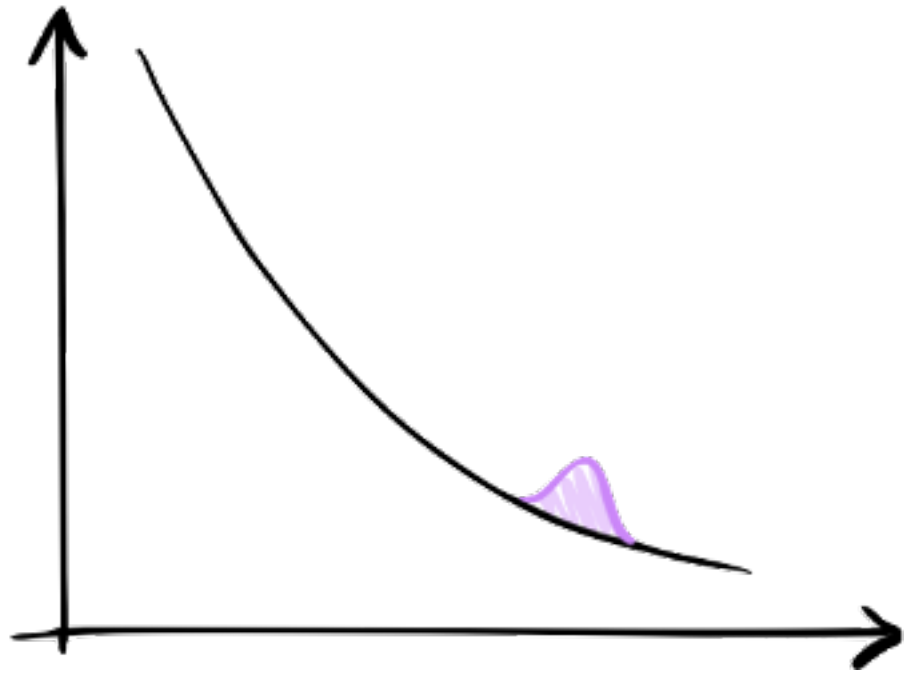


Analysis Strategy

- Compute an anomaly score for each large-R jet
- Define anomaly-enriched regions based on jet scores
- Use data driven methods to estimate background

Physics Search

- Look for deviations in the dijet spectrum
- Search for deviations from the background estimation
- Set limits on a benchmark signal pool



Background Estimation with ML

Making search robust

Signal Region definition

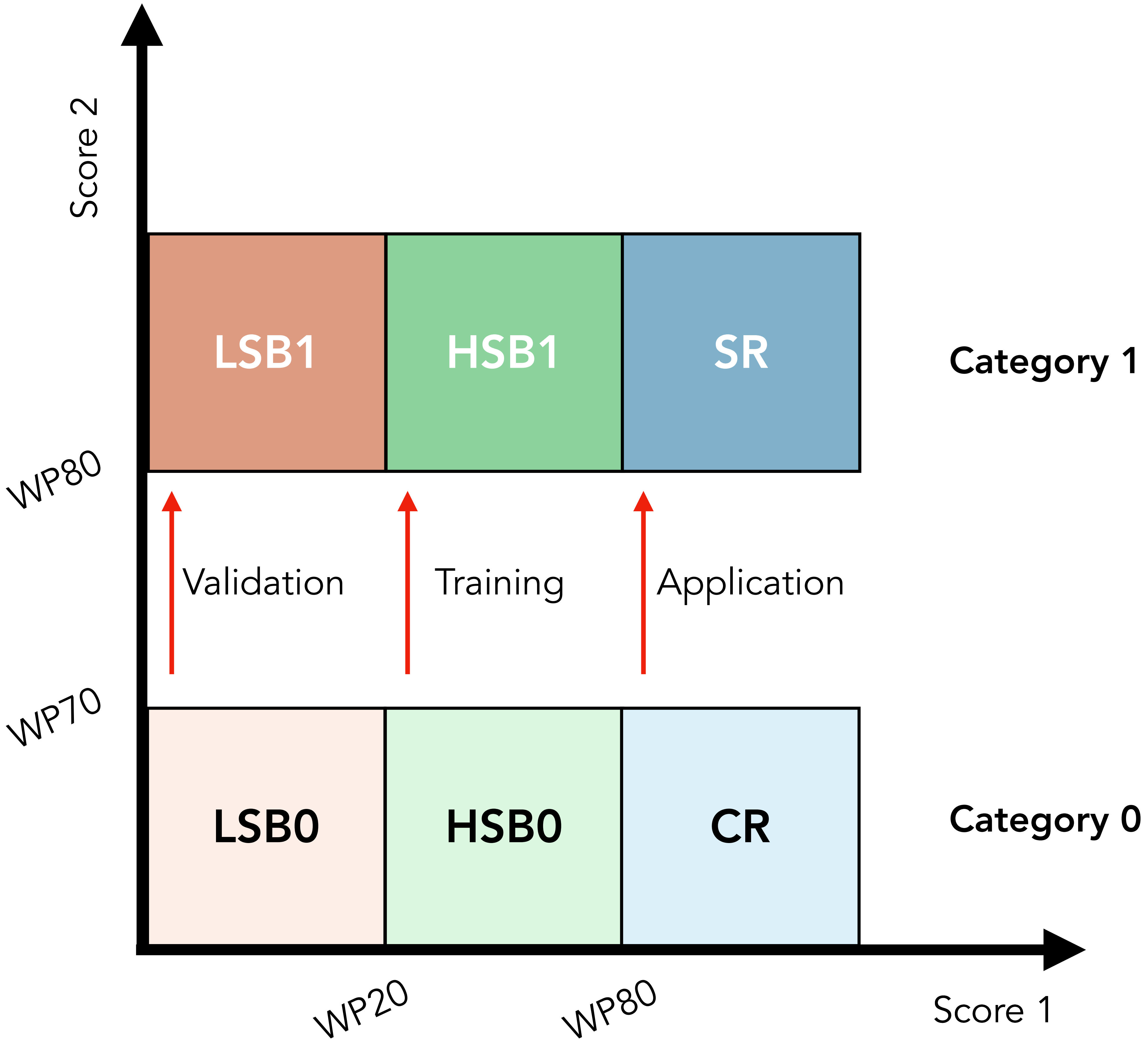
- Defined on a 2D plane built from two AD tagger scores
- Working Point (WP): defined on a fixed background rejection
  - ↪No signal used to optimize selection
  - ↪e.g. WP80: Cut on AS that reject 80% of background events

Background estimation

- Use event in category 0 to predict background in category 1through a data-driven transfer learned directly from data
- Likelihood ratio r(X) between two category learnt in High Sidebands by Direct Importance Estimation method

$$p_1(X) = r(X) \cdot p_0(X) \rightarrow r(X) = \frac{p_1(X)}{p_0(X)}$$

- Provides a fully data-driven multivariate background reweighting approach



DIE Method

TECHNICAL INFRASTRUCTURE (ATLAS)

Scaling graph-based AD to realistic datasets

EGAT training on the INFN Naples IBISCO GPU cluster

ARCHITECTURE

- **8 GPU** computer nodes
- NVIDIA **H100** and TESLA **V100** GPUs
- **100** Gb/s InfiniBand
- Parallel **Lustre** Filesystem
- **SLURM** workload manager
- **AlmaLinux** 9 OS

Training configuration

- Dataset size: 1.1 M events
- Model: Graph Neural Network
- Parameters: ~4 millions
- Training: CPU/GPU

CPU Training	GPU Training	
Time per epoch	Time per epoch	**31x**
6200 s	**200 s**	**speed-up** with GPUs

CONCLUSION

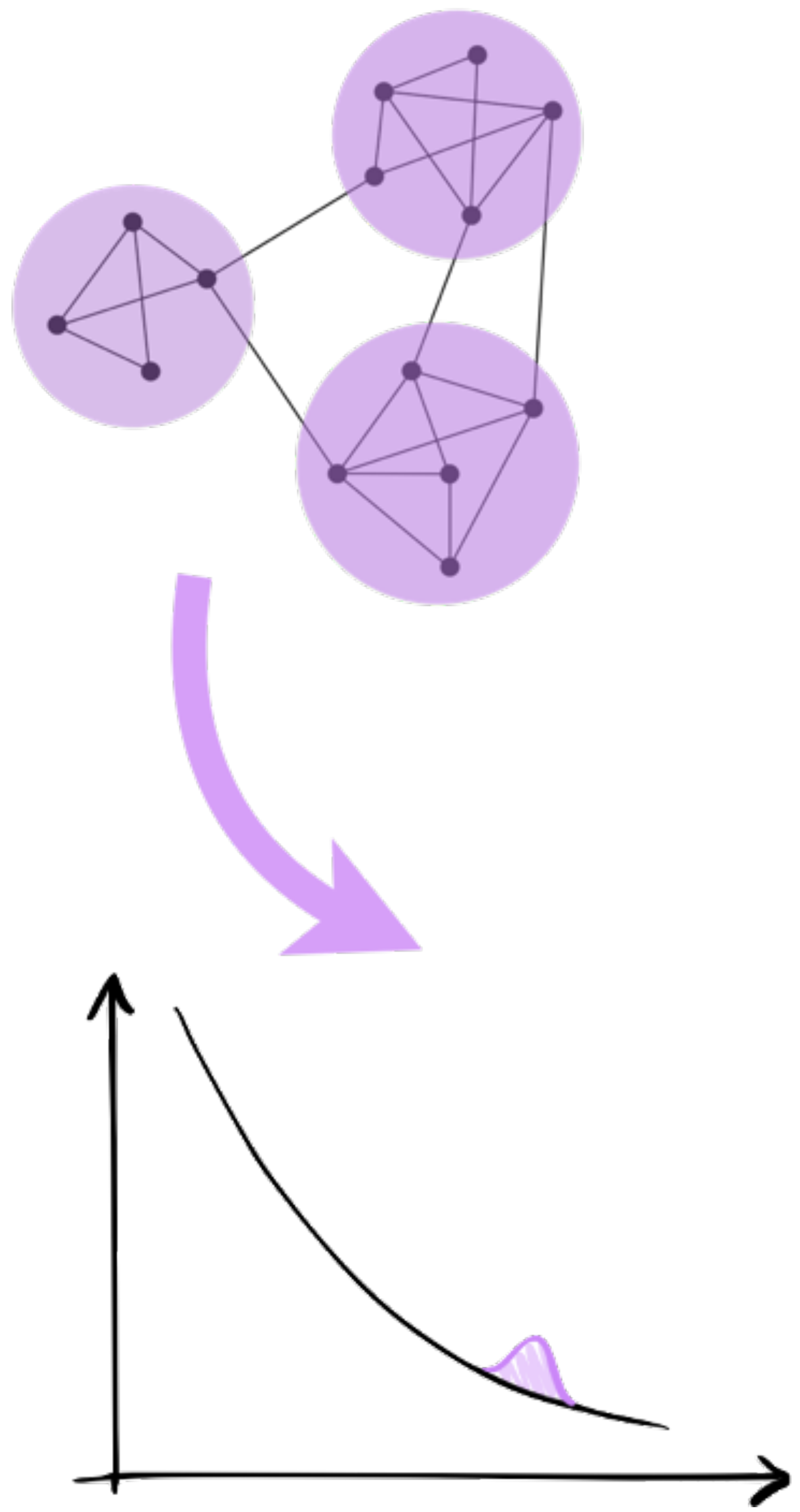
- Anomaly Detection provides a complementary approach to targeted BSM searches, with broad sensitivity to unexpected signatures

- Jet constituents provide a natural space for anomaly detection
  ↪different representations probe different aspects of jet substructure

- Graph-based AD shows promising performance on LHC Olympics
  ↪sensitivity extends across different benchmark topologies

- The next challenge is turning an anomaly tagger into an ATLAS search
  ↪signal-agnostic working points
  ↪data-driven background estimation

- These techniques are now being developed for searches with ATLAS Run-3 data

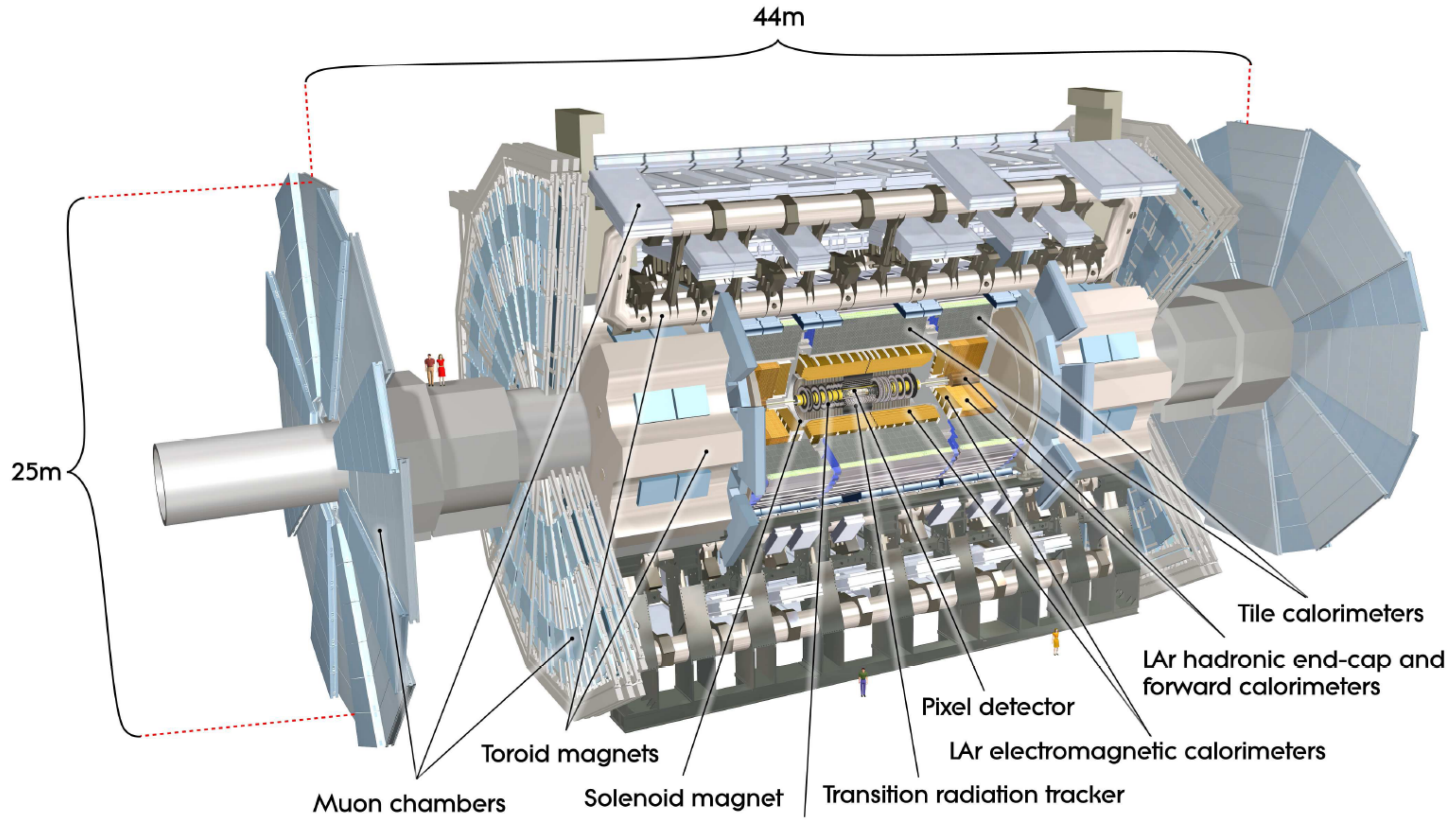


Backup

ATLAS: A TOROIDAL LHC APPARATUS

A multipurpose detector: Higgs and SM physics, BSM searches

- ATLAS is one of the two general-purpose detectors at LHC alongside CMS
- With its 7000-tonnes of weight, it is designed to exploit the full discovery potential of the LHC

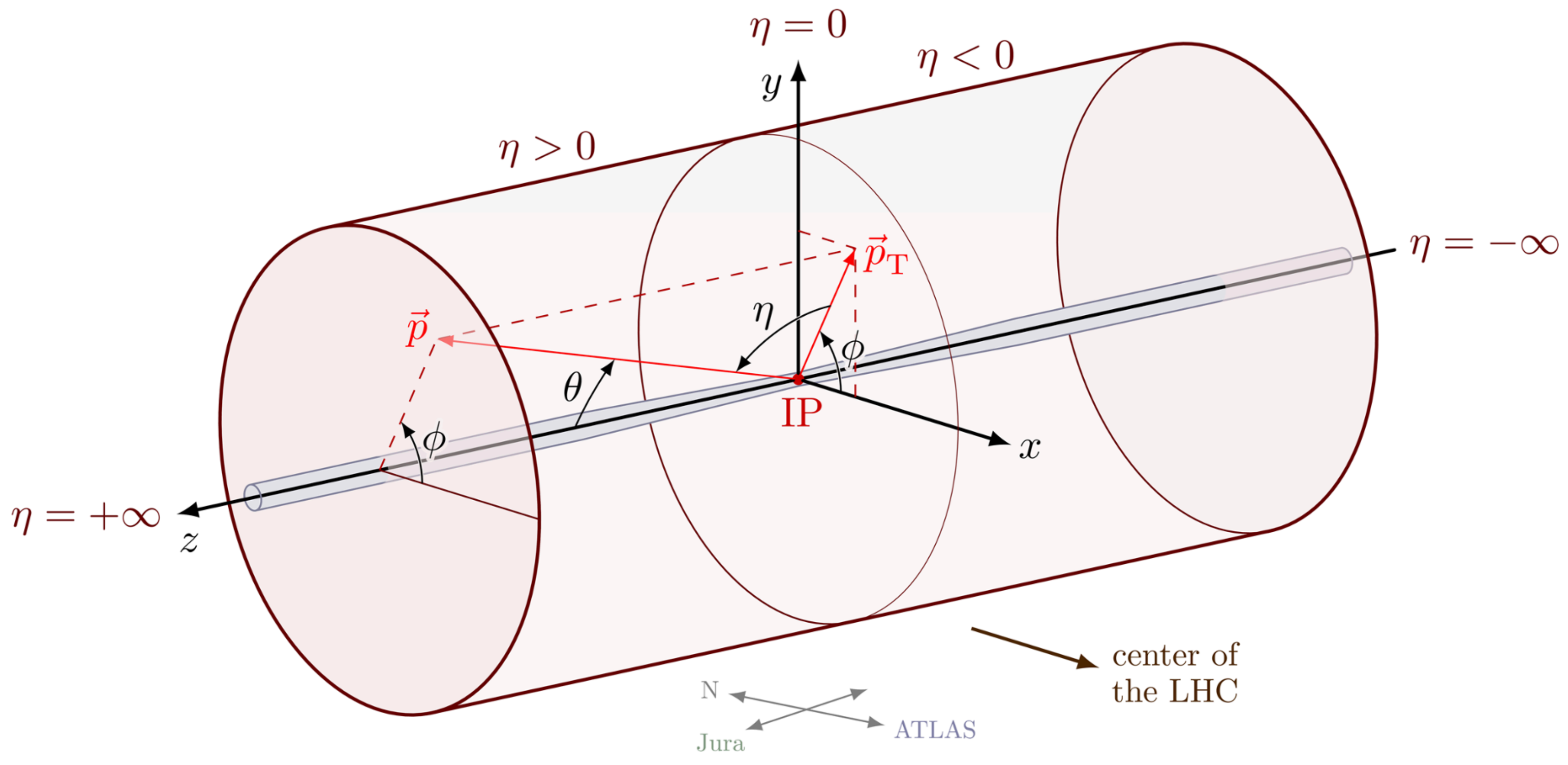


The Lorentz-invariant pseudo-rapidity is used instead of θ:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

Distance in the η-ϕ plane:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

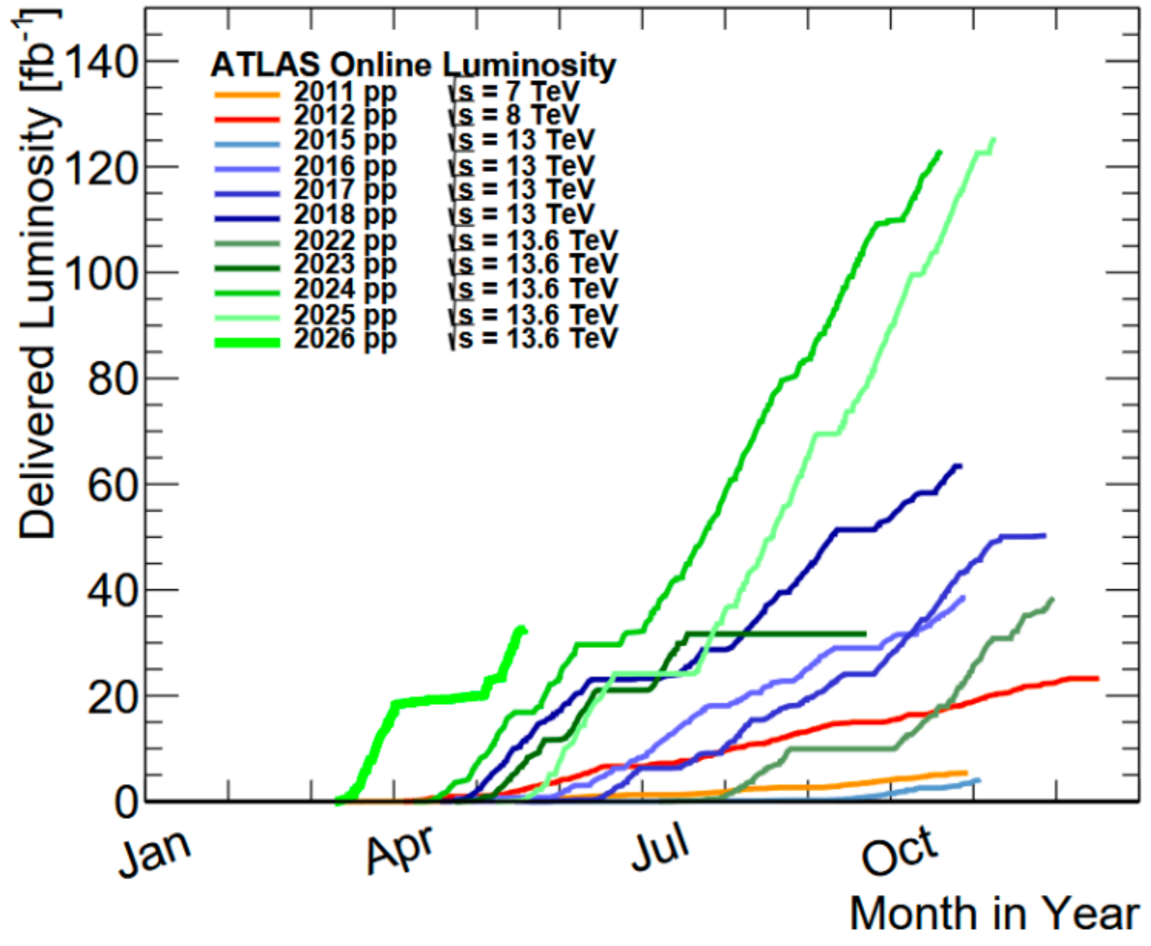


Luminosity recorded:

25 fb^{-1} @ 7,8 TeV

147 fb^{-1} @ 13 TeV

332 fb^{-1} @ 13.6 TeV



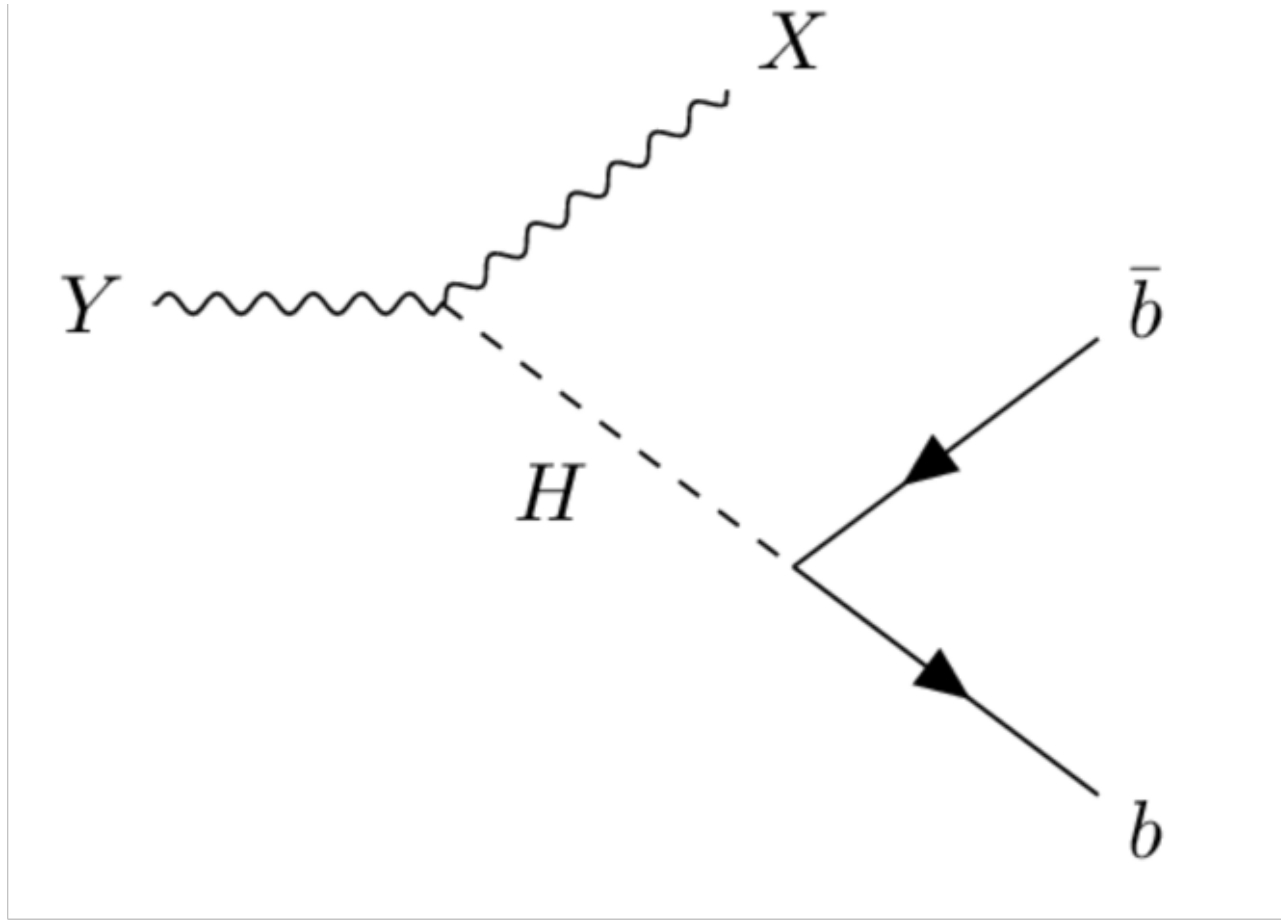
$Y \to XH$

The First Use of Unsupervised Learning on ATLAS Data

- The Y → XH analysis searches for heavy resonances decaying into a Higgs boson and new particle X in a fully hadronic final state [CONF - ICHEP 2022]
  - ↪X and H are highly boosted and their decay products collimated
  - ↪Reconstructed as 2 large-R jets in the final state
  - ↪Fully data driven, inclusive background estimation

- Y → XH analysis developed an unsupervised Variational Recurrent Neural Network as part of the search strategy, used to enhance a bump hunt in a model independent anomaly signal region

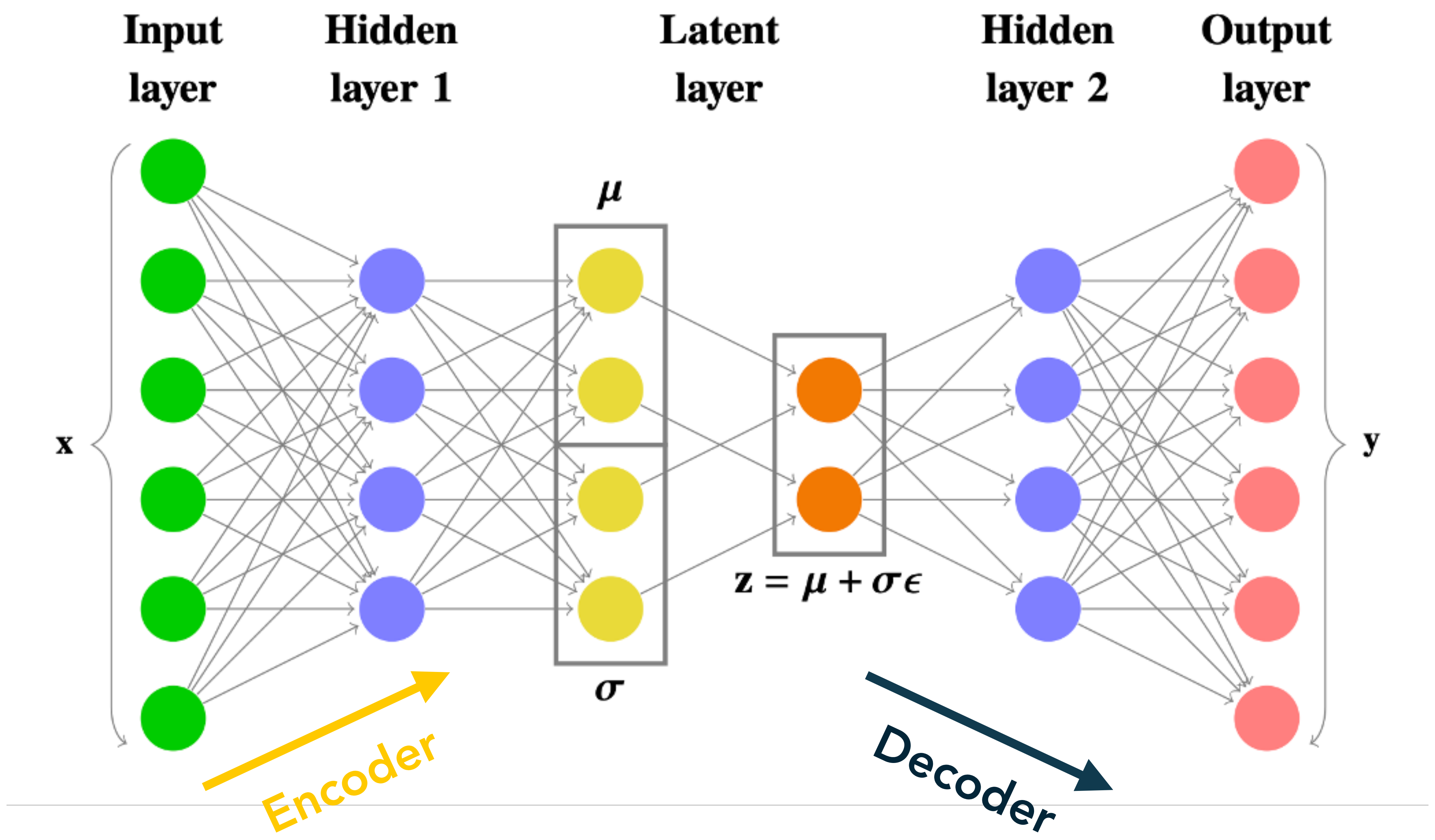
- The VRNN trained over the full Run-2 dataset
  - ↪Leading jet pT > 1.2 TeV; MJJ > 1.3 TeV



VARIATIONAL AUTOENCODERS

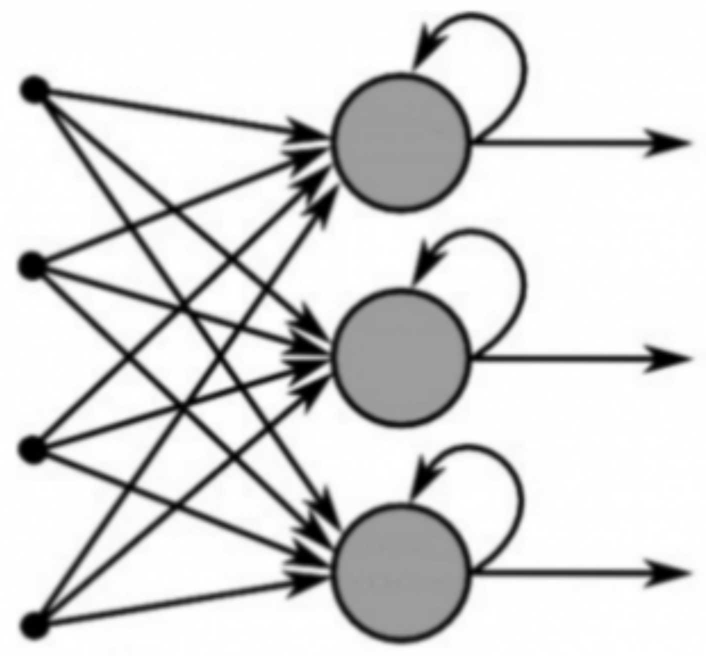
- An autoencoder is a generative model which **encodes** input into lower dimensional latent space to pick out it's most salient features, and then **decodes** from latent space while checking for reconstruction errors
  - ↪Assume that trained AE will be able to reconstruct samples similar to training samples
  - ↪Large reconstruction error when applied to "unseen" datasets

- A variational autoencoder encodes to a probability distribution in the latent space, which allows for Bayesian interference by sampling from this space
  - ↪VAEs offer a number of improvements [An,Cho]

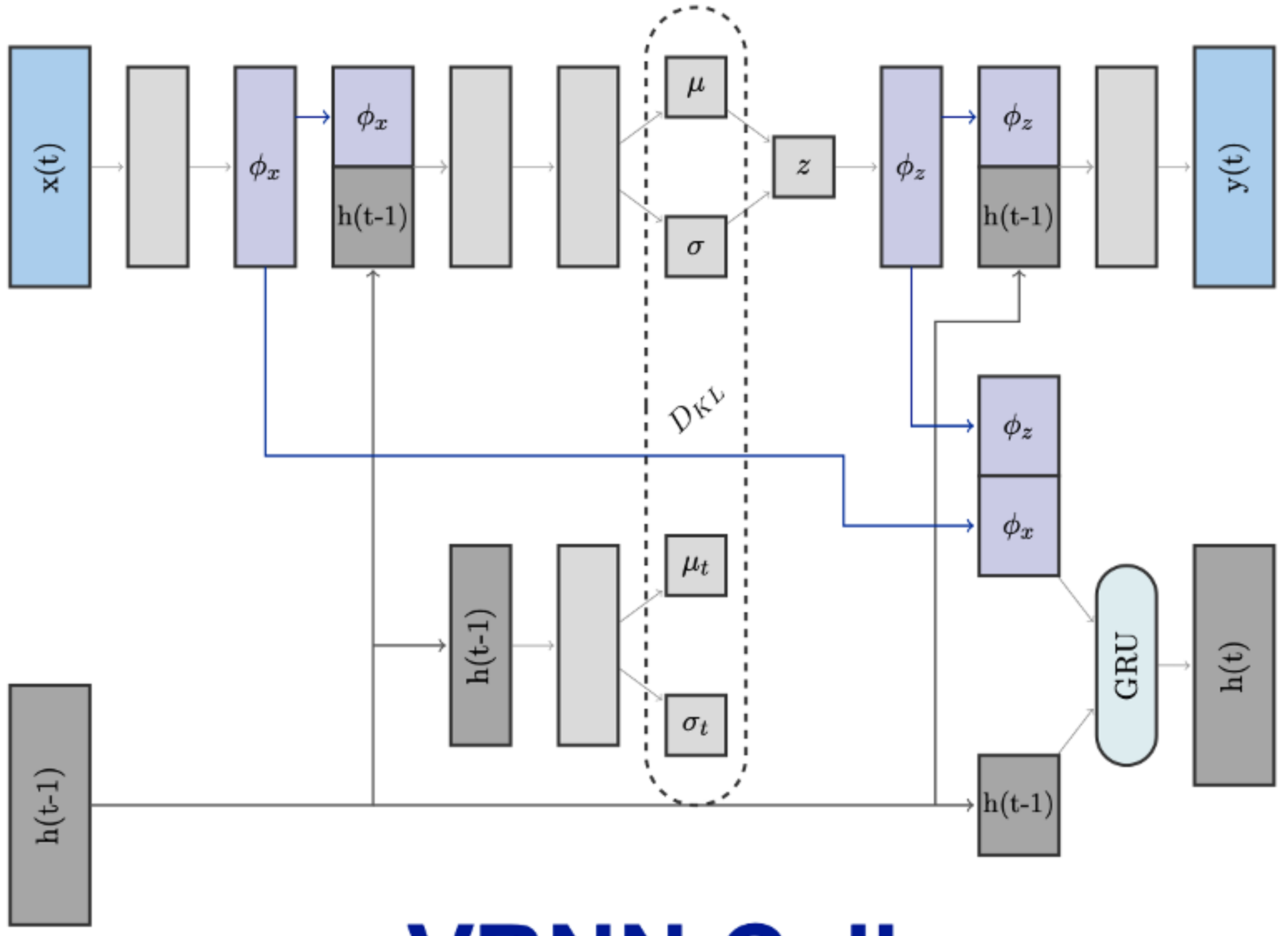


Variational Recurrent Neural Network

- Variational autoencoders are a fixed length architecture
  - ↪For anomalous jet substructure the desired input – jet constituents – is of variable length
- A recurrent architecture addresses this problem
  - ↪Recurrent networks break data into a sequence of features (time steps)
  - ↪The input to each cell is a fixed length feature
  - ↪The hidden state is updated at each time step. This allowed the hidden state to store the long term representation of the data in the sequence
- Variational autoencoder + recurrent architecture = **V**ariational **R**ecurrent **N**eural **N**etwork
  - ↪Recurrent neural network that updates a VAE latent space at each time step, accommodating variable-length input sequences



Recurrent Neural Network



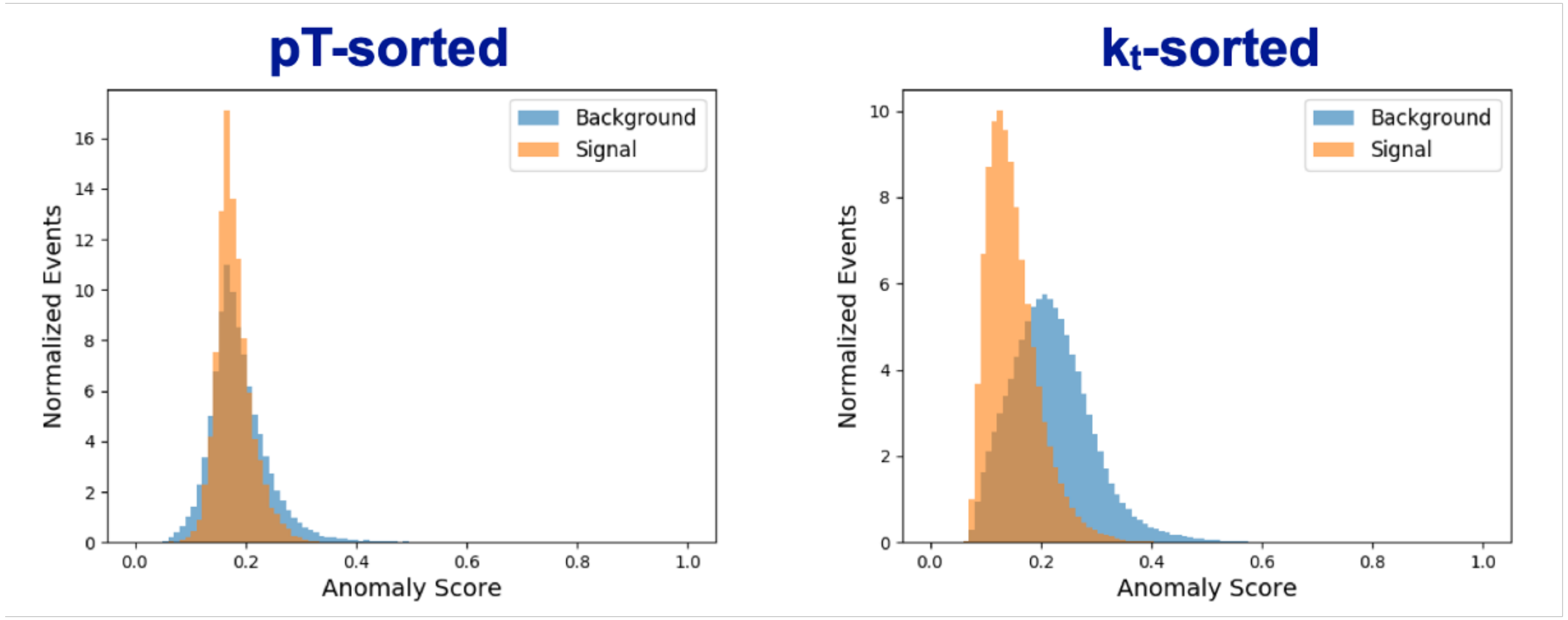
VRNN Cell

VRNN Paper

VARIATIONAL RECURRENT NEURAL NETWORK

Training setup

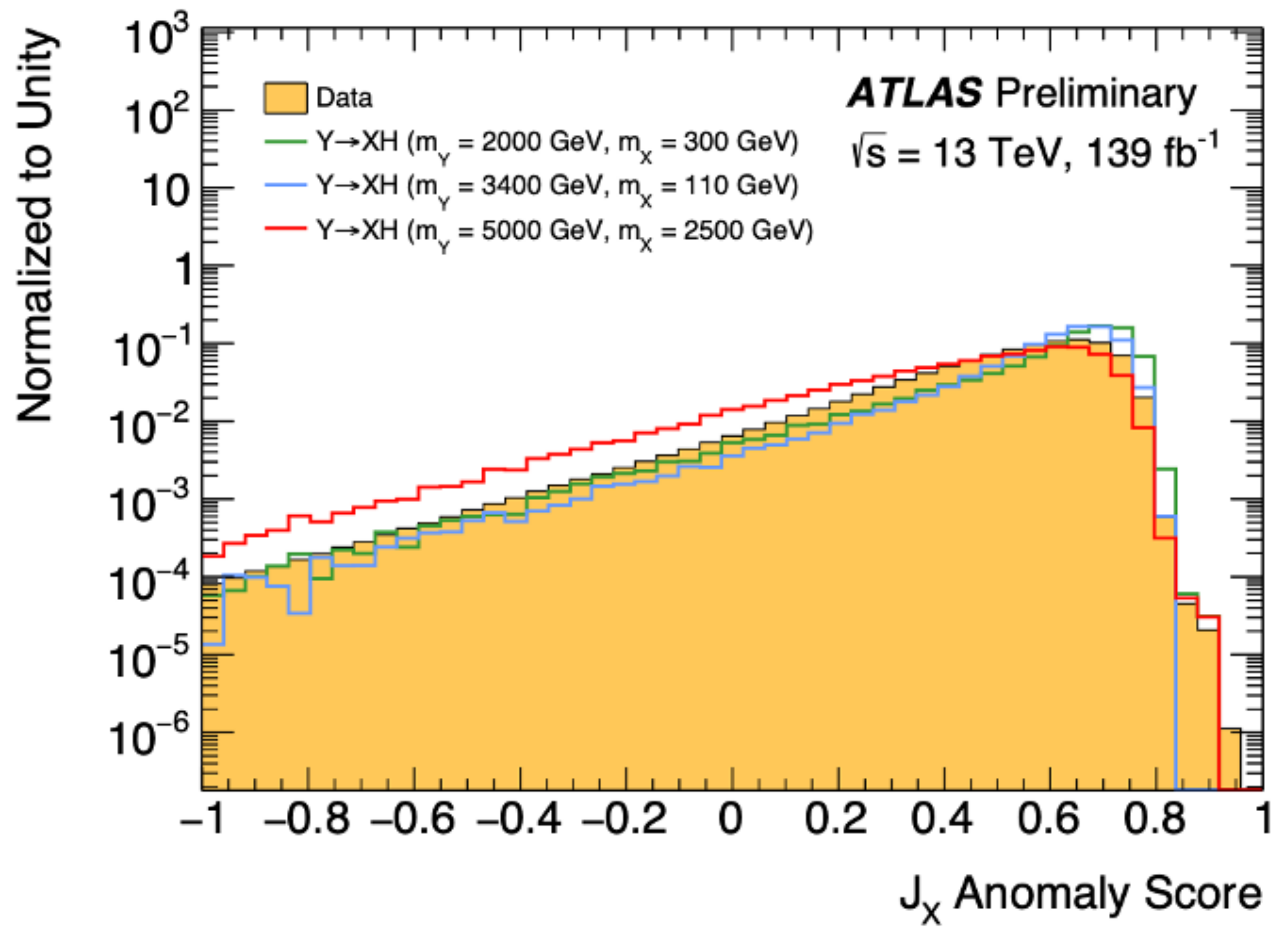
- In $Y \to XH$, ATLAS data is modeled as a sequence of jet constituents' 4 vectors (p_T, η, ϕ)
- The recurrent architecture is sensitive to the sequence modeling of the input data – in this case, the ordering of the jet constituents
- A k_T distance ordering is used:
  - ↪The nth constituent has the highest k_T distance relative to the n-1 constituent (starting with the highest p_T constituent)
- Mass and p_T information of the jet is removed during preprocessing, to avoid tagging on kinematics alone
- The network is conditioned on 4 HLVs; D_2, τ_{32}, and k_T splitting $Split_{12}$ & $Split_{23}$

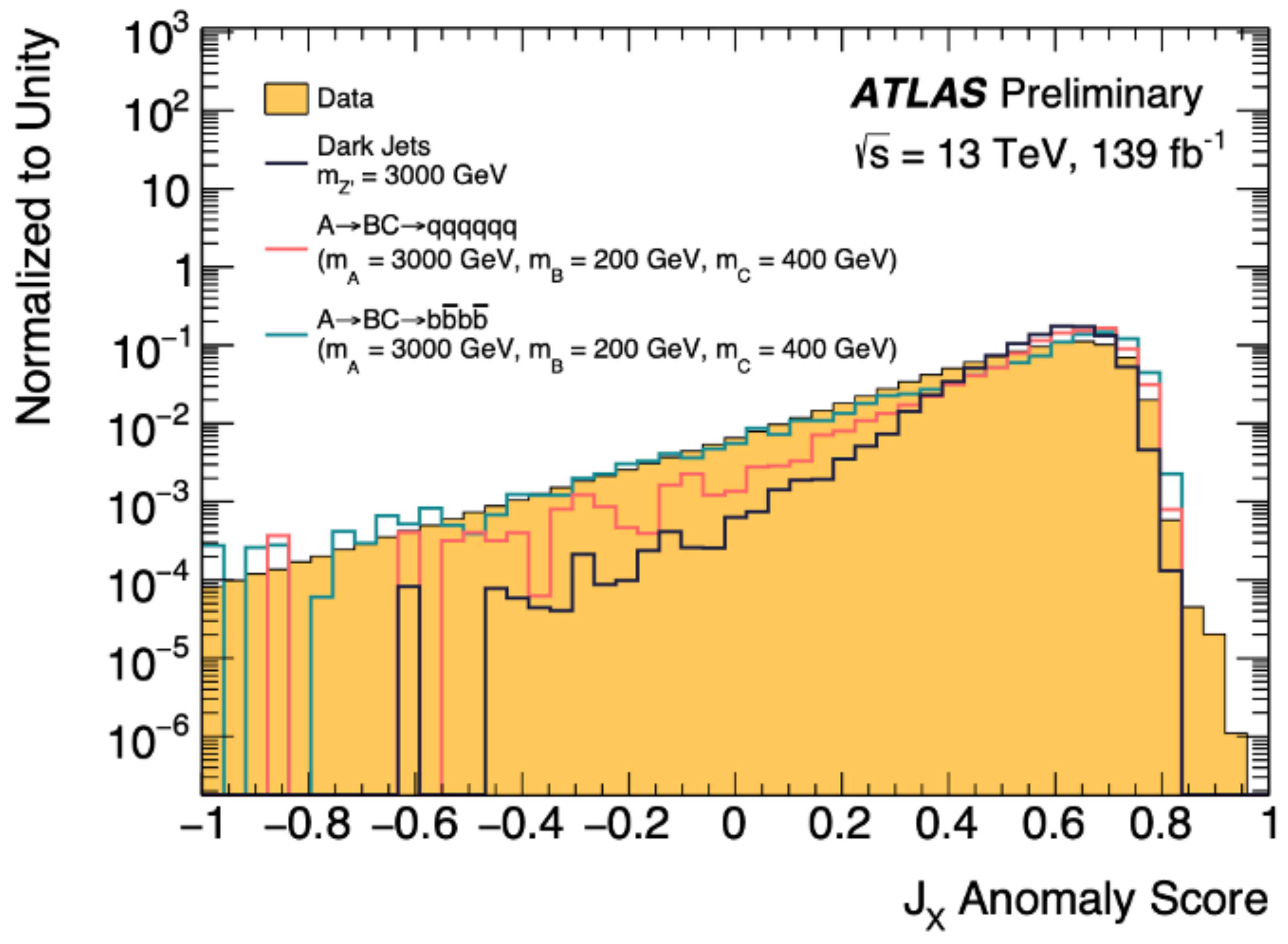


Variational Recurrent Neural Network

Anomaly Score

- The VRNN learns general features of the data and is able to separate background jets from jets with distinct substructure though an anomaly score (AS)
  - ↪AS is defined so that higher reconstruction error corresponds to a higher AS
- The analysis uses an optimized selection of AS > 0.5 to define the AS discovery region
  - ↪Two other regions (2-prong Merged and 2-prong Resolved) are used for exclusion results



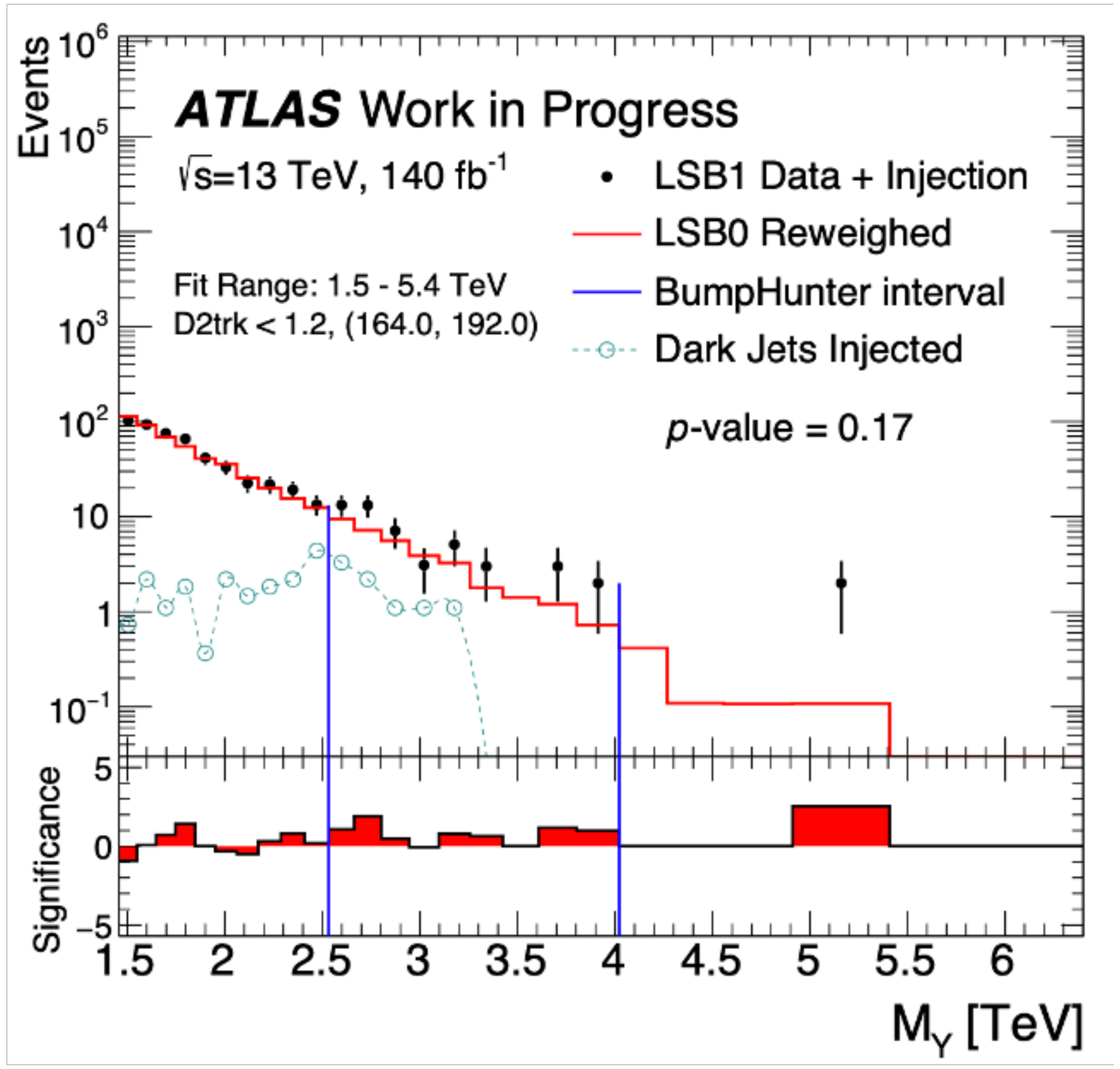


"Special" Signals
Dark Jets
3-prong
Heavy Flavor

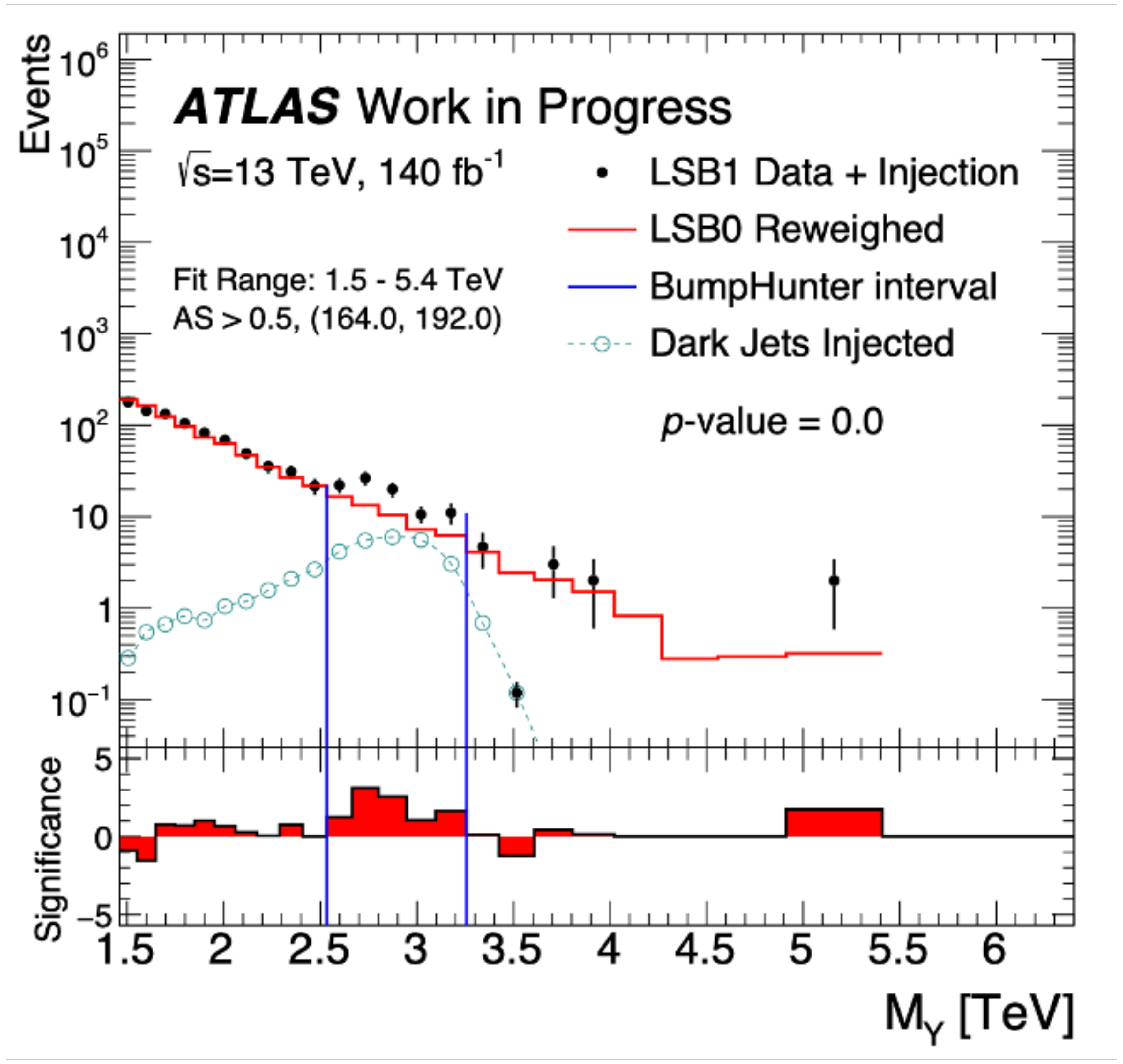
$Y \to XH$ RESULTS

Signal Injection in Anomaly Score Region

- D_2^{trk} looks for 2 prong substructure, which kills the dark jet signature
- The AS region allows sensitivity to generic new physics produced with a Higgs!



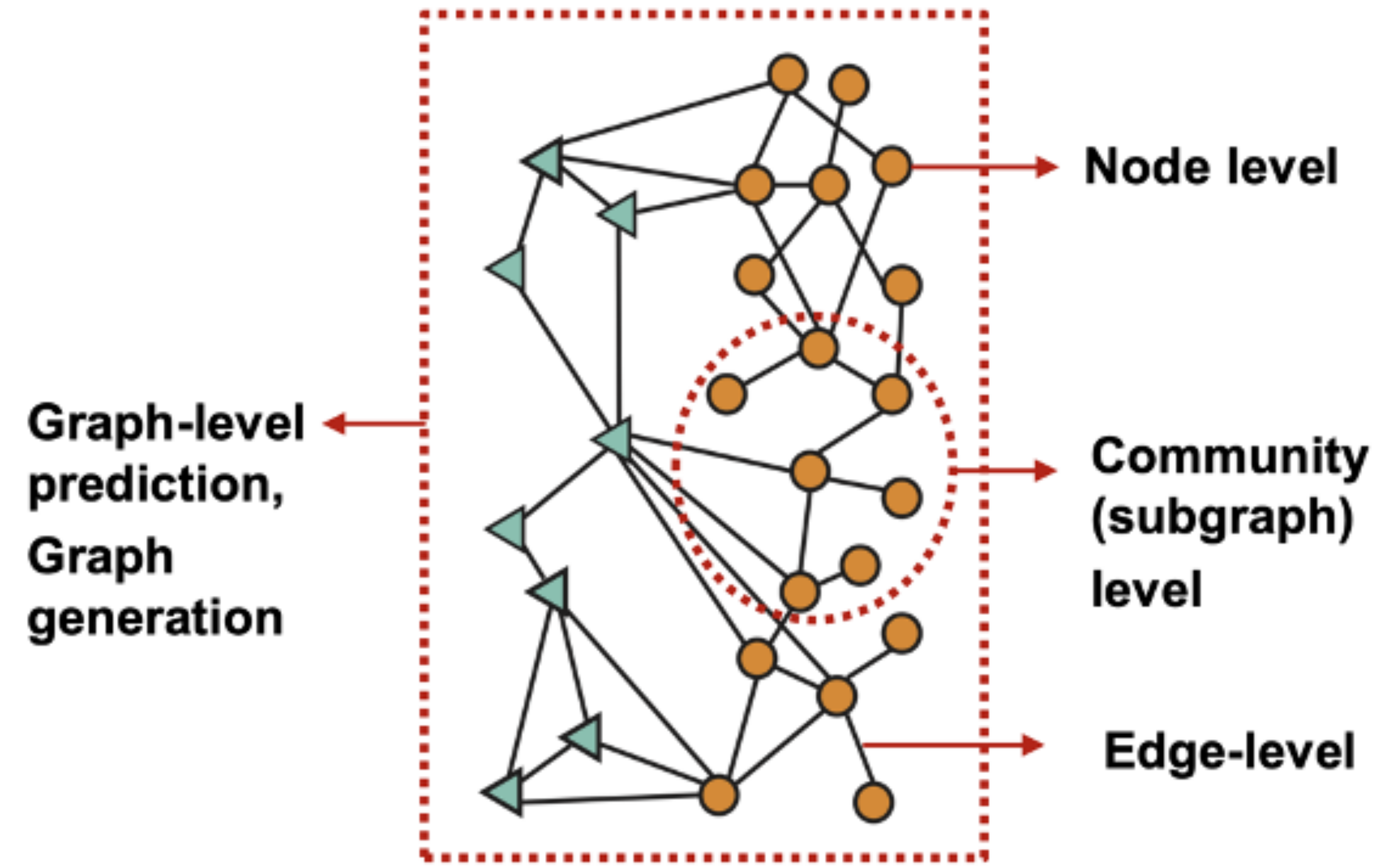
Dark Jet Injection in 2-Prong Merged region D_2^{trk}< 1.2)

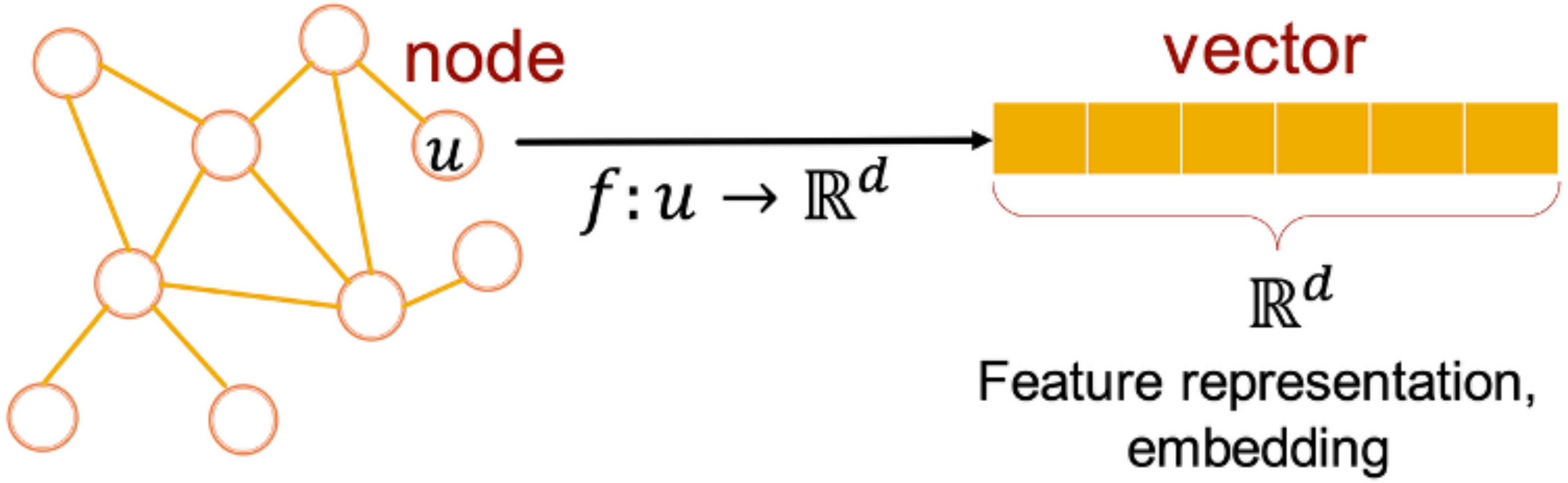


Dark Jet Injection in Anomaly Score region (AS> 0.5)

GRAPH NEURAL NETWORK

- Graph Neural Networks (GNNs) are ML architectures built specifically to make predictions on graphs, exploiting their relational nature
  - ↪The network during training learns the vector representation (embedding h_v) of each node of the input graphs
  - ↪Embedding of a target node depends in some way on what the embeddings of the other nodes are and from its structure





Several task levels, carried out by processing the final node embeddings in certain ways

Graph Neural Network

- The embeddings are updated at each layer by aggregating the information passed between the target node and the nodes from its closest neighbourhood → message passing
- G embedding is obtained by pooling the nodes embedding at the final layer into one global representation
  - ↪Global sum pooling: $h_G = Sum(h_v^L \in R^d, \forall v \in G)$
  - ↪Global mean pooling: $h_G = Mean(h_v^L \in R^d, \forall v \in G)$
  - ↪Global max pooling: $h_G = Max(h_v^L \in R^d, \forall v \in G)$