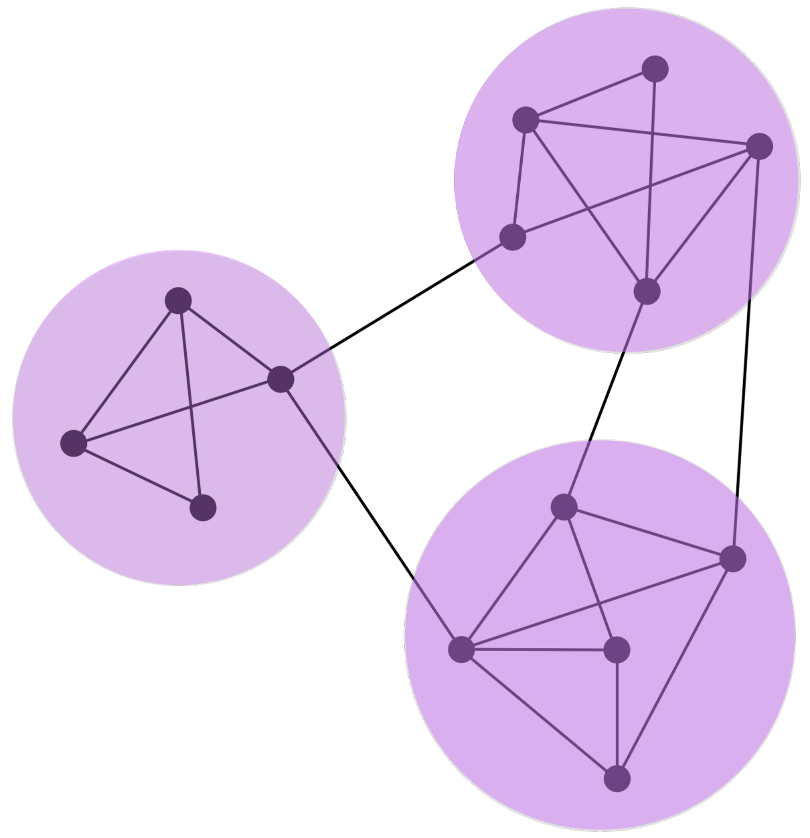
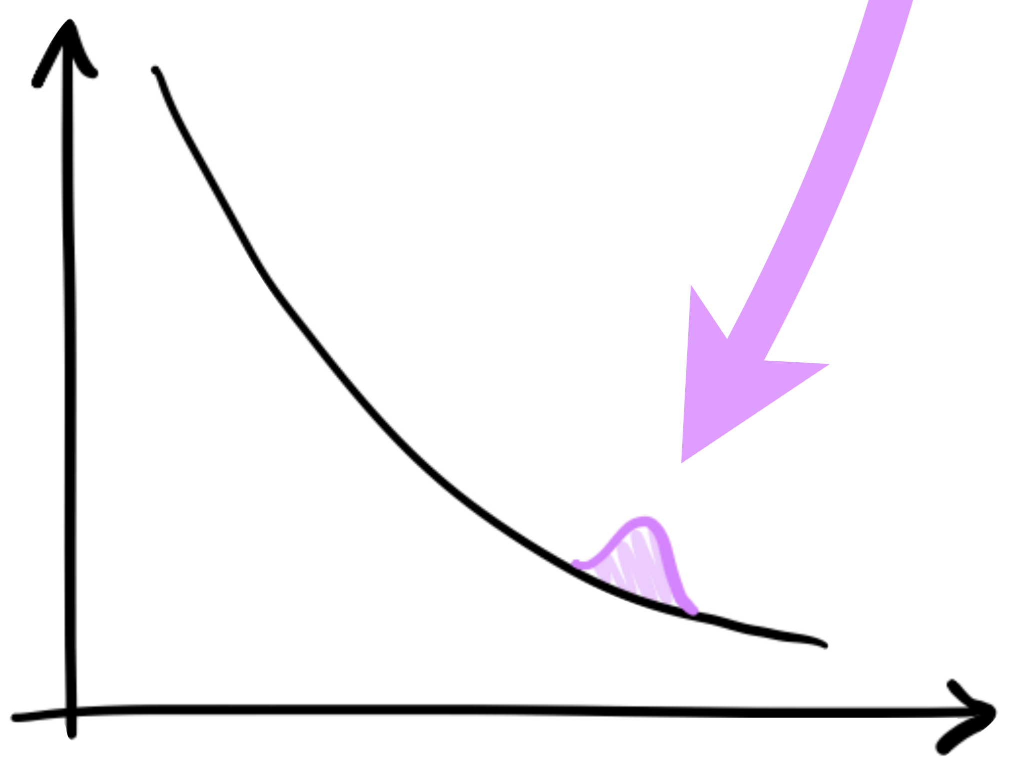


New Physics and how to find it (in ATLAS with ML tools)

Graziella Russo

Young@INFN
24/04/24



Structure of the talk

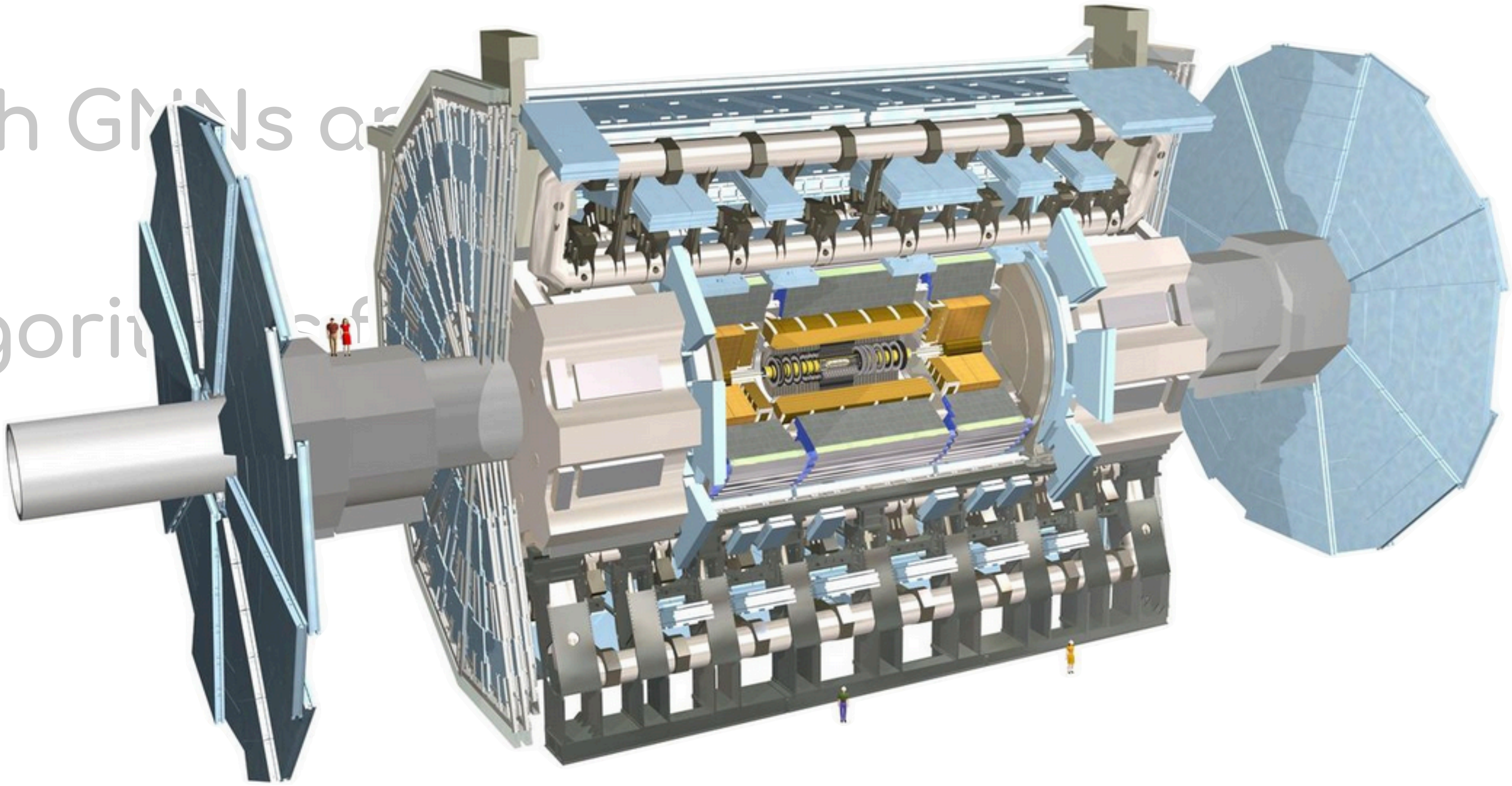
1. Basics of collision experiments (ATLAS)
2. Anomaly Detection with GNNs analysis
3. CNN-based trigger algorithms for ATLAS upgrade

Structure of the talk

1. Basics of collision experiments (ATLAS)

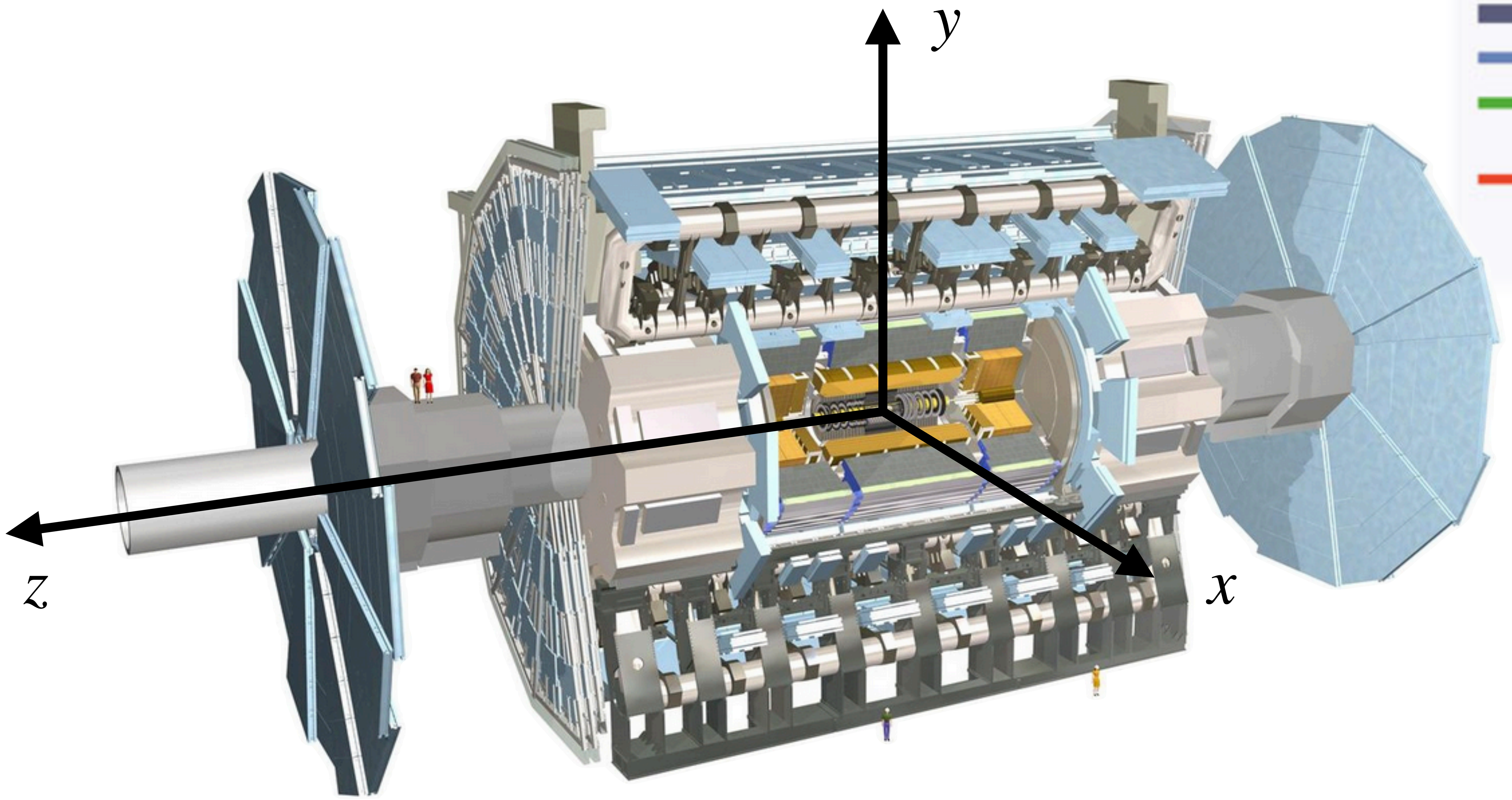
2. Anomaly Detection with GNNs

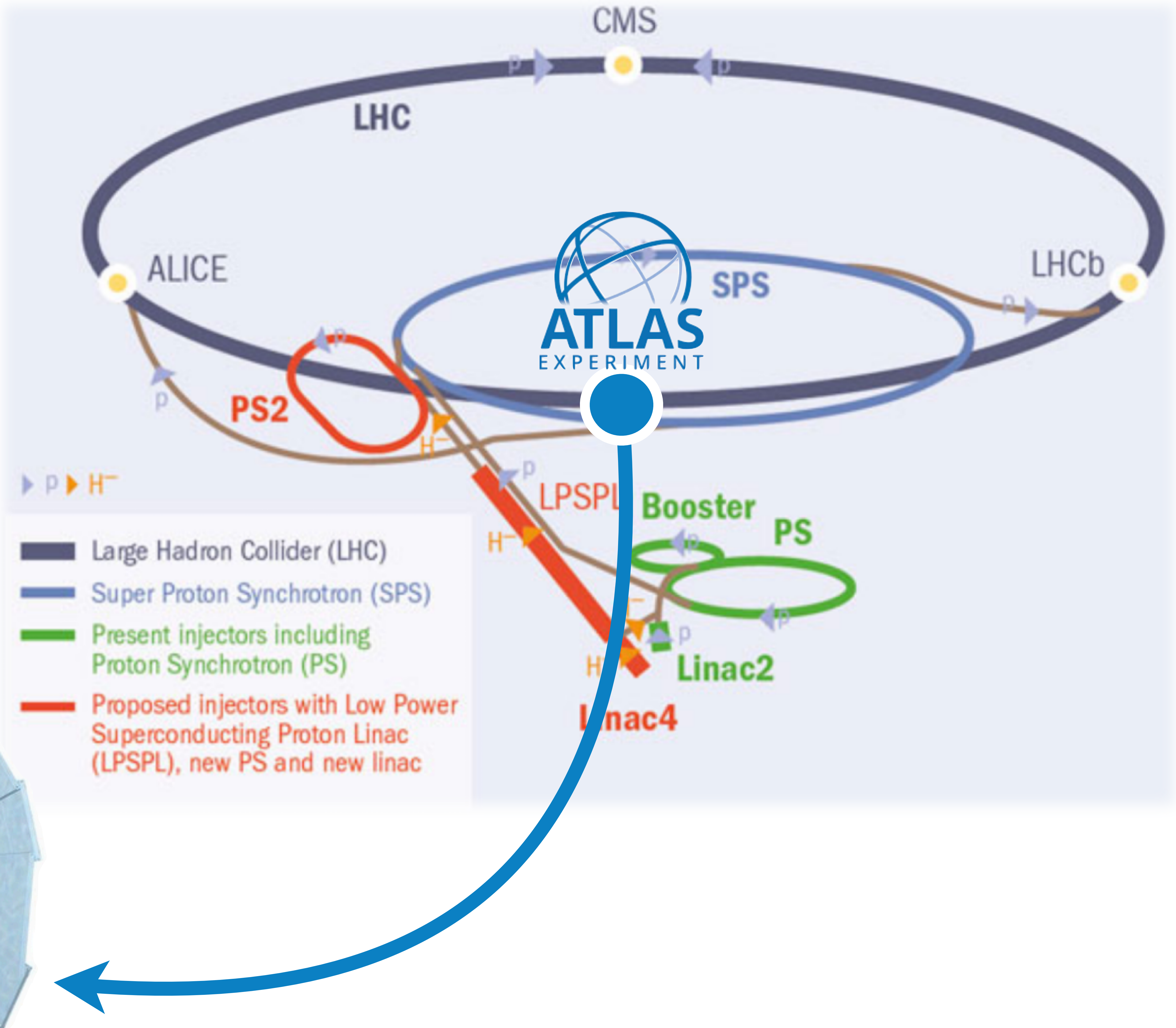
3. CNN-based trigger algori



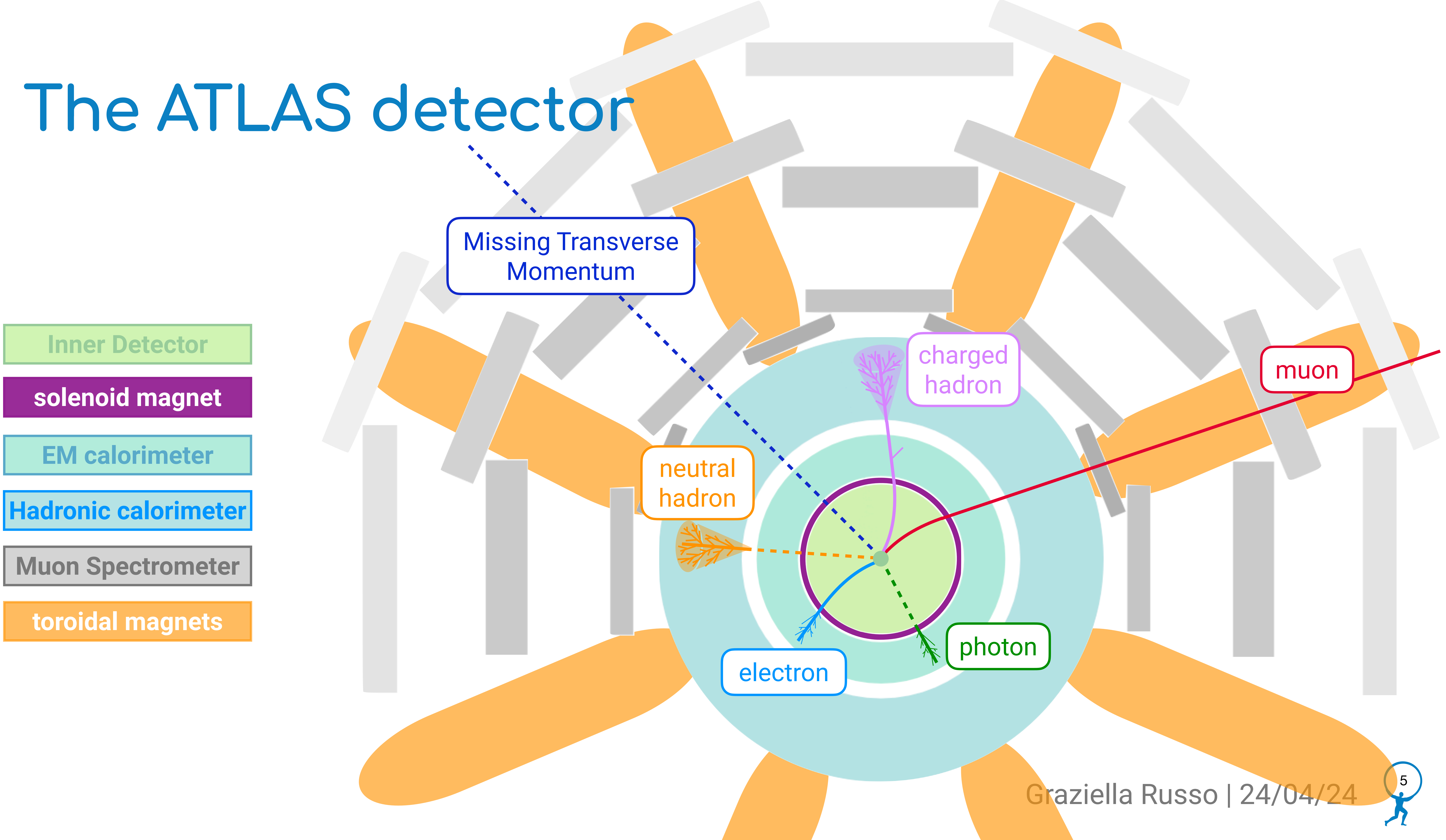
LHC accelerator

- LHC = Large Hadron Collider
- proton-proton collisions at $\sqrt{s} = 13.6\ TeV$
- 4 interaction points —> 4 experiments: **ATLAS**, CMS, Alice, LHCb





The ATLAS detector



The trigger system

$1.7 \cdot 10^9$ collisions every second @ LHC = 60 TB per second

1 over 1M events are stored

ATLAS trigger system

- **Level1**
  hardware trigger; only subset of detector are considered; 100k events accepted per second
- **High Level Trigger**
  software trigger; 1k events accepted per second

Starting from the theory

- up to 2012 the Standard Model was not complete
- 2012: discovery of the Higgs boson 

Dark Matter?

neutrino mass?

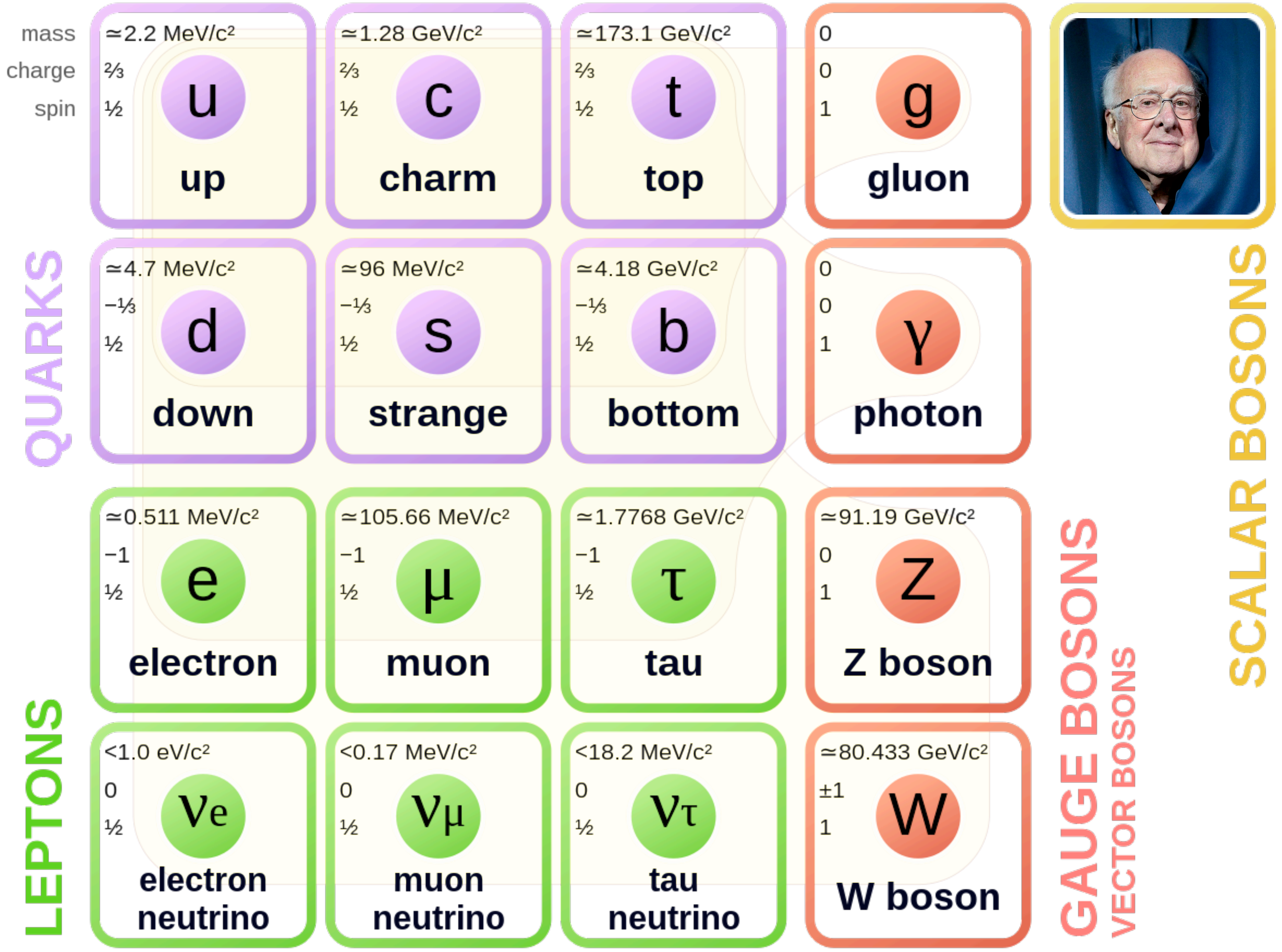
Dark Energy?

Still many ***unanswered questions***

strong CP problem

hierarchy problem

matter/antimatter asymmetry?

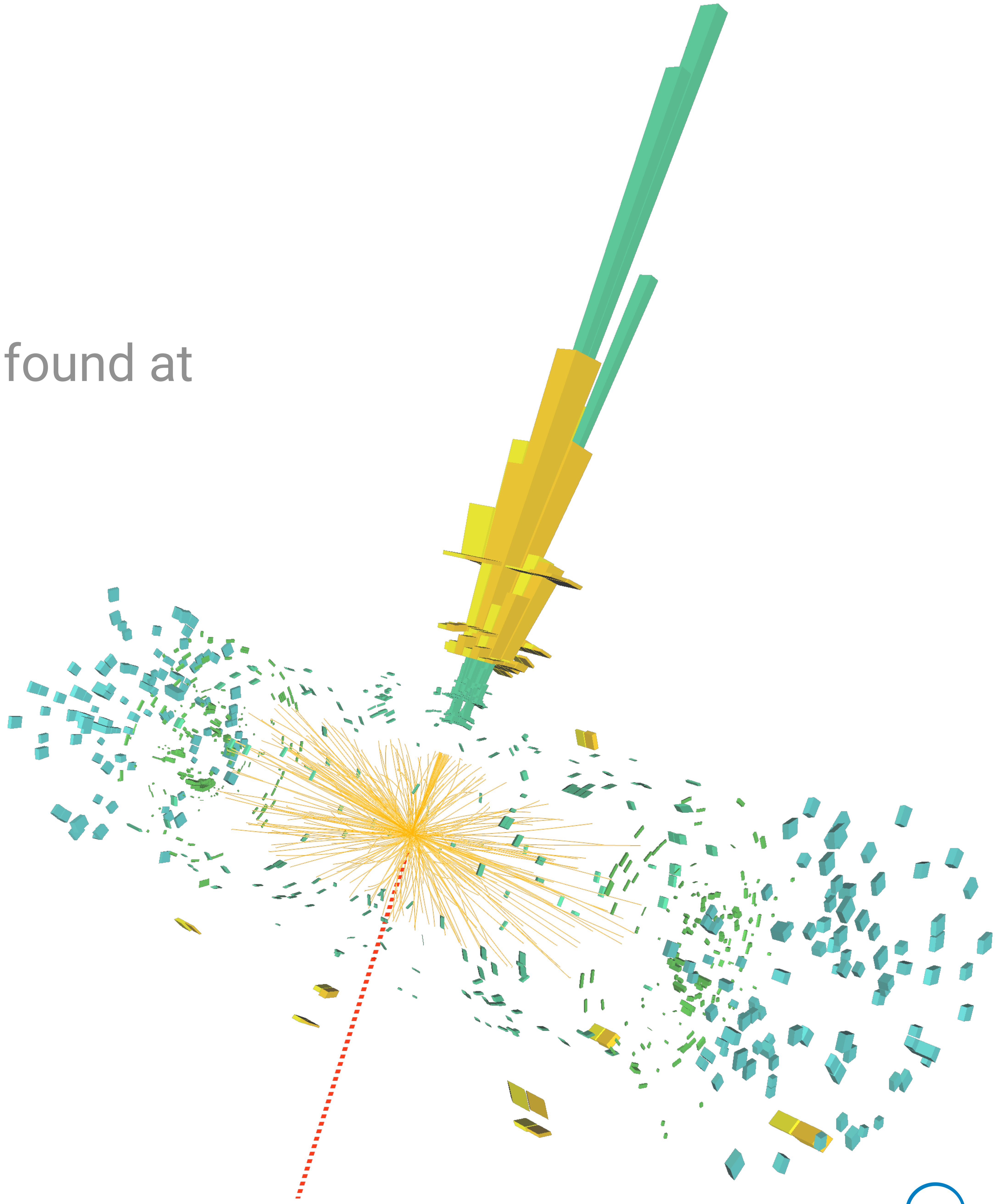


Where is New Physics?

from 2012 no evidence of New Physics (NP) has been found at LHC experiments

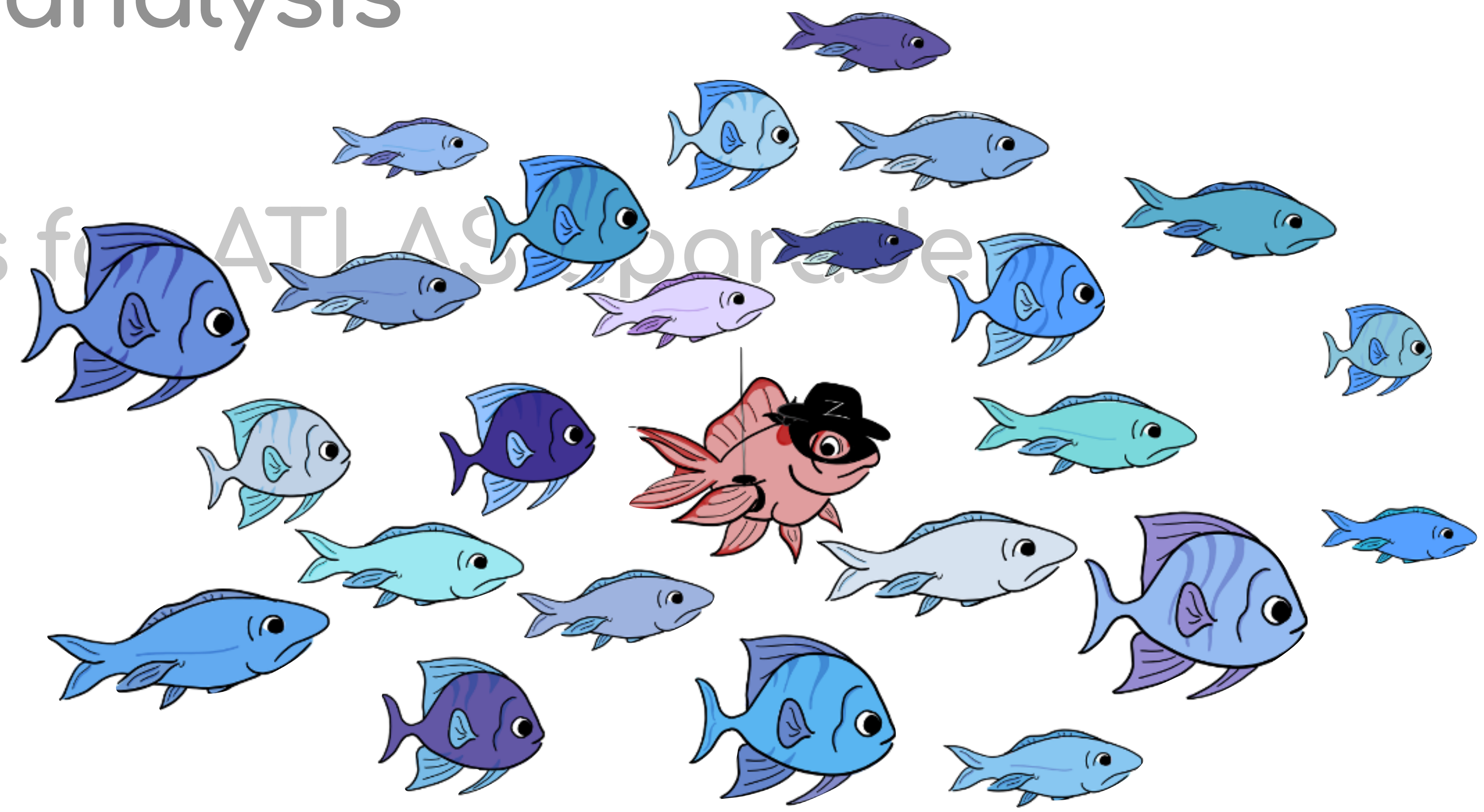
@ analysis level:
are we looking at the right discriminating variable?

@ trigger level:
are we storing all the interesting events and not discarding Beyond the Standard Model (BSM) events?



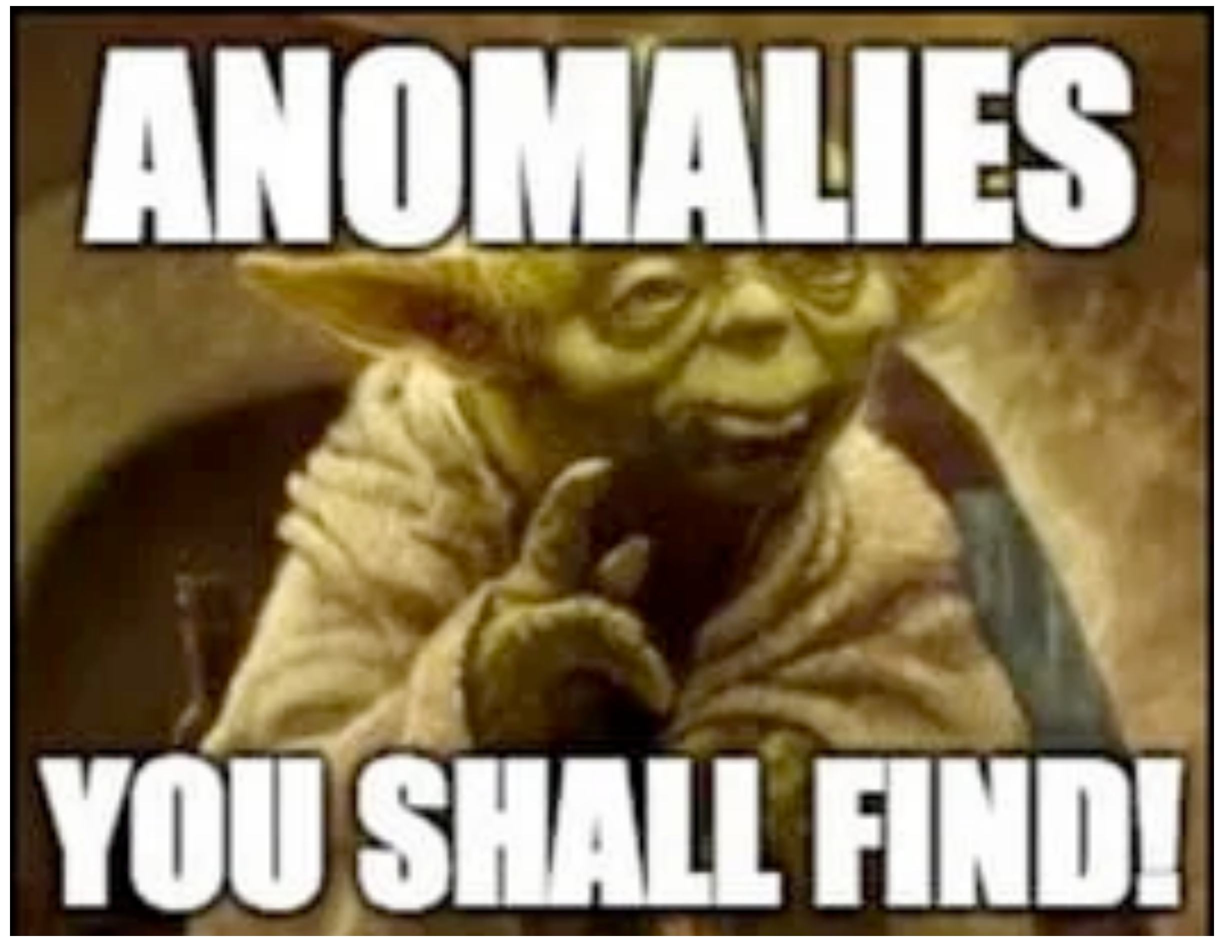
Structure of the talk

1. Basics of collision experiments (ATLAS)
2. Anomaly Detection with GNNs analysis
3. CNN-based trigger algorithms for ATLAS upgrade

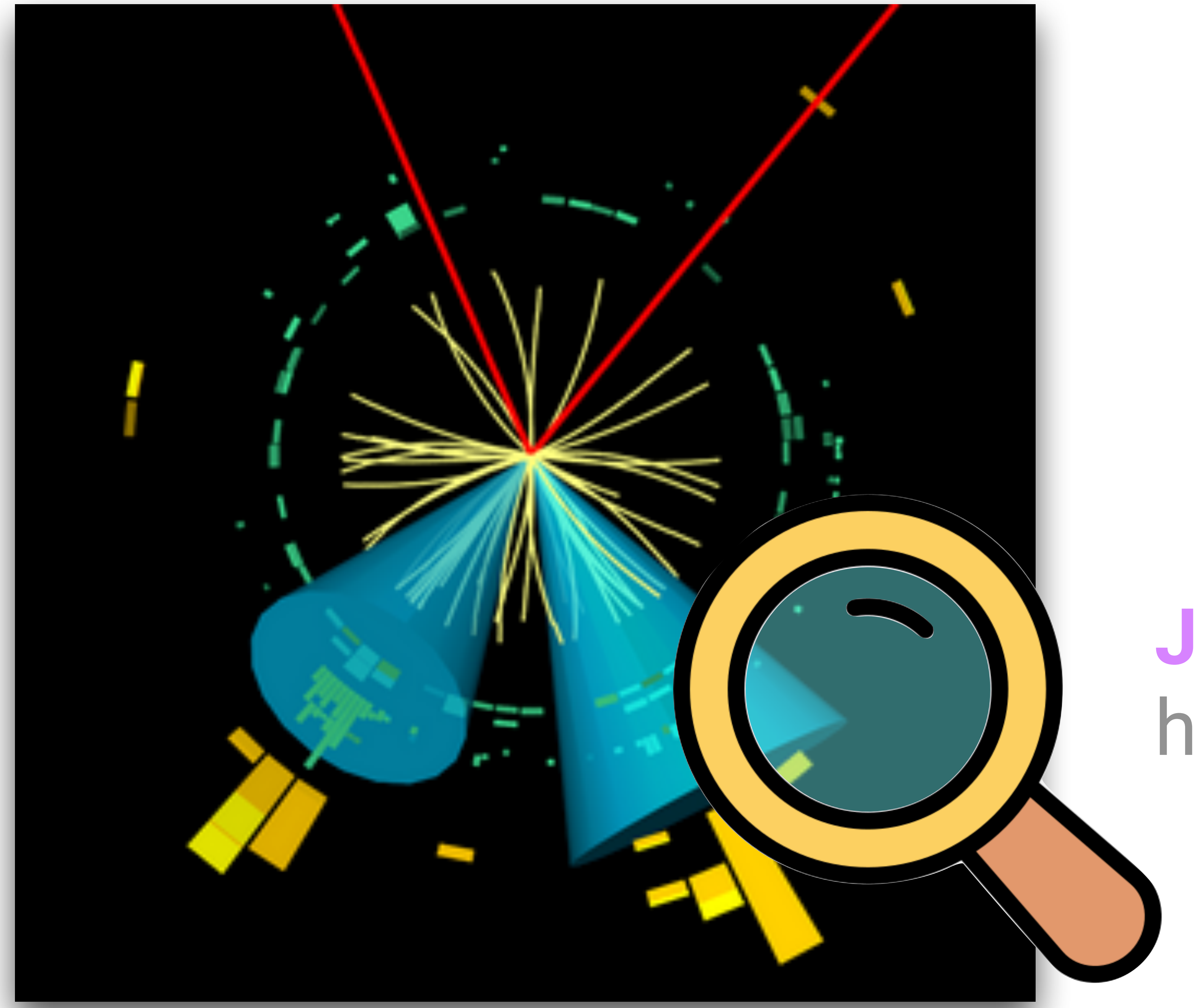


The idea behind

NP should consist in a *small deviation* from Standard Model data, otherwise we would have seen it before



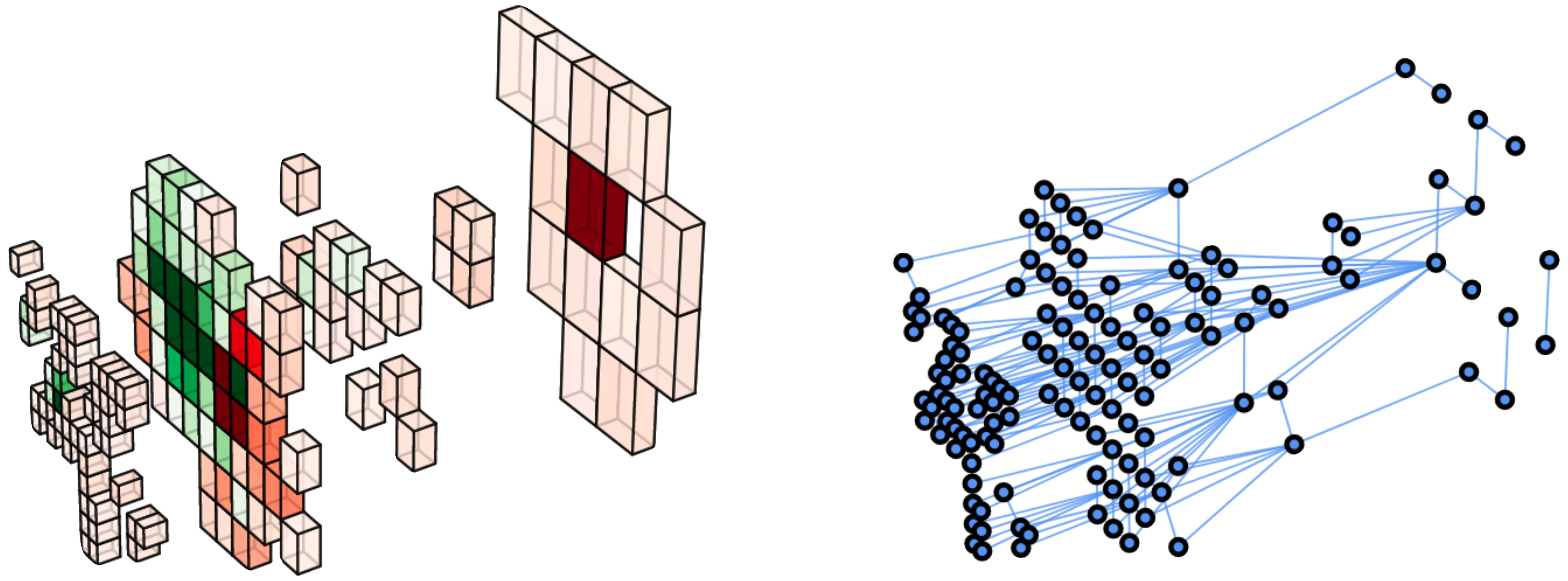
Anomaly Detection (AD) is an AI technique to identify anomalous behaviours among established, standard patterns. AD does not require a specific theoretical model, it looks directly on data

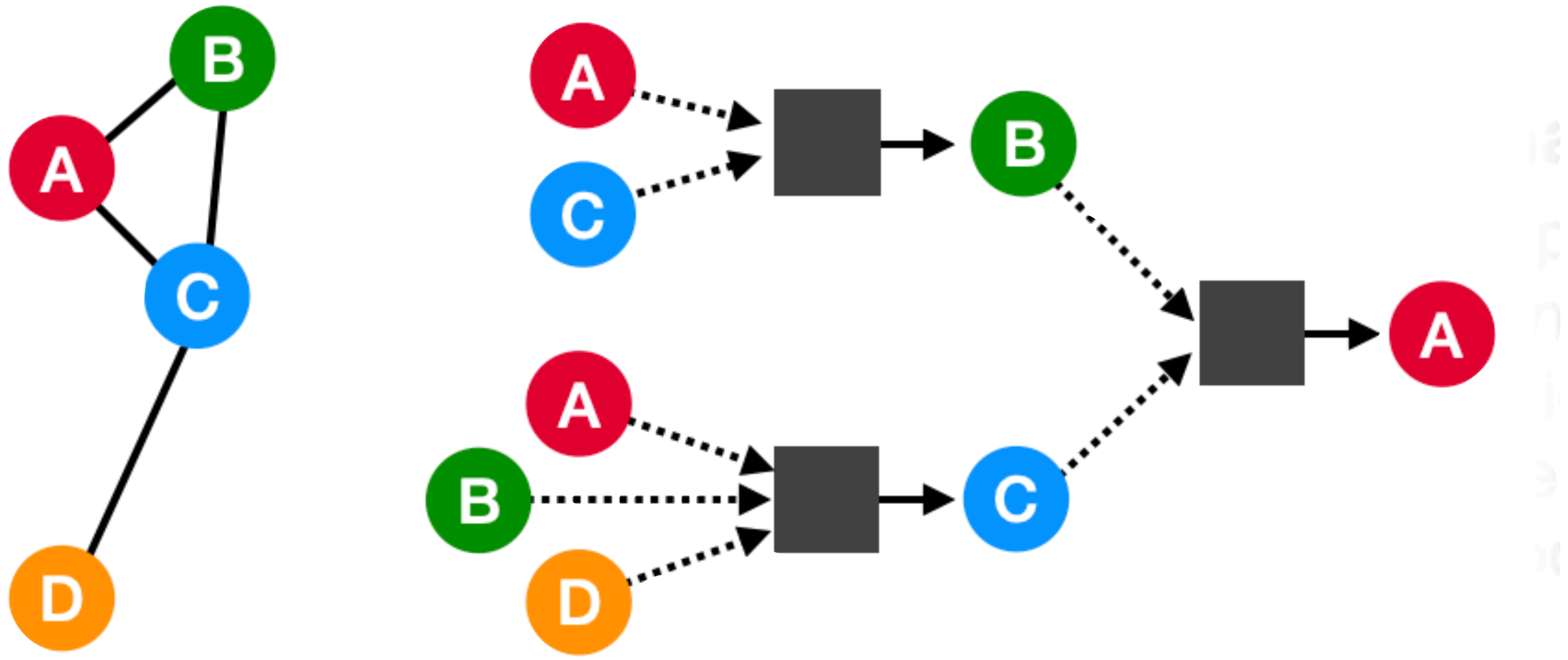


Jets have very complex substructure that can potentially hide NP

Graph Neural Networks (GNNs)

- GNN is any ML algorithm that handles graph structured data
- **message passing** is a useful way to embed the information from the neighbouring nodes





jets have very sparse structure, suitable for graph representation

Jets and graphs

A graph is a set of points (**nodes**) that can be connected by **edges**

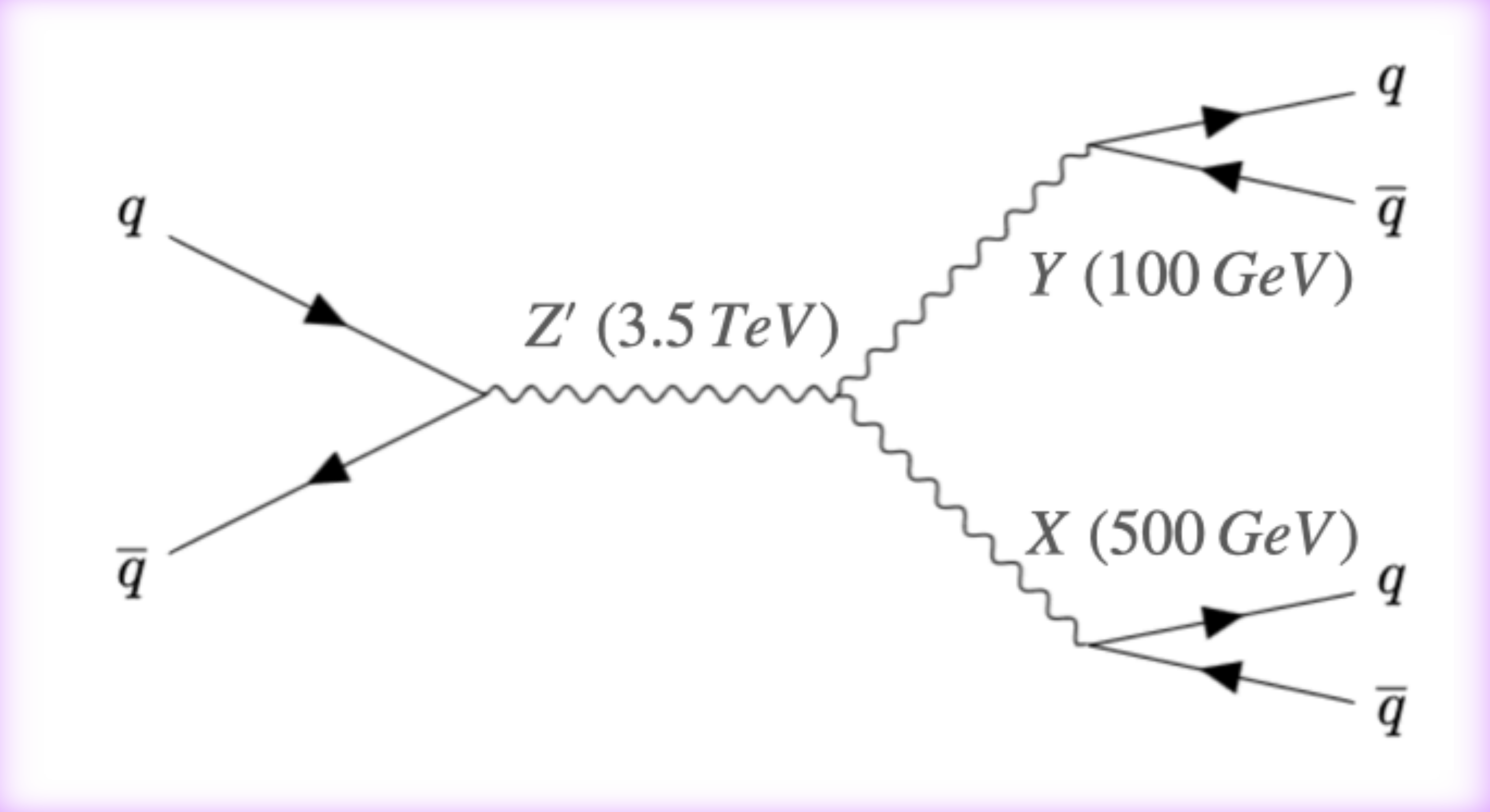
- **what is a node?**
  - each constituent is a node
  - node features: $p_T\, frac, \eta, \phi$
- **when are nodes connected?**
  - edge features $\dfrac{1}{\Delta R(const_i, const_j)}$
  - distance-based

Transformation **applied for data augmentation and model robustness reasons**

- ML algorithms rely a lot on the mass of the signature used in training, need to decorrelate AD score from S/B discriminants, e.g. mass
- rescaling of the four momenta ($m_{jet} = m_0 = 0.25\ GeV$)
- boost so that $E_{jet} = E_0 = 1\ GeV$
- further rotation of constituents along jet axis

toy model

- R&D *LHC Olympics* dataset
  - $Z' \longrightarrow XY \longrightarrow qqqq$ events
  - $m_{Z'} = 3.5\ TeV, m_X = 500\ GeV, m_Y = 100\ GeV$
  - reconstructed as large radius jets



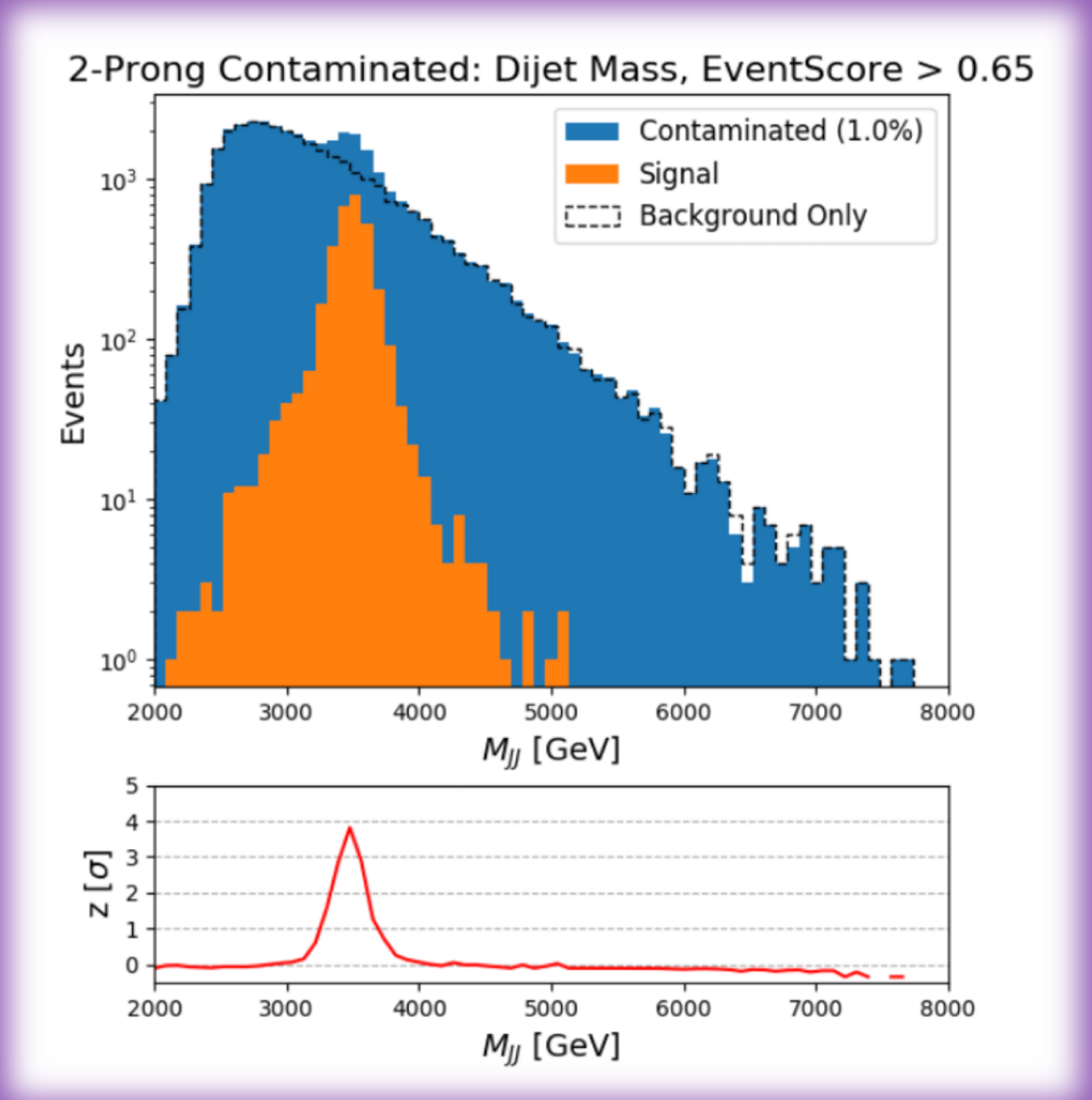
benchmarks

- ML tasks
  - output (jet-level) can be used to build event-level Anomaly Score (AS)
  - trained both supervised and unsupervised models

- to test the models, apply to benchmark physics
  - test on data for W boson rediscovery

final goal

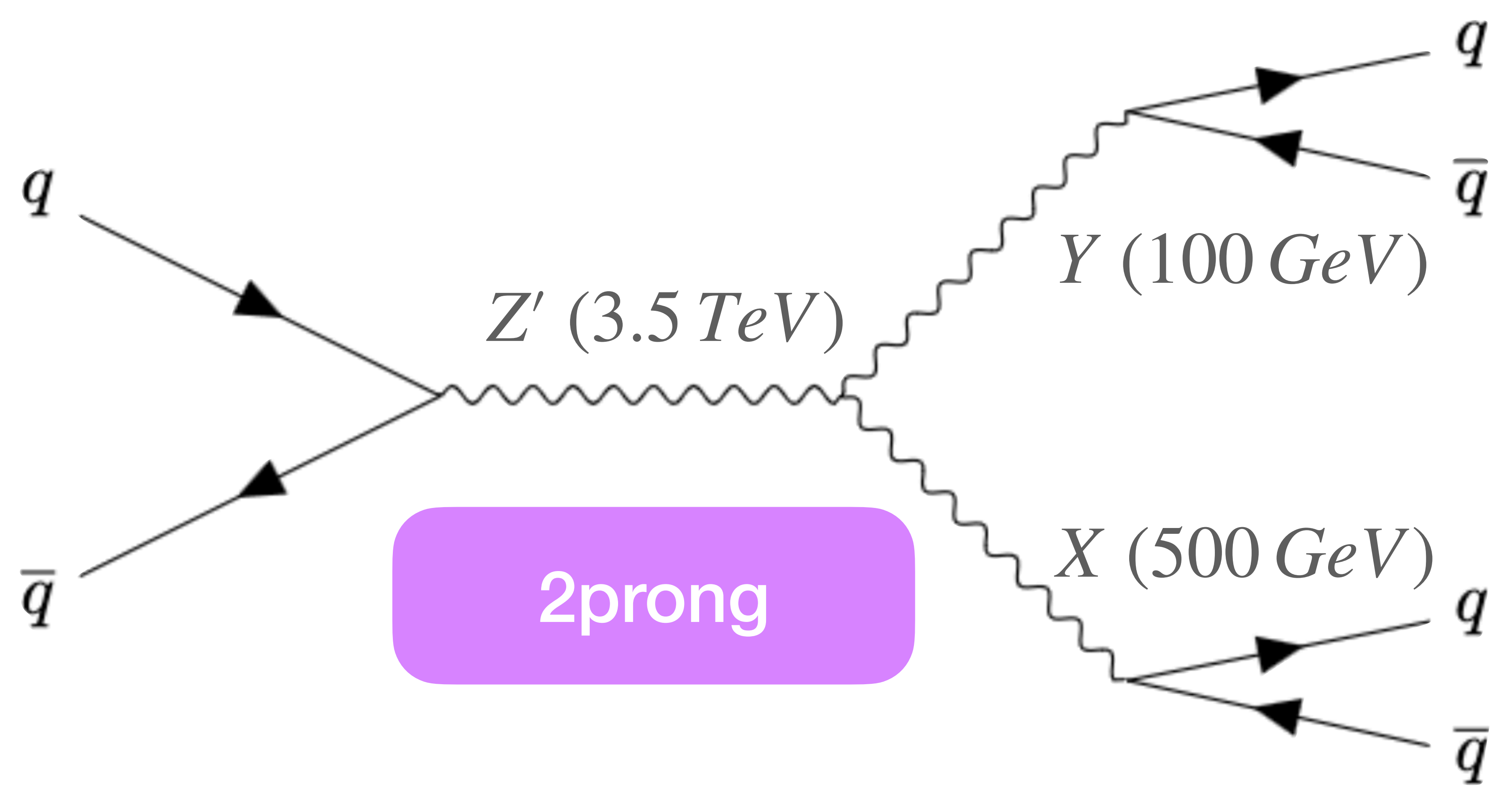
- Run3 di-jet fully hadronic search
  - completely model agnostic
  - 2 large-R jets
  - event selection based on AS cut
  - data-driven background estimation
  - bump hunt/framework fit for the AS distribution



ML dataset and models

LHC Olympics dataset contains:

- 1M QCD events (background) + 100k signal events
- **Event** = up to 2 large R jets
- **Jet** = up to 50 constituents
- **Constituent** = $[p_T\ frac, \eta, \phi]$, distance-based edges



Transformer

Graph Neural Networks

- Graph Isomorphism Network (GIN) layer
- Edge Graph Attention Transformer (EGAT) layer

GNN trainings

jet-level AD = model loss

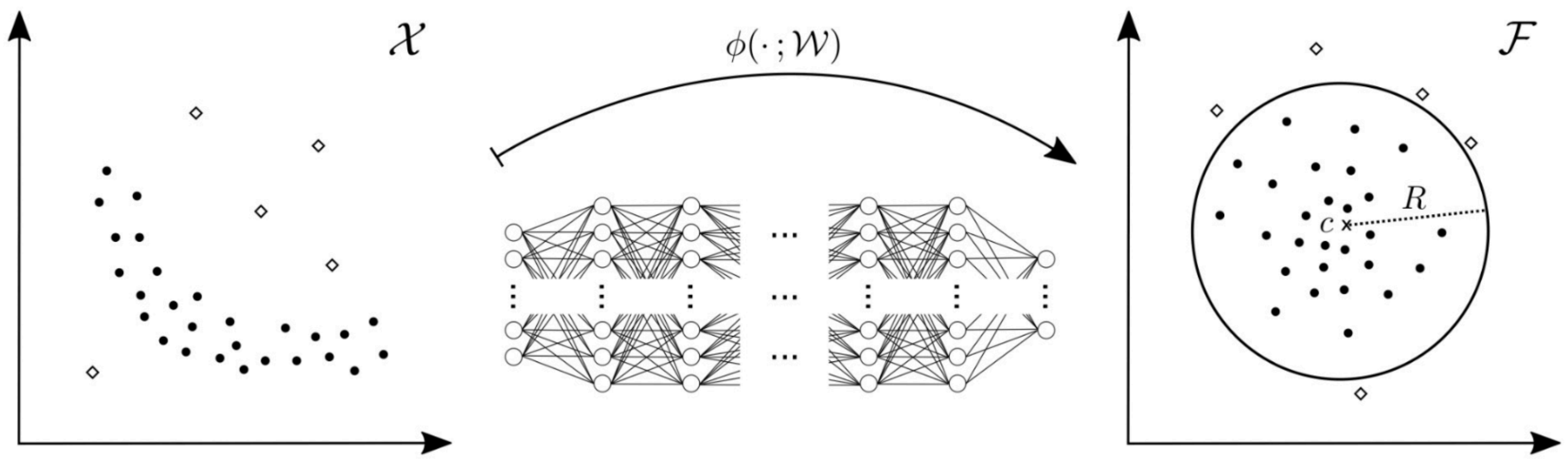
- using the DeepSVDD unsupervised loss
- optimizing the radius of hypersphere in the hidden representation space to contain all standard events

objective

$$\min_{W} \frac{1}{N}\sum_{i=1}^{N}\|\mathrm{GIN}(G_i;W)-\mathrm{c}\|^2+\frac{\lambda}{2}\sum_{l=1}^{L}\|W^l\|_F^2 \qquad W=\{W^1,\dots,W^l\}$$

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

Anomaly Score

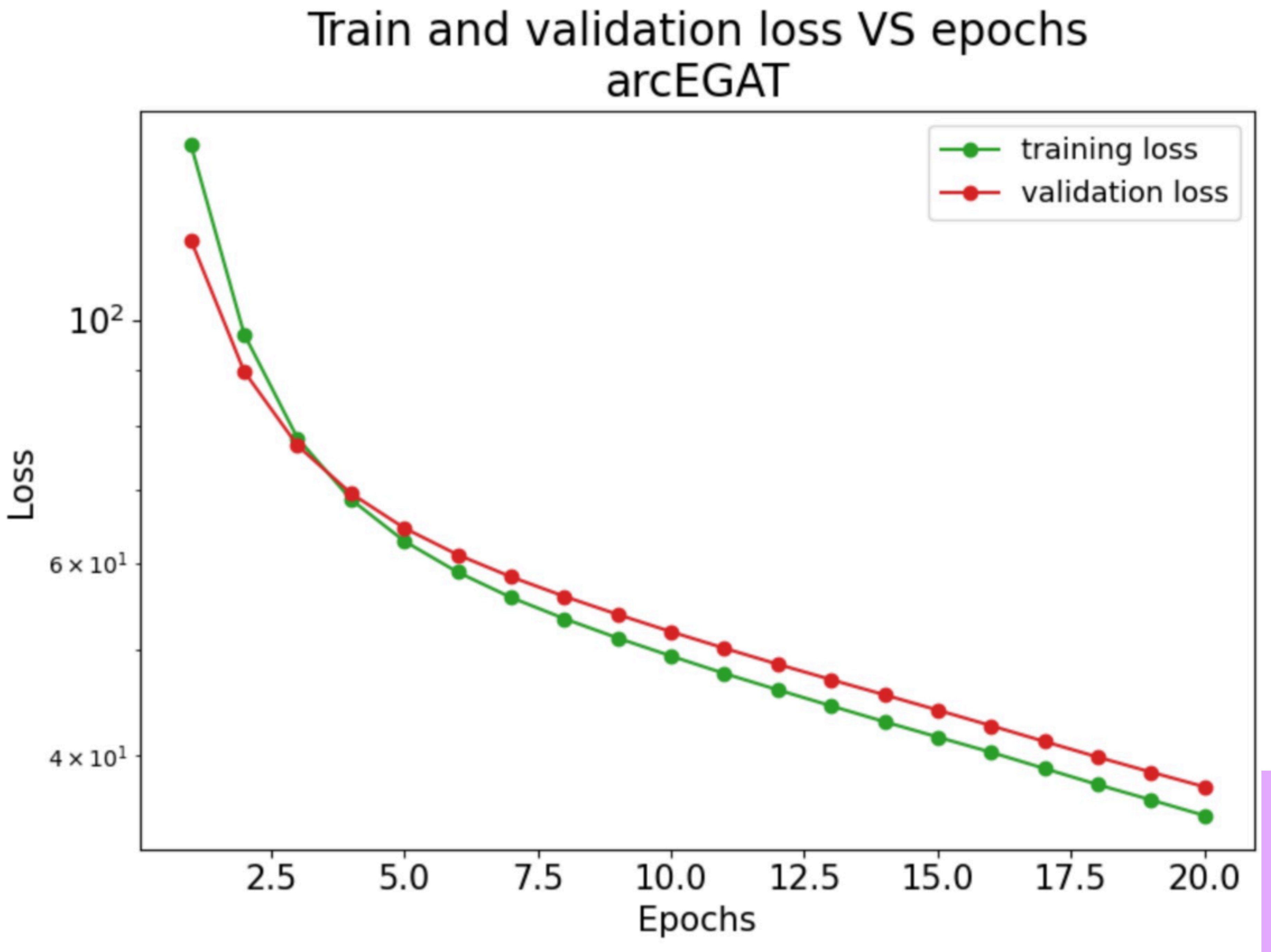


supervised = trained on balanced dataset with signal and background
unsupervised = trained only on background events

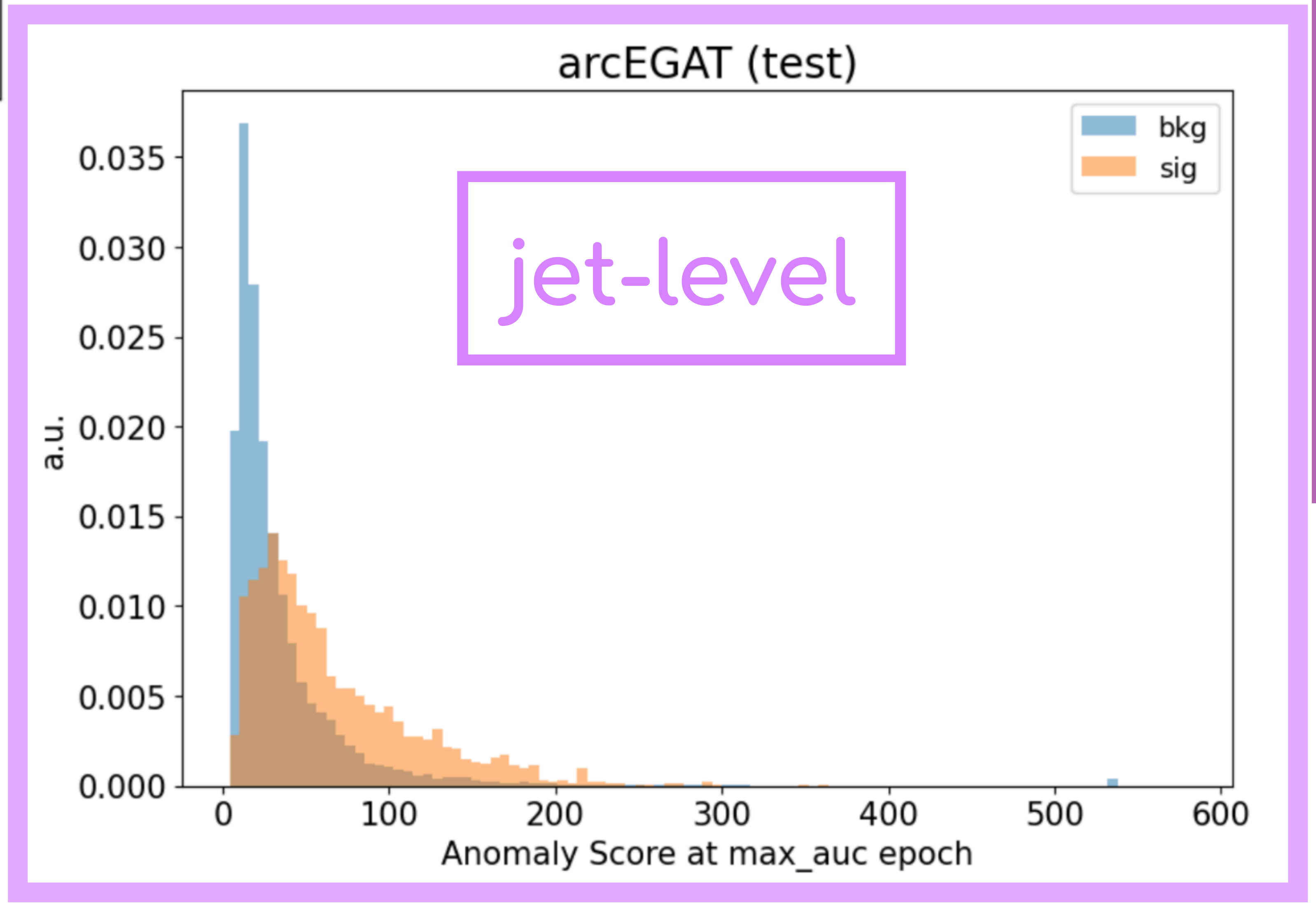
event-level AD strategy:

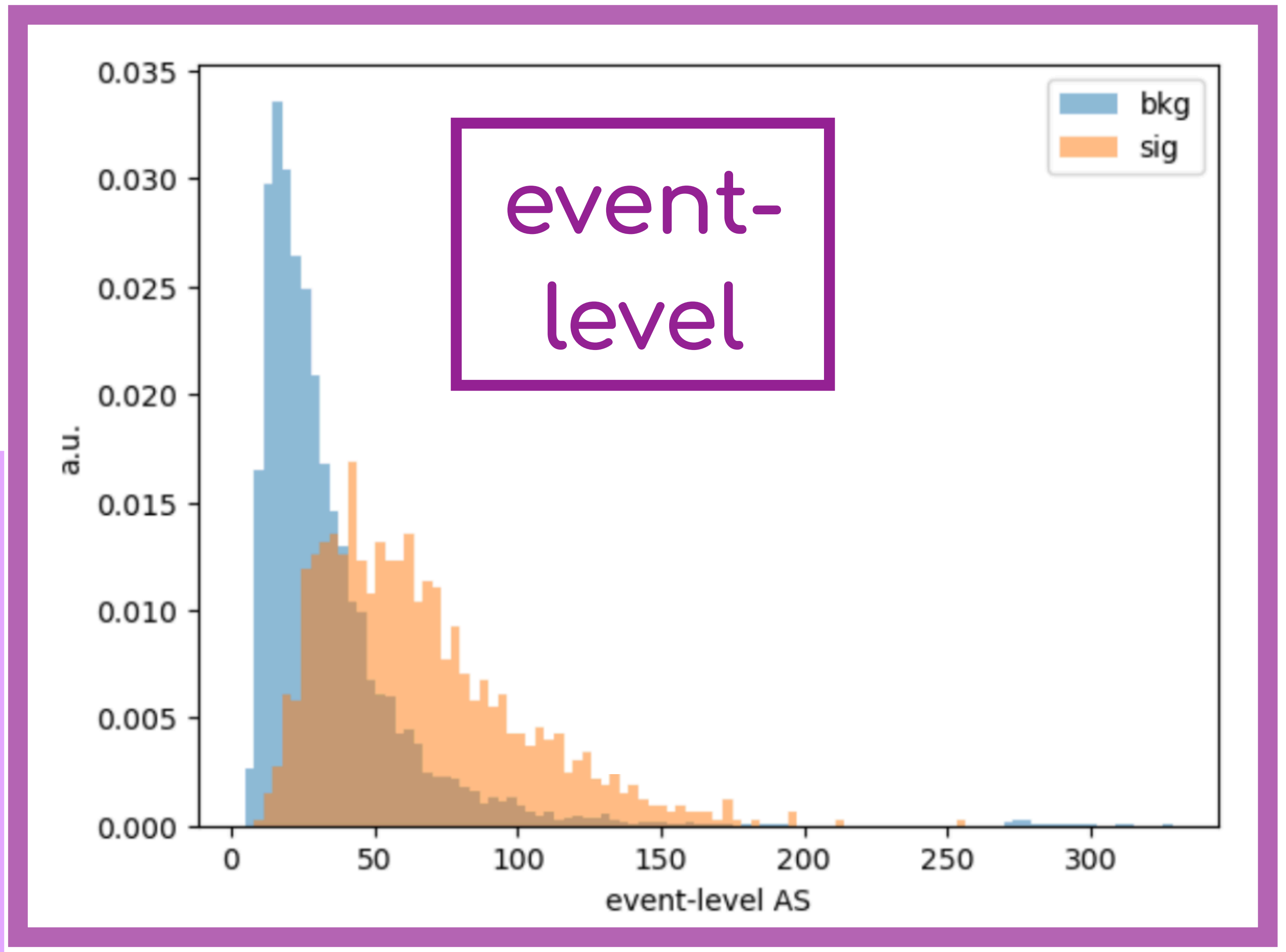
- possibility of recombining the AS of each jet of the event
- sum of the AS of single jets

UNsupervised EGAT training



- more stable training, no overtraining reached
- AUC jet-level: **75.5%**
- AUC event-level: **81.8%**





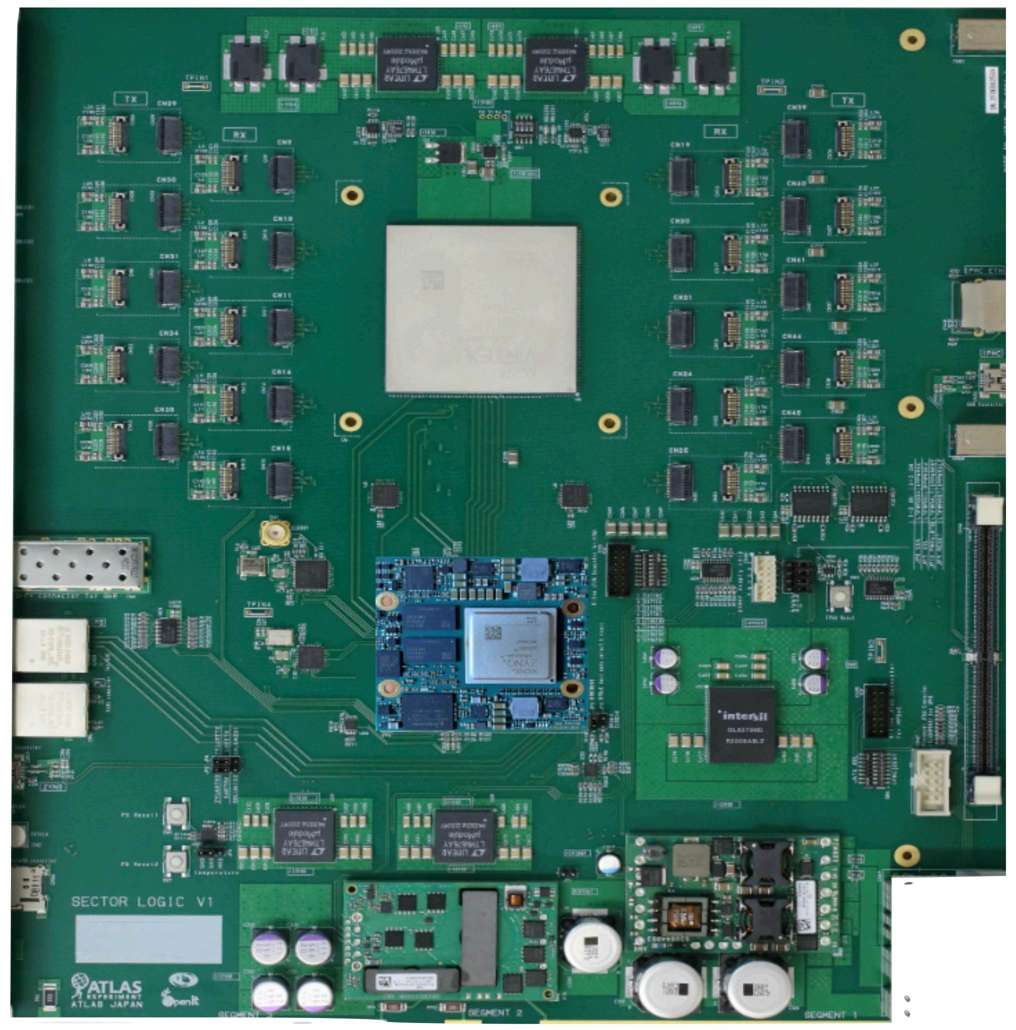
Model summary

Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *unsupervised*	GIN *unsupervised*	EGAT *unsupervised*
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

Model summary

Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *unsupervised*	GIN *unsupervised*	EGAT *unsupervised*
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

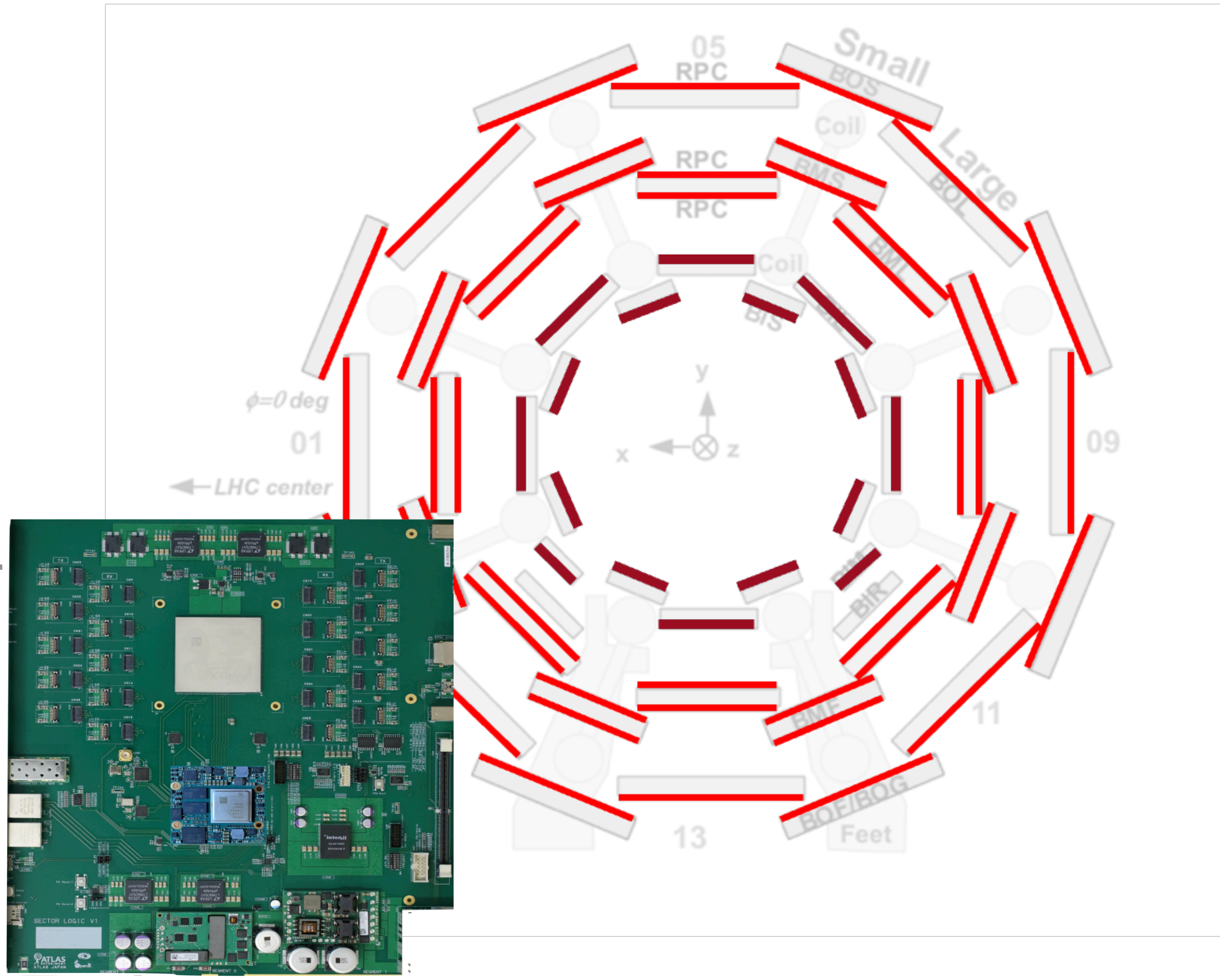
Structure of the talk

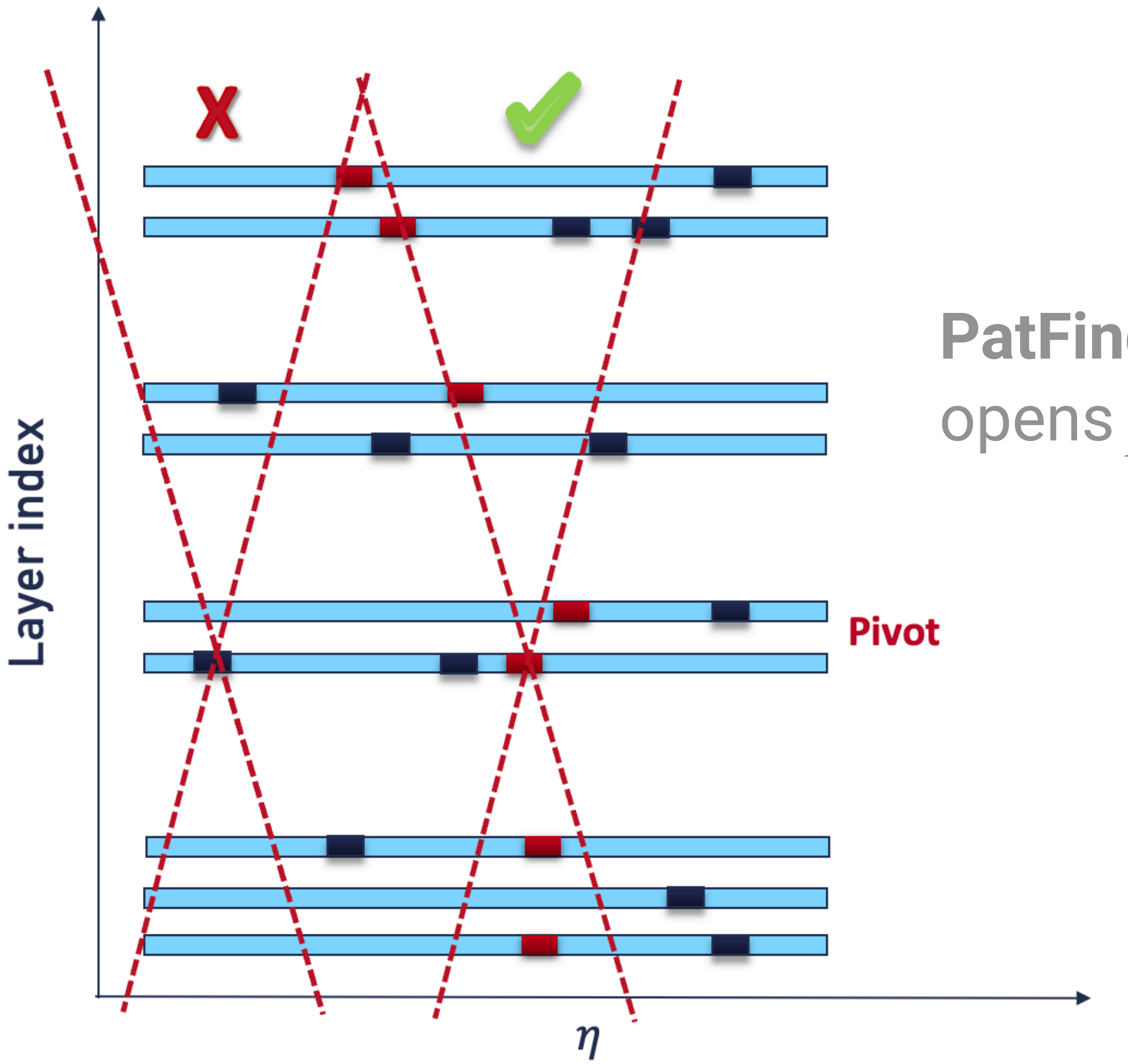


1. Basics of collision experiments (ATLAS)
2. Anomaly Detection with GNNs analysis
3. CNN-based trigger algorithms for ATLAS upgrade

L0 muon trigger

- based on 3 stations, 6 layers of Resistive Plate Chambers (RPCs)
- Phase II upgrade (2026-29): **additional RPC inner station** (4 stations, 9 layers) + Field Programmable Gate Array **(FPGA)-based sector logic** for fast inference $o(100\ ns)$





PatFinder, the traditional algorithm
opens p_T-dependent η and ϕ windows and searches for a muon pattern

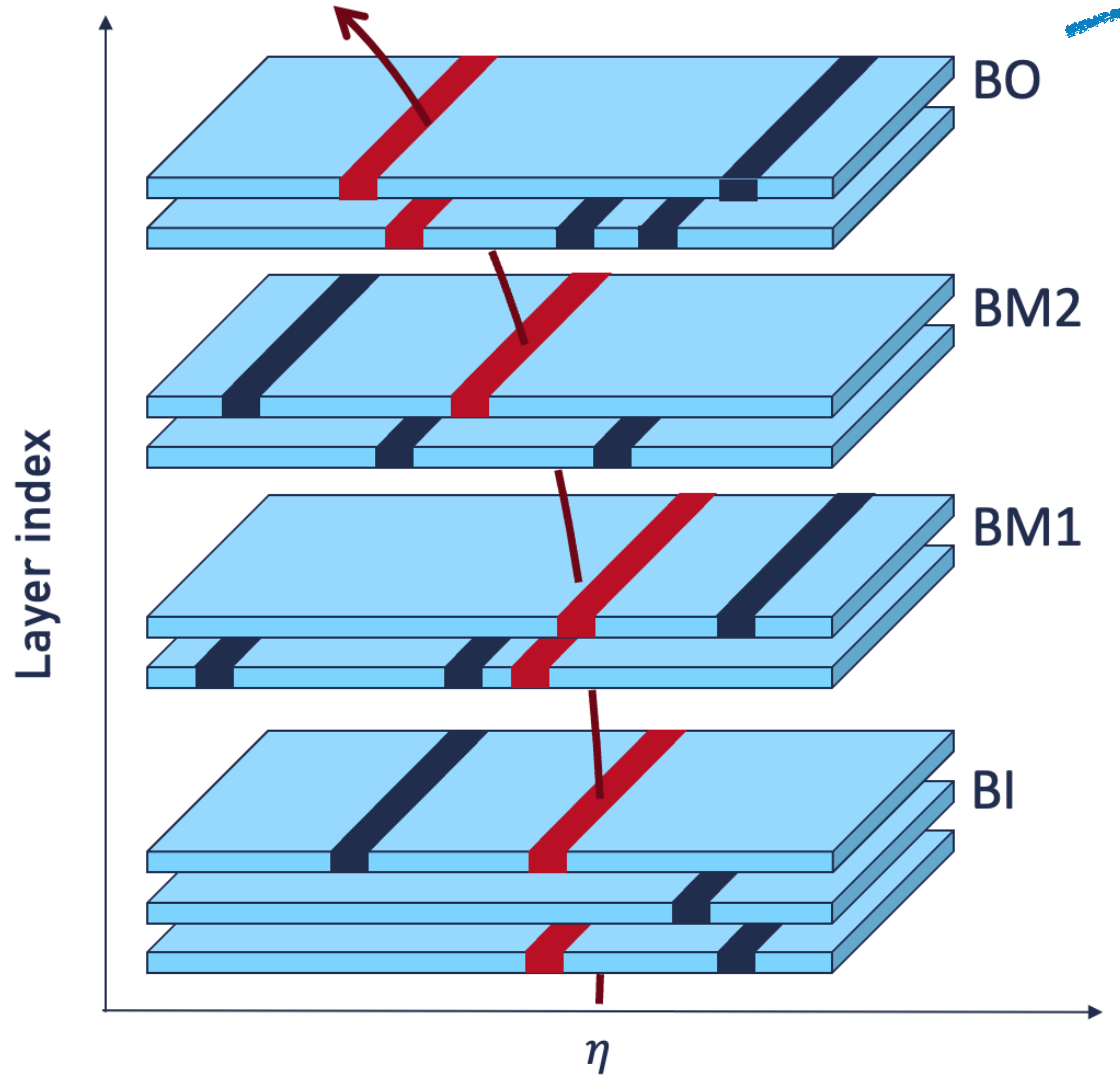
BUT relies on detector geometry!

- efficiency saturation
- windows fine tuning
- primary vertex muons only
- partial tracks?

ML approach

Trick: muon track reinterpreted as a black-and-white image

and used for a Convolutional Neural Network (CNN)



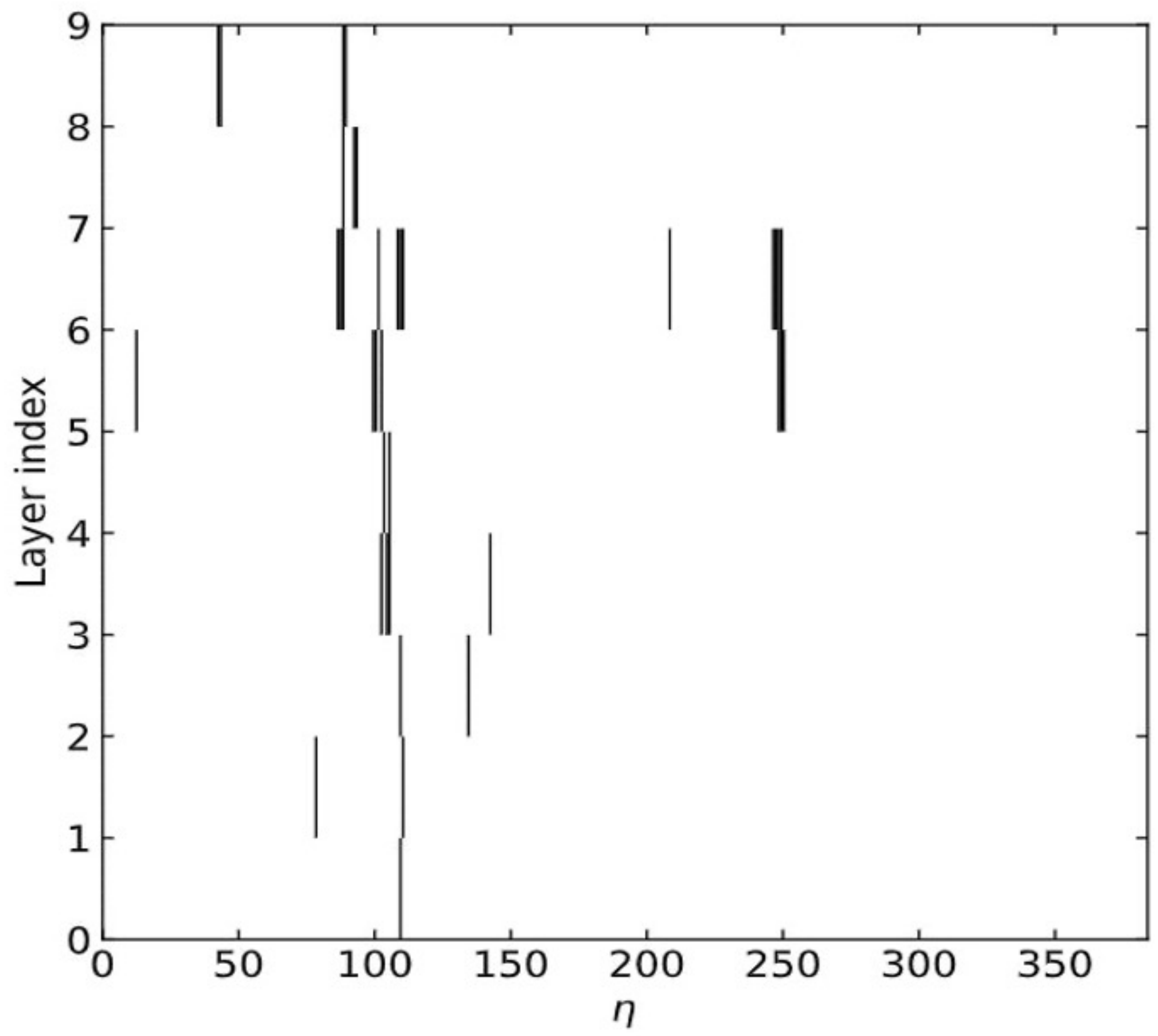
Dataset of $\sim$200k images with:

- random noise-only hits
- single muon tracks

  ($p_T \in [2.5, 20]$ GeV)

  + noise hits

Target label:
$p_T, \eta, (q)$



Compression techniques

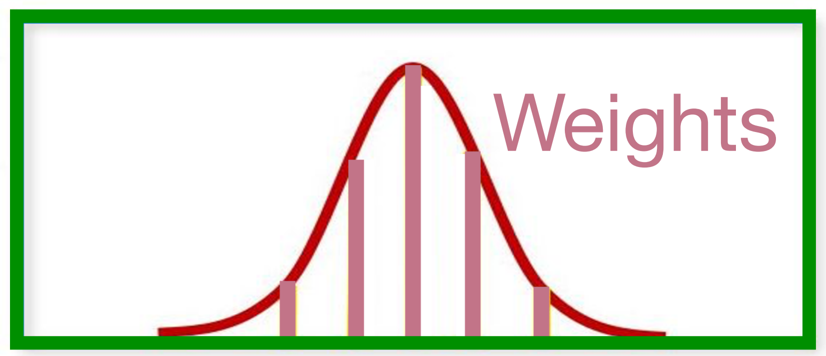
Experimental requirements

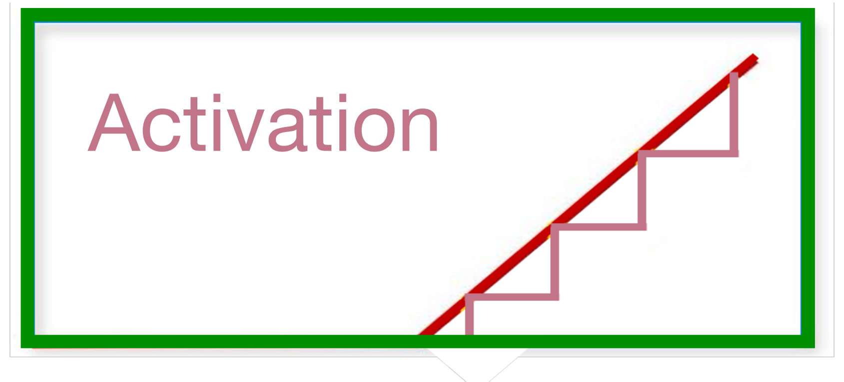
- Fit within the Virtex UltraScale+ 13 FPGA resources;
- Maximum latency (= time interval of algorithmic response) allowed of $\sim$400 ns;
- Fake efficiency (= trigger efficiency on noisy events) < 2‰

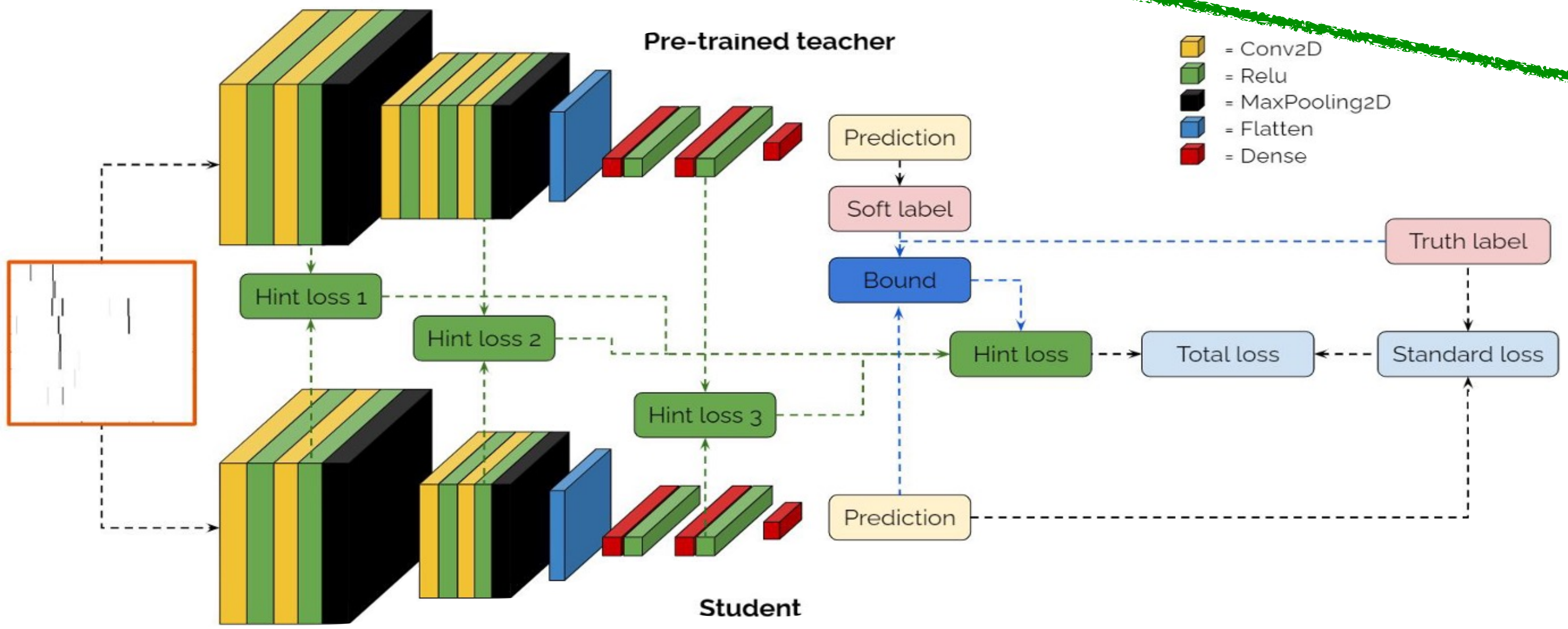
Train a CNN model (VGG-like)
+ help from compression techniques

Quantization

QKeras







Knowledge Distillation

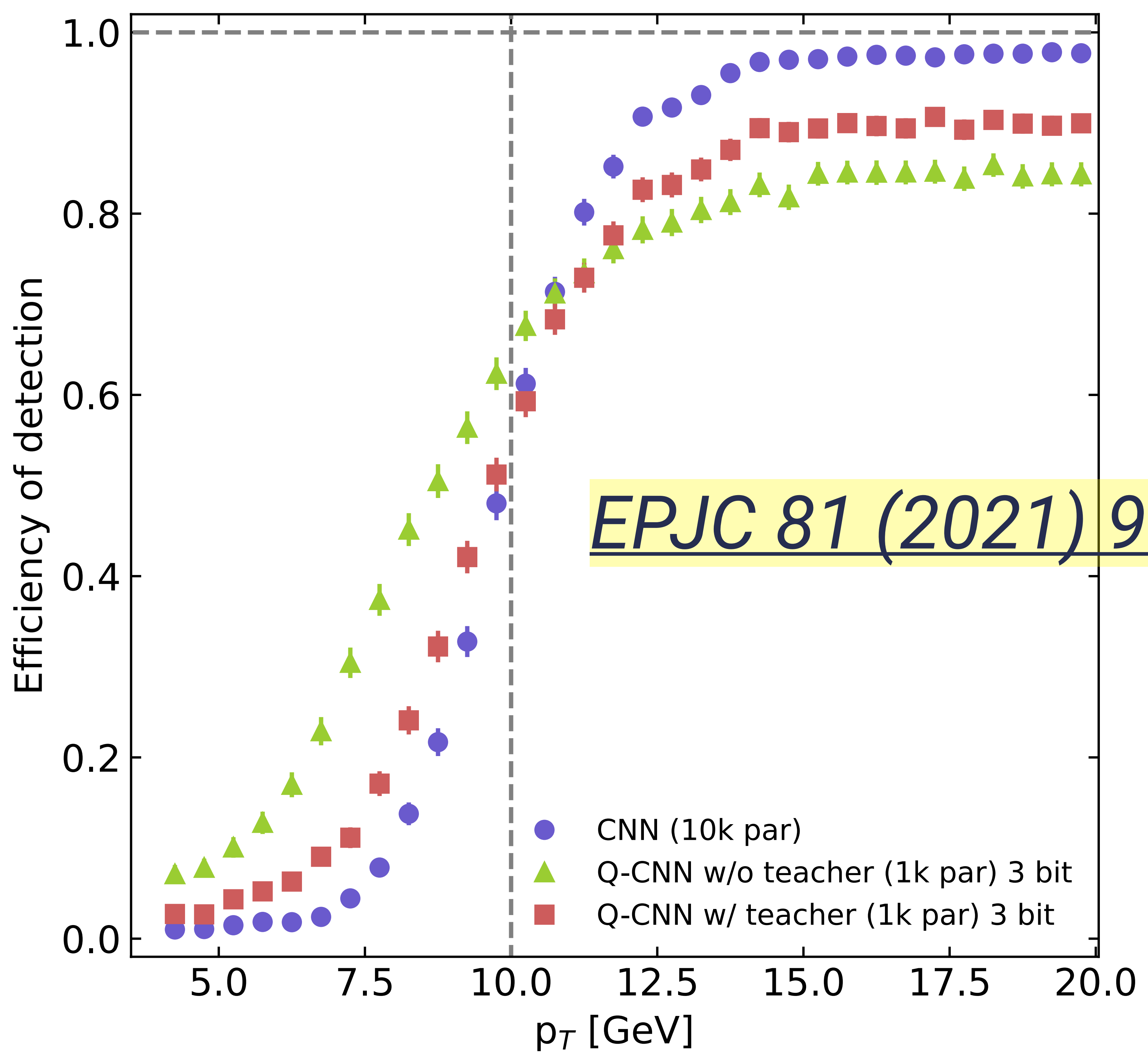
arXiv:1711.05852

arXiv:1804.03235

Single muon results

ultra-small CNN models
$\sim$700 parameters

Efficiency curve



EPJC 81 (2021) 969

synthesis on XCV13P FPGA performed by using *HLS4ML* library and by-hand VHDL implementation

	GPU Tesla V100	FPGA XCV13P with hls4ml	FPGA XCV13P with VHDL implementation
Latency	5 ms	438 ns	84 ns

Requirements are all satisfied thanks to the compression

- Resources occupancy: mix of quantization and KD
- Maximum latency: KD and VHDL implementation
- Fake efficiency < 2‰: quantization

Conclusion & perspectives

Anomaly Detection analysis

- model independent search, more general
- finalising the best ML model
- moving to Run3 ATLAS data, adding the tracks information

L0 ML muon trigger algorithm

- ML algorithm sensitive to larger range of physics
- successful single muon results, moving to 2 muons with p_T, η, charge and #muons prediction

LHC experiments are big data factories, large improvements in using ML tools


Thanks for the attention!
PETZL

More on the trigger system

ATLAS TDAQ public page

The billions of collisions in ATLAS have a combined data volume of more than 60 million megabytes per second – that's equivalent to **5400 simultaneous streams of 4K video**. However, only some of these events will contain interesting characteristics that might lead to new discoveries. To reduce the flow of data to manageable levels, ATLAS uses a special event selection system – the "trigger" – which picks events with distinguishing characteristics for physics analyses.

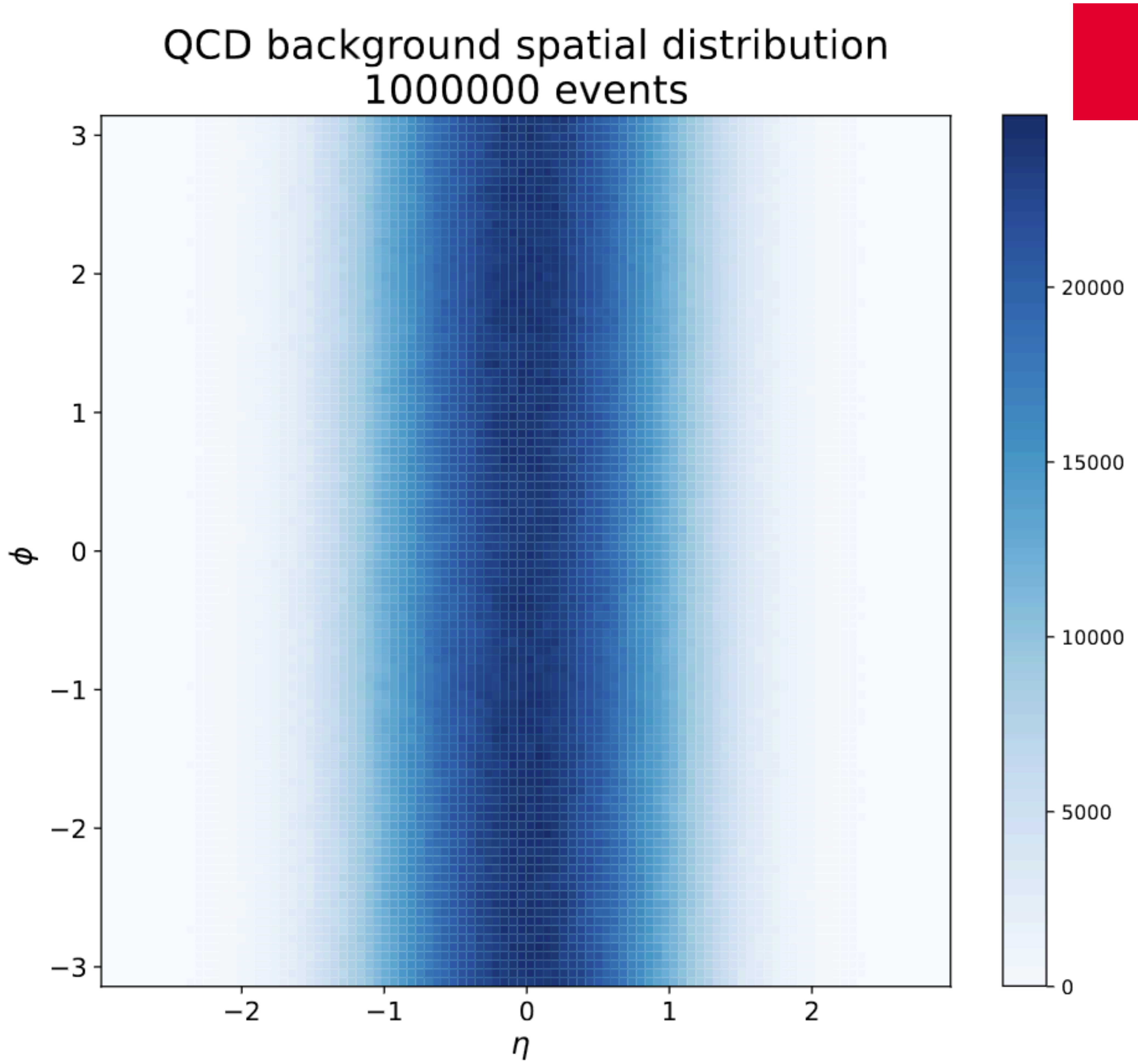
The ATLAS trigger system carries out the selection process in two stages.

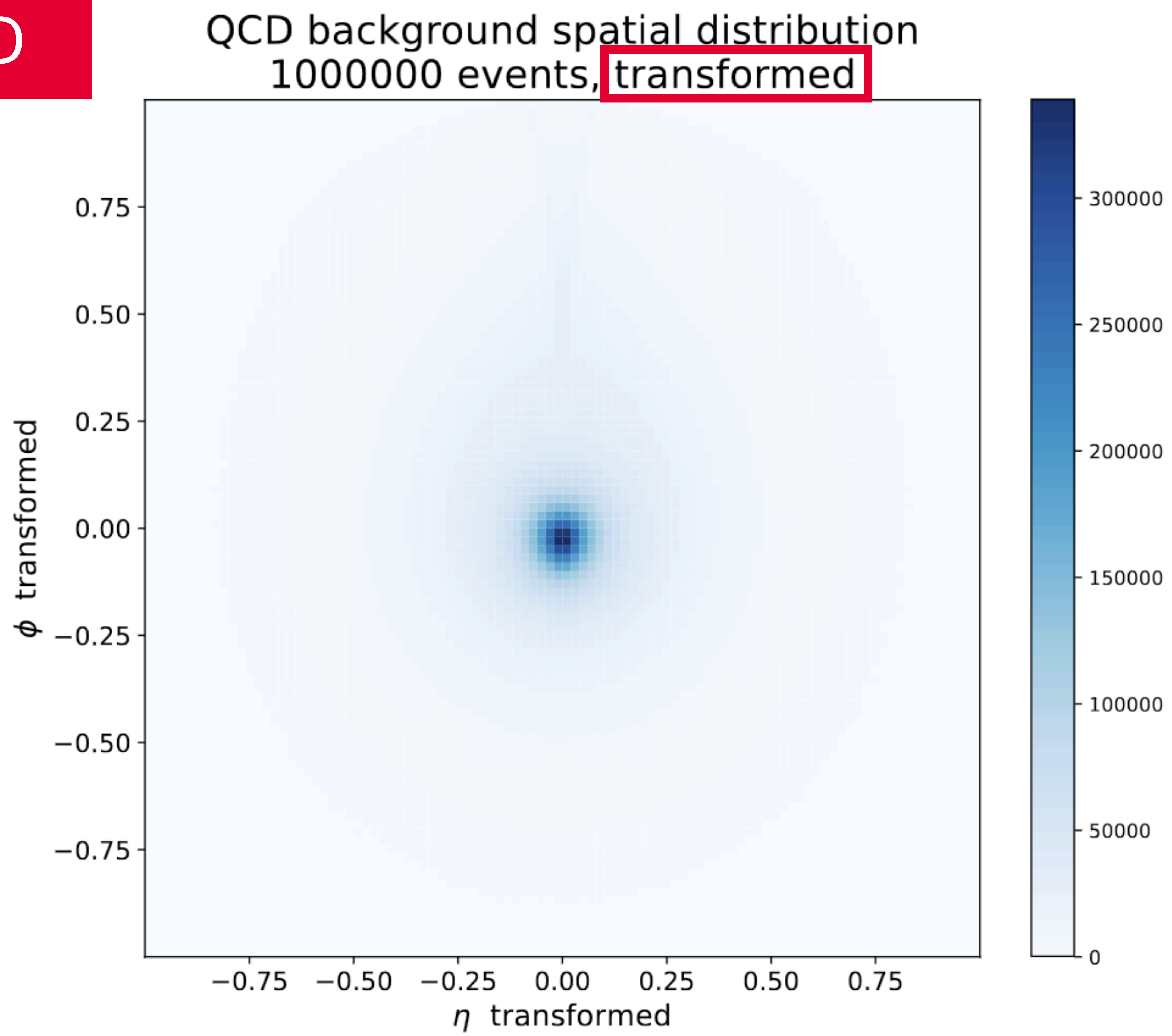
1. The first-level hardware trigger, constructed with custom-made electronics located on the detector, works on a subset of information from the calorimeters and the Muon Spectrometer. The decision to keep the data from an event is made less than **2.5 microseconds** after the event occurs. During this time the event data is kept in storage buffers. If the event is selected it is passed on to the second-level trigger, which can accept up to **100,000 events per second**.

2. The second-level software trigger operates from a large farm of about **40,000 CPU cores**. In just 200 microseconds, it conducts very detailed analyses of each collision event, examining data from specific detector regions. The second-level trigger finally selects about **1000 events per second** and passes them on to a data storage system for offline analysis.

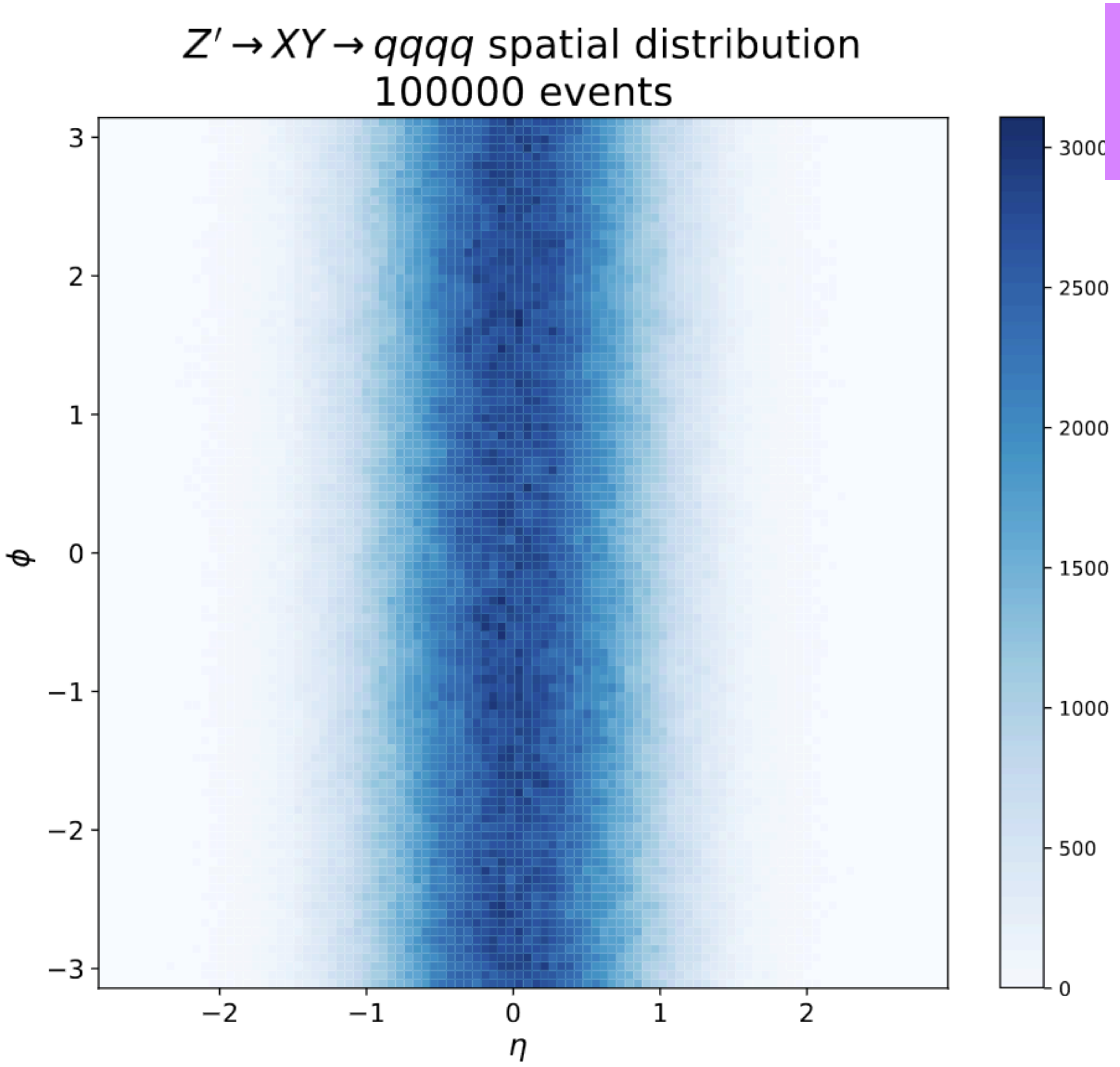
arXiv:1903.02032v3

QCD

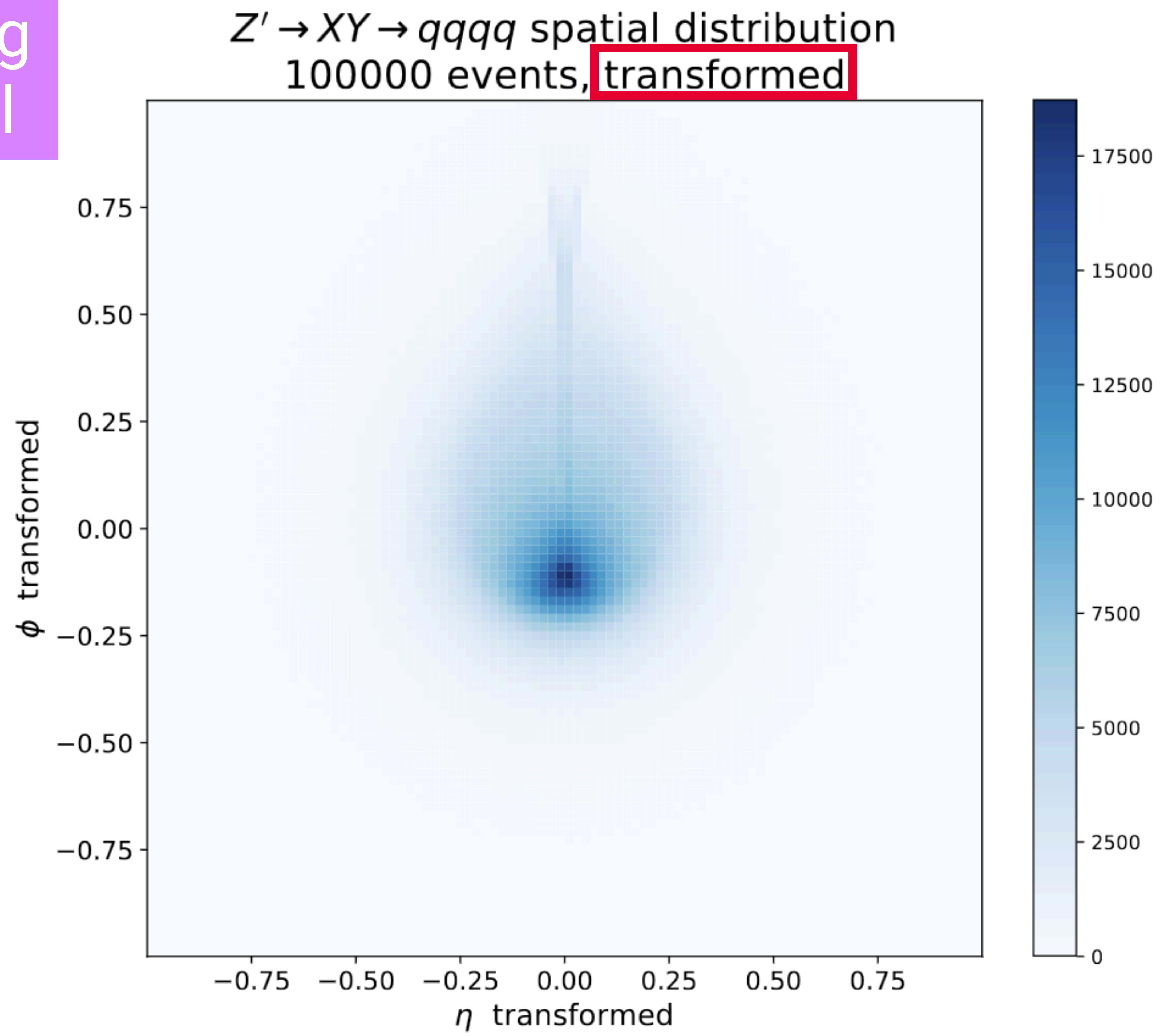




arXiv:1903.02032v3

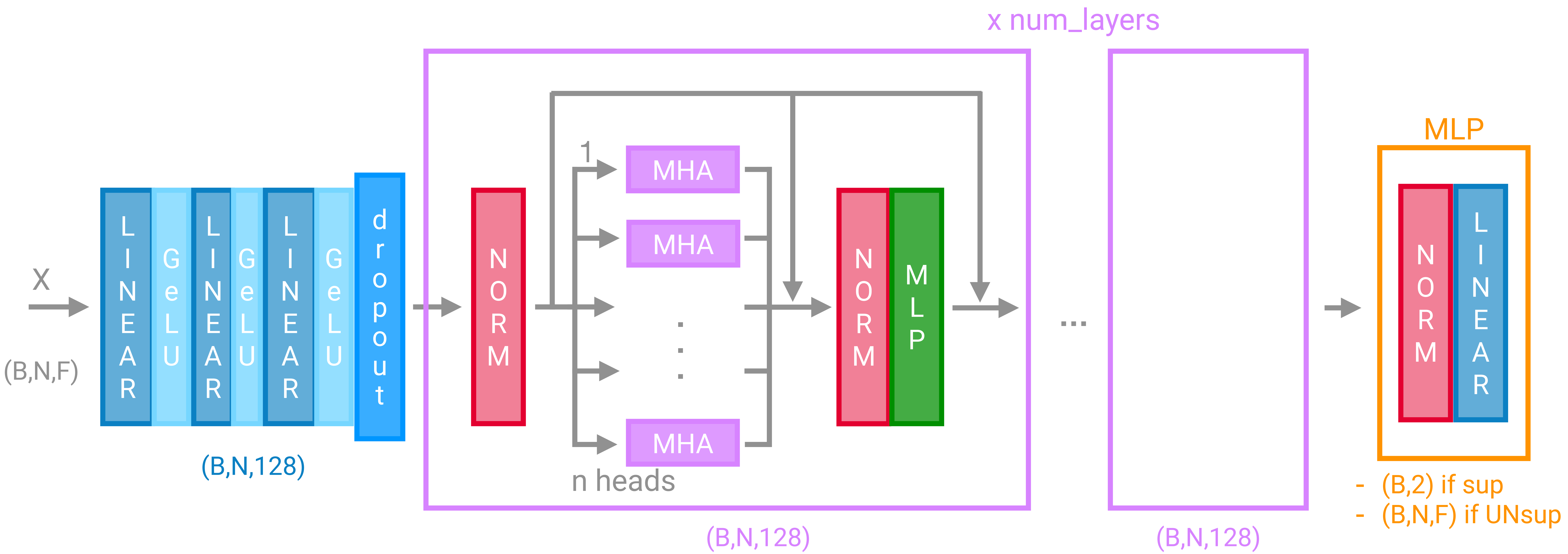


2prong signal



Transformer architecture

B = batch size
N = #constituents (50)
F = # features (3)



Transformer trainings

supervised loss:
Cross Entropy loss

Used as benchmark for comparison

$$L_{CE} = -(\boxed{y}\log(\boxed{p}) + (1 - y)\log(1 - p))$$

output of neural network

label of input

$$\textit{Anomaly Score} = Prob(y = 1)$$

un**supervised loss:**
MSE loss

actual AD approach

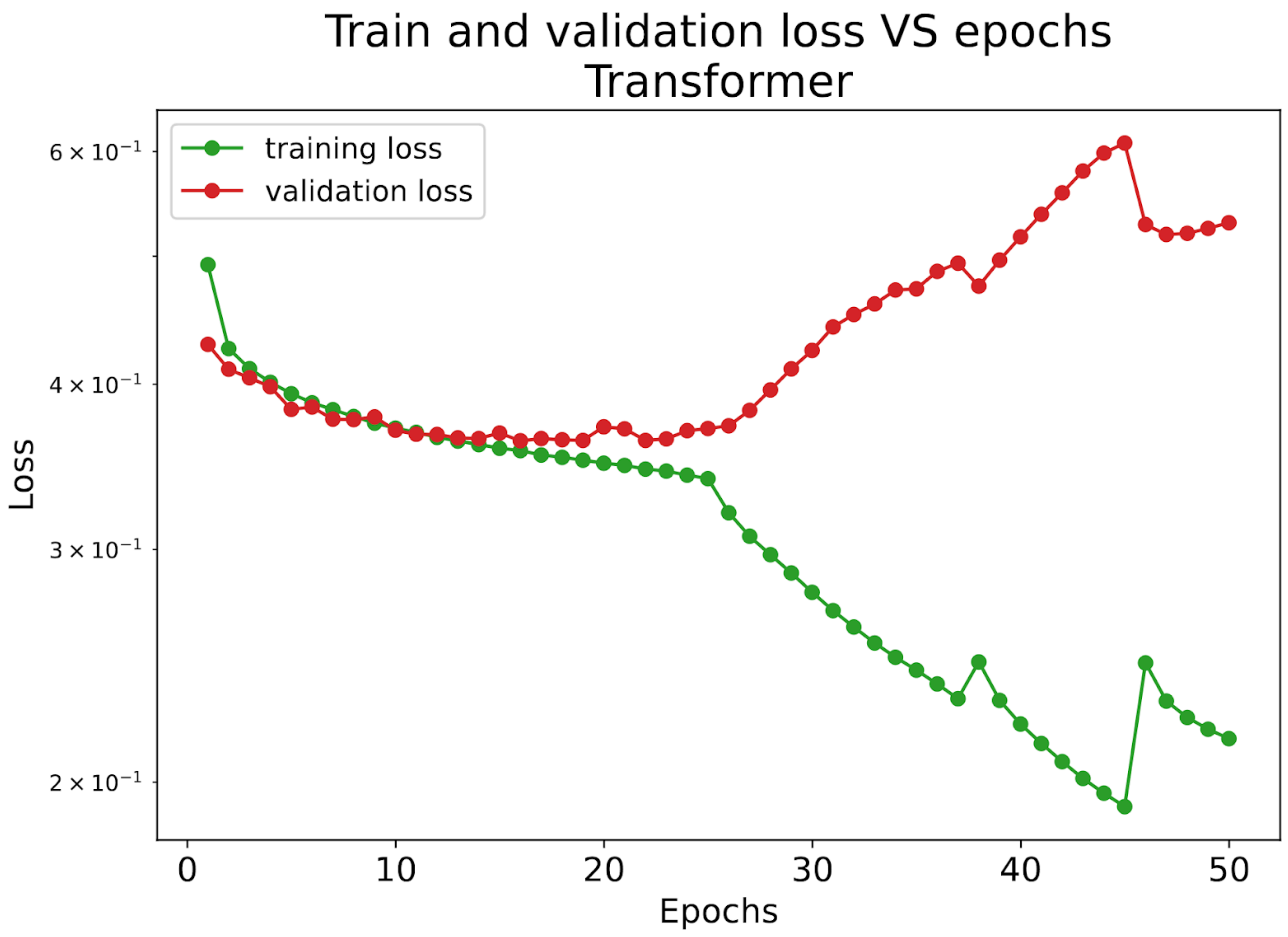
$$L_{MSE} = ||\boxed{x} - \boxed{y}||^2$$

input of neural network

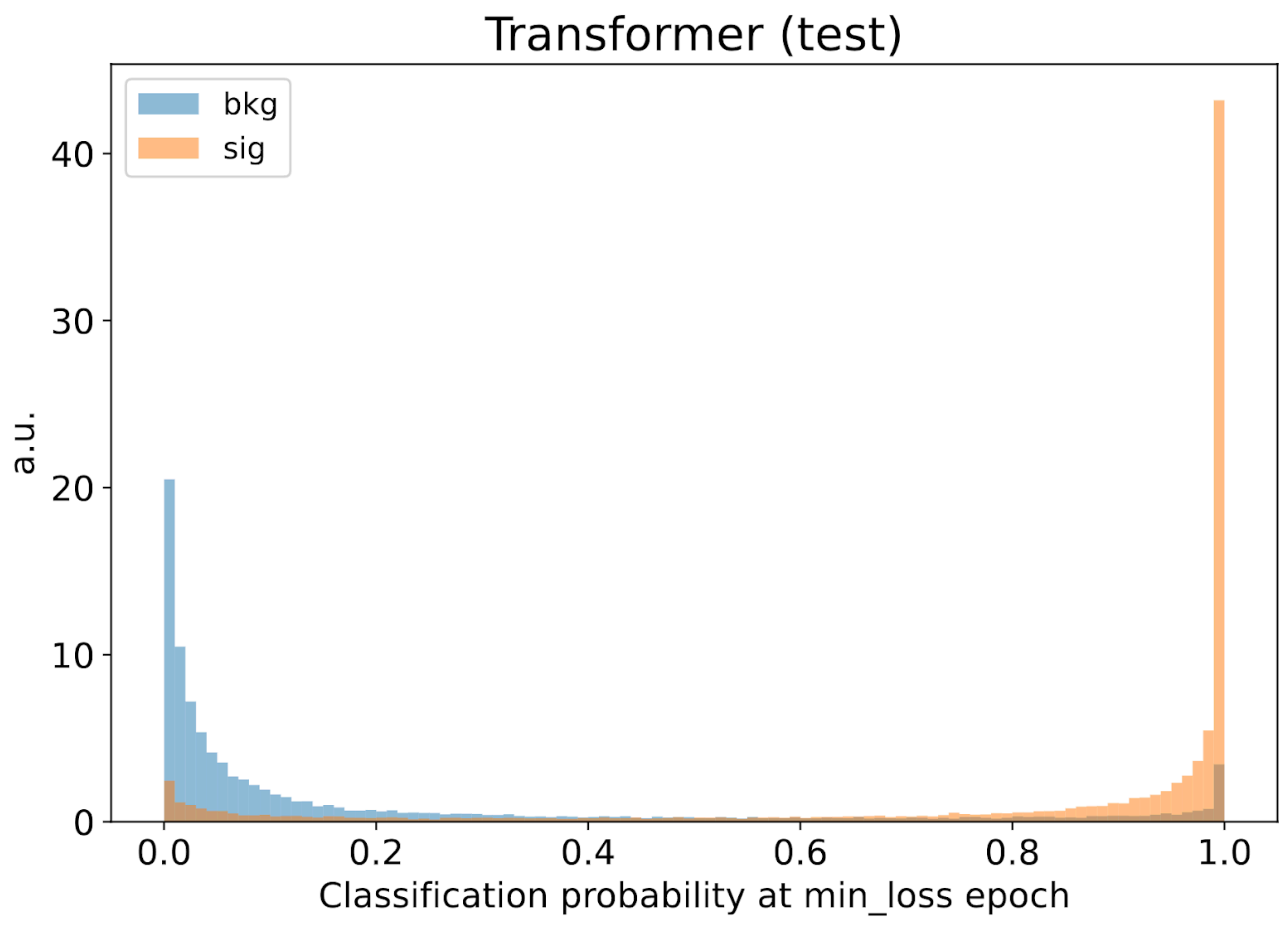
output of neural network

$$\textit{Anomaly Score} = ||x - y||^2$$

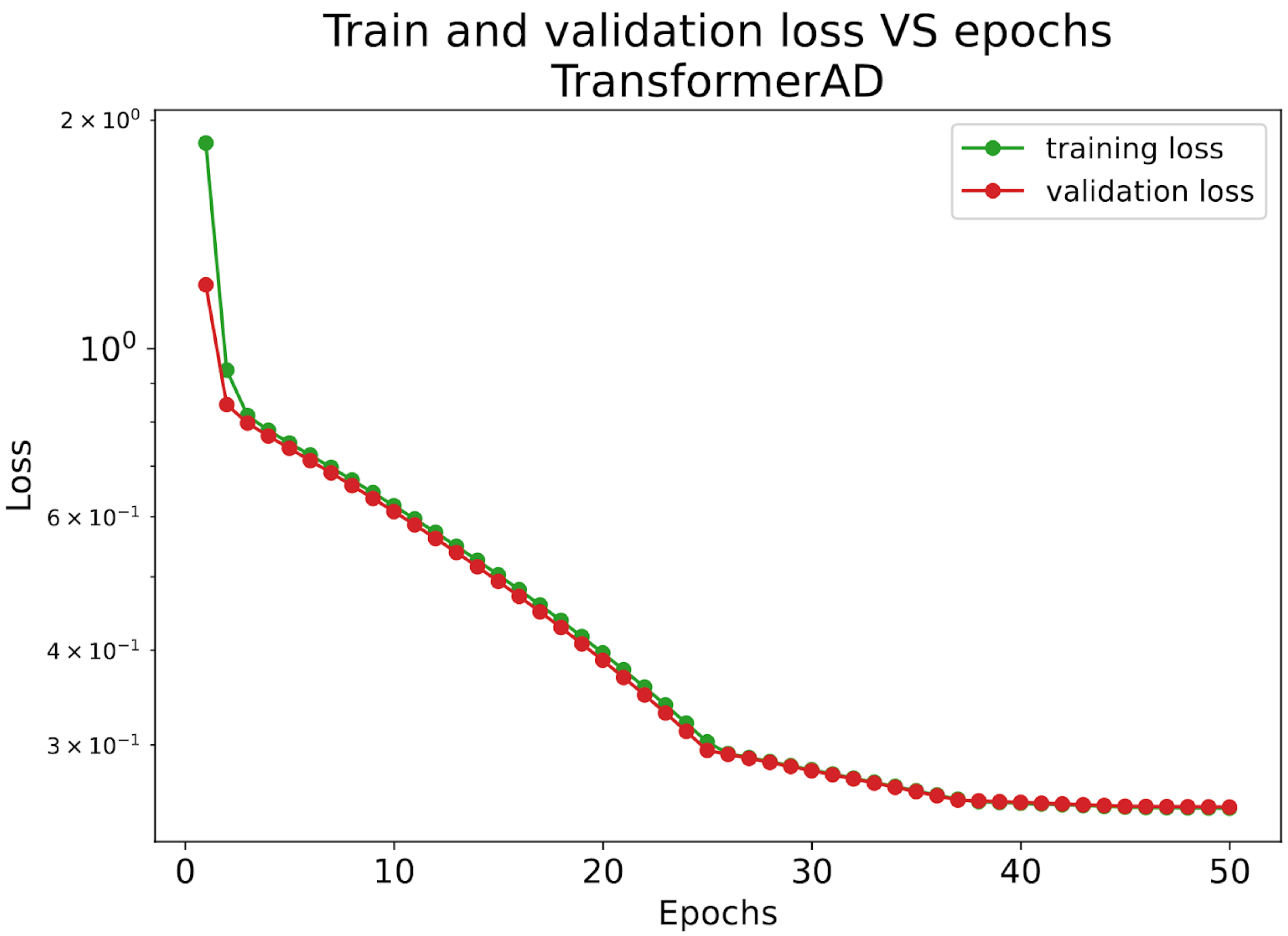
Supervised Transformer training



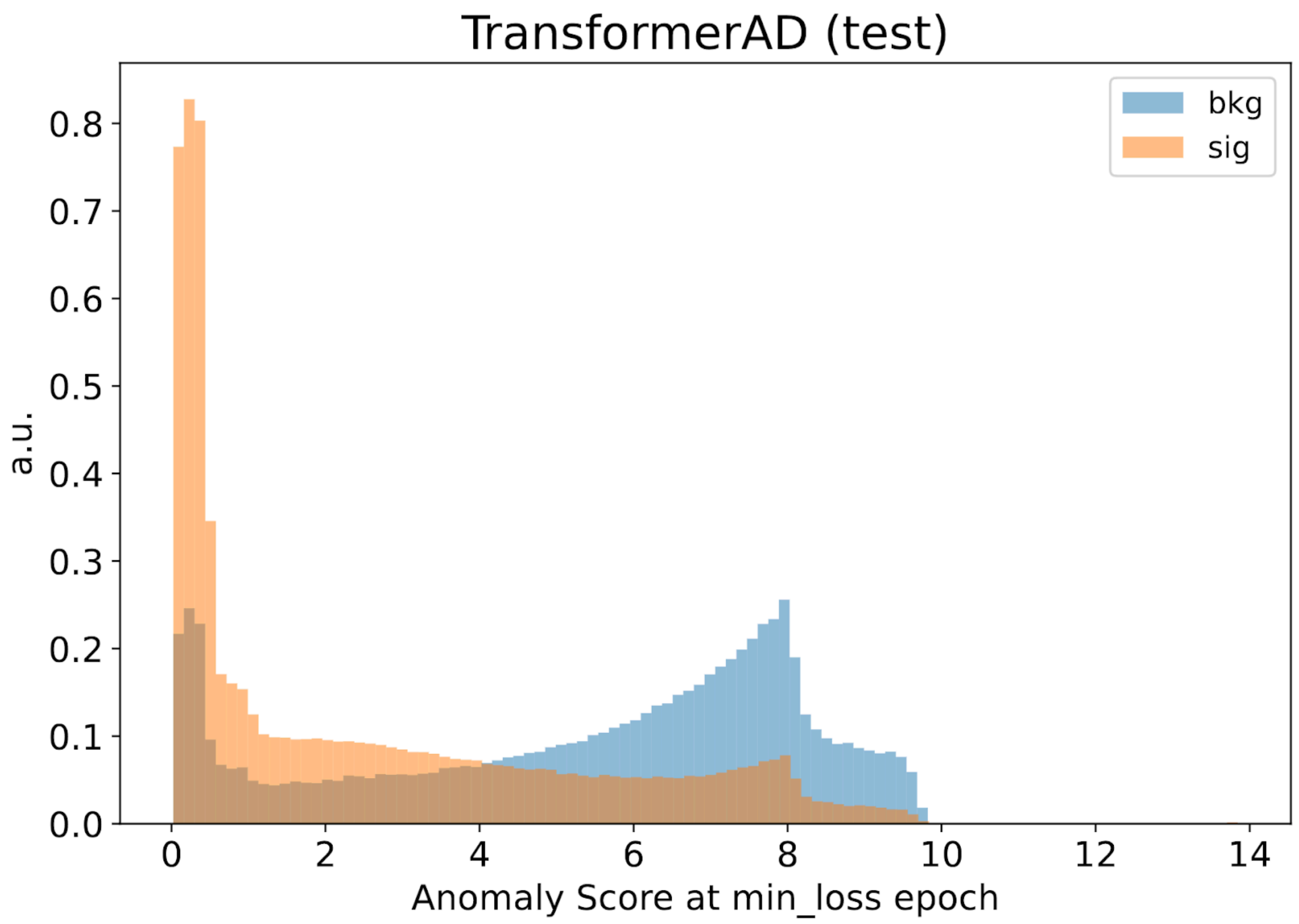
- unstable training, overtraining at epoch ~20 but epoch with validation minimal loss is reached before
- AUC: **88.3%** (2prong)



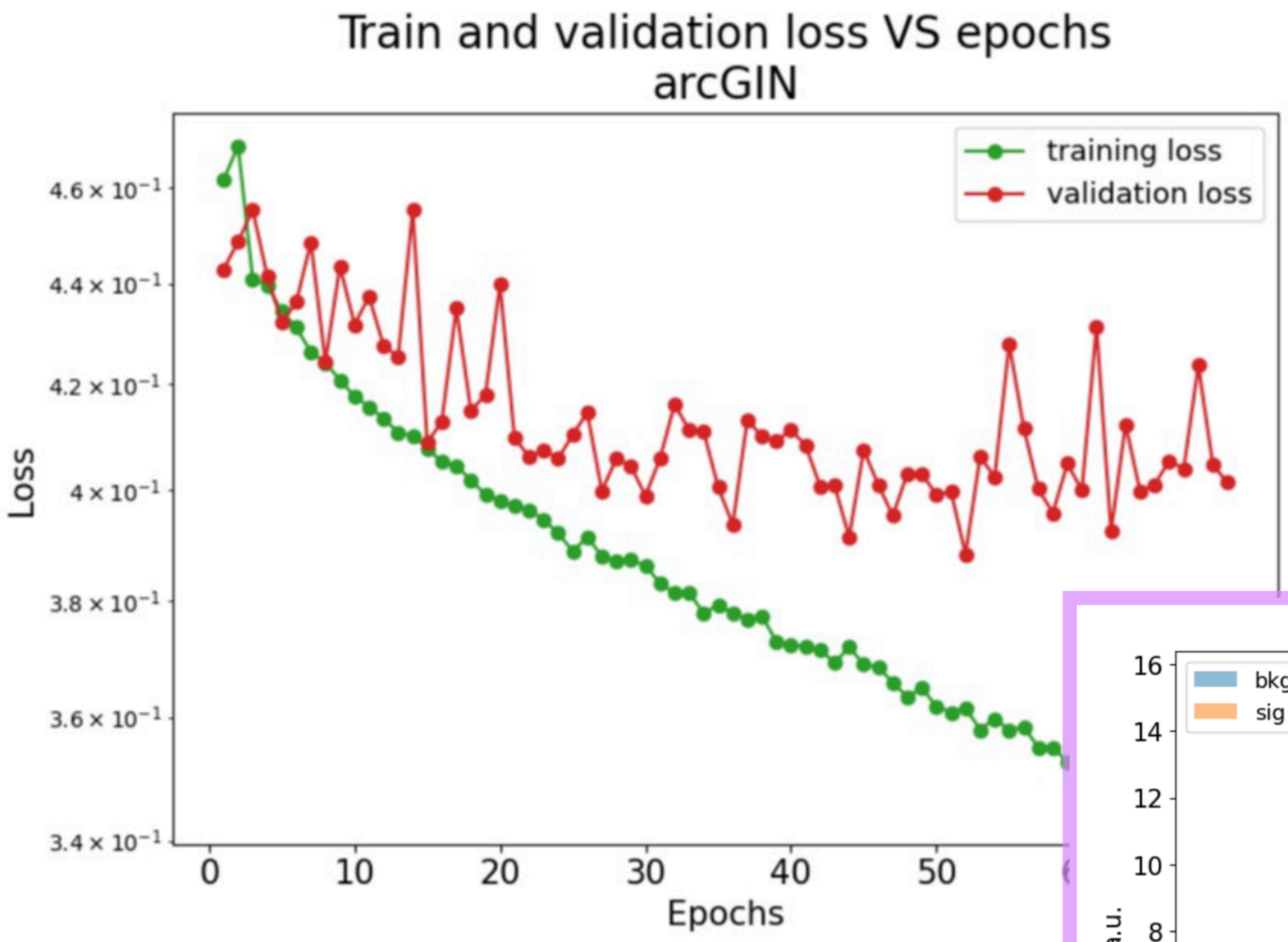
UNsupervised Transformer training



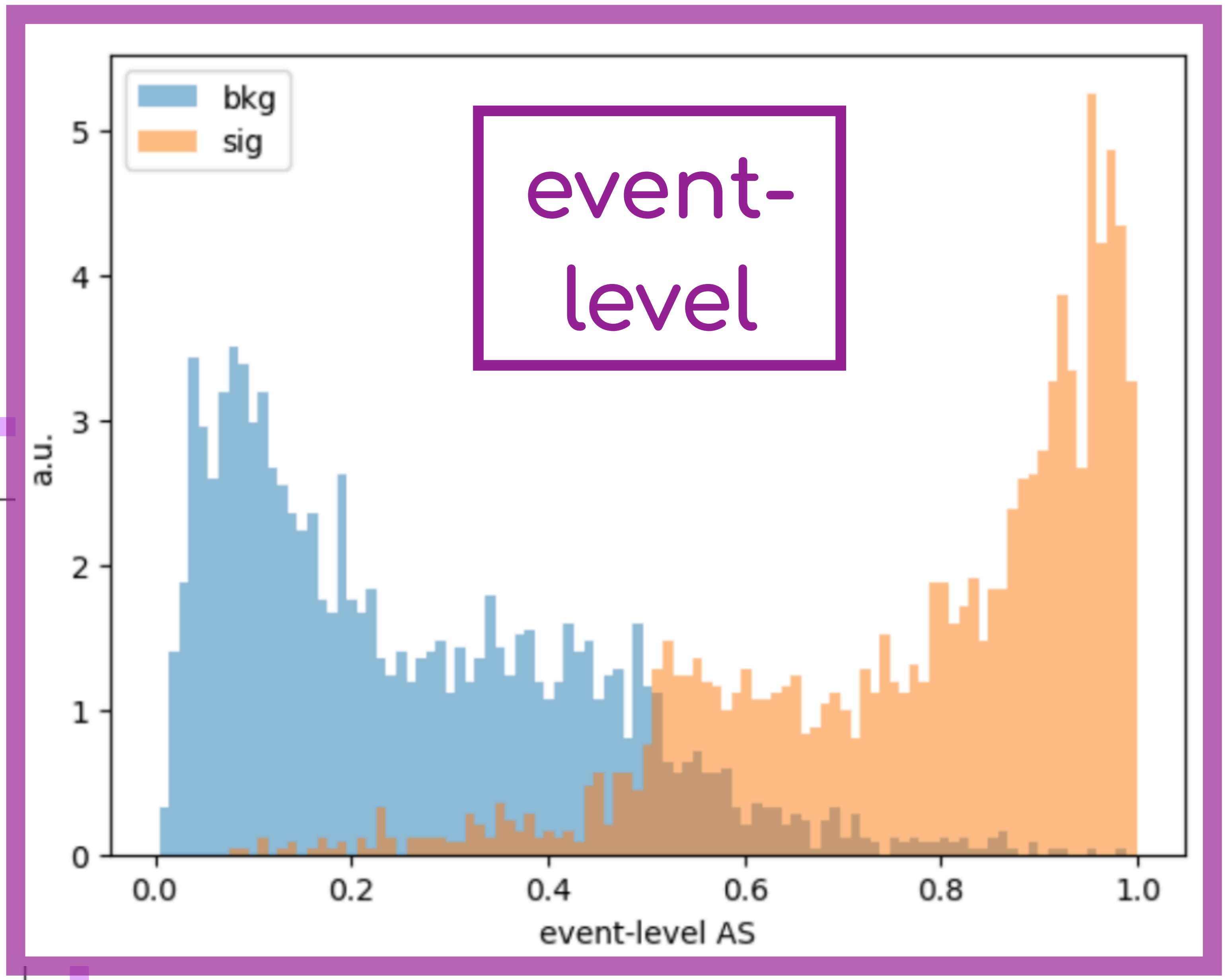
- stable training, almost no overtraining
- AUC: **75.5%** (2prong)

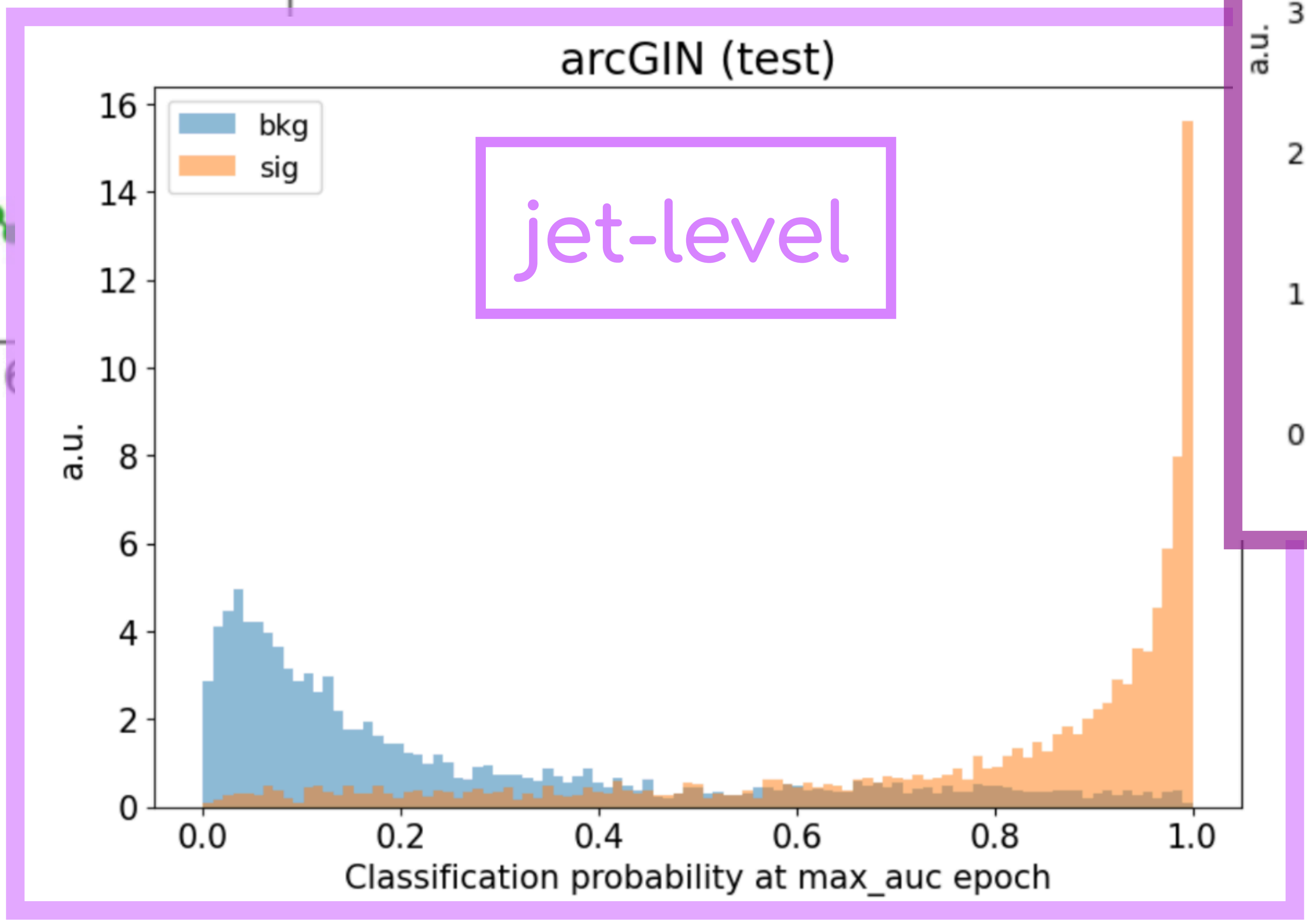


Supervised GIN training

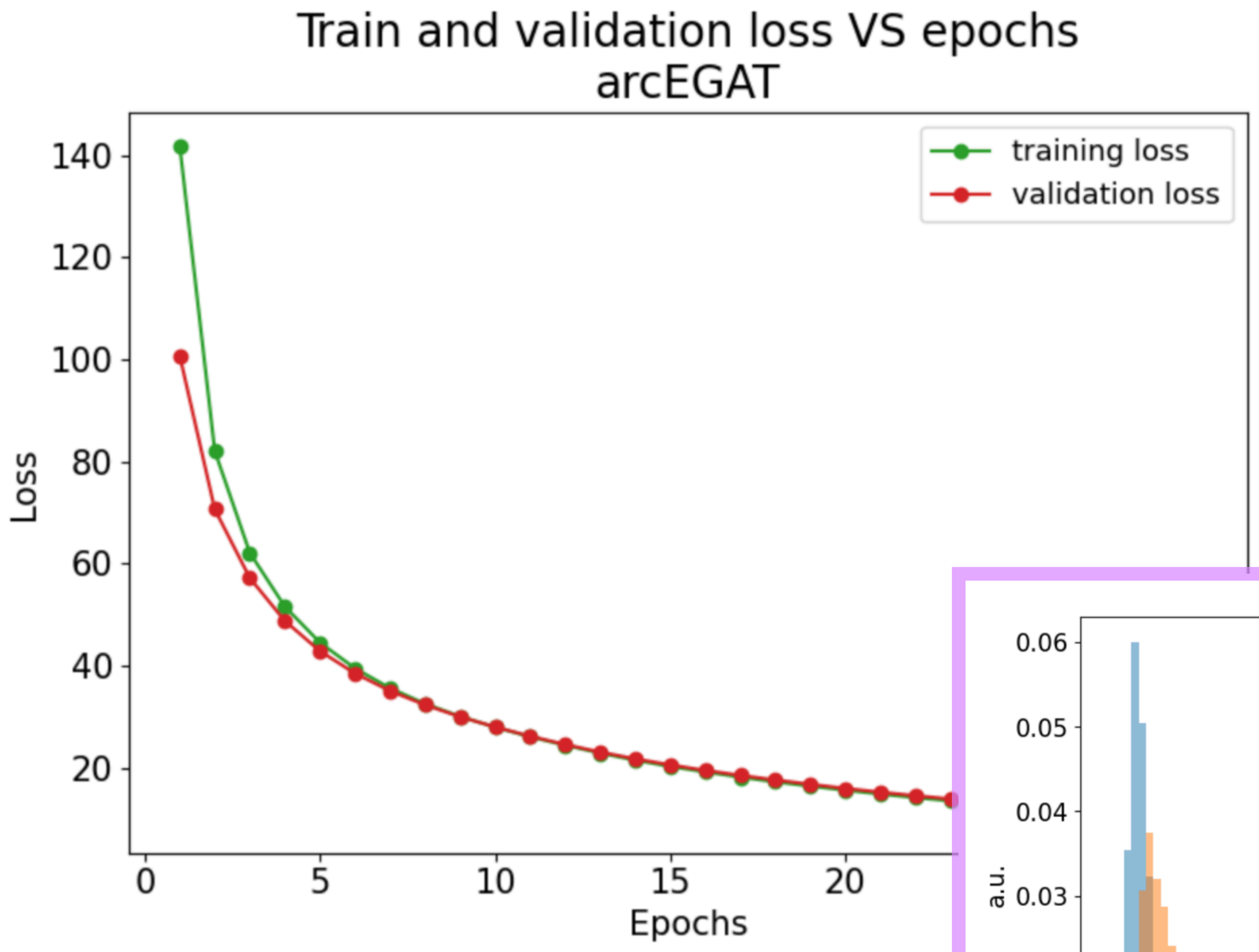


- quite stable training, very soon overtraining
- AUC jet-level: **90.2%** (2prong)
- AUC event-level: **96.5%** (2prong)

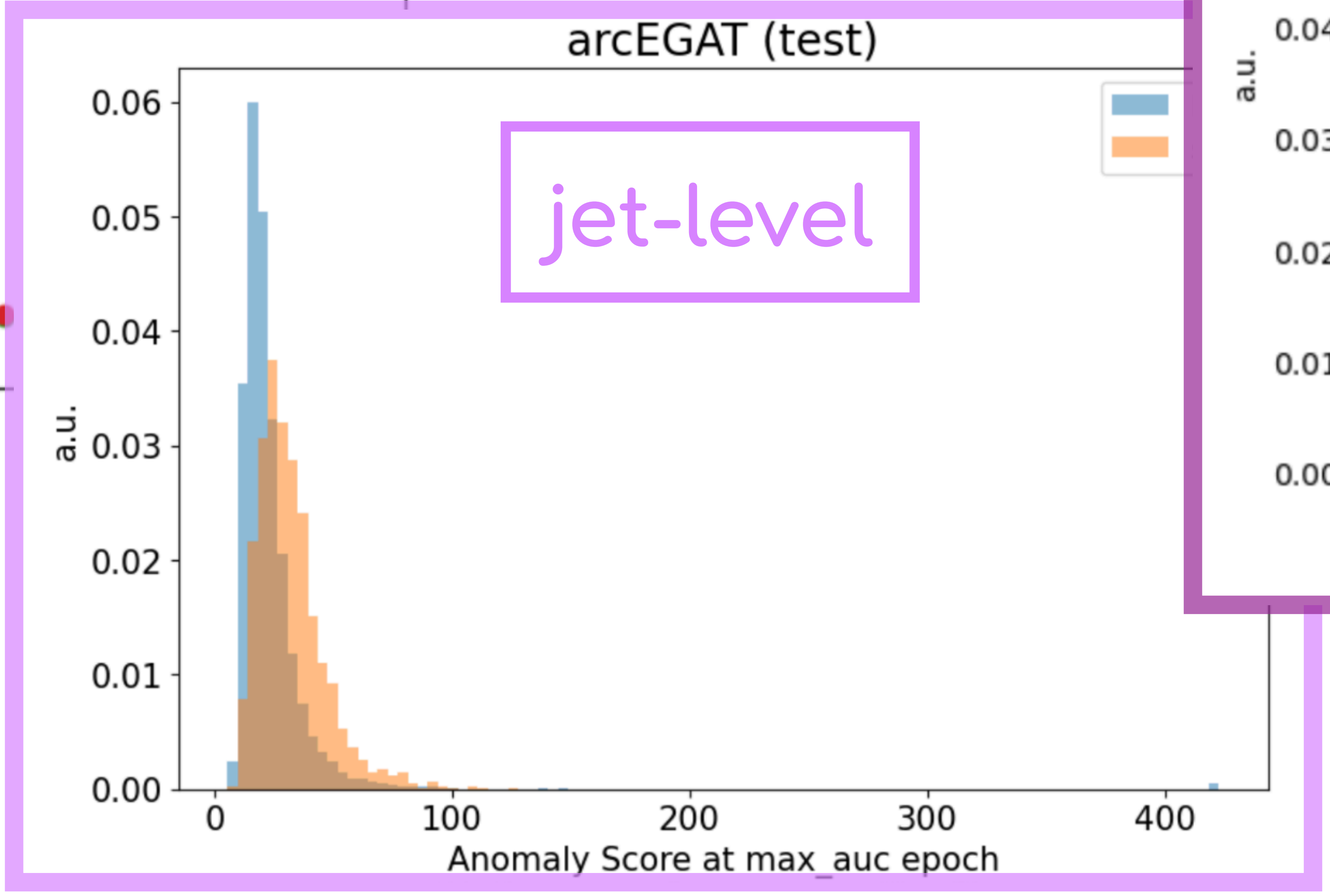


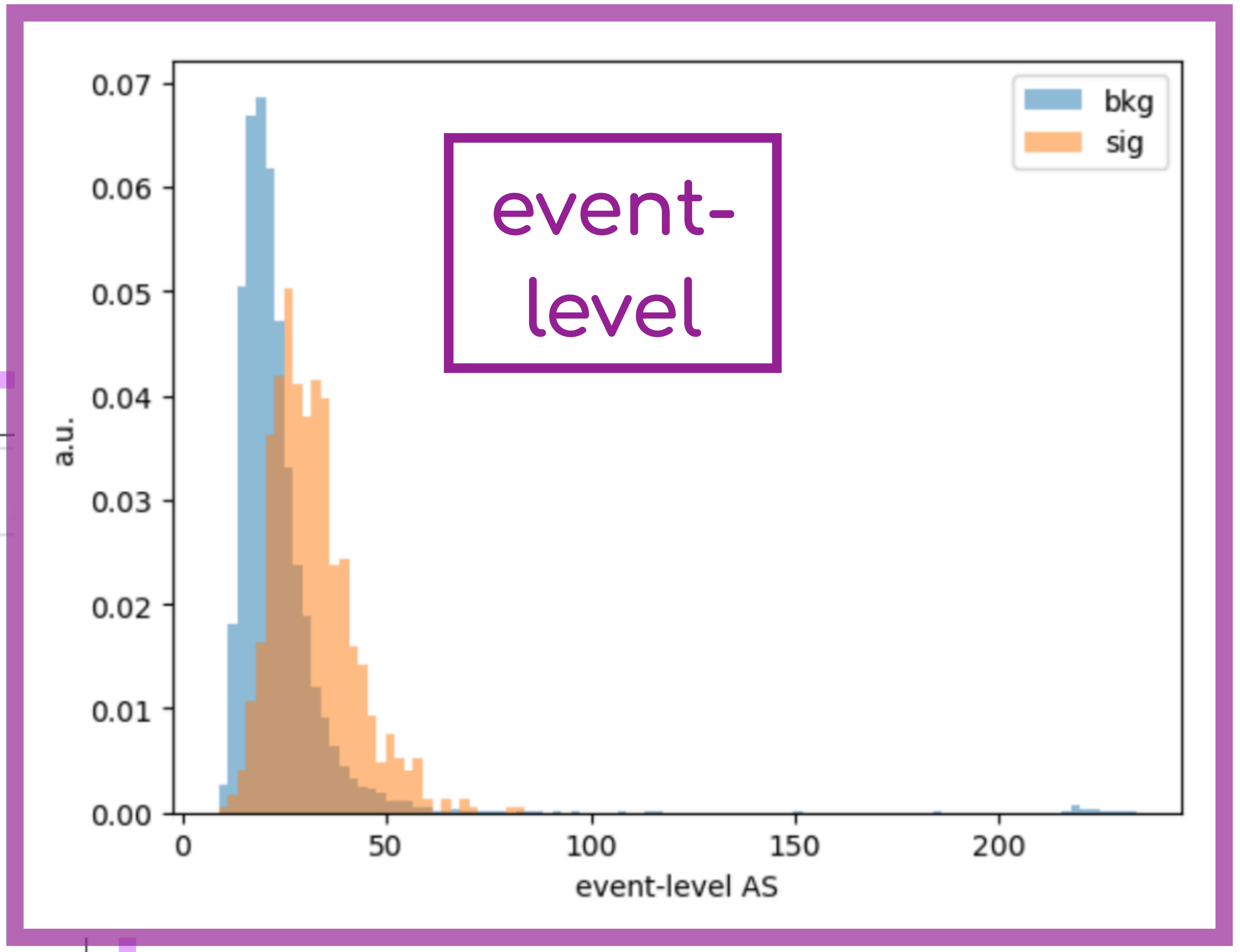


UNsupervised GIN training



- very stable training, almost no overtraining
- AUC jet-level: **73.7%** (2prong)
- AUC event-level: **79.6%** (2prong)





Supervised EGAT training

- more stable training, overtraining at epoch ~25
- AUC jet-level: **89.9%** (2prong)
- AUC event-level: **96.5%** (2prong)