Anomalies, representations and Self-Supervision

Luigi Favaro

Pheno2023
Pittsburgh - 09/05/2023
based on arXiv:2301.04660

UNIVERSITÄT HEIDELBERG
Zukunft. Seit 1386.

Model-agnostic searches & ML

Model-agnostic searches & ML

- Are we leaving stones unturned? Can we answer this question only via direct searches?
- **Anomaly searches**: define background from the data and find “anomalous” events

Model-agnostic searches & ML

- Are we leaving stones unturned? Can we answer this question only via direct searches?
- **Anomaly searches**: define background from the data and find "anomalous" events

a known problem in Machine Learning (**or not?**)
what we are looking for:

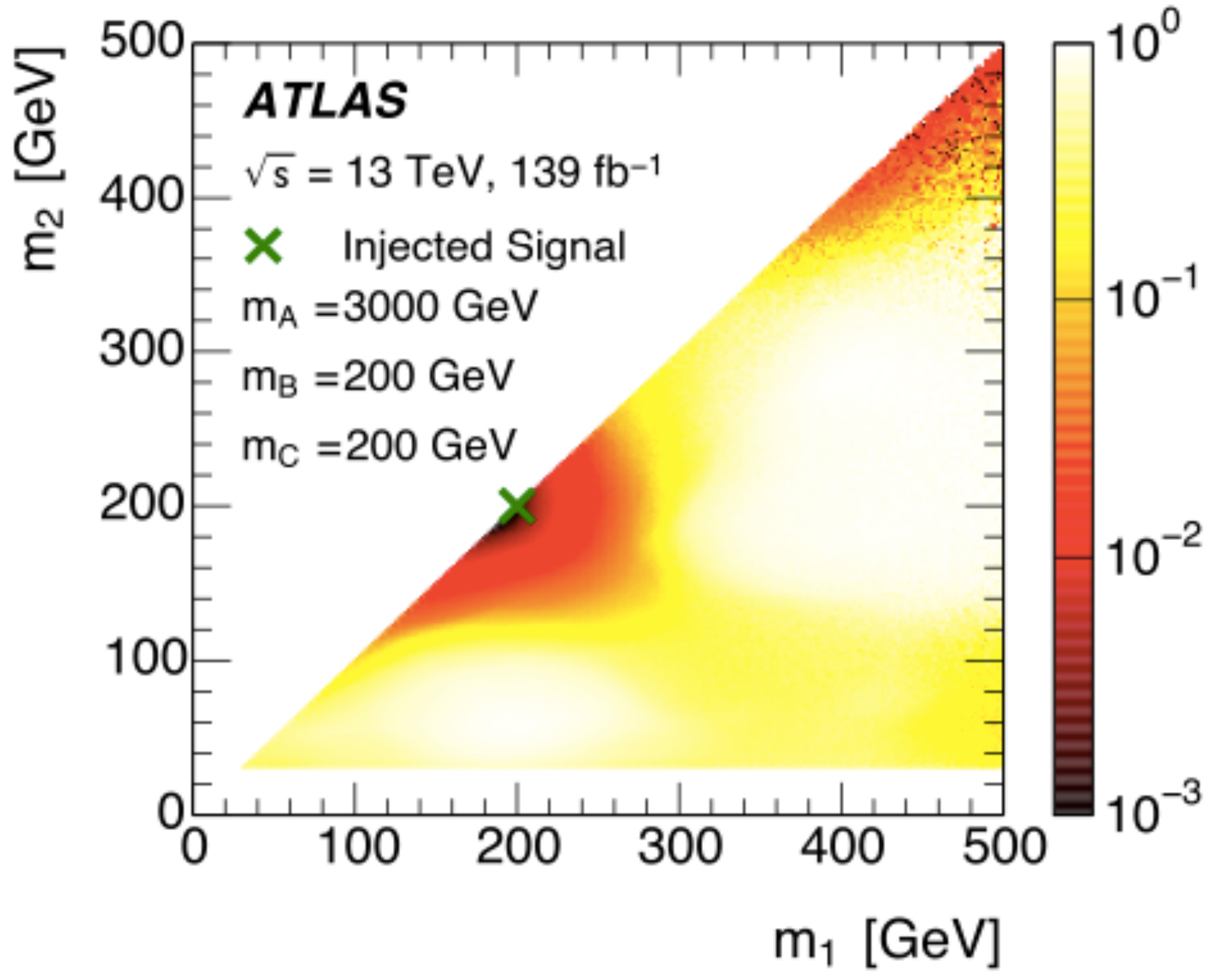
- robust anomaly detection tool
- looking for group anomalies
- level of agnosticism
- perform analysis (bump hunt, ABCD, …)

Model-agnostic searches & ML

- Are we leaving stones unturned? Can we answer this question only via direct searches?
- **Anomaly searches**: define background from the data and find "anomalous" events

a known problem in Machine Learning (**or not?**)
what we are looking for:

- robust anomaly detection tool
- looking for group anomalies
- level of agnosticism
- perform analysis (bump hunt, ABCD, …)

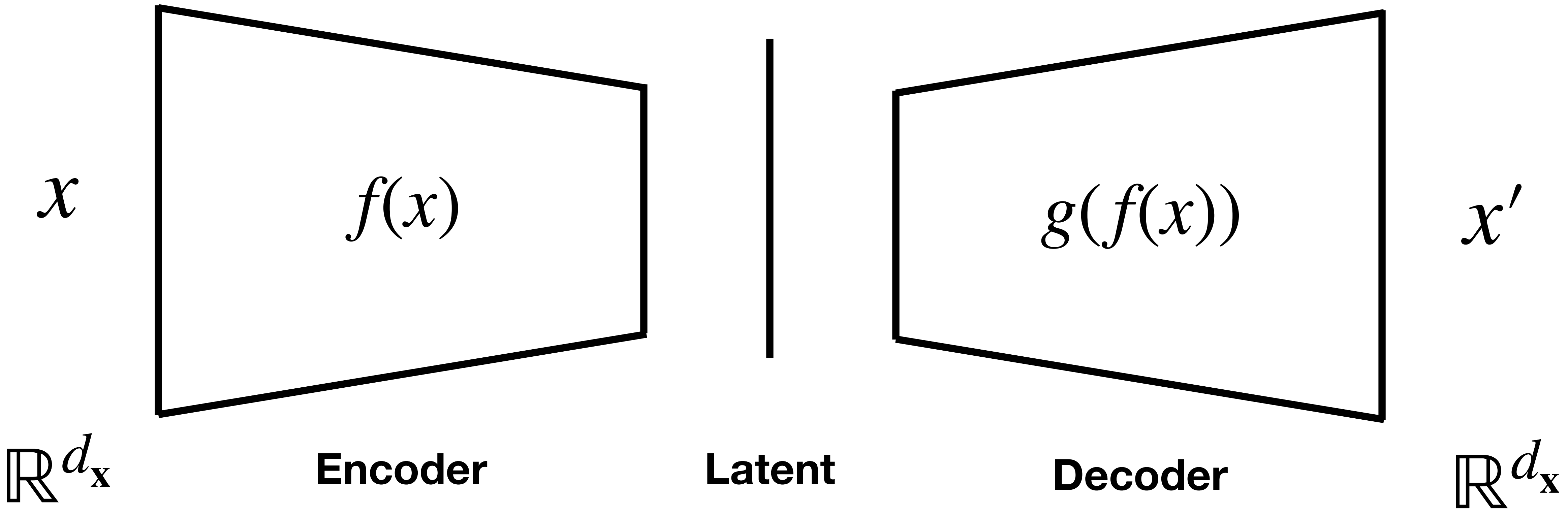


Already many interesting challenges/applications of ML techniques

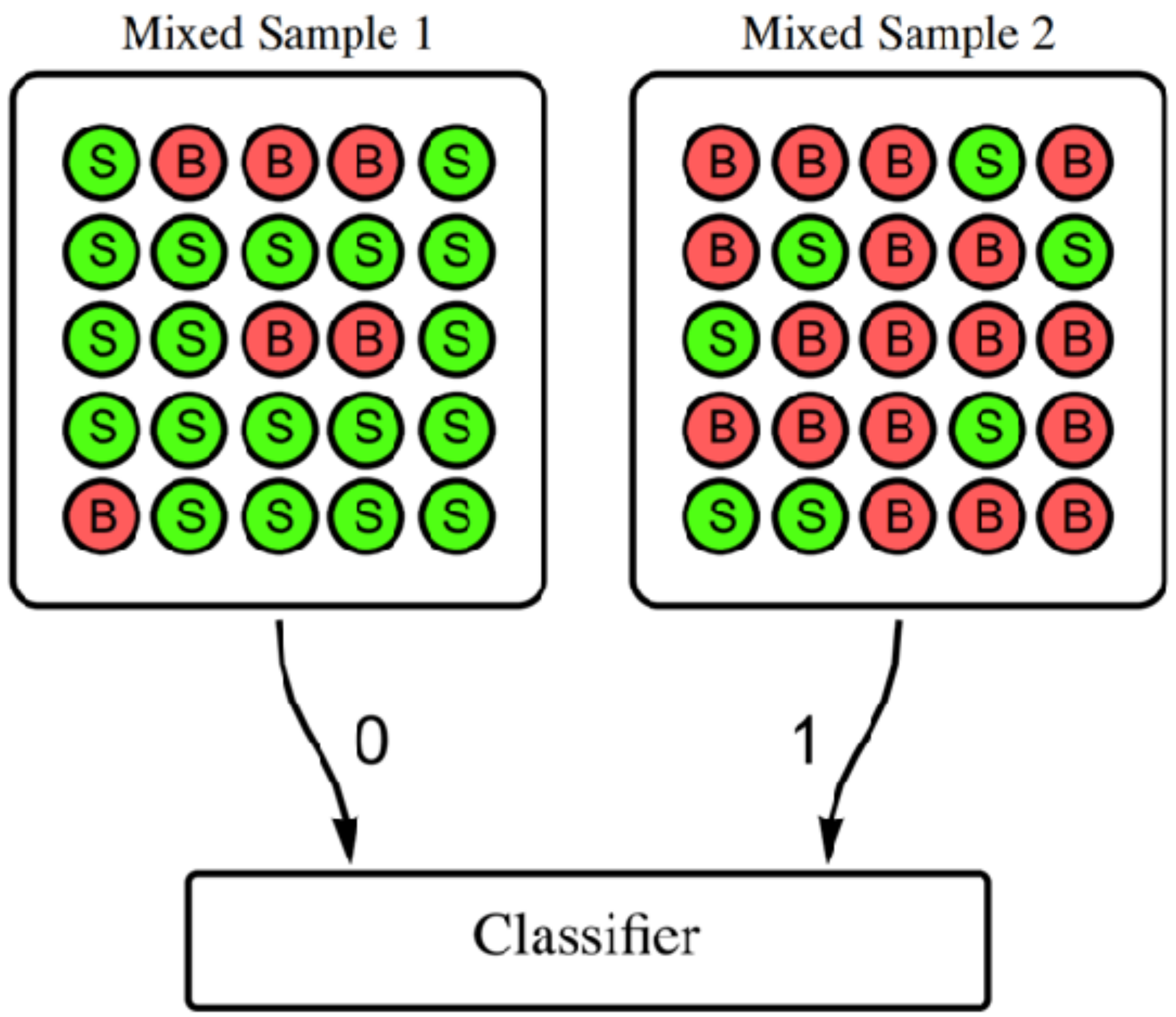
Model-agnostic searches & ML

Two big families:

Autoencoders (AE)



Classification without labels (CWOLA)



Autoencoders for HEP: questions

How do we define an anomaly?

Autoencoders for HEP: questions

How do we define an anomaly?

We can define an anomaly as an out of distribution (OOD) object

Anomaly score: $S = \{x \,|\, l(x) < \tau\}$

Auto-Encoders: use MSE as estimated density

Autoencoders for HEP: questions

How do we define an anomaly?

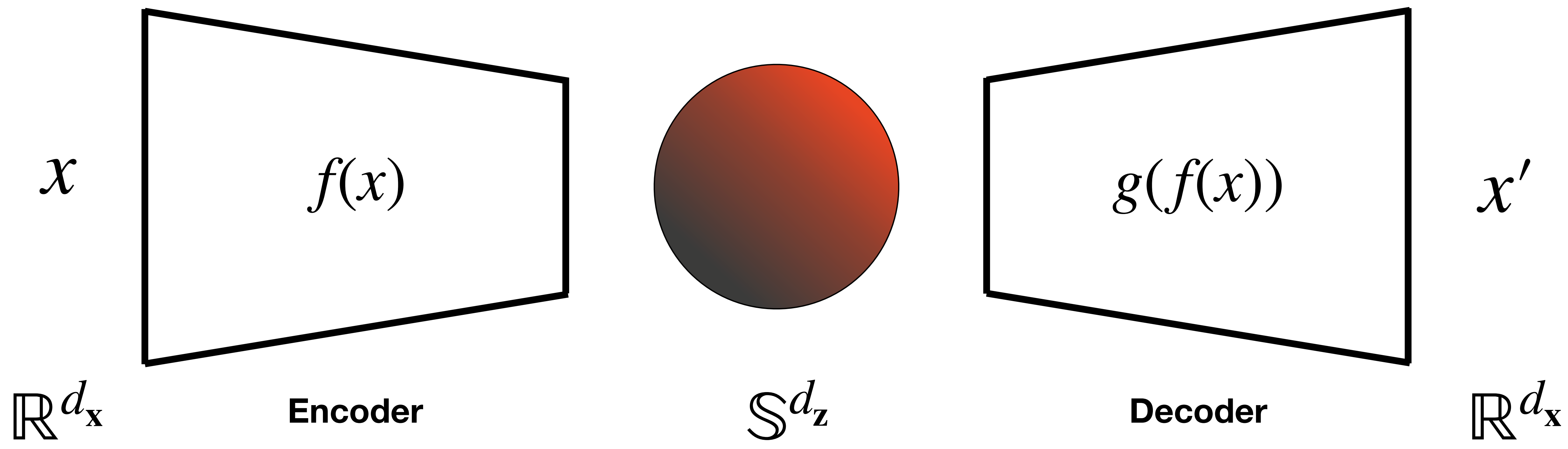
We can define an anomaly as an out of distribution (OOD) object

$$MSE(x, x') = ||x - x'||_2^2$$

Anomaly score: $S = \{x \,|\, l(x) < \tau\}$

not robust OOD estimator
see arXiv:2206:14225

Auto-Encoders: use MSE as estimated density



Autoencoders for HEP: questions

How do we define an anomaly?

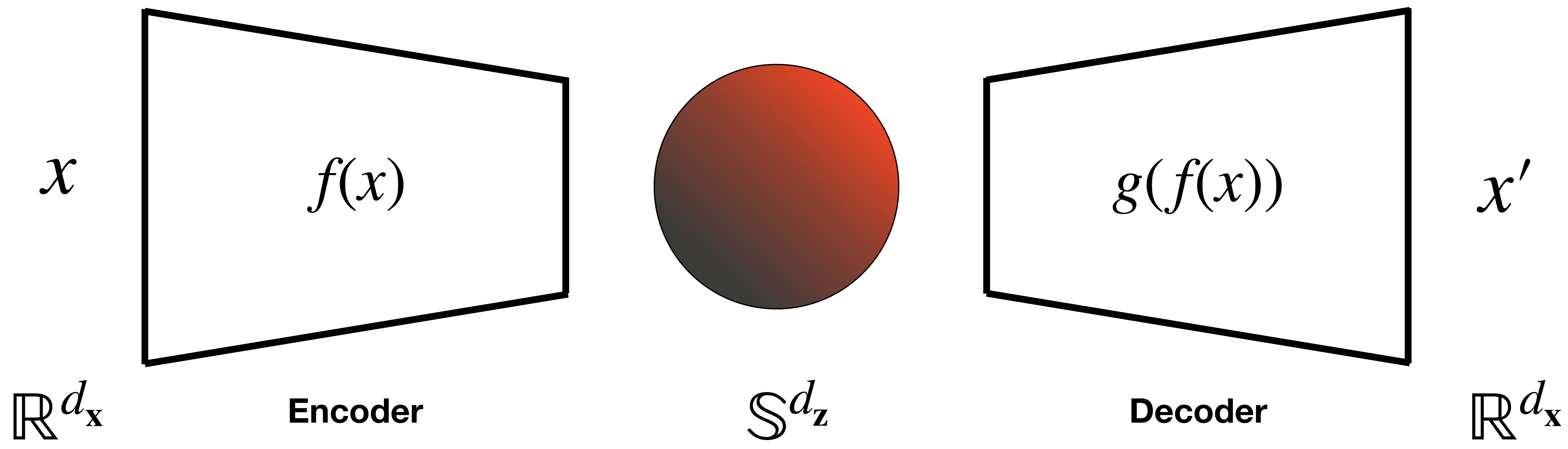
We can define an anomaly as an out of distribution (OOD) object

$$MSE(x, x') = ||x - x'||_2^2$$

Anomaly score: $S = \{x \,|\, l(x) < \tau\}$

not robust OOD estimator
see arXiv:2206:14225

Auto-Encoders: use MSE as estimated density

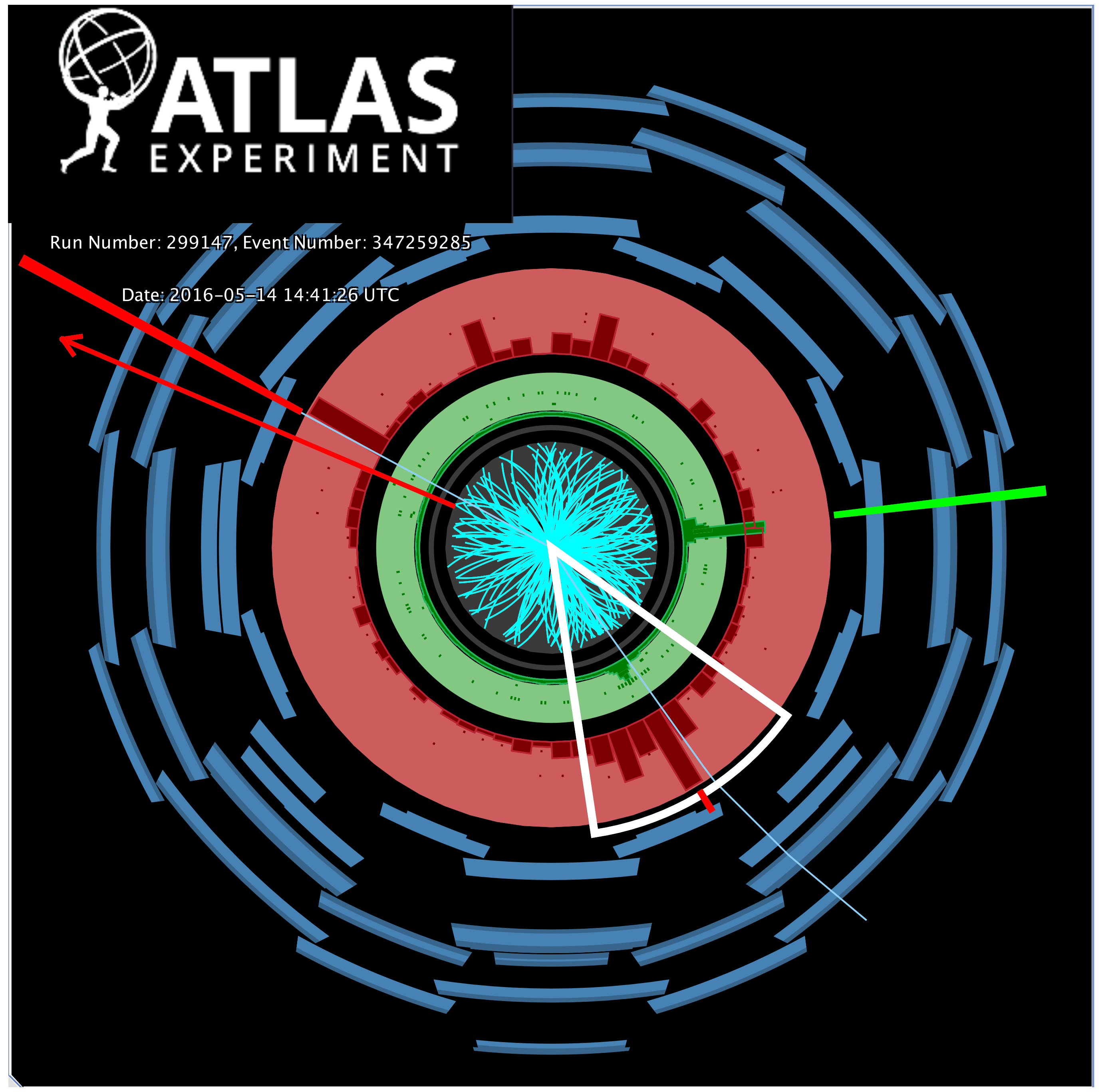


score is not invariant to data preprocessing

Autoencoders for HEP: questions

How to choose the best representation?

Example: LHC data has known symmetries ⟶ exploit them for better representations

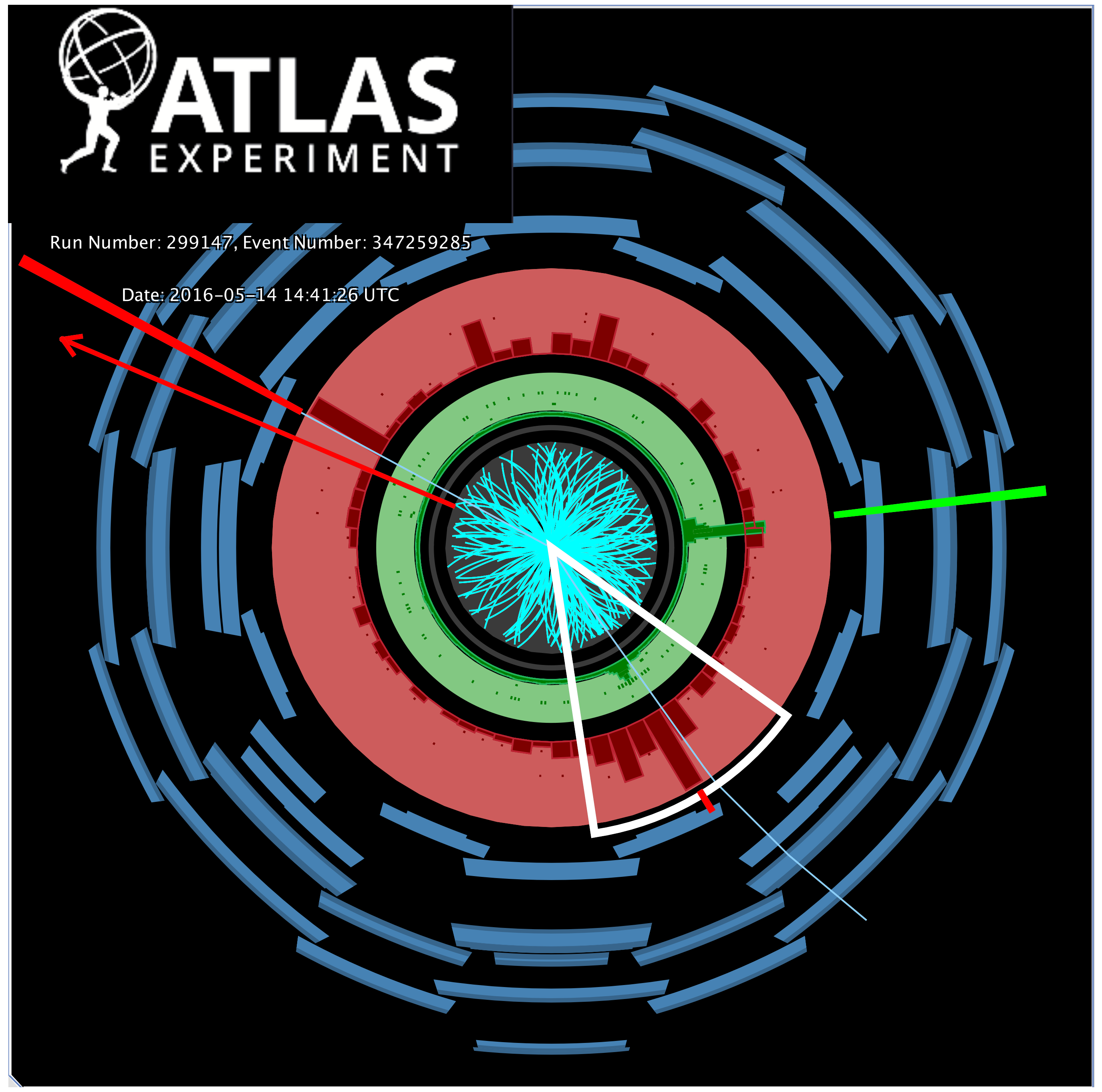


Reconstructed objects

Autoencoders for HEP: questions

How to choose the best representation?

Example: LHC data has known symmetries ⟶ exploit them for better representations



Reconstructed objects

Issue of auto encoding:

Latent space cannot be invariant to symmetries:

reconstruction of different events would not be possible

Autoencoders for HEP: questions

How to choose the best representation?

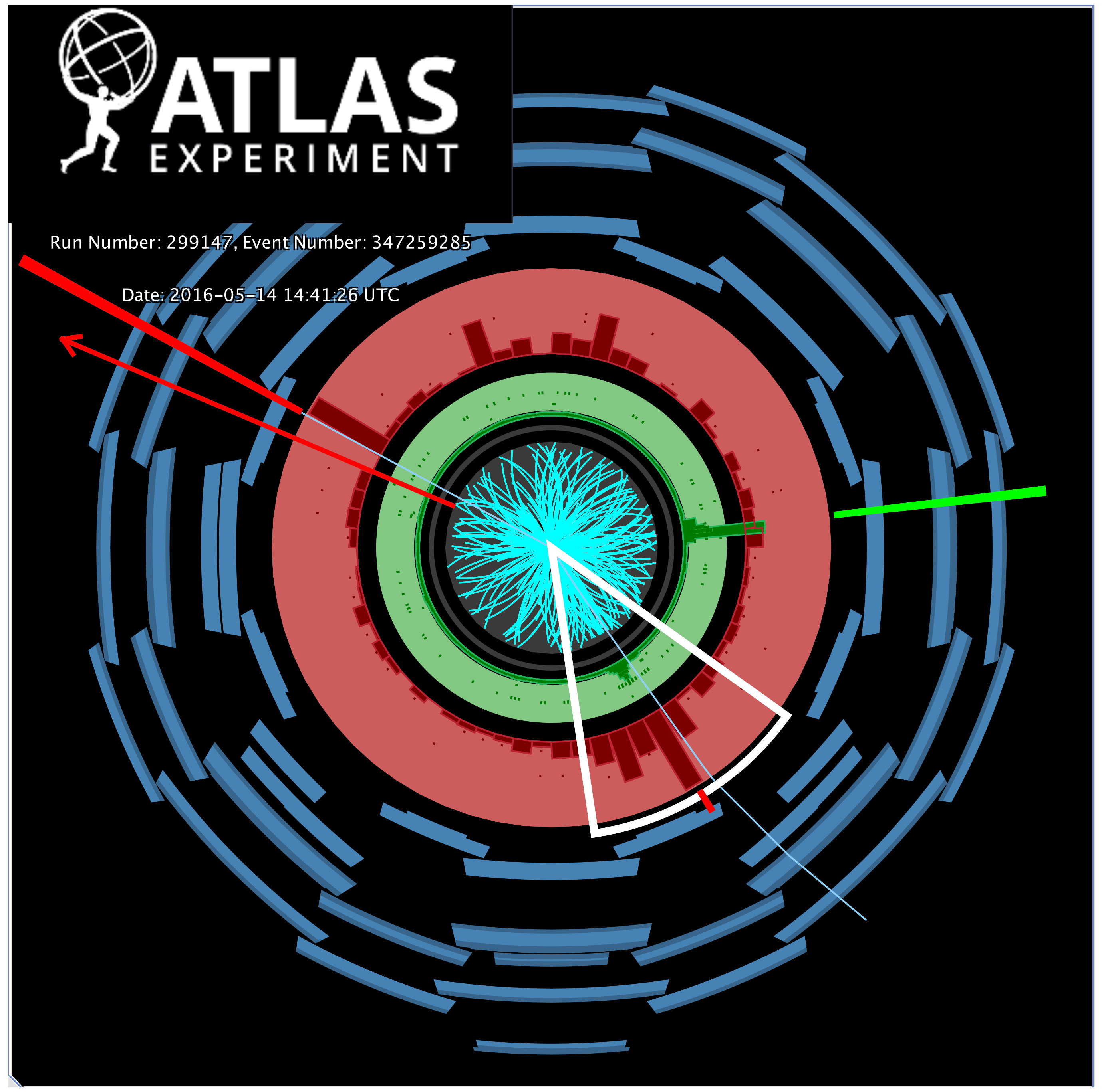
Example: LHC data has known symmetries ⟶ exploit them for better representations

Issue of auto encoding:

Latent space cannot be invariant to symmetries:

reconstruction of different events would not be possible

⟶ **preprocessing is necessary**



Reconstructed objects

Application at event-level

[Anomalies, representations, and self-supervision, Dillon B. et al. arXiv:2301.04660]

Application at event-level

Dataset: mixture of SM events

$W \to l\nu$ (59.2%)
$Z \to ll$ (6.7%)
$t\bar{t}$ production (0.3%)
QCD multijet (33.8 %)

BSM benchmarks

$A \to 4l$
$LQ \to b\nu$
$h_0 \to \tau\tau$
$h_+ \to \tau\nu$

[Anomalies, representations, and self-supervision, Dillon B. et al. arXiv:2301.04660]

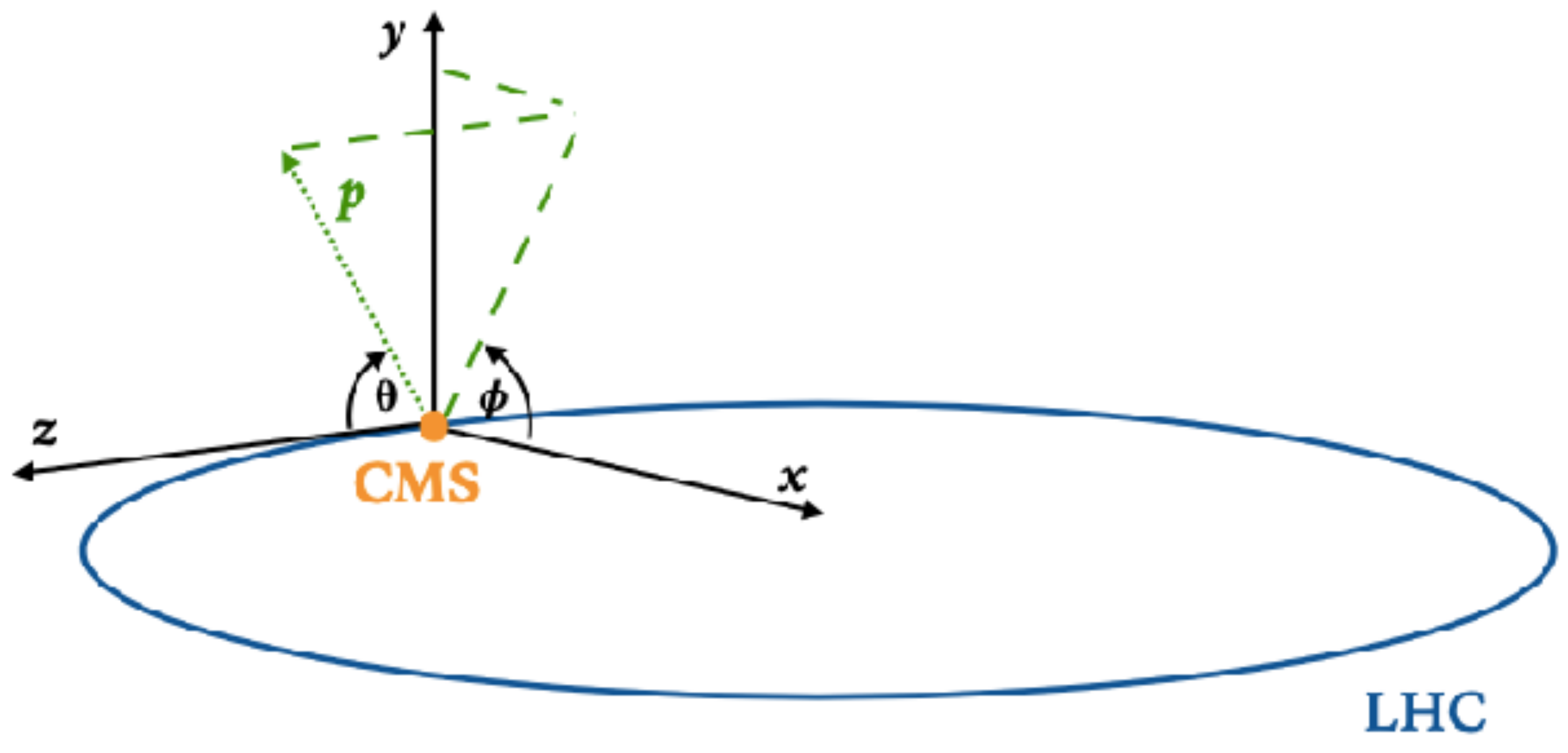
Application at event-level

Dataset: mixture of SM events

$W \to l\nu$ (59.2%)
$Z \to ll$ (6.7%)
$t\bar{t}$ production (0.3%)
QCD multijet (33.8 %)

BSM benchmarks

$A \to 4l$
$LQ \to b\nu$
$h_0 \to \tau\tau$
$h_+ \to \tau\nu$



The events are represented in format: (19, 3) entries

- 19 particles: MET, 4 electrons, 4 muons, and 10 jets
- 3 observables: p_T, η, ϕ
- $|\eta| < [3,\ 2.1,\ 4]$ for $e,\ \mu,\ j$ respectively

[Anomalies, representations, and self-supervision, Dillon B. et al. arXiv:2301.04660]

Self-supervision

Self-supervision

- Neural Networks are not invariant to physical symmetries in data
- Typically solved through “pre-processing”

Self-supervision

- Neural Networks are not invariant to physical symmetries in data
- Typically solved through “pre-processing”

Our goal: control the training to ensure we learn physical quantities

What the **representations** should have:

- invariance to certain transformations of the jet/event
- discriminative power

Self-supervision

- Neural Networks are not invariant to physical symmetries in data
- Typically solved through "pre-processing"

Our goal: control the training to ensure we learn physical quantities

What the **representations** should have:

- invariance to certain transformations of the jet/event
- discriminative power

- CLR: map raw data to a new representation/observables
- Self-supervision: during training we use pseudo-labels, not truth labels

Contrastive Learning framework

Contrastive Learning framework

Contrastive Learning paradigm:

- positive pairs: $\{(x_i, x_i')\}$ where x_i' is an augmented version of x_i
- negative pairs: $\{(x_i, x_j) \cup (x_i, x_j')\}$ for $i \neq j$

Contrastive Learning framework

Contrastive Learning paradigm:

- positive pairs: $\{(x_i, x_i')\}$ where x_i' is an augmented version of x_i
- negative pairs: $\{(x_i, x_j) \cup (x_i, x_j')\}$ for $i \neq j$

Augmentation: any transformation (e.g. rotation) of the original jet

Contrastive Learning framework

Contrastive Learning paradigm:

- positive pairs: $\{(x_i, x_i')\}$ where x_i' is an augmented version of x_i
- negative pairs: $\{(x_i, x_j) \cup (x_i, x_j')\}$ for $i \neq j$

Augmentation: any transformation (e.g. rotation) of the original jet

Train a Transformer-encoder network to map the data to a compact latent space, $f : \mathscr{J} \rightarrow \mathscr{R}$

Contrastive Learning framework

Contrastive Learning paradigm:

- positive pairs: $\{(x_i, x_i')\}$ where x_i' is an augmented version of x_i
- negative pairs: $\{(x_i, x_j) \cup (x_i, x_j')\}$ for $i \neq j$

Augmentation: any transformation (e.g. rotation) of the original jet

Train a Transformer-encoder network to map the data to a compact latent space, $f : \mathscr{I} \to \mathscr{R}$

Loss function:

$$\mathscr{L} = -\log \frac{exp(s(z_i, z_i')/\tau)}{\sum_{x \in batch} \mathbb{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

Defining augmentations

Defining augmentations

$$\mathscr{L} = -\log \frac{exp(s(z_i, z_i')/\tau)}{\sum_{x \in batch} \mathbb{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

Similarity measure:

$$s(z_i, z_j) = \frac{z_i \cdot z_j}{|z_i||z_j|}, \qquad z_i = f(x_i)$$

Defining augmentations

$$\mathscr{L} = -\log \frac{exp(s(z_i, z_i')/\tau)}{\sum_{x \in batch} \mathbb{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

alignment

uniformity

Similarity measure:

$$s(z_i, z_j) = \frac{z_i \cdot z_j}{|z_i||z_j|}, \qquad z_i = f(x_i)$$

Defining augmentations

$$\mathscr{L} = -\log \frac{exp(s(z_i, z_i')/\tau)}{\sum_{x \in batch} \mathbb{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

alignment

uniformity

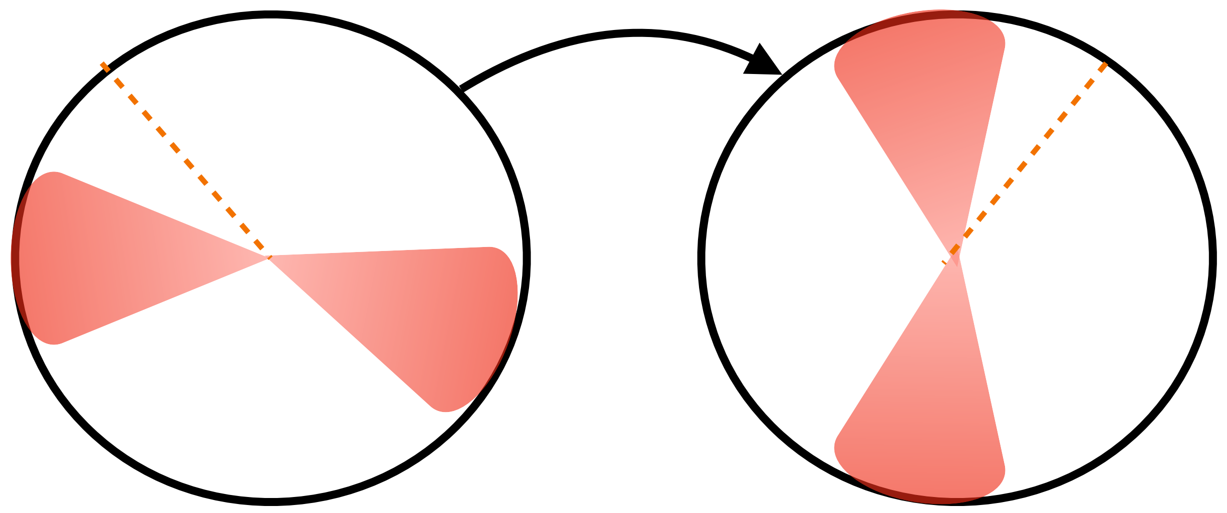
Similarity measure:

$$s(z_i, z_j) = \frac{z_i \cdot z_j}{|z_i||z_j|}, \qquad z_i = f(x_i)$$

Physical augmentations:

- azimuthal rotations
- η, ϕ smearing
- energy smearing

$$p_T \sim \mathcal{N}(p_T, f(p_T)), \qquad f(p_T) = \sqrt{0.052 p_T^2 + 1.502 p_T^2}$$



$$\eta' \sim \mathcal{N}\left(\eta, \sigma(p_T)\right)$$

$$\phi' \sim \mathcal{N}\left(\phi, \sigma(p_T)\right)$$

Self-supervision for anomaly detection

Can we train a transformer-encoder only on background data?

Possible, with no guarantee to learn representations sensitive to new physics

Introduce z^*, anomaly-augmented point

Self-supervision for anomaly detection

Can we train a transformer-encoder only on background data?

Possible, with no guarantee to learn representations sensitive to new physics

Introduce z^*, anomaly-augmented point

Loss function:

$$\mathscr{L}_{AnomCLR} = -\log \frac{exp(s(z_i, z_i') - s(z_i, z_i^*)/\tau)}{\sum_{x \in batch} \mathbb{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

$$\mathscr{L}_{AnomCLR+} = -\log e^{(s(z_i, z_i') - s(z_i, z_i^*))/\tau} = \frac{s(z_i, z_i^*) - s(z_i, z_i)}{\tau}$$

Enhancing discriminative features

Enhancing discriminative features

Representations may not be sensitive to BSM features:

- physical augmentations: alignment between positive pairs
- anomalous augmentations: discriminative power of possible BSM features

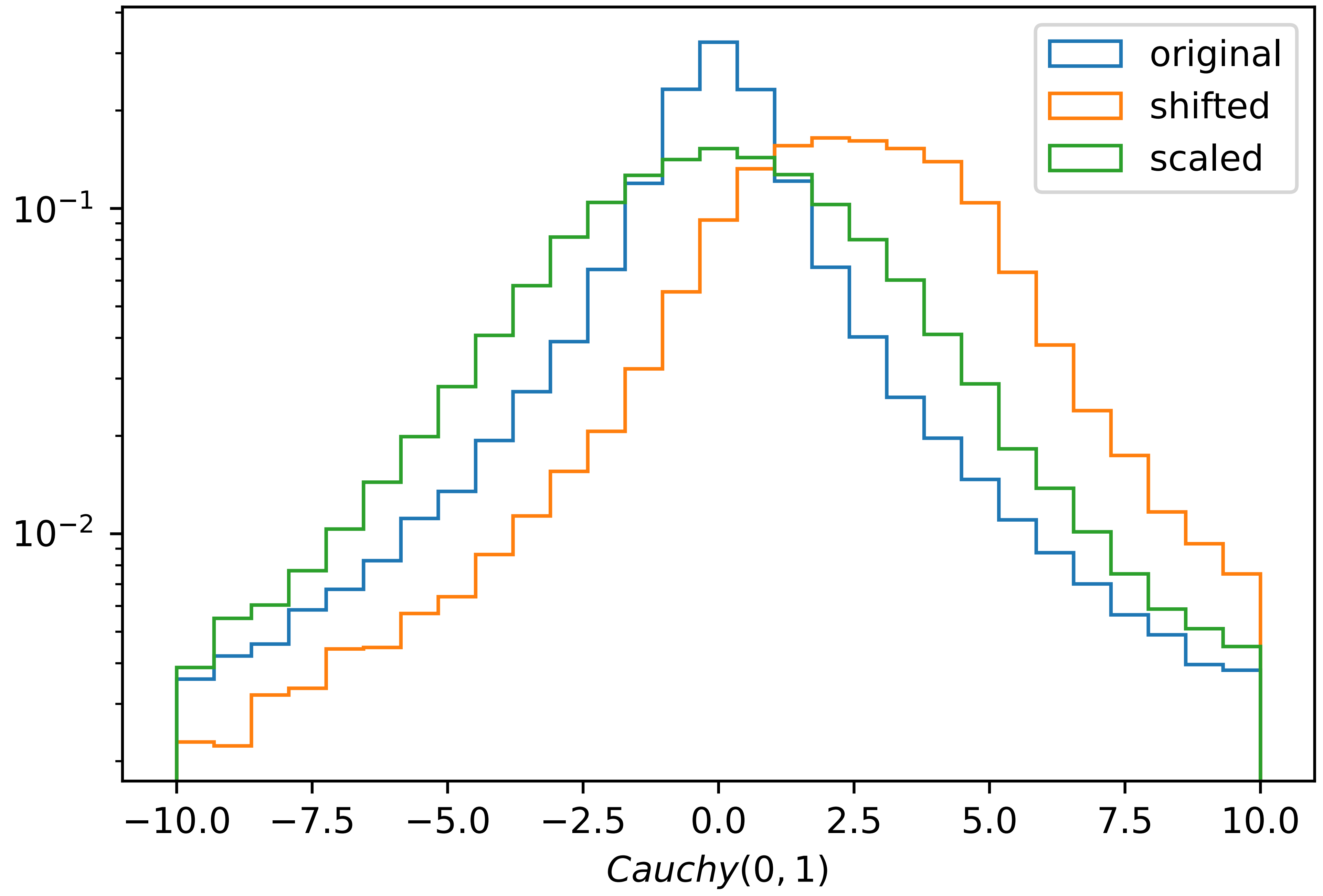
Enhancing discriminative features

Representations may not be sensitive to BSM features:

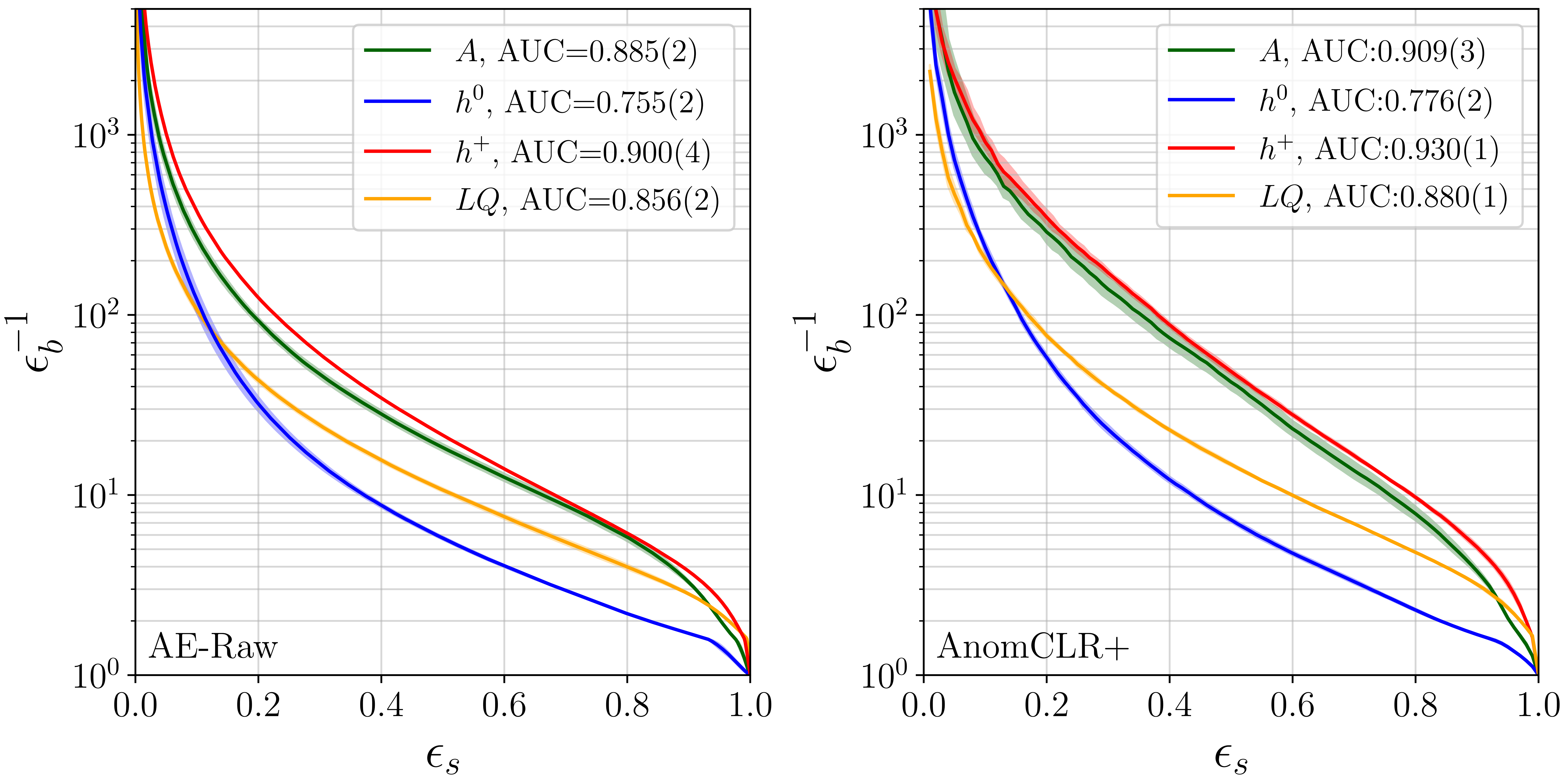
- physical augmentations: alignment between positive pairs
- anomalous augmentations: discriminative power of possible BSM features

Anomalous augmentations:

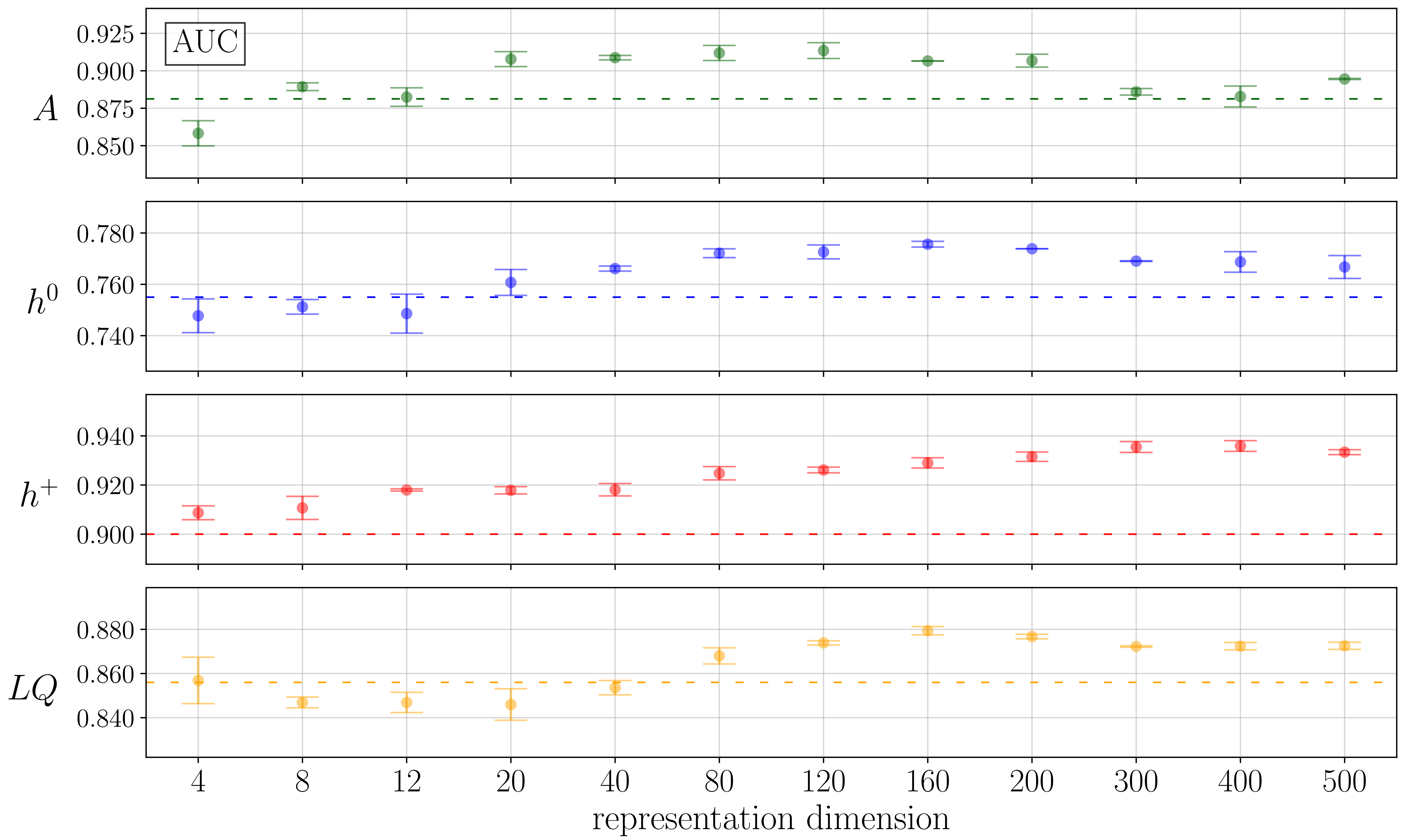
- multiplicity shifts:
  - add a random number of particles, update MET
  - split existing particles, keeping total p_T and MET fixed
- p_T and MET shifts

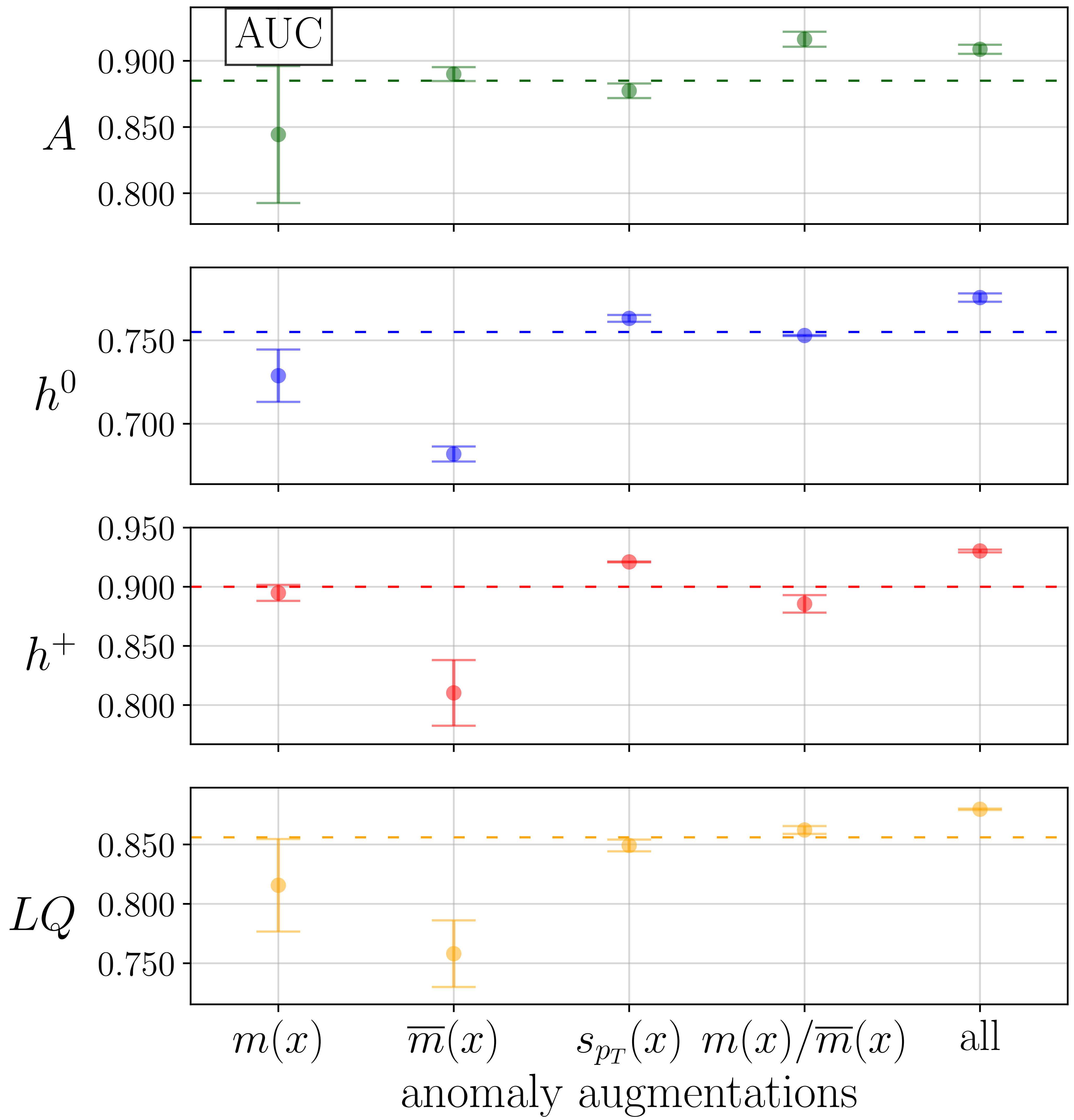


Results: improved sensitivity



Effect of anomalous augmentations



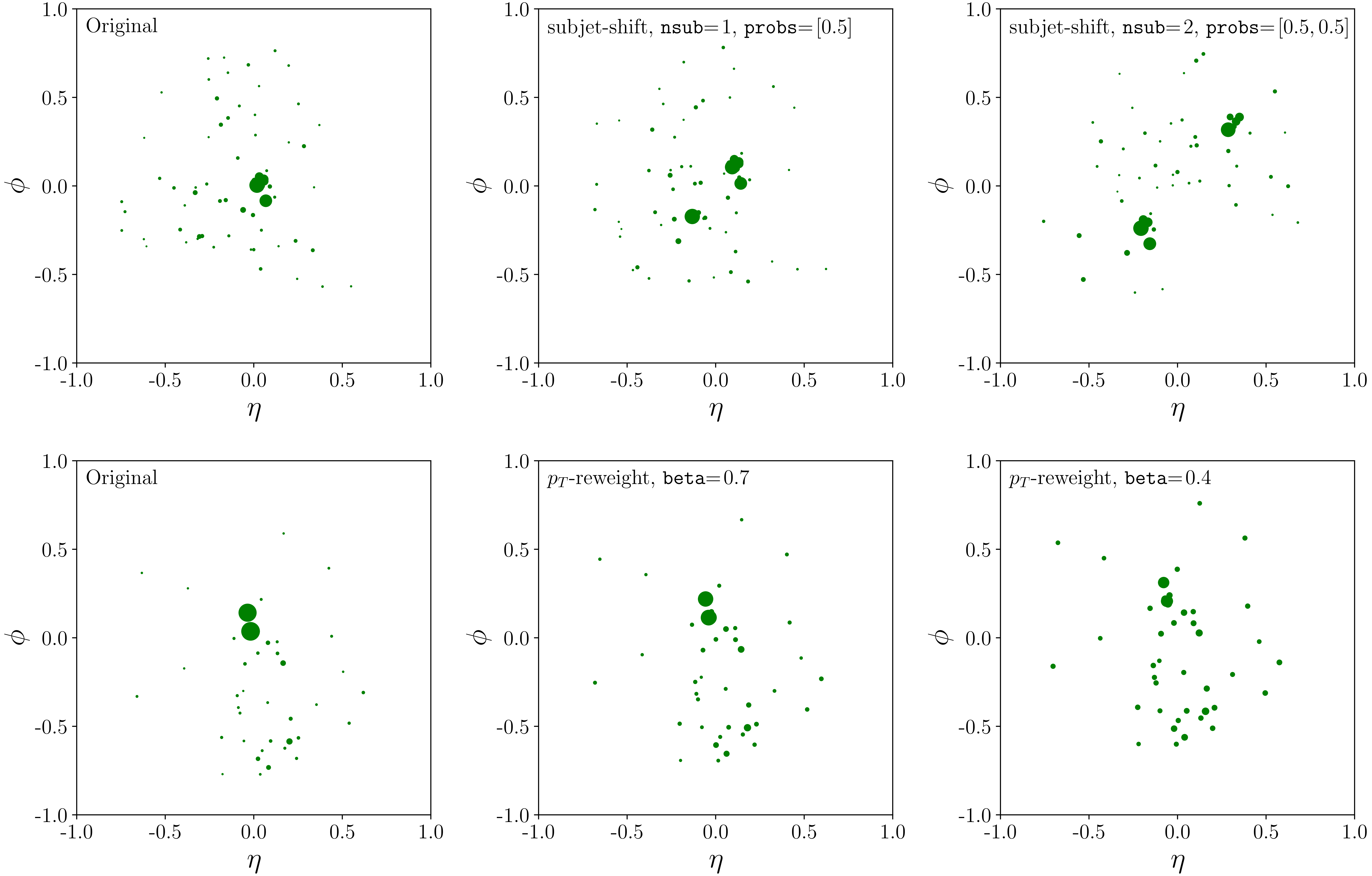


AnomalyCLR on Jets

preliminary

shift constituents
⟶ heavy decay

reweight constituents p_T
⟶ semivisible jets



Conclusions/Outlook

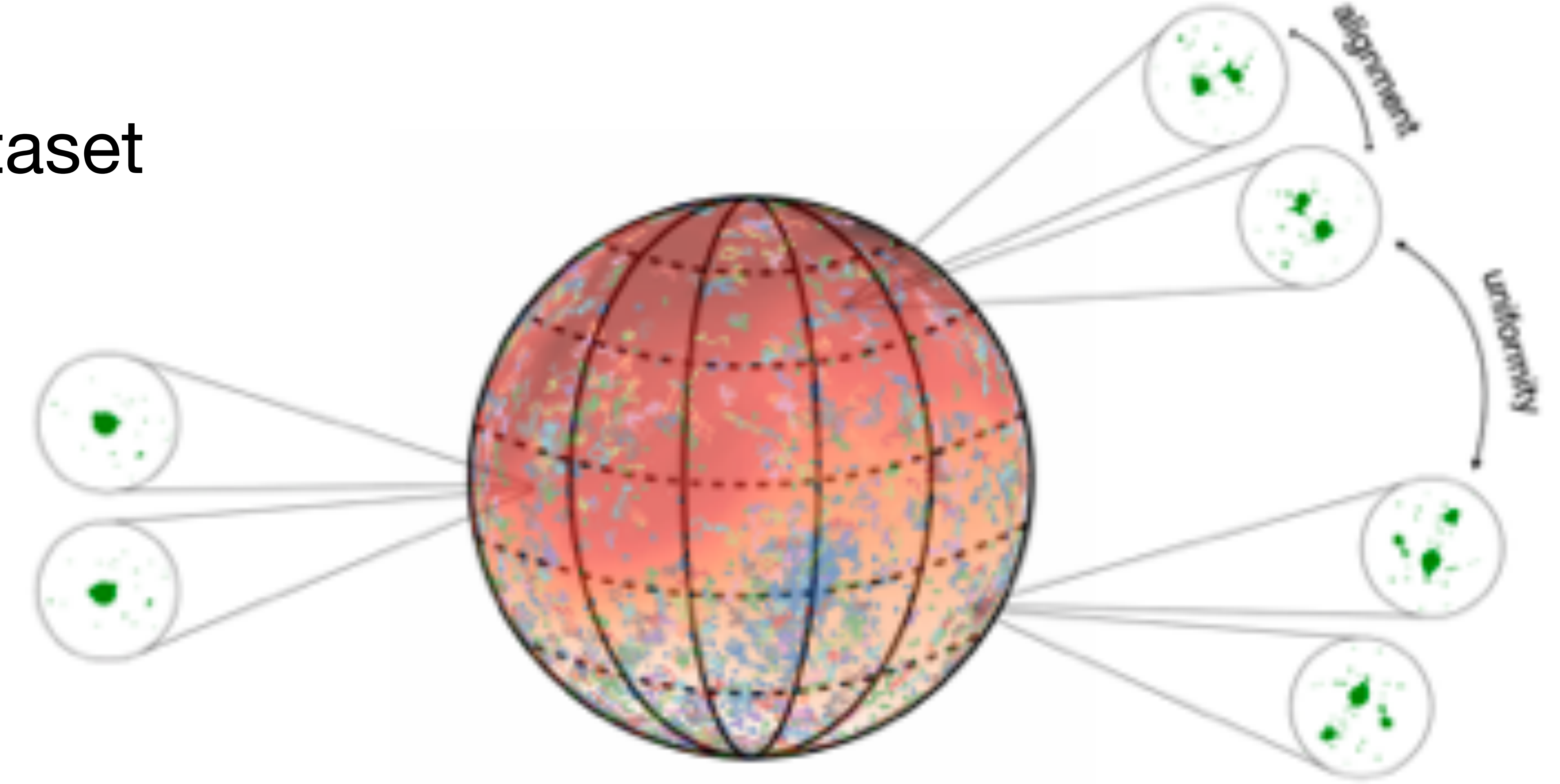
Unsupervised Machine Learning for NP searches can be a powerful tool for LHC physics

⟶ **Auto-Encoders (AE)** are simple and effective neural networks for AD

Self-supervision and CLR are a powerful tools to build representations for downstream tasks

AnomalyCLR ⟶ learn invariances, and representations with high discriminative power

Enhanced tagging performance tested on the ADC2021 dataset



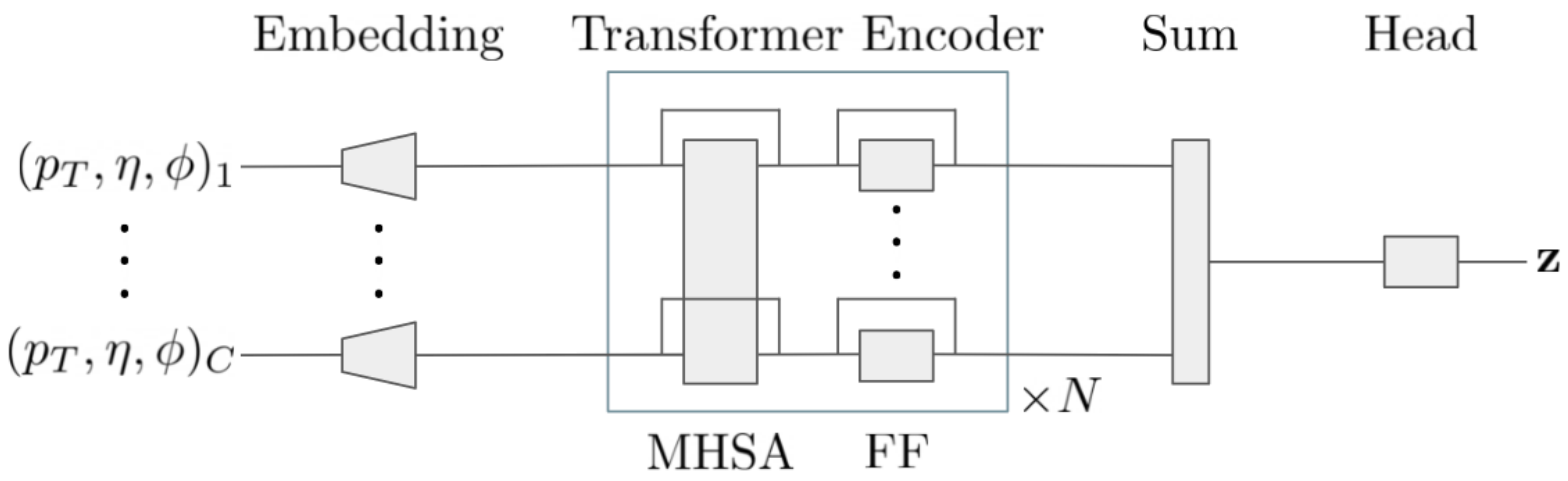
Future work:

Self-supervision for anomalous jet-tagging

Thanks for your attention!

Backup

Transformer Network



$$A(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

$$Multihead = Concat(head_{1\ldots N})W^O$$

Results: SIC CURVES