Caltech

UC San Diego

CMS

Lorentz Group Equivariant Autoencoders

ML4Jets2024

November 7, LPNHE, Paris, France

Hao, Z., Kansal, R., Duarte, J. *et al.* Lorentz group equivariant autoencoders. *Eur. Phys. J. C* **83**, 485 (2023).
https://doi.org/10.1140/epjc/s10052-023-11633-5, arXiv:2212.07347

Zichun Hao, Raghav Kansal, Javier Duarte, Nadezda Chernyavskaya

Outline

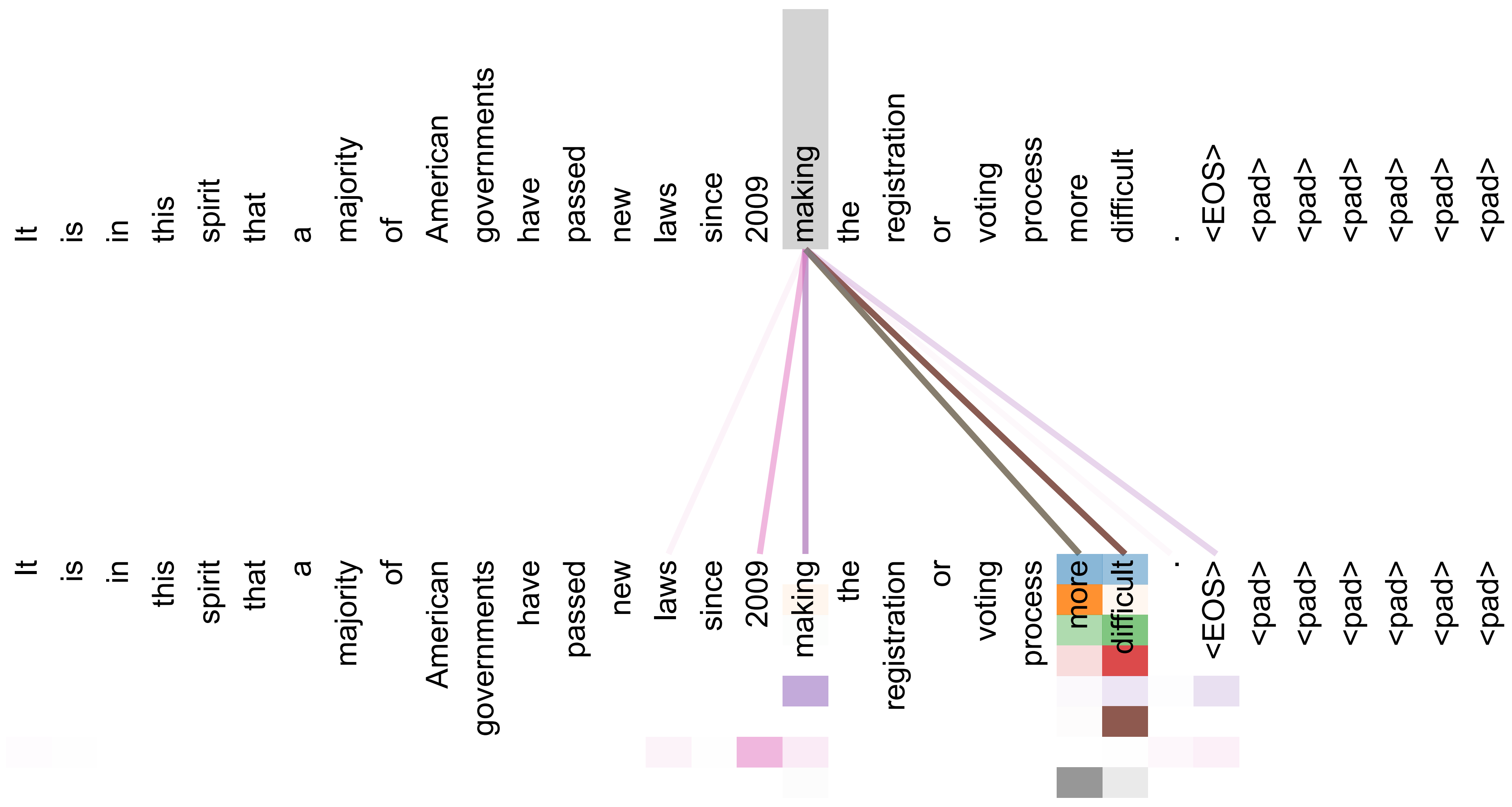
- Overview
- Experimental Results
- Conclusion

Overview

Embedding Inductive Biases

For Natural Languages

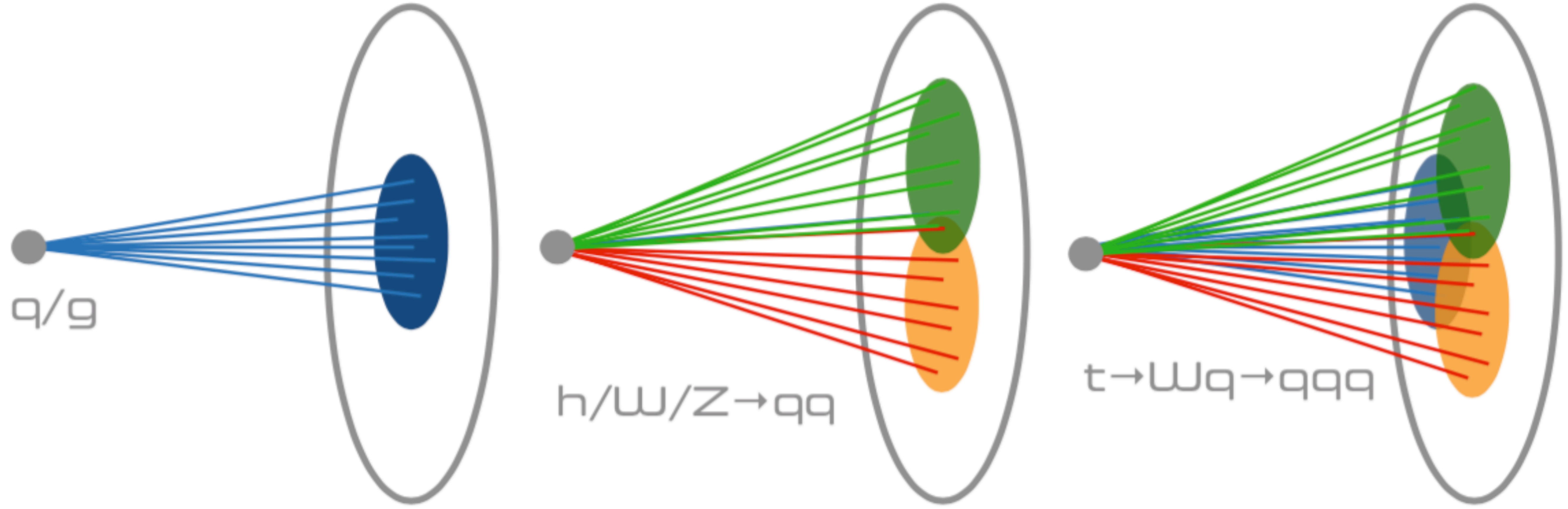
- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance in terms of **performance**, **interpretability**, and **data efficiency**.
  - The self-attention mechanism gives rise to transformers.



Embedding Inductive Biases

For HEP

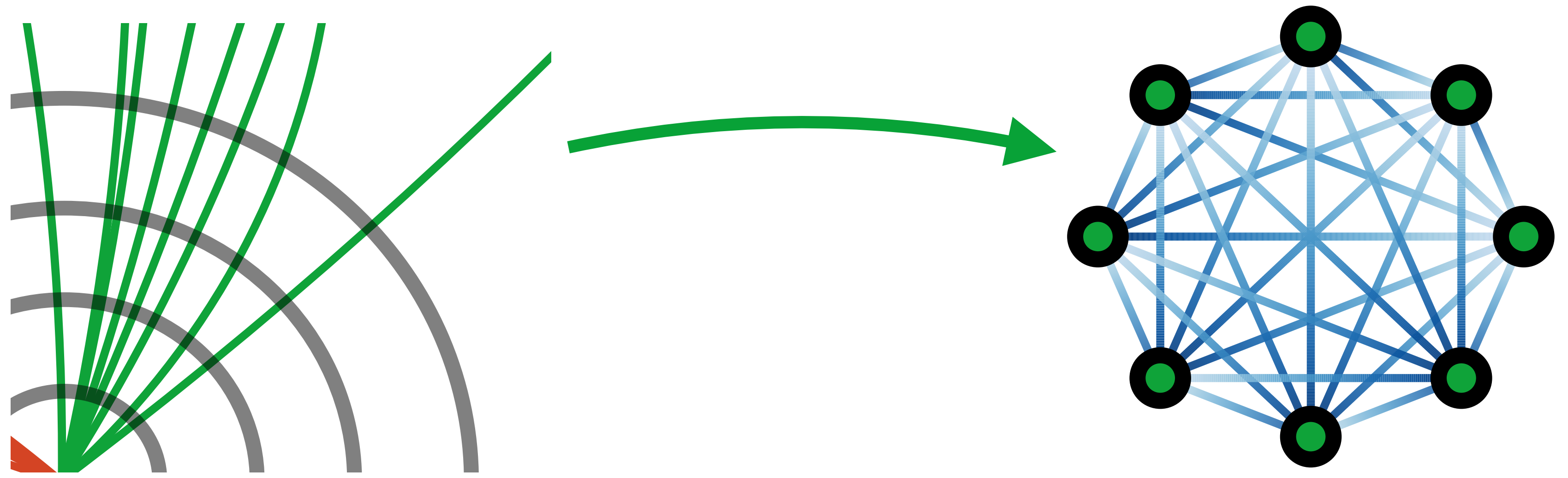
- What about HEP data like jets?



- One possible answer: graph neural networks (GNNs)

Graph Neural Networks

In HEP



- GNNs add **inductive biases and symmetries** into the neural network.
  - Mimics the structure of data in HEP: nodes as particles and edges as interactions.
  - Permutational symmetry: graphs have no sense of ordering.
  - Example: `ParticleNet` [arXiv:1902.08570] achieved by-then SOTA performance on jet tagging benchmarks.
- Another fundamental symmetry in HEP: (*approximate*) Lorentz group symmetry.
  - Example: `LorentzNet` [arXiv:2201.08187] and `PELICAN` [arXiv:2307.16506] show the advantages by achieving SOTA performance on jet tagging benchmarks.
  - An ablation study [arXiv:2208.07814] done to demonstrate the benefits of Lorentz-symmetry preservation even with detector effects

Lorentz Group Equivariant Autoencoder (LGAE)

Lorentz Group Network [arXiv: 2006.04780]

- Work on the irreducible representations (**irreps**) of the Lorentz group.
  - Examples: Lorentz scalars (e.g. mass) and 4-vectors (e.g. 4-momentum)
  - Input physical quantities and all intermediate features transform properly under the corresponding Lorentz transformation.
- **Graph** structure
  - Nodes as particles.
  - Edges as mutual and self interactions.

Lorentz Group Network

Lorentz Group Equivariant Message Passing (LMP) Layers

New node feature

Old node features*

Self interaction

Mutual interaction

$$\mathscr{F}_i^{(t+1)} = \mathrm{MixReps}\left(\mathscr{F}_i^{(t)} \oplus \mathrm{CG}\left[\left(\mathscr{F}_i^{(t)}\right)^{\otimes 2}\right] \oplus \mathrm{CG}\left[\sum_{j\neq i} f(p_{ij}^2) p_{ij} \otimes \mathscr{F}_j^{(t)}\right]\right).$$

Linear mixing
(Only mixing features from the same representation space)

Clebsch-Gordan decomposition

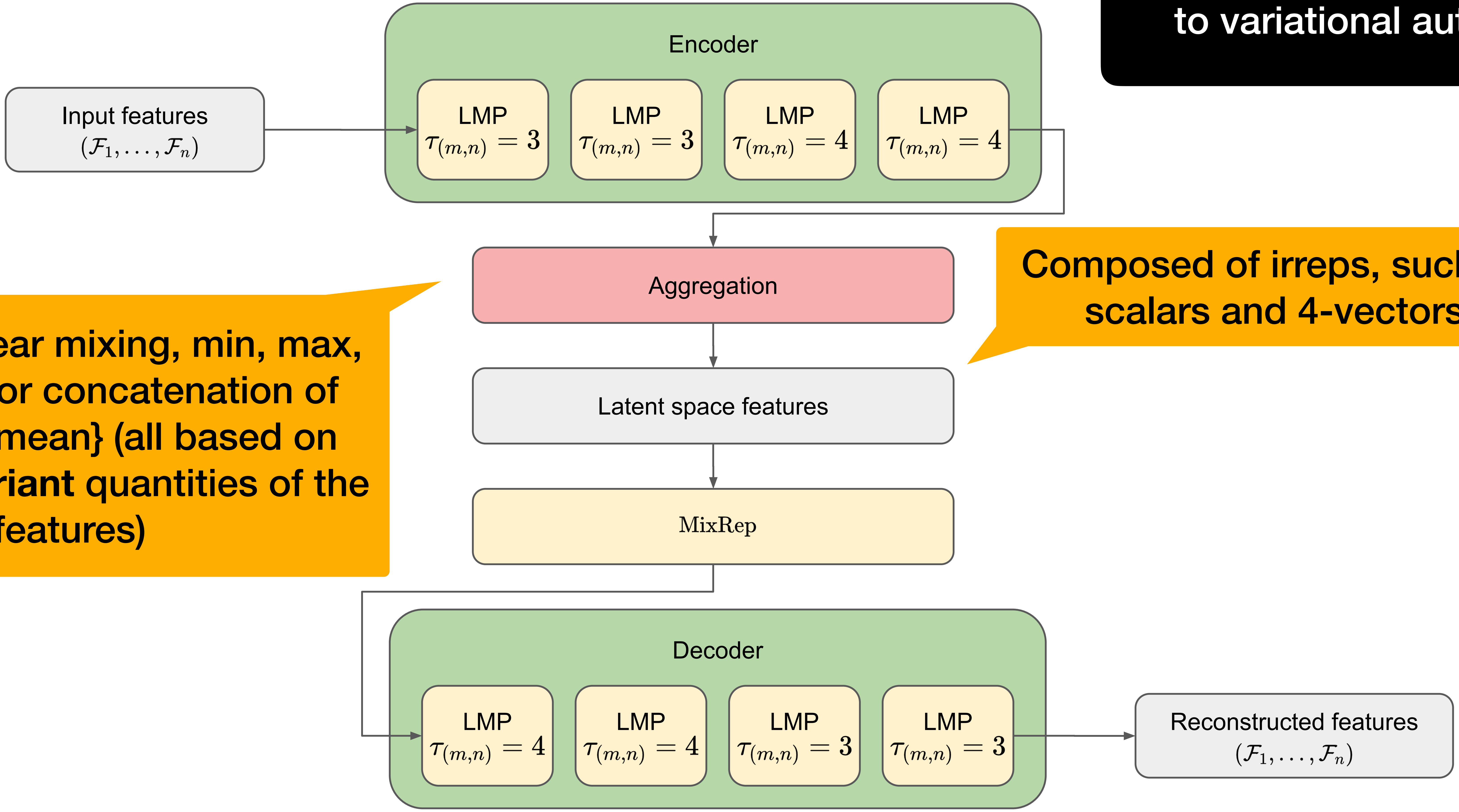
$p_{ij} \equiv p_i - p_j$

* Each (m,n) irrep space in $\mathscr{F}_i^{(t)}$ contains $\tau_{(m,n)}^{(t)}$ channels (similar idea with CNNs)

Lorentz Group Equivariant Autoencoder (LGAE)

Architecture

Applications: compressions, anomaly detections, data generations (if adapted to variational autoencoders), etc.



Jet-level: linear mixing, min, max, and mean, or concatenation of {min, max, mean} (all based on **Lorentz invariant** quantities of the features)

Composed of irreps, such as scalars and 4-vectors

Lorentz Group Equivariant Autoencoder (LGAE)

Autoencoders as Anomaly Detectors

- Trained to reconstruct background data.
- The autoencoder has **never** seen signal data.
  - Expect a **worse** reconstruction performance.
  - Use the reconstruction score (e.g. MSE) as an anomaly metric.
- Example: AXOL1TL (Level-1 Trigger at the CMS Experiment)

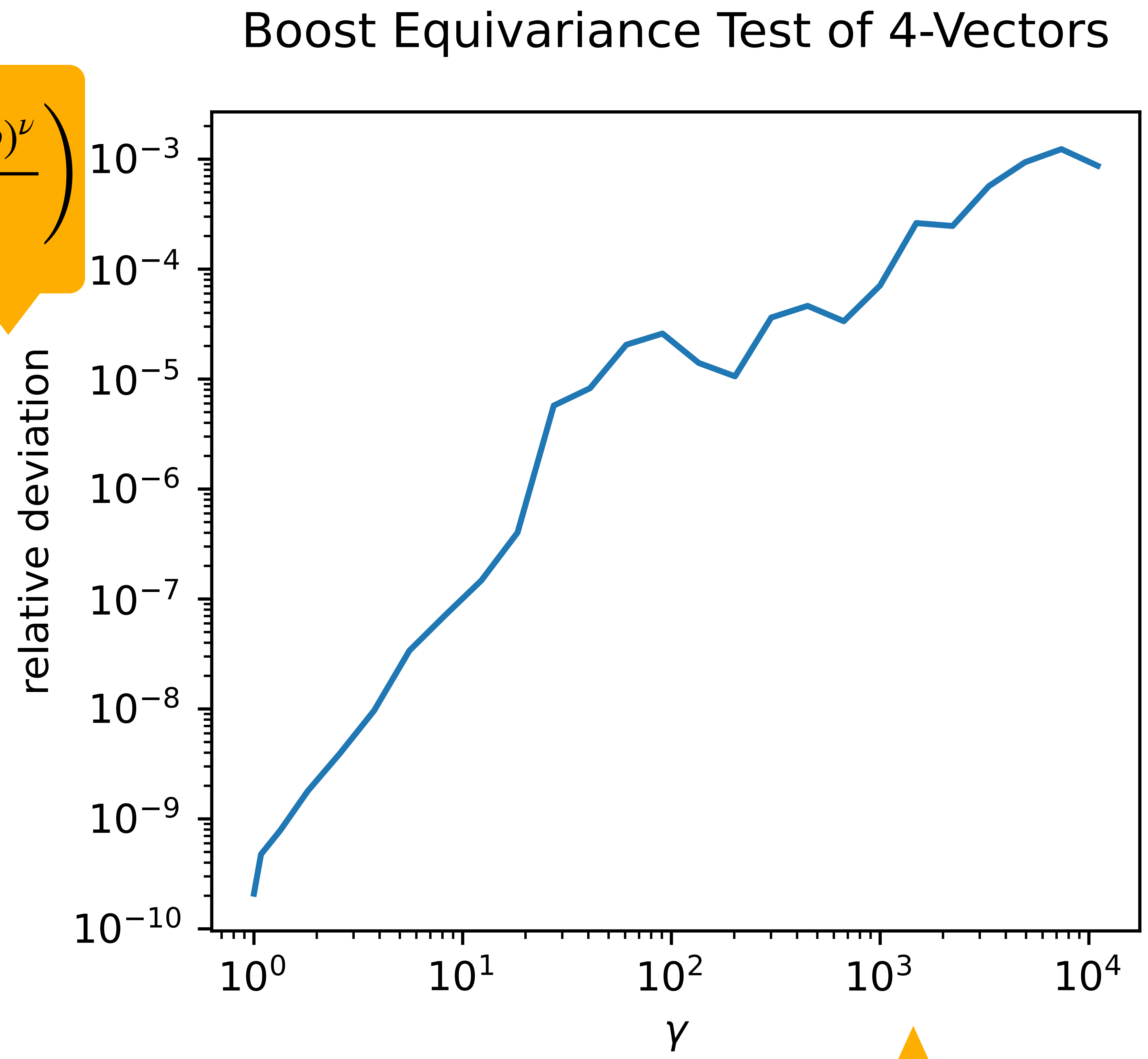
Experimental Results

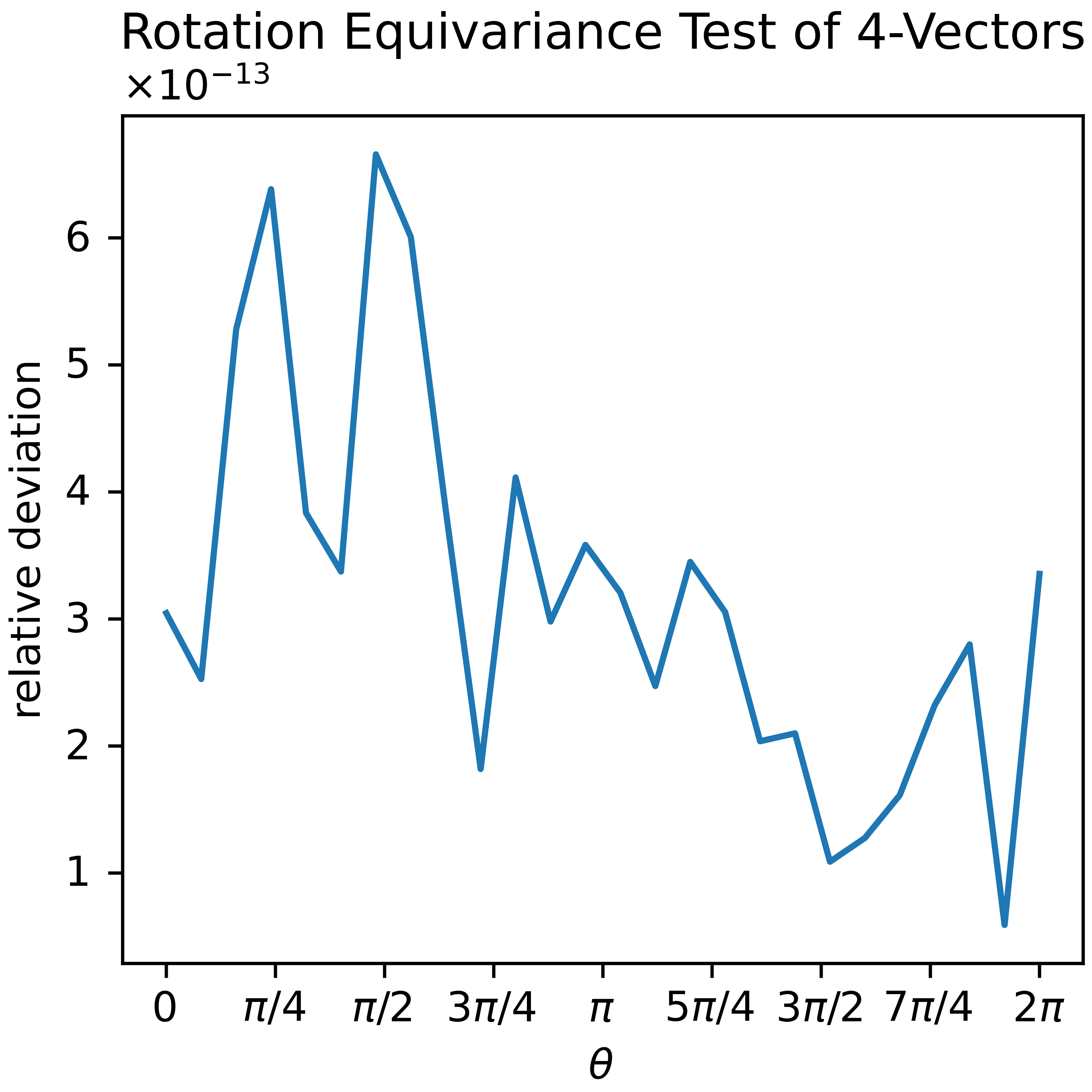
Experiment

Lorentz Group Equivariance Test

We expect $\mathrm{LGAE}(\Lambda^{\mu}{}_{\nu} p^{\nu}) = \Lambda^{\mu}{}_{\nu}\mathrm{LGAE}(p^{\nu})$

$$\underset{\substack{p \in J \\ \mu \in \{0,1,2,3\}}}{\mathrm{mean}} \left(\frac{\mathrm{LGAE}(\Lambda^{\mu}_{\nu} p^{\nu}) - \Lambda^{\mu}_{\nu}\mathrm{LGAE}(p)^{\nu}}{\Lambda^{\mu}_{\nu}\mathrm{LGAE}(p)^{\nu}} \right)$$





$v = 0.9999999917553856c$

Equivariant up to numerical errors

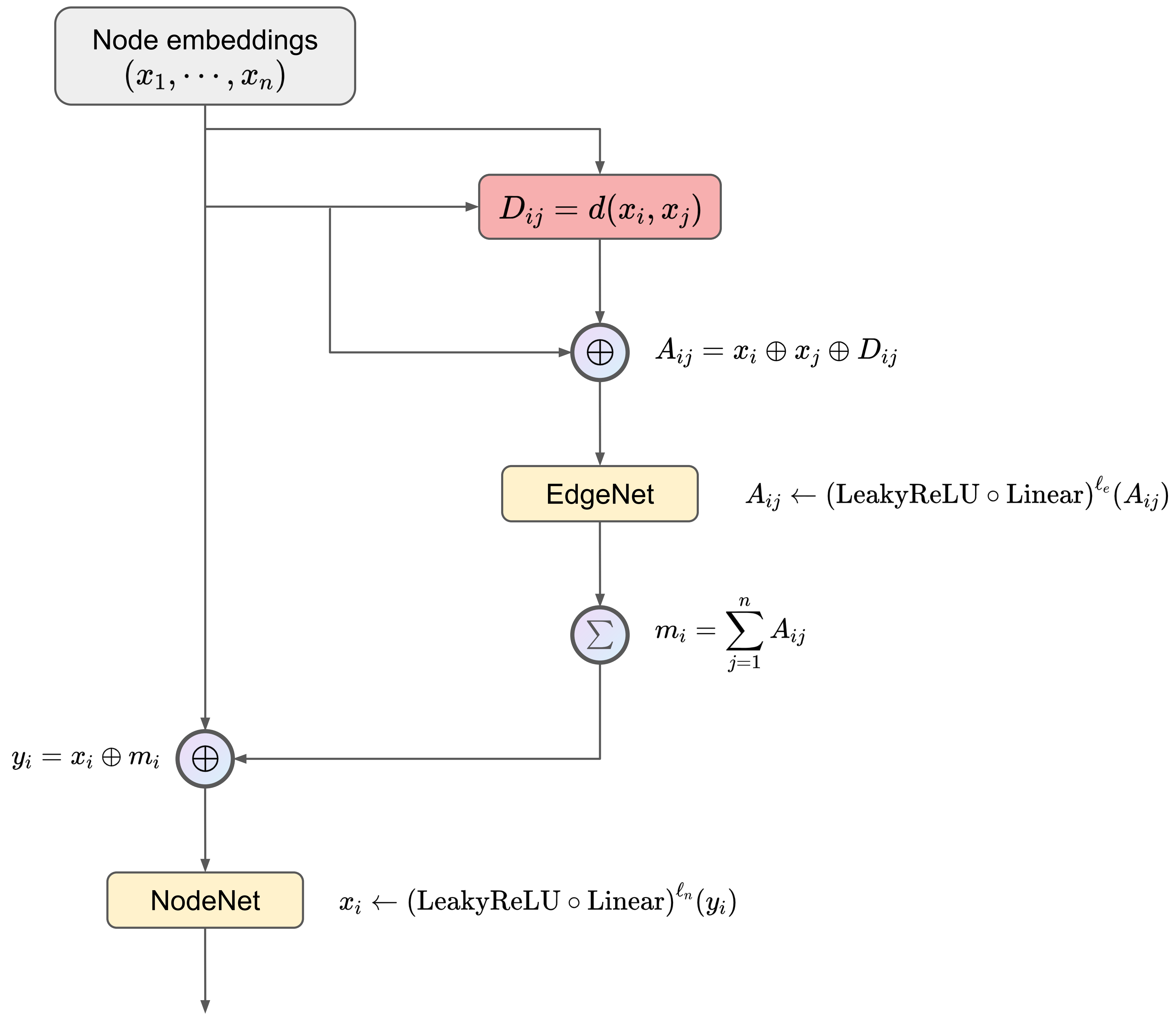
Description

Settings

- `JetNet` dataset (Detailed description: https://jet-net.github.io/jetnet/)
    - Gluon, top quark, light quark, W boson, and Z boson jets with $\mathcal{O}(1\,\mathrm{TeV})$ transverse momentum, produced in $13\,\mathrm{TeV}$ proton-proton collisions in a simplified detector, with 170k-180k jets per category.
- Training data: gluon and light quark jets (QCD) from the `JetNet` dataset.
- Signal jets for anomaly detection: top quark, W boson, and Z boson jets.
- Baseline models
    - Fully connected message-passing, graph neural network autoencoder (GNNAE) adapted from [arXiv:2012.00173] and [arXiv:2111.12849]
    - Convolutional neural network autoencoder (CNNAE)

Model

Baseline: GNNAE



- Fully connected message passing graph neural network adapted from arXiv:2012.00173
- Aggregation
    - Jet-level (**GNNAE-JL**): mean aggregation
        - Permutation invariant
    - Particle-level (**GNNAE-PL**): node-wise linear mixing, based on high-performing PGAE network [arXiv:2111.12849]
        - Permutation equivariant

GitHub Repo: https://github.com/zichunhao/gnn-jet-autoencoder

Model

Summary of Equivariance of Selected Models

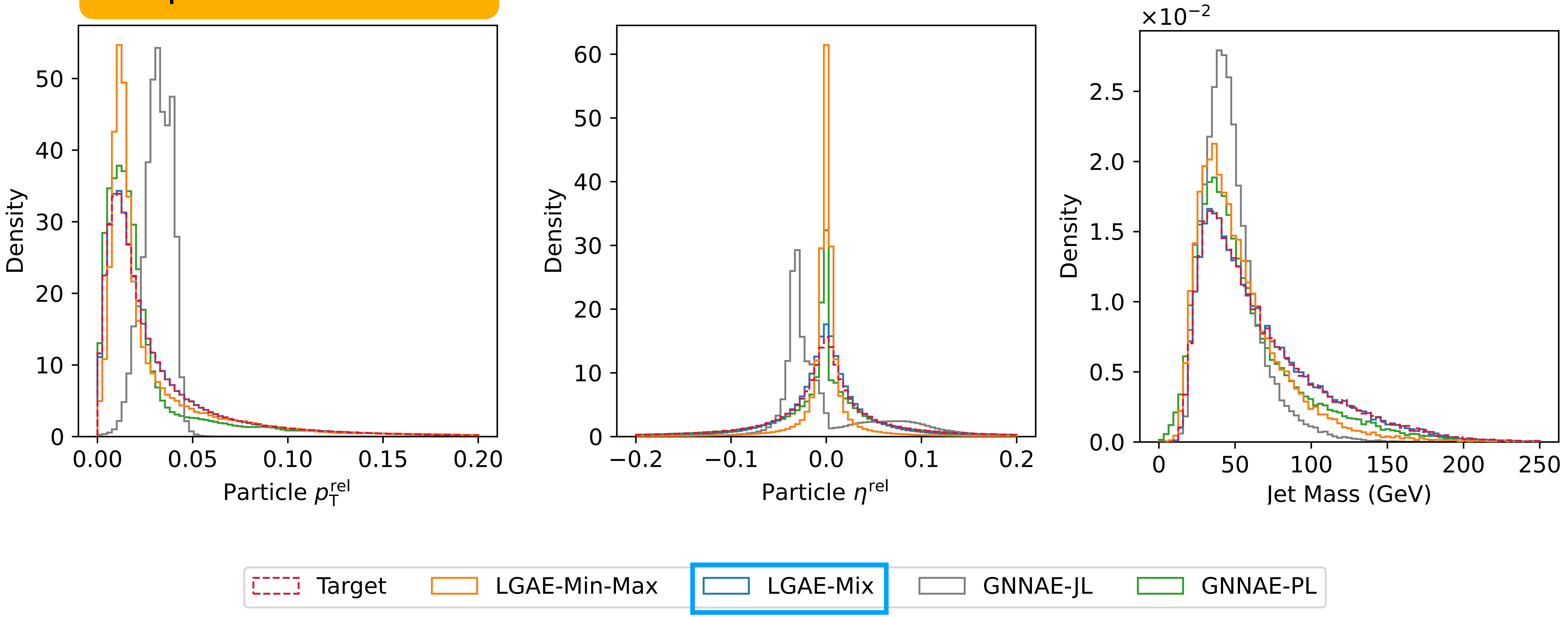
Model	Aggregation	Model's Name	Lorentz Symmetry	Permutation Symmetry
LGAE	Mix	LGAE-Mix	✓ (equivariance)	✗
	Min⊕Max	LGAE-Min-Max	✓ (equivariance)	✓ (invariance)
GNNAE	Particle level	GNNAE-PL	✗	✓ (equivariance)
	Jet level	GNNAE-JL	✗	✓ (invariance)

Reconstruction

Particle- and Jet-Level Features

Compression level $\approx 60\,\%$

LGAE-Mix has the best reconstruction performance in terms of the particle- and jet-level feature distribution



Reconstruction

Quantitative Measures

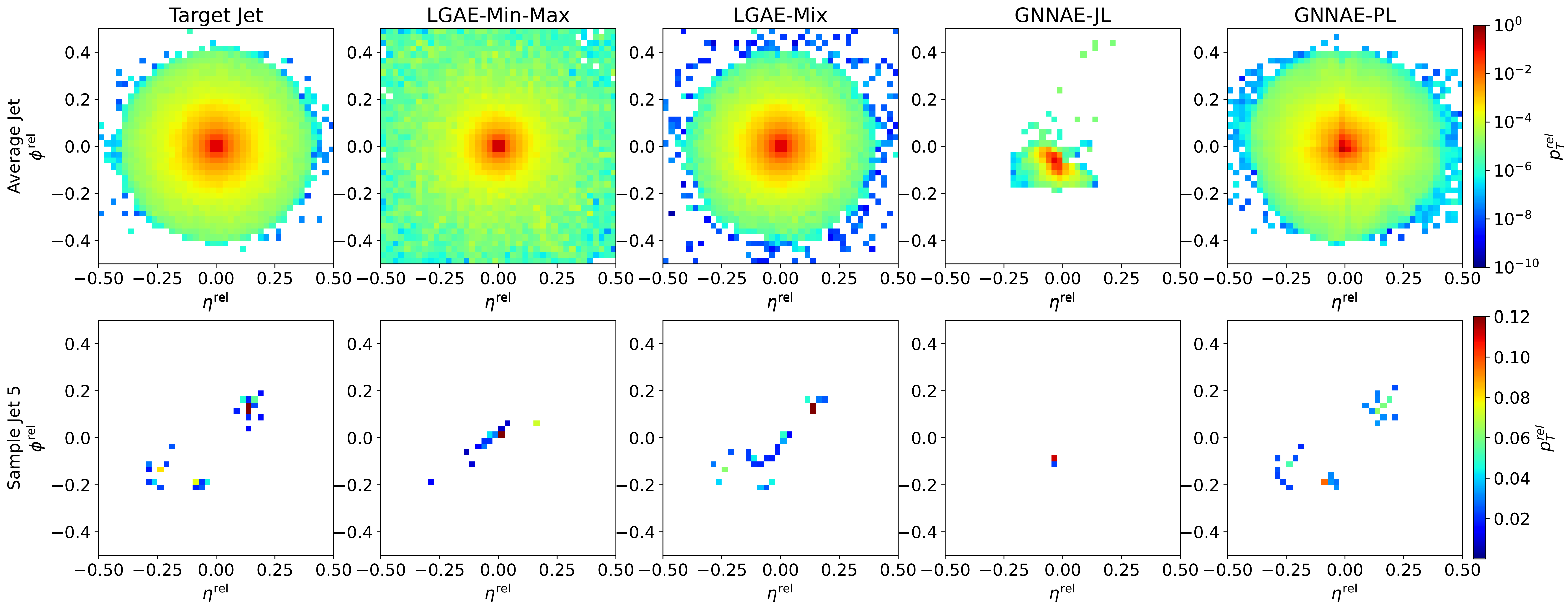
Model	Aggregation	Latent space	Jet mass		Jet p_T		Jet η		Jet ϕ	
			Median	IQR	Median	IQR	Median	IQR	Median	IQR
LGAE	Min-max	$\tau_{(1/2,1/2)} = 4$ (56.67%)	*0.096*	*0.134*	*0.097*	*0.109*	$< \mathbf{10^{-3}}$	*0.004*	$< \mathbf{10^{-3}}$	*0.002*
		$\tau_{(1/2,1/2)} = 7$ (96.67%)	−0.139	0.287	−0.221	0.609	$< \mathbf{10^{-3}}$	0.021	$< \mathbf{10^{-3}}$	0.007
	Mix	$\tau_{(1/2,1/2)} = 9$ (61.67%)	$< \mathbf{10^{-3}}$	**0.003**	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$
		$\tau_{(1/2,1/2)} = 13$ (88.33%)	$< \mathbf{10^{-3}}$	**0.003**	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$	$< \mathbf{10^{-3}}$
GNNAE	Jet-level	$\dim(L) = 45$ (50.00%)	0.326	*0.667*	*0.030*	*0.088*	*0.005*	*0.040*	*0.001*	0.021
		$\dim(L) = 90$ (100.00%)	3.7	2.6	*0.030*	0.089	0.292	0.433	0.006	0.021
	Particle-level	$\dim(L) = 2 \times 30$ (66.67%)	*0.277*	0.299	*0.037*	0.110	0.002	*0.010*	−0.001	0.005
		$\dim(L) = 3 \times 30$ (100.00%)	0.339	*0.244*	0.050	*0.094*	*−0.001*	0.011	$< \mathbf{10^{-3}}$	0.005
CNNAE	Linear layer	$\dim(L) = 55$ (61.67%)	−0.030	0.042	−0.021	0.017	$< \mathbf{10^{-3}}$	0.017	$< \mathbf{10^{-3}}$	0.003

Reconstruction

Reconstruction of Jet Images

Compression level $\approx 60\,\%$

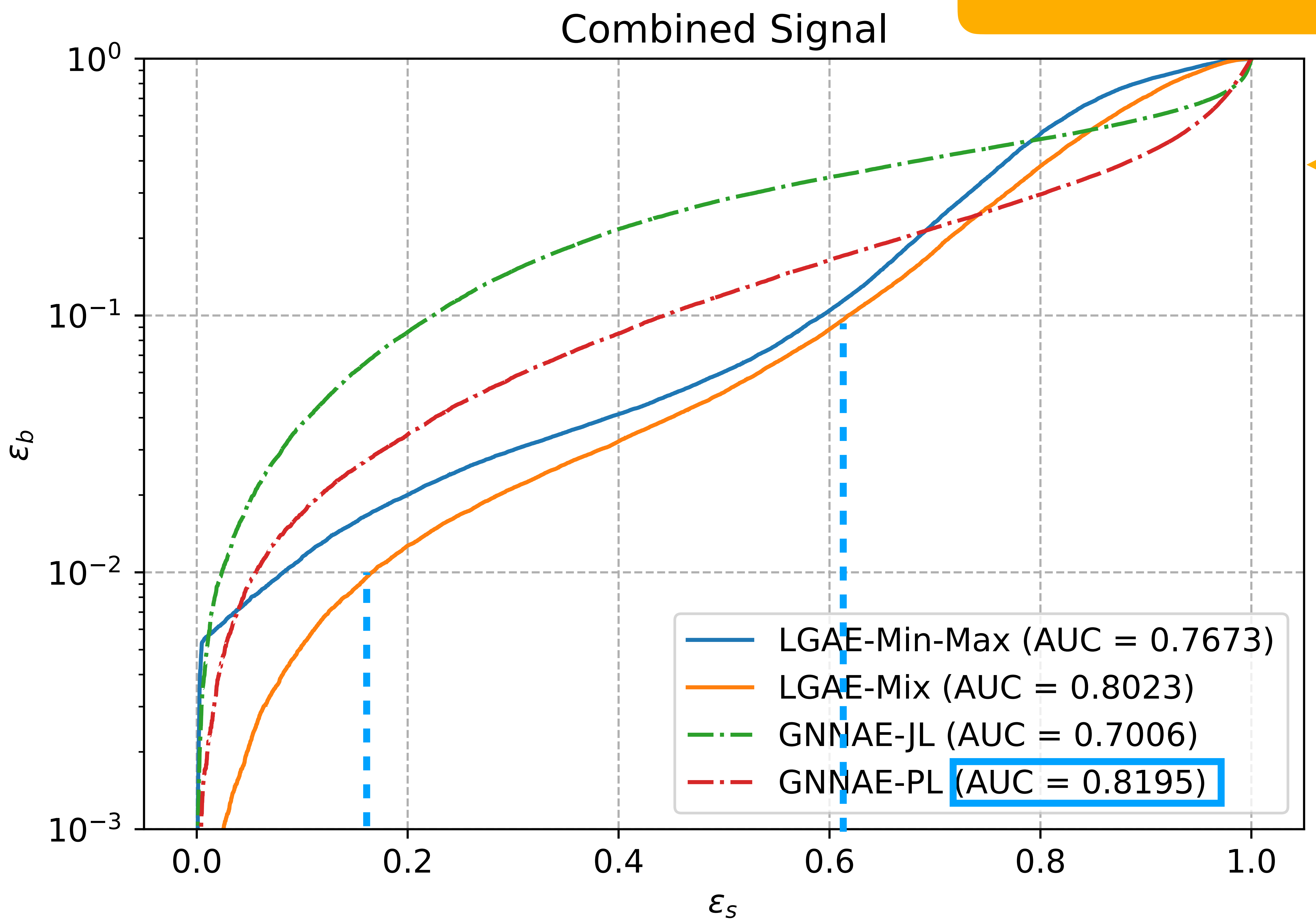
Symmetry is respected even though reconstruction is not ideal



Anomaly Detection

Tagging All Signals (Top, W, and Z Combined)

LGAEs have better ε_s at given ε_b's

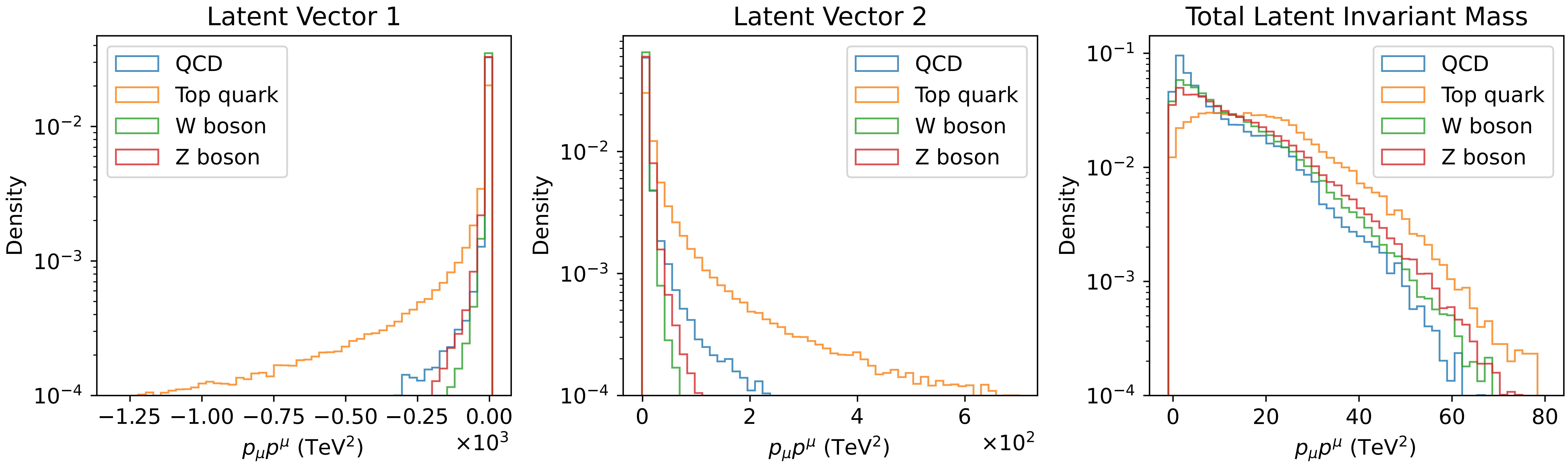


Score: MSE

Latent Space Analysis

Distributions of Derived Quantities

The representations are Lorentz scalars and 4-vectors!

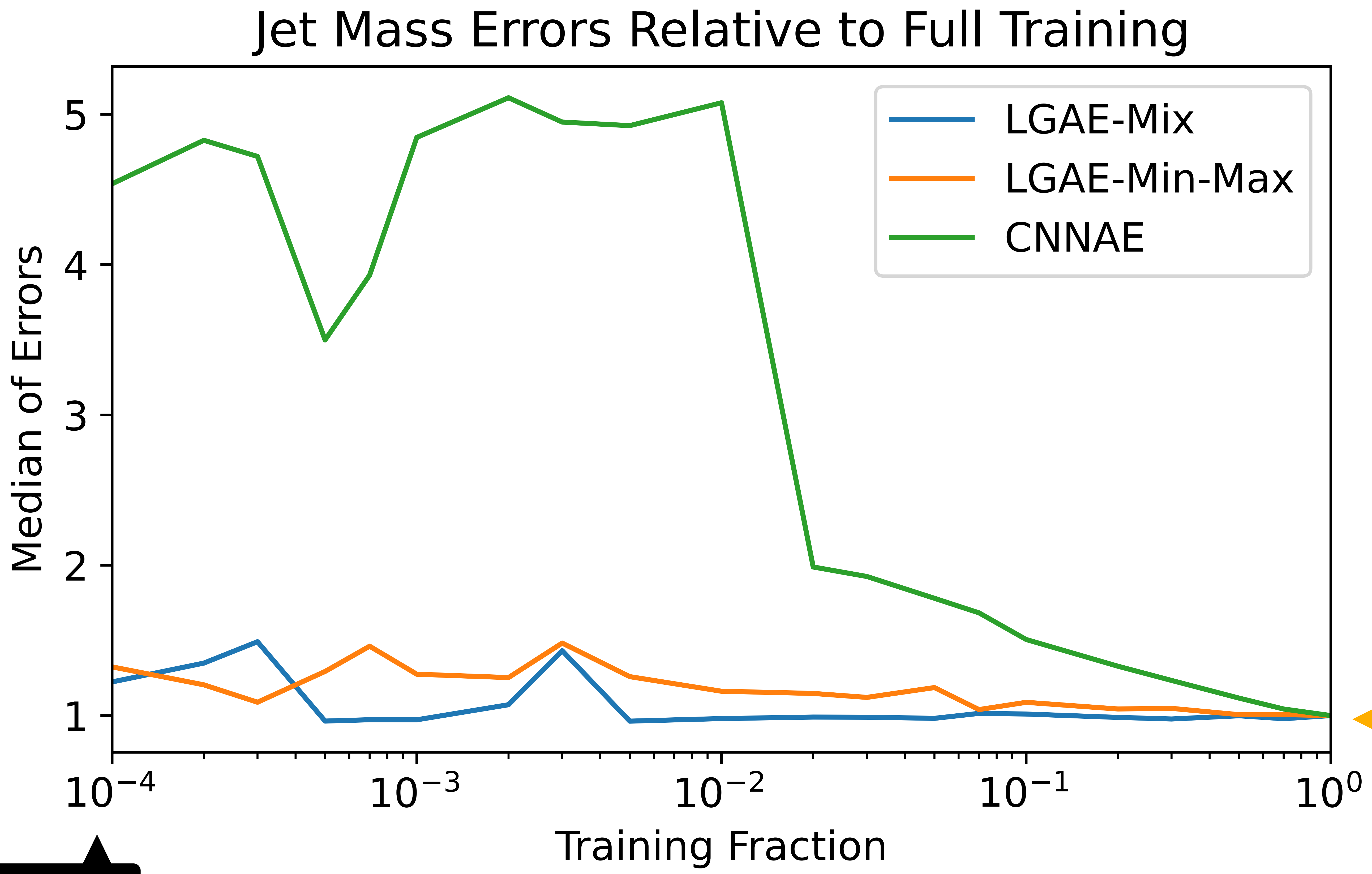


Possibly why top tagging has the best ROC curve

Data Efficiency

Generalizability: What If We Train the Model with Less Data?

$$\delta = \left| \frac{M_{\text{reco}} - M_{\text{true}}}{M_{\text{true}}} \right|$$



Normalized by the error with 100% training dataset

$\mathcal{O}(10)$ training jets

Conclusion

Conclusion

Takeaways and Next Steps

- Adding inductive biases and symmetry has shown to improve NNs in terms of performance, interpretability, and data efficiency.
- We embedded **Lorentz symmetry** into an autoencoder.
- LGAE-Mix model has a **better performance** in **reconstruction** and **anomaly detection** (in a HEP context) than the baseline GNNAEs.
- The LGAEs have a **promising interpretability** in latent space and more data efficient.
- Possible future works: further latent space analysis and LorentzNet-based autoencoders.

Conclusion

Funding Acknowledgement

- This work was supported by **US DOE** (No. DE-AC02-07CH11359, No. DE-SC0021187, and No. DE-SC0021396), **US NSF** (OAC-2117997), and the **European Research Council** (Grant Agreement No. 772369).

Backup

Lorentz Group

Irreducible Representations for Small (j^+, j^-)

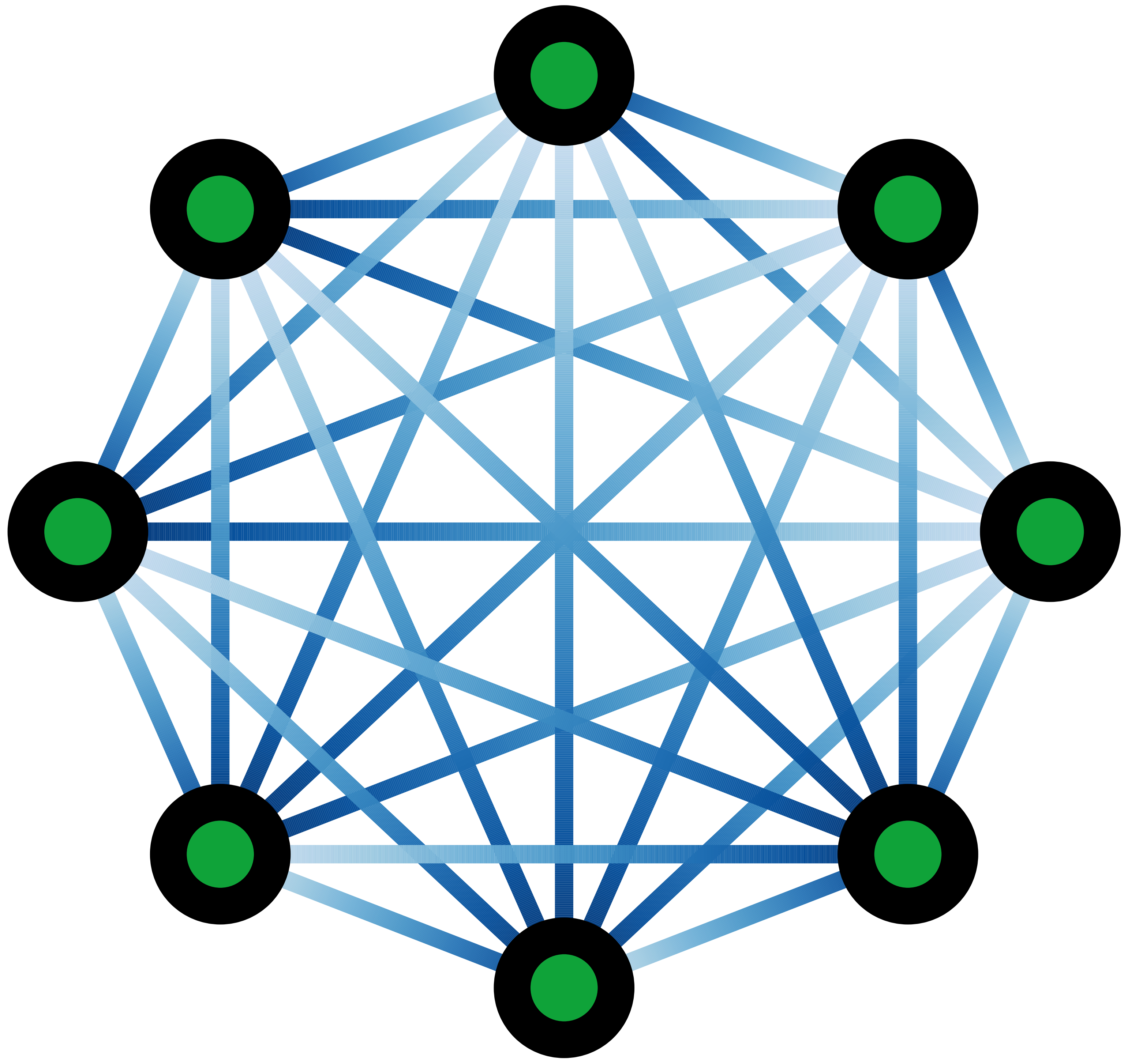
- Classified by two half integers: (j^+, j^-).

	$j^+ = 0$	$j^+ = \frac{1}{2}$	$j^+ = 1$	$j^+ = \frac{3}{2}$
$j^- = 0$	**Scalar** Dimension: 1	**Left-handed Weyl spinor** Dimension: 2	**Self-dual 4-form** Dimension: 3	Dimension: 4
$j^- = \frac{1}{2}$	**Right-handed Weyl spinor** Dimension: 2	**4-vector** Dimension: 4	Dimension: 6	Dimension: 8
$j^- = 1$	**Anti-self-dual 4-form** Dimension: 3	Dimension: 6	**Traceless symmetric tensor** Dimension: 9	Dimension: 12
$j^- = \frac{3}{2}$	Dimension: 4	Dimension: 8	Dimension: 12	Dimension: 16

Model

Graph Neural Networks

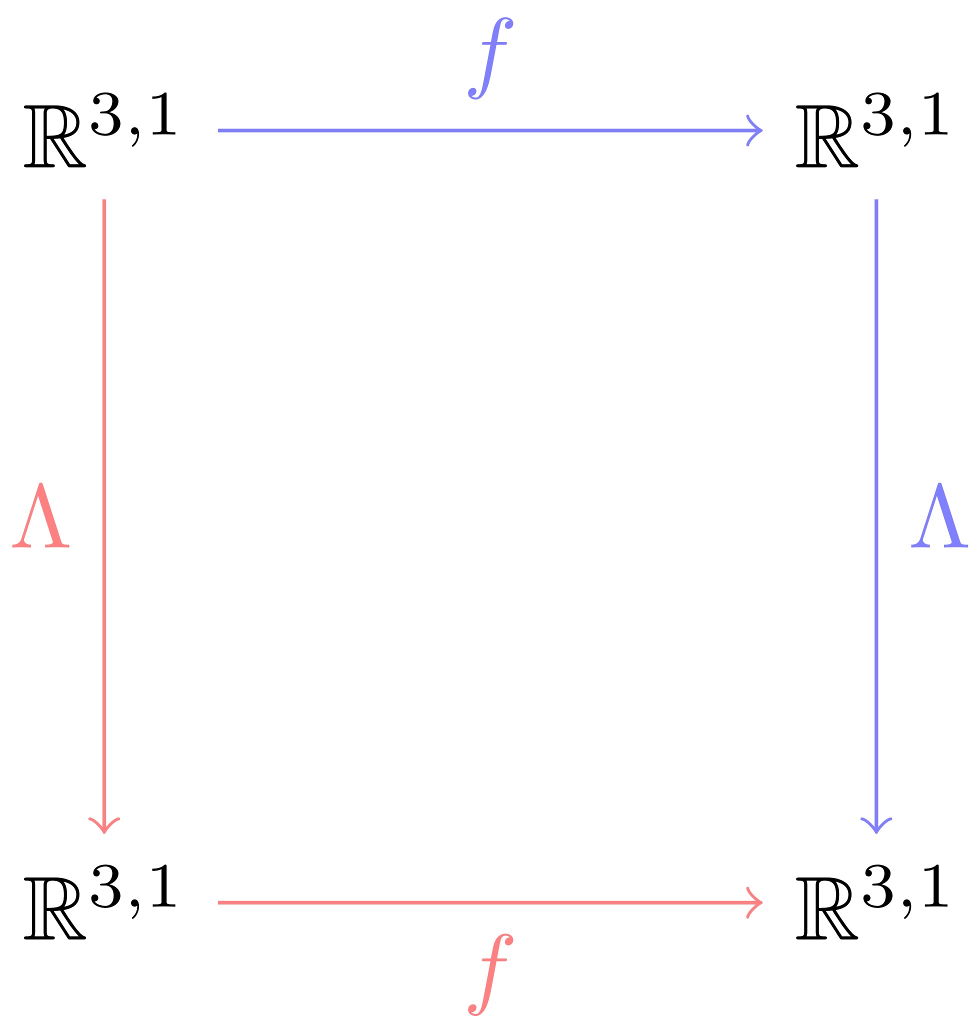
$p = [E, p_x, p_y, p_z] \equiv [p_\mathrm{T}, \eta, \phi, m]$



- $G = \{V, E\}$, possibly with global features
  - Node features $\mathbf{v}_i$: particle 4-momentum
  - Edge features $\mathbf{e}_{ij}$
    - distance between particles
    - interactions between particles
  - Graph (global) features $\mathbf{u}$: jet mass

Model

Embedding Lorentz Group Symmetry



$$\Lambda \cdot f(p) = f(\Lambda \cdot p)$$

- Method: **equivariance** with respect to the Lorentz group.
- Common approaches of achieving equivariance
  - Group convolutional kernels: generalization of CNN.
  - Fourier space: decomposition into **irreducible representations** (irreps).
- Advantages of achieving equivariance
  - Data efficiency
  - Interpretability

Model

Choices of Aggregation in LGAE

- Linear mixing (LGAE-Mix): concatenate nodes and linearly mix.
  - **Note**: We are imposing a specific order, so it breaks the permutation symmetry.
- Max/Min/Mean pooling.
  - Min/Max with respect to the Lorentz scalars.
  - Can concatenate these, such as min$\oplus$max and min$\oplus$max$\oplus$mean.

Experiment

Settings

- Loss functions
  - LGAE-Mix, GNNAE-PL, and CNNAE: MSE
  - LGAE-Min-Max and GNNAE-PL: Chamfer loss
    $$\mathcal{L}_{\text{chamfer}}(J_1, J_2) = \sum_{p_1 \in J_1} \min_{p_2 \in J_2} |p_1 - p_2|^2 + \sum_{p_2 \in J_2} \min_{p_1 \in J_1} |p_1 - p_2|^2.$$
  - Alternatives
    - Energy mover distance (EMD) [arXiv:1902.02346]: difficult computationally.
    - Hungarian loss (our implementation here): difficult to converge.

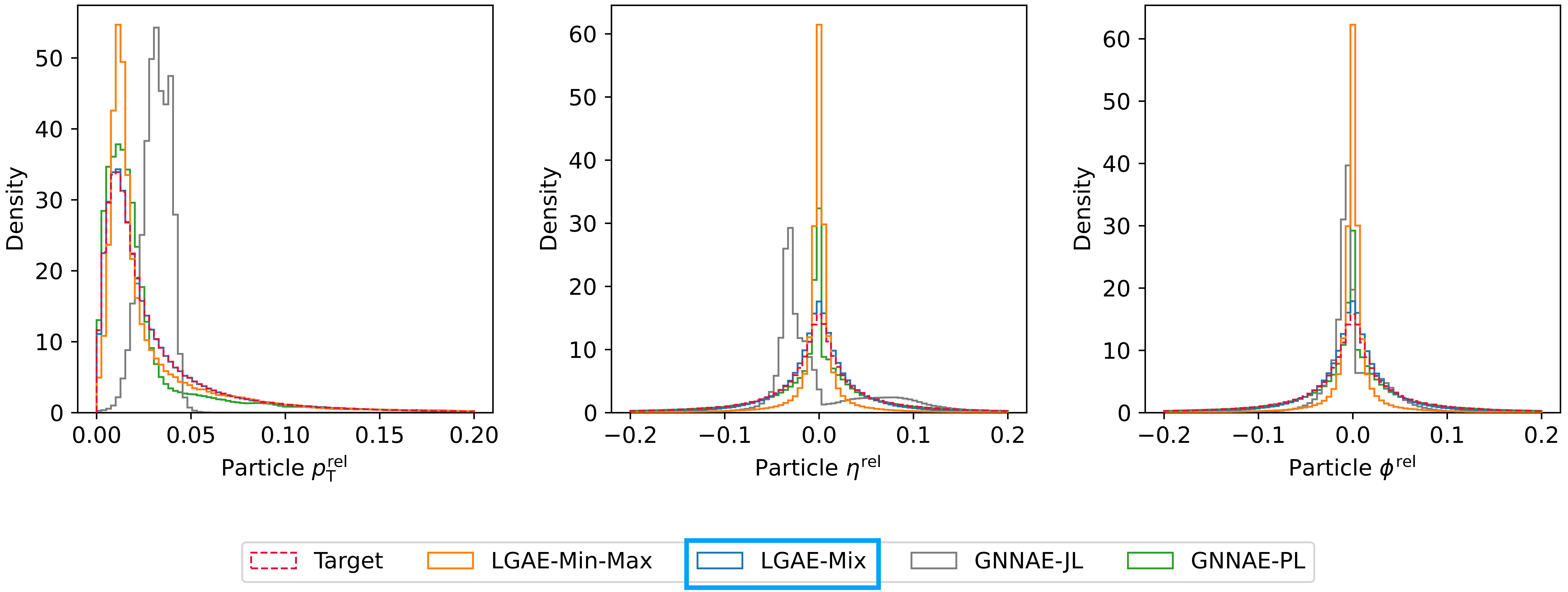
Experiment
LGAE Parameters

- Parameters to optimize: $\tau_{(m,n)}$ of each layer and the latent space.
  - Encoder: $\{\tau^{(t)}_{(m,n)}\}_{t=1}^{4} = (3,3,4,4)$.
  - Aggregation: {min-max, mix}.
  - Latent space dimension
    - $\tau_{(0,0)} = 1$
    - $\tau_{(1/2,1/2)} \in \{1,\ldots,14\}$
  - Decoder: $\{\tau^{(t)}_{(m,n)}\}_{t=1}^{4} = (3,3,4,4)$.

Experiment

Reconstruction of Particle Features

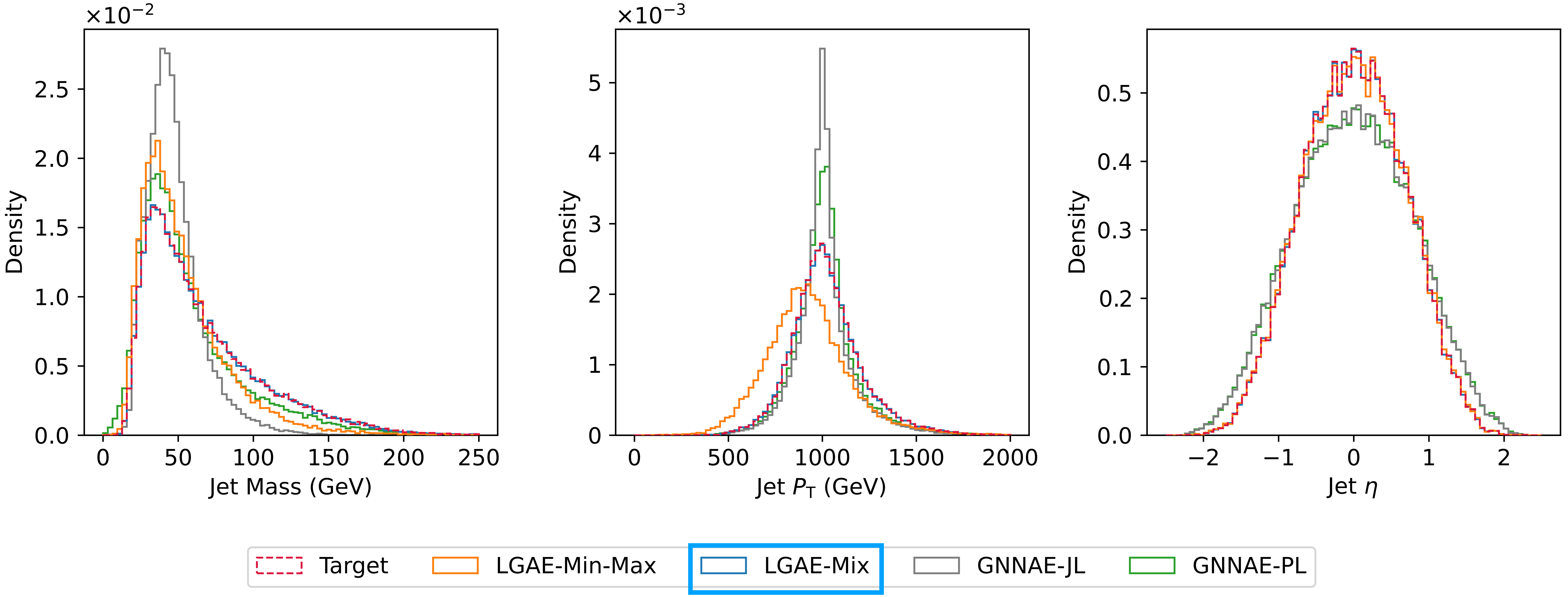
Compression level:
$\approx 60\,\%$



Experiment

Reconstruction of Jet Features

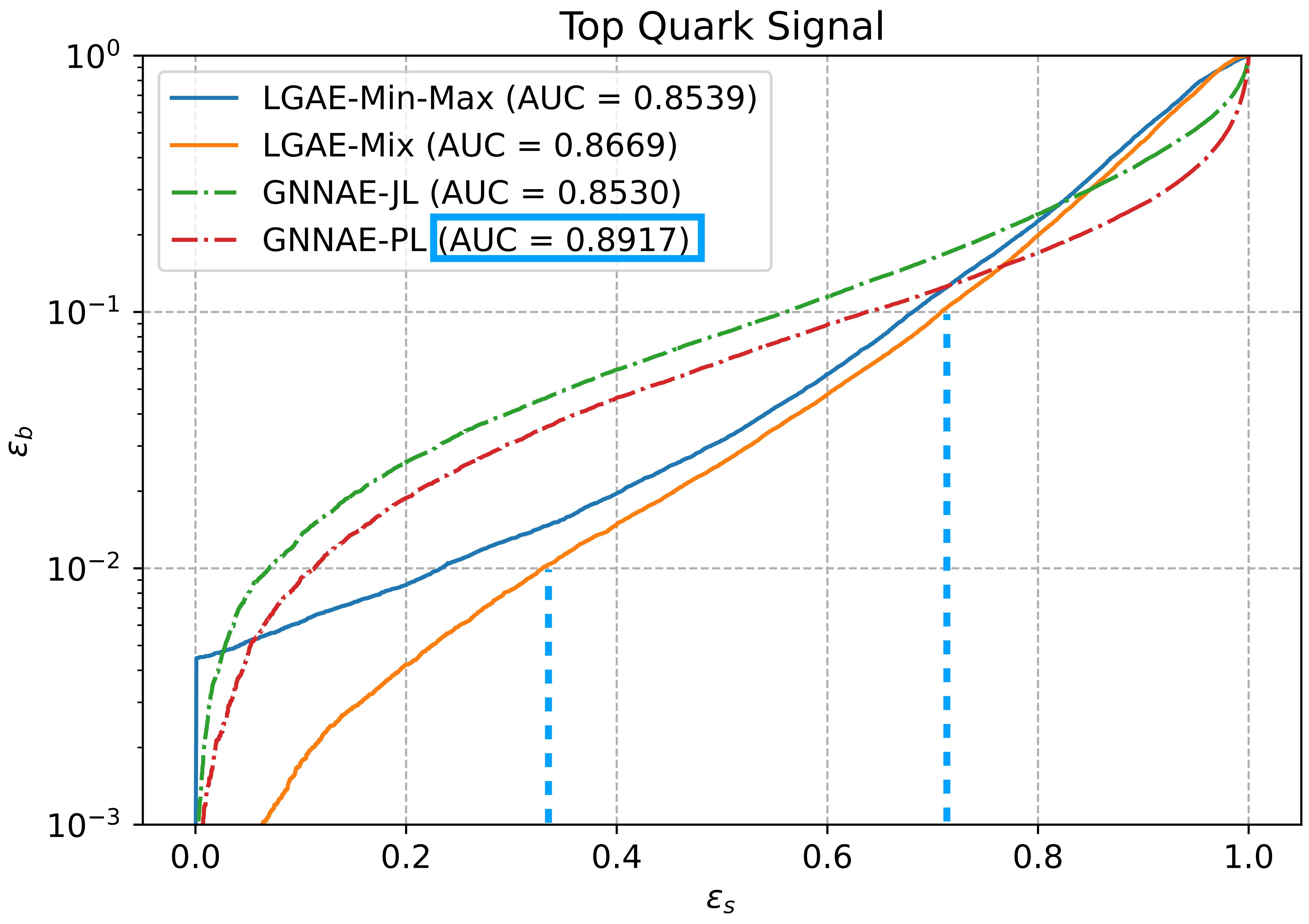
Compression level:
$\approx 60\,\%$



Experiment

Anomaly Detection: Top Tagging

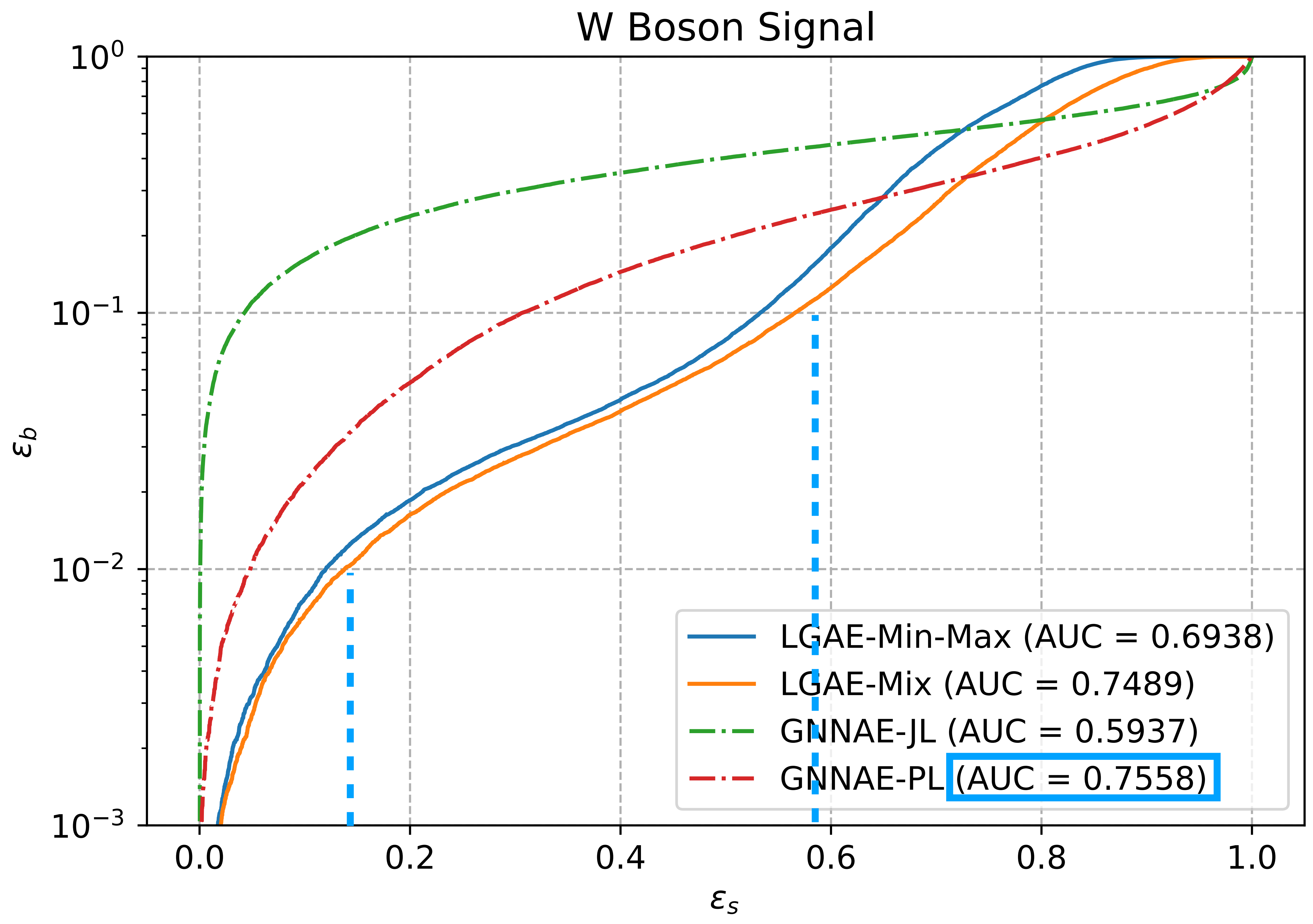
LGAEs have better ε_s at low ε_b



Experiment

Anomaly Detection: W Tagging

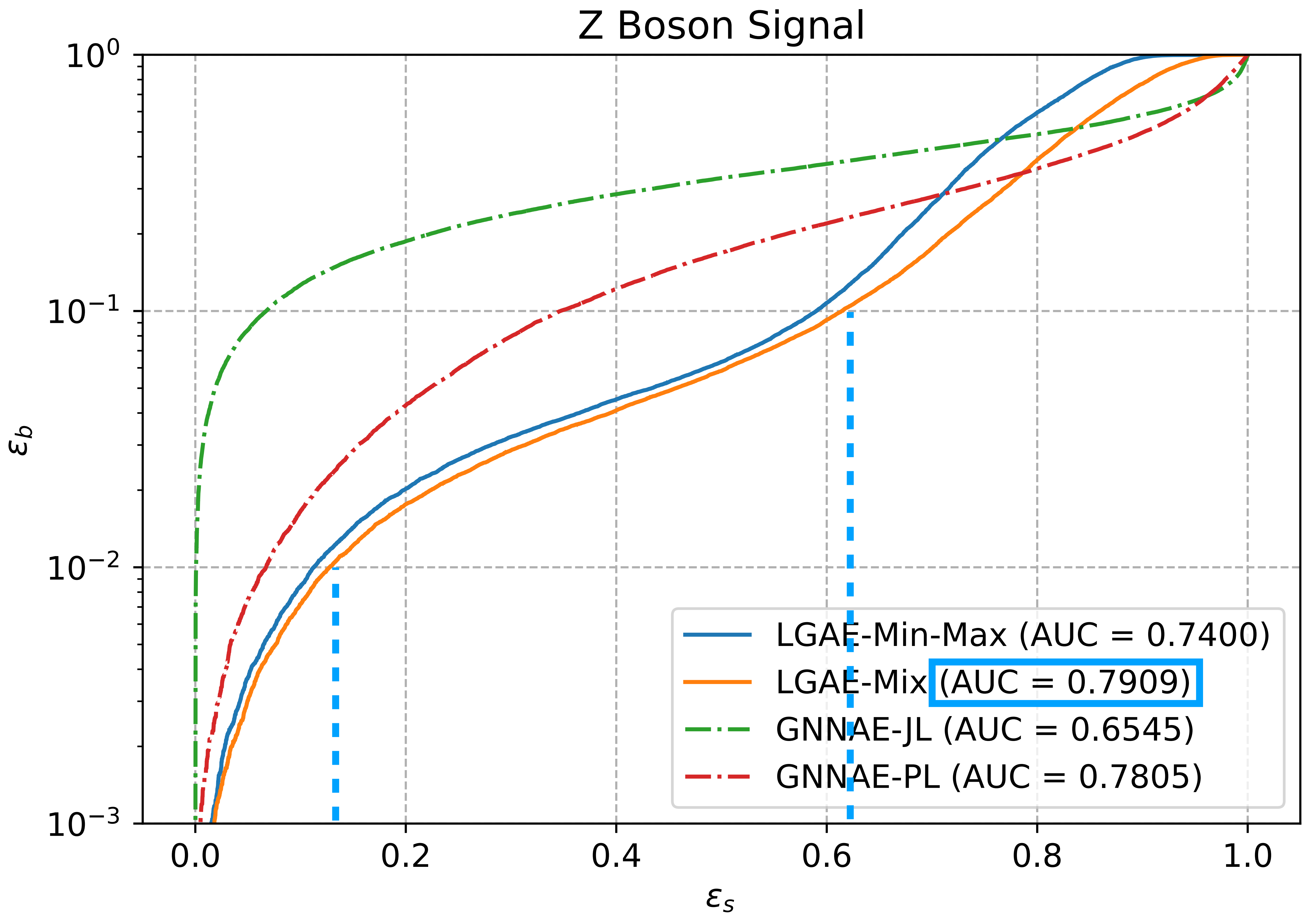
LGAEs have better ε_s at low ε_b



Experiment

Anomaly Detection: Z Tagging

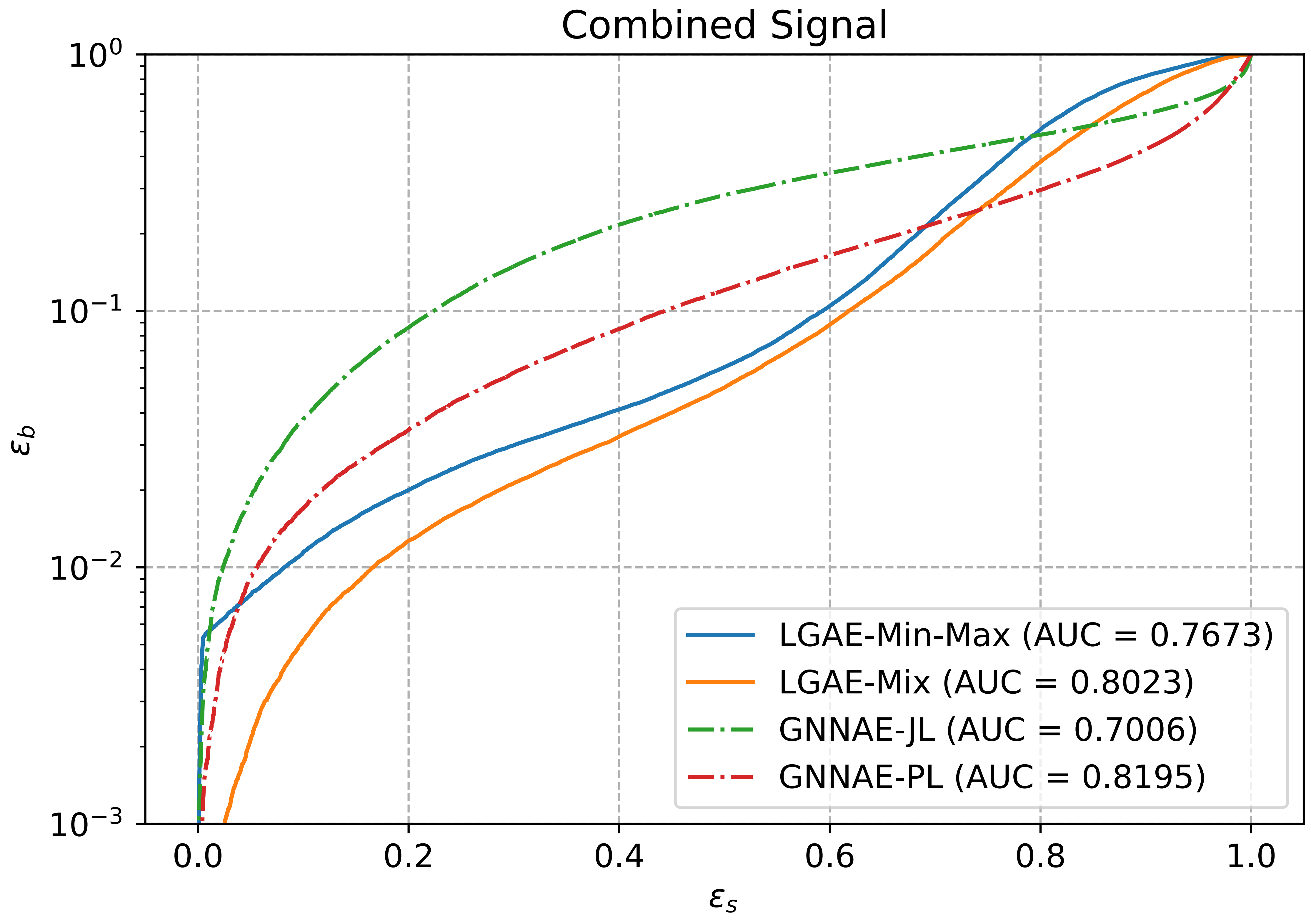
LGAEs have better ε_s at low ε_b

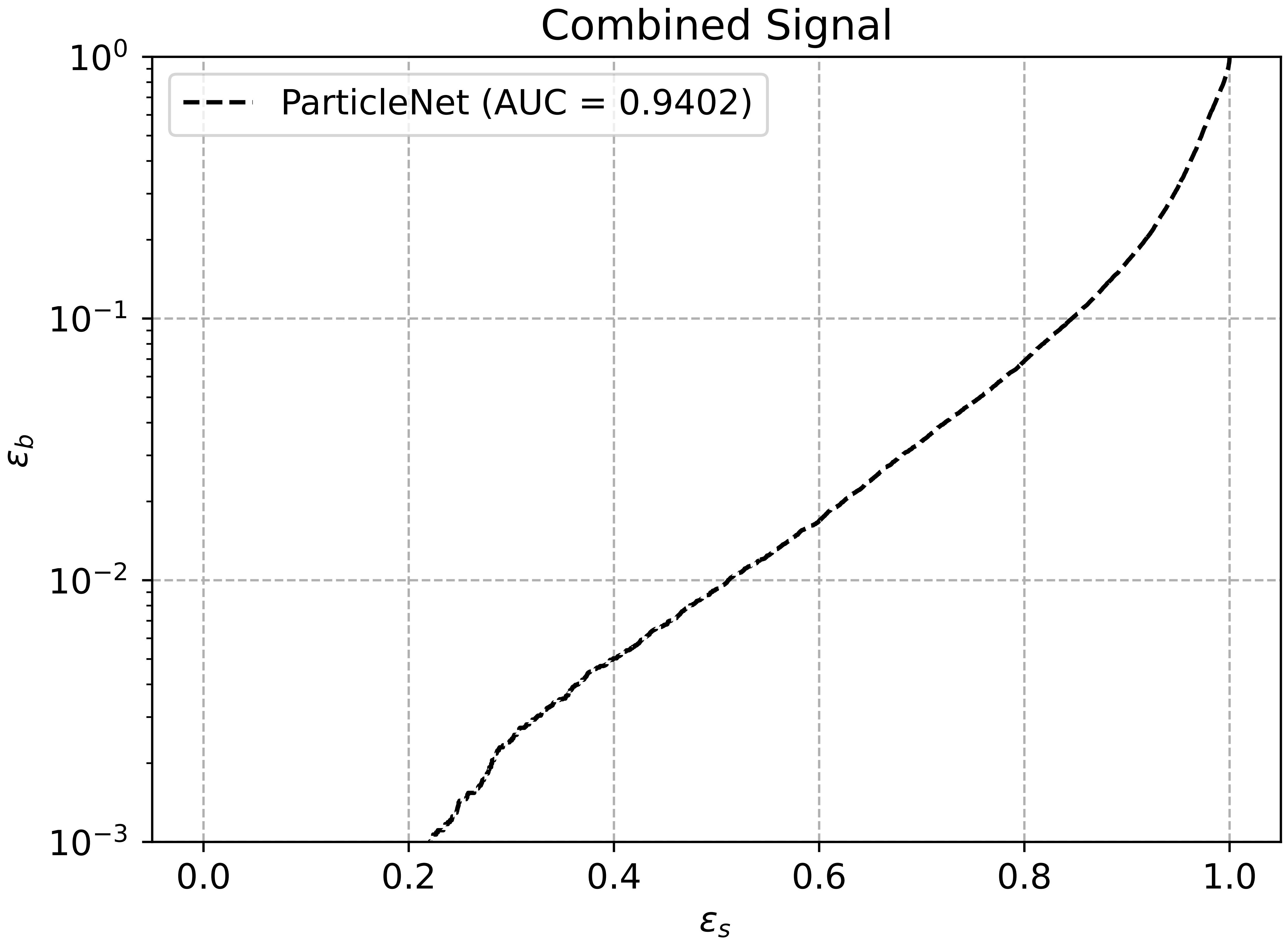


Experiment

Anomaly Detection: `ParticleNet`

Not as good as the SOTA supervised model, as expected



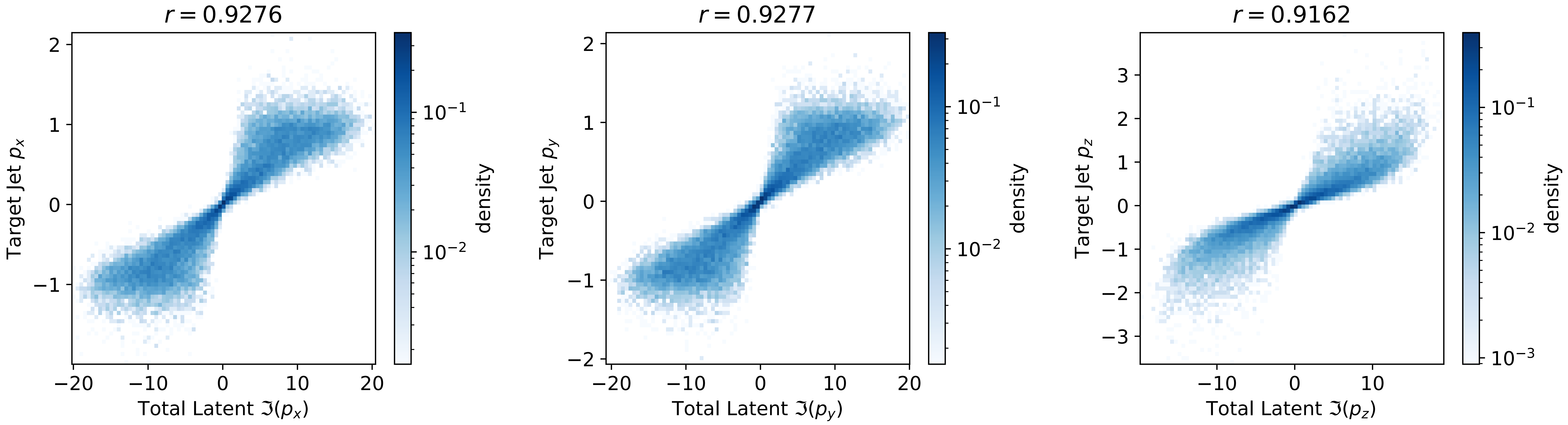


Experiment

Latest Space Analysis: Correlations

Jet 3-momentum encoded in the total latent 4-vector

Model: LGAE-Mix with 2 latent 4-vectors



No other strong correlations found → Possibly new useful quantities?