Self-supervised learning of jets using a realistic detector simulation

Kyle Cranmer, Etienne Dreyer, Eilam Gross, Nilotpal Kakati,
Dmitrii Kobilianskii, Patrick Rieck, Nathalie Soybelman

Hammers & Nails 2023
Ascona, Switzerland

representation space

“same”

representation space

“same”


"same"
"different"
representation space
"different"
"same"

Motivation

Given an adequate method of defining "sameness" pseudo-labels, self-supervised models can be trained to extract features without relying on explicit labels

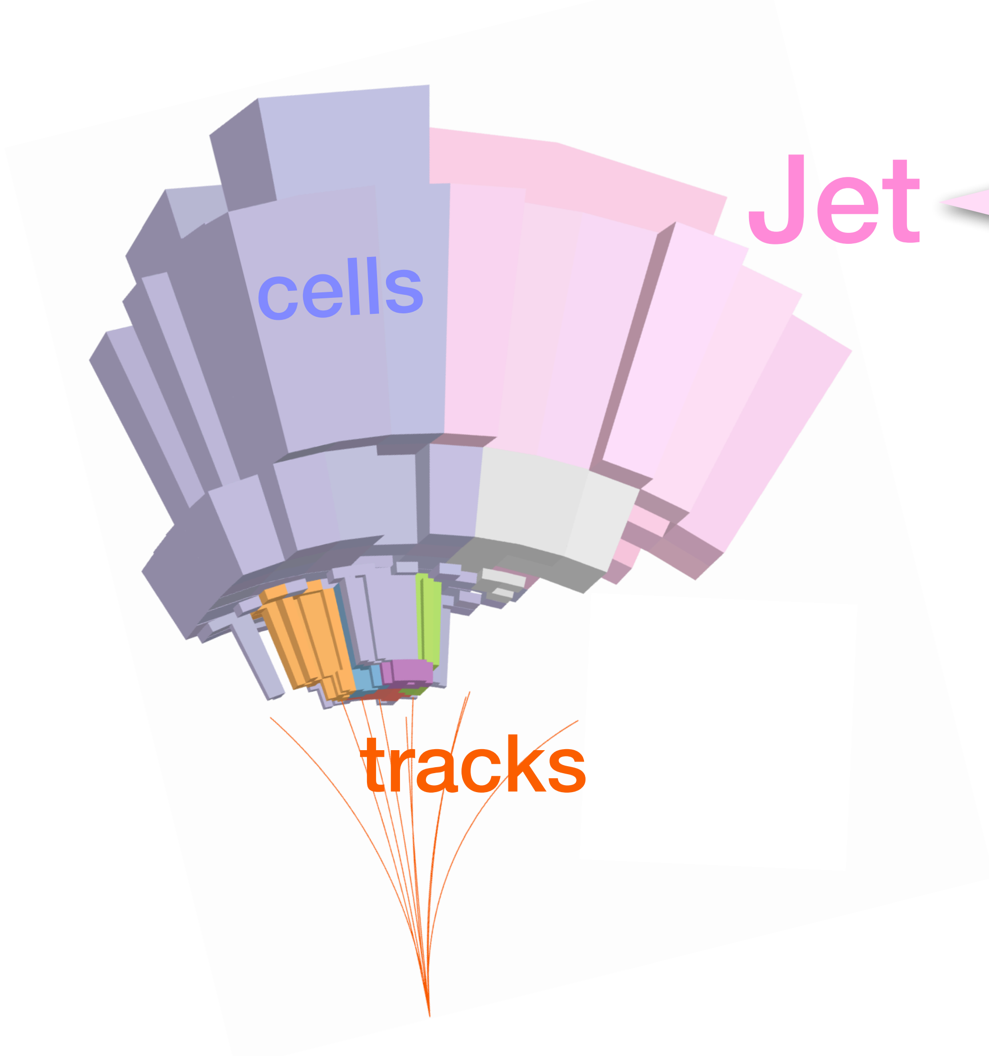
Motivation

Given an adequate method of defining "sameness" pseudo-labels, self-supervised models can be trained to extract features without relying on explicit labels

Potential perks for high energy physics…

- ✓ Training (pretext task) on huge datasets from real collisions
- ✓ Mitigate dependence of models on difference between simulation and reality
- ✓ Inject notions of "sameness" (e.g. symmetries) into learned representations [1]

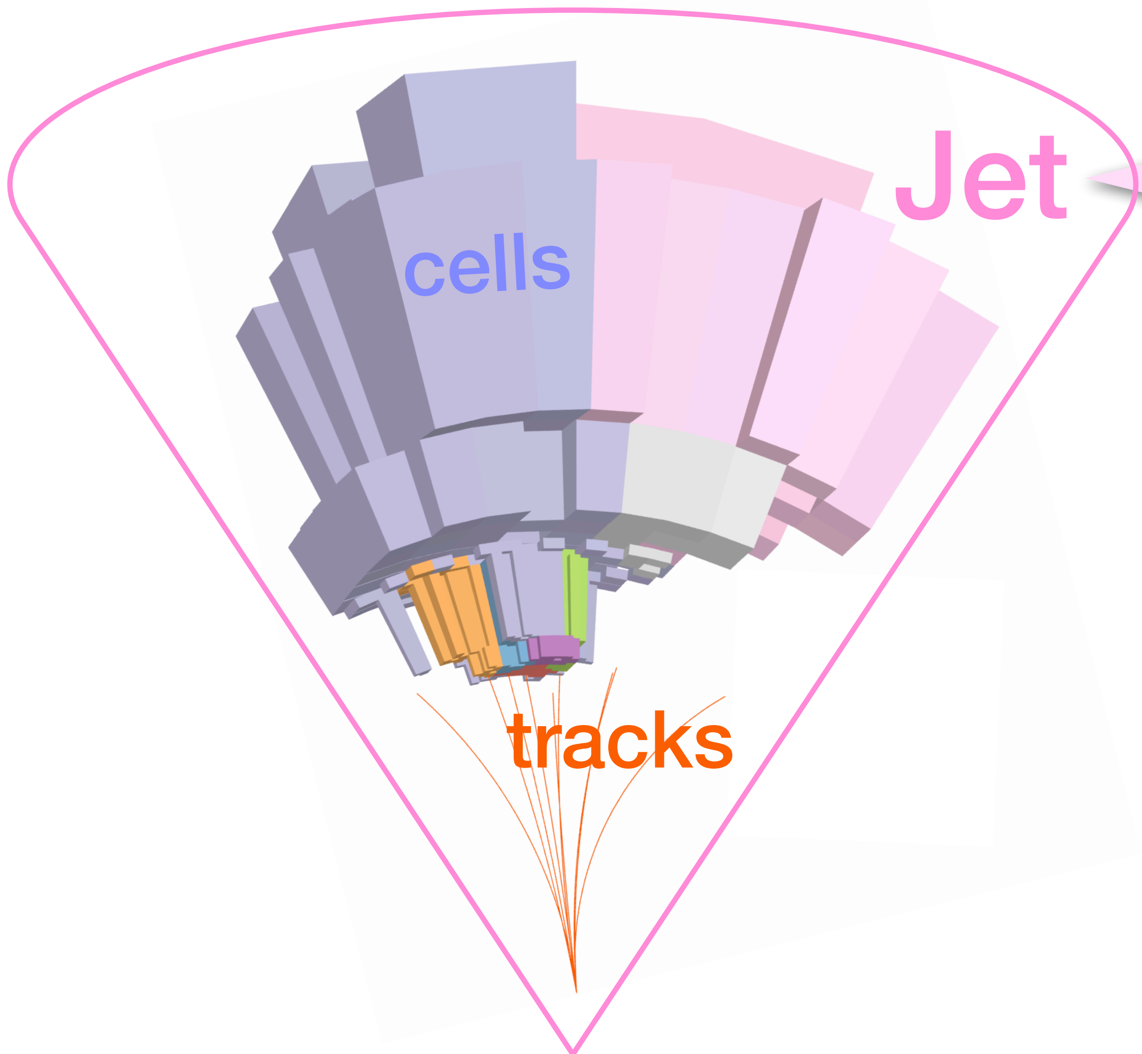
[1] Symmetries, Safety, and Self-Supervision. B. M. Dillon, G. Kasieczka, T. Plehn et al. (2021)



A correlated set of particles with several learnable properties:
(due to their composite nature)

- ✓ Process of origin (classification)
- ✓ Energy, mass (regression)
- ✓ Possible anomalous signatures

Contrastive learning of jets via detector replicas

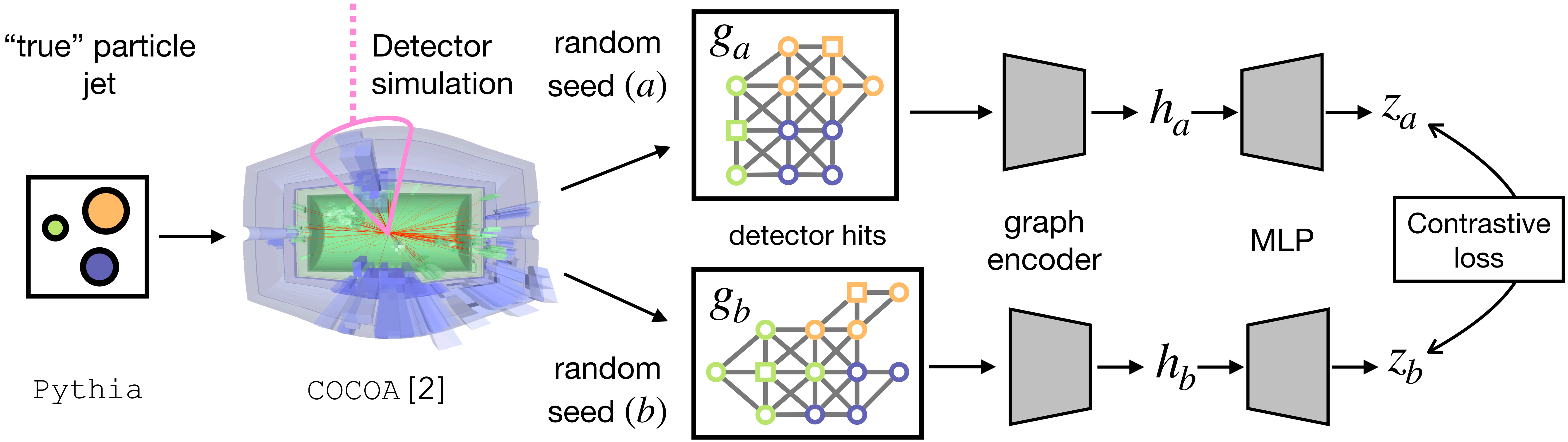


A correlated set of particles with several learnable properties:
(due to their composite nature)

- ✓ Process of origin (classification)
- ✓ Energy, mass (regression)
- ✓ Possible anomalous signatures

Note: similar effort by MIT/KIT/SLAC
(see their talk at BOOST)

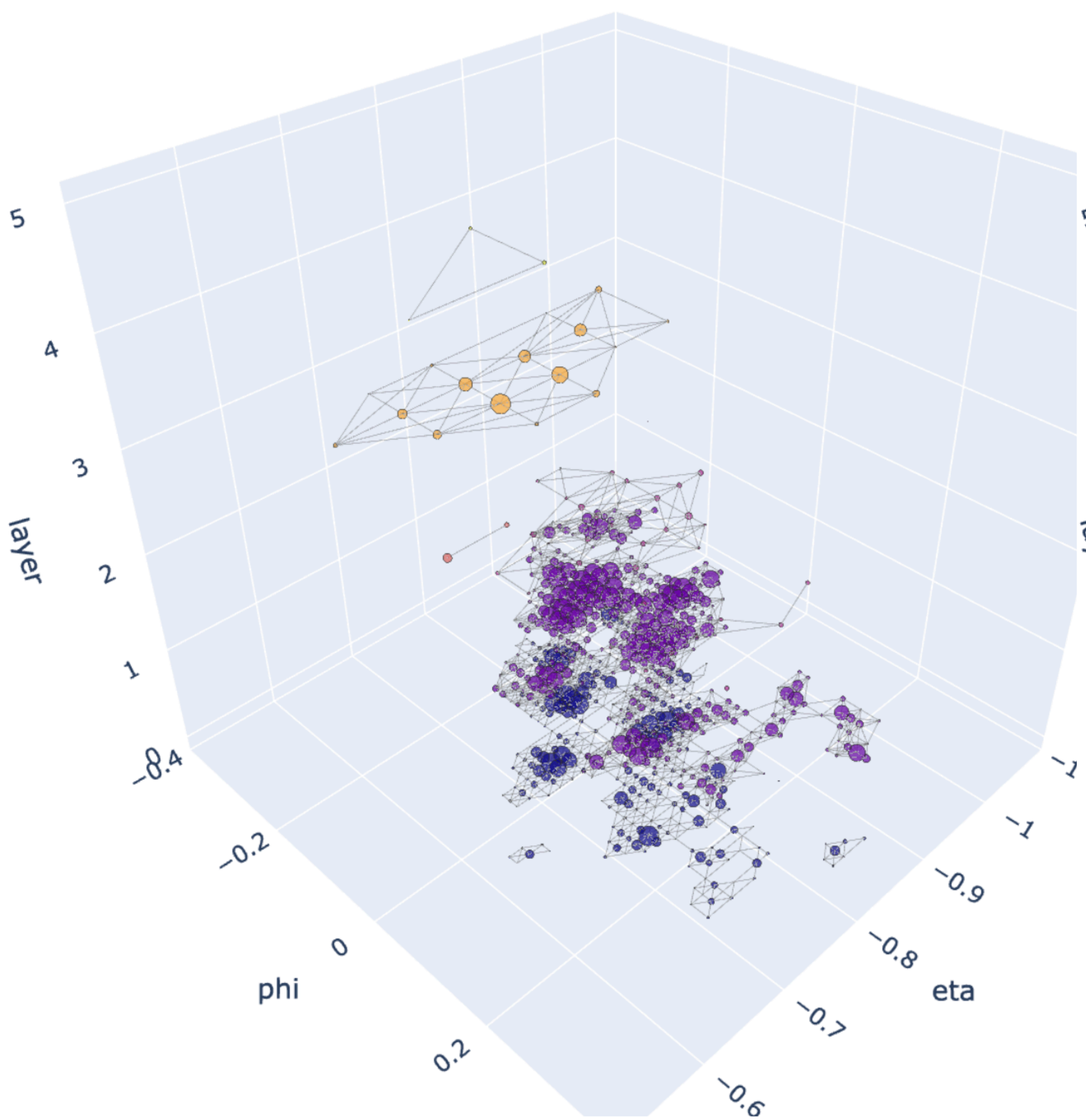
Contrastive learning of jets via detector replicas



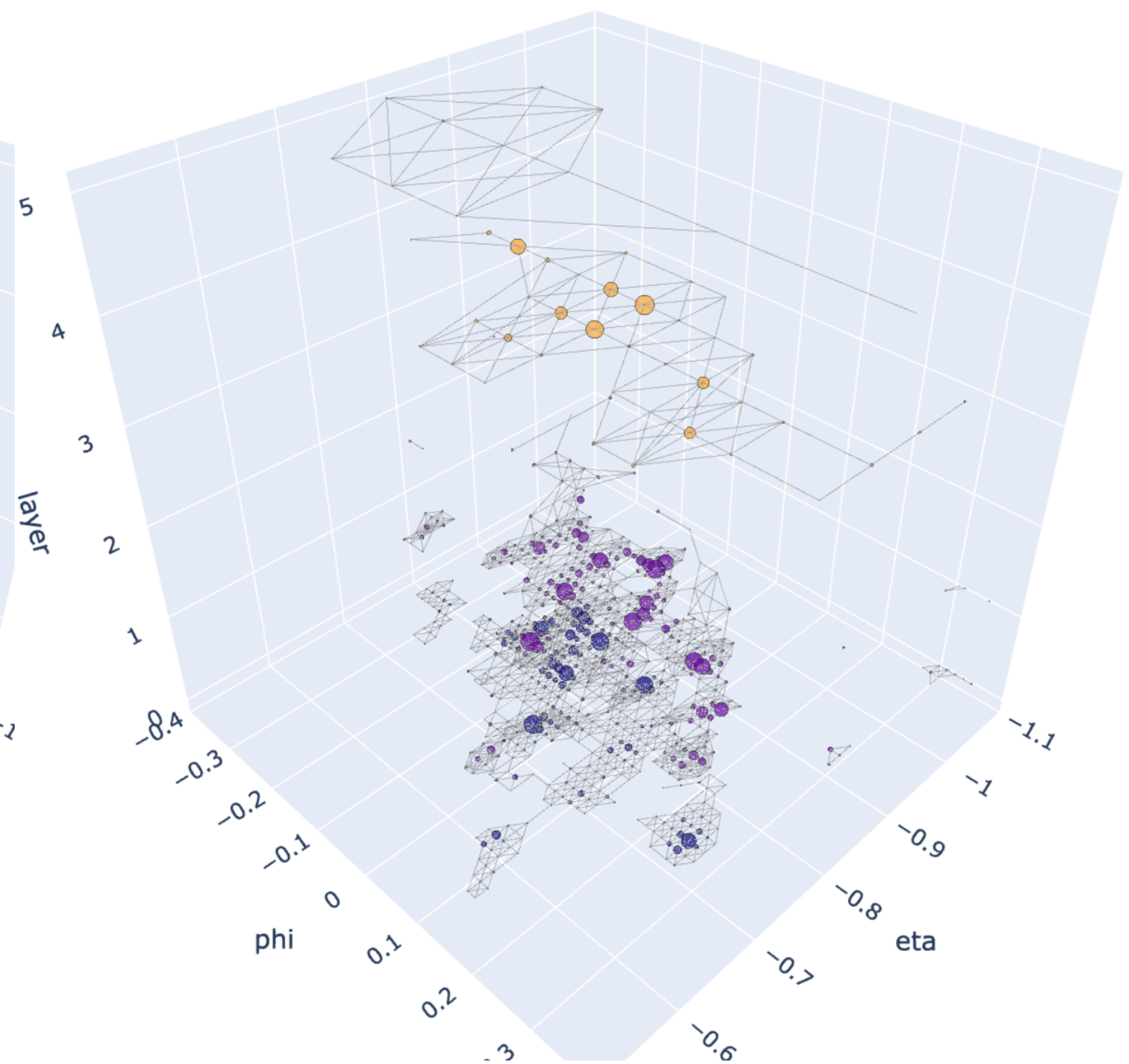
[2] Configurable calorimeter simulation for AI applications A. Charkin-Gorbulin et al, Mach. Learn.: Sci. Tech. (2023)

Example of a positive jet pair

random seed (a)

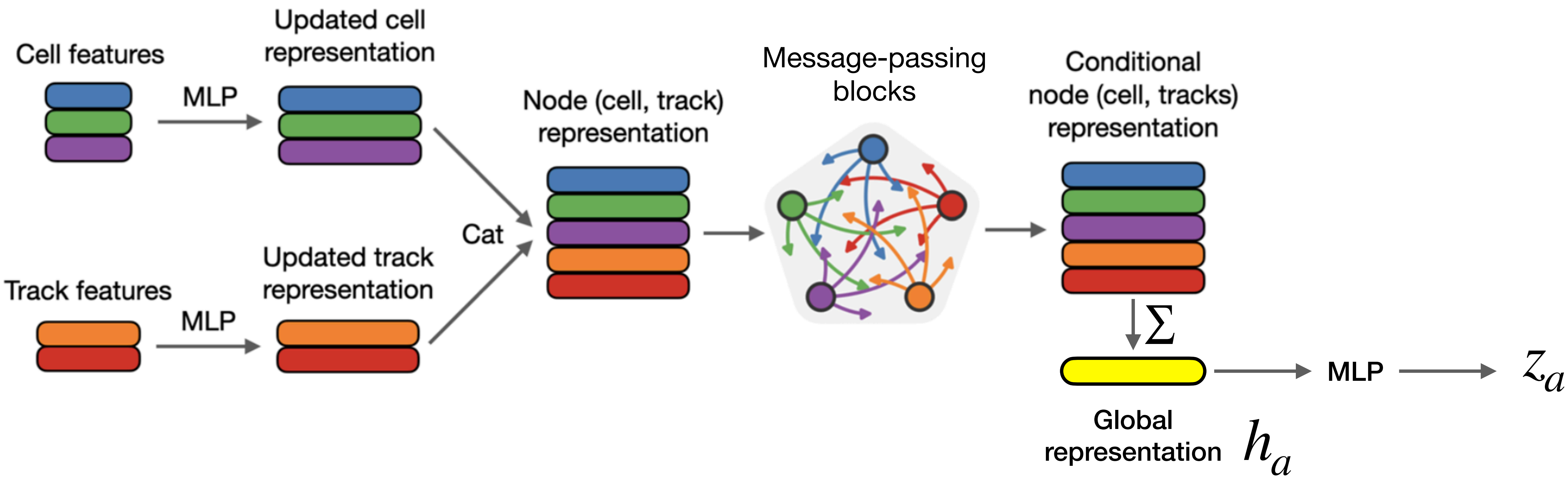


random seed (b)

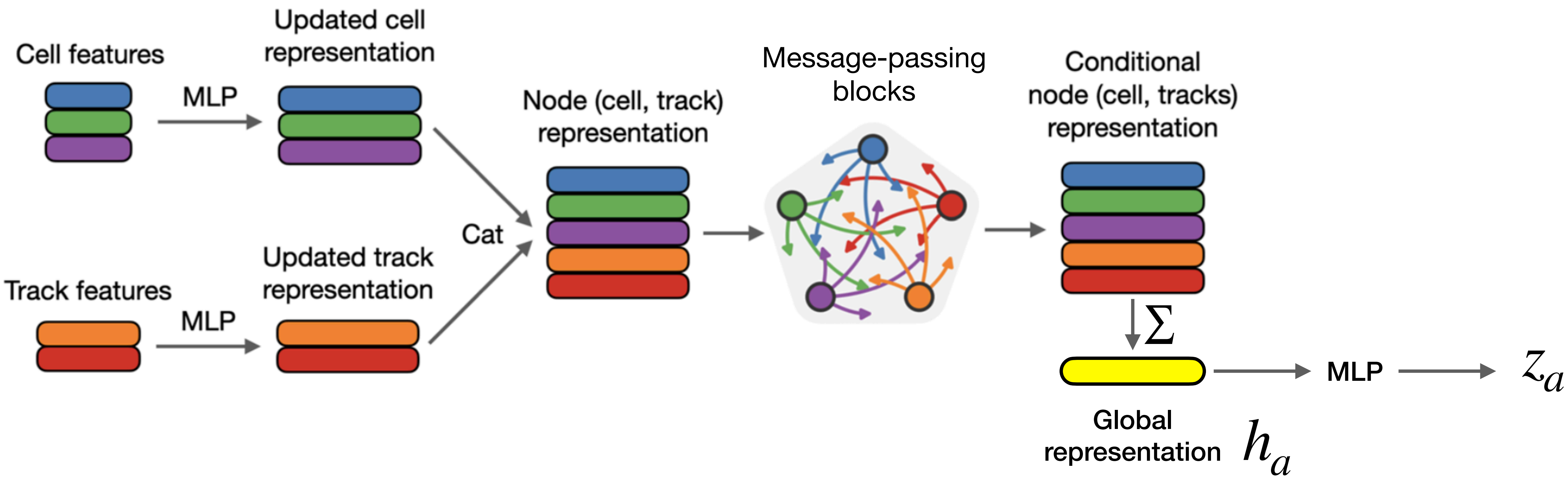


*inter-layer and track-cell edges not shown

Graph encoder "backbone" model



Graph encoder "backbone" model

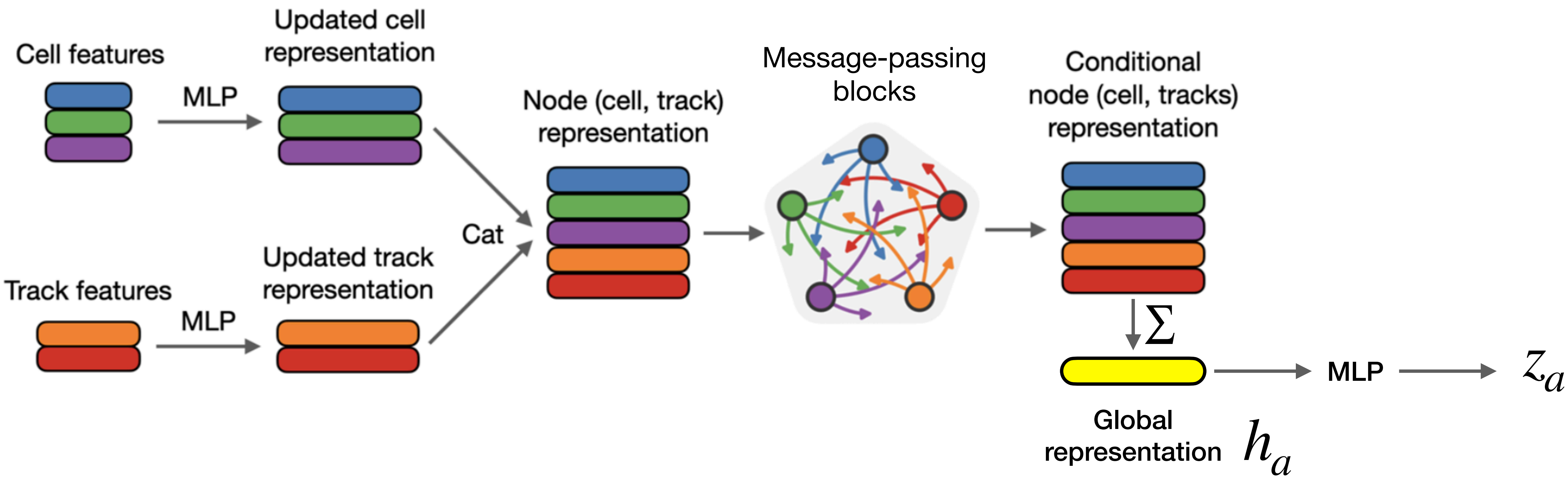


1. During SSL training, use "NT-Xent" loss [3] for each batch of $N = 300$ jet pairs:

$$L(z_a, z_b) = -\log \frac{\exp(\hat{z}_a \cdot \hat{z}_b/\tau)}{\sum_{i\neq a}^{2N} \exp(\hat{z}_a \cdot \hat{z}_i/\tau)} \qquad \text{where} \qquad \hat{z}_a := z_a/|z_a| \implies \hat{z}_a \cdot \hat{z}_b = \cos(\theta_{ab})$$

[3] A Simple Framework for Contrastive Learning of Visual Representations. T. Chen, G. Hinton. et al. (2020)

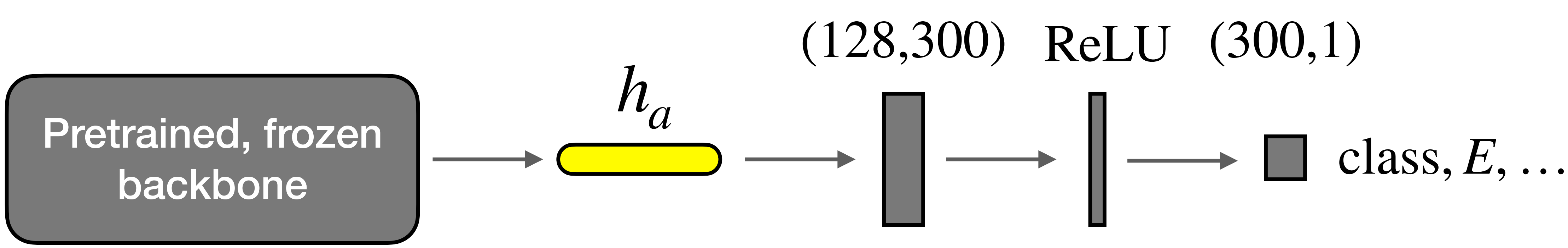
Graph encoder “backbone” model



1. During SSL training, use “NT-Xent” loss [3] for each batch of $N = 300$ jet pairs:

$$L(z_a, z_b) = -\log \frac{\exp(\hat{z}_a \cdot \hat{z}_b/\tau)}{\sum_{i\neq a}^{2N} \exp(\hat{z}_a \cdot \hat{z}_i/\tau)} \quad \text{where} \quad \hat{z}_a := z_a/|z_a| \implies \hat{z}_a \cdot \hat{z}_b = \cos(\theta_{ab})$$

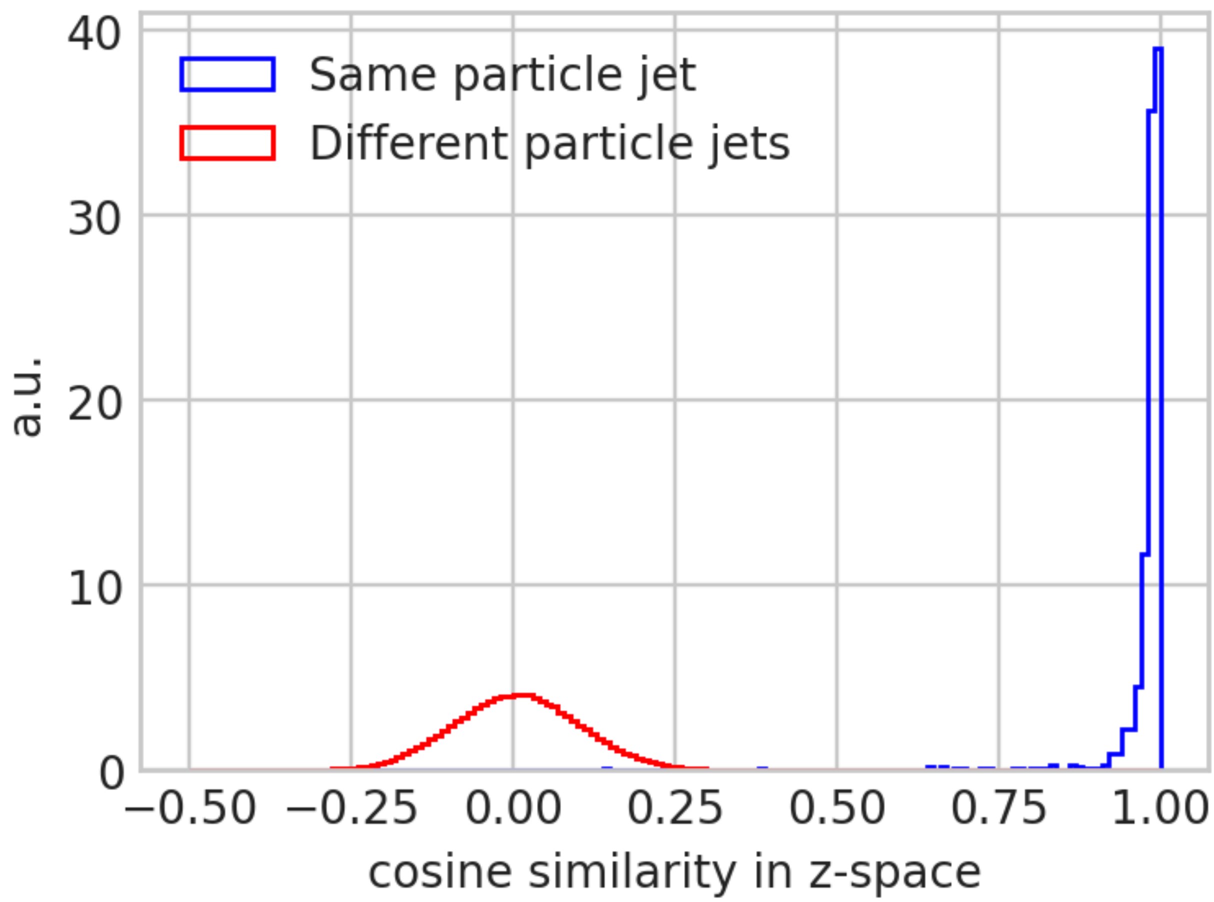
2. During downstream training, can freeze backbone and train single-layer perceptron



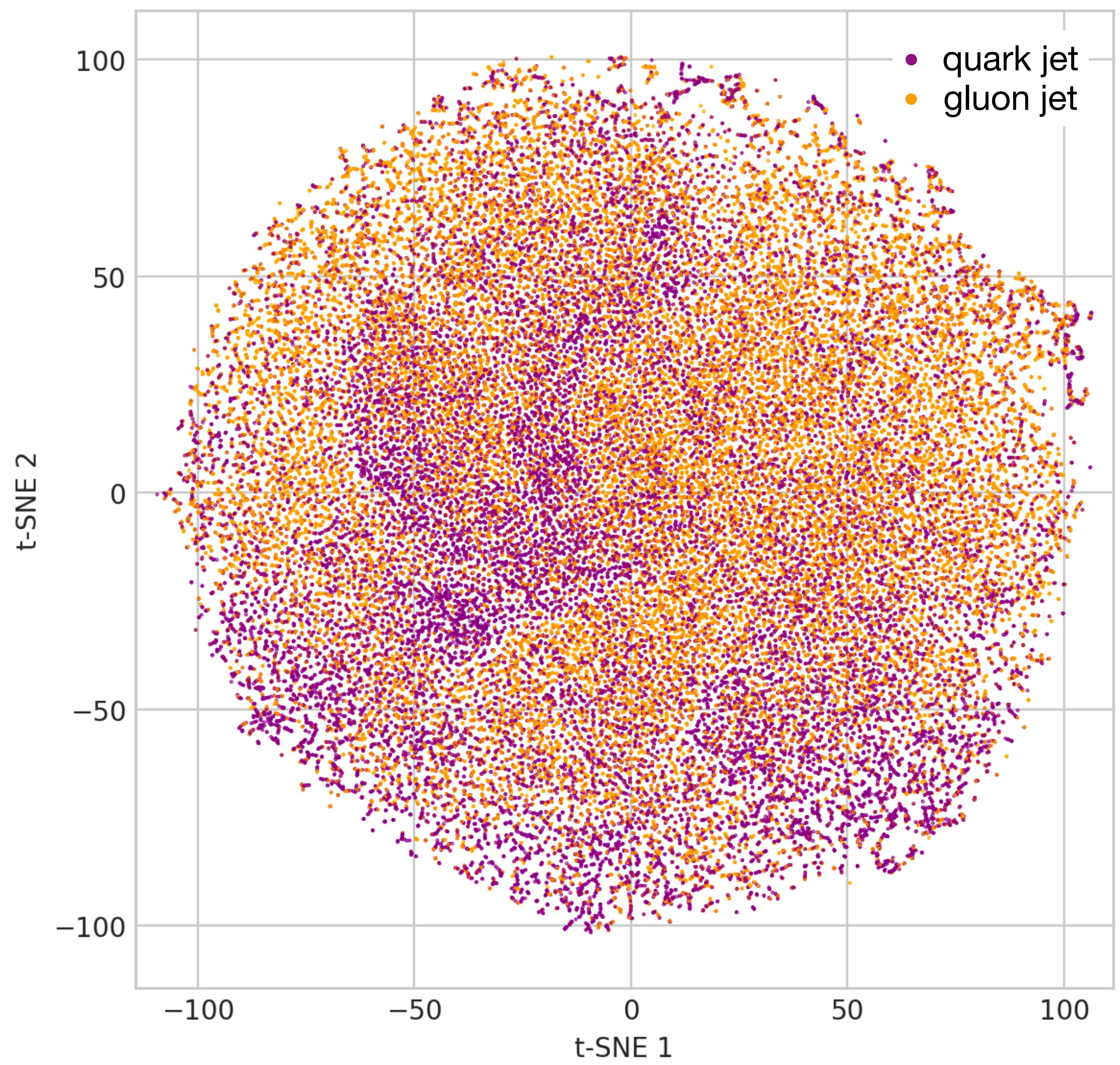
[3] A Simple Framework for Contrastive Learning of Visual Representations. T. Chen, G. Hinton. et al. (2020)

Probing the z contrastive space

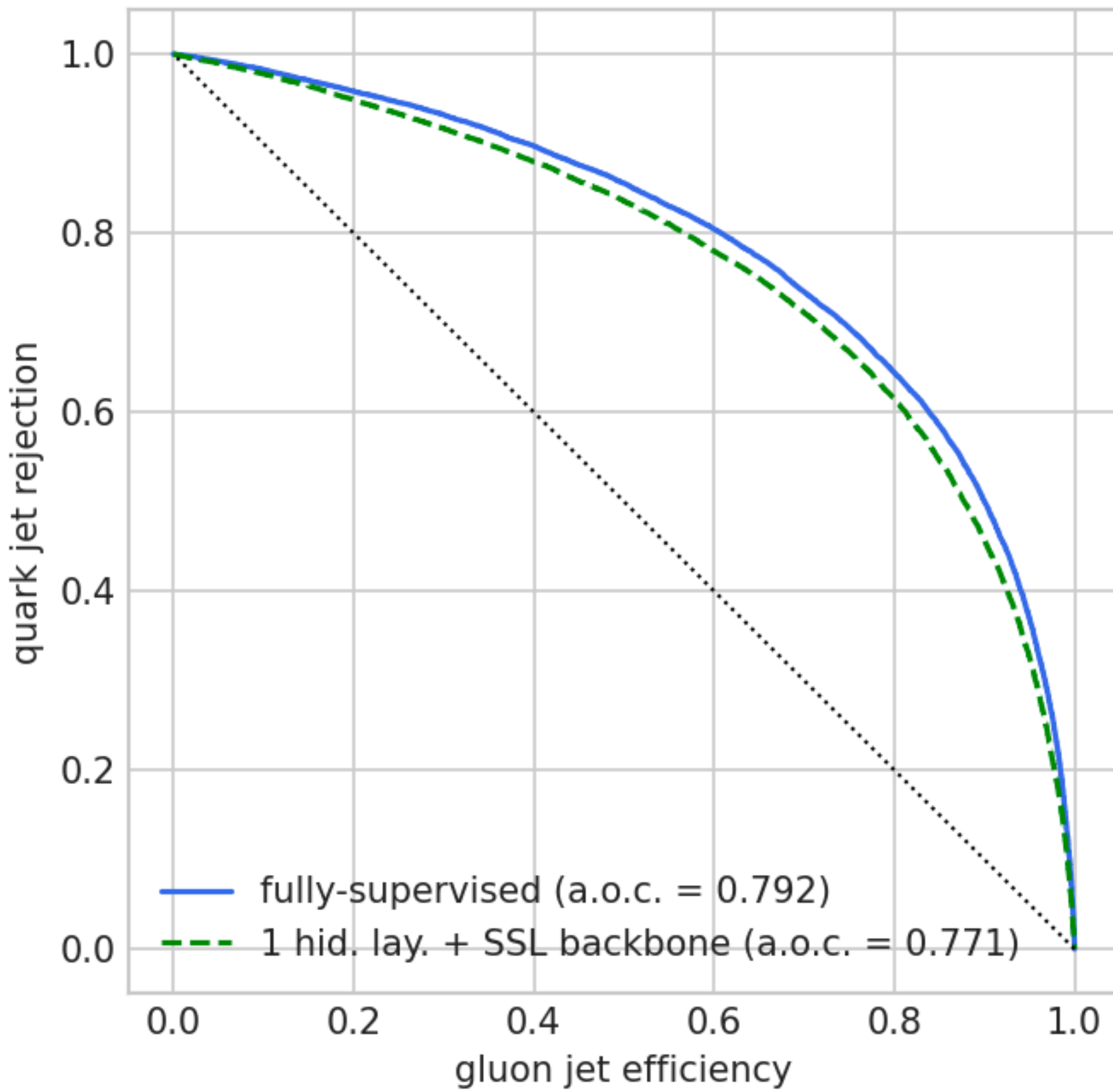
Contrast between jet pairs



t-SNE decomposition



Downstream task 1: q/g tagging

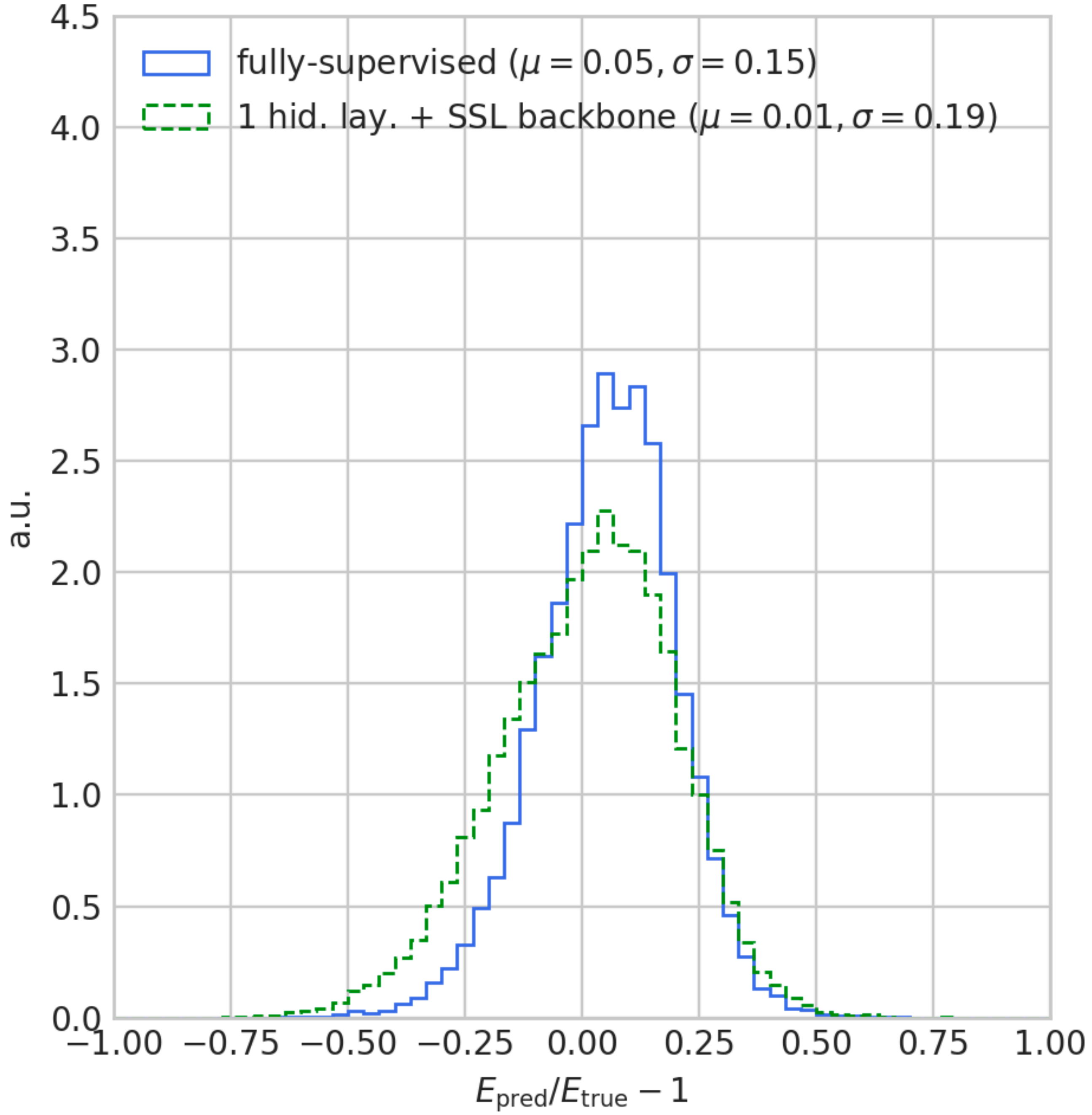


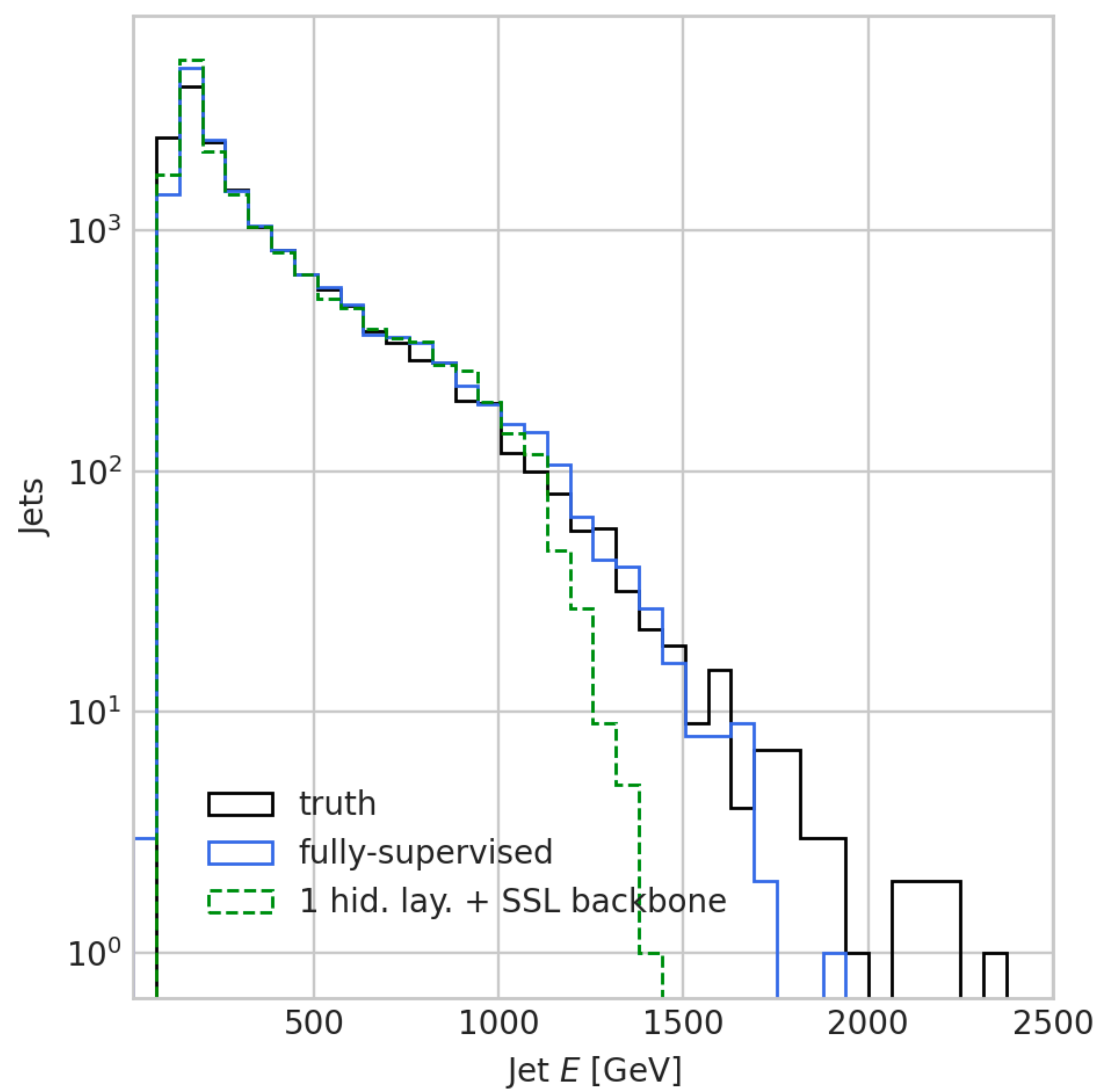
337k trainable parameters

vs.

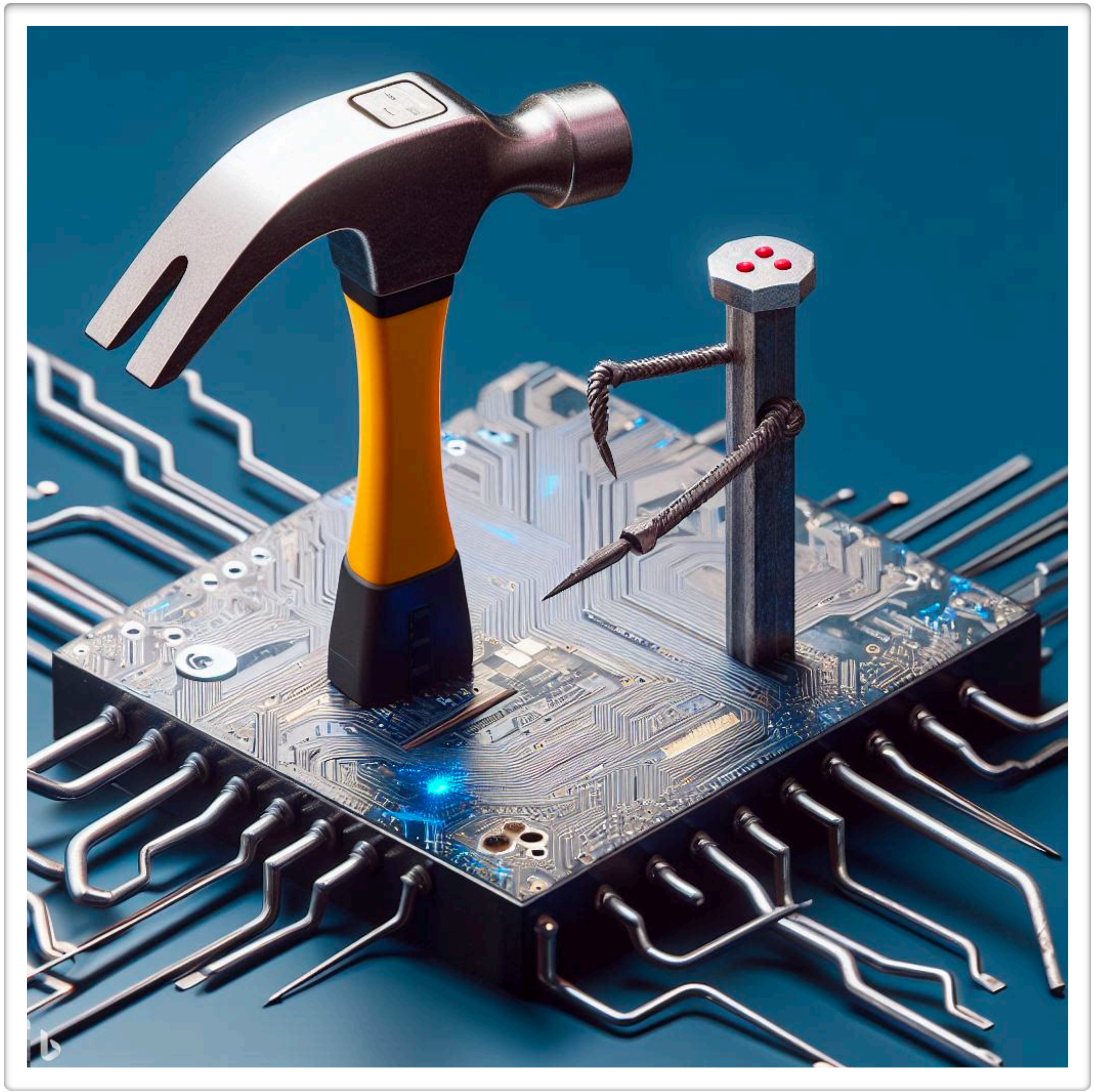
39k *trainable* parameters

Downstream task 2: jet E reg.





Thank you!



a hammer and a nail talking to each other using artificial intelligence