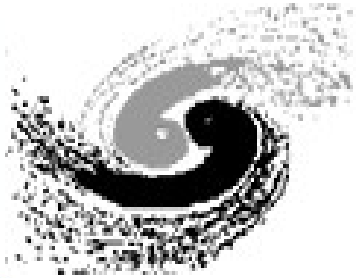
Spring 2026 IHEP ML Workshop
Multidisciplinary Building 122, IHEP, Beijing

Foundation Model at BESIII --- CLEAR (Contrastive Learning for Event Attention-based Representation)

Jingde Chen, Zijie Shang, Tong Liu, Ke Li

Institute of High Energy Physics, Chinese Academy of Sciences,
Experimental Physics Center, China



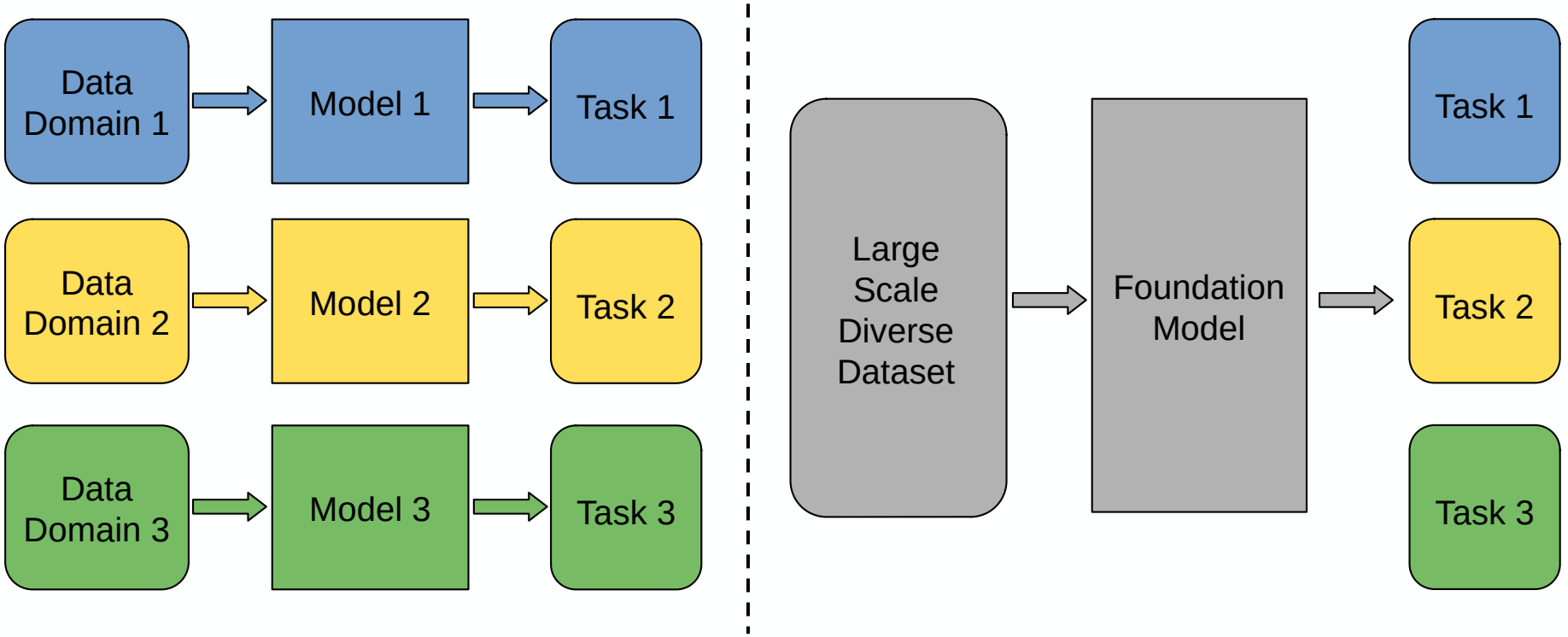
14th–16th April, 2026

Part 1: Introduction

Introduction

What is a Foundation Model:

Any model that is trained on broad data (generally using **self-supervision** at scale) that can be adapted (e.g., **fine-tuned**) to a wide range of **downstream tasks**.



Why Foundation Model:

- Traditionally, machine learning models are trained separately for **each specific task**.
- A Foundation Model is pre-trained once and serves as a shared foundation for **many downstream tasks**.

Introduction

Downstream Tasks in Foundation Model

★ Foundation Model ≈ Pretext Tasks + Loss Functions ★

Contrastive Learning | Adversarial Learning | Masked Modeling

Discriminative Tasks

Models are trained to classify event topologies and to distinguish between underlying physical processes using collider data.

(1). Event classification

(2). Particle identification and tagging

(3). Signal-background discrimination

(4). Anomaly detection

Generative Tasks

Models are trained to learn underlying data distributions and to generate realistic detector responses and physics events.

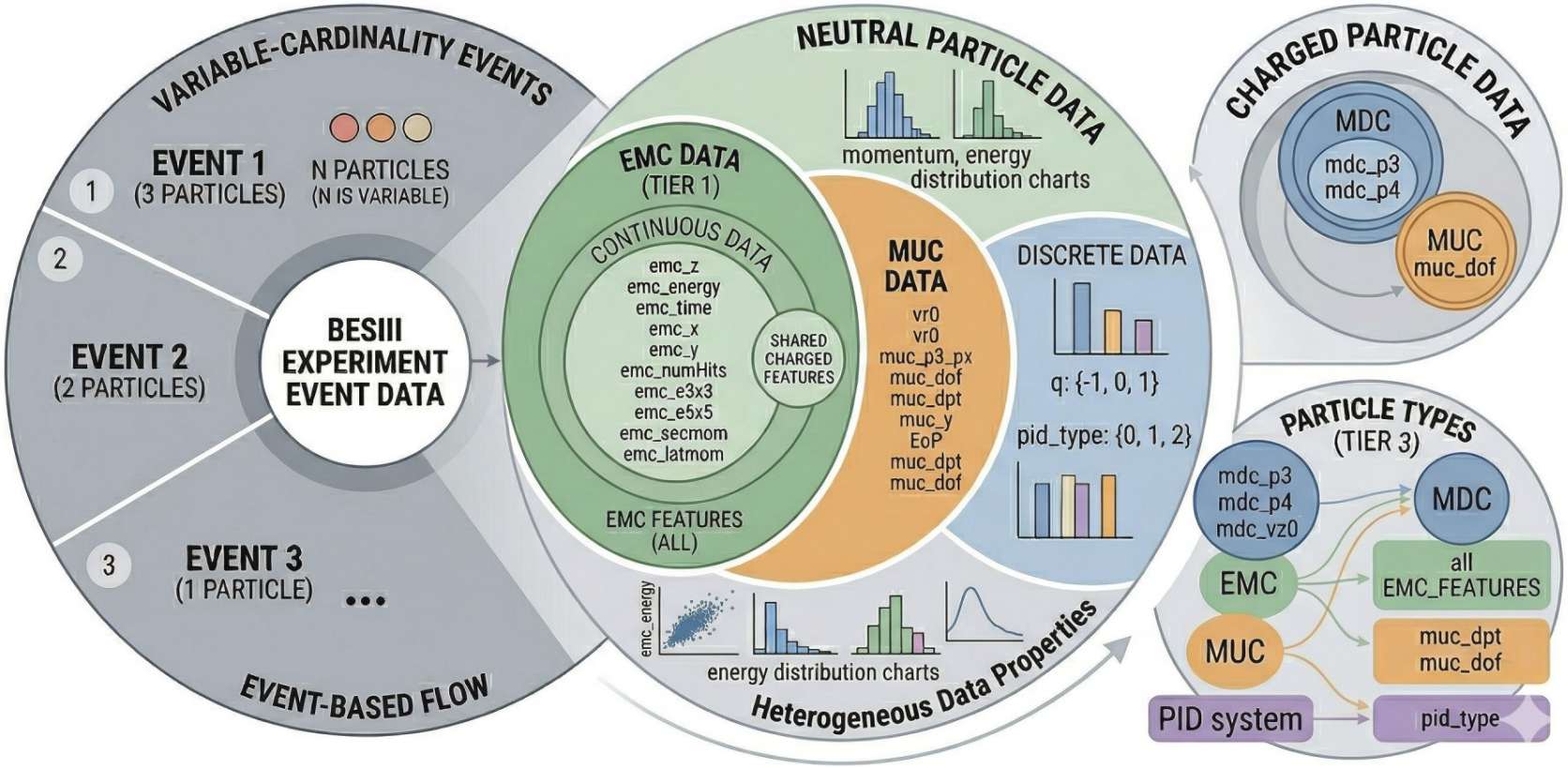
(1). Fast detector simulation

(2). Calorimeter shower generation

(3). Decay event reconstruction

Introduction

Intrinsic Heterogeneity in BESIII Data



- **Variable-cardinality** particle sets per event.
- **Inconsistent feature dimensionality** across charged and neutral particle types.
- Modality gaps across **sub-detectors (EMC, MUC, MDC)**.
- Coexistence of **discrete and continuous** data.

Figure is from Nano Banana 2.

Part 2: Research Method

Research Method

Training Data:

◎ Decay event: **J/ψ** decay (Inclusive MC data)

◎ Event classes: **More than 1000 decay classes** in J/ψ decay. **Only the top 100** most frequent **classes** are **used for model training**.

◎ Training dataset: 100 classes of J/ψ decay events (= **3M samples**; expected to scale to over **100M in future versions**)

◎ Input features:

⇒ **Kinematic**:

['q','mdc_p3_px', 'mdc_p3_py', 'mdc_p3_pz', 'vz0', 'vr0']

⇒ **Particle identification information**:

['pid_prob_type']

⇒ **Electromagnetic calorimeter** :

['emc_numHits', 'emc_e3x3', 'emc_e5x5', 'emc_energy', 'emc_x', 'emc_y', 'emc_z', 'emc_time', 'EoP', 'emc_secmom', 'emc_latmom']

⇒ **Muon Counter** :

['muc_dpt', 'muc_dof']

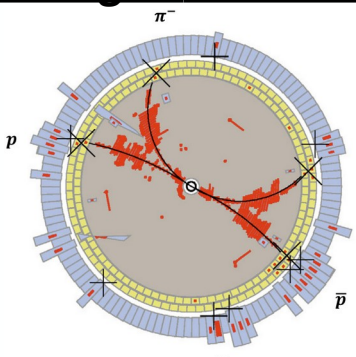
Downstream Tasks:

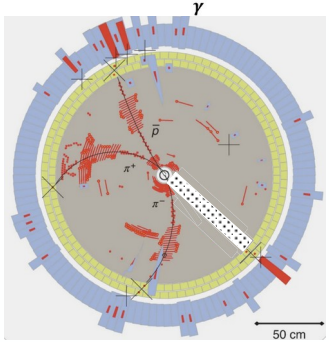
◉ Task 1 – Generative Tasks ⇒ Decay event **reconstruction**.

◉ Task 2 – Discriminative Tasks ⇒ **100 classes classification task.**

Research Method

Masking + Contrastive Learning:





Encoder

Embedding

Projection head

Contrastive Loss

Encoder

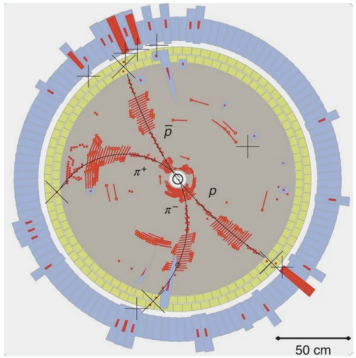
Embedding

Projection head

Reconstruction Loss

Decoder

Reconstruction head



◉ Contrastive Learning

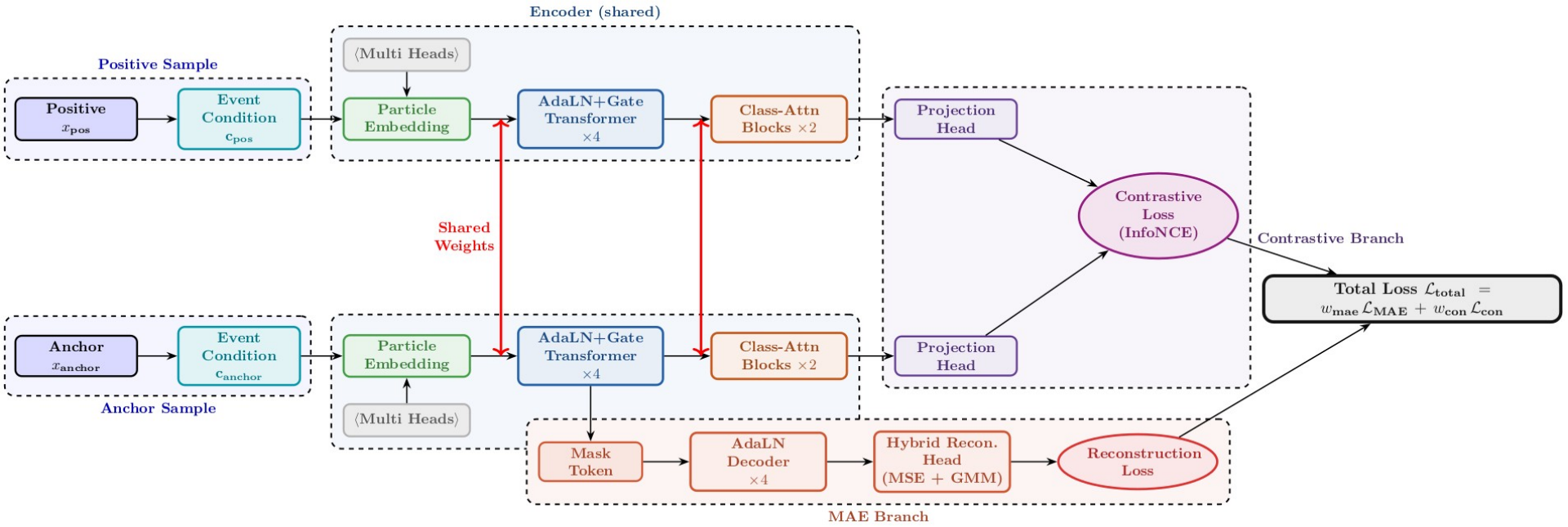
⇒ Brings representations of **same decay types closer**, and **push apart of different classes**.

◉ Masking Learning

⇒ Captures **inter-particle correlations** within decay events.

Research Method

CLEAR Structure:



Condition LayerNorm:

- Condition 1: Number of the total charged tracks.
- Condition 2: Number of the total neutral tracks.
- Condition 3: Number of the charged tracks without EMC hit.
- Condition 4: Number of the neutral tracks without secmom and latmom.

Hybrid Reconstruction Loss:

- MSE Loss: continuous features with **unimodal distributions.**
- GMM NLL Loss: features with bimodal / **U-shaped distributions.**

Part 3: Research Results

Research Results

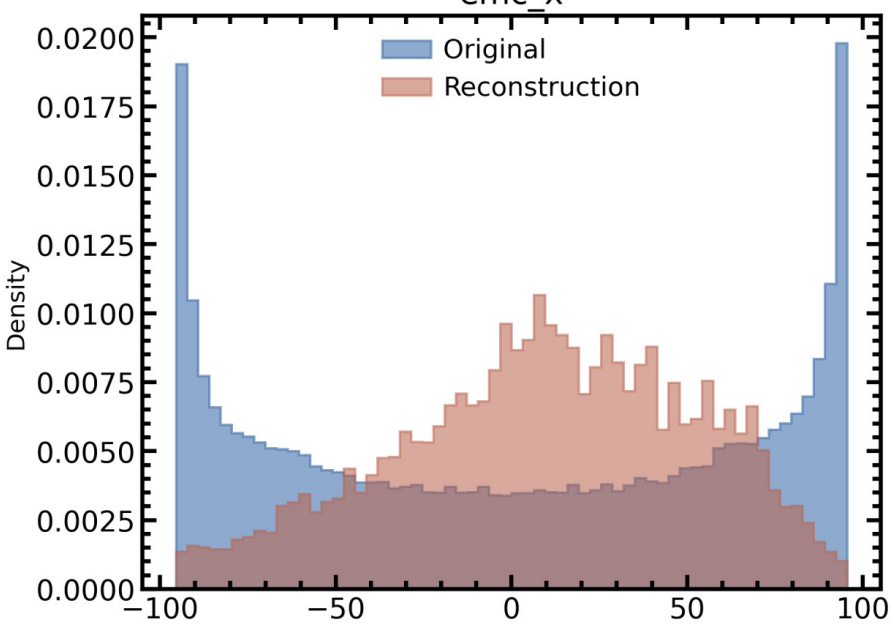
GMM for U-shape Distribution:

◎ **emc_x** exhibit **U-shaped distributions** due to the detector geometry — standard MSE or single-Gaussian reconstruction fails to capture the two separated modes.

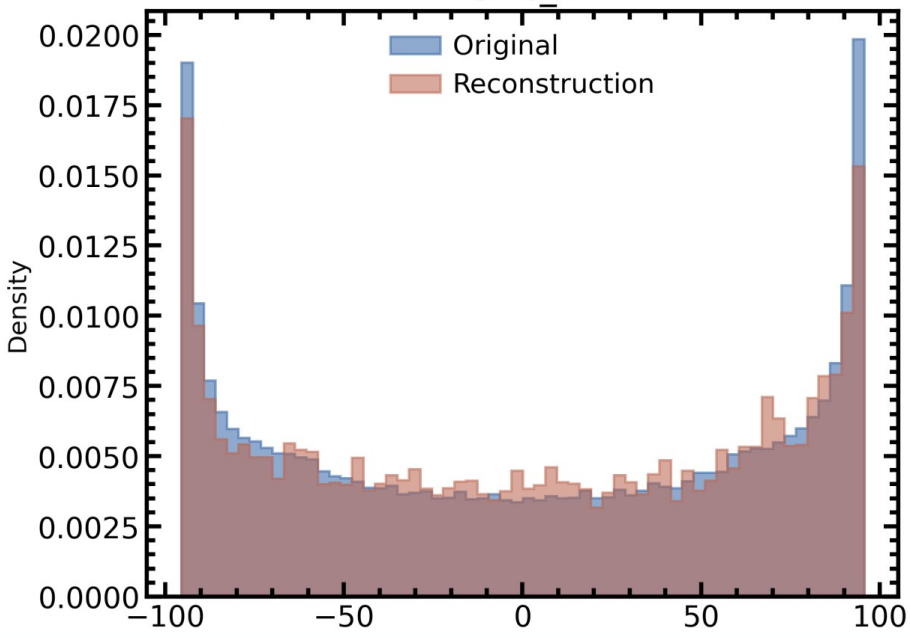
◎ A **GMM** head outputs mixture optimized **via negative log-likelihood (NNL)**.

◎ Each mode is captured by a dedicated Gaussian component.

Before



After



Research Results

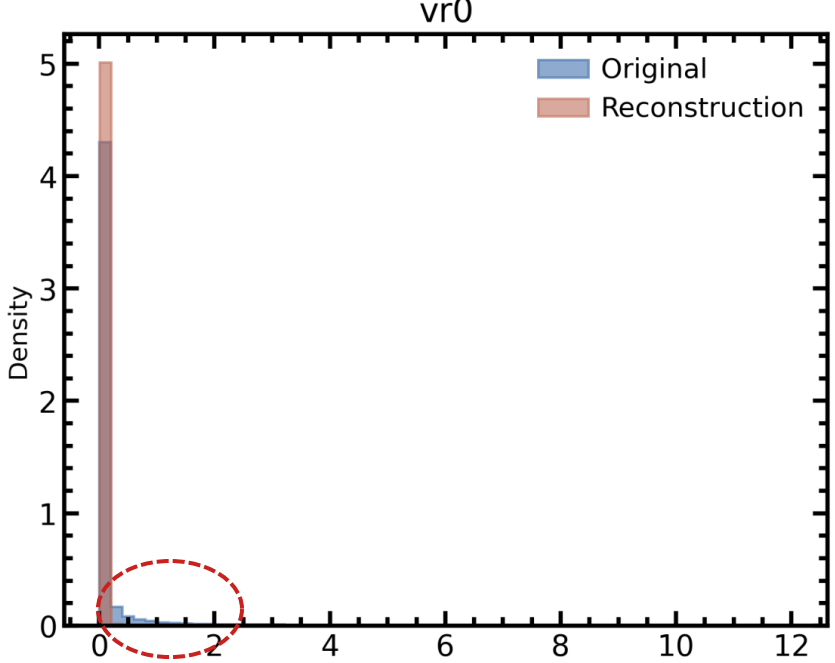
Zero-inflated log-normal:

◎ Features like **vr0** contain **a large spike at zero** plus **a heavy right tail** — a single Gaussian cannot model both components simultaneously

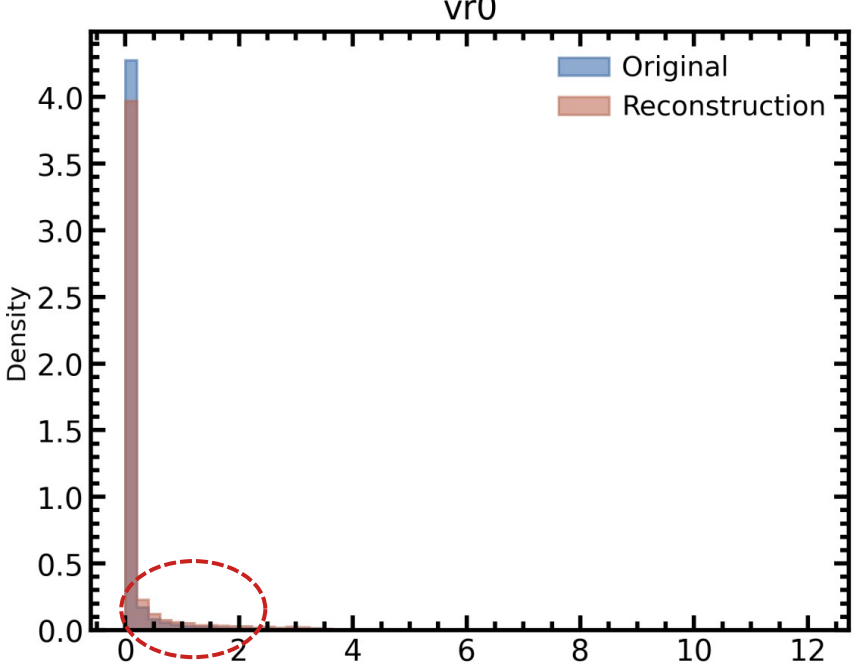
◎ **Zero-inflated modeling** refers to first estimating the **probability P(zero)** of observing a zero value, then **modeling the non-zero component separately**.

◎ **Zero** and **non-zero** regimes are **decoupled cleanly**.

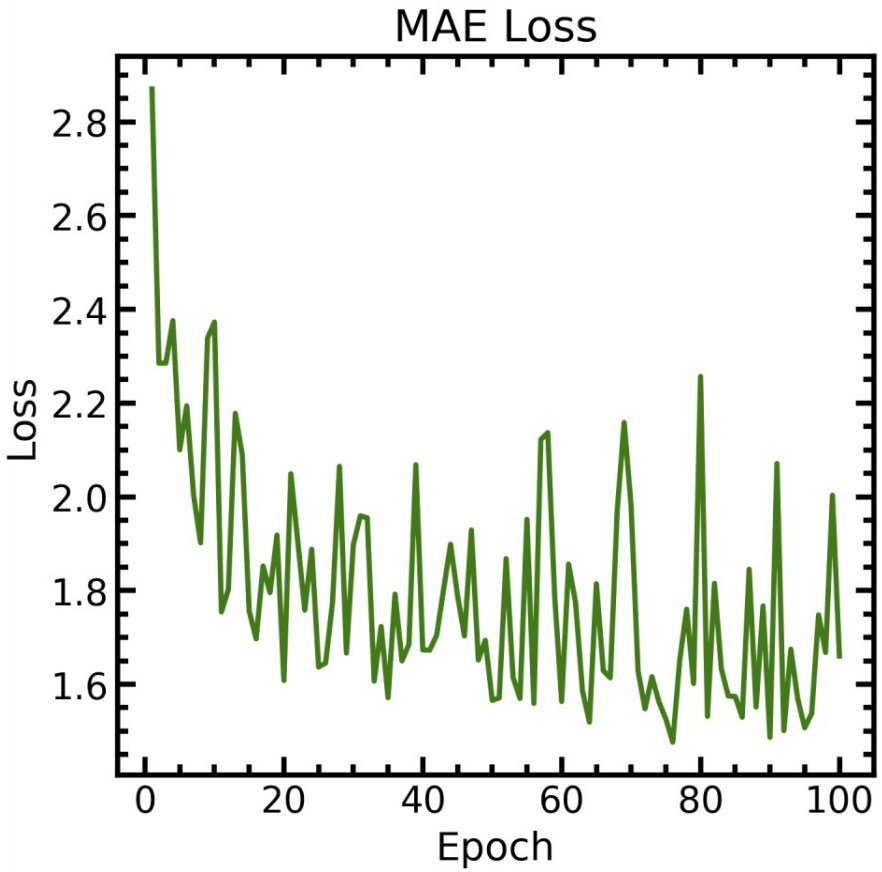
Before

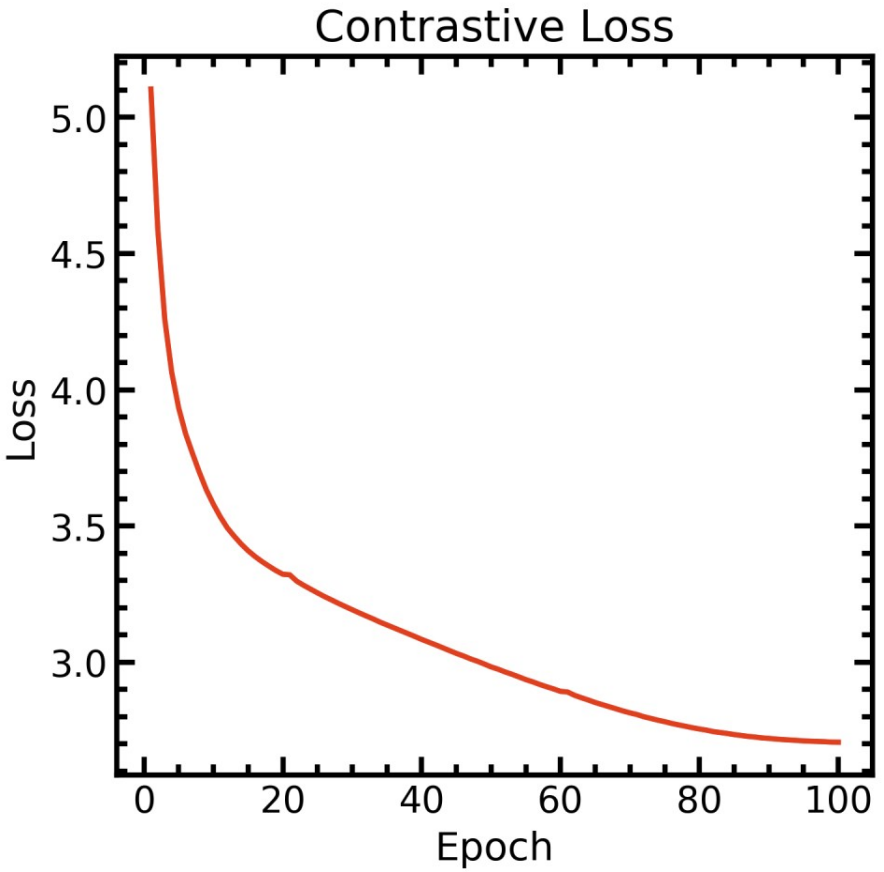


After



Research Results





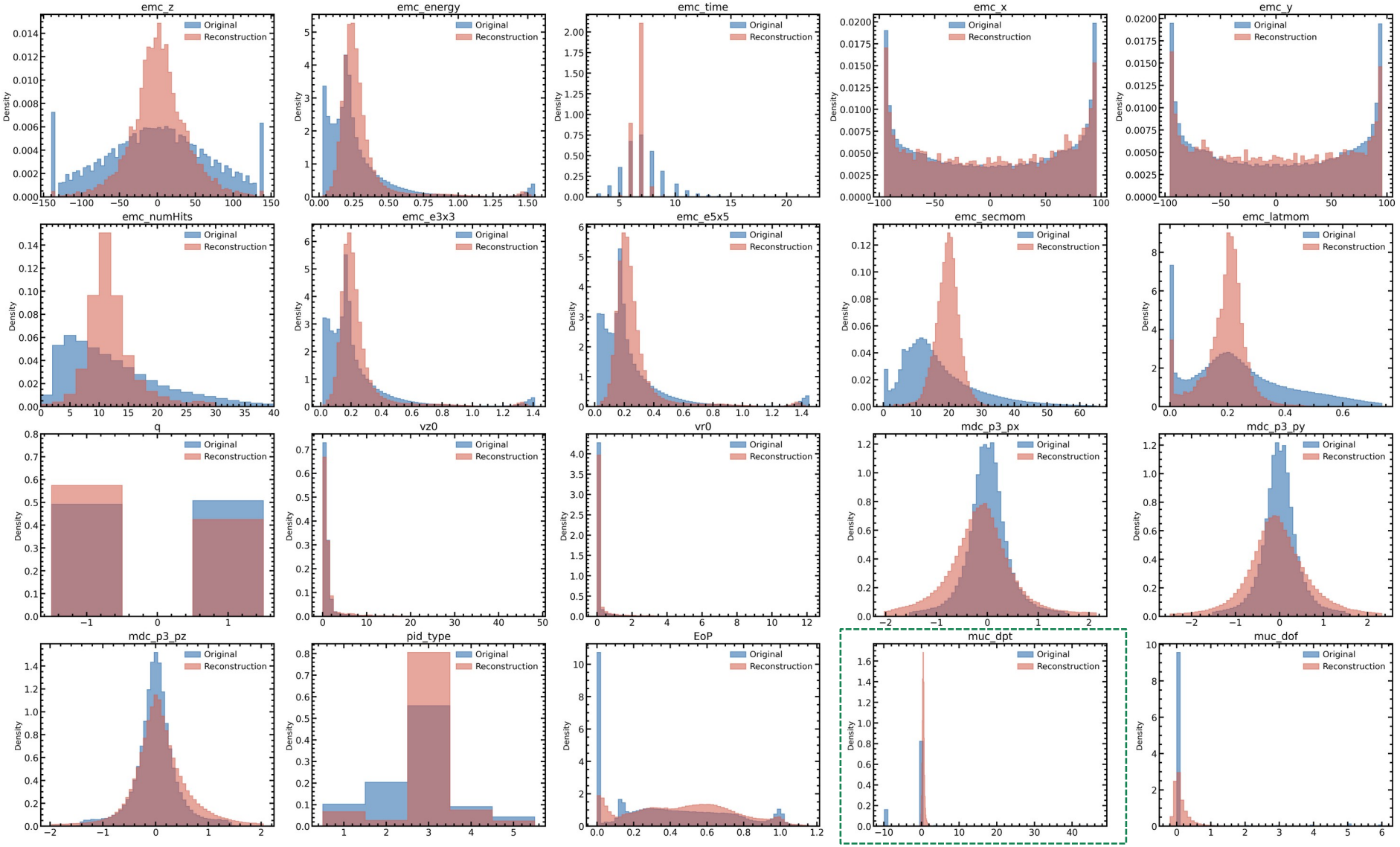
Training Results:

■ MAE Loss (Reconstruction Loss): Reconstruction loss decreases overall but exhibits **substantial instability** during training.

■ Contrastive Loss: Contrastive loss performs well, **benefiting from the supervised contrastive learning**. Future plans include exploring self-supervised contrastive way.

Research Results

Reconstruction Results:

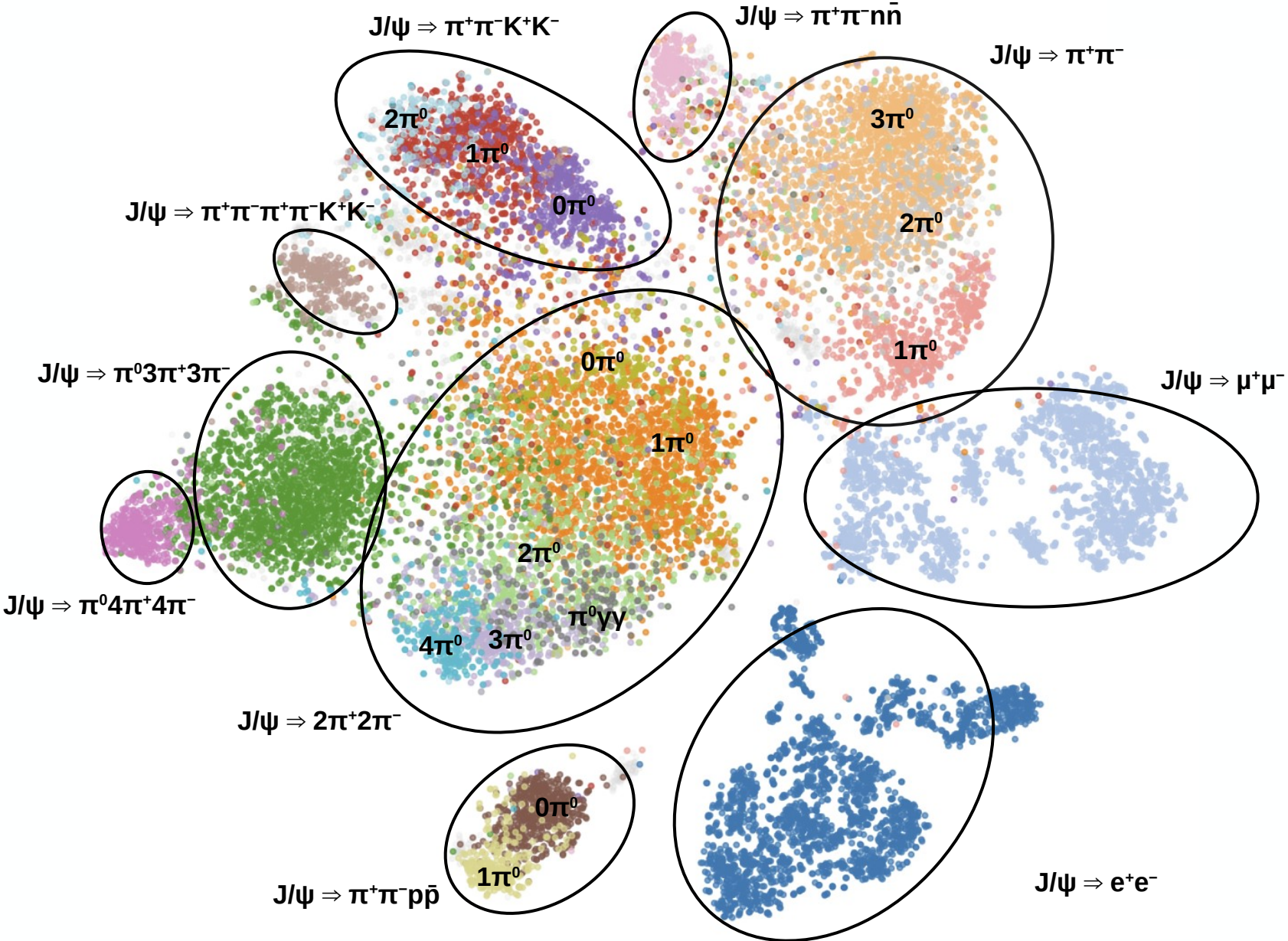


Research Results

Clustering Results:

Research Results

Clustering Results:



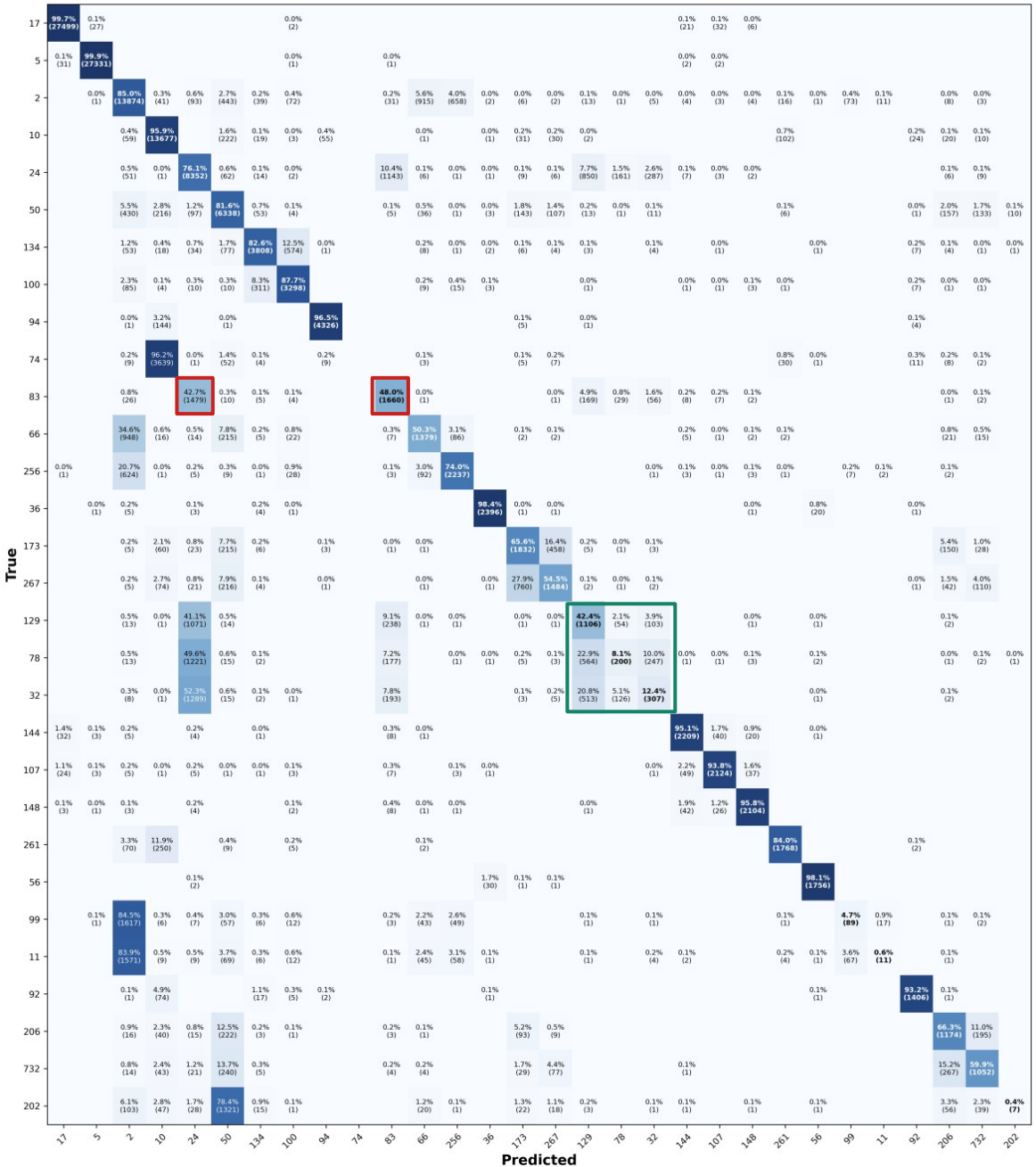
Research Results

Fine-tuning strategies:

- Strategy A: **Frozen encoder**
  ⇒ k-NN / 1L MLP (Logistic) / 2L MLP / 3L MLP (baseline)
- Strategy B: **CA-Only Fine-Tuning**
  ⇒ Freeze SA encoder; **unfreeze only Class Attention.**
- Strategy C: **End-to-End Fine-Tuning** with differential learning rate (LR).
- Strategy D: **Multitask Fine-Tuning**
  ⇒ Joint training with classification.
- Strategy E: **Two-Stage Fine-Tuning**
  ⇒ Stage 1: train CA + Head only until convergence.
  ⇒ Stage 2: unfreeze full model with very low LR.

Strategy	Acc (100-class)
[A] k-NN	0.5433
[A] MLP-1L(Logistic)	0.5979
[A] MLP-2L	0.6114
[A] MLP-3L	0.6116
[B] CA-Only	0.6527
[C] E2E	0.6911
[D] Multi-Task	**0.6929**
[E] Two-Stage	0.6853

Research Results



Classification results:

- C83 ⇒ C24: (42.7%)
- C83 = $\pi^0\pi^0\pi^+\pi^-$
- C24 = $\pi^0\pi^0\pi^0\pi^+\pi^-$

- C24 = $\pi^0\pi^0\pi^0\pi^+\pi^-$
- C122 = $\pi^0\pi^0 h_1(1170)$, **h_1 (1170)** ⇒ $\pi^0\rho^0$, $\rho^0 \Rightarrow \pi^+\pi^-$
- C72 = $\pi^0\pi^0 h_1(1170)$, **h_1 (1170)** ⇒ $\pi^-\rho^+$, $\rho^+ \Rightarrow \pi^0\pi^+$
- C30 = $\pi^0\pi^0 h_1(1170)$, **h_1 (1170)** ⇒ $\pi^+\rho^-$, $\rho^- \Rightarrow \pi^0\pi^-$

★ **neutral track (π^0)** and **resonance state**.

Research Results

Comparison with Standard Approaches:

Table 1: Performance Comparison on Different Decay Channels

	Foundation Model			Traditional Method		
Decay Channel	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
$J/\psi \to \pi^+\rho^-, \rho^- \to \pi^0\pi^-$	**0.85**	**0.88**	0.87	0.72	0.74	0.73
$J/\psi \to \pi^-\rho^+, \rho^+ \to \pi^0\pi^+$	**0.83**	**0.90**	0.86	0.71	0.74	0.72
$J/\psi \to \pi^0\rho^0, \rho^0 \to \pi^+\pi^-$	**0.87**	**0.92**	0.90	0.82	0.82	0.82
$J/\psi \to \pi^+\pi^-\pi^0$	0.74	**0.76**	0.74	0.74	0.70	0.71

◎ Compared with traditional physics-based approaches, foundation models **generally achieve higher precision and recall rate**.

◎ However, traditional methods can attain precision above 99% by sacrificing recall—**a trade-off that current foundation models are unable to match**.

Part 4: Conclusions

Conclusions

1. We successfully developed a **foundation model** for high-energy physics collider data — the **CLEAR (Contrastive Learning for Event Attention-based Representation)** — built upon a transformer-based architecture. The model achieves promising performance on **generative tasks** and **discriminative tasks**.

2. For generative tasks in high-energy physics, more principled data processing strategies (e.g., **zero-inflated modeling** and **Gaussian mixture modeling**) can significantly **improve generation quality**.

3. For discriminative tasks, foundation models **exhibit strong potential to surpass traditional physics-based approaches**.

Future Work:

1. Explore self-supervised contrastive learning frameworks.

2. Incorporate physics-informed priors to improve classification performance.

3. Investigate more advanced data processing strategies.

Thank you for listening!
Any question?