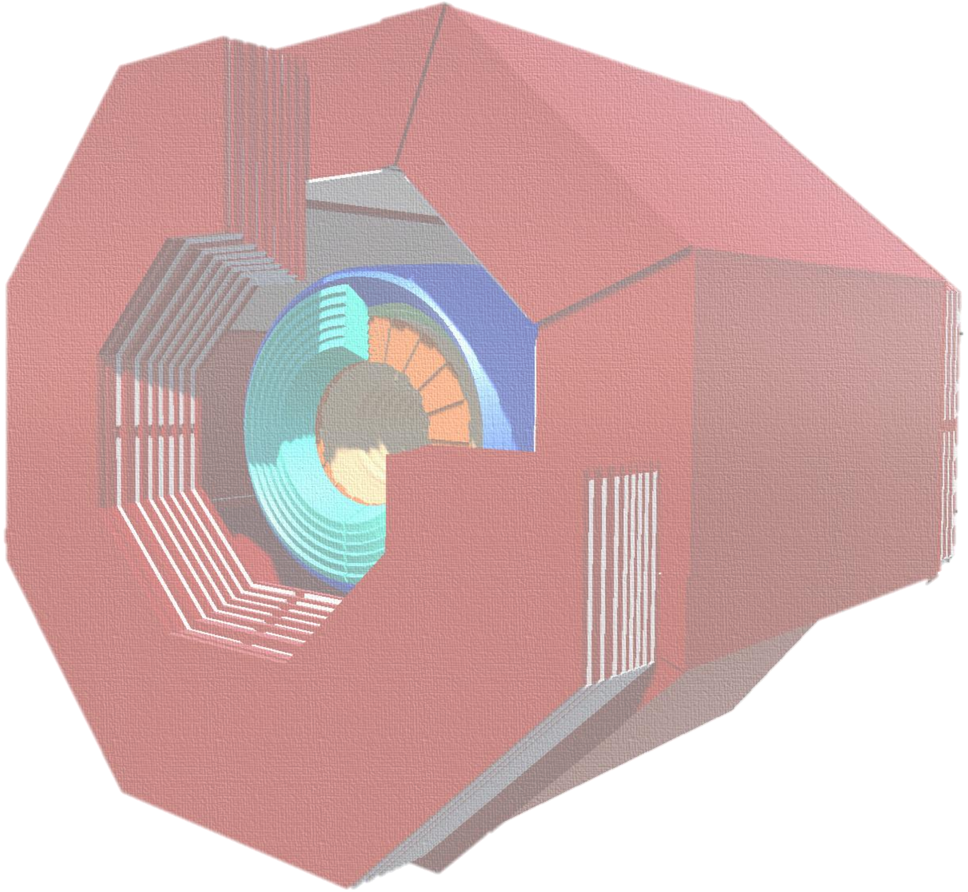
Learning transferable event representations for Λ_c^+ physics at BESIII

Kai-Xuan Huang(黄凯旋)
University of Chinese Academy of Sciences (UCAS)

Quantum Computing and Machine Learning Workshop 2026, Shenzhen
2026/08/19

Outline

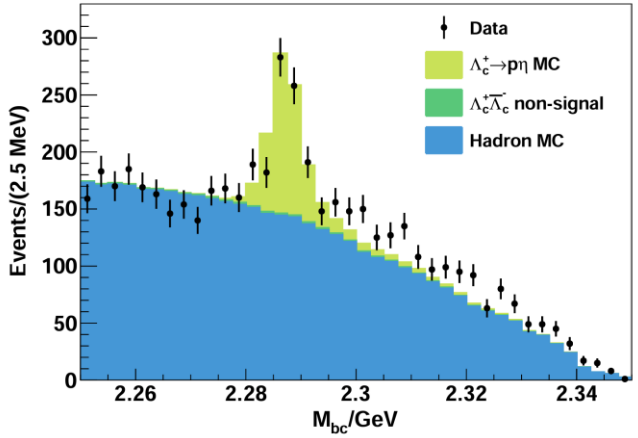


- Motivation
- Pre-training methods
  - Framework
  - Datasets
- Performance
  - $\Lambda_c^+\bar{\Lambda}_c^-$ event classification model
  - $\Lambda_c^+\bar{\Lambda}_c^-$ momentum-direction regression model
- Software package developed under the BOSS framework
- Summary

Motivation

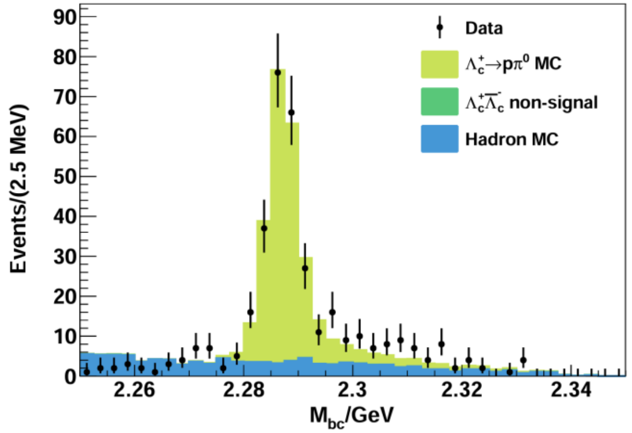
- Deep learning has excellent performance
  - In analyses: $D^+ \to \gamma e^+ \nu_e$, $\Lambda_c^+ \to p\eta'$, $\Lambda_c^+ \to p\pi^0 \ldots$
  - based on *Particle Transformer*[1]
  - **Training from scratch** (Training a model only for a decay channel)
  - Main 2 classification (signal, backgrounds events)

- Several limitations (**Training from scratch**)
  - Large training events
    - $> 10^5$ events ~ effective model
    - Insufficient events ~ degrade performance
  - Duplicated development
    - Each decay channel ~ a model
    - For Λ_c^+ physics, many channels ~ duplicated development
  - Expertise and GPUs resource demands
    - The *Weaver* toolkit, related training pipelines
    - Rely on GPU resources



$\Lambda_c^+ \to p\eta$, before score veto

Deep learning



$\Lambda_c^+ \to p\eta$, after score veto[2]

[1]: arXiv:2202.03772 [2]: Phys. Rev. D 111, L051101

Pre-training Method

- For Λ_c^+ physics, decay channels share:
  - Common detector environments
  - Similar event topologies
  - Comparable background compositions

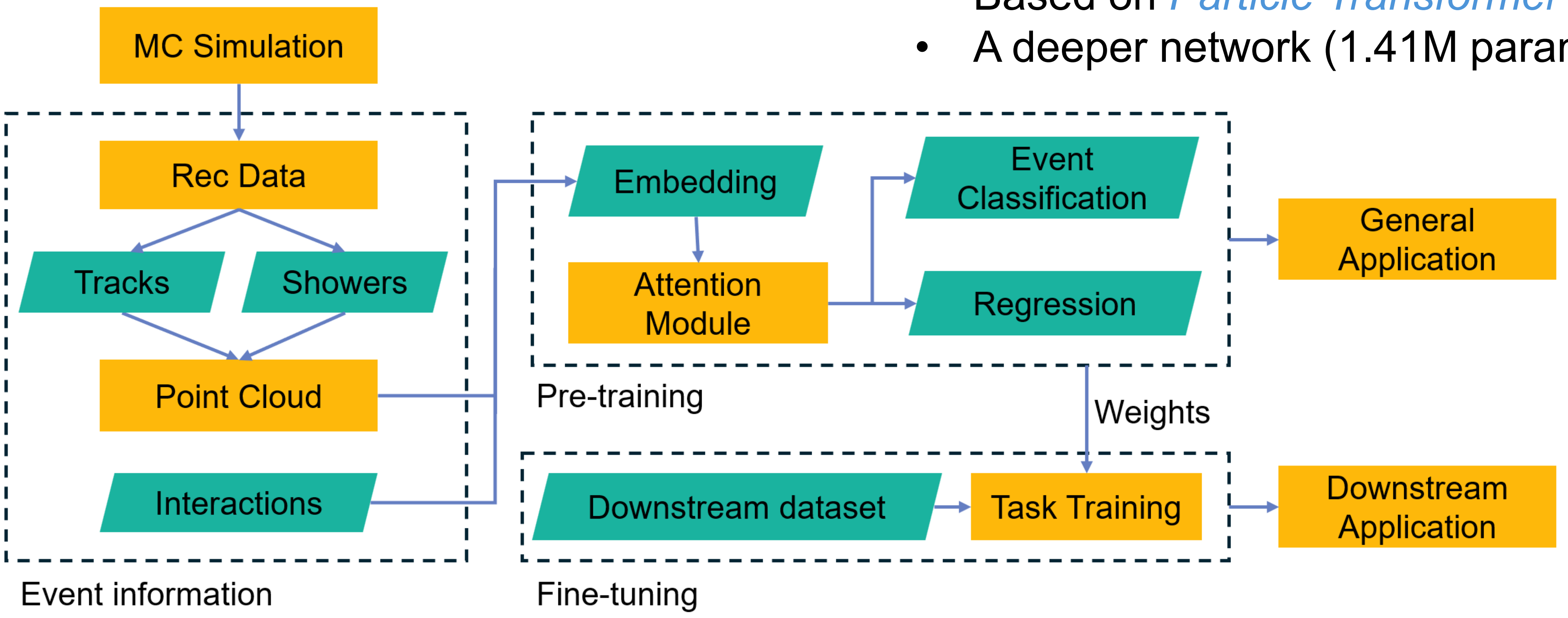
- General models **(Pre-training)**
  - Learn general representations from larger datasets
  - General application
    - Different Λ_c^+ decay channel
    - Online event classification during data taking
  - Adapted to downstream task **(Fine-tuning)**
    - Always better performance

- **Pre-training** applications
  - GPT (LLM) [1], ESM (protein sequences, biology) [2]
  - CMS Jet classification[3], ATLAS event classification[4]

[1]: *arXiv:2601.03267*
[2]: *Science 379.6637*
[3]: *arXiv:2408.09343*
[4]: *arXiv: 2412.10665*

Pre-training Models - Framework

- **Pre-training** model architecture
  - Based on *Particle Transformer*
  - A deeper network (1.41M parameters)



- Event input representation
  - Tracks, showers
  - Low-level detector response
  - Pairwise interaction

- **Fine-tuning**
  - Initialization with the pre-trained model weights

Pre-training Datasets

- **Pre-training** datasets
  - Event classification
    - 7.1×10^7 inclusive MC events
    - $\Lambda_c^+\bar{\Lambda}_c^-: D\bar{D}h: q\bar{q} = 1:1:1$
  - Momentum-direction regression
    - 2.4×10^7 $\Lambda_c^+\bar{\Lambda}_c^-$ inclusive MC events
- **Fine-tuning** datasets
  - 12 tagged decay channels of Λ_c^+
    - $7.2 \times 10^4 - 6.9 \times 10^7$ for different channel
  - $\Lambda_c^+ \to pKe\nu_e$ (only for regression)
    - 8.5×10^4 events
- Energy point
  - $4.600 - 4.700$ GeV

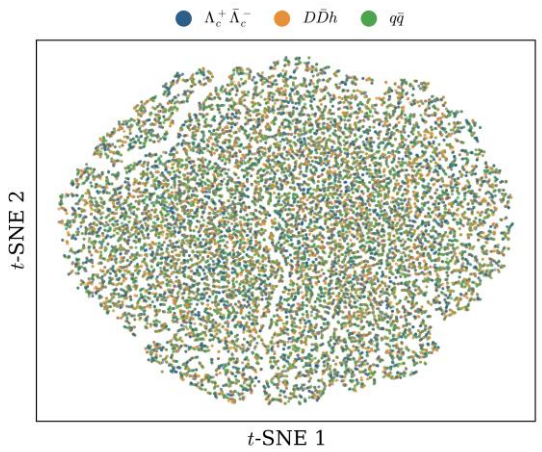
- **Event input features**

Features for charged tracks
• Polar and azimuth angle θ & ϕ
• Charge
• Momentum
• Helix parameters $d_\rho, d_z, \kappa, \phi_0, \lambda$
• $\chi_{dE/dx}$ in $e/\pi/K/p$ hypothesis
• χ_{TOF} in $\pi/K/p$ hypothesis

Features for showers
• Polar and azimuth angle θ & ϕ
• Deposited energy
• Hit number
• Time
• A_{20} & A_{42} moment
• Secondary & Lateral moment
• $E_{3\times3}$ & $E_{5\times5}$

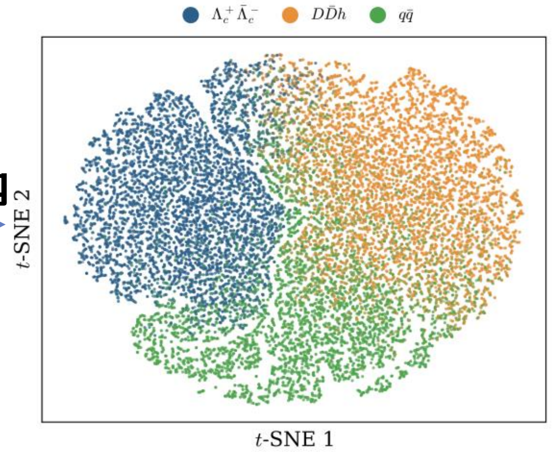
$\Lambda_c^+\bar{\Lambda}_c^-$ event classification model

- 3 event classification
  - $\Lambda_c^+\bar{\Lambda}_c^-$
  - $D\bar{D}h$ $(D^0\bar{D}^0, D^+D^-, D_s^+D_s^- \ldots)$
  - $q\bar{q}$
- Pre-training model
  - 6 * NVIDIA V100 GPUs ~ 4 Days
- For inclusive MC events
  - $\Lambda_c^+\bar{\Lambda}_c^-$,
    - 90.0% signal efficiency
    - Reject 97.0% background events
  - $D\bar{D}h$,
    - 90.0% signal efficiency
    - Reject 89.2% background events

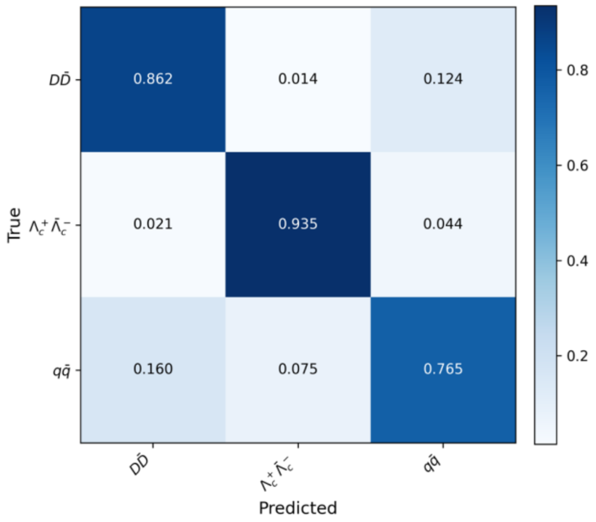


Visualization with **first** epoch

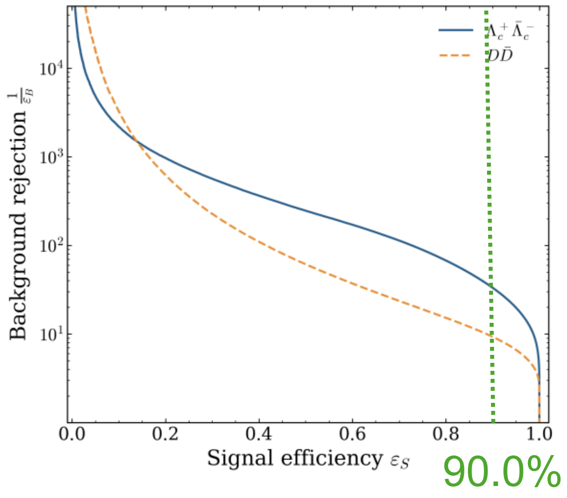
Pre-training



Visualization with **best** epoch



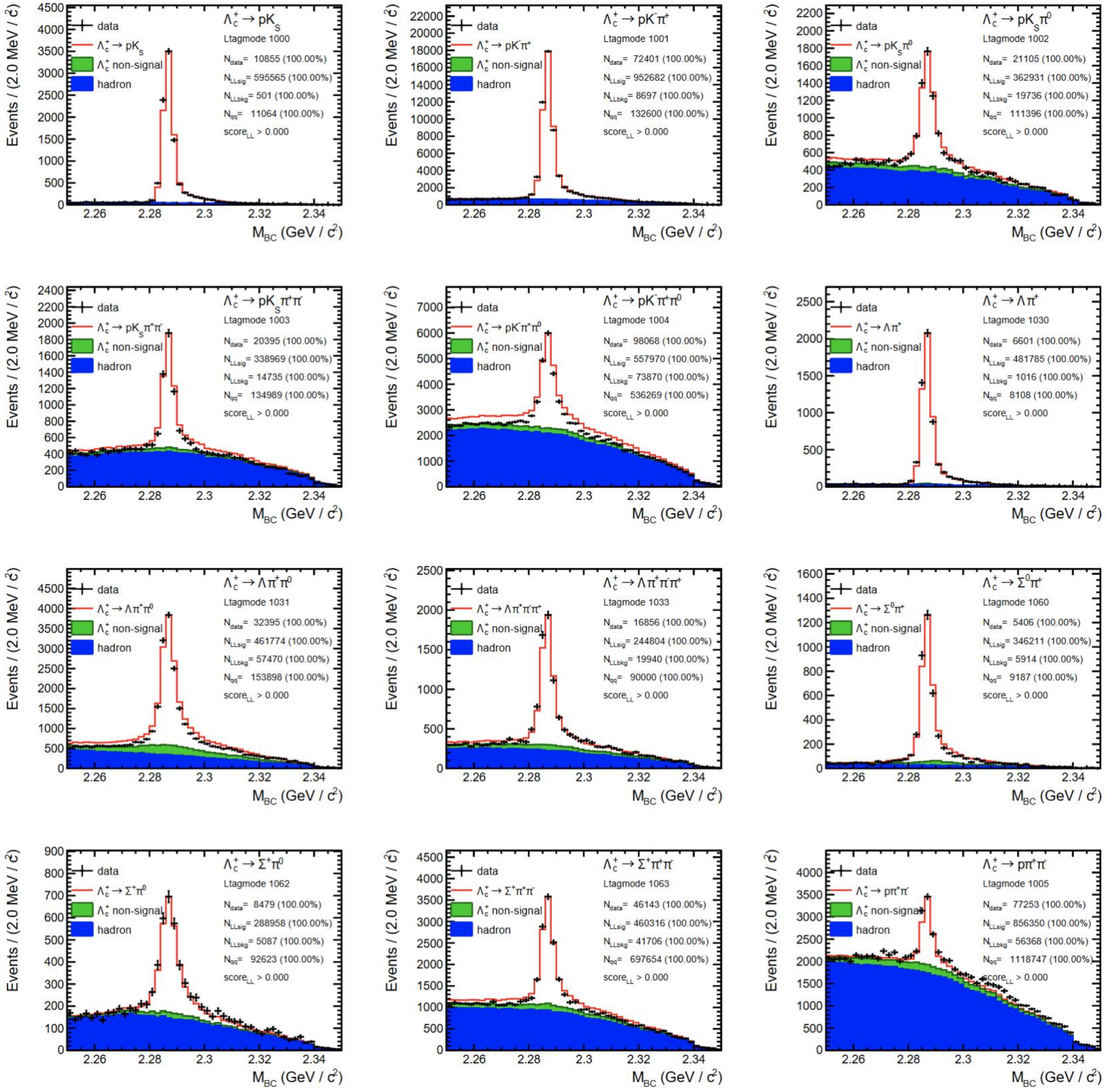
Accuracy comparison



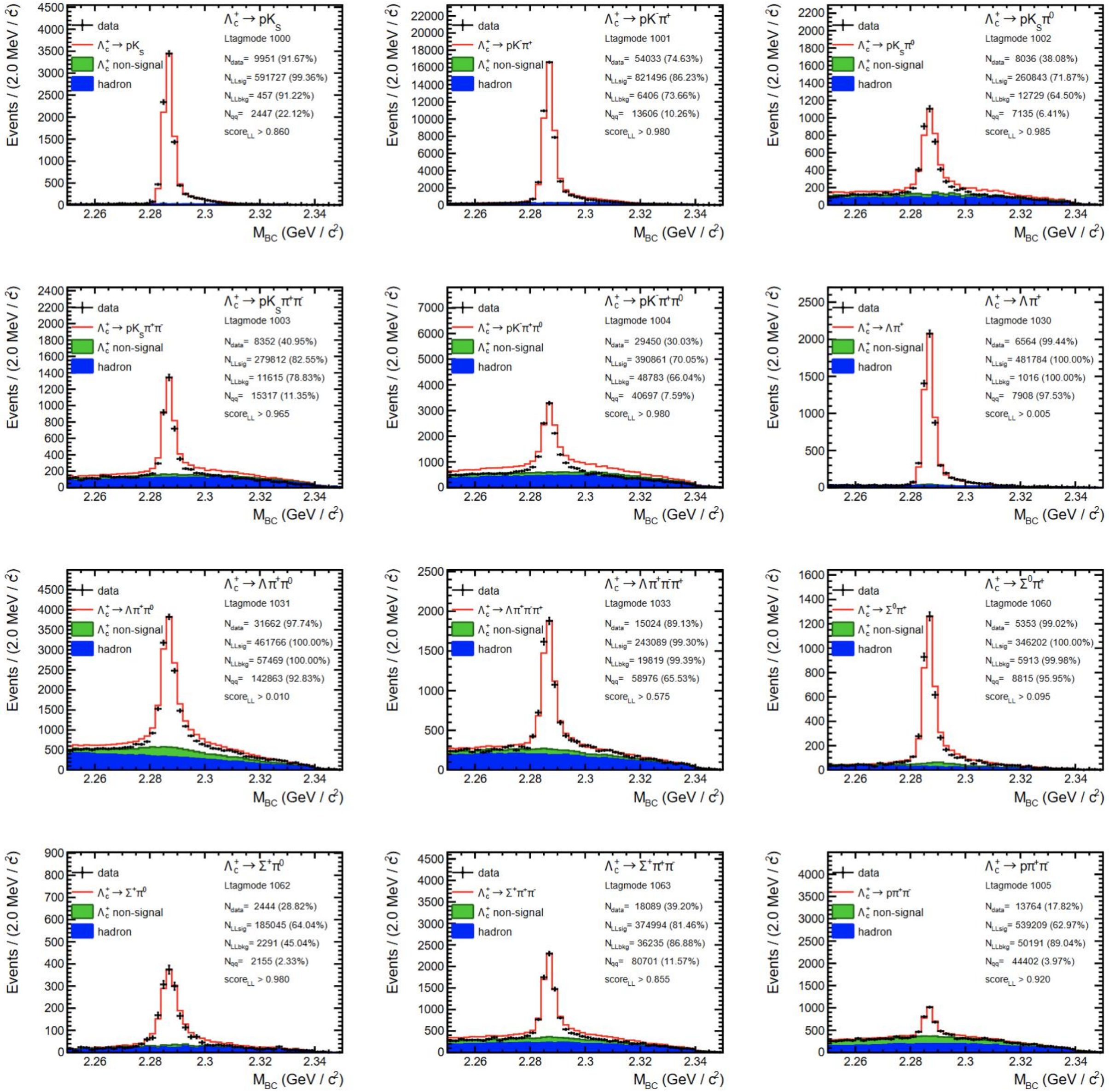
Background rejection with different signal efficiency

Performance on 12 Λ_c^+ tagged decay channels (1)

- *Flat background, no visible peaks*

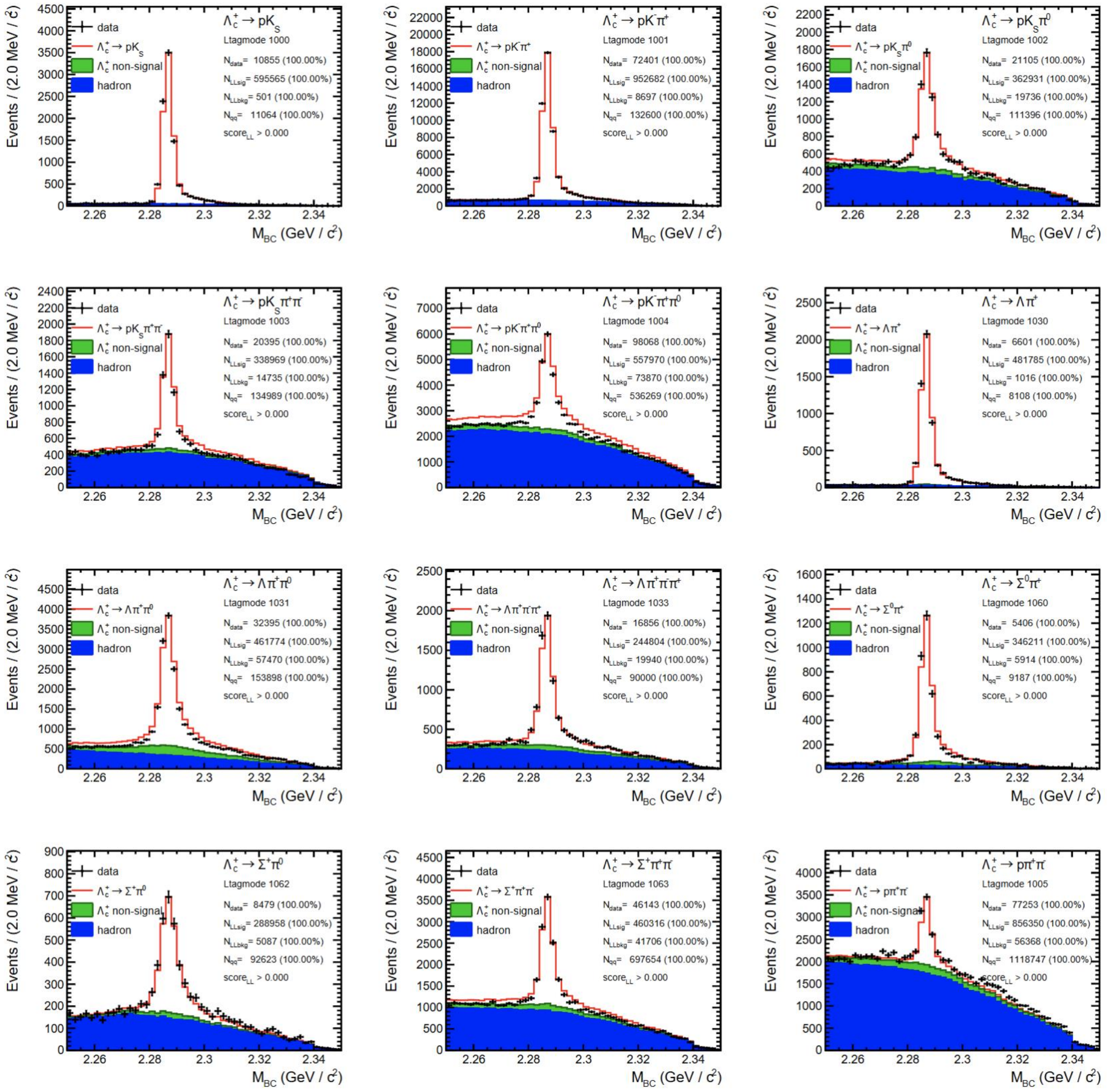


Conventional (cut-based) method

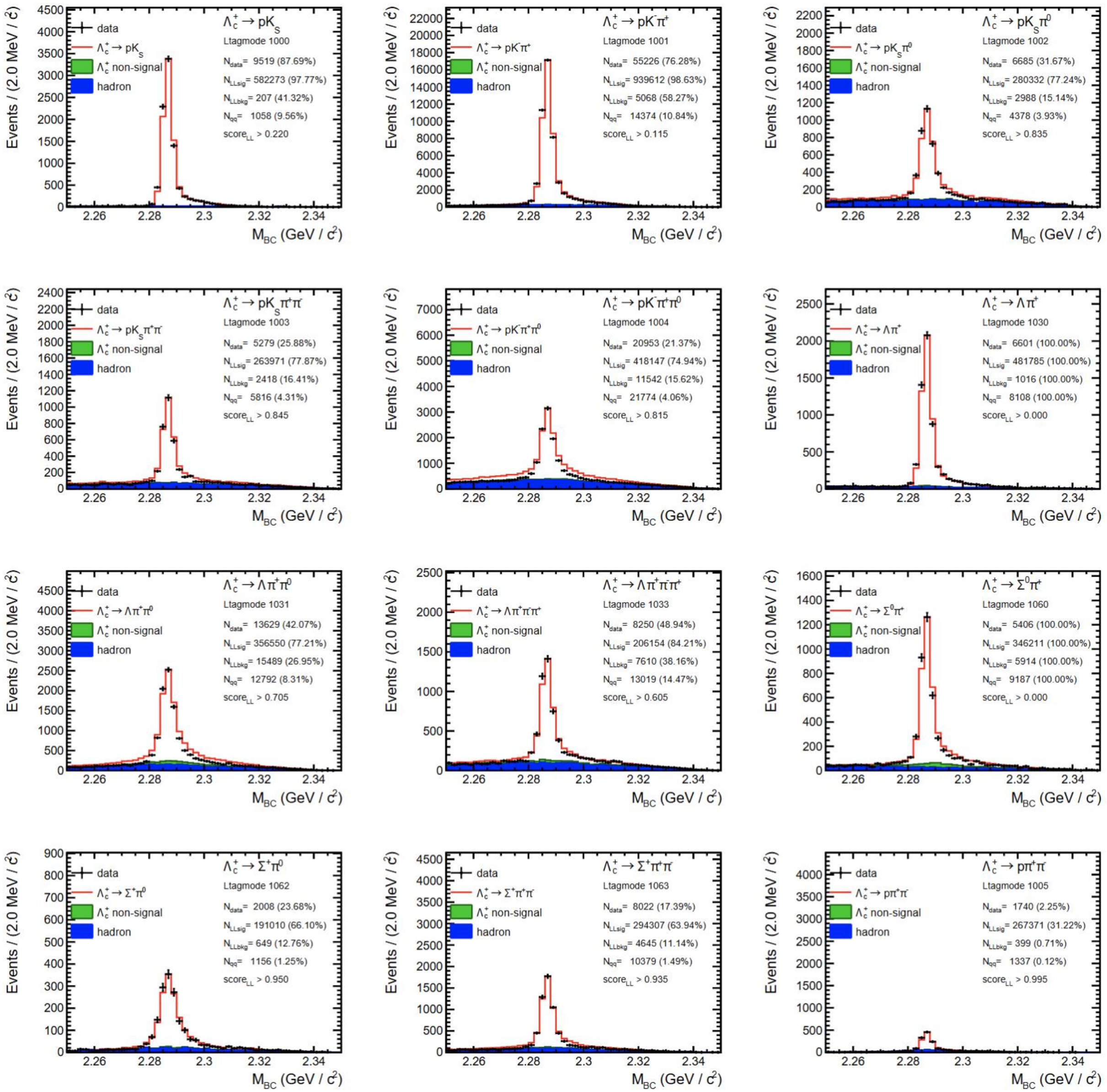


Pre-training method

Performance on 12 Λ_c^+ tagged decay channels(2)



Conventional (cut-based) method



Pre-training & Fine-tuning method

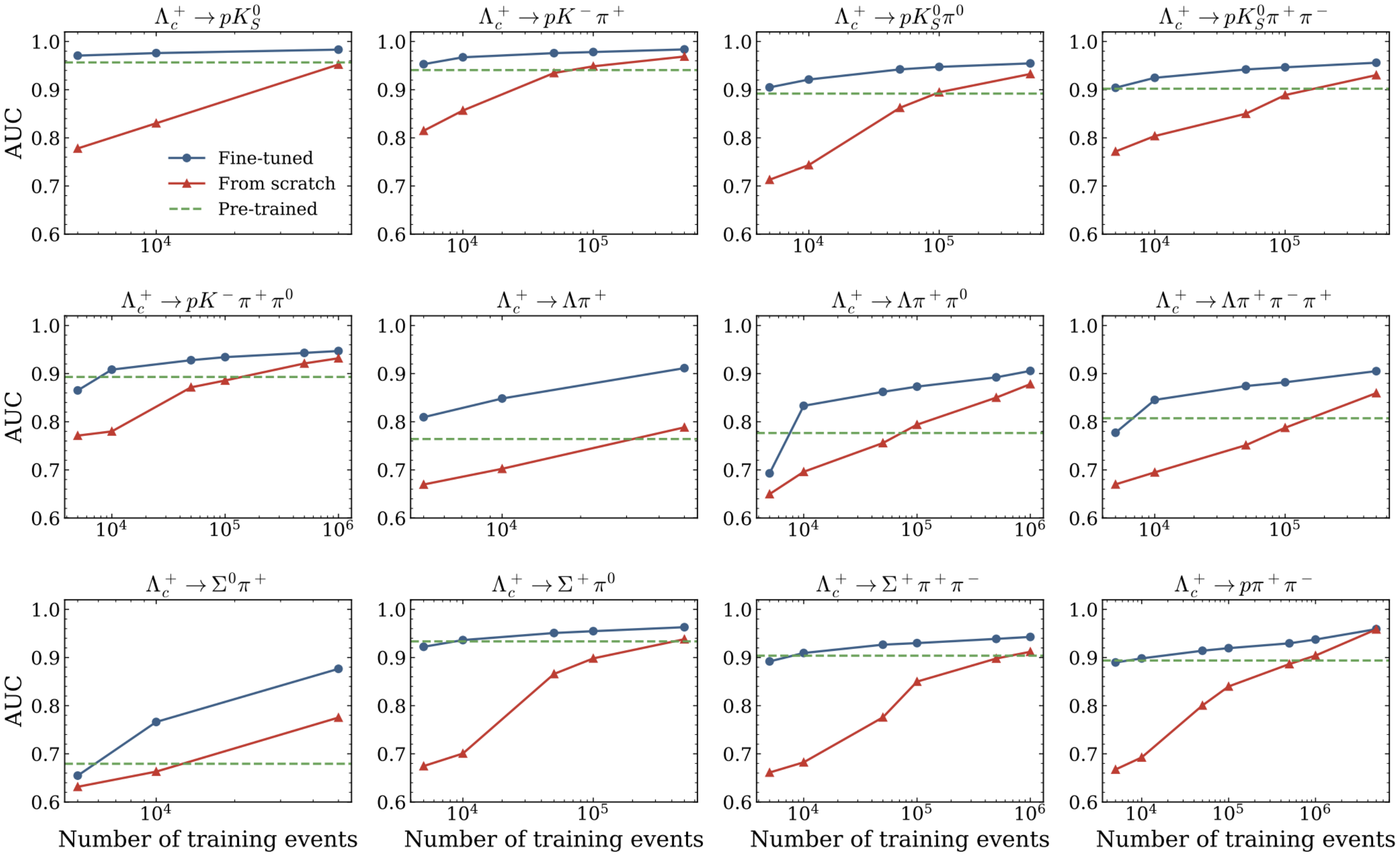
Performance on 12 Λ_c^+ tagged decay channels(3)

- AUC value
  - Evaluate the performance of event classification

1. Pre-training model achieves adequate performance

2. Low-statistics regime
   - ~ 5K–100K events
   - Pre-training ≥ Training from scratch
   - Fine-tuning ≫ Training from scratch

3. High-statistics regime
   - ~ several million events
   - Fine-tuning ≈ Training from scratch



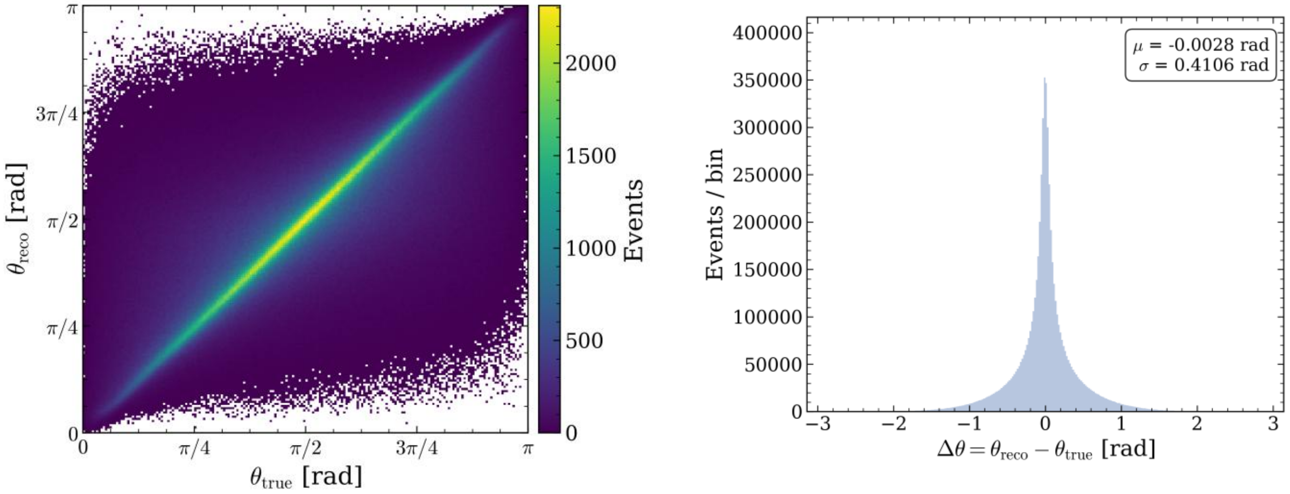
AUC comparison for $\Lambda_c^+ \to$ 12 tagged decay channels

$\Lambda_c^+\overline{\Lambda}_c^-$ momentum-direction regression model

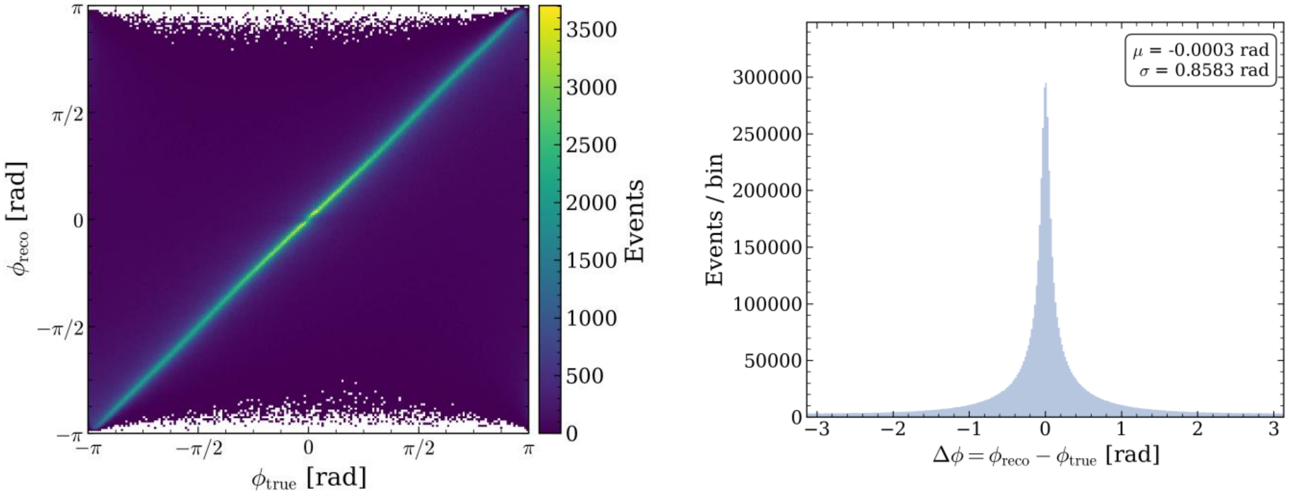
- $\Lambda_c^+\overline{\Lambda}_c^-$ momentum-direction regression
  - Spherical coordinate system
  - $\Lambda_c^+(\theta,\phi)$, $\overline{\Lambda}_c^-(\theta,\phi)$

- Pre-training
  - 6 * NVIDIA V100 GPUs ~ 4 Days

- **Pre-training** model performance
  - For $\Lambda_c^+\overline{\Lambda}_c^-$ inclusive MC events
  - Central bias μ
    - No significant central bias
    - $\mu(\Delta\theta)$ = –0.003 rad
    - $\mu(\Delta\phi)$ = 0.000 rad
  - Angular resolution σ
    - $\sigma(\Delta\theta)$ = 0.411 rad
    - $\sigma(\Delta\phi)$ = 0.858 rad



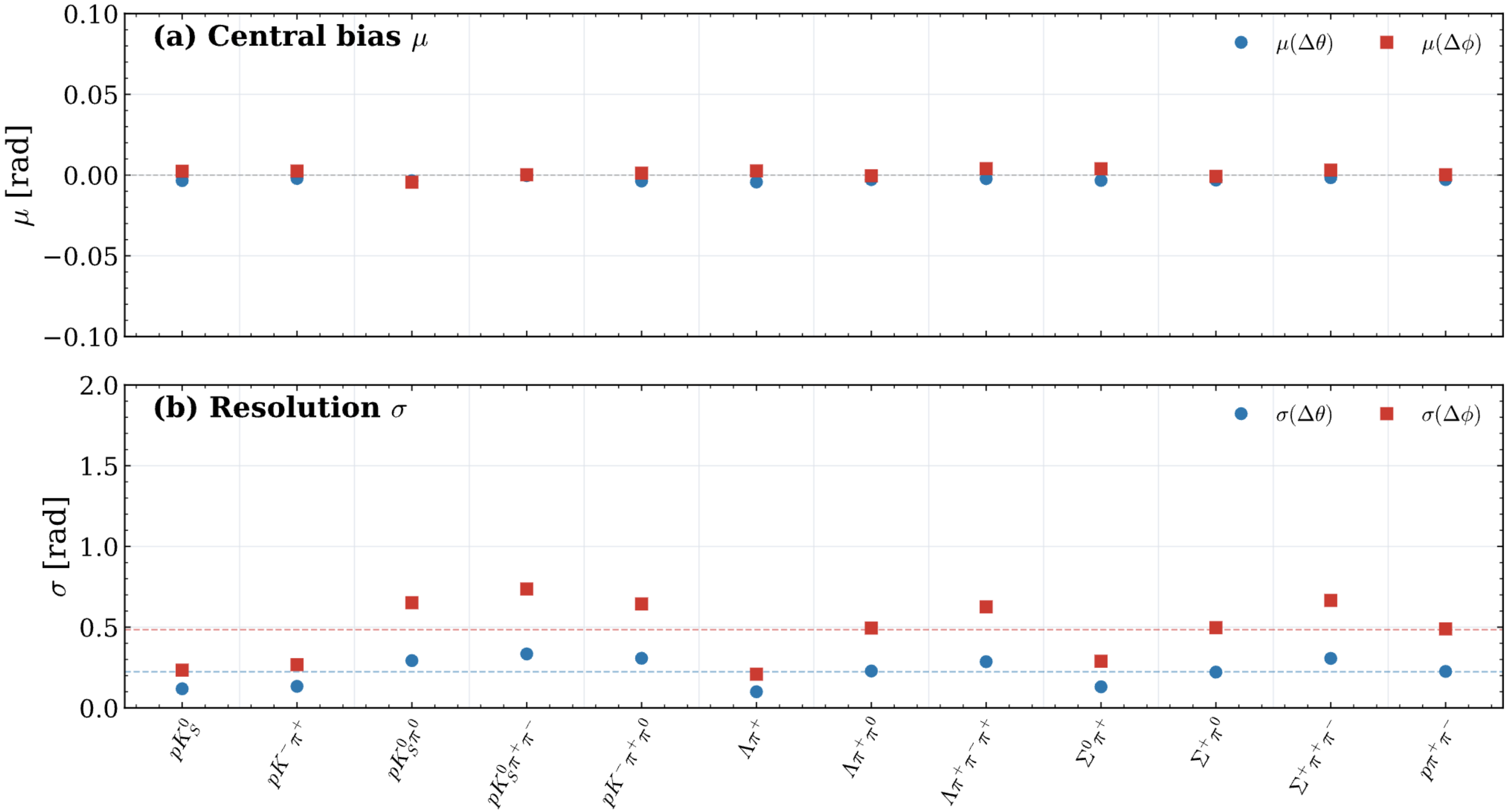
Comparison of the reconstructed angle $\boldsymbol{\theta}$ with the MC truth



Comparison of the reconstructed angle $\boldsymbol{\phi}$ with the MC truth

Performance on 12 Λ_c^+ tagged decay channels

- **Pre-training** model performance
  - Average value across all 12 tag channels
  - No significant central bias
    - $\mu(\Delta\theta)$ = −0.003 rad
    - $\mu(\Delta\phi)$ = 0.001 rad
  - Angular resolution
    - $\sigma(\Delta\theta)$ = 0.224 rad
    - $\sigma(\Delta\phi)$ = 0.484 rad

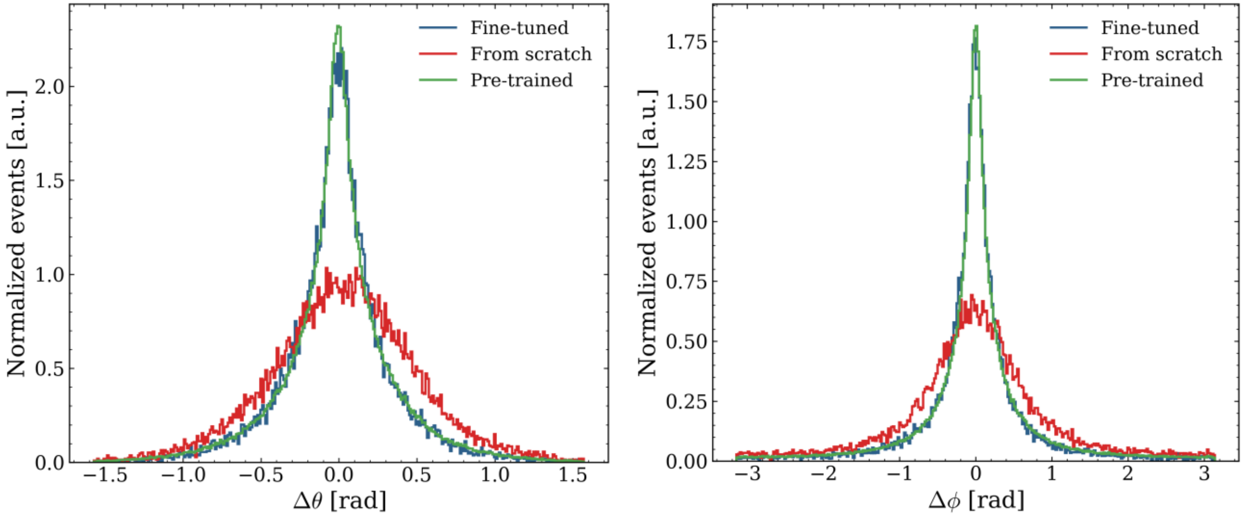


Central bias and resolution across the 12 channels

Application in the $\Lambda_c^+ \to \Lambda(1405/1520)\, e^+ \nu_e$

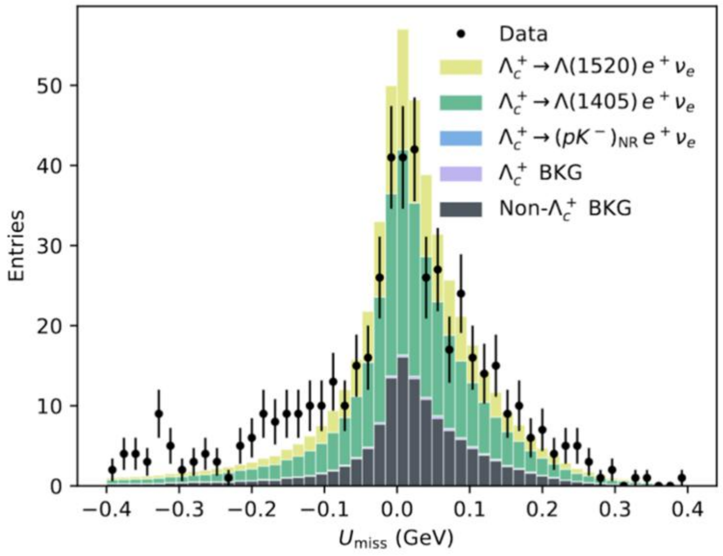
- Momentum-direction regression model
  - Mainly applicable to processes involving missing particles
  - Inferring $\Lambda_c^+ \bar{\Lambda}_c^-$ momentum information to reconstruct missing particle with double tag

  - $\Lambda_c^+ \to \Lambda(1405)\, e^+ \nu_e$, 8.5×10^4 events
  - Pre-training $\approx$ Fine-tuning $\gg$ Training from scratch



Comparison of $\Lambda_c^+ \to \Lambda(1405)\, e^+ \nu_e$ among three methods

- For $\Lambda_c^+ \to \Lambda(1405/1520)\, e^+ \nu_e$ physics analysis
  - Application to the U_{miss} distribution
  - Background events from pre-training model are randomly distributed



The U_{miss} distribution with Pre-training model

Next

- ***UcasMLTool:*** **A General Purpose Machine Learning Variable Extraction Toolkit**
  - Conveniently integrates into user analysis package
  - Automatically extracts event input features
  - Supports different types of physics analyses
  - Facilitate the use of *Particle Transformer* across a broad range of BESIII physics analyses
  - Currently undergoing internal testing

- ***Future development***
  - Facilitate online event classification during data taking
  - Other physics at BESIII
    - D_S mesons, $\Xi^0, \Xi^{\mp}$…
  - Can also be extended to STCF, CEPC …

Summary

- **Pre-training** models developed for Λ_c^+ physics

- Event classification Model
  - Pre-trained with $\Lambda_c^+\bar{\Lambda}_c^-, D\bar{D}h, q\bar{q}$ $(7.1\times10^7$ events)
  - Good performance with $\Lambda_c^+\bar{\Lambda}_c^-$ 90.0% signal efficiency, and reject 97.0% background events
  - Consistent performance across all 12 tagged decay channels
    - Low-statistics regime (~ 5K–100K events), Fine-tuning $\gg$ Training from scratch

- Momentum-direction regression
  - Pre-trained with $\Lambda_c^+(\theta,\phi)$, $\bar{\Lambda}_c^-(\theta,\phi)$ $(2.4\times10^7$ events)
  - Performance on 12 tagged decay channels
    - No significant central bias (< 0.003 rad), good angular resolution (<0.484 rad)
  - Good performance on $\Lambda_c^+ \rightarrow \Lambda(1405)\, e^+\nu_e$ $(8.5\times10^4$ events)

- UcasMLTool Toolkit
  - Facilitate the use of machine learning across a broad range of BESIII physics analyses

Thanks!