High-Granularity Calorimeters Using Machine Learning

Xiang li ,Hongbin Diao,Yunlong Zhang*

State key laboratory of particle detection and electronics

University of Science and Technology of China

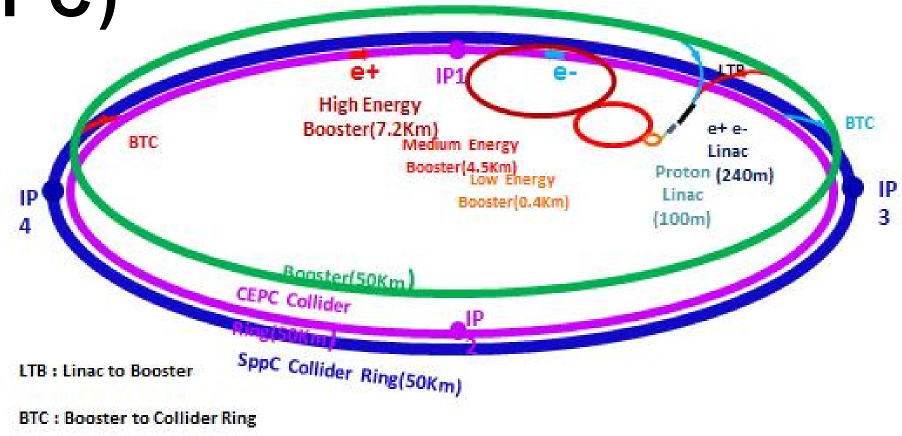
Outline

- Introduction and AHCAL Prototype
- Beam Tests and Digitization of Simulation
- Motivation
- CNN+LSTM
- Calib cell level parameter a,b
- Result

Introduction

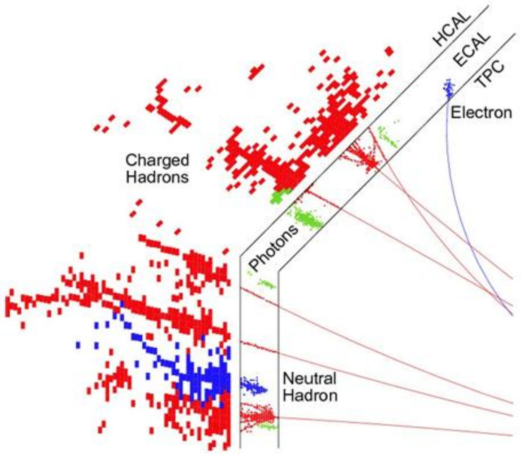
➢ Circular Electron Positron Collider (CEPC)

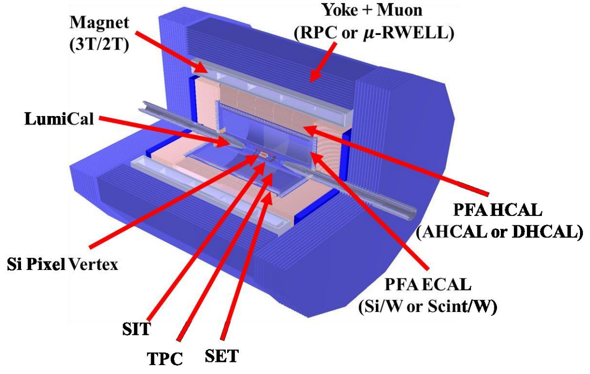
- ➢ $E_{cm} \approx 240$ GeV (higgs mode)
- ➢ luminosity ~ 8.3×10^{34} cm^{-2}s^{-1} (higgs mode)
- ➢ challenges: jet energy resolution < $\frac{30\%}{\sqrt{E[GeV]}}$



➢ Particle Flow Algorithm (PFA)

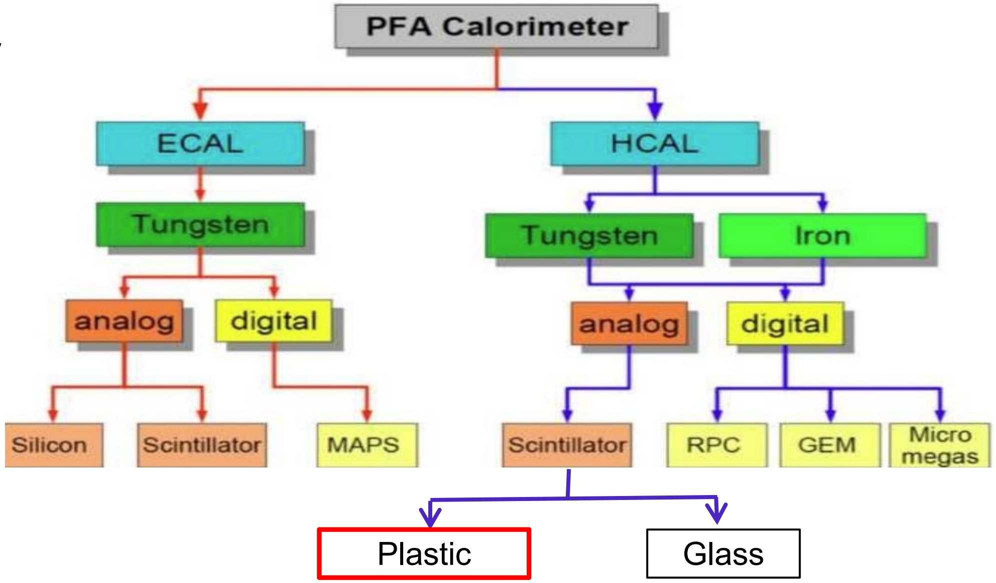
- ➢ High granularity
- ➢ Good track finding
- ➢ Good energy resolution



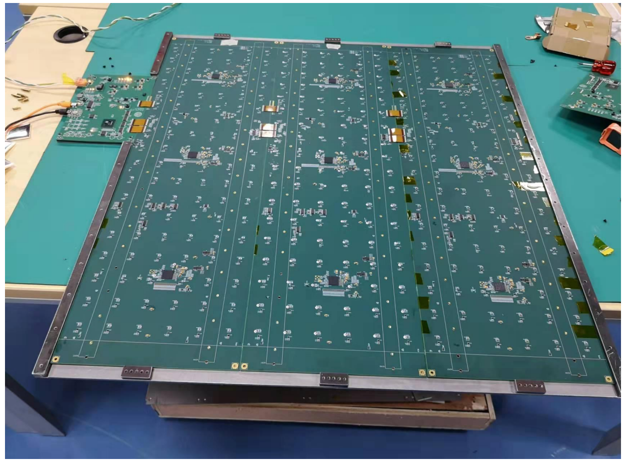


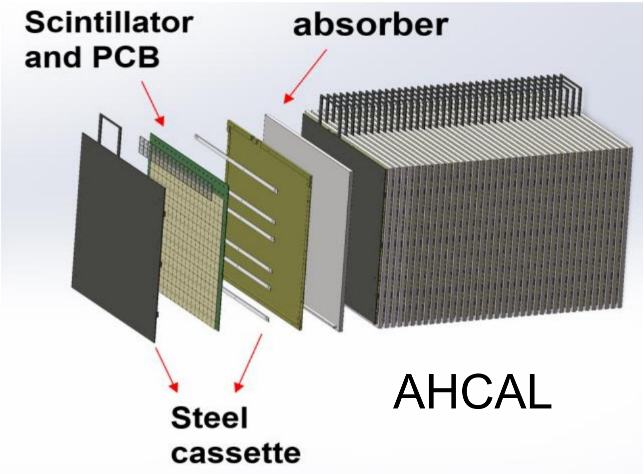
AHCAL Prototype

➢ An Analogue readout Hadronic Calorimeter prototype has been developed



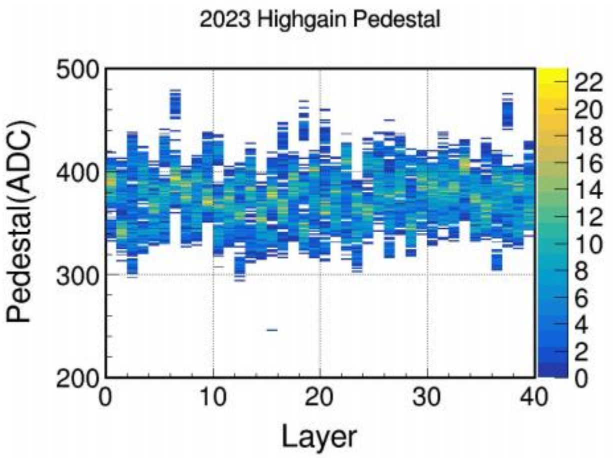
Total layers	40
Sensitive area	72cm*72cm
Granularity	4cm*4cm*0.3cm
Total channels	12960
NIL	4.6
Energy Resolution	<60%@ 1 GeV
Weight	5.0T

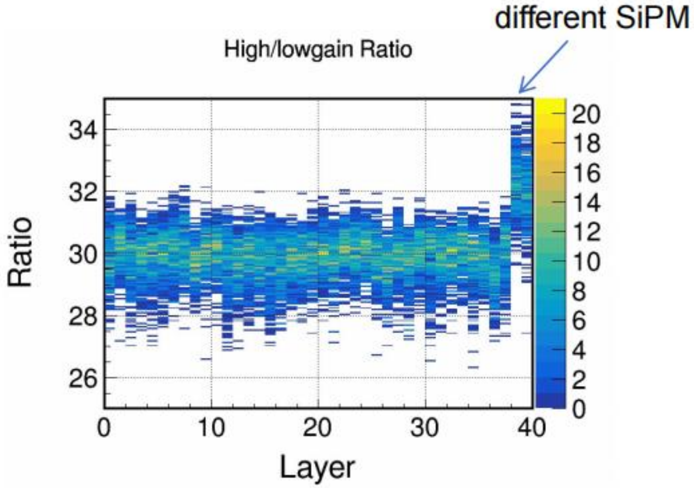


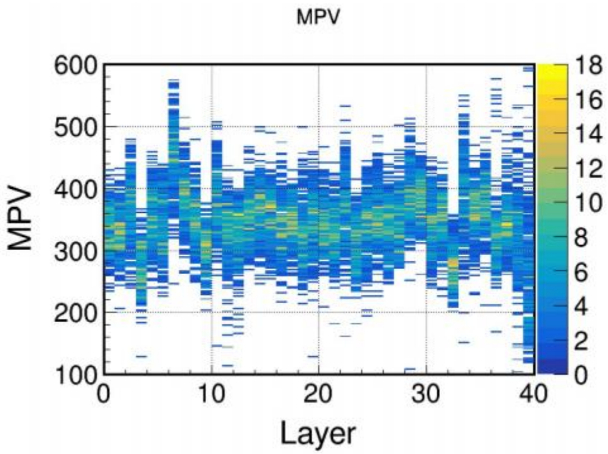


Beam Tests

- **beam tests have been done at CERN**
  - Electron (1-120GeV/c)
- **Parameters Calibration**
  - Pedestal Calibration
  - High-Lowgain Calibration
  - MIP Calibration
  - SPE Calibration







Digitization of Simulation

University of Science and Technology of China

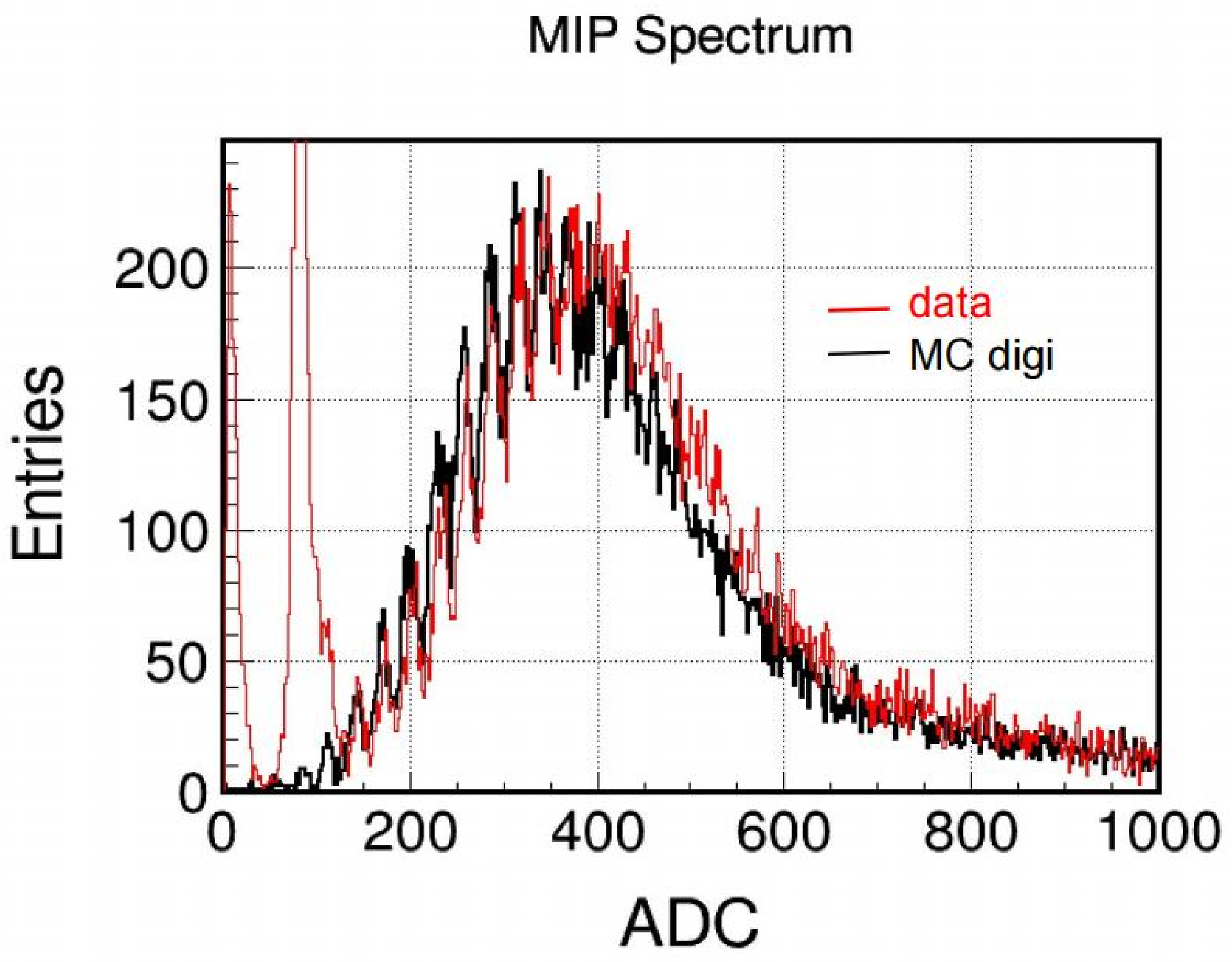
➢ **Scintillator** non-uniformity

- Energy to photon (poisson sampling)
- 4.8% gaussian smear from non-uniformity of scintillator
- Photon electron conversion (binomial sampling)

➢ **SiPM response**

- SiPM saturation and flacturation
- SiPM gain smear

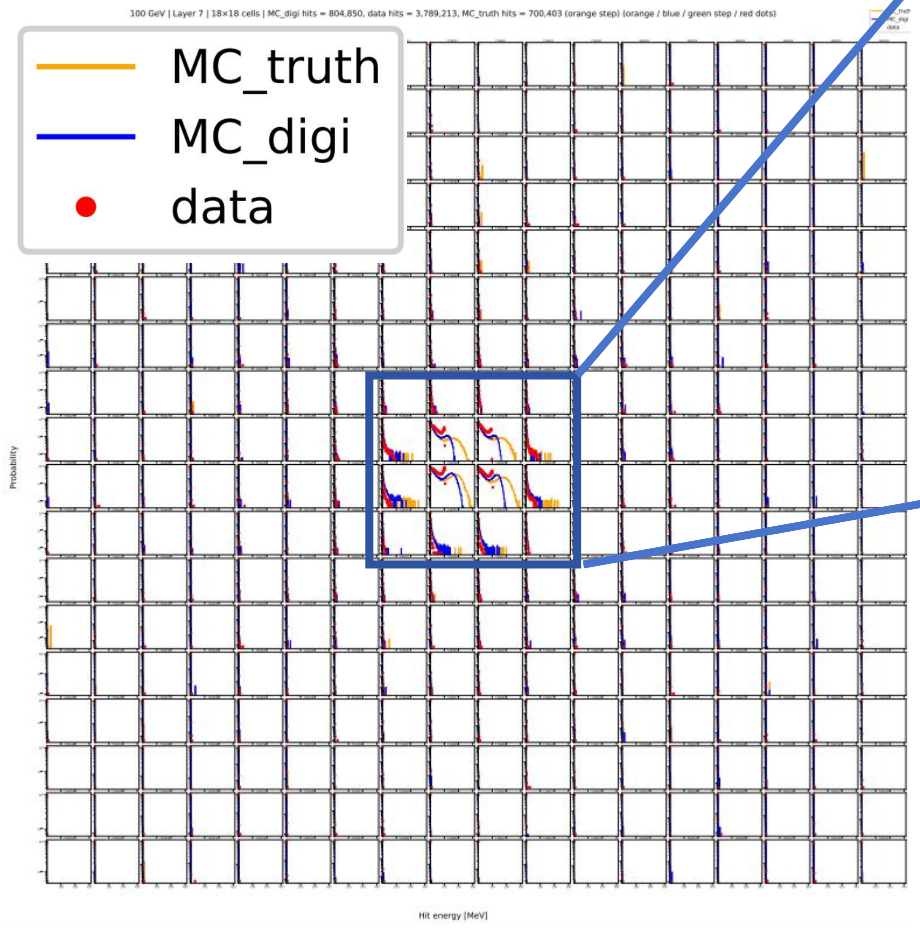
➢ **Convert SiPM signal to data format ADC**

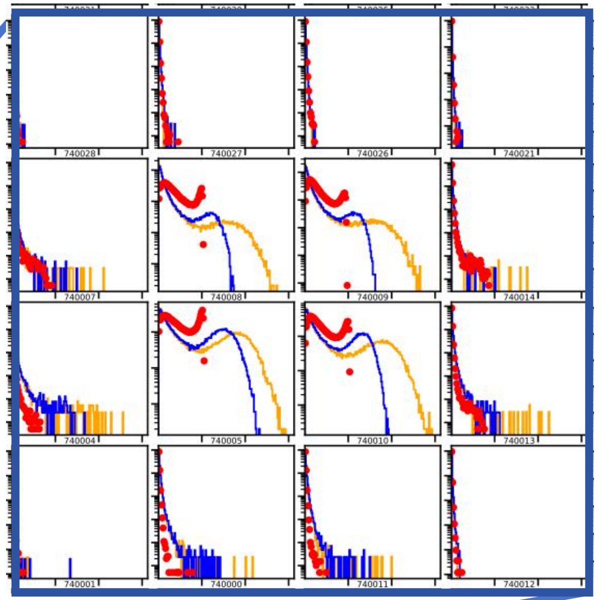


Cell level difference

- **Difference of MC_truth MC_digi Data**

100 GeV | Layer 7 | 18×18 cells





- MC_truth
- Raw MC simulation

- MC_digi
- Reconstruct enegy by applying MPV
- After SiPM non-linearity

- Data
- Beamtest data (SPS)

◆ Reason of diference

- Saturation effect and nonuniformity
- Electronic non-linearity
- Beam center and Beam size

....

Motivation

中国科学技术大学
University of Science and Technology of China

- Understand the detector response and strive to improve the simulation.
- Attempt to improve the HCAL e- energy resolution using ML
  - To Correct Saturation effect and nonuniformity:
  - **Key Hypothesis:** The discrepancy between MC_digi and Real Data mainly stems from electronics non-linearity

- Electronic non-linearity digitization function

$$\frac{1}{\left(\frac{1}{x^b}+\frac{1}{a^b}\right)^{1/b}}$$

- Inverse function to correct electronic non-linearity

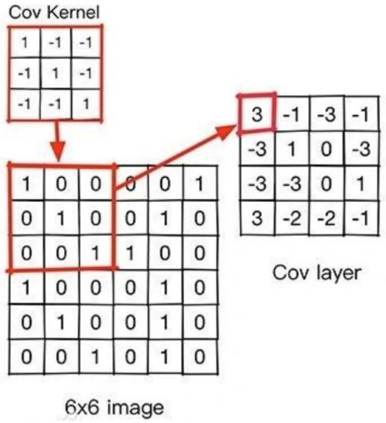
$$f^{-1}(x)=\frac{x}{\left[1-\left(\frac{x}{a}\right)^b\right]^{1/b}}$$

- Strategy
  - Using Hit leve machine learing model to calib cell level parameter a,b
  - Hit level data: Train a CNN+LSTM

CNN+LSTM

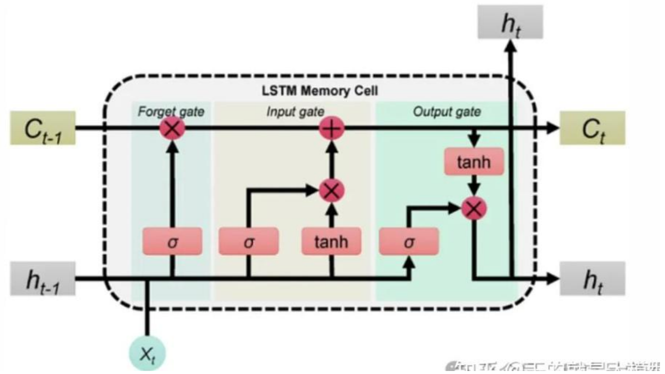
- **CNN (Convolutional Neural Network)**

Read the hit image of the super layer

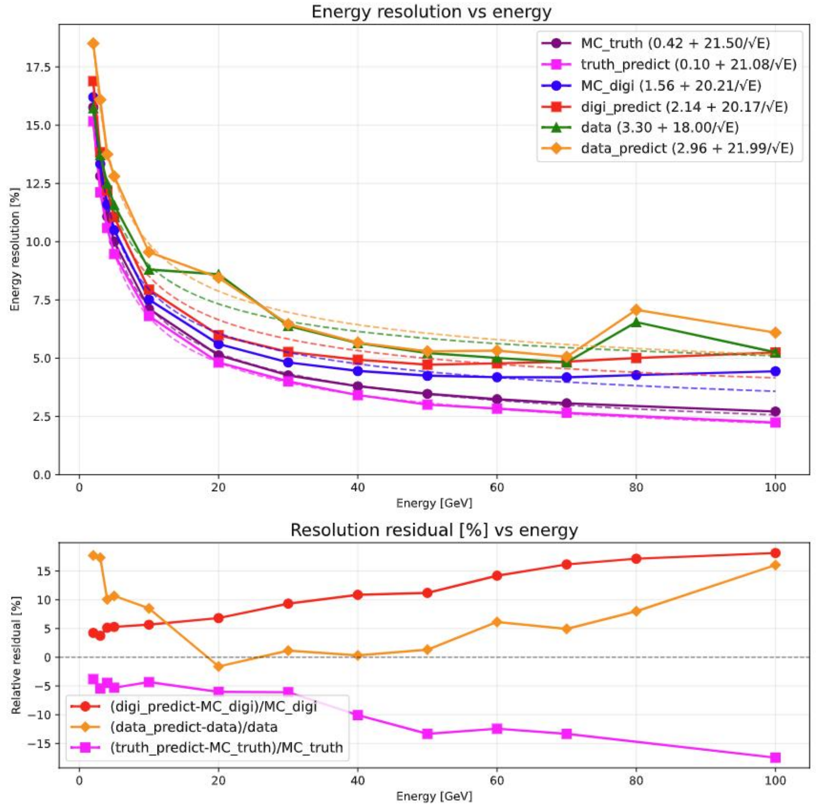


- **LSTM (Long Short-Term Memory)**

Learn the development of shower at different levels



Samples： MC_truth： (2000000) Continuous e-

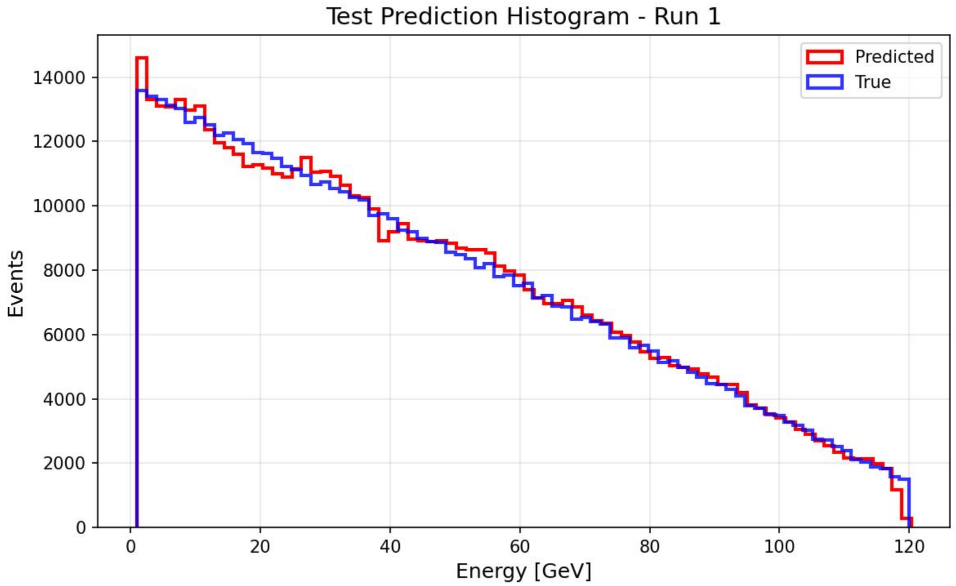


- **Key Performance:** Achieved a **10% improvement** in energy resolution for MC_truth.
- **Performance Drop on Realism:** Model performance degrades on **MC_digi** and **Real Data**

Calib cell level parameter a,b

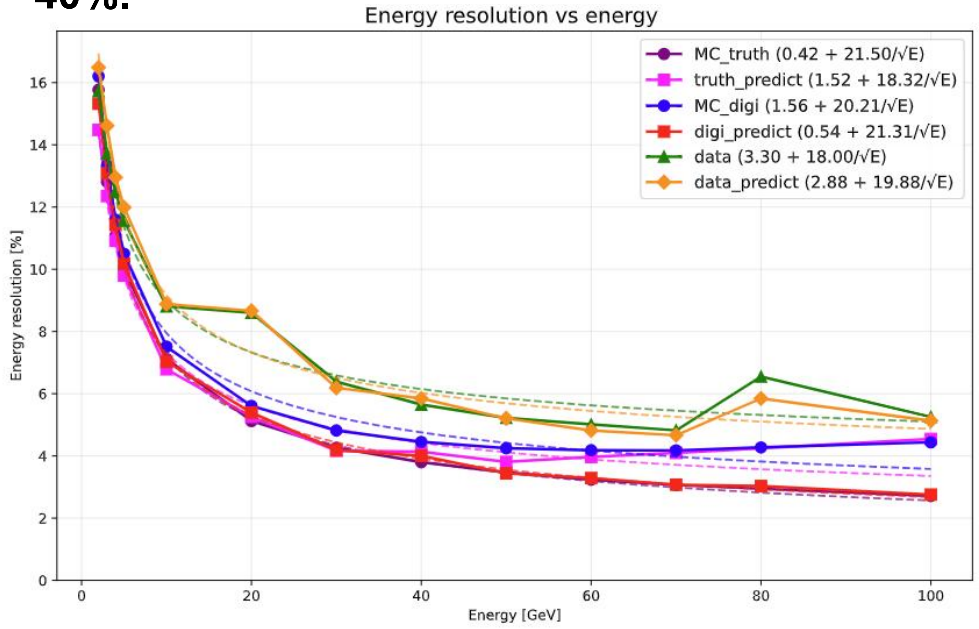
Step 1: Stabel CNN+LSTM model

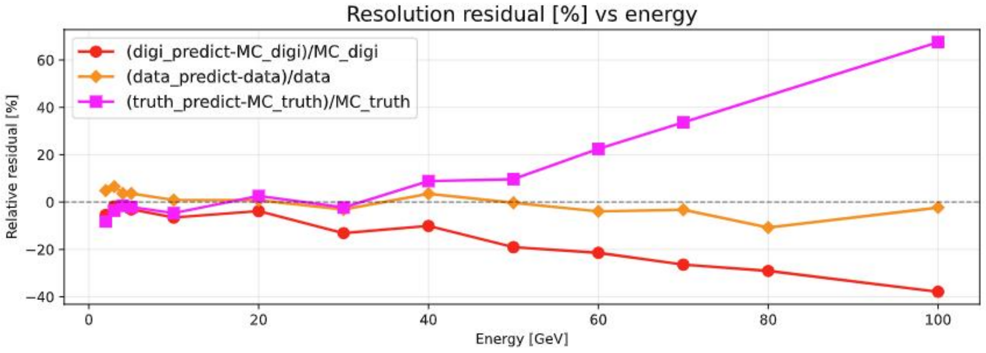
- ➢ Samples: MC_digi: 1-120GeV (12000000) Continuous e-



- Low-energy events produce fewer hits per event.--More events are needed
- As a result, the MC_digi dataset follows a trapezoidal event distribution.

- **Key Performance:** Achieved a **~20% improvement** in energy resolution for MC_digi. **Max improvement** arrive to **40%.**





Calib cell level parameter a,b

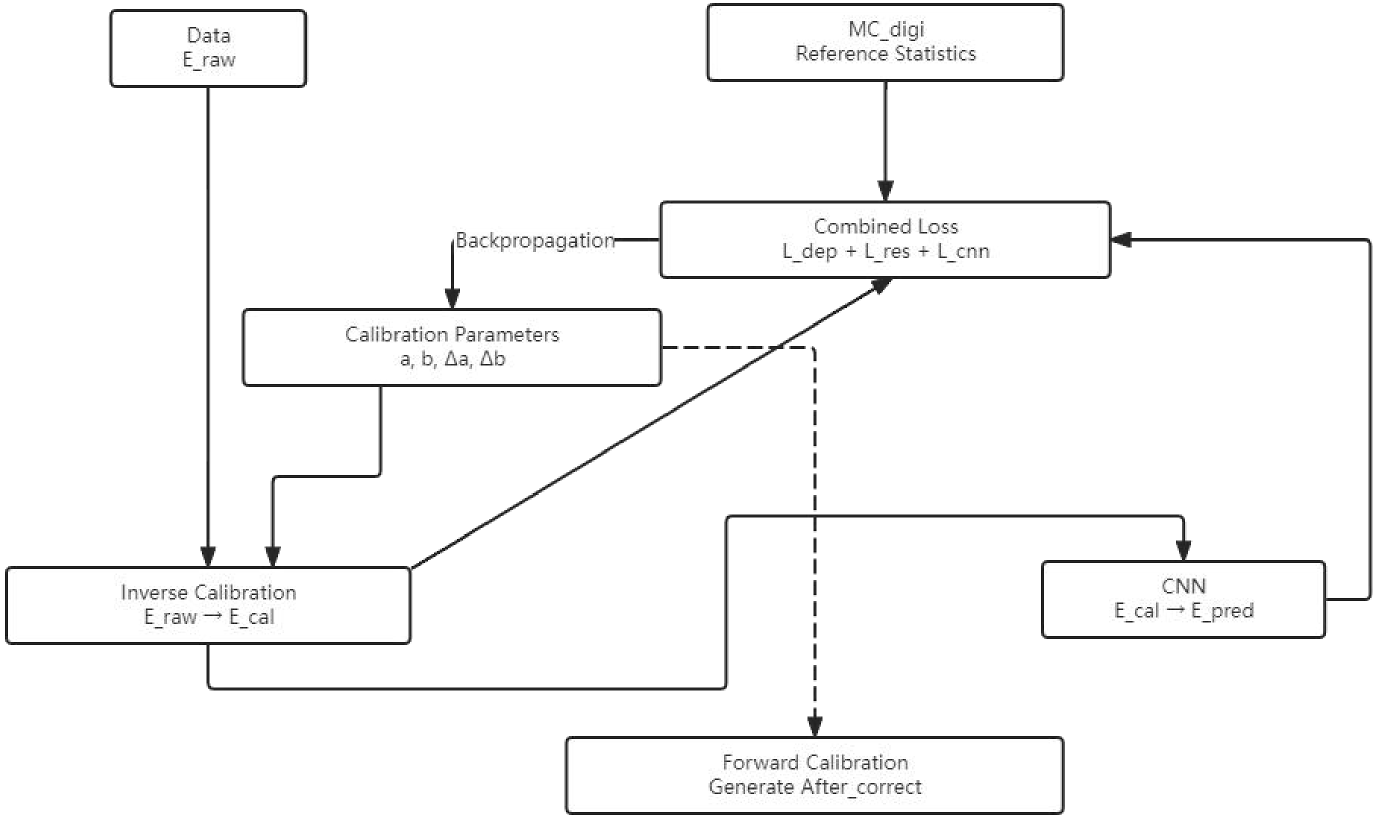
University of Science and Technology of China

Step 2: Using Inverse function to make data approch MC_digi.

$$f^{-1}(x) = \frac{x}{\left[1 - \left(\frac{x}{a}\right)^{b}\right]^{1/b}}$$

SPS e- data 80,100,120 GeV

MC_digi 80,100,120 GeV



Total Loss：

$$L = L_{ref} + \lambda_{cnn} L_{cnn} + \lambda_{dep} L_{dep}$$

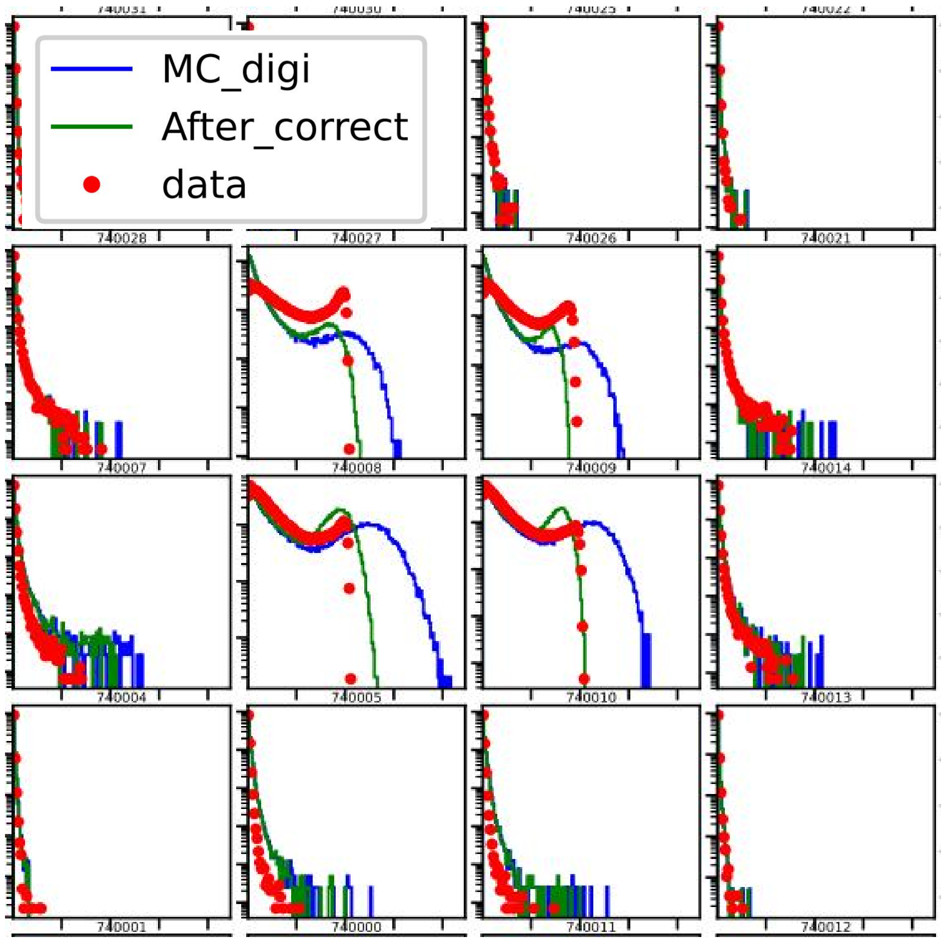
- ➢ Reference Loss： L_{ref}
  - Minimize the discrepancy with the MC_digi distribution.
- ➢ CNN Energy Prediction Loss : L_{cnn}
  - Evaluates the accuracy of the network's macroscopic physical prediction.
- ➢ Cell-level Deposition Loss： L_{dep}
  - Constrains the energy deposition distribution at the microscopic (per-cell) level.

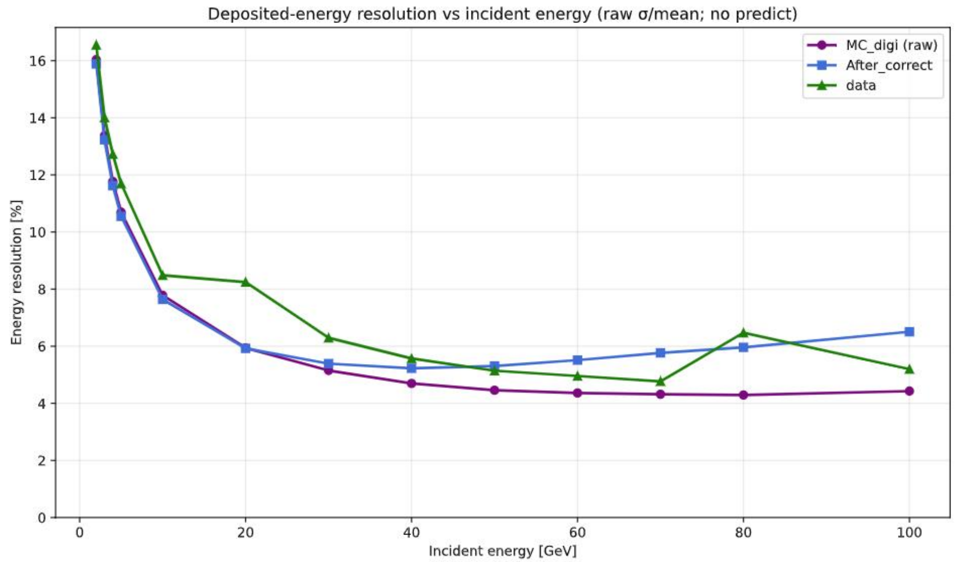
Calib cell level parameter a,b

Result：Apply parameter a,b

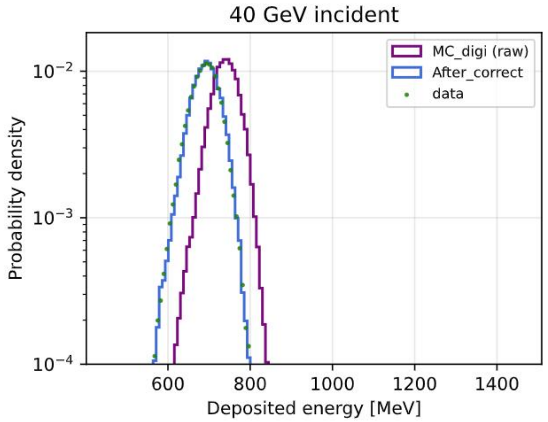
$$\frac{1}{\left(\frac{1}{x^b}+\frac{1}{a^b}\right)^{1/b}}$$

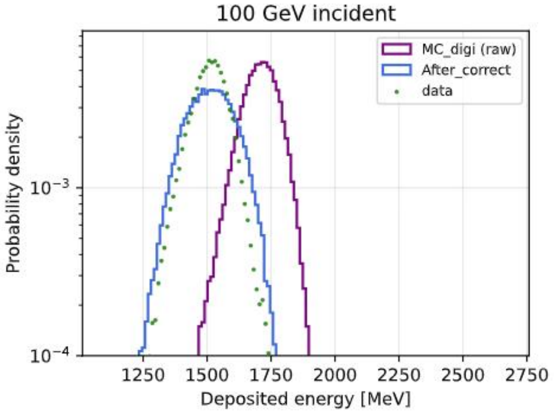
100 GeV | Layer 7 | Center 4×4 cells





- Differences in beam center and beam size **prevent a perfect match** between MC_digi_correct and data



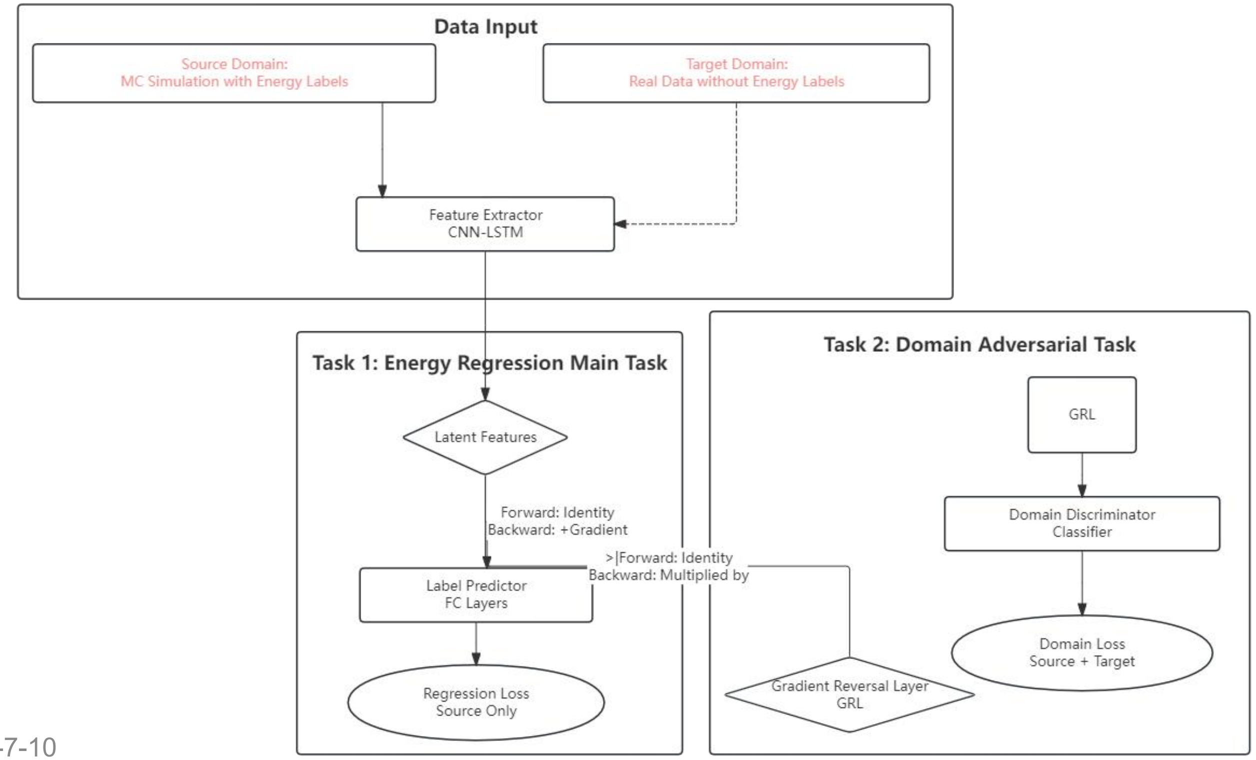


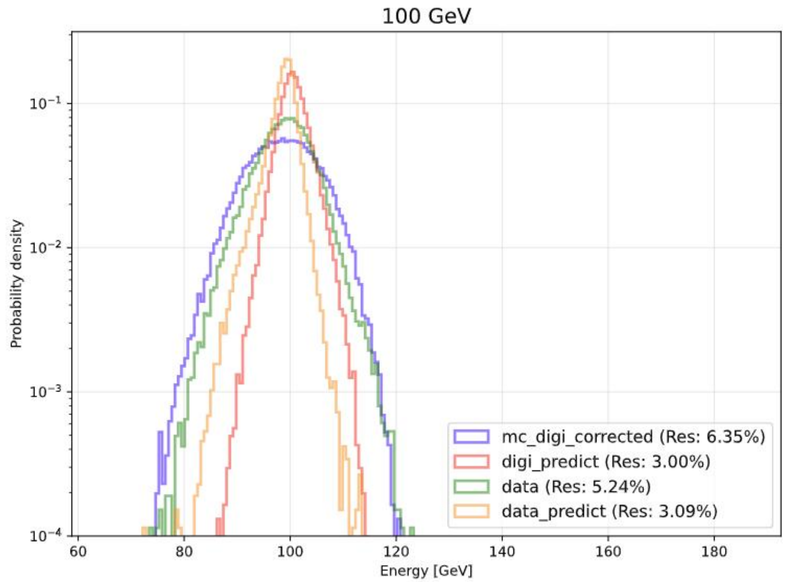
Retrain CNN+LSTM

Result: Retrain CNN+LSTM to improve energy resolution for data

- DANN (Domain-Adversarial Neural Network)
- Samples: MC_digi_corrected: 1-120GeV (400000) Continuous e-

  Data: PS and SPS [1-120] GeV Descret energy

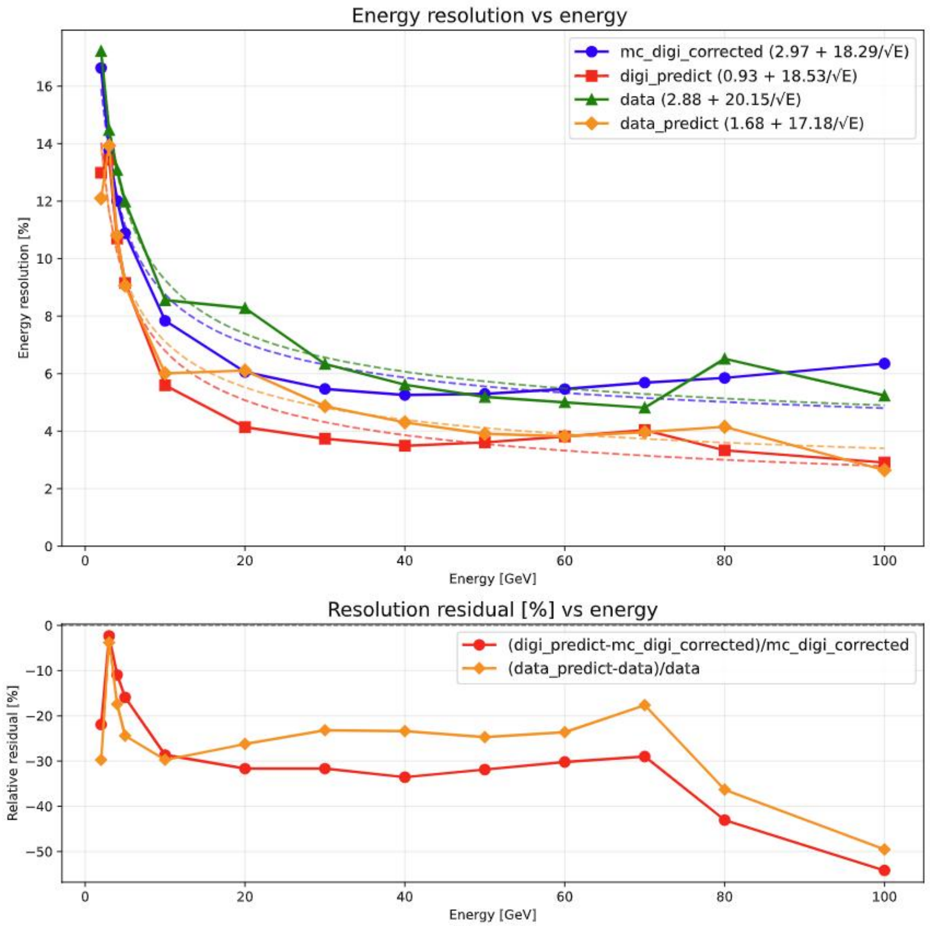




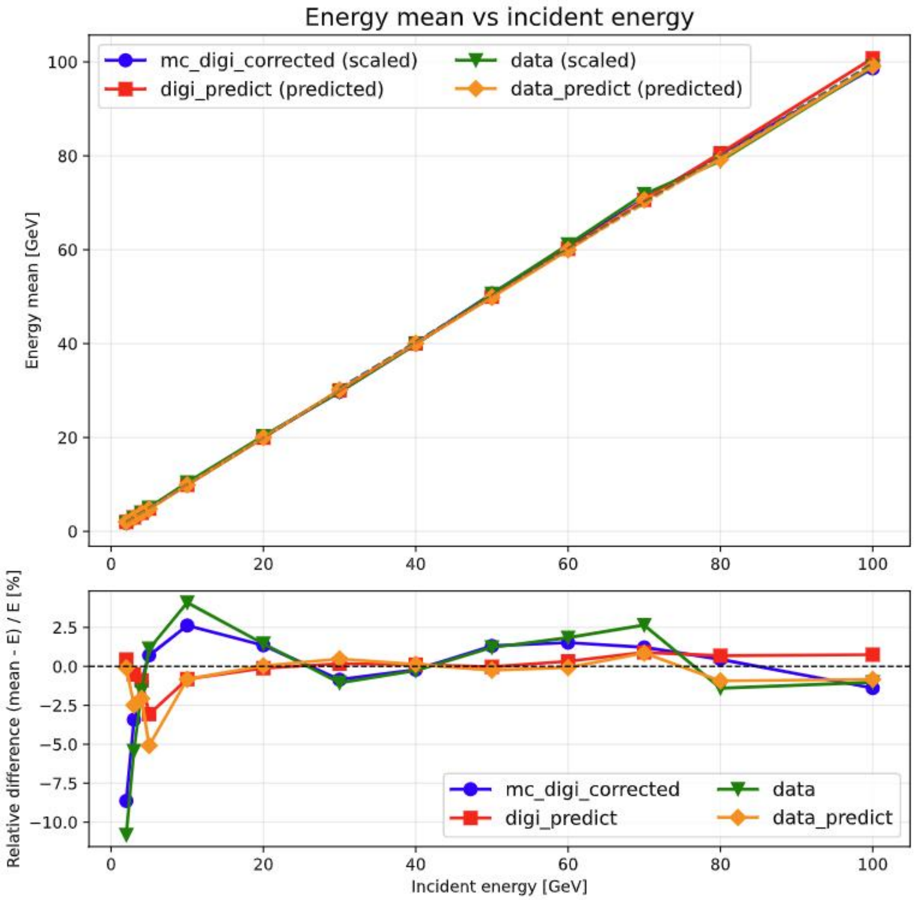
- Preserves Core Physics: Retains essential information for accurate energy prediction (satisfies Label Predictor).
- Acts as an "Eraser": Wipes away both MC "simulation fingerprints" and Data "real noise."

Retrain CNN+LSTM

- **Key Performance:** Achieved a **~20% improvement** in energy resolution for data. **Max improvement** arrive to **40%**



- **Key Performance:** The energy linearity has significantly improved.Linearity error **kept within 5%**



Summery

中国科学技术大学
University of Science and Technology of China

- **Background & Challenges**
  - Detector non-linearity and beam profile (center/size) differences caused a severe MC-Data "sim-to-real Gap."
- **Calibration & DANN Strategy**
  - Explored CNN+LSTM for CEPC AHCAL energy reconstruction.
  - Inverse Calibration: Applied at the cell level to correct electronics non-linearity, aligning real data with simulation.
  - DANN Integration: Addressed residual discrepancies between MC and Data.
- **Performance Breakthrough**
  - Resolution: Improved by ~20% for real data, with a maximum local improvement of 40%.
  - Linearity: Significantly enhanced, with the linearity error successfully kept within 5%.