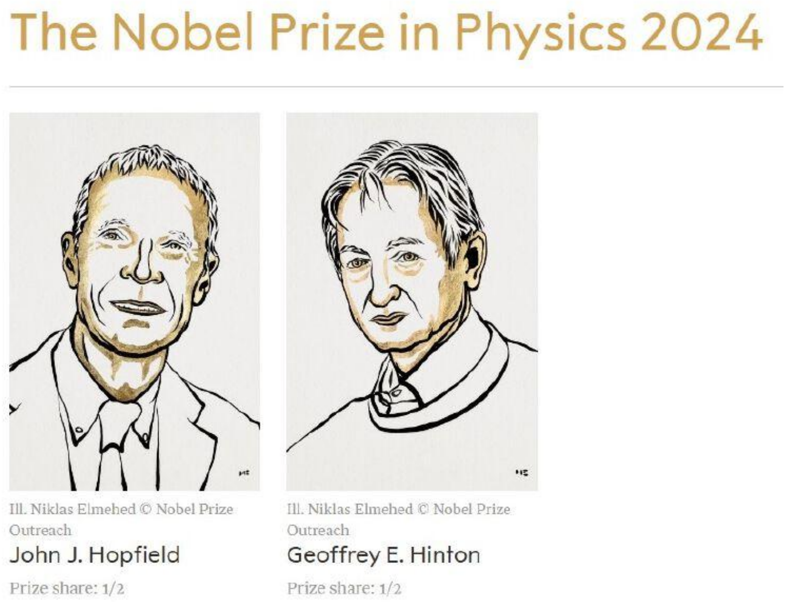
AI For LHAASO

之江实验室 天文计算研究中心
Zhejiang Lab, Research Center for Astronomical Computing
徐宸原 for LHAASO AI Study Group

目录

Content

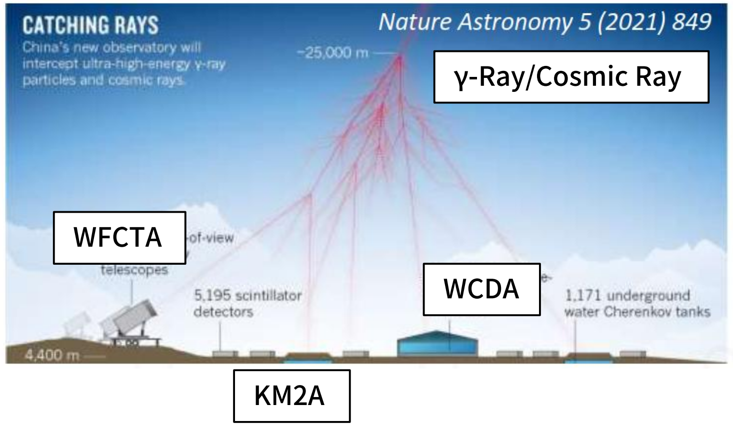
Background：Why AI ?



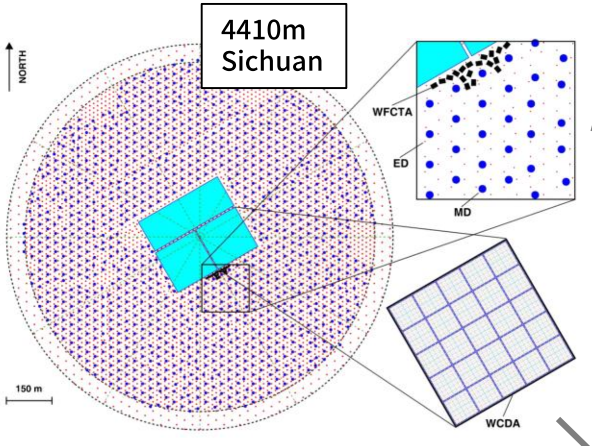
1980~：The party is over ?

2020~2030：
High Energy + AI ⇒ New party ?

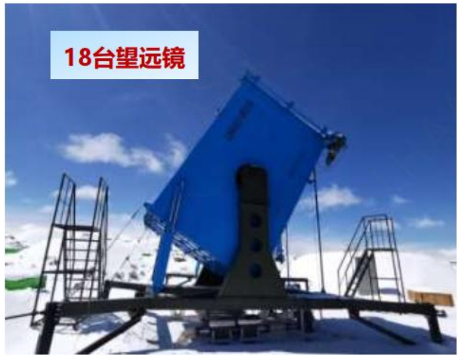
Background: LHAASO Experiment



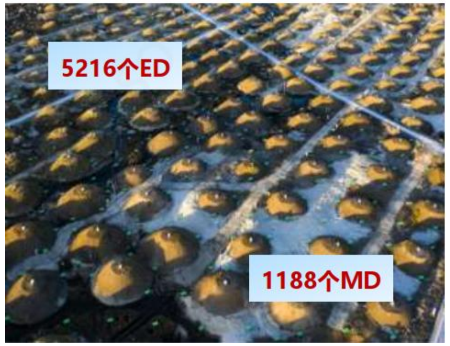
Large High Altitude Air Shower Observatory (LHAASO)

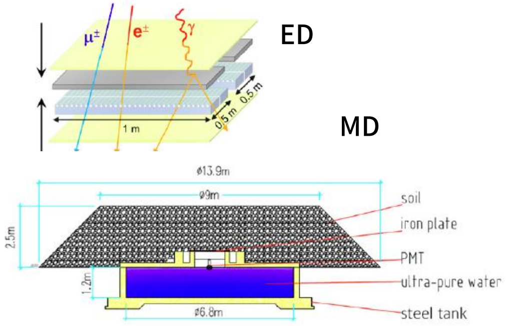


Wide Field-of-view Cherenkov Telescope Array (WFCTA)
⇒ 18 Telescopes

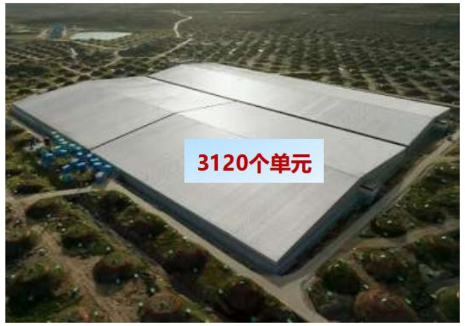


KiloMeter-square Array (KM2A)
⇒ 5216 Electron Detectors + 1188 Muon Detectors





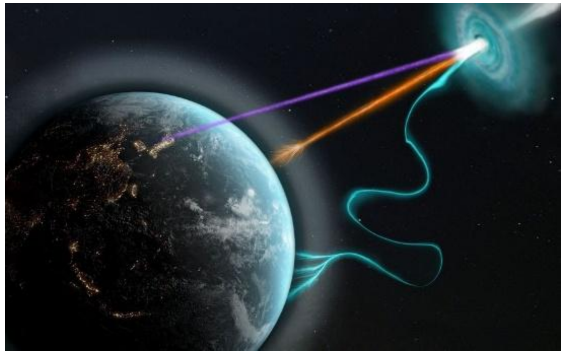
Water Cherenkov Detector Array (WCDA)
⇒ 3120 Detector Units



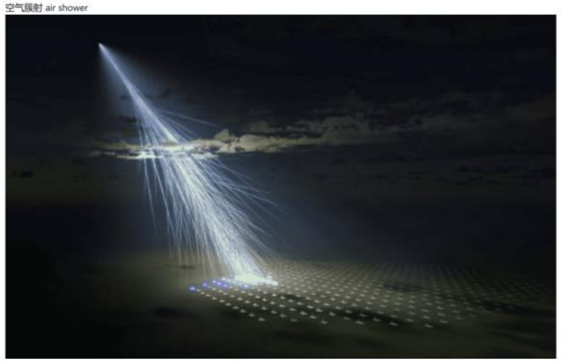
Background：High-Energy Astronomy

Source：SNR、PWN、AGN、BNS merge …

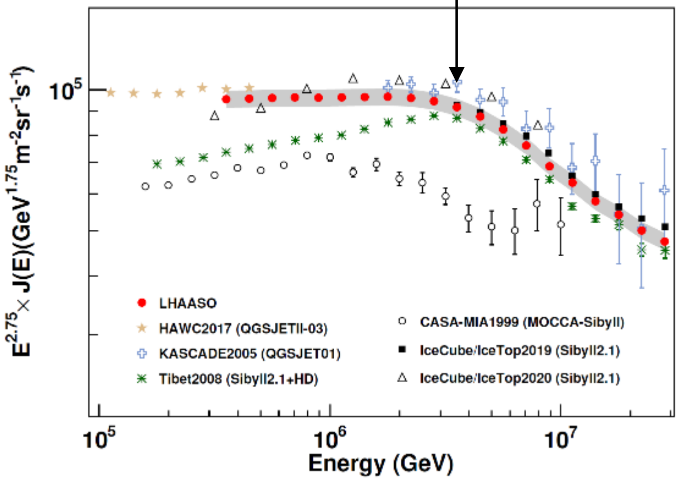
High-energy γ-rays
Atomic Nuclei
Neutrino
X-rays
Radio
…

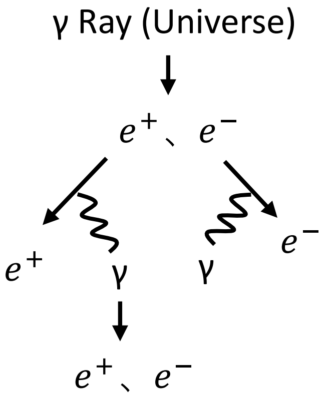


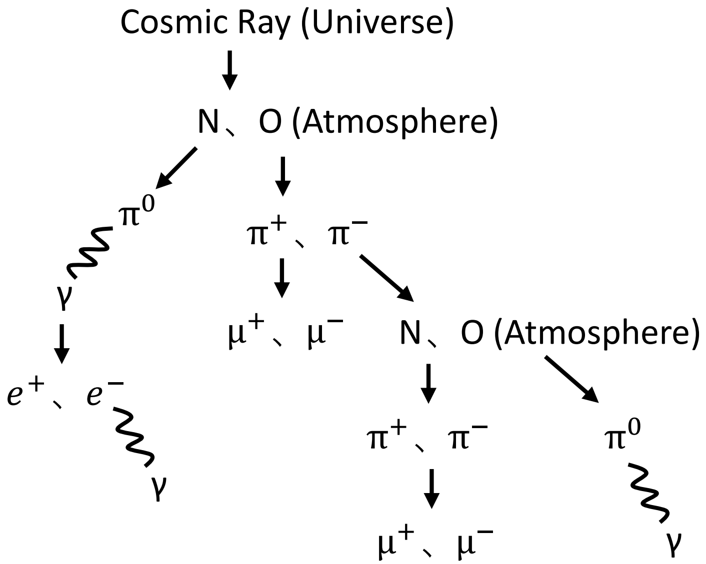
Extensive Air Shower



Knee in Energy Spectrum of Comic Ray [1]



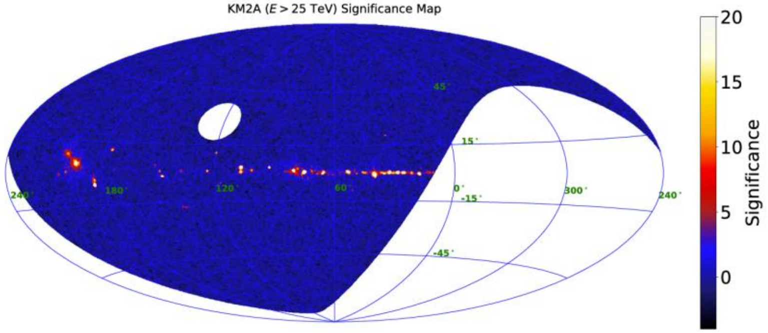




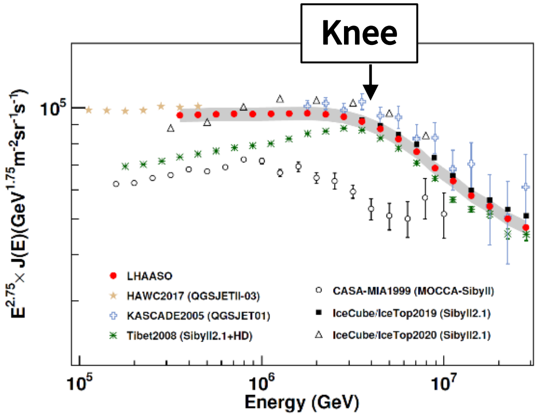
[1] Z. Cao, et al. (LHAASO Collaboration), Phys. Rev. Lett. 132, 131002 (2024)

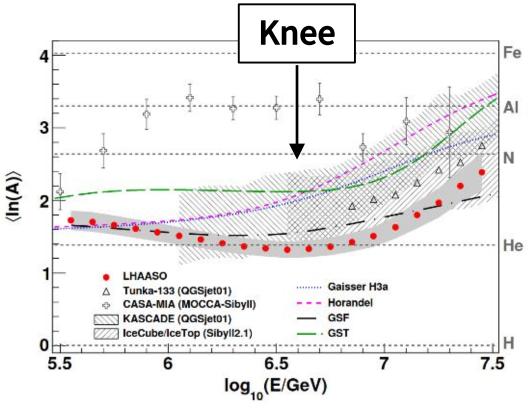
Background：What Can AI Do for LHAASO ?

Gamma-Ray Sources found by LHAASO [1]
⇒ Discover more sources with AI



Energy Spectrum/Mass of Comic Ray in the "Knee" [2]
⇒ More precise measurement with AI





Q：What is the Root Problem ?
A：Origin and acceleration mechanisms of High-Energy CR.
Q：What need to do?
A：More and more Precise γ-Ray Source Observations,
More accurate CR spectrum in 300 TeV~30 PeV.

This Report

First Step：Enhance Foundation Performance
1、AI for γ/CR Discrimination
2、AI for CR Composition
3、AI for Energy Reconstruction
4、AI for Directional Reconstruction

Future Step（? ）：
1、LLM Agent/AI for γ -Ray Source Analysis ?
2、Multi-wavelength Observations ?
⇒ Pulsar、SNR…

[1] Z. Cao, et al. (LHAASO Collaboration), The Astrophysical Journal Supplement Series, 271:25 (26pp), 2024
[2] Z. Cao, et al. (LHAASO Collaboration), Phys. Rev. Lett. 132, 131002 (2024)

目录

Content

AI for LHAASO：Main Participates + Current Status

Main Participates

	WCDA	KM2A	WFCTA
Event Classification	● ◆ ★ ▲ ⬟	◆ ★ ▲	★
Energy Reconstruction	▲	◆ ★ ▲	
Directional Reconstruction	▲ ⬟	◆ ★ ▲	
Unsupervised Pretraining	●		
Real Data		◆ ☗	
Simulation Data	■	◆ ☗	✶
Platform & Visualization	◀		

Status of AI Model

	WCDA	KM2A	WFCTA
Event Classification	○	○	○
Energy Reconstruction		△	
Directional Reconstruction		△	

中科院高能物理所
* 白云翔、姚志国
● 阮曼奇、程耀东、朱永峰
■ 姚志国、胡世聪、黄勇
◆ 吕洪魁、张笑鹏、谢天、黄志豪
☗ 李哲、刘祥廷
◀ 张易于、张航畅、席绍强
✶ 尹丽巧

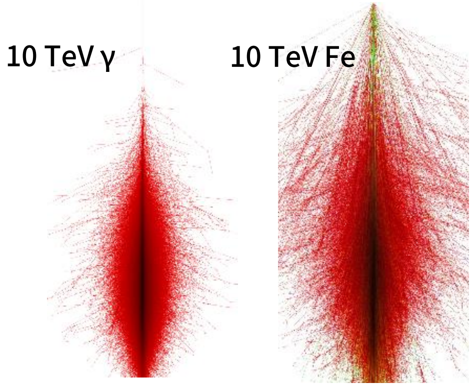
之江实验室
★ 陈华曦、徐宸原、王智、刘欢迎、刘畅

中科院自动化所
▲ 朱炳科、王潇潇、史千秋、杨辰

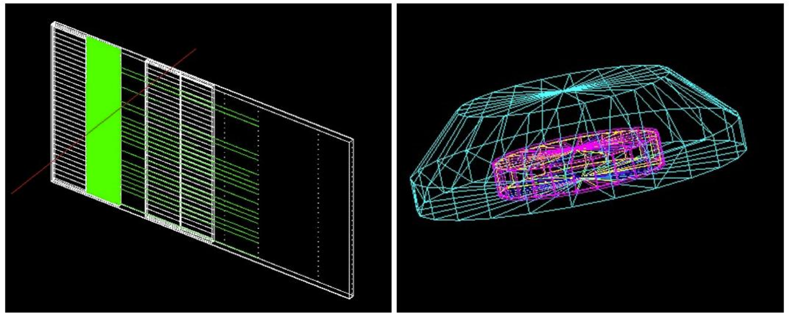
上海交通大学
⬟ 周浩、唐睿仪

AI for LHAASO: Simulation Data

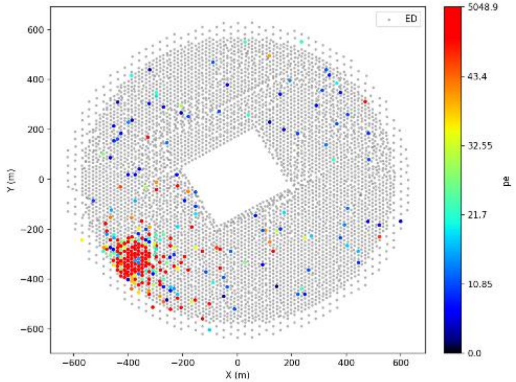
Corsika for extensive air showers [1]

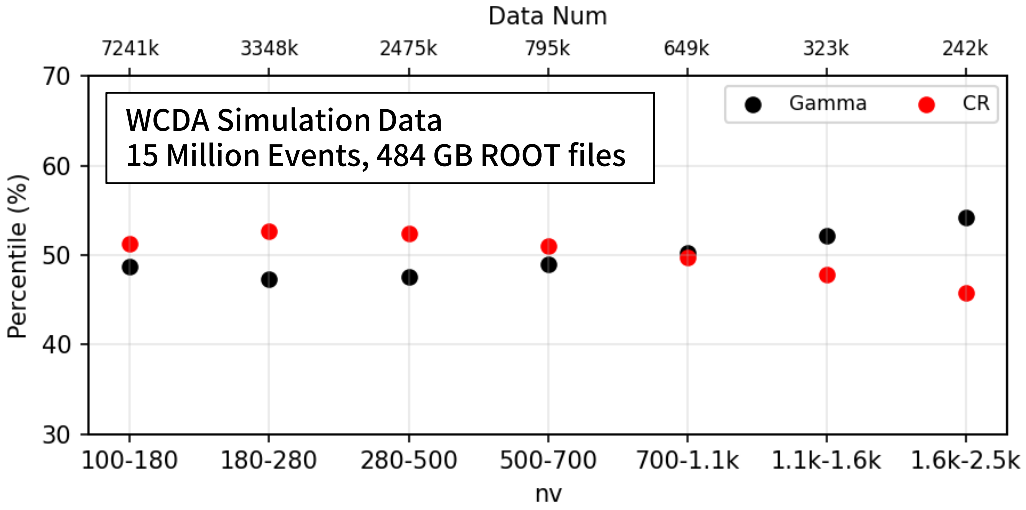


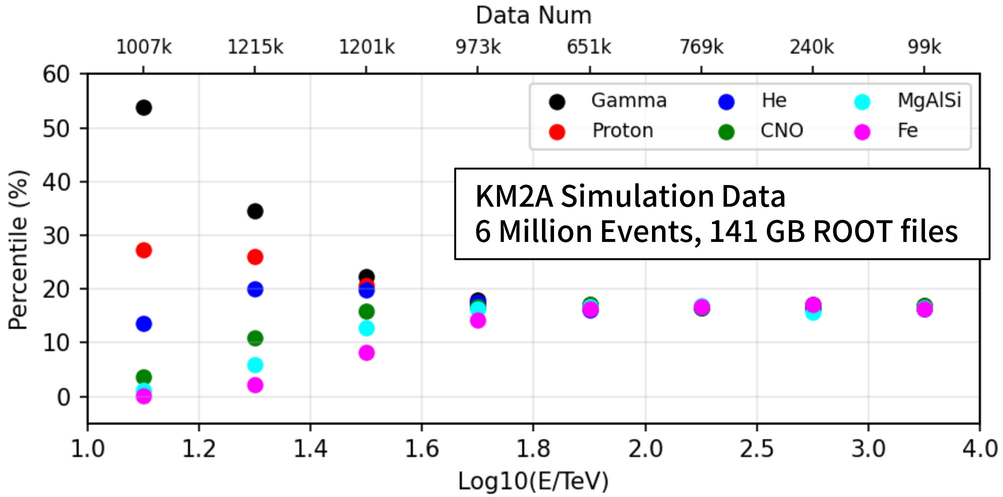
Geant4-based Detector Simulation [2]



LHAASO Event





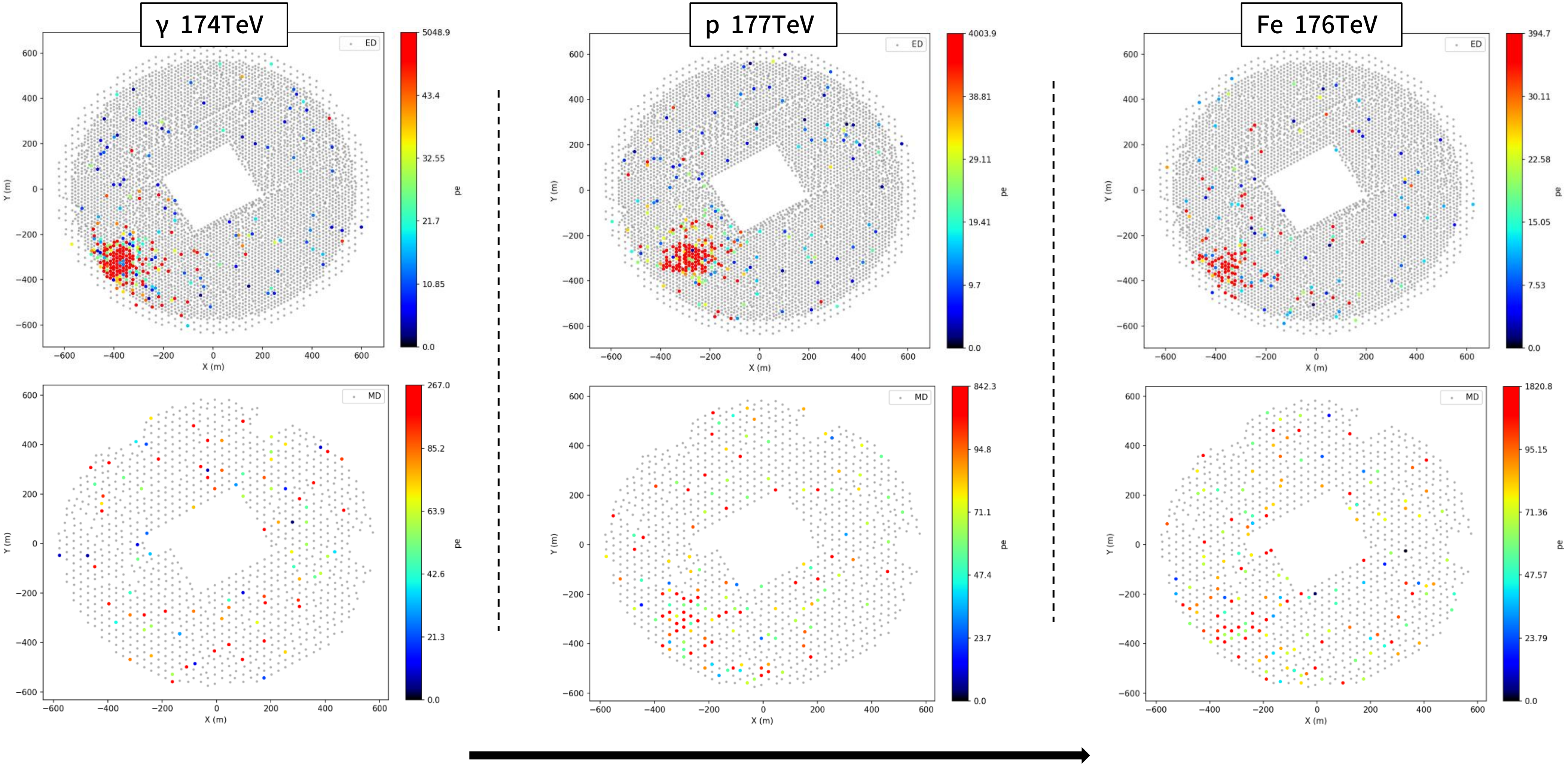


[1] D. Heck, J. Knapp, J.N. Capdevielle, G. Schatz, and T.Thouw, Report FZKA 6019 (1998), Forschungszentrum Karlsruhe, https://www.iap.kit.edu/corsika/70.php
[2] S. Chen, J. Zhao, Z. Li, et al., Detector simulation of LHAASO-KM2A with Geant4, ICRC 2019

目录

Content

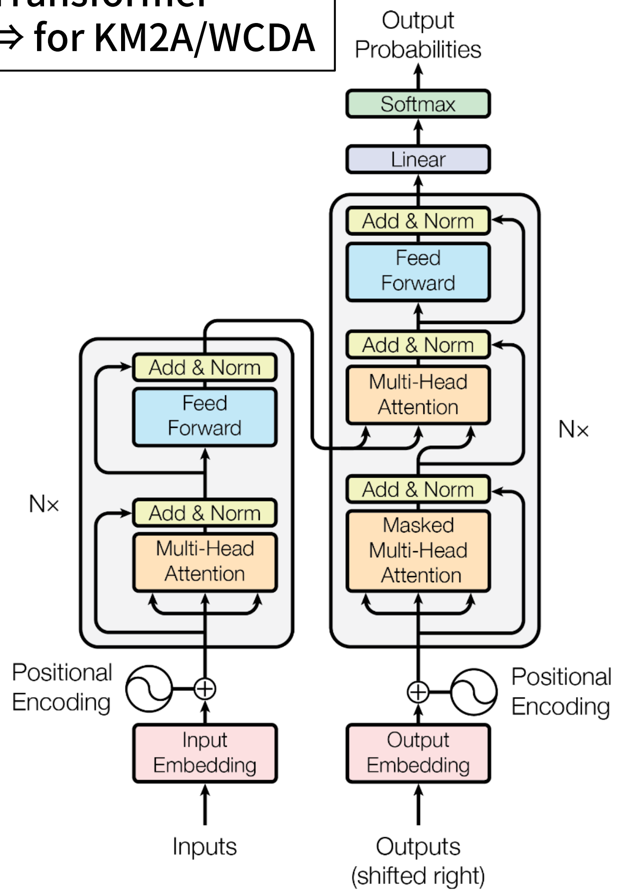
KM2A：Understanding the Data



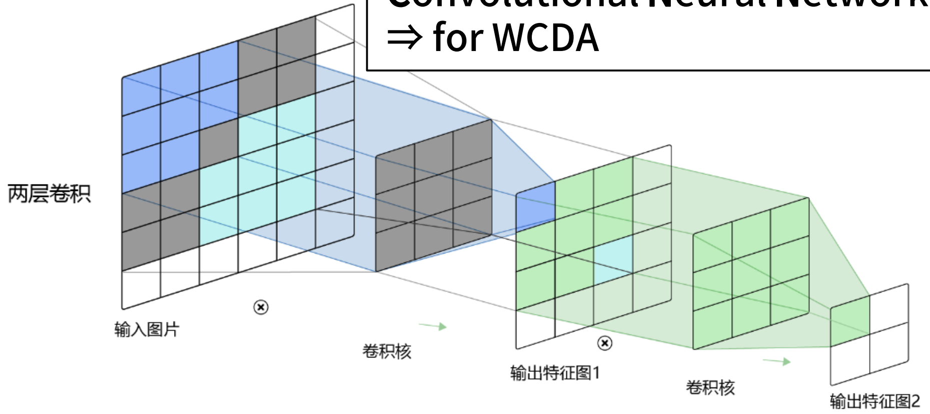
Heavier nuclei release more energy via muons (measured by MD)

AI for LHAASO：Potential Models

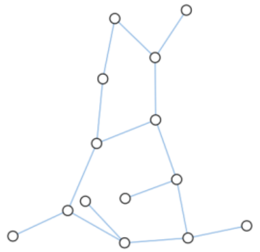
Transformer
⇒ for KM2A/WCDA



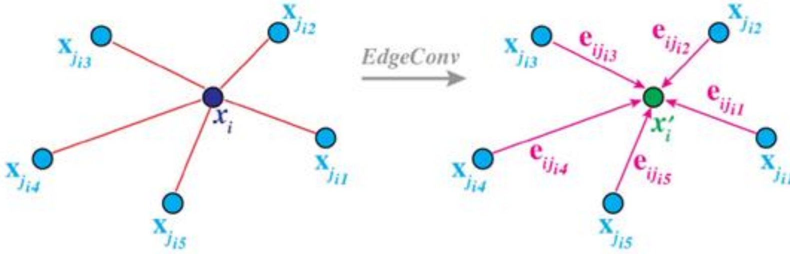
Convolutional Neural Network
⇒ for WCDA



Graph Neural Network



DGCNN / ParticleNet [1]
⇒ for KM2A/WCDA

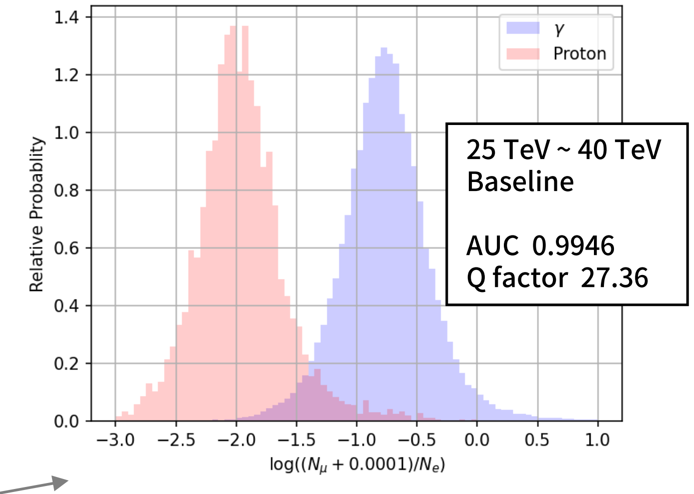


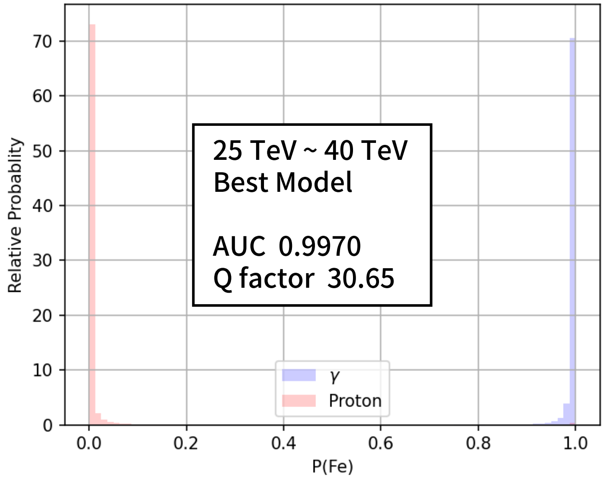
[1] H. Qu and L. Gouskos, Jet tagging via particle clouds, Phys. Rev. D 101, 056019 (2020)

KM2A：AI for γ/p Separation

Preliminary

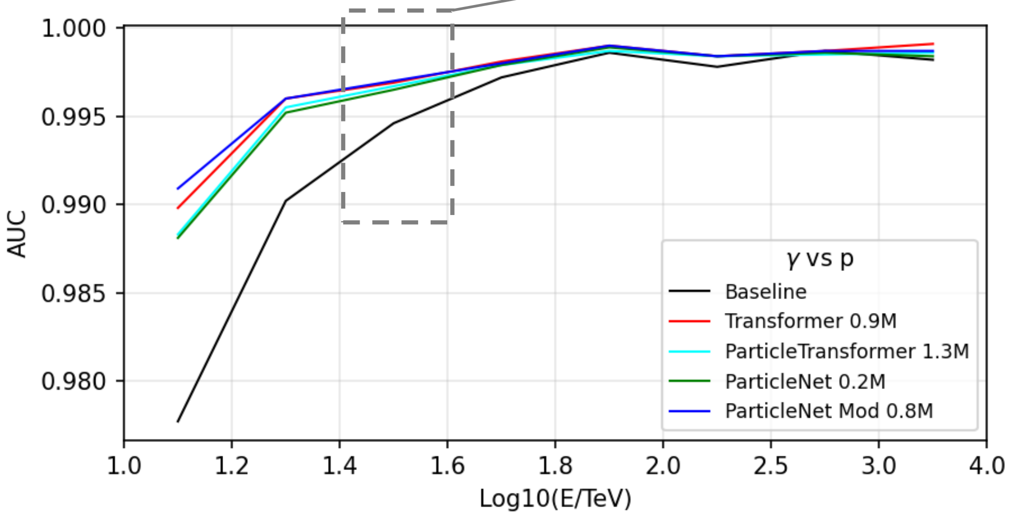
Event Selection
Core∈(200m, 500m)，Zenith<35°
Feature：Q, t, u, v, w, MD/ED
Baseline：$(N_\mu + 0.0001)/N_e$ μ/e counts in 40m



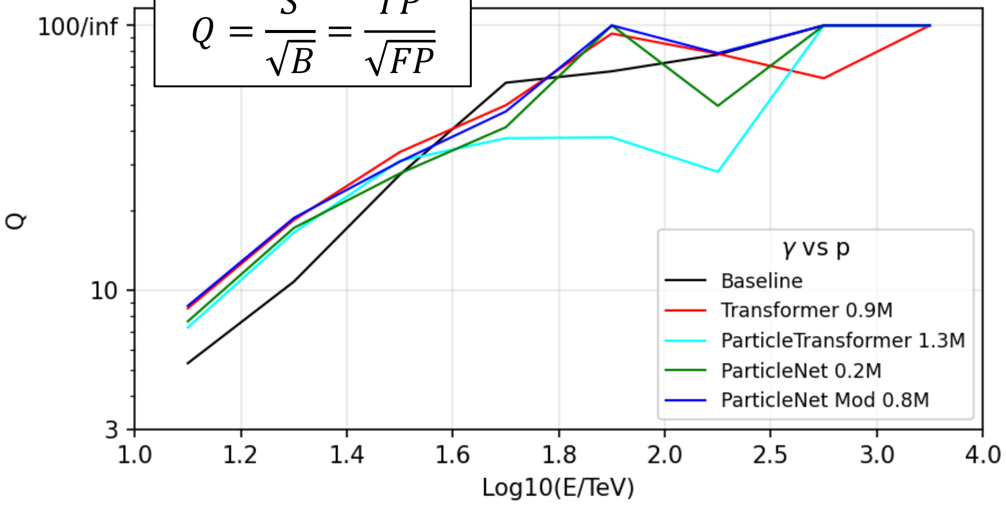


Best Model Boosts AUC & Q factor Under 100 TeV

***ParticleNet Mod：Only ED-MD edges**



$$Q = \frac{S}{\sqrt{B}} = \frac{TP}{\sqrt{FP}}$$



KM2A：AI for γ/p Separation

Preliminary

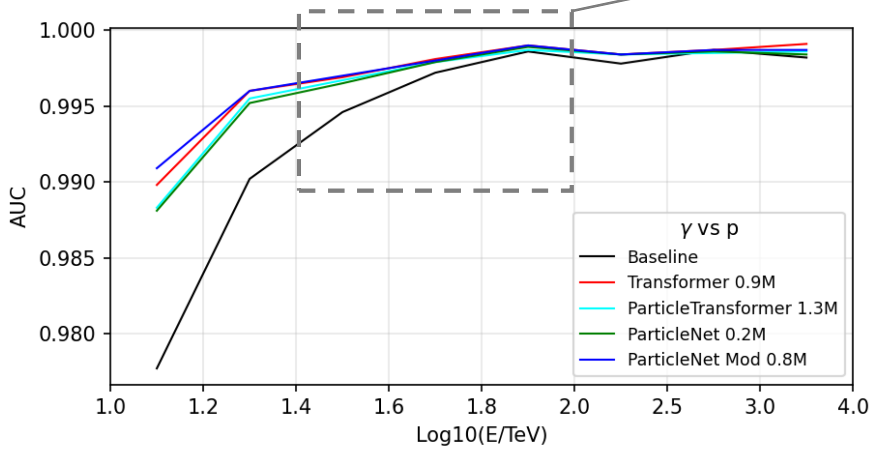
Real Data Validation:

- Selected events within ±7.5° of Crab Nebula (standard candle) + off data
- Used quality-filtered KM2A data from Jan 15, 2025

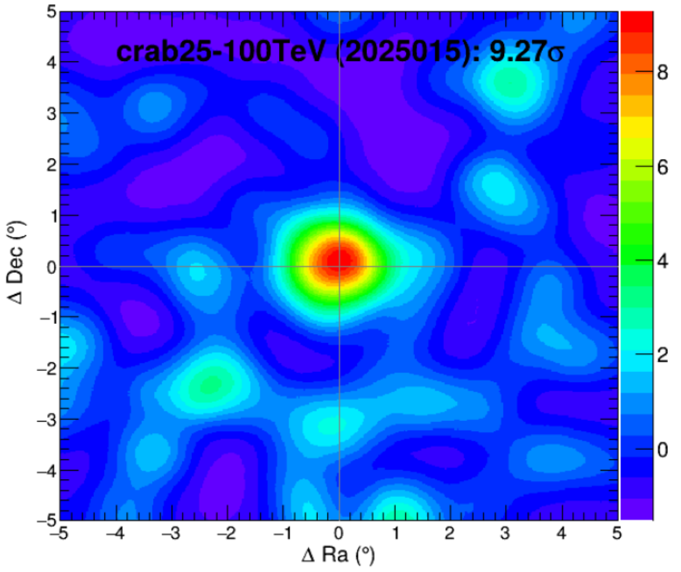
- Future optimization with expanded data collection planned
- Better performance can be expected for sources below 25 TeV

Jul 29 中科院高能物理所 张笑鹏
Progress in KM2A Background Rejection and Directional Reconstruction

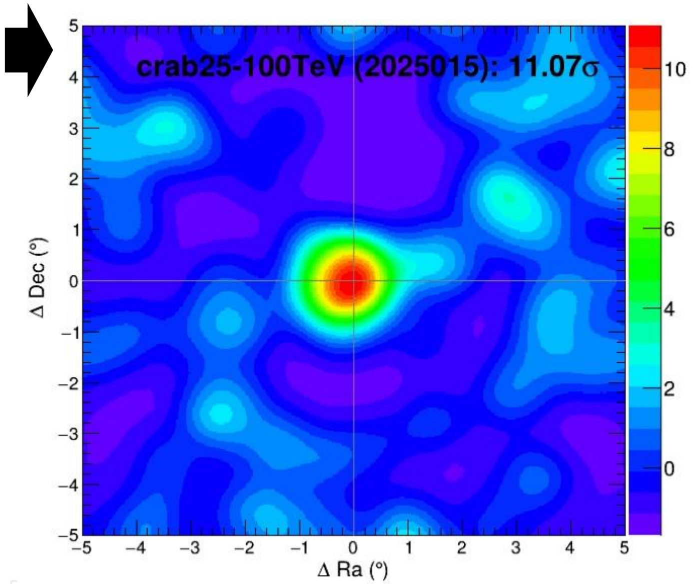
AI Model Based on Simulation Data



25 TeV ~ 100 TeV
Baseline 9.27σ



25 TeV ~ 100 TeV
AI Model 11.07σ

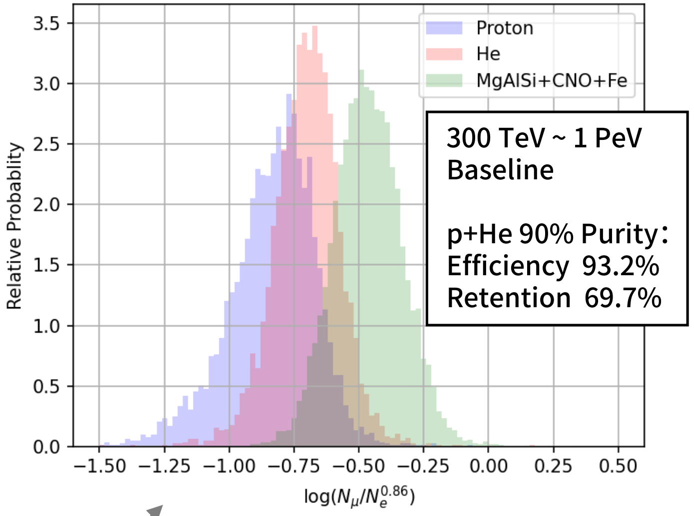


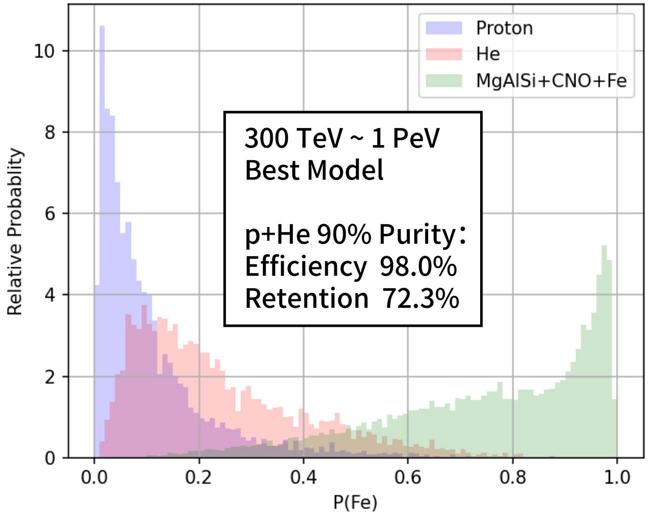
KM2A：AI for Cosmic Ray Separation

Preliminary

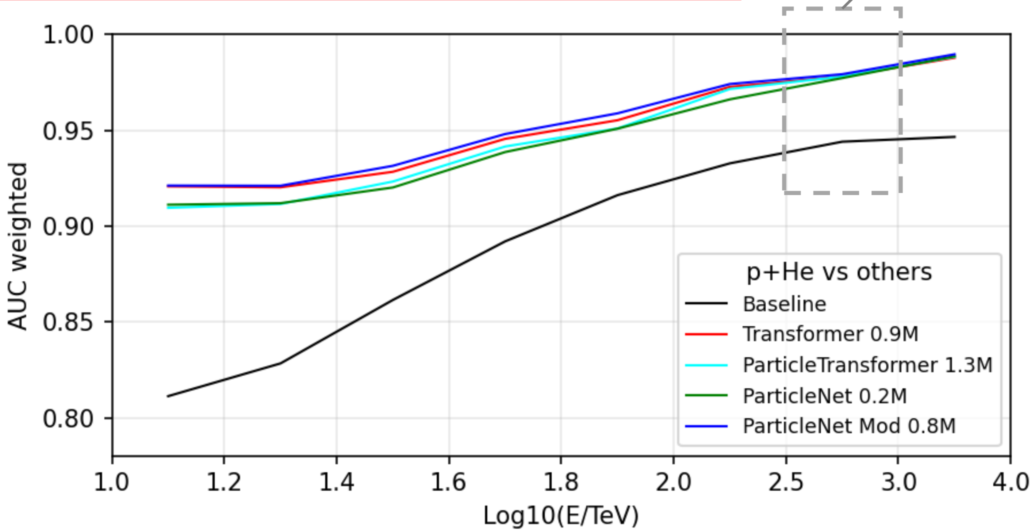
Event Selection
Core∈(200m, 500m)，Zenith<35°
Feature：Q, t, u, v, w, MD/ED
CR Weight：Horrandol
Baseline：$N_\mu/N_e^{0.86}$ μ/e counts in 40m

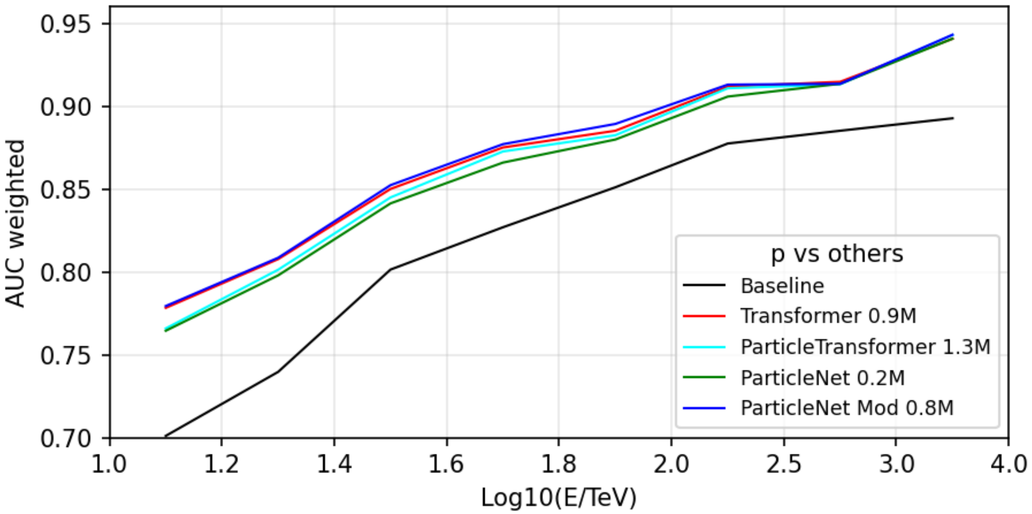
$$purity\ (纯度) = N_{select\_light}/N_{select}$$
$$efficiency\ (筛选效率) = N_{select\_light}/N_{light\_total}$$
$$retention(保留比) = N_{select}/N_{total}$$





Best Model Boosts AUC, Efficiency & Retention Ratio Across Full-Energy Spectrum.
***ParticleNet Mod：Only ED-MD edges**





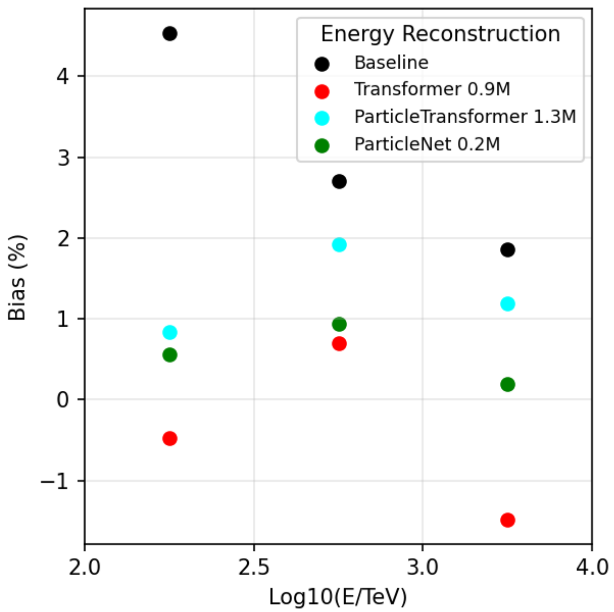
KM2A: AI for CR Energy Reconstruction

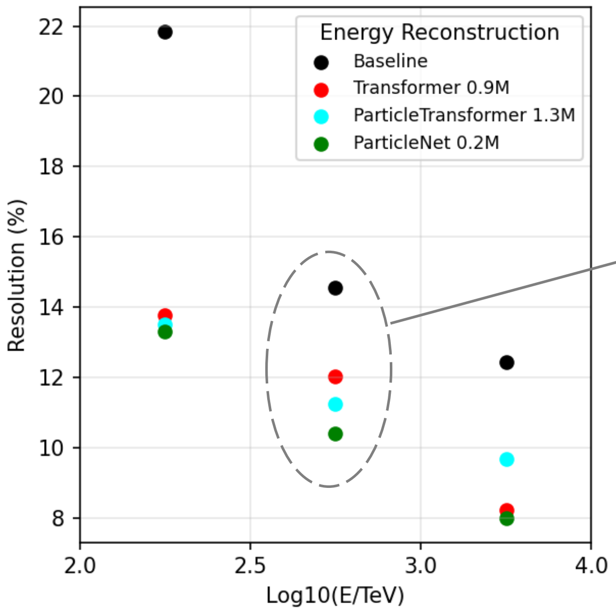
Preliminary

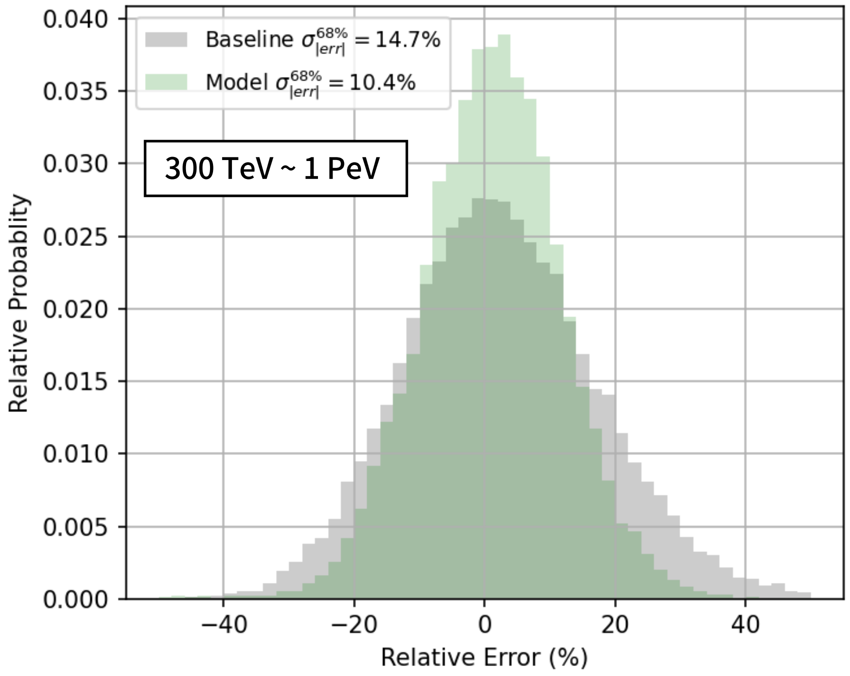
Event Selection
Core∈(200m, 500m), Zenith<35°
Feature: Q, t, u, v, w, MD/ED
Baseline [1]: μ/e counts in (40m, 200m)

$$N_{e\mu} = \frac{N_e}{A_{\rm ED}} + \frac{25 \cdot N_\mu}{A_{\rm MD}}. \qquad E_{\rm rec} = bN_{e\mu},$$

Best model improves CR energy resolution by >25% while controlling bias to <2%







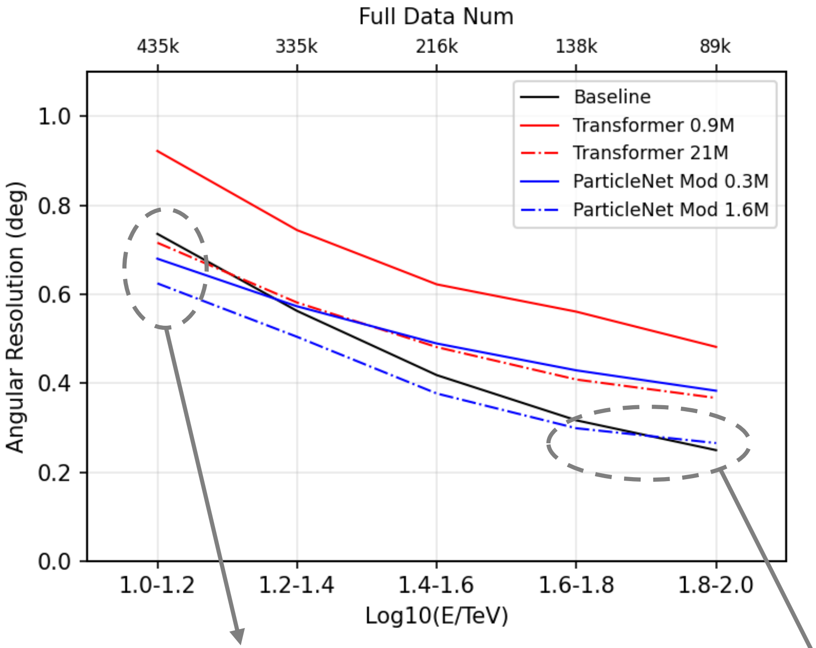
[1] H. Zhang, H. He and C. Feng, Phys. Rev. D 106, 123028 (2022)

KM2A: AI for γ Direction Reconstruction

Preliminary

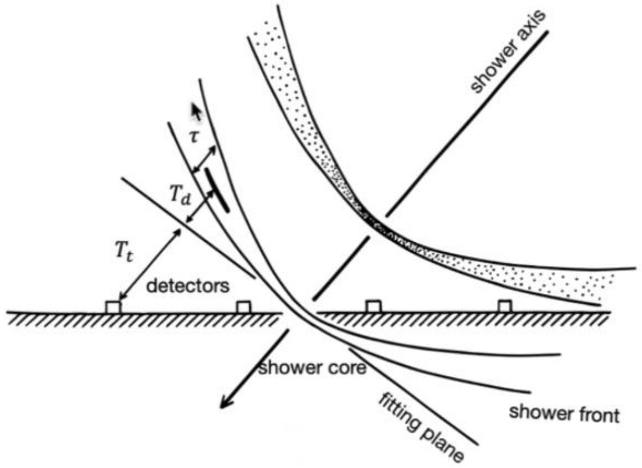
Best Model improves angular resolution by 10–20% at <60 TeV

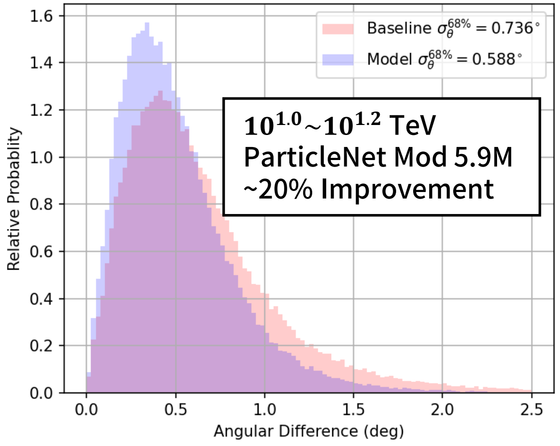
***ParticleNet Mod:**
1、only ED nodes
2、BN+ReLU ⇒ LN+GELU

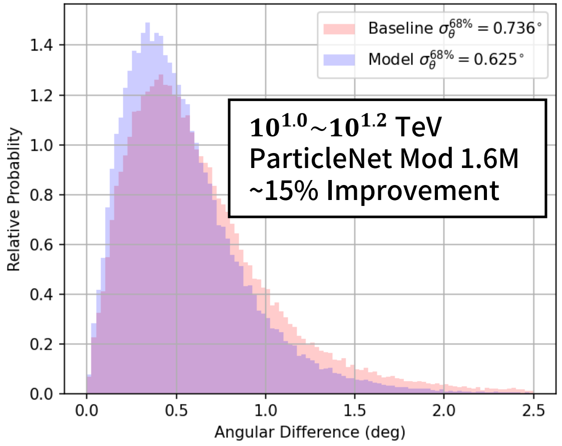


Event Selection
Core∈(200m, 500m), Zenith<35°
Feature: Q, t, x, y, z

Baseline:
Least-squares fitting of the shower front profile



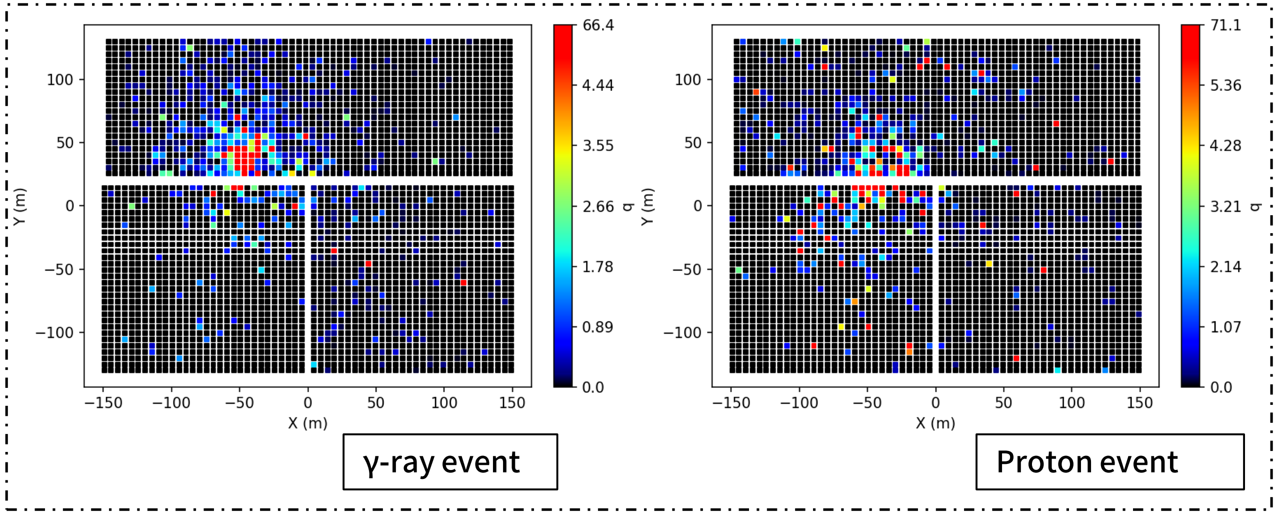


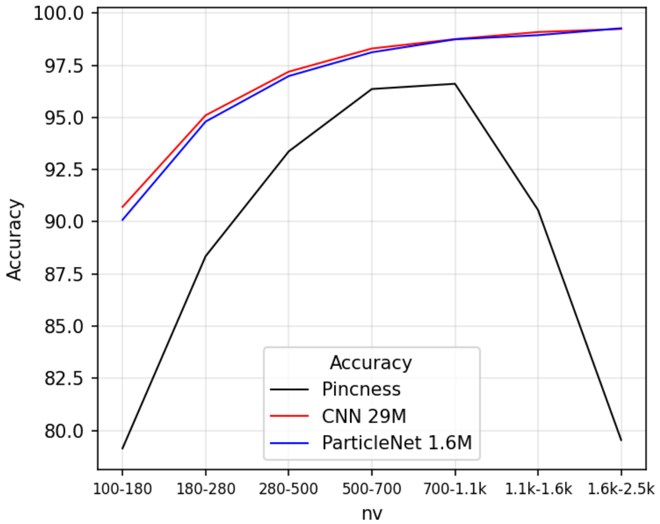


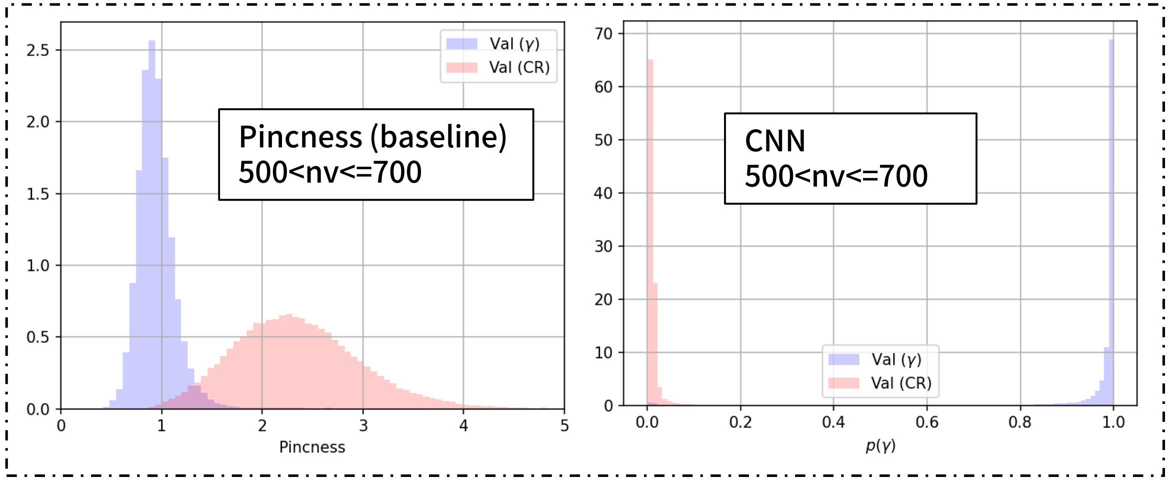
Plan to simulate more data for model training

WCDA: AI for γ/p Separation

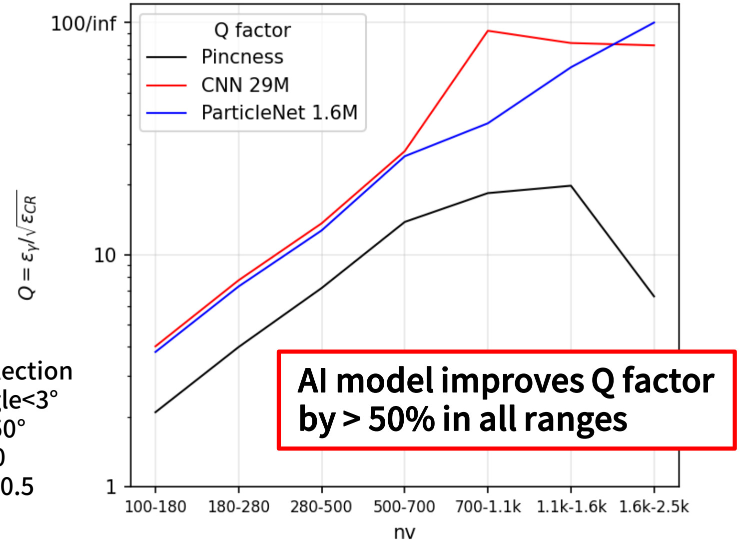
Preliminary







Event Selection
mc_dangle<3°
Zenith <50°
fitstat==0
vqsamp>0.5



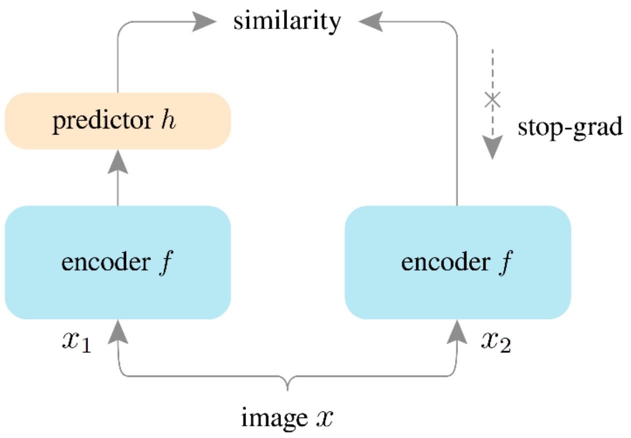
WCDA：AI for γ/p Separation

Preliminary

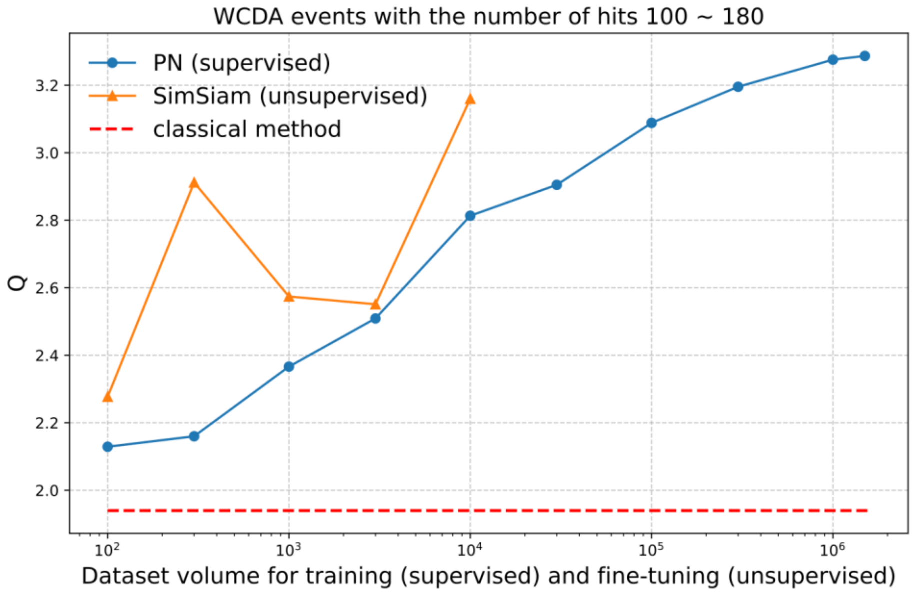
Jul 29 中科院高能物理所 朱永峰
Unsupervised Learning in γ/p classification

Self-Supervised SimSiam [1]:

After augmenting sample x_1 to x_2, both pass through a shared encoder.
A predictor then pulls close feature similarity between original and augmented instances.



Pre-training encoders with 3×10^6 unlabeled samples, then fine-tuning with sparse labels ⇒ outperforms supervised-only training.



[1] X. Chen and K. He, "Exploring Simple Siamese Representation Learning," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

目录

Content

Summary & Future

Good News:

- KM2A
  - ⇒ For background rejection, AI improves AUC at <100 TeV, and improves Q factor by 50% at <25 TeV
    - ⇒ Slight significance boost confirmed by Crab Nebula observations
  - ⇒ For CR composition, AI outperform baseline in all ranges
    - ⇒ In Pre-knee region, AUC up from 0.944 to 0.979, signal retention up from 69.7% to 72.3% with 90% p+He purity
  - ⇒ For CR energy reconstruction, AI improves resolution by ~25% in Knee and Pre-knee region
  - ⇒ For γ directional reconstruction, AI improves angular resolution by 10~20% at <60 TeV
- WCDA
  - ⇒ For background rejection, AI boosts Q factor by >50% in all ranges

Challenge:

- Real-data AI validation requires further testing and optimization

What we learned/found:

- Model Architecture > Parameter Scaling
  - ⇒ Point cloud modeling aligns best with physics
- Unsupervised Pretraining leverages unlabeled data effectively

Near Future:

- Crab Nebula validation/fine-tuning
- Joint reconstruction of KM2A+WFCTA / KM2A+WCDA ?

Far Future:

- LLM Agents for γ-source analysis

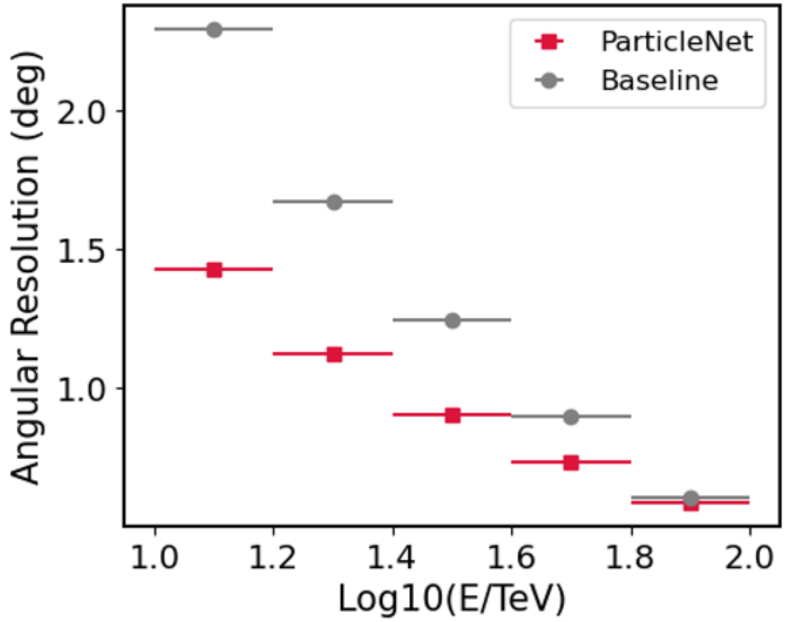
Q: Other ways to Integrate physics into model architecture for reconstruction ?

Welcome to Cooperate!
Welcome to Communicate!

Backup

KM2A：AI for Direction Reconstruction

Preliminary

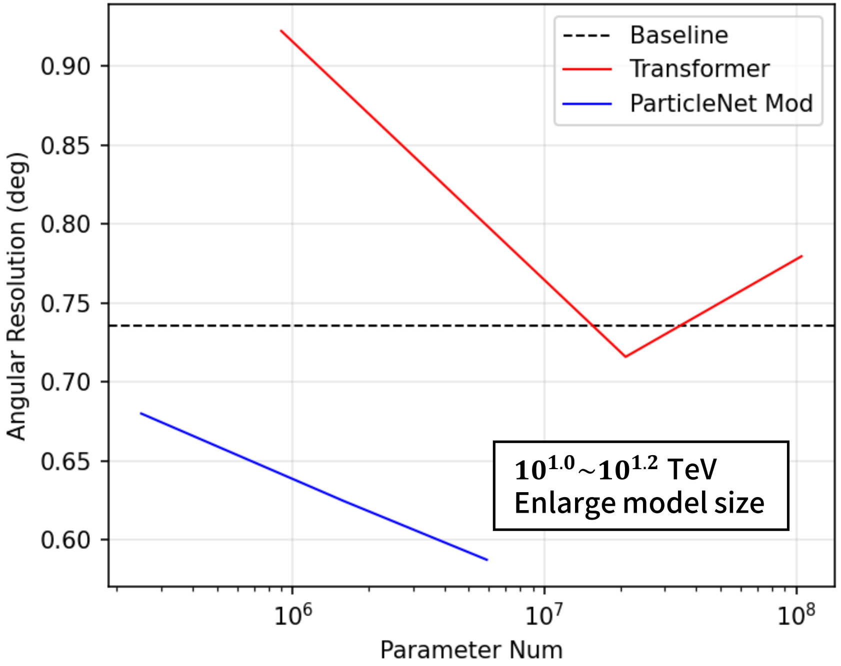


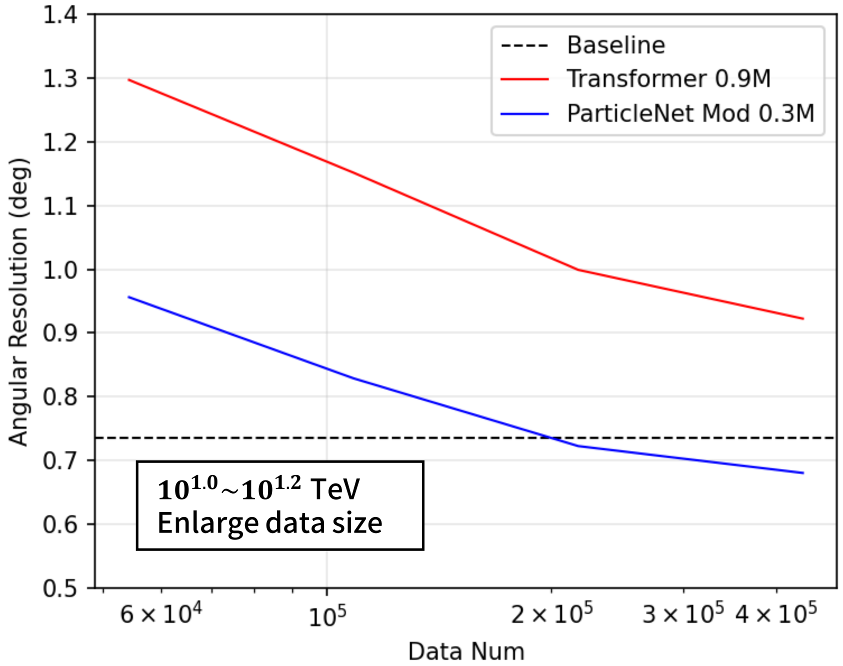
7月29日 中科院高能物理所 张笑鹏
KM2A背景排除和方向重建进展

铁核事件方向分辨率相对传统方法有较大提高
有望结合粒子分类模型，利用月影偏转进行实验验证

KM2A：Scaling Law in Direction Reconstruction

Preliminary

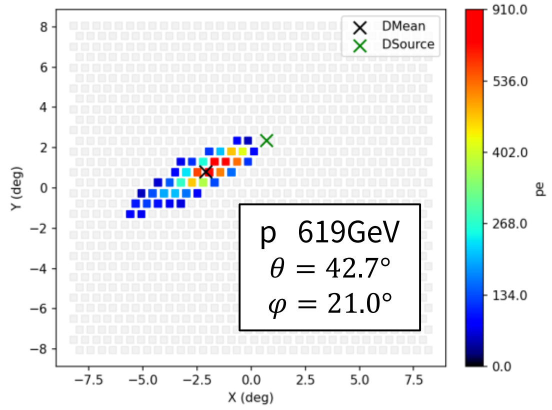


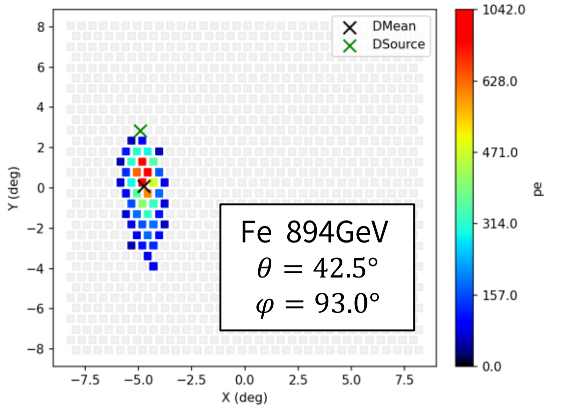


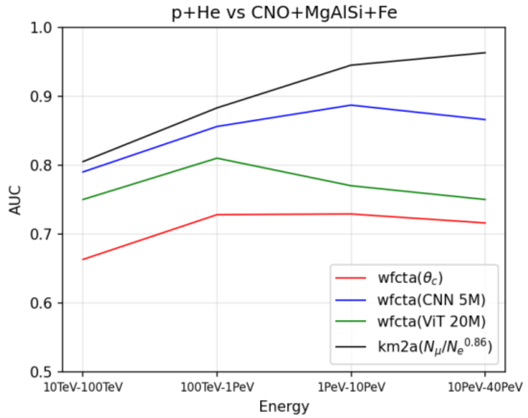
1、正确的模型结构比单纯增加模型尺寸更有效
2、数据量非常重要

WFCTA：AI for Cosmic Ray Separation

Preliminary







肉眼已无法辨别WFCTA的p/Fe

只靠WFCTA信息有多少效果?