PID studies with AHCAL on CERN SPS beam purity

Xin Xia
Institute of high energy physics
On behalf of the CEPC calorimeter working group

CALICE Collaboration meeting, Sep. 28 2023

Outline

- **Introduction**
- **Cut-based PID**
- **ML-based PID**
  - **BDT**
  - **ANN**
- **Comparison and validation**
- **Summary**

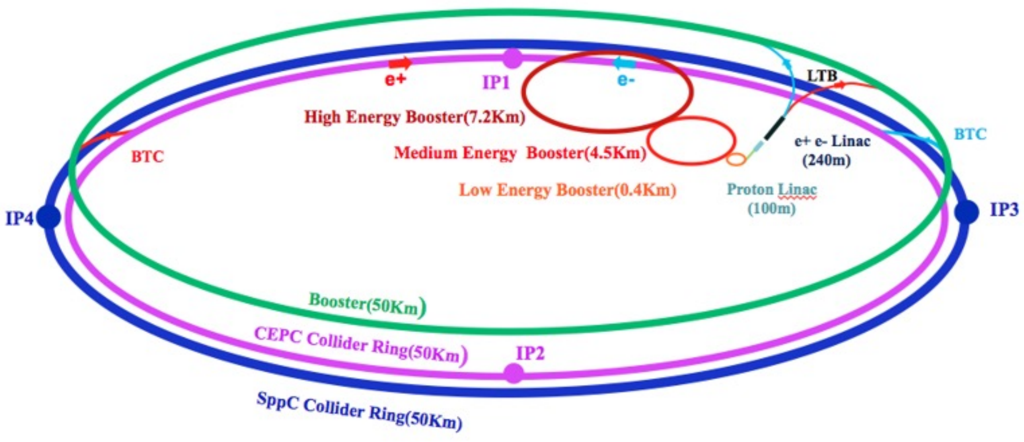
Introduction

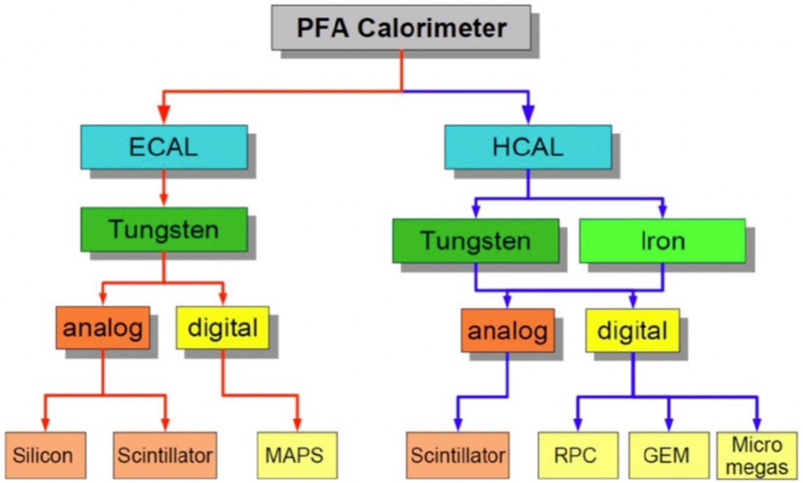
➢ **CEPC**

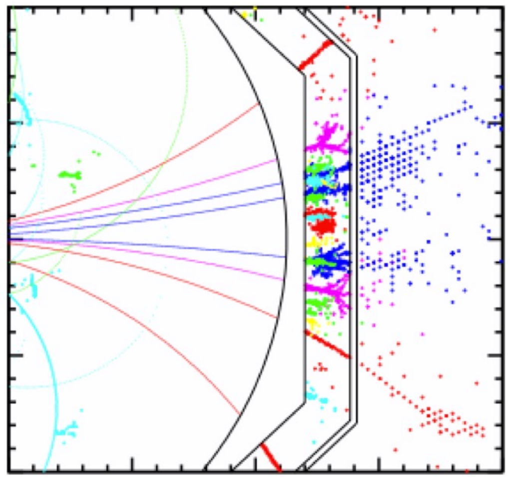
- Future electron-positron collider
- Physics goal: Precise measurements of the Higgs/EW/QCD
- Challenge: Jet resolution ~ $30\%/\sqrt{E}$, BMR ~ 3-4%

➢ **PFA-oriented calorimeter: various options explored in the CALICE collaboration**

- Focus in this talk: **AHCAL with high granularity**

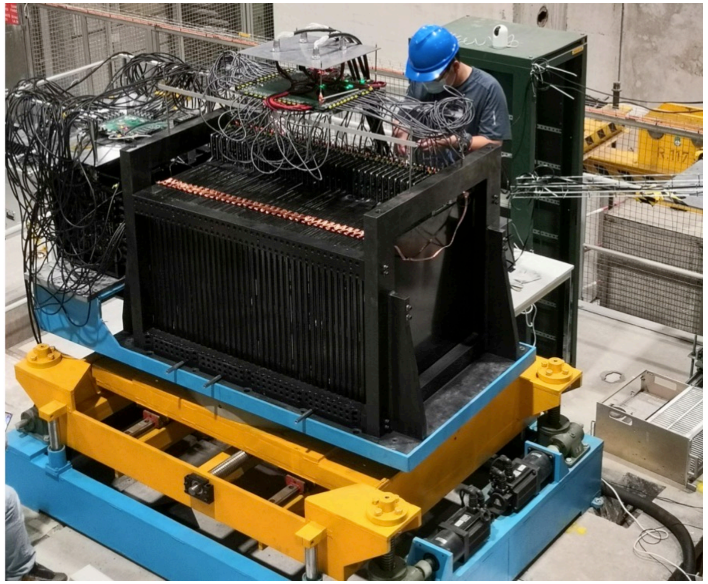


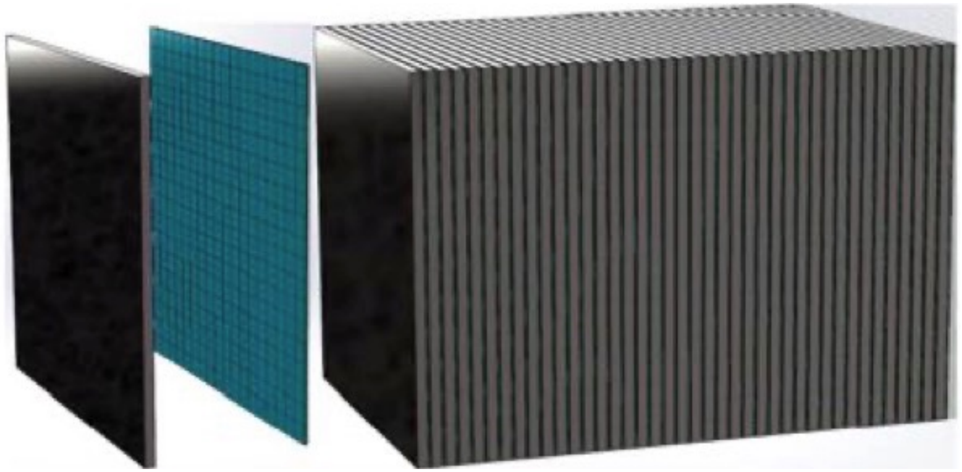


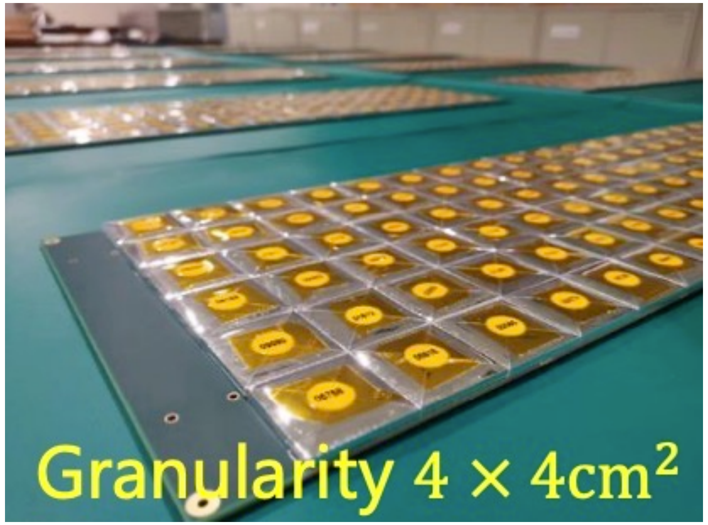


Introduction

- **CEPC AHCAL prototype configuration**
  - Active material: scintillator tiles + SiPM
  - Absorber: stainless steel
  - Transverse size 72×72 cm^2, 40 longitudinal layers (~4.6λ_0)
  - Granularity $4 \times 4\ \mathrm{cm}^2$, 12960 readout channels
  - ~5 ton in weight
  - Developed during 2018 – 2022







Introduction

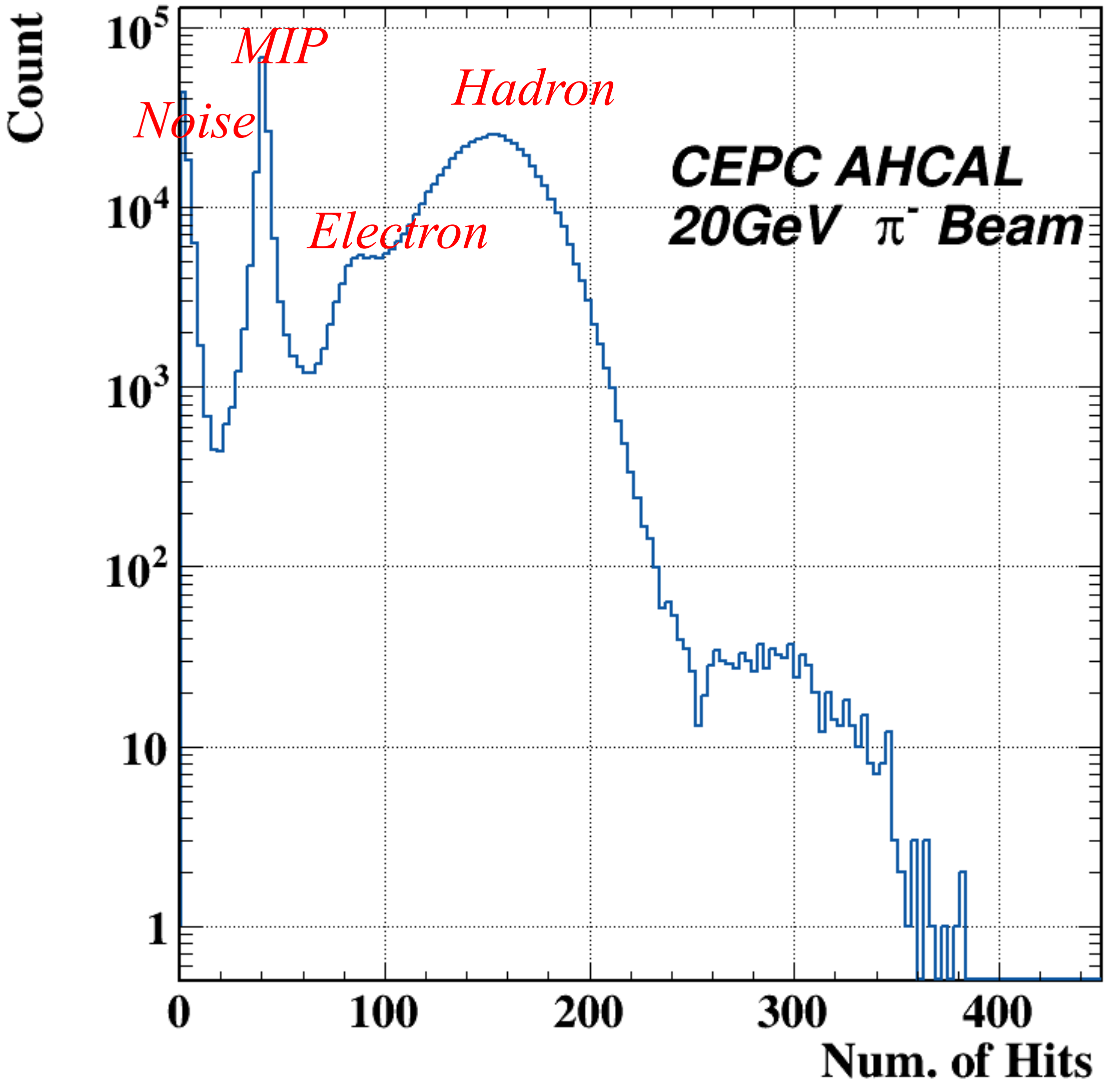
- **Beam test at CERN at SPS H2 beamline at 2023/05**
  - ~ 11 billion pion events from 10GeV to 120GeV
  - ~ 4 billion electron events from 10GeV to 250GeV

AHCAL Data List of Pion		
Momentum (GeV)	Total Runs	Number of Events
10	6	1027814
15	4	953774
20	4	749791
30	4	1081725
40	5	1038984
50	5	1258699
60	6	1034938
70	5	1058822
80	7	1073061
100	7	1286608
120	6	1036400

AHCAL Data List of Electron		
Momentum (GeV)	Total Runs	Number of Events
10	2	238139
20	1	573677
30	2	556570
40	1	421484
50	2	743810
60	1	377014
70	3	580935
80	1	273278
100	1	289511
120	1	242049
150	2	415991
250	1	121850

Introduction

- **The data contains:**
  - Noise, MIP, Electron, Hadron
  - Multiple injection events
  - …
- **In order to:**
  - Fully understand data components
  - Further quantify the detector response
- **PID method developed:**
  - Cut-based PID
  - BDT
  - ANN



Cut-based PID: Variables

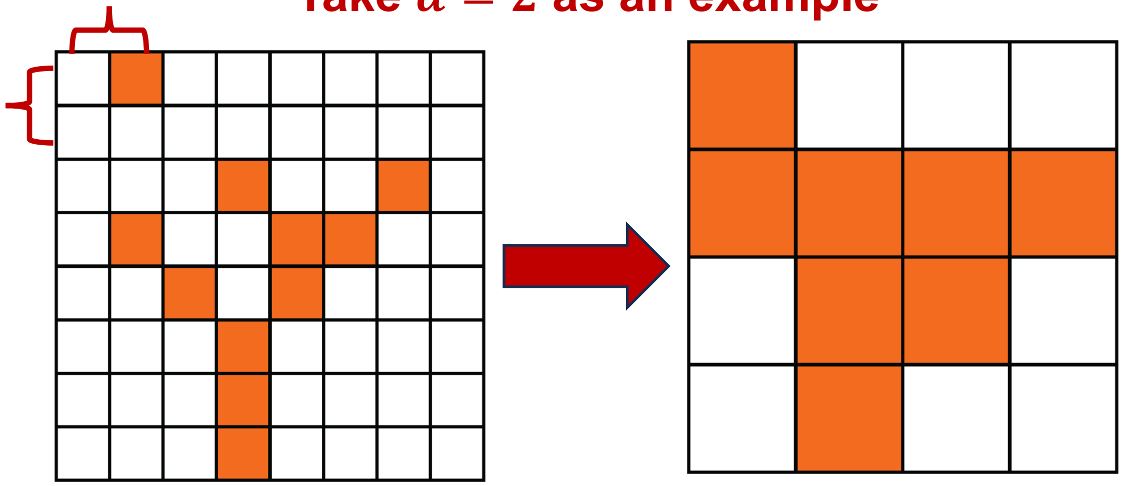
- **Fractal dimension**[1] **:**
  - Definition : $FD_\alpha = \left\langle \frac{\log(R_\alpha)}{\log(\alpha)} \right\rangle$, $R_\alpha = \frac{N}{N_\alpha}$,

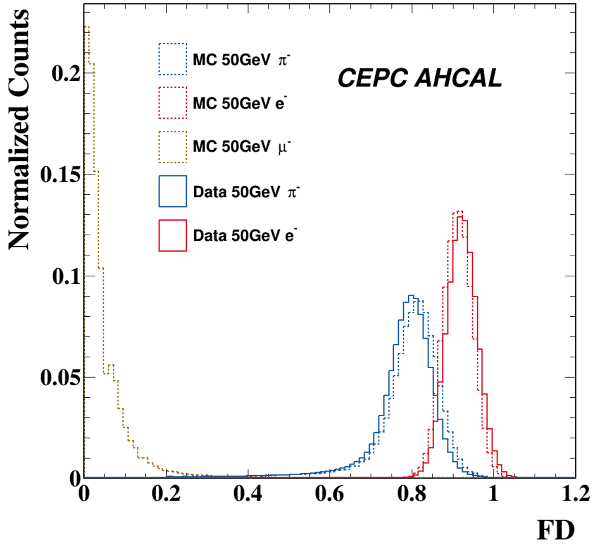
    $$\mathrm{FD} = 0.1 \cdot \sum_\alpha FD_\alpha,$$

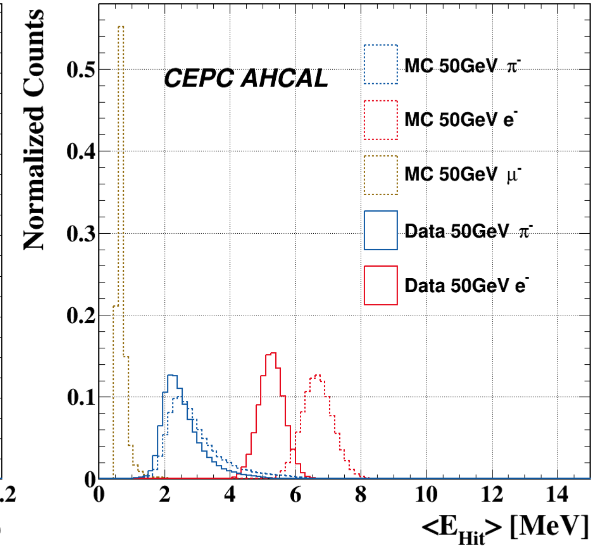
    $$\alpha = 2, 3, 4, 5, 6, 7, 8, 9, 10, 20$$

    - α : the scale at which the shower is analyzed ,
    - N : total number of hits
    - N_α : the number of hits at scale α
  - Dedicated for high granularity calorimeter
- **Average Hit Energy** : $< E_{Hit} >$
  - Definition : $< E_{Hit} > = {}^{Event\ Energy}/_{Num.\ of\ Hits}$
  - Progress in MC/Data consistency

Take $\alpha = 2$ as an example







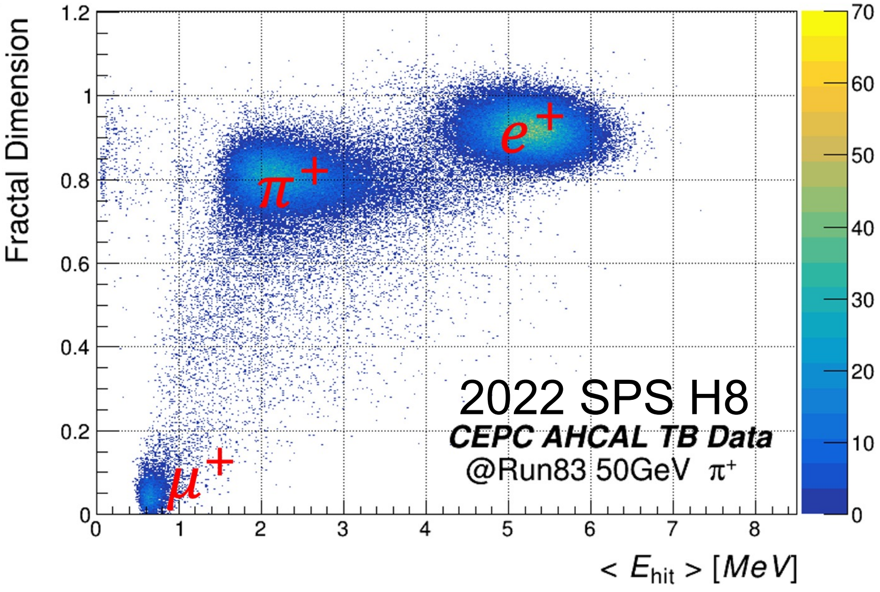
[1] M. Ruan et al, PRL 112, 012001

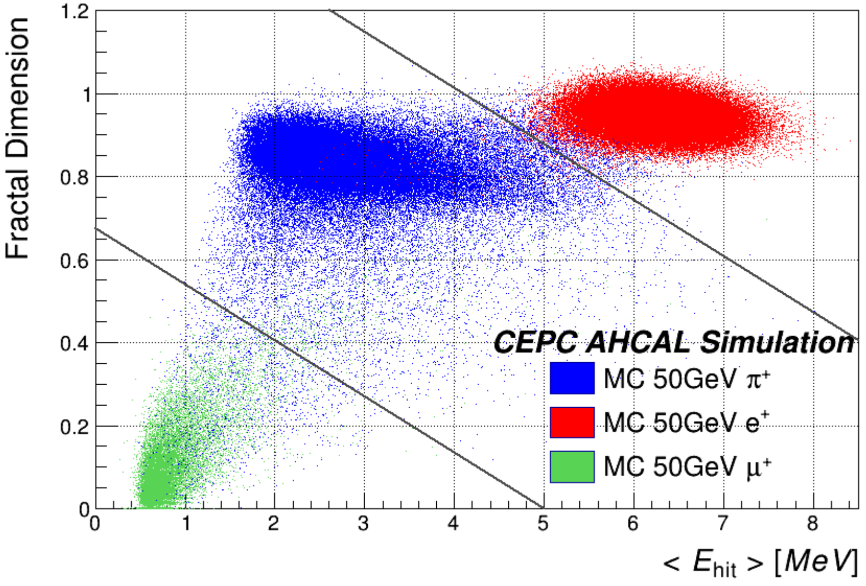
Cut-based PID: Performance

- **FD-<E_{Hit}> Diagram**
  - Shows good pattern recognition capability
  - Shows good PID capability
    - With artificial cut, for π, $\varepsilon = 94.1\%$, $p = 99.1\%$ @ 50 GeV MC sample

Efficiency @ 50 GeV MC

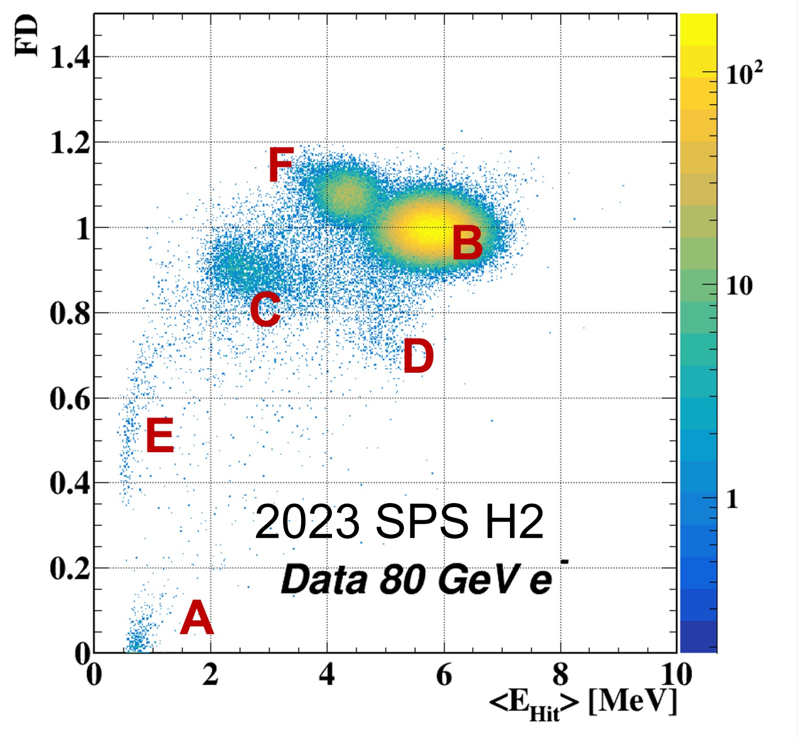
Truth \ ID	μ^+	π^+	e^+
μ^+	99.4%	0.6%	0
π^+	3.6%	94.1%	2.3%
e^+	0	0.3%	99.7%

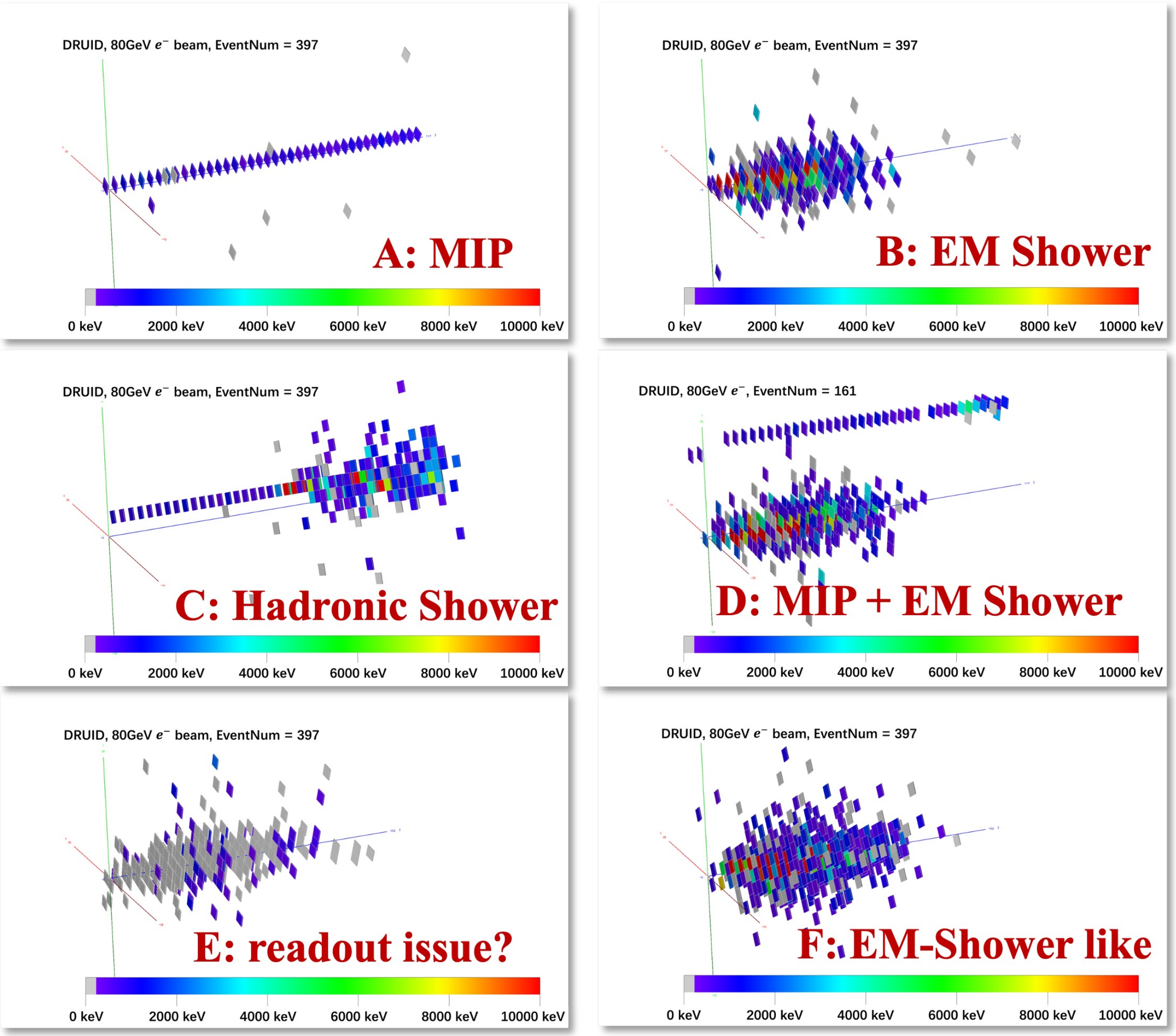




Cut-based PID: Data Components

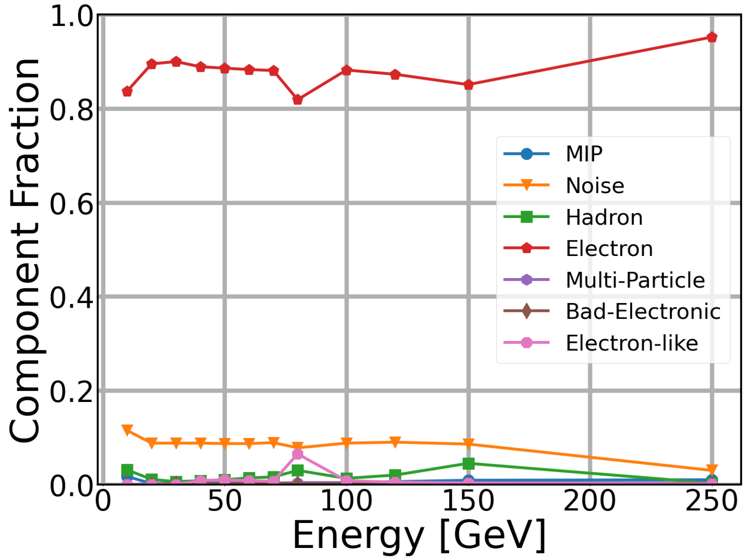
- **Use cut-based PID method, analyze data taken at SPS H2**
  - Observe different components
  - Calculate fractions by artificial cut (see backup)

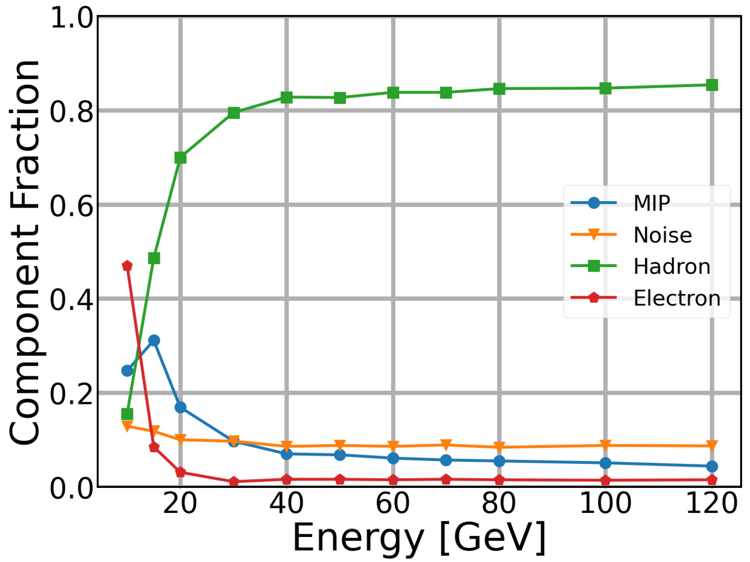




Cut-based PID: Purity Analysis

- **Beam purity @ SPS H2 beamline at 2023:**
  - Electron beam: purity > 80%
  - Pion beam:
    - increases with increasing energy
    - purity > 80% when Energy > 30 GeV
    - Lots of electron events mixed in low-energy
  - Noise is the dominant factor affecting beam purity.





ML-based PID: Overview

Siyuan Song

- **Method:** BDT and ANN
- **Sample:** MC + Data (tagged with Cherenkov/FD)

- **Monte Carlo Samples:** Employ Geant4 11.1.1 Toolkit with the $\mathrm{QGSP}_{\mathrm{BERT}}$ physics list.

Energy point	5 GeV		10 GeV		30 GeV		50 GeV		60 GeV		80 GeV		100 GeV		120 GeV	
	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source
Muon	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC
Electron	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC
Pion	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC	10k	MC

- **Test Beam Samples:** Pre-processed purer 2023 CERN SPS-H2 & PS-T9 test beam data.

Energy point	5 GeV		10 GeV		30 GeV		50 GeV		60 GeV		80 GeV		100 GeV		120 GeV	
	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source	#	Source
Muon	-	-	40k	Data	-	-	-	-	-	-	-	-	40k	Data	-	-
Electron	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data
Pion	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data	10k	Data

<u>Each sample set is split to a Train set and a Test set in a ration of 3:2.</u>
<u>Each sample set would be utilized to build classifiers.</u>

BDT-based PID

Siyuan Song

➢ **Boosting Decision Tree (BDT)**

- Apply Extreme Gradient Boosting
- Energy–independent variables

MC samples

Rank: Variable	Variable weight
1: Shower radius	0.377
2: Shower layers	0.232
3: Hits number	0.088
4: Fired layers	0.083
5: Shower start	0.080
6: Shower density	0.049
7: Z width	0.034
8: FD_6	0.017
9: FD_1	0.015
10: Shower layer ratio	0.014
11: Shower end	0.006
12: Shower length	0.006

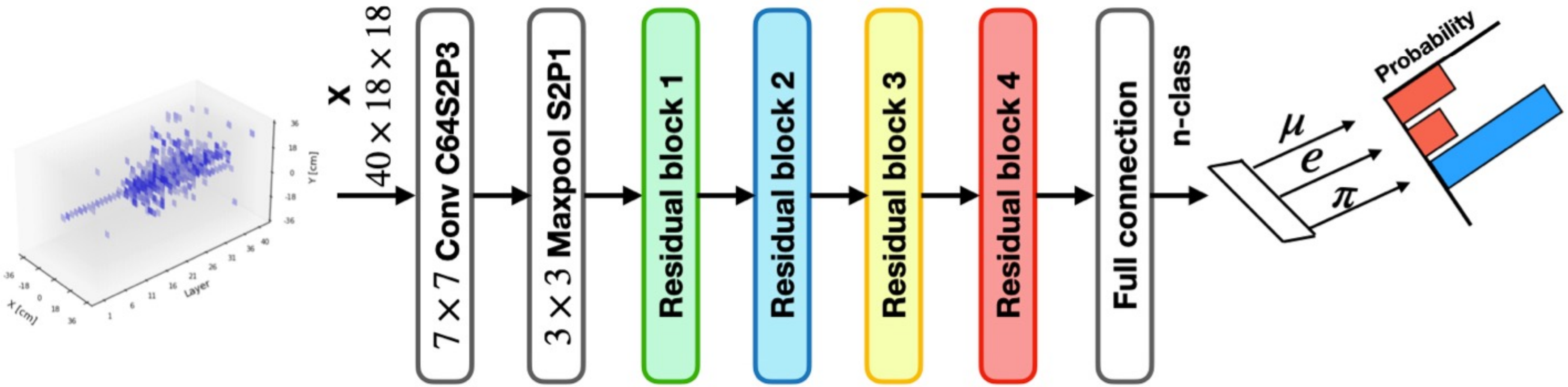
Data samples

Rank: Variable	Variable weight
1: Shower radius	0.379
2: Shower layers	0.228
3: Hits number	0.133
4: Shower density	0.058
5: Fired layers	0.058
6: Z width	0.042
7: Shower start	0.039
8: FD_6	0.019
9: FD_1	0.016
10: Shower layer ratio	0.010
11: Shower length	0.010
12: Shower end	0.008

ANN-based PID

Siyuan Song

- **Cell-based Artificial Neural Networks (ANN) [1]**
  - Input: spatial energy deposition in AHCAL ($18 \times 18 \times 40$)
  - output: probability of each particle type candidate
  - Architecture: take the advantage of the Residual Block
  - Make full use of high-dimensional input



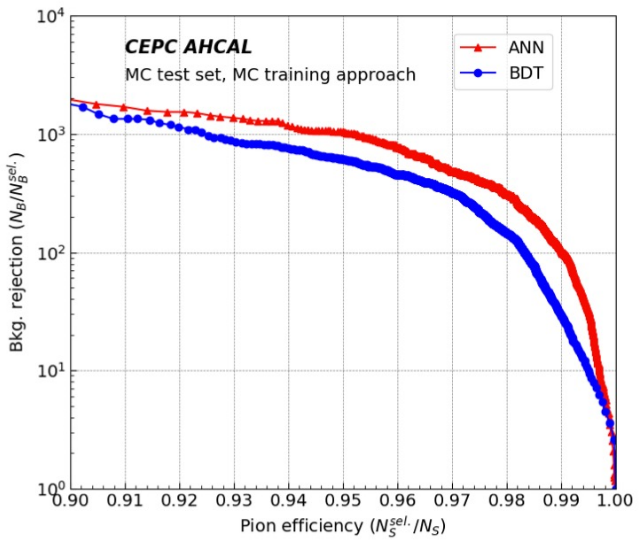
[1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

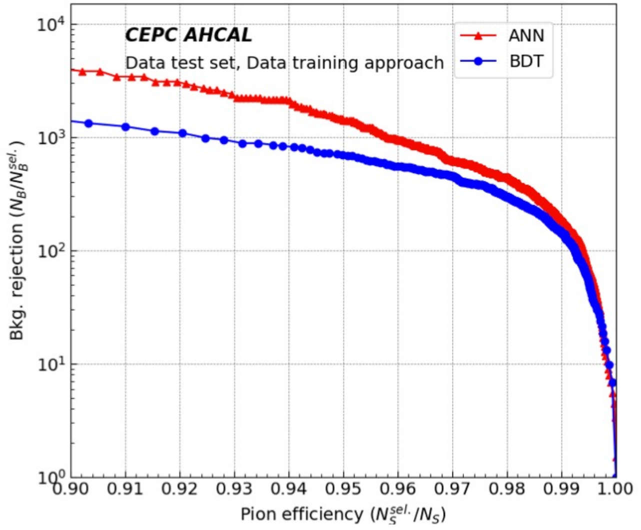
ML-based PID performance comparison

Siyuan Song

- ANN performs better than BDT in terms of background rejection and pion purity, for both MC samples and data.

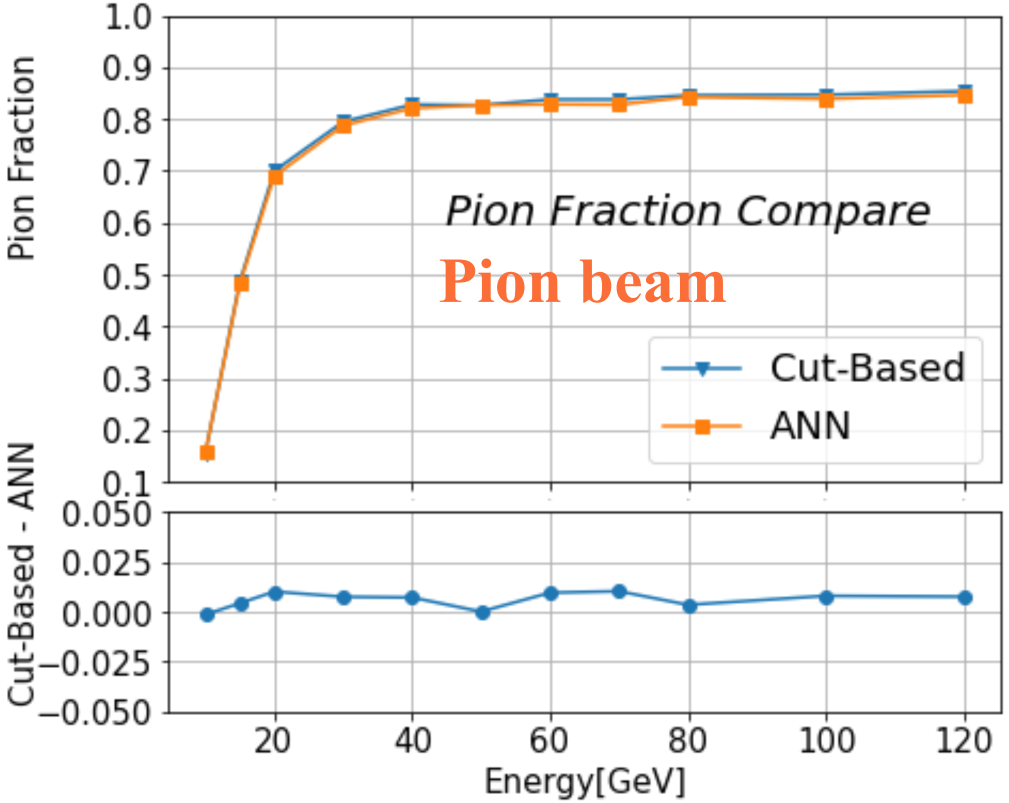
	Pion efficiency					
	90%		95%		99%	
	MC	Data	MC	Data	MC	Data
BDT bkg. rejection	1701.2	1448.5	617.4	691.6	29.6	143.0
ANN bkg. rejection	**2015.7**	**3811.2**	**1040.3**	**1408.5**	**103.9**	**187.8**
Improvement	**↑ 18.49%**	**↑ 163.12%**	**↑ 68.51%**	**↑ 103.65%**	**↑ 251.14%**	**↑ 31.37%**
BDT pion purity	0.998	0.998	0.996	0.996	0.923	0.983
ANN pion purity	**0.999**	**0.999**	**0.998**	**0.998**	**0.980**	**0.989**
Improvement	**↑ 0.05%**	**↑ 0.13%**	**↑↑ 0.21%**	**↑ 0.22%**	**↑ 6.20%**	**↑ 0.61%**

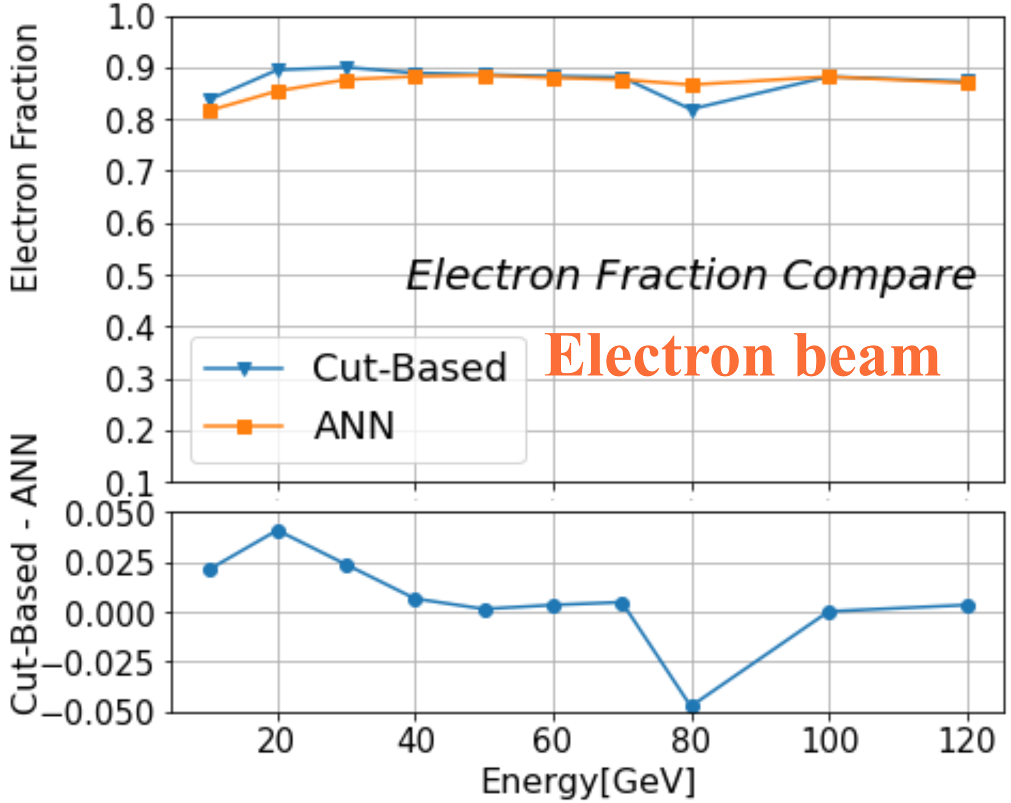




Comparison and Validation

- The difference of data purity between cut-based PID and ANN
  - within $\pm 1\%$ for pion beam
  - within $\pm 5\%$ for electron beam





Summary

- Three PID methods are developed
  - Cut-based: Fractal dimension, Average Hit Energy
  - BDT-based: Energy-independent variables
  - ANN-based: spatial energy deposition
- Performance comparison:
  - Cut-based: $\varepsilon = 94.1\%, p = 99.1\%$ (50GeV MC sample)
  - BDT-based: $\varepsilon = 95\ \%, p = 99.6\%$ (MC sample uniform ranged from 0-120 GeV)
  - ANN-based: $\varepsilon = 95\ \%, p = 99.8\%$ (MC sample uniform ranged from 0-120 GeV)
- The corresponding component fractions are estimated, and the results from cut-based PID and ANN are consistent.
  - > 80% for electron and pion beams with > 30 GeV

Thanks for your attention!

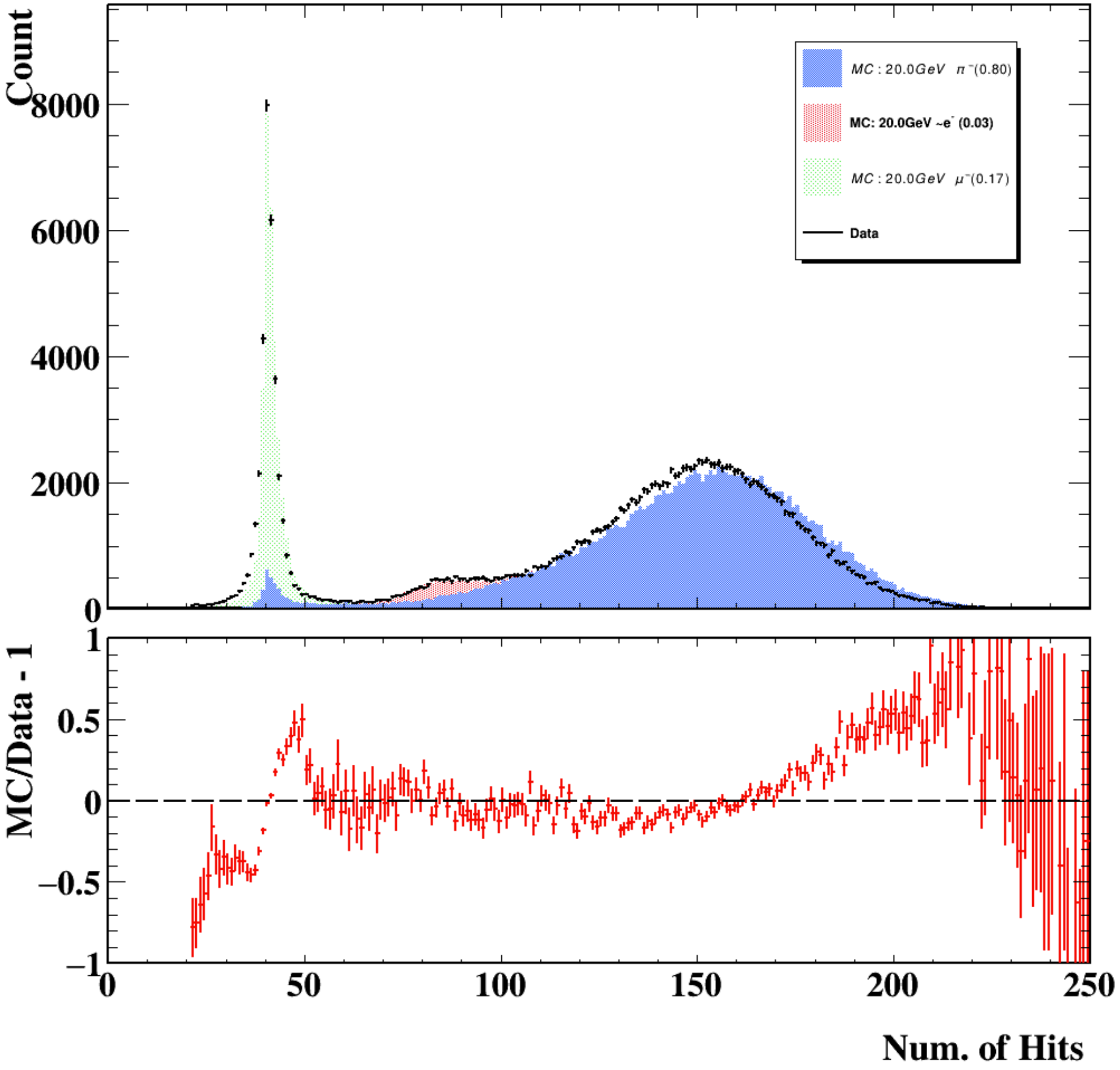
Xin Xia
Institute of high energy physics
On behalf of the CEPC calorimeter working group
CALICE Collaboration meeting, Sep. 28 2023

Backup !

Xin Xia

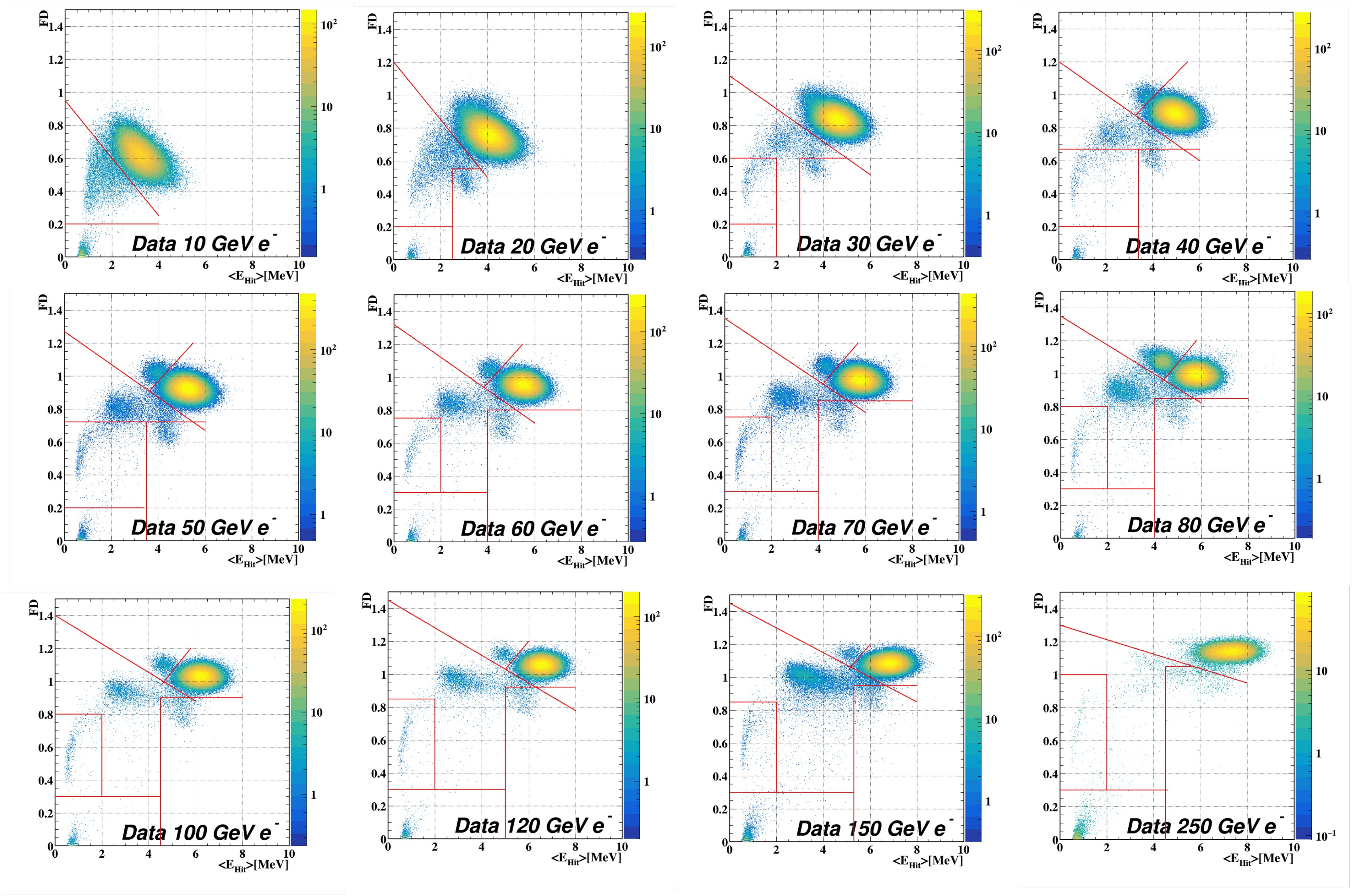
2023/9/26

Num. of hits

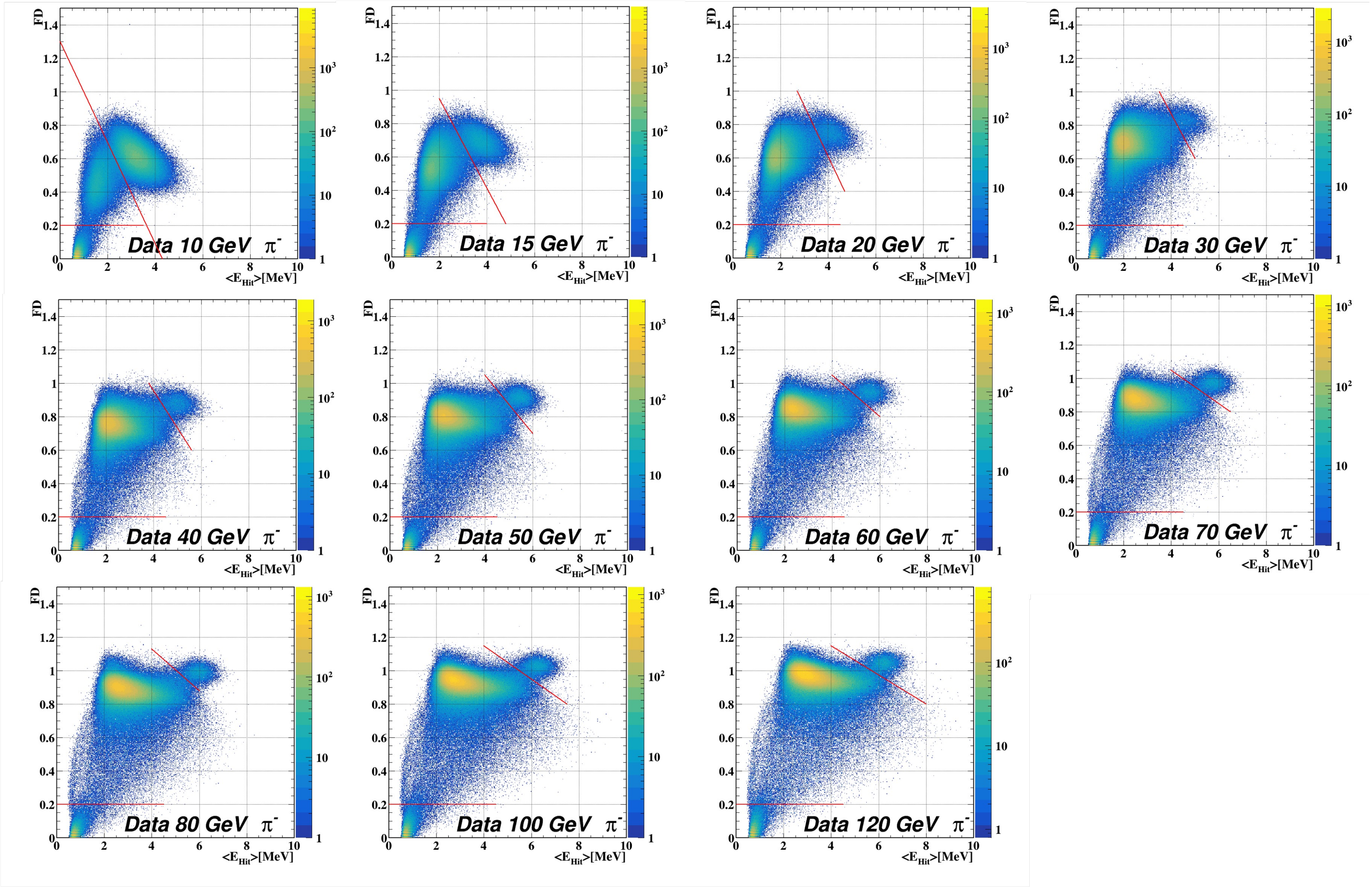


Count
8000
6000
4000
2000
0
MC : 20.0GeV π⁻(0.80)
MC: 20.0GeV ~e⁻ (0.03)
MC : 20.0GeV μ⁻(0.17)
Data
MC/Data - 1
1
0.5
0
−0.5
−1
0
50
100
150
200
250
Num. of Hits

Evolute versus Incident Energy: Electron Beam



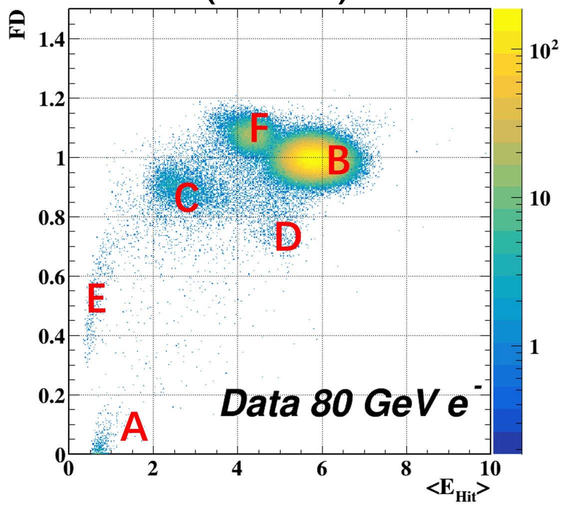
Evolute versus Incident Energy: Pion Beam

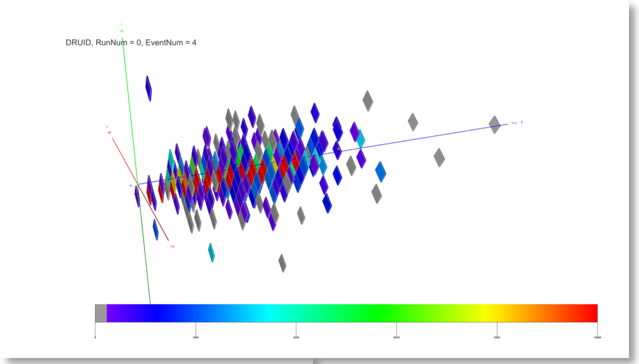


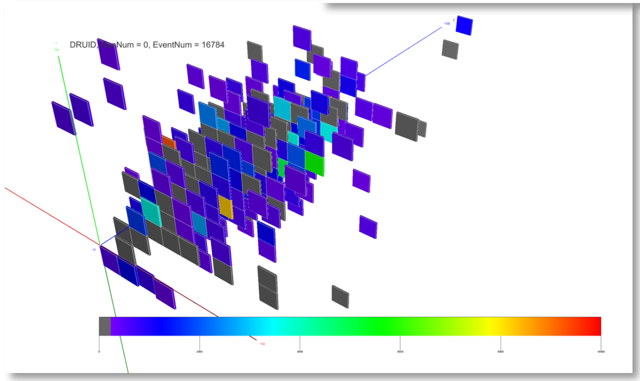
Component Highlight: E

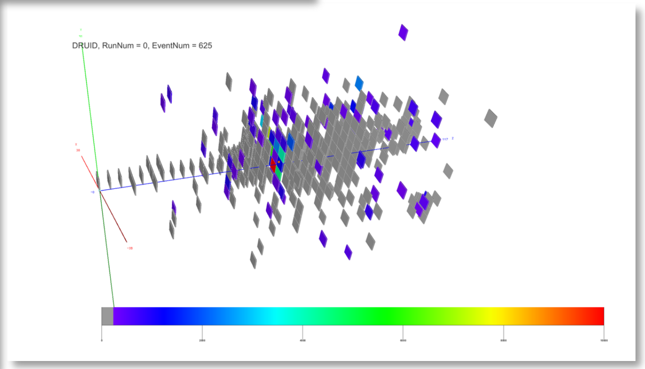
Characteristics:

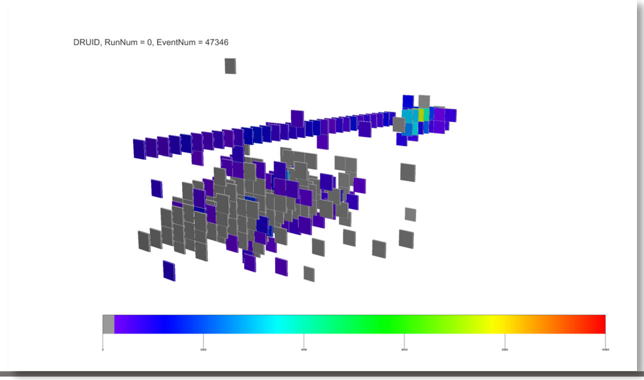
- Shape: Shower Like
- Energy: More low energy (< 0.5 MIP) hits in inner region of shower
- Found in:
  - Electron + pion beams
  - SPS H8 (2022) + SPS H2 (2023)

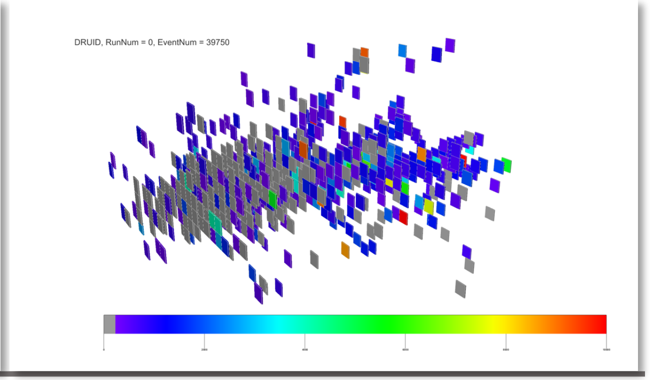








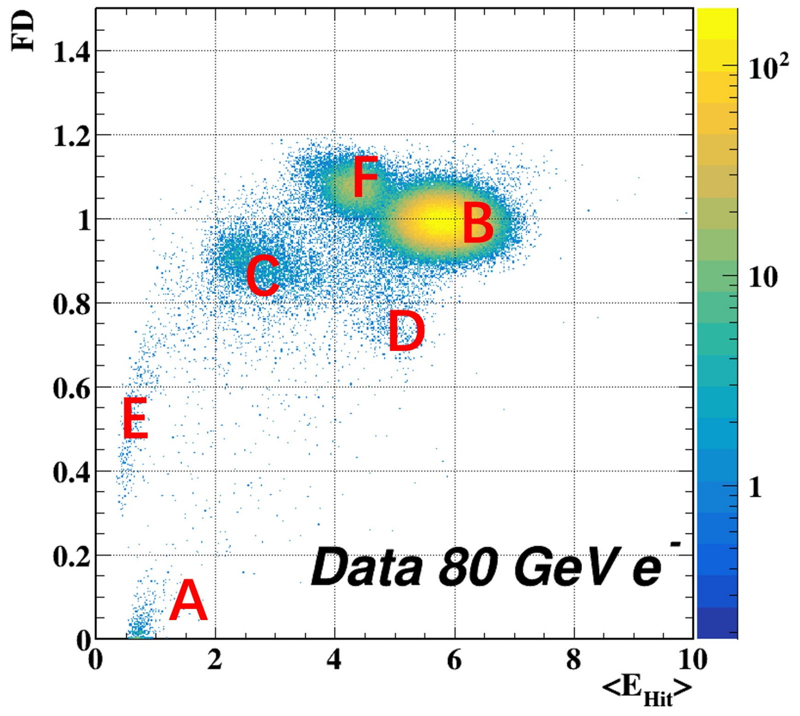


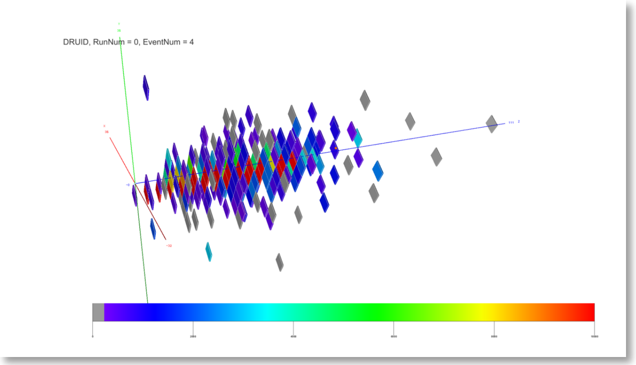


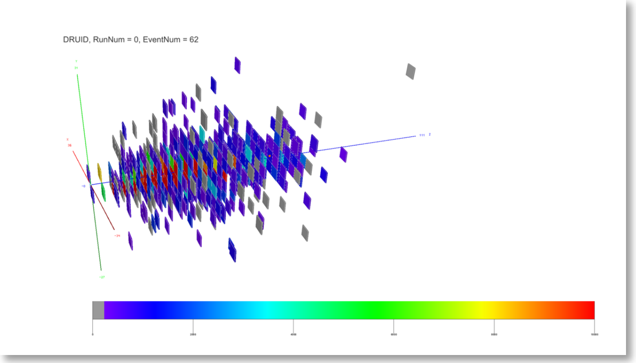
Component Highlight: F

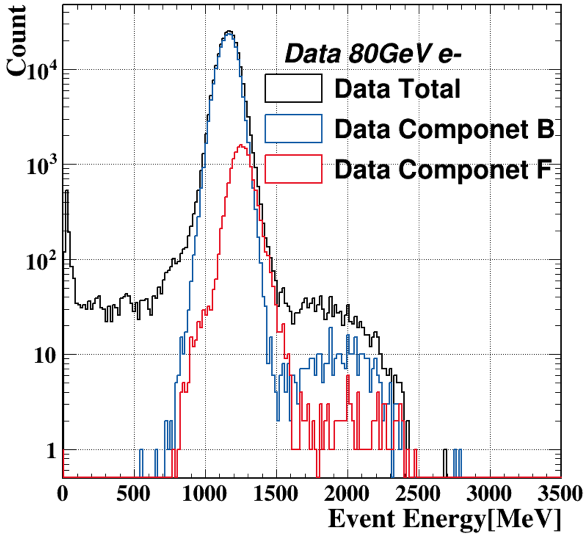
Characteristics:

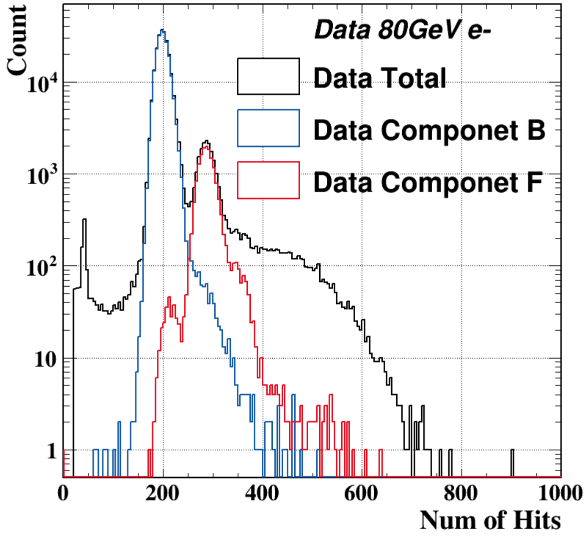
- Shape: EM. like
- More hits and higher energy deposition
- Only in electron beam taking at SPS H2 at 2023









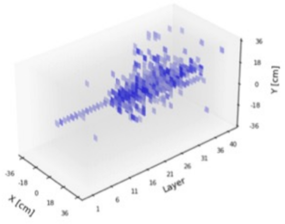


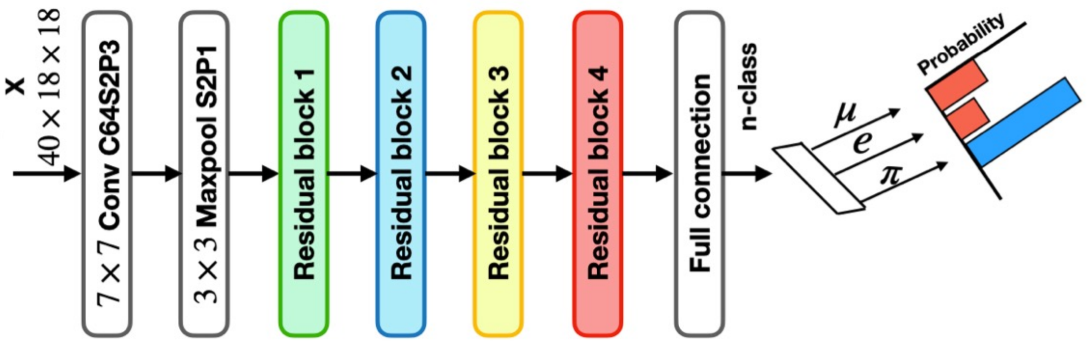
ANN architecture

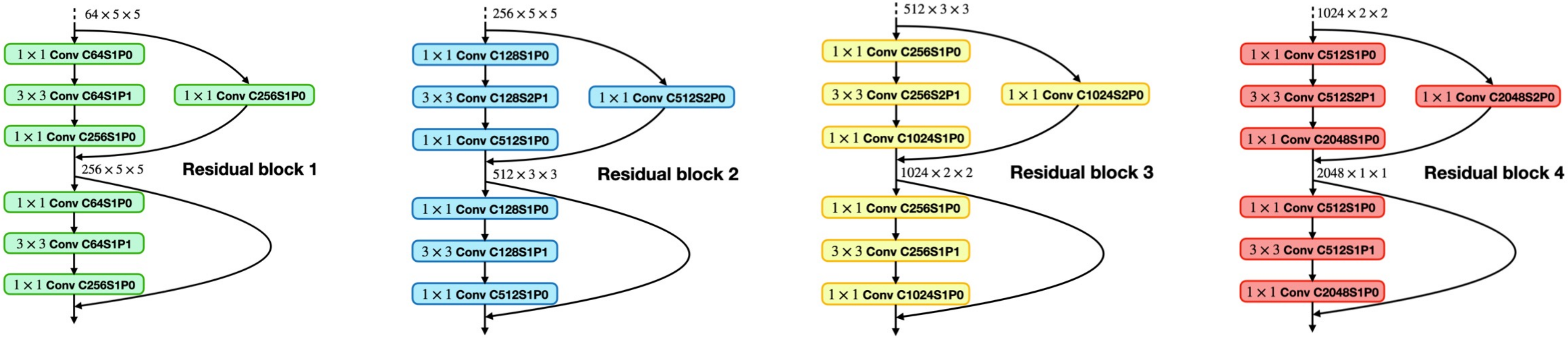
- **Architecture: take the advantage of the Residual Block**

Input: energy deposits in AHCAL $(18 \times 18 \times 40)$.

output: probability of each particle type candidate.







He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

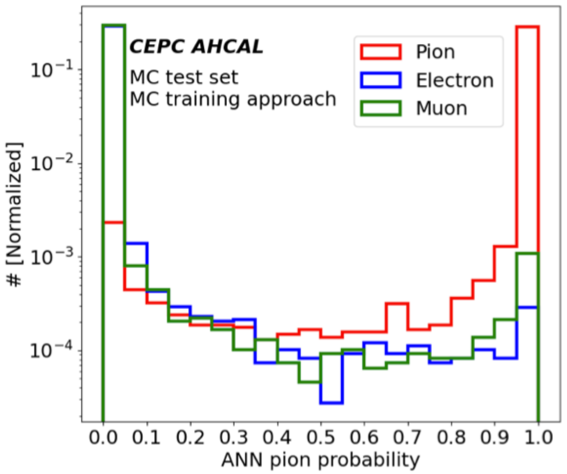
ANN-based PID: Performance

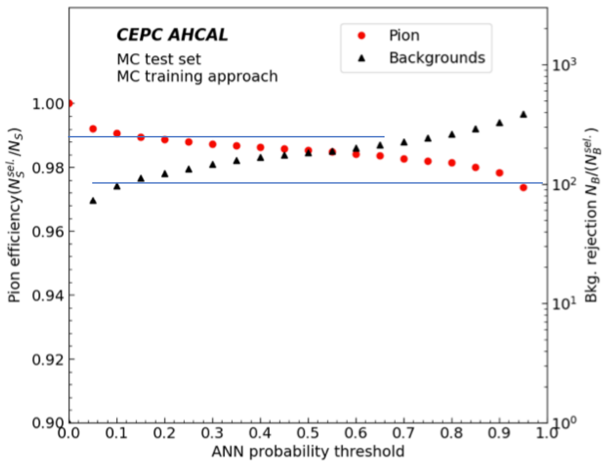
Siyuan Song

- ANN_{MC} **is trained on MC samples**
  - At 95% pion signal efficiency:

    Bkg. Rejection: 1040.3 ($N_{Bkg}/N_{sel.}$),

    pion purity: 99.8%

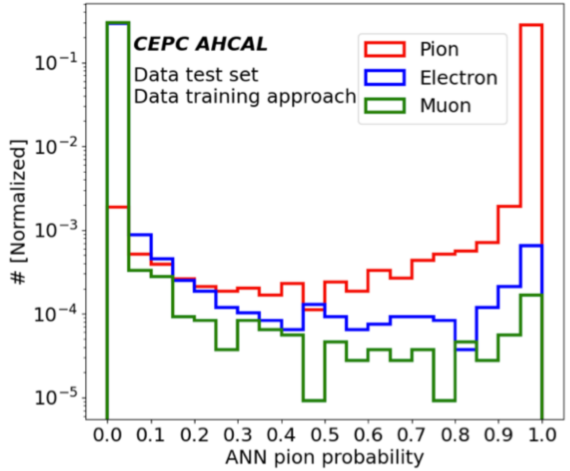


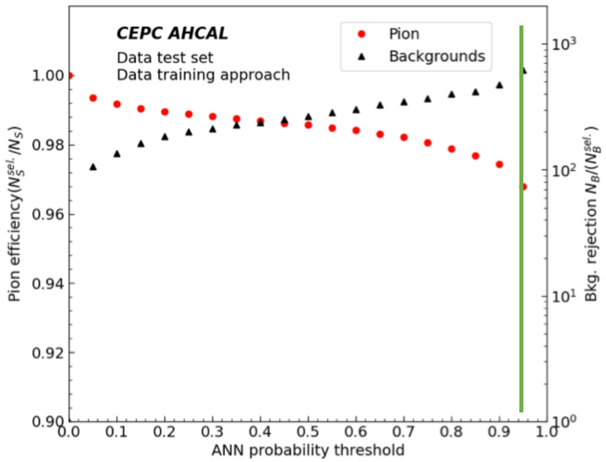


- ANN_{Data} **is trained on test beam data**
  - At 95% pion signal efficiency,

    Bkg. Rejection: 1408.5 ($N_{Bkg}/N_{sel.}$),

    pion purity: 99.8%





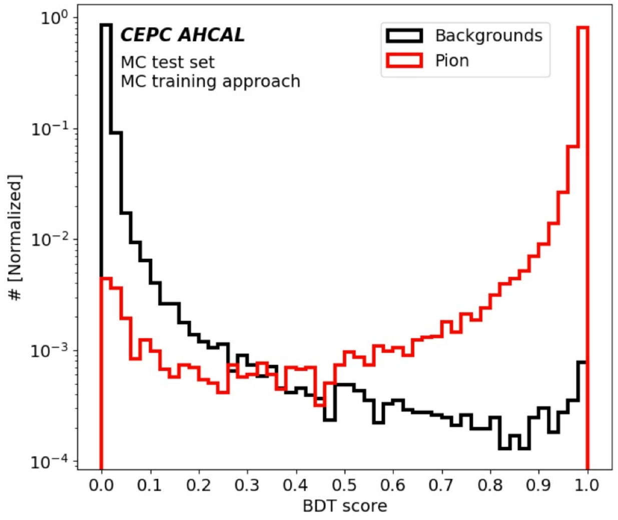
BDT-based PID: Performance

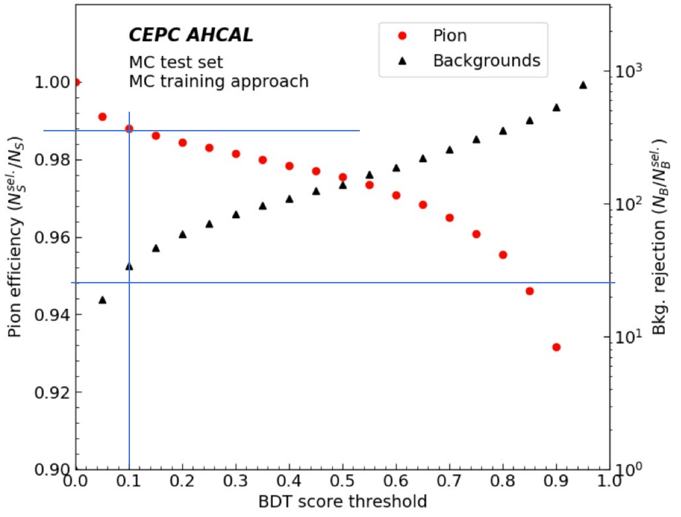
Siyuan Song

- BDT_{MC} **is trained on MC samples**
  - At 95% pion signal efficiency,

    Bkg. Rejection: 617.4 ($N_{Bkg}/N_{sel.}$)

    pion purity: 99.6%

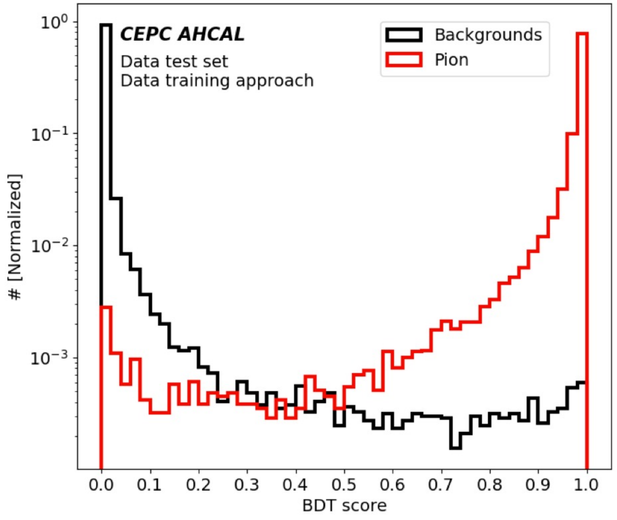


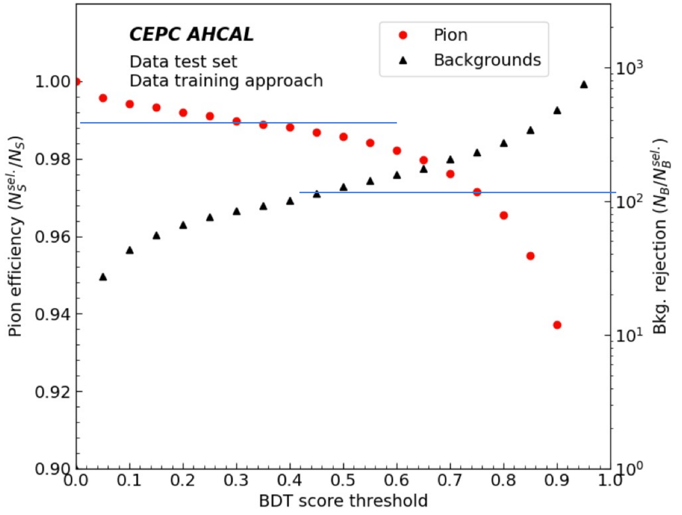


- BDT_{Data} **is trained on test beam data**
  - At 95% pion signal efficiency,

    Bkg. Rejection: 691.6 ($N_{Bkg}/N_{sel.}$)

    pion purity: 99.6%





Cherenkov

At 5 GeV, 20k pion from Cherenkov identification, 19.357k events are also identified by ANN. Ratio = 96.8%