CEPC Physics Performance Wednesday Working Meeting

1-1 correspondence reconstruction at CEPC

https://inspirehep.net/literature/2847034

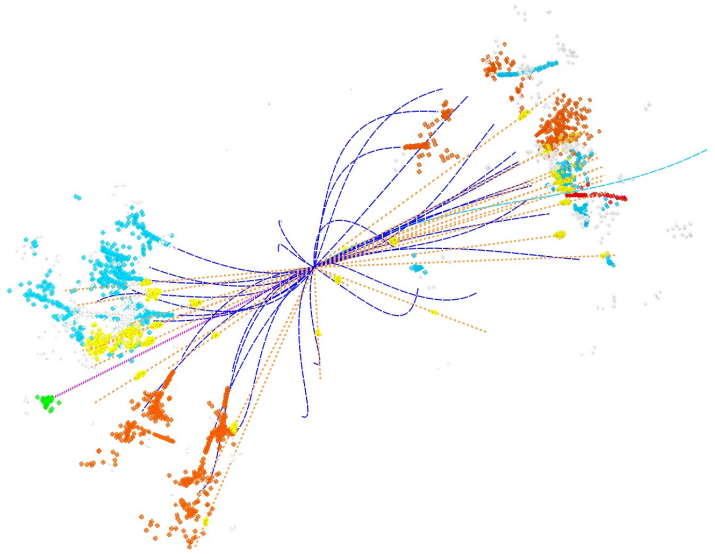
Yuexin Wang, Hao Liang, Yongfeng Zhu, Yuzhi Che,
Xin Xia, Huilin Qu, Chen Zhou, Manqi Ruan

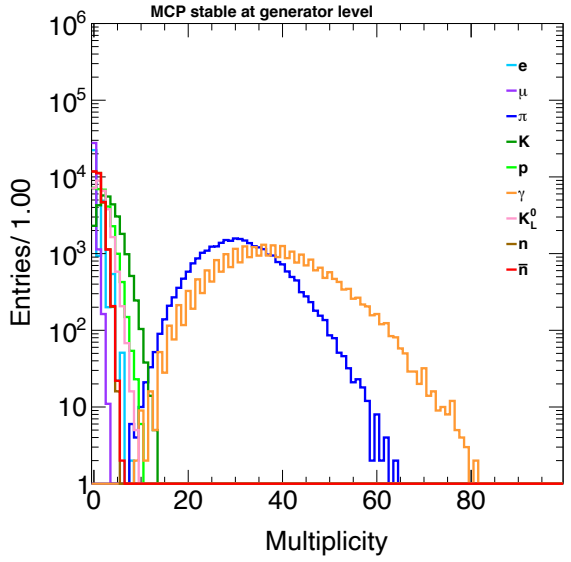
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

May 28, 2025

Particle reconstruction at e^+e^- Higgs factory

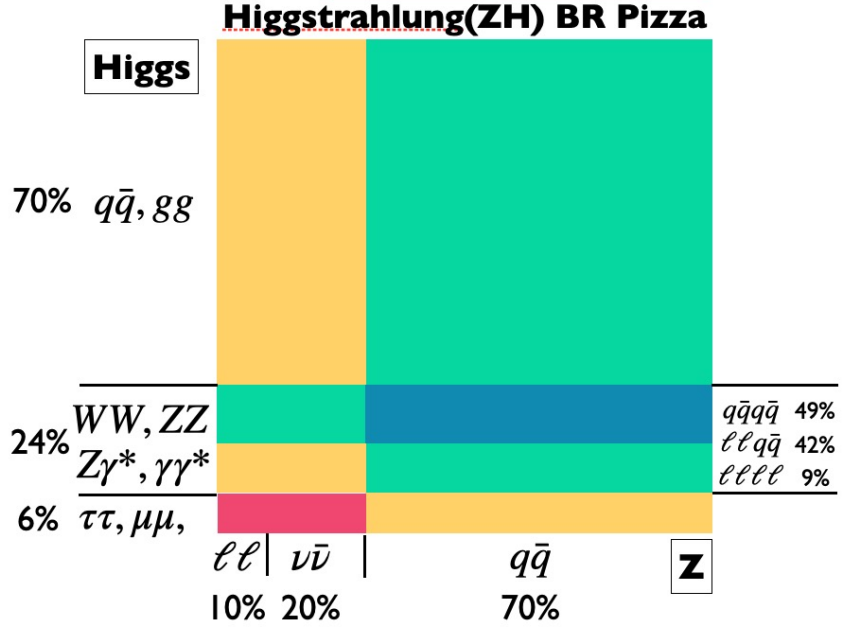
- **Reconstruction in collider experiments**
  - Interpret detector hits into observables (type, E, P, vertex…) of final state particles.
  - Ultimate goal: **1-1 correspondence** between truth & reconstructed particles.
  - Difficulty of achieving this goal highly depends on collision environment.
- **CEPC: e^+e^- Higgs factory (240 GeV)**
  - ~97% with hadronic/jet final states (ZH production)
  - particle multiplicity up to ~100 (di-jet)





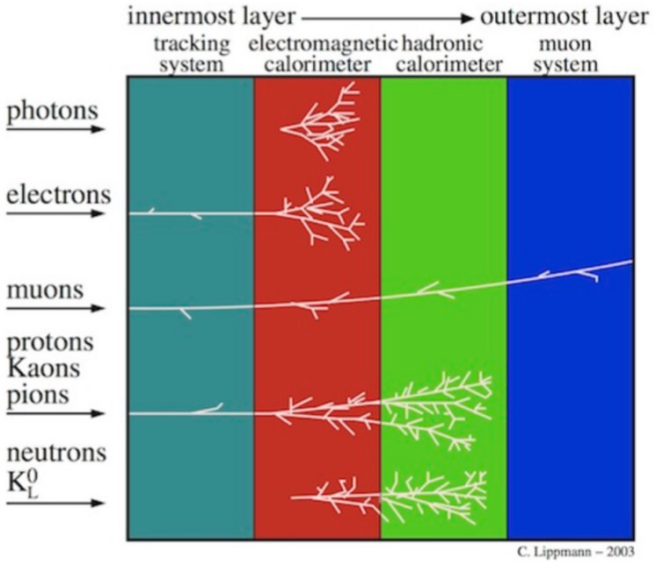
#jets	Probability
0	2.44%
2	29.73%
4	59.58%
6	8.23%

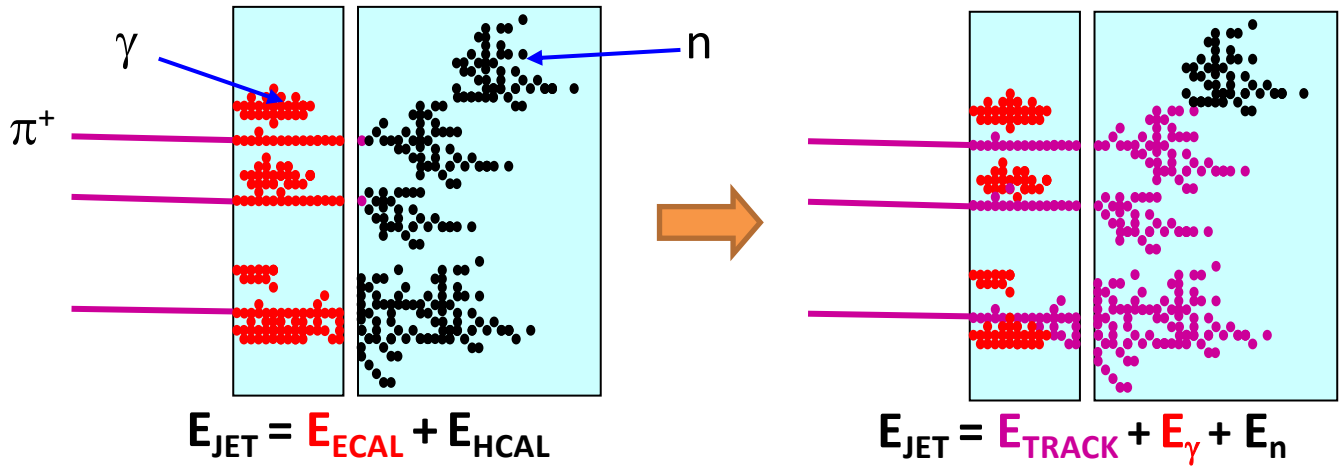
~ 97% with Jets!



Particle Flow reconstruction

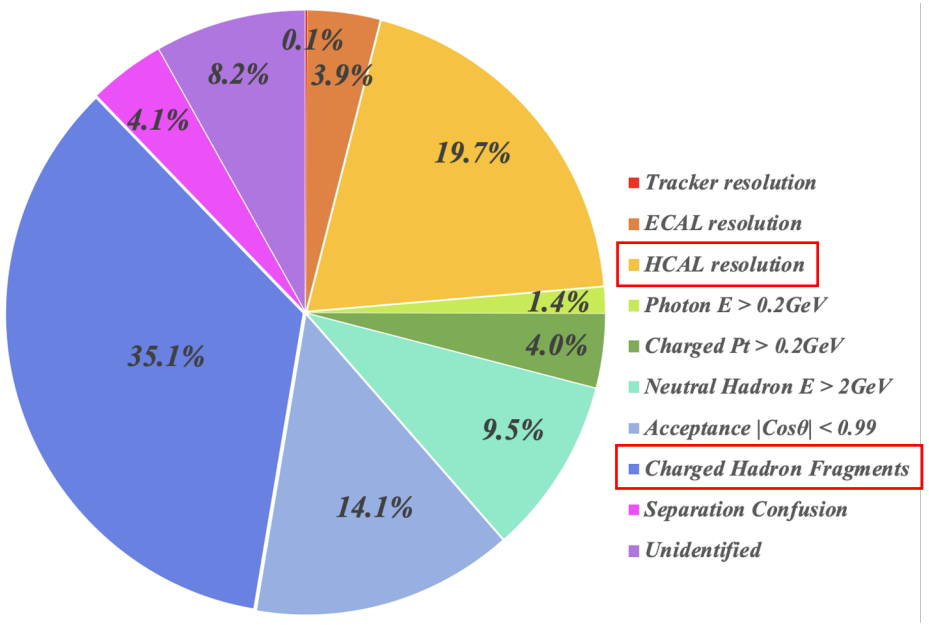
- **Basic idea**
  - efficiently separate and identify different types of particles
  - measure their energy/momentum with the optimal sub-detector
- **Significantly enhance jet energy measurement**
  - $E_{\mathrm{Jet}} = E_{\mathrm{Track}}(65\%) + E_{\gamma}(25\%) + E_{h^0}(10\%)$
  - Tracker momentum resolution (~0.1%) << Calorimeter energy resolution (~60%/√E)

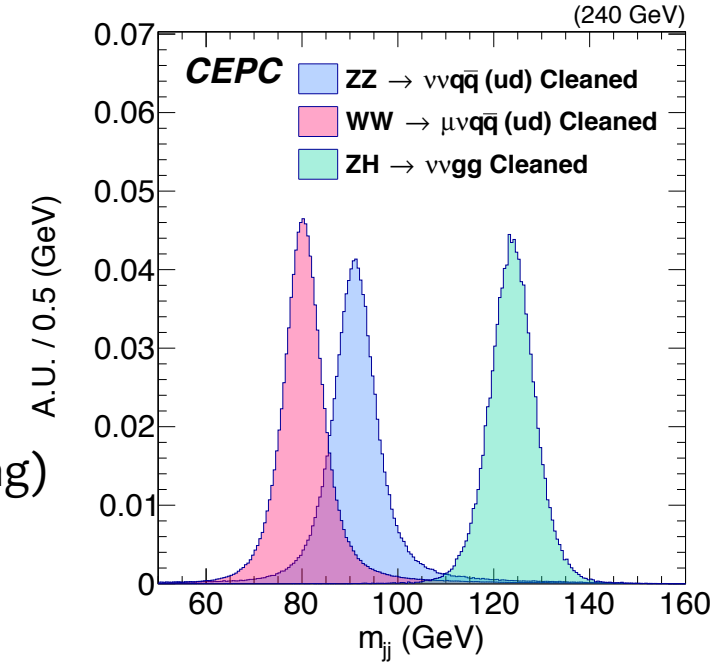


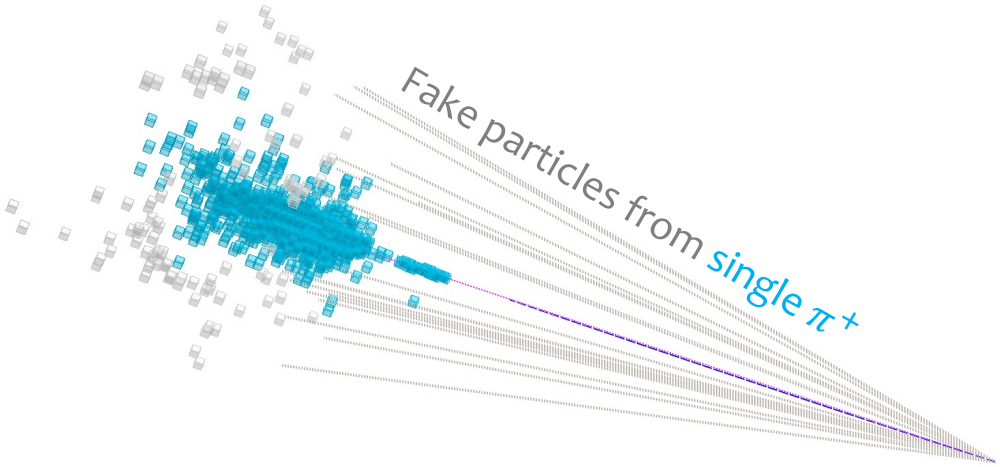


Particle Flow reconstruction

- **Performance at CEPC CDR phase**
  - Quantified by Boson Mass Resolution (BMR)
    - di-jet final state (at 240 GeV)
    - Baseline PFA: Arbor [arXiv: 1403.4784]
    - Baseline BMR: ~3.7%
  - Bottleneck (BMR decomposition pie-chart)
    - 1st: Fake particles (fragments) from $h^{\pm}$ (energy double counting)
    - 2nd: HCAL resolution

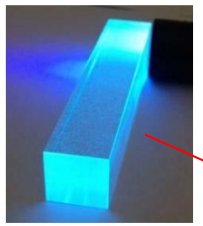




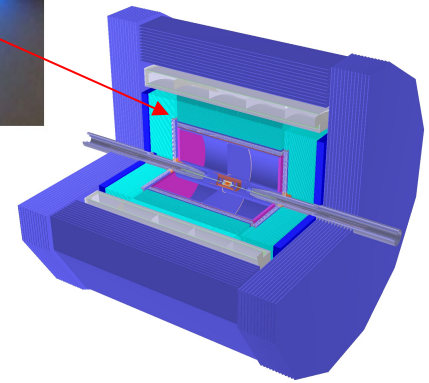


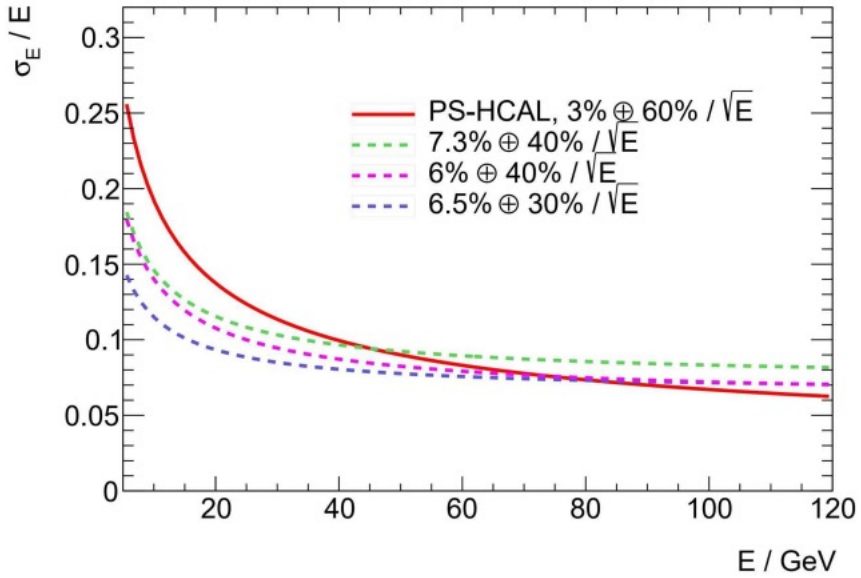
Improve hadronic resolution with Glass-HCAL

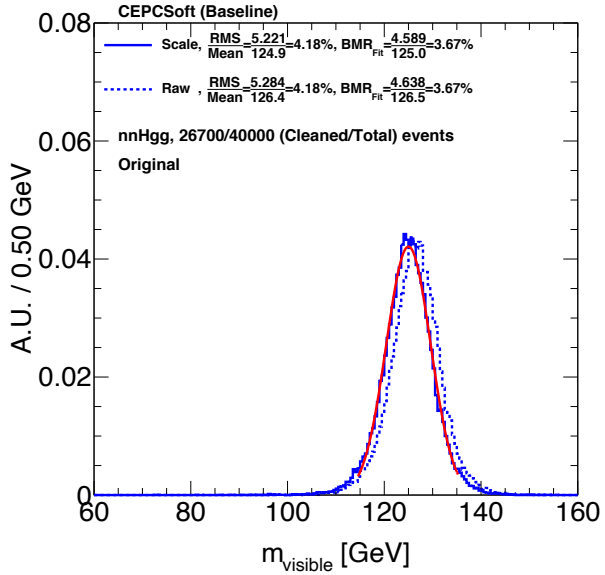
- **Glass Scintillator HCAL (GSHCAL)** [NIMA 1059 (2024) 168944]
  - Glass advantages
    - higher density → higher sampling fraction
    - doping with neutron-sensitive elements (Gd) : improve hadronic response
    - low cost
  - Significantly improvement in
    - Hadronic energy resolution: traditional 60%/√E → **30~40%/√E**
    - BMR: baseline 3.7% → **3.3%**

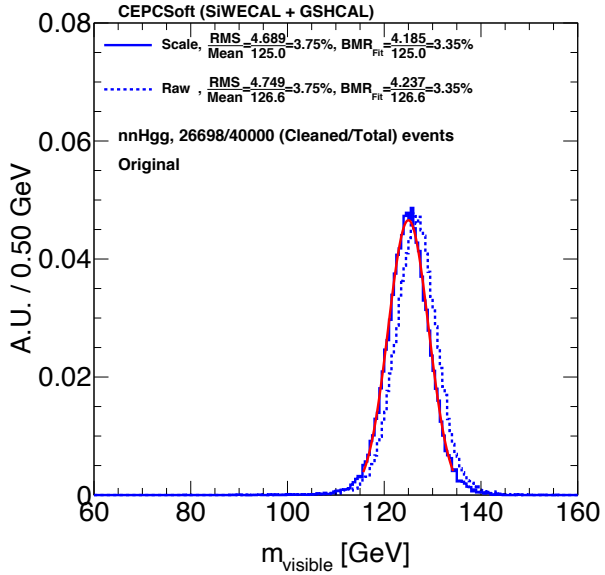


Glass Scintillator



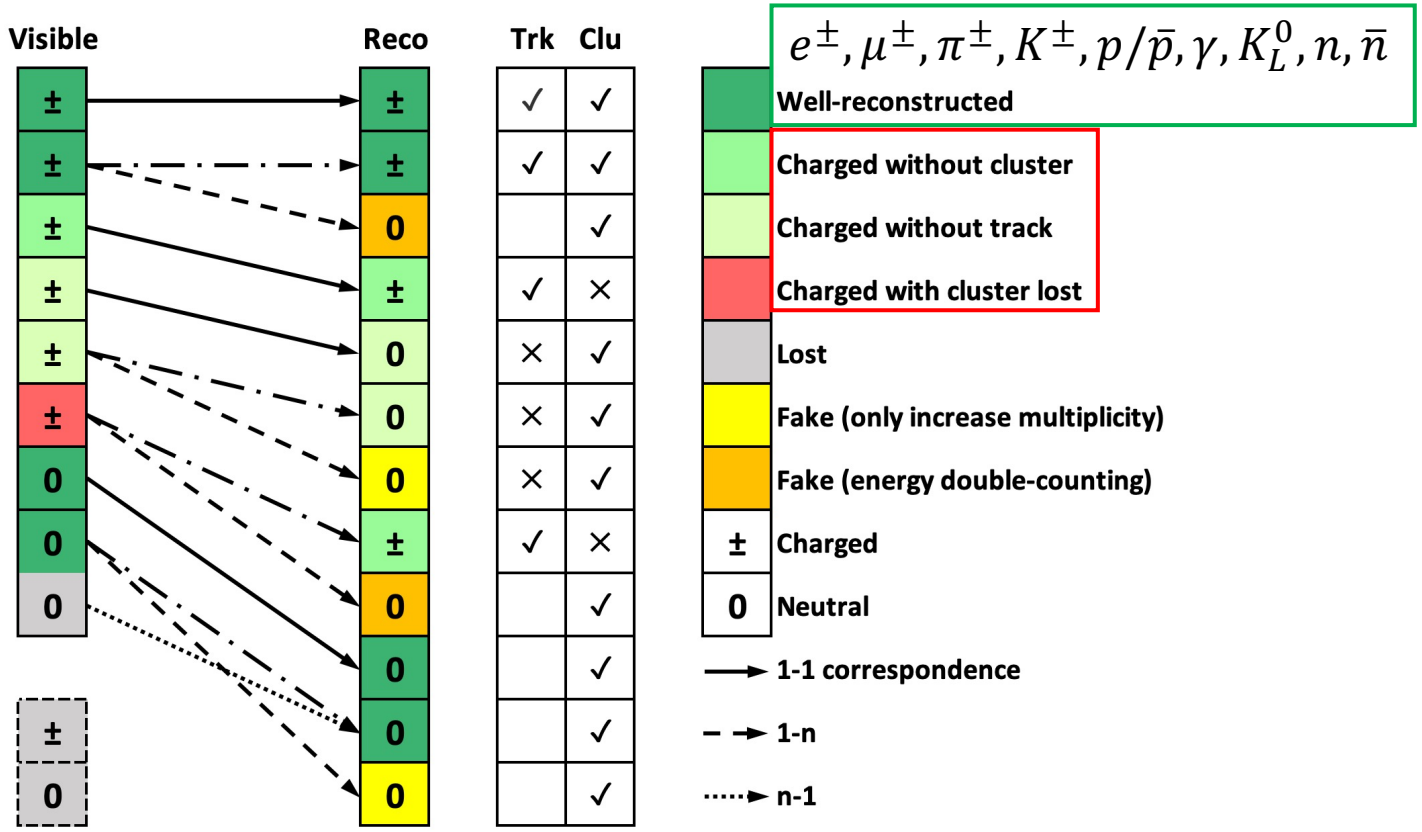






Particle mapping & classification of reconstructed particles

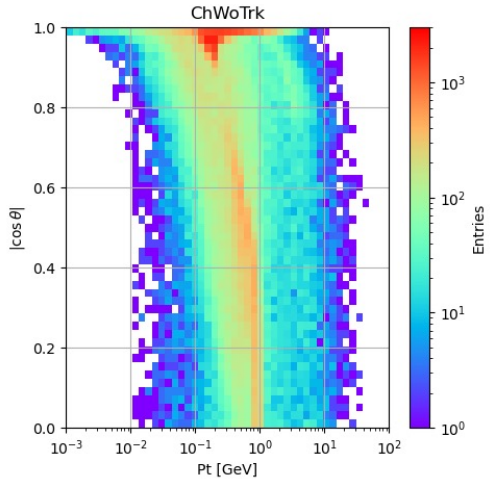
- **Build mapping between truth visible & reconstructed particles**
  - Reconstructed particles are composed of track/cluster hits
  - Truth link between hits and truth particles that generate the hits

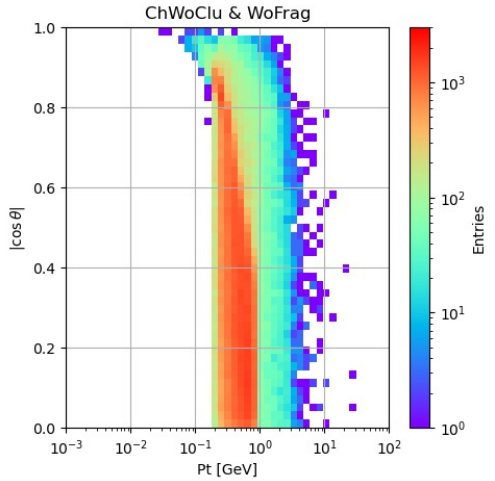


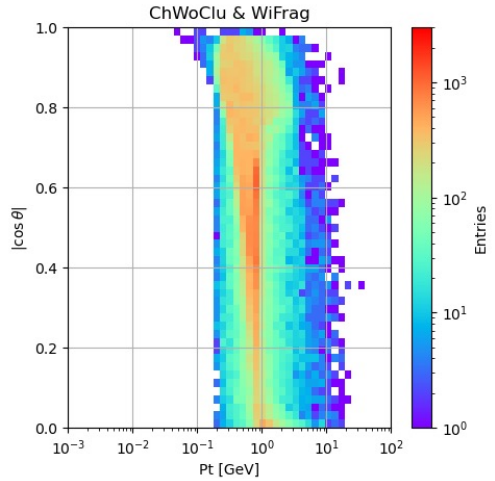
Charged particle without track/cluster

- **Almost in forward and low Pt regions**

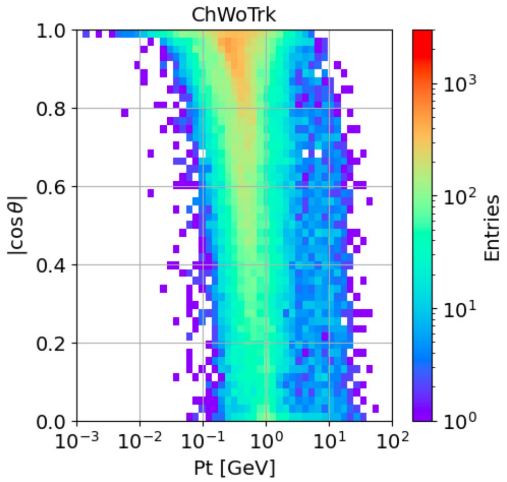
CDR

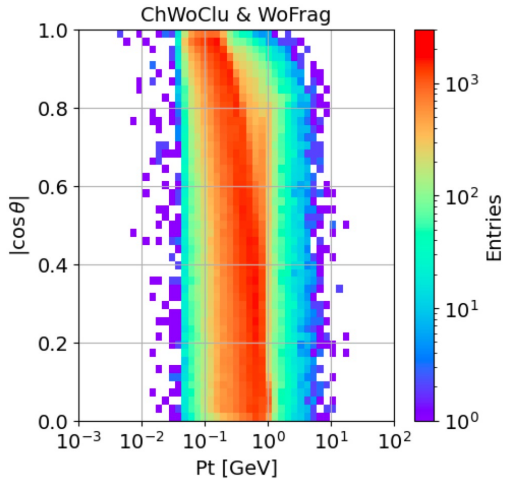


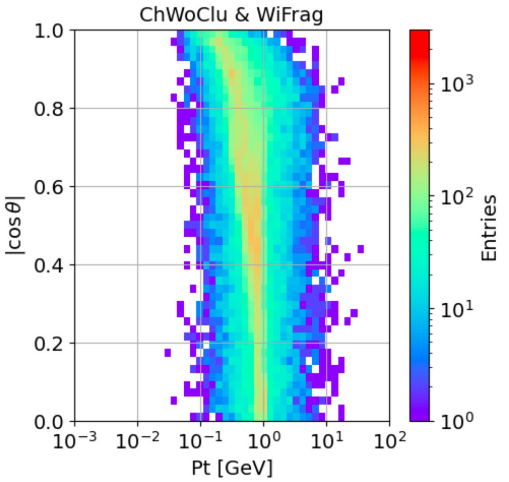




TDR

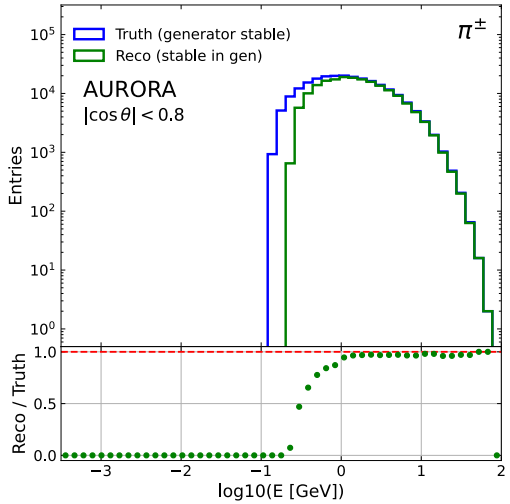


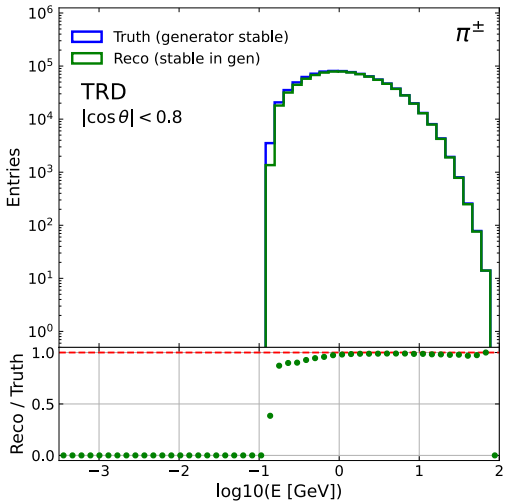


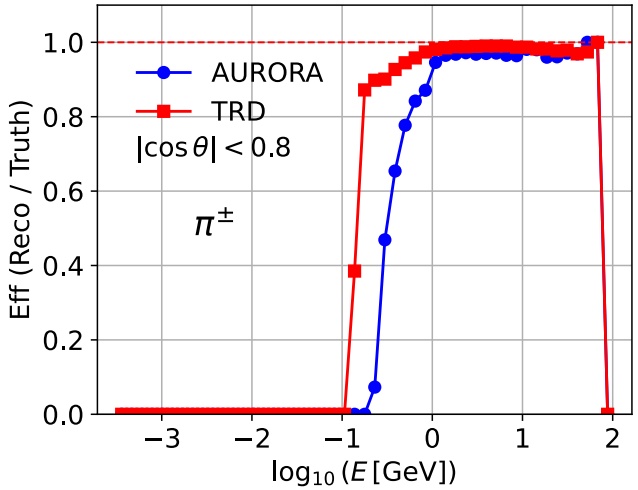


Particle reconstruction efficiency

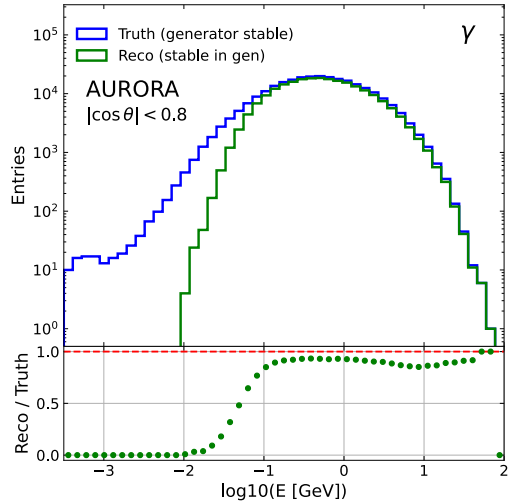
CDR

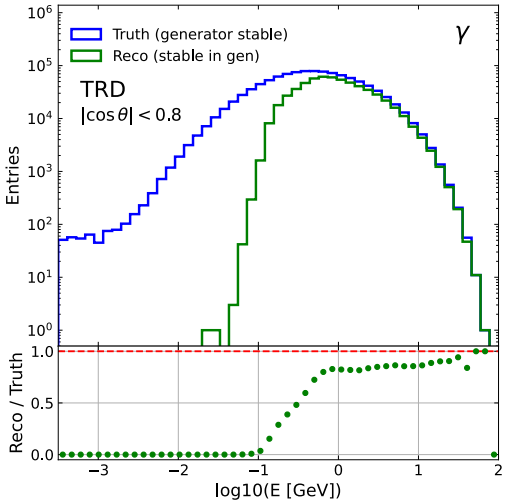


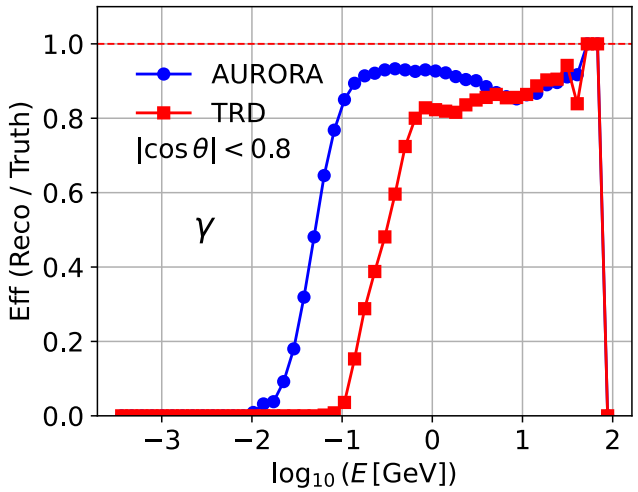




TDR

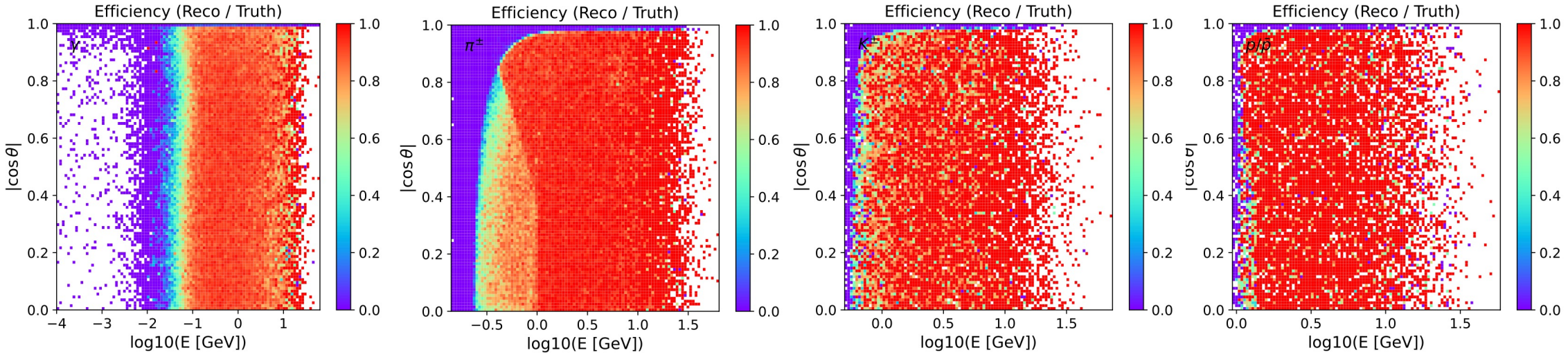




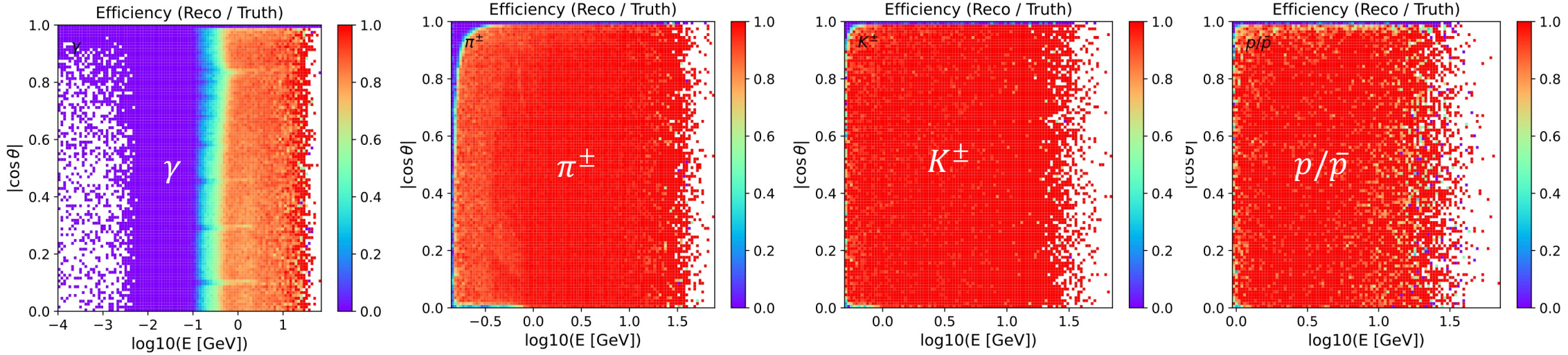


Particle reconstruction efficiency

CDR



TDR



Particle identification using machine learning

- **1-1 correspondence problem → supervised multiclass classification task**
- **ML model: Particle Transformer (ParT)** [arXiv: 2202.03772]
  - Inputs: ~55 particle features
    - Reconstructed particles: E, P, (θ, φ)
    - Track: #hits, P, endpoint, dE/dx
    - Cluster: #hits, E, shape variables, time spectrum…
  - Outputs
    - Likelihood (score) of each category
  - ~ 2.2M parameters
- **Sample**
  - Full simulation $\nu\nu H, H \to gg$ (240 GeV)
  - Statistics
    - CDR: 10^6 events
    - TDR: 2×10^5 events
  - Training : Validation : Test = 6:2:2

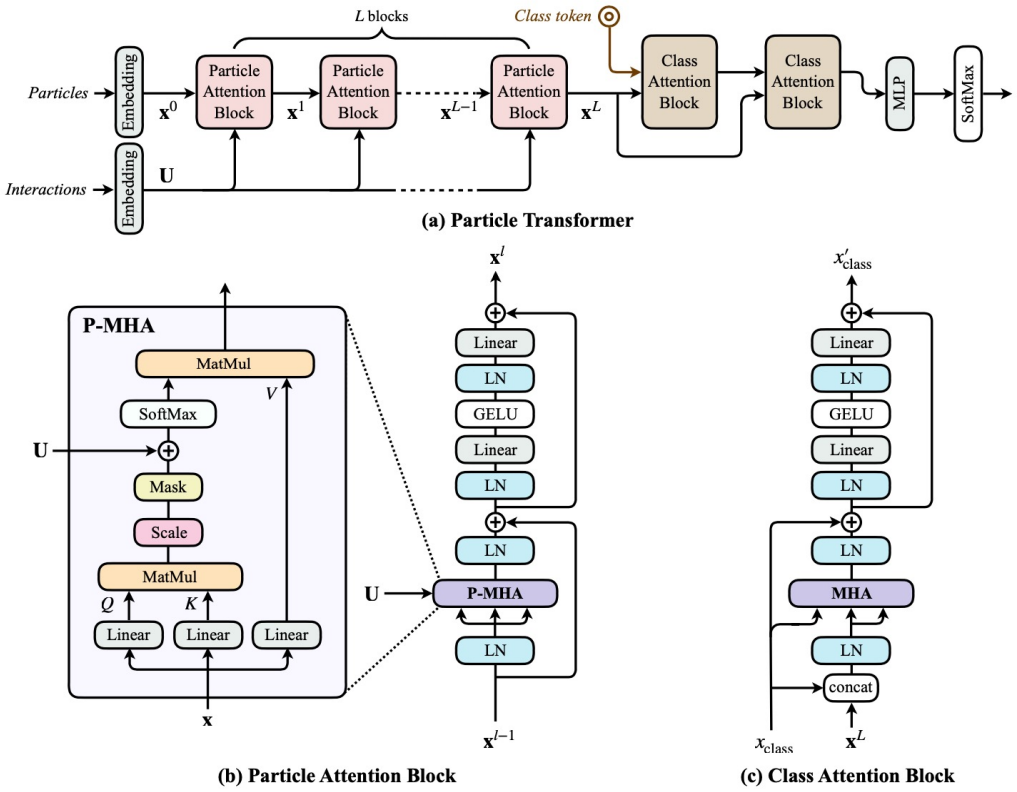


Figure 3. The architecture of (a) Particle Transformer (b) Particle Attention Block (c) Class Attention Block.

Input feature variables

➢ Input features at particle, track, cluster level

Table A.2: Input variables of ParT.

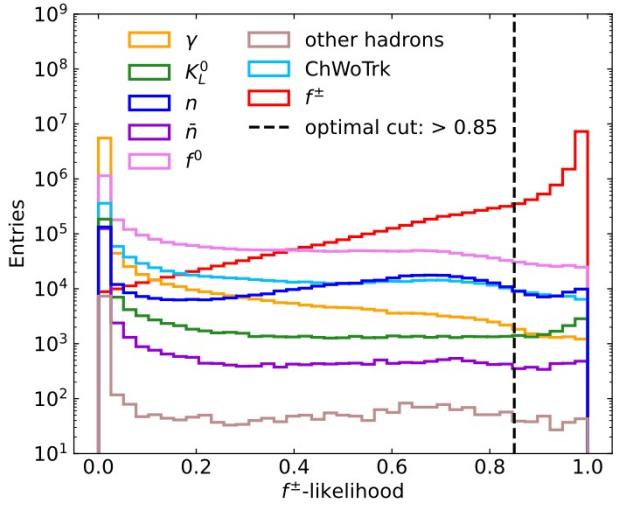
Object level	Observable variables
Reconstructed particle	4-momentum (E, p_x, p_y, p_z) Direction (θ, ϕ) Number of tracks and clusters
Track	Number of hits Endpoint position 3-momentum $(\lvert\vec{p}\rvert, p_x, p_y, p_z, p_T)$ $\mathrm{d}E/\mathrm{d}x$ (mean of 5–85% truncation and quartiles)
Cluster	Number of hits Energy Position of shower starting point Position of center of gravity Fractal dimension [56] Second moment (M_2) Distance between ECAL inner surface and shower starting point Distance between ECAL inner surface and center of gravity Distance between ECAL inner surface and the innermost hit Distance between ECAL inner surface and the outermost hit Maximum distance between cluster hits and the track helix (for charged particles) Maximum distance between cluster hits to the axis from the innermost hit to the center of gravity Average distance between cluster hits to the axis from the innermost hit to the center of gravity Hit time spectrum (the fastest time and quintiles)
Closest charged cluster	Minimum distance between cluster hits of each other Number of hits Energy Ratio of E_{cluster} to p_{track}

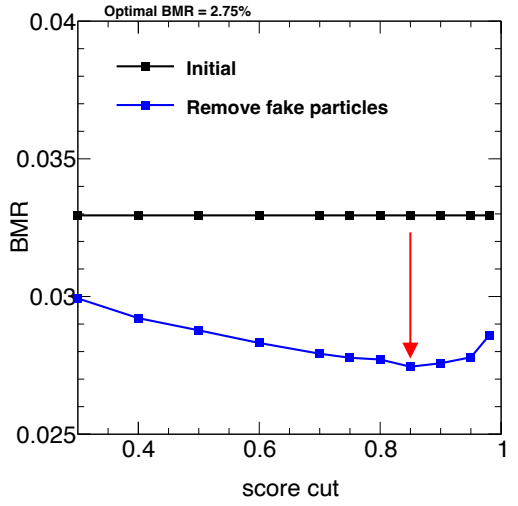
TDR ECAL has no time info.

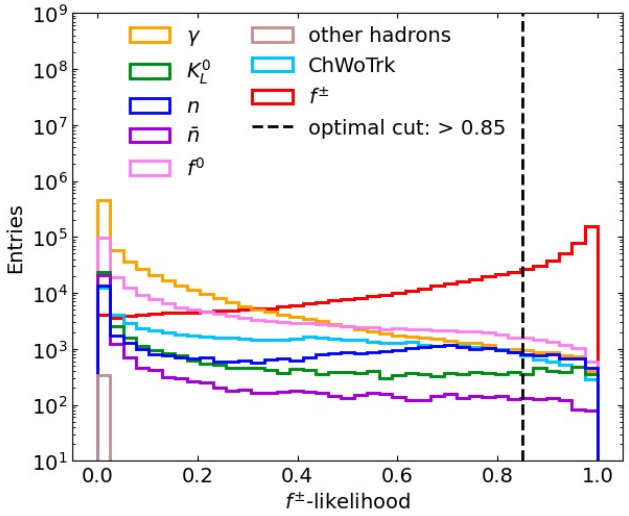
some variables are not yet implemented in TDR…

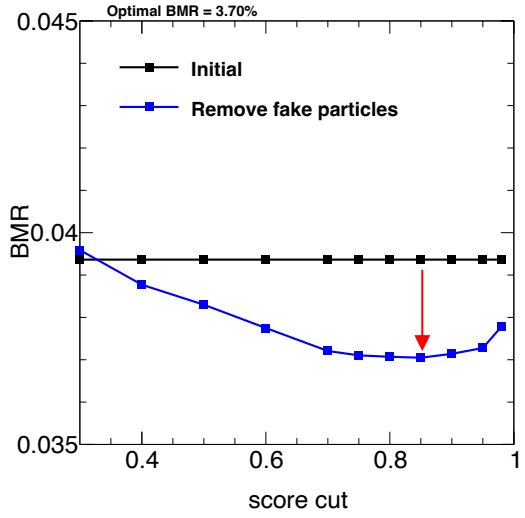
Fake particle identification and BMR

- **Optimal working point (wrt BMR)**
  - Score > 0.85
  - CDR: eff. ~77%, purity ~97.5%
  - TDR: eff. ~56%, purity ~95%



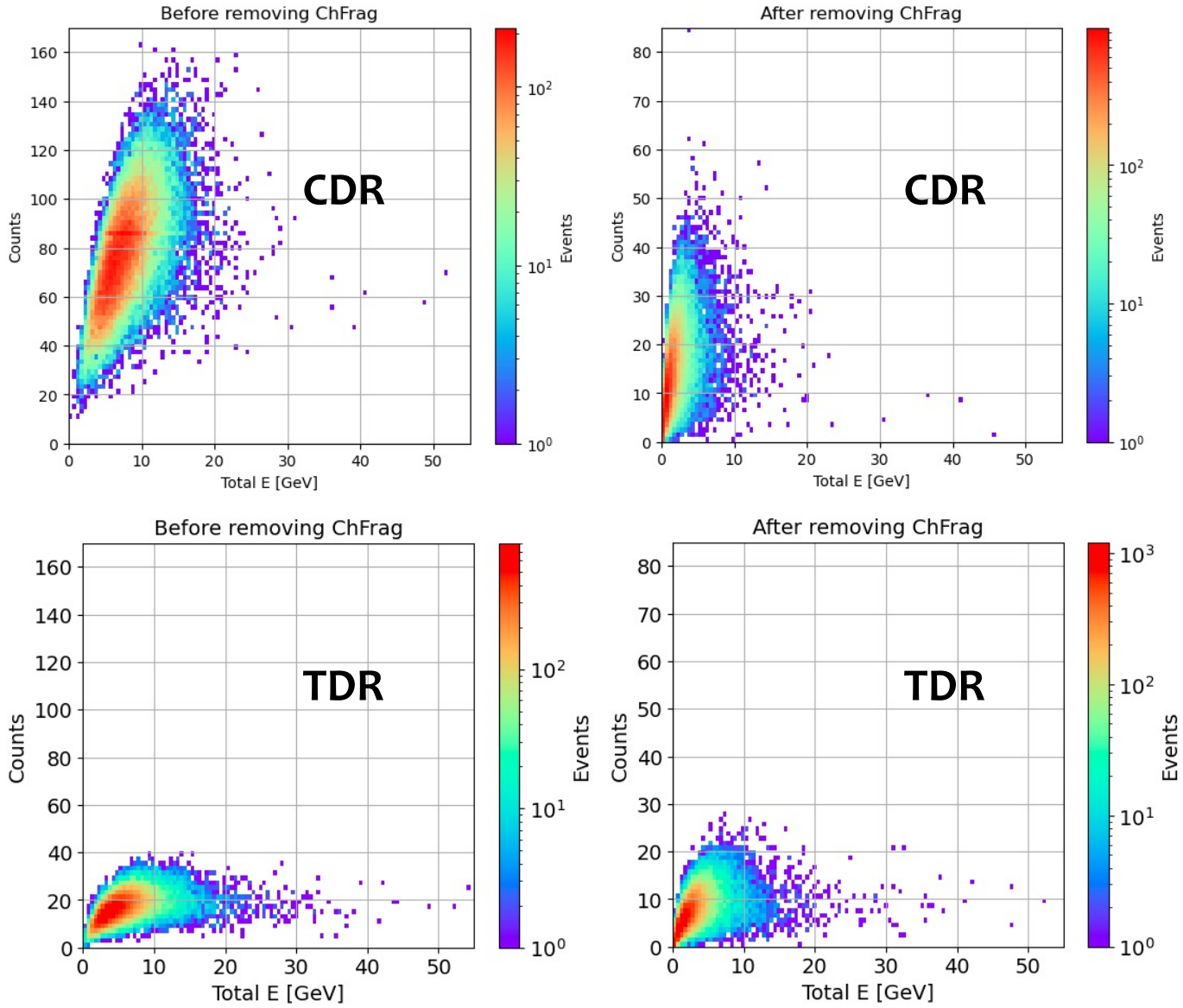






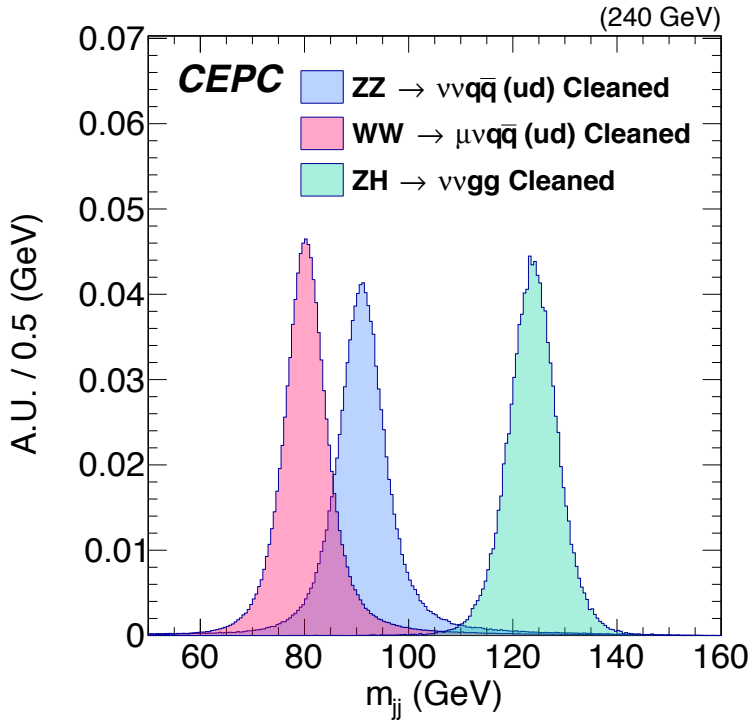
Fake particle identification and BMR

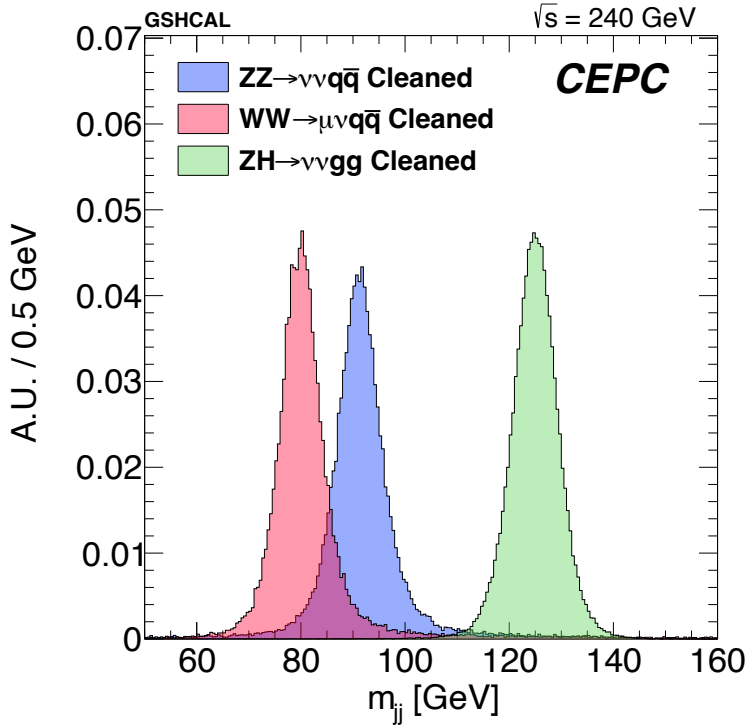
- **Fake particle suppression**
  - Average number
    - CDR: 75 → 10
    - TDR: 17 → 6
  - Double-counted energy
    - CDR: 6 → 0.6 GeV
    - TDR: 3.75 → 1.25 GeV

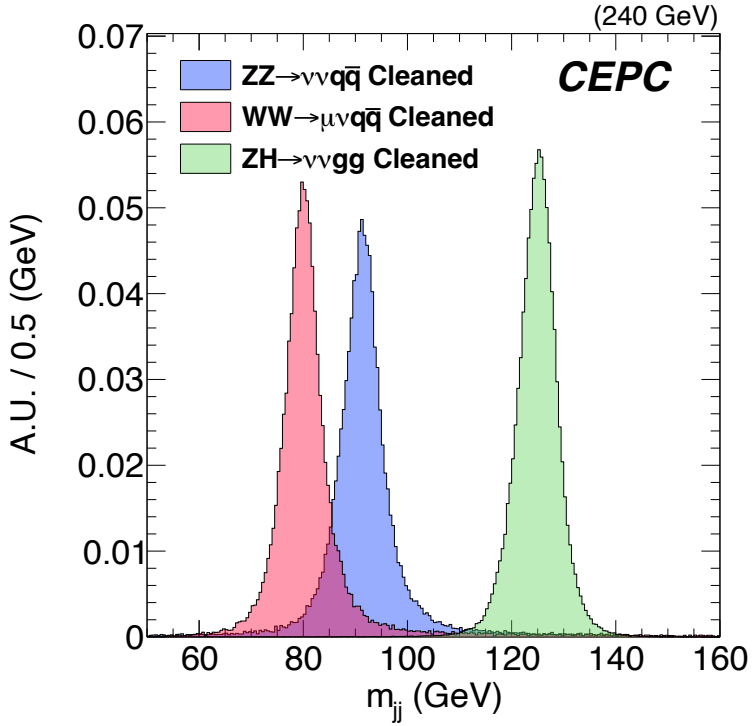


Fake particle identification and BMR

- **BMR improvement (CDR results only)**
  - CDR Baseline ~3.7%
  - Glass-HCAL ~3.3% (relative 10%)
  - Glass-HCAL + fake particle suppression (relative 15%) ~2.75%

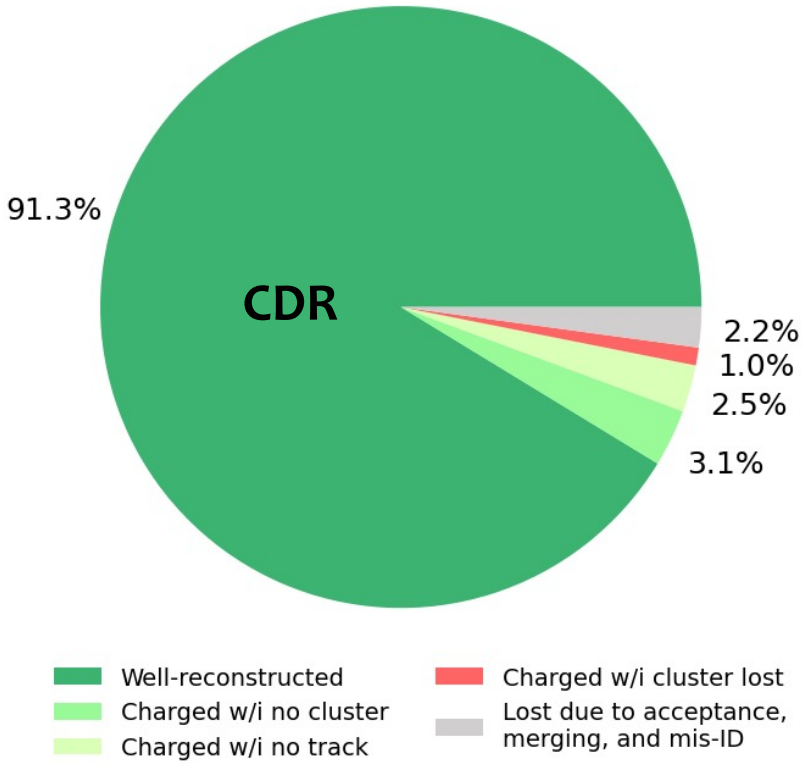


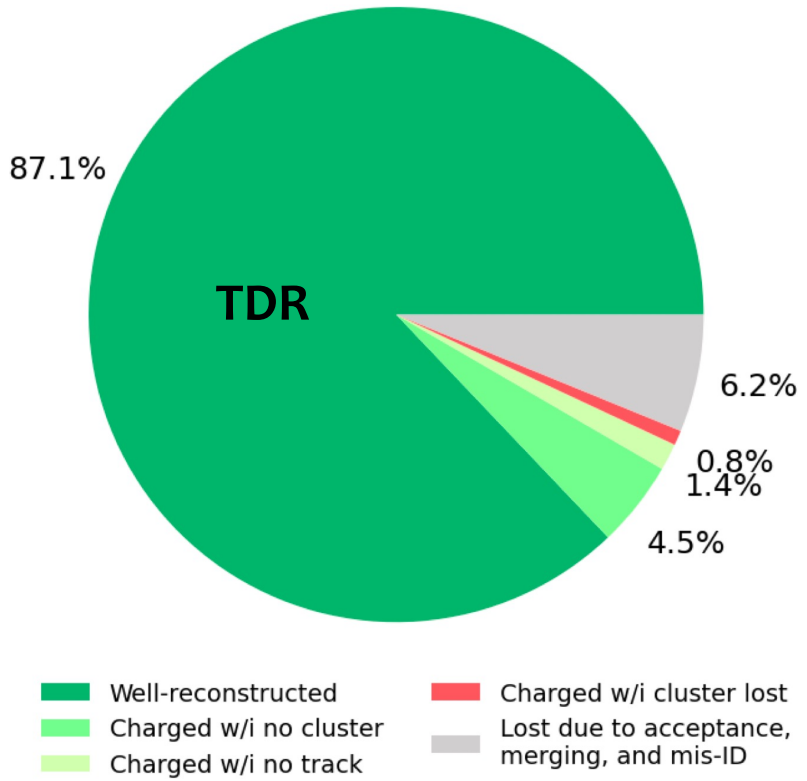




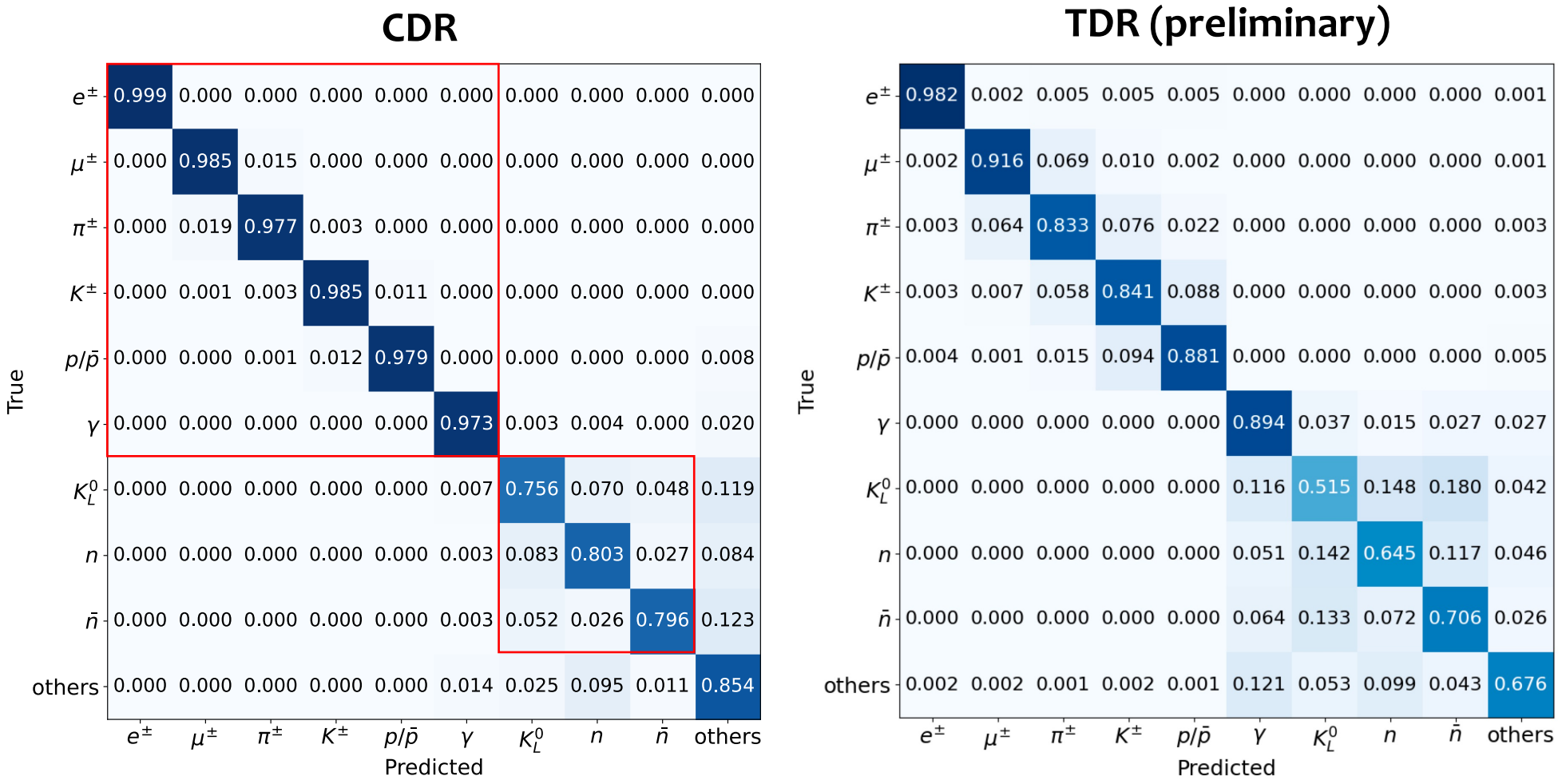
Energy fraction

- **Increased fractions in TDR**
  - Charged w/o cluster
  - Lost (mainly from particle merging due to ECAL long-bar configuration?)



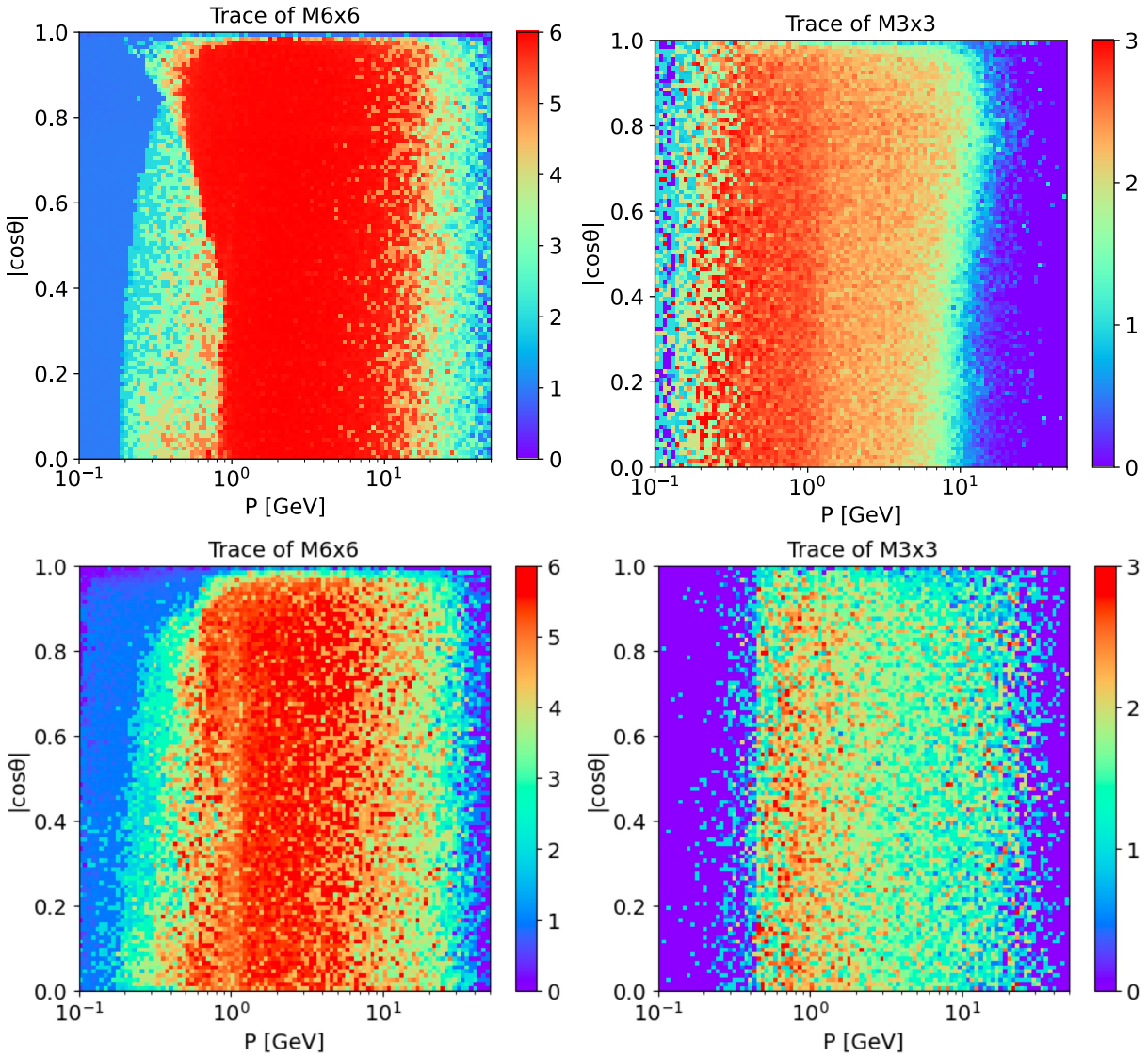


Simultaneous identification of particles



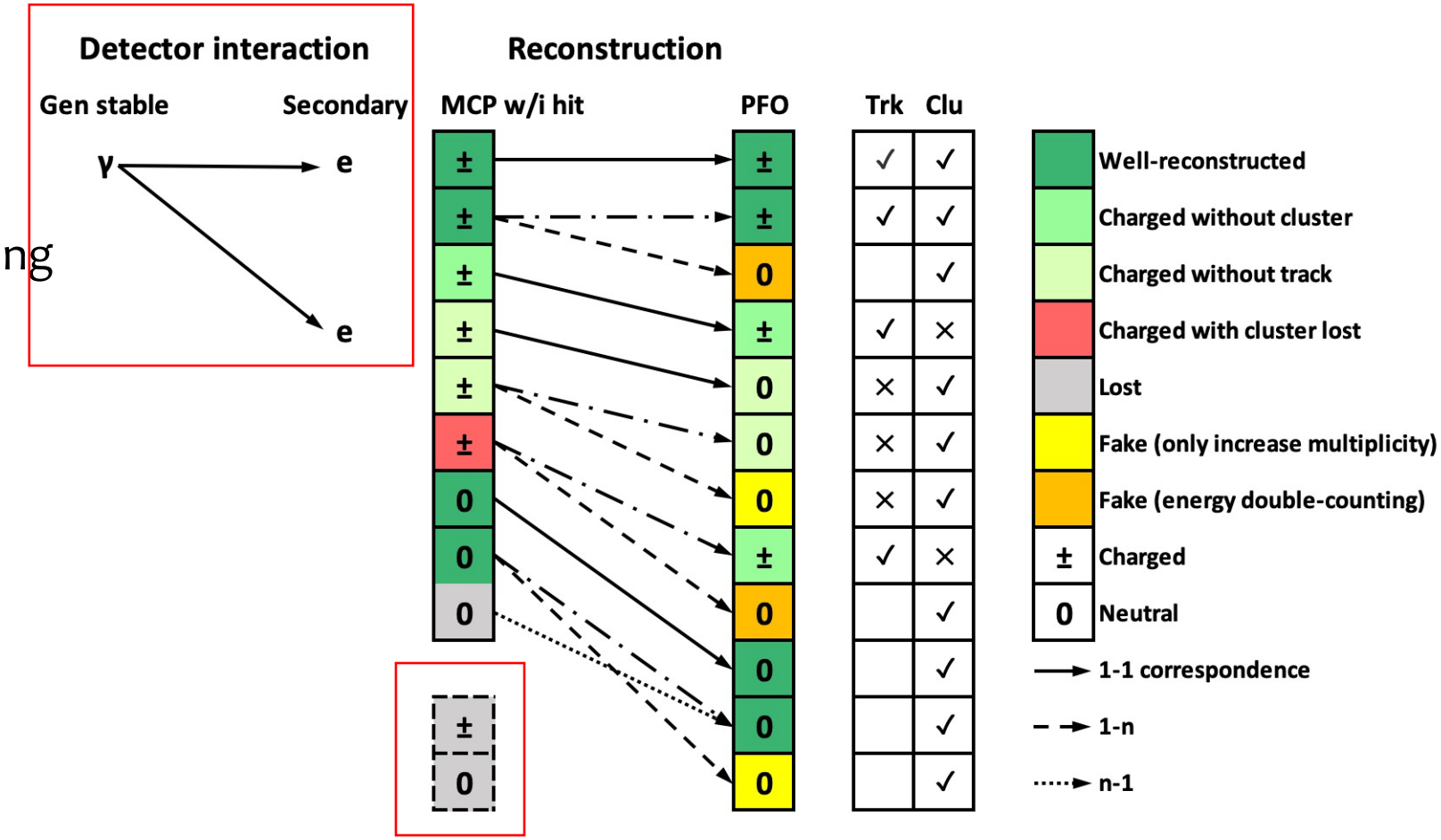
Simultaneous identification of particles

- **Differential efficiency**
  - M6x6 ($e^{\pm}, \mu^{\pm}, \pi^{\pm}, K^{\pm}, p/\bar{p}, \gamma$)
    - high eff. in wide range 1~10/20 GeV
  - M3x3 ($K_L^0, n, \bar{n}$)
    - high eff. in low energy range ~ 1 GeV



Future 3-stage particle mapping

- **Lost decomposition**
  - Acceptance
  - Particle merging
- **Secondary particles**
  - e.g. photon conversion
  - To avoid E double-counting

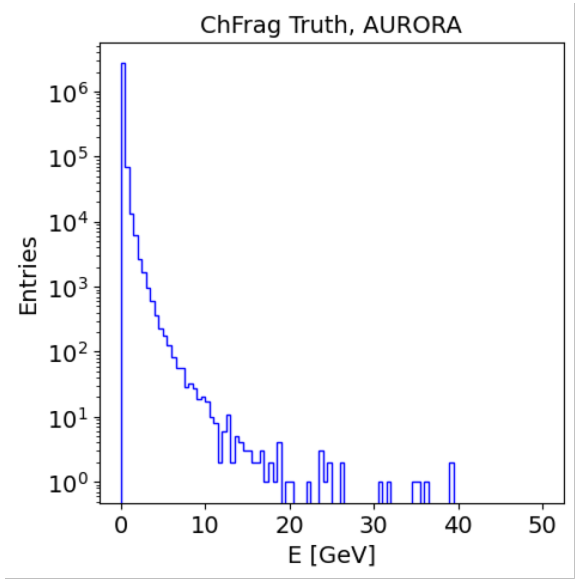


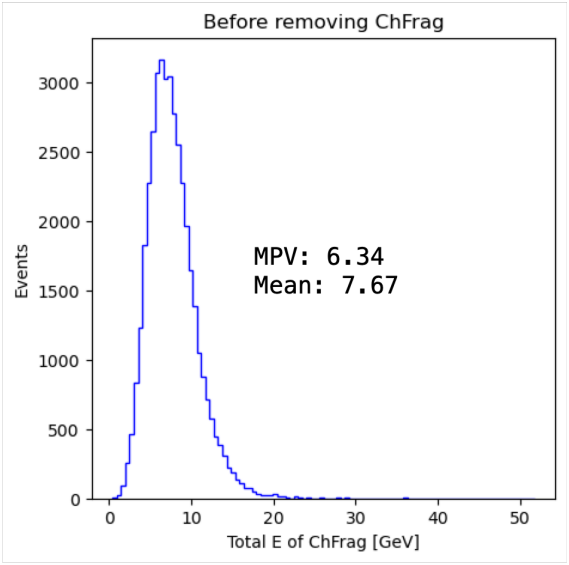
Summary

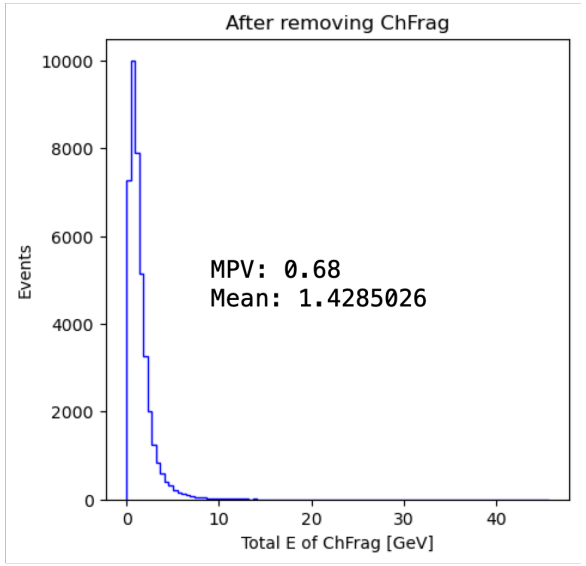
- **1-1 correspondence reconstruction = confusion-free PFA + efficient PID (unified in a same model)**
- **Particle mapping (truth link)**
  - especially in jet environment (with high multiplicity)
    - enable diagnosis on individual particles
    - provide more dimensions (not only BMR) for performance quantification
- **Standardization**
  - Key performance
    - BMR
    - Energy fraction
    - PID efficiency
  - Future 3-stage particle mapping

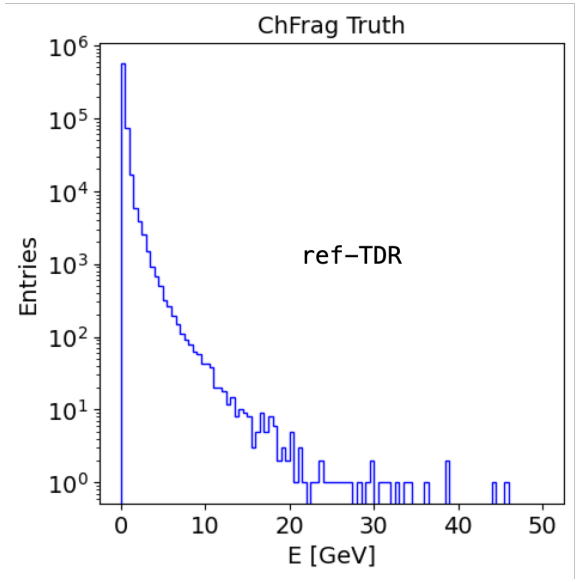
Thank you for your attention!

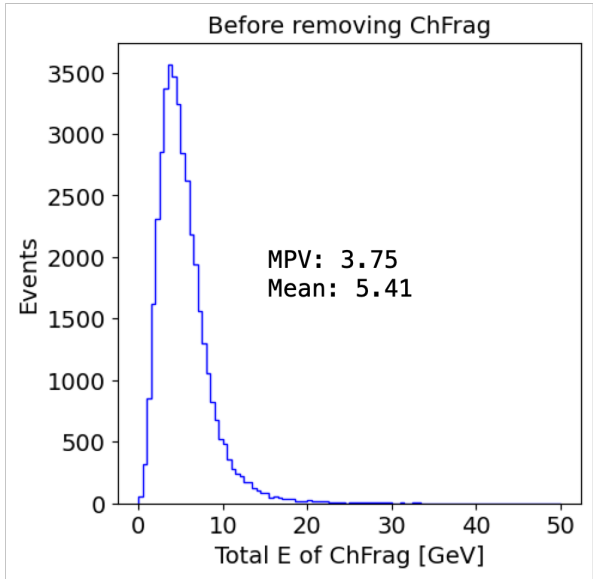
Double-counted energy (ChFrag) distribution

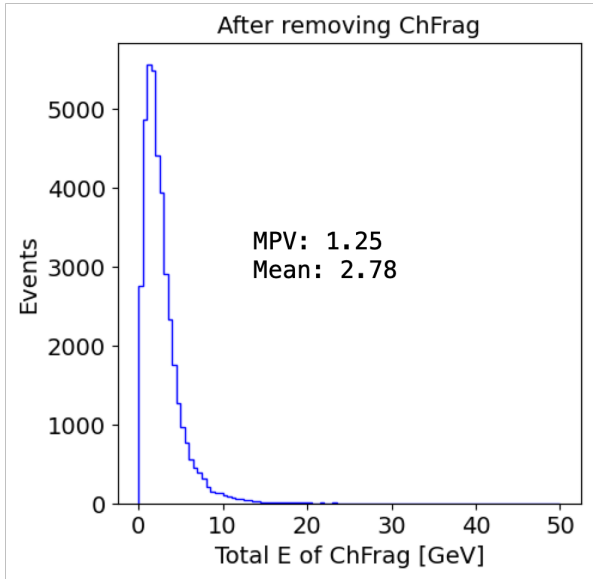






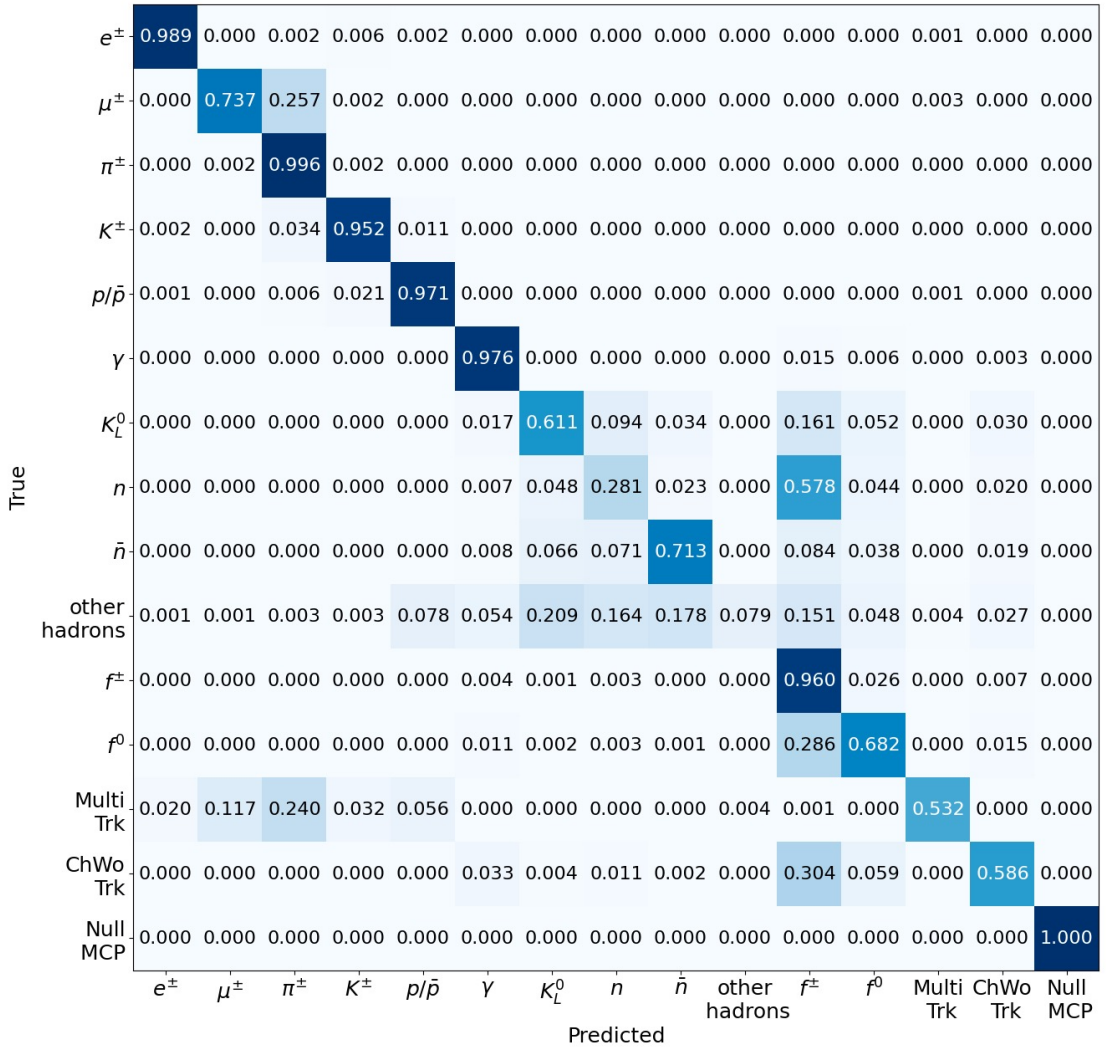




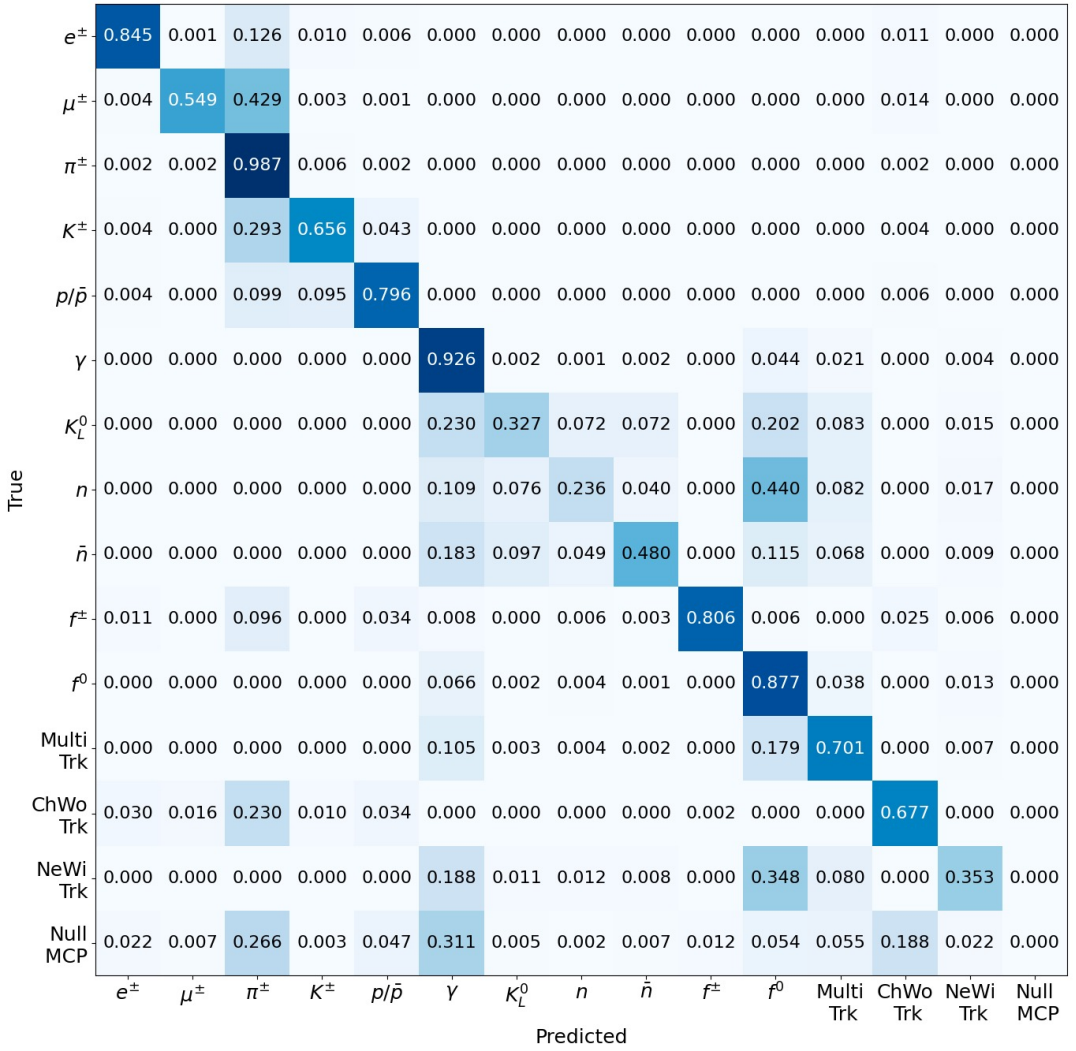


PID matrix using max score

CDR

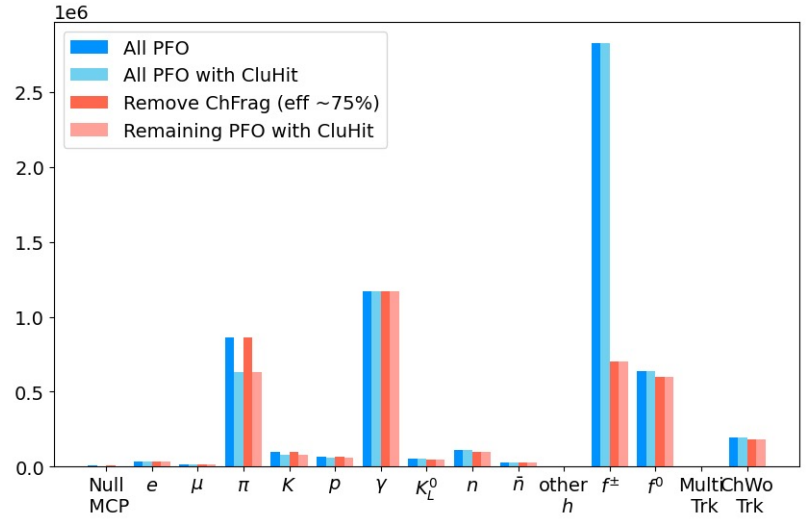


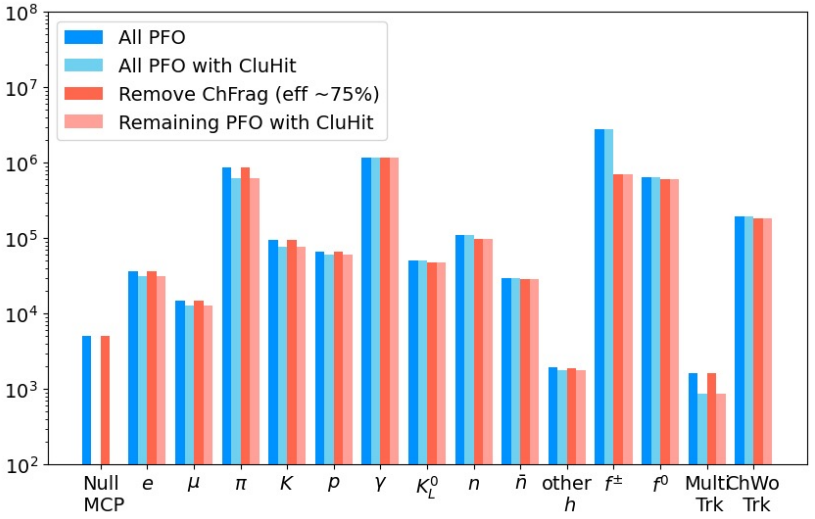
TDR



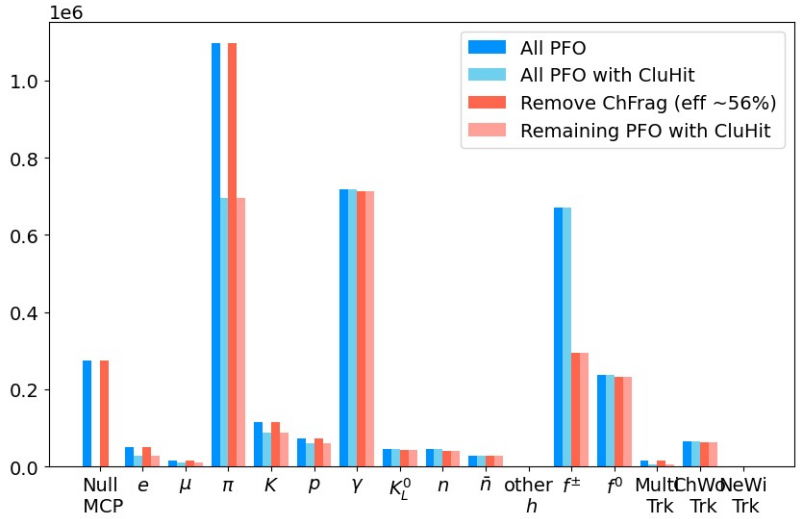
Particle (PFO) multiplicity

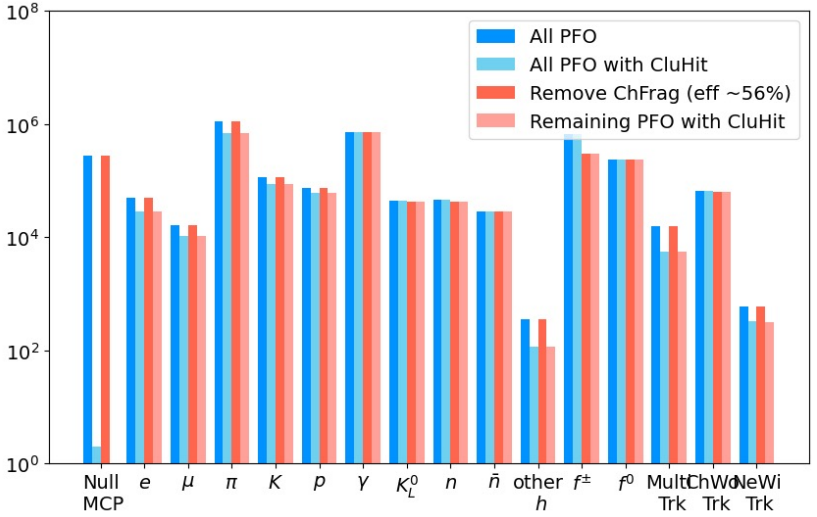
CDR



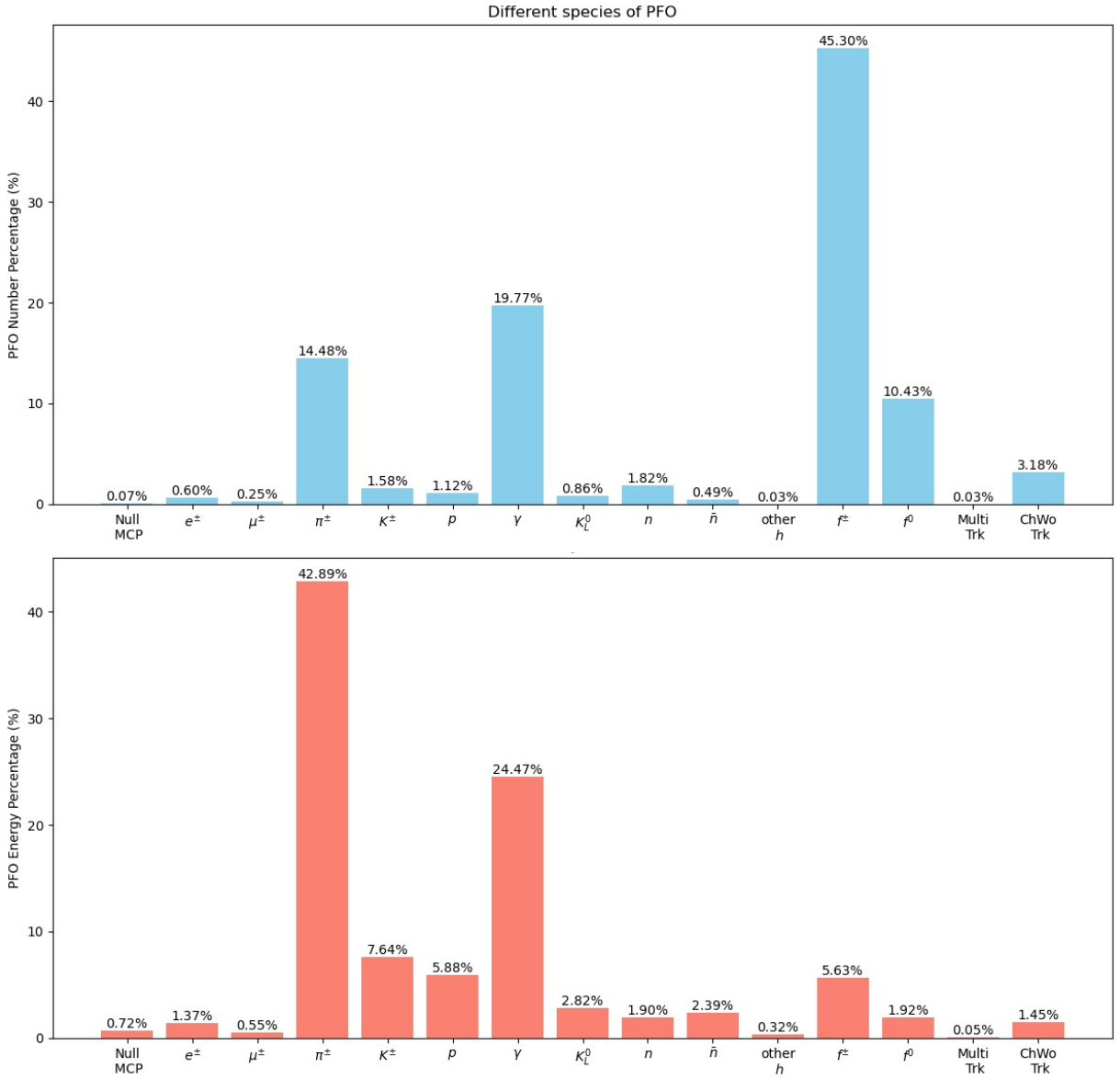


TDR



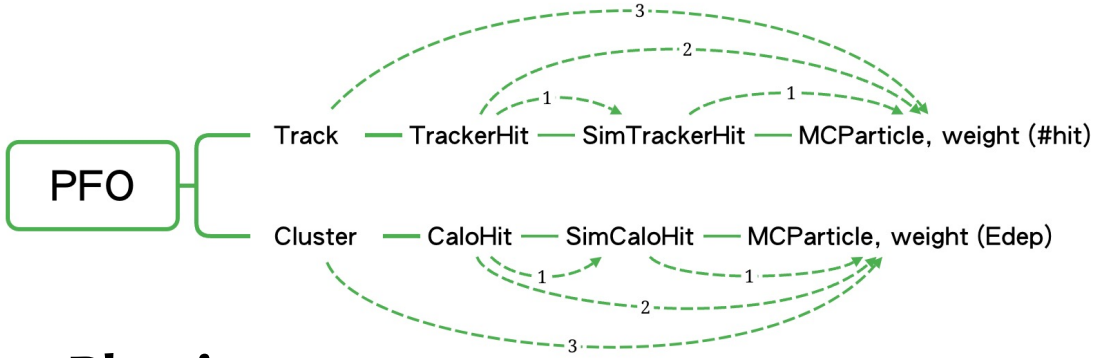


Reconstructed particle categories



Necessary Input

- **Reconstructed samples with truth info**
  - PFO -> Feature variables
    - Track, Track hits
    - Cluster, Cluster hits
  - Truth link -> Mapping & Truth label



- **Physics process**
  - $\nu\nu H, H \rightarrow gg$ (240 GeV)
- **Statistic**
  - 10^6 are expected for training ML model
  - but we can start with ~100 to examine dataset first
    - especially the truth link…

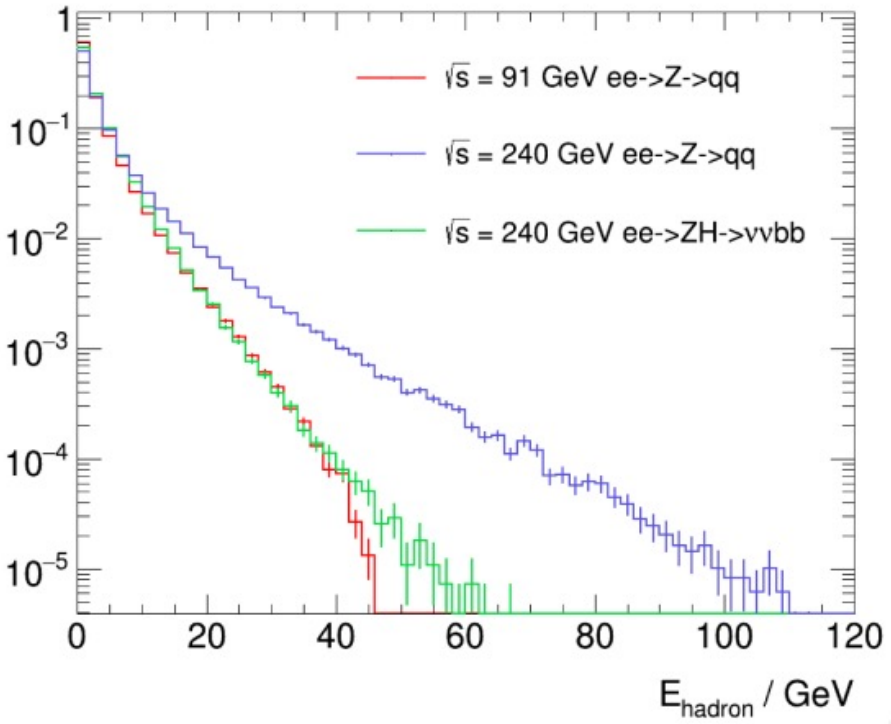
Table A.2: Input variables of ParT.

Object level	Observable variables
Reconstructed particle	4-momentum (E, p_x, p_y, p_z) Direction (θ, ϕ) Number of tracks and clusters
Track	Number of hits Endpoint position 3-momentum ($\lvert\vec{p}\rvert$, p_x, p_y, p_z, p_T) $\mathrm{d}E/\mathrm{d}x$ (mean of 5–85% truncation and quartiles)
Cluster	Number of hits Energy Position of shower starting point Position of center of gravity Fractal dimension [56] Second moment (M_2) Distance between ECAL inner surface and shower starting point Distance between ECAL inner surface and center of gravity Distance between ECAL inner surface and the innermost hit Distance between ECAL inner surface and the outermost hit Maximum distance between cluster hits and the track helix (for charged particles) Maximum distance between cluster hits to the axis from the innermost hit to the center of gravity Average distance between cluster hits to the axis from the innermost hit to the center of gravity Hit time spectrum (the fastest time and quintiles)
Closest charged cluster	Minimum distance between cluster hits of each other Number of hits Energy Ratio of E_{cluster} to p_{track}

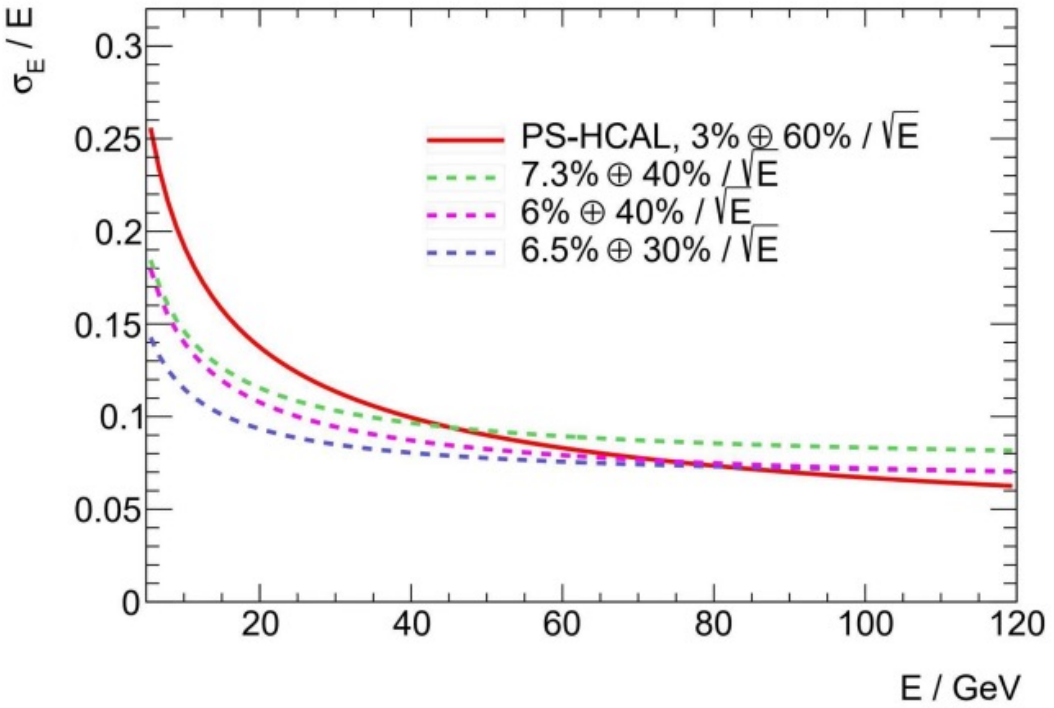
Detector geometry

Parameters	SiWECAL + SDHCAL (Baseline)	SiWECAL + GSHCAL	CSECAL + GSHCAL
ECAL Material	Si + W	Si + W	BGO (Homogeneous)
ECAL Transverse cell size	1×1 cm^2	1×1 cm^2	1×1 cm^2
ECAL Number of layers	30	30	27
ECAL Total thickness	24 X_0	24 X_0	24 X_0
ECAL Thickness/layer	Si 0.5 mm (30 layers) W 2.1 mm (20 layers) W 4.2 mm (10 layers)	Si 0.5 mm (30 layers) W 2.1 mm (20 layers) W 4.2 mm (10 layers)	10 mm
HCAL Material	GRPC	Glass + Steel	Glass + Steel
HCAL Transverse cell size	1×1 cm^2	2×2 cm^2	2×2 cm^2
HCAL Number of layers	40	48	48
HCAL Total thickness	5 λ	6 λ	6 λ
HCAL Thickness/layer	0.125 λ 3 mm GRPC + 3 mm Electronics + 20 mm Steel	0.125 λ 10 mm Glass + 13.85 mm Steel	0.125 λ 10 mm Glass + 13.85 mm Steel
HCAL Glass density	-	6 g/cm^3	6 g/cm^3

Glass-HCAL



- E_{hadron} < ~ 100 GeV, typically < 60 GeV



- Energy resolution of GS-HCAL is better than that of PS-HCAL for E < 80 GeV