Quantum Computing and Machine Learning Workshop 2025

1-1 correspondence reconstruction at electron-position Higgs factories

—— application of ML in HEP

https://inspirehep.net/literature/2847034

王悦心，梁浩，朱永峰，车逾之
夏欣，曲慧麟，周辰，阮曼奇

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

2025年8月22日

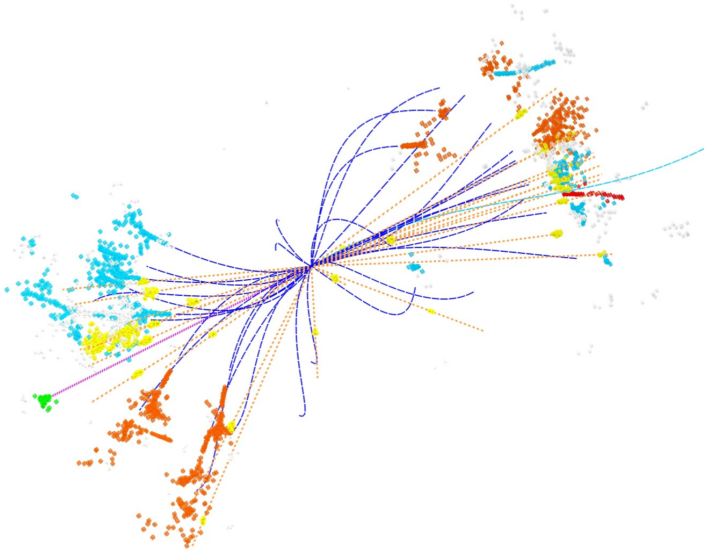
CEPC

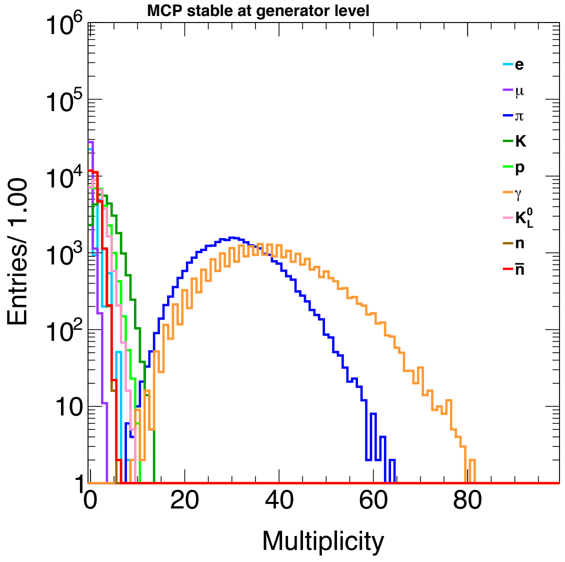
Particle reconstruction at e^+e^- Higgs factory

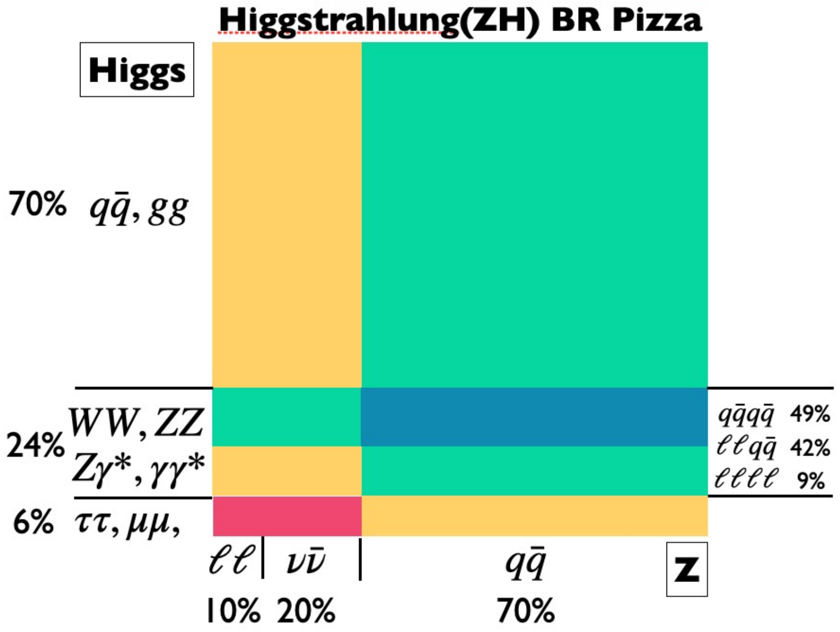
- **Reconstruction in collider experiments**
  - Interpret **detector hits** into observables (type, E, P, vertex…) of original **particles**.
  - Ultimate goal: **1-1 correspondence** between **truth & reconstructed** particles.
  - Difficulty of achieving this goal highly depends on collision environment.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

~ 97% with Jets!

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

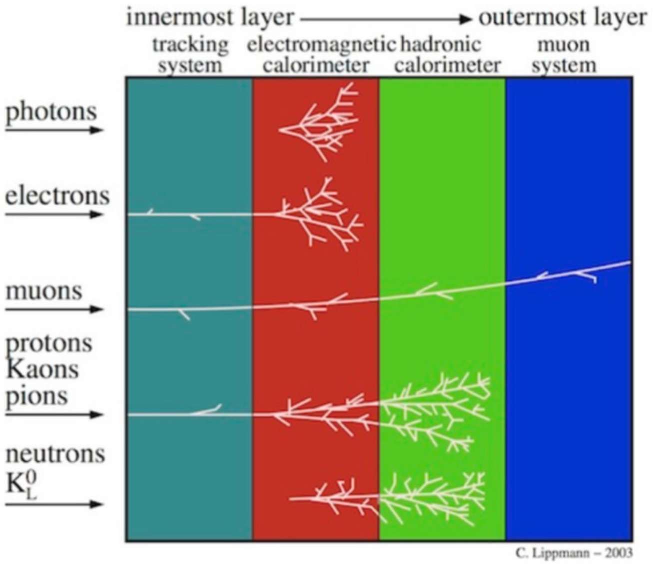


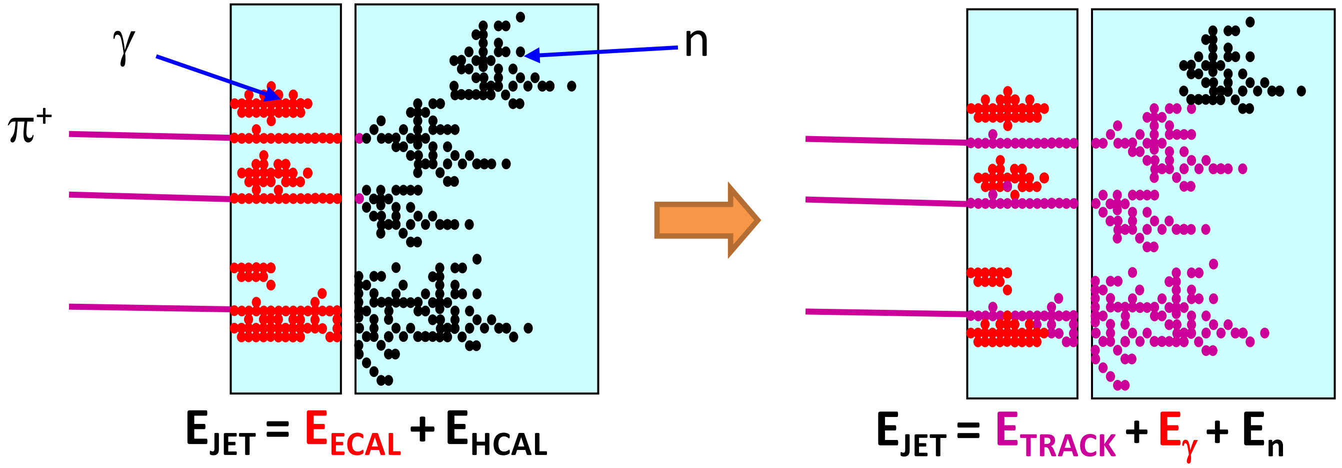




Particle Flow reconstruction

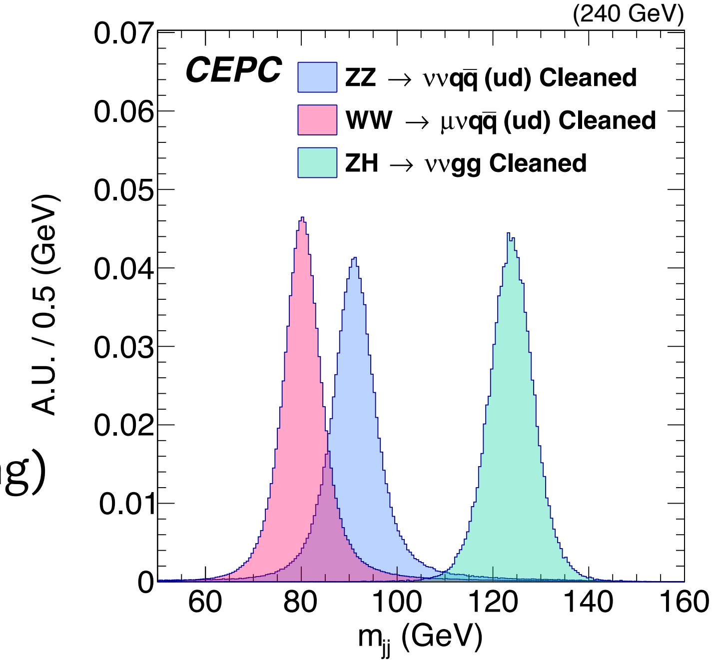
- **Basic idea**
  - efficiently separate and identify different types of particles
  - measure their energy/momentum with the optimal sub-detector
- **Significantly improve jet energy resolution**
  - $E_{\text{Jet}} = E_{\text{Track}}(65\%) + E_{\gamma}(25\%) + E_{h^0}(10\%)$
  - Tracker momentum resolution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

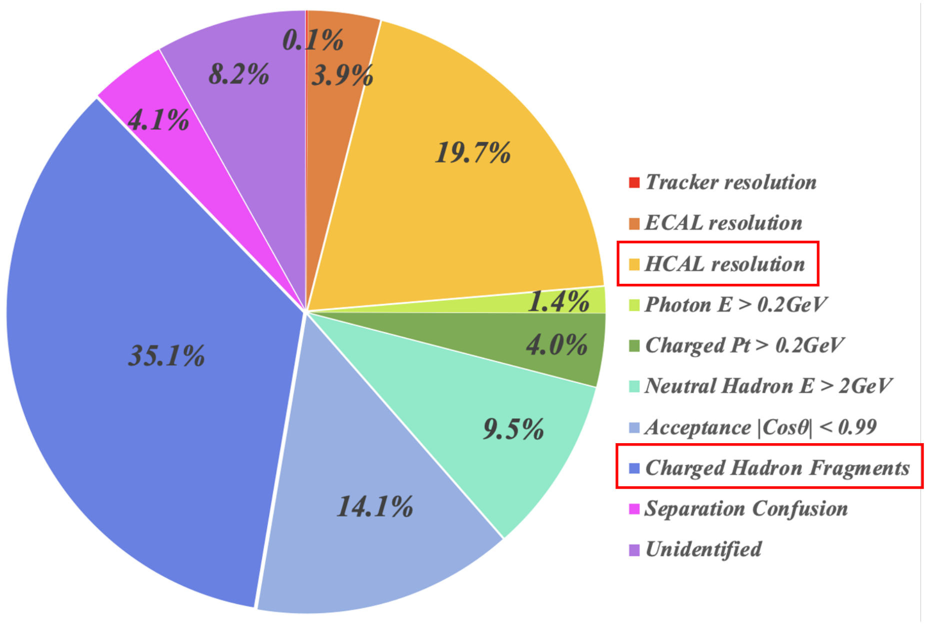


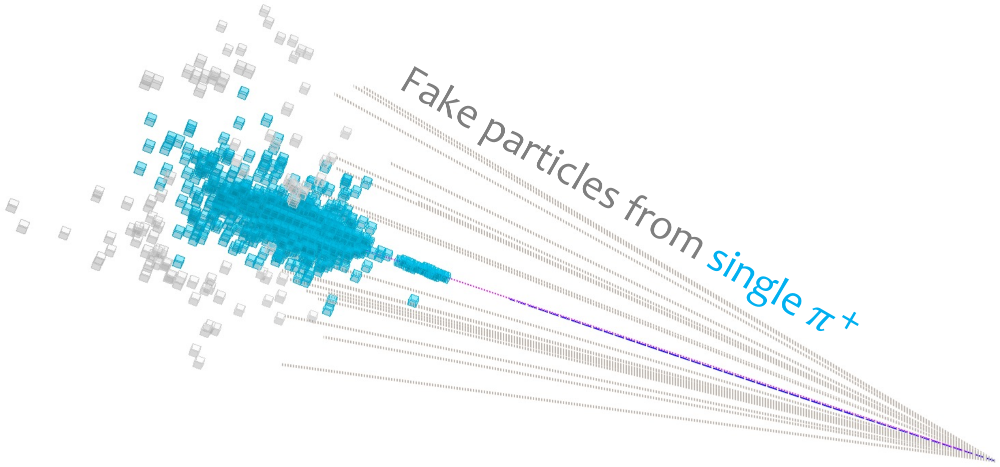


Particle Flow reconstruction

- **Performance at CEPC CDR phase**
  - Quantified by Boson Mass Resolution (BMR)
    - di-jet event (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

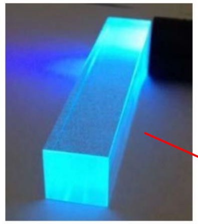


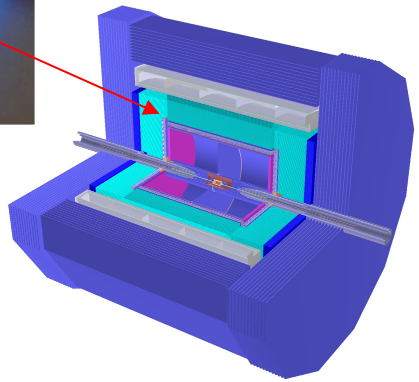


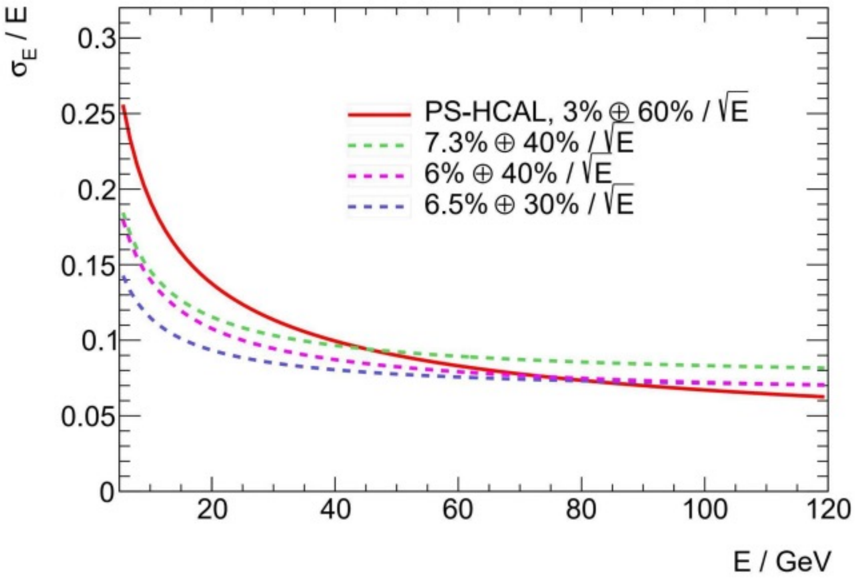


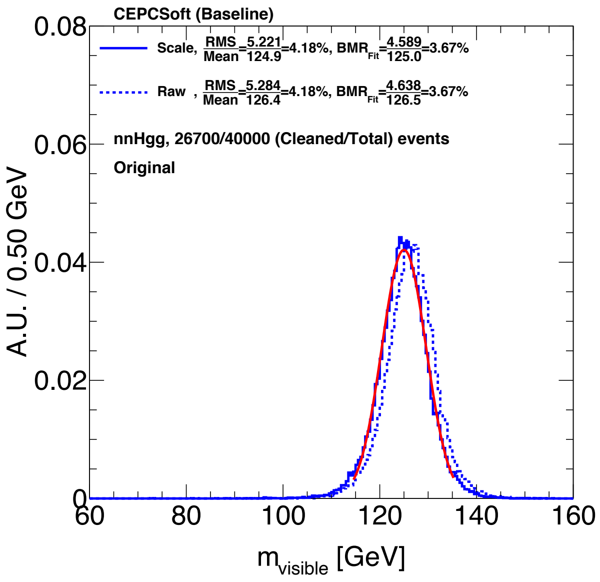
Improve hadronic resolution with Glass-HCAL

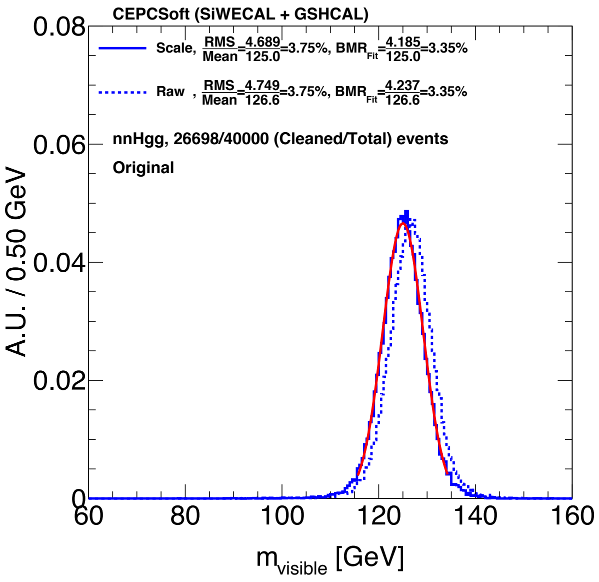
- **Glass Scintillator HCAL (GSHCAL)** [NIMA 1059 (2024) 168944]
  - Glass advantages
    - higher density → higher sampling fraction
    - doping with neutron-sensitive elements (Gd) : improve hadronic response
    - low cost
  - Significantly improvement in
    - Hadronic energy resolution: traditional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


Glass Scintillator



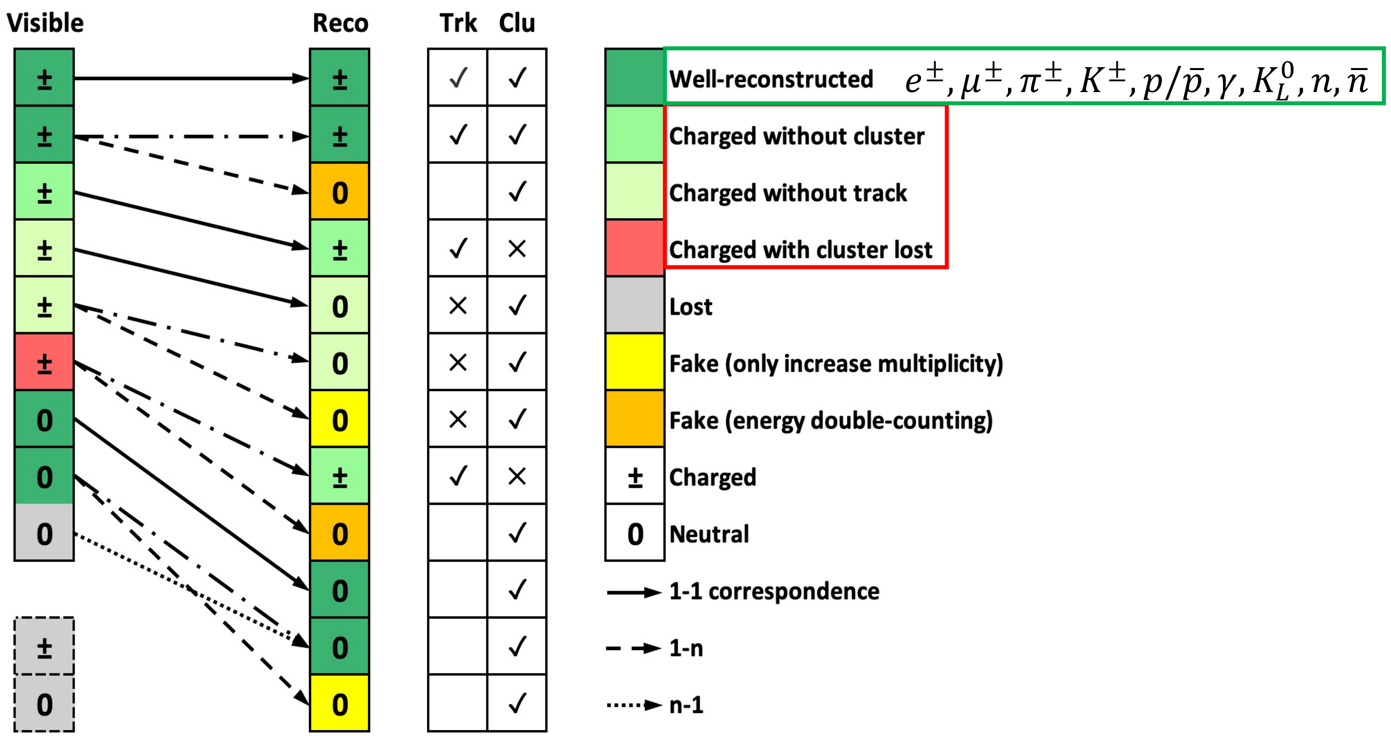






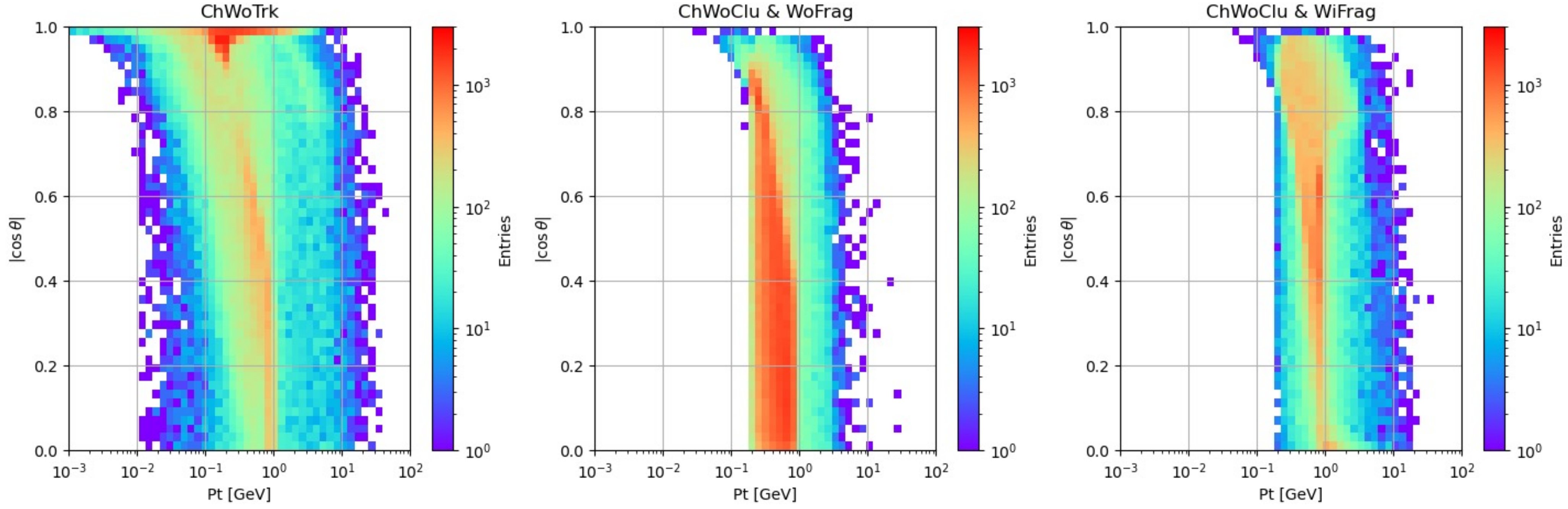
Particle mapping & classification of reconstructed particles

- **Build mapping between truth visible & reconstructed particles**
  - **Truth links** between truth particles and track/cluster hits that compose reco. particles
- **15 categories of reconstructed particles**
  - 1-1 correspondence preserved 10 categories: $e^{\pm}, \mu^{\pm}, \pi^{\pm}, K^{\pm}, p/\bar{p}, \gamma, K_L^0, n, \bar{n}, others$
  - Confusions: fake particles, charged w/o track, etc

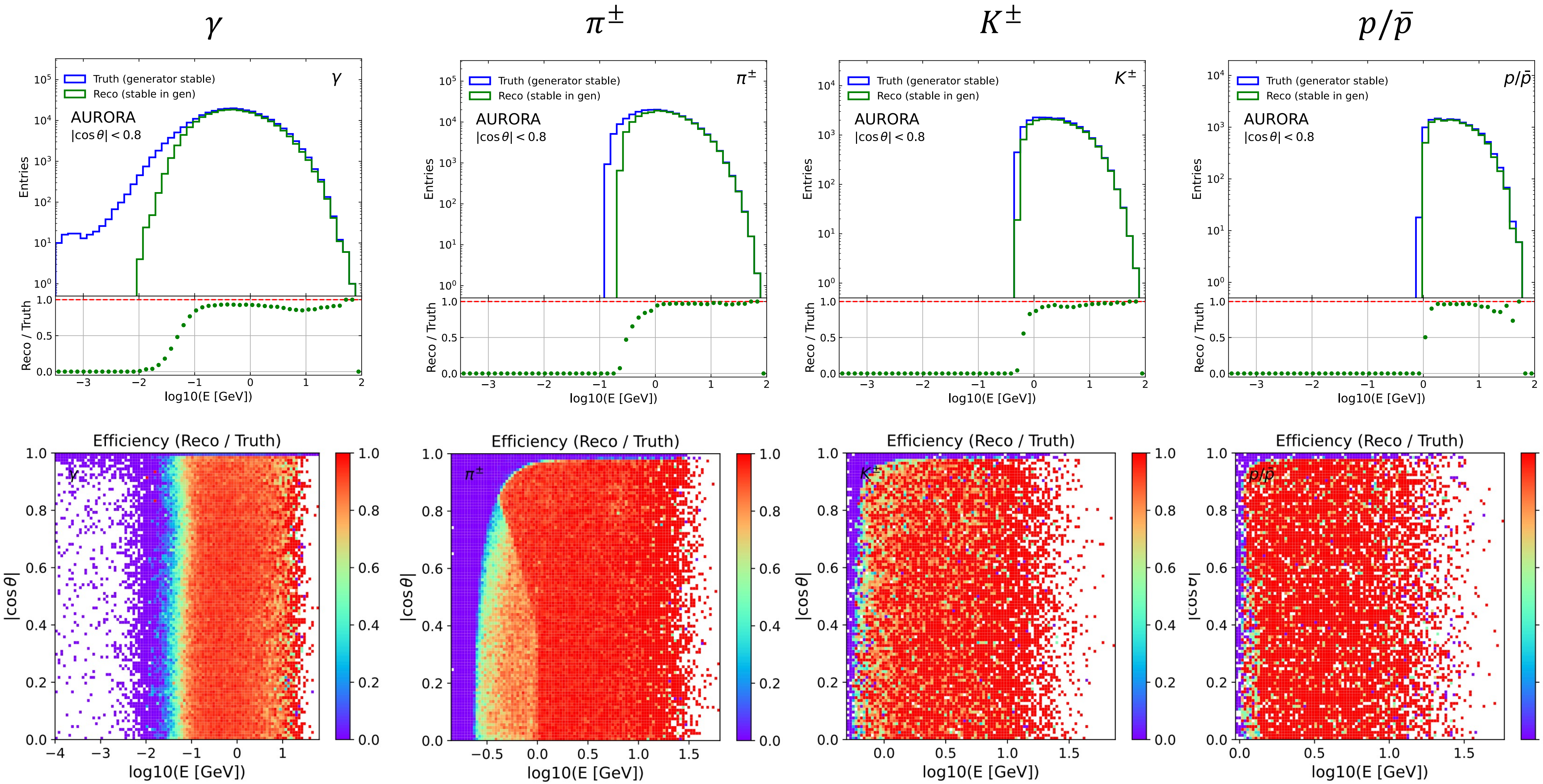


Diagnosis: charged particle w/o track or cluster

- **Charged particle w/o track or cluster**
  - Almost in forward and low Pt regions



Diagnosis: particle reconstruction efficiency



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 particle type
  - ~ 2.2M parameters
- **Sample**
  - Total 10^6 full simulation $\nu\nu H, H \to gg$ (240 GeV)
  - Training : Validation : Test = 6:2:2

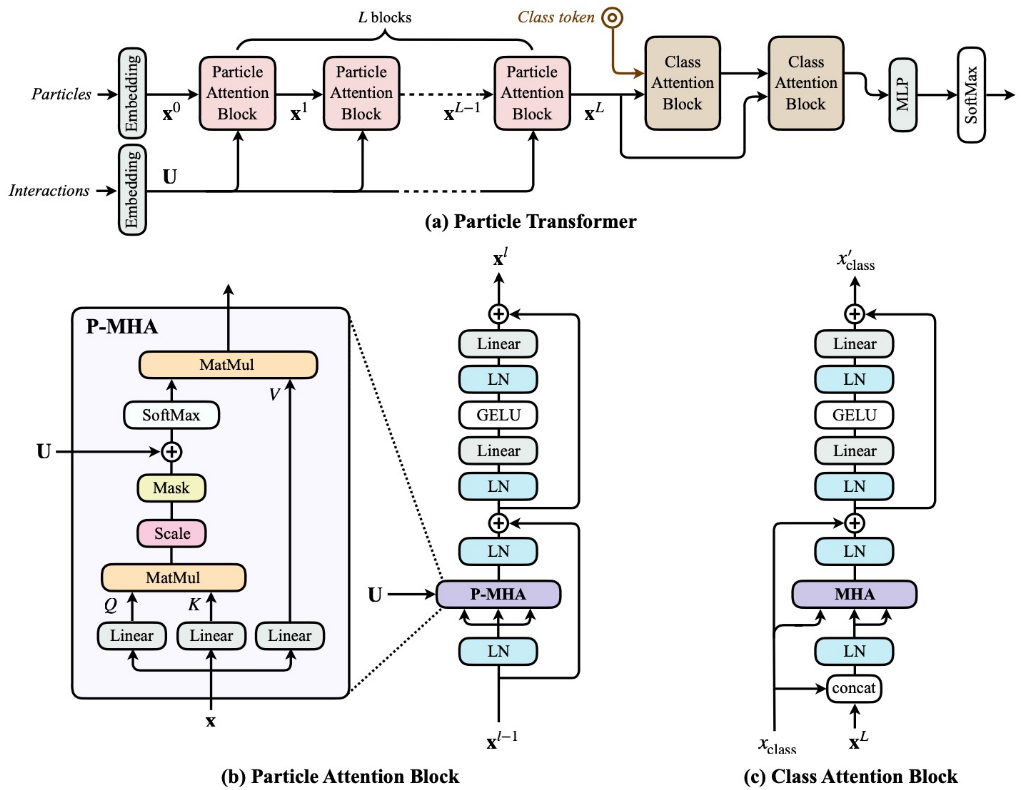


Figure 3. The architecture of (a) Particle Transformer (b) Particle Attention Block (c) Class Attention Block.

Input feature variables

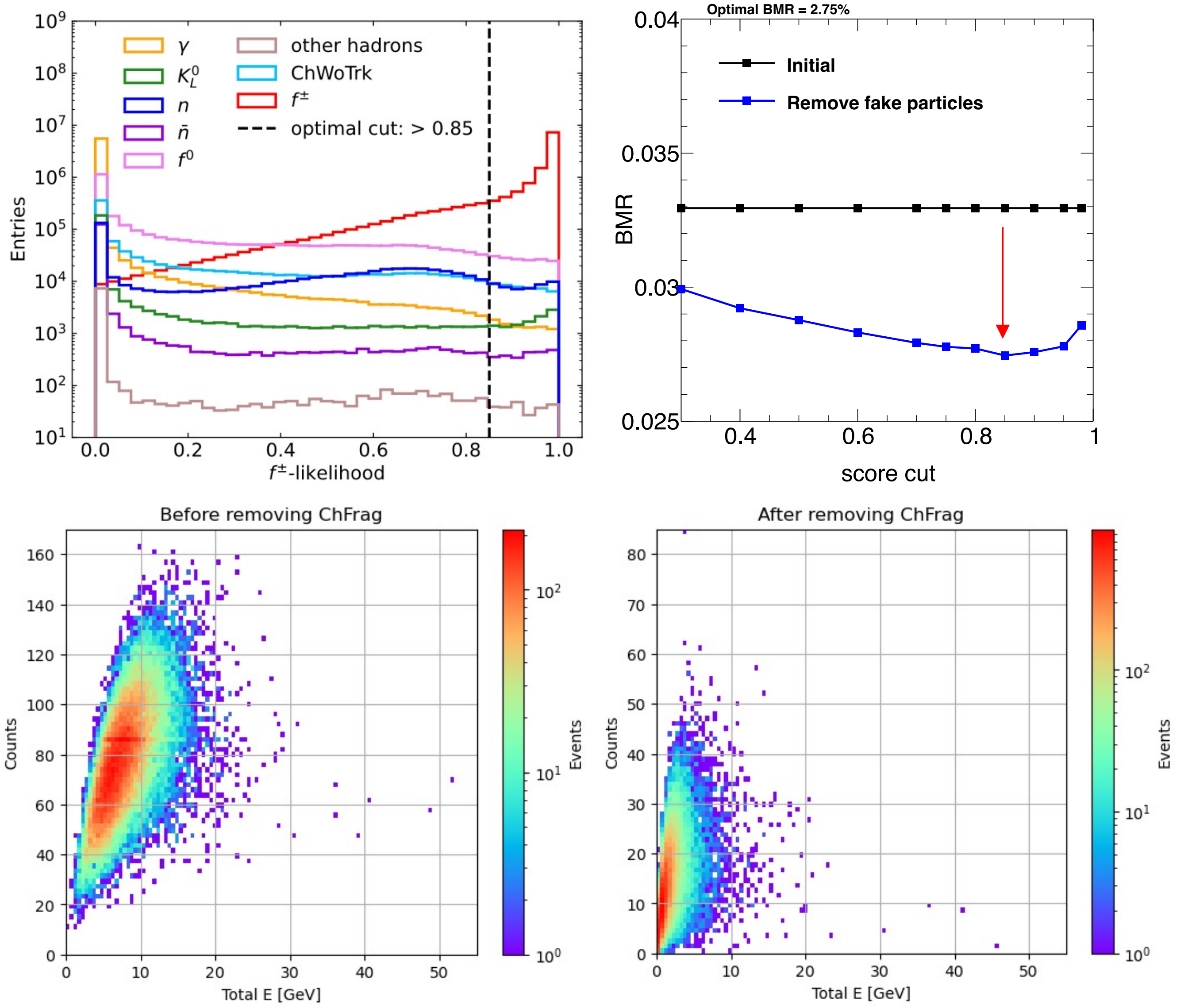
- **Input features at particle, track, cluster level**

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 ($\|\vec{p}\|, p_x, p_y, p_z, p_T$) $\mathrm{d}E/\mathrm{d}x$ (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) **5D calorimeter** Minimum distance between cluster hits of each other
Closest charged cluster	Number of hits Energy Ratio of E_{cluster} to p_{track}

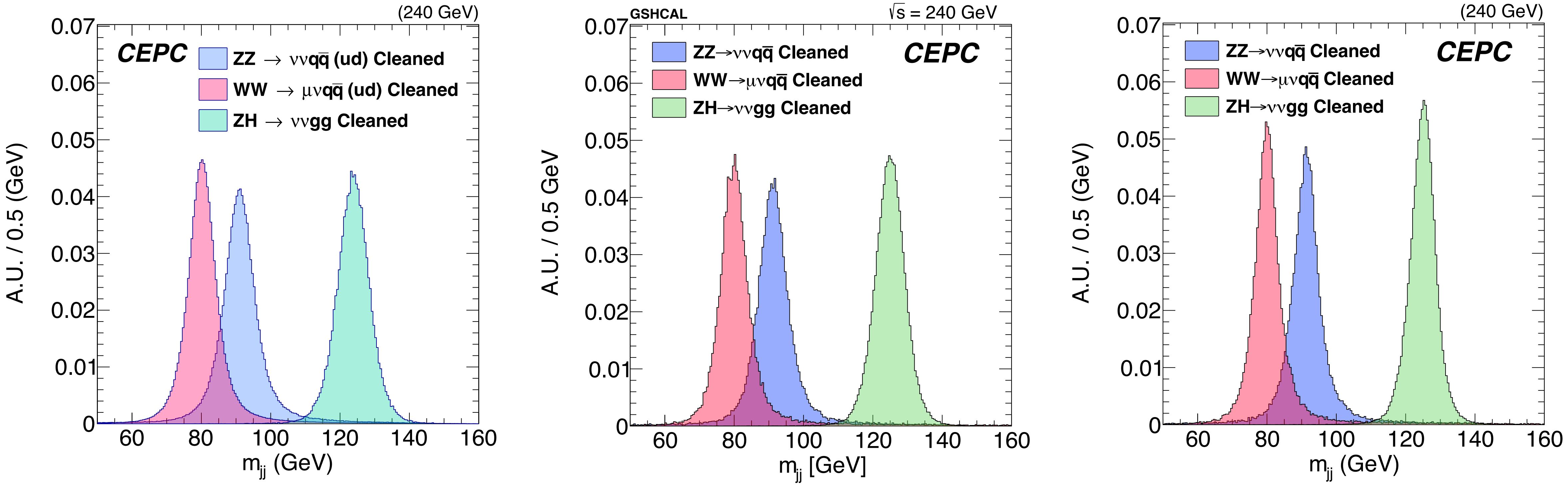
Fake particle identification and BMR

- **Optimal working point (wrt BMR)**
  - score > 0.85
  - efficiency ~77%, purity ~97.5%
- **Fake particle suppression**
  - average number: 75 → 10
  - double-counted energy: 6 → 0.6 GeV



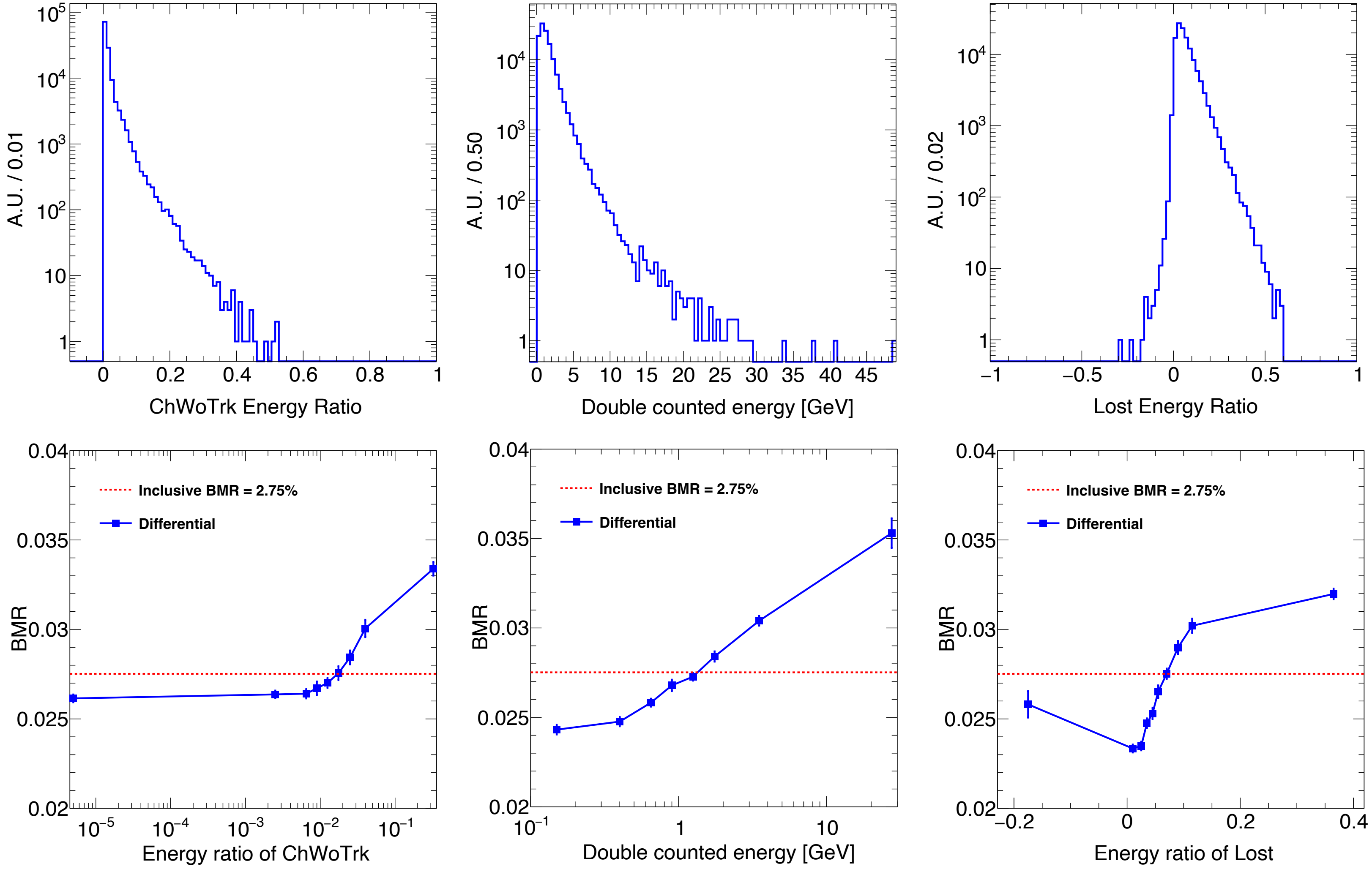
Fake particle identification and BMR

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



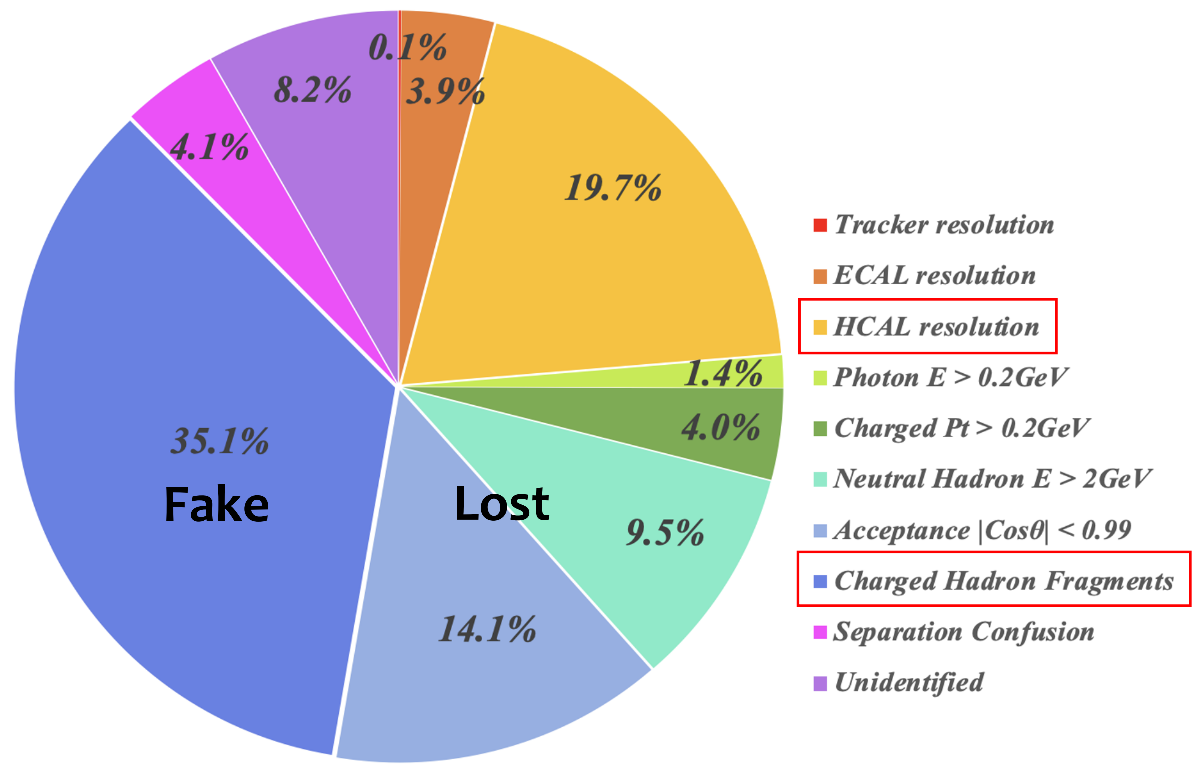
BMR decomposition

➢ **Differential BMR vs ChWoTrk, fake, lost particles**

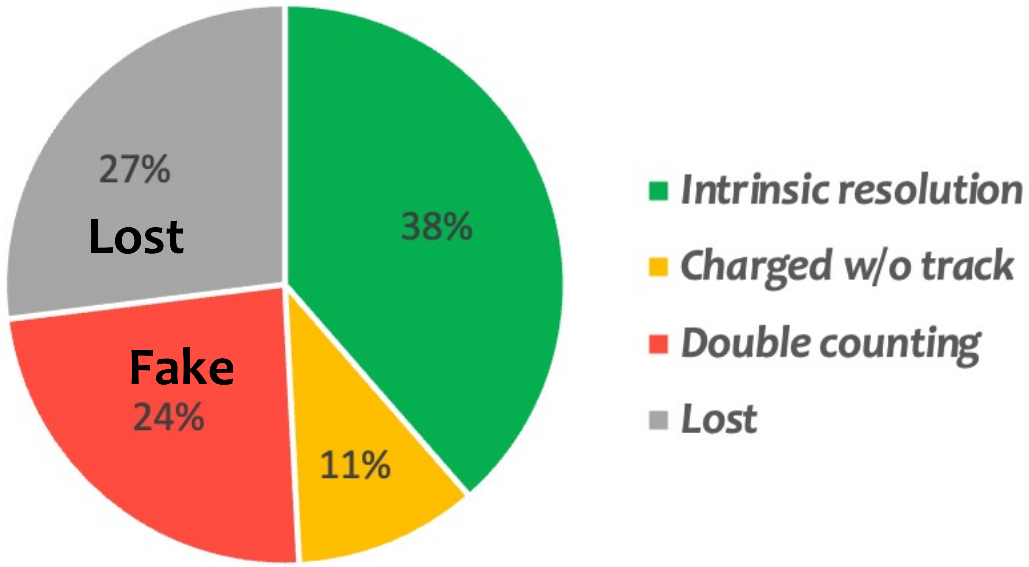


BMR decomposition

- **BMR 3.7% → 2.75%**
- **Bottleneck shift**
  - Fake particle → Intrinsic resolution & Lost particle



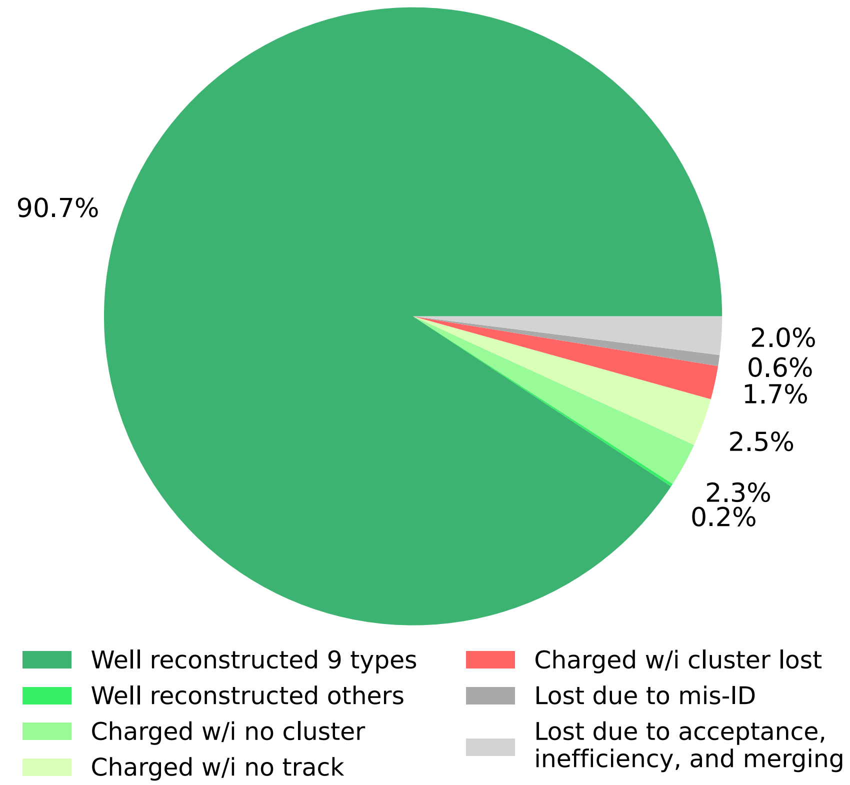
BMR ~3.7%

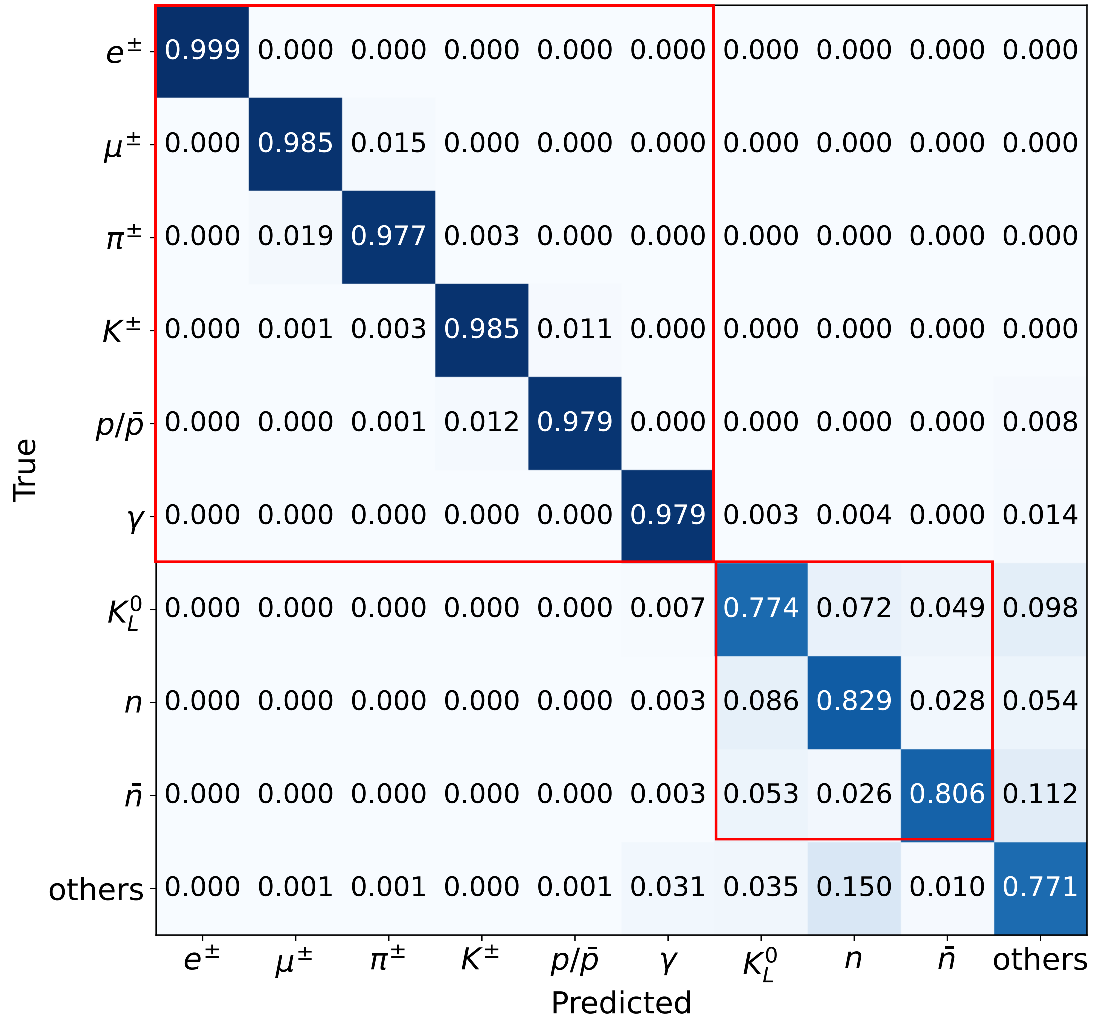


BMR ~2.75%

Simultaneous identification of particles

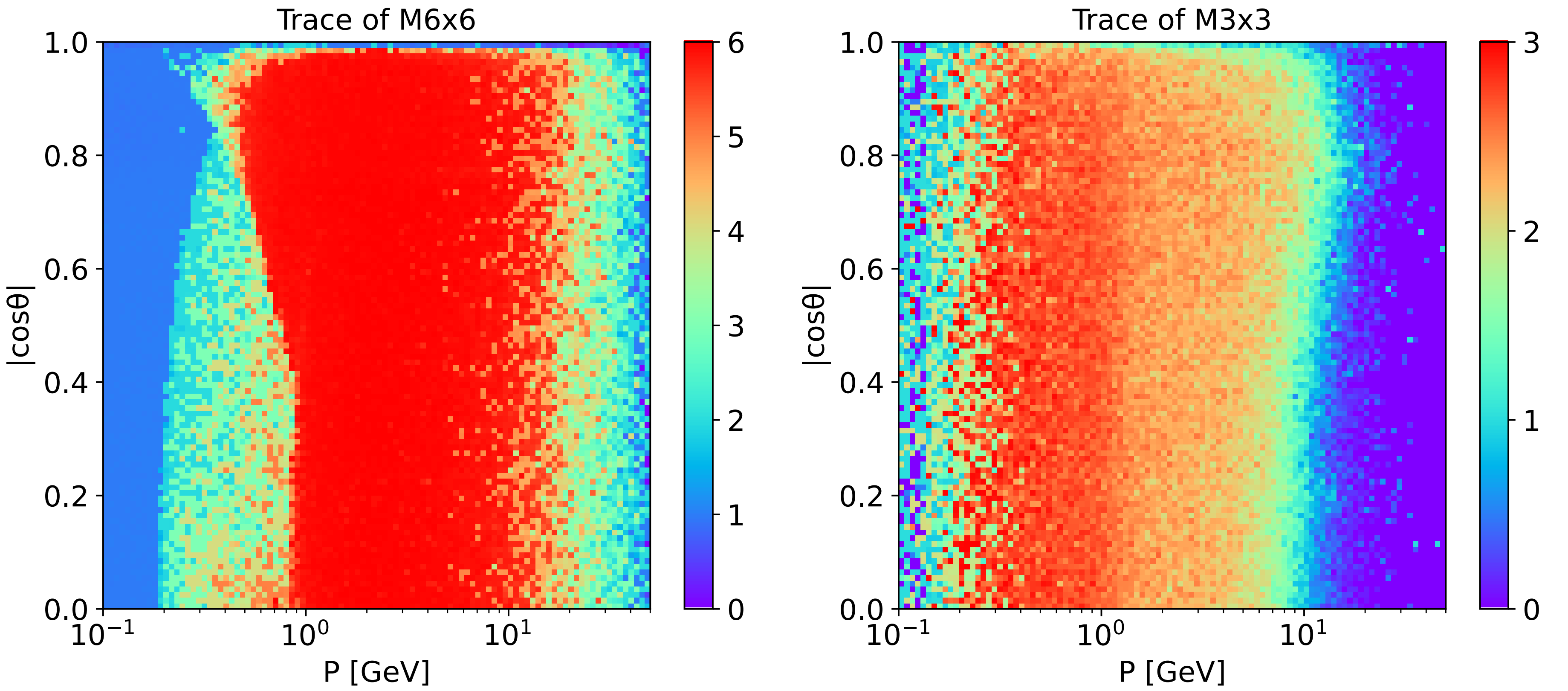
- **Energy fraction of 1-1 correspondence types ~90%**
- **Simultaneous identification efficiency**
  - M6x6 $(e^{\pm}, \mu^{\pm}, \pi^{\pm}, K^{\pm}, p/\bar{p}, \gamma)$: >97%
  - M3x3 $(K_L^0, n, \bar{n})$: >75~80%





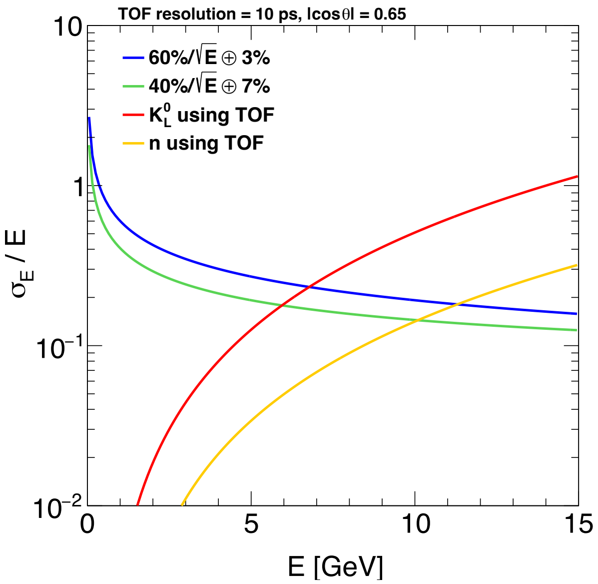
Simultaneous identification of particles

- **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

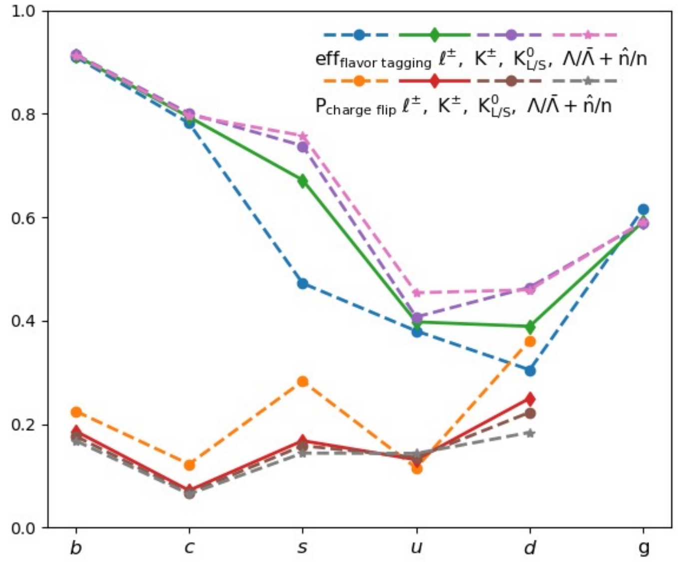


Prospects

- **TOF enhanced neutral hadron energy measurement**
  - efficient separation of K_L^0 & $n/\bar{n}$
    - TOF + PID (mass) → E
  - much more precise low-energy measurements

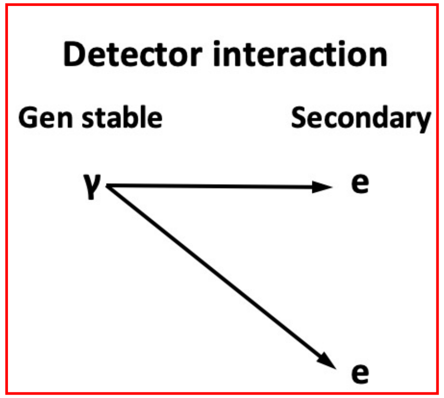


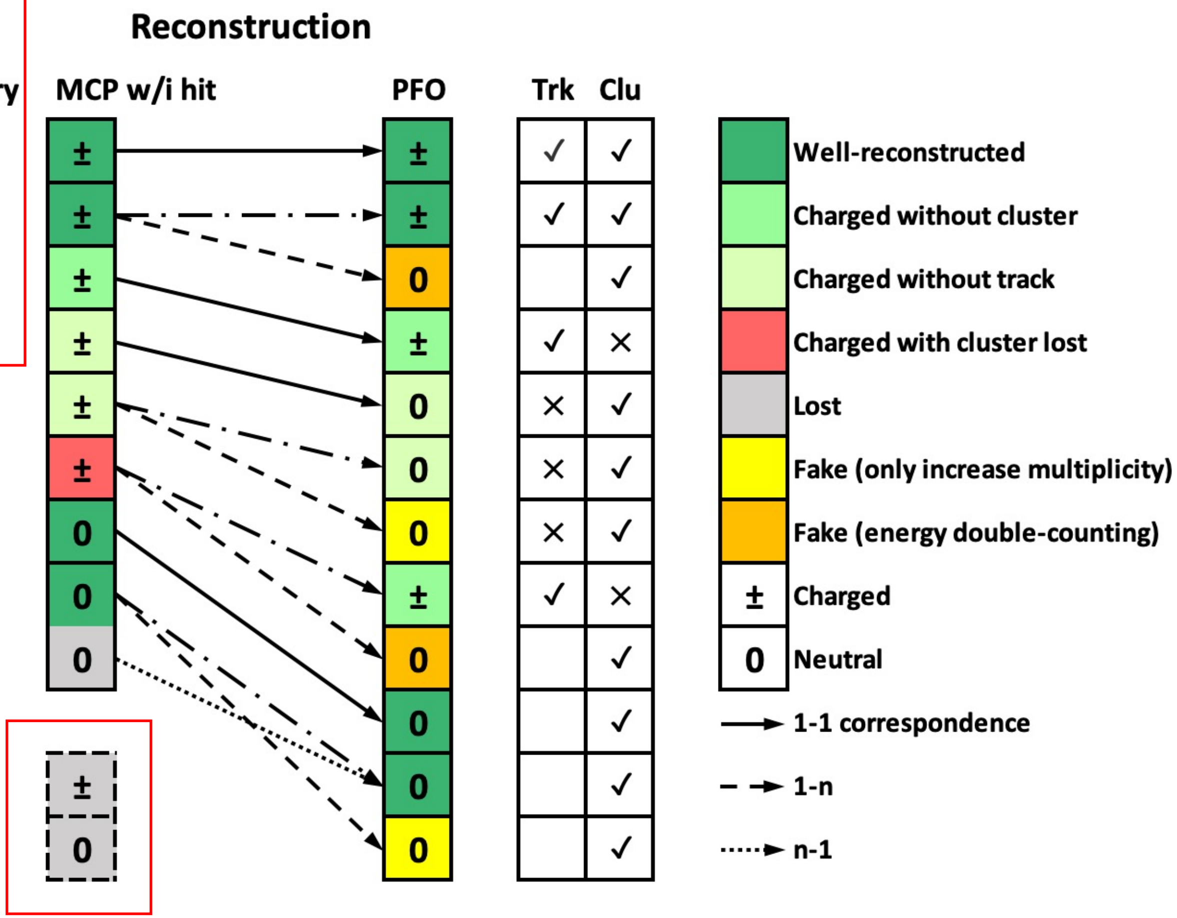
- **Jet origin ID** [PRL 132. 221802]
  - identification jets from 5 quarks + 5 anti-quarks + gluon
  - PID impacts significantly
    - especially improvement in light-quarks from $K^{\pm}, K^0_{L/S}, n, \Lambda$



Future 3-stage particle mapping

- **Lost decomposition**
  - Acceptance
  - Particle merging
- **Secondary particles**
  - e.g.
    - pre-interaction (upstream material)
      - photon conversion
    - back-scattering
    - nuclear interaction
    - …
  - To avoid E double-counting
- **3-stage particle mapping:**
  - More complete physics picture





Summary

- **1-1 correspondence reconstruction**
  - = confusion-free PFA + efficient PID
  - unified in a same ML model
- **Performance improvement via ML**
  - PFA confusion suppression: improve BMR by ~15% (3.3% → 2.75%)
  - PID: Extend traditional separated one to universal & simultaneous one of 9 categories
    - M6x6 ($e^{\pm}, \mu^{\pm}, \pi^{\pm}, K^{\pm}, p/\bar{p}, \gamma$): inclusive eff. > 97%
    - M3x3 ($K_L^0, n, \bar{n}$): inclusive eff. >75~80%
- **Prospects**
  - Extended PID, especially on neutral hadrons
    - TOF enhanced neutral hadron energy measurement
    - Further enhancement on jet origin ID
  - Improved 1-1 correspondence
    - More reliable and high-fidelity inputs for holistic event analysis
    - Make fast simulation modeling simpler and more accurate, facilitate pheno-studies

Summary

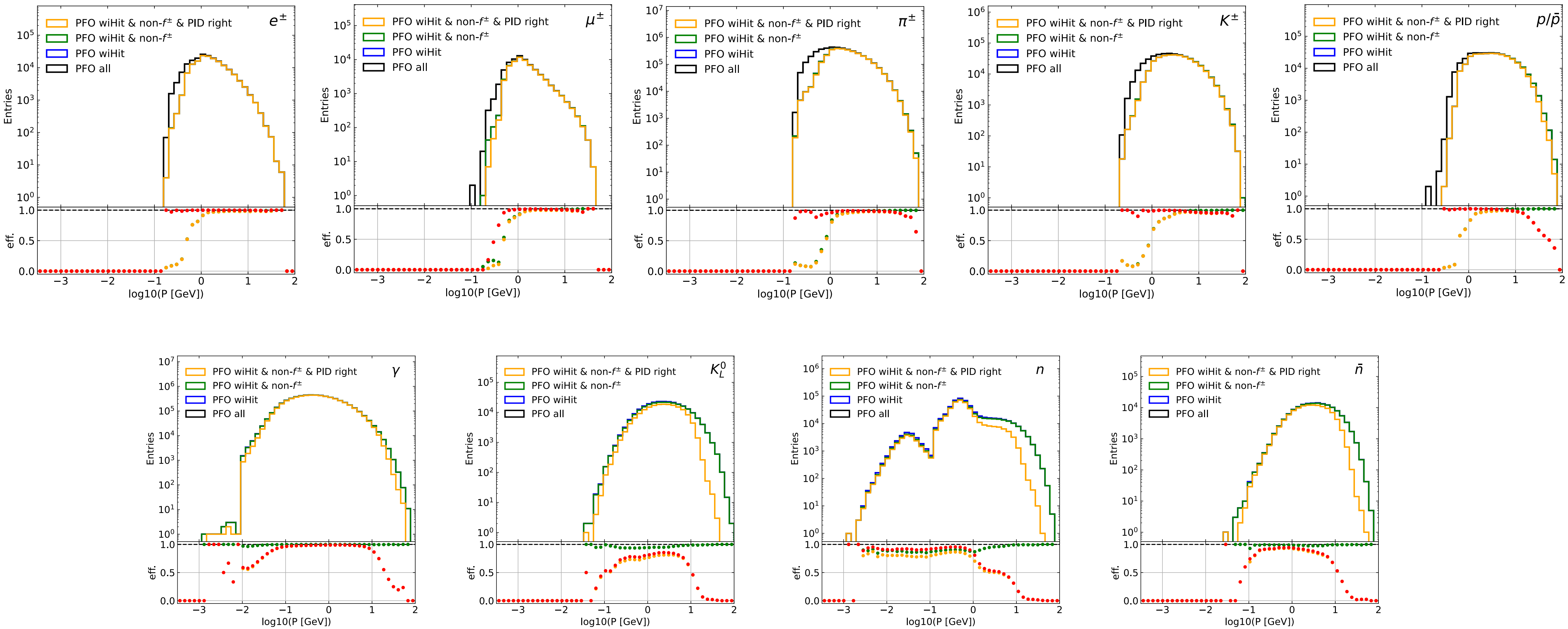
- **Key technique: particle mapping (truth link)**
  - Decomposition of PFA performance
  - Performance quantification of jet event (with high multiplicity)
    - enable diagnosis on individual particles
    - Beyond BMR: multi-dimensional metrics
  - Future 3-stage particle mapping, to further address
    - lost decomposition
    - secondary particles
- **Apply to different detector concepts, PFAs, collision environments**
  - ee Higgs factory (in progress)
    - CEPC ref-TDR
    - ILC ILD
  - from simulation to real data, e.g. BESIII, LHC…

Welcome collaborations!

Thank you for your attention!

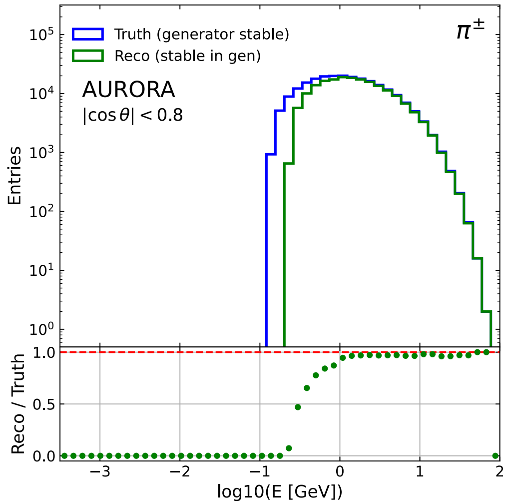
ParT PID efficiency

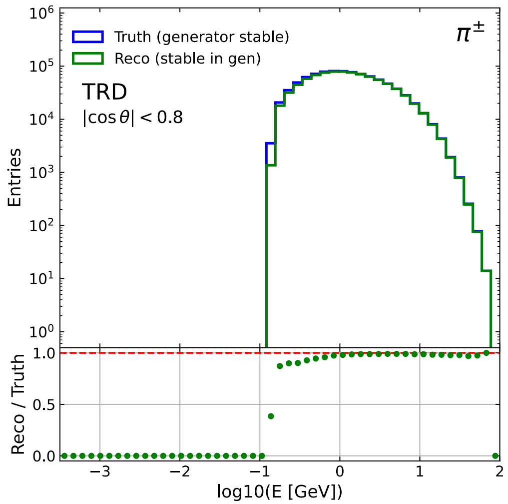
ParT PID eff: red dot = yellow / green

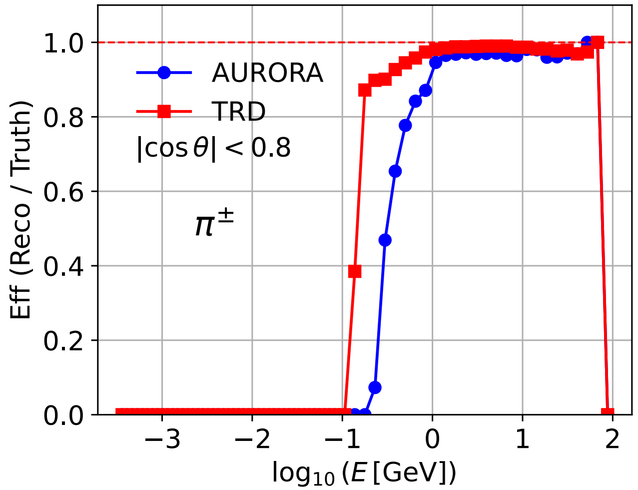


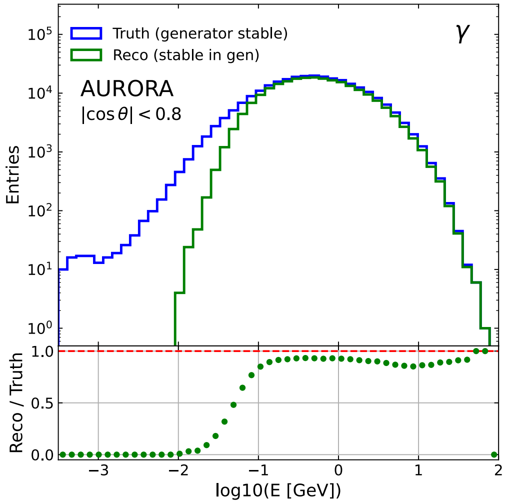
Particle reconstruction efficiency

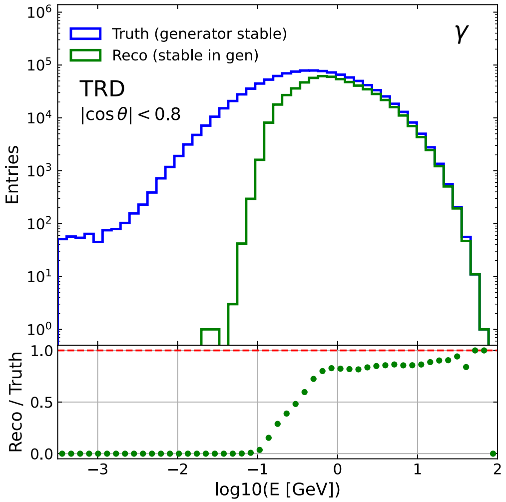
CDR | **TDR**

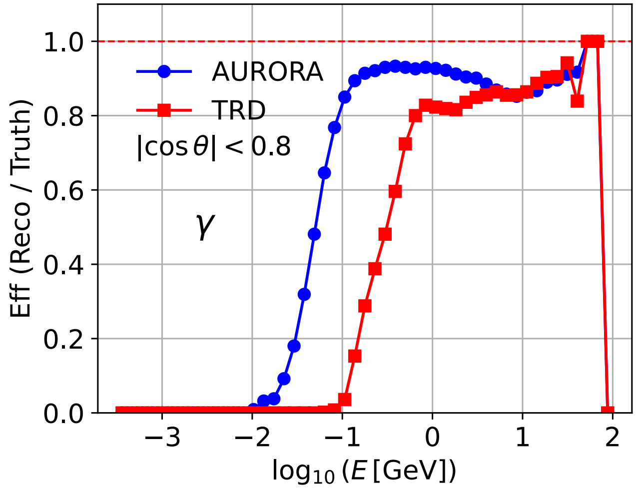








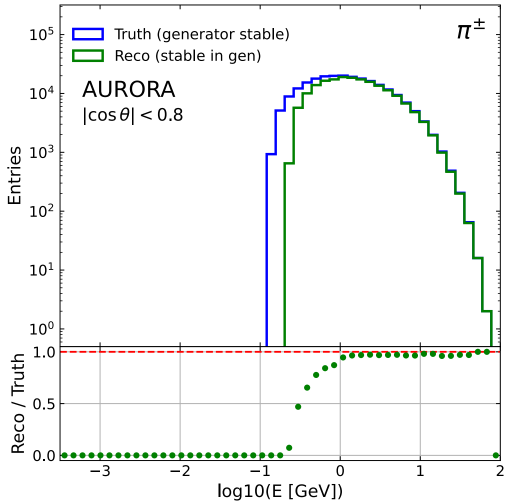


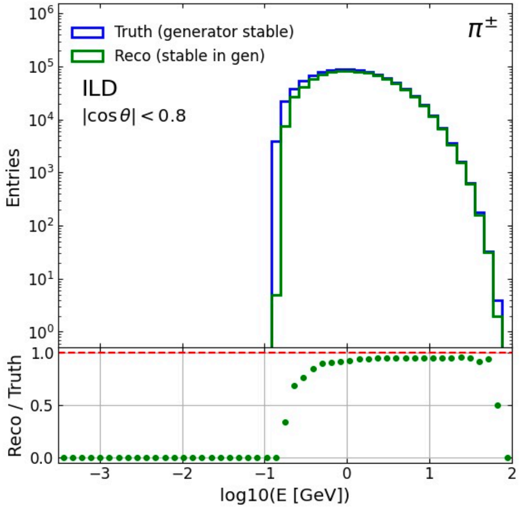


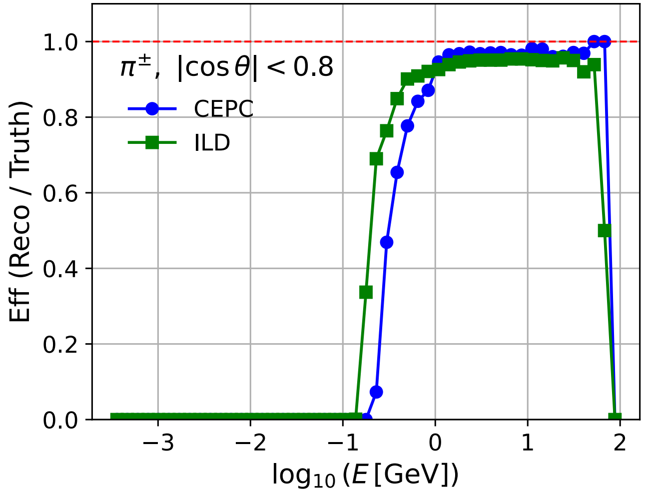
Particle reconstruction efficiency

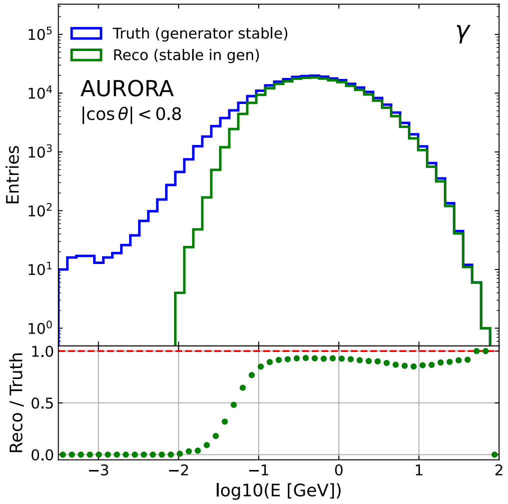
CDR

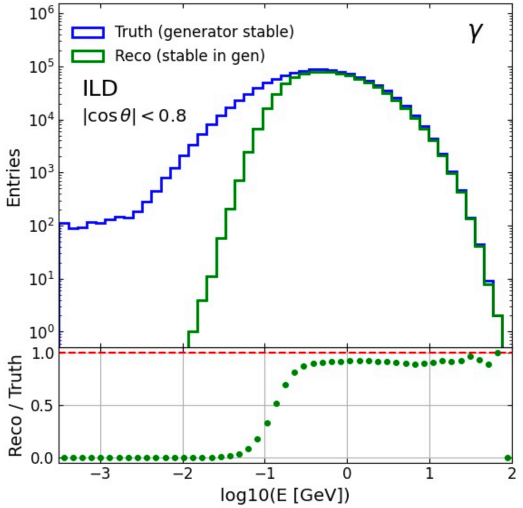
ILD

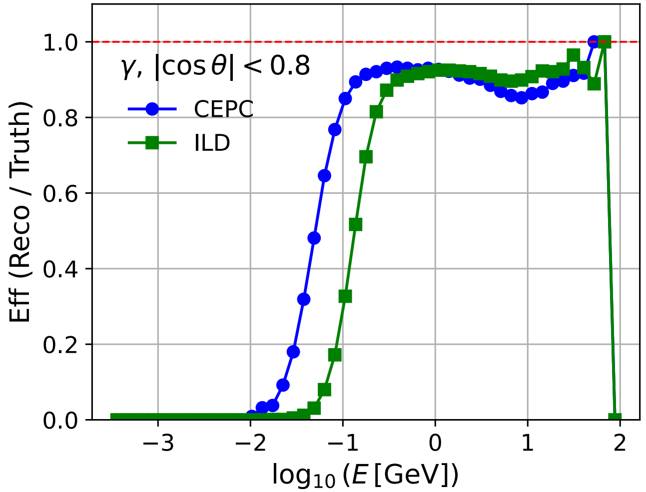






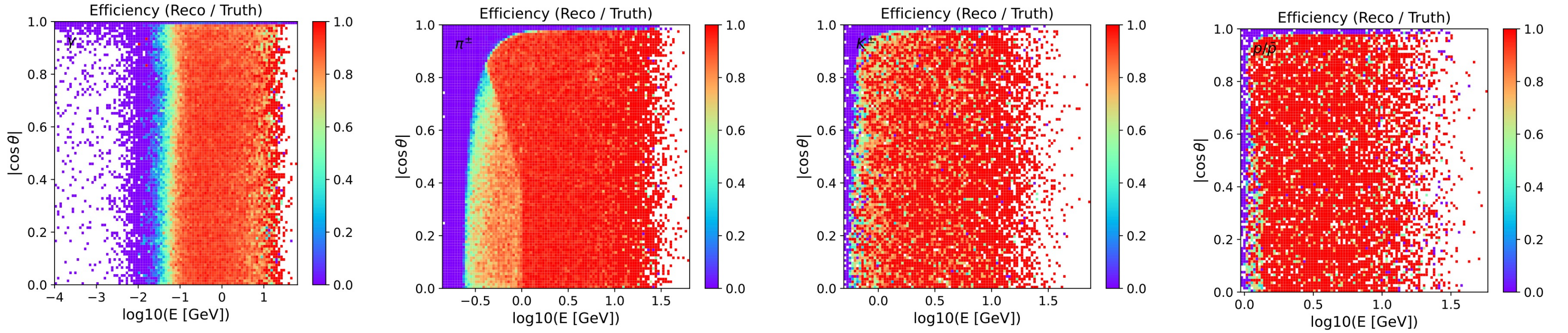




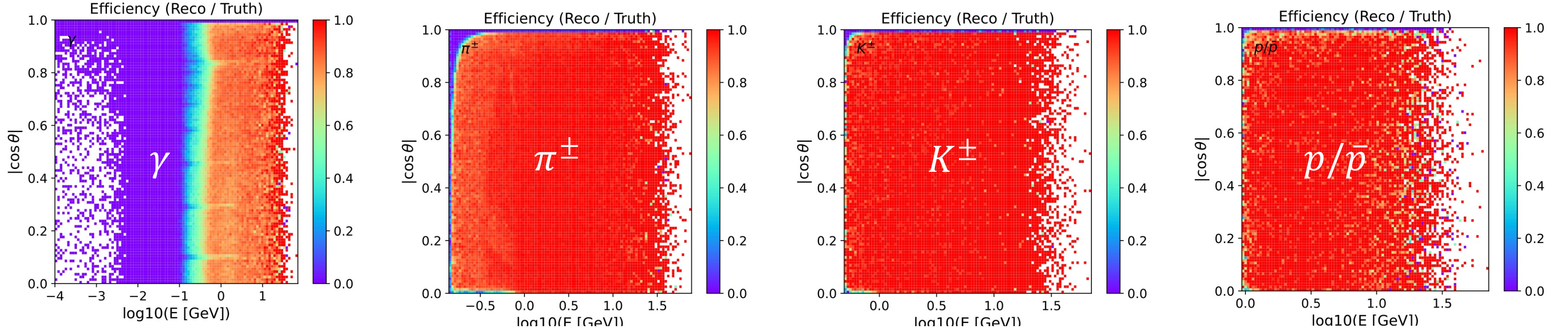


Particle reconstruction efficiency

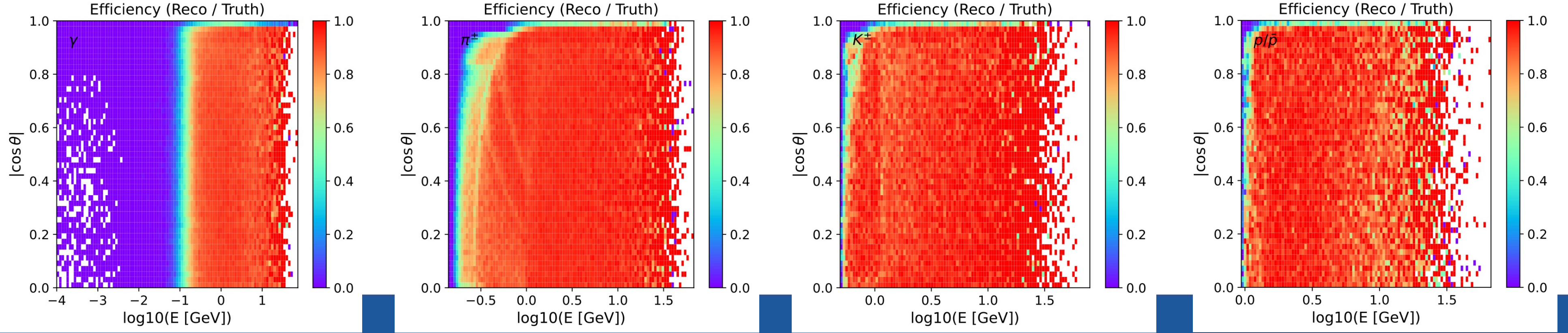
CDR



TDR



ILD

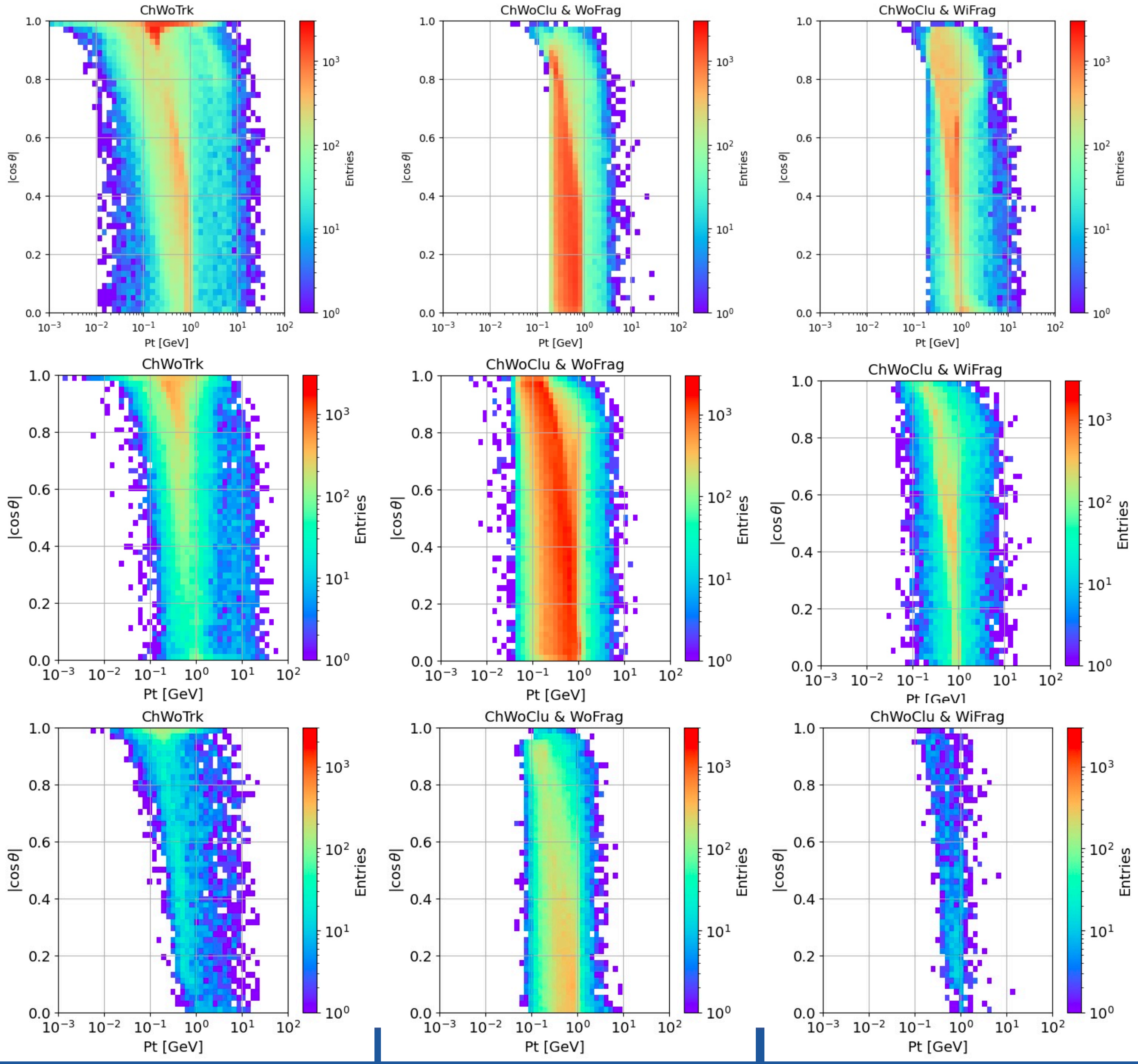


Charged particle without track/cluster

CDR

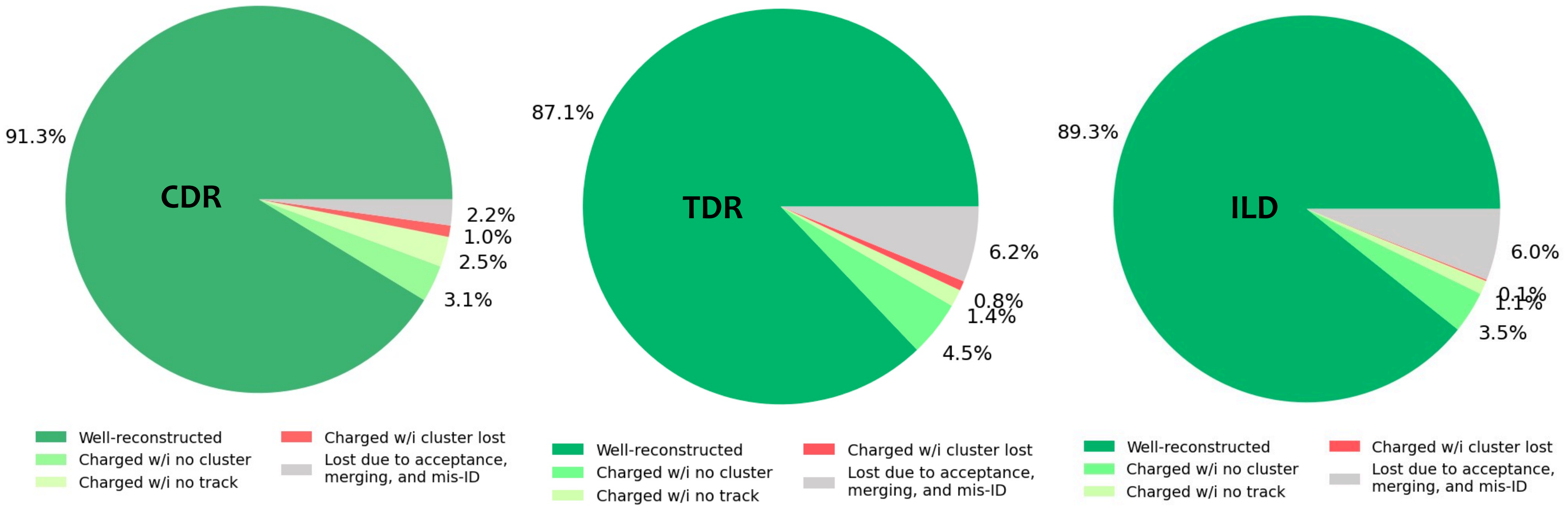
TDR

ILD



Energy fraction

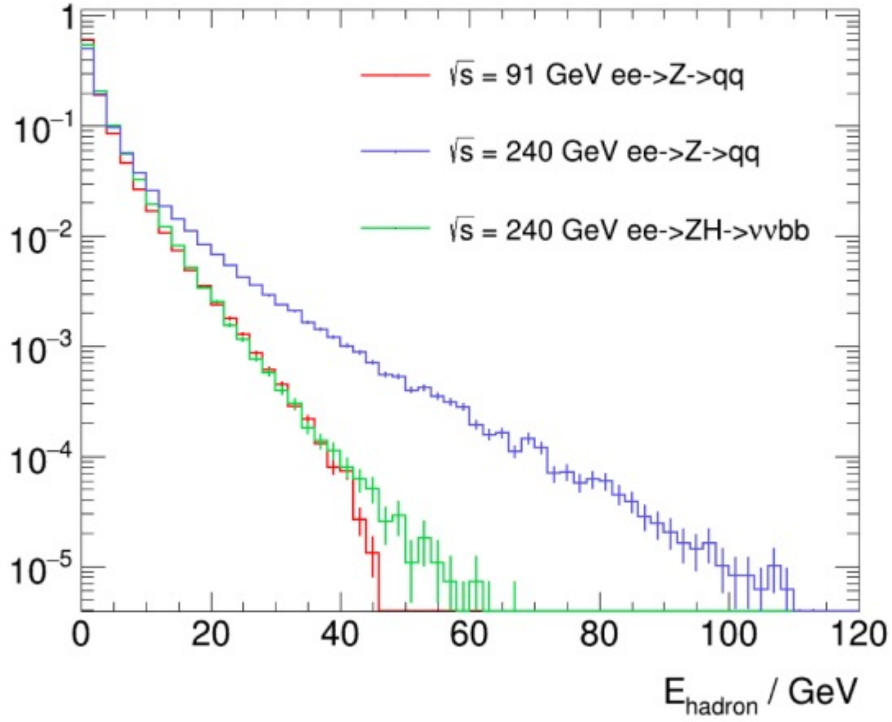
- **Increased fractions in TDR & ILD**
  - Charged w/o cluster
  - Lost (mainly from particle merging?)



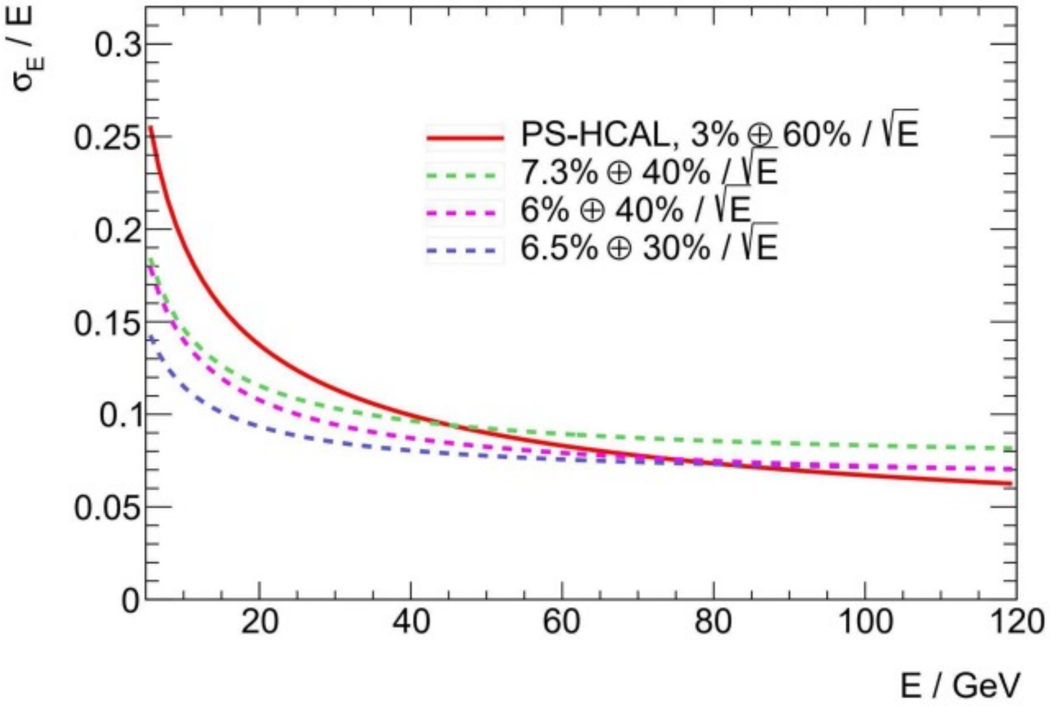
Detector geometry

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