Particle Identification Studies using a Highly Granular Crystal ECAL

Jack Rolph, Yong Liu, Baohua Qi, Zhiyu Zhao
on behalf of the AGORA/CEPC Calorimetry Team

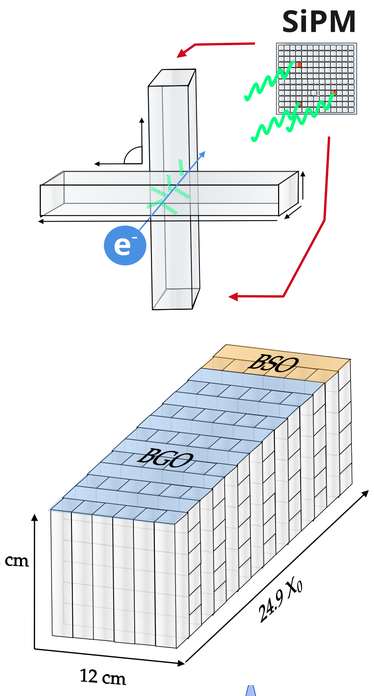
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

drdcalo

Overview

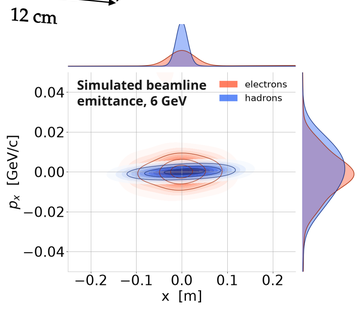
Highly Granular Crystal ECAL

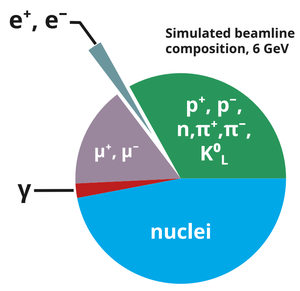
- **criss-crossing scintillating crystal bars [1]:**
  - CEPC detector volume: 1.5 × 1.5 × 40 cm³
  - prototype volume: 2 × 2 × 12 cm³
  - homogeneous;
  - crystal: BGO/BSO;
- **granularity:**
  - particle separation capability;
  - energy density information;
- **excellent EM energy resolution: stochastic term ~1-2%**



Particle Identification (PID)

- **Testbeam data contamination:**
  - Testbeam performed in June/July 2025;
  - Heavy contamination from hadrons;
  - Showering hadrons misidentified as electrons;
  - **influences energy resolution measurement;**
- **PID necessary for EM performance studies.**

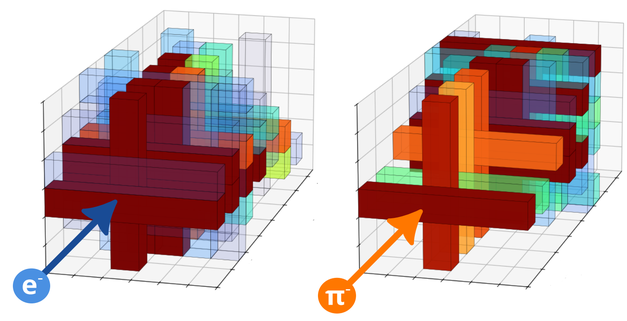




Summary of Studies

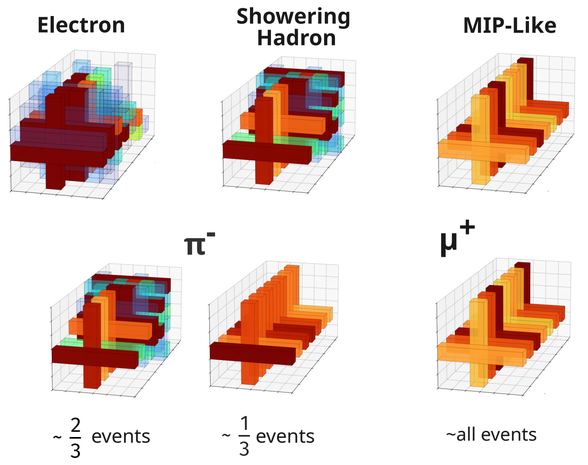
Simulation

- Geant4 simulation;
- DD4HEP geometry description;
- 1-10 GeV $e^-/\pi^-/\mu^+$ events;
- 130k μ^+ and 45k e^-/π events per energy;
- Detector effects (i.e., sensor noise).



PID: Unsupervised → Supervised

- **Three classes of event:**
  - electron;
  - showering-hadron-like;
  - MIP-like;
- **Problem: punch through hadrons:**
  - around 1/3 hadron events only ionize;
  - indistinguishable from μ^+/μ^-;
  - no trivial tag from simulation;
  - **empirical cut → classifier bias;**
- **Solution: unsupervised classification**
  - define measure where showering π^- is far from μ^+;
  - make cut to build tag;
  - **distill tag using simple classifier.**



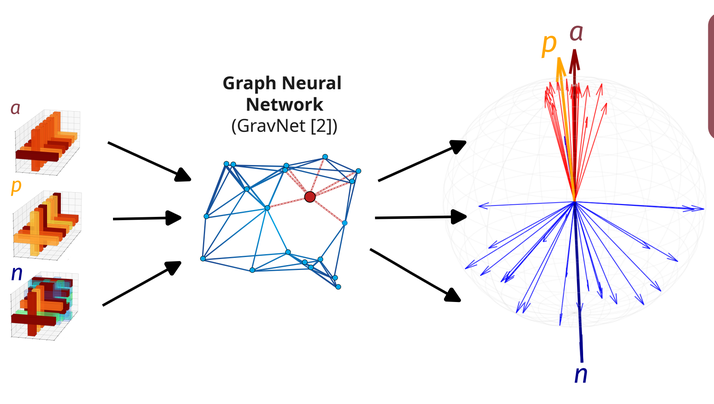


Key Results

Unsupervised PID: Contrastive Learning

- Sample two μ^+ events, and one π^- event;
  - **'anchor' (a_i)** : MIP-like reference event (μ^+);
  - **'positive' (p_i):** MIP-like event (μ^+);
  - **'negative' (n_i):** MIP-like / showering event (π^-);
- Encode each event to unit vector with neural network;
- Optimize triplet loss:

$$L(a,\ p,\ n) = \max\left\{d(a_i,\ p_i) - d(a_i,\ n_i) + \text{margin},\ \ 0\right\}, \qquad d(x_i,\ y_i) = \left\|\mathbf{x}_i - \mathbf{y}_i\right\|_2 = \sqrt{\sum_{k=1}^{D}\left(x_{i,k} - y_{i,k}\right)^2}$$

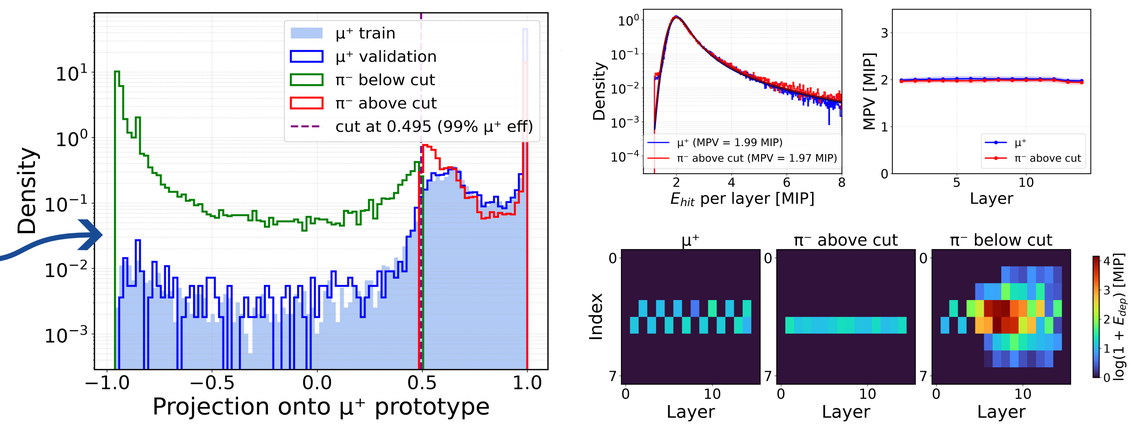


- Cosine similarity to average μ^+ vector ('prototype') from training data

→ 1D discriminator variable from 99% confidence interval of μ^+ projection.

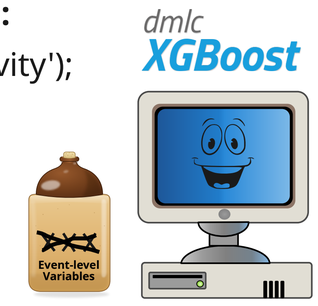
Key Result:

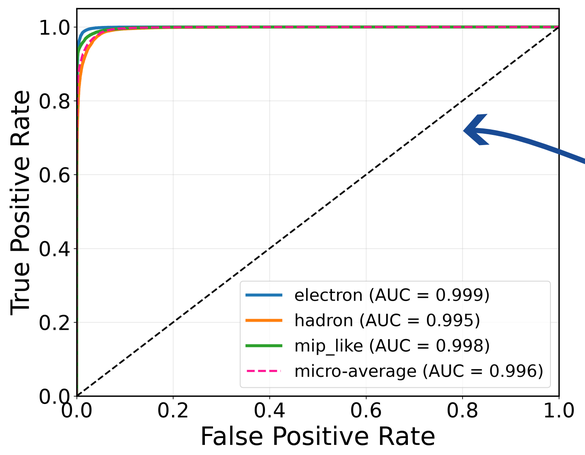
- **Red Region:**
  - Energy deposits Landau distributed;
  - ~ same MPV/layer as μ^+ ;
- **Green Region:**
  - Consistent with showering particle.



Supervised PID: Boosted Decision Tree

- **XGBoost gradient-boosted decision tree [3] (BDT) uses energy-independent event level variables:**
  - Energy-weighted mean position ('Center of Gravity');
  - Distance from Center of Gravity;
  - Fraction of energy per layer;
  - Moments of Central Log Ratio of bar energies;
- **BDT distills electron/showering hadron-like/ MIP-like from the unsupervised classifier.**



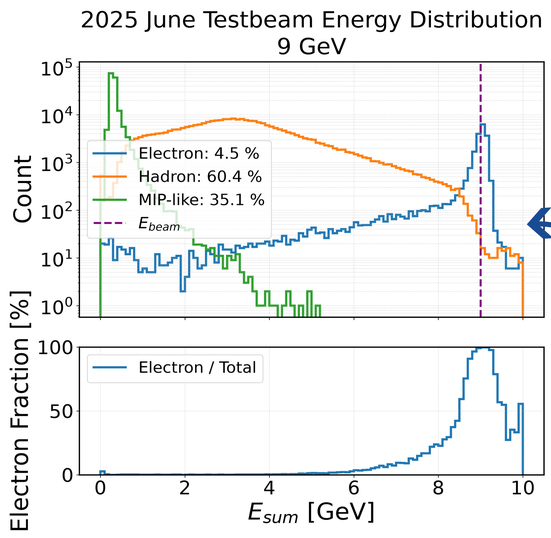


Key Result:

- Evaluation:
  e^-/μ^+: truth;
  π^-: unsupervised tag;
- Classifier performance almost perfect:

ROC AUC 0.996

- Consistent performance at all energies under test.
- Strong indication ECAL can be used for PID in Particle Flow.



Key Result:

- Applied to testbeam data;
- Classes of events follow expectations from experimental data;
- Electrons well-separated even though only 4.5% of total events.

[1] The CEPC Study Group, CEPC Technical Design Report -- Reference Detector, arXiv:2510.05260 [hep-ex] (2025), https://arxiv.org/abs/2510.05260.
[2] Qasim, S.R. et al., Learning representations of irregular particle-detector geometry with distance-weighted graph networks, Eur. Phys. J. C 79 (2019) 608, DOI: 10.1140/epjc/s10052-019-7113-9
[3] Chen, T. & Guestrin, C., XGBoost: A Scalable Tree Boosting System, Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (2016) 785–794, DOI: 10.1145/2939672.2939785