

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



Overview

Highly Granular Crystal ECAL

► criss-crossing scintillating crystal bars [1]:

- CEPC detector volume: $1.5 \times 1.5 \times 40 \text{ cm}^3$
- prototype volume: $2 \times 2 \times 12 \text{ cm}^3$
- homogeneous;
- crystal: BGO/BSO;

► granularity:

- particle separation capability;
- energy density information;

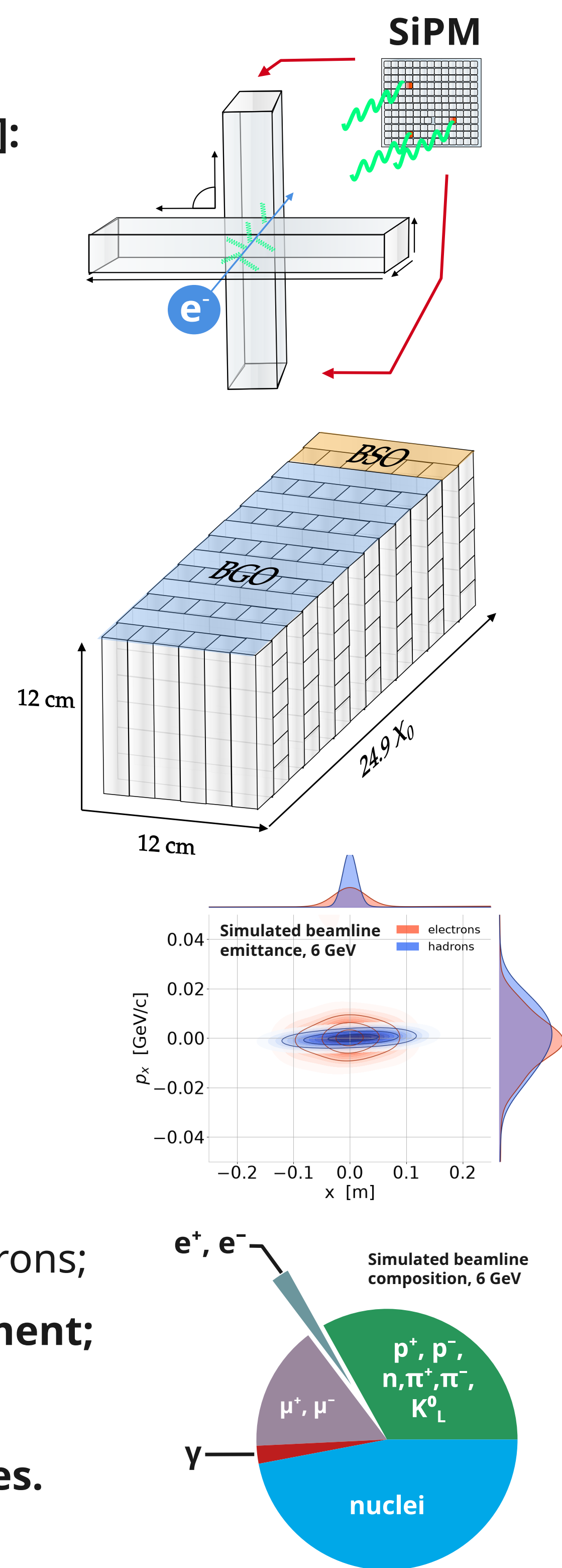
► excellent EM energy resolution: stochastic term ~1-2%

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:

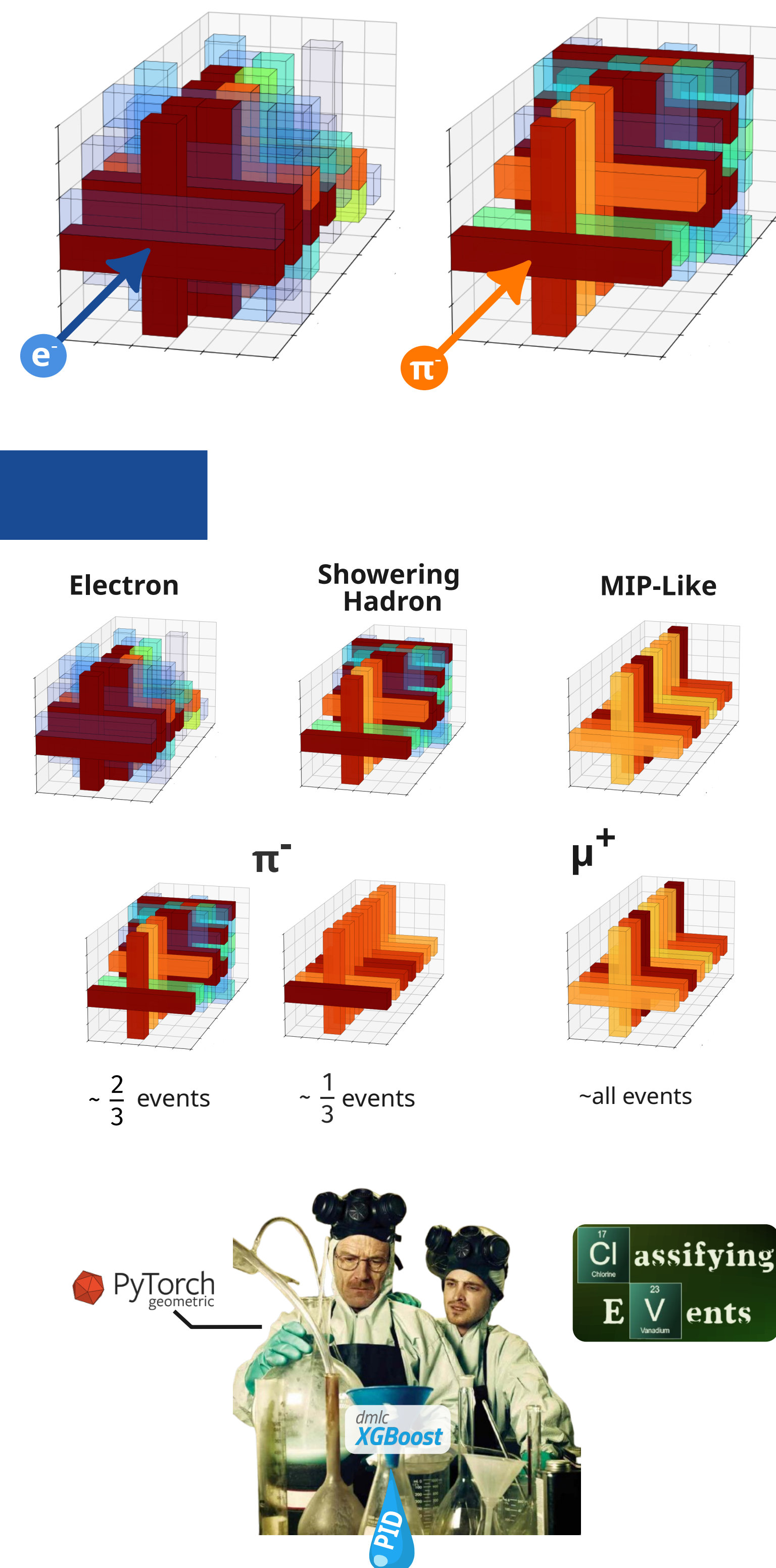
- electron;
- showering-hadron-like;
- MIP-like;

► 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.



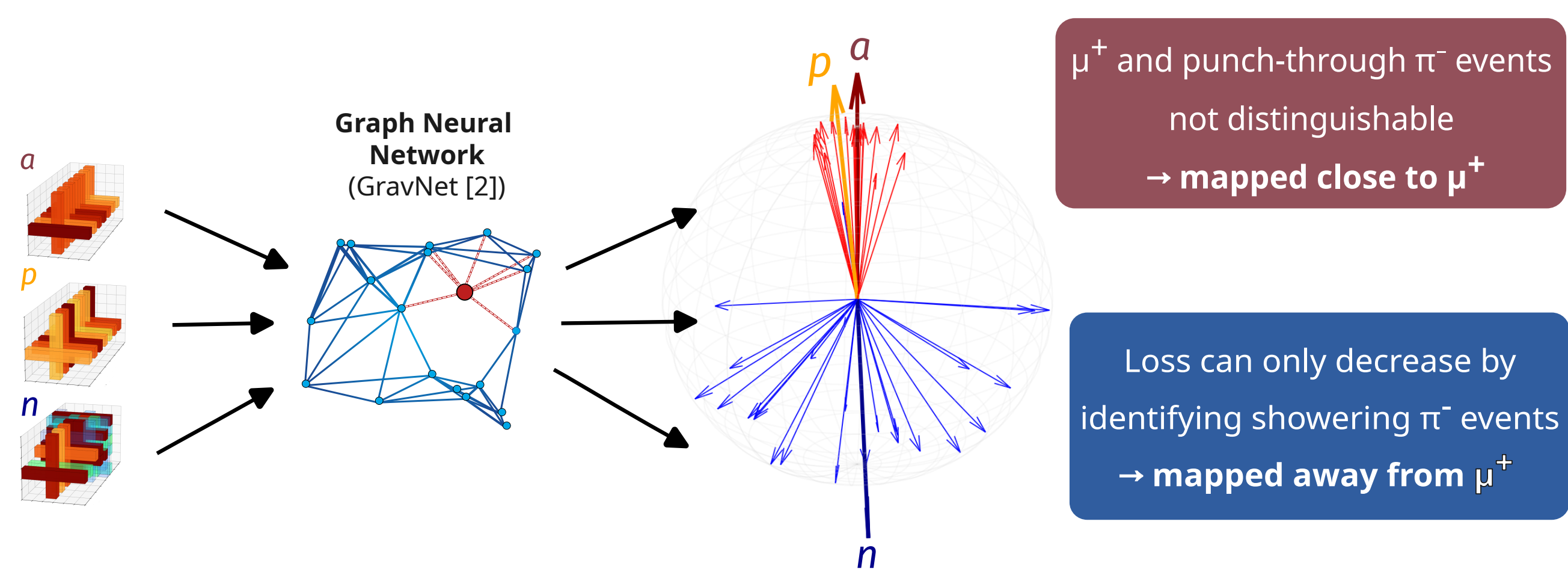
Key Results

Unsupervised PID: Contrastive Learning

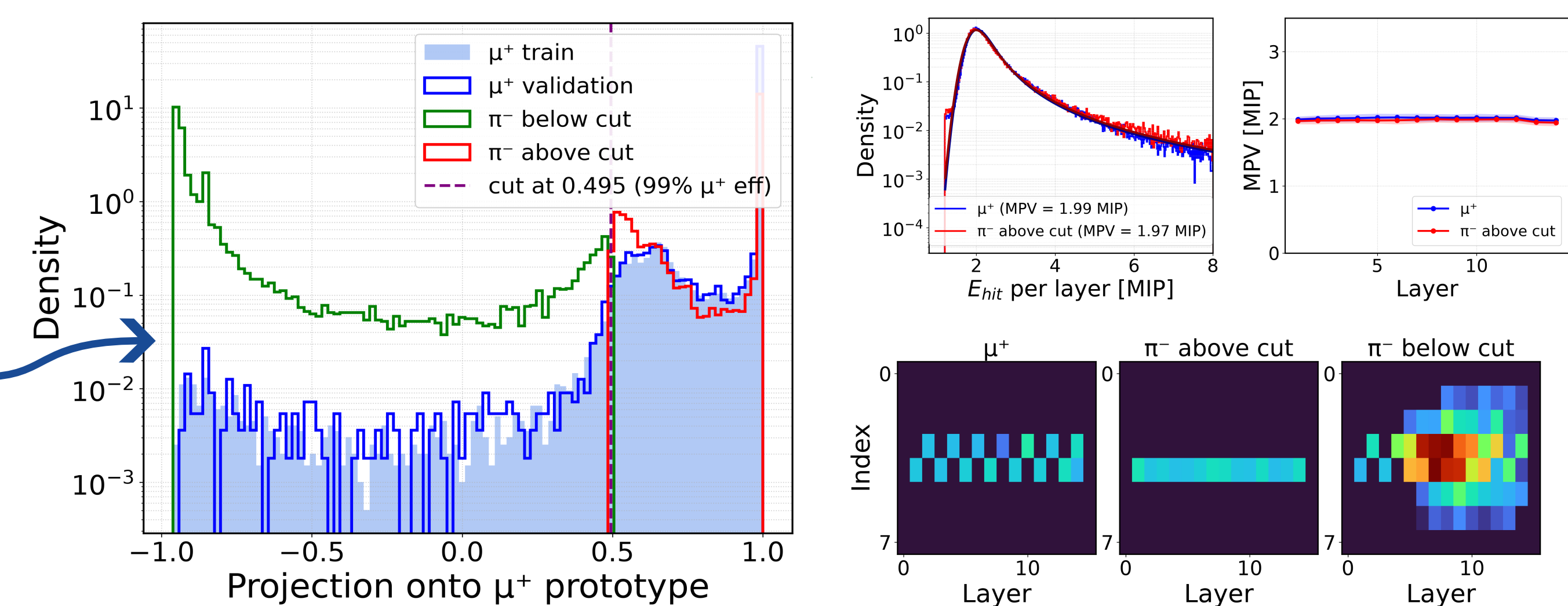
- Sample two μ^+ events, and one π^- event;
- 'anchor' (a_i): MIP-like reference event (μ^+);
- 'positive' (p_i): MIP-like event (μ^+);
- 'negative' (n_i): MIP-like / showering event (π^-);

- Encode each event to unit vector with neural network;
- Optimize triplet loss:

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



- Cosine similarity to average μ^+ vector ('prototype') from training data
→ 1D discriminator variable from 99% confidence interval of μ^+ projection.

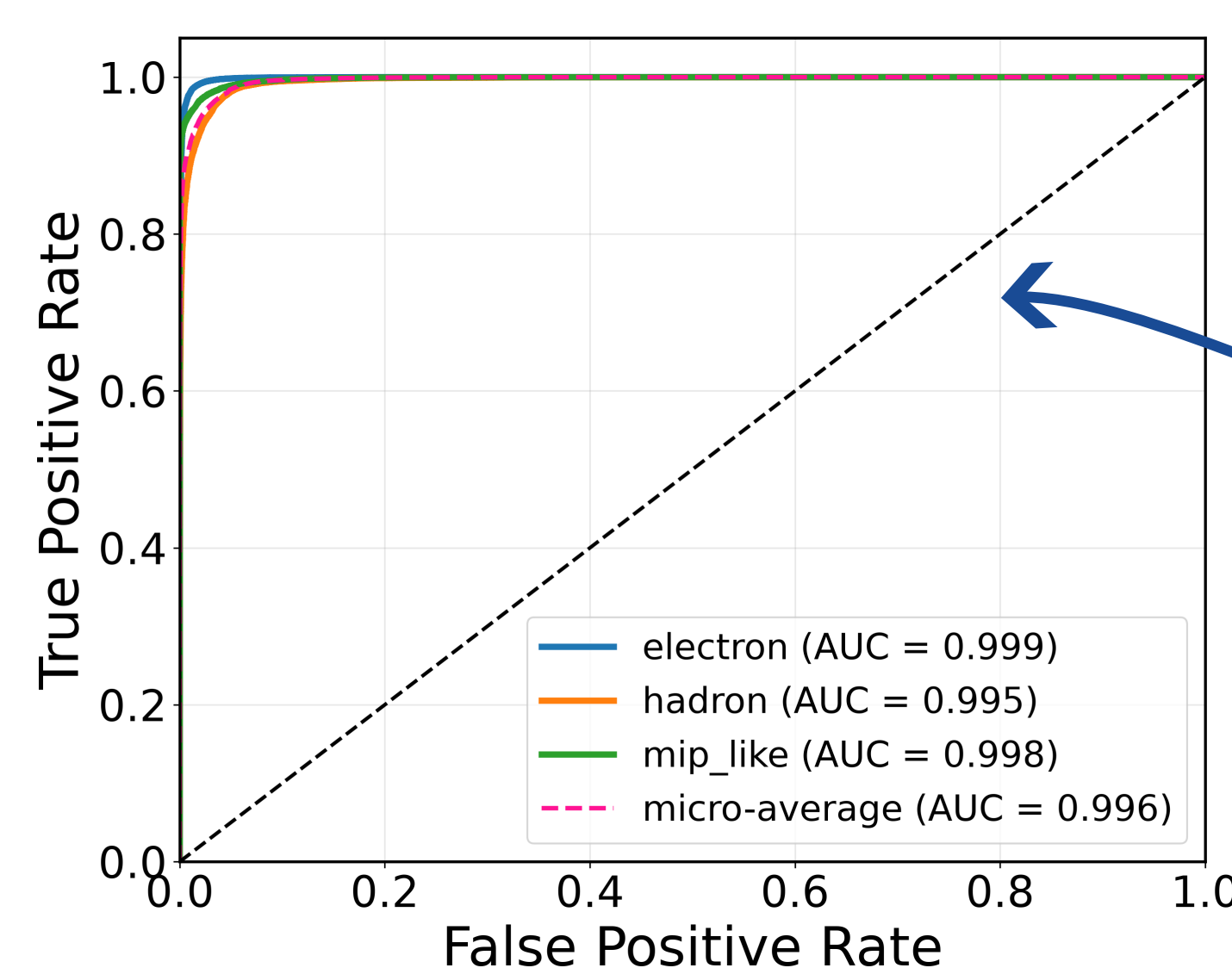


Supervised PID: Boosted Decision Tree

► 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.

