Reconstruction algorithm for ECAL

Sun Shengsen

on behalf of CEPC ECAL software group

CEPC Day, Oct. 28th 2021

中国科学院高能物理研究所

Institute of High Energy Physics Chinese Academy of Sciences

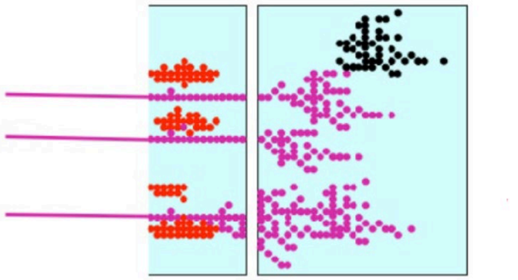
Introduction

- High precision Higgs / Z factory:
  - Requiring excellent jet energy resolution (3~4%) for W/Z separation and precise Higgs measurement.
  - Fine γ/π^0 separation for flavor physics study.
- Particle-Flow Approach (PFA):
  - Measure jet by its components: 60% charged particles, 30% photons, 10% neutral hadrons.
  - Final resolution:

$$\boldsymbol{\sigma_{Jet} = \sqrt{\sigma^2_{Track} + \sigma^2_{Had} + \sigma^2_{em} + \sigma^2_{Confusion}}}$$

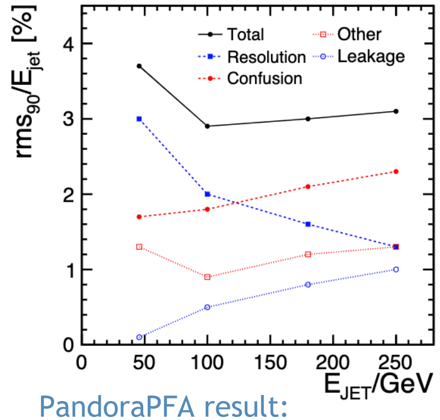
Component	Detector	Energy Fraction	Energy Resolution	Jet Energy Resolution
Charged Particles ($X^{\pm}$)	Tracker	$\sim 0.6\ E_J$	—	—
Photons (γ)	ECAL	$\sim 0.3\ E_J$	$\mathbf{0.15\sqrt{E_\gamma}}$	$\mathbf{0.08\sqrt{E_J}}$
			$\mathbf{0.03\sqrt{E_\gamma}}$	$\mathbf{0.016\sqrt{E_J}}$
Neutral Hadrons (h^0)	HCAL	$\sim 0.1\ E_J$	$0.55\sqrt{E_{h^0}}$	$0.17\sqrt{E_J}$

Jet E res.	W/Z sep
perfect	**3.1 σ**
2%	**2.9 σ**
3%	**2.6 σ**
4%	**2.3 σ**
5%	**2.0 σ**
10%	**1.1 σ**



$$\mathbf{E_{JET} = E_{TRACK} + E_{\gamma} + E_{n}}$$

- Avoid double counting
- Separate energy deposits

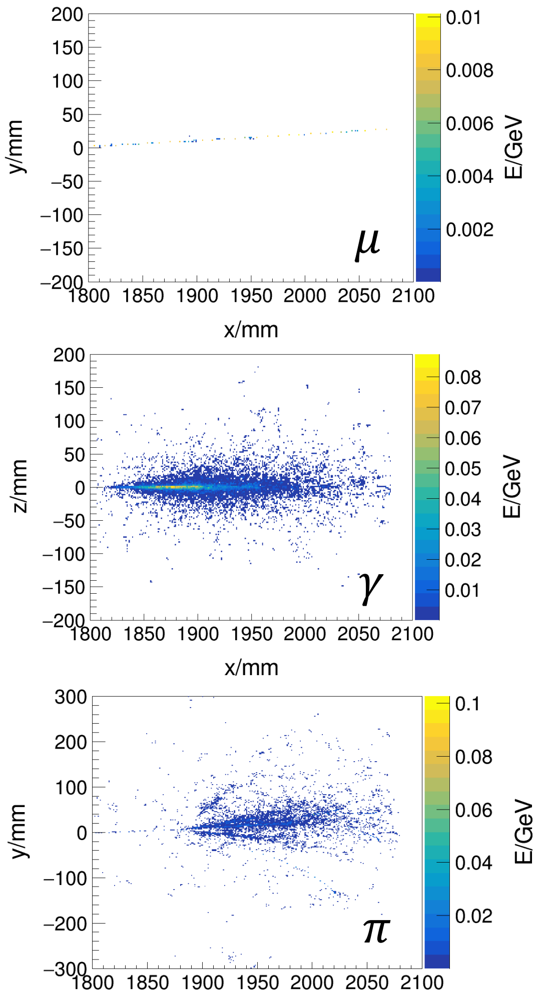


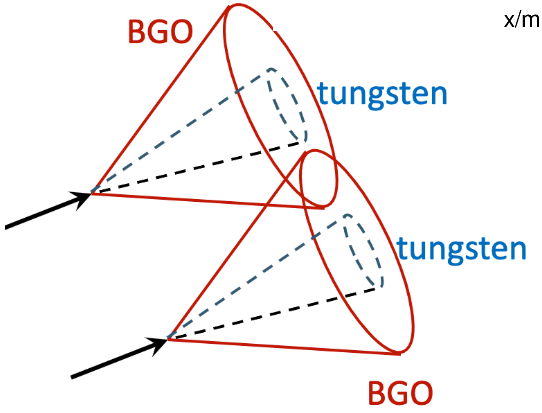
PandoraPFA result: confusion is important in JER

Introduction

- PFA requirement: Hardware + Software
  - Distinguish showers in calorimeter ⇨ high granularity ECAL/HCAL.
  - Minimize transverse spread of EM shower ⇨ small Moliere radius R_M ⇨ SiW sampling ECAL in ILD.
  - Separate EM and Hadronic showers longitudinally ⇨ large λ_I/X_0 ratio.
- Crystal ECAL:
  - Homogeneous structure ⇨ energy resolution $\sim 3\%/\sqrt{E} \oplus 1\%$.
  - Energy recovery for electrons.
  - Capability to trigger single photons ⇨ precise γ/π^0 reconstruction.
  - Larger Moliere radius ⇨ larger probability of shower overlap.
  - Smaller λ_I/X_0 ⇨ larger probability of hadronic shower in ECAL.
  - Exploit energy information and 3-D profile of shower

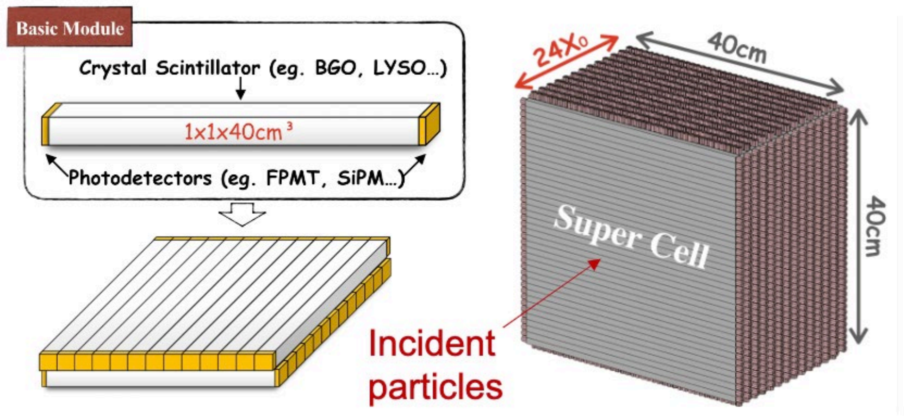
Material	X_0 /cm	R_M /cm	λ_I /cm	λ_I/X_0
W	**0.35**	**0.93**	**9.6**	**27.4**
BGO	**1.12**	**2.23**	**22.8**	**20.3**
Ratio	3.2	2.4	2.4	0.74

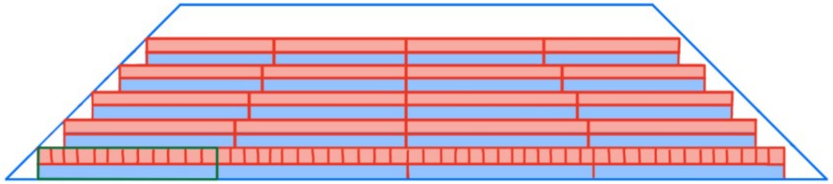


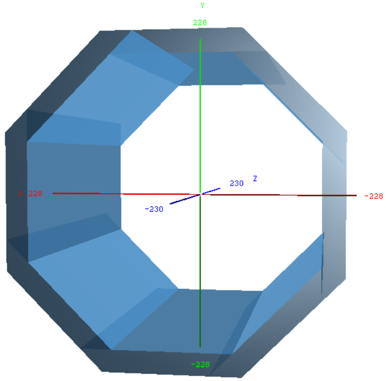


Introduction

- Crystal bar ECAL
  - Homogeneous BGO crystal.
  - Size: $1 \times 1 \times {\sim}40\ \mathrm{cm}^3$, double-sided readout.
  - Super cell module: ${\sim}40 \times {\sim}40 \times 2\ \mathrm{cm}^3$.
  - Time measurement at two ends for position along the bar.
  - Crossed arrangement in adjacent layers.
  - Save readout channels and minimize dead materials.
  - Full detector: $R = 1.8m, L = 4.6m, H = 28cm$, 8 same trapezoidal staves.
- Key issues:
  - Ambiguity caused by 2D measurement (ghost hit problem).
  - Identification of energy deposits from individual particles (confusion).







Simulation and digitization

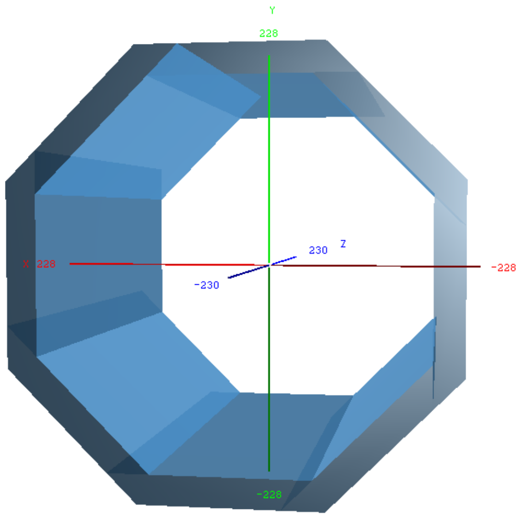
- Standalone ECAL geometry in CEPCSW
- Simulation is performed with GEANT4
  - Electromagnetic interactions.
- Simple digitization for one long crystal bar:
  - Readout information: 2-end Q and T.
  - Contribution from G4step i:

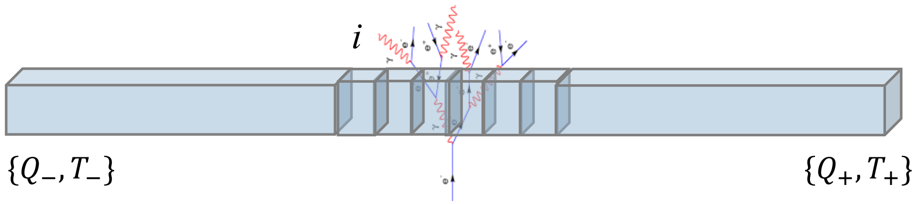
$$Q^i_\pm = E_0 \cdot e^{-\frac{\frac{L}{2}\pm z_i}{L_{Att}}}, \qquad T^i_\pm = T_0 + Gaus(z^i_\pm/v,\ \sigma_T).$$

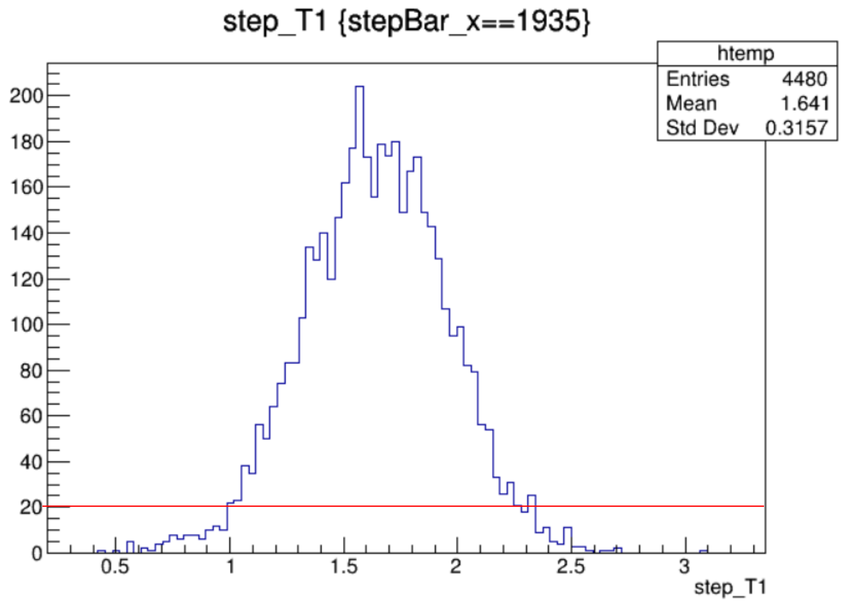
  - For the full bar:

$Q_\pm = \sum_{step} Q^i_\pm$, $T_\pm = T^k_\pm \mid (\sum^k_{i=1} Q^i_\pm > \epsilon Q^{tot}_\pm)$, $\epsilon = 5\%$.

Simplified condition: $L_{Att} = \infty$, so $Q_\pm = E_{tot}$.

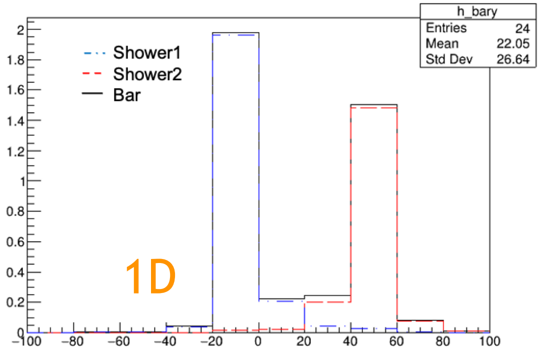


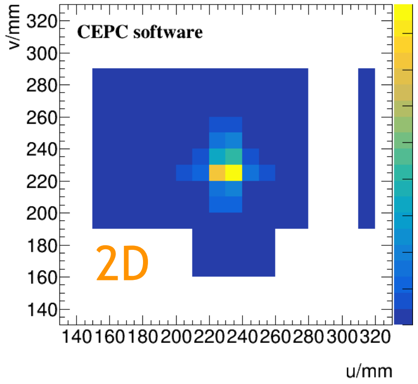


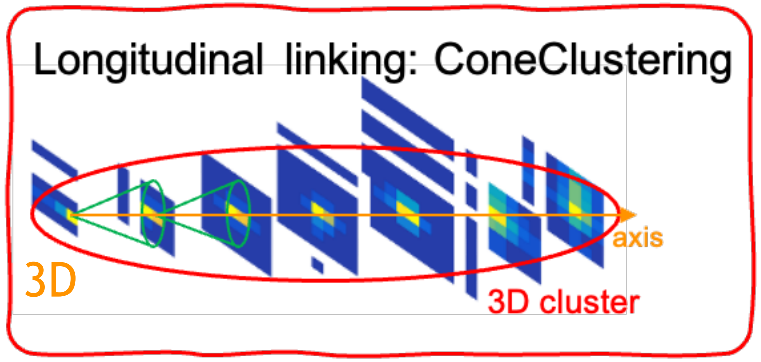


Reconstruction flow

- Reconstruction algorithm V0:
  - Iteration 0: roughly reconstruct clusters to remove the fluctuations.
  - Iteration 1:
    - 1D energy splitting.
    - 2D matching & ghost hits removal in each layer (X-Y plane). (chi2)
    - 3D cone clustering longitudinally.
    - Cluster ID
      - ID efficiency & mis-ID rate.
    - Re-cluster MIP/EM/Had showers.
    - Merge clusters.



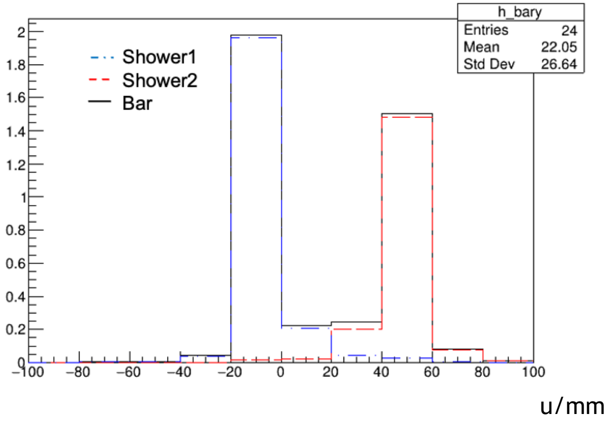


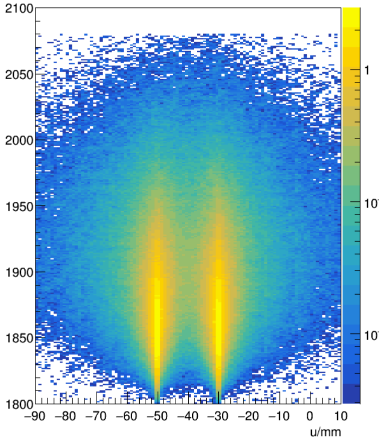


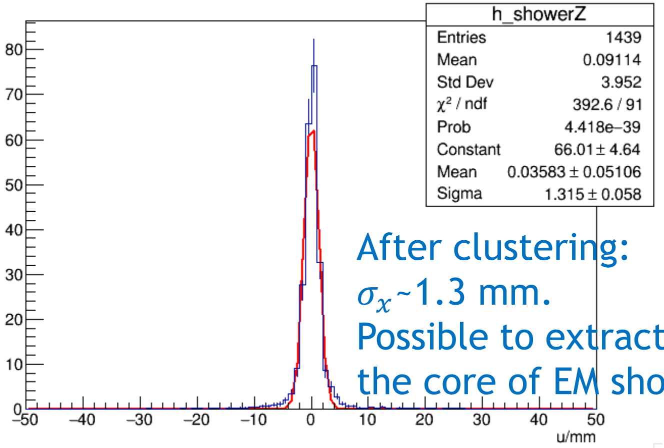
Reconstruction: 1D

Clustering and splitting in each layer:

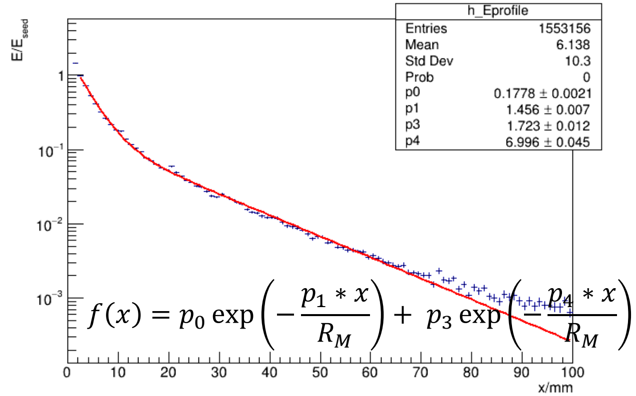
- Readout from detector: crystal bars (energy + time).
- Clustering:
  - Neighbor clustering.
  - Find local maximum and seed candidate.
- Energy splitting:
  - $N_{seed} \geq 2$ and second moment $S > S_{th}$ (0 for now).
  - Calculate the expected energy deposition of shower μ in bar i: $E_{i\mu}^{exp} = E_{\mu}^{seed} \times f(|x_i - x_{\mu}|)$.
  - Split bar i energy with expected value
    → Solve the overlap issue from large R_M.







After clustering: σ_x~1.3 mm.
Possible to extract the core of EM shower.

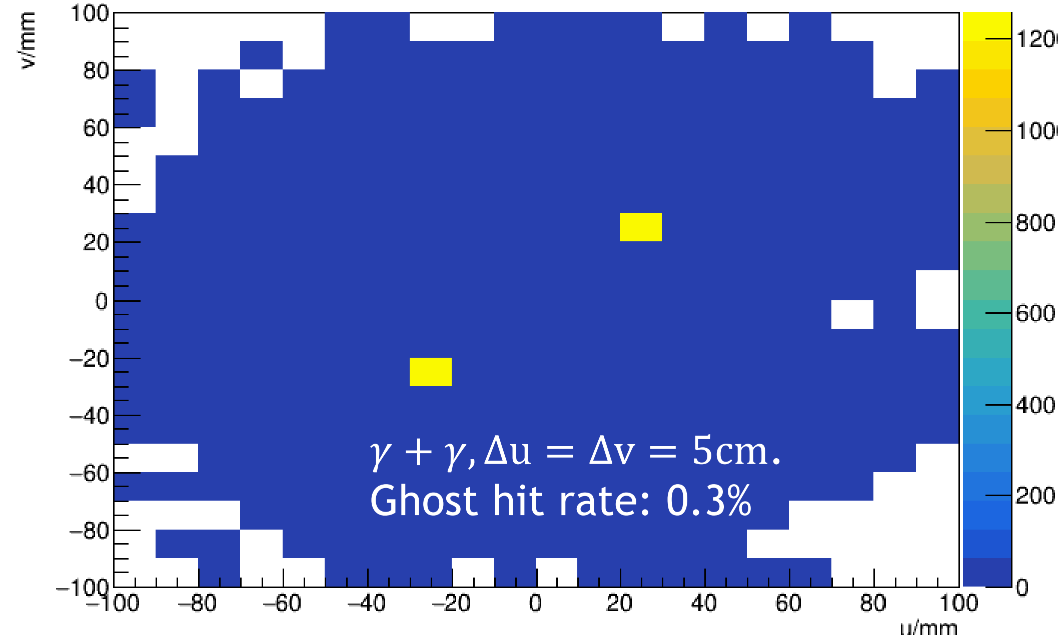


$$f(x) = p_0 \exp\left(-\frac{p_1 * x}{R_M}\right) + p_3 \exp\left(-\frac{p_4 * x}{R_M}\right)$$

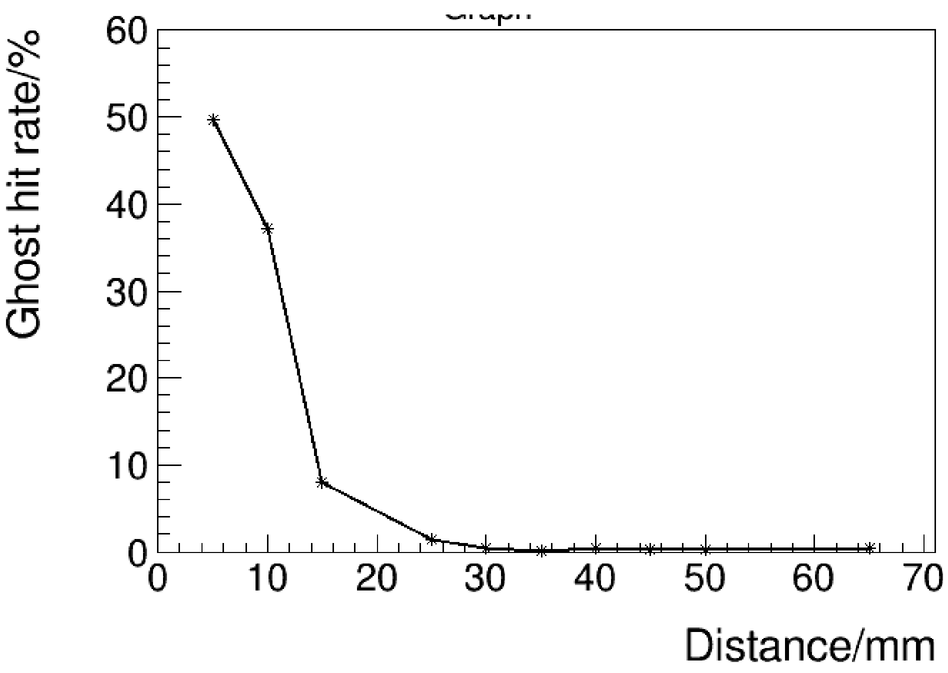
Reconstruction: 2D

Matching the crossing bars in 2 adjacent layers:

- Define a χ^2 with both energy and timing: $\chi^2_{point} = \chi^2_E + (\chi^2_{Tx} + \chi^2_{Ty})/2 = \frac{(E_X-E_Y)^2}{\sigma_E^2} + \left[\frac{(z_T-z_X)^2}{\sigma_S^2+\sigma^2_{z(t)}} + \frac{(z_T-z_Y)^2}{\sigma_S^2+\sigma^2_{z(t)}}\right]/2.$
- Reject the wrong combination with this χ^2.
- Define the ghost hit rate: $\frac{\#\text{shower in ghost position}}{\#\text{shower in correct position} + \#\text{shower in ghost position}}$.

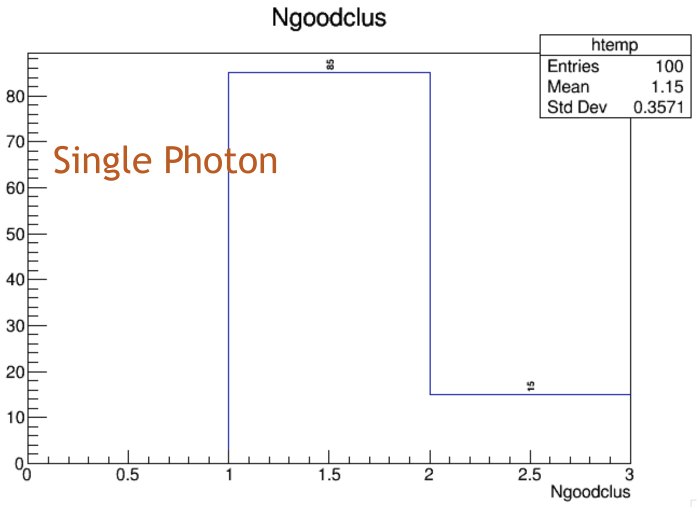


$\gamma+\gamma, \Delta u = \Delta v = 5\text{cm}.$
Ghost hit rate: 0.3%

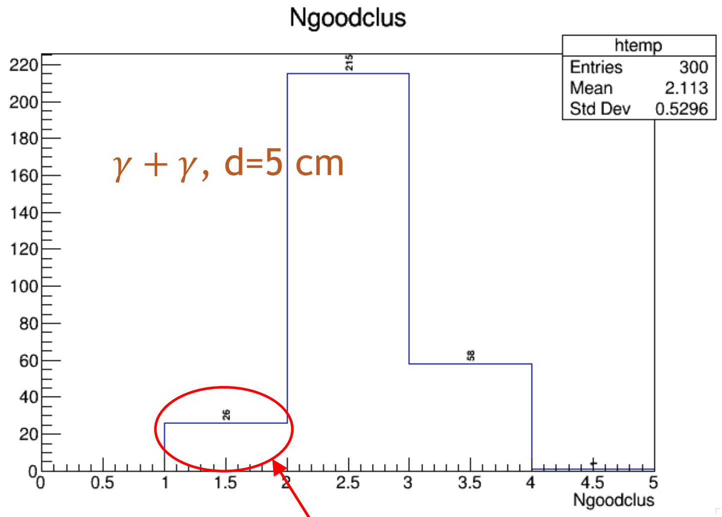


Reconstruction: 3D

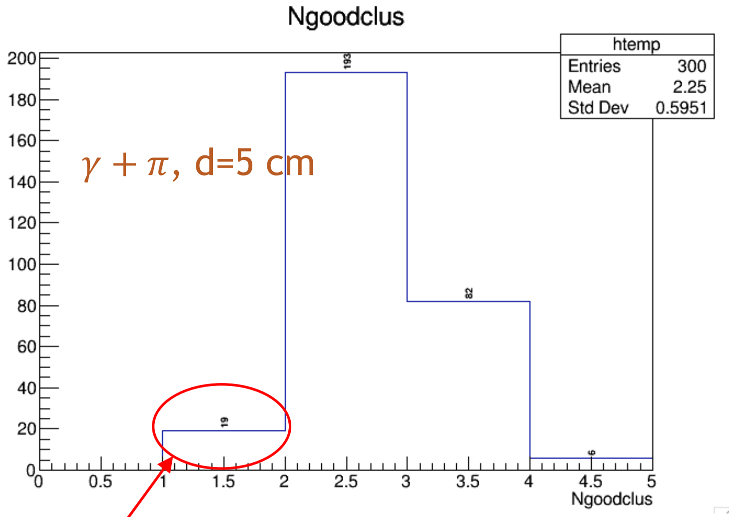
- Longitudinal linking for 3D cluster:
  - Cone-based clustering algorithm.
  - Get the very preliminary 3D structure, identify the cluster (MIP/EM/Hadron) based on this result.
  - Re-clustering: reconstruct MIP/EM clusters first, and then reconstruct hadronic clusters with a large R threshold.



Single Photon



$\gamma+\gamma$, d=5 cm

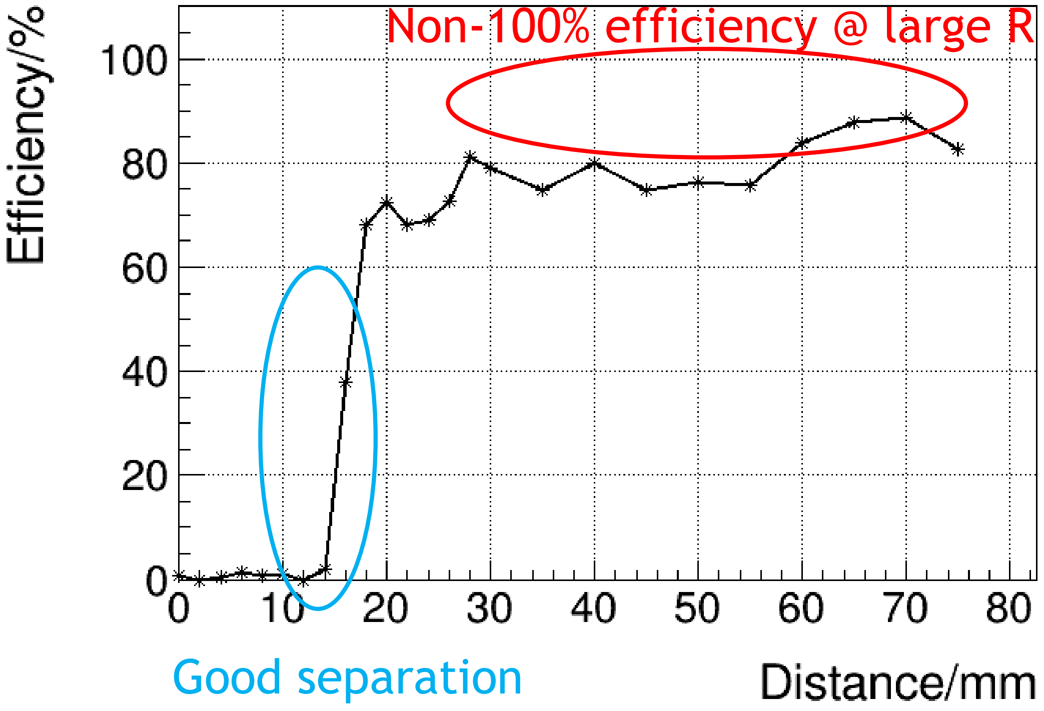


$\gamma+\pi$, d=5 cm

Issue need to solve: lost clusters (<10%).

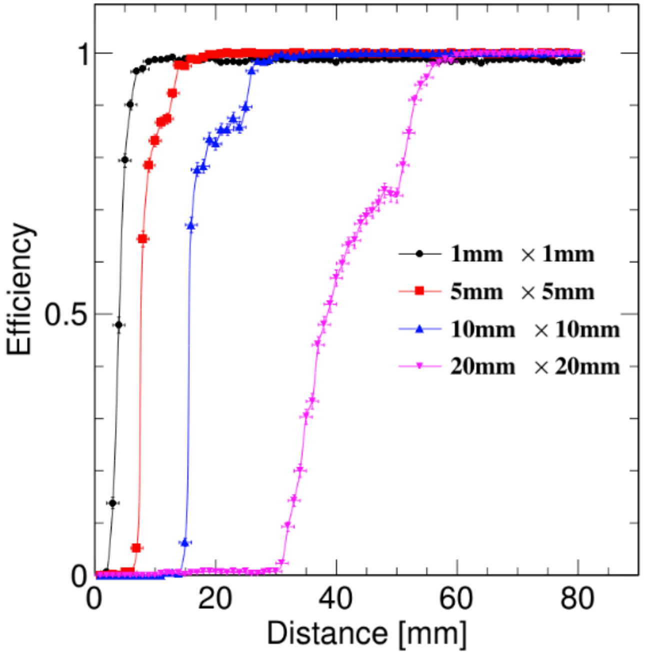
Performance

- Separation power:
  - Two 5 GeV γ shot in parallel, scan the distance.
  - At least 2 clusters, $|z_{rec} - z_{truth}| < 5mm$, $|E_{rec} - E_{truth}| < 2GeV$. $(E_{truth} = 5GeV)$



Non-100% efficiency @ large R

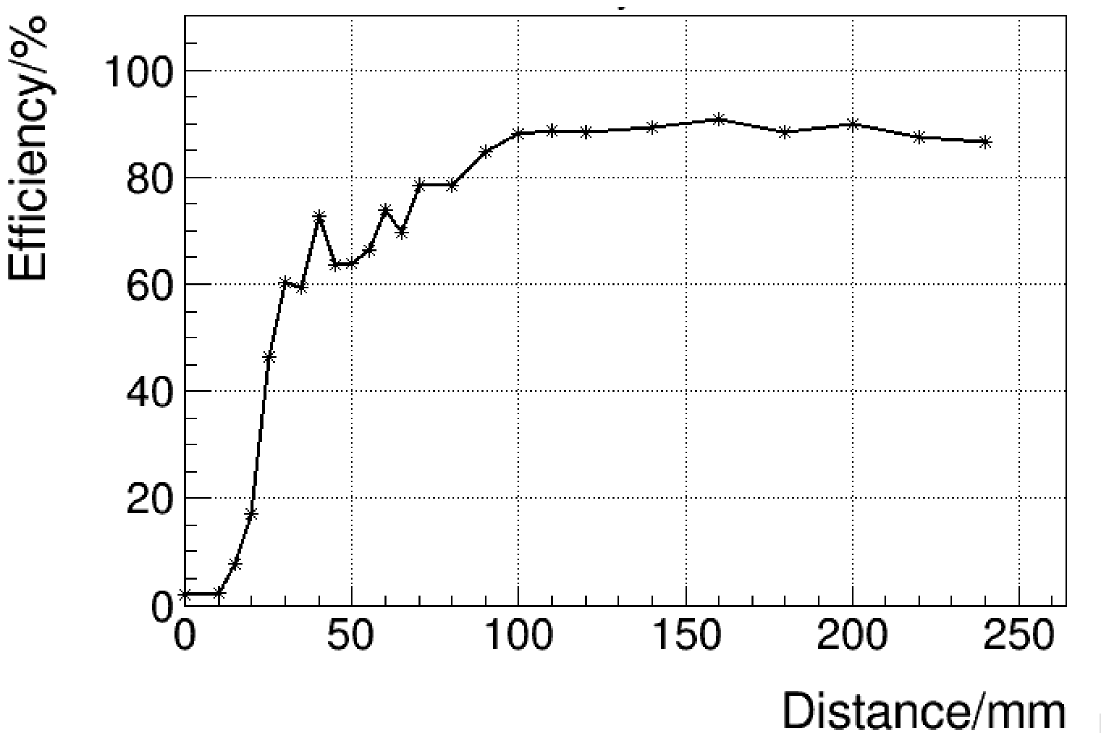
Good separation power @ small R. Limited by bar size.



Results from CEPC CDR: reconstruction efficiency of 2 pararrel 5 GeV photon.

Performance

- Separation power:
    - 10GeV π^- and 5 GeV γ shot in parallel, no B field, scan the distance.
    - at least 2 clusters, $\left|z_{rec}^{\gamma} - z_{truth}^{\gamma}\right| < 1cm, \left|z_{rec}^{\pi} - z_{truth}^{\pi}\right| < 2cm, \left|E_{rec}^{\gamma} - E_{truth}^{\gamma}\right| < 2GeV.$ $(E_{truth} = 5GeV)$



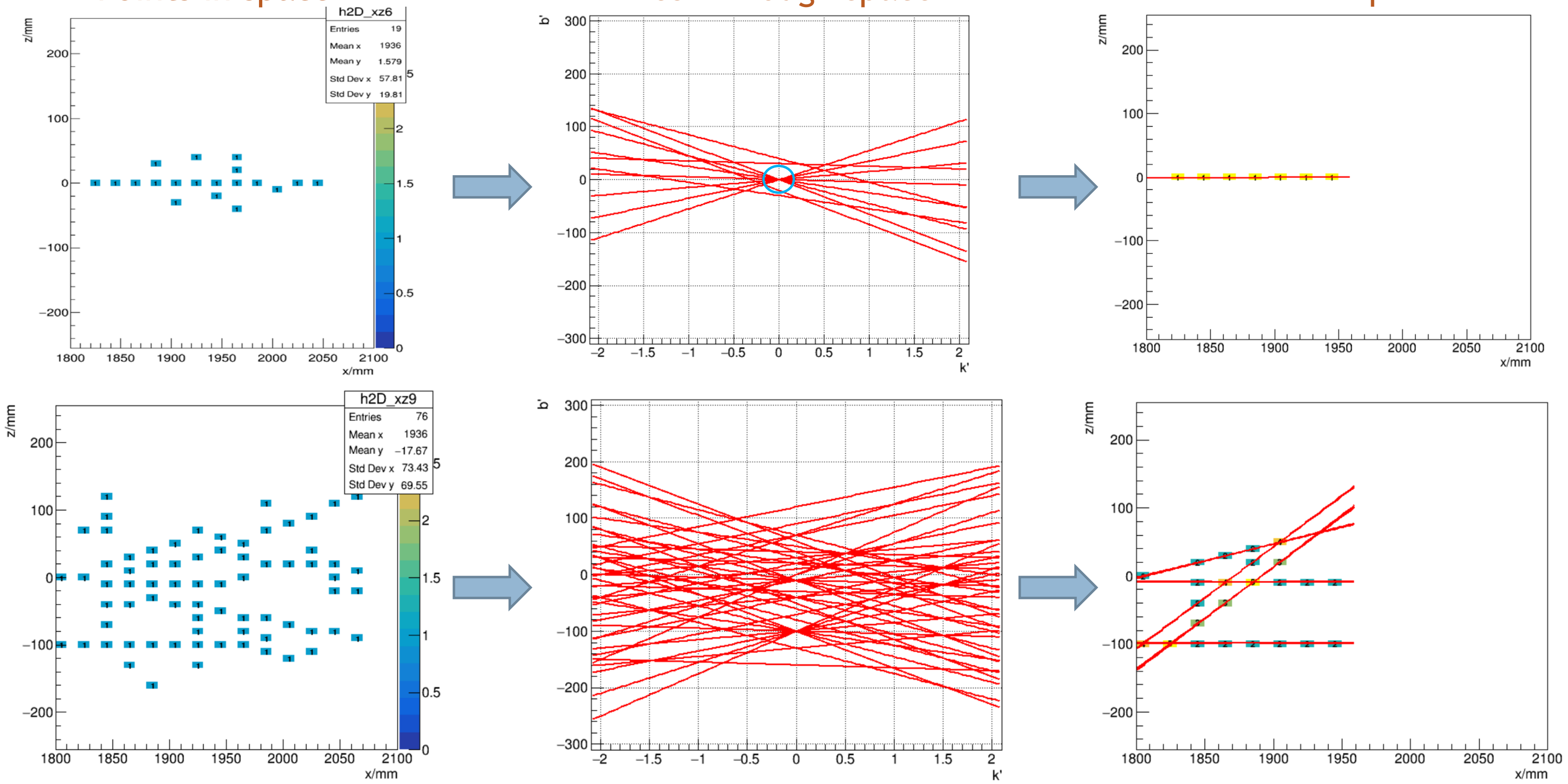
Reconstruction: scheme V1

- Use tracker idea in ECAL reconstruction: Hough transformation.
  - Consider from the simple case: EM shower core, no magnetic field.
  - Only use first 8 layers to avoid the shower in EM tail and complex hadronic shower.
  - Fit a "ECAL track" and match it with "tracker track". Details are still undergoing.

Points in space

Lines in Hough space

Back to space



single photon
5GeV

10 GeV pion
(z=0)
5 GeV photon
(z=100)

Summary

- Developing a new PFA software for crystal ECAL:
        - Traditional PFA: fine granularity + small R_M + less hits (sampling) for separation.
        - Crystal PFA: precise energy (homogeneous) + shower profile for separation.
    - χ^2 method for ghost hit removal is very efficient. ⇨ Ghost hit problem ✔☐
    - Energy splitting shows potential for particle separation. ⇨ Confusion ✔☐
    - Hadronic shower could be identified and would not influence the EM particle reconstruction.
    - Considering another scheme: fit “ECAL track”.

- Many details still need optimization:
    - Clustering efficiency,
    - Fragment absorption (cluster merging),
    - Cluster ID efficiency & mis-ID rate,
    - …

Backup

Cluster ID

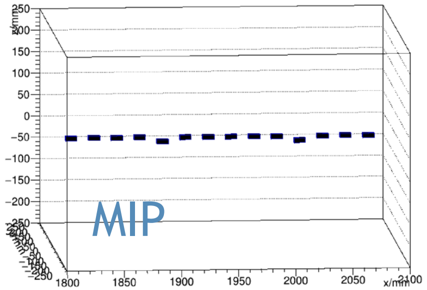
- Distinguish the MIP/EM/Had shower in crystal ECAL (after recon):
  - aveE: average energy in all layers.
  - stdDevE: standard derivation of energy in layers.

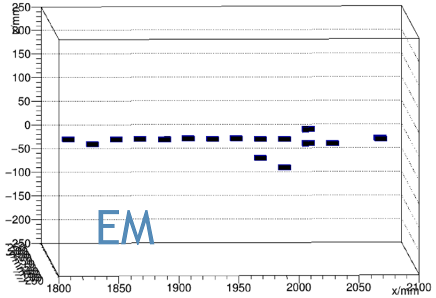
For MIP

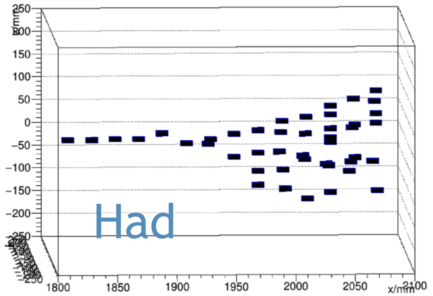
  - L_{start}: shower start layer, with $E_{this\ layer} > 0.1GeV\ (\sim 3MIP)$.
  - L_{end}: shower end layer.
  - showerMax: maximum width of this shower.
  - Alpha/Beta: fit longitudinal energy deposition with EM profile $\frac{\beta(\beta t)^{\alpha-1}e^{-\beta t}}{\Gamma(\alpha)}$, t is depth, α, β are fitted parameters.
  - Chi2: χ^2/N_{dof} for this longitudinal profile fitting.

For Had

- Considering a cut-based PID as the first step.

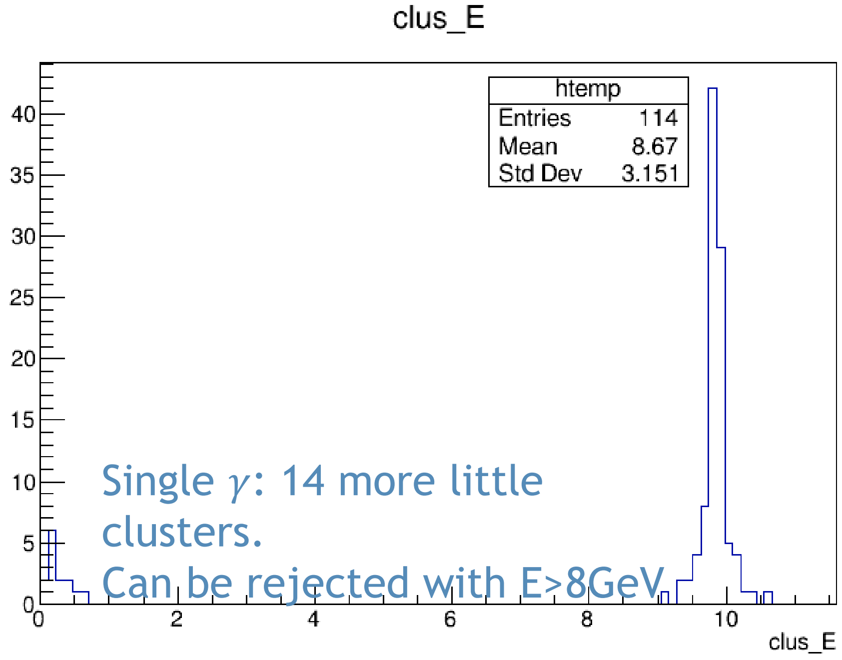






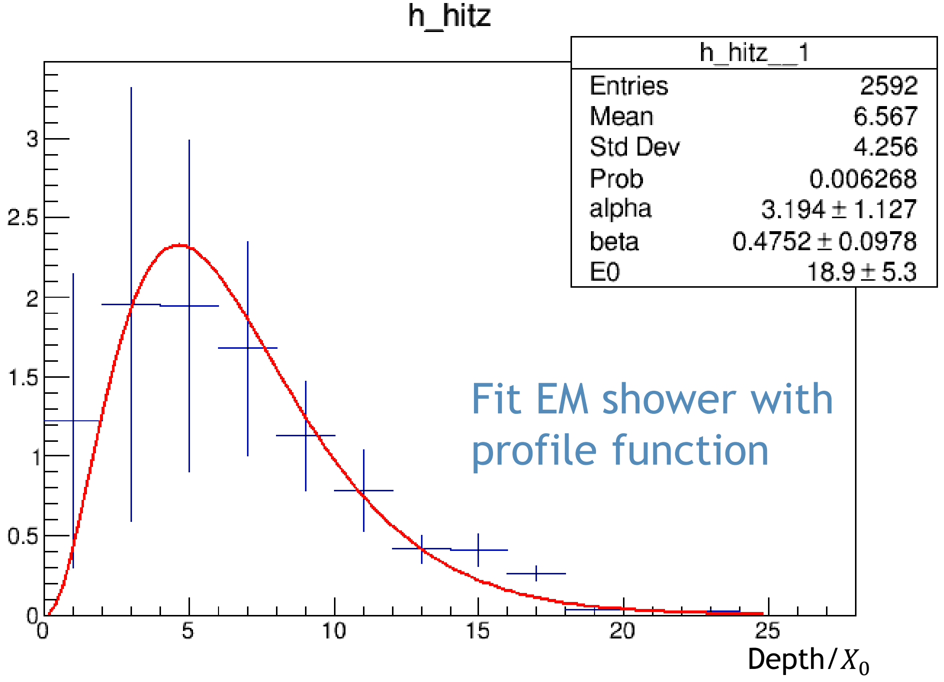
Cluster ID

- 100 events for each particle, 10GeV muon/photon/pion.
  - In Gam sample: require E>8GeV to remove wrongly reconstructed clusters.



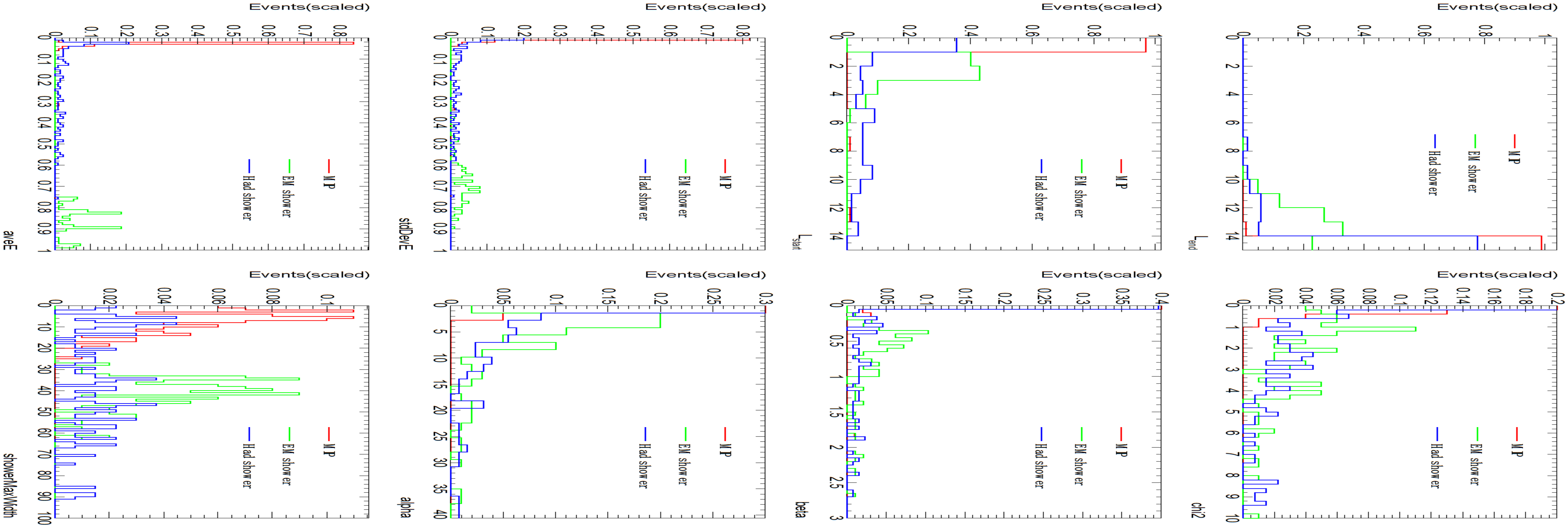
Single γ: 14 more little clusters.

Can be rejected with E>8GeV



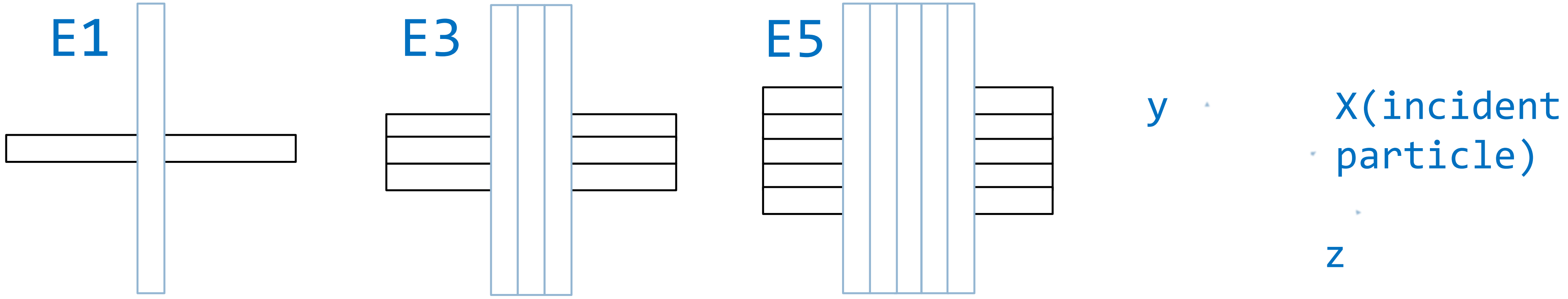
Fit EM shower with profile function

Cluster ID


Events(scaled)
MIP
EM shower
Had shower
L_{end}
L_{start}
stdDevE
aveE
chi2
beta
alpha
showerMaxWidth

PID using lateral shape

- For horizontal and vertical bars:
  - Lateral moment $LAT = \frac{\sum_{i=3}^{N} E_i r_i^2}{\sum_{i=3}^{N} E_i r_i^2 + E_1 r_0^2 + E_2 r_0^2}$
  - Second moment $SEC = \frac{\sum_{i=1}^{N} E_i r_i^2}{\sum_{i=1}^{N} E_i}$
  - Energy of the energetic crystal $E1$
  - Energy sum of 3×3 crystals $E3$
  - Energy sum of 5×5 crystals $E5$

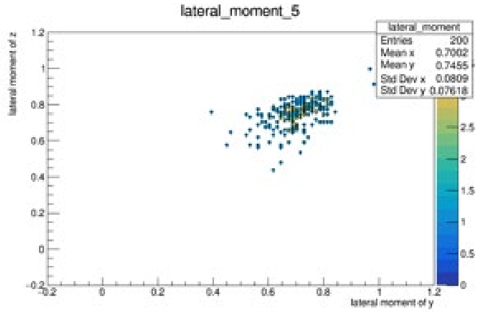


PID performance

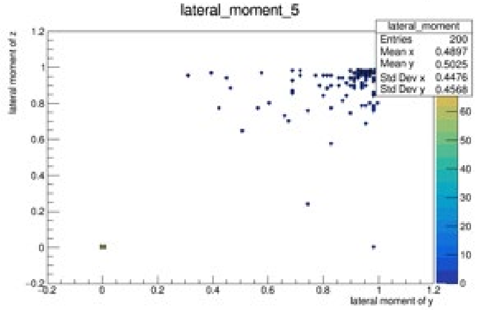
- Confusion matrix
  - lateral moment of super layer 5
  - $laty < 0.4$ && $latz < 0.4$: MIP
  - $laty \in [0.4,0.9]$ && $latz \in [0.4,0.9]$: EM
  - $laty > 0.9$ && $latz > 0.9$:Had
  - $others$: LOSS

PID Truth 10GeV	MIP	EM	Had
MIP	0.975	0.01	0
EM	0	0.99	0.01
Had	0.44	0.1	0.275

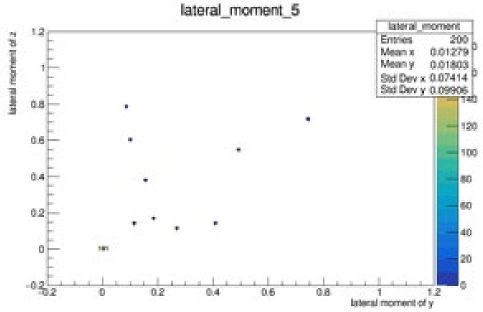
EM



Had

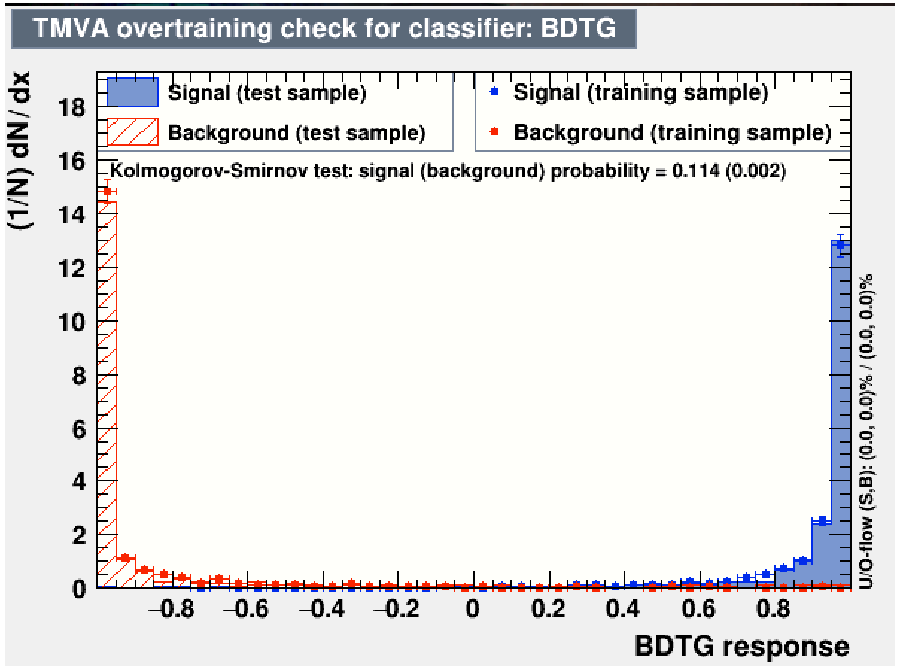


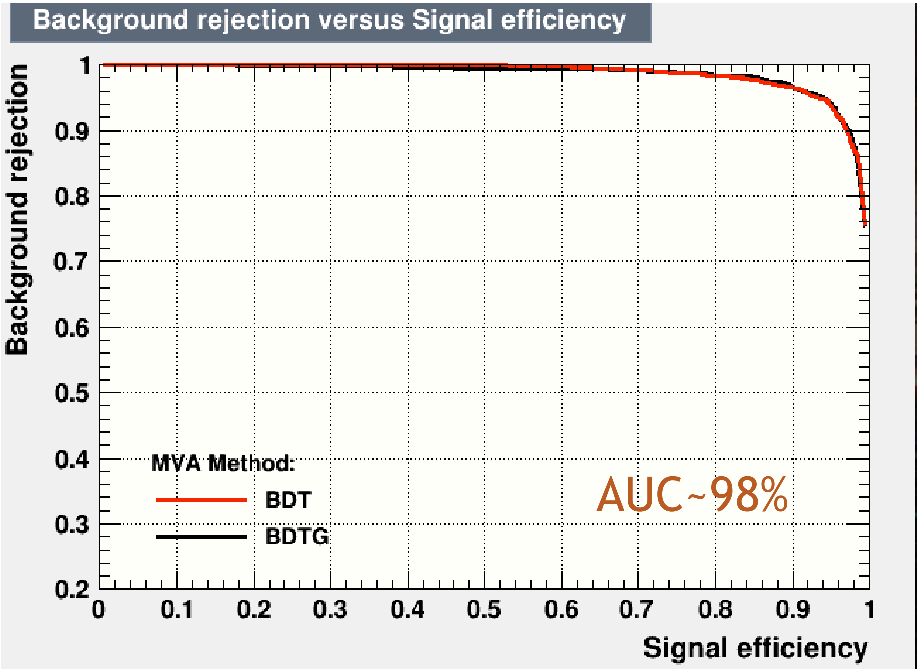
MIP



Cluster ID

- Potential with BDT:
  - 6 input variables for training, EM vs. Hadronic cluster: $L_{end}, L_{maxWidth}, L_{maxE}, E_{max}, width_{max}, \alpha, \beta$.
  - Training with BDTG: NTrees=900:MinNodeSize=2.5%:BoostType=Grad:Shrinkage=0.06:UseBaggedBoost:BaggedSampleFraction=0.5:nCuts=20:MaxDepth=3

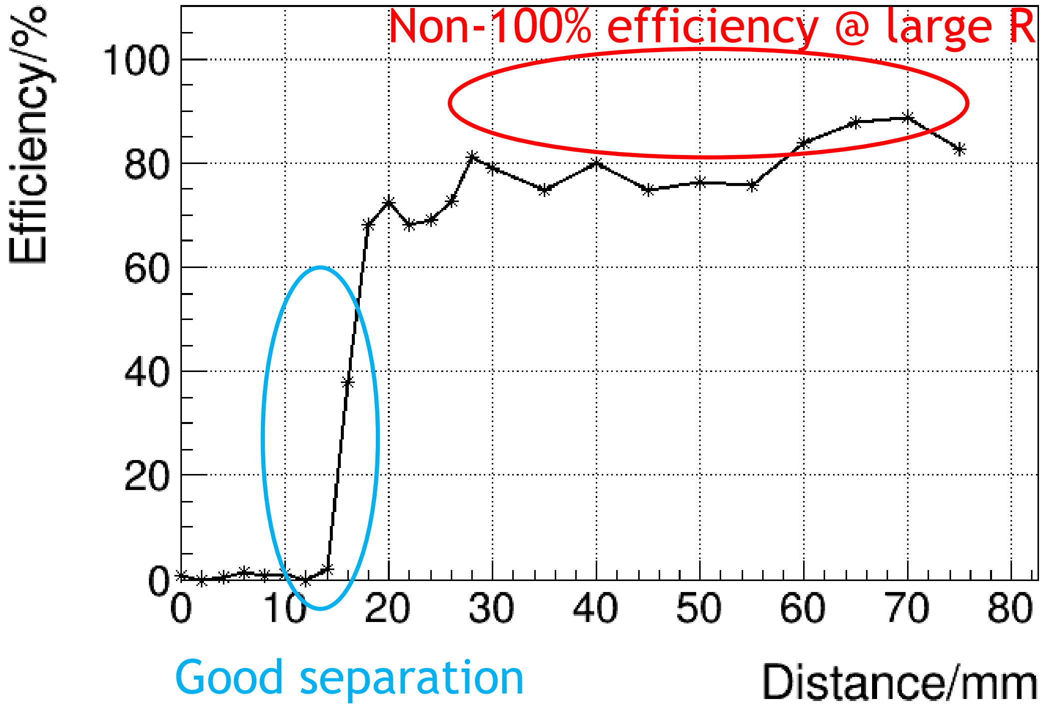




AUC~98%

Performance

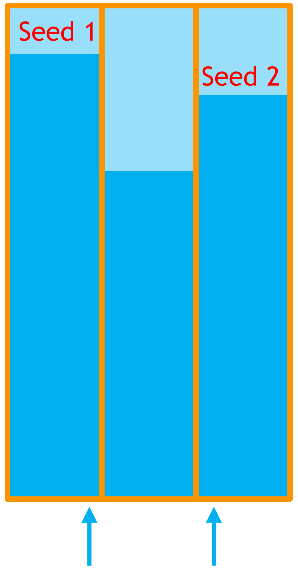
- Separation power:
  - Two 5 GeV γ shot in parallel, scan the distance.
  - At least 2 clusters, $|z_{rec} - z_{truth}| < 5mm$, $|E_{rec} - E_{truth}| < 2GeV$. $(E_{truth} = 5GeV)$



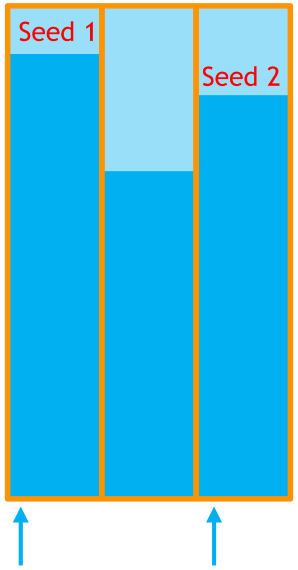
Non-100% efficiency @ large R

Good separation power @ small R.

Limited by bar size.



Most ideal case:
$R_{\gamma-\gamma} \sim > 1\ cm$



Worst case:
$R_{\gamma-\gamma} \sim 2\ cm$