Towards a clustering algorithm for CALICE

Chris Ainsley

University of Cambridge

<ainsley@hep.phy.cam.ac.uk>

General CALICE meeting

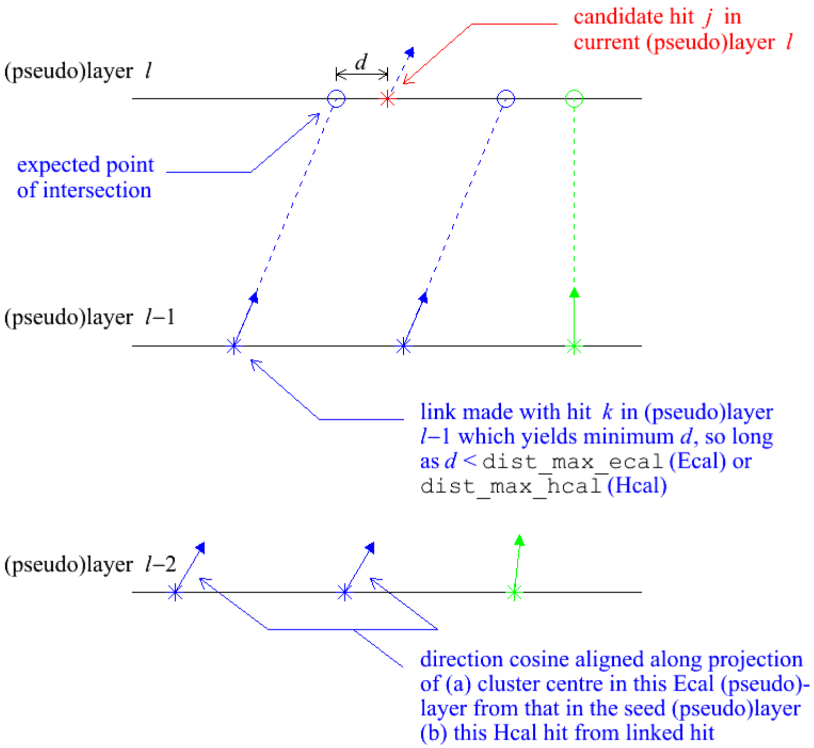
7-8 December 2004, DESY, Germany

Order of service

- Layer-by-layer approach to clustering.
- Application to a generalised calorimeter.
- Reconstructed event gallery for two close-by particles.
- How to quantify the two-particle separation “quality” and use it to optimise clustering cuts.
- Quality studies for nearby $\pi^+\gamma$, π^+n, $\pi^+\pi^+$ and nn.
- Quality dependence on hadronic shower model.
- Summary.

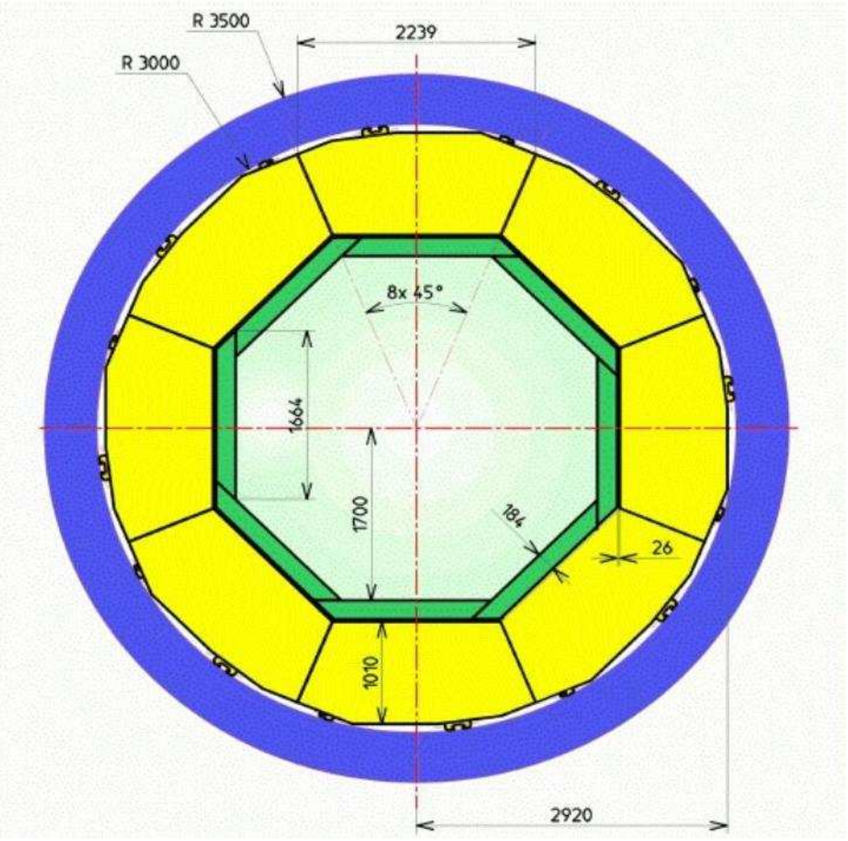
Layer-by-layer clustering: the algorithm

- Form clusters by tracking closely-related hits (> 1/3 mip) *layer-by-layer* through calorimeter:
  - for a given hit *j* in a given layer *l*, minimize the distance *d* w.r.t all hits *k* in layer *l*–1;
  - if *d* < `dist_max_ecal` (Ecal) or `dist_max_hcal` (Hcal) for minimum *d*, assign hit *j* to same cluster as hit *k* which yields minimum;
  - if not, repeat with all hits in layer *l*–2, then, if necessary, layer *l*–3, etc., right through to first layer of Ecal;
  - after iterating over all hits *j*, seed new clusters with those still unassigned;
  - if in Ecal, calculate weighted centre of each cluster's hits in layer *l* (weight by energy (analogue) or density (digital)) and assign a direction cosine to each hit along the line joining its cluster's centre in the seed layer (or (0,0,0) if it's a seed) to its cluster's centre in layer *l*;
  - if in Hcal, assign a direction cosine to each hit along the line from the hit to which each is linked (or (0,0,0) if it's a seed) to the hit itself;
  - try to recover backward-spiralling track-like, and low multiplicity 'halo', cluster fragments …



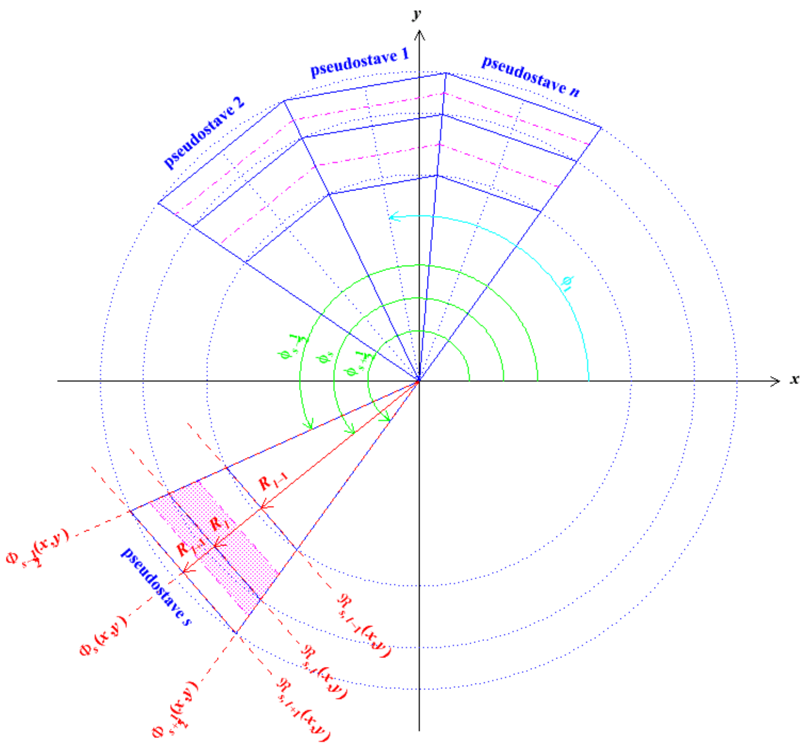
Layer-by-layer clustering in a generalised detector

- Approach requires layer index to vary smoothly: e.g. in CALICE, index changes abruptly
  - at stave boundaries in Ecal barrel (layers overlap at 45°);
  - at barrel/endcap boundaries in Ecal & Hcal (layers overlap at 90°).
- Scheme developed (see CERN, Durham talks) to overcome problem; extended to apply to any arbitrary *n*-fold rotationally-symmetric, layered calorimeter.
- Achieved by replacing layer index with *pseudolayer* index in regions where layer index discontinuities occur.
- Same-pseudolayer indexed hits defined by closed shells of *n*-polygonal prisms (e.g. CALICE: $n = 8$ $\Rightarrow$ octagonal prisms) coaxial with *z*-axis.
- Locations/orientations of shells automatically set by locations/orientations of real, physical, sensitive layers.
- Just takes *n* and layer-spacings in barrel and endcaps as input.

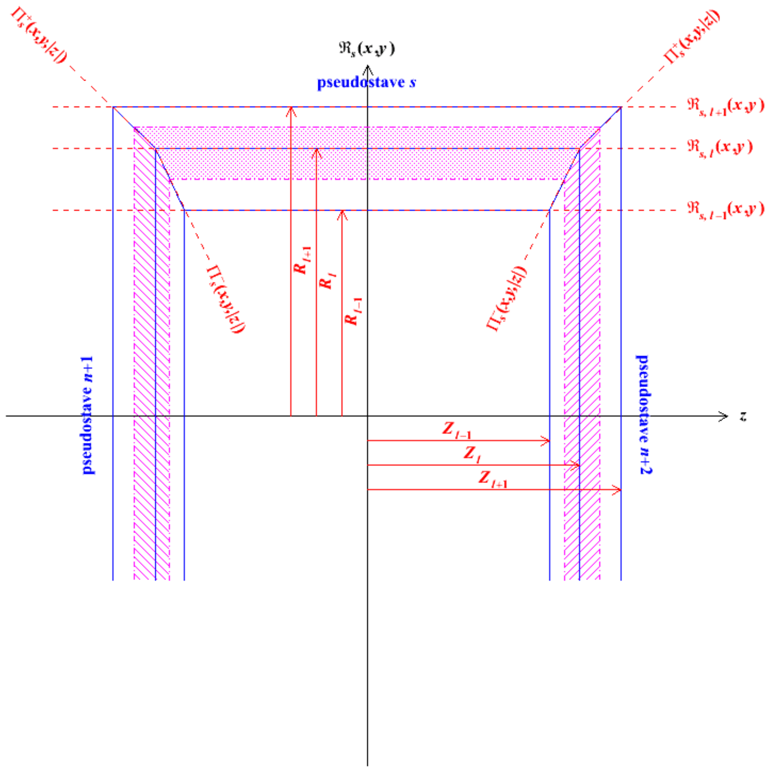


How the generalised detector shapes up

Transverse section



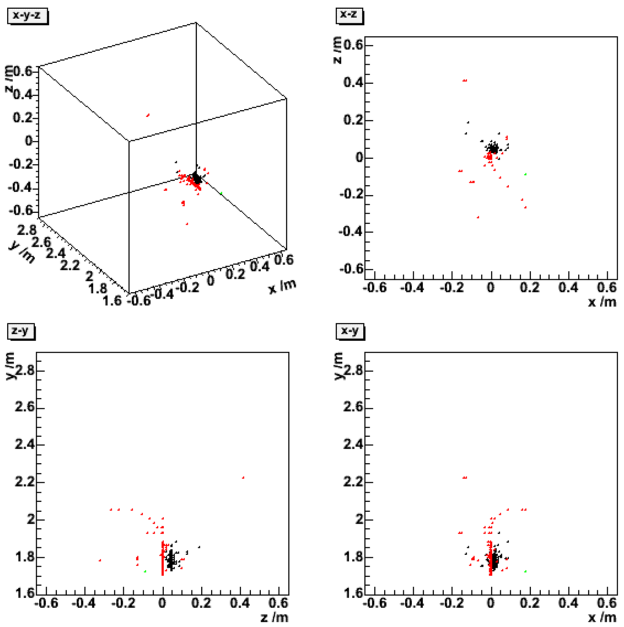
Longitudinal section

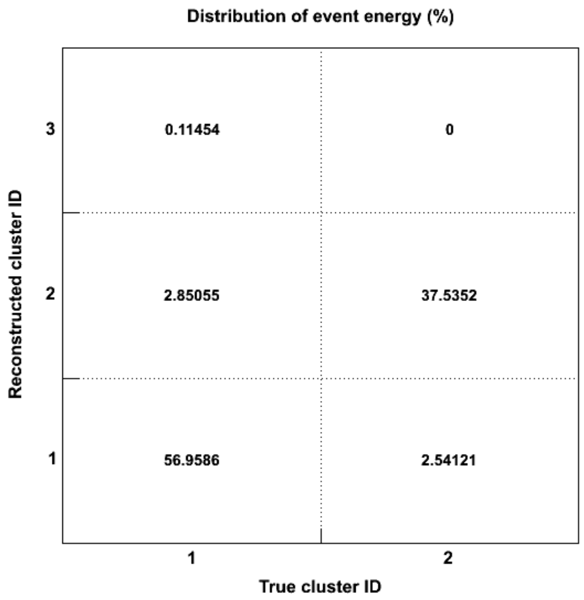


- Solid blue lines aligned along real, physical, sensitive layers.
- Dot-dashed magenta lines bound shell containing hits with same *pseudolayer* index, *l*.

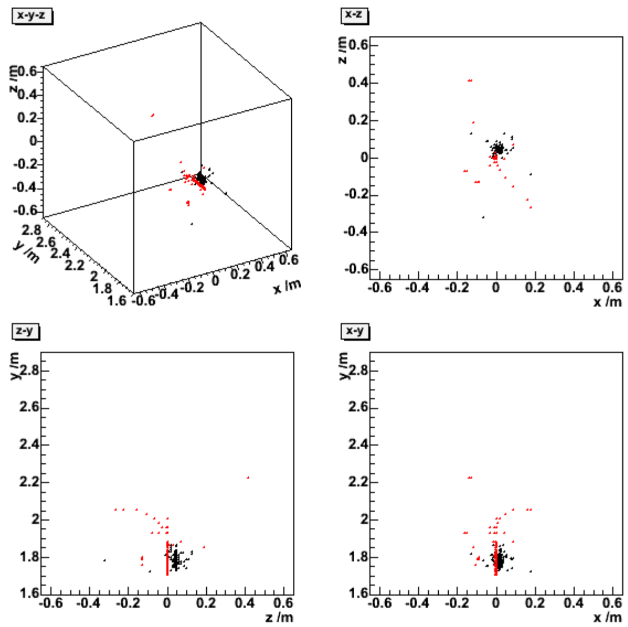
5 GeV $\pi^+\gamma$ event at 5 cm separation

Reconstructed clusters





True particle clusters

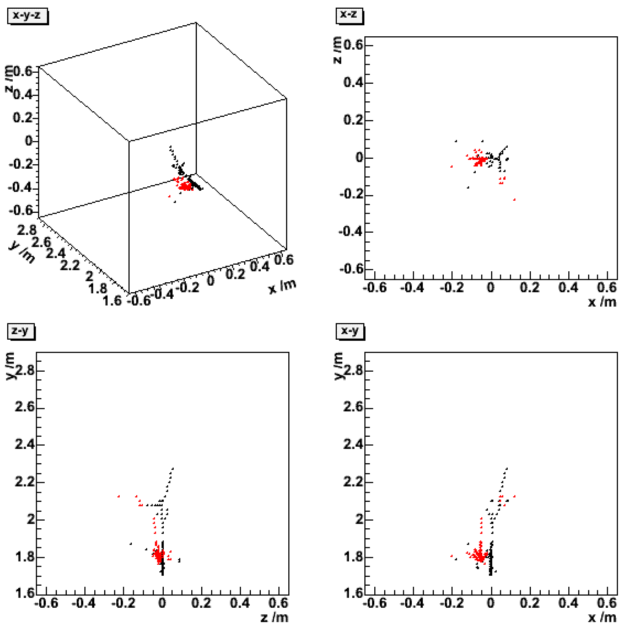


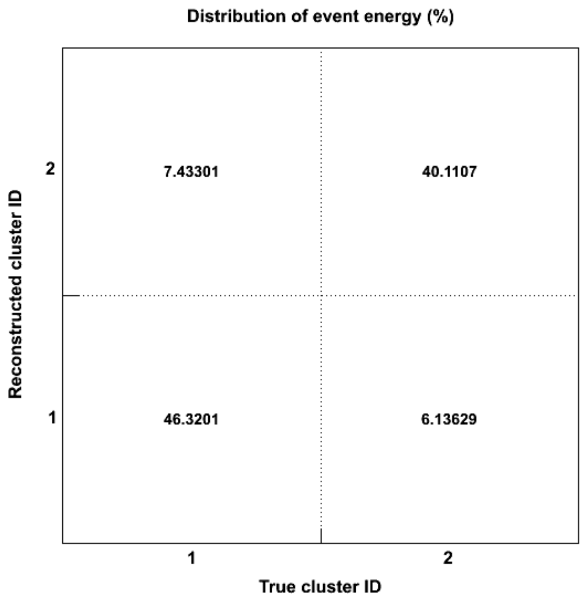
- Energy maps mostly *black* ↔ *black* (γ) and *red* ↔ *red* (π^+) .
- Quality = 57.0 + 37.5 = 94 %.

Chris Ainsley
<ainsley@hep.phy.cam.ac.uk>

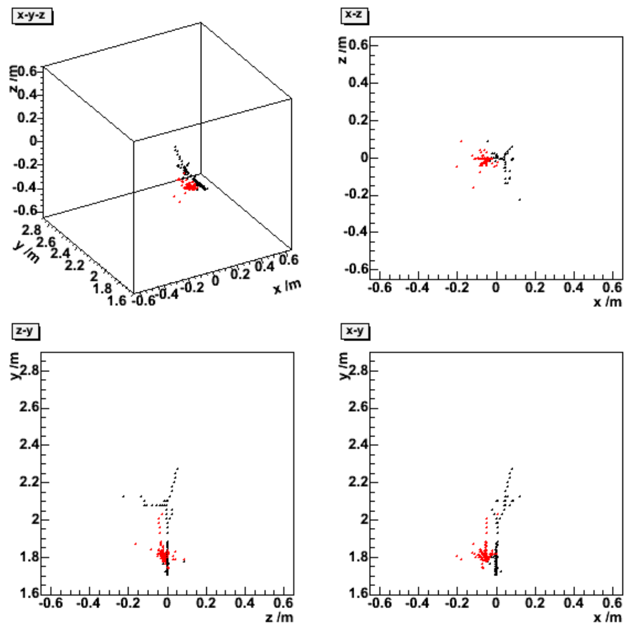
5 GeV π^+n event at 5 cm separation

Reconstructed clusters





True particle clusters

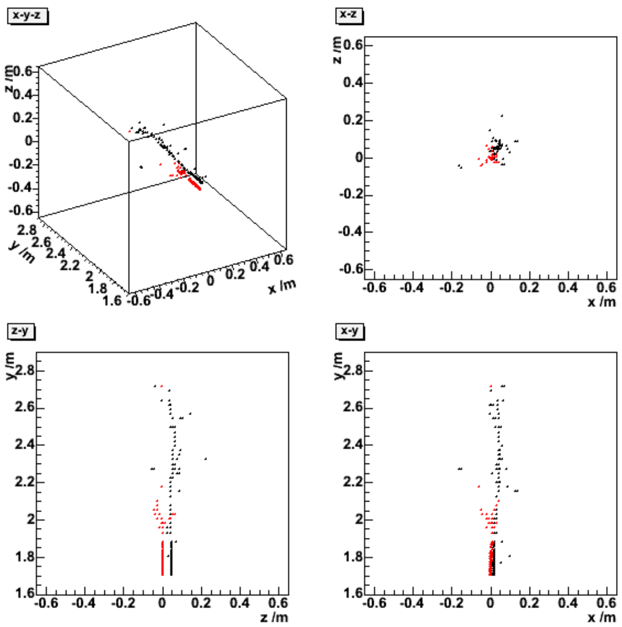


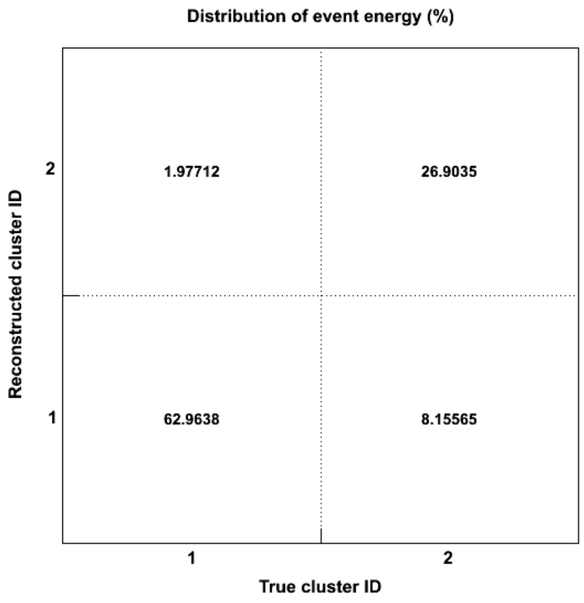
- Energy maps mostly *black* ↔ *black* (π+) and *red* ↔ *red* (n) .
- Quality = 46.3 + 40.1 = 86 %.

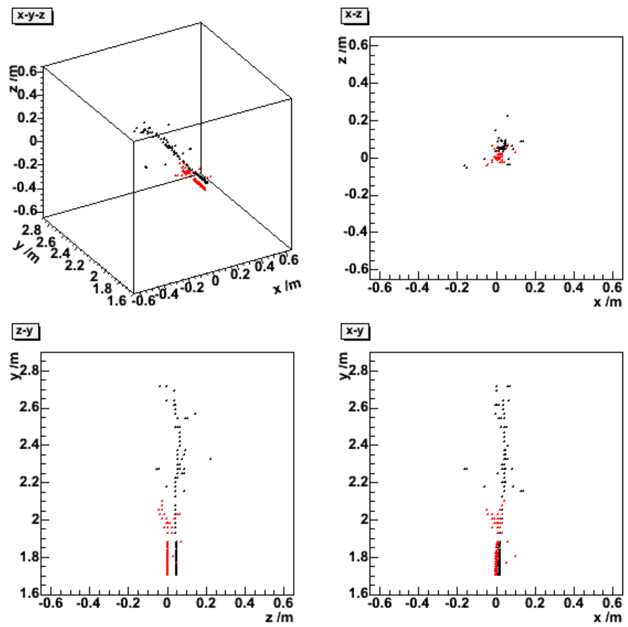
5 GeV $\pi^+\pi^+$ event at 5 cm separation

Reconstructed clusters

True particle clusters



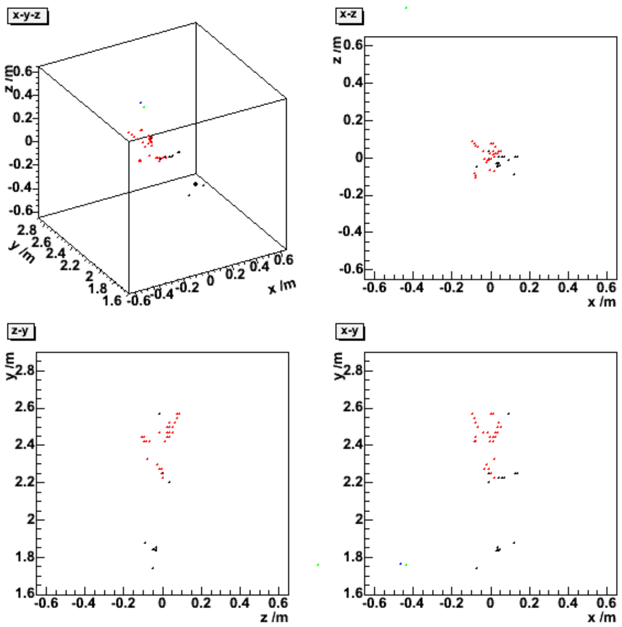


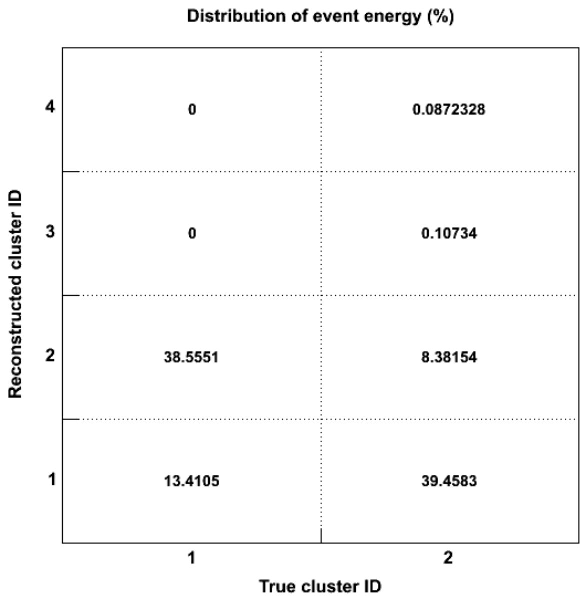


- Energy maps mostly *black* ↔ *black* and *red* ↔ *red*.
- Quality = 63.0 + 26.9 = 90 %.

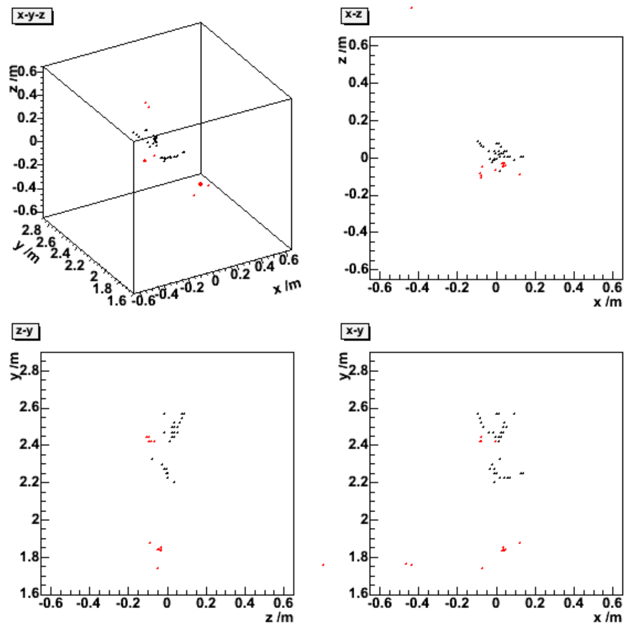
5 GeV nn event at 5 cm separation

Reconstructed clusters





True particle clusters



- Energy maps mostly *black* ↔ *red* and *red* ↔ *black* .
- Quality = 39.5 + 38.6 = 78 %.

Two-particle separation quality: definition

- Need to grade performance of clustering algorithm (in absence of full particle-flow algorithm).
- Want to optimise both:
  - *efficiency* – how closely true particle clusters correspond to reconstructed clusters; and
  - *purity* – how closely reconstructed clusters correspond to true particle clusters.
- Propose a figure of merit:

Quality = fraction of event energy that maps in a 1:1 ratio between reconstructed and true clusters.

- Combines efficiency and purity into a single, useful measure.
- For two equal-energy particles, expect
  - no clustering (*i.e.* "hit" = reconstructed cluster):

    ⇒ energy in true clusters divided between many reconstructed clusters;

    ⇒ quality → 0;
  - over-exaggerated clustering (i.e. "event" = reconstructed cluster):

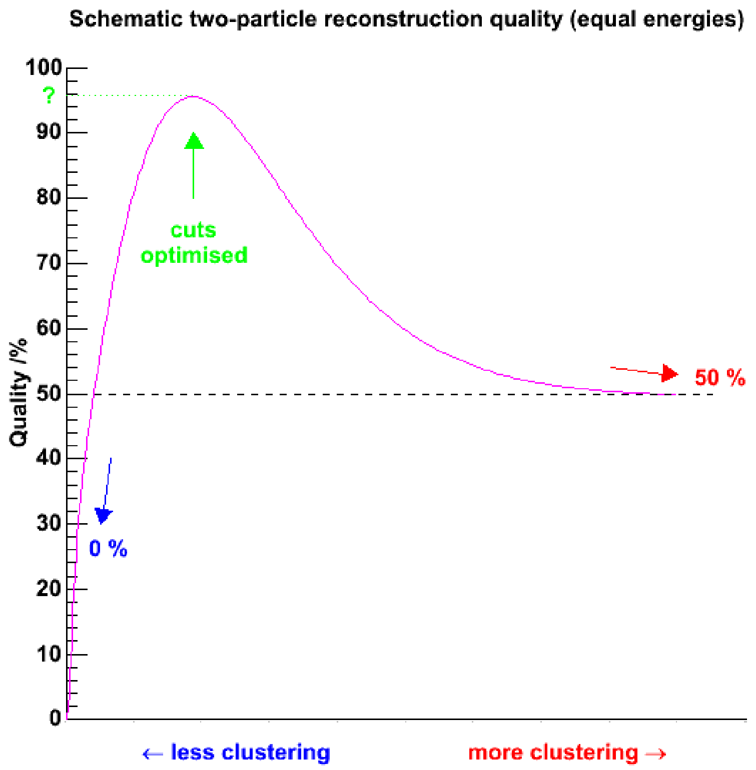
    ⇒ energy in single reconstructed cluster divided between two (equal-energy) true clusters;

    ⇒ quality → 50 %;
  - optimal clustering:

    ⇒ lies somewhere in between; where?

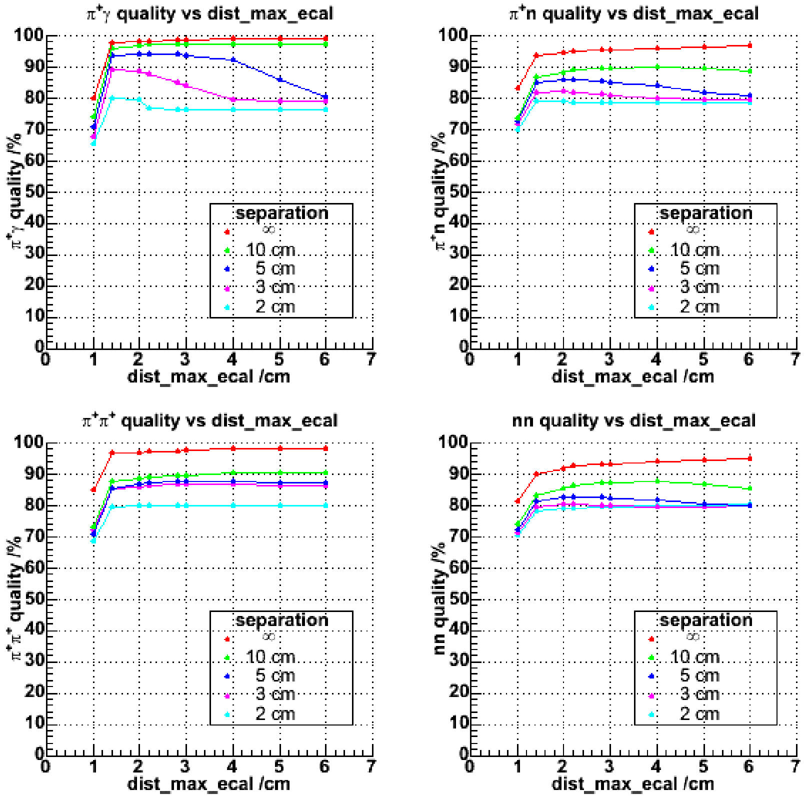
    ⇒ quality = ?
- Demonstrate principle by varying the `dist_max_ecal` and `dist_max_hcal` cuts.
- Energy calibrated (D09 detctor) according to:

$E = \alpha[(E_{\mathrm{Ecal;\,1\text{-}30}} + 3E_{\mathrm{Ecal;\,31\text{-}40}})/E_{\mathrm{mip}} + 20N_{\mathrm{Hcal}}]$ GeV.

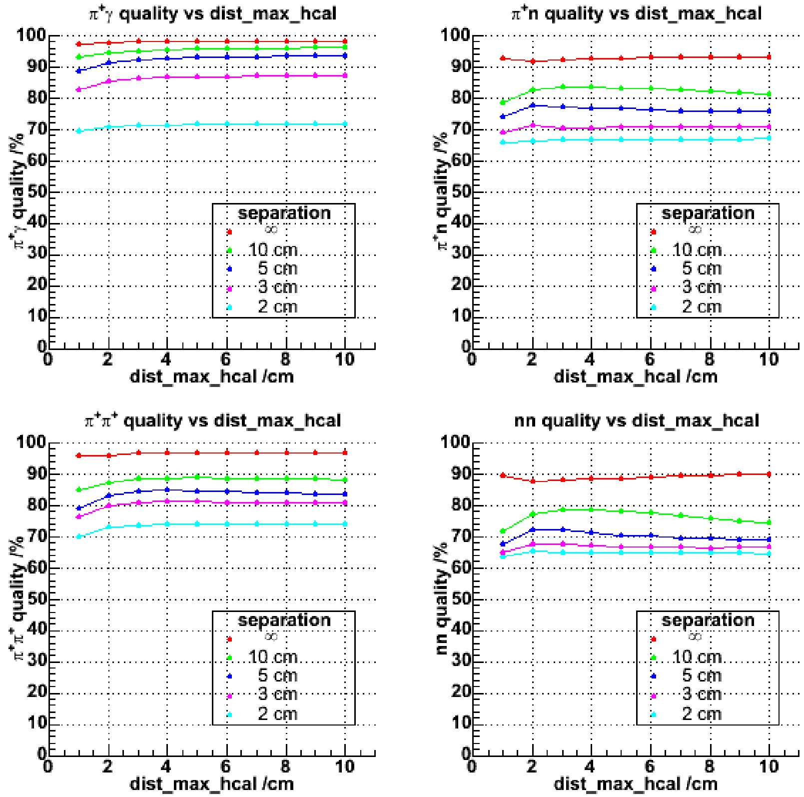


5 GeV two-particle quality vs clustering cuts

Ecal only



Ecal + Hcal

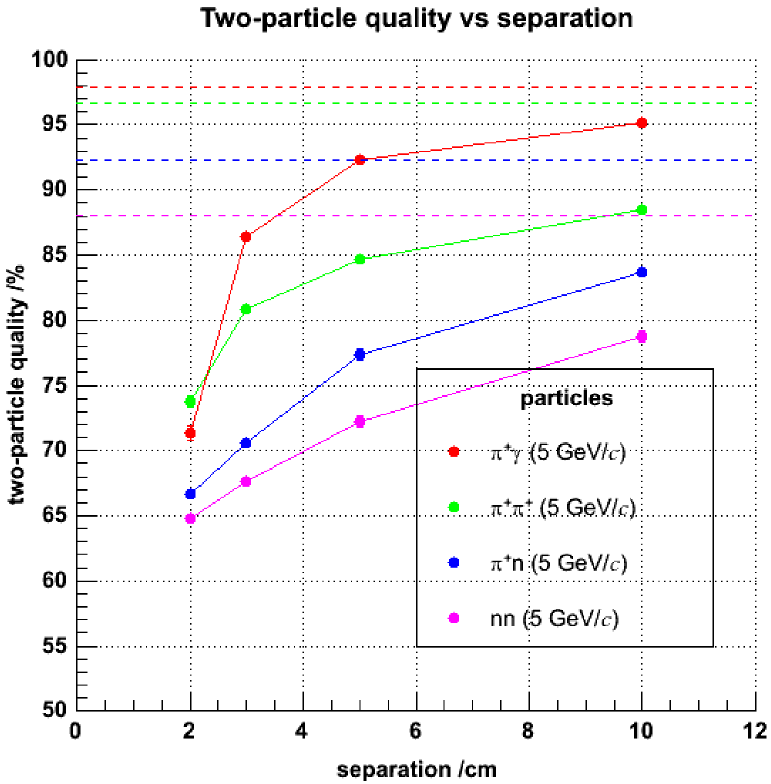


- Ecal quality peaks/plateaus (all particles/ separations) around `dist_max_ecal` = 2 cm.
- Physically reasonable (1×1 cm^2 cells). Fix it.

- Hcal quality peaks/plateaus (all particles/ separations) around `dist_max_hcal` = 3 cm .
- Again, physically reasonable. Fix this too.

5 GeV two-particle quality vs separation

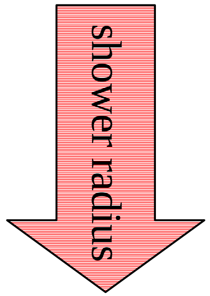
- Goal: to distinguish charged clusters from neutral clusters in calorimeters.
- Separation of $\pi^+\gamma$ and π^+n very important; that of $\pi^+\pi^+$ and nn less so (but still interesting).
- Quality improves with separation (naturally).
- $\pi^+\gamma$ separation at 5 GeV seems to be pretty good; π^+n is somewhat tougher (n by itself is tricky – dashed magenta line).
- Do things change much with energy / incident angle / other pairs of particles / pad-size / hadronic shower model…?



`dist_max_ecal` = 2.0 cm (fixed);
`dist_max_hcal` = 3.0 cm (fixed).

$\pi^+\pi^+$ quality vs hadronic shower model

- Survey by G. Mavromanolakis (see CERN, Durham talks) ⇒ different hadronic shower models give significant variations in predicted shower radius (≈ 35 % for 10 GeV π^+).
- Looked at dependence of quality on model for two 5 GeV π^+ separated by 5 cm:
  - LHEP 85.5 ± 0.4 %
  - QGSP_BIC 84.9 ± 0.3 %
  - LHEP_BERT 81.8 ± 0.4 %
  - LCPhys 81.8 ± 0.4 %.



- Quality decreases with increasing shower radius (as expected).
- Similar conclusions found with other separations; also for single π^+.
- Hadronic shower model impacts on pattern recognition predictions; ultimately significant for detector design.

Summary & outlook

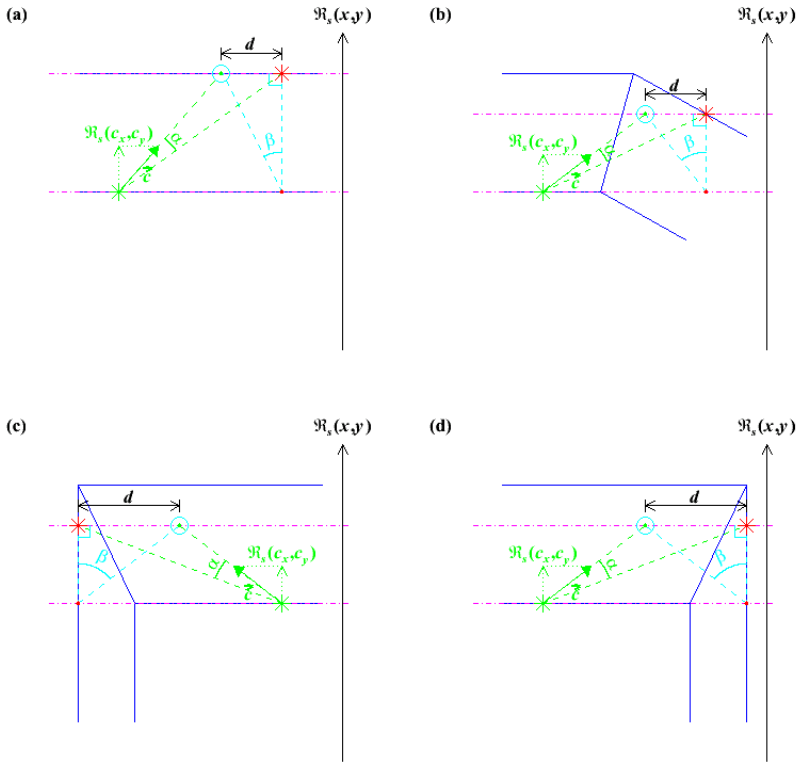
- R&D on clustering algorithm for CALICE on-going.
- Approach utilizes the high granularity of the calorimeter cells to "*track*" clusters (pseudo)layer-by-(pseudo)layer.
- Written in C++; LCIO (v1.3) compliant.
- Pseudolayer concept $\Rightarrow$ flexibility to cope with alternative layered geometries without having to recode algorithm itself.
- Introduced quality gauge to assess performance of algorithm w.r.t. charged/neutral cluster separation.
- Using it
  - to guide refinements to algorithm and optimise clustering cuts.
  - to compare relative merits of different detector layouts.
- If considered helpful, can aim to make code publicly available within ~6 months.

The end

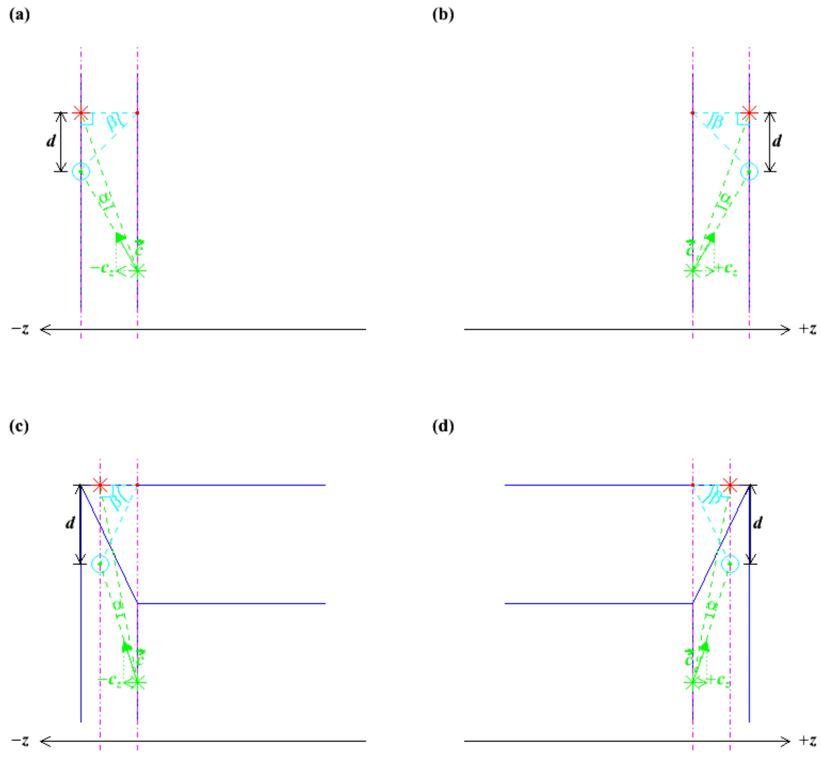
That's all folks…

Cluster-tracking between pseudolayers

From the pseudobarrel

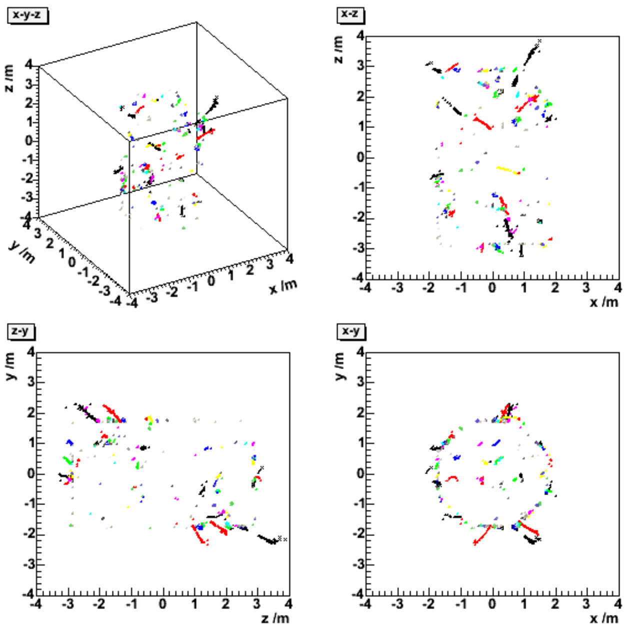


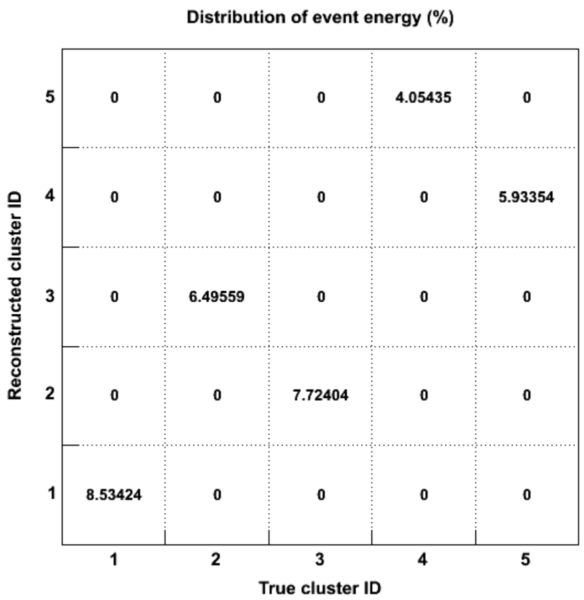
From the pseudoendcap



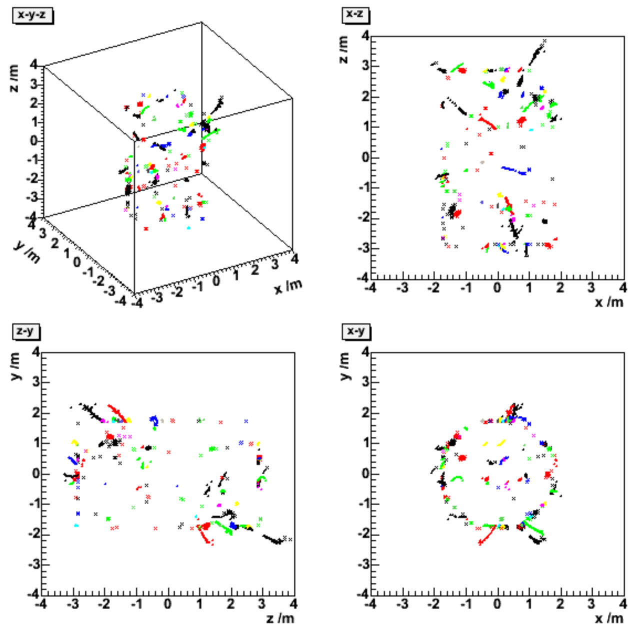
91 GeV $Z \rightarrow u,d,s$ jets event

Reconstructed clusters





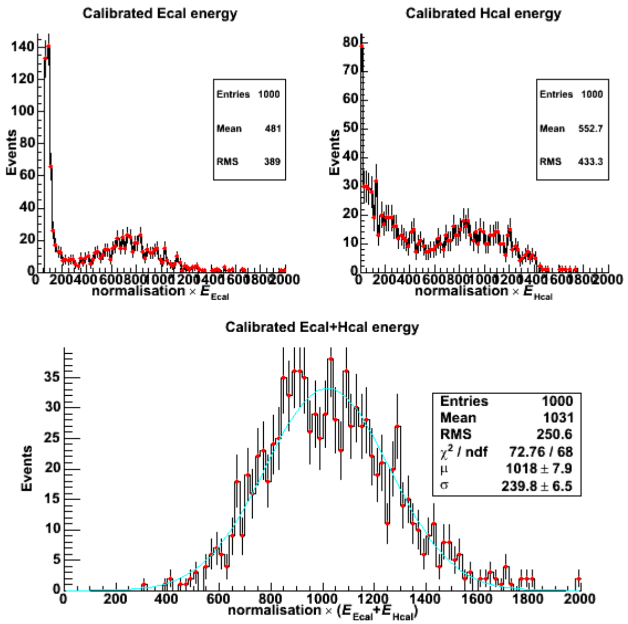
True particle clusters



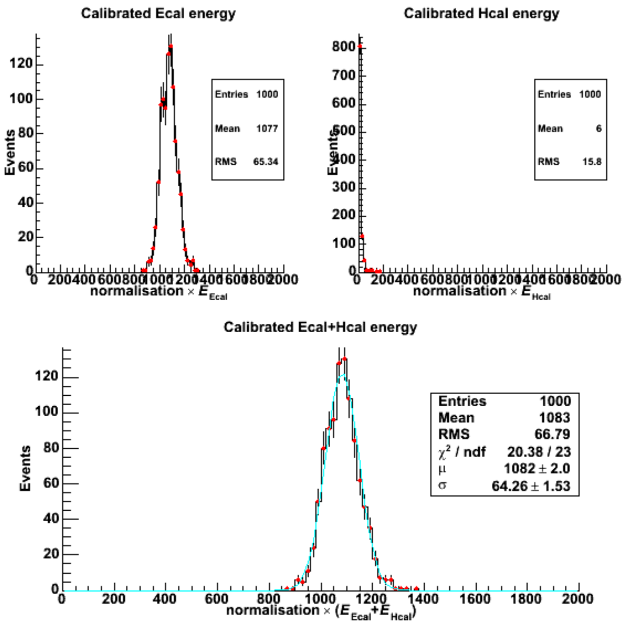
- Reconstruction in full detector (Si/W Ecal & RPC Hcal; 1×1 cm^2 cells).
- `dist_max_ecal` = 2.0 cm; `dist_max_hcal` = 3.0 cm.
- Good 1:1 correspondence between reconstructed and true clusters (5 highest energy clusters shown).

Calibration of π⁺, γ and n

π^+



γ



n