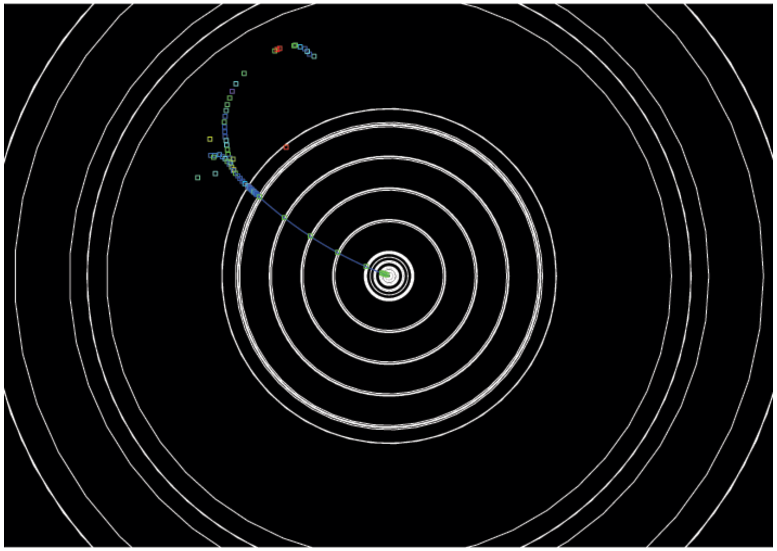
A hit density driven clustering algorithm for PFA study

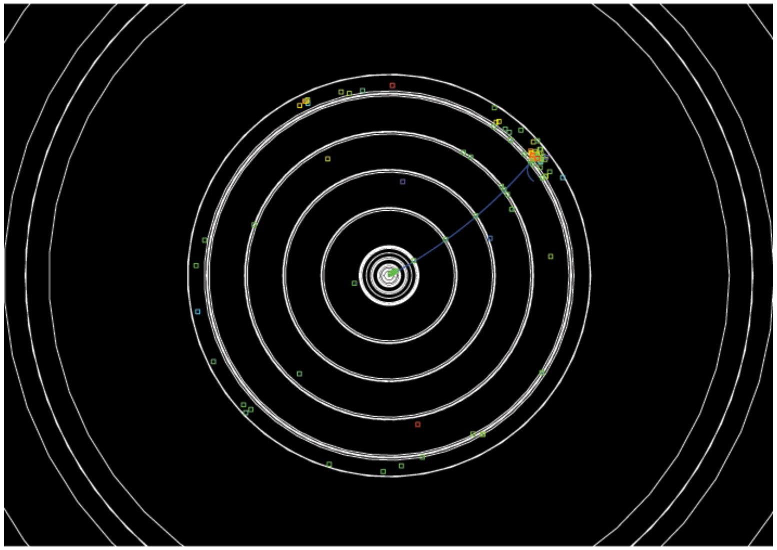
Lei Xia
ANL-HEP

SiD calorimeter WG

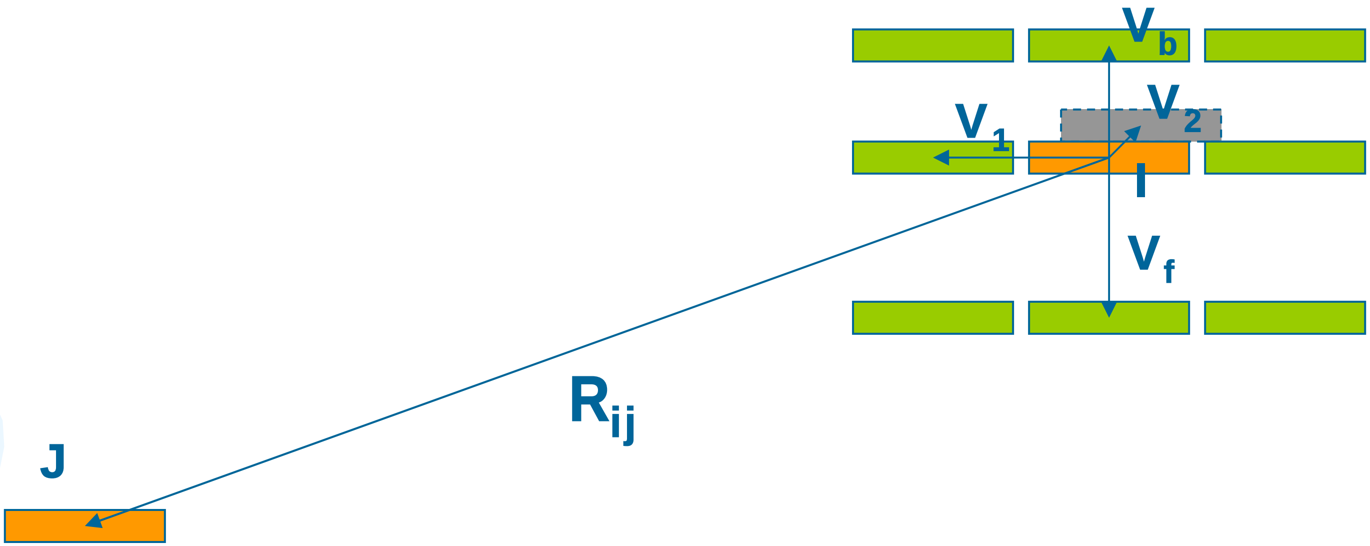
Introduction

- Goal: to have a clustering algorithm that can
  - Form clusters that can closely represent single particle shower
    - Pick up as many hits as possible for a single particle
    - Distinguish different particles
  - Treat ECal and HCal as one detector
    - Treat cell/layer structure properly
    - Cluster doesn't break up at boundaries
  - Adjustable parameter for PFA
- Reality: hadron showers have hits all over the detector
  - Impossible to pick up every hits of a shower without messing up different showers
  - Try to pick up only the central part of a shower, and deal with fragments later
  - Use hit density to drive the clustering
  - Using sidmay05 simulation, with RPC DHCal (projective geometry, barrel cell size 7.4 – 12.mm)





Clustering: hit density

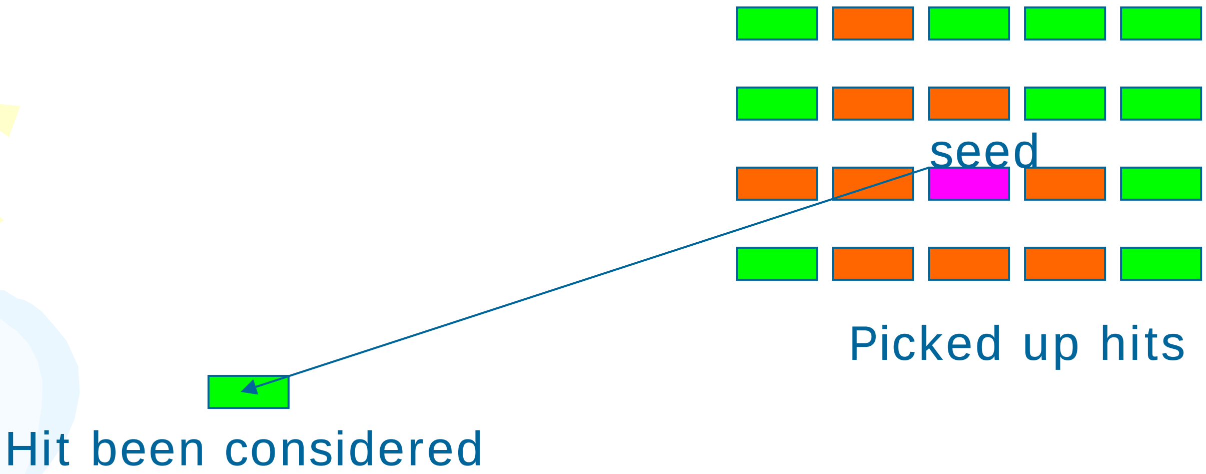


$$D_{ij} = e^{-(|(\hat{V}_1, R_{ij})|/|V_1|)} \times e^{-(|(\hat{V}_2, R_{ij})|/|V_2|)} e^{-((\hat{V}_3, R_{ij})/|V_3|)}$$

With $V_3 = V_f$ (if $(V_f, R_{ij}) > 0$) or V_b (if $(V_b, R_{ij}) > 0$)

- Try to find a two-point density function which can reflect the closeness of two hits (relative to local cell size)
- Considered hit density variation in different directions
- It is a very local density function, only nearby hits contribute
- So far, didn't consider hit energy weighting

Clustering: grow a cluster



- Find a cluster seed: hit with highest density among remaining hits
- Attach nearby hits to a seed with a tight cut on hit-seed density
- Attach additional hits with a tight cut on hit-seed cluster density
  - EM hits, D(hit,cluster) > 0.01
  - HAD hits, D(hit,cluster) > 0.001
  - Grow the cluster until no hits can be attached to it
- Find next cluster seed, until run out of hits

Hit efficiency: single particle

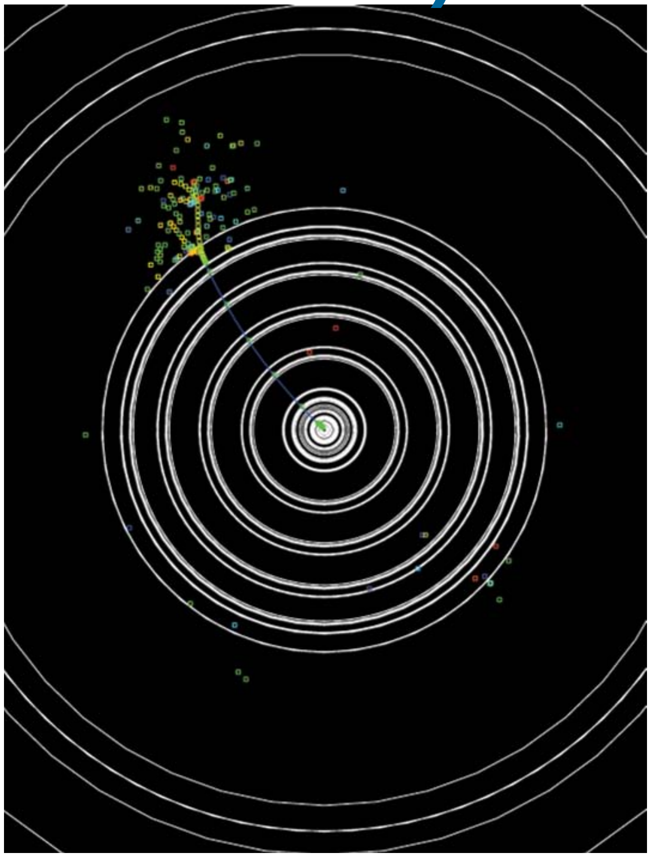
Particle	ECal hit efficiency	HCal hit efficiency	Overall hit efficiency	Overall energy efficiency
Photon (1GeV)	89%	43%	89%	91%
Photon (5GeV)	92%	54%	92%	96%
Photon (10GeV)	92%	61%	92%	97%
Photon (100GeV)	95%	82%	95%	>99%
Pion (2 GeV)	78%	59%	75%	71%
Pion (5 GeV)	81%	70%	79%	80%
Pion (10GeV)	84%	80%	83%	85%
Pion (20GeV)	85%	87%	88%	91%

- Typical electron cluster energy resolution ~ 21%/sqrt(E)
- Typical pion cluster energy resolution ~70%/sqrt(E)
- All numbers are for one main cluster (no other fragments are included)

Hit efficiency: single particle

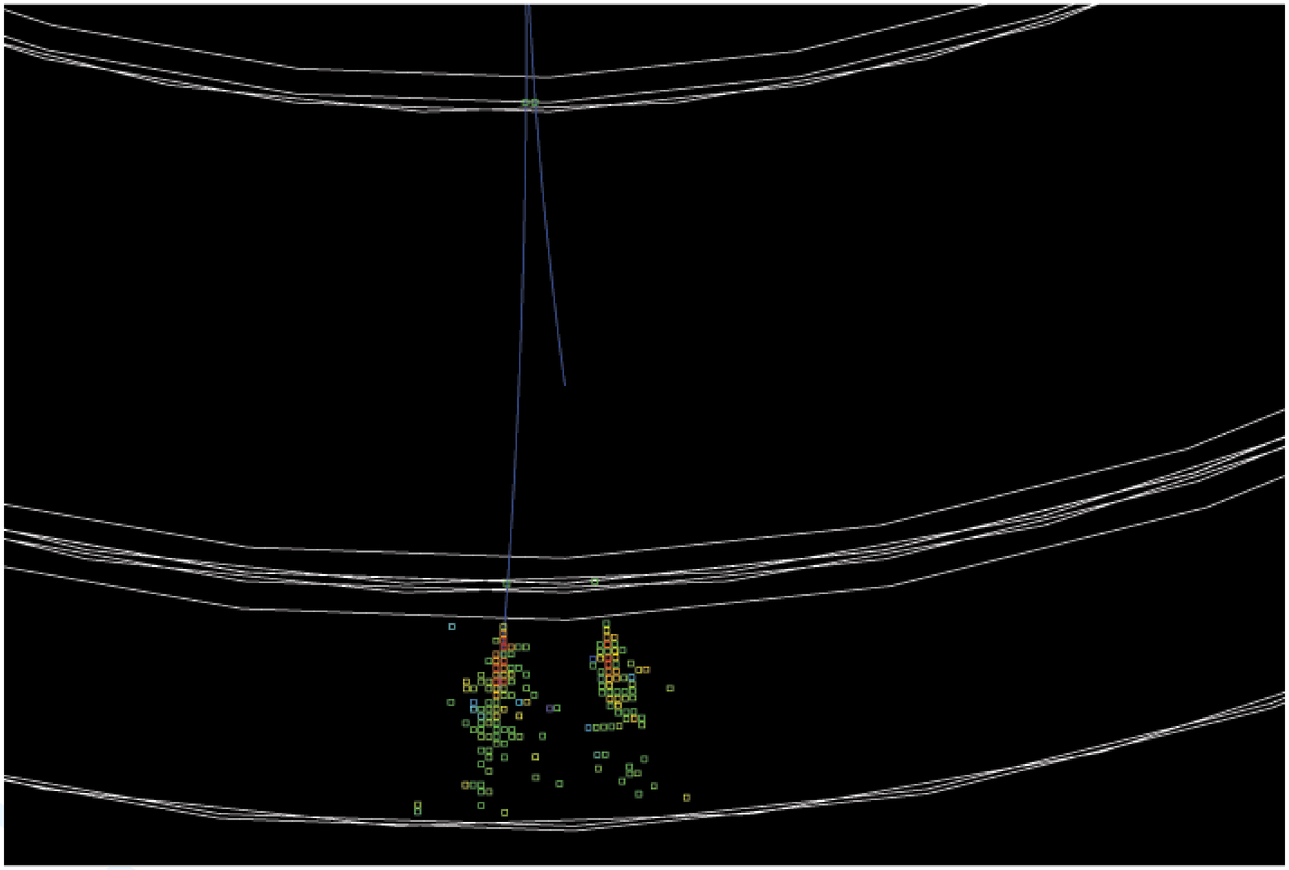
Very preliminary: from sidaug05_scinthcal (scint. DHCal, 7.4 – 12.mm barrel cell size)

Particle	ECal hit efficiency	HCal hit efficiency	Overall hit efficiency
Pion (2 GeV)	78%	30%	62%
Pion (5 GeV)	82%	47%	66%
Pion (10GeV)	86%	58%	72%
Pion (20GeV)	88%	71%	78%



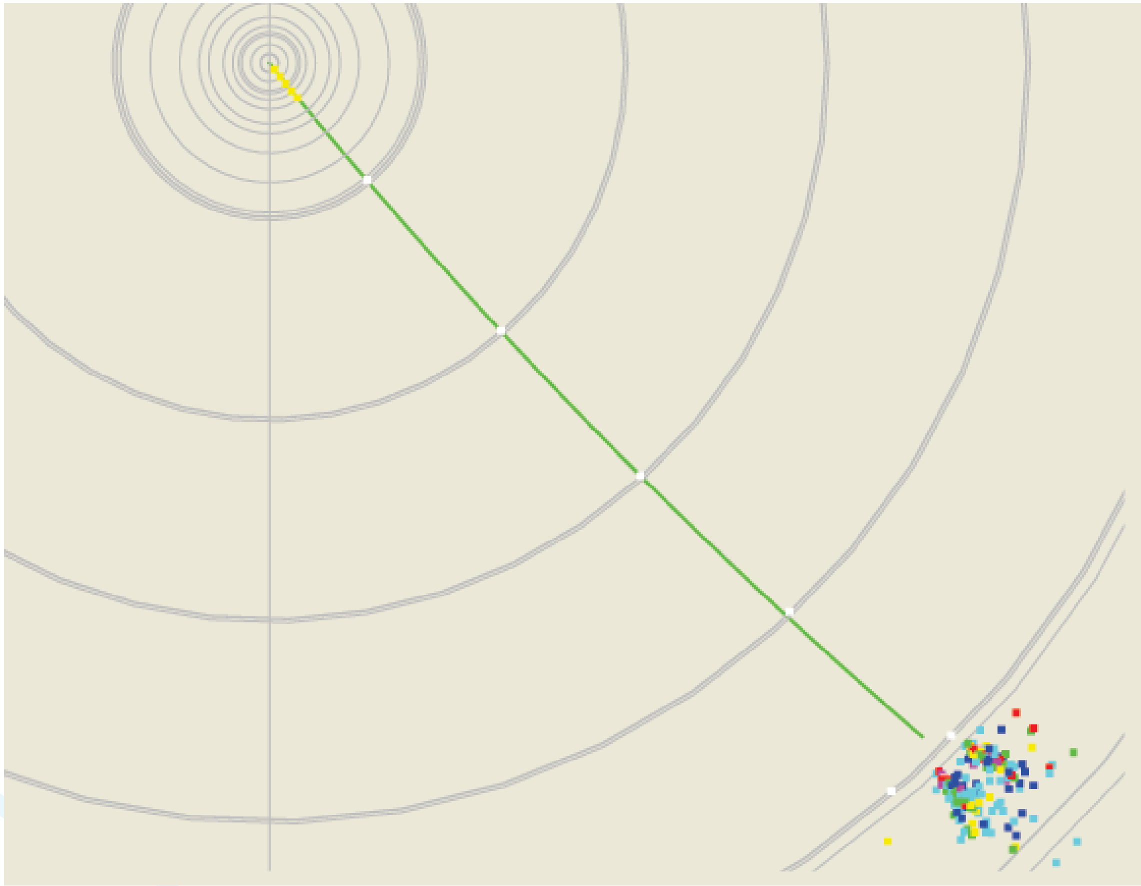
- All hadron calorimeter hits need to pass through a 0.3MeV cut

Shower separation...example



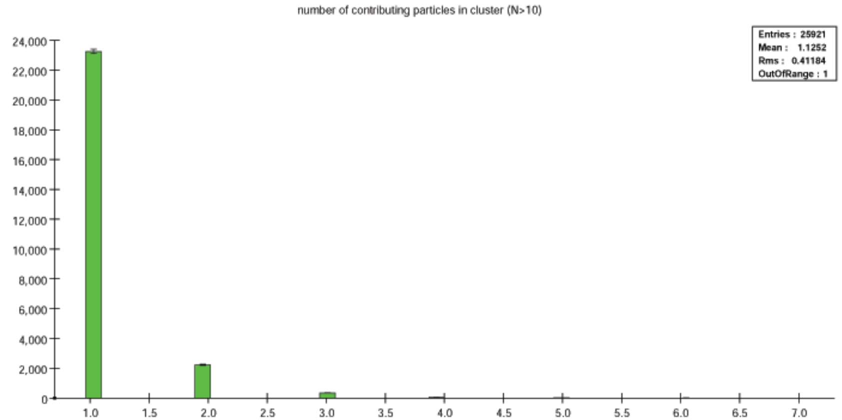
- 10 GeV photon converts to e+e- pair
- The two showers can be clearly separated

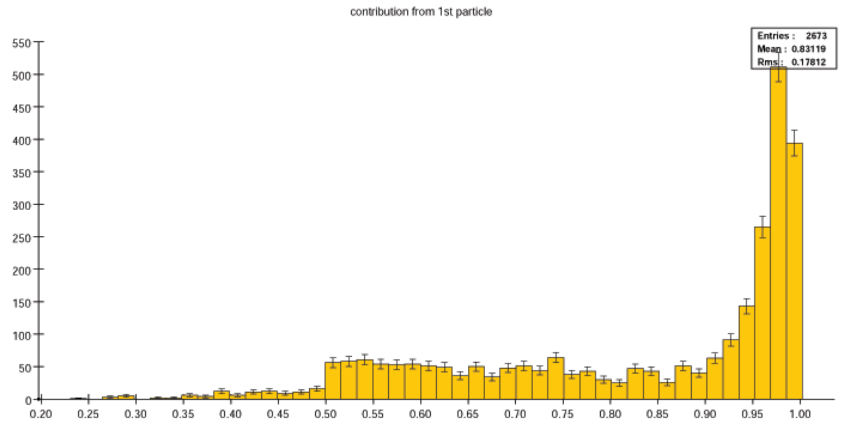
Shower separation...example

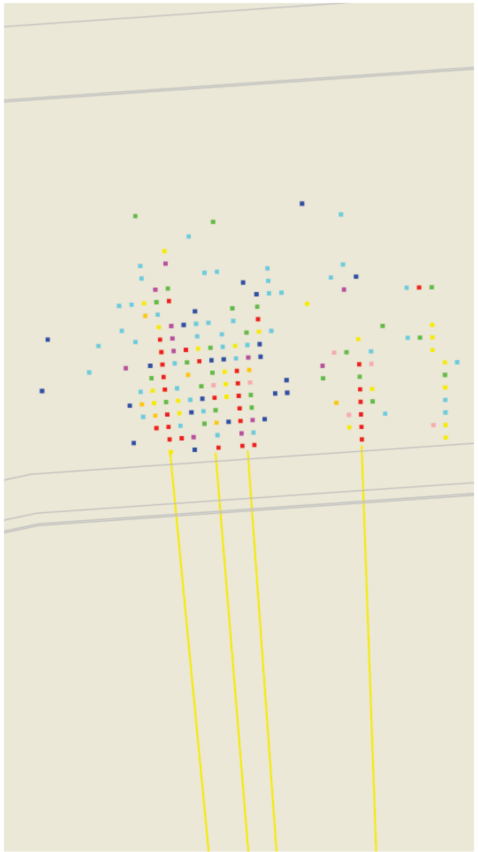


- 10 GeV e^- kick out a 5 GeV photon in tracker
- With appropriate density cut, the algorithm can separate the two showers
- Old picture from SDjan03 simulation

Z pole uds events: cluster purity

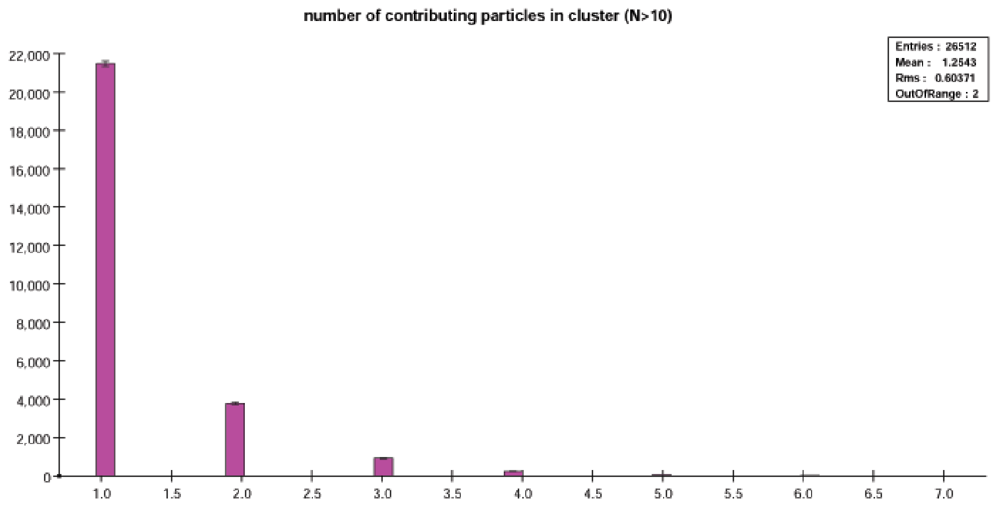


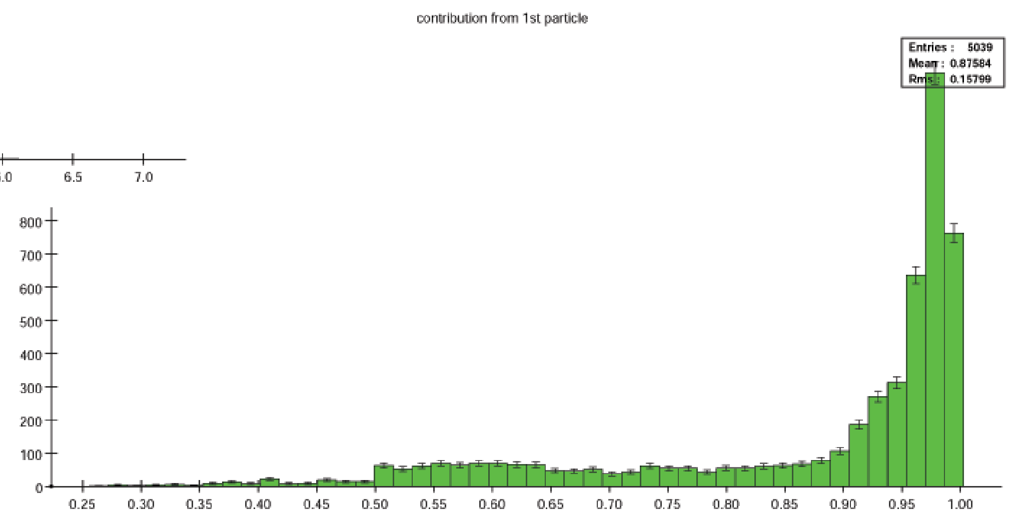




- Most of the clusters (89.7%) are pure (only one particle contributes)
- For the rest 10.3% clusters
  - 55% are almost pure (more than 90% hits are from one particle)
  - The rest clusters contain merged showers, part of them are 'trouble makers'
- On average, 1.2 merged shower clusters/Z pole event

Preliminary: sidaug05_scinthcal





- 81.0% of the clusters are pure (only one particle contributes)
- For the rest 19.0% clusters
  - 68% are almost pure (more than 90% hits are from one particle)
- On average, 1.6 merged shower clusters/Z pole event
- Projective geometry, 7.4 – 12.mm barrel cell size

Thing to work on…

- Better understanding of clustering
  - Hit efficiency -> new calibration(?)
  - Optimize density cut according to PFA performance
  - What to do with merged showers?
- Cluster ID
  - e/gamma, hadron, fragment
  - Clusters that contain multiple showers