Two Density-based Clustering Algorithms

L. Xia (ANL)

V. Zutshi (NIU)

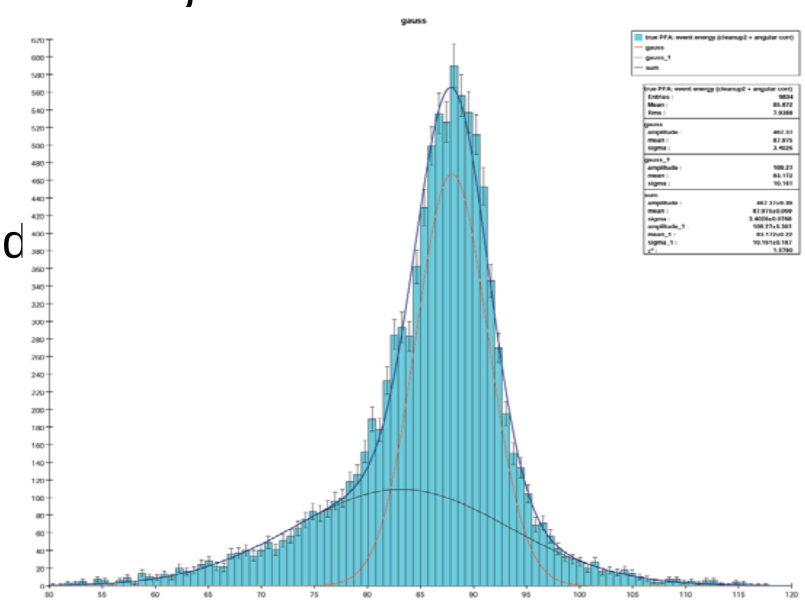
General Comments

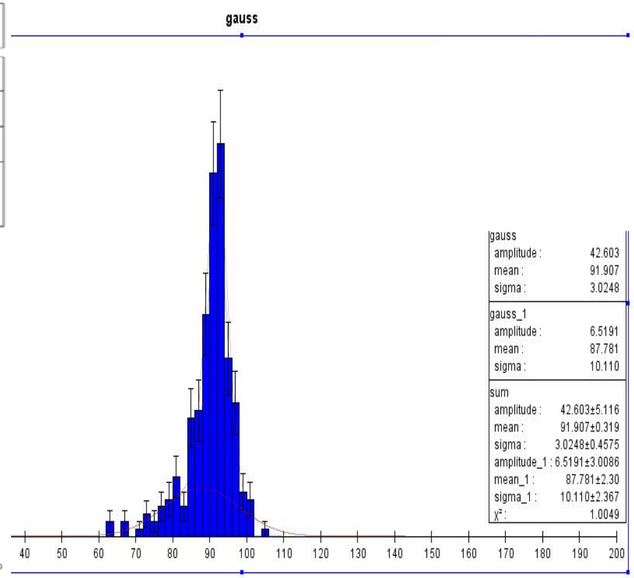
- Both are calorimeter first approaches

  clustering → track match → fragment….
- SiD geometry

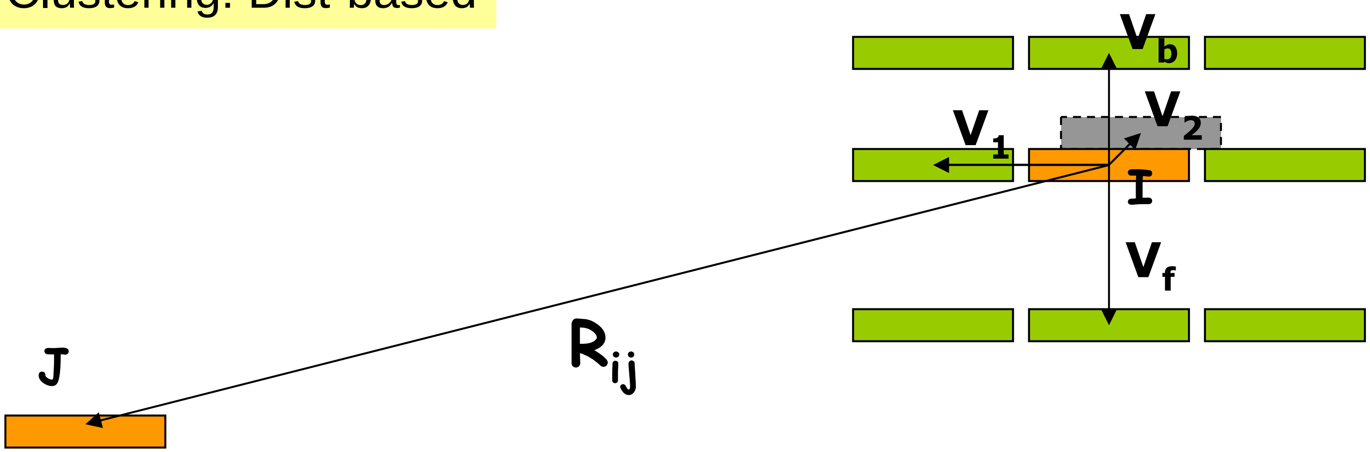
  Si-W ECAL, RPC or Scintillator HCAL

'Cheating' involved
in some steps





Clustering: Dist-based

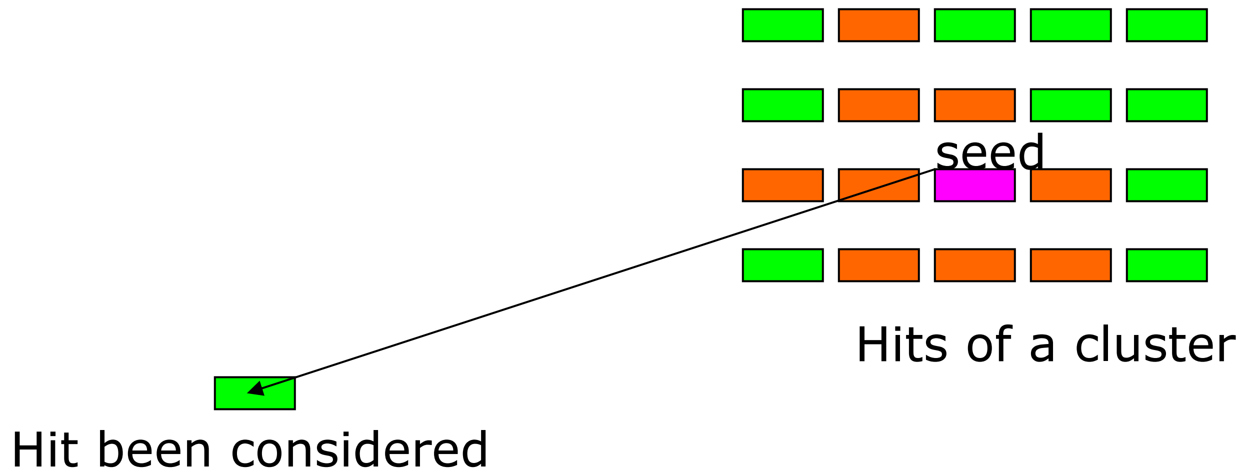


$$D_i = \sum_{\{j \neq i\}} \left(e^{-(|\vec{V}_1 \bullet \vec{R}_{ij}|/|\vec{V}_1|^2)} \times e^{-(|\vec{V}_2 \bullet \vec{R}_{ij}|/|\vec{V}_2|^2)} \times e^{-(|\vec{V}_3 \bullet \vec{R}_{ij}|/|\vec{V}_3|^2)} \right)$$

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

- Hit density reflects the closeness from one hit i to a group of hits {j}
  - {j} = {all calorimeter hits} to decide if hit i should be a cluster seed
  - {j} = {all hits in a cluster} to decide if hit i should be attached to this cluster
- Consider cell density variation by normalizing distance to local cell separation
  - Density calculation takes care of the detector geometry
  - Clustering algorithm then treat all calorimeter hits in the same way

Clustering: Dist-based



- Find a cluster seed: hit with highest density among remaining hits
- Attach nearby hits to a seed to form a small cluster
- Attach additional hits based on density calculation
  - i = hit been considered, {j} = {existing hits in this cluster}
  - EM hits, $D_i > 0.01$
  - HAD hits, $D_i > 0.001$
  - Grow the cluster until no hits can be attached to it
- Find next cluster seed, until run out of hits

Clustering: Grad-based

Define the <u>density</u> neighborhood of a cell i

Neighborhood → ± (nly,nz,nphi) window centered on the cell

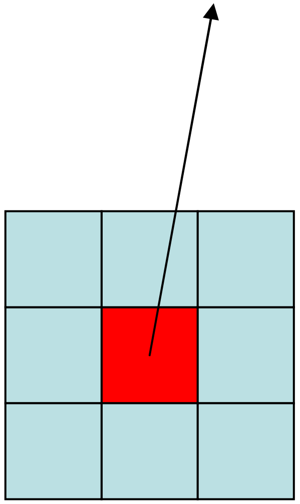
For each cell i calculate the 'density' D_i over its neighborhood

For each i calculate the 'gradient' $(D_j – D_i)/d_{ij}$ where j is in the <u>clustering</u> neighborhood

Find max $[\text{gradient}]_{ij}$

The magnitude and sign of $\max[\text{gradient}]_i$ determines what happens to cell i

Cell i



In this presentation simply occupancy or no. of cells above threshold is used

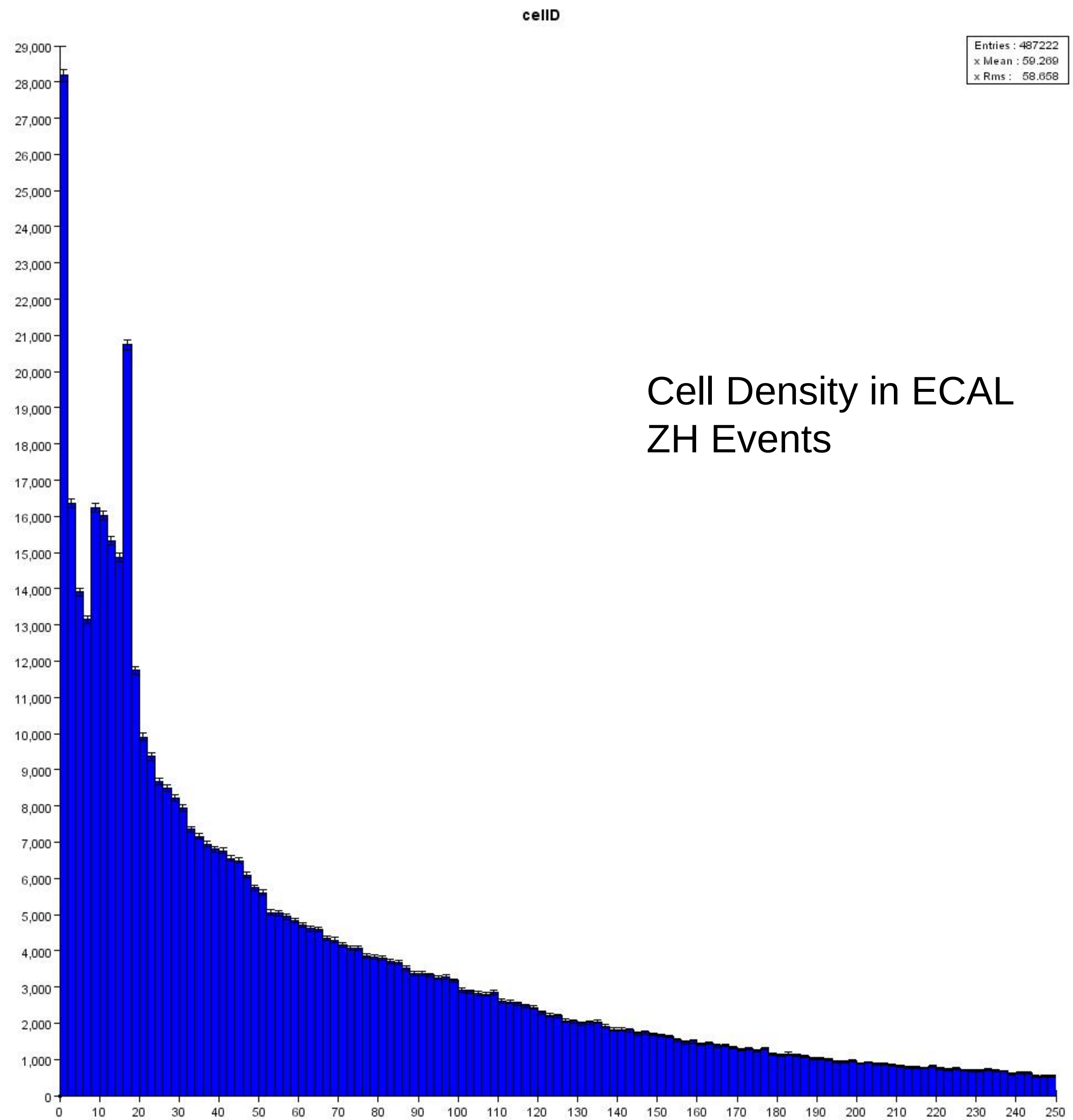
Clustering: Grad-based

- if $\max[\text{gradient}]_{ij} < 0$

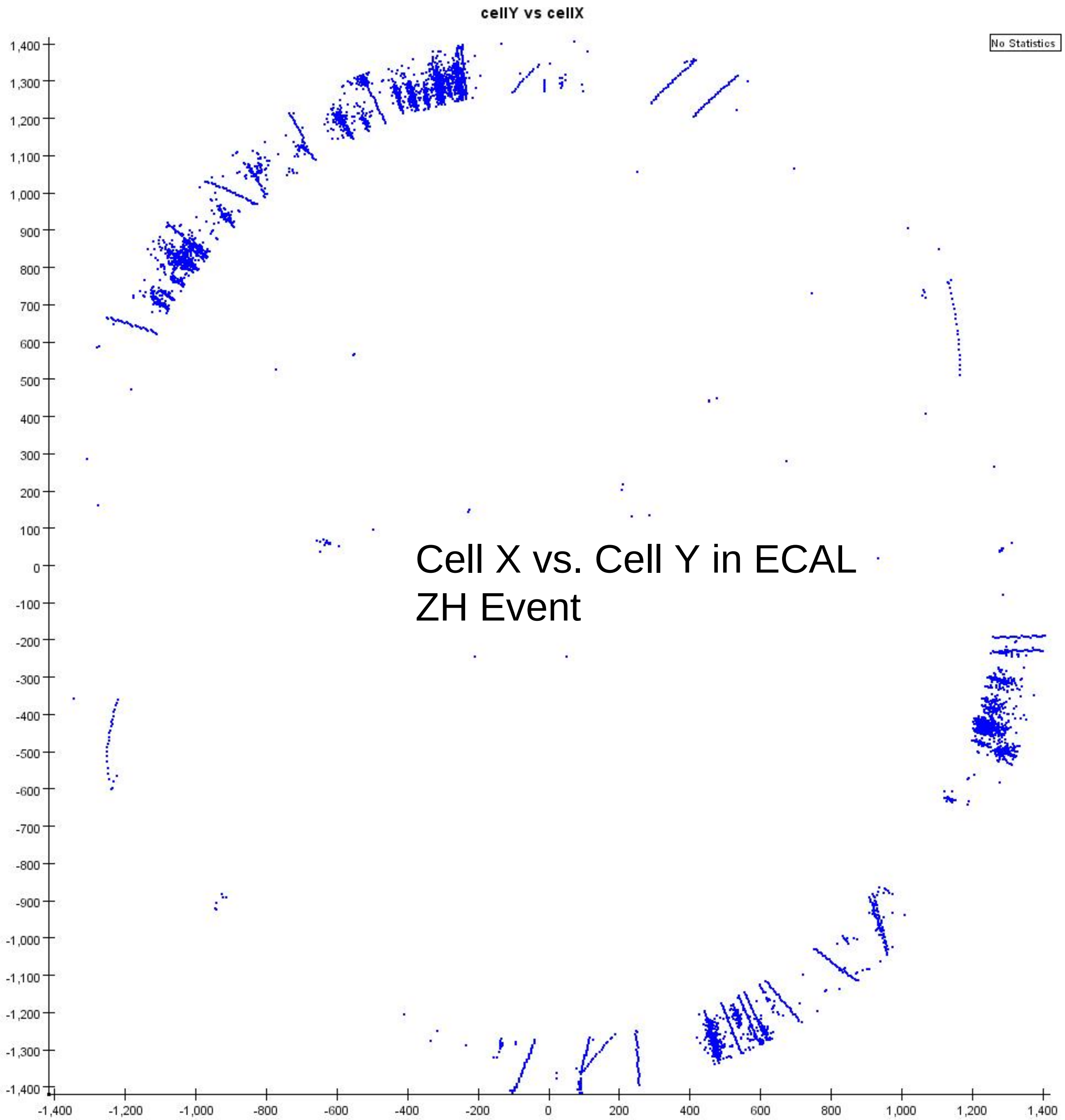
  i becomes the 'root' and starts a branch of its cluster
- if $\max[\text{gradient}]_{ij} > 0$

  j is the parent of i and attaches i to its branch
- if $\max[\text{gradient}]_{ij} == 0$

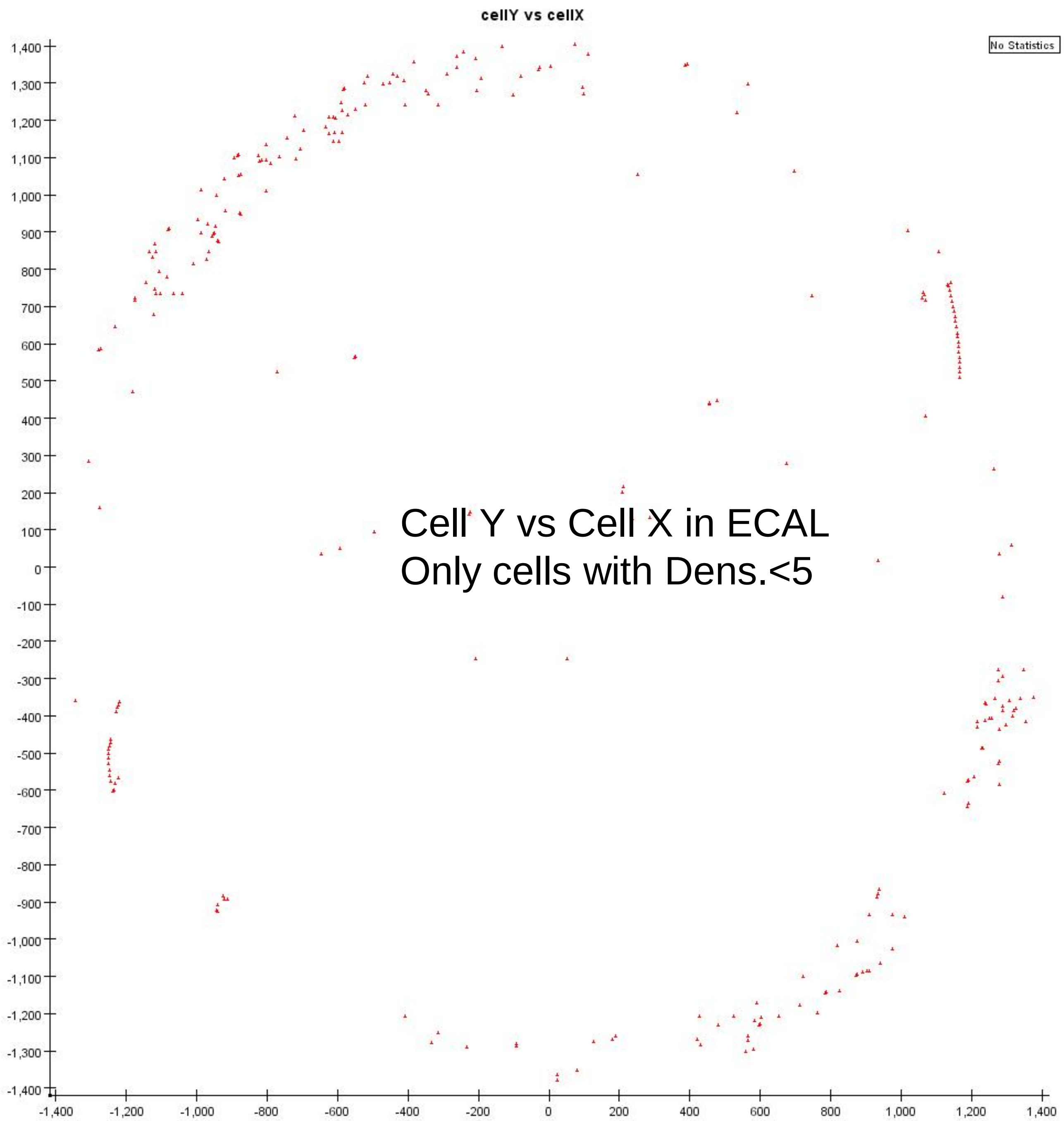
  attach i to the nearest j

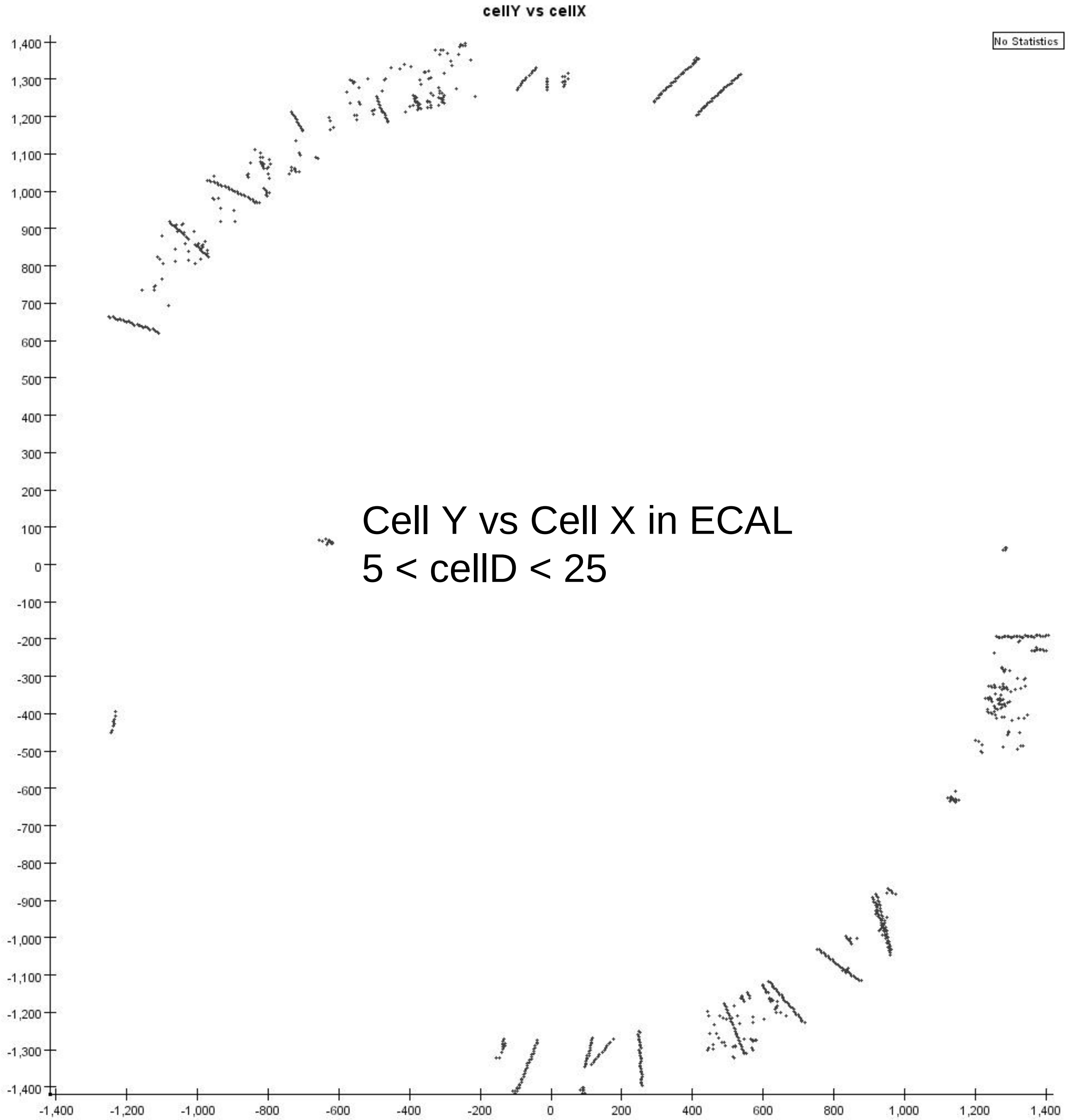


Cell Density in ECAL
ZH Events

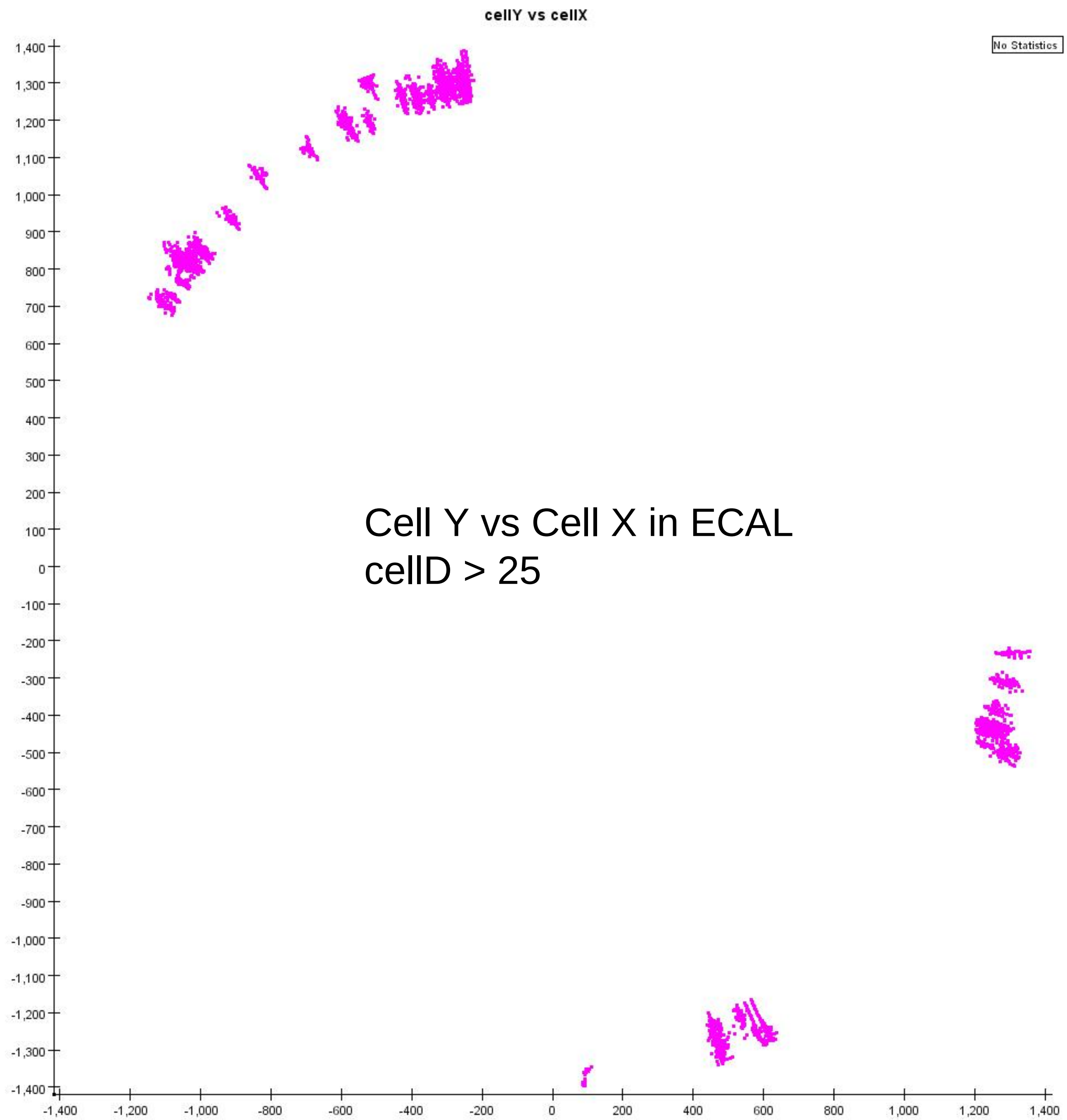


Cell X vs. Cell Y in ECAL
ZH Event


cellY vs cellX
No Statistics
Cell Y vs Cell X in ECAL
Only cells with Dens.<5
1,400
1,300
1,200
1,100
1,000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1,000
-1,100
-1,200
-1,300
-1,400
-1,400 -1,200 -1,000 -800 -600 -400 -200 0 200 400 600 800 1,000 1,200 1,400



Cell Y vs Cell X in ECAL
5 < cellD < 25


cellY vs cellX
No Statistics
Cell Y vs Cell X in ECAL
cellD > 25

Dist-based

clustering 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 fragments are included)

Grad-based

Z-pole Events

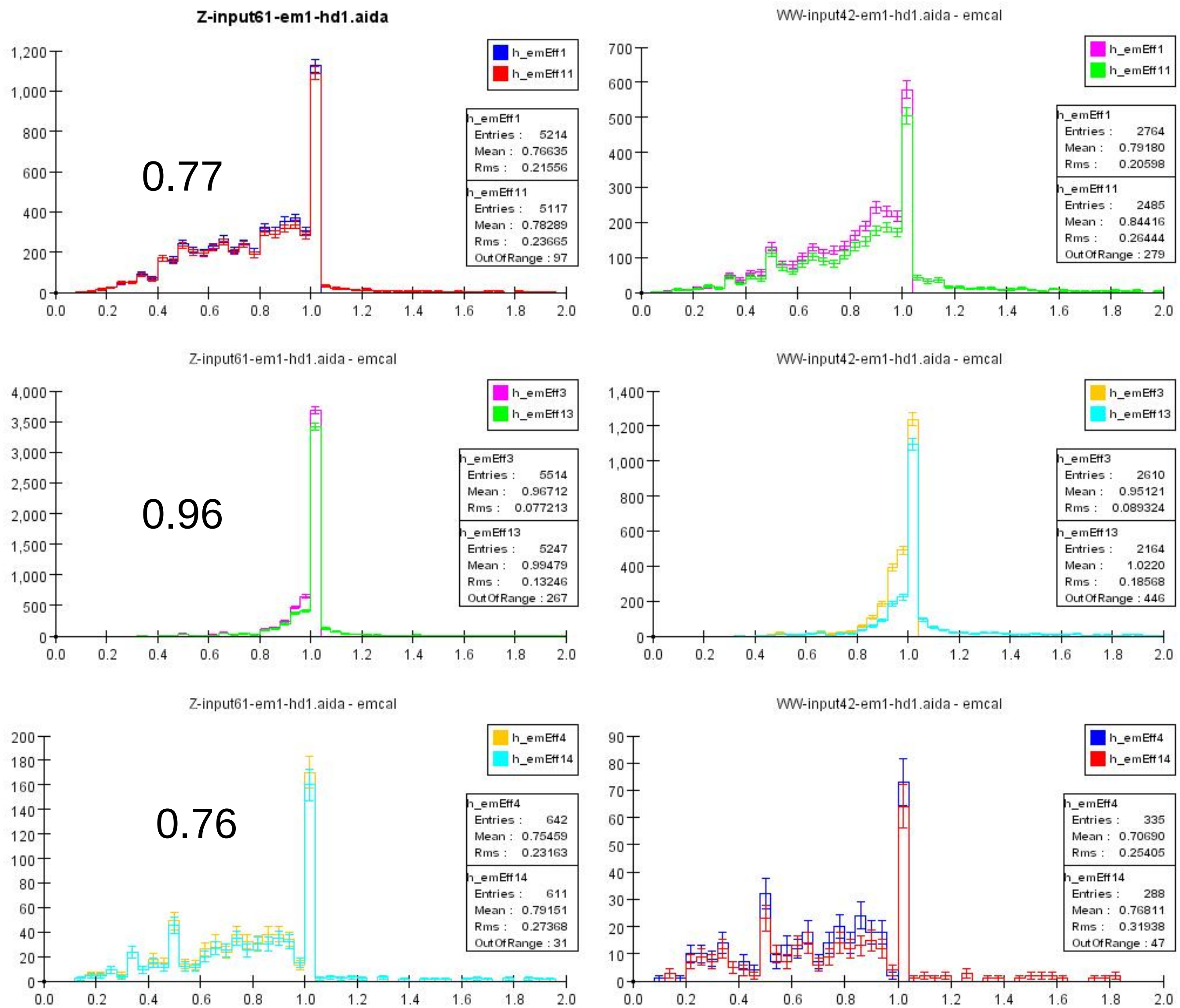
WW Events

Charged hadrons

Photons

Neutral hadrons

ECAL



Cluster Efficiency in Z-pole events

Grad-based

Z-pole Events

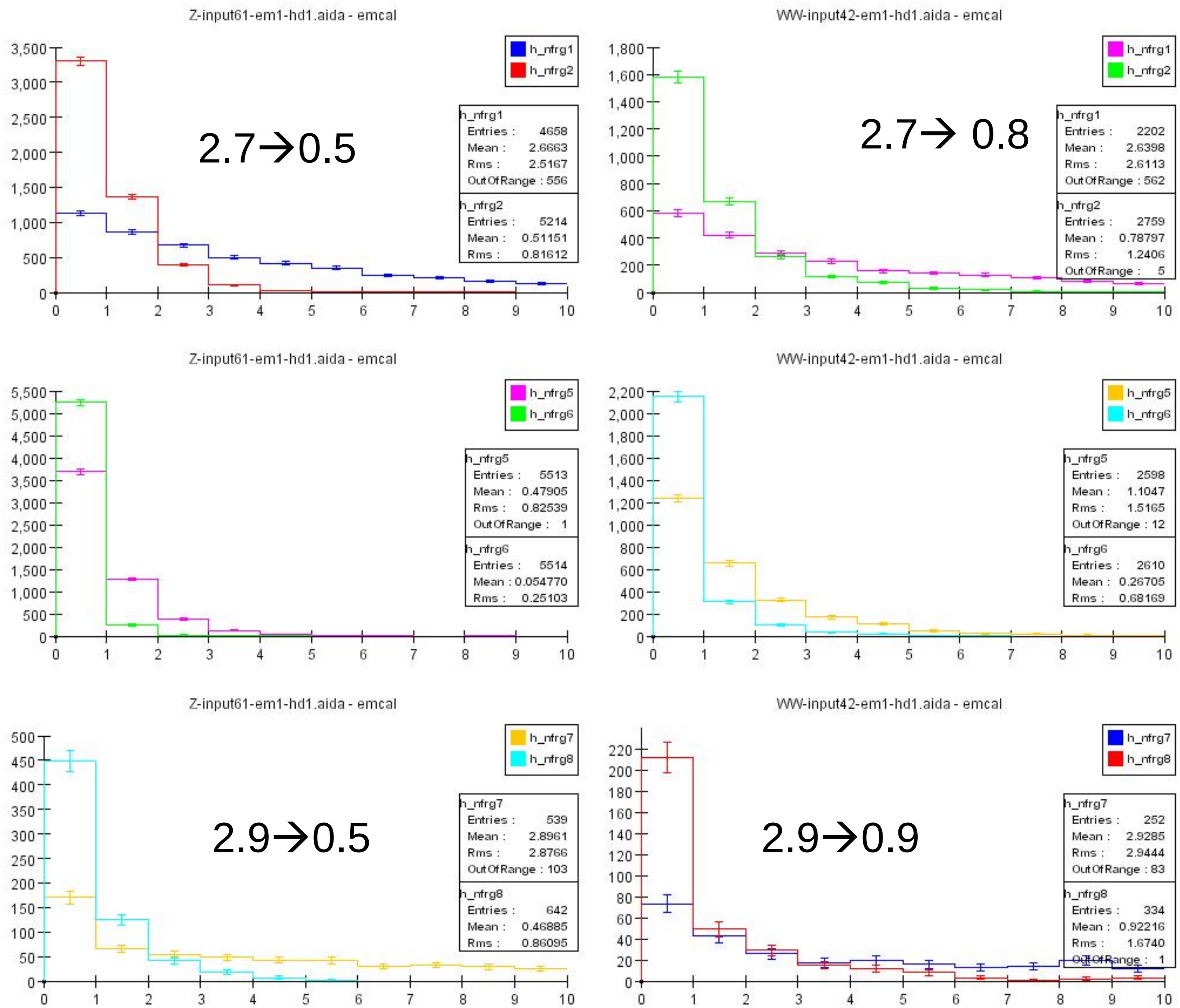
WW Events

Charged hadrons

Photons

Neutral hadrons

ECAL



No. of fragments w/ and w/o cut on fragment size

Grad-based

Z-pole Events

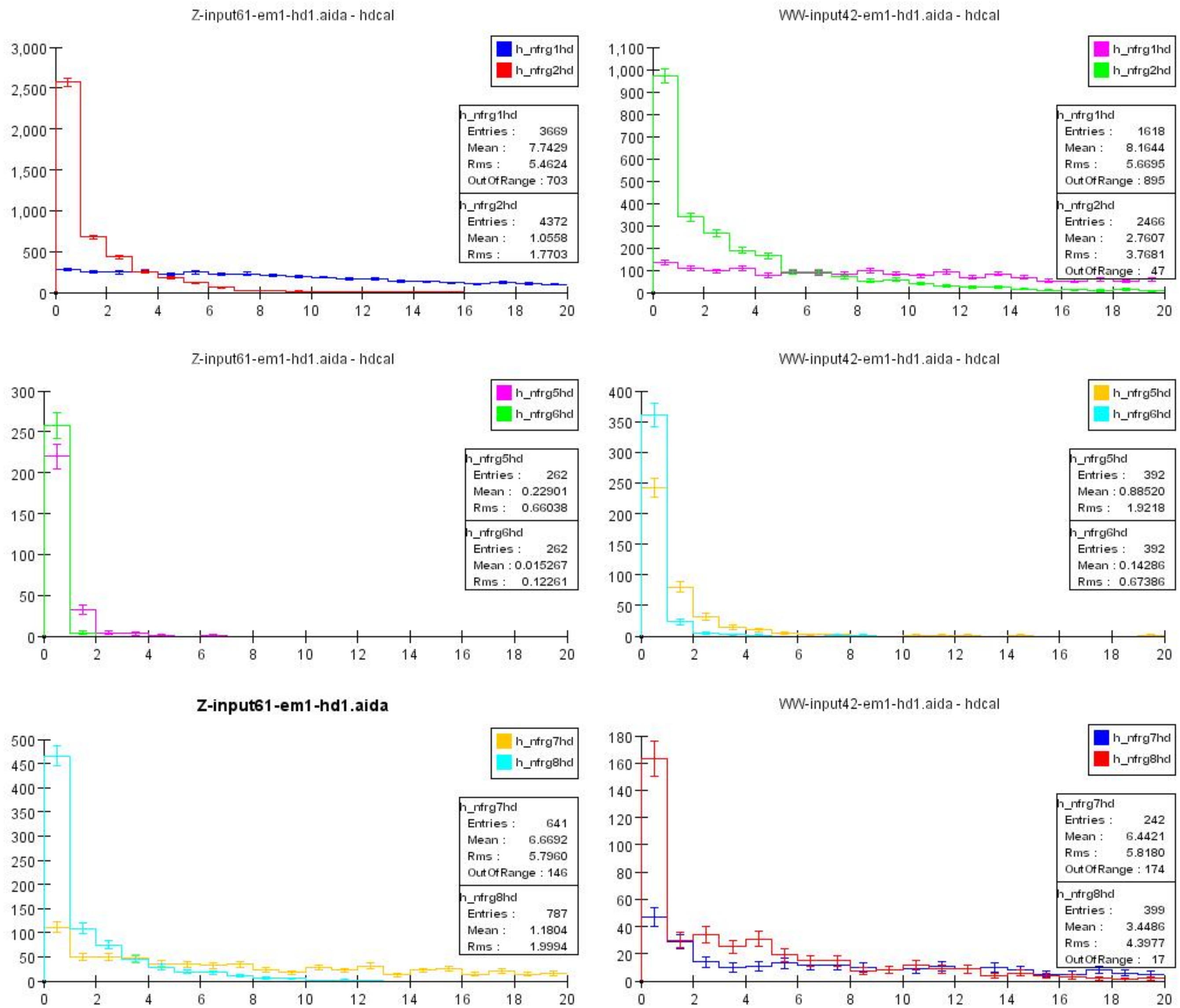
WW Events

Charged hadrons

Photons

Neutral hadrons

HCAL



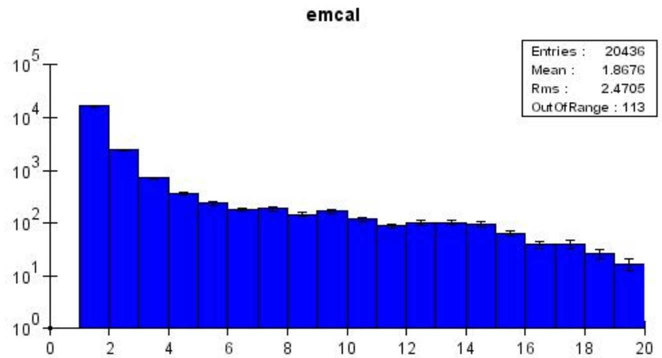
No. of fragments w/ and w/o cut on fragment size

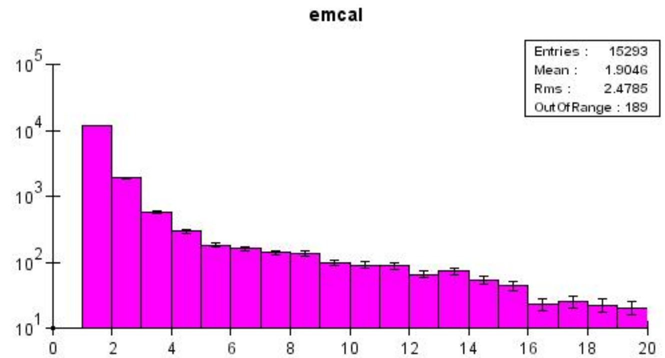
Grad-based

Z-pole Events

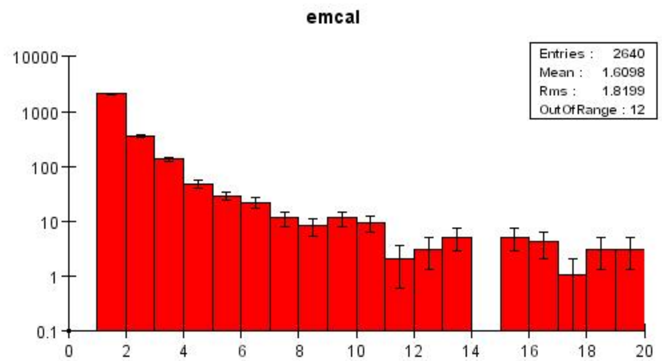
WW Events

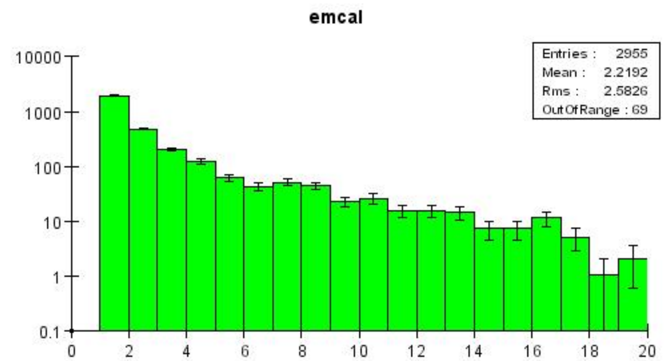
Charged hadrons





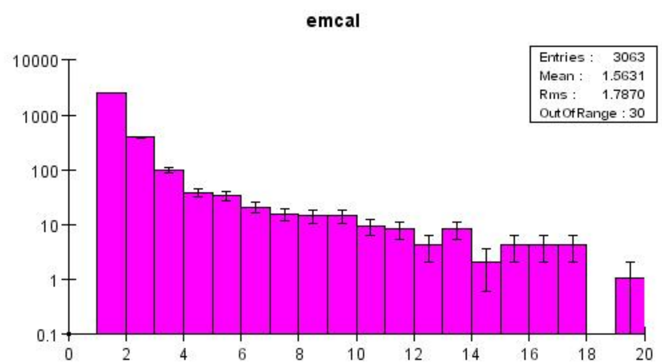
Photons

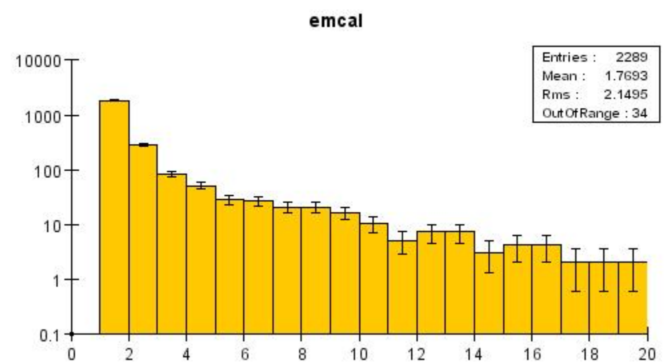




ECAL

Neutral hadrons





Fragment size

Grad-based

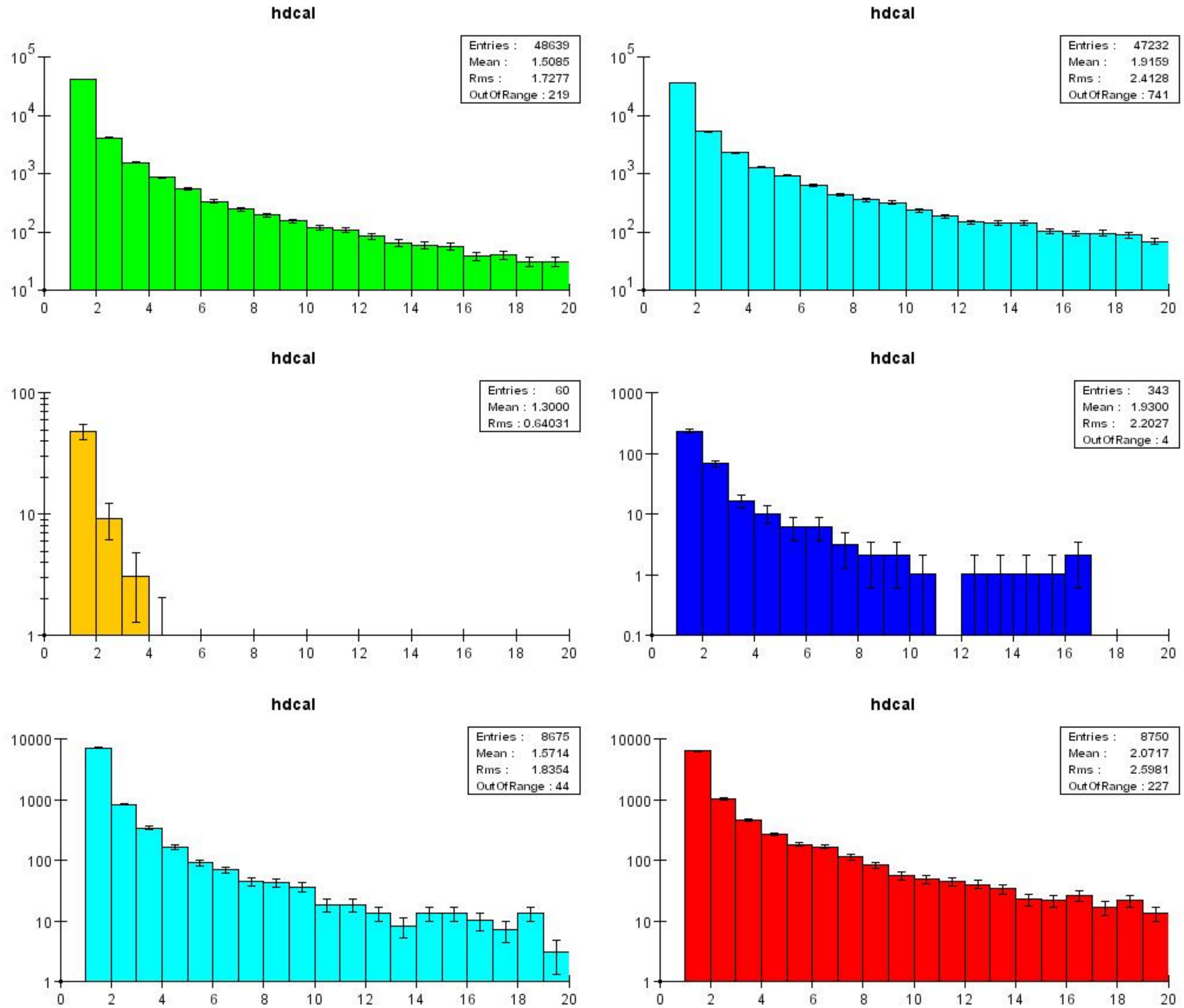
Z-pole Events

WW Events

Charged hadrons

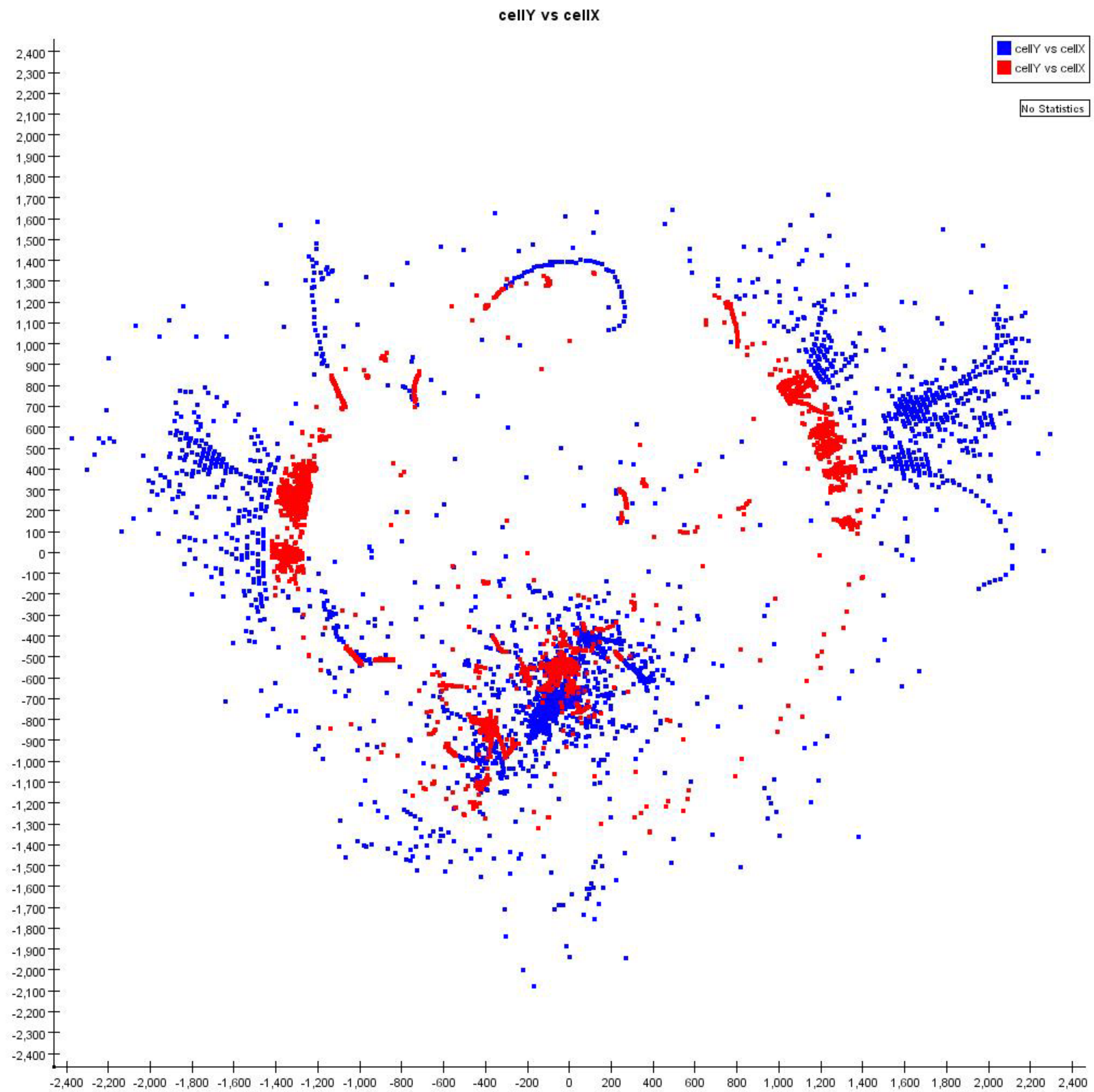
Photons

Neutral hadrons

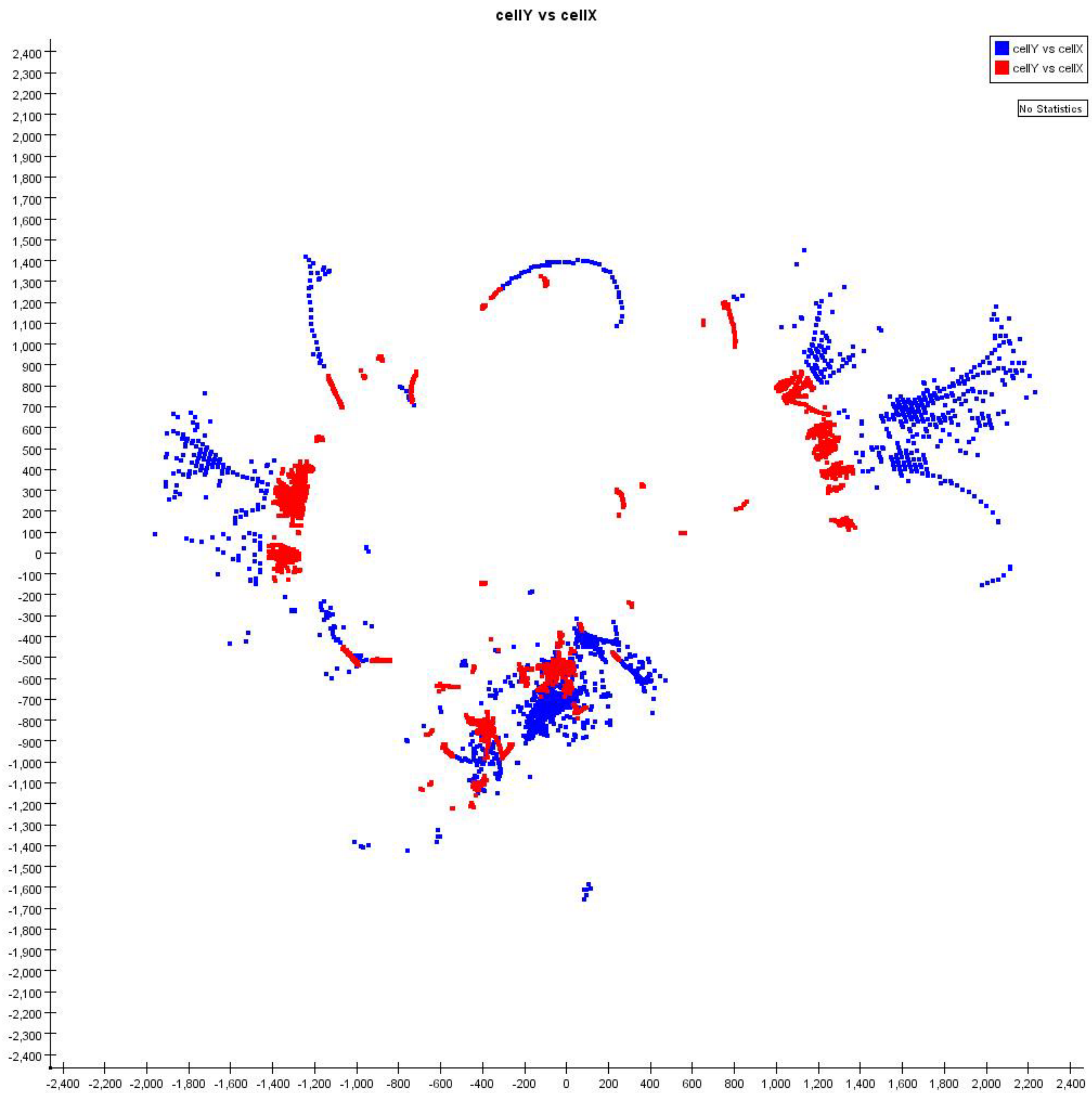


HCAL

Fragment size



cellY vs cellX
cellY vs cellX
cellY vs cellX
No Statistics
2,400
2,300
2,200
2,100
2,000
1,900
1,800
1,700
1,600
1,500
1,400
1,300
1,200
1,100
1,000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1,000
-1,100
-1,200
-1,300
-1,400
-1,500
-1,600
-1,700
-1,800
-1,900
-2,000
-2,100
-2,200
-2,300
-2,400
-2,400 -2,200 -2,000 -1,800 -1,600 -1,400 -1,200 -1,000 -800 -600 -400 -200 0 200 400 600 800 1,000 1,200 1,400 1,600 1,800 2,000 2,200 2,400

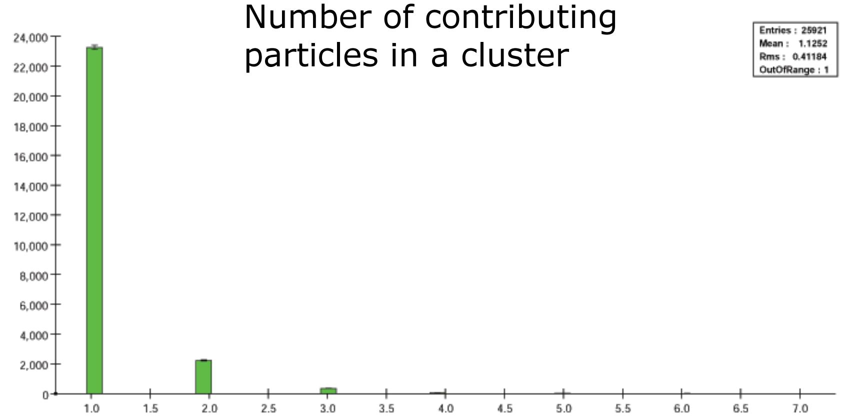


cellY vs cellX
cellY vs cellX
cellY vs cellX
No Statistics
2,400
2,300
2,200
2,100
2,000
1,900
1,800
1,700
1,600
1,500
1,400
1,300
1,200
1,100
1,000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1,000
-1,100
-1,200
-1,300
-1,400
-1,500
-1,600
-1,700
-1,800
-1,900
-2,000
-2,100
-2,200
-2,300
-2,400
-2,400 -2,200 -2,000 -1,800 -1,600 -1,400 -1,200 -1,000 -800 -600 -400 -200 0 200 400 600 800 1,000 1,200 1,400 1,600 1,800 2,000 2,200 2,400

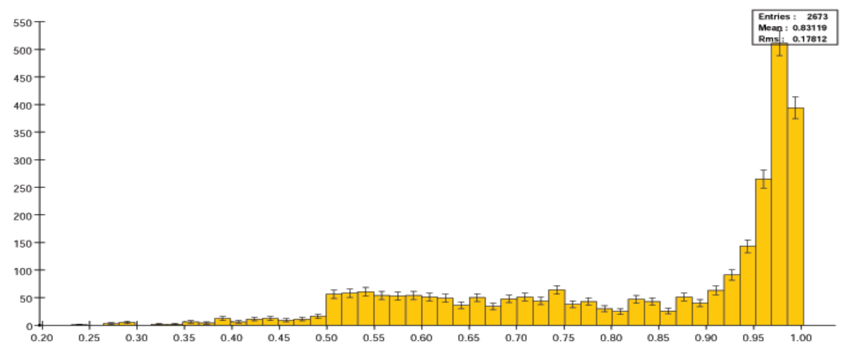
Dist-based

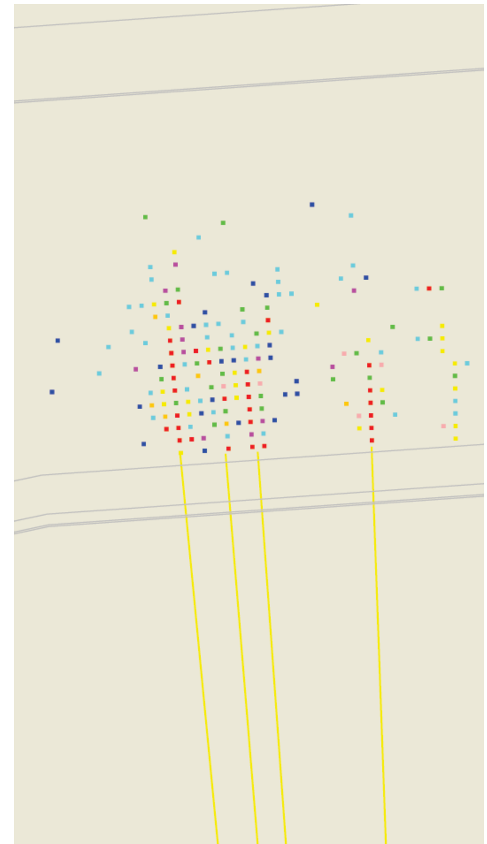
cluster purity : Z pole uds events

Number of contributing particles in a cluster



Fraction from largest contributor for clusters with multi-particles





- Most of the clusters (89.7%) are pure (only one particle contributes)
- For the remaining 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

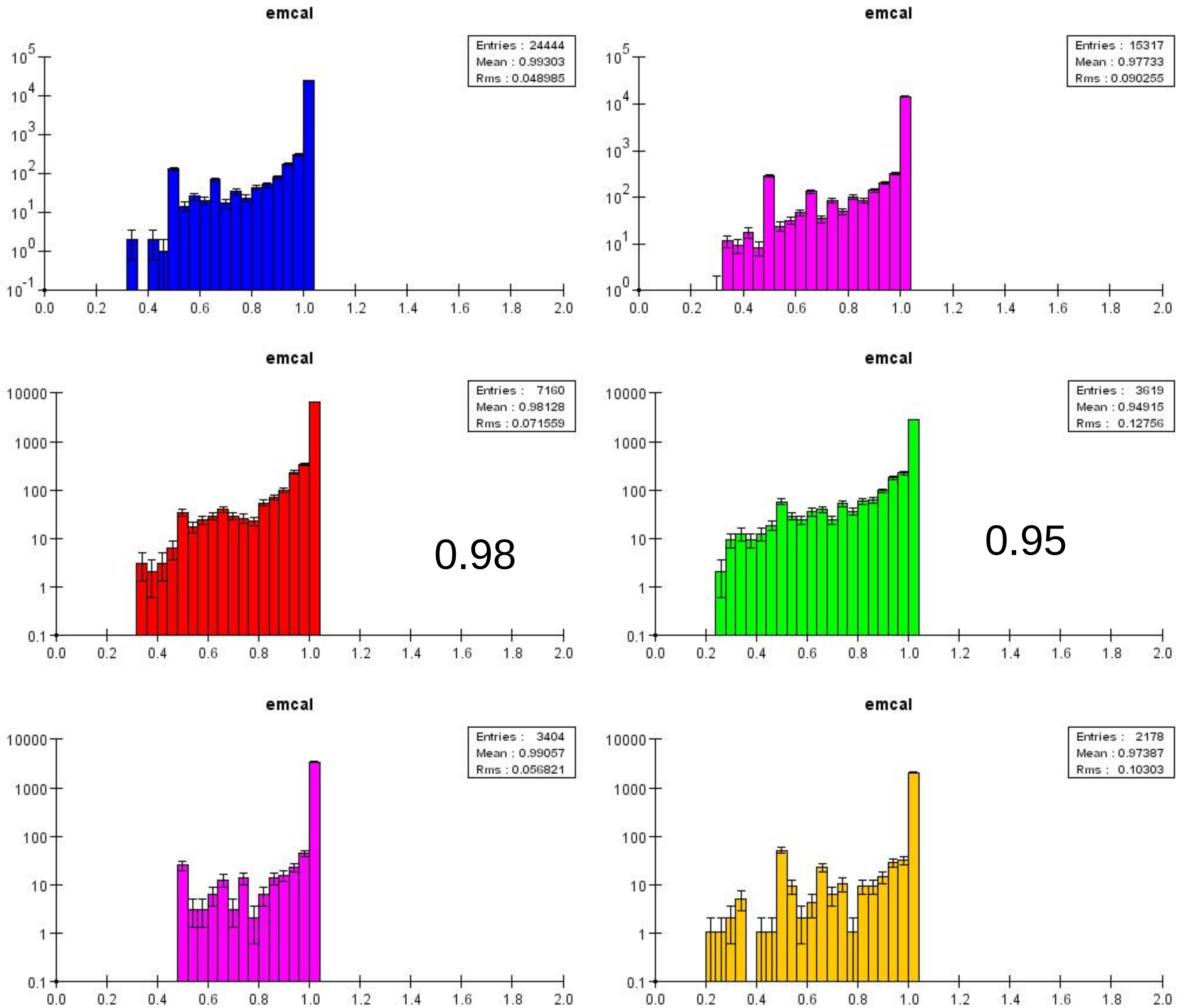
Grad-based

Z-pole Events | WW Events

Charged hadrons

Photons

Neutral hadrons



ECAL

purity

Grad-based

Z-pole Events

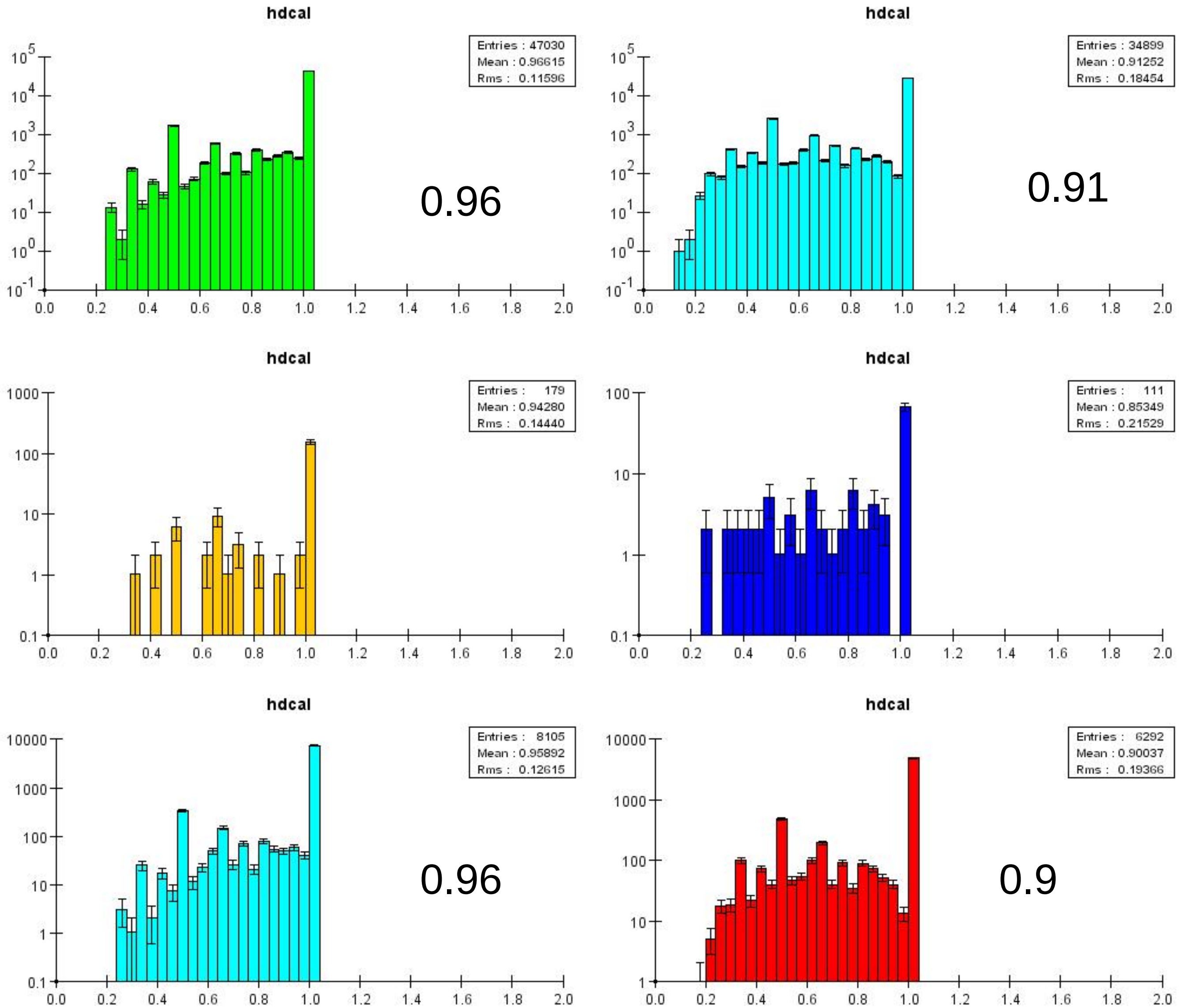
WW Events

Charged hadrons

Photons

Neutral hadrons

HCAL



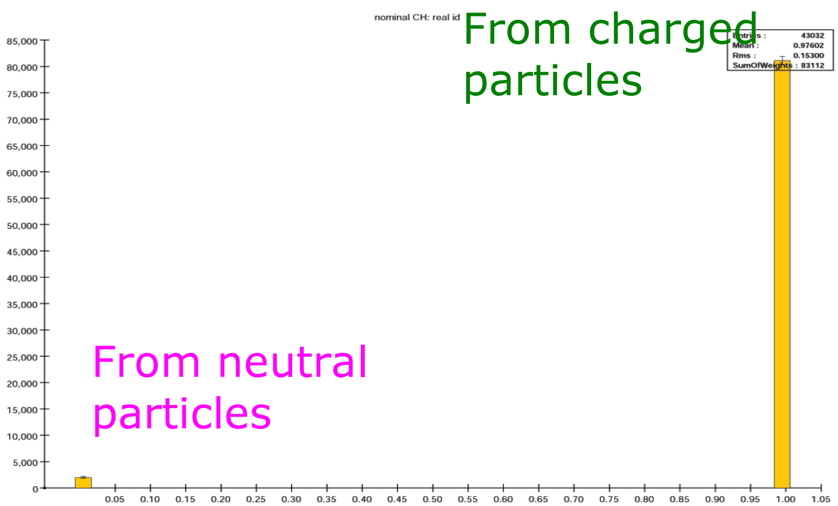
purity

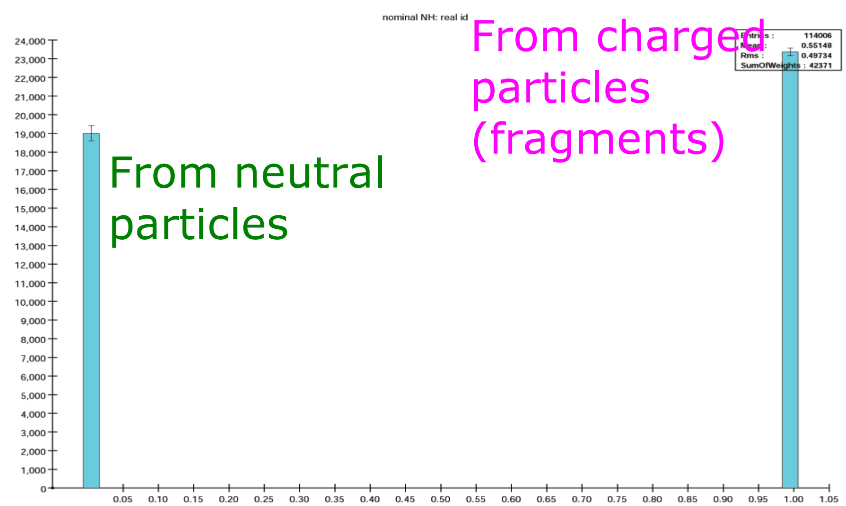
Dist-based

After track-cluster matching

Energy of matched clusters

Energy of clusters not matched to any track: neutral candidate

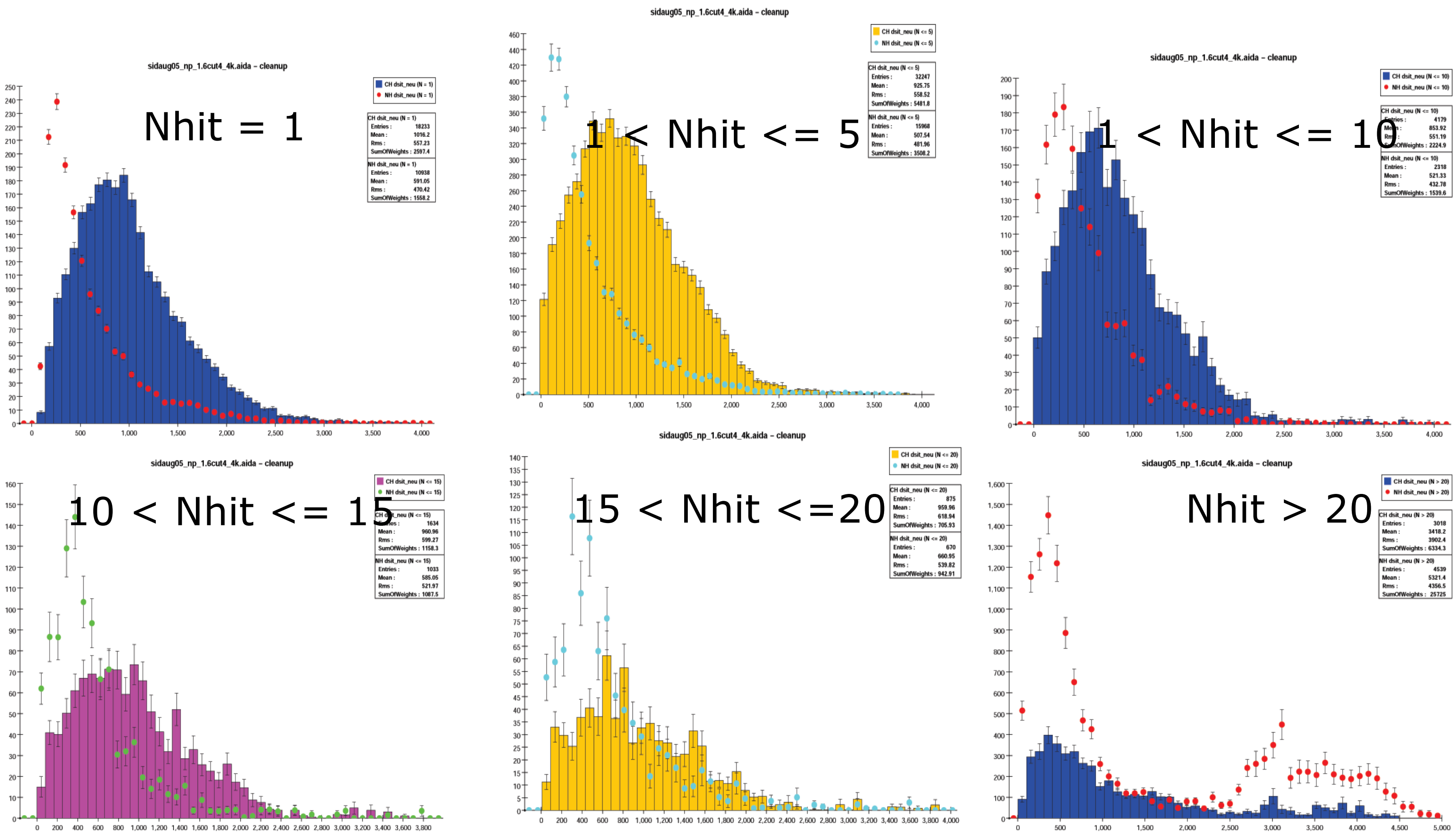




On average
~3% came from neutral

Energy from charged particles is more than real neutral
-- need to work on it!

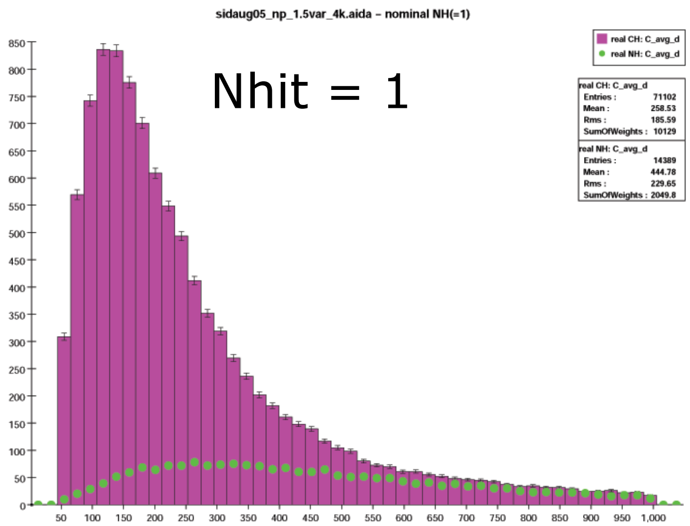
Fragment identification – variable2: distance to any neutrals

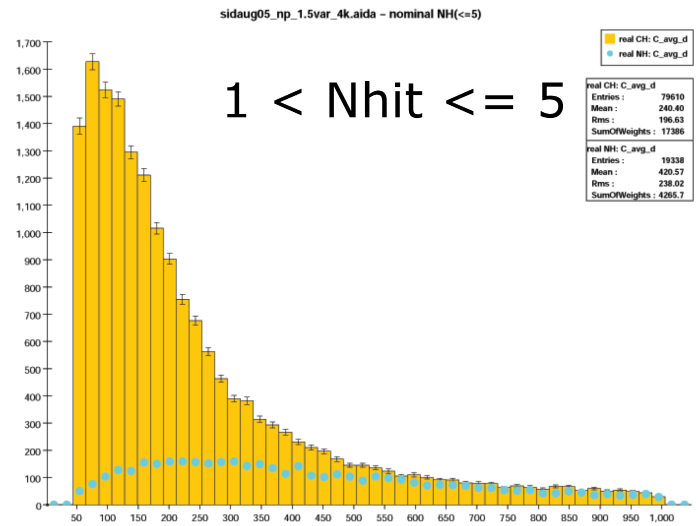


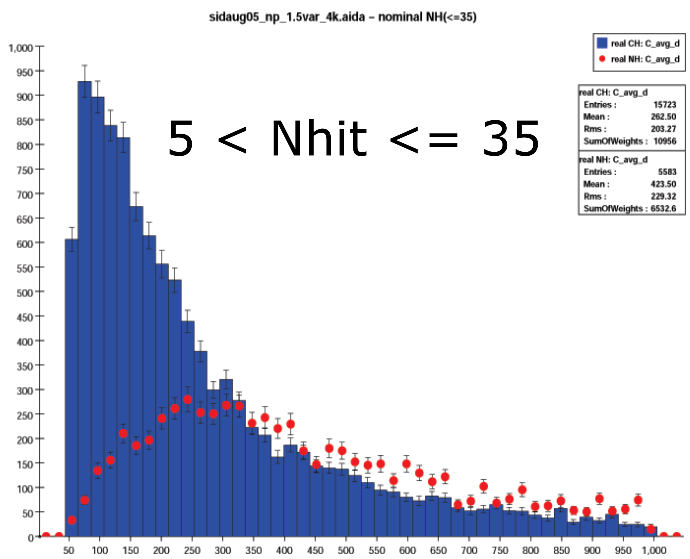
Neutral: any ‘big’ cluster not matched to any track

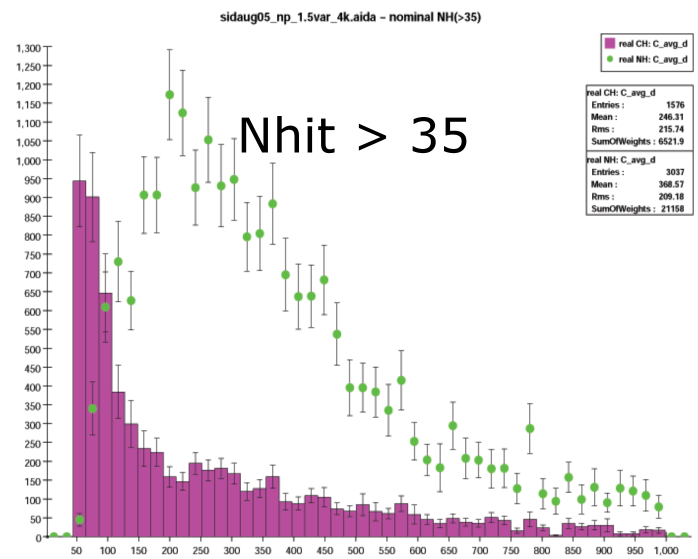
Fragment cluster (histogram) vs. real neutral cluster (dots)

Fragment identification - variable1: distance to any track



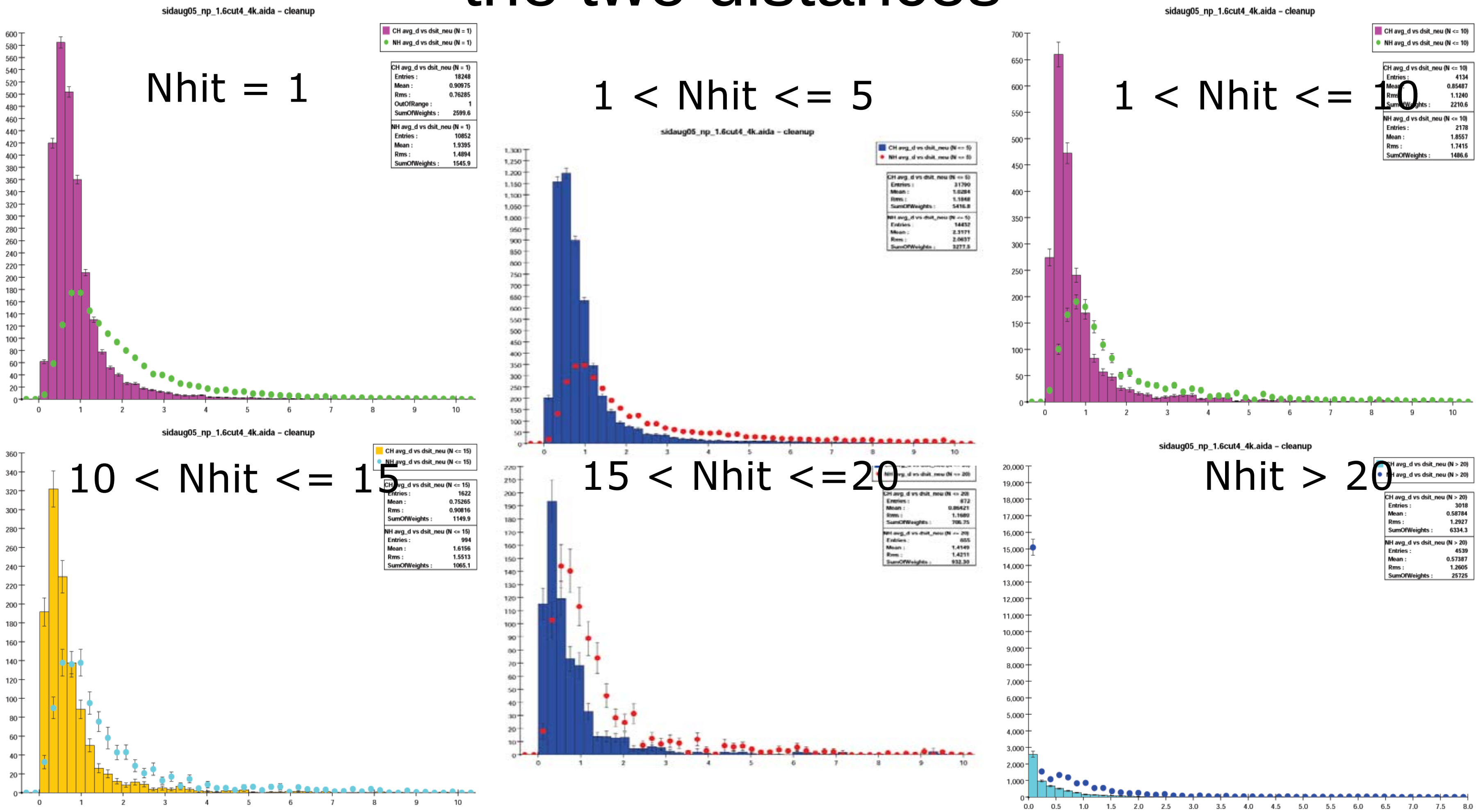






Dist-based

Fragment identification – variable3: ratio of the two distances

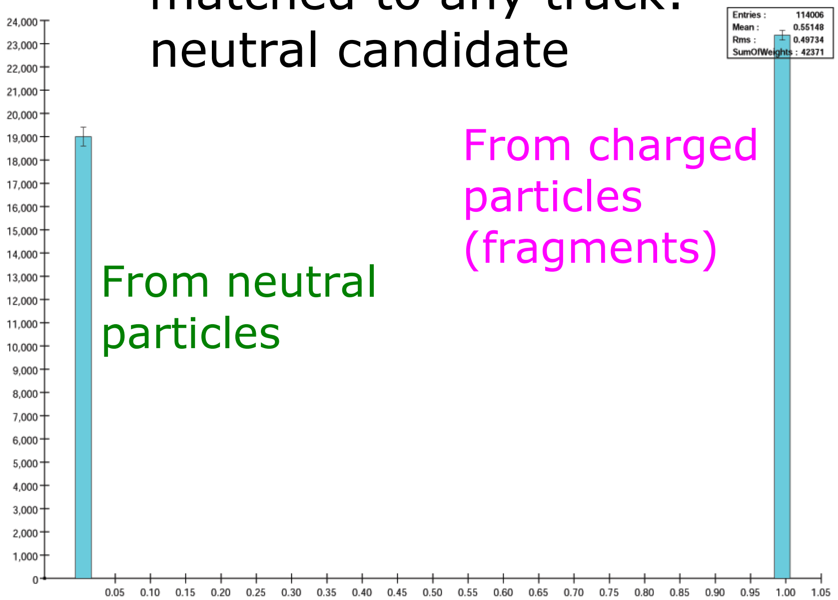


Fragment cluster (histogram) vs. real neutral cluster (dots)

Dist-based

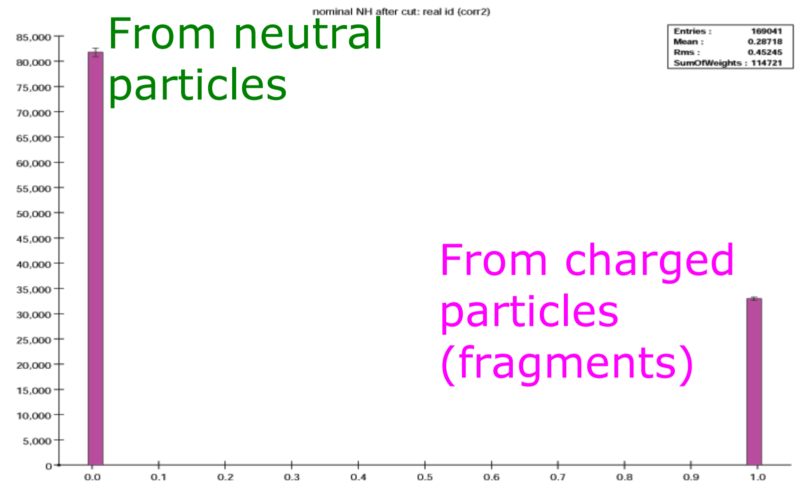
Fragment identification

Energy of clusters not matched to any track: neutral candidate



1 : 1.24

After removing identified fragments



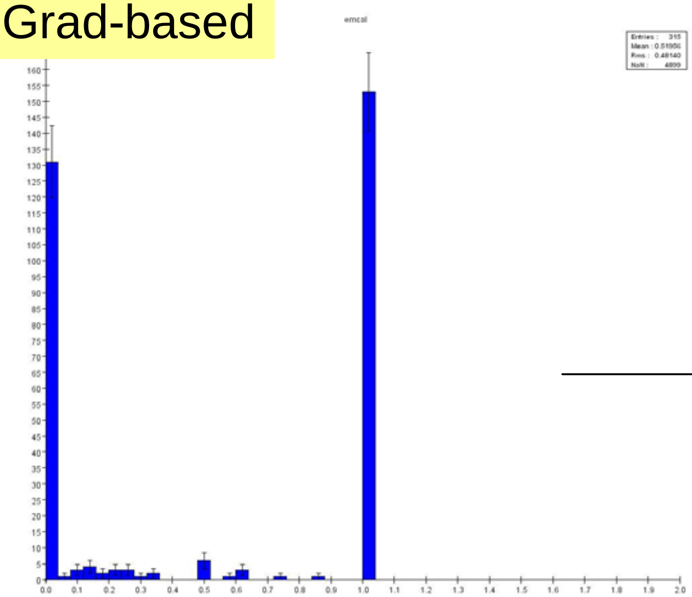
1 : 0.40

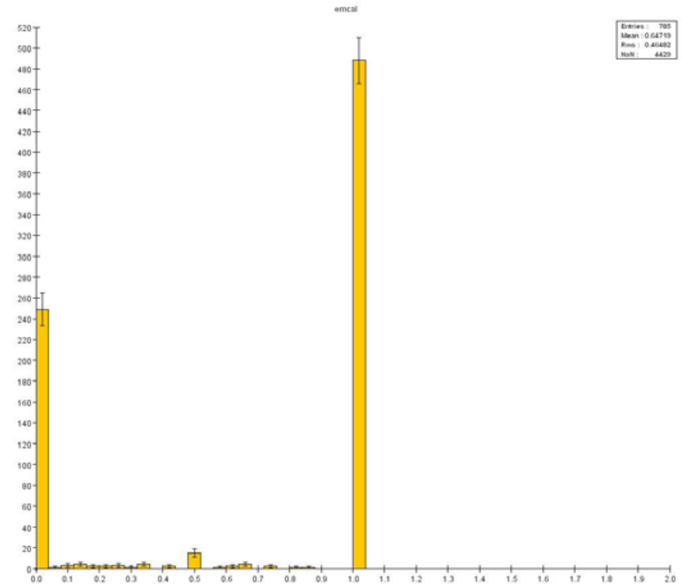
Eff(neu) ~ 88%

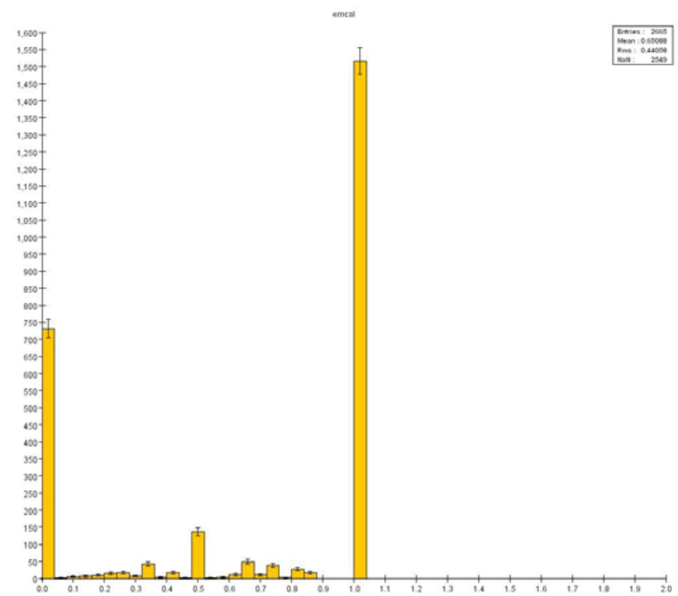
Use the three variables to identify fragments:

1. 72% of the energy from fragments is removed
2. Only lose 12% of real neutral energy

Grad-based





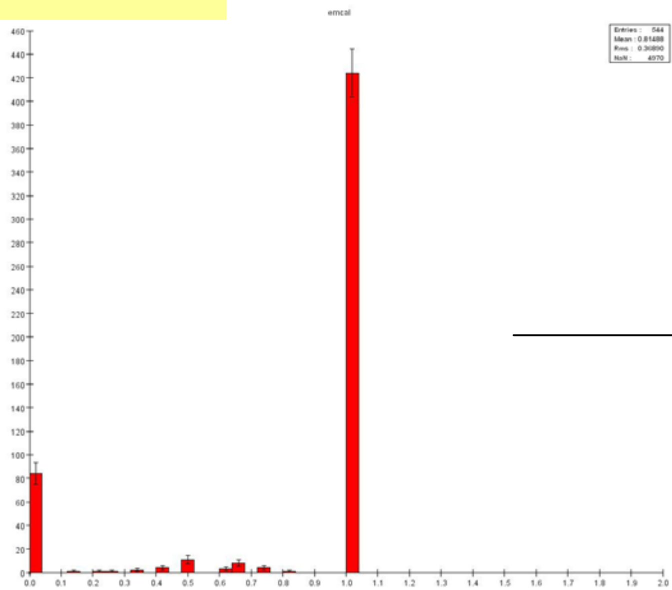


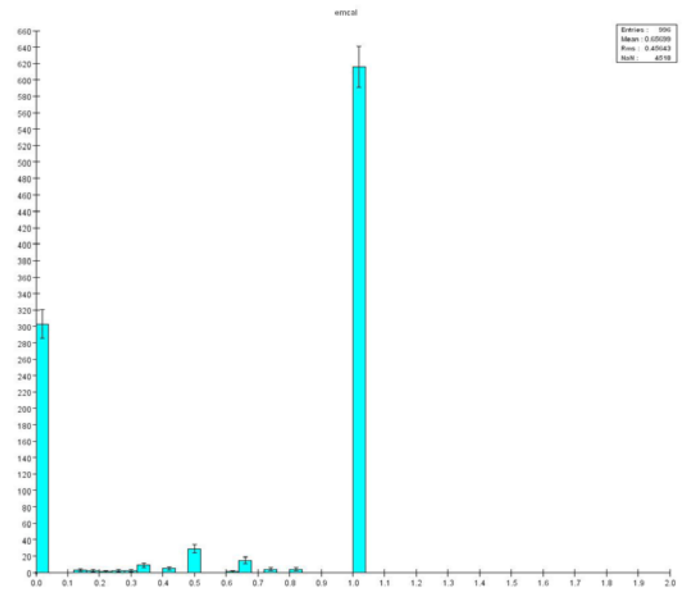
Fraction of cells attached
To main cluster which belong to it
Attachment is based on angular distance

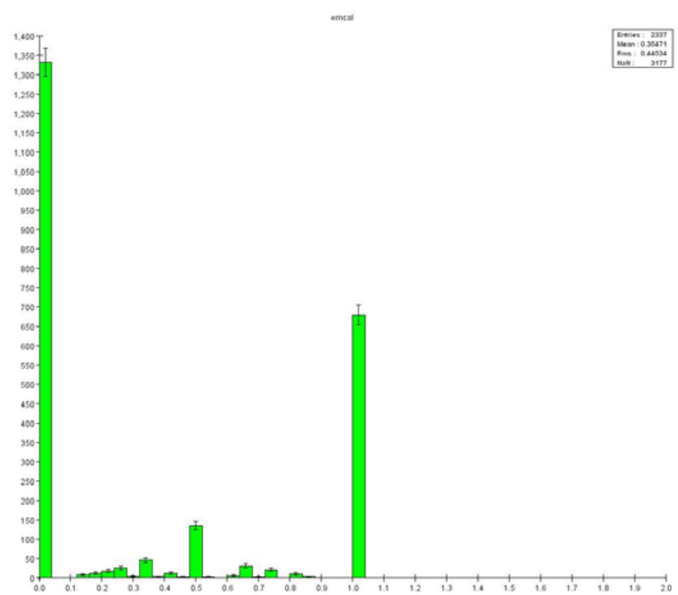
Z-pole events
Charged hadrons

→ Increasing dist

Grad-based





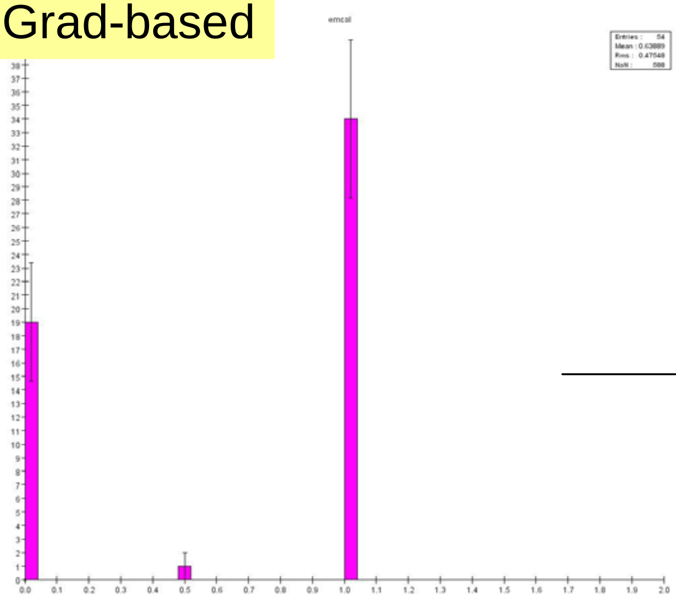


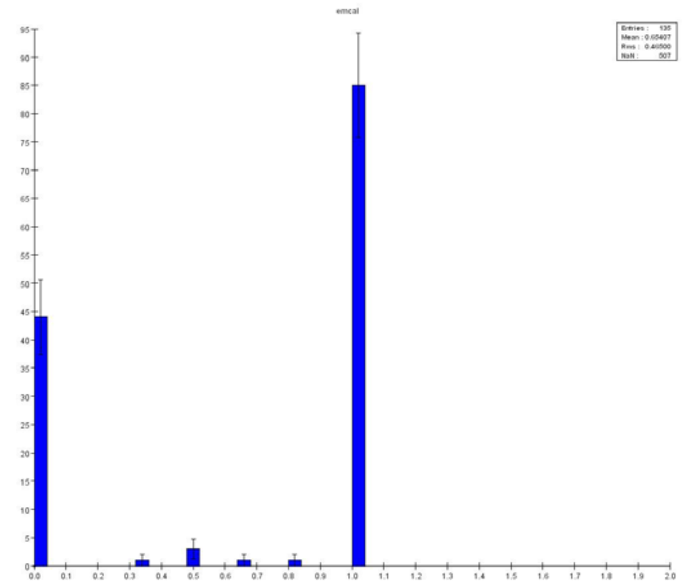
Fraction of cells attached
To main cluster which belong to it
Attachment is based on angular
distance

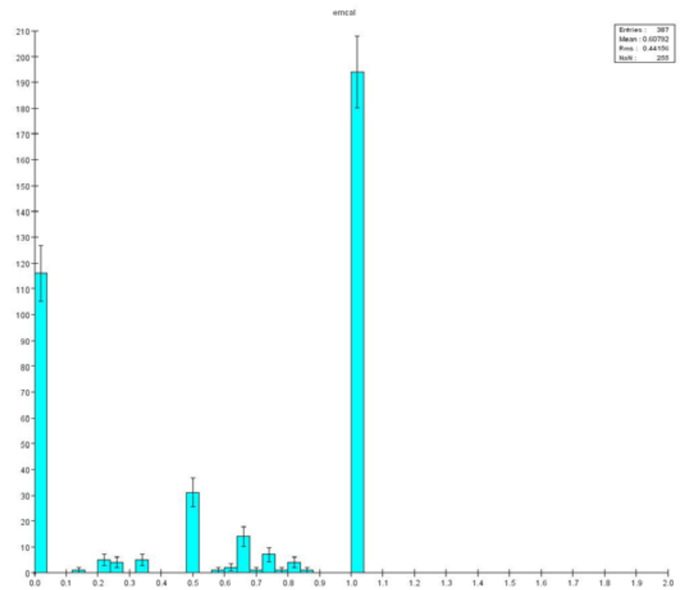
Z-pole events
Photons

→ Increasing dist

Grad-based







Fraction of cells attached
To main cluster which belong to it
Attachment is based on angular
distance

Z-pole events
Neutral hadrons
→ Increasing dist