Algoritmos de clusterización para partículas de muy bajo pT en ATLAS

Carmen Iglesias

Dpto. Física Atómica y Nuclear

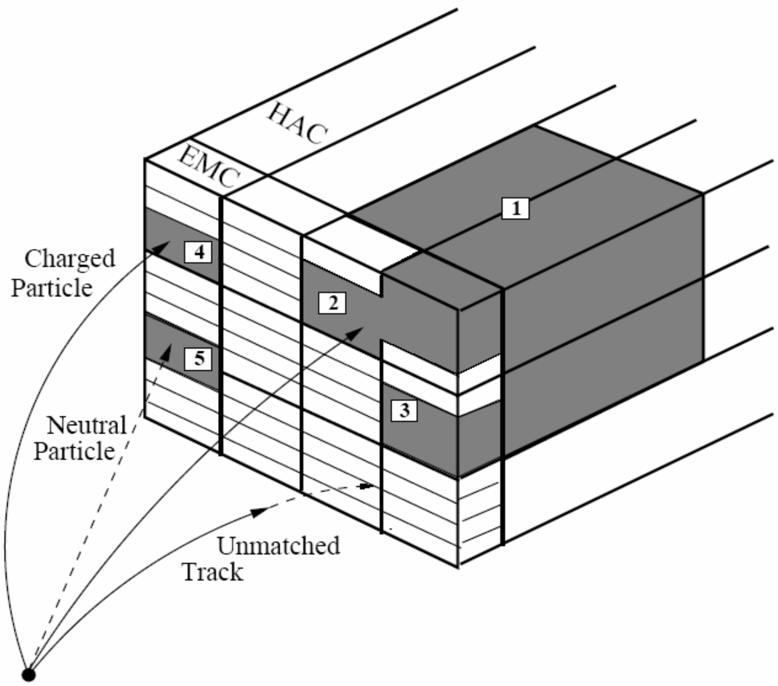
IFIC-Universidad de Valencia

XXX Reunión Bienal de la Real Sociedad Española de Física

Campus Universitario de Ourense de la Univ de Vigo, 12-16 Septiembre de 2005.

Why Clustering is useful for Energy Flow?

- The basic concept of the "Energy Flow" algorithm for jet-finding is to use the tracking detector for the measurement of charged particle momenta and the calorimeter for neutrals.
- We therefore have to reconstruct and subtract neutral clusters before identifying the charged particle's energy deposition in the calorimeter.



- We also need pattern recognition algorithms to associate energy deposition in calorimeter cells with particles.
- EM showers are energetic, very localized and highly correlated.
  - Clustering works well.
- Muons deposit only minimum ionization, but do so along their trajectory
  - Tracking in calorimeter.
  - MIP deposition minimal in any case.
- Hadron showers are broad and unconnected.
  - More difficult to handle.
- In complex events and within jets multiple particles will deposit energy in the same calorimeter cell, and showers will overlap
  - Good clustering is essential to resolve showers
  - A splitting/merging strategy is essential.
- Many cells are hit
  - An efficient algorithm is essential

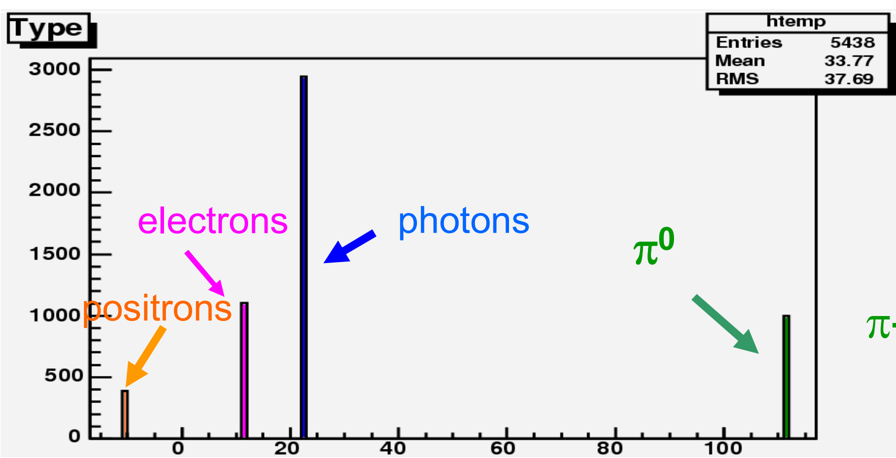
Samples used

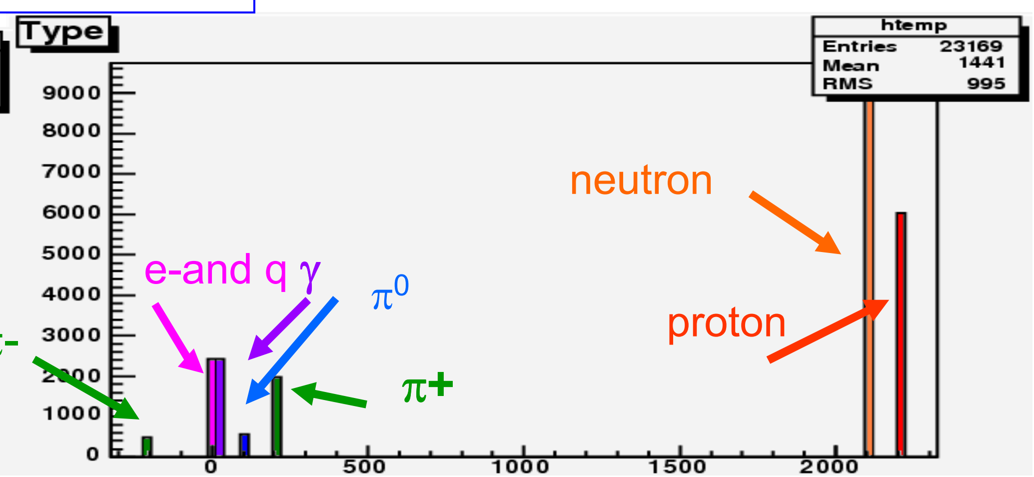
- DC1 samples of pions and neutrons (the main components of jets) at **very low E_T (p_T =1-30 GeV)**, because this is the range of E_T better to apply Energy Flow Algorithm.
- Used to generate ntuples with 1000 events at η=0.3 (central barrel) and ϕ=1.6 of :
  - π'^{0}s, to understand the behavior of photons inside the EM calorimeter.
  - π'$^+$s and neutrons, to know more about the hadronic shower.

  First, without electronic noise applied and later with it.

Shower composition

The shower of the π0 has only e.m. components!!!

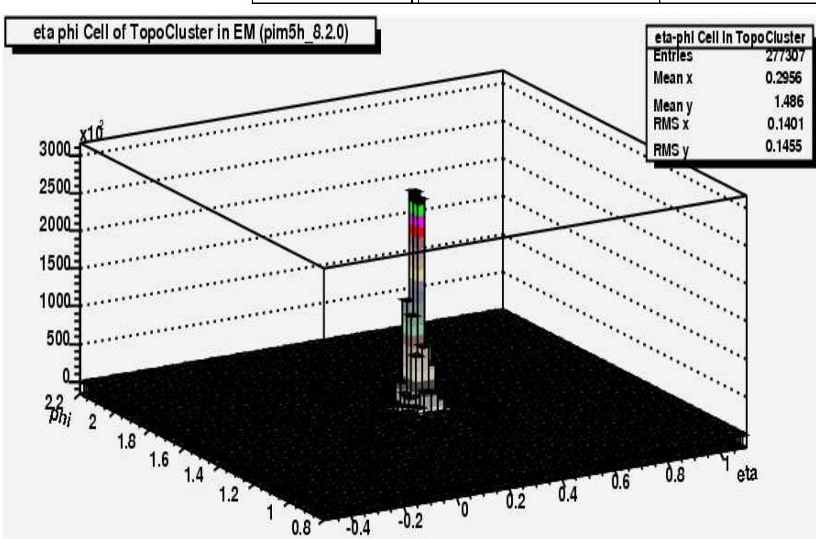


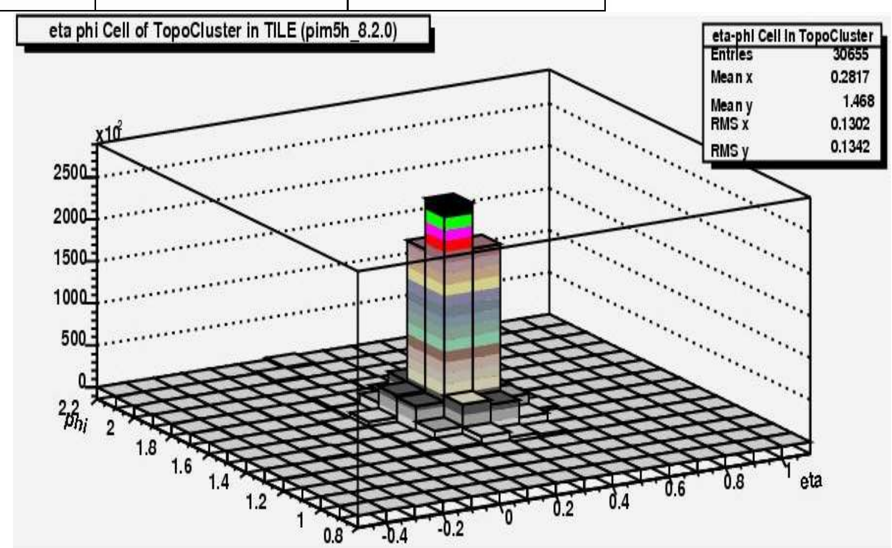


Total energy deposited

- For the $\pi 0$**'s**, as there are only e.m. particles we expect having all the ET deposited in the E.M calorimeter
- For $\pi +$**'s and neutrons** the situation is different. Although, for high pT particles their ET is usually deposited only in the HAD calorimeter, at very low energy, they also deposited their energy in the EM calorimeter (~40-50%) and this deposition increase with the ET of the particles

ET particles	Charged pions		Neutrons	
	ET in EM(%)	ET in Tile (%)	ET in EM (%)	ET in Tile (%)
1 GeV	84	16	66	34
3 GeV	70	30	57	43
5 GeV	62	38	56	44
10 GeV	55	45	53	47
30 GeV	47	53	49	51





INDEX

1)Compare Clustering Algorithms in ATLAS

3)Lower threshold for Seed and Neighbor cells
4)Cone algorithms
5)Topocluster analysis with Electronic Noise

Clustering Algorithms in ATLAS

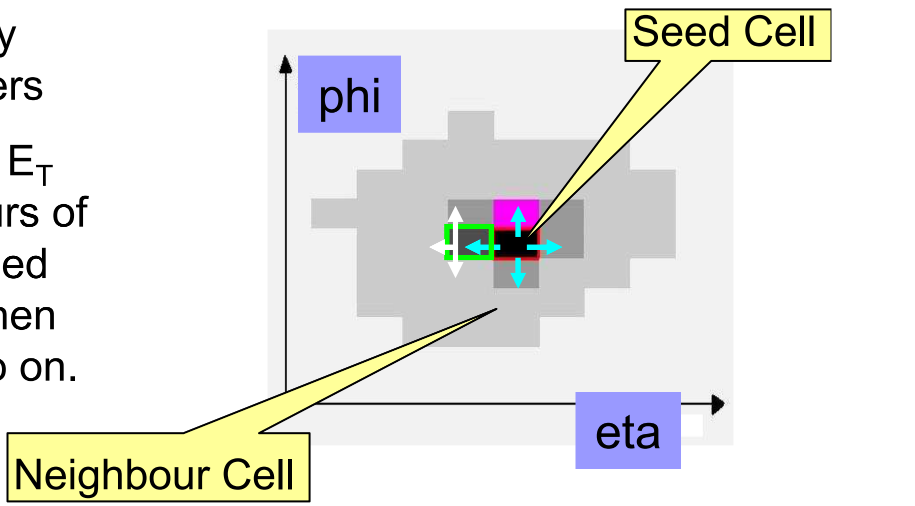
- **Sliding Window (SW) Clustering**
  - Simple search for local maxima of E_T deposit on a grid using a fixed-size "window" made up of a group of contiguous cells in η-ϕ space. Local maxima are found by moving the windows by fixed setps in η and ϕ.
  - Default value is **5 x 5 cells** in each cluster. Another values for SW clusters: **3x5 cells** (for unconverted photons) and **3x7 cells** (for electrons and converted photons).

- **EGAMMA Clusters**
  - Combines Inner detector tracks information with calorimeter clusters (SW) using the default value of **5 x 5 cells** in each cluster
  - Useful for the identification of the e.m objects (photons and electrons).

- **TopoCluster Algorithm**

For the reconstruction of hadronic shower, the energy depositions near by cells have to be merged to clusters

- Cluster is built around a **Seed Cell** which has an E_T above a certain threshold (Seedcut). The neighbours of the Seed Cell are scanned for their E_T and are added to the cluster if this E_T is above the neighborcut. Then the neighbors of the neighbors are scanned and so on.
- The cuts, which are made for the seed and the neighbour, depend on the noise in each cell



Clustering comparison

- First, calculate the ET deposited in all CELLs of the calorimeter and consider it as the "**reference Energy Flow**", i.e., the best resolution that could be reach for the most sophisticated algorithm taking into account the whole ET in all the calorimeter.
  - For **π0's**, compare the resolution of "reference Energy Flow" with the resolution of:
    - **Sliding Window Cluster/EGAMMA cluster**
    - **TOPOcluster in EM calorim**
  - For **π+'s and neutrons,** compare the resolution of "reference Energy Flow" with :
    - **TOPOcluster in EM and Tile**
    - PT of **TRACKS** from XKalman

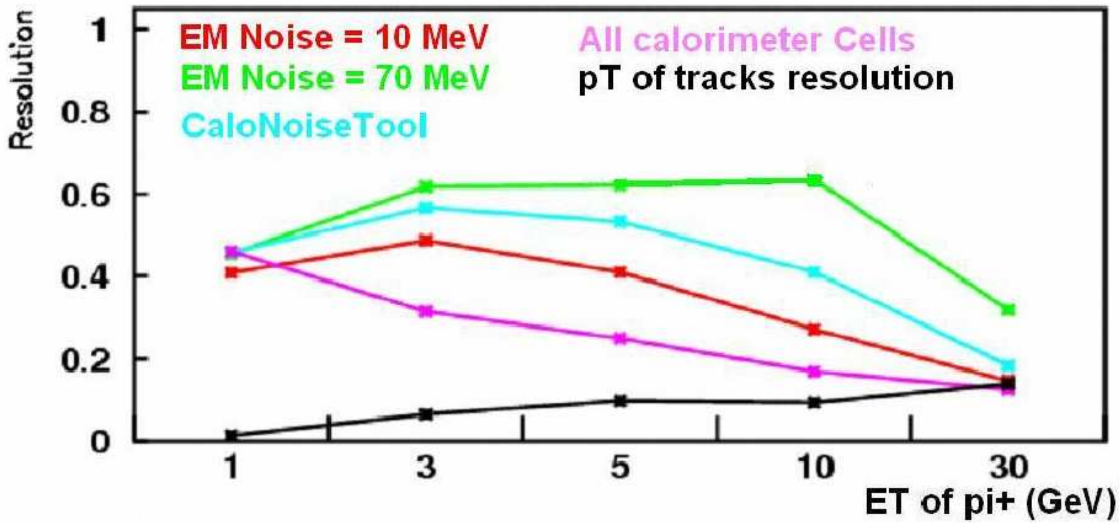
- Compare different ways of reconstructing **TopoCluster at VLE particles**, to find
  →the best ET resolution
  →the larger amount of ET deposited inside the cluster.
  - Use these thresholds:

    - for Seed Cell: $Seedcut = E/\sigma_{noise} = 30$
    - for Neighbor cells: $Neighborcut = |E/\sigma_{noise}| = 3$

  And checking different thresholds for EM Noise:
    - **EM Noise=10 MeV** (lower than realistic case, only useful for checking VLE particles)
    - **EM Noise=70 MeV** (Fix Value by default for EM cal)
    - **CaloNoiseTool=true** (package with a model for the electronic noise)

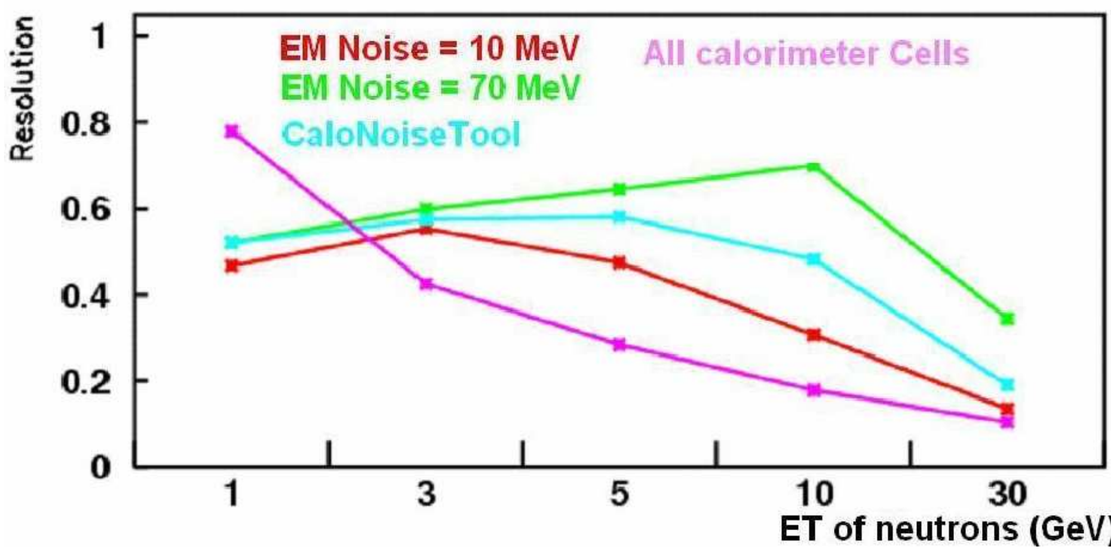
π+'s resolution

•Resolution from PT of **TRACKS** is the best result, but it get worse as the ET of particle increases.
•The best resolution for ET comes from the ET deposited in all calorimeter cells
•Around 30 GeV, ET resolution get better than PT resolution → **limit** of Energy Flow algo



neutrons resolution

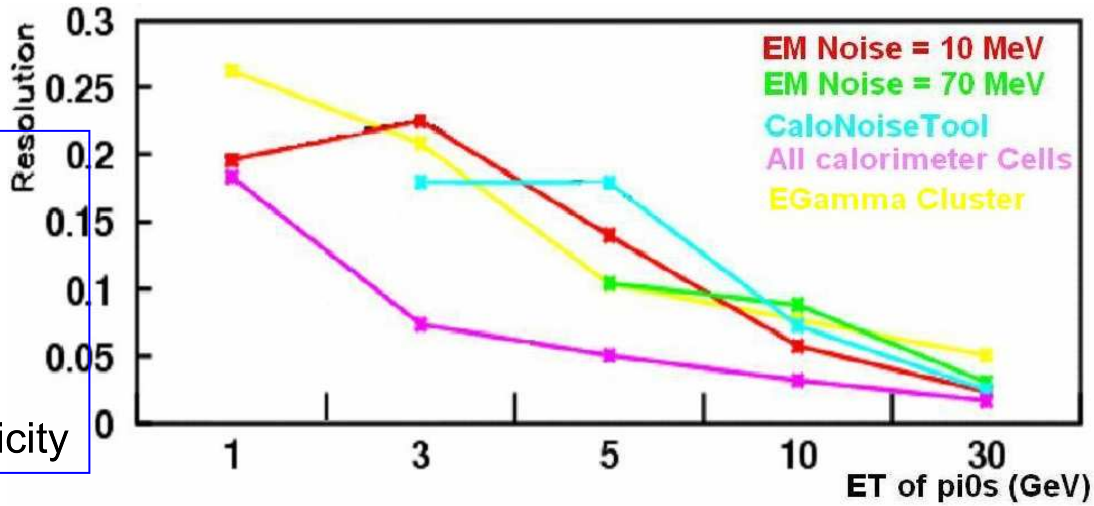
The worst result is **at 1 GeV**:
•ET very similar to the mass of neutron~940MeV.



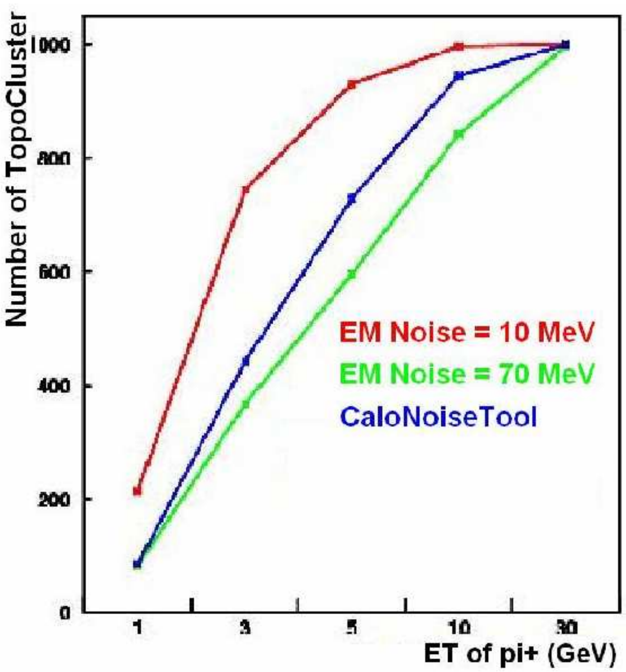
- For the TOPOclusters **CaloNoiseTool** is the most realistic simulation of Electronic Noise. The rest of the analysis will be done using it.

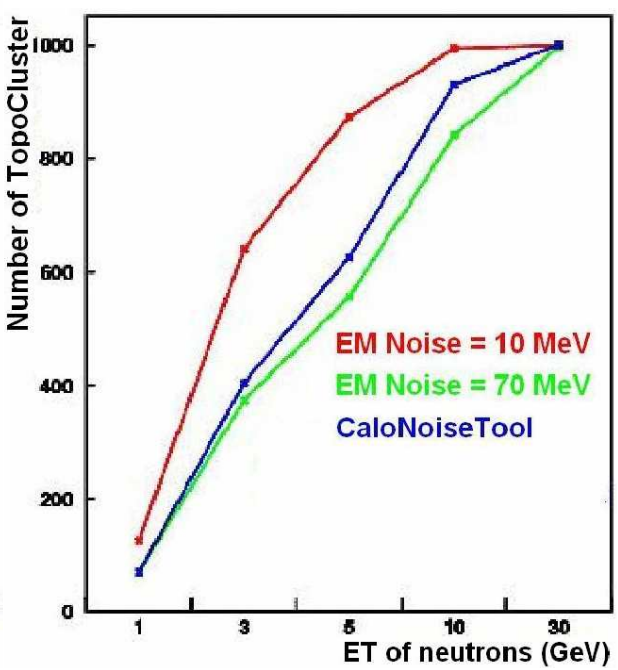
π0's resolution

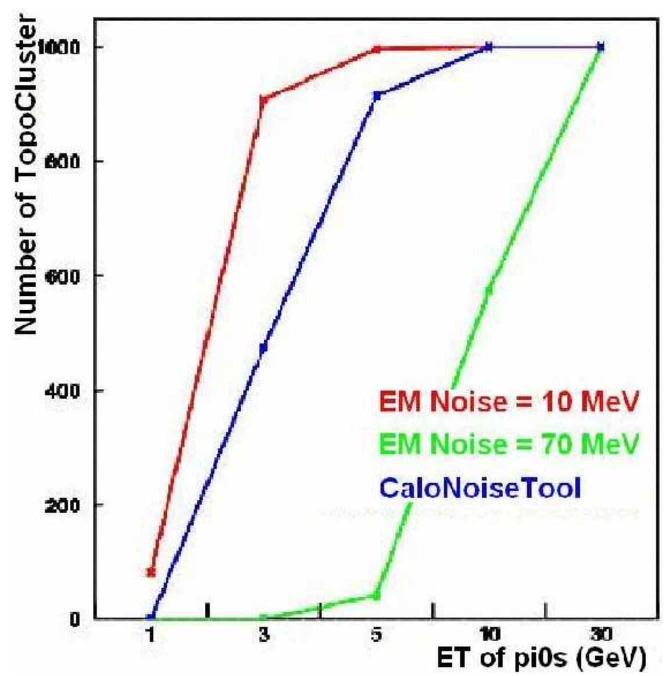
- **π0's** have better resolution than **π+'s and neutrons**
- For **Sliding-Window** clusters, always are obtained the same results as EGamma.
- TopoCluster non defined->low multiplicity



• At 1, 3 and 5 GeV TopoCluster results have non-sense-> Energy resolution increase instead of decreasing with ET. There is a loss in the deposited energy due to the low multiplicity of these clusters







INDEX

1)Compare Clustering Algorithms in ATLAS

2)Lower threshold for Seed & Neighbor cells

3)Cone algorithms

4)Topocluster analysis with Electronic Noise

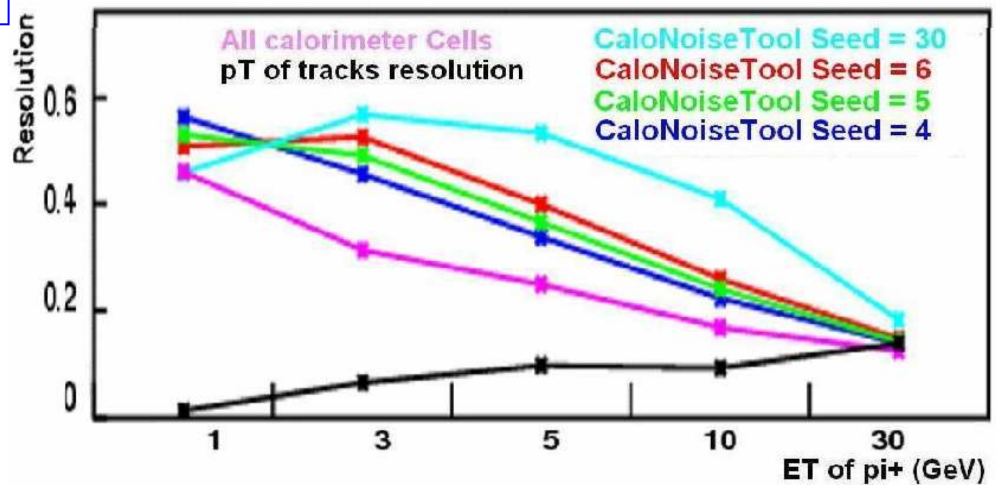
2)Lower threshold for Seed and Neighbor cells

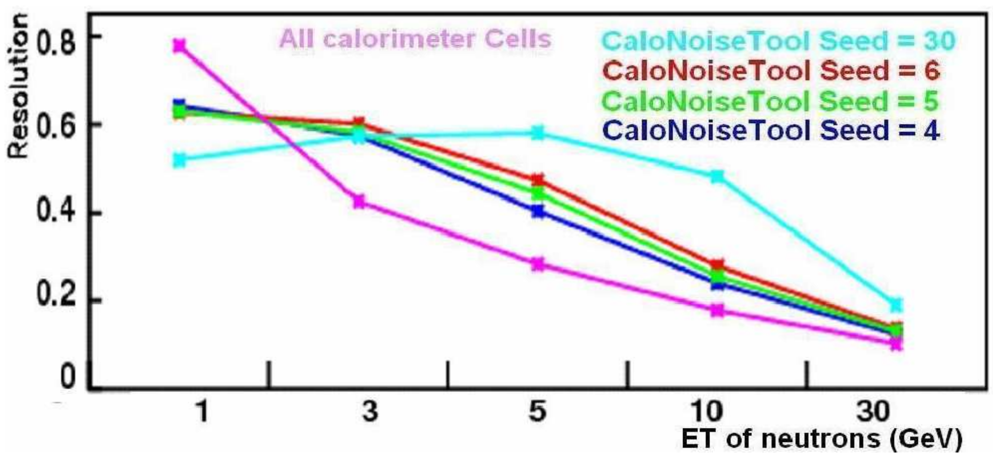
- Lost of ET deposited in TOPOcluster due to the low multiplicity of these clusters→ It's needed to move for lower cuts for the generation of TOPO.

  - Seed_cut: E/σ= 30 → 6, 5, 4…
  - Neigh_cut: E/σ= 3 → 3, 2.5, 2…

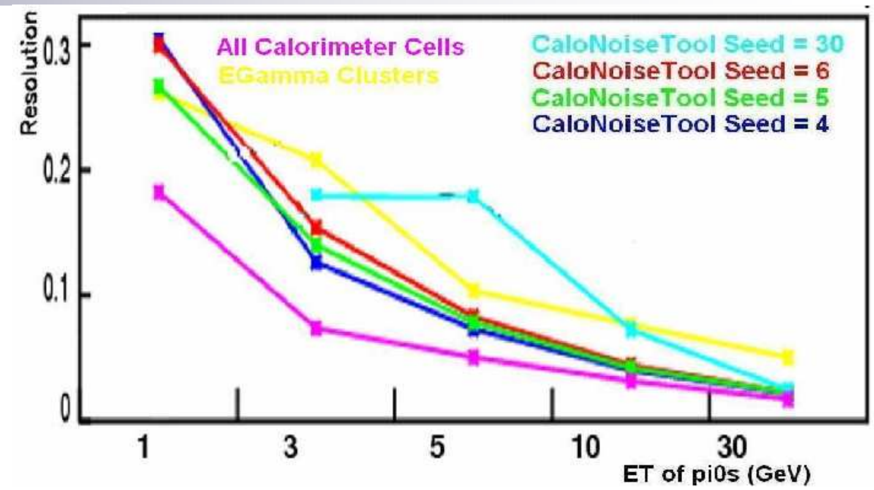
For π**+'s and neutrons**, the best resolution for TOPOcluster using CaloNoiseTool comes from Seed_cut=4 and Neigh_cut=2.

The behaviour of TOPOcluster resolution is more similar to the resolution of the ET deposited by all cells in the calorimeter

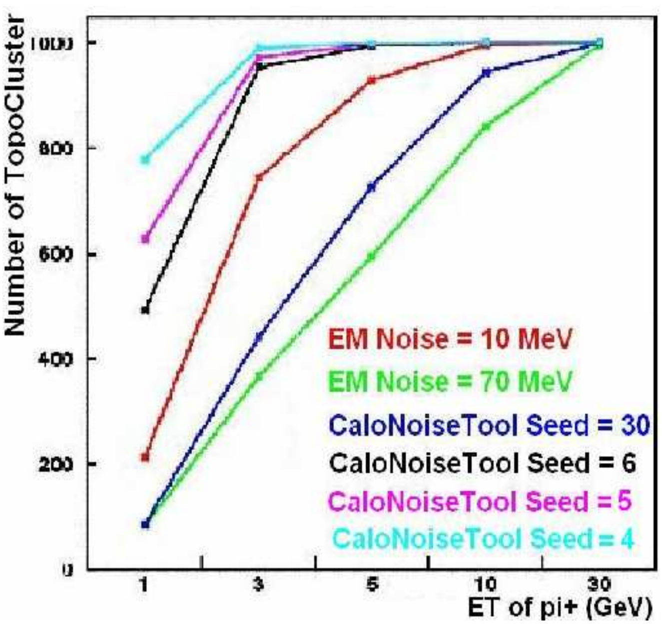


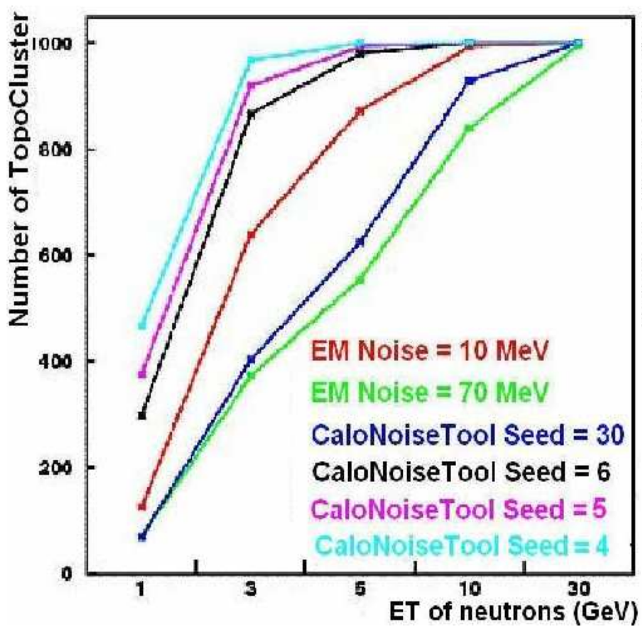


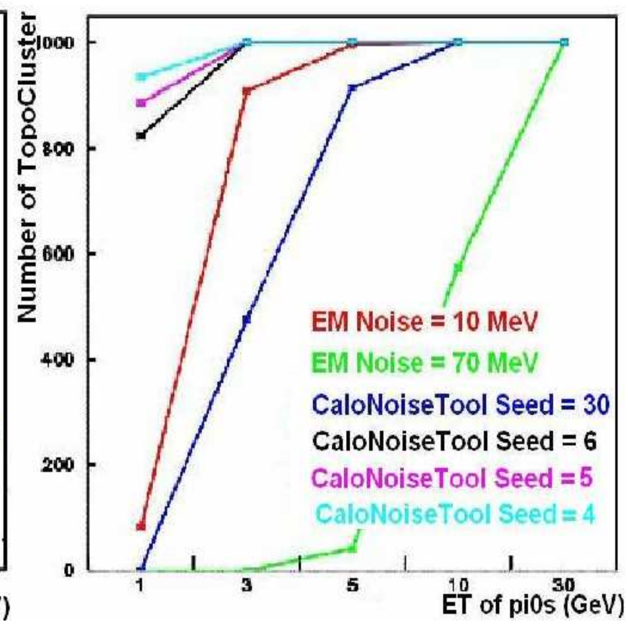
The resolution of TOPOclusters using CaloNoiseTool and Seed_cut=6, 5 o 4 is <u>even better than the resolution of EGamma.</u>



Using these new thresholds the low efficiency of TopoClusters for these single particles at 1-5GeV has been practically eliminated, mainly in π**0's** case. The worst results is for neutrons at 1 GeV, but it also improves with the changed cuts.







Deposited Energy

For π**+'s and neutrons,** changing the Seedcut from 30 to 4, a large increase in the deposited energy is obtained, mainly at 1-5 GeV (the ET is almost the double)

For π**0's,** with the new cuts, the Values of deposited ET for Topo are very similar to the Egamma one and competitive respect to the total energy in all the cells.

Charged pions (% of E_T)					
E_T (GeV)	TopoClusters				All Calo Cells
	Seedcut=30	Seedcut=6	Seedcut=5	Seedcut=4	
1	5.1	26.1	32.5	41.1	**65.7**
3	21.7	49.3	53.5	57.4	**72.9**
5	35.6	59.5	62.2	65.1	**76.1**
10	59.5	72.7	74.4	76.1	**83.4**
30	77.1	79.7	80.5	81.3	**84.6**

Neutrons (% of E_T)					
E_T (GeV)	TopoClusters				All Calo Cells
	Seedcut=30	Seedcut=6	Seedcut=5	Seedcut=4	
1	4.2	11.8	14.0	16.7	**28.4**
3	17.2	33.5	36.3	39.3	**51.3**
5	25.5	44.1	46.7	49.8	**60.4**
10	46.8	60.6	62.5	64.2	**72.2**
30	72.2	75.2	76.0	77.0	**81.1**

Neutral pions (% of E_T)						
E_T (GeV)	TopoClusters				Egamma clusters	All Calo Cells
	Seedcut=30	Seedcut=6	Seedcut=5	Seedcut=4		
1	0.0	52.8	59.8	67.8	76.8	**87.1**
3	36.4	83.7	85.2	86.6	81.1	**95.2**
5	76.2	90.4	91.0	91.8	91.0	**96.8**
10	93.4	94.6	95.0	95.4	95.3	**98.4**
30	97.5	97.7	97.8	97.9	97.8	**99.5**

INDEX

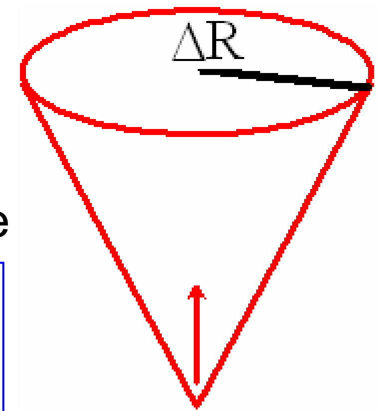
1)Compare Clustering Algorithms in ATLAS

2)Lower threshold for Seed & Neighbor cells

3)Cone Algorithms

4)Topocluster analysis with Electronic Noise

4)Cone algorithms



- Next, study the ET inside a cone with a radius $\Delta R=\sqrt{\Delta\eta^2+\Delta\phi^2}$
  - Different strategies are followed for the different type of particle

Neutral pions

- Cone's centred in η-ϕ coord of EGAMMA cluster
- Cone's centred in η-ϕ coord of TOPO cluster in EM cal
- Cone's centred in η-ϕ coord of TRUTH generated $\pi 0$

Charged pions

- Cone's centred in η-ϕ of TRUTH generated $\pi\pm$
- Cone's centred in η-ϕ of TRACK position at 2nd layer

Neutrons

- Cone's centred in η-ϕ of TRUTH generated neutrons

  - In principle, it's used a cone with $\Delta R<1.0$ → in this first contact, only it's required to select the cone algorithm with the best resolution.

For $\pi 0$'s and neutrons:

Cone's centered in η-ϕ coord of TRUTH

For $\pi\pm$'s:

Cone's centered in η-ϕ of TRACK position at 2nd layer

But with $\Delta R<1.0$ I'm taking into account more than one shower in the same cluster. It's needed to defined ΔR for each type of particle

Defined ΔR of the cone algorithm

- **For $\pi0$'s:**

  From "**Calorimeter Performance**" analysis the cluster size are (for E<100GeV):
  - Unconverted photons: 5x3 cells → $\Delta\phi$= 0.0625 $\Delta\eta$=0.0375 (ΔR<0.073)
  - Converted photons and electrons : 7x3cells → $\Delta\phi$= 0.0875 $\Delta\eta$=0.0375 (ΔR<0.095)

  For the reconstruction of the clusters from $\pi0$'s, will be used:
  - **ΔR <0.1** for starting, because I'm using very low ET
  - **$\Delta\phi$= 0.0875 $\Delta\eta$=0.0375** : 7x3cells
  - **$\Delta\phi$= 0.0625 $\Delta\eta$=0.0375** : 5x3 cells
  - **ΔR<0.0375**: 3x3 cells

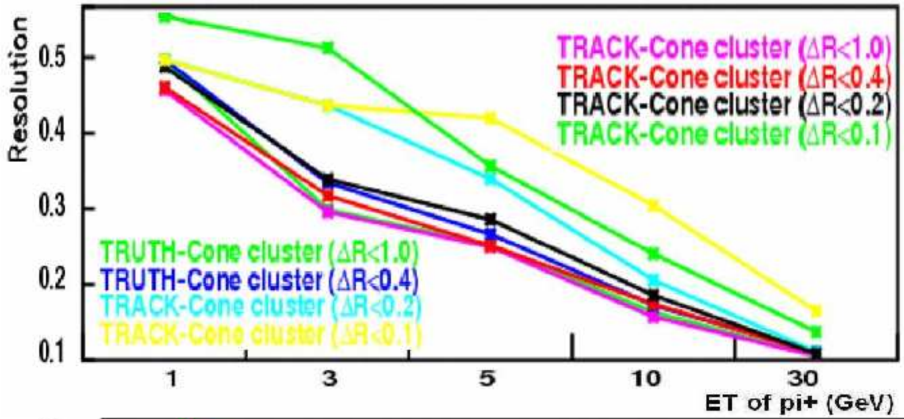
- **For $\pi\pm$'s**:

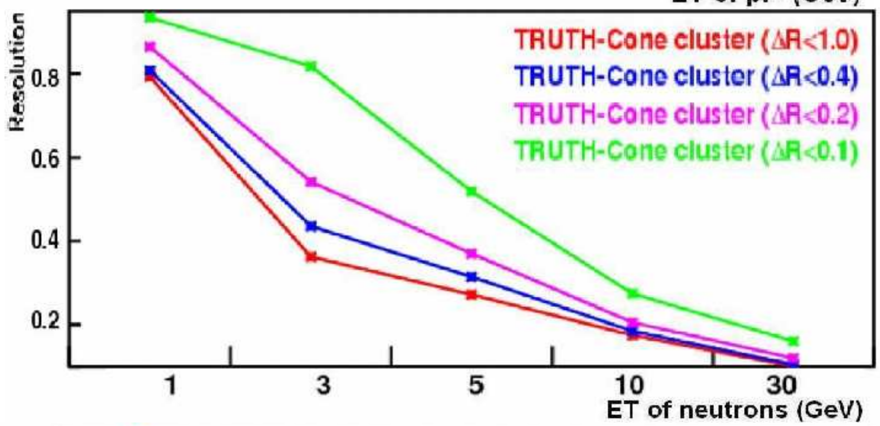
  From **LAr TestBeam analysis**, the cluster size for pions:
  - 7x7 cells (ΔR<0.12),
  - 9x7 cells (ΔR<0.16),
  - 11x11 cells (ΔR<0.20)…

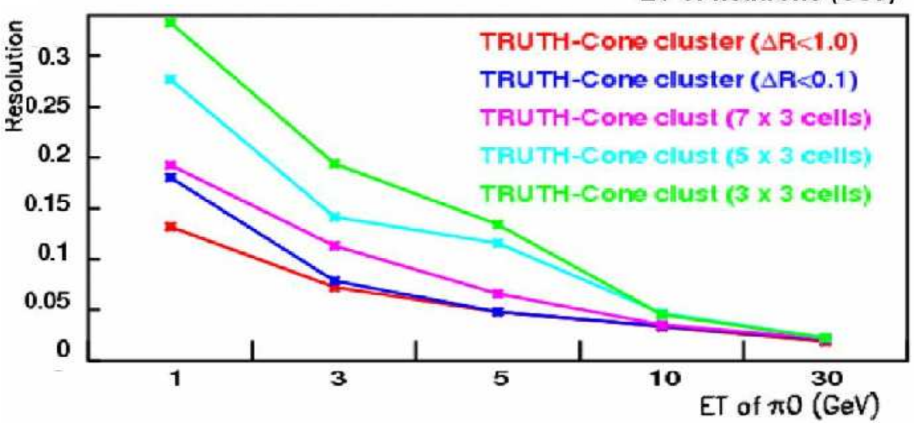
  For the reconstruction of the clusters from $\pi\pm$'s:
  - **ΔR <0.4**
  - **ΔR<0.2**
  - **ΔR <0.1**

- **For neutrons**: the shower of the neutrons must be so wide as the $\pi\pm$'s. So, in principle:
  - **ΔR>0.1, ΔR<0.2** and **ΔR<0.4**

ET Resolution with Cone algorithms





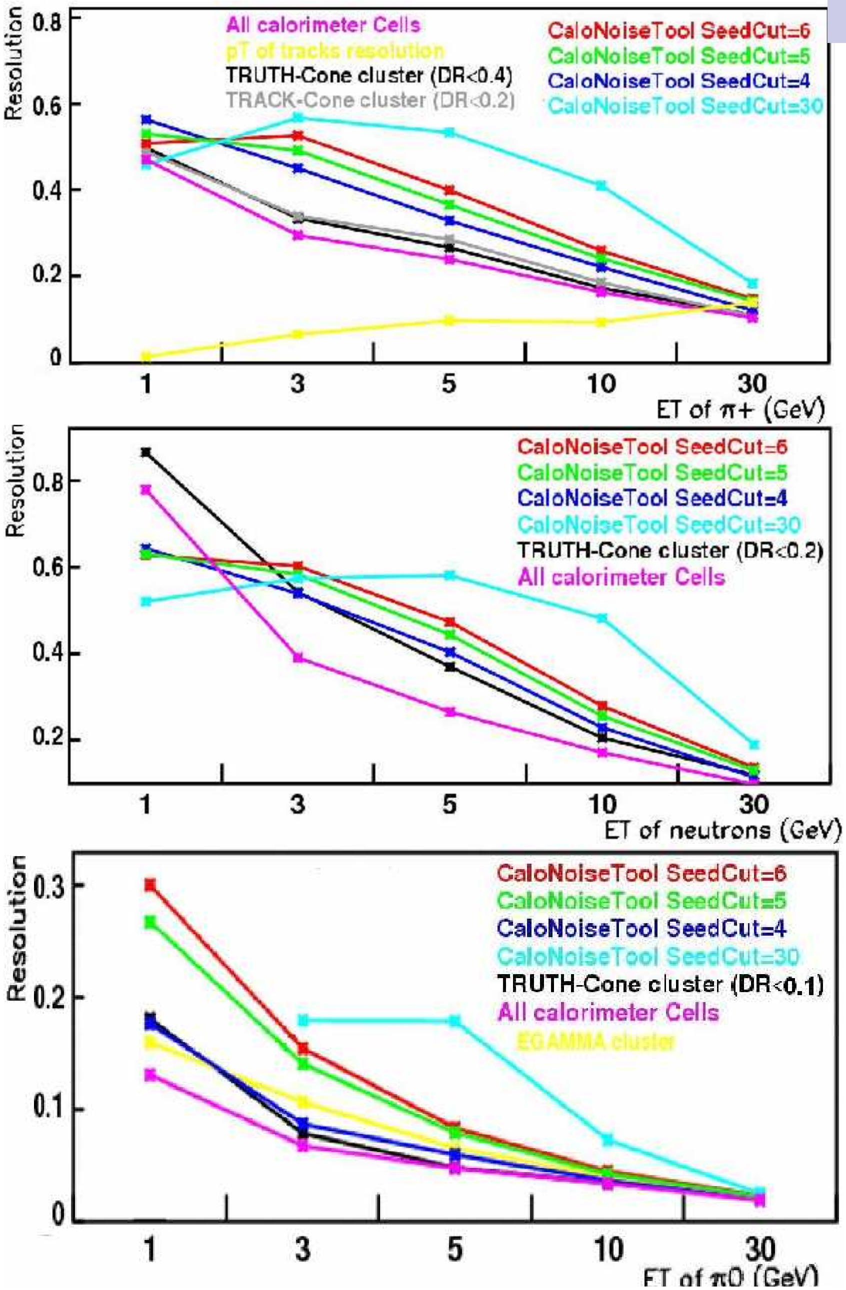


Always the best resolution is for **ΔR<1.0**, but it includes more than the shower of one particle.

For π±'s the best resolution for TRACK-cone with ΔR<0.4, but with **ΔR<0.2**. I have also a good resolution and it let me a better definition of the shower of one π±.
For neutrons: the best resolution with ΔR<0.4, but **ΔR<0.2** is still very good resolution.
In both cases, ΔR<0.1 is too strict to defined hadronic particles.

For π0's: Resolution with **ΔR<0.1** is the better. Clusters with 7x3 and 5x3 cells gives us good resolution but not so good.3x3 is too strict. They could be useful when elect noise will be applied

Clustering Algorithms Comparison



The best algorithm for the reconstruction of the clusters from single particles at very low ET (without electronic noise) is, in each case:

- **For $\pi\pm$'s:** Track-cone with $\Delta R<0.2$ (Truth-cone is close but with $\Delta R<0.4$)
- **For neutrons:** Truth-cone with $\Delta R<0.2$ in general, but TOPO with Seed_cut=4 and Neigh_cut=2 is very near and it's better at 1and 3 GeV.
- **For $\pi 0$'s:** Truth-cone with $\Delta R<0.1$.
  - EGAMMA-cluster give worse resolution, in general, than TOPO and Truth-cone, but gives the best resolution of all at 1 GeV.

Anyway, the results from TOPO algorithm with Seed_cut=4 and Neigh_cut=2 are very competitive for **neutrons and π0's**, for $\pi\pm$**'s** TOPO is a good algo but not enough, for the time being (it will be needed to test new versions of TopoCluster package in the newer release of Athena 8.2.0)

INDEX

1)Compare Clustering Algorithms in ATLAS

2)Lower threshold for Seed & Neighbor cells

3)Cone Algorithms

4)Topocluster analysis with Electronic Noise

Topocluster analysis with Electronic Noise

The energy deposited inside TopoCluster comes from the generated particles, but also from the electronic noise

	π±'s			neu			π0's		
E_T particles (GeV)	Mean of $\frac{ET_{cluster}}{ET_{generated}}$			Mean of $\frac{ET_{cluster}}{ET_{generated}}$			Mean of $\frac{ET_{cluster}}{ET_{generated}}$		
	No Noise	With Noise		No Noise	With Noise		No Noise	With Noise	
	7.8.0	8.2.0	8.2.0 (cut)	7.8.0	8.2.0	8.2.0 (cut)	7.8.0	8.2.0	8.2.0 (cut)
1	0.528	1.05	**0.437**	0.357	—-	**0.403**	0.678	1.217	**0.523**
3	0.581	0.871	**0.558**	0.406	0.681	**0.424**	0.866	1.135	**0.841**
5	0.652	0.841	**0.646**	0.498	0.687	**0.497**	0.918	1.084	**0.925**
10	0.761	0.864	**0.761**	0.643	0.749	**0.636**	0.954	1.038	**0.968**
30	0.813	0.851	**0.817**	0.110	0.811	**0.771**	0.919	1.010	**0.986**

Asking for a minimum value of ET in Seed Cell and Neighbor cells:

- Seed Cell >200MeV
- Neighbor cells >80MeV

a similar value of $\frac{ET_{cluster}}{ET_{generated}}$ without noise is obtained.

E_T (GeV)	Cells of π^+'s								
	No Cut Total	Cut Total	Times Size	No Cut EM	Cut EM	Times Size	No Cut Tile	Cut Tile	Times Size
1	182	12.8	**14.2**	170	11.4	**14.9**	11	1.4	**7.5**
3	257	38.7	**6.6**	233	32.0	**7.3**	24	6.7	**3.6**
5	311	62.6	**4.9**	276	50.5	**5.4**	35	12.1	**2.9**
10	380	97.0	**3.9**	330	77.3	**4.3**	50	19.6	**2.5**
30	513	168	**3.0**	432	130	**3.3**	81	38.1	**2.1**

- After these cuts, the size ot the Topocluster is up to 14 times smaller.
- This difference is more important for the EM calo because there the level of noise with respect to the signal is bigger.

E_T particles (GeV)	$\pi^\pm$'s E_T Resolution No Noise 7.8.0	$\pi^\pm$'s With Noise 8.2.0	$\pi^\pm$'s With Noise 8.2.0 (cut)	neu E_T Resolution No Noise 7.8.0	neu With Noise 8.2.0	neu With Noise 8.2.0 (cut)	π^0's E_T Resolution No Noise 7.8.0	π^0's With Noise 8.2.0	π^0's With Noise 8.2.0 (cut)
1	56.30	44.57	**60.71**	64.31	—-	**70.20**	30.52	35.50	—
3	45.65	38.87	**55.77**	57.54	46.19	**65.12**	12.66	20.72	**25.89**
5	33.82	30.65	**41.39**	40.41	34.04	**51.07**	7.35	14.35	**15.19**
10	22.41	20.84	**26.25**	23.98	22.43	**28.67**	4.02	7.77	**7.30**
30	13.91	13.36	**13.09**	12.39	11.94	**12.60**	2.21	3.11	**2.89**

The E_T resolution get worse with the application of these cuts→ there is a loss in energy reconstruction of the clusters. WHY?

Because we have applied a general threshold to the ET_{cell} for all calorimeter, and the electronic noise contribution is different in each layer of LAr and Tile.

- ❑ Seed Cell >200MeV
- ❑ Neighbor cells >80MeV

- EM Presampler: σ_{noise}= 60-150 MeV
- EM Front: σ_{noise}= 32-55 MeV
- EM Middle: σ_{noise}= 60-130 MeV
- EM Back: σ_{noise}= 70-100 MeV
- Tile A: σ_{noise}= 25-70 MeV
- Tile BC: σ_{noise}= 25-70 MeV
- Tile D: σ_{noise}= 25-65 MeV

Conclusions

- **WITHOUT NOISE**:
  - The best E resolution for VLE particles is obtained with cone algorithms
  - TopoCluster is a very competitive algorithm but doing the changes:
    - Using CaloNoiseTool to model th eEM Noise
    - Applying lower thresholds to Seed and Neighbor cells:
      - SeedCut=4 and NeighborCut =2

  TopoClusters is event better than EGamma cluster for π0's.

- **WITH NOISE:**
  - The E resolution get worse for TopoCluster
  - If we try to remove electronic noise, we also get a loss in energy from particles
    - It will be needed to applied ET thresholds in each layer of LAr and Tile