Clustering in the calorimeter for PFA

Dhiman Chakraborty

NORTHERN ILLINOIS UNIVERSITY

nicadd

2005 INTERNATIONAL LINEAR COLLIDER WORKSHOP

Stanford, California, USA 18-22 March, 2005

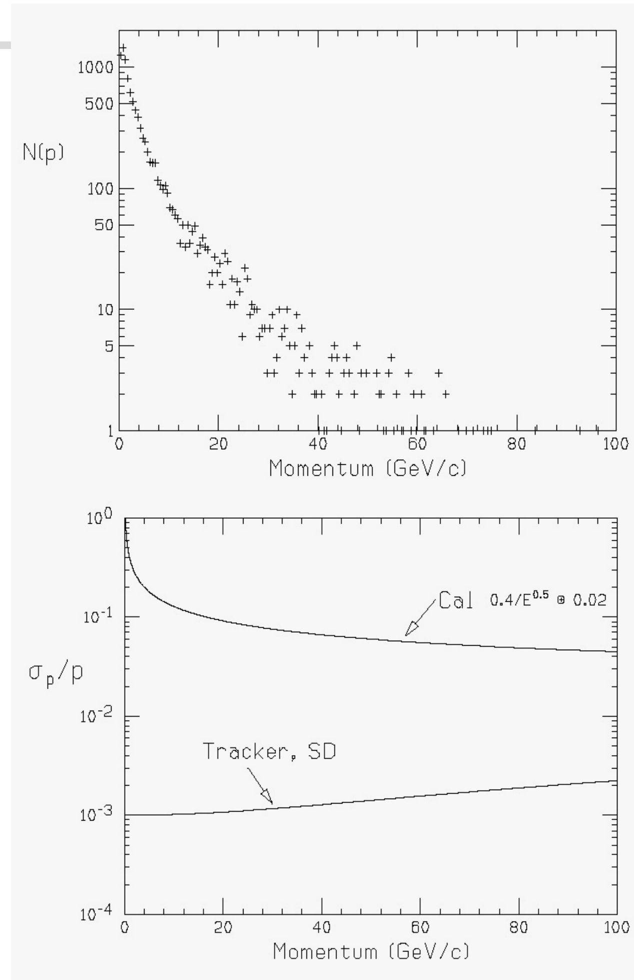
Introduction

- Primarily interested in exploring the digital hadron calorimeter option in general, with scintillator as the active material in particular.
- For digital algorithms and results for single particles, Refer to talks given at the LDC meeting: Paris, Jan 2005.
- Results are preliminary.

Particle-Flow Algorithm (PFA)

- Charged particles in a jet are more precisely measured in the tracker
- A typical jet consists of:
  - 64% charged particles,
  - 24% photons,
  - 11% neutral hadrons.
- Use tracker for charged,
- Calorimeter for neutrals only.
- Must be able to separate charged particle energy clusters from neutrals inside a jet in calorimeter

  ⇒ need fine 3d granularity.



The "SD" calorimeter

- ECal:
  - 30 layers, silicon-tungsten.
  - 5mm x 5mm transverse segmentation.
- HCal:
  - 34 layers, scintillator-steel
  - Transverse segmentation varied from 2 to 16 cm^2 (average in a projective geometry).
- Magnetic field: 5T
- Support structures, cracks, noise, x-talk, attenuation, inefficiencies,… not modelled.

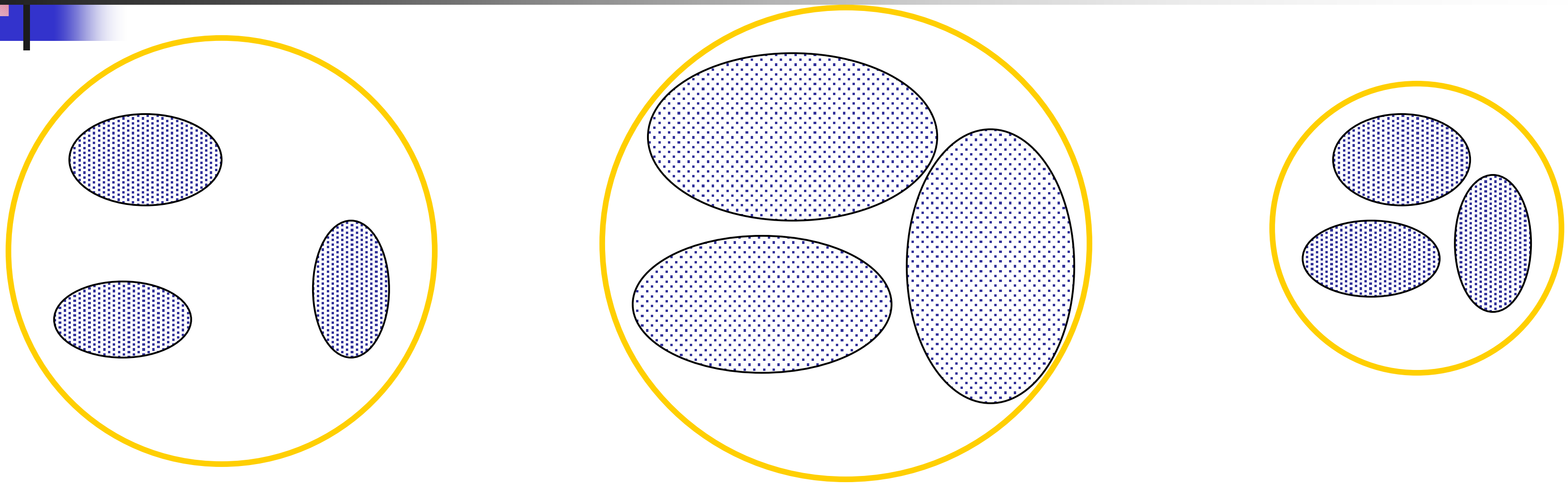
Clustering (reported in past)

- Seeds: maxima in local density:

$$d_i = \Sigma\ (1/R_{ij})$$

- Membership of each cell in the seed clusters decided with a distance function.
- Only unique membership considered.
- Calculate centroids.
- Iterate till stable within tolerance.

Separability of clusters



Best separability is achieved when width of a cluster is small compared to distances between clusters.

$$J = \mathrm{Tr}\{S_w^{-1}\ S_m\}$$

Separability of clusters (contd.)

where

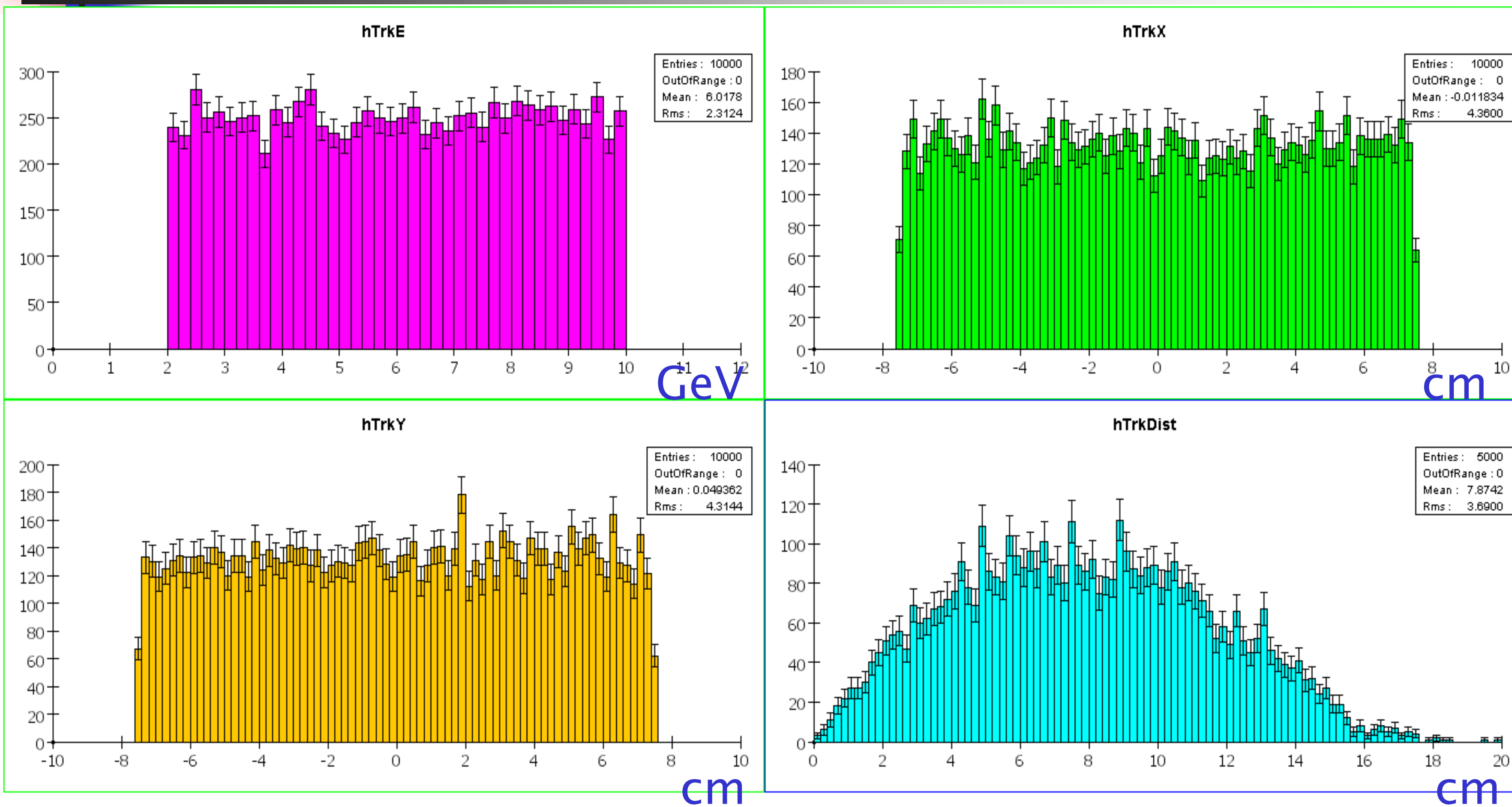
$S_w = S_i W_i S_i$

$S_i =$ covariance matrix for cluster c_i (in x, y, z)

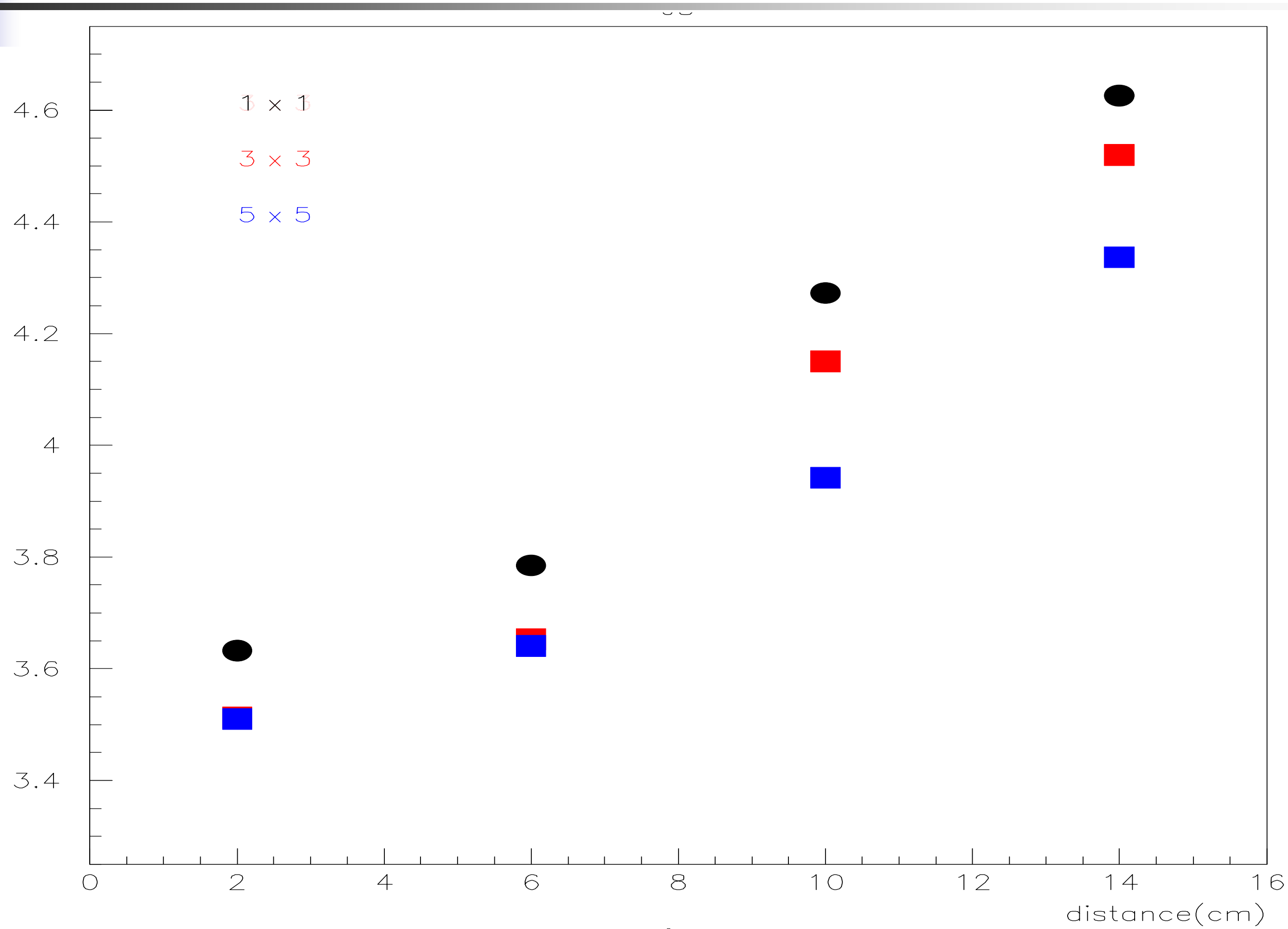
$W_i =$ weight of c_i (choose your scheme)

$S_m =$ covariance matrix w.r.t. global mean

Two (parallel) π^+'s in TB sim:



Two (parallel) π^+'s in TB prototype sim: separability (J) vs. track distance for different cell sizes



Another measure of separability

$$B = a\,(\mu_i-\mu_k)^T\,(\{S_i+S_k\}/2)^{-1}\,(\mu_i-\mu_k)\ +$$
$$b\,\ln\{(|(S_i+S_k)|/2)(|\,S_i\,||\,S_k\,|)^{-1/2}\}$$

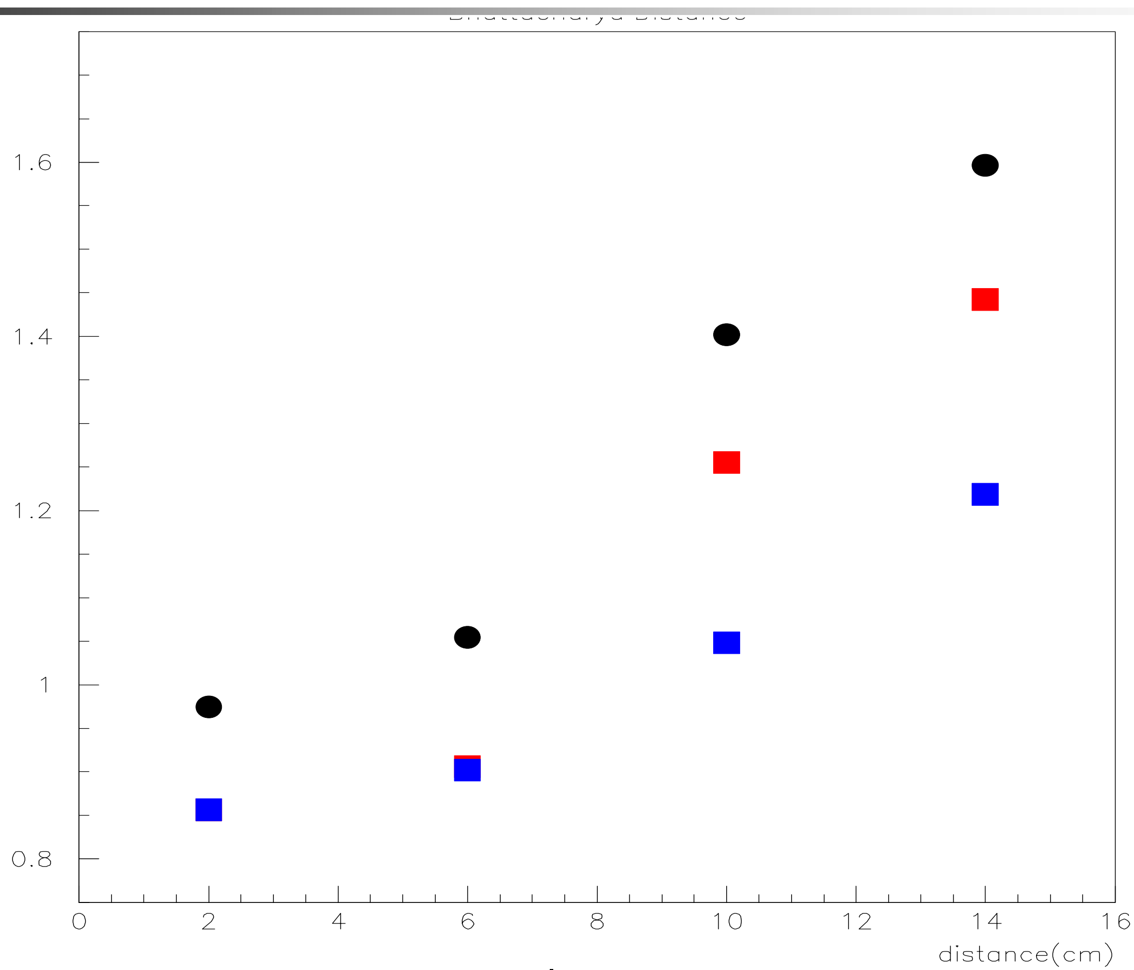
$a,b > 0.$

$\mu_i =$ mean, $S_i =$ covariance of i–th cluster

1st term: separation due to mean difference,

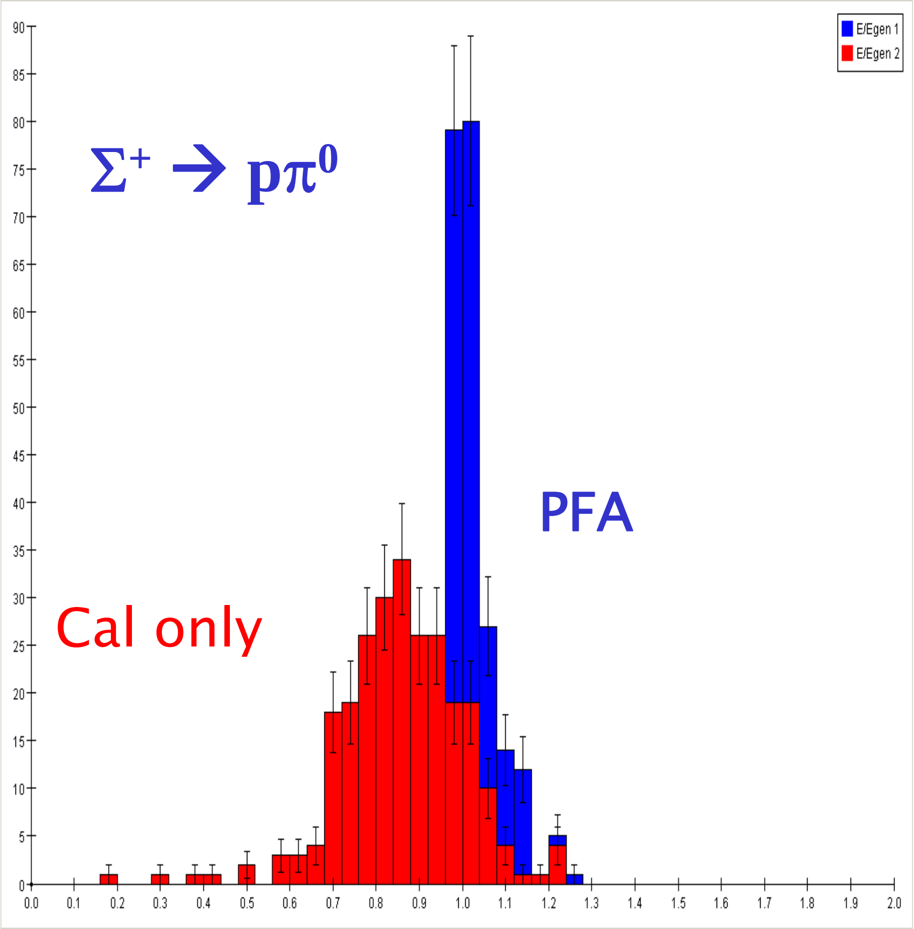
2nd term: separation due to covariance difference

Two (parallel) π^+'s in TB sim: separability (*B*) vs. track distance for different cell sizes



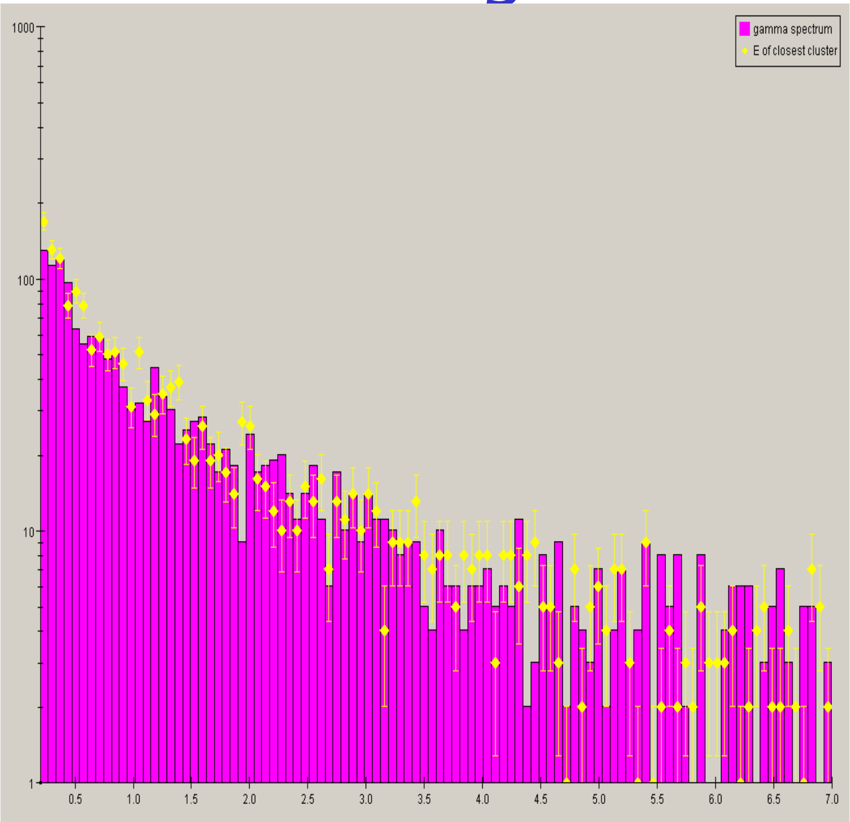
DHCal: Particle-flow algorithm (NIU)

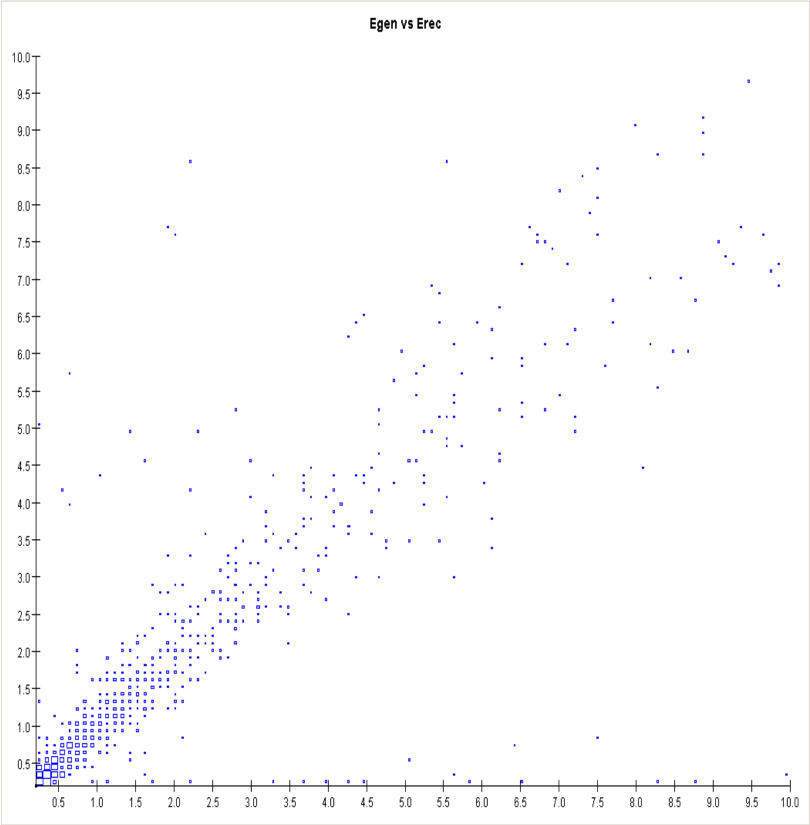
- Nominal SD geometry.
- Density-weighted clustering.
- Track momentum for charged,
- Calorimeter E for neutral particles.



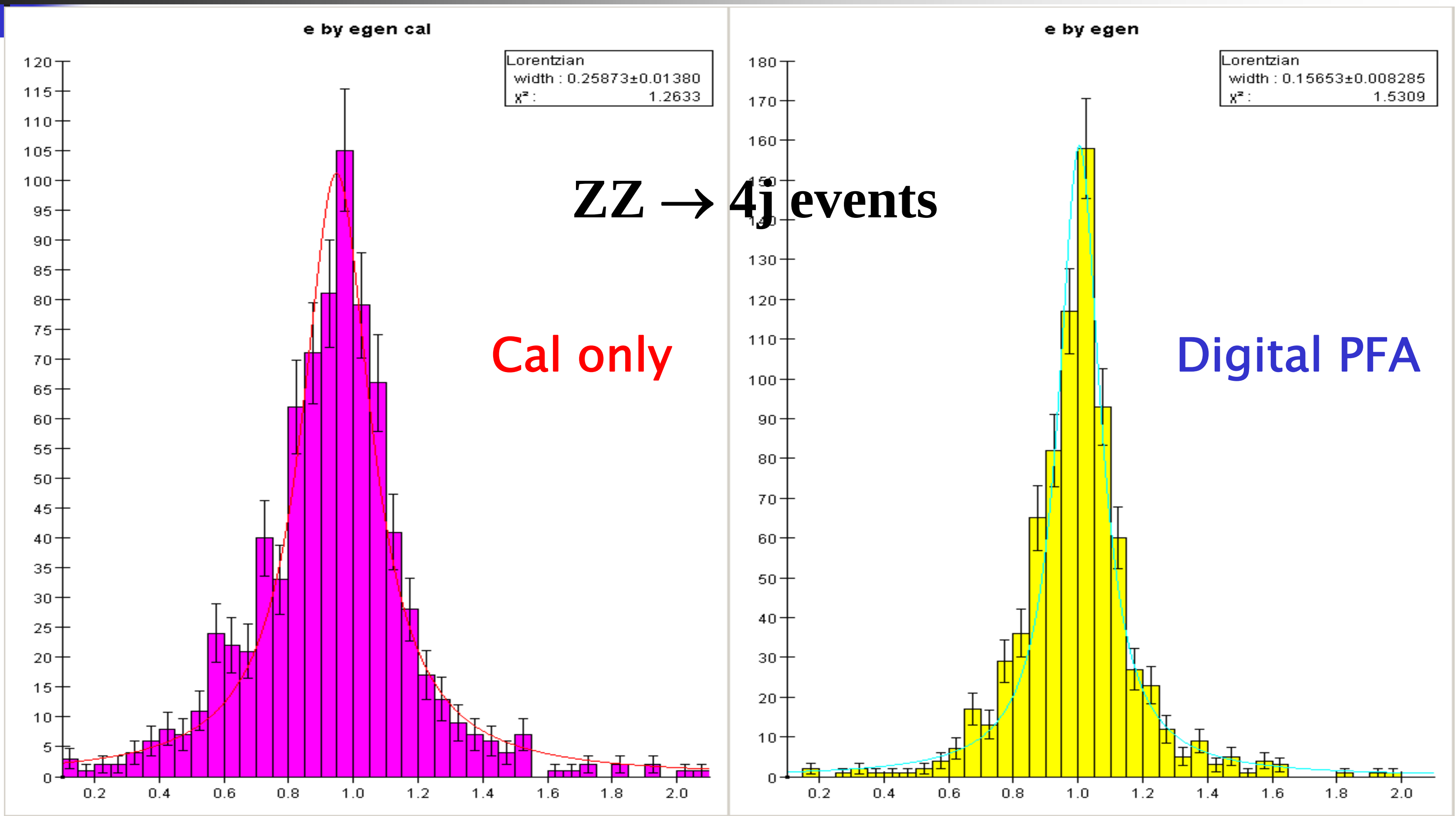
DHCal: Particle-flow algorithm (NIU) Photon Reconstruction inside jets

Excellent agreement with Monte Carlo truth:





DHCal: Particle-flow algorithm (NIU)
Reconstructed jet resolution



PFA Jet Reconstruction summary (past)

- Cone clustering in the calorimeters,
- Flexible definition of weight (energy- or density-based),
- Generalizable to form “proto-cluster” inputs for higher-level algorithms.
- Replace cal clusters with matching MC track, if any.
- Based on projective geometry.
- New clustering algorithms taking shape.

Current approach

- Detector geometry optimization
- Need to make sure that one is not studying the systematics of a particular algorithm
- Develop a suite of algorithms whose common performance features could be used as a guide to detector optimization
- Just starting….

Algorithm essence

- Define neighborhood for a cell
- Calculate neighbor density D for each cell i
- If(D_i==0) ?

  else

  calculate $(D_j - D_i)/d_{ij}$

  where j is in the neighborhood
- find max []

Algorithm essence (contd.)

- If max[] is –ve

  i starts a new cluster

  if max[] is +ve

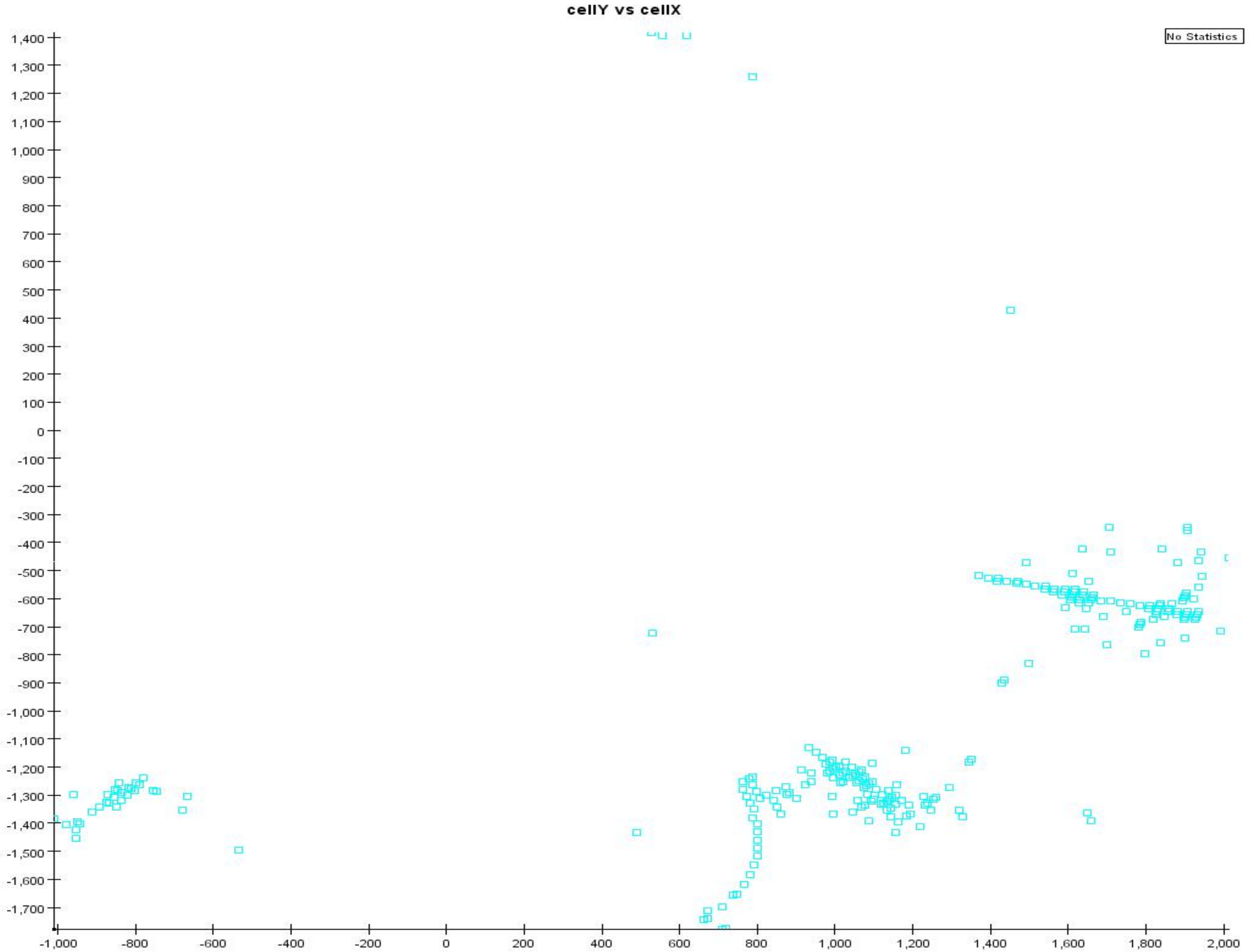
  j is the parent of i

  if max[] == 0

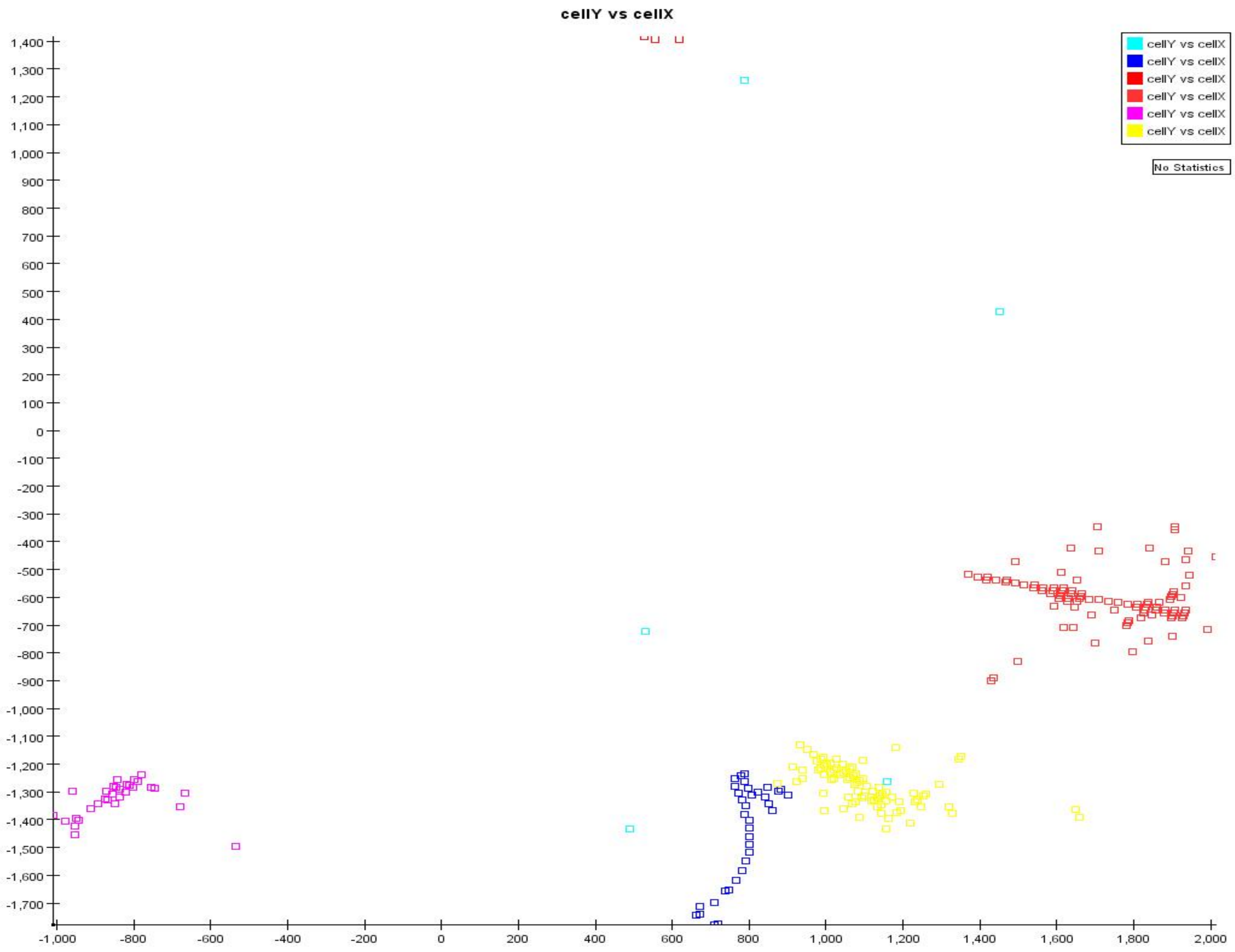
  avoid circular loop

  attach to nearest j

Single hadrons

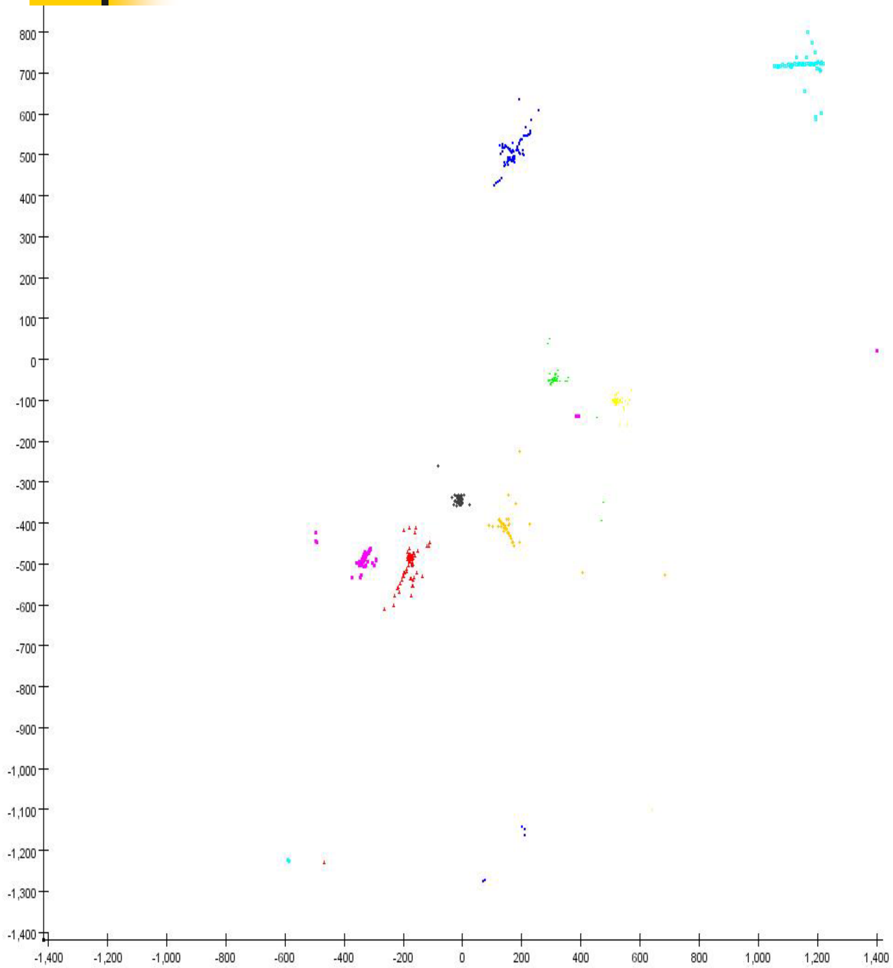


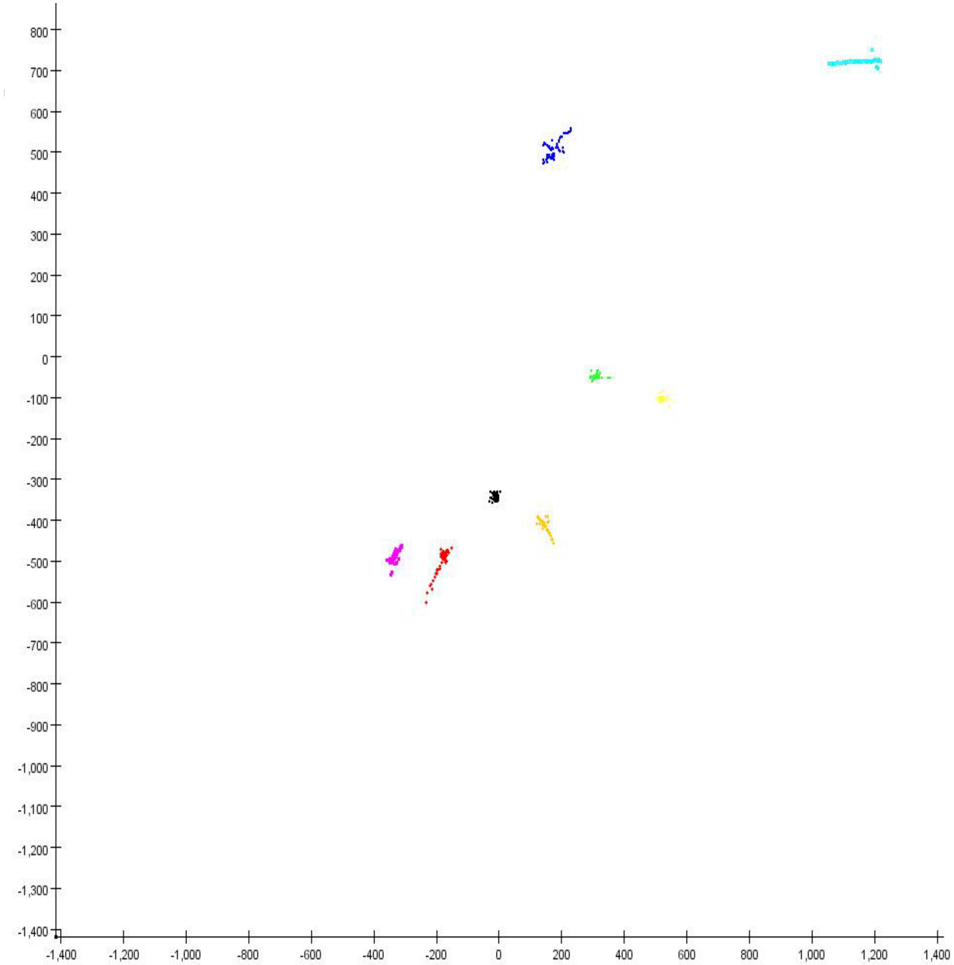
Single hadrons



Some early attempts at PFA

Z→qq Events


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

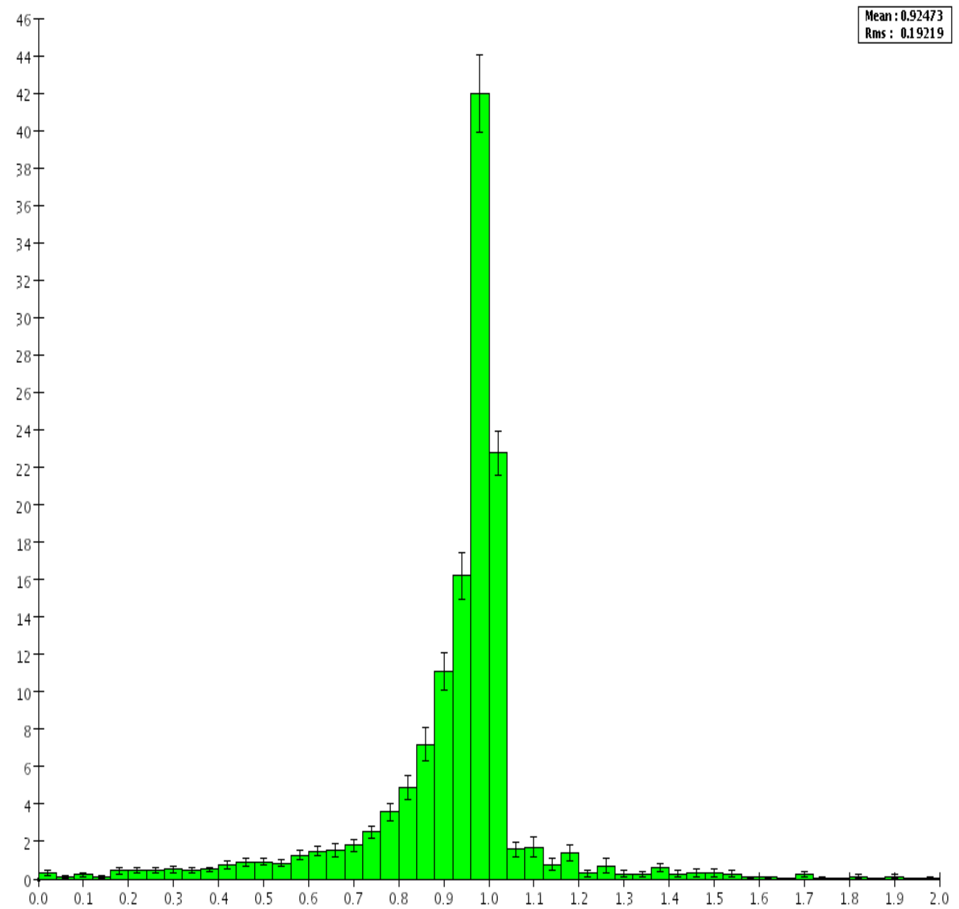

800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1,000
-1,100
-1,200
-1,300
-1,400
-1,400 -1,200 -1,000 -800 -600 -400 -200 0 200 400 600 800 1,000 1,200 1,400

The confusion term

- Internal to calorimeter.
- Reconstruct “gen” and “rec” clusters,
- A “gen” cluster is a collection of cells which are attached to a particular MCparticle. All detector effects are included in this cluster.
- Find centroids and match to nearest “rec” cluster, making sure that no cluster gets associated twice.
- Somewhat conservative.

Z→qq Events

- Calculate E/Egen for each generated cluster
- Enter into histogram with weight Egen/Etotal.



Objective goal

We should try to factorize algorithms so the higher level ones are as independent of hardware choices as possible. Interface design started.

- This is crucial for testing of ideas across detector designs.
- It is also important for international cooperation on algorithm development.

Work in progress

- Generalized “proto-clusters” to absorb geometry & technology details,
- Interfaces to existing pieces of code so they can be serialized in a standard manner (clean-up/ re-factorization with better OO design).

Work in progress

- Building a framework to facilitate algorithm and detector performance & geometry optimization studies.
- A number of algorithms coded, detailed evaluation underway.
- Expect to make substantial progress by Snowmass ‘05.