Particle Flow reconstruction based on the directed tree clustering algorithm

Dhiman Chakraborty, Guilherme Lima, Vishnu Zutshi
NICADD / Northern Illinois University

NORTHERN ILLINOIS
UNIVERSITY

Calor 2006 – Chicago
June 5-9, 2006

Talk outline

- Brief introduction and goals
  motivations for particle flow reconstruction
- Perfect PFA
  what can we expect?
- Real PFA
  how we are doing the real PFA
- Summary
  status and plans

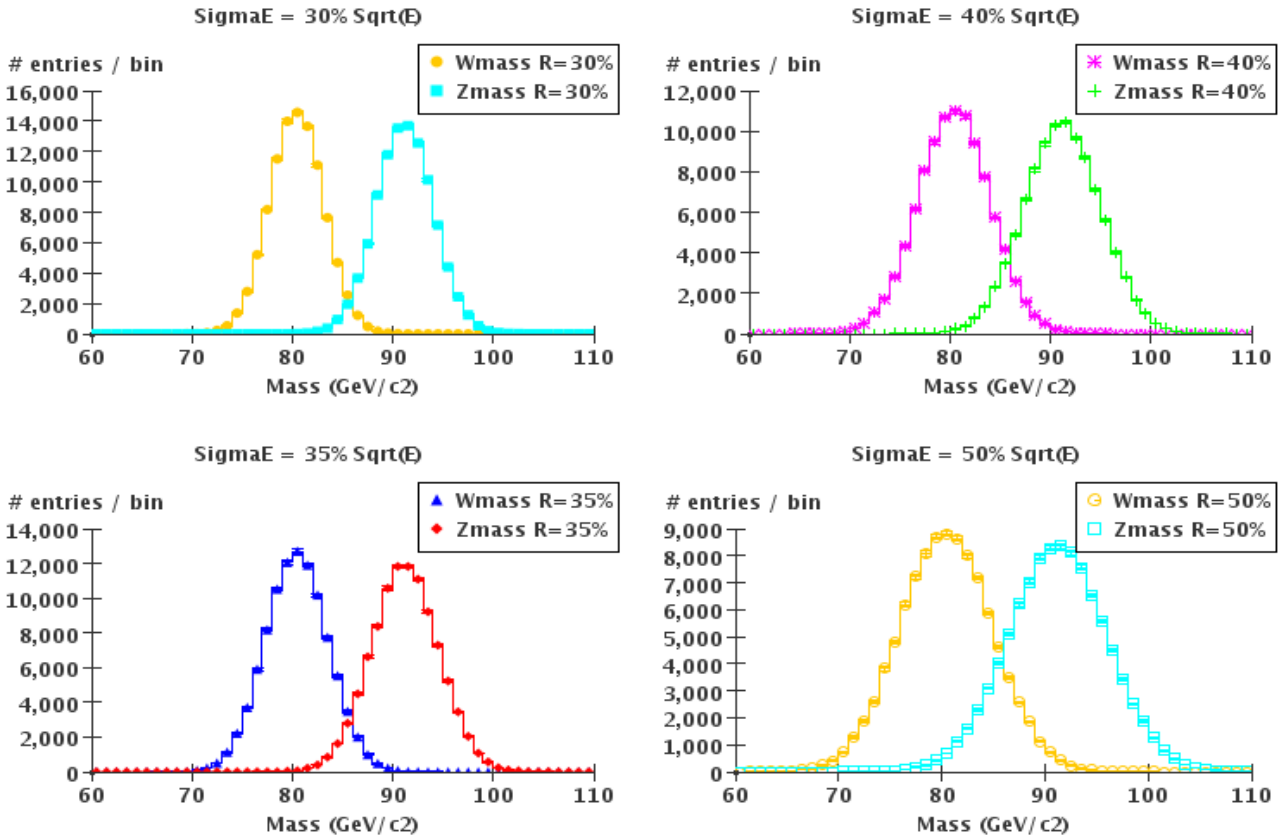
Jet resolution goals

- Physics: ability to resolve hadronic decays of Z and W bosons

Z-mass resolution goal:

$$\sigma_E / E = 30\% / \sqrt{E}$$

Most promising path seems to be through Particle Flow Algorithms (PFA).



PFA in a nutshell

- Basic idea: use high-granularity calorimeters to optimize the reconstruction of individual particles in every event (energy resolution)
- How?
  - Identify clusters from charged particles, replace them with track momentum (track extrapolation, MIP tracking in calorimeter, clustering)
  - Identify photons and optimize their energy resolution (clustering, EM-shower shape, optimize ECal resolution)
  - Neutral hadrons: remaining calorimeter hits/clusters

$$\sigma_E^2(\text{jet}) = \sigma_E^2(\text{charged}) + \sigma_E^2(\text{photons}) + \sigma_E^2(\text{neutrals}) + \sigma_E^2(\text{confusion})$$

negligible

detector dependent
(PFA used for design)

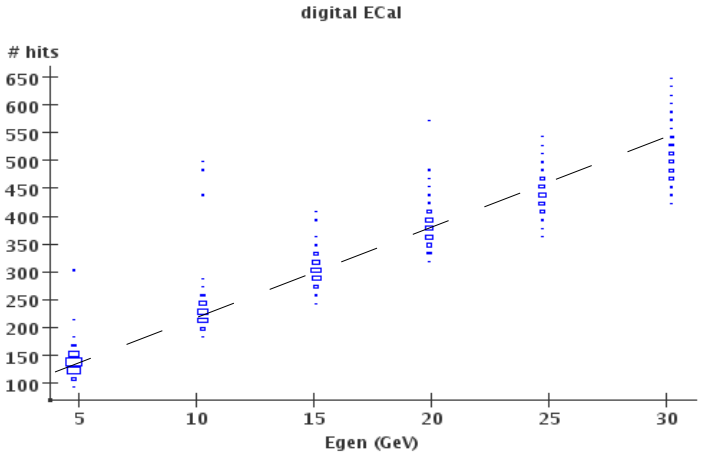
negligible for
a perfect PFA

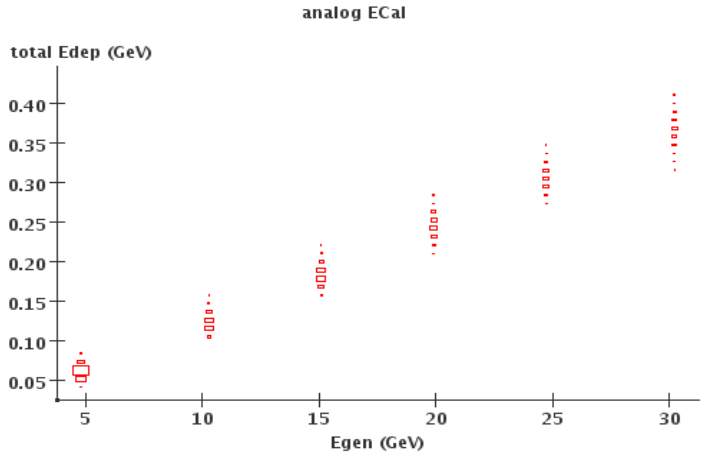
Perfect PFA: ingredients

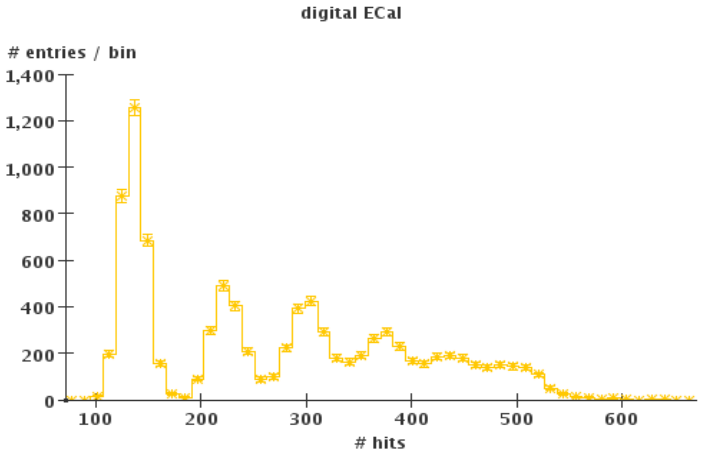
- Geometry: sidaug05_tcmt
  - ECal: 30 Si-W layers, each 3.75mm-thick (0.72 X_0), non-projective, 5x5mm^2 cells
  - HCal: 34 Sci-steel layers, each 28mm-thick (0.13 λ_I), non-projective, 10x10mm^2 cells
  - TCMT: 48 Sci-steel layers, each 28mm-thick, non-projective, 30x30mm^2 cells

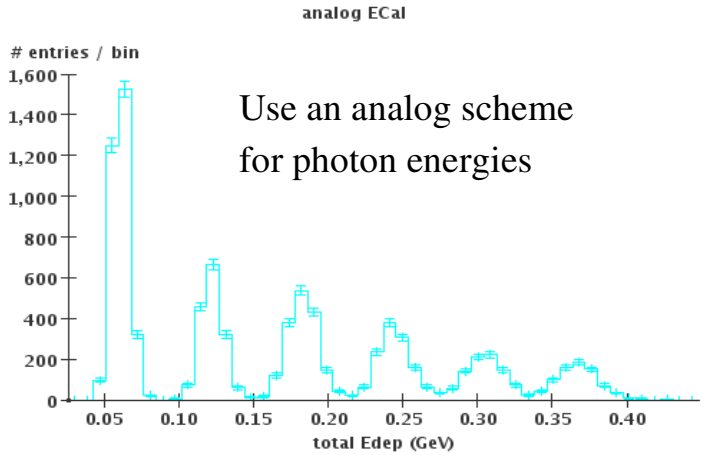
- Perfect clustering: uses all hits from each final state particle generated, no matter where the hits are located (far-flying hadronic debris are common)
- Energy reconstruction and corrections
  - Use smeared MC particles for tracks (tracking not yet available)
  - Neutral particle type: analog for photons, digital for hadrons
  - Good particle ID: use pion mass for all charged hadrons

Photons: analog vs. digital calorimeter



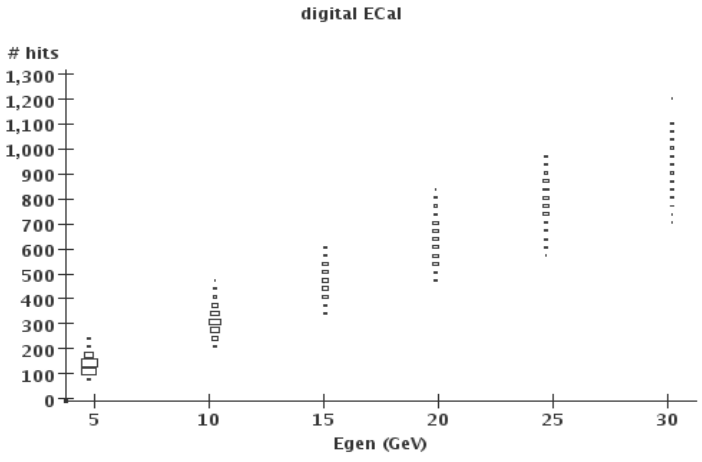


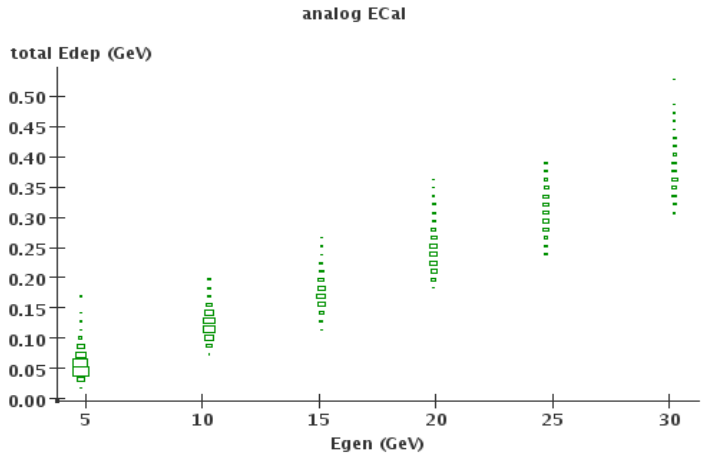


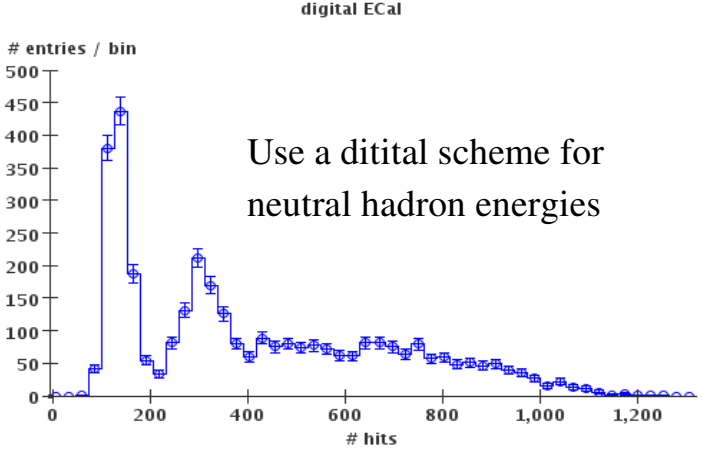


Use an analog scheme
for photon energies

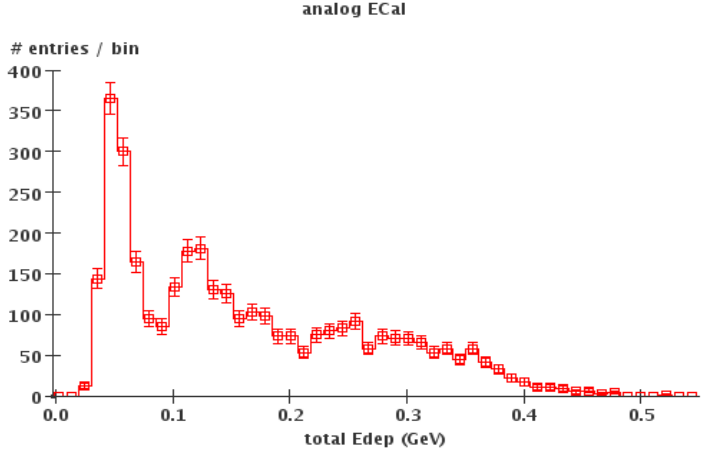
Neutrons: analog vs. digital calorimeter





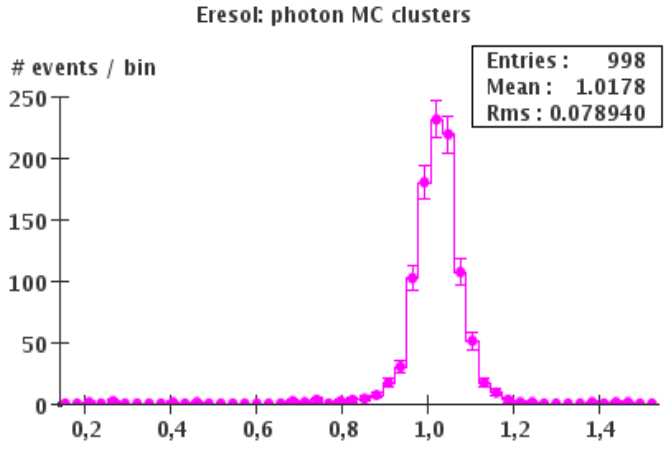


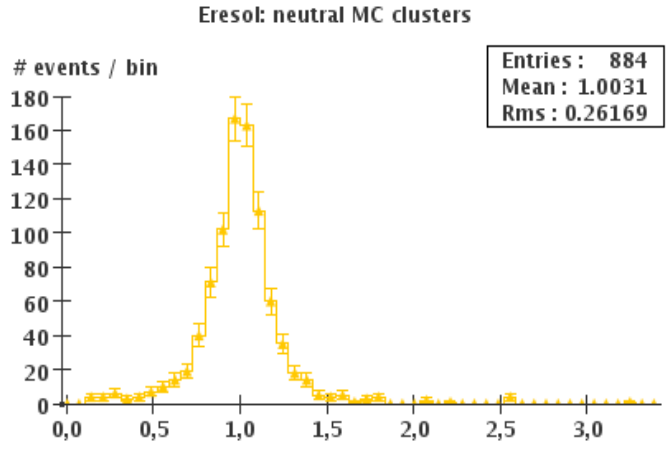
Use a ditital scheme for neutral hadron energies

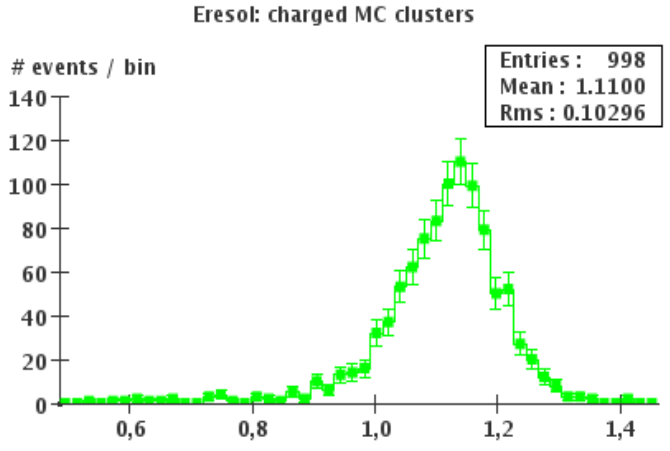


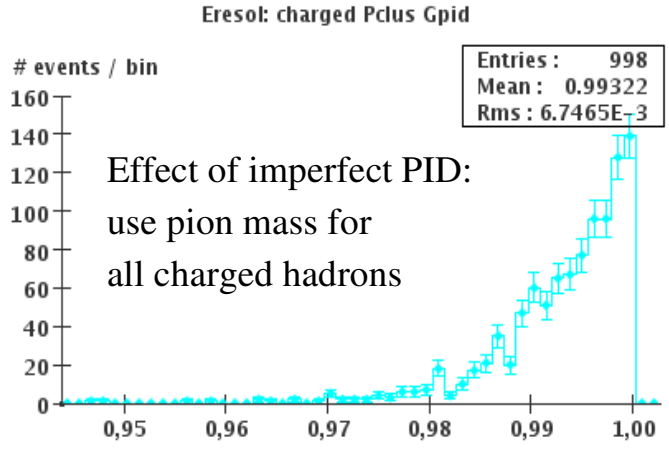
Perfect PFA

- Perfect clusters: all hits coming from each generated Final State particle
- Good (rather than perfect) Particle ID: use pion mass for all charged hadrons
- Calibrated by global particle type contributions



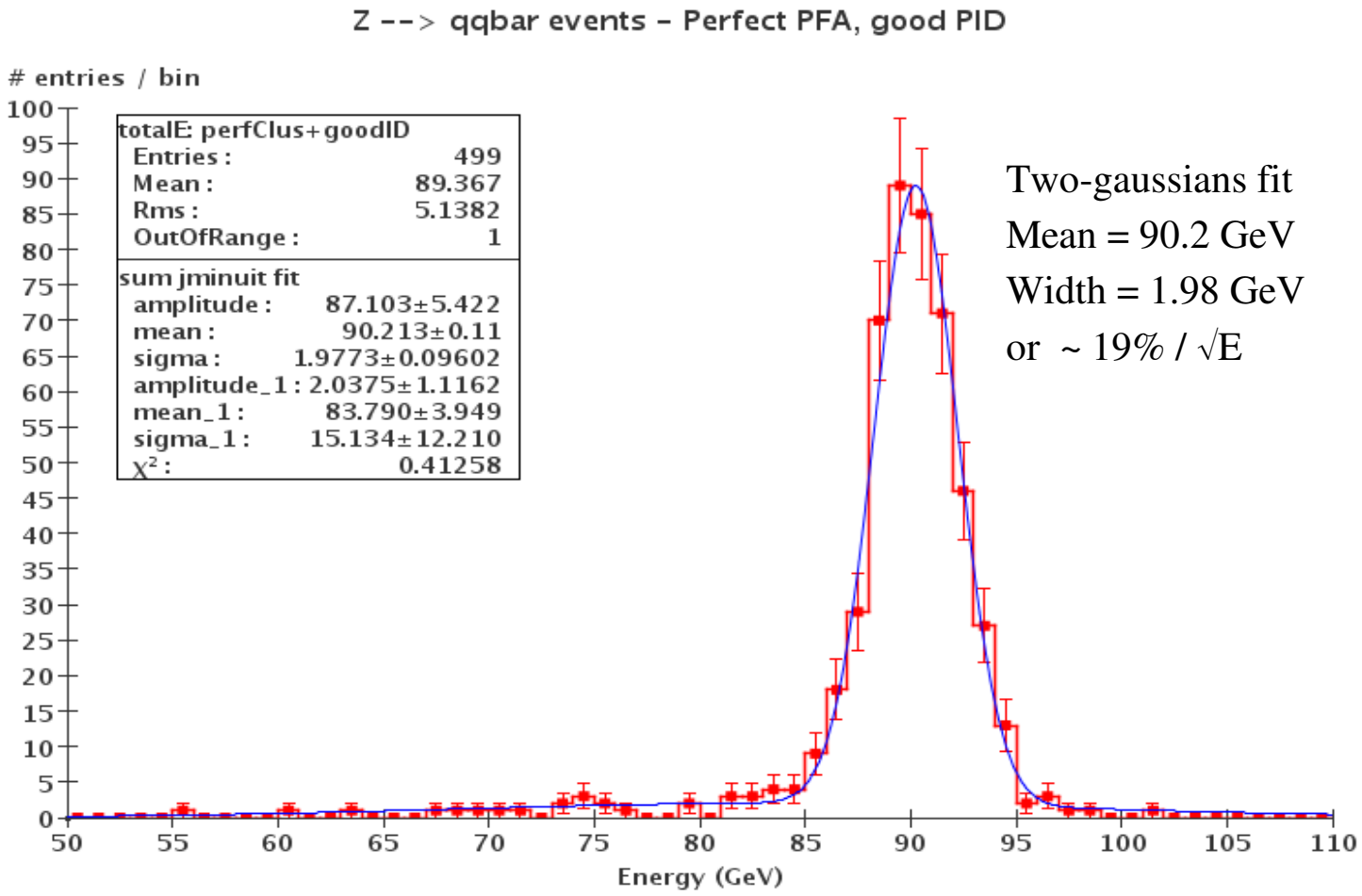






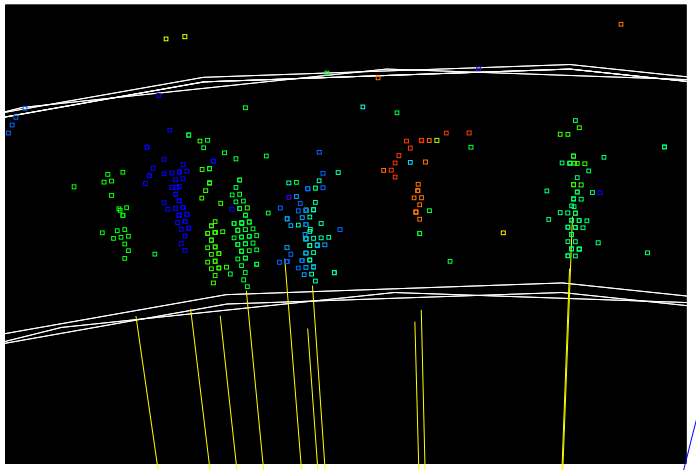
Perfect PFA: the ultimate, unattainable goal

- Perfect clusters: all hits coming from each generated Final State particle
- Good PID: use pion masses for all charged hadrons (incomplete Particle-ID)
- Calibrated by global particle type contributions
- Low-E tail is partially due to energy lost through beam pipe
- Further improvements are likely by using a more sophisticated calibration (angular and longitudinal corrections)



Two-gaussians fit
Mean = 90.2 GeV
Width = 1.98 GeV
or ~ 19% / √E

Directed Tree Clustering Algorithm



- Cal-only clustering developed at NIU (V.Zutshi):
  - density neighborhood (fixed, used to find hit densities D_i)
  - clustering neighborhood (adaptive, based on hit's density)
  - density gradient for cells i,j (j in the neighborhood of i)
    -->hit-density difference divided by distance d_{ij}:

    $$D_{ij} = (D_j - D_i)/d_{ij}$$

  - each cell attaches itself to the hit j with maximum hit-density gradient in its clustering neighborhood
  - Cells with local density maxima become cluster seeds (or directed tree roots)
- Hit selection: $E > E_{MIP} / 4$, and time < 100ns (applied before the clustering)

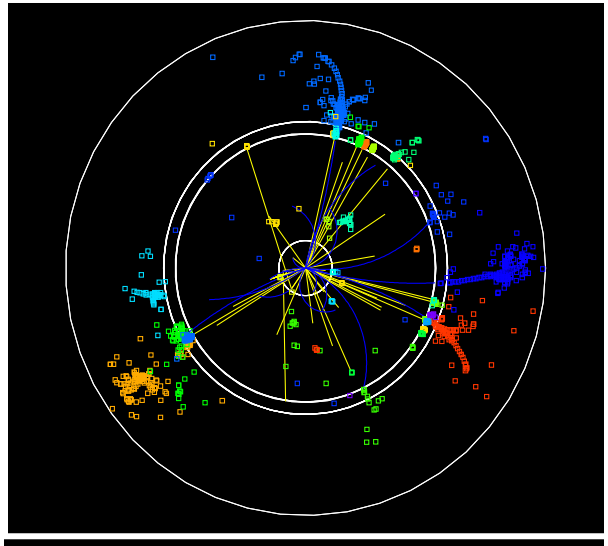
Developing a realistic Particle Flow Algorithm

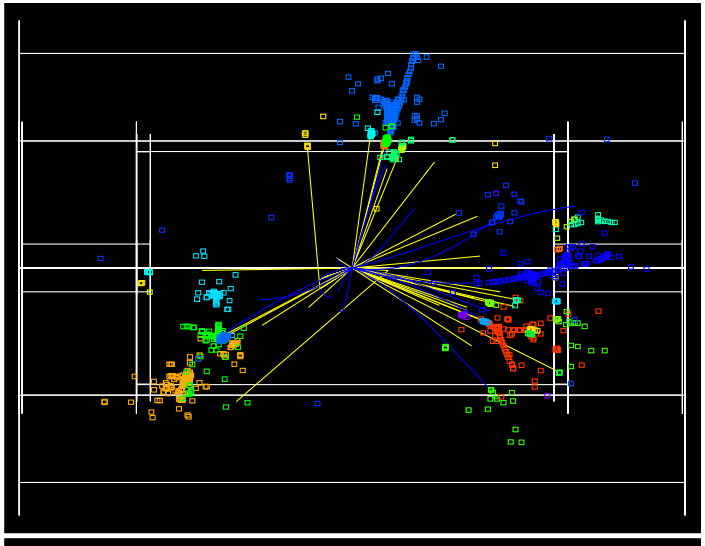
- Development based on:
  - Z --> q $\bar{q}$ (light quarks) on sidaug05_tcmt geometry
  - Data sample generation: SLIC v1.13 + Geant4 v8.0
  - Java-based framework (org.lcsim)

- Algorithm description:
  - Inputs: Directed tree clusters in ECal and HCal, reconstructed tracks
  - Track extrapolations --> track-cluster associations --> seeds for charged clusters
  - Photon-ID: use an Hmatrix (longitudinal cluster profile) --> seeds for photon clusters
  - Shower pattern recognition: algorithms to merge clusters based on cluster shapes and distances
  - Remaining clusters: sorted by size/energy --> seeds for neutral hadron clusters
  - Low multiplicity fragments are discarded (reduce confusion, but degrades final resolution)

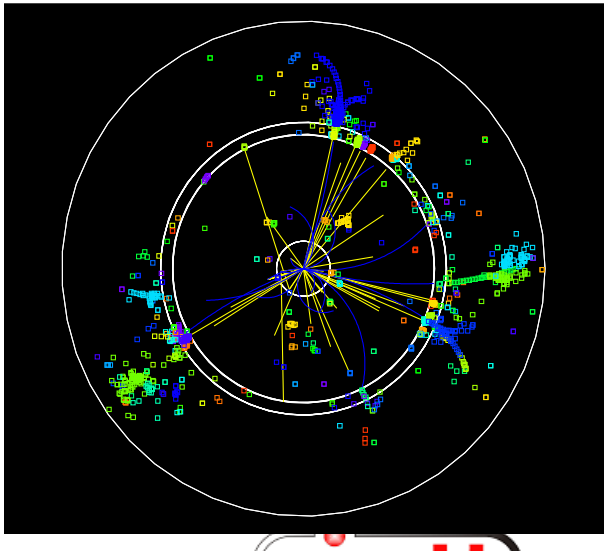
Z --> qqbar (uds) - Directed tree clusters

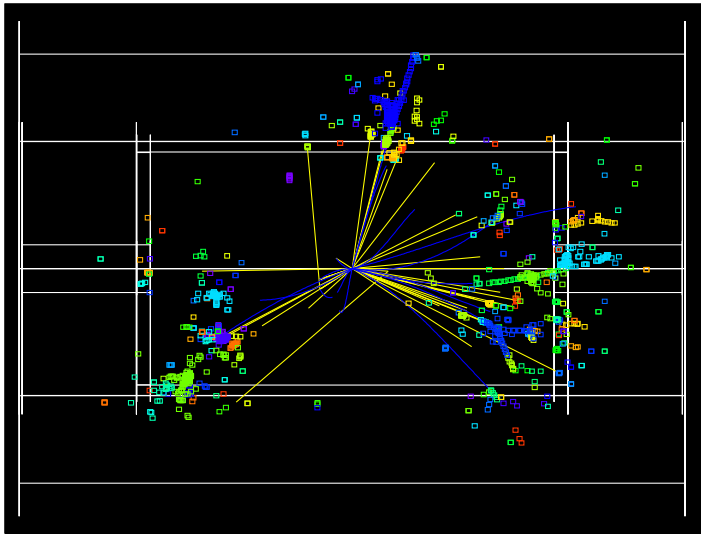
All generated clusters





All directed tree clusters

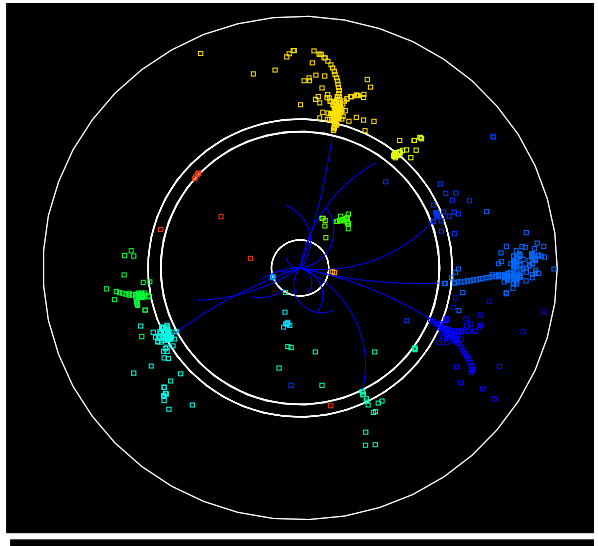


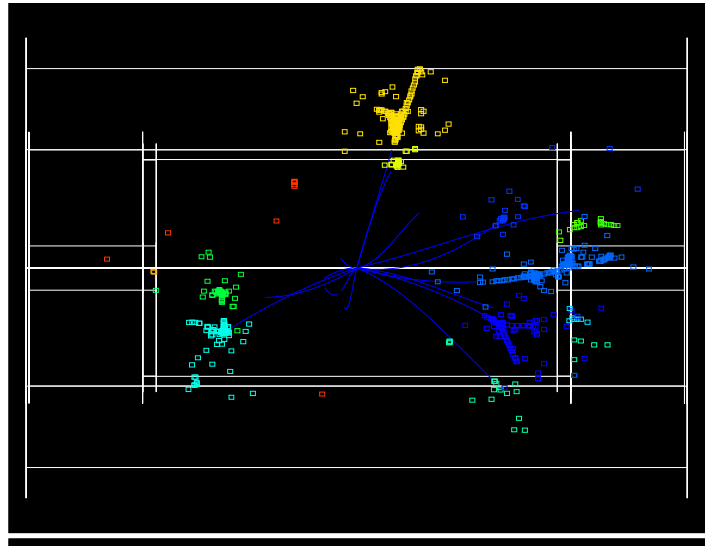


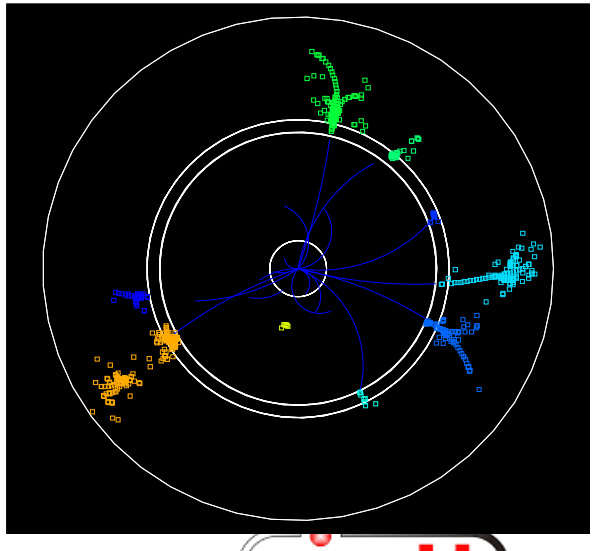
Clusters from charged particles

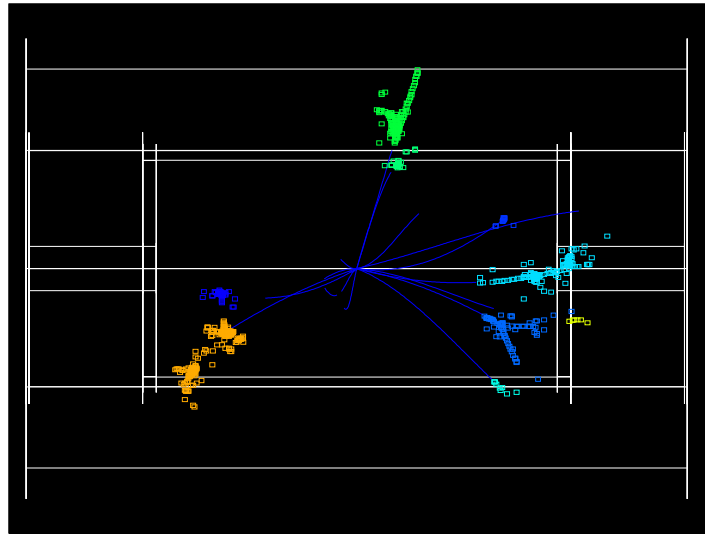
Generated charged clusters

Reconstructed charged clusters





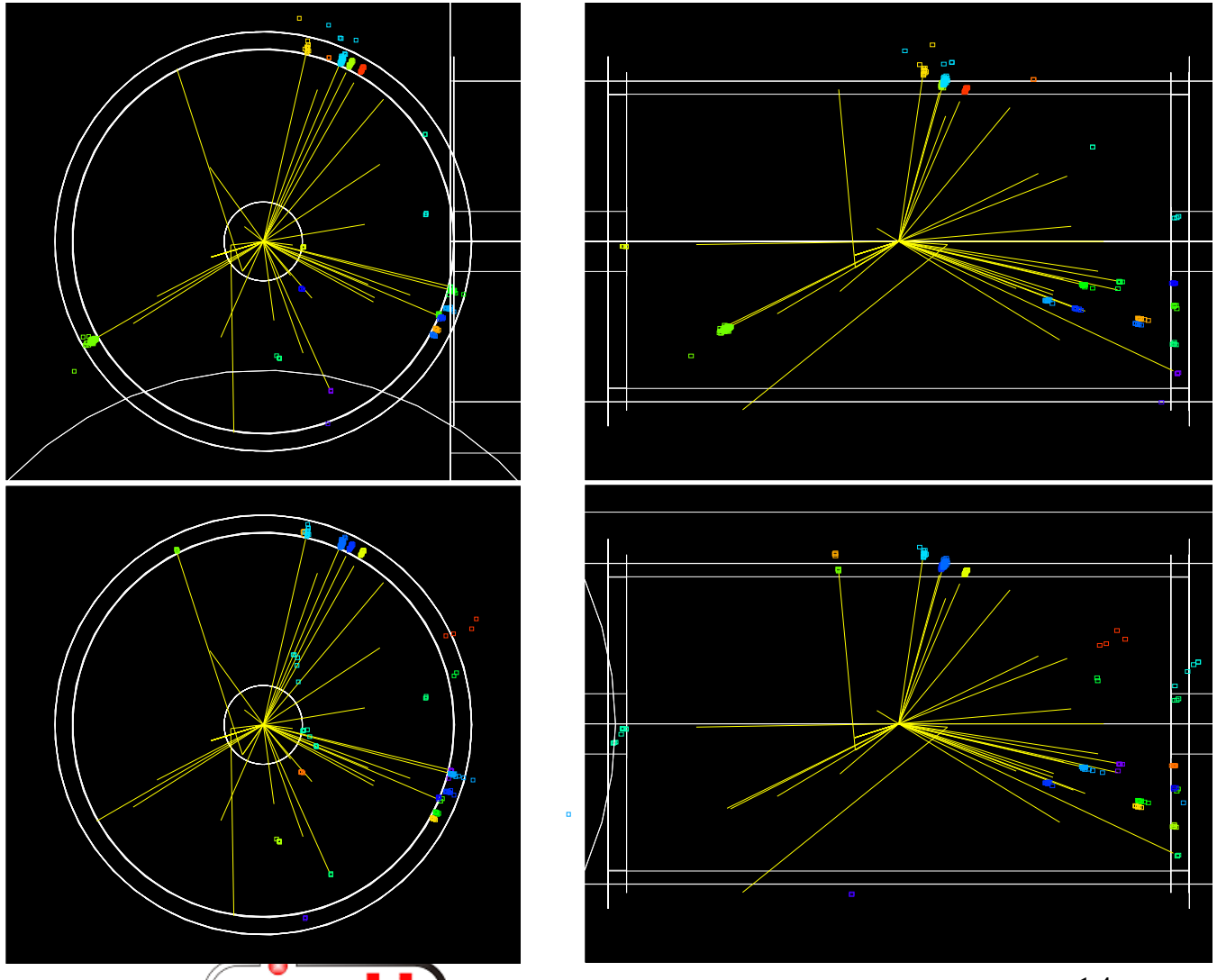




Clusters from photons

Generated photon clusters

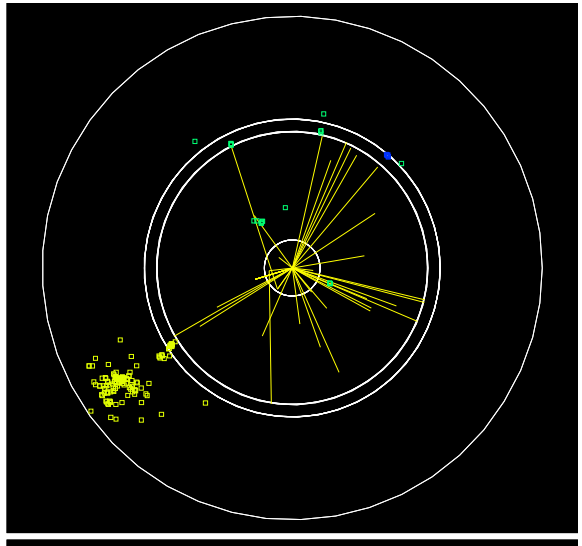
Reconstructed photon clusters

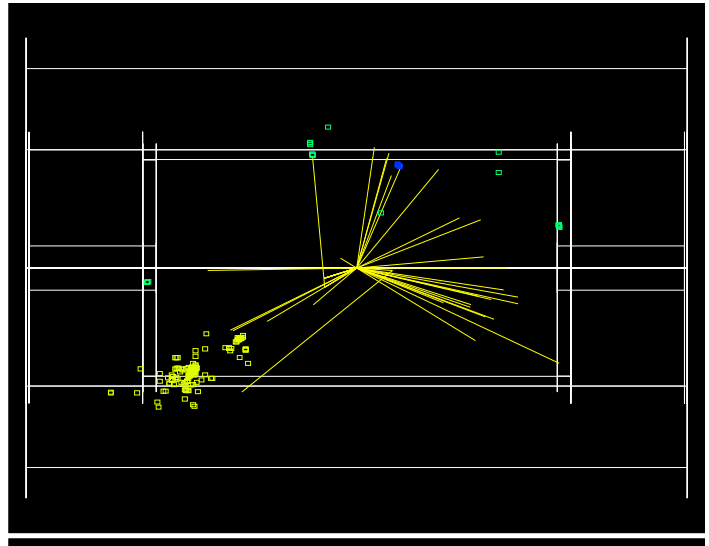


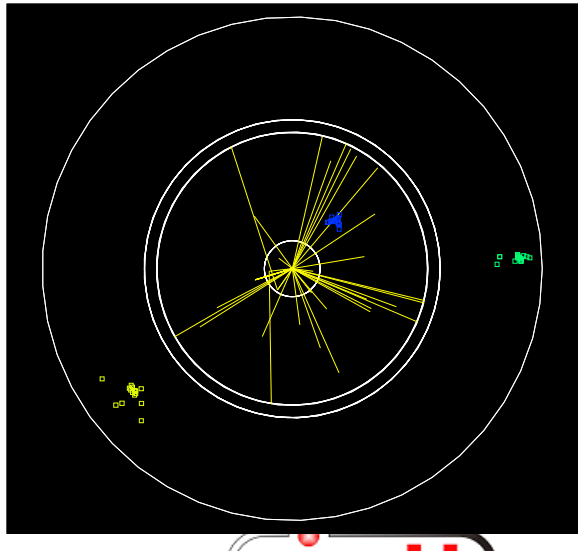
Clusters from neutral hadrons

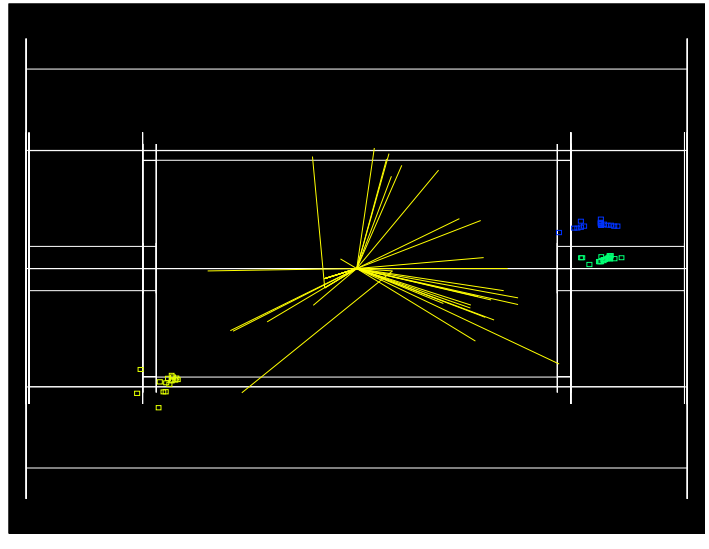
Generated neutral hadron clusters

Reconstructed neutral hadron clusters

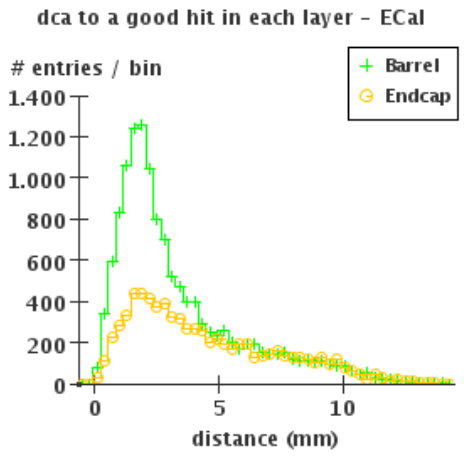


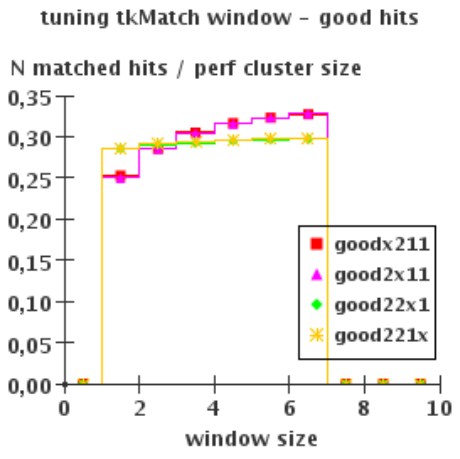


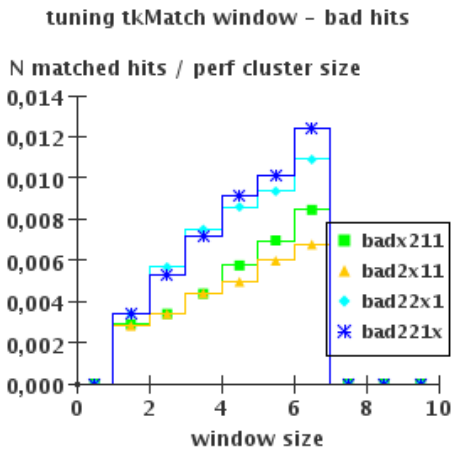


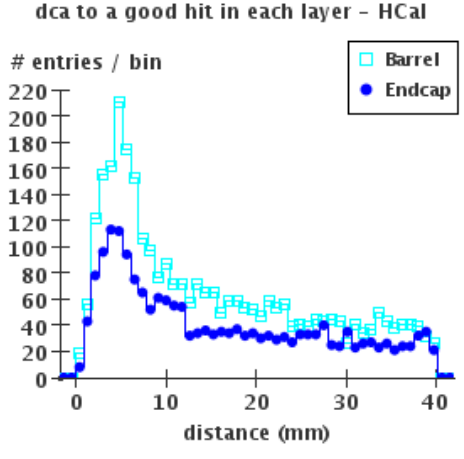


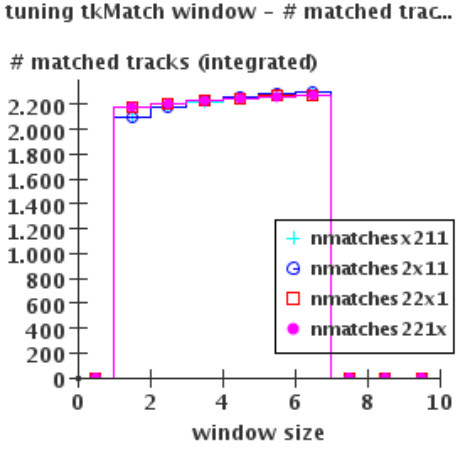
Optimizing the track matching window







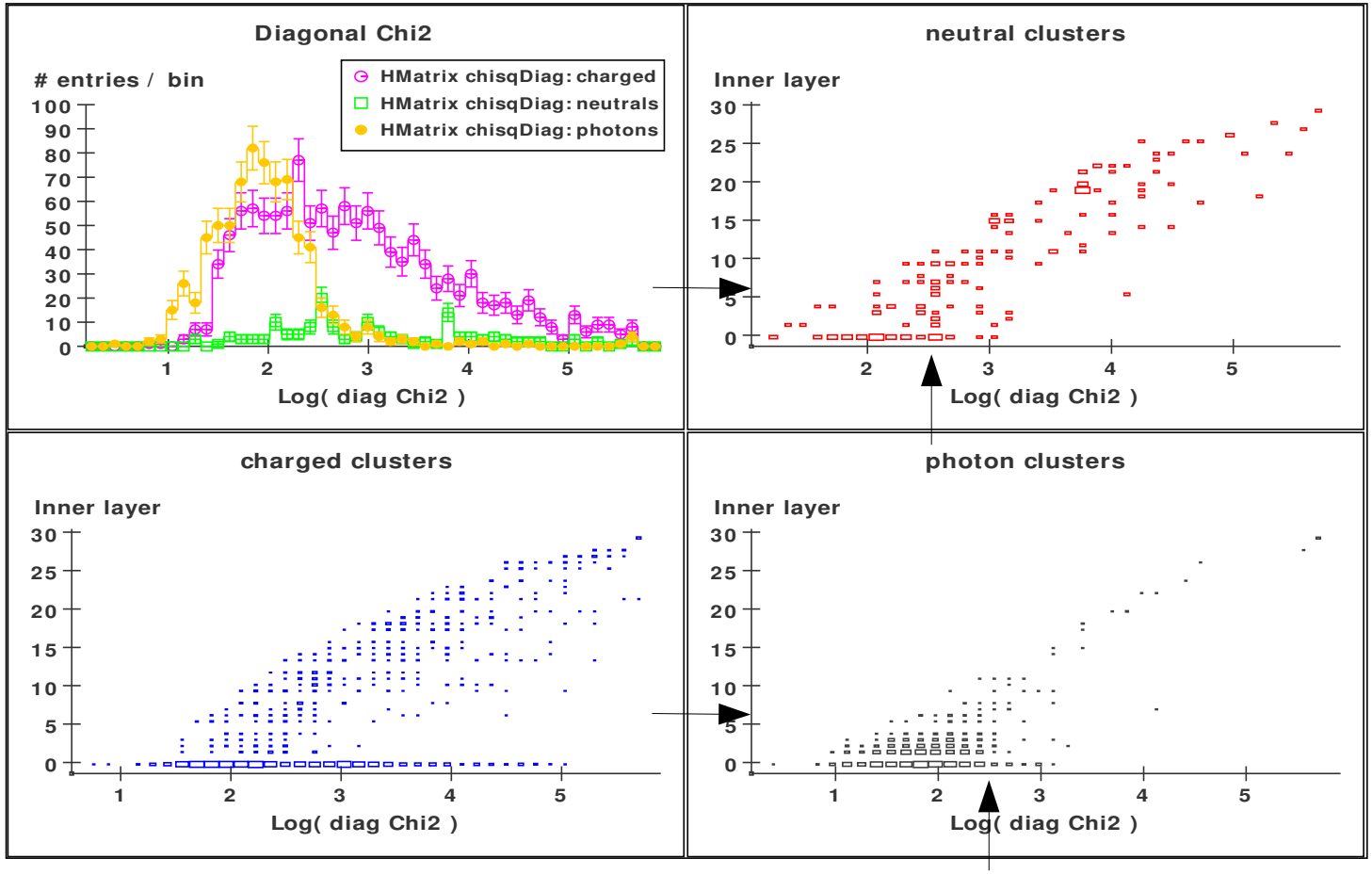




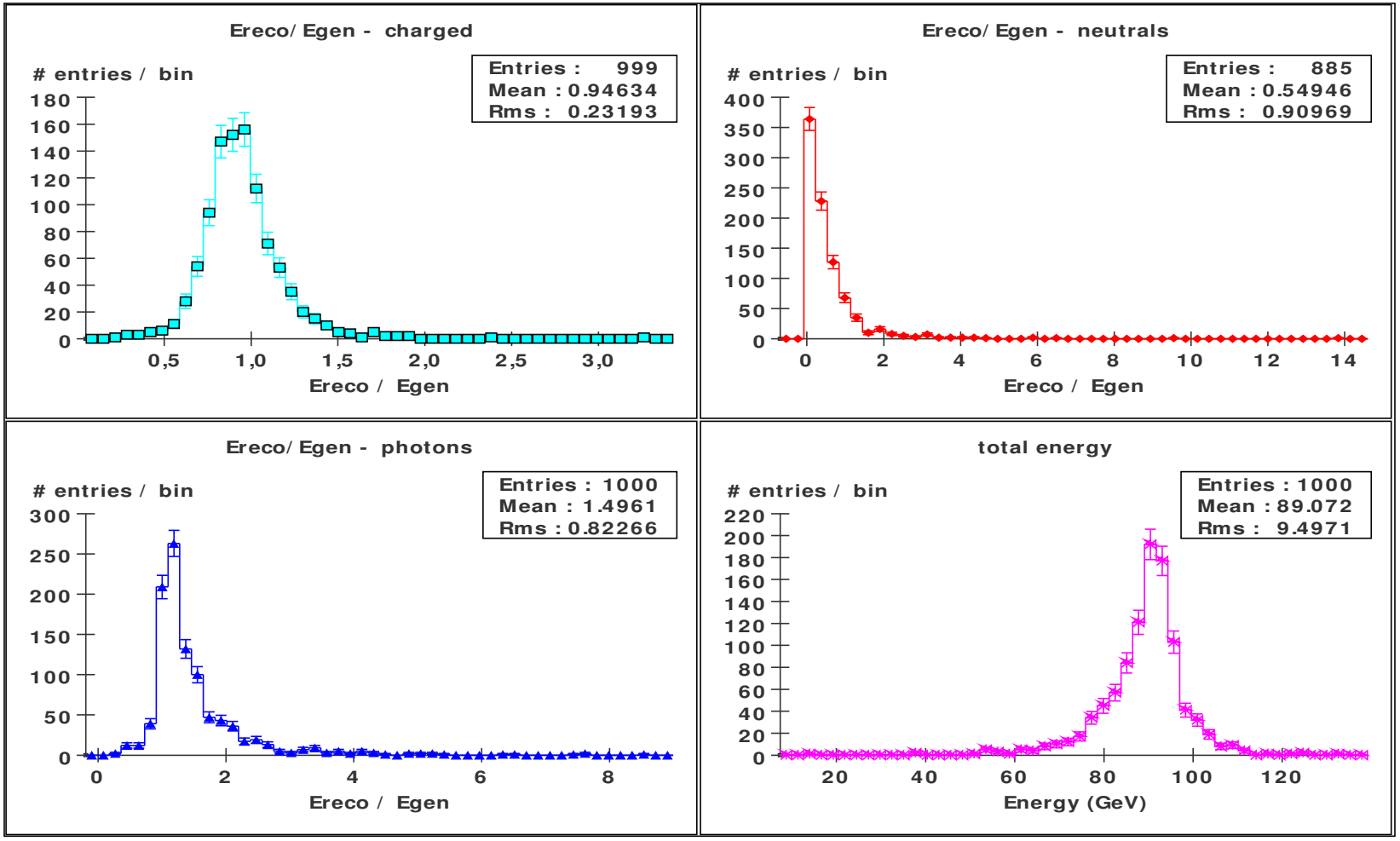
- A +/-1 window seems to be enough for good track matching in Ecal, but going to +/-2 still gets more good than bad hits
- Track matching window selected:
  +/-2 in ECal (each unit --> 5mm)
  +/-1 in HCal (each unit --> 10mm)

Photon-ID: longitudinal H-Matrix

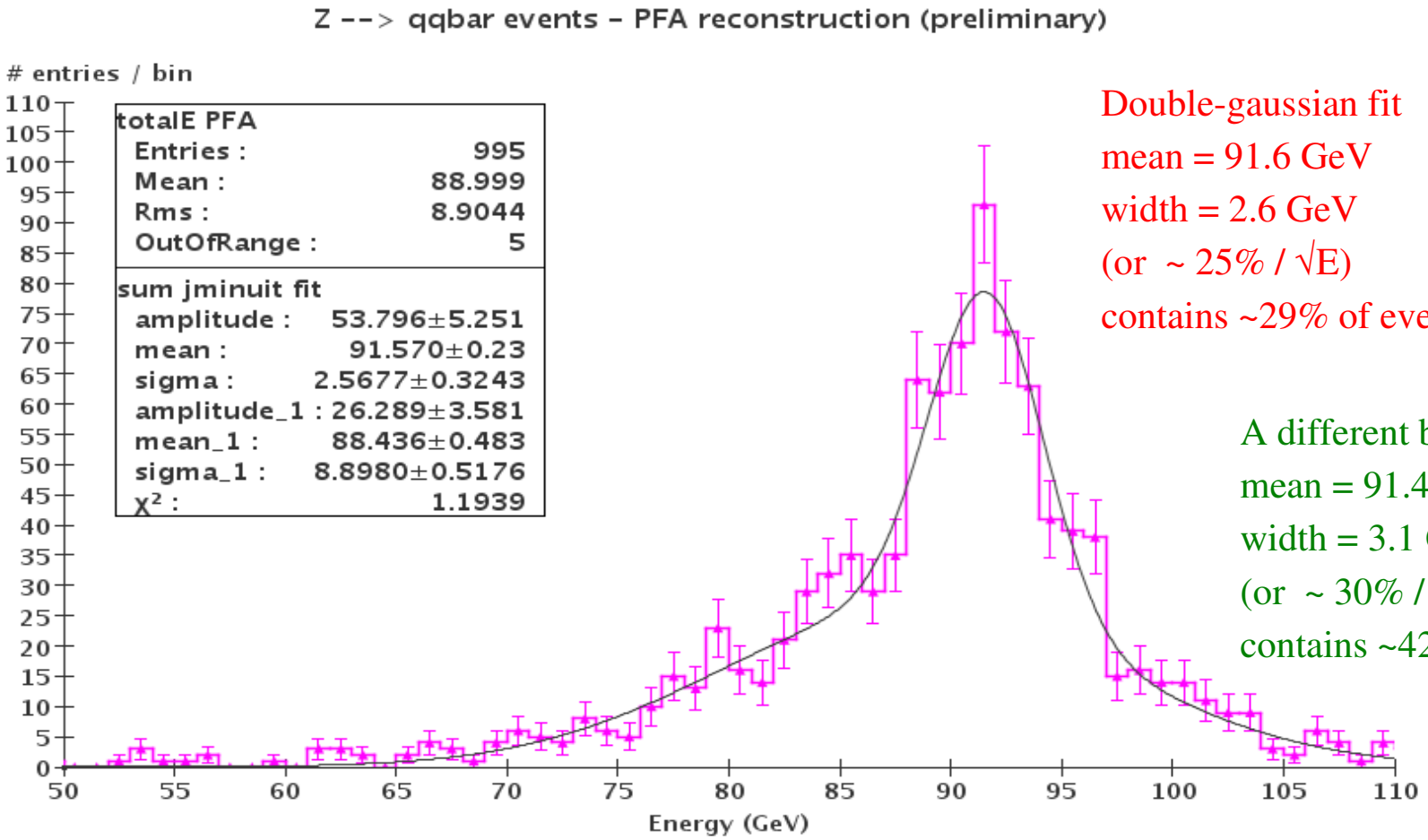
- H-Matrix is not yet fully optimized, so there is room for improvements
- Lowest layer hit in clusters provides further γ - h^0 discrimination
- Photon selection:
  $\log(\chi^2_{diag}) < 2.5$
  lowLayer <= 6



Current PFA result (preliminary)



Current PFA result (preliminary)



Double-gaussian fit
mean = 91.6 GeV
width = 2.6 GeV
(or ~ 25% / √E)
contains ~29% of events

A different binning:
mean = 91.4 GeV
width = 3.1 GeV
(or ~ 30% / √E)
contains ~42% of events

Summary

- All the basic tools needed for a full PFA are in place (ALCPG Java-based framework)
- Big effort to develop code which is independent of geometry
- Preliminary results are encouraging, but a lot of optimization still needed
- Things to do:
  - Investigate origin of misidentified clusters and how to improve cluster identification
  - Use tail catcher information to improve jet energy resolution (important for jetE > 75 GeV)
  - Further calibration corrections (dependency on energy, particle type, incidence angle and interaction layer)
  - Comparisons to other people's results (standard geometries)
  - Investigate PFA at higher energies and more complex physics processes (WW, ZH, etc)
  - Comparisons for different geometries, B-fields and technologies