PFA reconstruction with directed tree clustering

Dhiman Chakraborty
for the NICADD/NIU software group

NORTHERN ILLINOIS UNIVERSITY

Vancouver LC Workshop
July 21, 2006

Talk outline

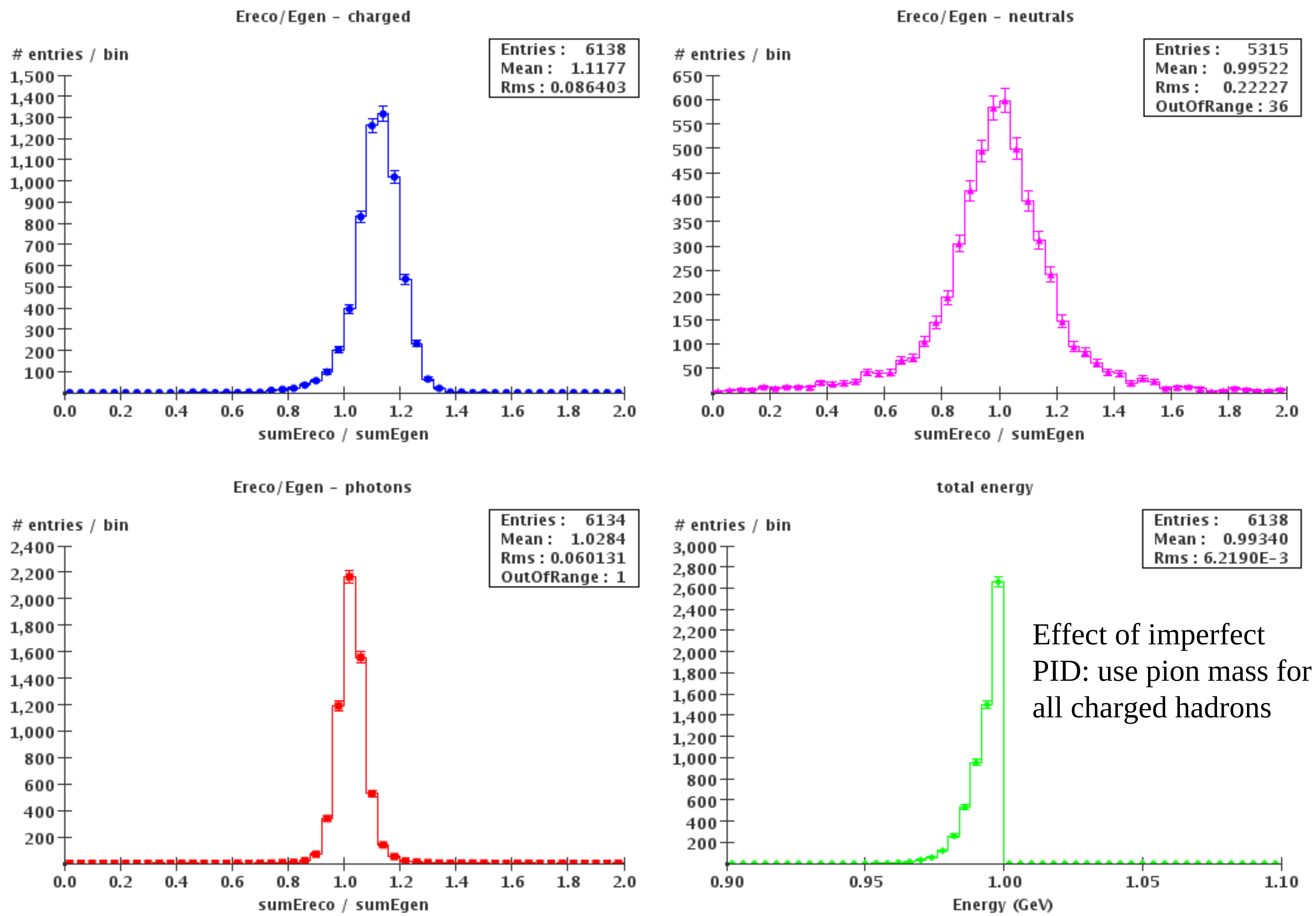
- Perfect PFA
  what resolutions can we expect?
- Real PFA
  how we are doing the real PFA
- Summary
  status and plans

Perfect PFA: ingredients

- Geometry: sidaug05_tcmt
  - ECal: 30 Si-W layers, each 3.75mm-thick (0.72 X_0), non-projective, 4x4mm^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 EM for photons, digital EM+HAD for hadrons
  - Good particle ID: use track momentum + pion mass for all charged hadrons

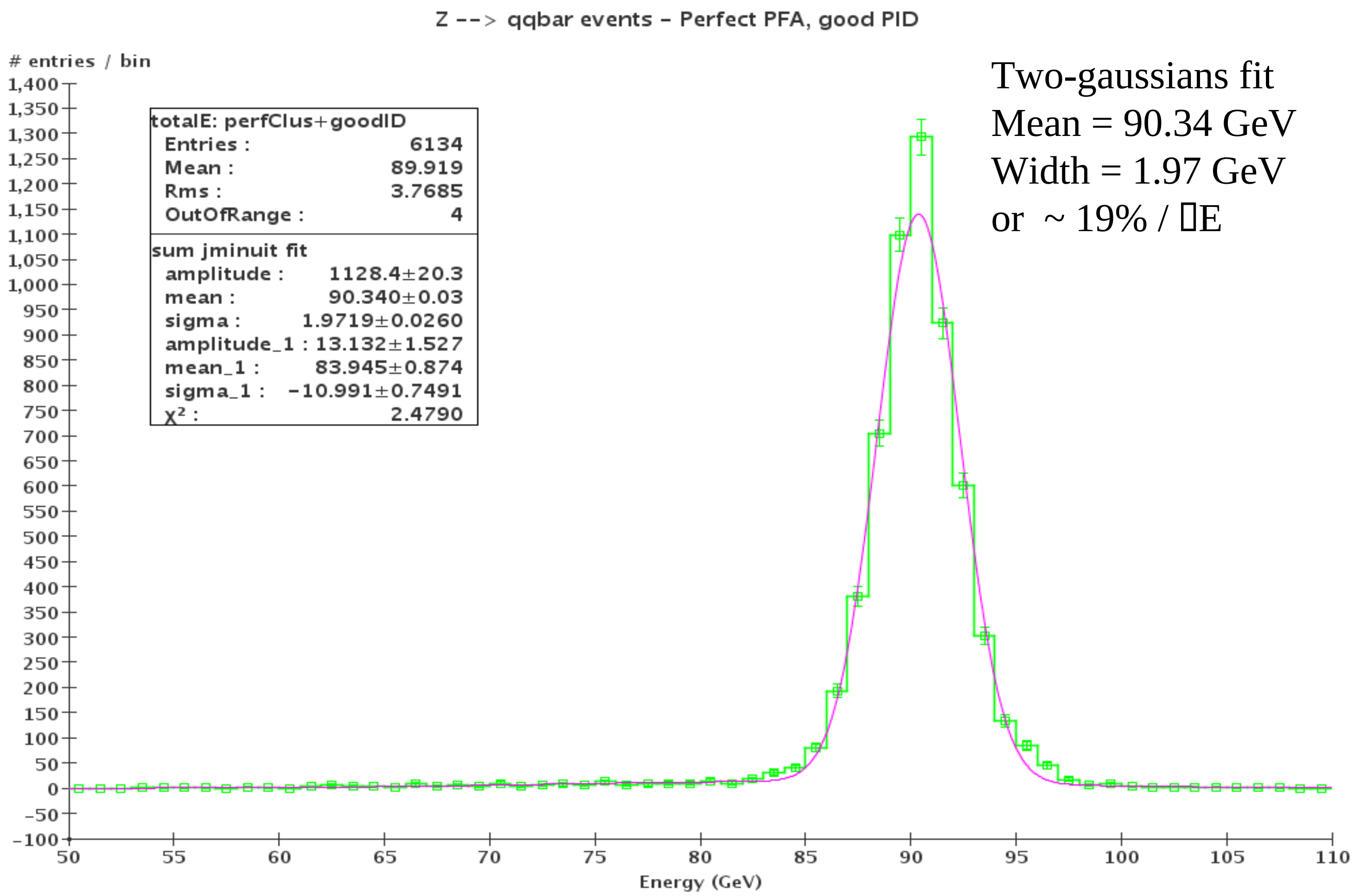
Global calibration

- 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: photon and neutral distributions scaled so that their mean ~1



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
- Barrel events selected
- Further improvements are likely by using a more sophisticated calibration schemes

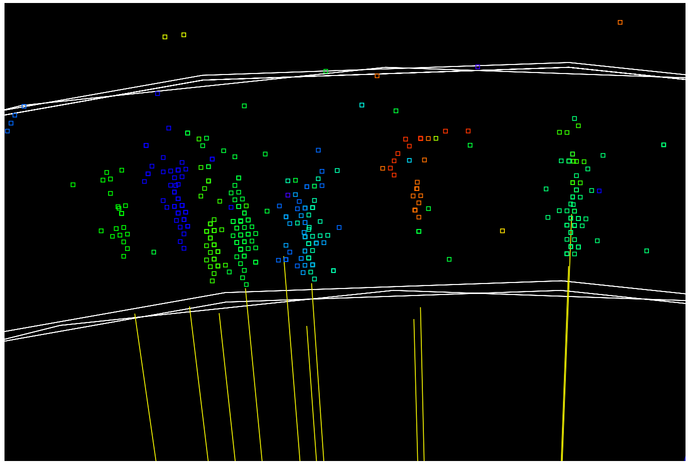


Two-gaussians fit
Mean = 90.34 GeV
Width = 1.97 GeV
or ~ 19% / √E

Perfect PFA: what is missing

- Studies with other geometries (make sure code works for different detector technologies and geometries)
- Include tail catcher hits (?)
- Use more sophisticated calibration schemes (including dependences on incidence angle, particle type, interaction layer, etc.)

Directed Tree Clustering Algorithm



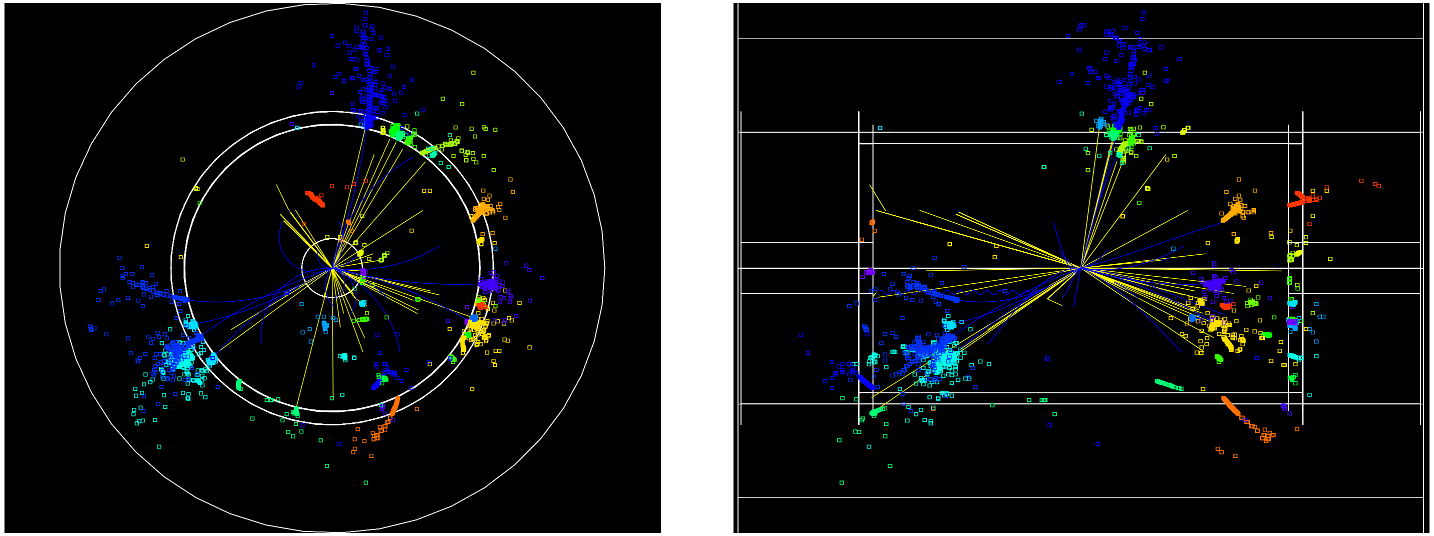
- Cal-only clustering developed at NIU (V.Zutshi):
  - density neighborhood (fixed, used to find hit densities D_i)
  - clustering neighborhood (adaptive, based on hit's density)
  - density gradient for cells i,j (j in the neighborhood of i)
    -->hit-density difference divided by distance d_{ij}: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

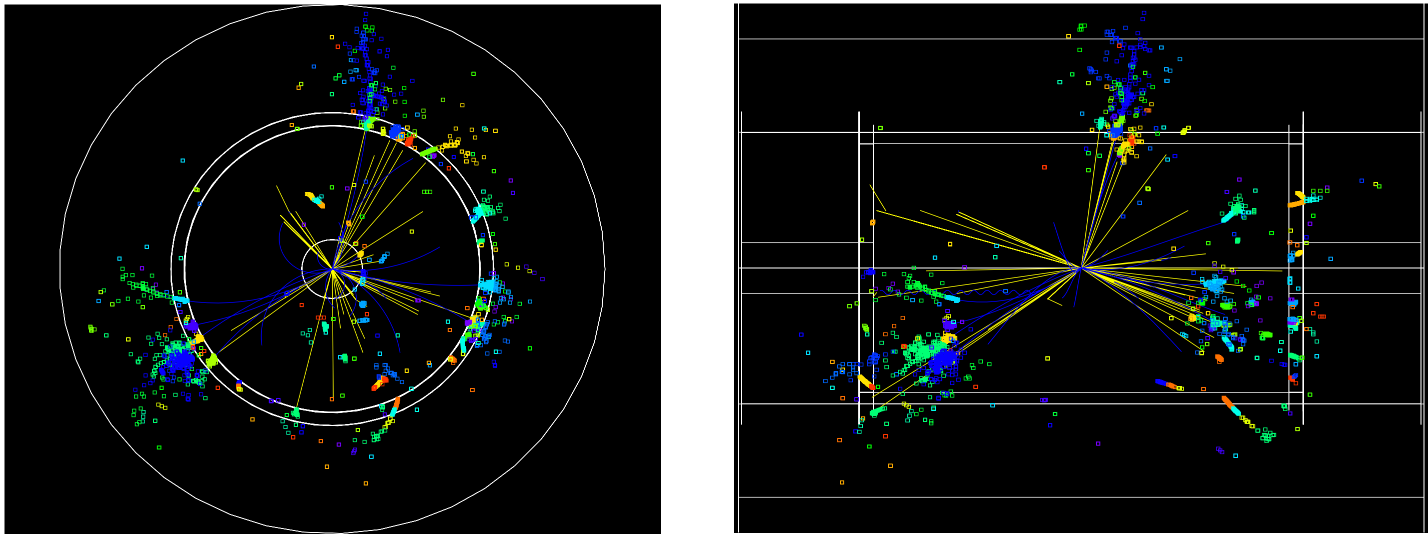
- Development based on:
  - Z --> q q (light quarks) on sidaug05_tcmt geometry
  - Data sample generation: SLIC v2r0p0 + Geant4 v8.0 generated at SLAC
  - Java-based framework (org.lcsim)
- Algorithm description:
  - Inputs: Directed tree clusters in ECal and HCal, reconstructed tracks
  - Track extrapolations --> track-cluster associations --> seeds for charged clusters
  - Photon-ID: use an Hmatrix (longitudinal cluster profile) --> seeds for photon and neutral hadron clusters
  - Shower pattern reco: algorithms to merge clusters based on cluster shapes and distances
  - Remaining clusters: sorted by size/energy --> seeds for extra neutral hadron clusters
  - Low multiplicity fragments are discarded (reduce confusion, but degrades final resolution)

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

All generated clusters

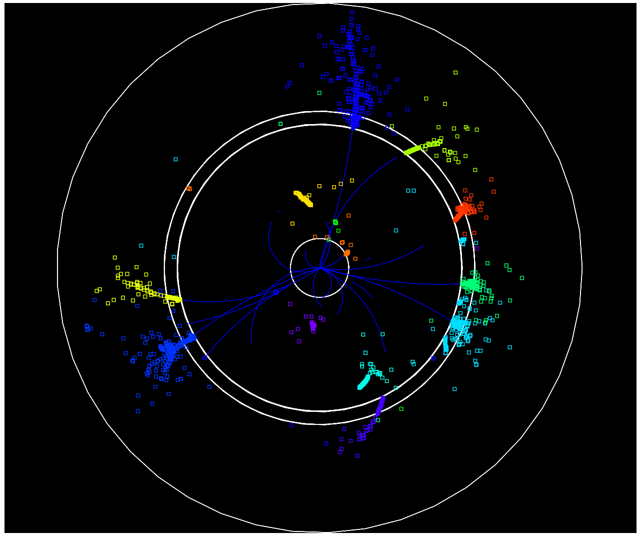


All directed tree clusters

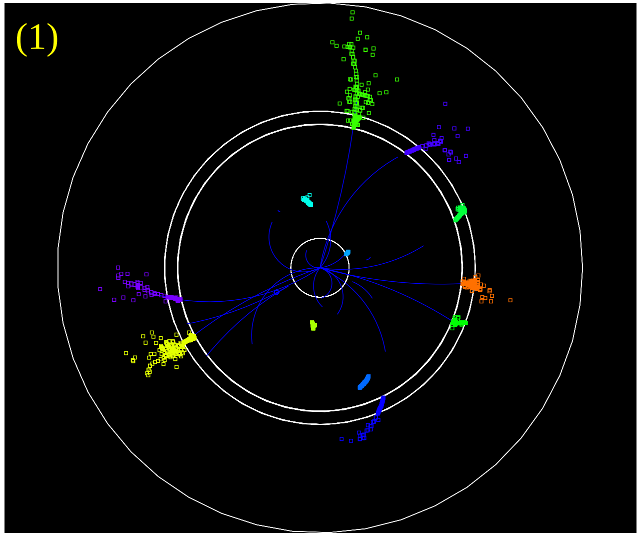


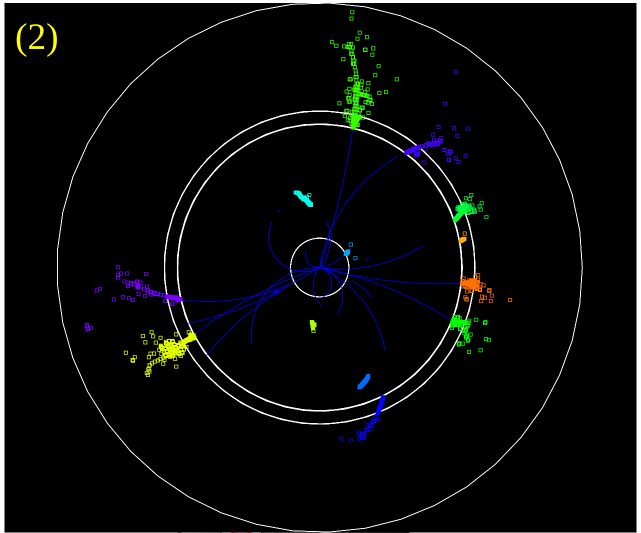
Charged cluster ID

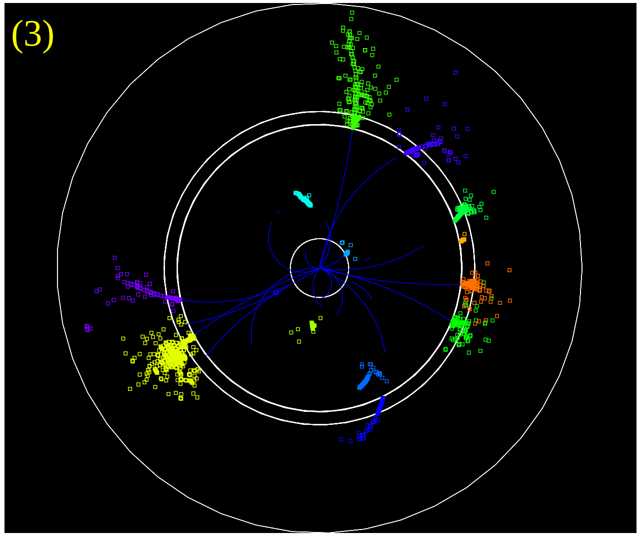
Track matching(1) (swimmer) + fragment association -- based on cluster shapes(2) and distances(3)



In this particular event, hits from a neutral hadron were matched to the track at 8 o'clock. There is a real charged particle there too.

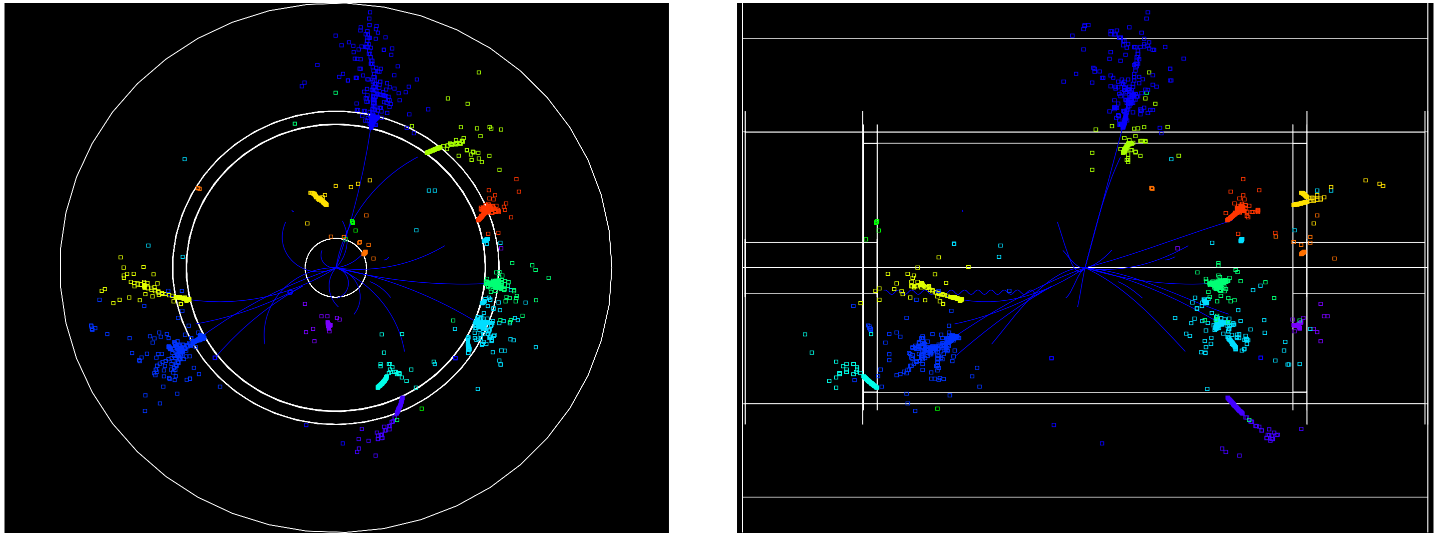




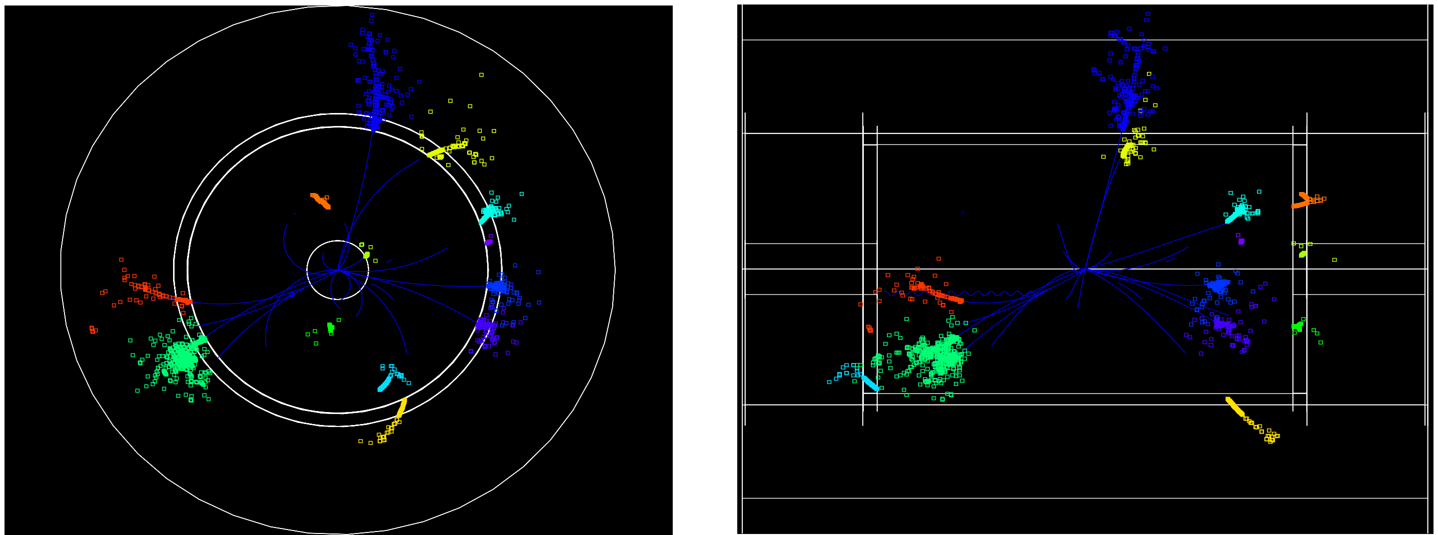


Clusters from charged particles

Generated charged clusters

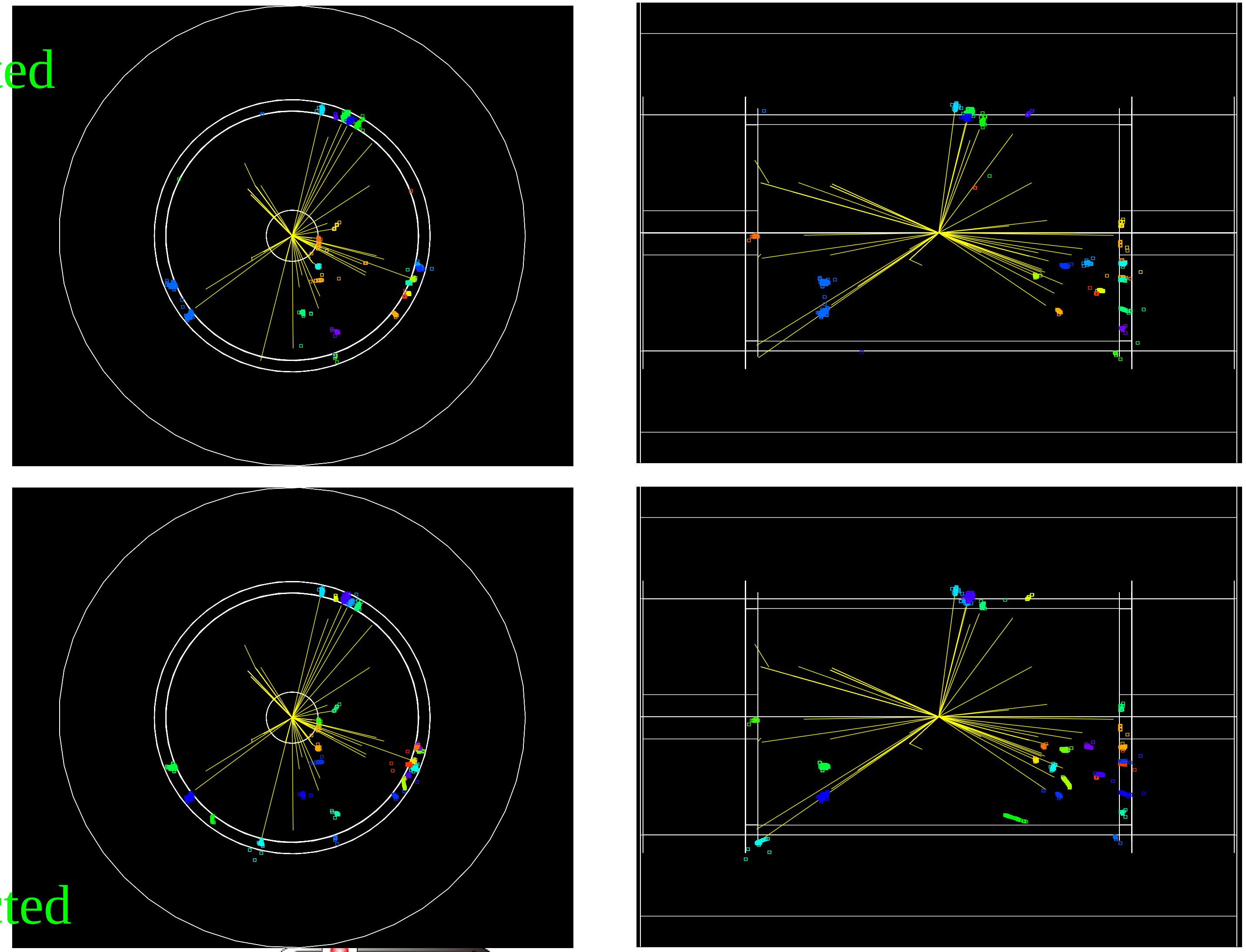


Reconstructed charged clusters



Photon Cluster ID

Generated

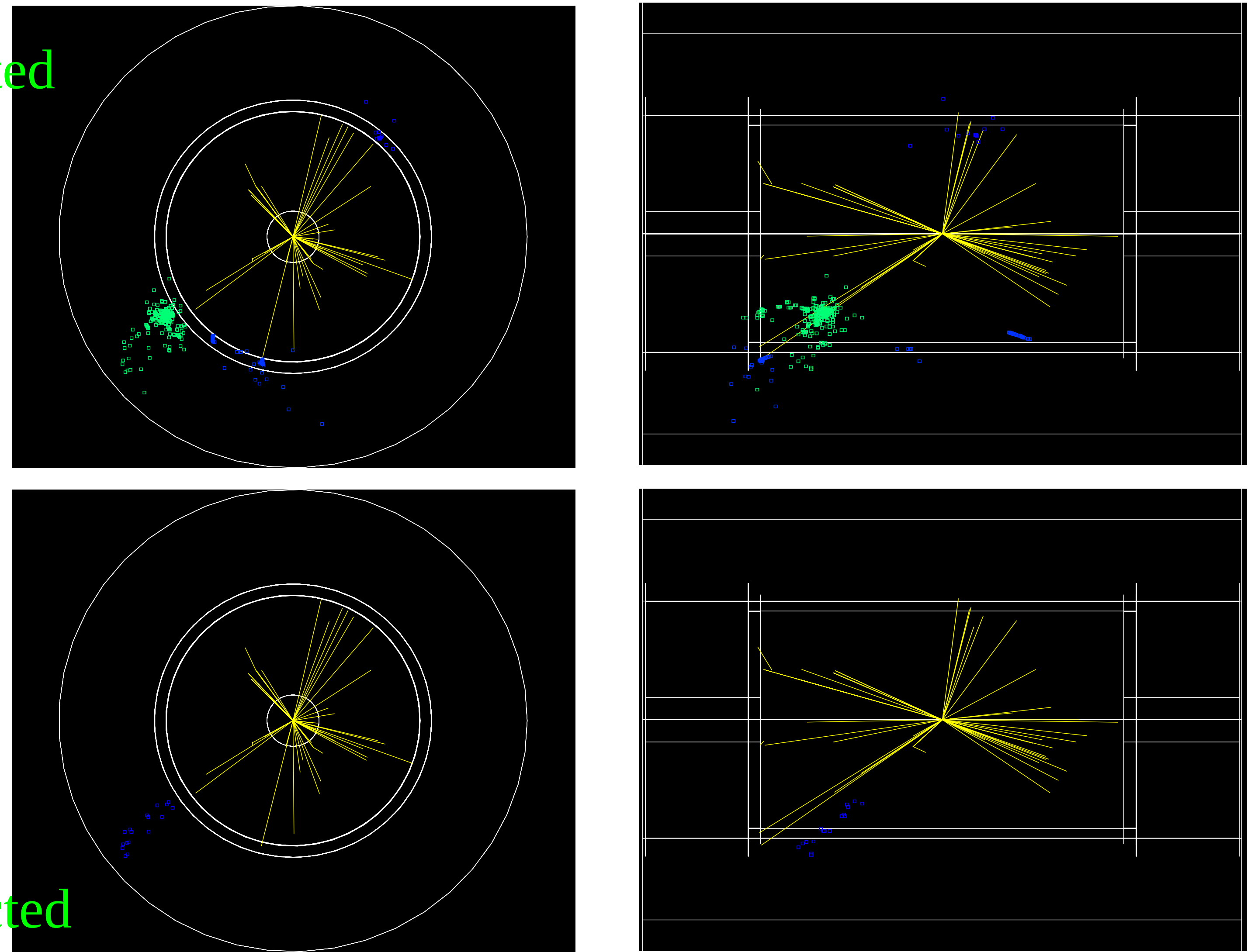


Reconstructed

Photon clusters are identified by a longitudinal H-matrix in association to lowest layer and non-MIP-stub requirements

Clusters from neutral hadrons

Generated



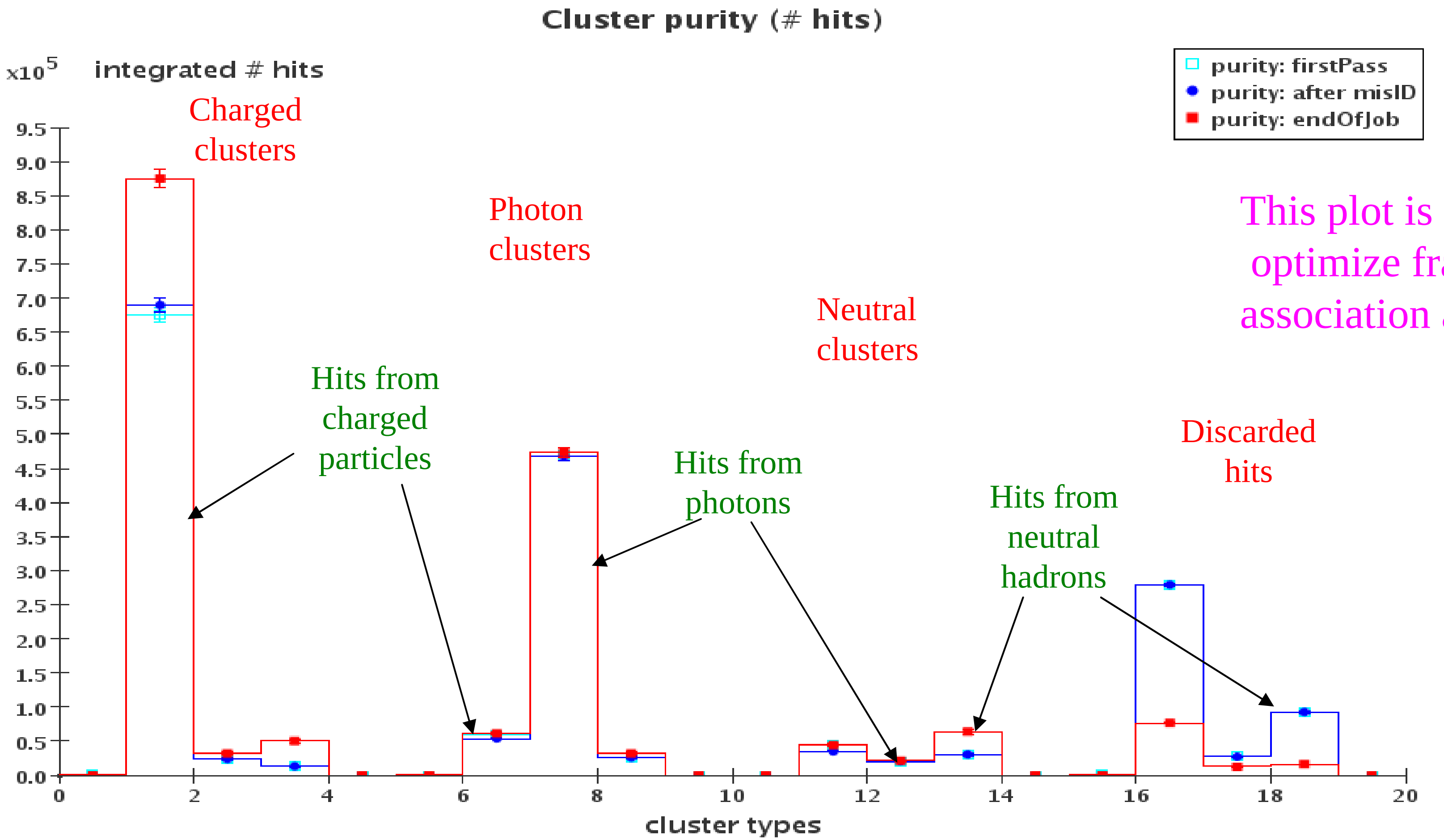
Reconstructed

Whatever is left from charged or photon ID is initially tagged as a neutral hadron cluster.

Fragment association is performed based on cluster shapes and distances.

Further work is needed to optimize the neutral cluster ID.

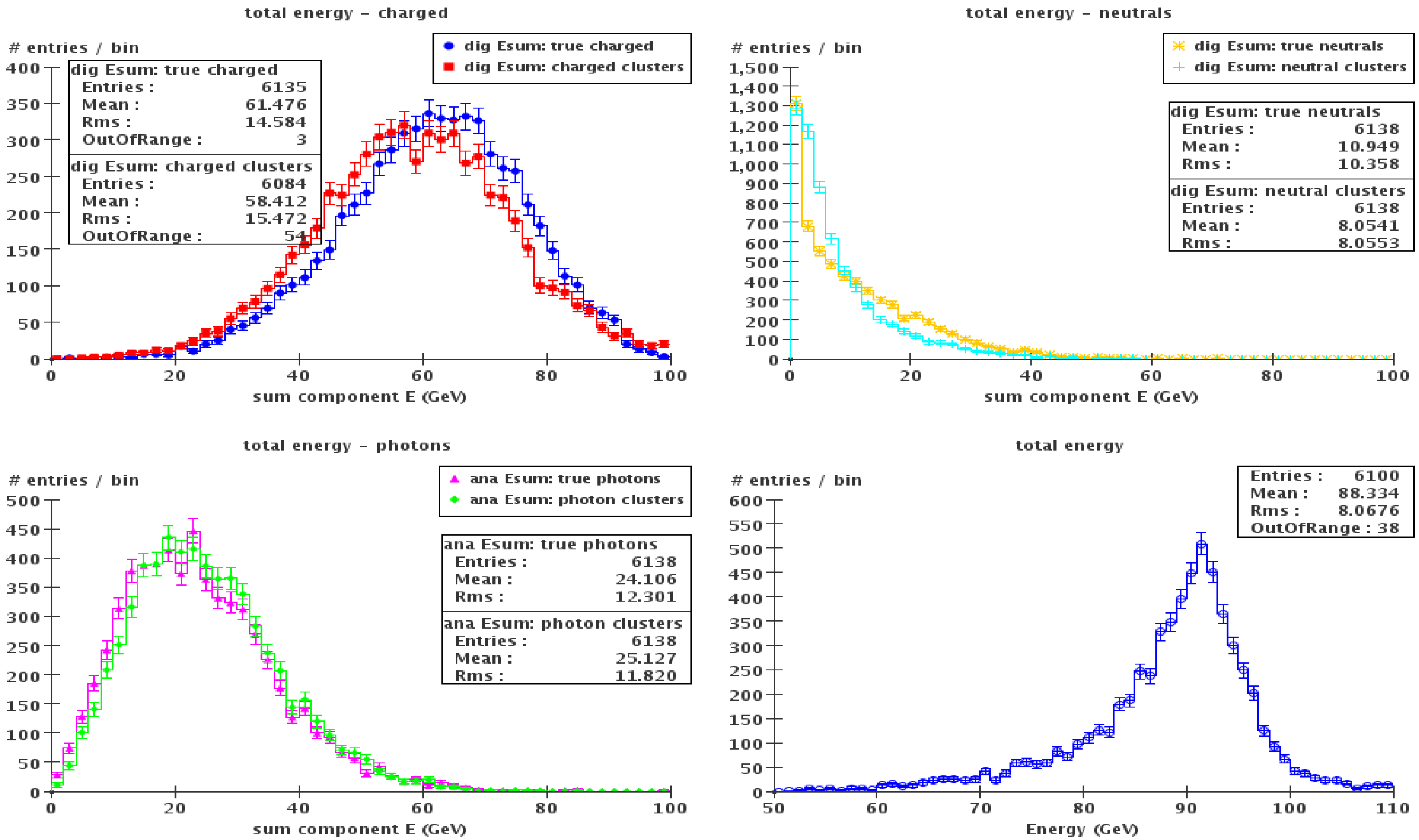
Cluster purity and fragment association



This plot is helpful to optimize fragment association algorithms

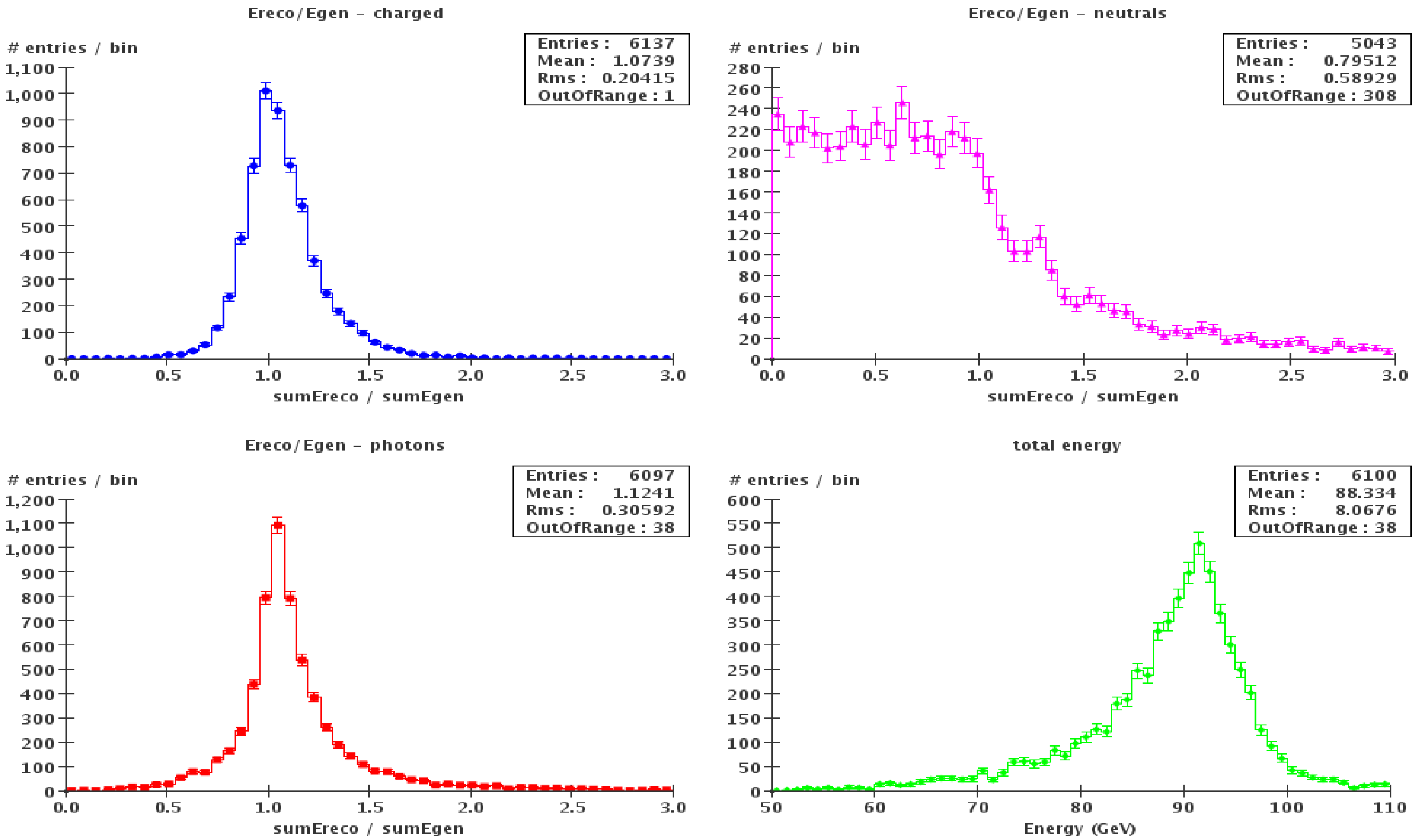
Current PFA results (preliminary)

Energy sums of each component in event

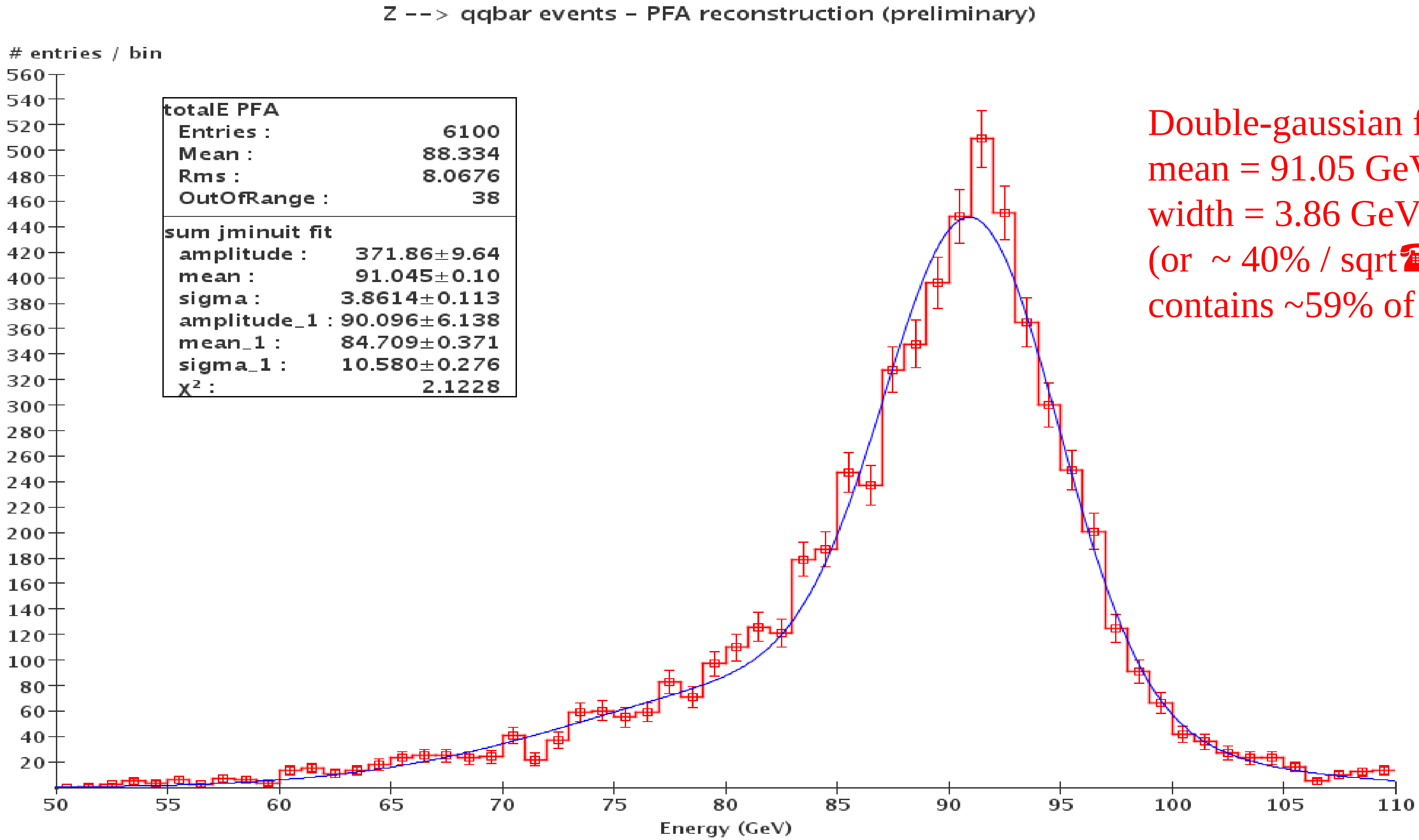


Current PFA results (preliminary)

Resolutions for the energy sums of each component in event



Current PFA result (preliminary)

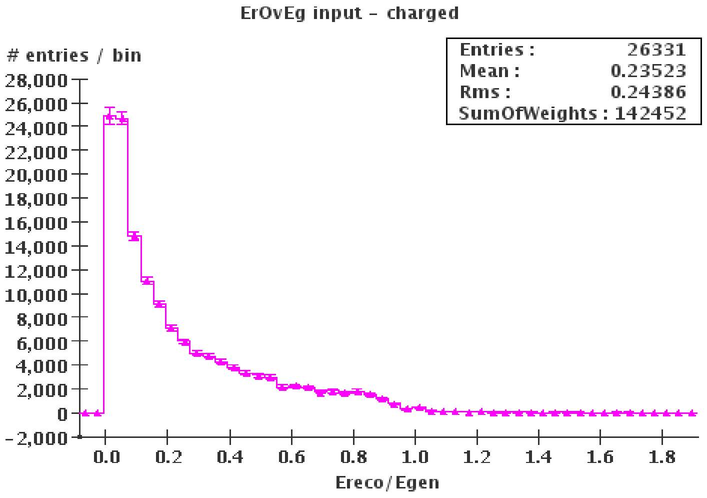


Double-gaussian fit
mean = 91.05 GeV
width = 3.86 GeV
(or ~ 40% / sqrt☎E))
contains ~59% of events

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 is still necessary**
- Things to do:
  - Investigate origin of misidentified clusters and how to improve cluster identification
  - Use tail catcher data 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)
  - Study PFA at higher energies and more complex physics processes (WW, ZH, etc)
  - Comparisons for different geometries, B-fields and technologies

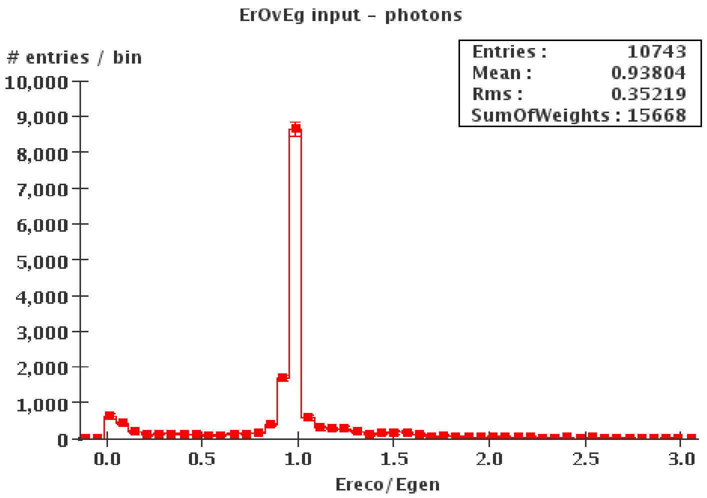
E_{rec}/E_{gen} for individual clusters (input to PFA)

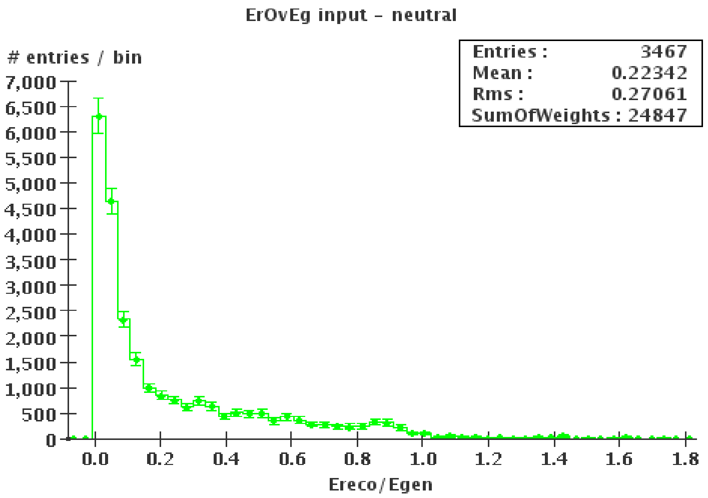


These plots (here and on next page) have one entry per cluster, which means several entries per MCParticle, due to their hits being split into several reconstructed fragments.

If all fragments are correctly associated, one would then get one entry per particle on next page, with peaks around 1 on the top plots. With perfect cluster-ID, there would be no entries to the bottom plots.

Photon hits do not have many fragments. This means that the directed tree clustering algorithm is doing a good job for the photons.





E_{rec}/E_{gen} for individual showers (result of PFA)

Photon hits do not have many fragments. This means that the directed tree clustering algorithm is doing a good job for photons, although some work is still needed on photon-ID.

For hadrons, fragment association (or shower pattern recognition) algorithms are required, and plots like these can be used to evaluate their performance and drive the development of the PFA algorithm.