Status of particle flow studies (short version)

Mat Charles
The University of Iowa

General approach

- Taking things step by step
- Get a working particle flow algorithm (even with lousy resolution)
- Find the limiting piece of the algorithm, then improve it until it no longer dominates resolution. Iterate.
- Don't expect to see 30%/√E today

Using SDFeb05_SciHcal throughout

General approach

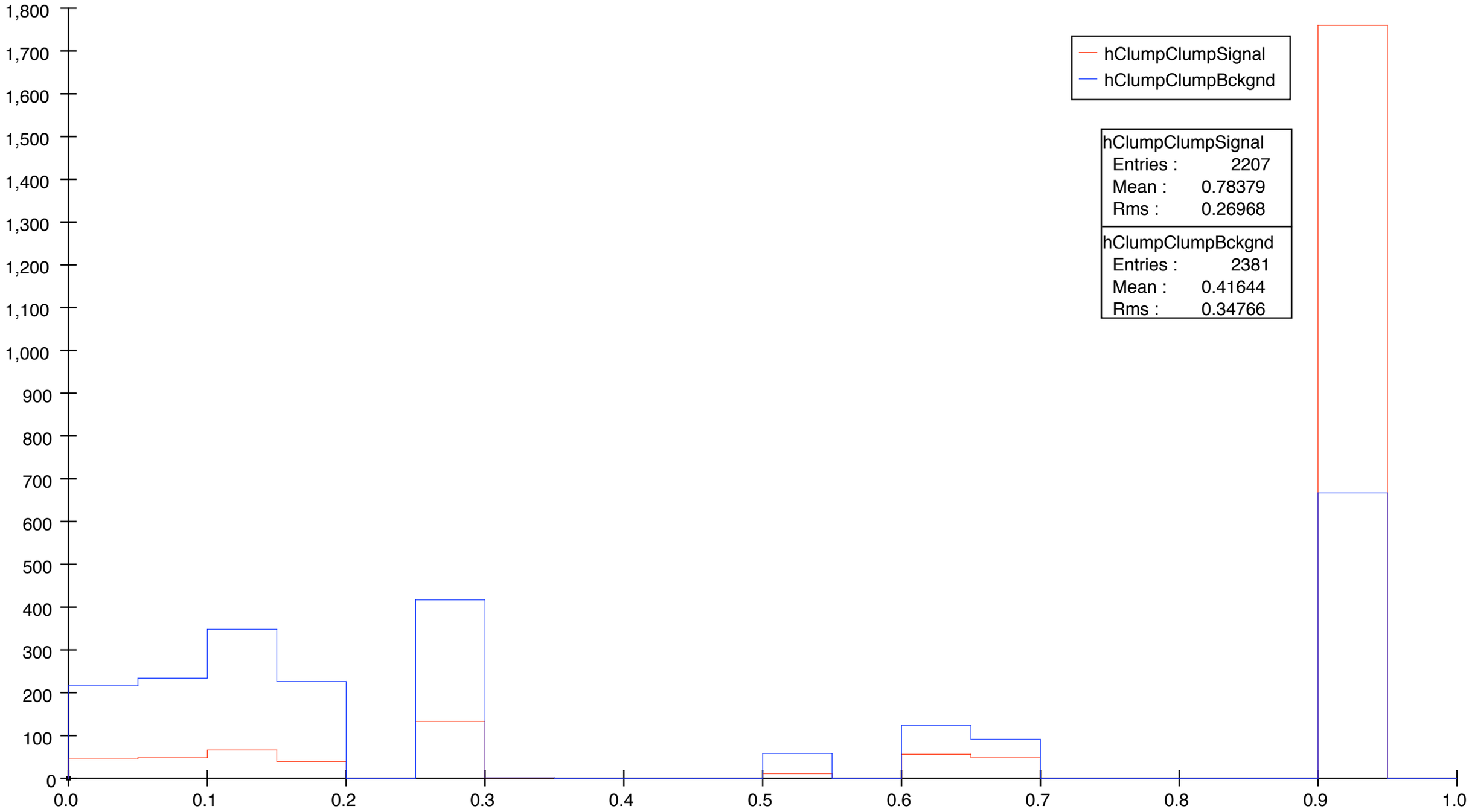
- Need to separate and identify hadronic clusters.
- Hadronic clusters have substructure (unlike typical EM clusters)
- Approach is to break down MST clusters into pieces, check whether they should really be linked, then reassemble them.
- Parts of the skeleton:
  - Clumps -- dense collections of hits
  - Track segments
- ... plus halo/fragments/satellites

List of likelihood variables

- Clump-Clump:
  - DOCA
  - Smallest distance from a hit in one cluster to a hit in the other
- Track-Clump:
  - DOCA
  - Smallest distance from a hit in one cluster to a hit in the other
- Track-Track:
  - DOCA
  - Smallest distance from track hit to POCA
  - Whether POCA is inside calorimeter
  - Extrapolating track to POCA (or joint CoE for parallel & disjoint tracks)...
    - # Layers where a hit is not found
    - Fraction of layers where a hit is not found (ignoring layers with a hit from cluster itself)

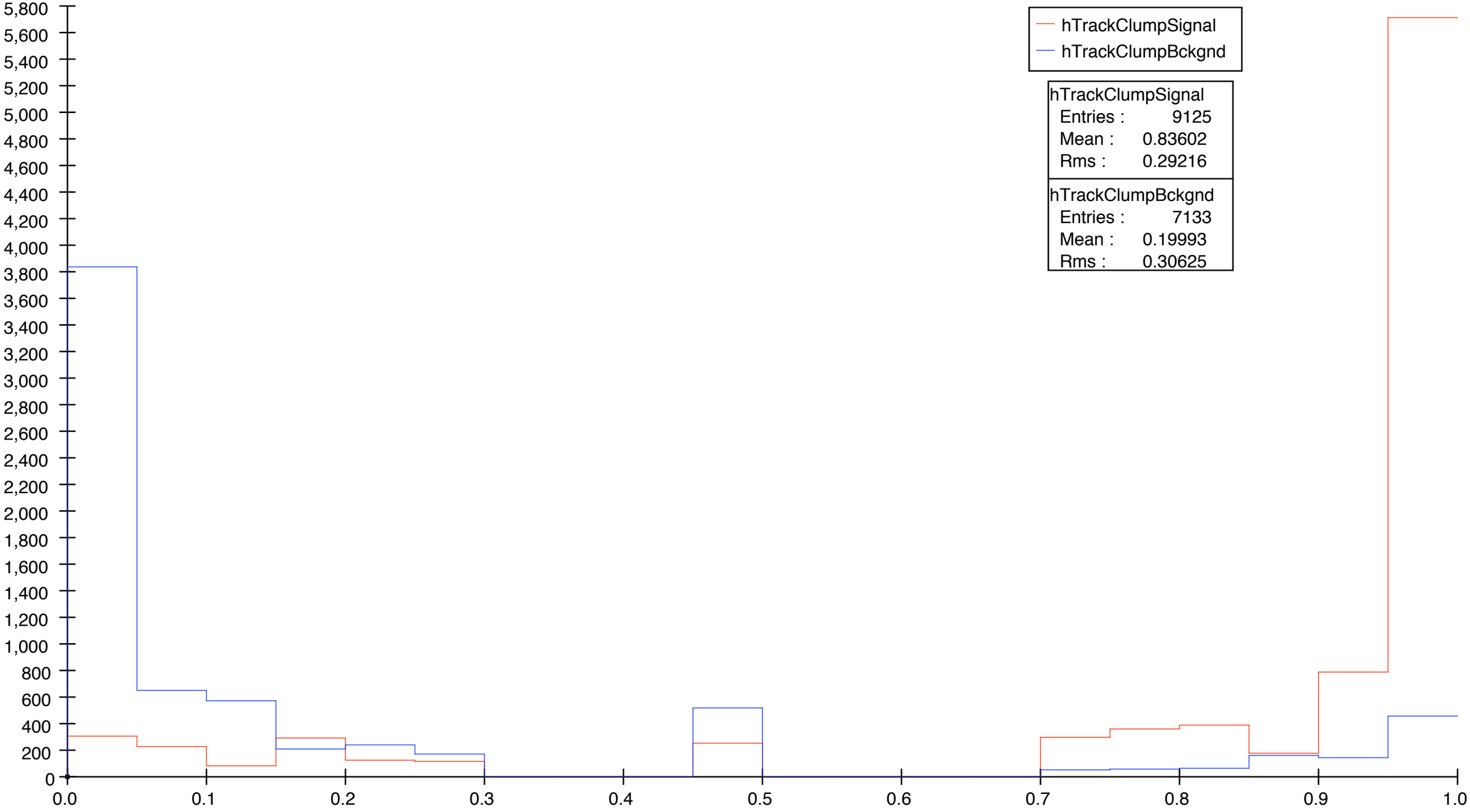
Check performance on a sample of 340 Z-pole ($Z^0 \rightarrow b\bar{b}$) events...

Performance: Clump-Clump links

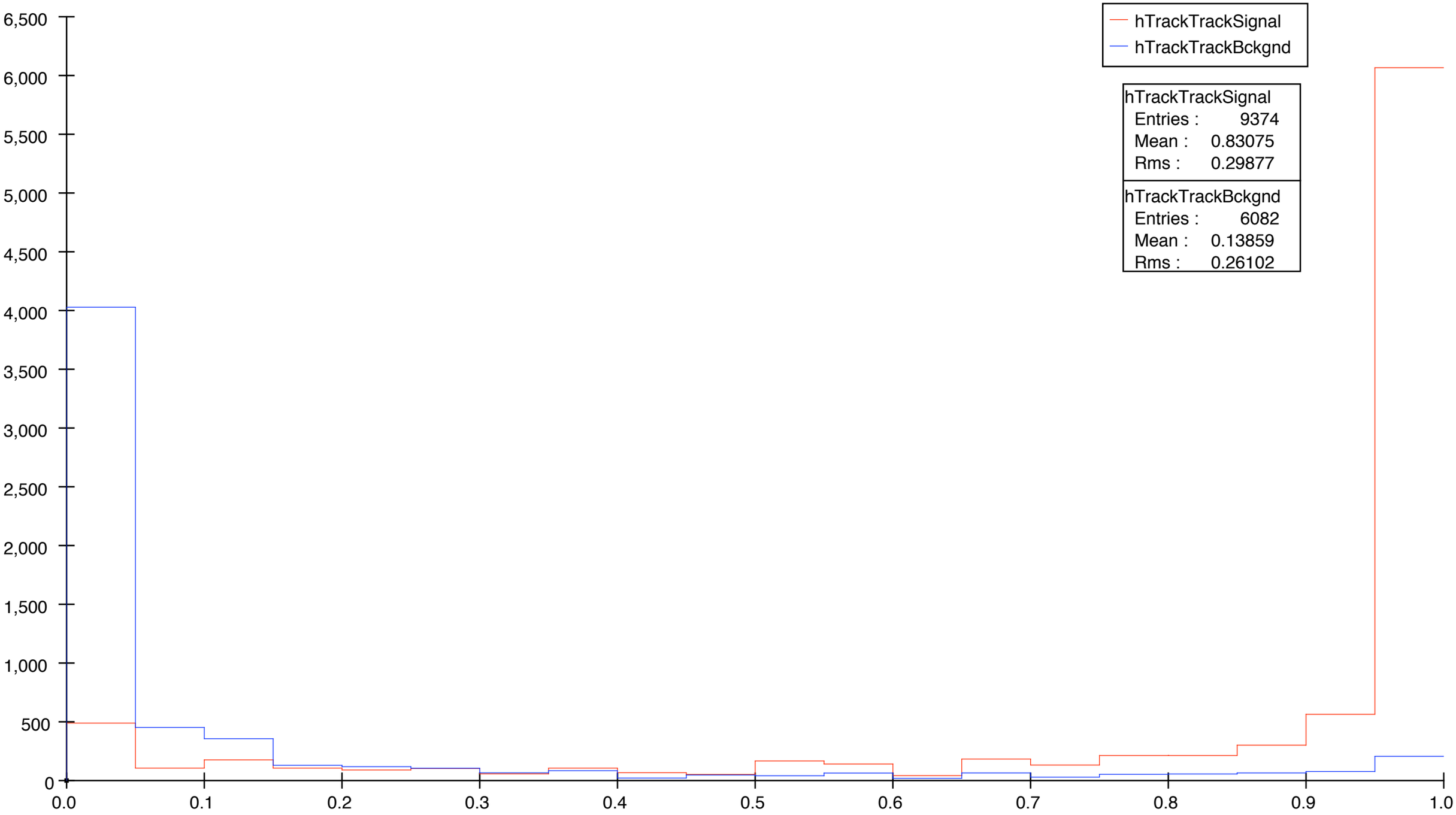


Likelihood

Performance: Track-Clump links



Performance: Track-Track links



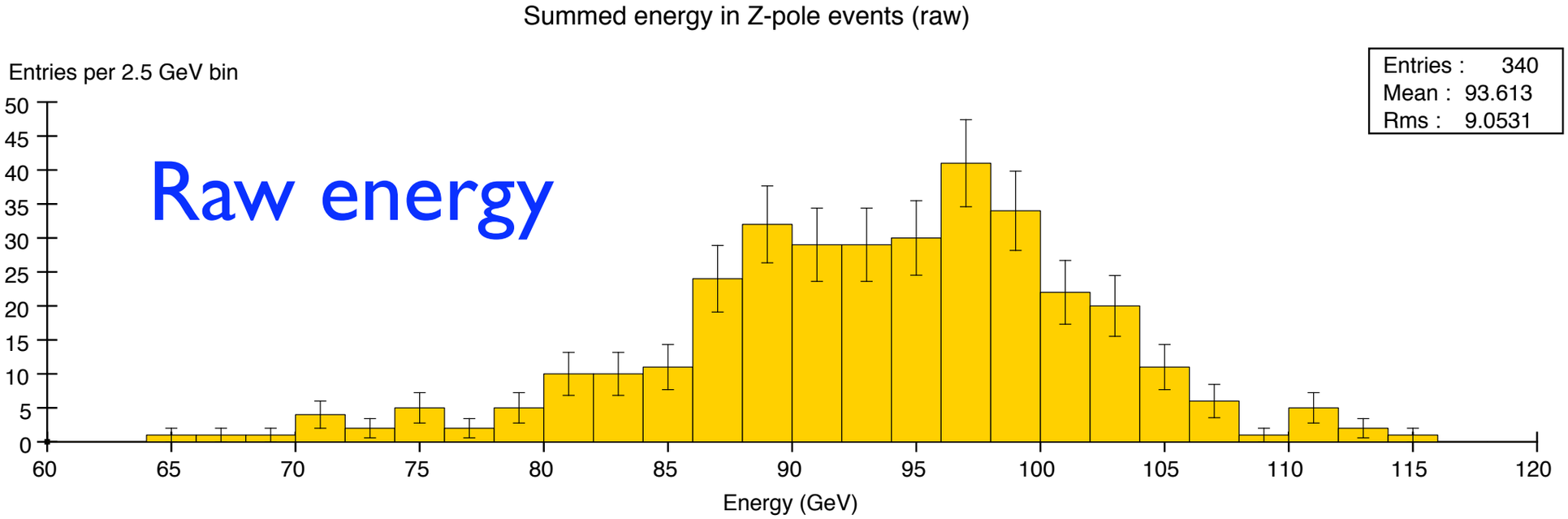
Likelihood

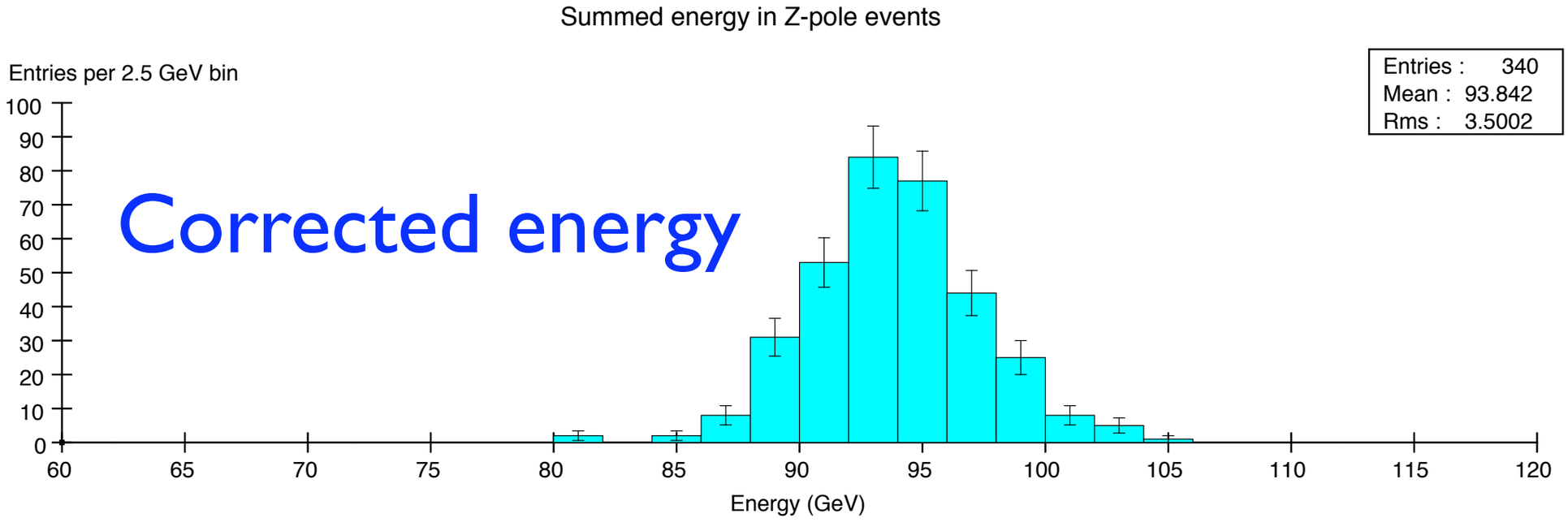
Where next?

- There are things that can be done to improve the algorithm
  - Ignoring correlations
  - Histogram binning is crude
    - Some variables not optimal or vulnerable to special cases
    - Linkage during training
- ... but for now it's good enough to study Particle Flow performance

Perfect PFA

How well could we do if the clustering algorithm were perfect?



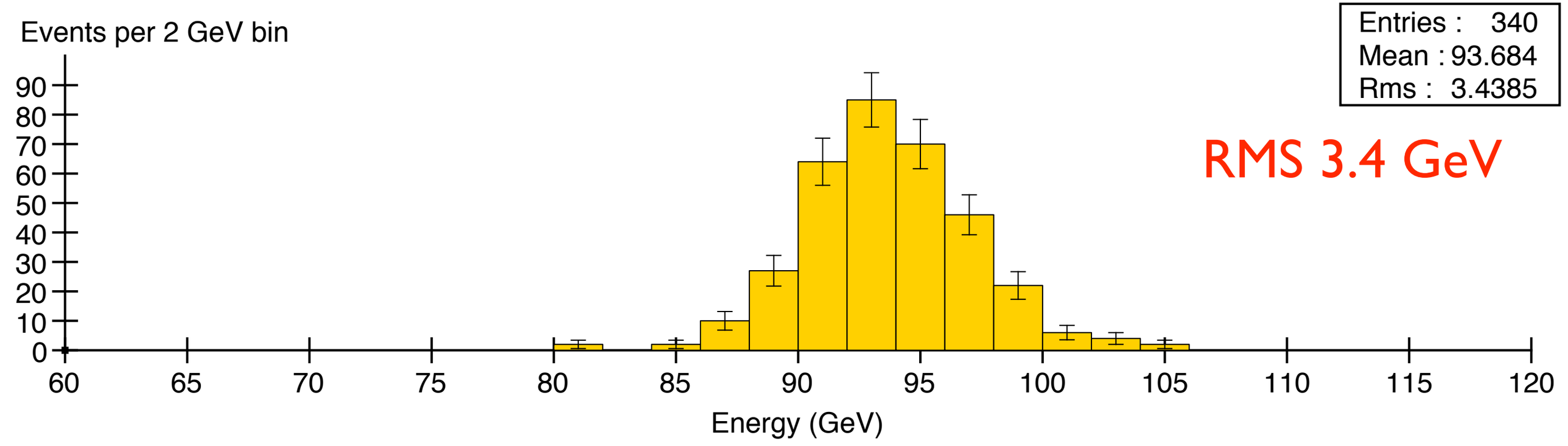


... so cheating got us 3.5 GeV.

Start from real algorithm, then add cheating...

+ Cheat to decide if it's a fragment (is it the dominant deposit for some MC particle?)
+ If it's a fragment, cheat to assign it to the parent cluster
+ For each cluster with > 0 charged tracks, energy = sum of track energies (using truth info)
+ For each cluster with 0 charged tracks, energy from calorimeter

Summed energy in Z-pole events

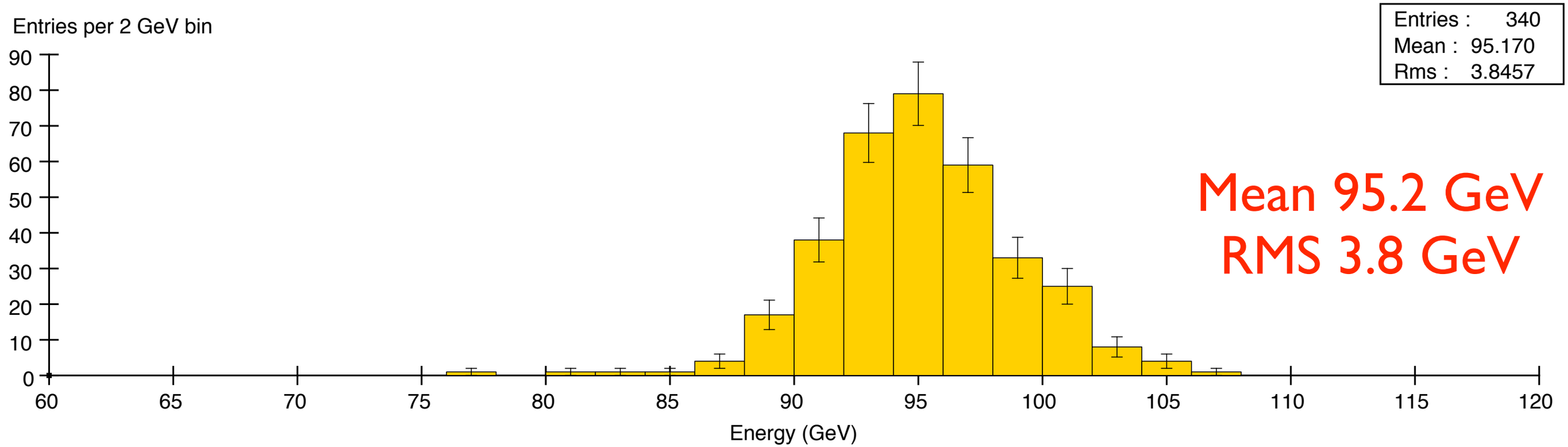


We are consistent with results from cheating completely.
(Perfect PFA gives mean 93.8, RMS 3.5 GeV)

Handle fragments a bit more realistically

Handling of fragments:

+ Cheat to decide which are fragments and which are not
+ Dumb guess: Assign fragments to nearest cluster (i.e. without cheating)



(Cheating gave mean 93.8, RMS 3.5 GeV)

RMS is OK. (Will it stay OK if we don't cheat at all on fragments?)
Mean is too high -- we are double-counting some charged energy.

Handle fragments a bit more realistically

Instead of cheating to decide what is/isn't a fragment, try to come up with criteria.

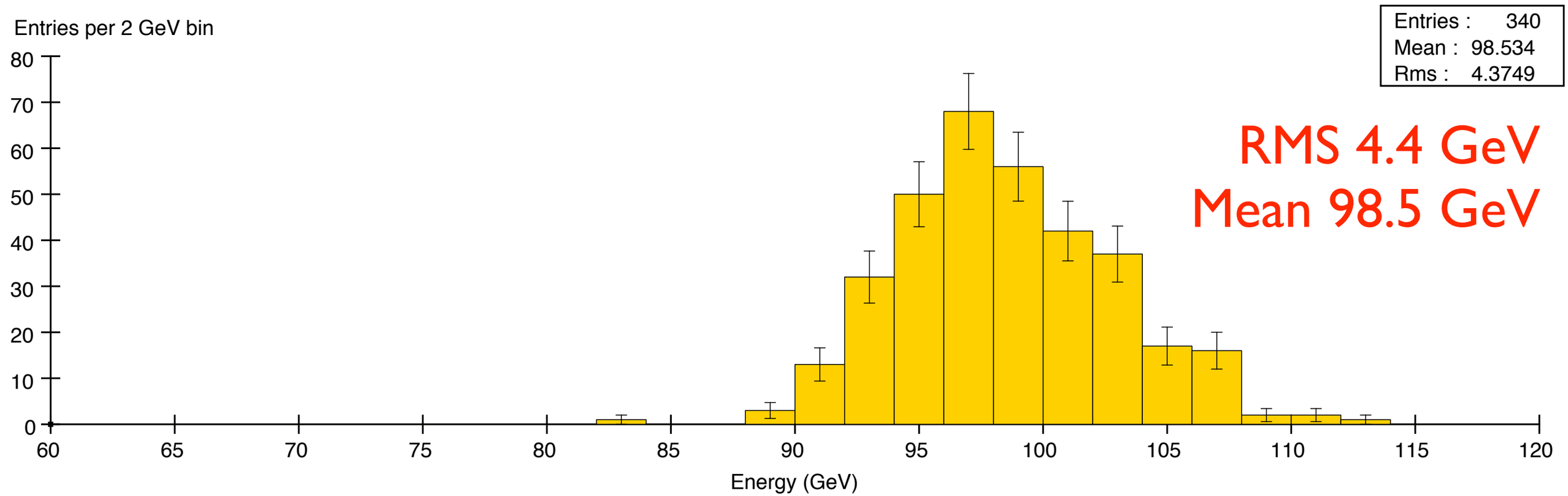
Ideas to do this properly:

+ Link fragments in with other clusters based on pointing/DOCA etc. (similar to before)
+ Likelihood selector comparing primary vs. secondary clusters
+ Careful when handling things like $K_S \rightarrow \pi^0 \pi^0$

... but for now start with a simplistic cut-based analysis.

Not cheating on fragment ID

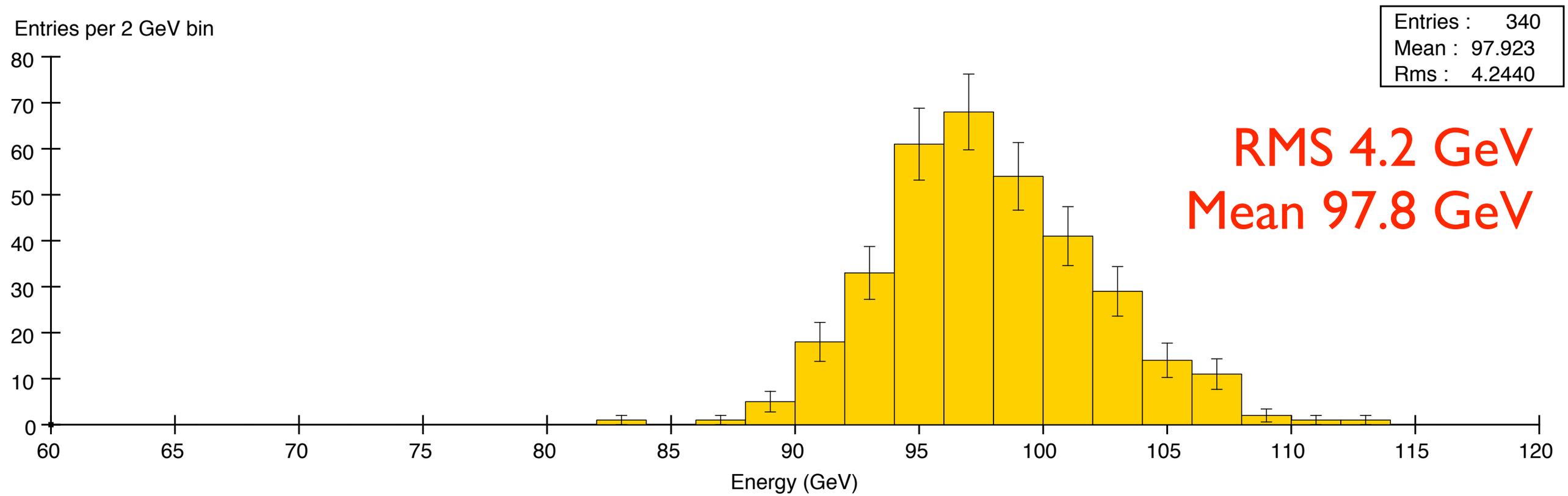
In fact, we are being very dumb about fragment ID.
We can certainly do better, but with these cuts we get:



Call it a fragment if no associated charged track
AND doca > 100 cm AND nHits < 10

Mean is higher again -- more double-counting of charged energy
(Cheating gives mean 93.8, RMS 3.5 GeV)

With a different set of fragment cuts I tried...



RMS 4.2 GeV

Mean 97.8 GeV

Call it a fragment if no associated charged track AND doca > 125 cm AND nHits < 25

Mean is higher again -- more double-counting of charged energy

(Cheating gives mean 93.8, RMS 3.5 GeV)

Summary

- Framework for likelihood selector is in place
  - Should be more careful with training. Histograms, selectors not really optimized
- With perfect PFA (but imperfect sampling fractions, tracks), RMS is 3.5 GeV, i.e. 37%/$\sqrt{E}$
- Using real structural algorithm but cheating on fragments, RMS is 3.4 GeV.
  - ... so likelihood-based structural algorithm is good enough for now
- Using v. naive fragment handling, RMS is ~ 4.3 GeV, i.e. 45%/$\sqrt{E}$
- Electromagnetics not handled properly at all
  - Not a big deal now, but will matter with proper sampling fractions
- Next: Convert to org.lcsim! Handle tracks properly

Used SDFeb05_SciHcal throughout

Words from Niels

Niels can't make it today, but he asked me to tell you about his plans.

- He's switching from hep.lcd to org.lcsim
  - Aims to reach the same level in org.lcsim as he was at Snowmass in hep.lcd within 1-2 weeks
- Add/compare cluster ID to his photos (HMatrix, NN, others?)
- Optimize further energy collection (EM fragments)
- Form lists of electromagnetic objects
  - Electrons, photons, pi0, merged pi0, merged Ks