Usage and Efficiency of SimpleClusterBuilder, ClusterCheater, and Similar Clustering Algorithms

Christopher Geraci
University of Colorado
High Energy Physics
International Linear Collider

Usage and Efficiency Study of Clustering Algorithms

- Description of SimpleClusterBuilder Algorithm
- Description of ClusterCheater Algorithm
- Comparitive Analysis of both clustering algorithms
- Plans for studies of more efficient clustering algorithms

SimpleClusterBuilder

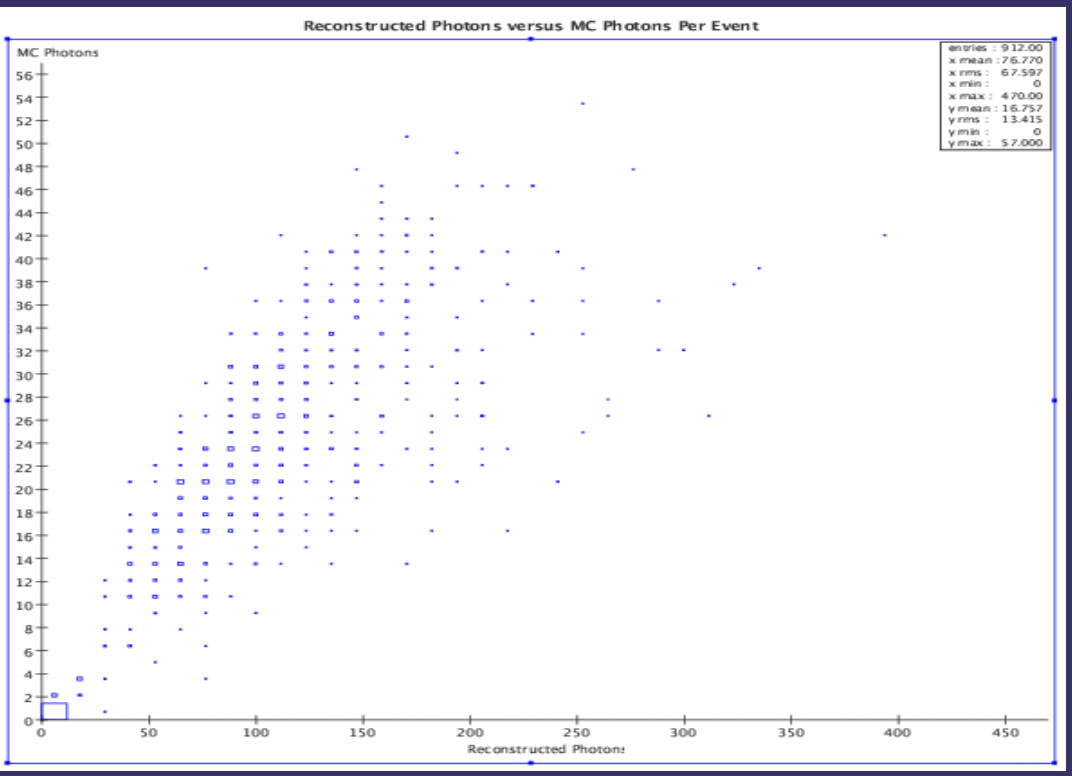
- Nearest neighbor clustering algorithm
- Uses delta-theta, delta-phi, and layers in order to determine which particles to group together as a cluster
- At defaults, this clustering algorithm is very inefficient and creates far too many clusters considering the number of involved Monte Carlo particles
- Defaults of delta-theta, delta-phi, and layers are all set to 1
- The goal of this study is to determine if it is possible to modify this clustering algorithms in order to use it with the LD offset geometry

ClusterCheater

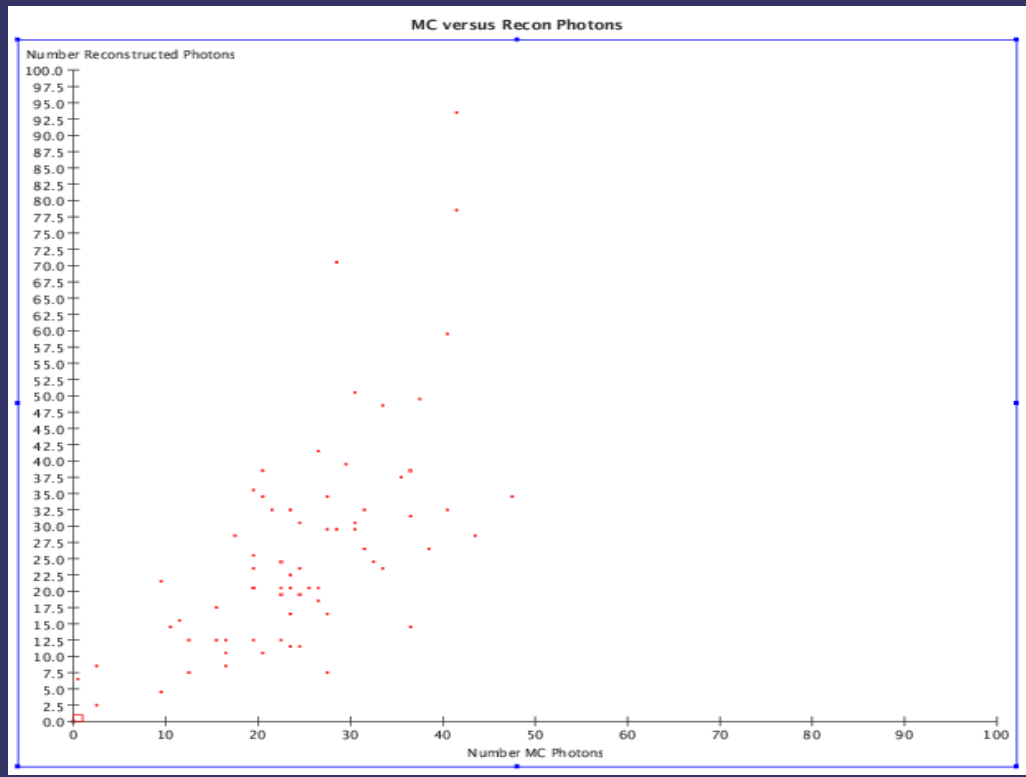
- Uses MC truth information in order to build clusters
- Associates all clusters spawned of a certain MC particle into one cluster
- Causes the average MC to recon cluster ratio to be approximately 1 meaning about 1 reconstructed cluster for every MC particle

1st Order Comparison of Number of MC Clusters versus Number of Reconstructed Clusters

There are approximately 5 clusters for every MC particle using SimpleClusterBuilder, whereas there is almost exactly one cluster using ClusterCheater

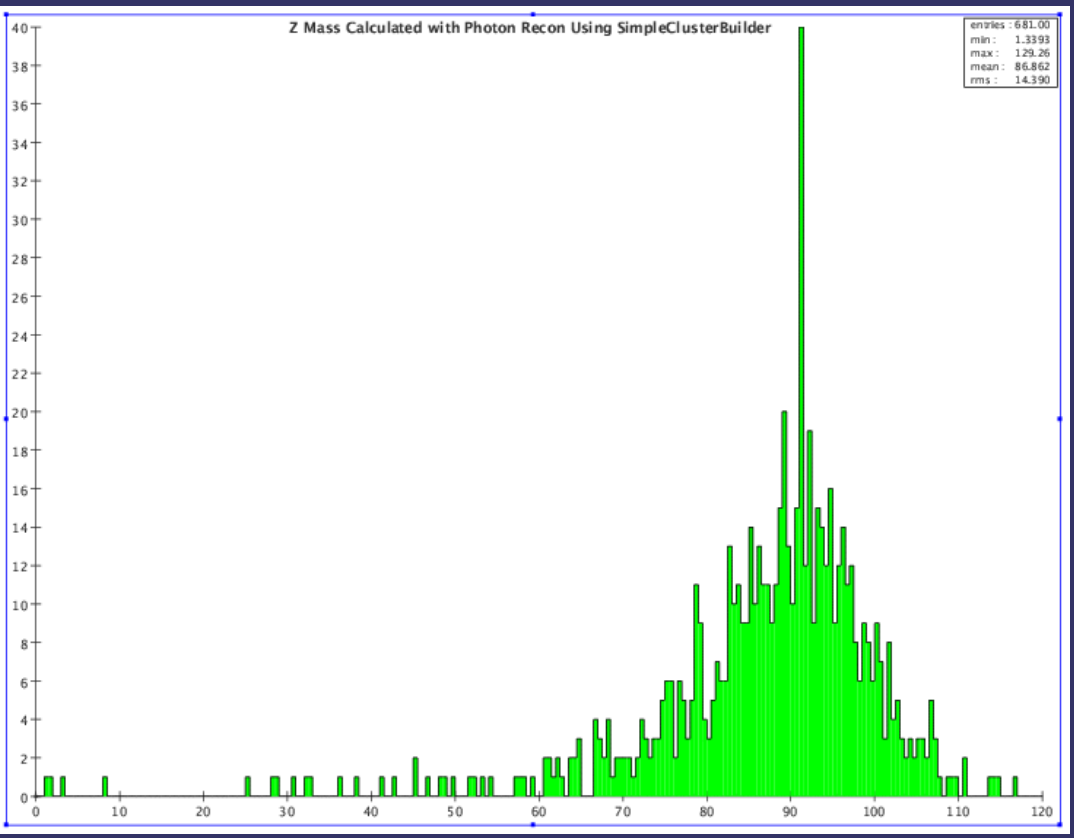


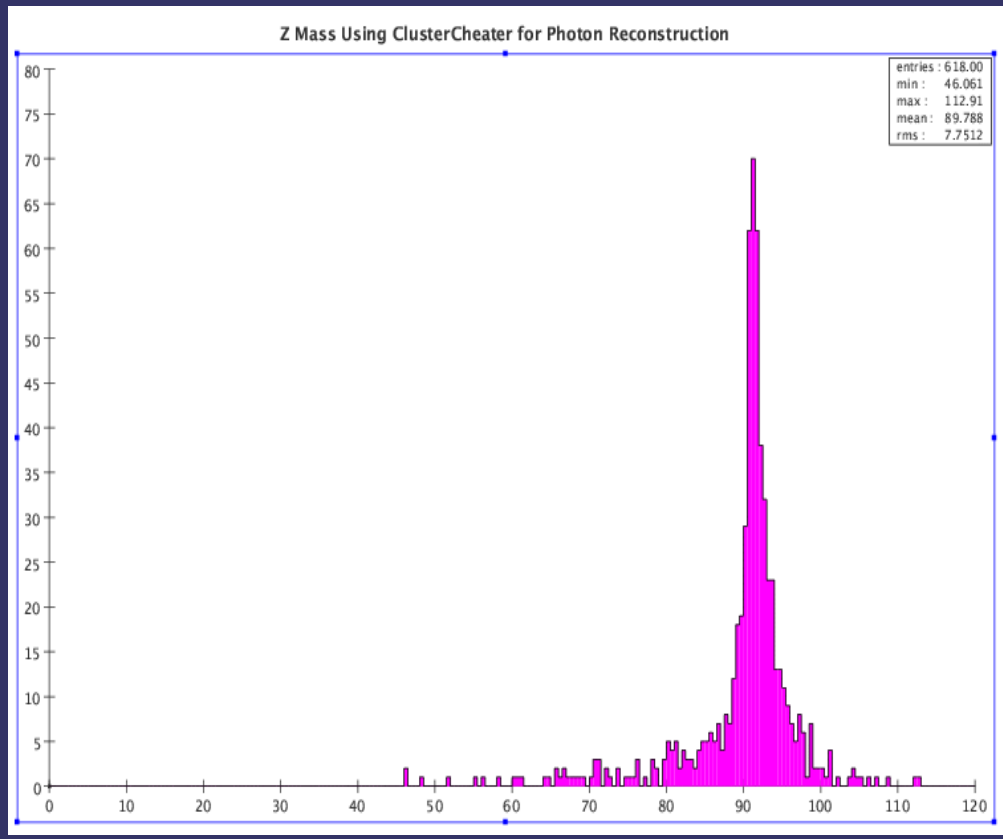
Number MC particles versus number of reconstructed photonic clusters for SimpleClusterBuilder



Number MC particles versus number of reconstructed photonic clusters for ClusterCheater

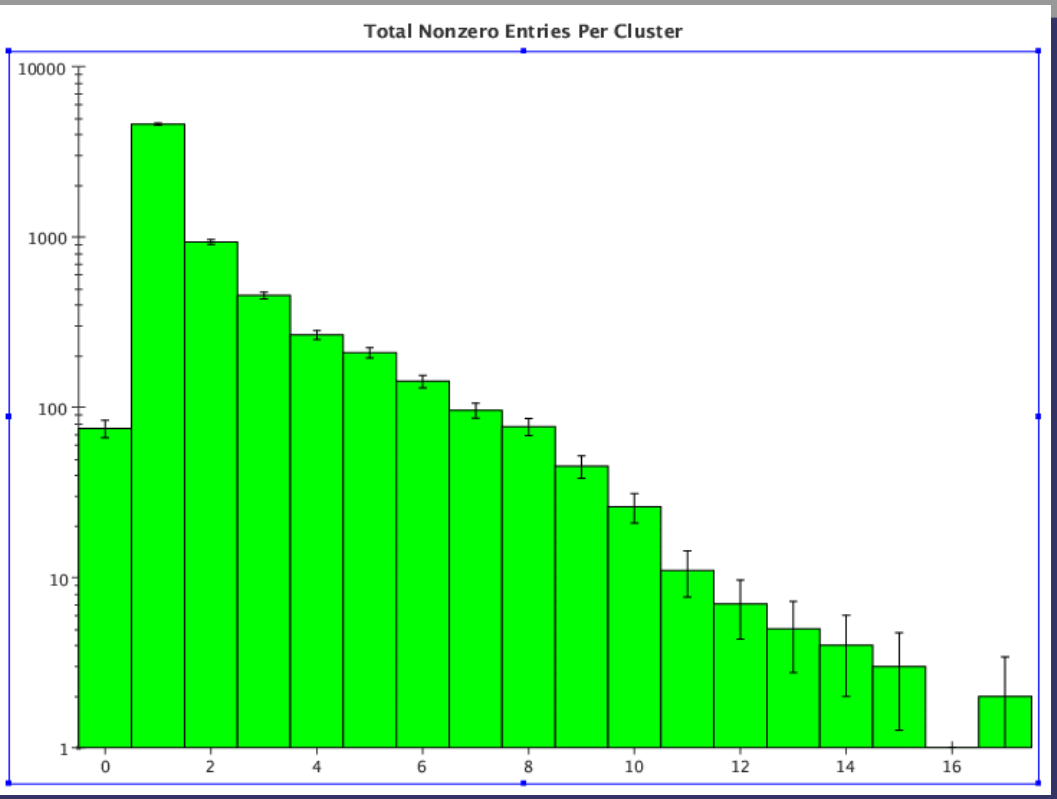
Z Mass Resolution Comparison

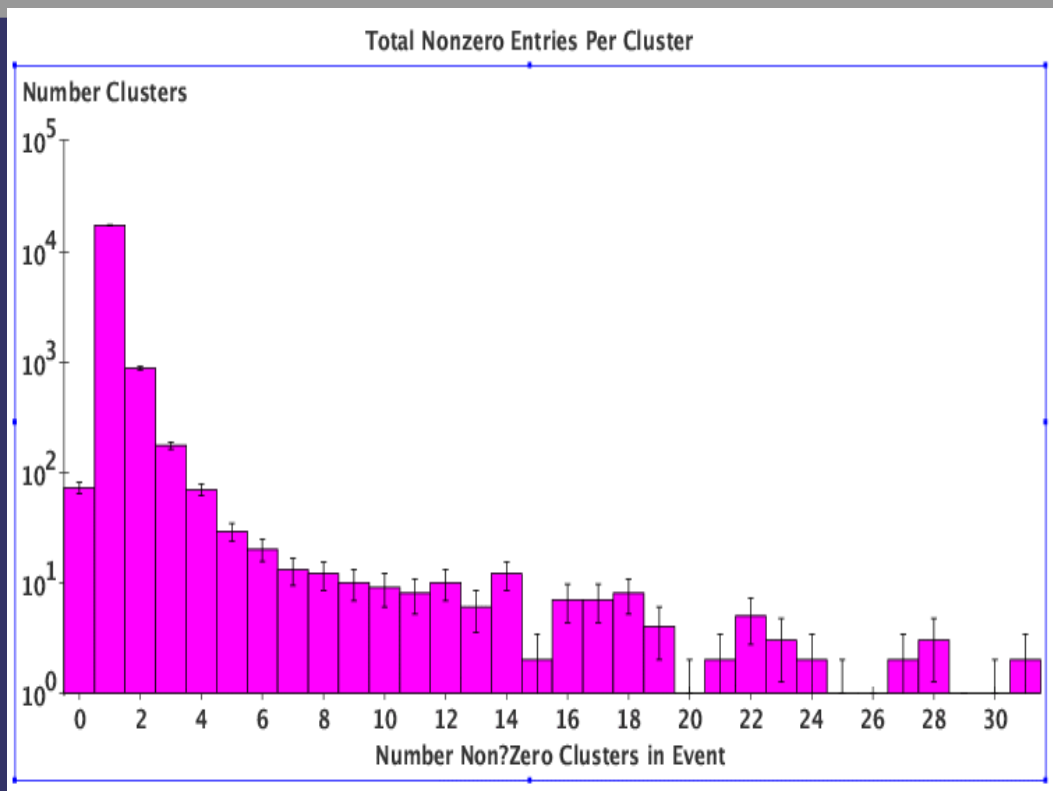




Because of the more efficient photon reconstruction using CC, the Z mass peak is much less sloppy and is shifted up.

Total number of non-Zero Energies Per Cluster



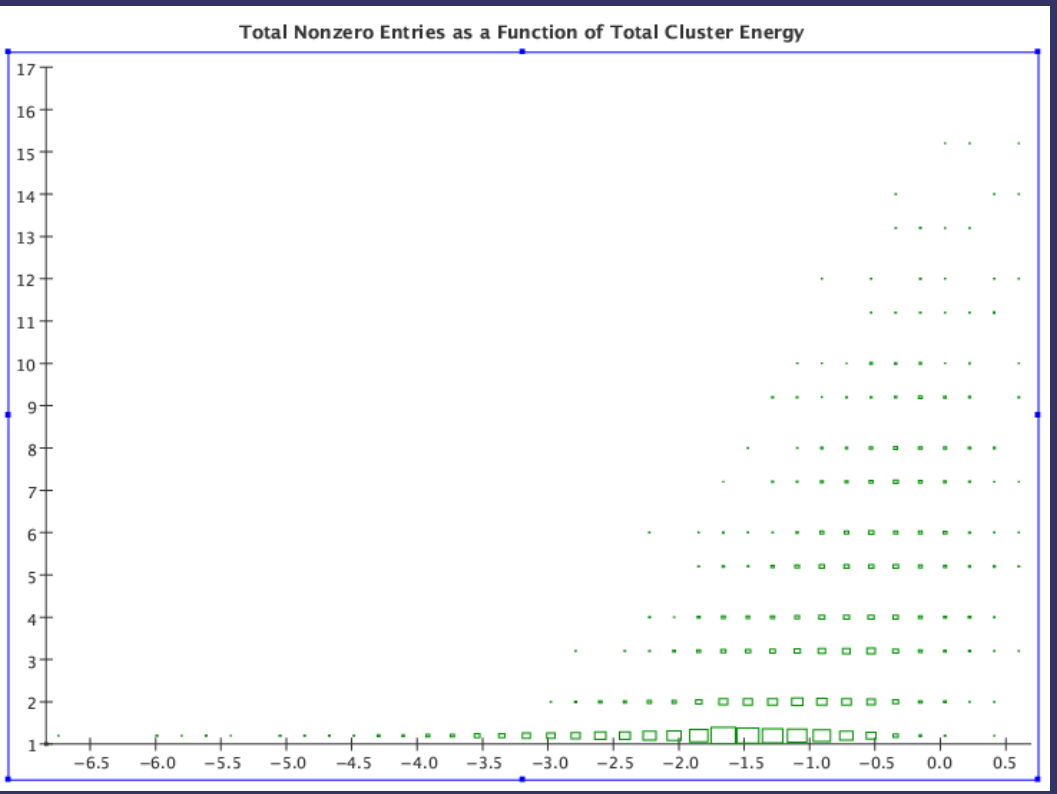


CC, with a relatively small range of non-zero entries per cluster

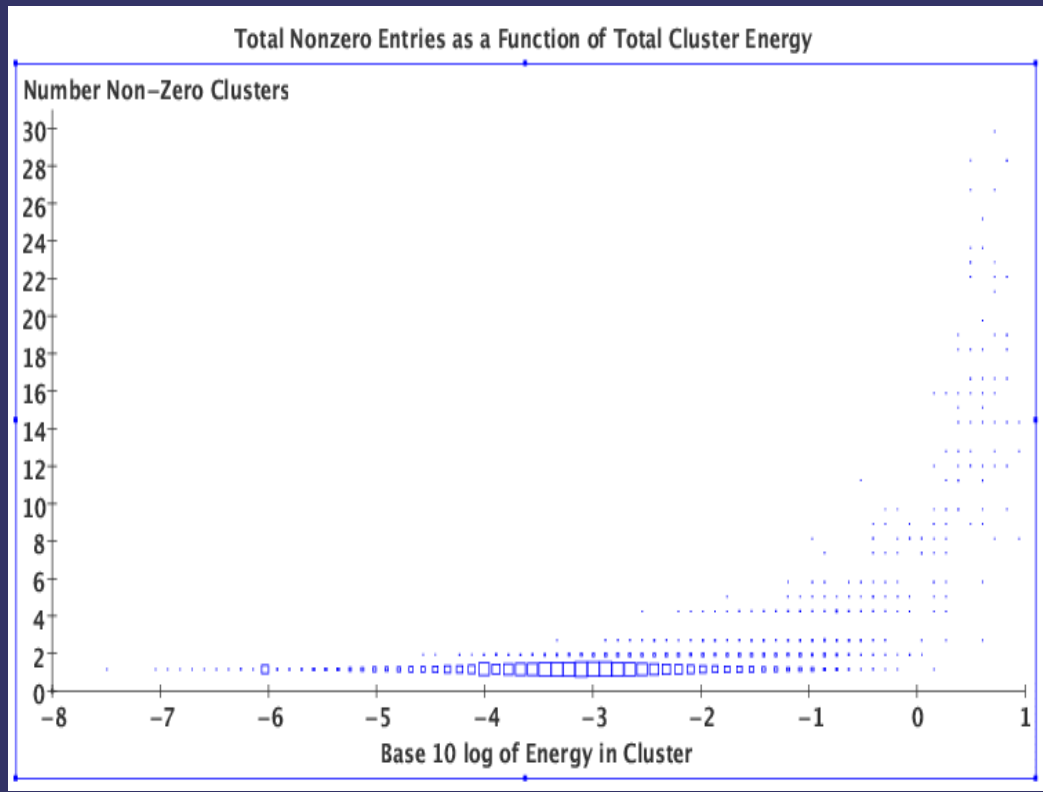
SCB, with a larger range of non-zero entries per cluster

With a high number of non-zero entries, it is more likely that the purity of the cluster will be compromised

Number of Non-Zero Entries per Cluster as a Function of Total Contributed Cluster Energy



CC Number of nonzero clusters versus total cluster energy



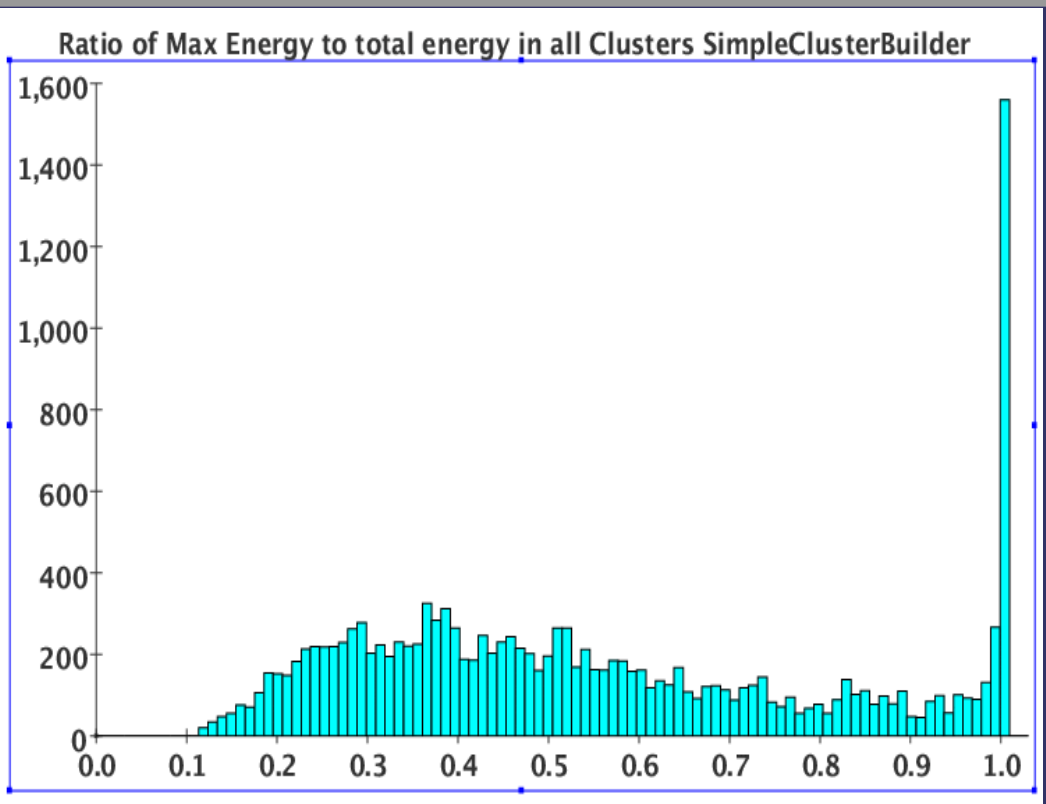
SCB number nonzero clusters versus total cluster energy

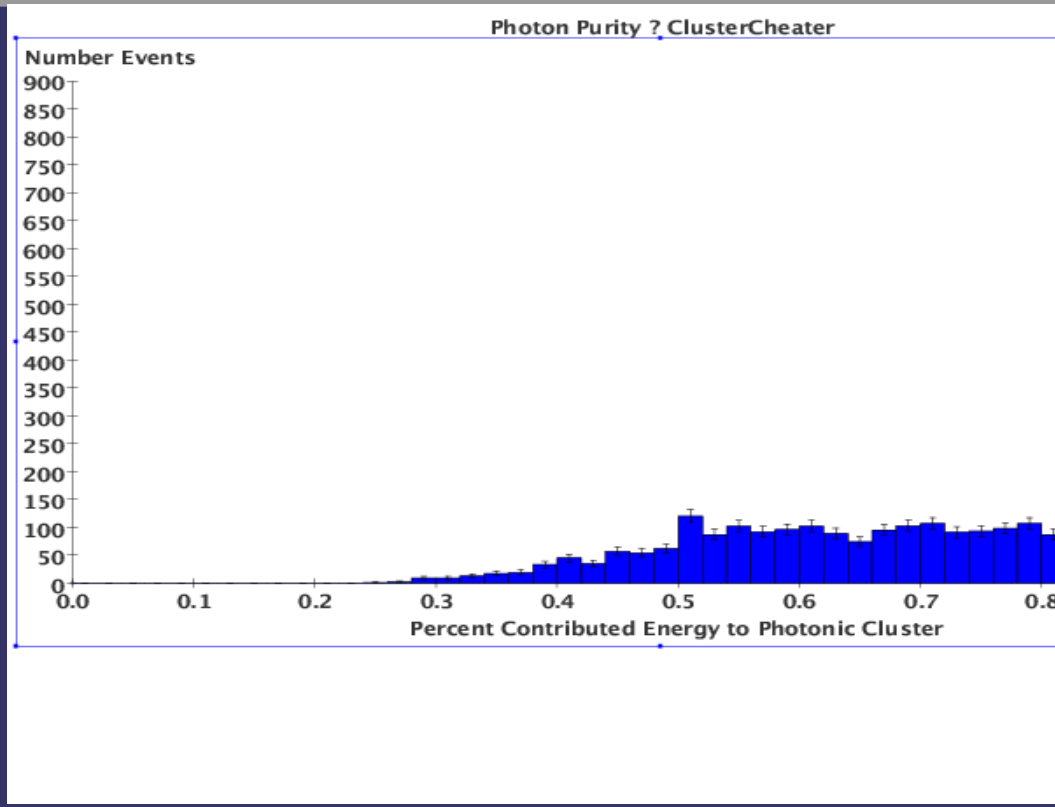
The number of nonzero clusters seems to depend on the energy in the cluster

Multiple Clusters Spawned of the Same MC Particle

- Since SCB does not associate clusters by MC particle information, there can be multiple clusters spawned from the same MC particle
- Excessive numbers of nonzero entries in a cluster contribute to poor resolution
- This inefficiency will be impossible to resolve in a real world clustering algorithm, so we cannot expect CC results from *any algorithm*

Ratio of Energy Contributed by Primary Photon to Energy of all Clusters Spawned of the MC Photon



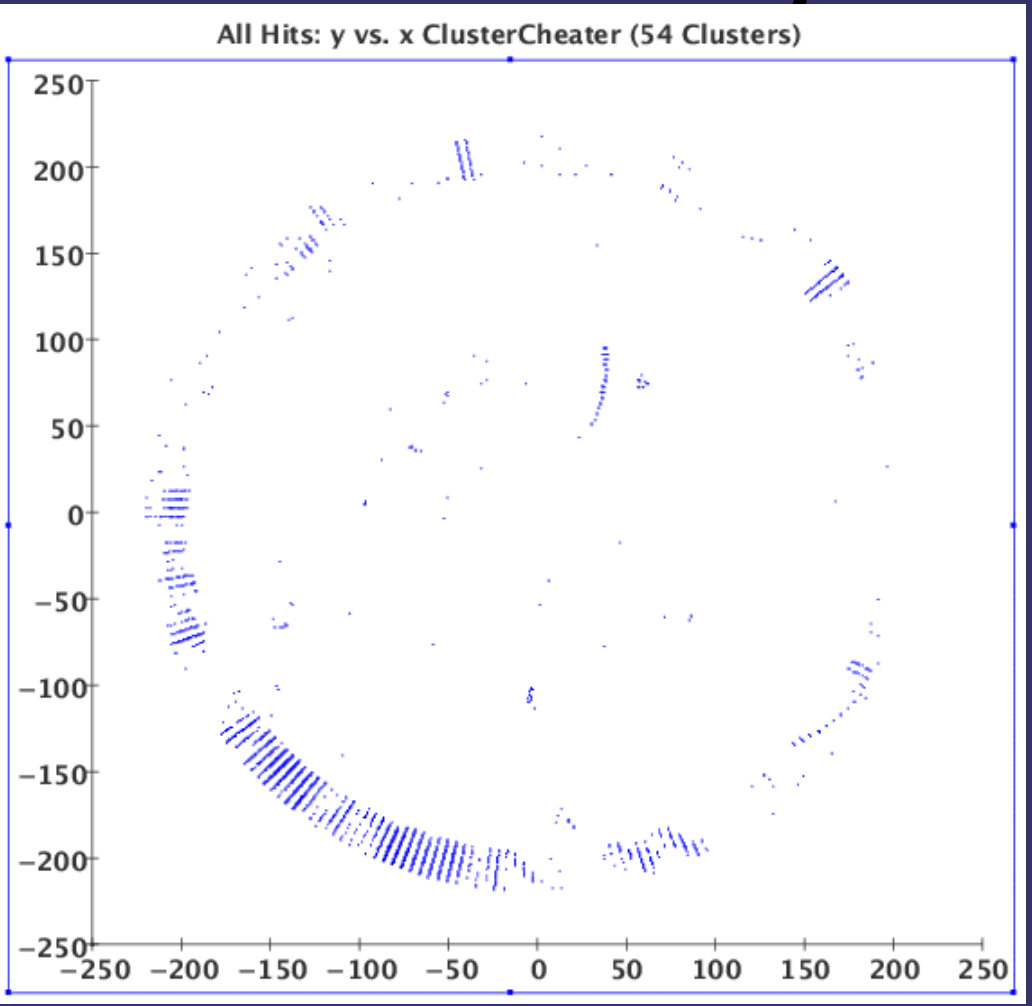


SCB: Significantly more clusters with worse resolution

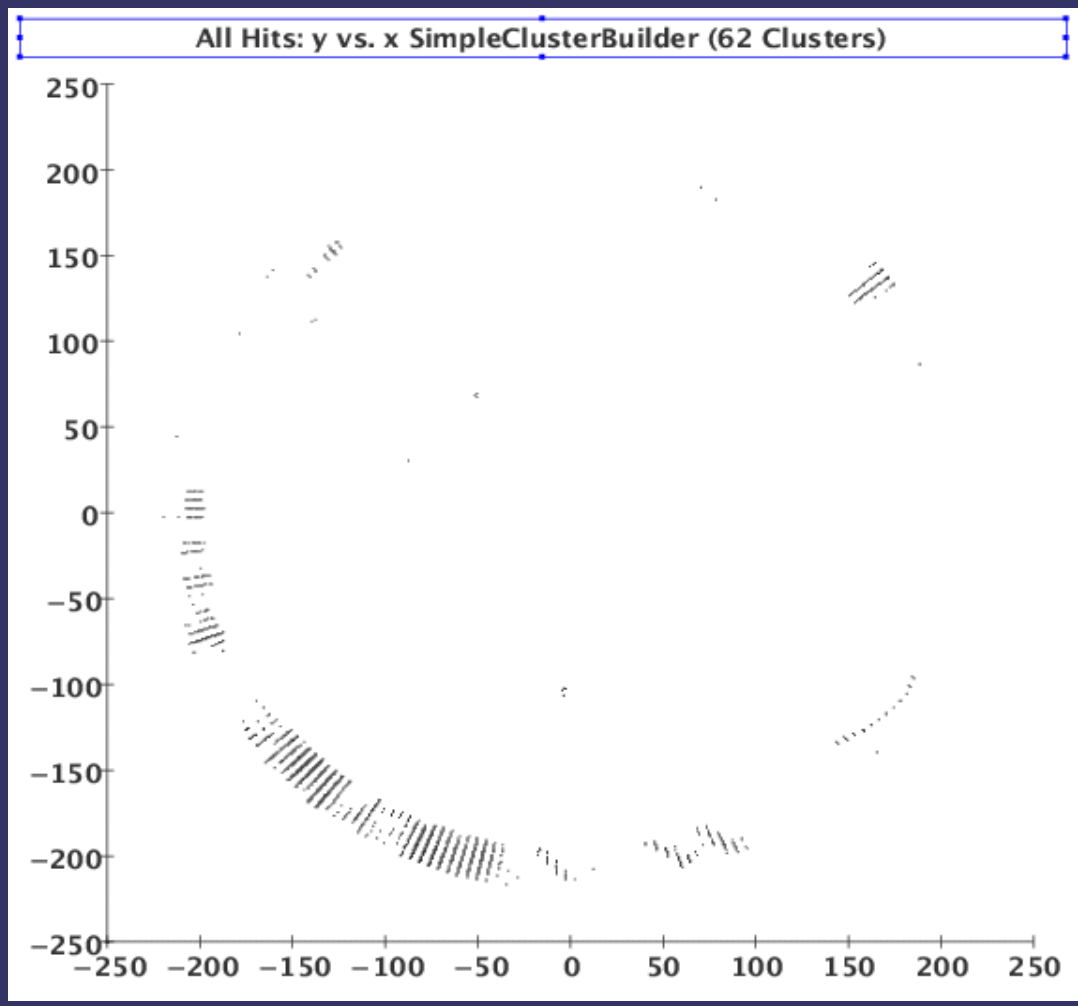
CC: Less clusters with much better resolution

The second peak for SCB occurs close to .2, which is concurrent with the MC particle versus reconstructed clusters shown earlier

XY Calorimeter Mapping for Hits: Cluster Cheater vs. SimpleClusterBuilder

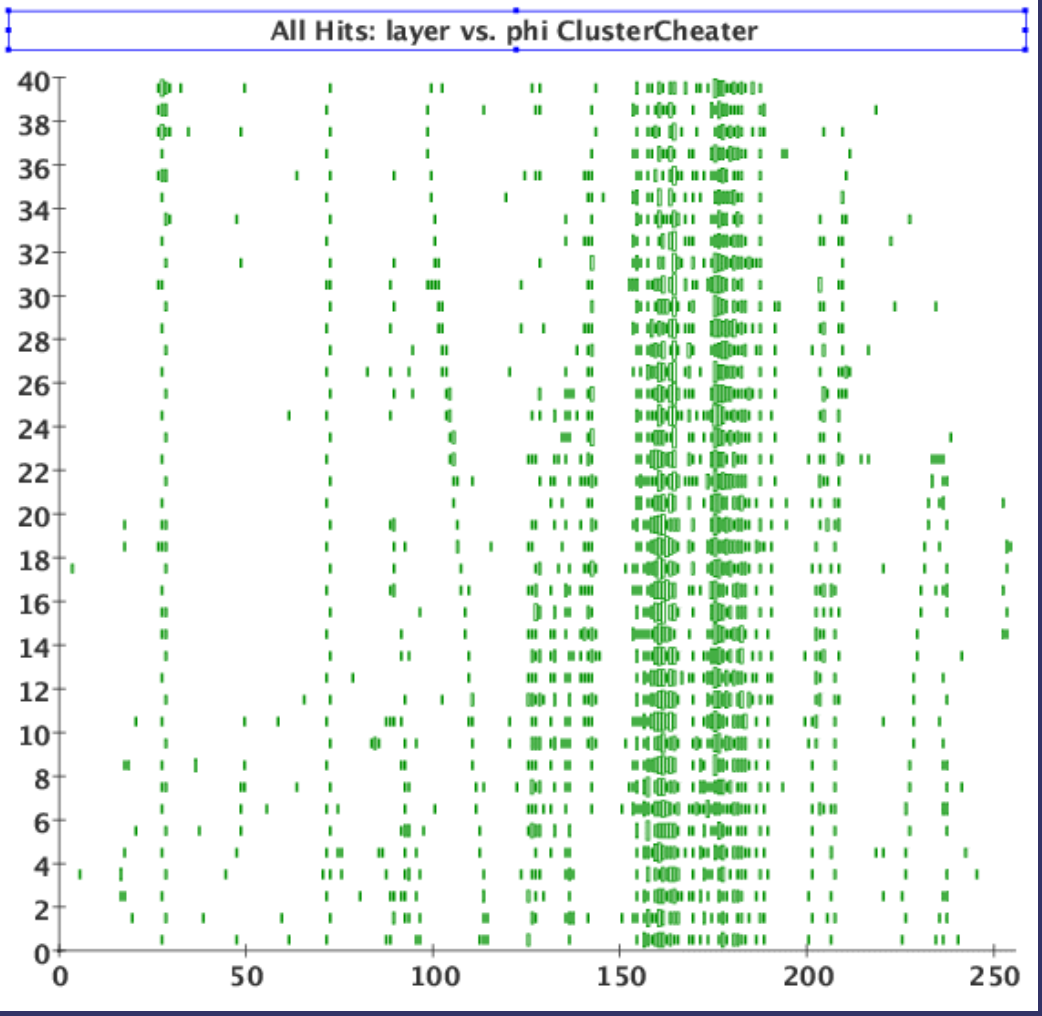


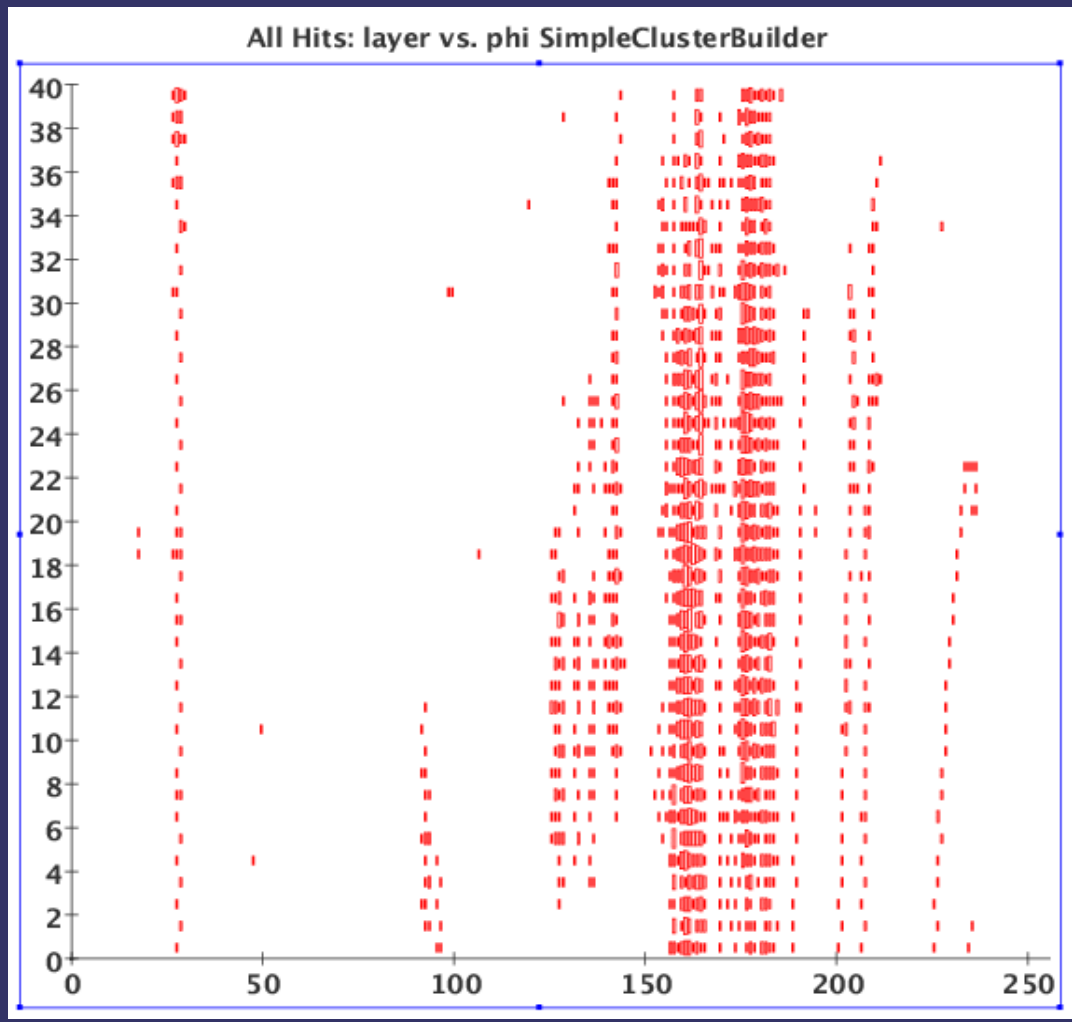
Y vs. X 2D calorimeter map for CC:
less clusters, many more visible hits



Y vs. X calorimeter map for SCB:
more clusters, far less visible hits

Layer Versus Phi Study for ClusterCheater and SimpleClusterBuilder

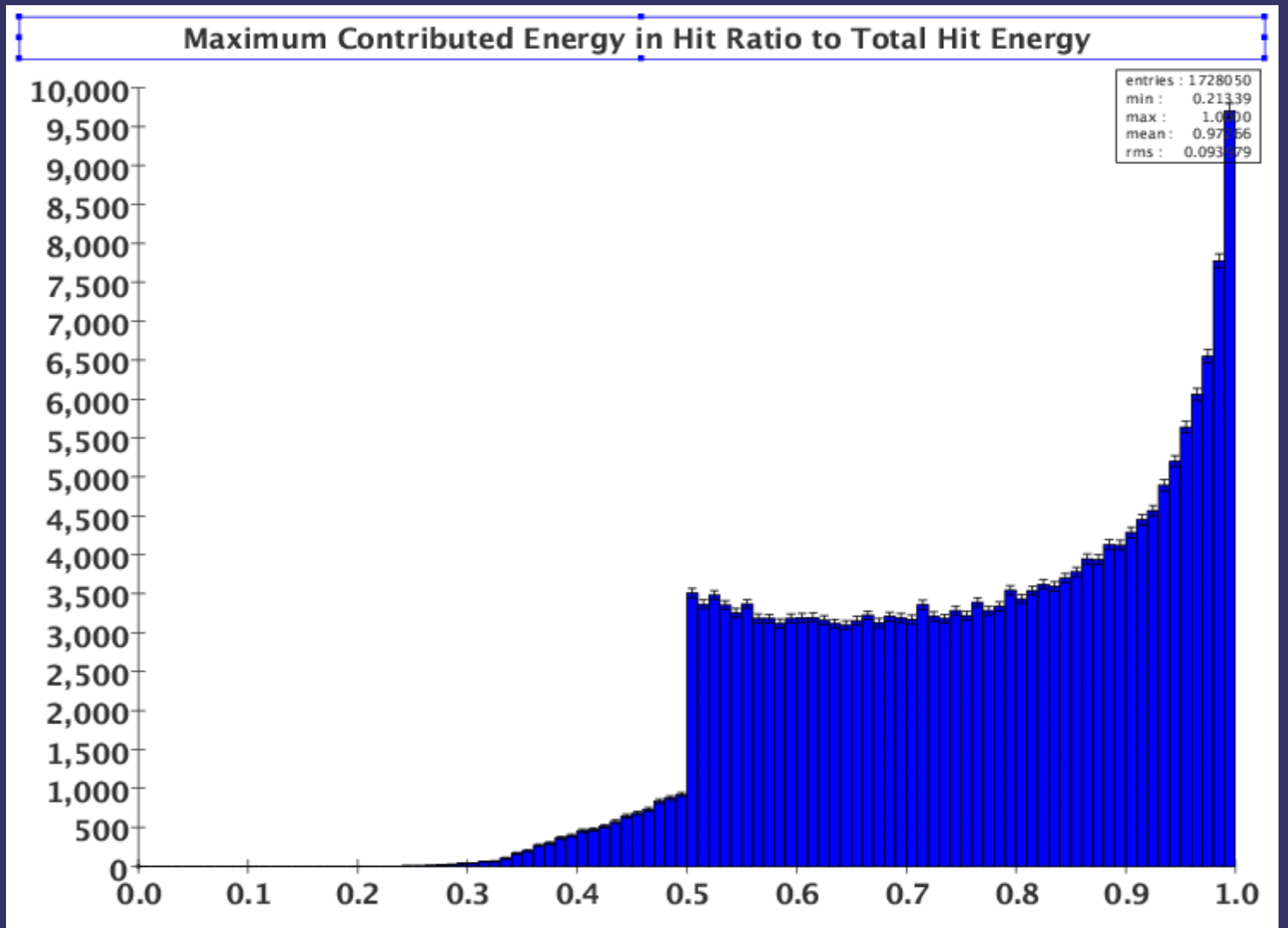




A closer look at the layer by layer hit mapping for SCB and CC reveals the virtual impossibility of a successful nearest-neighbor clustering algorithm, simple or otherwise

Percent contributed energy by Maximum Contributor in Hit

Hit purity for any clustering algorithm is relatively inefficient: resolution studies must be performed in order to determine the minimum threshold to create a cluster.



Conclusion

- Nearest neighbor clustering algorithms are not able to give sufficient resolution
- NEXT:
  - Study of RadialClusterBuilder
  - Combination of current “simple” algorithms and radial algorithm
- New possible clustering algorithms:
  - Seed hits
  - Connection hits