Studies of PFA Fundamentals

Ron Cassell – SLAC

SiD Workshop Jan. 28, 2008

PFA Reconstruction

- Why we need it soon (yesterday)
  - Detector design/optimization
  - Benchmarking
- The template approach
- Studying and improving common pieces
- Progress

Detector design

- Main design issue is the hadron calorimeter. (See next slide)
- Studies have been performed in the context of a Pflow reconstruction. (See Calorimeter R&D report)
- Need a “working” PFA to answer some of the critical questions

From Marty

- HCal Strategic Questions (assuming SiD stays with PFA!)
  - Plastic or gas detectors (i.e. relatively hydrogen rich vs. poor
  - Radiator choice (Stainless steel, Cu (brass?), Pb, W)
  - Number of sampling layers (currently 34 in 4 λ steel)
  - Total radiator thickness.
- SiD still needs to optimize R_{trkr}, $\cos(\theta_{trkr})$, and B, but:
  - Pandora parameterization indicates R_{trkr} and B are ~ok,
  - There is very little technical/cost phase space for making R_{trkr} or B larger.
  - The above questions need to be answered before much serious can be done with global optimization.
- SiD should concentrate on jets appropriate to the energy frontier – e.g ~180 GeV

Detector design (cont)

- In most cases, the sign (better or worse) is relatively easy.
- Since optimization is physics output vs cost, quantitative measurement is critical.
- Even with a high performance PFA, the answers may not be clear cut.

Benchmarking

- The need for physics analyses on full simulations with a full reconstruction is clear.
- Until a full PFA reconstruction that is “good enough” to give meaningful results exists, the analyses can be developed on PerfectPatternRecognition reconstruction.
- This is a true PFA reconstruction, without the confusion term.
- For a description, see http://www.slac.stanford.edu/~cassell/PPRPflow.ppt .

A modular approach

- Sets of nested Drivers, each with well defined input and output.
- Allows replacement of a piece with better performing algorithm.
- Good idea, implementation somewhat tricky.
- What are some of the common pieces?
- How do we measure their performance?

Some common pieces

- Tracking
- Clustering
- Photon finding
- Track/cluster association
- Neutral hadron reconstruction
- Calibration
- DigiSim

Some common performance definitions

- Data sample: ZZ->nunuqq at 500 GeV

  - plot (reconstructed – generated)Zmass

  - measures reconstruction of jet energy and direction (4vector) over a wide range of jet energies

  - eliminates missing energy from prompt neutrinos

  - eliminates jet finding

Common performance definitions (cont)

- The PFA output is ReconstructedParticles. Energy deposits in the calorimeters are assigned to each particle. While “correctness” is difficult to define and interpret on a particle by particle basis, all particles can be divided into 3 classes: charged tracks, photons and neutral hadrons. The origin of the energy deposits is also so divided, allowing investigation of efficiency and purity of the assignments on a particle, event or run basis.
- This is useful for examining the performance of pieces of the algorithm.

Tracking

- Convert hits in the trackers into reconstructed charged particles.
- Not complete, so we cheat.
- Which tracks are reconstructed and how well are important to PFA, so attempt to cheat in a “realistic” manner. (CheatReconDriver)
- Provides common tracklist for all PFA implementations, as well as a common definition of final state particles.
- Not considered a serious stumbling block for PFA development, since we expect actual tracking to closely match cheating.

Calibration

- Interface ClusterEnergyCalculator
- Implementations for photons and neutral hadrons
- See http://www.slac.stanford.edu/~cassell/Calorimeter_calibrations_acme.ppt for details.
- Will need to be redone if PFA results become sensitive

Clustering

- Rather than deal with the thousands of energy deposits in the calorimeters individually, we group the hits together.
- A variety of clusterers are available (all adhering to a Clusterer interface)
  - NearestNeighbor
  - FixedCone
  - MinimumSpanningTree
  - DirectedTree
  - Cheater
  - VariableCone (coming soon, work by Qingmin Zhang)
- Purity and efficiency are the big performance issues. The purity limits the jet resolution (introduces confusion) and the efficiency affects how well the association can be done.
- A variety of diagnostic packages exist for examining performance, but perhaps the most telling is the ClusterLevelCheater

ClusterLevelCheater

- Assign each cluster to the FS particle with the most energy in that cluster.
- Procedure:

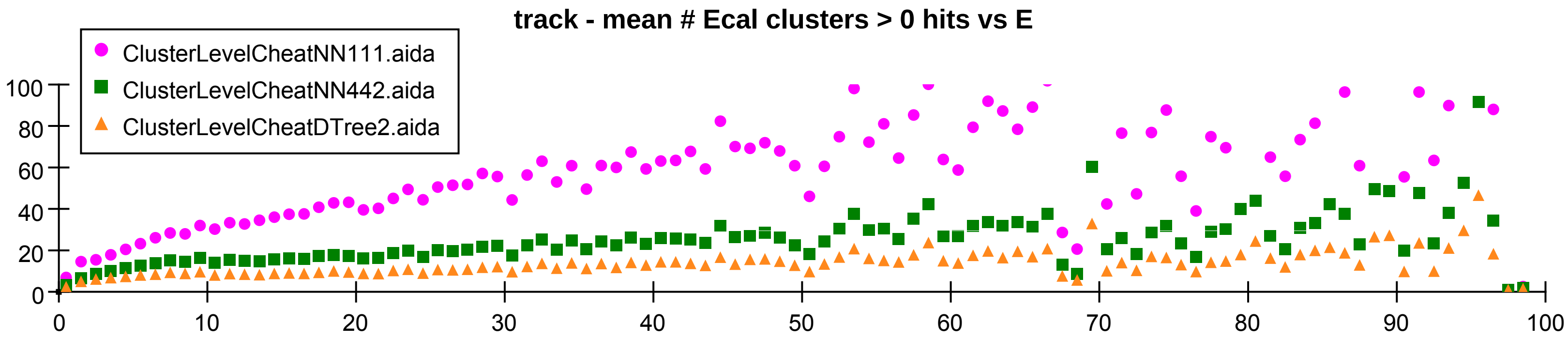
  - Use the Sid01 detector

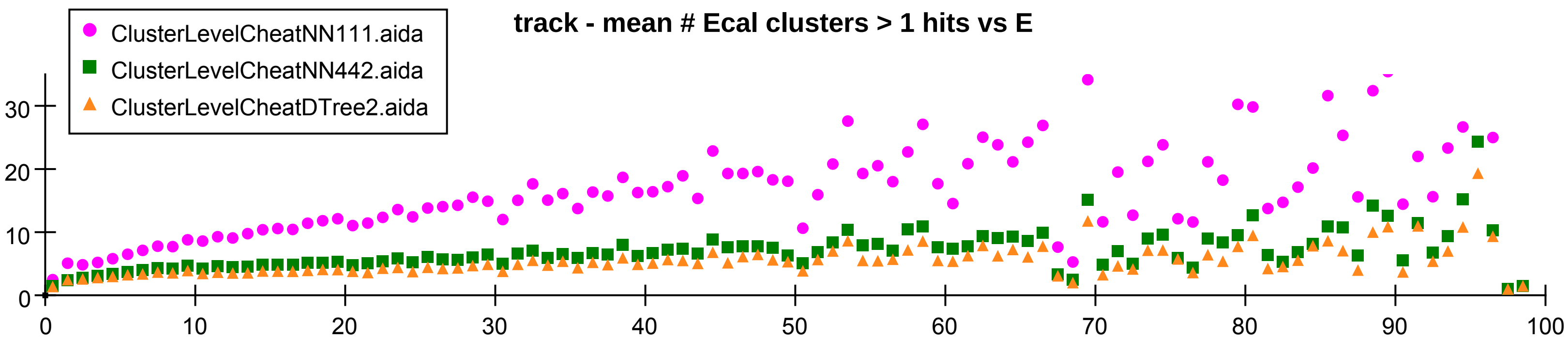
  - Use our favorite ZZ->nunuqq dataset

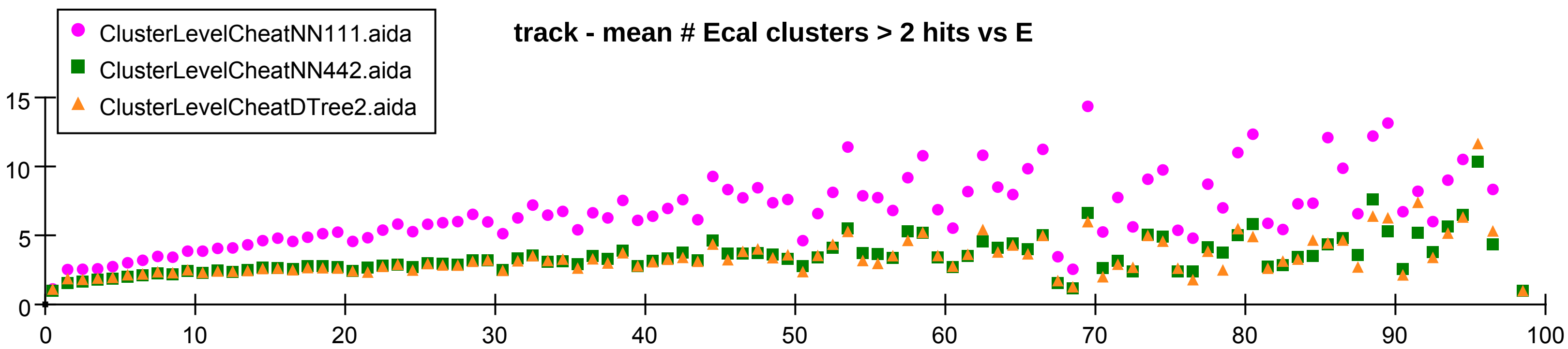
  - For each clusterer, use a CoreReclusterer (described later)

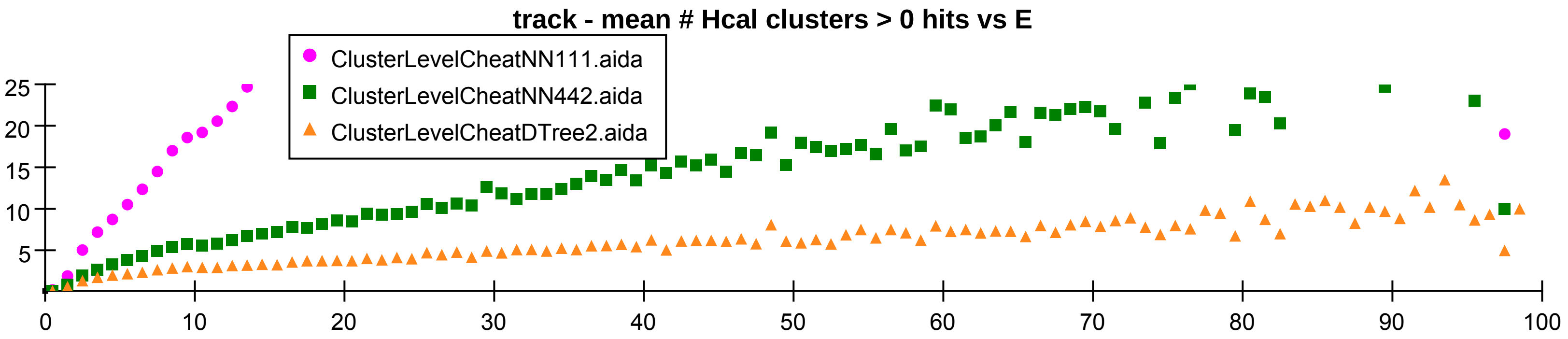
  - Make ReconstructedParticles as in the PPR

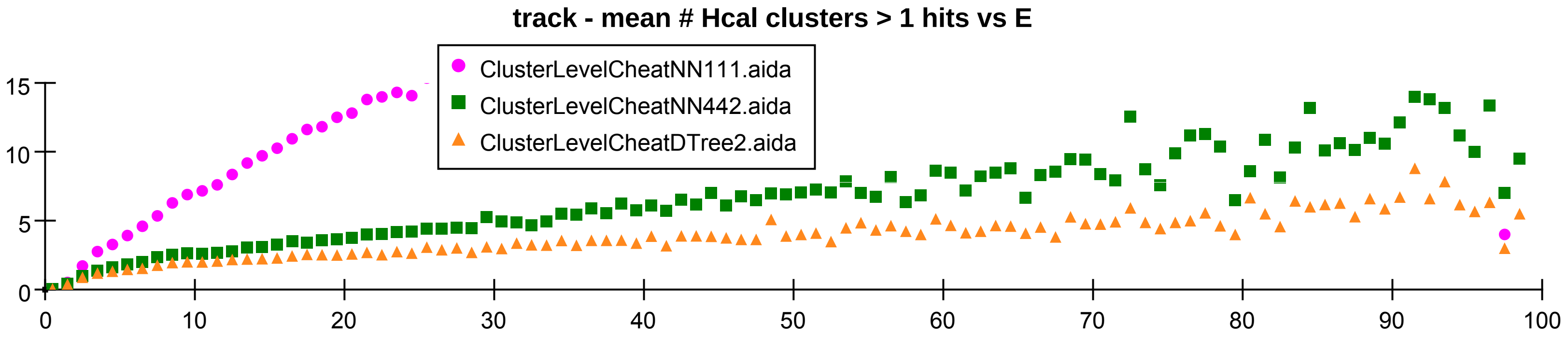
  - Look at the results

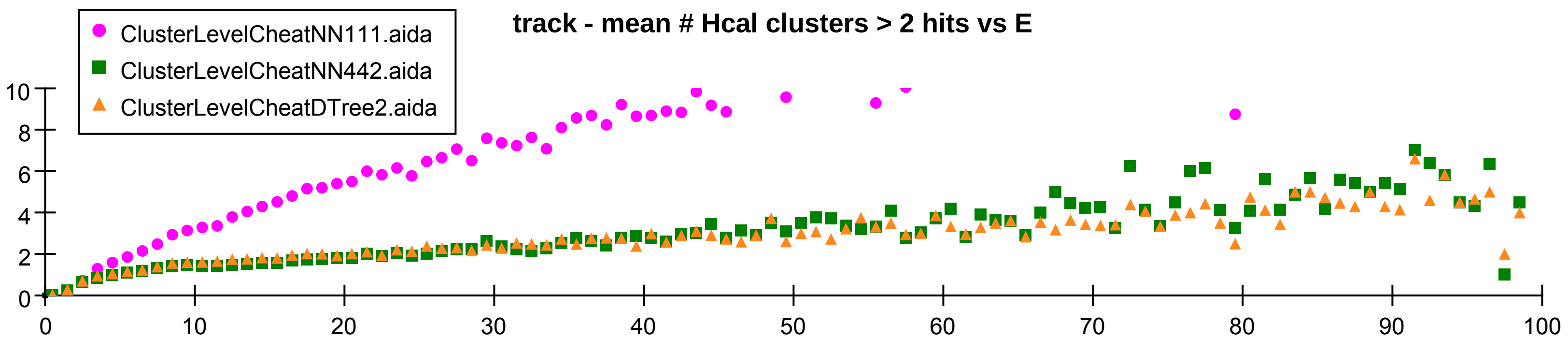

track - mean # Ecal clusters > 0 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
20
40
60
80
100
0 10 20 30 40 50 60 70 80 90 100


track - mean # Ecal clusters > 1 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
10
20
30
0 10 20 30 40 50 60 70 80 90 100


track - mean # Ecal clusters > 2 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
5
10
15
0 10 20 30 40 50 60 70 80 90 100


track - mean # Hcal clusters > 0 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
5
10
15
20
25
30
40
50
60
70
80
90
100


track - mean # Hcal clusters > 1 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
5
10
15
20
30
40
50
60
70
80
90
100


track - mean # Hcal clusters > 2 hits vs E
ClusterLevelCheatNN111.aida
ClusterLevelCheatNN442.aida
ClusterLevelCheatDTree2.aida
0
2
4
6
8
10
20
30
40
50
60
70
80
90
100

(Reconstructed – Generated) Zmass

	rms90(GeV)	mean90(GeV)	Rms90/(91.2+mean90)
PPR	2.597	-.497	2.86%
NN442	3.727	-.492	4.11%
NN442 > 1hit	3.676	-2.027	4.12%
NN442 > 2hit	3.719	-2.721	4.20%
NN111	2.834	-.524	3.13%
NN111 > 1hit	3.144	-4.497	3.63%
NN111 > 2hit	3.636	-6.536	4.29%
DT	3.125	-.333	3.44%
DT > 1hit	3.076	-.946	3.41%
DT > 2hit	3.084	-1.241	3.43%

Clustering (cont)

- The previous study was done before discovering a bug in the cheat reconstruction (bad programmer). This has been fixed, and while it affects the absolute numbers, should not affect the conclusions.
- What are the conclusions?

  - If you use only 1 clusterer, the DT is clearly superior.

  - If you use DT clusters, a good result can be obtained throwing away the small clusters. In fact, fix the bug and see how far we can take this.

(Reconstructed – Generated) Zmass

	rms90(GeV)	mean90(GeV)	Rms90/(91.2+mean90)
PPR	2.402	-.503	2.65%
DT > 0hit	2.787	-.491	3.07%
DT > 2hit	2.769	-1.384	3.08%
DT > 5hit	2.785	-1.977	3.12%
NN442 > 5hit	3.654	-3.880	4.18%
Mat	4.87	-5.37	5.67%

Calorimeter energy run summary

PPR	efficiency	purity	2.65%
Charged	.999	.999	
Neutral	.998	.998	
Photon	.999	.999	
DT > 5hit			3.12%
Charged	(.904) .966	.964	
Neutral	(.790) .866	.895	
Photon	(.972) .981	.968	
NN442 > 5hit			4.18%
Charged	(.846) .949	.957	
Neutral	(.711) .837	.820	
Photon	(.947) .983	.977	
Mat			5.67%
Charged	(.816) .894	.870	
Neutral	(.519) .581	.530	
Photon	(.763) .778	.878	

Photon finding

- Should be the easiest part
- A photon finder was written for Sid01.
- It isn’t very good, with ~ 85% eff and pur
- Was better than existed, used by Mat.
- Basic problem is overlapping photons from pi0s, spread the shape distributions, and make separation from hadrons difficult.
- New code exists, using a CoreReclusterer (next slide) to separate clusters by finding energy cores.
- Can get to 90% eff and pur, but to do much better will require integration of track/cluster matching.

CoreReclusterer

- Uses algorithm suggested by Norman.
- Within a cluster, raise the energy threshold of the hits and see if it separates into 2 or more "cores".
- If so, add all hits not in a core to the nearest core.
- Minimum energy and #hits requirements to be called a core are settable parameters, and have not been optimized.

Progress

- New results from our PFA development and from using PandoraPFA are in the next talks.
- The perfect pattern recognition PFA appears to be ready for use in benchmarking, with a test sample of Zhh events at 500 GeV processed and ready to look at. (as of yesterday)
- Tools for analyzing pieces of a PFA are providing some guidance as to how to proceed.

Summary

- We need a PFA reconstruction performing well enough to quantitatively measure physics performance vs detector parameters.
- We're not there yet.
- The template approach may speed up the process.
- The constraints on efficiency and purity for matching tracks to calorimeter hits are quite severe.
- Use the DirectedTree clusterer, at least to start.