PFA and combinatorics

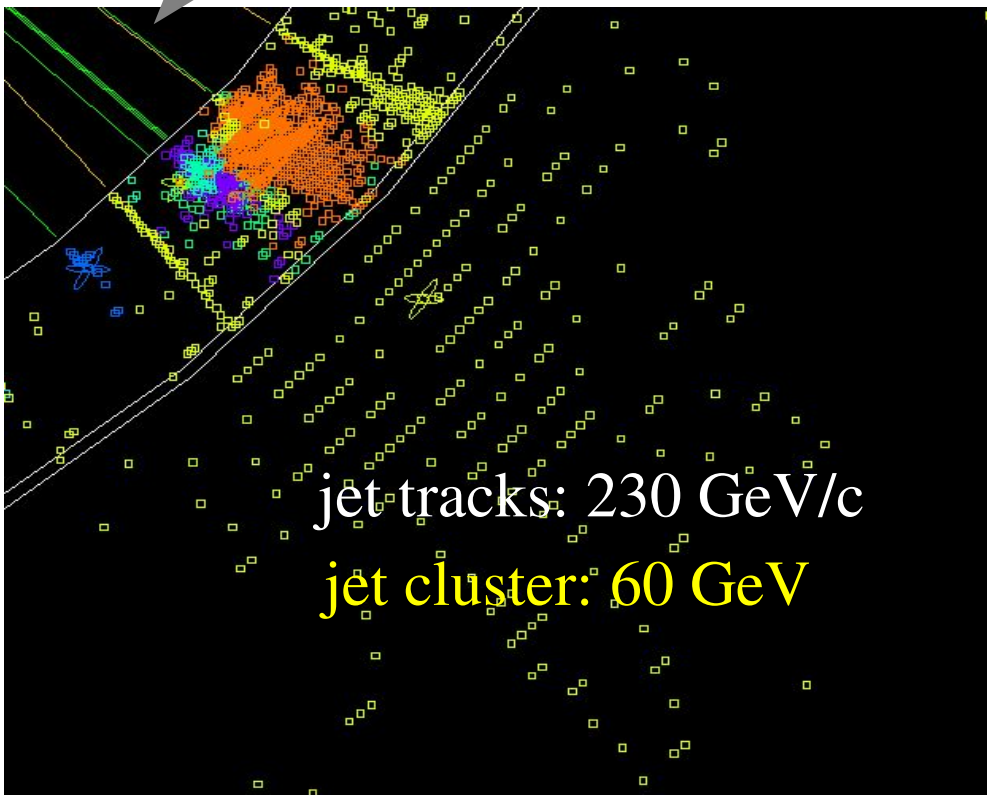
- Motivation
- Simulated annealing, traveling salesman problem
- Comparison to IOWA clustering
- Implementation on top of main clustering
- Further plan
- Conclusion

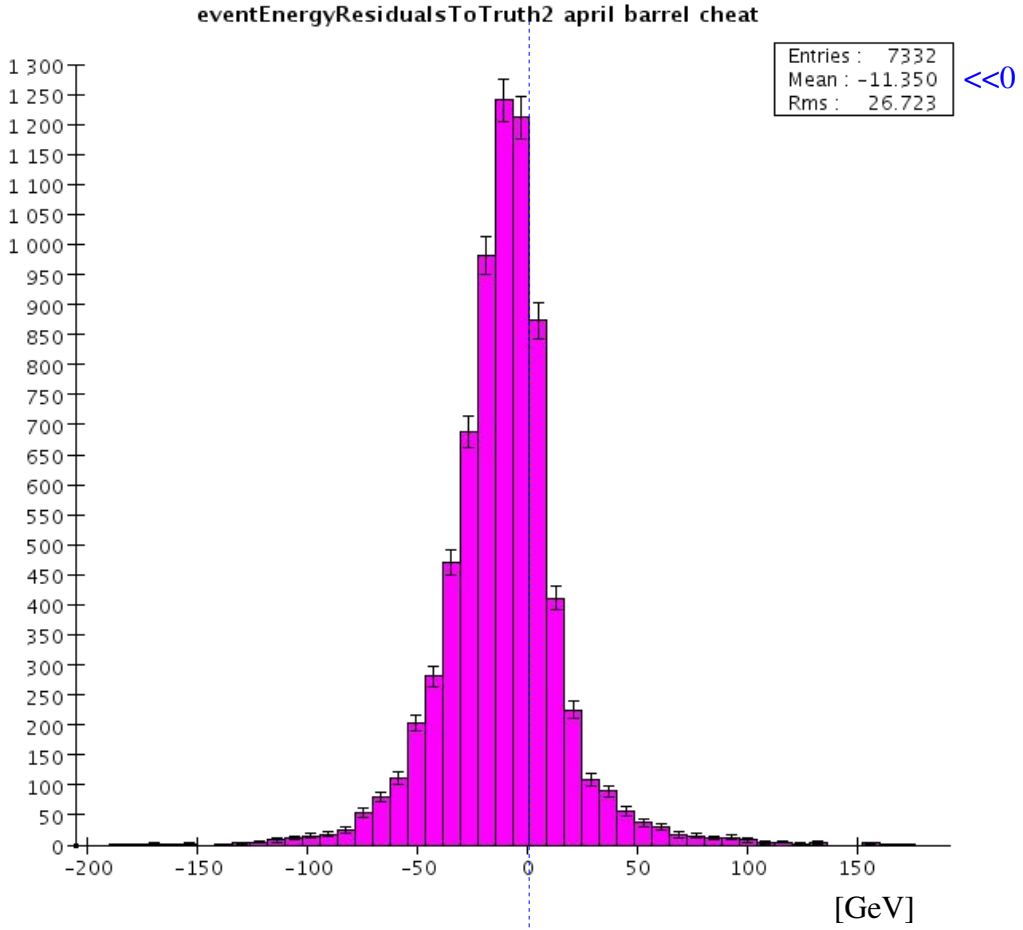
Motivation (500GeV qq)

(04/08/2010)

Event energy residuals from newest code version:

183GeV ch +fake 169GeV n





Simulated annealing

Physically motivated method to quickly approximate the global minimum of a function ("cost") of a discrete system.

Variant: Threshold Accepting (TA).

choose initial configuration;
for (threshold $T = T_0 \geq T_1 \geq T_2 \geq \ldots \geq T_n = 0$) {
    choose new configuration which is small perturbation of the old configuration;
    if (cost(new configuration) < cost(old configuration) + T)
        old configuration = new configuration;
}

(journal of computational physics 90, 161 1990)

Travelling Salesman Problem (TSP)

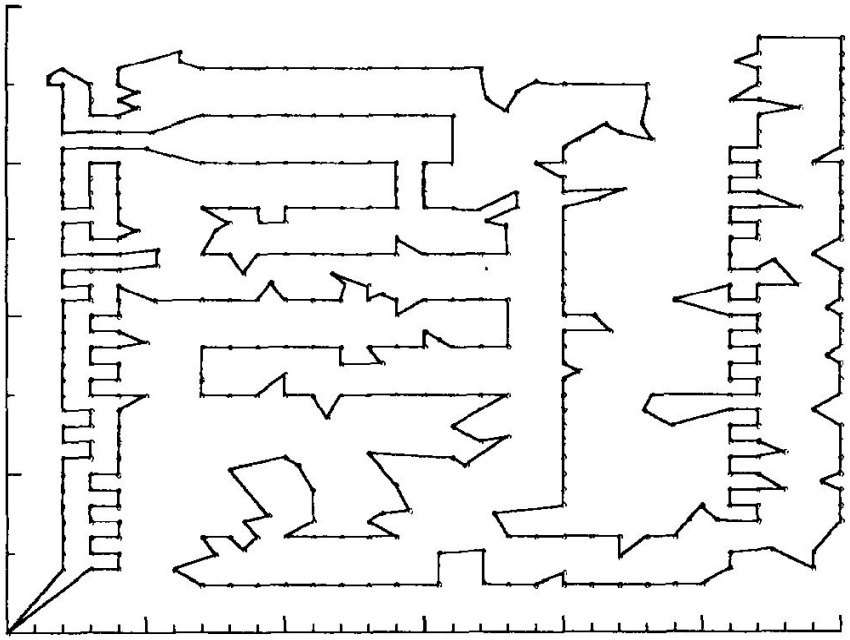


FIG. 1. Optimal tour for the 442-TSP by Holland. This tour has length 50.80.

Start in some city, visit every city exactly once going straight lines and return.
What is the shortest loop?

- "NLP complete": 10^6 $ for fast *exact* solution
- 1989: 442-cities TSP (~200'000 paths) solved *to excellent approximation* in seconds using TA
- Implemented small TSP: Simple, combinatorics "fitted"

Comparison to IOWA clustering

- 500 GeV qq: ~170 elementary clusters, 20'000 potential links (counting both directions!)
- Different goal: Find group of clusters, each containing preferably one seed cluster
- Complicated cost function, dependent on energies, momenta, DOCAs, proximities, ...
- Both problems are “global”, i.e. status of every possible link counts

Implementation on top of main clustering

For now: work on isolated showers, don't merge into PFA “jets”

Threshold Accepting

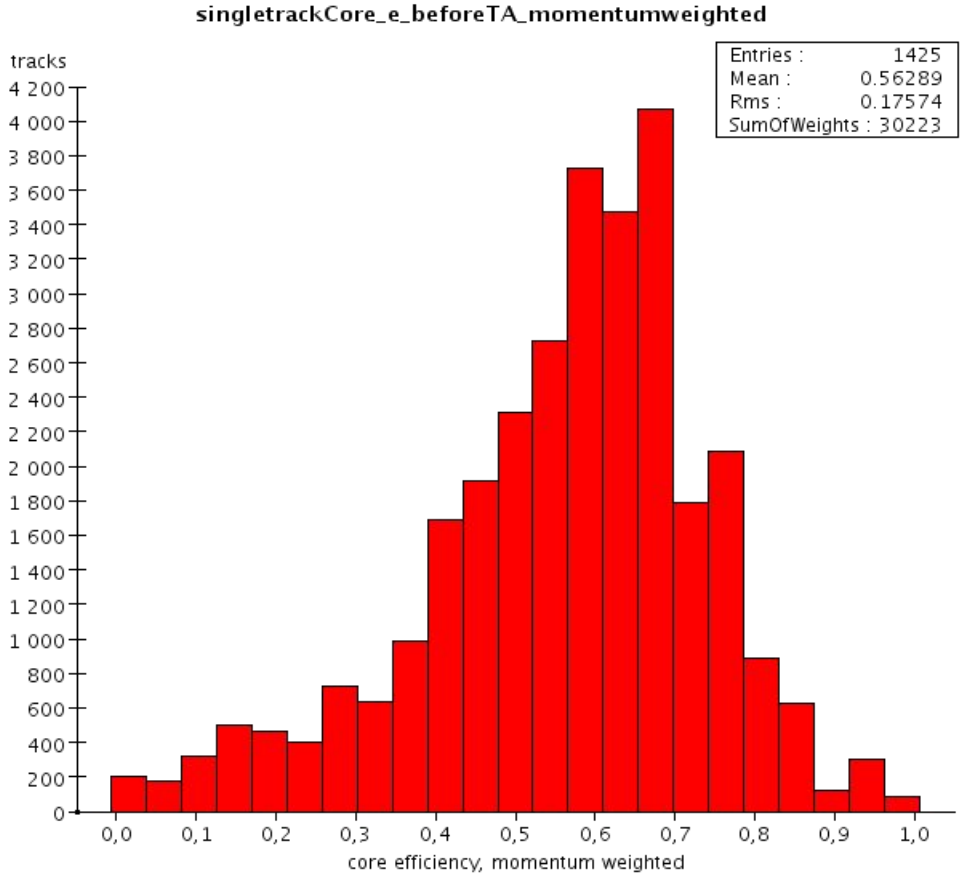
- Initial configuration: Result of performing clustering iterations
- Thresholds T = 9 (5 times), 8 (5 times), ... , 0 (5 times)
- Choose new configuration which is small perturbation of the old configuration: Randomly
  - exclude one elementary cluster from the shower (not the seed!)
  - or put one back in
- Cost function: Abs(E_{shower}-$P_{matched\ track}$)

this is not a PFA, but let's see what happens...

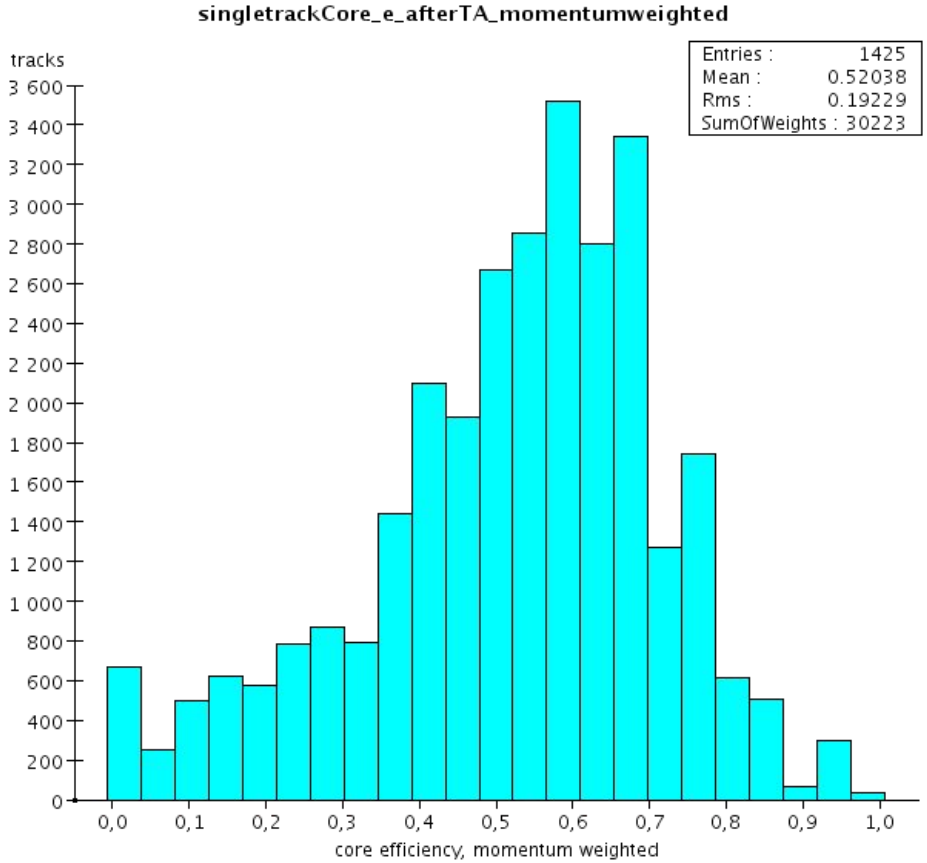
Implementation on top of main clustering

Momentum weighted core efficiency:

before TA



after TA

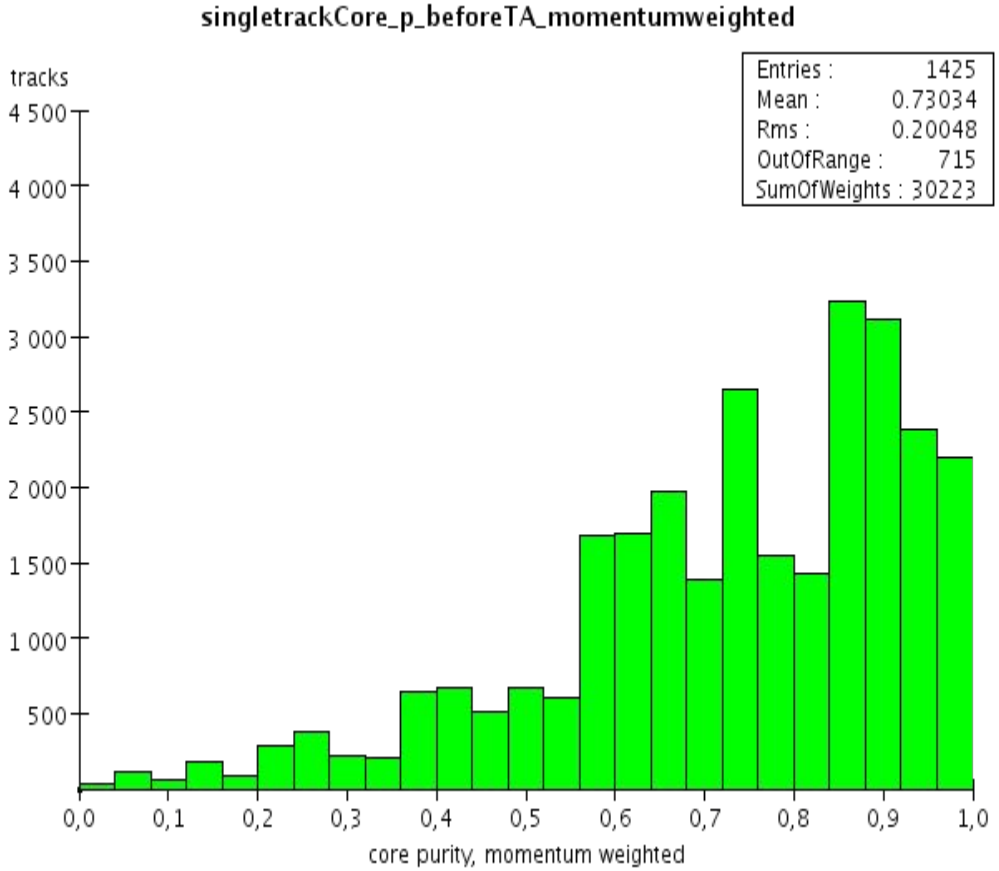


slight decrease, as expected.

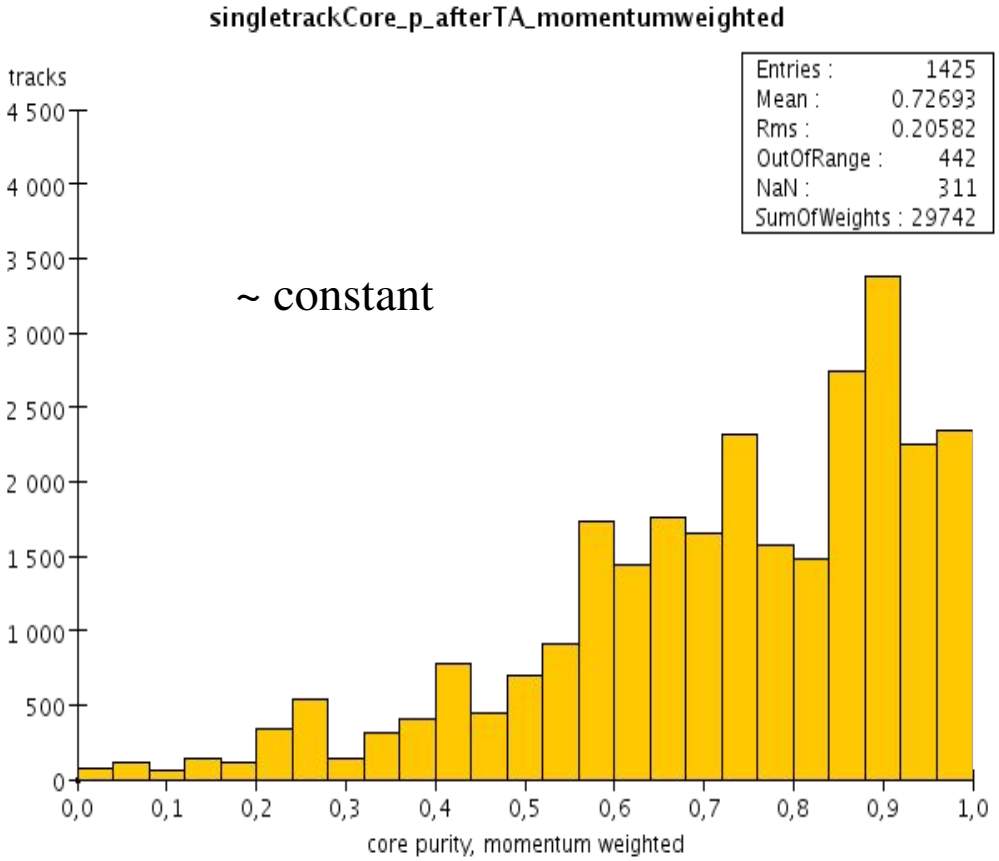
Implementation on top of main clustering

Core purity, momentum weighted:

before TA



after TA



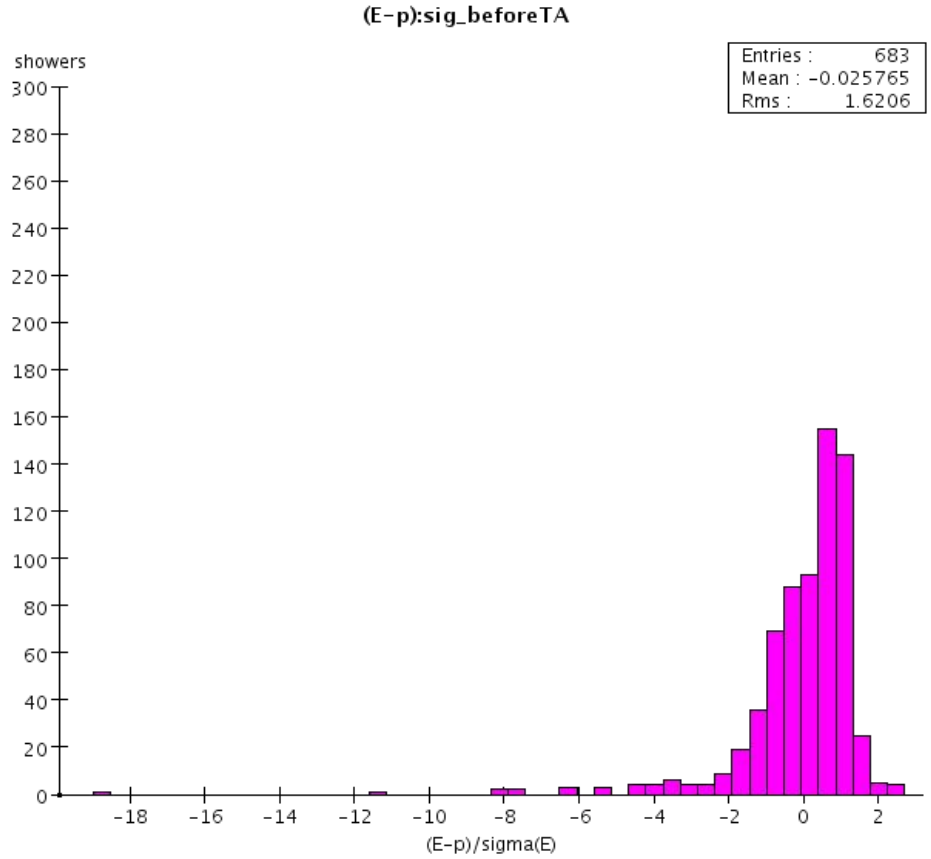
~ constant

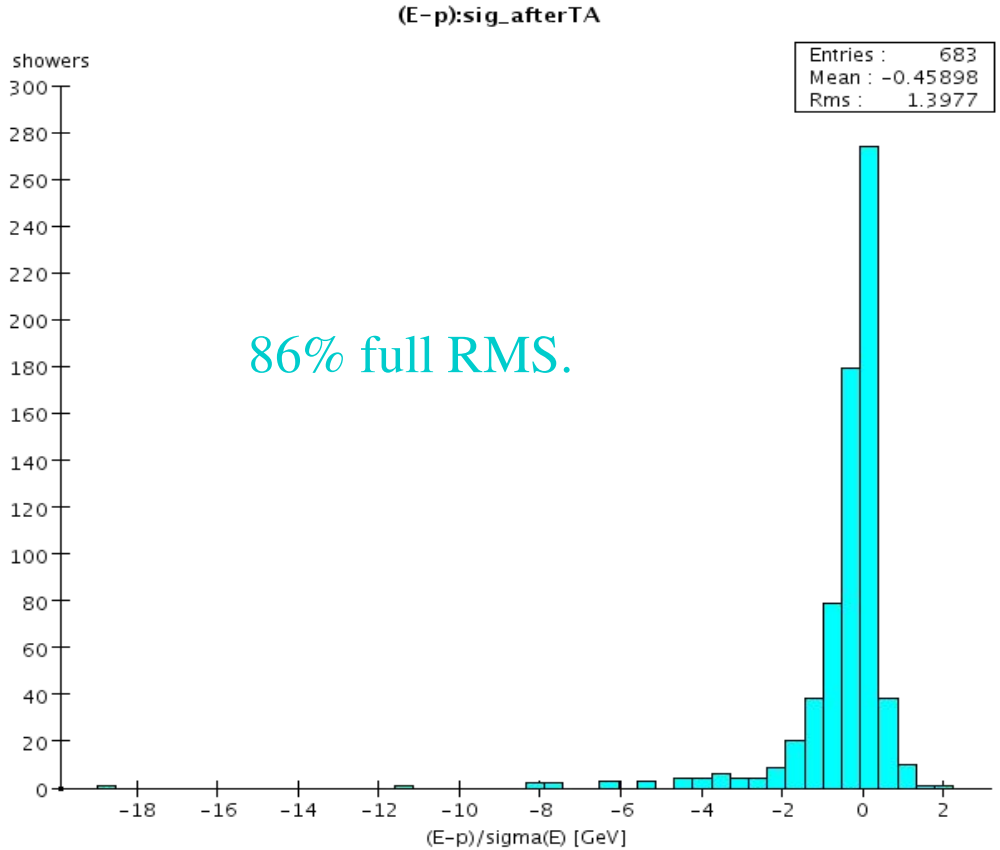
Physics beyond efficiency, purity: Checked 10 events in file -2.

Particle flow not deteriorated.

Implementation on top of main clustering

Energy resolution: before TA after TA





(didn't even process all showers)

Further plan: Extend application

- Increase purity by cost function depending on E, p, link-scores
- Further outward: Overrides, 2nd cone algorithm still necessary?
- Process multiple showers; allows resolution calculation from full PFA
  - Known problems at this point
  - Combinatorics harder: heuristics!
  - Study running time, optimize data structures
- Emulate, improve main clustering

Conclusion

- Confusion is of utmost importance for PFA resolution
- Good track\cluster balance gives low confusion
- As PFA complex: Some fit afterwards seems necessary to achieve this
- Simulated annealing is a way to do this