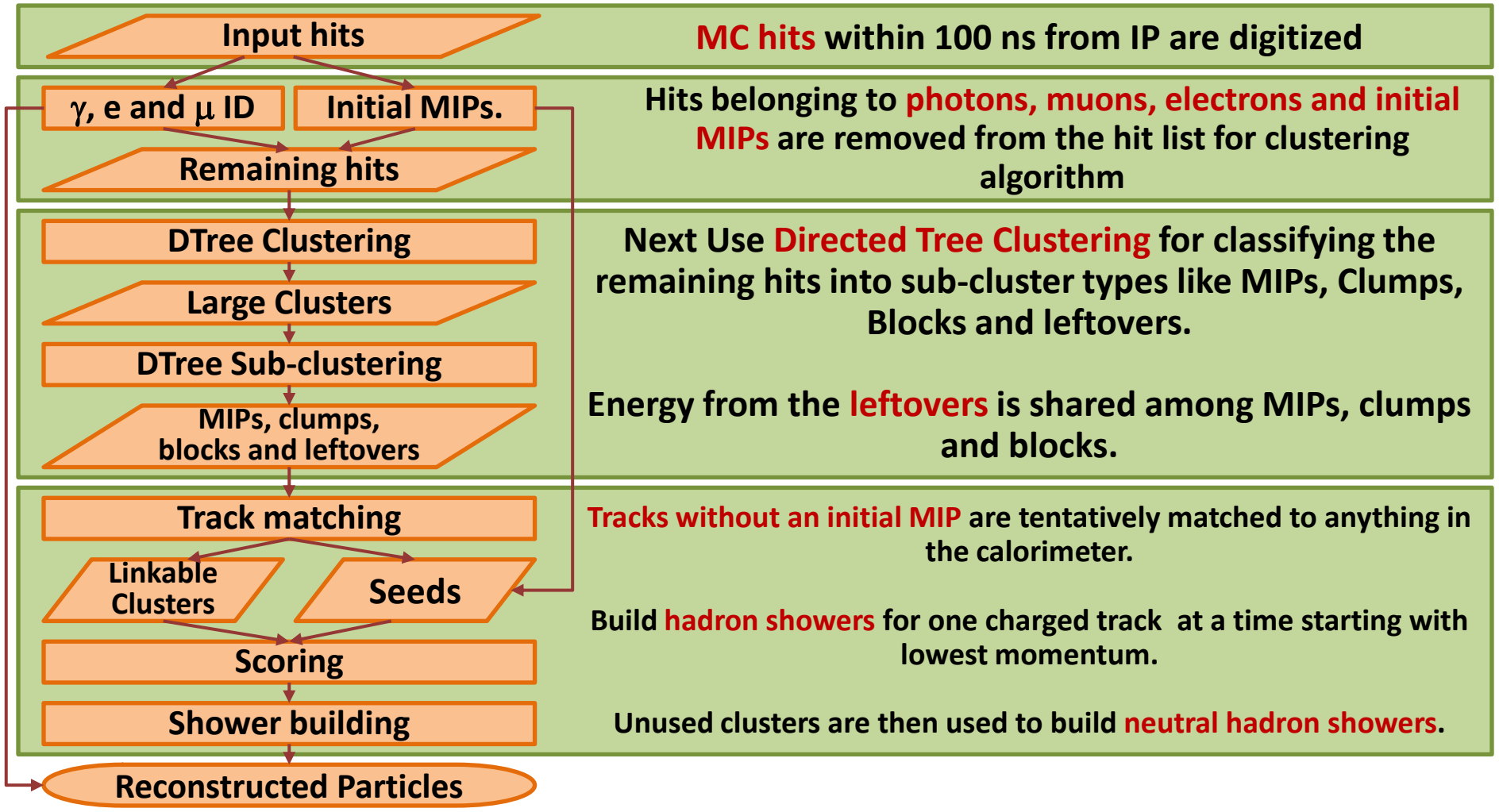
Status of the SiD/Iowa PFA

R. Cassell, M.Charles*, U. Mallik, Remi Zaidan

Nov 15, 2010

SiD Workshop, University of Oregon, Eugene

Basic Building Blocks of the (Iowa) PFA



Categorizing: DirectedTree Clustering

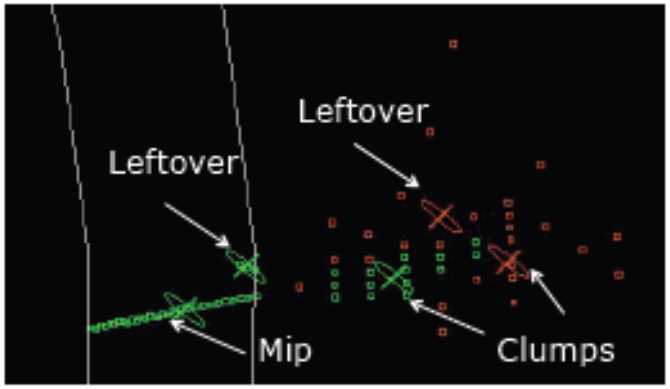
Ecal Digi Hits(Barrel, Endcap)			Hcal Digi Hits (Barrel, Endcap)		
Photon, Electron, Muon	DTree cluster (Ecal Barrel)	DTree cluster (Ecal Endcap)	Muon	DTree cluster (Hcal Barrel)	DTree cluster (Hcal Endcap)

DTree cluster

MIPs	A continuous sequence of single hits
Clumps	Group of hits with high density
Blocks	No structure, if(>= 20hits in ECal, >=15 hits in HCal)
Leftover	No structure, small number of hits (Share with others)

- Leakage

Some of high energy shower escapes Hcal, reaching Muon Detector. Adding the energy by using Muon Endcap as tail catcher give better resolution. (Currently not using Barrel)



Cluster Building

- Extrapolate (each) track to the ECAL surface
- FindSeed: sub-cluster directly connected to extrapolated track (other than MIPs)
- Each track typically has one seed
  Special cases: track without seed, or when it does not reach the calorimeter
- Now start connecting other sub-clusters to the seed of each track
- Start with lowest and then progressively higher momentum tracks
- Up to ten iterations until all track-cluster match satisfy (E – p) within tolerance

Connecting sub-clusters

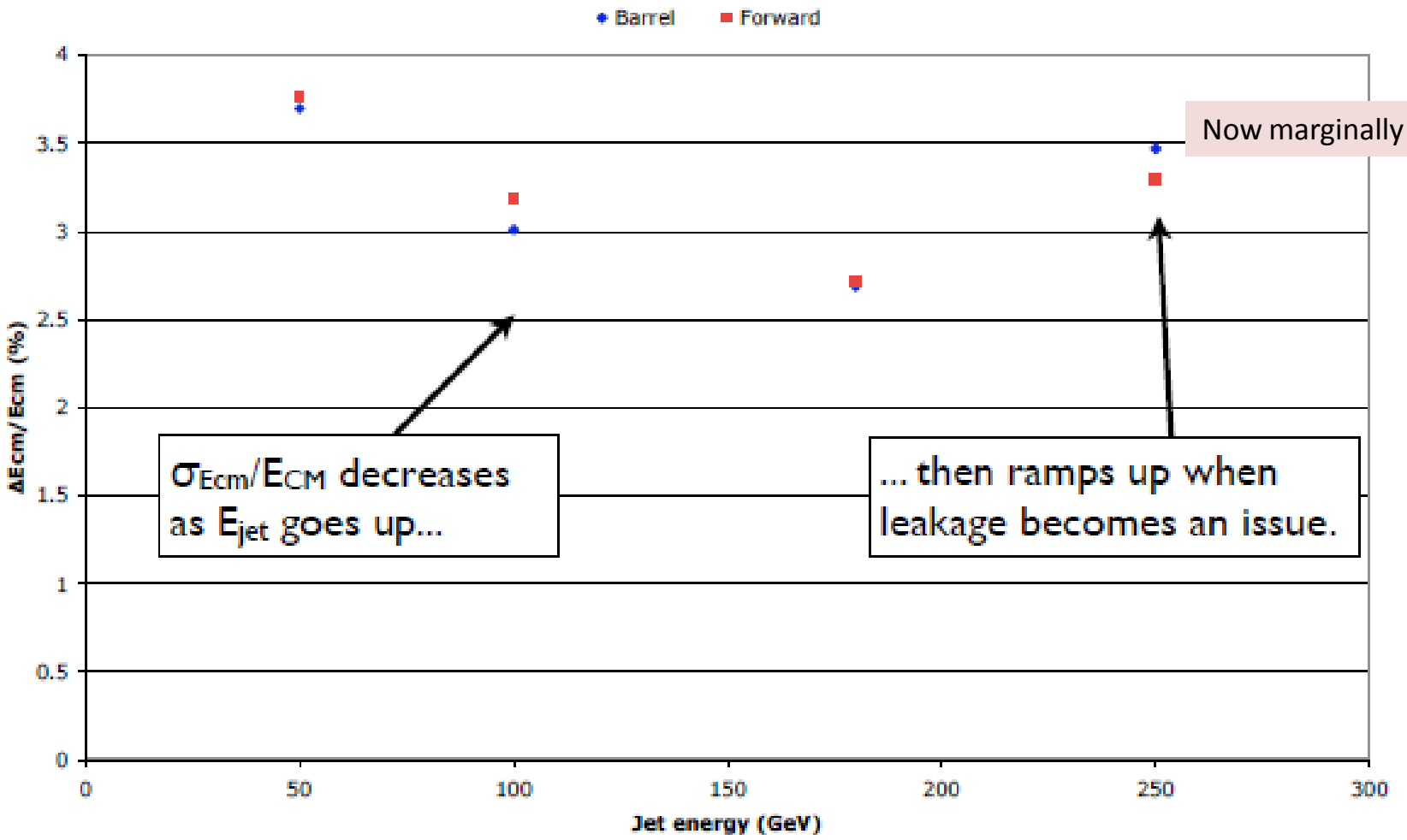
Scoring : (a poor man's) Probability of a link

Based on the sub-cluster type and geometric proximity a score between 0 and 1 is assigned between any two sub-clusters starting with the cluster in consideration
The higher the score the higher the probability of a link
To pick up secondary neutral hadrons, a cone-like algorithm is used

A cut-off threshold is obtained for an energy by tuning with events

Energy dependence

Performance at LOI, 4/2009



Now marginally improved

At higher energies resolution worsened partly due to leakage but also due to algorithm

Intermediate history

Threshold Accepting Method applied by Christoph Pahl
Showed initial promise; but ambitious

Zaidan took a good look at the PFA to understand in depth

- improved cone algorithm and applied a likelihood
- fixed several bugs
- modified the iterations, relaxed E/p chceks
- allowed sharing of hits by multiple tracks
- modified track-seed matching, using direction of track
- removed primary neutral hadron clusters
  to test algorithm for only charged tracks

No magic bullet found

Modification of each single step was being foiled by some later steps because on interdependence and built-in "cures" in the algorithm

Conclusion : No obvious "simple" solution

Meeting of the minds:

at end of September at CERN for a few days (Cassell, Charles, Mallik, Zaidan)

Plan of attack:

Develop diagnostic tools to measure success/failure quantitatively

Starting from the top:

Test sub-clustering immediately after Directed-tree clustering

Is sub-cluster purity good enough ? Evaluate

Are extensions of tracks and MIPS done with high efficiency ?

Is scoring done well enough ? Should use same algorithm

Is Photon finder efficiency good enough?

Is the Photon veto good enough (overlap with initial MIPs (seeds) or muons)

Is calibration good enough?

Where can we gain most : Where is the biggest problem (upstream)

Frequent meetings as needed

Diagnostics Tools: (Zaidan)

- track-seed matching
  match-quality, properties of unmatched tracks

- Dtree sub-cluster qualities and performance,
  purities and energy dependence

- link properties
  variables used in scoring (before and after first cone)

- shower properties
  efficiency, purity, energy-momentum comparison

Completed in early October

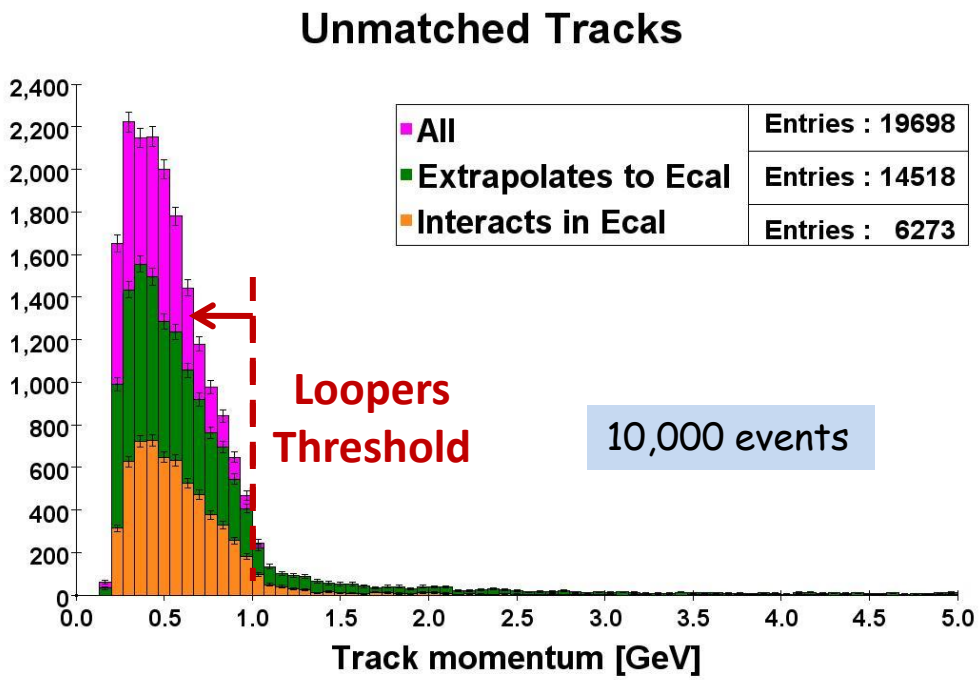
Some results from the studies:

Track-seed matching:

most of the seeds are MIPS

Seed	Match-rate
MIPs	84.5%
Leftovers	14%
Other	1.5%

Seeds with < 4 hits are 7.5%



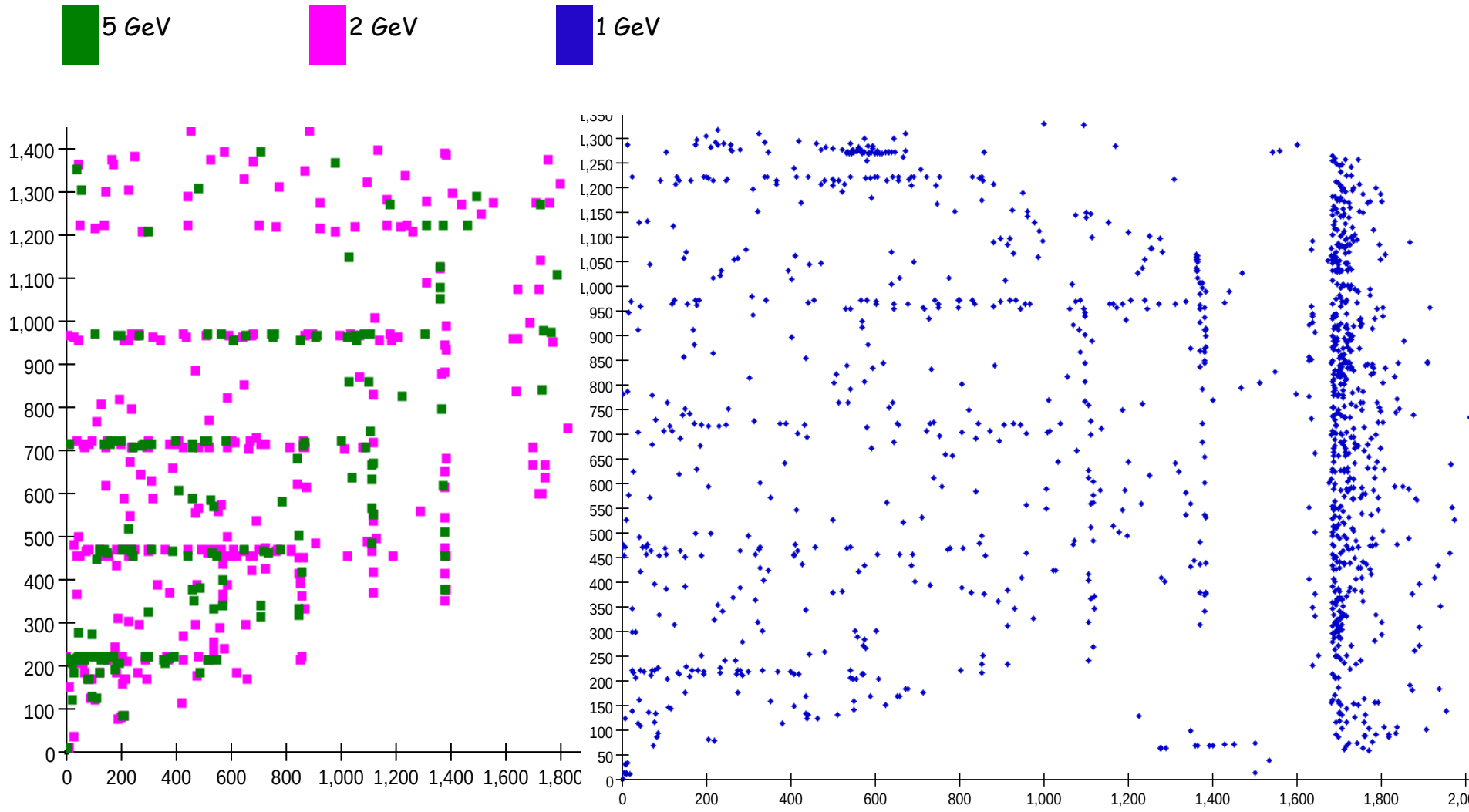
Fraction of unmatched tracks $\Rightarrow$ 18%

Of these most are loopers, fraction with p < 1 GeV $\Rightarrow$ 90%

Of these 18%, fraction which should reach Ecal $\Rightarrow$ 74%

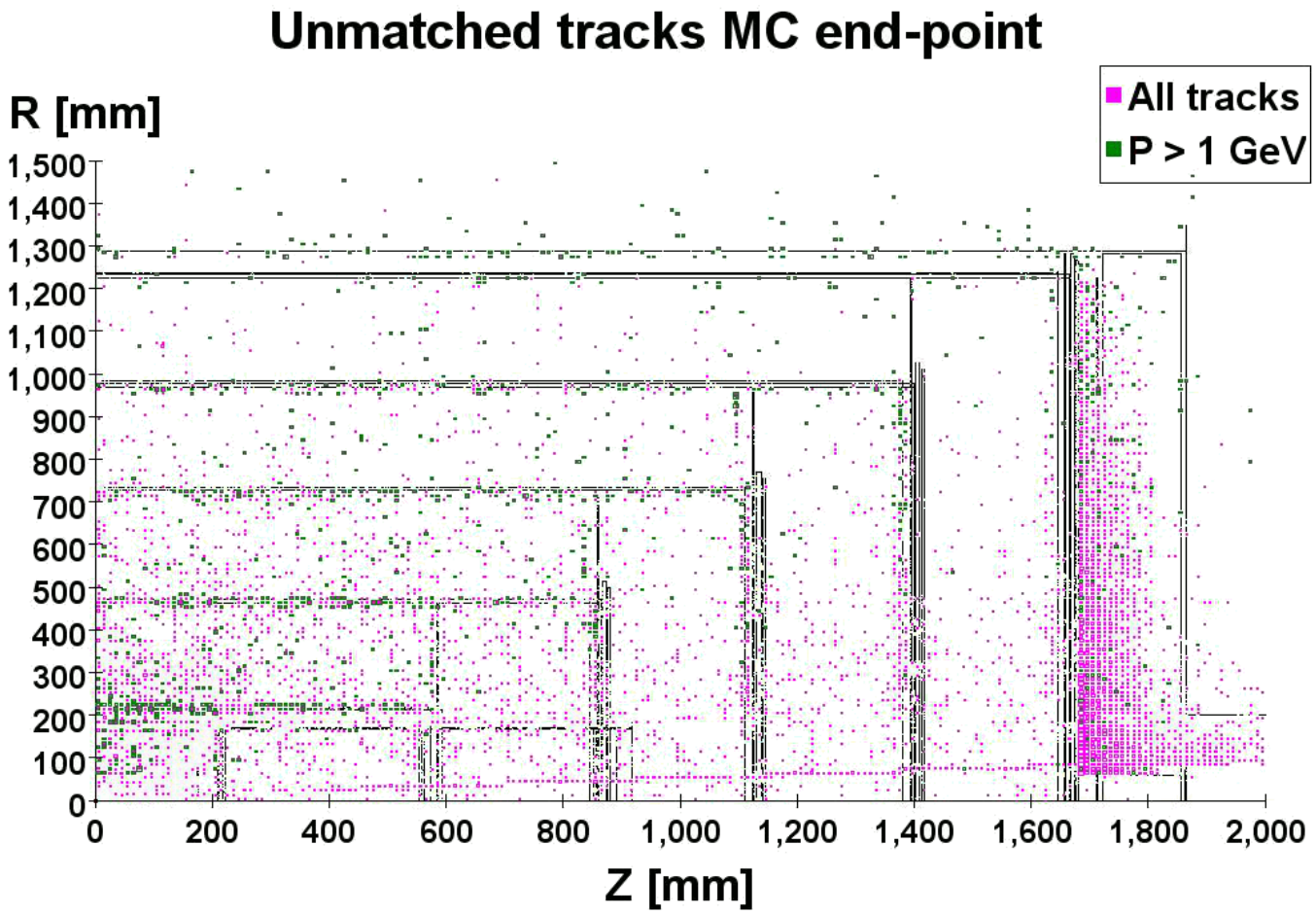
Only 32% actually reaches Ecal (int in tracker)

Few electrons get tagged as photons when there is a miss in match between a seed and a near-by track (0.03%)



End points of tracks with single pions (Cassell) at several energies

Track-Seed Matching



qqbar at 500 GeV, 500 events

Compare with single pions

track-seed matching looks better for single pions at a first glance

	Efficiency	Matching to MIP
qqbar	82%	84.5%
Single pion	99%	91.3%

Integrated spectrum

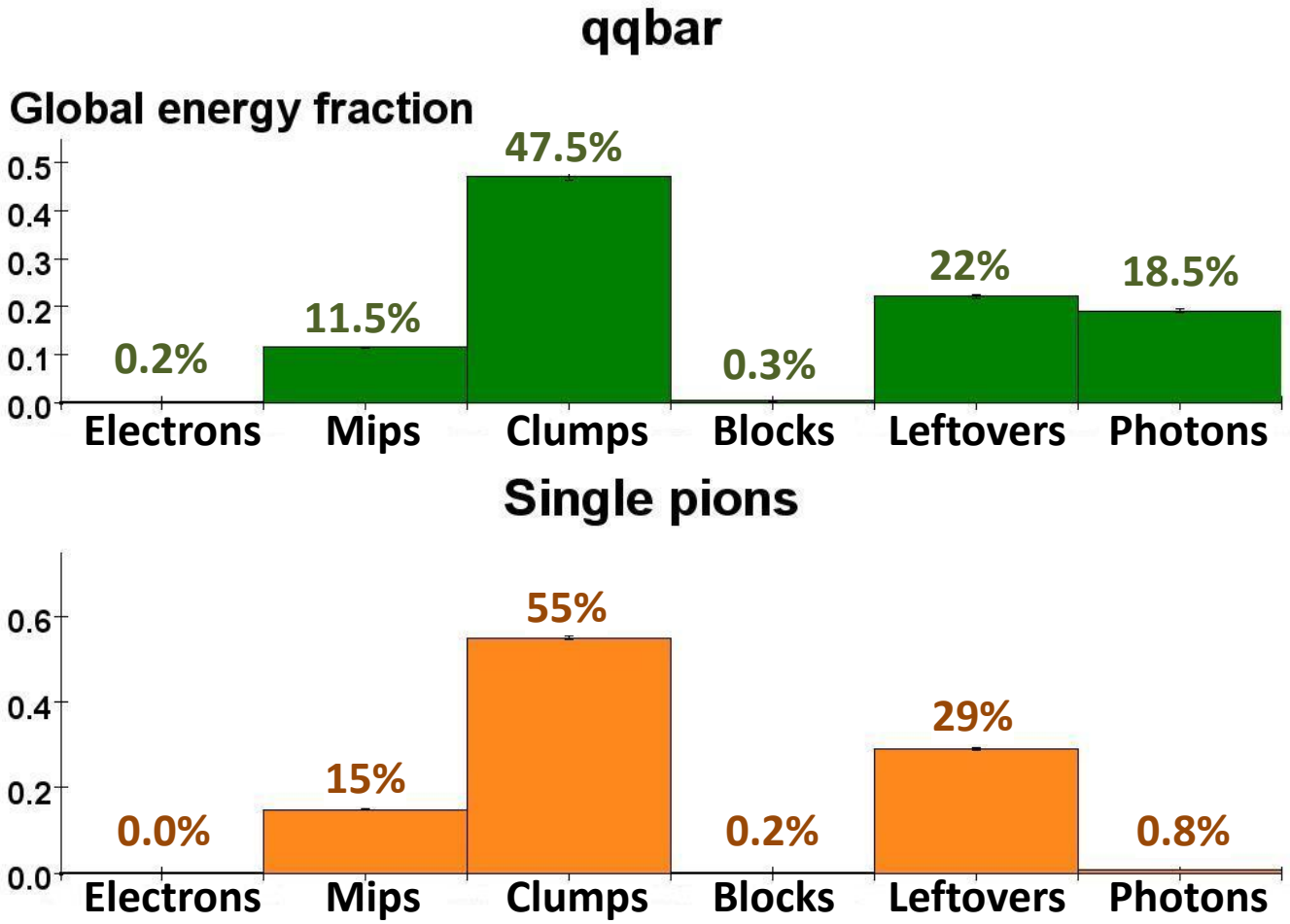
Momentum dependence between the qqbar and the single pion samples are different, no difference once it is properly accounted for

Dividing up into sub-clusters, comparison

energy fractions in total sample

Excess in clumps in qqbar w.r.t. single pions.

Similar distributions: no photons in single pions.

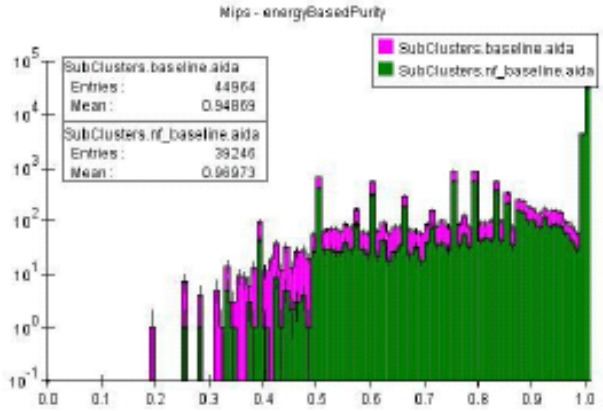


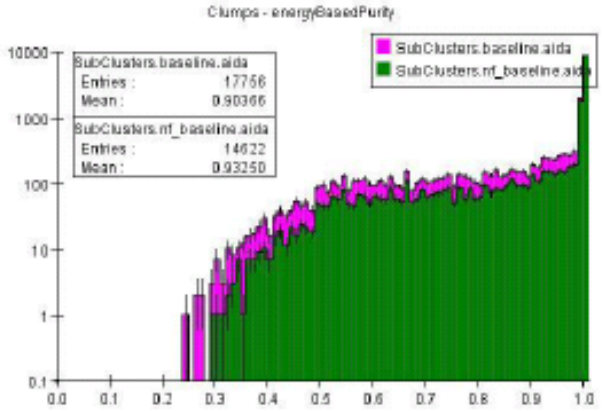
Sub-cluster purity, not energy weighted

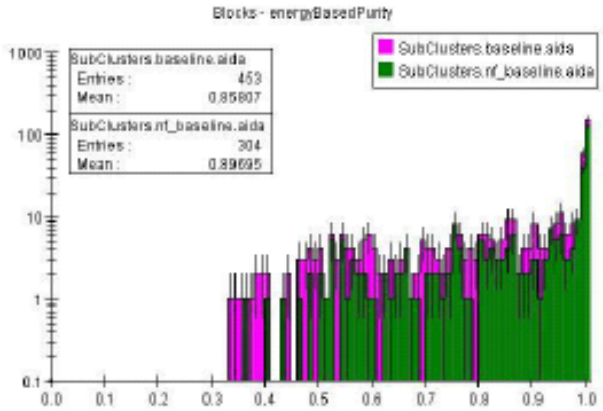
Are neutral hadrons causing a big problem ?
Compare performance with and without

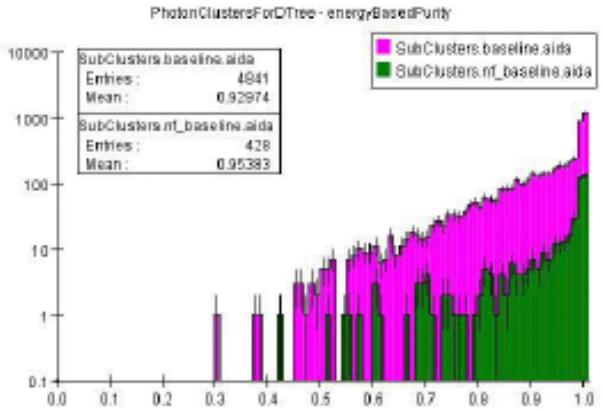
Purple: baseline
Green: Neutral filter applied

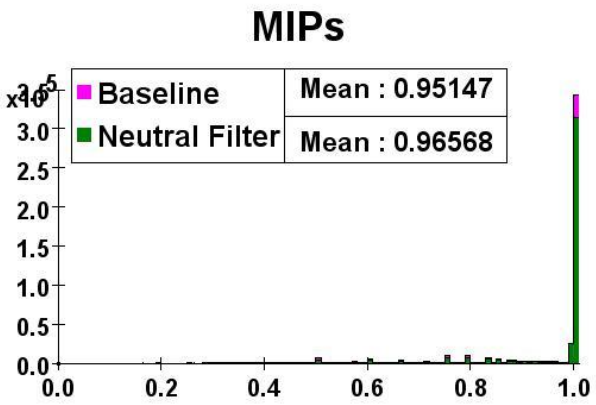
PFA	Baseline	Neutral filter
Mip	95%	97%
Clump	90%	93%
Block	86%	90%
Photon	93%	95%

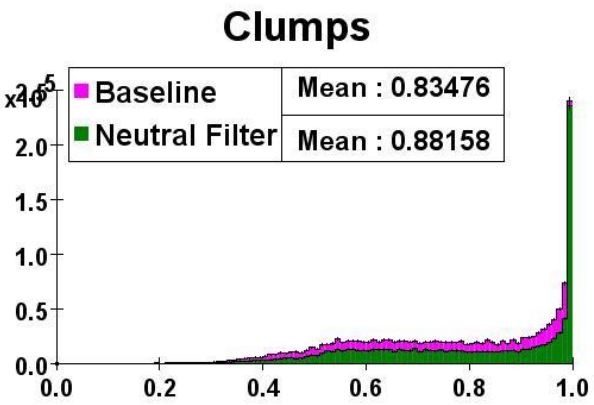




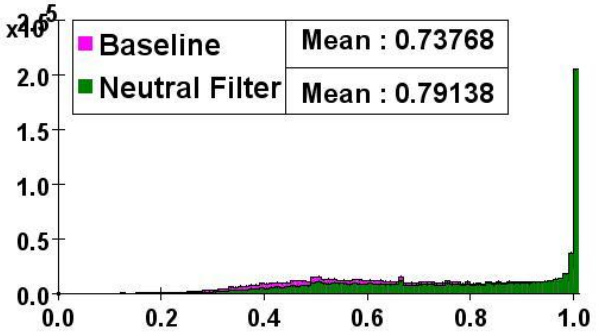


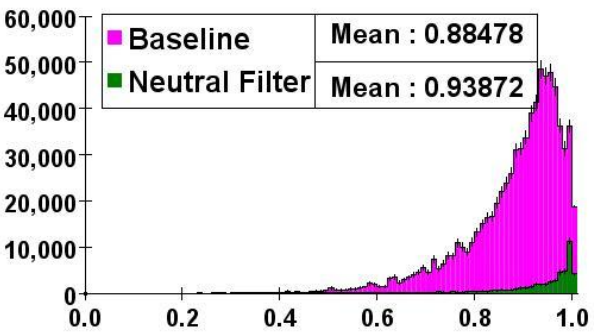


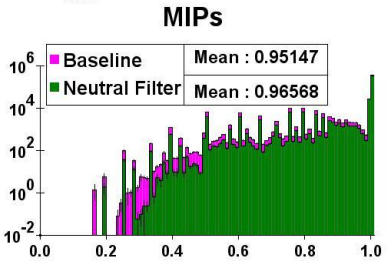


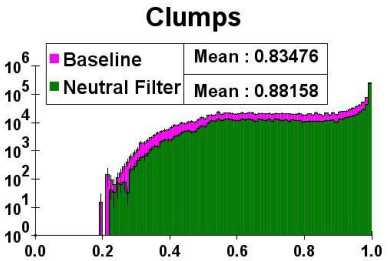


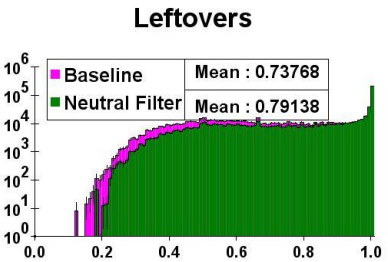
Sub-cluster purity, energy weighted

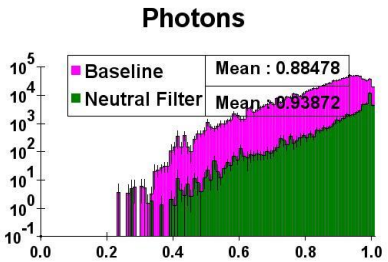












An intermediate summary

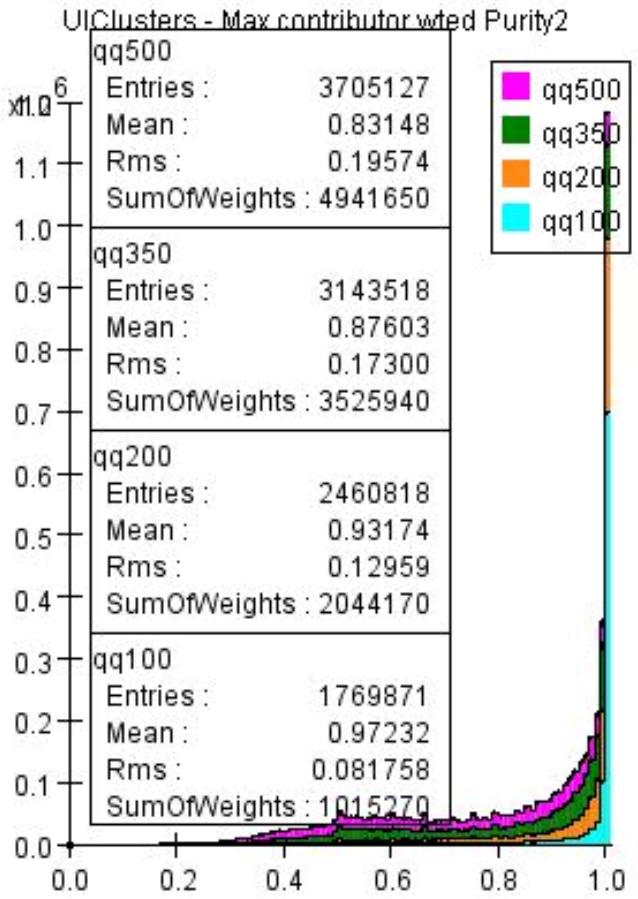
Energy distribution among the sub-clusters (in the data sample):
Clumps : 47% of energy
Leftovers : 22% of energy
Photons : 18% of energy
MIPs : 12% of energy

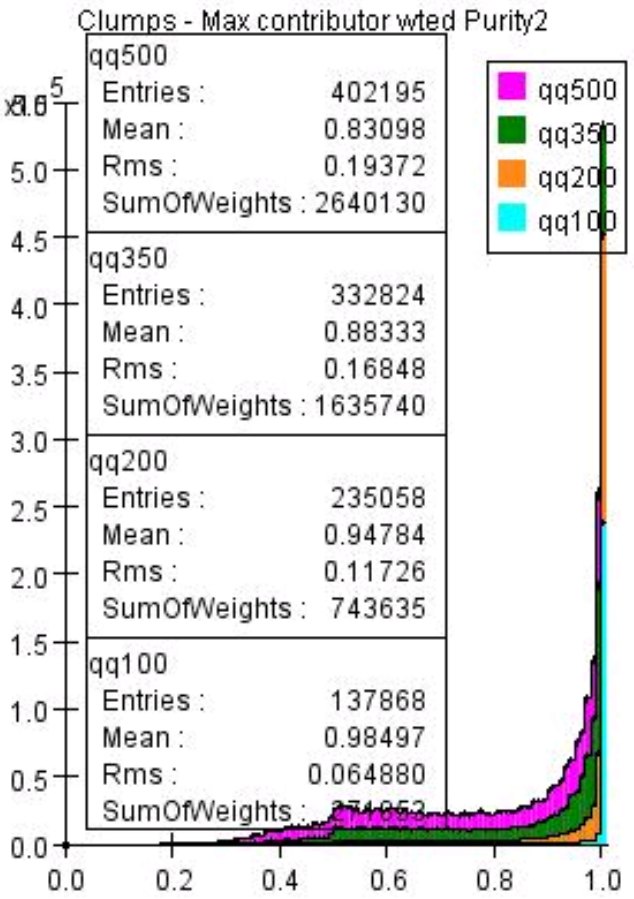
Purity according to sub-clusters, good enough ?

Subcluster	baseline		Neutral hadron filtered	
		Energy-weighted		Energy-weighted
Clumps	90%	84%	93%	88%
MIPs	95%	95%	97%	97%
Blocks	86%	82%	90%	86%
Photons	93%	89%	95%	94%

Neutral hadrons are adding some confusion but not that much

Energy-dependence of purity



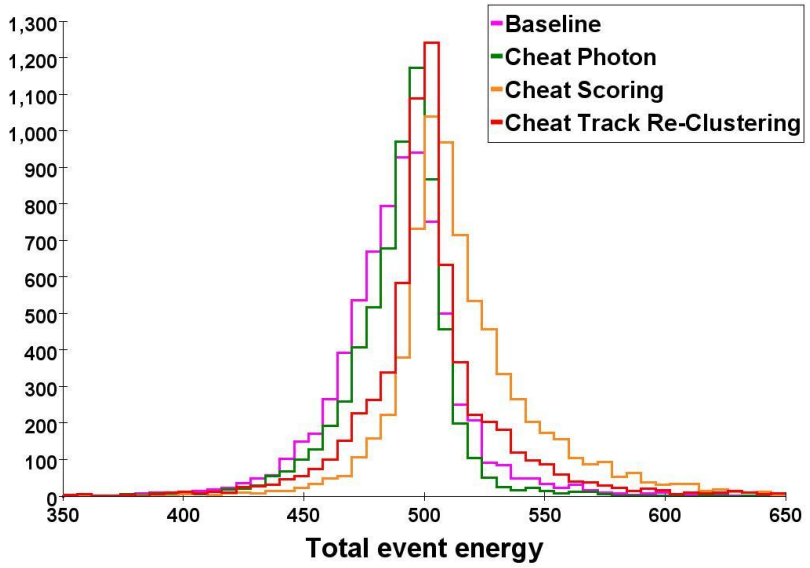


Cassell

As overlap of showers increases with energy, isolation gets to be challenging, purity suffers

Testing each piece step-by-step

Baseline
Cheat-scoring: simply check with MC if it is really a match (score 0 or 1)
Cheat photon finding: perfect photon finder, no photon veto
Cheat reclustering: associate sub-clusters if it belongs according to MC (ignore score) look at the sub-cluster and add to the track where it has dominant contrib
Cheat reclustering + perfect sharing : left-overs added to showers according to MC

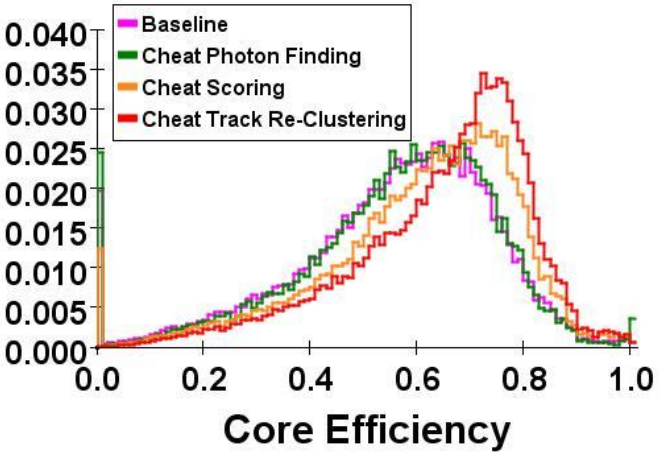


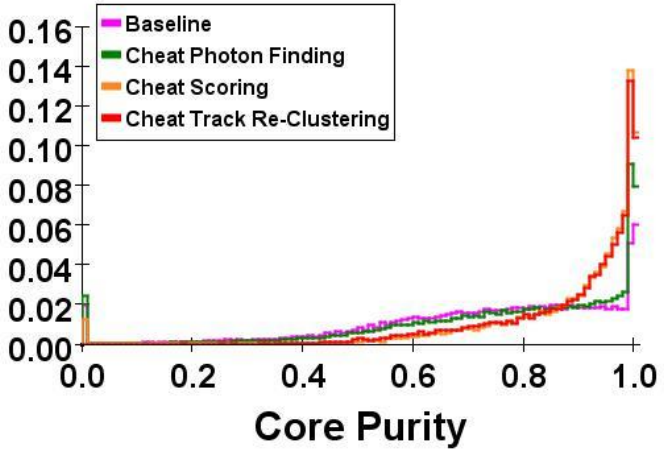
PFA	RMS90
Baseline	3.4%
Cheat Photon Finding	2.8%
Cheat Scoring	3.9%
Cheat Track Re-Clustering	3.6%

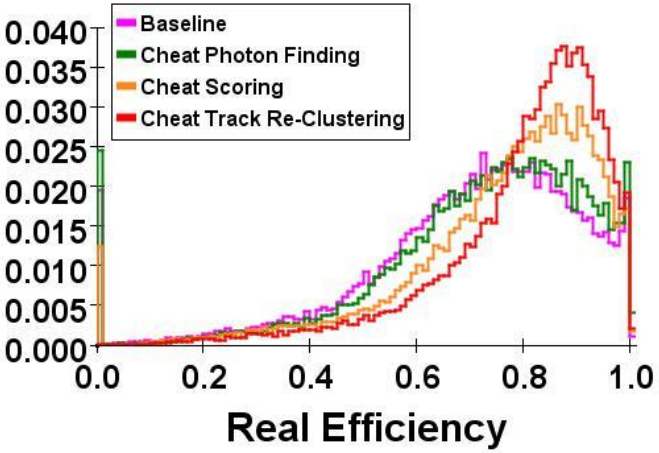
Shower reconstruction efficiency and purity

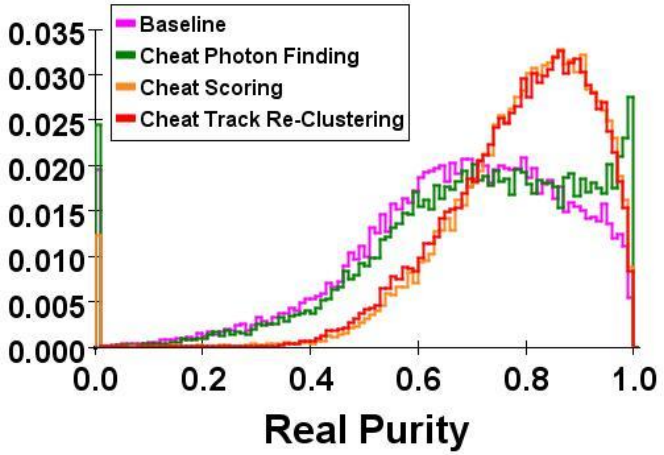
A = baseline ; B = cheat photon
C = cheat scoring;
D = cheat reclustering (perfect PFA)

PFA	A	B	C	D
Core (excluding shared)				
Eff	55	56	60	65
Pur	74	77	89	89
Real (including shared)				
Eff	70	72	76	80
Pur	68	70	79	79









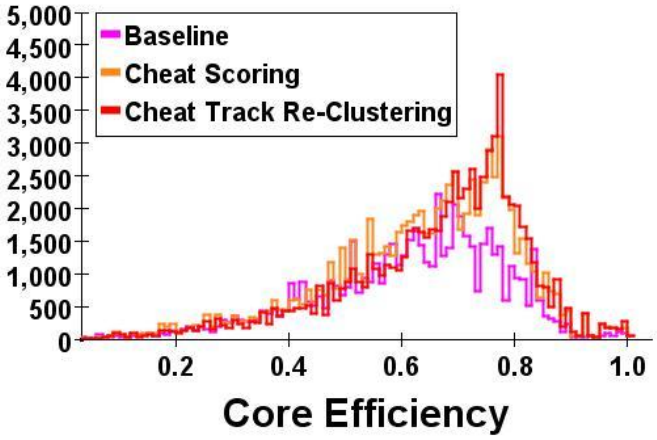
Perfect PFA + perfect shared hits do not enlighten further,
except 22% energy in shared hits is not negligible

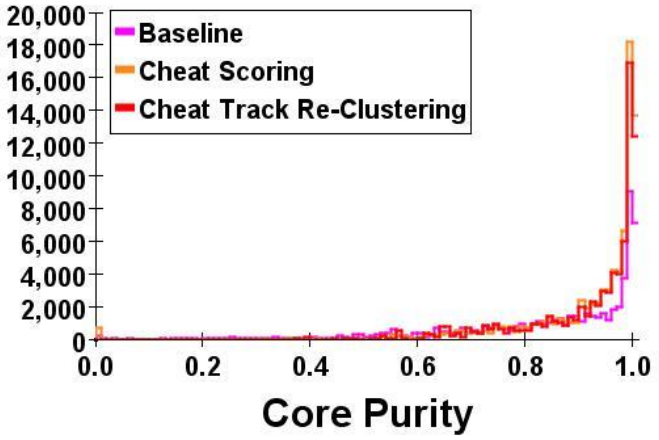
Shower reconstruction efficiency and purity

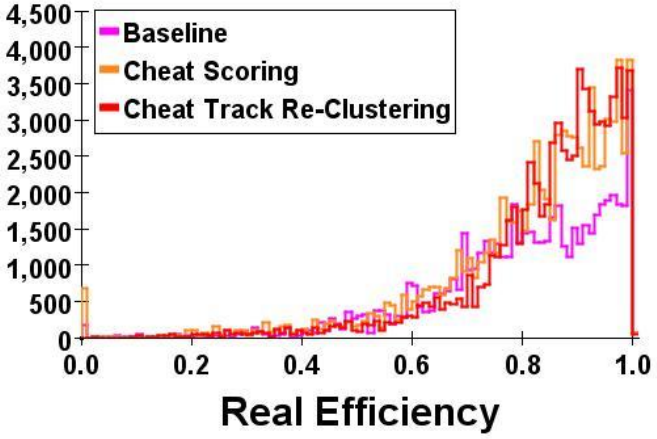
Neutral filter applied

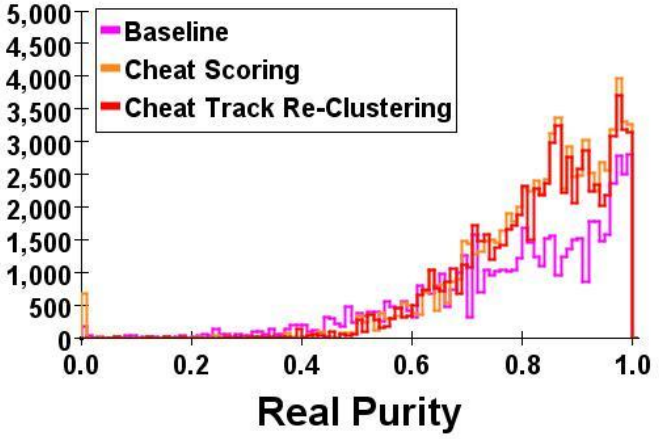
A = baseline ;
C = cheat scoring;
D = cheat reclustering (perfect PFA)

PFA	A	C	D
Core (excluding shared)			
Eff	61	63	66
Pur	85	92	92
Real (including shared)			
Eff	80	81	84
Pur	79	83	83









Conclusions: Optimize sub-clustering better (works OK at low energy)
Photon finder needs to be improved
(According to MC 26% of energy should be in the photons)
(Cheat photon finder shows large improvement)
Then asses if it should be scoring or algorithm

Where we are: Think we understand quite a few of the problems

Next steps : Where to start
: Attempt at isolation of sub-clusters ?
: Attempt at link-scoring ?

Remi will explain link scoring, resolution, effects of E/p balance

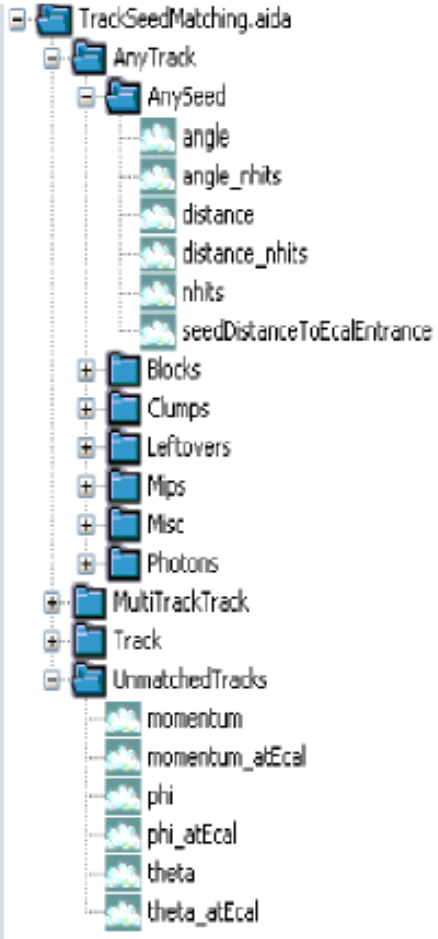
Garabed Halladjian joined the PFA effort on November 1, 2010

Back-ups

Track-Seed Matching: Definitions

THE UNIVERSITY OF IOWA

- Tracks are extrapolated to the innermost layer with hits from the seed:
  - Angle is computed between the seed direction from energy tensor calculation and the tangent to the extrapolated track.
  - Distance is computed between the track interception point and the closest hits in the cluster on the same layer.
- Seed distance to Ecal entrance is the depth of the innermost layer with hits from the seed.
- Plots are made per seed type and also separating simple from multiple tracks.
- Multiple tracks:
  - Angle and distance to seeds are computed using extrapolation results averaged on sub-tracks.
  - Angle is the maximum angle between "sub-tracks" at the extrapolation layer.
  - Distance is the maximum distance between interception points.
- For unmatched tracks:
  - Momentum, theta and phi are plotted for all unmatched tracks and for those that reach the Ecal.



Energy fractions

THE UNIVERSITY OF IOWA

Better track-seed matching for single pi in general

Higher matching efficiency:
qqbar: **82%**
single pi: **99%**

Higher matching to mip rate:
qqbar: **84.5%**
single pi: **91.3%**