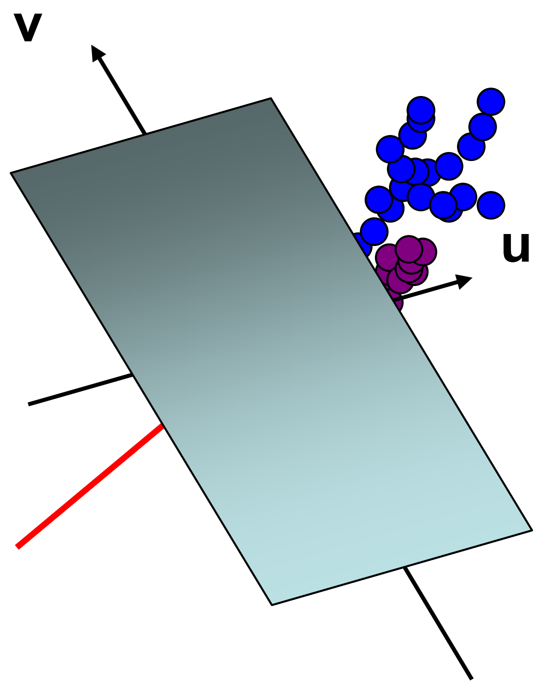
Particle Flow with PandoraPFA

Mark Thomson
University of Cambridge



This Talk:

❶ PandoraPFA: status
❷ PandoraPFA: the algorithm
❸ Performance (old vs. new)
❹ Comments
❺ PerfectPFA
❻ Detector Studies
❼ Deficiencies, random comments, outlook, and conclusions

❶ PandoraPFA Status

- ★ A lot of progress since LCWS06
- ★ Significant recent progress (last few months):
  - ♦ ValGrind – highlighted a few other minor issues
  - ♦ Reclustering significantly improved +
    fixed a couple of "features"
  - ♦ Added new "photonRecovery" code
  - ♦ Added new "fragmentRemoval" code
  - ♦ Version finalised last month and now in CVS repository (v01-01)

➡ **Much improved performance…**

All results here were obtained with CVS version

Please use CVS version

❷ PandoraPFA: the Algorithm

★ ECAL/HCAL reconstruction and PFA performed in a single algorithm
★ Keep things fairly generic algorithm
  ▪ applicable to multiple detector concepts
★ Use tracking information to help ECAL/HCAL clustering

★ This is a fairly sophisticated algorithm : ~9000 lines of code
★ Organic approach – "understanding of event" gradually grows

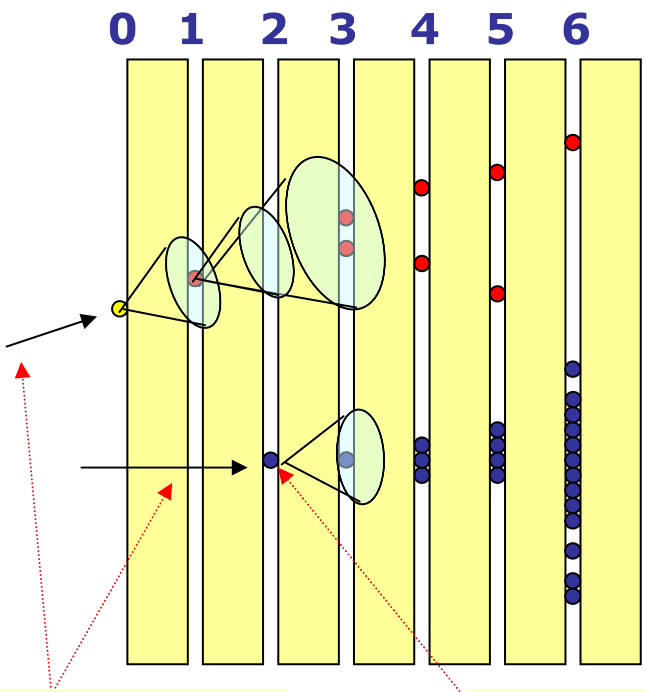
Eight Main Stages (was six):

i. Preparation (MIP hit ID, isolation, tracking)
ii. Loose clustering in ECAL and HCAL
iii. Topological linking of clearly associated clusters
iv. Courser grouping of clusters
v. Iterative reclustering
vi. Photon Recovery (NEW)
vii. Fragment Removal (NEW)
viii. Formation of final Particle Flow Objects (reconstructed particles) – not very sophisticated

Order inter-changable

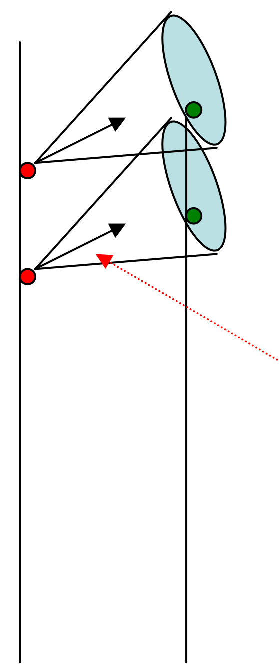
ii) ECAL/HCAL Clustering

- ★ Start at inner layers and work outward
- ★ Tracks can be used to "seed" clusters
- ★ Associate hits with existing Clusters
- ★ If no association made form new Cluster
- ★ Simple cone based algorithm



Initial cluster direction

Unmatched hits seeds new cluster



Simple cone algorithm based on current direction + additional N pixels

Cones based on either:
initial PC direction or
current PC direction

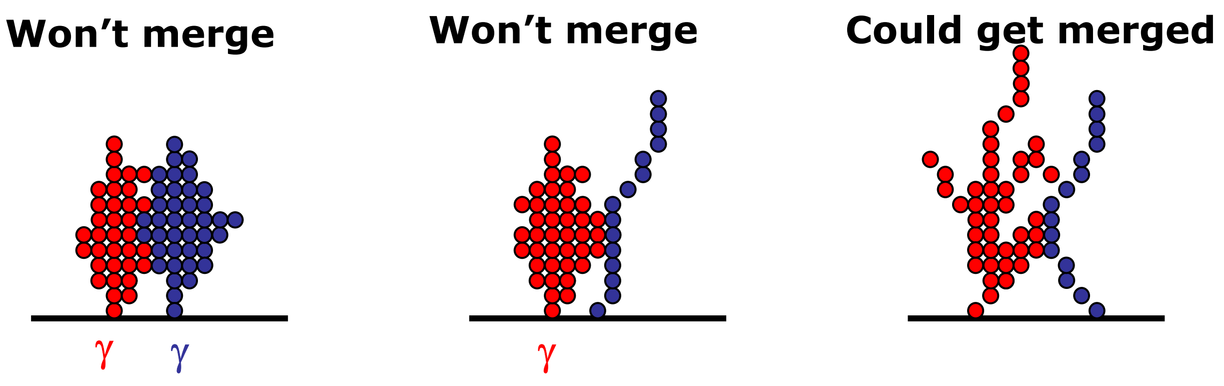
Parameters:
- cone angle
- additional pixels

iii) Topological Cluster Association

- By design, clustering errs on side of caution
  i.e. clusters tend to be split
- **Philosophy:** easier to put things together than split them up
- Clusters are then associated together in two stages:
  - 1) Tight cluster association – clear topologies
  - 2) Loose cluster association – fix what's been missed

Photon ID

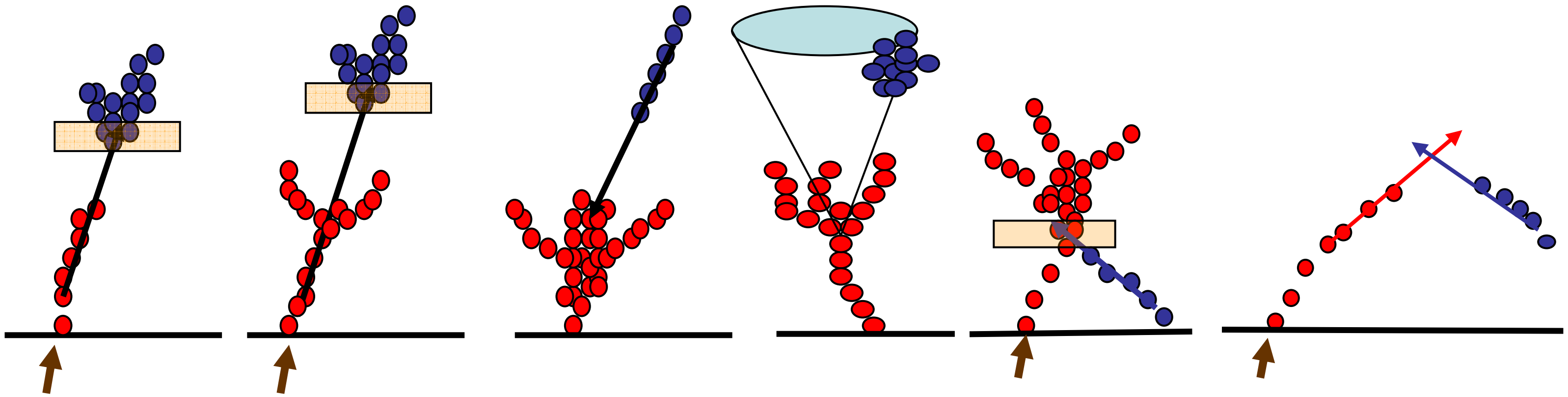
- Photon ID plays important role
- Simple "cut-based" photon ID applied to all clusters
- Clusters tagged as photons are immune from association procedure – just left alone



★ Clusters associated using a number of topological rules

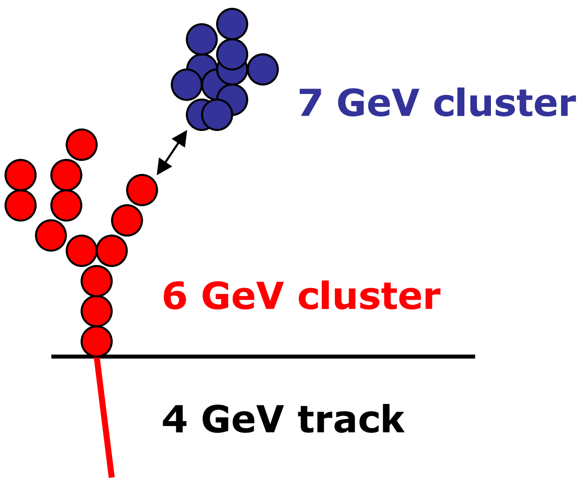
Clear Associations:

- Join clusters which are clearly associated making use of high granularity + tracking capability: very few mistakes



Less clear associations:

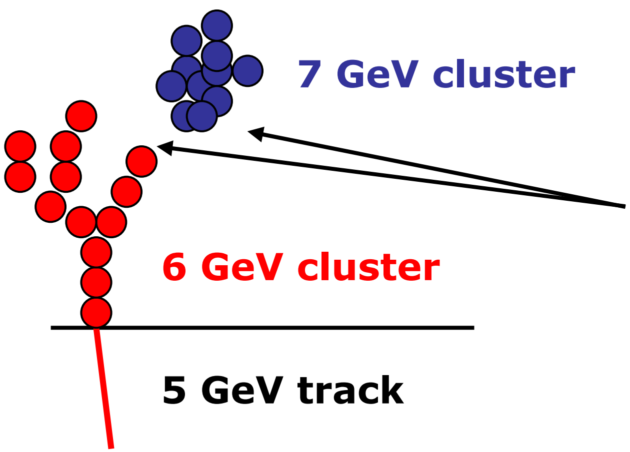
e.g. Proximity



Use E/p consistency to veto clear mistakes

iv) Cluster Association Part II

- Have made very clear cluster associations
- Now try “cruder” association strategies
- BUT first associate tracks to clusters (temporary association)
- Use track/cluster energies to “veto” associations, e.g.

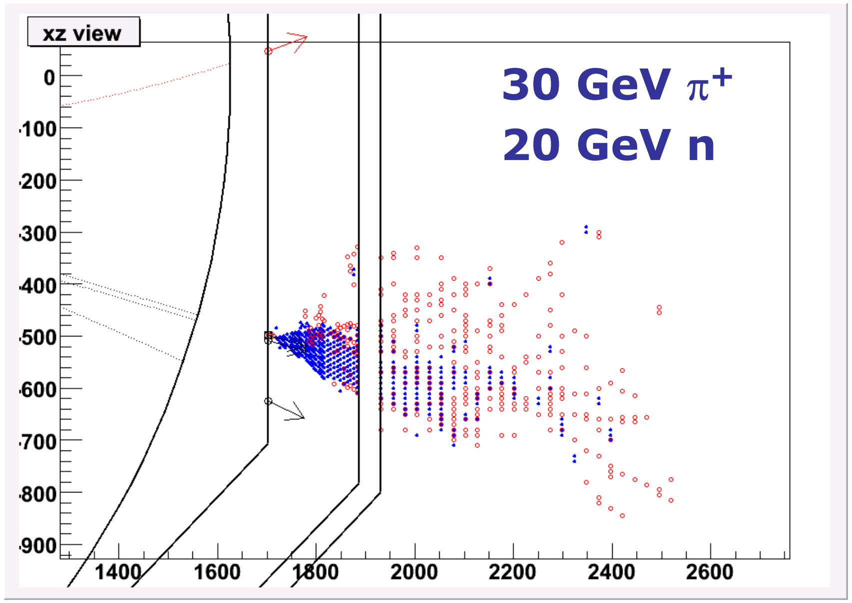


This cluster association would be forbidden if $|E_1 + E_2 - p| > 3\,\sigma_E$

Provides some protection against silly mistakes

v) Iterative Reclustering

★ Upto this point, in most cases performance is good – but some difficult cases…



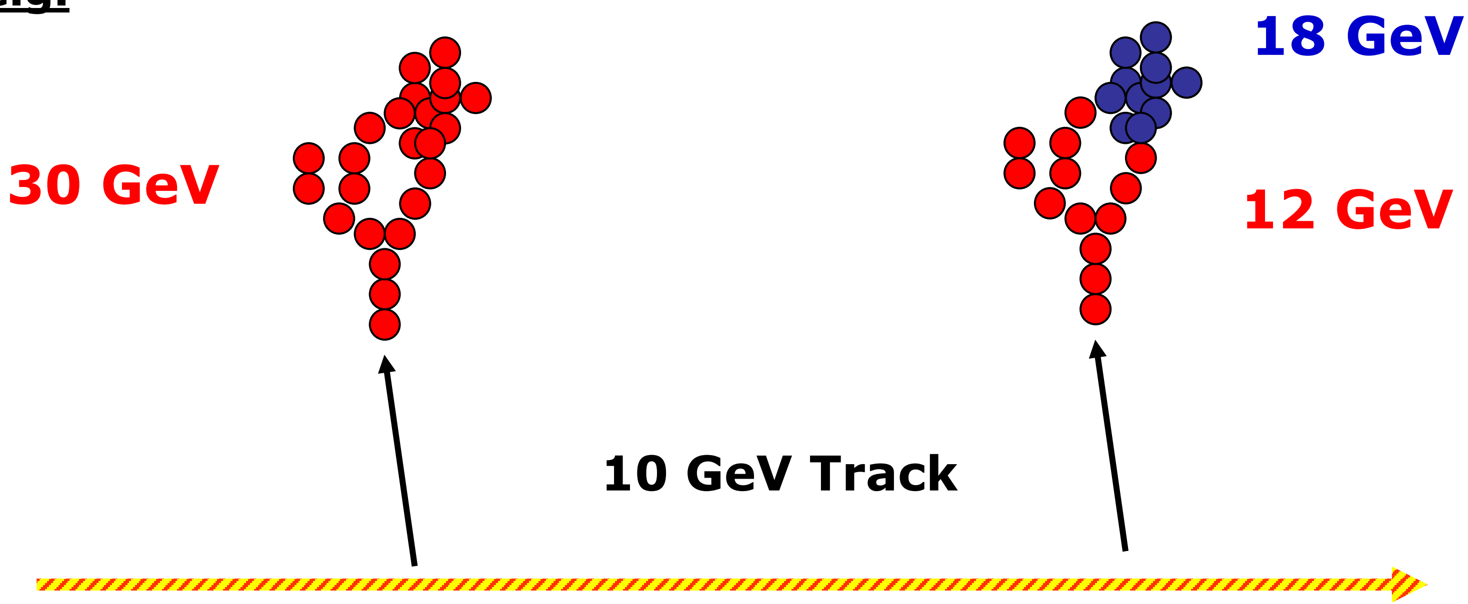
★ At some point hit the limit of "pure" particle flow
- ♦ just can't resolve neutral hadron in hadronic shower

The ONLY(?) way to address this is "statistically"

e.g. if have 30 GeV track pointing to 20 GeV cluster
SOMETHING IS WRONG

★ **If track momentum and cluster energy inconsistent : RECLUSTER**

e.g.



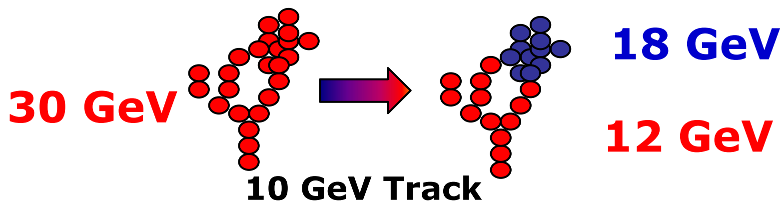
Change clustering parameters until cluster splits and get sensible track-cluster match

This is very important for higher energy jets

Iterative Reclustering Strategies

① Cluster splitting

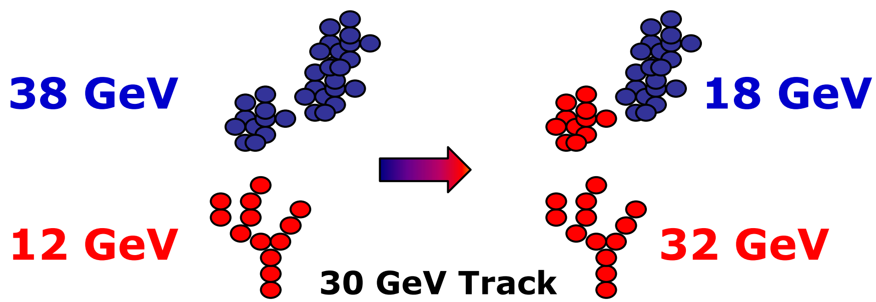
Reapply entire clustering algorithm to hits in "dubious" cluster. Iteratively reduce cone angle until cluster splits to give acceptable energy match to track



★ Could/Should plug in alt. clustering (to some extent this is now done)

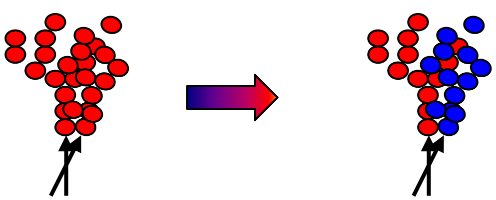
② Cluster merging with splitting

Look for clusters to add to a track to get sensible energy association. If necessary iteratively split up clusters to get good match.



③ Track association ambiguities

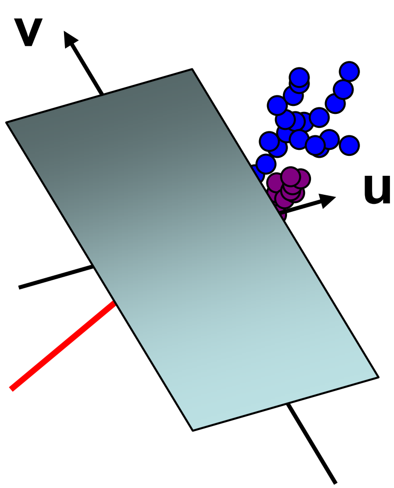
In dense environment may have multiple tracks matched to same cluster. Apply above techniques to get ok energy match.

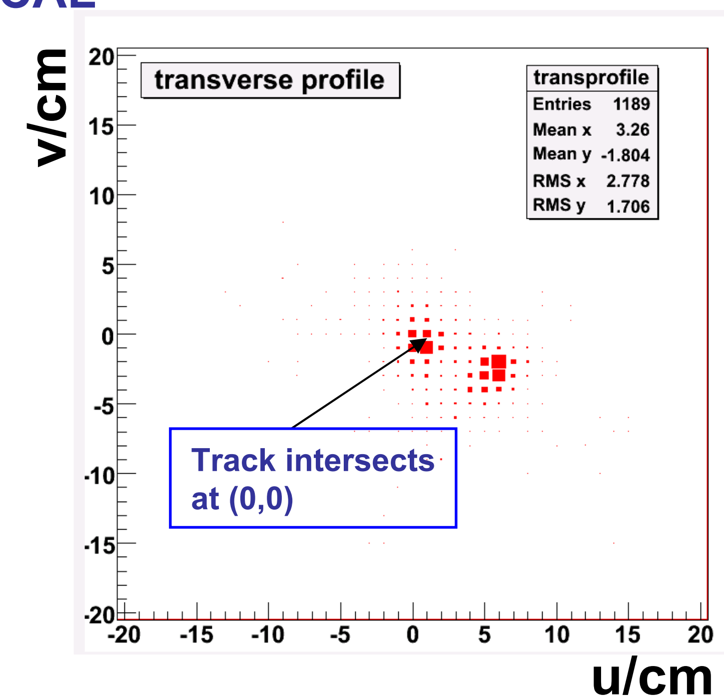


★ Significant improvements in Reclustering since Valencia ECFA meeting (including feature fixes)

vi) Photon Recovery (new in v01-01)

- Using cone clustering photons close to early showering charged hadrons can be merged into a single cluster.
- Introduce a new (very simple) algorithm to recover these
- For each cluster associated with a track:
  - project ECAL hits onto plane perpendicular to radial vector to point where track intersects ECAL
  - search for peaks…



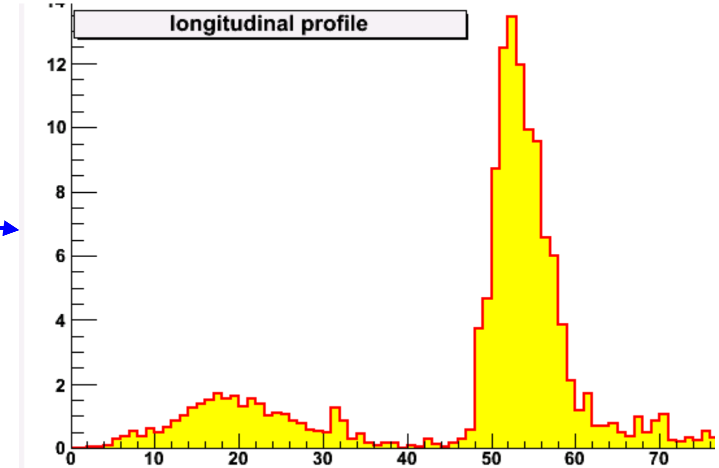


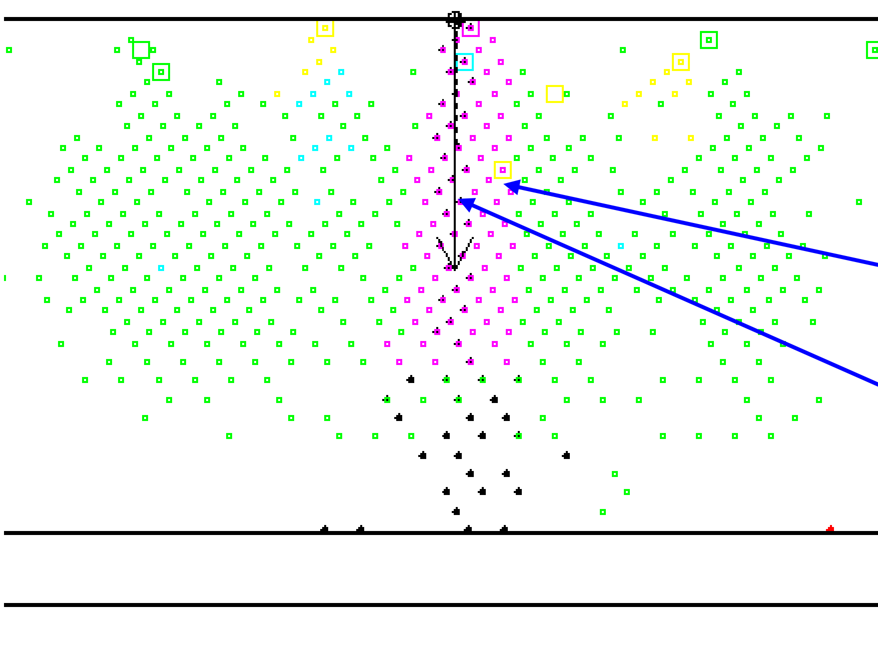
- If there is an isolated peak not associated with “track peak” make new photon cluster if track energy and remaining cluster energy still statistically compatible with track momentum

- Also look for photons where only a single peak is found 

- Implemented by looking at longitudinal profile of “shower”

- Currently very simple cuts – no attempt to identify EM shower profile **(to come)**





★ Enables resolution of overlapping MIPs and photons, e.g.

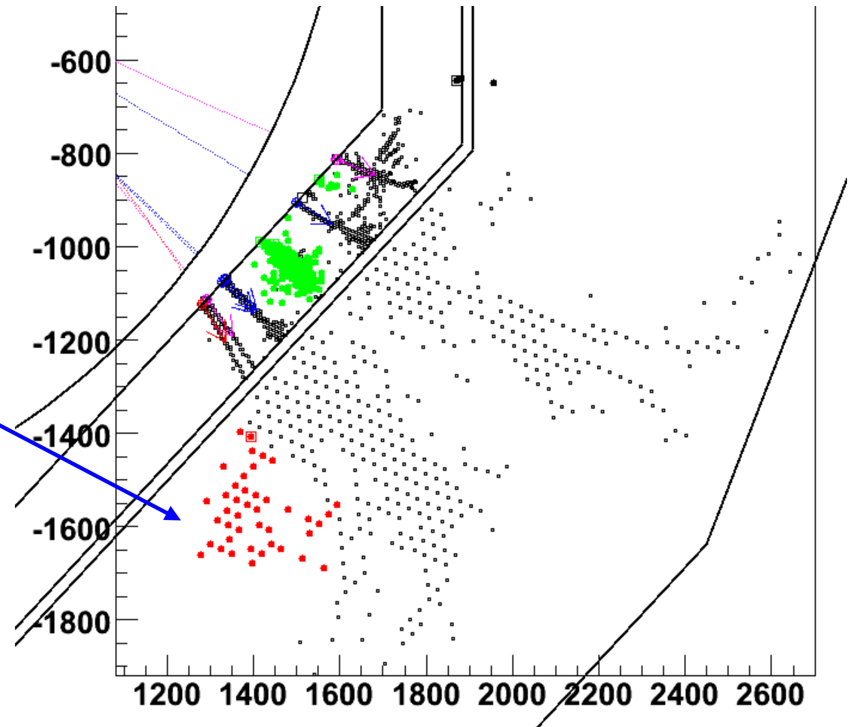
Recovered photon

Shared track/photon hits

vii) Fragment Removal (new in v01-01)

★ Final stage in PFA is to examine "neutral hadron" candidates
- true neutral PFO ?
- fragment from another cluster ?

e.g. is this a neutron/K_L ?

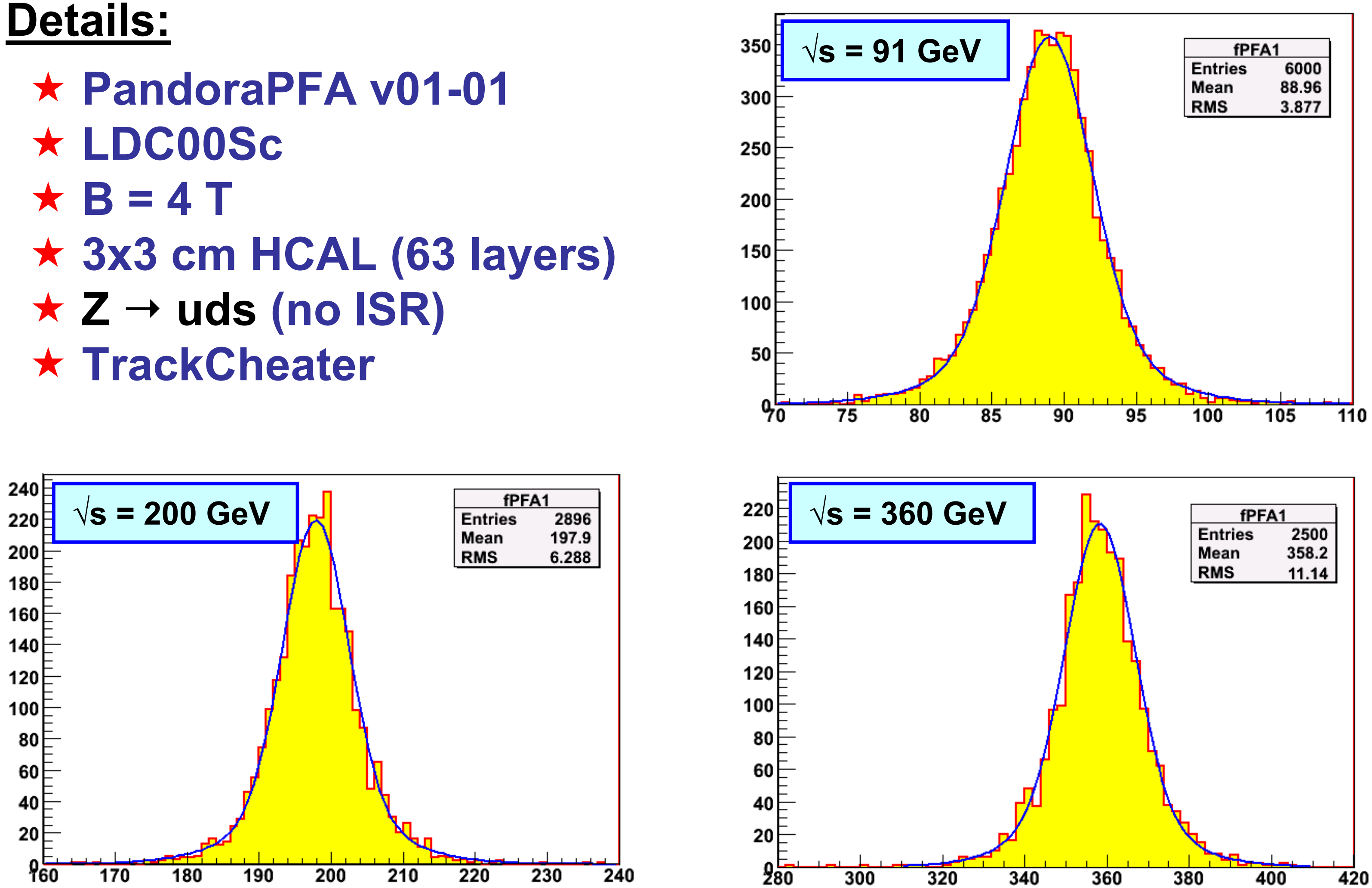


Procedure:

- Look at each non-photon neutral cluster
- Construct some "useful" variables:
  - contact layers
  - cone fraction
  - cluster direction
- Apply very simple (non-optimised) cuts
- Don't do anything too dumb – i.e. if associating fragment to another results in significant track momentum – cluster energy mismatch.

❸ Current performance

rms90



Details:

- ★ PandoraPFA v01-01
- ★ LDC00Sc
- ★ B = 4 T
- ★ 3x3 cm HCAL (63 layers)
- ★ Z → uds (no ISR)
- ★ TrackCheater

Current performance

rms90

E_{JET}	$\sigma_E/E = \alpha/\sqrt{(E/GeV)}$ $\lvert\cos\theta\rvert<0.7$	σ_E/E
45 GeV	0.295	4.4 %
100 GeV	0.305	3.0 %
180 GeV	0.418	3.1 %
250 GeV	0.534	3.3 %

For jet energies < 100 GeV ILC "goal" reached !!!

For jet energies ~ 200 GeV close to 40 %/$\sqrt{E(GeV)}$!!

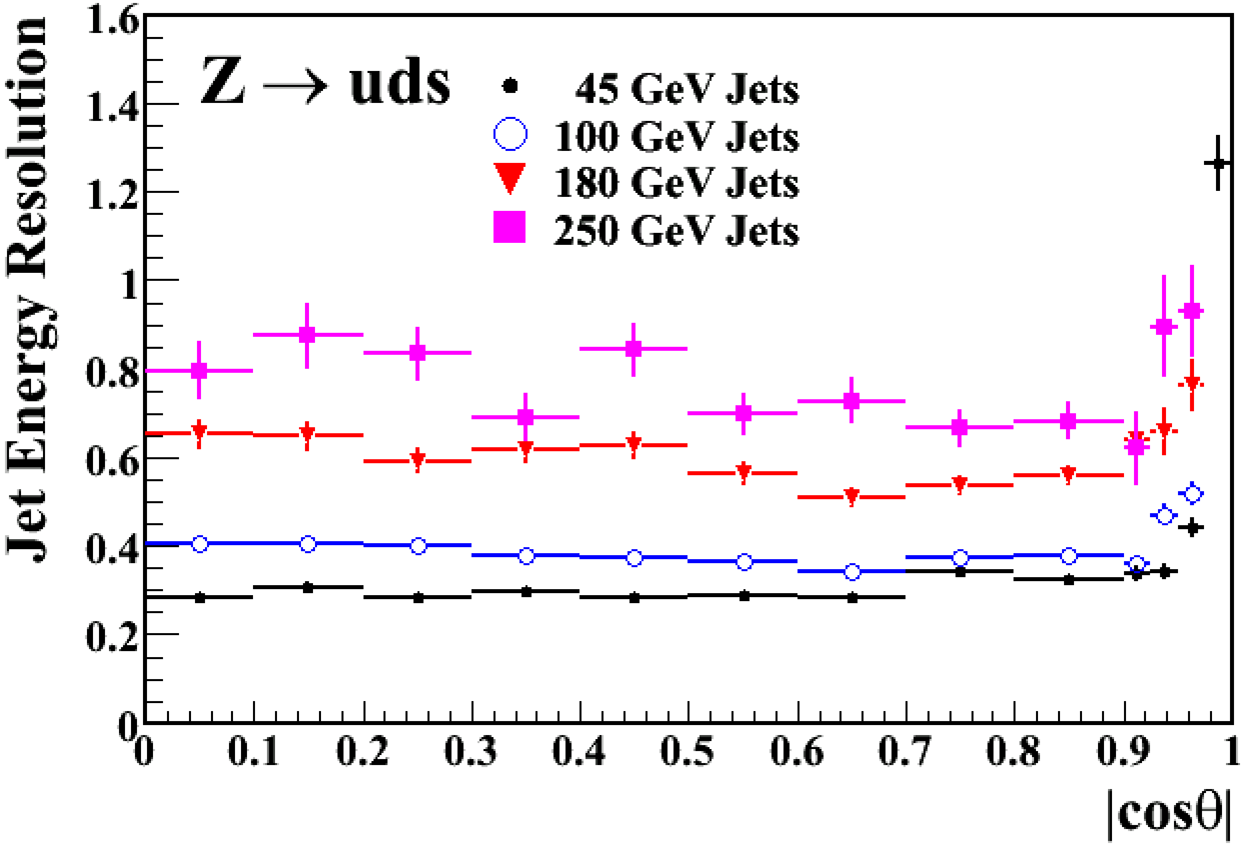
0.35 at LCWS06

Opinion:

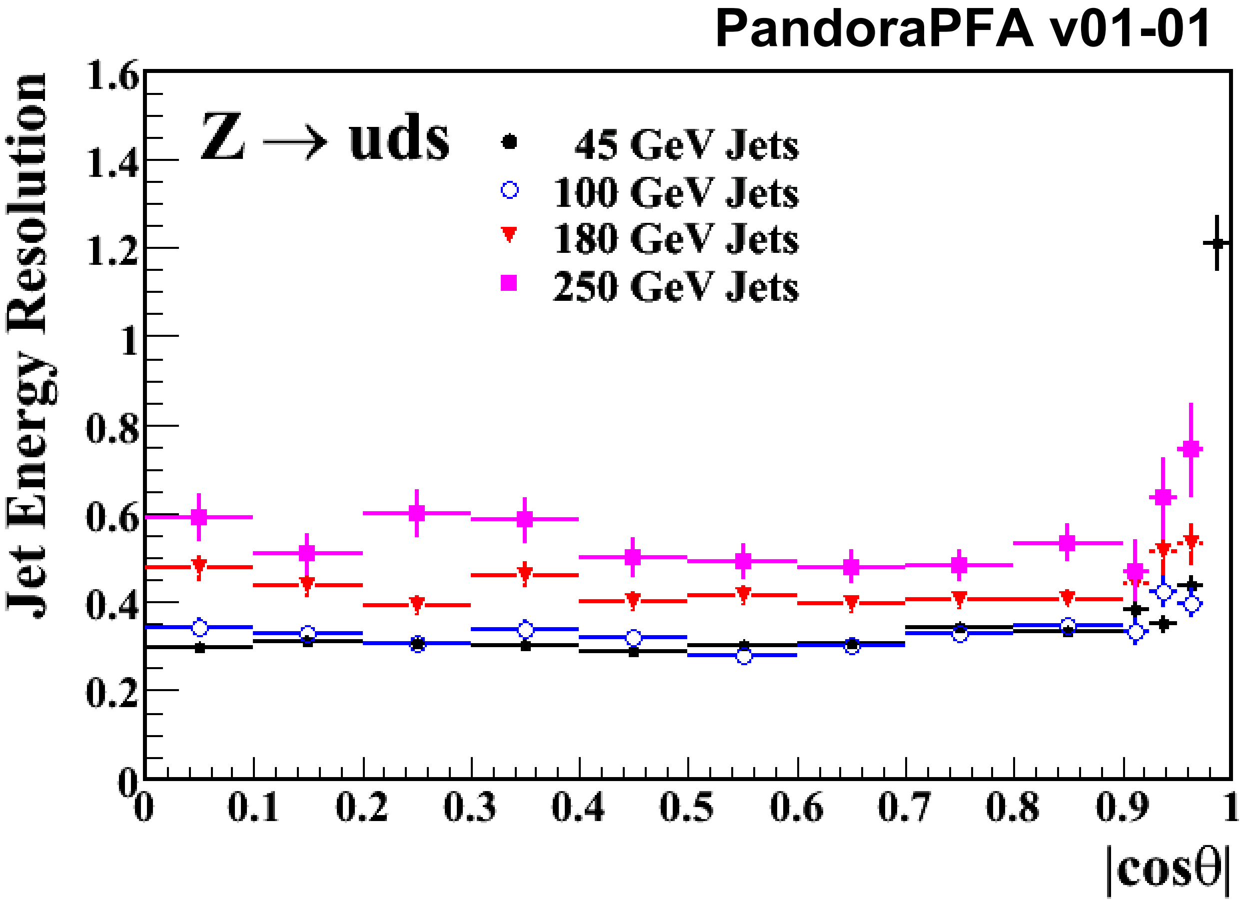
★Now convinced that PFA can deliver the required ILC jet energy performance

★The current code is not perfect (see later), things will get better

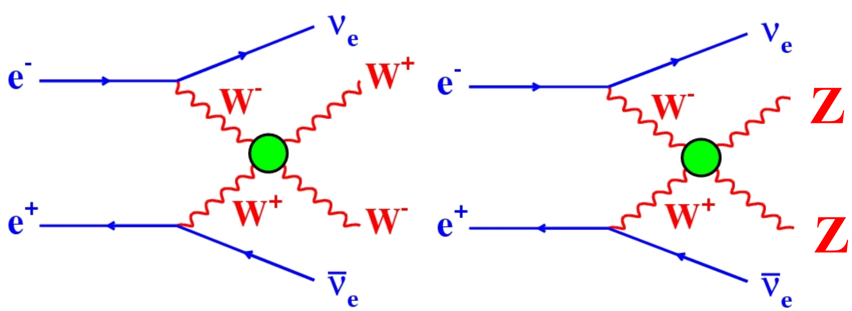
Angular Dependence (OLD)

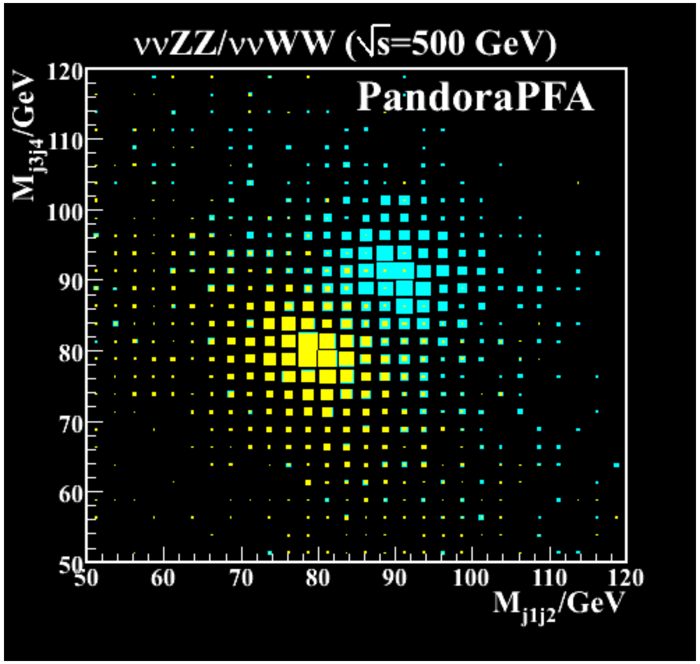


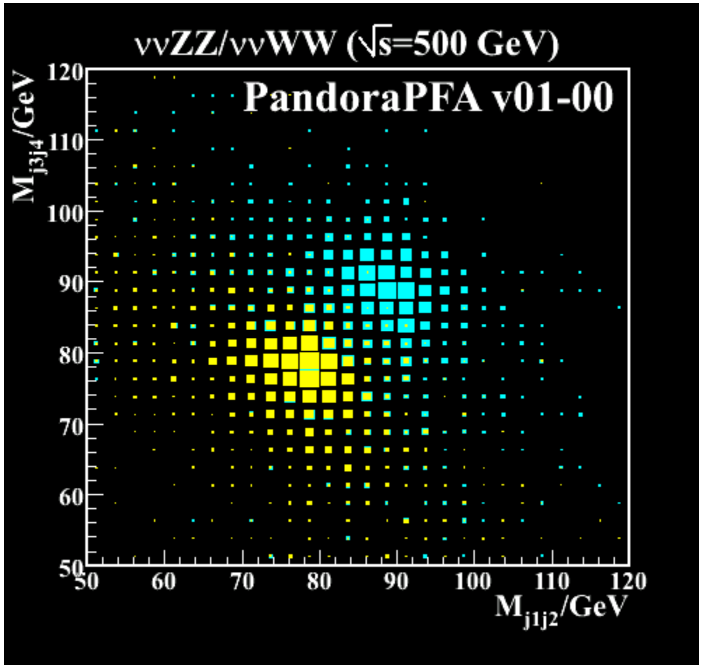
Angular Dependence (NEW)



for completeness…







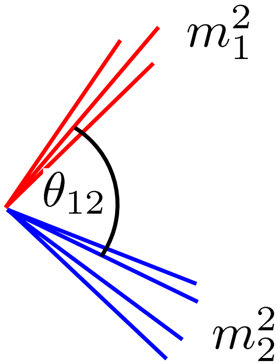
Visible improvement in WW/ZZ separation **(will return to this later)**

❹ Some Comments

What is the ILC goal for PFA ?

★ For a pair of jets have:

$$m^2 = m_1^2 + m_2^2 + 2E_1E_2(1 - \beta_1\beta_2\cos\theta_{12})$$



★ Assuming a single jet energy resolution of

$$\sigma_E/E = \alpha(E)/\sqrt{E(\mathrm{GeV})}$$

⇨ $$\sigma_m/m \approx \alpha(E_j)/\sqrt{E_{jj}(\mathrm{GeV})}$$

+ term due to θ_{12} uncertainty

★ For a Gauge boson mass resolution of order $\Gamma_{\mathrm{W/Z}}$

$$\frac{\sigma_m}{m} \approx \frac{2.5}{91.2} \approx \frac{2.1}{80.3} \approx 0.027$$

⇨ $$\alpha(E_j) < 0.027\sqrt{E_{jj}(\mathrm{GeV})}$$

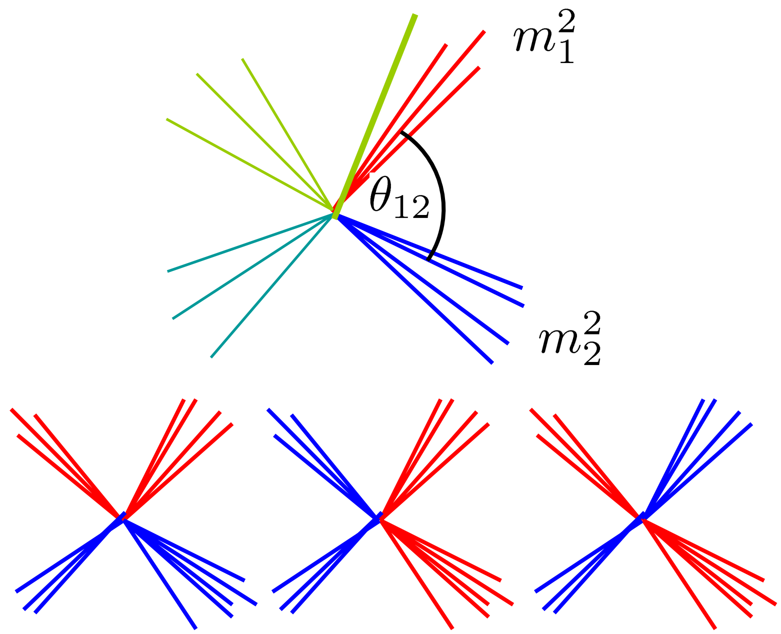
E_{jj}/GeV	$\alpha(E_j)$	σ_{Ej}/E_j
91	< 26 %	3.8 %
200	< 38 %	3.8 %
360	< 51 %	3.8 %
500	< 60 %	3.8 %

★ Don't take exact numbers too seriously – assumes equal sharing of energy

★ On this metric, low energy performance needs further work

Other effects:

- ★ jet angle measurement
- ★ jet finding
  - not perfect – will degrade mass resolution
- ★ jet pairing
  - multiple jet-pair combinations



Relative importance of these effects ?

❺ PandoraPerfectPFA

- ★ Recently added PerfectPFA option in Pandora (not yet in CVS)
  - <parameter name="PerfectPFA" type="int"> 1 </parameter>
- ★ Uses MC information to create the ProtoClusters
- ★ The rest of the algorithm is the same
- ★ Although very fresh, can already learn something…
- ★ Process same events/same analysis and compare PFA to perfect PFA
  - Note in these studies the tracks are the same "TrackCheater"

i) How close to being "Perfect" is PandoraPFA?

E_{JET}	$\sigma_E/E = \alpha/\sqrt{(E/GeV)}$ $|\cos\theta|<0.7$	
	PerfectPandora	PandoraPFA
100 GeV	0.220	0.305
180 GeV	0.305	0.418

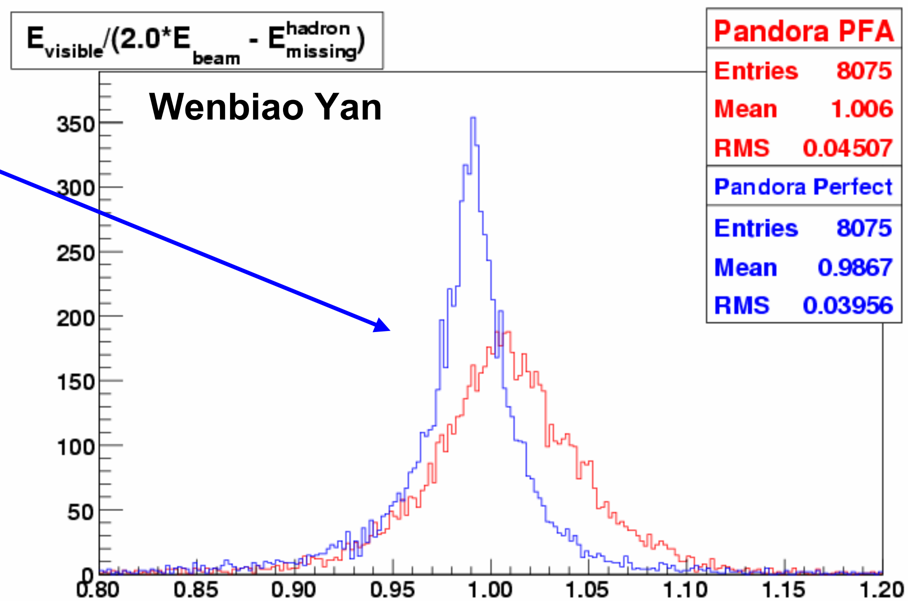
Still someway to go even for low energy jets – needs study

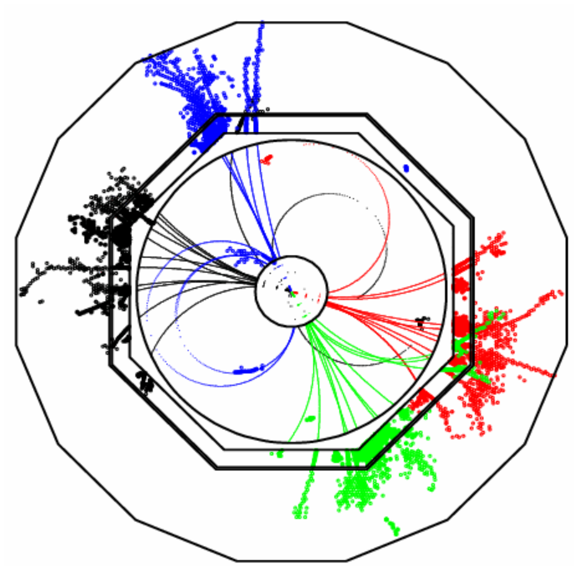
ii) PFA impact in a real physics process

e.g. $e^+e^- \rightarrow \nu\bar{\nu}W^+W^- \rightarrow \nu\bar{\nu}qqqq$ $\sqrt{s} = 800\,\mathrm{GeV}$

★ First compare visible energy from PFA with expected
(i.e. after removing neutrinos/forward tracks+clusters)

- PerfectPFA gives better energy resolution than PandoraPFA
(as expected)

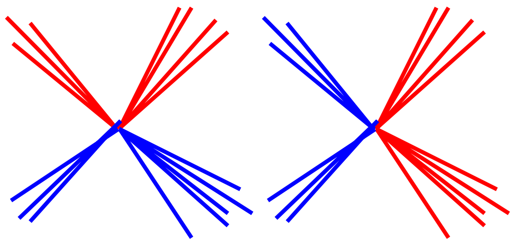




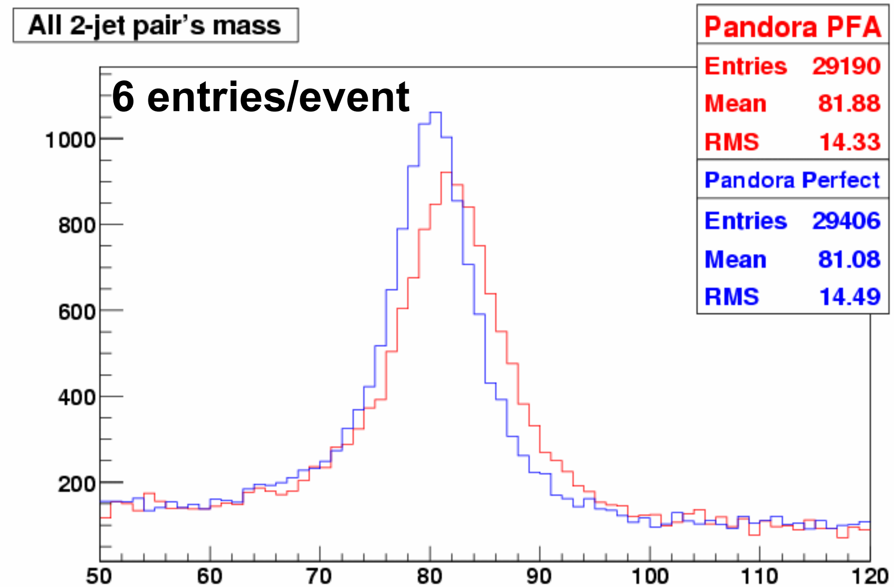
★ Does this difference make it through to a physics analysis (i.e. after jet finding/ jet pairing) ?

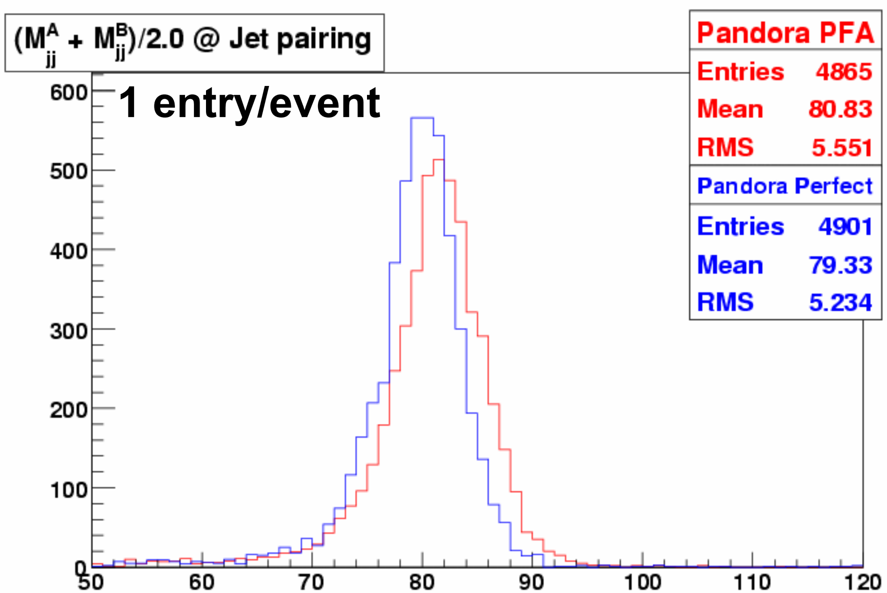
★Force event into 4 jets (Durham)

★Plot masses of the 2 Ws formed from the 3 possible jet-pairings



HERE: PandoraPFA ~ PerfectPFA





★Choose pairing with smallest mass difference

★Plot average mass of the 2 Ws

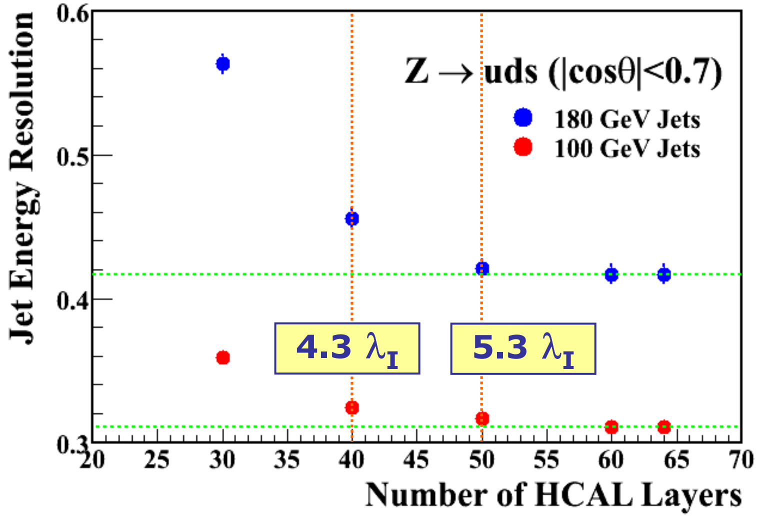
HERE: PandoraPFA ~ PerfectPFA

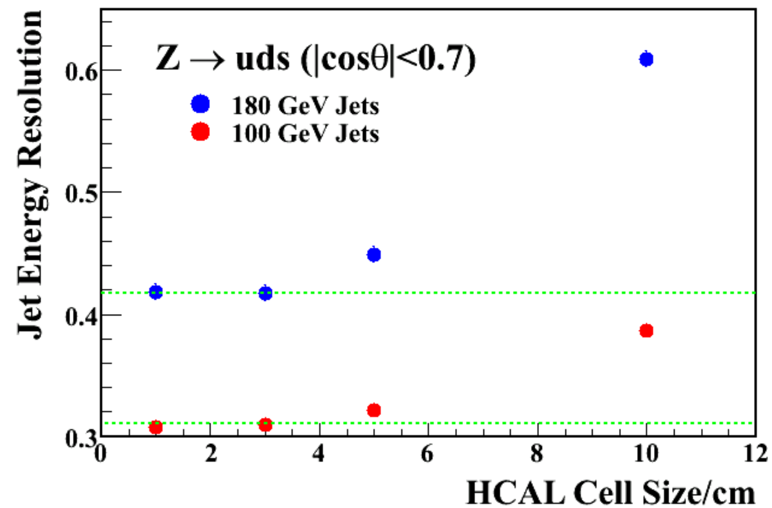
⇨ PandoraPFA performance not limiting analysis

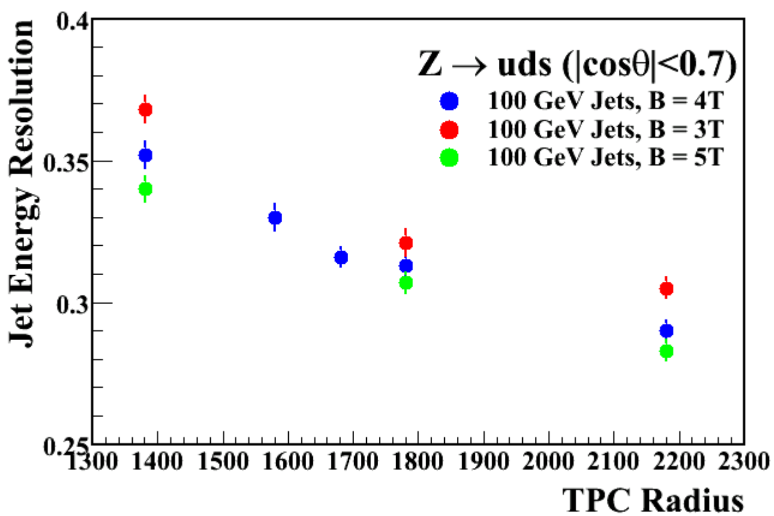
⇨ "Physics Ready PFA"

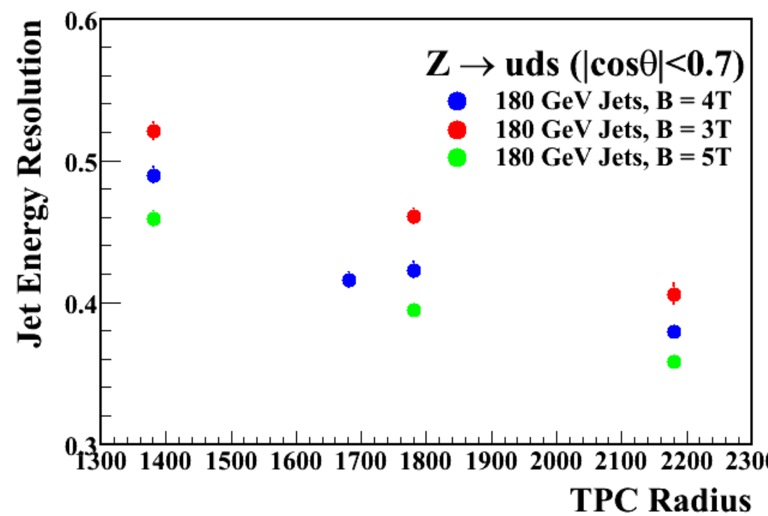
❻ Detector Optimisation Studies

★ Lots of progress… …no time









Parameterized Performance

- ★LDC Jet energy performance found to depend mainly on:
  - HCAL thickness
  - TPC Radius
  - B-field

EMPIRICAL

- ★Plots shown on previous page can be parameterized by:

100 GeV Jets

$$\alpha = 0.315 \left(\frac{B}{4}\right)^{-0.19} \left(\frac{R}{1.68}\right)^{-0.49} \left(1 + 6.3e^{-\frac{N}{8.0}}\right)$$

180 GeV Jets

$$\alpha = 0.42 \left(\frac{B}{4}\right)^{-0.31} \left(\frac{R}{1.78}\right)^{-0.61} \left(1 + 21.6e^{-\frac{N}{7.1}}\right)$$

- ★NOTE:
  - Different parameterization for different energy (increased dependence as confusion/leakage become more important)
  - Very different from naïve $B^{-1}R^{-2}$ dependence (much weaker)

Can start to think about starting to consider cost optimisation studies

- ★First trial run shown in LDC meeting

❼ Deficiencies, Random Comments, Outlook and Conclusions

Deficiencies:

★ PandoraPFA has evolved solely with the aim of improving performance … never overly concerned with niceties…

★ Moving to **LDCTracking** is the HIGHEST Priority

★ Very little has been optimised:

- Photon ID – very crude
- Photon Recovery – very crude
- Fragment Removal – very very crude

} Plenty of room for improvement

Random Comments:

★ PFA = much more than clustering

★ In developing code – learnt importance of:

- extreme care - do make any unnecessary mistakes
- use of track momentum – cluster energy to spot to PFA errors

★ Calibration – not trivial. Must check with single particles

★ HCAL energy resolution vs. reclustering ? Improved resolution would help resolve PFA errors…

Speed:

- ★ Some comments at this meeting about speed
- ★ Please be careful here – PandoraPFA is not "slow" but on some events it does use a lot of memory
- ★ "Rough benchmark numbers" (3GHz Pentium III, 1 GByte memory)
  - 200 GeV uds : 5 s/event
  - 500 GeV uds : 16 s/event
  - 500 GeV tt : 15 s/event

  (small samples run this a.m.)
- ★ Not fast, but not that slow. Could do with profiling…

Outlook:

- ★ User feedback very important
  - it is possible that there are still hidden bugs
  - help identify deficiencies

Conclusions:

- ★ PandoraPFA is not perfect
- ★ Things can only get better
- ★ Nevertheless, I believe it has been demonstrated that PFA can give ILC performance goals for physics at $\sqrt{s}$ = 500 GeV and $\sqrt{s}$ = 1 TeV

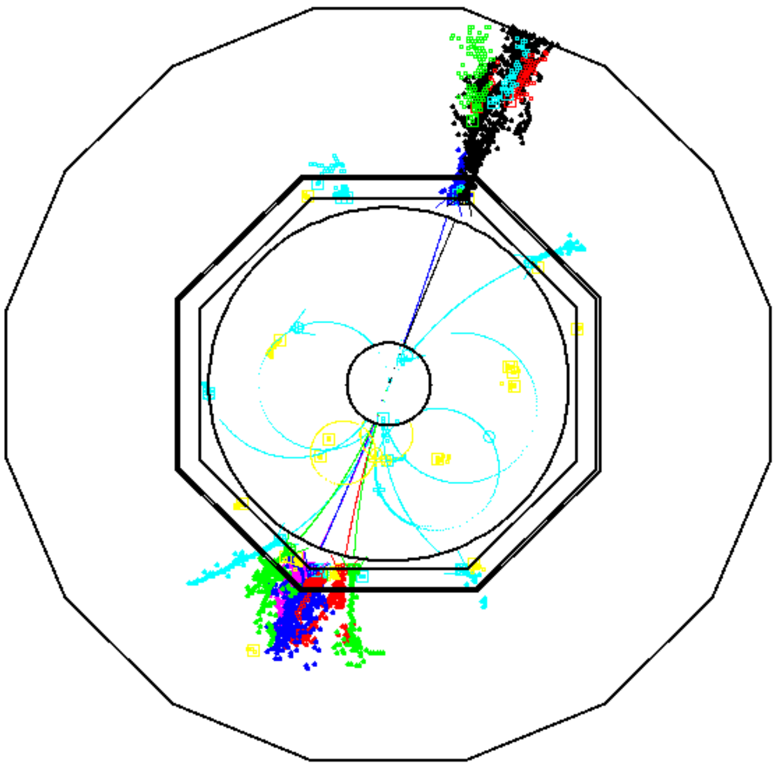
To further develop PandoraPFA need help…
How to get others involved?

End of Talk

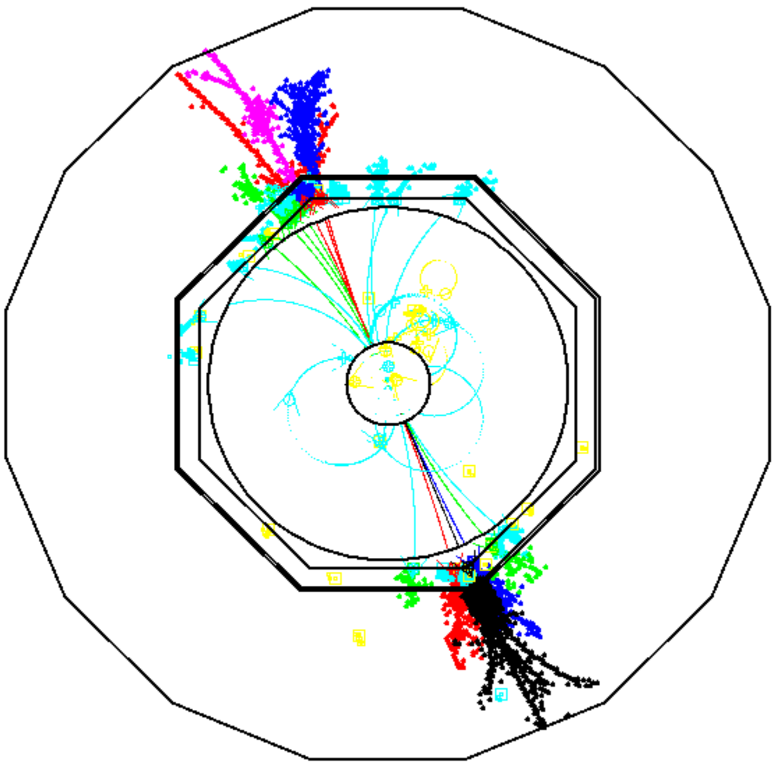
Backup : When PandoraPFA Goes Bad

- A few di-jet events at $\sqrt{s}$ = 360 GeV

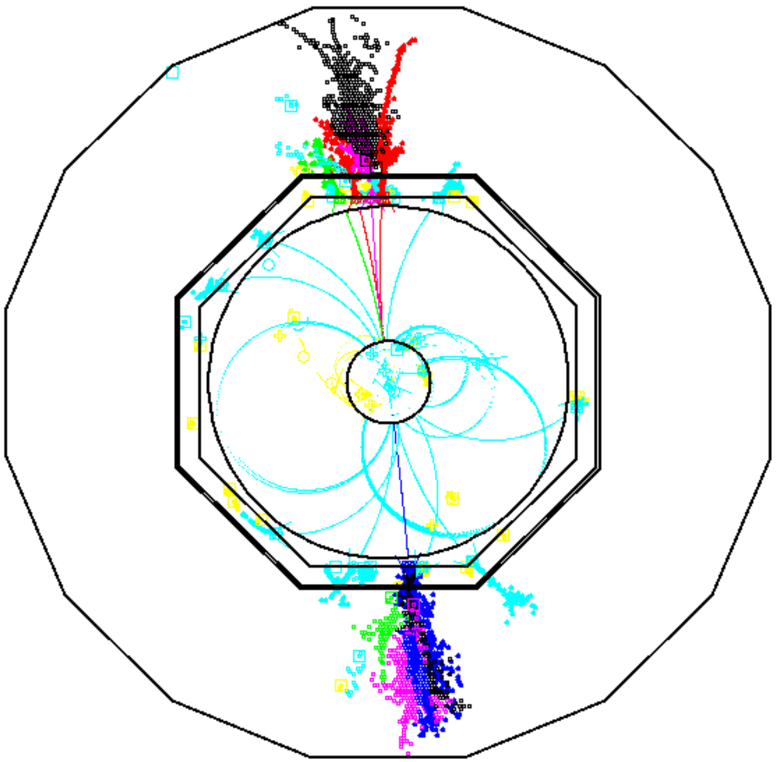
E_{reco} = 337 GeV



E_{reco} = 338 GeV



E_{reco} = 382 GeV



E_{reco} = 391 GeV