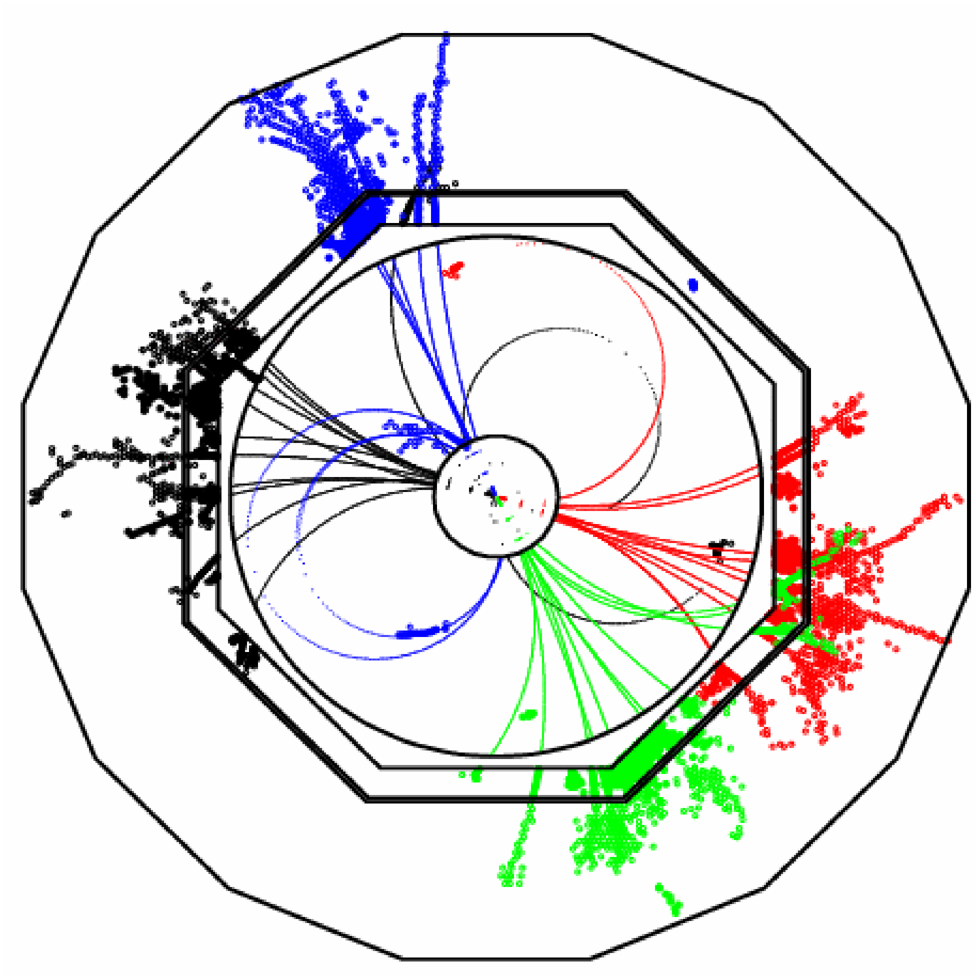
PandoraPFA

Mark Thomson
University of Cambridge



This Talk:

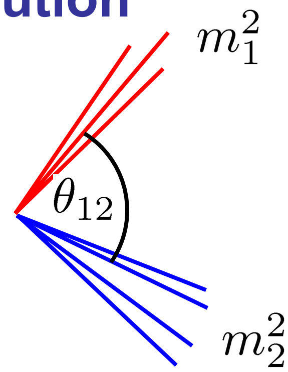
➊ PFA Goals revisited
➋ Algorithm Overview
➌ Status at LCWS07
➍ From LCWS to now
➎ On-going work
➏ Detector studies
➐ Some Comments
➑ Conclusions

❶ PFA Goals : revision

★Aim for jet energy resolution giving di-jet mass resolution similar to Gauge boson widths

★For a pair of jets have:

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



★For di-jet mass resolution of order $\Gamma_{\mathrm{W/Z}}$

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

⇒ $\sigma_{E_j}/E_j < 3.8\%$

+ term due to θ_{12} uncertainty

★Assuming a single jet energy resolution of normal form

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

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

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

E_{jj}/GeV	$\alpha(E_{jj})$
100	< 27 %
200	< 38 %

★ Typical di-jet energies at ILC (100-300 GeV) suggests jet energy resolution goal of $\sigma_E/E < 0.30/\sqrt{E_{jj}(\mathrm{GeV})}$

But Not The End

- ★ What jet energy resolution is really needed at the ILC ?
  - ★ NOT 30%/√E, NOT 3.8 %
  - ★ Ideally reach point where dominated by Z/W width
  - ★ NOT the same as

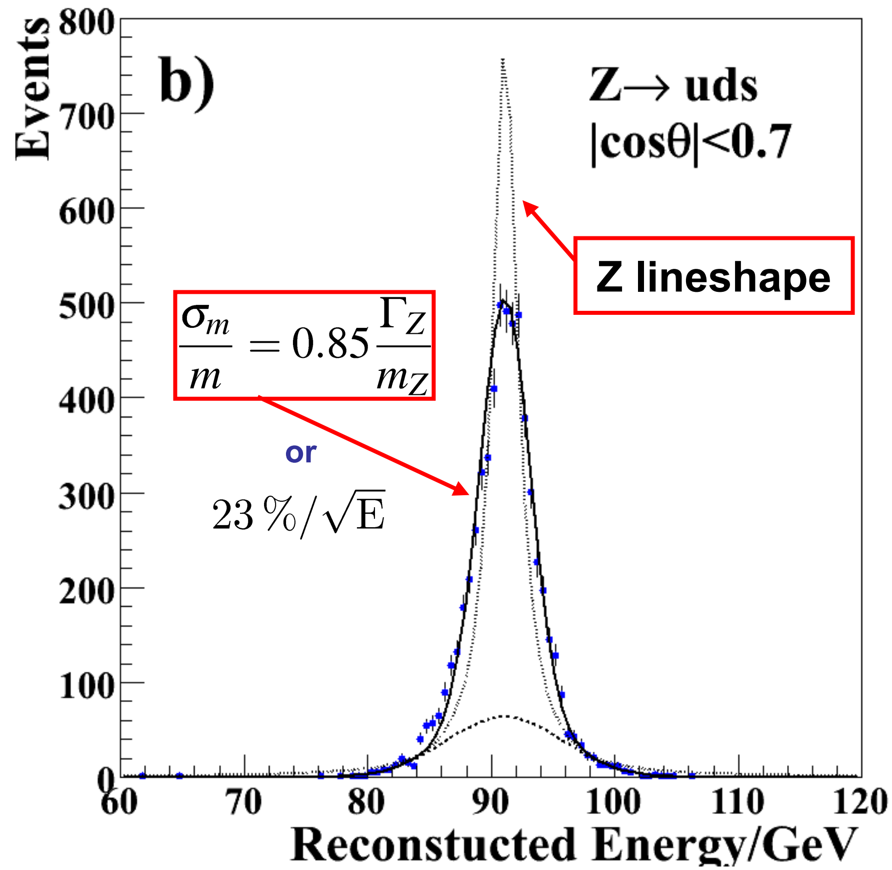
$$\frac{\sigma_m}{m} \sim \frac{\Gamma_Z}{m_Z}$$

FWHM

- ★ Aim for

$$\frac{\sigma_m}{m} < \frac{\Gamma_Z}{m_Z}$$

- ★ Significant advantages in further improvements ?



- ★ Push as hard as possible for best jet energy resolution
- ★ Ultimate criterion – "physics performance"…

❷ The PandoraPFA 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

★ Fairly "sophisticated" algorithm : 10^4 lines of code
  • of order 4 orders of magnitude less lines of documentation

Eight Main Stages:

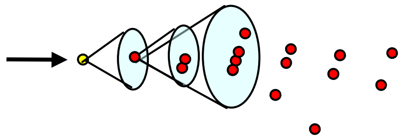
i. Preparation
ii. Loose clustering in ECAL and HCAL
iii. Topological linking of clearly associated clusters
iv. Courser grouping of clusters
v. Iterative reclustering
vi. Photon Identification/Recovery
vii. Fragment removal
viii. Formation of final Particle Flow Objects (reconstructed particles)

Algorithm overview

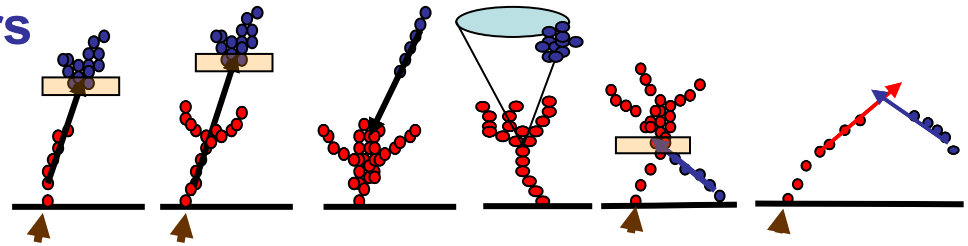
The Eight Main Stages:

i. Preparation/Tracking

ii. Loose clustering in ECAL and HCAL

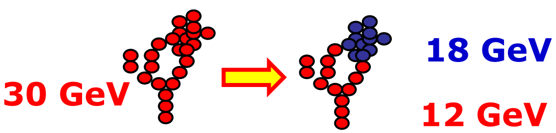


iii. Topological linking of clearly associated clusters

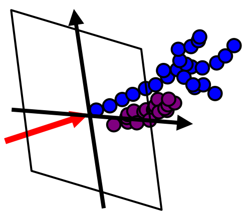


iv. Courser grouping of clusters

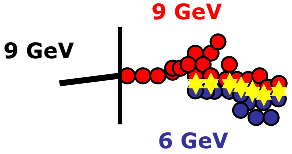
v. Iterative reclustering (using tracks)



vi. Photon Recovery



vii. Fragment Removal

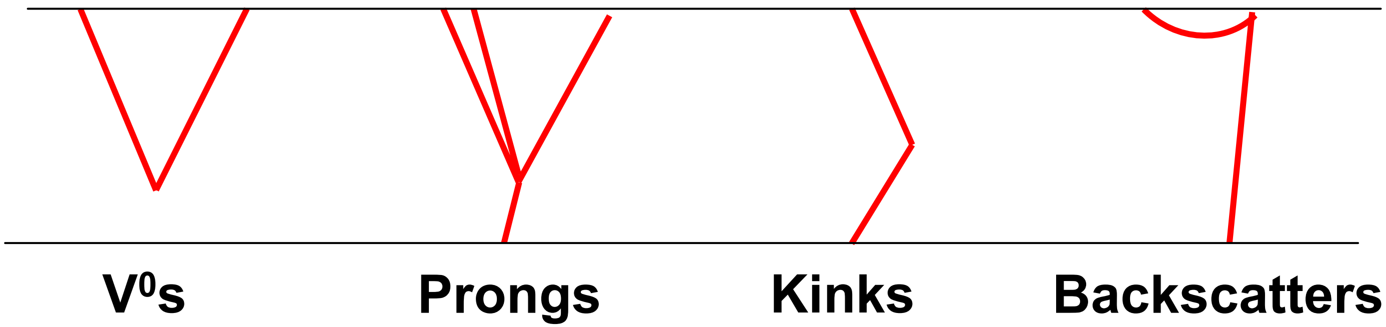


viii. Formation of final Particle Flow Objects

i) Tracking

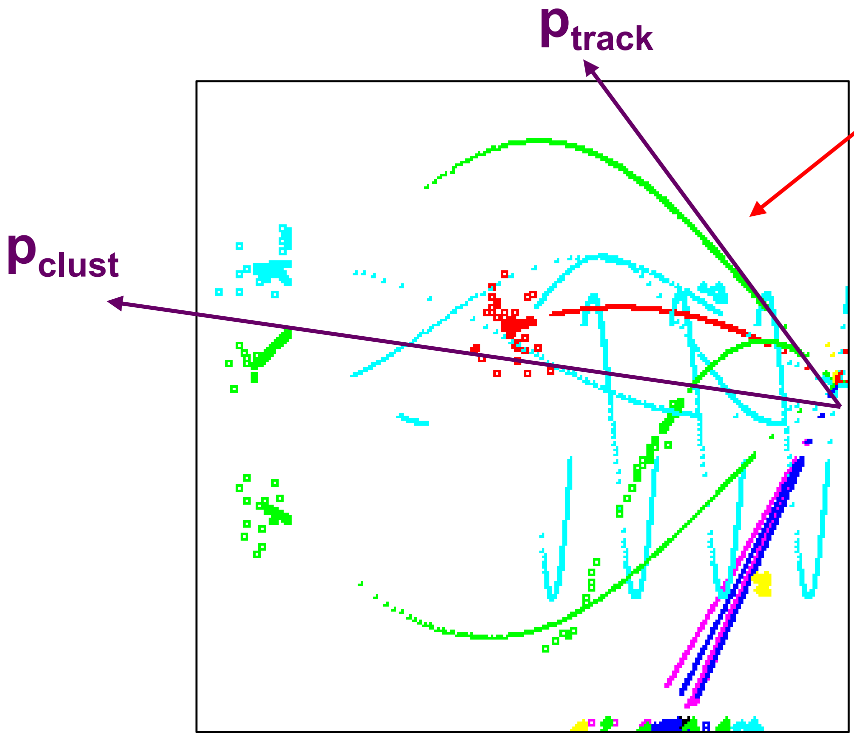
- ★The use of optimal use of tracking information in PFA is essential
- ★ Non trivial for looping tracks (even in a TPC)
- ★ Matching of tracks to endcap clusters is non-trivial
- ★ Probably spent at least as much time on tracking in PandoraPFA as clustering !
- ★ Big effort to use as many tracks in the event as possible
  - ★ helps particularly for lower energy jets
  - ★ motivation I : better energy resolution
  - ★ motivation II : correct measurement of direction
- ★**TPC-oriented:** take advantage of pattern recognition capability

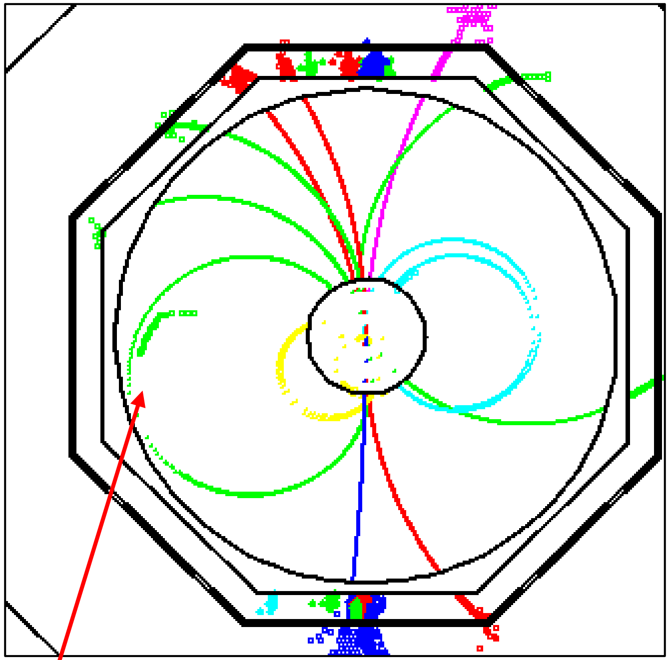
  (the algorithm would need modification for Si tracker)
- ★From fully reconstructed LDC tracks identify:



e.g. Tracking I : extrapolation

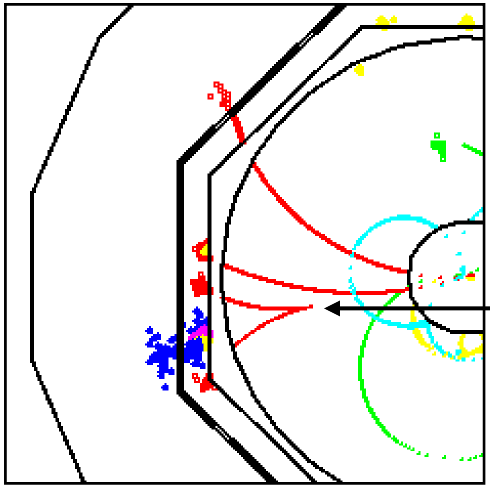
- If a track isn't matched to a cluster – previously track was dropped (otherwise double count particle energy)
- Not ideal – track better measured + direction





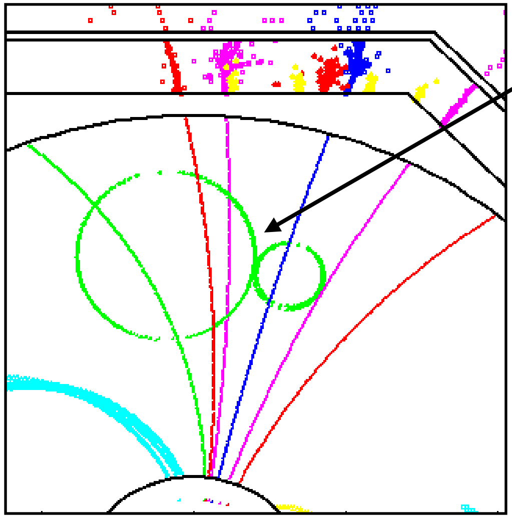
- e.g. try multiple (successively looser) track-cluster matching requirements e.g. "circle matching"
- Now only a few unmatched looping endcap tracks

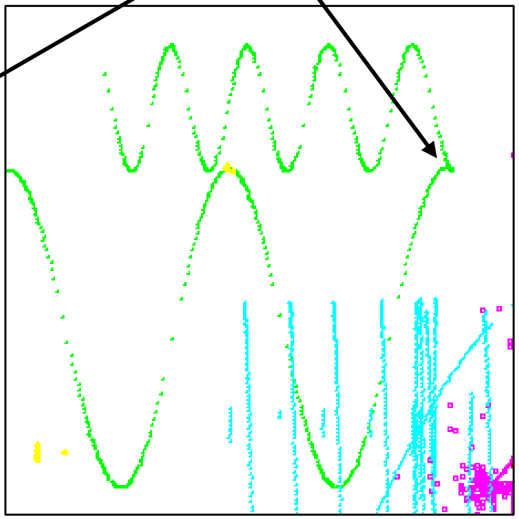
e.g. Tracking II : V^0s



- V^0 identification helps PFA as track momentum better measured than cluster energy
- Previously V^0 identification for the main topology
- Now extended to very low p_T tracks

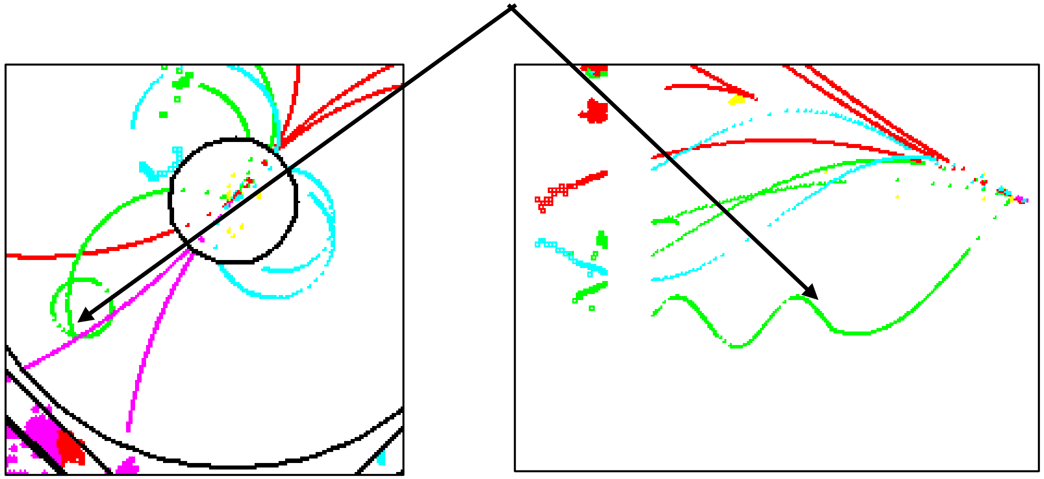
  (limited by low efficiency in Full Tracking code)
- Most important for lower energy jets



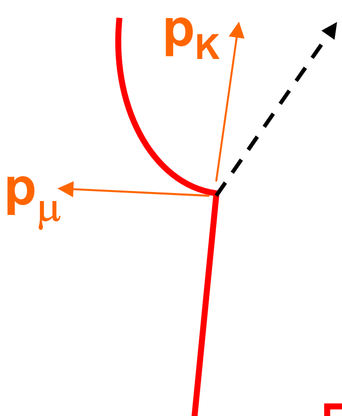


e.g. Tracking III: Kinks

- ★ Extended Kink finding to “loopers”



- ★ Improved (but still fairly crude) reconstruction missing energy

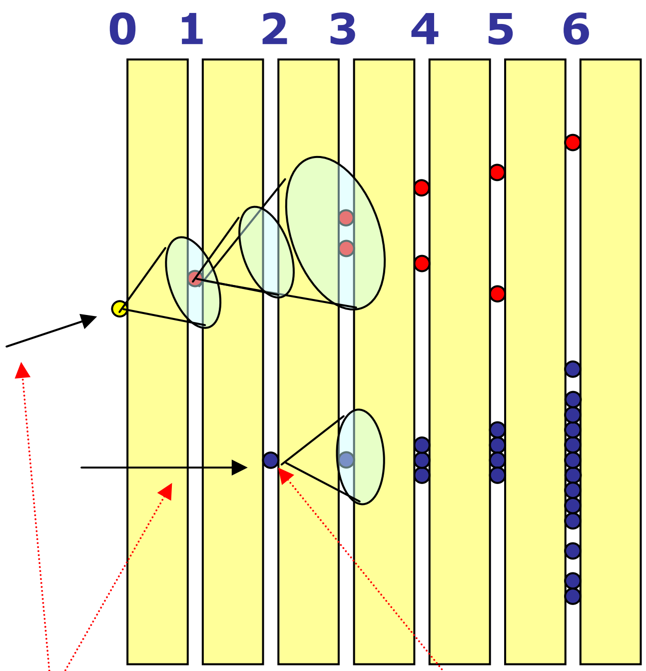


- ♦ Consider physics hypothesis, e.g. $K^{\pm} \rightarrow \mu^{\pm}\nu$
- ♦ Use Helix fits to start and end of tracks to reconstruct missing particle e.g. ν
- ♦ Can then reconstruct primary mass
- ♦ If consistent with hypothesis, e.g. m_K use primary track for PFO four-momentum

PandoraPFA reconstructs (some) neutrinos !

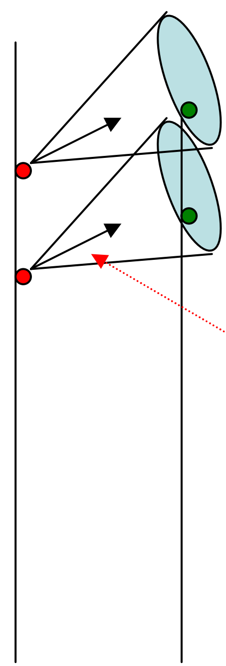
ii) ECAL/HCAL Clustering

- ★ 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

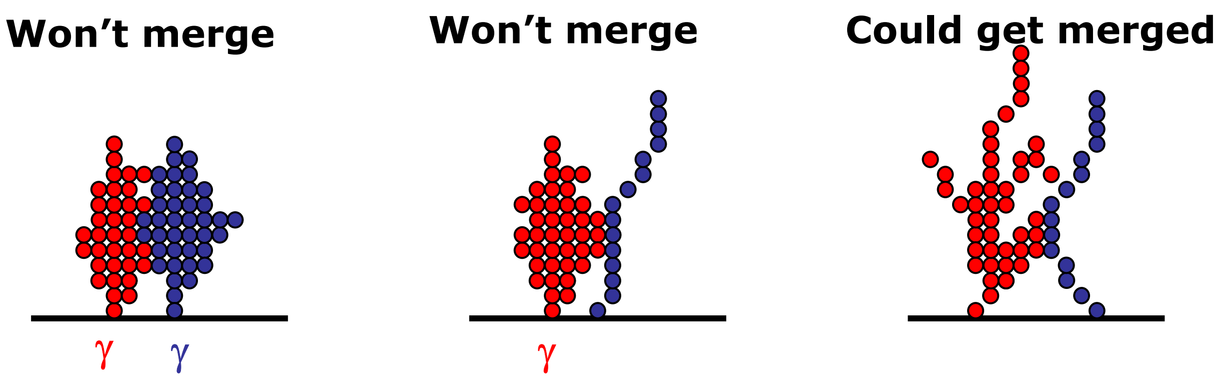
Parameters:
- cone angle
- additional pixels

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

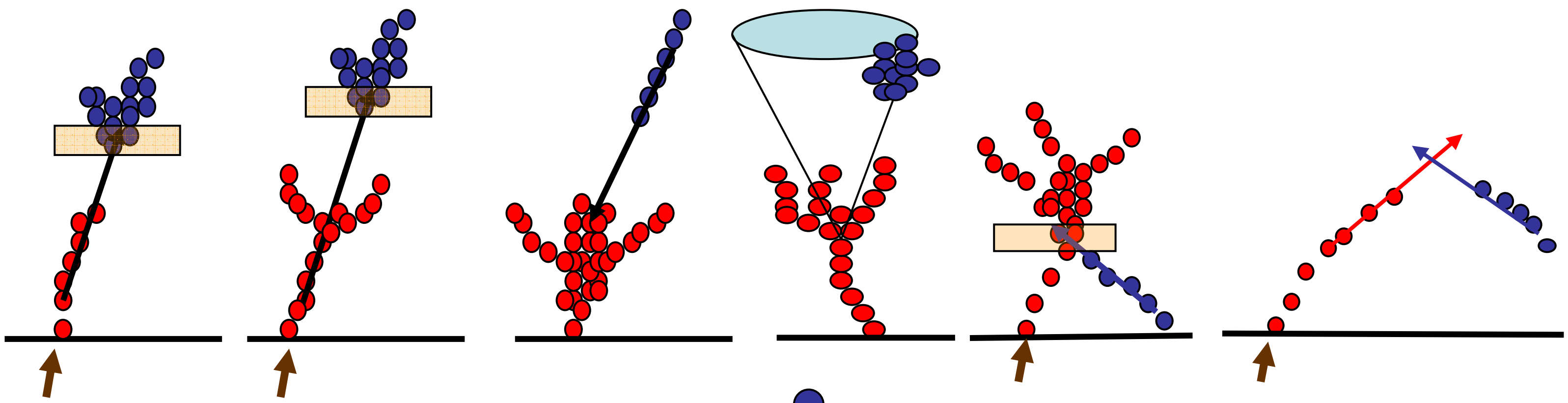
- Photon ID plays important role (but does not drive clustering)
- VERY SIMPLE "cut-based" photon ID applied to all clusters
- Clusters tagged as photons are immune from association procedure – just left alone



★ **Clusters associated using a number of topological rules**

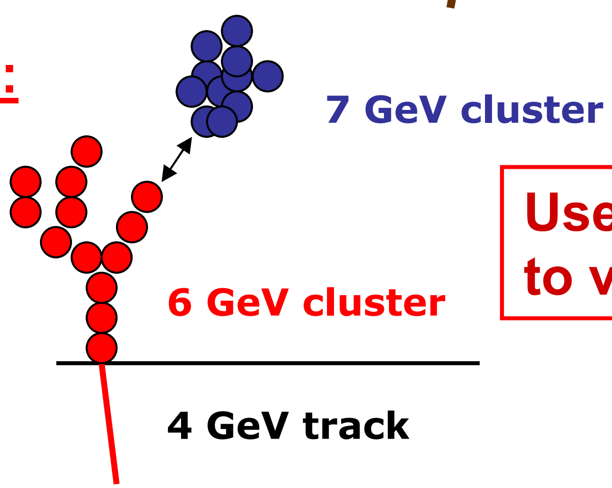
Clear Associations:

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



Less clear associations:

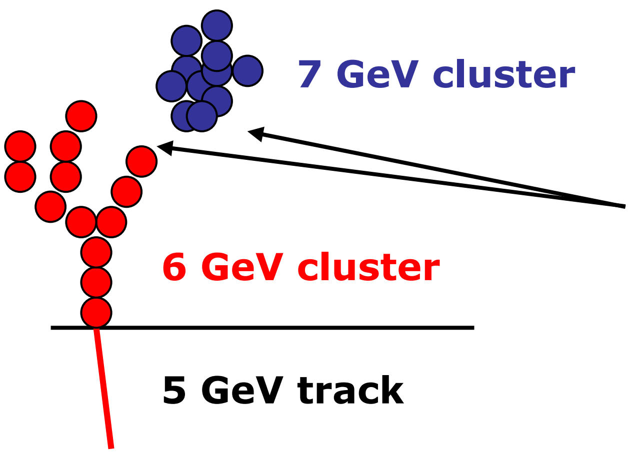
e.g. **Proximity**



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.

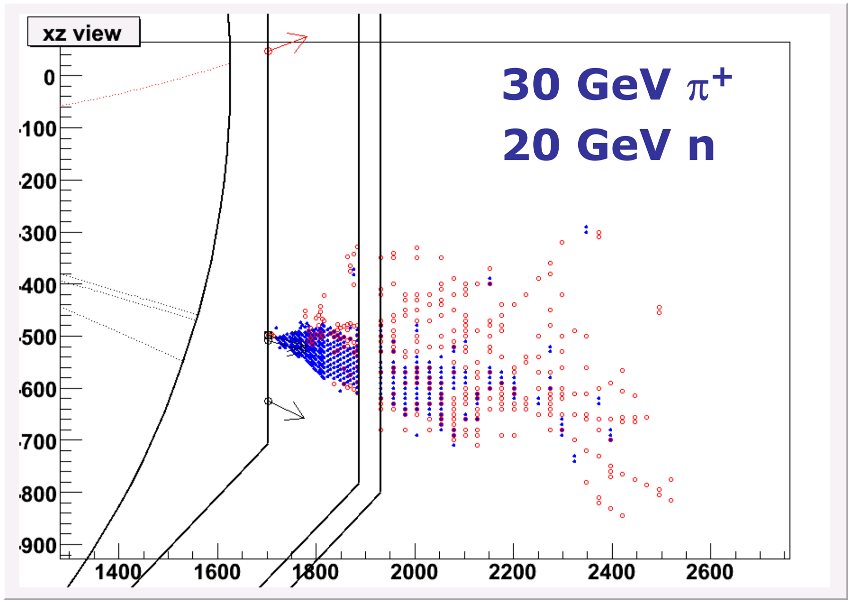


This cluster association would be forbidden if $|E_1 + E_2 - p| > 3\,\sigma_E$

Provides some protection against silly mistakes

v) Iterative Reclustering

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



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

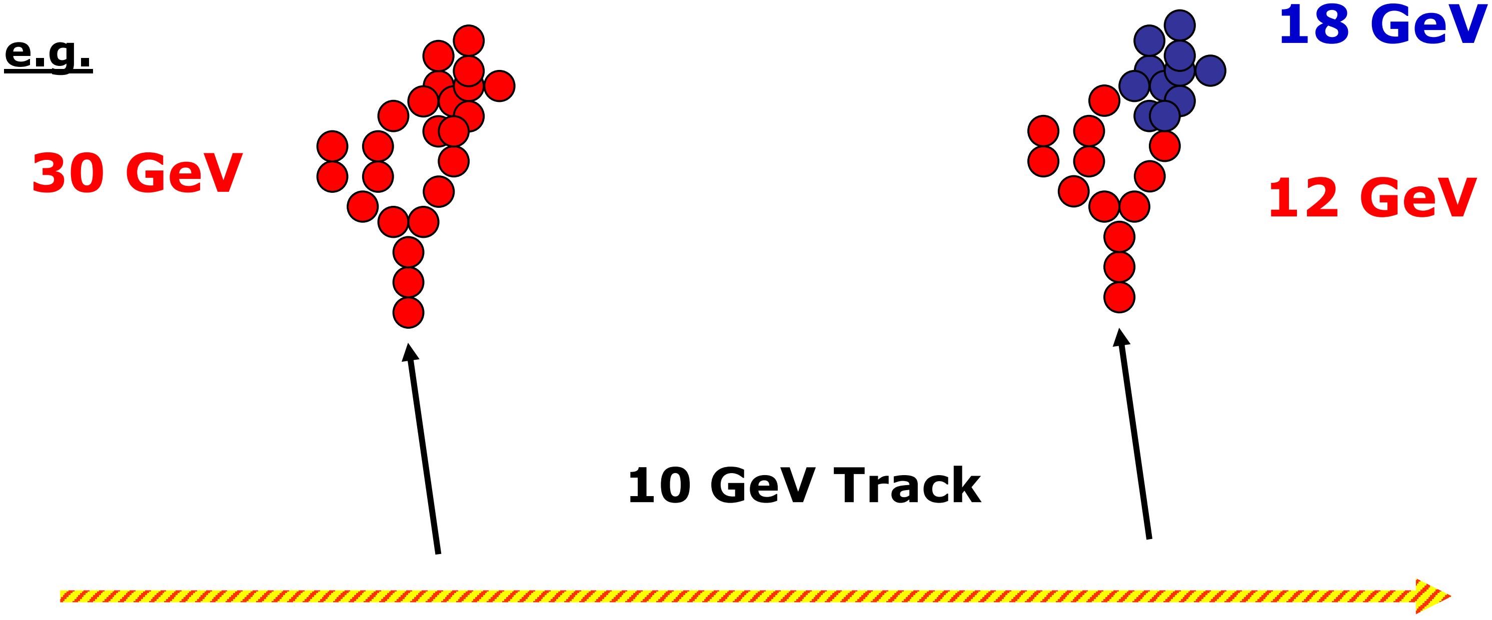
The ONLY(?) way to address this is "statistically"

→

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.

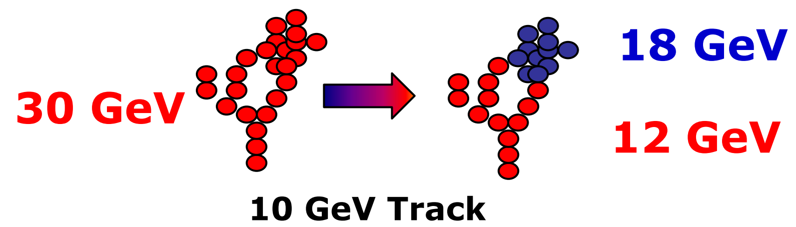


Change clustering parameters/alorithm until cluster splits and get sensible track-cluster match

Iterative Reclustering Strategies

① Cluster splitting

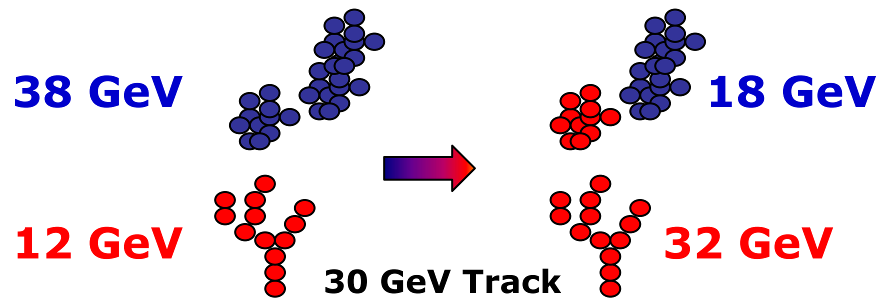
Reapply entire clustering algorithm to hits in "dubious" cluster. Iteratively reduce cone angle until cluster splits to give acceptable energy match to track



★ + plug in alternative clustering algorithms

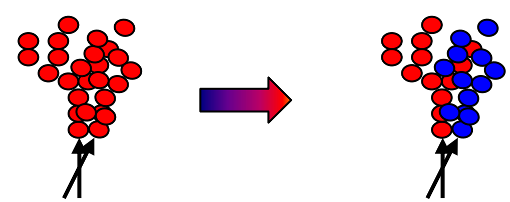
② Cluster merging with splitting

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.



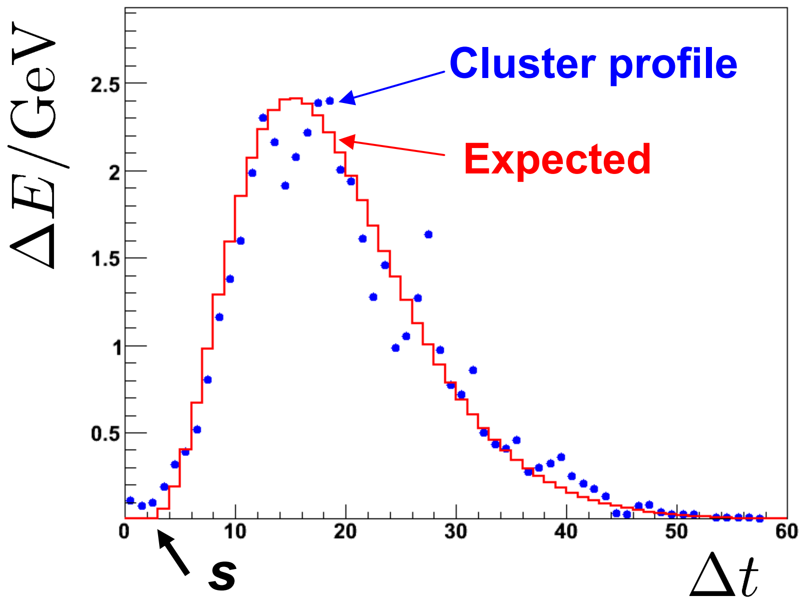
Very Important for higher energy jets

vi) Photon ID/Recovery

- Use simple cut-based photon ID in the early (CPU intensive) stages of PandoraPFA
- In the final stages, use improved photon ID based on the expected EM longitudinal profile for cluster energy E_0

$$\Delta E = E_0 \frac{(t/2)^{a-1} e^{-t/2}}{\Gamma(a)} \Delta t \qquad a = 1.25 + \frac{1}{2} \ln E_0/E_c$$

- Convert cluster into energy depositions per radiation length
  (use cluster to determine the layer spacing, i.e. geometry indep.)



- Shower Profile fixed by cluster energy
- But fit for best shower start, ***s***
- Normalise areas to unity and calc.

$$f = \sum_i |o_i - e_i|$$

- Gives a measure of fractional
- disagreement in obs/exp profiles
- Use ***f*** and ***s*** to ID photons

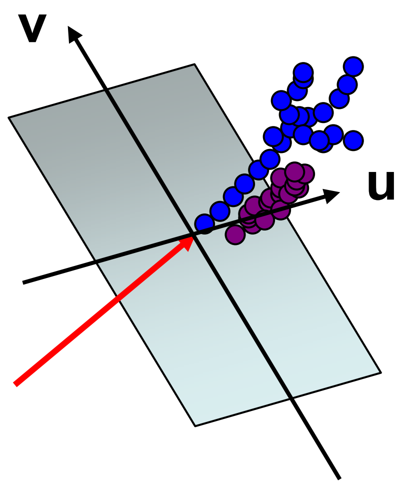
Photon Recovery

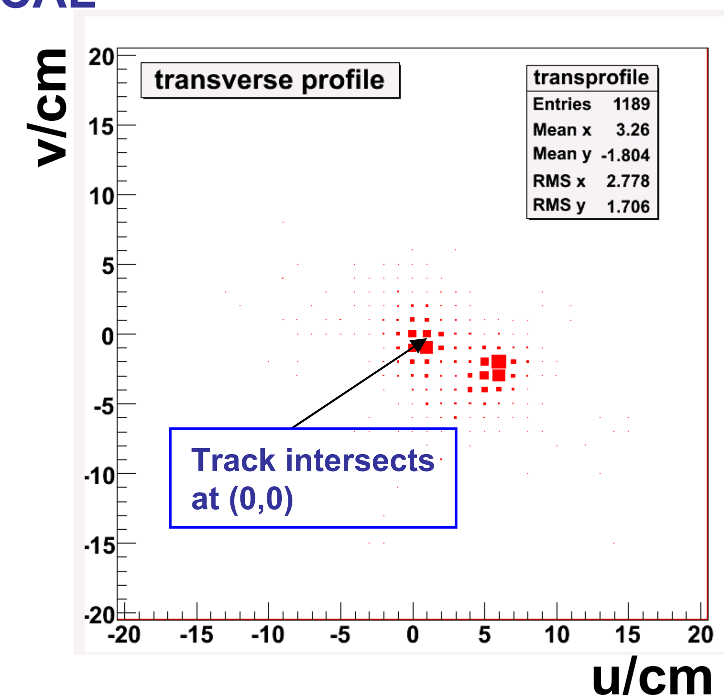
★ With cone clustering algorithm, photons close to early showering charged hadrons can be merged into a single cluster.

★ Use longitudinal + transverse profile to recover these

★ Essentially, 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…



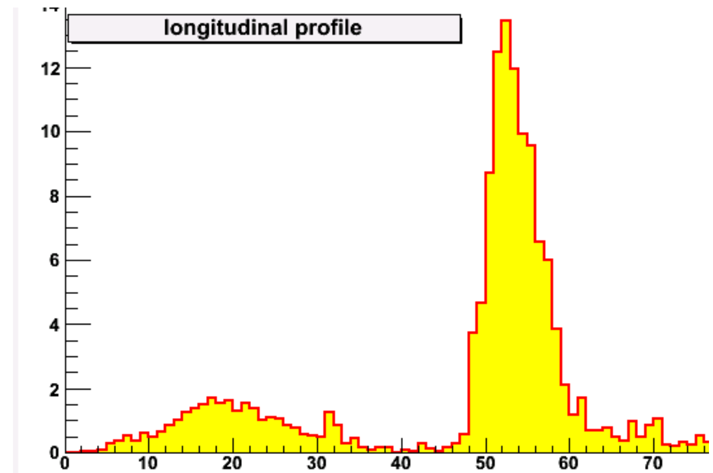


★ If there is an isolated peak not associated with "track peak" make new photon cluster if track energy and remaining cluster energy still statistically compatible with track momentum + cluster passes photonID

Use profiles to “dig out” photons overlapping with hadronic clusters:

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

- Implemented by looking at longitudinal profile of “shower”

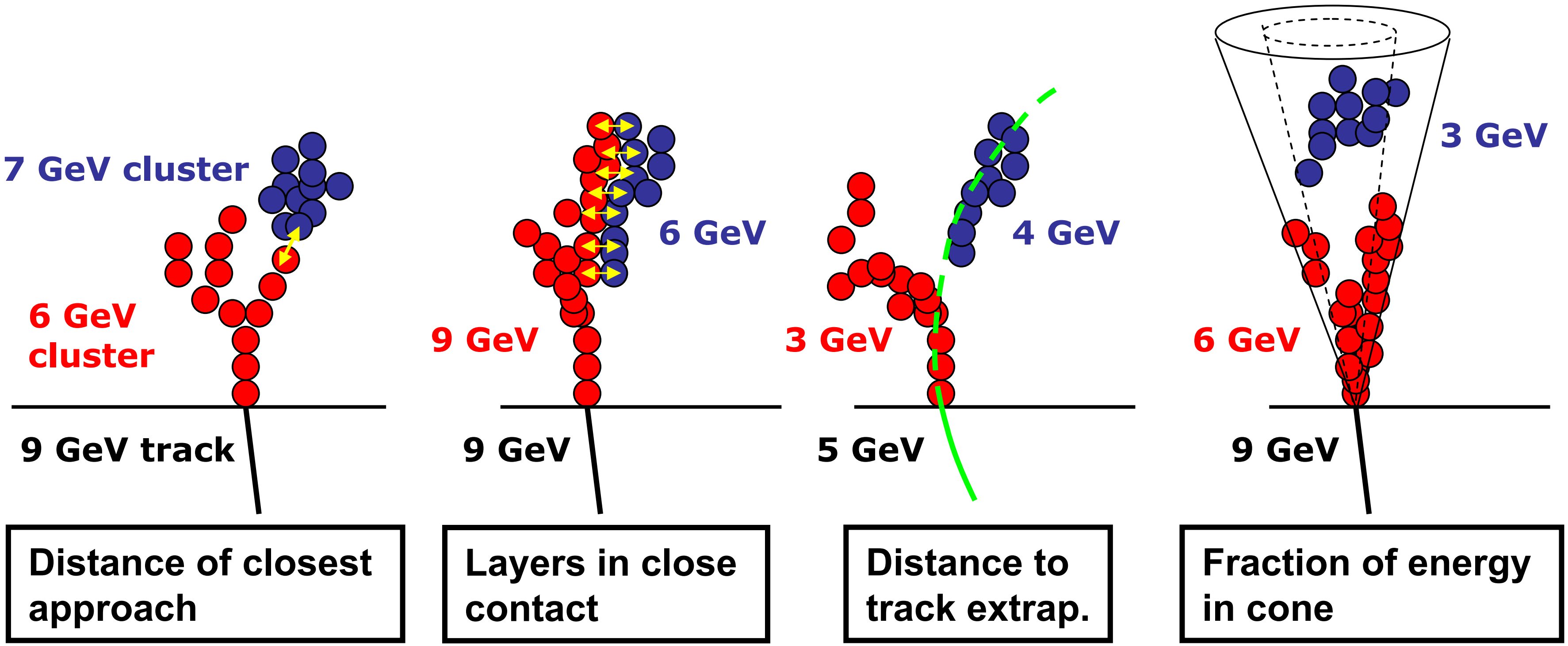


Only allowed if it results in acceptable track-cluster energy consistency…

NOTE: in PandoraPFA, photon identification is an “iterative”, rather than one-off process: different levels of sophistication applied at different stages of algorithm

viii) Fragment removal : basic idea

★ Look for "evidence" that a cluster is associated with another



★ Convert to a numerical evidence score **E**

★ Compare to another score "required evidence" for matching, **R**, based on change in E/p chi-squared, location in ECAL/HCAL etc.

★ If **E > R** then clusters are merged

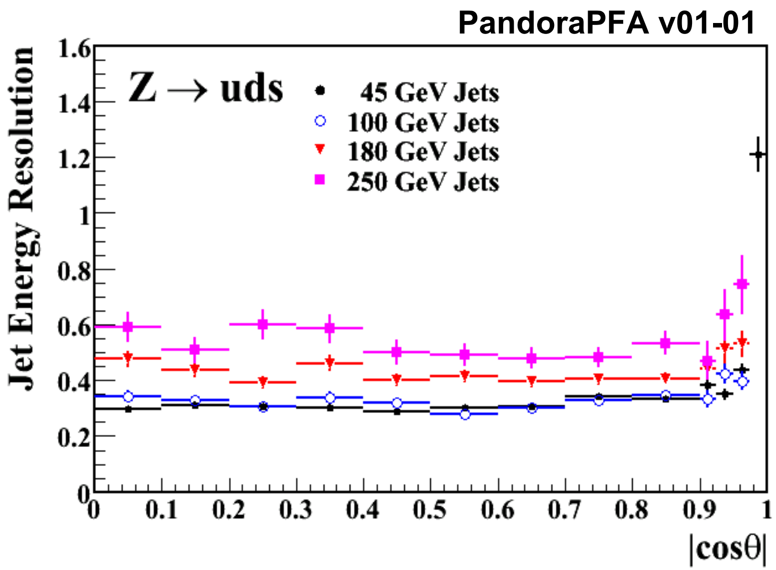
★ Rather ***ad hoc*** but works well – but works fairly well

❸ Status at LCWS07

★ Full simulation studies using the LDC ILC detector concept with the PandoraPFA algorithm. Use $Z \rightarrow u\bar{u}, d\bar{d}, s\bar{s}$ decays at rest to benchmark performance

rms90* PandoraPFA v01-01

E_{JET}	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $\lvert\cos\theta\rvert<0.7$	σ_E/E_j
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 **below** 100 GeV achieve $\sigma_E/E < 0.30/\sqrt{E_{jj}(\text{GeV})}$

★ Perhaps more importantly, for jet energies **above** ~75 GeV achieved

$$\sigma_{E_j}/E_j < 3.8\%$$

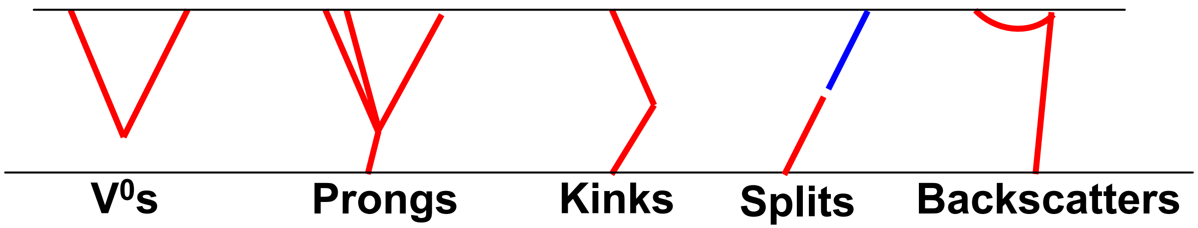
★ **Post-LCWS emphasis shifted to improving low energy performance,** important in likely initial phase of ILC at √s ~ 200-500 GeV

❹ From LCWS07 to RAL

Step 1: improve low energy performance

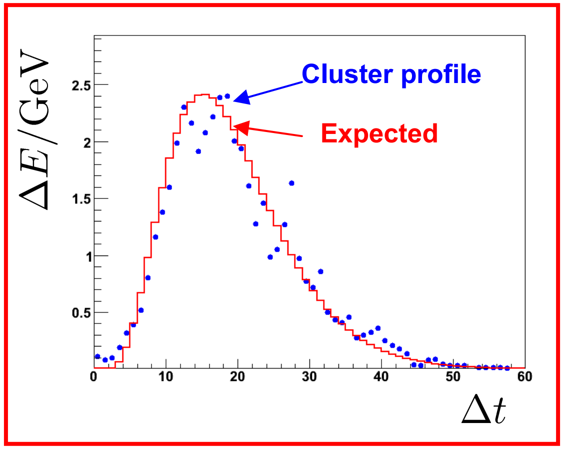
★ **Technical Improvements/bug fixes**
- ♦ reduced memory footprint (~ factor 2) by on-the-fly deleting of temporary clusters, rather than waiting to event end

★ **Improved track ID**

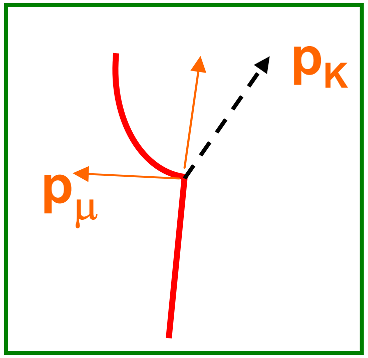


★ **New photon Identification**

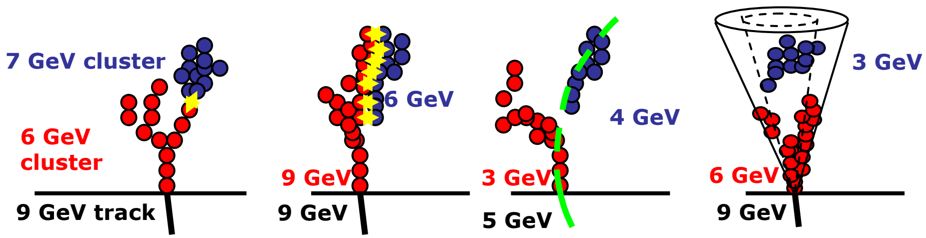
EM shower profile →



★ **Particle ID**
- ♦ Much improved particle ID : electrons, conversions, $K_S \to \pi^+\pi^-$, $\Lambda \to \pi^- p$ (no impact on PFA)
- ♦ Some tagging of $K^\pm \to \mu^\pm\nu$ and $\pi^\pm \to \mu^\pm\nu$ kinks →



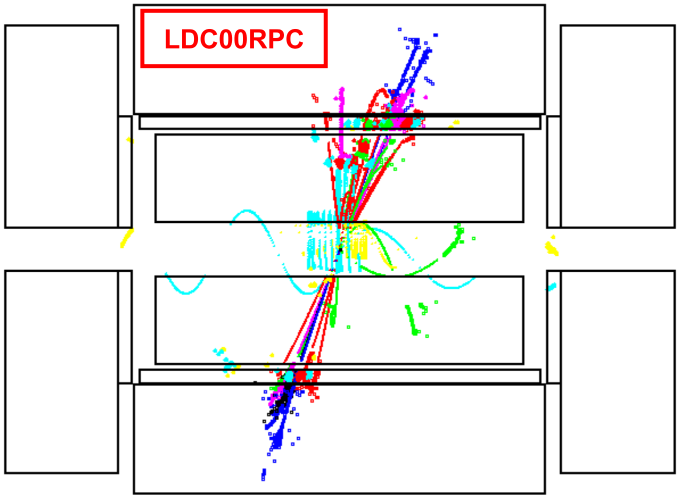
★ **More sophisticated identification of neutral fragments**



From LCWS to RAL cont.

Step 2: increase functionality

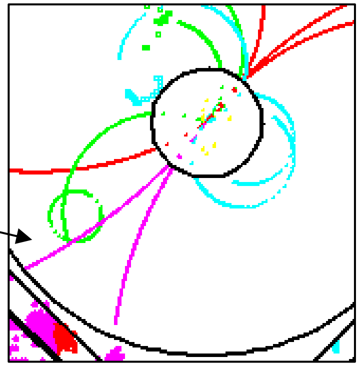
★ Now compatible with **digital HCAL** (and digital ECAL e.g. MAPs-based)



(PandoraPFAv02 +trackCheater)	E_{JET}	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $\lvert\cos\theta\rvert<0.7$
LDC00Sc	100 GeV	29.3 %
LDC00RPC	100 GeV	30.3 %

• very similar performance
(digital PFA is not fully optimised)

★ Can now use either **TrackCheater** or **FullLDCTracking**
- ♦ required rewrite of V^0 finding + tweaks for **kinks**
- ♦ note: **FullLDCTracking** does not find non-vertex curlers, i.e. reduced kink/V^0 efficiency



Step 3: compatibility with new LDC models "ILD ready"

★ Include LumiCAL, ECAL Plug. + include MUON hits (not yet used)
★ Made more robust – better error/warning reporting

LCWS → RAL: LDC00 (Tesla TDR)

Cheated Tracks

LCWS07 PandoraPFA v01-01

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

PandoraPFA v02-α

E_{JET}	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $\lvert\cos\theta\rvert<0.7$	σ_E/E_j
45 GeV	0.226	3.3 %
100 GeV	0.293	2.9 %
180 GeV	0.392	2.9 %
250 GeV	0.534	3.3 %

FullLDCTracking

PandoraPFA v02-01

E_{JET}	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $\lvert\cos\theta\rvert<0.7$	σ_E/E_j
45 GeV	0.235	3.5 %
100 GeV	0.306	3.1 %
180 GeV	0.427	3.2 %
250 GeV	0.565	3.6 %

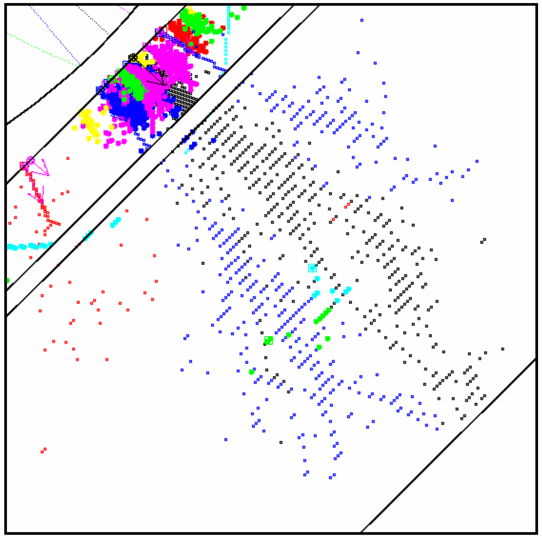
★ **For LDC00:**

- Only slight degradation when using FullLDCTracking
- Difference may be due to degraded kink finding
- Track quality cuts not fully optimised (i.e. how many hits required, use Si only tracks?)

Bottom Line…

- ★ Particle flow can achieve ILC "goal" of $\sigma_E/E_j < 3.8$ %
- ★ For lower energy jets Particle Flow gives unprecedented levels of performance, e.g. @ 45 GeV : 3.5% c.f. ~10% (ALEPH)
- ★ "Calorimetric" performance (α) degrades for higher energy jets
  + current code is not perfect - can do better
- ★ would like to investigate the ultimate limit of PFA calorimetry at higher energies…

PARTICLE FLOW CALORIMETRY WORKS !



… at least in simulation

Hadron Shower Models

★ People have rightly expressed concerns about sensitivity to hadron shower models…

★ First look: compare **LHEP** & **QGSP_BERT** models.

★ Large model differences
- 30 % in raw energy deposition
- longitudinal/transverse development

(PandoraPFAv02 +trackCheater)		E_{JET}	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $\lvert\cos\theta\rvert<0.7$
LDC00Sc	**QGSP_BERT**	**45 GeV**	**22.6 %**
LDC00Sc	**LHEP**	**45 GeV**	**23.2 %**
LDC00Sc	**QGSP_BERT**	**100 GeV**	**29.3 %**
LDC00Sc	**LHEP**	**100 GeV**	**30.2 %**

★ Differences rather small (+code not re-optimised for LHEP)

★ Sensitivity to Hadronic shower development may not be so large
- needs more study
- ultimately CALICE data will show the way

❺ Ongoing Work

★ Main emphasis of recent work
- Preparation for ILD mass simulation/reconstruction
- Make PandoraPFA fully compatible with new LDC detector model
  - significant changes to simulation (almost all sub-detectors)
  - including more realism…
- Won't discuss details here:
  - but be aware that some parts of code now LDC specific, e.g. energy corrections for ECAL gaps
  - all driven by switches in configuration – won't impact non-LDC studies (as long as run correctly – pay attention to warning/error messages)

BAD: realism in detector model → more complexity in software
GOOD: also gives insight into sub-detector design

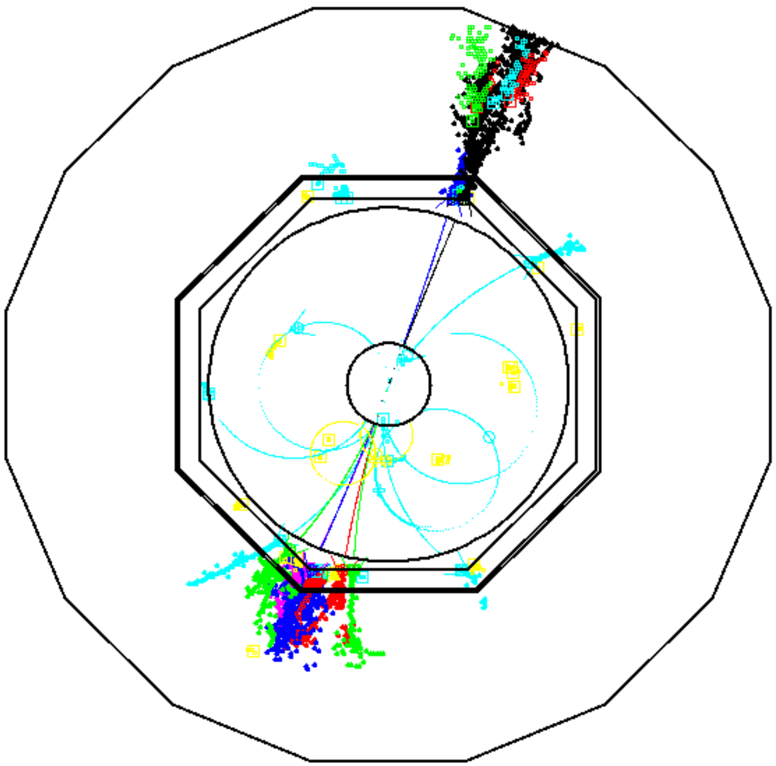
★ Next, how to improve PandoraPFA ?
- emphasis now on high energy performance

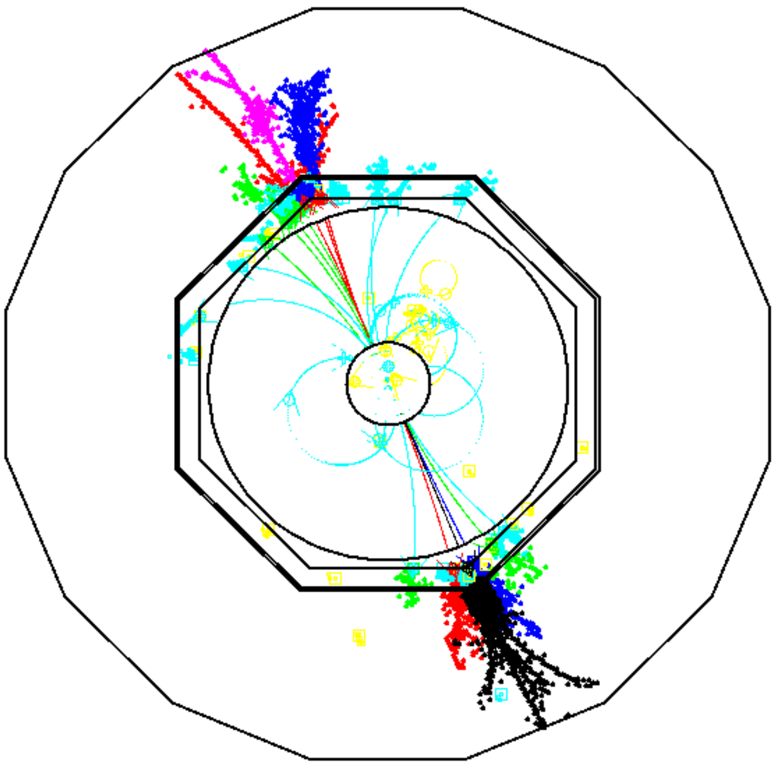
When PandoraPFA Goes Bad

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

i) Leakage

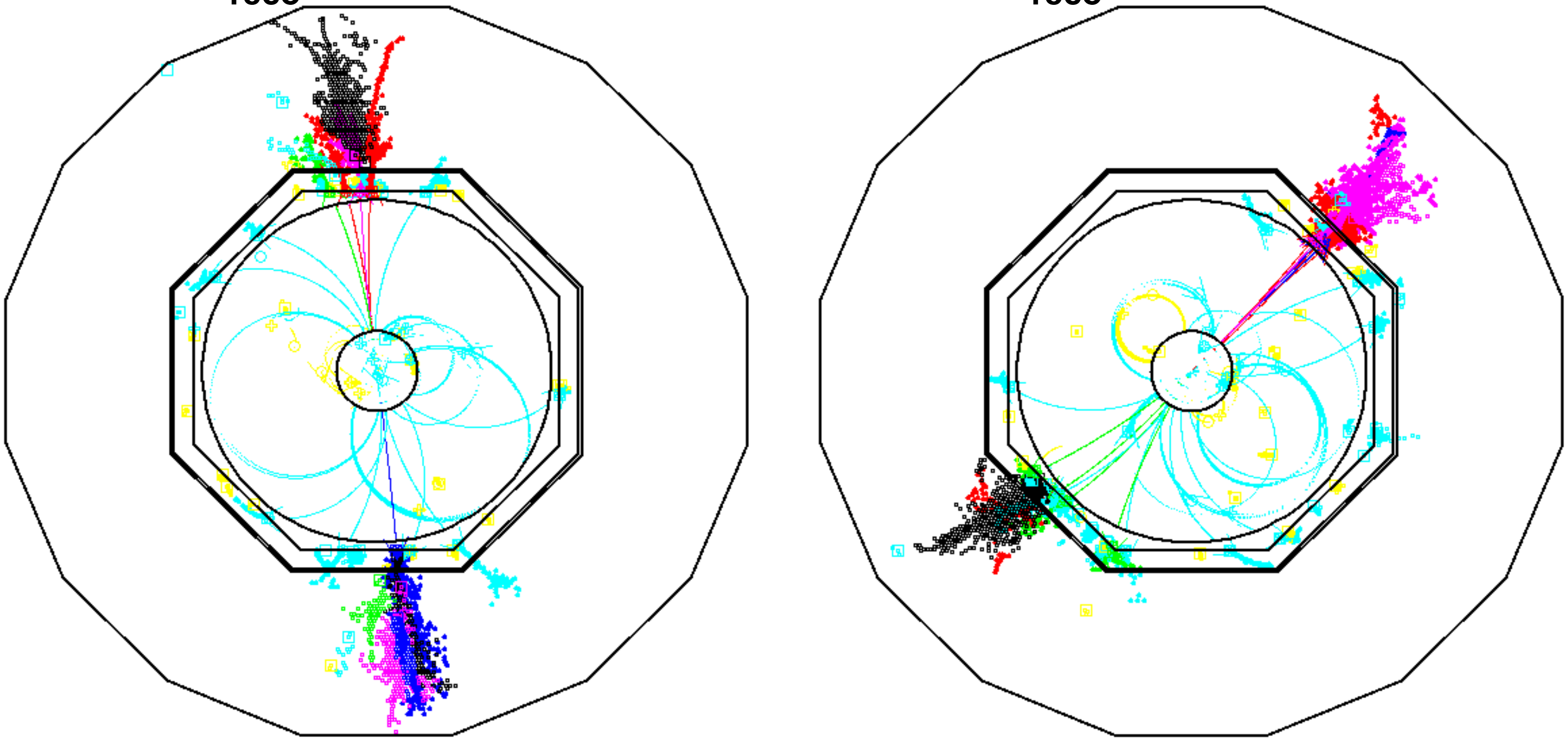
E_{reco} = 337 GeV **E_{reco} = 338 GeV**





ii) Confusion

$E_{reco} = 382$ GeV $E_{reco} = 391$ GeV

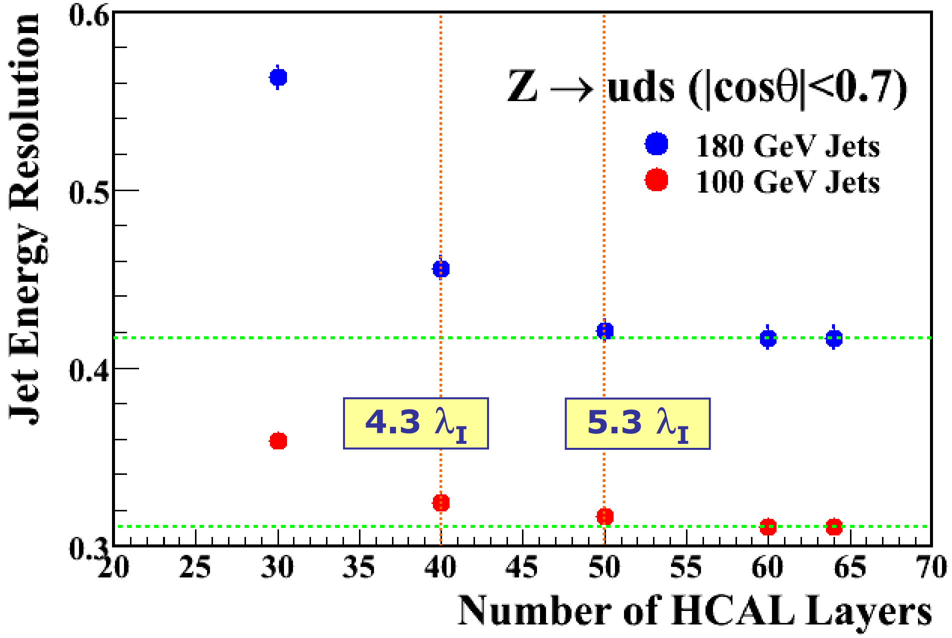


★ **How to improve performance ?**

- **plan to utilise detailed structure of hadronic showers…**
- **i.e. try different clustering algorithms in reclustering**
- **PandoraPFA being restructured for this and to interface to GLD strip clustering**

❻ Detector Optimisation (nothing new until ILD studies)

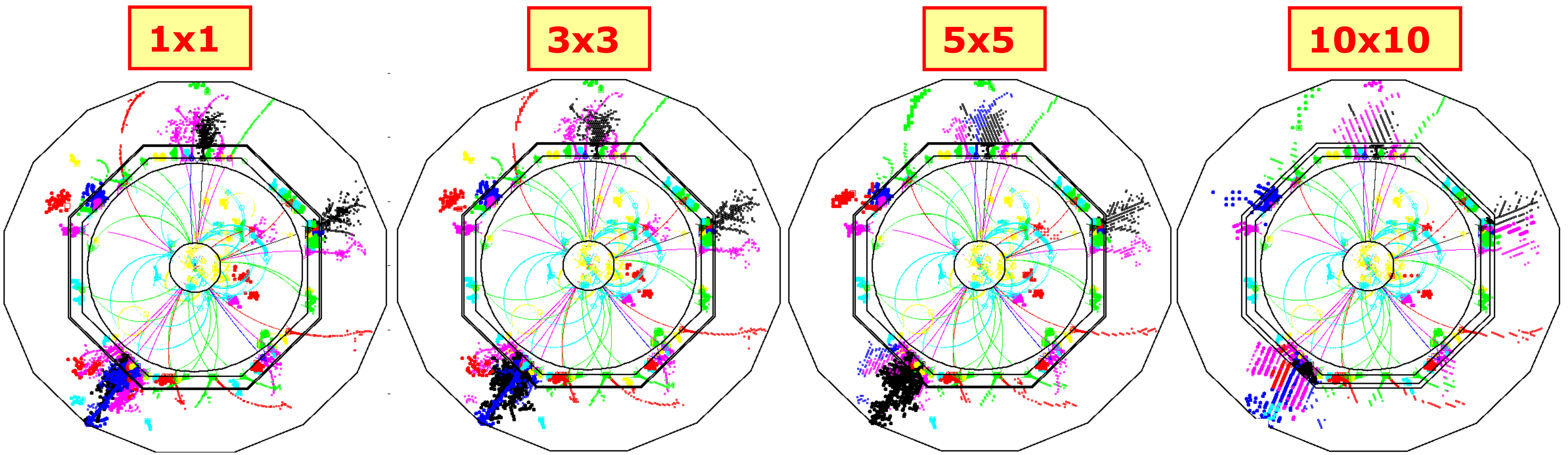
★ Investigated HCAL Depth (interaction lengths)
- Generated **Z→uds** events with a large HCAL (63 layers)
  - approx 7 λ_I
- In PandoraPFA introduced a configuration variable to truncate the HCAL to arbitrary depth
- Takes account of hexadecagonal geometry

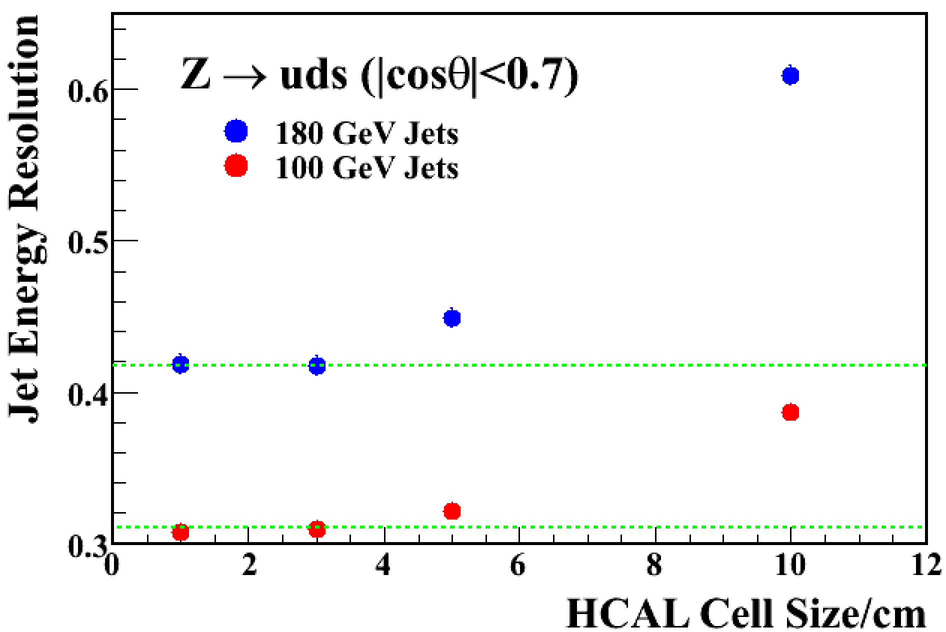


♦ HCAL leakage is significant for high energy
♦ Argues for ~ **5 λ_I HCAL**

NOTE: no attempt to account for leakage – i.e. using muon hits - this is a worse case

e.g. change HCAL tile size 1x1 → 10x10 mm²

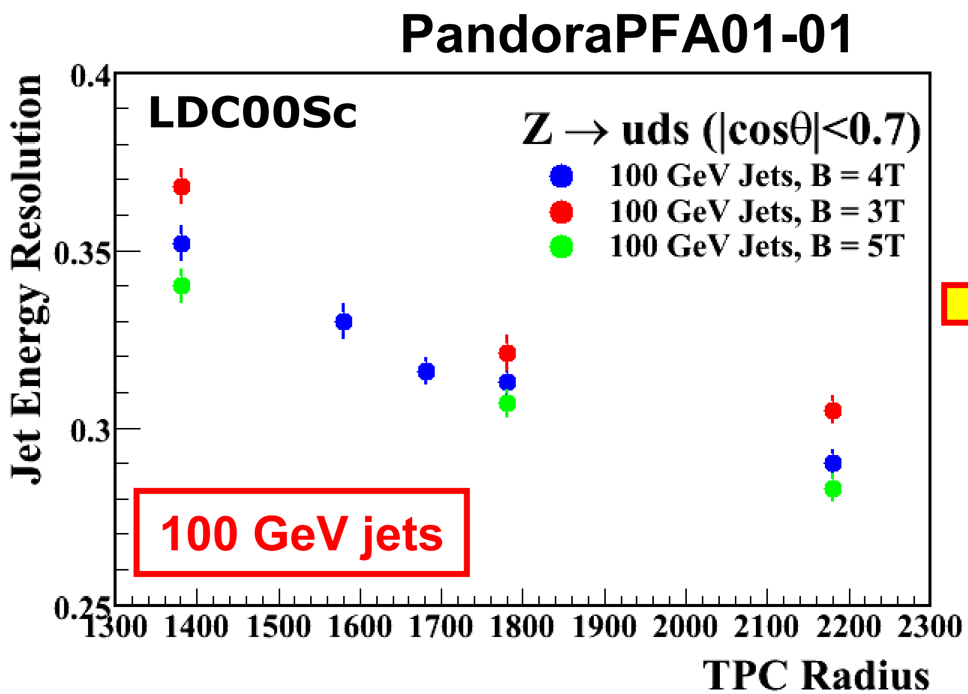




"Preliminary Conclusions"

- 3x3 cm² cell size
- No advantage → 1x1 cm²
  - physics ?
  - algorithm artefact ?
- 5x5 cm² degrades PFA
  - Does not exclude coarser granularity deep in HCAL

Radius vs Field



Radius more important than B-field

NOT BR^2 !!!!

- ★ These types of studies are interesting.
- ★ They highlight what matters for a particular PFA implementation
- ★ But how does this feed through to physics ?

PandoraPerfectPFA

- ★ **PerfectPFA option in Pandora**
  - <parameter name="PerfectPFA" type="int"> 1 </parameter>
- ★ **Uses MC information to create the ProtoClusters**
- ★ **The rest of the algorithm is the same**
- ★ **Useful tool**
- ★ **Process same events/same analysis and compare PFA to perfect PFA**

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

E_{JET}	$\sigma_E/E = \alpha/\sqrt{(E/GeV)}$ $\lvert\cos\theta\rvert<0.7$	
	PerfectPandora	**PandoraPFA**
100 GeV	**0.22**	**0.31**
180 GeV	**0.30**	**0.43**
	Cheated Tracks+PFA	Reco. Tracks+PFA

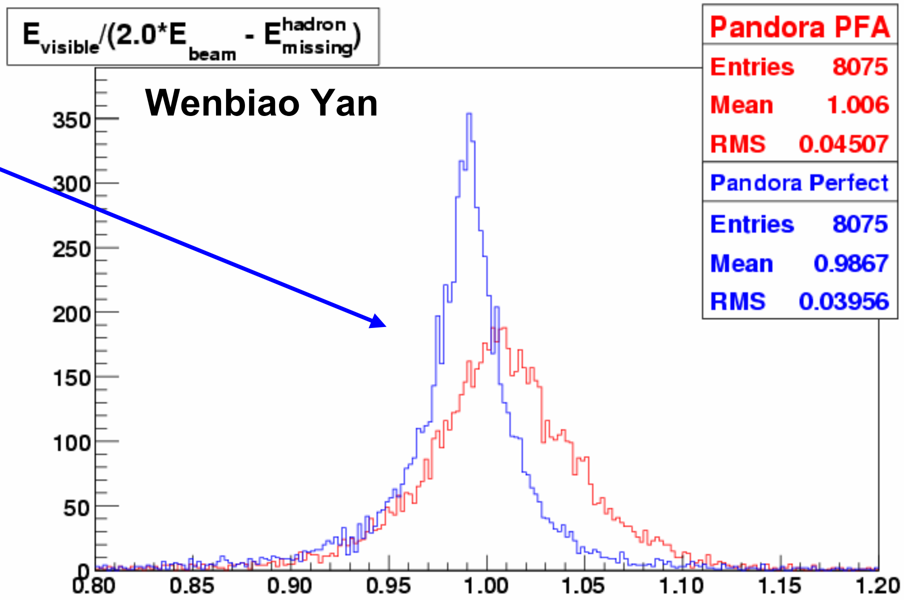
Still someway to go – needs study

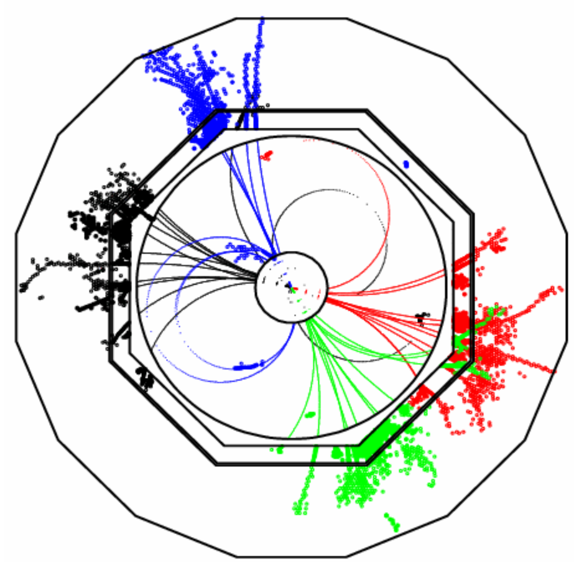
PFA impact in a real physics process

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)

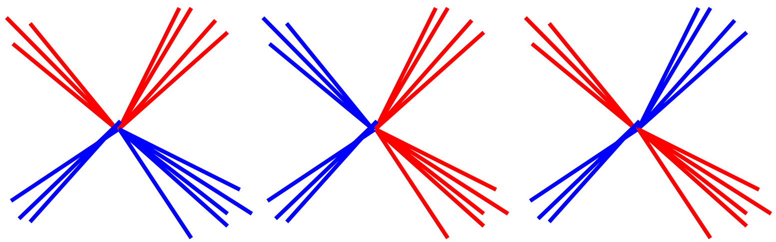




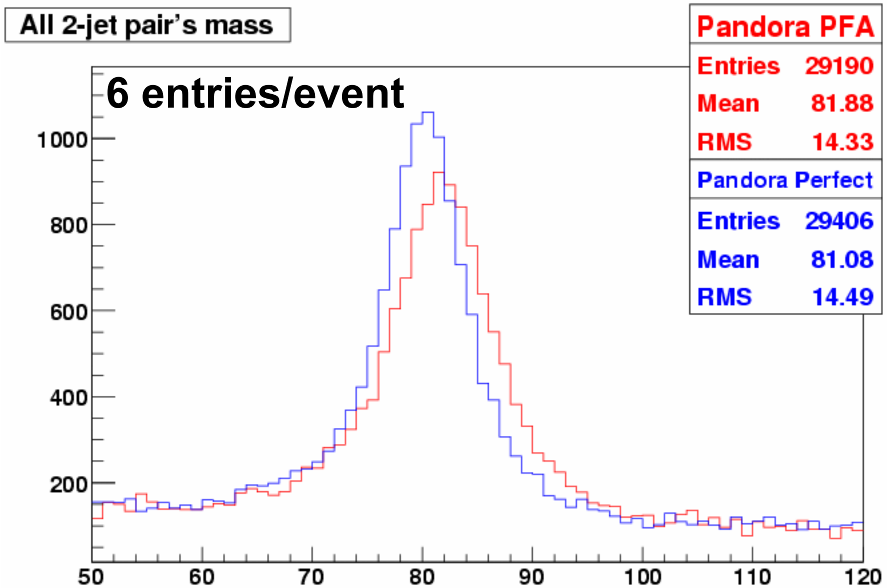
★ Does this difference make it through to a physics analysis (i.e. after jet finding/ jet pairing) ?

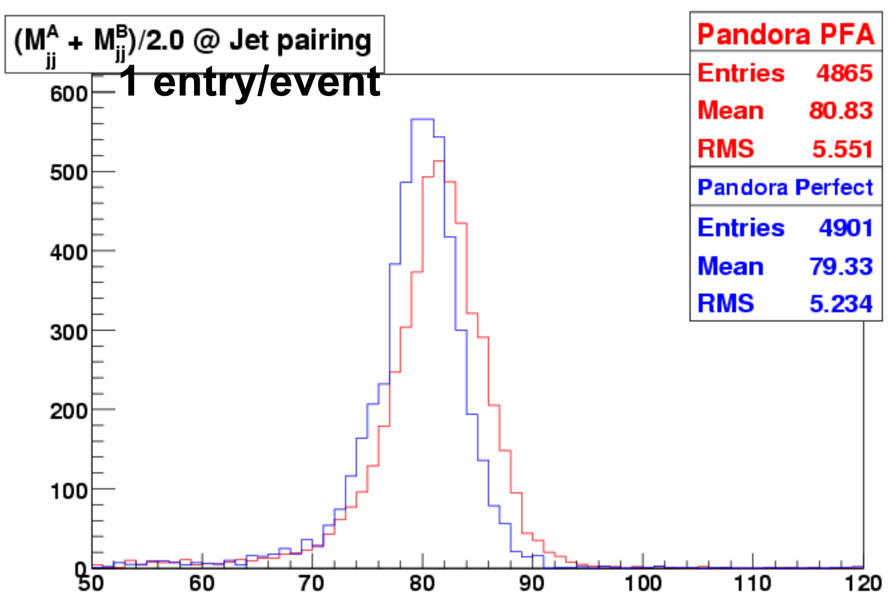
★Force event into 4 jets (Durham)

★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

⇒ Need real physics studies

❼ Random Comments

Deficiencies:

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

★ Very little has been optimised:
- Photon ID – good be better
- Photon Recovery – crude
- Fragment Removal – very crude

Plenty of room for improvement

Why does PandoraPFA work reasonably well ?

★ PFA = much more than clustering
- basic clustering algorithm developed in about a week shortly after Snowmass – essentially unchanged

★ Lessons learnt in developing code:
- advantages in having a single "coherent" approach
- always concentrated on optimising jet E performance, not photonID efficiencies etc.
- extreme care with all stages – avoid unnecessary mistakes
- great care needed in track/cluster matching
- use of track momentum – cluster energy to spot to PFA errors absolutely vital
- PFA reconstruction is an iterative process, use more sophisticated techniques as knowledge of event improves

❽ Summary/Outlook

Performance:

★ PandoraPFA with FullLDCTracking achieves good performance

$\sigma_{E_j}/E_j < 3.6\%$ for 45-250 GeV jets LDC00

★ Particle Flow approach to calorimetry at ILC now proven

★ Now want to investigate full potential of PFA for ILC/…

Towards ILD:

★ v02-01 works with latest LDC detector model

★ Current effort has shifted towards optimisation for LDC models to be used in ILD studies (will be used to process N Million events)

Medium-term (Summer) PFA optimisation/studies:

★ Implement new clustering algorithms for reclustering

★ Potential for many interesting detector studies

- Intend to start with detailed PFA / HCAL study
- New LDC HCAL model now in Mokka makes this possible

Finally, a couple of SiD-centric comments:

★ Possible lessons from PandoraPFA development

★ For SiD PFA performance – essential to use full track reconstruction