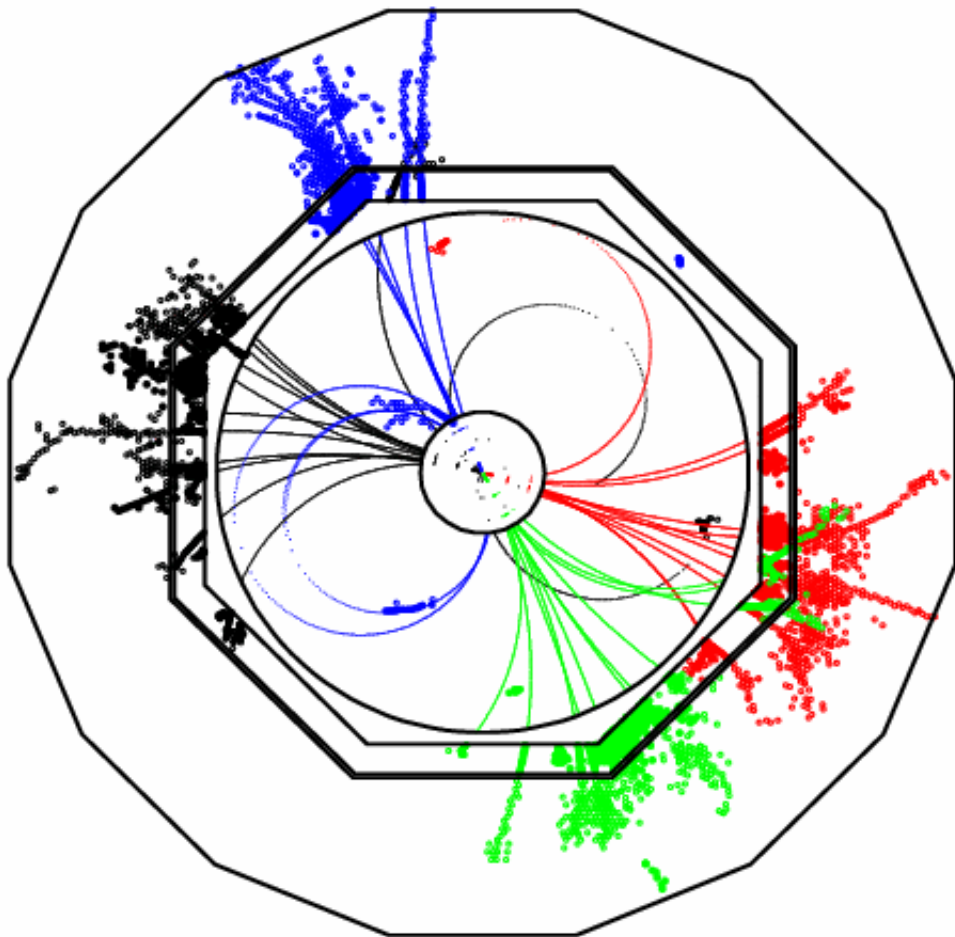


# Status of PandoraPFA

Mark Thomson  
University of Cambridge



## This Talk:

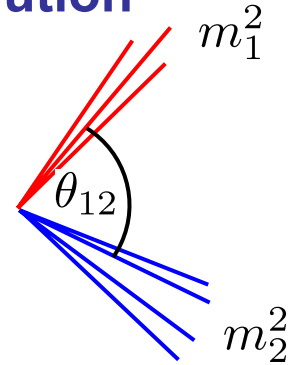
- ① PFA Goals
- ② Status at LCWS07
- ③ Recent Improvements
- ④ Current Performance
- ⑤ Summary
- ⑥ What next ?

# 1 Goals

★ 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  $\Gamma_{W/Z}$

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



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

+ term due to  $\theta_{12}$  uncertainty

★ Assuming a single jet energy resolution of normal form

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



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



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

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

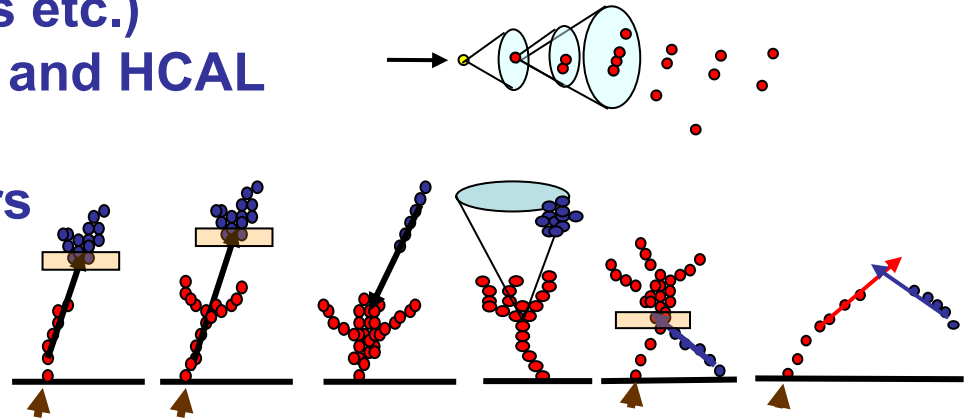
★ Typical di-jet energies at ILC (100-300 GeV)

suggests jet energy resolution goal of  $\sigma_E/E < 0.30/\sqrt{E_{jj}(\text{GeV})}$

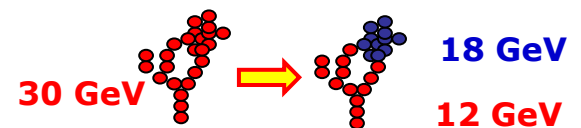
# 2 Status at LCWS07

## Eight Main Stages:

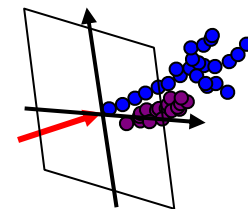
- i. Preparation (isolation cuts etc.)
- ii. Loose clustering in ECAL and HCAL
- iii. Topological linking of clearly associated clusters



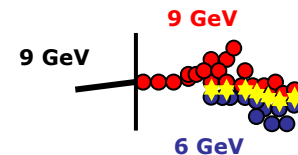
- iv. Coarser grouping of clusters
- v. Iterative reclustering



- vi. Photon Recovery (NEW for LCWS07)



- vii. Fragment Removal (NEW for LCWS07)



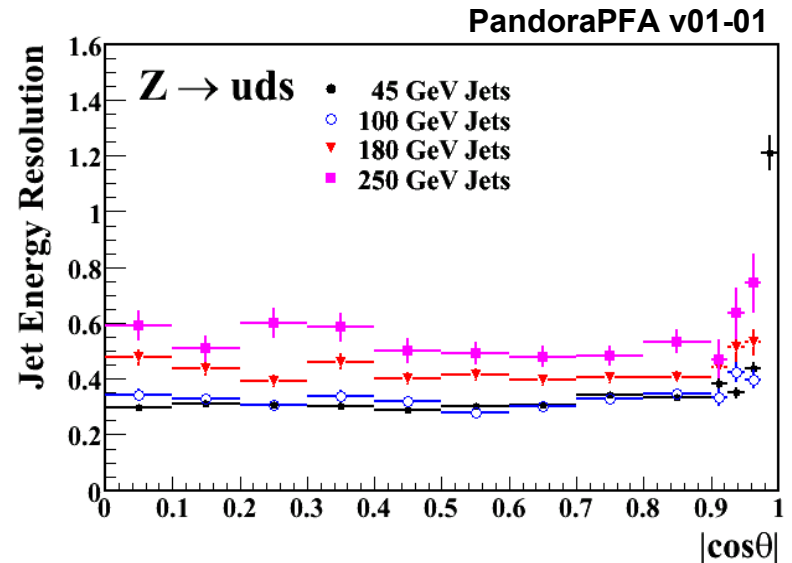
- viii. Formation of final Particle Flow Objects

# LCWS07 performance

- ★ 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\*

$E_{\text{JET}}$	$\sigma_E/E = \alpha/\sqrt{E_{jj}}$ $ \cos\theta  < 0.7$	$\sigma_E/E_j$
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 achieve

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

- ★ Recent emphasis shifted to improving low energy performance **which will be important in likely initial phase of ILC at  $\sqrt{s} \sim 200\text{-}500$  GeV**

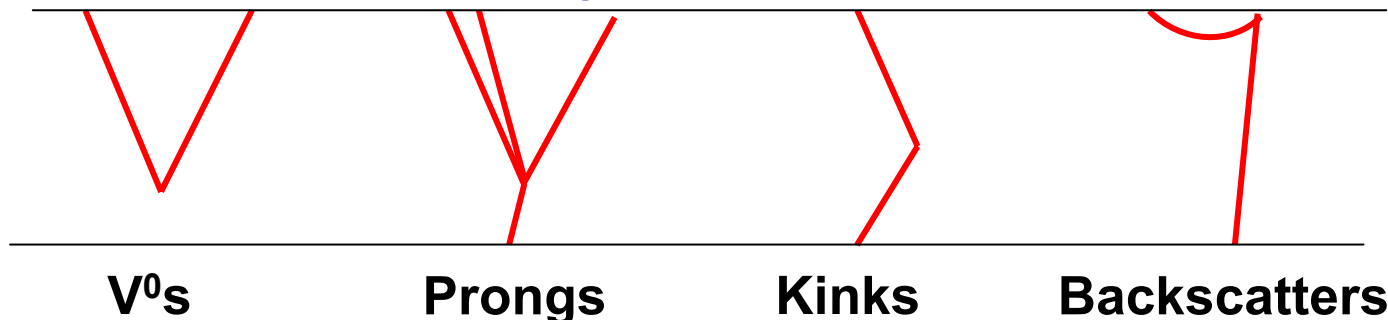
# ③ Recent Improvements

## Overview:

- ★ Technical Improvements
  - ◆ minor bug fixes
  - ◆ reduced memory footprint (~ factor 2) by on-the-fly deleting of temporary clusters, rather than waiting to event end
- ★ Use of tracks (still **TrackCheater**)
- ★ Photon Identification
  - ◆ EM cluster profile identification
- ★ Particle ID
  - ◆ Much improved particle ID : electrons, conversions,  
 $K_S \rightarrow \pi^+ \pi^-$ ,  $\Lambda \rightarrow \pi^- p$  (no impact on PFA)
  - ◆ Some tagging of  $K^\pm \rightarrow \mu^\pm \nu$  and  $\pi^\pm \rightarrow \mu^\pm \nu$  kinks
  - ◆ **No explicit muon ID yet**
- ★ Fragment Removal
- ★ “Calibration” – some interesting issues...

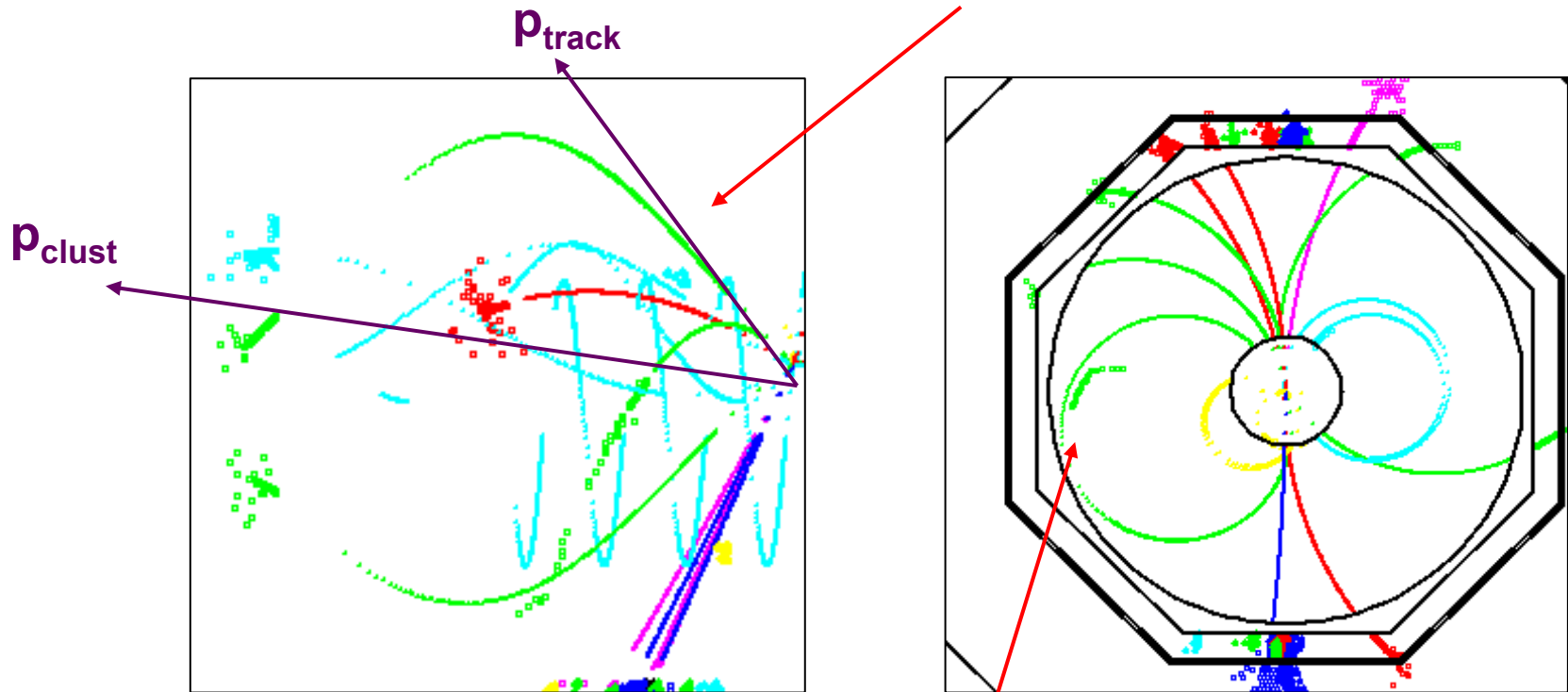
# ① Tracking

- ★ Still based on **TrackCheater** Pattern Recognition with fitted helices for momenta and reconstructed quantities
  - ◆ Will move to **LDCTracking** but want to determine “ideal” performance first
- ★ Previously only used vertex tracks (within 5cm of IP) and kinks +  $V^0$ s
- ★ 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  
(some aspects of code would need modification for Si tracker)
- ★ From cheated tracks identify:



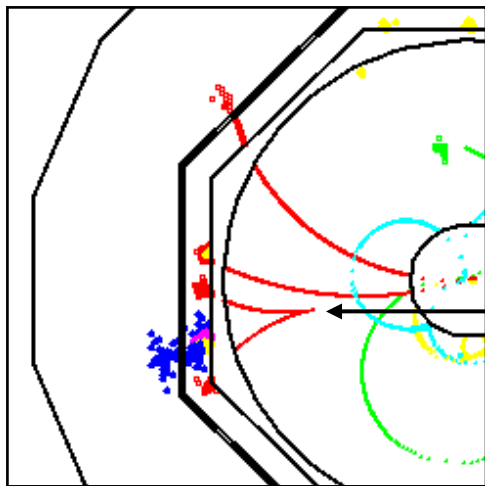
# e.g. Tracking I : extrapolation

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

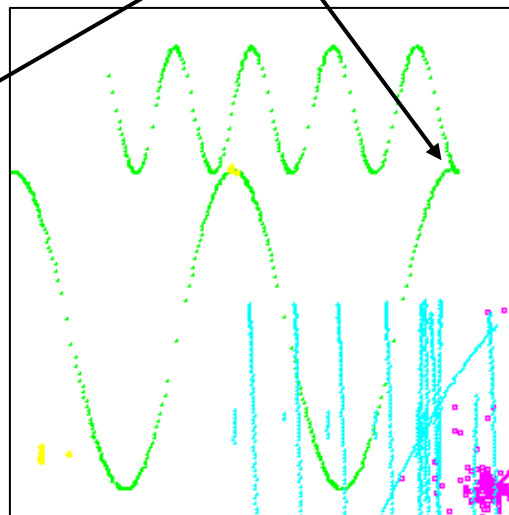
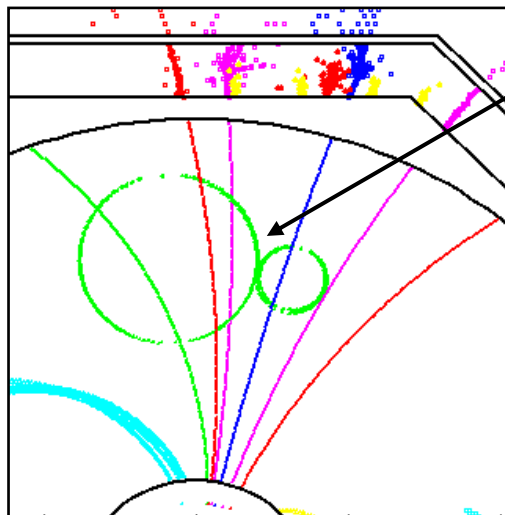


- ★ Now try multiple (successively looser) track-cluster matching requirements e.g. “circle matching”
- ★ As a result, fewer unmatched looping endcap tracks

# e.g. Tracking II : $V^0$ s

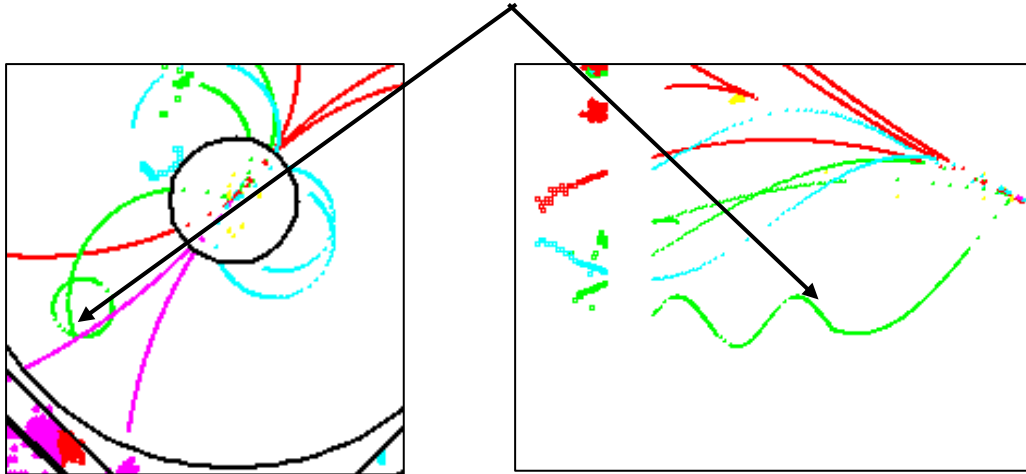


- ★  $V^0$  identification helps PFA as track momentum better measured than cluster energy
- ★ Previously  $V^0$  identification for the main topology
- ★ Now extended to lower  $p_T$  tracks
- ★ Most important for lower energy jets

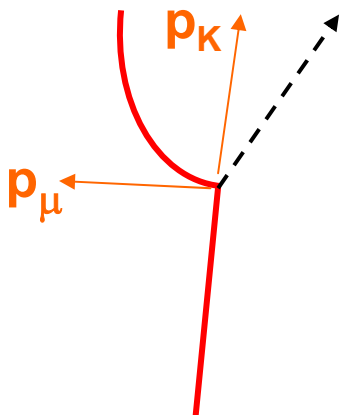


# e.g. Tracking III: Kinks

## ★ Extended Kink finding to “loopers”



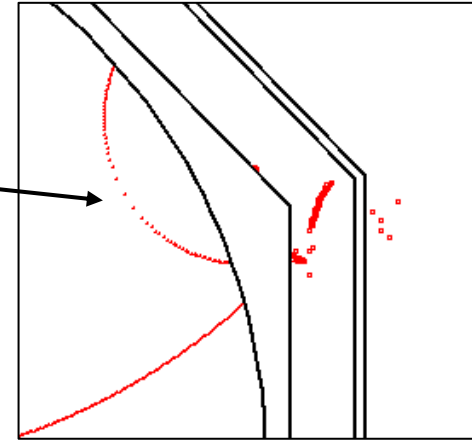
## ★ Improved (but still fairly **crude**) reconstruction missing energy



- ◆ 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.  $\nu$
- ◆ Can then reconstruct primary mass
- ◆ If consistent with hypothesis, e.g.  $m_K$  use primary track for PFO four-momentum

# Tracking: Summary

- ★ Also IDs back-scatters – although not yet used for PFA



- ★ Bottom line – most tracks identified as “something” and used appropriately in PFA
- ★ Mainly affects PFA performance for low energy jets ~1-2% absolute improvement. Will soon quantify this

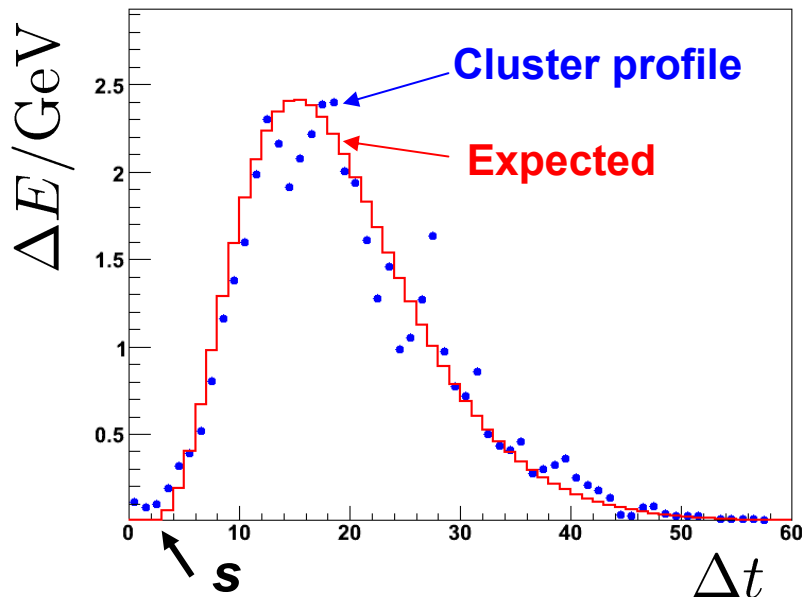
## ② Photon ID

- ★ Previously used simple cut-based photon ID
- ★ This is still used 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  $E_0$

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

$$a = 1.25 + \frac{1}{2} \ln E_0 / E_c$$

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



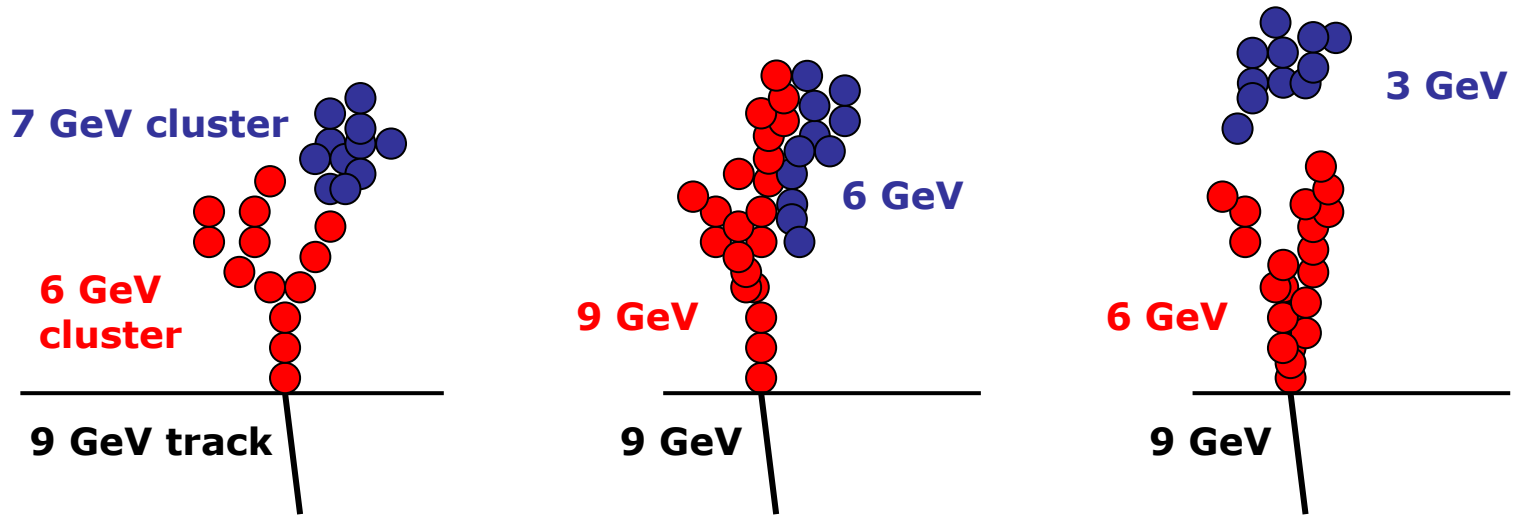
- ◆ Shower Profile fixed by cluster energy
- ◆ But fit for best shower start,  $s$
- ◆ Normalise areas to unity and calc.

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

- ◆ Gives a measure of fractional disagreement in obs/exp profiles
- ◆ Use  $f$  and  $s$  to ID photons
- ◆ Small improvement in PFA perf.

# ③ Fragment Removal

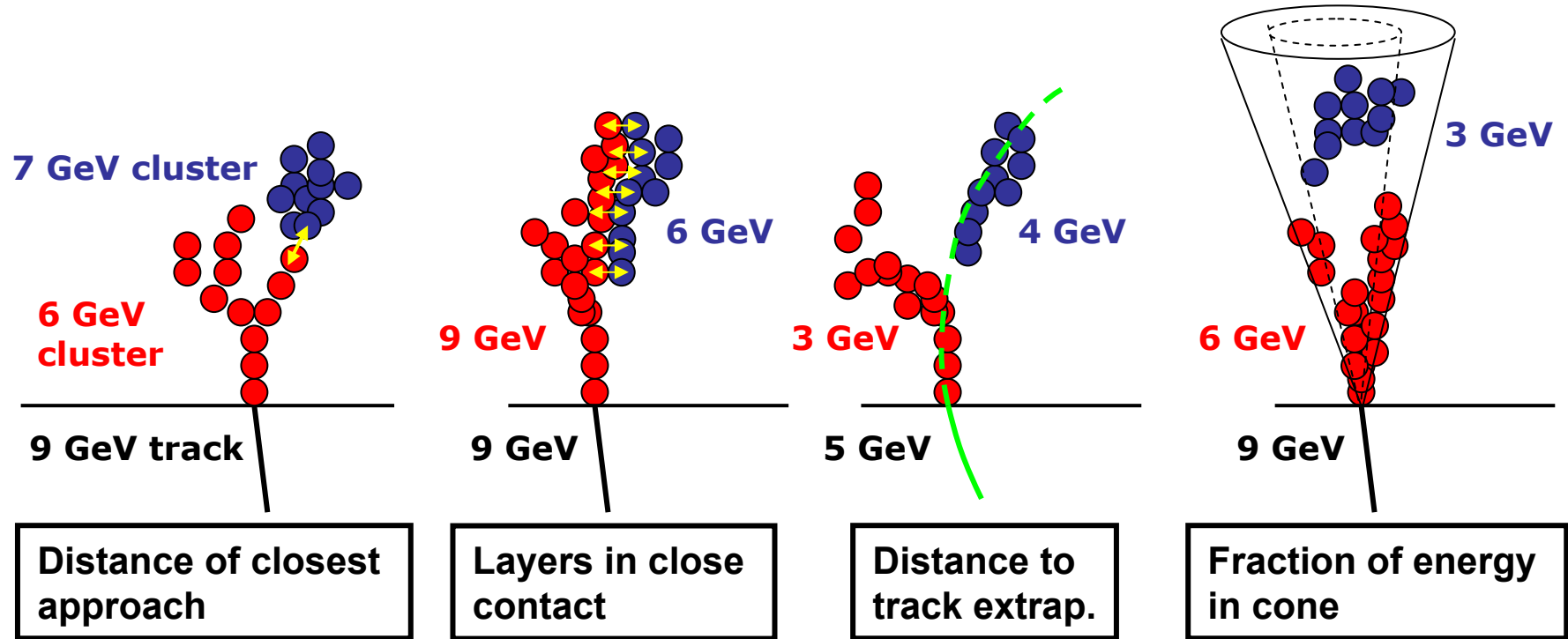
- ★ One of the final stages of PandoraPFA is to identify “neutral fragments” from charged particle clusters



- ★ Previously the code to do this was “a bit of a mess”
- ★ This has been significantly improved – but not yet optimised

# 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 (slight improvement wrt. previous)

## ④ “Calibration”

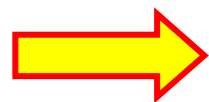
- ★ Current benchmark studies with LDC00Sc model
- ★ Wenbiao (see previous talk) pointed out non-linearity in calibration, i.e. energy dependence in  $E_{\text{PFA}}/E_{\text{TRUE}}$  for single  $K_L$ ,  $n$ , ...
- ★ Tracked down the origin of this effect to isolation cuts
- ★ PandoraPFA throws away isolated hits, i.e. hits which are not clustered and are  $> 10$  cm from a hit in an existing cluster

e.g. single 10 GeV  $K_L$

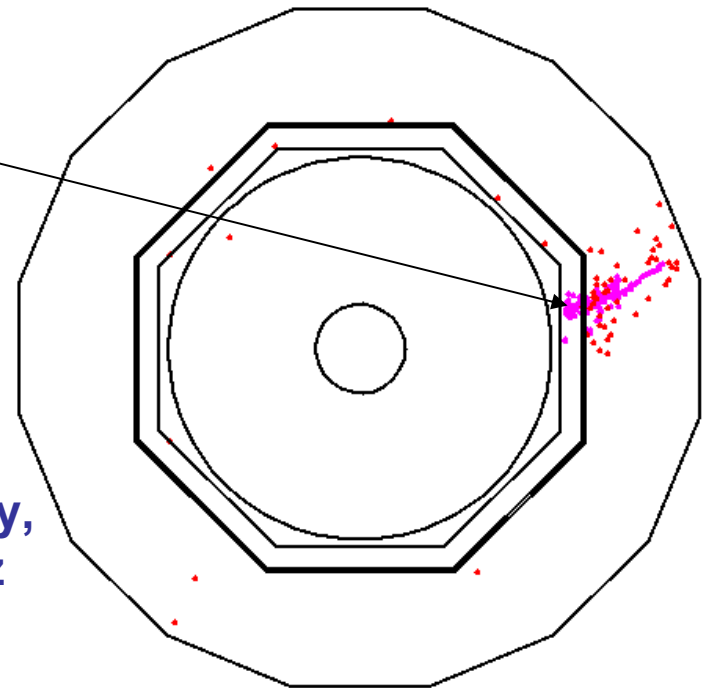
Clustered hits = magenta

Rejected hits = red

- At some level need to reject isolated hits for PFA as in dense environment not clear to which cluster they are associated
- The larger the cluster, i.e. higher energy, the more chance of picking up the fuzz

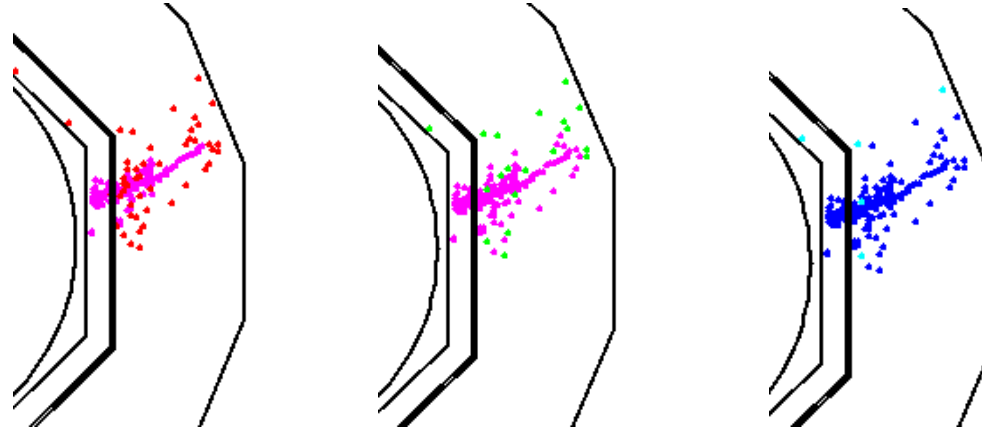


Non-linear response !



# “Calibration” cont.

- ★ Effect depends on cluster energy and isolation cut



Isolation cut: 10cm

25cm

50cm

**Fraction of energy rejected as isolated**

	5 GeV $K_L$	10 GeV $K_L$	20 GeV $K_L$
10 cm	16.1 %	12.7 %	6.7 %
25 cm	8.1 %	6.1 %	2.8 %
50 cm	3.6 %	2.7 %	1.1 %

$\Delta = 10 \%$

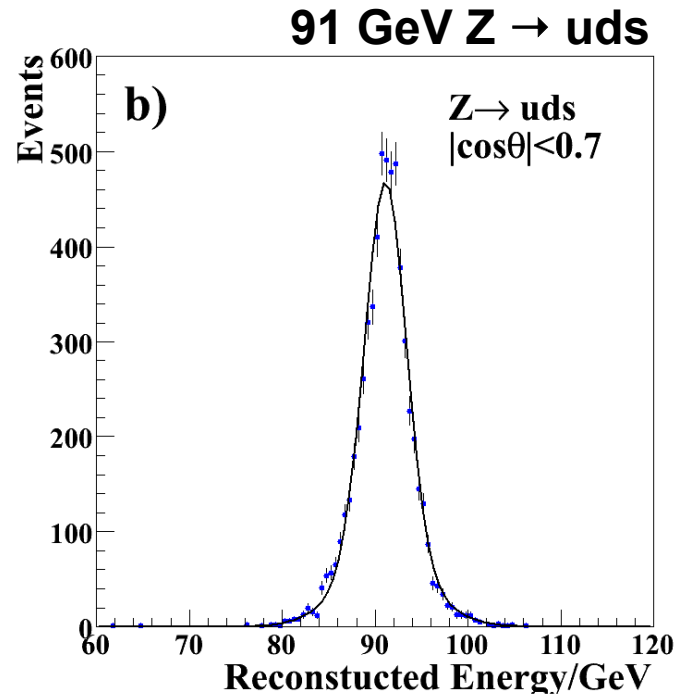
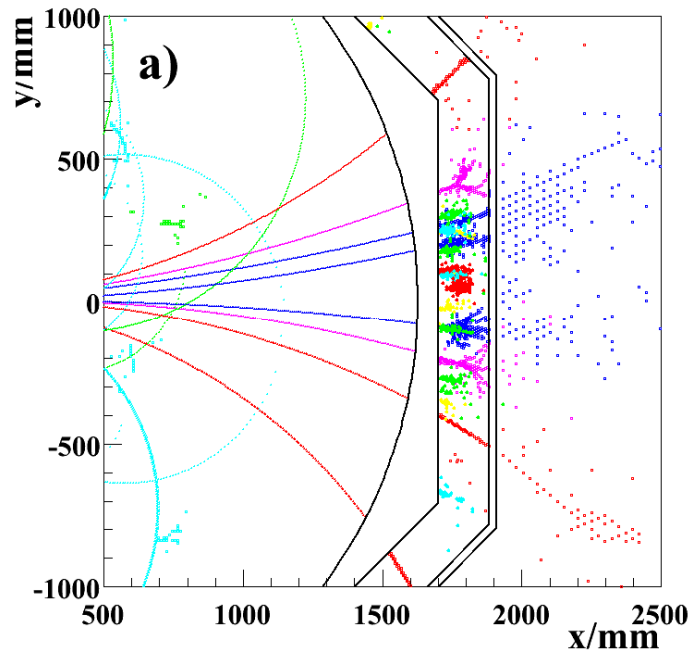
$\Delta = 5 \%$

$\Delta = 2.5 \%$

- ★ Non linearity degrades PFA performance
- ★ For now increase isolation cut to 25 cm (small improvement for PFA)
- ★ Best approach ?

# 4 Current Performance

- ★ As always use barrel  $Z \rightarrow uds$  (Mokka 6.1)
- ★ LDC00Sc model with increased 63 layer HCAL
- ★ NOTE: previously used older 91 GeV samples with 40 layer HCAL
- ★ Significant improvement in performance for low energy jets !



# Current Performance cont.

Caveat : work in progress, things will change

PandoraPFA v01-01

$E_{\text{JET}}$	$\sigma_E/E = \alpha/\sqrt{E_{\text{jj}}}$ $ \cos\theta  < 0.7$	$\sigma_E/E_j$
45 GeV	0.295	4.4 %
100 GeV	0.305	3.0 %
180 GeV	0.418	3.1 %
250 GeV	0.534	3.4 %

PandoraPFA v02- $\alpha$

$E_{\text{JET}}$	$\sigma_E/E = \alpha/\sqrt{E_{\text{jj}}}$ $ \cos\theta  < 0.7$	$\sigma_E/E_j$
45 GeV	0.227	3.4 %
100 GeV	0.287	2.9 %
180 GeV	0.395	2.9 %
250 GeV	0.532	3.4 %

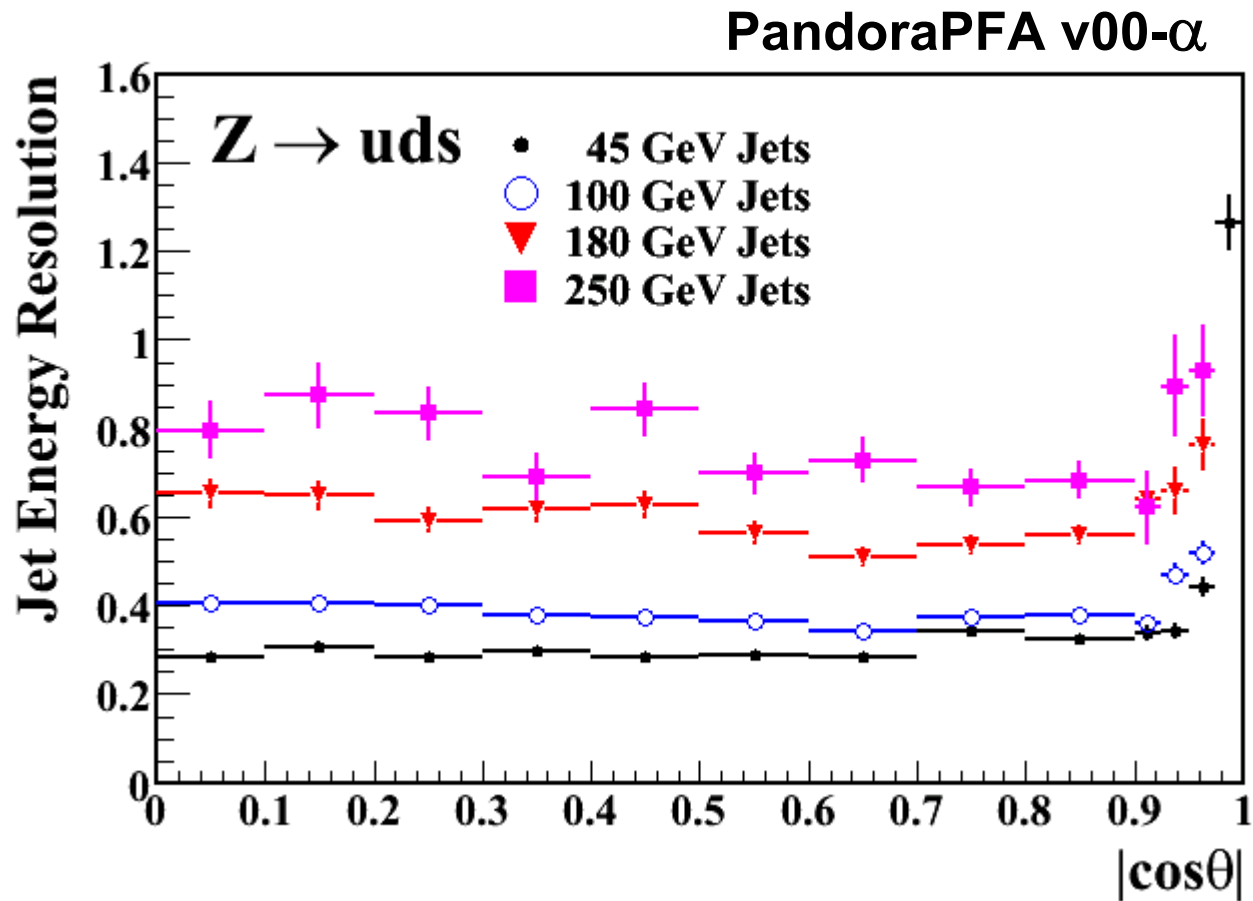
★ For 45 GeV jets, performance now equivalent to

$$23 \% / \sqrt{E}$$

★ For TESLA TDR detector “sweet spot” at just the right place  
100-200 GeV jets !

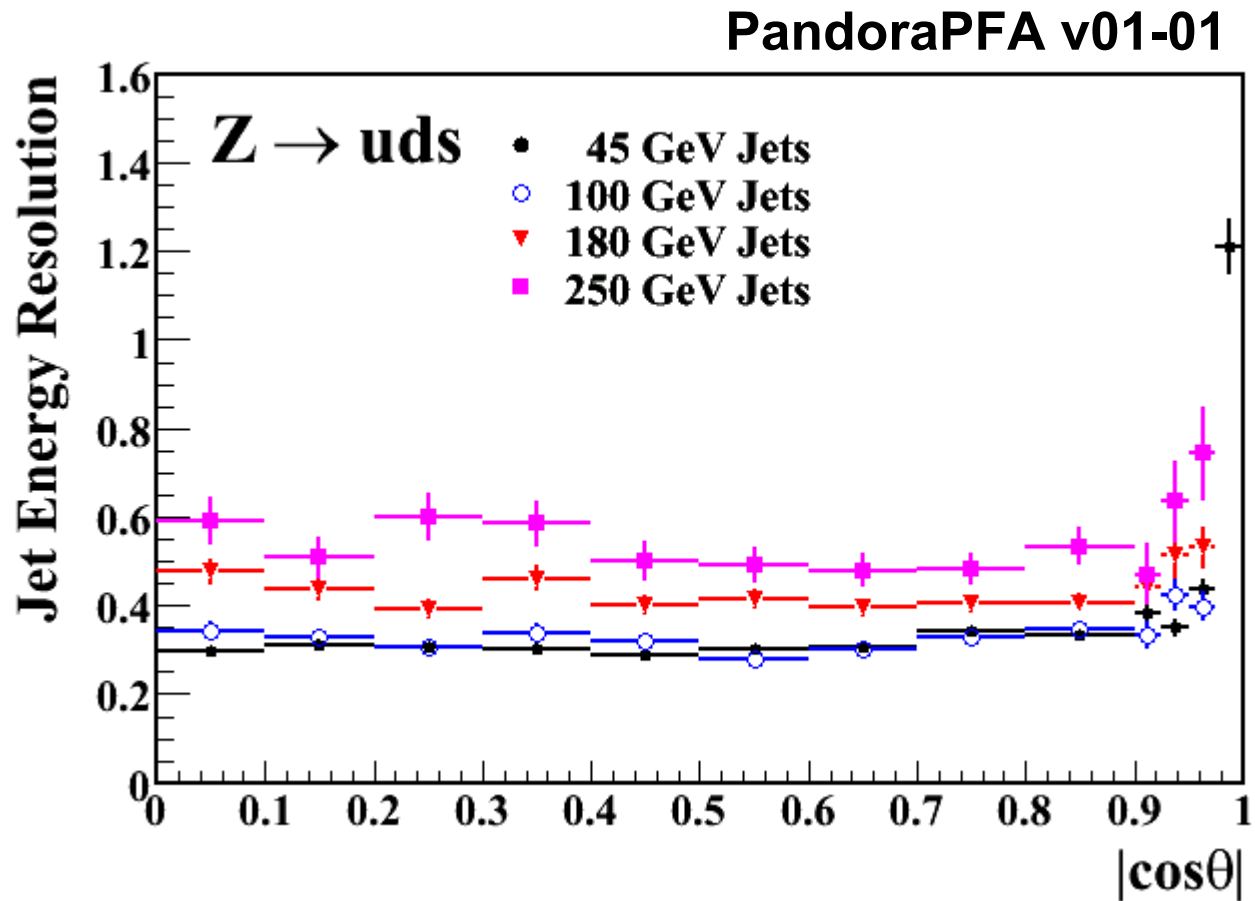
★ However, only modest improvements at higher energy...

# Evolution



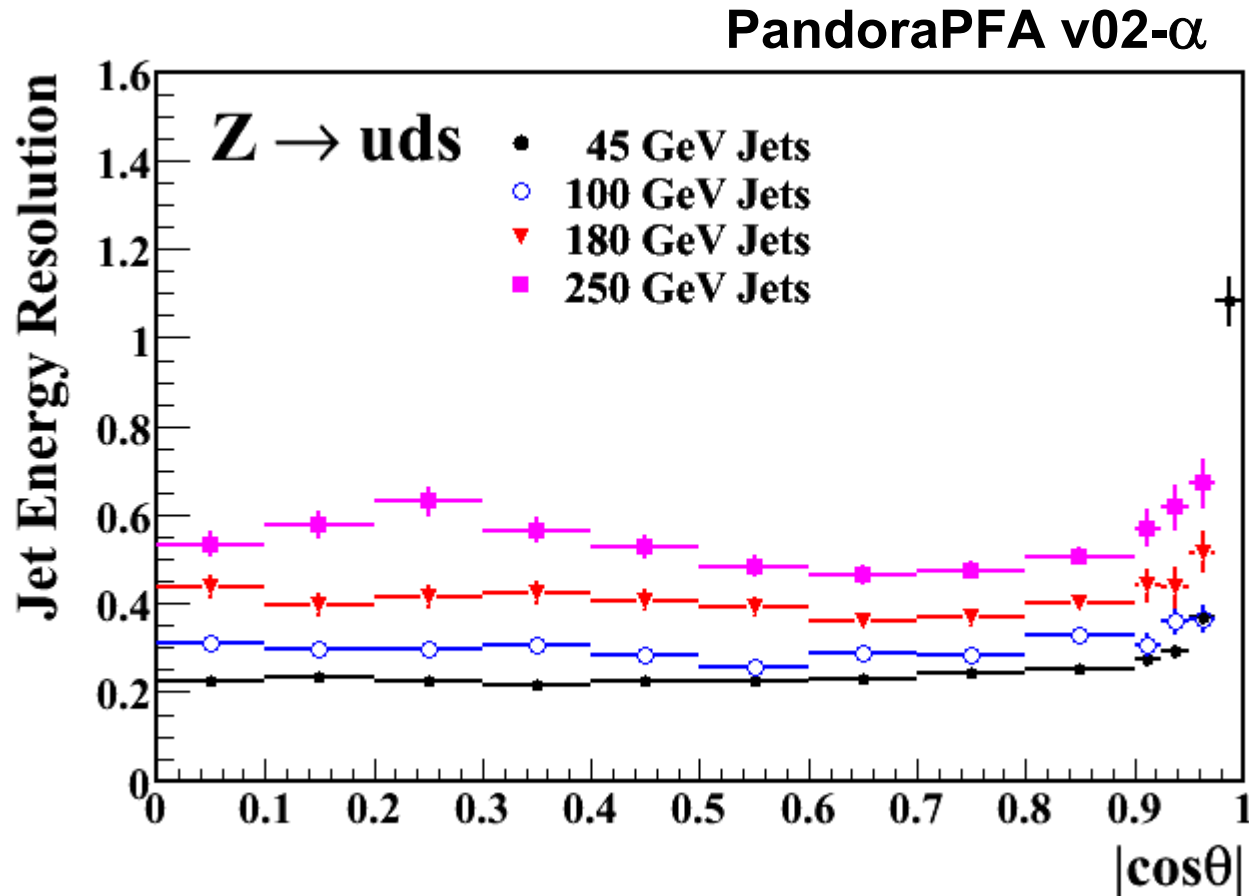
09/2006

# Evolution



06/2007

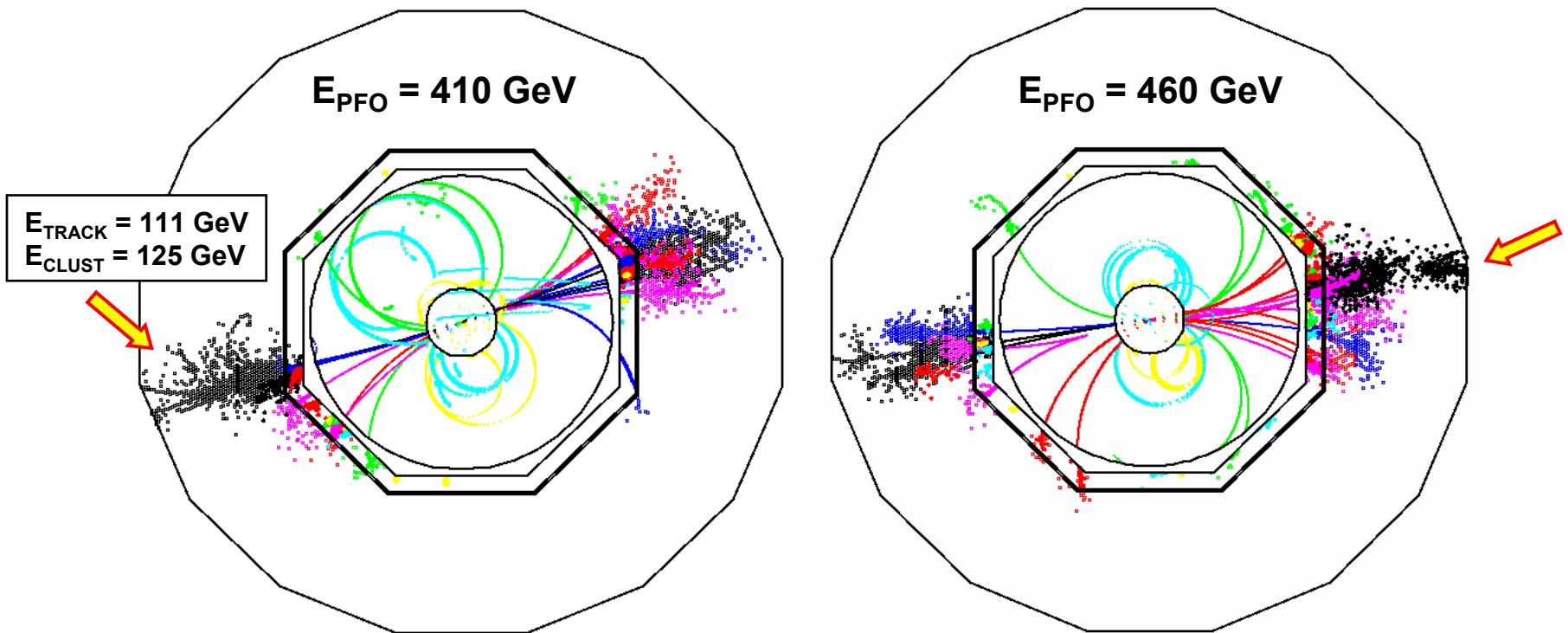
# Evolution



09/2007

# High Energy Performance

- ★ High energy performance stuck at  $53\%/\sqrt{E}$  – why ?
- ★ Don't fully understand – but there are some ideas
  - ◆ For 250 GeV jets hadronic showers are large and overlap, different clustering algorithm may help (in reclustering)
  - ◆ Leakage becomes (very?) important



- ★ Plenty of room for improvement – but not trivial

# 5 Summary

## Summary:

- ★ Concentrated on lower energy performance – major improvements !
- ★ Also improvements in structure of code
  - + almost certainly some new



- ★ Some small improvements for higher energy jets

## Perspective:

- ★ Development of high performance PFA is **highly** non-trivial
- ★ User feedback **very** helpful (thanks Wenbiao)
- ★ Major improvements on current performance possible
  - “just” needs effort + fresh ideas
- ★ PandoraPFA needs a spring-clean (a lot of now redundant code)
  - + plenty of scope for speed improvements
    - again needs new effort (I just don’t have time)

# ⑥ What Next

## Plans:

- ★ Optimisation of new code
  - ◆ Slow procedure... takes about 6 CPU-days per variation
  - ◆ Only small improvements expected – have found that the performance is relatively insensitive to fine details of alg.
- ★ More study of non-linear response due to isolation
  - Will look at **RPC** HCAL
- ★ Detailed study of importance of different aspects of PFA, e.g. what happens if kink finding is switched off...
- ★ Revisit high energy performance
- ★ Update code to use **LDCTracking**
- ★ Release version 02-00 on timescale of 1-2 months.