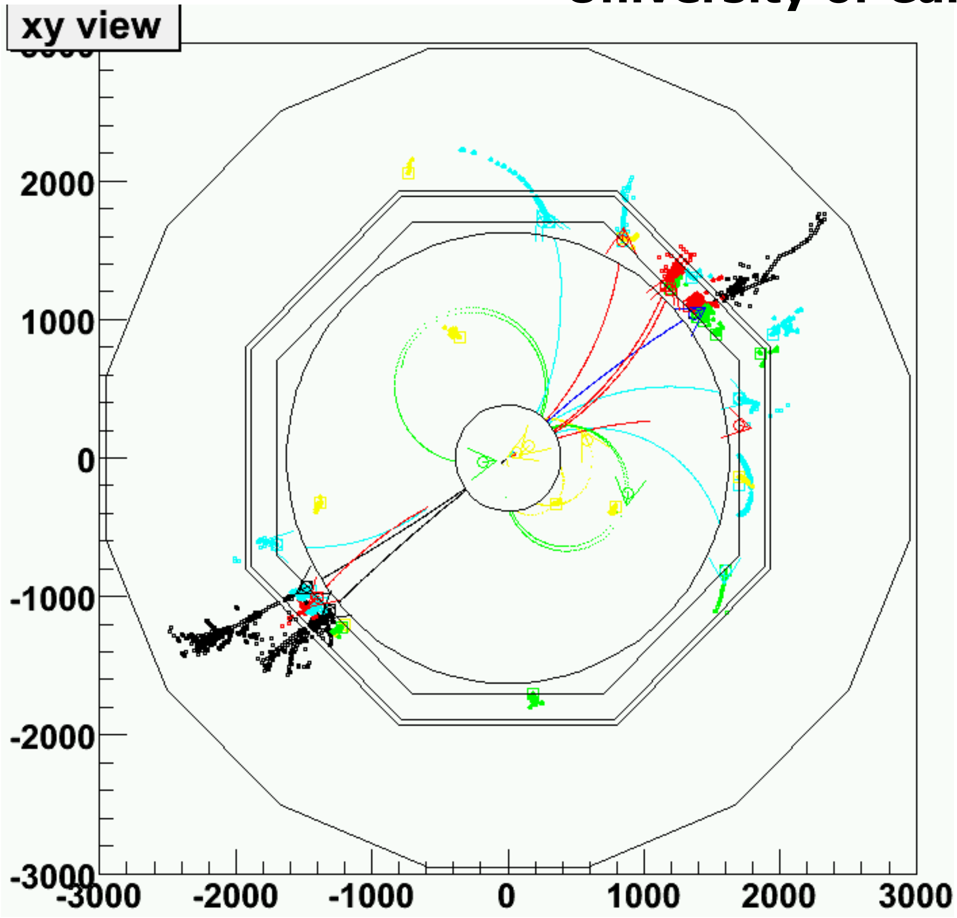
A Topologic Approach to Particle Flow "PandoraPFA"

Mark Thomson
University of Cambridge

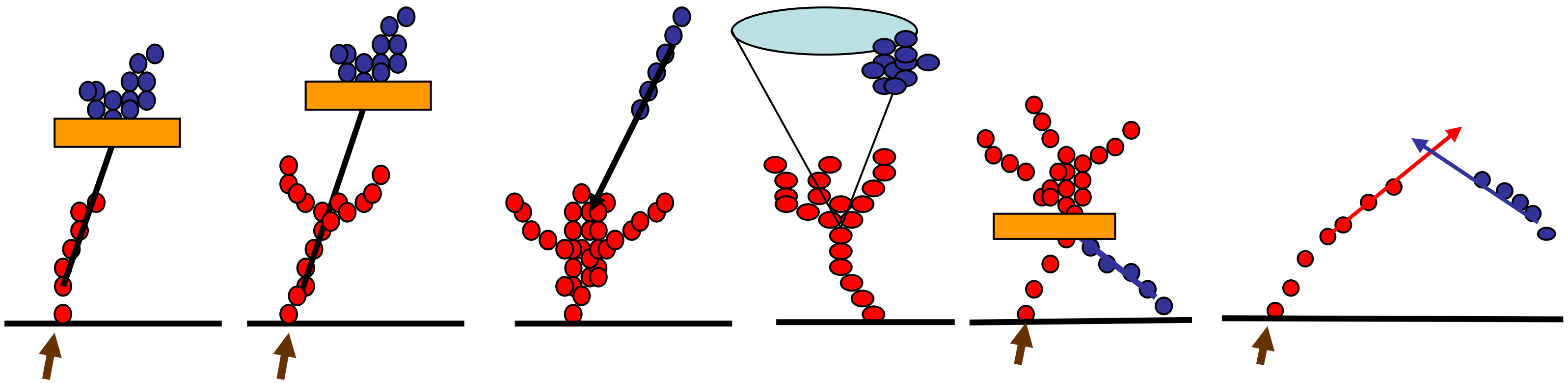


This Talk:

❶ Philosophy
❷ The Algorithm
❸ Some New Results
❹ Confusion
❺ Conclusions
❻ Outlook

❶ Philosophy

- ★ Work from the premise that PFA is not a pure ECAL/HCAL clustering problem
- ★ PFA and calorimeter clustering performed together
- ★ Start by applying loose clustering
- ★ Then join clusters using topology



★Algorithm defined by loose cluster + topological rules

Goals/Framework

★ Try to develop "generic" PFA which will take advantage of a high/very high granularity ECAL
★ Clustering and PFA performed in a single algorithm
★ Aim for fairly generic algorithm:
  - very few hard coded numbers
  - use GEAR to get basic geometry
★ Clustering uses tracking information
★ Initial clustering is fairly loose → ProtoClusters
★ Topological linking of ProtoCluster

<u>Runs in MARLIN framework using:</u>

- Marlin SimpleDigitisation
- Track finding/fitting : TrackCheater
- PFA Utility classes, e.g. Helix class for track extrap. (Alexei R.)

❷ The Algorithm

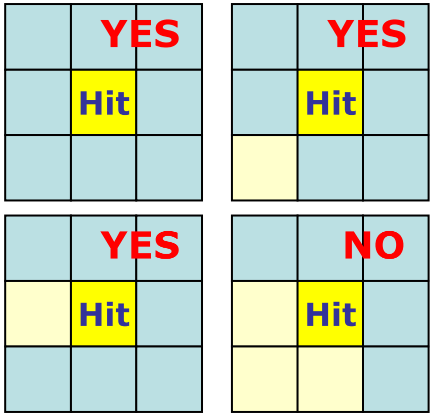
Overview:

- ★ **Preparation**
  - ★ Isolation cuts, hit ordering, track quality
- ★ **Initial clustering to form ProtoClusters**
  - ★ ProtoClusters are heavyweight objects:
    - ★ much more than a collection of hits
    - ★ know how to grow (configured when created)
    - ★ information about shape, direction, isPhoton,…
    - ★ can be configured to fragment tracks…
    - ★ +much more (not all used)…
- ★ **Cluster association/merging**
  - ★ Tight Topological linking of clusters
  - ★ Looser merging of clusters
  - ★ Track-driven merging
- ★ **PFA**
  - ★ Final track-cluster matching

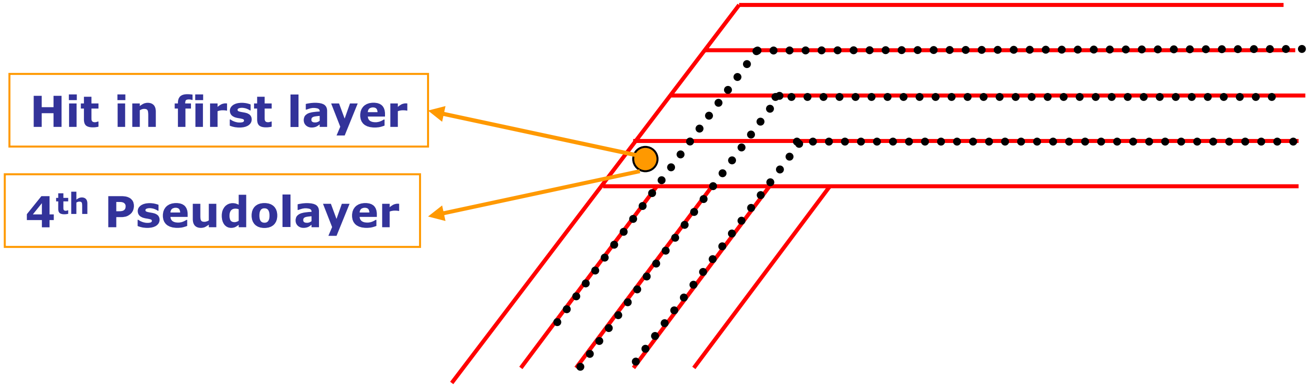
- The first part of this talk gives flavour of what's done in each stage skipping details

Preparation I: Extended Hits

- Create internal **ExtendedCaloHits** from **CaloHits**
- **ExtendedCaloHits** contain extra info:
  - pointer to original hit
  - pseudoLayer (see below)
  - measure of isolation for other hits
  - is it MIP like (to ID "tracklike objects")
  - actual layer (decoded from **CellID**)
  - Pixel Size (from GEAR)
- hits are now self describing

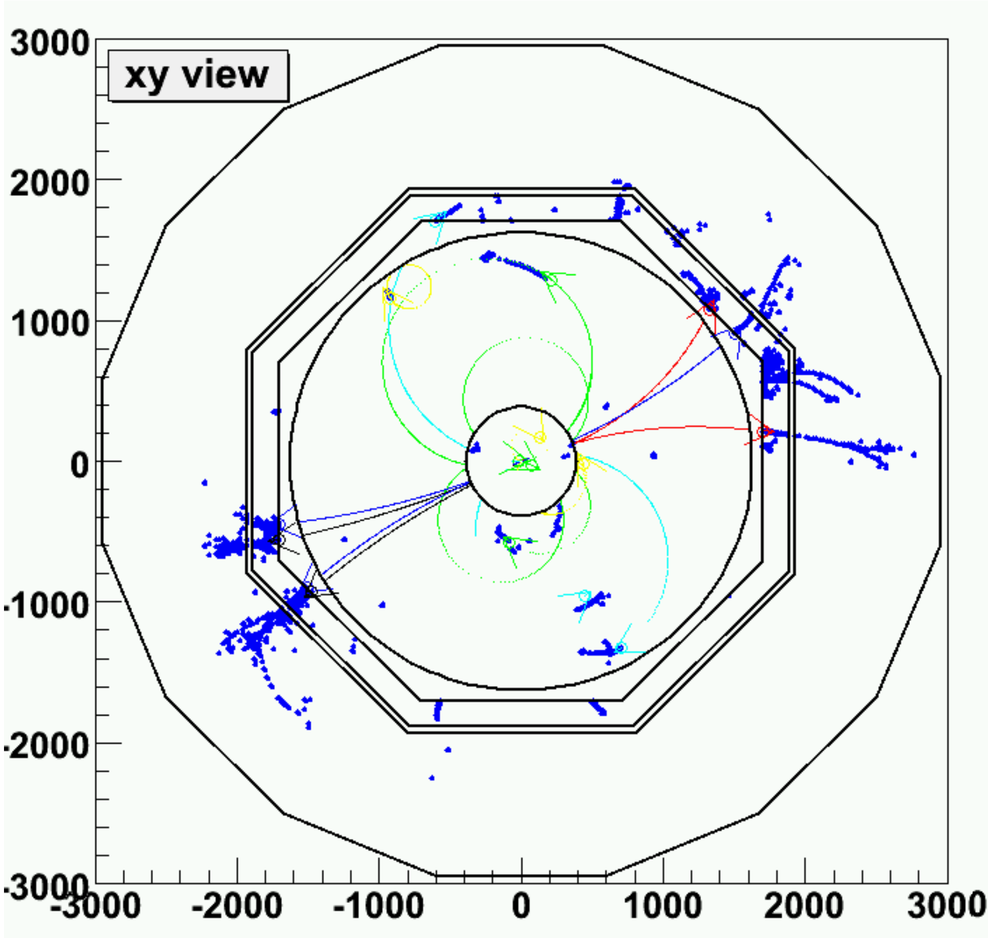


- Arrange hits into **PSEUDOLAYERS (e.g. Chris Ainsley's MAGIC)**
  - **i.e.** order hits in increasing depth within calorimeter
  - PseudoLayers follow detector geometry



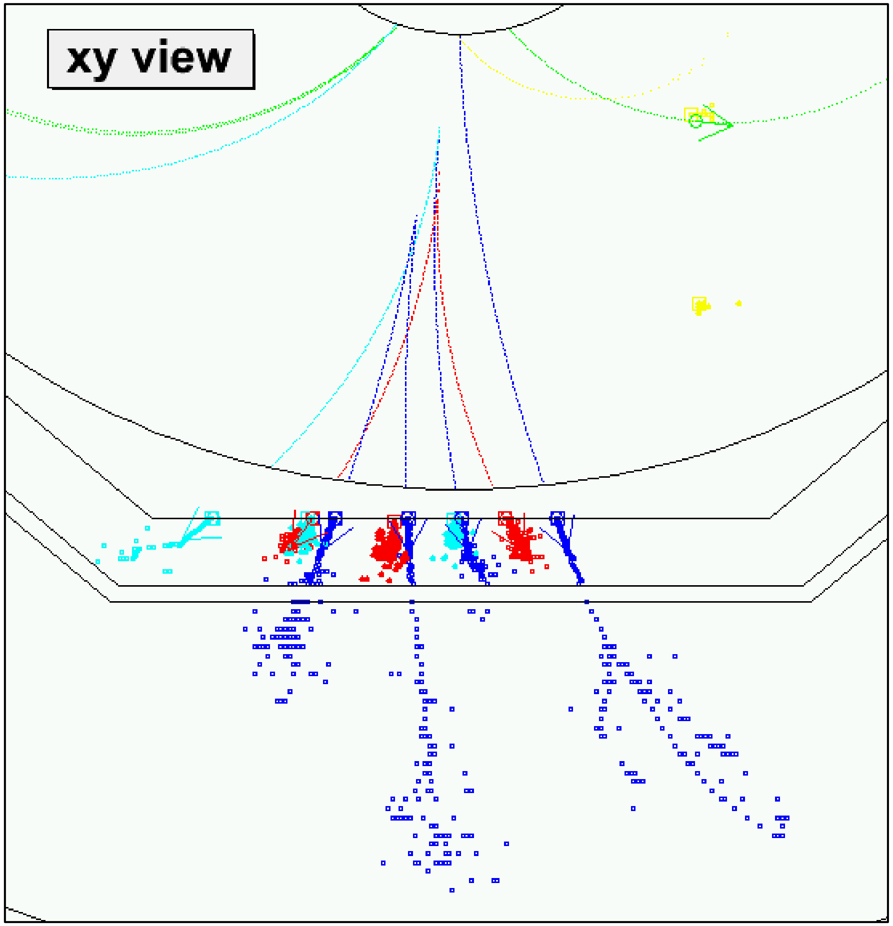
Preparation II: Isolation

- ★ Divide hits into isolated and non-isolated
- ★ Only cluster **non-isolated** hits
- ★ "Cleaner"/Faster clustering
- ★ Significant effect for scintillator HCAL



- ✦ Removal of isolated hits degrades HCAL resolution
- ✦ <u>e.g. D10scint:</u>
  50 %/√E/GeV →
  60 %/√E/GeV

Preparation III: Tracking



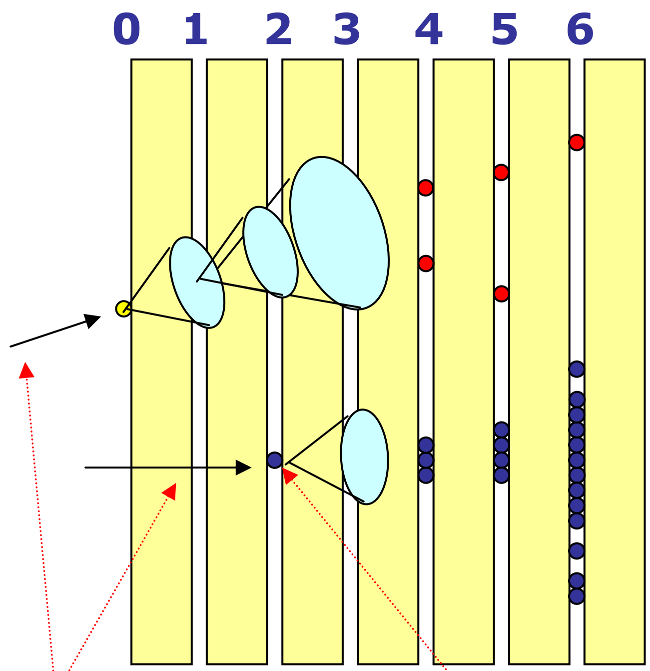
- ★ Use MARLIN **TrackCheater**
- ★ Tracks formed from MC Hits in **TPC/FTD/VTX**
- ★ HelixFit (Alexei R) ⇒ track params
- ★ Cuts (primary tracks):
  - ♦ $|d_0| < 5$ mm
  - ♦ $|z_0| < 5$ mm
  - ♦ >4 non-Si hits

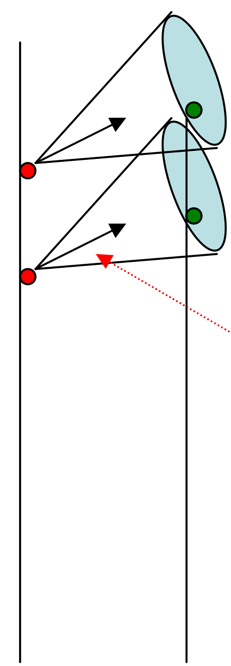
+ V_0 and Kink finding:

- ✦ Track resolution better than cluster
- ✦ Improves PFA performance by ~2 %

PandoraPFA Clustering II

- Start at inner layers and work outward
- Associate Hits with existing Clusters
- If multiple clusters "want" hit then Arbitrate
- Step back **N** layers until associated
- Then try to associate with hits in current layer (M pixel cut)
- If no association made form new Cluster
- + tracks used to seed clusters





Simple cone algorithm based on current direction + additional N pixels

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

Initial cluster direction

Unmatched hits seeds new cluster

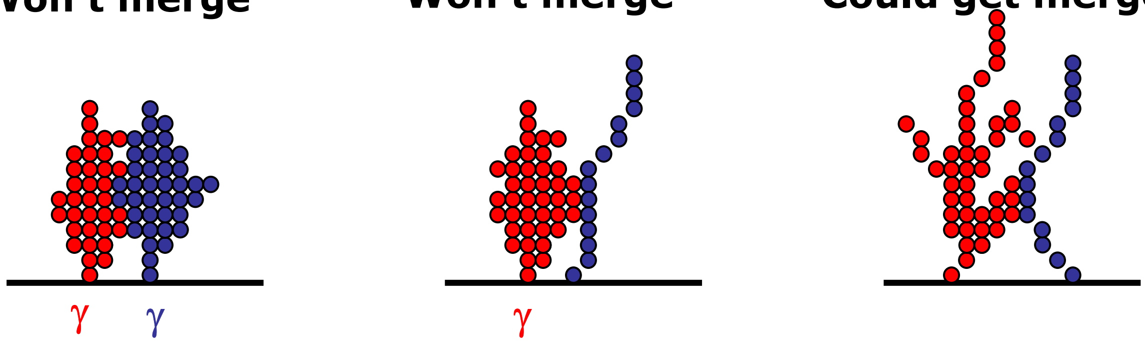
Cluster Association

- 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 – catches what's been missed but rather crude

Photon ID

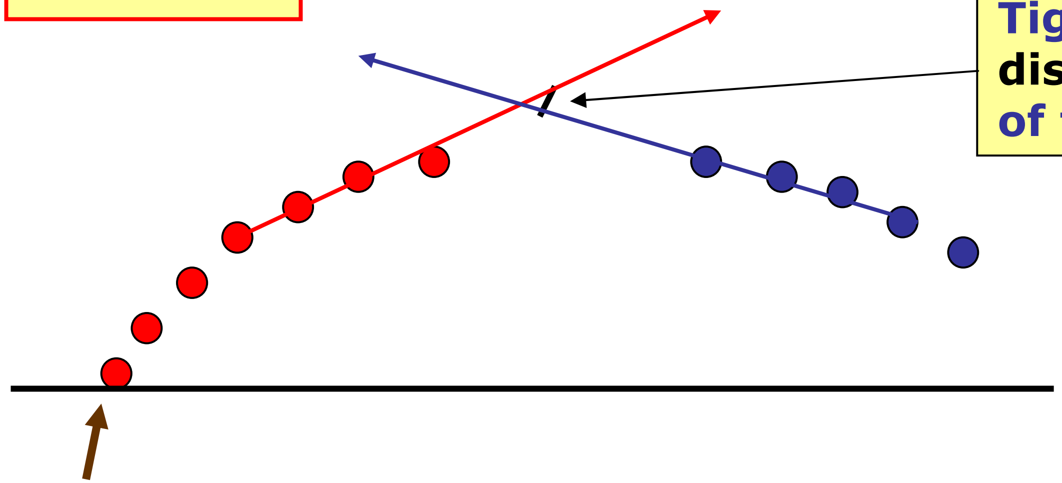
- Photon ID plays important role
- Simple "cut-based" photon ID applied to all clusters
- Clusters tagged as photons are immune from association procedure – just left alone

Won't merge **Won't merge** **Could get merged**



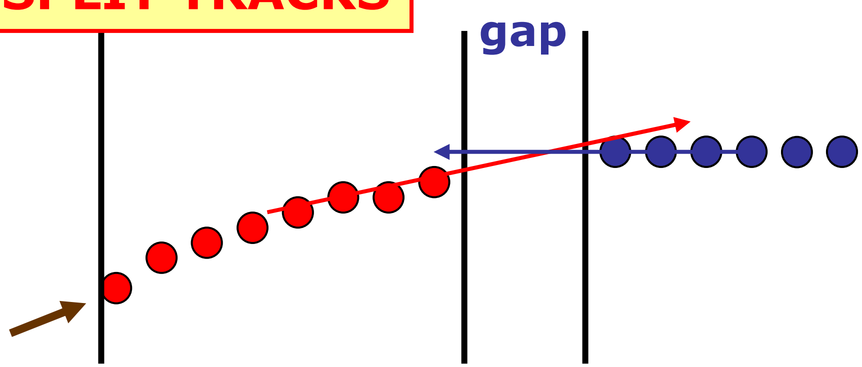
Cluster Association I : track merging

LOOPERS



Tight cut on extrapolation of **distance of closest approach** of fits to **ends** of tracks

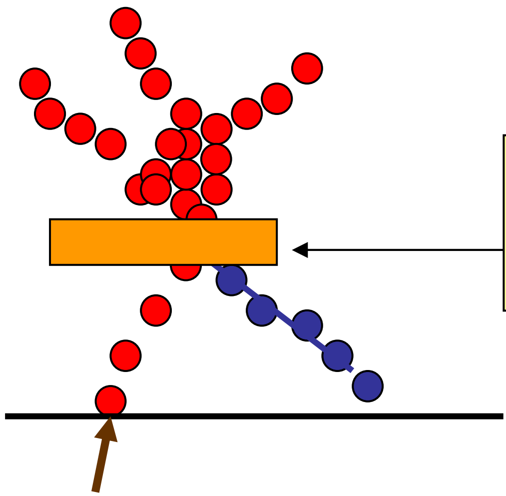
SPLIT TRACKS



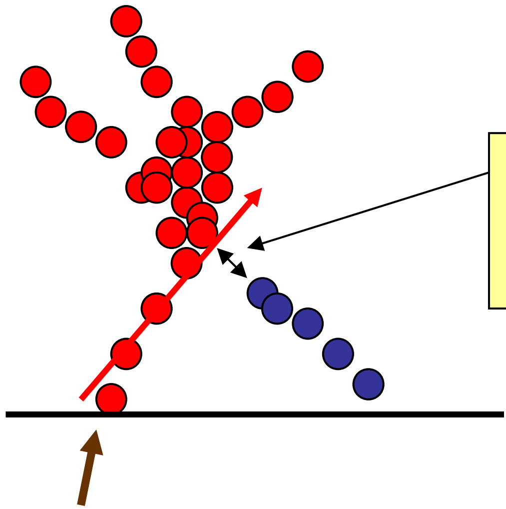
Tight cut on extrapolation of **distance of closest approach** of fits to **end of inner tracks** and **start of outer track**

Cluster Association II : Backscatters

- Forward propagation clustering algorithm has a major drawback: back scattered particles form separate clusters



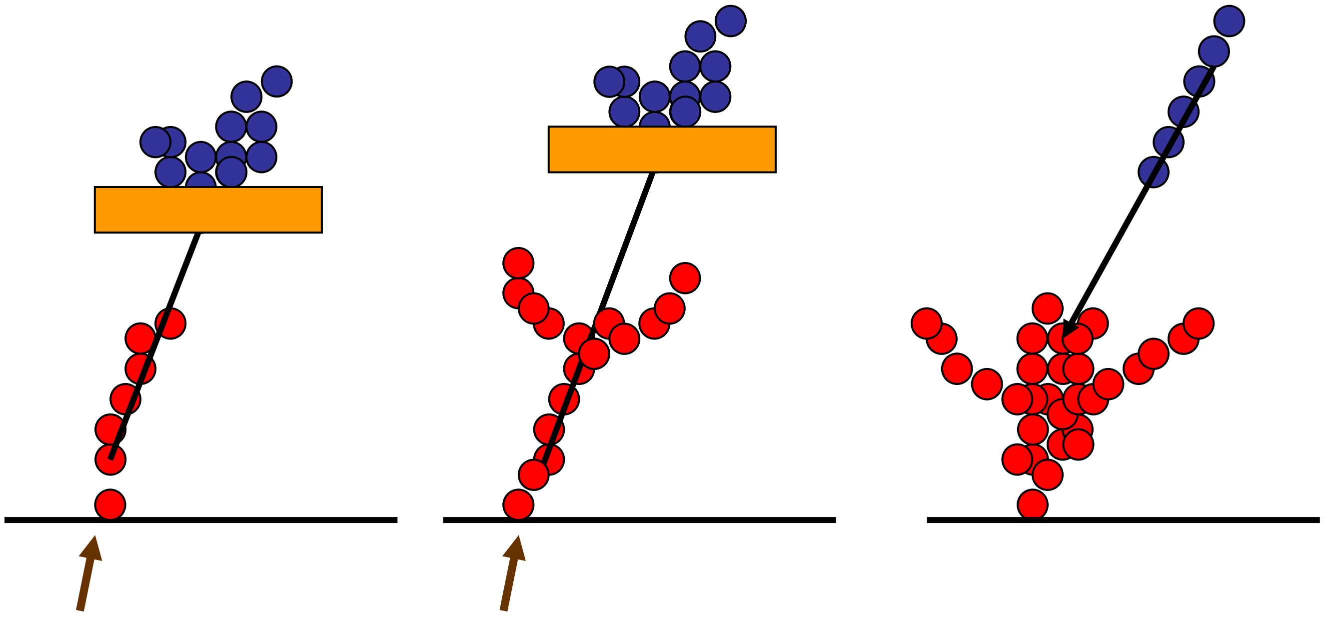
Project track-like clusters forward and check distance to shower centroids in subsequent N layers



Also look for track-like segments at start of cluster and try to match to end of another cluster

Cluster association III : MIP segments

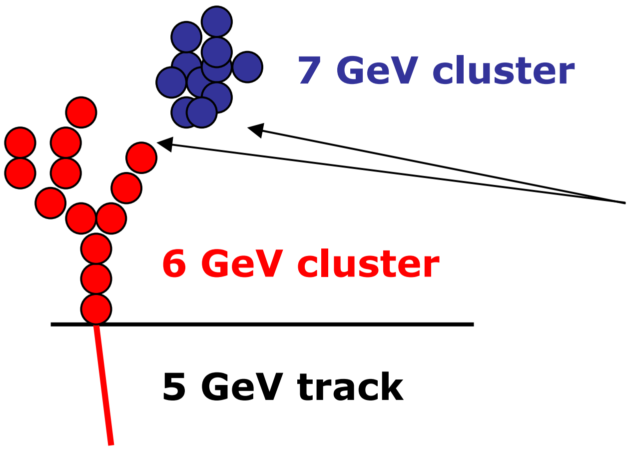
★ **Look at clusters which are consistent with having tracks segments and project backwards/forward**



★ **Apply tight matching criteria on basis of projected track [NB: + track quality i.e. chi2]**

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.

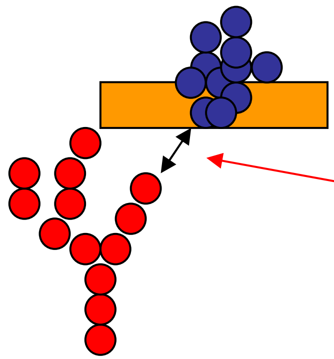


This cluster association would be forbidden if $|E_1 + E_2 - p| > 3\,\sigma_E$

Provides some protection against silly mistakes

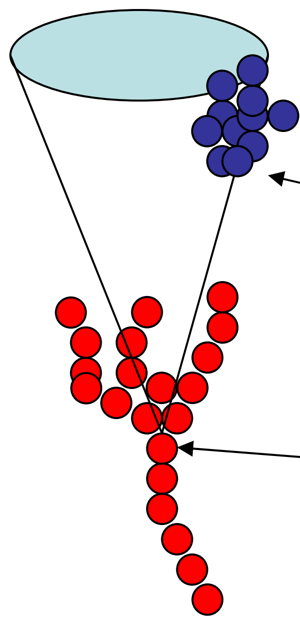
Sledgehammer Cluster Association

Proximity



Distance between hits -limited to first layers

Shower Cone



Associated if fraction of hits in cone > some value

Shower start identified

+Track-Driven Shower Cone

Apply looser cuts if have low E cluster associated to high E track

❸ Current Results

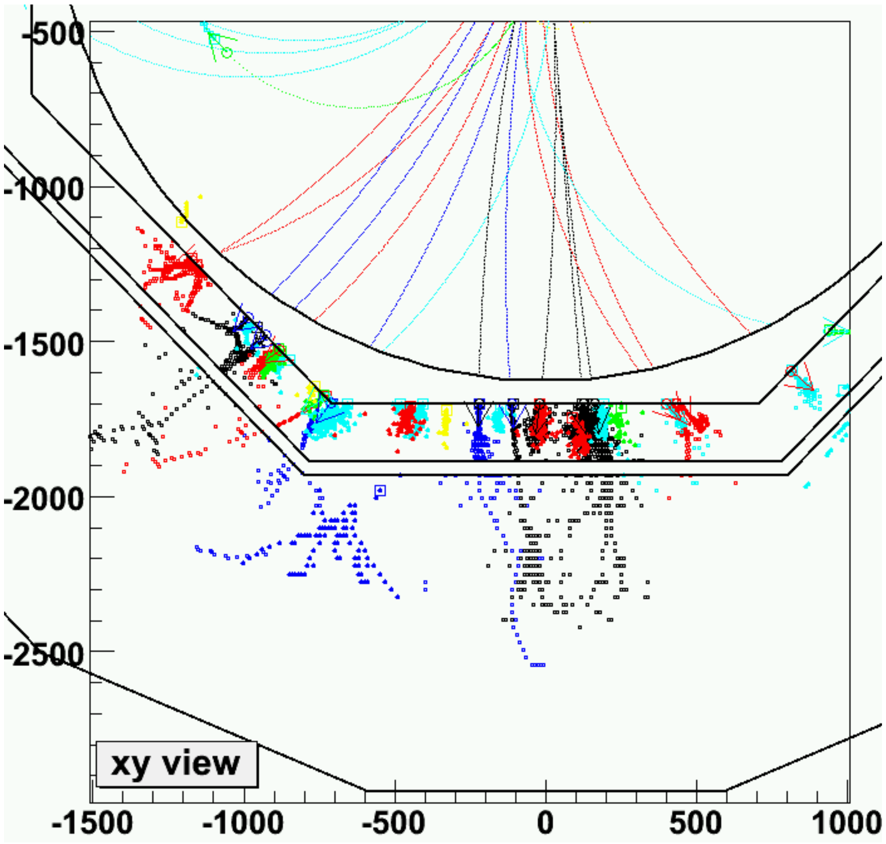
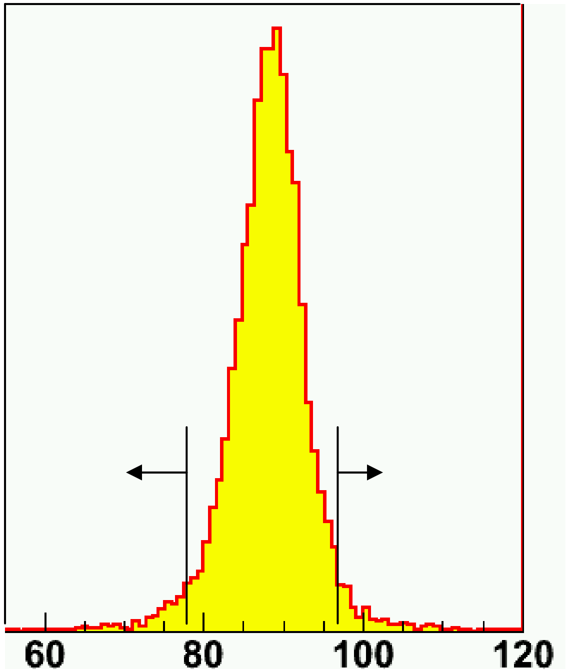


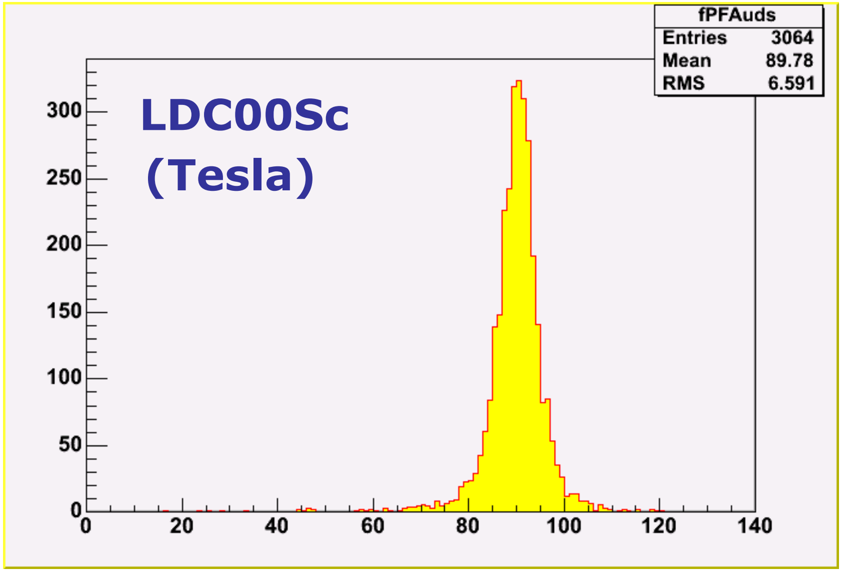
Figure of Merit:

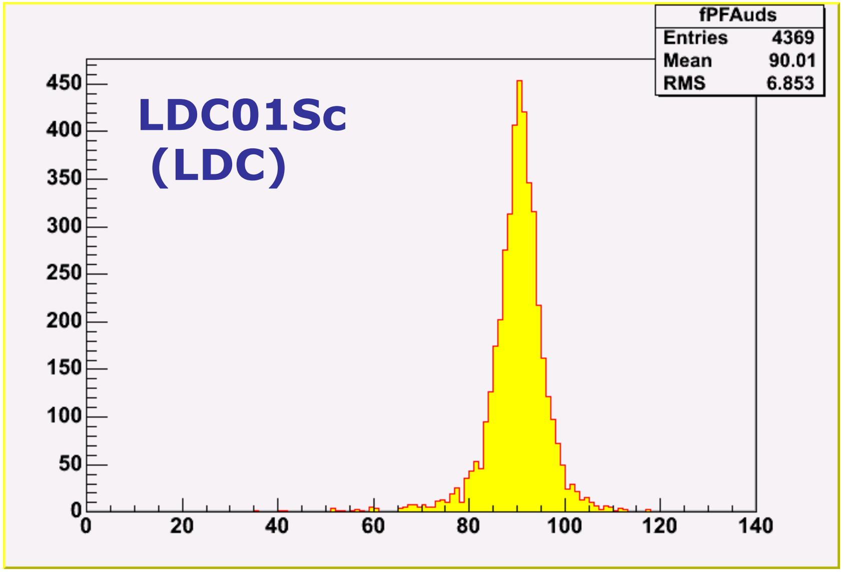


- ★**Find smallest region containing 90 % of events**
- ★**Determine rms in this region**

More robust than fitting double Gaussian

Preliminary Results : Z →uds events



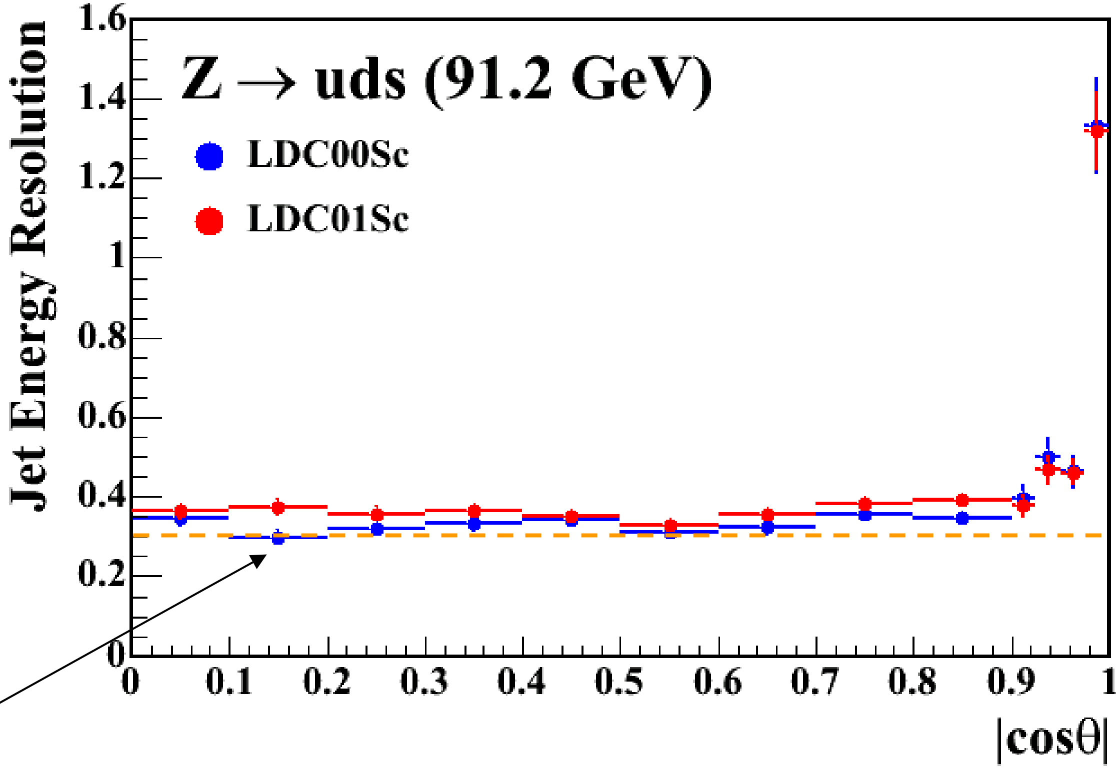


- **RMS of Central 90 % of Events**

	$\sigma_E/E = \alpha\sqrt{(E/GeV)}$
LDC00Sc	**35.3±0.6%**
LDC01Sc	**37.1±0.6 %**

Results : Z uds events
Angular dependence

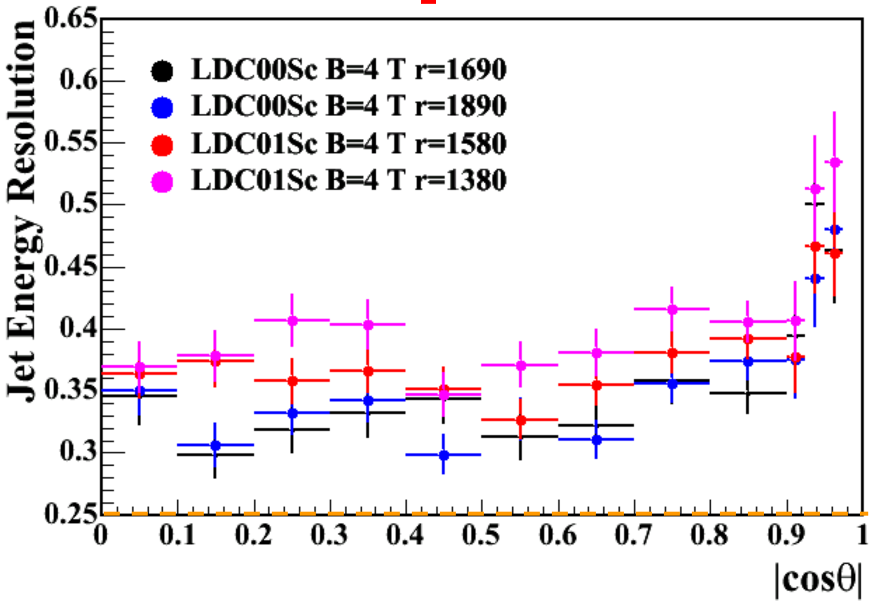
- **Plot resolution vs generated polar angle of qq system**



- **In barrel : 32-34 %/ √E(GeV)**

- **NOTE: approx 1.5 % improvement since Bangalore (= 8 hours work)**

Dependence on Radius



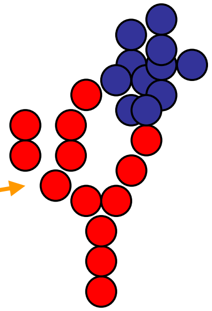
- ★ Some evidence that going to small radii gives worse performance
- ★ BUT… Z events + algorithm not finished
- ★ More in a moment…

Model	all angles	cosθ < 0.8
Tesla r_{tpc} = 1690, l_{tpc}=2730	35.3±0.6 %	32.9±0.7 %
Tesla r_{tpc} = 1890 l_{tpc}=2930	36.6±0.6 %	33.0±0.6 %
LDC r_{tpc} = 1580 l_{tpc}= 2200	37.1±0.6 %	36.0±0.6 %
LDC r_{tpc} = 1380 l_{tpc}= 2000	39.7±0.6 %	38.7±0.7 %

NOTE : All files copied from DESY using GRID tools

❹ Confusion...

★ Working towards an analysis of PFA performance...
★ e.g. Compare reconstructed PFOs with expected PFOs (from MC tree)
★ Look at MC PFOs that have been merged with hadronic shower from charged track



★ Starting to gain some interesting (?) insights (?)

Model (Z→uds @ 91.2 GeV)	cosθ < 0.8	Confusion (rms)	
		photon	n/K_L
LDC00Sc r_{tpc} = 1890mm	33.0±0.6 %	0.86 GeV	1.95 GeV
LDC00Sc r_{tpc} = 1690mm	32.9±0.7 %	1.13 GeV	1.86 GeV
LDC01Sc r_{tpc} = 1580mm	36.0±0.6 %	1.19 GeV	2.49 GeV
LDC01Sc r_{tpc} = 1380mm	38.7±0.7 %	1.34 GeV	2.92 GeV

Comments (within current implementation of PandoraPFA):

★ Confusion scales much as expected $\sim R^{-2}$
★ For ≥ Tesla-like radii confusion term does not dominate (at 91.2 GeV)
★ For < Tesla radii (i.e. LDC) confusion is an issue (even @ 91.2 GeV)
  - accounts (?) for degraded performance

Confusion: B-Field Dependence

Extremely hot off the press (4/4/06)
DO NOT TAKE TOO SERIOUSLY

Model	cosθ < 0.8	Confusion (rms)	
		photon	n/K_L
LDC00Sc B = 2T	37.4±0.7 %	1.74 GeV	2.98 GeV
LDC00Sc B = 4T	32.9±0.7 %	1.13 GeV	1.86 GeV
LDC00Sc B = 6T	32.8±0.7 %	0.72 GeV	1.42 GeV

Somewhat suspicious… but now using QGSP hadronic model

Comments (within current implementation of PandoraPFA):

- ★ Confusion appears to scale with B-field ~ $B^{-1/2}$
- ★ For Tesla Zs (at 91.2 GeV) going from 4→6T doesn't help
  - ♦ confusion not dominating

Confusion at Higher Energy Jets

Extremely hot off the press (4/4/06)
DO NOT TAKE TOO SERIOUSLY

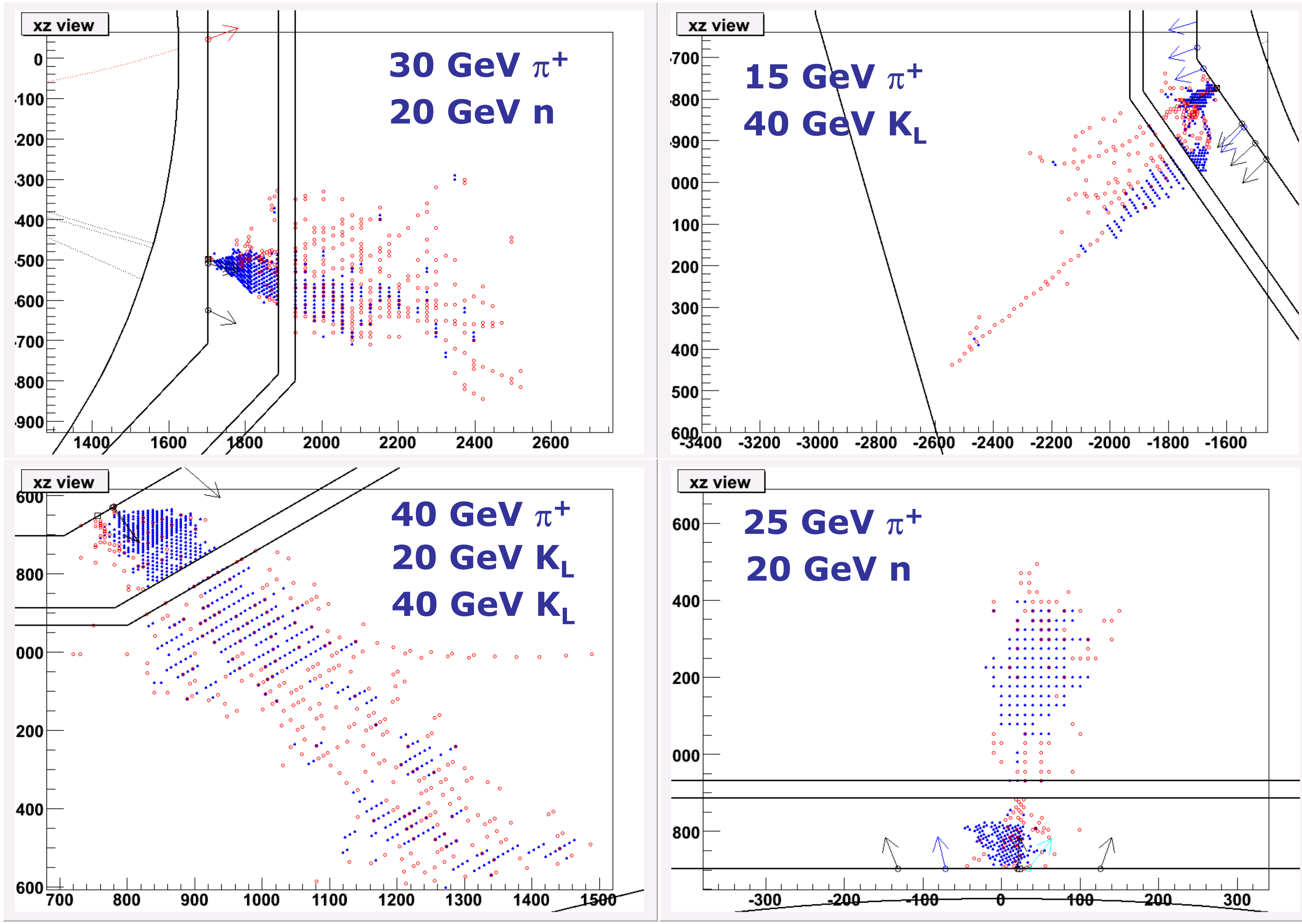
Model	cosθ < 0.8	Confusion (rms)	
		photon	n/K_L
LDC00Sc Z→uds @91 GeV	33 ± 1 %	1.1 GeV	1.9 GeV
LDC00Sc Z→uds @500 GeV	80 ± 6 %	13 GeV	16 GeV

Comments (within current implementation of PandoraPFA):

- ★ Confusion completely dominates for 250 GeV jets
- ★ However, algorithm can do significantly better
- ★ But some confusion is irreducible
- ★ At some energy PFA will fail in cores of jets
  - ★ at this point may need to resort to statistical subtraction

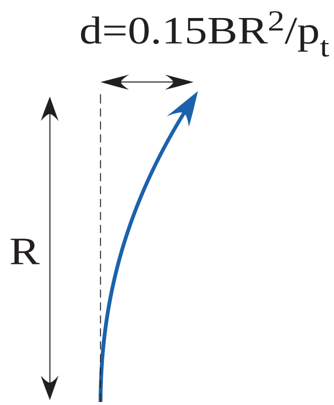
★ **What do the confused clusters look like?**

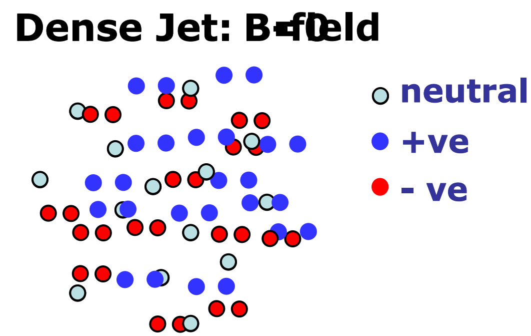
Confused ?


xz view
30 GeV π^+
20 GeV n
xz view
15 GeV π^+
40 GeV K_L
xz view
40 GeV π^+
20 GeV K_L
40 GeV K_L
xz view
25 GeV π^+
20 GeV n

❺ Conclusions

- ★ PandoraPFA starting to achieve OK perfromance
- ★ But only at 91.2 GeV
- ★ At 91.2 GeV confusion is not a big issue !
- ★ At higher energies confusion rules
- ★ In PandoraPFA "Confusion F.O.M" ~ $B^{1/2}R^2$





- ★ 30%/√E will soon be achieved for Z→uds with TESLA concept
  - confusion not dominating
- ★ For the smaller LDC concept – less clear (some confusion)
- ★ For 250 GeV jets, I doubt 30%/√E can be achieved !
- ★ PFA performance much more complex than α/√E
- ★ FINALLY bringing clarity (i.e. removing confusion) is not a pure clustering problem !

❻ Outlook

PandoraPFA can/will be improved:

- ✦ still a few features (i.e. does something silly)
- ✦ some problems with tracking extrapolation to endcap
- ✦ photon ID is quite basic
- ✦ + new ideas (for high density events)
- ✦ + ways to identify confused clusters

★ Code runs within **Marlin** framework and is "nearly" ready for release - BUT first optimise for higher energy jets

★ + code needs tidying up

- ✦ started with decent OO structure
- ✦ then grew organically…

★ Reluctant to release until performing real PFA…

★ **BUT** Aim to have complete algorithm before Summer