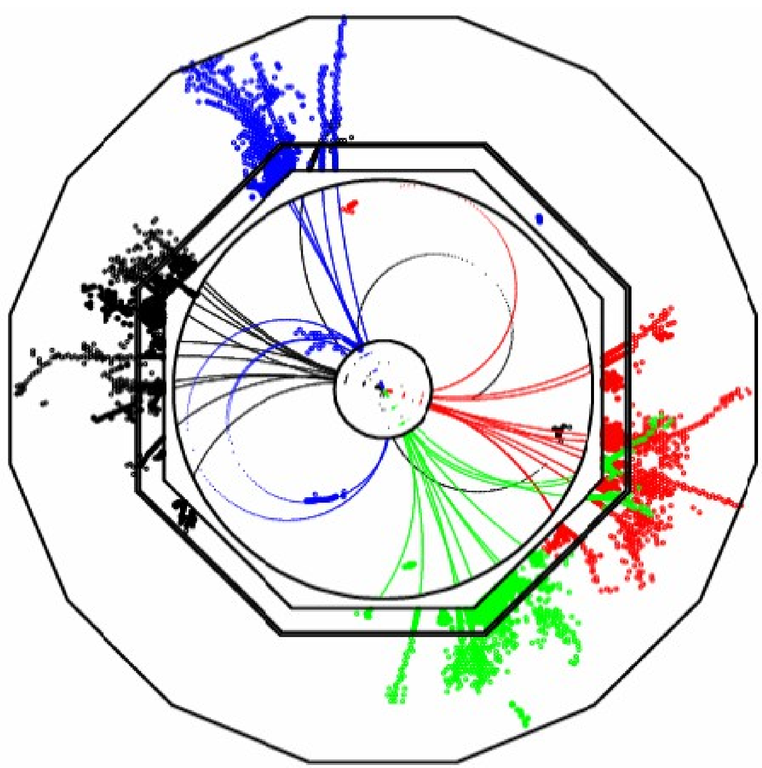
Pandora PFA - a brief introduction



Trong Hieu TRAN

Laboratoire Leprince-Ringuet,
Ecole polytechnique, CNRS/IN2P3

Pattern recoginition meeting @ LLR – 20/11/2012

Contents

- Energy resolution requirement at ILC
- Particle Flow
- PandoraPFA algorithm
- Pandora Performance
- Study of Ecal performance with reduced number of layers

Jet energy resolution requirement

- Aim to get jet energy resolution giving di-jet mass resolution similar to Gauge boson width

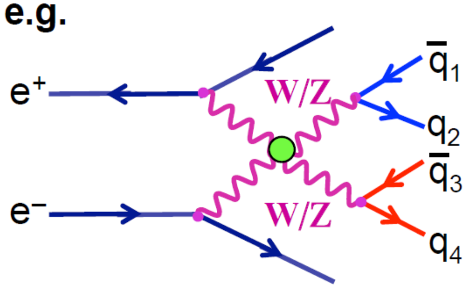
$$\sigma_m/m = 2.7\,\% \approx \Gamma_W/m_W \approx \Gamma_Z/m_Z$$

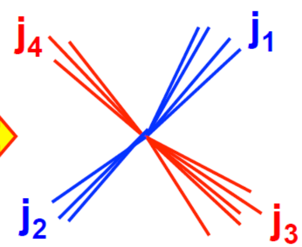
$$\frac{\sigma_E}{E} = \frac{\alpha}{\sqrt{\mathrm{E(GeV)}}} \oplus \beta$$

→ Stochastic term must be $\lesssim 30\,\%/\sqrt{\mathrm{E(GeV)}}$

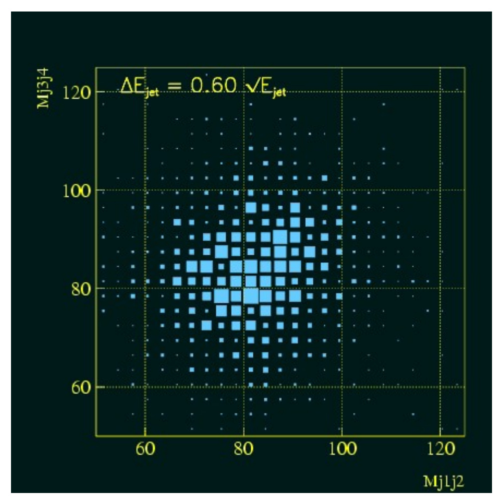
→ $\sigma_E/E < 3.8\%$

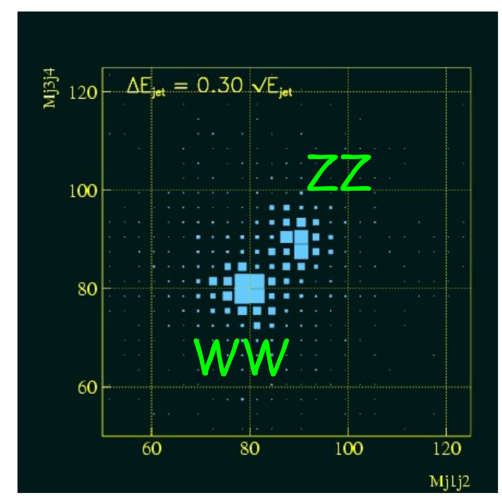
e.g.





- Reconstruction of two di-jet masses allows discrimination of WW and ZZ final states





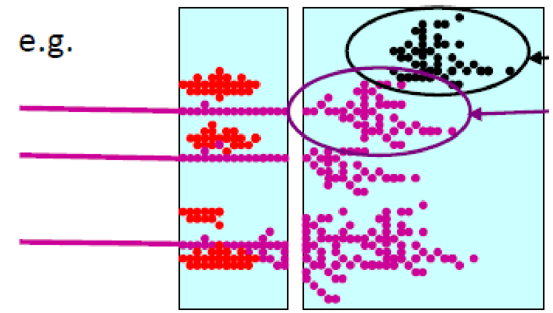
Particle Flow Algorithms

Very different from traditional approach (energy flow, e.g. H1, D0, …), aim to use the energy measured in calorimeters as little as possible.

The challenge of particle flow algorithms:

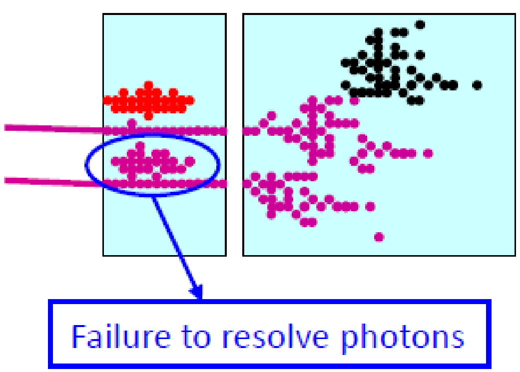
- Avoid double counting of energy from same particle
- Separate energy deposit from different particles

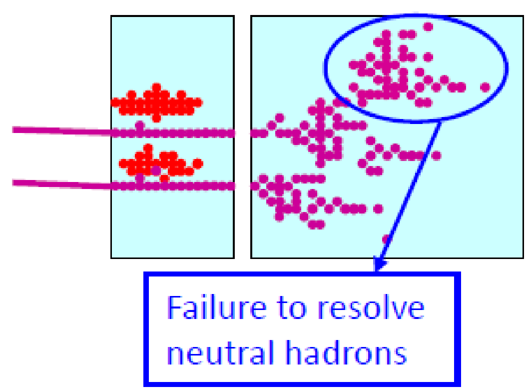
e.g.

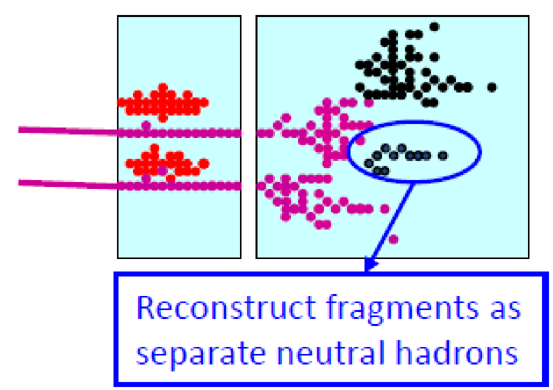


Level of mistakes: confusion, determines jet energy resolution, not intrinsic calorimetric performance

Three types of confusion:







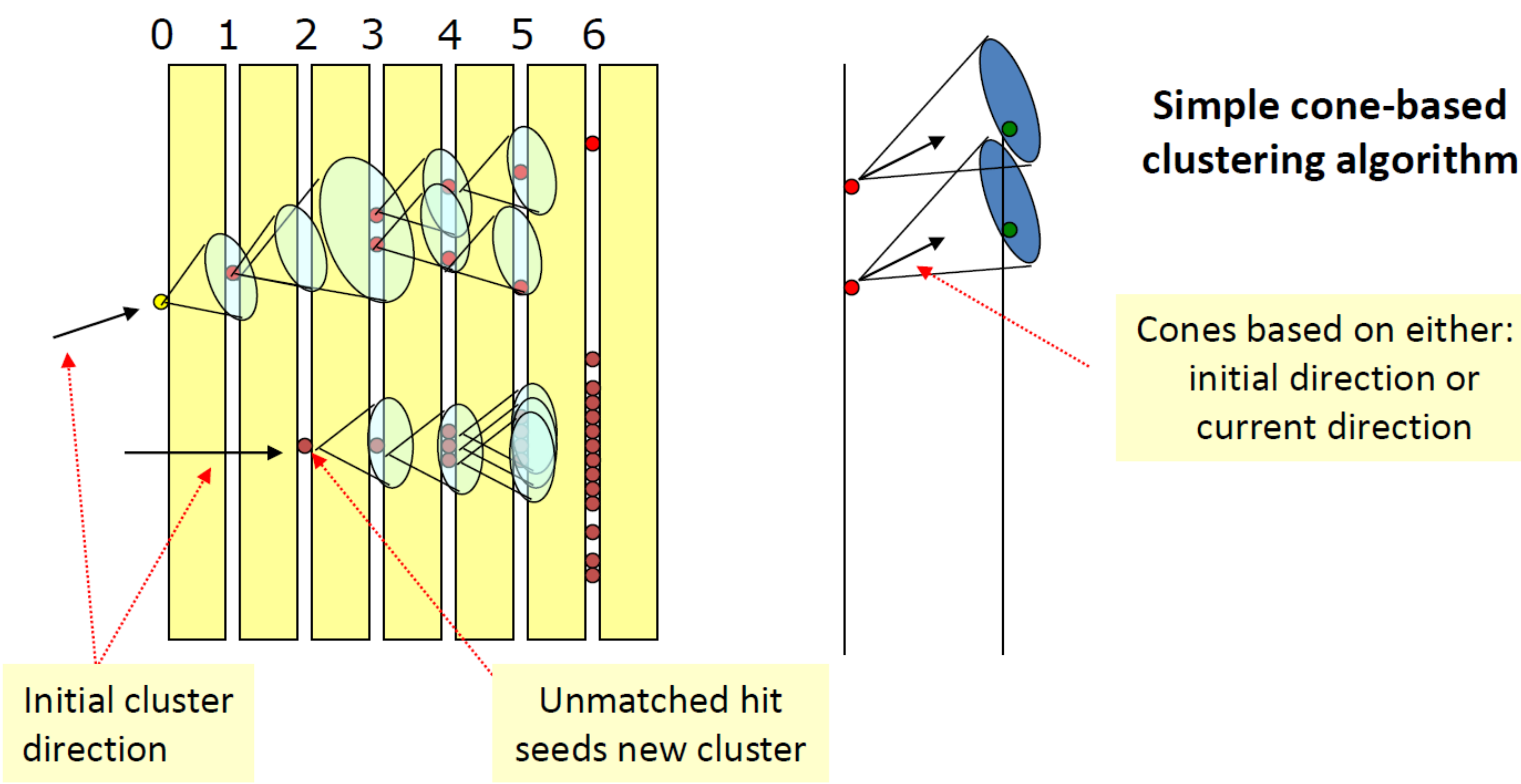
Pandora PFA overview

Eight main steps:

- Track Selection/ Topology:
  - track topologies such as kinks and decays of neutral particles in the detector volume are identified
- Calorimeter hit selection and ordering:
  - isolated hits, defined by proximity to others, are removed
  - remaining hits are ordered into pseudo-layers
- Clustering
  - Cone-based forward projective method. Seeding clusters using projections of rec. tracks onto the front face of ECAL
  - Photon clustering:
    - Consider ECAL hits to identify energy deposits from photons
    - Clustering on remaining hits
- Topological Cluster Merging
  - Apply only to clusters which have not been indentified as photon
- Statistical reclustering
  - For jet with E>50 GeV, attempts are made to recluser hits by applying clustering with different parameters until cluster splits, until $E_{clus} \sim E_{track}$
- Photon recovery and identification
  - More sophisticated photon id applied, improving photon tagging.
  - Recover cases where a primary photon is merged with had. shower
- Fragment removal: identify fragments of charged particle hadronic shower
- Formation of Particle Flow Objects (PFOs) – reconstructed particles

Cone Clustering

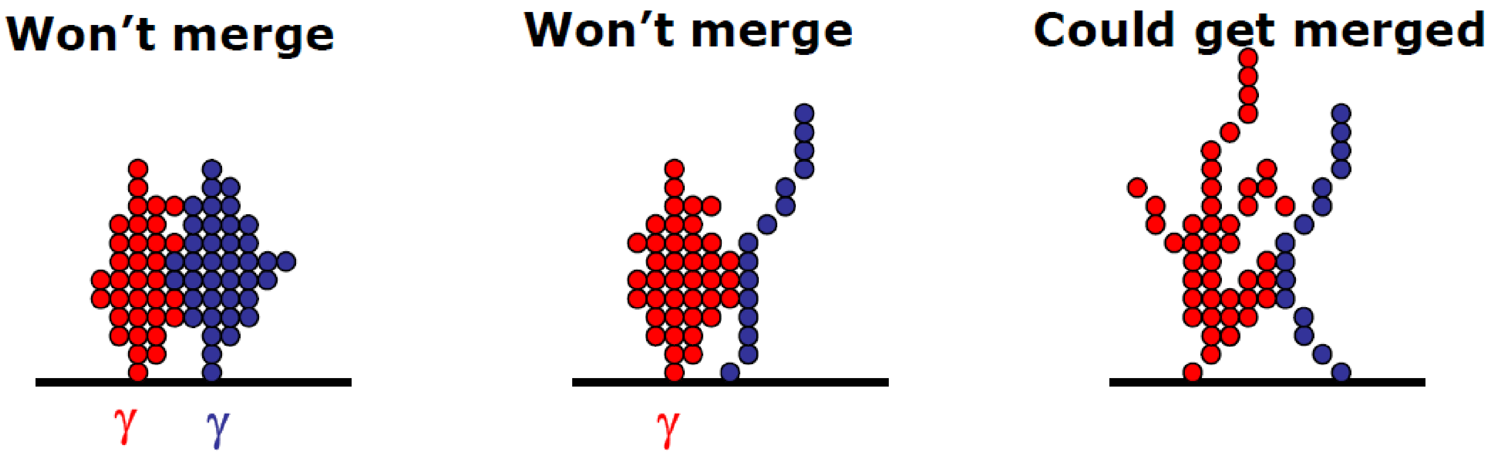
- Division of energy deposits into particles starts with simple cone-based clustering algorithm
- Clusters seeded by projections of inner detector tracks to surface of calorimeter
- Start at innermost layers and work outward, considering each calorimeter hit in turn
  - If hit lies within cone defined by existing cluster, and is suitably close, add hit to cluster
  - If hit is unmatched, use it to form a new cluster



Topological Cluster Association

- PandoraPFA philosophy: easier to put things together than split them up
- Clusters are associated together in two stages:
  - Tight cluster association – clear topologies
  - Loose cluster association – fix what has been missed

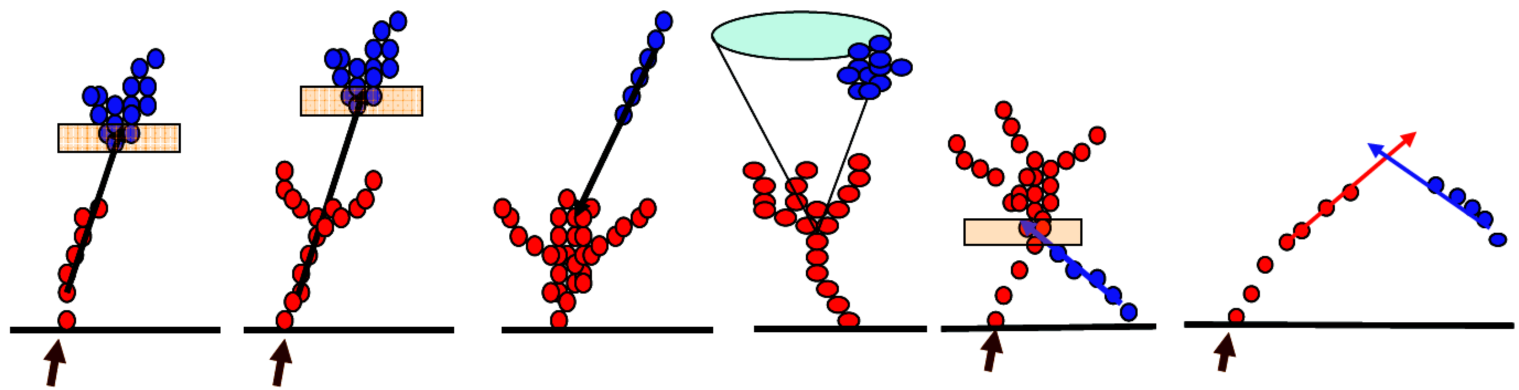
- Photon ID
  - Photon ID plays important role
  - Simple "cut-based" photon ID applied to all clusters
  - Cluster tagged as photons are immune from association procedure – just left alone



Topological Cluster Association (2)

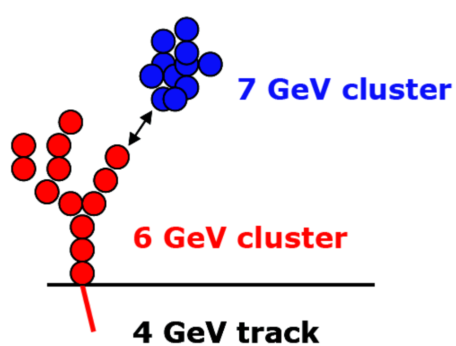
- Clusters associated using a number of topological rules

Clear associations:
 Join clusters which are clearly associated making use of high granularity + tracking capability: very few mistake



Less clear associations:

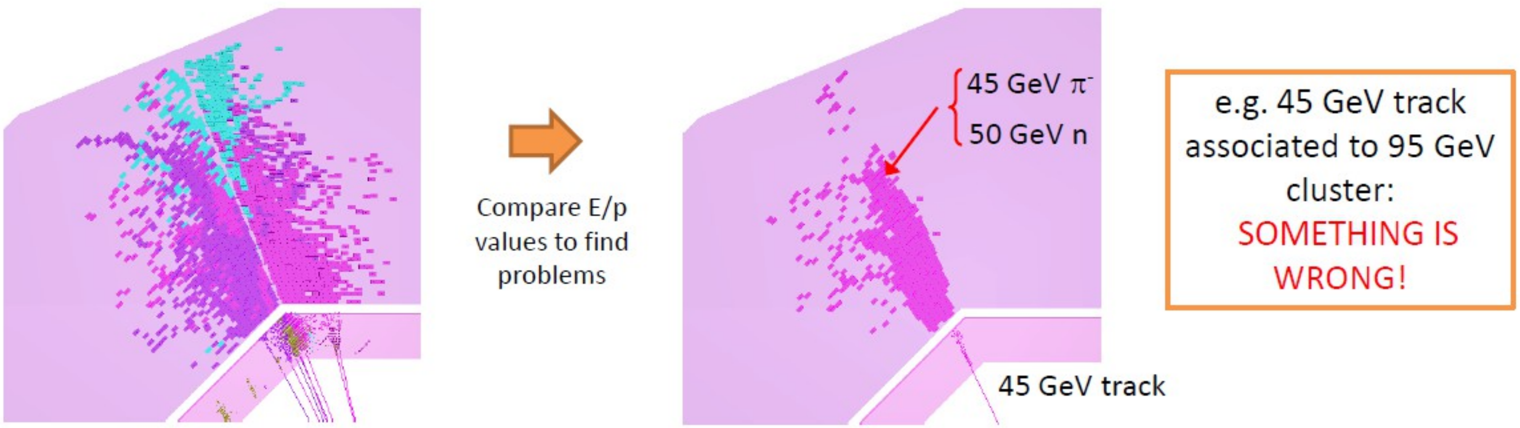
e.g. **Proximity**



Use E/p consistency to veto clear mistakes

Reclustering

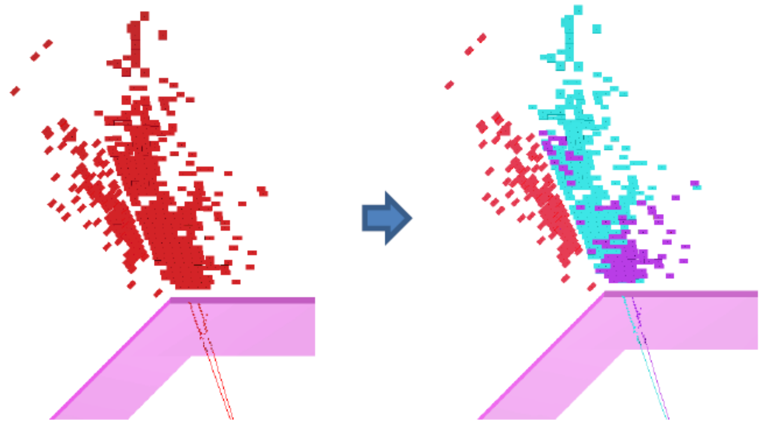
At some point, in high density jets (high energies), reach limit of “pure” particle flow.
•Cannot cleanly resolve neutral hadrons in hadronic showers.
•Use information from track-cluster associations to identify pattern-recognition problems:



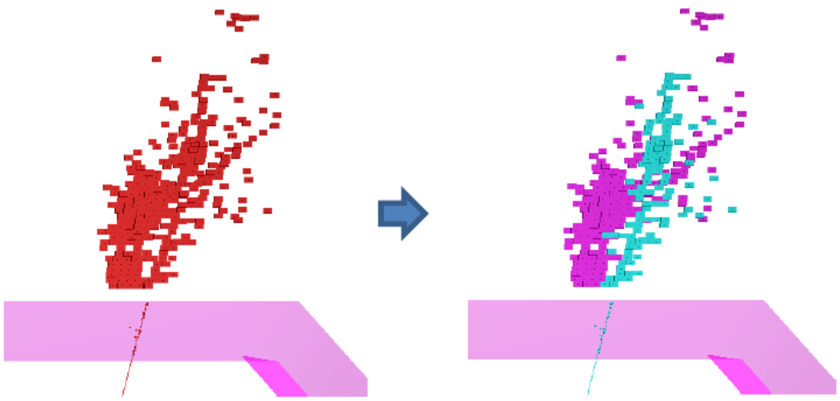
After topological associations
Find n absorbed into π⁻ cluster

Address the problem “statistically”; if we identify significant discrepancy between energy of a cluster and momentum of its associated track, choose to recluster.
•Alter clustering parameters, or change clustering algorithm entirely, until cluster splits in such a way that we obtain sensible track-cluster associations.

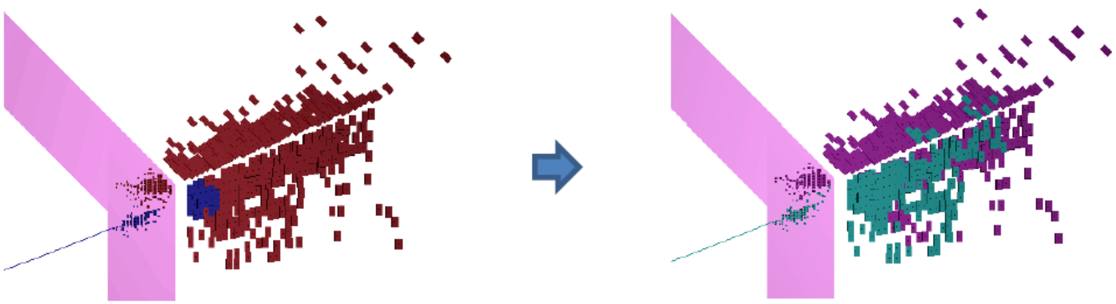
Reclustering strategies



1. Multiple tracks associated to single cluster – split cluster.



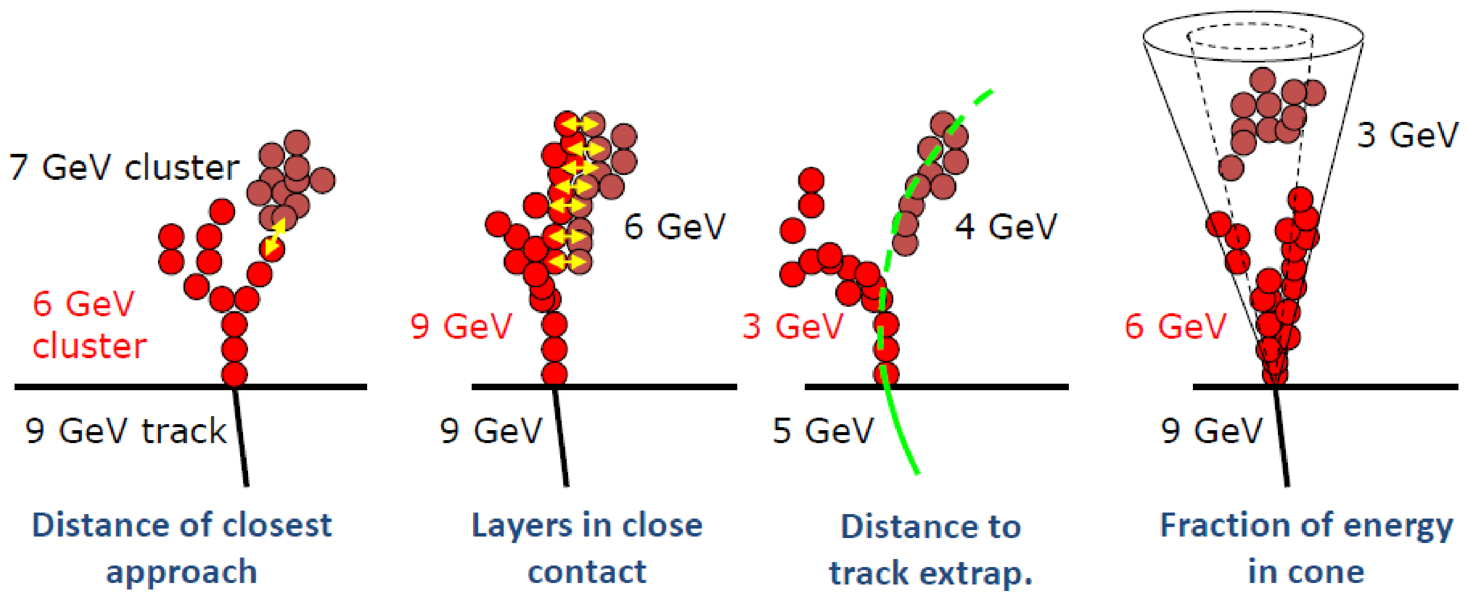
2. Cluster energy much greater than track momentum – split cluster.



3. Track momentum much greater than cluster energy – bring in nearby clusters and reconfigure.

Fragment Removal

Fragment removal algorithms aim to remove neutral clusters (no track-associations) that are really fragments of charged (track-associated) clusters. Relevant clusters are merged together.

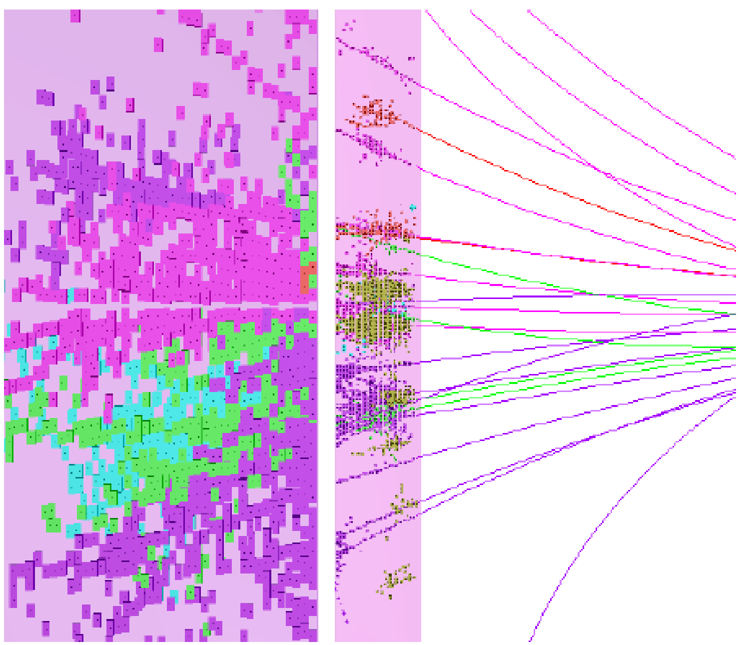


Look for evidence of association between nearby clusters. Evidence calculated as numerical score.
- Required evidence score also calculated, based on change in E/p $\chi 2$, location in calorimeters, etc.
- Clusters merged if collected evidence greater than required evidence. Ad hoc, but works well.

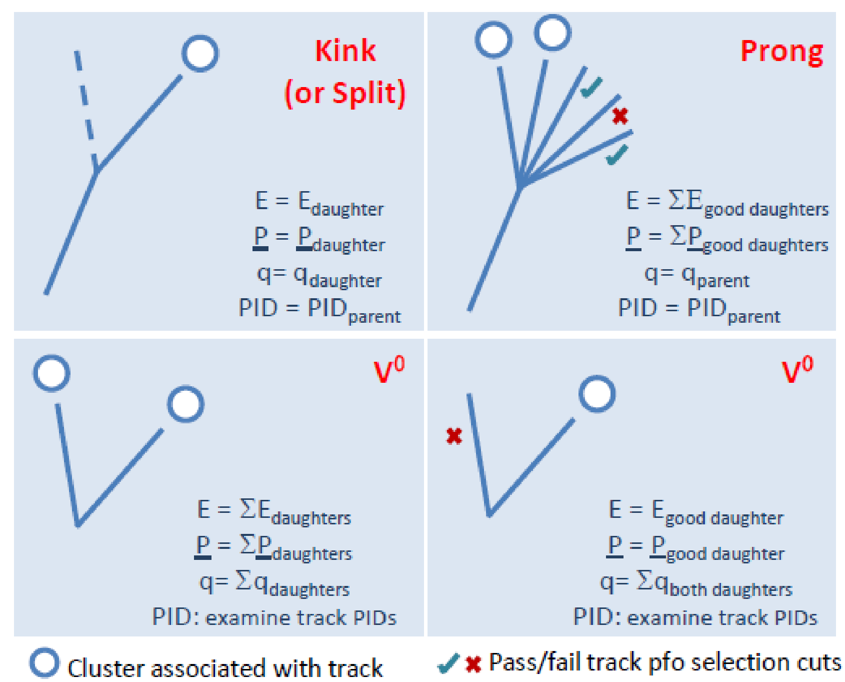
Track associations

- Use of tracking information is essential in PFA and major part of Pandora PFA
- The Pandora track-cluster association algorithms look for consistency between cluster properties and the helix-projected track state at the front face of the calorimeter:
  - Close proximity between cluster and track positions.
  - Consistent track and initial cluster directions.
  - Consistent track momentum and cluster energy.

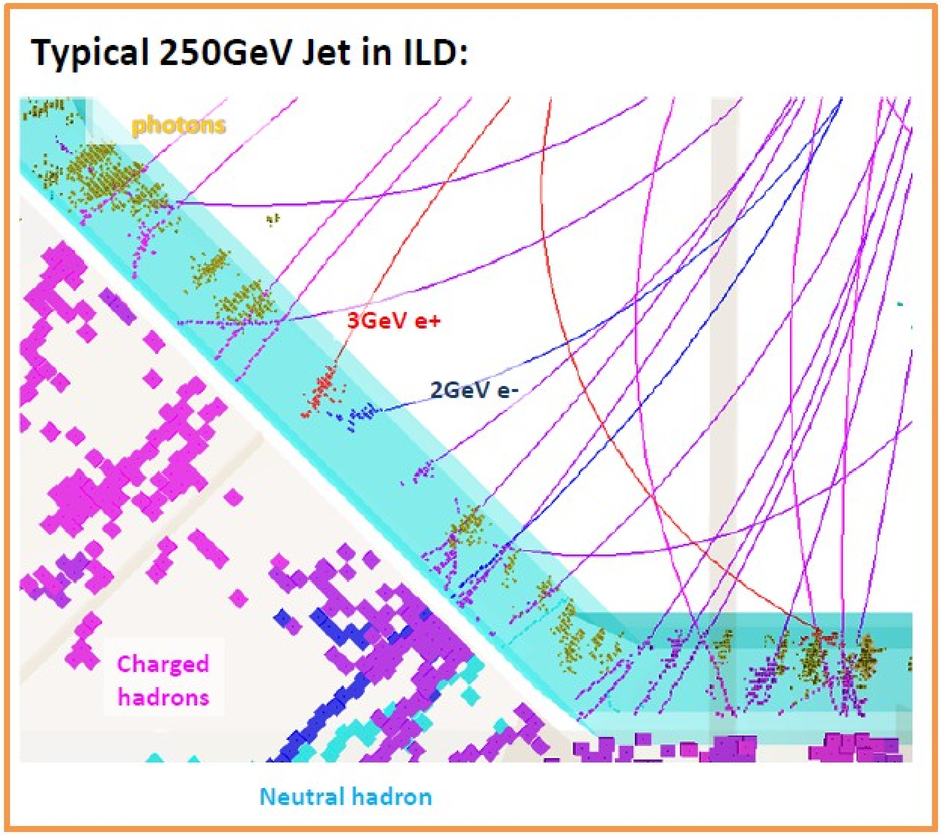
Clusters ↔ Tracks



Track-track relationship information is also used:



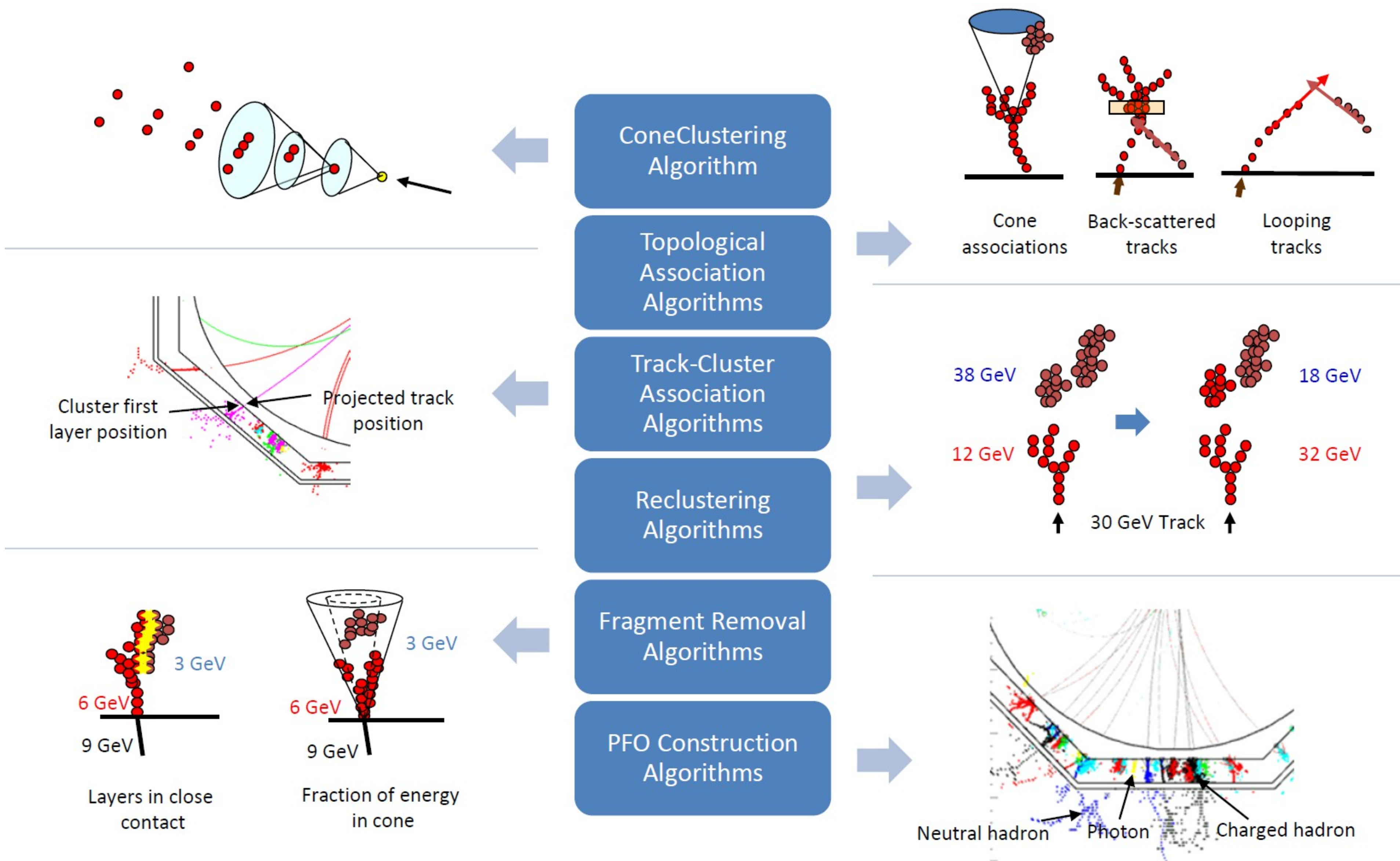
Particle Flow Objects



Particle flow objects built from track and (associated) clusters using series of simple rules:

- Obtain a list of reconstructed and identified particles, with measured energies.
- Calorimeter energy resolution not critical – most energy from tracks.
- Level of mistakes in associating hits with particles dominates the jet energy resolution.

Pandora Algorithms (illustrated)



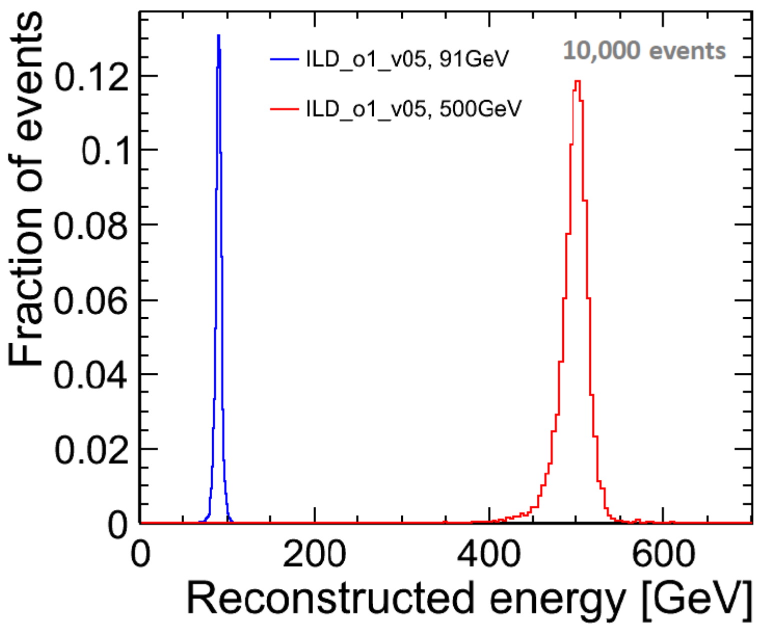
PandoraPFA performance

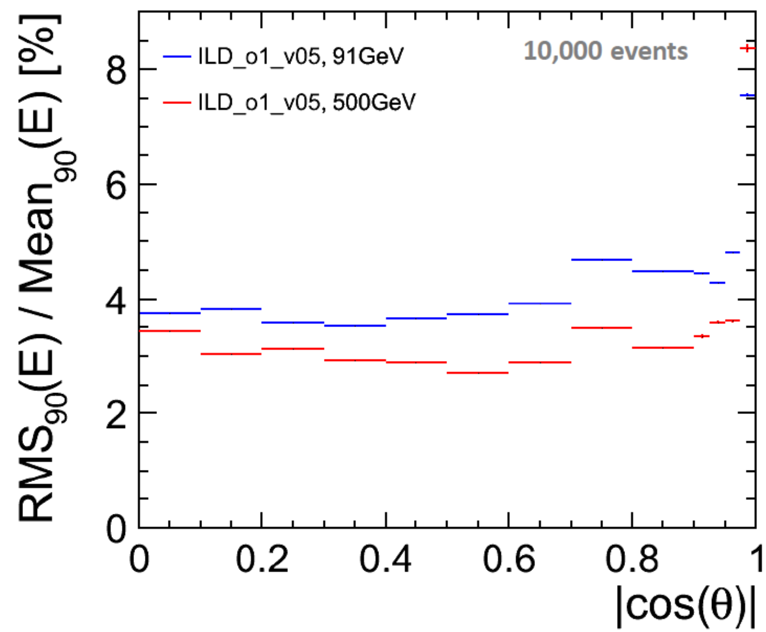
Jet energy resolution (for ILD)

- Study of jet energy resolution for ILD_o2_v05 using Z→qqbar events with Z decaying at rest (Z→uds)
- at CM energies: 91, 500 GeV

E_{jj} (= 2 * E_j)	91GeV	500GeV
ILD_o1_v05 v01-15-02-pre05	σ_E/E: 3.65 ± 0.05 mean: 90.49 GeV	σ_E/E: 2.97 ± 0.04 mean: 500.57 GeV

J. Marshall, ILD meeting 26/9/2012
(calibration constants were optimised for JER)



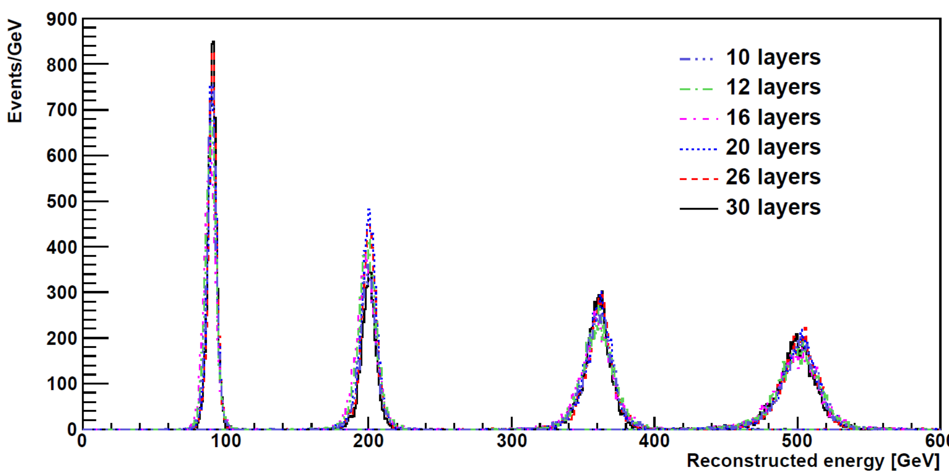


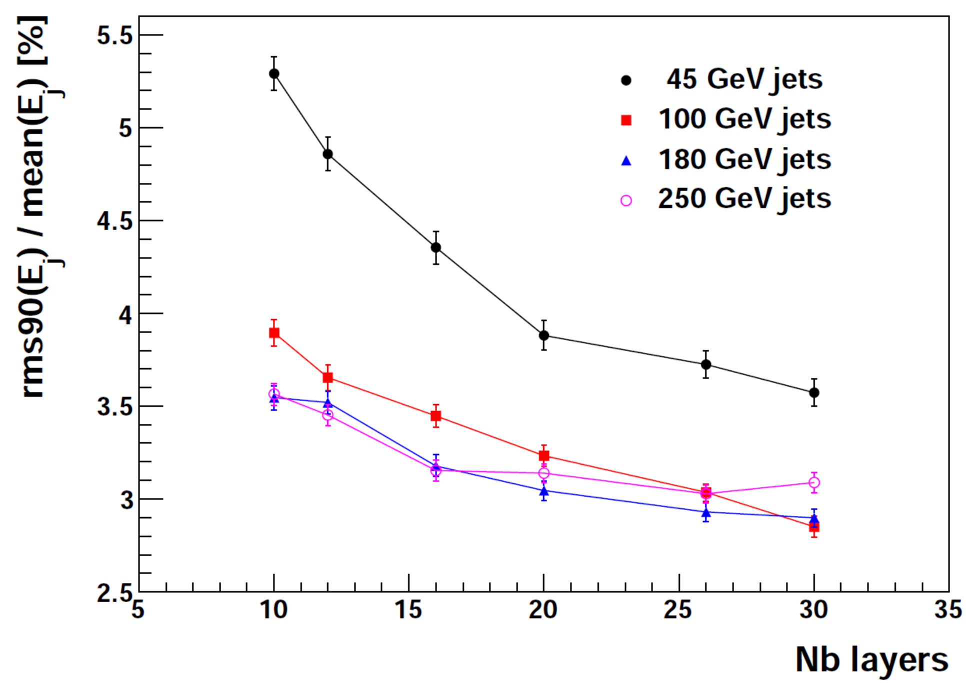
RMS90: RMS of the smallest region containing 90% of events

$$\frac{\mathrm{rms}_{90}(E_j)}{E_j} = \frac{\mathrm{rms}_{90}(E_{jj})}{E_{jj}}\sqrt{2}$$

Detector optimisation using PandoraPFA: ECAL with reduced number of layers

- Use of PandoraPFA to study the degradation of jet energy resolution by reducing the number of layers in ECAL (Z → uds events, CM energies: 91, 200, 360, 500 GeV)
- Keep total W thickness, only total Si surface is modified
- Since PandoraPFA is minimal-detector-dependant, the performance depends mostly on fraction of energy deposit in ECAL (only calibration constants need to be retuned, algorithms should remain the same for different ECAL models).





Summary

- PandoraPFA is a sophisticated algorithm and is likely the key for archieving ILC goal
- It had been applied in many studies for detector optimisation (B-field, TPC radius, Ecal number of layers, ECAL-HCAL cell size, ...
- Minimal detector dependence made PandoraPFA largely applicable in different detector concepts, e.g. ILC-ILD, CLIC-ILD, CLIC-SiD.

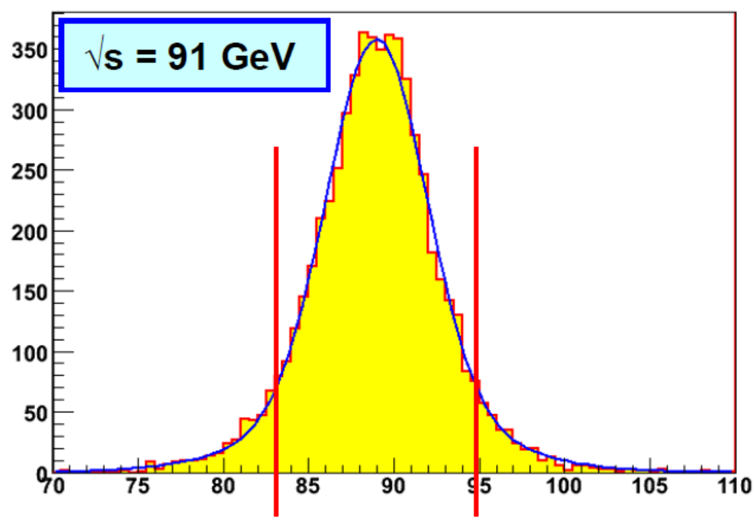
Extra-slides

RMS90

Figures of Merit:

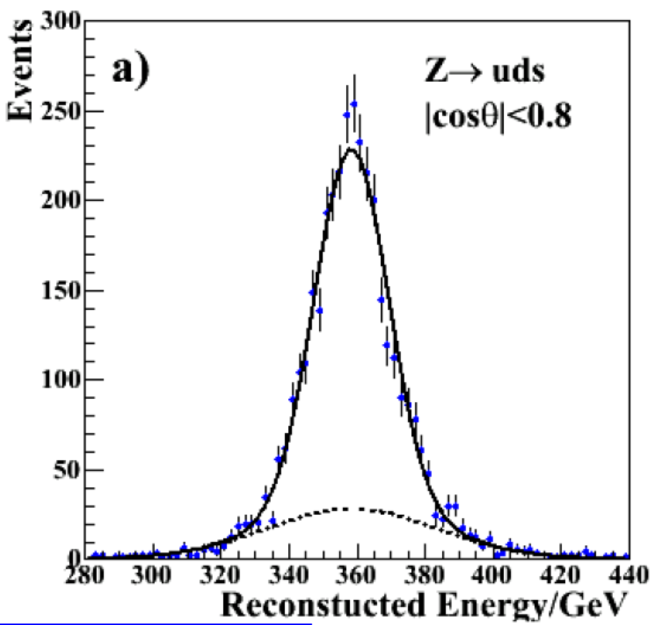
rms_{90}

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



σ_{80}

- Fit sum of two Gaussians with same mean. The narrower one is constrained to contain 80 % of events
- Quote σ of narrow Gaussian



It turns out that $rms_{90} \approx \sigma_{80}$

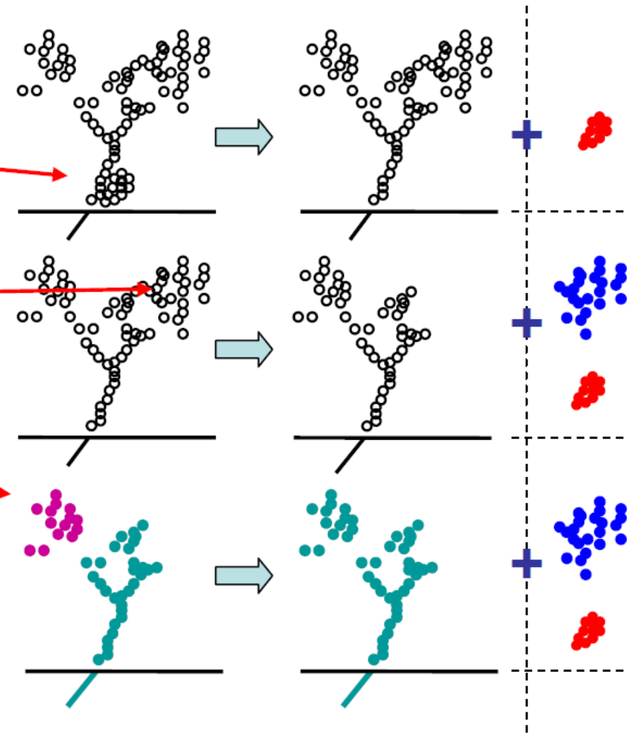
(need care when comparing to Gaussian resolution)

Understanding PFA Performance

Try to use various “perfect PFA” algorithms to pin down main performance drivers (resolution, confusion, …)
Use MC to “cheat” various aspect of PFlow

PandoraPFA options:

- **PerfectPhotonClustering**
  hits from photons clustered using MC info and removed from main algorithm
- **PerfectNeutralHadronClustering**
  hits from neutral hadrons clustered using MC info…
- **PerfectFragmentRemoval**
  after PandoraPFA clustering “fragments” from charged tracks identified from MC and added to charged track cluster
- **PerfectPFA**
  perfect clustering and matching to tracks



Understanding PFA Performance (2)

Contribution	σ_E/E			
	45 GeV	100 GeV	180 GeV	250 GeV
Calo. Resolution	3.1 %	2.1 %	1.5 %	1.3 %
Leakage	0.1 %	0.5 %	0.8 %	1.0 %
Tracking	0.7 %	0.7 %	1.0 %	0.7 %
Photons "missed"	0.4 %	1.2 %	1.4 %	1.8 %
Neutrals "missed"	1.0 %	1.6 %	1.7 %	1.8 %
Charged Frags.	1.2 %	0.7 %	0.4 %	0.0 %
"Other"	0.8 %	0.8 %	1.2 %	1.2 %

Comments:

- ★ For 45 GeV jets, jet energy resolution dominated by ECAL/HCAL resolution
- ★ Track reco. not a large contribution (Reco ≈ CheatedTracking)
- ★ "Satellite" neutral fragments not a large contribution
  - efficiently identified
- ★ Leakage only becomes significant for high energies
- ★ Missed neutral hadrons dominant confusion effect
- ★ Missed photons, important at higher energies

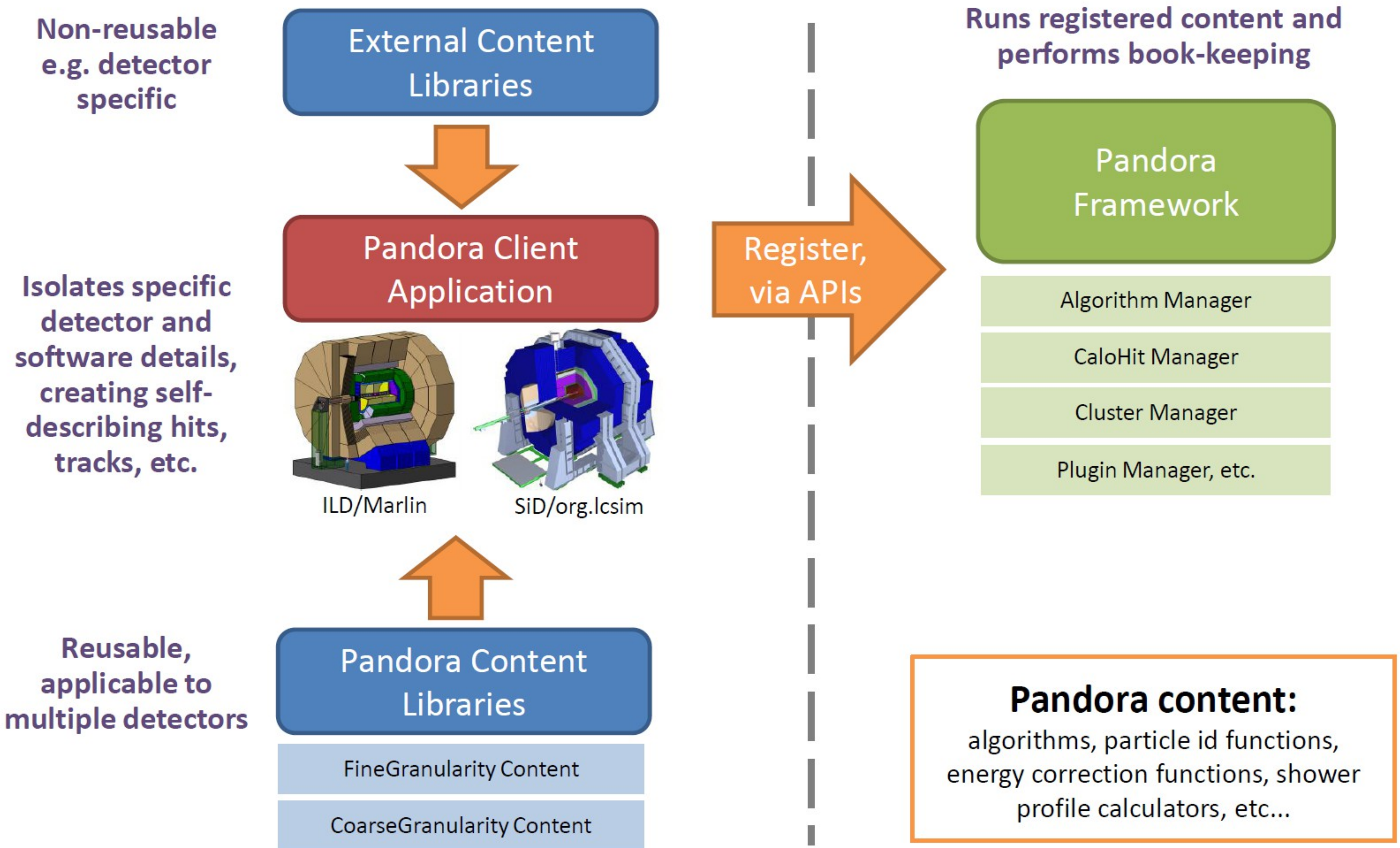
Physics lists dependency

Validation of PandoraPFA is studied with very different modelling of hadronic showers (physics lists in Geant4)

Physics List	Jet Energy Resolution			
	45 GeV	100 GeV	180 GeV	250 GeV
LCPhys	3.74 %	2.92 %	3.00 %	3.11 %
QGSP_BERT	3.52 %	2.95 %	2.98 %	3.25 %
QGS_BIC	3.51 %	2.89 %	3.12 %	3.20 %
FTFP_BERT	3.68 %	3.10 %	3.24 %	3.26 %
LHEP	3.87 %	3.15 %	3.16 %	3.08 %
χ^2	23.3 / 4	17.8 / 4	16.0 / 4	6.3 / 4
rms	4.2 %	3.9 %	3.5 %	2.5 %

Only a weak dependence < 5% (on the total, not just the hadronic confusion term)

Pandora Client App



Pandora Algorithms

Pandora Framework

- Algorithm Manager
- CaloHit Manager
- Cluster Manager
- Plugin Manager, etc.

APIs →

Pandora Algorithms

- Clustering Algorithms
- Cluster Merging Algorithms
- Track-Cluster Association Algs
- Fragment removal Algs, etc.

Owns named collections of Pandora objects: calo hits, tracks, clusters, PFOs.

Can perform memory management, as content can only be provided or accessed via APIs.

← APIs

Use APIs to access Pandora objects and carry out particle flow reconstruction.

Physics-driven code, with nested structure promoting re-use of code to perform specific tasks.

Currently available: **56** algorithms for fine granularity detectors, including clustering, visualization, etc.
6 algorithms for reconstruction in coarse granularity detectors.
Electron, **muon** and **photon** id functions; pseudo-layer calculators, and more.