T-SDHCAL Software : APRIL, a particle flow algorithm for future e^+e^- colliders

Tanguy Pasquier (speaker), Gérald Grenier, Imad Laktineh, Bo Li, Rémi Été

Université Lyon 1 - IP2I Lyon (FLC group)

16th France-China Particle Physics Network/Laboratory workshop (FCPPN/L 2025), Qingdao
July 24th, 2025

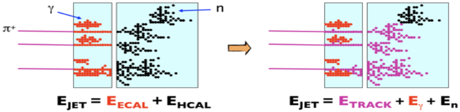
Outline

Contents

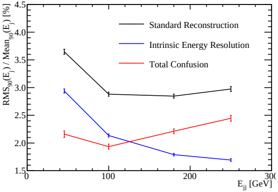
Particle flow calorimetry

Particle Flow Algorithm (PFA)

- PFA is the approach chosen for future Higgs factories
- Use optimal sub-detector for jet energy estimation :
  tracker ($\sim 60\%$), ECAL ($\sim 30\%$), HCAL ($\sim 10\%$).
- Separate energy depositions from close-by particles : high granularity is key point



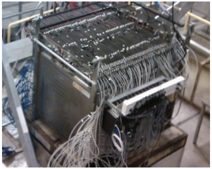
Extensive studies have been done with ILD detector option 1 (AHCAL, ILD baseline) and PandoraPFA algorithm.
At higher jet energy ($E \gtrsim 100$ GeV), dominant contribution to resolution is confusion.
See Steven Green, Cambridge University Thesis 2017

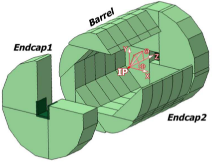


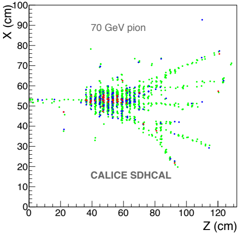
Semi-Digital HCAL

SDHCAL energy reconstruction

$$E_{reco} = \alpha_1 N_1 + \alpha_2 N_2 + \alpha_3 N_3$$







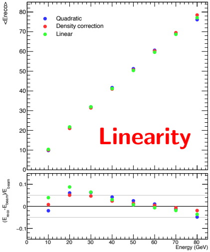
Options for E_{reco}

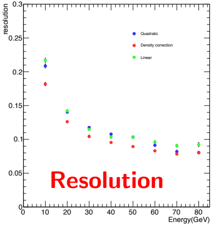
Thresholds(pC): 0.11, 5, 15

Quadratic $\alpha_i = a_i N_{hit}^2 + b_i N_{hit} + c_i$ (TB default)

Linear α_i constant (Pandora default).

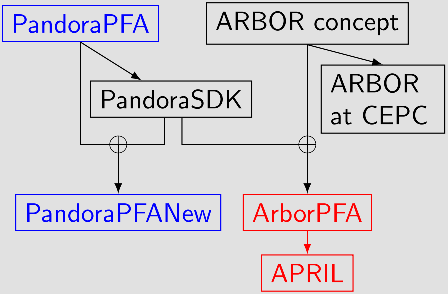
Density α_i = above × f(N neighbour hits)





PFA history

ILD PFA reconstruction



PFA strategy

Both PandoraPFA and APRIL, construct many small clusters then merge them.

- APRIL $\simeq$ ARBOR concept + PandoraSDK algorithms

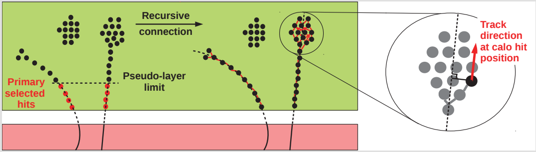
Contents

The APRIL algorithm

APRIL : Algorithm for Particle Reconstruction at ILC from Lyon.

The clustering strategy

1. Start from tracks
   (track driven clustering), extrapolate tracks in calorimeters → cluster hits close to the tracks.



   Clusters linked to tracks are defined as charged clusters.

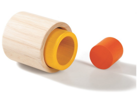
2. Perform Arbor like clustering with all hits (parameters set to avoid making big clusters).
3. Some hits remain unclustered → Nearby hits merging : remaining unclustered hits are clustered with mlpack DBSCAN (efficient Nearest Neighbour clustering)
4. Cluster merging : if $E_{track} > E_{cluster}$, merge nearby cluster.

Arbor like clustering

Graph theory : a shower is an oriented tree.

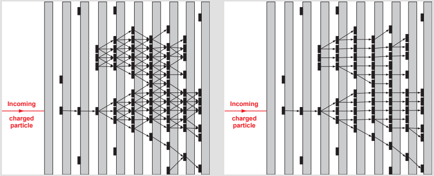
Orientation

- Rearrange hits in virtual nested cylinders (= pseudo layers)
- Count them from the inside.
- Forward direction = increase pseudo layer number.

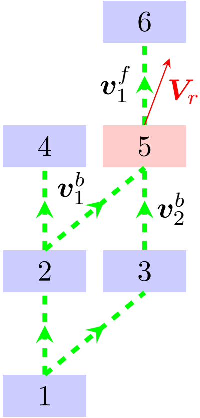


Process

1. Connect all neighbouring hits (use mlpack NeighborSearch).
2. Clean connectors = keep max one backward connection per hit.



Connector cleaning



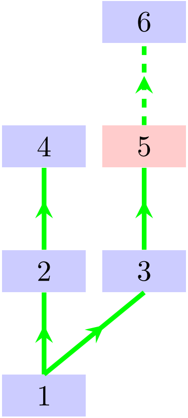
Reference direction

$$\mathbf{V}_r = w_b \times \sum_i \mathbf{v}_i^b + w_f \times \sum_j \mathbf{v}_j^f$$

- Can use different depth (max distance in pseudolayer) and weight in forward-backward direction.
- Define angle θ with respect to $\mathbf{V}_r$ and distance d between hits (in λ_I or X_0 unit).
- Keep connector with smallest

$$\kappa = \theta^{p_\theta} \times d^{p_d}$$

(if $\theta = 0$, smallest d)



Results

- Event samples: $e^+e^- \to q\bar{q}$, where $q = u, d, s$ ($|\cos\theta_q| < 0.7$)
- With ILD option 2 large (SDHCAL), Videau geometry, ILCSoft, linear reconstruction
- Jet energy resolution, JER $= \frac{\mathrm{RMS}_{90}(\mathrm{E}_j)}{\mathrm{Mean}_{90}(\mathrm{E}_j)} = \sqrt{2} \cdot \frac{\mathrm{RMS}_{90}(\mathrm{E}_{jj})}{\mathrm{mean}_{90}(\mathrm{E}_{jj})}$
- JER at 91.2 GeV: APRIL: 4.5%; Pandora: 4.5%; Perfect PFA: 3.28%

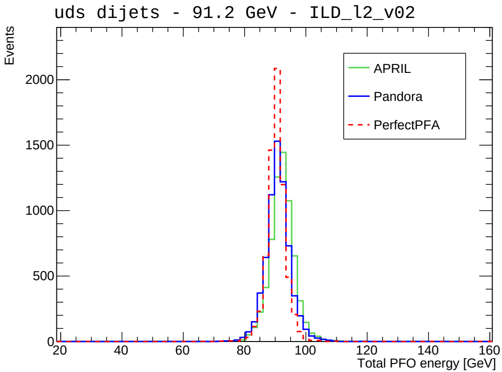
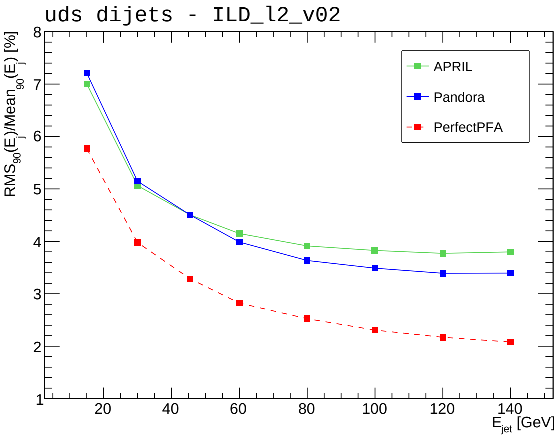


Figure: The energy of reconstructed PFO at $E_{\mathrm{CM}} = 91.2$ GeV.

Results (continued)



- APRIL better than Pandora at low energies for SDHCAL
- APRIL has no reclustering → crucial to counter confusion.

Reclustering / Cluster cutting

- Reclustering : Break the cluster and restart clustering with different parameters (Pandora strategy)
- Cluster cutting : cut the cluster and remove some parts

AMSTER algorithm

- Based on graph theory with Minimum Spanning Tree (MST)
- Cut the connections with bigger weights
- Implementation started and tested on simple cases

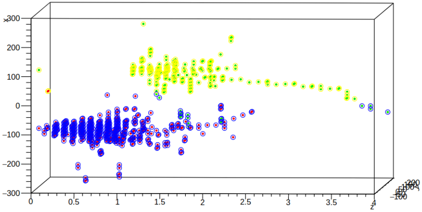
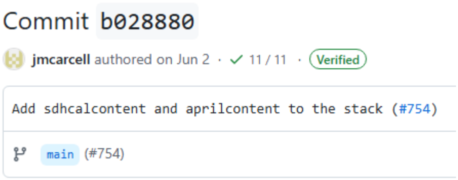


Figure: Event of a π^- (red) and a K^0 (green) in two clusters (blue and yellow)

T-SDHCAL Software inclusion

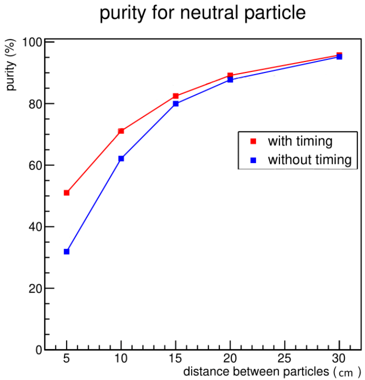
- SDHCALContent : Contains T-SDHCAL plugins (corrections, energy reconstruction formulas, ...) ▸ Git repo
- APRILContent : Contains the APRIL algorithms ▸ Git repo
- DDMarlinPandora update : Updated DDMarlinPandora able to run APRIL and SDHCALContent ▸ Git repo
- All recently added to the Key4hep stack ▸ Git repo
- Anyone can now run APRIL and/or the SDHCALContent plugins

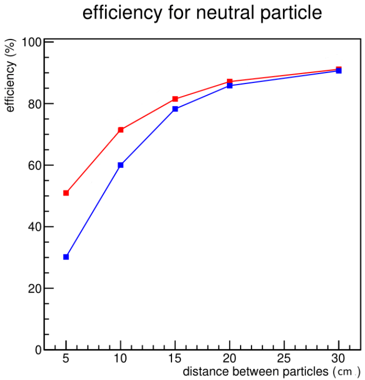


Contents

PFA with timing

- SDHCAL should allow timing segmentation < 100 ps thanks to MRPC $\rightarrow$ T-SDHCAL
- Possibility to follow the particles in "real time"
- Previous studies : timing improves separation
- **Goal** : Add timing to APRIL





Different applications

- Delete non-causal connectors between hits $(\beta > 1)$ ✓
- Late neutrons tagging to treat them separately ✓
- Pseudo layers ordered in timing and not in space ✓
- Other applications in discussion : cluster merging, seeds counting...
- → Ongoing studies to see the timing impact on energy resolution and showers separation

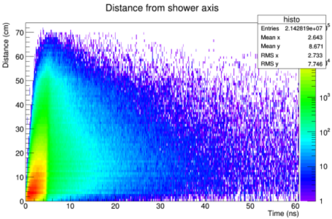


Figure: Simulated distribution of hits in the SDHCAL as function of their distance from the shower axis and the hit time, for all hits (left) and for all hits except those induced by neutrons (right)

Contents

Summary

- A particle flow algorithm implementing the ARBOR approach has been developed in the PandoraSDK framework.
- Available for anyone from the Key4hep stack
- Competing with PandoraPFA at low and intermediate jet energies in SDHCAL.
- Work started for cluster cutting algorithm for APRIL is expected to improve the reconstruction at higher jet energies.
- Next steps
  - Fully implement split cluster procedure (AMSTER)
  - Study the impact of timing on energy resolution and shower separation

Thank you for your attention !

Backup

Merging clusters distances

- Two methods possible

Barycenter projection

Find the parent with the smallest d_{proj} and smallest θ

d_{proj} Distance of closest approach between the parent barycenter and daughter axis.

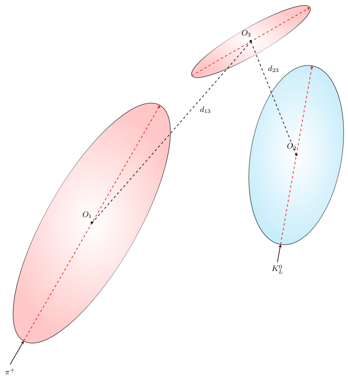
θ Angle between the 2 axes.

Axis crossing

Find the parent with the smallest d_{cross} and smallest d_{close}

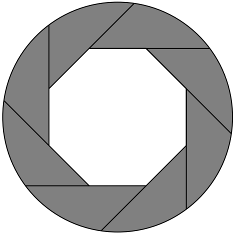
d_{cross} Distance of closest approach between the parent and the daughter axes.

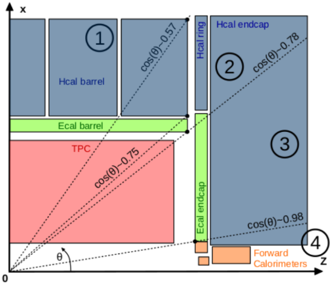
d_{close} Distance of closest approach between one of the parent cluster hit and the daughter axis.



Angle corrections

- **Goal** : implement angle corrections
- Purely geometric corrections
  - $N_{\text{hit}}^{\text{new}} = N_{\text{hit}} \times \text{Effect}$
  - Effect $\frac{1}{\cos\theta}$ for endcap
  - Effect $\frac{1}{\sin\theta}$ for barrel
  - Effect $\frac{1}{\cos\varphi}$ for barrel only
  - Videau geometry taken in consideration
- Created SDHCALContent for all SDHCAL related plugins ▸ Git repo
- Separating detector (SDHCAL, ILD option 2) from PFA (APRIL).





Determination of α_i parameters

- Several methods were tested:
  - Classical method: χ^2 minimization
  - Polynomial regression with sklearn ("ML")
  - "Split" methods
- Calibration performed using single KLong samples with energies between 5 GeV and 90 GeV with 1 GeV steps, in ILD_l2_v02 (SDHCAL)
- Calibration traditionnaly performed on the full energy range

Different methods

Classical method: χ^2 minimization

- $\chi^2 = \sum_{i=1}^{N} \frac{(E_{\text{mc,i}} - E_{\text{reco,i}})^2}{E_{\text{mc,i}}}$, where N is the number of events used
- Minimization performed using TMinuit and MIGRAD (ROOT)

Polynomial regression with sklearn ("ML")

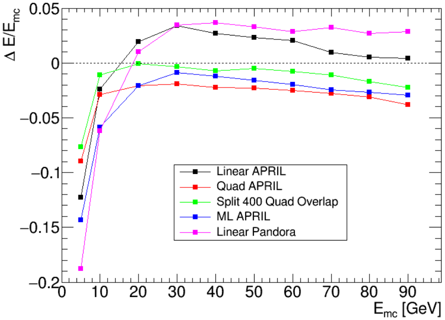
- Feed an ML algorithm (sklearn) with the events
- Force the algorithm to return a polynomial function with all possible combinations of terms up to a chosen degree : $N_1, N_1 N_2, N_1^2 ...$

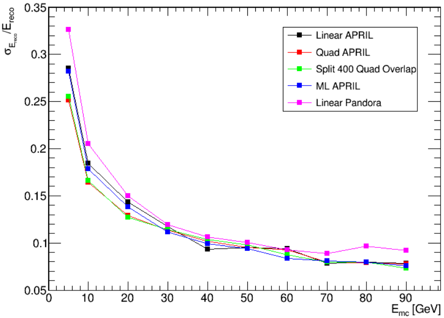
Split methods

- Difficulty to fit parameters that work for the whole energy range
- Tried to split the samples to have a formula for low N_{hit} and another for high N_{hit}

Comparison for the single KLongs

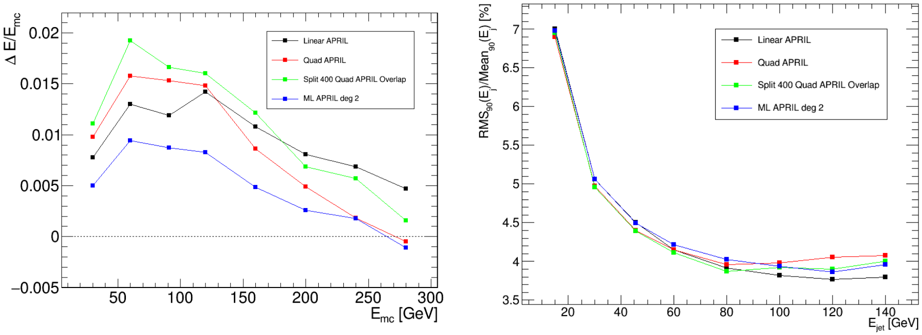
- Single KLongs with no angle
- Achieve good linearity and resolution with the different methods
- Split method is the best overall
- APRIL gives better results than Pandora on KLong





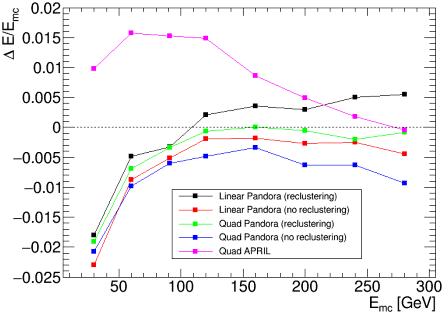
Comparison for dijets events

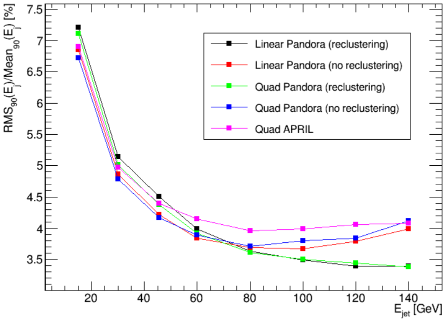
- Performance usually evaluated through mass reconstruction of dijets events (u, d, s) with $|\cos(\theta)| < 0.7$
- Use Mean_{90} and RMS_{90} on pfoEnergyTotal
- Good linearity for all methods : $< \pm 2\%$
- Need to fine tune the quadratic formula at high energy



Dijets comparison between APRIL and Pandora

- APRIL better than Pandora in resolution for $E_{jet} < 40$ GeV
- **Expected:** Reclustering improves resolution for $E_{jet} > 100$ GeV
- **Unexpected:** Reclustering impacts resolution negatively when $E_{jet} < 80$ GeV
- Hypothesis: Reclustering not optimized for semi-digital HCAL hits





Backup : Split method

- Change of regime between 20 GeV and 30 GeV $\rightarrow N_{hcal} \sim 400$
- Best results when changing formula at $N_{hcal} \sim 400$ with overlap of fitting range
  - Cluster with $N_{hcal} \leqslant 400$: Formula fitted on events with $N_{hcal} \leqslant 600$
  - Cluster with $N_{hcal} > 400$: Formula fitted on events with $N_{hcal} > 400$