An introduction to the particle flow algorithm for ILC, ArborPFA

Bo Li¹, Rémi Eté²

¹Institut de Physique Nucléaire de Lyon
²Deutsches Elektronen-Synchrotron

CHEF2017, Lyon
Outline

• Particle flow algorithm
• ArborPFA
• Optimization
• Conclusion
Particle flow algorithm

- Particle flow algorithm (PFA): an algorithm that tries to reconstruct individual final state particle from the record in detector.

- It can lead to measure jet energy accurately, and consequently improve the reconstruction and event identification.

- Jet energy resolution at ILC: $\sigma_E/E \lesssim 3.5\%$ from 50 to 500 GeV
ArborPFA

• The necessity of ArborPFA
  - Besides PandoraPFA, it is another code to validate and optimize the concept of particle flow algorithm for ILD.
  - It is also expected to implement some new ideas for PFA in this algorithm, such as timing and tracking.
  - MIP scale in Pandora is used in the algorithms which is not valid for SDHCAL.

• Brief history
  - Rémi built ArborPFA for SDHCAL prototype study,
  - then for ILD detector model (ILD_o2_v05); PandoraSDK as the framework. (arXiv:1506.05348)
  - For ILC physics research, the code is under optimization to improve jet energy resolution.
The algorithm of Arbor

- Arbor is an algorithm for clustering the hits in calorimeter based on tree topology (H. Videau, M. Ruan, arXiv:1403.4784)

- Vertex: hit in calorimeter space
- Connector: oriented link between two vertices
- Tree: the structure of linked vertices by connectors
  - the root vertex has no predecessor
  - all other vertices have only one predecessor
Hit clustering in ArborPFA

- Nearby hits connection
Hit clustering in ArborPFA (cont.)

- Nearby hits connection
- Reference vector

\[-\vec{C}_{ref} = w_{bck} \cdot \sum_b \vec{c}_b + w_{fwd} \cdot \sum_f \vec{c}_f\]
• Nearby hits connection

• Reference vector

\[ -\vec{C}_{\text{ref}} = w_{bck} \cdot \sum_b \vec{c}_b + w_{fwd} \cdot \sum_f \vec{c}_f \]

• The most probable connection is determined by order

\[ \kappa = \left( \frac{\theta}{\pi} \right)^{p_\theta} \cdot \left( \frac{\Delta}{\Delta_{\text{max}}} \right)^{p_\Delta} \]

\[ (p_\theta = 1, p_\Delta = 5) \]
Hit clustering in ArborPFA (cont.)

• Nearby hits connection

• Reference vector

\[-\vec{C}_{\text{ref}} = w_{\text{bck}} \cdot \sum_b \vec{c}_b + w_{\text{fwd}} \cdot \sum_f \vec{c}_f\]

• The most probable connection is determined by order

\[\kappa = \left(\frac{\theta}{\pi}\right)^{p_{\theta}} \cdot \left(\frac{\Delta}{\Delta_{\text{max}}}\right)^{p_{\Delta}}\]

\[(p_{\theta} = 1, p_{\Delta} = 5)\]

• Connector cleaning
Photon reconstruction

- Preparation of the event
  - CaloHits, tracks
  - Identification of $V_0$, kink and prong

- Reconstruction of photons
  - Clustering in ECAL
  - Remove hits from nearby track
Clustering of particles

- Track driven clustering

- Topological association for cluster fragments
Reclustering

• Reclustering is to correct cluster assignment based on the energy of track in cases of:
  - excessing energy
  - missing energy
  - track-cluster multiple association

• Create and identify the particle object

But, this is not real life…
Performance

- Performances of ArborPFA
  - non-linearity: 1 ~ 2%
  - JER: 5~7%

To make ArborPFA useful for ILC, we must improve the energy resolution.

But how?
Reclustering for missing energy

Maybe we can have a look at an event display…

- The photon cluster maybe should also be considered in the reclustering.
- It will be safe if we have a sophisticated cluster identification function.
- The reclustering process will go until no neutral cluster can be merged into the original cluster.
Neutral cluster identification

- The distribution of cluster property variables at 2 GeV

- The photon efficiency of identification for photon cluster and neutral hadronic cluster
After correction

- Reclustering of energy excessing also have similar issues. The main problem for jet energy resolution is on reclustering.
Current performance

- Problematic events (not all of them) have been fixed at 200 GeV

- The improvement in total energy range is still not enough, however,
  - JER at 200 GeV is close to the requirement at ILC.
  - The direction to optimize code seems reasonable.
Energy reconstruction

- The energy formulae of shower in SDHCAL

\[
E_{\text{quad}} = (\alpha_1 + \alpha_2 N_{\text{tot}} + \alpha_3 N_{\text{tot}}^2)N_1 \\
E_{\text{lin}} = \beta N_1 + \gamma N_3 \\
+ (\beta_1 + \beta_2 N_{\text{tot}} + \beta_3 N_{\text{tot}}^2)N_2 \\
+ (\gamma_1 + \gamma_2 N_{\text{tot}} + \gamma_3 N_{\text{tot}}^2)N_3
\]

- To apply the quadratic formula in PFA, we also need to investigate the performance for
  - nearby showers
  - shower in ECAL + HCAL

ILD_I4_v02 Pi+
Conclusion

• ArborPFA is created for the studies on prototype and ILD.
• The performance on jet energy resolution is improved, but we have still lots of issues to check and fix.
• ArborPFA is maintainable
  - using nice framework
  - code in high quality

Thanks for your attention!