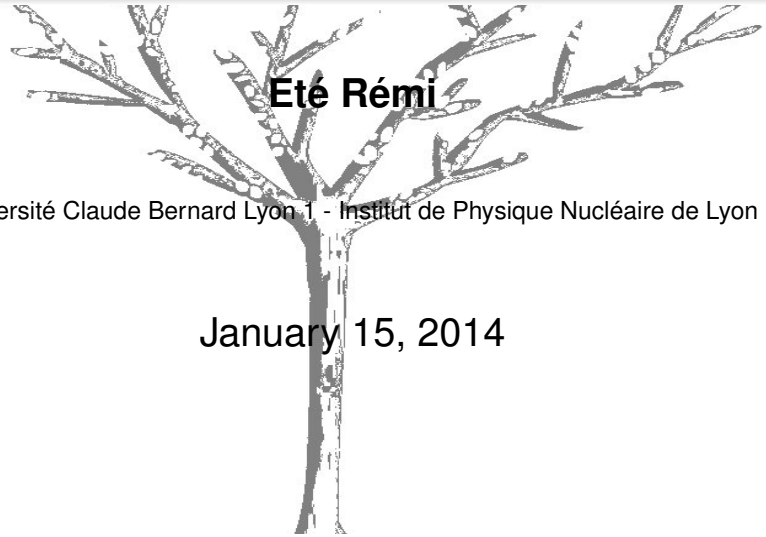
Arbor PFA approach
Single particle and separation of overlaid shower studies

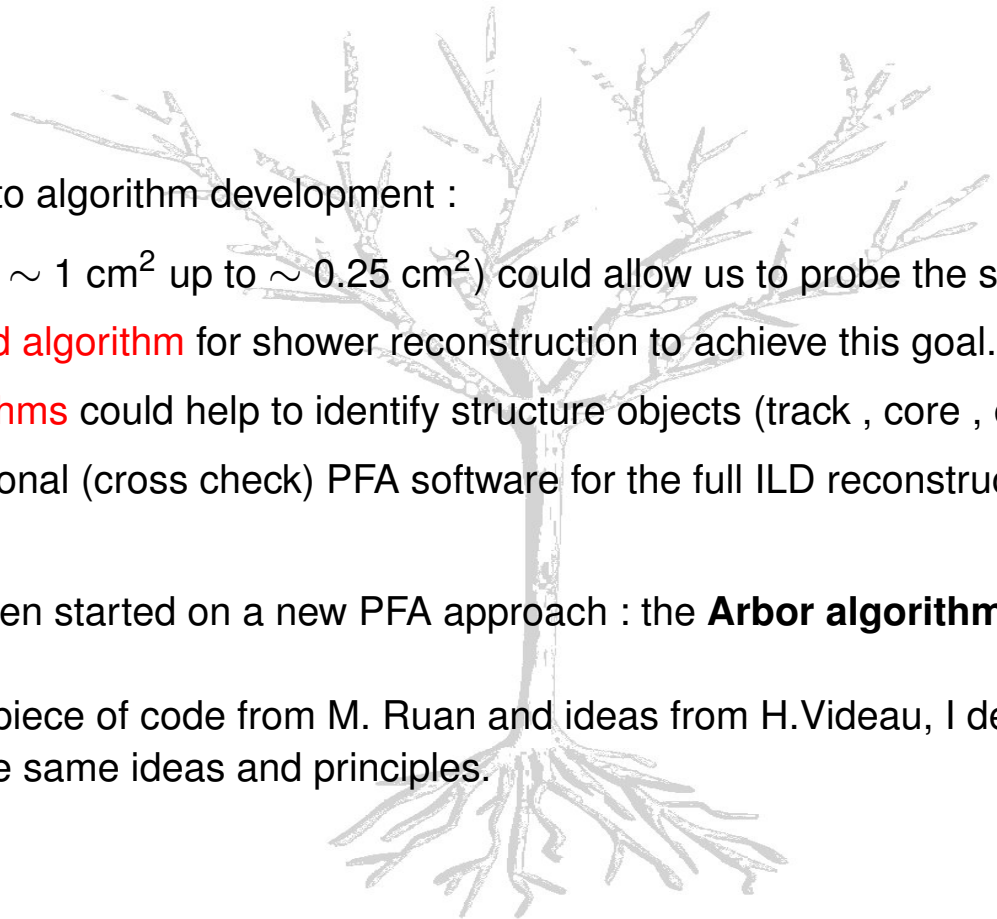


Eté Rémi

Université Claude Bernard Lyon 1 - Institut de Physique Nucléaire de Lyon

January 15, 2014

Motivations



Multiple reasons lead to algorithm development :

- Fine granularity ($\sim$ 1 cm^2 up to $\sim$ 0.25 cm^2) could allow us to probe the shower structure.
- Need a dedicated algorithm for shower reconstruction to achieve this goal.
- Additional algorithms could help to identify structure objects (track , core , edge , isolated hits, ...).
- Provide an additional (cross check) PFA software for the full ILD reconstruction.

A work has already been started on a new PFA approach : the **Arbor algorithm**.

Based on the original piece of code from M. Ruan and ideas from H.Videau, I developed an *Arbor-like* PFA approach based on the same ideas and principles.

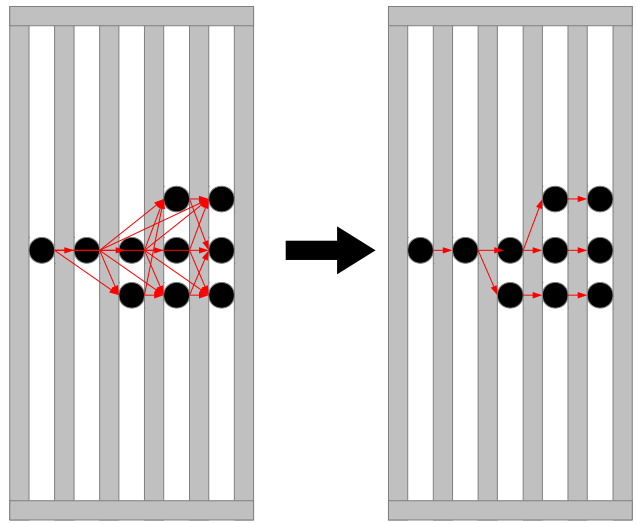
The Arbor Algorithm - ideas and principles

The Arbor algorithm is based on simple ideas.

It aims to connect objects between each others in order to cluster them.
The algorithm is run in multiple step in the following way :

- Arbor objects are created from input data (single calo hit, small clusters, track projection)
- For each object, initial connections are made in a region of interest (ROI), that is to say, we create connection from this objects to multiple other ones.
- An iteration is made on all objects and only one connector is kept in the backward direction.
- Clusters are created from a group of connected objects.

Remark : by keeping only one backward connector, we ensure that the solution for the branching is unique and all the objects are coming from one point (initial root) $\rightarrow$ **Arbor principle !**.



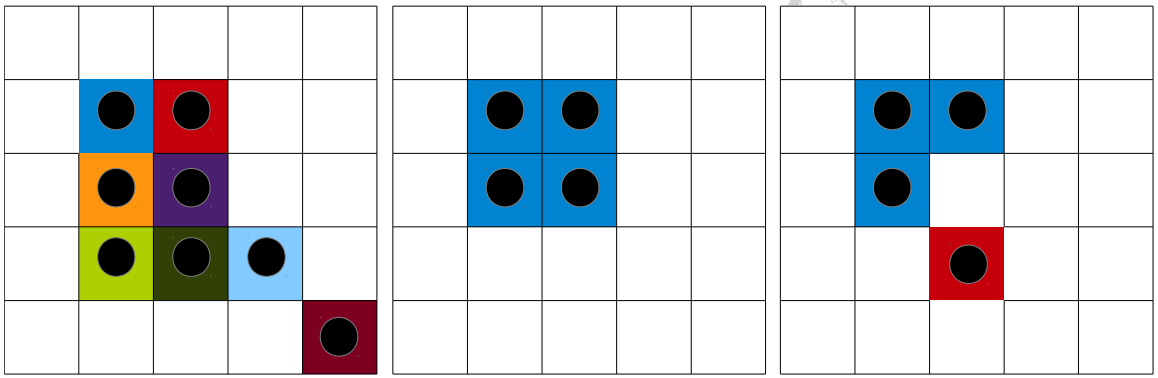
Arbor algorithm : more details

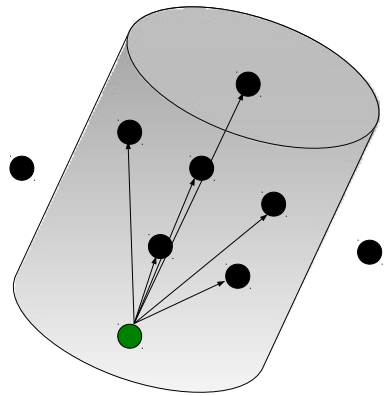
Object creation

An intra-layer clustering is first of all done. Clusters with a maximum size of $p_1 = 4$ are considered as input objects. Else, each calo hit is also considered as an input object for connection. This is time gaining for the multiple iterations and remove the dependency of GRPC multiplicity.

Initial connectors

Initial connection are made by looking at a maximum distance of $p_2 = 85.0$ mm in the forward direction (~ 3 layers) and $p_3 = 41.0$ mm in the transverse direction (4 cells). The cylinder axis used for these initials connections is the $\vec{r}$ vector (position of the object). Isolated objects can't creates forward connectors but can be connected in the backward direction.





Arbor algorithm : more details

Connector cleaning

For each object and its given list of backward connectors, we compute the mean backward direction. For each of these connectors, the order parameter :

$$\kappa = \theta^{p_4} . \Delta^{p_5}$$

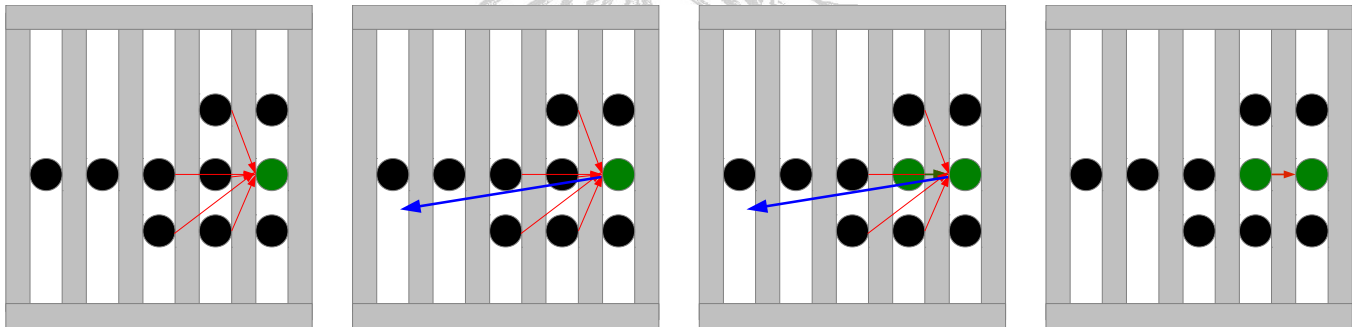
is computed and the connector that has the smallest κ one among the others is kept. ($p_4 = 1$ and $p_5 = 5$)

Small cluster merging

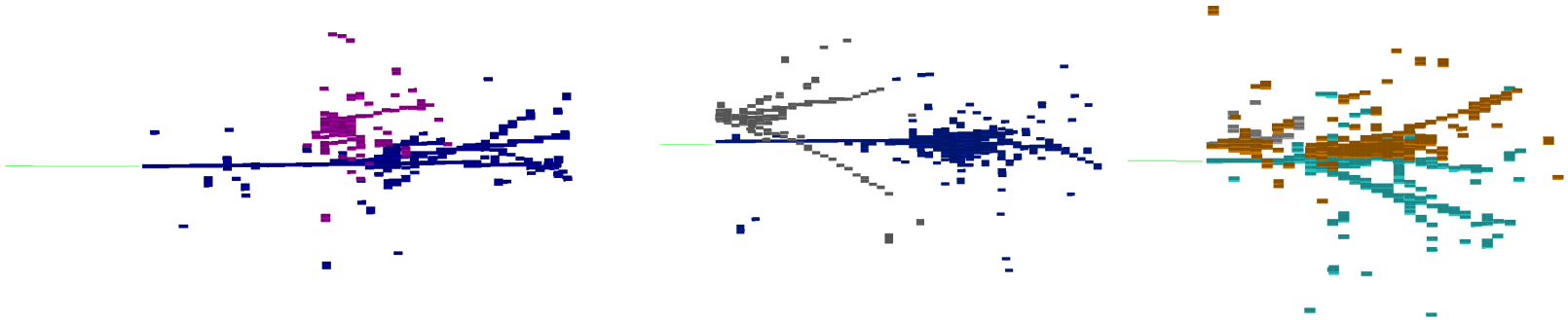
Cluster that have a size smaller than $p_6 = 15$ calo hits are merged in the closest one (closest calo hit distance).

Remark :

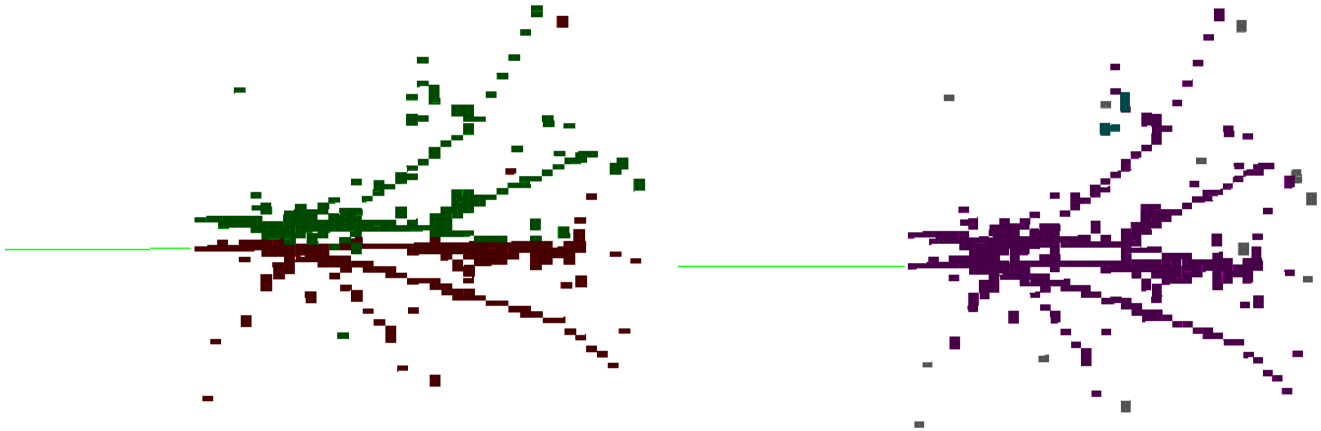
- All track-cluster associations are made with the same branching principle by considering track projection as an object to be connected only in the forward direction.
- A cut of 2 GeV is done while creating Particle Flow Objects.



Some event displays



10 GeV neutral/ 30 GeV charged - 5 cm - ArborPFA



ArborPFA

PandoraPFA

Data set

Both Geant4 simulation and SPS 2012 (SEPT - AUG) test beam data of SDHCAL prototype are used in these studies.
Latest SDHCAL digitizer version from Arnaud (see A. Steen talk on Friday) is used to produce calorimeter hits (includes inefficiency !).
Two studies have been investigated in order to check the validity and the efficiency of the algorithm.

- Single particle study : PFO multiplicity, reconstructed VS estimated energy comparison, missing energy identify as neutral particles, efficiency ...
- Overlaid particles (a 10 GeV neutral hadron and a 10 (or 30) GeV charged hadron) study : PFO multiplicity, reconstructed VS energy comparison for the neutral particle and the charged one, efficiency ...

For both cases, the ArborPFA processor and the PandoraPFA processor are run in order to compare them.
The PandoraPFA processor consists in the full reconstruction chain (60 reclustering included) without the muon and photon clustering since there is only SDHCAL. All algorithm parameters are taken from DBD ILD_01_vO5 settings.

The ArborPFA plug-in

The ArborPFA is implemented using PandoraSDK C++ package
The GIT repository is available at https ://github.com/rete/ArborPFA.git
All of these results can be reproduced using commit *d921c70e8137fe126f698af18985de77803894b1*

PFO Analysis

Definitions

Some definitions :

- Input energy : i.e energy of the particle thrown (10 GeV, 20 GeV, etc ...)
- Estimated energy : total energy coming from one particle computed using

$$E_{meas} = \alpha N_1 + \beta N_2 + \gamma N_3$$

with

$$\alpha = \alpha_1 + \alpha_2 . N_h + \alpha_4 . N_h^2 \quad , \quad \beta = \beta_1 + \beta_2 . N_h + \beta_3 . N_h^2 \quad , \quad \gamma = \gamma_1 + \gamma_2 . N_h + \gamma_3 . N_h^2$$

- Reconstructed energy : Energy of the PFO clusters estimated with the previous formula

For a perfect clustering algorithm the estimated energy and the reconstructed energy are equal. In order to compare the ArborPFA and PandoraPFA, both for TB data and simulation, we look at the difference between those two energies.

For the single particle study we define the efficiency as :

$$e_{single} = (nPFOs == 1) \; and \; (nChargedParticles == 1)$$

For the overlaid showers study, we define two efficiencies :

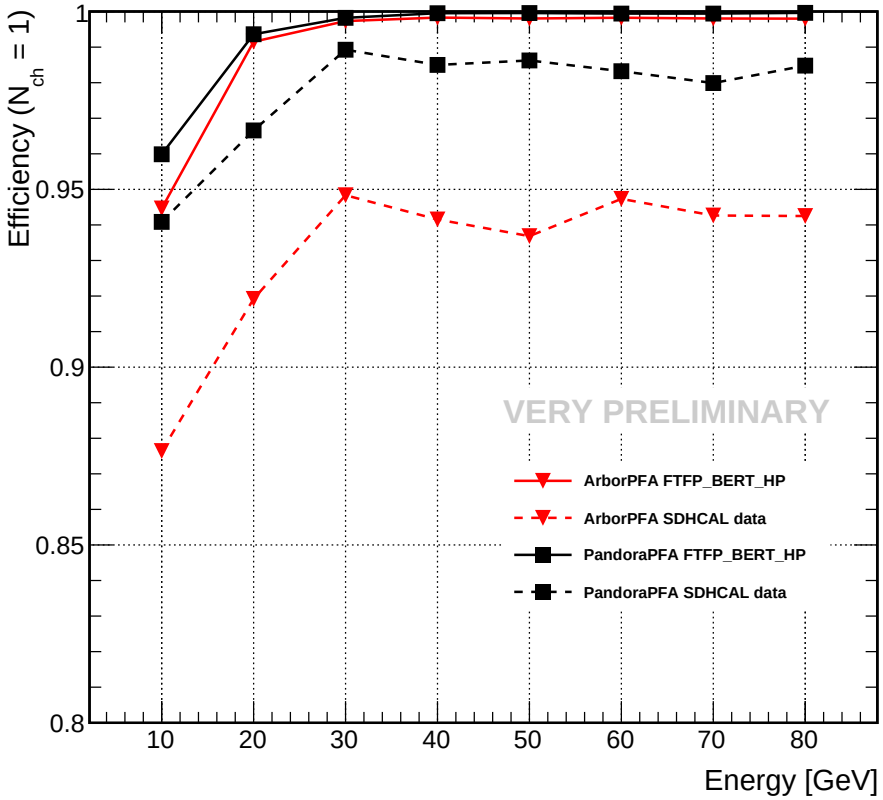
$$e_{overlay1} = (nChargedParticles == 1) \; and \; (nbOfNeutral == 1)$$

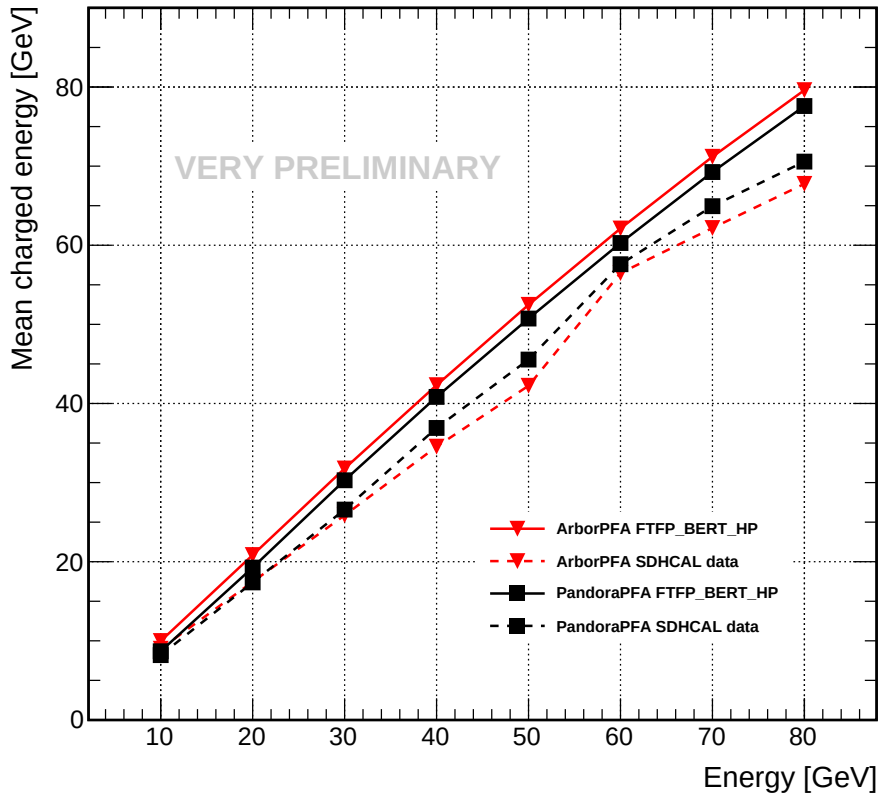
and

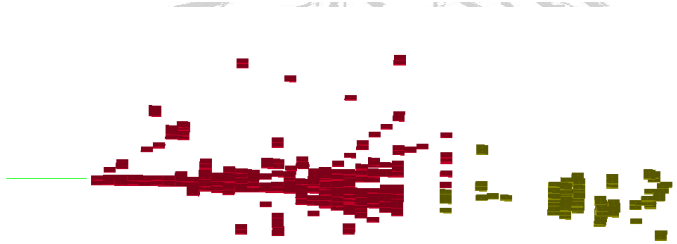
$$e_{overlay2} = (nChargedParticles == 1) \; and \; (nbOfNeutral \geq 1)$$

PFO Analysis

Single particle study

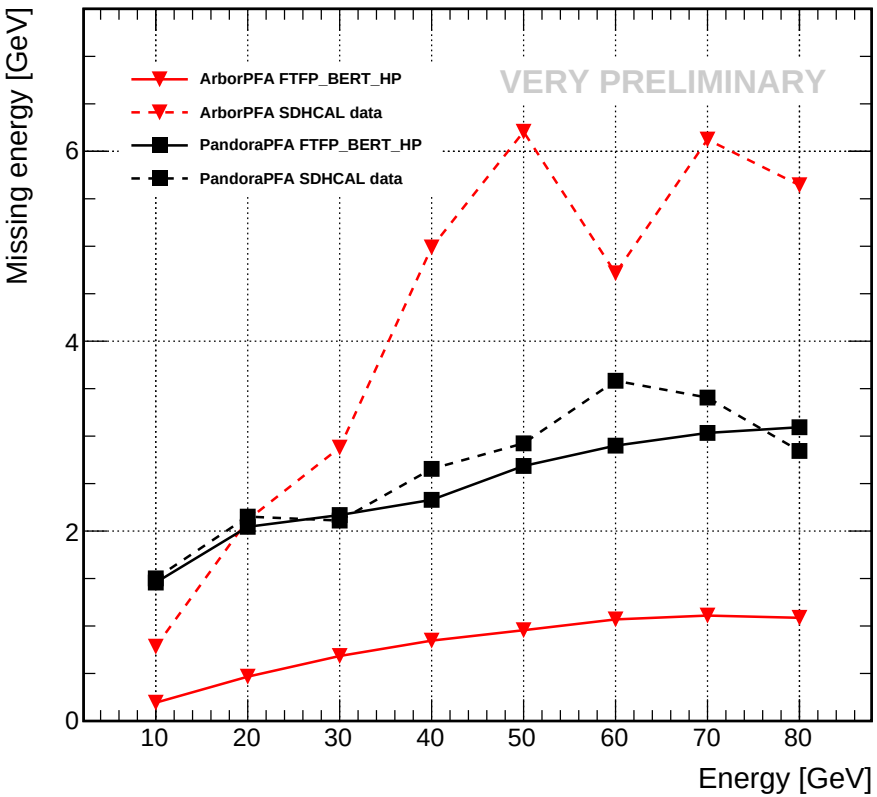


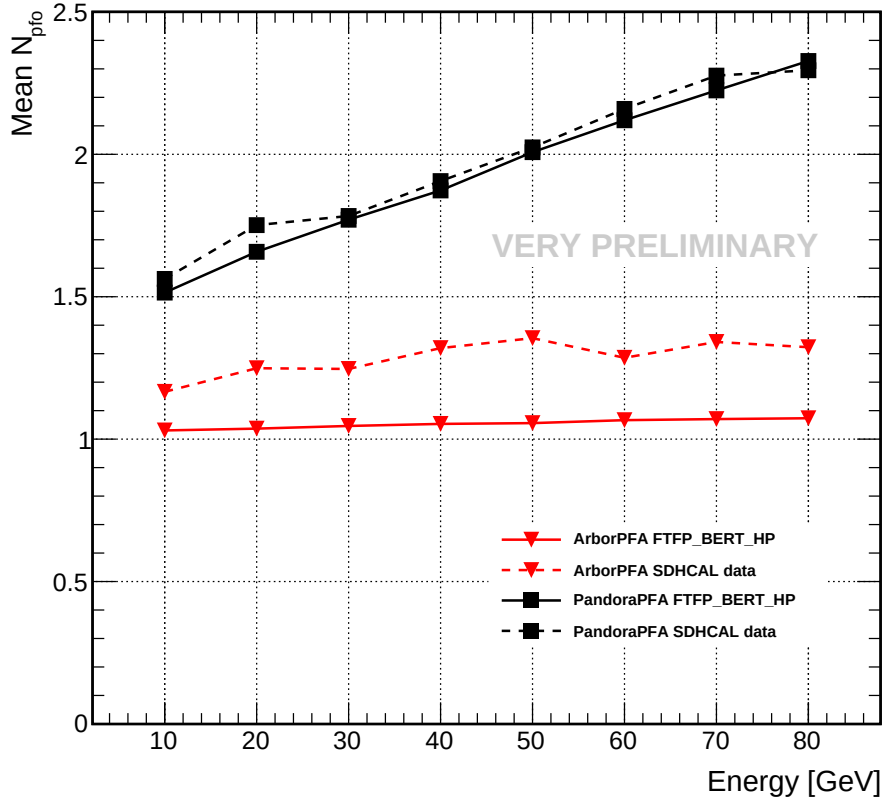




PFO Analysis

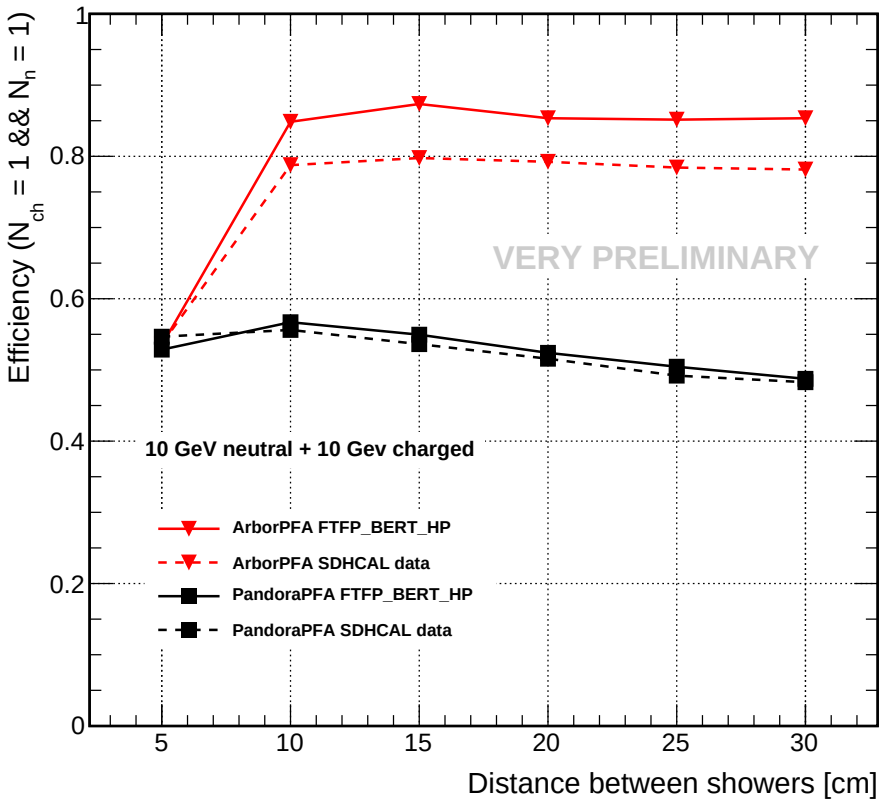
Single particle study

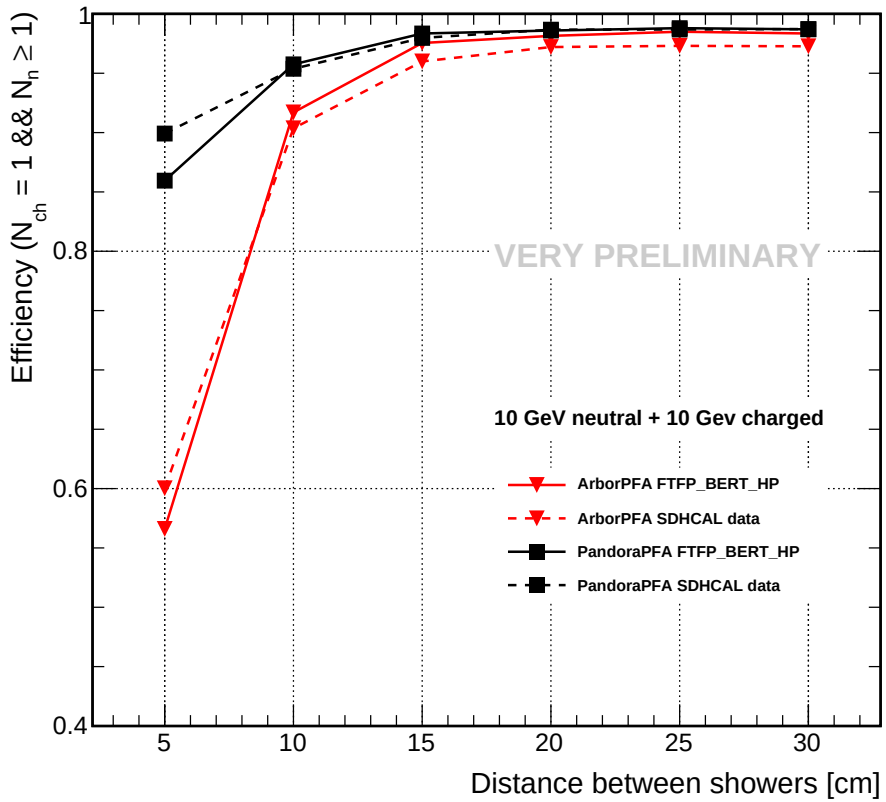




PFO Analysis

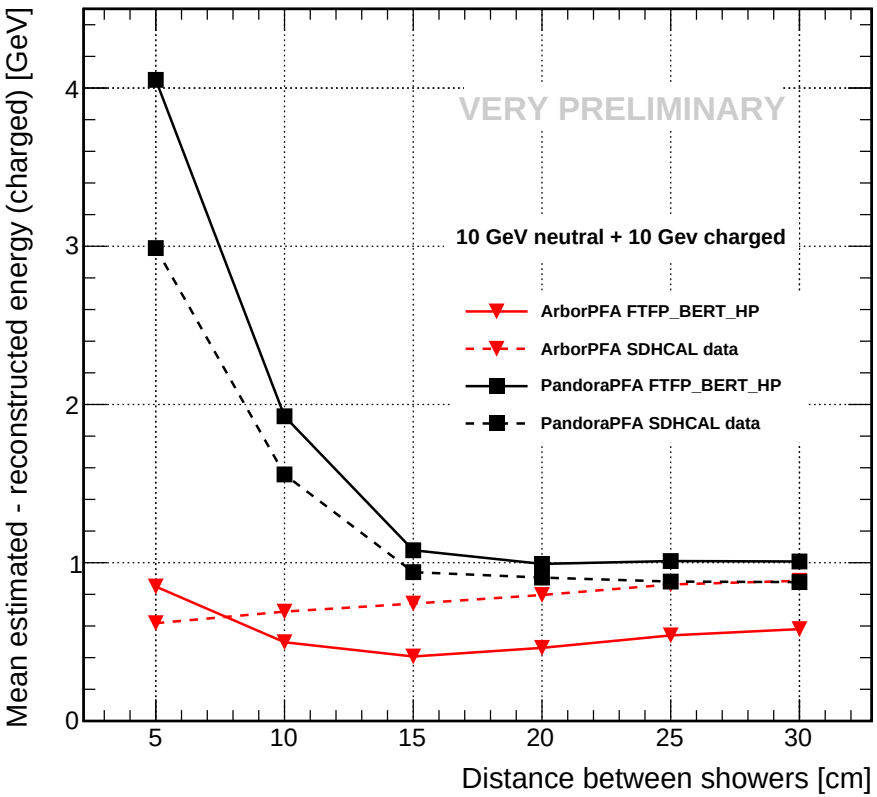
Overlaid Showers study - 10 GeV neutral + 10 GeV charged

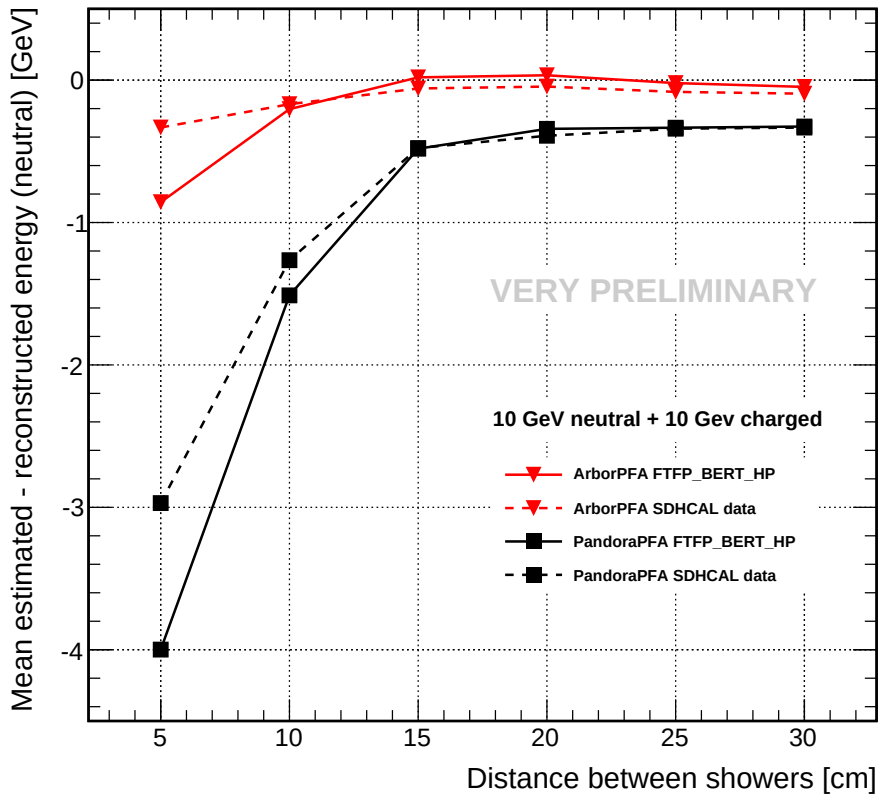




PFO Analysis

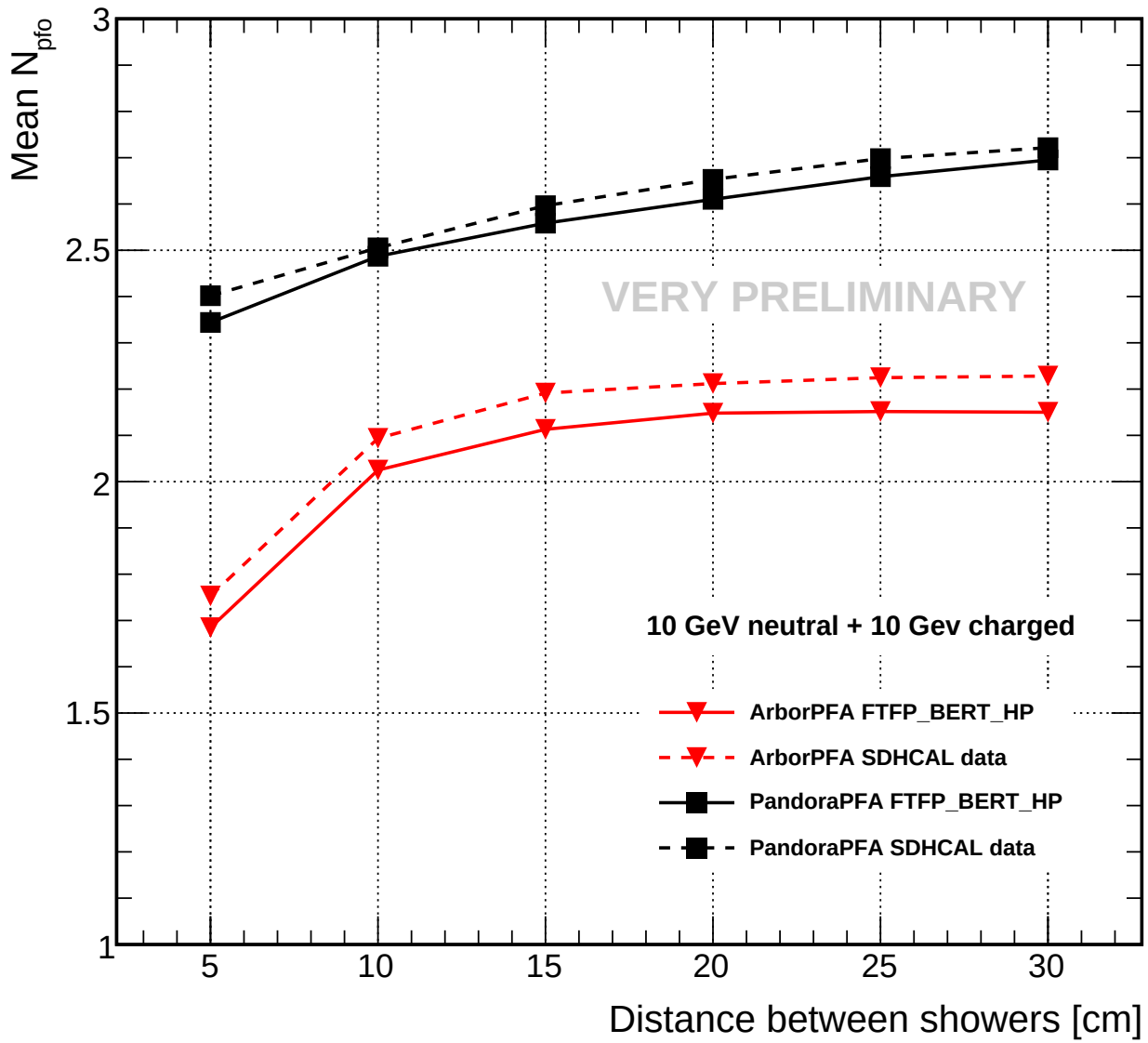
Overlaid Showers study - 10 GeV neutral + 10 GeV charged





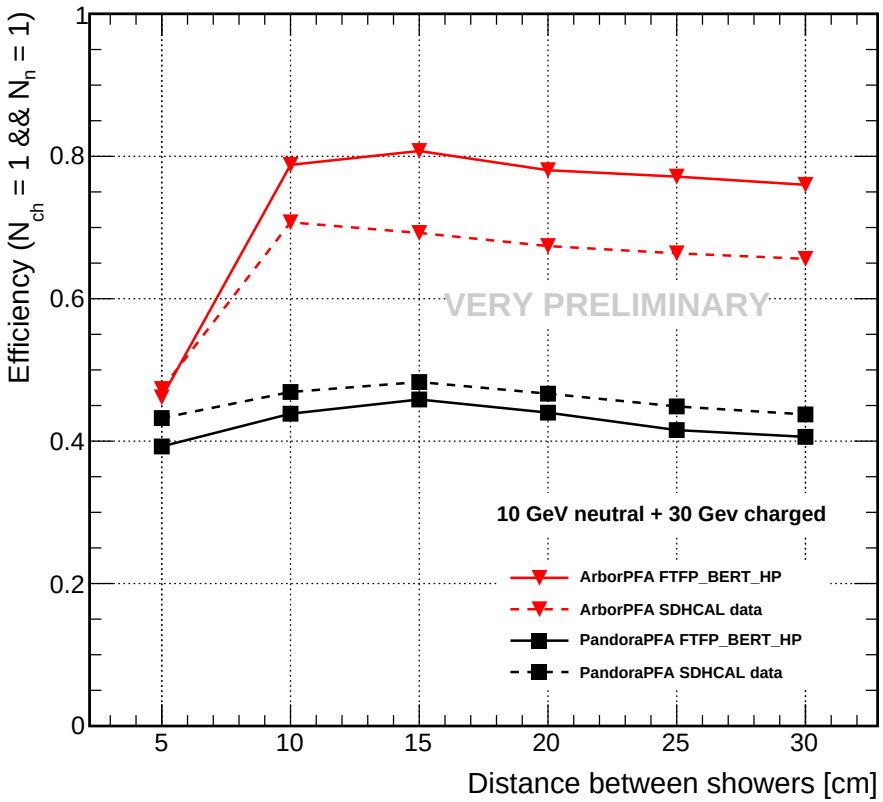
PFO Analysis

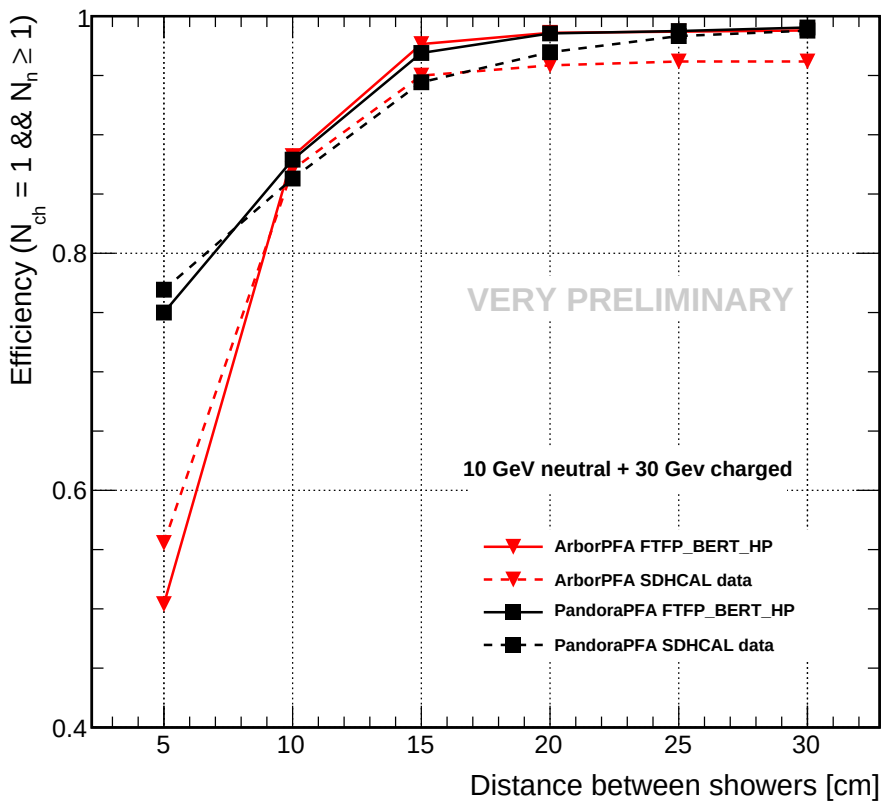
Overlaid Showers study - 10 GeV neutral + 10 GeV charged



PFO Analysis

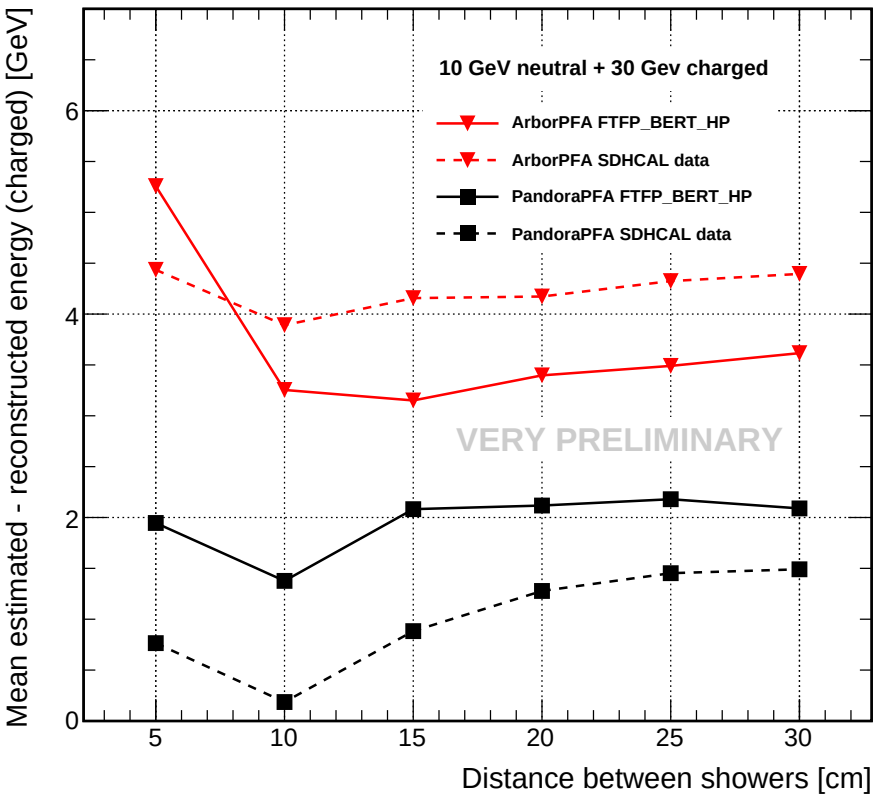
Overlaid Showers study - 10 GeV neutral + 30 GeV charged

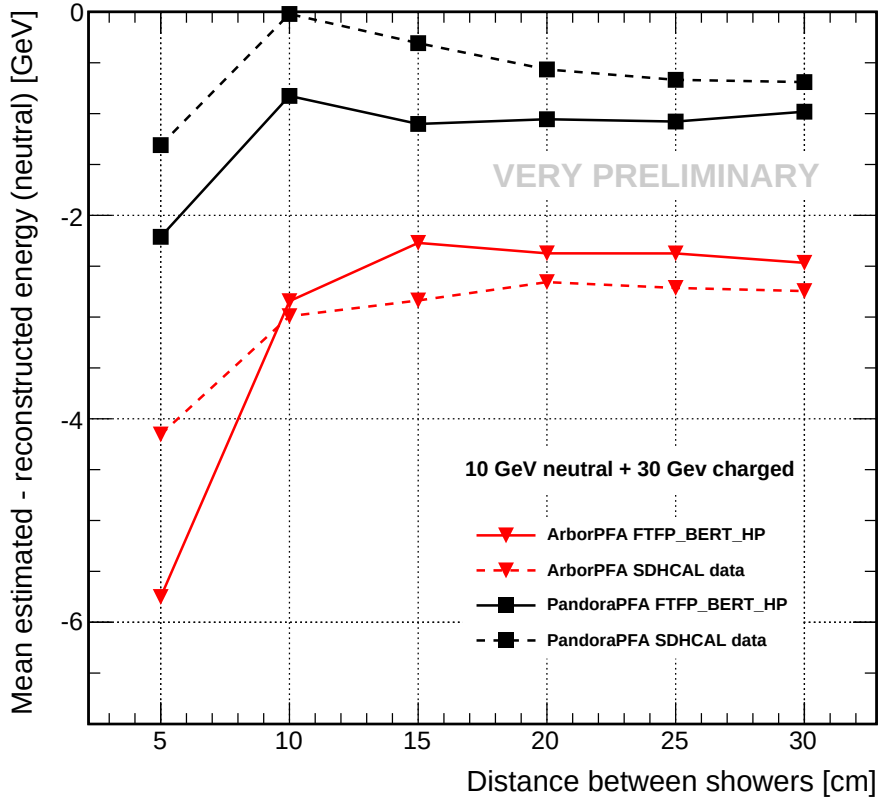




PFO Analysis

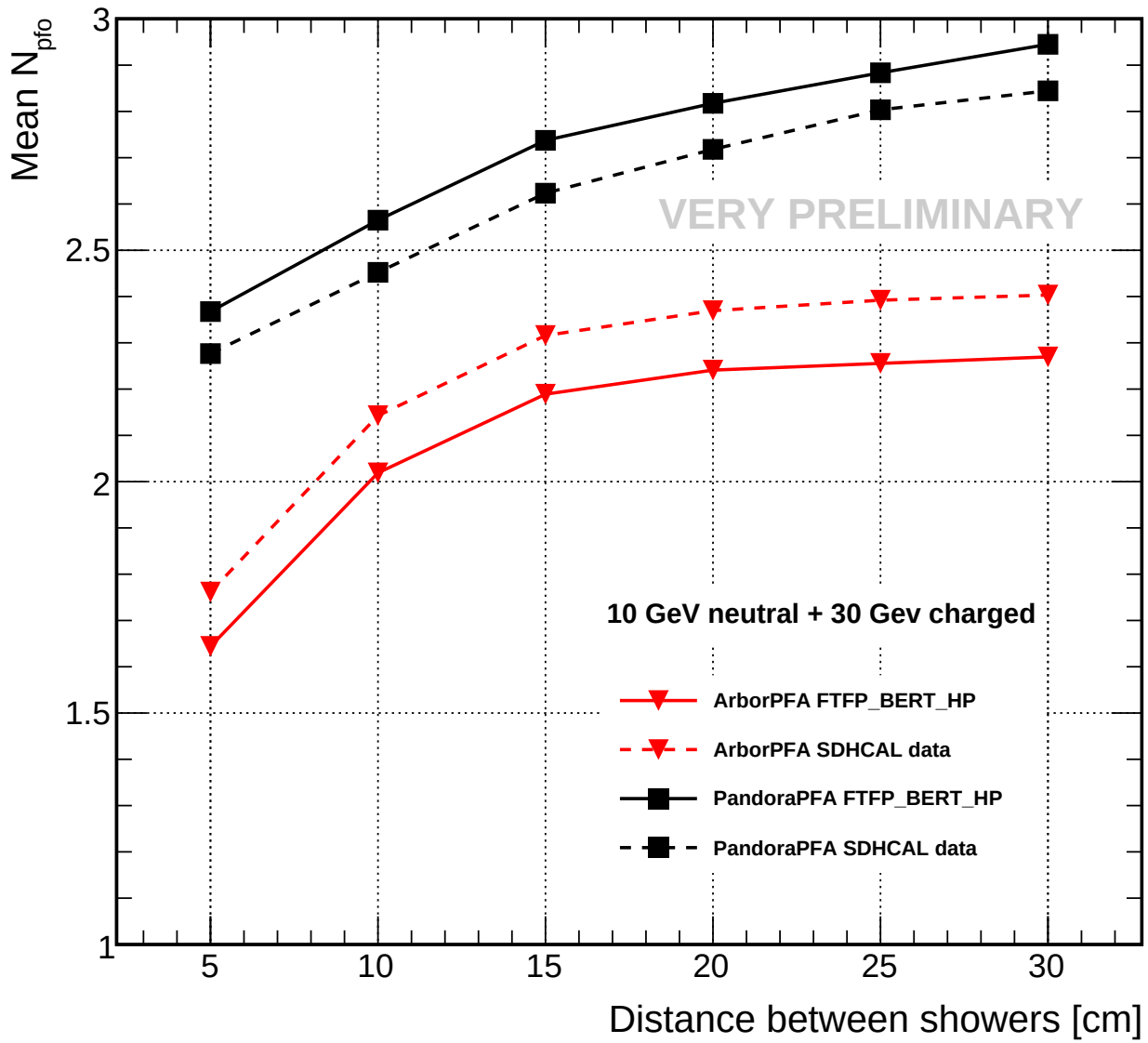
Overlaid Showers study - 10 GeV neutral + 30 GeV charged





PFO Analysis

Overlaid Showers study - 10 GeV neutral + 30 GeV charged



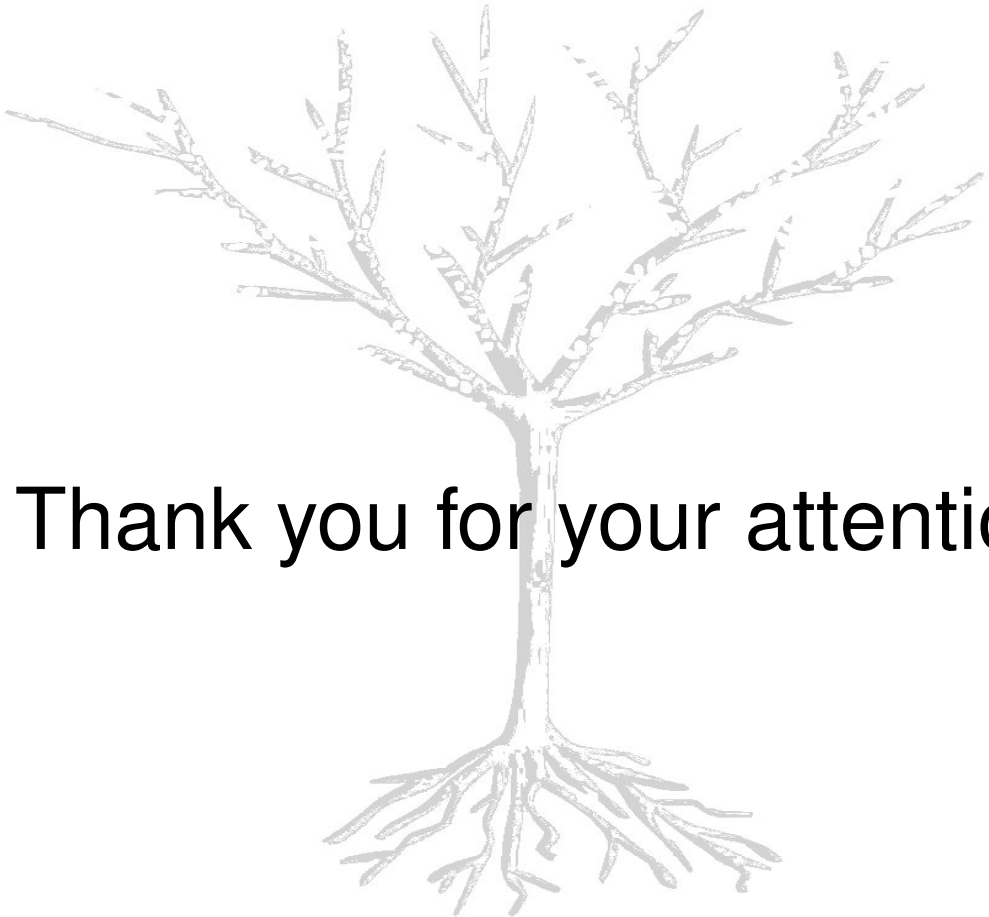
Conclusions and plans

Conclusions :

- Arbor branching decision is an ongoing work and has to be improved.
- PandoraPFA shows a bad efficiency in terms of PFO multiplicity but still has better energy performance.
- ArborPFA is more efficient when particle energies are similar for separation.

Plans :

- Arbor branching decision should be improved by adding a second connector cleaning with a track to cluster comparison (energy flow).
- Arbor algorithm has to be fine tuned. Maybe needs a reclustering procedure by looping over interesting parameter as PandoraPFA does.
- Next step is to obtain better results than PandoraPFA with SDHCAL technological prototype and then move to ECAL+SDHCAL combined analysis, then obtain better results, then ILD, better results, then ... Jackpot !
- ArborPandora tool development : Additionnal framework classes (ArborApi, ArborContentApi, ...), daughter algorithms dedicated to Arbor (multiple object creation algorithms, multiple connector cleaning algorithms, connector clustering), global tools for branches and connectors.



Thank you for your attention !