k4Clue: Empowering Future Collider Experiments with CLUE

<u>Erica Brondolin</u>,
Felice Pantaleo, Marco Rovere

2023 International Workshop on Future Linear Colliders, 17th May 2023

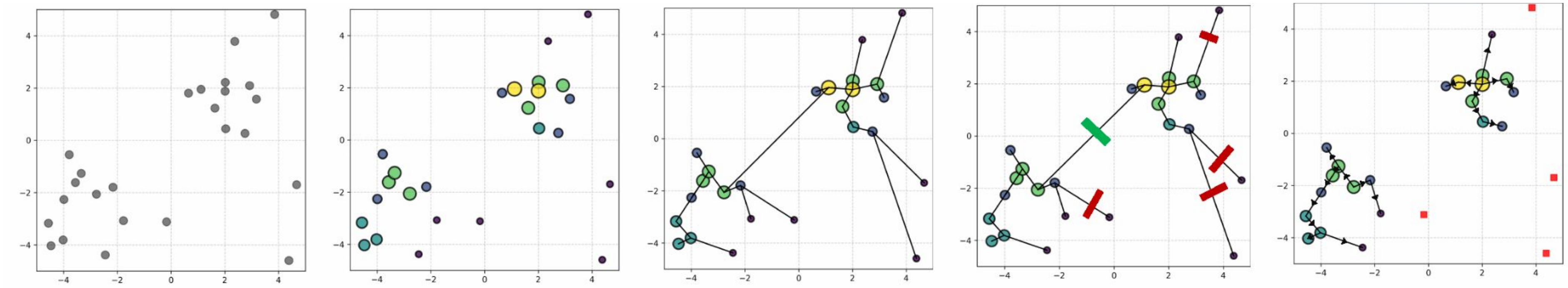
Introduction

CLUstering of Energy

- CLUE (**CLUstering of Energy**) is a fast density-based clustering algorithm for the next generation of sampling calorimeter with high granularity in HEP

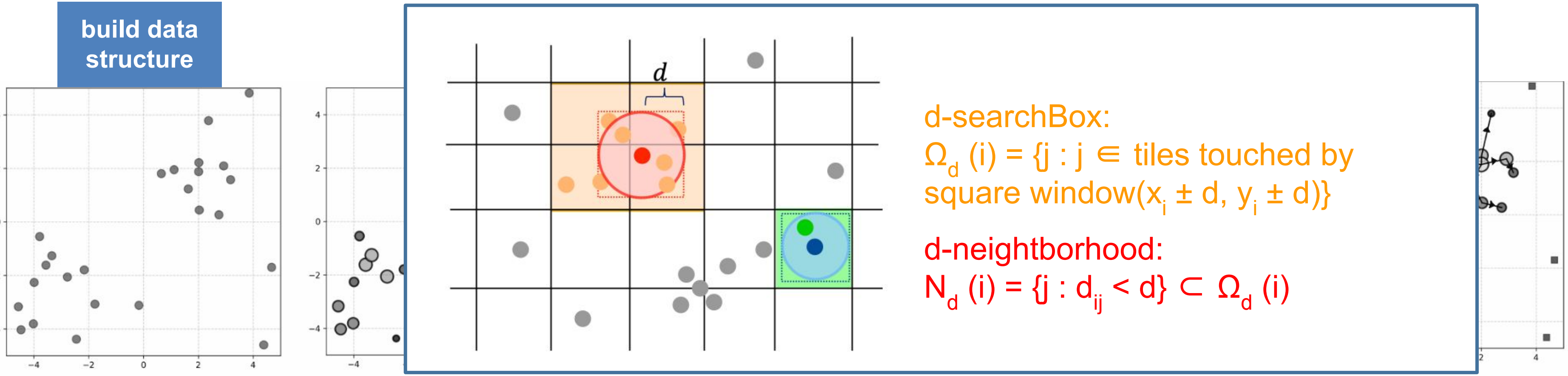
  [doi: 10.3389/fdata.2020.591315](https://doi.org/10.3389/fdata.2020.591315)

- It uses **energy density** - rather than individual cell energy - to establish **seeds, outliers, and followers** in 2D planes.
- **GPU-friendly**, i.e. suitable for the upcoming era of heterogeneous computing in HEP
- Standalone repo: CLUE - kalos gitlab.cern.ch



Step 1: Building Data Structure

- Querying neighborhood is a frequent operation in density-based clustering → **fast!**
- Build **Fixed-Grid Spatial Index** for hits on each layer:
  - Each tile in the grid hosts indices of hits inside it and has a fixed length of memory to store the hosted indices. It is independent by the detector granularity.
- To find the neighborhood hits $N_d(i)$ of *i*-hit, we only need to loop over hits in $\Omega_d(i)$



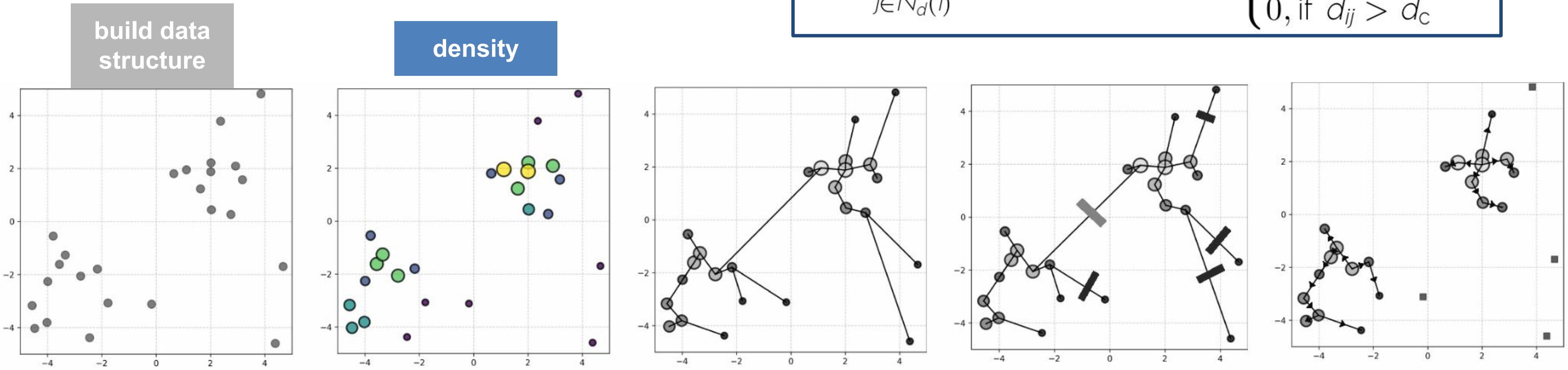
Step 2: Local energy density

- Calculate local energy density (ρ_i) in a distance (d_c)
  - Each hit *j* weighted by the deposited energy (E_j)
  - For each hit, calculate ρ_i

convolution kernel

k = 0.5

$$\rho_i = \sum_{j \in N_d(i)} E_j \times f(d_{ij})\,; \quad f(d_{ij}) = \begin{cases} 1, \text{if } i = j \\ k, \text{if } 0 < d_{ij} \le d_c \\ 0, \text{if } d_{ij} > d_c \end{cases}$$

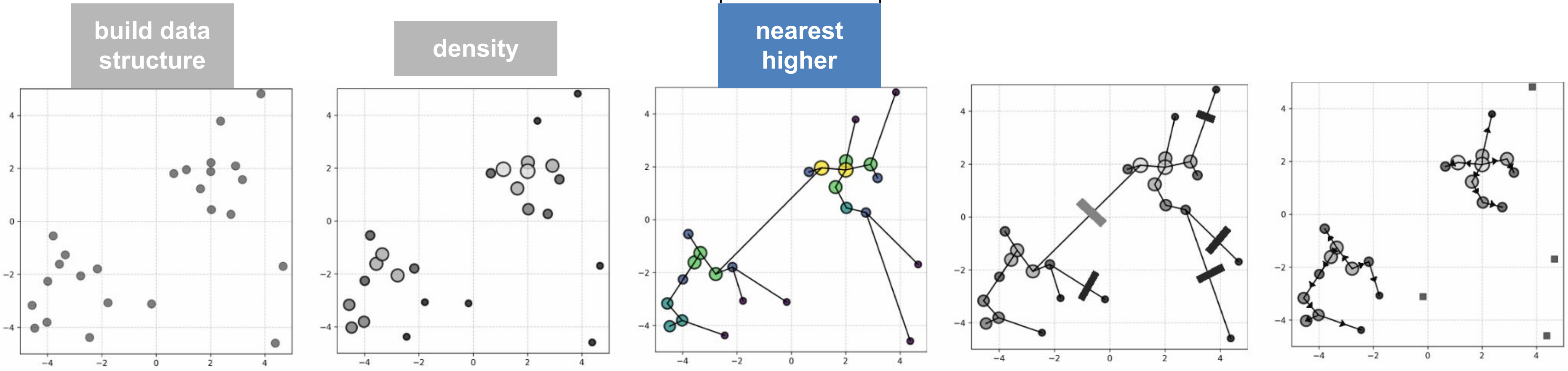


Step 3: Find "closest higher hit"

- Calculate "Nearest-Higher" hit within $N_{dm}(i)$
  - Define $d_m = o_f * d_c$
  - Find the closest hit with higher local energy density, nh_i

$$nh_i = \begin{cases} argmin_{j \in \hat{N}_{d_m}(i)} d_{ij}, \text{ if } |\hat{N}_{d_m}| \neq 0, \hat{N}_{d_m}(i) = \{j : j \in N_{d_m}(i), \rho_j > \rho_i\} \\ -1, \text{otherwise} \end{cases}$$

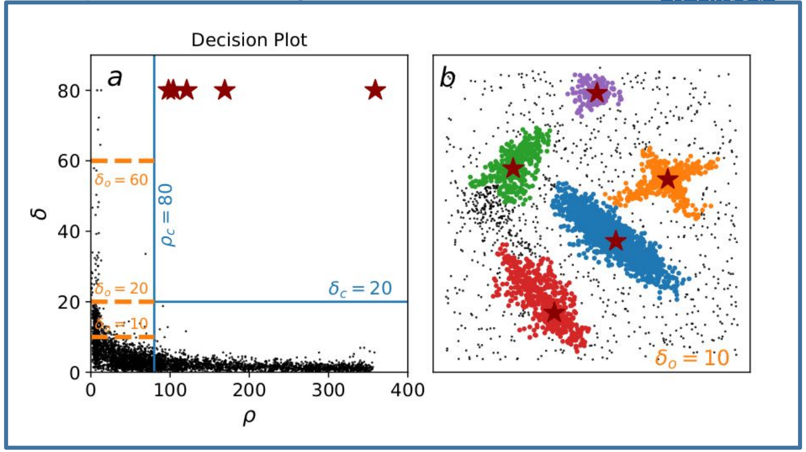
  - Calculate the separation distance $\delta_i = dist(i, nh_i)$

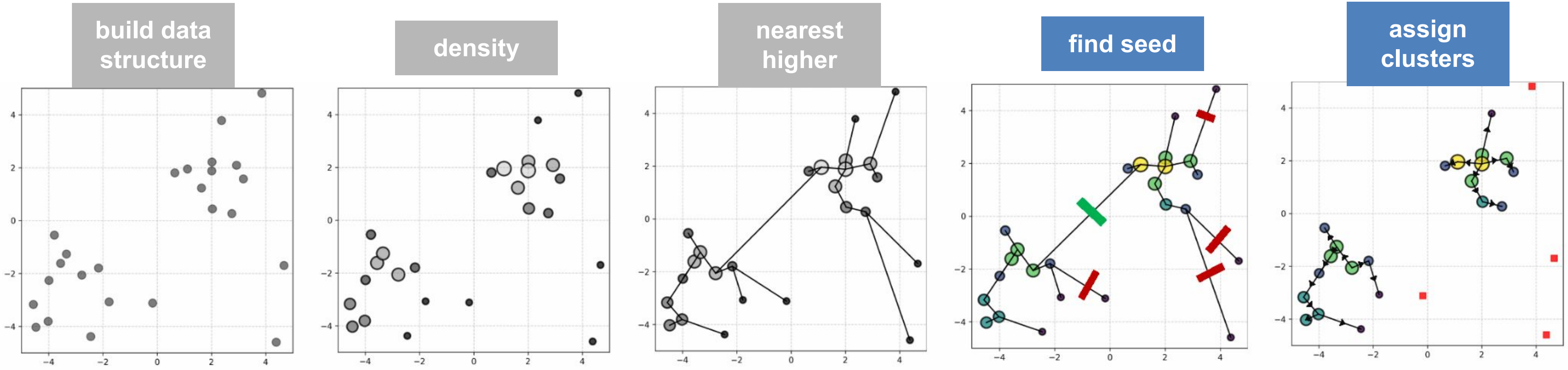


Step 4: Classify hits

- Promote as **seed** if $\rho_i > \rho_c$, $\delta_i > d_c$
- Demote as **outlier** if $\rho_i < \rho_c$, $\delta_i > o_f * d_c$
- Assign unique, progressive cluster ID to each cluster
  - **Followers** are defined and associated to their closest seed

Example of decision plot





Clustering procedure recap

Input parameters:

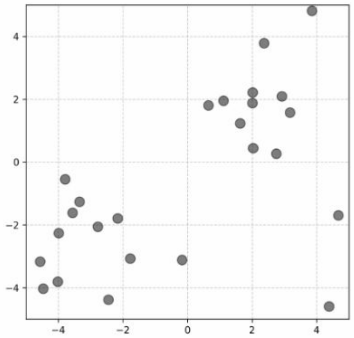
d_c : Critical Distance

o_f : Outlier Delta Factor

ρ_c: Minimum Local Density

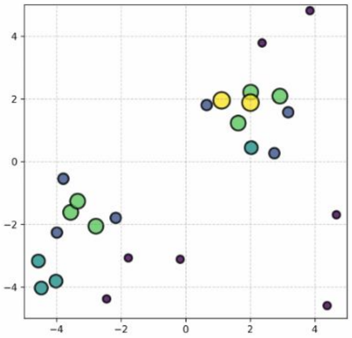
build data structure

Query the neighborhood of a point by looping over the points in N_d in the bins touched by the tiles intersected by d_c



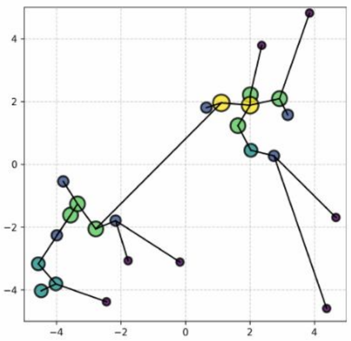
density

Hit position and energy used to calculate the hit's local energy density ρ_i and its distance δ_i to the nearest hit with higher local density



nearest higher

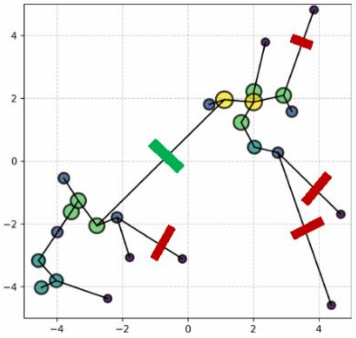
Define the nearest-higher of each hit as the hit with the local energy density higher then the hits itself and within a distance of $d_m = o_f \times d_c$



find seed

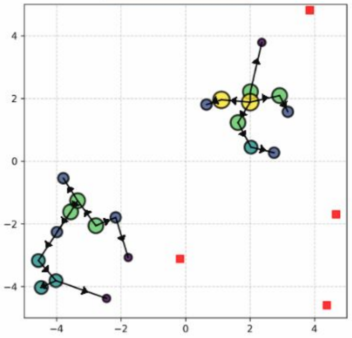
Use following criteria:

- seed: $\rho_i \geq \rho_c$ and $\delta_i \geq d_c$;
- outlier: $\rho_i < \rho_c$ and $\delta_i \geq (o_f \times d_c)$



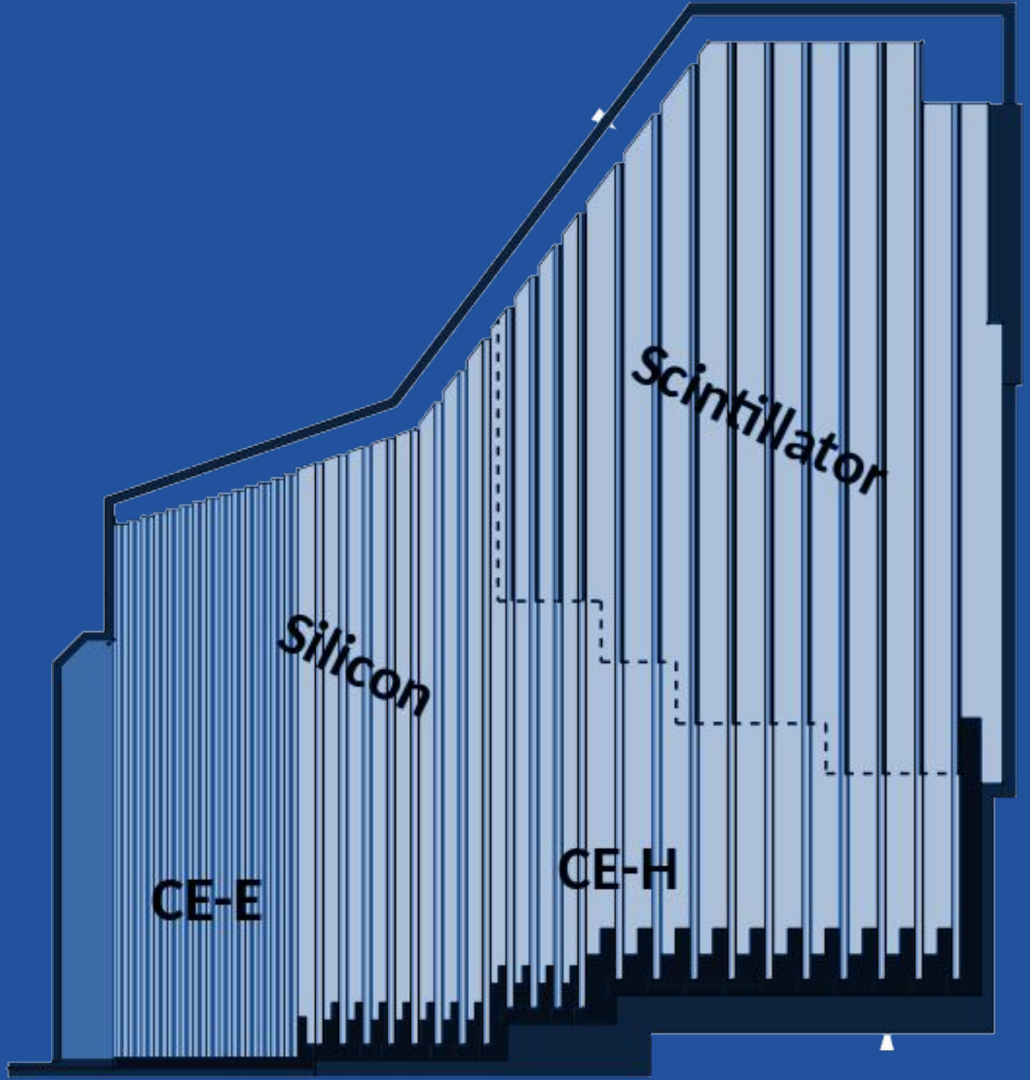
assign clusters

Register each remaining point as a follower to its nearest-higher



Erica Brondolin (erica.brondolin@cern.ch)

CLUE in the HGCAL reconstruction

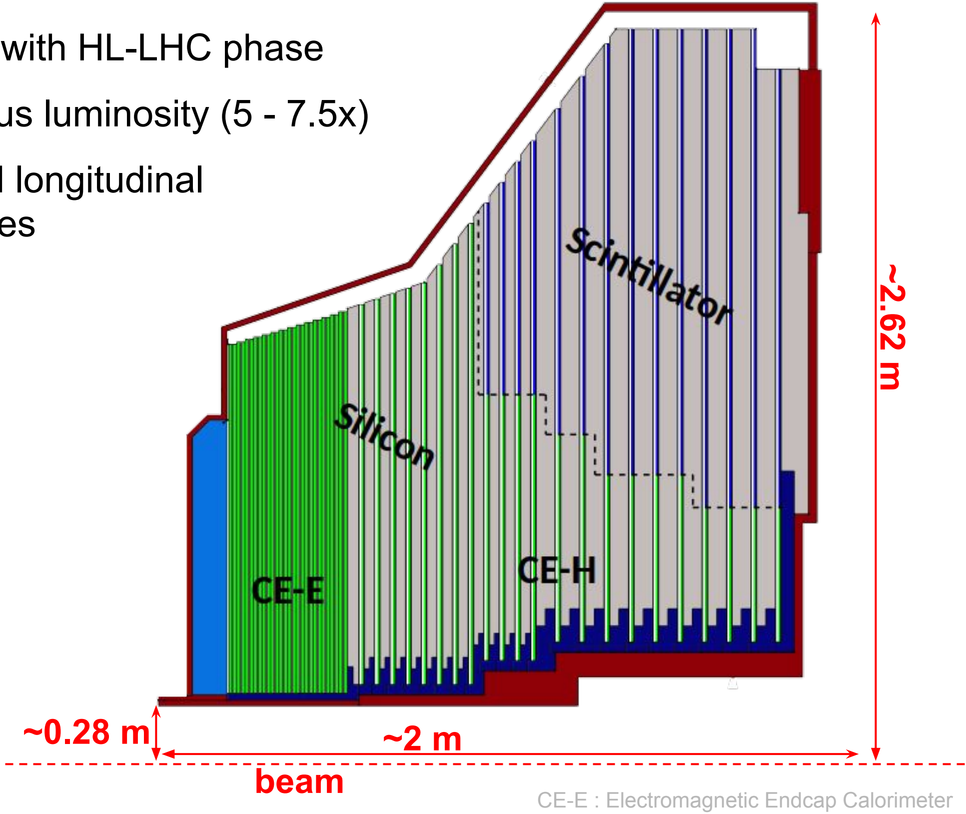


CMS High Granularity Calorimeter

- Phase-2 upgrade of CMS is needed to cope with HL-LHC phase
  - A significant increase in the instantaneous luminosity (5 - 7.5x)
- Imaging calorimeter with very fine lateral and longitudinal segmentation, and precision timing capabilities
  - Covering 1.5 < η < 3.0

Both endcaps	Silicon	Scintillators
Area	**~620 m²**	**~400 m²**
Channel size	0.5 - 1 cm^2	4 - 30 cm^2
#Modules	~30'000	~4'000
#Channels	**~6 M**	240 k
Op. temp.	-30 °C	-30 °C

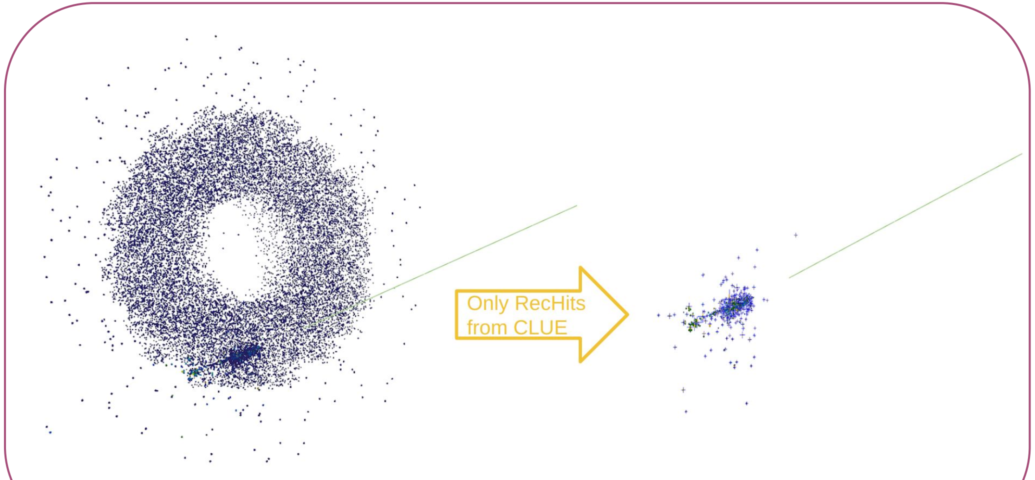
Ref.



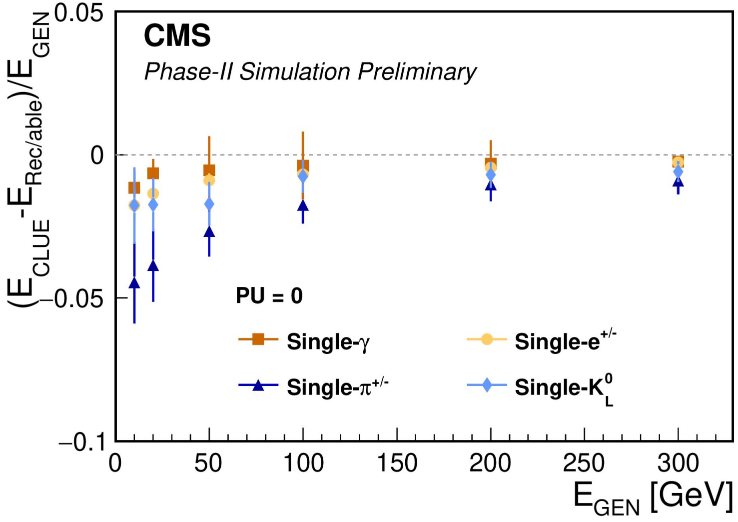
CE-E : Electromagnetic Endcap Calorimeter
CE-H : Hadronic Endcap Calorimeter

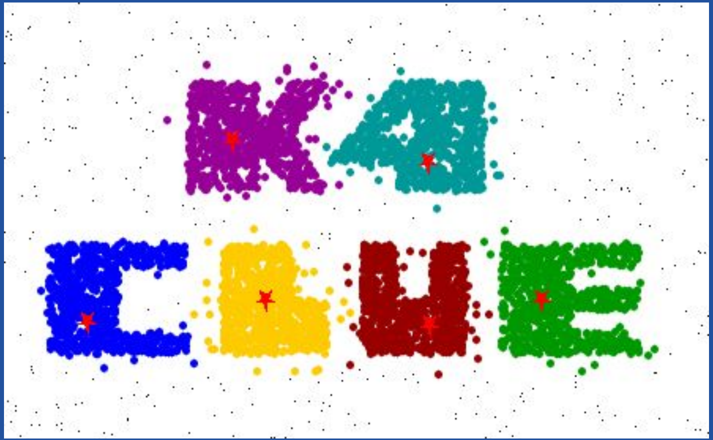
HGCAL Software Reconstruction

- The HGCAL reconstruction framework is **TICL (The Iterative Clustering)**
- It starts by calibrating deposited energy in individual cells, also called RecHits → an order of **10^5 RecHits** in the HGCAL detector for events @ 200 pileup
- CLUE clusters the RecHits in the same layer to produce **Layer Clusters (LCs)**



Single pion with E_{GEN} = 100 GeV with electronic noise enabled (no pileup)



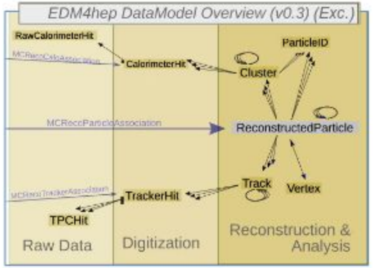
The  package

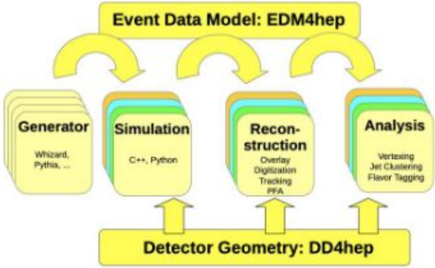
Key4hep in a nutshell

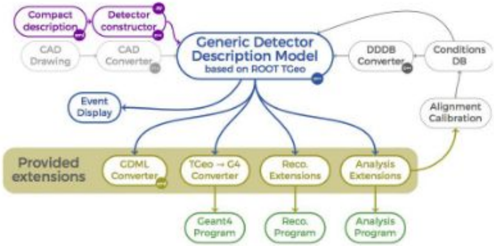
The common software vision: key4hep

key4hep is a huge ecosystem of software packages adopted by all future collider projects, complete workflow from generator to analysis

- Event data model: EDM4HEP for exchange among framework components
  - Podio as underlying tool, for different collision environments
  - Including truth information
- Data processing framework: Gaudi
- Geometry description: DD4hep, ability to include CAD files
- Package manager: Spack: `source /cvmfs/sw.hsf.org/key4hep/setup.sh`



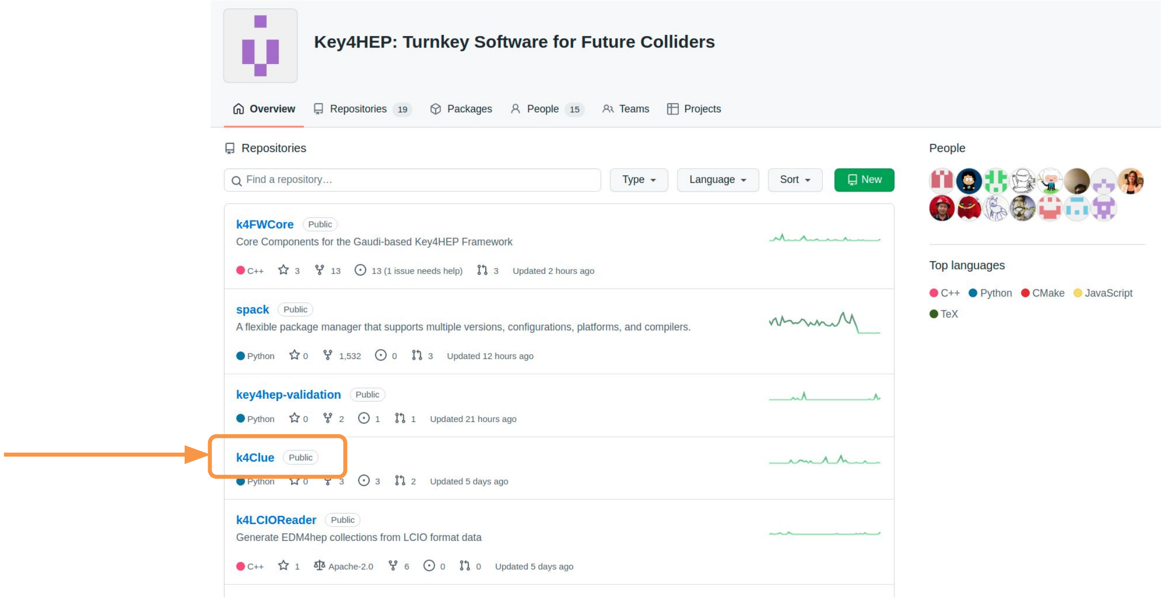




Armin Ilg (UZH) FullSim of IDEA using key4hep LCWS 2023 @ SLAC 17.05.2023 4 / 14

Integrating CLUE in Key4hep

- **k4Clue v01-00** (doi: [10.5281/zenodo.7851995](https://doi.org/10.5281/zenodo.7851995))
  - It's adapted to the common event data model, EDM4hep
  - It includes a wrapper class to run in the Gaudi software framework
  - It's included in the new Key4hep releases managed by Spack

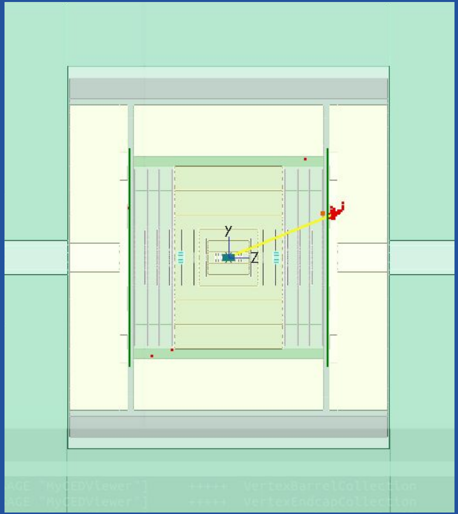


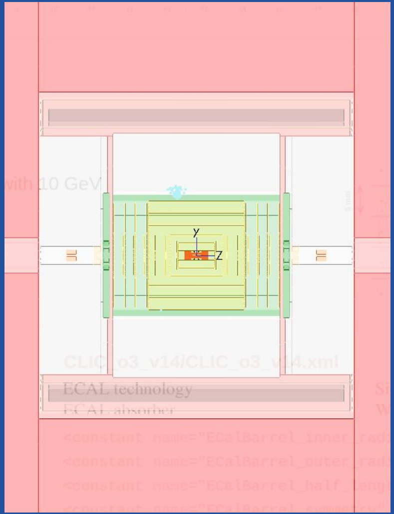
Additional features w.r.t. kalos/Clue

- **Cluster hits in the entire 4π detector region**
  - Definition of the tessellated space (`LayerTile`) in the standalone version defines coordinates and searches only in the transverse plane
  - Modified basic structure of the `LayerTile` and the search algorithm to to allow for the definition of a cylindrical surface

    $x \rightarrow r\Phi$ $y \rightarrow z$
- **Template CLUE algorithm classes**
  - To allow the possibility of defining several different calorimeter layouts
  - A dedicated documentation page in the package (`include/readme.md`) allows the user to follow a simple but detailed step-by-step procedure to introduce and test the preferred layout.
- **GitHub CI & EDM4hep Validation**
  - `edm4hep:CLUECalorimeterHit : CalorimeterHit` class with specific methods related to the CLUE algorithm

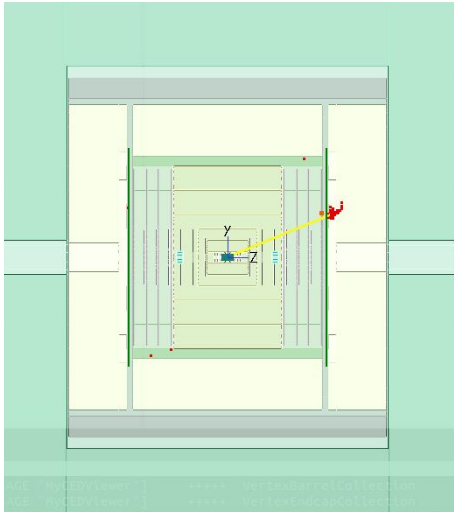
Performance evaluation
- CLD & CLICdet -

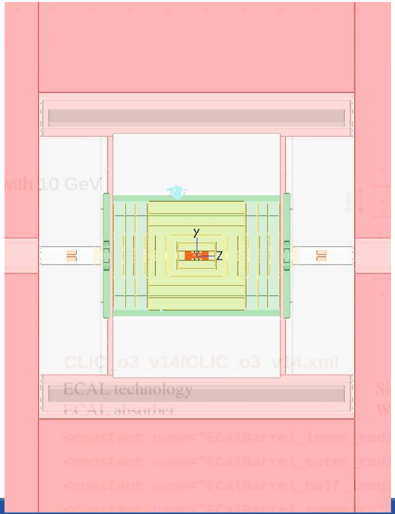




ECAL of CLICdet & CLD

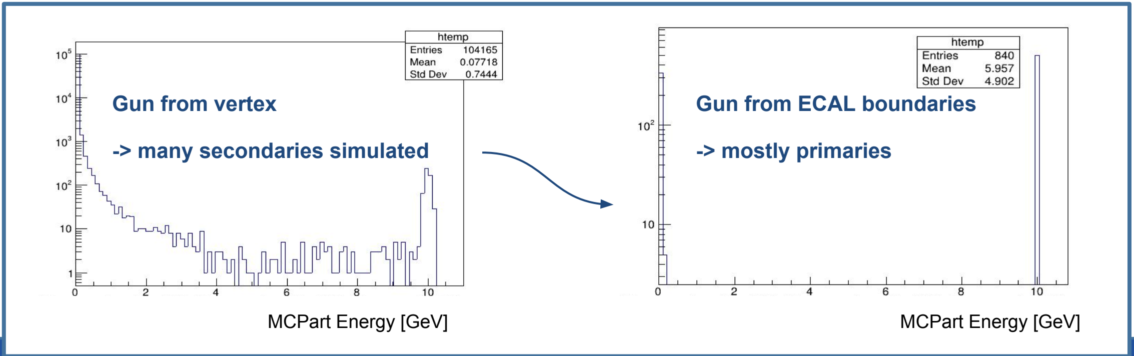
- 40 layers of 5x5 mm^2 Silicon cells & W
- The main difference between the two calorimeters lies in the layout parameters → To compensate for a lower detector solenoid field, the CLD design starts from a larger radius both in the barrel and in the endcap region w.r.t. CLICdet.
  - Further details in backup

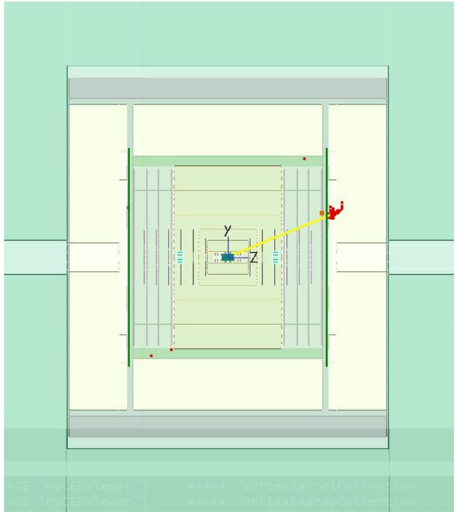


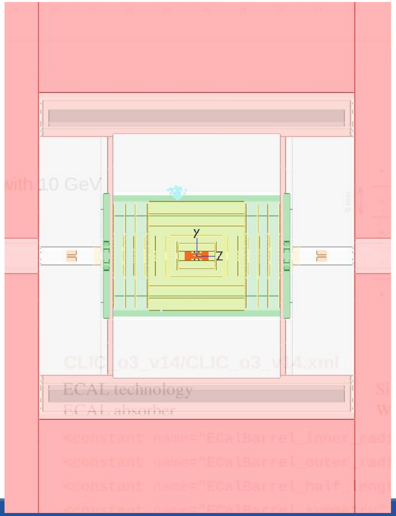


ECAL of CLICdet & CLD

- 40 layers of 5x5 mm^2 Silicon cells & W
- The main difference between the two calorimeters lies in the layout parameters → To compensate for a lower detector solenoid field, the CLD design starts from a larger radius both in the barrel and in the endcap region w.r.t. CLICdet.
  - Further details in backup
- 500 events of single gamma at 10 GeV generated perpendicular to the surface with [Geant4 General Particle Source]
  - Main reason: no conversion in the tracker volume

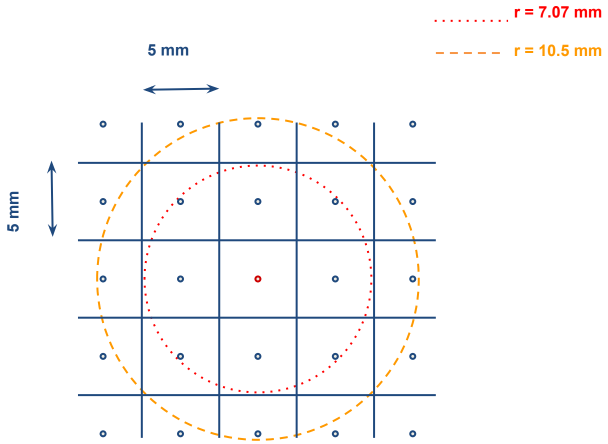






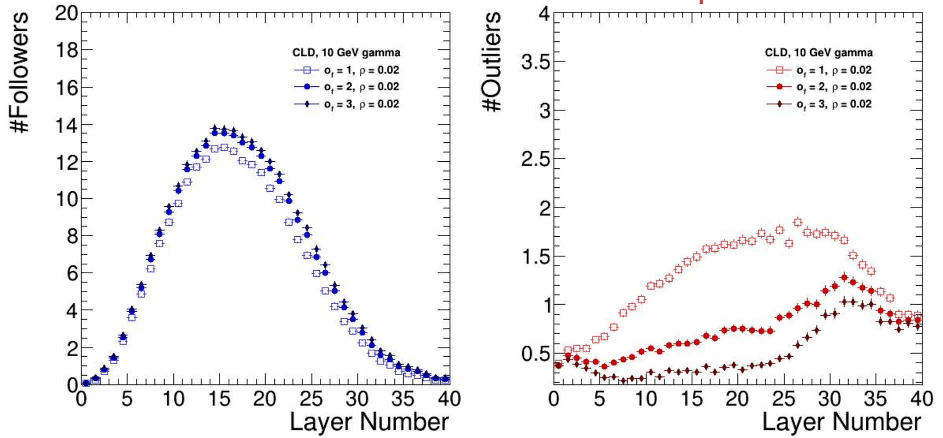
Parameters tuning

- Input parameters tuned for CLD
- Same ones tested also for CLICdet **(similar geometry, same granularity)**
- Critical Distance (d_c) is established by geometry granularity to contain (minimum) the close neighbors cells:
  - d_c = 15 mm

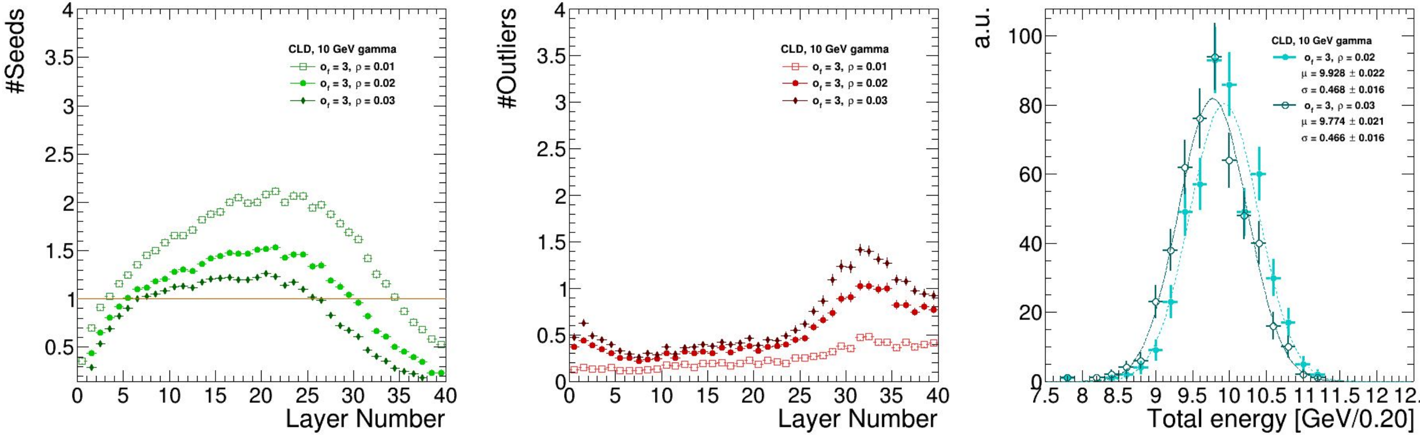


Parameters tuning

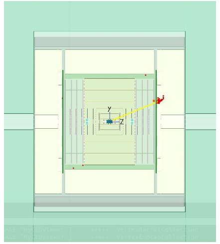
Varying the outlier factor (o_f)



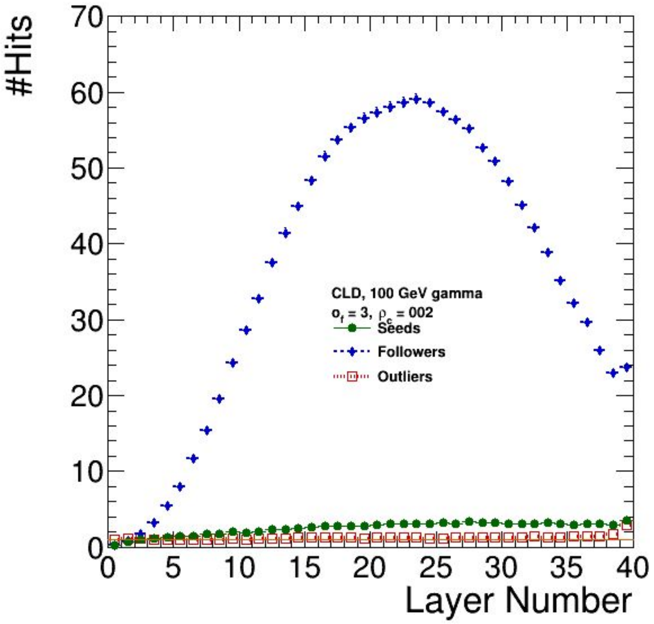
Varying the minimum local density (ρ_c)

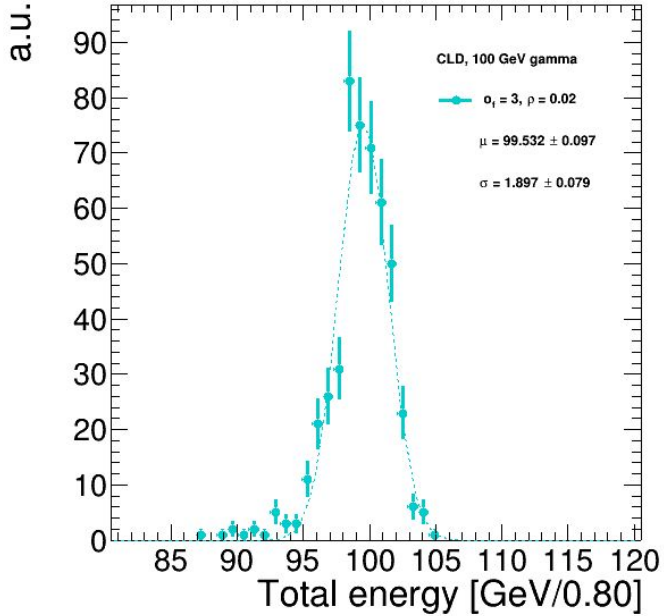


Higher energies



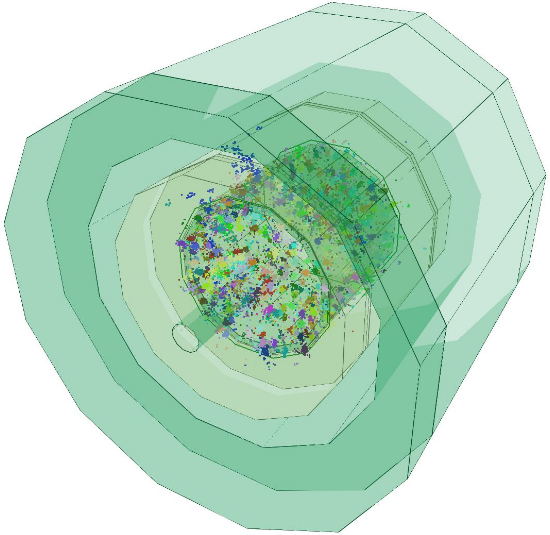
- 500 events with single gamma (from ECAL surface) at 100 GeV
- **d_c = 15.00, ρ_c = 0.02, o_f = 3.0**

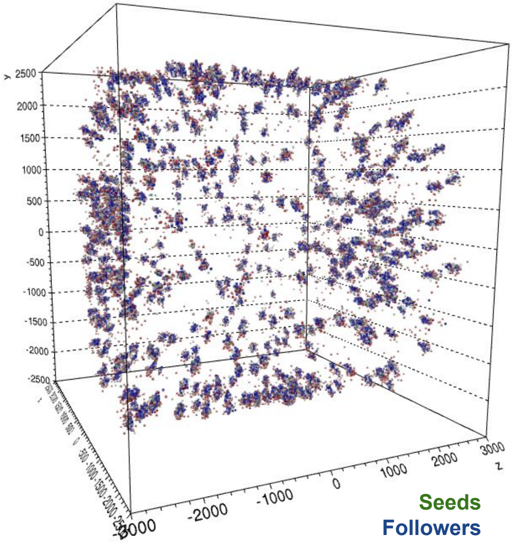


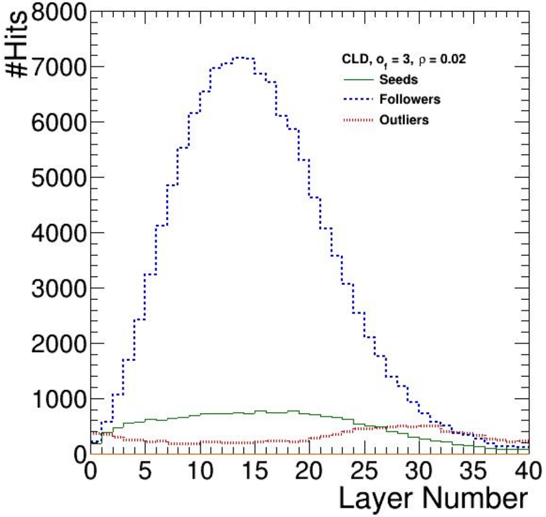


Multiple gamma event

- Produced with normal gun, i.e. particles generated from vertex
- 1 event with 500 single gammas each produced with 10 GeV
  Only simulated calorimeter hits are shown

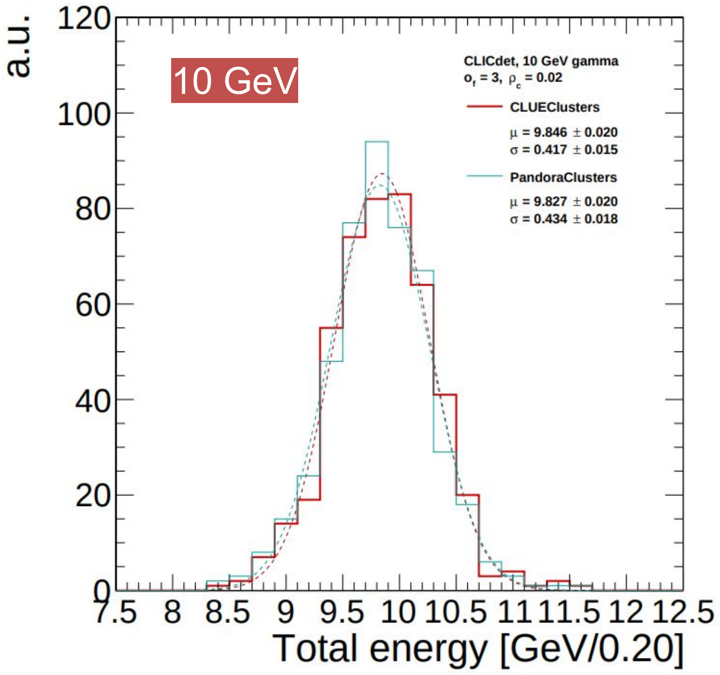


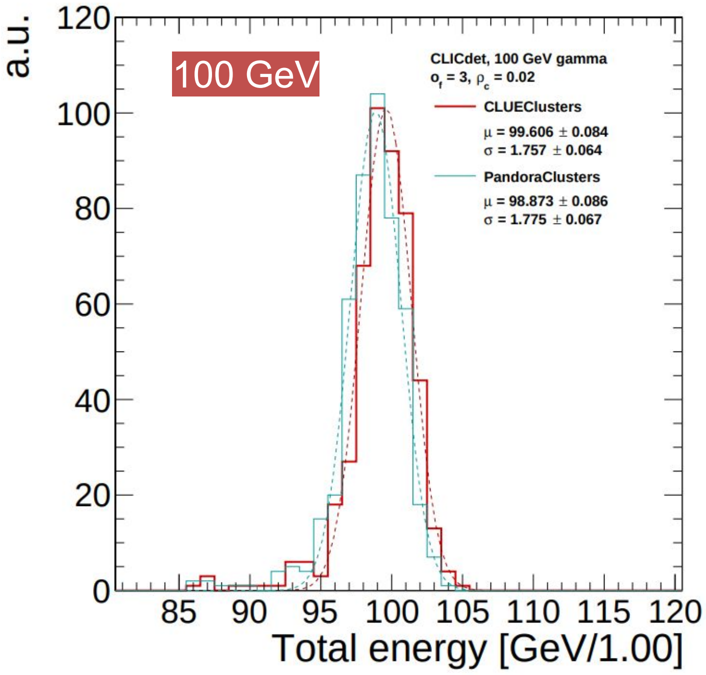




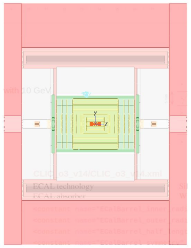
CLICdet results

- Using same input parameters selected for CLD

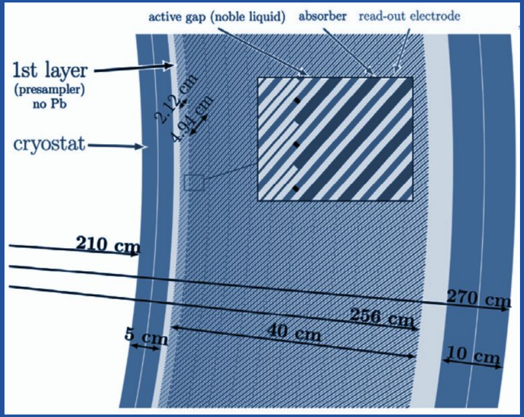




Comparison with **Pandora Clusters** not completely equitable comparison (it includes a dedicated calibration procedure), but comparable results in terms of energy linearity and resolution



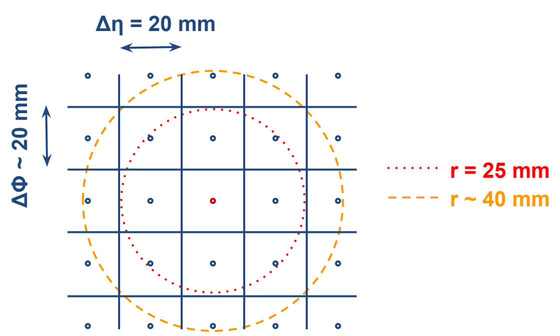
Performance evaluation
- Noble Liquid Calo -

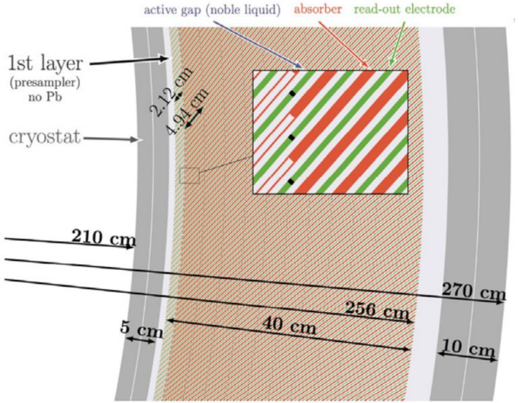


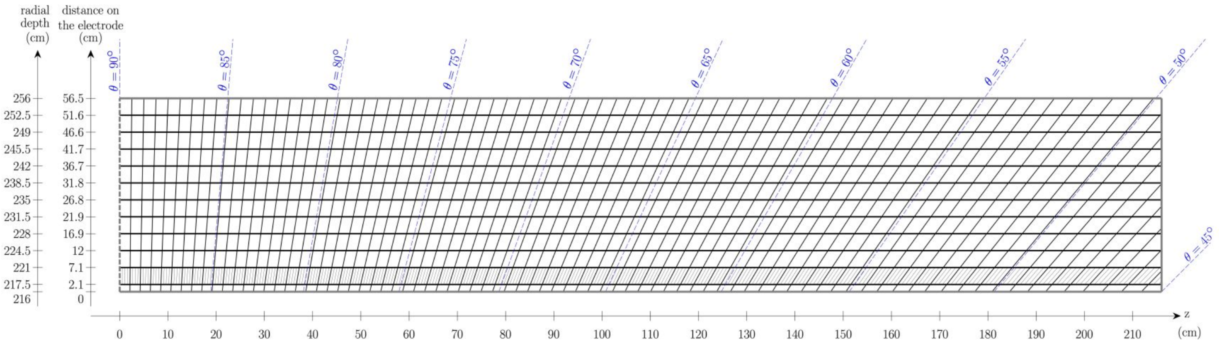
Noble Liquid ECAL for FCC-ee

- 12 layers, only barrel considered
  - cell size in Φ: 17.9 mm - 20.7 mm
  - cell size in η: ~ 20 mm
- Sample (if not stated otherwise):
  - 500 single gamma at 10 GeV
  - $\theta_{[min, max]}$ = [50, 130]

Critical distance choice: d_c = 40 mm







Parameters tuning

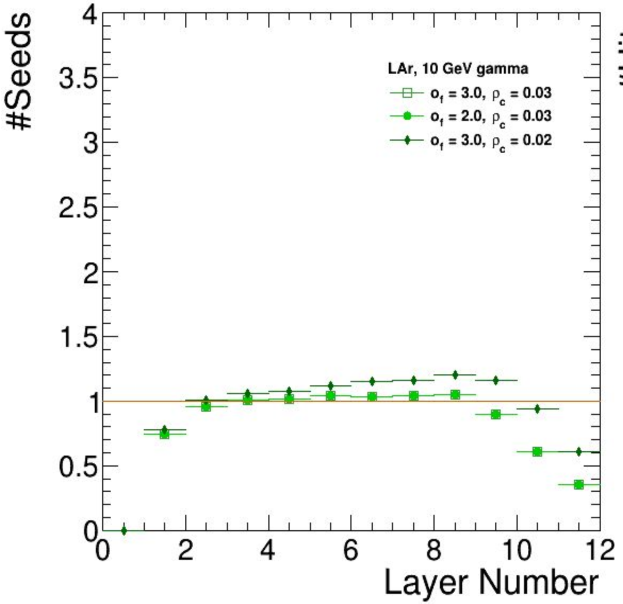
- 500 events with single gamma (from vertex) at **10 GeV**
- **d_c = 40.00, ρ_c = 0.03, o_f = 3.0**

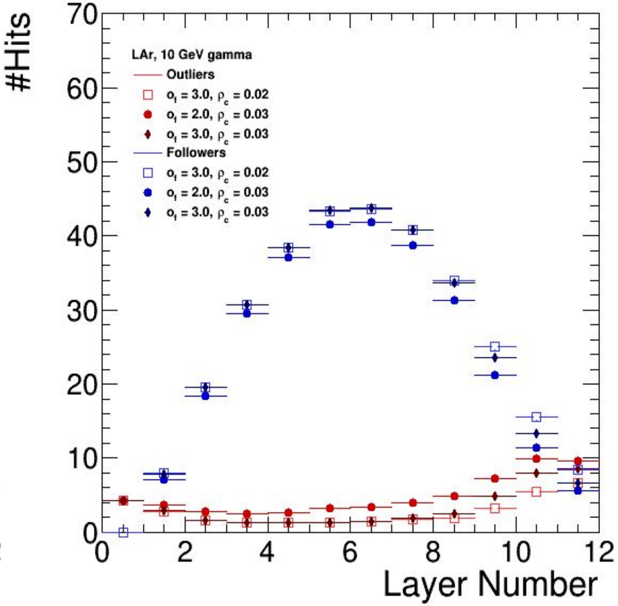
Input parameters:

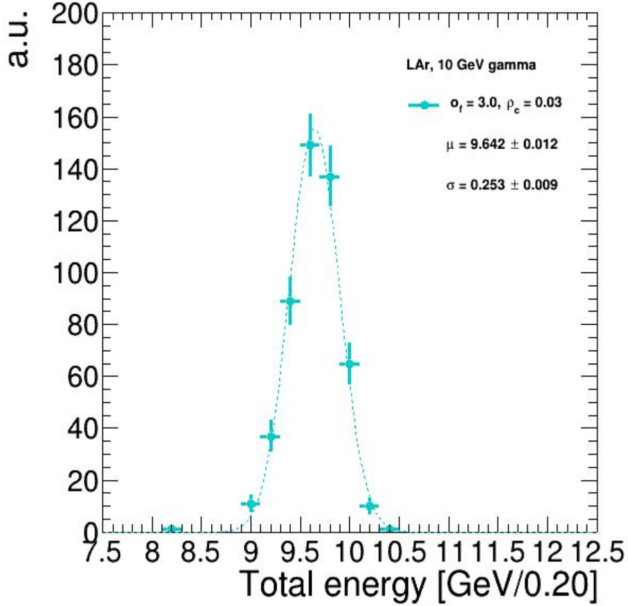
d_c : Critical Distance

o_f : Outlier Delta Factor

ρ_c : Minimum Local Density



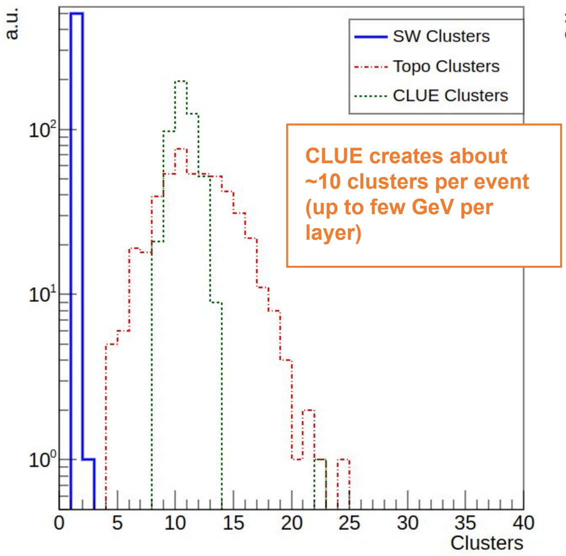


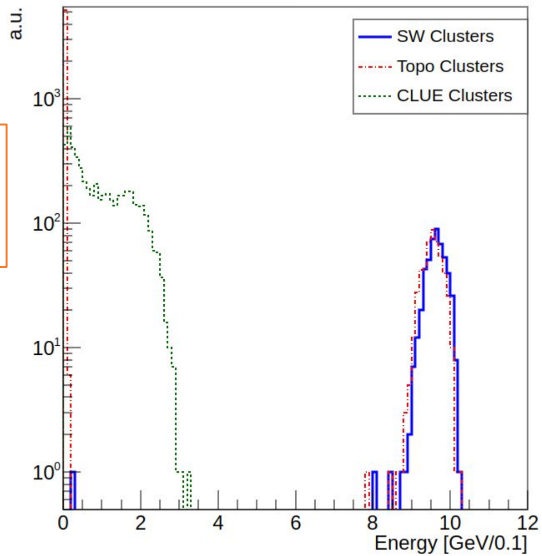


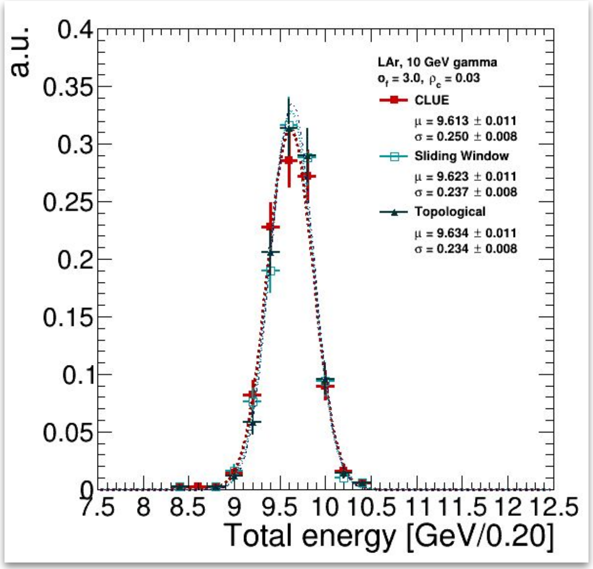
Comparison with other cluster algorithms

[Calorimeters for the FCC-hh](), Dec 2019

- **Sliding window**: It considers the calorimeter as a two-dimensional grid in η-ϕ space, neglecting the longitudinal segmentation of the calorimeter.
- **Topological clustering**: It starts with a seed cell and then adds topologically connected calorimeter cells

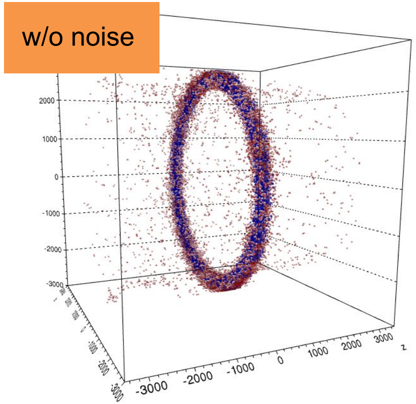


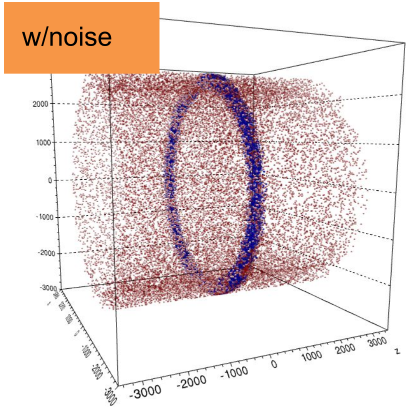




Noise in Liquid Argon Calorimeter

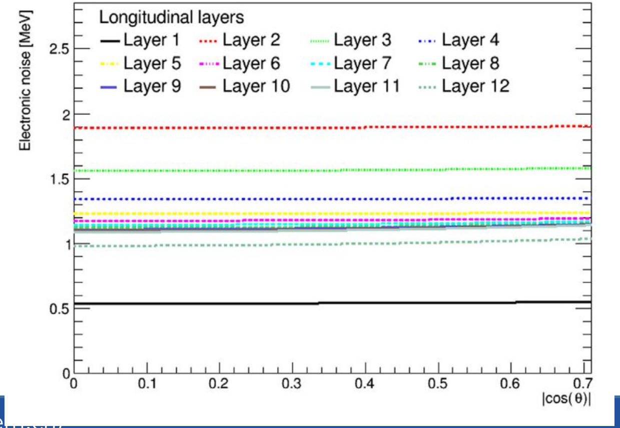
Signal produced only with θ ~ 90.25°





- High level of noise in the detector
- In the topoclustering, there is no filter directly at the beginning for the noise, but this is done using cuts in the algorithm itself
- The main observable is the cell significance ξ_{cell} which is defined as the absolute value of the ratio of the cell signal to the expected noise in this cell

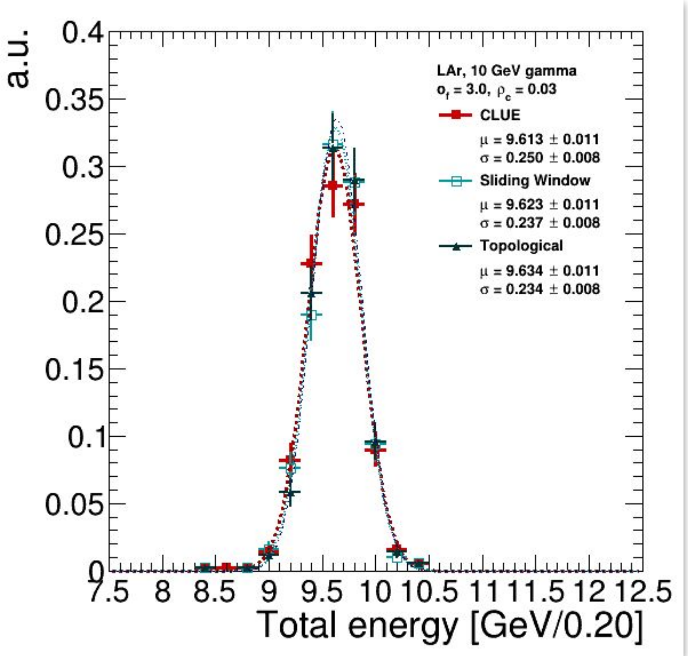
$$\xi_{\text{cell}} = \left| \frac{E_{\text{cell}}}{\sigma_{\text{cell}}^{\text{noise}}} \right|$$



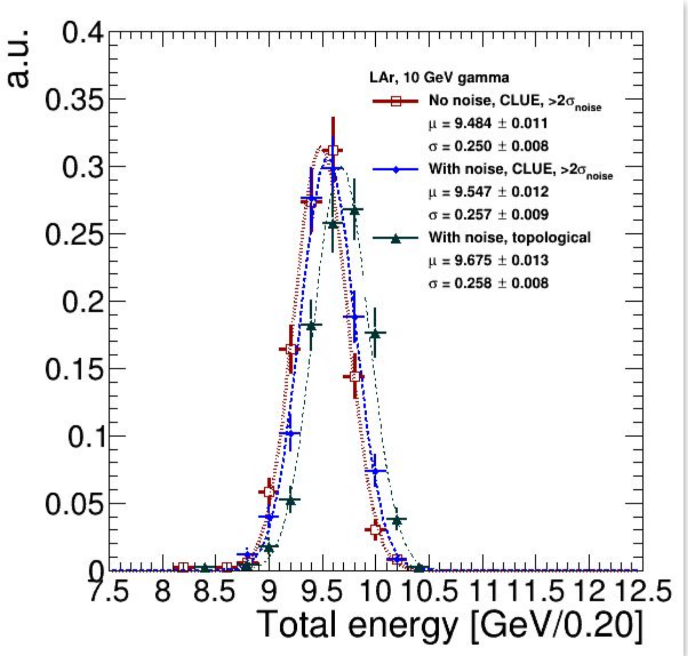
- CLUE hits w/noise **selected with filter of > $2\sigma_{noise}$**

Comparison with other cluster algorithms

No noise

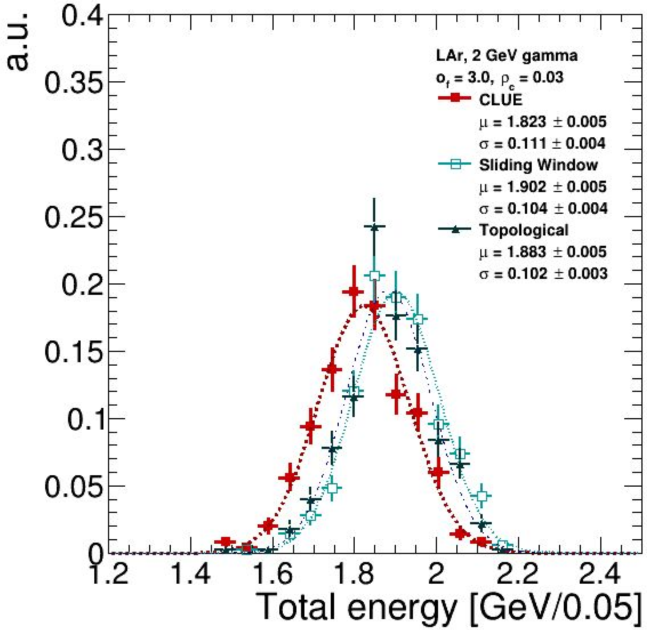


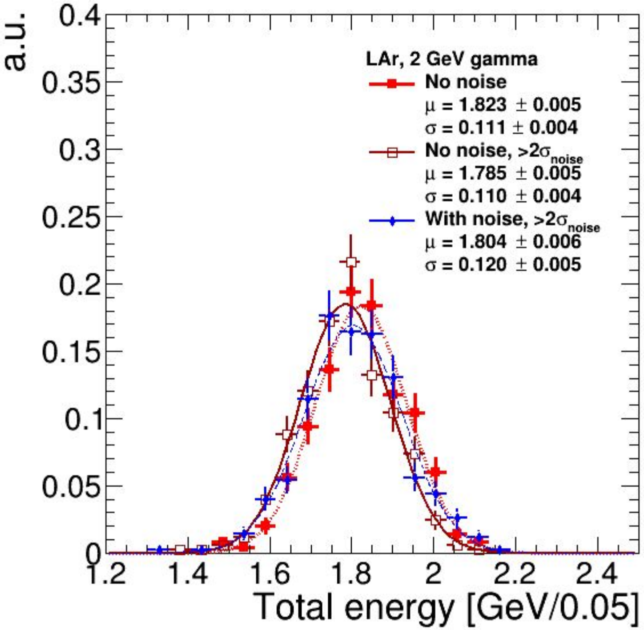
Including noise



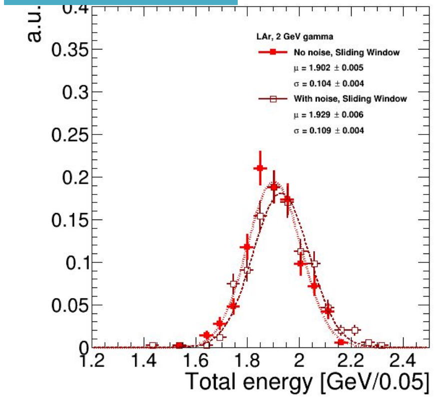
Low(er) energy

- 500 events with single gamma (from vertex) at **2 GeV**
  Motivated by flavor physics searches at Z peak

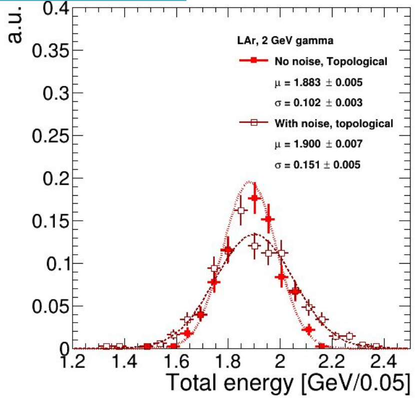




Sliding Window
cluster energy > 1GeV



Topological



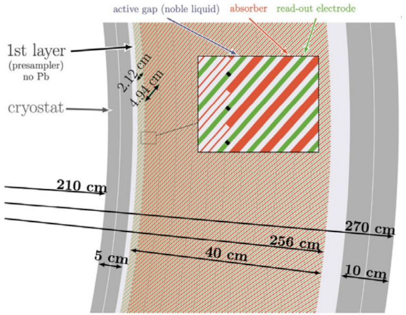
Conclusions & Outlook

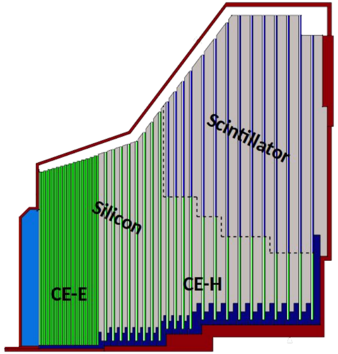
Conclusions

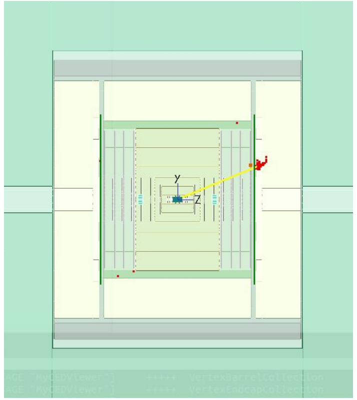
- k4Clue package (v01-00) has improved upon the standalone CLUE
  - Run on the full detector (barrel & endcap)
  - Adapted for different types of calorimeters
- Analysis on three different future calorimeters has demonstrated the good performance for single gamma events
  - Good performance even in the presence of noise
  - Compared favorably to other baseline algorithms

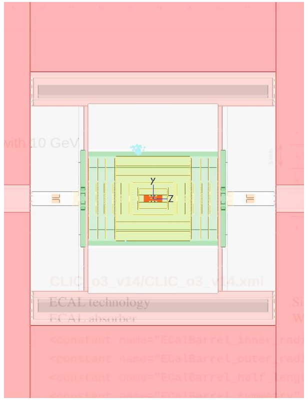
This work highlights the adaptability and versatility of the CLUE algorithm for a wide range of experiments and detectors, as well as its potential for future high-energy physics experiments beyond CMS

- Improvements from k4clue also under discussion to use the developments also in CMS (Phase-2 barrel region)









Conclusions

- Final article summarizing k4clue and its performance for future collider detectors almost ready
    - Computing time under study
- This research was supported by the CERN Strategic R&D Programme on Technologies for Future Experiments
- Special thanks go to the Key4hep team and the FCC-ee liquid calorimeter software experts for the support

EP-R&D
Programme on Technologies for Future Experiments

k4clue

E Brondolin[1], M Rovere[1], and F Pantaleo[1]

[1]CERN, Switzerland

May 15, 2023

Abstract

High granularity calorimeters have become increasingly crucial in modern particle physics experiments, and their importance is set to grow even further in the future. The CLUstering of Energy (CLUE) algorithm has shown excellent performance in clustering calorimeter hits in the High Granularity Calorimeter (HGCAL) developed for the Phase-2 upgrade of the CMS experiment. In this paper, we investigate the suitability of the CLUE algorithm for future linear collider experiments and test its capabilities outside the HGCAL software reconstruction. To this end, we developed a new package, k4Clue, which is now fully integrated into the Gaudi software framework and supports the EDM4hep data format for inputs and outputs. We demonstrate the performance of CLUE in three detectors for future colliders: CLICdet for the CLIC accelerator, and CLD and a second calorimeter based on noble liquid technology proposed for FCC-ee. We find excellent performance for single gamma events, even in the presence of noise and also compared with other baseline algorithms. This work highlights the adaptability and versatility of the CLUE algorithm for a wide range of experiments and detectors and the algorithm's potential for future high-energy physics experiments beyond CMS.

Contents

1 Introduction 1
2 CLUE 2
2.1 Clustering procedure 2
3 The k4Clue package 3
3.1 Integrating CLUE in Key4hep 3
3.2 k4CLUE additional features 3
3.3 Testing and validation 4
4 Performance evaluation 4
4.1 Performance for CLD and CLICdet detector 4
4.1.1 CLD calorimeter results 5
4.1.2 CLICdet calorimeter results 5
4.2 Results on Noble Liquid Calorimeter 7
5 Conclusions 9

1 Introduction

The High Energy Physics field is transitioning towards using calorimeters with higher lateral and longitudinal readout granularity due to their numerous advantages, such as providing more precise particle identification, detailed reconstruction of electromagnetic and hadronic showers, and improved separation of signals from pile-up events.

In the near future, the CMS high granularity calorimeter (HGCAL), developed for the HL-LHC era, will represent the next step in the generation of high-granularity calorimeters [1]. It is an imaging

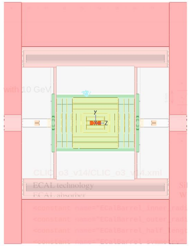
1

Backup

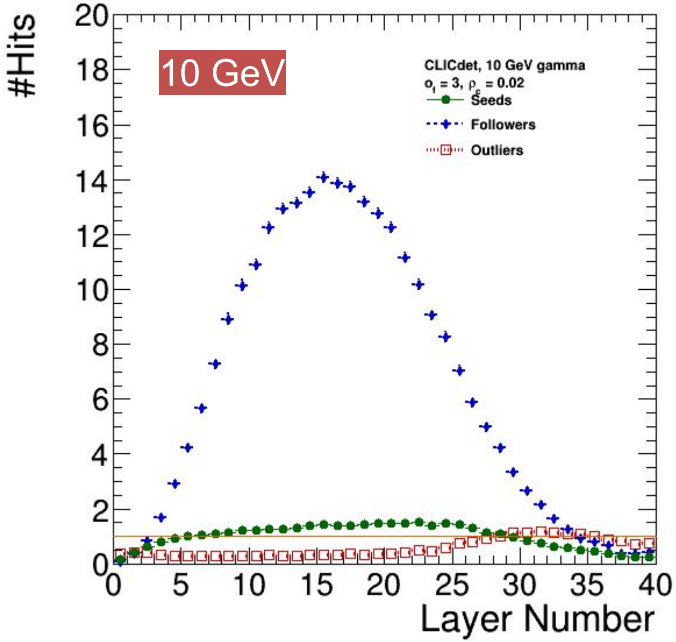
Integrating CLUE in Key4hep

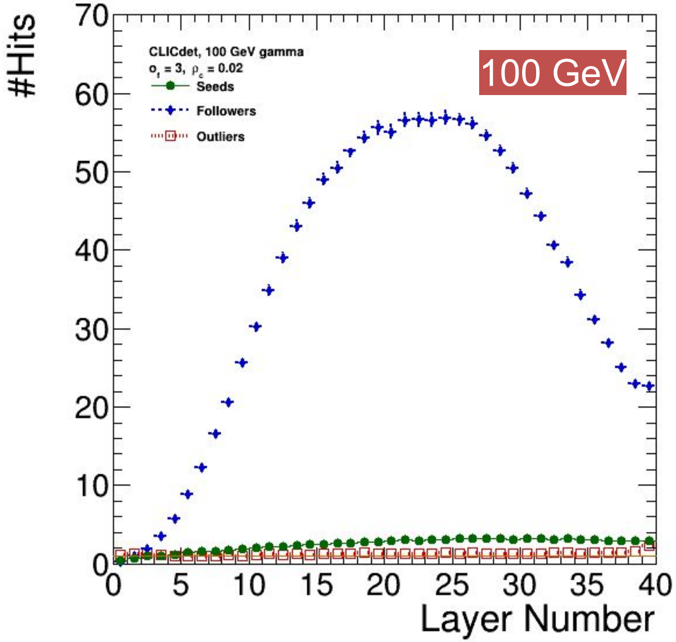
- **GitHub CI** to ensure that the modifications or additions to the software do not break the clusterization process
  - In the latest release was modified to focus on C++ code and `EDM4hep` data
  - Test on both current key4hep release and nightlies
- **Validation**
  - `edm4hep:CLUECalorimeterHit`
    - CalorimeterHit class with specific methods related to the CLUE algorithm
  - `CLUEHistograms` class to produce ntuples

CLICdet results



- Using same input parameters selected for CLD



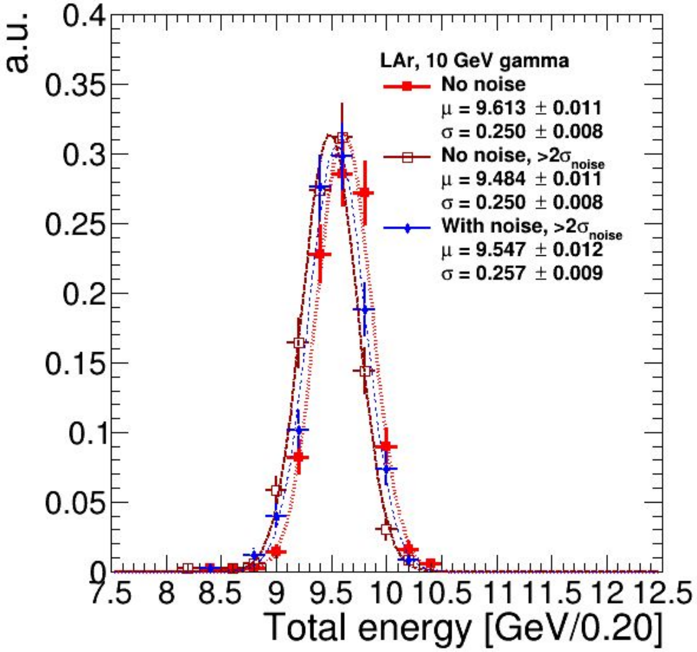


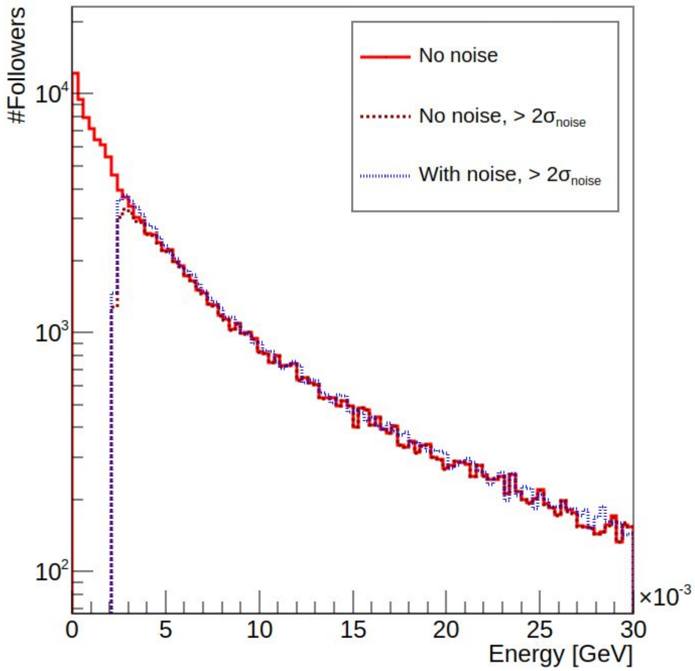
Comparison with **Pandora Clusters** not completely equitable comparison (it includes a dedicated calibration procedure), but comparable results in terms of energy linearity and resolution

Noble Liquid Calo

Pre-filtering

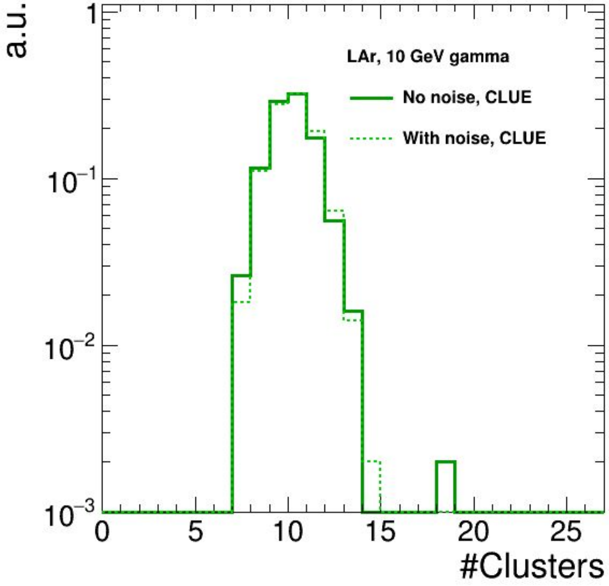
- CLUE hits w/noise selected with filter of > $2\sigma_{noise}$



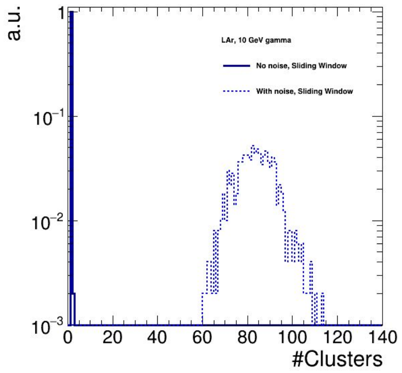


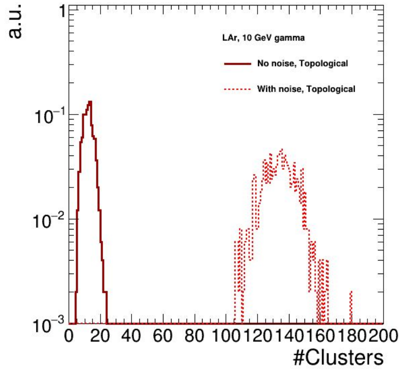
Noble Liquid Calo

Comparison with other cluster algorithms

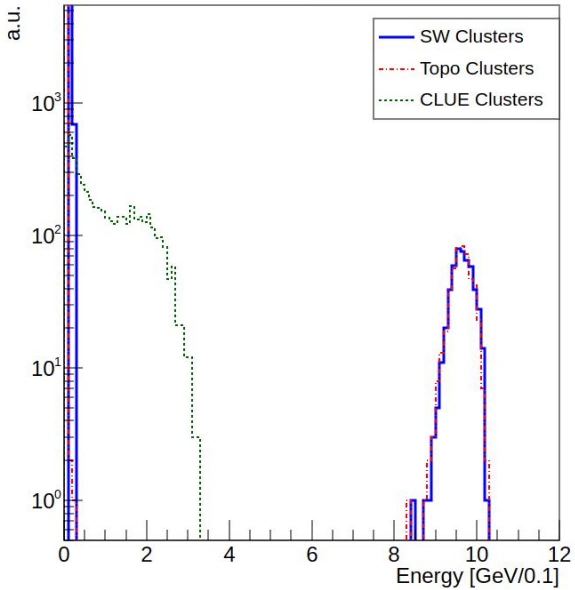


No significant effect on CLUE clusters - about ~10 per event (one per layer)





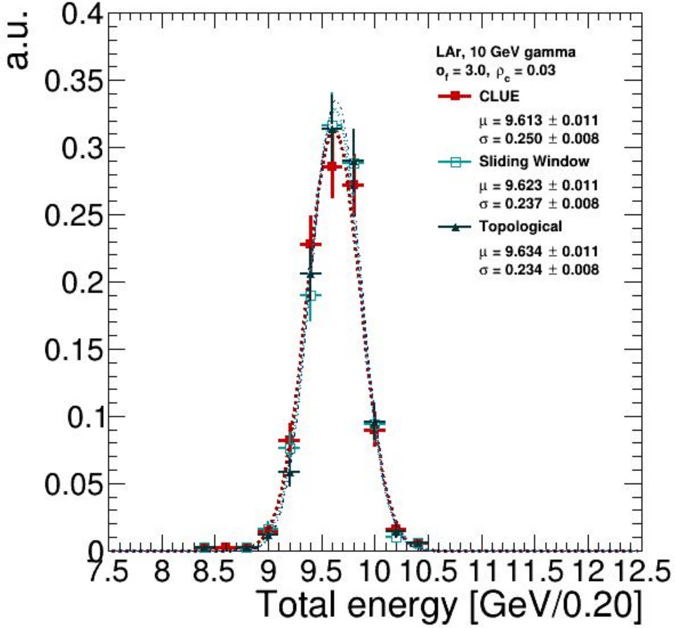
w/noise

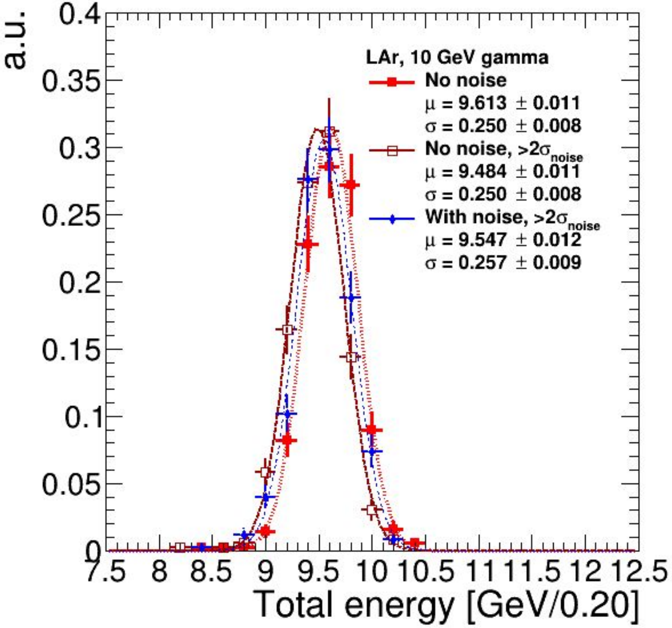


More than one SW and Topo cluster per event, but most of them with low energy

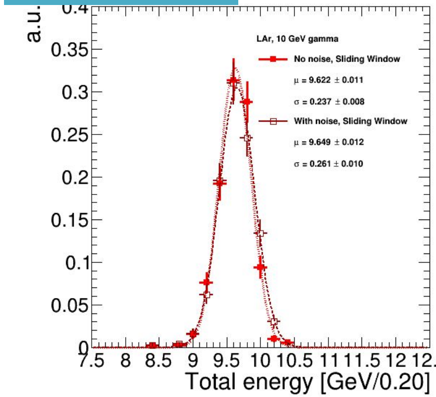
Noble Liquid Calo

Summary for 10 GeV gammas





Sliding Window cluster energy > 1GeV



Topological