The TICL reconstruction at the CMS Phase-2 High Granularity Calorimeter Endcap

Florian Beaudette[1], Shamik Ghosh[1], Benedikt Mayer[2], Abhirikshma Nandi[3], Felice Pantaleo[2], Wahid Redjeb[2,3], Marco Rovere[2], Alice Savona[4], Alexander Schmidt[3], Alessandro Tarabini[1], on behalf of the CMS collaboration

[1]Leprince Ringuet Laboratory, Paris, France
[2]CERN, European Organization for Nuclear Research, Meyrin, Switzerland
[3]RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
[4]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

E-mail: `wahid.redjeb@cern.ch`

Abstract. To sustain the harsher conditions of the high-luminosity LHC [1], the CMS Collaboration [2] is designing a novel endcap calorimeter. The new calorimeter will predominantly use silicon sensors to achieve sufficient radiation tolerance and will maintain highly granular information in the readout to help mitigate the effects of the pileup. In regions characterized by lower radiation levels, small scintillator tiles with individual SiPM on-tile readout are employed. A unique reconstruction framework (TICL: The Iterative CLustering) is being developed within the CMS Software CMSSW to fully exploit the granularity and other significant detector features and precision timing, with a view to mitigating pile up in the very dense environment of HL-LHC. The TICL framework was specifically designed with heterogeneous computing in mind, with algorithms and data structures optimized for execution on GPUs. Additionally, the framework includes geometry-agnostic data structures that enable efficient navigation and searching. The framework is composed of several key components, including seeding capabilities that leverage information from other detectors, dynamic cluster masking, energy calibration, and particle identification. To maximize flexibility, TICL allows for the iterative chaining of different combinations of modules.

1. Introduction

To handle high dose rates and pileup during the HL-LHC era, the CMS detector is undergoing upgrades, including the replacement of the CMS endcap with the High Granularity Calorimeter (HGCAL). The HGCAL features electromagnetic and hadronic sectors (CE-E and CE-H) with improved radiation hardness and increased granularity, offering excellent longitudinal and transverse resolution. The HGCAL includes 47 longitudinal samplings, with 26 in the electromagnetic sector and 21 in the hadronic sector. The electromagnetic and the major part of the hadronic sector will consist of silicon cells of size ($\approx 0.5 - 1\,\mathrm{cm}^2$), while the remaining part of the CE-H will adopt highly-segmented plastic scintillators ($\approx 4 - 30\,\mathrm{cm}^2$) in the region with lower radiation. The high granularity provides potential benefits for the reconstruction of physical objects and particle flow interpretation, enhancing the separation power between particle showers. A more detailed description can be found in the Technical Design Report [3]. With the expected high pileup environment during the HL-LHC phase, the online and offline reconstruction [4] of the High Granularity Calorimeter (HGCAL) is of utmost importance. However, this presents a challenging task for both physics and computing, particularly when faced with an average of 200 proton-proton collisions. The full combinatory approach of classical reconstruction algorithms is not viable due to the exploding time and memory usage. Nevertheless, the new HGCAL detector design provides an opportunity for the development of novel techniques and algorithms, such as clustering, graph theory, machine learning, and their combinations, which can leverage modern acceleration devices.

2. TICL: The Iterative CLustering

The *Iterative CLustering* (TICL) is a modular framework designed for the HGCAL reconstruction. Developed within CMSSW, TICL consists of multiple building blocks, as shown in Figure 1. It takes as input 2D clusters, known as *layer-clusters*, which are obtained from the clustering of energy deposits in HGCAL referred to as *rechits*. The outputs of TICL are physics objects that represent particle showers and the corresponding particle properties and identification probabilities. The core components of the TICL framework are the Seeding Region, the Pattern Recognition, and the Particle Flow reconstruction and interpretation.

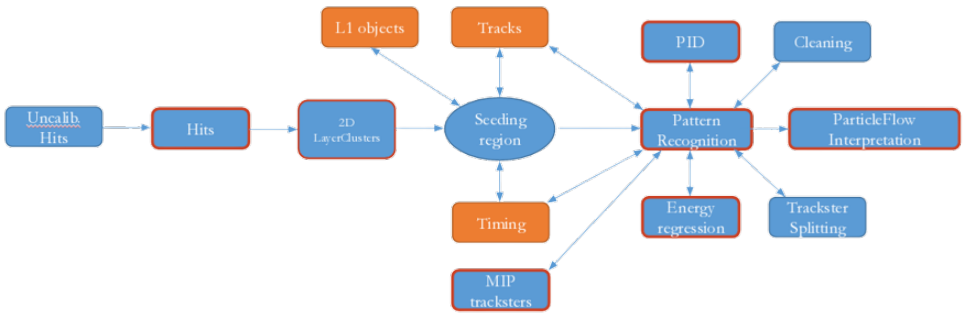


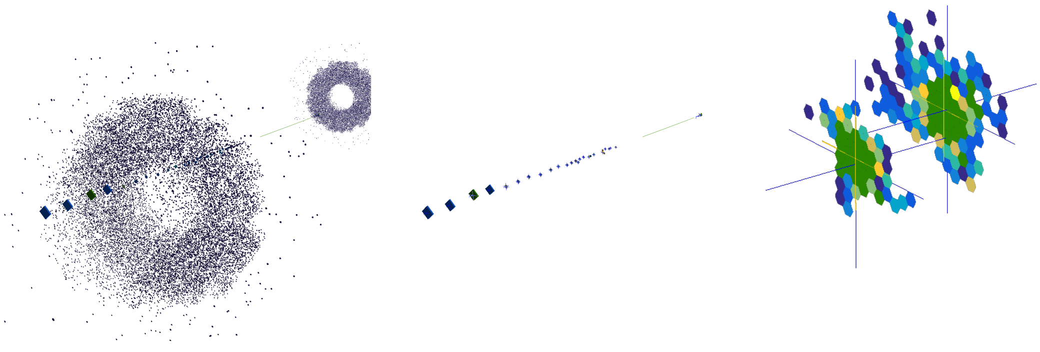
Figure 1: Diagram of the TICL framework. Each block corresponds to a different reconstruction step. Blocks with the red border are the steps targeted by the validation. Orange boxes represent physic objects coming from different detectors

2.1. Layer Clusters

The TICL framework is built around the layer clusters which are clusters of calibrated rechits on each HGCAL layer. This clustering is performed using CLUE [5], an energy-density-based clustering algorithm. CLUE is designed to create clusters with real signals and suppress pure-noise clusters, providing an advantage in effectively identifying signal hits. The algorithm identifies clusters' *seeds* with sufficient energy on each HGCAL layer, around which the reconstructed hits are aggregated to form the layer clusters. Figure 2 shows how CLUE only clusters the signal deposits.

2.2. Seeding Regions

Seeding regions aim to reduce the number of layer-clusters used in subsequent pattern recognition steps by limiting the input to clusters within specific volumes, based on seed information. When no seeding regions are defined, the reconstruction considers the entire solid angle of the HGCAL detector. Information from other detectors can be leveraged to identify volumes where the pattern recognition will be applied. Currently, TICL can use *Tracks* to build the seeding regions, exploiting the point, direction, and momentum obtained by their propagation to the innermost layer of HGCAL. The same procedure can be adopted by using the L1-Trigger



(a) (b) (c)

Figure 2: (a): 3D visualization of two HGCAL endcaps and the corresponding rechits (including detector noise), containing two simulated muons produced back-to-back with $p_T = 10\,\text{GeV}$. (b): The resulting layer clusters obtained by applying CLUE on the rechits. (c): An example of two reconstructed layer clusters on two different HGCAL layers. The color coding is proportional to the rechits' energy, the green representing the higher energy deposits [6].

seed. This approach is extremely useful to obtain a configuration that can run a faster regional reconstruction during the event selection at the CMS High Level Trigger (HLT).

2.3. Pattern Recognition

The inputs for the Pattern recognition step are the layer clusters available in the defined seeding regions. The Pattern Recognition algorithm connects the layer-clusters together to produce Direct Acyclic Graphs (DAG), where each node corresponds to a layer-cluster and the edges are the connection found between them. These DAGs are called *Tracksters* and they represent particle showers. The Pattern Recognition algorithms currently available are the following:

- **Cellular Automaton**: It starts building pairs of layer-clusters (*doublets*). The algorithm tries to connect doublets together, requiring them to be aligned and to be pointing to the corresponding seeding region. If the alignment conditions are satisfied they become neighbours of each other. Eventually, a Graph is obtained, and a Depth First Search is used for traversing the graph and building the final Trackster [7].
- **CLUE3D**: is a clustering algorithm that extends the capabilities of CLUE by clustering layer-clusters in the third, longitudinal dimension instead of individual rechits. Its purpose is to track the energy flow of showers in a more detailed manner.
- **FastJet** [8]: this algorithm can be used to aggregate layer-clusters to produce Jets and derive Tracksters from them.

In addition to the default CLUE3D algorithm, different pattern recognition algorithms can be used for different seeding regions. This step involves creating Tracksters as DAG and computing aggregated information, such as the barycenter position, shower direction, and timing information, utilizing Principal Component Analysis (PCA). The detector is divided into $\eta - \phi$ bins called *tiles* to facilitate fast navigation and search of nearby objects. These tiles cover the entire solid angle and currently have a dimension of $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$, allowing for a $\mathcal{O}(1)$ look-up. Tiles are also used in the subsequent Particle Flow reconstruction step described in the next section.

2.4. Particle Flow interpretation

Particle showers in the HGCAL detector typically consist of multiple sub-clusters, which may be produced by minimum ionizing particles in the case of hadron showers, or by bremsstrahlung for electromagnetic ones. Linking together sub-clusters associated with the same particle can significantly improve the final event description. However, the current pattern recognition algorithm may produce split tracksters, especially at the transition region between the CE-E and CE-H detectors where the characteristics of the two detectors are very different. The Particle Flow reconstruction aims to link sub-clusters from the same particle together and connect calorimetric energy deposits with the track objects entering the HGCAL detector, improving the overall event reconstruction. For this reason, the Linking step takes as input the Trackster collection produced by the Pattern Recognition and the reconstructed Tracks. The algorithm aims to simplify the problem by projecting tracks and tracksters onto a common surface, effectively reducing it to two dimensions. Geometrical criteria are then applied to identify neighbours and create a Graph of tracksters. This graph is iteratively analyzed, taking into account both energy and time compatibility, to form the final links between tracksters. Once the links are established, the algorithm generates a set of merged Tracksters and a collection of TICL Candidates, which serve as a proxy for CMS Particle Flow reconstruction. The TICL Candidates are processed to determine the shower direction, regressed energy, and particle identification.

3. Results

To evaluate the reconstruction performance of TICL, the reconstructed objects are compared against the simulation. For validation purposes, a collection of simulated tracksters is created, called *SimTracksters*. This object has the same data format as a reconstructed Trackster, and it is a DAG of all the layer clusters generated by the simulated particle entering HGCAL. In the following, the trackster *efficiency* is used as the evaluation metric.

Efficiency Number of reconstructed Tracksters that share at least 40% of their energy with a SimTrackter, divided by the total number of SimTracksters. It is important to notice that the shared energy is computed by accumulating the energy of the rechits in common between the reconstructed and simulated object.

3.1. TICL Performance

The TICL performance, which includes both Pattern Recognition and Linking step, has been evaluated for unconverted photons and charged pions in both 0 and 200 pileup conditions. The performance for photon reconstruction is extremely good in both 0 and 200 pileup conditions, as shown in Figure 3. A small decrease in efficiency can be seen at $\eta \approx 1.5$, which is due to shower leakage at the HGCAL boundary. This same decrease in efficiency can also be seen at $\eta \approx 3.0$ in the 200 pileup scenario, due to the high detector occupancy in the high η region. The charged pion reconstruction performance in the 0 pileup scenario is also good overall, but with a drop in efficiency in the low energy range, as shown in Figure 4. The pileup has a significant impact on the reconstruction performance of hadrons, especially in the high η region where the detector occupancy is very high. In Figures 3, 4 the efficiency as a function of η is computed between 1.5 and 3.0, while the efficiency versus energy, p_T and ϕ is computed in $\eta \in [1.7, 2.7]$, to disentangle the effects at the HGCAL boundary.

4. Conclusions and outlooks

TICL is a modular framework developed for the current HGCAL reconstruction in CMSSW, which is currently under active development. Its design provides developers with a flexible and user-friendly framework for experimenting with different algorithms and reconstruction

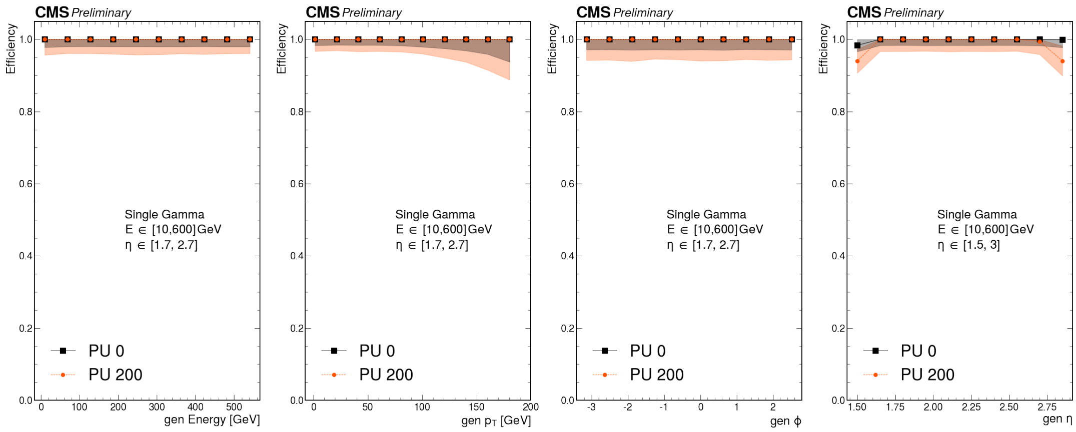


Figure 3: The reconstruction efficiency of the merged trackster collection obtained after the linking procedure compared between 0PU and 200PU, for energy, p_T, ϕ and η, for unconverted photons.

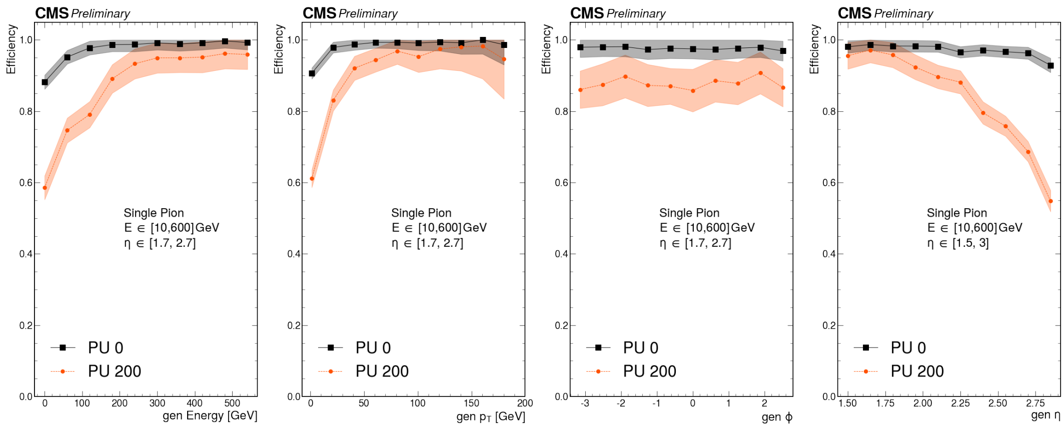


Figure 4: The reconstruction efficiency of the merged trackster collection obtained after the linking procedure compared between 0PU and 200PU, for energy, p_T, ϕ and η, for charged pions.

techniques, with common interfaces and validation steps at various points in the reconstruction chain. TICL has shown excellent performance in reconstructing electromagnetic objects in 200PU, and efforts are ongoing to improve hadron reconstruction. In terms of timing performance, the full HGCAL reconstruction using TICL takes less than 5% of the full CMS Phase 2 offline reconstruction on a single CPU core, for a $t\bar{t}$ event with 200 overlapping proton-proton collisions.

Acknowledgements
The work of W. Redjeb was sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA)

References

[1] 2020 *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report* CERN Yellow Reports: Monographs (Geneva: CERN) URL `https://cds.cern.ch/record/2749422`

[2] CMS Collaboration, The CMS Experiment at the CERN LHC 2008 *JINST* **3** S08004

[3] CMS Collaboration (CMS Collaboration) 2017 The Phase-2 Upgrade of the CMS Endcap Calorimeter Tech. rep. CERN Geneva URL `https://cds.cern.ch/record/2293646`

[4] CMS Collaboration 2021 The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger Tech. rep. CERN Geneva URL `https://cds.cern.ch/record/2759072`

[5] Rovere M, Chen Z, Di Pilato A, Pantaleo F and Seez C 2020 *Frontiers in Big Data* **3** 41

[6] Pantaleo F, Rovere M and on behalf of the CMS Collaboration 2023 *Journal of Physics: Conference Series* **2438** 012096 URL `https://dx.doi.org/10.1088/1742-6596/2438/1/012096`

[7] Pantaleo F 2017 *New track seeding techniques for the CMS experiment* Ph.D. thesis CERN

[8] Cacciari M, Salam G P and Soyez G 2012 *The European Physical Journal C* **72** 1–54