

A novel reconstruction framework for an imaging calorimeter for HL-LHC

Leonardo Cristella^{1,*} *on behalf of the CMS Collaboration*

¹CERN

Abstract. To sustain the harsher conditions of the high-luminosity LHC, the CMS collaboration is designing a novel endcap calorimeter system. 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 pileup. In regions characterised by lower radiation levels, small scintillator tiles with individual on-tile SiPM readout are employed. A unique reconstruction framework (TICL: The Iterative CLustering) is being developed to fully exploit the granularity and other significant detector features, such as particle identification and precision timing, with a view to mitigate pileup in the very dense environment of HL-LHC. The inputs to the framework are clusters of energy deposited in individual calorimeter layers. Clusters are formed by a density-based algorithm. Recent developments and tunes of the clustering algorithm will be presented. To help reduce the expected pressure on the computing resources in the HL-LHC era, the algorithms and their data structures are designed to be executed on GPUs. Preliminary results will be presented on decreases in clustering time when using GPUs versus CPUs. Ideas for machine-learning techniques to further improve the speed and accuracy of reconstruction algorithms will be presented.

1 Introduction

The significant increase in the instantaneous and integrated luminosity comes at the price of almost an order of magnitude increase in the number of multiple proton-proton collisions in the same or neighbouring bunch crossings (referred to as pileup), and the significant increase of the radiation levels. Both of these effects pose major challenges for the experiments, which need to be upgraded to cope with the harsher data taking conditions. One of the major CMS upgrades is the replacement of the current electromagnetic and hadronic endcap calorimeters with a high granularity calorimeter (HGCal) [3].

The design of the CMS endcap calorimeter upgrade was motivated by the physics requirements in this region, while preserving radiation tolerance under the harder HL-LHC conditions. The region covered by the endcap calorimeters ($1.5 < |\eta| < 3.0$) is essential for the success of the LHC physics program, where processes initiated by vector boson fusion and exotic signals play a major role. Therefore, the upgraded detector should provide the necessary capabilities to identify single objects with kinematic thresholds similar to the current ones, powerful jet flavour identification (e.g., quark-gluon separation), high- p_T particle

*e-mail: Leonardo.Cristella@cern.ch

identification (“tagging”) where the decay products of the initial particle are often merged into a single jet, and more.

Taking all these motivations under consideration, the most promising detector upgrade is an imaging calorimeter with very fine lateral and longitudinal segmentation, complemented by precision timing capabilities. HGCAL is a sampling calorimeter using extensively silicon sensors ($\sim 6\text{M}$ channels) as active material to achieve radiation tolerance, with the additional benefit of a very high readout granularity. In regions with lower radiation levels, small plastic scintillator tiles with individual SiPMs readout are employed.

2 HGCAL geometry

Each endcap consists of 50 sensor+absorber layers with a total thickness of about $10 \lambda_I$. The first 28 layers form the electromagnetic section, CE-E (about $25 X_0$ and $1.3\lambda_I$). The active element consists solely of silicon sensors of different thicknesses (120, 200, 300 μm) and cell sizes (~ 0.5 , $\sim 1.2 \text{ cm}^2$). The hadronic section, CE-H, is composed of 12 fine sampling layers followed by 10 layers with twice as thick absorbers. In the first eight layers of CE-H, silicon sensors of thickness of 200 or 300 μm , and size of a cell $\sim 1.2 \text{ cm}^2$, alone are used. In the remaining layers, some of the area at larger radius, where the radiation dose is smaller, is instrumented with scintillator tiles. In order to reliably operate the silicon sensors and the scintillator tiles after irradiation, the entire HGCAL detector will be operated at -30°C .

3 HGCAL Local Reconstruction

The HGCAL local reconstruction is designed to be fast and flexible. It proceeds by reconstructing the deposited energy in each single cells and calibrating that to an absolute electromagnetic scale. The product of these steps are the so called “RecHits”, which will serve as the building blocks of the particle shower reconstruction. For the time being, this step is performed on every triggered event globally in the whole HGCAL detector. Events with 200 pileup interactions at HL-LHC operation are expected to produce about $\sim O(10^5)$ RecHits in the HGCAL detector, constituting a very challenging task for the software reconstruction. In addition, the speed requirements at HLT make many of the most-efficient algorithms unsuitable for particle shower reconstruction in HGCAL. Taking these requirements into consideration, it is vital to develop novel reconstruction algorithms which are designed to explore the recent developments in computing resources (e.g., heterogeneous computing).

3.1 Layer cluster formation: the CLUE algorithm

One of the fundamental ingredients of the HGCAL reconstruction is the collection of RecHits in the same HGCAL layer that originate from the same particle, broadly known as clustering, to form the “Layer Clusters” (LC). The CLUE (CLUsters of Energy) [9] algorithm is a fast and GPU-friendly density-based algorithm designed for high granularity calorimeters that is fully compatible with the HGCAL geometry. It features linear scalability and easy parallelization, and aims at reducing the computational challenge of TICL reconstruction by one order of magnitude, as it builds small clusters (~ 10 hits per cluster).

To achieve fast performance, CLUE starts by organizing the RecHits by their proximity in a fixed grid in the $\eta - \phi$ space, and uses a spatial index for efficient neighbourhood queries. Therefore, for each HGCAL layer, a fixed-grid spatial index is constructed, registering each RecHit according to their $\eta - \phi$ coordinates. Then, the clustering procedure can be summarized in three main steps. First, CLUE calculates the local energy density of each RecHit,

defined as:

$$\rho_i = \sum_{j:j \in N_{d_c}(i)} \chi(d_{ij})E_j \quad (1)$$

where d_c is a cutoff distance that can be chosen based on the shower size and the lateral granularity of HGAL, E_j is the weight of point j , and $\chi(d_{ij})$ a convolution kernel, which in the current implementation has the form:

$$\chi(d_{ij}) = \begin{cases} 1 & d_{ij} = 0 \\ 0.5 & d_{ij} \leq d_c \\ 0 & d_{ij} > d_c \end{cases} \quad (2)$$

In the next step, CLUE computes for each RecHit the quantity δ , i.e. the distance to the closest RecHit with higher ρ , and establishes a connection between these two RecHits (important for parallelizing the algorithm).

At the final step, the RecHits are labelled as "seeds", "followers", and "noise". RecHits with $\rho > \rho_c$ and distance $\delta > \delta_c$ are promoted to be seeds, whereas RecHits with $\rho < \rho_c$ and distance $\delta > \delta_o$ are denoted to be "noise". All other RecHits are associated to their closest hit with higher local density, as their follower. The parameters ρ_c , d_c , δ_c , and δ_o are tuned based on physics arguments. The current configuration for silicon and scintillator sensors is summarized in Table 1. With the proposed tuning, the algorithm is extremely robust against noise, and is able to cluster almost all of the particle's deposited energy in the sensitive layers.

	ρ_c	d_c	δ_c	δ_o
silicon	$9 \times \sigma_{\text{Noise}}$	1.3 [cm]	1.3 [cm]	2.6 [cm]
scintillator	$9 \times \sigma_{\text{Noise}}$	$0.0315 [\eta \times \phi]$	$0.0315 [\eta \times \phi]$	$0.063 [\eta \times \phi]$

Table 1. Summary of the current values of CLUE's tunable parameters. The term σ_{Noise} refers to the standard deviation of the expected noise distribution.

4 Particle shower reconstruction in HGAL: "TICL"

The design of HGAL has great potential for the application of advanced pattern recognition techniques. The possibility of a five-dimensional (x, y, z , energy and time) particle shower reconstruction is ideally suited for particle flow algorithms. However, the large channel count and the severe pileup conditions are some of the challenges that require breakthroughs in many areas of the reconstruction chain to fully exploit the HGAL potential, without jeopardising the overall reconstruction timing. The success of this program relies on a coherent effort in all these areas which translates into designing a versatile reconstruction framework to explore, test and validate new approaches.

4.1 The iterative clustering framework

Motivated by the requirements discussed above, "The Iterative CLustering" (TICL) framework was developed. TICL is a modular reconstruction framework designed to fully exploit the HGAL potential by processing the LCs built by CLUE and returning particle properties

and identification probabilities. Figure 1 illustrates the TICL building blocks (components). The highly modular structure of TICL enables the study of different approaches for each step of the reconstruction chain, by simply modifying only the relevant parts. Moreover, TICL follows an iterative approach; separate TICL configurations (“iterations”) can be designed to reconstruct different particle species. Lastly, TICL is conceived with parallel processing in mind, suitable for the upcoming era of heterogeneous computing in high energy physics. In the remaining section, we present in more detail some of the key ingredients of TICL.

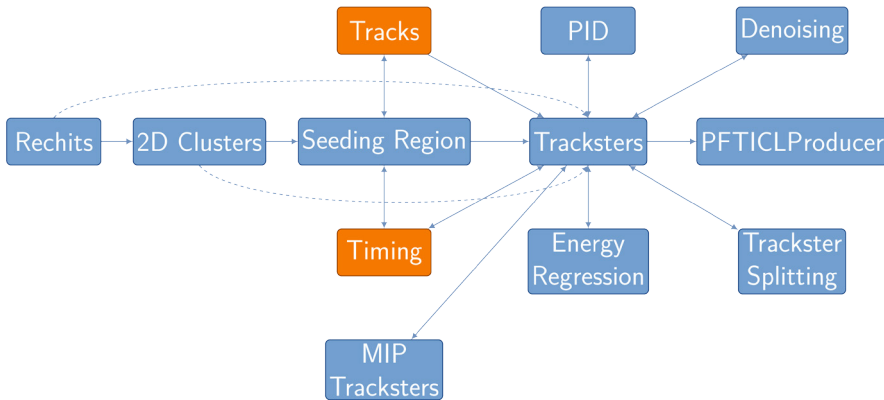


Figure 1. Illustration of the building blocks of TICL. Each box corresponds to a different component of the TICL framework, and the arrows indicate possible connections between the components. The dotted lines represent the alternative approaches to build Tracksters under study.

4.1.1 A TICL iteration

A fundamental TICL ingredient is the “iteration”, that combines information from the various TICL components to reconstruct the particle shower, which within TICL, is referred to as a “Trackster”. The skeleton of a TICL iteration can be summarised as follows:

- **Building blocks:** the LCs returned by CLUE
- **Seeding regions:** identify the spatial regions of interest and the layer clusters compatible with these regions. A seeding region can be global, i.e. it spans the full HGCAL acceptance, or local (e.g., a small region around a track propagated to the HGCAL entrance surface).
- **Pattern recognition:** the algorithm that links together layer clusters among different layers to reconstruct the particle shower i.e., a Trackster. Section 4.1.2 presents details of the current implementation.
- **Linking and classification:** identify the Trackster type and improve the energy measurement with the aid of traditional or machine-learning-based techniques.
- **Masking:** option to mask the layer clusters used in this iteration. This results in a significant reduction of the combinatorics in later iterations, which comes with the advantage of less computing requirements and improved reconstruction performance.

The overall goal of TICL is to follow an iterative approach: first reconstruct simpler objects (e.g., electrons), mask the layer clusters used in this iteration, then reconstruct more complicated ones.

4.1.2 A TICL Trackster

A TICL Trackster aims to link the LC associated to each TICL iteration. It is a Direct Acyclic Graph [6] created by a pattern recognition algorithm, which links LCs to form three-dimensional objects (i.e., the reconstructed particles' showers). Therefore, each vertex of the graph is a LC, and the connections between the vertices are the edges of the graph.

The pattern recognition method currently implemented in TICL is based on the Cellular Automaton (CA) algorithm [7]. The CA algorithm aims to group entities with similar properties (e.g., LCs associated with the same particle) by exploring information between the entity under question and the entities in its neighbourhood. Usually, a fixed rule is applied to all entities simultaneously. The CA implementation in HCGAL reconstruction can be streamlined in three steps.

The first step is responsible for the generation of “doublets”, i.e., connections between successive LCs. For a LC in the HCGAL layer N , LC_N , a search window in layer $N + 1$ is defined. The search window is obtained by projecting the spatial dimensions of LC_N in $\eta - \phi$ to layer $N + 1$. To account for the lateral shower evolution, in conjunction to the HCGAL design specifications, the search window is extended by $\Delta\eta \times \Delta\phi$ of 0.05 (0.1) for $|\eta| < 2.1$ ($|\eta| \geq 2.1$). This is graphically shown in Fig. 2 (a). LCs contained in the search region are connected to LC_N , and form doublets. Timing information [2] is used, whenever it is available, to reject LCs originating from pileup interactions. To account for detector inefficiencies or shower properties, the doublet formation could be carried out even for LCs belonging to non-consecutive layers. The maximum number of missing layers is a tunable parameter and can vary between iterations. Additional selection on the minimum number of RecHits to form a LC can be also applied to suppress LCs stemming from noise.

The second step in the pattern recognition is the doublet linking. Doublets are linked if two angular requirements are satisfied: an angular compatibility between each outermost doublet and the origin of the seeding region, and a minimal alignment condition between the two doublets. The two conditions are illustrated in Fig. 2 (b). The angular requirements may vary between different iterations.

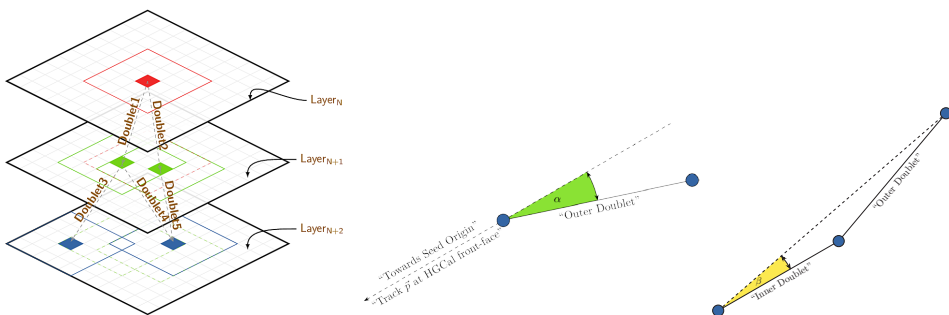


Figure 2. (left) Basic principle of the pattern recognition algorithm used in TICL. Illustration of the angular requirements between doublets; compatibility between the outermost doublet and the origin of the seeding region (centre), and minimal alignment condition between the two doublets (right).

The third and final step in the pattern recognition is to connect all doublets satisfying the above angular requirements to form a Trackster. Ideally a single Trackster should be created for each particle interacting with HCGAL. An event display of two tracksters originating from two electrons is shown in Fig. 3.

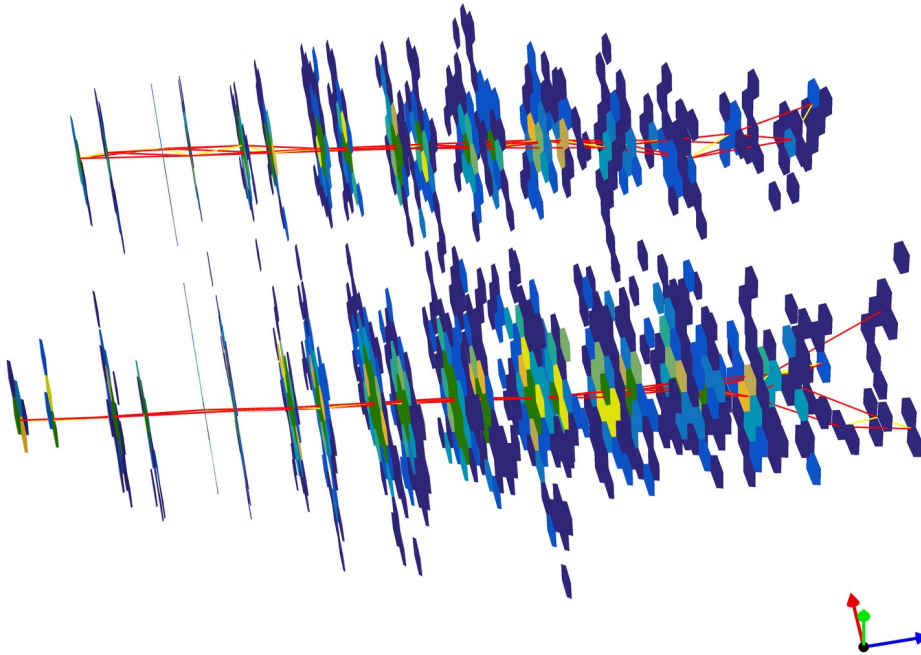


Figure 3. Event display of two tracksters originating from two electrons. The yellow (red) segments connect the edges of a doublet between two (non-)adjacent layers.

4.1.3 The current TICL configuration

There are currently five prototype iterations, which differ in the values adopted for the following parameters:

- layer range, to account for the different longitudinal extension of particle showers;
- minimum size of the input LCs, to suppress the noise;
- seeding region, for showers originating from a charged particle;
- maximum layer distance between a doublet edges;
- maximum time difference between two LCs in a doublet, to filter out LCs stemming from pile-up;
- maximum angle between each outermost doublet and the origin of the seeding region, to account for the different transverse extension of particle showers;
- maximum angle between two doublet directions;
- minimum number of layers in a trackster.

Given that LCs used to build a trackster in a given iteration are masked in the subsequent iterations, the order of the iteration execution is of crucial importance. Currently they follow this order:

1. a track-seeded electromagnetic iteration ("TrkEM"), targeting e^\pm ;
2. an electromagnetic iteration ("EM") targeting γ ;
3. a track-seeded iteration ("TrkHAD") targeting charged hadrons;
4. an hadronic iteration ("HAD") targeting neutral hadrons;
5. a MIP iteration ("MIP") focusing on particles that deposit a very small amount of energy in HGAL, such a muons.

The trackster provenance from different iterations allows to perform a preliminary Particle Flow interpretation of the TICL products:

- all tracksters reconstructed in the EM iteration are labelled as photons;
- all tracksters reconstructed in the HAD iteration are labelled as kaons long;
- tracksters from the TrkEM iteration:
 - if there is another trackster coming from the TrkHAD iteration that is seeded by the very same track, the tracksters are merged and labelled as charged hadron;
 - search the most compatible trackster (in $\eta - \phi$ and p_T space) with the seeding track and label as electron;
 - remaining additional trackster are labelled as photons;
- tracksters from the TrkHAD iteration are labelled as charged hadrons;
- all general tracks which have not been used by TICL are promoted as charged hadrons.

4.1.4 Runtime

An assessment of the current TICL configuration runtime within a realistic event reconstruction has been performed. It accounts for the 4.3% of the total time required by the CMS reconstruction of a $t\bar{t}$ event at a center-of-mass energy of 14 TeV in a PU-200 environment.

4.2 Particle shower identification

The final goal of TICL is to identify the type of particle that initiated the shower and precisely estimate shower energy. To this end, advanced machine learning (ML) techniques are utilised to develop a particle identification (PID) algorithm. The tracksters produced by TICL are used as inputs to a Convolutional Neural Network (CNN): each trackster is represented as a three-dimensional image of $50 \times 10 \times 3$, where each dimension represents the number of HGAL layers per endcap, the maximum number of LC on each layer ordered by decreasing energy, and the number of features (energy, η , ϕ) of each LC. In this representation, each pixel of the image corresponds to an LC that belongs to the Trackster. Zero-padding is applied in layers with less than 10 LC, whereas in layers with more than 10 LC, low energy LC are removed. A preliminary performance study has been conducted on a two-class model: particles generating an electromagnetic or hadronic shower. The dataset consisted of 24k events (12k per particle type) : 70% has been used for training, 20% for validation and the remaining 10% for testing. The CNN was trained for 15 epochs (passes of the algorithm

through the entire dataset), using the sum of categorical cross-entropy and mean squared error as loss function to account for particle ID and energy regression. In order to have the value of the two functions of the same order of magnitude during training, the energies of the Tracksters were normalised with respect to the data sample. The CNN was trained with Tensorflow [11] on a PU=0 sample, and applied both on a PU=0 and 200 sample. Results are shown in Figure 4 and demonstrate the robustness of the algorithm in the two extreme scenarios.

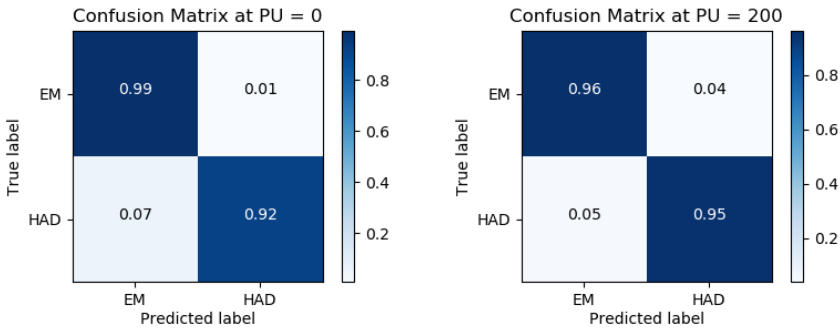


Figure 4. Confusion matrix of the CNN for a PU=0 (left) and PU=200 (right) scenario.

5 Conclusions

Reconstruction in High Granularity Calorimeters at HL-LHC poses many unprecedented challenges, therefore a novel reconstruction algorithm for imaging calorimeters has been developed in CMS: "TICL".

The current TICL configuration within a realistic environment ($\sqrt{s} = 14$ TeV at PU=200) accounts for the 4.3% of the total time required by the CMS reconstruction of a $t\bar{t}$ event. Being developed with parallelism in mind, the runtime is expected to improve on heterogeneous architectures (e.g. GPUs).

TICL provides a fertile ground for application of neural networks and other machine learning algorithms. First results from a CNN are extremely encouraging: single particle identification between EM and HAD is higher than 90% and stable against different PU scenarios.

Next developments will focus on the:

- systematic tuning of thresholds/cuts per TICL iteration,
- improvement of ParticleFlow-objects interpretation,
- local purification of Tracksters from PU contributions.

References

- [1] Apollinari G. et al, “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V.0.1”, CERN-2017-007-M.
- [2] CMS Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, CERN-LHCC-2015-010.
- [3] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, CERN-LHCC-2017-023.
- [4] M. Rovere, Z. Chen, A. Di Pilato, F. Pantaleo and C. Seez, “CLUE: A Fast Parallel Clustering Algorithm for High Granularity Calorimeters in High Energy Physics”, [arXiv:2001.09761 [physics.ins-det]].
- [5] Rodriguez, Alex and Laio, Alessandro, “Clustering by fast search and find of density peaks,” *Science* **344**, 1492–1496 (2014) [<https://science.sciencemag.org/content/344/6191/1492>].
- [6] K. Thulasiraman and M. N. S. Swamy, “Graphs: Theory and Algorithms,” John Wiley & Sons, Inc. (1992) [DOI:10.1002/9781118033104].
- [7] Francesco Berto and Jacopo Tagliabue, “Cellular Automata”, *The Stanford Encyclopedia of Philosophy*, 1095-5054 (2012)
- [8] CMS Collaboration, “A novel reconstruction framework for an imaging calorimeter for HL-LHC”, [CMS-DP-2020/027]
- [9] M. Rovere, Z. Chen, A. Di Pilato, F. Pantaleo and C. Seez, *CLUE: A Fast Parallel Clustering Algorithm for High Granularity Calorimeters in High Energy Physics*, *Frontiers in Big Data*, **3**, DOI:10.3389/fdata.2020.591315 (2020)
- [10] Z. Chen, A. Di Pilato, F. Pantaleo and M. Rovere, *GPU-based Clustering Algorithm for the CMS High Granularity Calorimeter*, *EPJ Web Conf.* **245**, DOI:10.1051/epjconf/202024505005 (2020)
- [11] *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015