Proceedings of CTD 2020
PROC-CTD2020-21
August 2, 2020

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

Bruno Alves, Leonardo Cristella, Loukas Gouskos,
Arabella Martelli, Felice Pantaleo, Marco Rovere

On behalf of the CMS Collaboration,
European Organization for Nuclear Research (CERN)
Esplanade des Particules 1, 1211 Meyrin, Switzerland

ABSTRACT

The CMS endcap calorimeter upgrade for the High Luminosity LHC in 2027 uses silicon sensors to achieve sufficient radiation tolerance, with the further benefit of a very high readout granularity. Small scintillator tiles with SiPM readout are used in regions permitted by the radiation levels. A novel iterative reconstruction framework (TICL) has being developed to fully exploit the granularity and other significant features of the detector such as precision timing. The inputs to the framework are clusters of energy deposited in individual calorimeter layers delivered by a density-based algorithm which has recently been developed and tuned. This talk will describe the approaches being considered and show selected first results.

PRESENTED AT

Connecting the Dots Workshop (CTD 2020)
April 20-30, 2020

1 Introduction

The first ten years of the LHC operation have been characterized by a remarkable success, exceeding initial expectations. The next big milestone on the accelerator front is the high-luminosity (HL) upgrade of the LHC [1]. At the HL-LHC a significant increase of the instantaneous luminosity is expected, which could reach values up to 7.5×10^{34} cm^{-2}s^{-1}, about 5-7 higher than the current one, leading to a total integrated luminosity at the end of the HL-LHC data taking of around 3000 fb^{-1}, 20$\times$ larger than the one collected thus far. This provides a great opportunity for sensitive tests of the standard model parameters and searches for new physics signals. However, the significant increase in the instantaneous and integrated luminosities comes at the price of almost an order of magnitude increase in the number of multiple proton-proton collisions in the same or neighboring 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. To this end, the CMS experiment has planned an extensive upgrade program [2]; the upgraded detectors are designed not only to sustain the HL-LHC conditions, but also to fully exploit the HL-LHC physics potential thanks to their novel capabilities. One of the major CMS upgrades is the replacement of the current electromagnetic and hadronic endcap calorimeters with a high granularity calorimeter (HGCAL) [3].

2 The CMS HGCAL upgrade

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 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 ($\sim$ 6M 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.

Each endcap consists of 50 sensor+absorber layers with a total thickness of about 10 λ_I. The first 28 layers form the electromagnetic section, CE-E (about 25 X_0 and 1.3λ_I). The active element consists solely of silicon sensors of different thicknesses (120, 200, 300 μm) and cell sizes ($\sim$0.5, $\sim$1.2 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 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 HCGAL detector will be operated at -30°C.

3 Particle reconstruction in HGCAL

The design of HGCAL 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 HGCAL potential, without jeopardizing 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.

3.1 The iterative clustering framework

Motivated by the requirements discussed previously, we developed "The Iterative CLustering" (TICL) framework. TICL is a modular reconstruction framework designed to fully exploit the HGCAL potential by processing the energy deposited by particles on the HGCAL sensors (RecHits) and returning particle properties and probabilities. Figure 1 illustrates the TICL building blocks (components). The highly modular structure of TICL enables us to study 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.

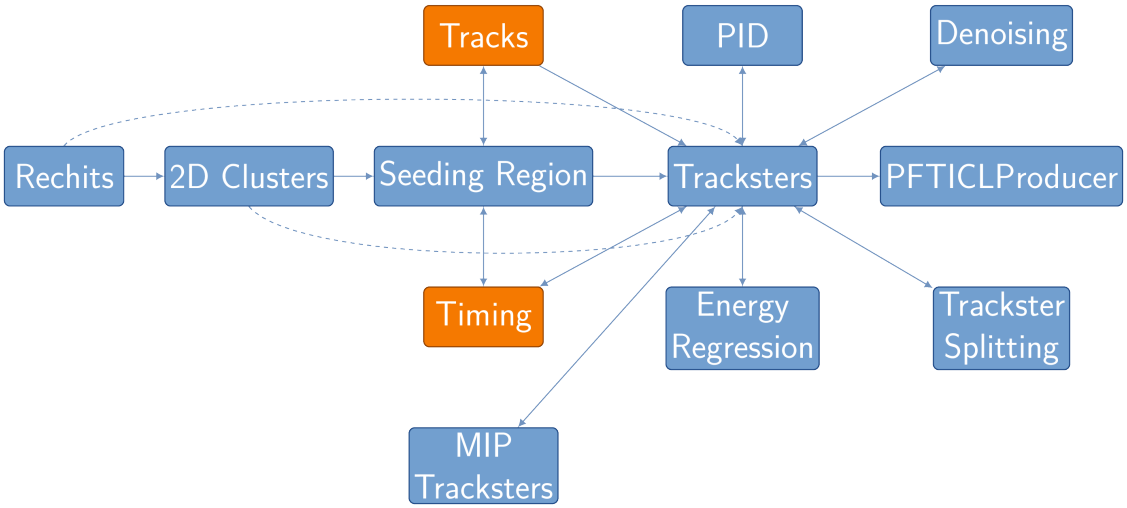


Figure 1: Illustration of the building blocks of TICL. Each box corresponds to a different component of the TICL framework, and the arrows indicate possible connections between the components.

3.1.1 Layer cluster formation ("2D clustering")

One of the fundamental ingredients of TICL is grouping RecHits in the same HGCAL layer that originate from the same particle, broadly known as clustering, to form the "Layer Clusters". The clustering relies on the "clustering by energy" (CLUE) algorithm [4]. CLUE, inspired by [5], is a fast and fully parallelizable algorithm, exploting the concept of local energy density flow.

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:

$$\rho = \sum_j \chi(d_{ij}) E_j, \tag{1}$$

where i is the current RecHit, j the RecHits within a distance d_c (including RecHit i), E_j the energy of each RecHit, and $\chi(d_{ij})$ a convolution kernel, which in the current implementation has the form:

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

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 $\rho \geq \rho_c$ and distance $\delta \geq \delta_c$ are promoted to be seeds, whereas RecHits with $\rho < \rho_c$ and distance $\delta \geq \delta_0$ are denoted to be "noise." All other RecHits are associated to their closest seeds. The parameters ρ_c, d_c, δ_c, and δ_0 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, while, as it will be seen in Section 4, is able to cluster almost all of the particle's deposited energy in the sensitive layers.

	ρ_c	d_c	δ_c	δ_0
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 one standard deviation of the expected noise distribution.

3.1.2 Particle shower reconstruction ("TICL iteration")

Another fundamental ingredient of TICL is the "iteration", which 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 summarized as follows:

- **Building blocks:** the layer clusters returned by CLUE
- **Seeding regions:** identify the 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 3.1.3 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.

3.1.3 A TICL Trackster

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

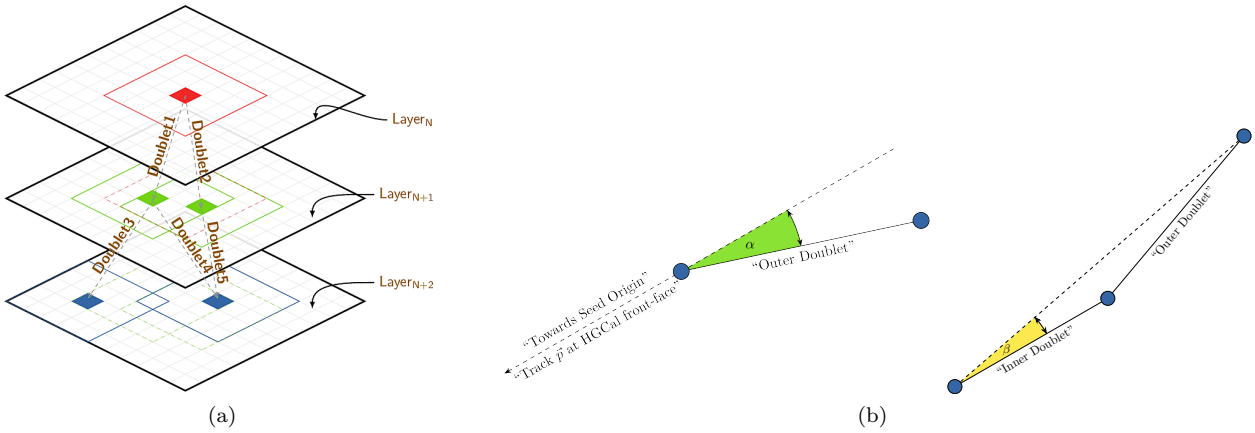
The pattern recognition method currently implemented in TICL is based on the Cellular Automaton (CA) algorithm [7]. The CA algorithm aims to group entities with similar properties (e.g., layer clusters associated with the same particle) by exploring information between the entity under question and the 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 layer clusters. For a layer cluster LC_N in the HGCAL layer N, a search window in layer $N+1$ is opened. 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 HGCAL 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). Layer clusters contained in the search region are connected to LC_N, and form doublets. Timing information is used, whenever it is available, to reject layer clusters originating from pileup interactions. To account for detector inefficiencies or the shower properties, the doublet formation could be carried out even for layer clusters 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 layer cluster can be also applied to suppress layer clusters 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.

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 HGCAL. There are currently four prototype iterations: the track-seeded iteration, targeting $e^{\pm}$ and charged hadrons; the electromagnetic iteration targeting γ; the hadronic iteration targeting neutral hadrons; finally, the MIP iteration focusing on particles that deposit a very small amount of energy in HGCAL, such a muons.



(a) (b)

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

4 TICL performance with single particles

Preliminary results of the single-particle energy response and resolution in events with no pileup interactions, for showers reconstructed using the latest TICL developments, are presented [8]. No corrections are applied for ECAL rear-leakage (for photons) or for hadronic non-compensation for pions (i.e. software compensation). Moreover, the results shown in this section use information solely from HGCAL, without combining information from other CMS subdetectors.

In order to access the performance of individual steps of the TICL reconstruction, different energy sums are compared:

- **Rec/able:** the reconstructable energy, obtained by summing the energy of all RecHits produced by the generated particle

- **Clustered:** the sum of the energy in the layer clusters returned by the CLUE algorithm having at least one RecHit associated with the generated particle
- **TICL:** the sum of the energy in the shower reconstructed using TICL
- **TICL (non-interacting):** the TICL sum where the showers are restricted to those resulting from pions that have not interacted in the tracker volume in front of HGCAL.

The TICL performance on electromagnetic shower reconstruction is evaluated using events with a single photon (γ). Figure 3 shows the energy response and resolution as a function of the generated γ energy. The γs are generated at $|\eta| \sim 1.9$ directly in front of HGCAL and thus reach HGCAL unconverted. It can be seen that CLUE manages to cluster almost all of the deposited energy. The showers reconstructed by TICL contain a large fraction of this, although the fraction varies with energy and the non-linearity induced will require correction. To compare the energy resolution consistently across the different energy sums, the resolution in each energy sum is corrected for the differences observed in the response. The correction has a negligible impact on the results. The energy resolution is fitted with a two-parameter function with the form: $\sigma_E/E = s/\sqrt{E} \oplus c$, where s (c) is the stochastic (constant) term. No noise term is included in the fit since its magnitude is negligible. The results of the fit are summarized in Table 2. The constant term for the "reconstructable" energy is higher than expected due to rear leakage which will have to be corrected for. This is confirmed by selecting events with almost no energy in the last layer of CE-E. In those cases, the constant term for the Rec/able case drops to 0.6%.

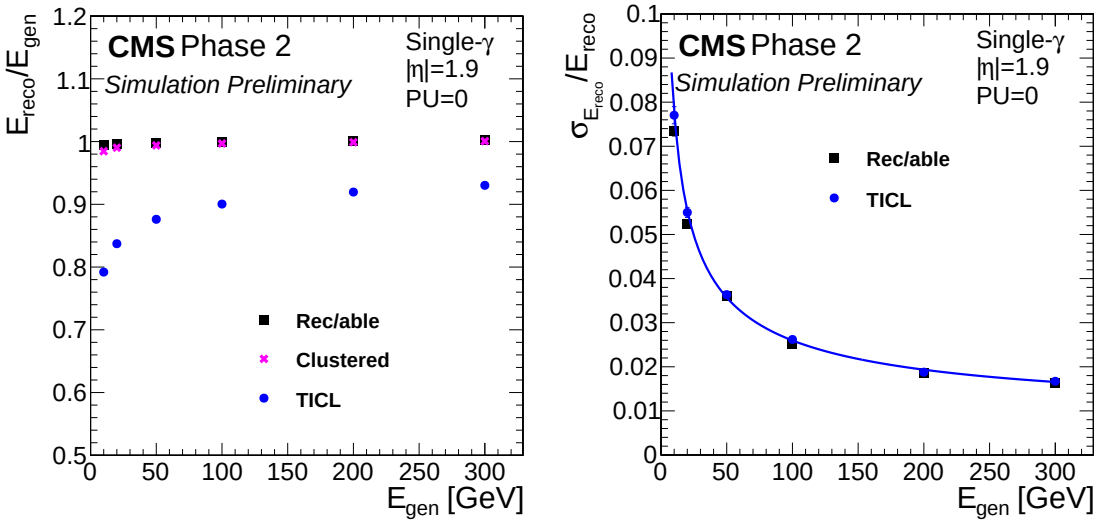


Figure 3: Energy response (left) and resolution (right) as a function of the generated γ energy. The different scenarios are detailed in the text.

The TICL performance in reconstructing hadronic showers is evaluated using events with single-π^+, generated at $|\eta| \sim 1.9$. Figure 4 shows the energy response and resolution as a function of the generated π^+ energy. The effect of hadronic non-compensation can be seen in the Rec/able energy, almost all of which is clustered. The showers reconstructed by TICL contain a large fraction of this energy. Similarly to the electromagnetic showers, the energy resolution as a function of energy is fitted with a two-parameter function of the same form. The results are summarized in Table 2.

The HGCAL layers are instrumented with different types of active elements, i.e., silicon sensors of different size and thickness in the CE-E, or even different types (silicon sensors and scintillator tiles) in the CE-H. The exact configuration depends on the layer depth and η. Therefore, it is important to evaluate the robustness of TICL reconstruction performance across η. Figure 5 shows the energy resolution of single$-\gamma$ and single$-\pi^+$ with $E = 100$ GeV, as function of the particle's η, where stable performance is observed.

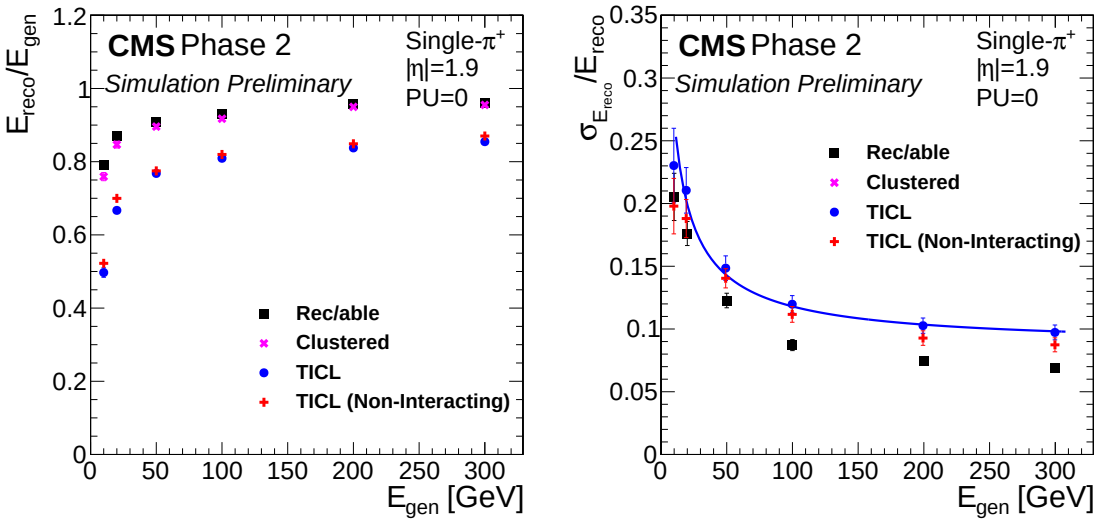


Figure 4: Energy response (left) and resolution (right) as a function of the generated π^+ energy. The different scenarios are detailed in the text.

	Rec/able	Clustered	TICL	TICL (non-interacting)
	Electromagnetic showers			
stochastic term [%]	23	23	25	-
constant term [%]	0.9	0.9	0.9	-
	Hadronic showers			
stochastic term [%]	72	72	80	72
constant term [%]	5.5	5.9	8.7	8

Table 2: Results of the two-parameter fit: $\sigma_E/E = s/\sqrt{E} \oplus c$, in the energy resolution achieved on electromagnetic and hadronic showers using the TICL framework. The different scenarios are detailed in the text. The constant term for the reconstructable energy in electromagnetic showers improved to 0.6% when events with no energy released beyond the last layer of CE-E are selected.

Residual differences in performance across η are due to the different HGCAL configurations discussed earlier in the section. Further tuning of the reconstruction algorithm is foreseen to account for these effects.

5 Conclusions

We present a novel reconstruction framework, "TICL", for imaging calorimeters like HGCAL. TICL has been designed as a modular framework to exploit the full potential of such detectors. Preliminary studies on single particle reconstruction in events with no pileup were presented showing very good performance. Additional activities, like heterogeneous computing and advanced machine learning techniques are also worked out in parallel.

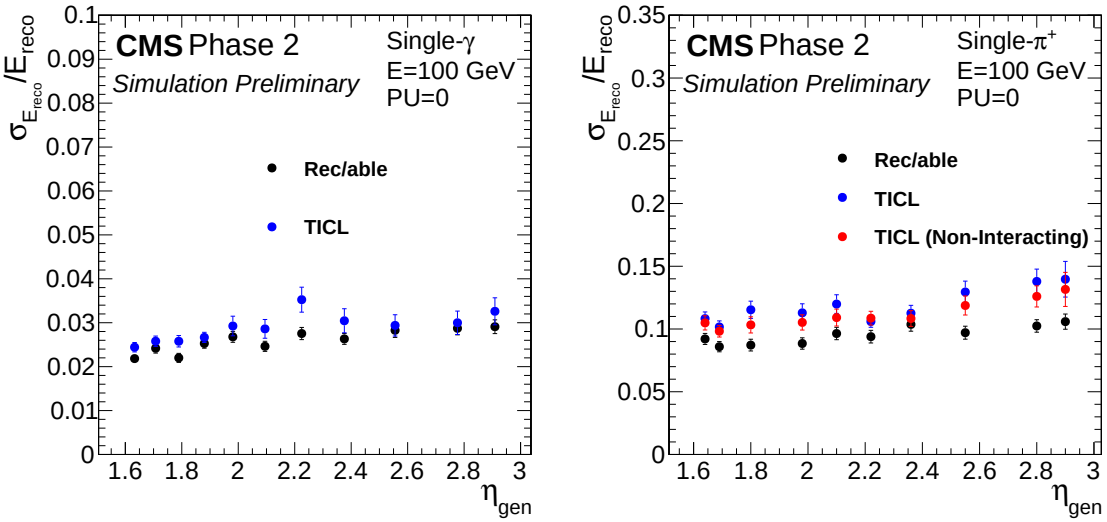


Figure 5: Relative energy resolution as a function of η of the generated γ (left) and π^+ (right). Both particles are generated with $E = 100$ GeV in a sample with no pileup interactions.

ACKNOWLEDGEMENTS

The authors would like to thank Chris Seez for the helpful discussions and suggestions received while developing TICL.

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]