Available on CMS information server

CMS DP -2024/124

The Compact Muon Solenoid Experiment

CMS Performance Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

02 December 2024 (v2, 16 December 2024)

The Iterative CLustering (TICL) (v5a) reconstruction at the CMS Phase-2 High Granularity Calorimeter Endcap

CMS Collaboration

Abstract

In this note we show the development and results achieved with the new version of The Iterative CLustering (TICL) (version 5a), being developed for particle flow reconstruction in CMS during Phase 2 of the LHC, using the High Granularity Calorimeter (HGCAL). Version 5a means that is the alpha version, the first pre-release of TICLv5 as its development is still ongoing and will be closed at the end of 2024. This new version consists in two new algorithms added for the Linking step, better usage of the timing information and a better TICLCandidate reconstruction. The first algorithm introduced is dedicated for the E/Gamma reconstruction, and it consists of a DNN that replaces the current Mustache Superclustering for the electron and photon superclustering. The second algorithm, instead, replaces the Geometrical Linking algorithm that was introduced with TICLv4. This new algorithm recovers the inefficiencies of the old Geometrical Linking, trying to collect as much energy as possible looking also around the core of the shower. Lots of improvements have been made also on the HGCAL Trackster time information and on its usage. The final TICLCandidate time resolution has improved and with it also the Track-to-Trackster linking.

The Iterative CLustering (TICL) (v5a) reconstruction at the CMS Phase-2 High Granularity Calorimeter Endcap

The CMS Collaboration

cms-dpg-conveners-hgcal@cern.ch

Abstract

In this note we show the development and results achieved with the new version of **T**he **I**terative **CL**ustering (TICL) (version 5a), being developed for particle flow reconstruction in CMS during Phase 2 of the LHC, using the High Granularity Calorimeter (HGCAL).
Version 5a means that is the alpha version, the first pre-release of TICLv5 as its development is still ongoing and will be closed at the end of 2024.
This new version consists in two new algorithms added for the Linking step [5], better usage of the timing information and a better TICLCandidate reconstruction.
The first algorithm introduced is dedicated for the E/Gamma reconstruction, and it consists of a DNN that replaces the current Mustache Superclustering [6] for the electron and photon superclustering.
The second algorithm, instead, replaces the Geometrical Linking [5] algorithm that was introduced with TICLv4. This new algorithm recovers the inefficiencies of the old Geometrical Linking, trying to collect as much energy as possible looking also around the core of the shower.
Lots of improvements have been made also on the HGCAL Trackster time information and on its usage. The final TICLCandidate time resolution has improved and with it also the Track-to-Trackster linking.

Glossary and Notation

- DetId : HGCAL detector Id, corresponds to a detector cell
- $\mathbf{fr}_{\mathbf{k}}^{\mathbf{reco}}$: is the fraction of the energy assigned by the pattern recognition to the reconstructed object (0 if the cell does not belong to the reconstructed object)
- $\mathbf{fr}_{\mathbf{k}}^{\mathbf{MC}}$: is the fraction of the energy belonging to the Monte Carlo object derived by the simulation
- ϵ_k: is the energy deposited in each single cell
- CP_i: CaloParticle i . The CaloParticle corresponds to the Monte Carlo Truth Information
- CE-E: calorimeter endcap electromagnetic section
- CE-H: calorimeter endcap hadronic section

Sim-to-Reco score

Shared DetID between two objects.

- Score = 1 completely unmatched objects
- Score = 0 completely matched objects

$$\text{score}_{i,j} = \frac{\sum_{\text{DetId}_k \in \text{CP}_i} \min\left(\left(fr_k^{reco} - fr_k^{MC}\right)^2, \left(fr_k^{MC}\right)^2\right) \times \epsilon_k^2}{\sum_{\text{DetId}_k \in \text{CP}_i} \left(fr_k^{MC}\right)^2 \times \epsilon_k^2}$$

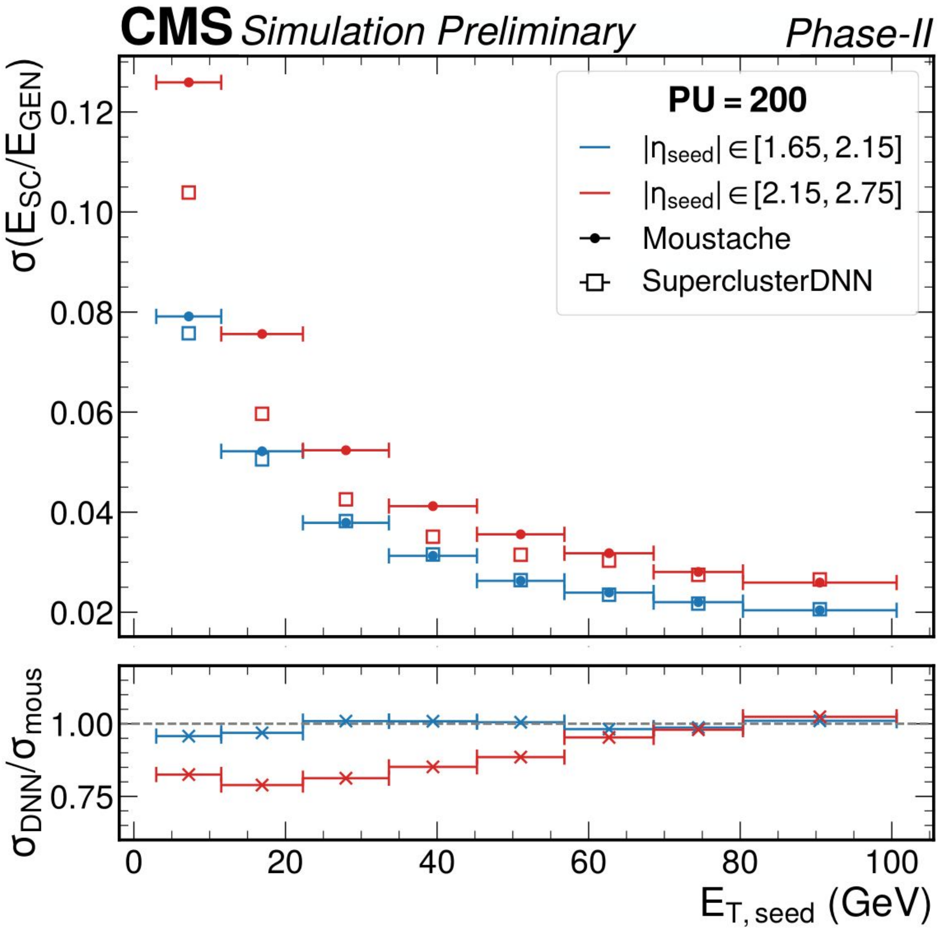
Efficiency: Number of Simulated objects that share at least the 50% of the energy with a single Reconstructed object, divided by the number of generated objects

Reco-to-Sim score

$$\text{score}_{i,j} = \frac{\sum_{\text{DetId}_k \in \text{recoObject}_i} \min\left(\left(fr_k^{reco} - fr_k^{MC}\right)^2, \left(fr_k^{MC}\right)^2\right) \times \epsilon_k^2}{\sum_{\text{DetId}_k \in \text{recoObject}_i} \left(fr_k^{reco}\right)^2 \times \epsilon_k^2}$$

Linking Step - HGCAL Super Clustering

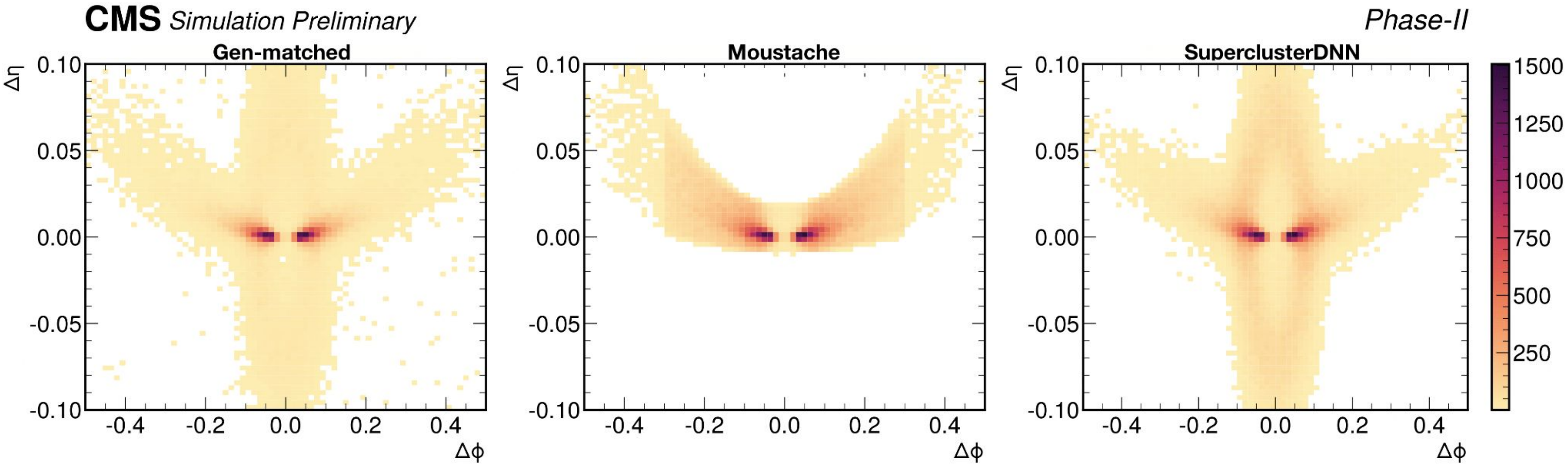
- **TICLv4 [5]**: SuperClustering algorithm used in ECAL: Mustache SuperClustering [6]
  - Prone to collect pileup (PU) if it falls in the seed **$\Delta\eta$-$\Delta\Phi$** window
  - Designed for homogeneous calorimeter
- **TICLv5**: DNN to keep or reject Tracksters within the seed window
  - Seed chosen as Mustache [6] algorithm → highest p_T trackster
    - Min seed p_T = 4GeV
    - Min trackster energy = 2GeV
    - Cut on explained variance on input trackster
  - Exploit HGCAL shower variables, mostly Trackster direction given by Principal Component Analysis (PCA)



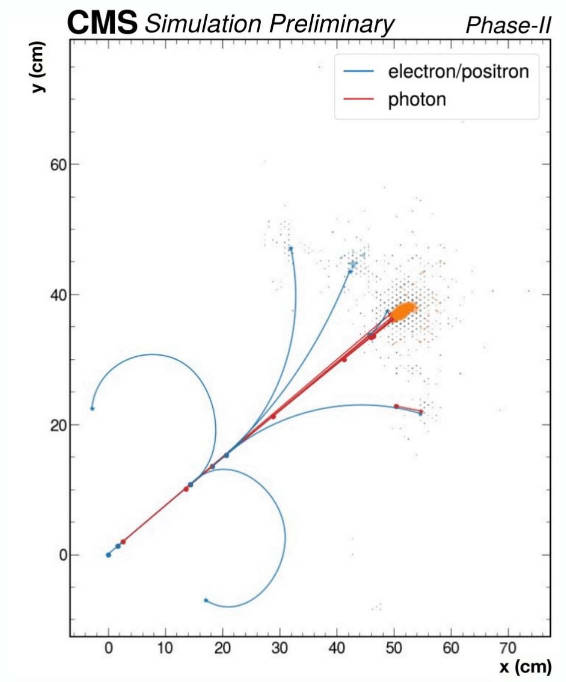
Caption: Resolution of the Supercluster energy (E_{SC}) divided with the gen-level particle energy (E_{GEN}) coming from generated electrons, as a function of the transverse energy of the highest energy trackster ($E_{T,seed}$) called the seed trackster.

Superclusters are formed with Moustache and SuperclusterDNN algorithms and their energy resolution is compared. The resolution is computed by fitting the E_{SC}/E_{GEN} distribution with a Cruijff function and taking σ = (σL + σR)/2, σL being the left sigma of the Cruijff function and σR the right one. The lower panel shows the ratio of the resolution between the two algorithms.

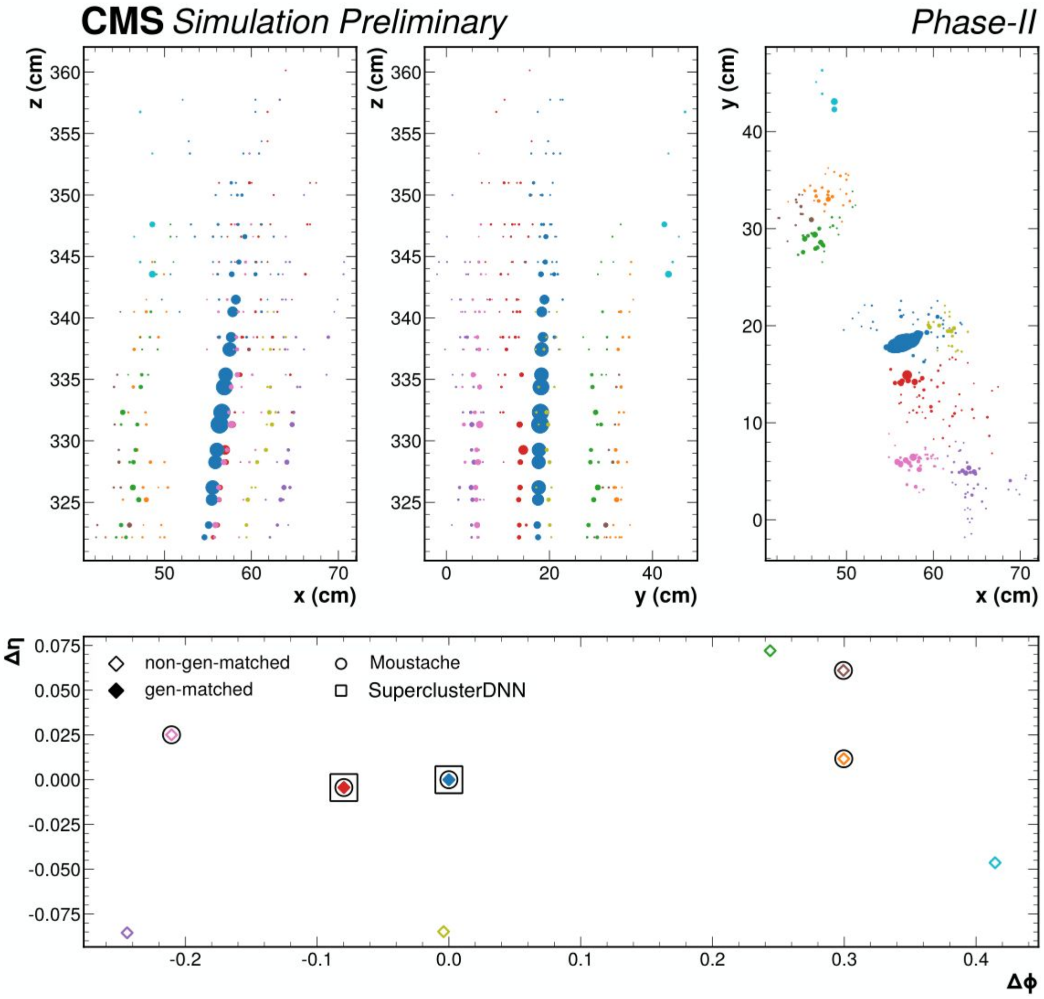
SuperclusterDNN utilises a combination of angular variables computed using trackster direction estimates extracted from Principal Component Analysis from cleaned tracksters (CMS-DP-2022/057). Tracksters around the seed trackster are associated together by the ML model and the supercluster is reconstructed. Utilising the 3D information from HGCAL, SuperclusterDNN shows improved energy resolution arising from improved trackster collection and pileup resilience, in both the high density and the low density regions of the calorimeter with respect to Moustache, which performs geometrical association based on the position and the energy of the seed and the nearby tracksters.



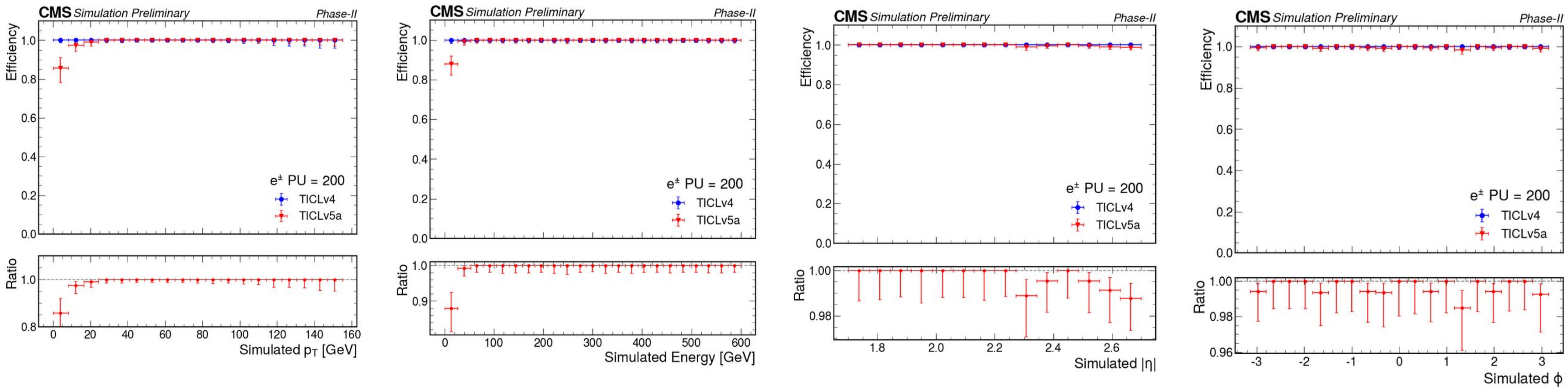
Caption: Distributions of the $\Delta\eta = |\eta| - |\eta_{seed}|$ versus $\Delta\varphi = \varphi - \varphi_{seed}$ variables for gen- matched trackster (left), tracksters forming superclusters using the Moustache algorithm (centre), and tracksters forming superclusters using the SuperclusterDNN algorithm (right) for electrons. The seed is defined as the trackster with the largest transverse energy.



Caption: Representation of the trajectories of electron emissions before reaching the HGCAL in the transverse plane. The electron is generated at (x,y,z) = (0,0,0). Each point represents either a bremsstrahlung emission or a pair production. The final points that are not connected to any other lines represent the entrance into the HGCAL, located at |z| ≃ 321 cm.



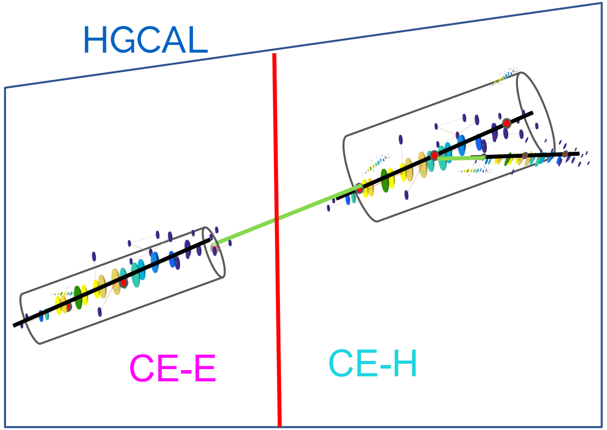
Caption: (Top) Event display in the HGCAL of an electron. Each circle represents a Layer Cluster (LC) and its radius is proportional to its energy. The LCs forming a trackster are identified with the same colour. (Bottom) Representation of the same event in the $\Delta\eta = |\eta| - |\eta_{seed}|$ versus $\Delta\varphi = \varphi - \varphi_{seed}$ plane. The gen-matched tracksters are identified with a filled diamond, while non-gen-matched tracksters are empty. If the trackster is part of the supercluster formed by Moustache, the corresponding diamond is surrounded by a black circle. If the trackster is part of the supercluster formed by the SuperclusterDNN, the corresponding diamond is surrounded by a black square. In this specific case, the gen-matched tracksters are captured by both the Moustache and the SuperclusterDNN algorithms. Three not gen-matched tracksters are clustered by Moustache but effectively rejected by the SuperclusterDNN. Kinematic information: E_{GEN} =119 GeV, $E_{Moustache}$ = 142 GeV, $E_{SuperclusterDNN}$ = 122 GeV.



Caption: The performance of Linked Tracksters is evaluated. The efficiency for the case of a single electron in PU 200, produced from the vertex, uniformly in the η range [1,7, 2.7] and uniformly in the energy range [10, 600] GeV. In blue the results for the current reconstruction (TICLv4), while in red the results for the new TICL reconstruction (TICLv5a) after the new Superclustering DNN.
The efficiency is shown against pT, energy, η, and Φ (from left to right).
Both TICL versions show excellent performance concerning the reconstruction efficiency. In the very low energy part < 20 GeV, TICLv5a shows some inefficiency that can be explained by the quality cut on the seed p_T (>4GeV) and on the minimum trackster energy (2 GeV) applied during the Superclustering.

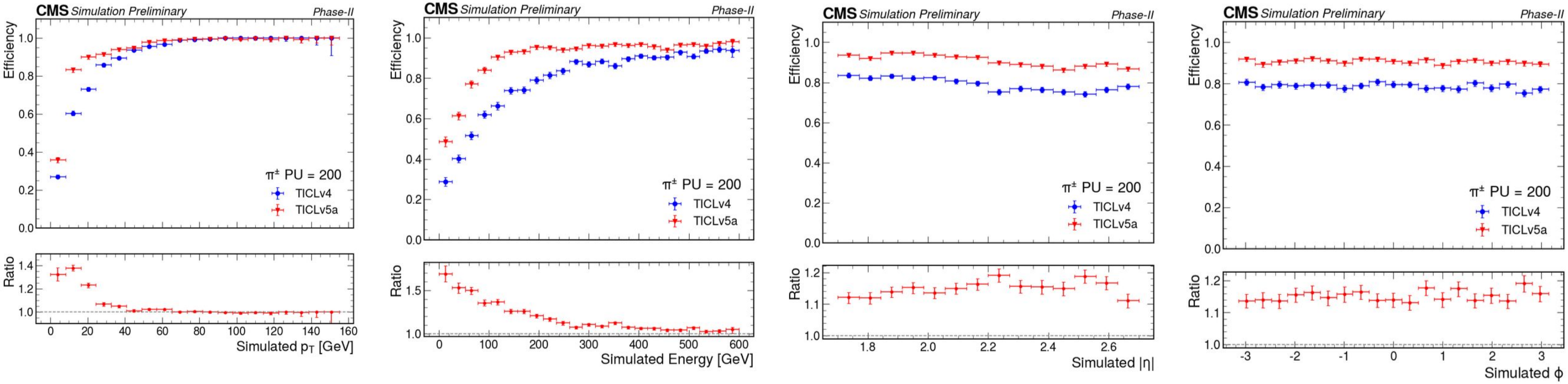
Hadronic Linking - Skeletons

- Hadronic Showers are quite fragmented
  - tens of energy blobs per each hadronic particle
- TICLv4 was doing a great job in connecting aligned energy blobs
  - But not designed to collect smaller Tracksters
- A first Skeleton Linking algorithm aims to improve the energy collection
  - Additional logic to gather fragments around the shower core
  - Looks at alignment between Tracksters' longitudinal axes
  - Collects small tracksters not aligned but close to the shower core
- TICL Graph introduced
  - Graph of connected tracksters
- To additionally improve the energy collection an additional collection of Tracksters has been introduced
  - Recovery Tracksters: Layer Clusters not belonging to a Trackster



CE-E: calorimeter endcap electromagnetic section

CE-H: calorimeter endcap hadronic section

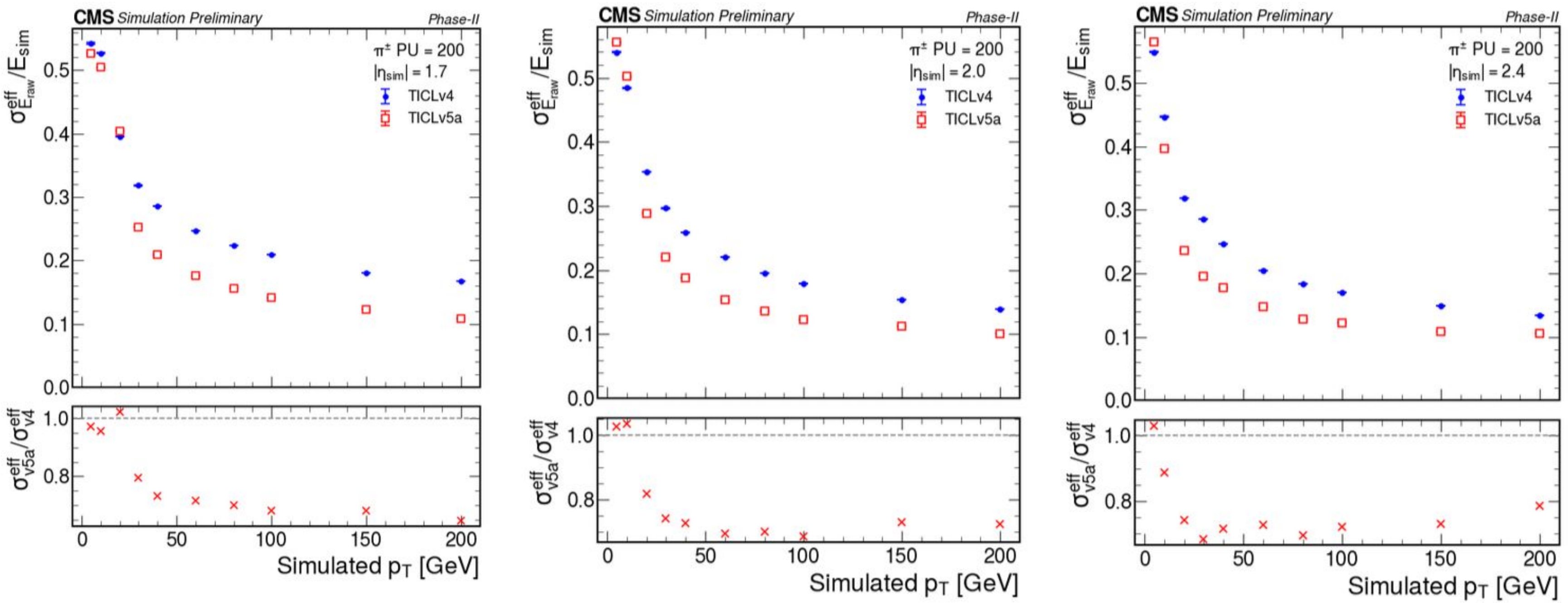


Caption: Linked Tracksters efficiency for the case of a single pion in PU 200, produced at the front face of the HGCAL, uniformly in the η range [1,7, 2.7] and uniformly in the energy range [10, 600] GeV. In blue the results for the Current reconstruction (TICLv4), while in red the results for the new TICL reconstruction (TICLv5a)
The efficiency is show against pT, energy, η, and Φ (from left to right)
The TICLv5a version shows quite big improvements in term of efficiency, especially in the energy region below 200GeV.

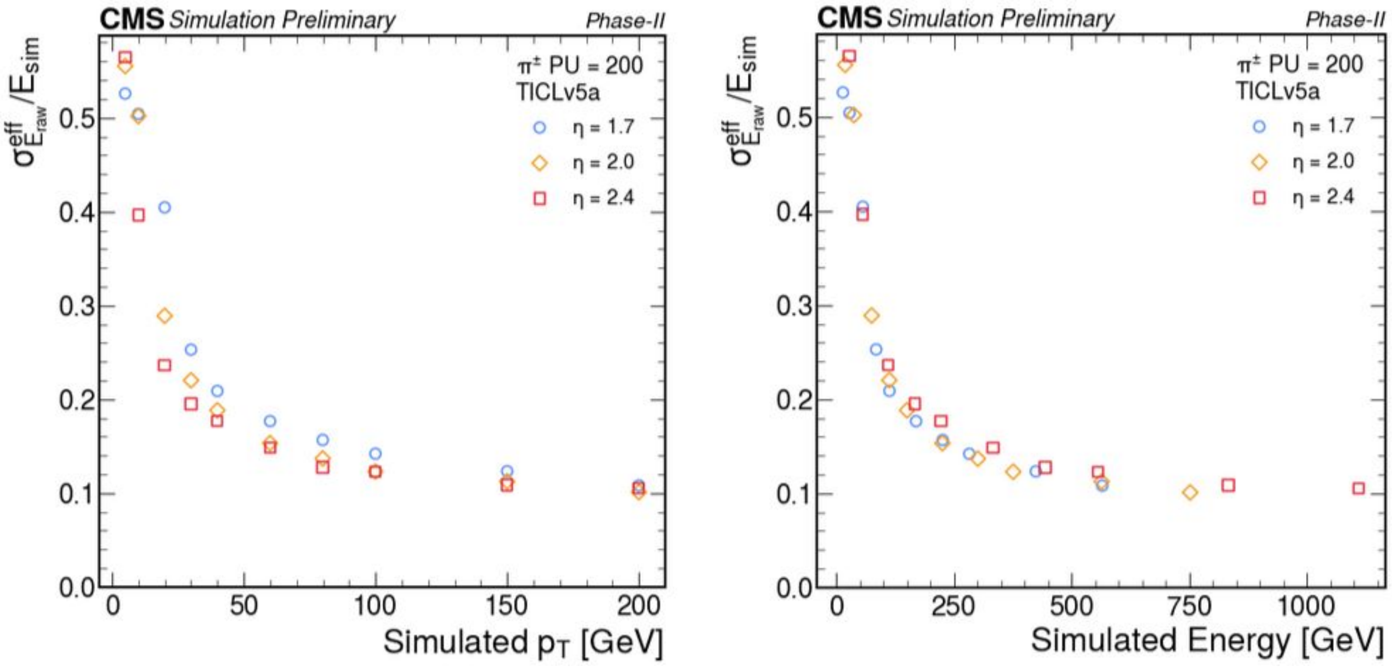
For the energy resolution we decided to adopt the sigma effective

Sigma Effective: Smallest width around the central value of a distribution that contains the 68.3% (one sigma) of the total area under the curve

Robust against outliers and long tails



Caption: Linked Tracksters raw deposited energy resolution for the case of a single pion in PU 200. The data has been generated for different bin in pT [2,5,10,20,40,60,80,100,150,200] GeV at three different eta values [1.7, 2.0, 2.4]. In blue the results for the Current reconstruction (TICLv4), while in red the results for the new TICL reconstruction (TICLv5a)
The raw deposited energy resolution has been computed using the sigma effective on the raw energy distribution, and it is shown in function of the transverse momentum.
The TICLv5a reconstruction shows an improvement between 20 and 30% for p_T > 20 GeV, while it remains almost identical in the low p_T region

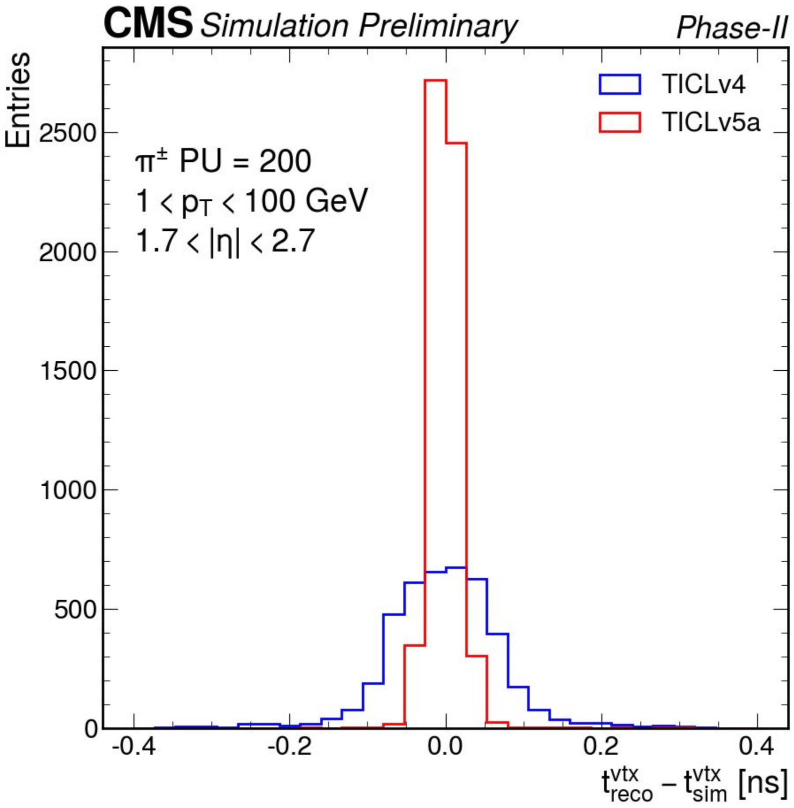


Caption: Linked Tracksters raw deposited energy resolution for the case of a single pion in PU 200 for the TICLv5a reconstruction. The data has been generated for different bin in p_T [2,5,10,20,40,60,80,100,150,200] GeV at three different eta values [1.7, 2.0, 2.4], respectively blue, orange and red.

The raw deposited energy resolution has been computed using the sigma effective on the raw energy distribution, and it is shown in function of the transverse momentum (left) and in function of the energy (right).

TICLCandidates and time information

- TICLCandidates are produced in the last module of TICL
  - Linking between track and trackster(s)
  - Vector of particle Id probabilities
  - Combination with Endcap Timing Layer (ETL) time
- TICLv4 was using the time at the origin assuming propagation in a straight line at the speed of light
  - Bigger errors and bias in the propagation
  - Affected the performance in the linking with the tracks
  - Combination with ETL time disabled
- TICLv5 is using the local time in HGCAL
  - Propagation to the vertex is done after the linking with the track
  - Better resolution
  - Combination with ETL time enabled

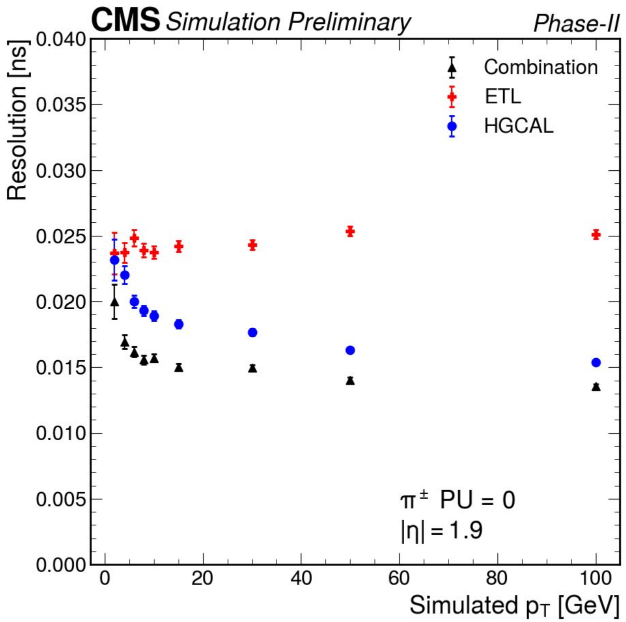


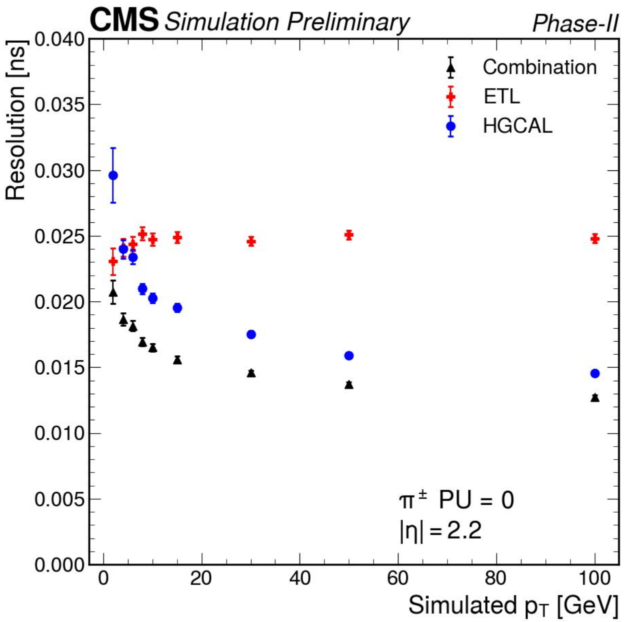
Caption: TICLCandidates time residuals for the case of a single charged pion in PU 200. Pions have been producedd in the p_T range [1,100] GeV and η range [1.7, 2.7], from HL-LHC beamspot.

The residuals have been computed subtracting the simulated vertex time, t^{vtx}_{sim}, from the reconstructed time of the TICLCandidate at the vertex, t^{vtx}_{reco}, restricted to charged candidates. For the association between the simulation and the reconstruction, a shared energy of at least 10% of the simulated trackster energy is required.

In blue there are the residuals of the current version, in red those of TICLv5a.

The improvement in the resolution comes from the use of local time in HGCAL and the propagation to the vertex using the track instead of a straight line to the origin. The number of entries for TICLv5a increased as well as a consequence of the improvements in the linking with the track.





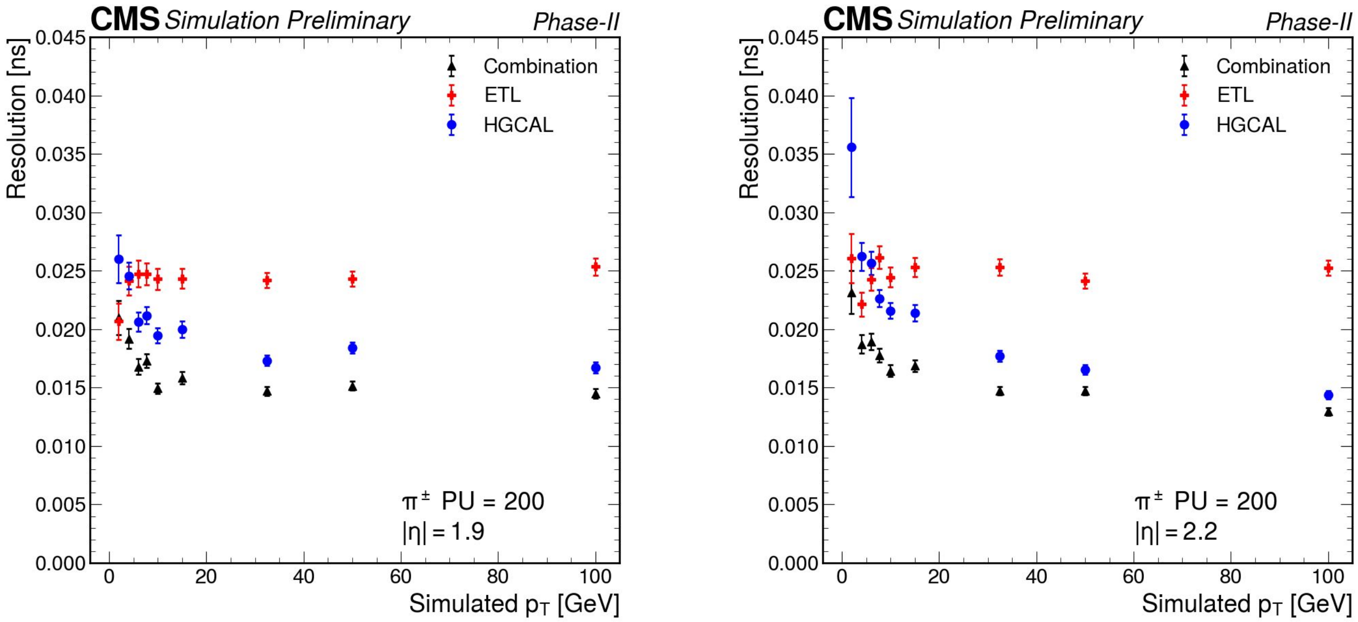
Caption: TICLCandidates time resolution for the case of a single pion in PU 0. The data has been generated for different bin in p_T [2,4,6,8,10,15,30,50,100] GeV at two different eta values [1.9, 2.2]. In red the results of the resolution of the ETL time associated to the TICLCandidate, in blue the HGCAL time, and in black the combination of the two (or only one if the other is not available) which corresponds also to the final TICLCandidate time.

Considering best associated TICLCandidate, not necessarily efficient.

The time resolution has been computed using the sigma of the Cruijff fit of the distribution of the reconstructed time minus the simulated time at the vertex, and it is shown as a function of the transverse momentum.

The ETL resolution does not depend on the energy. On the contrary, the HGCAL resolution improves at higher energies.

Caveat: the current simulation may give an over-optimistic floor for the timing resolution of the HGCAL.



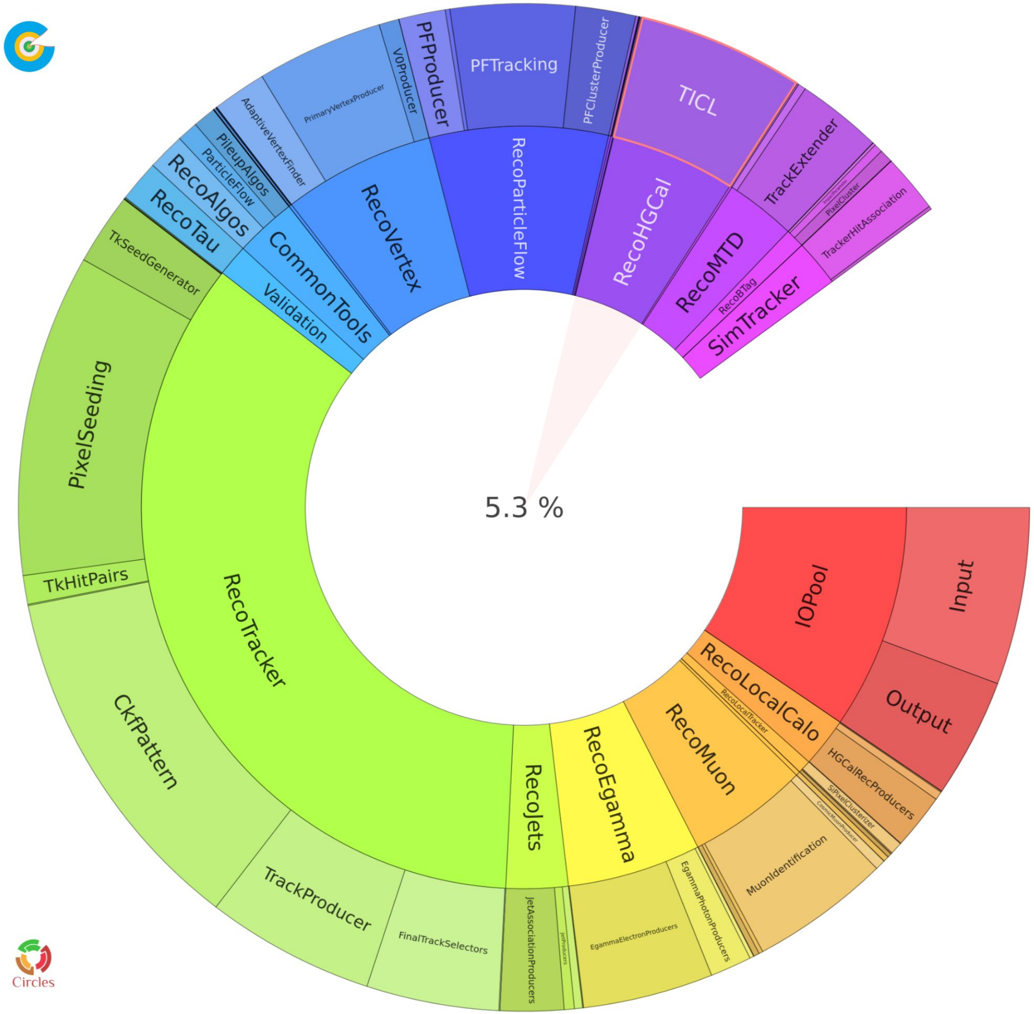
Caption: TICLCandidates time resolution for the case of a single pion in PU 200. The data has been generated for different bin in p_T [2,4,6,8,10,15,30,50,100] GeV at two different eta values [1.9, 2.2]. In red the results of the resolution of the ETL time associated to the TICLCandidate, in blue the HGCAL time, and in black the combination of the two (or only one if the other is not available) which corresponds also to the final TICLCandidate time.

Considering best matched TICLCandidate associated with the hard-scattering CaloParticle.

The time resolution has been computed using the sigma of the Cruijff fit of the distribution of the reconstructed time minus the simulated time at the vertex, and it is shown as a function of the transverse momentum.

The ETL resolution does not depend on the energy. On the contrary, the HGCAL resolution improves at higher energies.

Caveat: the current simulation may give an over-optimistic floor for the timing resolution of the HGCAL.



Real time per event measured with CMS software version CMSSW_14_2_0_pre1 running on sample of ttbar events with PU 200 on a single core of a AMD EPYC 9454 [7].

References

- [1] The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger
- [2] CLUE: A Fast Parallel Clustering Algorithm for High Granularity Calorimeters in High-Energy Physics
- [3] CERN-CMS-DP-2024-033: Use of HGCAL and MTD timing information in TICL reconstruction
- [4] CERN-CMS-DP-2022-057: Reconstruction of unconverted photons in the presence of high pileup at the CMS Phase-2 High Granularity Endcap Calorimeter
- [5] CERN-CMS-DP-2022-045: The TICL (v4) reconstruction at the CMS Phase-2 High Granularity Calorimeter Endcap
- [6] Electron and photon reconstruction and identification withthe CMS experiment at the CERN LHC
- [7] AMD-EPYC-9454