Deep learning techniques for energy clustering in the CMS Electromagnetic Calorimeter

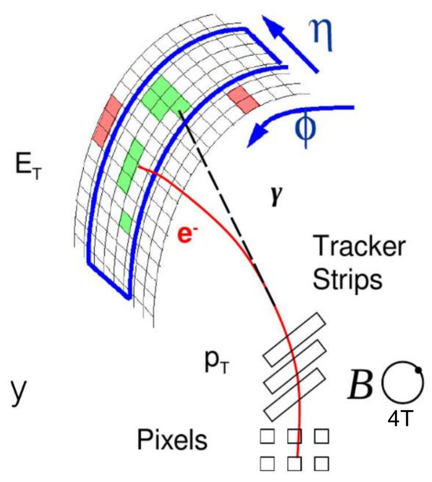
Badder Marzocchi, Polina Simkina, Davide Valsecchi, Jin Wang

ON BEHALF OF THE CMS COLLABORATION

SuperClustering in ECAL

Rechits → Clusters → SuperClusters

Reconstructed energy deposits in the $PbWO_4$ crystals of the calorimeter (rechits) left by particles.



Rechits are gathered together around the crystal with highest deposited energy to form **clusters**.

Each cluster represents a **single particle**.
→ Or several overlapping particles.

Due to bremsstrahlung and photon conversion before the ECAL, the individual clusters have to be combined together to form a **SuperCluster**.

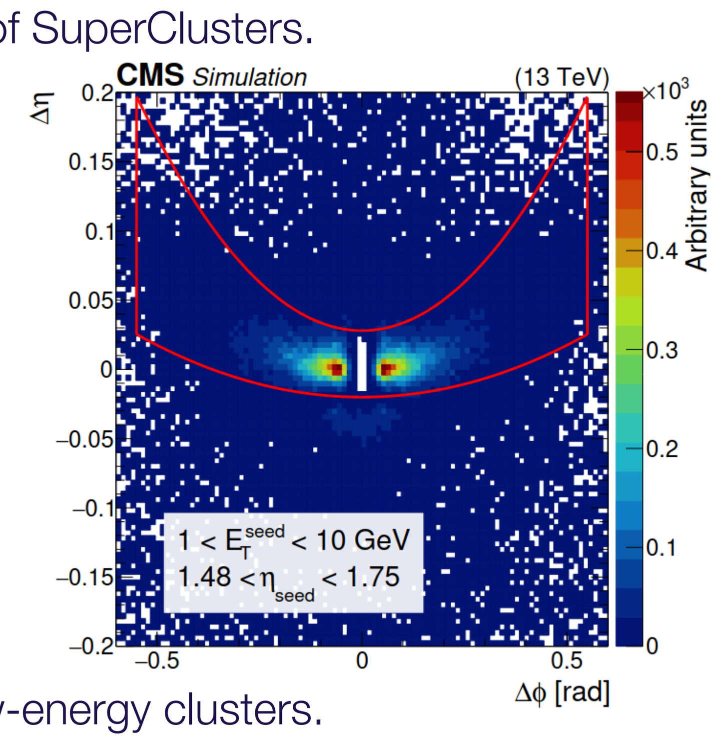
The energy of the initial particle can be reconstructed from the SuperCluster.

"Mustache" algorithm

The algorithm currently used in CMS for reconstruction of SuperClusters.

Purely **geometrical** approach:

- → All the clusters falling into the specified "mustache" shape are considered as part of the SuperCluster. The size of the area depends on energy and position of the seed.
- → "Mustache" shape due to the CMS magnetic field (spread along ϕ).



High efficiency: the algorithm is able to gather even low-energy clusters.

Downside: **suffers from pileup (PU) and noise contamination.**

Energy regression is further applied that can **correct PU and noise on average**.

https://iopscience.iop.org/article/10.1088/1748-0221/16/05/P05014

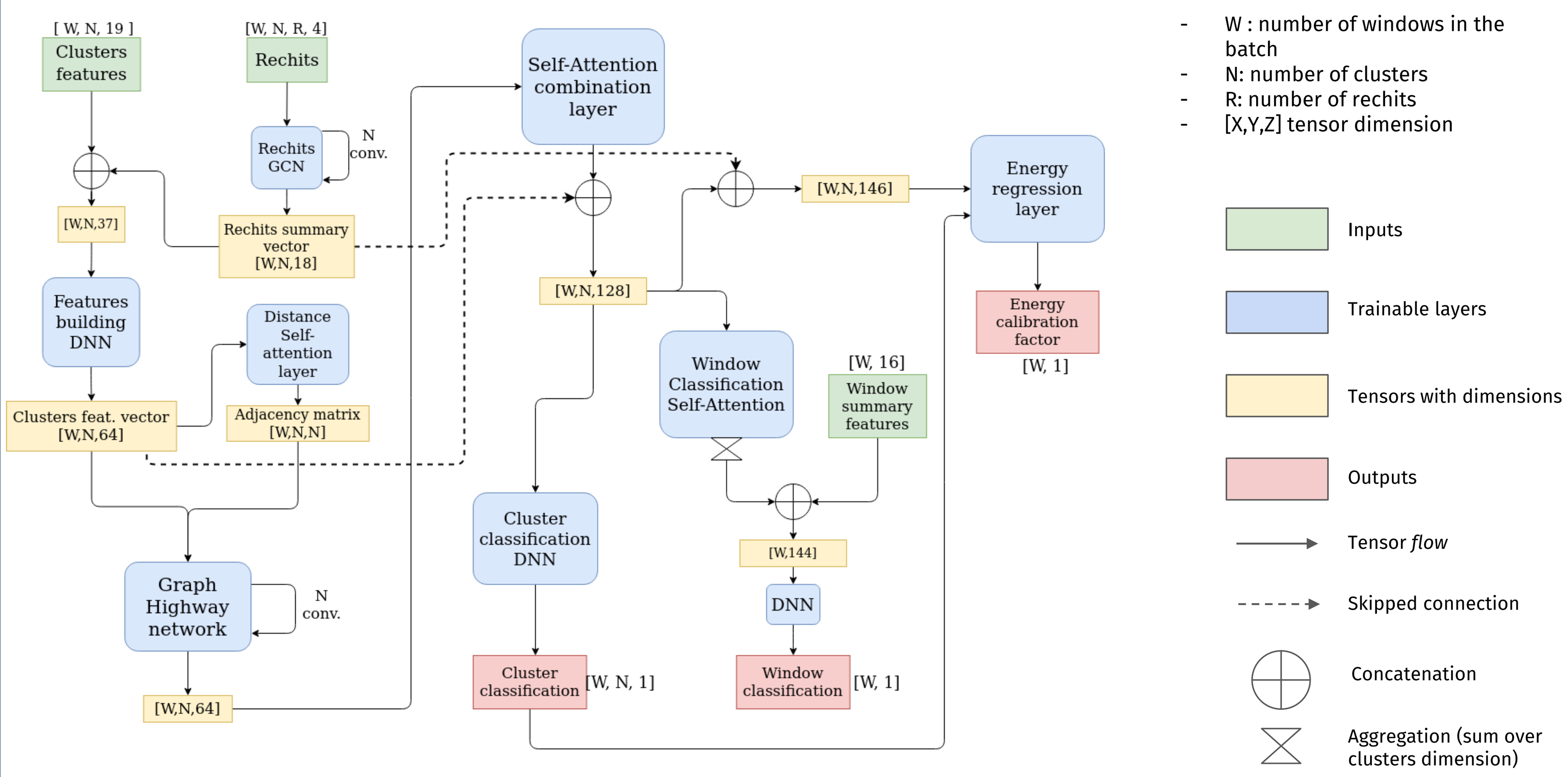
DeepSuperCluster model

Model architecture

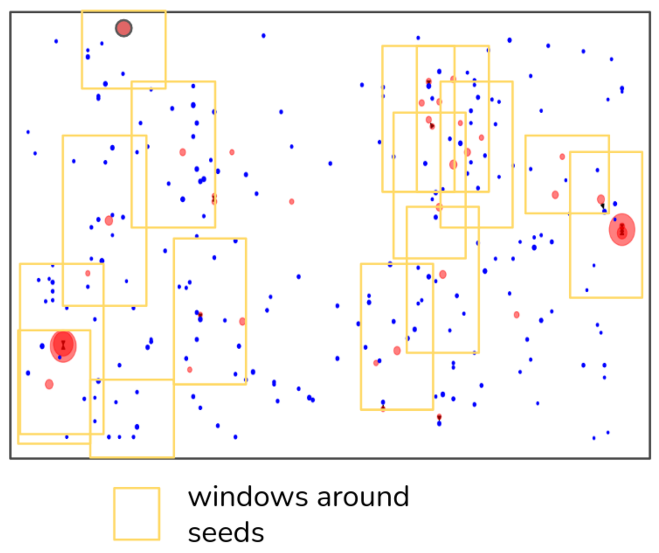
New **graph-based Machine Learning** algorithm for SuperClustering.

→ **Maintains the efficiency while improving PU and noise rejection.**

→ Graph NN are able to **aggregate the information between the neighbors.**



Dataset and Particle ID



Dataset for the training:

- Electrons and photons generated uniformly in p_T = [1,100] GeV.
- PU uniformly distributed between [55,75] interactions.
- Windows opened around all the clusters with E_T > 1 GeV (**seeds**).

Model Input: **Cluster information** *(E, E_T, η, ϕ, z, number of crystals, …)*, **list of rechits**, **summary window features** *(max, min, mean of the crystal variables)*.

Model Output: cluster classification (in/out of SC), particle classification, energy regression.

- → Same network to identify the **flavor of the particle**.
- → Extra dataset: sample containing **jets**.
- → Goal: classify jets/electrons/photons.
- → **Transfer Learning** was used to re-train only the ID part of the network to avoid performance degradation for electrons/photons.

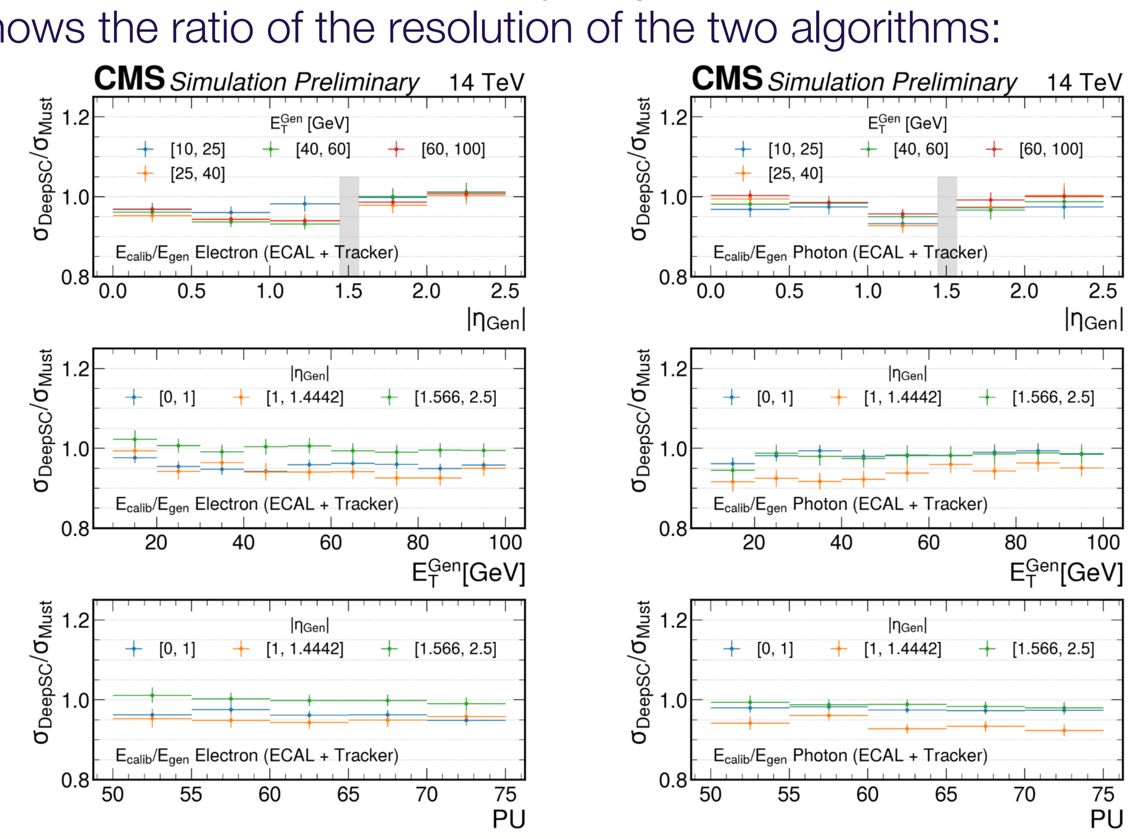
Results: Energy Resolution

Resolution of the reconstructed uncorrected **SuperCluster energy** (E_{Raw}) divided by the **true energy deposits** in ECAL (E_{Sim}) versus:

- the gen-level particle position $|\eta_{Gen}|$ (top)
- the transverse energy of the gen-level particle E_T^{Gen} (center)
- the number of simulated PU interactions (bottom)

More results in CMS-DP-2022-032

The resolution is computed as half of the difference between the 84% quantile and the 16% quantile (one σ) of the E_{Raw}/E_{Sim} distribution in each bin. The lower panel shows the ratio of the resolution of the two algorithms:



Significantly improved resolution, particularly for low E_T signals and at high PU.

Results: Particle Classification

- → Particle classification performance (DeepSC model) for jet vs. photon (left) and photon vs. electron samples (right).
- → Only ECAL variables are used.
- → **High performance for jet vs. photon discrimination.**