Deep learning techniques for energy clustering in CMS ECAL

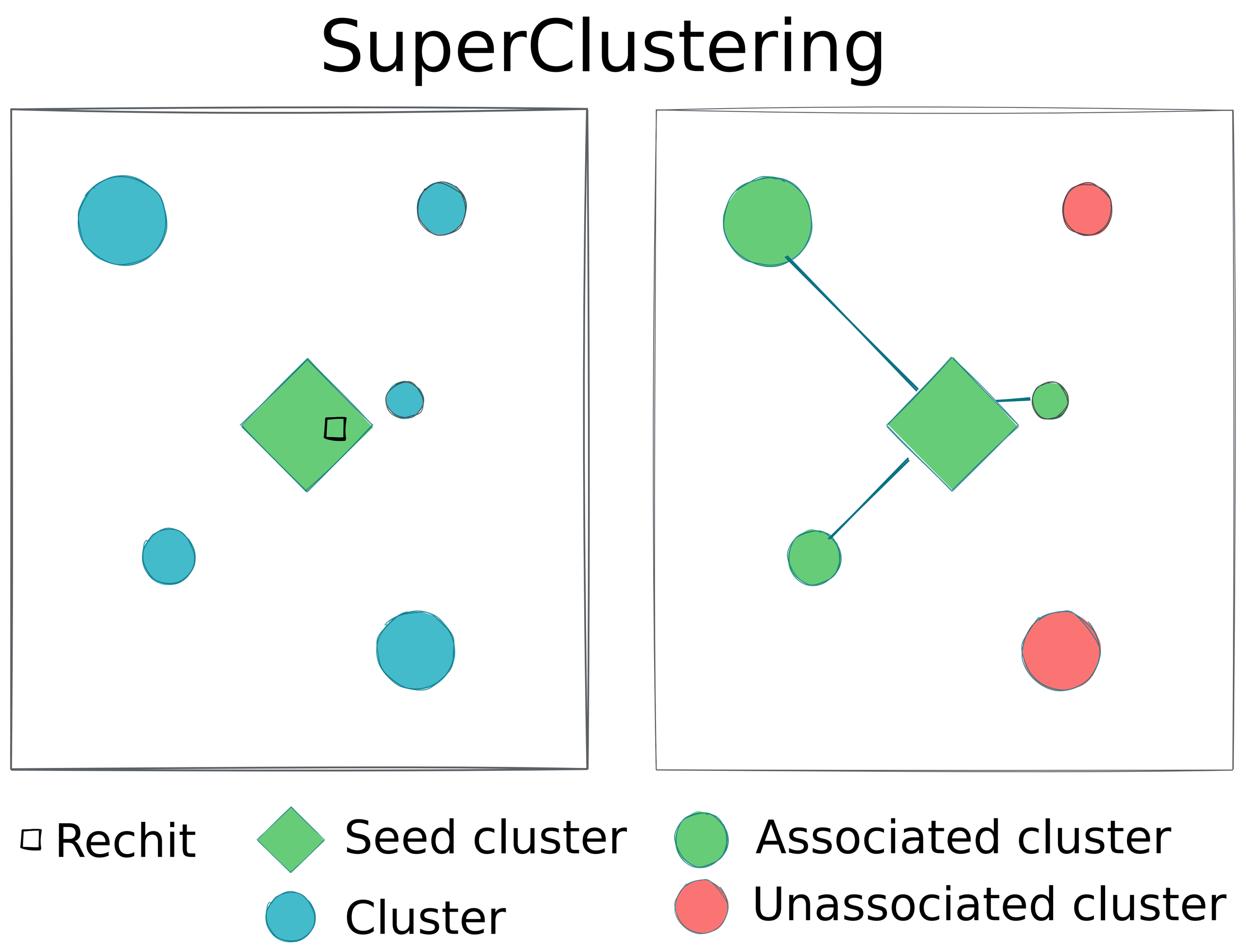
Davide Valsecchi on behalf of the CMS Collaboration
University of Milano-Bicocca, Italy - ETH Zurich, Switzerland

ACAT2021

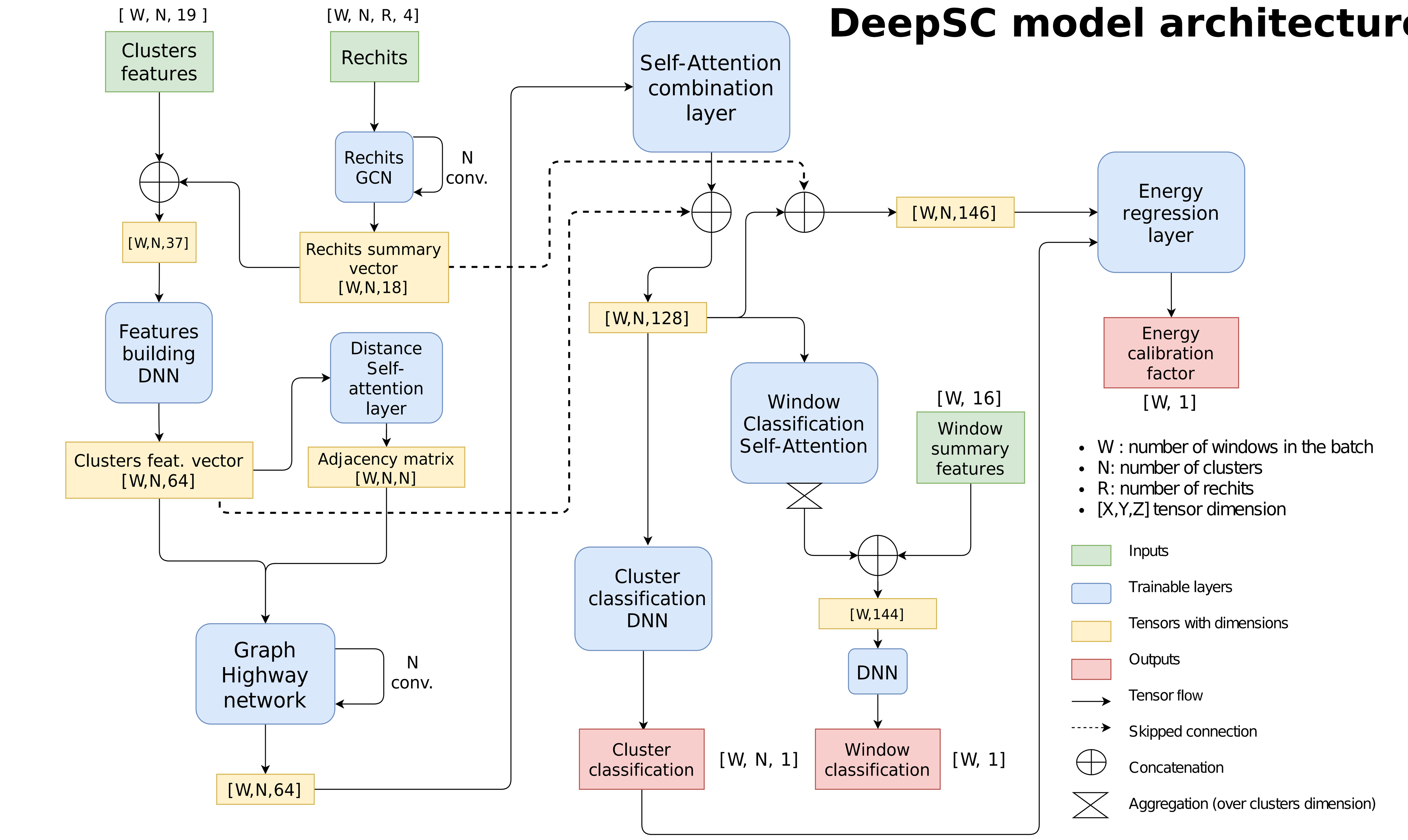
ECAL local reconstruction

Rechits → Clusters → SuperClusters

ECAL local energy reconstruction: first step of electron and photon reconstruction in CMS. The energies recorded in the ECAL crystals (**RecHits**) are first gathered together to form **clusters** of neighbouring deposits. CMS strong magnetic field (3.8T) + tracker detector material in front of ECAL = electron and photon fragmentation by bremsstrahlung and e^+e^- pair production. The **SuperCluster** algorithm gathers together clusters around energetic seeds to restore the optimal energy resolution.

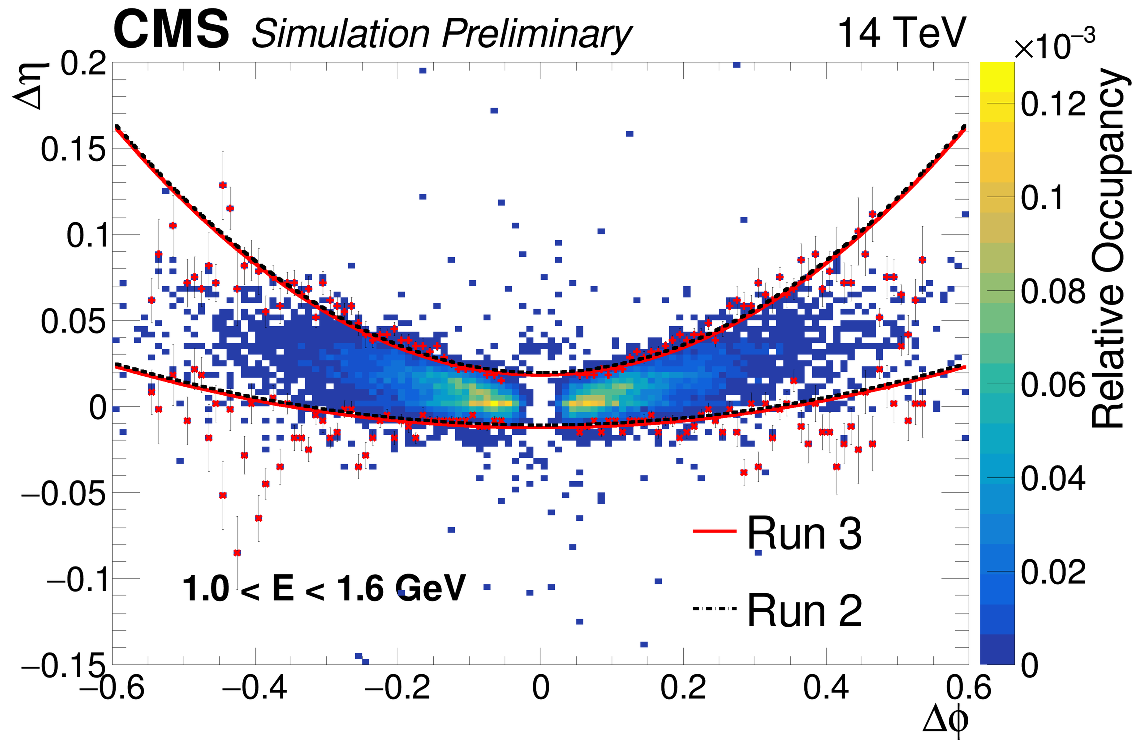


DeepSC model architecture



Blocks: features encoding, **dynamic graph-building**, **graph convolution** (Graph Highway Network) + **self-attention** layer, and multiple decoding outputs. Optimized to solve three different tasks: **cluster classification**, **window classification** and **energy correction** factor regression. The model is trained on a sample of 2M photons and electrons, generated uniformly in η and p_T=[1, 100] GeV, with the full CMS Monte-Carlo simulation at 14 TeV. A pileup scenario with the number of true interactions uniformly distributed between [55,75] is used.

SuperClustering algorithms



Mustache = standard CMS SuperClustering algorithm. It selects clusters inside of a Δη-Δφ parabolic region defined only by the seed position and the energy of the cluster. Highly efficient for signal, but does not filter pileup* (PU) and noise contamination.

DeepSC: a novel approach based on **Graph Neural Networks** has been developed, exploiting all the available information to improve the energy resolution and make the SuperClustering more robust against PU and noise.

*pileup = additional low energy p-p interactions overlapping with the hard interaction

DeepSC - Mustache performance comparison

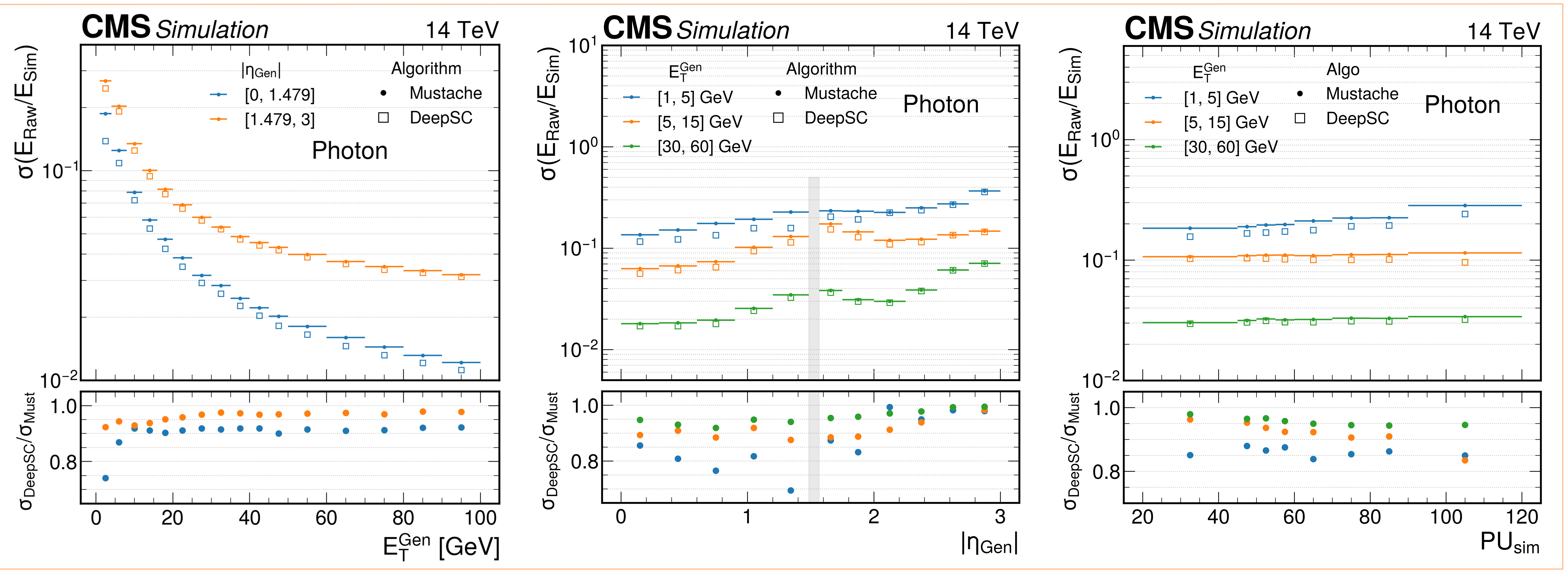
Resolution of the reconstructed uncorrected SuperCluster energy (E_{Raw}) divided by the true energy deposits in ECAL (E_{Sim}) versus:

- the transverse energy of the gen-level particle $\mathbf{E_T^{Gen}}$ (left)
- the gen-level particle position $|\boldsymbol{\eta}_{\mathbf{Gen}}|$ (center)
- the number of simulated $\mathbf{PU}_{\mathrm{Sim}}$ interactions (right)

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

$$\sigma_{\mathrm{DeepSC}}/\sigma_{\mathrm{Mustache}}$$

The DeepSC algorithm shows **significantly improved resolution**, particularly for low E_T signals and at high PU.



Ref: CMS-DP-2021/032

ACAT 2021

Contact: davide.valsecchi@cern.ch