Deep Learning techniques for energy clustering in the CMS electromagnetic calorimeter

Polina Simkina
On behalf of the CMS collaboration

CEA IRFU

TIPP 2023
5 September 2023

Outline

1. Introduction to Graph Neural Networks (GNN)
2. SuperClustering reconstruction in ECAL
    - Mustache algorithm
3. DeepSuperClustering model
    - Model description and architecture
4. Performance
    - Energy resolution
    - Particle Identification

Introduction to Graph Neural Networks

Neural Networks

Neural Networks are one of the widely used **Machine Learning** algorithms.

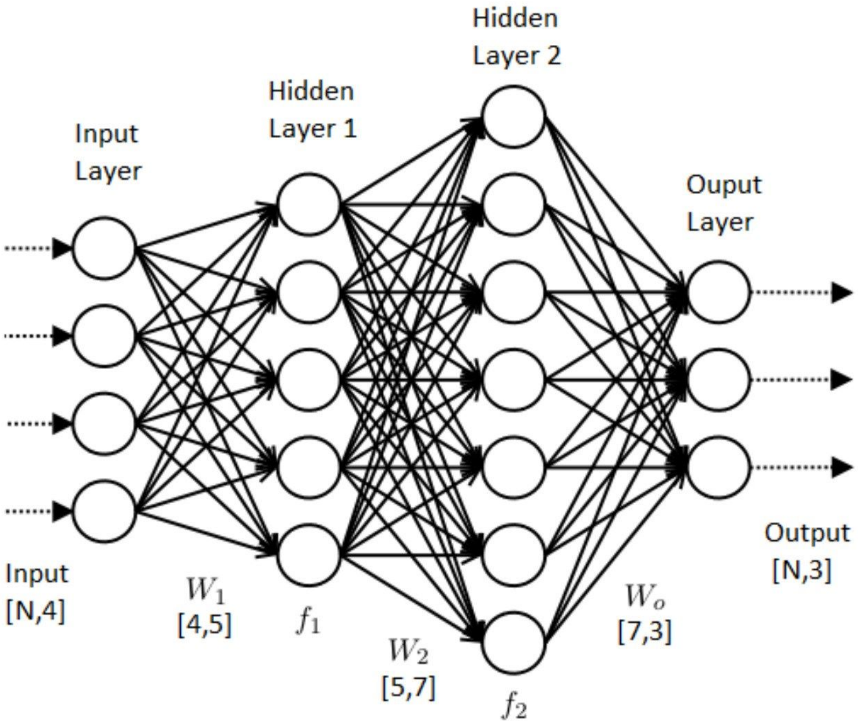
The simplest neural network consists of **an input, an output and one hidden layer.**

If the network has more than one hidden layer, it is called **Deep Neural Network**.

Network training:

- The input vector is multiplied by a **weight matrix** resulting in an input to a new **(hidden) layer**. This process then can be successively repeated with new layers (each time with a different weight matrix).

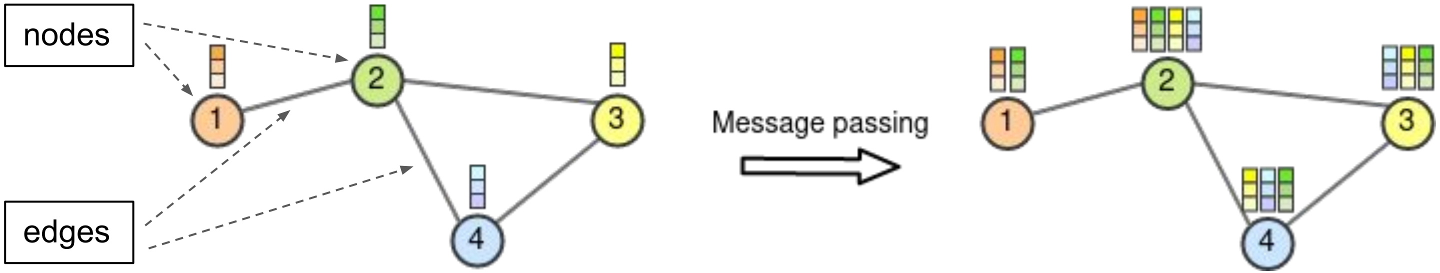
- The result can be extracted from the output of the last layer. It is compared with the "right" answer and based on the loss function (e.g. Mean Squared Error) the weights are adjusted using a method called **backpropagation**.



Example of Deep Neural Network

Graph Neural Networks (GNNs)

- Type of neural network that can operate on and analyze **graph structures**.
- Unlike other types of networks GNN can be easily applied on sparse data.
- A graph consists of **nodes** (contain features of the object) and **edges** (reflect the relationship between the nodes).
- In GNNs the information can be shared between the neighbors:
  - The vector features of each node are transformed into “messages” (e.g. using dense layers) that are sent to the neighbors (message-passing).
  - In this way, **each node learns information about its neighbors and itself**. The process is carried out in parallel and repeated several times.



Source: Graph Neural Networks

SuperClustering in ECAL

Electromagnetic CALorimeter

Homogeneous calorimeter.

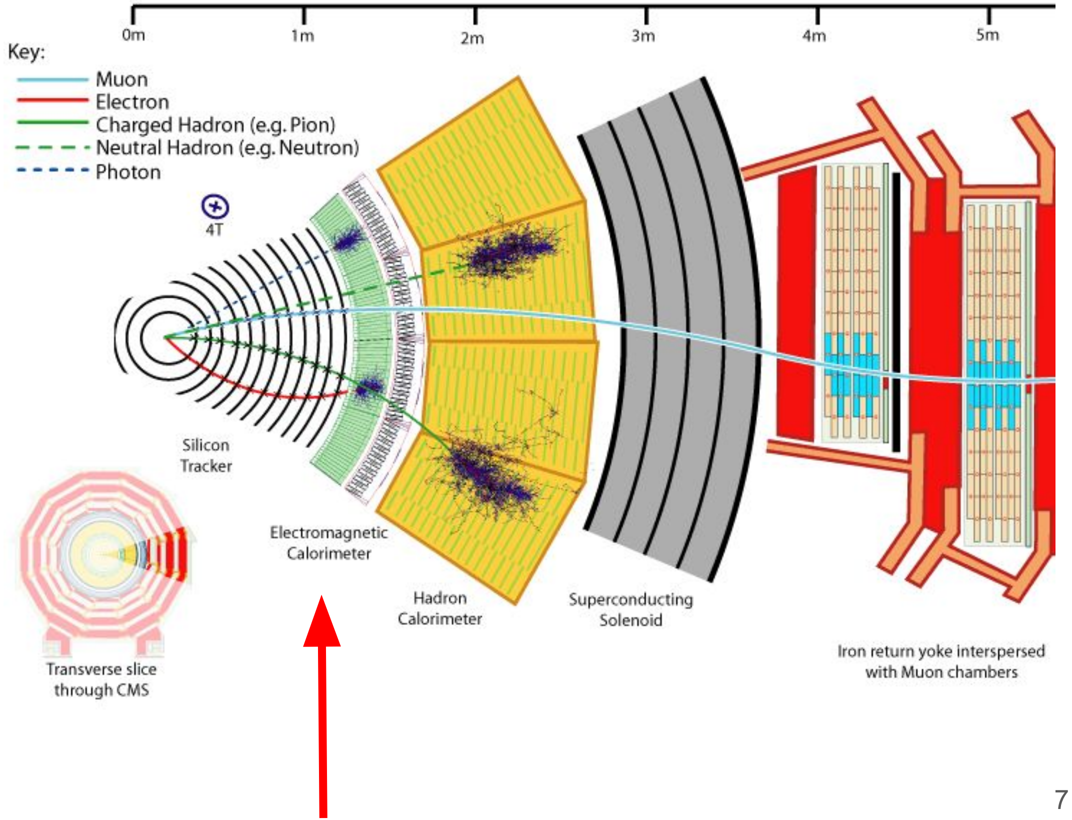
Around 76 000 **$PbWO_4$ crystals.**

Mainly used for the reconstruction of **electrons** and **photons,** as well as **jet reconstruction.**

Plays crucial role for **all physics analysis**, e.g. for Higgs decay channels:

$$H \to \gamma\gamma$$

$$H \to ZZ^{*} \to 4\ell$$

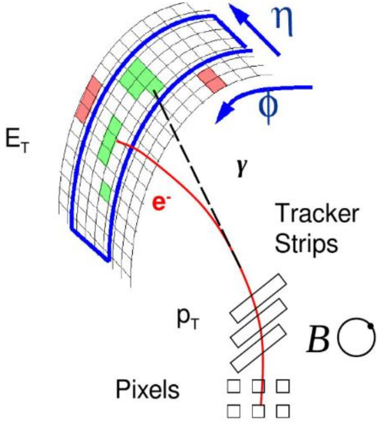


EM object reconstruction in ECAL

Part of CMS reconstruction framework – Particle Flow

Rechits → Clusters → SuperClusters

- **Energy deposits** left by particles in the $PbWO_4$ crystals of the calorimeter.

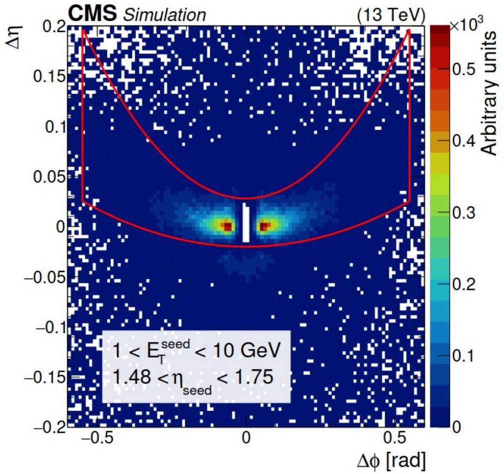


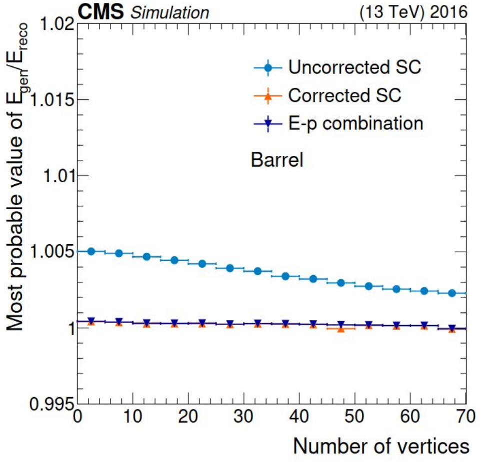
- Rechits are gathered together to form a **cluster** that represents a **single particle** or **several overlapping particles**.

- Because of bremsstrahlung or photon conversion before the ECAL, clusters have to be combined to form a **SuperCluster**.
- Currently a geometrical algorithm (Mustache) is used, a Boosted Decision Tree is applied for energy correction.
- The energy of the initial particle can be reconstructed from a SuperCluster.

Mustache SuperClustering

- The algorithm currently used in CMS for reconstruction of SuperClusters.
- Purely **geometrical** approach:
  - All the clusters falling into the specified “mustache” shape would be considered as part of the SuperCluster. The size of the area depends on energy and position of the seed (E_T > 1 GeV).
  - “Mustache” shape due to the CMS magnetic field (spread along φ) and geometric effect (spread along η).





- **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 to correct biases from **PU, noise, and energy losses.**

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

DeepSC model

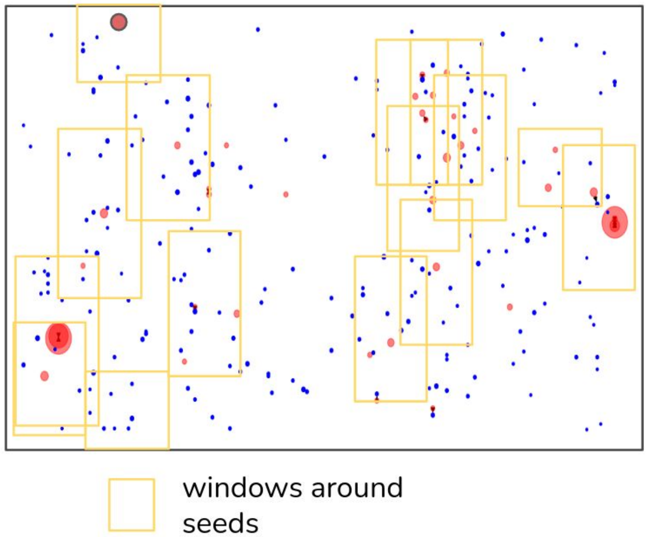
GNN for ECAL SuperClustering

New algorithm for SuperClustering: **DeepSuperCluster ML model**

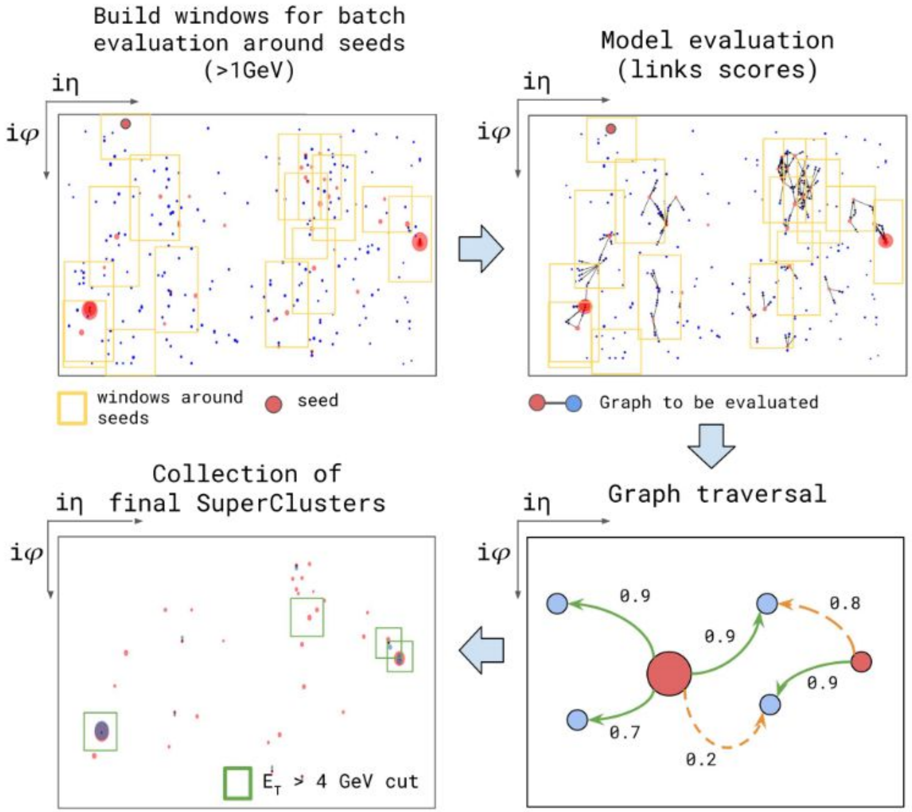
- Based on Graph Neural Network. It can receive and combine the information from all the clusters in the window.
- **Maintains the efficiency while improving PU and noise rejection.**

For the training and testing a dedicated Monte Carlo sample was generated:

- Electrons and photons are generated uniformly in p_T = [1,100] GeV.
- PU uniformly distributed between [55,75] interactions is used.

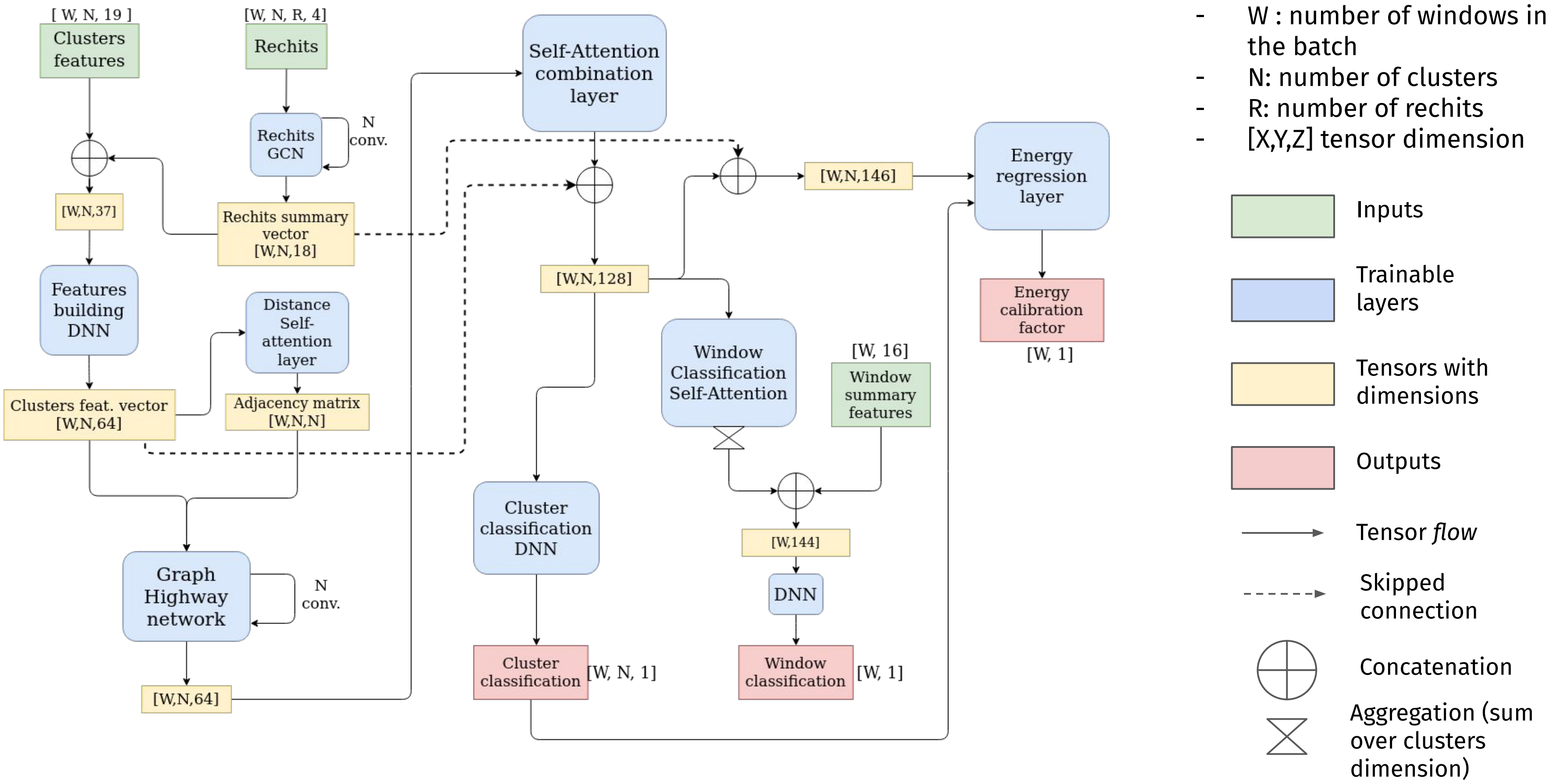


Model input



- Windows are opened around all the clusters with E_T > 1 GeV (**seeds**).
    - Window dimensions are η-dependent.
    - The model has to process each window and give a prediction for it.
- The **inputs** for the model are:
    - Cluster information (energy, position, etc.)
    - List of rechits for each cluster.
    - Summary window features.
- The **outputs**: cluster classification (in/out of SC), window classification (electron/photon/jet), energy regression.

GNN for ECAL SuperClustering architecture

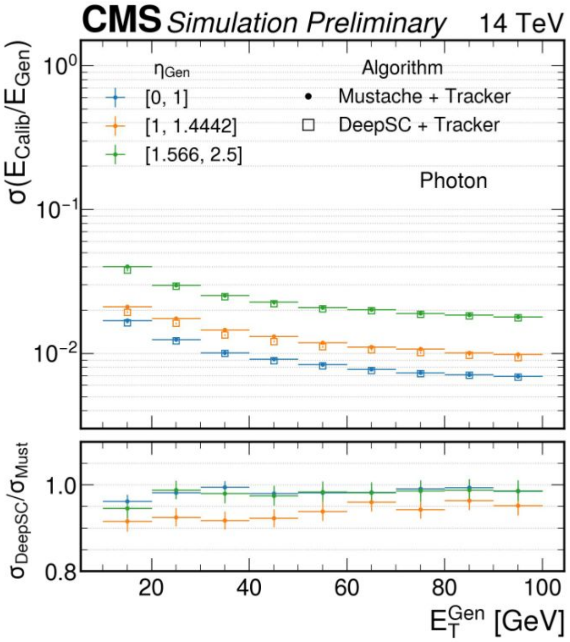


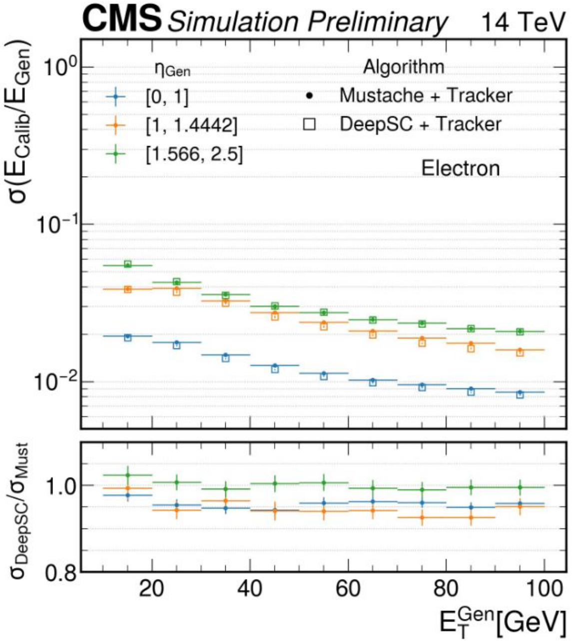
- W : number of windows in the batch
- N: number of clusters
- R: number of rechits
- [X,Y,Z] tensor dimension

Performance

Performance: energy resolution vs. energy

Resolution of the reconstructed calibrated **SuperCluster energy** (E_{Calib}) divided by the generated particle **energy** (E_{Gen}) versus the transverse energy of the gen-level particle E_T^{Gen}. ECAL energy is combined with the track energy for doing the calibration.





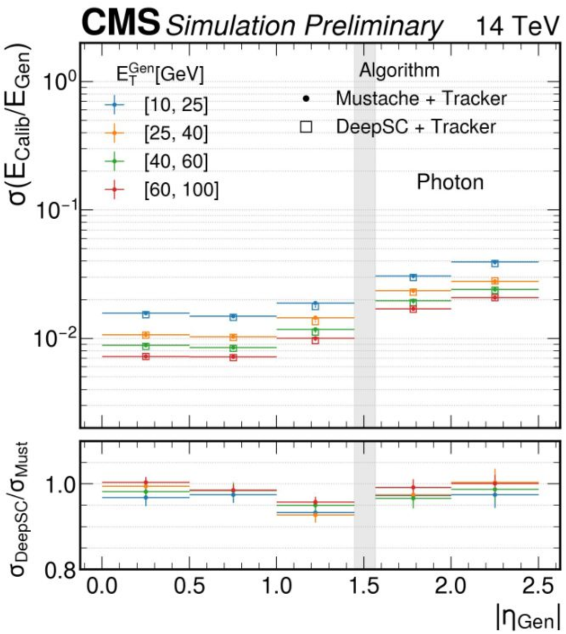
The resolution is computed by fitting a curjiff function.
The lower panel shows the ratio of the resolution of the two algorithms:

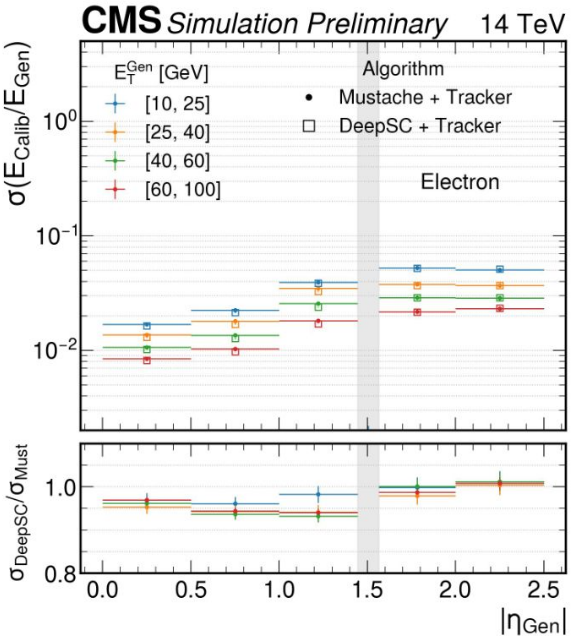
$$\sigma_{DeepSC} / \sigma_{Mustache}$$

The DeepSC algorithm shows **improved resolution**, particularly for low E_T.

Performance: energy resolution vs. eta

Resolution of the reconstructed calibrated **SuperCluster energy** (E_{Calib}) divided by the generated particle **energy** (E_{Gen}) versus the gen-level particle position $|\eta_{Gen}|$. ECAL energy is combined with the track energy for doing the calibration.





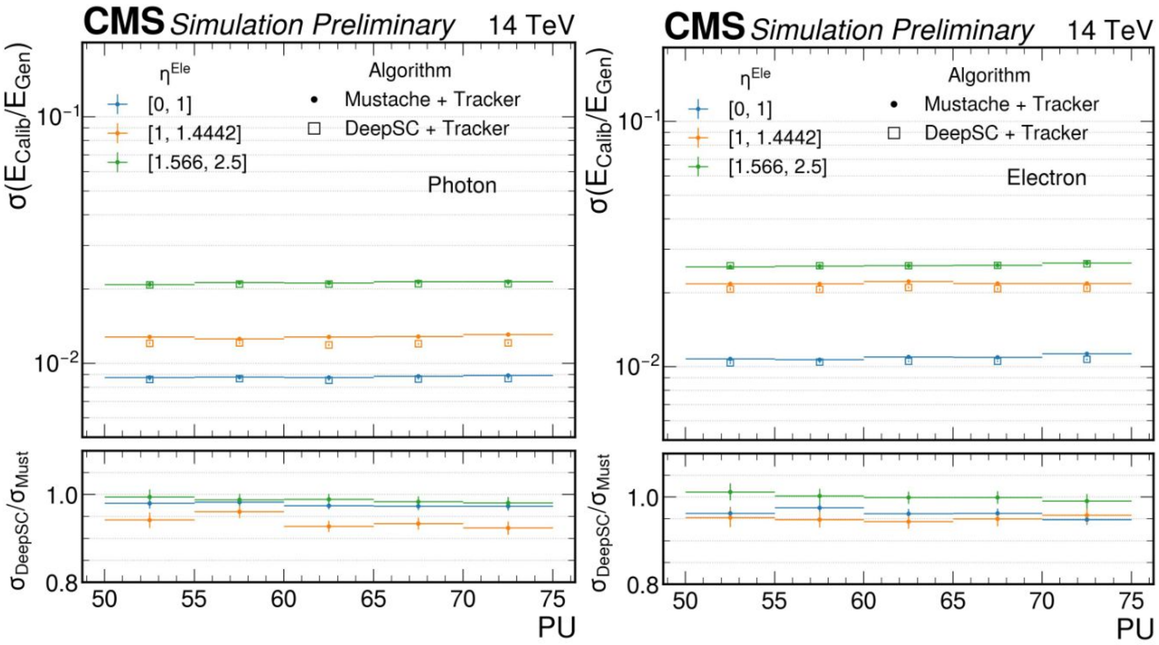
The resolution is computed by fitting a curjiff function.
The lower panel shows the ratio of the resolution of the two algorithms:

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

The DeepSC algorithm shows **improved resolution**, particularly for barrel region.

Performance: energy resolution vs. pileup

Resolution of the reconstructed calibrated **SuperCluster energy** (E_{Calib}) divided by the generated particle **energy** (E_{Gen}) versus the number of pile up (PU) events. ECAL energy is combined with the track energy for doing the calibration.



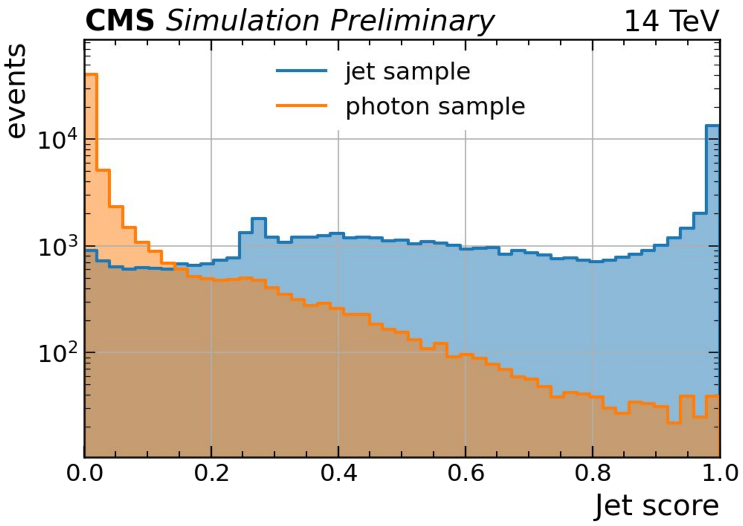
The resolution is computed by fitting a curjiff function.
The lower panel shows the ratio of the resolution of the two algorithms:

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

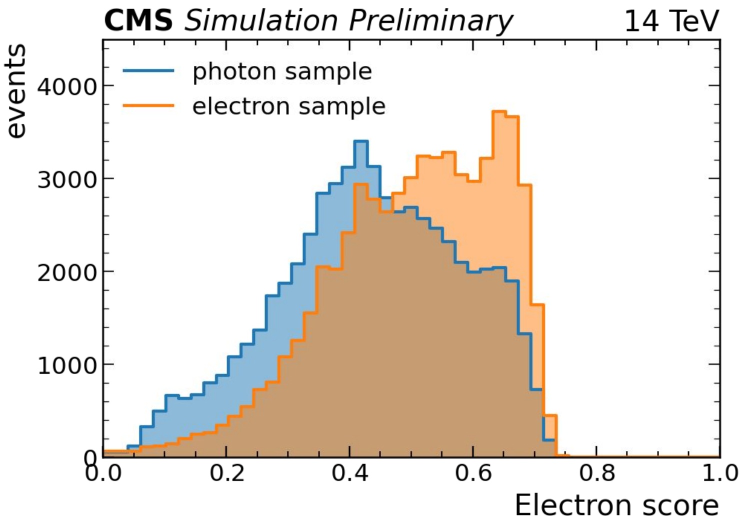
The DeepSC algorithm shows **improved resolution**, particularly for high PU.

Particle identification

- Same network can be used to identify the **type of the particle**.
- An extra sample containing **jets** was generated (same energy/PU as for electron/photon sample): 25% of energy in jets comes from photons but different topology between ECAL energy deposits coming from prompt electrons/photons and from jets.
- In order to avoid performance degradation for electrons/photons in terms of cluster selection, **Transfer Learning** was used to re-train only the ID part of the network.



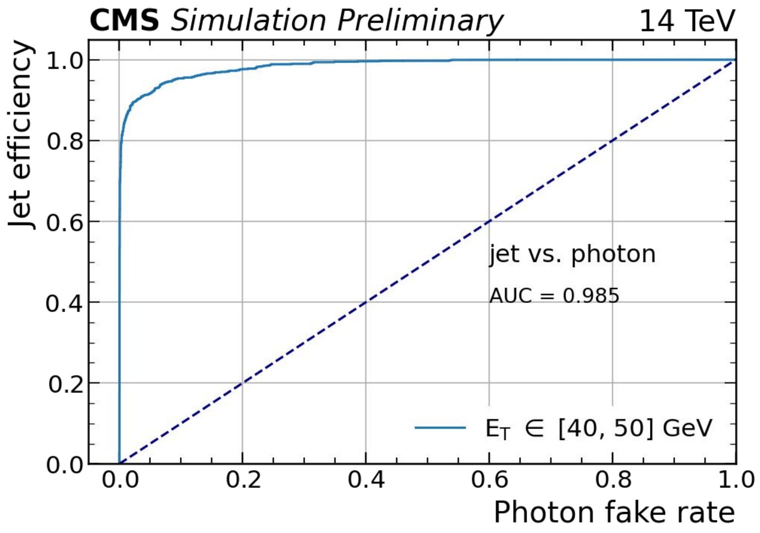
Likelihood to be predicted as jet (score) for the jet and photon sample.

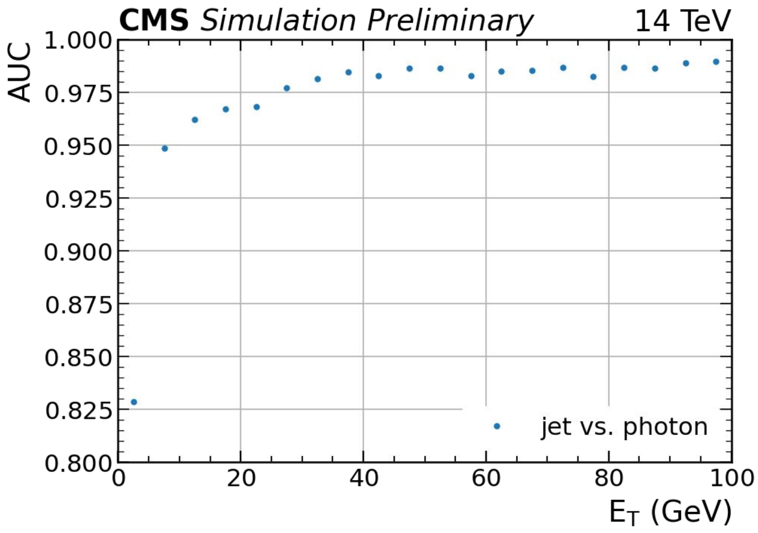


Likelihood to be predicted as electron (score) for the photon and electron sample.

Performance: jet vs. photon

- ROC curve obtained from the discriminator for jet vs. photon for E_T = [40, 50] GeV (left).
- Summary performance obtained by calculating **Area Under the ROC curve** (AUC) for different energy ranges (right).



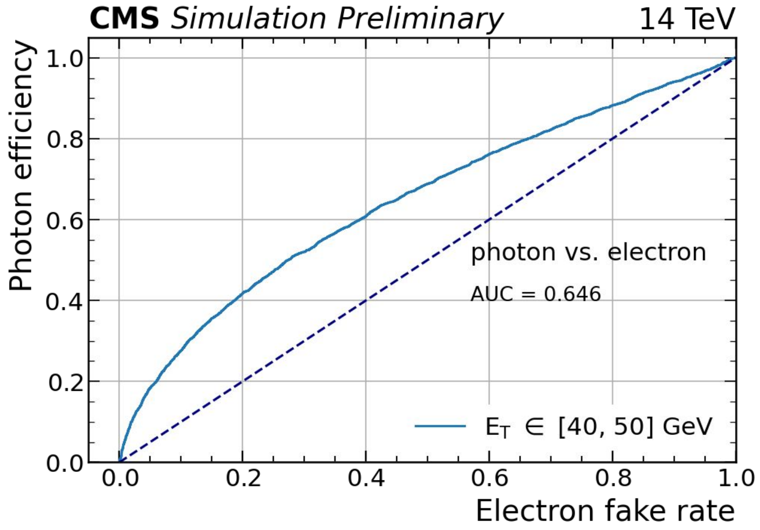


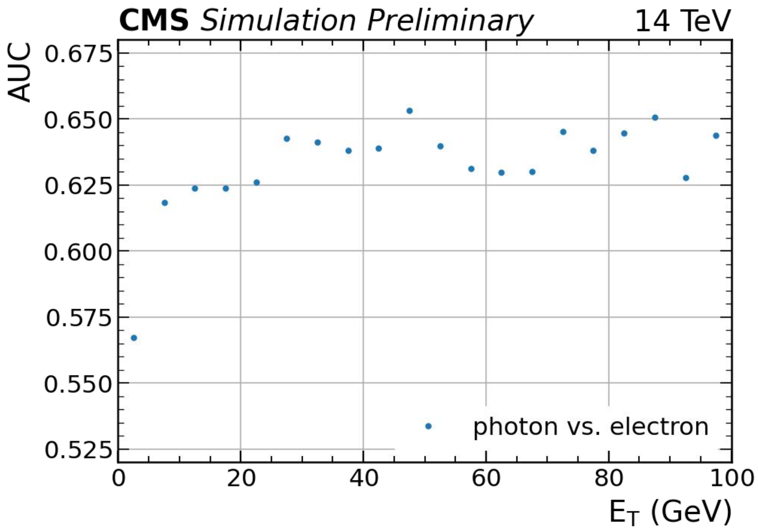
AUC levels ~98 % for jet vs. photon discrimination for high energy.

- The output of the model can be used in the global event reconstruction of CMS.

Performance: electron vs. photon

- ROC curve obtained from the discriminator for electron vs. photon for E_T = [40, 50] GeV (left).
- Summary performance obtained by calculating **Area Under the ROC curve** (AUC) for different energy ranges (right).





- AUC levels for photon vs. electron discriminator are ~63%.
- Only ECAL variables are used, can be beneficial when the track information is lost or not reconstructed.

Conclusion

- New calorimeter reconstruction algorithm based on Graph Neural Networks, DeepSC model, is presented.
- Outperforms the traditional approach in terms of energy resolution.
- DeepSC model is also able to perform particle identification based solely on the information from the ECAL.
    - Shows promising results for photon vs. jet discrimination.
    - Electron vs. photon discrimination can be additionally used for the cases where the track information is lost or not reconstructed.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No *800945* — NUMERICS — H2020-MSCA-COFUND-2017