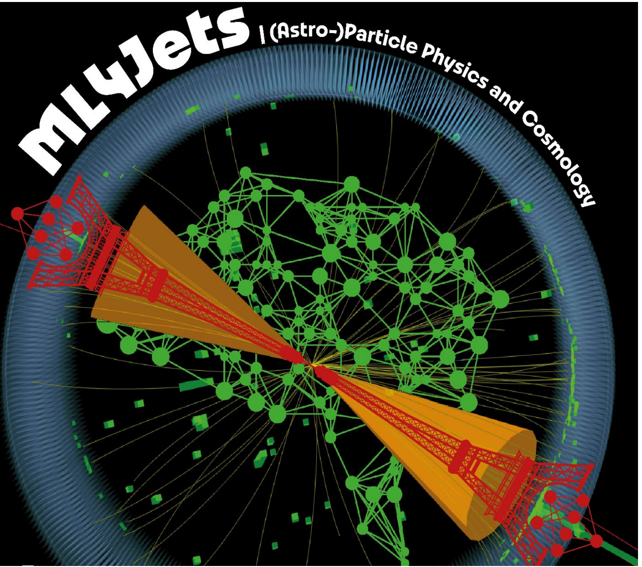
# ML assisted Event Reconstruction in the CMS Phase-2 High Granularity Calorimeter Endcap




**Théo Cuisset** on behalf of the **CMS collaboration**
*Laboratoire Leprince-Ringuet, École Polytechnique*

November 8$^{th}$, 2024
**ML4Jets**

Laboratoire Leprince-Ringuet
ÉCOLE POLYTECHNIQUE
INSTITUT POLYTECHNIQUE DE PARIS
CMS Compact Muon Solenoid

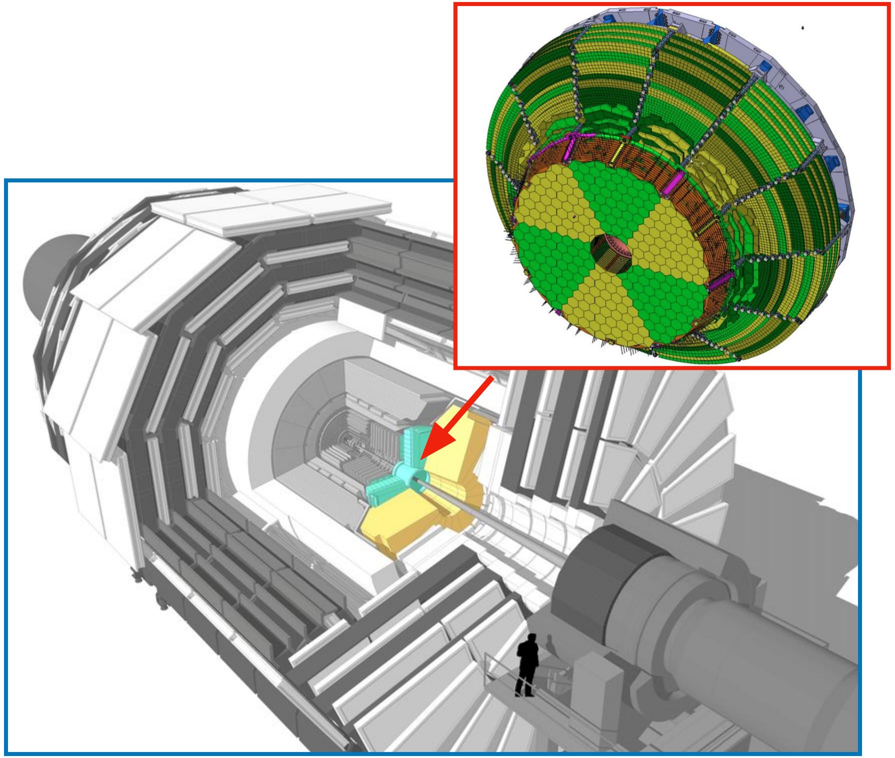
# The High Granularity Calorimeter (HGCAL)

- To enhance the potential for discoveries at the LHC, CERN plans to upgrade the accelerator to the **high luminosity LHC** (HL-LHC) by 2029.

- At the end of LHC Phase 1 operations, the end-cap calorimeters will have suffered irrecoverable radiation damage.

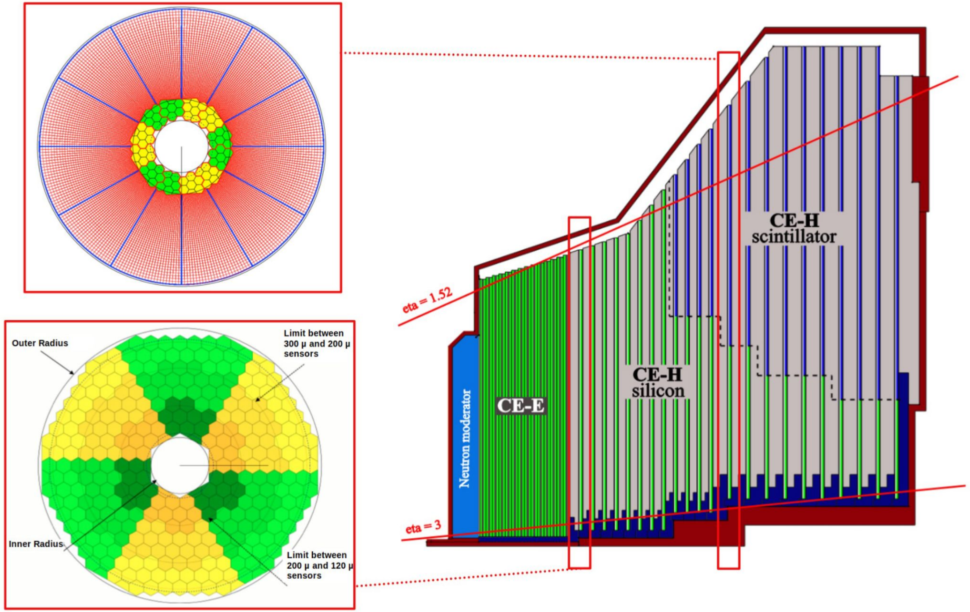
The CMS collaboration is replacing its endcap calorimeters with a new sampling

**High Granularity Calorimeter.**



*The CMS detector and position of HGCAL*

# The High Granularity Calorimeter (HGCAL)




*Longitudinal cross section of the upper half of one of the HGCAL (right), mixed silicon and scintillator layers (top left) and a full silicon layer (bottom left).*

- First **imaging calorimeter** to be implemented in high-energy physics.
- **Position, energy and timing** information.
- **26** layers of electromagnetic section and **21** layers of hadronic section.
- Designed for the HL-LHC target integrated luminosity of **3000 fb$^{-1}$.**

# Machine Learning in HGCAL

HGCAL : High granularity calorimeter designed for Particle Flow (➞ combining tracker and calorimeter information)

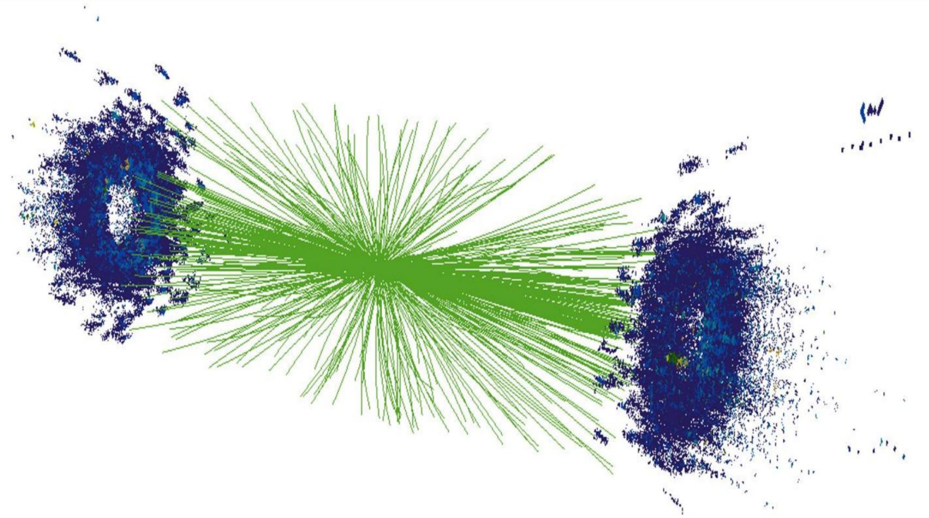
## Challenges:

- Environment from the **high pile-up** : number of interactions per bunch crossing will be increased to **140-200.**
- **6 million** read-out channels and 5D information from them.

Novel **machine-learning based algorithms** to tackle these challenges!

But : execution time needs to be kept at the **minimum** (especially for software trigger)

Hybrid approach : combining Machine Learning and traditional algorithms (highly optimized on GPUs)



*Simulated event display of HGCAL at 200 pileup*

# Reconstruction in HGCAL : TICL framework

TICL : The Iterative CLustering is a new framework is designed to fully exploit the high spatial resolution and precise timing information provided by HGCAL

**Rechits**
Sensor readout
***O*(1M)** per event

CLUE algorithm
on **GPU**

**Layer clusters**
Transverse section of a shower
***O*(100k)** per event

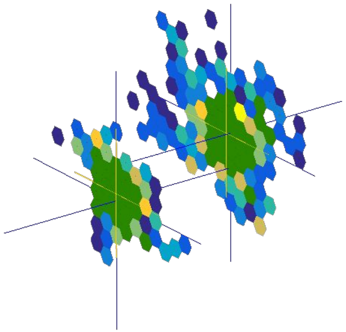
CLUE3D algorithm
soon on **GPU**

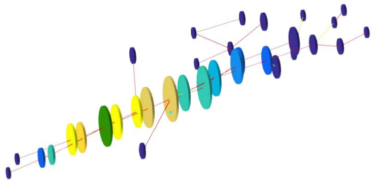
**Tracksters**
Full reconstructed shower of a particle entering the calorimeter
***O*(1k)** per event

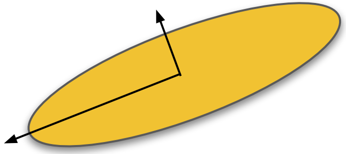
*Position x,y,z*
*Energy, time*

*Position x,y, layer*
*Energy*
*Average time*

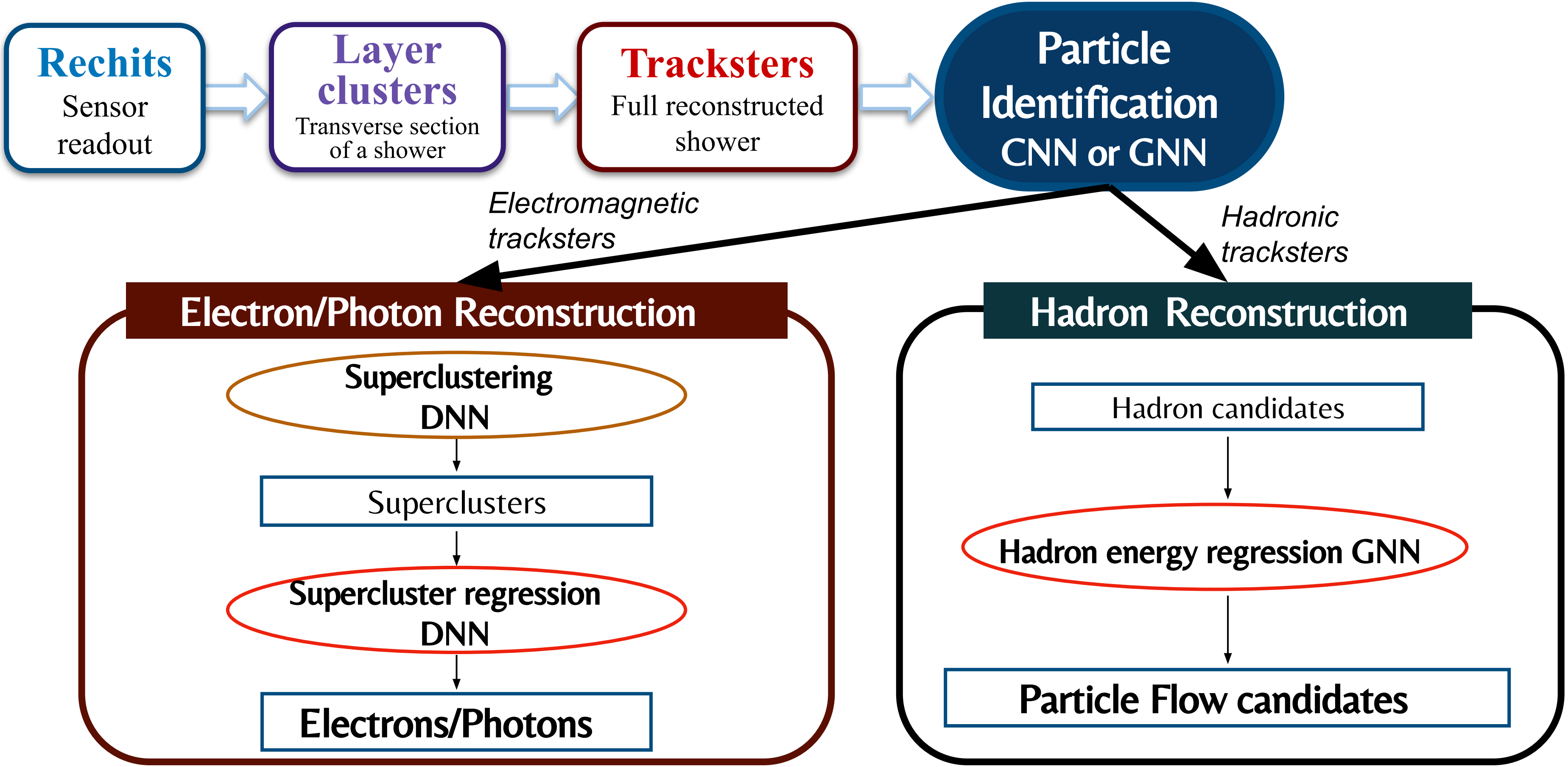
*Position (η-φ)*
*Pointing angle*
*Energy*
*Average time*
*Transverse & longitudinal spread*







# Reconstruction in HGCAL : TICL framework



# Particle identification

Aim : **separate tracksters between hadrons and electron/photon**
Electromagnetic objects are used to seed a dedicated electron track reconstruction
*(Gaussian Sum Filter, time consuming)*
➞ need to keep the number of electron seeds from HGCAL under control.

## Different approaches developed in HGCAL

### Classical approach

Compute **high-level features** :

- Energy fraction in hadronic compartment
- Longitudinal & transverse spread of shower

**Simple but limited performance**

### Layer cluster approach

Feed all layer clusters to a **CNN** or a **GNN**:

- layer cluster position x,y,z
- layer cluster energy E

O(100) inputs per trackster

**Good tradeoff between network size and performance**

### Reconstructed hits approach

Feed all the **raw detector hits** to a **GNN** :

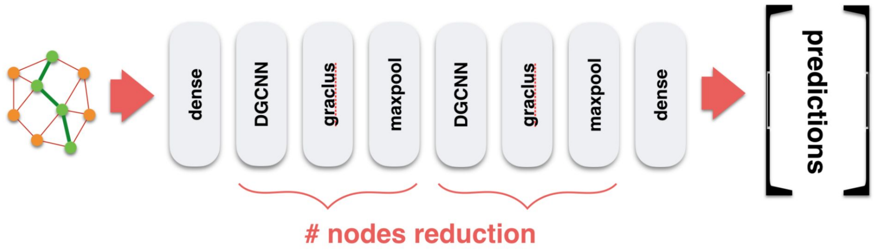
- hit position $x,y,z$
- hit energy $E$

O(10k) inputs per trackster
Time consuming

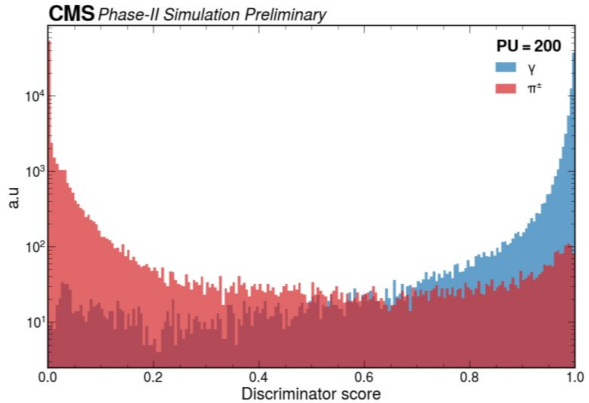
**Lets the network learn about the structure of a shower**

# Particle identification using GNN

**Graph neural network** using edge convolution and greedy clustering based pooling (Dynamic reduction Network)
Input as point cloud, dynamically learns the graph structure



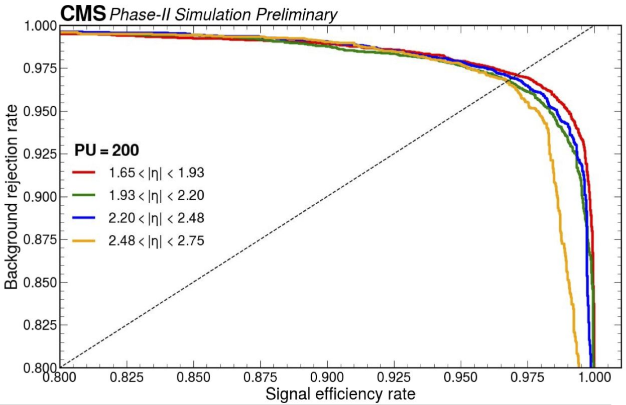

Dataset : unconverted photons & early showering pions (10GeV-1TeV energy) with 200 pileup




*Discriminator score*

*ROC curve*

Source : CMS DP-2022/002
arXiv:2003.08013




Good separation of pions versus photons.
Performance is better at lower η due to reduced pileup contamination of tracksters

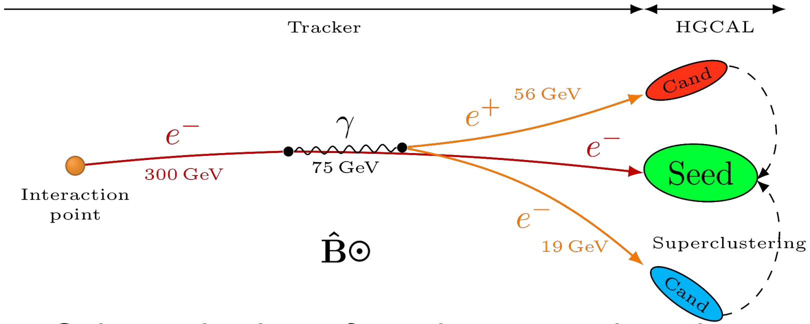
# Electron superclustering

An electron emits several **bremsstrahlung photons** whilst travelling through the tracker

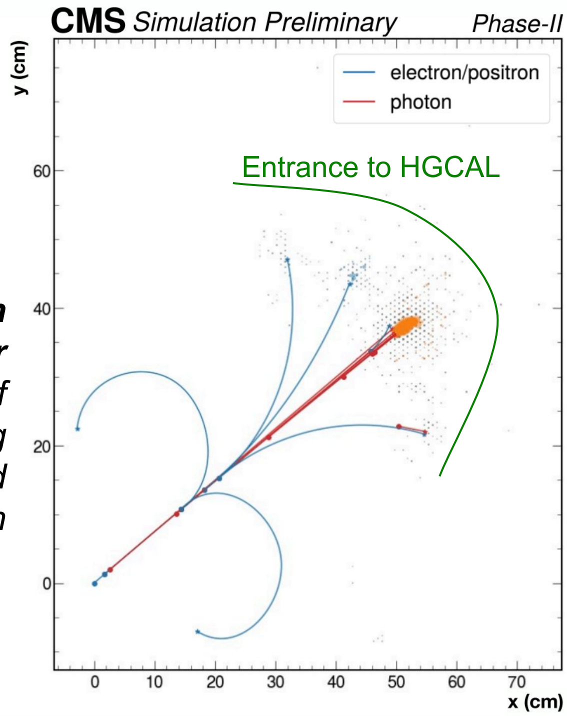
➙ Several particles enter the calorimeter, spread in φ due to the magnetic field.

They need to be reconnected to the parent electron, to:

- seed tracker electron reconstruction
- collect all the electron energy, minimizing pileup contribution




*Schematic view of an electron undergoing bremsstrahlung*

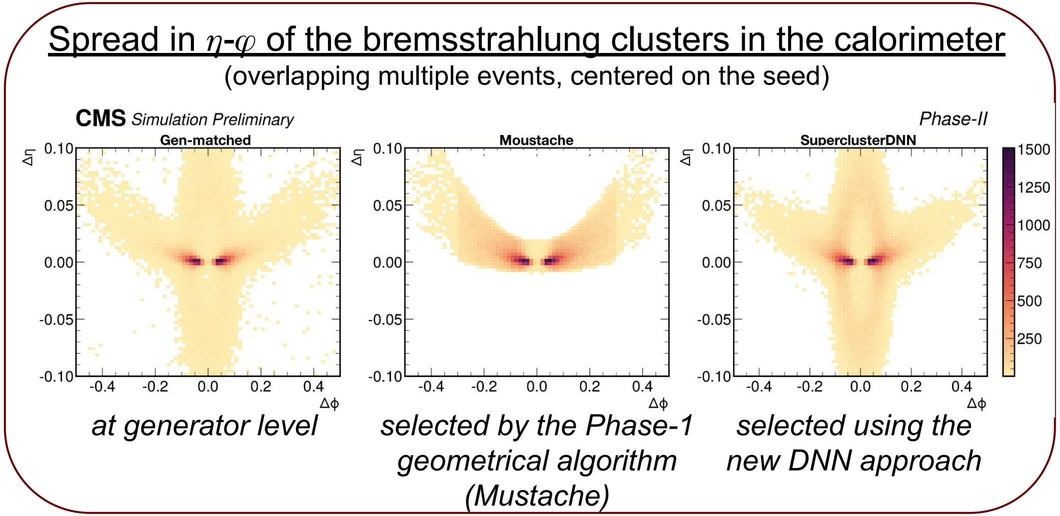


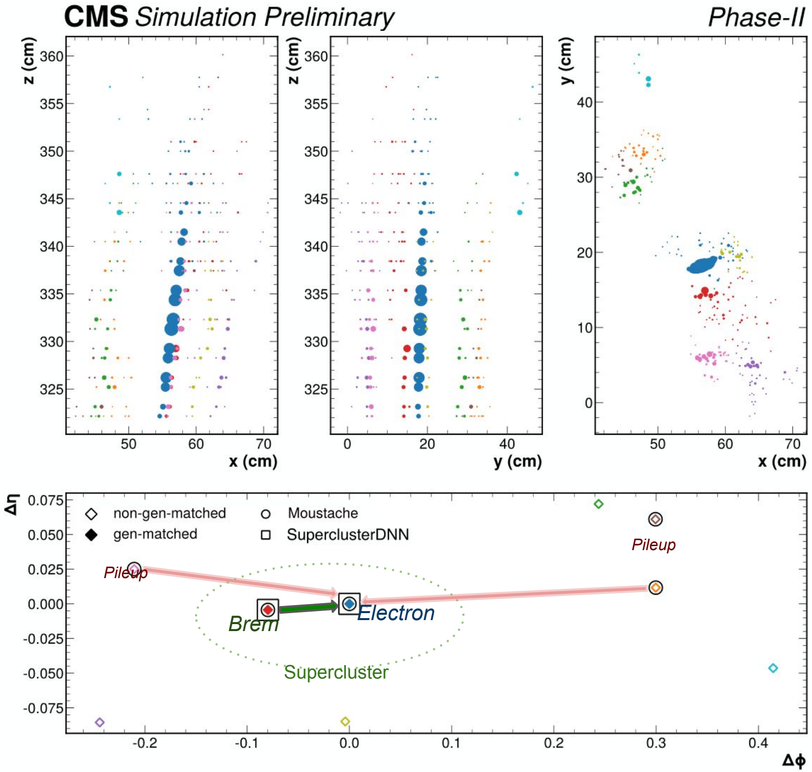

*Display of an* ***electron travelling in the tracker*** *until the entrance of HGCAL, undergoing bremmstrahlung and pair-production*

Input: **tracksters** of the event

Output : **superclusters** = collection of tracksters representing a calorimeter electron with all its brem deposits

# Electron superclustering algorithms



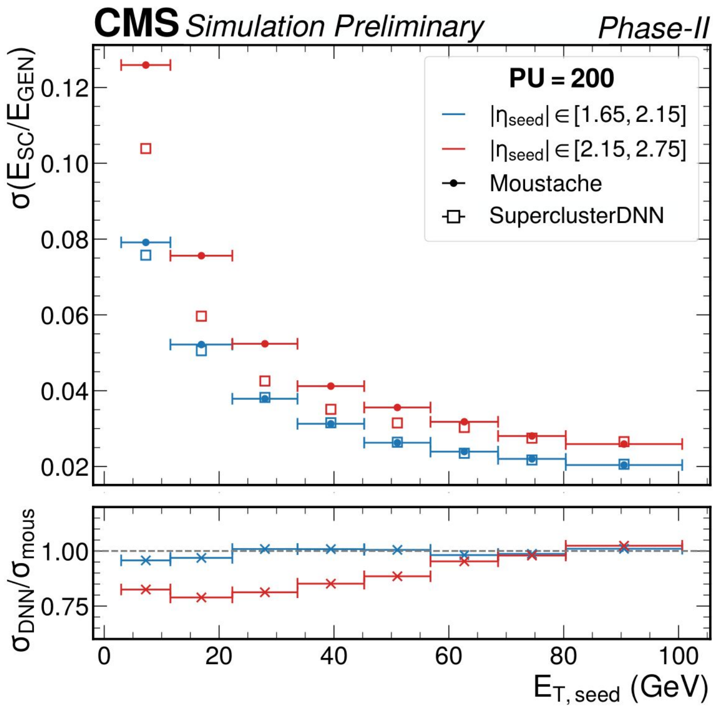

The Phase-1 algorithm uses solely η-φ information for superclustering.
A new **Deep Neural Network** was developed using the **imaging capabilities** of HGCAL.
Inference is run on **pairs of tracksters**, using position and angular variables.
Superclusters are built iteratively, placing a threshold on the score.

# Electron superclustering performance

<u>Dataset</u> : electrons in 200 pileup

**Supercluster energy** is compared to the **true Monte Carlo energy** in the calorimeter.
**Energy resolution plots** are obtained, comparing the superclustering DNN to the Moustache geometric algorithm.

The DNN shows **improved resolution at low energy and high η**, thanks to pileup rejection.

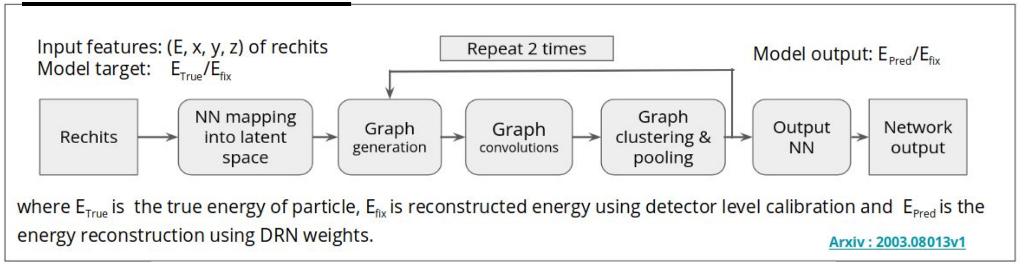


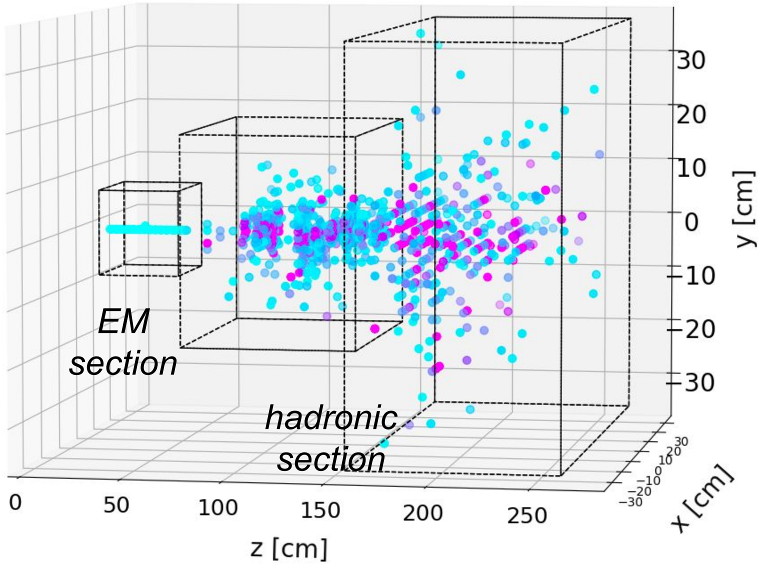

The Deep Neural Network allows for **better energy resolution** and **pileup rejection**, allowing better reconstruction efficiency and lower energy thresholds

# Hadron energy regression

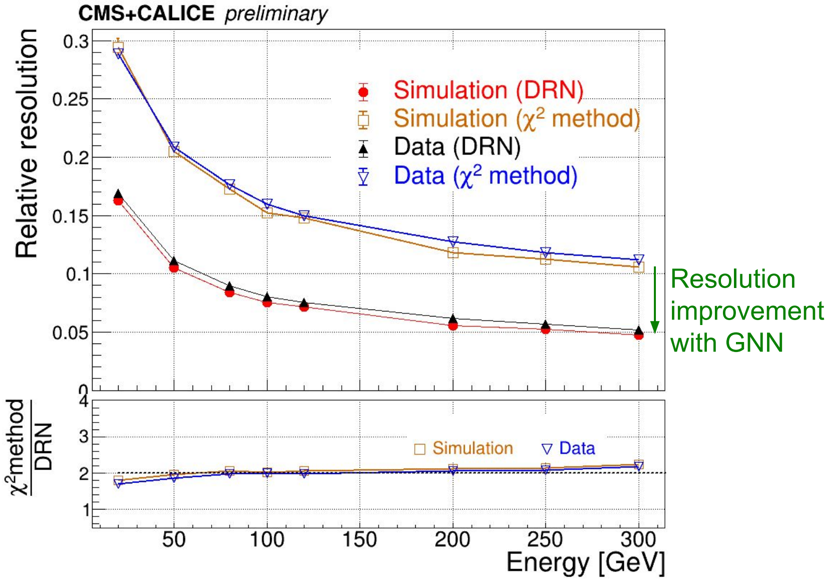
**Energy regression of hadronic shower** using a Graph Neural Network. Trained on simulation and tested on test beam data (pion beam).

## Model architecture




*Display of the input reconstructed hits to the GNN.*




*Energy resolution comparing classic calibration ($\chi^2$ method) and GNN regression, on test beam data and simulation.*

**→Significant resolution improvement.**

# Future ML developments in HGCAL

Many more Machine Learning applications in HGCAL are possible. Current directions include:

- Bringing most of the **reconstruction on GPU**, for increased speed (for software trigger) and reduced hardware costs.
- Use of **timing information** for pileup rejection using ML : timing measurements at a precision of 60 ps for a single hit (of enough energy), and down to 20 ps for a full shower
- Use of the **full set of reconstructed hits** in Graph Neural Network architectures.
- **ML in Particle Flow** : associating tracks to calorimeter tracksters, ambiguity resolution…

# Conclusion

HGCAL is a **high granularity** calorimeter, designed for **Particle Flow** reconstruction, combining calorimeter and tracker information.

Large **pileup** and 6M readout channels make reconstruction particularly challenging, requiring a combination of classical and ML algorithms for computing performance.

Variety of tasks where **Machine Learning** can be used : **Particle Identification**, **Electron Reconstruction**, **Energy Regression**….

**Machine learning will be a key ingredient to the success of HGCAL reconstruction !**