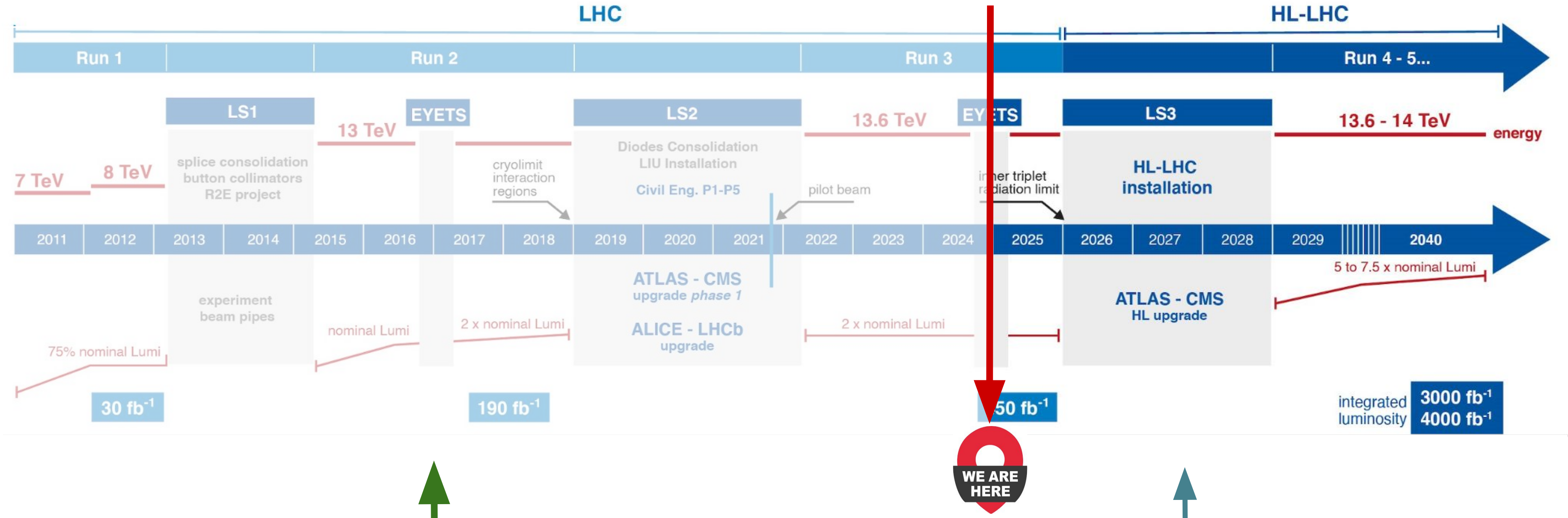
ML techniques for reconstruction developed and used for the future CMS High Granularity Calorimeter

Shamik Ghosh
Laboratoire Leprince-Ringuet, Ecole Polytechnique, CNRS

Polarized Perspectives: Tagging and Learning in the SM , 20-21 Feb 2025

Motivation: HL-LHC

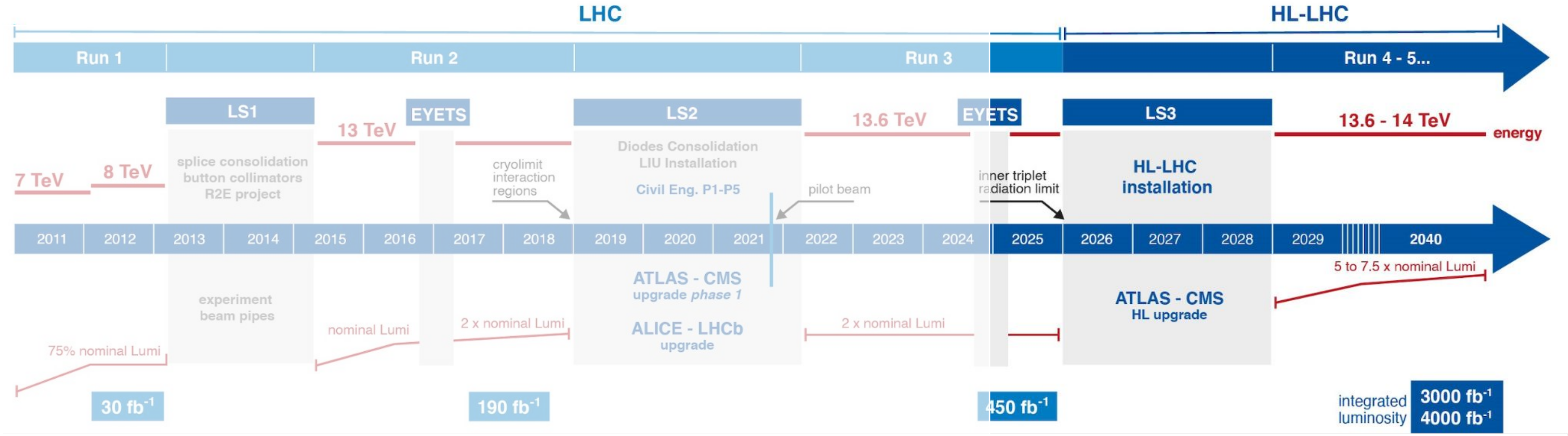


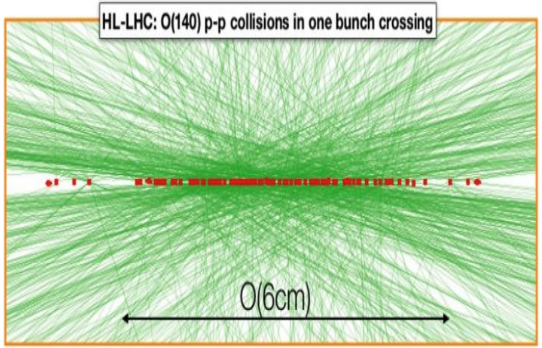
- Remarkable performance so far exceeding initial expectations
- But, things have just begun

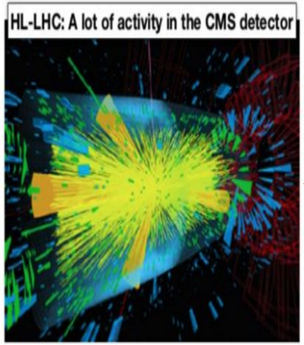
Significant increase in instantaneous luminosity

- 5×10^{34} (7.5×10^{34}) $cm^{-2}\ s^{-1}$ for 140 (200) PU in Run 4 (Run 5)
- Opportunity for Higgs boson precision studies, precision SM tests and BSM searches

Motivation: HL-LHC



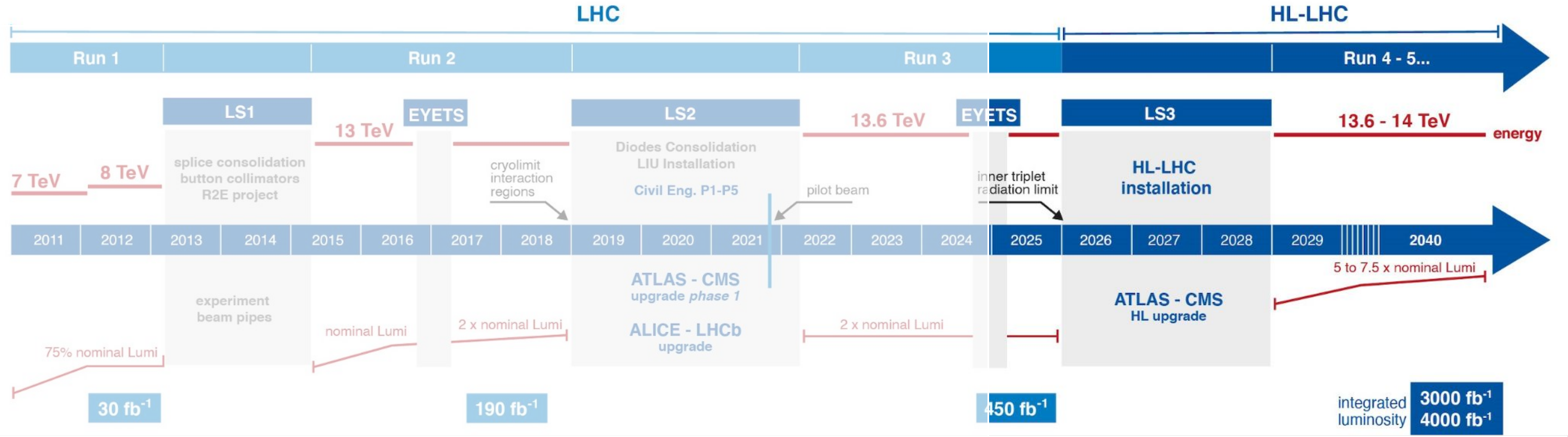




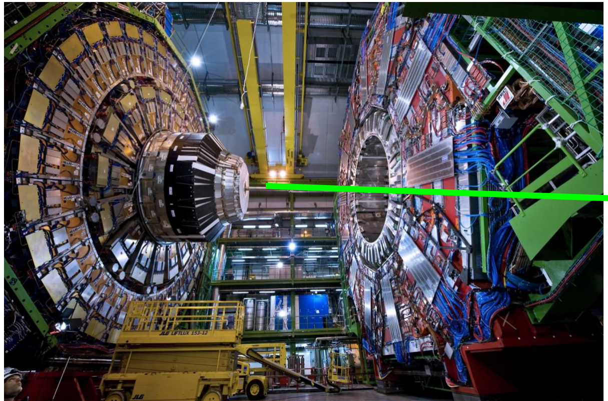
But we have to pay to play!!

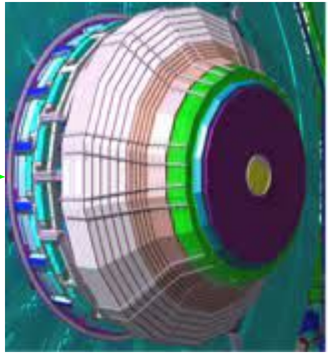
- High Pileup
  - ~200 collisions/BX (4-5x LHC)
- High Radiation Level
  - 1y @HL-LHC ~ 10 y @LHC

Motivation: HL-LHC



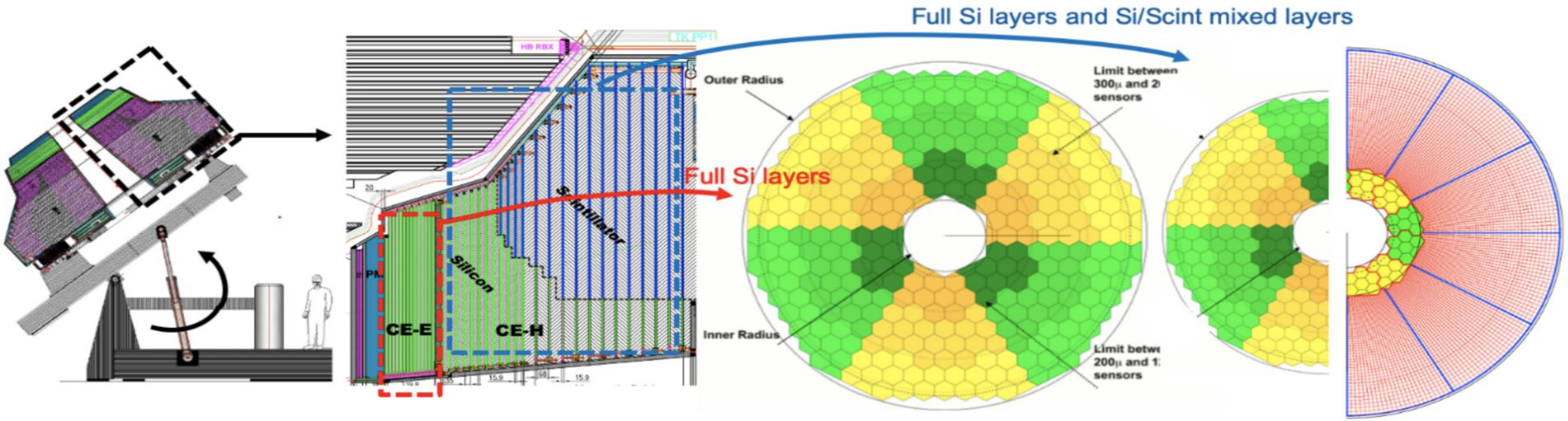
Existing endcap calorimeters will suffer the most → Replace with HGCal





CMS Phase-II Upgrade Endcap Calorimeter

High Granularity Calorimeter (HGCAL): granular and radiation hard endcap calorimeter replacement



Calorimeter Endcap Electromagnetic (CE-E)

- EM focused part

Active material

- 26 Layers of Si (cell size: 0.5-1 cm^2)

Passive material

- Pb, CuW, Cu
- 27.7 X_0

Calorimeter Endcap Hadronic (CE-H)

- HAD focused part (hybrid structure)

Active material

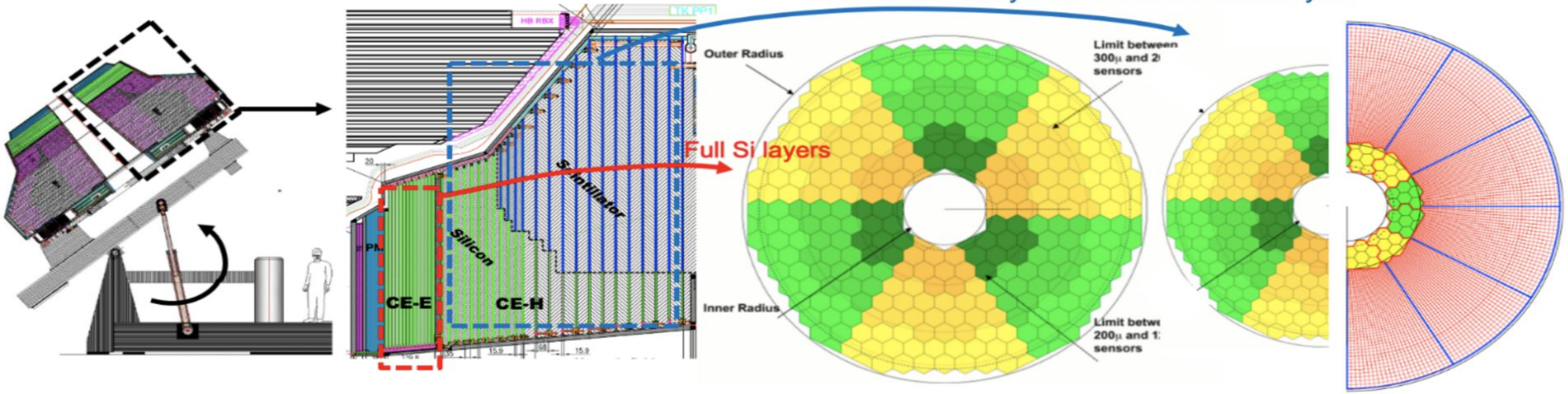
- 7 Layers of Si (cell size: 0.5-1 cm^2)
- 14 Layers of Si and plastic scintillator

Passive material

- Stainless Steel, Cu
- 10.0 λ

CMS Phase-II Upgrade Endcap Calorimeter

High Granularity Calorimeter (HGCAL): <u>granular</u> and <u>radiation hard</u> endcap calorimeter replacement



Calorimeter Endcap Electromagnetic (CE-E)

- EM focused part

Active material

- 26 Layers of Si (cell size: 0.5-1 cm^2)

Passive material

- Pb, CuW, Cu
- 27.7 X_0

~ 6M Si sensor channels

Calorimeter Endcap Hadronic (CE-H)

- HAD focused part (hybrid structure)

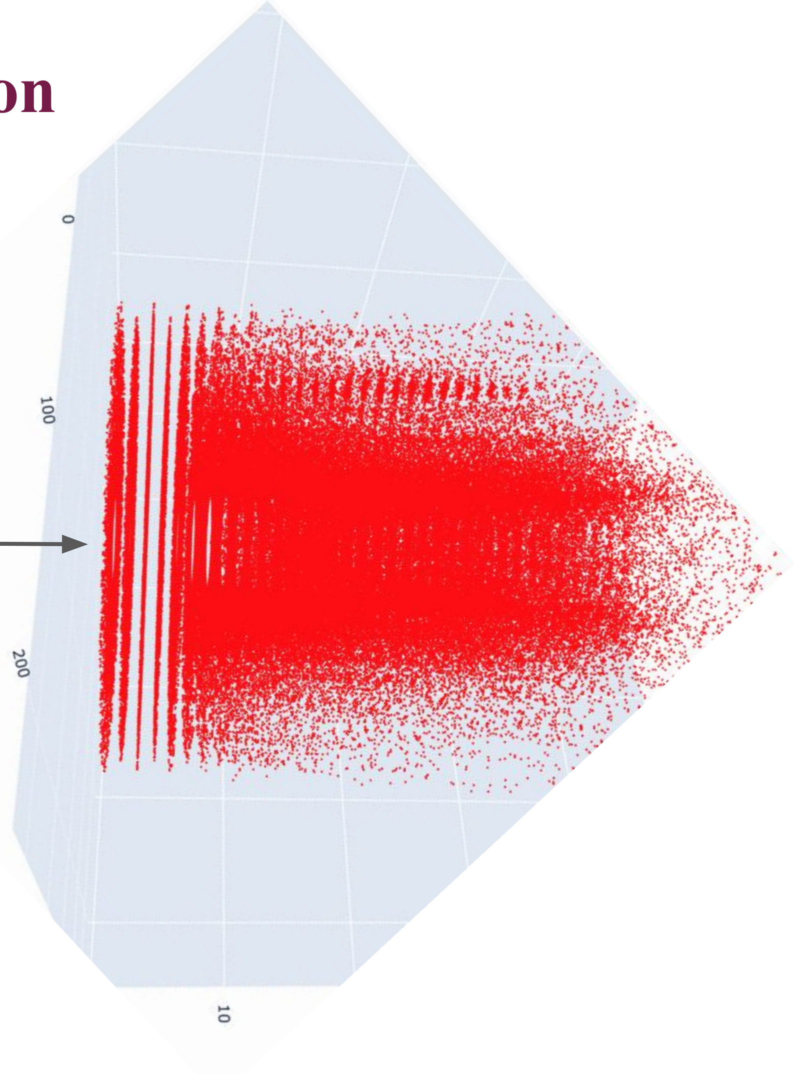
Active material

- 7 Layers of Si (cell size: 0.5-1 cm^2)
- 14 Layers of Si and plastic scintillator

Passive material

- Stainless Steel, Cu
- 10.0 λ

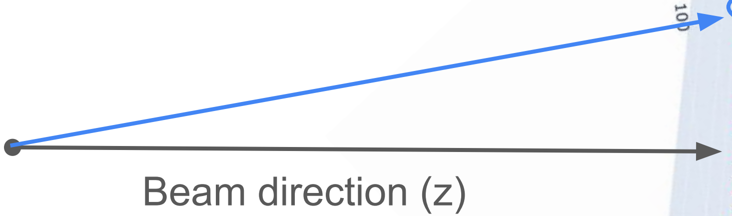
Extremely Granular Information

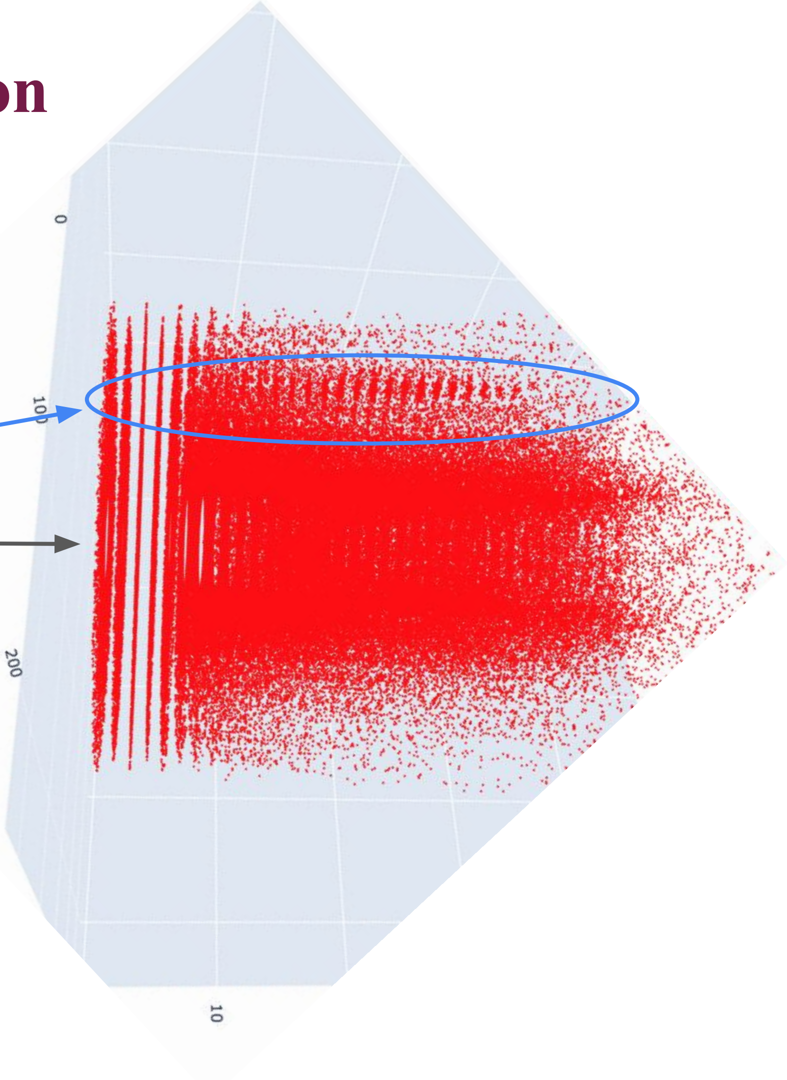


Beam direction (z)

Event display in 200PU

Extremely Granular Information





photon Event display in 200PU

Reconstrction Framework: TICL - The Iterative CLustering

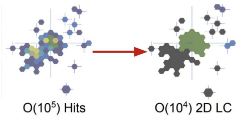
- Modular framework developed inside CMSSW
  - Allows customised iterations targeting specific objects (EM/HAD/MIP)
- Full reconstruction starting from rechits (x,y,z,E,t) to particle properties and identification probabilities
- Framework modern architecture friendly (GPU/FPGA)

Reconstrction Framework: TICL - The Iterative CLustering

- Modular framework developed inside CMSSW
  - Allows customised iterations targeting specific objects (EM/HAD/MIP)
- Full reconstruction starting from rechits (x,y,z,E,t) to particle properties and identification probabilities
- Framework modern architecture friendly (GPU/FPGA)

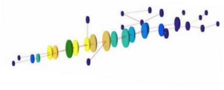
Rechits
Sensor readout
$O(10^5)$/event

CLUE algorithm



Layer Clusters
2D clusters per layer
$O(10^4)$/event

CLUE3D algorithm

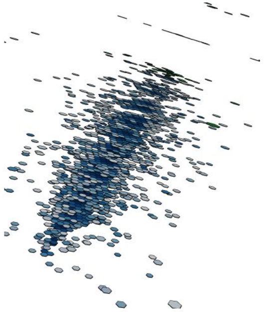


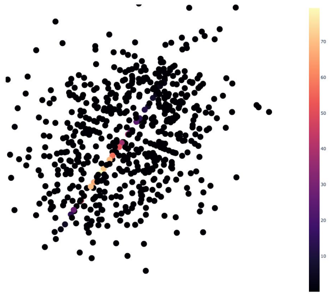
Tracksters
3D clusters (full showers)
$O(10^3)$/event

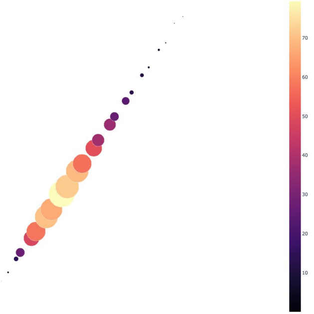
Higher level objects

Particle Flow

500 GeV Photon noPU







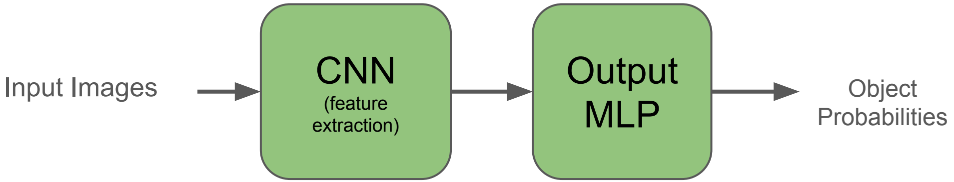
Particle Property Estimation

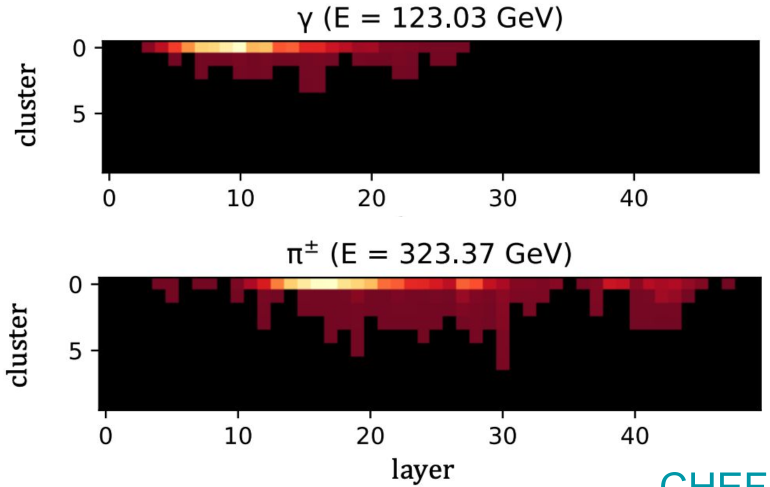
- Assign Identity to reconstructed 3D showers : Particle ID
- Estimate corrected properties: energy, position, timing
- Classical Approach
    - Compute variables describing transverse/ longitudinal spread of showers, electromagnetic and hadronic energy fractions etc
    - Simple but limited performance
- Machine Learning Approach
    - Learn useful representations from full shower information
    - Identify methods most suitable for our data

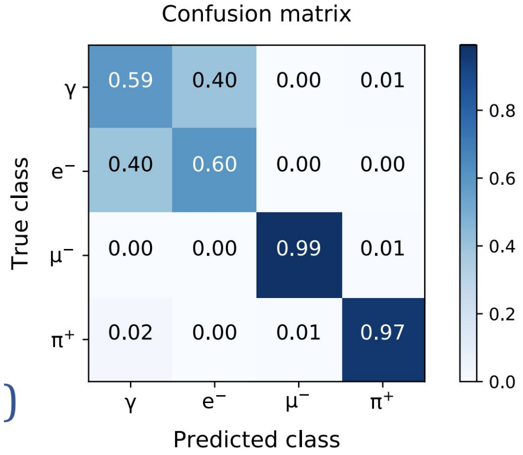
Particle Property Estimation: Particle ID

Distinguish electromagnetic from hadronic objects

- e/gamma VS pions
- Represent data as images
  - 2D clusters ordered by energy
- Images fed into a CNN



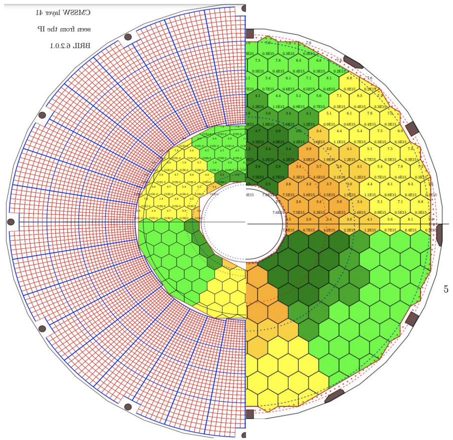


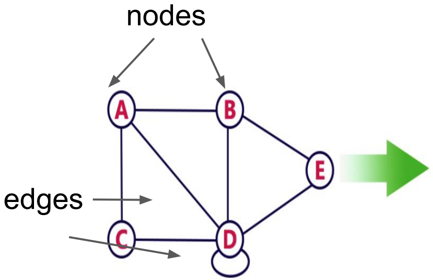


CHEF 2019 Antonio D.P.

Particle Property Estimation: Particle ID

- CNN approach misses granular information
- HGCAL generates heterogeneous data
  - Different sensor types/ sizes/ geometries
- Use graph data structures to handle this complexity
- Use Graph Neural Networks to build shower representations

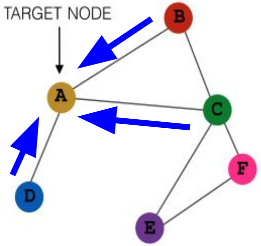




Adjacency Matrix

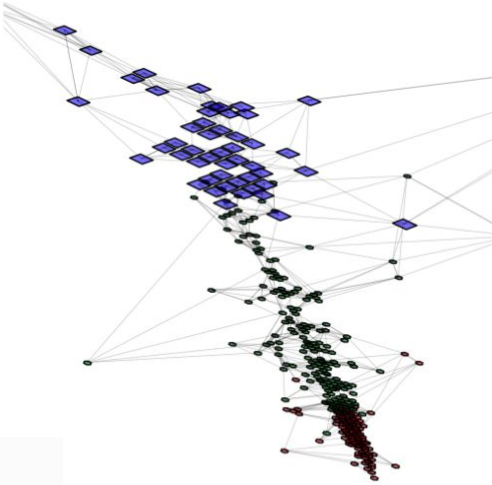
$$\begin{array}{c|ccccc} & A & B & C & D & E \\ \hline A & 0 & 1 & 1 & 1 & 0 \\ B & 1 & 0 & 0 & 1 & 1 \\ C & 1 & 0 & 0 & 1 & 0 \\ D & 1 & 1 & 1 & 1 & 1 \\ E & 0 & 1 & 0 & 1 & 0 \end{array}$$

Graph = (Nodes, Edges)
Nodes, Edges can have their own features



Message Passing in graphs

$$\mathbf{x}_i^{(k)} = \gamma^{(k)}\left(\mathbf{x}_i^{(k-1)}, \bigoplus_{j\in\mathcal{N}(i)} \phi^{(k)}\left(\mathbf{x}_i^{(k-1)}, \mathbf{x}_j^{(k-1)}, \mathbf{e}_{j,i}\right)\right),$$



Node Features
(x,y,z,Energy time, SensorID)

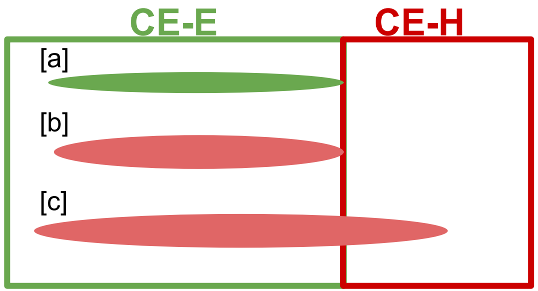
Adjacency using knn

Pion as a graph

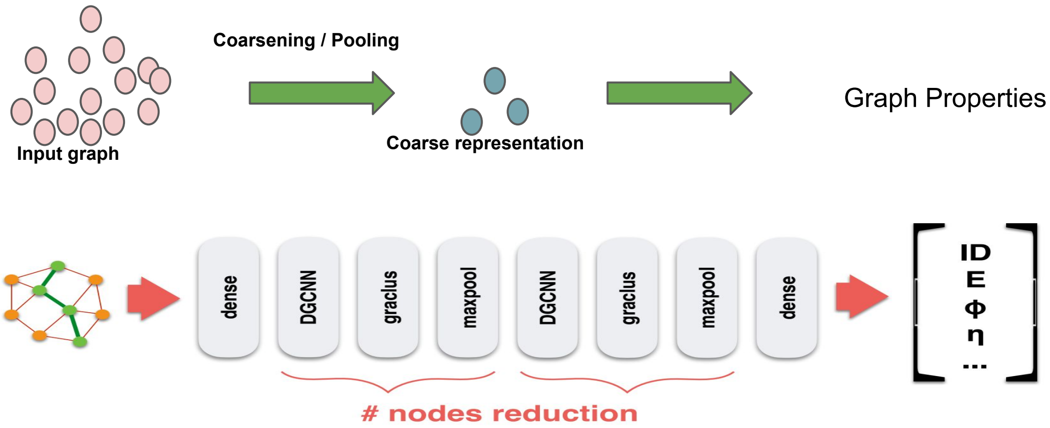
Particle Property Estimation: Particle ID

Distinguish electromagnetic from early showering pions

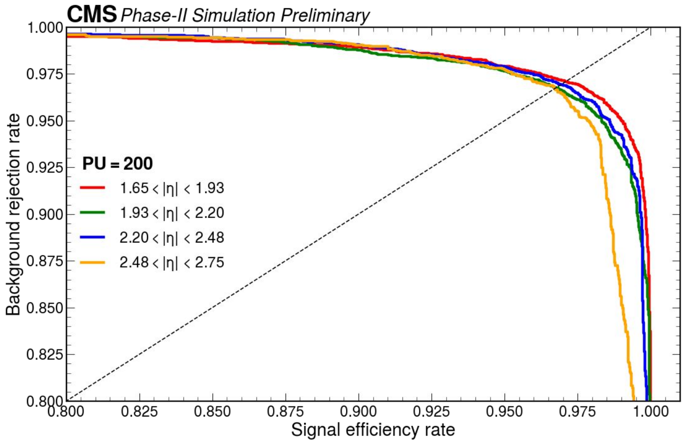
- Represent reconstructed shower as a graph
- Use optimised graph neural networks



[a] photon
[b,c] Early showering pion



Dynamic Reduction Network(DRN)
2003.08013

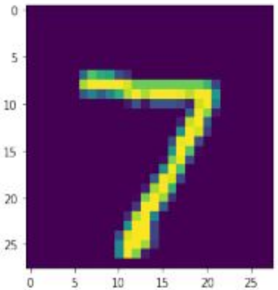


Good background rejections in 200PU

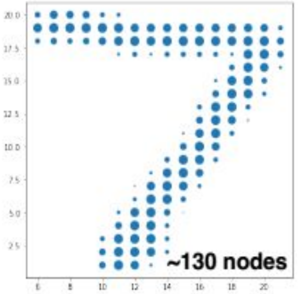
CMS-DP-2022_002

DRN

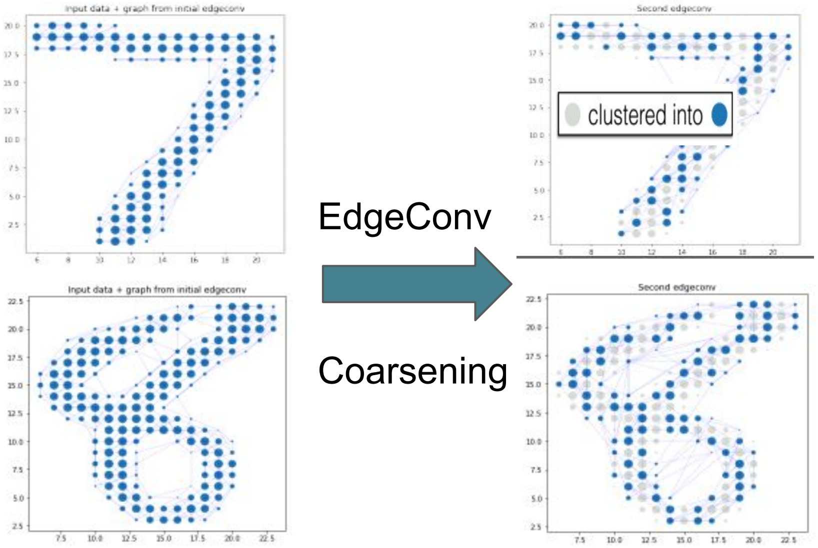
- Applied to MNIST



Keep only non-white pixels, use as nodes of a graph

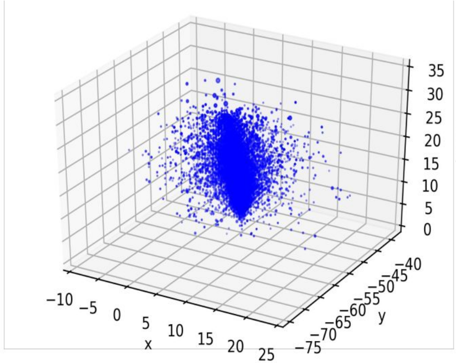


Accuracy = 0.9955 (Hidden dimension=256, k= 4)

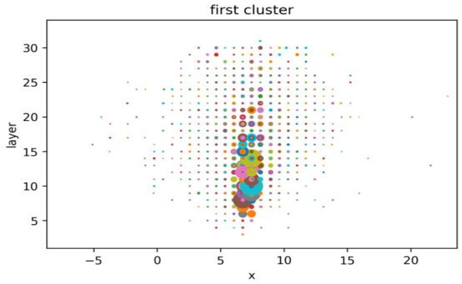


DRN on MNIST learns compact pooled representation to extract features of digits.

- **Applied to photon showers in HGCal**

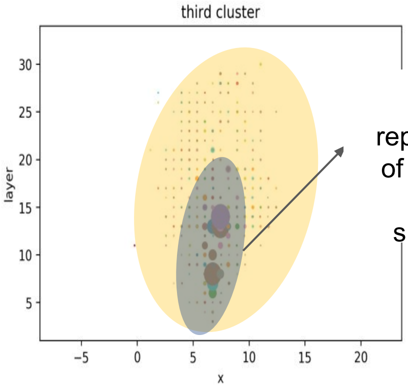


950GeV photon shower



First cluster : points start to coalesce

Cluster x3

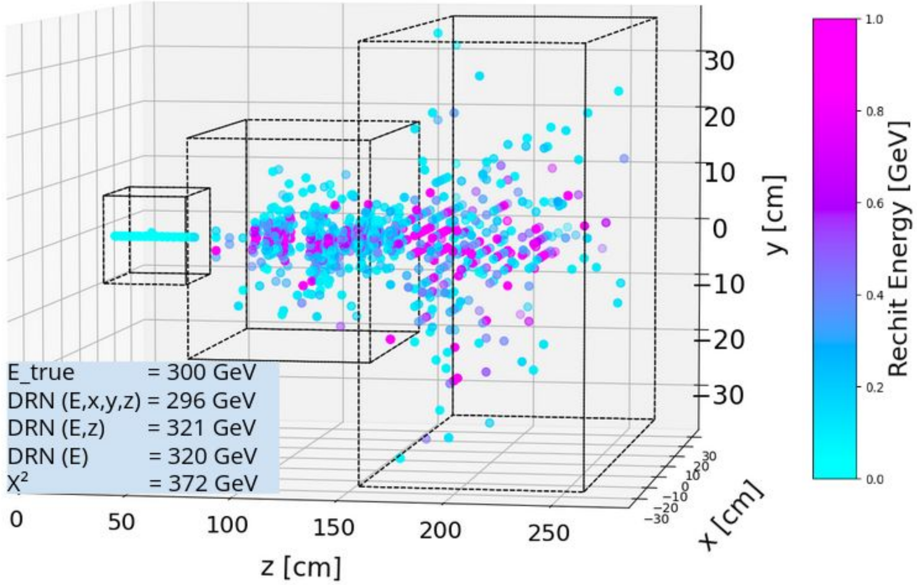


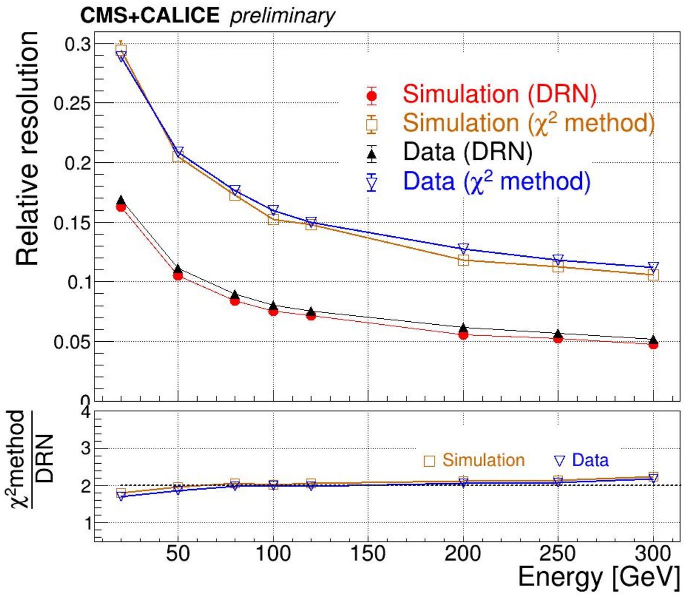
Coarse representation of condensed core with shower info

Particle Property Estimation: Energy Regression

Correct reconstructed energy

- Use pion test beam (2018) data
- Use Dynamic Reduction Network





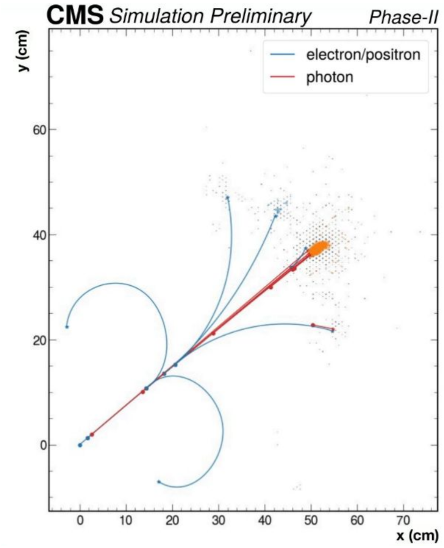
Significant improvement in energy resolution using GNN compared to classic calibration method

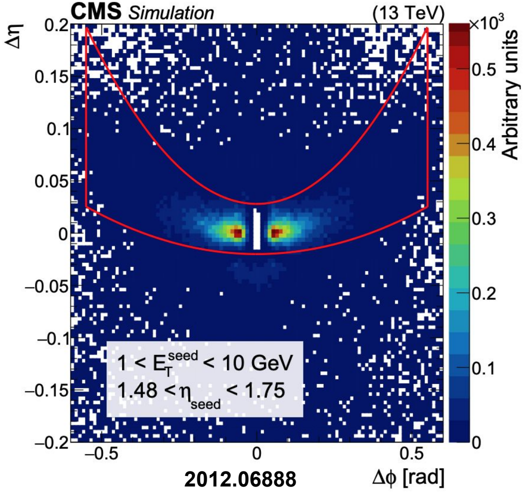
CMS-DP-2022_022

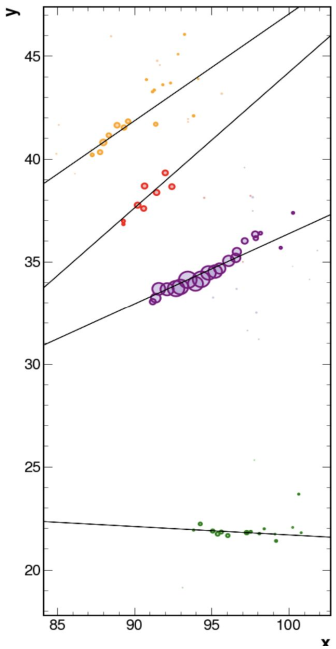
Linking : Higher Level Objects

Electron Superclustering

- Electrons and photons radiate and convert in the tracker and magnetic field spreads the trajectories in φ
  - Superclustering associates these showers
- Currently used algorithm is moustache
  - Geometrical algorithm in eta phi space
- HGCAL provides 3D tracking of showers
  - Utilise direction estimates from cleaned 3D showers using Principal Component Analysis





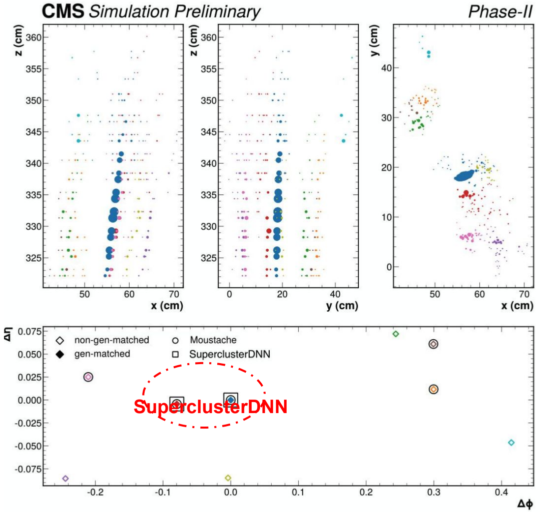


PCA for shower directions

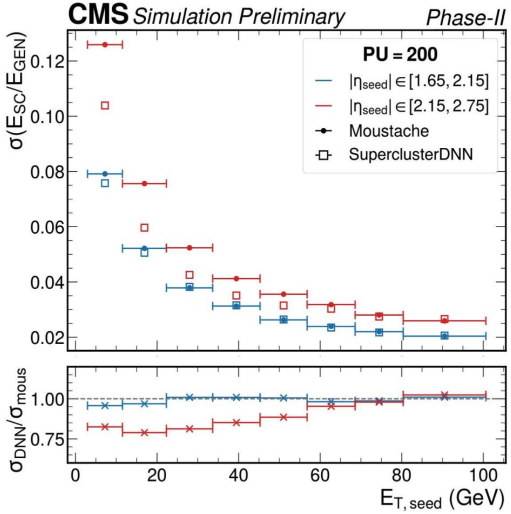
Linking : Higher Level Objects

For a pair of tracksters create variables based on position and angle → **Deep Neural Network** → Probability of coming from the same EM object

Build superclusters iteratively, placing threshold on the pair scores



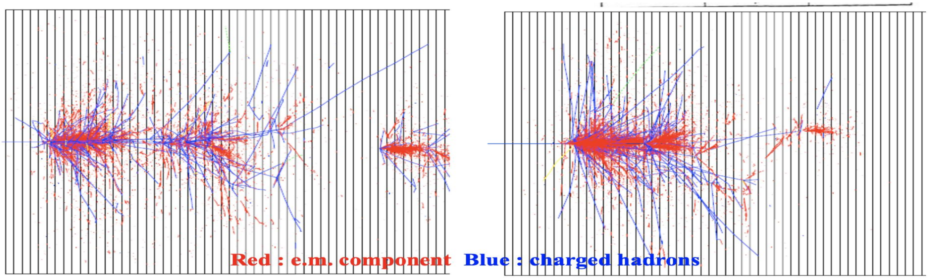
Novel algorithm correctly associates clusters coming from the same particle and rejects clusters from PU



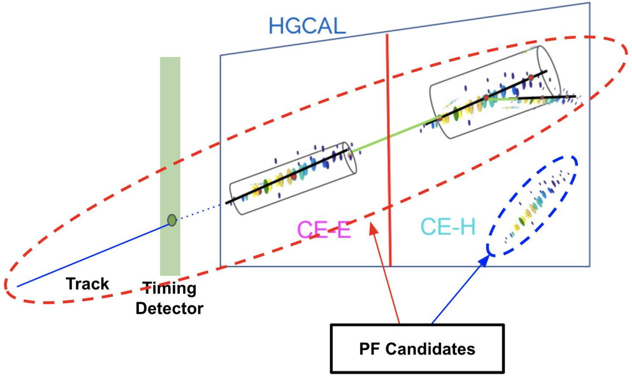
Improved resolution particularly in PU dominated high η

CMS-DP-2024/124

Work in Progress



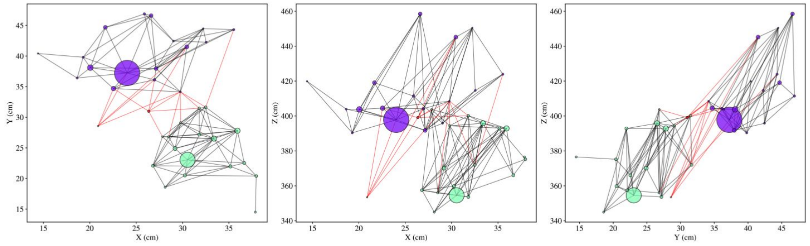
Hadron reconstruction: Linking is tricky



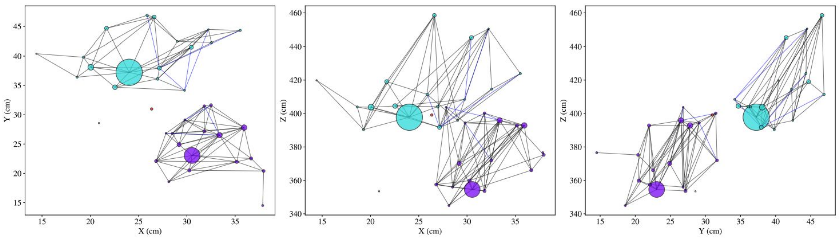
Current approach

- Collect tracksters in a neighbourhood
- Convert data to a graph with the tracksters as nodes and association as edges
- Use a GNN for performing edge prediction

Pair of nearby pions



Valid connections in black and invalid connections in red

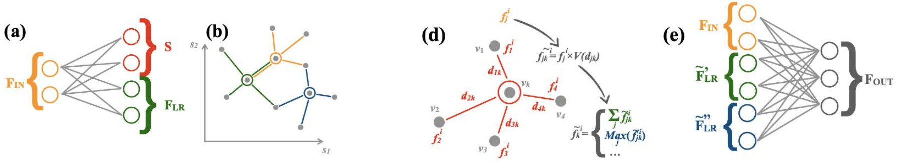


Collecting tracksters based on predicted connection scores

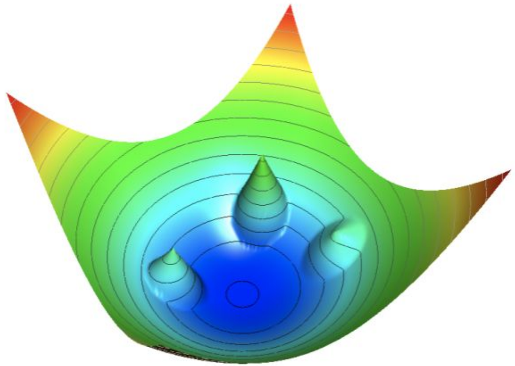
[Thesis of Jekatěrina J.]

End-to-End Reconstruction

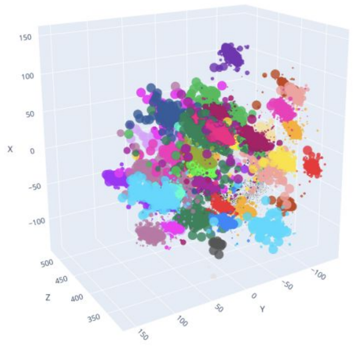
- Aim to estimate particle properties from sensor hits
- Use a GNN ([GravNet]) to learn latent space representation
- Use specialised losses ([Object Condensation Loss])



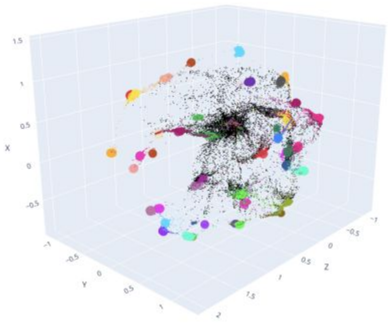
GravNet



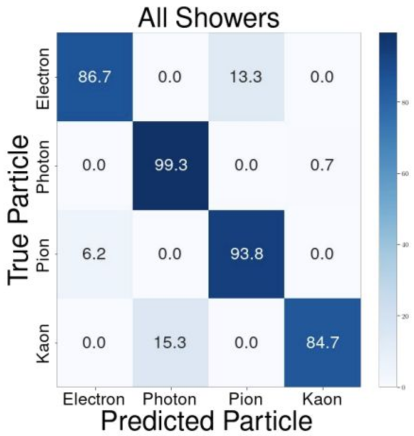
Object Condensation loss for a point with a minima at its matching point and maximas around nearby non-matching points



Shower in input space



Shower in clustering space



Promising PID performances

[CHEP 2024 Philipp Z.]

Conclusion

- HGCAL reconstruction is a challenging task
  - Overlapping showers in 200PU
- TICL is the current reconstruction framework inside CMSSW
- ML algorithms at key steps in the reconstruction workflow
  - A mix of traditional and ML algorithms
  - CNNs, GNNs, Transformers
- ML playing increasing role in higher level combinatorics
- Lots of interesting developments ongoing
- Alternative fully ML driven strategies also being developed

Thanks for your attention!

TICL : Computing performance

- HGCAL reconstruction currently takes only around 5% of the total Phase-2 CMS reconstruction time
- Further decrease expected with offloading of algorithms to GPUs