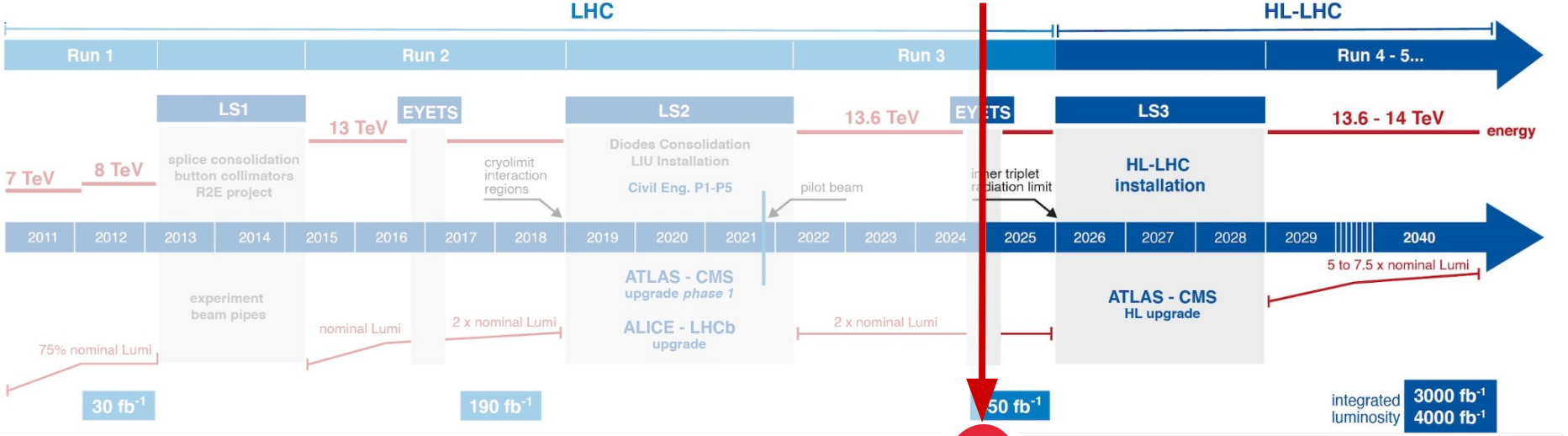
# Challenges and novel reconstruction techniques for the CMS High Granularity Calorimeter for HL-LHC

**Shamik Ghosh**
(Laboratoire Leprince-Ringuet, École Polytechnique)
on behalf of the CMS Collaboration

XXVI DAE-BRNS High Energy Physics Symposium 2024
19–23 Dec 2024

# Motivation: HL-LHC

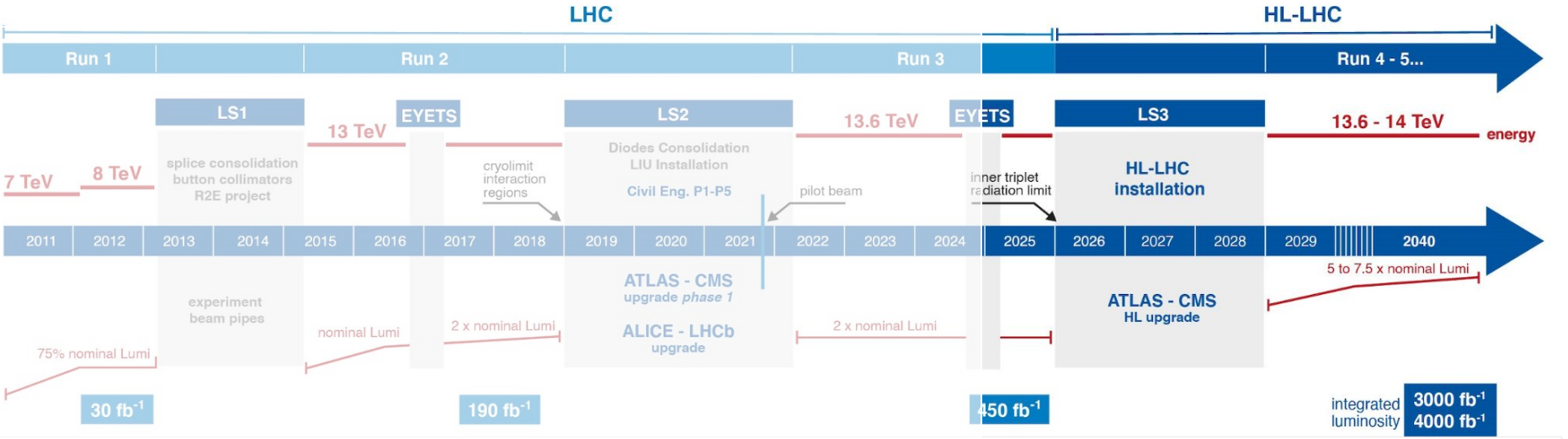


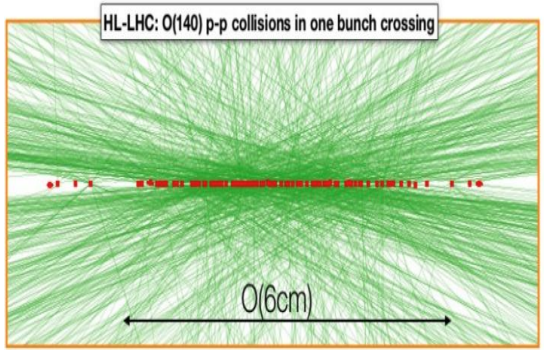

- Remarkable performance so far exceeding initial expectations
- But, things have just begun

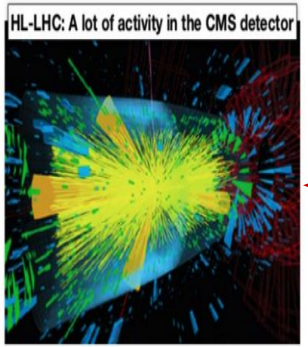
**Significant increase in instantaneous luminosity**

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

# Motivation: HL-LHC



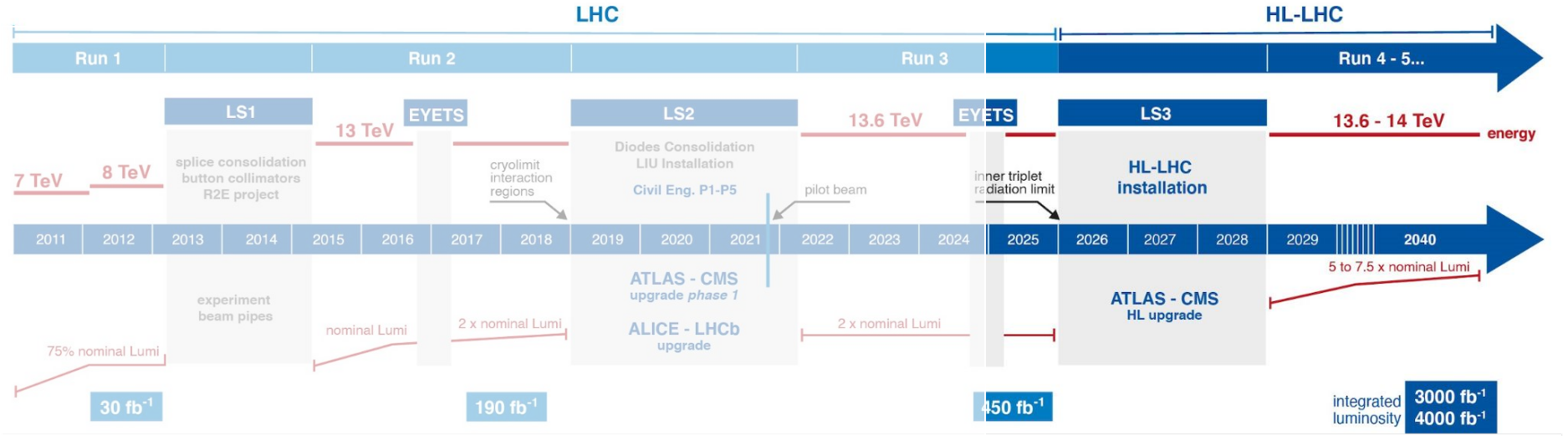



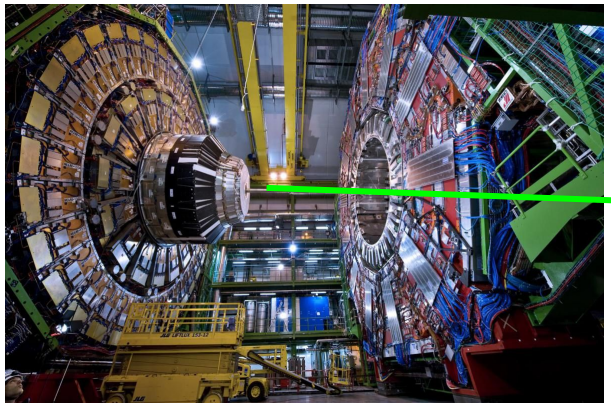

**But we have to pay to play!!**

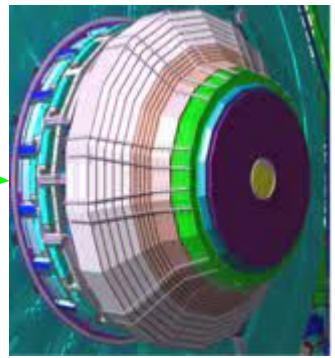
- High Pileup
  - ~200 collisions/BX (4-5x LHC)
- High Radiation Level
  - 1y @HL-LHC ~ 10 y @LHC

# Motivation: HL-LHC




**Existing endcap calorimeters will suffer the most → Replace with HGCal**

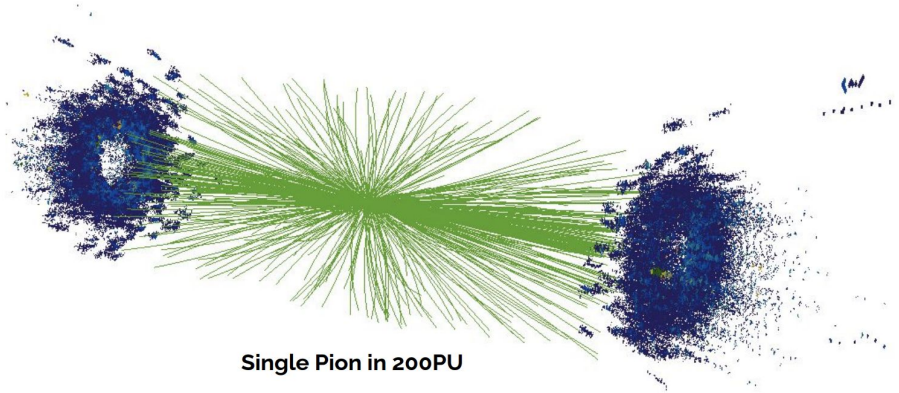


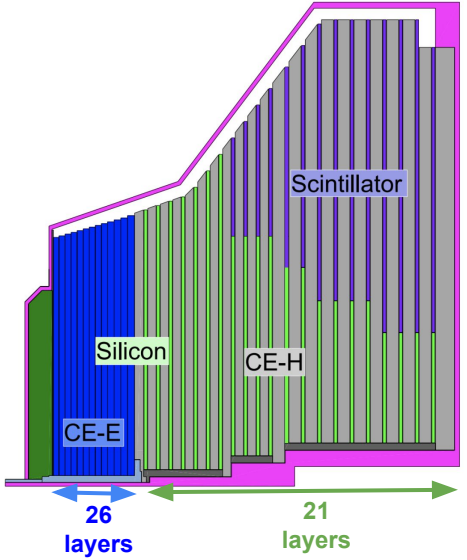


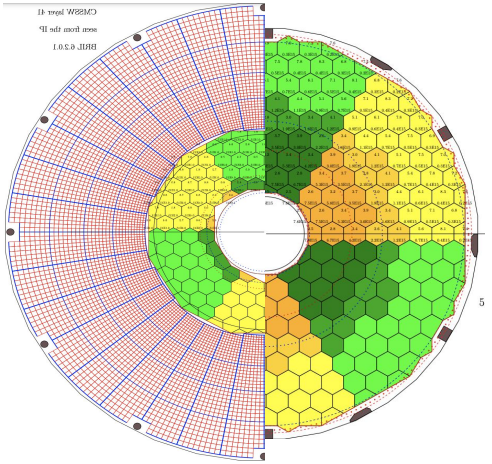
# Reconstruction in HGCAL

Reconstruction in HGCAL is a challenging task in high PU environments

- Extremely granular detector with ~47 layers of silicon+absorbers and scintillators
  - High density (0.5 $cm^2$)/ Low density (1.1 $cm^2$) silicon sensors
  - ~ 6M channels
  - 500k hits per event (x,y,z,E,t)
- Overlapping showers are the norm



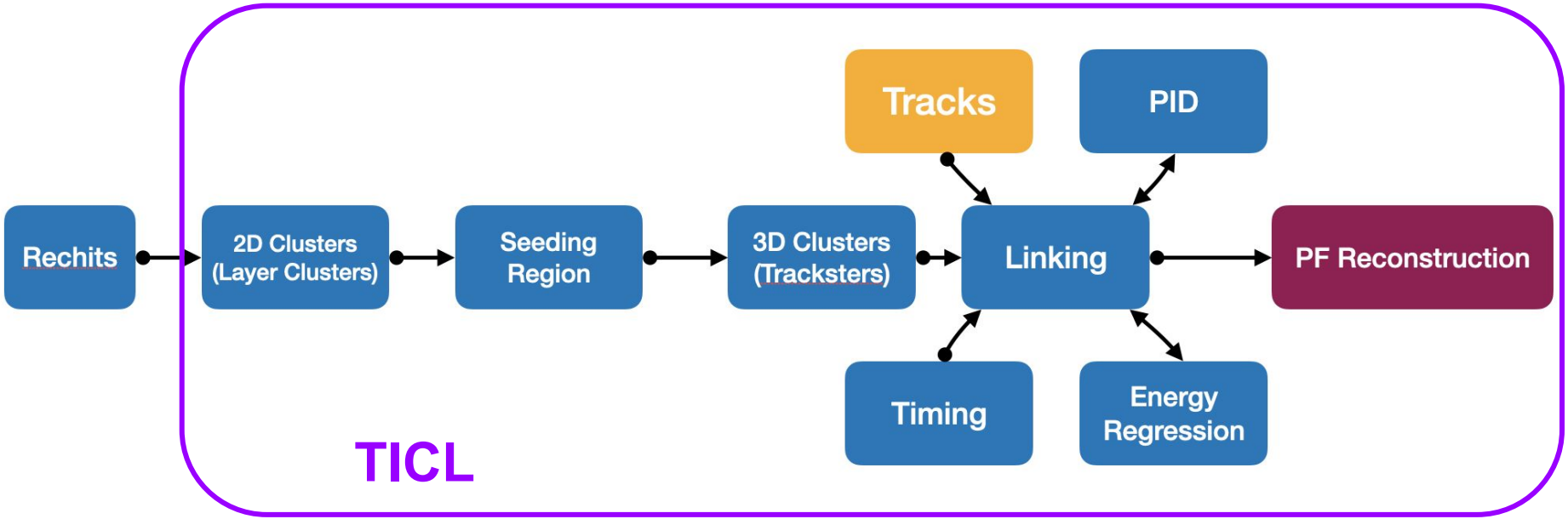



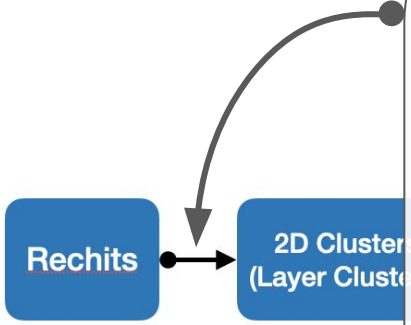
# TICL - The Iterative CLustering

- Modular framework developed inside CMSSW
  - Swap algorithms easily
  - Possibility of separate iterations targeting specific particles
  - Validation for monitoring performance w.r.t changes in release
  - Machine Learning algorithms in key steps
- Full reconstruction starting from rechits (x,y,z,E,t) to particle properties and identification probabilities
  - Particle flow candidates with global event description
  - Utilise tracking and timing information
- Framework modern architecture friendly (GPU/FPGA)
  - Data structures and algorithms built with parallelism in mind

# TICL workflow

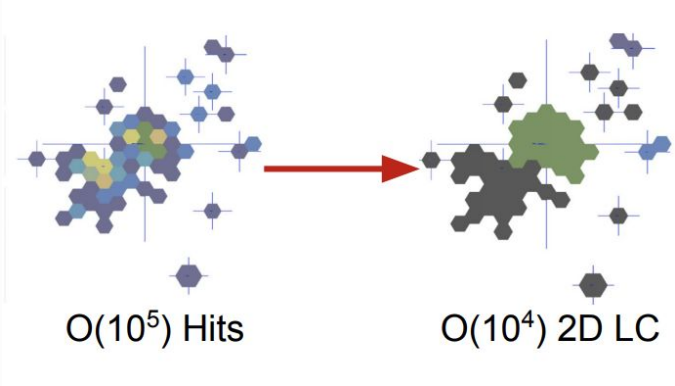


# TICL workflow

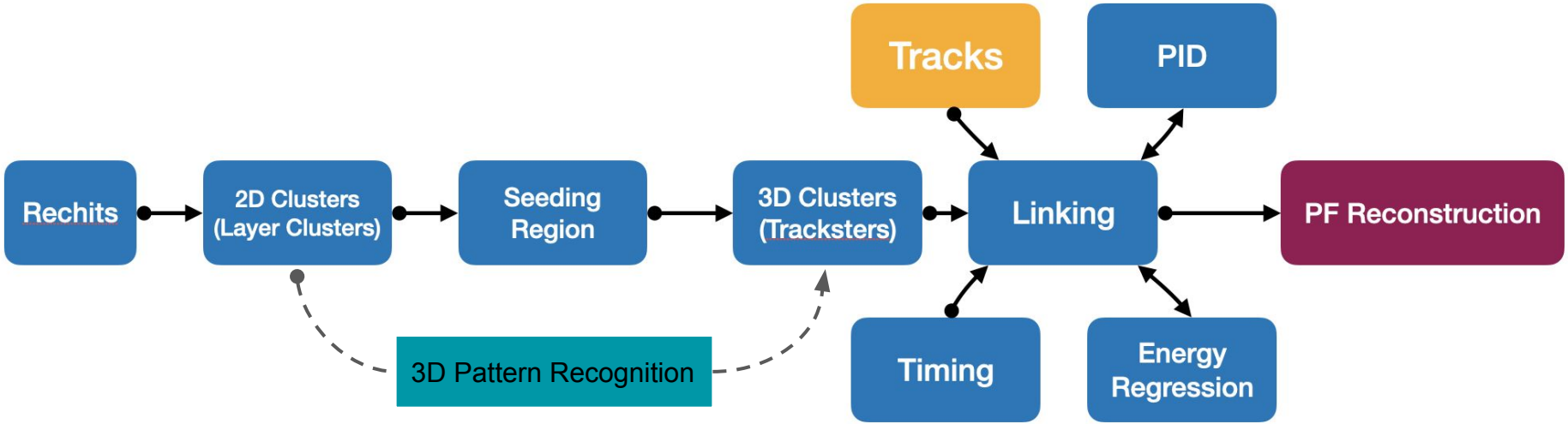



**CLUE : CLUstering of Energy**

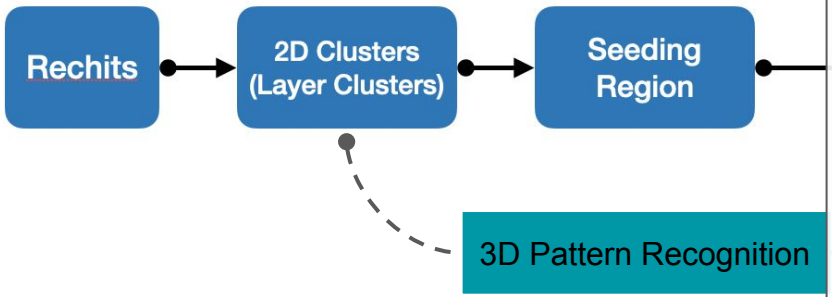
- Energy density based 2D per layer clustering algorithm
  - Noise reduction
- Produces 2D Layer Clusters (LCs) starting from rechits
  - Reduction of problem size by an order of magnitude
- Highly parallelizable
  - 300 events/second (ttbar 200 PU)



# TICL workflow



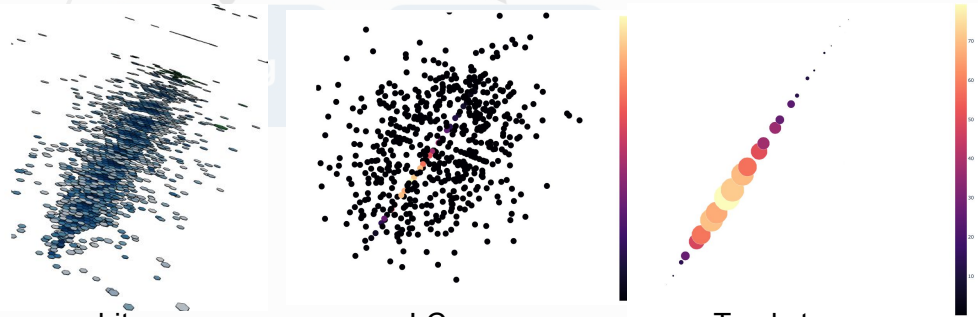
# TICL workflow




Pattern recognition algorithm connects 2D LCs to form 3D clusters

- Current algorithm is CLUE3D
  - Alternatively Cellular Automaton, Fastjet
- 3D clusters are graphs of LCs
  - Called tracksters
- Reduction of problem size by another order of magnitude
  - O($10^4$ ) LCs to O($10^3$ ) Tracksters

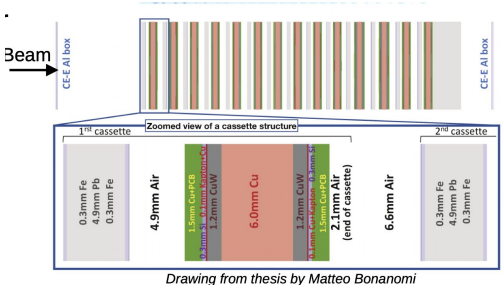
**500 GeV Photon noPU**



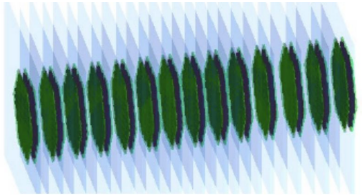
rechits LCs Trackster

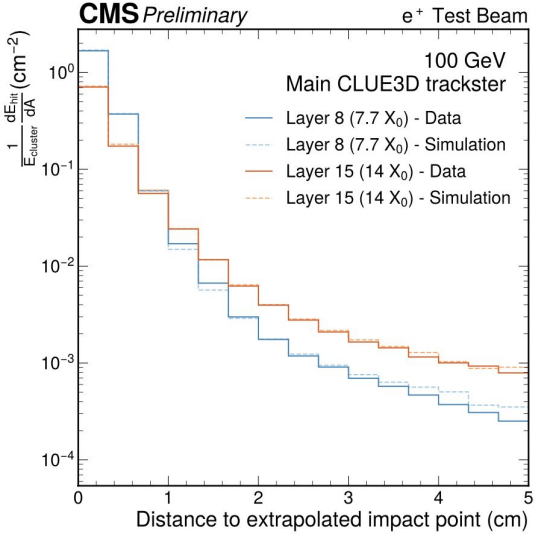
# Testing with 2018 test beam data

CMS-DP-2023_092

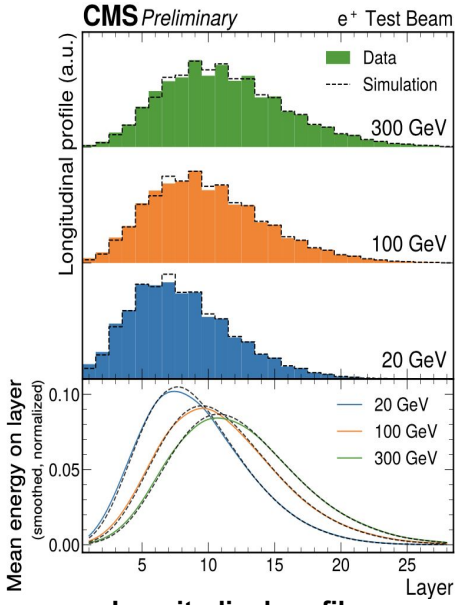


Drawing from thesis by Matteo Bonanomi

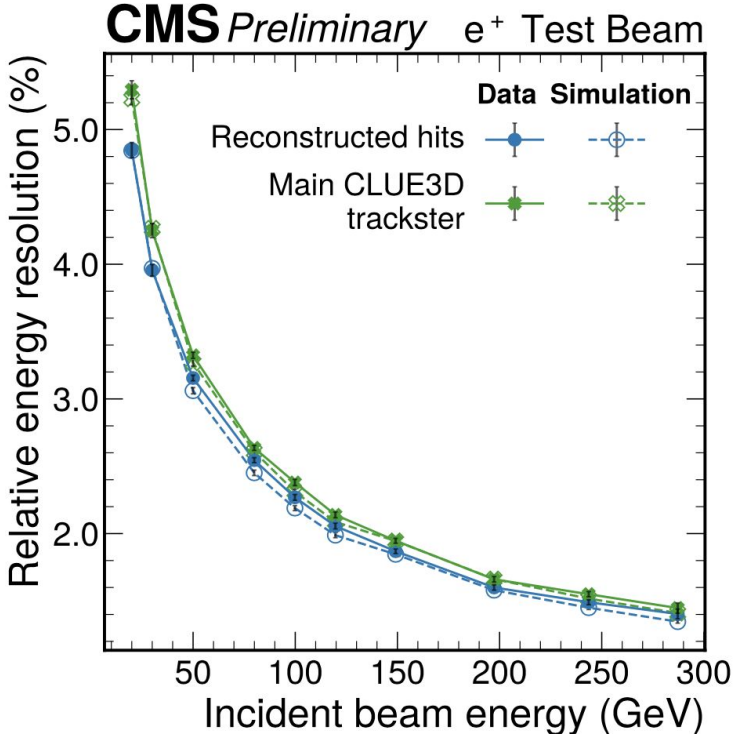




**Transverse profile**



**Longitudinal profile**

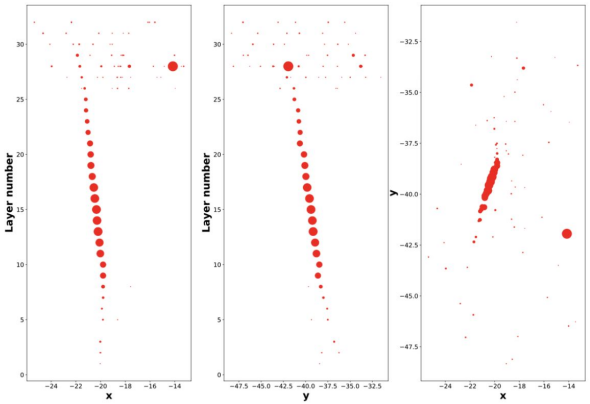


Energy resolution of clustered object slightly worse than reconstructed hits

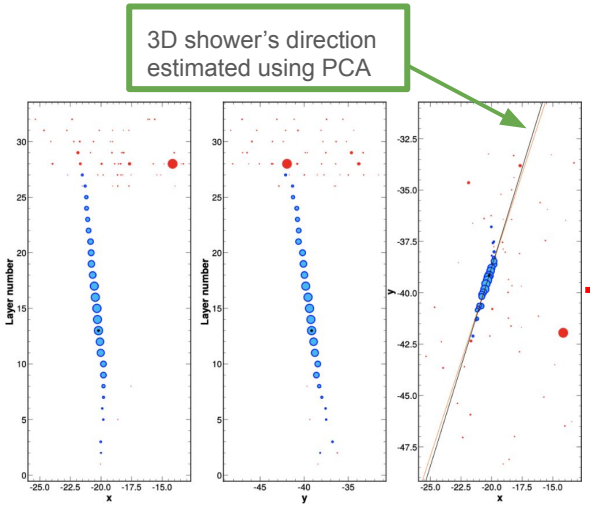
# Linking : Electromagnetic object reconstruction

CMS-DP-2022_057

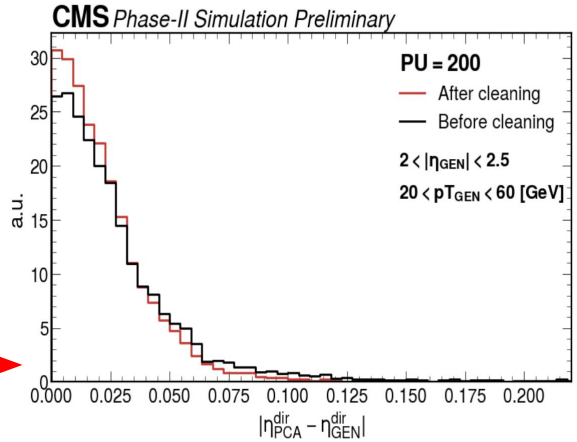
- Reconstructed 3D clusters have contaminations from PU/nearby objects
- Clean tracksters using algorithm based on shower geometry
  - Gain precise direction estimates using PCA




Tracksters before cleaning
[all hits in red are one object]



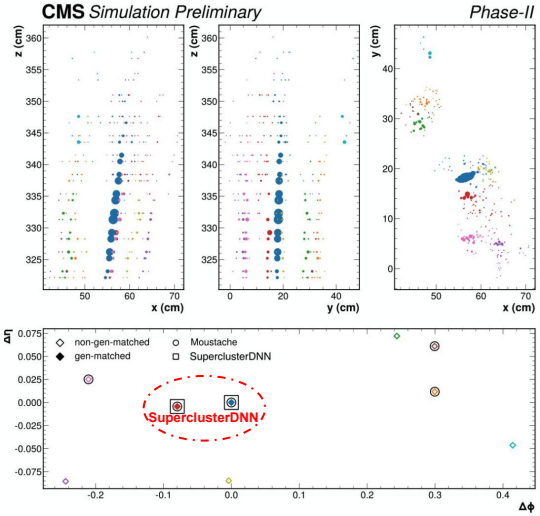

Tracksters after cleaning
[only hits in blue remain after cleaning]



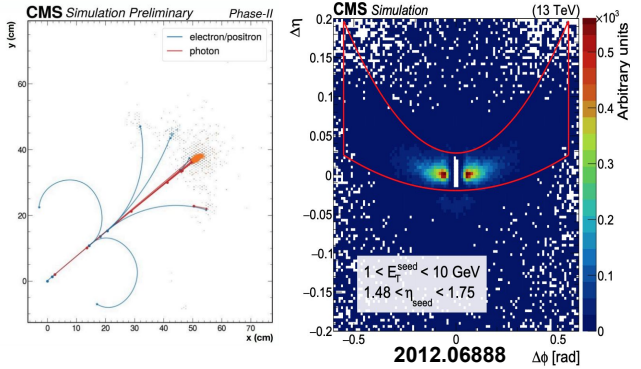

Trackster direction estimates before and after cleaning shows improvement after cleaning

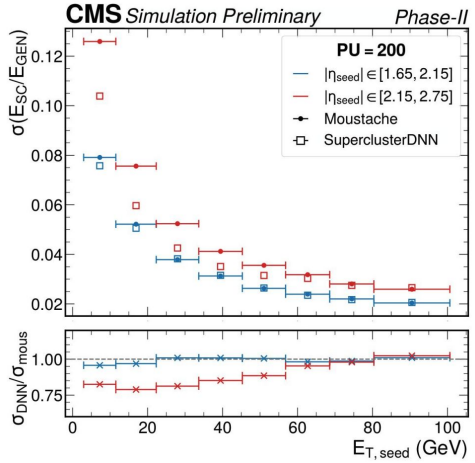
# Linking : Electromagnetic object reconstruction

- Electrons and photons radiate and convert in the tracker and magnetic field spreads the trajectories in φ
- Linking these clusters is called superclustering
  - Default algorithm called Moustache:  geometrical algorithm in η-φ space




Novel algorithm uses combination of angular variables constructed from directions estimated using  PCA fed into a Deep Neural Network

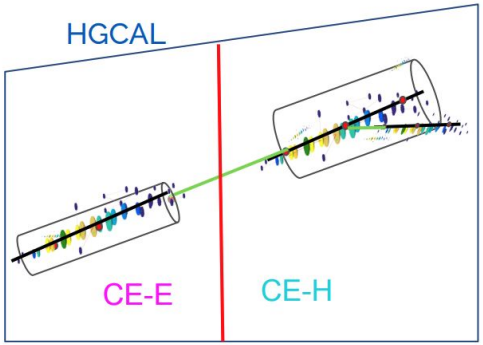


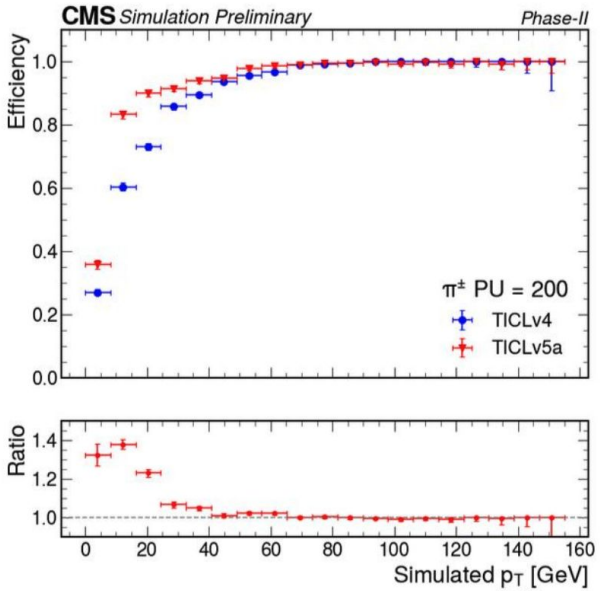



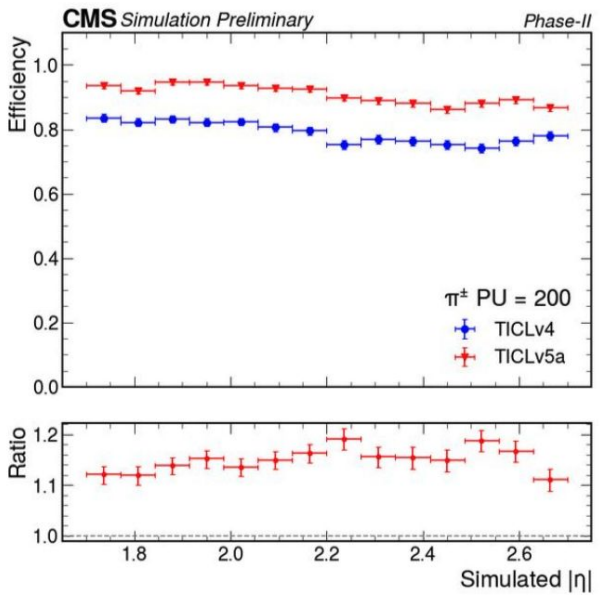

Improved resolution particularly in PU dominated high η

# Linking : Hadron reconstruction

- Hadrons produce multiple clusters inside HGCAL
- Previous algorithm used geometric linking
- New linking algorithm is called Skeletons
  - Utilises topology of clusters and PCA axes



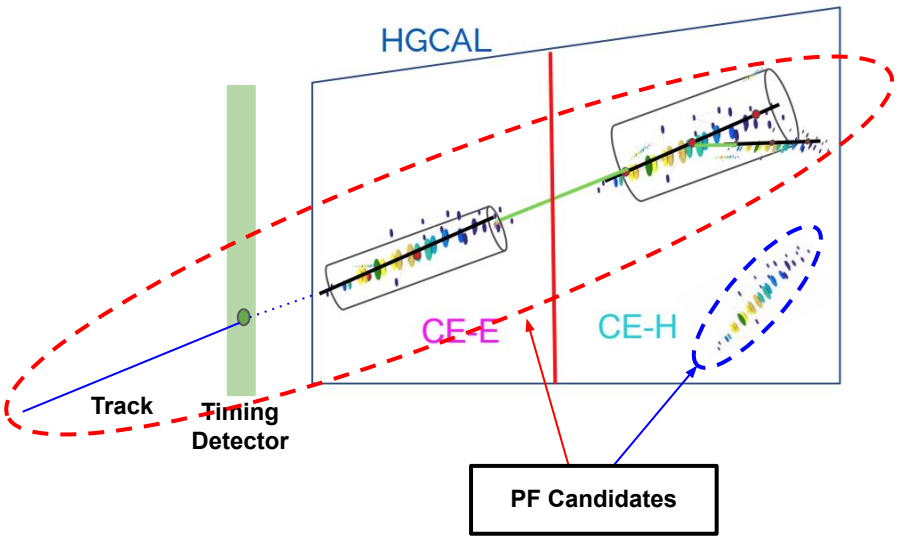



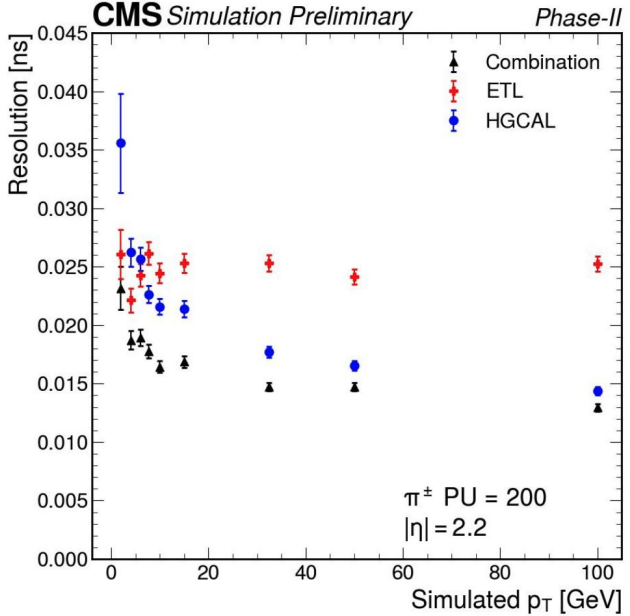




Improvement in efficiency due to to newer algorithms

# Linking : Particle Flow Candidates and timing

- Tracks and linked hadron clusters are further linked to form Particle Flow Candidates
  - Use timing, position, energy compatibility
- Timing information available from HGCAL and Endcap Timing Layer (ETL)
  - Combine timing information for PF Candidates
  - Time is propagated to primary vertex using charged or neutral particle hypotheses




~15ps time resolution

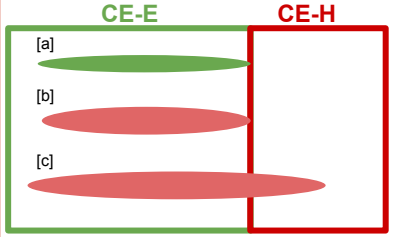
# Novel Machine Learning techniques

- Good particle Identification and regressions require good description of reconstructed showers
- Novel Graph Neural Networks (GNN) adapt to heterogenous detector data and learn powerful representations of particle showers

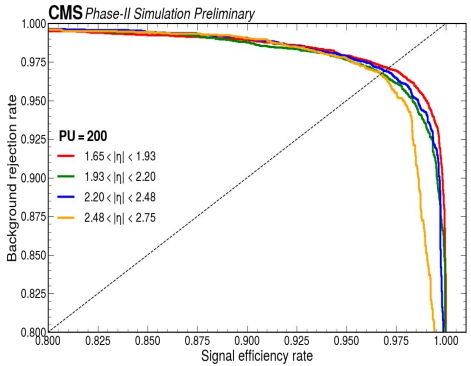
**Dynamic Reduction Network**

## Particle Identification

CMS-DP-2022_002



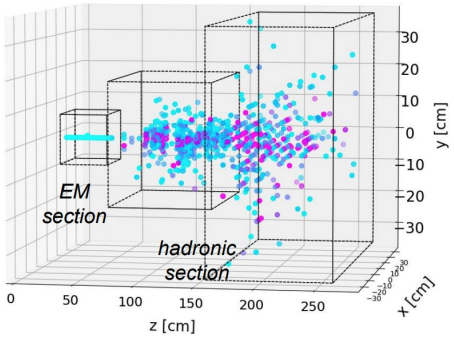

[a] photon
[b,c] Early showering pion



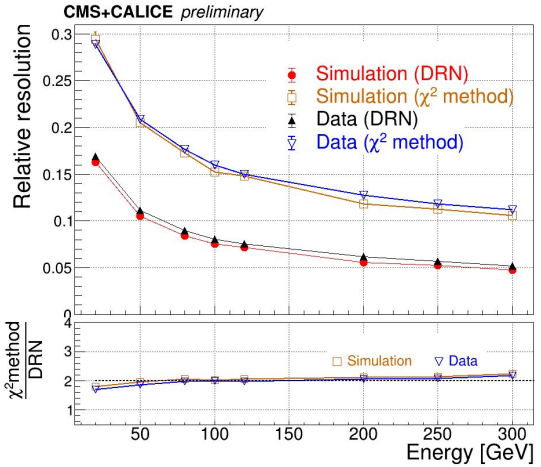

Good performance in separating photons from early showering pions in 200 PU

## Energy Regression

CMS-DP-2022_022



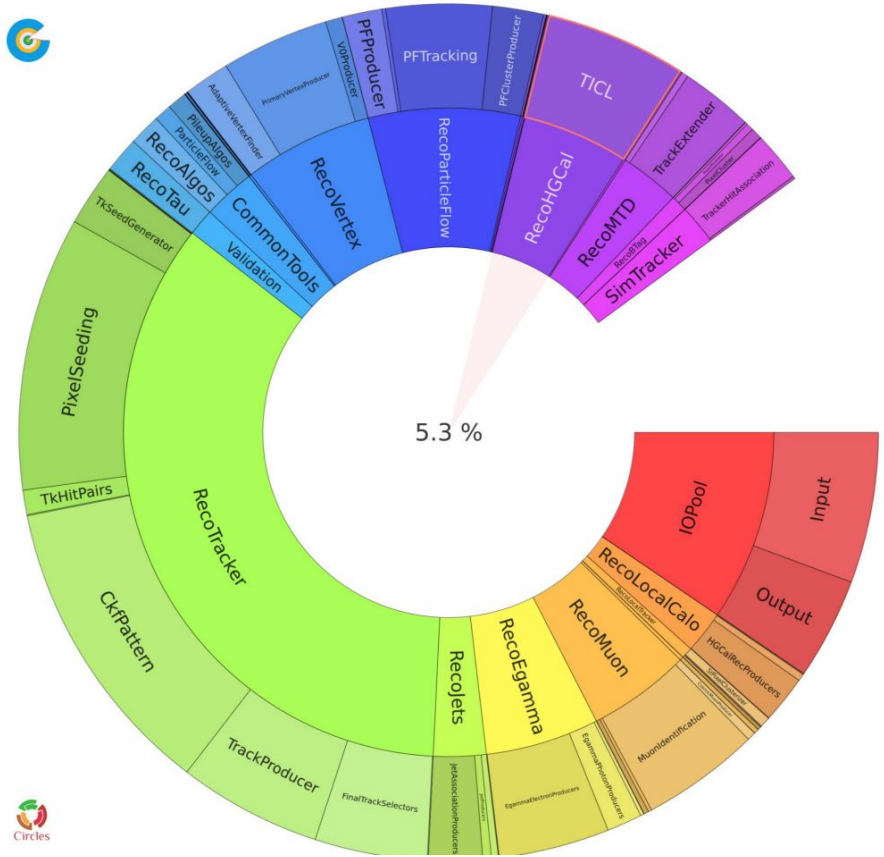

Reconstructed hits in 2018 pion test beam data are input to the GNN




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

# 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



# Conclusions

- Reconstruction in CMS High Granularity Calorimeter poses **unprecedented challenges**
  - “**Tracking**” detector with high granularity
- **TICL** is a highly modular and flexible framework developed in CMS for HGCAL reconstruction
  - Performs reconstruction from raw sensor hits to high level particle flow interpretations
  - Variety of pattern recognition algorithms can be **plugged in** and out
  - Different strategies for different particles
- Variety of strategies being actively explored for best performance in **200 PU**
  - Utilise **novel machine learning** techniques in key reconstruction steps
  - Identify parallelisable reconstruction tasks
- Reconstruction to become even faster with increasing GPU offload


Thank You