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

Shamik Ghosh (LLR) on behalf of the CMS Collaboration
Lepton - Photon 2021 (Manchester) January 11, 2022
Motivation: HL-LHC

- Remarkable performance so far exceeding initial expectations
- But, things have just begun with ~5% of total expected data collected

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
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 cm²)
- 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 cm²)
  - 14 Layers of Si and plastic scintillator
- Passive material
  - Stainless Steel, Cu
  - 10.0 $\lambda$
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 cm²)
  - Passive material
    - Pb, CuW, Cu
    - $27.7 \times X_0$
  - ~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 cm²)
    - 14 Layers of Si and plastic scintillator
  - Passive material
    - Stainless Steel, Cu
    - $10.0 \lambda$
Reconstruction in HGCAL

Tracks and clusters clearly identifiable by eye throughout most of detector.

the longitudinal shower footprint

Calorimeter clusters

high $p_T$ jet $O(500 \text{ GeV})$

Event display for technical proposal in 140 PU
Lindsey Gray, FNAL
Reconstruction in HGCAL

Reconstruction in HGCAL is a real challenge due to the granularity and high PU environment
- Imaging calorimeter with very fine lateral and longitudinal segmentation, and precision timing capabilities
- Naive reconstruction algorithms based on considering all possible combinatorics lead to memory/timing explosion
- Overlapping showers are frequent in high PU and require efficient algorithms to disentangle them

New techniques and algorithms to extract signals belonging to individual showers and properly identify them (clustering, linking, particle identification)
- Utilise modern computer architectures, graph theory, machine learning etc

Efficient workflows to utilise information from the tracking and timing detector
**TICL - The Iterative CLustering**

- Framework to produce 3D clusters and particle properties starting from HGCAL hits \((x, y, z, E, t)\)
  - Inspired from the successful CMS Run1 iterative tracking reconstruction strategy
- Framework is modular allowing for swapping of algorithms according to particle type
- Separate subdetector-based iterations and iterations using information from other subdetectors

---

**TICL Iterations**

- L1 object
- Tracks

**Hits**

- Pattern Recognition algorithm plugin

---

**Filter LC by Mask**

- Mask Selected LC
- Cleaning

---

- Filter based on
  - PID
  - cleaning

---

- Better shower property estimates
- Better subsequent iteration
CLUE [1] - CLUstering of Energy

- Energy density based 2D clustering algorithm
  - Removes noise
- Produces “Layer Clusters (LCs)” starting from hits
  - Dimensionality reduction by an order of magnitude

Fully parallelizable GPU-ready algorithm
- 0.8% of total offline reconstruction on single CPU core @PU200
- 2 orders of magnitude faster on small GPU NVIDIA T4

Pattern Recognition Algorithms

Pattern Recognition algorithms connect 2D LC’s to form 3D clusters called “tracksters”
Pattern Recognition Algorithms

Pattern Recognition algorithms connect 2D LC’s to form 3D clusters called “tracksters”

- Tracksters are Direct Acyclic Graphs
  - Nodes are layer clusters
  - Edges are defined according to connecting algorithm

- Currently available connecting algorithms
  - Cellular Automaton
  - CLUE3D (energy density based clustering using 2D layer clusters)
  - FastJet

- Tracksters are linked to form high level particles

Shower reconstructed as graphs using CA(center) and CLUE3D (middle/right)
Putting it all together : Object Reconstruction

Demonstrating by example : Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters

![Tracksters efficiency for different pattern recognition algorithms](image)
Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters

Background from hadrons in the presence of PU

[a] photon
[b,c] Early showering pion
Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions

Unconverted Photons and early showering pions in 200 PU

ROC curve (left) and discriminator scores (right) for PID based on Edge-convolution and greedy clustering based pooling [1]

Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions
- Cleaning tracksters to get better estimates of properties

Unconverted Photons and early showering pions in 200 PU

Cleaning tracksters to get rid of PU/nearby particle contributions based on shower geometry

Tracksters before cleaning [all hits in red are one object]
Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration

- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions
- Cleaning tracksters to get better estimates of properties

Unconverted Photons and early showering pions in 200 PU

Cleaning tracksters to get rid of PU/nearby particle contributions based on shower geometry

Tracksters after cleaning [only hits in blue remain after cleaning]
Putting it all together: Object Reconstruction

**Demonstrating by example: Electromagnetic iteration**
- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions
- Cleaning tracksters to get better estimates of properties

Unconverted Photons and early showering pions in 200 PU

![CMS Phase-II Simulation Preliminary](image1.png)

Ratio of trackster energy before and after cleaning as a function of generated particle energy shows very little difference

![CMS Phase-II Simulation Preliminary](image2.png)

Ratio of trackster and generated particle energy for $60 < E_{\text{GEN}} < 80$ GeV. Cleaning removes tails in resolution
Putting it all together: Object Reconstruction

Demonstrating by example: Electromagnetic iteration
- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions
- Cleaning tracksters to get better estimates of properties

Unconverted Photons and early showering pions in 200 PU

Ratio of trackster energy before and after cleaning as a function of generated particle energy shows very little difference

Trackster direction estimates before and after cleaning shows improvement after cleaning
Putting it all together: Object Reconstruction

**Demonstrating by example: Electromagnetic iteration**
- Iteration aimed at extracting EM objects first
  - Relatively less complex than HAD objects
  - Useful for electron/photon reconstruction
- Keep all LCs in the CE-E part of the detector + first few CE-H layers to capture leakage
  - Mask LCs deeper inside the detector
- Connect LCs using Cellular Automaton to form tracksters
- State of the art Graph Neural Networks based particle ID to distinguish from early showering pions
- Cleaning tracksters to get better estimates of properties

Tracksters and LCs filtered by the PID and/or cleaning fed back to LC collection for further HAD/MIP iterations
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
  - Variety of pattern recognition algorithms can be plugged in and out
  - Different strategies for different particles
- **CLUE** is an “imaging” density based GPU friendly density-based algorithm
  - Provides building blocks for pattern recognition algorithms
  - Reduces hit multiplicity without sacrificing performance
  - Designed with parallelism in mind
- Variety of strategies being actively explored for best performance in **200 PU**
  - Optimal Particle Flow interpretation requires robust particle ID/ energy regression/ PCA
  - Utilise **novel machine learning** ideas like Graph Neural Networks
  - Strategy for purifying objects from PU contributions
- Next steps:
  - Improve strategies for hadron reconstruction and PF-objects interpretations
References

- DPS note TBA
Thank You