Towards end-to-end particle flow in high granularity detectors

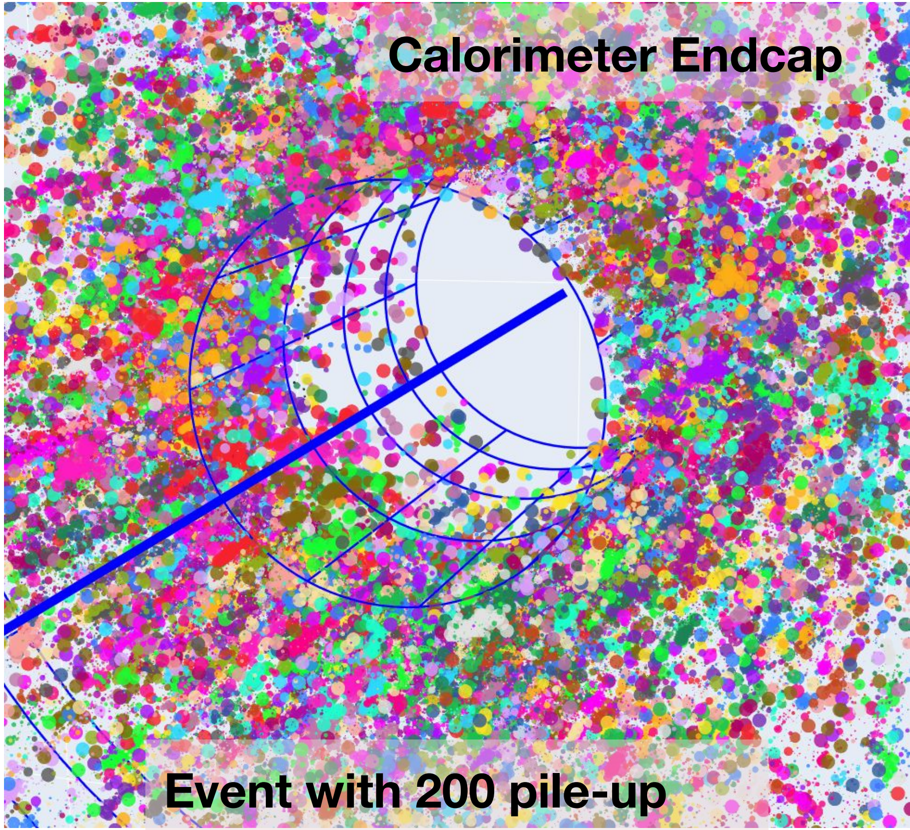
Philipp Zehetner[1,2] - Jan Kieseler[3] - Shah Rukh Qasim[4] - Dolores Garcia[2]

CHEP 2024
Offline Computing

[1] Ludwig Maximilian University of Munich
[2] CERN
[3] Karlsruhe Institute of Technology
[4] University of Zurich

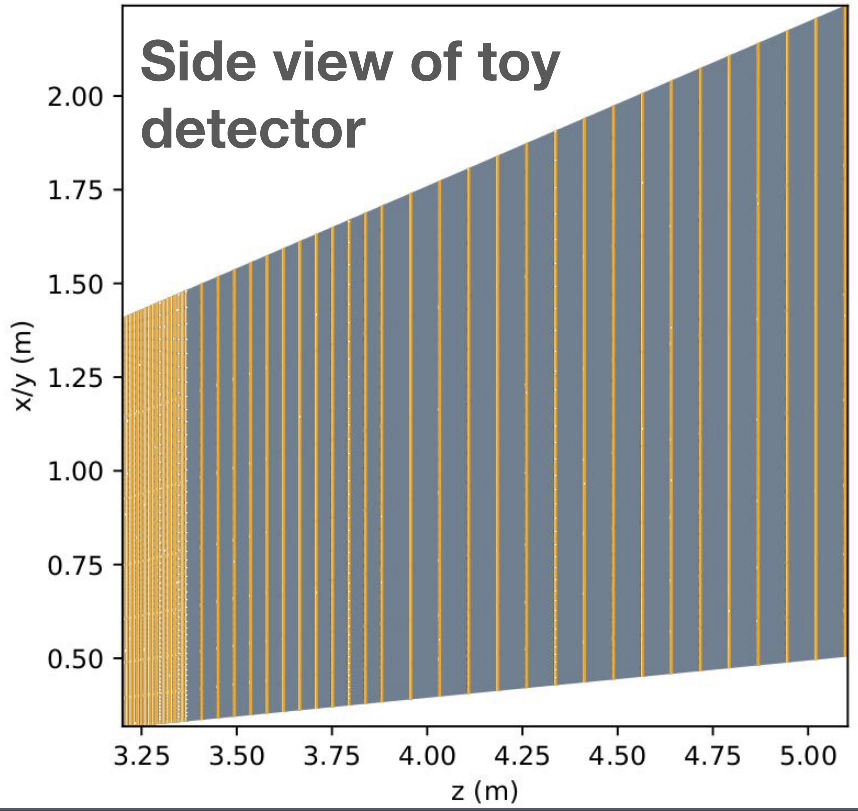
Overview

- New Highly Granular Calorimeters pose challenges due to their more complex events (~250k hits, ~5k showers / endcap)
- Track information can improve clustering performance in high pile-up environments
- Particle ID and correction factors for energies can be learnt on top of clustering
- “Precluster” algorithms speed up training time and inference time



Toy Detector

- Simulation of calorimeter endcap closely inspired by the CMS HGCAL
  - simpler geometry
  - slightly more readout channels
  - comparable complexity of events
- Covering $1.5 < \eta < 3.0$
- Silicon sensors rectangular in η and φ
- Tracks added in front of calorimeter
  - functionally like 'hits in front of calorimeter
  - include blurred true momentum
  - no fake tracks
- Standalone simulation with GEANT4



Previous Approach

Dynamic Graphs

GravNet Architecture

- Learn coordinates in latent **GravNet Space**
- Build edges to nearest neighbors
- Exchange information
- Significantly faster than comparable DGCNN

Object Condensation

Object Condensation Loss

- Learn coordinates and confidence in **Cluster Space**
- Build repulsive and attractive potentials scaled with confidence
- Minimize potentials

Challenges in Previous Approach

Dynamic Graphs

GravNet Architecture

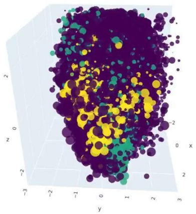
Clustering Loss

Object Condensation Loss

Large architectures leading to long training times (weeks)

Full 200 pile-up events too big to train (memory constraints)

Slow trainings only possible with incomplete data (30° slices of pile-up)



Presentation at Connecting The Dots

New Approach

Pre-clustering

"Snowflake"

Reduces input size by over 97% (factor ~30)

Dynamic Graphs

GravNet Architecture

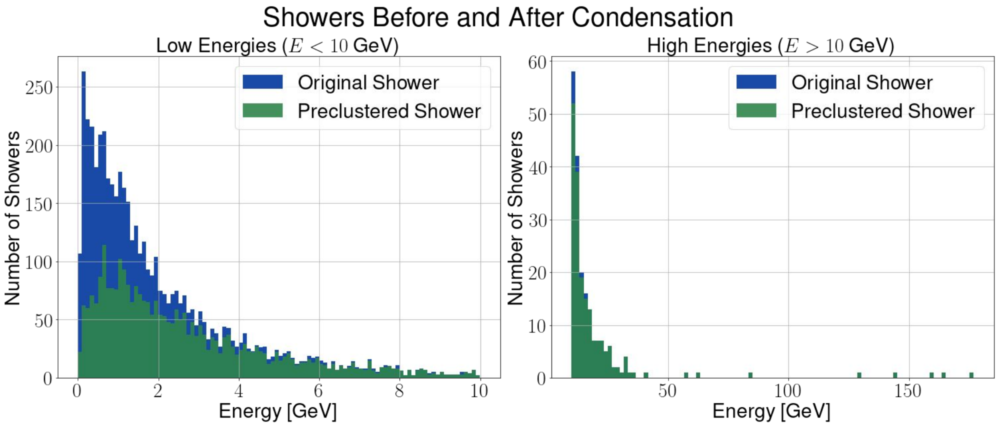
Clustering Loss

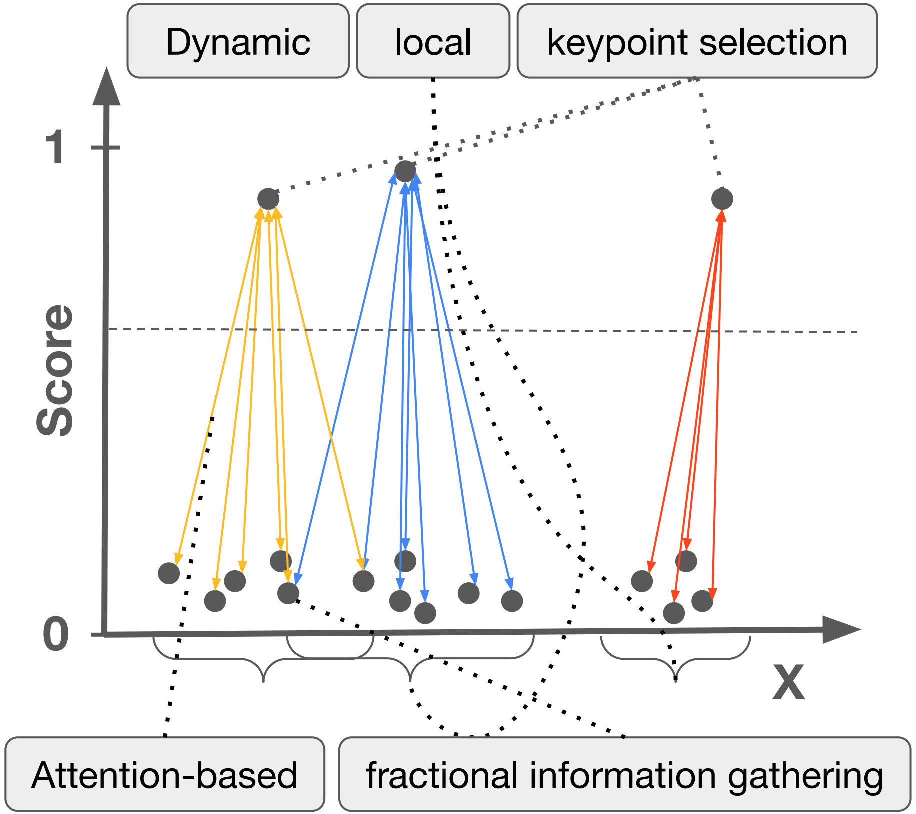
Object Condensation Loss

- fast turnaround in trainings > 7x
- faster inference times > 2x

- Training within days instead weeks
- Custom cuda kernels
- Inference time < 700 ms (A100)

Key Points:
- Reduce input data by condensing information into representative points
- Keep at least one point for every shower with E > 5 GeV
- Reduction from ~270k hits to 7k hits!

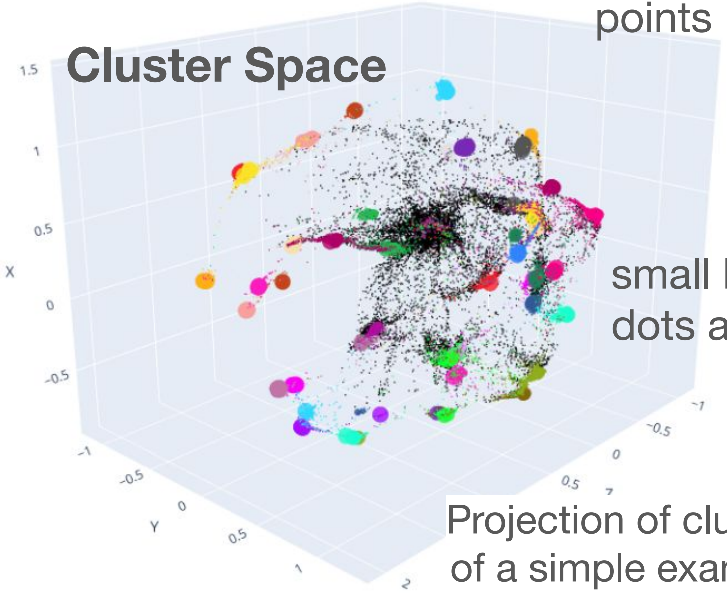




Creating and Matching Showers

Large colorful points are **condensation points**

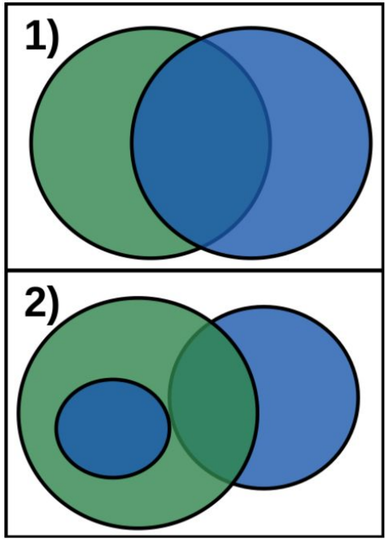
Showers are defined by **spheres** around condensation points



small black dots are **noise**

Projection of cluster space of a simple example event

Matching is performed based on showers' **intersection over union** (IoU) with a preference to matching on tracks



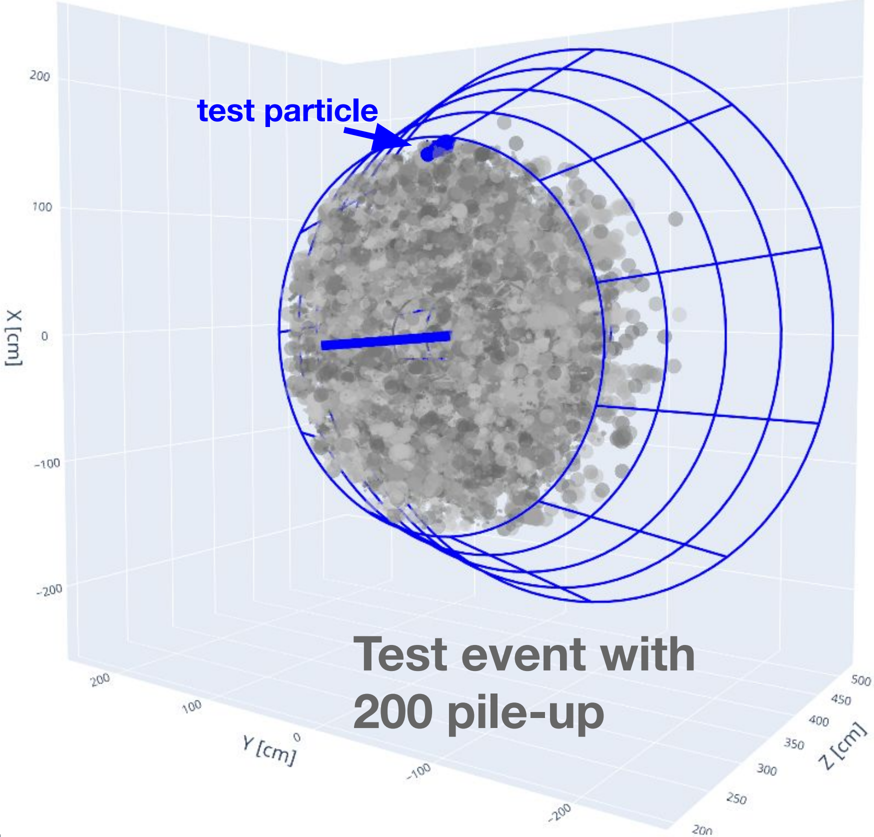
1. Showers matched if **IoU** > threshold
2. Two bad candidates for matching. Shower might be matched if it shares a track

Strict matching criteria lead to better resolution at the cost of efficiency

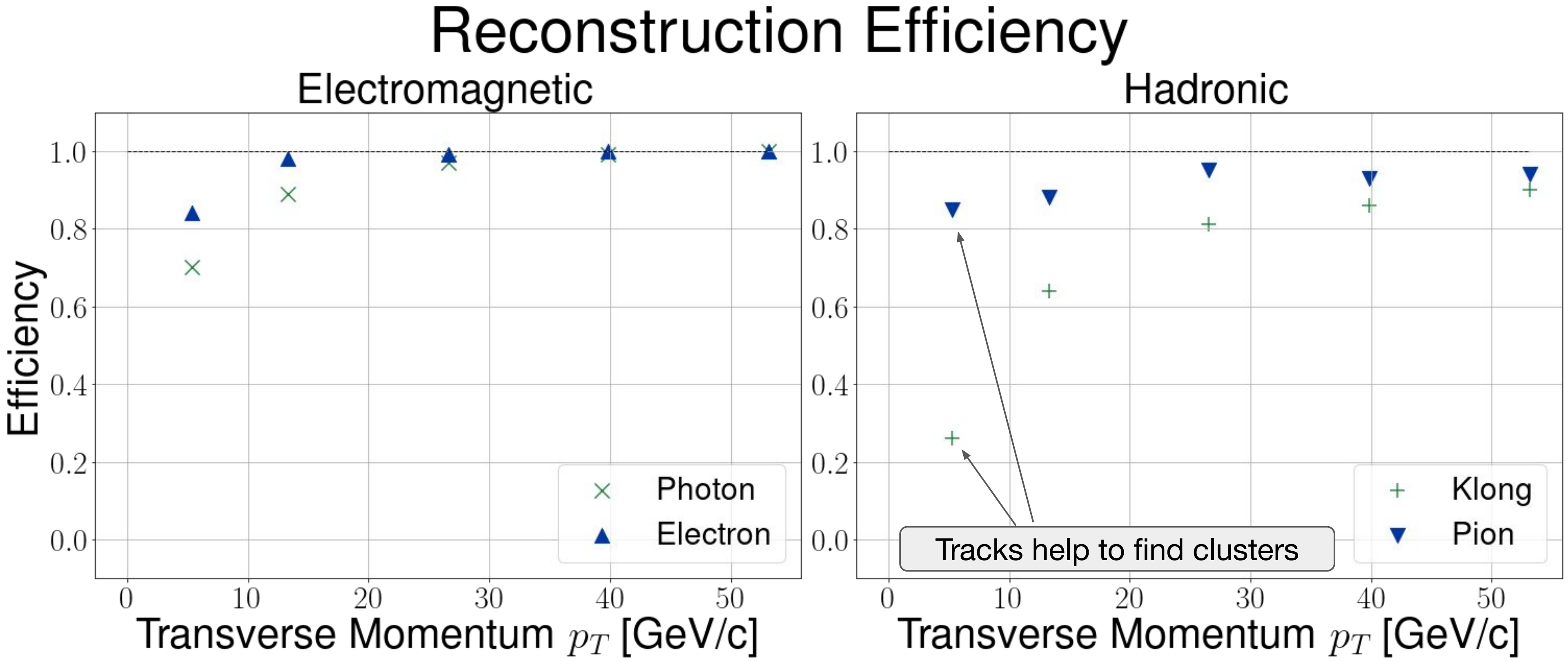
Test Data

Test data sets contain

- Single Photon, Electron, Pion, or K-long
- Random azimuthal direction at $\eta = 2.0$
- Monochromatic energies
  - E = 20 GeV $p_T = 5.3$ GeV/c
  - E = 50 GeV $p_T = 13.3$ GeV/c
  - E = 100 GeV $p_T = 26.6$ GeV/c
  - E = 150 GeV $p_T = 39.9$ GeV/c
  - E = 200 GeV $p_T = 53.2$ GeV/c
- Overlay of 200 minimum bias proton-proton collisions at 14 TeV, simulated with PYTHIA
- Gaussian noise representing detector noise



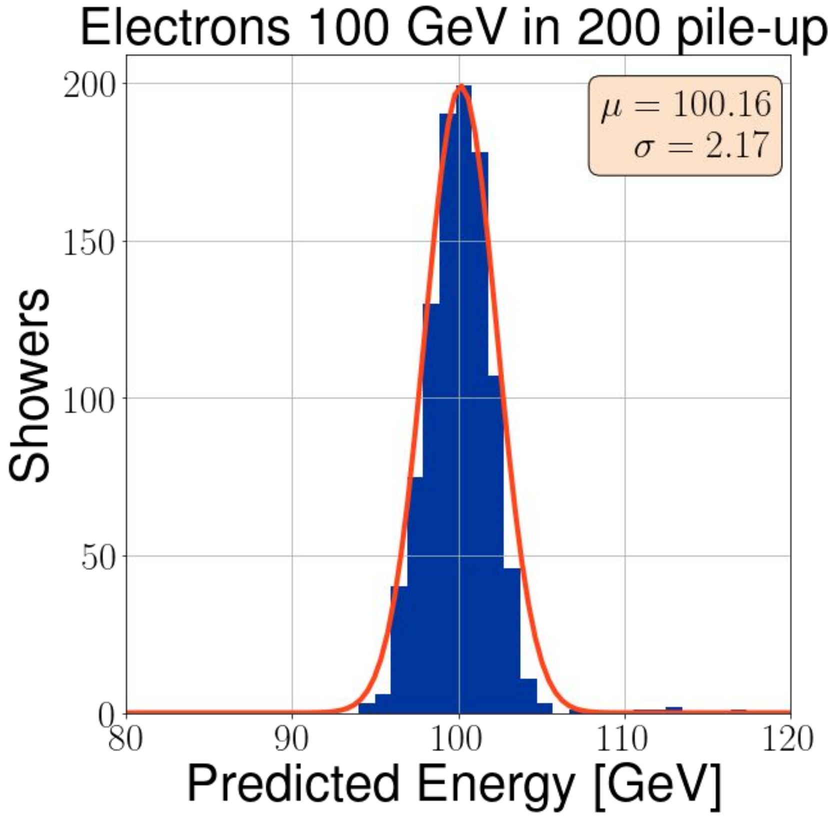
Efficiency


Reconstruction Efficiency
Electromagnetic
Hadronic
Efficiency
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Transverse Momentum p_T [GeV/c]
Photon
Electron
Klong
Pion
Tracks help to find clusters

Definition of Response & Resolution

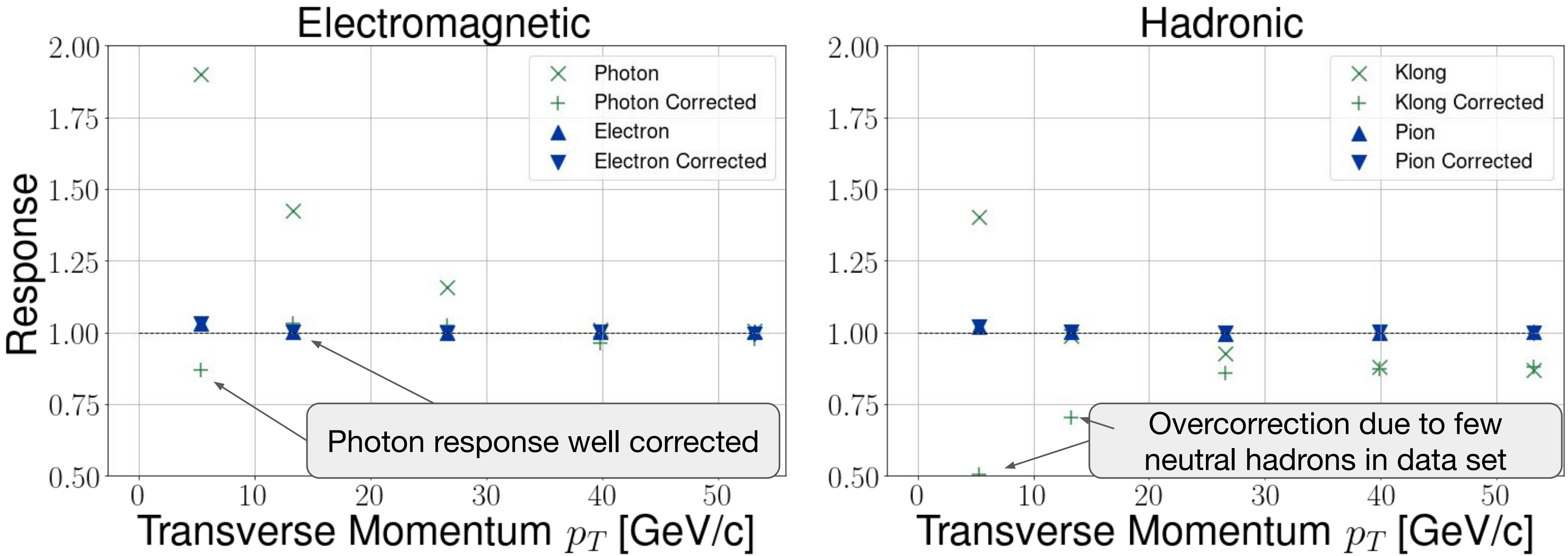
Response: $$\frac{\mu(E_{\text{pred}})}{E_{\text{true}}}$$

Resolution: $$\frac{\sigma(E_{\text{pred}})}{\mu(E_{\text{pred}})}$$



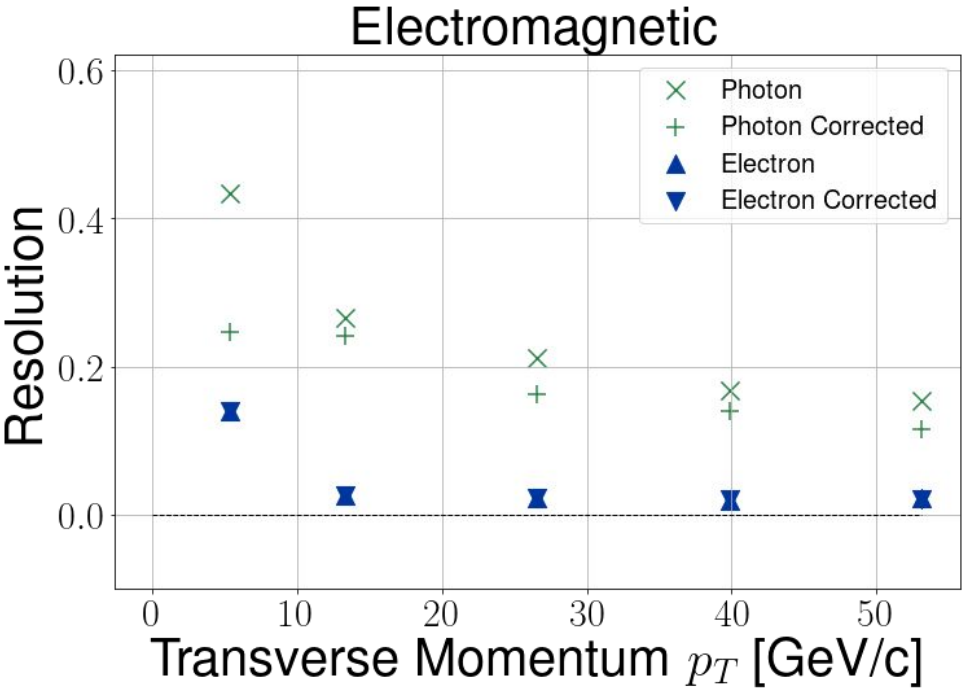
Response With Correction

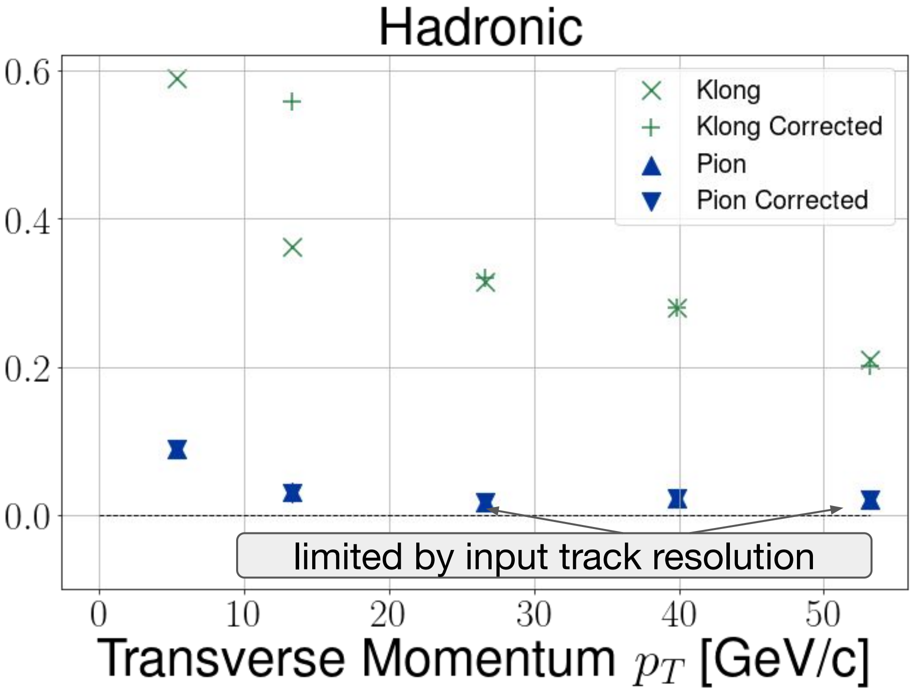
$$\text{Response} = \frac{\langle E \rangle}{E_{true}}$$



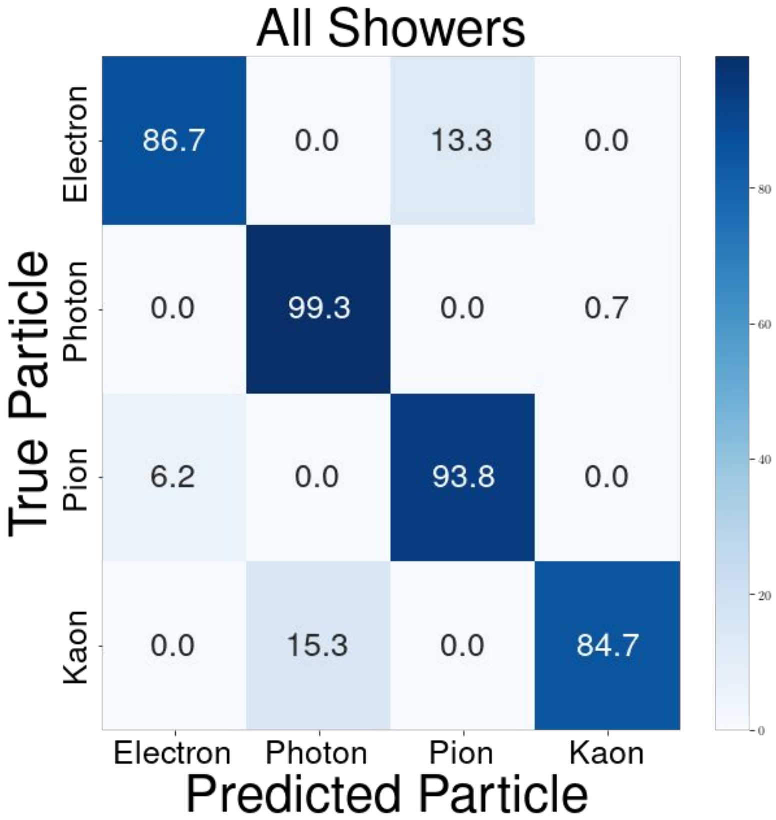
Resolution With Correction

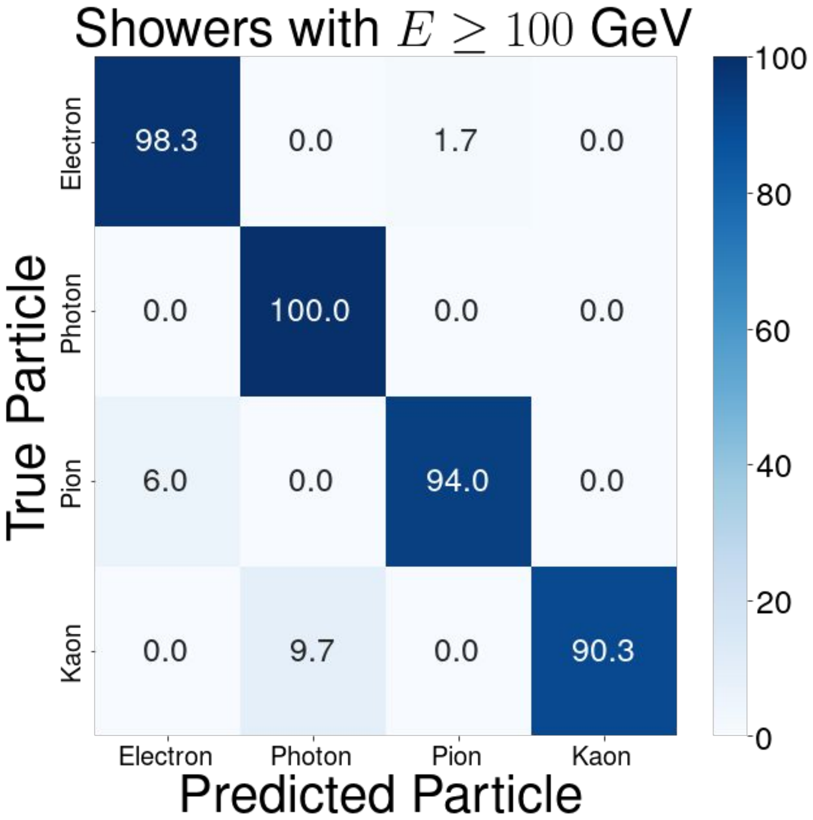
$$\text{Resolution} = \frac{\sigma(E)}{\langle E \rangle}$$





Particle ID


All Showers
True Particle
Electron
Photon
Pion
Kaon
86.7
0.0
13.3
0.0
0.0
99.3
0.0
0.7
6.2
0.0
93.8
0.0
0.0
15.3
0.0
84.7
Electron
Photon
Pion
Kaon
Predicted Particle
0
20
40
60
80


Showers with $E \geq 100$ GeV
True Particle
Electron
Photon
Pion
Kaon
98.3
0.0
1.7
0.0
0.0
100.0
0.0
0.0
6.0
0.0
94.0
0.0
0.0
9.7
0.0
90.3
Electron
Photon
Pion
Kaon
Predicted Particle
0
20
40
60
80
100

Summary

- Significant conceptual update of the end-to-end reconstruction algorithm for highly granular calorimeters

- This is the first application of the Snowflake preclustering model [publication in prep.]
    - allows much faster turn-around time while preserving necessary information
    - reducing input data dimensionality by a factor of 30
    - reduce training time by a factor of 7
    - more than double inference speed

- For the first time training with high-density full 200 PU events is possible

- For the first time, we perform a combined clustering and particle flow in high-density environments
    - Tracks significantly improve energy resolution
    - Neutrals approximately remain as before (not shown today)
    - Accurately predict particle ID

Acknowledgments

Thank you for you attention!

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

"This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA)"

We thank Ian Fisk and the Flatiron Institute for their support with access to their GPU cluster.

philipp.zehetner@cern.ch