

# Reconstruction with ML

**Dolores Garcia**, Michele Selvaggi, Jan Kieseler, Philipp Zehetner, Gregor Krzmanc, Brieuc Francois

# ML for very adaptive reconstruction

Adaptable reconstruction is crucial for a systematic design optimization:

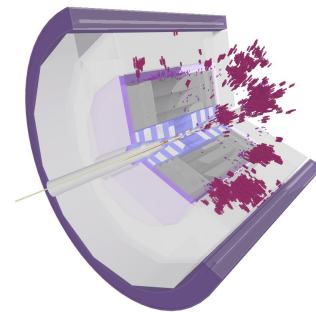
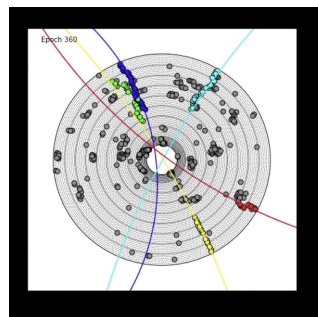
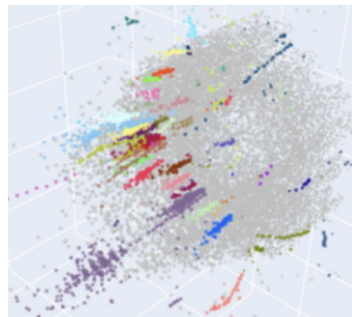
- No need for hand picked parameter tuning
- Costly conventional implementation
- Increased performance?

Problems that are similar:

- Calorimeters
- Tracking
- Combining information from multiple subdetectors
- Images of cars

Adaptive reconstruction algorithms are working: Belle II, HGCAL, CERN CMS

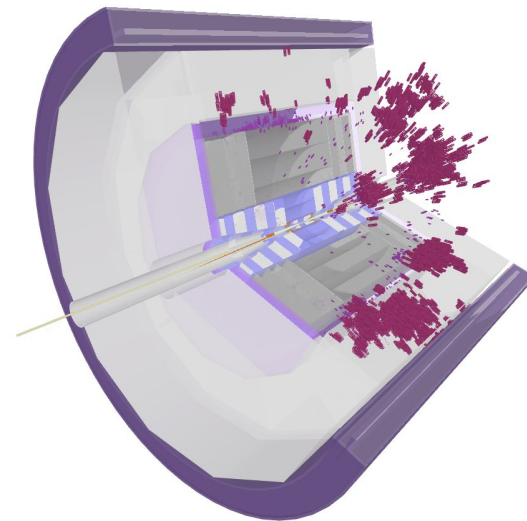
<sup>1</sup> ACKS : Michele Selvaggi, Gregor Krzmarc, Jan Kieseler, Philipp Zehetner



# ML for Particle Flow, a very adaptive reconstruction

The particle flow algorithm aims to **identify the produced particles in a collision through the combination of the information from the entire detector and provide best combined energy/momentum resolution**

- Hoping to achieve higher reconstruction performance: cluster merging, arbitration of track vs cluster energy
- Our approach is very adaptive to detector geometry
- First step: **focus on calorimeter clustering**



**A** Example of input data in the CLD detector

<sup>1</sup> ACKS : Michele Selvaggi, Gregor Krzmarc, Jan Kieseler, Philipp Zehetner

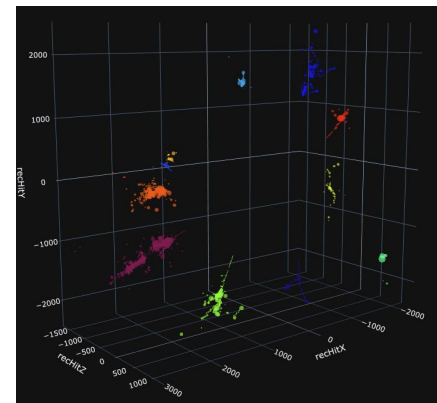
[1] Pata, J. Machine learning for particle flow reconstruction at CMS, presentation at CDS.

[2] Qasim, S. R., Chernyavskaya, N., Kieseler, J., Long, K., Viazlo, O., Pierini, M., & Nawaz, R. (2022). End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks. *The European Physical Journal C*, 82(8), 753.

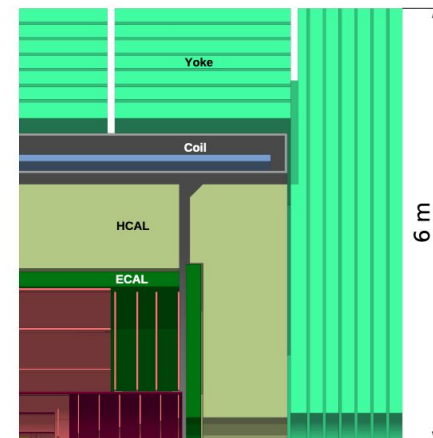
# Dataset

- Event generation:
  - Particle gun (10-15 particles)  $\rightarrow$  O(7000) hits
  - $E \in [0.5, 50]$  GeV
  - $p, n, K_L, \pi, e^{\pm}, \gamma$
  - FullSim CLD
  - Truth from gen [CalohitMCTruthLink]
- Training on 50k events (small dataset)

**A** Example of event from training set



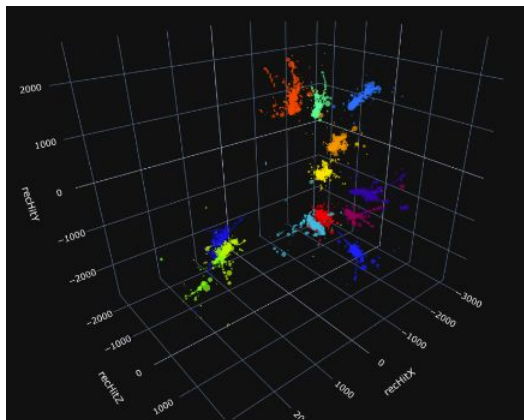
**B** Vertical cross section CLD [1]



<sup>1</sup> ACKS : Michele Selvaggi, Gregor Krzmann, Jan Kieseler, Philipp Zehetner

[1] Bacchetta, Nicola, et al. "CLD--A Detector Concept for the FCC-ee." *arXiv preprint arXiv:1911.12230* (2019).

# End-to-end approach



Input:

- A set of hits from different sensors (coordinates, type of hit, energy, A)
- Each one node in the graph O(600) per particle

CP + GNN

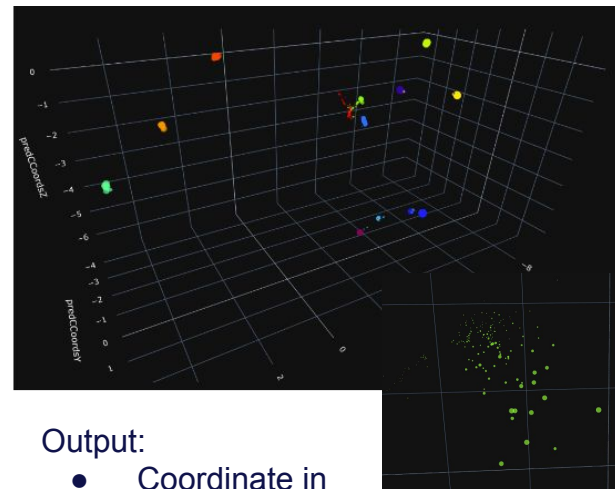
$$q_{\alpha k} = \max_i q_i M_{ik}.$$

$$\check{V}_k(x) = \|x - x_\alpha\|^2 q_{\alpha k}, \text{ and}$$

$$\hat{V}_k(x) = \max(0, 1 - \|x - x_\alpha\|) q_{\alpha k}.$$

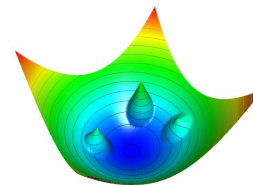
$$L_V = \frac{1}{N} \sum_{j=1}^N q_j \sum_{k=1}^K \left( M_{jk} \check{V}_k(x_j) + (1 - M_{jk}) \hat{V}_k(x_j) \right).$$

- **Each object 1 condensation point (CP)**
- **Repulsive + Attractive potentials for each CP**



Output:

- Coordinate in embedding space (3D>)
- Beta (q)
- Use clustering space to build showers

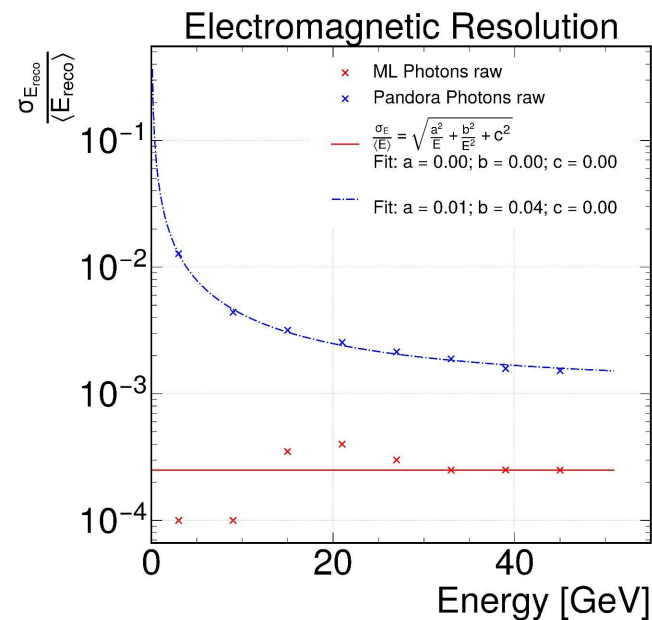
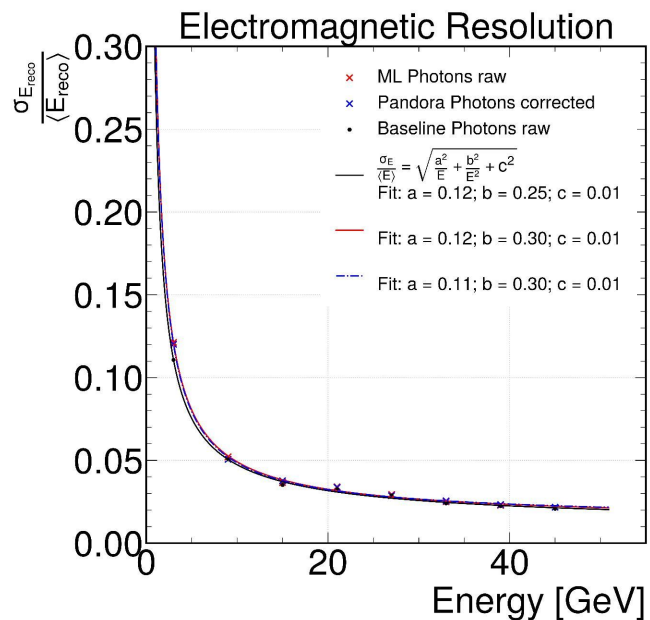


<sup>1</sup> ACKS : Michele Selvaggi, Gregor Krzmarc, Jan Kieseler, Philipp Zehetner

[1] Kieseler, J. (2020). Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph, and image data. *The European Physical Journal C*, 80, 1-12.

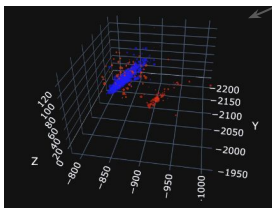
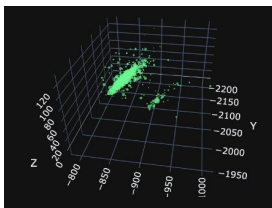
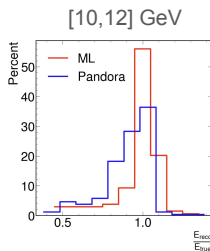
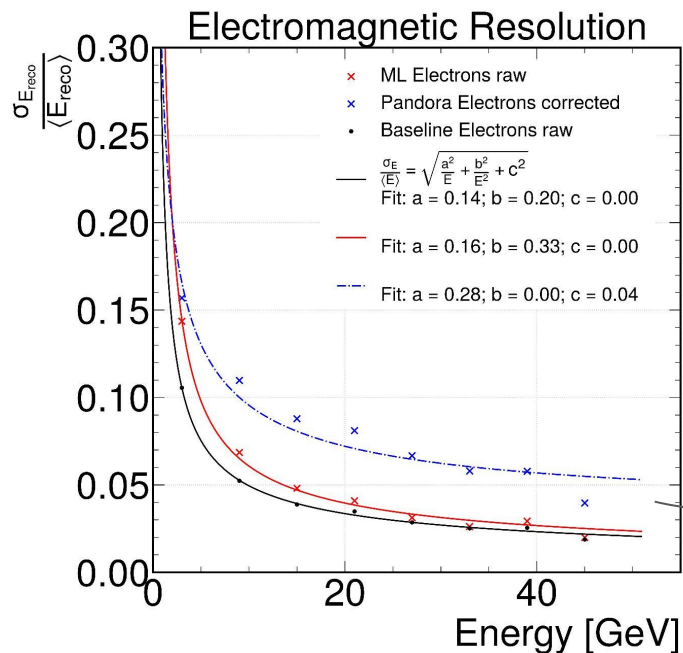
[2] Qasim, Shah Rukh, et al. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks." *The European Physical Journal C* 79.7 (2019): 1-11.

# Results:

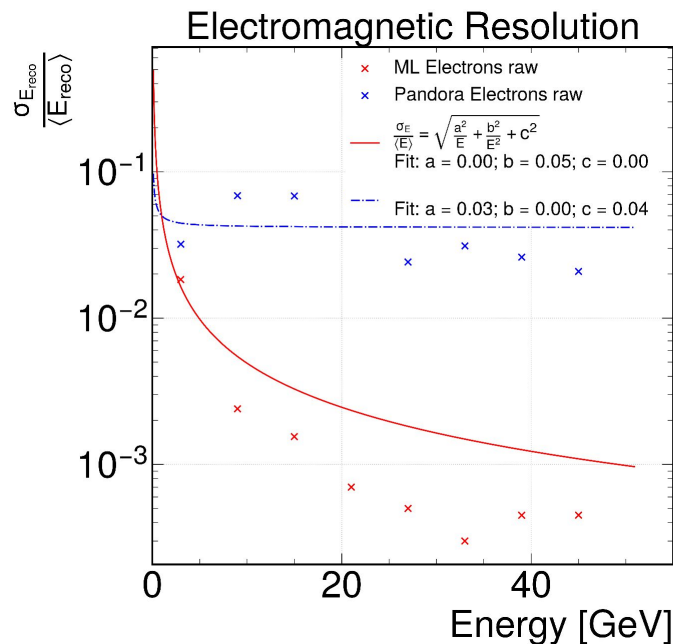


- The ML model has no energy correction applied
- Baseline is the sum of the hits energy

# Results:

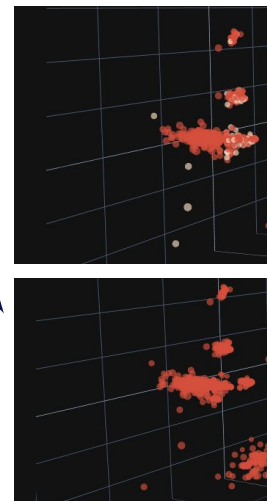
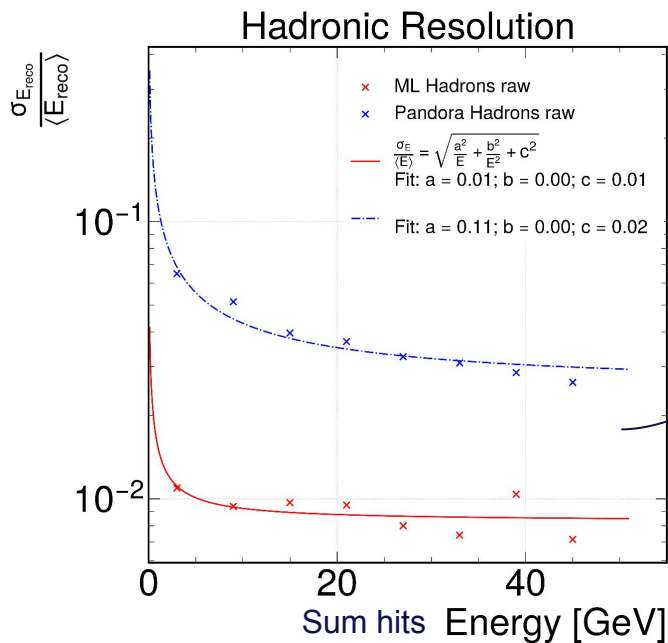
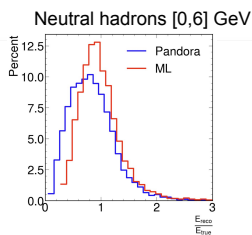
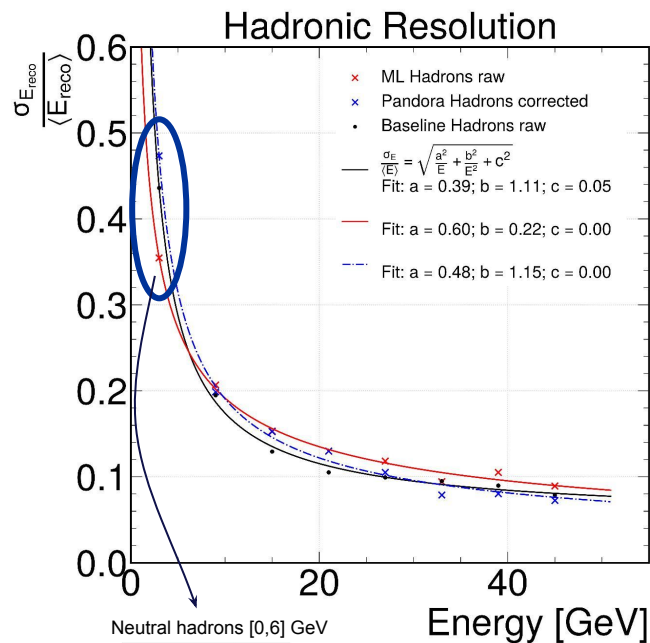


adds to the confusion term,  
as track energy is the sum of  
both



- The ML model has no energy correction applied
- Baseline is the sum of the hits energy

# Results:

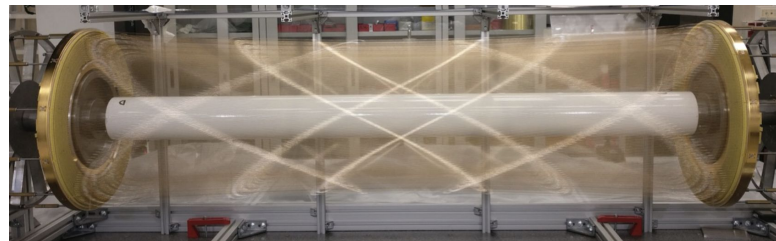


- The ML model has no energy correction applied



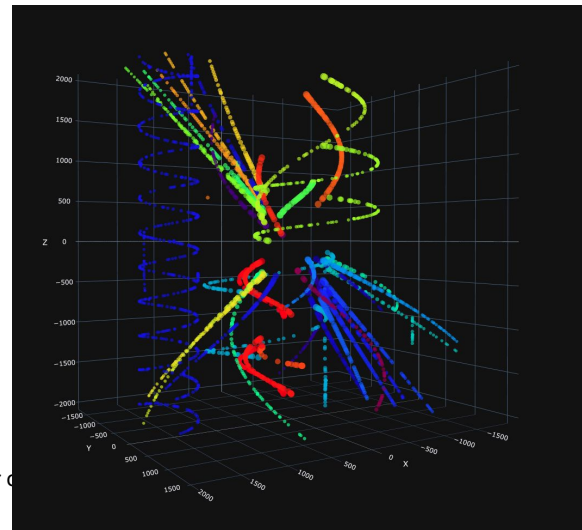
# Pattern recognition for the IDEA Drift Chamber

- Investigating Machine Learning based tracking for the IDEA detector (in addition to the conventional tracking)
- Simulating events with Pythia + ddsim + digi
- Dataset inputs of DCH:
  - Wire geometry:
    - Layer, superlayer
    - Stereo angle
    - → Coordinates
  - Hit:
    - Distance along the wire
    - Distance to the wire
- Hits from the vertex detector
- Exploiting GNN/ Point Cloud architectures



Top: Drift Chamber [1]

Bottom: example event from full sim

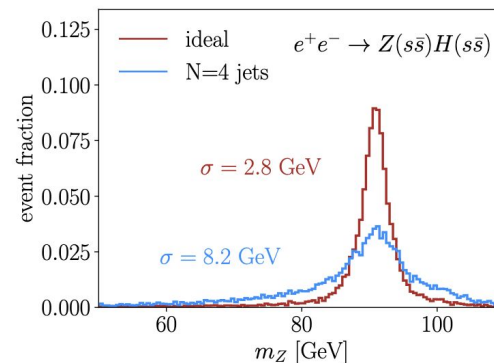
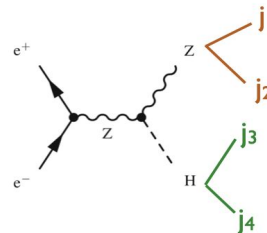


<sup>1</sup> ACKS : Michele Selvaggi, Brieuc Francois

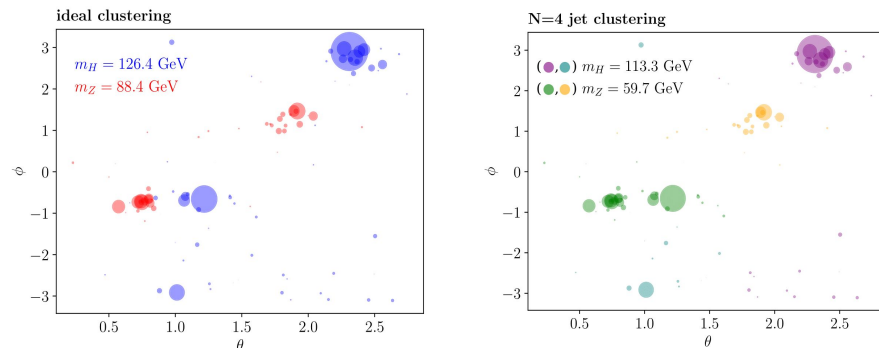
[1] Tassielli, G. F., Baldini, A. M., Cavoto, G., Cei, F., Chiappini, M., Chiarello, G., ... & Voena, C. (2020). The drift chamber of the IDEA detector. *Journal of Instrumentation*, 15(09), C09051.

# Clustering Color Singlets

- FCC-ee would serve as a Higgs factory, electroweak and top at highest luminosities
  - Measure Higgs particle properties and interactions in challenging decay modes
- Identification of color-neutral resonances relies on clustering final state into jets
- Calorimetry is expected to be much improved at future  $e^+e^-$  colliders, so that the 2-jet invariant mass resolution will be dominated not by detector resolution but rather by mis-clustering [1] (A)
- Jets are not well defined but color connection is physical, this may help **improve the mass estimation for color singlets (H,Z,W) and remove more background**



**A** Comparison of clustering performance vs ideal reconstruction



**B** Example of miss clustering

<sup>1</sup> ACKS : Michele Selvaggi

[1] Fujii, K., Grojean, C., Peskin, M. E., Barklow, T., Gao, Y., Kanemura, S., ... & Murayama, H. (2020). ILC study questions for snowmass 2021. *arXiv preprint arXiv:2007.03650*.

[2] Gallicchio, J., & Schwartz, M. D. (2010). Seeing in color: jet superstructure. *Physical review letters*, 105(2), 022001.

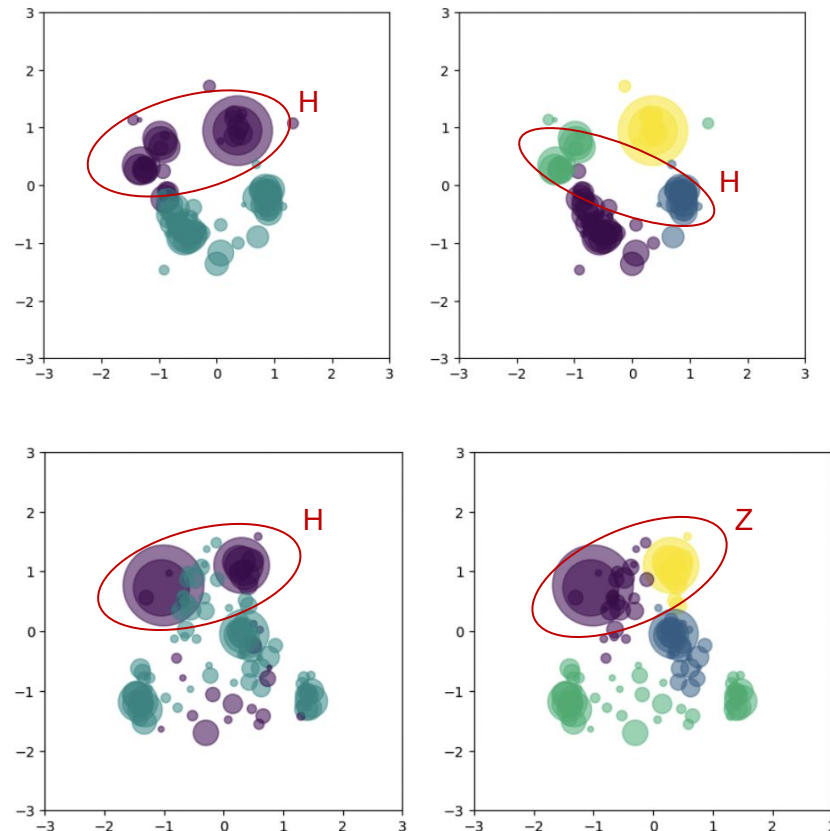
# Clustering Color Singlets

Loss in performance can be due to:

- Miss matching of jets pairs
- Miss clustering of soft particles leading to degraded resolution

Possible solutions:

- Parameter tuning (BAO)
- Optimize distance metrics?: piecewise continuous function, hard optimization problem
- **End-to-end approach**



**A** Mismatching of jets pairs

<sup>1</sup> ACKS : Michele Selvaggi

[1] Fujii, K., Grojean, C., Peskin, M. E., Barklow, T., Gao, Y., Kanemura, S., ... & Murayama, H. (2020). ILC study questions for snowmass 2021. *arXiv preprint arXiv:2007.03650*.

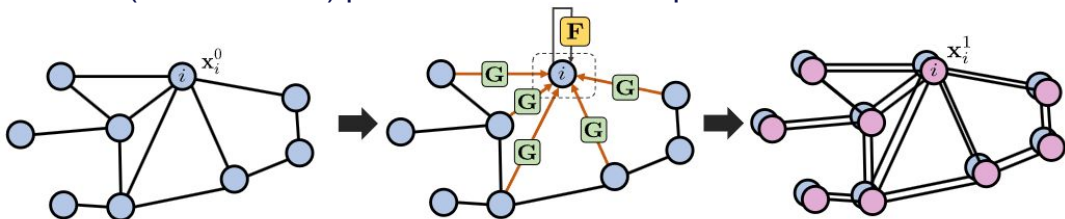
[2] Gallicchio, J., & Schwartz, M. D. (2010). Seeing in color: jet superstructure. *Physical review letters*, 105(2), 022001.

# Clustering Color Singlets

- **GNN** - Node classification (instantiation) problem, permutation invariant and equivariant
- Arch: FC - Graph **Transformer** [1]
- **Results:**
  - Similar performance to classical approach
  - **Baselines:**
    - **Chi-squared**  

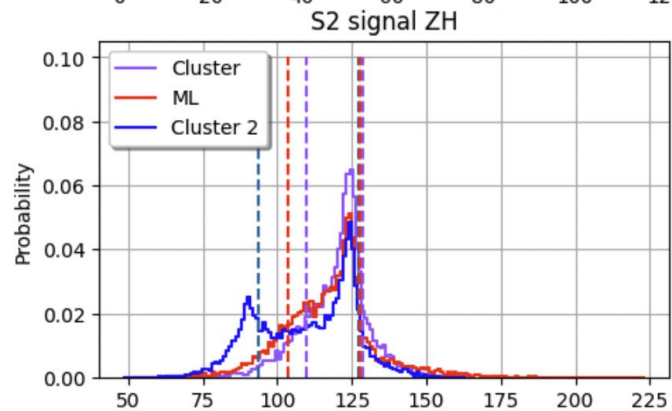
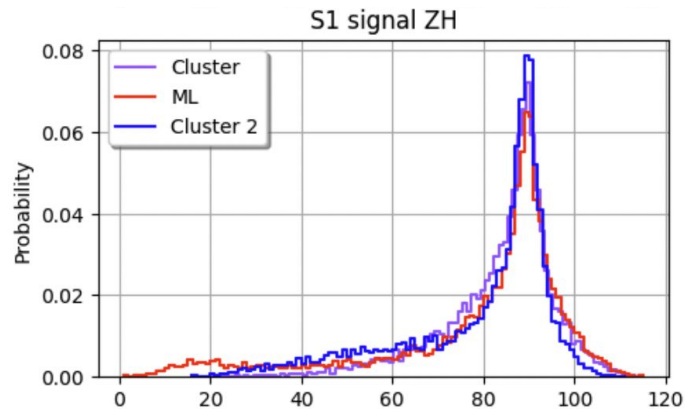
$$\chi^2 = 1/\sigma_H (M_{1/2} - M_H)^2 + 1/\sigma_Z (M_{2/1} - M_Z)^2$$
    - **Z only**  

$$\chi^2 = (M_{1/2} - M_Z)^2$$
- Wiring is important, using information about the ordering (tree structure) performance can be improved



<sup>1</sup> ACKS : Michele Selvaggi

[1] Dwivedi, V. P., & Bresson, X. (2020). A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*.



A. Mass distributions of signal

# Summary and next steps

## MLPF

MLPF allows for a detector agnostic calorimeter clustering with similar performance to Pandora

Next steps:

- Tackling PF (adding tracks and energy correction)
- Evaluation on more complex datasets with physics events and jet metrics

## IDEA wire chamber tracking

Created dataset, pipeline and started first trainings

Next steps:

- Link the geometry to access wire coordinates
- Improve architecture and study performance
- Add parameter estimation for TrackStates

## Clustering

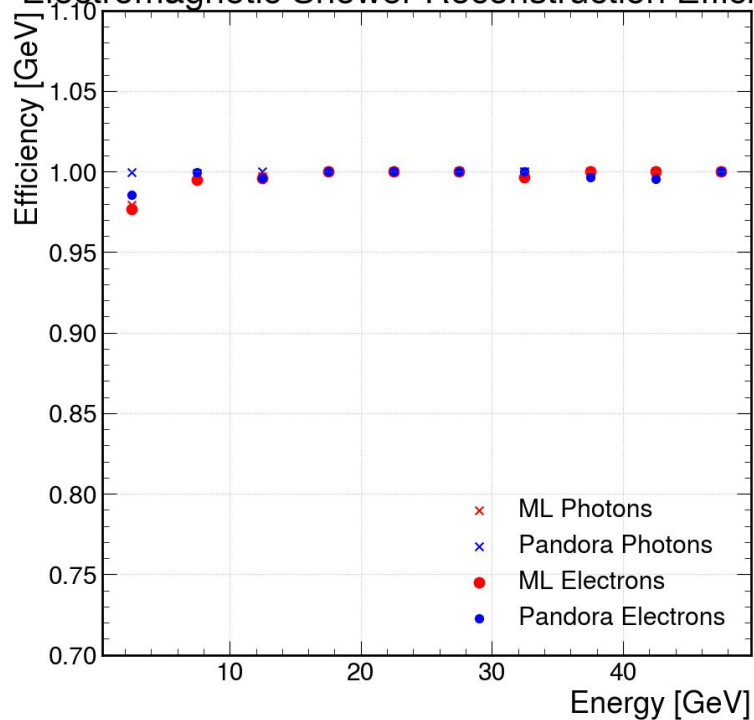
Color singlet clustering seems promising with ML but more research is necessary

Thanks to the Key4hep team!



Thank you

# Electromagnetic Shower Reconstruction Efficiency



# Hadronic Shower Reconstruction Efficiency

