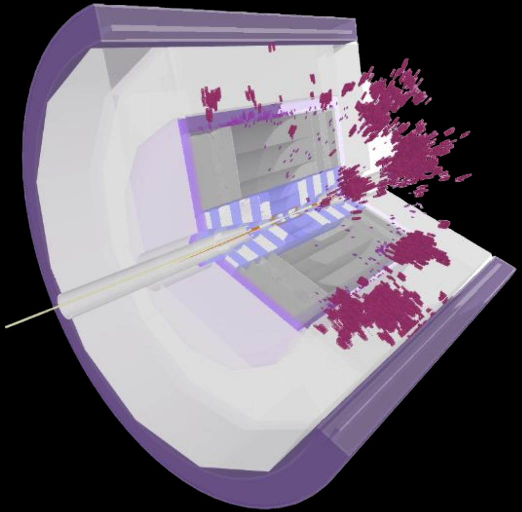
ML-based Particle Flow for CLD

Gregor Krzmanc, Dolores Garcia, Michele Selvaggi

3rd ECFA Workshop, Paris, 9 October 2024

ML-based Particle Flow

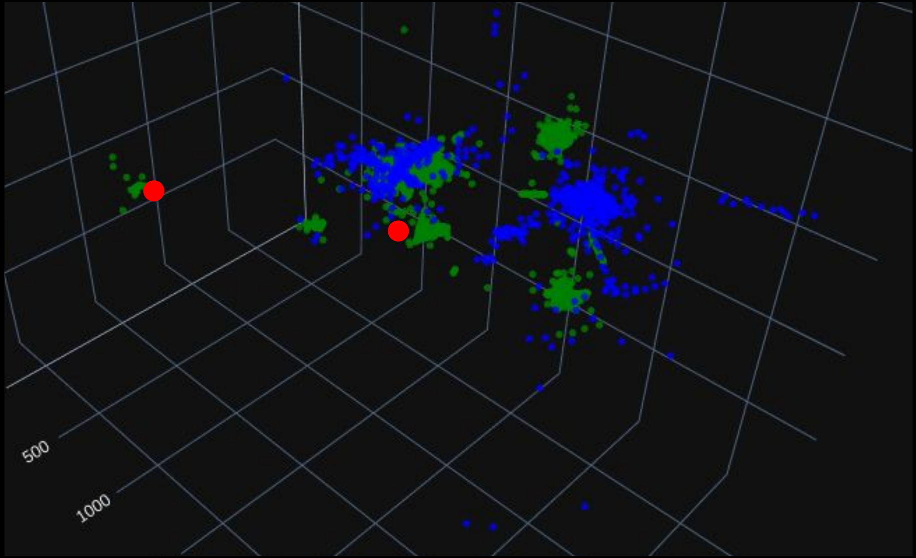


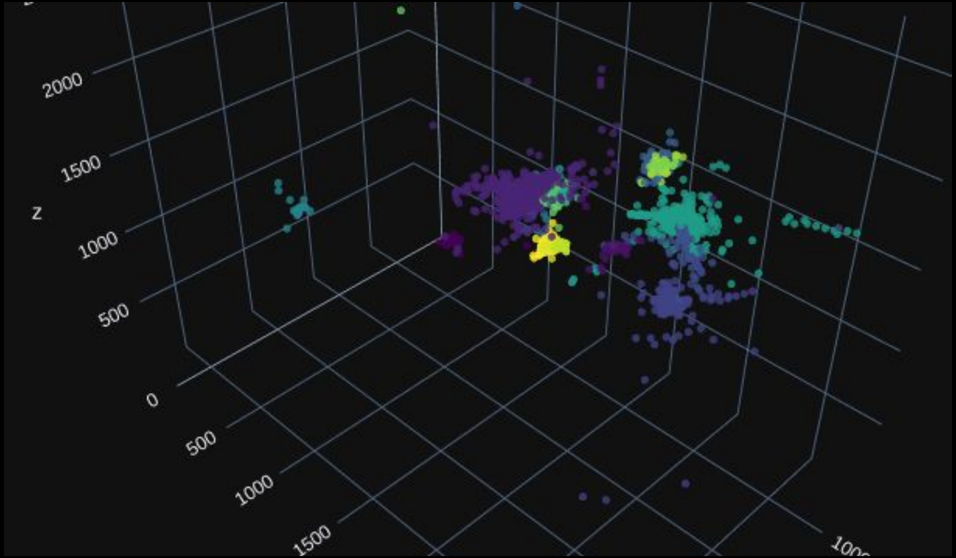
- Build PF candidates out of hits and tracks
- Interesting problem for ML: Variable number of inputs (hits, tracks) and outputs (PF candidates)
- Similar problems: image segmentation, tracking…

- Dolores' talk tomorrow: **End-to-end ML-based reconstruction for FCC-ee**

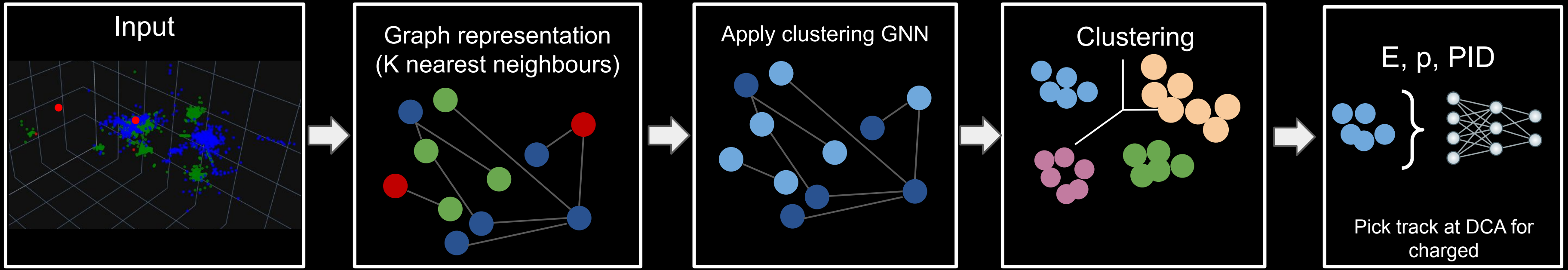
Dataset - CLD fullsim

fitted track
ECAL hit
HCAL hit



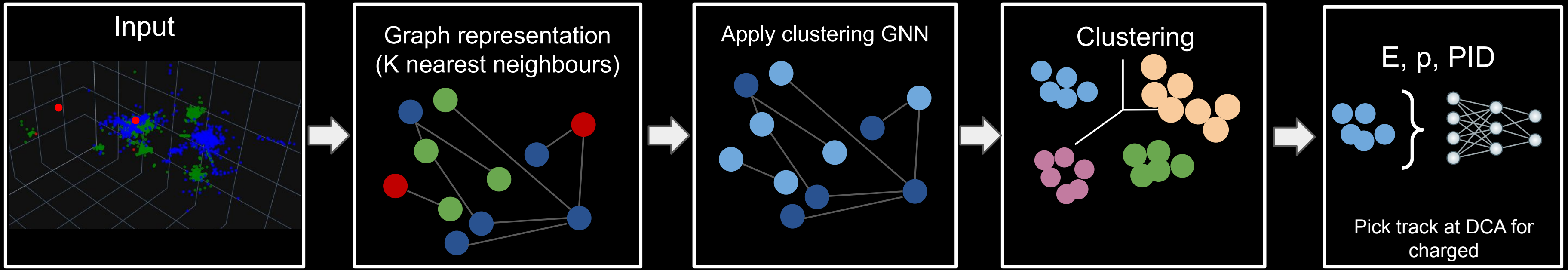


Model



- Input: set of hits (ECAL, HCAL) + fitted tracks
- Graph Transformer
- Use HDBscan to build clusters - PF Candidates
- Simple PID: photon/CH/NH/e
- Output: PFCandidates

Model

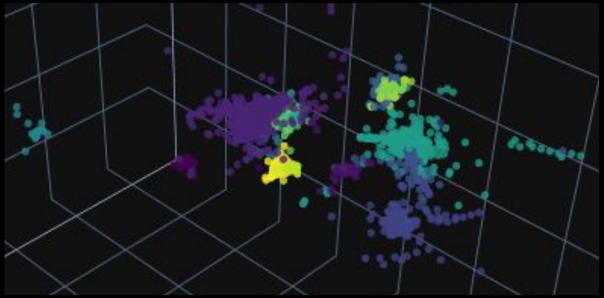


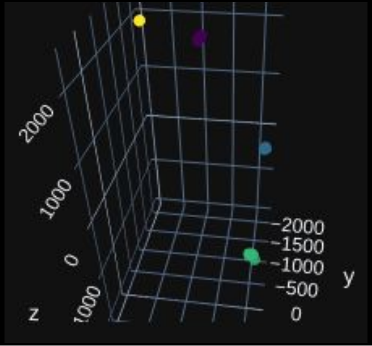
- Input: set of hits (ECAL, HCAL) + fitted tracks
- Graph Transformer
- Use HDBscan to build clusters - PF Candidates
- Simple PID: photon/CH/NH/e
- Output: PFCandidates

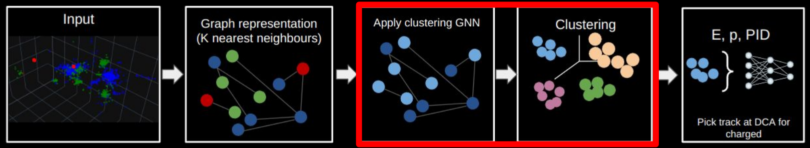
Similar approach for each detector! Can we avoid tuning of the Pandora parameters for each detector?

Datasets

10-15 particle gun	$K_S \to \pi^0 \pi^0 \to 4\gamma$
E ~ [0.5, 50] GeV, from (0,0,0) $\Delta r = 0.5$ p, n, K_L, $\pi^\pm$, $e^\pm$, γ	E ~ [0.5, 50] GeV Displaced vertex; $c\tau$ ~ 2.6 cm

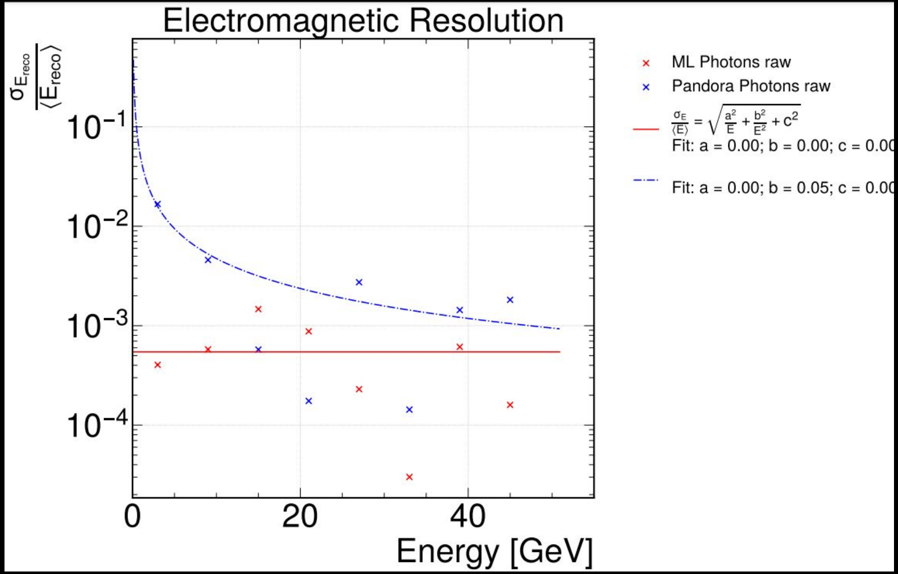


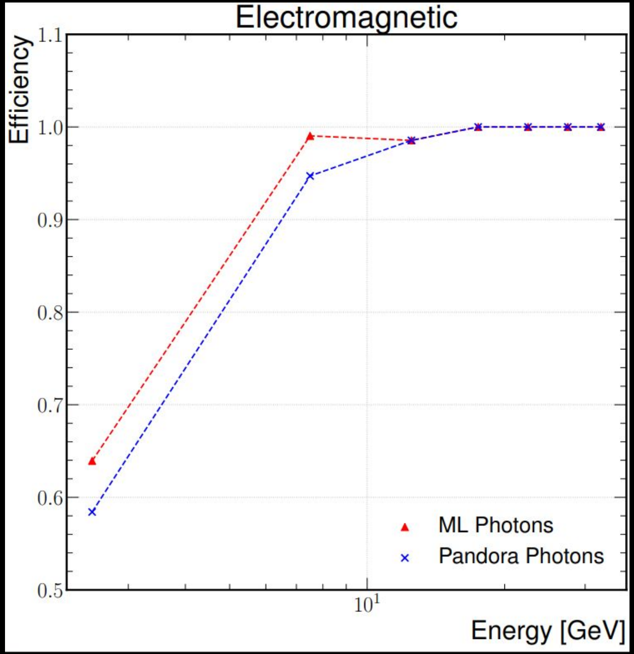


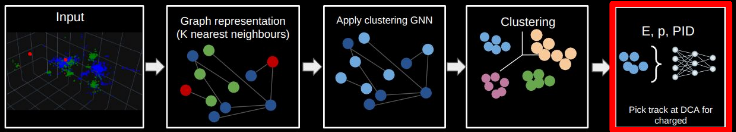


Results: clustering

- Improved efficiency of clustering o the 10-15 particles dataset

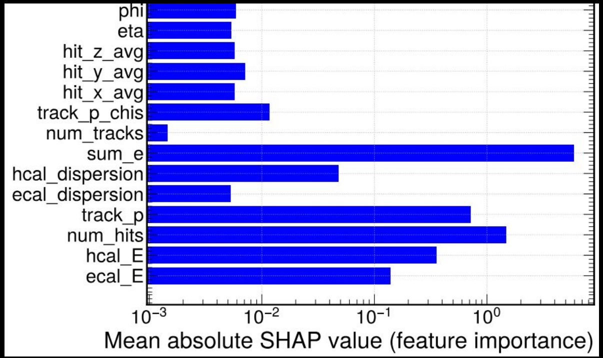




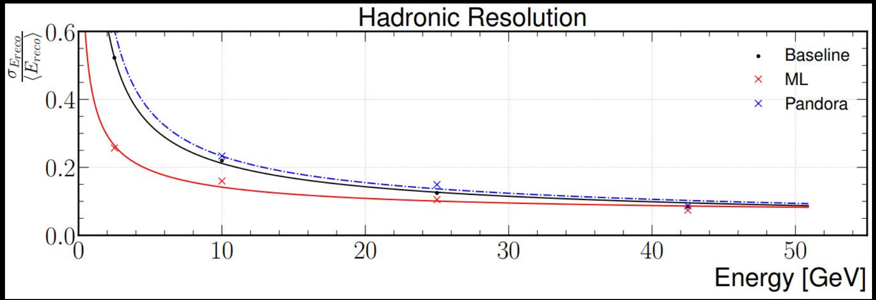


Results: energy correction

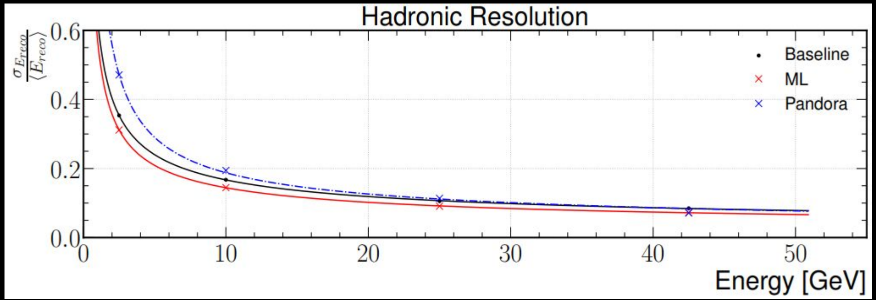
- Better clustering, energy correction for neutral hadrons leads to better mass resolution
- Key: high-level features (% energy in ECAL/HCAL; sum of E_{hits}) complementing the complex shower geometry information extracted by the GNN
- Better clustering is key - sum of the hits the most important feature



neutrons

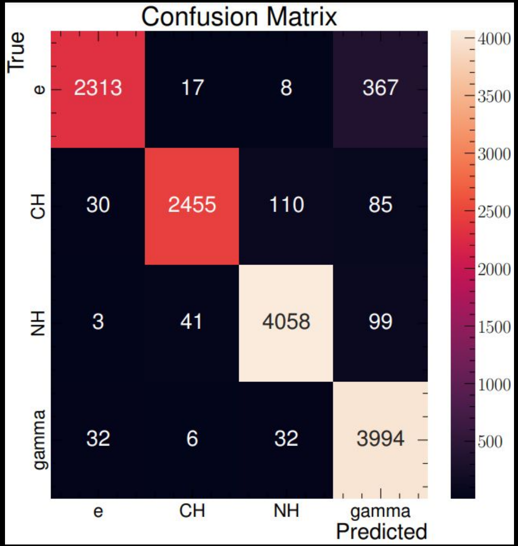


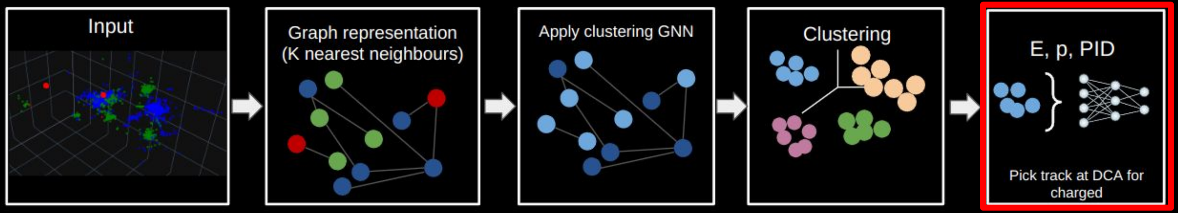
K_L



Results: PID

- Simplified PID: electron, CH (assign pion mass), NH (assign neutron mass), photon
- Predict unit vector p/|p| + energy correction + PID
- Separate model heads for charged (containing a track) and neutral (containing no track) particles
- All classes in both heads (to accounts for errors in clustering)

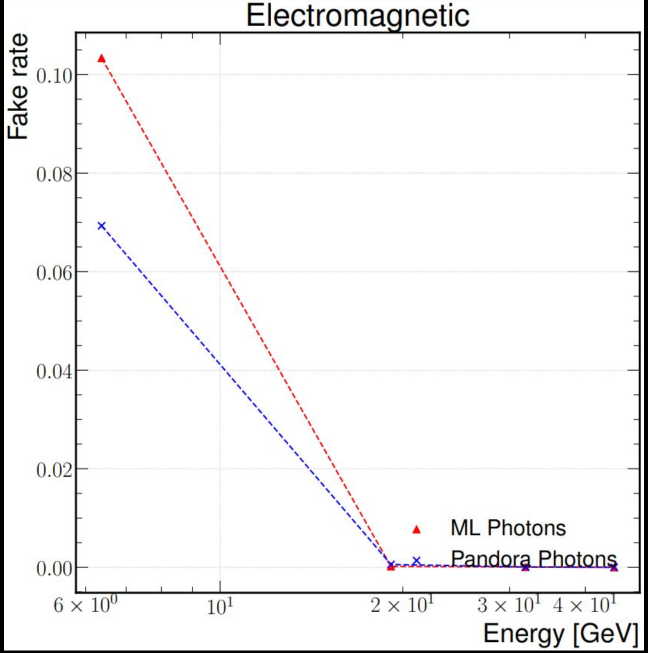


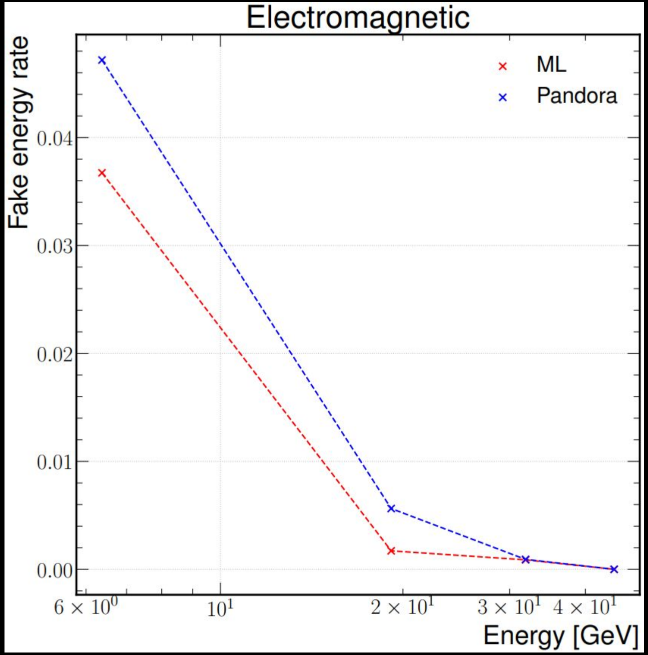


What about fake rate?

- Lots of fakes!
- But: less energy stored in them

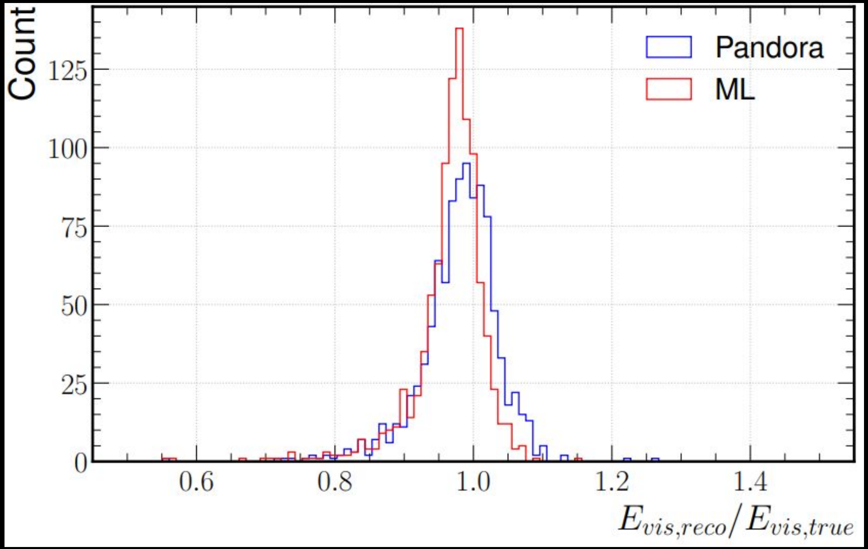
Fake energy rate = E of fakes / Total E

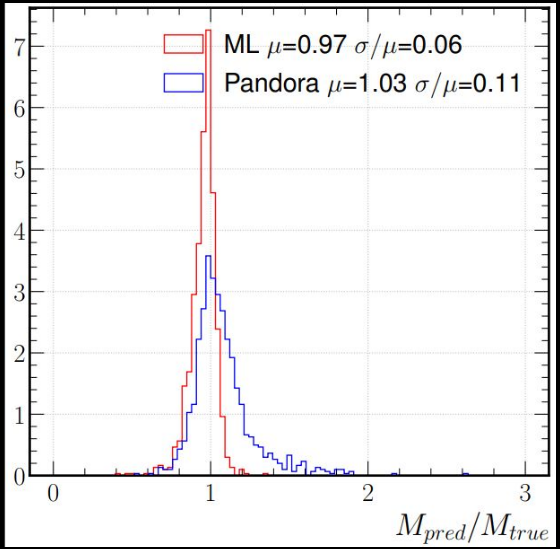




Results - 10-15 particles dataset

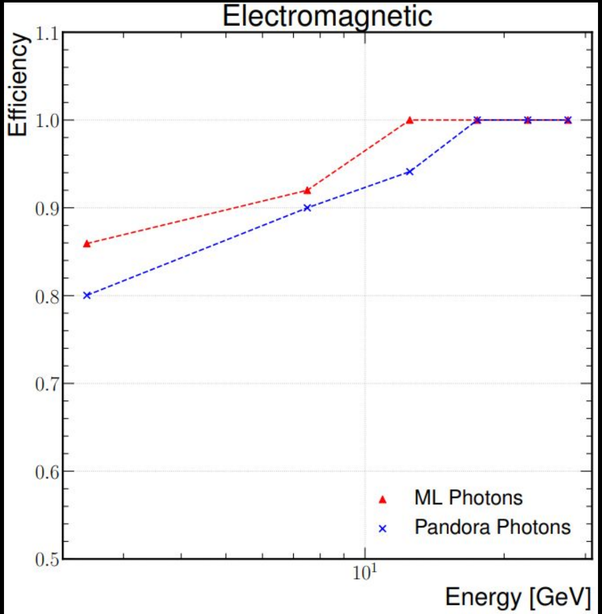
- Better clustering efficiency, energy correction for neutrals leads to better invariant mass resolution (here, particles species are present equally in the same quantity - not the case for physics events!)

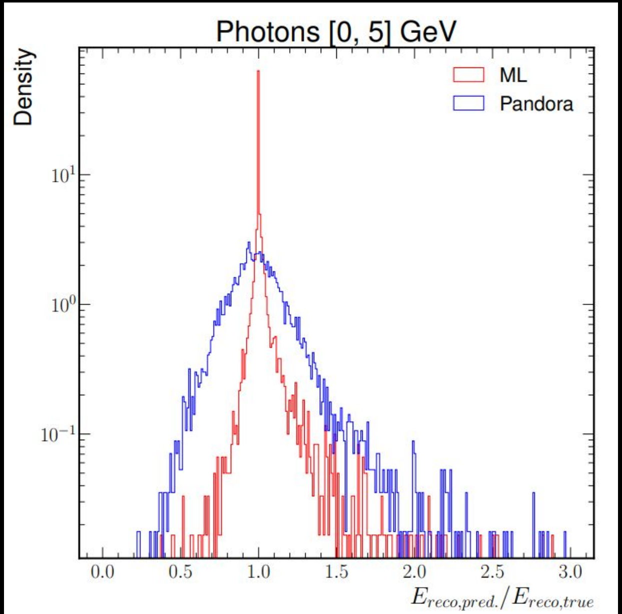


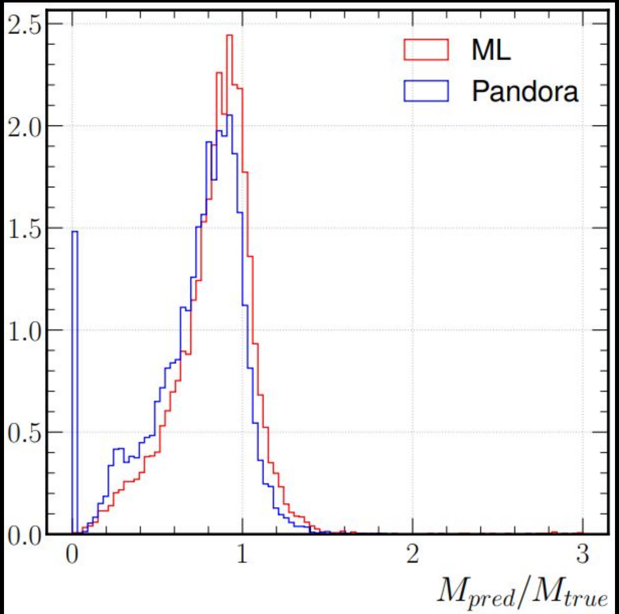


Results: $K_S \rightarrow \pi^0 \pi^0$

- Not feasible to do direction regression - weighted average of the hits from (0,0,0) is a better solution - similar to Pandora





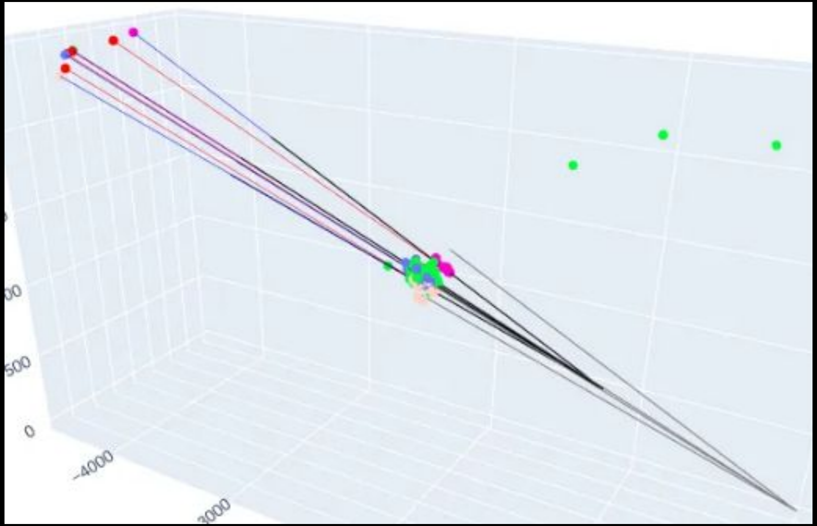


Future work

- Demonstrate the model on physics events
- Improve angular resolution with ML
- Integrate MLPF into FCC software

$K_S \rightarrow \pi^0 \pi^0$ dataset

- Still highly collimated photons



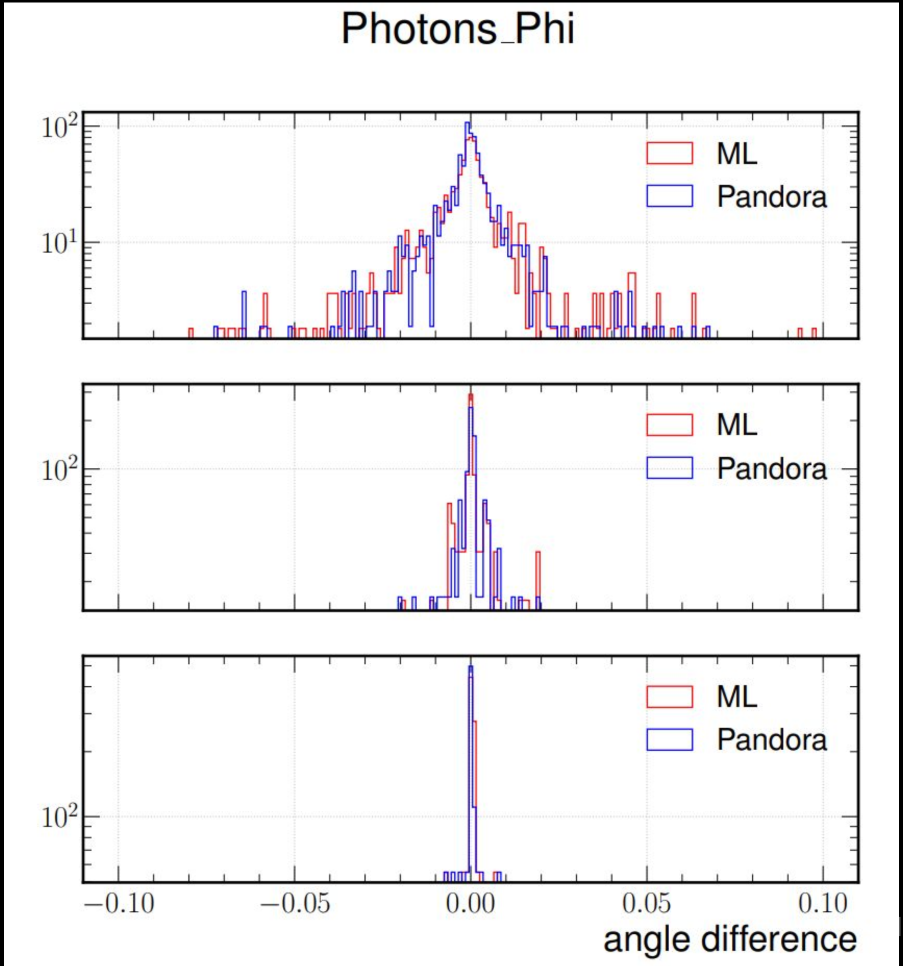
Angular resolution (K_S dataset)

- Weighted average

[0,5] GeV

[5,15] GeV

[15,50] GeV



Angular resolution (K_S dataset)

- Weighted average

[0,5] GeV

[5,15] GeV

[15,50] GeV