# Machine Learning-Based Tracking Algorithm and Particle Flow Reconstruction

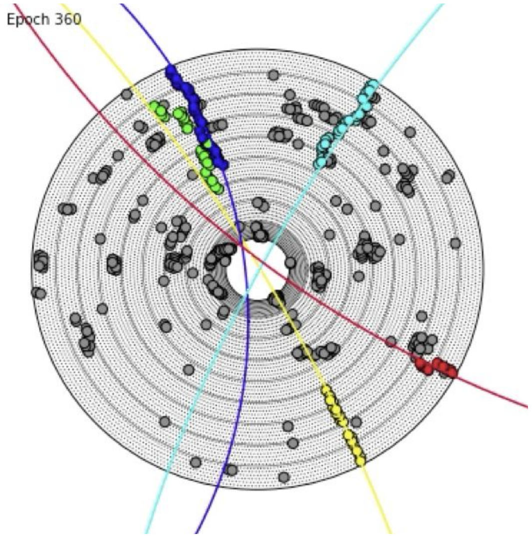
Andrea De Vita, Dolores Garcia, Gregor Krzmanc, Brieuc Francois, Michele Selvaggi

# Table of Contents

# Geometric Graph Neural Network Based Track Finder

# From Current Challenges to a Novel Approach

Traditional track finding methods are often complex and detector-specific, limiting their adaptability. To address this, we propose the **Geometric Graph Track Finding (GGTF)** method, an end-to-end detector-agnostic approach.



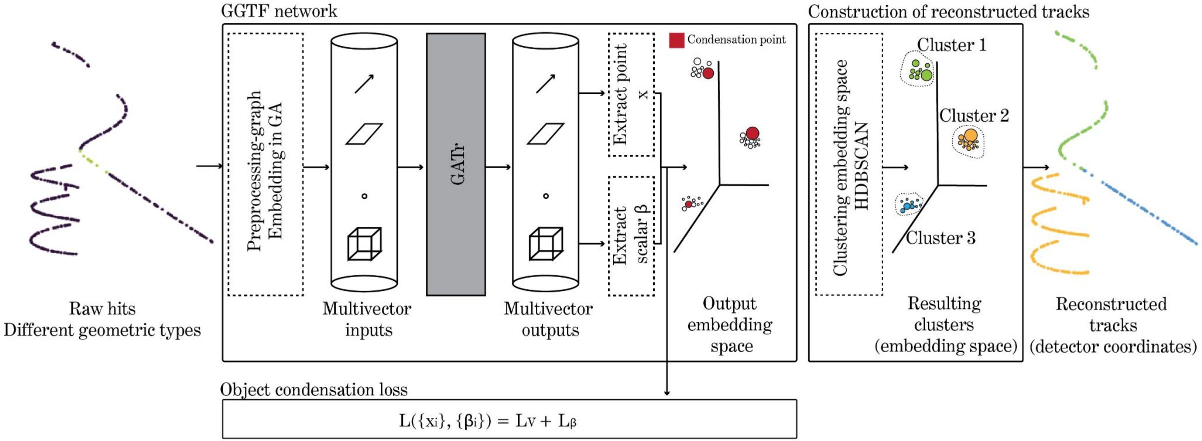

Key features include:

- **Compatibility** with multiple sub-detectors and tracking technologies.
- **Independence from detector geometry** and material specifications.
- **No reliance on analytical trajectory parameterization**.

# Insights into the Neural Network Design

**Inputs :** For each event, a set of hits is collected from various tracking components. The input to the pipeline is the aggregated collection of all hits, denoted as $X=\{X_v, X_i, X_o, \ldots)$ (vertex, inner tracker, outer tracker, …)

**Outputs** : For each event, a set of track is returned. Each track is a collection of hits from different tracking components.

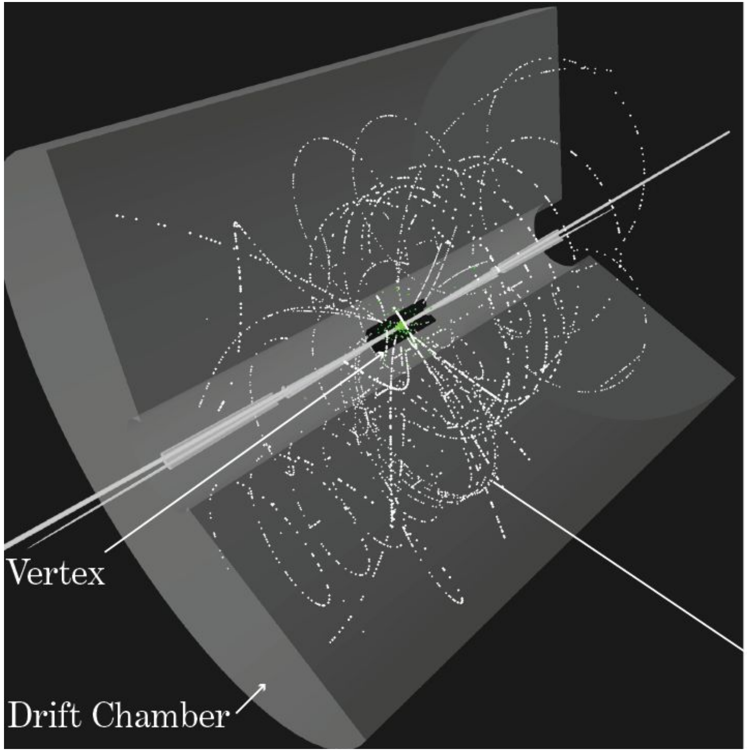



Garcia, D., Francois, B., & Selvaggi, M. (2024). Geometric Graph Neural Network based track finding. CERN.

# Tracking Efficiency

Each reconstructed track is matched to a ground truth particle to which it shares the largest number of hits.

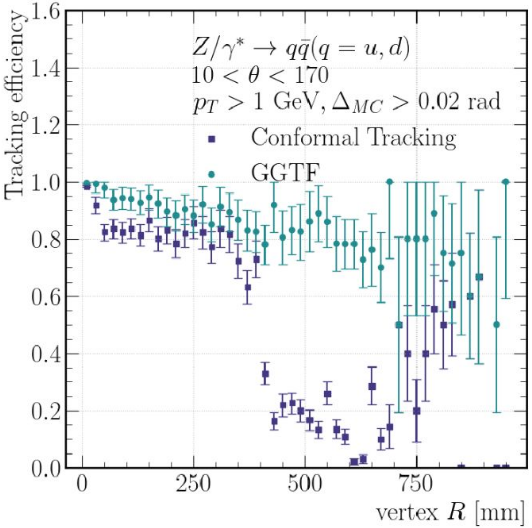
**Tracking Efficiency** is the probability to reconstruct a track and it can be defined as the percentage of reconstructable charged particles matched to a reconstructed track with at least 4 hits.

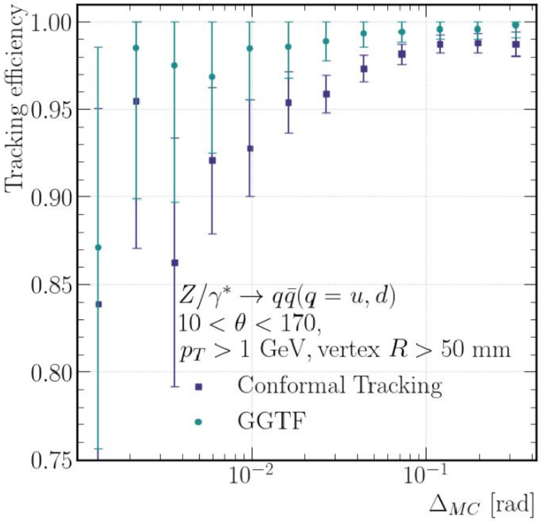


# Performance Analysis on CLD

Results show that **GGTF is better than the baseline in reconstructing displaced tracks**, thanks to its ability to overcome the limitations of the conformal tracking algorithm, which assumes all tracks originate from a single common point.

Tracking efficiency improves as the distance to the nearest track increases. **GGTF has good performance in detecting closely spaced tracks**, outperforming the baseline in these cases.

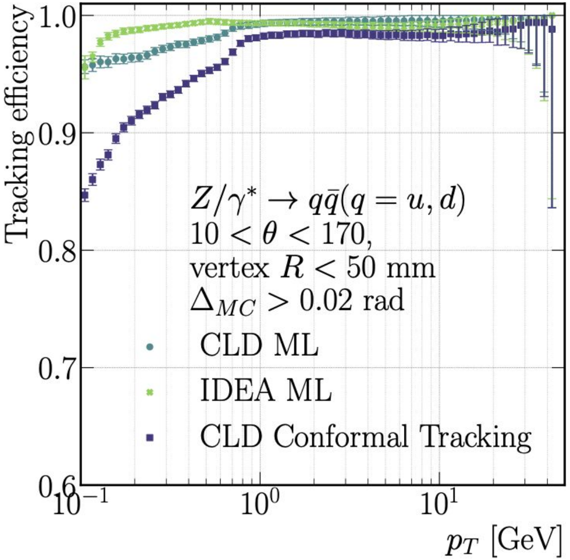


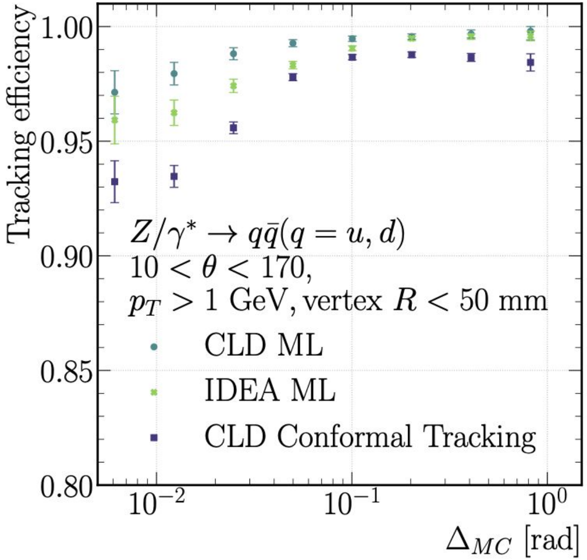



# Comparison between CLD and IDEA

Without background, a comparison of CLD and IDEA shows that IDEA achieves better tracking efficiency at low $p_T$. This is due to the drift chamber design, which allows particles to **traverse a larger portion of the tracker**.

CLD, on the other hand, ensures better identification in the case of **closely spaced tracks**.

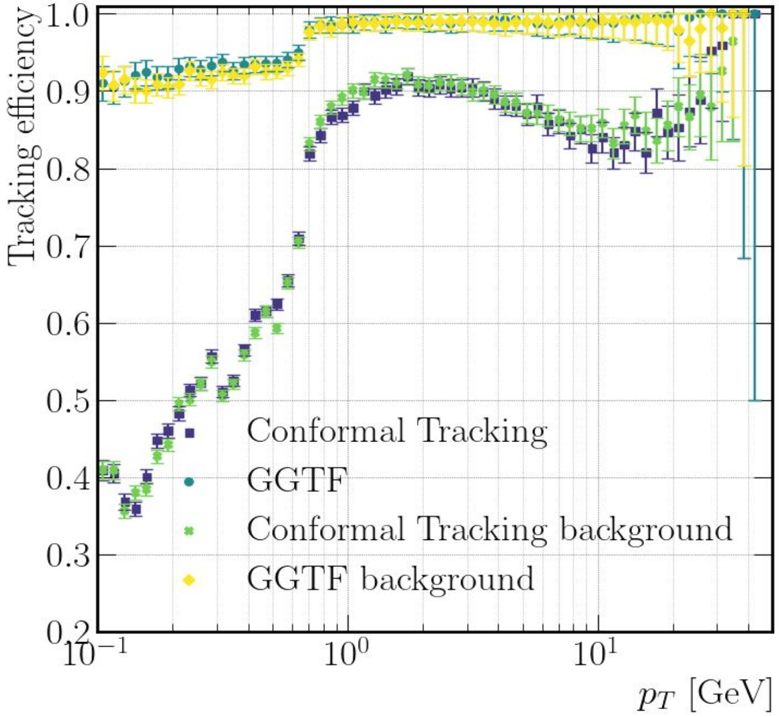


# Background Studies on CLD

The results comparing the CLD with and without background show that the **same tracking performance** can be achieved.

This result is valid for both conformal tracking and GGTF and it confirms that tracking efficiency is higher with GGTF over the entire $p_T$ range.



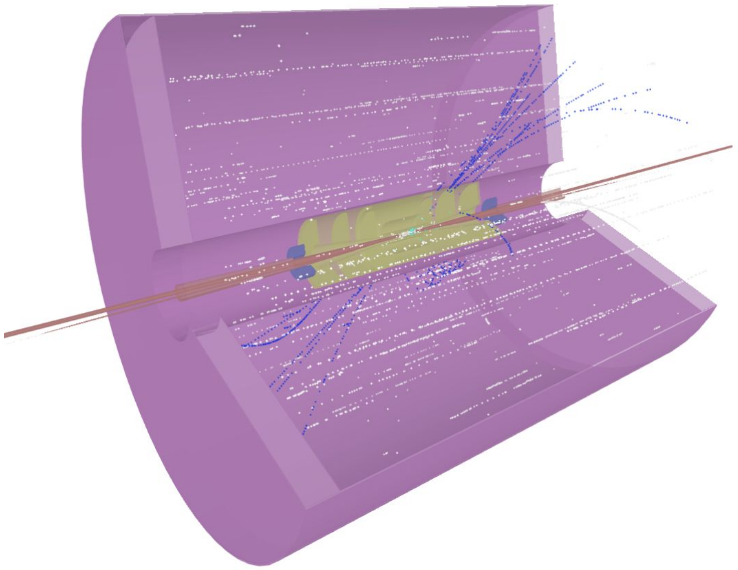

Tracking Efficiency : the percentage of reconstructable charged particles with both ratios, the track hit purity and the track hit efficiency, above 50%.

Number of fake tracks : 5.37% (ML) and 5.47% (CT)

# Background Studies on IDEA

Future studies will evaluate tracking performance with background on IDEA, focusing on challenges posed by **low-energy noise particles** in the drift chamber.

These particles create horizontal lines of **loopers** with a very small radius. To prevent the identification of false tracks, they must be detected and removed before pattern recognition.



# Machine learning-based Particle Flow at the FCC-ee

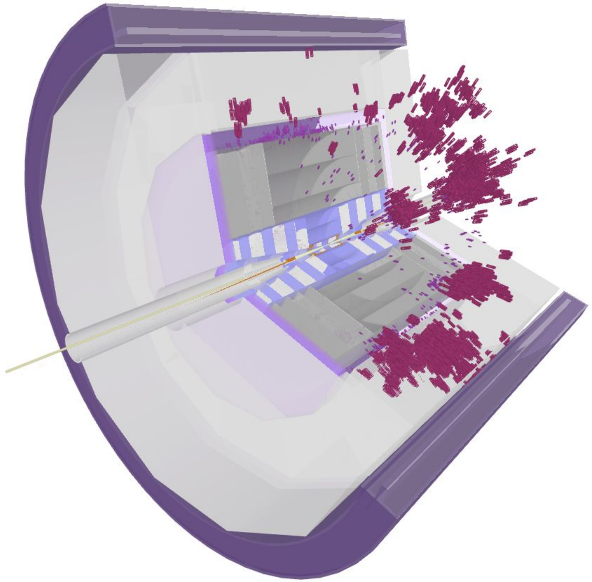
# What is Particle Flow?

More details in the Anna Zaborowska's slides from this workshop.

Method to reconstruct **stable** particles in the event (e, γ, μ, charged and neutral hadrons) using the **information from all sub-detectors**.

Main idea: leverage the **most precise sub-detector** that measures a particle:

- trackers – for charged particles
- calorimetry – for neutral particles.

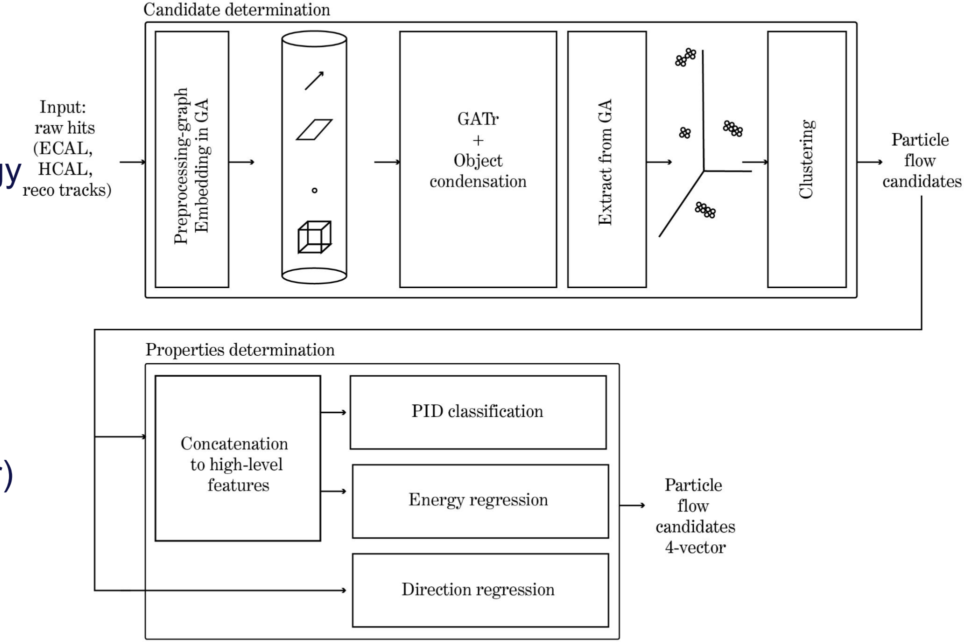


# Insights into the Neural Network Design

**Inputs**: A set of hits from the ECAL, HCAL, and muon system, along with tracks $X=\{X_{ECAL},X_{HCAL},X_{Muon},X_T\}$. Each hit includes energy (E), position (x,y,z), and detector type (ECAL, HCAL, muon subdetectors, tracker).

**Outputs**: The pipeline performs two main steps:

1. Detects particle flow candidates.
2. Determines the properties (PID, four-vector) for each candidate.




Garcia, D., Krzmanc, G., de Cosa, A., & Selvaggi, M. (2024). Machine Learning based Particle Flow. CERN.

# Metrics Overview: Building Blocks for Evaluation

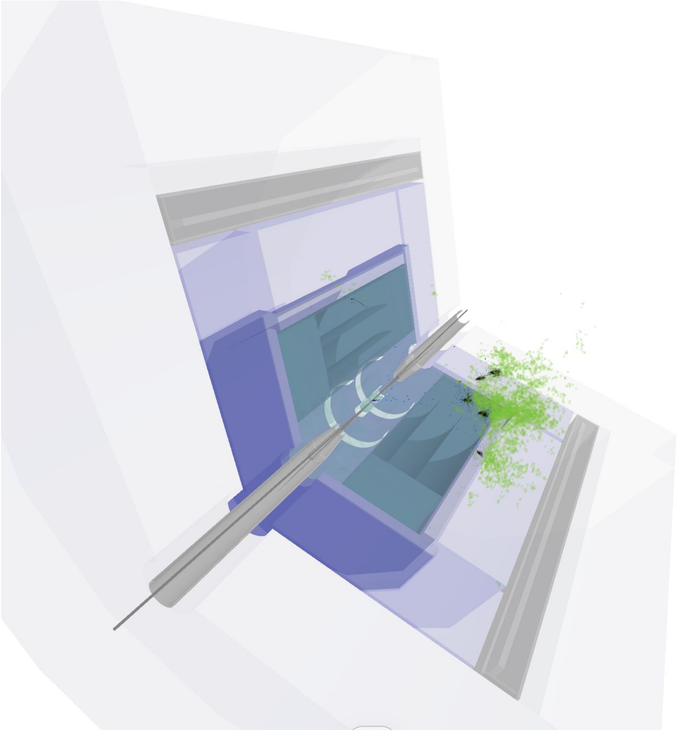
The following features allowing for comparison of MLPF with Pandora are computed.

- **Energy resolution** is estimated by fitting a Gaussian PDF N(μ, σ) onto the histogram of $E_{model}/E_{true}$ and reporting the variance divided by mean σ/μ.
- **Angular resolution** is estimated by looking at the difference in angles Φ and θ, defined according to the CMS coordinate system.
- **Efficiency**: the fraction of particles that are reconstructed and matched to one of the true showers with respect to the total number of true showers.
- **Fake rate**: the fraction of particles that are reconstructed without matching to a true shower with respect to the total number of true showers.
- **Confusion Matrix** : matrix reporting the PID step performance.

# Dataset - Jet-like Particle Gun

- Generates **10–15 stable particles** per event.
- **Energy range**: 0.5 GeV < E < 50 GeV
- **Particles included**: $\pi^+$ , pi$^+$, $K_{L,}$, n, γ, e$^+$, μ , K$^+$, p
- Balanced classes: Charged hadrons,Neutral hadrons, Photons, Muons, Electrons.

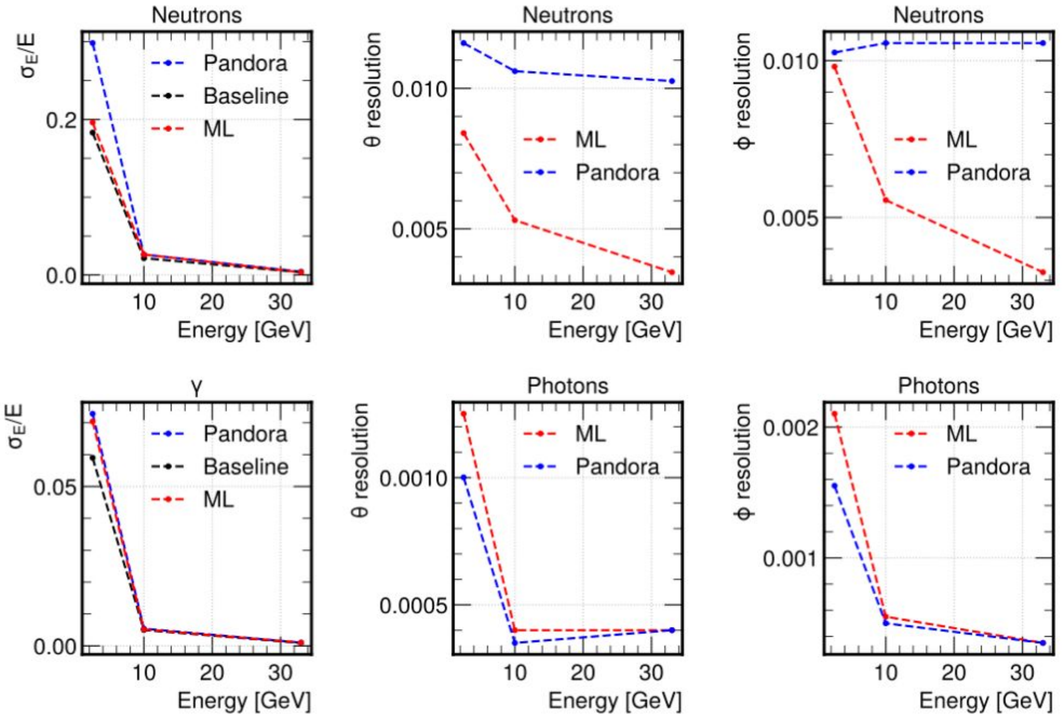
It offers a controlled approximation of general physics events, excluding many exotic or long-lived particles by default but allowing their inclusion if needed.



# Results - Energy and Angular Resolution

The MLPF shows improved energy resolution for the neutral hadrons.

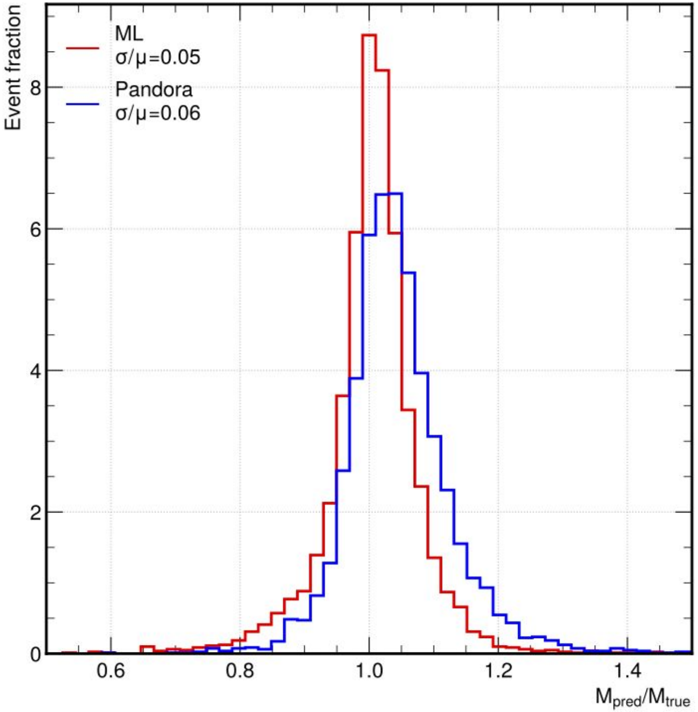
The MLPF model also shows equivalent angular resolution to Pandora, with improved resolution for neutral hadrons.



# Results - Reconstructed Mass

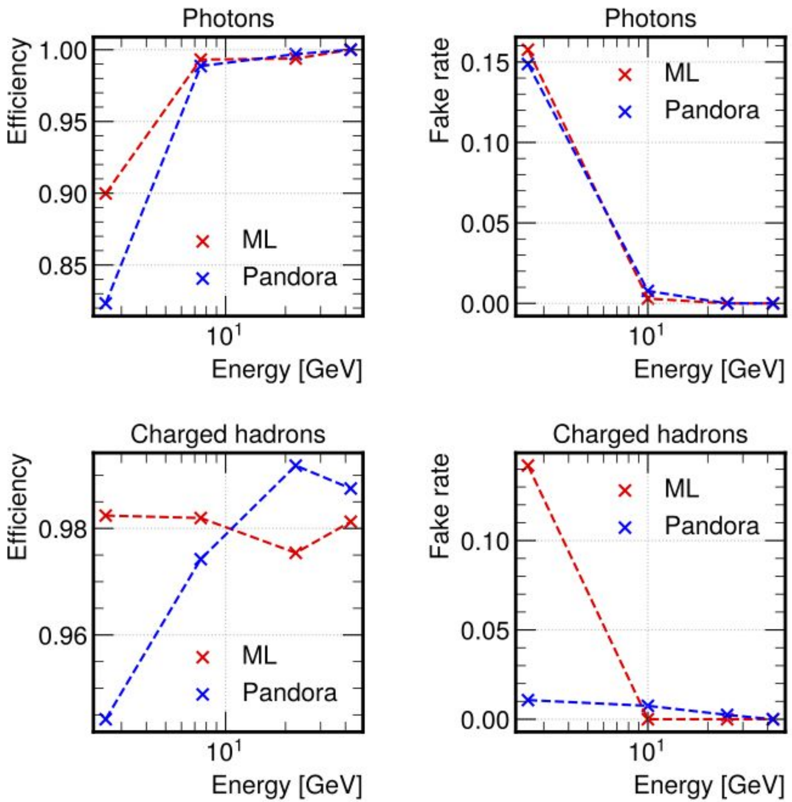
These results demonstrate that the ML method outperforms Pandora in mass reconstruction. Specifically, the **$m_{pred}/m_{true}$ ratio** obtained with the ML method has a **mean value closer to 1** compared to Pandora's results.

Additionally, the **mass relative resolution σ/μ is smaller** when applying the ML method.



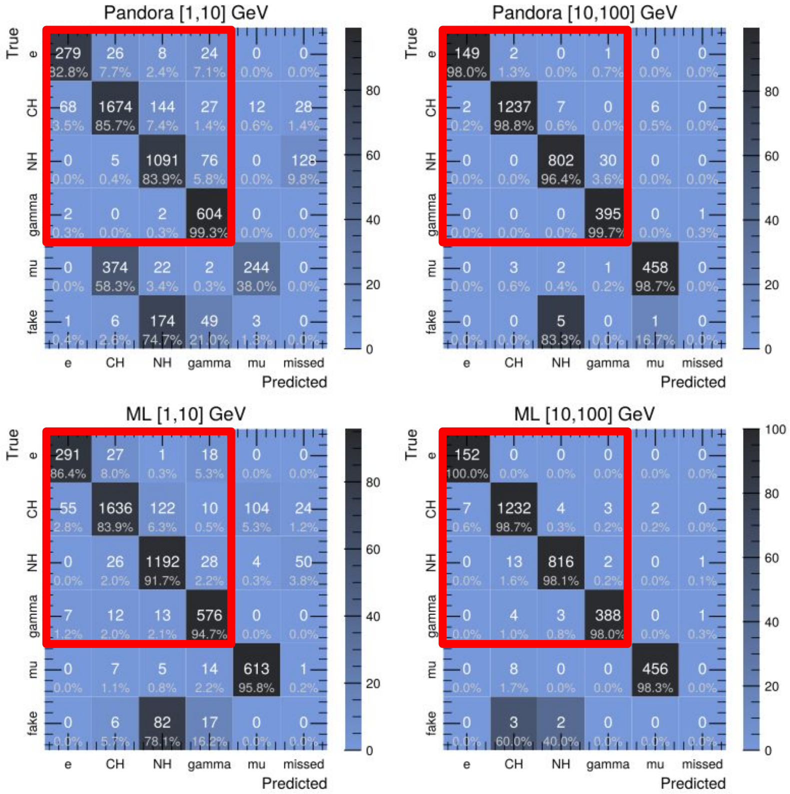
# Results - Efficiency and Fake Rates

- **Low Energies**:
    - MLPF achieves **higher efficiency** compared to Pandora.
    - **Fake rate** remains similar, expect for charged hadrons (due to fake tracks).
- **High Energies**:
    - Efficiency decreases due to challenges in clustering **close-by showers**.



# Results - Particle Identification ( I )

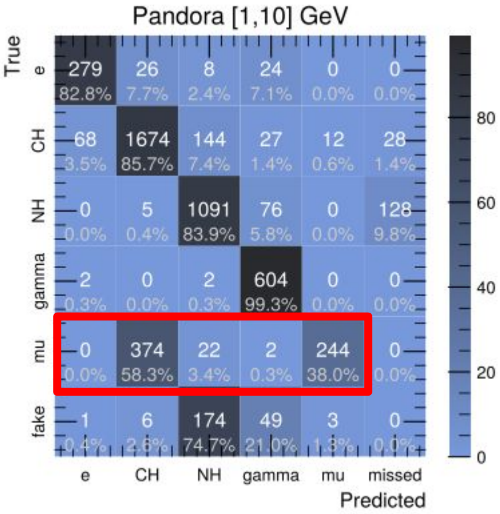
In the electron, charged hadrons, neutral hadrons and gamma region the overall performance of MLPF is compatible with Pandora, but the PID also shows **improved performance for MLPF** with 6% misidentification rate of pandora of neutral hadrons into photons.

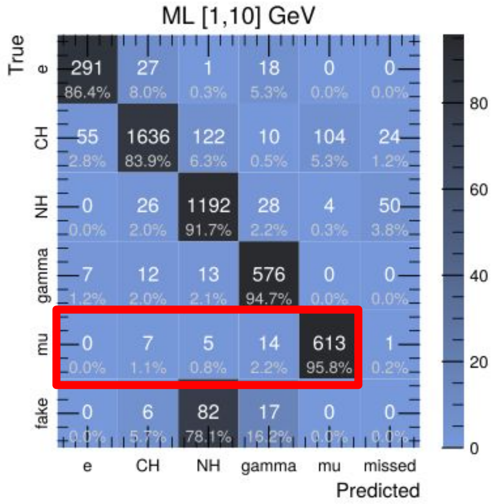


# Results - Particle Identification ( II )

Muons in the range of 1-10 GeV are correctly **classified by the MLPF 95%** of the times (38% for pandora).

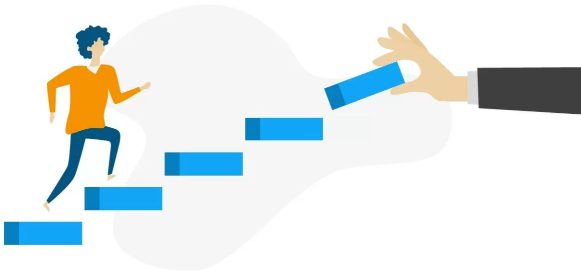
The analysis highlights challenges with lower-energy muons, which often leave only a **few hits in the muon detectors**. Due to this limited information, the Pandora algorithm tends to misclassify these muons as charged hadrons.



# Summary

# Summary and Next Steps



**Geometric Graph Neural Network Based Track Finder**:

- **GGTF Method**: An end-to-end, detector-agnostic approach that shows encouraging results for both CLD and IDEA detector concepts.
- **Background Robustness**: GGTF demonstrates consistent performance with and without background in the CLD detector concept.
- **Next Steps**: Background studies on IDEA to evaluate the effect of drift chamber noise on GGTF performance.

**Machine learning-based Particle Flow at the FCC-ee:**

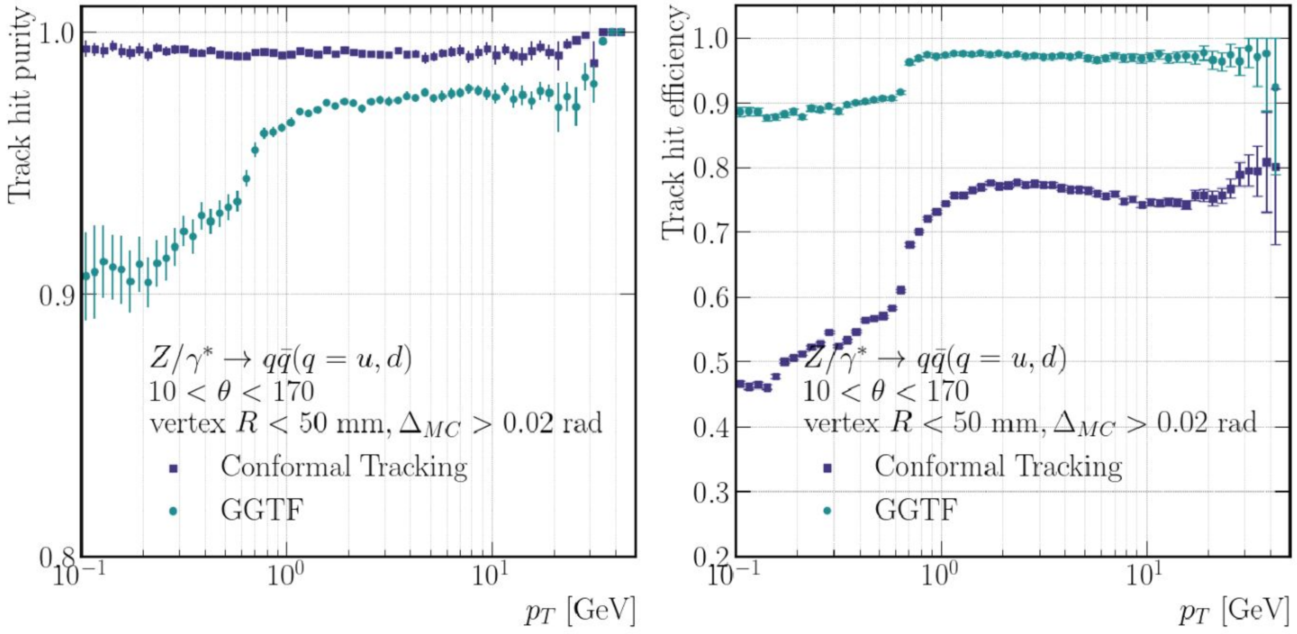
- **MLPF**: An end-to-end, detector-agnostic approach with similar performance to Pandora.
- **Next Steps**: Add the possibility of discarding classifications made by the PID module in the event that the assignment is not sure, but it has ambiguities with other particles.

# Thank you for your attention.

# Backup

# Performance Analysis on CLD

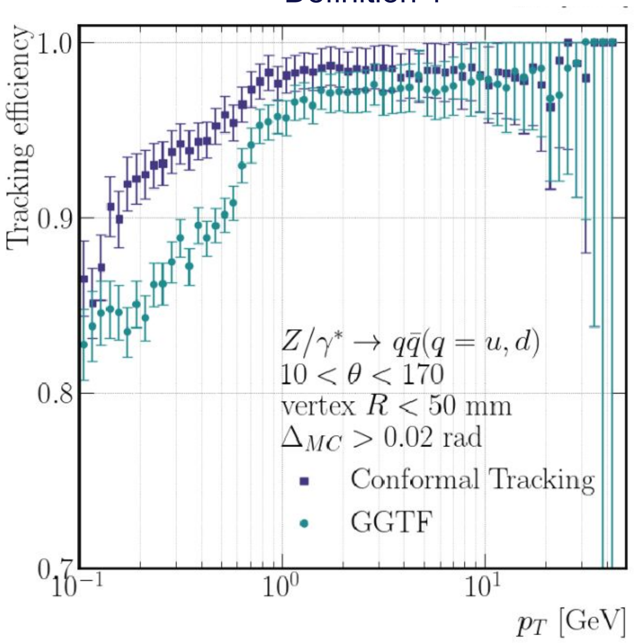
## Track Hit Purity & Track Hit Efficiency



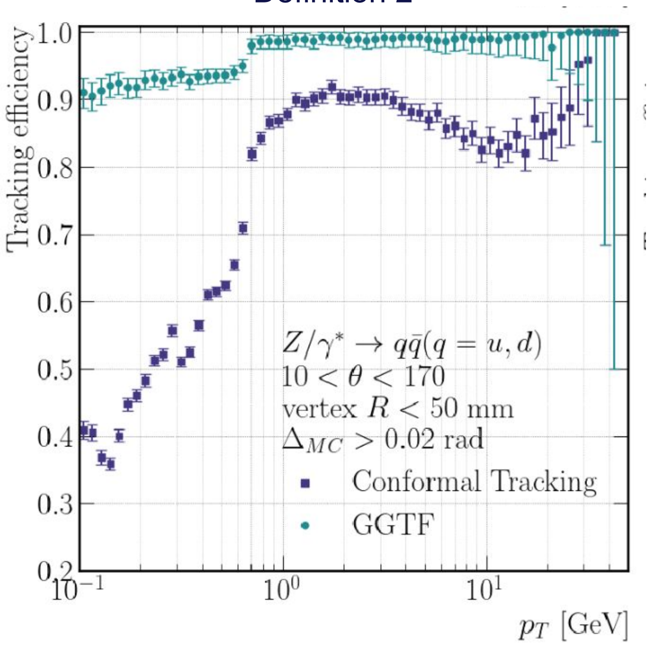
# Performance Analysis on CLD

## Tracking Efficiency with Definitions 1 and 2

Definition 1




Definition 2



# Performance Analysis on IDEA

## Tracking Efficiency & Track Hit Efficiency