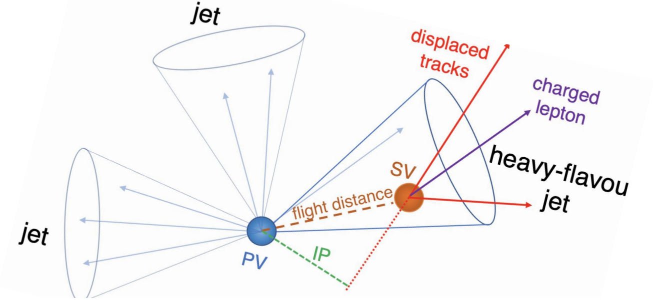
CERN

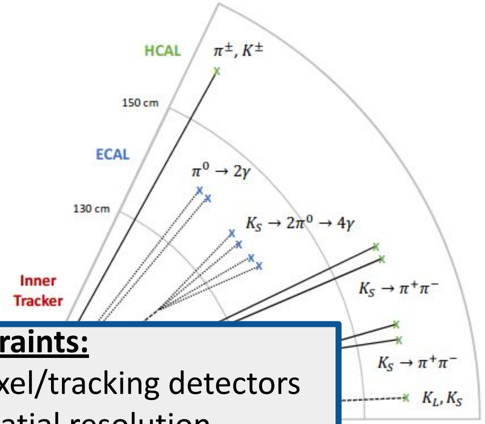
Event reconstruction using ML for FCC-ee

Dolores Garcia

Motivation, s & b tagging



- Flavour tagging essential for the e^+e^- program
  - Measure Higgs particle properties and interactions in challenging decay modes
  - Precise determination of top properties
  - QCD physics: modeling hadronization…
- **b-tagging:**
  - Large lifetime, (2-3) mm for ~50 GeV boost
  - Displaced vertices/tracks (Large impact parameters)
  - Large track multiplicity
  - Presence of non-isolated e/μ (20 (10)% in B (C) decays)
- **s-tagging:**
  - Large Kaon content:
    - as tracks (K/pi separation ToF, dEdx, dNdx)
    - Neutral Kaons:
      - K_S 2 tracks
      - K_L ToF vs n



Detector constraints:

Need power pixel/tracking detectors

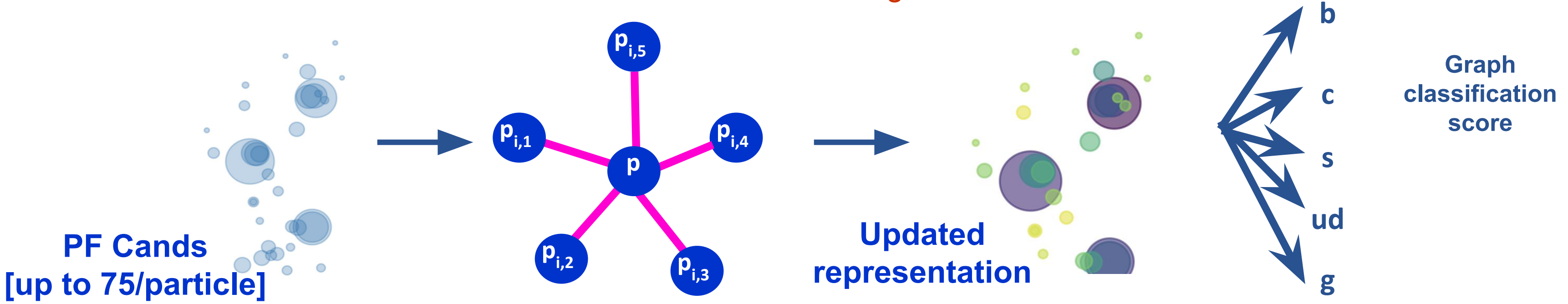
- Good spatial resolution
- As little material as possible
- Precise track alignment
- Timing detectors
- Charged energy loss (gas/silicon)

ACKS : Michele Selvaggi, Loukas Gouskos, Franco Bedeschi

Graph based tagger

[O(50) properties/particle]
x [~50-100 particles/jet]
~O(1000) inputs/jet

- Particle based jet tagging using all information from PF reconstruction
- Jet representation: *"Point Cloud"* → *"Particle Clouds"*
  - Treat the jet as an unordered set of particles
- Algorithm design: Graph Neural Networks
  - Particle cloud represented as a graph
    - Each particle: **node** of the graph; Connections between particles: the **edges**
- Follow a hierarchical learning approach
  - First learn local structures → then move to more global ones

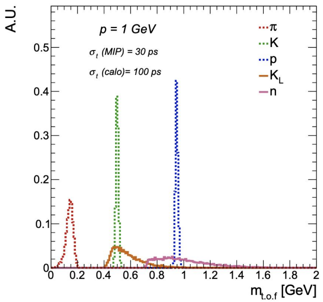


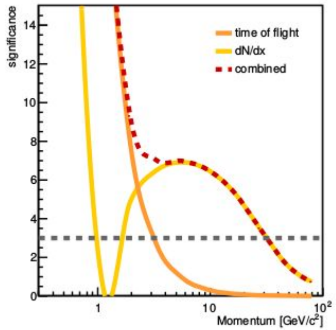
ACKS : Michele Selvaggi, Loukas Gouskos, Franco Bedeschi

[1] Qu, H., & Gouskos, L. (2020). Jet tagging via particle clouds. *Physical Review D*, *101*(5), 056019.
[2] Qu, H., Li, C., & Qian, S. (2022, June). Particle transformer for jet tagging. In *International Conference on Machine Learning* (pp. 18281-18292). PMLR.

Input variables

- Kinematic
- Displacement (important for b-tagging)
- Identification:
  - Number of ionization clusters (dN/dx)
  - ToF results in good K/π separation at low-momenta
  - Module added to Delphes





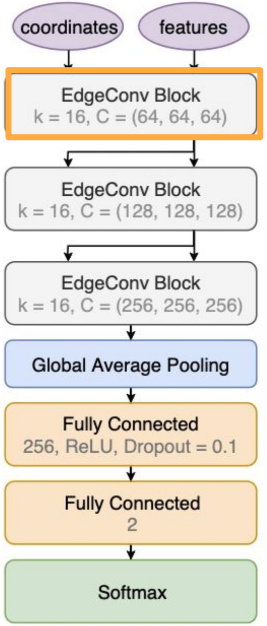
Variable	Description
	Kinematics
$E_{\text{const}}/E_{\text{jet}}$	energy of the jet constituent divided by the jet energy
θ_{rel}	polar angle of the constituent with respect to the jet momentum
ϕ_{rel}	azimuthal angle of the constituent with respect to the jet momentum
	Displacement
d_{xy}	transverse impact parameter of the track
d_z	longitudinal impact parameter of the track
SIP_{2D}	signed 2D impact parameter of the track
$\text{SIP}_{\text{2D}}/\sigma_{\text{2D}}$	signed 2D impact parameter significance of the track
SIP_{3D}	signed 3D impact parameter of the track
$\text{SIP}_{\text{3D}}/\sigma_{\text{3D}}$	signed 3D impact parameter significance of the track
d_{3D}	jet track distance at their point of closest approach
$d_{\text{3D}}/\sigma_{d_{\text{3D}}}$	jet track distance significance at their point of closest approach
C_{ij}	covariance matrix of the track parameters
	Identification
q	electric charge of the particle
$m_{\text{t.o.f.}}$	mass calculated from time-of-flight
dN/dx	number of primary ionisation clusters along track
`isMuon`	if the particle is identified as a muon
`isElectron`	if the particle is identified as an electron
`isPhoton`	if the particle is identified as a photon
`isChargedHadron`	if the particle is identified as a charged hadron
`isNeutralHadron`	if the particle is identified as a neutral hadron

Architectures

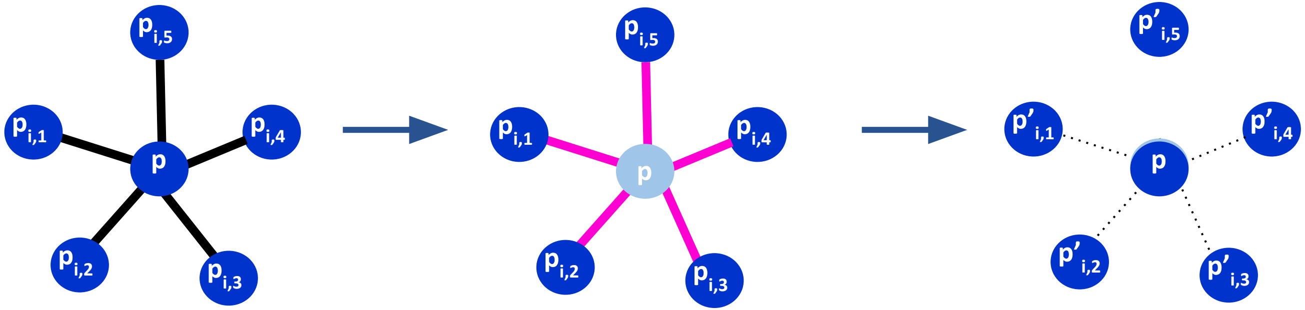
- ParticleNeT [1]:
  - Find the ***k*-nearest neighbors** of each point
  - Design a permutation invariant **convolution operation**
    - Define an **edge feature** function → **aggregate** edge features w/ a symmetric func.
  - Dynamically update graph
- Other architectures: Particle Transformer [2] FC transformer

particles distributed in η-φ

from local to global



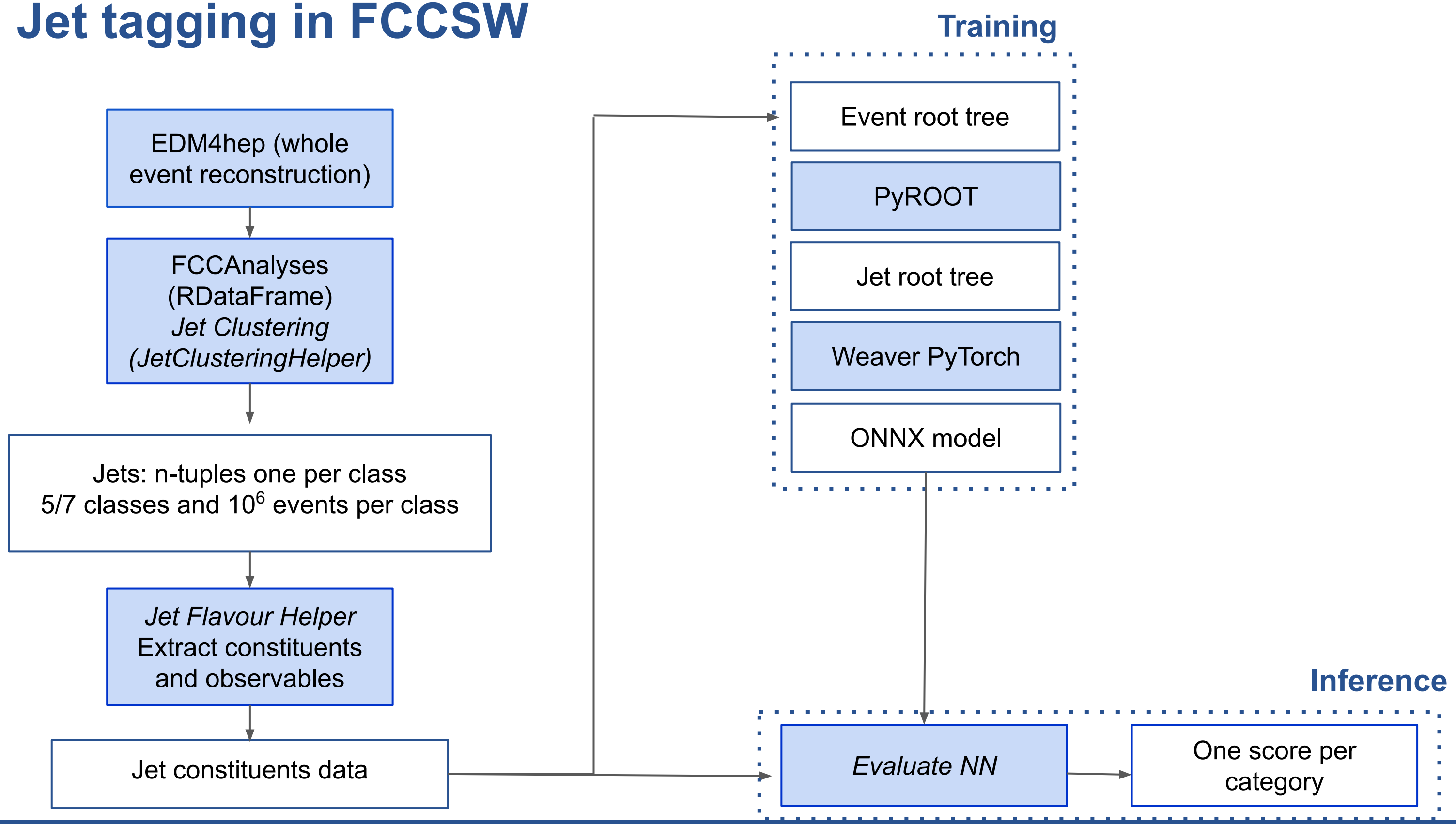
k-Nearest Neighbors **Convolution operation** **Update Graph**



ACKS : Michele Selvaggi, Loukas Gouskos, Franco Bedeschi

[1] Qu, H., & Gouskos, L. (2020). Jet tagging via particle clouds. *Physical Review D*, *101*(5), 056019.
[2] Qu, H., Li, C., & Qian, S. (2022, June). Particle transformer for jet tagging. In *International Conference on Machine Learning* (pp. 18281-18292). PMLR.

Jet tagging in FCCSW

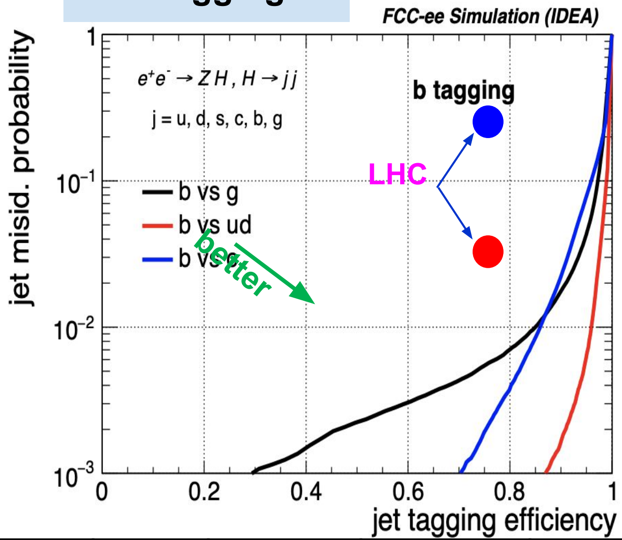


ACKS : Michele Selvaggi, Loukas Gouskos, Franco Bedeschi

Results

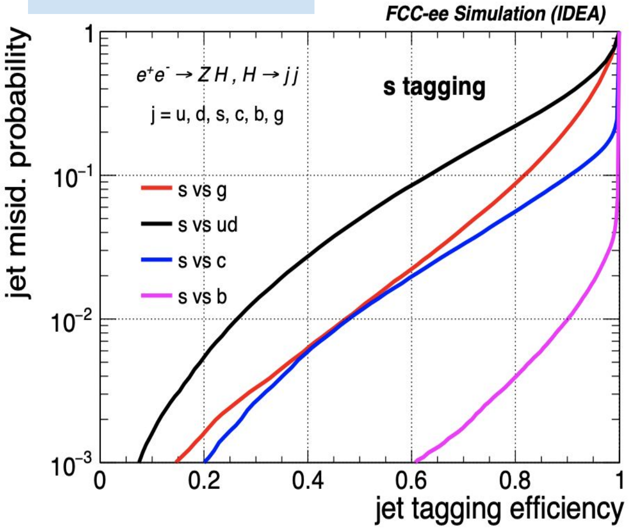
Some examples: (Also new classes for u/d/tau and different detector configurations)

b-tagging



WP	Eff (b)	Mistag (g)	Mistag (ud)	Mistag (c)
Loose	90%	2%	0.1%	2%
Medium	80%	0.7%	<0.1%	0.3%

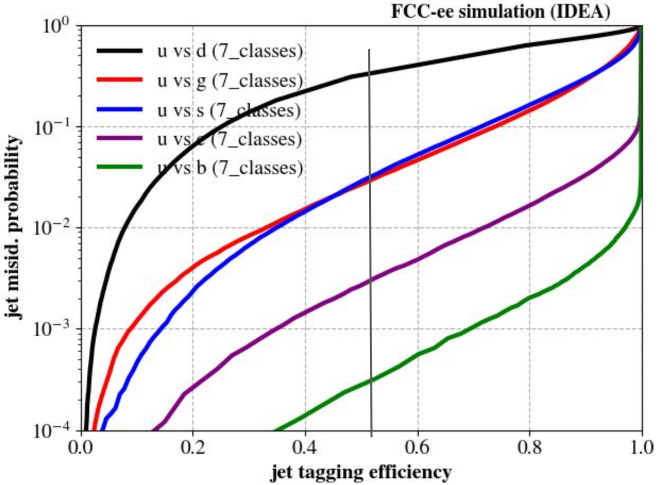
s-tagging



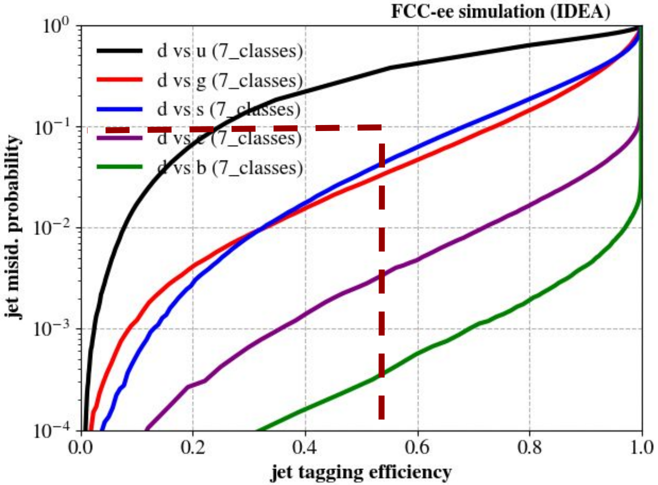
WP	Eff (s)	Mistag (g)	Mistag (ud)	Mistag (c)	Mistag (b)
Loose	90%	20%	40%	10%	1%
Medium	80%	9%	20%	6%	0.4%

Tagger update: up and down

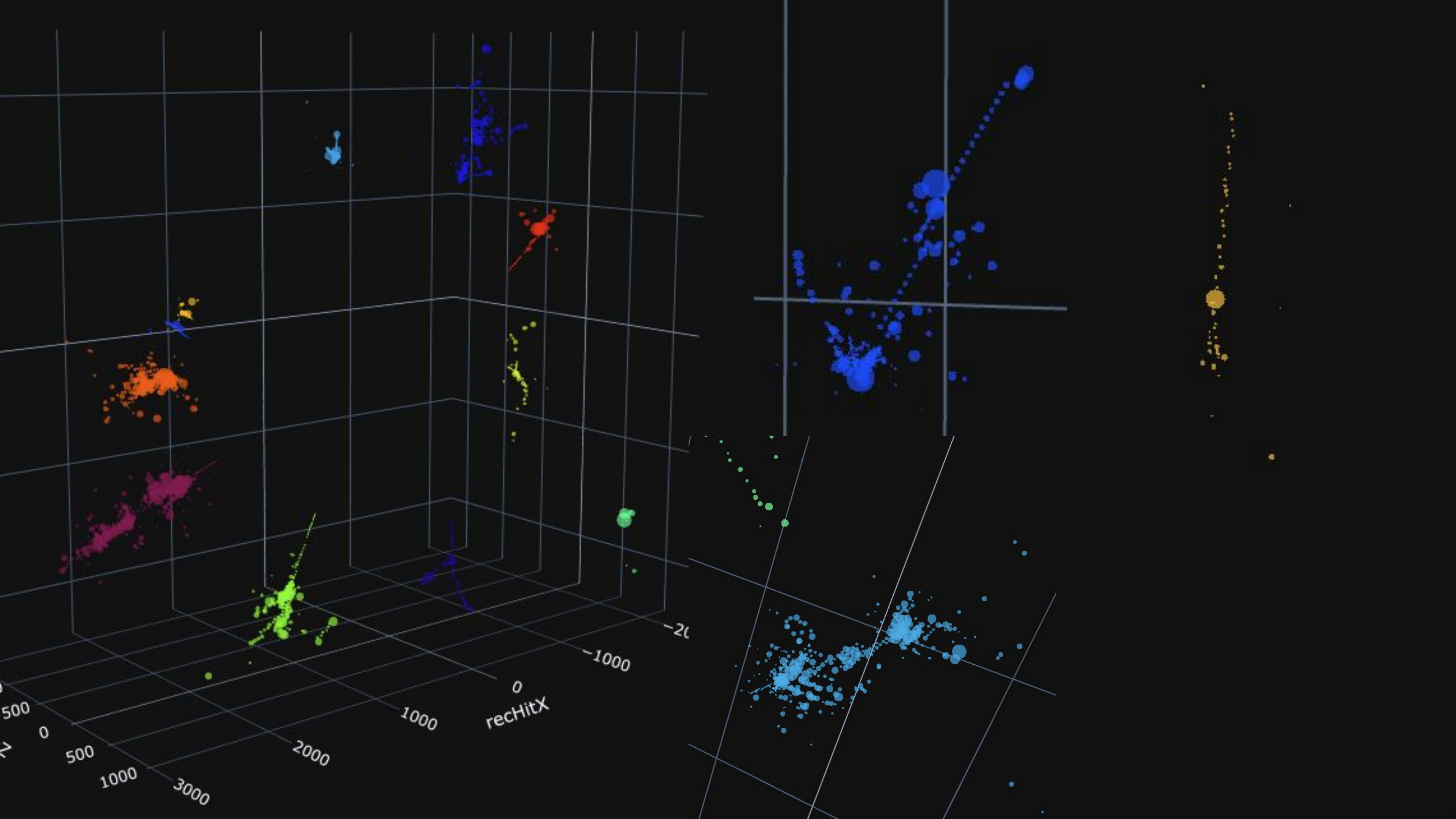
Up-tagging



Down-tagging

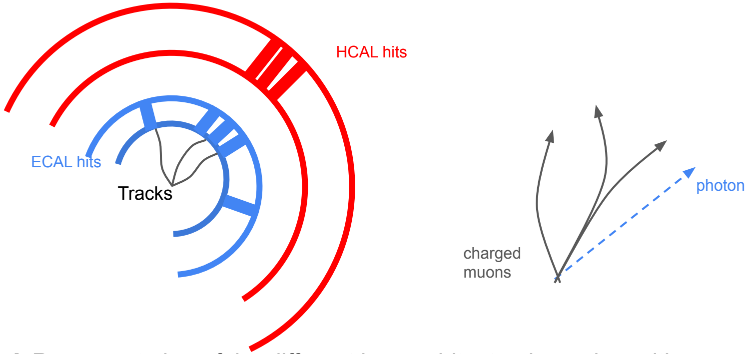


- Up vs Down discrimination seems possible thanks to jet charge
- 30% bkg eff at 50% signal (better than random coin toss)

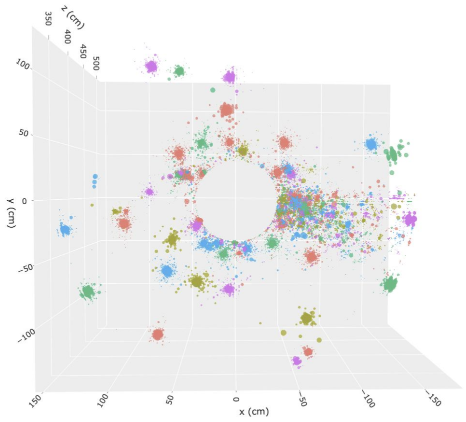

recHitX
0
1000
2000
3000
-1000
500
0
500
1000

MLPF: Motivation

- 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
- First step: **focus on calorimeter clustering**



A Representation of the different layers, hits, tracks and resulting particles (reproduced from [1])



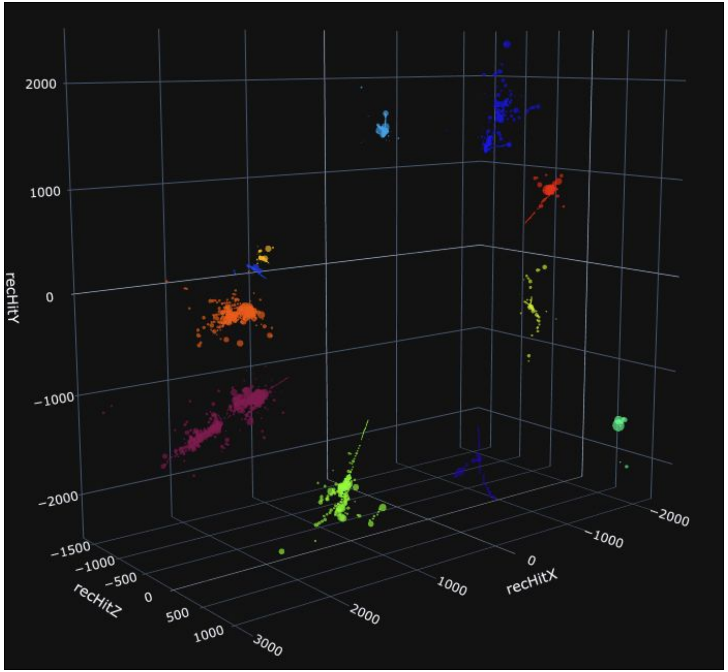
B Example of an event, the shower of secondary particles generated by an individual particle is labelled with one colour [2]

ACKS : Michele Selvaggi, **Gregor Krzmanc,** Jan Kieseler, Philipp Zehetner

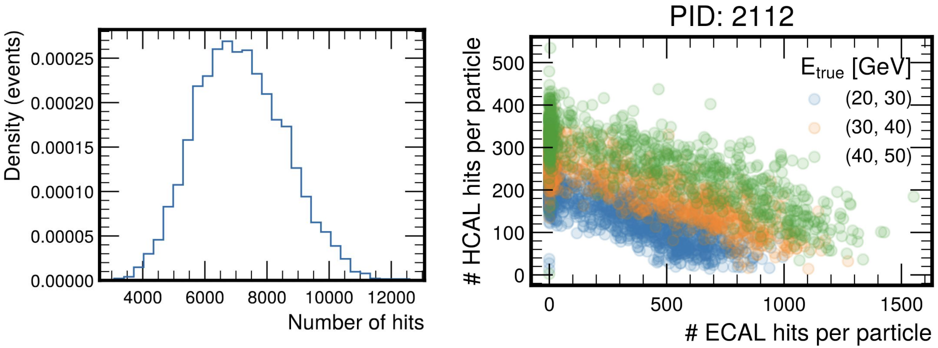
[1] Pata, J. Machine learning for particle flow reconstruction at CMS, presentation at CDS.
[2] Qasim, S. R., Chernyavskaya, N., Kieseler, J., Long, K., Viazlo, O., Pierini, M., & Nawaz, R. (2022). End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks.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

Training Data

A Example train event - 15 particles



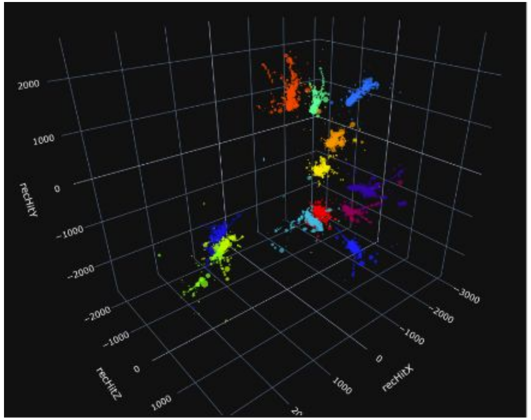
- Event generation:
  - Use particle gun (0-15 particles)
  - $E \in [0.5, 50]$ GeV
  - p, n, K_L, π
- FCC-ee O(100)
- Simulation and reconstruction: Key4HEP turnkey + Geant4 (CLIC pipeline)



B Number of hits per event (left) and #hits ECAL vs HCAL (right)

ACKS : Michele Selvaggi, **Gregor Krzmanc,** Jan Kieseler, Philipp Zehetner

Architecture: Object condensation (End-to-End approach)



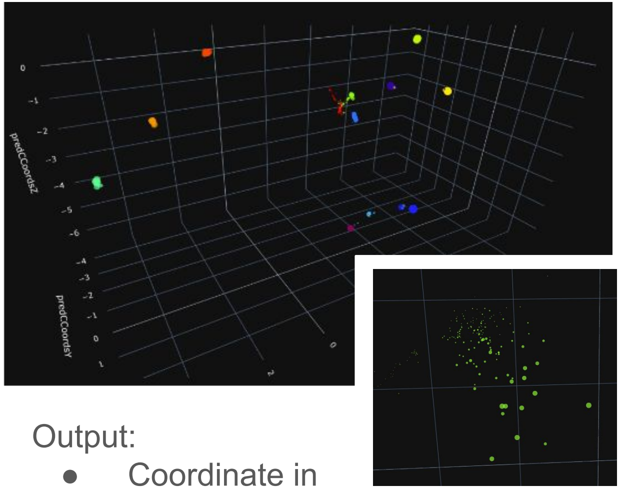
CP

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

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

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

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



Input:

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

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

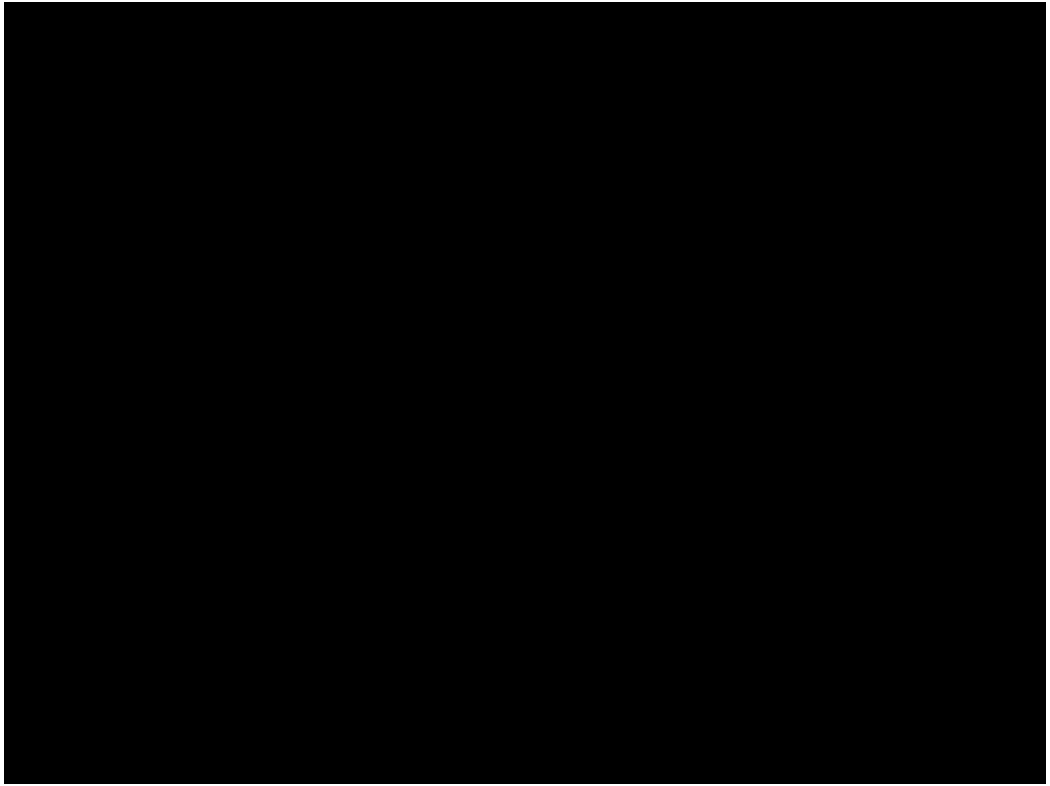
Output:

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

ACKS : Michele Selvaggi, Gregor Krzmanc, 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.

Architecture: Object condensation (End-to-End approach)



ACKS : Michele Selvaggi, Gregor Krzmanc, Jan Kieseler, Philipp Zehetner

Architecture: Gravnet Model

- Input: a set of hits from different sensors (coordinates, type of hit, energy), each one node in the graph O(600) per particle

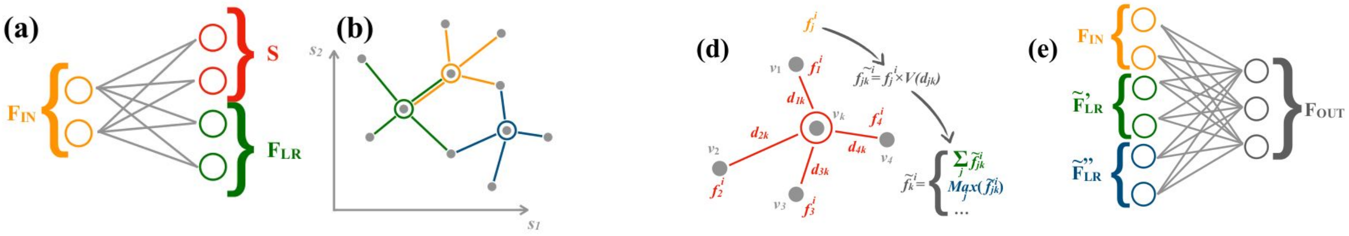
- Graph representation with **no given graph structure**
- **Dynamically** compute edges in embedding space with knn

a) Transform input features $\mathbf{F_{IN}}$ into
   - transformed features $\mathbf{F_{LR}}$
   - latent coordinates **S**

b) Build graph using coordinates **S**

d) Aggregate weighted features
   - Weights depending on distance
   - Aggregation typically is *mean* or *max*

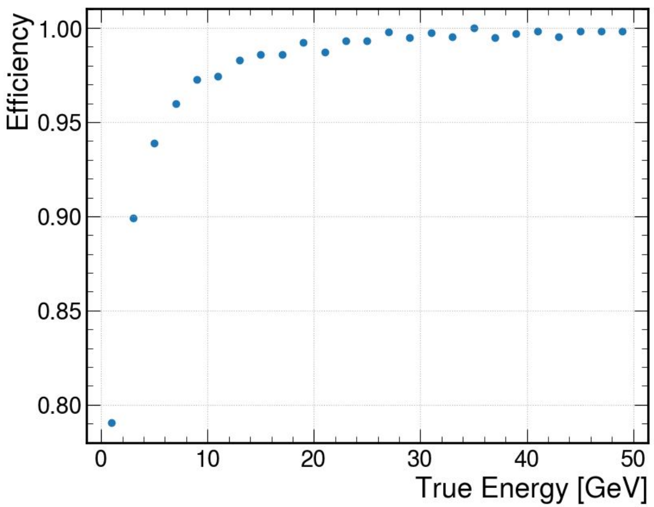
e) Concatenate the new features

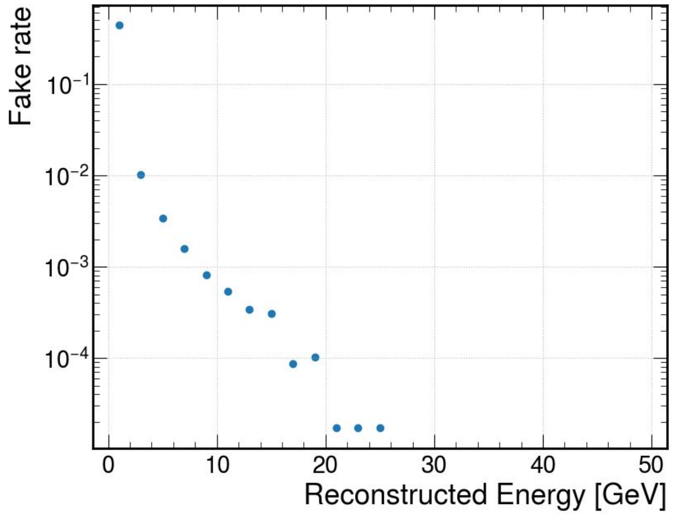


ACKS : Michele Selvaggi, Gregor Krzmanc, Jan Kieseler, Philipp Zehetner

[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.

Efficiency and fake rate

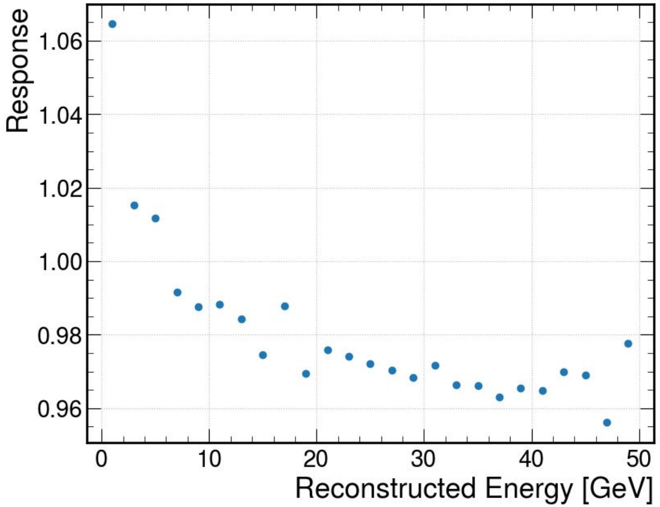




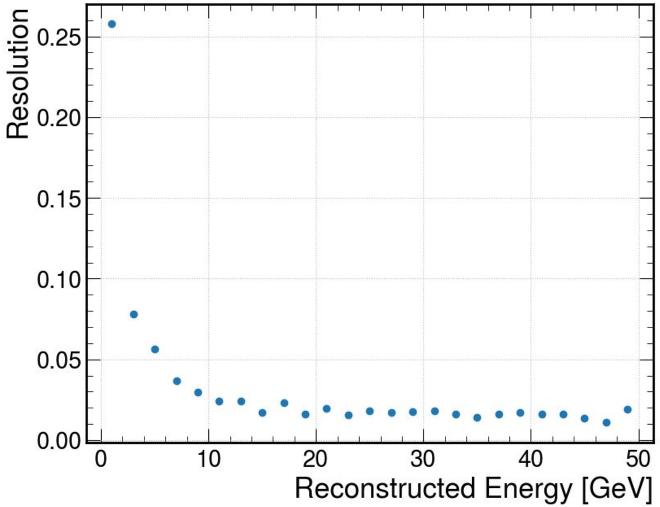
- Efficiency approaches 100% with high p_T
- Adding tracks will improve efficiency

- Most fakes with E< 1 GeV
- Other clustering methods in the embedding space can improve fakes
- Resulting from bad beta distributions

Response and Resolution (clustering metrics)



- Evaluated on reco values (for clustering evaluation)

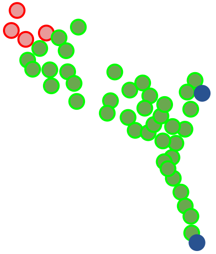


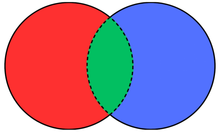
- Resolution performance must be improved for low energies
- Can be improved with better clustering in embedding space

ACKS : Michele Selvaggi, Gregor Krzmanc, Jan Kieseler, Philipp Zehetner

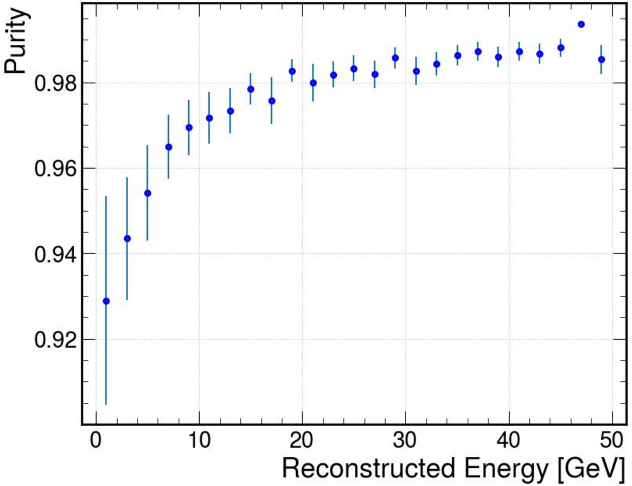
Calorimeter clustering - Results

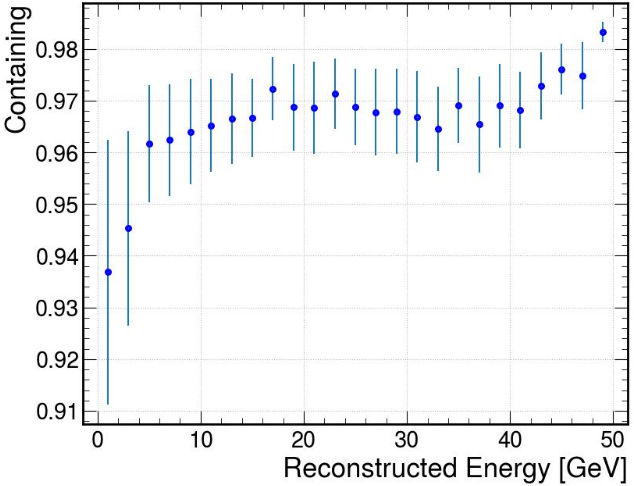
- Containing: percentage of reco energy that belong to the reconstructed particle **(G+R)/(G+B)**
- Purity: Percentage of reco energy **contained** in reconstructed cluster **(G)/(G+R)**





True
Reco





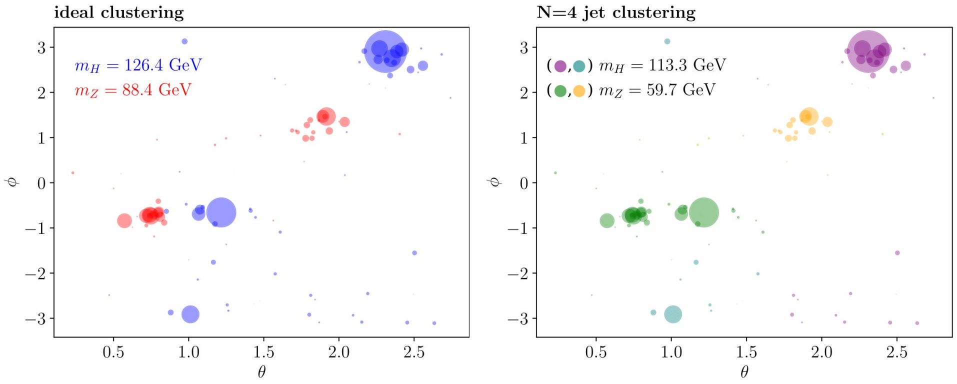
ACKS : Michele Selvaggi, Gregor Krzmanc, Jan Kieseler, Philipp Zehetner

Clustering Color Singlets

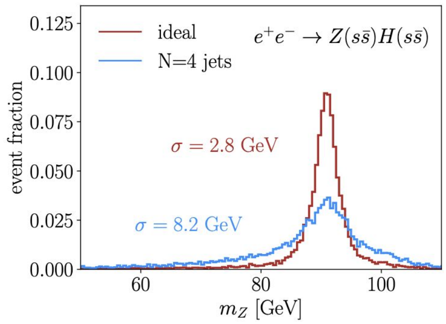
- Identification of color-neutral resonances relies on clustering final state into jets
- Calorimetry is expected to be much improved at future e+e− colliders, so that the 2-jet invariant mass resolution will be dominated not by detector resolution but rather by mis-clustering [1]
- Jets are not well defined but color connection is physical, this may help **improve the mass estimation for color singlets** (H,Z,W) **and remove more background**

Errors can be due to:

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



A Miss clustering example ZH→ssss



B Comparison of clustering performance vs ideal reconstruction

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.

Summary and Outlook

Summary:

- Promising performance, we will soon compare to PFA (baseline for CLD)
- Demonstrated generalization over different types of events (for now kept particle number low)
- Fast execution time, linear scaling with number of hits

Ongoing work and next steps:

- Add tracks as inputs to the graph
- Regress particle properties
- Try heterogeneous graph architectures
- Compare to the performance of PFA

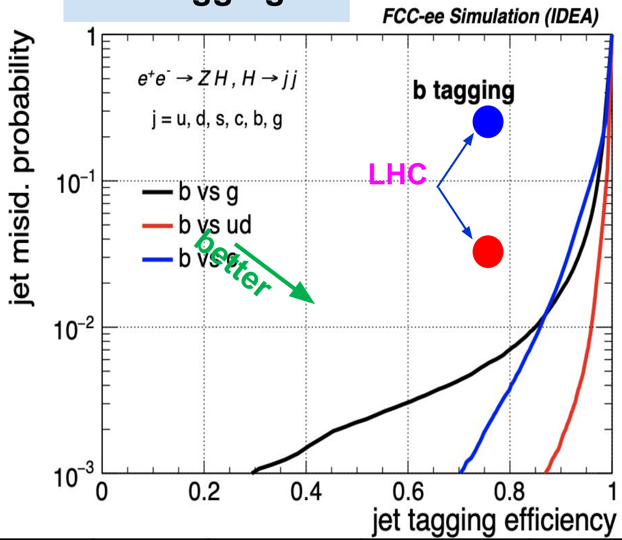
ACKS : Michele Selvaggi, Gregor Krzmanc, Jan Kieseler, Philipp Zehetner

www.cern.ch


FCC
SUISSE
FRANCE
Genève
Annecy
LHC

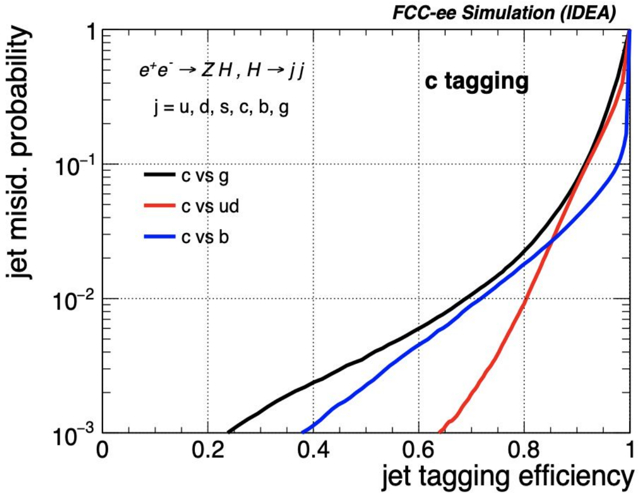
Results

b-tagging



WP	Eff (b)	Mistag (g)	Mistag (ud)	Mistag (c)
Loose	90%	2%	0.1%	2%
Medium	80%	0.7%	<0.1%	0.3%

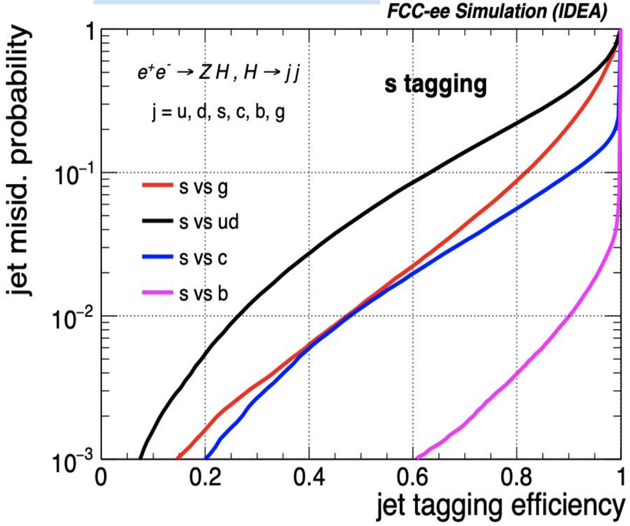
c-tagging



WP	Eff (c)	Mistag (g)	Mistag (ud)	Mistag (b)
Loose	90%	7%	7%	4%
Medium	80%	2%	0.8%	2%

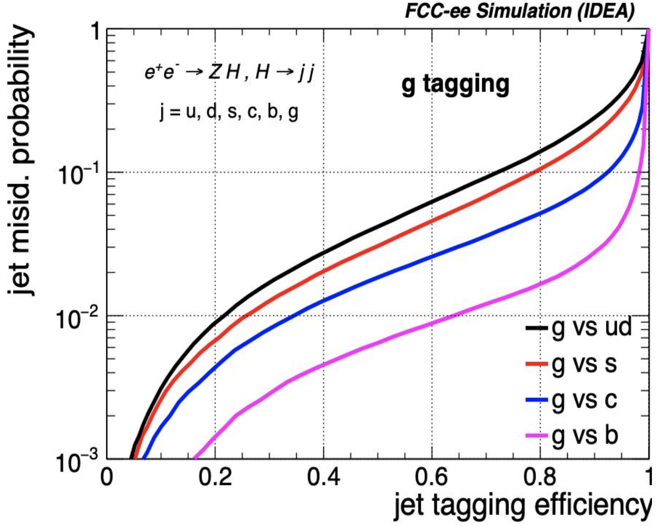
Results

s-tagging



WP	Eff (s)	Mistag (g)	Mistag (ud)	Mistag (c)	Mistag (b)
Loose	90%	20%	40%	10%	1%
Medium	80%	9%	20%	6%	0.4%

g-tagging



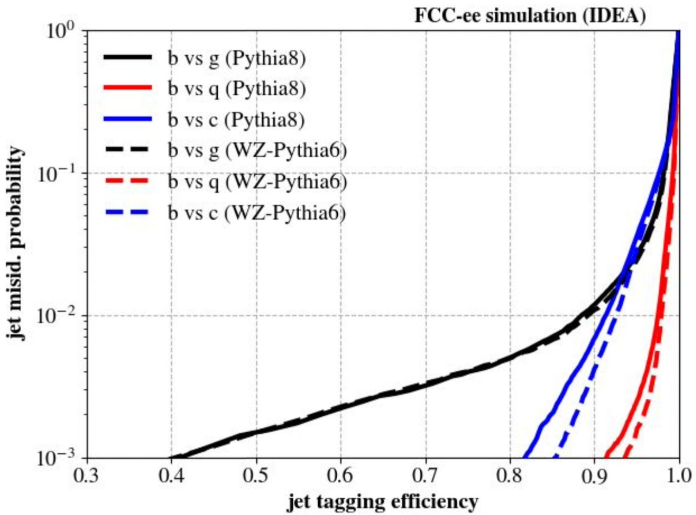
WP	Eff (g)	Mistag (ud)	Mistag (c)	Mistag (b)
Loose	90%	25%	7%	2.5%
Medium	80%	15%	5%	2%

Robustness

ParticleNet-ee trained using ***Pythia 8*** samples

- tested on ***Pythia 8*** [solid lines]
- tested on ***WZ-Pythia6*** [dashed lines]

b-tagging



gluon-tagging