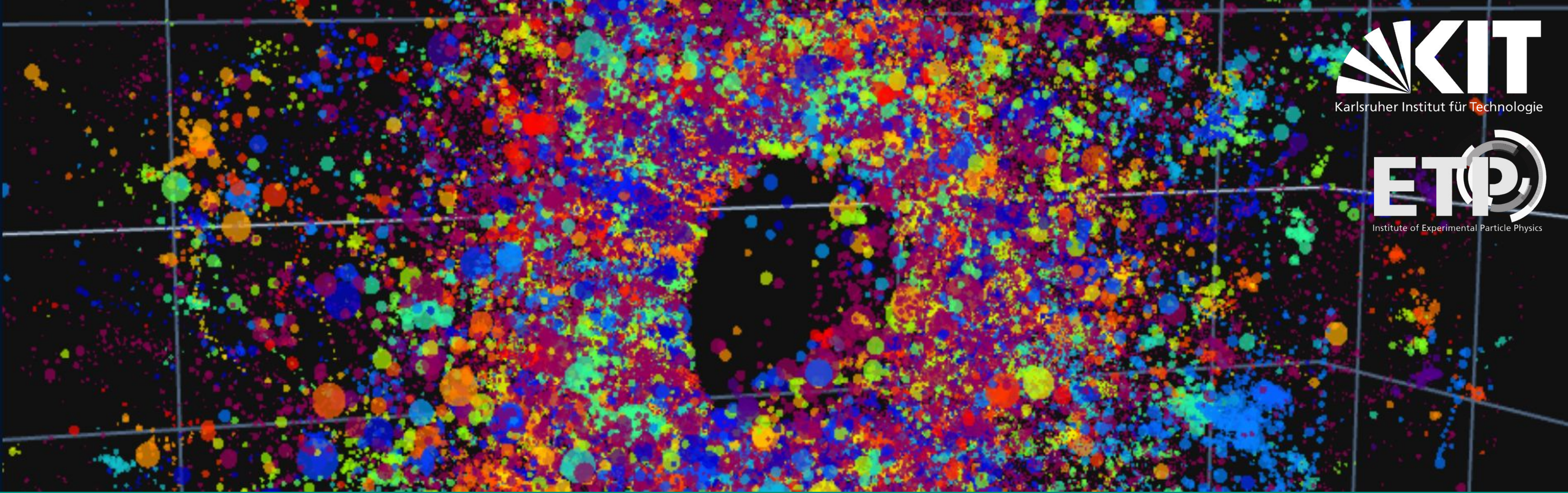


Karlsruher Institut für Technologie

Institute of Experimental Particle Physics

Towards Robust End-to-End Machine Learning Particle Flow

Thesis Presentation - ILD Analysis/Software Meeting

Katharina Schäuble, Alessandro Brusamolino, Ulrich Einhaus, Jan Kieseler

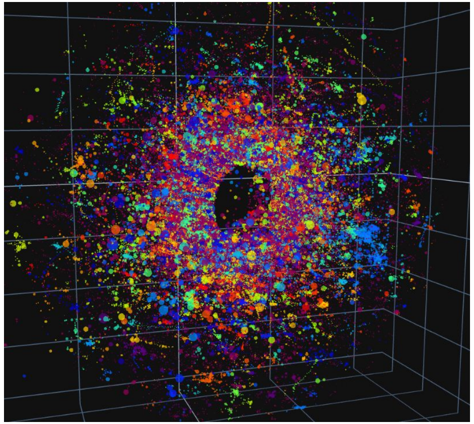
20/05/26

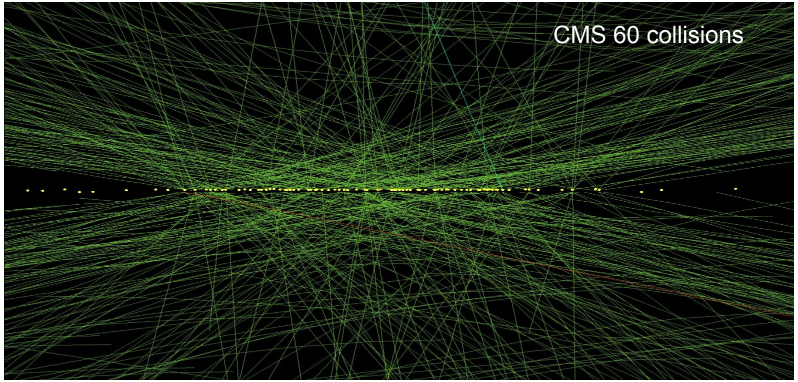
Why do we need ML-based Reconstruction?

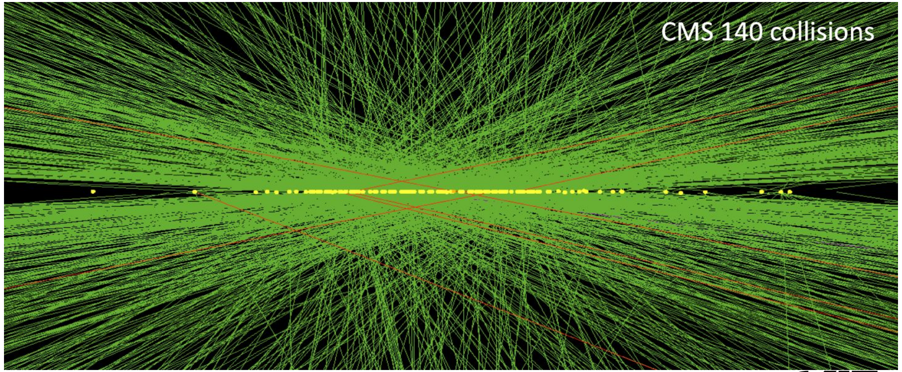
Challenges for Reconstruction Algorithms

- **New detector design** with higher granularity and more complex geometry
- HL-LHC challenges reconstruction by increased **pile-up**
  - From ~60 to ~140 collisions per bunch crossing
- **Computing challenges** due higher rates and more complex events

→ Need for new reconstruction algorithms

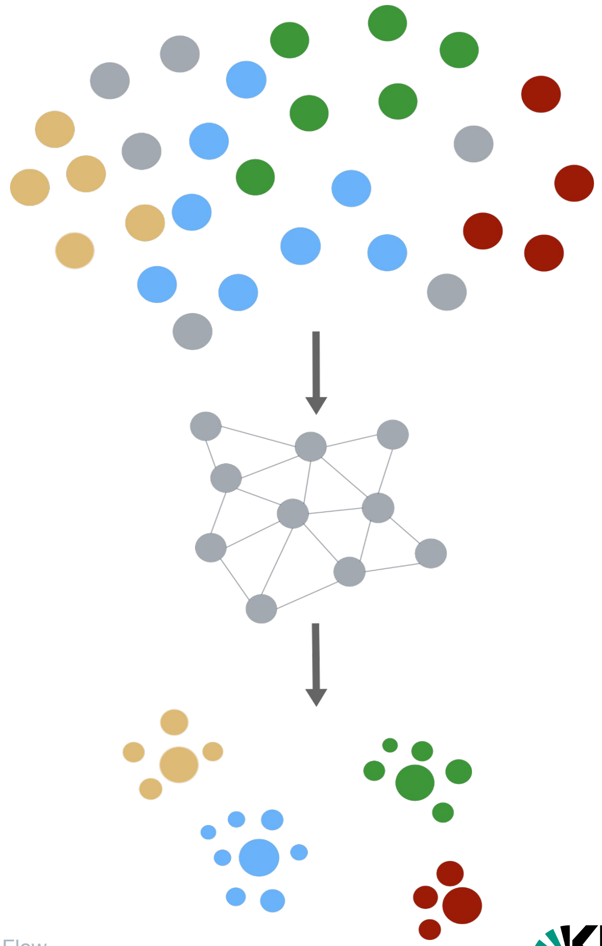






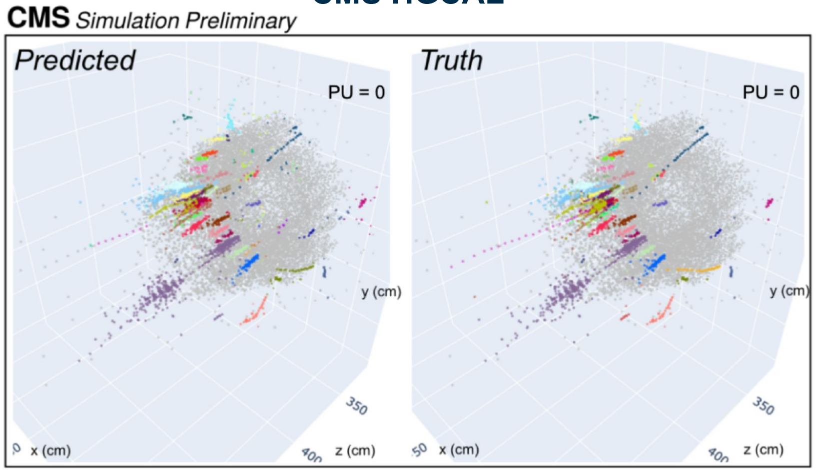
End-to-End Reconstruction

- State of the art ML-based reconstruction using GravNet and Object Condensation arxiv:2204.01681
- Traditional approach uses at least two steps: clustering first, then regression for object properties
- End-to-end reconstruction: from detector hits directly to particle candidates with respective properties



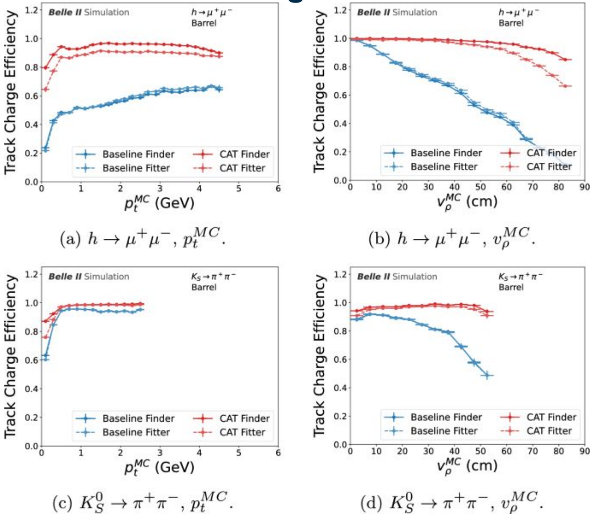
Applications of End-to-End ML-based Reconstruction

CMS HGCAL



CERN-CMS-DP-2022-004

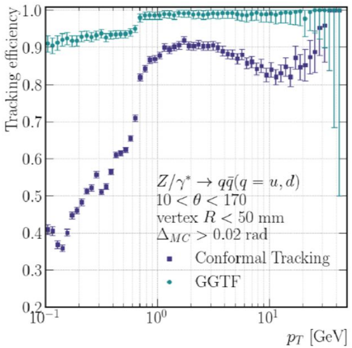
Belle II Tracking - CAT Finder



(a) $h \to \mu^+\mu^-$, p_t^{MC}.
(b) $h \to \mu^+\mu^-$, v_ρ^{MC}.
(c) $K_S^0 \to \pi^+\pi^-$, p_t^{MC}.
(d) $K_S^0 \to \pi^+\pi^-$, v_ρ^{MC}.

L. Reuter, et al., arxiv:2411.13596

…and many more
S.R. Qasim,et. al., arXiv:1902.07987
J.Kieseler, arxiv:2002.03605, EPJC
S.R. Qasim, N. Chern. J. Kieseler., arXiv:2204.01681,
S. Battarchaya, et et al., arXiv:2203.01189,
S. Qasim, K. Long, J.K., et al., arXiv:2106.01832
I. Iiyama, .. J.K., et al, arXiv:2008.03601



FCC
CLD Tracking

D.Garcia, et al.
10.1051/epjconf/202533701125

Reconstruction algorithm already proven successful across various applications and is adaptable for future detector concepts

Towards Robust End-to-End Machine Learning Particle Flow

- Reconstruction algorithm successfully adapted to various applications

→ *How well does the algorithm work?*

→ *How can GravNet modification improve reconstruction performance?*

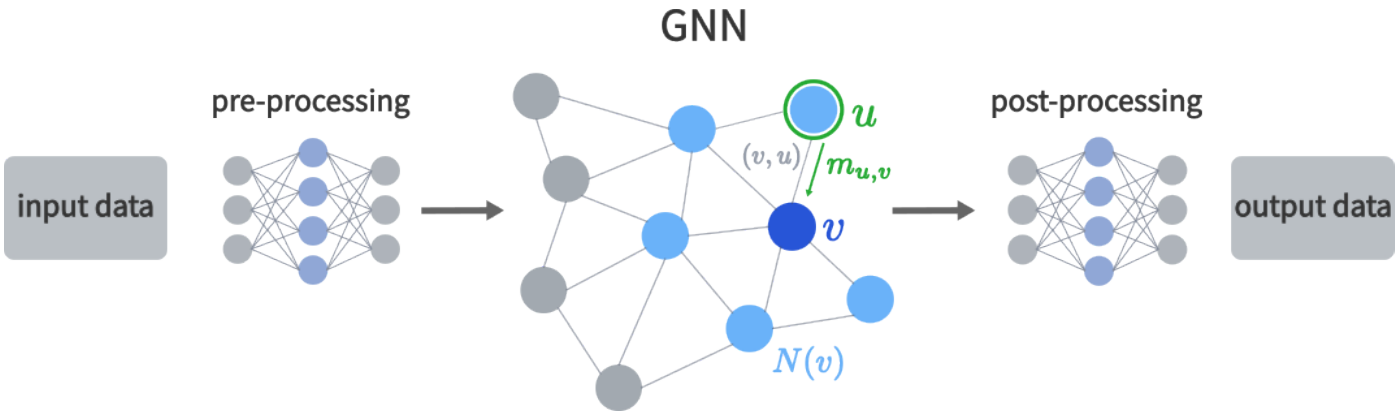
→ *How to further enhance algorithm robustness?*

How does End-to-End Reconstruction work?

GravNet & Object Condensation

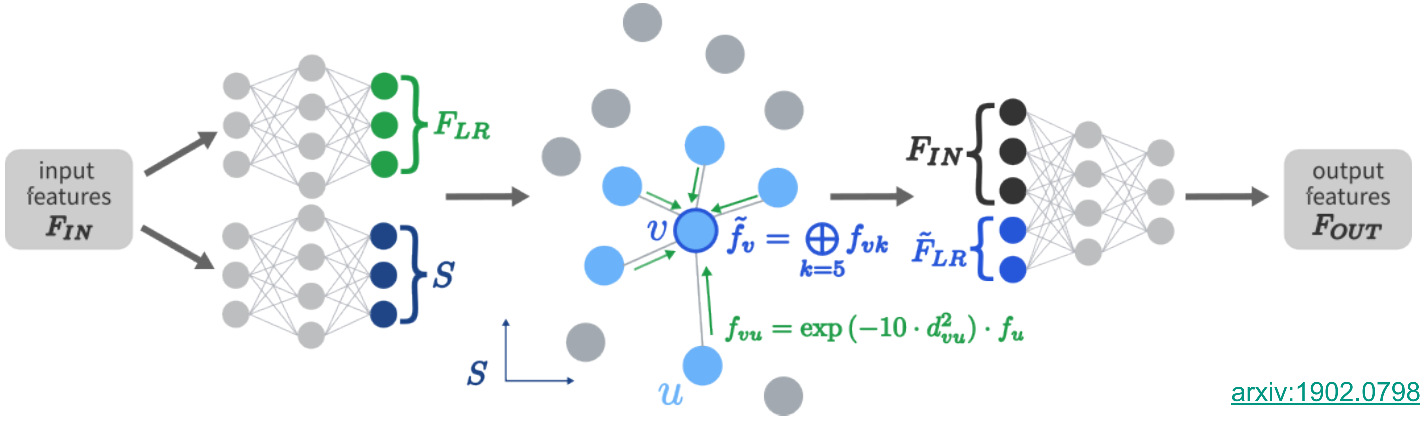
A Brief Introduction to Graph Neural Networks

- GNNs apply Neural Networks to graph-like structures
  - e.g. detector hits as point clouds
- Graph G(V,E) consisting of a set of vertices V and edges E connecting them
  - Both may be attributed with features (e.g. hit energy, timing information)
- Information exchange by aggregating messages from neighboring nodes ⟶ **message passing**
- Connecting all vertices is computationally expensive
  - edges can be learned instead → GravNet



GravNet

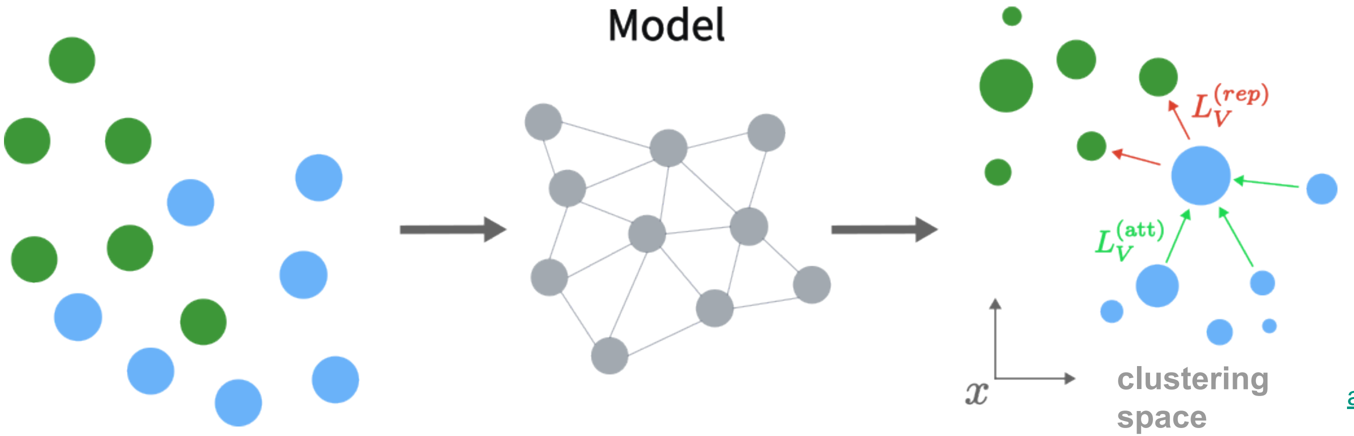
- GravNet: dynamic distance-weighted GNN
- Learnable embedding space and features for building graph topology
- Builds edges dynamically between nearest neighbors
  - Local exchange of information
- Use Euclidean distance between nodes as weights for message passing
  - Distance with Gaussian potential as edge weight ⟶ rotation invariant in GravNet space
  - Combine features of nearest neighbors with different aggregator functions



arxiv:1902.07987

Object Condensation

- Attractive potential for all hits belonging to an object, repulsive potential towards all other hits
- Hit-wise condensation score β for best object representation
- Beta loss to encourage one hit per object ⟶ **condensation point**
  - Gather all hits around condensation point



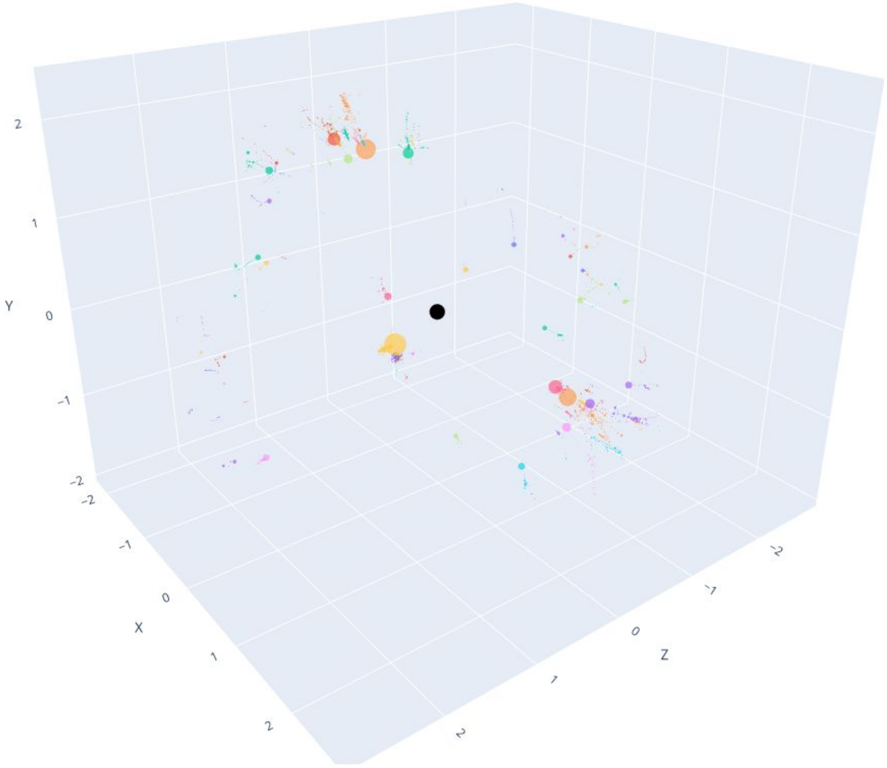
arxiv:2002.03605

What can the Reconstruction Algorithm be applied to?

ILD Detector

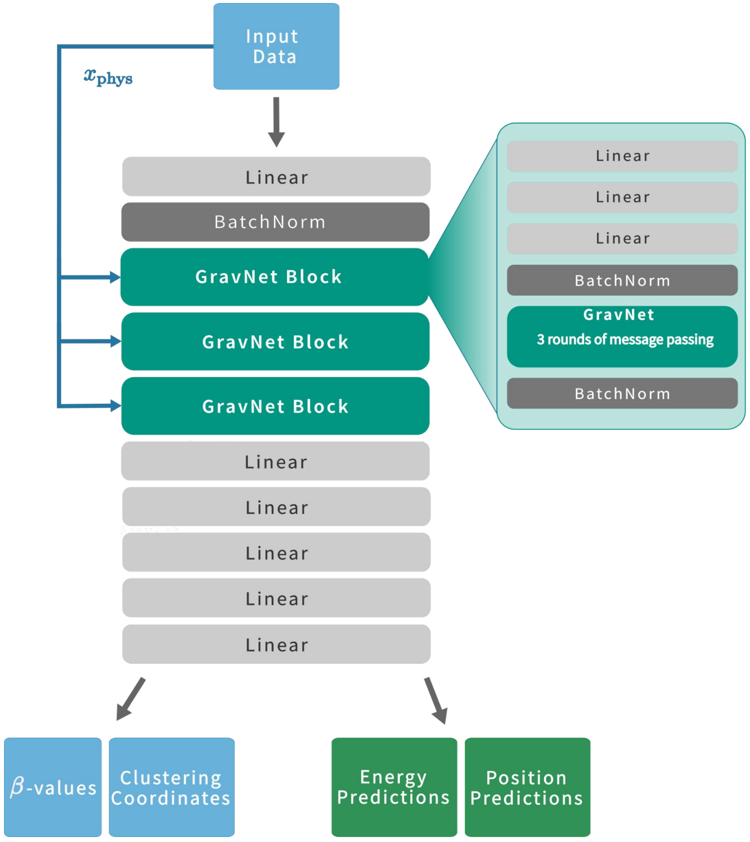
ILD Dataset

- Proposed ILD design optimized for Particle Flow analyses
- Dataset of $e^+e^- \longrightarrow q\bar{q}$ events at 250 GeV
- Truth definition: reconstruction targets are defined at the boundary between the tracking system and calorimeter
  - If multiple showers deposited energy in one cell assign to shower with highest deposition
  - for each shower assign unique index
- Hits in calorimeter and include track as additional hit at boundary between tracker and calorimeter



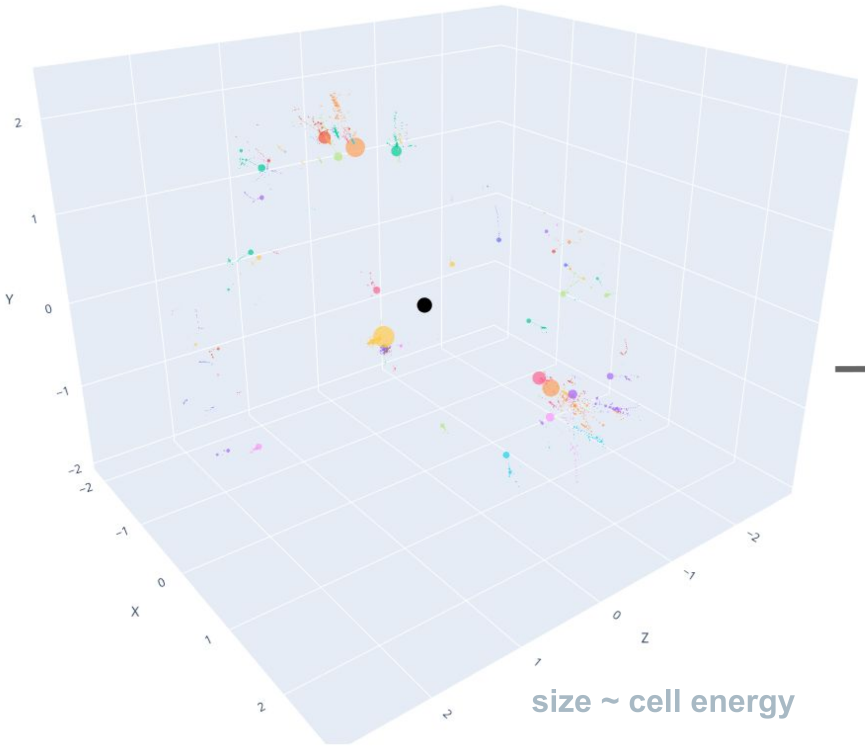
Model and Training

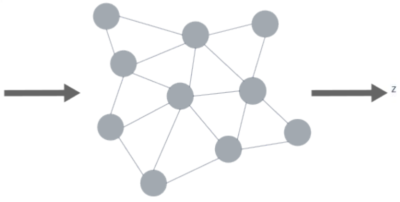
- Model inputs: x, y, z, r_{xy} , θ, η, E, *isTrack* (calo or tracker hit)
- Model outputs: coordinates in cluster space, β scores, energy predictions, position predictions
- Losses:
  - Object Condensation losses (repulsive and attractive potential losses, β loss)
  - Energy loss $L_E = \ln\left(\frac{(E_{true} - E_{pred})^2}{E_{true}} + 1\right)$
  - Position loss $L_{pos} = \sum_i^3 (x_{true}^{(i)} - x_{pred}^{(i)})^2$



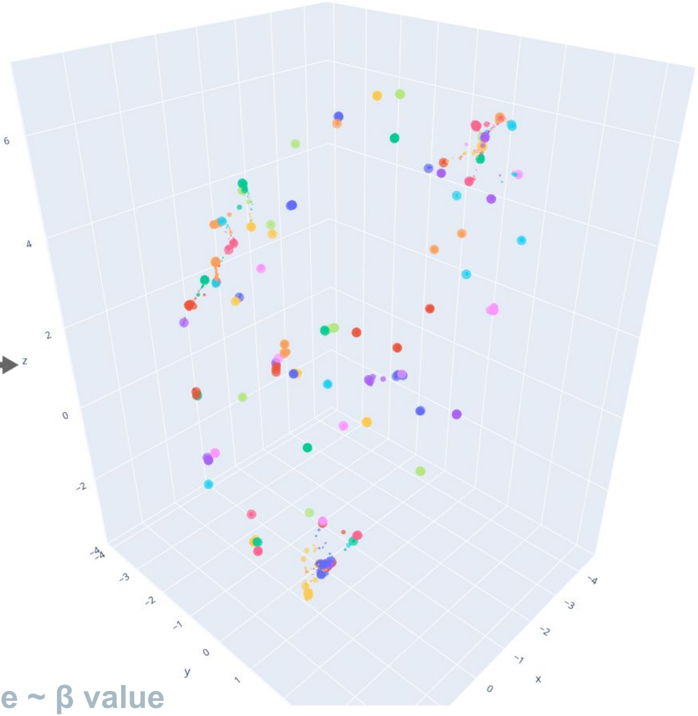
Clustering Results

Model Input





Model Output in Clustering Space

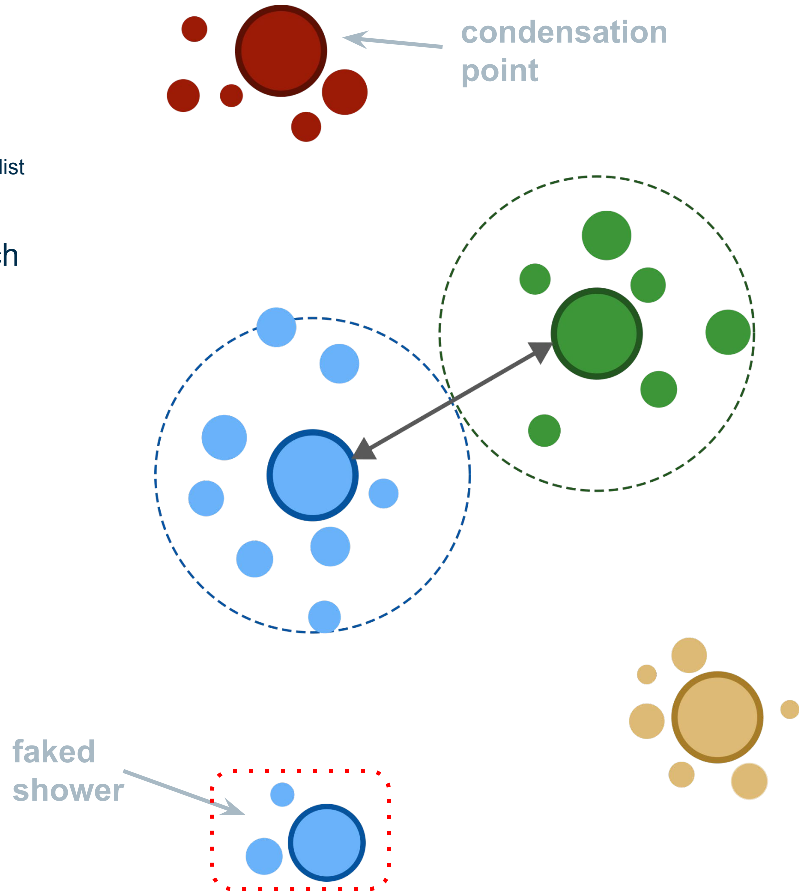


size ~ cell energy

size ~ β value

Clustering Results

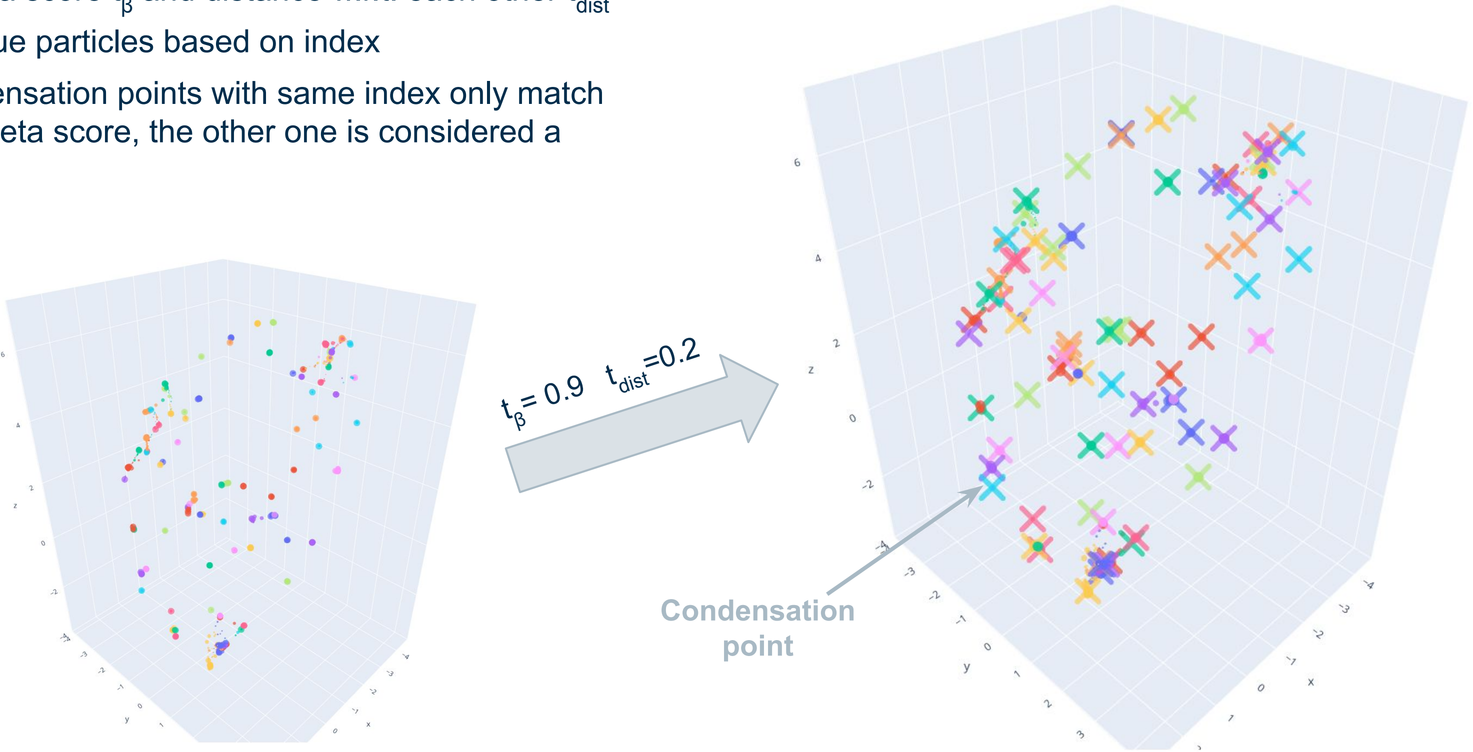
- Select condensation points from clustered hits
  - Thresholds on beta score t_β and distance w.r.t. each other t_{dist}
- Match predicted to true particles based on index
  - For multiple condensation points with same index only match one with highest beta score, the other one is considered a faked shower



Clustering Results

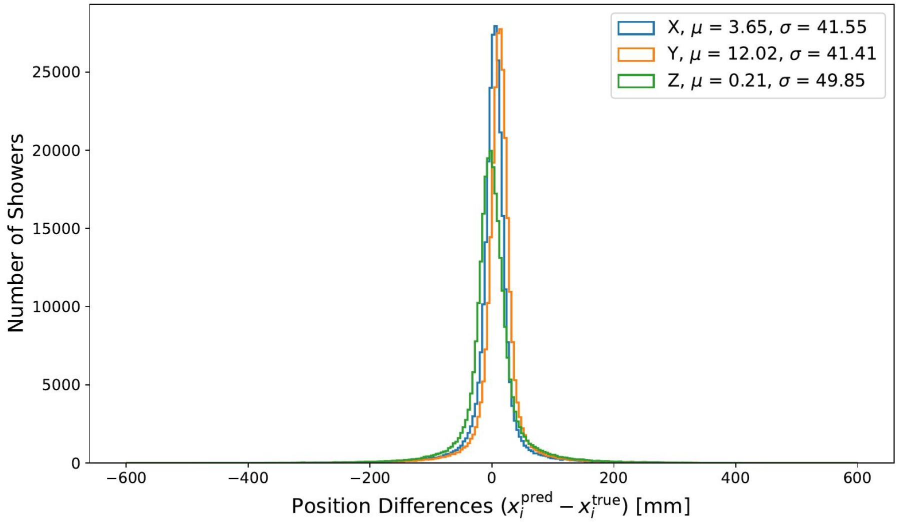
- Select condensation points from clustered hits
  - Thresholds on beta score t_β and distance w.r.t. each other t_{dist}
- Match predicted to true particles based on index
  - For multiple condensation points with same index only match one with highest beta score, the other one is considered a faked shower

Clustering Space



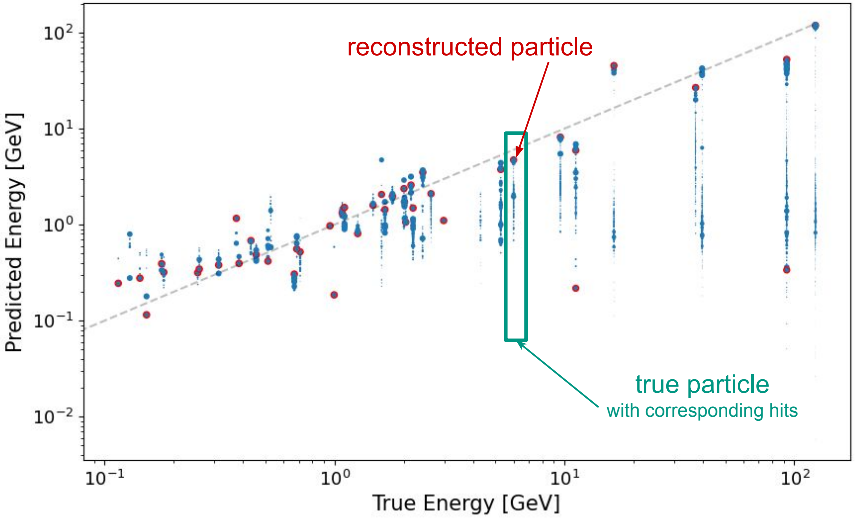
Reconstruction Performance

- Select condensation points above minimal beta score t_β and distance w.r.t. each other t_{dist}
- Match predicted to true particles based on index
  - If multiple showers with same index are reconstructed only match one with highest beta score, the other one is considered a faked shower
- Position predictions yield resolution ~O(cm)



Reconstruction Performance

- Select condensation points above minimal beta score t_β and distance w.r.t. each other t_{dist}
- Match predicted to true particles based on index
  - If multiple showers with same index are reconstructed only match one with highest beta score, the other one is considered a faked shower
- Position predictions yield resolution ~O(cm)
- Energy predictions:
  - Model tends to predict higher values for low-energy showers and lower energies for energetic showers



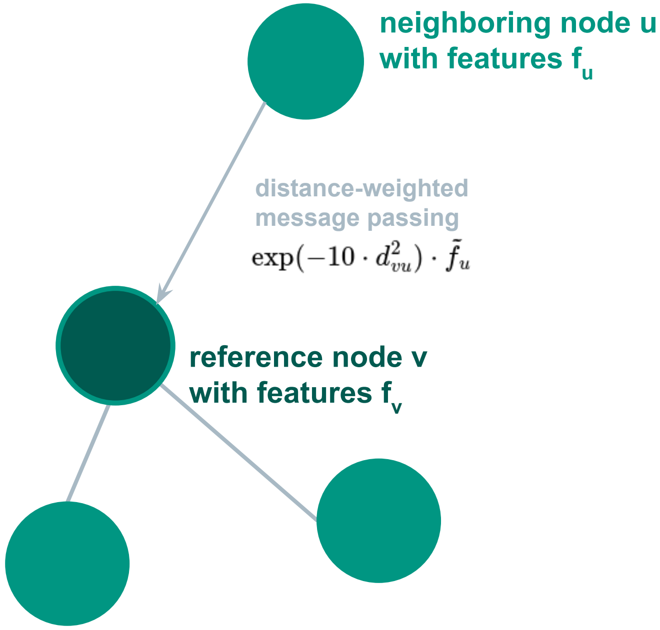
How can we improve GravNet for Reconstruction?

Translation Invariance & Self-Attention

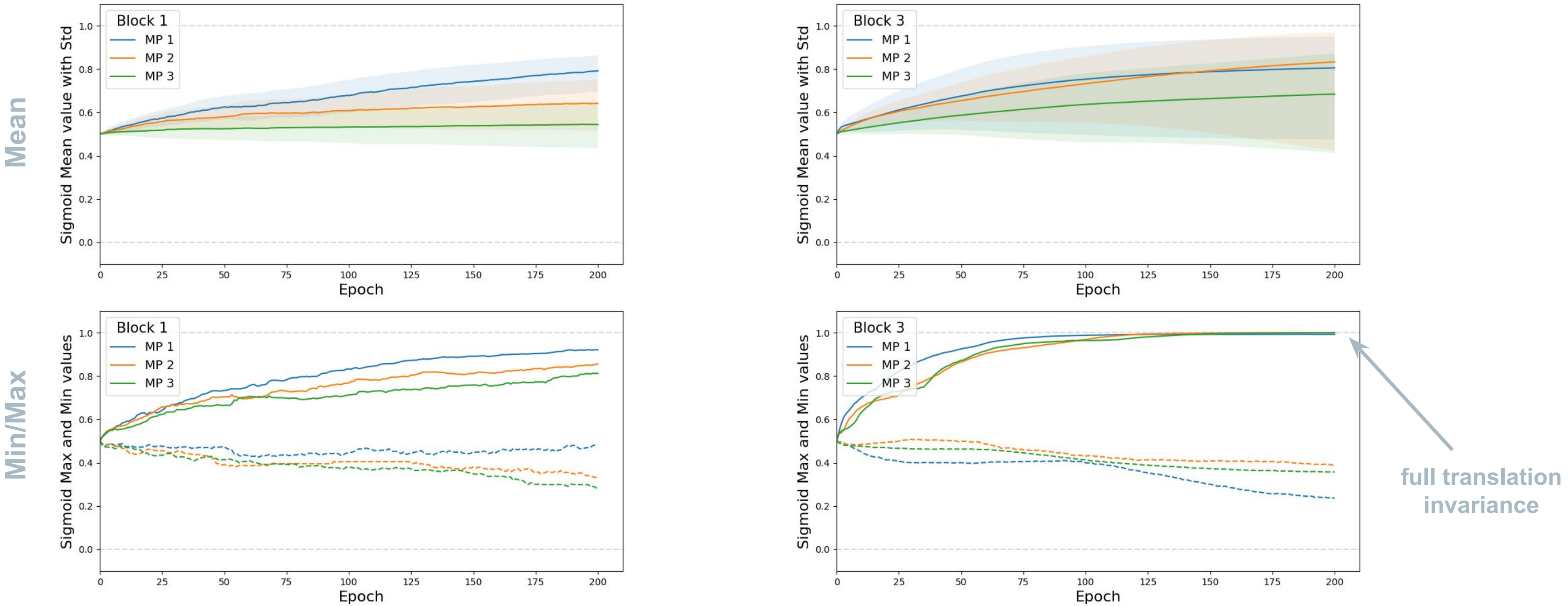
Translation Invariant GravNet

- Showers show similar structure independent of detector location
  - Embed this into network architecture
- Make message passing translation invariant in the GravNet feature space
  - Only pass relative information
- Distance-weighted message passing $\exp(-10 \cdot d_{vu}^2) \cdot \tilde{f}_u$
  - 'Standard' $\tilde{f}_u = f_u$
  - Translation invariant $\tilde{f}_u = f_u - f_v$
- Make translation invariance learnable for the model by introducing it as a parameter $\tilde{f}_u = f_u - \mathrm{sigmoid}(a) \cdot f_v$
  - Apply sigmoid to keep parameter between [0,1]
  - Let model choose between full translation invariance and none

vector



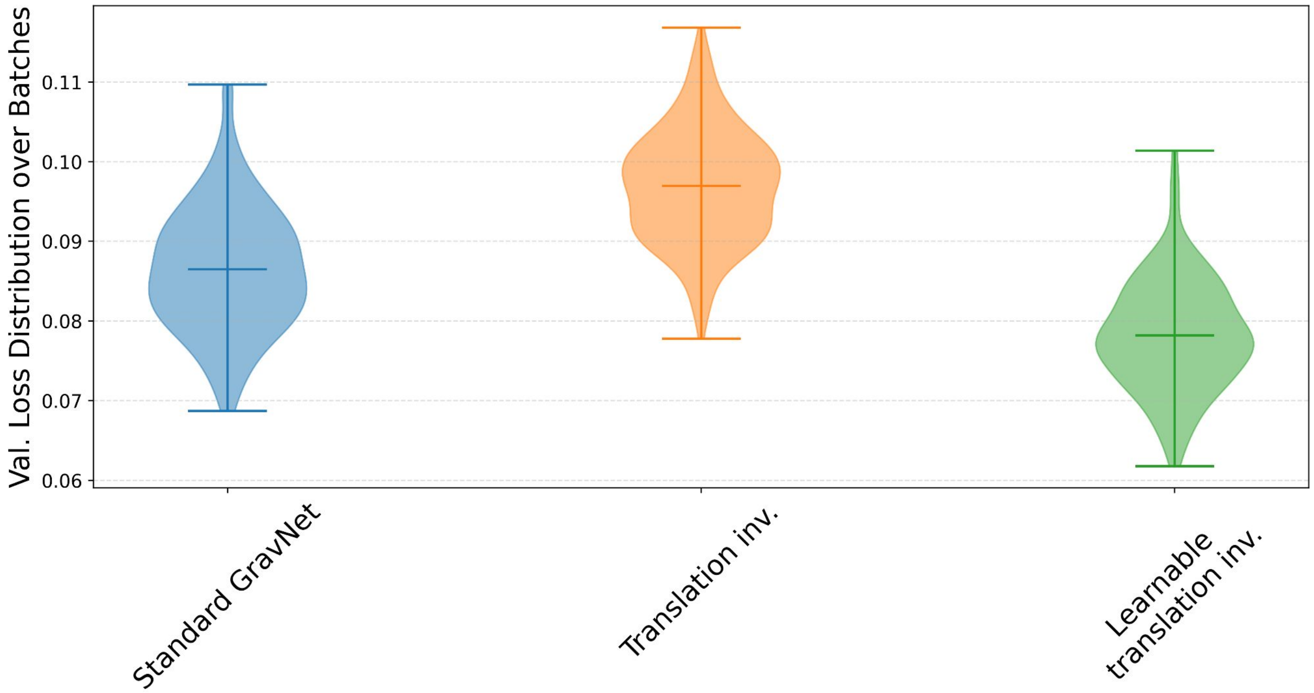
Learnable Translation Invariance



- Learnable translation invariance parameter introduced for every message passing step (of each GravNet block) individually

→ GravNet can explicitly learn translation invariance, but is not optimal choice for all parameters

Learnable Translation Invariance - Comparison

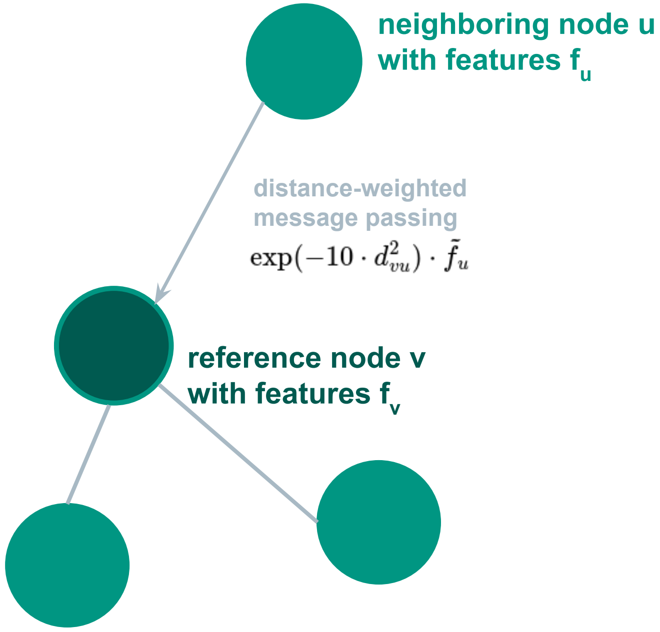


Introducing (learnable) translation invariance yields no significant improvement

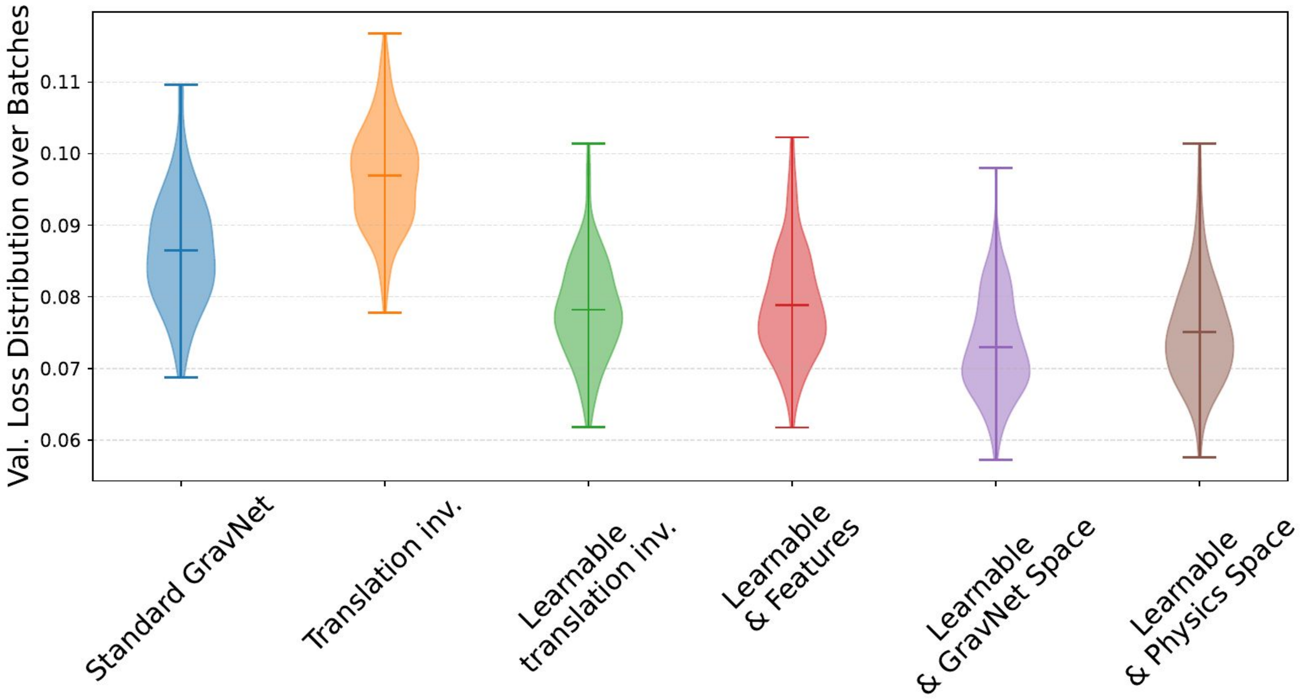
Translation Invariant GravNet

- Make message passing translation invariant by only passing relative information
- Distance-weighted message passing $\exp(-10 \cdot d_{vu}^2) \cdot \tilde{f}_u$
    - 'Standard' $\tilde{f}_u = f_u$
    - Translation equivariant $\tilde{f}_u = f_u - f_v$
    - Learnable parameter $\tilde{f}_u = f_u - \text{sigmoid}(a) \cdot f_v$ vector
    - Depending on features $\tilde{f}_u = f_u - \text{sigmoid}(a \cdot f_v) \cdot f_v$
    - Depending on GravNet space $\tilde{f}_u = f_u - \text{sigmoid}(A \cdot s_v) \cdot f_v$
    - Depending of physical space $\tilde{f}_u = f_u - \text{sigmoid}(A \cdot x_v) \cdot f_v$

matrix



Learnable Translation Invariance - Full Comparison

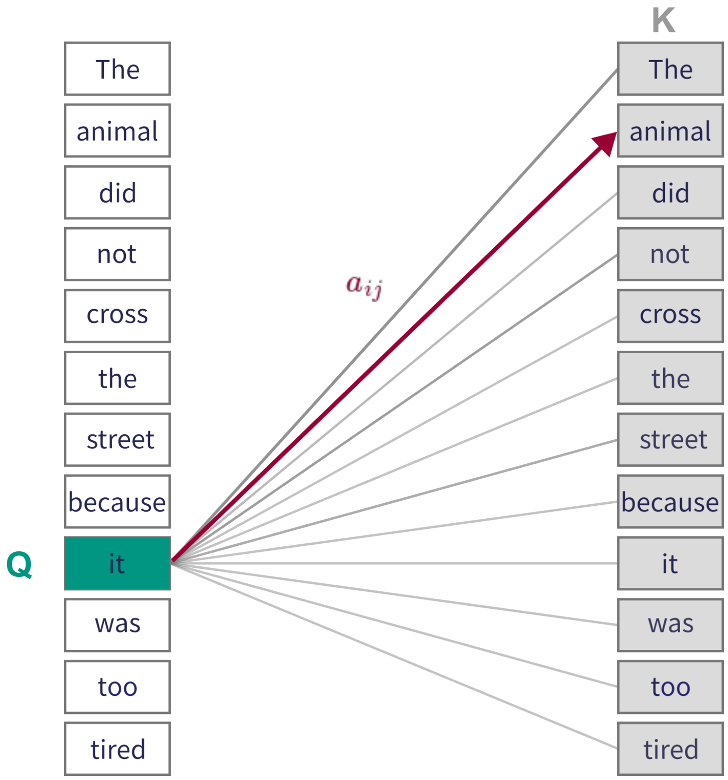


➡ Introducing (learnable) translation invariance yields no significant improvement

A Brief Introduction to Attention

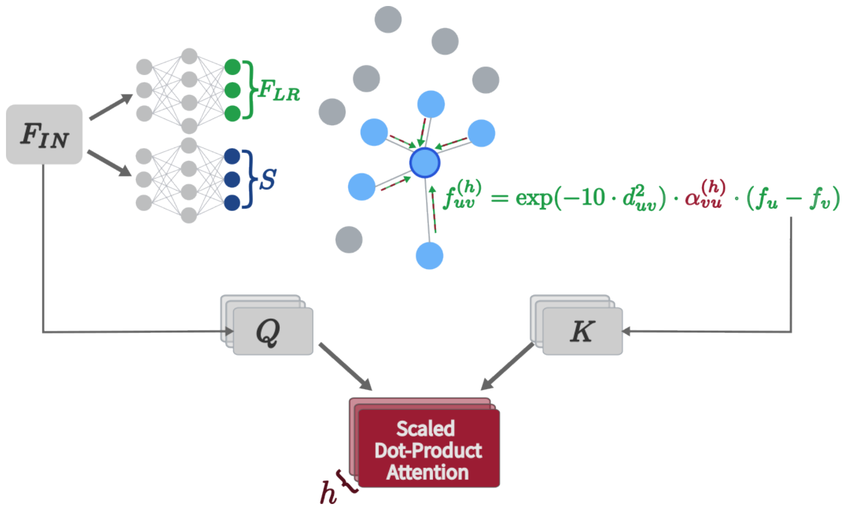
- Attention is a powerful operation that forms basis of transformers like ChatGPT
- Goal: determine importance of each element in a sequence relative to all other elements
- Self-attention: assign learnable weights based relevance of elements to each other
- Originally introduced for Natural Language Processing and used in Transformer architecture (arxiv:1706.03762)
- Attention score determined is by key and query pairs
  - Scaled dot-product attention $a(Q, K) = \frac{Q \cdot K}{\sqrt{d_{\text{att}}}}$
- GravNet's distance-weighted message passing as form of attention

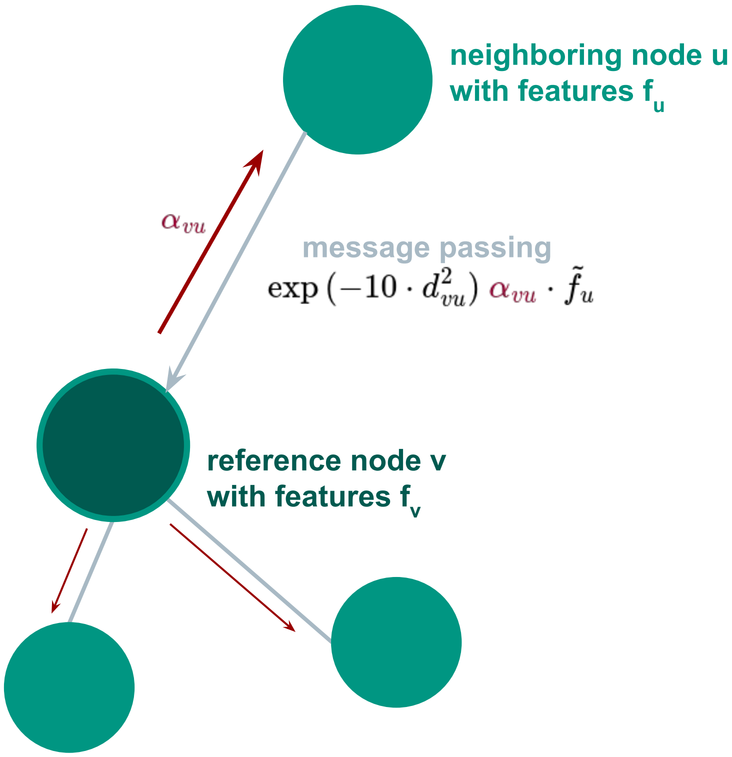
$$\exp(-10 \cdot d_{vu}^2) \cdot \tilde{f}_u$$



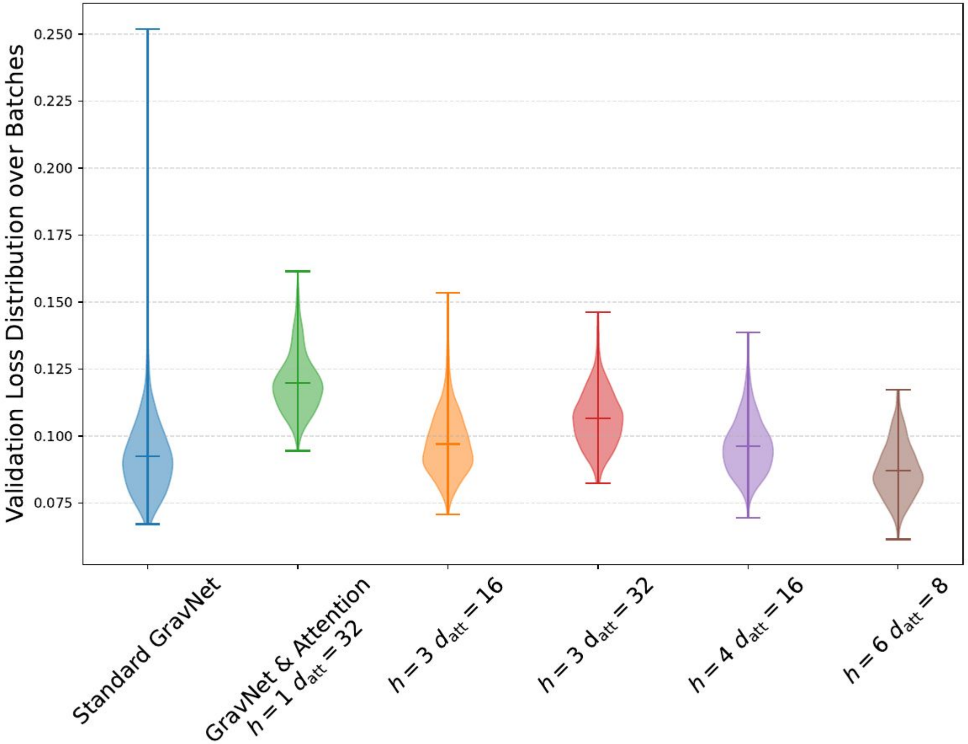
GravNet with Self-Attention

- Combine distance weight with learned attention score
  - Each node attending to its neighbors
- Compute multiple representations of queries and keys in parallel
  - Multi-head attention





GravNet with Attention



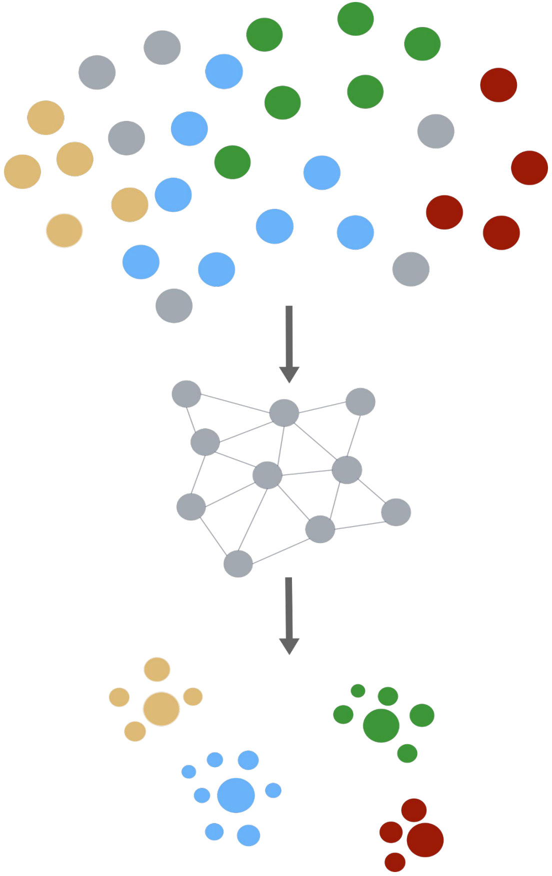
Standard GravNet yields comparable results across different combinations of heads and attention dimensions - and is significantly less resource-demanding!

How can we improve GravNet for reconstruction?

- GravNet-based algorithm can be applied to highly-granular detector concepts like ILD
- No significant benefit from explicit translation invariance
- GravNet proves sufficiently powerful; with self-attention increasing resource demands but offering no significant improvement (so far)

- Strictly local information exchange makes GravNet generalisable

➡ *How can the algorithm's robustness be improved?*



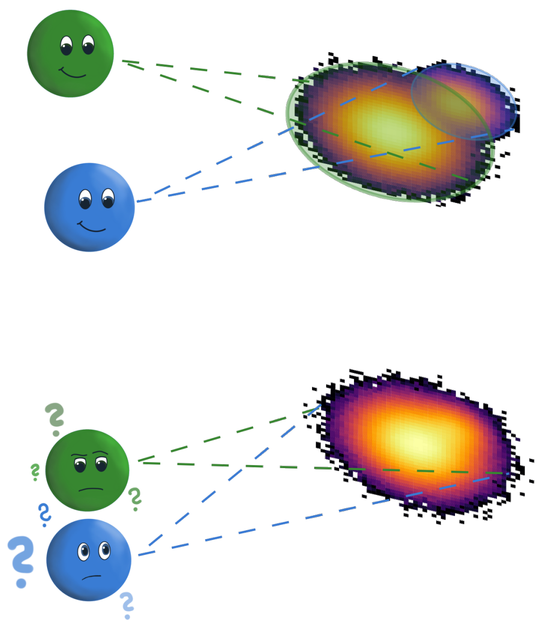
How can we improve the Reconstruction Target?

Truth Definition and Merging

Truth Definition

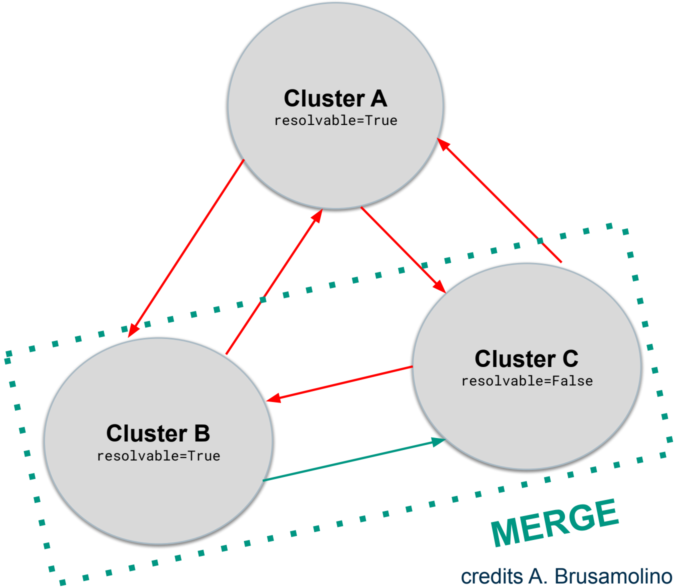
- Need for a robust reconstruction algorithm
  - Reliance on prior assumptions reduces ability to generalise to different physics scenarios
- Truth information is crucial for reconstruction algorithms
  - Impact on performance and development of reconstruction algorithms
- Define truth that embeds detector constraints, relying on what we consider resolvable by the detector
- Redefining reconstruction targets reduces sample bias and improves algorithm robustness

→ Merged truth information

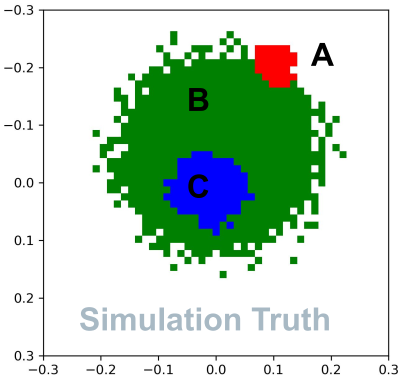


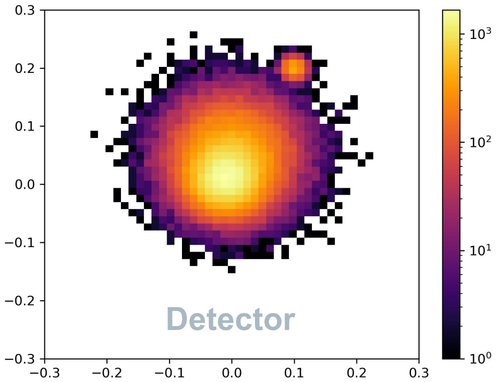
Shower Merging Algorithm

- Resolvable showers: high amount of energy that is not shared with other showers
- Merge non-resolvable showers if they share a significant amount of energy
- Compute resolvability and connection scores based on hits and energy fractions
- Build a graph with showers as node and edges
  - Only edges between non-resolvable showers and from resolvable to non-resolvable showers are valid for merging
  - Sorting edges by connection type ⟶ ensures IRC safety



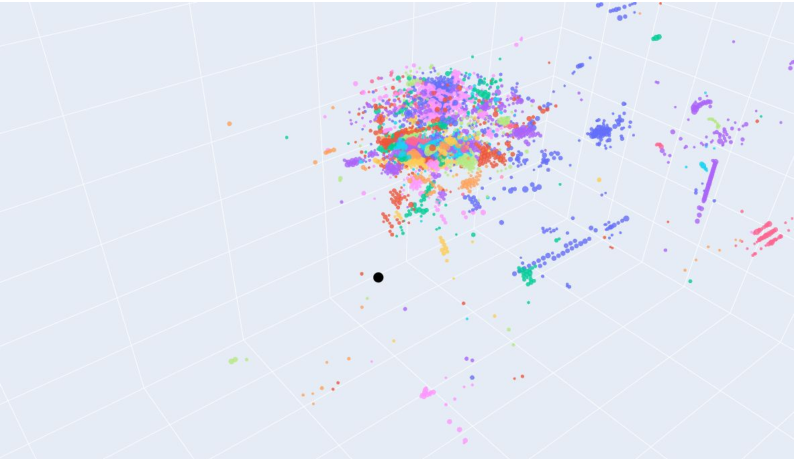
credits A. Brusamolino





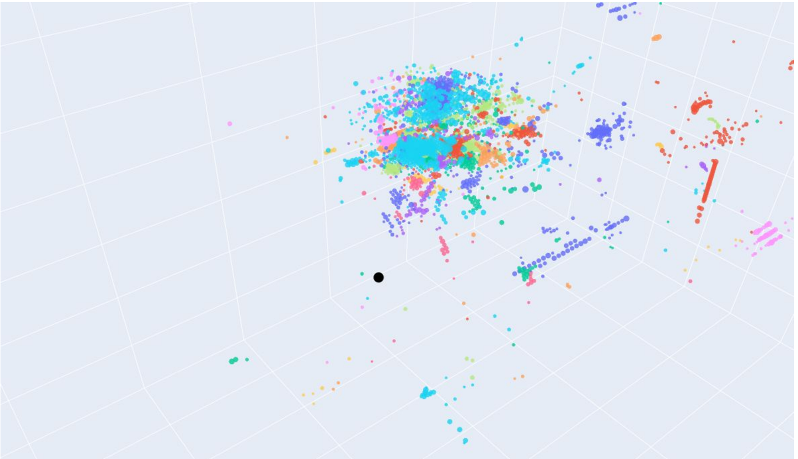
DICE Dataset

- DICE: GEANT4 based simulated detector inspired by CMS calorimeters, tracker only serves as interaction material
- To evaluate merging algorithm create “simple” single jet dataset from top quarks
- Train models on datasets with different resolvability thresholds
  - Resolvability score = average cell energy fraction of a shower



unmerged event

Shower Merging

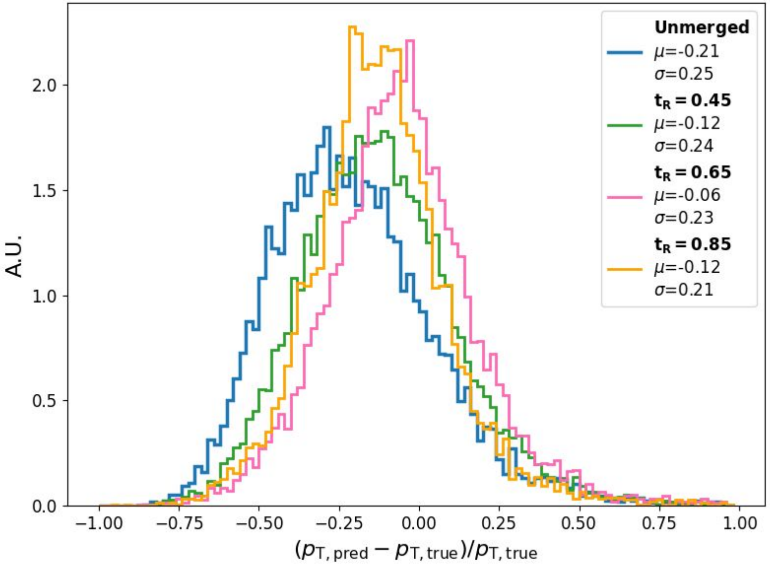


high resolvability threshold (0.85)

Merging Evaluation

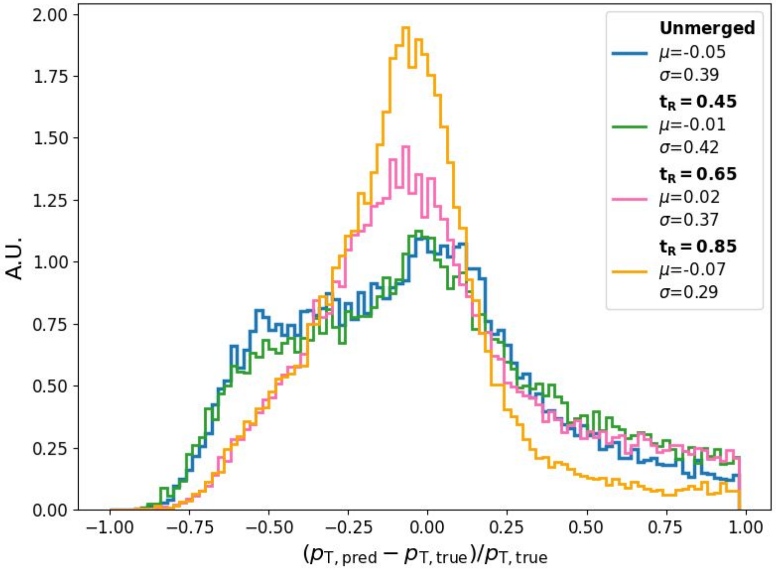
- Evaluate reconstruction performance on jet metrics
  - Not affected by shower-level truth definitions
- Compare performance of merged truth definitions with different resolvability thresholds to unmerged truth

→ Merging consistently outperforms models trained on unmerged truth



Generalisation Capabilities Beyond Training Data

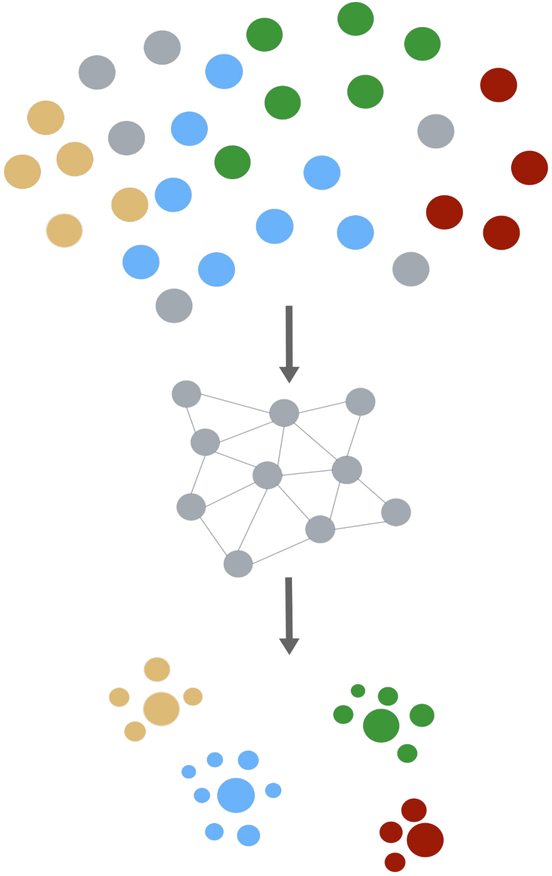
- Merging algorithm aims for improved model robustness when extrapolating to unseen physics
- Evaluate trained models on different physical processes
  - Studied tau single-jet events
- Models trained on merged truth definition can outperform unmerged truth



→ Detector-aware truth definition improves model robustness

Thesis Summary

- Reconstruction algorithm widely applicable and can be used for new detector designs → ILD detector
- GravNet architecture does not seem to profit from extensions through translation invariance or resource-demanding self-attention (on the ILD dataset)
- Generalisability can be enhanced by detector-aware truth definition, yielding improved physics performance



Thank you for your attention!