

# Hierarchical Graph Neural Networks for Particle Track Reconstruction

Ryan Liu<sup>1, 2</sup>, Paolo Calafiura<sup>1</sup>, Steven Farrell<sup>1</sup>, Xiangyang Ju<sup>1</sup>,  
Daniel Thomas Murnane<sup>1</sup>, Tuan Minh Pham<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA

<sup>2</sup>University of California, 366 Physics North, MC 7300, Berkeley, CA 94720, USA

E-mail: ryanliu@lbl.gov

**Abstract.** We introduce a novel variant of GNN for particle tracking—called Hierarchical Graph Neural Network (HGNN). The architecture creates a set of higher-level representations which correspond to tracks and assigns spacepoints to these tracks, allowing disconnected spacepoints to be assigned to the same track, as well as multiple tracks to share the same spacepoint. We propose a novel learnable pooling algorithm called GMPool to generate these higher-level representations called “super-nodes”, as well as a new loss function designed for tracking problems and HGNN specifically. On a standard tracking problem, we show that, compared with previous ML-based tracking algorithms, the HGNN has better tracking efficiency performance, better robustness against inefficient input graphs, and better convergence compared with traditional GNNs.

## 1. Introduction

In the upcoming High Luminosity Phase of the Large Hadron Collider (HL-LHC) [1, 2], the average number of inelastic proton-proton collisions per bunch  $\langle\mu\rangle$  (pile-up) is expected to reach 200 in the new silicon-only Inner Tracker (ITk). This will pose a significant challenge in track reconstruction due to the limited computational resources [3]. Since charged particle reconstruction (“particle tracking”) dominates the CPU resources dedicated to event offline reconstruction, a new and efficient algorithm for event reconstruction becomes an urgent need. The HEP.TrkX project [4] and its successor the Exa.TrkX project [5] have studied Graph Neural Networks (GNNs) for charged particle tracking, and excellent performance on the TrackML dataset [6] has been demonstrated in Refs. [7, 8] and more recently on ITk simulation, referred to as GNN4ITk [9].

However, despite the success of GNN-based tracking algorithms, there is much in these techniques that can be improved. In particular, GNN tracking suffers from two types of errors: (1) **broken tracks** (one true track split into multiple segments) and (2) **merged tracks** (a track contains spacepoints of multiple particles). In its nature, the GNN4ITk tracking pipeline prototype [9] is a process of reducing the number of edges; starting from a graph constructed for example by a multi-layer perceptron (MLP) embedding model, filter MLP and GNN edge classifiers are applied to filter out fake edges (i.e. connecting two spacepoints of distinct particles). Thus, broken tracks are more difficult to remove than merged tracks since they can only be resolved by including more edges during the graph construction stage. As such, the pipeline is very sensitive to the efficiency of the graph constructed. Furthermore, the nature of

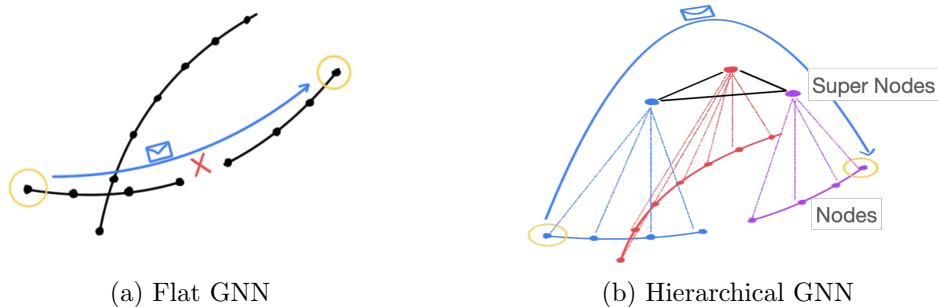


Figure 1: The HGNN can not only shorten the distance between two nodes and effectively enlarge the receptive field but also pass messages between disconnected components

message-passing neural networks [10] utilized in the GNN4ITk pipeline, precludes the passing of information between disconnected components, such as the two ends of a broken track. Broken tracks not only limit the performance of edge-cut-based algorithms but also inhibit the full capability of the message-passing mechanism.

In this paper, we present a novel machine learning model called Hierarchical Graph Neural Network (HGNN)<sup>1</sup> for particle tracking to address the aforementioned problems. Similar to the pooling operation often used in Convolutional Neural Networks (CNN), the HGNN pools nodes into clusters called “super-nodes” to enlarge the “receptive field” of nodes to resolve the problem that a “flat” GNN cannot pass messages between disconnected components. Unlike the case of image processing where pooled pixels are already arranged on a 2D grid, the pooled super-nodes cannot use a graph induced by the original graph since disconnected components will remain disconnected. Thus we propose to utilize a K-nearest-neighbors (KNN) algorithm to build the super-graph among super-nodes to facilitate message passing between super-nodes. Furthermore, the HGNN offers us a new approach to track building, as defining a bipartite matching between nodes (spacepoints) and super-nodes (tracks). We measure the performance of this matching procedure against several baselines and show that it can not only recover broken tracks, but also produces fewer fakes tracks from merging.

## 2. Related Work

### 2.1. The GNN4ITk Pipeline for Charged Particle Tracking

The GNN4ITk pipeline [8, 9] aims to accelerate particle tracking by utilizing geometric deep learning models. The pipeline as implemented can be divided into four steps: firstly, graph construction takes place to build a graph on the input point-cloud. With one possible construction technique, an MLP is trained to embed spacepoints into a high-dimensional space such that spacepoints belonging to the same particle gets closer in space; a fixed radius graph is then built and passed to a “filter” MLP. The filter takes in spacepoint doublets and prunes the graph down by a  $O(10)$  factor in the number of edges. A graph neural network is used to prune the graph further down. Finally, the tracks are built by running a connected components algorithm on the pruned graphs, and ambiguities are resolved by a walk-through algorithm based on topological sorting.

### 2.2. Graph Pooling Algorithms

As discussed in section 1, the pooling algorithm is a crucial piece of the HGNN architecture. Graph pooling has long been studied in the context of graph neural networks as generating graph

<sup>1</sup> The code now available on github

representations require some global pooling operation. Ying *et al.* introduced DiffPool [11], which pools the graph by aggregating nodes according to weights generated by a GNN. DiffPool pools the graph to a fixed number of super-nodes, and the pooled graph has a dense adjacency matrix. Lee *et al.* proposed SAGPool [12], which pools a graph by selecting top-k rank nodes and uses the subgraph induced. However, SAGPool does not support soft assignment, i.e. assigning a node to multiple super-nodes. The granularity is completely defined by the hyperparameter  $k$  and thus also pools to a fixed number of super-nodes. Diehl proposed EdgePool [13], which greedily merges nodes according to edge scores. It is capable of generating a graph that is sparse and variable in size. These pooling algorithms and their features are presented in table 1, along with our proposed pooling technique, described in section 3.1.

Table 1: Graph Pooling Algorithms

Tracking Goal	Feature	DiffPool	SAGPool	EdgePool	GMPool (ours)
Subquadratic scaling	Sparse	✗	✓	✓	✓
End-to-end trainable	Differentiable	✓	✓	✓	✓
Variable event size	Adaptive number of clusters	✗	✗	✓	✓
Many hits to many particles relationship	Soft assignment	✓	✗	✗	✓

### 2.3. Hierarchical Graph Neural Networks

Hierarchical structures of graph neural networks have been studied in the context of many graph learning problems; some of them utilize deterministic pooling algorithms or take advantage of preexisting structures to efficiently create the hierarchy [14, 15, 16, 17, 18], while the others [19, 20, 21] create the hierarchy in a learnable fashion. Compared with solely graph pooling operations [11], by retaining both pooled and original representations one has the capability of simultaneously performing node predictions and learning cluster-level information. Furthermore, as shown in [20], introducing hierarchical structures can solve the long-existing problem of the incapability of capturing long-range interactions in graphs. Empirical results also show that Hierarchical GNNs have better convergence and training stability compared with traditional flat GNNs.

## 3. Model Architecture

In order to build the model, there are several challenges that must be tackled, namely, pooling the graph, message passing in the hierarchical graph, and designing a loss function for such a model. In the following section, we introduce our proposed methods for each of them.

### 3.1. Gaussian Mixture Pooling

In order to provide the features in table 1, we propose a method that leverages the connected components algorithm and Gaussian Mixture Model. The algorithm takes a set of node embeddings as input. The embeddings are then used to calculate edge-wise similarities defined as  $s_{ij} = \tanh^{-1}(\vec{v}_i \cdot \vec{v}_j)$ . We hypothesize that the graph consists of two types of edges, in-cluster edges and out-of-cluster edges. Then, given the distribution of node similarities, we fit a Gaussian Mixture Model (GMM) to obtain the estimation of the in-cluster and out-of-cluster distributions  $p_{in}(s)$  and  $p_{out}(s)$ . An example distribution is plotted in fig. 3b. We then solve for  $s_{cut}$  by  $\ln(p_{in}(s_{cut})) - \ln(p_{out}(s_{cut})) = r$ , where  $r$  is a hyperparameter defining the resolution

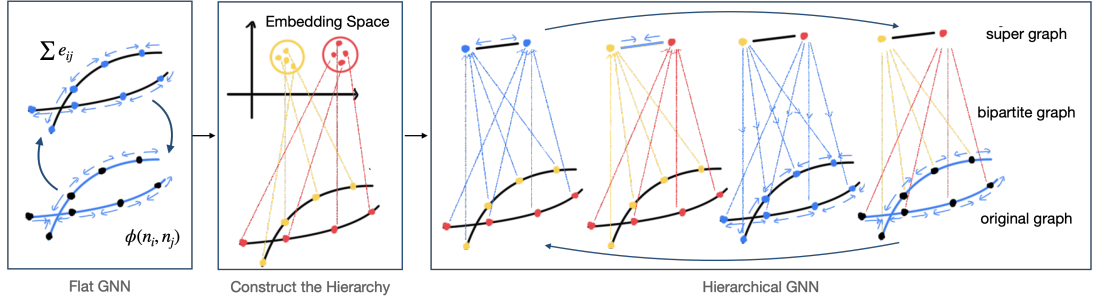


Figure 2: A schematic overview of the HGNN architecture. A flat GNN encoder is used to transform features and embed spacepoints. A pooling algorithm (GMPool) follows to build the hierarchy using the embedded vectors. Finally, hierarchical message passing is applied iteratively to obtain final representations of both nodes and super-nodes.

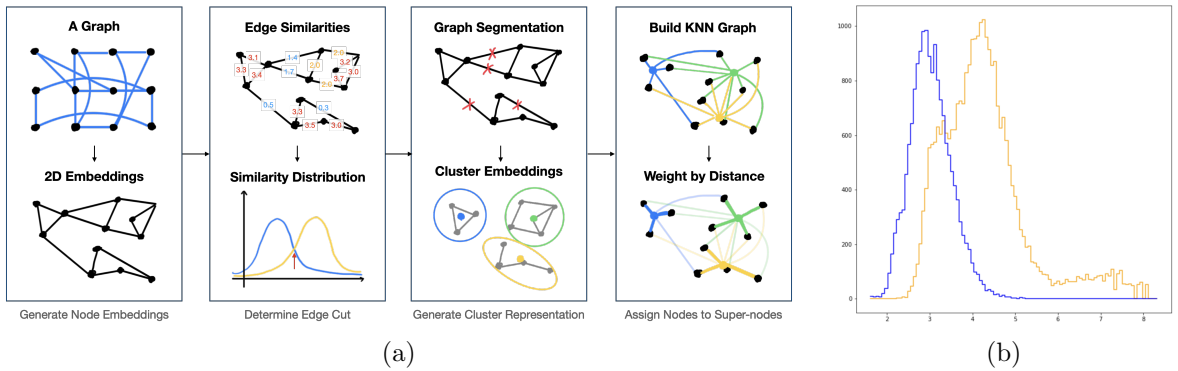


Figure 3: (a): schematic overview of the GMPool algorithm. (b): Distribution of edge similarities. Edges connecting spacepoints of the same particle are colored in yellow and otherwise blue.

of the pooling algorithm. The  $s_{cut}$  value that gives the best separation of in- and out-of-cluster Gaussians is chosen, and edges with scores below this value are cut. The connected components algorithm follows, and the components  $C_\alpha$  of the cut graph are regarded as super-nodes.

To construct super-edges, first super-node embeddings are defined as the centroid of each of the connected components in the embedding space, i.e.  $\vec{V}_\alpha = \frac{\vec{V}'_\alpha}{\|\vec{V}'_\alpha\|_{L_2}}$  where  $\vec{V}'_\alpha = \frac{1}{N(C_\alpha)} \sum_{i \in C_\alpha} \vec{v}_i$ . To connect nodes with super-nodes, similar to the method used in [22], we maintain the sparsity by constructing the bipartite graph with the k-nearest neighbors algorithm. The differentiability can be restored by weighting each of the edges according to the distance in the embedding space, i.e. and  $w_{i\alpha} = \frac{\exp(v_i \cdot V_\alpha)}{\sum_{\alpha \in \mathcal{N}(i)} \exp(v_i \cdot V_\alpha)}$ . Finally, node features are aggregated to be super-node features according to the graph weights. The super-graph construction is identical except that the k-nearest neighbor search has the same source and destination set. Thanks to its edge-cut nature, the GMPool has sub-quadratic time complexity and runs in milliseconds on our graphs.

### 3.2. Hierarchical Message Passing Mechanism

In general, it is possible to stack arbitrarily many pooling layers to obtain a hierarchy of arbitrary height. However, the nature of tracking problems suggests that a spacepoint-particle hierarchy will be sufficient for tracking problems. Thus, the pooling layer in this work is kept to be of two levels. For each of the nodes, we update it by aggregating adjacent edge features, super-nodes

features weighted by bipartite graph weights, and its own features. For each of the super-nodes, it is updated by aggregating super-edge features weighted by super graph weights, node features weighted by bipartite graph weights, and its own features. For edges and super-edges, their update rule is identical to the one used in interaction networks.

### 3.3. Bipartite Classification Loss

At this point, the architecture of HGNN is possible to train on traditional tasks such as node-embedding thanks to GMPool’s differentiability. This feature is useful for apples-to-apples comparisons between flat and hierarchical GNNs under the same training regimes. However, to exploit the full potential of the HGNN, we propose a new training regime for it specifically. The most natural way of doing track labeling with HGNN is to use super-nodes as track candidates. For each of the spacepoint-track pairs (bipartite edges), a score is produced to determine if it belongs to a specific track. A maximum-weight bipartite matching algorithm is used to match tracks to super-nodes to define the “truth” for each of the bipartite edges. The loss is given by the binary cross-entropy loss defined by the matched truth. An auxiliary hinge embedding loss is also used for the first warm-up epochs to help the embedding space stably initialize.

## 4. Results

### 4.1. Dataset

In this paper, the dataset used to report the performance of HGNN is that of the TrackML Challenge[6]. The TrackML dataset contains events of simulated proton-proton collisions at  $\sqrt{s} = 14\text{TeV}$  with pile-up  $\langle\mu\rangle = 200$ . Details can be found in [6]. The HGNN has been evaluated in two scenarios; the first scenario is called TrackML-full and contains 2200 filter-processed events, each with approximately  $O(7k)$  particles and  $O(120k)$  spacepoints. In addition to that, an extensive test of robustness has been done on Bipartite Classifiers, using a simplified dataset TrackML-1GeV. We take the subgraph induced by removing any track below  $p_T = 1\text{GeV}$ . Such an event typically consists of  $O(1k)$  particles and  $O(10k)$  spacepoints.

### 4.2. Evaluation

The evaluation metric is tracking efficiency and purity. A particle is matched to a track candidate if **(1)**: the track candidate contains more than 50% of the spacepoints left by the particle and **(2)**: more than 50% of the spacepoints in the track candidate are left by the particle. A track is called reconstructable if it **(1)** left more than 5 spacepoints in the detector and **(2)** has  $p_T \geq 1\text{GeV}$ . The tracking efficiency and fake rate (FR) are thus defined as:

$$\text{Eff} := \frac{N(\text{matched, reconstructable})}{N(\text{reconstructable})} \quad \text{FR} := 1 - \frac{N(\text{matched})}{N(\text{track candidates})}$$

### 4.3. Experiments

We evaluate four models on the TrackML-full dataset. **(1)**: Embedding Flat GNN (E-GNN), **(2)**: Embedding Hierarchical GNN (E-HGNN), **(3)**: Bipartite Classifier Hierarchical GNN (BC-HGNN), **(4)**: Edge Classifier Flat GNN (EC-GNN). The first two serve for apples-to-apples comparisons between flat and hierarchical GNNs - the loss function is the same as the hinge embedding loss used for the metric learning graph construction; tracks candidates are selected by applying a spatial clustering algorithm (H-DBSCAN). The third model represents the state-of-the-art hierarchical GNN for particle tracking; the last one is identical to the GNN4ITk pipeline, and serves as a baseline. The performance of a truth-level connected-components (Truth-CC) track builder are also reported; this takes in filter-processed graphs and prunes them down with ground truth. It is a measure of the graph quality and also an upper bound of edge classifier flat GNN performance. The timing results are obtained on a single Nvidia A100 GPU. To test

robustness against edge inefficiency, we remove 0%, 20%, 30%, and 40% of the edges and train the Bipartite Classifier model to compare it with the Truth-CC.

Table 2: TrackML-Full experiment results. Comparison between embedding models shows that hierarchical structure can enhance the expressiveness of GNNs. Comparing Bipartite Classifiers with the Truth CC, we can see that Bipartite Classifiers can recover some of the tracks that cannot be reconstructed by edge-based GNNs<sup>2</sup>. The timing results also show that HGNN scales to large input graphs of HL-LHC events competitively with other embedding GNNs

Models	E-GNN	E-HGNN	BC-HGNN	EC-GNN	Truth-CC
Efficiency	94.61%	95.60%	<b>97.86%</b>	96.35%	97.75%
Fake Rate	47.31%	47.45%	<b>36.71%</b>	55.58 %	57.67%
Time (sec.)	2.17	2.64	1.07	<b>0.22</b>	0.07

Table 3: TrackML-1GeV extensive robustness test results. We can see that Bipartite Classifiers (BC) are very robust against inefficiencies, whereas edge-based GNN’s performance is strongly influenced by missing edges.

Percent Edge Removed	0%	10%	20%	30%	40%	50%
BC Efficiency	98.55%	98.39%	97.68%	96.63%	95.10%	92.79%
BC Fake Rate	1.23%	1.55%	2.13%	3.10%	4.75%	7.31%
Truth-CC Efficiency	98.72%	96.21%	92.31%	85.81%	77.26%	64.81%
Truth-CC Fake Rate	5.87%	15.53%	24.40%	33.48%	42.99%	53.12%

## 5. Conclusion

In this paper, we introduced a novel graph neural network called a hierarchical graph neural network. We also proposed a new learnable pooling algorithm called GMPool to construct the hierarchy. The architecture successfully resolved the issues of GNN being incapable of capturing long-range interactions and the GNN particle tracking pipeline being sensitive to graphs’ efficiency. Creating higher-level representations both shortens the distance between distant nodes in graphs and offers new methods of building track candidates. Empirical results demonstrate that Hierarchical GNNs have superior performance compared with flat GNNs. The hierarchical GNN is available at <https://github.com/ryanliu30/HierarchicalGNN> and has been integrated into the common framework of the GNN4ITk pipeline [23].

## 6. Acknowledgements

This research was supported in part by: the U.S. Department of Energy’s Office of Science, Office of High Energy Physics, under Contracts No. DE-AC02-05CH11231 (CompHEP Exa.TrkX). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

<sup>2</sup> The fake rates given here are not intended as final track building performance, but rather as a model-to-model comparison. A walkthrough algorithm such as in the GNN4ITk pipeline, as well as track cleaning and ambiguity resolution can greatly reduce final fake rates.

## References

- [1] Apollinari G, Béjar A I, Brüning O, Fessia P, Lamont M, Rossi L and Taviani L 2017 High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1 Tech. rep. Geneva URL <https://cds.cern.ch/record/2284929>
- [2] Lyndon E and Philip B 2008 *Journal of Instrumentation* **3** S08001 URL <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>
- [3] Collaboration A 2022 ATLAS Software and Computing HL-LHC Roadmap Tech. rep. CERN Geneva URL <https://cds.cern.ch/record/2802918>
- [4] Farrell S, Calafiura P, Mudigonda M, Prabhat, Anderson D, Vlimant J R, Zheng S, Bendavid J, Spiropulu M, Cerati G, Gray L, Kowalkowski J, Spentzouris P and Tsaris A 2018 Novel deep learning methods for track reconstruction URL <https://arxiv.org/abs/1810.06111>
- [5] Ju X, Farrell S, Calafiura P, Murnane D, Prabhat, Gray L, Klijsma T, Pedro K, Cerati G, Kowalkowski J, Perdue G, Spentzouris P, Tran N, Vlimant J R, Zlokapa A, Pata J, Spiropulu M, An S, Aurisano A, Hewes J, Tsaris A, Terao K and Usher T 2020 Graph neural networks for particle reconstruction in high energy physics detectors URL <https://arxiv.org/abs/2003.11603>
- [6] Kiehn M, Amrouche S, Calafiura P, Estrade V, Farrell S, Germain C, Gligorov V, Golling T, Gray H, Guyon I, Hushchyn M, Innocente V, Moysse E, Rousseau D, Salzburger A, Ustyuzhanin A, Vlimant J R and Yilmaz Y 2019 *EPJ Web of Conferences* **214** 06037
- [7] Biscarat C, Caillou S, Rougier C, Stark J and Zahreddine J 2021 *EPJ Web of Conferences* **251** 03047 URL <https://doi.org/10.1051/2Fepjconf/2F202125103047>
- [8] Ju X, Murnane D, Calafiura P, Choma N, Conlon S, Farrell S, Xu Y, Spiropulu M, Vlimant J R, Aurisano A, Hewes J, Cerati G, Gray L, Klijsma T, Kowalkowski J, Atkinson M, Neubauer M, DeZoort G, Thais S, Chauhan A, Schuy A, Hsu S C, Ballow A and Lazar A 2021 *The European Physical Journal C* **81** URL <https://doi.org/10.1140/2Fepjc/2Fs10052-021-09675-8>
- [9] Caillou S, Calafiura P, Farrell S A, Ju X, Murnane D T, Rougier C, Stark J and Vallier A 2022 ATLAS ITk Track Reconstruction with a GNN-based pipeline *Connecting The Dots* (Princeton, United States) URL <https://hal.archives-ouvertes.fr/hal-03793565>
- [10] Battaglia P, Pascanu R, Lai M, Jimenez Rezende D *et al.* 2016 *Advances in neural information processing systems* **29**
- [11] Ying R, You J, Morris C, Ren X, Hamilton W L and Leskovec J 2018 Hierarchical graph representation learning with differentiable pooling URL <https://arxiv.org/abs/1806.08804>
- [12] Lee J, Lee I and Kang J 2019 Self-attention graph pooling *International conference on machine learning* (PMLR) pp 3734–3743
- [13] Diehl F 2019 *arXiv preprint arXiv:1905.10990*
- [14] Zhang Z, Zhuang F, Zhu H, Shi Z, Xiong H and He Q 2020 *Proceedings of the AAAI Conference on Artificial Intelligence* **34** 9612–9619 URL <https://ojs.aaai.org/index.php/AAAI/article/view/6508>
- [15] Li Z, Shen X, Jiao Y, Pan X, Zou P, Meng X, Yao C and Bu J 2020 Hierarchical bipartite graph neural networks: Towards large-scale e-commerce applications *2020 IEEE 36th International Conference on Data Engineering (ICDE)* pp 1677–1688
- [16] Guille A and Attali H 2022 Document classification with hierarchical graph neural networks *18th International Workshop on Mining and Learning with Graphs* URL [https://openreview.net/forum?id=ribpghC\\_s0](https://openreview.net/forum?id=ribpghC_s0)
- [17] Xia Y, Xia C Q, Pan X and Shen H B 2021 *Nucleic Acids Research* **49** e51–e51 ISSN 0305-1048 (*Preprint* <https://academic.oup.com/nar/article-pdf/49/9/e51/38000370/gkab044.pdf>) URL <https://doi.org/10.1093/nar/gkab044>
- [18] Zhong Z, Li C T and Pang J 2022 *Data Mining and Knowledge Discovery* 1–28
- [19] Gao H and Ji S 2019 Graph u-nets *international conference on machine learning* (PMLR) pp 2083–2092
- [20] Rampásek L and Wolf G 2021 Hierarchical graph neural nets can capture long-range interactions *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE) pp 1–6
- [21] Chen C, Li K, Wei W, Zhou J T and Zeng Z 2021 *IEEE Transactions on Circuits and Systems for Video Technology* **32** 240–252
- [22] Qasim S R, Kieseler J, Iiyama Y and Pierini M 2019 *The European Physical Journal C* **79** 1–11
- [23] Atkinson M J, Caillou S, Clafiura P, Collard C, Farrell S A, Huth B, Ju X, Liu R, Minh Pham T, Murnane D c a, Neubauer M, Rougier C, Stark J, Torres H and Vallier A gnn4itk URL <https://github.com/GNN4ITkTeam/CommonFramework>