A Graph Neural Network for B decays reconstruction at Belle II

J Cerasoli[1,2], F Bernlochner[3], T Boeckh[3], G Dujany[1], P Goldenzweig[4], M Götz[5,6], J Kahn[5,6], L Reuter[4], A Streit[6], O Taubert[5,6], A Thaller[7] and I Tsaklidis[3]

[1]Université de Strasbourg, CNRS, IPHC, UMR 7178, 67037 Strasbourg, France
[3]Physikalisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn, Germany
[4]Insitute for Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[5]Helmholtz AI, Germany
[6]Steinbuch Centre for Computing (SCC), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[7]Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

E-mail: `jacopo.cerasoli@iphc.cnrs.fr`

Abstract. The Belle II experiment has unique features that allow to study B meson decays with neutrinos in the final state. It is possible to deduce the presence of such particles from the energy-momentum imbalance obtained after reconstructing the companion B meson produced in the event. This task is complicated by the thousands of possible final states in which B mesons can decay, and is currently performed at Belle II by the Full Event Interpretation, an algorithm based on Boosted Decision Trees and limited to specific, hard-coded decay processes. In recent years, graph neural networks have proven to be very effective tools to describe relations in physical systems, with applications in a range of fields. Particle decays can be naturally represented in the form of rooted directed acyclic tree graphs, with nodes corresponding to particles and edges representing the parent-child relations between them. In this work, we present a graph neural network approach to generically reconstruct B decays at Belle II by exploiting the information on the final-state particles alone, without formulating any prior assumption about the nature of the decay. Preliminary results show that the graph neural network approach outperforms the Full Event Interpretation by a factor of about 2 in terms of number of decays whose topologies are correctly reconstructed.

1. Introduction

The Belle II detector [1] is a particle physics experiment located at the SuperKEKB accelerator [2] in Tsukuba, Japan, where e^+e^- pairs are collided at the center-of-mass energy of the $\Upsilon(4S)$ resonance, producing pairs of B mesons. The experiment has unique features that allow to study B meson decays with invisible particles in the final state, for example neutrinos. As shown in figure 1, the presence of such particles in signal decays (signal-side) can be deduced by the energy-momentum imbalance obtained after reconstructing the companion B meson produced in the event (tag-side). However, such a task is complicated by the presence of

[2] On behalf of the Belle II software group.

thousands of decay modes in which the B can decay. The reconstruction of the tag-side decay is currently performed at Belle II with the Full Event Interpretation (FEI) [3]. The FEI is an algorithm based on Boosted Decision Trees (BDT) which uses a hierarchical approach with six stages, as shown in figure 2. It constructs final-state particle candidates using the information on the reconstructed tracks and clusters in the event, and combines them to intermediate particles until a B candidate is formed. The output of the final BDT is interpreted as a likelihood for

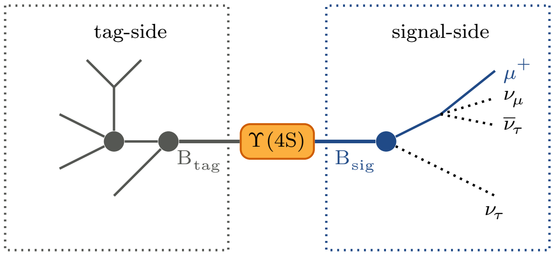


Figure 1. Schematic representation of a $\Upsilon(4S)$ decay into two B mesons [3].

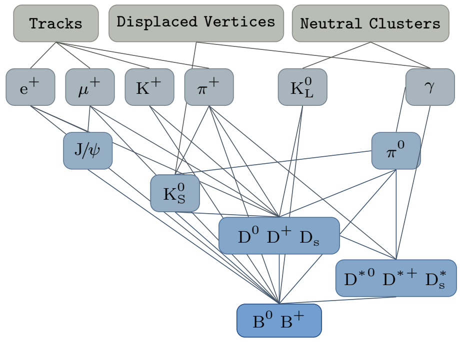


Figure 2. Schematic view of the six stages of the FEI [3].

the reconstructed candidate to be a real B meson decay. The decay modes are hard-coded in the algorithm, which limits the number of reconstructible decay channels to about 15% of the total branching fraction, resulting in a reconstruction efficiency of about 1%.

In this paper, a novel approach based on deep Graph Neural Networks (GNN) to inclusively reconstruct the tag-side B meson, which allows to increase the reconstruction efficiency, is presented. Section 2 describes the proof of concept using a simulated toy particle decay dataset, while Section 3 describes a first application to a fully simulated Belle II dataset.

2. Particle decay reconstruction with deep Graph Neural Networks

GNNs are a class of deep learning algorithms that act on *graphs*, i.e. sets of independent entities (nodes) and their relationships (edges). A particle decay can be naturally described by a rooted directed acyclic tree graph, where nodes represent particles and edges the parent-child relations between them. The adjacency matrix $\mathbf{A}$ of a graph of n nodes is an $n \times n$ binary matrix with $A_{ij} = 1$ if two nodes i and j are connected by an edge and $A_{ij} = 0$ otherwise. However, the approach of using a GNN to reconstruct the adjacency matrix requires the total number of nodes in the graph to be known *a priori*, while only final-state particles are detected by experiments, the total number of intermediate particles being unknown. To overcome this difficulty, a novel approach based on the Lowest Common Ancestor (LCA) matrix [4, 5] is used. The LCA of two nodes a and b is defined as the farthest node from the root that is an ancestor of both a and b. Figure 3 shows an example of B decay with its adjacency and LCA matrices. In the following, two representations of the LCA matrix are employed: the LCAG, in which each ancestor is replaced with its corresponding generation in the tree, and the LCAS matrix, in which each ancestor belongs to a predefined class (5 for B mesons, 4 for D^* mesons, 3 for D mesons, 2 for K^0_S, 1 for π^0 and J/ψ and 0 if a common ancestor can not be identified).

For the proof of concept [6], the training dataset is simulated using the Phasespace library [7] and reflects the relative mass scales and decay multiplicities observed in nature. The network used is the encoder component of the Neural Relational Inference (NRI) model [8], shown in figure 4. Its building blocks consist of node-to-edge and edge-to-node message passing modules, followed by MultiLayer Perceptrons (MLPs) containing two linear layers with Exponential Linear

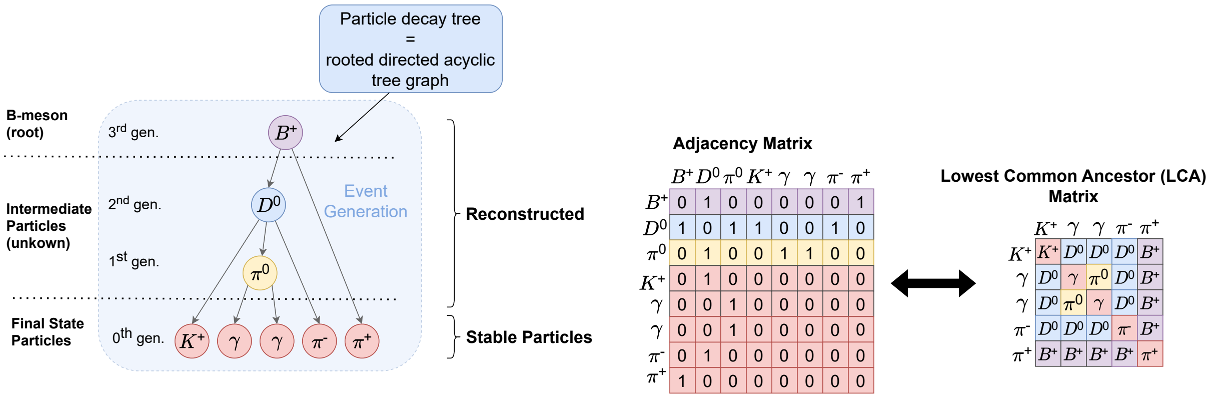


Figure 3. Example of B decay described in terms of its adjacency and LCA matrices [5].

Unit (ELU) activations. The model takes as input the four-momenta of the simulated final-state particles. The target of the training is the LCAG matrix, where each entry is treated as an individual classification task with a softmax followed by an argmax function after the last layer. The metric used to evaluate the performances is the rate of perfectly reconstructed LCAG matrices in the dataset. Figure 5 shows the performances of the NRI model applied to the simulated dataset. The model is able to correctly predict the LCAG matrix for an average of 47.7% of decay trees in an independent sample, which increases to 60.9% and 94.2% for decay trees with ≤ 10 and ≤ 6 final-state particles.

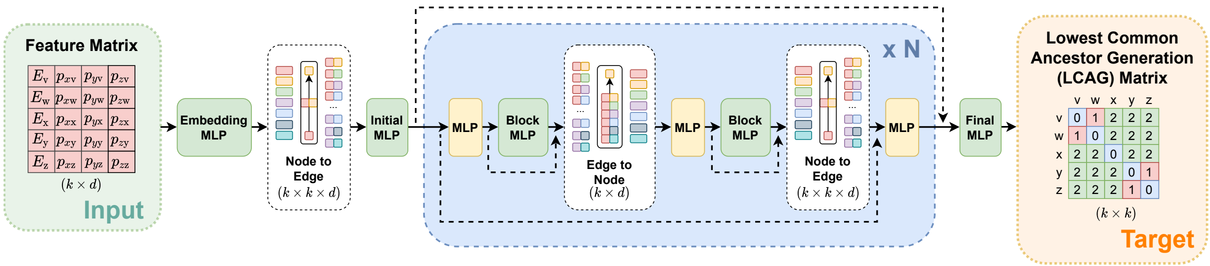


Figure 4. Schematic representation of the encoder component of the NRI model. The dashed arrows indicate skip connections [6].

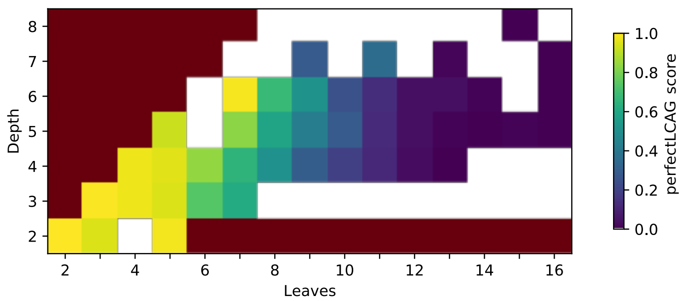


Figure 5. Performances of the NRI model as a function of the number of final-state particles (leaves) and the depth of the decay tree. Dark red indicates disallowed topologies [6].

3. Application to Belle II simulation

In order to reconstruct the LCA matrix, and hence the topology, of B decays in the Belle II experiment, a new model called Graph-based Full Event Interpretation (graFEI) is developed. A schematic view of the model is shown in figure 6. The model is based on *graph network blocks* [9] and its input is a fully-connected graph where nodes represent the final-state particles detected by the experiment. The input graph contains node-level (four-momenta, charged particle identification information, impact parameters of the tracks, neutral clusters information), edge-level (angle between pairs of particles' momenta) and global-level (number of particles in the final state) attributes. The input graph is passed through a series of graph network blocks, each in turn composed of three sub-blocks dedicated to the update of edge, node, and global attributes. Each sub-block combines the input information into an MLP with one hidden layer and ELU activation. The input information of the edge sub-block is the edge attributes, the node attributes of the two nodes connected to each edge and the global attributes. The output of the sub-block is then used as input for the following node and global sub-blocks. The input information of the node sub-block is the node attributes, the average of edge attributes for edges connected to the same node and the global attributes. The output of the sub-block is then used as input for the following global sub-block. The input information of the global sub-block is the global attributes, the average of edge attributes over the graph and the average of node attributes over the graph. The updated edge-level attributes are used to predict the LCAS matrix of the

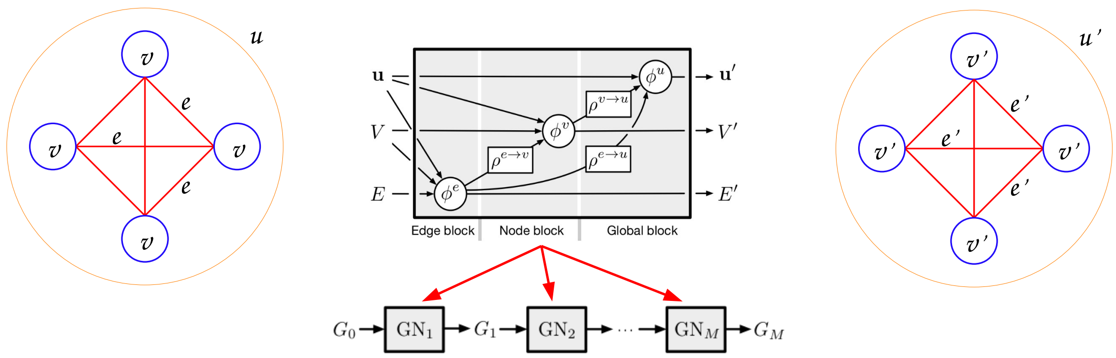
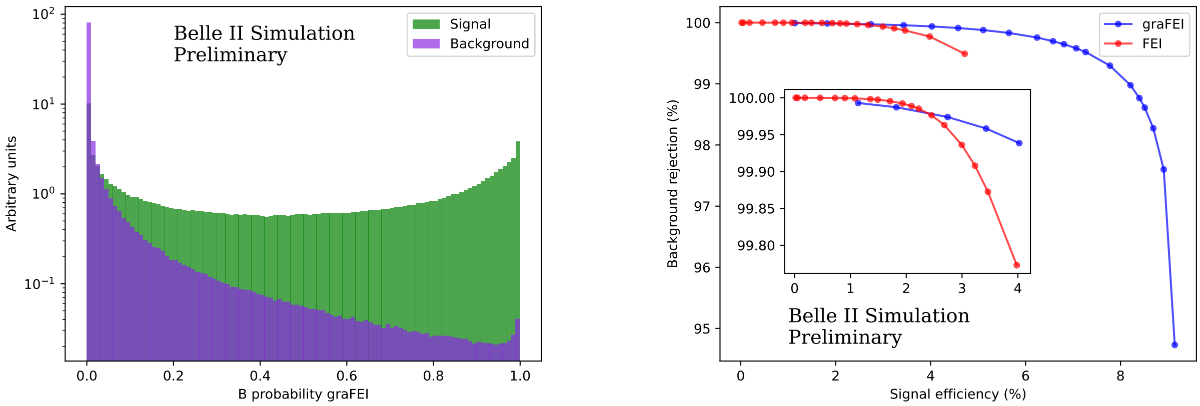


Figure 6. Schematic view of the graFEI model. MLPs are indicated with ϕ whereas ρ indicates *aggregation functions*, i.e. averages of edges connected to the same node ($\rho^{e\to v}$), edges over the graph ($\rho^{e\to u}$) or nodes over the graph ($\rho^{v\to u}$). Figure adapted from [9].

decay by applying a softmax followed by an argmax function. The model is trained on about 9 million simulated *monogeneric* $\Upsilon(4S) \to B^0(\to \nu\bar{\nu})\bar{B}^0(\to X)$ decays, where one B^0 meson decays into a pair of undetected neutrinos, while the companion B^0 meson is decayed into any of the possible B^0 decays included in the Belle II simulation framework. The model is able to correctly predict the LCAS matrix for an average of 18.6% of decay trees in an independent monogeneric sample. It has to be noticed that, contrary to the simulated sample used for the training of the NRI, resolution effects which degrade the performances of the model are present in the Belle II simulation.

A comparison of the performances of the graFEI and FEI is performed using an independent monogeneric signal sample and a background sample obtained selecting a random subset of tracks from simulated generic $\Upsilon(4S)$ decays. In order to derive from the graFEI a quantity able to discriminate between signal and background events, the softmax probabilities corresponding to the predicted classes are multiplied over the LCAS matrix. The distribution of such a quantity and a comparison of the two algorithms are shown in figure 7. For both the graFEI and FEI it is required that, in addition to having a perfectly reconstructed decay topology, there are no tracks in the event not belonging to the decay tree. In these conditions, the graFEI shows a maximum signal efficiency of 9.1% with a corresponding background rejection of 94.7%, while the FEI shows a maximum signal efficiency of 4.7% with a corresponding background rejection of 99.5%.



(a)

(b)

Figure 7. (a) Discriminating quantity derived from graFEI softmax probabilities used to separate signal events from random combinations of tracks coming from B^0 decays. (b) Background rejection as a function of signal efficiency for graFEI and FEI.

4. Conclusions

A novel approach based on GNNs allows to inclusively reconstruct particle decays exploiting the information on the final-state particles alone. Results obtained on a simulated Phasespace dataset show the ability of the encoder part of the NRI model to correctly predict the LCA matrix of a large fraction of decay trees. Extending this approach to a simulated Belle II dataset suggests that GNNs enable a two times higher maximum signal efficiency than the current reconstruction algorithm.

Acknowledgments

This work is supported by the Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) du CNRS (France), the Centre de Calcul de l'IN2P3 (CC-IN2P3), the French Agence Nationale de la Recherche (ANR) under Grant ANR-21-CE31-0009 (Project FIDDLE) and the Helmholtz Association Initiative and Networking Fund under the Helmholtz AI platform grant and the HAICORE@KIT partition.

References

[1] Abe T *et al.* (Belle II collaboration) 2010 (*Preprint* `1011.0352`)
[2] Akai K, Furukawa K and Koiso H 2018 *Nucl. Instrum. Meth.* **A907** 188 (*Preprint* `1809.01958`)
[3] Keck T *et al.* 2019 *Comput. Softw. Big Sci.* **3** 6 (*Preprint* `1807.08680`)
[4] Tsaklidis I 2020 *Demonstrating learned particle decay reconstruction using Graph Neural Networks at Belle II* Master's thesis Université de Strasbourg
[5] Reuter L 2022 *Full Event Interpretation using Graph Neural Networks* Master's thesis Karlsruhe Institute of Technology (KIT)
[6] Kahn J *et al.* 2022 *Machine Learning: Science and Technology* **3** 035012 (*Preprint* `2208.14924`)
[7] Navarro A P and Eschle J 2019 *Journal of Open Source Software* **4** 1570
[8] Kipf T *et al.* 2018 (*Preprint* `1802.04687`)
[9] Battaglia P *et al.* 2018 (*Preprint* `1806.01261`)