RESEARCH

GNN for Deep Full Event Interpretation and Hierarchical Reconstruction of Heavy-Hadron Decays in Proton–Proton Collisions

Julián García Pardiñas[1,2] · Marta Calvi[1] · Jonas Eschle[3] · Andrea Mauri[4] · Simone Meloni[1] · Martina Mozzanica[1] · Nicola Serra[3]

Abstract

The LHCb experiment at the Large Hadron Collider (LHC) is designed to perform high-precision measurements of heavy-hadron decays, which requires the collection of large data samples and a good understanding and suppression of multiple background sources. Both factors are challenged by a fivefold increase in the average number of proton–proton collisions per bunch crossing, corresponding to a change in the detector operation conditions for the LHCb Upgrade I phase, recently started. A further tenfold increase is expected in the Upgrade II phase, planned for the next decade. The limits in the storage capacity of the trigger will bring an inverse relationship between the number of particles selected to be stored per event and the number of events that can be recorded. In addition the background levels will rise due to the enlarged combinatorics. To tackle both challenges, we propose a novel approach, never attempted before in a hadronic collider: a Deep-learning based Full Event Interpretation (DFEI), to perform the simultaneous identification, isolation and hierarchical reconstruction of all the heavy-hadron decay chains per event. This strategy radically contrasts with the standard selection procedure used in LHCb to identify heavy-hadron decays, that looks individually at subsets of particles compatible with being products of specific decay types, disregarding the contextual information from the rest of the event. Following the DFEI approach, once the relevant particles in each event are identified, the rest can be safely removed to optimise the storage space and maximise the trigger efficiency. We present the first prototype for the DFEI algorithm, that leverages the power of Graph Neural Networks (GNN). This paper describes the design and development of the algorithm, and its performance in Upgrade I simulated conditions.

Keywords GNN · High-energy-physics · LHCb · Full event interpretation · Event reconstruction · Trigger

✉ Julián García Pardiñas
julian.garcia.pardinas@cern.ch

Marta Calvi
marta.calvi@cern.ch

Jonas Eschle
jonas.eschle@cern.ch

Andrea Mauri
a.mauri@cern.ch

Simone Meloni
sim.meloni@gmail.com

Martina Mozzanica
martina.mozzanica@uni-hamburg.de

Nicola Serra
nicola.serra@cern.ch

1   Dipartimento di Fisica "G. Occhialini", Università di Milano Bicocca and INFN Sezione di Milano-Bicocca, Piazza della Scienza 3, Milano 20126, Italy

2   Experimental Physics Department, European Organization for Nuclear Research (CERN), Espl. des Particules 1, Meyrin 1211, Switzerland

3   Department of Physics, University of Zürich, Winterthurerstrasse 190, Zürich 8057, Switzerland

4   Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Introduction

The Large Hadron Collider beauty Experiment (LHCb) is one of the four large experiments at the proton–proton collider LHC, at CERN [1]. It is dedicated to the study of beauty (b) and charm (c) hadron decays, performing high-precision measurements to test the validity of the Standard Model (SM) of particle physics and identify possible signatures of the presence of physics beyond the SM. To push the precision frontier, LHCb needs to record as many heavy-hadron decays as possible. One way to increase that quantity for a given period of data collection is to increment the average number of proton–proton collisions that happen in each event (bunch crossing). During the LHC Run 1 and Run 2 periods, between 2010 and 2018, each LHCb event contained an average of around one visible proton–proton collision, producing a flow of tens of particles to be reconstructed. The experiment has now undergone its Upgrade I, with the installation of new sub-detectors and a new data-collection software to allow the processing of events with around five visible proton–proton collisions each. These will be the conditions for the ongoing Run 3 and for the future Run 4. In a decade from now, the Upgrade II [2] of LHCb will prepare the experiment to face another tenfold increase in proton-collision multiplicity [3] to fully exploit the High-Luminosity (HL-LHC) Phase of the LHC during Runs 5 and 6. The approximated expected object multiplicities per event in the different conditions are shown in Table 1. Beyond upgraded sub-detectors, the much larger event complexities bring unprecedented challenges to LHCb, both for data-collection and for the eventual measurements. New strategies need to be devised and implemented to tackle those challenges and hence maximise the future physics reach of the experiment.

So far, the entire data flow of the LHCb experiment has been based on an exclusive approach, i.e. it was sufficient for a set of particles to be compatible with a certain type of decay to be identified as a signal candidate. While this approach has its merits, it ignores all the remaining particles produced in the collision during its selection process, which contain important information on the underlying physics process. Exceptions to this exclusive approach are found in flavour tagging algorithms [4] and isolation studies [5, 6]. However, both cases examine the rest of the event in relation to a specific candidate, e.g. flavour tagging aims to infer the flavour of the heavy-hadron associated with a given signal candidate. While technically very challenging, significantly more information could be gained by an inclusive study of all the particles in the event. This would not only add discriminating power to disentangle true signal decays from multiple sources of background but also allow for the identification and separation of groups of particles corresponding to multiple heavy-hadron decays in the event, all of which can be used for subsequent physics analyses. The gains of such an inclusive approach compared to the individual study of signal candidates become stronger with increasing event complexities, as the larger combinatorics problem makes it more complicated to identify and isolate signals [5, 6].

The individual study of heavy-hadron decays is also at the core of the LHCb strategy for data collection. The trigger of the experiment aims at discerning between events that contain a signal decay and those that do not, by means of a combination of exclusive and partially inclusive [7, 8] particle selections. In previous LHC runs, the disk space available to store the information for the selected events was large enough to allow persisting all the objects in the event in many cases. This provided the flexibility to study offline particles other than the ones composing the signal candidate that triggered the event, which is a crucial feature for signal-background separation in many analyses and for allowing the study of modes not considered when the trigger selections were made. This situation is completely different in the HL-LHC era. First of all, the fraction of events containing decays of interest will saturate to around 100%, with each event typically containing several heavy-hadron decays. Second, the event sizes will be much larger than in the past due to increased particle multiplicity. This implies that the potential datasets to be collected are huge, while the available disk space is limited and imposes tight constraints. A trigger strategy based on selecting events in those conditions necessarily leads to a signal inefficiency, impacting the potential physics reach of the experiment. Consequently, the trigger paradigm needs to shift from deciding "which events are interesting?" to "which parts of the event are interesting?". Minimising the average event size will directly translate into maximising the number of events LHCb can record. When

Table 1 Approximate average quantities per event for the different LHCb run conditions, as estimated from the simulation used in this work

LHCb period	Num. vis. pp collisions	Num. tracks	Num. b hadrons	Num. c hadrons
Runs 1–2	~ 1	~ 50	$\ll 1$	$\ll 1$
Runs 3–4 (Upgrade I)	~ 5	~ 150	$\ll 1$	~ 1
Runs 5 (Upgrade II)	~ 50	~ 1000	~ 1	~ 5

Only objects in LHCb geometrical acceptance are considered

doing so, the trigger needs to ensure that the relevant particles (those produced in heavy-hadron decays) are amongst those to be kept for offline analysis, otherwise also impacting the potential physics reach. These problems are already partially present in the current LHCb Upgrade I, as anticipated in Ref. [9]. In preparation, LHCb has developed a framework that allows the persistency of part of the event, named turbo stream [10] (e.g. the persistency of the information associated to the identified signal candidates and the set of reconstructed particles associated to the same proton–proton collision). During the ongoing Run 3, for instance, about two-thirds of the recorded events follow such turbo data processing model. However, at present, there is no nominal strategy in LHCb to systematically identify which parts of the event may be interesting for physics analysis. This is a very complicated task affected by large particle combinatorics and a huge variability of types of signal decays.

To tackle the previous challenges, we propose a new algorithm to perform a Deep-learning based Full Event Interpretation (DFEI) at LHCb. This innovative approach, which targets an inclusive analysis of the entire event, represents a shift of paradigm with important applications both at the trigger level and at the offline analysis level. The algorithm takes as input all the reconstructed particles in an event and aims to identify which of them originate from the decays of heavy-hadrons and at reconstructing the hierarchical decay chains through which they were produced. The possibility of accomplishing this difficult task leverages on some of the most recent developments in the field of machine learning. At the trigger level, DFEI can identify the part of each event which is interesting for physics analyses, allowing to safely discard the rest of the event and hence minimise the storage required. As an additional benefit, an automated identification and classification of the decay chains could eventually replace the need for cut-based exclusive selections that need to be designed and carefully tuned independently for each signal decay type. At the offline analysis level, DFEI can offer a common tool for physicists to identify and classify the different types of backgrounds contributing to a broad spectrum of possible decays of interest. Leveraging the information from all the correlations in the event can enhance the background rejection power in many cases, increasing the precision of future LHCb measurements.

This document describes the conceptualisation, construction, training and performance of the first prototype of the DFEI algorithm. The prototype is specialised for reconstructed charged particles produced in beauty-hadron decays. Extensions to include reconstructed neutral particles and charm-hadron decays can be considered in the future. All the studies are done using simulated datasets that emulate proton–proton collisions in the LHCb Run 3 environment. These datasets have been produced with a custom simulation framework and made publicly available for future

benchmarking. The algorithm is based on a composition of Graph Neural Network (GNN) models, designed to handle the complexity of high-multiplicity events in a computationally efficient way. Regarding the paper organisation, the state of the art is first presented in Sect. "Related Work". The development of the DFEI prototype is described in Sect. "Methods" , starting with an introduction to GNN models in Sect. "Usage of Graph Neural Networks", followed by the description of the employed dataset in Sect. "Dataset" for which additional details are provided in App. A, the structure of the algorithm in Sect. "Structure of the Algorithm", and finally the training in Sect. "Training". The performance of the algorithm is described in detail in Sect. "Results". In particular, the quality of the reconstruction is first evaluated at the event level, in Sect. "Event-Level Performance", and then at the exclusive-signal level, in Sect. "Decay-Level Performance". A timing study is presented in Sect. "Timing Studies" (with additional details provided in App. B). The results are discussed in Sect. "Discussion" and future prospects are presented in Sect. "Future Work". Finally, the conclusions are summarised in Sect. "Conclusion".

Related Work

Even though the problem addressed in this paper is unique, it shares similarities with a variety of past efforts at the technical and/or scientific level. In this section, we conduct a review of those efforts and place our approach in the context of the field.

The first and so far only use of a machine learning-based approach on the full set of reconstructed tracks within LHCb is Ref. [11], where the authors employed a probabilistic model based on decision trees for the inclusive flavour tagging of signal beauty hadrons. The combined processing of all the event information demonstrated better results compared to a combination of multiple classical flavour tagging algorithms, each using only a subset of the reconstructed particles in the event. The task of flavour tagging is much simpler than the explicit decay chain reconstruction attempted by DFEI. Regarding isolation tools, past LHCb efforts [5, 6] are restricted to multivariate classifiers that aim to predict whether individual particles from the rest of the event originate or not from the same heavy-hadron decay as a signal candidate. The decision is based on a combination of features from the signal candidate and the extra particle, fully disregarding any correlation with the other particles in the event. Concerning trigger-oriented applications, the authors in Ref. [12] presented a study of the full information in the event in terms of the activity in the different LHCb sub-detectors, hence at a level prior to the reconstruction of the stable particles, which are considered as input in DFEI. Using machine learning techniques, they successfully

managed to predict the number of reconstructible proton-proton collisions per event.[1] They also studied the possible classification of events between those containing (at least) one b-hadron decay and those that do not, but this turned out to be a very complicated task when looking only at the sub-detector activity information.

Regarding other LHC experiments, a type of full event reconstruction is done in CMS [13] and ATLAS [14], through the usage of the particle flow algorithm. The implementation uses all the final state particles for a global event description, significantly improving the performance of jet reconstruction with respect to the previous baseline that used basic geometric cones to cluster particles. In order to further improve the performance, an approach with a GNN [15] was proposed in CMS, which takes all particles of an event as input and predicts variables such as particle identification and transverse momentum for each particle. While similar to DFEI at a technical level, the particle flow algorithm does not attempt to explicitly reconstruct the decay chains for all the relevant decays of interest.

The task of decay-chain reconstruction is conceptually close to the hierarchical reconstruction of jets, for which a variety of algorithms based on GNN were developed [16–30]. The ultimate goal of those algorithms, however, is typically focused on inferring quantities of the jet overall, for example doing a flavour tagging of the jet to determine the initial particle, and reconstructing the jet to infer its kinematics. The jet substructure is only studied to the extent in which it's useful for those purposes. The limitations of those algorithms for the task of reconstructing all the ancestors in particle decay chains are reviewed in detail in Ref. [31].

The effort closest to the one presented in this paper is being conducted at the Belle II experiment, where the FEI algorithm [32] was developed for exclusive tagging of B-decays. This constitutes a similar approach to the one presented in this paper but with a different goal and in a simpler environment. As Belle II is a hermetic detector situated at an electron-positron collider, the event is a fully reconstructible system with known initial states and significantly less tracks, making the task of inference less challenging. In addition, only two species of b hadrons are studied, B^0 and B^+ mesons,[2] while LHCb is interested in all b-hadron species (for example B_s and B_c mesons, and Λ_b baryons) as well as c-hadron decays. From a probabilistic point of view, the FEI algorithm at Belle II is based on a fixed set of different boosted decision tree classifiers, one for each

considered decay type. This approach would be unfeasible at LHCb, given the much larger variability in terms of different signal decay topologies, further compounded by the fact that a fraction of the particles produced in the decays may fall outside the LHCb geometrical acceptance, and hence not be reconstructed in the detector. Recently, an extension to the FEI algorithm based on GNN was proposed [31, 33], showing a better performance than the previous implementation. This resembles the approach presented in this paper, but in a very different environment, as has been discussed.

As exemplified by previous efforts, GNNs have become popular for replacing other machine learning algorithms within particle physics experiments [34, 35], as they can naturally capture the structure and spatial sparsity of the problem. A challenge, however, is the GNN's performance in deployment, such as in real-time computing for trigger purposes. Achieving a fast inference with GNNs would require sparse operations and standards of representing such operations in protocols that would allow the automatic optimisation of the networks. This is a matter of broad interest and front-line research. Very recently, there have been multiple successful efforts in this direction within other CERN experiments [36–42], for example by reducing the complexity of the networks and using FPGAs or GPUs as hardware accelerators.

Methods

Usage of Graph Neural Networks

Machine learning and especially neural networks usage in particle physics has been growing exponentially in the last decade [43]. The major motivation to explore new and increasingly complex machine learning techniques is to optimally incorporate the structure of the underlying problem into the model itself. This includes incorporating variable input sizes, representing different types of connections between inputs and embedding invariances into the architecture. Graph Neural Networks are a class of neural networks built around the concept of a graph, which is an unordered and variable-sized collection of nodes ($v \in V$), edges connecting those nodes ($e \in E$), and possibly a vector of graph-level features (**u**). The relations between the nodes occur in a high-dimensional latent space, allowing for a more complete description of the data. This architecture is especially well-fitted to capture problems with sparse connections and invariance under input permutation, as is the case for the set of reconstructed particles in a collision event.

In general, GNNs implement "graph-to-graph" transformations, by the application of multiple layers that operate on the graph constituents. At each layer, input vectors of features at the node, edge and/or graph level are used and

---

[1] The DFEI algorithm assumes the proton-proton collision points have already been reconstructed, and usses information of their measured positions as input, as discussed in the following sections.

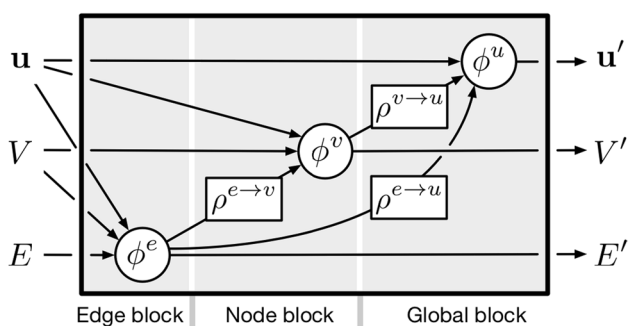[2] Charge conjugation is implied throughout this paper.

Fig. 1 Graph processing block at each message-passing step, as presented in Ref. [45]

returned to the next layer, with the output of the last layer fulfilling the goal of a certain task. This work is based on the usage of message-passing GNNs [44], in which the information is propagated through the graph at each layer by exchanging information between adjacent nodes. Specifically, we use the so-called full GN block in Ref. [45], depicted in Fig. 1. This block is composed of three feature-update functions, ϕ^v, ϕ^e and ϕ^u, and three information-aggregation functions, $\rho^{e \to v}$, $\rho^{e \to u}$ and $\rho^{v \to u}$. Each of the three update functions is implemented by a multilayer perceptron (MLP), and the aggregation functions are element-wise summations.

These blocks are then applied multiple times, and what is returned is a new representation of the graph. As the final layer, each output element–nodes, edges or the entire graph—has a sigmoid for binary activation function and a softmax for multilabel classification. The output is then up to interpretation as further described in Sect. "Structure of the Algorithm".

Dataset

The DFEI prototype is trained on simulated data. Since the LHCb simulation samples are restricted to internal member access only, and there is no publicly available dataset that fully captures the essence of the problem at hand, we have created a new simulation environment and produced datasets which we have made publicly available [46]. The datasets are generated with PYTHIA8 [47] and EvtGen [48], replicating the particle-collision conditions expected for the LHCb Run 3. In addition, an approximate emulation of the LHCb detection and reconstruction effects is applied, as described in App. A. In the generated dataset, each event is required to contain at least one b-hadron, which is subsequently allowed to decay freely through any of the standard decay modes present in PYTHIA8. In these conditions, around 40% of the events contain more than one b-hadron decay within LHCb acceptance, and the maximum observed number of

b-hadrons is five. All the studies presented in this paper refer only to reconstructed particles that have been produced inside the LHCb geometrical acceptance and in the Vertex Locator region (as defined in App. A). Other particles are not considered, which also implies they are not included as part of the ground truth heavy-hadron decay chains.

A total of 100,000 simulated events have been used to develop this first prototype of the DFEI algorithm. They are divided into: training dataset (40,000 events), test dataset (10,000 events) and evaluation dataset (50,000 events). In addition to this inclusive dataset, several other smaller samples (of a few thousand events each) have also been generated simulating specific signal decay types. These decay types have been chosen to be representative of the most common signal topologies studied in physics analyses at LHCb, and are used to evaluate the performance of DFEI focused on typical use cases. These samples contain only events in which all the particles originating from each of the considered exclusive decays have been produced inside the LHCb geometrical acceptance and in the Vertex Locator region.

The input features used in the DFEI GNN modules are described in the following. Regarding geometrical variables, a cartesian right-handed coordinate system is adopted, with the z axis along the beam, the y pointing upwards and the x axis parallel to the horizontal.

- Node variables:

    - Transverse momentum (p_T): component of the three-momentum transverse to the beamline.
    - Impact parameter (IP) with respect to the associated primary vertex (PV): distance of closest approach between the particle trajectory and its associated primary vertex (proton-proton collision point), defined as the one with the smallest IP for the given particle amongst all the primary vertices in the event.
    - Pseudorapidity (η): spatial coordinate describing the angle of a particle relative to the beam axis, computed as $\eta = \operatorname{arctanh}(p_z / \|\mathbf{p}\|)$.
    - Charge (q): since only charged reconstructed particles are considered, the charge can only take the values 1 or -1.
    - O_x, O_y, O_z: cartesian coordinates of the origin point of the particle.
    - p_x, p_y, p_z: cartesian coordinates of the three-momentum of the particle.
    - PV_x, PV_y, PV_z: cartesian coordinates of the position of the associated primary vertex.

- Edge variables:

    - Opening angle (θ): angle between the three-momentum directions of the two particles.

- Momentum-transverse distance ($d_{\perp\mathbf{P}}$): distance between the origin point of the two particles projected onto a plane which is transverse to the combined three momentum of the two particles.
- Distance along the beam axis (Δ_z): difference between the z-coordinate of the origin points of the two particles.
- *FromSamePV*: Boolean variable indicating whether the two particles share the same associated primary vertex.
- *IsSelfLoop*: Boolean variable indicating whether the edge is connecting a particle with itself or not (i.e. it connects two different particles).

Structure of the Algorithm

A full-sized event graph with all necessary features can grow quite large, possibly exceeding the available computing resources in online applications. To improve the scalability of the algorithm, a sequential approach is adopted: several event pre-filtering steps are applied before the decay chain reconstruction is performed, with tunable thresholds depending on the available resources. It should be noted that the GNN-based FEI algorithm used at Belle II uses a dense network, that would not scale well to LHC conditions. Each collision event is transformed into a graph, where the charged reconstructed particles are represented as nodes and the relations between them are represented as edges. Edges are established between particles that either share the same associated primary vertex or have an opening angle smaller than a given threshold, to reduce the graph size and the time needed to build it. Requiring that the edge selection keeps 99% of the connections between particles that originate from the same b-hadron decay corresponds to choosing a threshold value for θ of 0.26 rad. This requirement removes around 11% of all the other connections. Further tuning of this parameter goes beyond the scope of this paper, which privileges a loose preselection in order not to compromise the subsequent performance of the algorithm.

The input graph is passed subsequently through three GNN modules, built using the `graph_nets` library [45]. The modules are schematically represented in Fig. 2 and described in the following. The input features used by each module are specified in Table 2.

- *Node pruning (NP)*. The first GNN module has the goal of removing most of the particles (nodes) that have not been produced in the decay of any b-hadron. It mostly exploits the fact that particles produced in the decay of a b-hadron typically have large IP and p_T values. Since the main contributing factor to the prediction of each node in this case comes from the same node's features, self-loop connections are included in the graphs. The model
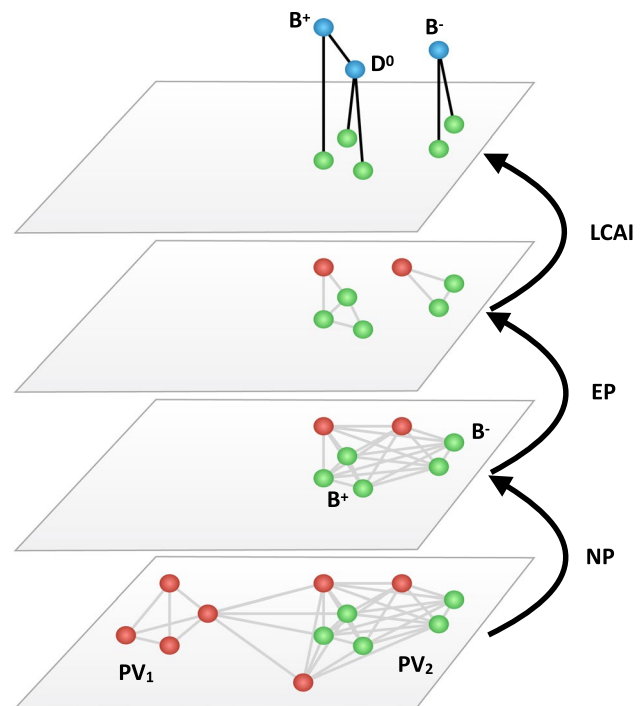

Fig. 2 Schematic representation of an event processing by the algorithm. Green (red) graph nodes represent particles originated in the decay chain of a b-hadron (from the rest of the event). The reconstructed ancestors are represented in blue

is trained using binary cross-entropy loss function to predict whether a node originates from beauty hadrons or not. Nodes with an output score below a certain threshold are removed from the graph.

- *Edge pruning (EP)*. The output graph of the previous step still has a large number of edges, which are further reduced by a second GNN module. This one aims to remove edges between particles that do not share the same beauty-hadron ancestor. Amongst other relations, this exploits the fact that particles coming from the same b-hadron decay tend to be closer in space and their three-momenta tend to form a small opening angle. The model is trained using binary cross-entropy loss function to predict whether an edge connects two particles from the same beauty-hadron decay. Edges with an output score below a certain threshold are removed from the graph.
- *Lowest common ancestor inference (LCAI)*. Finally, a third GNN module takes the output of the previous algorithm, and aims at inferring the so-called "lowest common ancestor" of each pair of particles (a technique similar to the recently proposed LCA-matrix reconstruction for the Belle II experiment [33]). The limited coverage of the LHCb geometrical acceptance and the fact that only charged reconstructed particles are considered in this prototype implies that a large fraction of the decay chains can only be partially reconstructible.

Table 2 Input variables used by each of the DFEI modules

Model	Node variables	Edge variables	Global variables
NP	p_T, IP, η, q	*FromSamePV*, θ, $d_{\perp\mathbf{P}}$, Δ_z, *IsSelfLoop*	*NumParts*
EP	p_T, IP, η, q	*FromSamePV*, θ, $d_{\perp\mathbf{P}}$, Δ_z	*NumParts*
LCAI	$O_{(x,y,z)}$, $p_{(x,y,z)}$, $PV_{(x,y,z)}$, q	*FromSamePV*, θ, $d_{\perp\mathbf{P}}$, Δ_z	*NumParts*

In addition, the total number of stable particles per event (*NumParts*) is included as an input variable. Where relevant, edges connecting two different particles are differentiated from edges connecting a particle with itself (self-loops) by the Boolean variable *IsSelfLoop*

Table 3 Hyperparameters used in the construction and training of the different GNN modules

Model	# Layers	Latent size	# GN blocks	Batch size	Learning rate	# Training steps
NP	3	50	3	32	$5 \cdot 10^{-4}$	500
EP	4	100	5	32	10^{-4}	500
LCAI	5	100	5	128	10^{-3}	$2 \cdot 10^5$

To circumvent this limitation, the target decay chains for this prototype are not the ones output by the PYTHIA8 simulation but a "topological" version of them, constructed from the separable decay vertices in the decay chain. In practice, this amounts to a transformation of the ground truth decay chain, removing the ancestors that either correspond to very-short-lived resonances or do not have enough charged-particle descendants to allow the formation of a vertex. From a technical perspective, the GNN module performs a multi-class classification on the edges. The model is trained using multi-class cross-entropy loss function outputting a score associated to the "topological" LCA relation between the two connected particles, e.g. particles that share an ancestor at the lowest level will have 1st order LCA (class-1), particles that share an ancestor at the next-to-lowest level will have 2nd order LCA (class-2), etc. The fraction of edges with a ground-truth order larger than 3 in the simulation is very small, so the target classes considered are class-1, class-2 and class-3. In addition, a class with an LCA value of 0 is included (class-0), to identify the case in which the two particles do not originate from the same decay chain. As a side product of the addition of this last class, the LCAI provides a final step of node filtering, by allowing to remove fully disconnected particles (those whose edges are all predicted to have an LCA value of 0).

Each of the previous modules uses independent MLPs for the node-, edge- and global-update functions introduced in Sect. "Usage of Graph Neural Networks". Each MLP is composed of a certain number of layers, all of which have the same latent size. The number of GN block iterations is also configured separately for each module. The hyperparameters chosen for this prototype are written in Table 3.

The output of the DFEI processing chain can be directly translated into a set of selected charged reconstructed

Table 4 Cumulative average efficiencies on the total number of nodes and edges in the graph after each pre-filtering step, illustrating the graph reduction power achieved in each case

Filtering step	Node eff	Edge eff
Edge pre-selection	$\sim$100 %	$\sim$89 %
NP	$\sim$29 %	$\sim$6 %
EP	$\sim$27 %	$\sim$2 %

particles and their inferred ancestors, with the predicted hierarchical relations amongst them.

Training

The training is done in stages, following the algorithm sequence. Each model is trained in a supervised way, using a weighted cross-entropy as loss function, where the weights (corresponding to the inverse of the number of elements in each true class) compensate for the imbalance across classes present in the dataset. The minimisation is done using Adam with the hyperparameter configuration reported in Table 3.

Thresholds are defined for the output score of the NP and EP models, as those resulting into a $\sim$ 99% efficiency of selecting the desired nodes and edges, respectively. This loose requirement is chosen to minimise the potential negative impact on the performance of the subsequent steps. The working point corresponds to a $\sim$ 70% background rejection power for nodes from the NP algorithm and a $\sim$ 68% background rejection power for edges from the EP algorithm. In this setup, the ROC AUC for the NP module is 0.977, and the one for the EP module is 0.974. A consistent performance is observed between the training and test samples, showing no overtraining of these modules. The average reduction of the total event size after each processing step is shown in Table 4.

The training of the LCAI module requires significantly more training iterations than the previous steps, given the much higher complexity of the task. A certain level of over-training is found for the least populated classes, and the training is stopped once the average classification accuracy for the test sample reaches a plateau. Since the goal of this paper is demonstrating the feasibility of the approach by presenting a first working prototype, rather than obtaining the maximum possible performance, we leave the improvements in the training as future work.

Results

In this section, the performance of the current DFEI prototype is described, both at an event level (relevant for trigger) and at an individual-decay-chain level (relevant for trigger and offline analysis).

Event-Level Performance

Different metrics are defined and evaluated in the following section to characterise the performance of DFEI at event level, from multiple perspectives.

Event-size-reduction capabilities. Three different quantities are studied, as a function of the particle multiplicity per event: efficiency of selecting particles from a b-hadron (H_b) decay, efficiency of rejecting particles from the rest of the event (background), and total number of selected particles in the event. The obtained values are shown in Figs. 3 and 4. The average efficiency for selecting particles truly produced in b-hadron decays is 94%, and the average background rejection power is 96%. The selection efficiency of
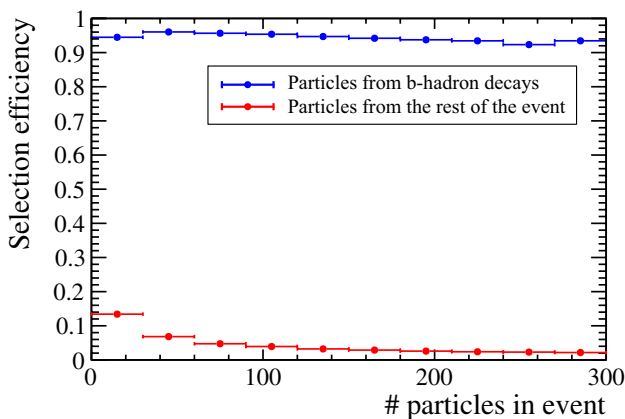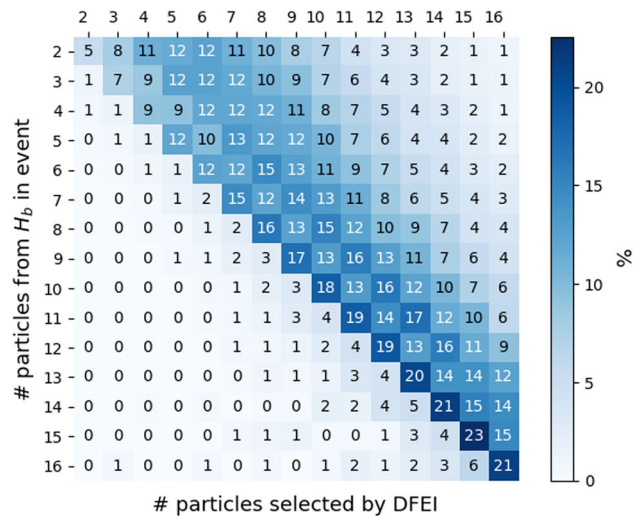
Fig. 4 Confusion matrix for the true vs. predicted number of particles from b-hadron decays per event, computed in terms of percentages normalised for each row (true value) and shown for the square subregion corresponding to a number of particles between 2 and 16

particles from b-hadron decays is found to be independent on the total number of particles in the event. The average number of selected particles per event is ~10, from the initial number of ~140. A good event reduction is obtained irrespectively of the number of particles originating from b-hadron decays per event, as demonstrated by the linear behaviour of the confusion matrix presented in Fig. 4. For the set of selected particles per event, an average purity of 60% is found, defined as the number of selected particles that truly originate from b-hadron decays over the total number of selected particles.

Quality of the decay-chain reconstruction. Apart from helping in background suppression in offline analysis, being able to accurately reconstruct and classify the decay chains in an event can allow DFEI to allow a further level of automation to the LHCb trigger, as introduced in Sect. "Introduction".

A first metric that can serve for characterising the overall understanding of the event in this regard, and be used for benchmarking purposes, is the fraction of events in which DFEI achieves a perfect event reconstruction (PER). For an event to fulfil this condition, all the b-hadron decays in the event need to have been found, all the charged reconstructed particles produced in them been selected, the associated "topological" decay chains been exactly reconstructed, and all the particles from the rest of the event been removed. An example of a PER case found by DFEI in the evaluation dataset is shown in Figs. 5 and 6, from the points of view of the ancestor-chain reconstruction and of the reconstructed-particle

Fig. 3 Average particle-selection efficiency as a function of the total number of particles per event, shown separately for (blue) particles originating from a b-hadron decay and (red) particles from the rest of the event
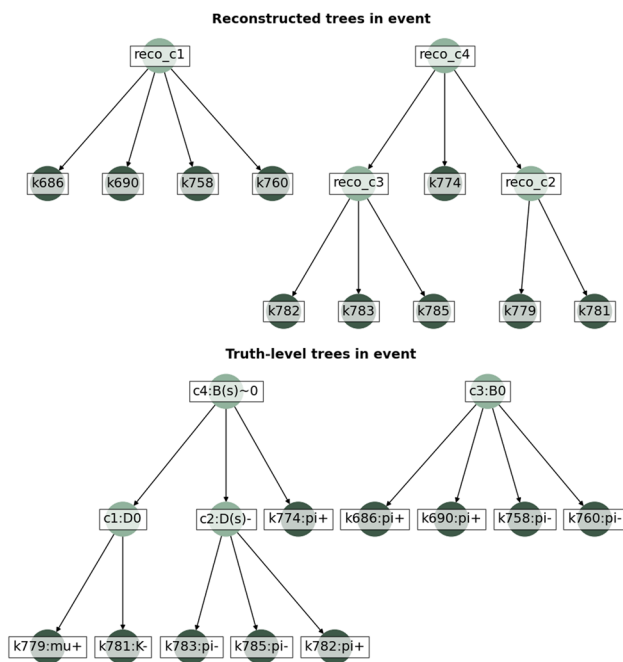
Fig. 5 Example of a PER from the evaluation dataset. The (top) reconstructed and (bottom) ground truth b-hadron decay chains in the event are shown. Apart from the reconstructed particles produced in those decays, the event contains 106 particles from the rest of the event (not shown for simplicity), all of which are correctly removed by DFEI. The dark-green (light green) circles represent the reconstructed particles (topological ancestors). The key (k) numbers correspond to unique identifiers for each reconstructed particle produced in the simulation. The cluster (c) numbers correspond to unique identifiers assigned to each ancestor during the construction of the decay chains. The true identity of the particles is shown in the ground-truth case
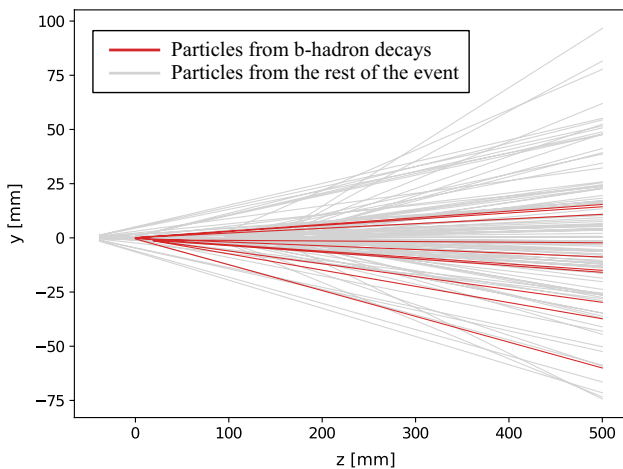
Fig. 6 Example of a PER from the evaluation dataset, same as in Fig. 5. Two-dimensional view of the charged reconstructed particle trajectories in the proton–proton interaction region. Red lines represent particles produced in b-hadron decays, that DFEI has correctly selected, and gray lines represent particles from the rest of the event, that DFEI has correctly removed

filtering, respectively. The average fraction of PER found in the evaluation dataset is $(2.14 \pm 0.07)\%$.

It should be noted that the PER is an extremely challenging case, and that even a partially good reconstruction can be used for trigger purposes. For example, the selection of extra particles from the rest of the event will break the conditions for a PER, but will not impact the efficiency for selecting all the particles produced in b-hadron decays.

Decay-Level Performance

The performance shown in the previous section refers inclusively to all the heavy-hadron decays per event. For each of them, it considers an average over all the known b-hadron species and their known decay types. In this section, the DFEI output is processed to obtain predictions for individual decays. First, all the true decay chains of a certain type are identified in the simulation dataset, taking note of the events in which they were produced. Then, DFEI is run for each of those events, outputting a set of candidate decay chains (connected sub-graphs) per event. Each true decay chain is finally compared with the corresponding set of candidate decay chains, to classify the DFEI reconstruction into one of the following mutually exclusive categories:

- Perfectly reconstructed decay: all the reconstructed particles originating from the b-hadron decay have been predicted to be part of the same connected sub-graph, which is disconnected from all the other particles in the event, and the "topological" ancestor decay chain has been perfectly reconstructed.
- Wrong hierarchy: same as before, but there is at least one mistake in the reconstruction of the "topological" ancestor decay chain.
- Not isolated: all the reconstructed particles originating from the b-hadron decay have been predicted to be part of the same connected sub-graph, but there is at least one extra particle from the rest of the event which is also contained in that sub-graph. This category does not consider the specific "topological" decay chain reconstruction of the sub-graph, and is solely based on the association with extra particles.
- Partially reconstructed: not all of the reconstructed particles originating from the b-hadron decay have been predicted to be part of the same connected sub-graph. As before, this category does not consider the specific "topological" decay chain reconstruction, and is solely based on the impossibility to group all the desired particles in a single sub-graph. It should be noted that this type of reconstruction does not necessarily imply an overall inefficiency in selecting the particles from the b-hadron decay since they can have been selected in multiple sub-graphs.

Table 5 Decay-level performance of DFEI for the inclusive (H_b) case and for several exclusive decay types

Decay mode	Perfect (%)	Wrong hierarchy (%)	Not iso. (%)	Part. reco. (%)
Inclusive H_b decay	4.6 ± 0.1	5.9 ± 0.1	76.0 ± 0.2	13.4 ± 0.1
$B^0 \to K^{*0}[K^+\pi^-]\mu^+\mu^-$	35.8 ± 0.7	19.2 ± 0.6	44.9 ± 0.7	<0.02
$B^0 \to K^+\pi^-$	38.0 ± 0.7	–	54.7 ± 0.7	7.2 ± 0.4
$B_s^0 \to D_s^-[K^-K^+\pi^-]\pi^+$	32.8 ± 0.7	7.1 ± 0.4	53.7 ± 0.8	6.4 ± 0.4
$B^0 \to D^-[K^+\pi^-\pi^-]D^+[K^-\pi^+\pi^+]$	22.7 ± 0.6	22.4 ± 0.6	54.9 ± 0.8	<0.02
$B^+ \to K^+K^-\pi^+$	35.7 ± 0.7	10.2 ± 0.4	46.4 ± 0.7	7.7 ± 0.4
$\Lambda_b^0 \to \Lambda_c^+[pK^-\pi^+]\pi^-$	21.7 ± 1.0	8.9 ± 0.7	36.8 ± 1.2	32.6 ± 1.1
$B_s^0 \to J/\psi[\mu^+\mu^-]\phi[K^+K^-]$	26.9 ± 0.6	20.5 ± 0.5	52.5 ± 0.6	<0.02

In the cases in which the fraction is measured to be zero, the frequentist Wilson upper limit [49] for a 68% coverage is provided. In the other cases, we show the binomial uncertainty related to fraction of events

(a) True decay chain.

(b) Perfect reconstruction.

(c) Wrong hierarchy.

(d) Not isolated.

(e) Partially reconstructed.
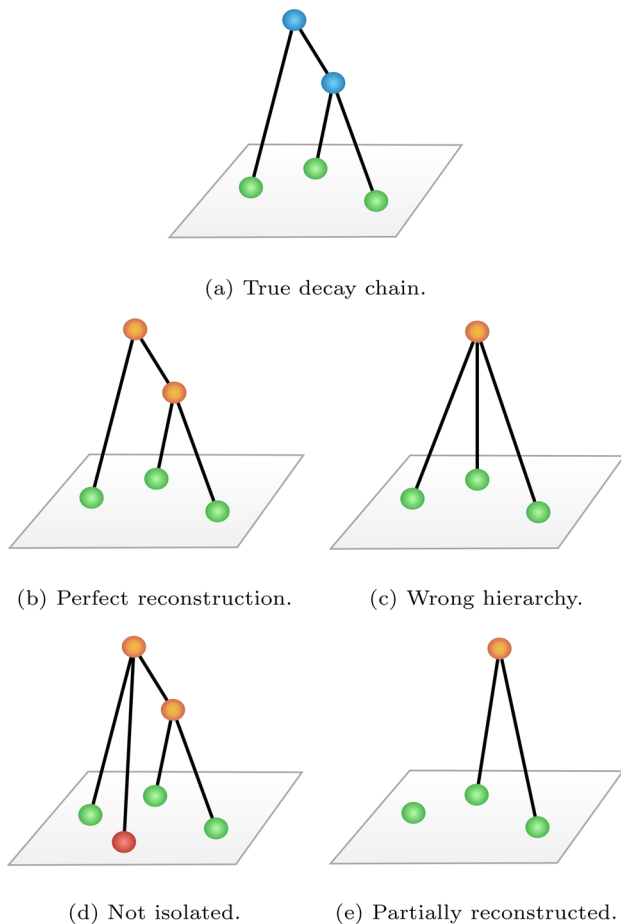
Fig. 7 Examples for the types of decay-chain reconstruction defined in the text. Green (red) nodes represent particles created in the true decay chain (from the rest of the event). Blue (orange) nodes represent true (candidate) ancestors

Examples of the different types of reconstruction for a given true decay chain are shown in Fig. 7.

The decay-level performance is first computed in an inclusive way using the evaluation dataset, by measuring individually the response for all the b-hadron decays contained in the simulation and then looking at the fraction of decays reconstructed in each of the four possible categories. The numbers are reported in Table 5. Complementary to the inclusive case, the DFEI response is evaluated in a second stage restricted to specific decay types, by using the additional datasets introduced in Sect. "Dataset". The resulting numbers are also reported in Table 5. Those modes are representative of the most typical case studies of LHCb, with the inclusive sample also containing decays to many particles and more complicated decay topologies, for which the reconstruction is more challenging.

The performance evaluated on the exclusive modes is significantly better than the inclusive case, with fractions of perfectly reconstructed decays in the range 20–40%. The comparative study of the performance on the different exclusive modes helps to understand which cases are easier or harder for DFEI to reconstruct, and in general to analyse the dependencies of the DFEI response. The most complicated cases are found to be $B^0 \to D^-[K^+\pi^-\pi^-]D^+[K^-\pi^+\pi^+]$ (with two three-particle vertices very separated in space, given the long lifetime of the D^+ meson), and $\Lambda_b^0 \to \Lambda_c^+[pK^-\pi^+]\pi^-$ (with a single π^+ that needs to be associated to a spatially separated three-particle vertex. The difference in performance between the second of the previous decays and $B_s^0 \to D_s^-[K^-K^+\pi^-]$, which has a similar topology, is due to the Λ_c^+ flying more on average than the D_s^-, due to a significantly larger Lorentz boost. The fraction of partial reconstruction is below 10% in all the exclusive cases except for the $\Lambda_b^0 \to \Lambda_c^+[pK^-\pi^+]\pi^-$ decay, which translates into an efficiency for selecting all the reconstructed particles produced in those decays above 90%.

Timing Studies

Detailed timing studies and an optimisation of the inference speed of the DFEI algorithm are out of the scope of this paper, and are left for future research. However, a first, simplified, timing study of the current prototype is shown in this section. The first motivation for the study is to understand
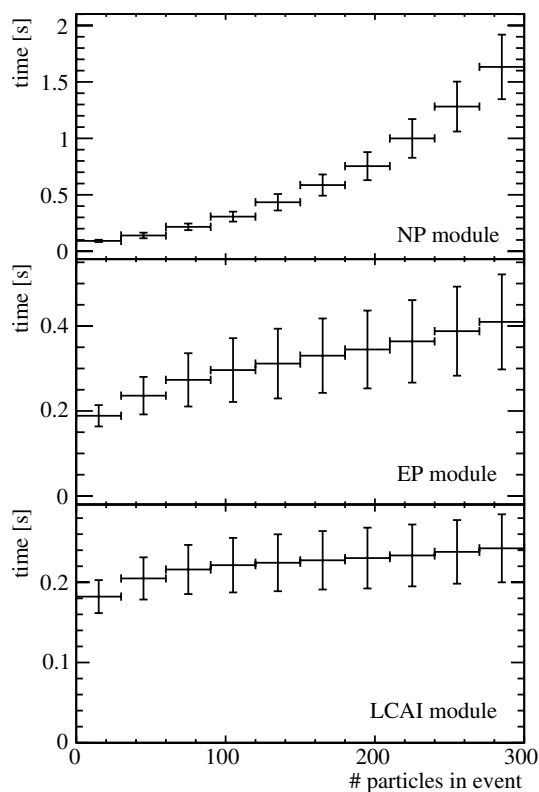
Fig. 8 Average evaluation time per event of the different DFEI modules as a function of the total number of reconstructed particles in the event. The error bars correspond to the standard deviation

the scalability of the response with the object multiplicity per event. The second goal is to estimate how the current event-processing rate achievable by the algorithm compares with the requirements to run DFEI in the LHCb Run 3 trigger. Since the algorithm runs over reconstructed tracks, at the moment the target would be the Run 3 HLT2 trigger, which runs on CPU. As explained in App. B, this would imply a processing rate in the ballpark of 500 Hz per computing node (the precise target number would depend on internal LHCb considerations).

The timing study is done on a CentOS Linux 7 (Core) x86 architecture, using a 2.2 GHz Intel Core Processor (Broadwell, IBRS). No parallelisation scheme is employed. The average computing time required for the evaluation of the NP, EP and LCAI modules as a function of the total number of particles in the event is computed and reported in Fig. 8. In this configuration, the NP is both the slowest module and the one that presents the strongest scaling as a function of the event size, hence the one that can profit the most from a future optimisation in terms of timing. The average of the combined NP + EP + LCAI times is approximately 1 s per event.

The time needed to create the input graph[3] for each of the three modules and to post-process their output (i.e. filtering nodes and edges and interpreting the predicted LCA values in terms of reconstructed decay chains) is not included in the previous study. From all these auxiliary tasks, the only one that does not have a processing time significantly under 1 s is the graph construction of the NP module, that requires an average of 2 s per event.

Taking into account these first studies, a strategy to speed up the full algorithm in order to meet the trigger constraints is outlined in Sect. "Future Work".

Discussion

The proposed approach for a multi-heavy-hadron-decay reconstruction of b-hadron decays in a hadronic environment is the first of its kind. To allow the benchmarking of future efforts in this new scenario, all the datasets used for the training and evaluation performance of DFEI have been made publicly available [46]. In this section, the performance obtained with this first prototype is discussed, in reference to the global context.

On a first stage, the reconstructed-particle selection capabilities can be compared with previous studies in LHCb. The closest case study, reported in Ref. [10], considers the subset of reconstructed particles that have been selected by a standard LHCb inclusive trigger algorithm, and attempts to discern whether each of the other particles in the event has been produced in the same b-hadron decay or not. By combining vertex-quality requirements and the output of a multivariate algorithm trained on individual-particle features, the authors estimate an approximate selection efficiency for particles from the same b-hadron decay of 90% for an approximate background rejection power of 90%. That study is based on official LHCb simulation, which contains material-interaction backgrounds and fake-track backgrounds, not included in the simulated dataset used in this paper. Both simulations, however, aim at representing inclusive b-hadron decays in LHCb Run 3-like conditions. The performance of DFEI (94% selection efficiency for particles from b-hadron decays and 96% background rejection power) is similar and numerically higher, within the caveats of the comparison. Most importantly, it shows a powerful discrimination consistently for all the b-hadron decays present in the event at the same time, instead of focusing on an individual decay. It should be noted that the strategy presented in Ref. [10] is not used in production by LHCb. The difference between the two approaches will only increase in the much harsher object-multiplicity

---

[3] The values of the input features are assumed to be already available at the time DFEI is evaluated, as is the case in the datasets used in this paper.

conditions expected for LHCb Upgrade II. The almost flat response found in DFEI for the particle-selection efficiencies as a function of the number of particles in the event also suggests good prospects for the Upgrade II conditions.

On a second level, regarding decay-chain reconstruction, DFEI has demonstrated for the first time that this kind of reconstruction can be done successfully both in a hadronic environment and in a multi-decay-chain scenario. Given the novelty of the approach, the performance at this level can only be partially compared with the one achieved by the FEI algorithm at the Belle II experiment, and with significant caveats. On one side, as explained in Sect. "Introduction", the reconstruction in LHCb is a much more difficult task than in Belle II. On the other side, the DFEI prototype for LHCb makes use of several, previously introduced, simplifications: omitting particles produced outside the geometrical acceptance, not including neutral reconstructed particles and reconstructing only the "topological" decay chains, not the full ones. Keeping the previous caveats in mind, the fraction of perfect decay reconstruction obtained in this paper can be approximately compared to the so-called tag-side reconstruction efficiency determined in Ref. [32] using a Belle simulated dataset, which is of the order of few per cent for semileptonic decays and of few per mille for hadronic decays. The conclusion that can be drawn from this comparison is that DFEI manages a level of reconstruction of decay chains in a hadronic environment which is in the ballpark of that achieved in Belle (II), hence demonstrating not only the feasibility but also the competitiveness of a Full Event Interpretation approach at the LHC.

Concerning offline analysis applications, a study of different possible types of DFEI reconstruction for specific ground truth decay chains is reported in Sect. "Decay-Level Performance". A technically similar but conceptually different study could be done on a collision dataset, focusing this time on the DFEI prediction for the reconstructed particles output by any standard LHCb analysis preselection. Those preselections aim at identifying particles that are compatible with having been produced in a specific type of decay chain, which are denoted as signal candidates. The DFEI output can be used to classify each signal candidate in one of the following categories: "signal" (if the reconstruction matches the expected decay chain), background with a different resonance structure (if the selected reconstructed particles are deemed to be correct but the predicted hierarchy is not), background from decays with extra particles (some of which are not part of the signal candidate) and combinatorial background (where the candidate particles are predicted to originate from multiple sources). This implies that DFEI could virtually be used in every LHCb analysis to suppress/study the different possible types of contributing backgrounds with a potentially higher background separation power, by leveraging all the information in the event.

Future Work

The work in this paper opens the door for multiple future research lines. Natural follow up steps are detailed performance studies on official LHCb simulation and Run 3 collision data. This will allow to assess the impact of the DFEI reconstruction in a broad spectrum of decay distributions, to understand the potential needs for further optimisations/calibrations of the algorithm. Another natural continuation is the extension of the developments and studies to Upgrade II conditions. Additionally, the DFEI functionality is expected to be expanded, to include neutral reconstructed particles, charm-hadrons decays and particle-identity information. This can bring potential new complementary applications of DFEI, such as providing enhanced flavour-tagging capabilities to LHCb.

Regarding speeding up the inference, a design optimisation of the NP module, for example substituting the GNN by a combination of independent multi-variate classifiers per particle or by a nearest-neighbour selection in a learnt embedding, together with an overall hyperparameter optimisation can bring large reductions to the evaluation time. Significant additional speed ups can be gained by converting the full DFEI pipeline into C++ [50, 51] (which is by itself a technical requirement to run DFEI in the current LHCb trigger). The combination of the suggested improvements gives a good hope to achieve the target event-processing rate discussed in Sect. "Timing Studies". Finally, regarding the utilisation of DFEI in the LHCb Upgrade II, the inference of the GNN modules could become much faster by the usage of GPUs [41, 42, 50] or FPGAs [37–40] as hardware accelerators in the trigger system. For example, in Ref. [50], graphs of order 100,000 nodes are segmented with GNNs (in a technically similar way to this work) in less than 1 s.

Conclusion

This is the first proof of concept for an inclusive event processing at the LHC in a high-multiplicity environment focused on the identification and explicit reconstruction of all the heavy-hadron decay chains in the event. It is heavily based on deep learning and uses GNNs to optimally capture the event structure. To keep the approach computationally scalable, the algorithm is divided into three stages: node pruning removes all the nodes that are not associated with a heavy-hadron decay, edge pruning removes all the edges that do not share the same ancestor inference and finally the lowest-common-ancestor that predicts the hierarchical decay relations of particles, allowing to completely reconstruct all decays. The algorithm has been trained using a simulated dataset that emulates LHCb Run 3 conditions, and

is specialised for beauty hadron decays and charged reconstructed particles.

The algorithm is able to separate between particles originating from b-hadron decays and those from the rest of the event better than previous approaches in similar conditions at LHCb. The resulting fraction of perfectly reconstructed b-hadron decay chains is in the ballpark of the one obtained by the FEI algorithm in an electron-positron environment, showing not only the feasibility but also the competitiveness of this approach at the LHC.

The performance of DFEI is studied in detail both at the global event level and at the individual b-hadron decay level, using both inclusive and exclusive samples containing typical decays of interest for LHCb. A particularly good performance is found in the exclusive modes, in terms of both the efficiency of a perfect decay-chain reconstruction (in the range 20–40%) and the efficiency to identify all the reconstructed particles originated from the decay (above 90% in most of the cases).

The application of the algorithm for data analysis at offline level is discussed, explaining how DFEI can be used as a common tool to identify and classify different types of background. These capabilities can already be explored with the Run 3 dataset, which is currently being collected. In terms of charged reconstructed particles, the current DFEI algorithm achieves a 14× event reduction factor in Run 3 conditions, for a 94% efficiency in the selection of particles from b-hadron decays in the event. For illustration, if this kind of performance was achieved in Upgrade II conditions and all the event information was solely related to charged particles, the saving factor would translate into a 14× larger integrated luminosity that could be recorded, compared to storing the full event information. This shows the strong potential of the DFEI approach, while accurate estimates of the gaining factor in Upgrade II conditions will be the focus of future research.

To be used in the trigger, the DFEI algorithm needs to be able to process events at high rate. A first timing study of the DFEI algorithm is performed and several steps towards achieving the target event-processing rate are identified.

Finally, the successful development of the DFEI prototype opens the door to future research towards expanding its functionality and use cases in LHCb and can inspire similar developments in other LHC experiments for the HL-LHC Phase.

Appendix A: Construction of the dataset

To approximately emulate the topology of a Run 3 event in LHCb, single proton-proton collisions at a centre-of-mass energy of 13 TeV are first generated with PYTHIA8, using an inclusive `softQCD` interaction model. Then, several collisions are combined in each event, such that their number follows a poisson distribution with an average of 7.6, corresponding to the average multiplicity expected in Run 3 conditions [52]. For all the studies done on inclusive b-hadron decays, at least one collision that has produced b-hadrons is included in each event. For the studies done on exclusive decays of interest, the dataset generated in the previous configuration is reused, substituting a collision that produced an inclusive b-hadron decay by a new one, that produced the specified exclusive decay. This new collision is generated by combining the PYTHIA8 and EvtGen generators.

To place those collisions in the space, a coordinate system is defined with its centre on the nominal collision point, considered here to be the centre of the Vertex Locator of LHCb. The true position of each proton-proton collision point is sampled from a three-dimensional gaussian distribution, centred on the origin of coordinates, with widths of 0.05 mm, 0.05 mm and 10 mm, respectively along the x, y and z axes (see Ref. [52] for discussions on the expected beam geometry).

Charged stable particles (pions, kaons, protons, electrons and muons) produced in the collisions are only kept if their pseudorapidity is in the range $1.9 \leq \eta \leq 4.9$, corresponding to the geometric acceptance of LHCb [1], and if their origin position along the z direction is within a distance of ±500 mm from the origin of coordinates, which emulates an approximate coverage of the Vertex Locator [53]. The measurement of the relevant properties for each particle performed by the LHCb detection and reconstruction process is emulated by modifying the particle properties generated by PYTHIA8, as discussed in the following.

As a first step, the measurement of primary vertices is considered. Those which result in a number of charged particles less than four are considered not to be reconstructible, and hence are discarded. For all the others, a gaussian smearing is applied to their position in each of the three dimensions. The resolution for that smearing as a function of the total number of charged particles in the collision is assumed to be the same as the one measured by LHCb in Run 2 as a function of the number of tracks, that is reported in Fig. 5 of Ref. [54]. The resolution for the x and y coordinates is assumed to be identical.

In a second step, the determination of the origin point of each particle is studied, which in real life would correspond to the measurement of the position of the first hit from the associated track in the Vertex Locator. The Vertex Locator is segmented into 52 measurement planes along the z-direction [53], which in this study are approximated to be equally spaced for simplicity. The z-coordinate of the origin point is therefore assigned to that of the closest plane to the true origin position, looking in the positive direction of the z axis. The values for the x and y coordinates are determined by obtaining the true position of the particle at the given z

plane, assuming constant velocity, and applying a gaussian smearing with a resolution of 8.5 μm in both the *x* and *y* directions (see Ref. [55] for discussions on expected resolutions in Run 3 conditions).

Finally, the measurement of the three-momentum of the particles is emulated. The momentum slope in the *x* and *y* directions relative to the *z* axis is smeared using a gaussian function, using the momentum-dependent resolution reported in Fig. 1 of Ref. [55]. The modulus of the momentum is smeared using a gaussian function assuming a relative resolution of 0.4% [56].

It should be noted that this emulation does not include additional particles produced in material interactions within the detector or fake particles resulting from wrongly reconstructed tracks, both of which are present in the official LHCb simulation.

Appendix B: Data taking conditions at LHCb Run 3

In Run 3, the trigger system of LHCb is fully software based and composed of two consecutive levels, HLT1 and HLT2 [57]. The first level performs a partial reconstruction of charged particles, reconstructs primary vertices and performs muon identification, pre-selecting the events to reduce a 30 MHz input rate to 1 MHz [58]. The raw event information for each passing event is temporarily written to a disk buffer, which allows to perform the next trigger steps in an asynchronous way, and hence with an enlarged computing-time budget. Running on the data in the disk buffer, the HLT2 level performs a full reconstruction of the objects in each event, followed by a combination of inclusive and exclusive selections. Those selections primarily identify interesting events, but can also be used to decide which elements inside them (particles, raw-event information, etc.) will be stored for future processing [10]. The HLT2 output is saved on a permanent tape storage. Before being moved to disk storage, the only one which is accessible for data analysis, the events on the tape undergo a further filtering stage offline [59].

The DFEI algorithm can ideally be run at the HLT2 stage, if its inference is fast enough. This would imply event-processing frequencies per computing node in the ballpark of the current HLT2 sequence, which amounts to around 500 Hz excluding selection algorithms [60]. Determining the accurate time requirements is outside the scope of this work, since it would require including the algorithm in the LHCb HLT2 sequence and testing it in realistic data-taking conditions. It should be noted that, if this requirement eventually turned out to be too difficult, DFEI could run instead in the Run 3 offline filtering stage before data is sent to disk storage. This would however require persisting the information of all the reconstructed particles to tape.

Data availability All the simulation datasets used for the studies reported in this paper are available at https://doi.org/10.5281/zenodo.7799170.

Declarations

Competing interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Alves AA Jr et al (2008) The LHCb detector at the LHC. JINST 3:S08005
2. Aaij R et al (2018) Physics case for an LHCb Upgrade II—opportunities in flavour physics, and beyond, in the HL-LHC era. arXiv:1808.08865
3. Albrecht J, et al (2019) Luminosity scenarios for LHCb Upgrade II. Tech. Rep., CERN, Geneva. http://cds.cern.ch/record/2653011
4. Fazzini D (2018) Flavour tagging in the LHCb experiment. *PoS* LHCP 2018, 230
5. Aaij R et al (2017) Measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+\mu^-$ decays. Phys Rev Lett 118:191801 arXiv:1703.05747
6. Aaij R et al (2018) Test of lepton flavor universality by the measurement of the $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays. Phys Rev D 97:072013 arXiv:1711.02505

7. Gligorov VV, Williams M (2013) Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. JINST 8:P02013 arXiv:1210.6861

8. Likhomanenko T et al (2015) LHCb topological trigger reoptimization. J Phys Conf Ser 664:082025

9. Fitzpatrick C, Gligorov VV (2014) Anatomy of an upgrade event in the upgrade era, and implications for the LHCb trigger. Tech. Rep., CERN, Geneva. http://cds.cern.ch/record/1670985

10. Aaij R et al (2019) A comprehensive real-time analysis model at the LHCb experiment. JINST 14:p04006 arXiv:1903.01360

11. Likhomanenko T, Derkach D, Rogozhnikov A (2016) Inclusive flavour tagging algorithm. J Phys Conf Ser 762:p012045 arXiv:1705.08707

12. Bourgeois D, Fitzpatrick C, Stahl S (2018) Using holistic event information in the trigger. LHCb-PUB-2018-010. arXiv:1808:00711

13. Sirunyan AM et al (2017) Particle-flow reconstruction and global event description with the CMS detector. JINST 12:P10003 arXiv:1706.04965

14. Aaboud M et al (2017) Jet reconstruction and performance using particle flow with the ATLAS detector. Eur Phys J C 77:466 arXiv:1703.10485

15. Pata J, Duarte J, Vlimant J-R, Pierini M, Spiropulu M (2021) MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks. Eur Phys J C 81:381 arXiv:2101.08578

16. ATLAS Collaboration (2022) Graph Neural Network Jet Flavour Tagging with the ATLAS Detector. ATL-PHYS-PUB-2022-027 https://cds.cern.ch/record/2811135

17. Huang A et al (2023) Heterogeneous graph neural network for identifying hadronically decayed tau leptons at the high luminosity LHC. JINST 18 (2023) 07, P07001 arXiv:2301:00501

18. Ma F, Liu F, Li W (2022) A jet tagging algorithm of graph network with HaarPooling message passing. arXiv:2210:13869

19. Mokhtar F, Kansal R, Duarte J (2022) Do graph neural networks learn traditional jet substructure? In 36th Conference on Neural Information Processing Systems. arXiv:2211.09912

20. Ju X, Nachman B (2020) Supervised jet clustering with graph neural networks for Lorentz boosted bosons. Phys Rev D 102:075014 arXiv:2008.06064

21. Atkinson O et al (2022) IRC-safe graph autoencoder for unsupervised anomaly detection. Front Artif Intell 5:943135 arXiv:2204.12231

22. Murnane D, Thais S, Wong J (2023) Semi-equivariant GNN architectures for jet tagging. J Phys Conf Ser 2438:012121 arXiv:2202.06941

23. Gong S et al (2022) An efficient Lorentz equivariant graph neural network for jet tagging. JHEP 07:030 arXiv:2201.08187

24. Konar P, Ngairangbam VS, Spannowsky M (2022) Energy-weighted message passing: an infra-red and collinear safe graph neural network algorithm. JHEP 02:060 arXiv:2109.14636

25. Verma Y, Jena S (2021) Jet characterization in heavy ion collisions by QCD-Aware Graph. Neural Netw 2103:14906

26. Dreyer FA, Qu H (2021) Jet tagging in the Lund plane with graph networks. JHEP 03:052 arXiv:2012.08526

27. Guo J, Li J, Li T, Zhang R (2021) Boosted Higgs boson jet reconstruction via a graph neural network. Phys Rev D 103:116025 arXiv:2010.05464

28. Qu H, Gouskos L (2020) ParticleNet: jet tagging via particle clouds. Phys Rev D 101:056019 arXiv:1902.08570

29. Moreno EA et al (2020) JEDI-net: a jet identification algorithm based on interaction networks. Eur Phys J C 80:58 arXiv:1908.05318

30. Shlomi J et al (2021) Secondary vertex finding in jets with neural networks. Eur Phys J C 81:540 arXiv:2008.02831

31. Kahn J et al (2022) Learning tree structures from leaves for particle decay reconstruction. Mach Learn Sci Technol 3:035012. https://doi.org/10.1088/2632-2153/ac8de0

32. Keck T et al (2019) The full event interpretation: an exclusive tagging algorithm for the Belle II experiment. Comput Softw Big Sci 3:6 arXiv:1807.08680

33. Tsaklidis I, Goldenzweig P, Ripp-Baudot I, Kahn J, Dujany G (2020) Demonstrating learned particle decay reconstruction using Graph Neural Networks at BelleII. Ph.D. thesis, school-Strasbourg, Université de Strasbourg, Karlsruhe, Strasbourg (2020). Presented on 19 06 2020

34. Ju X, et al. (2020) Graph neural networks for particle reconstruction in high energy physics detectors. In 33rd Annual Conference on Neural Information Processing Systems. arXiv:2003.11603

35. Shlomi J, Battaglia P, Vlimant J-R (2020) Graph neural networks in particle physics. Mach Learn Sci Technol 2:021001. https://doi.org/10.1088/2632-2153/abbf9a

36. Thais S, et al. (2022) Graph neural networks in particle physics: implementations, innovations, and challenges. In Snowmass 2021. arXiv:2203.12852

37. Que Z, et al. (2022) LL-GNN: low latency graph neural networks on FPGAs for particle detectors. arXiv:2209.14065

38. Elabd A (2022) et al. Graph neural networks for charged particle tracking on FPGAs. Front Big Data 5: 828666. arXiv:2112.02048

39. Heintz A et al (2020) Accelerated charged particle tracking with graph neural networks on FPGAs. In 34th Conference on Neural Information Processing Systems. 2012.01563

40. Iiyama Y et al (2020) Distance-weighted graph neural networks on FPGAS for real-time particle reconstruction in high energy physics. Front Big Data 3:598927 arXiv:2008.03601

41. Ju X et al (2021) Performance of a geometric deep learning pipeline for HL-LHC particle tracking. Eur Phys J C 81:876 arXiv:2103.06995

42. Pata J et al (2023) Machine learning for particle flow reconstruction at CMS. J Phys Conf Ser. 2438:012100 arXiv:2203.00330

43. Albertsson K et al (2018) Machine learning in high energy physics community white paper. J Phys Conf Ser 1085:022008 arXiv:1807.02876

44. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE (2017) Neural message passing for quantum chemistry. In Precup D, Teh YW (eds) Proceedings of the 34th international conference on machine learning, vol. 70 of Proceedings of machine learning research, 1263–1272 (PMLR, 2017). https://proceedings.mlr.press/v70/gilmer17a.html

45. Battaglia P, et al. (2018) Relational inductive biases, deep learning, and graph networks. arXiv. https://arxiv.org/pdf/1806.01261.pdf

46. Pardiñas JG, et al. (2023) Dataset of paper "GNN for deep full event interpretation and hierarchical reconstruction of heavy-hadron decays in proton-proton collisions". https://doi.org/10.5281/zenodo.7799170

47. Bierlich C, et al. (2022) A comprehensive guide to the physics and usage of PYTHIA 8.3. arXiv:2203.11601

48. D. J. Lange (2001) The EvtGen particle decay simulation package. Nucl. Instrum. Meth. A462 (2001) 152

49. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22:209

50. Lazar A et al (2023) Accelerating the Inference of the Exa. TrkX Pipeline. J Phys Conf Ser 2438:012008 arXiv:2202.06929

51. An S et al (2023) C++ code generation for fast inference of deep learning models in root/tmva. J Phys Conf Ser 2438:012013. https://doi.org/10.1088/1742-6596/2438/1/012013

52. LHCb collaboration (2013) LHCb VELO Upgrade Technical Design Report. Tech. Rep. CERN-LHCC-2013-021, LHCB-TDR-013 https://cds.cern.ch/record/1624070

53. Hennessy K (2016) Lhcb velo upgrade. Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment 845: 97–100 (2017). https://www.sciencedirect.com/science/article/pii/S016890021630290X. Proceedings of the Vienna Conference on Instrumentation

54. Aaij R et al (2019) Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. JINST 14:P04013 arXiv:1812.10790

55. Billoir P, De Cian M, Günther PA, Stemmle S (2021) A parametrized Kalman filter for fast track fitting at LHCb. Comput Phys Commun 265:108026 arXiv:2101.12040

56. LHCb tracker upgrade technical design report (2014)

57. LHCb trigger and online upgrade technical design report (2014)

58. LHCb Collaboration CM (2020) LHCb Upgrade GPU high level trigger technical design report. Tech. Rep., CERN, Geneva. https://cds.cern.ch/record/2717938

59. Skidmore N, Rodrigues E, Koppenburg P (2022) Run-3 offline data processing and analysis at LHCb. *PoS* EPS-HEP2021, 792

60. HLT2 reconstruction throughput and Forward Tracking performance for Run 3 of LHCb (2022). http://cds.cern.ch/record/2810226