GNN for Deep Full Event Interpretation and hierarchical reconstruction of heavy-hadron decays in proton-proton collisions

Julián García Pardiñas\textsuperscript{1,2}, Marta Calvi\textsuperscript{1}, \textbf{Jonas Eschle}\textsuperscript{3}, Andrea Mauri\textsuperscript{4,5}, Simone Meloni\textsuperscript{1}, Martina Mozzanica\textsuperscript{1}, Nicola Serra\textsuperscript{3}

\textsuperscript{1} University and INFN Milano-Bicocca (Italy)  
\textsuperscript{2} CERN (Switzerland)  
\textsuperscript{3} University of Zürich, Switzerland  
\textsuperscript{4} Imperial College London (UK)  
\textsuperscript{5} NIKHEF (The Netherlands)
The LHCb detector

Single-arm forward spectrometer, studying the decays of beauty and charm hadrons. Very broad physics program. → To be maintained and expanded in future LHC runs.

Excellent vertexing capabilities, momentum resolution and PID performance.
Current trigger

OR between decay modes

Store whole event*
Event information and size

Current trigger
OR between decay modes
Store whole event
→ Store only PV + Tracks?

Full event VS disk space
Evolution in the LHCb trigger

- \( \ll 1 \) signal
- \( \sim 50 \) tracks

Which events are interesting?

\[ \downarrow \]

Trigger strategy: signal based.

\[
\begin{align*}
\text{Instantaneous luminosity} \\
2010 & \quad 2020 & \quad 2030 & \quad 2040
\end{align*}
\]

- Upgrade I
- Upgrade II

\( \sim 5 \) signals
- \( \sim 1000 \) tracks

Which parts of the event are interesting?

New era for LHCb!

[\text{LHCb-PUB-2014-027}]
Evolution in the LHCb trigger

Which events are interesting?

Which parts of the event are interesting?

Trigger strategy: signal based.

New era for LHCb!

[9LCb-PUB-2014-027]

Fully software trigger, CPU + GPU [JINST 14 (2019) 04, P04006].

Online alignment and calibration, offline-quality online reconstruction.

Key developments that enable more ambitious trigger strategies.
Evolution in the LHCb trigger

**Reconstruction difficulty ~ n²**

- Hits & tracks combinatorics
- Pile-up

Which events are interesting?
Which parts of the event are interesting?

<table>
<thead>
<tr>
<th>Year</th>
<th>Signal</th>
<th>Tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>~ 0.5</td>
<td>~ 140</td>
</tr>
<tr>
<td>2020</td>
<td>~ 5</td>
<td>~ 1000</td>
</tr>
<tr>
<td>2030</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>2040</td>
<td>~</td>
<td></td>
</tr>
</tbody>
</table>

New era for LHCb!

[LHCb-PUB-2014-027]

Trigger strategy:

- Signal based.

Fully software trigger, CPU + GPU [JINST 14 (2019) 04, P04006].

- Online alignment and calibration,
  offline-quality online reconstruction.

Key developments that enable more ambitious trigger strategies.

～ 0.5 signals
～ 140 tracks
Evolution in the LHCb trigger

**Reconstruction difficulty ~ n²**

- track & decay
- Hits & tracks combinatorics
- Pile-up

↓

Trigger strategy: signal based.

**Storage space ~ n²**

- Event size ~ n
- Trigger rate ~ n

*more signal events on avg.*

~ 0.5 signals
~ 140 tracks

Fully software trigger, CPU + GPU [JINST 14 (2019) 04, P04006].

⇒ Online alignment and calibration,
offline-quality online reconstruction.

Key developments that enable more ambitious trigger strategies.
Evolution in the LHCb trigger

Which events are interesting?

\[
\ll 1 \text{ signal} \\
\sim 50 \text{ tracks}
\]

\[\downarrow\]

Trigger strategy: signal based.

\[\sim 0.5 \text{ signals} \\
\sim 140 \text{ tracks}\]

\[\Rightarrow \text{Online alignment and calibration,} \]
\[\text{offline-quality online reconstruction.}\]

\[\Rightarrow \text{Fully software trigger, CPU + GPU} \textbf{[JINST 14 (2019) 04, P04006].}\]

\[\Rightarrow \text{New era for LHCb!} \textbf{[LHCb-PUB-2014-027]}\]

\[\Rightarrow \text{Which parts of the event are interesting?}\]

\[\Rightarrow \text{Key developments that enable more ambitious trigger strategies.}\]
Fighting the scaling

Density of proton bunches

2010  2020  2030  2040

Upgrade I  Upgrade II

Need for ML

≪ 1 signal per event
Low noise

Multiple signals per event
High noise
Facing the new era with machine learning

Novel approach proposed

DFEI: Deep-learning based Full Event Interpretation

“Maximally efficient” trigger.
Facing the new era with machine learning

Novel approach proposed

DFEI: Deep-learning based Full Event Interpretation

“Maximally efficient” trigger.

Similar developments in other experiments

Full Event Interpretation algorithm at an e+e- collider

GNNs for trigger purposes
Facing the new era with machine learning

Novel approach proposed

DFEI: Deep-learning based Full Event Interpretation

Event

Input

Graph neural

Target

Belle II case

Only $B^0/B^\pm$ hadrons.
e+e- environment.
Hermetic detector.

LHCb case

All b,c-hadron species.
pp environment.
Non-hermetic detector.

Similar developments in other experiments

Full Event Interpretation algorithm at an e+e- collider

GNNs for trigger purposes
First prototype of DFEI for LHCb, focused on b-hadron decays and charged stable particles.
Decays and graph structures

Decay

Global \((n \text{ vars})\): event information
\(n\text{Tracks}, \ldots\)

Nodes \((m_{\text{event}} \text{ vars})\): track variables
\(\text{momentum}, (\text{PID}), \ldots\)

Edges \((m^2_{\text{event}})\): track relations
\(\text{angle}, \text{DOCA}, \ldots\)

Graph structures

Representation of objects with relations

Arbitrary, sparse/dense relations
Mathematically expression

Adjacency matrix defines graph completely

Comparison of data structure

Images/text special case of graph

GNN: “generalized” CNN

Propagate and transform information
GNN training

e' v' u' updated (by DNN) e, v, u

BLUE updated by BLACK not utilizing GREY

Node Aggregation
Aggregating information from neighbors
GNN training

Node Aggregation
Aggregating information from neighbors

Transformation of graph
Input graph (X), return graph (y)

Output graph

Node/edge/global features
Different interpretations depending on application

e' v' u' updated (by DNN) e, v, u

BLUE updated by BLACK not utilizing GREY

Updated by DNN e, v, u
DFEI Overview

1. Node pruning
2. Edge pruning
3. LCA inference
Input graph construction

Nodes: all the charged particles in the event.
  ➔ On average \(~140\).

Edges: connect particles which are topologically close (see backup for details).
  ➔ On average \(~10\,000\).

\(PV_1, PV_2\): different proton-proton primary vertices.
1\textsuperscript{st} module: node pruning

**Signal nodes:** particles from a b-hadron (any of them)

**Background nodes:** particles from the rest of the event

\textbf{Target}

\textbf{Input}

- \( p_T \): transverse momentum
- \( \text{ETA} \): pseudorapidity
- \( \text{PV} \): associated primary vertex
- \( \text{IP} \): impact parameter with respect to the PV
- \( q \): charge

Opening angle, distance (between origins) along the beam axis, “transverse distance” (see backup), from same PV (boolean)
**Signal edges:** pairs of particles with the same b-hadron ancestor

**Background edges:** any other pair of particles

2\textsuperscript{nd} module: edge pruning

Target

Input

Same as before
3rd module: Lowest Common Ancestor (LCA) inference

From [BELLE2-MTHESIS-2020-006]:
(see also [James Kahn et al 2022 Mach. Learn.: Sci. Technol. 3 035012])

Problem reduced to multi-class classification on edges.
3rd module: Lowest Common Ancestor (LCA) inference

From [BELLE2-MTHESIS-2020-006]:
(see also [James Kahn et al 2022 Mach. Learn.: Sci. Technol. 3 035012])

Problem reduced to multi-class classification on edges.

For the prototype, use as target a simplified version of the decay chain, based on the reconstructible vertices.

• Very-short-lived resonances merged with the previous ancestor.
• Resonances with less than two charged descendants merged with the previous ancestor.
Three momentum, origin-point coordinates, PV coordinates

Same as before

3rd module: Lowest Common Ancestor (LCA) inference

LCA value

0 (unrelated)
1
2
3

B⁺
D⁰
B⁻
3rd module: Lowest Common Ancestor (LCA) inference

Target

Input

LCA value

3 2 1 0 (unrelated)

Three momentum, origin-point coordinates, PV coordinates

Same as before
Overview and training

1. Node pruning
2. Edge pruning
3. LCA inference

Blue: reconstructed ancestors.
Green: particles from a b-hadron
Red: particles from the rest of the event
Dataset:
- PYTHIA-based simulation, Run 3-like conditions, **approximated emulation of LHCb reconstruction.**
- Events required to contain at least one **b-hadron (inclusive decay).**

**PYTHIA configuration**
- Proton-proton collisions at 13 TeV
- Avg. number of collisions per event: 7.6

**Emulated “Reconstruction”**
*using public Run 3 expectations*
- Within LHCb acceptance
- Origin point of the tracks
- Three-momentum of the tracks.
- Position of the primary vertices.

**Dataset published**
*for benchmarking & comparison*
10.5281/zenodo.7799169
Training

Dataset:
- PYTHIA-based simulation, Run 3-like conditions, approximated emulation of LHCb reconstruction.
- Events required to contain at least one b-hadron (inclusive decay).

Filter the data
Choose a threshold (~99% signal eff.)
Train the node-pruning module
Training

Dataset:
• PYTHIA-based simulation, Run 3-like conditions, approximated emulation of LHCb reconstruction.
• Events required to contain at least one $b$-hadron (inclusive decay).
Training

Dataset:
- PYTHIA-based simulation, Run 3-like conditions, approximated emulation of LHCb reconstruction.
- Events required to contain at least one b-hadron (inclusive decay).

Train the LCA inference module

Filter the data

Choose a threshold (~99% signal eff.)

Train the edge-pruning module

Filter the data

Choose a threshold (~99% signal eff.)

Train the node-pruning module

~140 nodes
~10000 edges

~30 nodes
~300 edges

~20 nodes
~75 edges
Motivation

The algorithm

Performance

Outlook

Run3-like conditions
Consistent performance with different number of signals

"single-b-hadron-signal" approach performance comparable to the envisaged nominal LHCb strategy for Run 3 [JINST 14 (2019) 04, P04006]

LHCb: 90% sig eff, 90% bkg rej. power
DFEI: 94% sig eff, 96% bkg rej. power

DFEI capability #1
Powerful event size (~ x14) reduction in a multi-signal environment.
Performance: perfect event reconstruction (PER)

Real example of a perfectly reconstructed simulation event.

Simulated event

DFEI output

• PER efficiency similar to tag-side efficiency for Belle (II)
• Can be easily extended for more target variables!

DFEI capability #2
Automatised and inclusive reconstruction of decay chains.
Timing studies

Scaling

current implementation

Implementation

Currently Python & TensorFlow
flexible for experimenting

TMVA SOFIE implementation (WIP)
Fast Inference System

Possible speed improvements

Simplification of layers, especially first
Approximate convolutions etc.

Hardware accelerators such as FPGA, GPU,...
WIP for GNNs in general
Summary

Increased particle multiplicities for LHCb Upgrades I and II bring big challenges, both for trigger and offline analysis.

Paradigm change in trigger: “which events?” → “which parts of the event?”

Deep-learning based Full Event Interpretation

Online application:
- Safely discard rest of event, minimal loss for analyses
- Hierarchical reconstruction of heavy-hadron decay chains

Offline application:
- Tool for powerful background classification & suppression

First prototype of the DFEI algorithm based on GNN focused on b-hadron decays and charged stable particles. Very promising performance in realistic conditions!
Next steps

Algorithmic optimization and architectural choices
- Accuracy and useful information (separation, signal channels, ...)
- Expansion in functionality (neutrals, PID, ...) 

Extensive performance studies, crucial for calibration
- In simulation
- In real data

Integration in LHCb trigger
- Export into ROOTs TMVA SOFIE, finishing GNN implementation
- Study usage of hardware accelerators for Upgrade II (FPGA, GPU, ...)
Backup slides
Signal-based trigger vs Full Event Interpretation (FEI)

**Signal based**

The current LHCb trigger is an **OR between many decay-mode selection lines**.

Since Run2, to reduce the event size, some lines **store only parts of the event which are related** to the specific signal.  

[**JINST 14 (2019) 04, P04006**](#)

E.g.: store the signal + the tracks in the same primary vertex (PV).

**FEI**

New proposal: try to **reconstruct the b- and c- hadron decay chains in the event**, in a hierarchical-clustering manner (cluster → unstable particle), **and discard the rest**.

**Advantages:**

- **Exploit extra correlations** between objects in the event.
- **Bandwidth oriented**: focus on storing as much “useful” information as possible.
- Case of several signals per event as an integral part of the approach.
- Establishment of a basis for an expanded functionality of the trigger: inclusive selections, study of anomalous events ...
Cut-based edge pruning

Define two adequate topological variables for each edge (pair of particles)

- Angle
- Transverse distance

Combined 3-momentum

\[ p_1 \]
\[ p_2 \]

Transverse distance

Signal edges (from same b-ancestor)

Background edges (all other edges)

Applied veto

This veto reduces on average 60% of the total number of edges in the graph. It also reduces connections between signal tracks, but it only leaves ~2% of the signal tracks fully disconnected.
Training dataset: emulating Run3 conditions

Particle collision&decay

The training and performance studies are currently done using PYTHIA, with the following configuration:

- Proton-proton collisions at 13 TeV.
- Average number of collisions per event: 7.6.
- Selecting events with at least one b-hadron produced (inclusive decay).

“Detection and reconstruction”

We require all the tracks and the b-hadrons to be inside the LHCb geometrical acceptance.

In addition, we emulate the reconstruction of the following quantities, using publicly available expectations for the LHCb performance in Run3 (see backup):

- Origin point of the tracks (first measurement in the Vertex Locator).
- Three-momentum of the tracks.
- Position of the primary vertices.
Example of decay-tree simplification used in the prototype

Original chain of ancestors:

$$\pi^+ \leftarrow \rho(770)^0 \leftarrow \varphi(1020) \leftarrow D^+ \leftarrow B^0 \leftarrow B^{*0}$$

Simplified chain of ancestors (based on reconstructible vertices):

$$\pi^+ \leftarrow D^+ \leftarrow B^0$$
Performance: timing

**Simplistic study** (no parallelisation, no hardware accelerators*, algorithm to be further optimised), to **understand which are the slowest parts of the algorithm and how they scale with the total number of particles per event**.

The slowest part is the **node pruning**, which also has the strongest dependency on the number of particles. → Many possible ways of optimisation.

The processing time of the subsequent algorithms is quite stable regarding changes in event complexity.

(*) Study done on a darwin-x86_64 architecture with a 2.8 GHz Intel Core i7 processor.