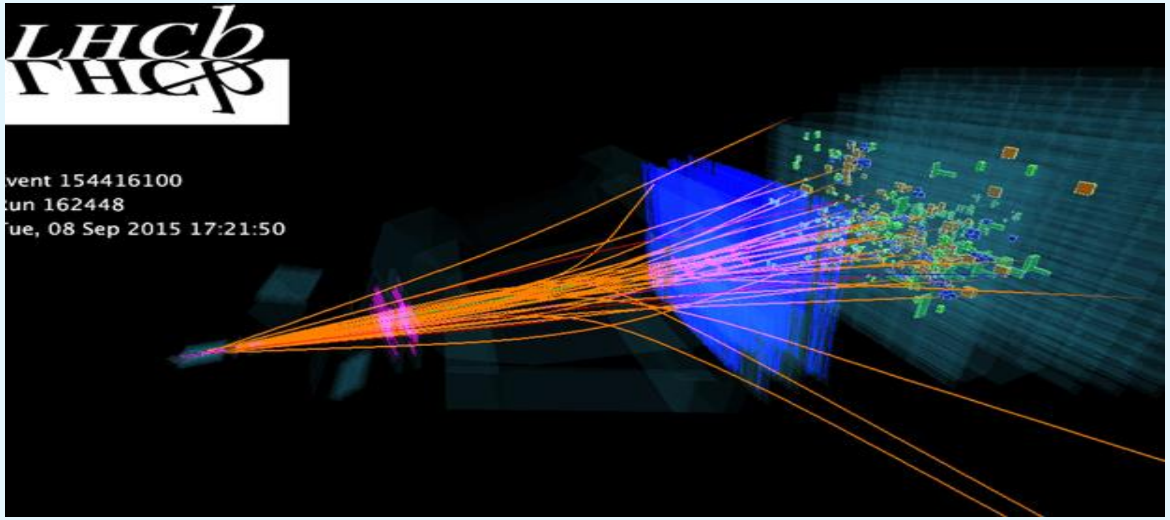
New developments and applications of a Deep-learning-based Full Event Interpretation (DFEI) in proton-proton collisions



Marta Calvi[1], **Simone Capelli**[1], Jonas Eschle[2], Julian Garcia Pardinas[3], Abhijit Mathad[3], Andrea Mauri[4], Nicola Serra[5], Rafael Silva Coutinho[2], Felipe Luan Souza De Almeida[2], William Sutcliffe[5], Azusa Uzuki[5]

University of Milano-Bicocca & INFN[1], Syracuse University[2], CERN[3], Imperial College[4], University of Zurich[5]

5/02/2024

1st European AI for Fundamental Physics Conference

Outline (I)

Introduction

Motivation
Full Event
Interpretation

DFEI
Prototype

Towards
production

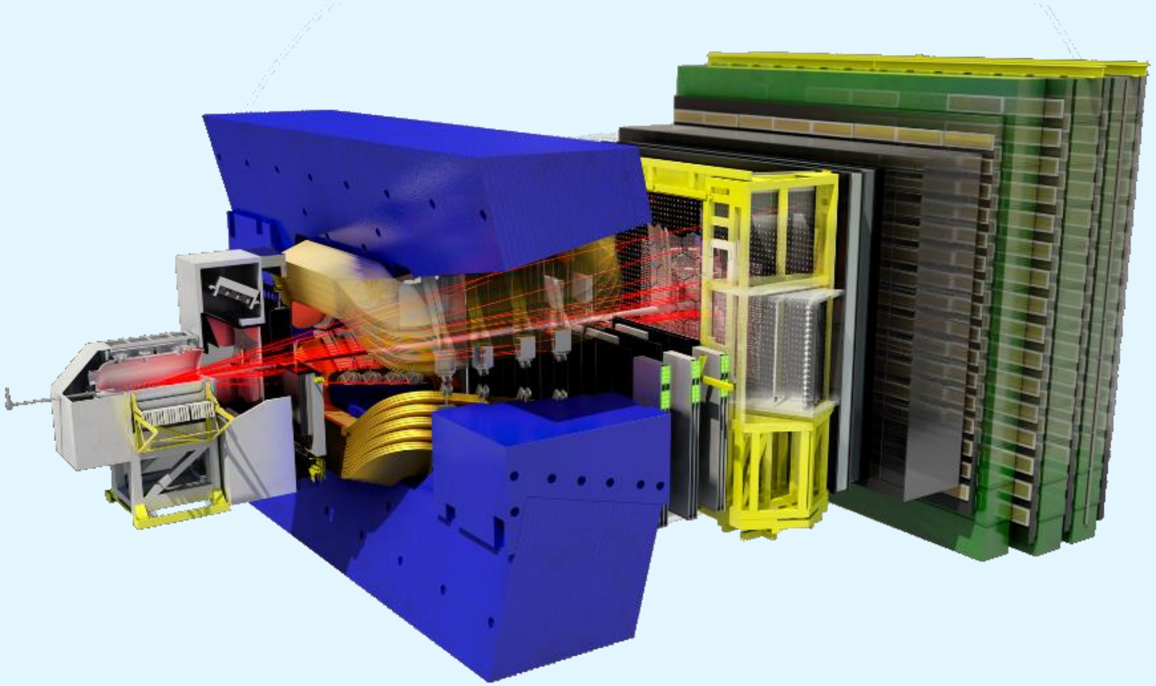
Summary

Motivation

- R&D project with two main applications: **Trigger** and **offline analysis**

- It could be used at LHCb, a single-arm forward spectrometer, designed to studying the **decays of beauty and charm hadrons**, rare decays and CPV measurements

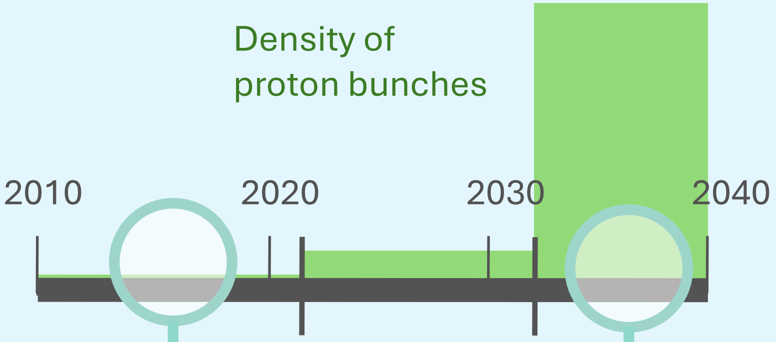
- Very **broad physics program**, to be **maintained** and **expanded** in future LHC runs
- **Increased particle multiplicities** for LHCb Upgrades I and II bring big challenges



Paradigm shift

Which events are interesting?

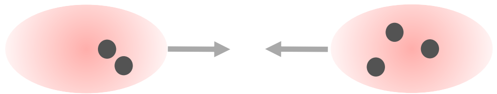
“Signal-centric” trigger strategy



Which *parts* of the event are interesting?

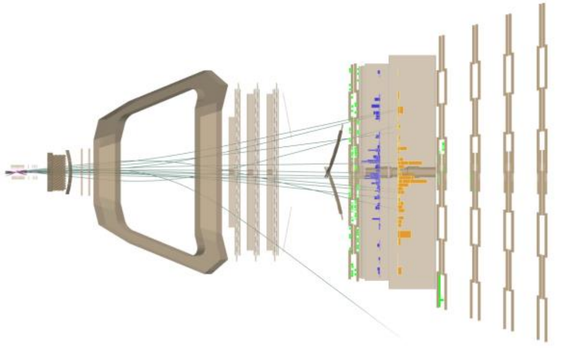
⇩

New trigger paradigm & higher background level

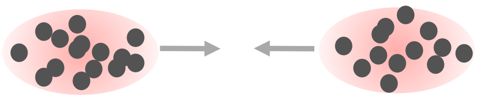


$\ll$ **1 signal per event, ~50 tracks**

Low noise

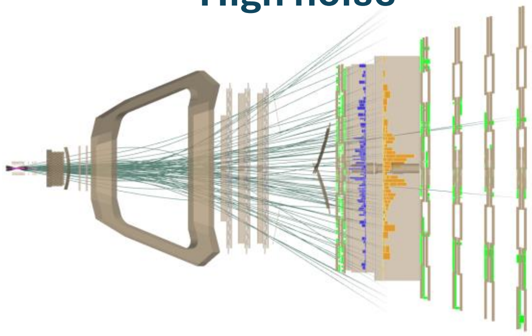


Need for ML



Multiple signals per event, ~1000 tracks

High noise



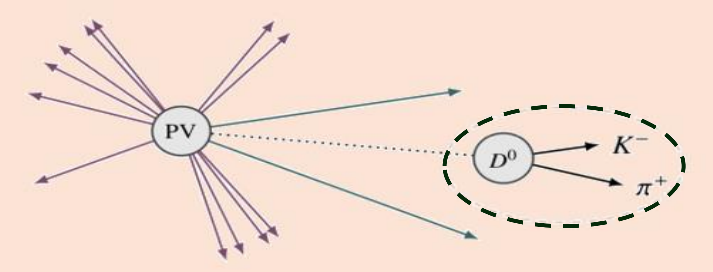
Signal-based vs Full Event Interpretation

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 that are related** to the specific signal [JINST 14 (2019) 04, P04006]

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



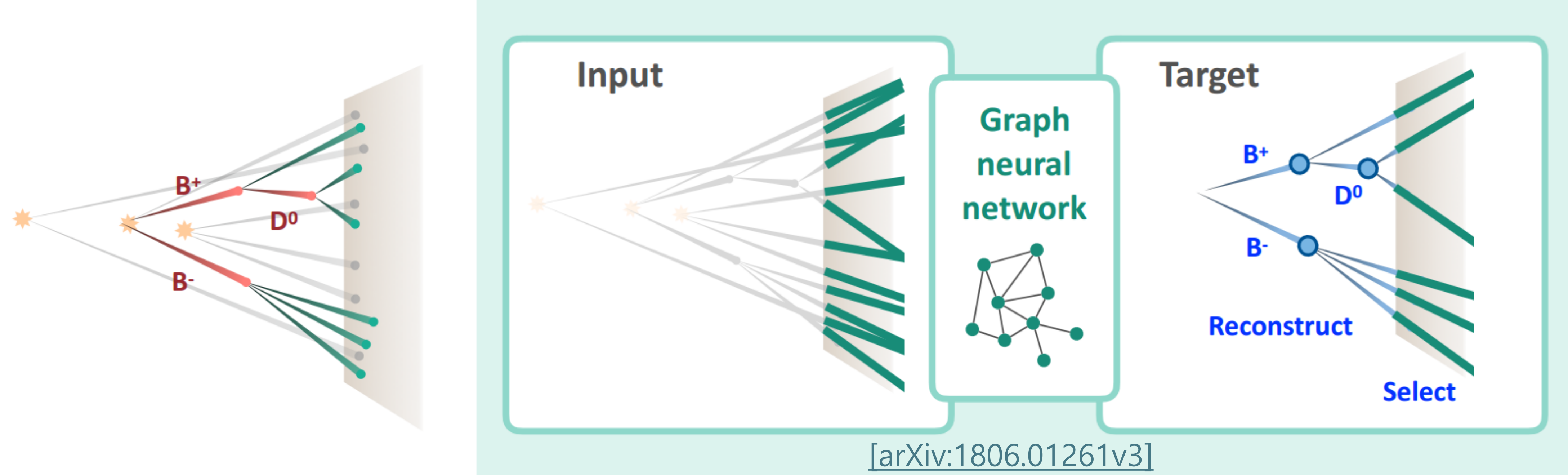
FEI

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, ...

Deep-learning based Full Event Interpretation (DFEI)



- **Input features**: *charged particles* and their measured *properties* (**nodes**) and their *relations* (**edges**)
- **Hierarchical, automatized and inclusive** reconstruction of heavy-hadron decay chains

1) **Trigger**: Safely discard rest of event, with minimal loss for analyses → **powerful event size reduction tool in a multi-signal environment**

2) **Analysis** tool for **background classification & suppression and for inclusive studies**

Algorithm Modules

[Comput.Softw.Big Sci. 7 (2023) 1, 12]

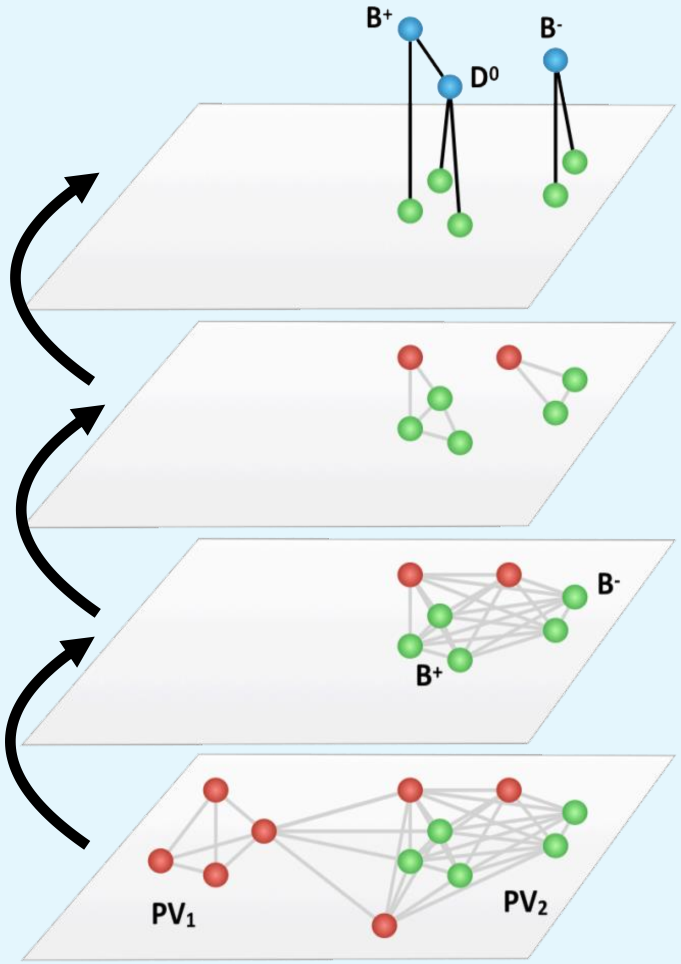
Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event

The Cooking Recipe

3. “Lowest common ancestor” inference

2. Edge pruning

1. Node pruning



Algorithm Modules

Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event

Goal: Remove most of the nodes not produced in a b-hadron decay

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

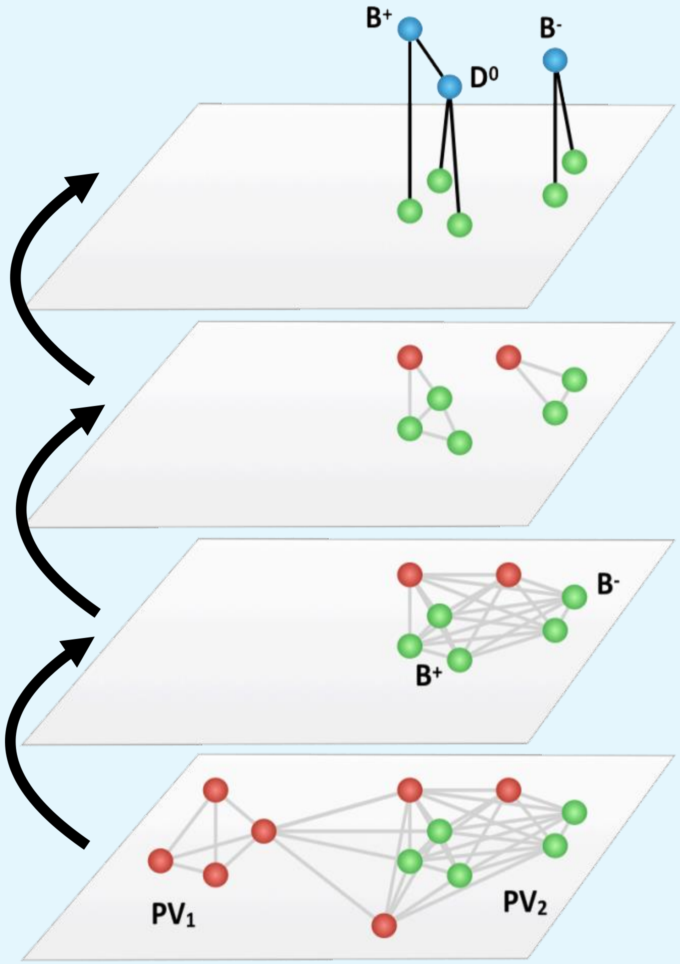
Background nodes: particles from the rest of the event

Cut @ 99% ~70% BKG rejection

3. "Lowest common ancestor" inference

2. Edge pruning

1. Node pruning



Algorithm Modules

Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event

Goal: Remove connections between nodes not produced in the same b-hadron decay

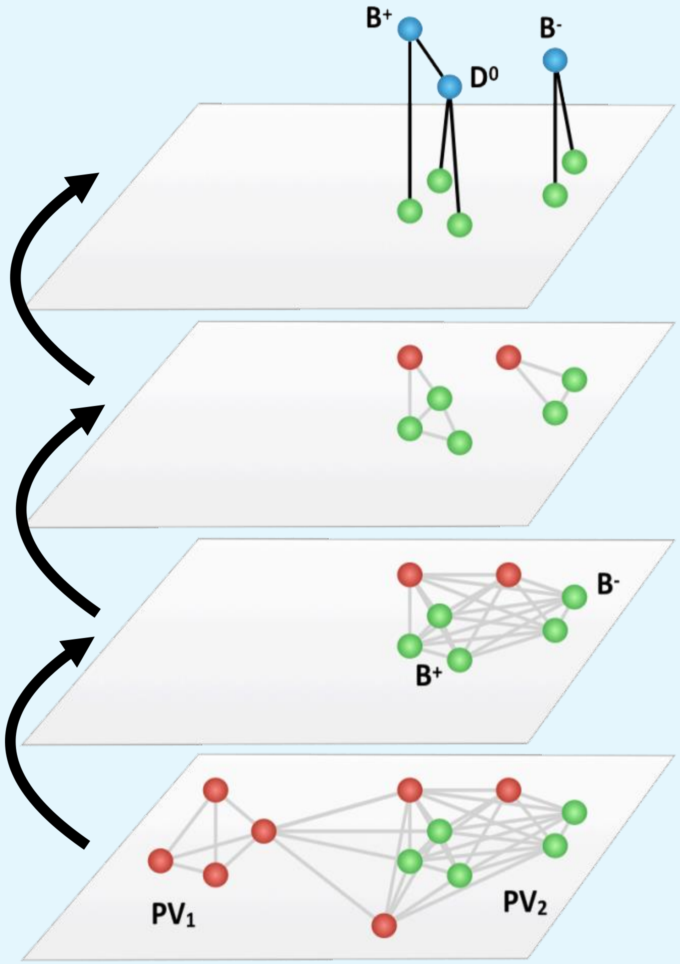
Signal nodes: pairs of particles with the same b-hadron ancestor
Background nodes: any other pair of particles

Cut @ 99% ~68% BKG rejection

3. "Lowest common ancestor" inference

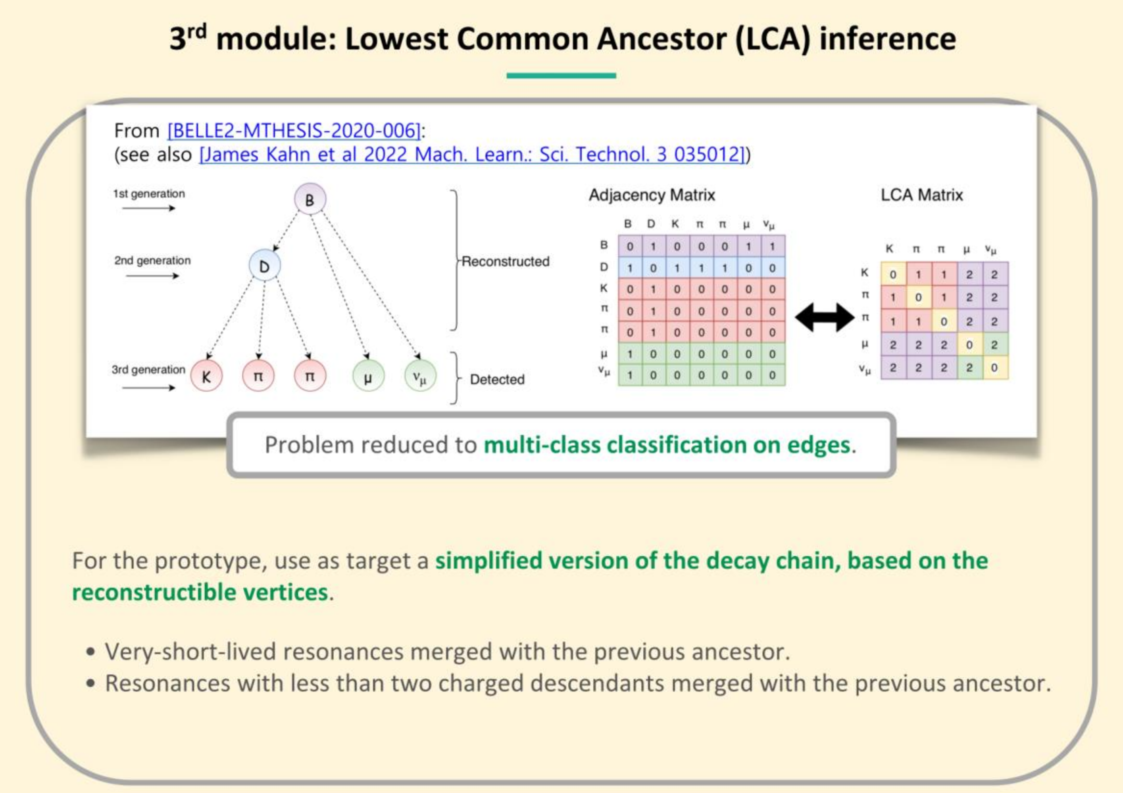
2. Edge pruning

1. Node pruning



Algorithm Modules

Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event



3[rd] module: Lowest Common Ancestor (LCA) inference

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

Adjacency Matrix

	B	D	K	π	π	μ	ν_μ
B	0	1	0	0	0	1	1
D	1	0	1	1	1	0	0
K	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
μ	1	0	0	0	0	0	0
ν_μ	1	0	0	0	0	0	0

LCA Matrix

	K	π	π	μ	ν_μ
K	0	1	1	2	2
π	1	0	1	2	2
π	1	1	0	2	2
μ	2	2	2	0	2
ν_μ	2	2	2	2	0

Problem reduced to **multi-class classification on edges**.

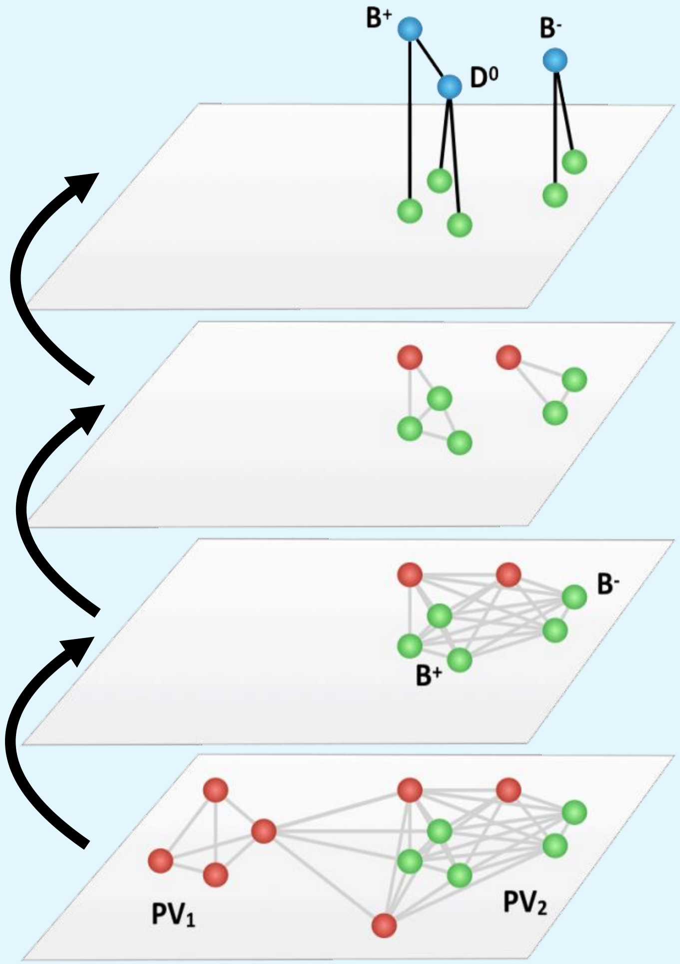
For the prototype, use as target a **simplified version of the decay chain, based on the reconstructible vertices**.

- Very-short-lived resonances merged with the previous ancestor.
- Resonances with less than two charged descendants merged with the previous ancestor.

3. "Lowest common ancestor" inference

2. Edge pruning

1. Node pruning



Outline (II)

DFEI Prototype

- Event-size reduction in trigger
- BKG reduction in Analysis
- Search for new states

Introduction

Towards production

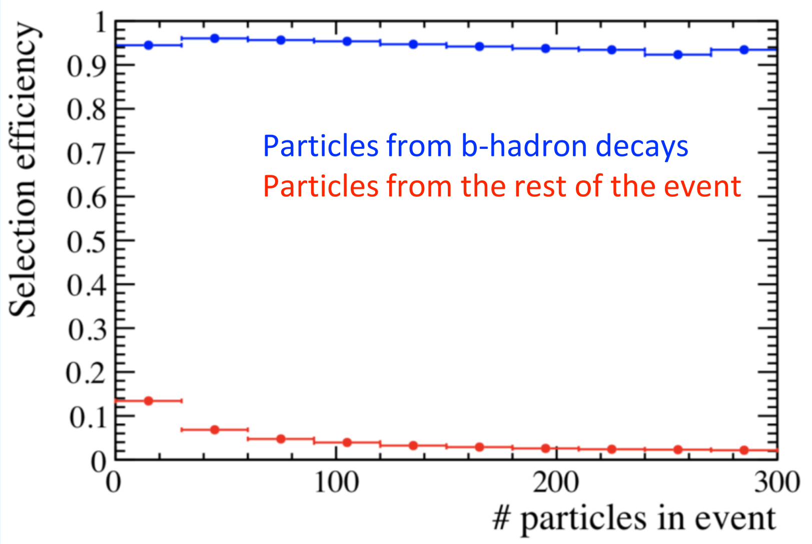
Summary

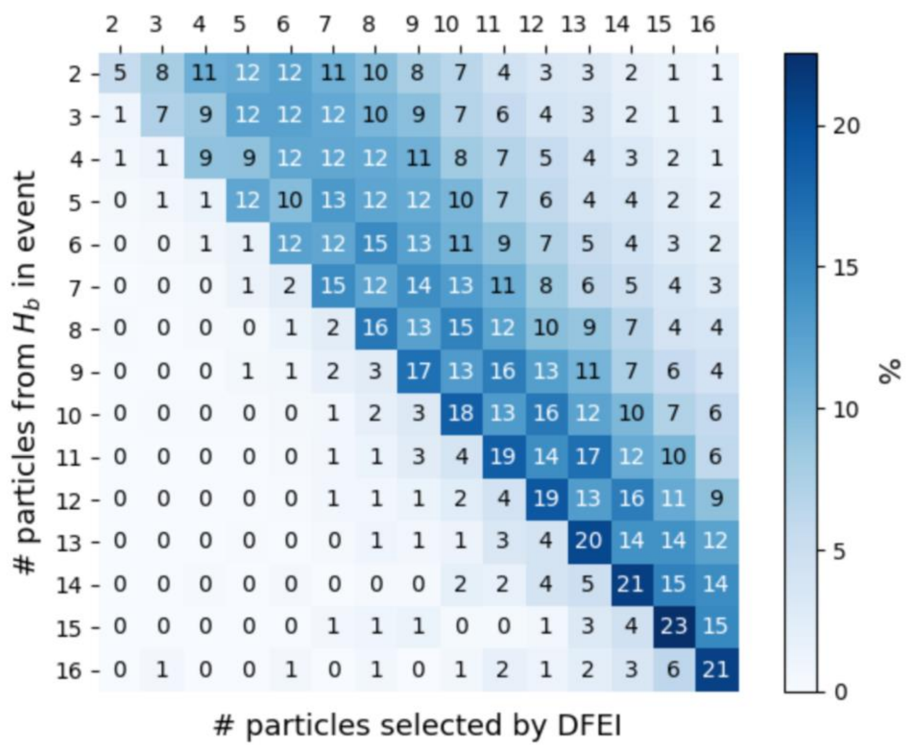
Trigger performance: event-size reduction

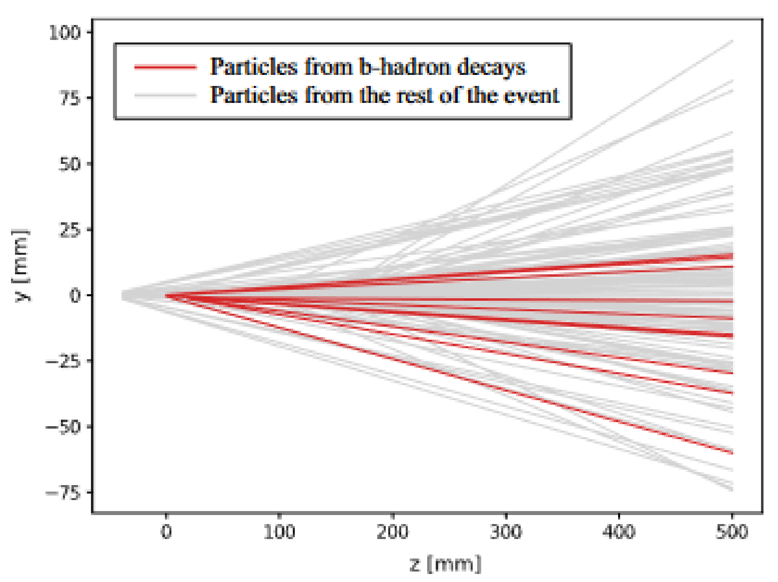
Algorithm based on **Graph Neural Networks**, for the moment restricted to b-hadrons and charged stable particles

- Trained on **custom simplified simulation** in Run3-like conditions, with ~ 140 particles per event [https://zenodo.org/records/7799170]

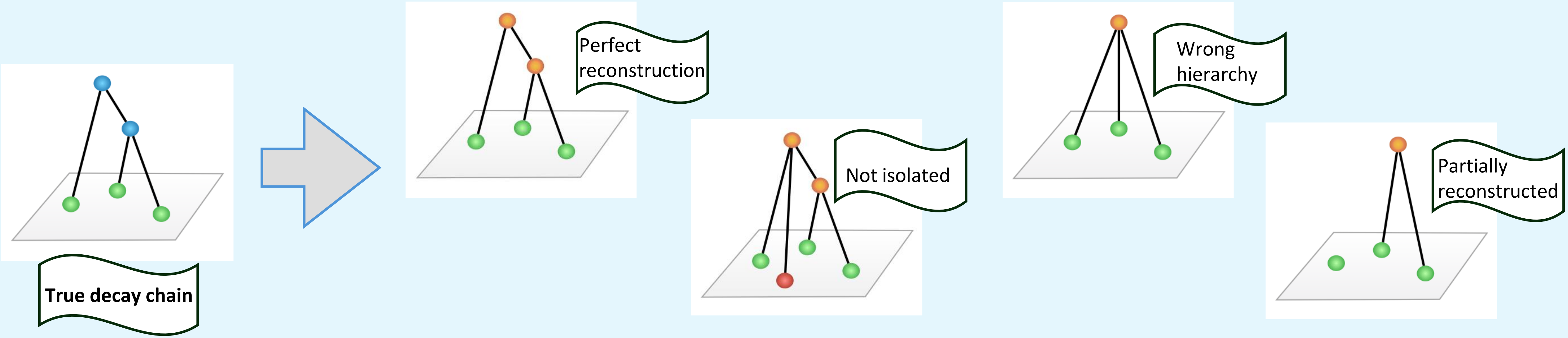
Background rejection 96%	Signal efficiency: 94%	Average selected particles: ~10







Data filtering: decay-chain reconstruction



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

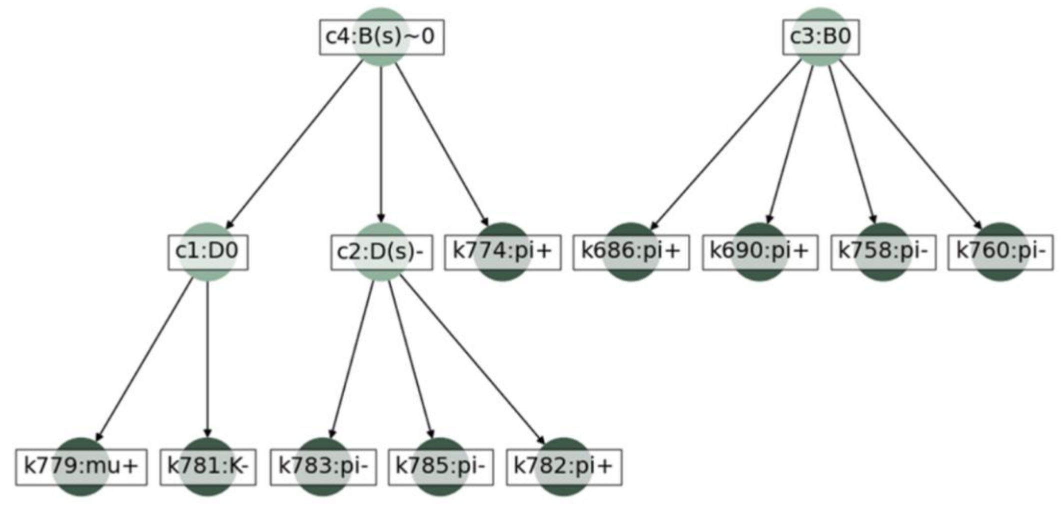
Standard vs DFEI-based analysis

Standard exclusive analysis → Search for **specific candidate** → Filter through data

DFEI analysis → Look what is **reconstructed** in the events → Characterise decay chains, **standard** or **exotic**

Possible **contamination** from poor or **partial reconstructed** decays

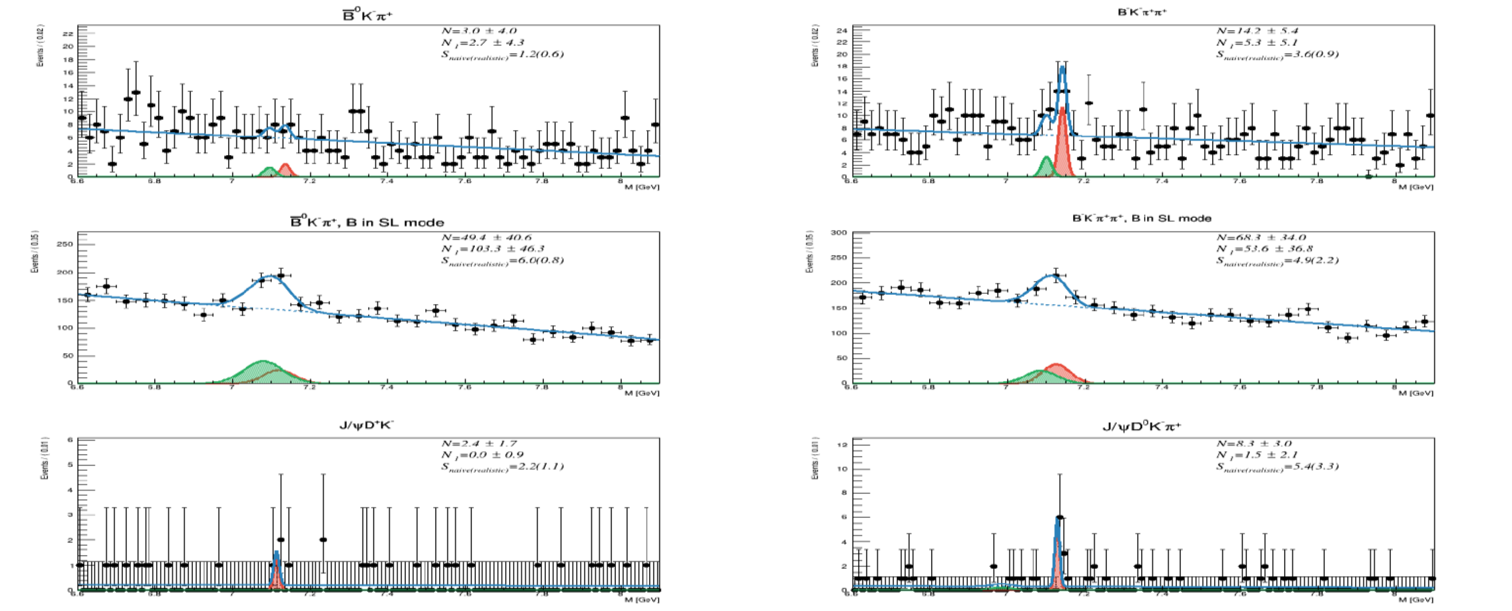
Reconstructed trees in event



Search for exotic states

- LHCb has found ~50 new exotic states:
  could benefit from **more inclusive searches**, and DFEI could do that
- Current ideas for searching for exotic states include **simultaneous analysis** of multiple exclusive decay chains

E.g.: the search for the tetraquark T_{bc} could involve the simultaneous mass fit of 20 to 40 channels!



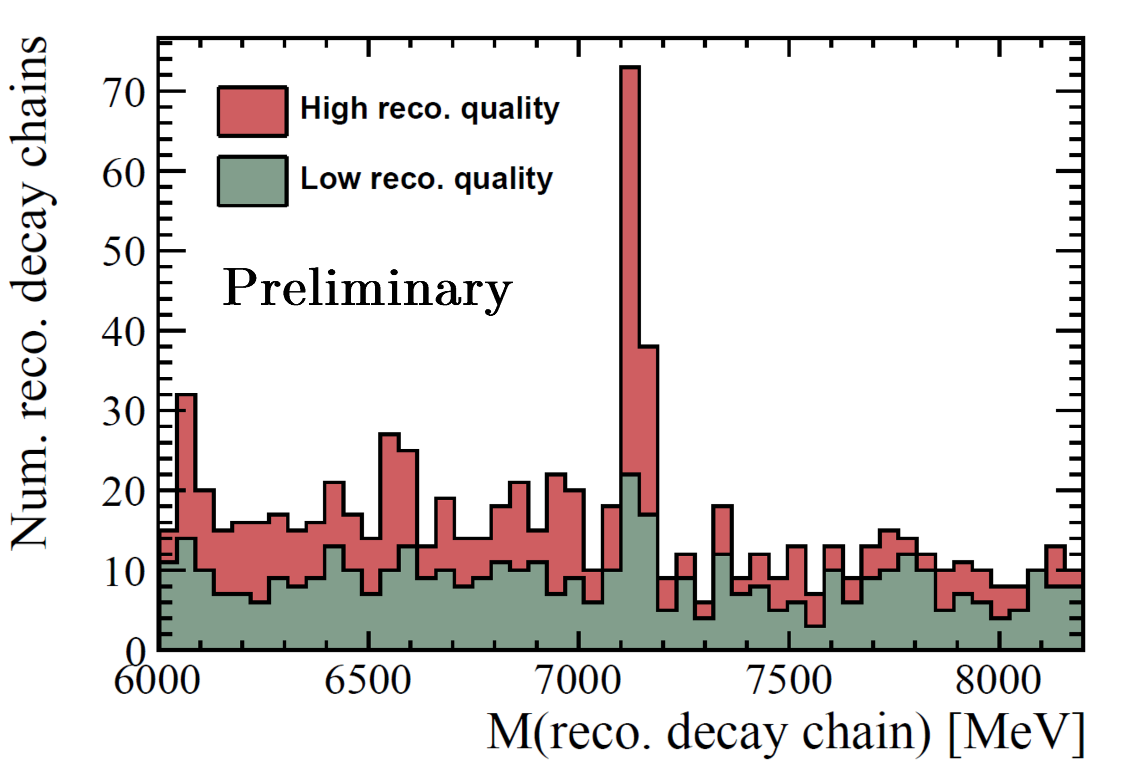
Search for Tbc – prospects for Run3, I. Polyakov,
Hunting for the charming beauty tetraquark Tbc: LHCb meets theory, 5 October 2023, CERN

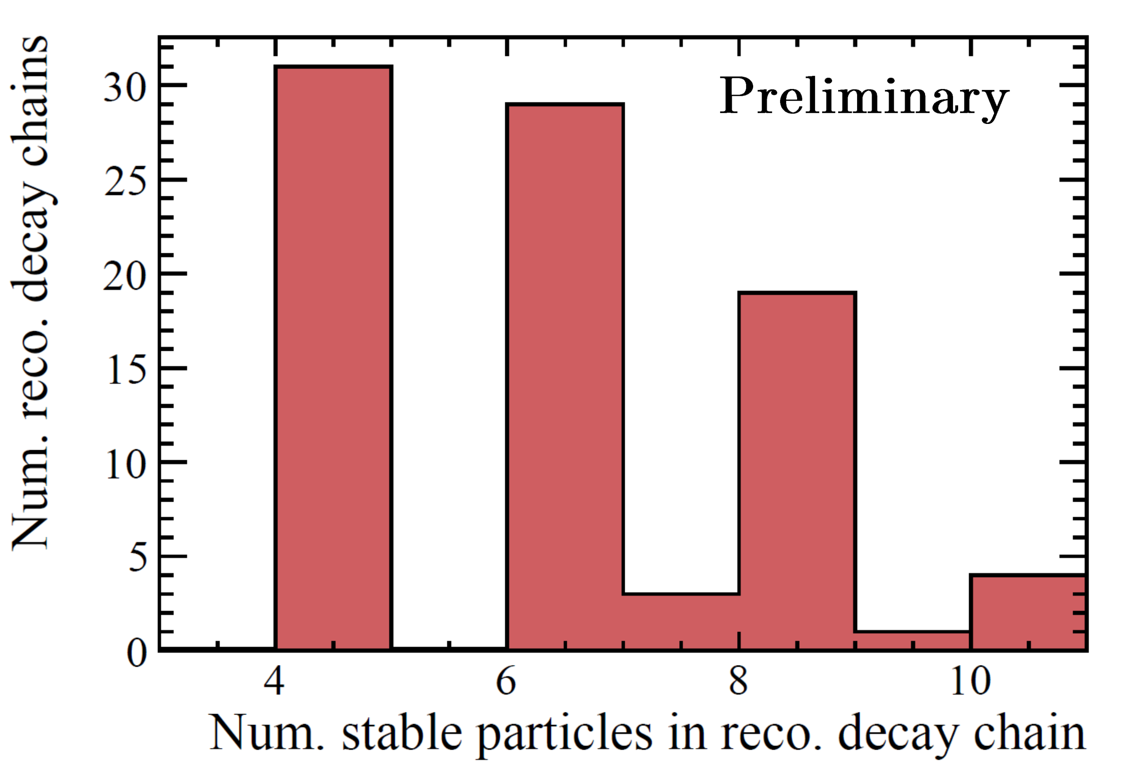
decay channel	$\varepsilon_{tot} \times \mathcal{B}$ $[10^{-9}]$	Expected yield
fully reconstructed channels		
D^0D^0	7.8	3.1
$D^0D^+\pi^-$	9.2	3.7
$D^0D^0\pi^+\pi^-$	3.4	1.4
$D^+D^+\pi^-\pi^-$	3.5	1.4
sum		10
$J/\psi D^+K^-$	2.3	0.9
$J/\psi D^0K^-\pi^+$	3.1	1.2
sum		2.1
$\overline{B}^0K^-\pi^+$	32.9	13.2
$B^-K^-\pi^+\pi^+$	33.6	13.4
$\overline{B}^0K^-\pi^+\pi^+\pi^-$	6.7	2.7
sum		29
$\overline{B}^0{}_{SL}K^-\pi^+$	188	75
$B^-{}_{SL}K^-\pi^+\pi^+$	94	38
$\overline{B}^0{}_{SL}K^-\pi^+\pi^+\pi^-$	61	24
sum		137
$D^0D^+\mu^-\nu$	56	22
$D^0D^0\pi^+\mu^-\nu$	43	17
$D^0D^+\mu^- + X$	163	65
$D^0D^0\mu^- + X$	108	43
sum		147
$\overline{B}^0K^-\mu^+\nu$	24	9.5
$B^-K^-\pi^+\mu^+\nu$	16	6.5
sum		16
$D^0K^-\pi^+$	68	27
D^+K^-	134	54
$D^0\pi^+\pi^-$	2.5	1
$D^+\pi^-$	4.9	2
sum		84

Search for exotic states

- DFEI can **simultaneously reconstruct** the **different decay chains**, allowing for a **more inclusive search** for exotic states

- **Simplified Pythia-based simulation** sample of several weakly decaying modes of T_{bc} analysed simultaneously using DFEI

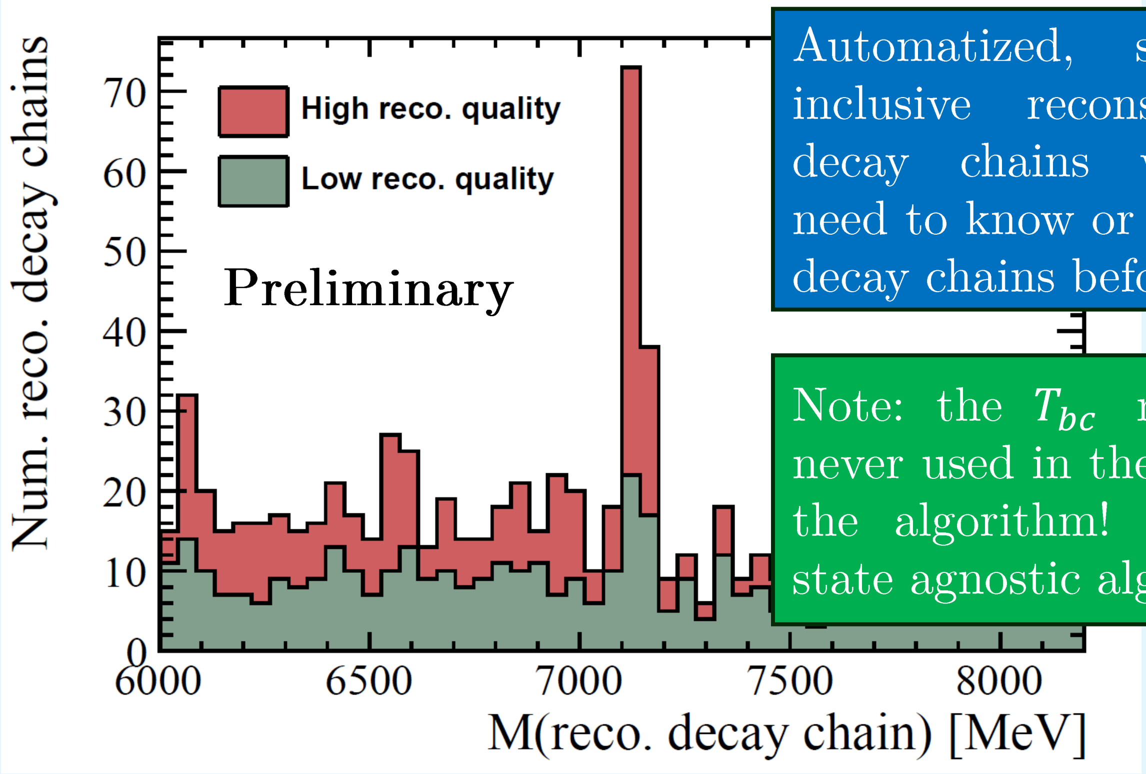




Search for exotic states

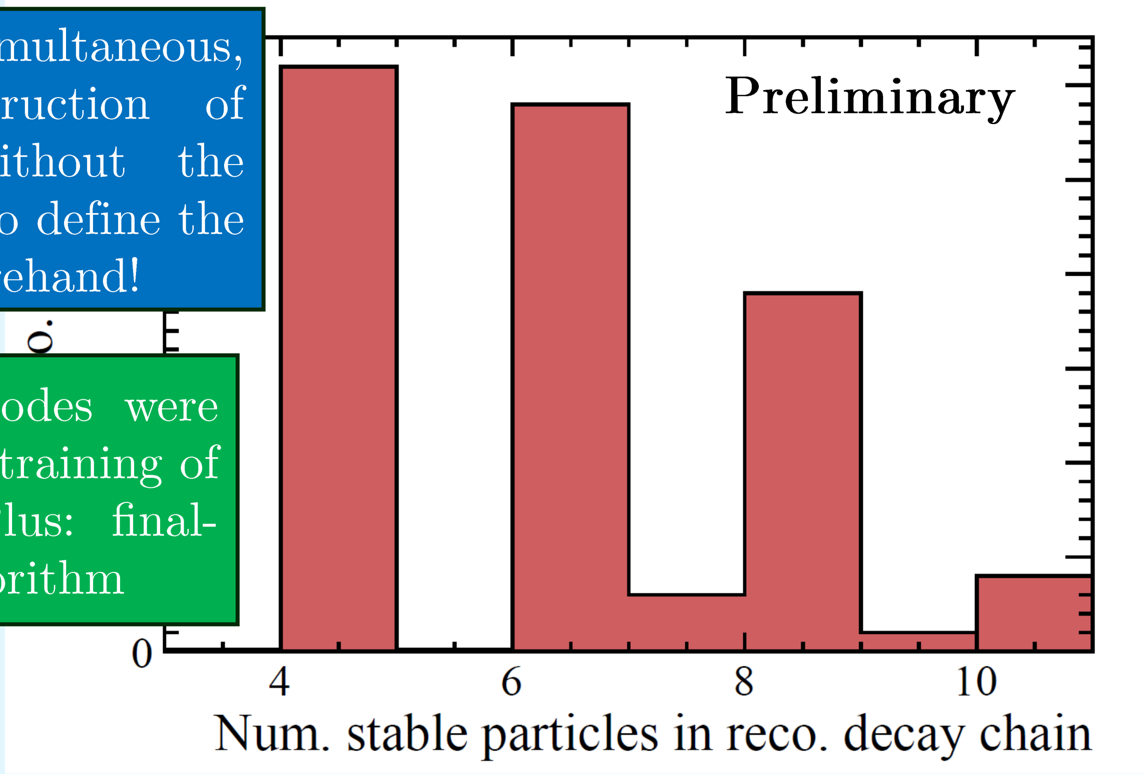
- DFEI can **simultaneously reconstruct** the **different decay chains**, allowing for a **more inclusive search** for exotic states

- **Simplified Pythia-based simulation** sample of several weakly decaying modes of T_{bc} analysed simultaneously using DFEI



Automatized, simultaneous, inclusive reconstruction of decay chains without the need to know or to define the decay chains beforehand!

Note: the T_{bc} modes were never used in the training of the algorithm! Plus: final-state agnostic algorithm



Outline (III)

- Introduction
- DFEI Prototype
- Towards production
  - Speed-up
  - C++ Inference pipeline
- Summary

First speed-up round and C++ inference pipeline

First DFEI prototype: Evaluation pipeline on python with TensorFlow

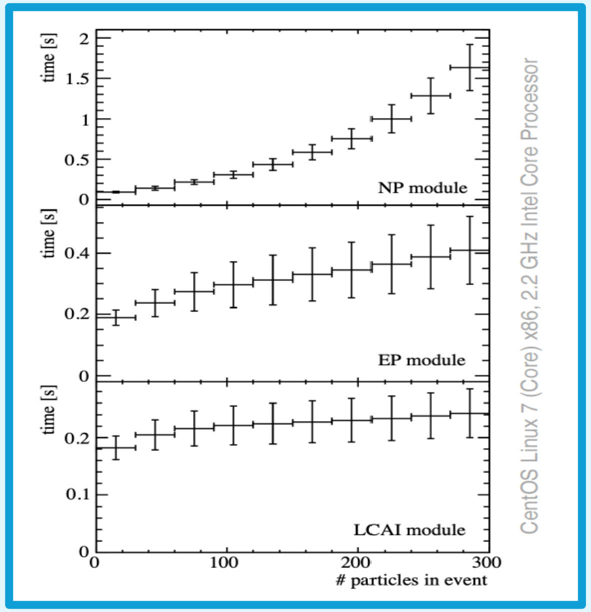
Quadratic scaling of the inference time with the track multiplicity	Overall evaluation time on the order of few seconds per event on CPU (dominated by NP)

Substitute the GNN models of the NP and EP by BDT models to improve scaling

Evaluated independently for each node/edge	Cut at 99% signal efficiency for nodes/edges

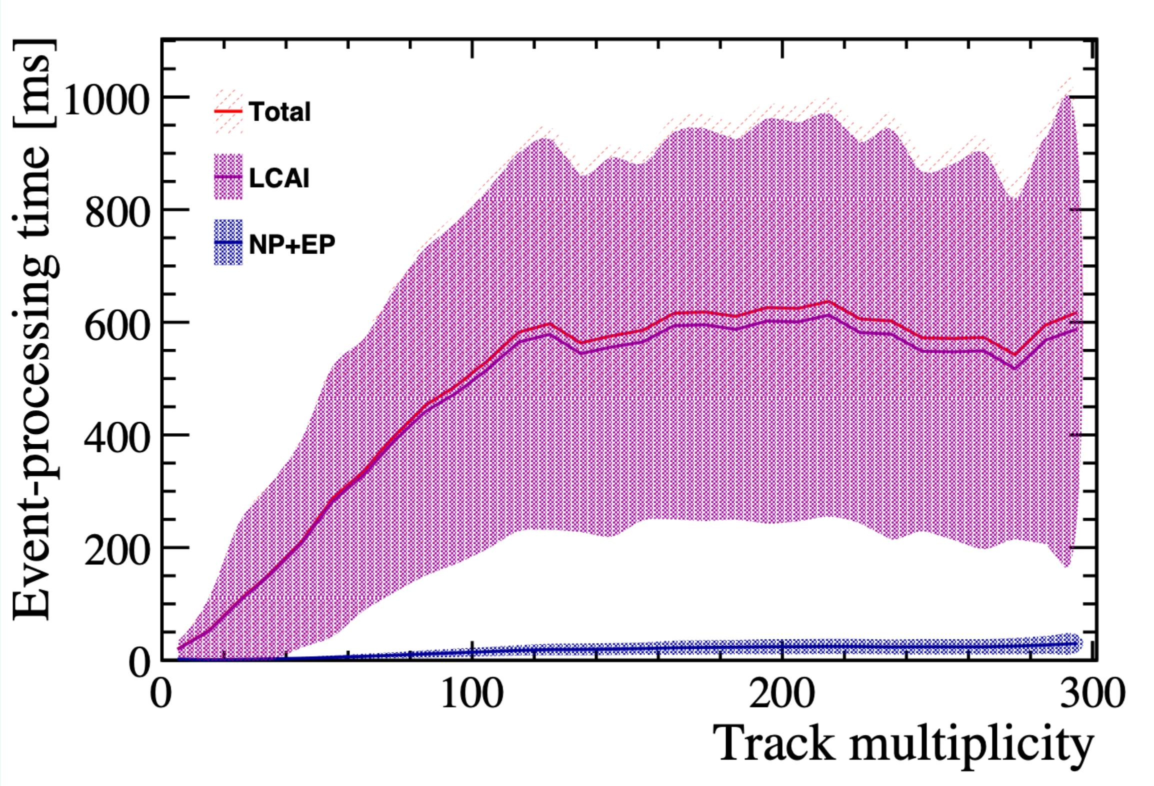
C++ inference pipeline that takes as input

BDT-based NP, BDT-based EP, GNN-based LCAI	Using the C API of CatBoost and TMVA::SOFIE for the LCAI GNN inference

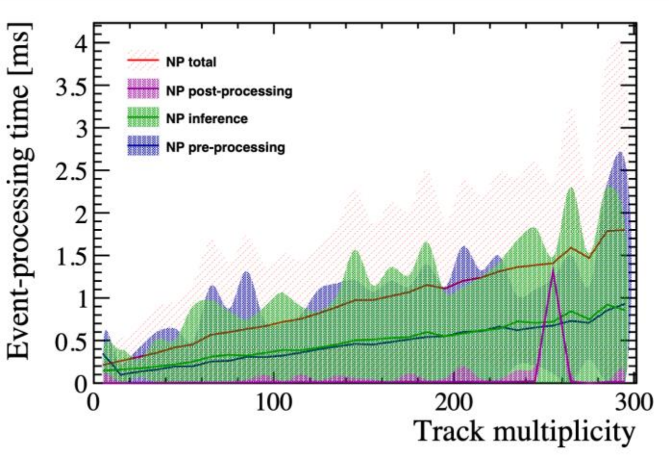


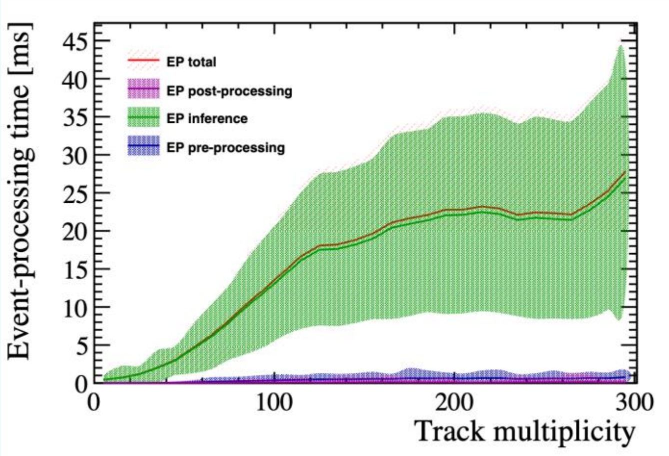
First speed-up round and C++ inference pipeline

CentOS Linux 7 (Core) x86 2.8 GHz Intel Core Processor



- ✓ **Excellent scaling** achieved thanks to the **fast pre-filtering** by the **NP** and **EP**
- ➢ **Time now dominated** by the **LCAI** algorithm
- ➢ Significant overall **speed up**





- ➢ Next: physics performance studies for the new configuration, **hyper-parameter tuning**
- ➢ In parallel: study of **GNN architecture** developments to gain **further speed-ups**

Outline (IV)

- Introduction
- DFEI prototype
- Towards production
- Summary

Summary

1) Developed a **prototype**, published in paper [Comput.Softw.Big Sci. 7 (2023) 1, 12]

2) We are exploring **applications** in **three** different domains:

   - **Trigger**, **data filter** and **inclusive analysis** with very promising results [https://zenodo.org/records/7799170]

3) On the way from prototype **towards production**

   - Developed a C++ pipeline and **improved the scaling and timing** of the algorithm

- **Stay tuned**: further developments are on the way!

Thank you!

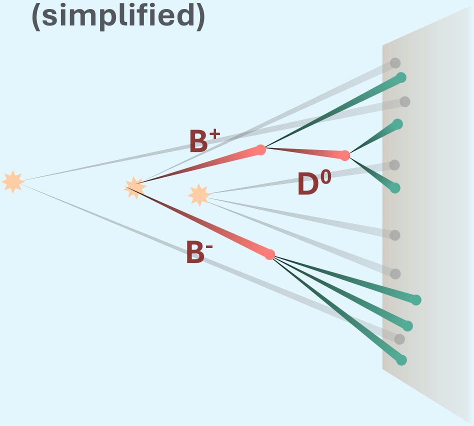
Backup

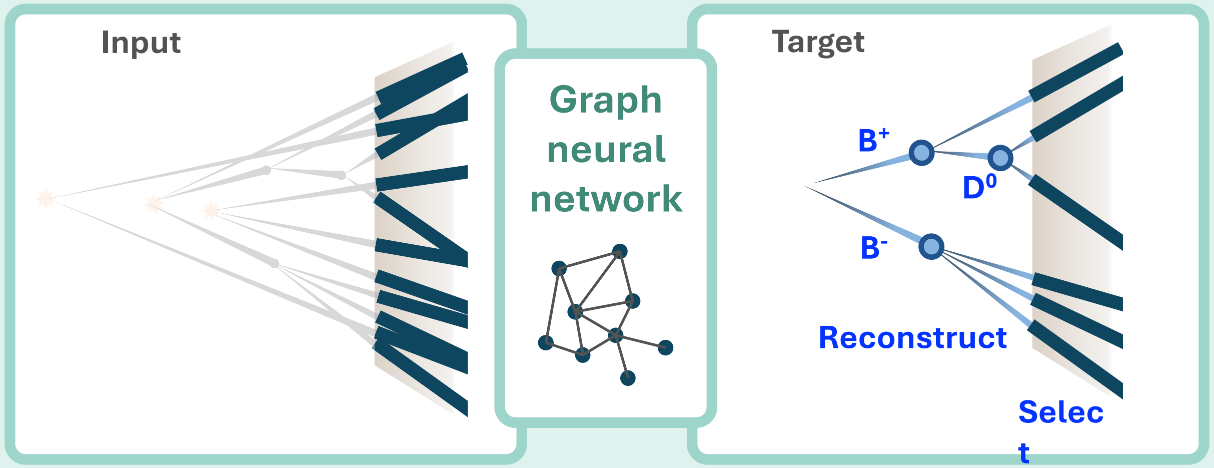
Facing the new era with machine learning

Novel approach proposed

DFEI:
Deep-learning based
Full Event Interpretation

LHCb event (simplified)





"Maximally efficient" trigger.

Similar developments in other experiments

Full Event Interpretation algorithm at an e+e- collider
[Comput.Softw.Big Sci. 3 (2019) 1 6], BELLE2-MTHESIS-2020-006].

GNNs for trigger purposes
[see e.g. Eur.Phys.J.C 81 (2021) 5, 381, Frontiers in Big Data 3 (2021) 44].

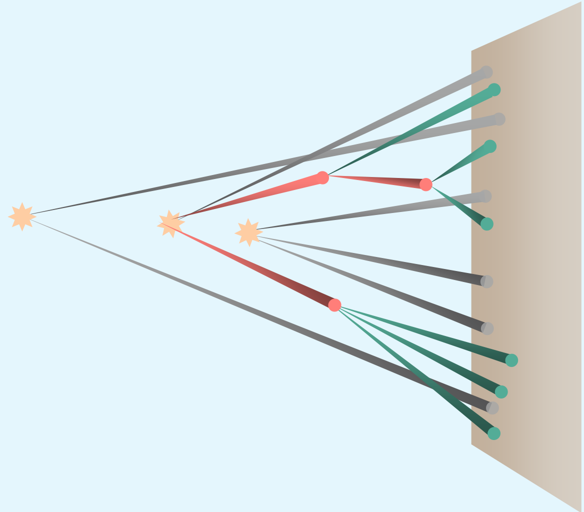
Decays and graph structures

Event

Global: event information
nTracks, ...

Nodes: track variables
momentum, (PID), ...

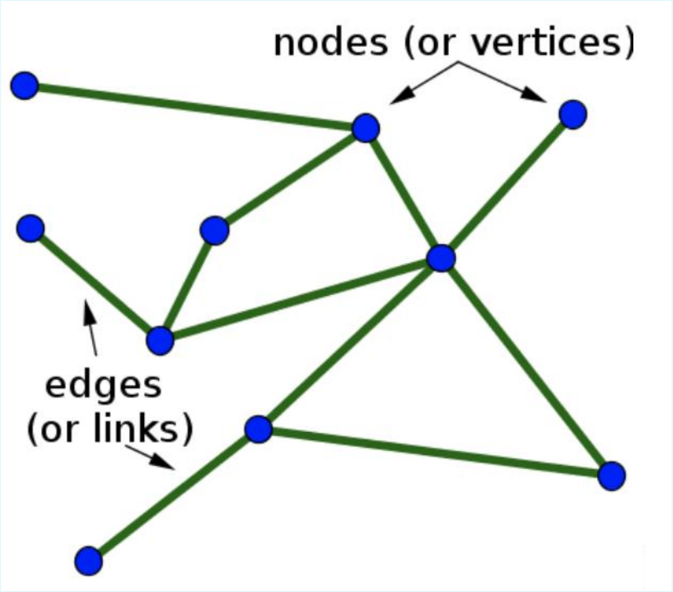
Edges (# nodes^2!): track *relations*
angle, DOCA, ...



Graph structures

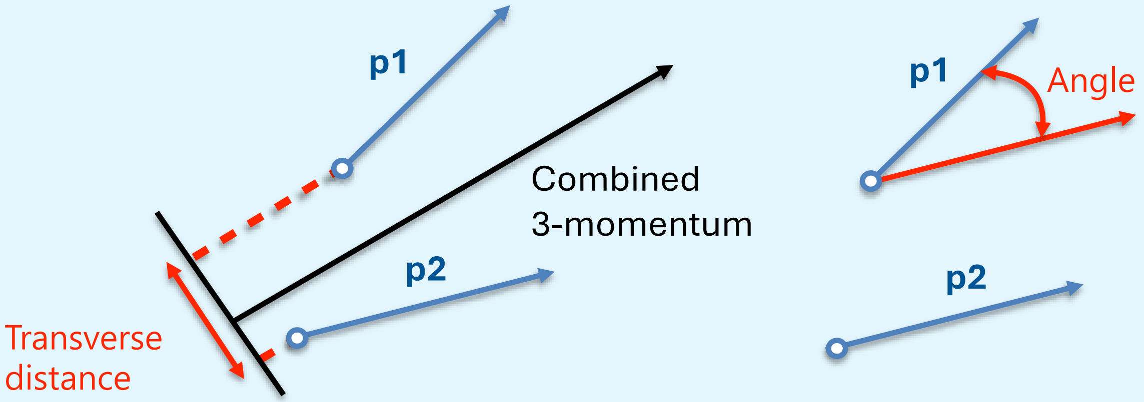
Representation of objects with relations

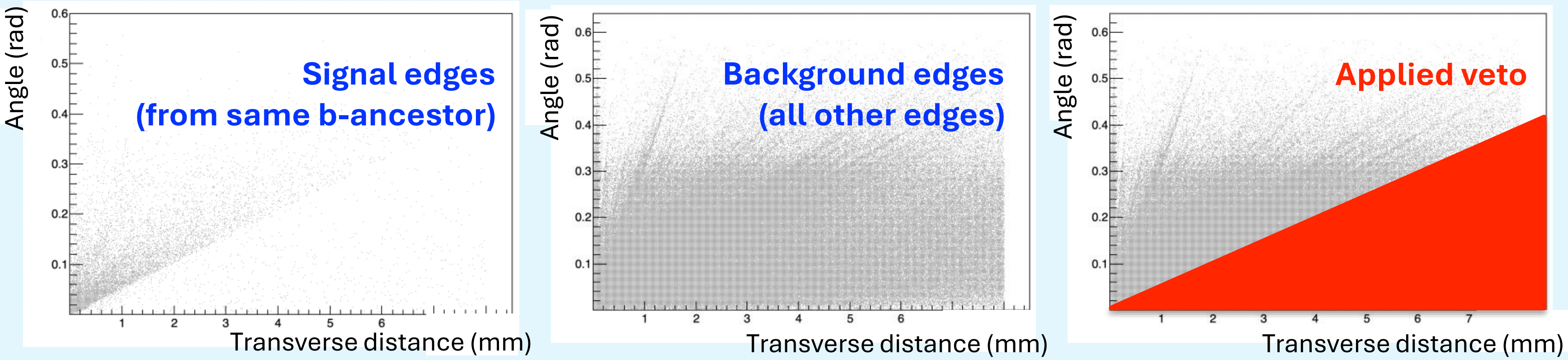
Arbitrary, sparse/dense relations



Cut-based edge pruning

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





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.

C++ inference pipeline

Information of the charged stable particles in an event.

NP algorithm:

1. **Pre-processing:** read information on particle-level quantities in the event, compute derived quantities.
2. **Inference:** evaluate the NP CatBoost BDT model, using the CatBoost C API.
3. **Post-processing:** apply the node filter.

EP algorithm:

4. **Pre-processing:** compute particle-pair-level quantities.
5. **Inference:** evaluate the EP CatBoost BDT model, using the CatBoost C API.
6. **Post-processing:** apply the edge filter.

LCAI algorithm:

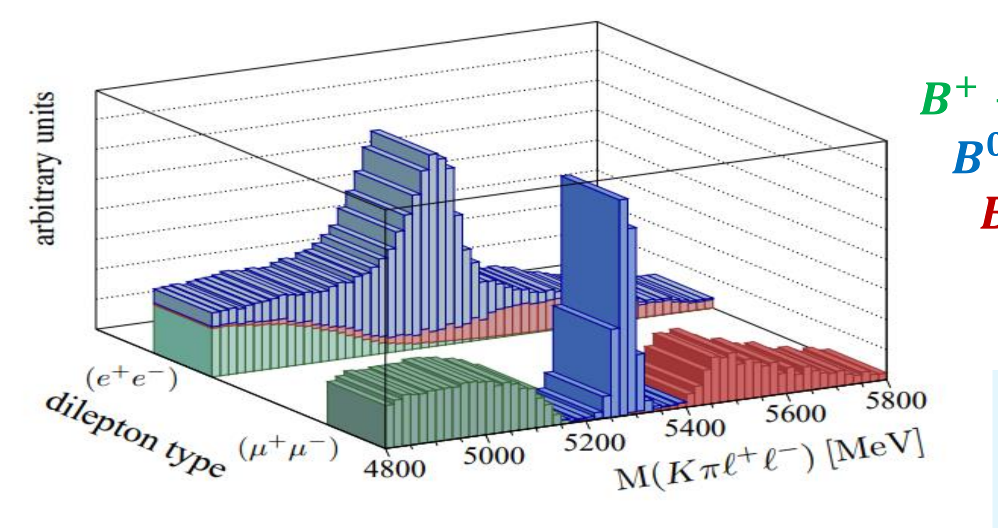
7. **Pre-processing:** construct the input graph, combining particle-level, particle-pair-level and global information in the event.
8. **Inference:** evaluate the LCAI GNN model, using TMVA::SOFIE.

Prediction of the hierarchical relations for all pairs of particles.

Example case: $B^0 \to K^+\pi^-\ell^+\ell^-$

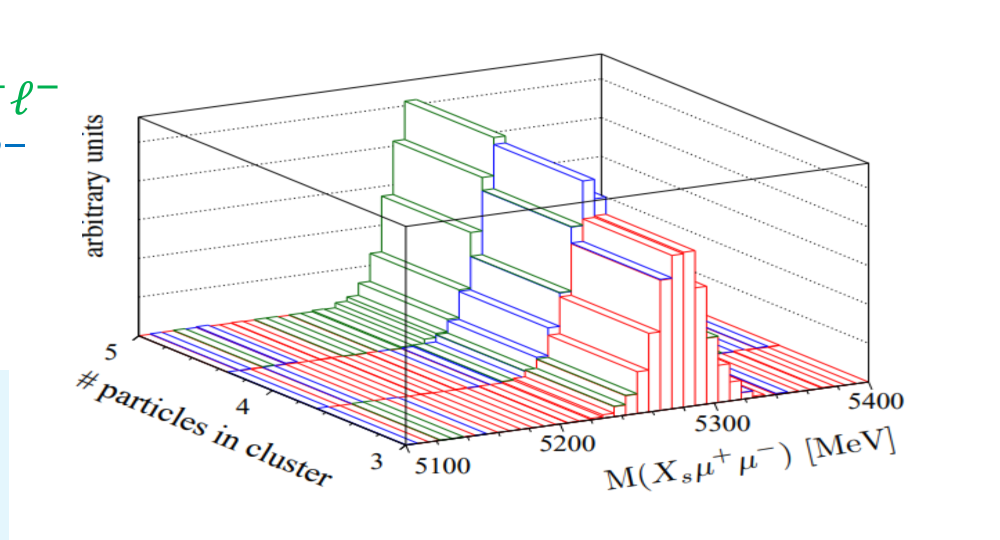
- Partially- and over-reconstructed backgrounds are challenging to disentangle from the signal in exclusive, conventional analysis → loose of sensitivity
  - Particularly important for electron modes, impacting R_X measurements

- With DFEI, these contributions are fully reconstructed and classified in clusters of different numbers of particles



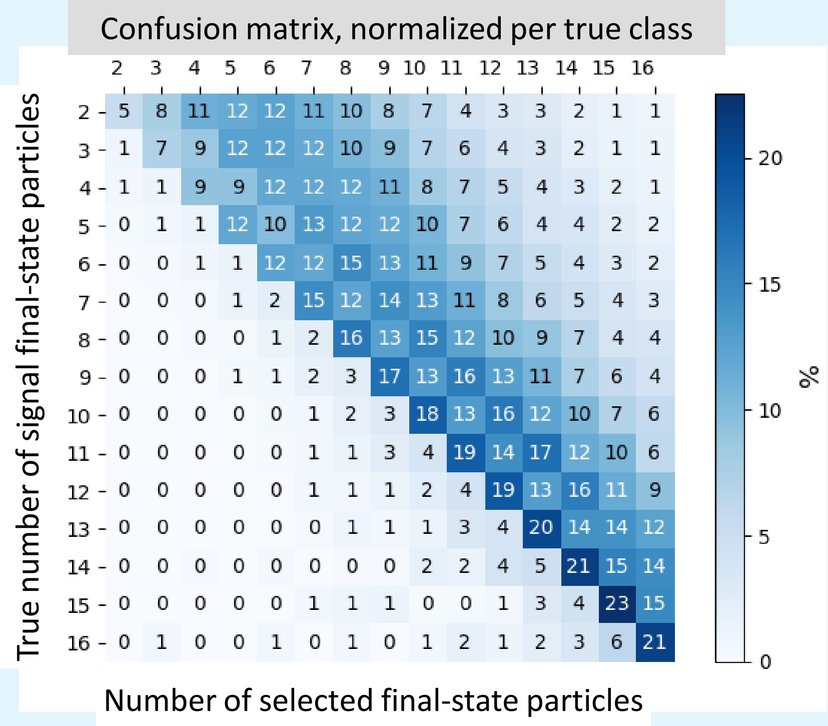
Conventional, exclusive analysis

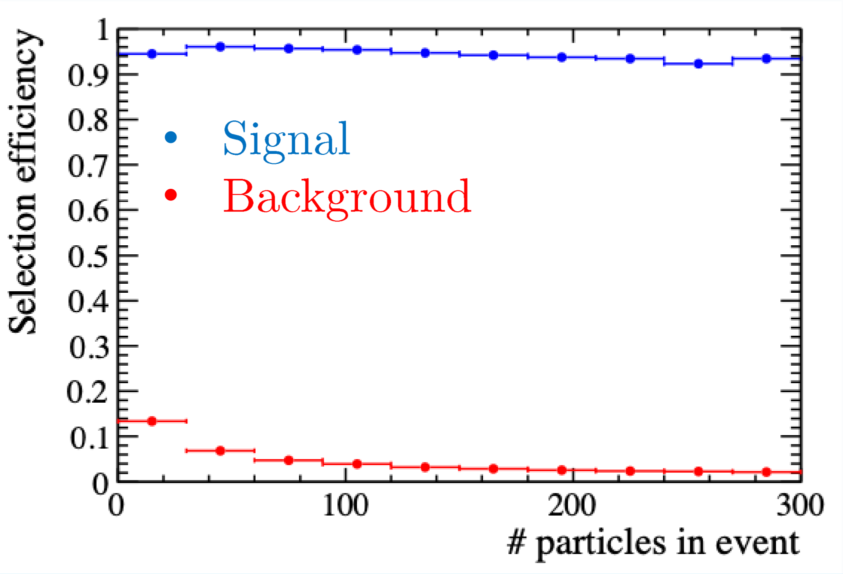
$B^+ \to K^+\pi^-\pi^+\ell^+\ell^-$
$B^0 \to K^+\pi^-\ell^+\ell^-$
$B^+ \to K^+\ell^+\ell^-$



With DFEI inclusive reconstruction

Performance: final-state particle filtering





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]

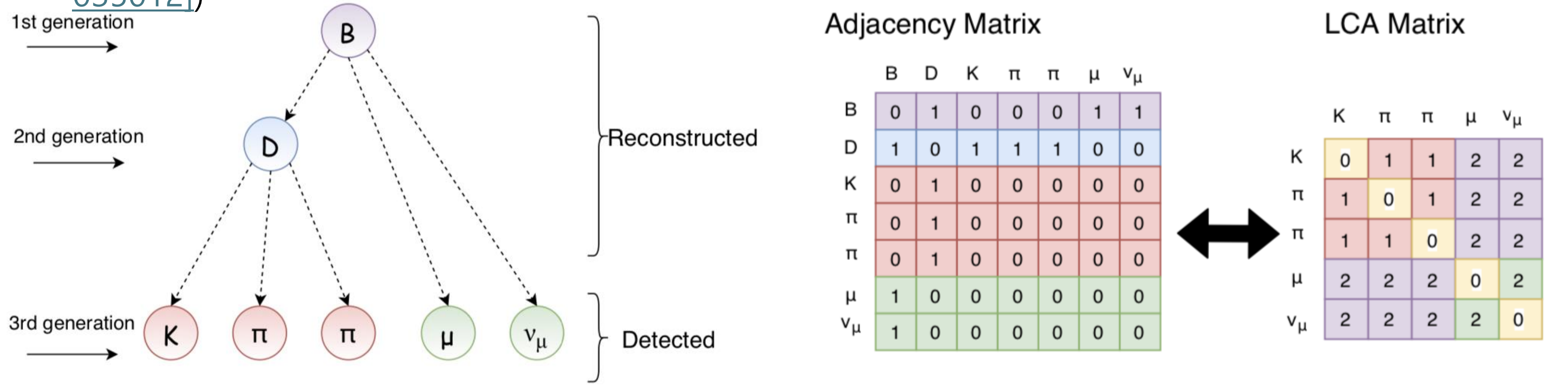
LHCb: 90% sig eff, 90% bkg rej. power
DFEI: 94% sig eff, 96% bkg rej. power

DFEI selects all of them simultaneously

DFEI capability #1
Powerful event size (~ x14) reduction in a multi-signal environment.

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.

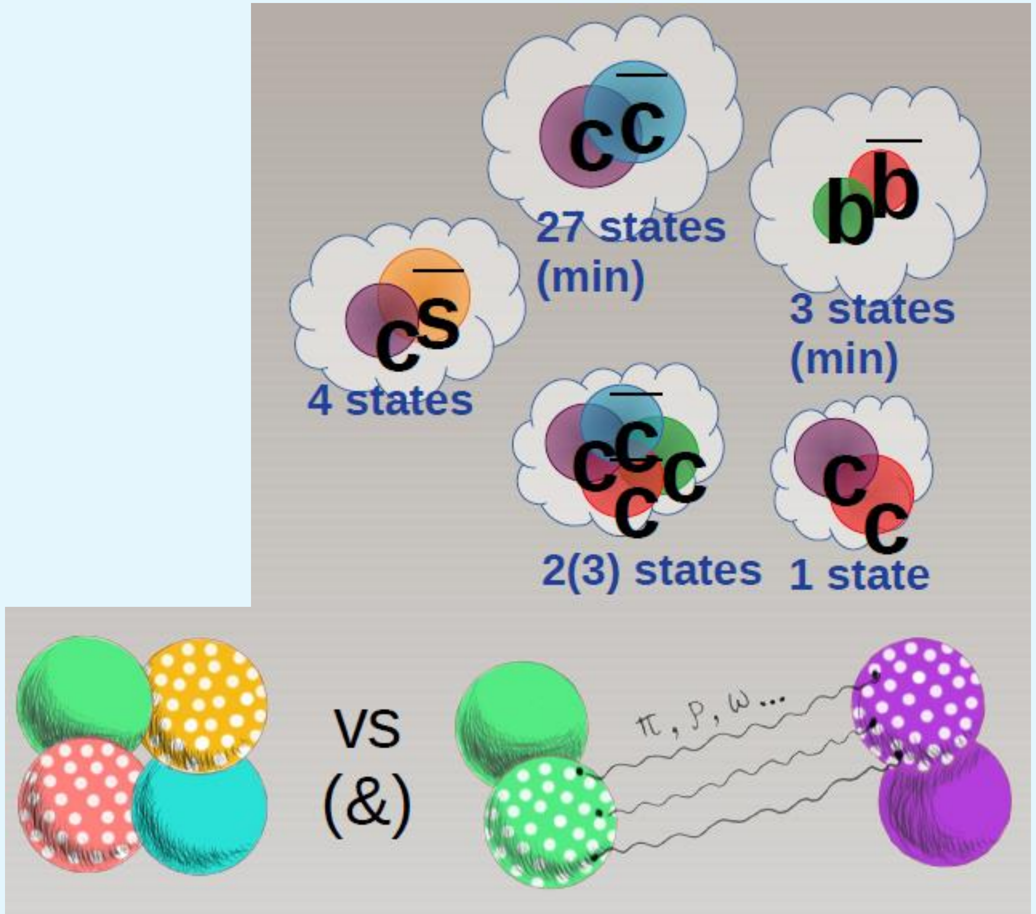
Decay-level performance

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

Different types of decay reconstruction
- wrong hierarchy: correct tracks but wrong hierarchy
- Not isolated: additional tracks that do not belong to the decay
- missing tracks of the true decay

Fraction of perfect signal reconstruction approximates the tag side efficiency for FEI at Belle (II) (order a few percent for semileptonic decays and a few per mille for hadronic decays.)

Search for exotic states



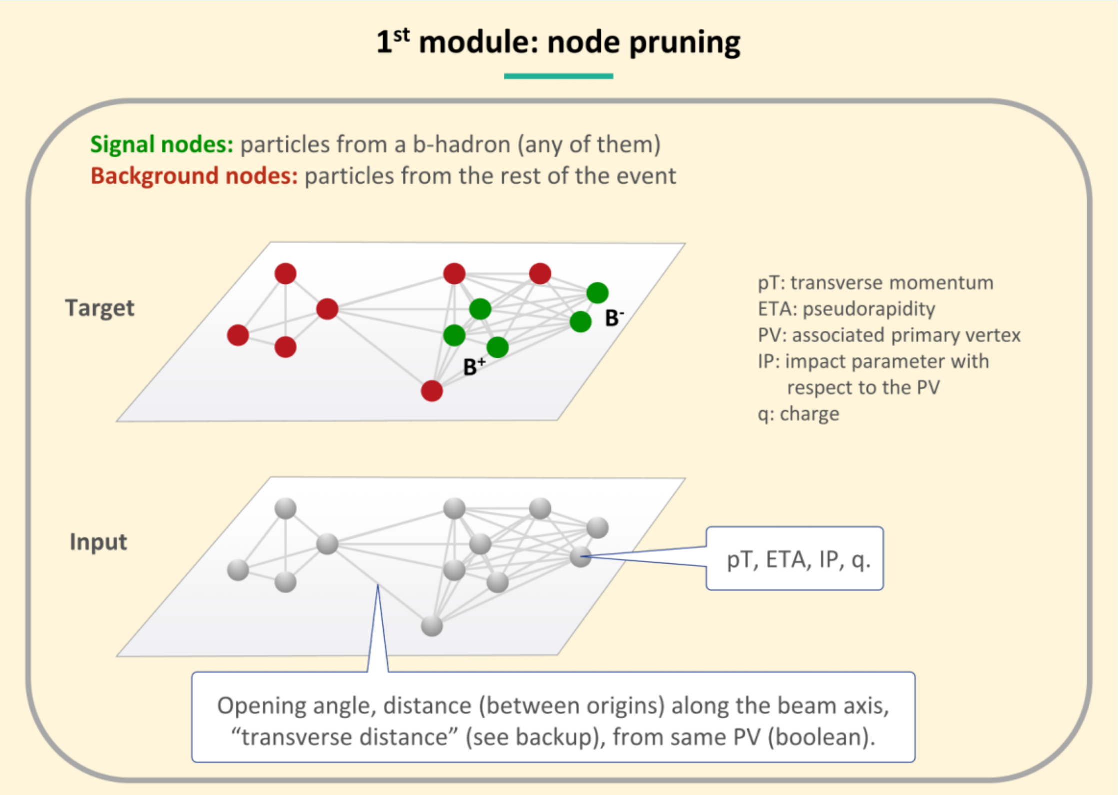
- Several states observed (50 since 2003)
- But: we still don't understand their nature
  - Bound or molecular states?
- Need unambiguous experimental evidence
- Other doubly-heavy states $[QQ\overline{ud}]$:
  - T_{bb} $[bb][\overline{ud}]$ → ~10^{-3} events in Run3&4
  - T_{bc} $[bc][\overline{ud}]$ → may be below $\bar{B}D$ threshold, but opposite expectations in some molecular models

Karliner, Rosner, 2017, Semay, SIlvestre-Brac, 1994, Carames, Vijande, Valcarce, 2019, Meng et al., 2021
Li, Sun, Liu, Zhu, 2012, Liu et al., 2019, Hudspith et al., 2020

Search for Tbc – prospects for Run3, I. Polyakov,
Hunting for the charming beauty tetraquark Tbc: LHCb meets theory, 5 October 2023, CERN

Algorithm Modules

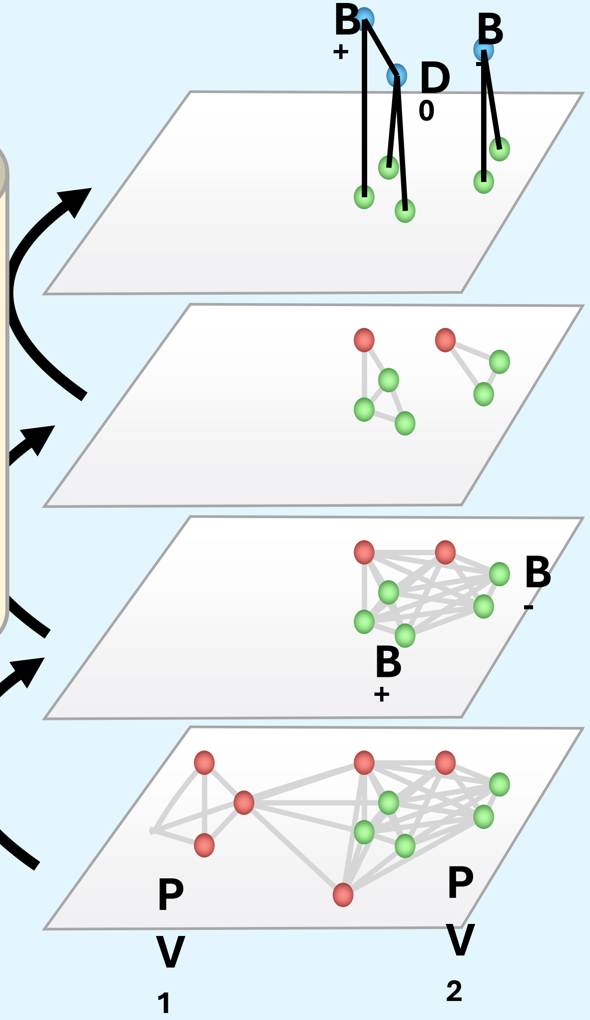
Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event



Goal: Remove most of the nodes not produced in any b-hadron decay

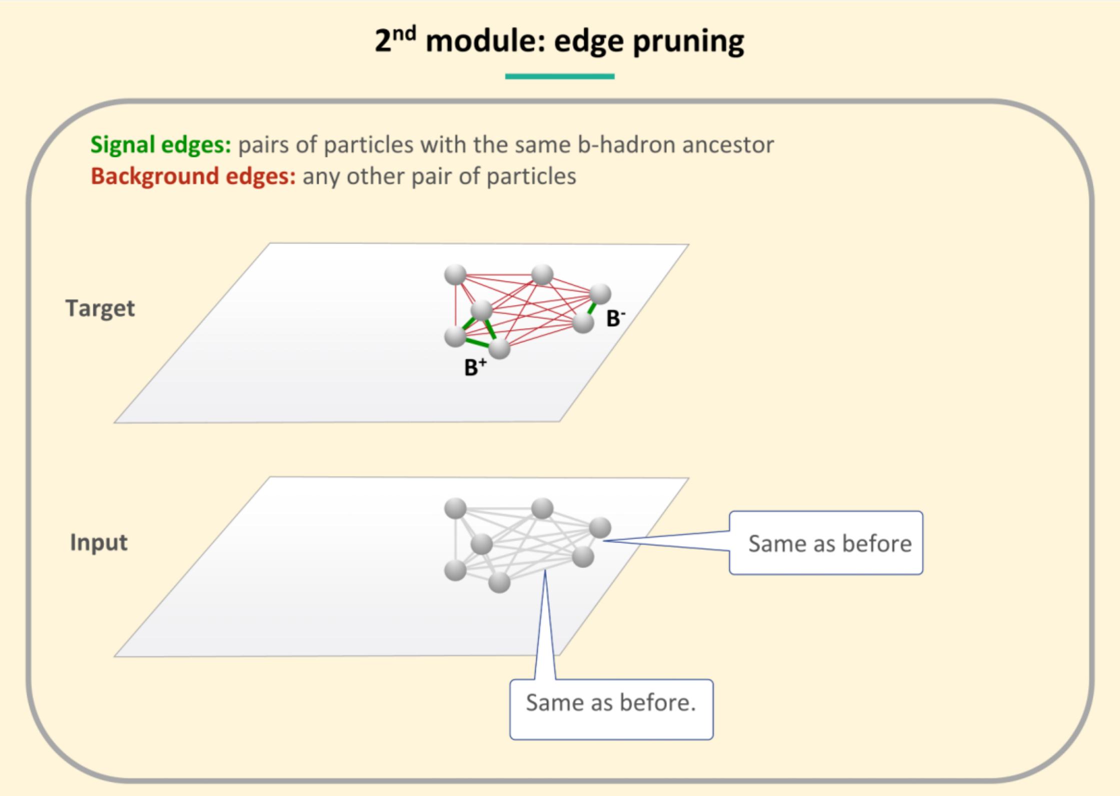
Cut @ 99% ε_{signal}
~70% BKG rejection

1. Node pruning



Algorithm Modules

Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event

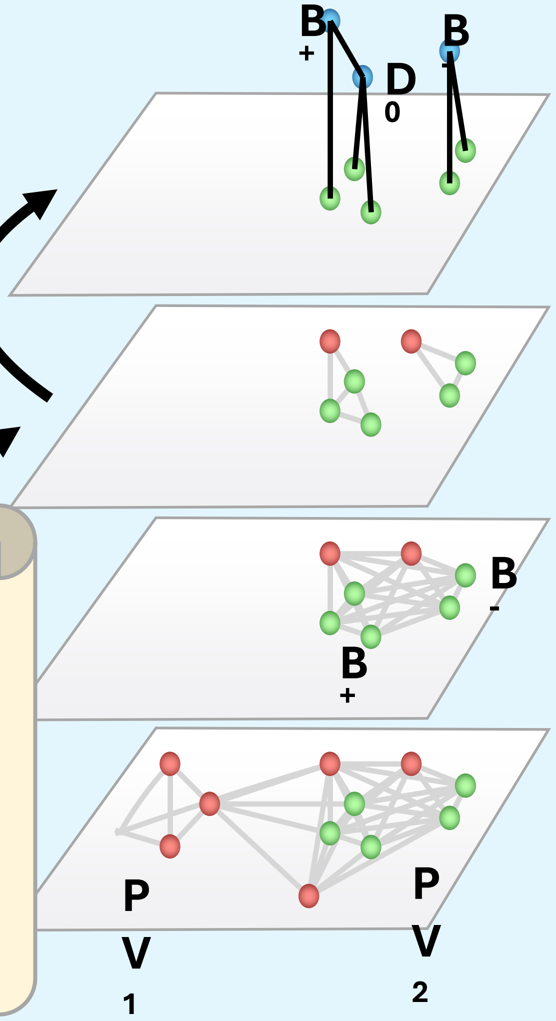


3. "Lowest common ancestor" inference

2. Edge pruning

Goal: Remove connections between particles that don't share the same ancestor

Cut @ 99% ε_{signal}
~68% BKG rejection

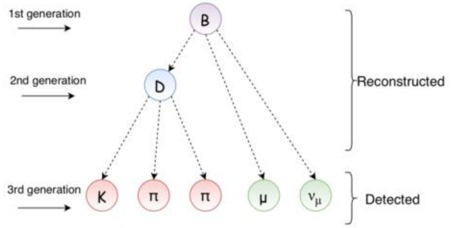


Algorithm Modules

Blue: reconstructed ancestors
Green: particles from a b-hadron
Red: particles from the rest of the event

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])



Adjacency Matrix

	B	D	K	π	π	μ	ν_μ
B	0	1	0	0	0	1	1
D	1	0	1	1	1	0	0
K	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
μ	1	0	0	0	0	0	0
ν_μ	1	0	0	0	0	0	0

⟷

LCA Matrix

	K	π	π	μ	ν_μ
K	0	1	1	2	2
π	1	0	1	2	2
π	1	1	0	2	2
μ	2	2	2	0	2
ν_μ	2	2	2	2	0

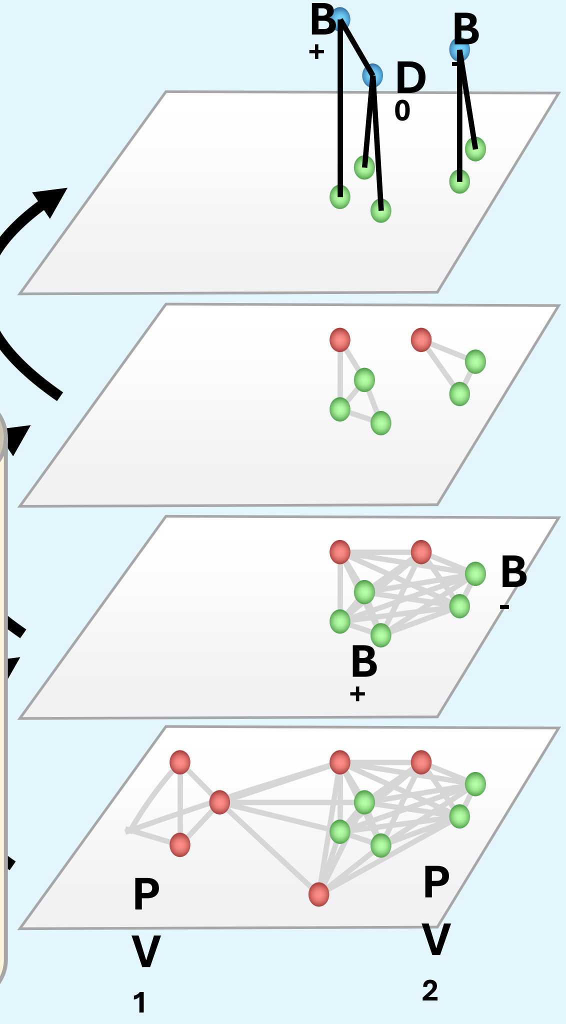
Problem reduced to **multi-class classification on edges.**

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

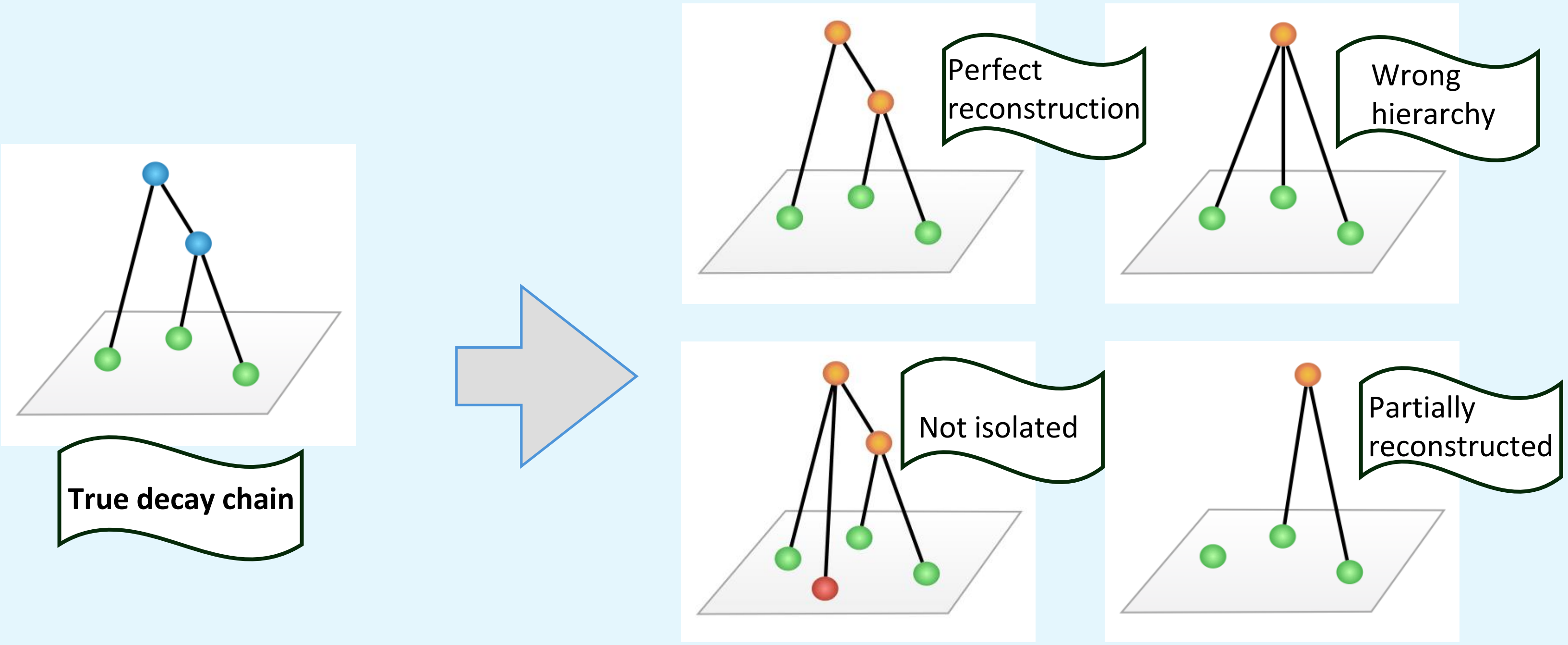
- Very-short-lived resonances merged with the previous ancestor.
- Resonances with less than two charged descendants merged with the previous ancestor.

3. "Lowest common ancestor" inference

Goal: Multi-class classification of the edges based on the level of the shared ancestor:
0) Different decay chain
1) B children
2) B grand-children



Physics performance: decay-chain reconstruction



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