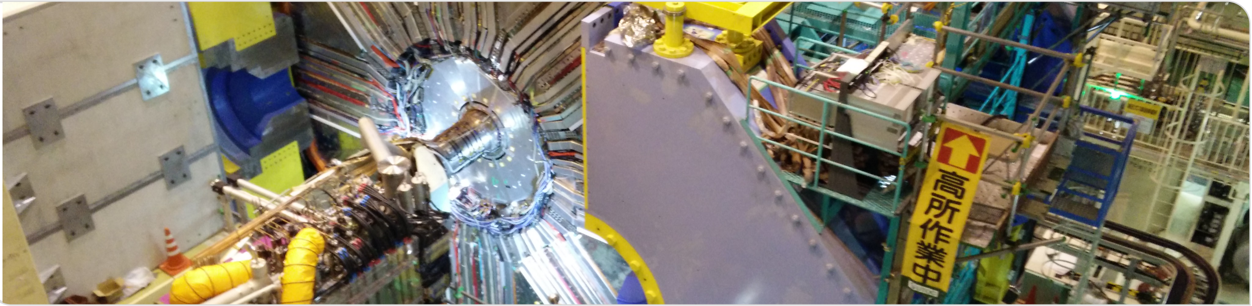
graFEI: Full Event Interpretation using Graph Neural Networks at Belle II

Lea Reuter, James Kahn, Torben Ferber on behalf of BaumBauen collaboration | Monday 28th February, 2022



Project members

KIT (ETP)

L. Reuter, T. Ferber, P. Goldenzweig

KIT (Helmholtz AI and SCC)

J. Kahn, O. Taubert, M. Götz

Uni-Bonn

I. Tsaklidis, T. Boeckh, F. Bernlochner

UoS (IPHC)

G. Dujany, A. Thaller, I. Ripp-Baudot

Particle Decay Trees

Why is b-tagging important?

- Reconstruct tag-side to constrain signal decays with neutrinos ν
- Maximise information extracted from each event

Challenging task:

- Large amount of possible decay channels
- High multiplicity in event decays

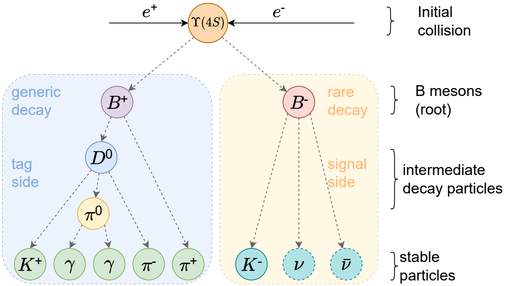
$\rightarrow$ **analytical solution intractable**

Clean Detection Environment at Belle II

event interpretation

$\equiv$

reconstructing signal- and tag-side

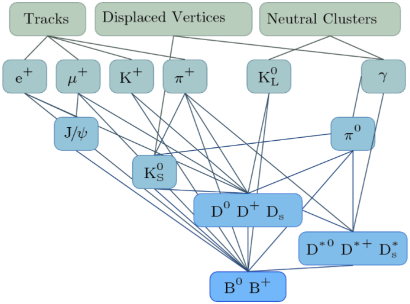


Goal:
identify structure of decay process up to B particles

The Belle II Full Event Interpretation

Currently used: Full Event Interpretation (**FEI**) algorithm as Exclusive Tagging Algorithm

- $\mathcal{O}(10000)$ distinct decay chains
- Hierarchical approach with six distinct stages
- Multivariate classification (BDTs)



efficiency	$B^{\pm}$ (%)	B^{0} (%)
Hadronic	0.76	0.46
Semileptonic	1.80	2.04

$\rightarrow$ **Efficiency**: $\frac{\textbf{correctly reconstructed}\text{ B decays}}{\text{all B decays}}$

[1] *The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment*, Keck et al. (2018), arxiv:1807.08680

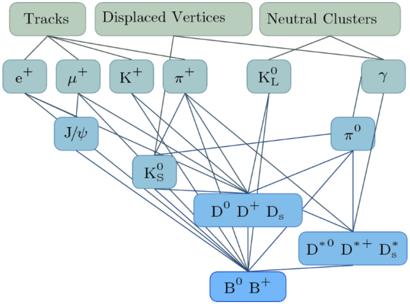
The Belle II Full Event Interpretation

Currently used: Full Event Interpretation (**FEI**) algorithm as Exclusive Tagging Algorithm

- $\mathcal{O}(10000)$ distinct decay chains
- Hierarchical approach with six distinct stages
- Multivariate classification (BDTs)

efficiency	$B^{\pm}$ (%)	B^{0} (%)
Hadronic	0.76	0.46
Semileptonic	1.80	2.04

→ **Efficiency**: $\frac{\textbf{correctly reconstructed}\text{ B decays}}{\text{all B decays}}$



Limitations

- Hard coded decay channels **restrict branching fraction** to $\sim 15\%$
- Hierarchical structure leads to **error propagation**

[1] *The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment*, Keck et al. (2018), arxiv:1807.08680

Next Generation FEI

Karlsruhe Institute of Technology

- **End to end** trainable network
- Data driven **learning by example** to exploit full MC coverage

Next Generation FEI

- **End to end** trainable network
- Data driven **learning by example** to exploit full MC coverage

Framing the Problem

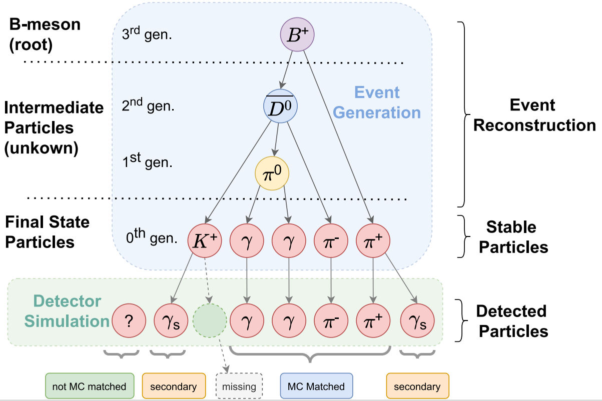
Have: Final state particles (FSPs)

Want: Entire decay structure

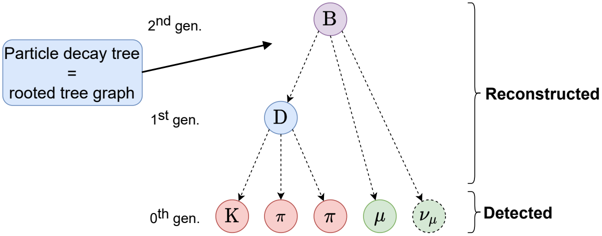
Challenges of event reconstruction:

- **Variable number** of final state particles
- Unknown **tree structure** (depth and breadth)
- Unknown number of **intermediate particles**
- Undetected FSPs
- Missing ground truth

Solution: Encode tree as FSP relations

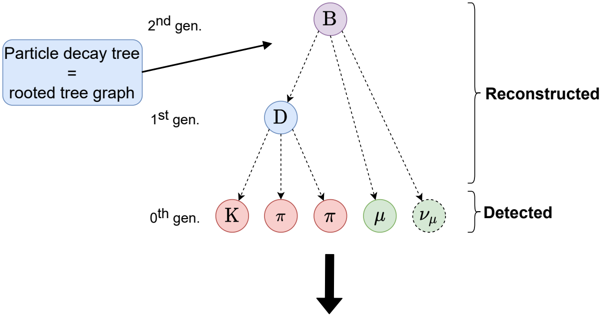


Graph-based Full Event Interpretation (graFEI)



Graph-based Full Event Interpretation (graFEI)

Karlsruhe Institute of Technology

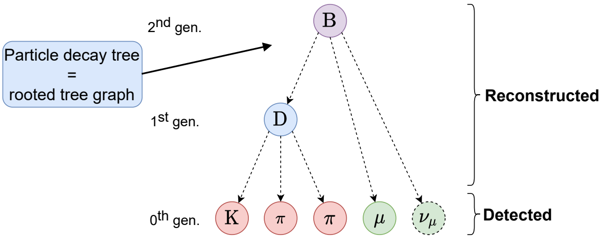


Adjacency Matrix

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

Graph-based Full Event Interpretation (graFEI)

Karlsruhe Institute of Technology



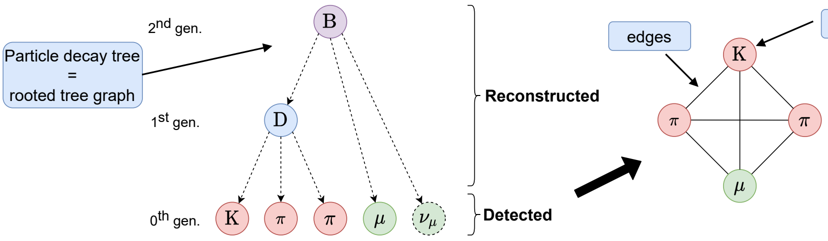
Adjacency Matrix

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

Lowest Common Ancestor (LCA) Matrix

	K	π	π	μ
K	K	D	D	B
π	D	π	D	B
π	D	D	π	B
μ	B	B	B	μ

Graph-based Full Event Interpretation (graFEI)



Feature Matrix

E_K	E_π	E_π	E_μ
p_{xK}	$p_{x\pi}$	$p_{x\pi}$	$p_{x\mu}$
p_{yK}	$p_{y\pi}$	$p_{y\pi}$	$p_{y\mu}$
p_{zK}	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

Adjacency Matrix

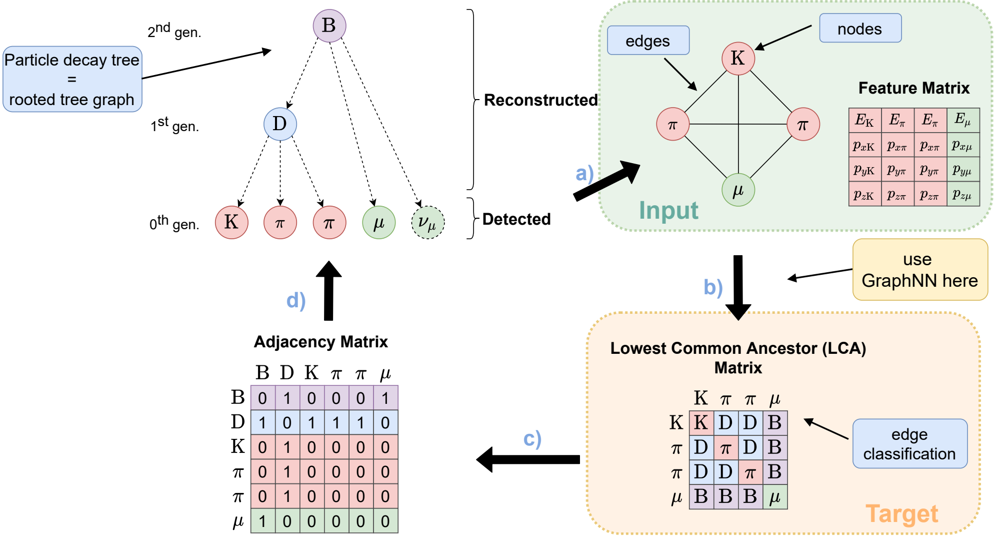
	B	D	K	π	π	μ
B	0	1	0	0	0	1
D	1	0	1	1	1	0
K	0	1	0	0	0	0
π	0	1	0	0	0	0
π	0	1	0	0	0	0
μ	1	0	0	0	0	0

Lowest Common Ancestor (LCA) Matrix

	K	π	π	μ
K	K	D	D	B
π	D	π	D	B
π	D	D	π	B
μ	B	B	B	μ

edge classification

Graph-based Full Event Interpretation (graFEI)



Feature Matrix

E_K	E_π	E_π	E_μ
p_{xK}	$p_{x\pi}$	$p_{x\pi}$	$p_{x\mu}$
p_{yK}	$p_{y\pi}$	$p_{y\pi}$	$p_{y\mu}$
p_{zK}	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

Lowest Common Ancestor (LCA) Matrix

	K	π	π	μ
K	K	D	D	B
π	D	π	D	B
π	D	D	π	B
μ	B	B	B	μ

Adjacency Matrix

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

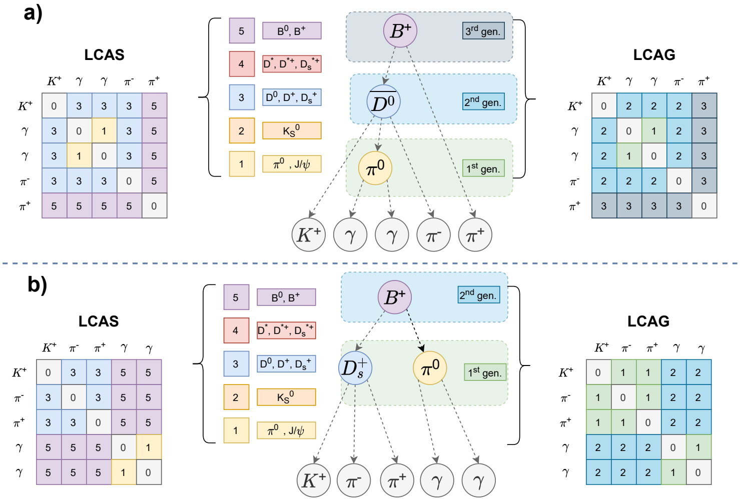
Lowest Common Ancestor Matrix Representations

Karlsruhe Institute of Technology

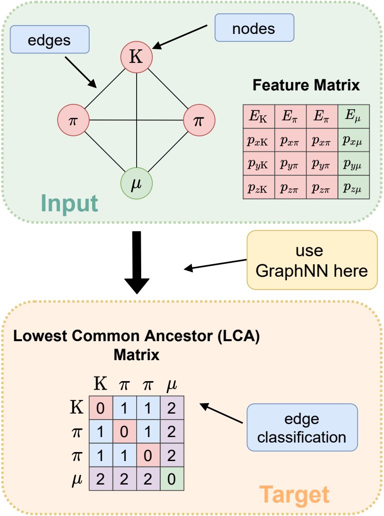
Need appropriate representation for machine learning

LCAG generational representation of intermediate particles

LCAS stage-view representation inspired by FEI stages



Interaction learning



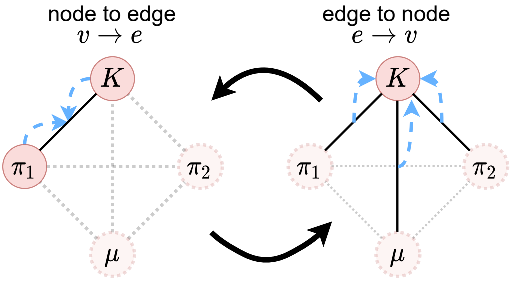
Feature Matrix

E_K	E_π	E_π	E_μ
p_{xK}	$p_{x\pi}$	$p_{x\pi}$	$p_{x\mu}$
p_{yK}	$p_{y\pi}$	$p_{y\pi}$	$p_{y\mu}$
p_{zK}	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

Lowest Common Ancestor (LCA) Matrix

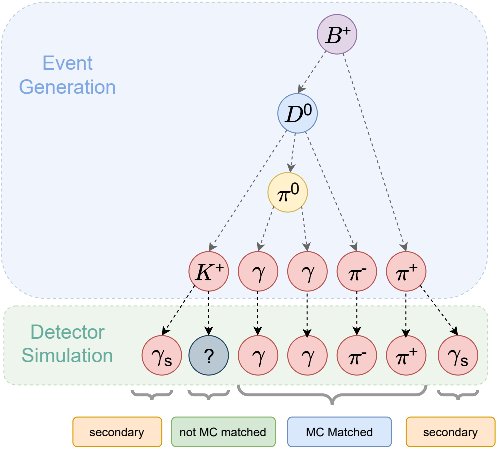
	K	π	π	μ
K	0	1	1	2
π	1	0	1	2
π	1	1	0	2
μ	2	2	2	0

Message passing mechanism



All combined in **Neural Relational Inference for Interacting Systems** (arxiv:1802.04687)

Building the training LCA



True LCA

	K^+	γ	γ	π^-	π^+
K^+	-1	2	2	2	3
γ	2	-1	1	2	3
γ	2	1	-1	2	3
π^-	2	2	2	-1	3
π^+	3	3	3	3	-1

Training LCA

	γ	γ	π^-	π^+	γ_s	γ_s	K^+
γ	-1	1	2	3	0	0	0
γ	1	-1	2	3	0	0	0
π^-	2	2	-1	3	0	0	0
π^+	3	3	3	-1	0	0	0
γ_s	0	0	0	0	-1	0	0
γ_s	0	0	0	0	0	-1	0
K^+	0	0	0	0	0	0	-1

Detected Belle II simulated data includes:

- missing particles
- unmatched particles
- secondary particles
- duplicate particles

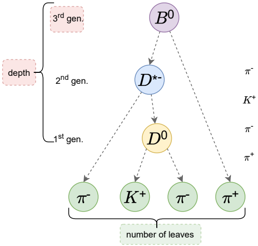
Build Training LCA with:

(-1) To tell PyTorch to ignore it (diagonal and padding)

(0) As background class (unmatched/secondary)

Single decay: training LCA distribution

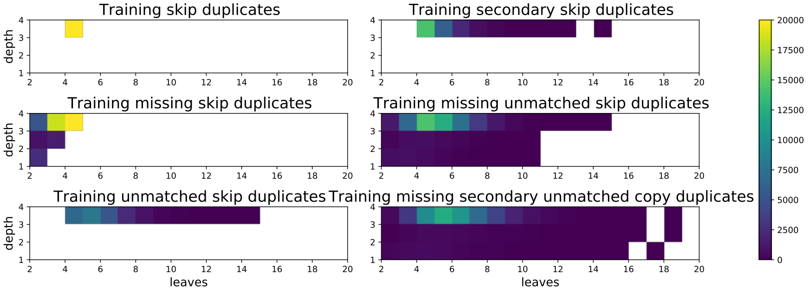
example B^0 decay



True LCA

	π^-	K^+	π^-	π^+
π^-	-1	2	2	3
K^+	2	-1	1	3
π^-	2	1	-1	3
π^+	3	3	3	-1

Dataset distrubution gets smeared when including unmatched, missing, secondary particles and including duplicates



Features:

- Particle IDs
- Kinematics
- Vertex information
- Basic particle information
- ECL Cluster variables

Goal: comparison on Belle II simulated data

Compare performance of **B-tag reconstruction** when the **B-sig** $B^0 \to \nu\bar{\nu}$ for **FEI** and **graFEI**, using simulated samples including detector effects and beam background

FEI

- **tagging efficiency**: fraction of reconstructed B^0 decays of all decays
- **tag-side efficiency**: fraction of **correctly** reconstructed B^0 decays of all decays
- **purity**: fraction of correctly reconstructed decays out of all reconstructed decays

graFEI

- **valid-tree efficiency**: fraction of B^0 decays with a rooted, directed, acyclic, predicted tree
- **accuracy**: amount of particle edges that get classified correctly (independet of B^0 decays)
- **perfect LCA**: fraction of B^0 decays with a **correctly** predicted LCA
- **purity**: fraction of perfect LCA out of all decays with valid trees

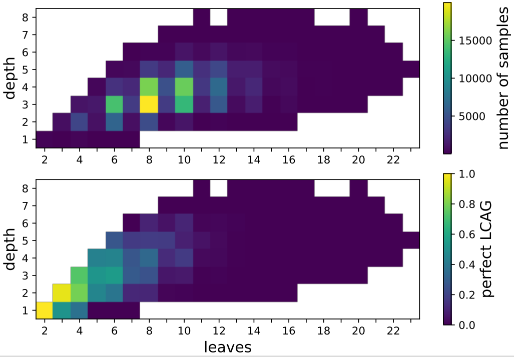
Training graFEI on mono-generic hadronic B-decays

mono-generic: one B^0-decay per sample, where the B^0-meson decays according to Belle II phasespace

best-case ideal scenario : 21.6% perfect LCAG

Accuracy 62.5%

Trained on 1.8 million samples



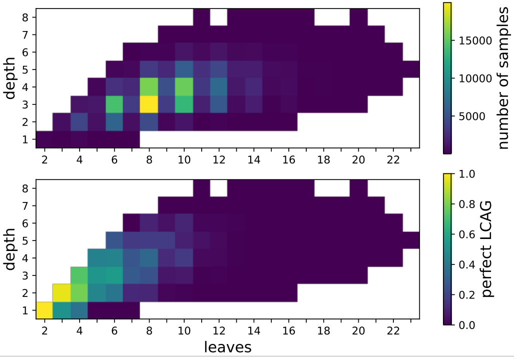
Training graFEI on mono-generic hadronic B-decays

mono-generic: one B^0-decay per sample, where the B^0-meson decays according to Belle II phasespace

best-case ideal scenario : 21.6% perfect LCAG

Accuracy 62.5%

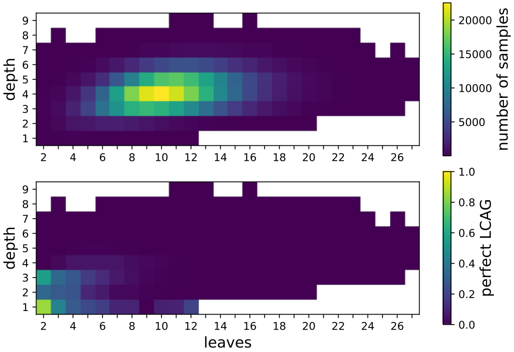
Trained on 1.8 million samples



realistic scenario: 1.8% perfect LCAG

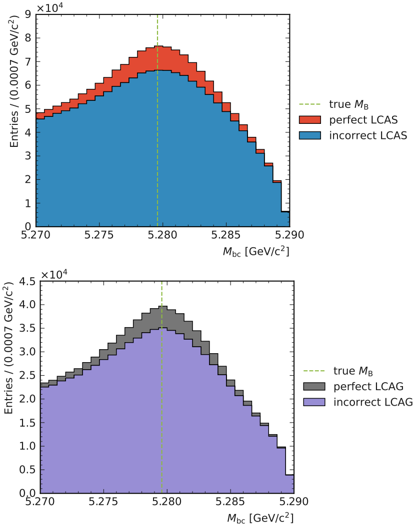
Accuracy 31.3%

Trained on 37 million samples



Comparison on mono-generic hadronic B-decays

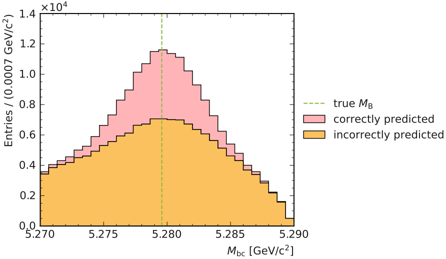
graFEI



Evaluated on 12 million events for **B-tag reconstruction** when the **B-sig** $B^0 \to \nu\bar{\nu}$ on beam-constrained mass M_{bc}

$$M_{\mathrm{bc}} = \sqrt{(E_{\mathrm{beam}}^{\mathrm{CMS}})^2/c^4 - (\boldsymbol{p}_{\mathrm{B}}^{\mathrm{CMS}})^2/c^2} \qquad (1)$$

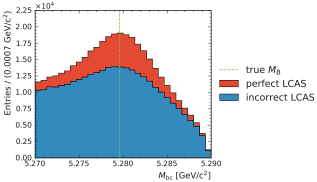
FEI

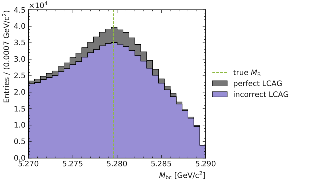


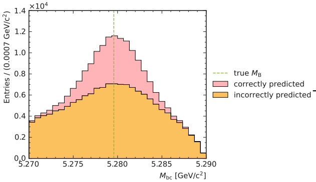
	FEI (%)	graFEI (%)		
metrics		LCAS	LCAG	metrics
tagging efficiency	1.87	13.96	6.89	valid-tree
tag-side efficiency	**0.41**	**1.32**	**0.54**	**perfect LCA**
purity	21.92	9.46	7.84	purity

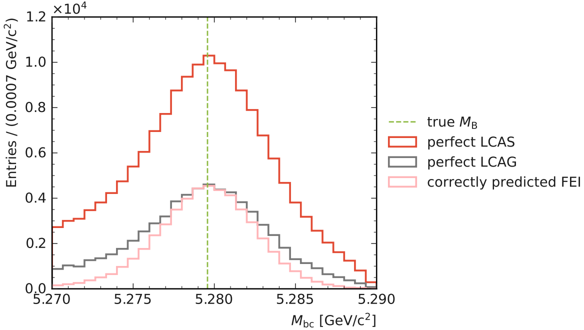
Comparison on mono-generic hadronic B-decays

Karlsruhe Institute of Technology









	FEI (%)	graFEI (%)			
metrics		LCAS	imp. LCAS	LCAG	metrics
tagging efficiency	1.87	13.96	3.19	6.89	valid-tree
tag-side efficiency	**0.41**	**1.32**	**0.66**	**0.54**	**perfect LCA**
purity	21.92	9.46	20.7	7.84	purity

Summary and Outlook

Summary

- Strategy for tree reconstruction with **minimal assumptions** about ancestors
- Can model **all trajectories**, not only hard-coded events
- End-to-end trainable
  - Data-driven solution
  - **No hierarchical error propagation**
- Promising first results on mono-generic decays: improvement from 0.41% to 1.32%

$\rightarrow$ paving the way for the development of an end-to-end trainable, graph-based reconstruction algorithm

Submission to ICML!

Summary and Outlook

Summary

- Strategy for tree reconstruction with **minimal assumptions** about ancestors
- Can model **all trajectories**, not only hard-coded events
- End-to-end trainable
  - Data-driven solution
  - **No hierarchical error propagation**
- Promising first results on mono-generic decays: improvement from 0.41% to 1.32%

$\rightarrow$ paving the way for the development of an end-to-end trainable, graph-based reconstruction algorithm

Submission to ICML!

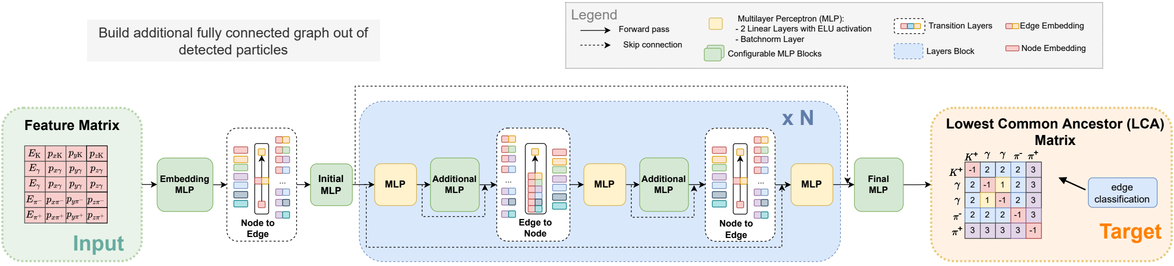
Outlook

- Tailor Trainings to analyses and improve final state particle selection
- Distinguish between signal and background (implement signal probability)
- Input feature optimization

Backup

Model: Neural Relational Inference[1] Encoder

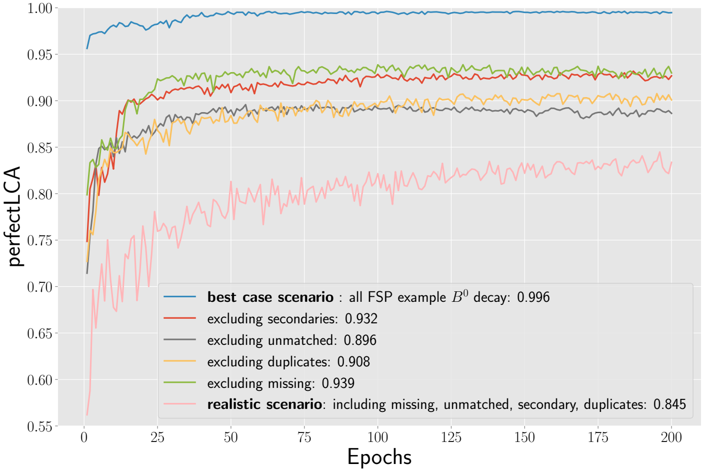
Karlsruhe Institute of Technology



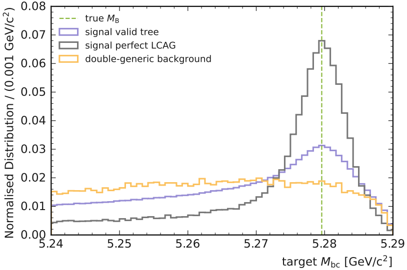
- **Iterative** node $v \leftrightarrow$ edge e **update** and hierarchical agglomeration
- Implementation in `PyTorch` and `Optuna`
- Experiments using `NVIDIA V100`, `NVIDIA V100s` and `NVIDIA A100` GPUs provided by TOpAS the `NVIDIA A100s` in HoreKa

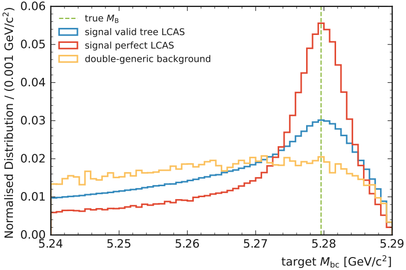
[1] *Neural Relation Inference or Interacting Systems*, Kipf et al. (2018), arXiv:1802.04687v2

Ablation Study on Experimental Effects



Double-Generic Mixed Background Decays





	background double-generic decays		single decays			
metrics in %	LCAS	LCAG	LCAS		LCAG	
	valid-tree	valid-tree	perfect	valid-tree	perfect	valid-tree
Mbc constrained	1.15	0.89	1.32	13.96	0.54	6.89
Mbc constrained and no class-label 0 leaves	0.0068	0.0012	0.77	6.24	0.31	2.31