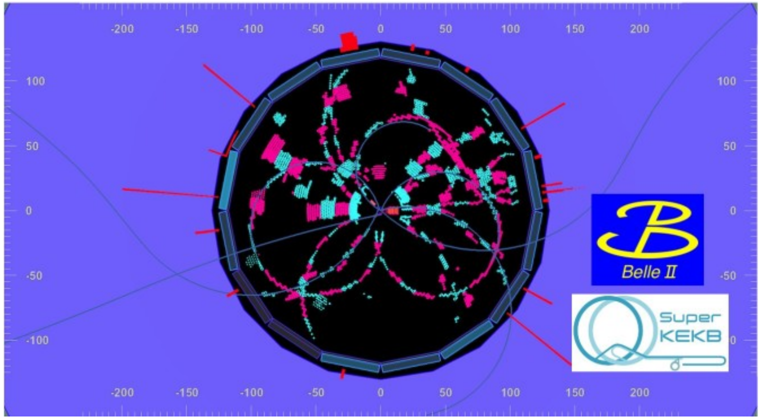
B decays reconstruction at Belle II using Graph Neural Networks

IN2P3/IRFU Machine Learning workshop - 28/09/2022

Jacopo Cerasoli
CNRS - IPHC

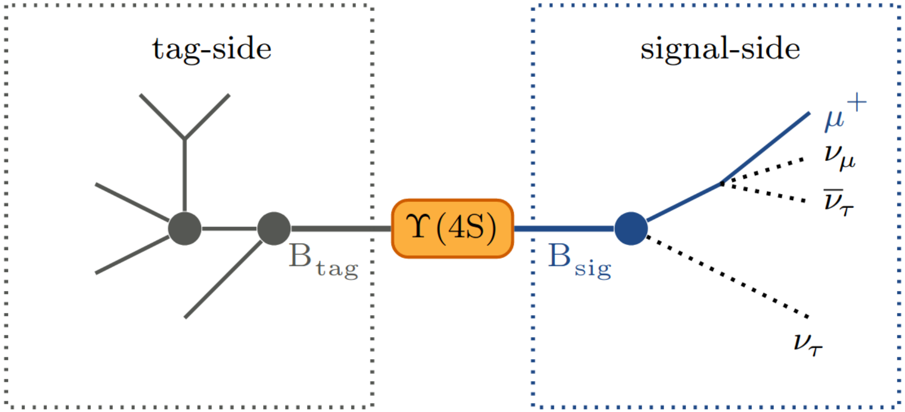
Belle II [arXiv:1011.0352]

- **Multi-purpose detector @ SuperKEKB** accelerator
- Collisions at center-of-mass energy of 10.58 GeV
  - $\sigma(e^+e^- \to \Upsilon(4S)) \sim 1$ nb
  - $\mathcal{B}(\Upsilon(4S) \to B\bar{B}) \gtrsim 96\%$

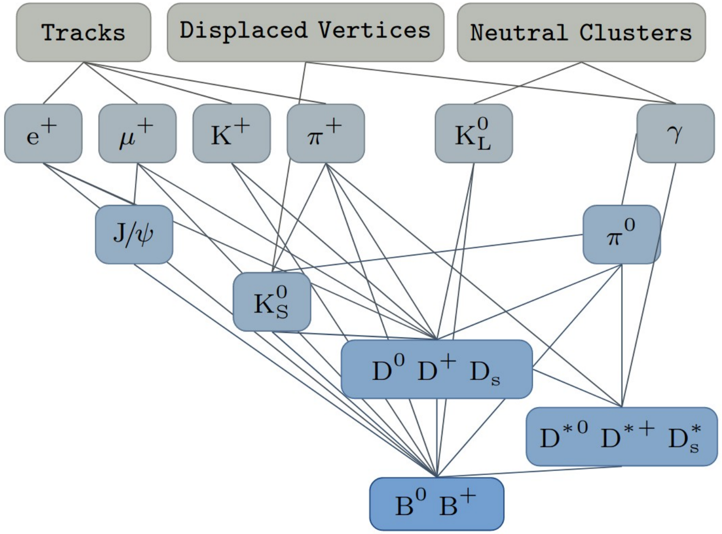


- **Will collect 50 ab^{-1}** at the end of operation (now ~ 430 fb^{-1})
- Instantaneous luminosity world record: **4.7 x 10^{34} cm^{-2} s^{-1}** (June 2022)
- **Important complementarity with LHCb**
  - Independent checks if new physics discovered

Full Event Interpretation [arXiv:1807.08680]

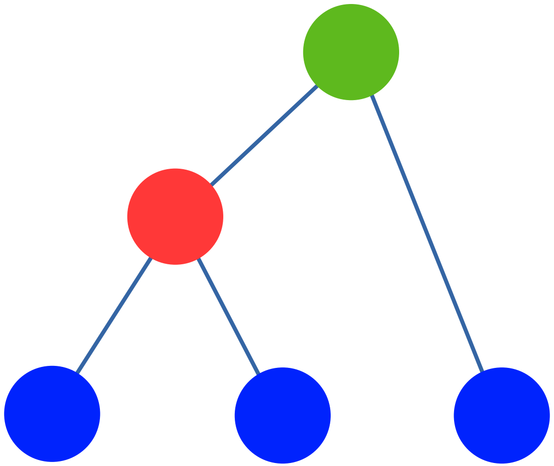


- Interested in **final states with neutrinos**
  - Reconstruct *tag-side* to constrain the kinematics



- Present solution: **Full Event Interpretation (FEI)**
  - Hierarchical approach based on BDTs
  - B reconstructed in **more than 10k modes**!
  - Overall reconstruction efficiency $\mathcal{O}(1\%)$
  - Output of final stage interpreted as "B probability"
- Cons: decay modes hard-coded, ~ 85 % B decays not considered

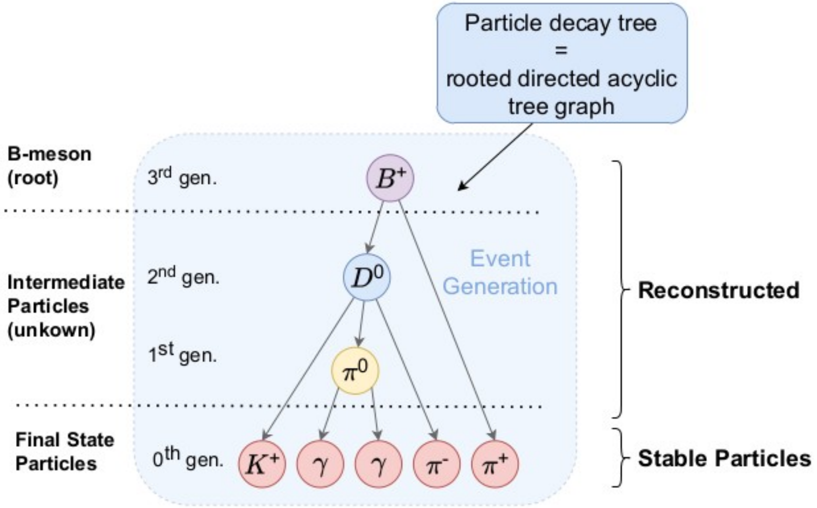
B reconstruction using Graph Neural Networks



- **Particle decays are naturally described by tree graphs**
- Goal: develop **graph-based Full Event Interpretation (graFEI)**

- Today's menu:
  - Proof of concept: *Learning tree structures from leaves for particle decay reconstruction*, James Kahn et al 2022 [arXiv:2208.14924] (see also Ilias Tsaklidis' and Lea Reuter's master theses)
  - First (preliminary) results on Belle II simulated dataset (see also Arthur Thaller's talk)

Lowest Common Ancestor (LCA) matrix



Adjacency Matrix

	B^+	D^0	π^0	K^+	γ	γ	π^-	π^+
B^+	0	1	0	0	0	0	0	1
D^0	1	0	1	1	0	0	1	0
π^0	0	1	0	0	1	1	0	0
K^+	0	1	0	0	0	0	0	0
γ	0	0	1	0	0	0	0	0
γ	0	0	1	0	0	0	0	0
π^-	0	1	0	0	0	0	0	0
π^+	1	0	0	0	0	0	0	0

Lowest Common Ancestor (LCA) Matrix

	K^+	γ	γ	π^-	π^+
K^+	K^+	D^0	D^0	D^0	B^+
γ	D^0	γ	π^0	D^0	B^+
γ	D^0	π^0	γ	D^0	B^+
π^-	D^0	D^0	D^0	π^-	B^+
π^+	B^+	B^+	B^+	B^+	π^+

LCAG

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

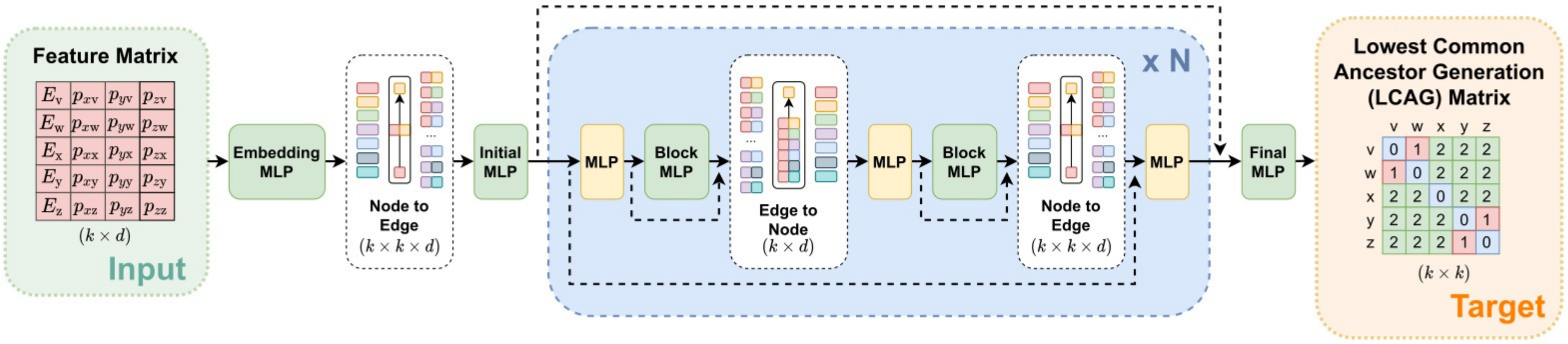
LCAS

	K^+	γ	γ	π^-	π^+
K^+	0	3	3	3	5
γ	3	0	1	3	5
γ	3	1	0	3	5
π^-	3	3	3	0	5
π^+	5	5	5	5	0

5	B^0, B^+
4	D^*, D^{*+}, D_s^{*+}
3	D^0, D^+, D_s^+
2	K_S^0
1	π^0 , J/ψ

L. Reuter. *Full Event Interpretation using Graph Neural Networks* (link)

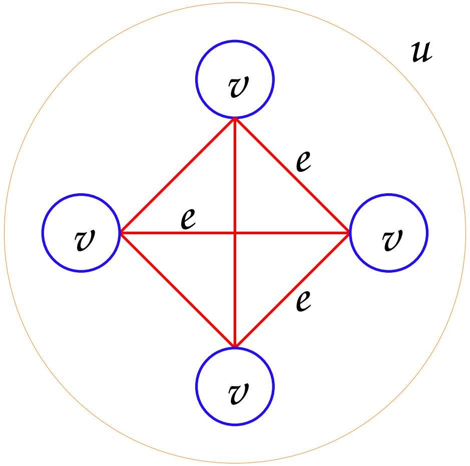
graFEI on Phasespace dataset [arXiv:2208.14924]

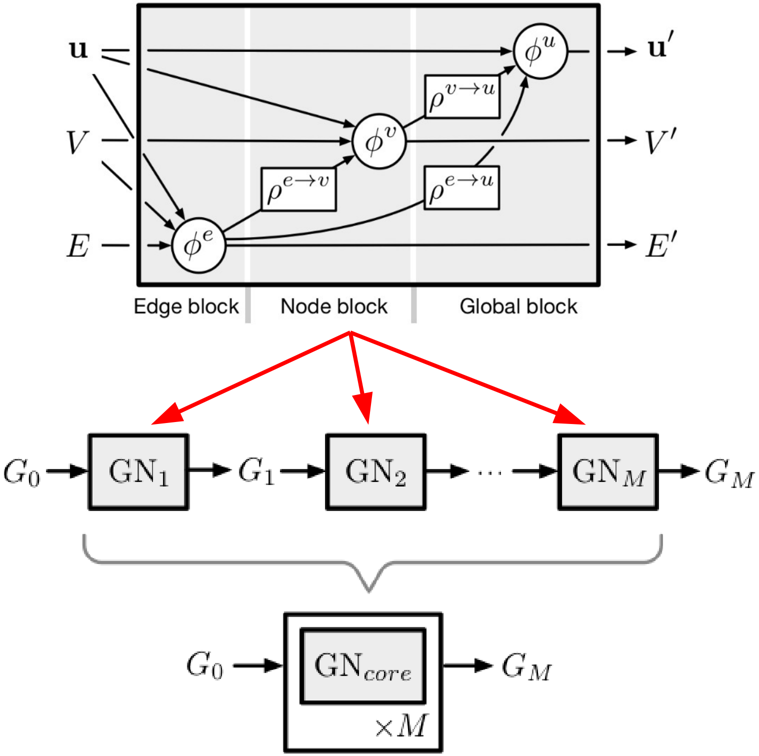


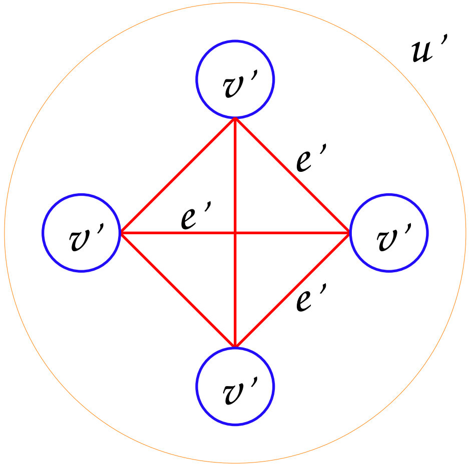
- Neural Relational Inference (NRI) model [arXiv:1802.04687]
- Dataset generated with Phasespace library
- 4-momentum used as input feature
- Average 47.7 % perfectly predicted LCAG on Phasespace dataset (60.9 % for decays with up to 10 leaves, 94.2 % up to 6 leaves)

GraFEI on Belle II simulated dataset – Graph network block

- "Geometric" graph neural network [arXiv:1806.01261]
- Input graph is transformed with series of GN blocks, output graph has same structure with updated values





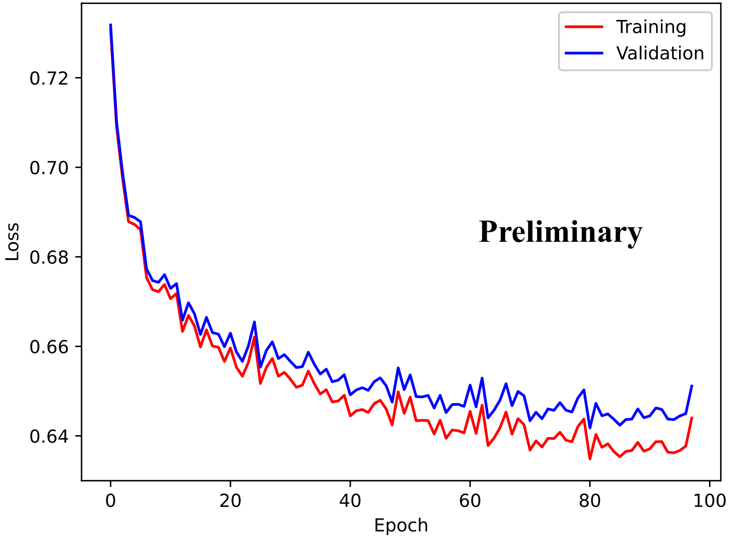


$$\begin{aligned} \mathbf{e}'_k &= \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right) & \bar{\mathbf{e}}'_i &= \rho^{e\to v}\left(E'_i\right) \\ \mathbf{v}'_i &= \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right) & \bar{\mathbf{e}}' &= \rho^{e\to u}\left(E'\right) \\ \mathbf{u}' &= \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}\right) & \bar{\mathbf{v}}' &= \rho^{v\to u}\left(V'\right) \end{aligned}$$

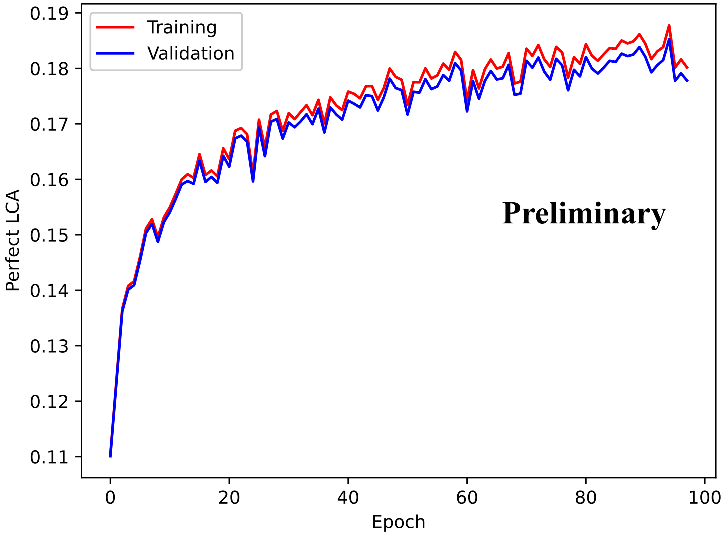
GraFEI on Belle II simulated dataset – Training

- Training done with monogeneric $\Upsilon(4S) \to B^0(\to \nu\nu)\bar{B}^0(\to \mathrm{X})$ simulated Belle II sample
  - Node-level features: particle IDs, 4-momentum, M, charge, impact parameter, photon cluster features
  - Edge-level feature: angle between pairs of particles' momenta
  - Global feature: number of leaves (i.e. final state particles)

$$\text{Cross-entropy} = -\frac{1}{N}\sum_i \log \frac{e^{x_{i,y_n}}}{\sum_c e^{x_{i,y_c}}}$$



$$\text{Perfect LCA} = \frac{\text{Perfectly predicted LCAs}}{\text{Total predicted LCAs}}$$

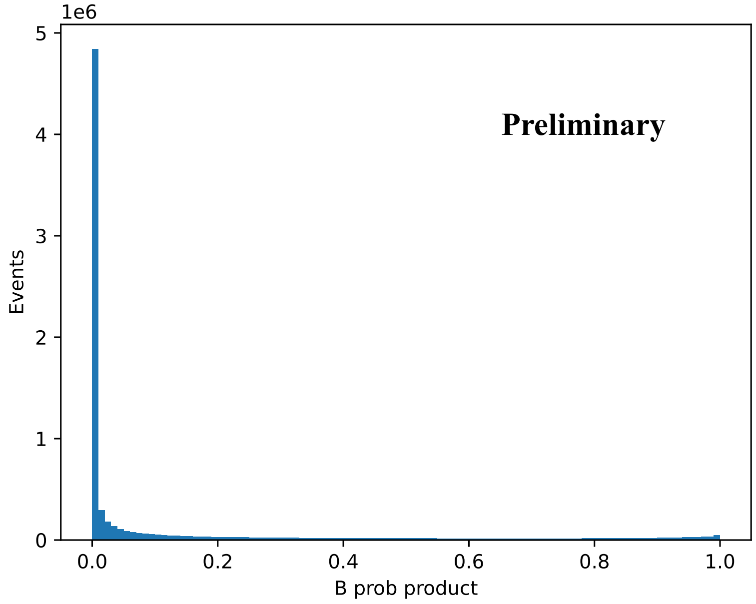


Max perfect LCA = 18.6 % (epoch 94)

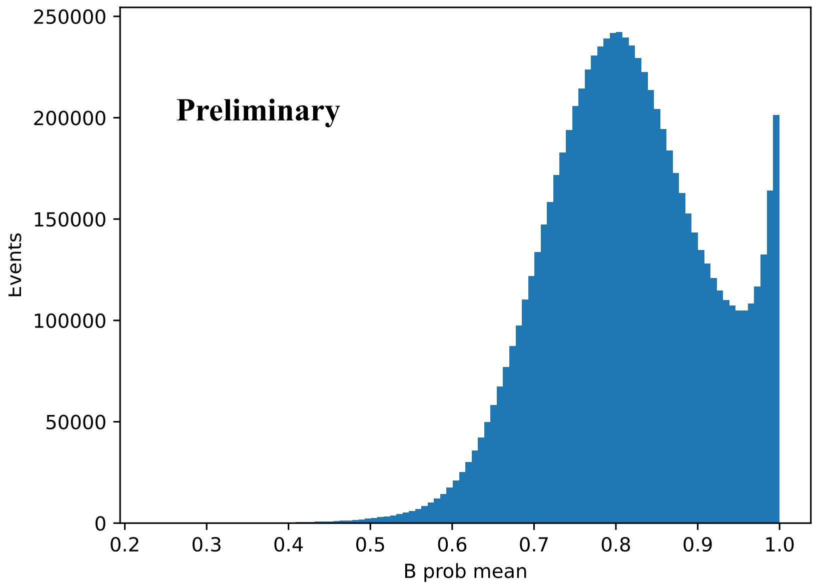
GraFEI on Belle II simulated dataset – B probability

- Having a **definition of "B probability"** analogous to FEI is desirable
  - Each LCA element has corresponding probability of belonging to the predicted class

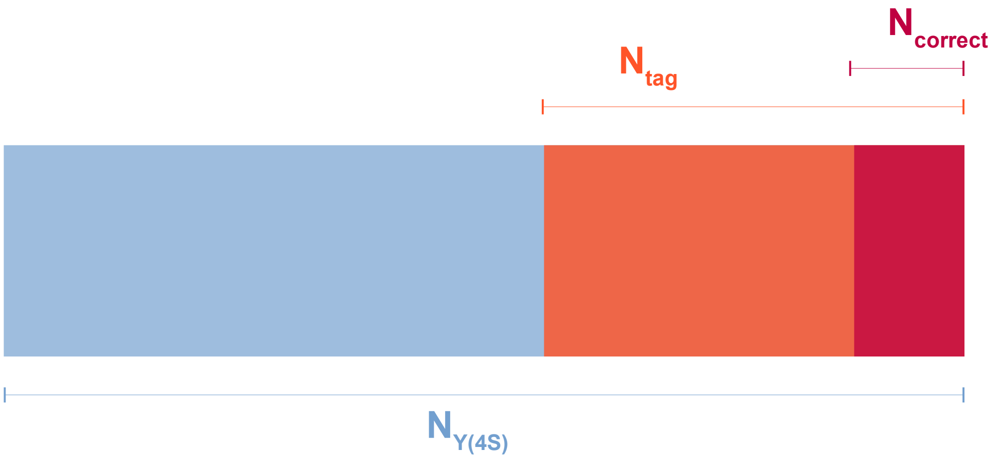
Product of class probabilities



Arithmetic mean of class probabilities



GraFEI on Belle II simulated dataset – Comparison with FEI



- **Tag-side efficiency** = $N_{correct}$ / $N_{Y(4S)}$
- **Purity** = $N_{correct}$ / N_{tag}

	FEI	graFEI
N_{tag}	# Events reconstructed	# Events with valid tree
$N_{correct}$	# Correctly reconstructed events*	# Events with perfectly predicted LCAS

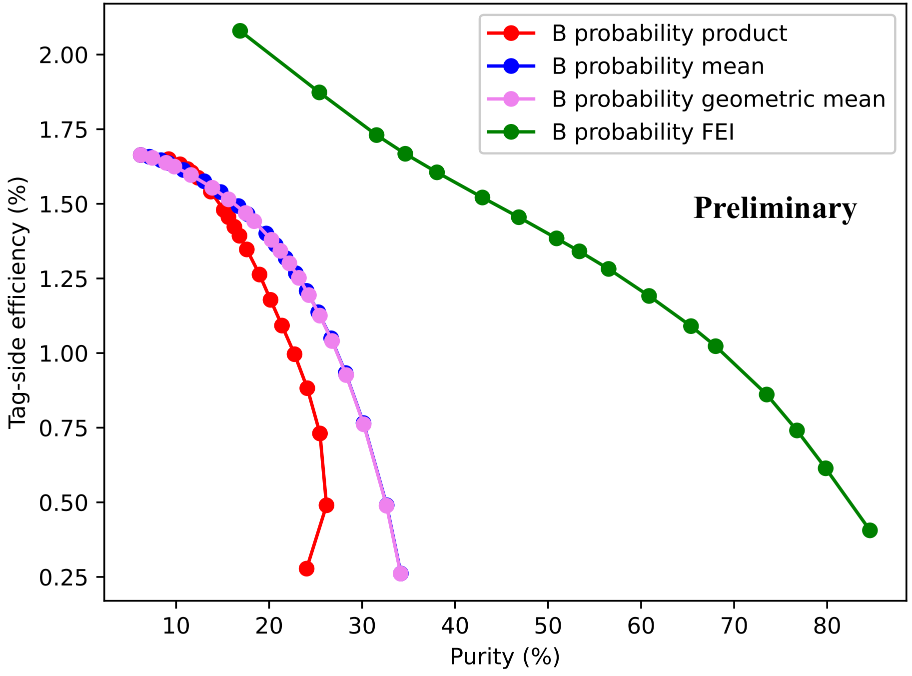
- *graFEI does not make predictions on masses of final state particles, so truth-matching criteria are relaxed accordingly

GraFEI on Belle II simulated dataset – Comparison with FEI

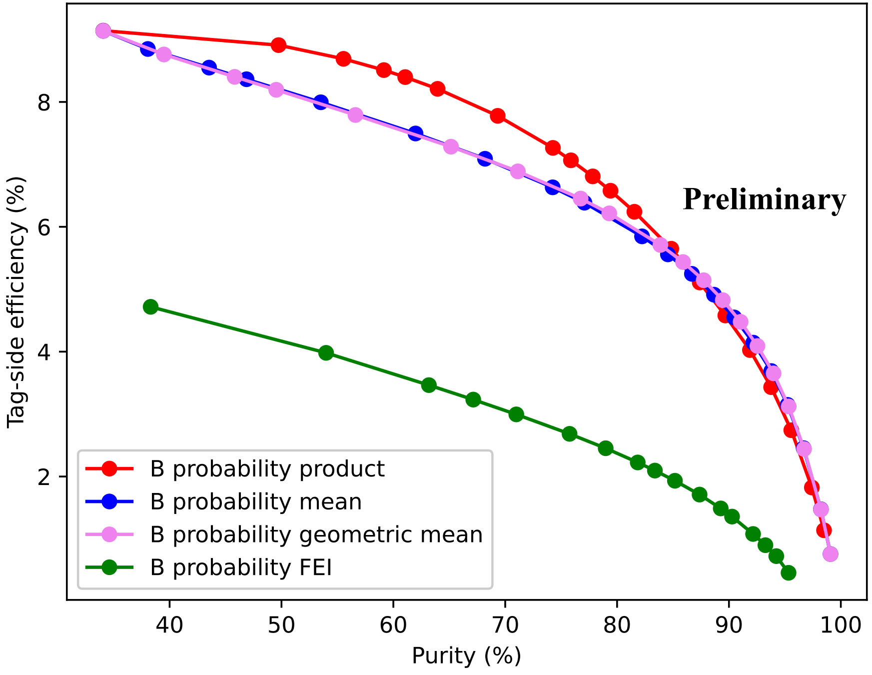
- First evaluation of the **performances on** $\Upsilon(4S) \rightarrow B^0(\rightarrow \nu\nu)\bar{B}^0(\rightarrow \mathrm{X})$ **MC signal**
- Once tag side is reconstructed, events with no extra tracks are selected

Tag-side efficiency = $N_{correct}$ / $N_{\Upsilon(4S)}$

Purity = $N_{correct}$ / N_{tag}



FEI outperforms graFEI when **all the final state particles are reconstructed**



graFEI outperforms FEI when **some final state particles might be missing**

Future prospects

- Short-term plans:
  - **Investigate performances including background MC**
  - **Make predictions on B momentum:** $\mathrm{Loss} = -\frac{1}{N}\sum_i \log \frac{e^{x_{i,y_n}}}{\sum_c e^{x_{i,y_c}}} + \alpha||\vec{p} - \vec{p}_{\mathrm{true}}||^2$
  - **Make prediction on masses of final state particles?**

- Longer-term plans:
  - Optimization of various aspects of the algorithm (input variables, hyper-parameters, architecture, …)
  - Full reconstruction of Υ(4S) decay?
  - ...

Stay tuned!

BACKUP

Hyper-parameters

- NRI:
  - Droput rate = 0.3
  - Batch size = 128
  - Learning rate = 0.001
  - Feed-forward width = 2048
  - Number hidden layers = 1
  - Additional initial/final layers = 0
  - Number of blocks = 1

- graFEI:
  - Droput rate = 0.3
  - Batch size = 128
  - Learning rate = 0.001
  - Feed-forward width = 512
  - Number hidden layers = 1
  - Number of ML = 1

GraFEI on Belle II simulated dataset – Lost particles