Graph-based Full Event Interpretation: a graph neural network for event reconstruction in Belle II

GraFEI - CHEP 2024

Merna AbuMusabh, Jacopo Cerasoli, Giulio Dujany, Corentin Santos

Institut Pluridisciplinaire Hubert Curien
University of Strasbourg

22 October 2024

Table of Contents

1. Experimental context

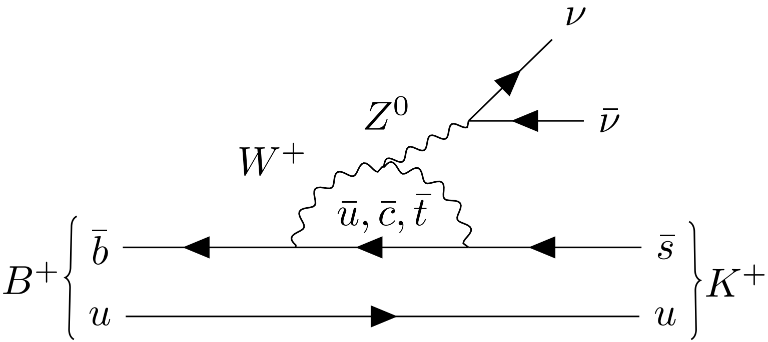
2. Graph-based Full Event Interpretation: the GRAFEI

3. Feasibility study with the search for $B^+ \to K^+\nu\bar{\nu}$ at Belle II

4. GRAFEI future

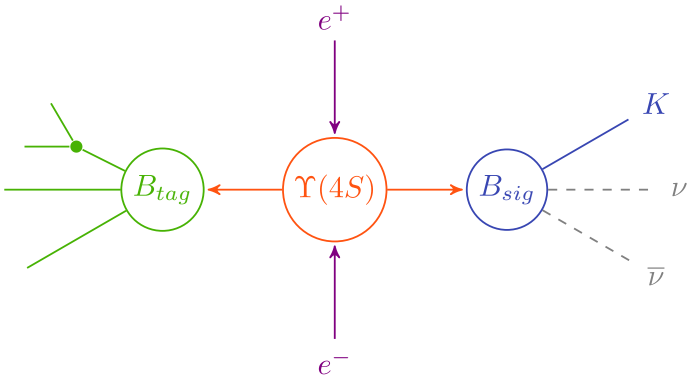
Motivations: B-meson decays with invisible particles

- **New physics** particles may appear in the loop and modify probability
- Possible significant discrepancy from the standard model prediction
- May have presence of new physics invisible particles in place of ν



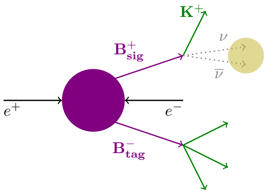
Motivations: B-meson decays with invisible particles

- Tag side needed to infer information about the signal side (*e.g.* $B \to K \nu \bar{\nu}$)

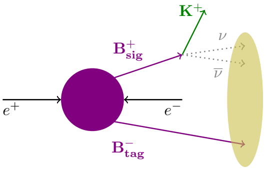


Tagged Anaysis strategies

Hadronic Tagged Analysis (HTA)



Inclusive Tagged Analysis (ITA)



Tagging efficiency

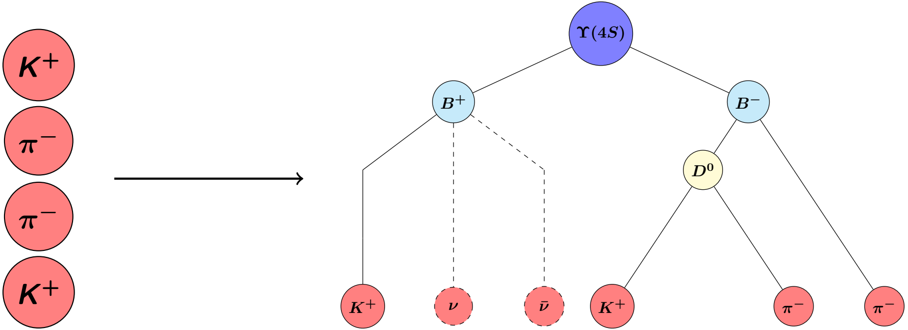
$\mathcal{O}\,(0.1\%)$ $\mathcal{O}\,(10\%)$

Tagging purity

$80\% - 20\%$ $\mathcal{O}\,(1\%)$

Deep learning algorithm: GRAFEI

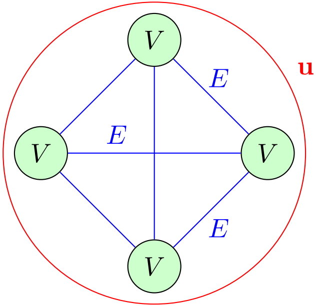
- Based on a deep graph neural network
- Reconstructs the $\Upsilon(4S)$ decay via the final state particles
- Trained over $\Upsilon(4S)$ decays $\longrightarrow$ No need to hard-code decay channels



Overview of the GRAFEI's principles

- Takes as input a fully connected graph representing final state particles

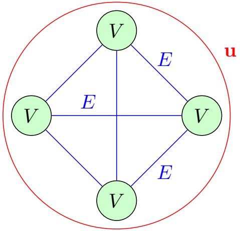
  Nodes = quantities associated to the particles (p_T, p_z, $\mathrm{d}r$, $\mathrm{d}z$, charge, ...)
  Edges = relations between the particles (angle between tracks, distance of closest approach)
  Global = the whole event

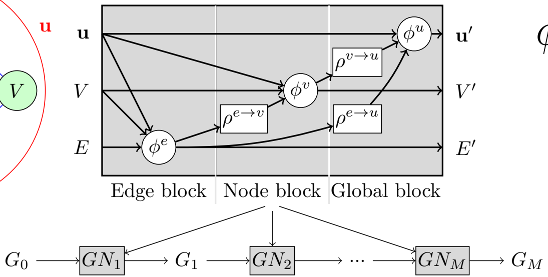


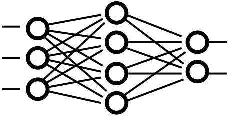
Overview of the GRAFEI's principles

- Takes as input a fully connected graph representing final state particles
- Uses update functions, ϕ, and aggregation functions, ρ

ϕ = Multilayer perceptrons, update the features of the graph per block
ρ = Compute average of the updated features, pass it to the next block





$\phi =$ 

$$\rho = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Overview of the GRAFEI's principles

- Takes as input a fully connected graph representing final state particles
- Uses update functions, ϕ, and aggregation functions, ρ
- Returns a fully connected graph from which we can get mass hypotheses and **Lowest Common Ancestor** (LCAS) matrix elements

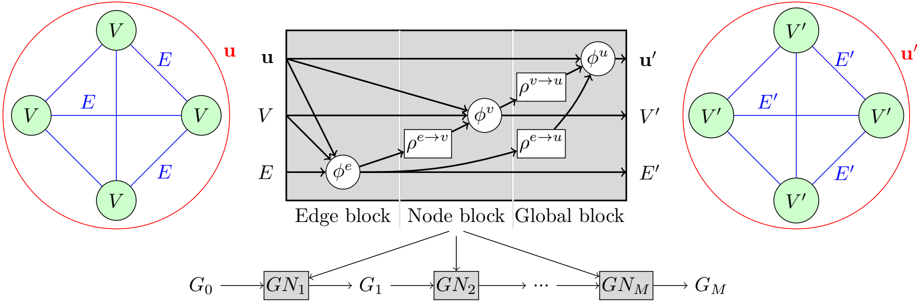
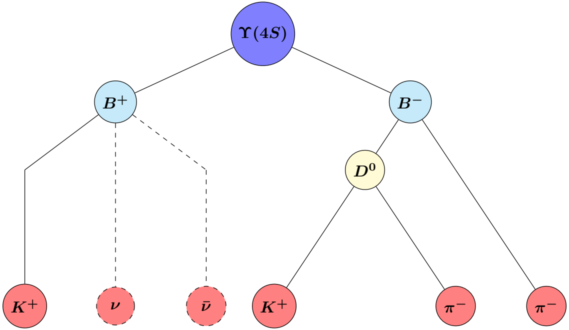


Figure: inspired from arXiv:1806.01261

Lowest Common Ancestor Matrix

- **LCAS** matrix = representation of a decay tree[2,3]
- Rows and columns = FSPs
- Elements = **Lowest Common Ancestor** (LCA)
- Identify them via a **class** between 0 and 6

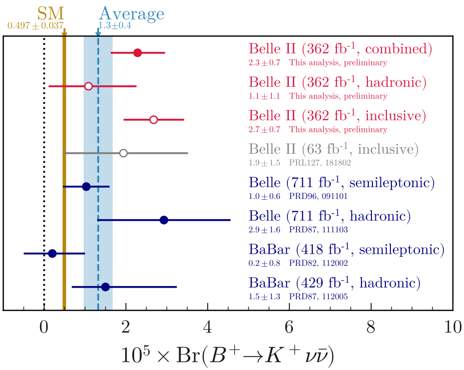


	K^+	π^-	π^-	K^+
K^+		3	5	6
π^-	3		5	6
π^-	5	5		6
K^+	6	6	6	

Implemetation in software framework

- Use PyTorch Geometric for architecture[4]
- Used Optuna for hyperparameter optimization[5]
- Public repository available on Github
- Documentation available on basf2 Sphinx

State of the art



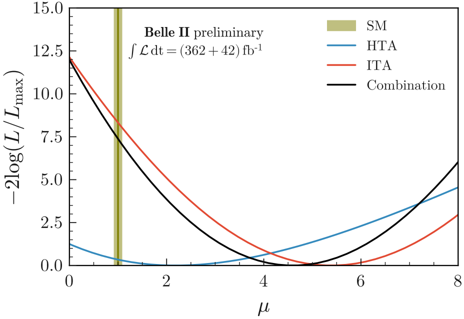


Figure: from arXiv:2301.06990

$$\mu = \frac{\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu} | exp\right)}{\mathcal{B}(B^+ \to K^+\nu\bar{\nu} | SM)}$$

GTA Workflow

Let's define the GRAFEI-based Tagged Anaysis (GTA):

1. Reconstruction + Preselection:
   - LCAS must contain only signal-side Kaon and B_{tag}
   - Cut on GRAFEI *probability*, or *B probability*, BGEOM, *derived from cross-entropy*
   - Apply other *preselection cuts* (more details in the backup)
2. Train classifier
3. Apply *signal region cut*: BDT output > 0.8

GTA Workflow

Let's define the GRAFEI-based Tagged Anaysis (GTA):

1. Reconstruction + Preselection:
   - LCAS must contain only signal-side Kaon and B_{tag}
   - Cut on GRAFEI *probability*, or *B probability*, BGEOM, *derived from cross-entropy*
   - Apply other *preselection cuts* (more details in the backup)
2. Train classifier
3. Apply *signal region cut*: BDT output > 0.8

GTA Workflow

Let's define the GRAFEI-based Tagged Anaysis (GTA):

1. Reconstruction + Preselection:
   - LCAS must contain only signal-side Kaon and B_{tag}
   - Cut on GRAFEI *probability*, or *B probability*, BGEOM, *derived from cross-entropy*
   - Apply other *preselection cuts* (more details in the backup)
2. Train classifier
3. Apply *signal region cut*: BDT output > 0.8

Comparison between analyses

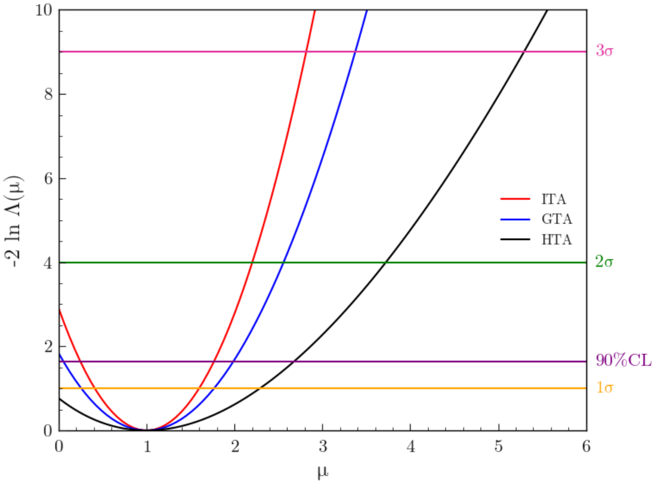
Signal purity:

$$\mathcal{P}^{\text{sig}} = \frac{N_{\text{sig}}}{N_{\text{bkg}} + N_{\text{sig}}}$$

Comparing efficiencies and signal purities:

	$\varepsilon[\%]$	$\mathcal{P}^{\text{sig}}[\%]$
HTA	0.4	3.5
ITA	8	0.8
GTA	**2.7**	**1.3**

Comparison between analyses



N.B.: No systematic uncertainties considered in this study.

Conclusion and perspectives

- GRAFEI is a new approach to inclusive reconstruction
- GTA has a greater *signal purity* than ITA with a higher efficiency than HTA
- Underline ability to classify rare signal events among numerous background events
- Will improve by optimizing GTA and considering statistical and systematic uncertainties

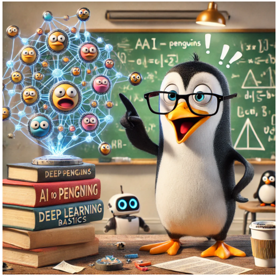
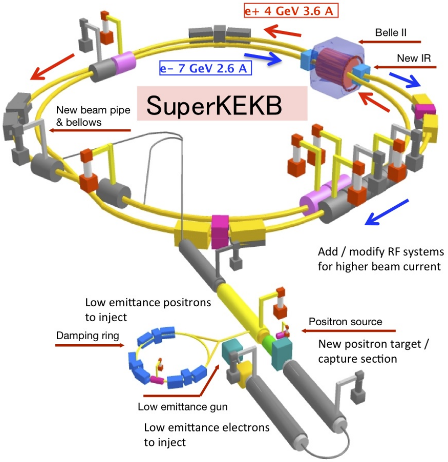
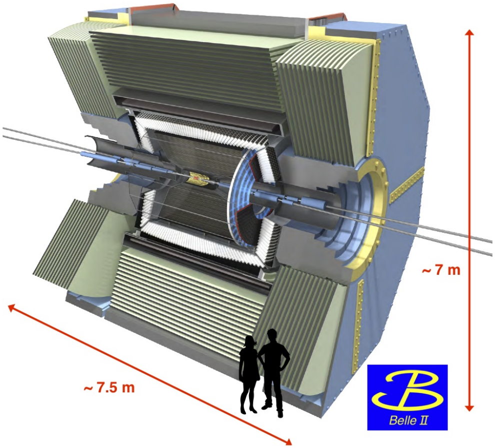


Figure: Generated by Dall-E - "Could you please generate a picture of a penguin manipulating an artificial neural network ?"

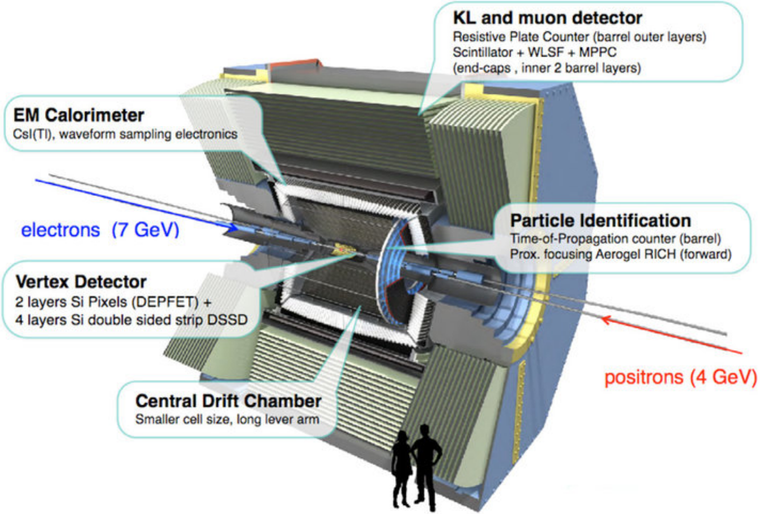
5 Supplementary Material

Belle II experiment



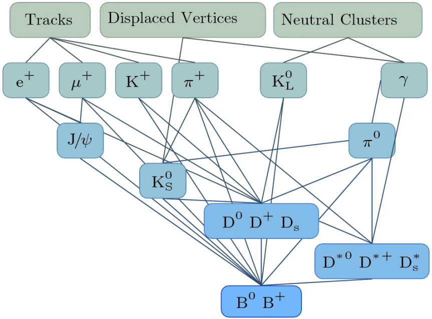


Belle II subdetectors



Full Event Interpretation (FEI) algorithm

- Reconstructs the tag side based on a six stages approach using Boosted Decision Trees[7]
- Need to hard-code decay channels $\longrightarrow$ $\approx 15\%$ of B decays considered



GRAFEI input features

- **Node**: $(\mathcal{L}_{\text{particle}})_{\text{particle}\in\{e,p,\mu,K,\pi\}}$, $p_t.p_z$, $\mathrm{d}r$, $\mathrm{d}z$, CLUSTERNHITS, CLUSTERTIMING, CLUSTERE9E25, charge
- **Edge**: distance of closest approach (doca), cosine of the polar angle $\cos(\theta)$
- **Global**: NONE

GRAFEI hyperparameters optimization

Hyperparameters	Minimum value	Maximum value	Mode
Learning rate	1×10^{-5}	1	LOGUNIFORM
Batch size	32	512	CATEGORICAL
Hidden layer's dimension	64	1024	CATEGORICAL
Number of hidden layers	1	3	INT
Number of GN blocks	1	3	INT
Dropout	0	0.5	UNIFORM

Table: Table summarizing the hyperparameters explored by OPTUNA. "Mode" refers to how the set will be explored: UNIFORM and LOGUNIFORM are self-explanatory since they refer to the distributions of the same name; INT corresponds to a uniform distribution with integer step; CATEGORICAL means that the hyperparameters will be selected in a specified list of values. For "Batch size" and "Hidden layer's dimension", the values can only be powers of 2, meaning that the sets are $[32, 64, 128, 256, 512]$ and $[64, 128, 256, 512, 1024]$. More information about OPTUNA in reference [5].

GRAFEI hyperparameters optimized

Hyperparameters	Value
Learning rate	1×10^{-5}
Batch size	32
Hidden layer's dimension	64
Number of hidden layers	1
Number of GN blocks	1
Dropout	0

GRAFEI probability

$$\textsc{BGeom} = \left[- \prod_{i=1}^{I} \log \left(\frac{exp\left(x_{i,y_i}\right)}{\sum_{c=1}^{C} exp\left(x_{i,c}\right)} \right) \right]^{\frac{1}{I}}$$

where I are the elements of the LCA matrix' lower triangle, C is the number of classes, $(x_{i,c})_{i \in I, c \in [\![1,C]\!]}$ is the model's output for the i-th event and the c-th class, and y_i is the true class of the i-th event.

Pre-training cuts

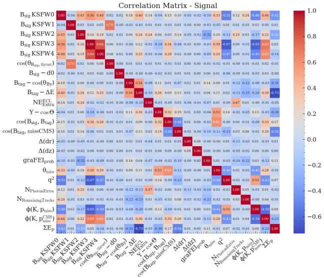
- $\textsc{BGeom} > 0.2$
- $\mathcal{L}_K$ (sig-side FSP) > 0.9
- $\mathit{cos}\left(\theta_{B_{\text{sig, thrust}}, ROE_{\text{thrust}}}\right) > 0.9$

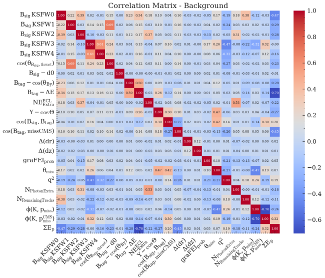
GTA classifier

- Optimized hyperparameters:

BDT parameter	Value
Tree depth	5
Learning rate	0.1547
Number of estimators	600
Subsample Ratio	0.8912

- Input variables:





Workflow performances

Let's define the GRAFEI-based Tagging Anaysis (GTA) and its performances:

Selection	Number of signals after the cut	$\varepsilon_{\textbf{sig}}[\%]$
GRAFEI event reconstruction	26940105	27.295 ± 0.005
Preselection cuts	14429334	53.607 ± 0.010
Signal Region	2671942	18.517 ± 0.010

$\Longrightarrow$ **Total efficiency [%]**: $\varepsilon^{\text{tot}} = 2.709 \pm 0.002$

[1] P. W. Battaglia et al., *Relational inductive biases, deep learning, and graph networks*, (2018), arXiv:1806.01261.

[2] J. Kahn et al., *Learning Tree Structures from Leaves For Particle Decay Reconstruction*, arXiv:2208.14924 [physics] (2022), arXiv:2208.14924.

[3] L. Reuter, *Full Event Interpretation using Graph Neural Networks*, MA thesis (Karlsruhe Institute of Technology (KIT), 2022), https://publish.etp.kit.edu/record/22115.

[4] M. Fey and J. E. Lenssen, *Fast graph representation learning with pytorch geometric*, (2019), arXiv:1903.02428.

[5] *The optuna package, used for hyperparameters optimization*, https://optuna.org/.

[6] Belle II Collaboration, *Evidence for $B^+ \to K^+ \nu\bar{\nu}$ decays*, (2024), arXiv:2301.06990.

[7] T. Keck et al., *The Full Event Interpretation*, Computing and Software for Big Science **3**, 6 (2019), arXiv:1807.08680.