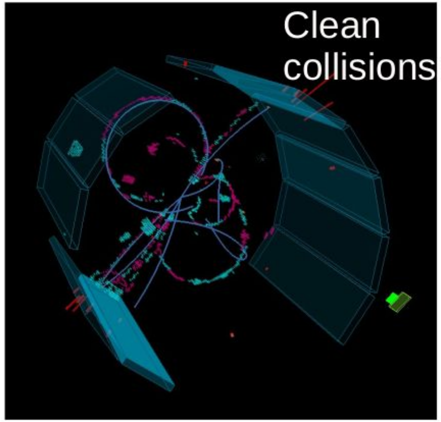
Search for $B^+ \to K^+ \, \nu \, \nu$ at Belle II and B-tagging using Deep Learning

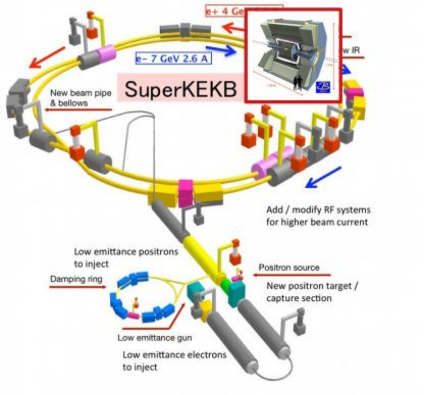
Lucas Martel, Jacopo Cerasoli
CNRS - IPHC

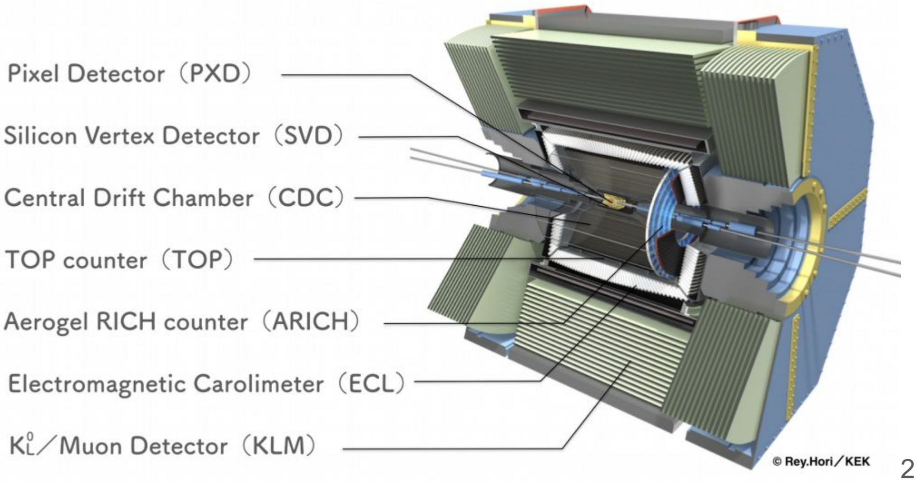
GDR-InF annual meeting, Lyon
03/11/2022

The Belle II experiment

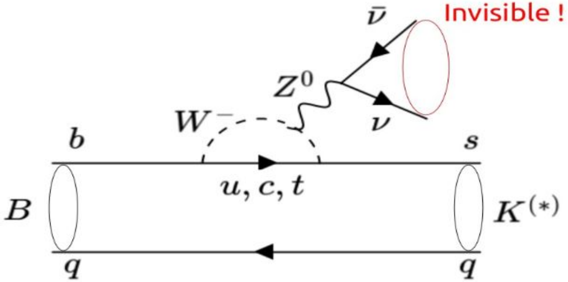
- International collaboration based in Japan.
- Data taking since 2019.
- Asymmetric e^+e^- collider @ 10.58 GeV
- Highest instantaneous luminosity in the world ($> 4.1 \times 10^{34}$ $cm^{-2}s^{-1}$)
- Goal is to reach 50 ab^{-1}
- Strengths: rare and partially invisible decays + precision measurements.







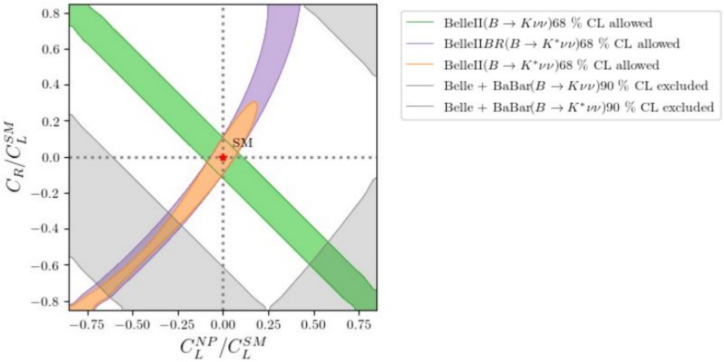
Motivation



- FCNC b →$s\nu\nu$ transition (allowed in the SM, but suppressed).

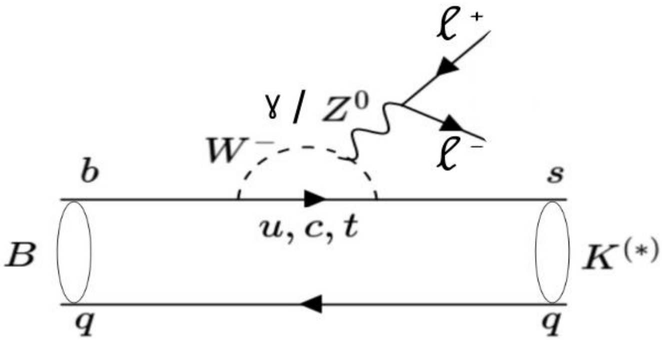
- Good probe for BSM physics:
  - Theoretically clean (no radiative effect from photon wrt b →s l l transitions).
  - Rare ($\mathcal{B}(B^+ \to K^+\nu\nu) \sim 10^{-5}$) but deviation would be a clear sign of BSM physics.

- Interesting:
  - Allows to constrain Wilson coefficients C_L and C_R for effective theories.
  - Input for BSM physics models (Z’, leptoquarks, SUSY).
  - Allows for DM searches (invisible final state).



Constraints on Wilson coefficients with existing measurements and target Belle II measurements at 50 ab^{-1}

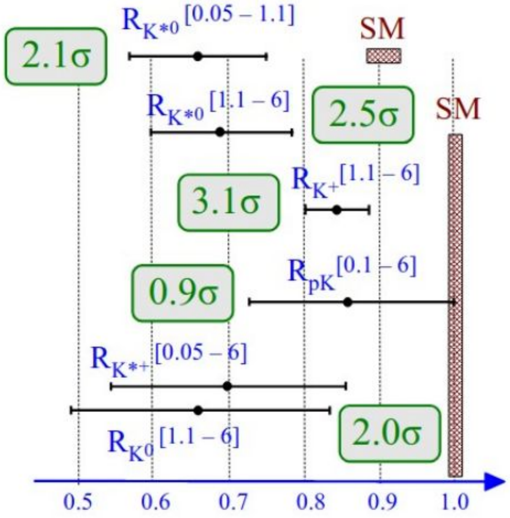
Motivation



- FCNC b →s$\nu\nu$ transition (allowed in the SM, but suppressed).

- Good probe for BSM physics:
  - Theoretically clean (no radiative effect from photon wrt b →s l l transitions).
  - Rare ($\mathcal{B}(B^+ \to K^+\nu\nu) \sim 10^{-5}$) but deviation would be a clear sign of BSM.

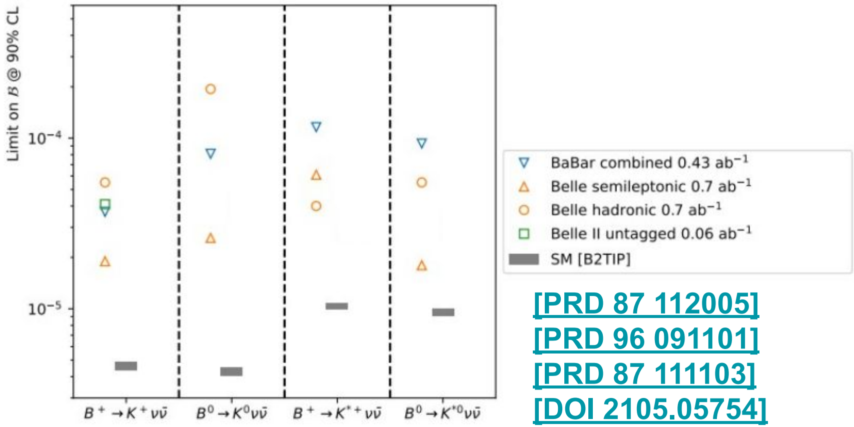
- Interesting:
  - Allows to constrain Wilson coefficients C_L and C_R for effective theories.
  - Input for BSM physics models (Z', leptoquarks, SUSY).
  - Allows for DM searches (invisible final state).
  - Close to b →s l l transitions where tensions with SM are already seen.



[PRL 127 052302]

Experimental challenges

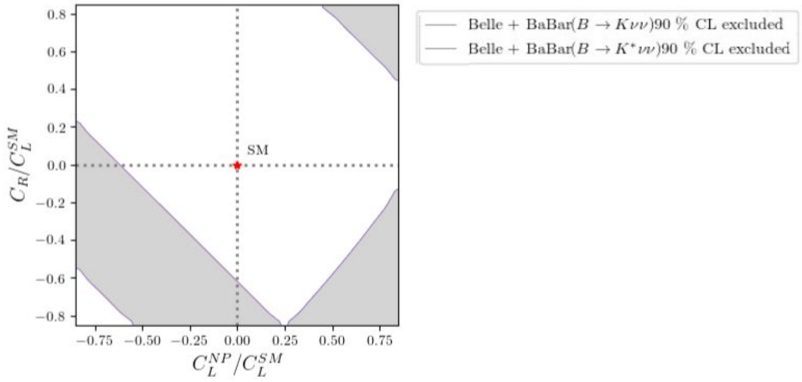
- No observations yet, but upper limits on branching fraction set by Belle, BaBar and Belle II.
- Rare: $\mathcal{B}(B^+ \to K^+ \nu\nu) \sim 10^{-5}$.
- Partially invisible final state.



[PRD 87 112005]
[PRD 96 091101]
[PRD 87 111103]
[DOI 2105.05754]

Belle II is the only experiment able to make a first measurement of this process.

High luminosity, clean environment, good hermiticity of the detector.

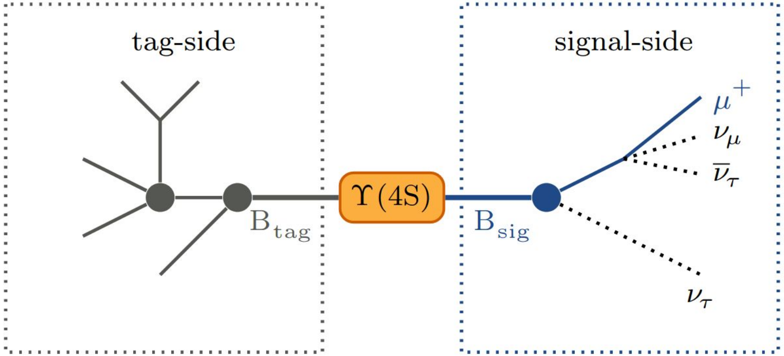


Constraints on Wilson coefficients with existing measurements

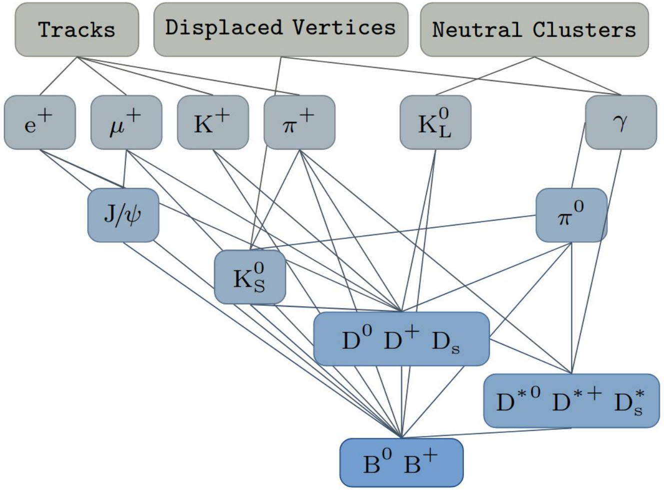
Need for tagging:
Full Event Interpretation

[arXiv:1807.08680]

- Interested in final state with **missing energy**
    - Need to 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 ~ 1%
    - Output of final stage interpreted as “B probability”
    - Decay modes hard-coded, majority of B decays not considered



See talk by Karim!

Analysis overview

Search for $B^+ \to K^+ \nu\nu$ decays using the full pre-shutdown Belle II dataset (400 fb^{-1})

- Analysis strategy:
  - Reconstruct $B^+ \to K^+ \nu\nu$ against hadronic FEI B_{tag}.
  - Train classifier to separate signal from background and define signal region based on BDT output.
  - Binned fit of two components (signal and background).
  - Use profile likelihood to compute branching fraction or set upper limit.

Event selection

- Hadronic FEI Btag selection: Loose selection on tracks and calorimeter clusters + requirement on B_{tag} mass.

- $K^{\pm}$ signal candidates: Must be “good”, i.e come from good tracks and satisfy tight PID requirement.

- $Y(4S) \rightarrow B^{+}{}_{sig}B^{-}{}_{tag}$ reconstructed from K^{+} candidate and B_{tag}.

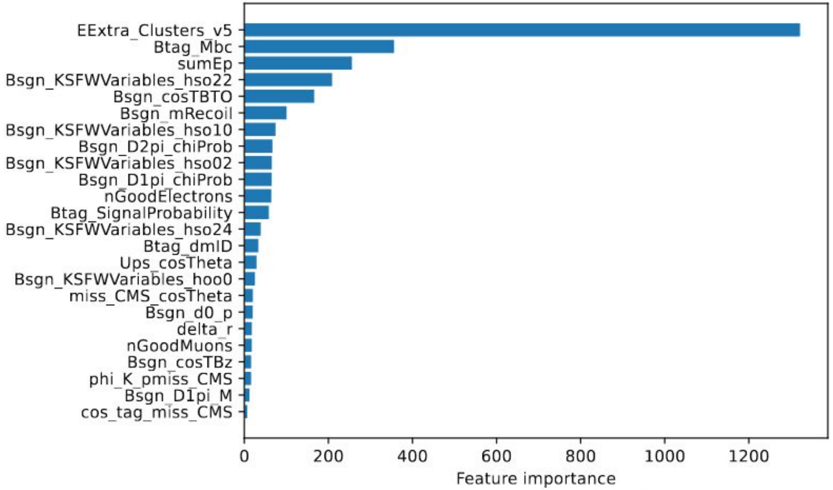
- No **good tracks** and at most one **pi0** in the rest of event.

- Best candidate based on B_{tag} FEI probability.

- Signal efficiency of this selection ~10^{-2}.

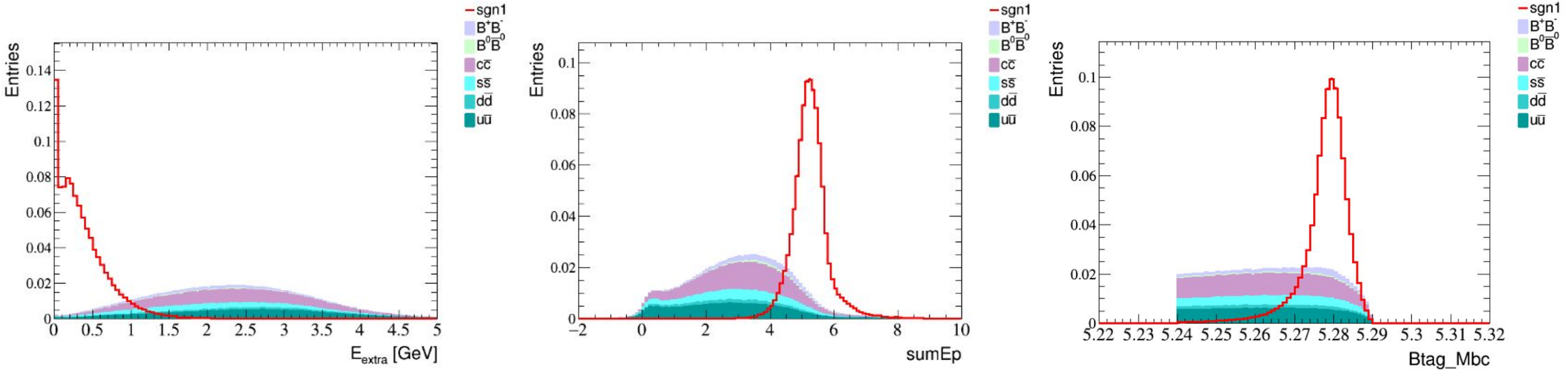
BDT overview

- BDT based on XGBoost trained on 1ab^{-1} of simulated bkg events and 50M simulated signal events



- Variables used in the training:
  - Continuum suppression (KSFW moments, $\cos\theta_{TBTO}$, …)
  - Signal K^+ kinematics (E_K, p_K, …)
  - D meson suppression variables
  - Missing variables (E_{miss}, p_{miss}, …)

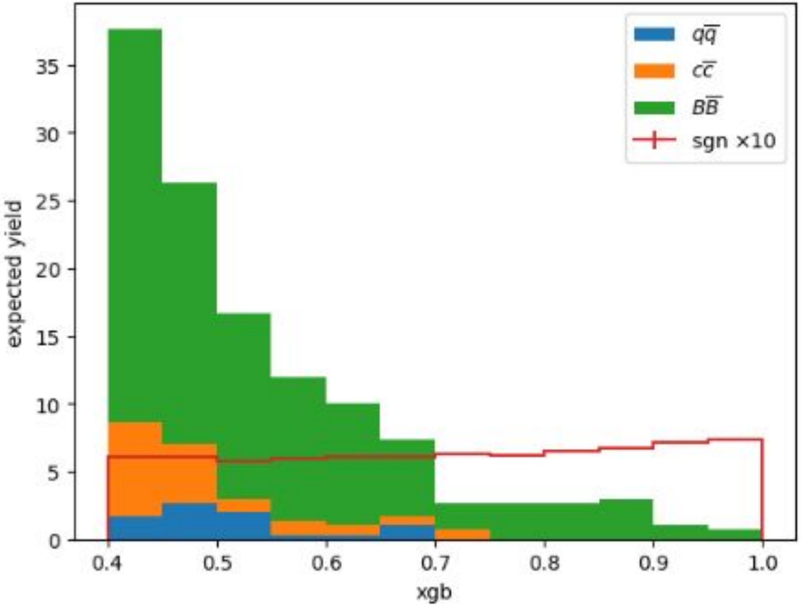
- Pre/post processing so that signal input variables and classifier output uniformly distributed between 0 and 1.

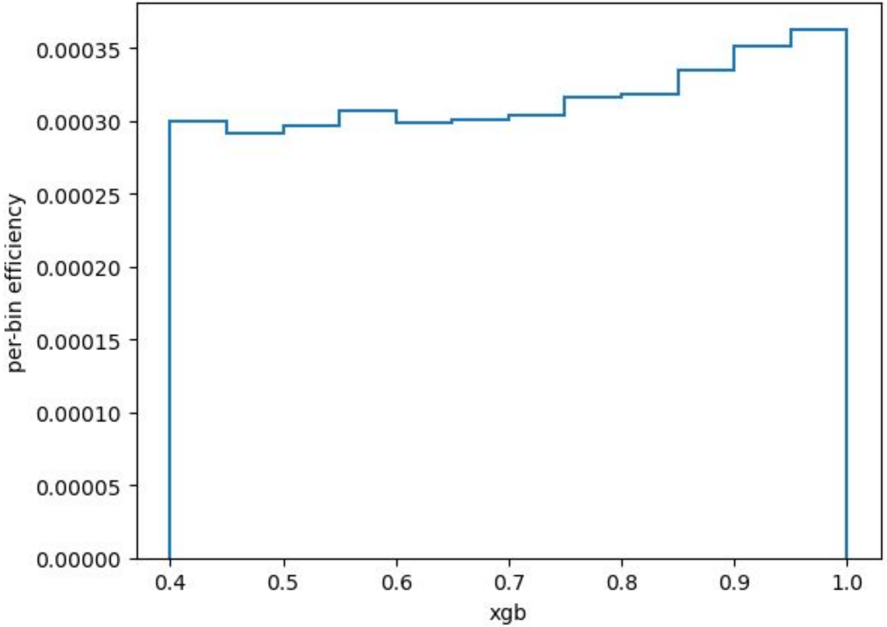


BDT output performance

- BDT trained to separate between signal and background.
- Cut on BDT output > 0.4

Cut	Cumulative sgn efficiency	Bkg yield
Reconstruction + preselection	$(57.60 \pm 0.11) \times 10^{-4}$	219903
BDT cut + single candidate	$(37.98 \pm 0.09) \times 10^{-4}$	172





Computing limit (I)

- We set upper limits on Br($B^+ \to K^+ \boldsymbol{\nu\nu}$) using profile likelihood method:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\nu}})}{L(\hat{\mu}, \hat{\nu})}$$

Constrained best fit

Unconstrained best fit

- We can find 90% CL interval with:

$$-\ln\lambda(\mu) = CDF^{-1}_{\chi^2_1}(0.9)/2 = 1.35$$

- We use as likelihood:

yields

Systs.

$$\prod_{c\in\text{channels}}\prod_{b\in\text{bins}_c}\text{Pois}\left(n_{cb}\middle|\, N_{cb}(\mu,\nu)\right)\prod_{\nu} c_\nu\left(a_\nu\middle|\,\nu\right)$$

Computing limit (II)

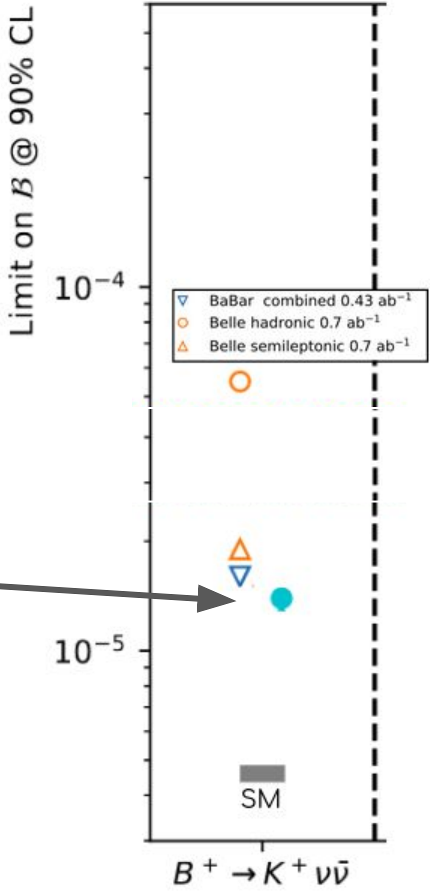
- Preliminary conservative estimate of systematics:
  - 1% on Lσ
  - 10% on selection efficiency
  - 30% on background yield

- We fit signal and background in 12 BDT output bins using pyhf and compute limit using profile likelihood.

- $\boxed{\mathrm{Br}(B^+ \to K^+ \boldsymbol{\nu\nu}) < 1.54 \times 10^{-5}\ @90\%\mathrm{CL}}$

- improvement wrt Belle full reconstruction

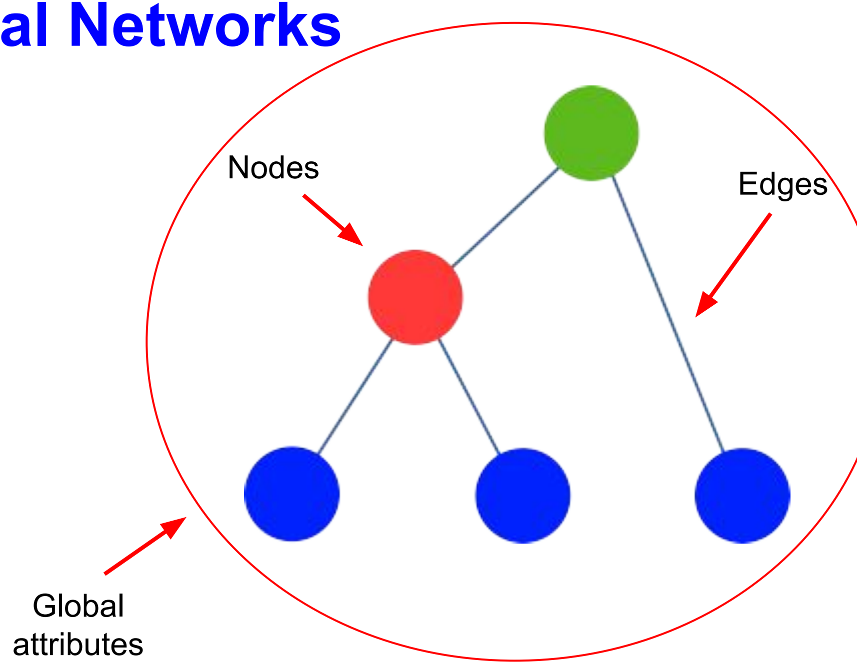
Projection expecting a dataset similar to BaBar (~0.4 ab^{-1})



Towards B tagging using deep learning

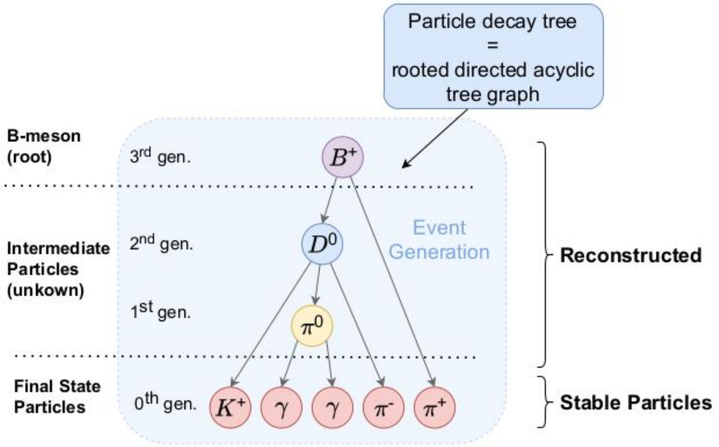
B reconstruction using Graph Neural Networks

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

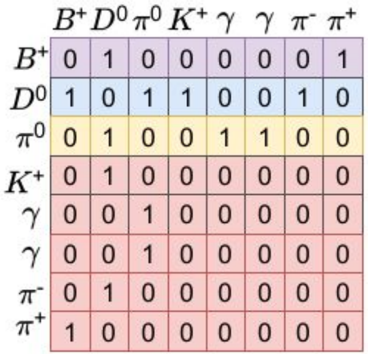


- Proof of concept: *Learning tree structures from leaves for particle decay reconstruction* (see also backup, Ilias Tsaklidis' and Lea Reuter's master theses)
- Today's menu: first (preliminary) **results on Belle II simulated dataset**

Lowest Common Ancestor (LCA) matrix


Particle decay tree = rooted directed acyclic tree graph

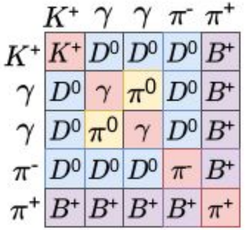
Adjacency Matrix



	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

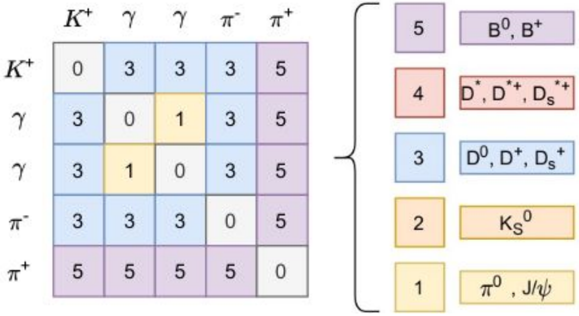
↔

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^+

LCAS

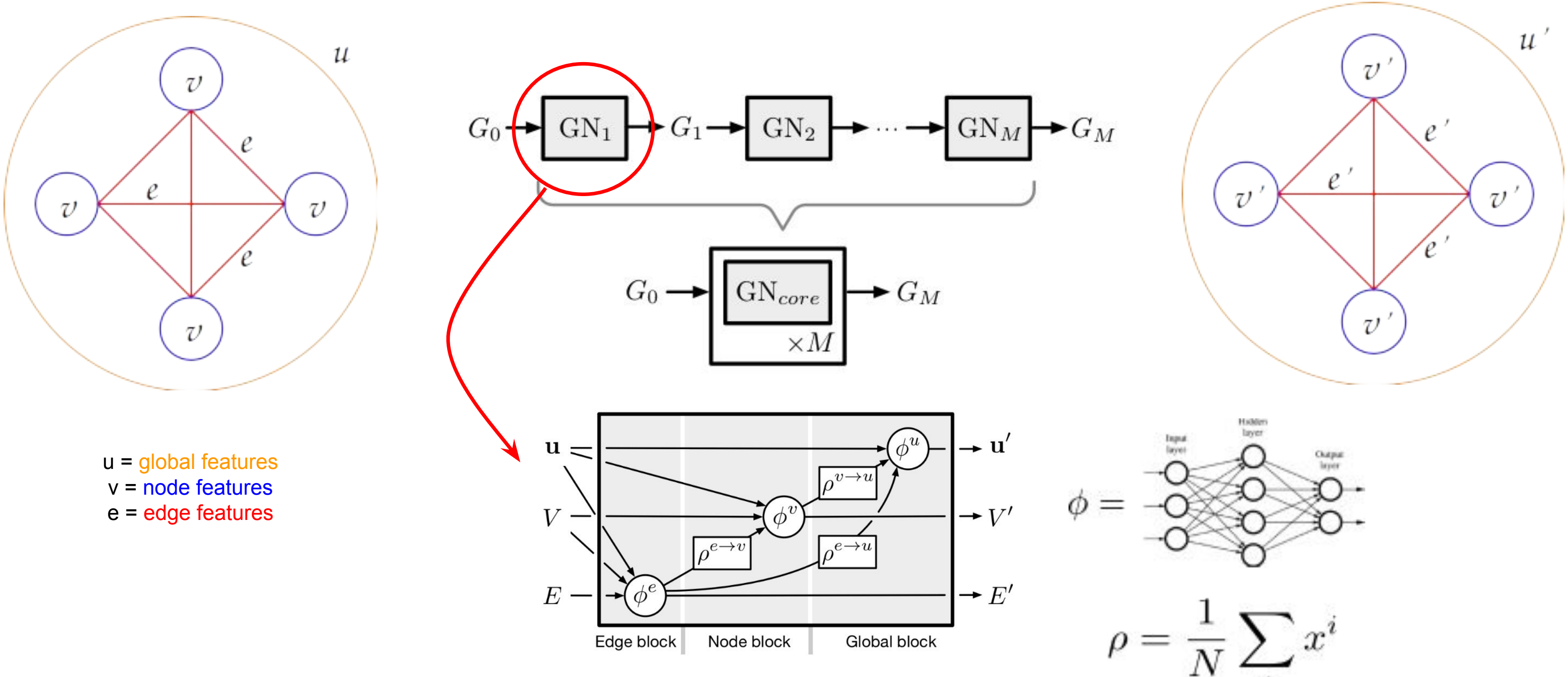


	K^+	γ	γ	π^-	π^+
K^+	0	3	3	3	5
γ	3	0	1	3	5
γ	3	1	0	3	5
π^-	3	3	3	0	5
π^+	5	5	5	5	0

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

graFEI on Belle II simulated dataset

- Model based on graph network blocks
- We input a fully connected graph, output graph has same structure with updated attributes
- Updated edge values used to **predict LCAS matrix**



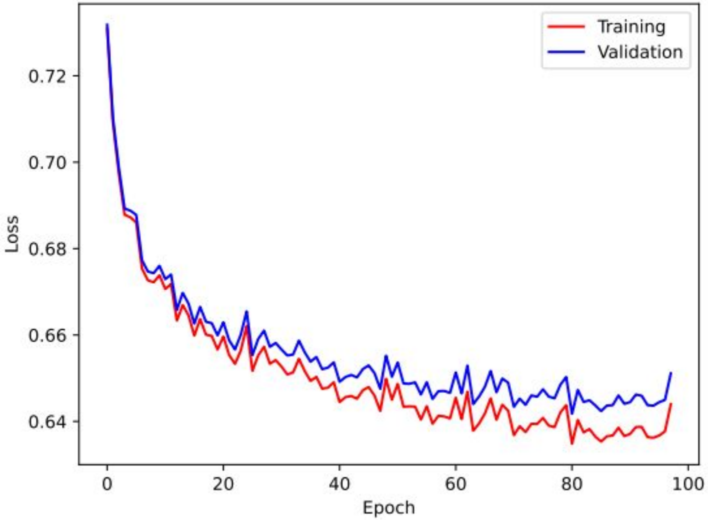
u = global features
v = node features
e = edge features

$$\rho = \frac{1}{N}\sum_i x^i$$

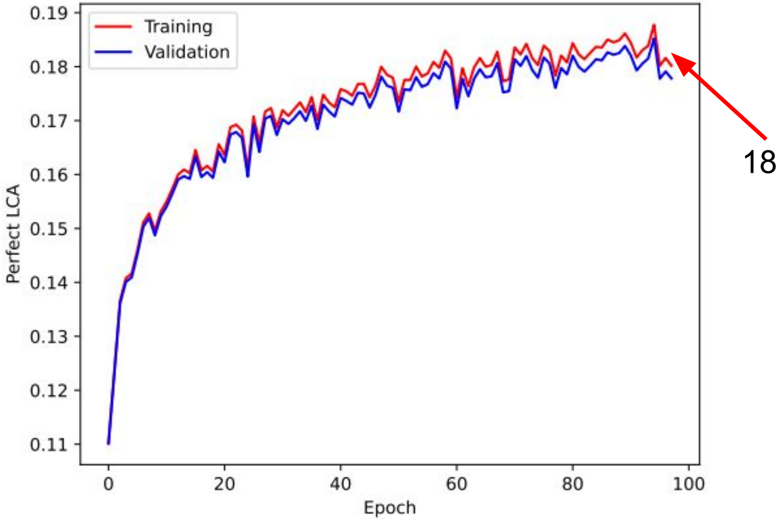
graFEI on Belle II simulated dataset: training

- Training done with **monogeneric $\Upsilon(4S) \to B^0 (\to X)\, B^0 (\to \nu\nu)$ MC** sample
    - Node-level features: particle IDs, 4-momentum, mass hypothesis, charge, impact parameters, ECL cluster variables
    - Edge-level feature: angle between pairs of particles' momenta
    - Global feature: number of final state particles

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



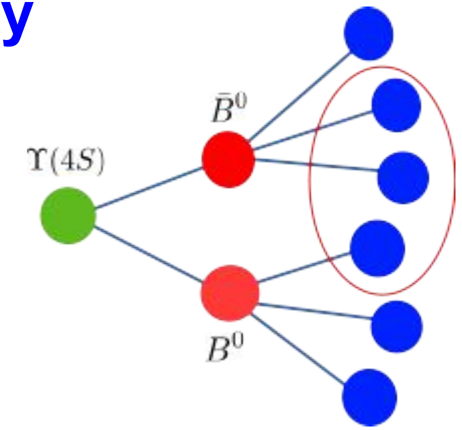
$$\text{Perfect LCA} = \frac{\#\ \text{perfectly predicted LCAS}}{\#\ \text{total predicted LCAS}}$$



18.6 %

graFEI on Belle II simulated dataset: B probability

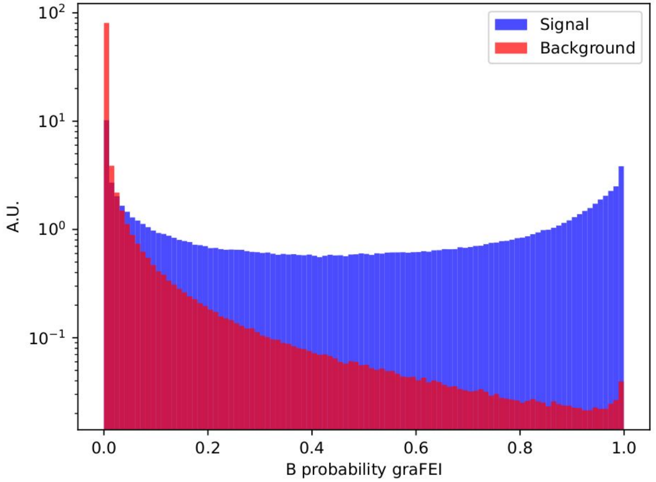
- Signal: monogeneric $Y(4S) \rightarrow B^0\ (\rightarrow X)\ B^0\ (\rightarrow \nu\nu)$ MC sample
- Background: random tracks coming from different B decays



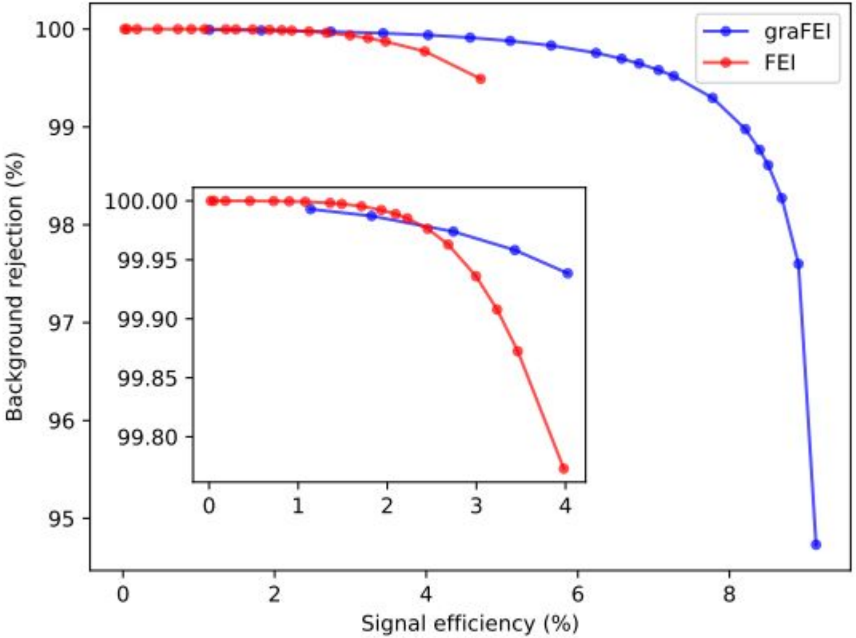
- Having a definition of “**B probability**” analogous to FEI is desirable
  - Each LCA element has a corresponding probability of belonging to the predicted class
  - Product of class probabilities defined as B probability

$$\mathrm{LCA} = \begin{pmatrix} 0 & 3 & 5 \\ 3 & 0 & 5 \\ 5 & 5 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0.62 & 0.31 \\ 0.62 & 0 & 0.76 \\ 0.31 & 0.76 & 0 \end{pmatrix} \rightarrow 0.146$$



graFEI on Belle II simulated dataset: comparison with FEI



- Signal efficiency = # well reco decays / # total decays
- Well reco decays with graFEI means **decays with perfectly reconstructed LCAS**
- graFEI doesn’t make predictions on masses of final state particles (yet) and doesn’t consider intermediate resonances
- To make a fair comparison, same requirements are applied for the FEI (hence higher-than-usual efficiency)

- Performances competitive with FEI
- Factor ~2 more efficiency with higher background
- Algorithm still to be optimized: **room for improvements!**

In summary

- $B^+ \to K^+ \nu\nu$ search using hadronic tagging at Belle II ongoing.
- Hot topic in the wake of tensions seen in b → s l l decays.
- Would allow to provide additional constraints on $\mathcal{B}(B^+ \to K^+ \nu\nu)$ in addition to Belle and BaBar measurements.
- Belle II analysis already published on reduced data sample using inclusive tagging approach →two complementary analyses.
- Analysis on track for all $B \to K^{(*)} \nu\nu$ channels, combination of tagged and untagged measurements on all channels will provide useful inputs for BSM physics models.
- New B-tagging algorithm based on Deep Neural Networks is being developed
  - Early results are promising, stay tuned!

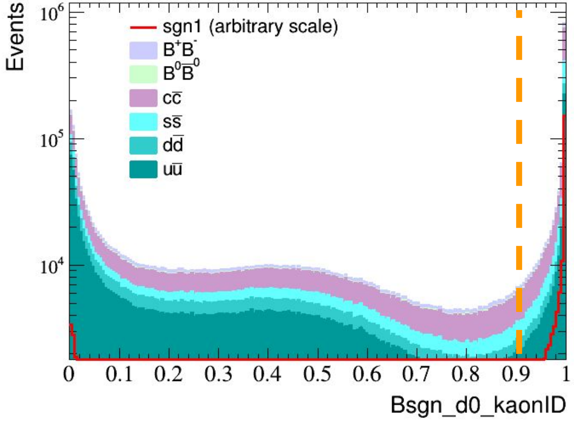
Backup

Additional selection before BDT

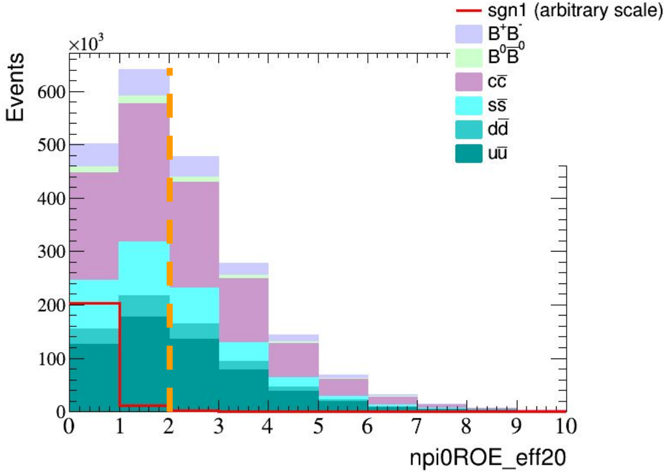
- **KaonID >0.9**
- **npi0 in the ROE of the Y(4S) <2**

Selection stage	efficiency
Skim + reco	0.015
KaonID	0.011
npi0ROE	0.011
best candidate selection	0.006

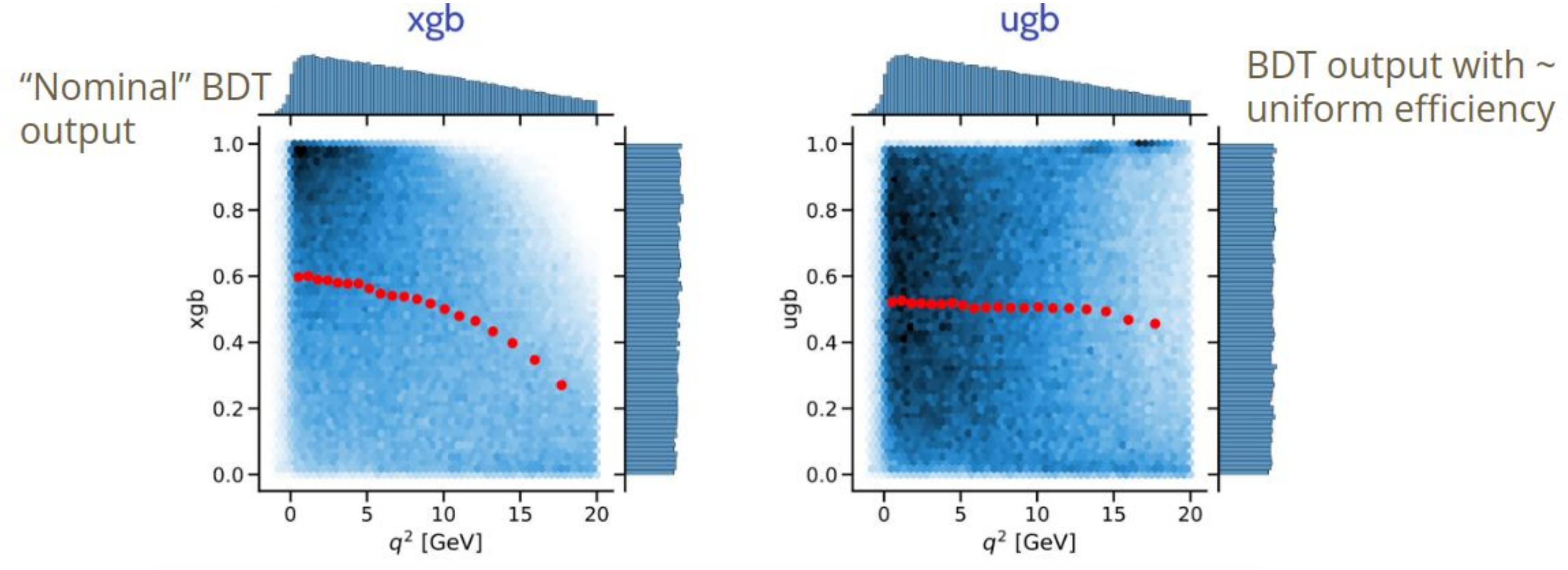
kaonID just after the reconstruction



npi0 after the reconstruction and kaonID cut



Efficiency vs q^2



- Can train XGBoost to flatten efficiency as a function of q^2

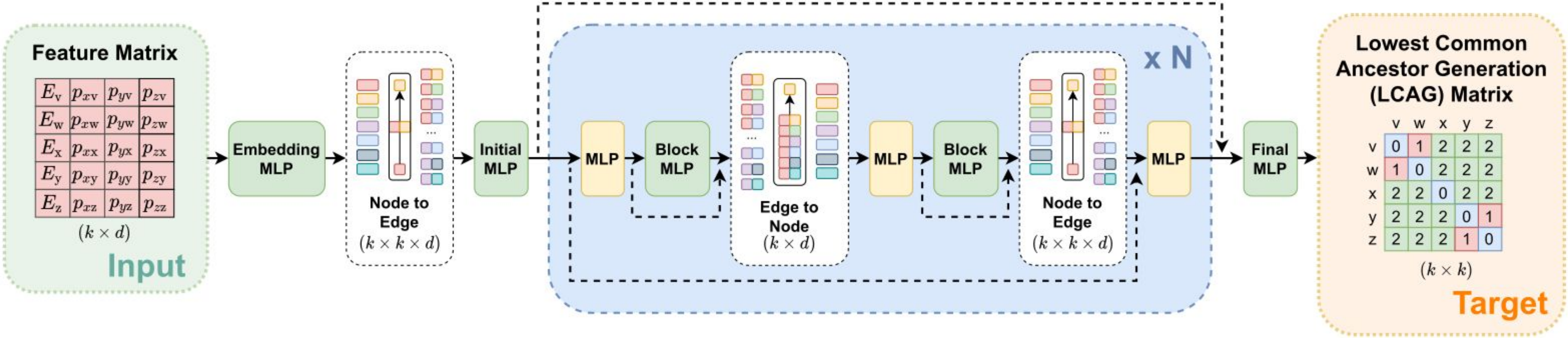
Control samples

- Signal efficiency validation: embedded sample
- Background validation: several control samples to study the data/MC agreement in the BDT input variables and background normalization
    - qqbar background validation: off-resonance data
    - generic background validation: wrong charge sideband, B^+ ->J/ψ K^+ , J/ψ->$\mu\mu$, ee

Embedding method

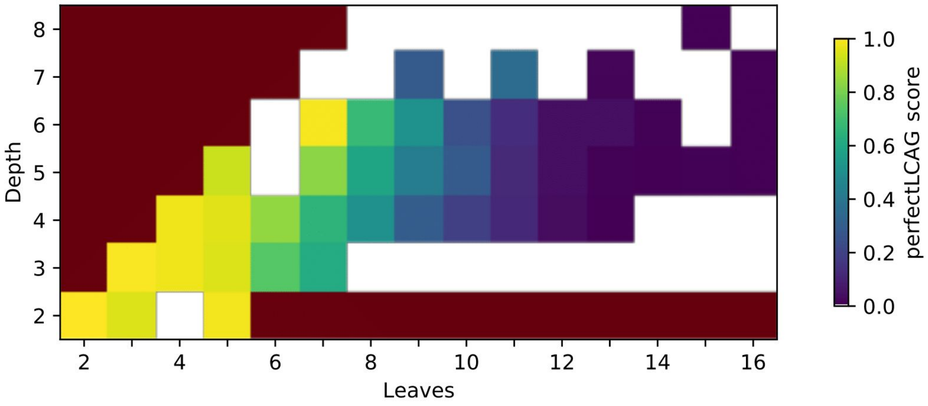
- Identify B decay by a clean hadronic tag
- Remove $B^+ \to K^+ J/\psi$ from the event
- Insert signal $B^+ \to K^+ \nu\nu$ decay instead

graFEI on Phasespace dataset



E_v	p_{xv}	p_{yv}	p_{zv}
E_w	p_{xw}	p_{yw}	p_{zw}
E_x	p_{xx}	p_{yx}	p_{zx}
E_y	p_{xy}	p_{yy}	p_{zy}
E_z	p_{xz}	p_{yz}	p_{zz}

	v	w	x	y	z
v	0	1	2	2	2
w	1	0	2	2	2
x	2	2	0	2	2
y	2	2	2	0	1
z	2	2	2	1	0



- Neural Relational Inference model (NRI)
- Dataset generated with Phasespace library
- 4-momentum used as input
- Average 47.7 % perfectly predicted LCAG matrices

graFEI hyperparameters

- Activation function: elu
- Droput rate = 0.3
- Batch size = 128
- Learning rate = 0.001
- Hidden layer size = 512
- Number hidden layers = 1
- Number of GN blocks = 3 (encoder + intermediate + decoder)