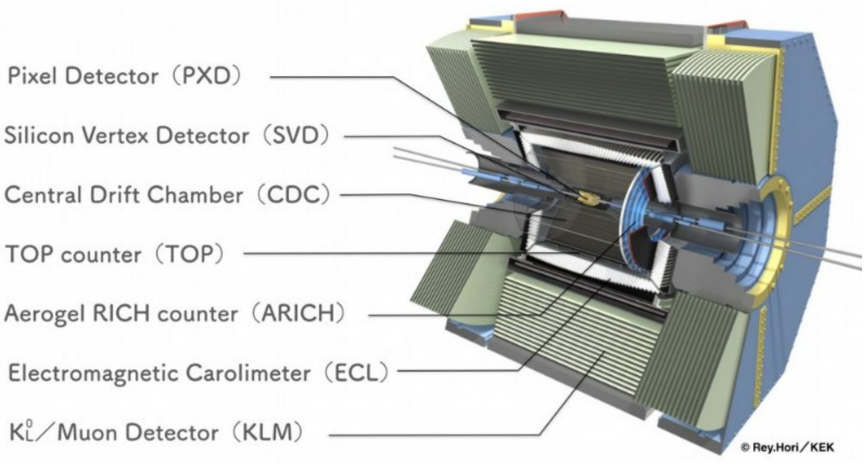
Full Event Interpretation and beyond

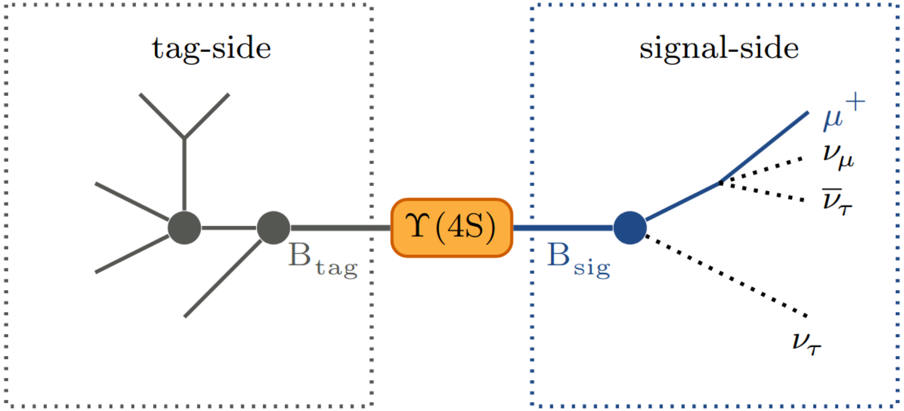
Belle II Summer Workshop
Duke University – 26/07/2023

Jacopo Cerasoli
jacopo.cerasoli@iphc.cnrs.fr

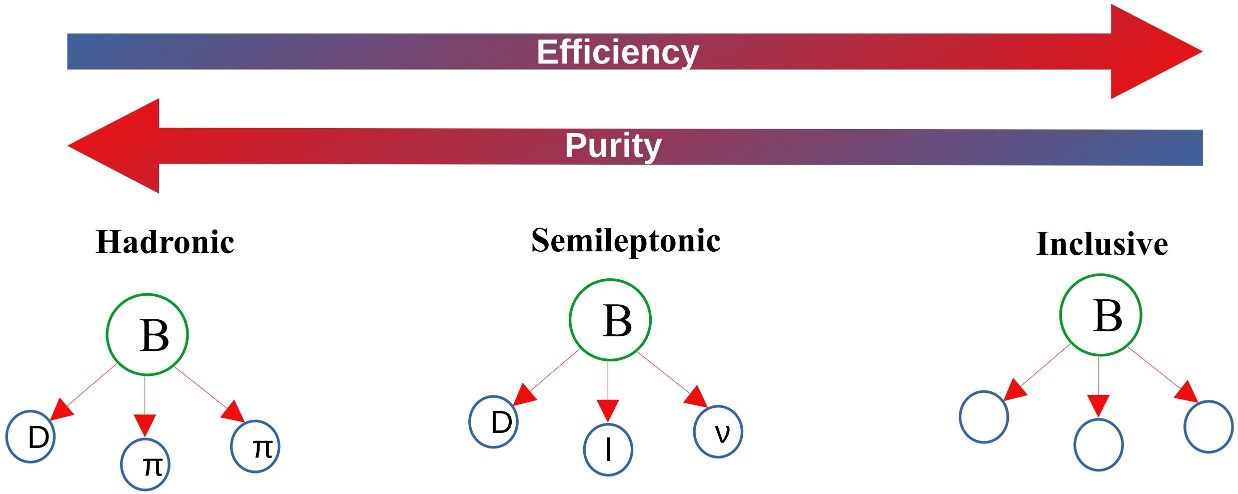
Introduction

- Many Belle II measurements involve processes with missing energy:
  - $B \to K^{(*)} \nu \nu$
  - $B \to D \tau (\to X \nu) \nu$
  - $B \to l \nu \gamma$
  - …
- Quite some unique features at Belle II:
  - Knowledge of **initial 4-momentum**
  - Good **detector hermeticity**
  - **BR(Υ(4S) → *B B*) ~ 100 %**
- → Reconstruct *tag-side B* to constrain kinematics on signal-side





How to reconstruct the B_{tag} ?



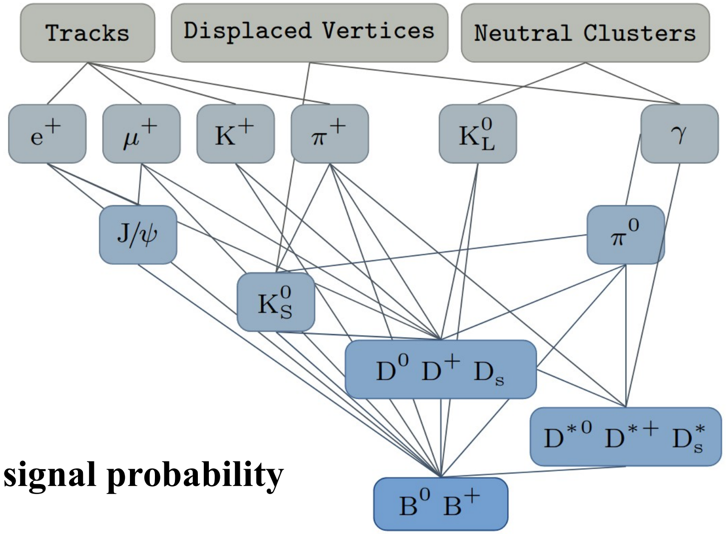
~ 0.5 % efficiency
~ 10 % purity

~ 2 % efficiency
~ 5 % purity

~ 100 % efficiency

- **Full Event Interpretation (FEI)** algorithm for HAD/SL tagging at Belle II:
  - Hierarchical approach based on BDTs
  - Trained on MC $\Upsilon(4S) \to B\,B$ events

Full Event Interpretation [Comput Softw Big Sci 3, 6 (2019)]



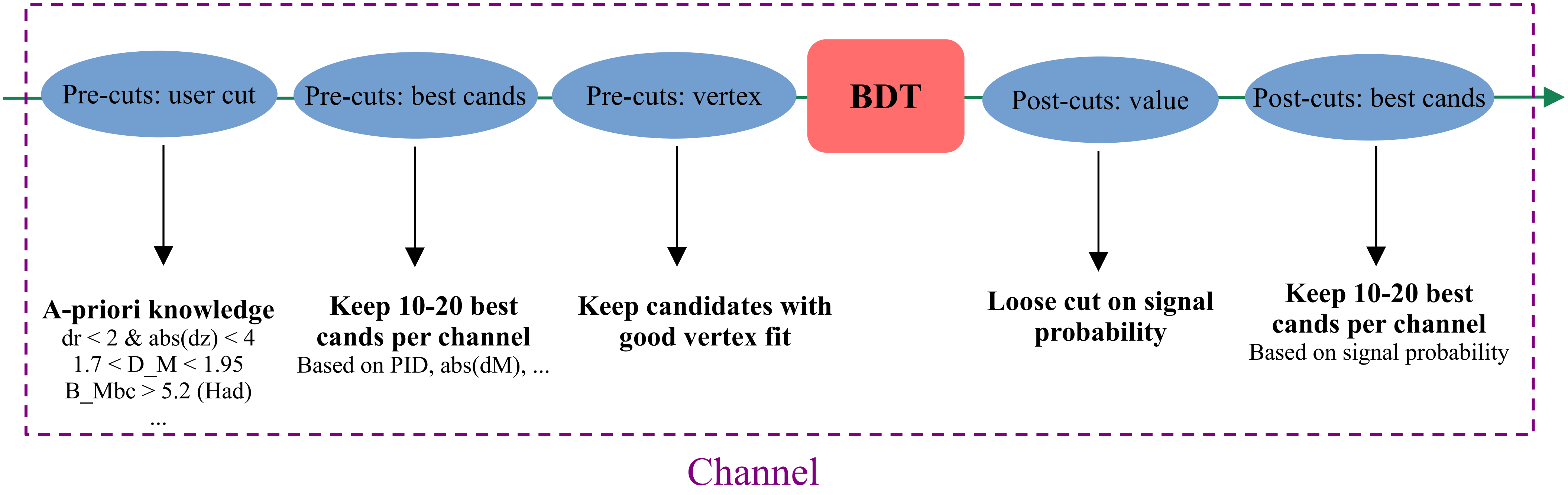
1. Reconstruct final state particles using detector information
2. Combine final state particles into intermediates
3. Combine intermediates and FSPs into *B* candidates

- At each stage a BDT combines the information on candidate into a single number: **signal probability**
- Candidates from different decays are treated equally in following steps
- Signal probability is available to the next BDT

- FEI reconstructs *B* decays in ~ **10k modes**
- Last BDT interpreted as "B probability"
- Overall performances: ~ 1-2 % efficiency at ~ 5-10 % purity

Reducing the combinatoric

- Intermediate cuts applied to reduce combinatoric and save computing time



Training

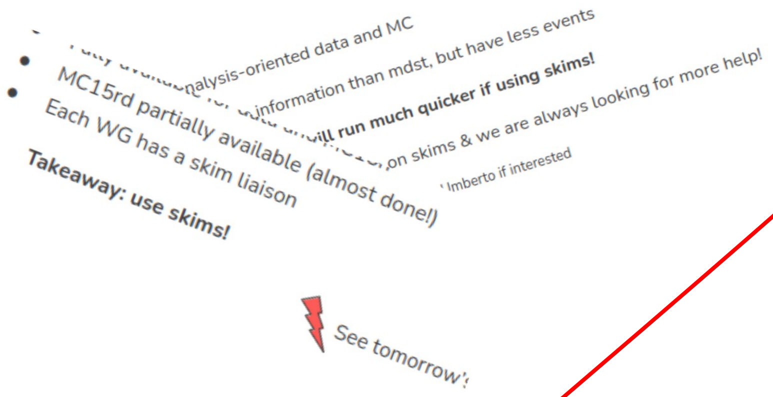
- Each BDT trained to **discriminate signal from incorrect candidates**
- Latest training performed with 200 fb^{-1} run independent MC15 (~ 220 M *BB* pairs)

Input variables

FSPs (charged + photons)	Intermediates	B candidates
• PIDs • p, p_T, p_z • dr, dz • chiProb • Pre-cut ranking based on PID • clusterReg • clusterNHits • clusterTiming • clusterE9E25 • p_T, E, p_z • Pre-cut ranking based on E	• Combination of inv. masses of decay products • Angle between decay products • Momenta of decay products • chiProb of decay products • Signal probability of decay products • Angle between momentum and vertex vector • Vertex χ^2 • Energy released Q, dQ • Mass difference wrt nominal value	• All variables used for intermediates (including signal probability) • ΔE • Flight distance and its significance • dr, dz, dx, dy

Example from basf2 tutorial

Trevor's slides:



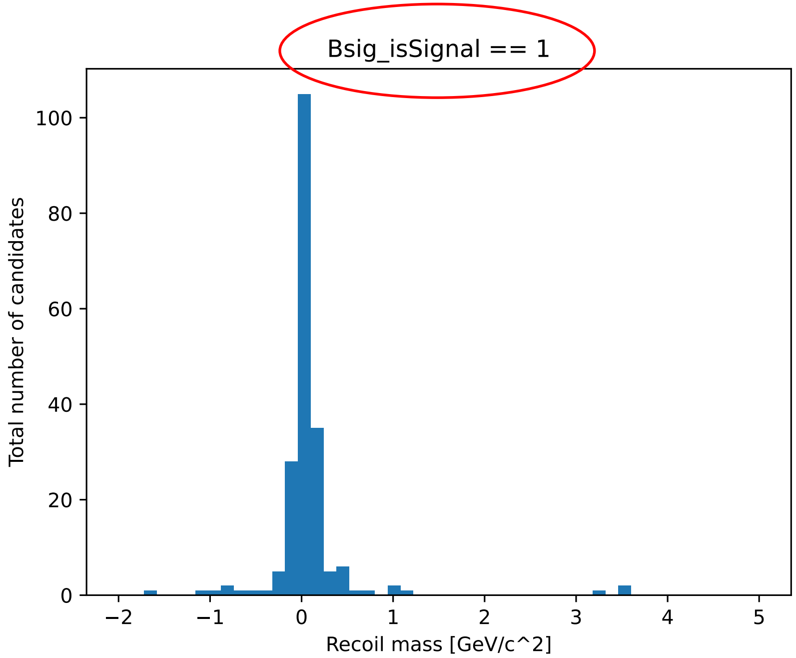
- $B^0 \to \pi^-/\rho^-\ l+\ \nu$ MC with **FEI skim**
- Signal-side $B^0 \to \pi^-\ \mu^+\ \nu$ selection
- Υ(4S) reconstruction
- Rest of event building

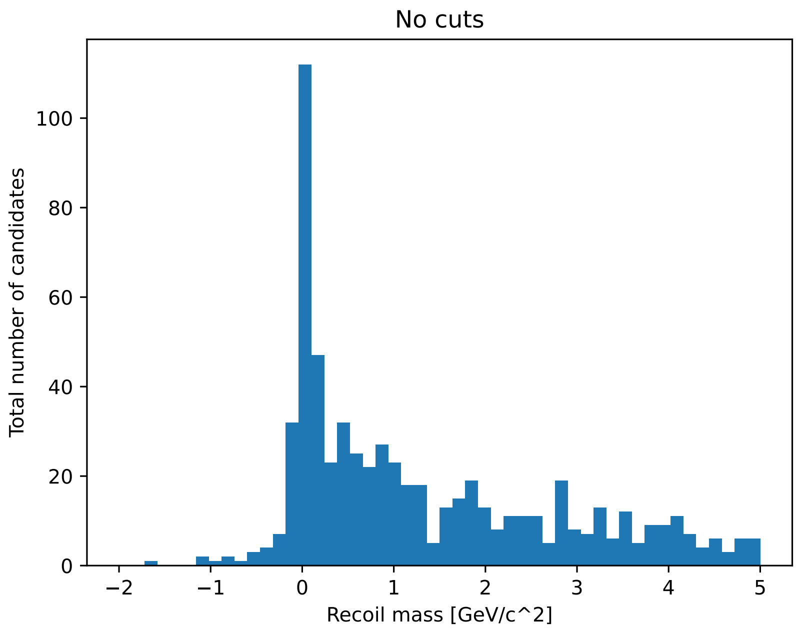
https://software.belle2.org/development/sphinx/online_book/basf2/fei.html

```python
import basf2 as b2
import modularAnalysis as ma
from variables import variables as vm

main = b2.Path()

ma.inputMdst(
    "./fei_skimmed_xulnu.udst.root",
    path=main,
)

good_track = (
    "dr < 0.5 and abs(dz) < 2 and nCDCHits > 20 and thetaInCDCAcceptance"
)

ma.fillParticleList("mu-", "muonID > 0.9 and " + good_track, path=main)
ma.fillParticleList("pi-", "pionID > 0.5 and " + good_track, path=main)

ma.reconstructDecay("B0:signal -> pi- mu+ ?nu", cut="", path=main)

ma.reconstructDecay(
    "Upsilon(4S):opposite_cp -> B0:generic anti-B0:signal", cut="", path=main
)
ma.reconstructDecay(
    decayString="Upsilon(4S):same_cp -> B0:generic B0:signal",
    cut="",
    path=main,
)
# Combine the two Upsilon(4S) lists to one. Note: Duplicates are removed.
ma.copyLists(
    outputListName="Upsilon(4S)",
    inputListNames=["Upsilon(4S):opposite_cp", "Upsilon(4S):same_cp"],
    path=main,
)

ma.buildRestOfEvent("Upsilon(4S)", path=main)
track_based_cuts = "thetaInCDCAcceptance and pt > 0.075 and dr < 2 and abs(dz) < 4"
ecl_based_cuts = "thetaInCDCAcceptance and E > 0.05"
roe_mask = ("my_mask", track_based_cuts, ecl_based_cuts)
ma.appendROEMasks("Upsilon(4S)", [roe_mask], path=main)

ma.matchMCTruth(list_name="Upsilon(4S)", path=main)
```

Example from basf2 tutorial

- We look at the ***B^0* recoil mass**
- Should peak at 0 for signal events (neutrino)
- Broad tail in full sample (background events)





- Let's see if we can do better with the FEI

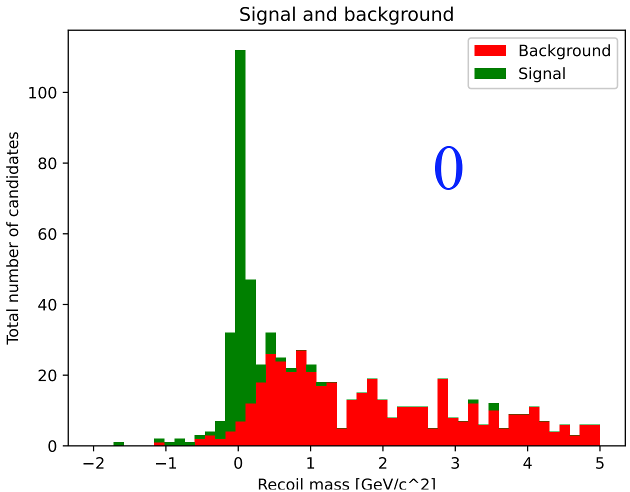
Example from basf2 tutorial

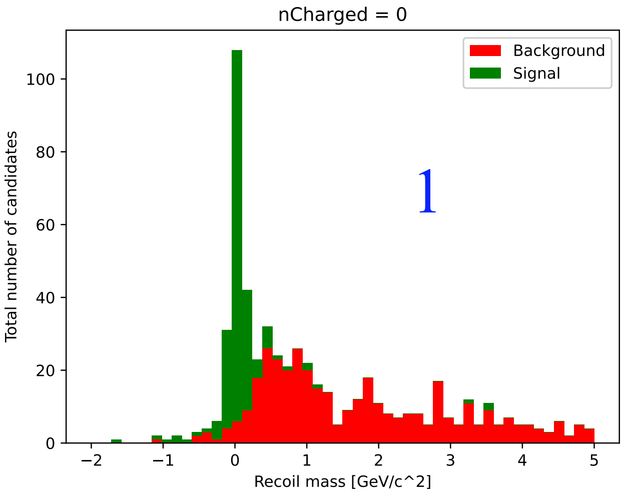
0) Initial dataset

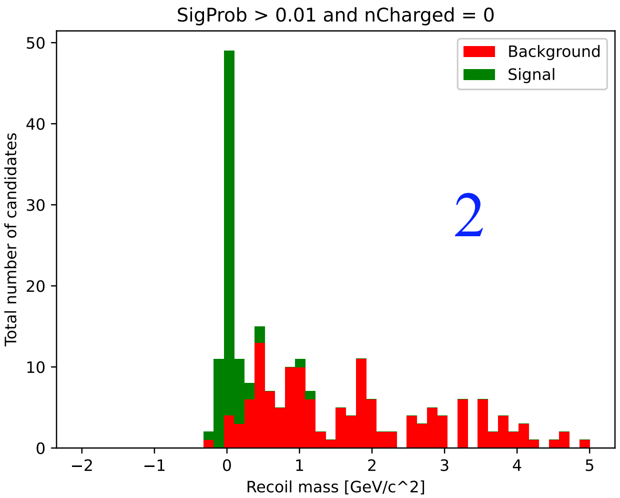
1) We can start requiring 0 charged tracks in the ROE (**completeness constraint**)

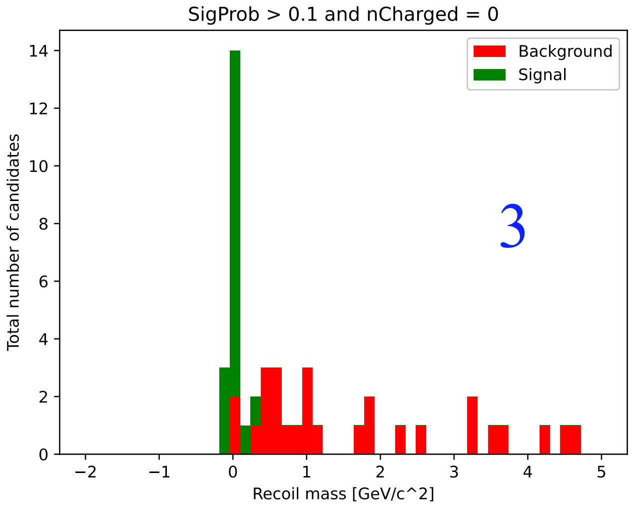
2) Then we cut on signal probability > 0.01

3) Finally on signal probability > 0.1

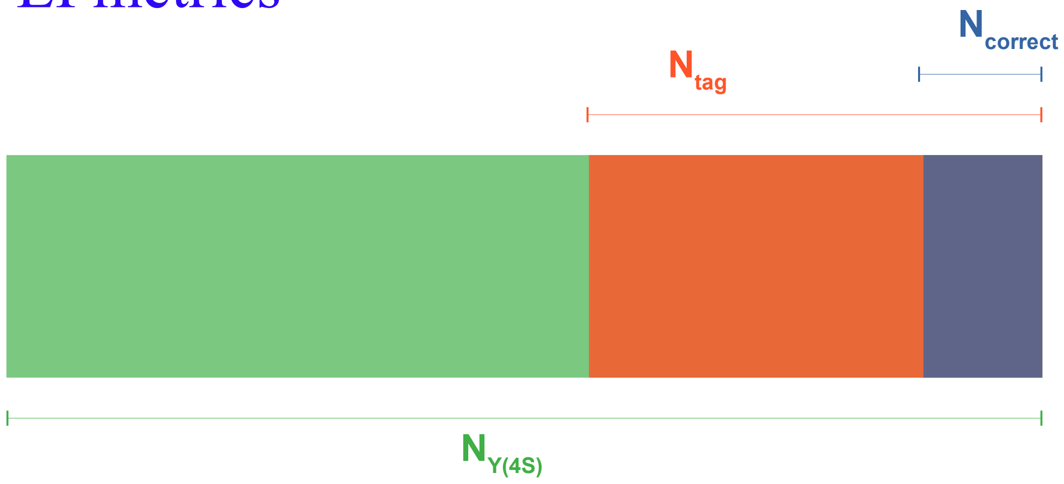








FEI metrics



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

BELLE2-NOTE-PH-2019-031
W. Sutcliffe, F. Bernlochner

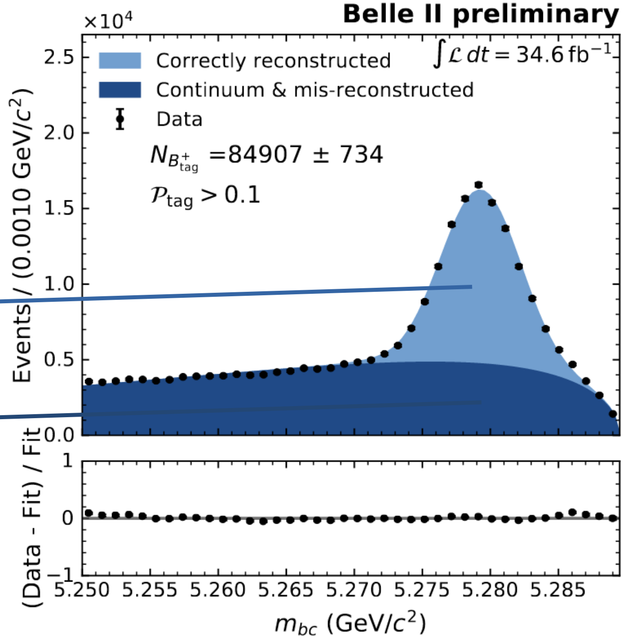
$N_{correct}$

+

$N_{incorrect}$

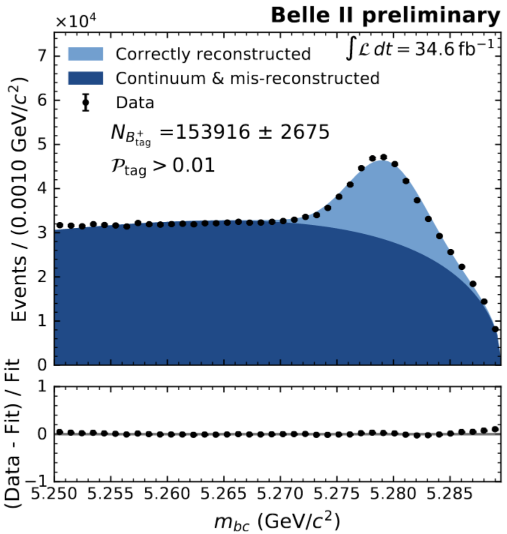
=

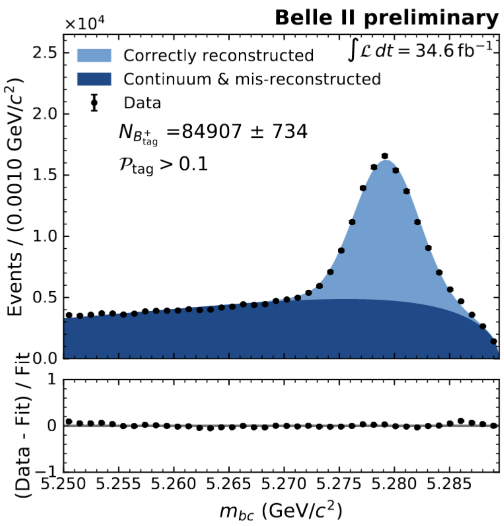
N_{tag}

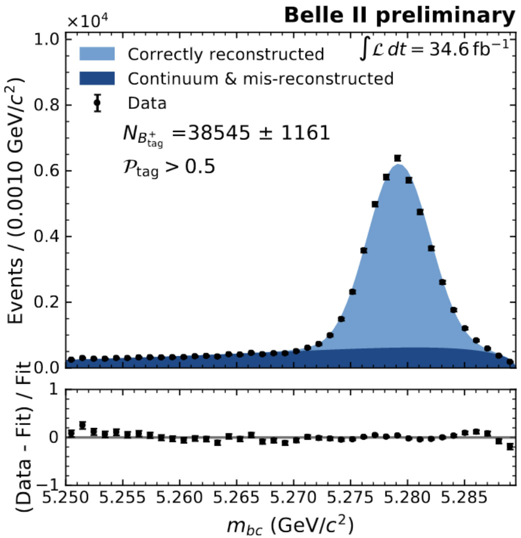


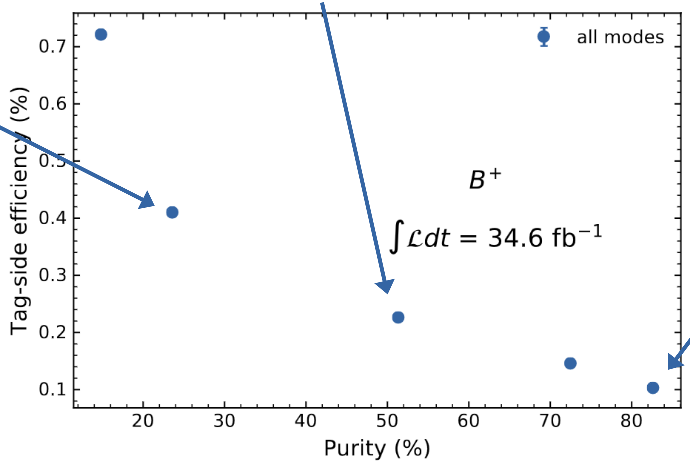
FEI performances (Had)

BELLE2-NOTE-PH-2019-031
W. Sutcliffe, F. Bernlochner









FEI calibration

- Efficiency does not agree between simulation and data because of:
  - **Branching fractions** of *B* and intermediate decays not well known
  - **Data-MC differences** of BDT variables
  - Tag-side particles have **no corrections** applied

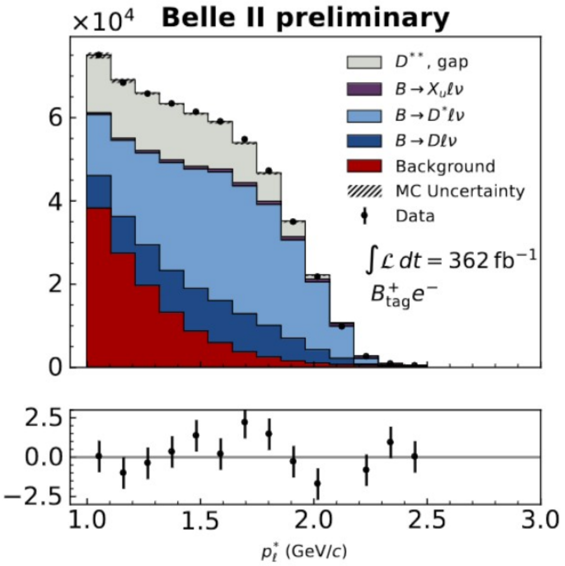
- → Need to perform a *calibration* of the FEI:
  - Measure the **yield of *B* decays in data and MC** and extract a correction factor N_{data}/N_{MC}
  - Perform the calibration for various values of signal probability, tag-side decay mode and possibly other quantities

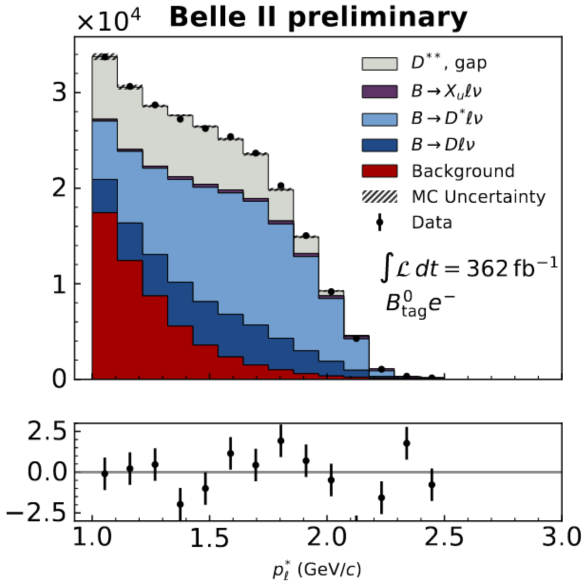
Had FEI calibration with $B \to X\, l\, \nu$ samples

BELLE2-NOTE-PH-2023-008
W. Sutcliffe, F. Bernlochner

- Perform calibration using inclusive $B \to X\, l\, \nu$ decays ($l = e, \mu$)
  - Perform ***B*-tagging using hadronic FEI**
  - **Select lepton** from ROE with $p^*_l > 1$ GeV
  - Perform **binned fit to p^*_l** to get yields in data and MC

B^0 modes
Sig prob > 0.001

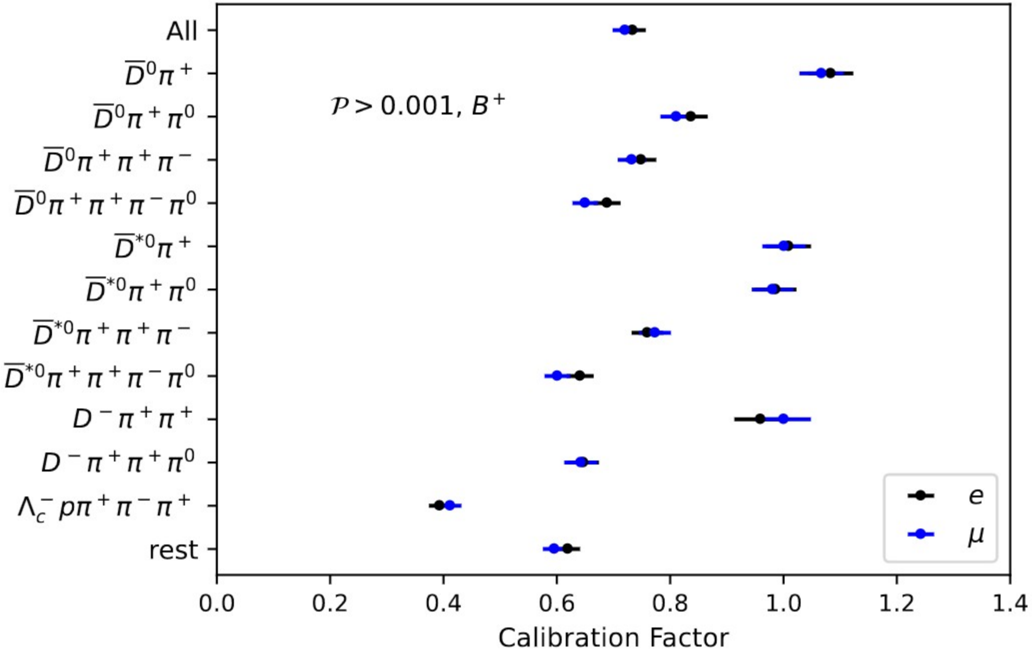


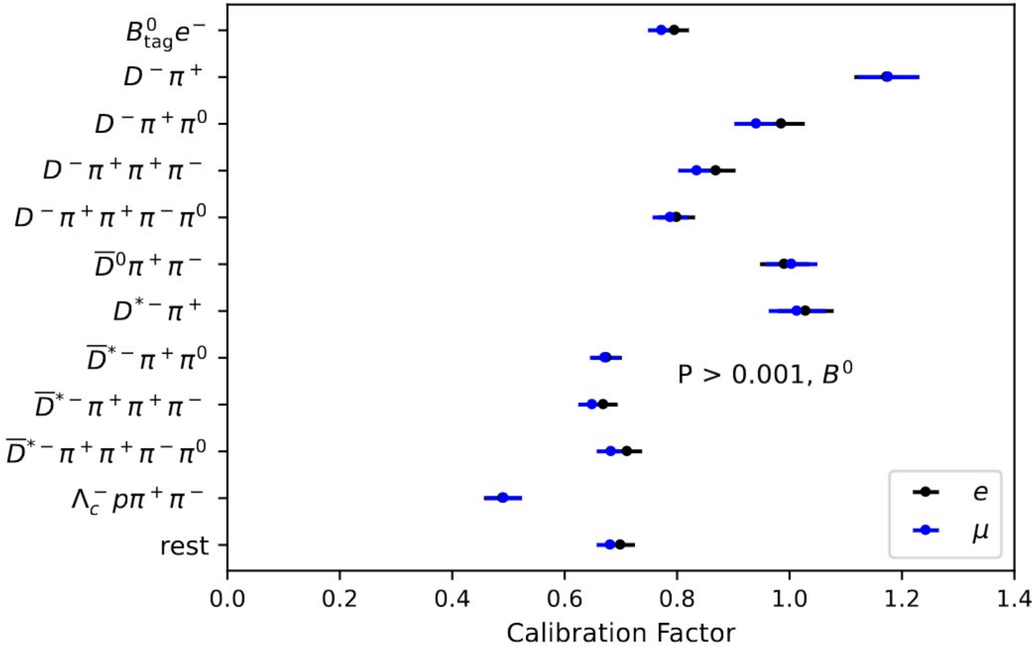


Had FEI calibration with $B \to X\, l\, \nu$ samples

BELLE2-NOTE-PH-2023-008
W. Sutcliffe, F. Bernlochner

- **Mode-by-mode calibration factors for B^+ and B^0 and p > 0.001**, using electron and muon channels
- Systematic and consistent tension between electron and muon channels
  - Still investigating, maybe due to $K \to \mu$ fake rate or background in muon channel



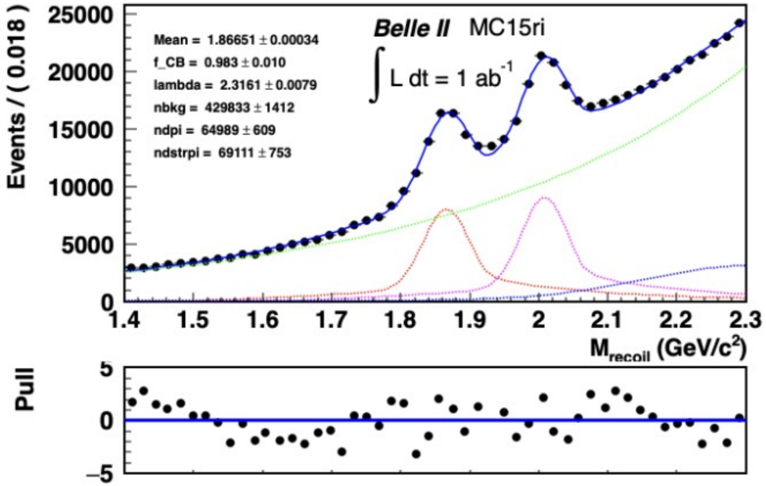


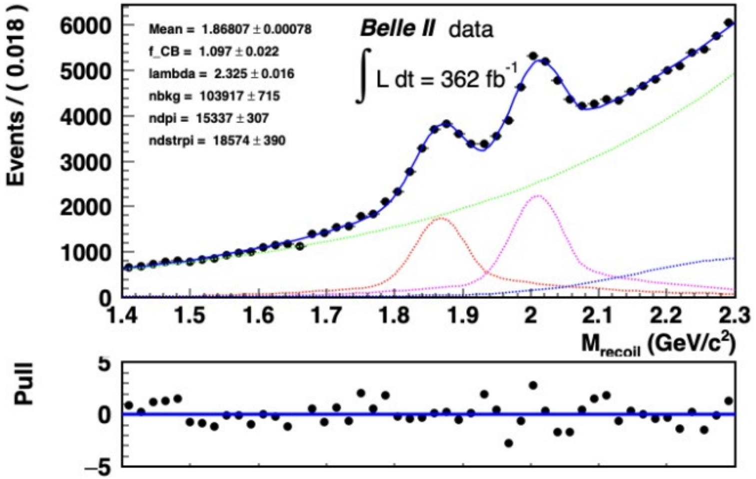
Had FEI calibration with $D^{(*)}\pi$ samples

BELLE2-NOTE-PH-2023-004
M. Liu, N. Rout, K. Trabelsi, V. S. Vobbilisetti

- Calibrate the FEI by partially reconstructing $B \to D^{(*)}\pi$ decays
  - Perform ***B*-tagging using hadronic FEI**
  - **Select pion with highest momentum** from ROE
  - Fit **recoil mass** of the system

B^+ modes
Sig prob > 0.001

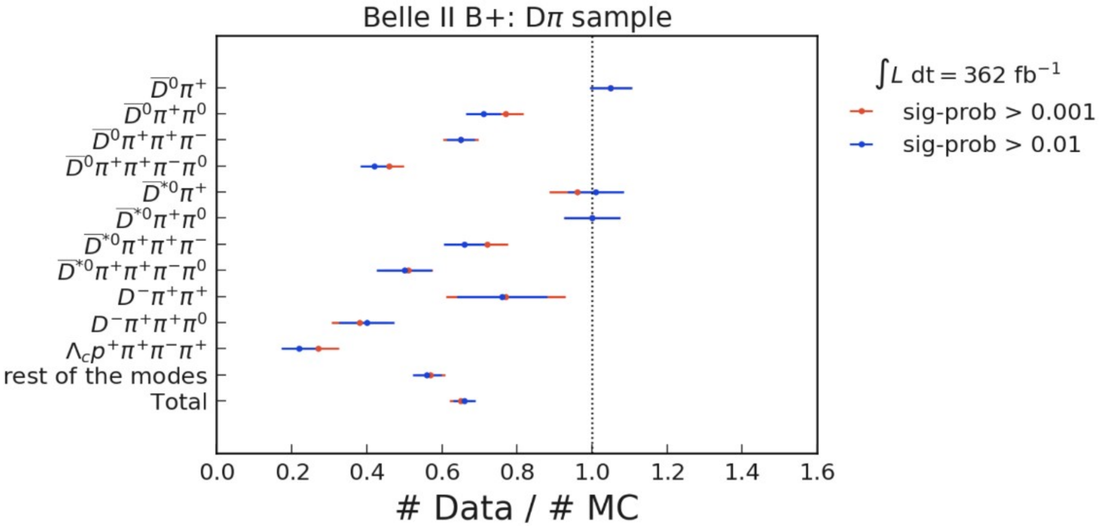


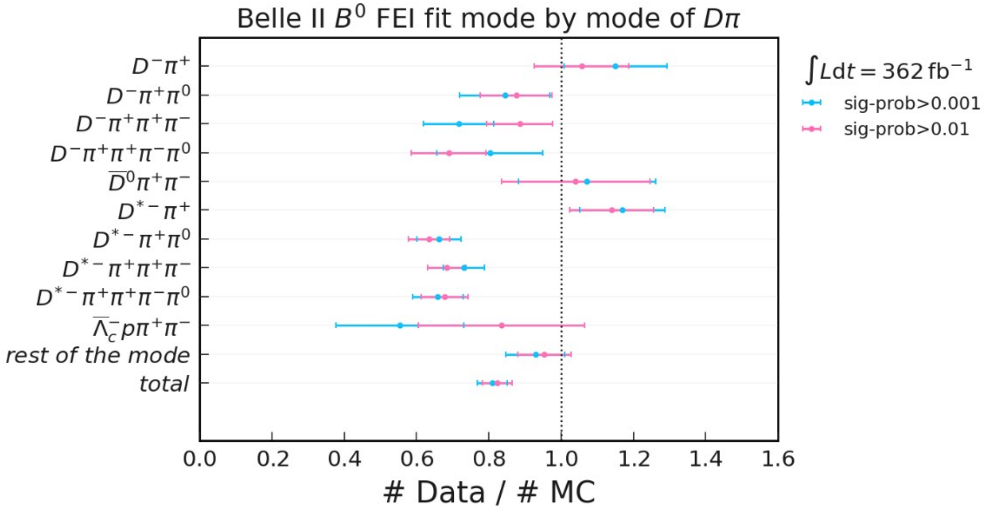


Had FEI calibration with $D^{(*)}\pi$ samples

BELLE2-NOTE-PH-2023-004
M. Liu, N. Rout, K. Trabelsi, V. S. Vobbilisetti

- **Mode-by-mode calibration factors for B^+ and B^0** using $D\pi$ samples



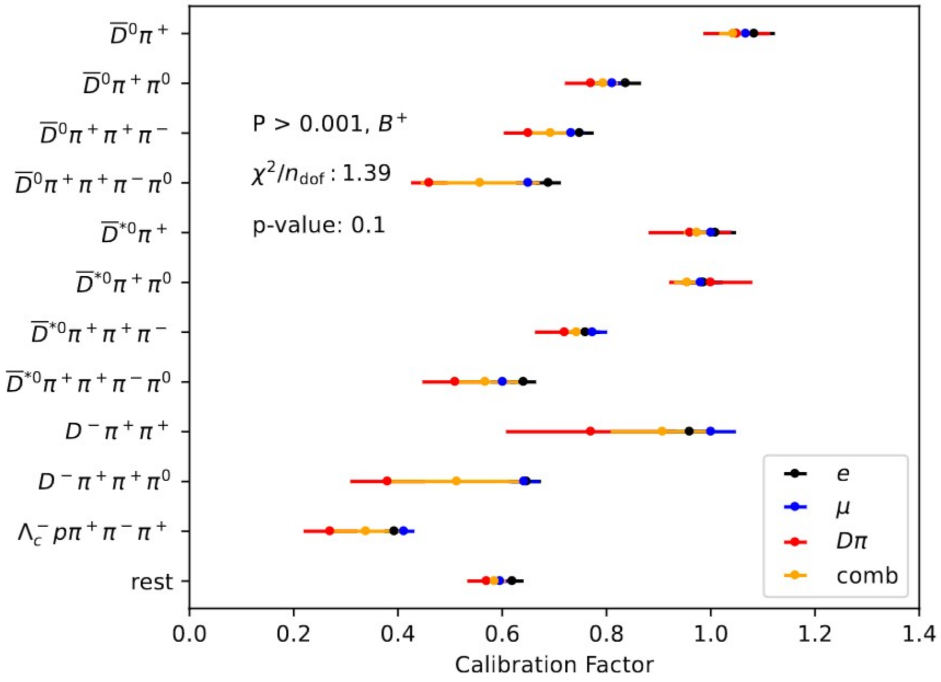


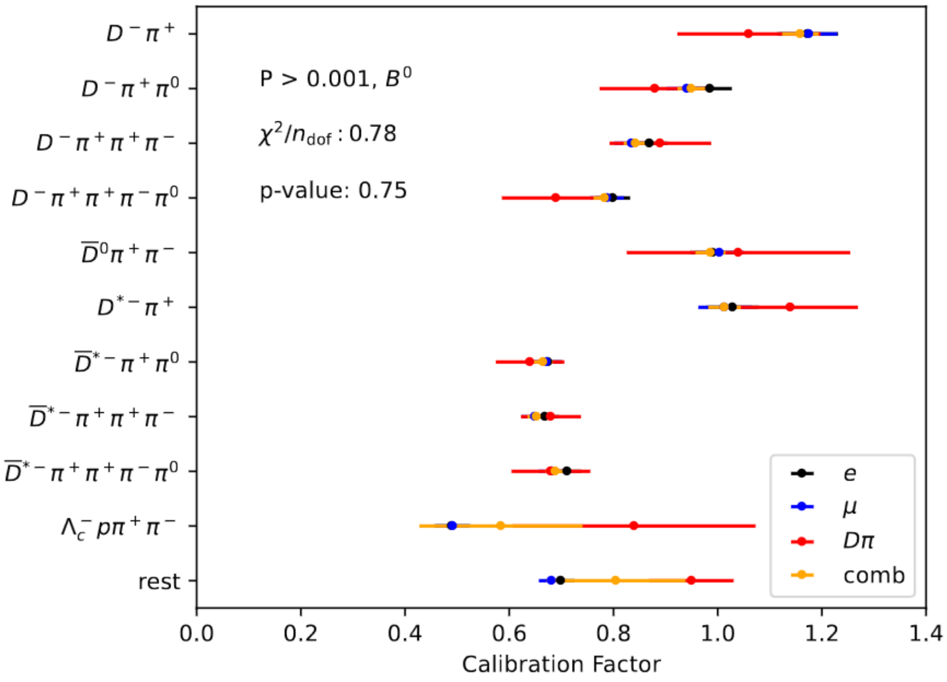
Combination

BELLE2-NOTE-PH-2023-029
M. Liu, N. Rout, K. Trabelsi,
V. S. Vobbilisetti, W. Sutcliffe,
F. Bernlochner

- Combined calibration factors and example application jupyter notebook can be found on kekcc:

  /hsm/belle2/bdata/users/sutclw/fei_calibration/hadronic_FEI_calibration_factors

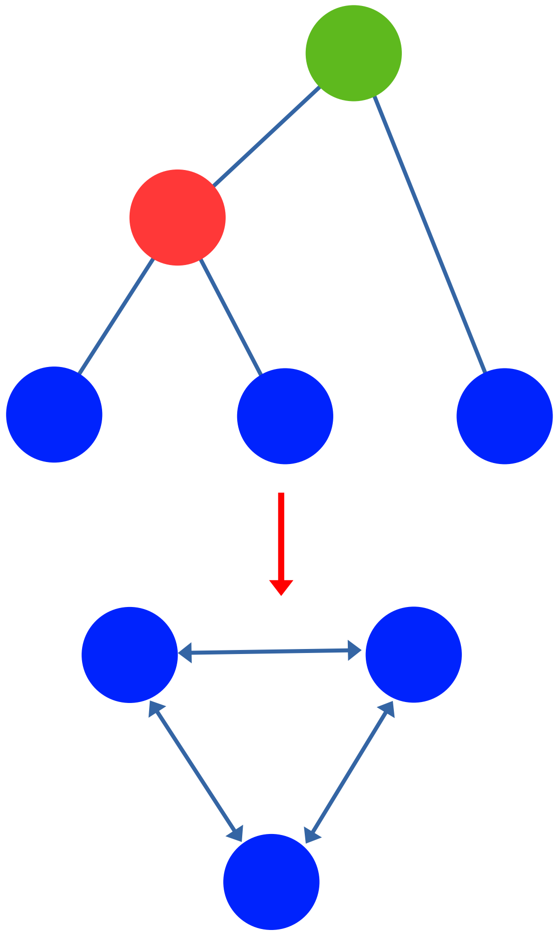




What comes next?

B reconstruction using Graph Neural Networks

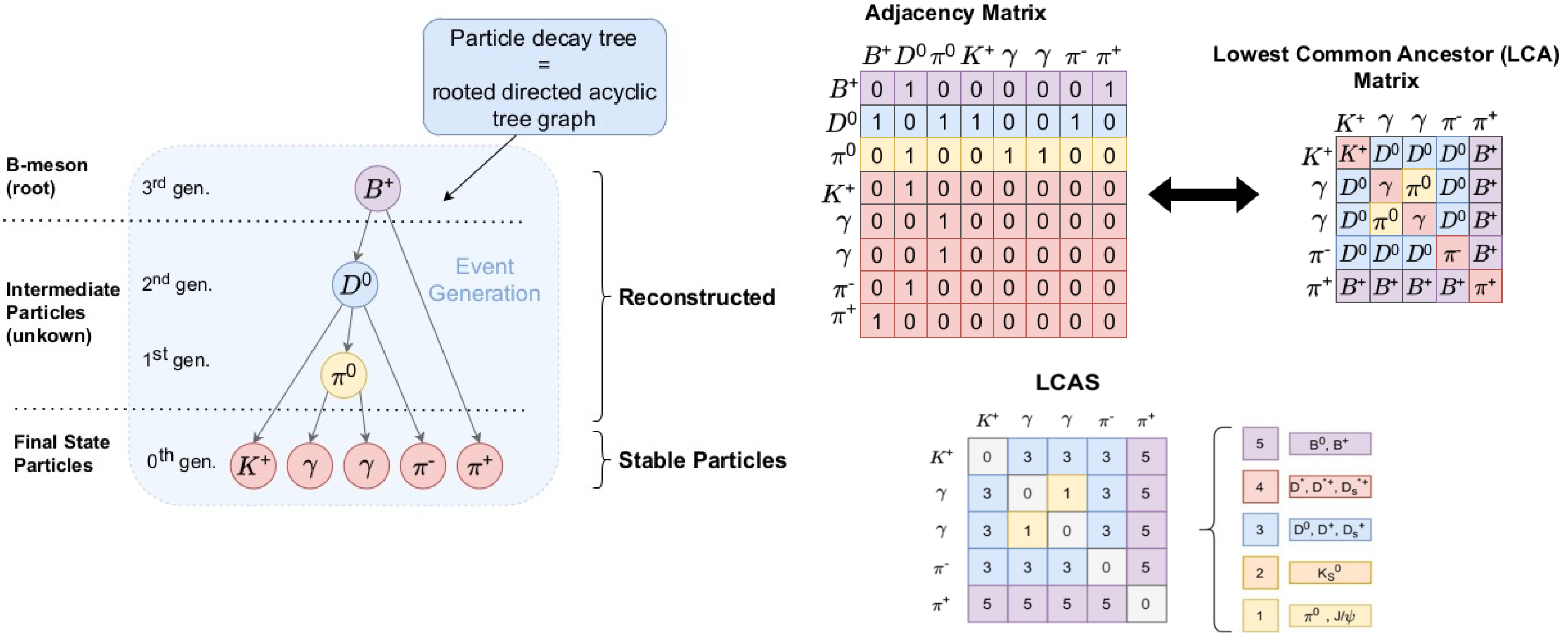
- Main limitation of FEI: **channels need to be hard-coded**
- Decay tree can be encoded in **rooted directed acyclic tree graph**
- Goal: use graph neural networks to inclusively reconstruct B_{tag}
  - **→ Graph-based Full Event Interpretation (graFEI)**



- Challenges:
  - We only have information on FSPs
  - Variable number of FSPs
  - Unknown number of intermediates
- **Solution: encode decay tree as FSP relations**

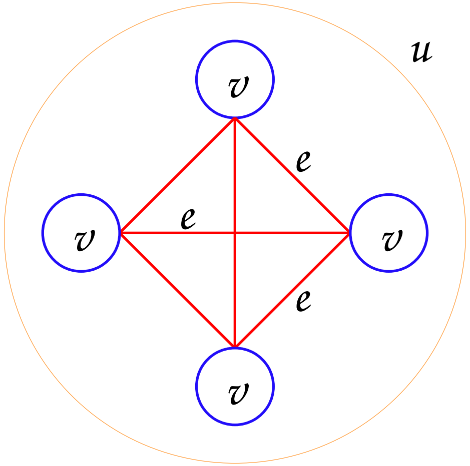
Lowest Common Ancestor (LCA) matrix

- Based on:
  - Ilias Tsaklidis' and Lea Reuter's master theses
  - *Learning tree structures from leaves for particle decay reconstruction*, Kahn et al 2022 Mach. Learn.: Sci. Technol. 3 035012

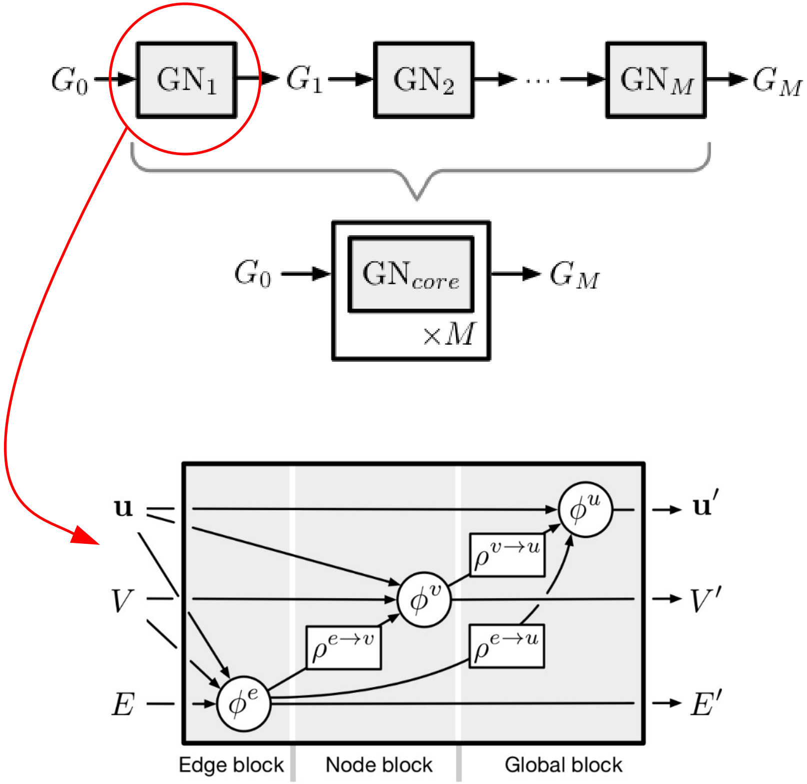


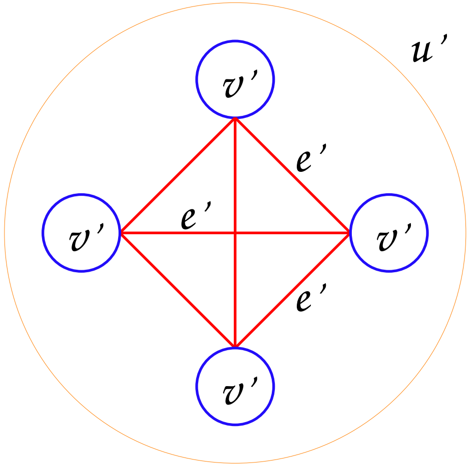
GraFEI – Model description

- Model based on *graph network blocks* [arXiv:1806.01261](https://arxiv.org/abs/1806.01261)
- We input a **fully connected graph**, output graph has same structure with updated attributes

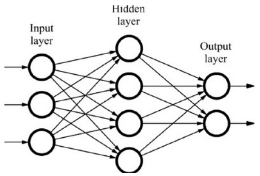


u = global features
v = node features
e = edge features



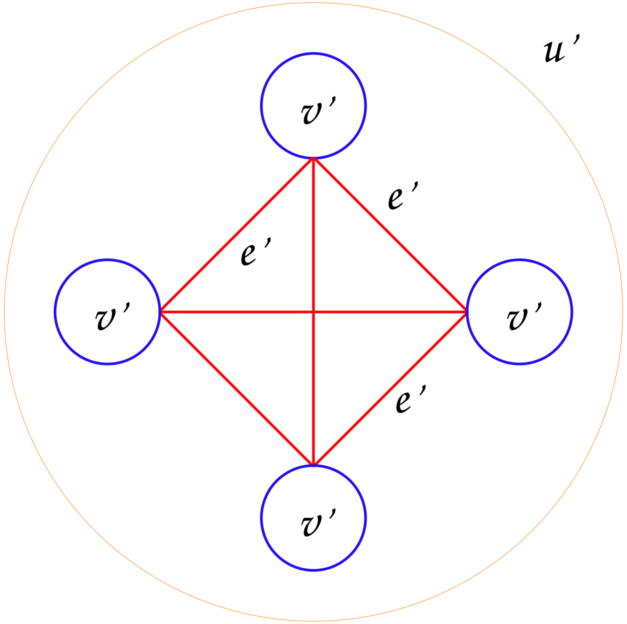


$\phi =$



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

GraFEI - Loss function



$$\text{Loss} = \text{LCA} + \alpha \cdot \text{Particle IDs}$$

6-classes cross-entropy:

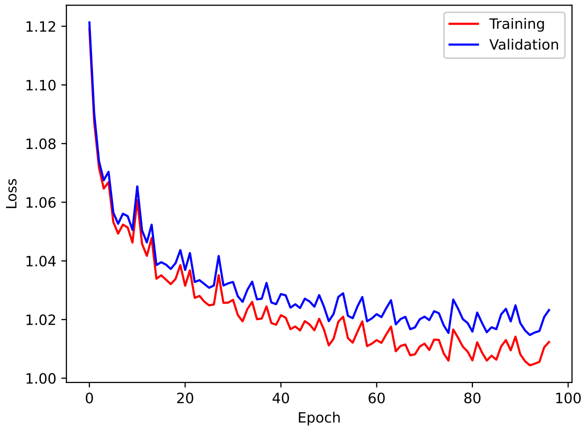
- 5 : B^0
- 4 : $D^{(\pm)}_{(S)}{}^*$
- 3 : $D^{(\pm)}_{(S)}$
- 2 : $K_S{}^0$
- 1 : π^0, J/ψ
- 0 : background

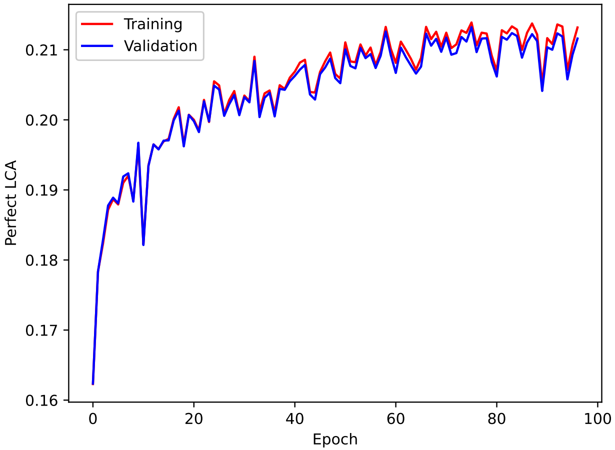
7-classes cross-entropy:

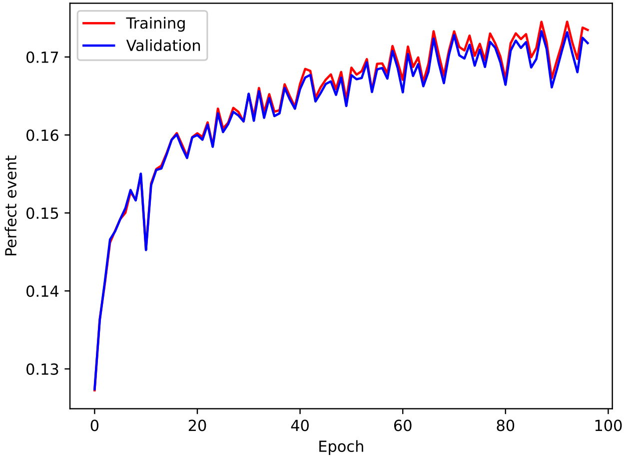
- 6 : γ
- 5 : p
- 4 : K
- 3 : π
- 2 : μ
- 1 : e
- 0 : other

GraFEI – Training

- Model trained on mixed MC: ~3M decays for training + ~150k for evaluation
- Input features:
  - Node-level: particle IDs, pt, pz, charge, dr, dz, clusterNHits, clusterTiming, clusterCharge
  - Edge-level: cos(θ) between momenta, DOCA between tracks



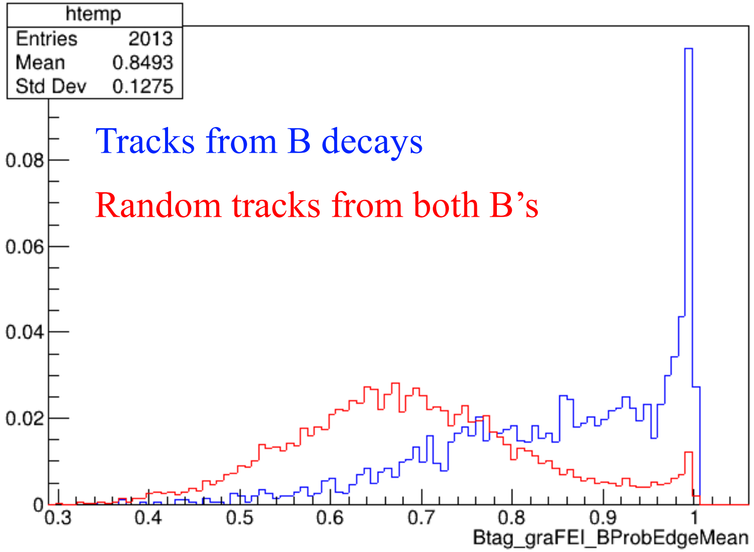




GraFEI - B probability

- Having a **definition of "B probability"** analogous to FEI is needed
  - Each LCA element has a corresponding probability of belonging to the predicted class
  - Arithmetic mean of class probabilities defined as B probability

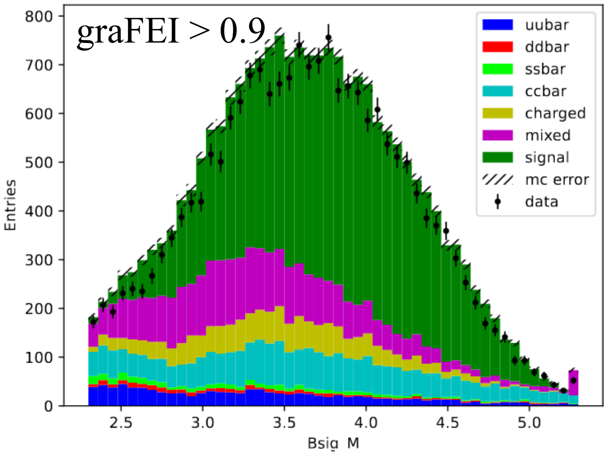
$$\mathrm{LCA} = \begin{pmatrix} 0 & 3 & 5 \\ 3 & 0 & 5 \\ 5 & 5 & 0 \end{pmatrix} \longleftrightarrow \begin{pmatrix} 0 & 0.62 & 0.31 \\ 0.62 & 0 & 0.76 \\ 0.31 & 0.76 & 0 \end{pmatrix} \to 0.563$$

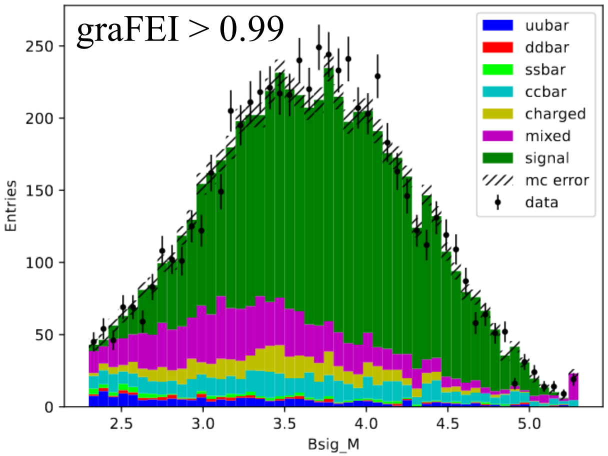


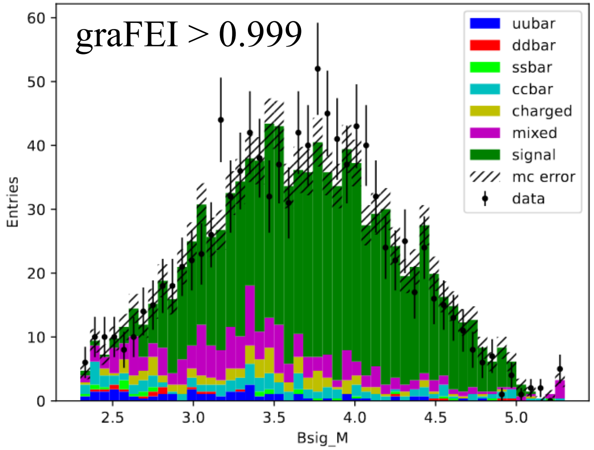
GraFEI in action – Comparison with FEI

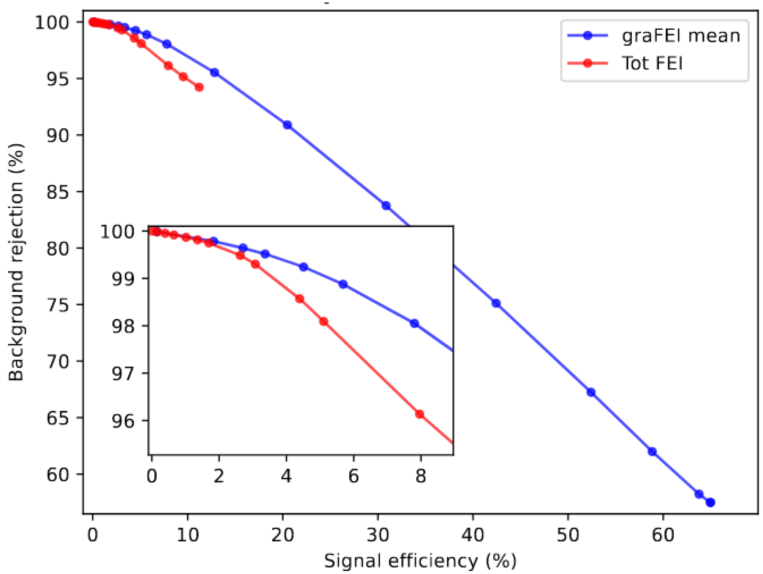
- Applied on tag-side of $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu$ candidates

Preliminary!





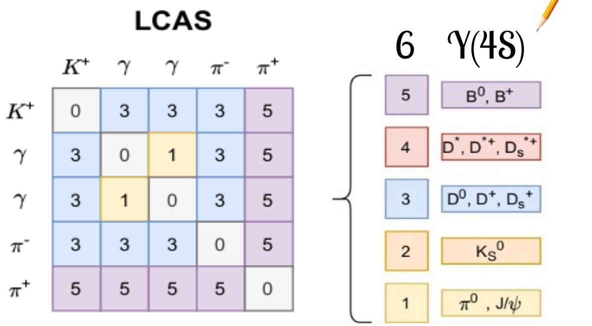


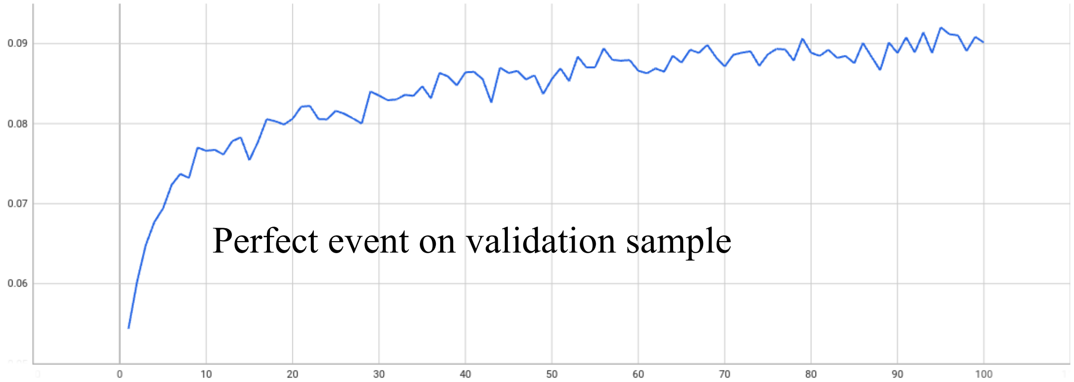


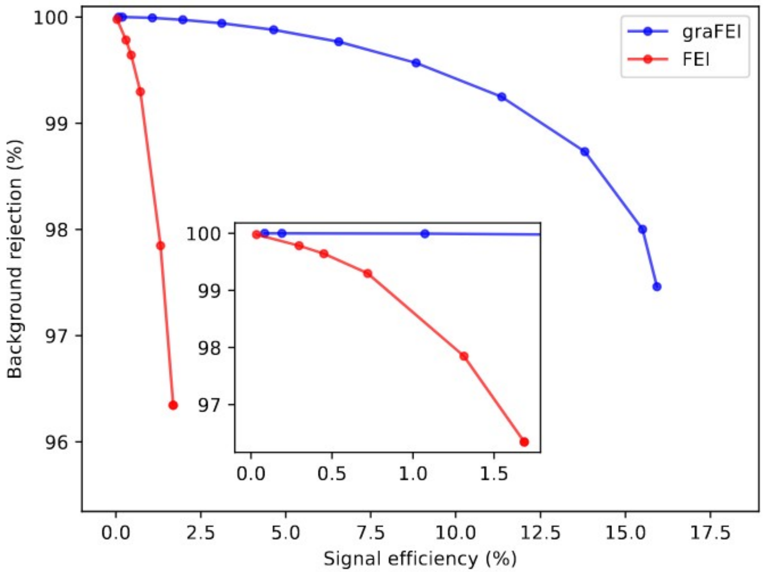
GraFEI – Full $\Upsilon(4S)$ reconstruction

Preliminary!

- Train the model on signal MC to **reconstruct Υ(4S) candidates**
  - Used ~ 3M $B^0 \rightarrow K^{*0} \nu\nu$ signal MC events
  - ~ 2 days training on a GPU Nvidia V100







- Currently writing an internal note and implementing the code into basf2
- **If you want to join us just send me a mail!**

Conclusions

- Take-home messages:
  - Belle II features allow to **reconstruct the partner *B* produced in the event**
  - Reconstruction performed with **Full Event Interpretation** (use FEI skim)
  - Overall performances: ~ 1-2 % efficiency at ~ 5-10 % purity
  - Calibration needed, performed as a function of decay mode and signal probability cut

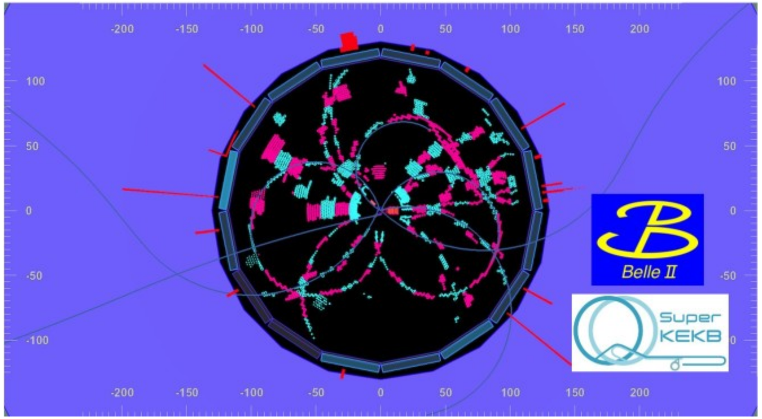
- **GraFEI could be a possible extension**:
  - Based on graph neural network, reconstructs decay topology and particle IDs
  - Seems powerful when trained on signal MC, more investigation and documentation ongoing

BACKUP

Belle II [arXiv:1011.0352]

- **Multi-purpose detector @ SuperKEKB** accelerator
- Focus on B, charm and τ physics
- 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)



graFEI on `phasespace` dataset [arXiv:2208.14924]

- “Perfect world” simulation generated with `phasespace` library
- Comparison of two GNN models:
  - Neural Relational Inference (NRI) [arXiv:1802.04687]
  - Transformer encoder
- Hyperparameter optimisation with **Optuna**
- **4-momentum used as input feature**
- Average 47.7 % perfectly predicted LCAG with NRI
  - 60.9 % for decays with up to 10 leaves, 94.2 % up to 6 leaves