Search for the $B \to K^{(*)}\nu\bar{\nu}$ decays at Belle II

Corentin Santos
Under the supervision of Isabelle Ripp-Baudot and Giulio Dujany

IPHC,
CNRS,
University of Strasbourg,

11 June 2025

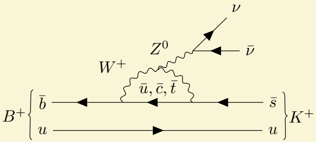
Motivations

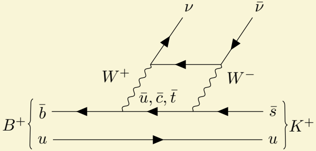
Why $B \to K^{(*)}\nu\bar{\nu}$?

- $b \to s\nu\bar{\nu}$ transition = **Flavor Changing Neutral Current**
- **Rare decay** (suppressed by Glashow–Iliopoulos–Maiani mechanism): Branching Ratio (BR) $\sim \mathcal{O}\left(10^{-5}\right)$
- **Well described** by the Standard Model (SM)

$\Longrightarrow$ **Sensitive to high energy new physics** contributions [arXiv:2301.06990]

$\Longrightarrow$ Since neutrinos are never measured, also **sensitive to light new particles** [arXiv:2503.19025]





Definition (Signal strength)

$$\mu = \frac{BR(B \to K^{(*)}\nu\bar{\nu})_{\text{measured}}}{BR(B \to K^{(*)}\nu\bar{\nu})_{\text{expected}}} \tag{1}$$

Decay channels

$B \longrightarrow K^{(*)}\nu\bar{\nu}$ is a generic term for many possible decays:

- $B^{\pm} \longrightarrow K^{\pm}\nu\bar{\nu}$
- $B^{0} \longrightarrow K^{0}_{S}\nu\bar{\nu}$
- $B^{\pm} \longrightarrow \left(K^{*\pm} \longrightarrow K^{\pm}\pi^{0}\right)\nu\bar{\nu}$
- $B^{\pm} \longrightarrow \left(K^{*\pm} \longrightarrow K^{0}_{S}\pi^{\pm}\right)\nu\bar{\nu}$
- $B^{0} \longrightarrow \left(K^{*0} \longrightarrow K^{\pm}\pi^{\mp}\right)\nu\bar{\nu}$

$\longrightarrow$ Focus on the $B^{\pm} \longrightarrow K^{\pm}\nu\bar{\nu}$ decay channel in this presentation, but the goal of the analysis is to measure the BR of all the decays listed above.

Experimental constraints

- Two neutrinos in a 3-body decay = cannot simply reconstruct the signal
- Lack of kinematic constraints $\implies$ **Complete angular acceptance**
- Only signature is a K $\implies$ **Clean environment**

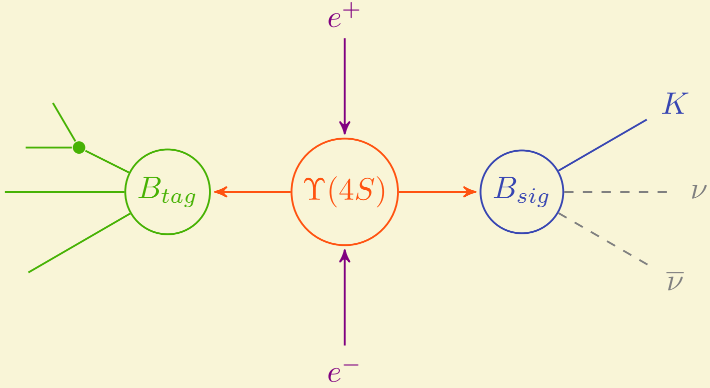
Only experiment capable of measuring the BR of this decay: **Belle II**

Why Belle II?

- e^+e^- collisions $\longrightarrow$ **Clean environment** ✓
- Current integrated luminosity $\mathcal{L} = 365\ \mathrm{fb}^{-1}$ $\longrightarrow$ **High statistics** ✓
- Detector covering a 4π solid angle $\longrightarrow$ **Complete angular acceptance** ✓

What about the almost invisible signal?

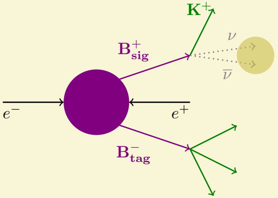
Handling the invisible signal



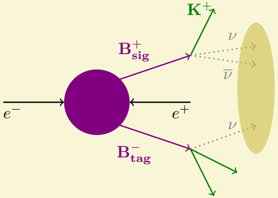
$\longrightarrow$ We infer information about the B_{sig} from the B_{tag} since they are *entangled*

Tagging strategies

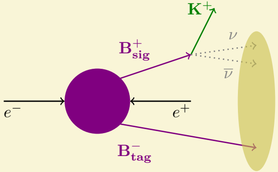
Hadronic Tagged Analysis (HTA)



Semileptonic Tagged Analysis (STA)



Inclusive Tagged Analysis (ITA)



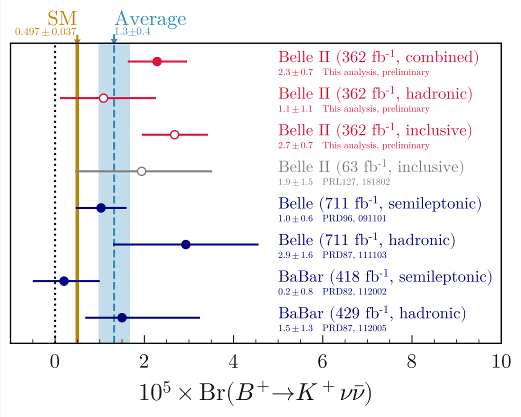
Tagging efficiency

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

Tagging purity

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

Previous $B^+ \to K^+\nu\bar{\nu}$ analyses



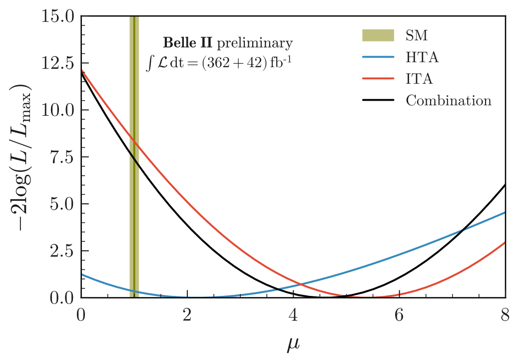


Figure: from arXiv:2311.14647

Goal: Carry an independent measurement using *semileptonic tagging*

Why semileptonic tagging?

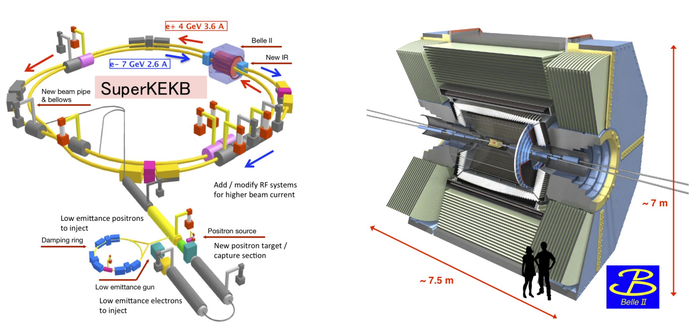
- Higher efficiency than HTA
- Higher purity than ITA
- Complementary to the other two Belle II analyses
- Complementary to the Belle and BaBar analyses

Analysis	Uncertainty on the BF naively scaled to 362 fb^{-1}
Belle HTA	2.29×10^{-5}
Belle STA	0.80×10^{-5}
BaBar HTA	1.41×10^{-5}
BaBar STA	0.81×10^{-5}
Belle II HTA	1.20×10^{-5}
Belle II ITA	0.71×10^{-5}

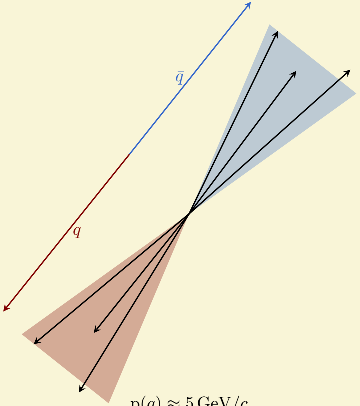
Table: Comparison of the uncertainties on the branching fraction of the various analyses, scaled to 362 fb^{-1}

The Belle II experiment

- International collaboration
- Taking data on the SuperKEKB accelerator, in Japan, since 2019
- e^+e^- collisions at $\sqrt{s} = 11$ GeV
- Record for highest instantaneous luminosity: 5.1×10^{34} cm^2/s

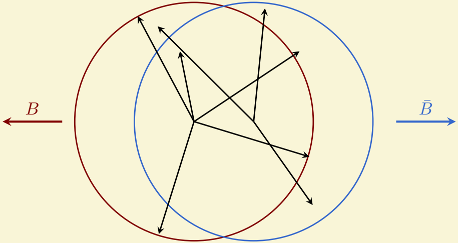


Event geometries



$\mathrm{p}(q) \approx 5\,\mathrm{GeV}/c$

$e^+e^- \to q\bar{q} \quad (q \in \{u, d, s, c\})$

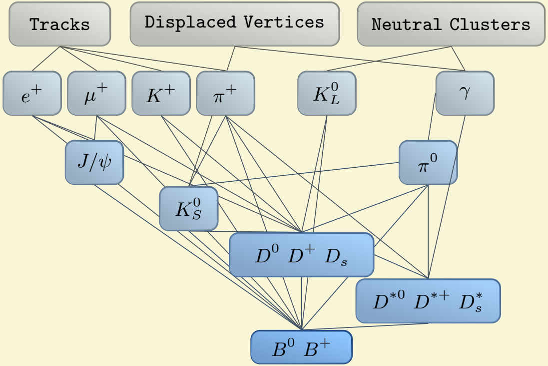


$\mathrm{p}(B) \approx 0.3\,\mathrm{GeV}/c$

$e^+e^- \to \Upsilon(4S) \to B\bar{B}$

Full Event Interpretation (FEI)

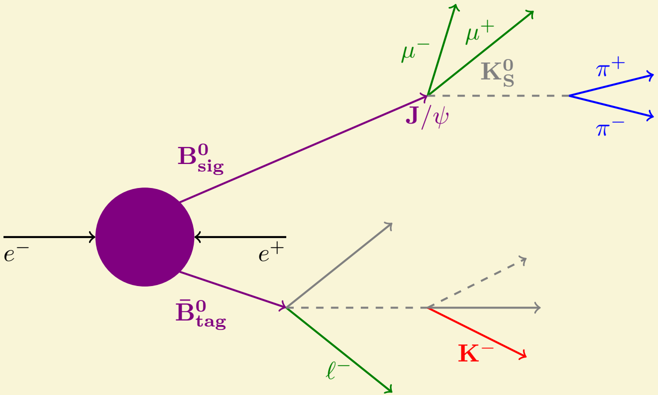
- **Tagging algorithm**: reconstruct the tag side
- Based on **Boosted Decision Trees** (BDT)
- Can be *hadronic* or **semileptonic**
- For the semileptonic, we consider the following B_{tag} decays:
  - $B \longrightarrow De\nu$
  - $B \longrightarrow D\mu\nu$
  - $B \longrightarrow D^{*}e\nu$
  - $B \longrightarrow D^{*}\mu\nu$
  - $B \longrightarrow D^{(*)}\ell\nu\pi$, $\ell \in \{e, \mu\}$, also available, but not used



About tagging

- **Tagging** is an important technique when no direct information on the signal side is available
- Essential for the $B \to K^{(*)}\nu\bar{\nu}$ search, since quasi-invisible

$\longrightarrow$ Can also be used for study of CP violation in complex B decays



CP-asymmetry

- Want to measure the *CP*-asymmetry

Definition (*CP*-asymmetry)

$$a^{CP} = \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{2}$$

$\longrightarrow$ Need to know the *flavour* of the B meson at the time of its decay (B^0 or $\bar{B}^0$)

- **Self-tagged** decays: one of the final state particles gives away the flavour of the B meson

- *What if the signal is not self-tagged ?*

CP-asymmetry

- Want to measure the *CP*-asymmetry

Definition (*CP*-asymmetry)

$$a^{CP} = \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{2}$$

$\longrightarrow$ Need to know the *flavour* of the B meson at the time of its decay (B^0 or $\bar{B}^0$)

- **Self-tagged** decays: one of the final state particles gives away the flavour of the B meson

- *What if the signal is not self-tagged ?*

CP-asymmetry

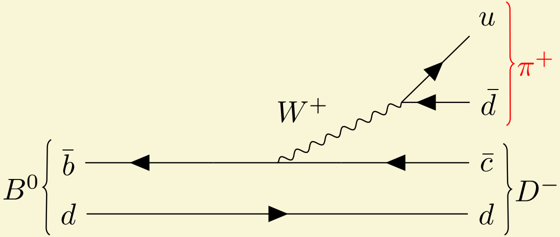
- Want to measure the *CP*-asymmetry

Definition (*CP*-asymmetry)

$$a^{CP} = \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{2}$$

$\longrightarrow$ Need to know the *flavour* of the B meson at the time of its decay (B^0 or $\bar{B}^0$)

- **Self-tagged** decays: one of the final state particles gives away the flavour of the B meson



- *What if the signal is not self-tagged ?*

CP-asymmetry

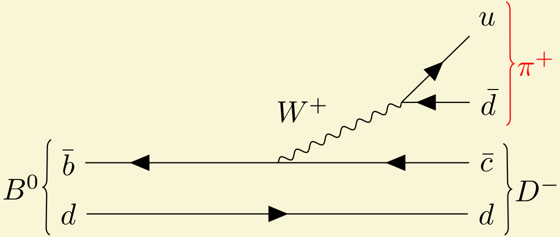
- Want to measure the CP-asymmetry

Definition (CP-asymmetry)

$$a^{CP} = \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{2}$$

$\longrightarrow$ Need to know the *flavour* of the B meson at the time of its decay (B^0 or $\bar{B}^0$)

- **Self-tagged** decays: one of the final state particles gives away the flavour of the B meson

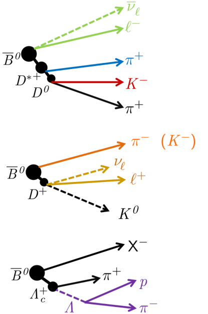


- *What if the signal is not self-tagged ?*

Tagging (again)

- Can rely on self-tagging B_{tag} decays
- 13 signature categories were define:

Categories	Targets
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
KinLepton	e^-
Intermediate KinLepton	ℓ^+
Kaon	K^-
KaonPion	K^-, π^+
SlowPion	π^+
FastHadron	π^-, K^-
MaximumP	ℓ^-, π^-
FSC	ℓ^-, π^+
Lambda	Λ
Total= 13	



- Categories $\longrightarrow$ Classification $\longrightarrow$ Classifier $\longrightarrow$ Machine Learning

$\Longrightarrow$ Introducing the FLAVOUR TAGGER

Parameters

- Flavour tagger characterized by its **tagging efficiency**:

Definition (Tagging efficiency)

$$\varepsilon = \frac{N^{\text{tag}}_{B^0} + N^{\text{tag}}_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{3}$$

- Mistagging quantified by the **wrong tag fraction**, w:

Definition (Wrong tag fraction)

$$N^{\text{tag}}_{B^0} = \varepsilon \left((1 - w) N_{B^0} + w N_{\bar{B}^0} \right) \tag{4}$$

$$N^{\text{tag}}_{\bar{B}^0} = \varepsilon \left(w N_{B^0} + (1 - w) N_{\bar{B}^0} \right) \tag{5}$$

Observed vs True *CP*-asymmetry

- Observed *CP*-asymmetry written as

$$a^{CP}_{\text{obs}} = \frac{N^{\text{tag}}_{B^0} - N^{\text{tag}}_{\bar{B}^0}}{N^{\text{tag}}_{B^0} + N^{\text{tag}}_{\bar{B}^0}} = (1 - 2w)\, a^{CP}$$

Definition (Dilution factor)

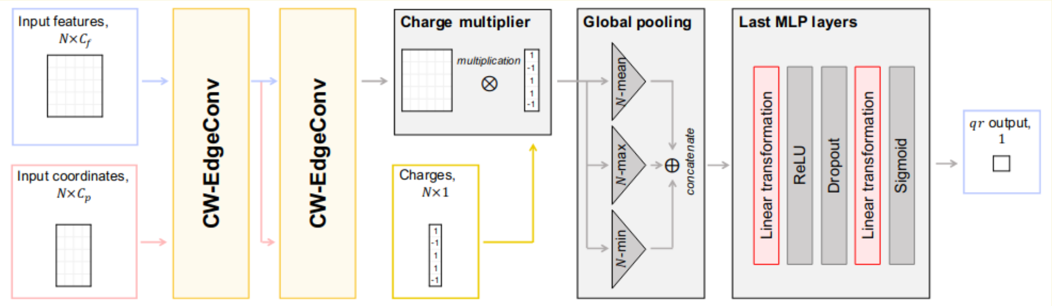
$$r \equiv |1 - 2w| \in [0, 1] \tag{6}$$

Interpretation:

- If $w = \frac{1}{2}$, then $r = 0$ and it gives us no information
- If $w = 0$ or $w = 1$, then $r = 1$ and it gives us the full information

Flavour Tagger algorithm: GFlaT

- Based on a Graph Neural Network [arXiv:2402.17260]
- Procedure:
  1. Reconstructs and classifies the tracks among the 13 categories with a probability
  2. Takes into consideration all the probabilities and computes qr, where $q \in \{-1, 1\}$ is the flavour of the particle



Calibration strategy: ersatz

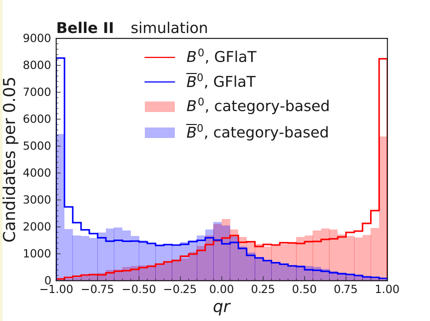
Goal:

- Implement a calibration strategy for the FLAVOUR TAGGER in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the FLAVOUR TAGGER on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

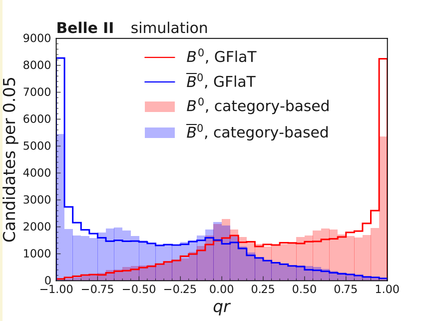
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the Flavour Tagger on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

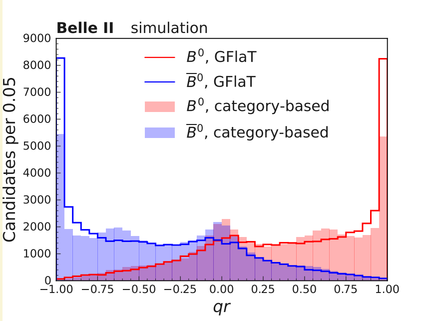
Goal:

- Implement a calibration strategy for the FLAVOUR TAGGER in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the FLAVOUR TAGGER on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

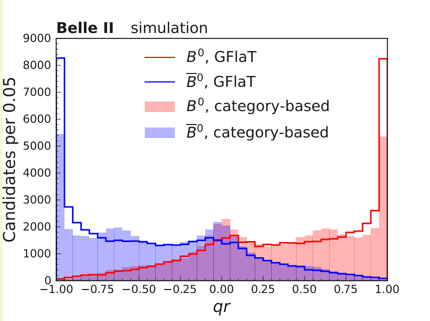
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the Flavour Tagger on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

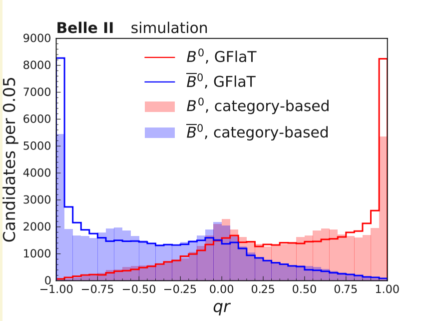
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the Flavour Tagger on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

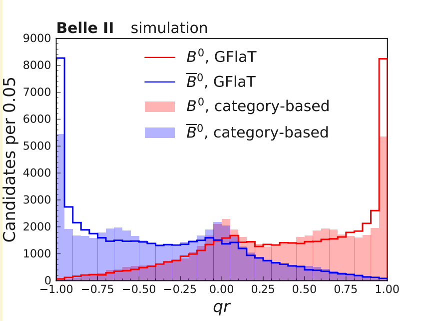
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the Flavour Tagger on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency



Calibration strategy: ersatz

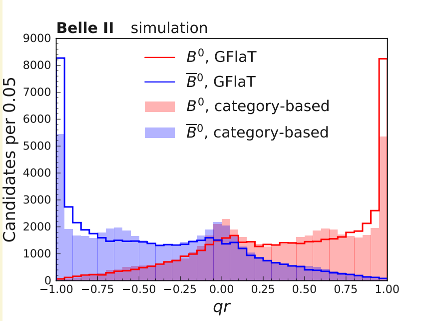
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimize the calibration process

Strategy:

- Want to measure the tagging efficiency and the wrong tag fraction

1. Work with events with self-tagged decays for the B_{sig}
2. Use the Flavour Tagger on those events to get qr
3. Create 7 bins of $|qr|$
4. Fit the overall ΔE PDF
5. Extract the N^{tag} for B^0 and $\bar{B}^0$
6. Compute the wrong tag fraction and the tagging efficiency

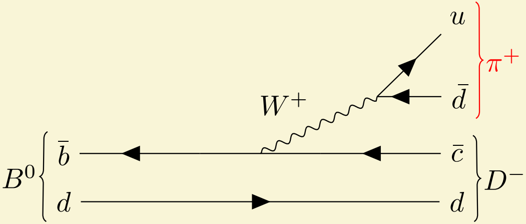


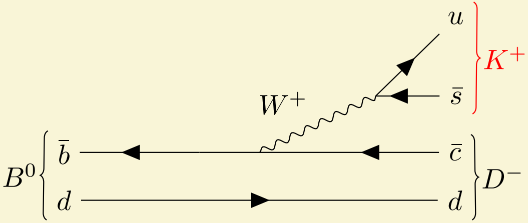
Calibration strategy: signal

- Need self-tagged decays:

$$B^0 \to D^-\pi^+$$
$$B^0 \to D^{*-}(\to \bar{D}^0(K^+\pi^-)\pi^-)\pi^+$$
$$B^0 \to D^{*-}(\to \bar{D}^0(K^+\pi^-\pi^0)\pi^-)\pi^+$$
$$B^0 \to D^{*-}(\to \bar{D}^0(K^+\pi^-\pi^-\pi^+)\pi^-)\pi^+$$

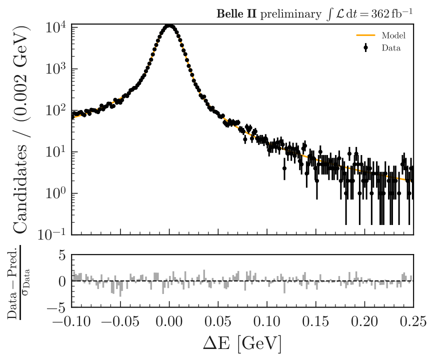
- Also consider $B \to D^{(*)-}K^+$ equivalents, called "BDK"s

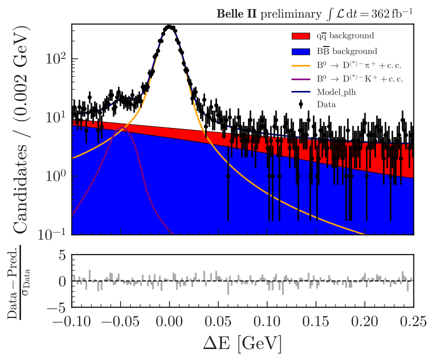




Calibration strategy: extracting the yields

- Important kinematic variable for signal selection: $\Delta E = E_B - E_{\text{beam}}$
- Fit the ΔE distributions on MC for the signals $BD\pi$ and BDK, as well as the $B\bar{B}$ and $q\bar{q}$ backgrounds
- Sum all the shapes and fit the overall ΔE PDF
- Extract the yield





Calibration strategy: wrong tag fraction

- Need to discriminate between:
  - Same or Opposite Flavours (SF or OF)
  - Flavour of the B_{tag} (+ or -)
- Define 7 bins of $|qr|$
- Compute the wrong tag fractions $w_i^{\pm}$ for each bin $i \in [\![0, 6]\!]$

$$w_i^{\pm} = \frac{f_i^{\pm} - R}{(1 - R)\left(1 + f_i^{\pm}\right)}$$

with

$$f_i^{\pm} = \frac{n^{\pm}_{\mathrm{OF},i}}{n^{\pm}_{\mathrm{SF},i}}, \quad R = \frac{\chi_d}{1 - \chi_d}, \quad \chi_d = 0.1860 \pm 0.0011[2]$$

and deduce

$$\boxed{w_i = \frac{1}{2}\left(w_i^{+} + w_i^{-}\right)}$$

Calibration strategy: tagging efficiency

- Define a tagging efficiency for each bin $i \in [\![0, 6]\!]$

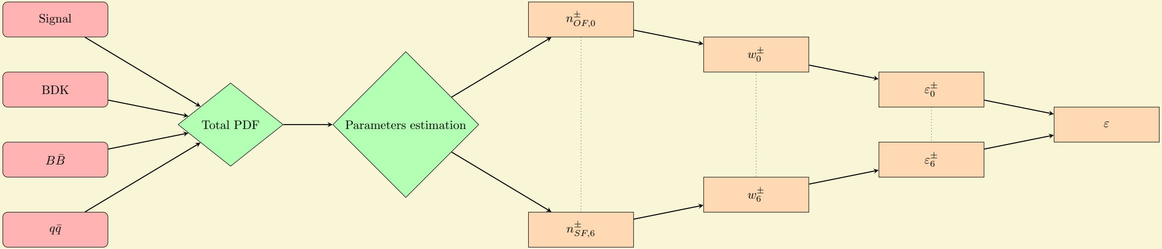
$$\varepsilon_i^{\pm} = \frac{n_{\mathrm{OF},i}^{\pm} + n_{\mathrm{SF},i}^{\pm}}{N_{B^0}^{\mathrm{tag}} + N_{\bar{B}^0}^{\mathrm{tag}}}$$

and compute the *effective* tagging efficiency

Definition (Effective tagging efficiency)

$$\varepsilon = \sum_{i=0}^{6} \varepsilon_i \left(1 - 2w_i\right)^2 \tag{7}$$

Calibration workflow



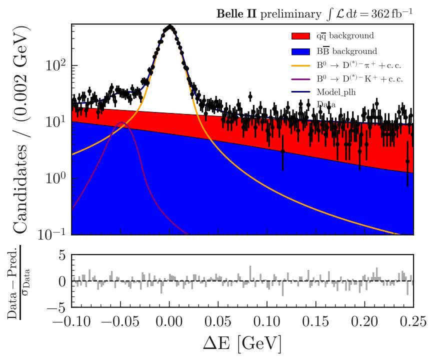
Calibration optimization: simultaneous fit

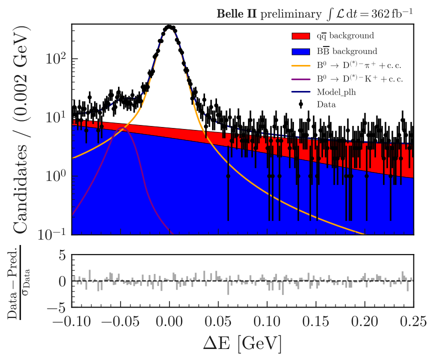
- **Before this work:**
  - Fits of the various shapes (signal, BDK, $B\bar{B}$, $q\bar{q}$) were done sequentially
  - Fits were independent on the the bins and the categories
- **Idea:**
  - Parallelize the fits of the shapes
  - Simultaneously fit the 7 bins and the 4 categories

Calibration optimization: backgrounds studies

- $q\bar{q}$: reduced continuum background
  - Via a cut on a geometric variable ✓
  - Via a BDT (worked on MC but not on data)
- $B\bar{B}$: studied the bump observed at $\Delta E = 0$

⟶ due to signal events wrongly considered as background at MC level





Calibration optimization: automatization

- Automatize the processes using b2luigi
- **Pros:** Parallelize processes and easy to use
- **Whole workflow automated**:
  - Reconstruction of the events, management of the dataset and pre-processing
  - Calibration strategy and visualization

Calibration optimization: summary

- Parallelized the fits of the shapes
- Transform the independent fits into a simultaneous fit
- Reduced the continnum background
- Studied the $B\bar{B}$ background
- Automatized the workflow using B2LUIGI

Results:

- Wrong tag fractions (bin-by-bin): 48%, 42%, 35%, 24%, 18%, 11%, 3%
- Effective tagging efficiency on data: $\varepsilon = 36.90 \pm 0.8\%$

Analysis flow

1. Preselection
2. Apply data/MC corrections
3. Train and apply Boosted Decision Tree (BDT) for selection
4. Compute systematics uncertainties
5. Fit to get the signal strength μ

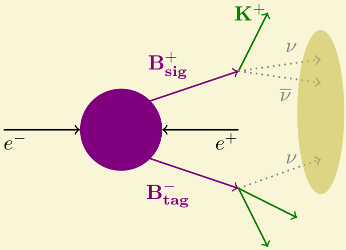
Control samples:

- For $q\bar{q}$ continuum: off-resonance data and MC
- For $B\bar{B}$ background: wrong charged and pion-enriched sample
- For signal: embedding

Reconstruction and preselection

Overview:

1. Reconstruct the B_{tag} using FEI
2. Reconstruct the signal-side Kaon (see details on the cut in the appendix 1)
3. Combine the corresponding tracks
4. Remove all extra "*good*" tracks and remaining π^0, Λ, K_S^0



Corrections overview

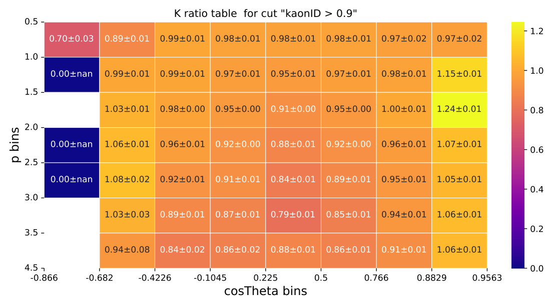
Corrections	Status
Particle Identification (PID)	✓
FEI	✓
Signal Form Factors	*to be improved*
Continuum	✓
ECLCluster[1] Extra Energy (EEXTRA)	*to be improved*
Specific backgrounds	*to be improved*

Table: Table of the corrections and their status

[1]ECL = Electromagnetic Calorimeter

FEI and PID corrections

- PID and FEI corrections based on recommendations
- May switch to SYSVAR



mode	labels	cal factor	error
mode0	$De\nu$	1.18	0.09
mode1	$D\mu\nu$	1.03	0.08
mode2	$D^*e\nu$	0.97	0.07
mode3	$D^*\mu\nu$	0.91	0.06

Table: Calibration factors and errors for FEI, with sgn probability cut > 0.004, on B^+

Continuum correction (BDTc)

Computed by reweighting the MC to the data based on the BDTc output in the off-resonance sample, $weight = \frac{BDTc}{1-BDTc}$.

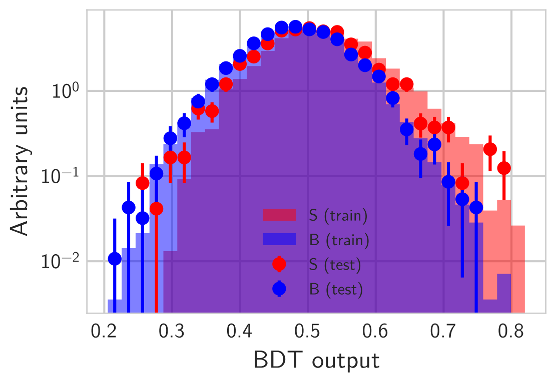


Figure: BDT output for offres data (in red) and offres MC (in blue)

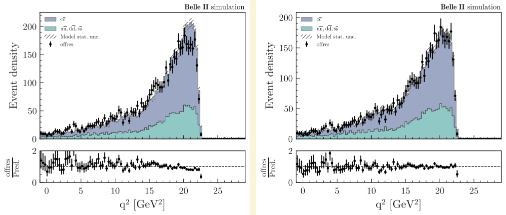
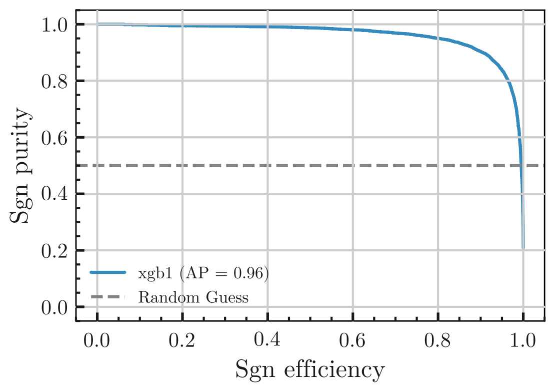


Figure: Plots of q^2 before and after correction

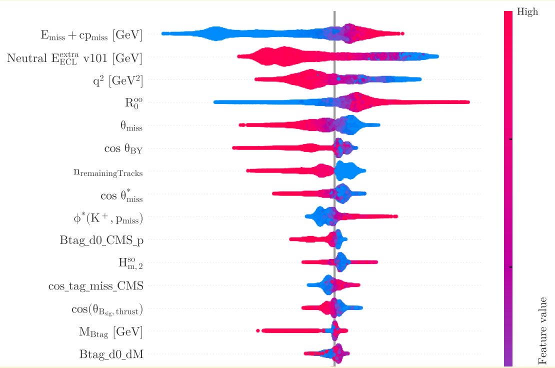
Selection

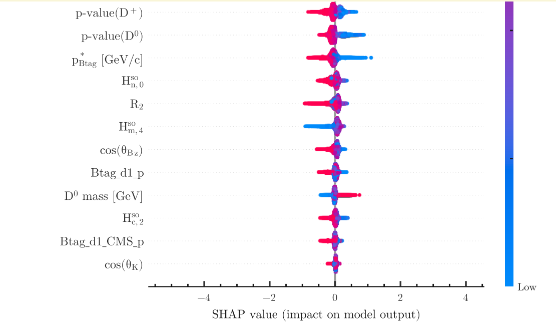
- Using XGBCLASSIFIER
- Using Shapley values for feature importance
- F_1-score for signal on validation data: 97%
- **Signal region**: BDT output > 0.4
- Efficiency after cut on BDT output > 0.4: 77.5×10^{-4}



BDT input variables

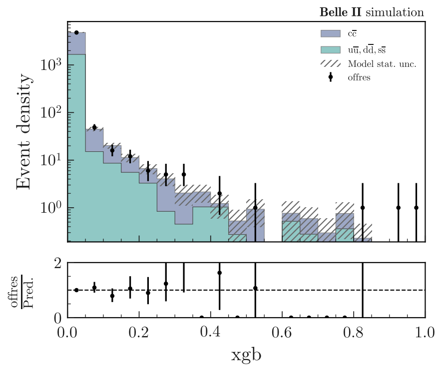
- The color gradient indicates the value of the feature
- The x-axis is the Shapley value (see Slide 74); the more on the right(left) it is, the more it contributes to the signal(background) prediction

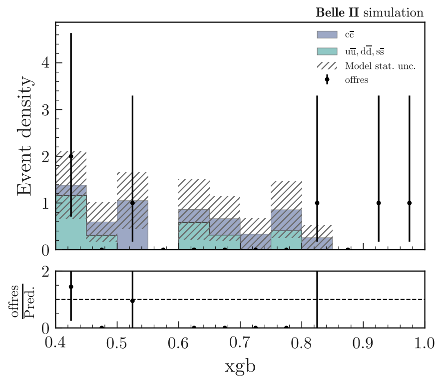




Selection on sidebands (offres)

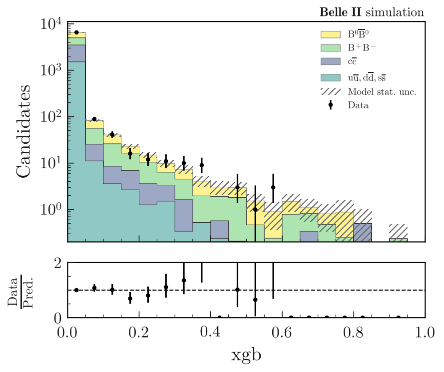
Sideband	data/MC in preselection	data/MC in signal region
offres	1.114 ± 0.021(4900data events)	0.7 ± 0.5(6data events)

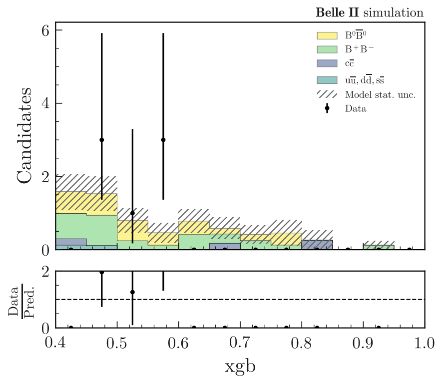




Selection on sidebands (wc)

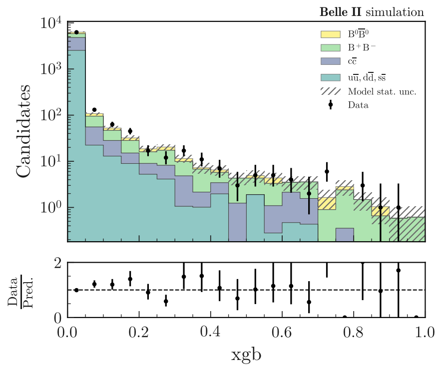
Sideband	data/MC in preselection	data/MC in signal region
wc	0.963 ± 0.019(6721data events)	1.1 ± 0.5(7data events)

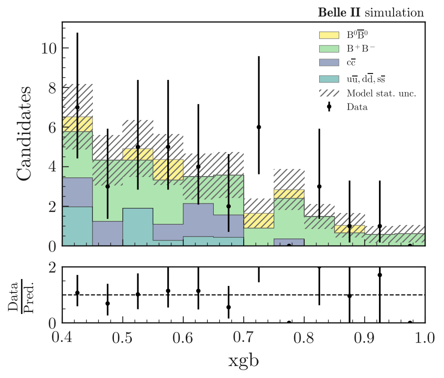




Selection on sidebands (kid)

Sideband	data/MC in preselection	data/MC in signal region
kid	1.101 ± 0.012(6588data events)	1.8 ± 0.5(37data events)





New possible sideband: extra tracks

- Based on "Measurement of $B \to \tau\nu$ Branching Ratio with Hadronic FEI Tagging in the Run1 dataset"
- Will use events with extra tracks to validate the E^{extra} corrections on $B\bar{B}$
- Still need to work on the details of the selection

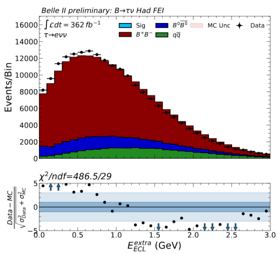


Figure: EEXTRA distribution for the extra tracks sideband

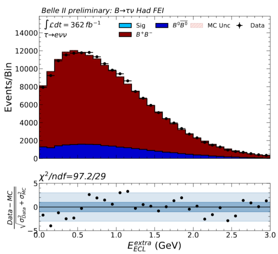
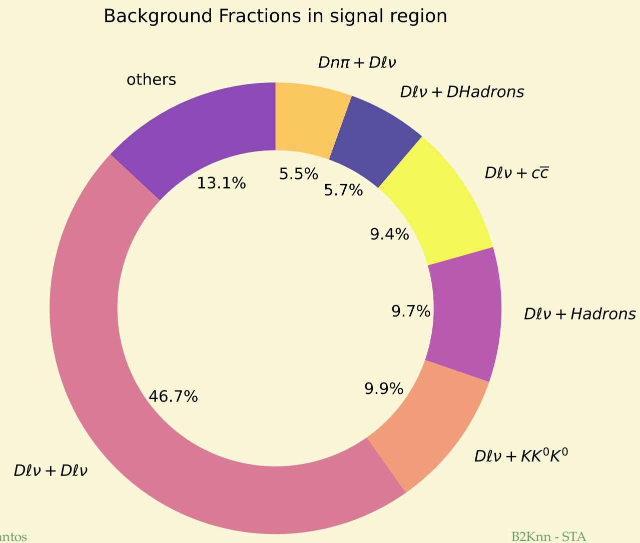


Figure: EEXTRA distribution for the extra tracks sideband after continuum substraction and bin-by-bin correction

Backgrounds



Specific corrections and systematics for the following backgrounds:

- $D \longrightarrow K_L^0 X$
- $B \longrightarrow D^{**}$
- $B^+ \longrightarrow K^+ K_L^0 K_L^0$
- $B^+ \longrightarrow K^+ n\bar{n}$

Methodology

- Create histograms in signal region (BDT > 0.4) with 12 bins
- Run a maximum likelihood fit with pyhf on the data
- Compute the profile likelihood ratio with the signal strength μ as the parameter of interest
- Get the expected error on BR and expected upper limits

Definition (Likelihood function)

$$\mathcal{L}(\mu, \theta|\mathbf{n}) = \frac{1}{Z}\prod_{b=1}^{12} \mathrm{Poisson}\left(n_b|\nu_b(\mu, \theta)\right) p(\theta) \tag{8}$$

with:

$\theta \in \mathbb{R}^N$ is a vector of N nuisance parameters;

n_b is the number of events in bin b, ν_b the expected number of events in bin b;

Z a normalization constant;

p is a prior acting as a constraint on the nuisance parameters.

Next steps

The whole machinery is present for the analysis flow, we now plan to optimize and study in details each steps:

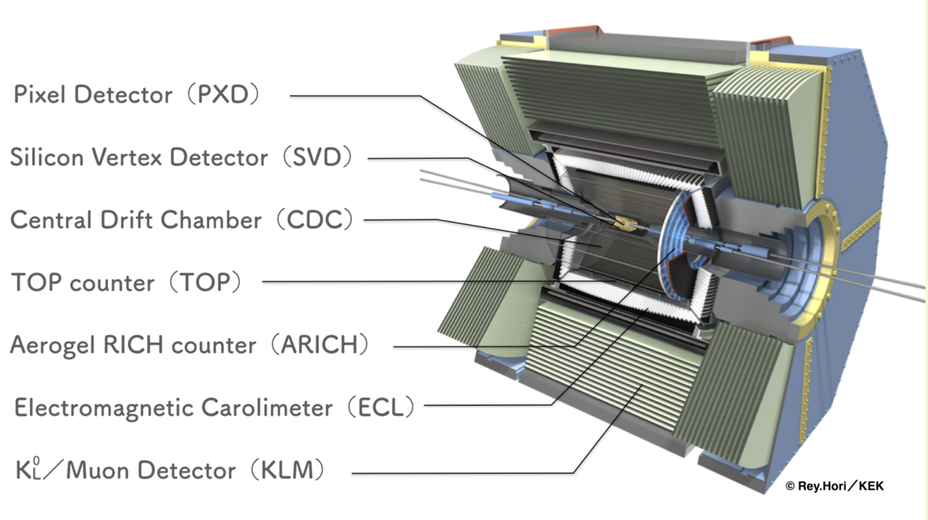
- Detailed study of the backgrounds
- Embedding samples for signal efficiency validation
- Optimize BDTs features and hyperparameters
- Improve systematics evaluations

Trainings, conferences and publications

- Scientifc trainings: 54h
- Transversal trainings: 55h
- Internation conference already attended: CHEP 2024
- Other conferences expected in Winter 2026
- Article: proceedings of CHEP 2024 [arxiv:2503.09401]
- Poster: soon presenting at the Congrès Annuel de la Société Française de Physique (SFP) 2025

6 Supplementary Material

Belle II subdetectors



6 Supplementary Material

Preselection

"Good Track":

- $p_t > 0.2$ GeV/c
- $E < 5.5$ GeV
- thetaInCDCAcceptace
- $dr < 2$ cm
- $|dz| < 4$ cm

K^+ :

- "Good Track"
- $dr < 0.5$ cm
- $|dz| < 2$ cm
- Kaon ID > 0.9[a]

π^0 :(pi0eff30_May2020)

- InvM $\in]0.120, 0.145[$ GeV/c^2
- daughterDiffOfPhi $\in]-1.5, 1.5[$ rad
- daughterAngle < 1.4 rad

[a] will switch to Neural Network PIDs when ready

Preselection

ROE Mask: for signal and tag

- "Good Track"
- `clusterReg==1 and E > 0.08 GeV`

or `clusterReg==2 and E > 0.03 GeV`

or `clusterReg==3 and E > 0.06 GeV`

- clusterNHits > 1.5
- $|\text{clusterTiming}|$ `< 200 ns`
- `0.2967 < clusterTheta < 2.6180 rad`

FEI cuts:

- $\cos(\theta_{BY}) \in]-4, 3[$
- `cosTBTO < 0.9`
- `SignalProbability > 0.004`

Rest Of Event

ROE Cluster Cuts: (for EExtra)

- "Good Track"
- `clusterReg == 1 and clusterE > 0.100` GeV

or `clusterReg == 2 and clusterE > 0.060` GeV

or `clusterReg == 3 and clusterE > 0.150` GeV

- `0.2967 < clusterTheta < 2.6180`

Two masks studied for now:

- `minC2TDist>50` cm
- `beamBackgroundSuppression > 0.5` and `fakePhotonSuppression > 0.5`

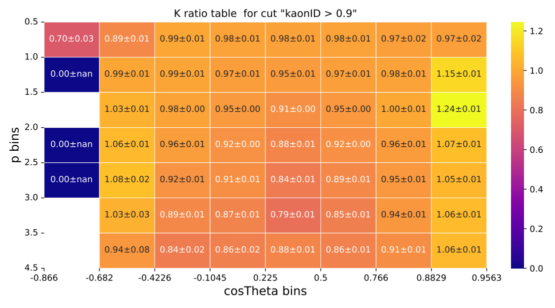
Ask for:

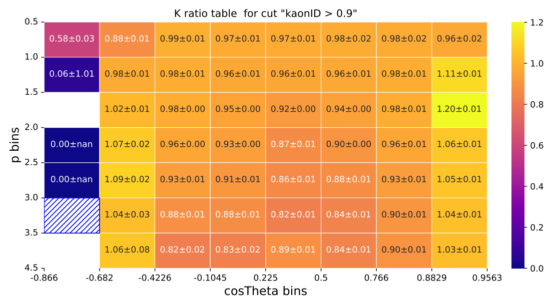
- No extra good track in the event
- No extra K_S, Λ or π^0 in the event

6 Supplementary Material

- Preselection cuts
- PID corrections
- Selection
- EExtra corrections
- Shapley values
- Backgrounds

PID corrections - Kaon

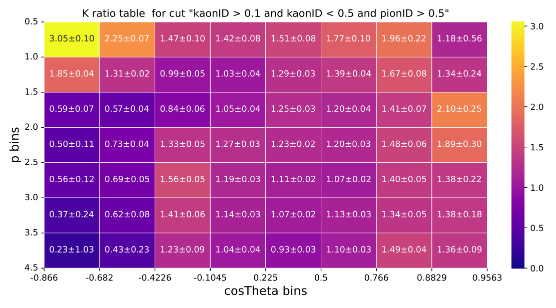


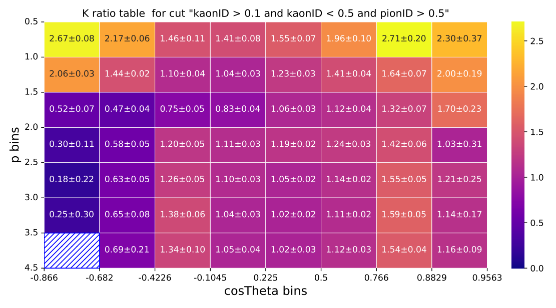


$\longrightarrow$ Cut: $\cos(\theta_K) > -0.682$

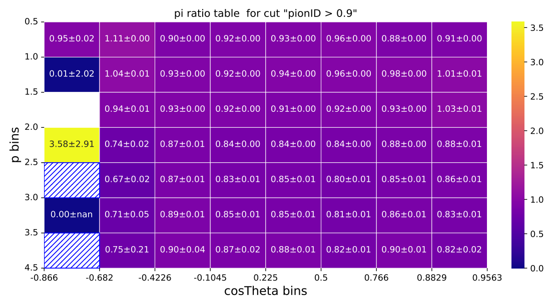
$\longrightarrow$ Efficiency of 99.9%

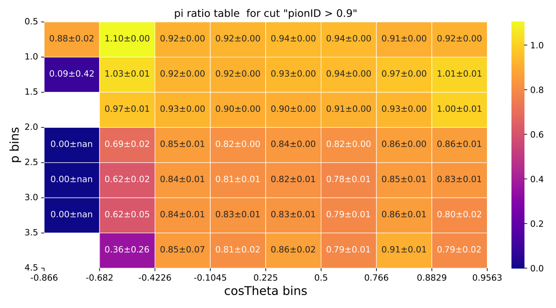
PID corrections - Kaon sideband



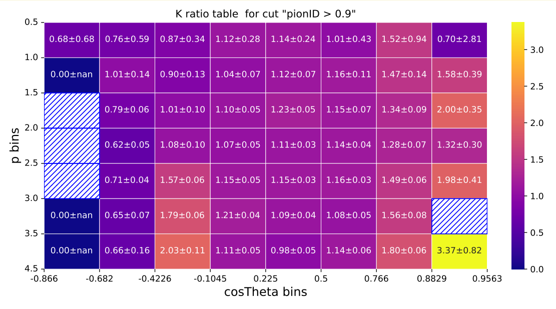


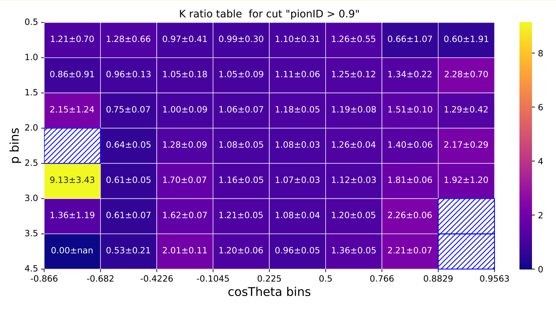
PID corrections - Pion sideband



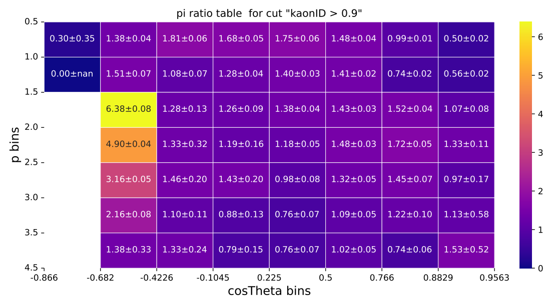


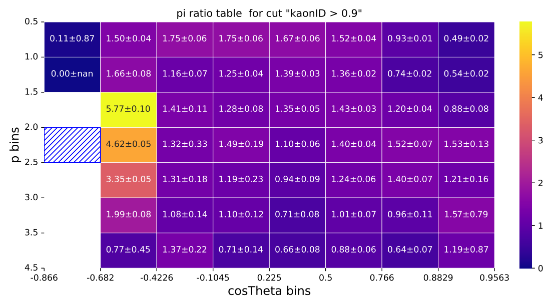
PID corrections - Pion sideband



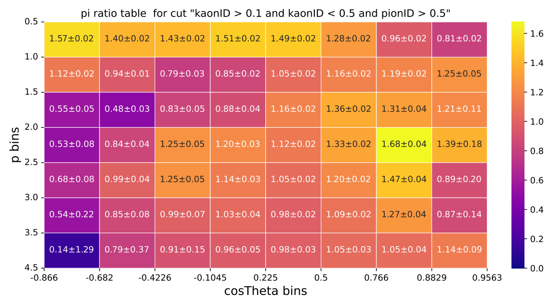


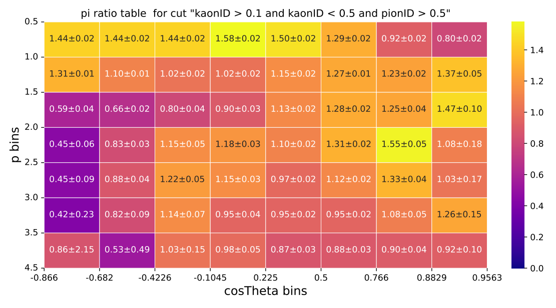
PID corrections - Pion





PID corrections - Pion





6 Supplementary Material

- Preselection cuts
- PID corrections
- Selection
- EExtra corrections
- Shapley values
- Backgrounds

Continuum suppression pre-BDT

Summary:

Cut	Efficiency after Signal probability > 0.01	Purity [%]	Sigma minuit [$\times 10^{-5}$]
Original cuts	74.2	12	0.4
$-1.75 < \cos(\theta_{BY}) < 1.1$	66.1	14	0.4
$-1.9 < \cos(\theta_{BY}) < 1.2$	65.6	13	0.4
nRemainingTracks	58.4	14	0.5
R2	53.1	13	0.4
Signal probability > 0.01	65.0	14	0.4
Signal probability > 0.1	26.5	17	0.6

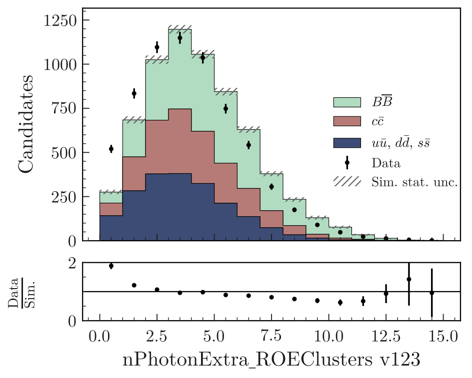
Table: Table summarizing the cuts applied and their consequences on the various metrics

6 Supplementary Material

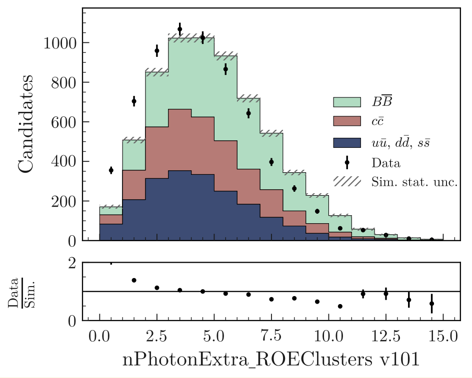
EExtra corrections (in pion-enriched sideband, before corrections)

Computed by reweighting the MC to the data on nPhotons in the wrong charged control sample.

Using BDT variables:



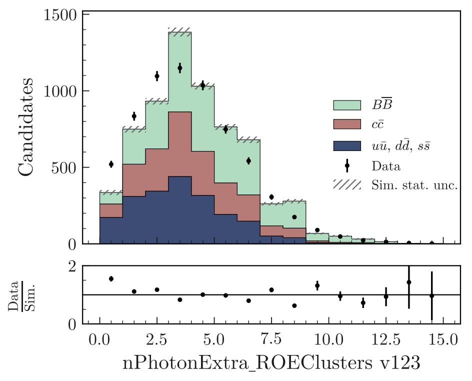
Using minC2TDist:



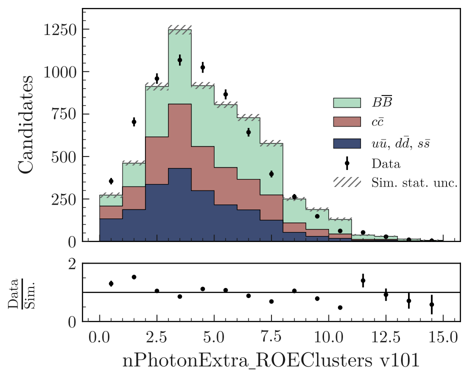
EExtra corrections (in pion-enriched sideband, after corrections)

Computed by reweighting the MC to the data on nPhotons in the wrong charged control sample.

Using BDT variables:



Using minC2TDist:

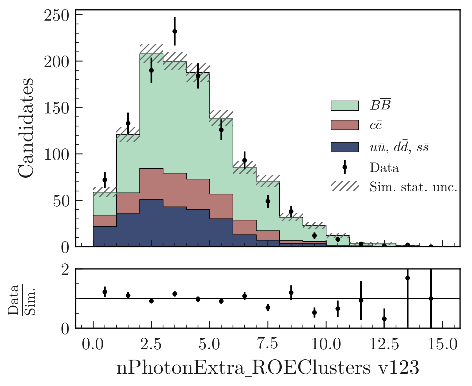


$\longrightarrow$ Need to study the wrong charged control sample in details

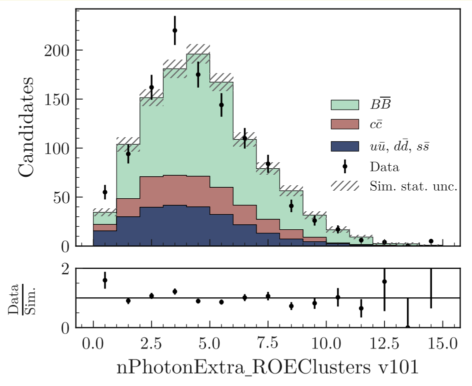
EExtra corrections (in wrong charged pion-enriched sideband before corrections)

Computed by reweighting the MC to the data on nPhotons in the wrong charged control sample.

Using BDT variables:



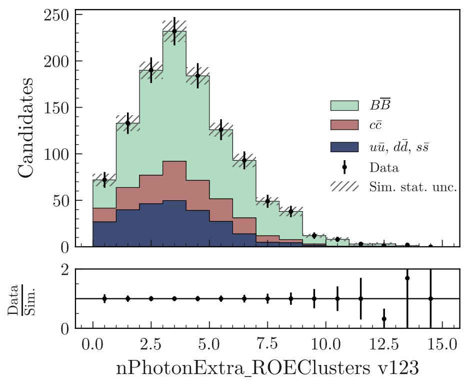
Using minC2TDist:



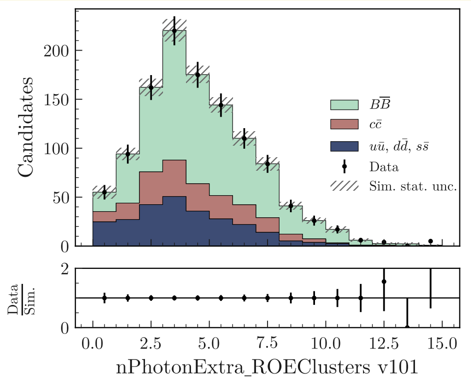
EExtra corrections (in wrong charged pion-enriched sideband after corrections)

Computed by reweighting the MC to the data on nPhotons in the wrong charged control sample.

Using BDT variables:



Using minC2TDist:



6 Supplementary Material

Shapley values - Explanation

- Method to attribute the prediction of a model to its features
- Originates from cooperative game theory:
  - Imagine a set N (of n players) and a gain function v that assigns a value to each coalition S
  - Question: How to fairly distribute the total gain among the players?
  - Answer: Shapley values

$$\varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

Shapley values - Example

```python
import xgboost as xgb
import shap as sh

# ------ XGBoost Classifier ------ #
bdt = xgb.XGBClassifier(**param)
bdt.fit(X_train, y_train, sample_weight=weights_train)

# ------ Shapley Values ------ #
explainer_xgb = sh.TreeExplainer(bdt)
explanation = explainer_xgb(X_test)

# ------ SHAP Interpreter Plot ------ #
sh.plots.beeswarm(explanation,max_display=len(branches))
```

Listing: Example Python Code

6 Supplementary Material

Backgrounds

Channel 1 sl charged evt type	occurence (398 evts)
misID K_{sig}	3.77% (15)
at least 1 K_L	24.62% (98)
$D\ell\nu + D\ell\nu$	46.74% (179)
$D\ell\nu + KK^0K^0$	9.92% (38)
$D\ell\nu + Hadrons$	9.66% (37)
$D\ell\nu + c\overline{c}$	9.40% (36)
$D\ell\nu + DHadrons$	5.74% (22)
$Dn\pi + D\ell\nu$	5.48% (21)
$D\ell\nu + DD$	2.09% (8)
$Dn\pi + DHadrons$	1.04% (4)
$D\ell\nu + D\tau\nu$	1.04% (4)
$Dn\pi + c\overline{c}$	0.78% (3)
$Dn\pi + Dn\pi$	0.26% (1)
$Dn\pi + D\tau\nu$	0.26% (1)
$KK^0K^0 + DD$	0.26% (1)
$DHadrons + DHadrons$	0.26% (1)
other	6.78% (27)

[1] Belle II Collaboration, *Evidence for $B^+ \to K^+ \nu\bar{\nu}$ decays*, (2024), arXiv:2311.14647.

[2] *B^0-$\bar{B}^0$ Mixing*, https://pdg.lbl.gov/2024/reviews/rpp2024-rev-b-bar-mixing.pdf.