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

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

IPHC,
CNRS,
University of Strasbourg,

17 September 2026

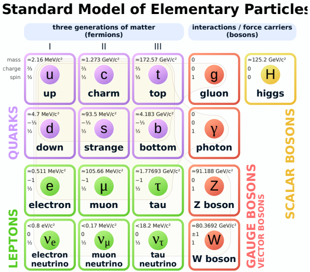
Outline

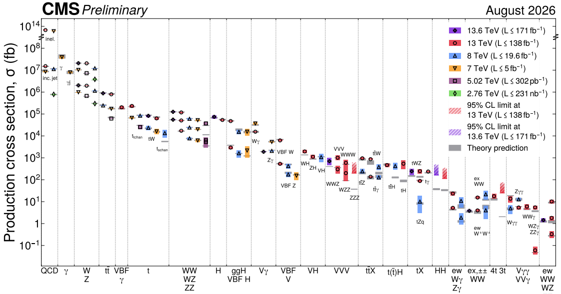
Outline

The standard model of particle physics

Standard Model (SM) of particle physics:

- Mathematical framework
- Describes elementary particles and their interactions
- Predicts interaction probabilities (cross-section, **branching fraction/ratio**, . . .)
- Successful in its predictions

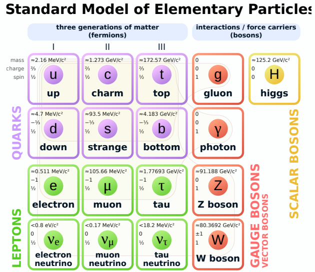




The standard model of particle physics

Standard Model (SM) of particle physics:

- Mathematical framework
- Describes elementary particles and their interactions
- Predicts interaction probabilities (cross-section, **branching fraction/ratio**, . . .)
- Successful in its predictions

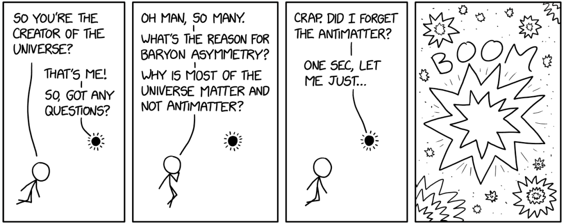


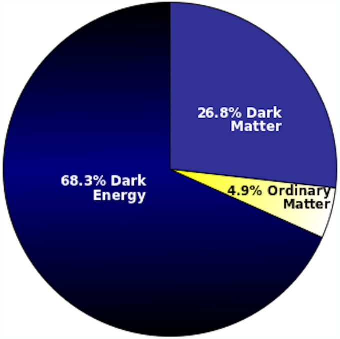
electron g factor g_{e^-}	
Numerical value	$-2.002\ 319\ 304\ 360\ 92$
Standard uncertainty	$0.000\ 000\ 000\ 000\ 36$
Relative standard uncertainty	1.8×10^{-13}
Concise form	$-2.002\ 319\ 304\ 360\ 92(36)$

Beyond the standard model of particle physics

SM cannot explain *e.g.*:

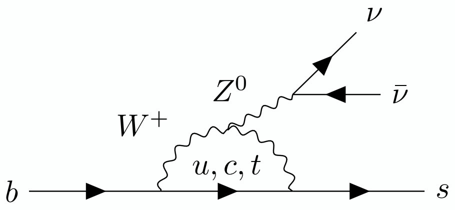
- Matter/anti-matter asymmetry of the universe
- Dark matter/energy



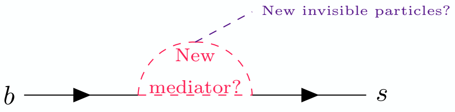


$b \to s\nu\bar{\nu}$ transitions

- Sensitive to:
  - **High energy new physics** contributions [Bečirević, Piazza, Sumensari, 2023]
  - **Light new particles** [Bolton, Fajfer, Kamenik, Novoa-Brunet, 2025]
- $b \to s$ transition = **Flavour changing neutral current**
- **Rare decay** (suppressed by Glashow–Iliopoulos–Maiani mechanism): Branching ratio $\leq \mathcal{O}\left(10^{-5}\right)$
- **Well described** by the Standard Model (SM)

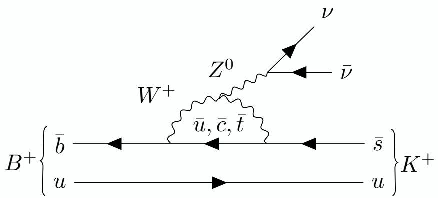


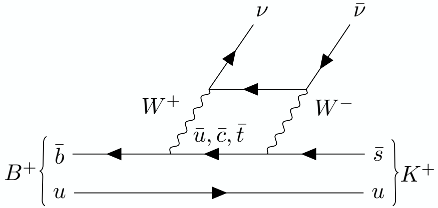
+



Many possible $b \longrightarrow s\nu\bar{\nu}$ decays:

- $B^{\pm} \longrightarrow K^{\pm}\nu\bar{\nu}$
- $B^{0} \longrightarrow K^{0}_{S}\nu\bar{\nu}$
- $B^{\pm} \longrightarrow K^{*\pm}\nu\bar{\nu}$
- $B^{0} \longrightarrow K^{*0}\nu\bar{\nu}$
- ...

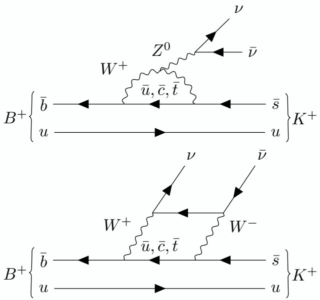




$\longrightarrow$ Focus on the $B^{\pm} \longrightarrow K^{\pm}\nu\bar{\nu}$ decay channel

The $B \to K^{(*)}\nu\bar{\nu}$ decays - Standard Model

Decay	**SM Prediction** $[\times 10^{-6}]$
$B^{\pm} \to K^{\pm}\nu\bar{\nu}$	4.44 ± 0.30
$B^{0} \to K^{0}_{S}\nu\bar{\nu}$	2.05 ± 0.14
$B^{\pm} \to K^{*\pm}\nu\bar{\nu}$	9.8 ± 1.4
$B^{0} \to K^{*0}\nu\bar{\nu}$	9.05 ± 1.37



- For K modes: 7% precision, mainly due to CKM matrix elements uncertainties
- For K^* modes: 14% precision, mainly due to form factor uncertainties

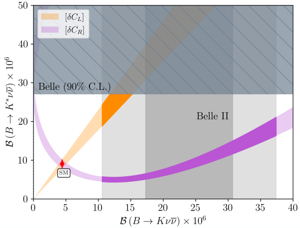
The $B \to K^{(*)}\nu\bar{\nu}$ decays - New physics models

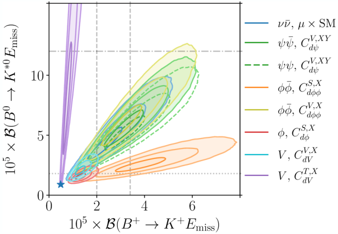
Heavy new particles contributions[Bečirević, Piazza, Sumensari, 2023]:

- Z' mediator
- Leptoquark mediator
- ...

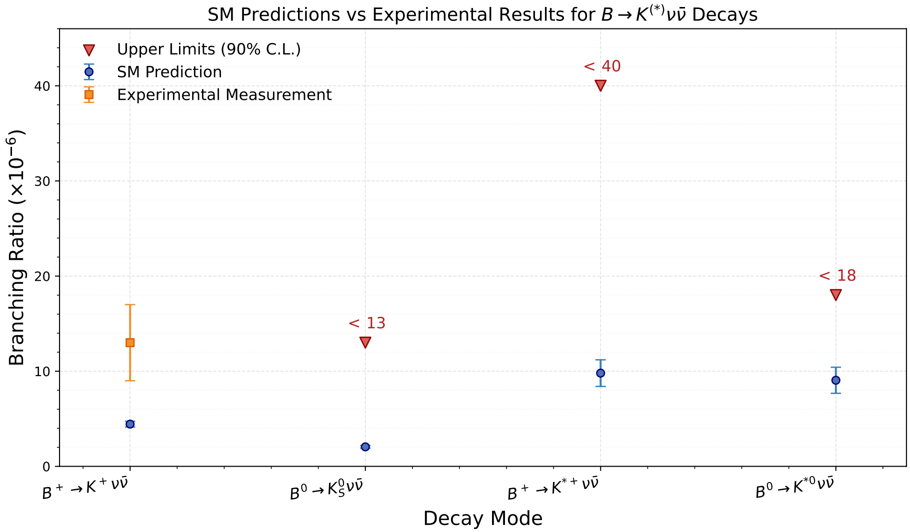
Light new particles contributions[Bolton, Fajfer, Kamenik, Novoa-Brunet, 2025]:

- Axion like particles
- Scalar
- ...



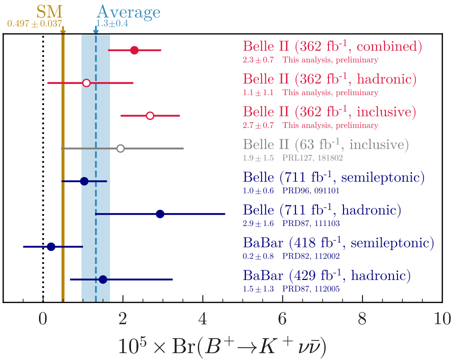


The $B \longrightarrow K^{(*)}\nu\bar{\nu}$ decays - State of the art



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

Three possible strategies: **hadronic (HTA), semileptonic (STA), inclusive (ITA)**



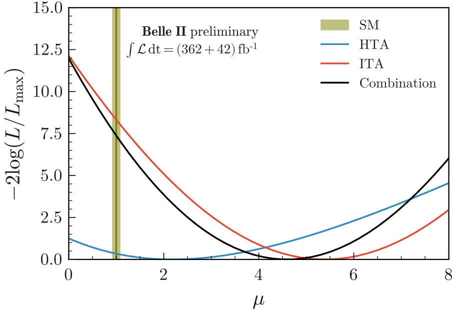


Figure: $\mu = \frac{BF}{BF_{SM}}$ [PhysRevD11.112006]

Goal: Carry an independent measurement using **semileptonic tagging strategy**

Why semileptonic tagging strategy?

- Higher efficiency than hadronic method
- Higher purity than inclusive method
- Complementary to the other two Belle II analyses
- Complementary to the Belle and BaBar analyses

Analysis	**Uncertainty naively scaled to** $487.72\ \mathrm{fb}^{-1}$ $(\times 10^6)$
BaBar HTA	12
BaBar STA	7
Belle HTA	19
Belle STA	7
Belle II ITA	6
Belle II HTA	10
Belle II HTA + ITA	6

Outline

Experimental constraints

- Two neutrinos in a 3-body decay = **Cannot simply reconstruct the signal**

☐ Lack of kinematic constraints $\Longrightarrow$ **Complete angular acceptance**

☐ Only signature is a K $\Longrightarrow$ **Clean environment**

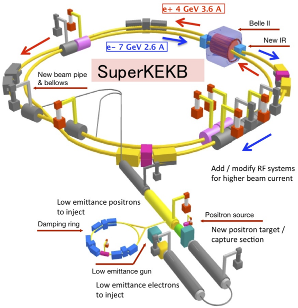
☐ Rare decay $\Longrightarrow$ **High statistics**

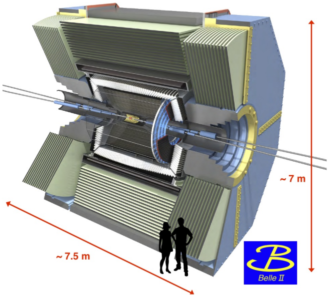
Only experiment capable of measuring the branching ratio of this decay: **Belle II**

Outline

Why Belle II?

- International collaboration
- Taking data on the SuperKEKB accelerator, in Japan, since 2019
- $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ (*B*-factory) $\longrightarrow$ **Clean environment** ☑
- Detector covering a $\approx 4\pi$ solid angle $\longrightarrow$ Almost **complete angular acceptance** ☑
- Integrated luminosity at the ab^{-1} level $\longrightarrow$ **High statistics** ☑

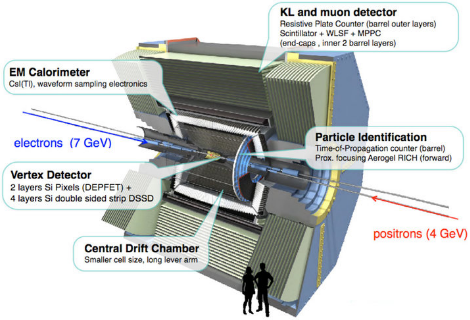




The Belle II detector

Composed of layered sub-detectors, designed for specific tasks

- Tracking:
  - Vertex detector
  - Central Drift Chamber
- Particle Identification (PID):
  - Time Of Propagation
  - Aerogel Ring-Imaging Cherenkov
- Clustering:
  - Electromagnetic Calorimeter (ECL)
- K_L^0 and μ detector

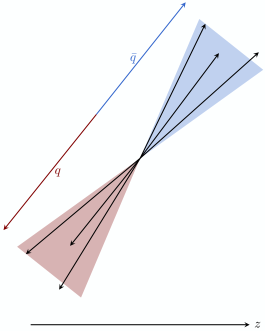


Events geometry

- e^+e^- collision $\Rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ **OR** $q\bar{q}$
- How to discriminate?

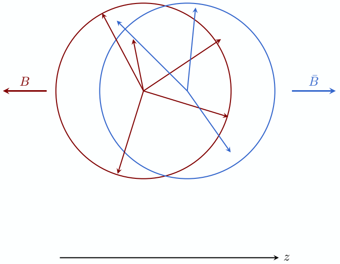
  $\Longrightarrow$ Geometry of the decay

Continuum



$p(q) \approx 5\,\mathrm{GeV}/c$
$e^+e^- \rightarrow q\bar{q} \quad (q \in \{u, d, s, c\})$

B$\bar{\text{B}}$ pair

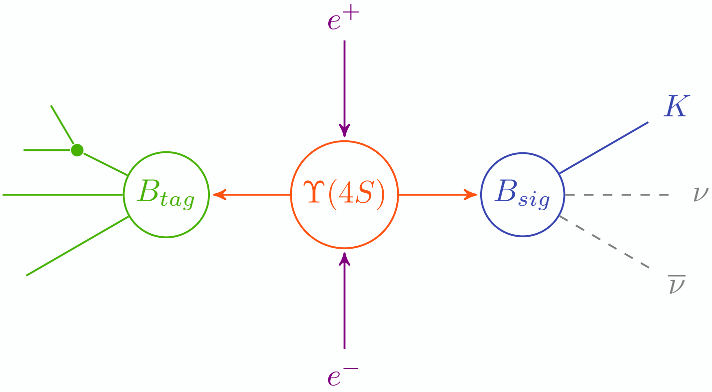


$p(B) \approx 0.3\,\mathrm{GeV}/c$
$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$

Outline

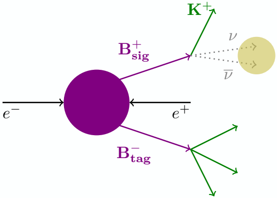
Handling invisible particles

- Only two particles are produced: signal B_{sig} and partner B_{tag}
  $\implies$ Infer information about B_{sig} from B_{tag}

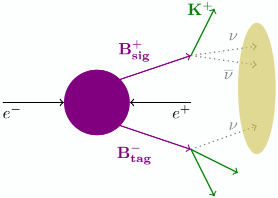


Selecting the signal - Tagging strategies

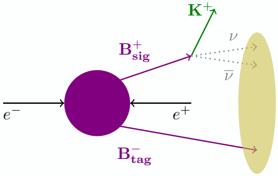
Hadronic Tagged Analysis (HTA)



Semileptonic Tagged Analysis (STA)



Inclusive Tagged Analysis (ITA)



Tagging efficiency $\longrightarrow$

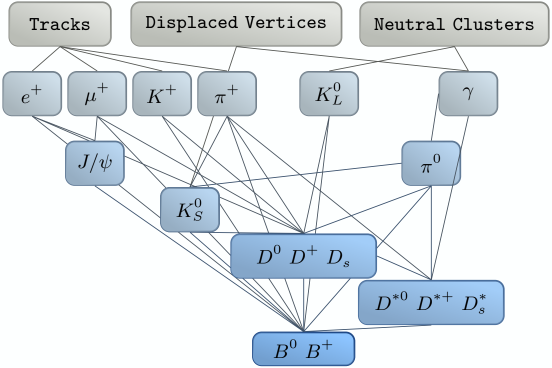
$\mathcal{O}\,(0.1\%)$ $\qquad$ $\mathcal{O}\,(10\%)$

Tagging purity $\longleftarrow$

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

Full Event Interpretation (FEI)

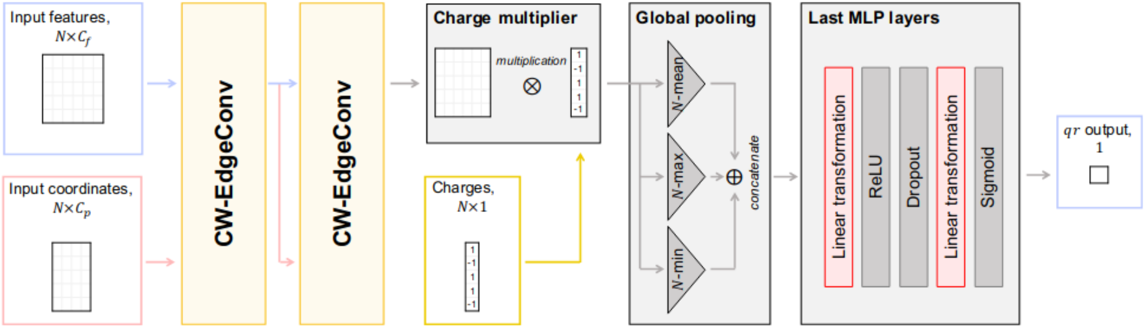
- **Tagging algorithm**: reconstruct the tag side
- Based on **Boosted Decision Trees** (BDT)
- Can be hadronic or **semileptonic**
- For the semileptonic, we consider the following B_{tag} decays:
  - $B \longrightarrow De\nu$
  - $B \longrightarrow D\mu\nu$
  - $B \longrightarrow D^{*}e\nu$
  - $B \longrightarrow D^{*}\mu\nu$



Outline

CP-violation studies

- Need to know the *flavour* of the B meson at the time of its decay (B^0 or $\bar{B}^0$)
  $\Longrightarrow$ **Flavour identification** algorithm
- Role of the `GFlaT`, a graph neural network



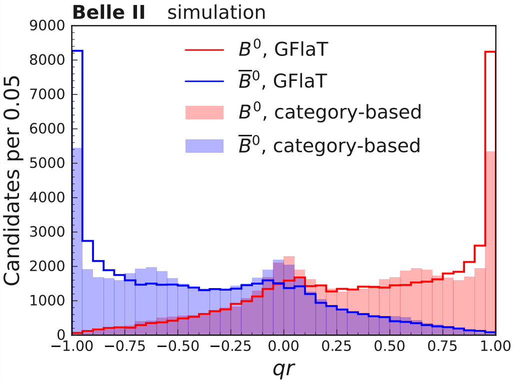
Parameters

- `GFlaT` characterised by its **identification efficiency**:

$$\varepsilon = \frac{N^{\mathrm{id}}_{B^0} + N^{\mathrm{id}}_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}}$$

- Misidentification quantified by the **misidentification fraction**, w:

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



- Work with the **effective identification efficiency** $\varepsilon_{eff}(\varepsilon, w)$

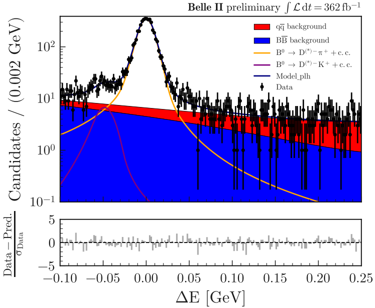
Calibration and optimisation

Goal:

- Implement a calibration strategy for the `GFlaT` in a *time-integrated* way
- Automatise and optimise the calibration process

My work:

- Parallelised the fits necessary for the calibration
- Transformed the independent fits into a simultaneous fit
- Reduced the continuum background
- Automatised the workflow using B2LUIGI



Results:

- Effective identification efficiency on data: $\varepsilon = 37.3 \pm 1.0\%$

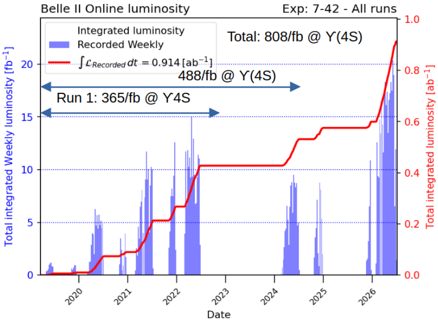
Outline

Analysis overview

- Analysis is **blinded**: not looking at signal in data sample $\longrightarrow$ **Monte-Carlo (MC) simulation**

  $\Rightarrow$ Need to control data/simulation discrepancies (**control samples**)

- **Dataset:** 488fb^{-1} at the $\Upsilon(4S)$ resonance
- **Simulated samples:**
  - $B\bar{B}$ and $q\bar{q}$: $4\times$ dataset luminosity each
  - $B^{\pm} \to K^{\pm}\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, $B^{\pm} \to K^{*\pm}\nu\bar{\nu}$, $B^0 \to K^{*0}\nu\bar{\nu}$ channels: 10×10^6 events each

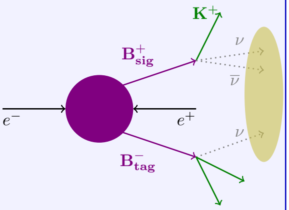


- **Important variable:** q^2, squared invariant mass of the $\nu\bar{\nu}$ system

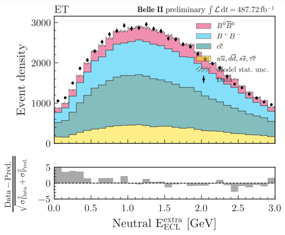
$$q^2 = (E_\nu + E_{\bar{\nu}})^2 - (\vec{p}_\nu + \vec{p}_{\bar{\nu}})^2$$

Analysis flow overview

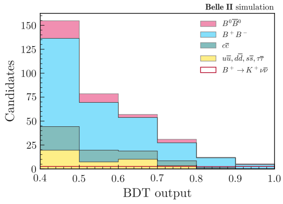
1. Reconstruction and preselection:



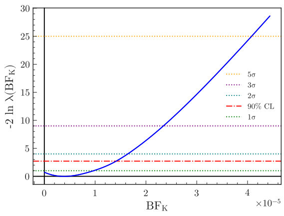
2. Data/simulation corrections:



3. Background suppression:



4. Signal extraction:



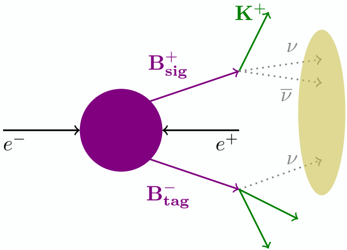
Focus on $B^\pm \to K^\pm \nu\bar{\nu}$

Outline

Reconstruction and preselection

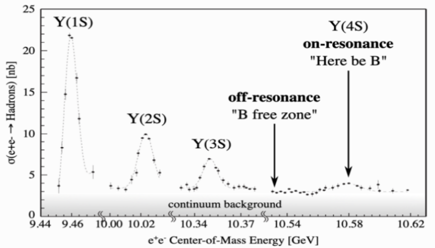
Overview:

1. Reconstruct the B_{tag} using FEI
2. Reconstruct the signal-side K
3. Combine the corresponding tracks
4. Remove events with extra "*good*" tracks or remaining π^0, Λ, K_S^0



Control samples - backgrounds

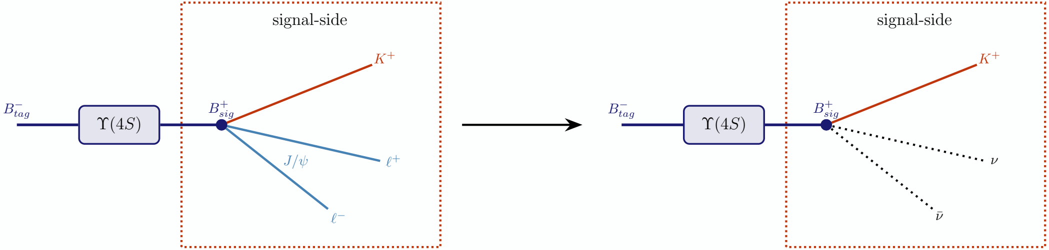
- For $q\bar{q}$: **off-resonance** sample
  - 60 MeV below the $\Upsilon(4S)$ resonance
  - 59 fb^{-1} of data, 4× for simulation



- For $B\bar{B}$: **wrong charge** and **extra tracks** samples
  - **Wrong charge**: the two B mesons must have the same charge
  - **Extra tracks**: at least one extra "*acceptable*" track in the rest of event

Control samples - signal

- For signal: **embedding**
  1. Choose a channel with good kinematics knowledge and a K *e.g.* $B^+ \longrightarrow K^+ J/\psi\,(\longrightarrow \ell^+\ell^-)$
  2. Reconstruct in simulation and data
  3. Remove the signal-side ($B^+ \longrightarrow K^+ J/\psi$) from reconstructed samples
  4. Replace with our simulated signal $B^\pm \longrightarrow K^\pm \nu\bar{\nu}$



Outline

Corrections overview

Simulation used for:

- Training of the BDT for background suppression
- **Create the template for signal extraction**

$\Longrightarrow$ Corrections of the simulation are **necessary**

Corrections	**Method**	**Applied to**
PID	Syscorr framework	All samples
FEI	Calibration from $B \to D^* \ell \nu$	Signal and $B\bar{B}$
Signal Form Factors	Correct from latest theory paper	Signal
Backgrounds	Correct branching fractions to PDG value	$B\bar{B}$
Specific backgrounds	Correct q^2 distribution ($B \to KK_L^0 K_L^0$ and $B \to Kn\bar{n}$)	$B\bar{B}$
$q\bar{q}$ normalisation	Correct $\frac{data}{MC}$ from off-resonance sideband	$q\bar{q}$
Badly-tagged $B\bar{B}$ normalisation	Correct $\frac{data}{MC}$ from extra-tracks sideband	$B\bar{B}$
Backgrounds shape corrections	Train BDTs to separate data and MC in sidebands	$B\bar{B}$ and $q\bar{q}$

Special backgrounds - $B^{\pm} \longrightarrow K^{\pm} n\bar{n}$ and $B^{\pm} \longrightarrow K^{\pm} K_L^0 K_L^0$

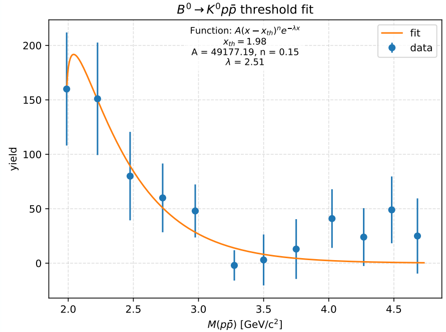
- Some modes can easily mimic our signal, *e.g.* $B^{\pm} \longrightarrow K^{\pm} n\bar{n}$ and $B^{\pm} \longrightarrow K^{\pm} K_L^0 K_L^0$

  $\Longrightarrow$ Need proper q^2 description $\longrightarrow$ correct q^2 distribution
- No measurement for $B^{\pm} \longrightarrow K^{\pm} n\bar{n}$ and $B^{\pm} \longrightarrow K^{\pm} K_L^0 K_L^0$

  $\Longrightarrow$ Use isospin partners

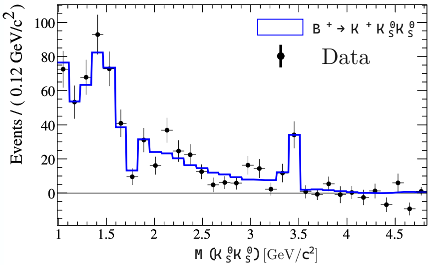
$B^\pm \to K^\pm n\bar{n}$

- $B^0 \to K^0 p\bar{p}$: measured by BaBar [PhysRevD76.092004]
- Model: $A\,(x - x_{th})^n\, e^{-\lambda x}$



$B^\pm \to K^\pm K^0_L K^0_L$

- $B^\pm \to K^\pm K^0_S K^0_S$: measured by BaBar [PhysRevD85.112010]
- Reweight to obtain a similar binned q^2 distribution as BaBar

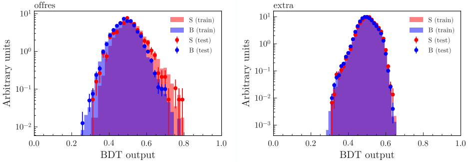


Backgrounds shape corrections

- **Goal**: correct the shape of the variables which will be used for the background suppression
  $\Longrightarrow$ **Issue:** 18 variables = 18 weights
- **Idea**: use a **multi-variate** algorithm (18 inputs $\longrightarrow$ 1 output)
- **Method**: train a BDT to separate data and simulation
  - BDT output = $P\,(\text{data})$ = probability to be a data event
  - $1 - P\,(\text{data}) = P\,(\text{simulation})$ = probability to be a simulation event
  - Weight:

$$w_{BDT_{sb}} = \frac{P\,(\text{data})}{P\,(\text{simulation})}$$

- Train two BDTs: one on off-resonance (for $q\bar{q}$) sample **and** one on extra tracks sample (for $B\bar{B}$)

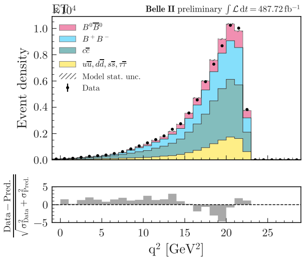


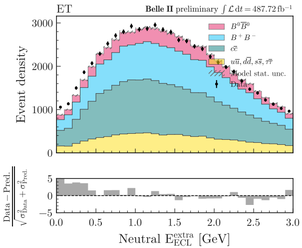
Other corrections

- **PID and FEI**: calibration factors from the collaboration
- **Signal form factors**: reweight q^2 distribution
- **Backgrounds**: compute weight to update branching ratio to state of the art
- $q\bar{q}$ **normalisation**: compute weight from off-resonance sample to unity
- **Badly-tagged** $B\bar{B}$ **normalisation**: compute weight from badly-tagged extra-tracks sample

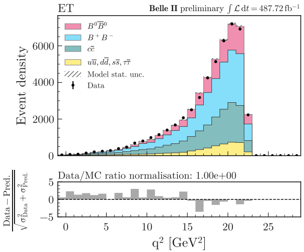
Corrections validation

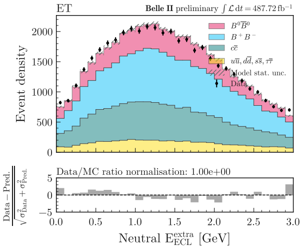
Before corrections





After corrections

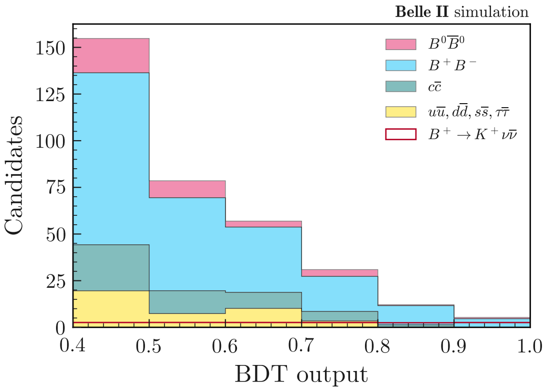




Outline

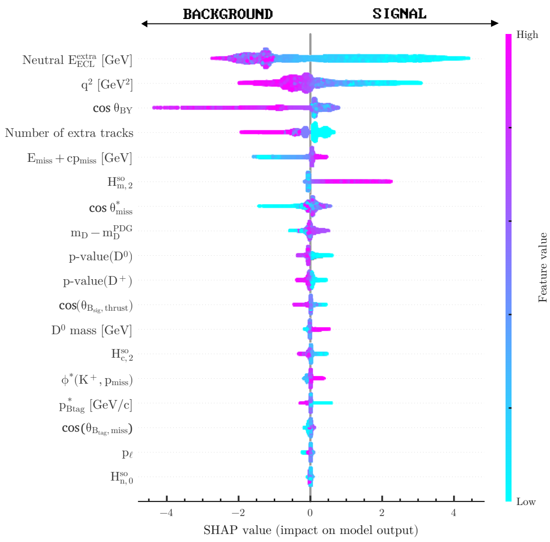
Selection

- Input: 18 variables
- Chosen using Shapley values for feature importance
- **Signal region**:
  - BDT output > 0.4
  - Signal efficiency: 0.7%
  - Purity: 6%



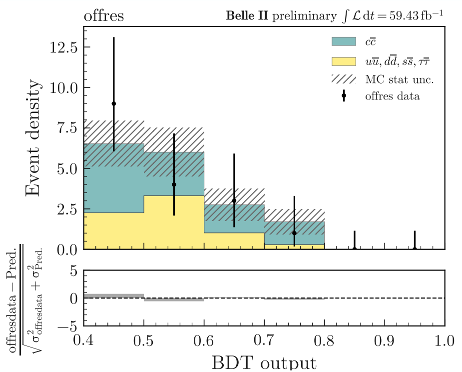
BDT input variables

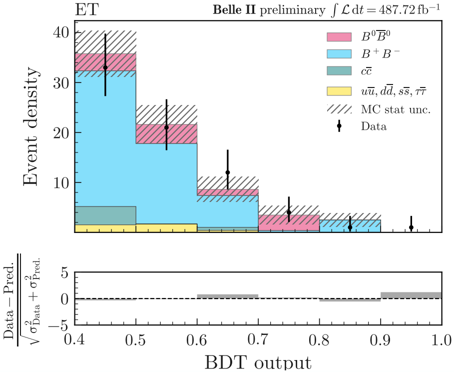
- The color gradient indicates the value of the feature
- The x-axis is the **Shapley value**
  - Shapley value > 0 = contributes to the "signal" prediction
  - Shapley value < 0 = contributes to the "background" prediction



Validation

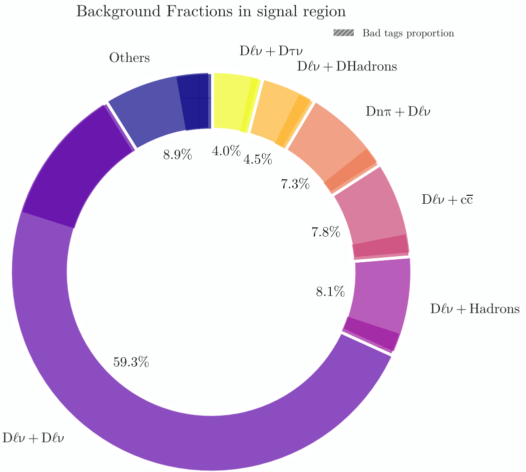
Sideband	data/MC signal region	Valid? ($\leq 2\sigma$)
offres	1.25 ± 0.35 (18 data evts)	✓
wc	1.18 ± 0.24 (39 data evts)	✓
extra	1.32 ± 0.20 (78 data evts)	✓





Outline

Backgrounds

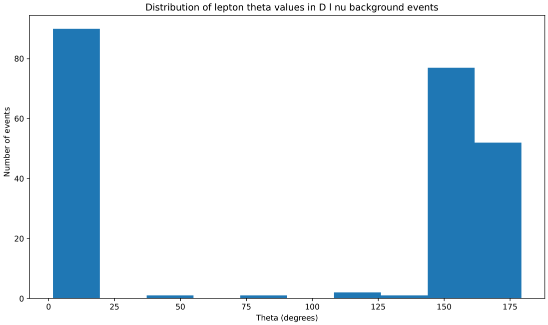


$B \longrightarrow D\ell\nu$ studies

- **Most important decay:** $B \longrightarrow D\ell\nu$
  - For 488fb^{-1} $\approx 20 \times 10^6$ events (≈ 1171 for $B^\pm \longrightarrow K^\pm \nu\bar{\nu}$)
  - How does it mimic our signal $B^\pm \longrightarrow K^\pm \nu\bar{\nu}$?
  - Can we reduce its contribution?
- Full decay is: $B^\pm \longrightarrow D^{(*)0}(\longrightarrow K^\pm X^\mp)\ell^\pm\nu$

  $\Rightarrow$ X^- and ℓ are lost/not reconstructed

$\Longrightarrow$ It is deemed as **irreducible**



Outline

Statistical model - context

- **Goal:** Extract the branching fraction $\mathcal{B}\left(B^{\pm} \to K^{\pm}\nu\bar{\nu}\right)$

  $\Rightarrow$ We work with the **signal strength** μ

$$\mu = \frac{\mathcal{B}\left(B^{\pm} \to K^{\pm}\nu\bar{\nu}\right)}{\mathcal{B}\left(B^{\pm} \to K^{\pm}\nu\bar{\nu}\right)_{SM}}$$

- **Contributions:**
  - **Signal:** $B^{\pm} \to K^{\pm}\nu\bar{\nu}$
  - **Generic backgrounds:** $B^{+}B^{-}$, $B^{0}\bar{B}^{0}$, light-$q\bar{q}$ $(u\bar{u}, d\bar{d}, s\bar{s}, \tau\bar{\tau})$, $c\bar{c}$
  - **Crossfeeds:** $B^{0} \to K_{S}^{0}\nu\bar{\nu}$, $B^{\pm} \to K^{*\pm}\nu\bar{\nu}$, $B^{0} \to K^{*0}\nu\bar{\nu}$
- 8 contributions

Statistical model - framework and strategy

- Create histograms in signal region (BDT > 0.4) with $b = 6$ bins
- Run a maximum likelihood fit

Definition (HistFactory framework)

$$\mathcal{L}(\mu, \theta|\mathbf{n}) = \frac{1}{Z}\prod_b \mathrm{Poisson}(n_b|\nu_b(\mu, \theta))\,\pi(\theta)$$

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 normalisation constant
- π is a prior acting as a constraint on the nuisance parameters

- Compute the profile likelihood ratio with the signal strength μ as the parameter of interest
- Get the expected uncertainty on branching fraction and expected upper limits

Statistical model - systematic uncertainties

Source	Contribution(s)	Treatment	Number of nuisance parameters
PID	All samples	Bin-correlated	22
Signal Form Factors	Signal	Bin-correlated	3
Continuum	$q\bar{q}$	Bin-correlated	1
Backgrounds	$B\bar{B}$	Bin-correlated	77
Specific backgrounds	$B\bar{B}$	Bin-correlated	5
Sgn eff (incl. FEI)	Signal	Normalisation	1
$q\bar{q}$ normalisation	$q\bar{q}$	Normalisation	2
$B\bar{B}$ normalisation	$B\bar{B}$	Normalisation	2
Crossfeed normalisation	$B\bar{B}$	Normalisation	3
B meson counting	Signal, $B\bar{B}$ and crossfeed	Contribution-correlated	1
MC statistical	All samples	Uncorrelated	8×6

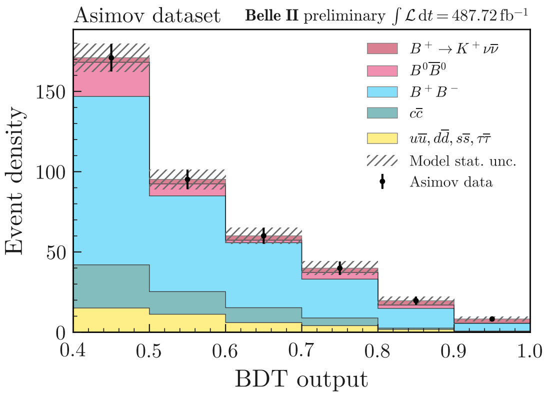
Statistical model - uncertainties contributions

- Based on [Pinto,Wu,2023]
- For **statistical uncertainty**:
  - For each bin, vary the content n_b by $\pm\sqrt{n_b}$
  - Repeat the fit leaving everything free $\longrightarrow$ get a new μ_{var}
  - Compute $\Delta\mu = \mu_{var} - \mu_{nom}$
  - Sum in quadrature to get the full statistical uncertainty
- For **systematic uncertainty**:
  - For nuisance parameter, vary the value a_i by $\pm\sigma_i$
  - Repeat the fit leaving everything free $\longrightarrow$ get a new μ_{var}
  - Compute $\Delta\mu = \mu_{var} - \mu_{nom}$
  - Sum in quadrature to get the full systematic uncertainty

Outline

Asimov fit - results

- **Asimov dataset:** observed data = expected data
- Allows us to
  - Get an estimate of the correlations
  - Get an estimate of the uncertainties



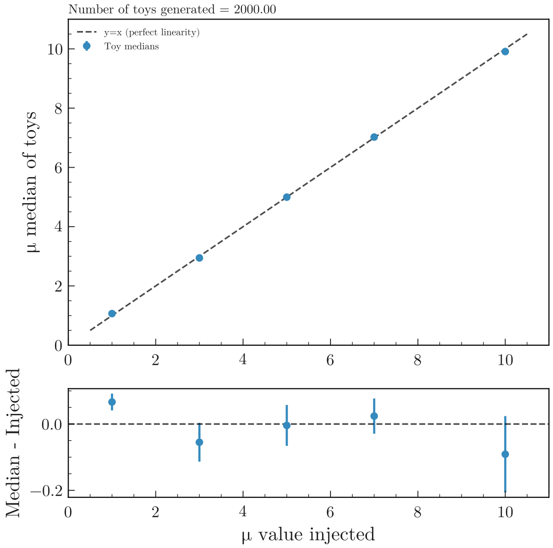
Asimov fit - uncertainties decomposition

- **Statistically dominated**
- Most important systematics:
  - Limited simulation statistics
    - $\longrightarrow$ More simulation
  - K_L^0 reconstruction efficiency
  - Continuum shape correction
    - $\longrightarrow$ Could improve with a better optimisation of the BDT

Systematic Category	Absolute contribution on branching fraction ($\times 10^{-6}$)
MC statistics	2.2
K_L^0	1.0
Data/MC off-resonance sideband	0.8
Background normalization	0.6
Signal efficiency from embedding	0.5
Backgrounds branching fractions	0.4
$B^\pm \to K^\pm K_L^0 K_L^0$	0.3
$N_{B\bar{B}}$	0.2
Data/MC ET sideband	0.1
Form factors	0.1
PID	0.1
$D \to K_L^0$	0.0
$KX(\to n\bar{n})$	0.0
$B^\pm \to K^\pm n\bar{n}$	0.0
Total systematic	**2.7**
Total statistical	**4.3**
Total	**5.1**

Discussions - closure test

- Test different signal injections to see how the model behaves
- For each signal injection:
  - Generate 2000 toys datasets
  - Perform a fit
  - Compare the median to the injected signal
- Test at various signal strength: $[1, 3, 5, 7, 10]$



Discussions - sensitivity at $\mu = 4.6$

- **Goal:** Compare sensitivity to previous Belle II analysis if same observation
- **Test:** Injection of signal $\mu = 4.6$ ($\mathcal{B}(B^{\pm} \rightarrow K^{\pm}\nu\bar{\nu}) = 2.3 \times 10^{-5}$)
- Uncertainty from previous analysis: 7.0×10^{-6}
- **Uncertainty from this analysis:** $\sigma_{BF} = 7.0 \times 10^{-6}$
  - On par with previous analysis

Systematic Category	**Absolute contribution on branching fraction** ($\times 10^{-6}$)
Signal efficiency from embedding	2.3
MC statistics	2.1
Background normalization	1.7
K_L^0	1.4
Data/MC off-resonance sideband	0.9
$N_{B\bar{B}}$	0.6
Backgrounds branching fractions	0.5
Form factors	0.4
$B^{\pm} \rightarrow K^{\pm}K_L^0K_L^0$	0.3
Data/MC ET sideband	0.2
PID	0.2
$D \rightarrow K_L^0$	0.0
$KX(\rightarrow n\bar{n})$	0.0
$B^{\pm} \rightarrow K^{\pm}n\bar{n}$	0.0
Total systematic	**4.0**
Total statistical	**5.7**
Total	**7.0**

Outline

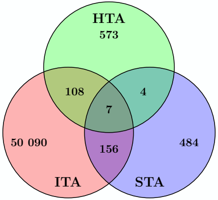
Summary

- The strategy is working and gives encouraging results
- Current sensitivity is on par with state of the art and better than previous occurence of semileptonic tagged analyses

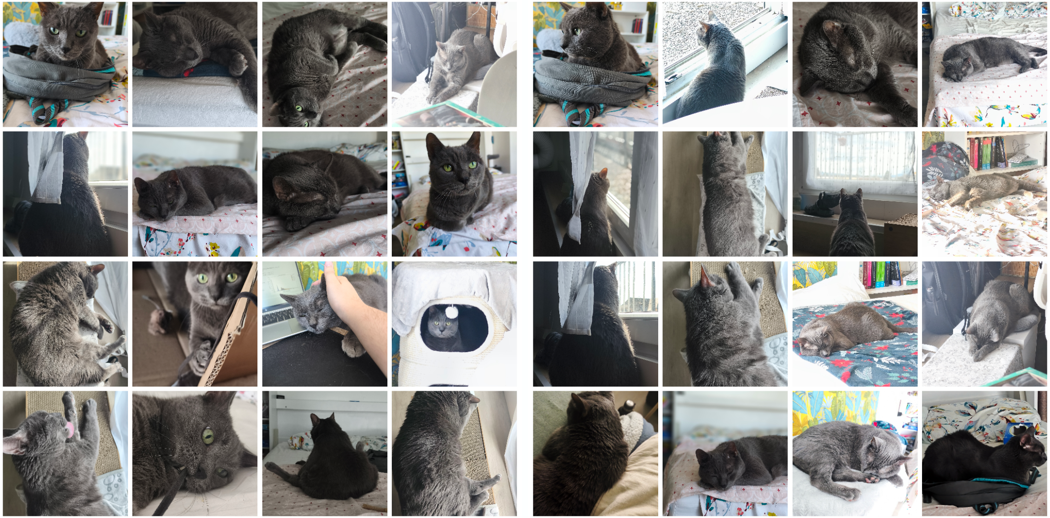
Analysis	**Absolute uncertainty** $(\times 10^6)$	**Naively scaled to 487.72 fb**$^{-1}$ $(\times 10^6)$
BaBar HTA	13	12
BaBar STA	8	7
Belle HTA	16	19
Belle STA	6	7
Belle II ITA	7	6
Belle II HTA	12	10
Belle II HTA + ITA	7	6
This analysis	5	5

Prospects

- **Short-term** prospects:
  - Add other $B^{\pm} \longrightarrow X_{su}\nu\bar{\nu}$ and $B^0 \longrightarrow X_{sd}\nu\bar{\nu}$ as crossfeeds
  - **Unblinding** = results on real data
  - Combined fit with $B^0 \longrightarrow K^{*0}\nu\bar{\nu}$ channel
- **Long-term** prospects:
  - Combination with other analyses on the same channels (inclusive and hadronic tagging)
  - Improvement expected thanks to higher statistics for future analyses



DOUDOU



Outline

Outline

Effective field theory

Effective field theory:

- Approximation of the theory at lower energy scale
- Two ingredients:
  - Operators O (encode lower energy scale)
  - Wilson coefficients C (encode higher energy scale)
- Hamiltonian of the system, energy μ:

$$\mathcal{H}_{eff} = \sum_{i=1}^{N} C_i(\mu)\, O_i(\mu)$$

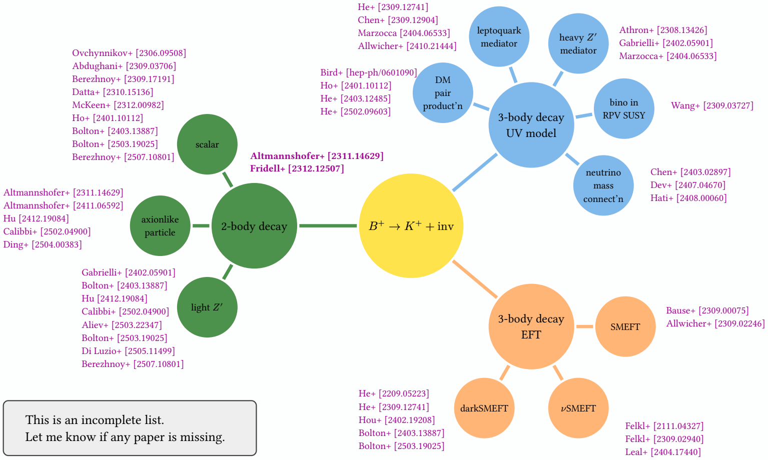
$B \to K^{(*)}\nu\bar{\nu}$ in the SM

- **Important:** choice of $\left|V_{tb}V_{ts}^{*}\right|$

$$|V_{tb}V_{ts}^{*}| \times 10^{3} = \begin{cases} 41.4 \pm 0.8\,, & (B \to X_c\ell\bar{\nu}) \\ 39.3 \pm 1.0\,, & (B \to D\ell\bar{\nu}) \\ 37.8 \pm 0.7\,, & (B \to D^{*}\ell\bar{\nu}) \end{cases}$$

- Choose $B \to D\ell\nu$ as in the paper [Bečirević, Piazza, Sumensari, 2023]
- Using the inclusive value gives branching fraction predictions 1.5σ greater

Proposed new physics explanations



This is an incomplete list.
Let me know if any paper is missing.

Outline

Reconstruction - tracks and clusters

Tracks:

- Charged particles leave *hits* in the VXD and CDC
- Bundle hits together via pattern recognition algorithms
- Extrapolate VXD and CDC tracks together
- Obtain trajectory, momentum and charged via a Kalman filter

Clusters:

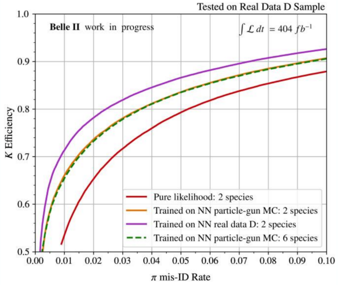
- Neutral particles leave *energy deposits* in the ECL
- Bundle deposits together via nearest neighbour algorithm
- Extrapolate CDC tracks to the reconstructed cluster

	13.				21.	
1.2	34.	1.0		1.0	0.6	0.9
	3.4	1.4	0.6	12.	9.8	1.2
	0.9					
9.5		0.2		93.		1.0
1.0						
		0.4	15.	1.4	0.9	
		0.7	2.1			

Reconstruction - particle identification

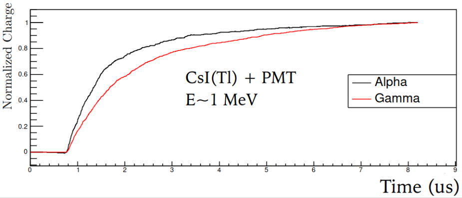
Charged particles identification:

- Each subdetector gives a likelihood for each particle hypotheses
- Likelihoods and momenta are used as inputs for a neural networks
- Gives a probability for each charged particle hypotheses: K, π, e, μ, p, d

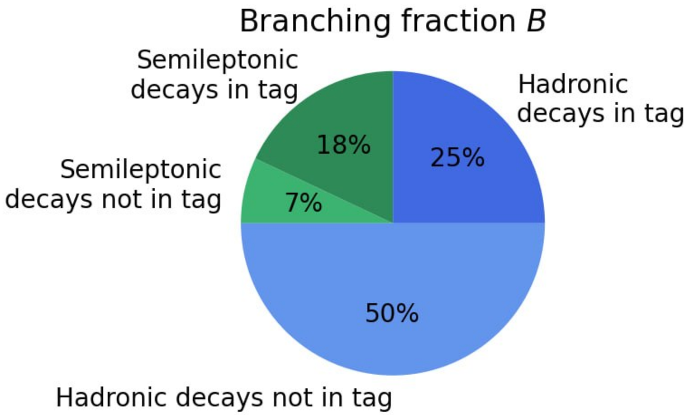


Neutral particles identification:

- Look at shower shapes and use *pulse shape discrimination*



Full Event Interpretation



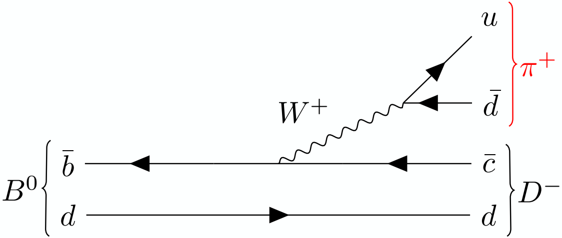
CP-asymmetry

- Want to measure the *CP*-asymmetry:

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

$\longrightarrow$ 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^+
Kinetic Lepton	ℓ^-
Intermediate Kinetic Lepton	ℓ^+
Kaon	K^-
Kaon-Pion	K^-, π^+
Slow Pion	π^+
Fast Hadron	π^+, K^-
Maximum p	ℓ^-, π^-
Fast-Slow-Correlated	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)

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

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

Definition (Wrong tag fraction)

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

Observed vs True CP-asymmetry

- Observed CP-asymmetry written as

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

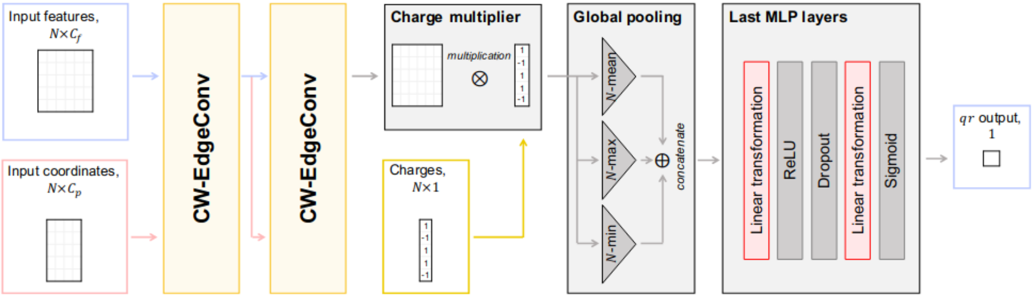
Definition (Dilution factor)

$$r \equiv |1 - 2w| \in [0, 1]$$

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

- 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



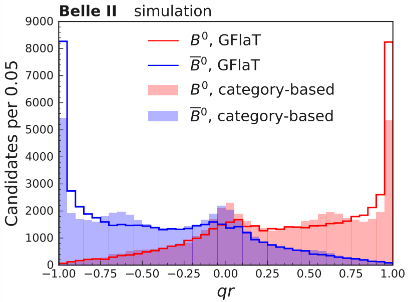
Calibration strategy: ersatz

Goal:

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

Strategy:

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



Calibration strategy: ersatz

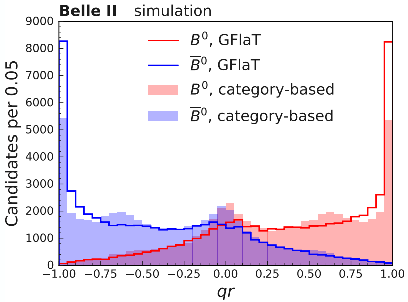
Goal:

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



Calibration strategy: ersatz

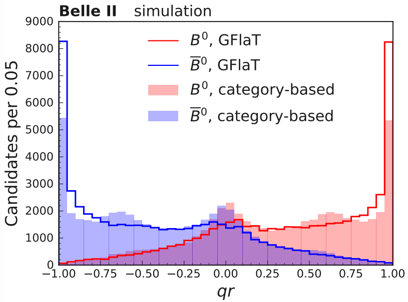
Goal:

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



Calibration strategy: ersatz

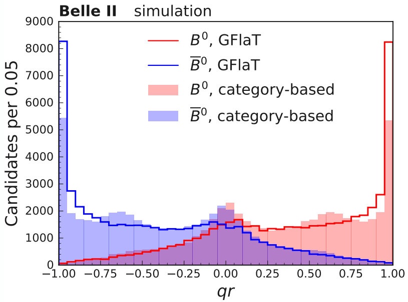
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimise 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 $|qr|$



Calibration strategy: ersatz

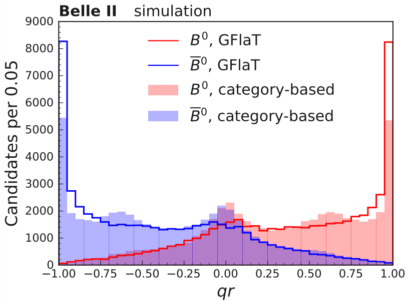
Goal:

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



Calibration strategy: ersatz

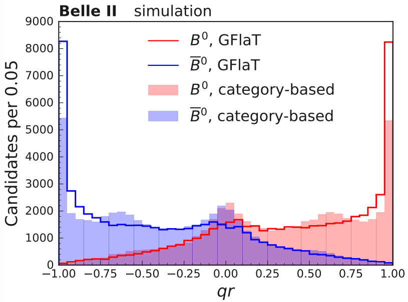
Goal:

- Implement a calibration strategy for the Flavour Tagger in a *time-integrated* way
- Automatize and optimise 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 $\bar{B}^0$



Calibration strategy: ersatz

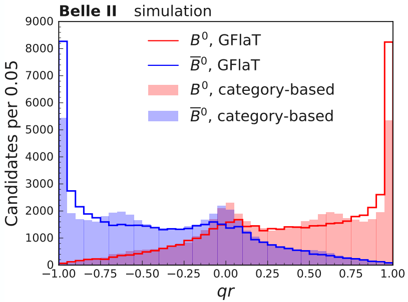
Goal:

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

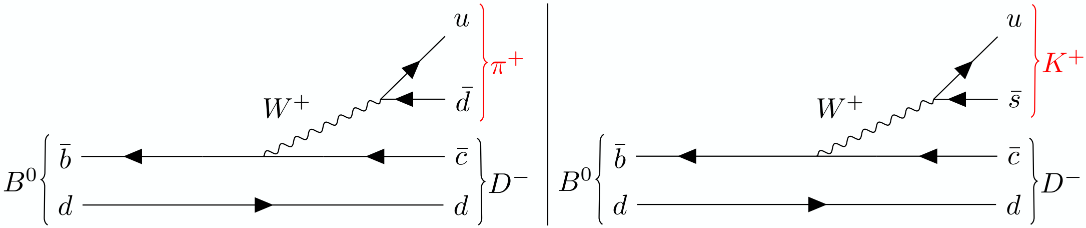


Calibration strategy: signal

- Need self-tagged decays:

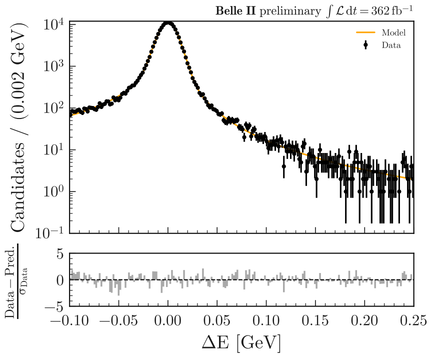
$$B^0 \longrightarrow D^- \pi^+$$
$$B^0 \longrightarrow D^{*-}(\longrightarrow \bar{D}^0(K^+\pi^-)\pi^-)\pi^+$$
$$B^0 \longrightarrow D^{*-}(\longrightarrow \bar{D}^0(K^+\pi^-\pi^0)\pi^-)\pi^+$$
$$B^0 \longrightarrow D^{*-}(\longrightarrow \bar{D}^0(K^+\pi^-\pi^-\pi^+)\pi^-)\pi^+$$

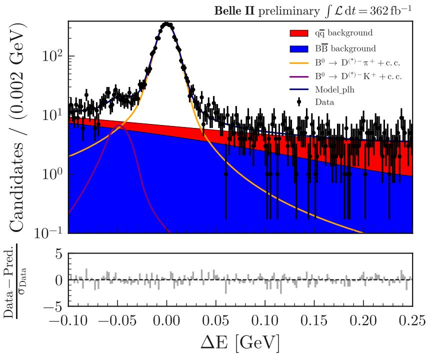
- Also consider $B \longrightarrow D^{(*)-}K^+$ equivalents, called "BDK"s



Calibration strategy: extracting the yields

- Important kinematic variable for signal selection: $\Delta E = E_B - E_{\mathrm{beam}}$
- 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 $i \in [\![0, 6]\!]$

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

with

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

and deduce

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

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

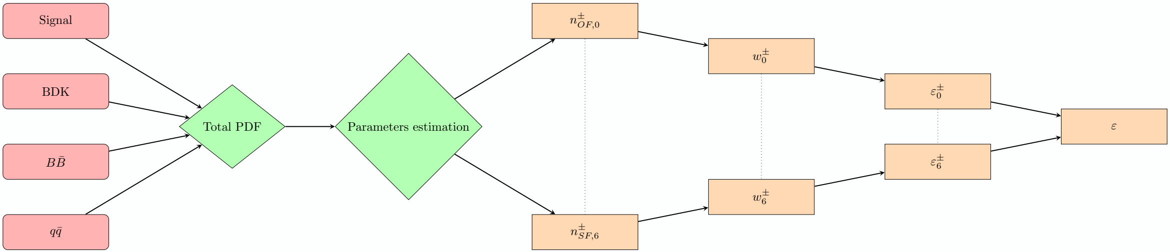
$$\varepsilon_i^{\pm} = \frac{n^{\pm}_{\mathrm{OF},i} + n^{\pm}_{\mathrm{SF},i}}{N^{\mathrm{tag}}_{B^0} + N^{\mathrm{tag}}_{\bar{B}^0}}$$

and compute the *effective* tagging efficiency

Definition (Effective tagging efficiency)

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

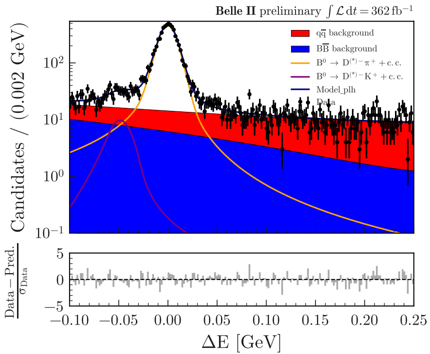
Calibration workflow

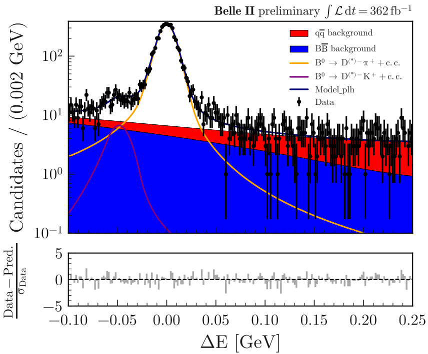


Calibration optimisation: parallelisation and automation

- **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
- 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

- $q\bar{q}$: reduced continuum background
  - Via a cut on a geometric variable ✓
  - Via a BDT (worked on MC but not on data)





Calibration optimisation: summary

- Parallelised the fits of the shapes
- Transformed the independent fits into a simultaneous fit
- Reduced the continnum background
- Automatised the workflow using B2LUIGI

Results:

- Effective tagging efficiency on data: $\varepsilon = 37.3 \pm 1.0\%$

Calibration optimisation: summary

Quantity	MC [%]	Data [%]
w_1^+	47.06 ± 0.78	46.96 ± 1.51
w_1^-	47.98 ± 0.78	48.87 ± 1.47
w_2^+	41.70 ± 0.68	39.36 ± 1.29
w_2^-	41.99 ± 0.68	43.80 ± 1.32
w_3^+	32.54 ± 0.63	32.57 ± 1.20
w_3^-	33.07 ± 0.63	35.52 ± 1.14
w_4^+	22.59 ± 0.72	22.27 ± 1.34
w_4^-	23.93 ± 0.73	28.99 ± 1.40
w_5^+	15.77 ± 0.76	14.12 ± 1.46
w_5^-	17.19 ± 0.78	19.16 ± 1.44
w_6^+	11.71 ± 0.64	11.38 ± 1.19
w_6^-	8.72 ± 0.61	12.62 ± 1.16
w_7^+	3.39 ± 0.41	1.83 ± 0.78
w_7^-	3.24 ± 0.40	4.50 ± 0.81

Outline

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`

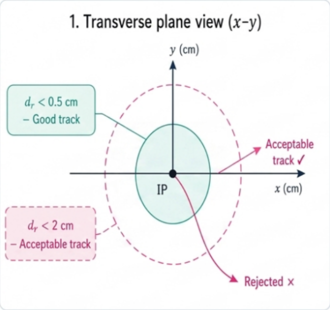
Two masks studied for now:

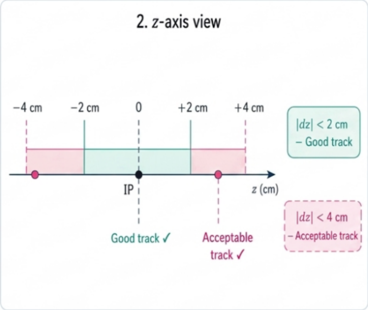
- `minC2TDist>50` cm
- `beamBackgroundSuppression > 0.5` and `fakePhotonSuppression > 0.5`

Ask for:

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

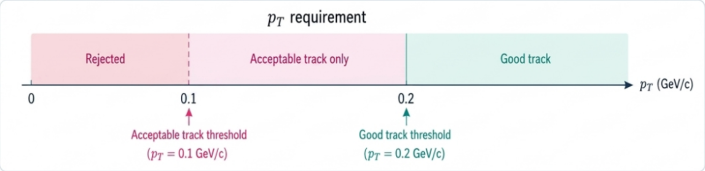
Extra tracks control sample





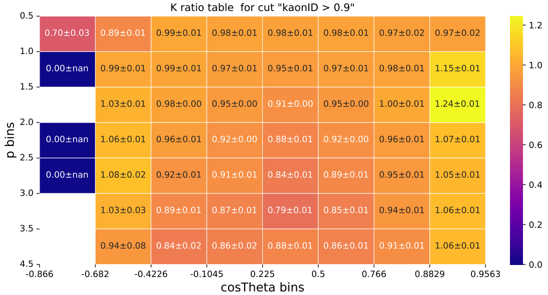
3. Summary of track quality tiers

Requirement	Good track	Acceptable track
d_r	< 0.5 cm	< 2 cm
$\|dz\|$	< 2 cm	< 4 cm
p_T	> 0.2 GeV/c	> 0.1 GeV/c
CDC acceptance	yes	yes
Used for	Signal reconstruction	Sideband definition



FEI and PID corrections

- PID and FEI corrections based on recommendations from collaboration
- PID calibrated with $D^{*+} \longrightarrow D^0\,(\longrightarrow K^-\pi^+)\,\pi^+$
- FEI calibrated with $B^+ \longrightarrow D^{*0}\ell^+\nu_\ell$



Mode	Calibration factor	Uncertainty
$De\nu$	1.14	0.11
$D\mu\nu$	1.10	0.11
$D^*e\nu$	0.87	0.10
$D^*\mu\nu$	0.850	0.098

Branching fractions corrections

- Branching fractions used for simulation are not always up-to-date

$\Longrightarrow$ Compute **weight** $w_{BF} = \frac{BF_{PDG}}{BF_{simulation}}$

- Exceptions:
  - $B \to X_s \nu\bar{\nu}$, $B^\pm \to K^\pm K_L^0 K_L^0$ special backgrounds
  - PYTHIA modes and modes not present in the PDG $w_{BF} = 1$
  - $B^\pm \to K^\pm \pi^+ \pi^-$, $B^\pm \to K^\pm K^\pm K^0$ $w_{BF} = 1$
- **Global correction for not corrected decays:**

$$w_{\text{global BF}} = \frac{1 - \sum_{i \in \text{corrected}} \text{BF}_i^{\text{PDG}}}{\sum_{i \in \text{not corrected}} \text{BF}_i^{\text{simulation}}}$$

  - For B^+B^-: $w_{\text{global BF}} = 1.07$
  - For $B^0\bar{B}^0$: $w_{\text{global BF}} = 1.00$

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

Efficiency vs q^2