UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE

Institut Pluridisciplinaire Hubert Curien (IPHC), UMR 7178

THÈSE présentée par

Corentin SANTOS

soutenue le: **17 Septembre 2026**

pour obtenir le grade de:	**Docteur de l'Université de Strasbourg**
Discipline/Spécialité:	Physique des particules élémentaires

Recherche des désintégrations $B \to K^{(*)}\nu\bar{\nu}$ à l'expérience Belle II

THÈSE dirigée par:

Mme RIPP-BAUDOT Isabelle	Dr, Institut Pluridisciplinaire Hubert Curien

RAPPORTEURS:

M. CHARLES Matthew	Dr, Laboratoire de Physique Nucléaire et Hautes Énergies
M. SUMENSARI Olcyr	Dr, Laboratoire de Physique des 2 infinis, Irène Joliot-Curie

---

AUTRES MEMBRES DU JURY:

M. DUJANY Giulio	Dr, Institut Pluridisciplinaire Hubert Curien
Mme COLLARD Caroline	Dr, Institut Pluridisciplinaire Hubert Curien
Mme SERRANO Justine	Dr, Centre de Physique des Particules de Marseille

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

This thesis details the strategy for the measurement of the $B \to K^{(*)}\nu\bar{\nu}$ decays' branching fraction, focusing on $B^+ \to K^+\nu\bar{\nu}$, at the Belle II experiment, using 487.7 fb^{-1} of data at the $\Upsilon(4S)$ resonance. This experiment operates on an electron-positron collider, SuperKEKB, at KEK, Tsukuba, Japan, which specialises in the production of B mesons. The $B \to K^{(*)}\nu\bar{\nu}$ decays are important probe for physics beyond the Standard Model, as the underlying process is a $b \to s\nu\bar{\nu}$ quark transition, *i.e.* a Flavour Changing Neutral Current. Thanks to the accurate prediction of its branching ratio in the Standard Model, a precise measurement of the decay's branching fraction could reveal tensions and evidence for New Physics. Because of the two neutrinos in the decay, we have to make use of the partner B, B_{tag}, coming with the signal B meson, B_{sig}, in the $\Upsilon(4S) \to B_{\text{sig}}\, B_{\text{tag}}$, making Belle II the only experiment for this analysis. Especially, we select B_{tag} decaying into Semileptonic mode, reconstructed using the `Full Event Interpretation` algorithm developed by the Belle II collaboration. This analysis does not use data; instead, it uses simulation. To control any mismodelling, we define control samples for which we can look both at data and simulation, and correct any mismodelling.

Using 10 million events of $B^+ \to K^+\nu\bar{\nu}$ decays, as well as $B\bar{B}$ and $q\bar{q}$ samples corresponding to 1948 fb^{-1}, we obtain a first estimate of the expected total uncertainty at 5×10^{-6} in the hypothesis of a branching fraction equal to its Standard Model value, and 7×10^{-6} for a branching fraction equal to the one obtained by the current state of the art. This measurement is thus expected to be as accurate as the one obtained by the previous Belle II measurements, proving its importance as an independent measurement as well as its relevance for combination.

Recherche des désintégrations $B \to K^{(*)}\nu\bar{\nu}$ à l'expérience Belle II

Cette thèse détaille la stratégie mise en œuvre pour mesurer la fraction d'enbranchement des désintégrations $B \to K^{(*)}\nu\bar{\nu}$, en se concentrant sur le processus $B^+ \to K^+\nu\bar{\nu}$, dans le cadre de l'expérience Belle II , à partir de 487,7 fb^{-1} de données à la résonance $\Upsilon(4S)$. Cette expérience est menée sur le collisionneur électron-positron SuperKEKB, au KEK, à Tsukuba, au Japon, spécialisé dans la production de mésons B. Les désintégrations $B \to K^{(*)}\nu\bar{\nu}$ constituent un outil important pour l'étude de la physique au-delà du modèle standard, car le processus sous-jacent est une transition de quark $b \to s\nu\bar{\nu}$, c'est-à-dire un changement de saveur par courant neutre. Grâce à la prédiction précise de son rapport de branchement dans le Standard Model, une mesure précise de la fraction de branchement de ces désintégrations pourrait révéler des tensions et apporter des preuves en faveur de nouvelle physique. En raison de la présence de deux neutrinos dans la désintégration, nous devons utiliser le B, B_{tag}, associé au méson signal B, B_{sig}, dans la réaction $\Upsilon(4S) \to B_{\mathrm{sig}}\, B_{\mathrm{tag}}$. En particulier, nous sélectionnons les B_{tag} se désintégrant selon des modes semi-leptonique, reconstitués à l'aide de l'algorithme `Full Event Interpretation` développé par la collaboration Belle II . Cette analyse n'utilise pas de données, mais repose sur une simulation. Afin de contrôler toute erreur de modélisation, nous définissons des échantillons de contrôle pour lesquels nous pouvons examiner à la fois les données et la simulation, puis éventuellement corriger

En utilisant 10 millions d'événements de désintégrations $B^+ \to K^+\nu\bar{\nu}$, ainsi que des échantillons $B\bar{B}$ et $q\bar{q}$ correspondant à 1948 fb^{-1}, nous obtenons une première estimation de l'incertitude totale attendue de l'ordre de 5×10^{-6} dans l'hypothèse d'une fraction de branchement égale à sa valeur Standard Model, et de 7×10^{-6} pour une fraction de branchement égale à celle obtenue par l'état actuel de l'art. Cette mesure devrait donc être aussi précise que celle obtenue par les précédentes mesures de Belle II, ce qui prouve son importance en tant que mesure indépendante ainsi que sa pertinence pour la combinaison des résultats.

Contents

Prelude

This thesis takes place in the context of the Standard Model (SM) of particle physics. This theoretical framework, describing elementary particles and their behaviour, is considered very successful when tracing back its history and the predictions that came from it. The model proved its predicting abilities on many occasions, *e.g.* with the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments [1].

However, a few years prior to the discovery of the Higgs boson, two independent teams observed the first oscillations of neutrinos [2, 3] . This is proof that neutrinos possess a mass, while it was not included by the SM. This is not its only limitation. The model is still unable to explain the origin of the matter/anti-matter asymmetry, nor the origin of dark matter whose presence is deduced from cosmological and astrophysical observations [4]. This entices us to keep pushing the boundaries of the SM and search for physics beyond the standard model.

Theorists describe many places to look for when searching for New Physics (NP). Processes of interest are Flavour Changing Neutral Current (FCNC) $b \to s$ transitions, rare processes, suppressed in the SM, for which some theories predict a different probability compared to the SM due to the inference of new phenomena.

Among the decays depending on FCNC processes, a particular one gained attention thanks to recent results from the Belle II collaboration: $B \to K^{(*)}\nu\bar{\nu}$ [5]. In this analysis, they found evidence for this decay but also pointed to a possible deviation from the SM predictions, as predicted by NP scenarios.

The $B \to K^{(*)}\nu\bar{\nu}$ decays are experimentally challenging to observe because of the two neutrinos in the final particles. Because of the impossibility to directly reconstruct 2 of the 3 final state particles and the rarity of the decay, no significant observation of the decay was made before Belle II.

However, its experimental complexity is compensated by the very precise prediction for the decay probability. With a precise measurement of this process, one would get a powerful probe for NP, both to identify possible models or to constrain others.

The goal of this thesis is to present the current status of the measurement the $B^+ \to K^+\nu\bar{\nu}$ decay branching fraction at the Belle II experiment, collecting data at the SuperKEKB collider, using a different approach compared to the previous analysis: a Semileptonic (SL) tagged analysis.

This work is organised as follows:

- Chapter 1 answers the questions "why?" by introducing the theoretical elements necessary for the comprehension of the importance for this measurement in the context of the SM and search for NP;
- Chapter 2 describes "where" this analysis is taking place, the experimental apparatus we will use and extract the data from;
- Chapter 3 explains "how" we will carry this analysis, the toolkit that will help us from reconstruction of the particles in the events to the extraction of the measurement;
- Chapter 4 is an intermezzo presenting the work performed on the graph neural network-based flavour tagger of Belle II;
- Chapter 5 finally exploits all the introduced elements to define an strategy to extract the "what", the branching ratio of the $B^+ \to K^+\nu\bar{\nu}$ decay;

This analysis relies on a collective work with other Belle II groups from Perugia and Bonn, both working on $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ decays. My contribution was Chapter 4 and Chapter 5. In particular, results from Chapter 4 are all based on my own work. For Chapter 5 the code is adapted from the previous hadronic tagged analysis. The work on Section 5.7.7 was done by the team in Perugia, and the Section 5.4.6 and Section 5.7.3 were done by Merna Abumusabh, a colleague in Strasbourg who will defend her thesis in 2027.

1

The why: scientific context on the search for $B \to K^{(*)}\nu\bar{\nu}$

Contents

The Standard Model (SM) of particle physics is a theoretical framework that describes the electromagnetic, weak, and strong nuclear interactions between elementary particles. Based on the principles of Quantum Field Theory (QFT), it has been tested extensively and has been able to describe the observations of particle physics experiments with great accuracy. However, there are several phenomena that the SM is not able to explain, such as the existence of Dark Matter (DM) or the matter-antimatter asymmetry in the universe. For reasons discussed later on, tensions with the SM have been previously observed when quark flavour transitions occur, such as in the $b \to sl^+l^-$, $b \to c\tau\nu$, or $b \to s\nu\bar{\nu}$ transitions. We will study the last one in this thesis, via the prism of the $B \to K\nu\bar{\nu}$ decay. In this chapter, we will first introduce the theoretical framework behind the SM and its limitations (Section 1.1), which will lead us to the formulation of the SM as an Effective Field Theory (EFT) (Section 1.2) and the study of the $B \to K\nu\bar{\nu}$ decay (Section 1.3). We will then describe NP models which could intervene in the $B \to K\nu\bar{\nu}$ decay and the experimental constraints on these models (Section 1.4). Finally, we will present the state of the art in the measurement of the $B \to K\nu\bar{\nu}$ decay (Section 1.5).

1.1 *The standard model of particle physics*

This summary of the SM is inspired by [6]

The SM provides mathematical tools to describe the interactions between elementary particles, which can be *fermions*, with half-odd spin, or *bosons*, with integer spin. The elementary bosons, or *gauge bosons*, act as mediators of the fundamental interactions: *electromagnetic*, *weak* and *strong* interactions. The elementary fermions are, in this theory, 12 particles (and 12 anti-particles) forming multiplets of the groups $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$.

The $SU(2)_L \otimes U(1)_Y$ gauge group describes the electroweak interaction, mediated respectively by the $(W^i)_{i\in[\![1,3]\!]}$ and B bosons, acting on the *weak isospin* $T \in \mathbb{R}^3$ and the *weak hypercharge* Y. At lower energy, this symmetry is spontaneously broken by the Higgs mechanism[1] [7, 8, 9], leading to the appearance of the $W^\pm$ and Z bosons, and the photon γ, which fields are defined as such:

[1] or, to be more complete, the Anderson-Brout-Englert-Guralnik-Hagen-Higgs-Kibble-'t Hooft mechanism

$$W^\pm = \frac{1}{\sqrt{2}}\left(W^1 \mp iW^2\right), \quad Z = W^3 \cos\theta_W - B\sin\theta_W, \tag{1.1}$$

$$A = W^3 \sin\theta_W + B\cos\theta_W \tag{1.2}$$

where θ_W is the weak mixing, or Weinberg, angle [10]. One also needs to redefine the generator of $U(1)_Y$ as $Q = T_3 + \frac{Y}{2}$, which is the electric charge operator. Additionally, this symmetry-breaking mechanism leads to the appearance of the Higgs boson H, which is the only scalar elementary particle in the SM.

The $SU(3)_C$ gauge group describes the strong interaction, and acts only on particles with a *colour charge* C. This group, being of dimension 8, has 8 gauge bosons, called *gluons* $(g_i)_{i\in[\![0,8]\!]}$. Elementary fermions can then be separated into two categories:

- *Quarks*, with a colour charge, which are grouped in three generations, each containing one quark with electric charge $Q = \frac{2}{3}$ and one with $Q = -\frac{1}{3}$, as follows:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.3}$$

  and their anti-particle equivalents. As they interact with the strong nuclear force, quarks are never observed in isolation, but always in bound states called hadrons (except for the top quark because of its lifetime), since free particles must always have a "null" colour charge. Particles composed of a quark and an anti-quark are called mesons, and those consisting of three quarks are called baryons.

- *Leptons*, without a colour charge, which are also grouped in three generations, containing a charged lepton and a neutrino, as follows:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \tag{1.4}$$

and their anti-particle equivalents.

All fermions interact weakly, and charged fermions also interact electromagnetically. From the way they act under $SU(2)_L$, one can write the fermions as *weak-isospin doublets*

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \quad \begin{pmatrix} u' \\ d' \end{pmatrix}_L \quad \begin{pmatrix} c' \\ s' \end{pmatrix}_L \quad \begin{pmatrix} t' \\ b' \end{pmatrix}_L \tag{1.5}$$

and *weak-isospin singlets*

$$e^-_R \quad \mu^-_R \quad \tau^-_R \quad u'_R \quad c'_R \quad t'_R \quad d'_R \quad s'_R \quad b'_R \tag{1.6}$$

with the L and R subscripts indicating the chirality of the particles, left-handed or right-handed, respectively. The right-handed neutrinos are not included in the minimal version of the SM.

The primes in the expressions above indicate that the quarks' "*weak*" eigenstates are not the same as their "*mass*" eigenstates [11]. The two bases can be related, in a particular basis, by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix, $V_{\text{CKM}} \in M_3(\mathbb{C})$ [12]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.7}$$

$$\begin{pmatrix} u' \\ c' \\ t' \end{pmatrix} = \begin{pmatrix} u \\ c \\ t \end{pmatrix} \tag{1.8}$$

We can also write this change of basis for the other quarks' states, relating (u, c, t) and (u', c', t') via the CKM matrix, and $(d, s, b) = (d', s', b')$.

This matrix encodes much information. Firstly, due to physical reasons[2], we need only three mixing angles $(\theta_{12}, \theta_{13}, \theta_{23}) \in \mathbb{R}^3$ and one phase $\delta_{13} \in \mathbb{R}$ in order to fully parametrise this matrix. The phase is interpreted as the *CP-violating* phase and is an important parameter of the SM to measure because of its implication in phenomena such as the *matter-antimatter asymmetry* [13]. Secondly, the squared modulus of the matrix elements is the probability for a quark to change flavour.

[2] The number of free real parameters of a unitary $n \times n$ matrix is n^2 ($2n^2 - n^2$ because of the constraints from the unitarity). Three of those parameters are angles, while the other six are phases. However, the 4-vector currents are invariant under a global phase transformation, so all the phases can be absorbed into a single one.

The unitary nature of the CKM matrix implies that:

$$\sum_{i \in \{u,c,t\}} V_{ij} V_{ik}^* = \delta_{jk}, \quad (j,k) \in \{d,s,b\}^2 \tag{1.9}$$

$$\sum_{i \in \{d,s,b\}} V_{ji} V_{ki}^* = \delta_{jk}, \quad (j,k) \in \{u,c,t\}^2 \tag{1.10}$$

where δ is the Kronecker symbol.

Let us develop Equation (1.9) for $j = b$ and $k = d$:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.11}$$

$$\iff 1 + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0 \tag{1.12}$$

as $V_{ud}V_{ub}^* \neq 0$. The last relation describes a *unitarity triangle* in the complex plane represented in Figure 1.1, with two of its vertices at $C = (0,0)$ and $B = (1,0)$ respectively. The last vertex is defined to be at $A = (\bar{\rho}, \bar{\eta}) \in \mathbb{R}^2$. Consequently, the lengths of the triangle sides are:

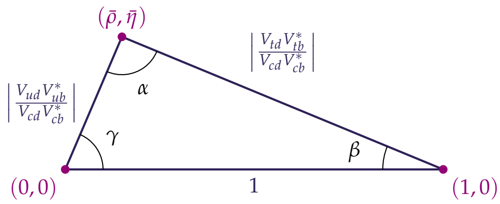


Figure 1.1: Representation of a unitarity triangle and its parametrisation.

$$\bar{AB} = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right| \tag{1.13}$$

$$\bar{AC} = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right| \tag{1.14}$$

$$\bar{BC} = 1 \tag{1.15}$$

and the angles, also parametrised by the CKM matrix elements, are:

$$\alpha = \arg\left(\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \tag{1.16}$$

$$\beta = \arg\left(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right) \tag{1.17}$$

$$\gamma = \arg\left(\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right) \tag{1.18}$$

Hence, the measurement of the sides and angles of the unitarity triangle allows us to test the SM and search for NP.

1.2 *The beauty in the standard model*

To work with processes occurring at energy scales much lower than the mass of the mediating particles, we can build an EFT, *i.e.* an approximation of a more fundamental theory valid at a lower energy scale. An EFT is built by integrating out the high-energy degrees of freedom from the full theory. This results in an effective Hamiltonian expressed as a series of local operators O multiplied by their corresponding *Wilson coefficients* C. The operators must respect the symmetries of the lower energy scale theory, and the Wilson coefficients encode the effects of the higher energy physics. The effective Hamiltonian, at a given energy scale $\mu \in \mathbb{R}_{>0}$, can be expressed as:

$$\mathcal{H}_{\text{eff}} = \sum_{i=1}^{N} C_i(\mu) O_i(\mu) \tag{1.19}$$

where $N \in \mathbb{N}^*$ is the order of the series expansion. For a weak decay, occurring at an energy scale much lower than the mass of the W boson m_W, the EFT is known as the *weak effective theory*.

1.3 *The $B \to K^{(*)} \nu\bar{\nu}$ decays in the standard model*

As one can see from the unitarity and the transitions represented by the CKM matrix elements, change of flavour between same electric charge quarks, also called FCNC, cannot occur at tree level. Consider a transition from a b quark to an s quark. It is forbidden at tree level but could occur at loop level with the exchange of a W boson, *e.g.* as presented in Figure 1.2. The amplitude of this transition is proportional to the product of the $b \to j$ and $j \to s$ amplitudes and squared mass of the intermediate quark m_j^2, where $j \in (u, c, t)$. Taking the sum of the amplitudes for all possible intermediate quarks:

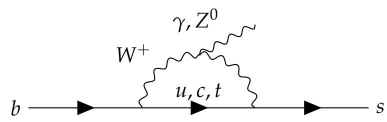


Figure 1.2: Feynman diagram representing the $b \to s$ transition at one-loop level.

$$\mathcal{M}_{b \mapsto s} \propto \sum_{j \in \{u,c,t\}} V_{bj} V_{js}^* m_j^2 \tag{1.20}$$

Hence, returning to equation Equation (1.9), if $m_u = m_c = m_t$, we can factorize m_j^2 for $j = u$ in Equation (1.20), and the amplitude of the $b \mapsto s$ transition is null. If such a symmetry of masses exists, FCNC would be forbidden at the one-loop level. However, this is not the case, and instead the amplitude of the transition is dominated by the contribution of the heaviest quark, the t quark [3]. Thus, FCNC processes in the SM are highly suppressed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [14]. This rarity makes FCNC processes very interesting to study in order to search for NP, because new particles could contribute in the loops, or involve new mechanisms that may allow FCNC at tree level, modifying the amplitude of those transitions significantly.

[3] The leading order in Equation (1.20) is $m_c^2 - m_t^2$, but m_t is much larger than m_c.

Notable examples of FCNC processes are the $B \to K^{(*)} \nu\bar{\nu}$ decays. The $B \to K^{(*)} \nu\bar{\nu}$ decays are rare decays of the B meson into a $K^{(*)}$ meson and a pair of neutrinos. In terms of quark content, it corresponds to a $b \to s\nu\bar{\nu}$ transition, hence it involves an FCNC. The leading order Feynman diagrams contributing to this process in the SM are the one-loop *penguin* and *box* diagrams, as shown in Figure 1.3 for $B^+ \to K^+\nu\bar{\nu}$. This decay is particularly interesting because it is suppressed in the SM, as explained in the previous section, and thus sensitive to NP effects. In the following, we will talk about the $B \to K\nu\bar{\nu}$ and $B \to K^*\nu\bar{\nu}$ decays generically, but it can actually correspond to four different modes:

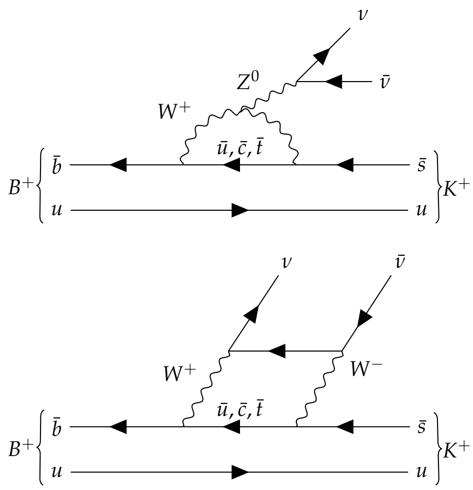


Figure 1.3: Feynman diagrams representing the $B^+ \to K^+\nu\bar{\nu}$ transition at one-loop level in the SM, with penguin (above) and box (below) diagrams.

- $B^\pm \to K^\pm\nu\bar{\nu}$
- $B^0 \to K^0_S\nu\bar{\nu}$
- $B^\pm \to K^{*\pm}\nu\bar{\nu}$
- $B^0 \to K^{*0}\nu\bar{\nu}$

For the $b \to s\nu\bar{\nu}$ decay, with $\mu \approx m_b$ [15]:

$$\mathcal{H}_{\text{eff}} \approx -\frac{4G_F}{\sqrt{2}} V^*_{ts} V_{tb} C^{\text{SM}}_L O^\nu_L + h.c. \tag{1.21}$$

The operator O^ν_L is defined as [16]:

$$\begin{aligned} O^\nu_L &= \frac{1}{2}\frac{\alpha}{4\pi}(\bar{s}b)_{V-A}(\bar{\nu}\nu)_{V-A} && (1.22)\\ &= \frac{e^2}{16\pi^2}\left(\bar{s}\gamma_\mu \frac{1-\gamma^5}{2} b\right)\left(\bar{\nu}\gamma^\mu(1-\gamma^5)\nu\right) && (1.23)\end{aligned}$$

where $(\gamma^\mu)_{\mu\in[\![0,3]\!]}$ are the gamma matrices, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and the Wilson coefficient C^ν_L is:

$$C^{\text{SM}}_L = -\frac{X_t}{\sin{(\theta_W)}^2} \tag{1.24}$$

where $X_t = 1.469(17)(2)$ [17] is is the Inami–Lim function, including higher-order QCD and electroweak corrections [18], and $\sin{(\theta_W)}^2 = 0.22348(4)$ [19]. In this effective theory, we can compute the branching ratio of the $B \to K\nu\bar{\nu}$ decay, which is given by Fermi's golden rule:

$$\mathcal{B}\left(B \to K\nu\bar{\nu}\right) = N\tau_B |\langle K\nu\bar{\nu}| \mathcal{H}_{\text{eff}} |B\rangle|^2 \rho \tag{1.25}$$

where N is a normalization factor, τ_B is the B meson lifetime, ρ is the phase space factor and $\langle K\nu\bar{\nu}| \mathcal{H}_{\text{eff}} |B\rangle$ is the matrix element of the effective Hamiltonian between the initial and final states. A powerful observable for studying those decays is the differential branching ratio. For $B \to K\nu\bar{\nu}$, to fully describe the kinematics of the decay, we need the squared invariant mass of the neutrino pair, denoted q^2, over which we can define the differential branching ratio to also study the shape of the distribution in search for NP.

The differential branching ratio for $B \to K\nu\bar{\nu}$ is given by [20]:

$$\frac{\mathrm{d}\mathcal{B}\left(B \to K\nu\bar{\nu}\right)}{\mathrm{d}q^2} = \mathcal{N}_K\left(q^2\right)\left|C_L^{\mathrm{SM}}\right|^2 |V_{tb}V_{ts}^*|^2 f_+^2(q^2) \tag{1.26}$$

with

$$\mathcal{N}_K\left(q^2\right) = \frac{G_F^2\alpha_{\mathrm{EW}}^2}{256\pi^5 m_B^3}\lambda\left(m_B^2, m_K^2, q^2\right)^{3/2}\tau_B \tag{1.27}$$

where α_{EW} is the fine structure constant, λ is the Källén function and f_+ is the $B \to K$ vector form factor, which encapsulates the parametrisation of the $B \to K$ transition matrix element. In the case of the $B \to K$ transition, the form factor is computed via lattice quantum chromo-dynamics [21, 22, 23]. One can also parametrise the form factor using a z-expansion [24], leading to:

$$f_+(q^2) = \frac{1}{1-\frac{q^2}{m_+^2}}\left[\alpha_0 + \alpha_1 z(q^2) + \alpha_2 z(q^2)^2 + \frac{-\alpha_1 + 2\alpha_2}{3} z(q^2)^3\right] \tag{1.28}$$

with

$$z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}} \tag{1.29}$$

where $t_+ = (m_B + m_K)^2$, $t_0 = (m_B + m_K)(\sqrt{m_B} - \sqrt{m_K})^2$ and $m_+ = m_B + 0.046$ GeV is the resonance mass. The values of the α parameters, and their correlation matrix for $B^+ \to K^+\nu\bar{\nu}$ (indicated by the + sign) are obtained after fitting the form factor f_+ with the described parametrisation [20]:

$$\alpha_0^+ = 0.4742 \pm 0.0062 \tag{1.30}$$
$$\alpha_1^+ = -0.894 \pm 0.051 \tag{1.31}$$
$$\alpha_2^+ = -0.44 \pm 0.14 \tag{1.32}$$

and

$$C_{\alpha,+} = \begin{pmatrix} 1 & -0.2904 & -0.0347 \\ -0.2904 & 1 & 0.7757 \\ -0.0347 & 0.7757 & 1 \end{pmatrix} \tag{1.33}$$

Back to the branching ratio predictions, using Equation (1.26), integrating over q^2, one can compute the branching ratio. For numerical applications, we use $V_{ts}^* V_{tb} = (39.3 \pm 1.0) \times 10^{-3}$, as specified in [20]. Hence, we obtain:

$$\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right)_{\mathrm{SM}} \approx (5.06 \pm 0.31) \times 10^{-6} \tag{1.34}$$

The branching ratio for $B^0 \to K_S^0\nu\bar{\nu}$ can be derived from Equation (1.26) similarly, since most of the parameters will be the same.

Hence, just by taking into account the different lifetimes, we get:

$$\mathcal{B}\left(B^0 \to K^0_S \nu\bar{\nu}\right)_{\mathrm{SM}} = \frac{1}{2}\frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right)_{\mathrm{SM}} \tag{1.35}$$

$$\approx (2.05 \pm 0.14) \times 10^{-6} \tag{1.36}$$

One can do a similar work for $B \to K^*\nu\bar{\nu}$ branching ratios. Starting from the formula [20]:

$$\frac{\mathrm{d}\mathcal{B}\left(B \to K^*\nu\bar{\nu}\right)}{\mathrm{d}q^2} = \mathcal{N}_{K^*}\left(q^2\right)\left|C_L^{\mathrm{SM}}\right|^2 |V_{tb}V_{ts}^*|^2 \mathcal{F}(q^2) \tag{1.37}$$

with

$$\mathcal{N}_{K^*}\left(q^2\right) = \frac{G_F^2\alpha_{\mathrm{EW}}^2}{128\pi^5 m_B^3}\lambda\left(m_B^2, m_{K^*}^2, q^2\right)^{1/2} q^2\tau_B\left(m_B + m_{K^*}\right)^2 \tag{1.38}$$

However, $\mathcal{F}$ is not the form factor, it is a combination of 3 form factors because of the vector nature of the K^* meson:

$$\mathcal{F}\left(q^2\right) = A_1^2\left(q^2\right) + \frac{32 m_B^2 m_{K^*}^2}{q^2\left(m_B + m_{K^*}\right)^2}A_{12}^2\left(q^2\right) + \frac{\lambda\left(m_B^2, m_{K^*}^2, q^2\right)}{\left(m_B + m_{K^*}\right)^4}V^2\left(q^2\right) \tag{1.39}$$

where A_1, A_{12} and V are the (axial and vector) form factors, computed using light-cone sum rules [25].

Using the differential branching fraction from Equation (1.37), the results are:

$$\mathcal{B}\left(B^+ \to K^{*+}\nu\bar{\nu}\right)_{\mathrm{SM}} \approx (10.86 \pm 1.43) \times 10^{-6} \tag{1.40}$$

$$\mathcal{B}\left(B^0 \to K^{*0}\nu\bar{\nu}\right)_{\mathrm{SM}} \approx (9.05 \pm 1.37) \times 10^{-6} \tag{1.41}$$

Additionaly, the branching fractions predicted for $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$ are taking into account the *long-distance contribution* $B^+ \to \tau^+\left(\to K^{(*)+}\bar{\nu}_\tau\right)\nu_\tau$. We need to subtract this contribution which gives us, after computation [20, 26]:

$$\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right)_{\mathrm{SM}} \approx (4.44 \pm 0.30) \times 10^{-6} \tag{1.42}$$

$$\mathcal{B}\left(B^+ \to K^{*+}\nu\bar{\nu}\right)_{\mathrm{SM}} \approx (9.8 \pm 1.4) \times 10^{-6} \tag{1.43}$$

We can see here that those decays are indeed rare in the SM, with branching ratios of the order of 10^{-6}. This makes them sensitive to NP effects, which could modify those branching ratios significantly. Moreover, the theoretical predictions are precise, with uncertainties of about 6% for $B \to K\nu\bar{\nu}$, mainly due to the CKM matrix elements uncertainties, and 13% for $B \to K^*\nu\bar{\nu}$, mainly due to the form factor uncertainties. This is because the $B \to K^{(*)}\nu\bar{\nu}$ decays are theoretically clean, as they do not suffer from long-distance

QCD effects that can plague other rare decays. This makes them ideal probes for testing the SM and searching for NP effects.

1.4 *New physics models in the $B \to K^{(*)}\nu\bar{\nu}$ decays*

To take into account the effects of NP in $B \to K^{(*)}\nu\bar{\nu}$ decays, we can modify the effective Hamiltonian in Equation (1.21):

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \left(\left(C_L^{\text{SM}} + C_L^{\text{NP}} \right) O_L^{\nu} + C_R^{\text{NP}} O_R^{\text{NP}} \right) + h.c. \tag{1.44}$$

It must be noted that we assume Lepton Flavour Universality (LFU) in this effective theory, *i.e.* that the potential NP produced will couple universally to the three flavours of neutrinos, without any preference. It is possible to consider LFU-violation by introducing different Wilson coefficients for each neutrino flavour. Compared to the Standard Model, we added a right-handed operator O_R^{NP} and the corresponding Wilson coefficient C_R^{NP} for the neutrino. The operator O_R^{NP} is defined as [15]:

$$O_R^{\text{NP}} = \frac{e^2}{16\pi^2} \left(\bar{s}\gamma_\mu P_R b \right) \left(\bar{\nu}\gamma^\mu \left(1-\gamma_5\right) \nu \right) \tag{1.45}$$

where $P_R = \frac{1+\gamma_5}{2}$ is the right-handed projection operator. For convenience, we can define two real parameters from the Wilson coefficients:

$$\epsilon = \frac{\sqrt{|C_L^{\text{SM}} + C_L^{\text{NP}}|^2 + |C_R^{\text{NP}}|^2}}{C_L^{\text{SM}}} \tag{1.46}$$

$$\eta = -\frac{\text{Re}\left(\left(C_L^{\text{SM}} + C_L^{\text{NP}} \right) C_R^{\text{NP}*} \right)}{|C_L^{\text{SM}} + C_L^{\text{NP}}|^2 + |C_R^{\text{NP}}|^2} \tag{1.47}$$

The parameter ϵ quantifies the strength of the NP contribution relative to the SM, while η measures the interference between the left- and right-handed contributions. In the SM-only case, we have $\epsilon = 1$ and $\eta = 0$, making the measurements of those parameters a check for presence of NP. Moreover, using those new parameters, we can define the *signal strength*:

$$\mu_K = \frac{\mathcal{B}(B \to K\nu\bar{\nu})}{\mathcal{B}(B \to K\nu\bar{\nu})_{\text{SM}}} = (1 - 2\eta)\,\epsilon^2 \tag{1.48}$$

$$\mu_{K^*} = \frac{\mathcal{B}(B \to K^*\nu\bar{\nu})}{\mathcal{B}(B \to K^*\nu\bar{\nu})_{\text{SM}}} = (1 + \kappa\eta)\,\epsilon^2 \tag{1.49}$$

where κ is computed from the form factors of the $B \to K^*$ hadronic matrix element and is about 1.34 [15].

Hence, measurements of the various $B \to K^{(*)}\nu\bar{\nu}$ decay branching ratios allow us to constrain the parameters ϵ and η, thus constraining the Wilson coefficients. We can then pre-emptively ask ourselves what kind of NP models we are sensitive to, and how they can affect our measurements.

At the vertex level, possible NP contributions include the inference of new heavy mediators in the decay, such as Z' [27], another massive neutral boson involved in the flavour changing, or such as leptoquarks [28]. The indirect nature of their production in the decay makes it possible to consider mediators with masses much higher than the energy scale of the decay.

Moreover, the $B \to K^{(*)}\nu\bar{\nu}$ decays are also a great test of LFU. Even considering that the reconstruction of the dineutrino system is *flavour-blind*, it is possible to construct a band of allowed values for the branching ratio of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ decays outside of which LFU would be violated [29], as shown in Figure 1.4. Indeed, in the framework of the SMEFT, one can relate dilepton decays to dineutrino decays, hence providing a bridge between the flavour-blindness and the actual lepton flavours. It is then possible to rewrite the expression of the effective Hamiltonian and the branching fractions in order to obtain the band presented [30, 31],

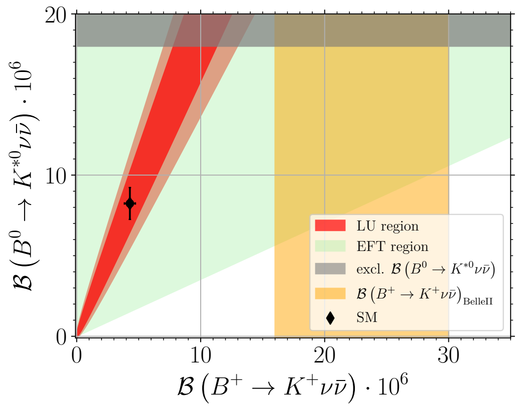


Figure 1.4: Branching ratio of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ decays with the band of allowed values if LFU is verified in red. The orange band corresponds to the current state of the art measurement of the $B^+ \to K^+\nu\bar{\nu}$ branching ratio with its 90% Confidence Level (CL) [29].

At lower energy, we can consider any invisible particle in the final state. Indeed, current collider experiments do not detect the neutrinos produced by the decay. This means that any measurement of $B \to K^{(*)}\nu\bar{\nu}$ branching ratio includes all $B \to KX$ modes, where X is any invisible particle, such as dark matter or light new particles [32].

Further discussions on the possible NP contributions to $B \to K^{(*)}\nu\bar{\nu}$ decays can be found in [33].

1.5 *State of the art*

The $B \to K^{(*)}\nu\bar{\nu}$ decays are powerful probes for various NP models. Another point of interest is also the experimental challenges they pose. Indeed, the presence of two neutrinos in the final state, with a single kaon as our visible particle, makes it impossible to reconstruct the decay kinematics fully; hence, it is a very challenging measurement, and only a few experiments can perform it.

Various experiments have searched for the $B \to K^{(*)}\nu\bar{\nu}$ decays so far; the most recent ones are: Belle, BaBar, and Belle II. All three experiments are e^+e^- colliders operating at the $\Upsilon(4S)$ resonance, which decays almost exclusively into $B\bar{B}$ pairs. More details about why we need those specific apparatus are given in Chapter 2.

Because of the inherent difficulty of the task, most of the results are upper limits, summarised in Table 1.1. The *tagging* refers to the strategy used to perform the analysis. A detailed explanation of what each type of tagging is and how they work can be found in Section 3.2.

The latest results come from the Belle II collaboration, using a luminosity of 362 fb^{-1}. It established evidence for the $B^+ \to K^+\nu\bar{\nu}$ decay, with a significance of 3.5 standard deviations with respect to the background-only hypothesis, and a 2.7 standard deviation above the SM prediction [5], with a branching ratio of:

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (2.3 \pm 0.7) \times 10^{-5} \tag{1.50}$$

The current status of the experimental results is summarised in Figure 1.5.

Experiment	$\mathcal{L}$ [fb^{-1}]	Tagging	Mode	Limit at 90% CL
BaBar [34]	418	SL	K^+	$< 1.3 \times 10^{-5}$
			K^0	$< 5.6 \times 10^{-5}$
BaBar [35]	429	HAD	K^+	$< 3.7 \times 10^{-5}$
			K^0	$< 8.1 \times 10^{-5}$
			K^{*+}	$< 11.6 \times 10^{-5}$
			K^{*0}	$< 9.3 \times 10^{-5}$
BaBar [35]	429	HAD + SL	K^+	$< 1.6 \times 10^{-5}$
			K^0	$< 4.9 \times 10^{-5}$
			K^{*+}	$< 6.4 \times 10^{-5}$
			K^{*0}	$< 12 \times 10^{-5}$
Belle [36]	711	HAD	K^+	$< 5.5 \times 10^{-5}$
			K^0_S	$< 9.7 \times 10^{-5}$
			K^{*+}	$< 4.0 \times 10^{-5}$
			K^{*0}	$< 5.5 \times 10^{-5}$
Belle [37]	711	SL	K^+	$< 1.9 \times 10^{-5}$
			K^0_S	$< 1.3 \times 10^{-5}$
			K^{*+}	$< 6.1 \times 10^{-5}$
			K^{*0}	$< 1.8 \times 10^{-5}$
Belle II [38]	63	INC	K^+	$< 4.1 \times 10^{-5}$

Table 1.1: Summary of the main experimental results for the $B \to K^{(*)}\nu\bar{\nu}$ decays. The limits are given at 90% Confidence Level (CL). The tagging used is Semileptonic (SL), Hadronic (HAD) and Inclusive (INC) tagging, as well as combinations of measurements.

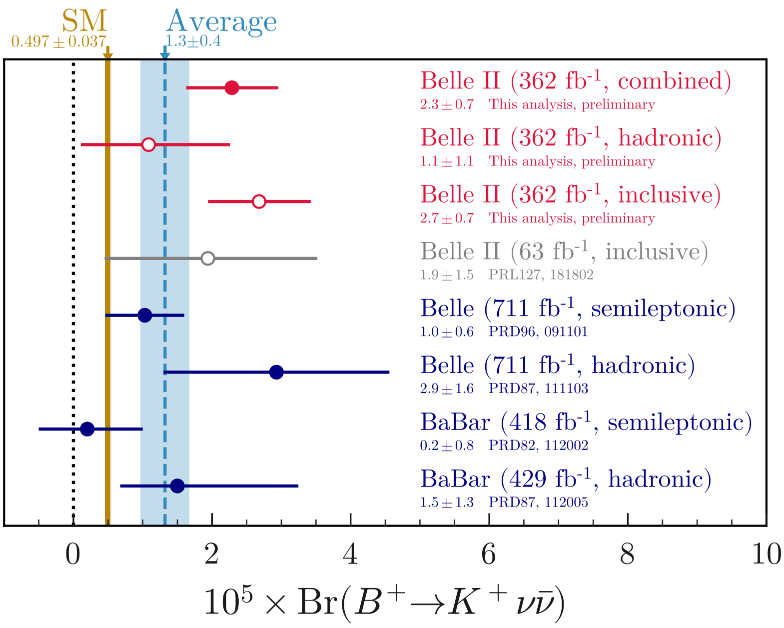


Figure 1.5: Summary plot presenting the current experimental results for the measurements of the $B \to K^{(*)}\nu\bar{\nu}$ decays [5].

2

The where: experimental setup

Contents

In the following chapter, especially in Section 2.1, we will describe what apparatus could allow us to measure the $B \to K^{(*)}\nu\bar{\nu}$ decay and what key performances are needed. This will naturally lead us to present the SuperKEKB accelerator and the Belle II detector. Belle II is an experiment built on the SuperKEKB accelerator in Tsukuba, Japan. It is designed to study the properties of B mesons. The experiment aims to investigate the phenomenon of Charge-Parity (CP) violation and to search for signs of NP beyond the SM. We will first describe the SuperKEKB accelerator, in Section 2.2, which will allow us to talk more about the collisions of interest and their pros regarding the $B \to K^{(*)}\nu\bar{\nu}$ decays. We will then present the Belle II detector, in Section 2.3, which is used to detect and characterise the particles produced in these collisions. We will finally cover the `Belle II Analysis Software Framework` (Section 2.4), and how the simulation (Section 2.5) or the reconstruction (Section 2.6) are done using this software framework.

2.1 *The challenges: building the ideal experiment*

The $B \to K^{(*)}\nu\bar{\nu}$ decays are characterised by two neutrinos in the final state and a single kaon as the only visible particle, which makes their reconstruction and selection very challenging. While a two-body decay or a three-body decay with one invisible particle can be fully reconstructed kinematically using energy and momentum conservation, this is not possible for a decay with two invisible particles. In this state, the signal decay cannot be reconstructed without additional constraints. The first challenge is therefore to find a way to constrain the signal kinematics. An ideal environment to study such decays would be an e^+e^- collider operating at the $\Upsilon(4S)$ resonance, which decays almost exclusively into $B\bar{B}$ pairs. In this case, the presence of two B mesons in the event can be used to constrain the kinematics of the signal decay, from the B_{sig}, by reconstructing the second B meson in the event, referred to as the B_{tag}.

Moreover, $e^+e^- \to \Upsilon(4S)$ colliders solve another challenge for the analysis: as our only reliable signal signature is a single kaon, we need a clean environment that can only be provided by lepton-lepton colliders, since lepton-hadron or hadron-hadron colliders would produce many extra particles. By knowing the four-momentum of the $\Upsilon(4S)$ and with only the B_{sig} and the B_{tag} in the event, we could infer the four-momentum of the signal decay and reconstruct its kinematics. This technique, known as tagging, is a powerful tool for studying decays with multiple invisible particles in the final state.

The third challenge is the low kinematic constraints on the signal-side kaon. To reliably identify the signal decay, it is crucial to have a hermetic detector with excellent particle identification capabilities. The detector's hermeticity ensures that all visible particles in the event are reconstructed, so that missing energy and momentum can be properly attributed to the neutrinos from the signal decay. Excellent particle identification is needed to separate the signal kaon from other particles, such as pions, which may be present in the event.

Finally, since missing energy is a key variable in the analysis, it is crucial to have a detector with excellent calorimetric capabilities to accurately measure the energy of all visible particles in the event, especially neutral particles like photons and neutrons, which do not leave tracks in the tracking system. Any mismeasurement of the energy of these particles can lead to incorrect estimates of the missing energy and significantly impact the sensitivity of the analysis.

From this description, only one experiment currently in operation meets all these requirements: the Belle II experiment at the SuperKEKB e^+e^- collider in Japan. The following sections will describe how Belle II meets these challenges and how the detector is used in the $B \to K^{(*)}\nu\bar{\nu}$ analysis.

2.2 *The accelerator: SuperKEKB*

As specified, SuperKEKB is an electron-positron collider located at the KEK laboratory in Tsukuba, Japan [39]. It is designed to operate at the centre-of-mass energy of 10.58 GeV, hence producing $\Upsilon(4S)$ mesons that decay into $B\bar{B}$ pairs. Electrons are accelerated in the high-energy ring to an energy of 7.007 GeV, while positrons are accelerated in the low-energy ring to an energy of 4.000 GeV, before colliding at the Interaction Point (IP) inside the Belle II detector. Because of the energy asymmetry between the two beams, the resulting $\Upsilon(4S)$ mesons are produced with a Lorentz boost:

$$\beta\gamma = \frac{1}{2}\frac{E_{e^-} - E_{e^+}}{\sqrt{E_{e^-}E_{e^+}}} \approx 0.28 \tag{2.1}$$

which helps in the reconstruction of decay vertices, since otherwise both B mesons would decay too close to each other to be distinguished.

Process	Cross section (nb)
$e^+e^- \to \Upsilon(4S)$	1.110
$e^+e^- \to u\bar{u}(\gamma)$	1.61
$e^+e^- \to d\bar{d}(\gamma)$	0.40
$e^+e^- \to s\bar{s}(\gamma)$	0.38
$e^+e^- \to c\bar{c}(\gamma)$	1.30
$e^+e^- \to e^+e^-(\gamma)$	3.00×10^2
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.148
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.919
$e^+e^- \to \gamma^+\gamma^-(\gamma)$	4.99

Table 2.1: Main processes occurring at the SuperKEKB collider at a centre-of-mass energy of 10.58 GeV, along with their cross sections [41].

While the $e^+e^- \to \Upsilon(4S)$ is the process of interest, other processes, called beam-beam backgrounds, also occur at the same centre-of-mass energy, as reported in Table 2.1. Additionaly, the collider can operate at different centre-of-mass energies, between the $\Upsilon(1S)$ (9.46 GeV) and the $\Upsilon(11020)$ (11.00 GeV) [19], and can run up to 11.24 GeV [40]. Events where the collider does not operate at the $\Upsilon(4S)$ resonance are referred to as *off-resonance* events. Those types of events are used to study the background processes that do not involve $B\bar{B}$ production, such as the continuum processes $e^+e^- \to q\bar{q}(\gamma)$, where $q = u, d, s, c$. Other highly probable processes, including Touschek scattering and beam-gas interactions, can occur and are called single-beam machine-induced backgrounds.

Finally, another important aspect of SuperKEKB is its targeted high integrated luminosity. Indeed, the higher the integrated luminosity, the higher the statistics, which is crucial for rare decays such as $B \to K^{(*)}\nu\bar{\nu}$. Currently, SuperKEKB has reached a peak luminosity of 5.1×10^{34} cm^{-2}s^{-1} [42]. The entire integrated luminosity registered by Belle II as of the 16th of December 2025 is shown in Figure 2.1, amounting to a total of 589.89 fb^{-1} [43].

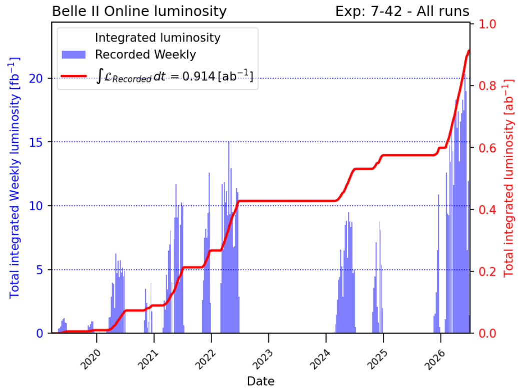


Figure 2.1: Total integrated luminosity recorded per day and overall by the Belle II detector since the start of data taking in 2019 up to the 4th of July 2026 [43].

2.3 *The detector: Belle II*

The Belle II detector is located at the IP of SuperKEKB to detect and characterise the particles produced in the e^+e^- collisions. To account for the Lorentz boost, the detector is slightly asymmetric, with forward and backward regions (endcaps), and the forward endcap being slightly larger. Within its barrel-shaped structure, it is composed of several sub-detectors arranged in layers around the IP, each designed to measure different properties of the particles for their detection and identification. From the innermost to the outermost layers, the main sub-detectors of Belle II are [44]:

- The Vertex Detector (VXD), comprises two devices (Pixel Detector (PXD) and Silicon Vertex Detector (SVD)), used to measure the decay vertex positions.
- The Central Drift Chamber (CDC), a large cylindrical drift chamber that provides tracking and momentum measurement for charged particles.
- A Particle Identification (PID) system, composed of two detectors:
  - The Time Of Propagation (TOP) counters in the barrel region, based on Cherenkov radiation to identify charged particles based on their velocity.
  - The Aerogel Ring-Imaging Cherenkov (ARICH) detector in the forward endcap, also based on Cherenkov radiation for particle identification.
- The Electromagnetic Calorimeter (ECL), an electromagnetic calorimeter used to measure the energy of electrons and photons.
- The K_L and Muon detector (KLM), a detector for K_L^0 mesons and muons.

A schematic view of the sub-detectors is presented in Figure 2.2.

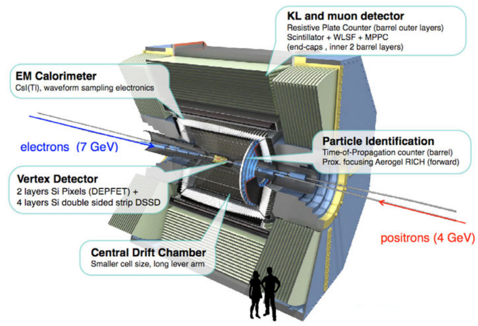


Figure 2.2: Diagram of the Belle II detector and its sub-detector components [45].

Those sub-detectors are enclosed in a superconducting solenoid that provides a magnetic field of 1.5T, which allows the measurement of the charge and momentum of charged particles by observing their curvature in the magnetic field. Finally, Belle II is equipped with a triggering scheme to reject uninteresting events.

We will now give more details about the PID system and ECL, and how they are useful for the $B \to K^{(*)}\nu\bar{\nu}$ analysis.

2.3.1 *Particle identification system (TOP and ARICH)*

As specified, the role of the Time Of Propagation (TOP) and Aerogel Ring-Imaging Cherenkov (ARICH) detectors is to identify the charged particles produced in the collisions. In our case, we want to have good discrimination between kaons and pions. Both detectors rely on the Cherenkov effect. Indeed, when a charged particle travels through the quartz of the TOP, or the aerogel of the ARICH, at a speed greater than the speed of light in that medium, it emits Cherenkov photons. The general principle of particle identification is to measure the emission angles of the Cherenkov photons and to deduce the velocity of the particle using:

$$\beta = \frac{1}{n \cos \theta_C} \tag{2.2}$$

where n is the refractive index of the aerogel and θ_C is the Cherenkov angle. Combining this velocity measurement with the momentum measurement from the CDC, one can determine the mass of the particle and hence its identity.

In the case of the TOP, the Cherenkov photons are internally reflected within a quartz radiator and detected by an array of photodetectors at the end of the bar. The particle type is then determined by comparing the photons' hit patterns, both temporal and spatial, in the TOP module with Probability Density Function (PDF) templates for different particle hypotheses (e.g., pion and kaon). A schematic view of the TOP principle and particle identification is shown in Figure 2.3. In order to have a proper separation between kaons and pions, a minimum single photon temporal resolution of 100ps is necessary. This is largely achieved by the TOP module which has a timing accuracy of approximately 50ps [46]. The TOP detector has an 85% kaon identification efficiency for a 10% pion misidentification rate [46].

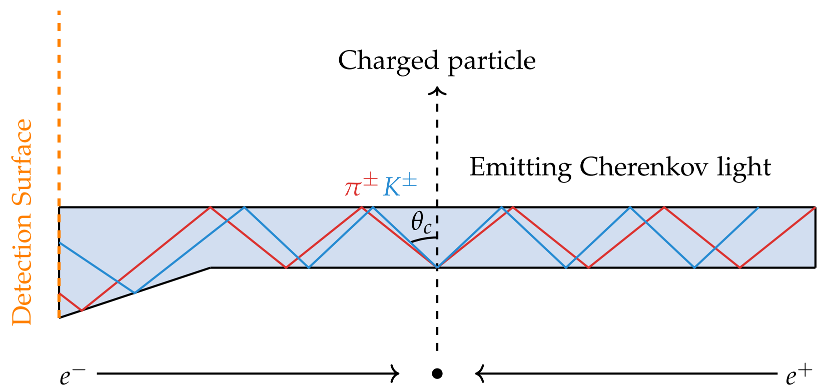


Figure 2.3: Schematic representation of Cherenkov radiation paths from a kaon and a pion passing through the TOP module. Adapted from [47]

For the ARICH, the Cherenkov photons are emitted in the aerogel and directly detected by an array of photodetectors (Hybrid Avalanche Photo-Detectors, HAPDs). Then, a similar method as for the TOP is used to identify the particle by comparing the detected photons' hit patterns with PDF templates for different particle hypotheses. A schematic view of the ARICH principle is shown in Figure 2.4. The ARICH detector has a 95% kaon identification efficiency with a 10% pion misidentification rate [48].

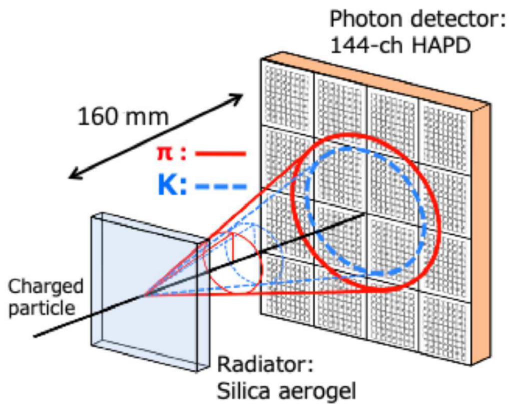


Figure 2.4: Schematic representation of the ARICH detector principle [49].

2.3.2 *Electromagnetic calorimeter*

The ECL is designed to detect photons, measure their energy and angular coordinates, as well as to identify electrons with respect to other hadrons like pions. It is composed of 8736 thallium-doped cesium iodide (CsI(Tl)) crystals arranged in a barrel and two endcaps, covering a polar angle range from 12.4° to 155.1°, which represents about 90% of the solid angle in the centre-of-mass frame [41]. When a high-energy particle enters the ECL, it interacts with the CsI(Tl) crystals and produces an electromagnetic or hadronic shower, a cascade of secondary particles that deposit their energy in the crystals. The scintillation light produced by the CsI(Tl) crystals is then detected by photodiodes attached to each crystal, converting the light into an electrical signal proportional to the energy deposited. However, the energy is generally deposited over multiple crystals, forming a cluster.

Only about 0.2% of the photons fail to leave a detectable signal in the ECL [50], for an energy resolution $\frac{\sigma_E}{E}$ of 4% at 100 MeV and 1.6% at 8 GeV. The low number of undetected photons and the good energy resolution is of key interest in the $B \to K^{(*)}\nu\bar{\nu}$ analysis, where we want to reconstruct events with missing energy due to the neutrinos, so we should not omit any energy contribution from neutral particles.

2.4 *The software framework:* BASF2

Once the data is collected by the Belle II detector, it needs to be processed and analysed. For this purpose, the Belle II collaboration has developed a software framework called BASF2 [51]. This framework is designed to handle all aspects of data processing, from raw data acquisition to high-level physics analysis. It is built using modern programming practices and is modular, allowing for easy integration of new algorithms and tools. Modules are written either in `C++` or `Python`, but external libraries can also be interfaced, like `Pythia`

[52] and EvtGen [53] for simulation, or ROOT [54] for data storage and handling.

2.5 The simulation: modelling the physics

Monte Carlo (MC) methods are widely used at Belle II to simulate physics processes and detector response. The BASF2 framework provides a comprehensive suite of tools for detector simulation, using the Geant4 software [55], and for generating MC events. The simulation begins by generating the physics process of interest, such as the e^+e^- collision, and the immediate particle decays, before passing the particles through a detailed model of the Belle II detector.

Cross-sections of possible events in the e^+e^- collision at the 10.58 GeV centre-of-mass energy were already given in Table 2.1. After the e^+e^- collision, three generators are used depending on the event type:

- EvtGen 1.3 [53] is used for exclusive B and D meson decays.
- KKMC 4.15 [56] is used for generating τ pairs, while TAUOLA handles their decay [57].
- PYTHIA 8.2 [58] is used for inclusive decays and continuum production.

Beam-beam backgrounds, such as Bhabha scattering and two-photon events [1], are also simulated to model detector background conditions, using BABAYAGA.NLO [59]. Finally, non-radiative four fermion processes [2] are generated using AAFH [60].

[1] *e.g.* $e^+e^- \to e^+e^-\ (\gamma)$ and $e^+e^- \to \gamma\gamma(\gamma)$

[2] *e.g.* $e^+e^- \to e^+e^-e^+e^-$ and $e^+e^- \to e^+e^-\mu^+\mu^-$

An exclusive, non-exhaustive list of inclusive decays to be generated, along with their branching ratios, is provided in a decay table. Any final state not included in the decay table is handled by PYTHIA, while making sure to avoid double counting.

Single-beam machine-induced backgrounds are directly extracted from the detectors, run-by-run, and overlaid on the physics events to accurately represent the detector conditions during data taking [61].

After that, the 4-momenta of the generated particles are passed to Geant4, which simulates their interactions with the detector material, including energy loss, multiple scattering, and secondary particle production.

2.6 The reconstruction: from the bits to the decay

The reconstruction process at Belle II involves converting the raw data collected by the detector, *i.e.* the position and electric charge when it is relevant, also called low-level objects, into meaningful physics objects, such as tracks, clusters, and identified particles, also called high-level objects. This process is crucial for analysing the events and extracting the underlying physics. The BASF2 framework provides a comprehensive suite of tools for reconstructing the events from the raw data. This processing is done for both real data and simulated data, ensuring consistency between the two. For simulated data, one also has access to the true information, which provides details about the generated particles and their interactions with the detector, allowing for deeper analyses.

2.6.1 Tracks reconstruction

The first step in the reconstruction process is to identify and reconstruct the tracks of charged particles. This is done using the information from the VXD and CDC detectors. The hits in these detectors are clustered together to form track candidates by applying pattern recognition algorithms, a step called *track finding*. Different algorithms are used for VXD and CDC since the features of the hits are different in each detector. Both track candidates in the VXD and CDC are then merged according to the proximity of their extrapolated trajectories to form complete tracks. The new track candidates are then fitted using a Kalman filter algorithm, the deterministic annealing filter, to determine the particle's trajectory, momentum, and charge [62, 63].

2.6.2 Calorimeter reconstruction

The next step is to reconstruct the energy deposits in the ECL. The reconstruction starts with an array of energy deposits, from which we only take crystals with an energy deposit above a certain threshold (typically 10 MeV). All nearest neighbours with an energy deposit above 0.5 MeV are added, while merging the regions if two connected regions share common crystals. We then search for the local maximum in a given connected region, from which we consider the shower started, allowing us to compute the timing at which the particle arrived, as well as its energy. This is how clusters in the calorimeter are reconstructed. The potential tracks from the CDC are then extrapolated to the ECL before associating a given track candidate to a cluster in the ECL [50].

2.6.3 Charged particle identification

As specified in Section 2.3.1, the PID system provides useful information for charged particle identification. Moreover, for charged hadrons, measurements of the stopping power $\left(\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x}\right\rangle\right)$ from the SVD and CDC are also used, since it also depends on the particle momentum. For the electrons, additional information from the ECL is used, while for muons, we also rely on the dedicated KLM.

Each detector provides an independent likelihood for each charged particle hypothesis [3], which can then be combined into a global likelihood per hypothesis $i \in \{K, \pi, e, \mu, p, d\}$, written $\mathcal{L}_i$, and tested via a likelihood ratio test to determine the identity of the particle:

$$\mathrm{PID}_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j} \tag{2.3}$$

where j spans over all the charged particle hypotheses mentioned.

[3] *i.e.* kaon, pion, electron, muon, proton and deuteron

However, a recent approach using machine learning techniques has been developed to improve the charged particle identification performance at Belle II [64]. The idea is to use a Neural Network (NN) to which we give the log likelihoods for each charged particle hypothesis from each sub-detector, charge and track momentum as inputs, and the algorithm returns as outputs the probabilities for all hypotheses. The results shown in Figure 2.5 indicate an improvement in the performance of the kaon identification efficiency versus pion mis-identification rate when using the NN approach compared to the previous global likelihood method, even when training on all six hypotheses.

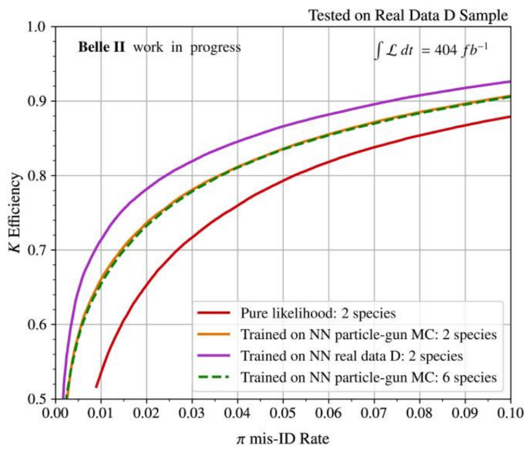


Figure 2.5: Kaon identification efficiency versus pion mis-identification rate for the global likelihood method (no NN), the NN trained on simulation and on data for kaon-pion hypotheses, as well as the NN trained on simulation but with all 6 hypotheses as output [64].

In the following analysis, we will use these new PIDs as recommended by the collaboration and since it performs very well on kaon and pion separation, which is important for our study.

2.6.4 Neutral particle identification

For neutral particles, namely photons and neutral hadrons, the identification is done using ECL information (and KLM for the K_L). First, every track matched to a cluster is removed from the list of clusters in the ECL, as they correspond to charged particles. The photon-hadron separation is then done by looking at the shower shapes and by using pulse shape discrimination. Indeed, the scintillation pulse is different for high-energy particles compared to photons, as one can see in Figure 2.6. This observation allows us to define a new variable for hadronic cluster identification [65]. The discrimination power of this new variable is shown in Figure 2.7, showing a clear discrimination of photons with respect to neutral hadrons.

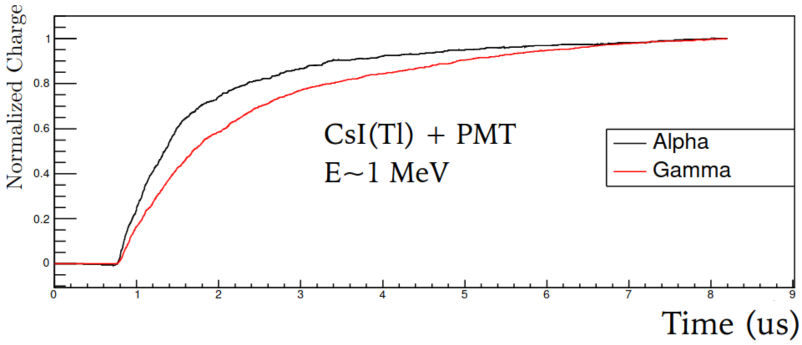


Figure 2.6: Comparison of the scintillation pulse shapes for photons and alpha particles in CsI(Tl) crystals [66].

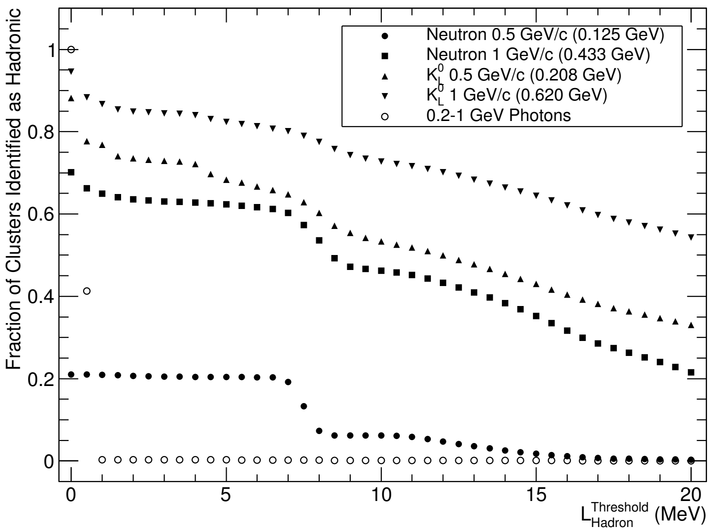


Figure 2.7: Hadron identification efficiency with the new cluster variable, for clusters with a light output above the threshold $L^{\text{Threshold}}_{\text{Hadron}}$, and asking for a total light output greater than 50MeV [65].

3

The how: analysis toolbox

Contents

In this chapter, we introduce all the tools required to carry out the $B \to K^{(*)}\nu\bar{\nu}$ analysis with the data and simulations obtained by the Belle II experiment. The first tool of interest is binary classifiers and, more specifically, Boosted Decision Tree (BDT), which are used to discriminate between two classes of events, generally signal and background (Section 3.1). Then, we introduce the `Full Event Interpretation` (FEI) algorithm (Section 3.2), which is used to reconstruct the B_{tag} meson in the event that allows us to do the tagging and infer properties on the B_{sig}. After that, we define useful variables to discriminate signal and backgrounds, which we will use as variables in our BDT (Section 3.3). For further studies and systematic uncertainties, we need to give some elements of theory on covariance matrices (Section 3.4). We then proceed to explain the concept of blind analysis (Section 3.5), and how this impacts the way we do ours. Finally, we explain the tools for signal extraction (Section 3.6) and how to do systematic inferences, such as limit determination (Section 3.7) and uncertainties estimation (Section 3.8).

3.1 Binary classification

For the analysis, we will need to perform binary classification. The obvious example will be for signal and background separation, but we will have other uses for it. Binary classification is the task of discriminating between two classes. For the generalities, let's call the class of a given event $\mathcal{C}$ and the two classes 0 and 1. Each event can be represented by a set of $N_v \in \mathbb{N}^*$ features $x \in \mathbb{R}^{N_v}$, and its class $\mathcal{C} \in \{0, 1\}$.

We want an algorithm that, given the features of an event, can predict its class $\mathcal{C}'$. In this analysis, we will use the BDT algorithm. These are a type of supervised learning algorithm, which means they need a training dataset and a *loss function* to optimise. The loss function quantifies the "distance" between the predicted class $\mathcal{C}'$ and the true class $\mathcal{C}$. Typically, for binary classification, the loss function is the binary cross-entropy, which is defined as [67]

$$L(\mathcal{C}, \mathcal{C}') = -\frac{1}{N} \sum_{i=1}^{N} \left[\mathcal{C}_i \log\left(\mathcal{C}'_i\right) + (1 - \mathcal{C}_i) \log\left(1 - \mathcal{C}'_i\right) \right], \tag{3.1}$$

This choice is motivated by the fact that it is the negative log-likelihood of the binomial distribution, which is exactly what we want to model. To discuss BDT specifically, we first need to introduce decision trees.

3.1.1 Decision tree algorithm

The decision tree algorithm is a simple algorithm that can be used for classification. It recursively splits the data into subsets in a dichotomous way, based on the value of a single feature at each node, generally obtained by fitting a simple model to the data [67]. This procedure is repeated until a stopping criterion is met, such as a maximum depth of the tree or when no further splits are beneficial. The terminal nodes of the tree are called leaves, and they are assigned a weight computed from the training samples that reach that leaf. An illustration of a decision tree is shown in Figure 3.1.

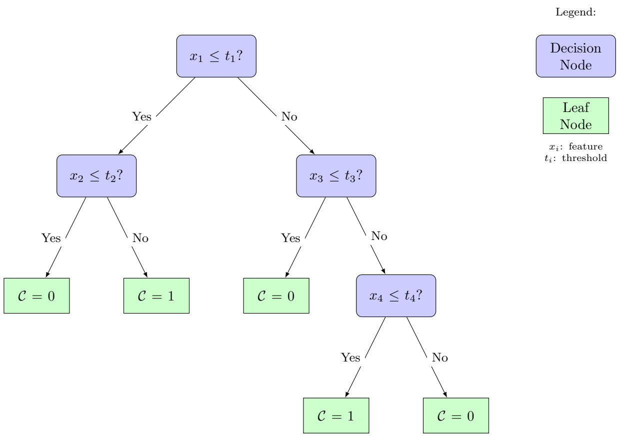


Figure 3.1: Schematic representation of a decision tree. Each node represents a split of the data based on the value of a single feature.

Finally, for a given event with features x, the estimated class $\mathcal{C}'$ is a function of the weight $w(x) \in \mathbb{R}$ of the leaf that x falls into. Typically, this function is a sigmoid function:

$$\mathcal{C}'(x) = \frac{1}{1 + e^{-w(x)}} \tag{3.2}$$

which is between 0 and 1, and can be properly interpreted as a probability.

However, decision trees are not very powerful on their own, and they are prone to overfitting, *i.e.* being too fine-tuned to the training data. This is why the BDT algorithm comes in.

3.1.2 *Boosted decision trees*

Boosting is a technique that combines multiple *weak learners* to build a stronger learner. In the case of boosting, this is done in a sequential way, meaning that each new learner is trained to correct the residual errors of the previous learner. Moreover, data are weighted so that the misclassified events are given more importance in the next iteration. The final prediction is then a weighted sum of the predictions of all the learners. In the case of BDT, the weak learners are decision trees.

In the following, we will not use BDT strictly speaking but *gradient-boosted* decision trees from the `XGBoost` library [68]. Instead of simply passing the residuals to the next learner, this algorithm passes both the gradient and the Hessian of the loss function.

3.1.3 *Performances*

There are two main performance metrics we want to look at for the BDT. The first one is the quality of the classification task itself.

The second one is the overfitting, which is the phenomenon where the model performs well on the training data but poorly on unseen data.

For the classification task[1], there are many possible metrics. All of them are defined via the *confusion matrix*, which is a 2×2 matrix that contains the number of true positives TP, true negatives TN, false positives FP and false negatives FN. The metric of interest for us is the F_1 score:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{3.3}$$

[1] i.e. the qualification of the event as a "background" or a "signal"

For overfitting, use a *test* dataset, which is not used for training and is used to evaluate the model's performance. If the performance on the test dataset is significantly worse than the performance on the training dataset, conclude that there is overfitting.

3.1.4 *Hyperparameters*

The parameters defining the structure and the training of the BDT are called hyperparameters. The hyperparameters of interest for this analysis are the following:

- The maximum depth of the trees (`max_depth`)
- The number of trees to be built (`n_estimators`)
- The learning rate, which controls how much each new tree contributes to the final prediction (`learning_rate`)
- The subsample ratio of the training dataset, which controls the fraction of the training data that is used to grow each tree, to limit overfitting (`subsample`)
- The minimum loss reduction required to make a further partition on a leaf node of the tree (`gamma`)
- Regularisation parameter to control the complexity of the trees and prevent overfitting (`reg_lambda`)

As each of these hyperparameters controls both the training time and the performance of the model, such as overfitting, we need to properly tune them. This is done by using the `Optuna` package [69], and by optimising the hyperparameters to maximise the F_1 score.

3.1.5 *Feature importance*

We use the *Shapley values* [70] and the `SHAP` package [71] for feature importance.

The Shapley values are a concept from cooperative game theory, answering the question: how to fairly reward players in a coalition for their contributions? If we consider a set N of $n \in \mathbb{N}$ players, a reward function v and a coalition of players $S \subseteq N$, the Shapley value of a player $i \in S$ is defined as:

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] \tag{3.4}$$

This can be interpreted as the average contribution of player i to all possible coalitions of players. To apply it to BDTs, we can simply replace the players with the features of our model, and the reward with the model's prediction. The higher the Shapley value, in absolute value, the greater the impact of an input feature for the BDT to decide if the event is more "signal-like" or "background-like". The sign then indicates if it is the former or the latter. The ranking is done by computing the correlation between the feature values and their Shapley values, in other words, how many times did a feature was decisive when the BDT made a decision. This is best seen in "*beeswarm*" plots, where all the training data points are shown, feature per feature, with the correlation between the Shapley value and the actual feature value. Examples are shown in Figure 5.4, Figure 5.5 and Figure 5.7.

This then gives us a way to quantify the contribution of each feature to the model's prediction in a way that is both reliable and explainable.

3.2 *Full Event Interpretation*

As explained in Section 2.1, the knowledge of the four-momentum of the $\Upsilon(4S)$ and the presence of only two B mesons in the event allow us to constrain the kinematics of the signal decay by reconstructing the B_{tag}. Any reconstructed object, *i.e.* tracks and clusters, not assigned to the B_{tag} should thus be attributed to the B_{sig}. At Belle II, this *tagging* is performed by the FEI algorithm [72].

Tagging refers to the process of reconstructing the B_{tag} and inferring the properties of the B_{sig}. This is an *exclusive* approach by construction, as the B_{tag} is reconstructed in specific decay channels. The other type of approach is called *inclusive* "tagging", even if it is not really tagging as it does not reconstruct the B_{tag}. There are two types of exclusive tagging: *hadronic* and *semileptonic*. Both approaches are possible with the FEI.

The algorithm is built on a hierarchical six-stage approach, as illustrated in Figure 3.2. It starts with tracks, clusters, and vertices information, which are further used to reconstruct the final state particles, such as $\pi^{\pm}$, $K^{\pm}$, $\gamma \dots$ Then intermediate particles are reconstructed with these final state particles, and it iterates with six stages in total until the B meson reconstruction. Each candidate can come from multiple decay channels, *e.g.* B^+ may come from $B^+ \to D^0\pi^+$ or $B^+ \to D^0\pi^+\pi^0$, and the same for the D^0 and so on. Each candidate is assigned a probability of being properly reconstructed, called `SignalProbability`, estimated by a multivariate classifier. For each possible final state particle and decay channel of a given intermediate state particle, a BDT is trained. Following the hierarchical approach, the output of the classifiers of the previous stage is given as input to the next stage until a single probability is assigned to each B candidate.

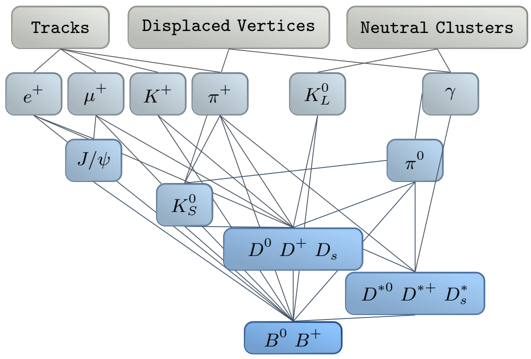


Figure 3.2: Schematic overview of the FEI six stages. Each node represents a multivariate classifier, and each edge represents a relationship between them. This builds a complete network of classifiers [72].

There are three main figures of merit for the FEI:

- **Tagging efficiency**: the fraction of $\Upsilon(4S)$ events which can be tagged, *i.e.* for which a B_{tag} candidate is reconstructed;
- **Tag-side efficiency**: the fraction of $\Upsilon(4S)$ events for which the tag is correct;
- **Tag-side purity**: the fraction of tagged $\Upsilon(4S)$ events with a correct tag.

In theory, accounting for all the intermediate particles and the combinatorics, the FEI can reconstruct $\mathcal{O}\left(10^5\right)$ different decay chains. However, in the context of this analysis, only some *semileptonic* decay channels of the B_{tag} are considered: $B \to D^{(*)}\ell\nu_\ell$ and $B \to D^{(*)}\ell\nu_\ell\pi^0$ with $\ell \in \{e, \mu\}$. The tag-side efficiency for the FEI on these semileptonic channels is about 1.8% [72].

Each tagging approach (hadronic, semileptonic and inclusive) offers advantages and disadvantages. This can be reduced to a trade-off between tag-side purity, the probability of the B_{tag} being correct, and tag-side efficiency, due to the kinematic constraints. This is summarised in Figure 3.3.

3.3 *Discrimating variables*

In this section, we will describe important variables for the analysis. These variables will be useful to discriminate between signal and background, either by themselves or used as features for a BDT.

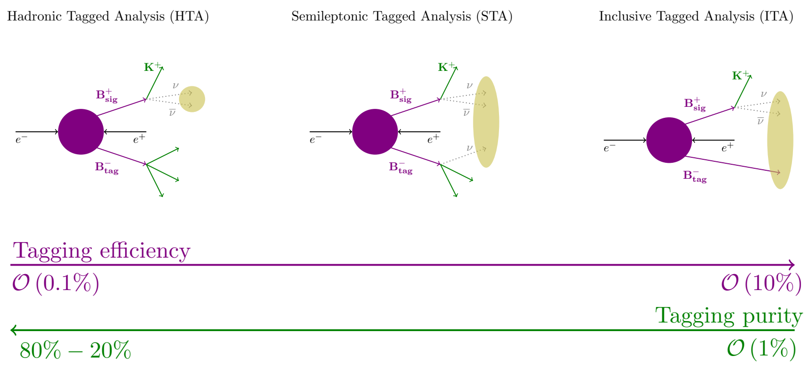


Figure 3.3: Tagging efficiency vs purity for the different tagging approaches. The hadronic approach offers the highest purity and the highest kinematic constraints, while the inclusive approach offers the highest efficiency. The semileptonic approach is in between, offering a good compromise between the two.

3.3.1 *Angular distribution of a semileptonic decay*

As specified in Section 3.2, this analysis relies on semileptonic decays of the B_{tag}. Hence, we will have a system of the type $B \rightarrow Y + \nu$ where Y is the non-neutrino system, in our case $D^{(*)}\ell$. Because of energy and momentum conservation:

$$(p_B - p_Y)^2 = m_\nu^2 \tag{3.5}$$

where p_B and p_Y are the four-momenta of the B_{tag} and the hadronic system, respectively, and m_ν is the mass of the neutrino. Let's consider $m_\nu = 0$. Then Equation (3.5) becomes:

$$m_B^2 + m_Y^2 - 2E_B E_Y + 2|\vec{p}_B||\vec{p}_Y| \cos(\theta_{BY}) = 0 \tag{3.6}$$

where $\cos(\theta_{BY})$ is the cosine of the angle between the directions of the B_{tag} and the hadronic system. Rearranging the equation, we can express $\cos(\theta_{BY})$ as:

$$\cos(\theta_{BY}) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|} \tag{3.7}$$

This variable is very useful to discriminate $D\ell\nu$ events from backgrounds, because for events where the missing particle is not a neutrino, $\cos(\theta_{BY})$ can take unmathematical values (i.e. outside the range $[-1, 1]$), since $(p_B - p_Y)^2$ would not be equal to zero. However, due to resolution effects, even for signal events, this variable can take values outside the range $[-1, 1]$.

3.3.2 *Topological variables*

$B\bar{B}$ events have a very different topology from continuum events. Indeed, $B\bar{B}$ events are produced almost at rest in the $\Upsilon(4S)$ frame, since the $\Upsilon(4S)$ mass is just above the $B\bar{B}$ production threshold, so their decays are isotropic. While for $q\bar{q}$ events, since they are produced with a higher momentum, they have a more jet-like topology. This is illustrated in Figure 3.4. Hence, variables based on the event topology, such as *Fox-Wolfram* moments [73], are very useful for signal and background separation.

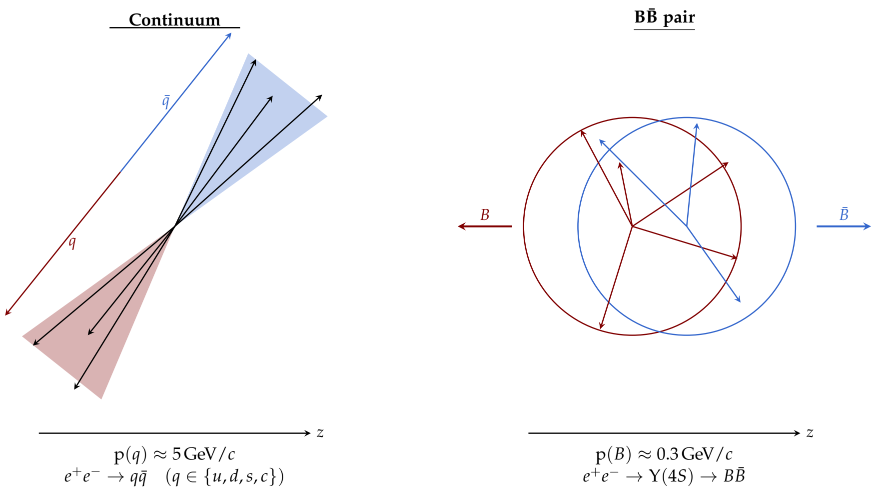


Figure 3.4: Schematic representation of the topology of a $B\bar{B}$ event and a $q\bar{q}$ event in the $\Upsilon(4S)$ frame. The $B\bar{B}$ event has an isotropic topology, while the $q\bar{q}$ event has a more jet-like topology.

Fox-Wolfram moments are defined, for N particles with momenta $(\vec{p}_i)_{i\in[\![1,N]\!]}$, as [74]:

$$H_l = \sum_{(i,j)\in[\![1,N]\!]^2} |\vec{p}_i||\vec{p}_j| P_l\left(\cos\left(\theta_{ij}\right)\right) \tag{3.8}$$

where P_l is the Legendre polynomial of order $l \in \mathbb{N}$, and θ_{ij} is the angle between the momenta of particles i and j. A useful variable directly derived from the Fox-Wolfram moments is the ratio

$$R_2 = \frac{H_2}{H_0} \tag{3.9}$$

which will be used in the analysis for continuum suppression.

To further improve the discrimination power of the Fox-Wolfram moments, the Belle collaboration has developed a modified version of them using Fisher discriminants, called Kakuno-Super-Fox-Wolfram (KSFW) variables [74]. To define them, we first need to separate the particles in the event into two categories: B_{sig} daughters (denoted s, for signal) and the Rest Of Event (ROE), *i.e.* the particles left after the B_{sig} reconstruction (of symbol o, for other). There are then 2 possibilities:

- so moments, between the B meson daughters and the ROE
- oo moments, between the particles in the ROE

There are then 2 categories of ROE particles: charged c and neutral n. To those, we can add a third category corresponding to the missing momentum of the event m[2]. The *so* KSFW variables are then defined, for a particle of category $k \in \{c, n, m\}$, as [75]:

[2] obtained by subtracting the momentum of all reconstructed particles from the total momentum

$$H_{k,l}^{so} = \frac{1}{Z} \sum_{i \in s} \sum_{j \in k} C_{ij}^{l} |\vec{p}_j| P_l \left(\cos\left(\theta_{ij}\right)\right) \tag{3.10}$$

where:

- $Z = 2\left(\sqrt{s} - E_B^*\right)$ is a normalisation factor, with $\sqrt{s}$ the centre-of-mass energy and E_B^* the energy of the B meson in the centre-of-mass frame
- $C_{ij}^{l} \in \{-1, 0, 1\}$ is the product of the charge of particle i and the charge of particle j if l is odd, and 1 otherwise

and the rest of the notations are the same as in Equation (3.8). Similarly, the *oo* KSFW variables are defined as:

$$H_{k,l}^{oo} = \frac{1}{Z^2} \sum_{i \in k} \sum_{j \in k} C_{ij}^{l} |p_i||p_j| P_l \left(\cos\left(\theta_{ij}\right)\right) \tag{3.11}$$

We can add another type of topological variables based on the *thrust* of the particles. The thrust axis $\vec{T}$ is defined as the unit vector along which the projection of the momenta of the particles is maximised [74]. From this thrust axis we can define `cosTBTO`, defined between 0 and 1, which is the cosine of the angle between the thrust axis of the B meson candidate and the thrust axis of the ROE. For $B\bar{B}$ events, the thrust axis of the B meson candidate and the thrust axis of the ROE are uncorrelated, so `cosTBTO` is expected to be uniformly distributed[3]. On the other hand, for $q\bar{q}$ events, the two axes are more likely to be aligned, so `cosTBTO` is expected to peak at 1. This clear separation between the two types of event makes this variable very useful for continuum reduction.

[3] as it is distributed on a sphere parametrised by $(r, \cos(\theta), \phi)$, so the uniformity will be on the cosine of the angle, not the angle itself

3.4 *Covariance matrices*

For future studies, we will need to use the covariance matrix and some notions from random matrix theory. Let's imagine we have m events with n features each. The covariance matrix of the random array $X = (x_1 \dots x_m)$, with $x_i \in \mathbb{R}^n$ for $i \in [\![1, m]\!]$, is a $n \times n$ matrix defined as:

$$\Sigma = \frac{1}{m} \left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T \right] \tag{3.12}$$

Since Σ is a symmetric matrix with real elements, the spectral theorem applies, and we can diagonalise it. Hence, there exists an orthogonal matrix P, of unit eigenvectors $(p_i)_{i\in[\![1,n]\!]}$, and a diagonal matrix S^2 of ordered (from smallest to largest) eigenvalues $\left(\sigma_i^2\right)_{i\in[\![1,n]\!]}$ such that:

$$\Sigma = PS^2P^T = \sum_{i=1}^{n} \sigma_i^2 p_i p_i^T \tag{3.13}$$

This decomposition is typical of *principal component analysis*, which is a technique used to reduce the dimensionality of a dataset while preserving as much of the variance as possible. If the eigenvalues are distributed evenly, we can only keep t eigenvectors with the largest eigenvalues, considering only the diagonal elements for the others:

$$\Sigma \approx \sum_{i=n-t}^{n} \sigma_i^2 p_i p_i^T + \operatorname{diag}\left(\sum_{i=1}^{n-t} \sigma_i^2 p_i p_i^T\right) \tag{3.14}$$

The number t of eigenvectors to keep can be defined either arbitrarily or as the number of eigenvalues above a certain threshold σ_{th}^2. The threshold value can be computed via the *Marchenko-Pastur* distribution, which is the limiting distribution of the eigenvalues of a covariance matrix when the number of samples m and the dimension n go to infinity [76]:

$$\sigma_{\text{th}}^2 = \sigma^2 \left(1 + \sqrt{\frac{n}{m}}\right)^2 \tag{3.15}$$

where σ^2 is the median of the eigenvalues of the covariance matrix.

3.5 *Blind analysis*

To avoid bias from the analyst, it is usual to perform a *blind* analysis [74]. Many blinding techniques exist. For this analysis, we develop the procedure using only simulated data. This way, we can optimise the strategy, from the reconstruction to the signal extraction, without biasing it. However, this exposes us to the risk of overfitting the analysis to the simulated data, which may not perfectly represent the real data. To mitigate this risk, we can use *control samples*, *e.g. sidebands*, which are datasets for which we do not expect to find any signal, and on which we can control any discrepancy between data and simulation. From these control samples, we can derive corrections to apply to the simulation, and associate systematic uncertainties with them. Our control samples are defined in Section 5.3. After checking the validity of our analysis strategy, with various test cases, we can finally unblind the data and perform the measurement.

3.6 *Signal extraction*

After optimisation of the analysis strategy, the signal extraction can be performed. The *parameter of interest* for us is the branching ratio of $B^+ \to K^+\nu\bar{\nu}$, but we prefer to express it as the *signal strength* μ:

$$\mu = \frac{\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right)}{\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right)_{\text{SM}}} \tag{3.16}$$

The signal extraction is performed using a *binned maximum likelihood fit* to the BDT output distribution, above a certain threshold that we will define later. Since building a parametric statistical model can be difficult in this analysis, we will construct a *template*. The principle is to build a histogram of the BDT output distribution for the signal and each background contribution from simulation. Some parameters of the model are constrained from this template, and then we perform a fit to the data to get the parameter of interest [77]. The template model is based on the `HistFactory` formalism [78], and the software used is `pyhf` [79].

Let's denote $N_b \in \mathbb{N}^*$ the number of bins in the BDT output distribution, and ν_i the number of expected events in the i-th bin, derived from simulation, which is typically the sum of the signal and $n_b \in \mathbb{N}^*$ background contributions:

$$\nu_i\left(\mu,\theta\right) = \sum_{c\in\{s,b_1,\dots,b_{n_b}\}} \nu_{i,c}\left(\mu,\theta\right) \tag{3.17}$$

with $\nu_{i,c}\left(\mu,\theta\right)$ the expected number of events in the i-th bin for contribution c and $\theta \in \mathbb{R}^N, N \in \mathbb{N}^*$, is a vector of *nuisance parameters*. Nuisance parameters are parameters that are not of interest, but can affect the measurement of the parameter of interest and have uncertainties. Among those nuisance parameters, we find n_m multiplicative nuisance parameters θ' and n_a additive nuisance parameters $\tilde{\theta}$:

$$\theta = \left(\theta'_1, \dots, \theta'_{n_m}, \tilde{\theta}_1, \dots, \tilde{\theta}_{n_a}\right) \tag{3.18}$$

We can then rewrite Equation (3.17) as:

$$\nu_i\left(\mu,\theta\right) = \sum_{c\in\{s,b_1,\dots,b_{n_b}\}} K_c\left(\theta'\right)\left(\nu^0_{i,c} + \Delta_{i,c}\left(\tilde{\theta}\right)\right) \tag{3.19}$$

where $\nu^0_{i,c}$ is the *nominal* expected number of events in the i-th bin for contribution c. $K\left(\theta'\right)$ and $\Delta_{i,c}\left(\tilde{\theta}\right)$ are multiplicative and additive variations of the expected number of events in the i-th bin for contribution c due to the nuisance parameters. We can make the dependence in θ' and $\tilde{\theta}$ explicit:

$$K_c\left(\theta'\right) = \prod_{j=1}^{n_m} \theta'_j \kappa_c^j \tag{3.20}$$

$$\Delta_{i,c}\left(\tilde{\theta}\right) = \sum_{j=1}^{n_a} \tilde{\theta}_j \delta_{i,c}^j \tag{3.21}$$

with κ_c^s and $\delta_{i,c}^j$ being the variation of the expected number of events in the i-th bin for contribution c due to specific nuisance parameter. The variations are derived from *auxiliary measurements*, *i.e.* additional measurements done in different experimental conditions specifically to constrain the nuisance parameters. For example, it could be the histogram of the BDT output with a higher or lower number of a given decay mode to constrain the nuisance parameter associated with the branching fraction of this mode. Depending on their value, the additive nuisance parameters can be:

- **Uncorrelated** if there is one $\tilde{\theta}_j$ for each contribution c and each bin i
- **Bin-correlated** if there is one $\tilde{\theta}_j$ for each contribution c and all bins i
- **Contribution-correlated** if there is one $\tilde{\theta}_j$ for several contributions c and each bin i

Similarly, the multiplicative nuisance parameters can be:

- **Uncorrelated** if there is one θ'_j for each contribution c
- **Contribution-correlated** if there is one θ'_j for several contributions c

Now, let's call $(n_i)_{i\in[\![1,N_b]\!]}$ the number of observed events in the i-th bin. Then, the likelihood of our model (μ, θ) after observing $(n_i)_{i\in[\![1,N_b]\!]}$ is:

$$\mathcal{L}\left(\mu, \theta | \left(n_i\right)_{i\in[\![1,N_b]\!]}\right) = \frac{1}{Z} \prod_{i=1}^{N_b} \text{Pois}\left(n_i | \nu_i\left(\mu, \theta\right)\right) \pi\left(\theta\right) \tag{3.22}$$

with $\text{Pois}\left(n_i | \nu_i\left(\mu, \theta\right)\right)$ the Poisson probability of observing n_i events given an expectation of $\nu_i\left(\mu, \theta\right)$ events, $\pi\left(\theta\right)$ the prior probability density function of the nuisance parameters θ and Z a normalisation factor. While each nuisance parameter can have its own prior distribution, we will assume that they are all independent and follow a standard normal distribution, giving the total prior:

$$\pi\left(\theta\right) = \prod_{j=1}^{N-n_b} \text{Gauss}\left(\theta_j | 0, 1\right) \prod_{c=1}^{n_b} \text{Gauss}\left(\mu_c | 1, \sigma_c\right) \tag{3.23}$$

with Gauss $(x|\mu,\sigma)$ the Gaussian probability density function of x with mean μ and standard deviation σ. For the background strength, the σ_c are hyperparameters of our statistical model.

Finally, the signal strength μ can be estimated using a profile likelihood method. The idea is to maximise the likelihood with respect to the nuisance parameters θ for each value of μ, and then find the value of μ that maximises the profile likelihood:

$$\hat{\mu} = \underset{\mu}{\operatorname{argmax}} \left\{ \max_{\theta} \left[\mathcal{L}\left(\mu, \theta | \left(n_i\right)_{i \in [\![1, N_b]\!]}\right) \right] \right\} \tag{3.24}$$

Technically, we will actually minimise twice the negative log-likelihood, which is more convenient for numerical optimisation:

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \left\{ \min_{\theta} \left[-2 \ln \left(\mathcal{L}\left(\mu, \theta | \left(n_i\right)_{i \in [\![1, N_b]\!]}\right) \right) \right] \right\} \tag{3.25}$$

3.7 *Limits determination*

The following section is based on reference [80].

After obtaining an estimate of the signal strength $\hat{\mu}$, we can test its compatibility with various hypotheses such as: "*background-only*", $\mu = 0$, or SM, $\mu = 1$.

To test this compatibility, the best test statistic, as proved by the Neyman-Pearson lemma [81], is the profile likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\theta}_\mu | \left(n_i\right)_{i \in [\![1, N_b]\!]}\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\theta} | \left(n_i\right)_{i \in [\![1, N_b]\!]}\right)} \tag{3.26}$$

where $\hat{\theta}_\mu$ maximises the likelihood for a given value of μ, and $\hat{\theta}$ is the profile likelihood estimate of the nuisance parameters for the best-fit value $\hat{\mu}$. From this definition, it is clear that $0 \leq \lambda(\mu) \leq 1$, and that the closer $\lambda(\mu)$ is to 1, the more compatible the data are with the hypothesis μ.

As specified in the previous section, we actually prefer to work with the negative log-likelihood, so we can define a new test statistic:

$$t_\mu = -2 \ln (\lambda(\mu)) \tag{3.27}$$

for which higher values correspond to less compatibility with the hypothesis μ. The p-value of the hypothesis μ can then be computed as:

$$p_\mu = P\left(t_\mu \geq t_{\mu,\text{obs}} \mid \mu\right) = \int_{t_{\mu,\text{obs}}}^{+\infty} f\left(t_\mu | \mu\right) dt_\mu \tag{3.28}$$

where $t_{\mu,\text{obs}}$ is the observed value of the test statistic, and $f\left(t_\mu|\mu\right)$ is the probability density function of t_μ under the hypothesis μ. The density function $f\left(t_\mu|\mu\right)$ is typically approximated using Wilk's theorem [82], which states that in the asymptotic limit, t_μ follows a chi-squared distribution with one degree of freedom. Then, if the p-value is smaller than a predefined threshold α, the hypothesis μ is rejected at a CL of $1-\alpha$. From this, we can define the *confidence interval* of μ at a CL of $1-\alpha$ as the set of all values of μ that are not rejected at a CL of $1-\alpha$:

$$\text{CI}_{1-\alpha} = \left\{\mu|p_\mu > \alpha\right\} \tag{3.29}$$

In the case where no significant signal is observed, we can instead set an *upper limit* on μ at a CL of $1-\alpha$. We just need to do the same development as with the test statistic t_μ but with the following test statistic:

$$q_\mu = \begin{cases} t_\mu & \text{if } \hat{\mu} \leq \mu \\ 0 & \text{otherwise} \end{cases} \tag{3.30}$$

Replacing t_μ with q_μ in equation 3.28 gives us the p-value of the upper limit, and the upper limit at a CL of $1-\alpha$ is the largest value of μ such that $p_\mu \leq \alpha$.

Moreover, to give a more physically meaningful result when setting an upper limit with μ going to zero (i.e. when the signal is very small), we can use the *CLs method* [83]. The idea is to define a new test statistic from the p-values of the signal-plus-background and background-only hypotheses:

$$\text{CL}s = \frac{P\left(q_\mu \geq q_{\mu,\text{obs}}|\mu\right)}{P\left(q_\mu \geq q_{\mu,\text{obs}}|0\right)} \tag{3.31}$$

Then, the upper limit at a CL of $1-\alpha$ is the largest value of μ such that $\text{CL}_s \leq \alpha$.

Finally, the significance is also a commonly used quantity to quote when giving results from test statistics. It is defined as the number of standard deviations that a Gaussian distribution would need to fluctuate to give a p-value equal to the one obtained from the test statistic:

$$Z = \Phi^{-1}\left(1-p_\mu\right) \tag{3.32}$$

with Φ^{-1} the quantile function of a standard Gaussian distribution. Conventionally, a significance of $Z=3$ is considered as *evidence* for a signal, while a significance of $Z=5$ is considered as a *discovery*.

3.8 Uncertainties estimation

It is important to estimate the statistical and systematic uncertainties of the measurement. To do so, we use the method proposed in [84]. We need to define two types of datasets first: the *measurements* $\vec{n}$, which are the results of the experiment, so the N_b bins of the histogram in our case; and the *auxiliary measurements* $\vec{a}$ which are measurements made especially for the nuisance parameters. The idea is to build the total uncertainty σ_μ^2 as a sum in quadrature of the statistical and systematic uncertainties. As μ is a function of $\vec{n}$ and $\vec{a}$, using the uncertainty propagation formula and assuming no correlation between the elements of $\vec{n}$ with themselves and similarly for $\vec{a}$, we can write [84]:

$$\sigma_{\hat{\mu}}^2 = \sum_i \left(\frac{\partial \mu}{\partial n_i}\right)^2 \sigma_{n_i}^2 + \sum_j \left(\frac{\partial \mu}{\partial a_j}\right)^2 \tag{3.33}$$

where the first term corresponds to the statistical uncertainty and the second term to the systematic uncertainty. We can see the partial derivatives as the *sensitivity* of $\hat{\mu}$ to the observables. So to compute the contribution of an uncertainty source to the total uncertainty, one can:

1. Shift the corresponding measurement or auxiliary measurement by one standard deviation;
2. Repeat the fit leaving everything free;
3. Compute the difference between the new value of $\hat{\mu}$ and the nominal value.

For example, to compute the contribution of a Gaussian-distributed nuisance parameter with auxiliary measurement a_i, we can replace it in our statistical model by $a_i + \sigma_{a_i}$ and repeat the fit with no other change. This will give us a new $\hat{\mu}^{[a_i]}_{\text{var, }+}$, and then

$$\Delta\hat{\mu}^{[a_i]}_{+} = \hat{\mu}^{[a_i]}_{\text{var, }+} - \hat{\mu}_{\text{nominal}} = \sigma_\mu^{[a_i]} \tag{3.34}$$

This can also be done with a negative shift, and the final contribution of this uncertainty source is the average of the absolute values of the positive and negative shifts:

$$\sigma_\mu^{[a_i]} = \frac{\left|\Delta\hat{\mu}^{[a_i]}_{+}\right| + \left|\Delta\hat{\mu}^{[a_i]}_{-}\right|}{2} \tag{3.35}$$

By summing all the contributions in quadrature, we can get the total uncertainty $\sigma_{\hat{\mu}}$.

4

Intermezzo: calibration of the `Flavour Tagger`

Contents

As presented in Section 1.1, CP-violation is a key ingredient to explain the matter-antimatter asymmetry in the universe. The Belle II experiment, thanks to its design described in Section 2.3, is well suited to perform precise measurements of CP-violation in the quark sector. Indeed, as we explain more in detail in Section 4.1, to measure CP-violation in B mesons, one needs to know the *flavour*, *i.e.* whether it is an B^0 or $\bar{B}^0$ meson, at the time of its production. In the following chapter, we present an algorithm called `GFlaT` which is used to determine the flavour of B mesons to perform CP-violation studies. In Section 4.1, we define the main quantity for CP-violation studies at Belle II as well as the difficulties in measuring it. This will lead us to the description of *flavour tagging* algorithms in Section 4.2. To be used, these algorithms need to be calibrated, which is the topic of Section 4.3, and in Section 4.4 we discuss the challenges and the upgrades to the calibration of the `GFlaT` algorithm. Finally, in Section 4.5, the results of the calibration with the upgrades and the prospects for flavour tagging at Belle II are presented.

4.1 CP-violation studies at Belle II

There are three types of CP-violation:

- in *decay*, also called direct CP-violation, occurs when a particle's branching fraction to a final state differs

  from that of its antiparticle to the corresponding final state.

- in *mixing*, happens when the rate of oscillation from the B^0 to $\bar{B}^0$ is different from the rate of oscillation from the $\bar{B}^0$ to B^0.

- in *interference* between decay and mixing, the combination of the other two, which can occur when a neutral particle and its antiparticle both share a common final state.

Analyses at Belle II can be performed to study these different types of CP-violation. The phenomenon is related to the flavour changing of the quarks via weak interaction. In the SM, the quark mixing and the CP-violation are parameterised via the three angles and the one phase of the CKM matrix, as described in Section 1.1, thus the importance for precise measurement of those parameters. There are various ways to study and quantify CP-violation. One of them is to measure the CP-asymmetry. The CP-asymmetry in the direct CP-violation is given by:

$$a^{CP} = \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \tag{4.1}$$

where N_{B^0} and $N_{\bar{B}^0}$ are the number of B^0 and $\bar{B}^0$ mesons at the time of their production, respectively. Hence, the problem is now to determine this flavour. An easy case is when the decay of the B meson is *self-tagged*, *i.e.* when a final state particle can give us almost surely the flavour of the decaying B meson. For example, in the decay $B \to D\pi$, the electric charge of the pion can give us the flavour of the decaying B, as one can see from the quark content in Figure 4.1.

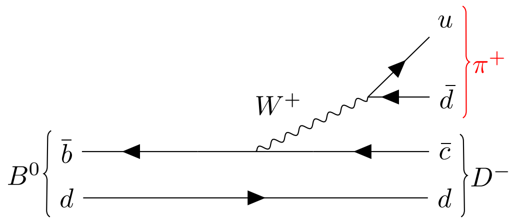


Figure 4.1: Feynman diagram of the $B^0 \to D^-\pi^+$ decay. The charge of the pion can give us the flavour of the initial B meson since only the $\bar{b}$ quark changes to a $\bar{c}$ quark and the d quark is a spectator.

However, in most cases, the decay we want to study is not self-tagged, and so we need to find another strategy to determine the flavour of the B meson. To do so, we can exploit another property of the collisions at Belle II: the B mesons are *entangled*. Indeed, the B mesons are produced in pairs from the decay of the $\Upsilon(4S)$ resonance, which is a bound state of a b and $\bar{b}$ quark. Thus, we can infer properties of the B_{sig} from the B_{tag}, in particular its flavour. Hence, if the B_{tag} decays in a self-tagged mode, we can determine its flavour and thus the flavour of the B_{sig} at the same moment.

We can classify the self-tagged modes in 13 different categories, shown in Table 4.1. The "target" is the particle whose charge gives us the flavour of the decaying B meson. The kinematic information is also relevant to determine the process that produced the target particle. For example, the $b \to c(u)\ell^-\bar{\nu}_\ell$ transitions will produce a statistically higher momentum ℓ^- (Electron, Muon and Kinetic Lepton categories), while $b \to c \to s(d)\ell^+\nu_\ell$ transitions will generate a statistically lower momentum ℓ^+ (Intermediate Electron, Intermediate Muon and Intermediate Kinetic Lepton categories). This categorisation makes the usage of a *multivariate classifier* a suitable choice for a *flavour tagging algorithm*.

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	

Table 4.1: Categories of self-tagged modes and their corresponding targets. It is to be noted that the categories can overlap, *e.g.* the "Kinematic Lepton" category with both the "Electron" and "Muon" categories. "Intermediate" indicates that the lepton is a secondary lepton, not coming directly from a B. "Fast Hadrons" are kaons and pions from a W meson (*e.g.* in $b \to cW^-$ transitions). "Maximum p" searches for high momentum particles. "Fast-Slow-Correlation" is for event with both a slow pion and a high momentum particle (*e.g.* $\bar{B}^0 \to D^{*+}W^-$). Recreation from [85].

4.2 *Flavour tagging algorithms*

In order to understand how the `flavour tagger` algorithm works, we need to introduce the quantities used to estimate its performance first. The first one is the *tagging efficiency* ε:

$$\varepsilon = \frac{N_{B^0}^{\text{tag}} + N_{\bar{B}^0}^{\text{tag}}}{N_{B^0} + N_{\bar{B}^0}} \tag{4.2}$$

where $N_{B^0}^{\text{tag}}$ and $N_{\bar{B}^0}^{\text{tag}}$ are the number of B^0 and $\bar{B}^0$ mesons that are tagged, respectively. Of course, the flavour tagging algorithm can also make mistakes, and thus we need to introduce the *wrong tag fraction* w defined as:

$$N_{B^0}^{\text{tag}} = \varepsilon\left(\left(1-w\right)N_{B^0} + wN_{\bar{B}^0}\right) \tag{4.3}$$

$$N_{\bar{B}^0}^{\text{tag}} = \varepsilon\left(wN_{B^0} + \left(1-w\right)N_{\bar{B}^0}\right) \tag{4.4}$$

Hence, we can write the "observed" CP-asymmetry, obtained from the `flavour tagger` algorithm, as:

$$\begin{aligned} a_{\text{obs}}^{CP} &= \frac{N_{B^0}^{\text{tag}} - N_{\bar{B}^0}^{\text{tag}}}{N_{B^0}^{\text{tag}} + N_{\bar{B}^0}^{\text{tag}}} \\ &= (1-2w)\,\frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} \\ &= (1-2w)\,a^{CP} \end{aligned} \tag{4.5}$$

The factor in front of the real CP-asymmetry is called the *dilution factor* $r \in [0,1]$:

$$r = |1-2w| \tag{4.6}$$

It can be interpreted like:

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

Any dilution factor below 1 means that the algorithm gives us some information, but not the full one.

With these definitions, we can now introduce the `flavour tagger` algorithm. It comprises one multivariate classifier for each of the 13 categories defined in Table 4.1, which are trained to determine the probability for the given particle to be the target particle of the self-tagged category. After reconstruction of the B_{sig}, the remaining clusters and tracks are mainly from the B_{tag}. The first step of the algorithm is thus to reconstruct the particles from the information in the ROE and to determine for each of those particles the categories it belongs to, as well as to compute the probability described above. After this, the particle with the highest probability is selected as the target candidate and is given to another BDT which is trained to determine the dilution factor r for this candidate and the predicted flavour of the B_{tag}. The final output of the algorithm is the dilution factor r times the flavour q [85]. A schematic view of the algorithm flow is shown in Figure 4.2.

This was the first flavour tagging algorithm used at Belle II, also called *category-based*. In 2024, a second algorithm, called `GFlaT`, was developed later on [86]. Instead of BDTs, it uses a graph neural network architecture by assimilating the problem as a cloud of particles. Otherwise, the output is the same as for the category-based algorithm. The comparison of the qr distribution for the two algorithms is shown in Figure 4.3. We can see that the `GFlaT` algorithm gives us a better separation between the different values of qr, which is a good indication of its better performance. Especially, we

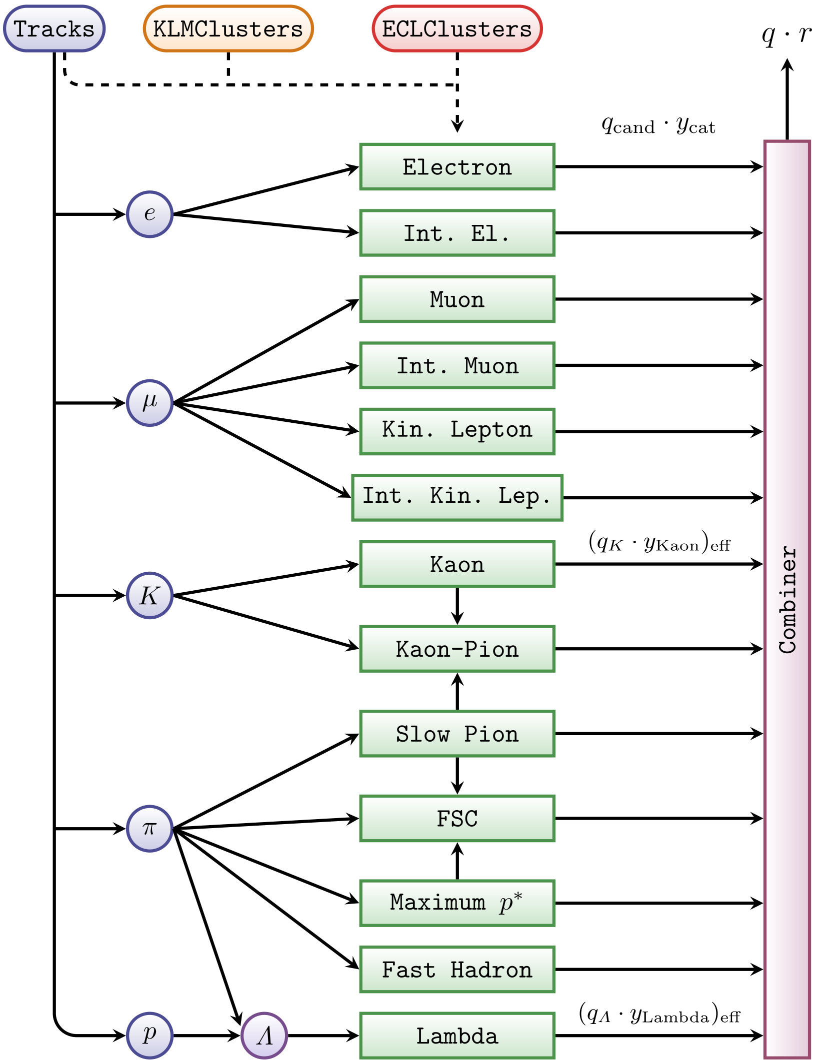


Figure 4.2: Schematic representation of the `flavour tagger` algorithm [85].

can see the suppression of the peaks at 0 and around $qr = \pm 0.65$, as well as a greater concentration of the values around $qr = \pm 1$.

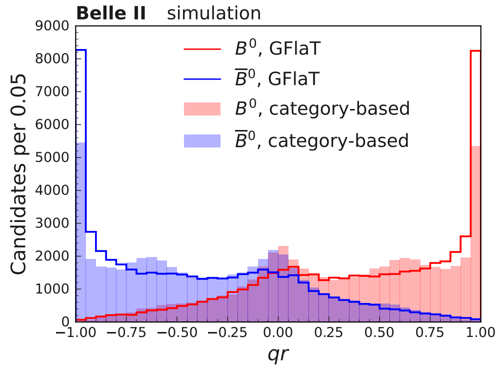


Figure 4.3: Comparison of the qr distribution for the category-based and GFlaT algorithms [86].

4.3 Calibration of the GFlaT algorithm

We need to calibrate the GFlaT algorithm before using it to perform physics measurements. This is a necessary step to ensure the reliability of the algorithm and to give the uncertainty associated with its usage.

There are three parameters of interest for the calibration: the efficiency ε defined in Equation (4.2), the wrong tag fraction w defined in Equation (4.3), and the difference in performances (efficiency and wrong tag fraction) for B^0 and $\bar{B}^0$ mesons.

The first step of the calibration is to select decays of the B_{sig}, for which we can check the validity of the algorithm predictions. Those decays need to be self-tagged and abundant enough to facilitate calibration. The decays selected for the calibration are the following:

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

as well as their equivalent, but with a kaon instead of the pion coming from the B meson directly. As described in Figure 4.1, those are indeed self-tagged decays with the pion or the kaon acting as probe. The reason for the addition of the $B^0 \to D^{(*)-}K^+$ modes is that they are similar to the $B^0 \to D^{(*)-}\pi^+$ decays.

The main variable used for signal selection is the energy difference between the energy of the reconstructed B_{sig} and half the beam energy in the centre of mass frame, defined as:

$$\Delta E = E_B^{\text{reco}} - \frac{\sqrt{s}}{2} \tag{4.7}$$

The choice of this variable is motivated by the fact that for correctly reconstructed B mesons, the energy should be equal to the beam energy in the centre-of-mass frame, and thus ΔE should be close to zero. For background events, ΔE will take a wider range of values, forming a sort of continuum. Finally, for $B^0 \to D^{(*)-}K^+$ decays, we also expect a peak, but at a slightly lower value ($\Delta E < 0$) due to the higher mass of the kaon compared to the pion.

We select events with $\Delta E \in [-0.1, 0.25]$ GeV and $R_2 < 0.4$, which is a common selection for this variable. Additionally, we apply a selection on the beam-constrained mass,`Mbc`, defined as:

$$M_{\mathrm{bc}} = \sqrt{\left(\frac{\sqrt{s}}{2}\right)^2 - |\vec{p}_B^{\,\mathrm{reco}}|^2} \tag{4.8}$$

i.e. the mass of the reconstructed B meson assuming that its energy is equal to the beam energy. We require a `Mbc` value between 5.27 and 5.30 GeV.

We then model the ΔE distribution of each component with the following functions:

- $B^0 \to D^{(*)-}\pi^+$ and $B^0 \to D^{(*)-}K^+$: double-sided crystall-ball, parametrised by the mean μ, the width σ, the tail parameters α_L and α_R and the tail exponent parameters n_L and n_R;
- $B\bar{B}$ background: Chebychev polynomial of degree 2;
- $q\bar{q}$ background: exponential law.

The second step is to perform a fit in MC to the ΔE distribution for each sample ($B^0 \to D^{(*)-}\pi^+$ and $B^0 \to D^{(*)-}K^+$) and for the background components, from which we extract the parameters and fix the shape for the following step.

After obtaining the shapes for each contribution from MC, we perform an unbinned maximum likelihood fit on this same variable. We extract the yield of the signal events in the sample, which will be used to compute the wrong tag fraction and efficiency. This fit is done for various categories. Each category is defined by:

- The flavour of the B_{tag}, with a superscript $^+$ if the B_{tag} is tagged as a $\bar{B}^0$ and a superscript $^-$ if it is tagged as a B^0.
- The relative flavour of the B mesons at the time of their relative decays, which can be either Same Flavour (SF) or Opposite Flavour (OF).
- The bin of $|qr|$, defined by a number from 0 to 6, with the binning being: $([0, 0.1], [0.1, 0.25], [0.25, 0.45], [0.45, 0.6], [0.6, 0.725], [0.725, 0.875], [0.875, 1])$. This binning is standard for CP-violation analyses at Belle II.

Thus, we define 28 categories in total, for which we can extract the signal yield. An example of the ΔE distribution for one of the categories is shown in Figure 4.4. We can see the peak around $\Delta E = 0$ corresponding to the $B^0 \to D^{(*)-}\pi^+$ decays, and the smaller peak around $\Delta E = -0.05$ GeV corresponding to the $B^0 \to D^{(*)-}K^+$

decays. The remaining distribution corresponds to the background events.

For a given bin of $|qr|$ i, we can then compute the wrong tag fraction for each flavour of B_{tag} $w_i^{\pm}$. A formula can be derived by seeing that the number of B^0 coming from an SF event is related to the total number of B^0 meson via

$$N_{SF}^{-} = N_{tot}^{-}\left(w^{-} + \chi_d\right) \tag{4.9}$$

where χ_d is the time-integrated mixing probability of the B^0 meson, equal to 0.1860 ± 0.0011 [87]. In other words, B^0 mesons from SF events are either due to mixing or to an error on the tagging. We can then develop the expression, as $N_{tot}^{-} = N_{OF}^{-} + N_{SF}^{-}$:

$$\begin{aligned} N_{SF}^{-} &= \left(N_{OF}^{-} + N_{SF}^{-}\right)\left(w^{-} + \chi_d\right) \\ \Leftrightarrow \left(N_{OF}^{-} + N_{SF}^{-}\right) w^{-} &= \left(1 - \chi_d\right) N_{SF}^{-} - \chi_d N_{OF}^{-} \\ \Leftrightarrow w^{-} &= \frac{\left(1 - \chi_d\right) N_{SF}^{-} - \chi_d N_{OF}^{-}}{\left(N_{OF}^{-} + N_{SF}^{-}\right)} \\ \Leftrightarrow w^{-} &= \frac{1 - \frac{\chi_d N_{OF}^{-}}{(1-\chi_d) N_{SF}^{-}}}{\frac{\left(N_{OF}^{-} + N_{SF}^{-}\right)}{(1-\chi_d) N_{SF}^{-}}} \end{aligned}$$

We can then generalise to any of the categories defined before. Finally, by defining

$$f_i^{\pm} = \frac{N_{OF,i}^{\pm}}{N_{SF,i}^{\pm}} \tag{4.10}$$

$$R = \frac{\chi_d}{1 - \chi_d} \tag{4.11}$$

we get

$$w_i^{\pm} = \frac{1 - Rf}{(1 - R)\left(f_i^{\pm} + 1\right)} \tag{4.12}$$

From the separation between the w_i^{+} and w_i^{-}, we then get the average wrong tag fraction w_i and the wrong tag fraction asymmetry Δw_i as:

$$w_i = \frac{w_i^{+} + w_i^{-}}{2} \tag{4.13}$$

$$\Delta w_i = w_i^{+} - w_i^{-} \tag{4.14}$$

The same separation can be done for the efficiency $\varepsilon_i^{\pm}$:

$$\varepsilon_i^{\pm} = \frac{N_{OF,i}^{\pm} + N_{SF,i}^{\pm}}{N^{\text{tagged}}} \tag{4.15}$$

with the same average efficiency ε_i and efficiency asymmetry $\Delta\varepsilon_i$.

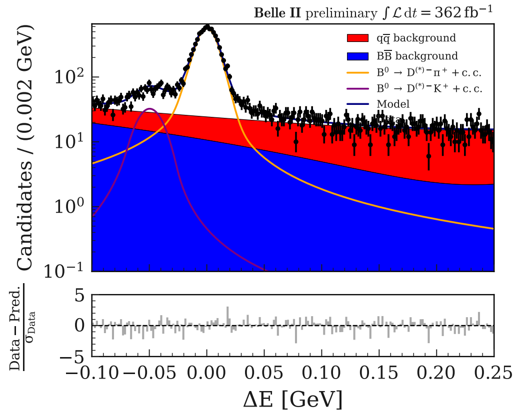


Figure 4.4: Total ΔE distribution for the 1st bin of $|qr|$ in the OF^{-} case on MC.

Finally, we compute the *effective* tagging efficiency ε:

$$\varepsilon = \sum_{i=1}^{6} \varepsilon_i \left(1 - 2w_i\right)^2 \tag{4.16}$$

which is the figure of merit for the tagging algorithm. A summary of the workflow of the calibration procedure is shown in Figure 4.5.

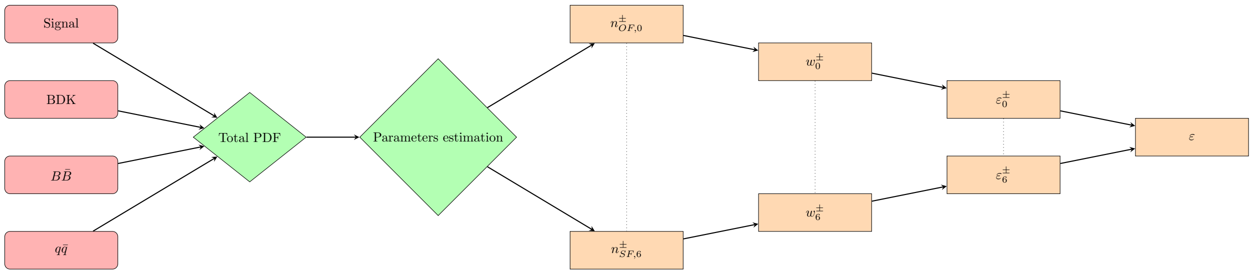


Figure 4.5: Workflow of the calibration procedure.

4.4 *Upgrade of the* `GFlaT` *calibration*

As explained in Section 4.3, to extract the signal yields from the ΔE distribution, we have to perform a fit for each of the 28 categories (2 flavour categories $\times$ 2 relative flavours categories $\times$ 7 bins of $|qr|$). Making those fits sequentially and independently is time-consuming. One could just parallelise the fits, but we would then maybe miss some correlations between the categories and thus underestimate the uncertainties or not use the full information available. To solve this issue, we can perform a simultaneous fit of all the categories. This is what we switched to.

To verify the impact of this change, we can check the correlations between the signal yields of the different categories. They were evaluated at about 10^{-10}, which is negligible. From this, we decided to keep the simultaneous fit as the default method to extract the signal yields, as it is much faster. Moreover, since there are no correlations, there is no need for uncertainty propagation when computing the wrong tag fractions and the efficiencies.

An important background component is the $q\bar{q}$ background. The first way to reduce it is by using a variable such as R^2 or `cosTBTO` defined in Section 3.3. A selection on $R^2 < 0.4$ is already applied, but by selecting events with `cosTBTO < 0.8`, the $q\bar{q}$ count is further reduced by 86% on the MC sample while keeping 84% of the signal events. The impact of this selection on the ΔE distribution is shown in Figure 4.6. The second way to reduce it would be to use a BDT trained to separate the $q\bar{q}$ from the other events. This technique was effective, as it reduced the $q\bar{q}$ count by 98%, however it was very ineffective on data, probably due to the features used for the training not being well modelled in MC. Thus, we decided to keep the additional cut `cosTBTO` as the default method to reduce the $q\bar{q}$ background.

Lastly, the whole procedure, as described in Section 4.3 and in Figure 4.5, is managed using the `b2luigi` package[88], which allows us to easily parallelise the different steps.

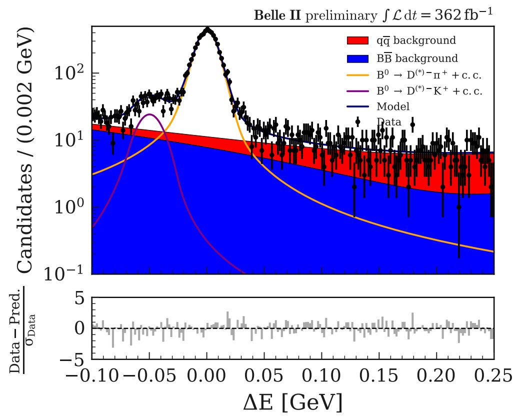


Figure 4.6: Total ΔE distribution for the 1st bin of $|qr|$ in the OF$^-$ case on MC with the selection based on `cosTBTO`.

4.5 *Results and prospects for flavour tagging at Belle II*

The results of the calibration are presented in Table 4.2. The wrong tag fractions and efficiencies obtained in data and MC are similar to those obtained in the previous calibration[86]. With this work, we thus developed and automatised a calibration procedure of the `GFlaT` for study of *CP*-violations. To bring some further prospects on flavour tagging at Belle II, we are currently working on a flavour tagging algorithm based on a *transformer* architecture [89], which already shows promising results with an increase in efficiency, for example.

Quantity	MC [%]	Data [%]
ε	38.2 ± 0.5	37.3 ± 1.0
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

Table 4.2: `GFlaT` calibration parameters obtained in MC and data.

5

The what: analysis strategy and results

Contents

After developing the motivation for the study of the $B^+ \to K^+\nu\overline{\nu}$ decay in Chapter 1, giving the tools for this study in Chapter 2 and Chapter 3, we can now develop our strategy for the measurement of the $B^+ \to K^+\nu\overline{\nu}$ decay's branching ratio. We will perform a counting experiment, using the techniques described in Section 3.6 and Section 3.7. Before doing so, we have to define our dataset (Section 5.1) and the reconstruction strategy (Section 5.2). As we will do a blind analysis with a template fit, we need control samples (Section 5.3) to correct for discrepancies between our simulation and the real data (Section 5.4). Using a BDT, we then perform the final selection on the BDT output to get the signal region in which we will extract the signal (Section 5.5). To validate our signal region, we need to cross-check via the control samples and the studies of the backgrounds (Section 5.6). We explicitly describe our statistical model with its systematic uncertainties (Section 5.7) and perform the Asimov fit (Section 5.8), final step for the testing of the statistical model and before the unblinding. Finally, we discuss the obtained results and compare them with the previous analyses to give some context and make explicit the situation of the analysis relative to the state of the art (Section 5.9).

5.1 Dataset and Simulated Samples

We use data collected during Run 1 and Run 2 of Belle II operation, between 2019 and 2024, corresponding to an integrated luminosity of 487.72 fb^{-1} at the $\Upsilon(4S)$ resonance and 59.43 fb^{-1} 60 MeV below the resonance (off-resonance sample).

For the simulation, we have 10M events for the $B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$, while there are 20M events for the $B^+ \to K^{*+}\nu\bar{\nu}$ mode. For the generic $B\bar{B}$ and $q\bar{q}$ samples, we generated 4 times the dataset luminosity. As this analysis focuses on $B^+ \to K^+\nu\bar{\nu}$, the other signal modes will be used for crossfeed studies and will be considered as $B\bar{B}$ backgrounds when not specified. Additionally, we use specific samples for background studies, such as $B \to X_s\nu\bar{\nu}$, $B \to \tau\nu$ and $B \to K^+K^0_LK^0_L$.

5.2 *Reconstruction and preselection*

To save computing time, we first reject events containing more than 12 clean tracks (defined as tracks with dr $<$ 2 cm and $|\mathrm{d}z| < 4$ cm), as we do not expect that many tracks. This cut has a 100% efficiency, and is mainly there for technical reasons to avoid too much combinatorics for the FEI algorithm. Events passing this pre-filter are then subject to reconstruction and selection criteria described in the following sections.

5.2.1 *B_{tag} reconstruction*

The B_{tag} candidates are reconstructed using the semileptonic FEI with the following 8 modes: $B \to D^{(*)}\ell\nu\,(\pi)$, with $\ell \in \{e, \mu\}$, as described in Section 3.2. We select B_{tag} with the properties:

- $-4 < \cos(\theta_{BY}) < 3$
- $\log_{10}(\texttt{SignalProbability}) > -2.4$
- $p^*_\ell > 1.0$ GeV/c
- $\texttt{cosTBTO} < 0.9$

Both for $\cos(\theta_{BY})$ and $\log_{10}(\texttt{SignalProbability})$ we tested different selections. For the latter, two alternative working points were tested: $\log_{10}(\texttt{SignalProbability}) > -2, -1$. For $\cos(\theta_{BY})$, we considered tighter ranges inspired from other semileptonic tagged analyses, namely: $[-1.75, 1.1]$ and $[-1.9, 1.2]$. No improvements in signal sensitivity with respect to the nominal choice were observed.

The population of the different B_{tag} decay modes is shown in Figure 5.1. It can be seen that the modes with a π show lower purity with respect to the others. Since their contribution does not improve the signal sensitivity, we retain only B_{tag} candidates reconstructed in $D^{(*)}\ell\nu$ final states.

After reconstruction of the B_{tag}, we have multiple possible candidates, in average 3.15, each assigned with a probability of being the correct B_{tag}. We then do a *best candidate selection* by ranking the FEI candidates based on their `SignalProbability` and keeping the one with the highest value.

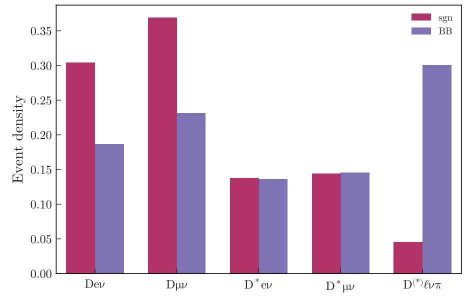


Figure 5.1: Signal and $B\bar{B}$ background composition of the B_{tag} candidates reconstructed in the semileptonic FEI. The modes with a π show lower purity with respect to the others.

5.2.2 *Signal candidate and Rest-of-the-event*

For the tracks not used in the B_{tag} nor the K^+ reconstruction, we define two selections:

- "GoodTracks" with d$r < 0.5$ cm, $|\text{d}z| < 2$ cm, $p_t > 0.2$ GeV/c, $E < 5.5$ GeV in the CDC acceptance,
- "AcceptableTracks" with looser requirements (d$r < 2$ cm, $|\text{d}z| < 4$ cm, $p_t > 0.1$ GeV/c).

For $B^+ \to K^+\nu\bar{\nu}$, the signal kaon must satisfy "`GoodTracks`" requirements and have `KaonIDNN` > 0.1. We then apply a selection on the invariant mass of the reconstructed B with $0 < M < 6$ GeV/c^2.

Once the B_{sig} and B_{tag} candidates are reconstructed, we can build the ROE, *i.e.* the tracks and clusters that are attached to neither the B_{sig} or the B_{tag}. For the ROE tracks we apply the AcceptableTracks selection, while for the cluster we have different energy cuts depending on the ECL region: $E > 0.1$ GeV for the forward region, $E > 0.06$ GeV for the barrel, and $E > 0.15$ GeV for the backward region. Additionally, for the ROE clusters, we require a minimum cluster-to-track distance `minC2TDist` > 20 cm as well as the polar angle of the cluster `clusterTheta` $\in [0.2967, 2.6180]$, which is the CDC acceptance. While the energy thresholds help reduce machine background photons, the `minC2TDist` cut suppress the contamination from clusters deposited by charged tracks, for which track-cluster matching has failed.

After both the B_{tag} and B_{sig} candidates are reconstructed, we can reconstruct the $\Upsilon(4S)$ candidate by combining the two B mesons. The final $\Upsilon(4S)$ selection enforces event cleanliness by requiring no additional K^0_S, Λ or π^0, as well as the remaining tracks to not contain events with AcceptableTracks, meaning that remaining tracks must have $p_t < 0.1$ GeV/c, d$r >$ 2cm and $|\text{d}z| > 4$ cm.

We finally ask that $\theta_{miss} \in [0.3, 2.6]$, θ_{miss} being the polar angle associated to the missing momentum vector, so that the missing

Category	Cuts
(Good)Tracks	*nCleanedTracks*(dr < 2 cm and $\|$d$z\|$ < 4 cm) < 12
	thetaInCDCAcceptance
	$p_t > 0.2$ GeV, $E < 5.5$ GeV/ c^2
	dr < 0.5 cm, $\|$d$z\|$ < 2 cm
Clusters	$(E_{\text{forward}}, E_{\text{barrel}}, E_{\text{backward}}) > (0.1, 0.06, 0.15)$ GeV
	`minC2TDist` > 20 cm
	`clusterTheta` $\in [0.2967, 2.6180]$
K^+	GoodTracks
	`KaonIDNN` > 0.9
ROE	No extra K^0_S, Λ and π^0
	No AcceptableTracks ($p_t < 0.1$ GeV/c, dr > 2 cm and $\|$d$z\|$ > 4 cm)

Table 5.1: Table summarising the selection criteria at reconstruction level.

energy system is within the CDC acceptance. We apply a first continuum suppression cut by selecting events with $R_2 < 0.4$ and the cut on the K^+ identification is further tightened to `KaonIDNN` > 0.9. Table 5.1 summarises the reconstruction and selection.

5.3 *Control samples*

To validate the model, we need control samples. For $q\bar{q}$ backgrounds, we use an off-resonance sample with both data and MC. For $B\bar{B}$ backgrounds, we have two control samples: one where we accept extra `AcceptableTracks` but no `GoodTracks` in the ROE, *i.e.* we require at least one track in the ROE with $p_t \in [0.1, 0.2]$ GeV/c or d$r \in [0.5, 2]$ cm or $|$d$z| \in [2, 4]$ cm; and another one for which we require the two B mesons to have the same charge. They are respectively called the "*Extra Tracks*" (ET) and "*Wrong Charge*" (WC) control samples or sidebands. For the signal efficiency validation, we use the *embedding* method, as already used by the previous Belle II $B \to K\nu\bar{\nu}$ analyses [5].

The idea of embedding is to first choose a channel of which we have good knowledge of the kinematics and with a kaon (*e.g.* $B^+ \to K^+ J/\psi$). Then, we reconstruct a sample of data and a simulation of this channel, and we remove the signal-side from the events, leaving only the ROE. We then embed our simulated signal-side into the ROE, and we can then reconstruct the event as if it were a signal event.

5.4 Data-simulation discrepancy corrections

As the background suppression and signal extraction will rely on the MC, it is crucial to correct the simulation to match the data as much as possible. The following corrections are all applied before the training of the BDT for background suppression. The different corrections are summarised in Table 5.4.

5.4.1 PID correction

As the detector effects are not perfectly simulated, we need a correction for the PID. We use the `Systematic Corrections Framework` to produce PID tables with respect to the transverse momentum p_T and the polar angle $\cos(\theta)$. The correction is calibrated from the $D^{*+} \to D^0 (\to K^-\pi^+) \pi^+$ control mode. Tables are produced for the efficiency on kaons and the pion fake rate, on a p_T range of $[0.2, 5.5]$ GeV and a $\cos(\theta)$ range of $[-0.87, 0.9563]$. The tables are shown in Appendix A.1.

5.4.2 FEI correction

For the FEI correction, we use calibration factors extracted by the collaboration using the $B^+ \to D^{*0}\ell^+\nu$ control mode. There is one calibration factor per mode used for the B_{tag}. The factors we used are presented in Table 5.2.

B_{tag} **mode**	**Calibration factor**
$B^+ \to D^0 e^+ \nu_e$	1.14(11)
$B^+ \to D^0 \mu^+ \nu_\mu$	1.10(11)
$B^+ \to D^{*0} e^+ \nu_e$	0.87(10)
$B^+ \to D^{*0} \mu^+ \nu_\mu$	0.850(98)

Table 5.2: Table summarising the calibration factors used for each B_{tag} mode.

5.4.3 Branching fraction correction

The branching fractions used to generate the MC samples are not always up-to-date. We thus need to correct them to the most up-to-date values from the Particle Data Group (PDG) [19]. The weights are computed as the ratio of the 2024 PDG value to the value used in the simulation. We only apply them on B_{sig} decays.

There are exceptions: if the decay is an inclusive $B \to X_s \nu\bar{\nu}$, with the X_s being an hadronic system containing an s-quark and that is not a K or a K^*, decay, or if the final state particles are $KK_L^0 K_L^0$ or $Kn\bar{n}$, then we do not compute a weight for the branching fraction correction since we correct them separately, as explained in the following section. Likewise, if the decay does not appear explicitly in the simulation decay table, we consider it as a decay generated via the `PYTHIA` package, and we apply a weight of 1. For more information on the simulation, please refer to Section 2.5.

Moreover, some decays include resonant and non-resonant contributions, *e.g.* $B^0 \to K^+K^-K^0$. In nature, because of the interference between the resonances and the non-resonant contributions, it is possible that the branching fraction of the non-resonant contribution is larger than the total branching fraction of the inclusive decay. However, in the simulation this is not taken into account, so the resonances and non-resonant contributions should be additive. We decided not to correct the branching fractions for the decays $B^+ \to K^+\pi^+\pi^-$ and $B^0 \to K^+K^-K^0$ to the PDG value. Instead, we keep the value present in the simulation and increase the associated uncertainty on the lower limit to cover this discrepancy.

In total, our sample contains 3984 unique decays. Decays generated by `PYTHIA` represent about 83% of the decays.

5.4.4 *Global branching fraction correction*

As we only correct a part of all the branching fractions from the Belle II simulation, the sum of all the branching fractions is not unitary anymore. To rectify this, we apply a global correction for the events we do not correct with the previous method. The weight is computed from the sum of corrected branching fractions to their PDG values and the sum of the branching fractions we do not correct to their value in the simulation:

$$w_{\text{global BF}} = \frac{1 - \sum_{i \in \text{corrected}} \text{BF}_i^{\text{PDG}}}{\sum_{i \in \text{not corrected}} \text{BF}_i^{\text{simulation}}} \tag{5.1}$$

The weight is calculated separately for $B^0\bar{B}^0$ and B^+B^- events, yielding a correction of 1.00 for the $B^0\bar{B}^0$ events and about 1.07 for the B^+B^- events.

5.4.5 *Specific background correction*

Certain decays mimic very well the $B^+ \to K^+\nu\bar{\nu}$ decay. This is the case for the $B^+ \to K^+n\bar{n}$ and $B^+ \to K^+K_L^0K_L^0$ decays, as both have a charged kaon and two neutral particles which are difficult to detect and identify properly.

For $B^+ \to K^+n\bar{n}$ decays, we would like to correct the simulation branching fraction just like previously. In the simulation, the indicated value is 5.7×10^{-6}. As there is no measurement of this branching fraction, we have to use isospin symmetry to compute it from the measurement of $B^0 \to K^0p\bar{p}$, with a branching fraction of $2.79(32) \times 10^{-6}$. However, it was observed that at the $p\bar{p}$ production threshold, the production probability is enhanced [90].

To take this into account, we further fit the mass of the $p\bar{p}$ system from $B^0 \to K^0 p\bar{p}$ analysis with a threshold plus an exponential law. The fitted function is of the form $A\,(x - x_{th})^n\, e^{-\lambda x}$, so we have 3 parameters to extract (A, n, λ). The fit and the values extracted can be seen in Figure 5.2. The q^2 distribution of $B^+ \to K^+ n\bar{n}$ events is then reweighted to match the fitted distribution, considering the two neutrons system as the two neutrinos system in our decay of interest.

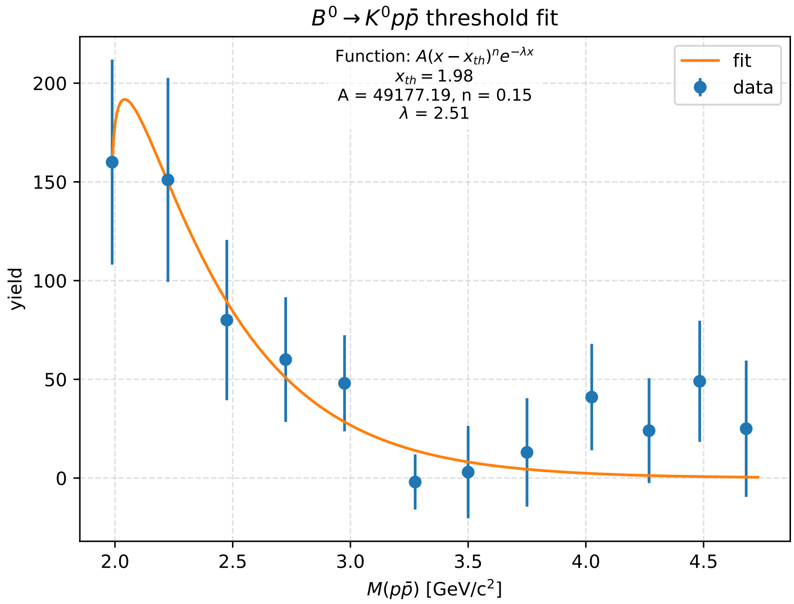


Figure 5.2: Fit of the $p\bar{p}$ mass distribution from $B^0 \to K^0 p\bar{p}$ decays, using a threshold function and an exponential law. Data are from [90].

For $B^+ \to K^+ K_L^0 K_L^0$ decays, we directly correct the q^2 distribution, originally generated via a phase-space model, to match the one from the BaBar measurement of $B^+ \to K^+ K_S^0 K_S^0$ [91].

For the $B \to X_s \nu\bar{\nu}$ decays, we intend to correct the mass distribution of the X_s and the q^2 distribution following the $B \to X_s \ell\ell$ analysis done by Belle II that is yet to be published.

5.4.6 *Form factor correction*

The $B \to K^+ \nu\bar{\nu}$ sample was generated using an outdated model in which the decay amplitude is described by means of the form factors provided in [15]. The sample is then reweighted according to the SM prediction from [20], using the formula and values presented in Section 1.3. The corrections here are close to unity, and so have little to no impact.

5.4.7 Background normalisation

We apply a correction to the normalisation of the different backgrounds, namely continuum and $B\bar{B}$.

The continuum normalisation correction is computed from the off-resonance data-MC ratio, yielding a correction of about 0.96.

For the $B\bar{B}$ normalisation factor, we only compute it for the "badly-tagged" ET events, and apply it only on badly-tagged events. We define "well-tagged" events as an event with a B meson and *either*:

- 0 wrong daughter
- 1 wrong daughter **but** it is a γ

The weight is then computed as:

$$w_{B\bar{B}} = \frac{N_{\text{data}} - \left(N_{q\bar{q}} + N_{\text{well-tagged}}\right)}{N_{\text{badly-tagged}}} \tag{5.2}$$

The correction is about 2.28. The well-tagged $B\bar{B}$ events normalisation is checked on the embedding sample and is compatible with 1, so we do not need further correction.

5.4.8 Backgrounds shape correction

To take into account the differences in shape for data and simulation of backgrounds, for variables which will be used in the background suppression, we train two BDTs, one on off-resonance (BDT_{offres}) and one ET (BDT_{ET}) sideband, respectively, to separate between data events and MC events. The goal of this training is to use the BDT output to compute the data/MC correction. The weights are computed as:

$$w_{BDT_{sb}} = \frac{BDT_{sb}}{1 - BDT_{sb}} \tag{5.3}$$

with $sb \in \{\text{offres}, \text{ET}\}$. The heuristic is the following: if the input variables from the two datasets are close enough in shape and values, the BDT should not be able to distinguish the two and hence should give an output of 0.5, so the weight would be 1; any discrepancy in the input variables would have an impact on the BDT output, which would give a weight different than unity. As one can see from Figure 5.3, the BDT output on the data is indeed close to 0.5, which means that most of the weights will be 1, or close to.

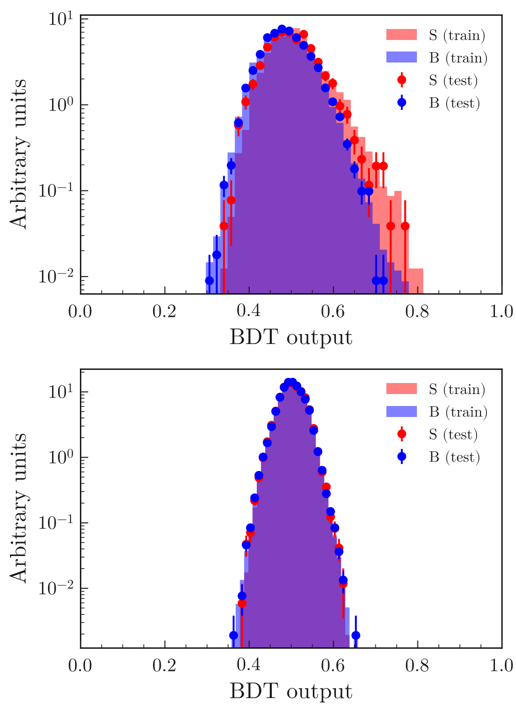


Figure 5.3: Training output of the BDTs trained on off-resonance (top) and ET (bottom) samples.

The hyperparameters of the BDT_{sb} and their values are summarised in Table 5.3. A detailed description of the hyperparameters can be found in Section 5.5.1. The features used for the BDT_{sb}, ranked by their importance, are summarised in Figure 5.4 and Figure 5.5. The importance is quantified by their *Shapley value* computed using the `SHAP` package, as detailed in Section 3.1.5. A description on how to read them is provided in Section 5.5.

Hyperparameter	Value
`max_depth`	2
`learning_rate`	0.05
`subsample`	0.5
`n_estimators`	200

Table 5.3: Table summarising the hyperparameters of the BDT_{sb} and their values.

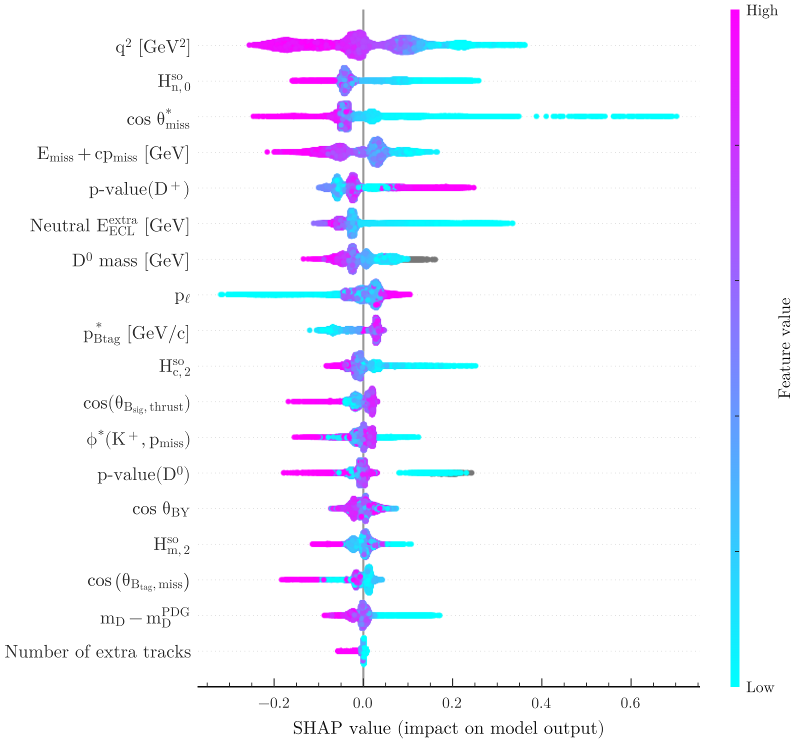


Figure 5.4: Shapley values for the BDT_{offres} features. A description of the variables can be found in Appendix A.2.

The BDTs are trained one after the other, with first the BDT_{offres}, from which we apply the weights on the continuum backgrounds before training the second BDT.

5.4.9 *Summary and results of the corrections*

We can check the impact of the corrections on the data/MC agreement. For the variables of interest for the background suppression,

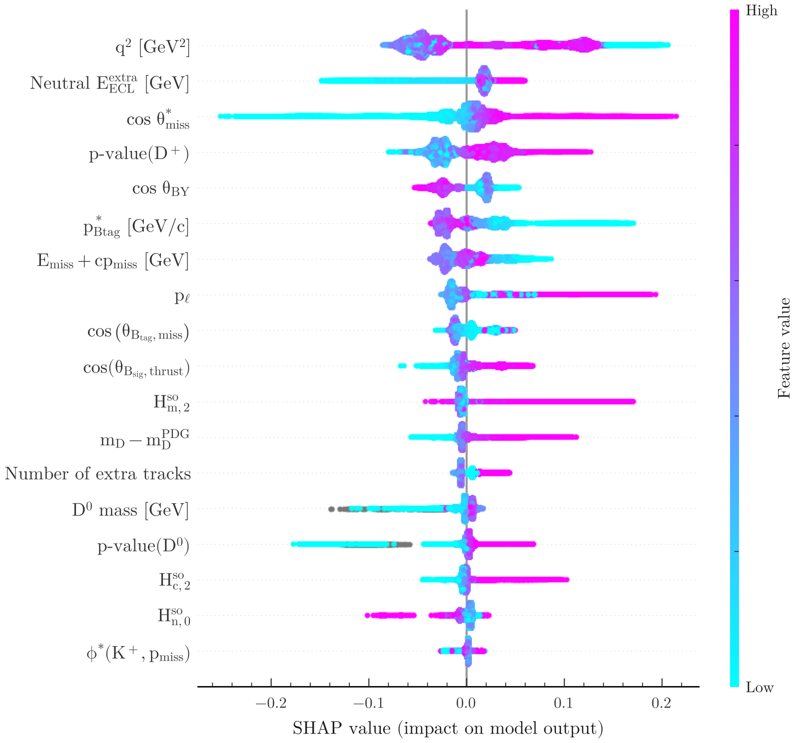


Figure 5.5: Shapley values for the BDT_{ET} features. A description of the variables can be found in Appendix A.2.

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 ($B \to K K^0_L K^0_L$ and $B \to K n \bar{n}$)	$B\bar{B}$
$q\bar{q}$ 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 $q\bar{q}$

Table 5.4: Table of the corrections and the samples they are applied to

we show the distribution for data and MC in Figure 5.8 and Appendix A.3.

To verify the pertinence of the control samples, we can also check the distribution of the backgrounds in MC for both signal sample and control samples. We see in Figure 5.6 that the control samples do replicate the peak in the background distribution for q^2 of the signal sample after the corrections. Improvement is still possible, *e.g.* via a better training of the BDTs from Section 5.4.8.

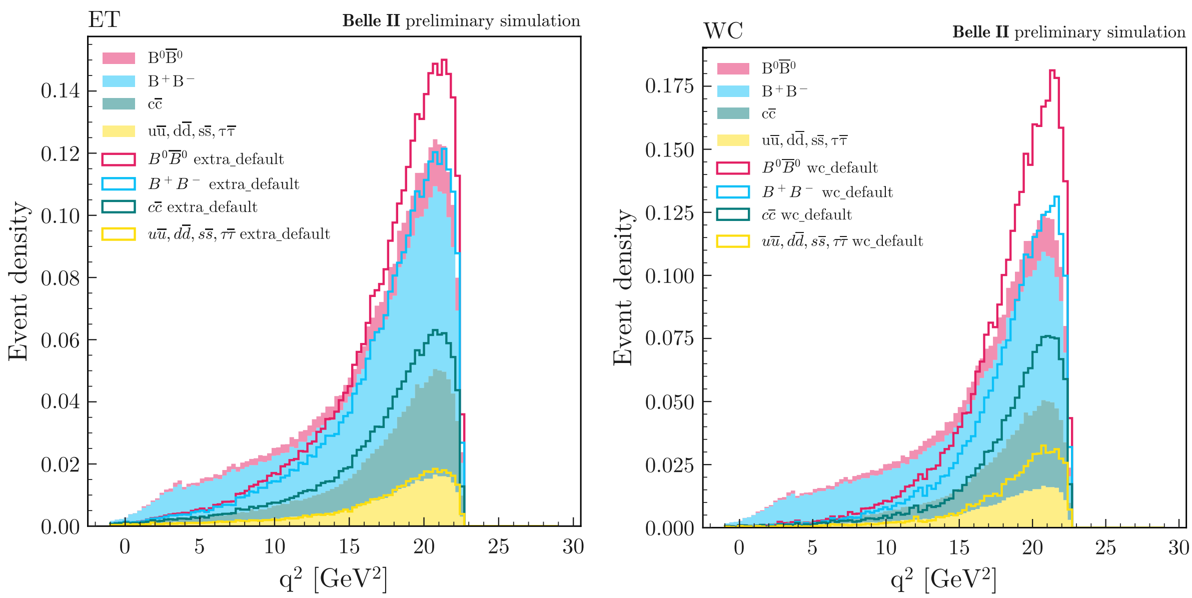


Figure 5.6: Plots of the background MC comparison between the ET and WC sidebands (contour only) and the signal sample (filled) for q^2.

5.5 *Background suppression and signal selection*

5.5.1 *BDT training*

We train a multivariate binary classifier to separate signal from background events. More specifically, we use a BDT from `XGBoost`. The training is performed on a mixture of signal and background MC samples, where the signal sample is weighted to be 300 times lower than the background sample using the `scale_pos_weight` hyperparameter. This 300 is the approximate number of expected background events for 1 signal event.

The other hyperparameters of interest for this analysis are the following:

- The maximum depth of the trees (`max_depth`)
- The number of trees to be built (`n_estimators`)

- The learning rate, which controls how much each new tree contributes to the final prediction (`learning_rate`)
- The subsample ratio of the training dataset, which controls the fraction of the training data that is used to grow each tree, to limit overfitting (`subsample`)
- The minimum loss reduction required to make a further partition on a leaf node of the tree (`gamma`)
- Regularisation parameter to control the complexity of the trees and prevent overfitting (`reg_lambda`)

As each of these hyperparameters controls both the training time and the performance of the model, such as overfitting, we need to properly tune them. This is done by using the `Optuna` package [69], and optimising the hyperparameters to maximise the F_1 score, as defined in Section 3.1. The optimised hyperparameters and their values are summarised in Table 5.5.

Hyperparameter	$B^+ \to K^+ \nu \bar{\nu}$
`max_depth`	7
`learning_rate`	0.06
`subsample`	0.9
`n_estimators`	290
`gamma`	0.5
`reg_lambda`	0.005

Table 5.5: Table summarising the hyperparameters of the BDT and their value.

For the training, we use 18 variables described in Appendix A.2. Their importance are reported in Figure 5.7. The x-axis indicates the actual Shapley value, as explained in Section 3.1.5. The colour spectrum shows the feature value. Each point represent a data point given for the training. This shows how strongly the feature values are correlated to their Shapley value, allowing us to rank them. The y-axis corresponds to the actual ranking of the features using the Shapley values. The most important variable is the `Neutral` $E_{\mathtt{ECL}}^{\mathtt{extra}}$, *i.e.* the energy in the ECL associated to photons from the ROE, followed by q^2. The left side of Figure 5.8 shows the discriminating potential of the variable, while the right side shows the agreement between data and MC in the ET sideband with the corresponding data-MC ratio comparison and the pulls.

Both variables exhibit strong discriminating power with good data-MC agreement. The other variables are presented in Appendix A.3.

Overfitting is controlled in two ways:

- First, we split our training sample into two, depending on the *parity of the event number*, and we train a BDT for each of the two samples;
- Second, we check the agreement between the training and the testing sample for each BDT.

This strategy allows us to identify any potential overfitting. The BDTs are then applied to the samples with opposite event number

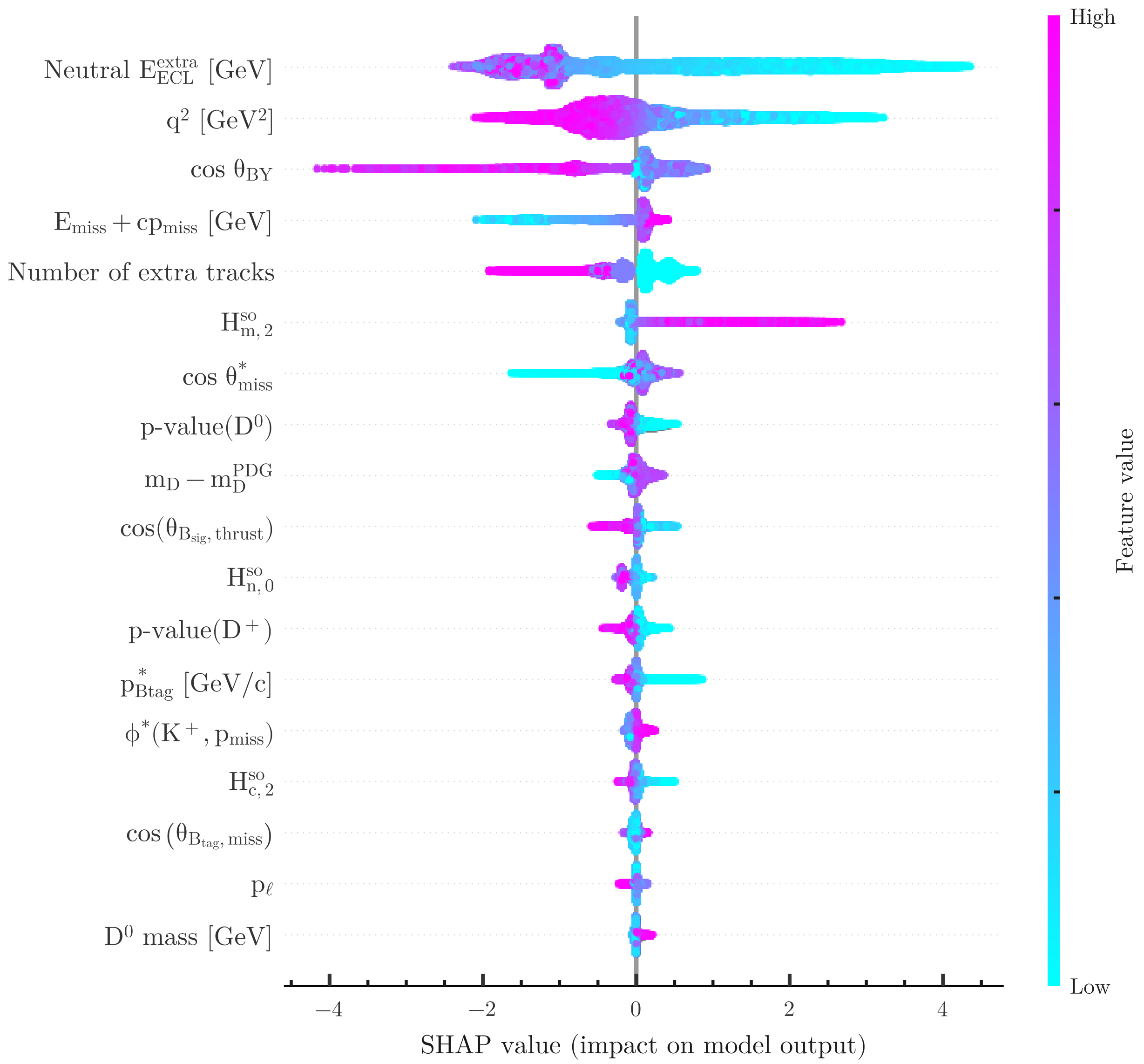


Figure 5.7: Importance of the variables used for the BDT training as measured by the Shapley values for the BDT, shown in a beeswarm plot.

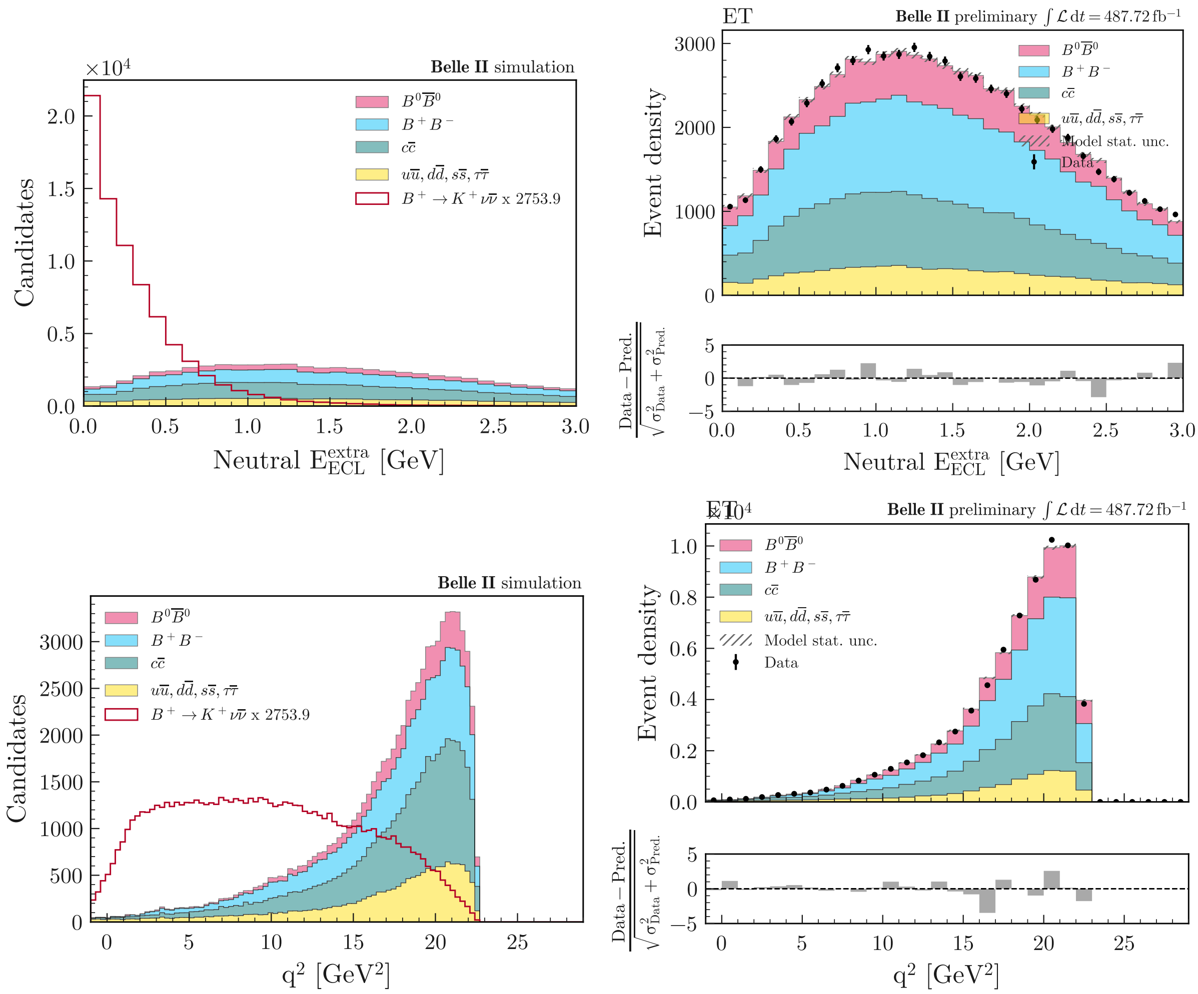


Figure 5.8: On the left, distribution of Neutral $E_{\mathrm{ECL}}^{\mathrm{extra}}$ and q^2 in MC for signal and backgrounds. On the right, distribution of Neutral $E_{\mathrm{ECL}}^{\mathrm{extra}}$ and q^2 in data and MC for the ET sideband.

parity (*e.g.* the BDT trained on the training sample with even event numbers will be used for classification of the events with odd event numbers). The overfitting of each BDT individually is checked by comparing the BDT output distribution for the training and testing sample, as shown in Figure 5.9.

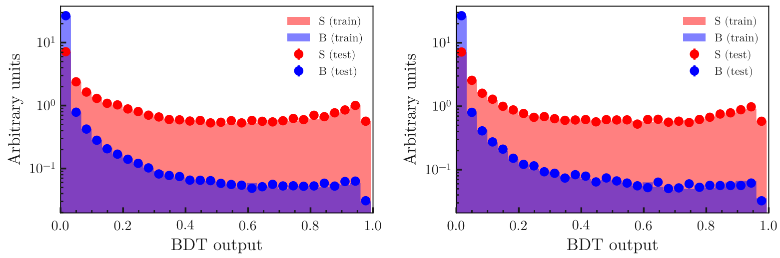


Figure 5.9: Overfitting check for the BDTs training from the training and testing samples, with the even event training sample (left) and odd event training sample (right).

The agreement between the two BDTs can be seen in Figure 5.10 and is considered as good.

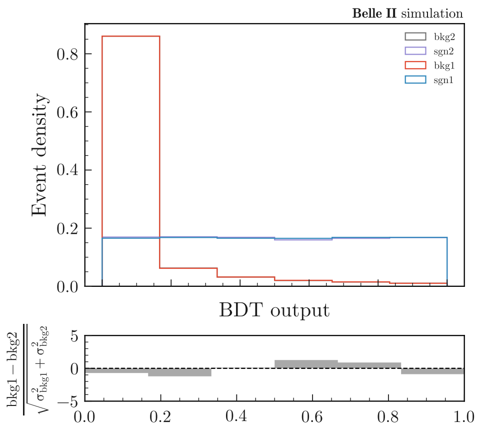


Figure 5.10: Overfitting check for the BDT training from the two-sample training (1 is for the even event sample and 2 is for the odd event sample).

Finally, the BDT output is modified such that signal events are flattened. This is done by using a *quantile transformer*, namely the `QuantileTransformer` from `scikit-learn` [92]. This transformation makes the signal region selection easier, as signal events are uniformly distributed between 0 and 1. The results and the impact of the transformation on the background events can be observed in Figure 5.11. The signal events show no sign of overfitting, even after the transformation.

5.5.2 *Signal region selection*

After the training, the transformation and the overfitting checks, we apply the BDT to all the samples. The Signal Region (SR) is then defined by selecting events with a BDT output above 0.4. The choice for this value was not dictated by any figure of merit, but a trade-off between having enough background events in the signal

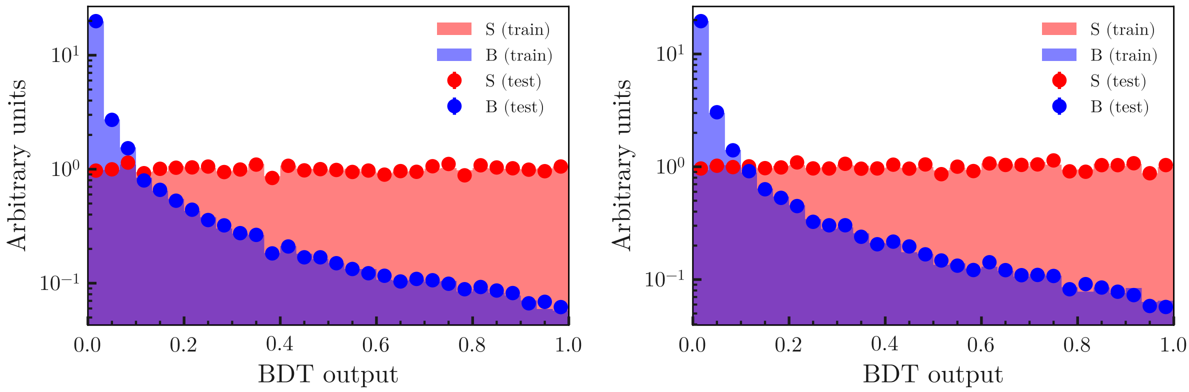


Figure 5.11: Signal BDT output flattening, with the even event training sample (left) and odd event training sample (right).

region, the purity and the efficiency. Moreover, having a "simple" value (which we are less likely to get when optimising a figure a merit) helps to have clear and simple bin boundaries, which helps with interpretability. However, we did try selections at 0.3 and 0.5, and they gave similar and worse results, respectively, on the limit in the final fit compared to 0.4. The studies regarding the alternative selections are shown in Appendix A.4.

The final SR for MC can be seen in Figure 5.12, from which we can look at the expected number of signal and background events per bin. There are 6 even bins in total. The data-MC agreement in the signal region for the various backgrounds is checked in the off-resonance sideband, for continuum only, and ET sideband for all contributions. There is a good agreement in the shapes for the two sidebands, especially in the four first bins, but a slight difference in the last two bins. As they are within a reasonable margin (2σ in the pulls), we still consider the agreement to be good. This can be observed in Figure 5.13.

Finally, we can check the signal efficiency and the expected background yields evolution after the various selection steps. This information is summarised in Table 5.6. The final signal efficiency in the signal region is 70×10^{-4}, which is almost two times higher than the one obtained by the previous hadronic tagging analysis. We can also compute the signal purity in the signal region, which is around 6%, and is also almost two times higher than the one obtained by the previous tagging analysis [5].

Cut	Cumulated sgn eff [$\times 10^{-4}$]	$B^0\overline{B}^0$ yield in 0.49 ab^{-1}	$B^+\overline{B}^-$ yield in 0.49 ab^{-1}	ccbar yield in 0.49 ab^{-1}	uds yield in 0.49 ab^{-1}
reconstruction + preselection	119.43 ± 0.34	8830	23364	29156	15661
BDT>0.4	70.74 ± 0.27	38	236	57	40
BDT>0.9	11.50 ± 0.11	0	5	1	0

Table 5.6: Table summarising the signal efficiency and background yields after the various selection steps.

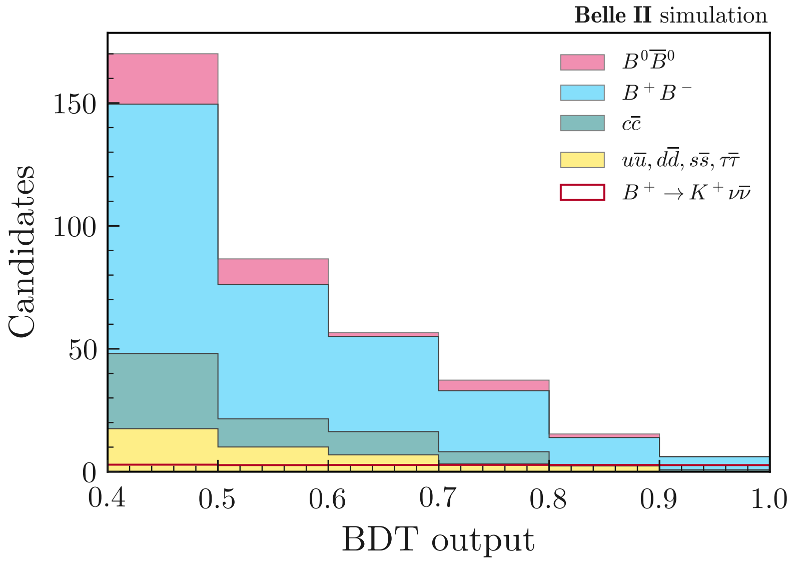


Figure 5.12: BDT output distribution for signal and background events in MC, in the signal region.

5.6 Background studies

We study the backgrounds in the signal region and check the MC truth information. The data/MC agreement before and after selection is checked using the three control samples defined in Section 5.3.

5.6.1 Background composition

In the signal region, we count 77 unique B decays, 10 of which are modes generated by the `PYTHIA` package. A sample of the most important modes is shown in Figure 5.15 for the charged modes and for the neutral modes. The most important mode is $B \to D\ell\nu$, $\ell \in e, \mu$, which represents half of our backgrounds. With our most important background being the $B \to D\ell\nu$ decay, we need to understand:

- How does it mimic the signal decay $B^+ \to K^+\nu\overline{\nu}$?
- Can we reduce its contribution in the SR?

Let's start by answering the first question. By looking at the truth information and all the particles in the final state, we obtain that the full decays are of the form $B^+ \to D^0(\to K^+X^-)\ell\nu$ where X^- is a charged hadron. In order to mimic the signal decay, it is necessary that neither the lepton nor the X^- are reconstructed. For the lepton, we did indeed observed that it is lost outside of the CDC acceptance ($\theta \in [17°, 148°]$), making it impossible to reconstruct, as one can see in Figure 5.14.

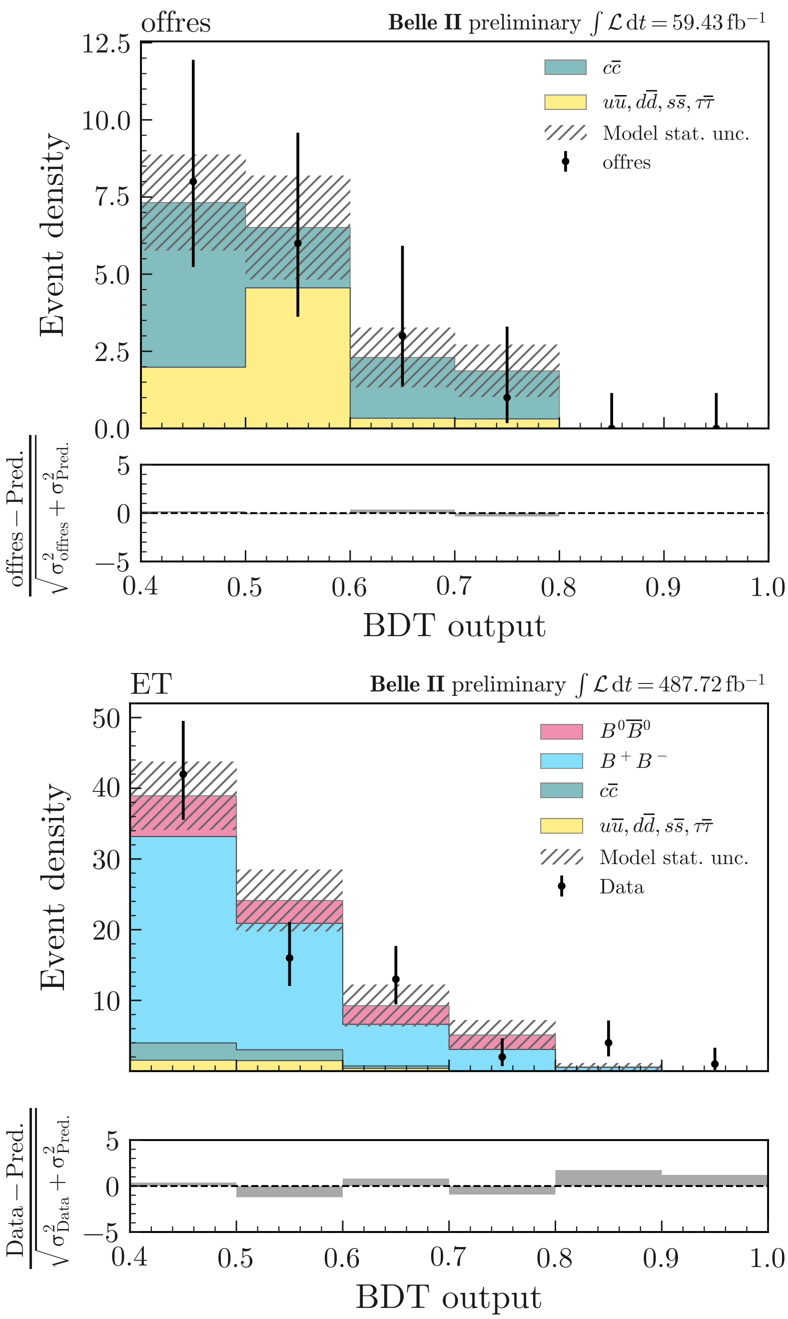


Figure 5.13: BDT output distribution for data and MC in the off-resonance sideband (above) and the ET sideband (below), in the signal region.

For the charged hadron accompanying the K^+, we verify the reconstructibiliy of the particle by looking at the truth information. We find that the X^- has no track found in about 92% of the events.

This two facts explain how the $B \to D\ell\nu$ decay can mimic the signal decay and indicates that it is a very difficult background to reduce, and we in fact consider it as irreducible.

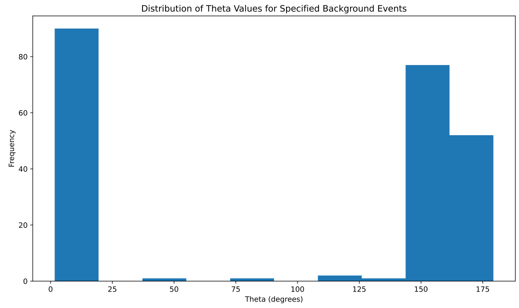


Figure 5.14: Angular distribution of the ℓ in the $B \to D\ell\nu$ backgrounds. The CDC covers an angle between 17 and 148°.

5.6.2 *Background validation*

To validate the background modelling in the SR, we use the same metrics as proposed in Section 5.4.9.

The data-MC agreement on the BDT output was already presented in Figure 5.13 and was deemed satisfactory. Regarding the data-MC ratio, they are shown in the signal region for all control samples in Table 5.7. The condition for a satisfactory agreement is that the ratio is within 2σ of unity, where σ is the uncertainty on the ratio. This condition is satisfied by all control samples.

Sideband	data/MC presel	data/MC signal region
offres	1.026 ± 0.017 (6181 data evts)	1.25 ± 0.35 (18 data evts)
wc	1.003 ± 0.014 (12195 data evts)	1.18 ± 0.24 (39 data evts)
extra	1.027 ± 0.010 (68520 data evts)	1.32 ± 0.20 (78 data evts)

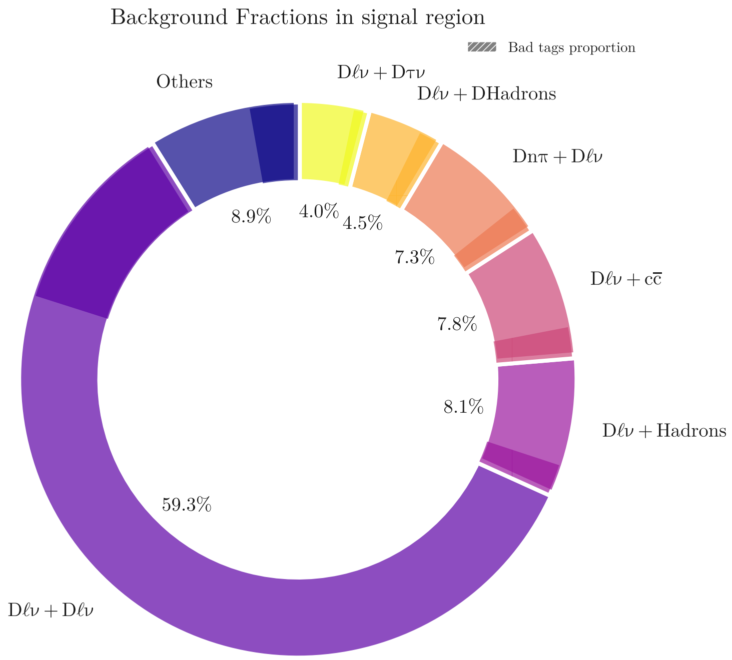
Table 5.7: Data-MC agreement on all three sidebands before and after the BDT selection.

Finally, the pertinence of the control samples in the SR is checked by looking again at the distribution of the backgrounds in MC for both signal sample and control samples. The distributions are shown in Figure 5.16.

5.7 *Systematic uncertainties*

Corrections are naturally uncertain as they are based on data and simulation from control samples. Hence, we need to assign a systematic uncertainty for each correction we apply. Afterwards, these systematic uncertainties will be implemented in the final fit as nuisance parameters, in order to constrain them and propagate their effect on the final result, as explained previously in Section 3.6. The uncertainty sources and treatment of nuisance parameters are summarised in Table 5.8.

From this point on, the crossfeed contributions will be considered separately from the $B\bar{B}$ contributions as 3 extra contributions.



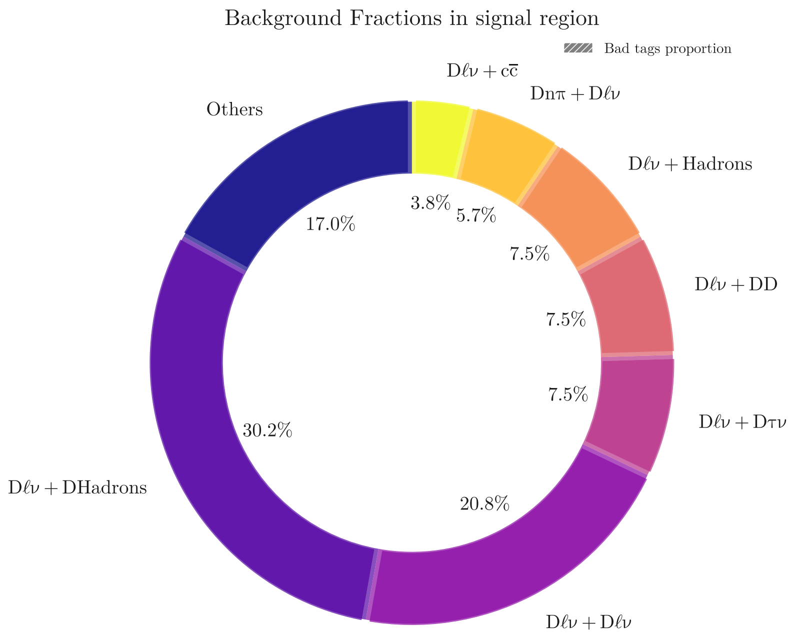


Figure 5.15: Pie charts representing the proportions of different B decay modes in the backgrounds of our SR. The first part of the mode corresponds to the B_{tag} while the second is the alleged B_{sig}. The dashed part indicates the proportion of wrongly tagged events. The top pie chart shows the B^+B^- modes while the bottom one shows the $B^0\bar{B}^0$ modes. Since we are looking for a charged mode, it is expected that the remaining $B^0\bar{B}^0$ modes in the SR are badly-tagged.

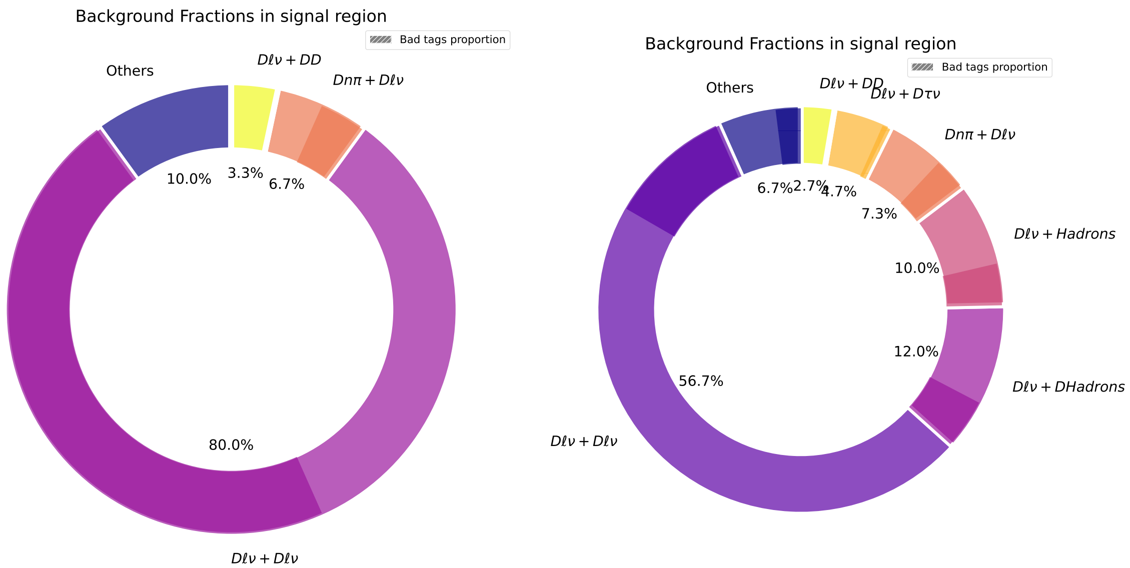


Figure 5.16: Pie charts representing the proportions of different B decay modes in the backgrounds of our SR. The left plot is for the WC sideband and the right plot is for the ET sideband.

5.7.1 *Systematic uncertainties treatment*

As specified, the systematic uncertainties are implemented as nuisance parameters in the fit based on the `HistFactory` formalism (cf. Section 3.6). In this formalism, modifiers of the expected values in each bin of the histogram we want to fit are constrained by auxiliary measurements. Those auxiliary measurements for bin-correlated uncertainties are alternative histograms, named "high" and "low" data, corresponding to $\pm 1\sigma$ for the Gaussian associated to the nuisance parameter. Using interpolating functions with the auxiliary measurements, the `pyhf` package then computes those modifiers.

5.7.2 *PID systematic*

Systematic uncertainties on the PID corrections are taken from the same table we extract the corrections from. They are presented beside the corrections in Figure A.1. The uncertainty we quote is then the relative uncertainty on the correction. While this process is event-based, the fact that two events from different bins in the SR or from different contributions (signal, B^+B^-, $B^0\bar{B}^0$, $q\bar{q}$ or $c\bar{c}$) may share the same bin in the PID table can introduce correlations. To deal with this, we use a toy MC method. We generate 500 replicas of the relative uncertainties on the PID correction, distributed according to a LOGNORMAL centred at 0 and with a σ equal to the relative uncertainty if it is smaller than 1, and equal to 1 otherwise. From those replicas, we can compute the covariance matrix between each contribution and each bin in the SR, giving a total of 48 cases

(8 contributions and 6 bins in the SR), making the covariance matrix 30×30.

In order to decorrelate the systematic uncertainties, we can use the technique introduced in Section 3.4. In our case, we end up keeping 22 eigenvectors, which makes a reduction of about 54% on the number of nuisance parameters in the final statistical model.

The auxiliary measurements for the 22 nuisance parameters introduced are then computed from the corresponding eigenvectors. We multiply the eigenvector by the nominal histogram (*i.e.* the histogram used for the fit) for the corresponding contribution. The high or low data is then obtained by adding or subtracting this product to the nominal histogram.

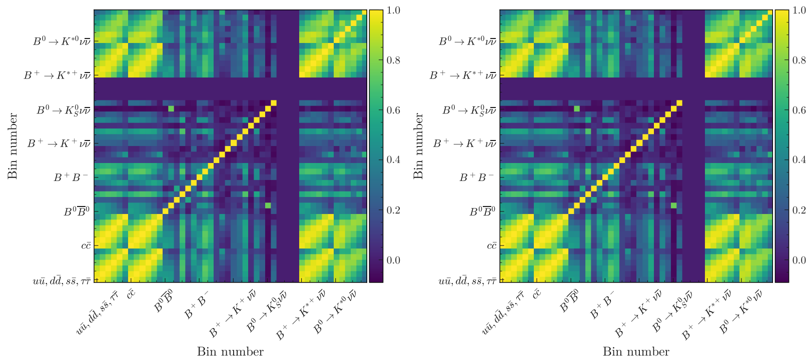


Figure 5.17: Covariance matrix of the PID systematic uncertainty. The left matrix is the full covariance matrix, while the right matrix is the covariance matrix after keeping only the 22 most important eigenvectors. We can see that this new matrix perfectly reproduces the main features of the original one.

5.7.3 Signal MC form factor systematics

The form factor systematics are calculated by reweighting simulated events using eigenvector variations of the form factor covariance matrix, presented in Equation (1.33), which are propagated via eigendecomposition. Since the form factor has three fit parameters, we have three form factor nuisance parameters. The residual of the variations in q^2 relative to the nominal form factor model is taken as the systematic uncertainty.

5.7.4 Data-MC discrepancy

For the nuisance parameter related to the data-MC discrepancies, the auxiliary measurements are derived from the alternative histograms obtained by removing the respective corrections. We then compute the difference between the nominal and the alternative histograms, before adding or subtracting this difference to the nominal histogram to obtain the high and low data:

$$\text{auxiliary data} = \{\text{nominal} \pm (\text{nominal} - \text{alternative})\} \tag{5.4}$$

5.7.5 *Branching fraction correction*

For the systematic associated with the branching fraction correction, we consider 3 cases:

- If the branching fraction exists in the PDG and has an uncertainty, we compute the relative uncertainty using those values;
- If the branching fraction exists in the PDG but is inclusive, with resonant and non-resonant contributions, we take the PDG uncertainty for the upper limit, but we take a 100% relative uncertainty for the lower limit;
- if the decay mode is generated via PYTHIA, we take a 100% relative uncertainty for the lower limit, but we do not put a limit on the upper bound (technically, we put a limit of 10000% relative uncertainty if the decay populates more than one bin in the SR, otherwise 1000%).

The second point is motivated by how those modes are handled by the Belle II simulation, as explained in Section 5.4.3. For the third point, we argue that a simple 100% symmetric relative uncertainty is not enough, as the correction on the branching fraction can be way higher than 2 times the original value. To be as conservative as possible, we thus choose not to put any upper limit on the relative uncertainty (or to put one absurdly big enough to be considered as such). Alternatively, for backgrounds present in only one bin in the SR, we reduce the uncertainty for fit stability reasons.

Additionally, we also extract the histogram corresponding to the BDT output for each mode. As explained in the corresponding section, we only consider cases where the B_{sig} decays into the mode. The high and low datasets are then the nominal histogram for which we vary the number of events of the given mode according to the relative uncertainty:

$$\text{auxiliary data} = \left\{ \text{nominal} \pm \sigma_{\pm}^{\text{rel}}(mode) \times \text{histogram(mode)} \right\} \quad (5.5)$$

This gives us a total of 77 nuisance parameters for the branching fraction corrections.

5.7.6 *Specific background corrections*

The same process is applied for the specific background corrections. We introduce 5 systematic sources corresponding to 5 cases:

- $B \to K n \bar{n}$ decays, with a 100% relative uncertainty;

- $B \to KK_L^0K_L^0$, decays with a 20% relative uncertainty;
- events with $D \to K_L^0$, with a 40% relative uncertainty;
- events with a K_L^0, with a 100% relative uncertainty.

As explained previously, we only look at the signal side for those decays.

The 100% relative uncertainty for the $B \to Kn\bar{n}$ decays is motivated by multiple uncertainties regarding the derivation of the correction for this mode, such as the imperfect modelling of the threshold enhancement or the isospin symmetry assumption between $B^+ \to K^+n\bar{n}$ and $B^0 \to K^0p\bar{p}$ [90].

The choice for the 20% relative uncertainty for the $B \to KK_L^0K_L^0$ decays is based on the precision of the data-driven checks performed in [91] and the isospin symmetry assumption between $B^+ \to K^+K_L^0K_L^0$ and $B^+ \to K^+K_S^0K_S^0$. For the events with $D \to K_L^0$, we take a 40% relative uncertainty. This number is actually the sum of the 30% correction applied by the previous analyses [5] and the 10% relative uncertainty they quote. Finally, for the events with a K_L^0, the quoted relative uncertainty is inflated to 100% from the sum of the correction (17%) and the systematic uncertainty (50%) used by the previous Belle II analysis.

The auxiliary measurements for those 5 systematic sources are then derived just like for the branching fraction correction.

5.7.7 Signal efficiency validation

We repeat the analysis pipeline up to the SR selection, for both the data and simulation embedding. We then compute the selection efficiency at this stage for both samples, and we compute the ratio. With this ratio, we can check the data-MC agreement of the selection efficiency, which is compatible with 1, and we also get an associated uncertainty of 11%. This uncertainty is used in the fit as a normalisation systematic for the signal contribution.

5.7.8 Normalisations

Regarding the uncertainties related to the normalisations, we introduce first one nuisance parameter per contribution except for the signal. For $B\bar{B}$ and $q\bar{q}$ contributions, the systematic uncertainties are derived from the relative uncertainty remaining from the data-MC ratio in the signal region combined from WC and ET sidebands:

$$\sigma^{\text{rel}}_{data/MC} = \frac{\sqrt{\frac{1}{\sigma^2_{ET}} + \frac{1}{\sigma^2_{WC}}}}{\frac{r_{ET}}{\sigma^2_{ET}} + \frac{r_{WC}}{\sigma^2_{WC}}} \tag{5.6}$$

The auxiliary data are then computed as $1 \pm \sigma^{\text{rel}}_{data/MC}$. For each crossfeed contribution, the auxiliary data are defined as:

$$\text{low data} = 0 \tag{5.7}$$

$$\text{high data} = 1 + \frac{\text{Current experimental upper limit}}{BF_{theo}} \tag{5.8}$$

Additionally, we introduce a nuisance parameter correlated between the signal and the two $B\bar{B}$ contributions. This nuisance parameter is related to the systematic uncertainty on the B meson counting. The values are:

$$f_{\pm} = 0.5113 \pm 0.0108 \tag{5.9}$$

$$N_{\Upsilon(4S)} = (520.6 \pm 7.6) \times 10^6 \tag{5.10}$$

$$N_B = 2 \times N_{\Upsilon(4S)} \times f_{\pm} = (532.4 \pm 13.7) \times 10^6 \tag{5.11}$$

$$\sigma^{\text{rel}}_{N_B} = \frac{13.7}{532.4} = 2.6\% \tag{5.12}$$

where $f_{\pm}$ is the probability of an $\Upsilon(4S)$ to decay into a B^+B^- pair.

5.7.9 MC statistical uncertainty

Finally, we also introduce a nuisance parameter for the MC statistical uncertainty. This nuisance parameter is uncorrelated between each contribution and each bin. There are two inputs for this nuisance parameter:

- The Poisson uncertainty on the number of events in each bin.
- The 26 remaining eigenvectors from the PID systematic uncertainty, as explained in Section 5.7.2.

The two inputs are summed in quadrature to give the total uncertainty for the MC statistical uncertainty, but the PID part is negligible.

5.8 Signal Extraction

5.8.1 Method

As indicated in Section 3.6, our statistical model is based on the `HistFactory` formalism, using the `pyhf` and `cabinetry` packages

Source	Contribution(s)	Treatment	Number of NP
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

Table 5.8: Summary of the systematic uncertainties implemented, with the contribution they affect, the way they are treated and the corresponding number of nuisance parameters.

[93]. We build a template binned likelihood of the binned BDT output distribution, with 6 bins in the range of the SR, $[0.4, 1.0]$. There are 8 contributing processes: the signal, the B^+B^-, the $B^0\bar{B}^0$, $c\bar{c}$, $uds\tau$ (includes together $u\bar{u}$, $d\bar{d}$, $s\bar{s}$ and $\tau\bar{\tau}$) as well as the crossfeeds from the 3 others modes ($B^0 \to K^0_S\nu\bar{\nu}$, $B^+ \to K^{*+}\nu\bar{\nu}$, $B^0 \to K^{*0}\nu\bar{\nu}$).

The parameter of interest is the signal strength μ. For the systematic uncertainties, we consider the 169 nuisance parameters described in the previous section and summarised in Table 5.8.

To estimate the parameter of interest, we perform a binned maximum likelihood fit. First, we perform an Asimov fit to verify the model and its implementation.

The fit is realised in two steps:

- We perform an initial fit using the `scipy` optimiser, via sequential least squares programming, to get a first estimate of the parameters. This is necessary for computational reasons.
- We then perform a second fit using the `iminuit` optimiser [94], as it is more accurate and gives us an uncertainty on the parameter of interest, as well as a covariance matrix, via the `HESSE` algorithm.

5.8.2 *Asimov fit results*

We first perform a fit on the Asimov dataset, *i.e.* the dataset for which the observed data is equal to the expected data. The expected data per bin is given in Table 5.9. The expected number of signal yields is computed from the theoretical branching ratios.

Bin number	expected signal yields	expected $B^0\bar{B}^0$ yields	expected B^+B^- yields	expected $c\bar{c}$ yields	expected uds yields	expected $B^0 \to K^0_S \nu\bar{\nu}$ yields	expected $B^+ \to K^{*+}\nu\bar{\nu}$ yields	expected $B^0 \to K^{*0}\nu\bar{\nu}$ yields
1	2.9	20	1e+02	31	17	0.00016	0.036	0.44
2	2.7	10	55	11	10	0.00018	0.027	0.33
3	2.7	1.3	39	9.4	6.8	0	0.022	0.23
4	2.8	4.1	25	5	3.2	0	0.018	0.19
5	2.8	1.3	11	0.26	2.3	0	0.015	0.095
6	2.7	0	5.3	0.78	0	0	0.01	0.059

Table 5.9: Expected yields for the various contributions in the signal region.

Using this dataset, we can estimate the correlations between the parameters and the impact of the uncertainties on the signal strength. The method to compute the impact of the uncertainties is described in Section 3.8. We find an estimate for the uncertainty of 5.2×10^{-6}, or 1.2 in units of the signal strength μ. The individual impact of each contribution is given in Table 5.10. The uncertainty is statistically dominated. For the systematic uncertainties, the main impact comes from the limited MC statistics, followed by the background normalisation. The correlation matrix of the statistical model's parameters is shown in Figure 5.18, where we only represent the parameter with at least one correlation above 0.2 in absolute value.

Systematic Category	**Absolute contribution on branching fraction ($\times 10^{-6}$)**
StatMC	2.2
KL	1.0
Continuum	0.8
bkg norm	0.6
sgn eff + tracking	0.5
Backgrounds	0.4
KKLKL	0.3
nBB	0.2
Data/MC extra	0.1
FF	0.1
PID	0.1
D2KL	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**2.7**
Total statistical	**4.3**
Total	**5.1**

Table 5.10: Impacts of the various types of nuisance parameters on the final systematic uncertainty.

In the correlation matrix: `charged_err`$[i]$ corresponds to the MC statistical error associated to the charged mode B^+B^- in bin i; `norm_charged` is the systematic uncertainty associated to the B^+B^- normalisation; `-20423,14,-13` is the $B^+ \to \bar{D}_1\,(2430)^0\,\mu^+\nu_\mu$ decay, `-2112,2112,321` is the $B^+ \to K^+ n\bar{n}$ mode and `-421,311,-311,211` is $B^0 \to \bar{D}^0 K^0 \bar{K}^0 \pi^+$.

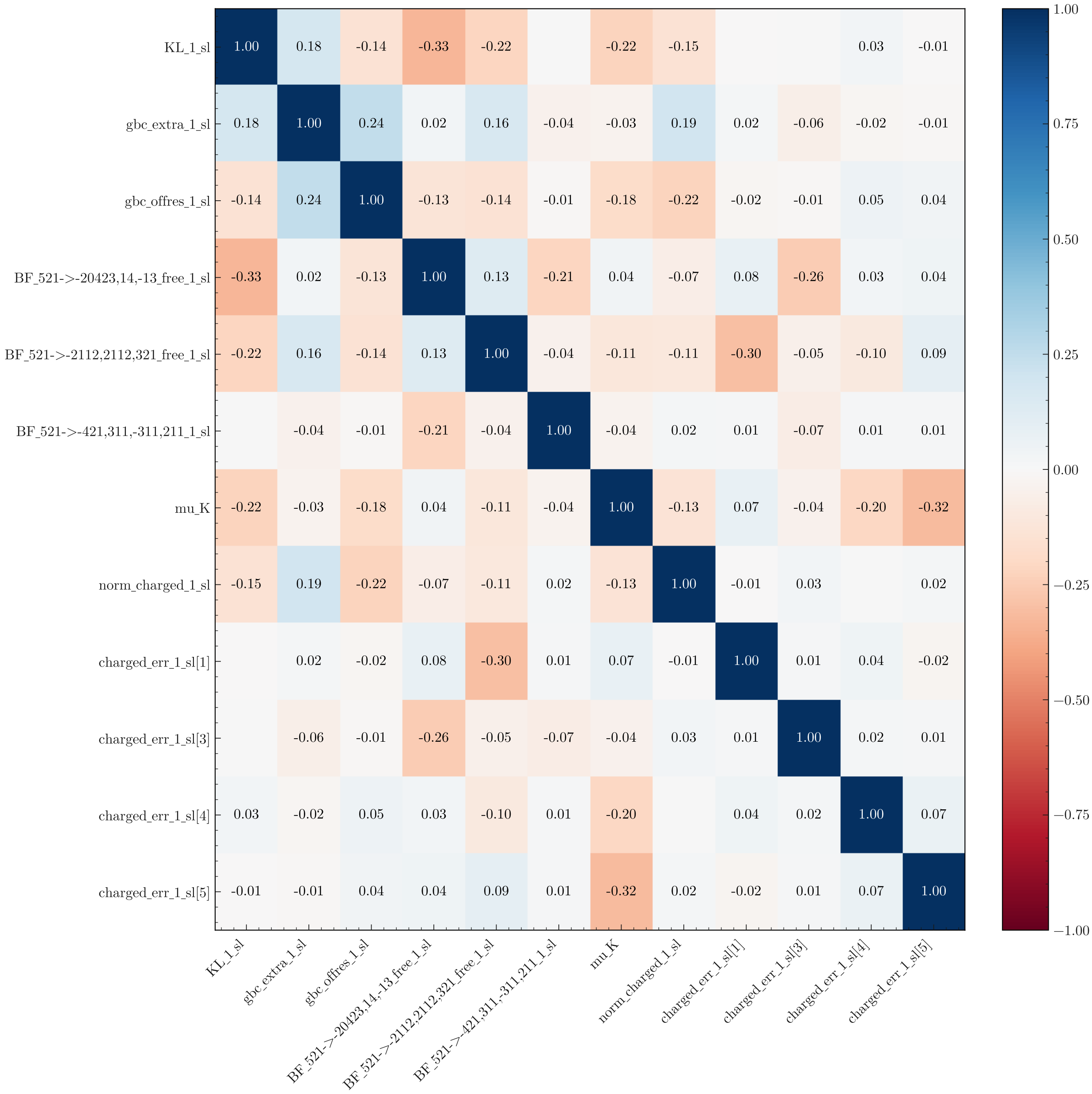


Figure 5.18: Correlation matrix of the nuisance parameters. Only the nuisance parameters with at least one correlation above 0.2 are shown.

The fitted profile negative log-likelihood is shown in Figure 5.19, with some CL of interest. We also show the results of the CL_s method in Figure 5.20, described in Section 3.7.

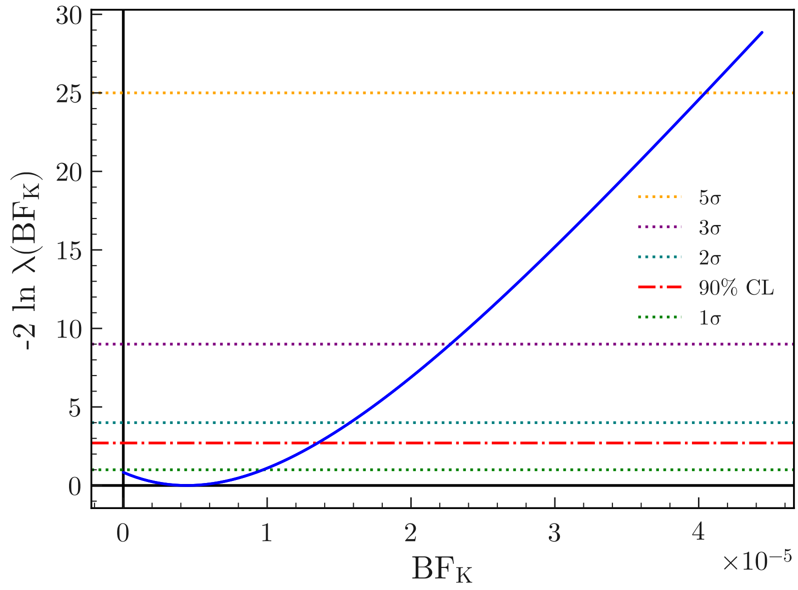


Figure 5.19: Profile negative log-likelihood on the branching fraction, with an injected signal strength of 1, with the various lines representing different CL. One can then simply read the corresponding CL by checking the intersection with the likelihood function.

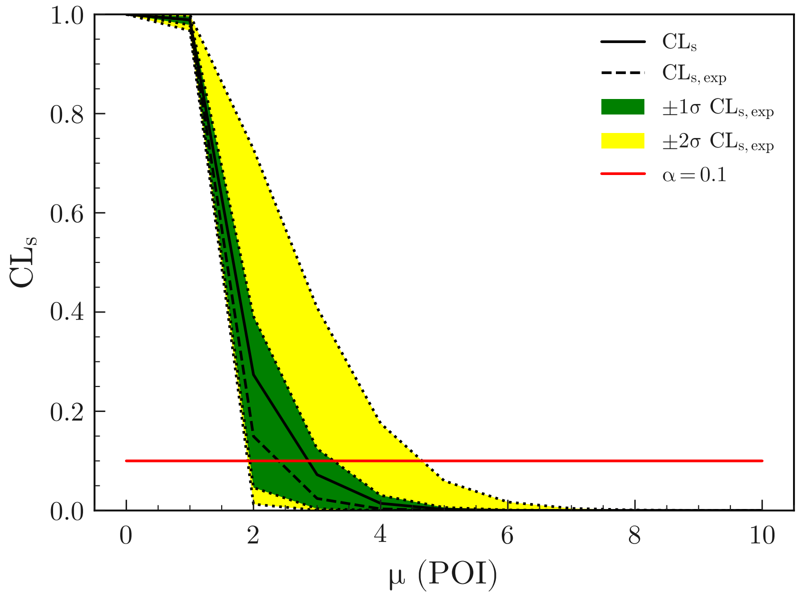


Figure 5.20: Plot showing the CL_s scan with the $\pm 1\sigma$ and $\pm 2\sigma$ belts as well as the 90% confidence level line.

5.8.3 Signal injection test

The final test for our statistical model is to perform a signal injection test. With this test, we verify that the model is not biased and that it is able to recover the injected signal strength. We perform this test by generating toy datasets, varying both the nuisance parameters via Gaussian distributions and the expected number of events in each bin via a Poisson distribution. We then perform a fit on each of these toy datasets, and we compare the median of the fitted signal strength with the injected one. To verify the robustness of the model for different signal strength hypotheses, we perform this test for different injected signal strengths $\mu_{\text{inj}} \in [1, 3, 5, 7, 10]$. We do not expect the observed branching ratio to be more than 10 times the SM prediction because of the current upper limits in the branching ratio value, so we do not investigate higher values. The results of this test are shown in Figure 5.21. We find that the model is unbiased across the full range of injected signal strengths, which validates our statistical model and its implementation.

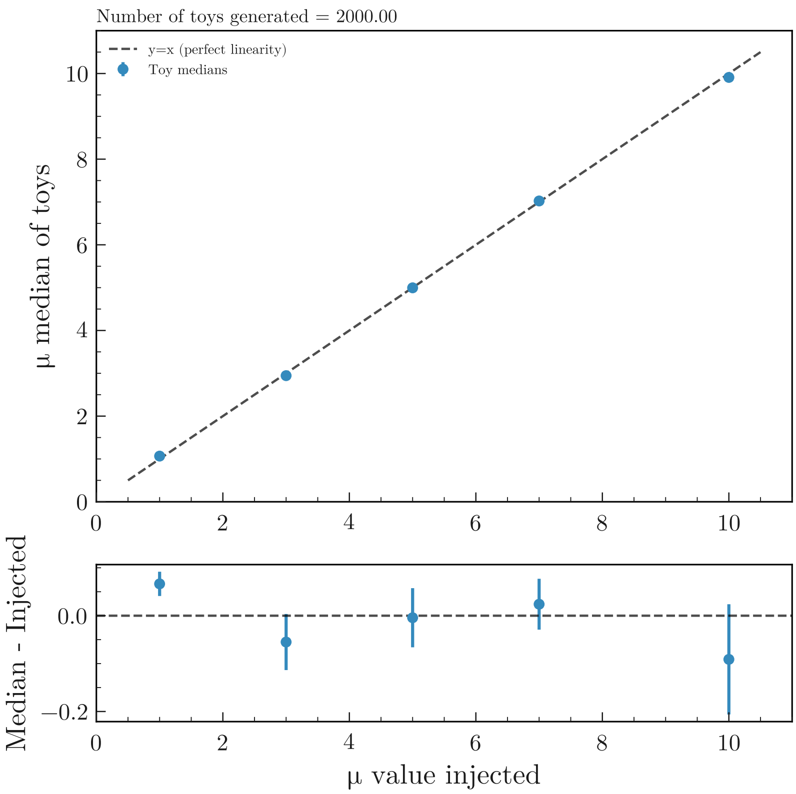


Figure 5.21: Results of the signal injection test. The median of the fitted signal strength is shown as a function of the injected signal strength. We can clearly see how they follow the dashed line, which represents the unbiased case.

5.9 *Comparison with the state of the art and discussion*

5.9.1 *Comparison to the previous analyses*

In this section, we will discuss the results we obtained, comparing them with the previous analyses.

Firstly, instead of fixing the signal strength μ to 1 in the Asimov dataset, we can try to fix it to the value found in the previous

Belle II analysis to compare and cross-check the results. The value obtained then was $\mu = 4.6 \pm 1.3$. [5] By injecting this number into the same procedure as described in the previous section, we obtain a new Asimov dataset, and we can fit it, looking at the same metrics of interest (Table 5.11, Figure 5.22). Moreover, we can check the significance of the result, as computed in Section 3.7, and we get 2.43. This preliminary expected significance indicates that we would not get evidence for 4.6 times the predicted SM value of the branching ratio, so we would not be able to confirm the results from the previous analysis. However, comparing the obtained uncertainty for this fit with the uncertainty from the previous Belle II analysis [5], we can see that we are totally on par. It proves the importance of this independent SL tagged analysis for the tension previously observed.

Systematic Category	**Absolute contribution on branching fraction** ($\times 10^{-6}$)
sgn eff + tracking	2.3
StatMC	2.1
bkg norm	1.7
KL	1.4
Continuum	0.9
nBB	0.6
Backgrounds	0.5
FF	0.4
KKLKL	0.3
Data/MC extra	0.2
PID	0.2
D2KL	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**4.0**
Total statistical	**5.7**
Total	**7.0**

Table 5.11: Impacts of the various types of nuisance parameters on the final systematic uncertainty, with an injected μ of 4.6.

Secondly, we can compare the estimated uncertainty from our Asimov fit (with $\mu = 1$) with the uncertainties obtained by the previous analyses, as presented in Figure 1.5. The results are summarised in Table 5.12. We can see that we have a smaller absolute

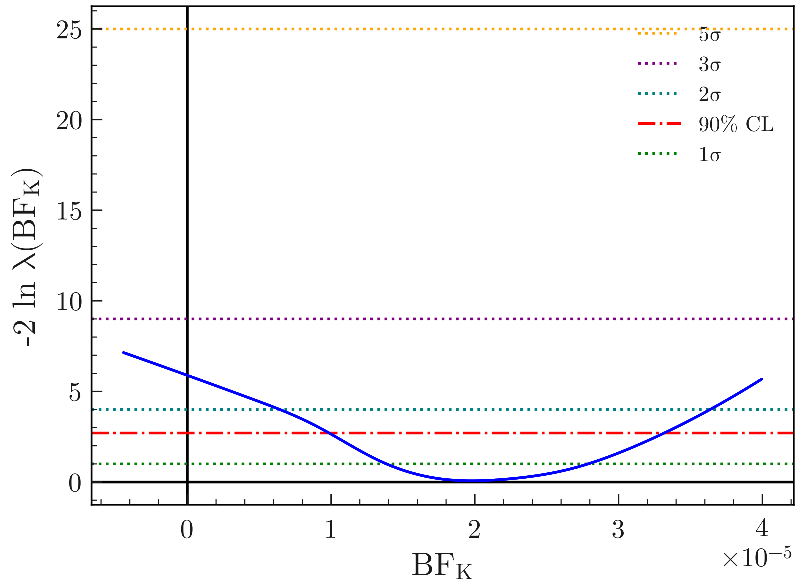


Figure 5.22: Profile negative log-likelihood on the branching fraction, with an injected signal strength of 4.6, with the various lines representing different CL. We can see a non-parabolic behaviour on the left-side of the minimum that has not yet been explained. If we had a more Gaussian-like likelihood, we would expect a way better significance.

uncertainty than other analyses, making this SL tagged analysis a pivotal analysis for a precise measurement of the $B \to K^{(*)}\nu\bar{\nu}$ decays. Even when naively scaling the uncertainties to the luminosity used for this analysis, we can see that our expected uncertainty stays lower than the others. This study has two limitations:

- Our uncertainty was obtained on an Asimov dataset considering a signal strength of 1, but we know it increases with the signal strength;
- The scaling for the uncertainties was done by naively multiplying the absolute uncertainty by $\sqrt{\frac{\text{Luminosity of the analysis}}{\text{Luminosity for our analysis}}}$, which does give a rough idea but is not correct.

Analysis	Absolute uncertainty $(\times 10^6)$	Naively scaled to 487.72 fb^{-1} $(\times 10^6)$
BaBar Hadronic (HAD) [35]	13	12
BaBar SL [34]	8	7
Belle HAD [36]	16	19
Belle SL [37]	6	7
Belle II Inclusive (INC) [5]	7	6
Belle II HAD [5]	12	10
Belle II HAD + INC [5]	7	6
This analysis	5	5

Table 5.12: Comparison of the previous analyses' uncertainties with our expected uncertainty.

5.9.2 *Discussion on the combination with other analyses*

There are two main combinations to be done with the results from our analysis.

As specified in the introduction, our team worked on both $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$. The goal in the end is to do a combined fit with both channels, to explore the $\left(\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right), \mathcal{B}\left(B^0 \to K^{*0}\nu\bar{\nu}\right)\right)$ space as described in Section 1.4.

Furthermore, we will combine with two other Belle II analyses on the same channels: the INC and HAD tagged analyses. Those two are direct continuations of the ones performed in [5]. To properly perform this combination, as the datasets are the same, we need to have an understanding of the overlap. It is shown in Figure 5.23. From it, we can compute the relative number of overlap for the INC (ITA) and our analysis (STA):

$$\frac{ITA \cap STA}{ITA} = 0.3\% \tag{5.13}$$
$$\frac{ITA \cap STA}{STA} = 25\% \tag{5.14}$$

We can see how independent the three analyses are, which is again a reminder of the importance of our work.

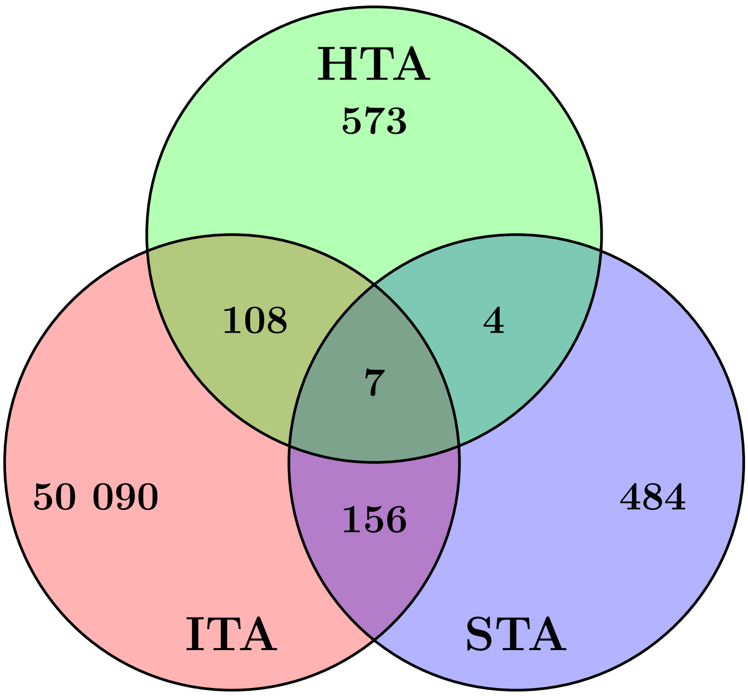


Figure 5.23: Wenn diagram showing the number of events in the SR of the INC (ITA) and HAD (HTA) tagged analyses from [5] and in our work (STA) for run 1, scaled to a luminosity of 1 ab^{-}1, with the overlaps. The numbers of events unique to each analysis were obtained from data for the two previous analyses and from MC for ours, as we did not unblind yet. Overlaps were obtained from data for all analyses, but using a ratio of events and considering a data-MC ratio of 1 for our analysis to avoid any unblinding.

Overview and outlook

In this thesis, we detailed the search for $B^+ \to K^+\nu\overline{\nu}$ at the Belle II experiment using a SL tagged analysis.

Chapter 1 has opened this work with theoretical and contextual aspects of the search, motivating the need for this measurement in the context of the SM and physics beyond the SM.

Chapter 2 has described Belle II, the experimental apparatus used for the search. We explained in more detail why it is the only place in the world where this analysis can be done, thanks to the cleanness of the events and the hermicity of the detector.

Chapter 3 has introduced the toolbox necessary to carry out the tasks we carried, such as BDTs, important discriminating variables and the statistical knowledge to extract a signal or compute a limit and an uncertainty.

Chapter 4 has explored the calibration of an important tool for the study of CP-violation at Belle II, the `GFlaT`. We especially mentioned upgrades, with a gain in flexibility and computing time thanks to the parallelisation.

Finally, Chapter 5 has detailed the full workflow of the $B^+ \to K^+\nu\overline{\nu}$ analysis, starting from the reconstruction of the 487.72 fb^{-1} of data, exploiting the features of the detector described in Chapter 2, handling the correction of the simulation, the training of a BDT and the statistical aspects with the tools introduced in Chapter 3 and finally discussed the results in the light of the state of the art given in Chapter 1. We obtained an expected uncertainty of 5.2×10^{-6} for an injected signal corresponding to the SM expectation, and 7×10^{-6} for an injected signal equal to the branching fraction obtained by the previous Belle II analysis, both values being better or on-par with the state of the art. Combining this precision with the low overlap with the previous Belle II analysis, we once again reaffirm the importance of this independent measurement using a SL tagging method, whether it is for cross-checking the previous measurements or combining to get a more significant result.

Yet, it is not the end of this work. As explained previously, we did not unblind the analysis yet, so all results are solely simulation-based. Despite the very good data-simulation agreement observed in the control samples, unblinding is a crucial step in the analysis, as unexpected mismodelisation may appear. Moreover, we still have to continue the analysis on the $B^0 \to K^{*0}\nu\bar{\nu}$ channel in order to, *in fine*, do a combined analysis, effectively constraining the $\left(\mathcal{B}\left(B^+ \to K^+\nu\bar{\nu}\right), \mathcal{B}\left(B^0 \to K^{*0}\nu\bar{\nu}\right)\right)$ space in search for possible NP.

Lastly, other analyses are currently ongoing at Belle II on the same channels, also on the $B^0 \to K^0_S\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$ modes, using the previously cited methods. Those analyses should also be combined with ours to give a more precise and less biased estimate of the measurement or limit on the branching fractions of the $B \to K^{(*)}\nu\bar{\nu}$ decays.

Finally, with the new data being collected by the Belle II experiments, and the improvement of the statistical and analysis tools, the search can only continue and improve, leading to exciting new results. In the longer term, the recent European Strategy for Particle Physics update is predicted to allow a gain of one order of magnitude in precision for the measurement of $b \to s\nu\bar{\nu}$ transitions probabilities [95].

A

Appendix

A.1 PID tables

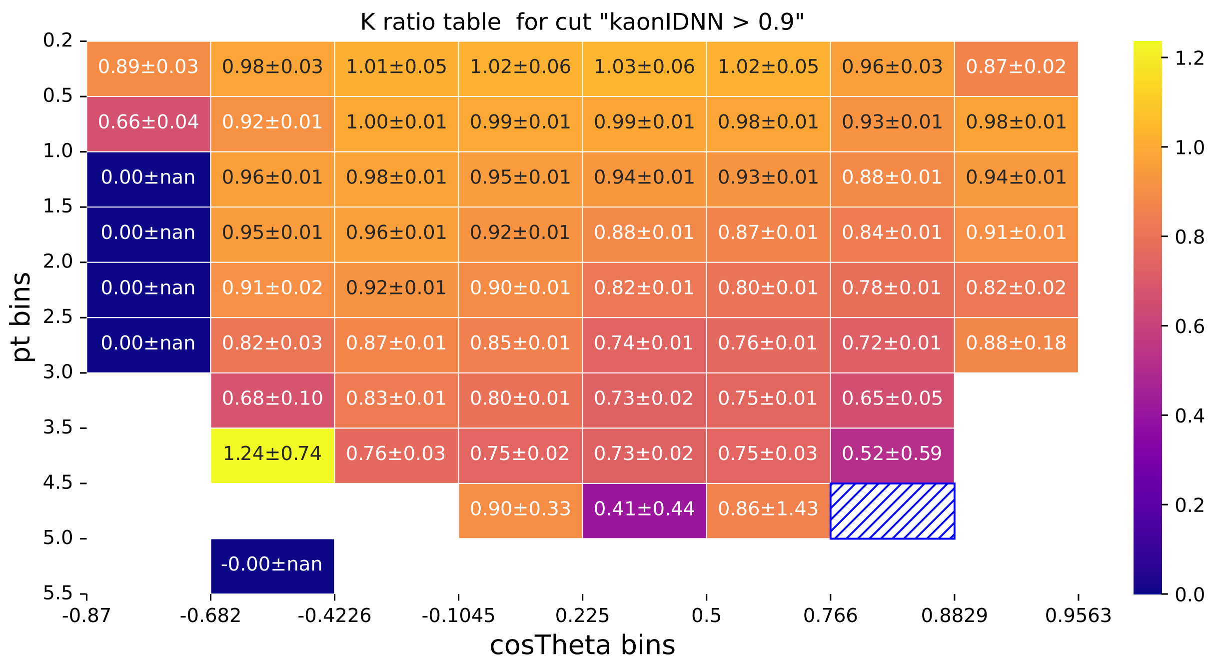


Figure A.1: Data-MC ratio efficiency of the PID for K^+ as a function of p_T and $\cos(\theta)$.

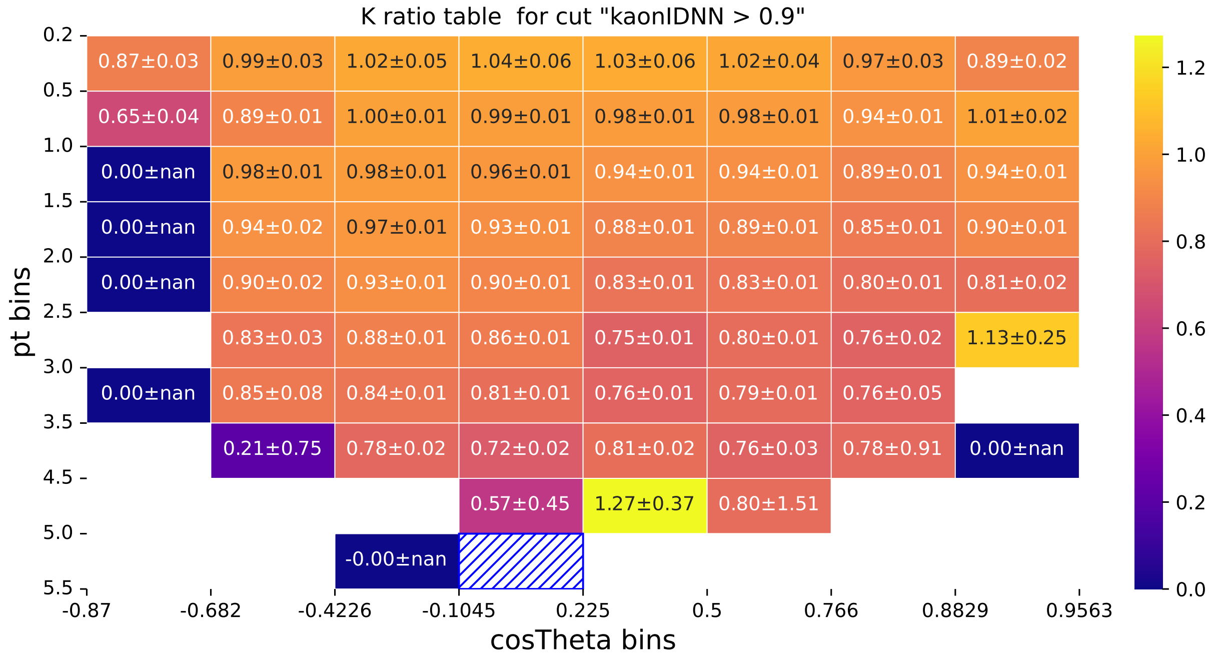


Figure A.2: Data-MC ratio efficiency of the PID for K^- as a function of p_T and $\cos(\theta)$.

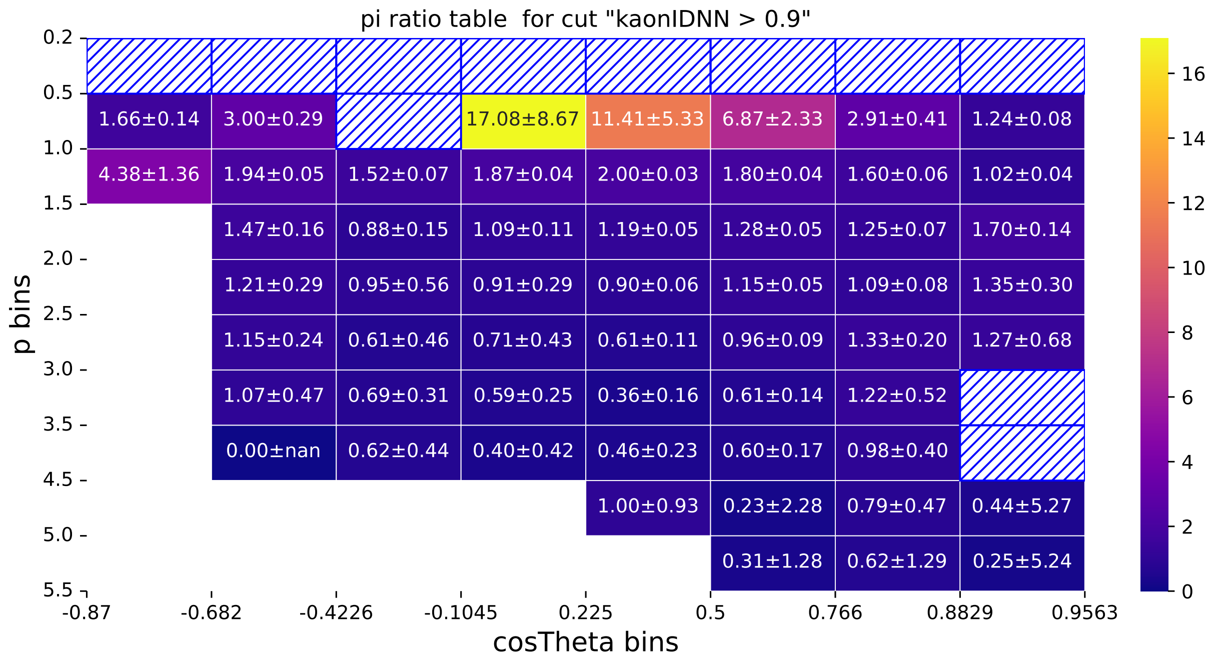


Figure A.3: Data-MC ratio π fake rate of the PID for K^+ as a function of p_T and $\cos(\theta)$.

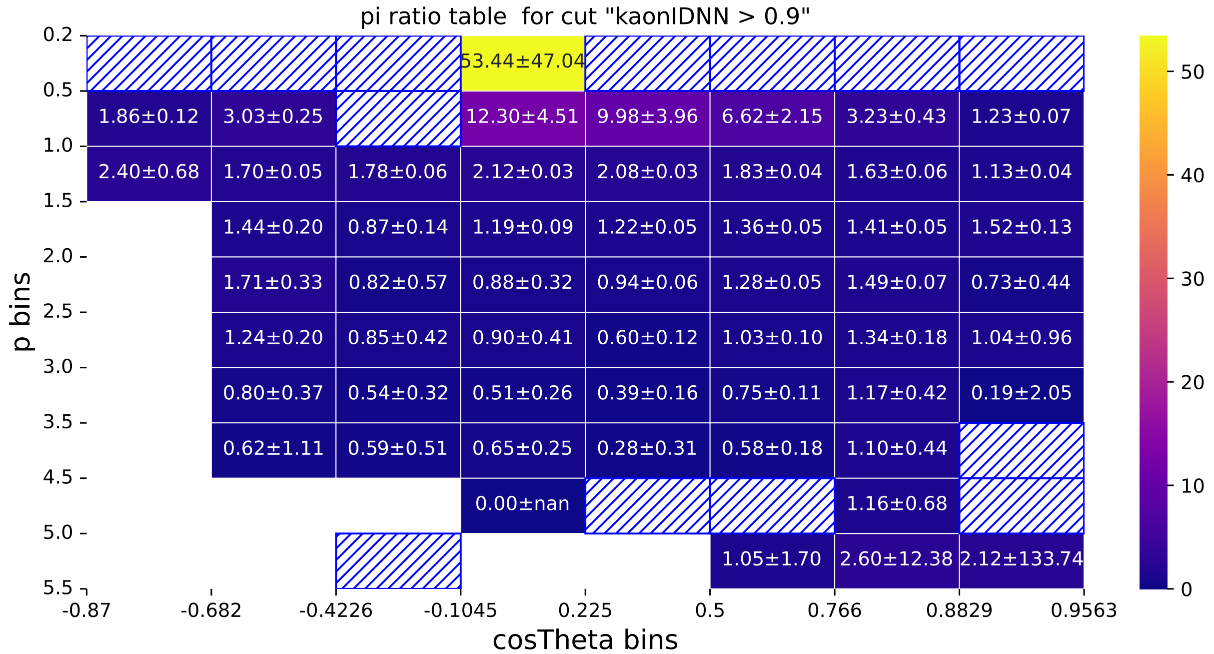


Figure A.4: Data-MC ratio π fake rate of the PID for K^- as a function of p_T and $\cos(\theta)$.

A.2 BDT variables

In this section, we will give an explanation for each of the 18 variables used in the BDT:

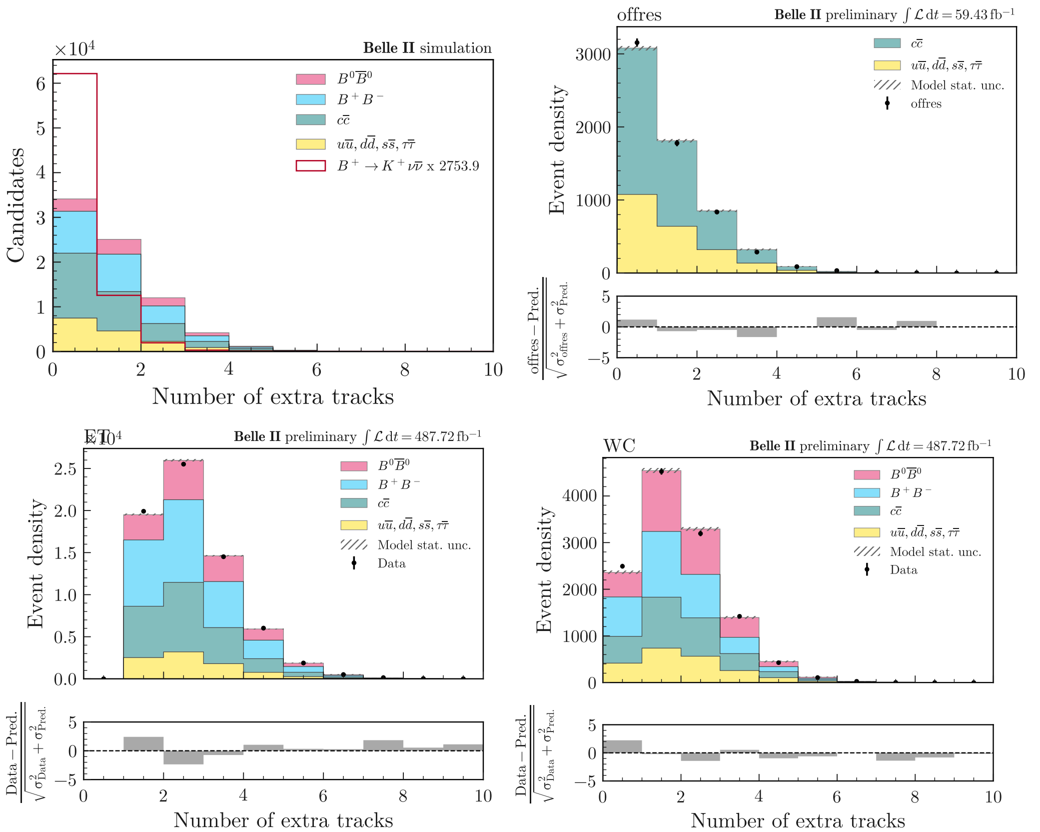
- Number of extra tracks: number of tracks in the ROE
- $E_{miss} + cp_{miss}$: Sum of the missing system energy with the energy carried by the missing system momentum
- $H^{so}_{m,2}$: a KSFW moment as defined in Section 3.3
- q^2: the mass of the two neutrinos system
- $H^{so}_{c,2}$: a KSFW moment as defined in Section 3.3
- Neutral E^{extra}_{ECL} [GeV]: Extra energy coming from neutral particles in the ECL
- $H^{so}_{n,0}$: a KSFW moment as defined in Section 3.3
- $\cos(\theta_{BY})$: as defined in Section 3.3
- $p^*_{B_{tag}}$: momentum of the B_{tag} in the CMS frame
- $m_D - m_D^{PDG}$: mass of the reconstructed D meson coming from the B_{tag} minus its nominal mass
- p_ℓ: momentum of the ℓ from the B_{tag}

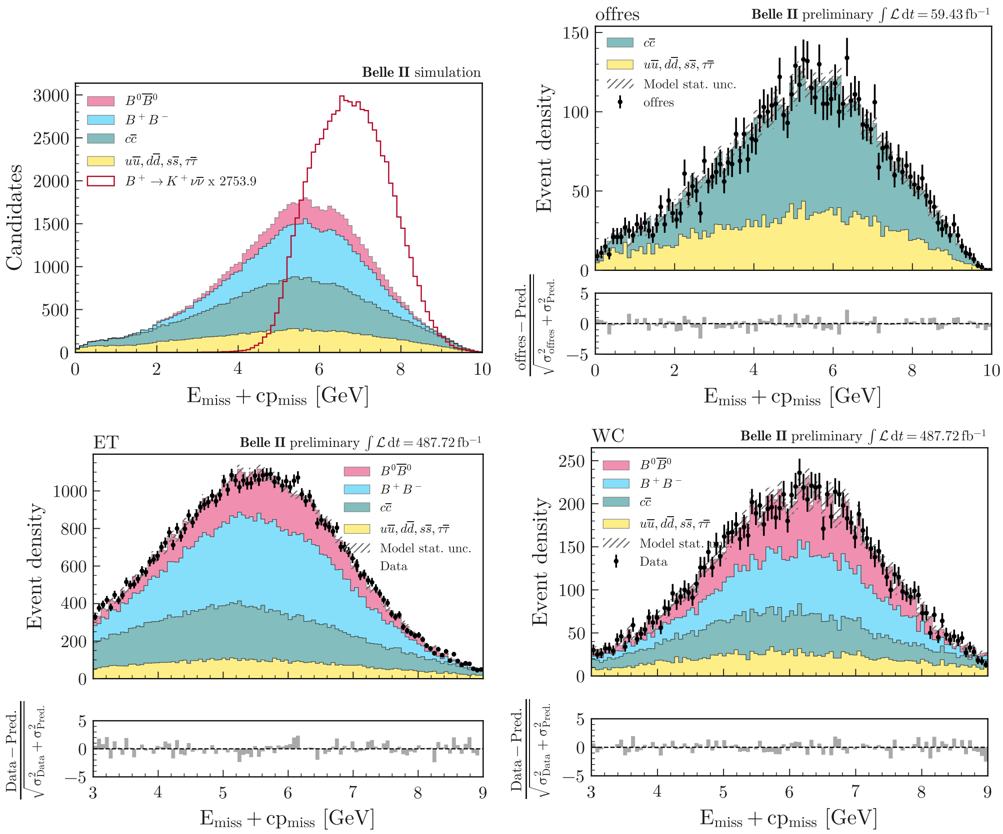
- $\cos\left(\theta^*_{miss}\right)$: polar angle of the missing energy system in the CMS frame
- $D^0 mass$: mass of the $K^+\pi^-$ system, where π^- is in the ROE; for D^0 meson discrimination
- $\cos\left(\theta_{B_{tag},miss}\right)$: polar angle between the B_{tag} and the missing energy system in the $\Upsilon(4S)$ frame
- $\cos\left(\theta_{B_{sig},thrust}\right)$: `cosTBTO`, as defined in Section 3.3
- p-value(D^+): p-value of the hypothesis test on the $K^+\pi^0$ system, to check if it is a D^+
- p-value(D^0): p-value of the hypothesis test on the $K^+\pi^-$ system, to check if it is a D^0
- $\phi^*(K^+, p_{miss})$: azimuthal angle between the signal K^+ and the momentum of the missing energy system

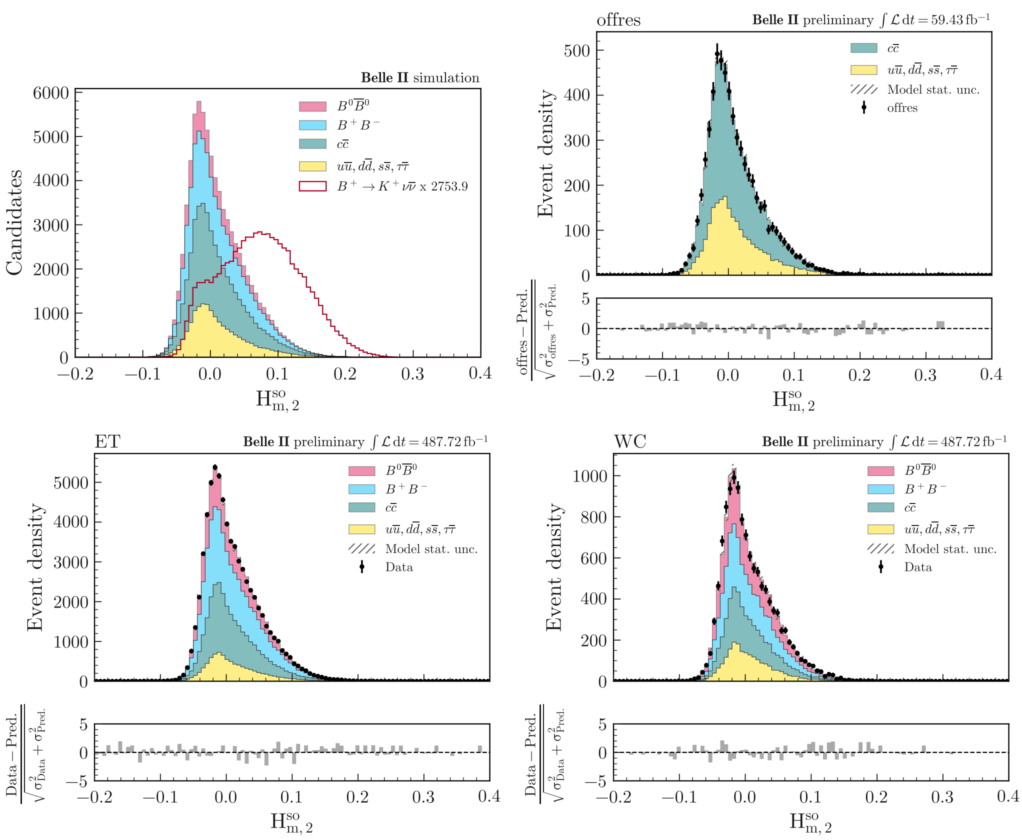
A.3 Validation of the data-MC agreement in the control samples

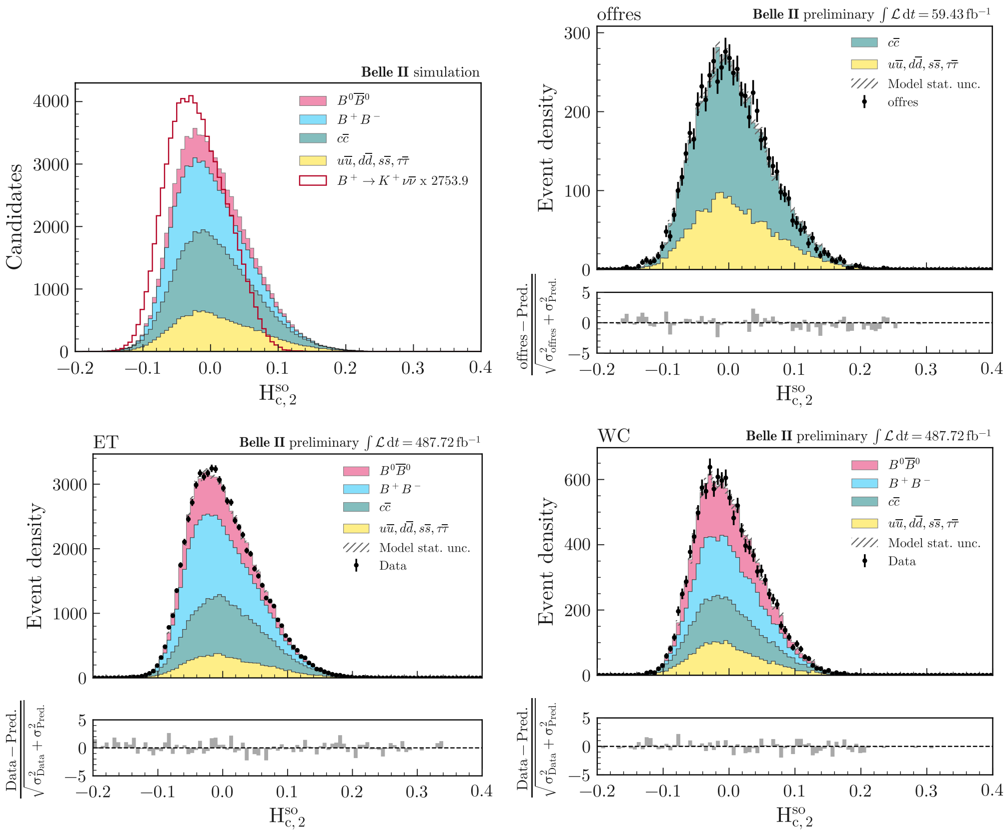
Distributions of the variables used in the BDTs. The upper left plot shows the signal versus background distributions in the default MC sample, to illustrate the discriminating power of the variable. The three other plots show the data-MC agreement in the three control samples used. On each page, the plots are represented in the following way:

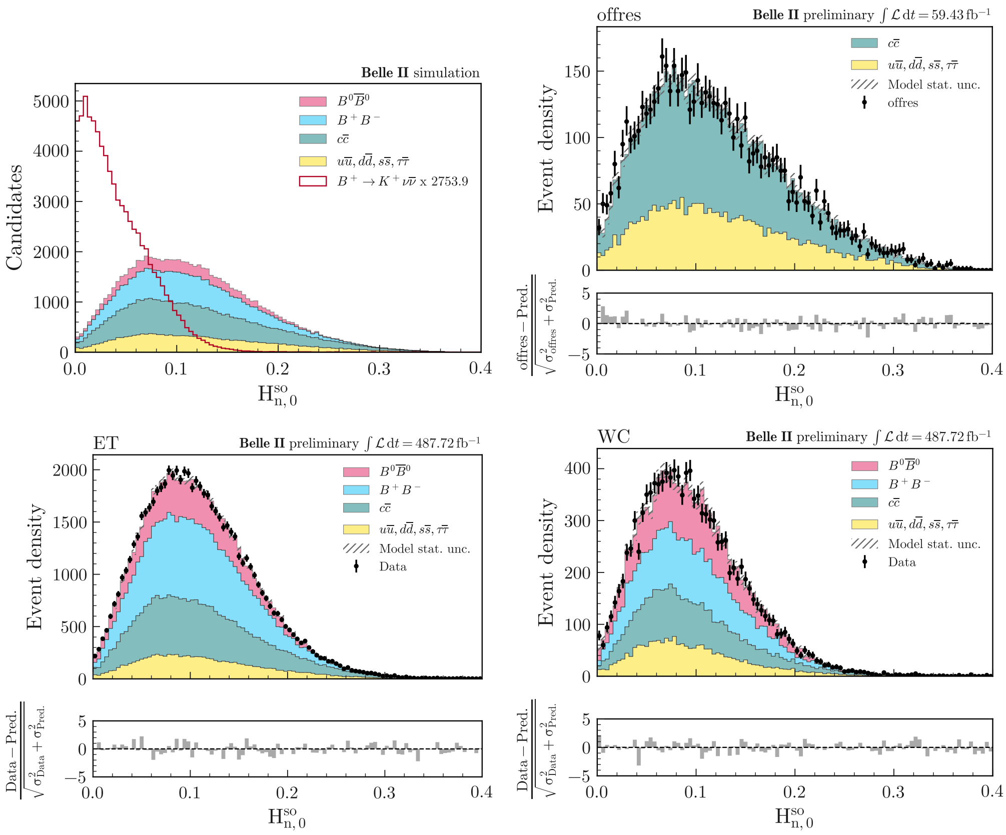
Signal vs background in default MC sample	Data-MC comparison in off-resonance
Data-MC comparison in ET sideband	Data-MC comparison in WC

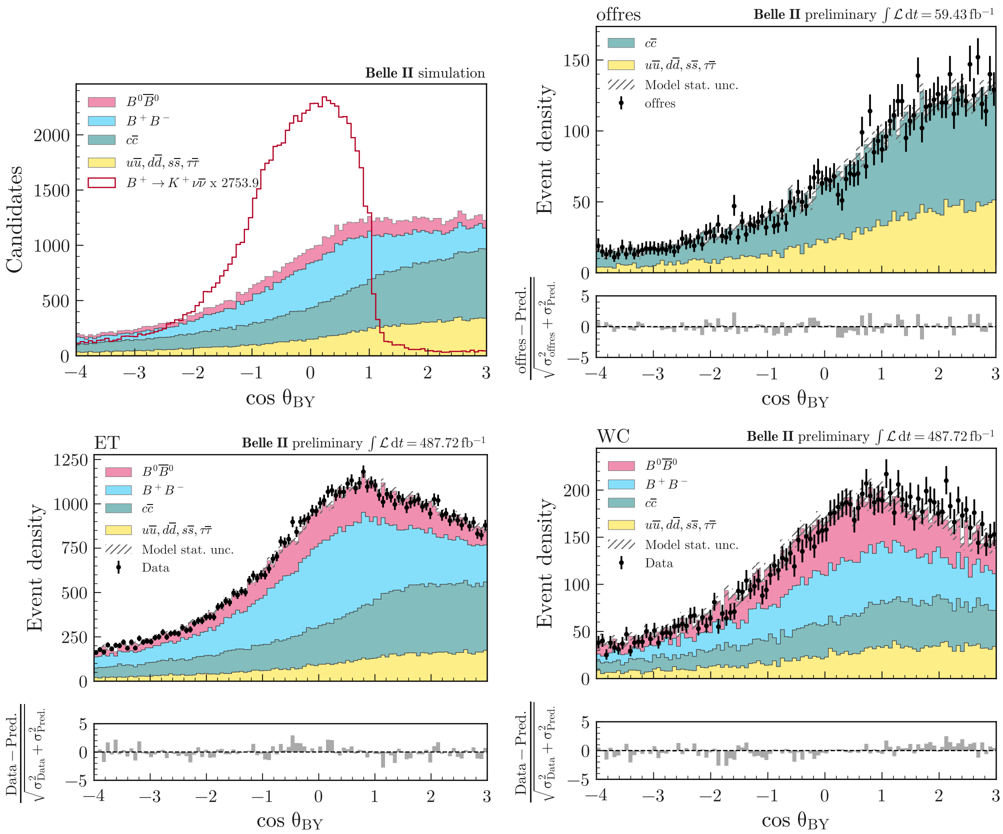


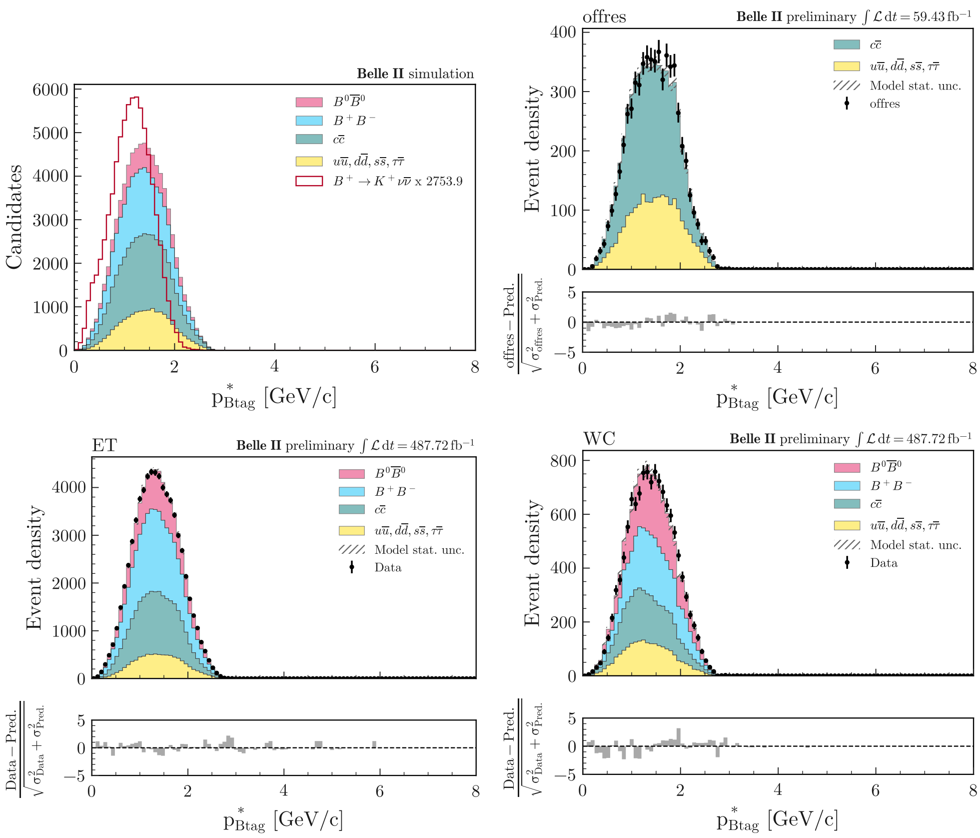

Belle II simulation
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+ \nu\overline{\nu}$ x 2753.9
Model stat. unc.
offres
Data
Candidates
Event density
$\frac{\mathrm{offres-Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}}+\sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{Data-Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}}+\sigma^2_{\mathrm{Pred.}}}}$
$\mathrm{E_{miss} + cp_{miss}}$ [GeV]

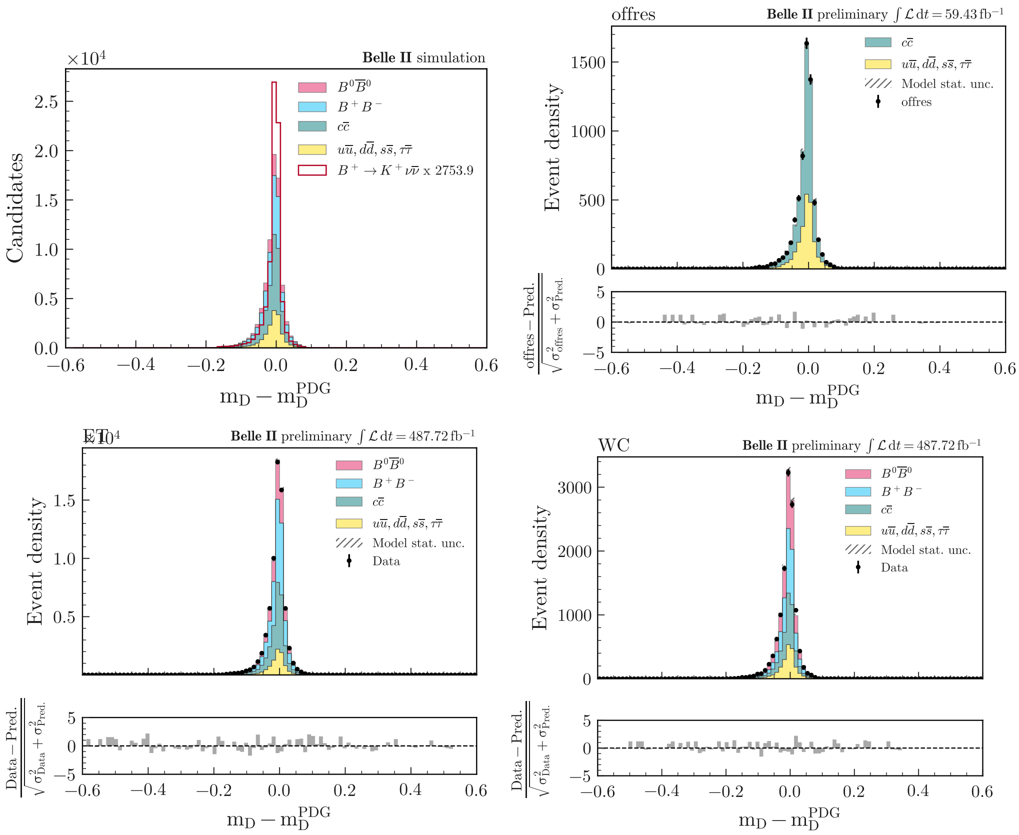

Belle II simulation
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+ \nu\overline{\nu}$ x 2753.9
Model stat. unc.
offres
Data
Candidates
Event density
$\frac{\mathrm{offres-Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}}+\sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{Data-Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}}+\sigma^2_{\mathrm{Pred.}}}}$
$\mathrm{H^{so}_{m,2}}$

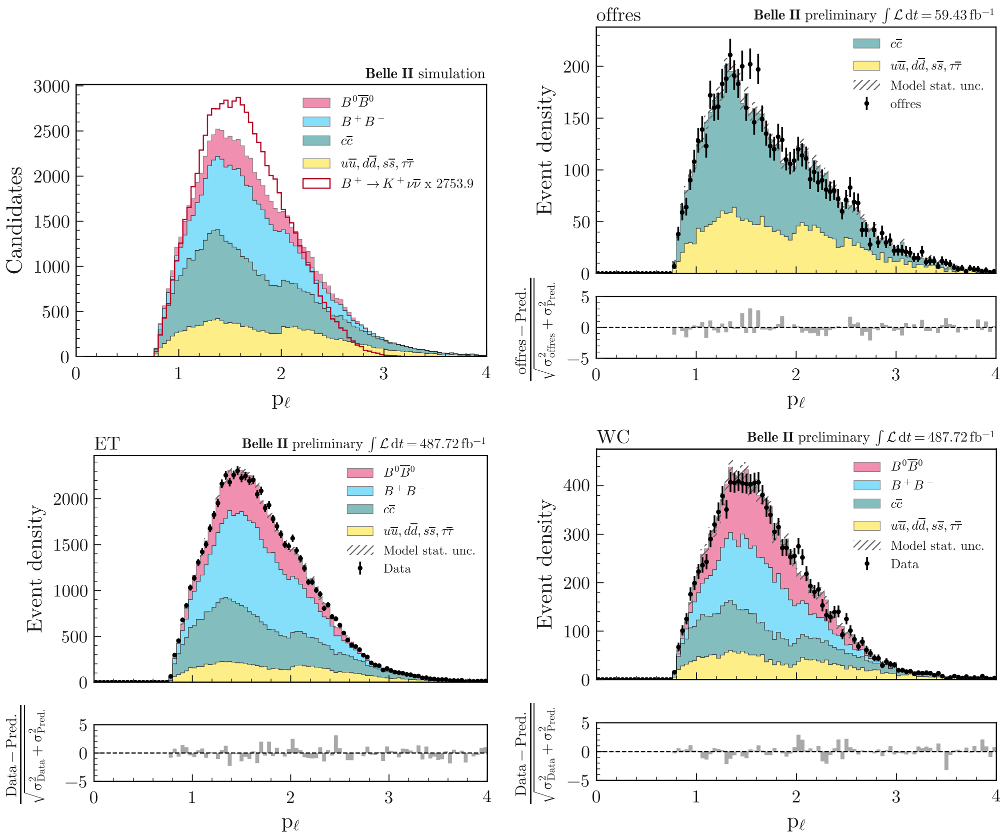

Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+\nu\overline{\nu}$ x 2753.9
$H^{so}_{c,2}$
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
ET
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\text{Data} - \text{Pred.}}{\sqrt{\sigma^2_{\text{Data}} + \sigma^2_{\text{Pred.}}}}$
$\frac{\text{offres} - \text{Pred.}}{\sqrt{\sigma^2_{\text{offres}} + \sigma^2_{\text{Pred.}}}}$

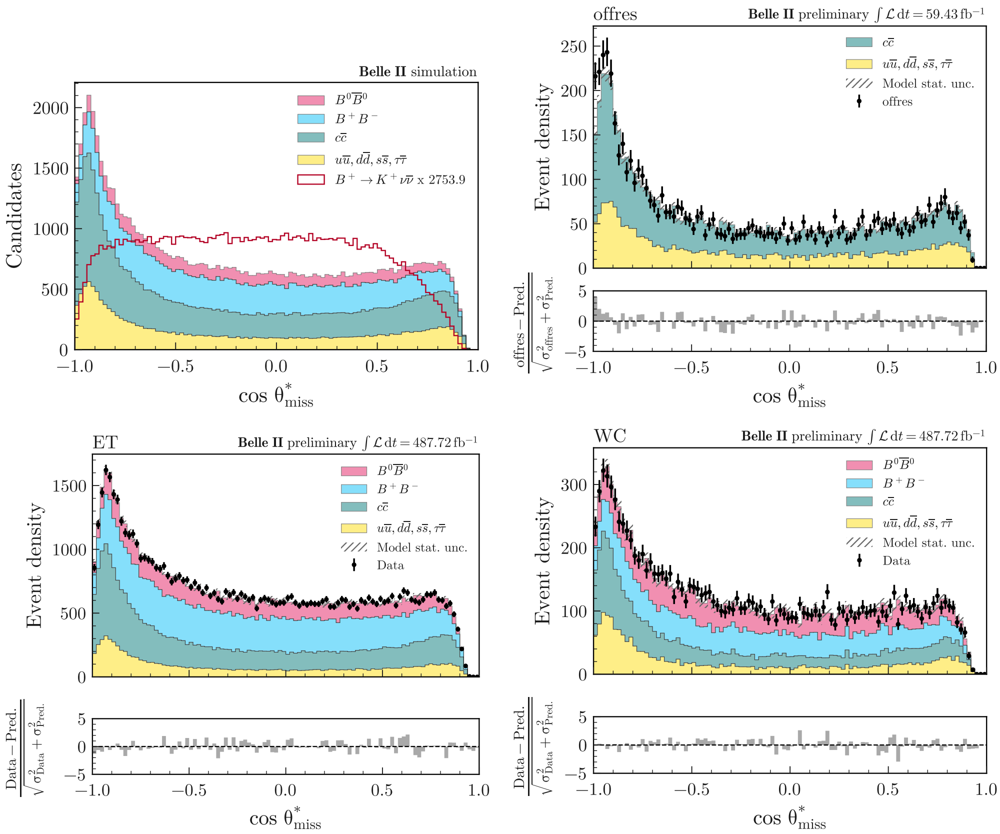

Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+\nu\overline{\nu}$ x 2753.9
$H^{so}_{n,0}$
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
ET
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\text{Data} - \text{Pred.}}{\sqrt{\sigma^2_{\text{Data}} + \sigma^2_{\text{Pred.}}}}$
$\frac{\text{offres} - \text{Pred.}}{\sqrt{\sigma^2_{\text{offres}} + \sigma^2_{\text{Pred.}}}}$

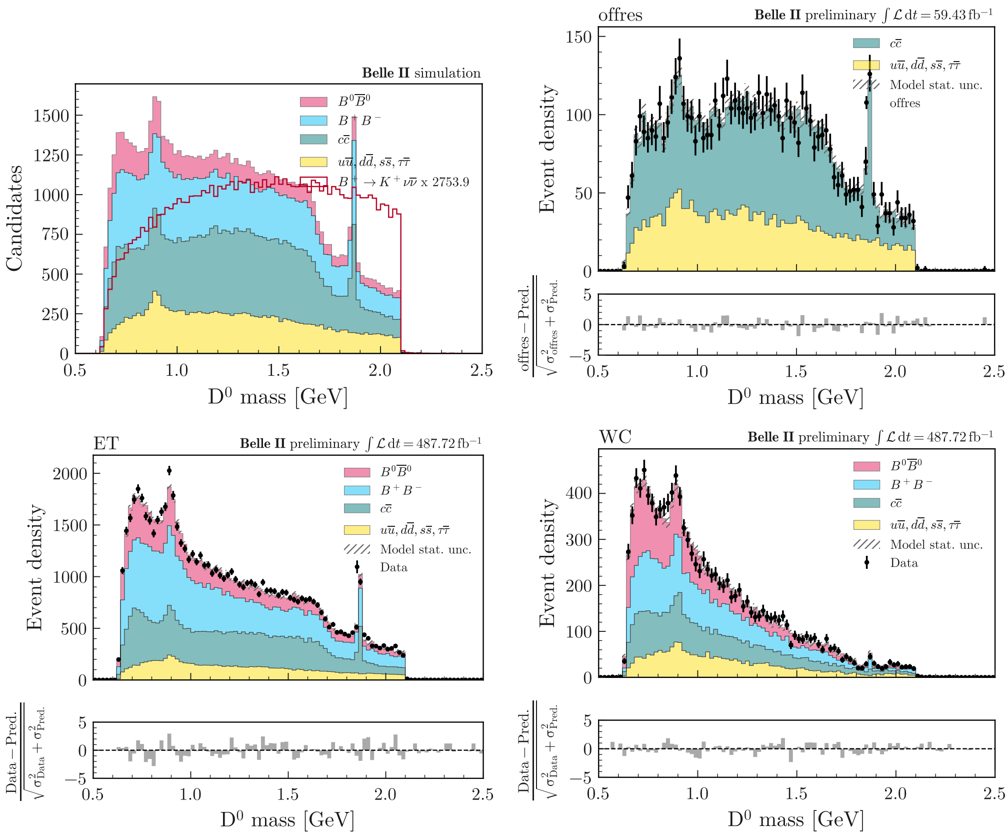

Belle II simulation
Candidates
cos θBY
offres
Belle II preliminary ∫ℒdt = 59.43 fb⁻¹
Event density
Model stat. unc.
ET
Belle II preliminary ∫ℒdt = 487.72 fb⁻¹
Data
WC
Belle II preliminary ∫ℒdt = 487.72 fb⁻¹

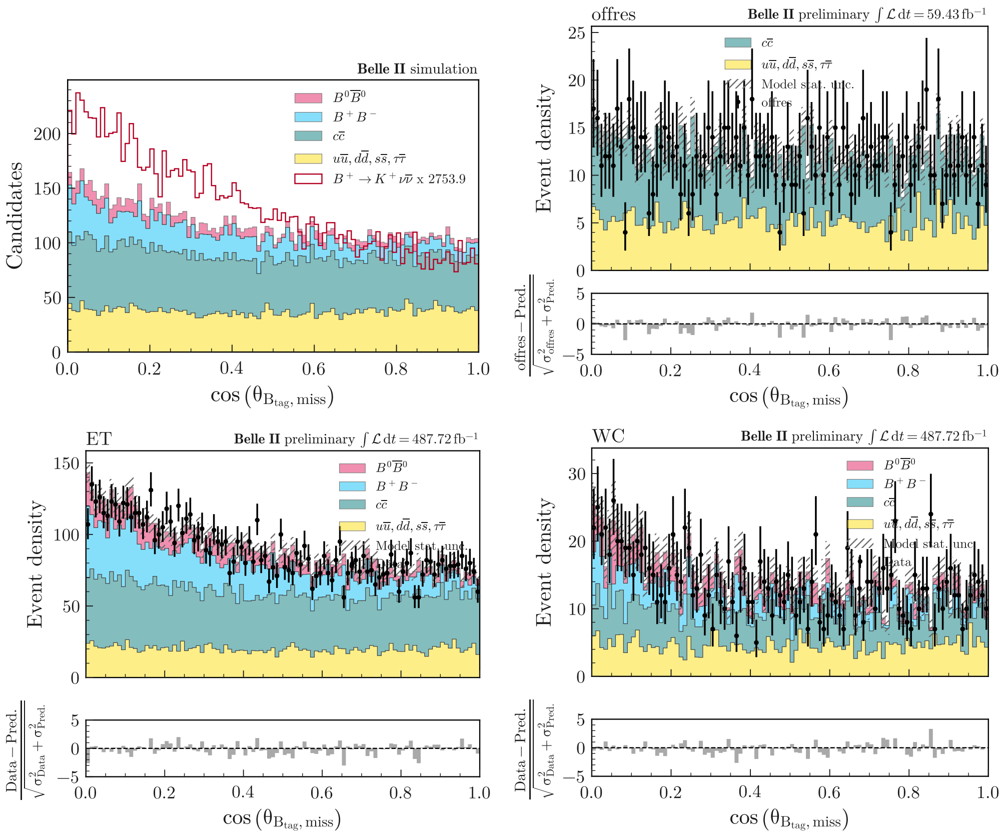

Belle II simulation
Candidates
p*Btag [GeV/c]
offres
Belle II preliminary ∫ℒdt = 59.43 fb⁻¹
Event density
Model stat. unc.
ET
Belle II preliminary ∫ℒdt = 487.72 fb⁻¹
Data
WC
Belle II preliminary ∫ℒdt = 487.72 fb⁻¹

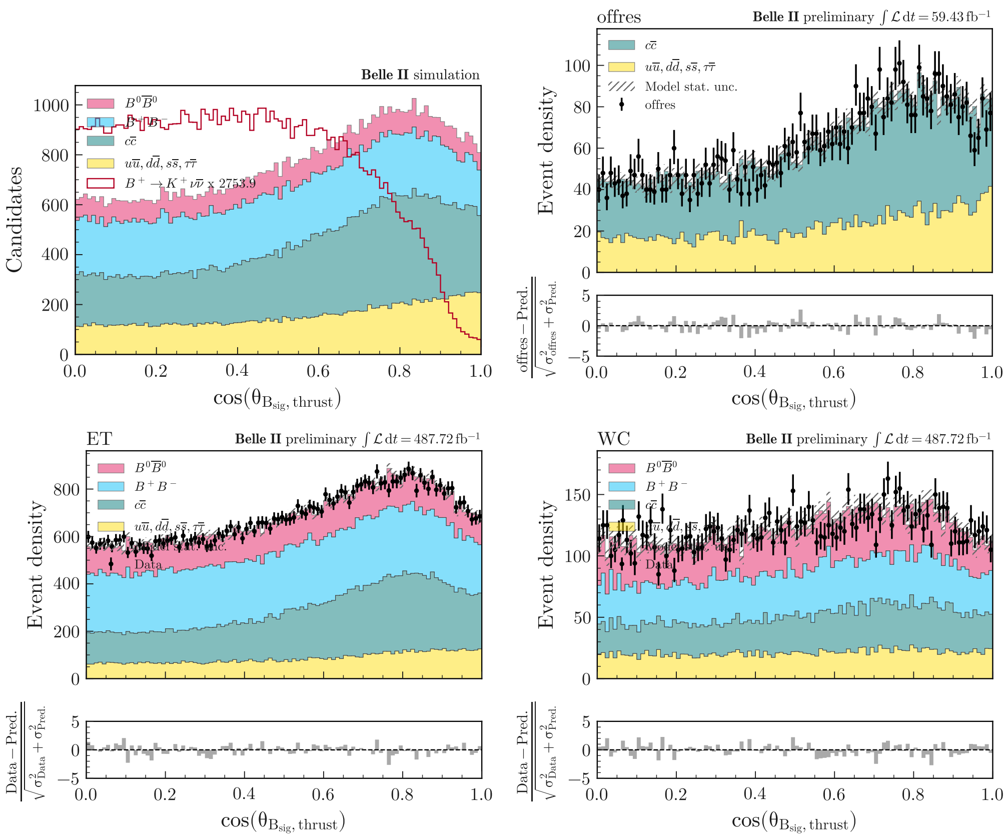

Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\bar{c}$
$u\bar{u}, d\bar{d}, s\bar{s}, \tau\bar{\tau}$
$B^+ \to K^+\nu\bar{\nu}$ x 2753.9
$\mathrm{m_D} - \mathrm{m_D^{PDG}}$
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
WC
$\frac{\mathrm{offres-Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}} + \sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{Data-Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}} + \sigma^2_{\mathrm{Pred.}}}}$

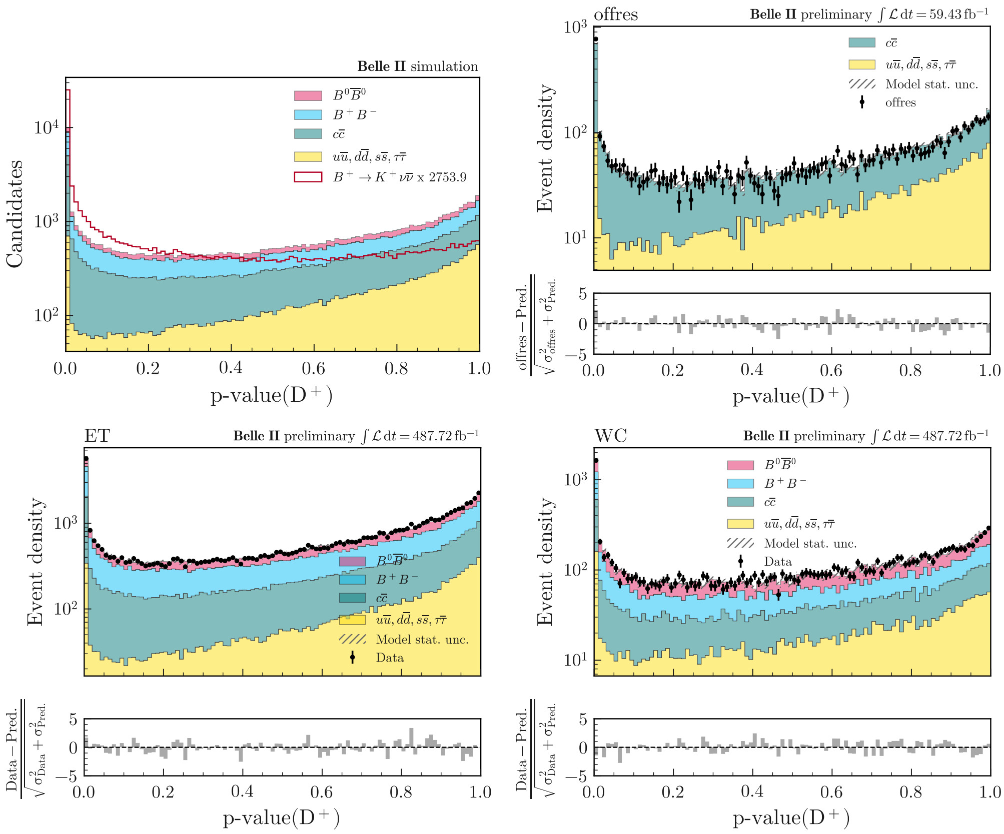

Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\bar{c}$
$u\bar{u}, d\bar{d}, s\bar{s}, \tau\bar{\tau}$
$B^+ \to K^+\nu\bar{\nu}$ x 2753.9
p_ℓ
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
WC
$\frac{\mathrm{offres-Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}} + \sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{Data-Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}} + \sigma^2_{\mathrm{Pred.}}}}$


Belle II simulation
Candidates
cos θ*miss
offres
Belle II preliminary ∫ℒ dt = 59.43 fb⁻¹
Event density
Model stat. unc.
ET
WC
Belle II preliminary ∫ℒ dt = 487.72 fb⁻¹
Data

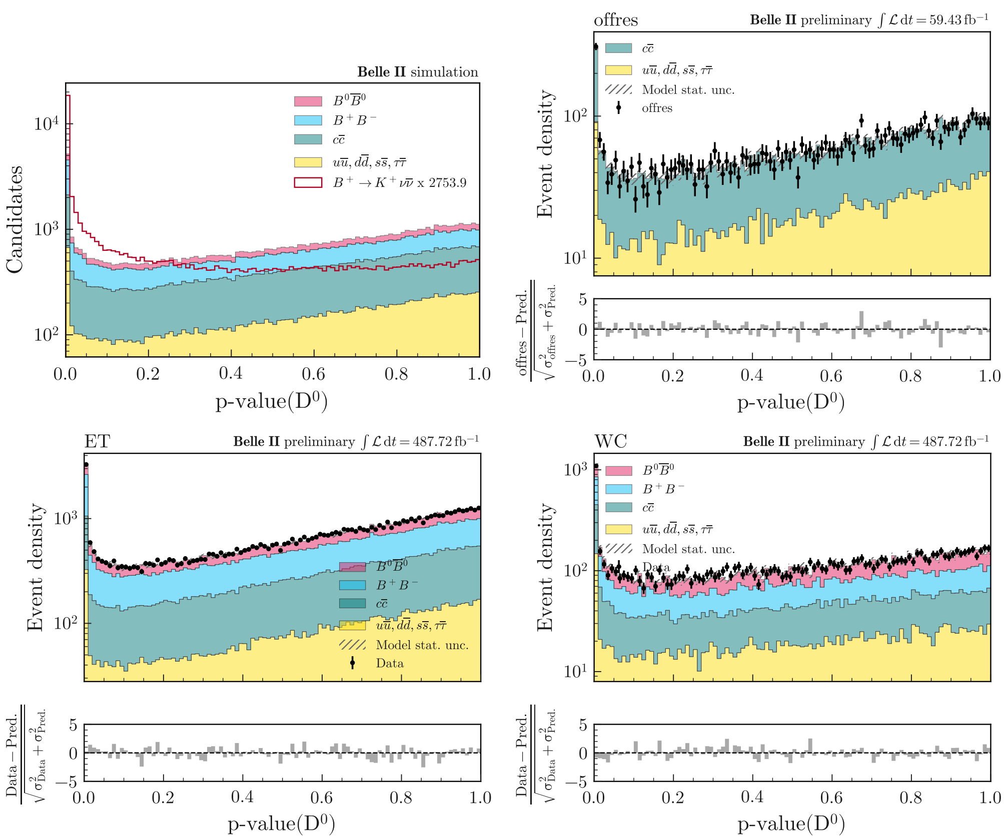

Belle II simulation
Candidates
D⁰ mass [GeV]
offres
Belle II preliminary ∫ℒ dt = 59.43 fb⁻¹
Event density
Model stat. unc.
ET
WC
Belle II preliminary ∫ℒ dt = 487.72 fb⁻¹
Data


Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+ \nu\overline{\nu}$ x 2753.9
$\cos(\theta_{B_{tag},\,miss})$
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Model stat. unc.
Event density
ET
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\mathrm{Data}-\mathrm{Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}}+\sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{offres}-\mathrm{Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}}+\sigma^2_{\mathrm{Pred.}}}}$

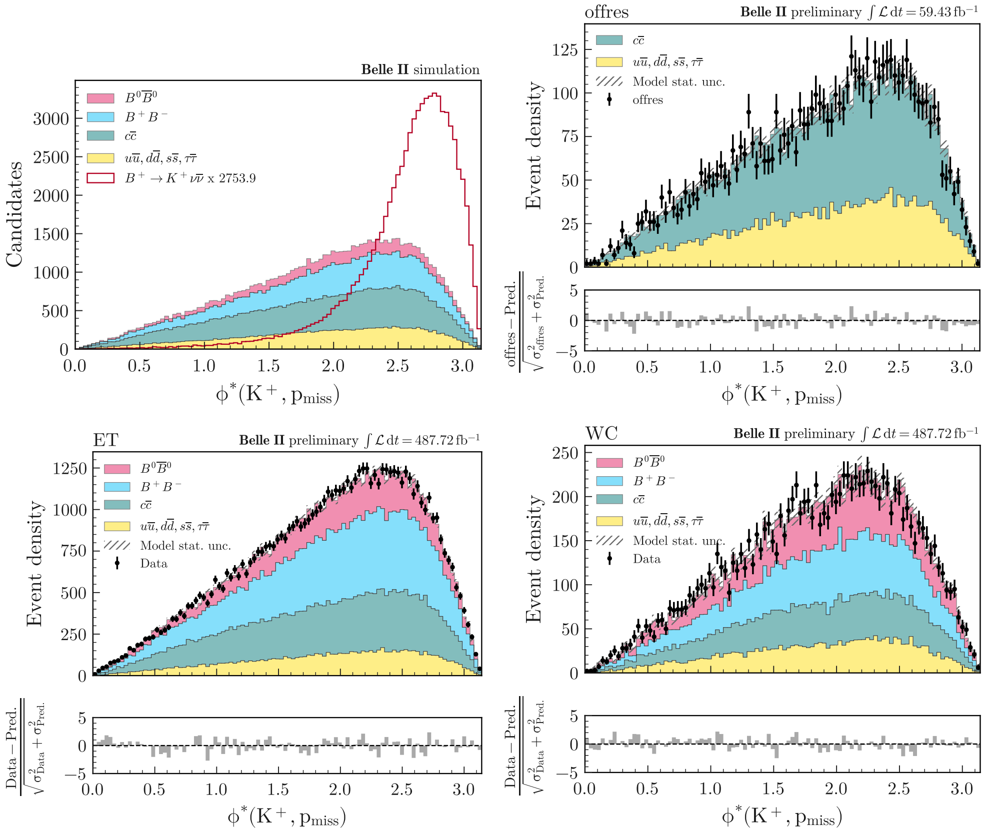

Belle II simulation
Candidates
$B^0\overline{B}^0$
B^+B^-
$c\overline{c}$
$u\overline{u}, d\overline{d}, s\overline{s}, \tau\overline{\tau}$
$B^+ \to K^+ \nu\overline{\nu}$ x 2753.9
$\cos(\theta_{B_{sig},\,thrust})$
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Model stat. unc.
Event density
ET
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\mathrm{Data}-\mathrm{Pred.}}{\sqrt{\sigma^2_{\mathrm{Data}}+\sigma^2_{\mathrm{Pred.}}}}$
$\frac{\mathrm{offres}-\mathrm{Pred.}}{\sqrt{\sigma^2_{\mathrm{offres}}+\sigma^2_{\mathrm{Pred.}}}}$



Belle II simulation
Candidates
p-value(D$^+$)
$B^0\overline{B}^0$
B^+B^-
$c\bar{c}$
$u\bar{u}, d\bar{d}, s\bar{s}, \tau\bar{\tau}$
$B^+ \to K^+ \nu\bar{\nu}$ x 2753.9
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
offres
$\frac{\text{offres} - \text{Pred.}}{\sqrt{\sigma^2_{\text{offres}} + \sigma^2_{\text{Pred.}}}}$
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\text{Data} - \text{Pred.}}{\sqrt{\sigma^2_{\text{Data}} + \sigma^2_{\text{Pred.}}}}$
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$



Belle II simulation
Candidates
p-value(D^0)
$B^0\overline{B}^0$
B^+B^-
$c\bar{c}$
$u\bar{u}, d\bar{d}, s\bar{s}, \tau\bar{\tau}$
$B^+ \to K^+ \nu\bar{\nu}$ x 2753.9
offres
Belle II preliminary $\int \mathcal{L}\,dt = 59.43\,\mathrm{fb}^{-1}$
Event density
Model stat. unc.
offres
$\frac{\text{offres} - \text{Pred.}}{\sqrt{\sigma^2_{\text{offres}} + \sigma^2_{\text{Pred.}}}}$
ET
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$
Data
$\frac{\text{Data} - \text{Pred.}}{\sqrt{\sigma^2_{\text{Data}} + \sigma^2_{\text{Pred.}}}}$
WC
Belle II preliminary $\int \mathcal{L}\,dt = 487.72\,\mathrm{fb}^{-1}$



A.4 *Alternative signal region selection tests*

In this section, we show the tests realised to define alternative signal region selections, *i.e.* a signal region at $BDT > 0.3$ and another at $BDT > 0.5$.

A.4.1 $BDT > 0.3$

By loosening the selection of the signal region, we accept more events which are essentially backgrounds. The purity went from 6% to 3%, and the efficiency slightly increased by 17%, as one can see in Table A.1. As we have more events, there is a minor improvement on the statistical uncertainty, but the systematic is worse, as we have more backgrounds compared to before. However, we see no impact on the total uncertainty shown in Table A.2 compared to the SR at 0.4.

Cut	Cumulated sgn eff [$\times 10^{-4}$]	$B^0\overline{B}^0$ yield in 0.49 ab^{-1}	$B^+\overline{B}^-$ yield in 0.49 ab^{-1}	ccbar yield in 0.49 ab^{-1}	uds yield in 0.49 ab^{-1}
reconstruction + preselection	119.43 ± 0.34	8829	23050	29156	15661
BDT>0.3	82.00 ± 0.29	66	404	120	79
BDT>0.9	11.58 ± 0.11	0	5	0	0

Table A.1: Table summarising the signal efficiency and background yields after the selection of the SR at 0.3.

Systematic Category	**Absolute contribution on branching fraction** ($\times 10^{-6}$)
StatMC	1.9
bkg norm	1.5
KL	1.3
Continuum	0.6
sgn eff + tracking	0.5
Backgrounds	0.4
KKLKL	0.3
nBB	0.2
Data/MC extra	0.1
FF	0.1
PID	0.1
D2KL	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**2.9**
Total statistical	**4.2**
Total	**5.1**

Table A.2: Impacts of the various types of nuisance parameters on the systematic uncertainty in the SR selected at 0.3.

A.4.2 $BDT > 0.5$

By tightening the selection of the signal region, we reject more background events. Compared to the 0.3 case, the efficiency decreased by the same amount, 17%, as one can compute from Table A.3, but the purity is exactly the same as the nominal case. Hence, the total uncertainty will worsen, as one can see in Table A.4.

Cut	Cumulated sgn eff [$\times 10^{-4}$]	$B^0\overline{B}^0$ yield in 0.49 ab^{-1}	$B^+\overline{B}^-$ yield in 0.49 ab^{-1}	ccbar yield in 0.49 ab^{-1}	uds yield in 0.49 ab^{-1}
reconstruction + preselection	119.61 ± 0.34	8829	23050	29156	15661
BDT>0.5	58.17 ± 0.24	15	138	29	23
BDT>0.9	11.41 ± 0.11	0	5	0	0

Table A.3: Table summarising the signal efficiency and background yields after the selection of the SR at 0.5.

Systematic Category	**Absolute contribution on branching fraction ($\times 10^{-6}$)**
StatMC	2.0
bkg norm	1.8
KL	1.5
Continuum	0.7
sgn eff + tracking	0.5
nBB	0.3
Data/MC extra	0.3
KKLKL	0.3
Backgrounds	0.2
D2KL	0.1
FF	0.1
PID	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**3.2**
Total statistical	**4.4**
Total	**5.5**

Table A.4: Impacts of the various types of nuisance parameters on the systematic uncertainty in the SR selected at 0.5.

List of acronyms

ARICH Aerogel Ring-Imaging Cherenkov. 24, 25, 26

ATS Analyse avec tagging semileptonique. 121

BDT Boosted Decision Tree. 31, 32, 33, 34, 35, 36, 41, 42, 55, 58, 62, 65, 66, 67, 68, 69, 72, 73, 74, 75, 76, 80, 83, 93, 97, 99, 122, 123, 124, 125, 126

BR Branching Ratio. 120, 122, 123, 126, 128, 129

CDC Central Drift Chamber. 24, 25, 28, 29, 60, 61, 74, 76

CKM Cabibbo-Kobayashi-Maskawa. 11, 12, 13, 16, 48

CL Confidence Level. 18, 19, 44, 88, 91

CP Charge-Parity. 21

DM Dark Matter. 9

ECL Electromagnetic Calorimeter. 24, 26, 28, 29, 60, 69, 97

EFT Effective Field Theory. 9, 12, 13, 18

ET "*Extra Tracks*". 61, 65, 68, 69, 71, 73, 75, 78, 82, 99

FCNC Flavour Changing Neutral Current. 2, 7, 13, 120, 1

FEI `Full Event Interpretation`. 2, 3, 31, 35, 36, 59, 60, 62, 67, 122, 123, 1

GIM Glashow-Iliopoulos-Maiani. 13

HAD Hadronic. 19, 91, 92

INC Inclusive. 19, 91, 92

IP Interaction Point. 23, 24

KLM K_L and Muon detector. 24, 29

KSFW Kakuno-Super-Fox-Wolfram. 38, 39, 97

LFU Lepton Flavour Universality. 17, 18

MC Monte Carlo. 27, 53, 54, 55, 56, 65, 122, 123, 124

NN Neural Network. 29

NP New Physics. 2, 7, 8, 9, 12, 13, 14, 16, 17, 18, 21, 94, 120, 1

OF Opposite Flavour. 53, 54, 56

PDF Probability Density Function. 25, 26

PDG Particle Data Group. 62, 63, 67

PID Particle Identification. 24, 29, 62, 67, 78, 79, 82, 95, 96, 97, 123

PXD Pixel Detector. 24

QFT Quantum Field Theory. 9

ROE Rest Of Event. 38, 39, 50, 60, 61, 69, 97, 98, 122, 123

SF Same Flavour. 53, 54

SL Semileptonic. 2, 8, 19, 90, 91, 93

SM Standard Model. 2, 3, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 21, 43, 48, 64, 89, 90, 93, 120, 121, 122, 123, 126, 129, 1

SR Signal Region. 72, 73, 74, 76, 77, 78, 79, 80, 81, 83, 92, 107, 108, 109

SVD Silicon Vertex Detector. 24, 29

TOP Time Of Propagation. 24, 25, 26

VXD Vertex Detector. 24, 28

WC "*Wrong Charge*". 61, 68, 78, 82, 99

Bibliography

[1] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 7).

[2] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562 (cit. on p. 7).

[3] Q. R. Ahmad et al. "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. arXiv: nucl-ex/0204008 (cit. on p. 7).

[4] Douglas Clowe et al. "A direct empirical proof of the existence of dark matter". In: *Astrophys. J. Lett.* 648 (2006), pp. L109–L113. DOI: 10.1086/508162. arXiv: astro-ph/0608407 (cit. on p. 7).

[5] Belle II Collaboration et al. "Evidence for $B\hat{}\{+\}\backslash toK\hat{}\{+\}\nu\bar{\nu}$ decays". In: *Physical Review D* 109.11 (June 2024). arXiv:2311.14647 [hep-ex], p. 112006. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.109.112006. URL: http://arxiv.org/abs/2311.14647 (cit. on pp. 7, 19, 20, 61, 73, 81, 90–92, 122, 127).

[6] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, Oct. 2013. ISBN: 978-1-107-03426-6, 978-1-139-52536-7. DOI: 10.1017/CBO9781139525367 (cit. on p. 10).

[7] F. Englert et al. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 10).

[8] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 10).

[9] G. S. Guralnik et al. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on p. 10).

[10] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on p. 10).

[11] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531 (cit. on p. 11).

[12] Makoto Kobayashi et al. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652 (cit. on p. 11).

[13] Andrei D Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (May 1991), p. 392. DOI: 10.1070/PU1991v034n05ABEH002497. URL: https://doi.org/10.1070/PU1991v034n05ABEH002497 (cit. on p. 11).

[14] S. L. Glashow et al. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1285 (cit. on pp. 13, 120).

[15] Andrzej J. Buras et al. *$B\to K^{(*)}ν\barν$ decays in the Standard Model and beyond*. arXiv:1409.4557 [hep-ph]. Dec. 2014. DOI: 10.48550/arXiv.1409.4557. URL: http://arxiv.org/abs/1409.4557 (cit. on pp. 14, 17, 64).

[16] Gerhard Buchalla et al. “Weak Decays Beyond Leading Logarithms”. In: *Reviews of Modern Physics* 68.4 (Oct. 1996). arXiv:hep-ph/9512380, pp. 1125–1244. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.68.1125. URL: http://arxiv.org/abs/hep-ph/9512380 (cit. on p. 14).

[17] Joachim Brod et al. “Two-loop electroweak corrections for the $K \to \pi\nu\bar{\nu}$ decays”. In: *Physical Review D* 83.3 (Feb. 2011). ISSN: 1550-2368. DOI: 10.1103/physrevd.83.034030. URL: http://dx.doi.org/10.1103/PhysRevD.83.034030 (cit. on p. 14).

[18] T. Inami et al. “Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes k(L) —> mu anti-mu, K+ —> pi+ Neutrino anti-neutrino and Ko <—> anti-Ko”. In: *Prog. Theor. Phys.* 65 (1981). [Erratum: Prog.Theor.Phys. 65, 1772 (1981)], p. 297. DOI: 10.1143/PTP.65.297 (cit. on p. 14).

[19] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001 (cit. on pp. 14, 23, 62, 121, 123).

[20] Damir Bečirević et al. “Revisiting $B\to K^{(\ast)} ν\barν$ decays in the Standard Model and beyond”. In: *The European Physical Journal C* 83.3 (Mar. 2023). arXiv:2301.06990 [hep-ph], p. 252. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11388-z. URL: http://arxiv.org/abs/2301.06990 (cit. on pp. 15, 16, 64, 123).

[21] Jon A. Bailey et al. “$B \to K l^+ l^-$ Decay Form Factors from Three-Flavor Lattice QCD”. In: *Phys. Rev. D* 93.2 (2016), p. 025026. DOI: 10.1103/PhysRevD.93.025026. arXiv: 1509.06235 [hep-lat] (cit. on p. 15).

[22] Y. Aoki et al. “FLAG Review 2021”. In: *Eur. Phys. J. C* 82.10 (2022), p. 869. DOI: 10.1140/epjc/s10052-022-10536-1. arXiv: 2111.09849 [hep-lat] (cit. on p. 15).

[23] W. G. Parrott et al. “B→K and D→K form factors from fully relativistic lattice QCD”. In: *Phys. Rev. D* 107.1 (2023), p. 014510. DOI: 10.1103/PhysRevD.107.014510. arXiv: 2207.12468 [hep-lat] (cit. on p. 15).

[24] Bhubanjyoti Bhattacharya et al. “Model independent determination of the axial mass parameter in quasielastic neutrino-nucleon scattering”. In: *Physical Review D* 84.7 (Oct. 2011). arXiv:1108.0423 [hep-ph], p. 073006. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.84.073006. URL: http://arxiv.org/abs/1108.0423 (cit. on p. 15).

[25] Aoife Bharucha et al. “$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules”. In: *JHEP* 08 (2016), p. 098. DOI: 10.1007/JHEP08(2016)098. arXiv: 1503.05534 [hep-ph] (cit. on p. 16).

[26] Lukas Allwicher et al. “Understanding the first measurement of B(B→K$\nu\nu^-$)”. In: *Phys. Lett. B* 848 (2024), p. 138411. DOI: 10.1016/j.physletb.2023.138411. arXiv: 2309.02246 [hep-ph] (cit. on p. 16).

[27] Lukas Allwicher et al. “Probing third-generation New Physics with K→$\pi\nu\nu^-$ and B→K(*)$\nu\nu^-$”. In: *Phys. Lett. B* 861 (2025), p. 139295. DOI: 10.1016/j.physletb.2025.139295. arXiv: 2410.21444 [hep-ph] (cit. on p. 18).

[28] David Marzocca et al. “Implications of $B \to K\nu\bar{\nu}$ under rank-one flavor violation hypothesis”. In: *Eur. Phys. J. C* 84.11 (2024), p. 1217. DOI: 10.1140/epjc/s10052-024-13534-7. arXiv: 2404.06533 [hep-ph] (cit. on p. 18).

[29] Rigo Bause et al. “Implications of an enhanced B→K$\nu\nu^-$ branching ratio”. In: *Phys. Rev. D* 109.1 (2024), p. 015006. DOI: 10.1103/PhysRevD.109.015006. arXiv: 2309.00075 [hep-ph] (cit. on p. 18).

[30] Rigo Bause et al. “Lepton universality and lepton flavor conservation tests with dineutrino modes”. In: *Eur. Phys. J. C* 82.2 (2022), p. 164. DOI: 10.1140/epjc/s10052-022-10113-6. arXiv: 2007.05001 [hep-ph] (cit. on p. 18).

[31] Rigo Bause et al. “Interplay of dineutrino modes with semileptonic rare B-decays”. In: *JHEP* 12 (2021), p. 061. DOI: 10.1007/JHEP12(2021)061. arXiv: 2109.01675 [hep-ph] (cit. on p. 18).

[32] Patrick D. Bolton et al. *Signatures of Light New Particles in $B\to K^{(*)} E_{\rm miss}$*. arXiv:2403.13887 [hep-ph]. Feb. 2025. DOI: 10.48550/arXiv.2403.13887. URL: http://arxiv.org/abs/2403.13887 (cit. on p. 18).

[33] Michael Schmidt. *BSM for $B \to K\nu\bar{\nu}$ & related anomalies*. 2025. URL: https://indico.belle2.org/event/14981/contributions/98595/ (cit. on p. 18).

[34] P. del Amo Sanchez et al. “Search for the Rare Decay $B \to K\nu\bar{\nu}$”. In: *Phys. Rev. D* 82 (2010), p. 112002. DOI: 10.1103/PhysRevD.82.112002. arXiv: 1009.1529 [hep-ex] (cit. on pp. 19, 91).

[35] J. P. Lees et al. “Search for $B \to K^{(*)}\nu\bar{\nu}$ and invisible quarkonium decays”. In: *Phys. Rev. D* 87.11 (2013), p. 112005. DOI: 10.1103/PhysRevD.87.112005. arXiv: 1303.7465 [hep-ex] (cit. on pp. 19, 91).

[36] O. Lutz et al. “Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample”. In: *Phys. Rev. D* 87.11 (2013), p. 111103. DOI: 10.1103/PhysRevD.87.111103. arXiv: 1303.3719 [hep-ex] (cit. on pp. 19, 91).

[37] J. Grygier et al. “Search for $\boldsymbol{B \to h\nu\bar{\nu}}$ decays with semileptonic tagging at Belle”. In: *Phys. Rev. D* 96.9 (2017). [Addendum: Phys.Rev.D 97, 099902 (2018)], p. 091101. DOI: 10.1103/PhysRevD.96.091101. arXiv: 1702.03224 [hep-ex] (cit. on pp. 19, 91).

[38] F. Abudinén et al. “Search for B+→K+$\nu\nu^-$ Decays Using an Inclusive Tagging Method at Belle II”. In: *Phys. Rev. Lett.* 127.18 (2021), p. 181802. DOI: 10.1103/PhysRevLett.127.181802. arXiv: 2104.12624 [hep-ex] (cit. on p. 19).

[39] Kazunori Akai et al. “SuperKEKB Collider”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (Nov. 2018). arXiv:1809.01958 [physics], pp. 188–199. ISSN: 01689002. DOI: 10.1016/j.nima.2018.08.017. URL: http://arxiv.org/abs/1809.01958 (cit. on pp. 23, 120).

[40] Yukiyoshi Ohnishi et al. "Accelerator design at SuperKEKB". In: *PTEP* 2013 (2013), 03A011. DOI: 10.1093/ptep/pts083 (cit. on p. 23).

[41] E. Kou et al. "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). arXiv:1808.10567 [hep-ex], p. 123C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. URL: http://arxiv.org/abs/1808.10567 (cit. on pp. 23, 26).

[42] *SuperKEKB/Belle II Complete 2024 Operations*. Jan. 2025. URL: https://www2.kek.jp/ipns/en/news/7015/ (cit. on p. 23).

[43] *Luminosity*. URL: https://www.belle2.org/research/luminosity/ (cit. on pp. 24, 121).

[44] T. Abe et al. "Belle II Technical Design Report". In: (Nov. 2010). arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 24, 120).

[45] *Belle II Detector*. URL: https://www.belle2.org/about_us/belle_ii_detector/ (cit. on p. 24).

[46] H. Atmacan et al. "The imaging Time-of-Propagation detector at Belle II". In: *Nucl. Instrum. Meth. A* 1080 (2025), p. 170627. DOI: 10.1016/j.nima.2025.170627. arXiv: 2504.19090 [hep-ex] (cit. on p. 25).

[47] Kazuki Kojima. "The operation and performance of the TOP detector at the Belle II experiment". In: *PoS* EPS-HEP2021 (2022), p. 803. DOI: 10.22323/1.398.0803 (cit. on p. 25).

[48] I. Adachi et al. "Aerogel RICH Counter at the Belle II Detector". In: *Nucl. Instrum. Meth. A* 1089 (2026), p. 171592. DOI: 10.1016/j.nima.2026.171592. arXiv: 2512.19146 [hep-ex] (cit. on p. 26).

[49] M. Yonenaga et al. "Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data". In: *PTEP* 2020.9 (2020), 093H01. DOI: 10.1093/ptep/ptaa090. arXiv: 2008.06251 [physics.ins-det] (cit. on p. 26).

[50] *The ECL: an overview*. Nov. 2023. URL: https://indico.belle2.org/event/9402/contributions/66287/attachments/25104/37129/ECL%20overview%201-Nov-2023.pdf (cit. on pp. 26, 28).

[51] T. Kuhr et al. "The Belle II Core Software". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. DOI: 10.1007/s41781-018-0017-9. arXiv: 1809.04299 [physics.comp-ph] (cit. on p. 26).

[52] H. U. Bengtsson. "The Lund Monte Carlo for High p_T Physics". In: *Comput. Phys. Commun.* 31 (1984), p. 323. DOI: 10.1016/0010-4655(84)90018-3 (cit. on p. 27).

[53] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan et al., pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4 (cit. on p. 27).

[54] R. Brun et al. "ROOT, an object oriented data analysis framework". In: *CERN School of Computing (CSC 2000)*. 2000, pp. 11–42. DOI: 10.5170/CERN-2000-013.11 (cit. on p. 27).

[55] S. Agostinelli et al. "GEANT4 - A Simulation Toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 27).

[56] S. Jadach et al. "The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions". In: *Comput. Phys. Commun.* 130 (2000), pp. 260–325. DOI: 10.1016/S0010-4655(00)00048-5. arXiv: hep-ph/9912214 (cit. on p. 27).

[57] N. Davidson et al. "Universal Interface of TAUOLA Technical and Physics Documentation". In: *Comput. Phys. Commun.* 183 (2012), pp. 821–843. DOI: 10.1016/j.cpc.2011.12.009. arXiv: 1002.0543 [hep-ph] (cit. on p. 27).

[58] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on p. 27).

[59] C. M. Carloni Calame et al. "The BABAYAGA event generator". In: *Nucl. Phys. B Proc. Suppl.* 131 (2004). Ed. by Marco Incagli et al., pp. 48–55. DOI: 10.1016/j.nuclphysbps.2004.02.008. arXiv: hep-ph/0312014 (cit. on p. 27).

[60] Frits A. Berends et al. "Monte Carlo Simulation of Two Photon Processes. 2. Complete Lowest Order Calculations for Four Lepton Production Processes in electron Positron Collisions". In: *Comput. Phys. Commun.* 40 (1986), pp. 285–307. DOI: 10.1016/0010-4655(86)90115-3 (cit. on p. 27).

[61] Giovanni Gaudino. *Run Dependent Monte Carlo at Belle II*. 2025. URL: https://indico.in2p3.fr/event/33627/contributions/155250/ (cit. on p. 27).

[62] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4 (cit. on p. 28).

[63] R. Frühwirth et al. "Track fitting with ambiguities and noise: A study of elastic tracking and non-linear filters". In: *Computer Physics Communications* 120.2 (1999), pp. 197–214. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(99)00231-3. URL: https://www.sciencedirect.com/science/article/pii/S0010465599002313 (cit. on p. 28).

[64] Daniel Greenwald et al. *Particle identification at Belle II using Neural Networks*. 2023. URL: https://docs.belle2.org/pub_data/documents/2297/ (cit. on p. 29).

[65] S. Longo et al. "Hadronic vs Electromagnetic Pulse Shape Discrimination in CsI(Tl) for High Energy Physics Experiments". In: *JINST* 13.03 (2018), P03018. DOI: 10.1088/1748-0221/13/03/P03018. arXiv: 1801.07774 [physics.ins-det] (cit. on pp. 29, 30).

[66] Savino Longo. *A new particle identification method with the Belle II calorimeter using pulse shape discrimination in CsI(Tl)*. 2018. URL: https://indi.to/8pqwC (cit. on p. 30).

[67] Trevor Hastie et al. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001 (cit. on p. 32).

[68] Tianqi Chen et al. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145/2939672.2939785 (cit. on p. 33).

[69] Takuya Akiba et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019 (cit. on pp. 34, 69, 124).

[70] Lloyd S. Shapley. *Notes on the N-Person Game II: The Value of an N-Person Game*. Santa Monica, CA: RAND Corporation, 1951. DOI: 10.7249/RM0670 (cit. on p. 34).

[71] Scott M Lundberg et al. "A Unified Approach to Interpreting Model Predictions". In: Curran Associates, Inc., 2017. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf (cit. on p. 34).

[72] T. Keck et al. "The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 6. DOI: 10.1007/s41781-019-0021-8. arXiv: 1807.08680 [hep-ex] (cit. on pp. 35, 36, 122).

[73] Geoffrey C. Fox et al. "Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41 (1978), p. 1581. DOI: 10.1103/PhysRevLett.41.1581 (cit. on p. 38).

[74] A. J. Bevan et al. "The Physics of the B Factories". In: *Eur. Phys. J. C* 74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex] (cit. on pp. 38–40).

[75] Cyrille Praz. "Search for B→K decays with a machine learning method at the Belle II experiment". Dissertation, University of Hamburg, 2022. Dissertation. Hamburg: University of Hamburg, 2022, pp. 1–165. DOI: 10.3204/PUBDB-2022-04919. URL: https://bib-pubdb1.desy.de/record/482717 (cit. on p. 39).

[76] Zhidong Bai et al. "Sample Covariance Matrices and the Marčenko-Pastur Law". In: *Spectral Analysis of Large Dimensional Random Matrices*. New York, NY: Springer New York, 2010, pp. 39–58. ISBN: 978-1-4419-0661-8. DOI: 10.1007/978-1-4419-0661-8_3. URL: https://doi.org/10.1007/978-1-4419-0661-8_3 (cit. on p. 40).

[77] Hans Dembinski et al. "A new maximum-likelihood method for template fits". In: *The European Physical Journal C* 82.11 (Nov. 2022), p. 1043. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-11019-z. URL: https://doi.org/10.1140/epjc/s10052-022-11019-z (cit. on p. 41).

[78] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. DOI: 10.17181/CERN-OPEN-2012-016. URL: https://cds.cern.ch/record/1456844 (cit. on pp. 41, 125).

[79] Lukas Heinrich et al. *pyhf: v0.7.6*. Version 0.7.6. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6. DOI: 10.5281/zenodo.1169739. URL: https://doi.org/10.5281/zenodo.1169739 (cit. on pp. 41, 126).

[80] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on p. 43).

[81] Jerzy Neyman et al. "IX. On the problem of the most efficient tests of statistical hypotheses". In: *Philosophical Transactions of the Royal Society of London, Series A: Containing Papers of a Mathematical or Physical Character* 231.694-706 (Feb. 1933), pp. 289–337. ISSN: 0264-3952. DOI: 10.1098/rsta.1933.0009. eprint: https://royalsocietypublishing.org/rsta/article-pdf/231/694-706/289/240090/rsta.1933.0009.pdf. URL: https://doi.org/10.1098/rsta.1933.0009 (cit. on p. 43).

[82] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://doi.org/10.1214/aoms/1177732360 (cit. on pp. 44, 127).

[83] A L Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/0954-3899/28/10/313 (cit. on pp. 44, 127).

[84] Andrés Pinto et al. "Uncertainty components in profile likelihood fits". In: *Eur. Phys. J. C* 84.6 (2024), p. 593. DOI: 10.1140/epjc/s10052-024-12877-5. arXiv: 2307.04007 [physics.data-an] (cit. on pp. 45, 127).

[85] Fernando Abudinén. "Development of a B^0 flavor tagger and performance study of a novel time-dependent CP analysis of the decay $B^0 \to \pi^0\pi^0$ at Belle II". Oct. 2018. URL: http://nbn-resolving.de/urn:nbn:de:bvb:19-230032 (cit. on pp. 49–51).

[86] I. Adachi et al. "New graph-neural-network flavor tagger for Belle II and measurement of sin $2\phi1$ in B0→J/ψKS0 decays". In: *Phys. Rev. D* 110.1 (2024), p. 012001. DOI: 10.1103/PhysRevD.110.012001. arXiv: 2402.17260 [hep-ex] (cit. on pp. 50, 52, 56).

[87] S. Navas et al. "Review of Particle Physics". In: *Phys. Rev. D* 110 (3 Aug. 2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.110.030001 (cit. on p. 54).

[88] Alexander Heidelbach et al. *belle2/b2luigi: v1.2.8*. Version v1.2.8. Mar. 2026. DOI: 10.5281/zenodo.18893185. URL: https://doi.org/10.5281/zenodo.18893185 (cit. on p. 56).

[89] Ashish Vaswani et al. "Attention is All you Need". In: *Neural Information Processing Systems*. 2017. URL: https://api.semanticscholar.org/CorpusID:13756489 (cit. on p. 56).

[90] Bernard Aubert et al. "Evidence for the B0 —-> p anti-p K*0 and B+ —> eta(c) K*+ decays and Study of the Decay Dynamics of B Meson Decays into p anti-p h final states". In: *Phys. Rev. D* 76 (2007), p. 092004. DOI: 10.1103/PhysRevD.76.092004. arXiv: 0707.1648 [hep-ex] (cit. on pp. 63, 64, 81).

[91] J. P. Lees et al. "Study of CP violation in Dalitz-plot analyses of B0 —> K+K-K0(S), B+ —> K+K-K+, and B+ —> K0(S)K0(S)K+". In: *Phys. Rev. D* 85 (2012), p. 112010. DOI: 10.1103/PhysRevD.85.112010. arXiv: 1201.5897 [hep-ex] (cit. on pp. 64, 81).

[92] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 72).

[93] Alexander Held et al. *scikit-hep/cabinetry: v0.6.0*. Version v0.6.0. Sept. 2023. DOI: 10.5281/zenodo.8360833. URL: https://doi.org/10.5281/zenodo.8360833 (cit. on p. 83).

[94] Hans Dembinski et al. *scikit-hep/iminuit*. Version v2.32.0. Nov. 2025. DOI: 10.5281/zenodo.17565861. URL: https://doi.org/10.5281/zenodo.17565861 (cit. on p. 83).

[95] Jorge de Blas et al. "Physics Briefing Book: Input for the 2026 update of the European Strategy for Particle Physics". In: (Nov. 2025). DOI: 10.23731/CYRM-2025-008. arXiv: 2511.03883 [hep-ex] (cit. on p. 94).

Résumé en français

Contexte scientifique

Ma thèse s'inscrit dans le cadre de la recherche sur la physique au-delà du SM via des expériences menées sur des collisionneurs. Le SM est actuellement la théorie la plus aboutie pour décrire la physique des particules. Au cours de son histoire, les prédictions du SM ont été comparées aux résultats expérimentaux et se sont révélées exactes. Bien que cette théorie soit très fructueuse, certains points restent à élucider, tels que la nature de la matière noire et de l'énergie noire, ou l'origine de l'asymétrie matière/antimatière. Pour répondre à ces questions, les physiciens des collisionneurs disposent de deux méthodes : soit utiliser des collisionneurs plus puissants, soit augmenter la précision des mesures afin de comparer les prédictions avec les résultats expérimentaux. Celles-ci sont décrites comme les frontières de l'énergie et de l'intensité.

Une puissante sonde de la frontière d'intensité est la désintégration du méson B en un méson K et deux neutrinos, $B \to K^{(*)}\nu\bar{\nu}$. En effet, ces désintégrations correspondent à des transitions $b \to s\nu\bar{\nu}$, impliquant ainsi un FCNC. Nous savons que, dans le SM, le FCNC ne peut se produire qu'au niveau des boucles, avec les diagrammes de Feynman les plus simples présentés dans Figure A.5, et que les boucles sont fortement supprimées par le mécanisme de Glashow-Iliopoulous-Maiani [14]. Par conséquent, les désintégration $B \to K^{(*)}\nu\bar{\nu}$ sont rares, avec un Branching Ratio (BR) de l'ordre de 10^{-5}. La combinaison de cette rareté et les deux neutrinos dans l'état final en font une désintégration difficile à détecter. Mais nous disposons d'une prédiction précise du BR dans le SM, car il s'agit d'un canal théoriquement propre. De plus, certains modèles théoriques prédisent la présence de NP dans les boucles quantiques qui modifierait le BR, ou l'apparition de particules de matière noire à la place des deux neutrinos. En combinant tous ces éléments, nous pouvons voir que les désintégrations $B \to K^{(*)}\nu\bar{\nu}$ constituent une puissante sonde de la NP.

L'expérience Belle II [44], menée au collisionneur SuperKEKB [39] à Tsukuba, au Japon, recherche les NP en fournissant des données avec la luminosité instantanée la plus élevée jamais atteinte

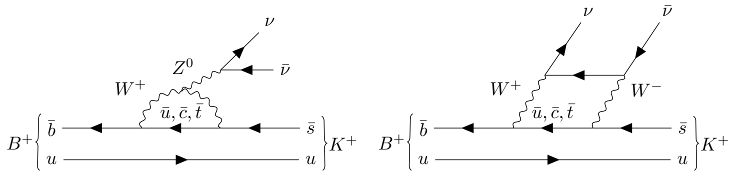


Figure A.5: Principaux diagrammes de Feynman pour la désintégration $B \to K^{(*)}\nu\bar{\nu}$.

[43]. C'est là que des paires e^+e^- entrent en collision avec une énergie dans le centre de masse permettant de produire le méson $\Upsilon(4S)$, un méson qui se désintègre presque toujours en une paire $B\bar{B}$ [19], ce qui en fait un lieu idéal pour étudier la physique des B. Les caractéristiques de Belle II, telles que son herméticité et l'environnement propre qu'elle offre, en font la seule expérience actuelle permettant d'étudier les désintégrations de B avec des particules invisibles , en particulier $B \to K^{(*)}\nu\bar{\nu}$.

Le fait que le méson B qui nous intéresse, appelé B_{sig}, soit produit exclusivement avec un méson B partenaire, appelé B_{tag}, constitue une caractéristique essentielle pour l'analyse des désintégrations $B \to K^{(*)}\nu\bar{\nu}$. En effet, les deux neutrinos de l'état final ne peuvent pas être détectés ; nous devons donc trouver un moyen indirect de reconstituer la désintégration B_{sig}. La stratégie consiste à déduire des informations sur le B_{sig} à partir de la reconstitution du B_{tag}. Cette technique est connue sous le nom de *tagging*. On peut définir deux types de stratégies de tagging : le tagging *exclusif* et le tagging *inclusif*. Le tagging exclusif consiste à choisir un ensemble de désintégrations pour reconstruire le B_{tag}, avec soit des modes *hadroniques*, soit des modes *semi-leptoniques*. Pour le tagging inclusif, on prend en compte de manière inclusive toutes les désintégrations possibles pour le B_{tag}.

Ma thèse porte sur l'*Analyse avec tagging semileptonique (ATS)*. Le choix du tag semi-leptonique s'explique par les analyses précédentes menées à Belle II à l'aide d'analyses hadroniques et inclusives (HTA et ITA), ainsi que par les analyses de Belle et BaBar réalisées à l'aide de la ATS. Mon objectif est de mener une recherche indépendante sur $B \to K^{(*)}\nu\bar{\nu}$. Elle servira de vérification solide des résultats obtenus par les analyses précédentes de Belle II, car comme on peut le voir sur la figure A.6, les résultats antérieurs de ATS sont plus compatibles avec l'hypothèse du SM seul.

Le terme $B \to K^{(*)}\nu\bar{\nu}$ désigne l'ensemble des canaux $B^{\pm,0} \to K^{(*)\pm,0}\nu\bar{\nu}$, mais le présent document se concentre sur la désintégration $B^{\pm} \to K^{\pm}\nu\bar{\nu}$.

Dans le document qui suit, nous présenterons l'analyse *aveugle*, ce qui signifie que nous n'examinerons pas les données de la région de signal tant que le flux de travail d'analyse n'aura pas été

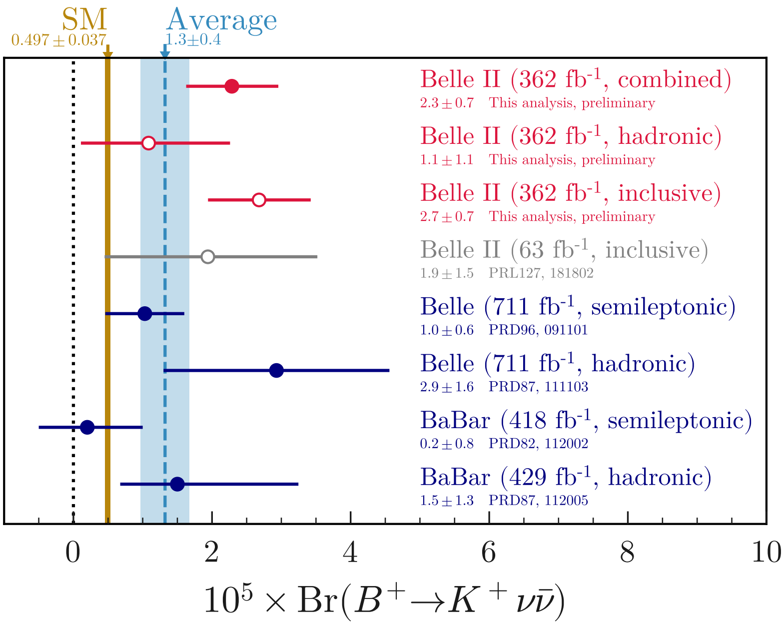


Figure A.6: Valeurs des BR mesurés par Belle II, mesurés lors d'expériences précédentes et valeurs prédites par le SM [5].

entièrement défini et testé sur la simulation MC. Pour vérifier la concordance entre les données et la simulation, nous utilisons des échantillons de contrôle qui seront spécifiés le moment venu.

Reconstruction

Pour commencer la mesure du BR de la désintégration $B^+ \to K^+\nu\bar{\nu}$, nous reconstruisons d'abord le B_{tag} en tenant compte des désintégrations $D^{(*)}\ell\nu$, avec $\ell \in \{e, \mu\}$. Ensuite, nous reconstruisons le kaon à partir de la désintégration du côté du signal. La reconstruction du B_{tag} est effectuée à l'aide de l'algorithme FEI, un algorithme basé sur des BDT [72]. Nous combinons le kaon signal et le $D\ell$ provenant du B_{tag} pour former notre événement et considérons tout le reste comme ROE. Comme il ne doit y avoir aucune autre particule liée au signal dans le ROE, nous exigeons qu'il n'y ait pas de traces supplémentaires dans l'événement, ni de π^0, K_S^0 et Λ supplémentaires.

De plus, nous appliquons une présélection souple afin de réduire d'ores et déjà le nombre d' événements de fond.

Échantillon de contrôle

Afin de corriger les problèmes de modélisation, ainsi que pour vérifier la pertinence de celle-ci, nous faisons usage d'échantillons de contrôle. Nous en avons quatre:

- Pour le continuum, événements $q\bar{q}$, nous utilisons un échantillon *hors résonance*, *c'est-à-dire* ne se situant pas à la résonance $\Upsilon(4S)$, de sorte que nous n'avons que des événements du continuum;

- Pour les événements $B\bar{B}$, nous avons deux échantillons: un échantillon où la charge du B_{sig} est erronée et un échantillon pour lequel nous acceptons des traces en plus dans le reste de l'événement;

- Pour l'efficacité sur le signal, nous utilisons la technique de "l'intégration", *i.e.* nous choisissons un canal dont on connaît bien la cinématique et qui comporte un kaon (*e.g.* $B^+ \to K^+ J/\psi$). Ensuite, nous reconstruisons un échantillon de données et une simulation de ce canal, puis nous supprimons la partie « signal » des événements, ne conservant que le ROE. Nous intégrons ensuite notre partie "signal" simulée dans le ROE et pouvons alors reconstruire l'événement comme s'il s'agissait d'un événement de signal.

Corrections des simulations

Pour cette analyse, nous nous appuyons sur des simulations MC afin d'entraîner les BDT et d'obtenir le nombre attendu d'événements dans la région de signal pour l'ajustement du modèle. Comme les simulations MC ne sont pas parfaites, nous devons corriger les écarts entre les données et les simulations MC pour les variables utilisées dans le processus de sélection. Parmi ces divergences possibles, nous appliquons une correction pour les rendements PID, la normalisation du continuum $q\bar{q}$ et $B\bar{B}$, ainsi que pour le BR des modes de désintégration dans les bruits de fond. D'autres corrections sont nécessaires pour l'étalonnage du FEI et les facteurs de forme (la paramétrisation des éléments de matrice nécessaires aux calculs SM) de notre signal.

Les corrections PID et FEI sont effectuées à l'aide de tables de correction calculées par la collaboration. La correction des facteurs de forme, nécessaire pour tenir compte des facteurs de forme obsolètes utilisés pour la simulation MC, est tirée de [20]. Le BR des modes de désintégration dans les bruits de fond est corrigé pour correspondre à la valeur rapportée dans le PDG de 2024 [19]. En ce qui concerne les fonds spécifiques, nous corrigeons directement les événements $B \to Kn\bar{n}$ et $B \to KK_L^0K_L^0$, qui peuvent facilement imiter notre signal et dont la distribution cinématique est mal modélisée dans la simulation MC. Ensuite, nous corrigeons la normalisation de nos contributions de bruits de fond, $q\bar{q}$ et $B\bar{B}$ à partir de nos échantillons de controle. Pour le $q\bar{q}$, nous avons calculons la normalisation de l'échantillon hors-résonance à partir du rapport data/MC. Pour le $B\bar{B}$, il y a deux cas de figures: pour les événements avec un B_{tag} "mal-taggés", *i.e.* pour lequel au moins une fille n'est pas correctement reconstruite et elle n'est pas un photon, nous corrigeons en nous servant conjoitement des échantillons avec la mauvaise charge et avec la trace supplémentaire; pour les autres événements, nous pourrions utiliser l'échantillon avec intégration, mais nous avons vérifier que le rapport est compatible avec l'unité, donc aucune nécessité d'une correction supplémentaire sur la normalisation, Finalement, nous corrigeons la forme des distributions des variables qui seront utilisées pour le BDT. Pour cela, nous entrainons deux BDT à distinguer les données et les événements MC sur l'échantillon hors-résonance puis sur l´échantillon avec une trace supplémentaire.

Selection de la région de signal

Étant donné que la sélection revient à choisir une région de signal présentant une forte concentration de signal et une faible concentration de bruit de fond, nous sommes confrontés à une tâche de classification. Nous utilisons un BDT comme classificateur binaire pour séparer le signal du bruit de fond. Le BDT est entraîné à l'aide de 18 variables, et les hyperparamètres ont été optimisés à l'aide de `Optuna` [69]. Nous définissons ensuite notre région de signal en aplatissant la contribution du signal et en sélectionnant les événements dont la valeur à la sortie du BDT est supérieure à 0,4. L'efficacité de la coupure dans la région de signal est de 70×10^{-4}, soit environ deux fois l'efficacité obtenue par l'analyse HTA, avec une pureté de 6%, là aussi deux fois supérieure à ce qui avait été obtenu durant la précédente analyse. Le rapport data/MC dans la région de signal, montré dans la Table A.5, est considéré comme satisfaisant du fait que les rapports soient à moins de 2σ de l'unité.

Sideband	data/MC presel	data/MC signal region
offres	1.026 ± 0.017 (6181 data evts)	1.25 ± 0.35 (18 data evts)
wc	1.003 ± 0.014 (12195 data evts)	1.18 ± 0.24 (39 data evts)
extra	1.027 ± 0.010 (68520 data evts)	1.32 ± 0.20 (78 data evts)

Table A.5: Rapport du nombre d'événements data/MC dans chaque échantillon de contrôle, ainsi que avant et après la sélection de la région de signal. "offres" correspond à l'échantillon hors-résonance, "wc" est l'échantillon avec la charge erronée et "et" est l'échantillon avec les traces supplémentaires. La compatibilité data/MC est considérée comme satisfaisante dans la région de signal.

Étude des bruits de fond

Une fois la région de signal sélectionnée, nous avons déjà éliminé la majeure partie du fond $q\bar{q}$, mais il subsiste encore quelques fonds $B\bar{B}$, comme le montre la figure A.7. Nous pouvons donc étudier les fonds restants et voir si certains d'entre eux apparaissent fréquemment, afin de comprendre pourquoi ils imitent notre signal et d'essayer de les réduire davantage. Comme on peut le voir sur la figure A.8, les principales composantes du bruit de fond sont les événements $D\ell\nu + D\ell\nu$, où $\ell \in \{e, \mu\}$, le premier $D\ell\nu$ correspondant au B_{tag} et le second le B_{sig}. Bien que nous soyons satisfaits que notre B_{tag} soit correctement reconstitué, le fait est que le deuxième $D\ell\nu$ imite notre signal $K^+\nu\bar{\nu}$. Nous avons observé que la plupart d'entre eux étaient en réalité des D se désintégrant en un méson K accompagné de quelques mésons π. Nous avons vérifié que nous perdons le ℓ en dehors de la zone d'acceptance du détecteur et que le pion compagnon n'est pas reconstituable. Nous avons conclu à l'irréductibilité de ce bruit de fond.

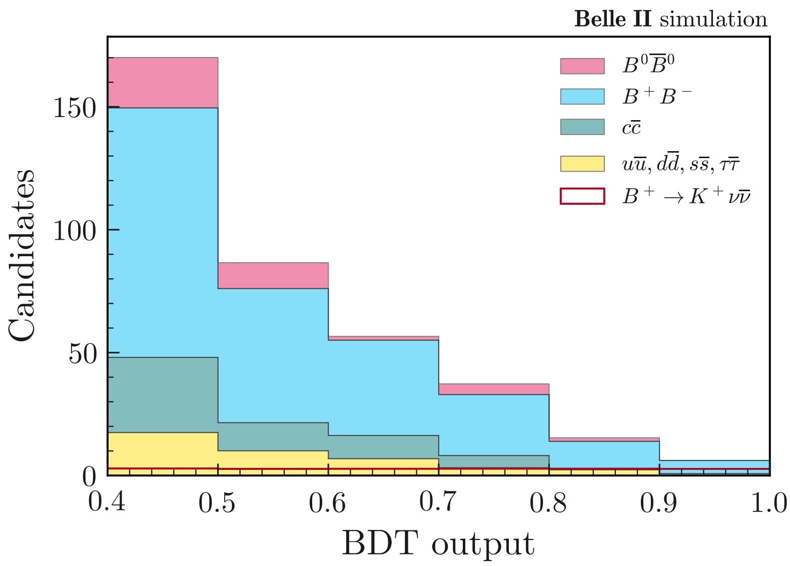


Figure A.7: Distribution de la sortie du BDT dans la région de signal.

Évaluation des incertitudes systématiques

La dernière étape avant l'ajustement consiste à évaluer les incertitudes systématiques de l'analyse. Nous avons une incertitude systématique par correction. D'autres incertitudes systématiques sont prises en compte, telles que l'incertitude sur le nombre de paires $B\bar{B}$, ou la statistique limitée de notre simulation. Pour chacune de ces sources, nous pouvons définir un *paramètre de nuisance* qui sera utilisé dans l'ajustement pour prendre en compte et contraindre l'incertitude, en suivant le formalisme `HistFactory` [78]. Au total, notre modèle statistique compte ainsi 168 paramètres.

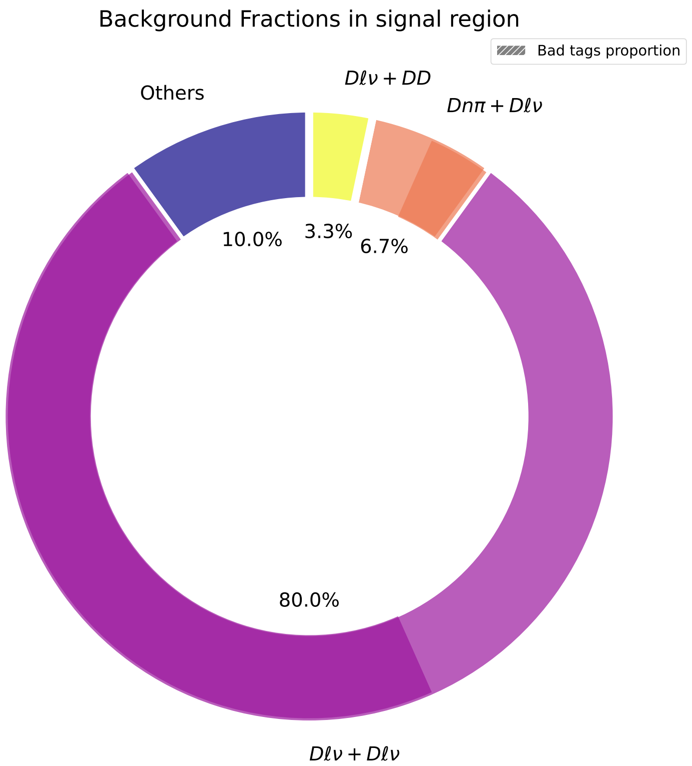


Figure A.8: Fractions de bruit de fond dans la région de signal.

Ajustement et extraction du BR

La dernière étape de l'analyse consiste à extraire le BR de la désintégration $B^+ \to K^+\nu\bar{\nu}$. Comme nous nous intéressons à l'écart éventuel par rapport à la prédiction du SM, nous préférons travailler avec l'*intensité du signal* μ, notre *paramètre d'intérêt* :

$$\mu = \frac{\mathrm{BR}(B^+ \to K^+\nu\bar{\nu})}{\mathrm{BR}_{\mathrm{SM}}(B^+ \to K^+\nu\bar{\nu})}$$

Afin d'obtenir l'intensité du signal puis le BR, nous commençons par créer un histogramme à 6 bins dans la région de signal du BDT. Nous obtenons ensuite le nombre attendu d'événements dans chaque bin et pour chaque type de contribution grâce à la simulation. À partir de ces chiffres, nous pouvons enfin effectuer un ajustement par bins par la méthode du maximum de vraisemblance pour extraire l'intensité du signal μ mais aussi les paramètres de nuisances résumés. Pour plus de détails sur cette partie et les aspects techniques de l'ajustement, veuillez vous reporter à la documentation officielle de `pure-python HistFactory` (`pyhf`) [79]. Nous pouvons ensuite effectuer un ajustement de la vraisemblance de profil sur μ pour obtenir la meilleure valeur d'ajustement $\hat{\mu}$ et l'intervalle de confiance associé, en utilisant le théorème de Wilks

[82] et la méthode CLs [83]. Le graphique issu de la méthode CLs est présenté dans Figure A.9.

L'ajustement est d'abord effectué sur l'ensemble de données Asimov, qui est l'ensemble de données pour lequel le nombre d'événements dans chaque classe est égal au nombre attendu d'événements. Cela nous permet d'obtenir la matrice de corrélation des différents paramètres du modèle statistique ainsi qu'une estimation de l'incertitude.

L'impact de chaque source d'incertitude est estimé à l'aide de la méthode décrite dans [84]. Les incertitudes estimées sont résumées dans Table A.6.

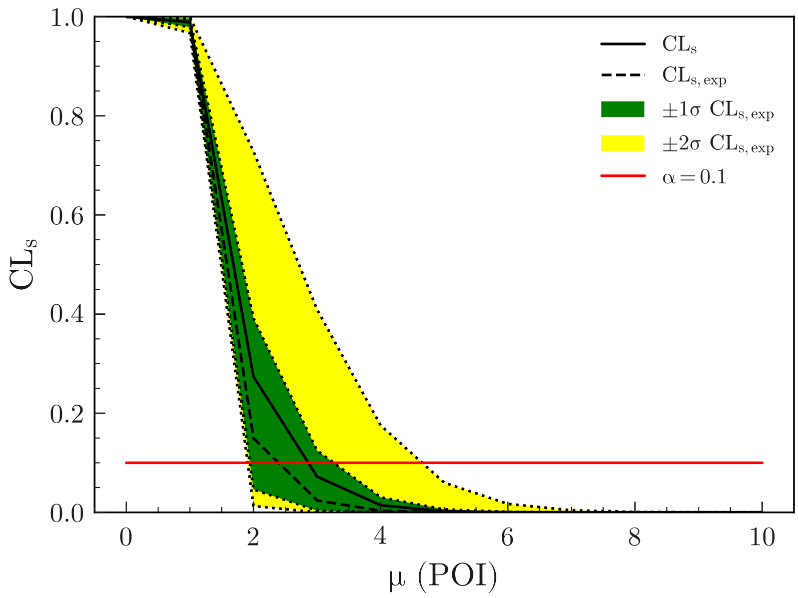


Figure A.9: Intervalle de confiance attendu en fonction de l'intensité du signal μ ajusté sur le jeu de données Asimov.

Afin de contrôler le biais de notre modèle statistique, nous réalision un test de linéarité: nous générons des modèles dont on fait varier les différents paramètres de nuisance dans leurs incertitudes, puis nous injecton du signal au travers de μ, puis nous répétons la procédure d'ajustement et extraction du signal. Nous testons en injectant du signal à [1, 3, 5, 7, 10], 10 étant une valeur plafond aux vues de la limite haute actuelle. Le test, représenté Figure A.10, montre clairement que notre modèle n'est pas biaisé.

Finalement, le dernier test que nous faisons consiste à injecter comme signal celui obtenu par la précédente analyse, à savoir un μ de 4.6 [5], puis de comparer l'incertitude obtenue et la significance dans le cas où nous devrions tester l'hypothèse "bruit de fond uniquement". L'incertitude pour ce test est montrée dans Table A.7 et est au même niveau que celle obtenue dans l'analyse citée. Pour ce qui est de la significance du test, nous obtenons une valeur de

Systematic Category	Absolute contribution on branching fraction ($\times 10^{-6}$)
StatMC	2.2
KL	1.0
Continuum	0.8
bkg norm	0.6
sgn eff + tracking	0.5
Backgrounds	0.4
KKLKL	0.3
nBB	0.2
Data/MC extra	0.1
FF	0.1
PID	0.1
D2KL	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**2.7**
Total statistical	**4.3**
Total	**5.1**

Table A.6: Résumé des incertitudes attendues sur le BR et leurs impacts

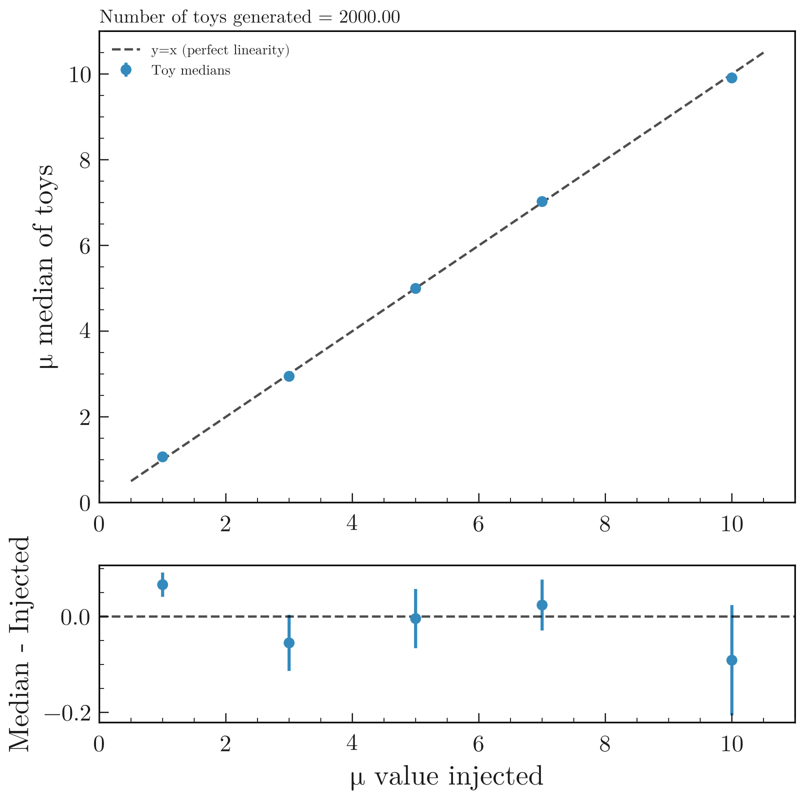


Figure A.10: Test de linéarité: on fait varier notre modèle statistique, puis on y injecte du signal et on répéte la procédure d'ajustement. La droite en pointillée représente le cas d'un modèle non-biaisé. 2000 modèles alternatifs ont été généré pour chaque hypothèse de signal.

2.43, ce qui ne nous donne pas assez pour proclamer une preuve de l'existence de cette désintégration au-delà d'un simple bruit de fond.

Systematic Category	**Absolute contribution on branching fraction ($\times 10^{-6}$)**
sgn eff + tracking	2.3
StatMC	2.1
bkg norm	1.7
KL	1.4
Continuum	0.9
nBB	0.6
Backgrounds	0.5
FF	0.4
KKLKL	0.3
Data/MC extra	0.2
PID	0.2
D2KL	0.0
KX2nn	0.0
Kneutrons	0.0
Total systematic	**4.0**
Total statistical	**5.7**
Total	**7.0**

Table A.7: Résumé des incertitudes attendues sur le BR et leurs impacts avec un signal injecté égal à celui obtenu par l'analyse précédente (4.6).

Résumé et prochaines étapes

Pour résumer, nous avons mis en place une analyse pour mesurer le BR de la désintégration $B^+ \rightarrow K^+\nu\bar{\nu}$ à l'aide d'une stratégie de tagging semi-leptique à l'aveugle pour éviter l'introduction de certains biais venant de l'expérimentateur. L'optimisation de la selection nous donne une efficacité de signal d'environ 70×10^{-4}, soit environ deux fois l'efficacité obtenue par l'analyse HTA, ainsi qu'une pureté de 6%, soit là aussi quasiment deux fois celle de l'analyse HTA. Nous avons déjà une bonne estimation de notre incertitude finale, à 5×10^{-6} pour l'hypothèse du SM et 7×10^{-6} pour l'hypothèse d'un rapport d'embranchement 4.6 fois supérieur à la valeur prédite dans le SM, ce qui est clairement compétitif avec les valeurs de l'état de l'art. Avant de pouvoir ajuster nos vraies données, nous devons encore ajouter des échantillons de simulation correspondant aux désintégrations $B \rightarrow X_s\nu\bar{\nu}$, un bruit de fond qui pourrait s'avérer important.

Finalement, cette analyse promet une précision excellente qui saura permettre de vérifier indépendamment les résultats obtenus précédemment mais aussi d'apporter une contribution majeure dans le cadre d'une combinaison avec les autres résultats.

Université de Strasbourg

IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Corentin SANTOS

Recherche des désintégrations $B \to K^{(*)}\nu\bar{\nu}$ à l'expérience Belle II

Résumé

Cette thèse détaille la stratégie pour mesurer la fraction de branchement des désintégrations $B \to K^{(*)}\nu\bar{\nu}$, en se concentrant sur $B^+ \to K^+\nu\bar{\nu}$, dans l'expérience Belle II avec 487,7 fb^{-1} de données à la résonance $\Upsilon(4S)$. Ces désintégrations, via la transition $b \to s\nu\bar{\nu}$, sont une sonde sensible de nouvelle physique grâce à la prédiction précise du modèle standard. La présence de deux neutrinos impose la reconstruction du B associé (B_{tag}) dans la réaction $\Upsilon(4S) \to B_{\text{tag}}\, B_{\text{sig}}$. Les B_{tag} semi-leptoniques sont reconstruits par l'algorithme FEI. L'analyse repose sur la simulation; des échantillons de contrôle sont définis pour corriger les erreurs de modélisation.
Avec 10 millions d'événements $B^+ \to K^+\nu\bar{\nu}$ et des échantillons $B\bar{B}/q\bar{q}$ correspondant à 1948 fb^{-1}, l'incertitude totale attendue est estimée à 5×10^{-6} (hypothèse SM) et 7×10^{-6} (valeur actuelle de l'art). Cette mesure sera aussi précise que les précédentes de Belle II, confirmant son importance comme mesure indépendante et pour la combinaison des résultats.

Mots-clés: Belle II, SuperKEKB, $B \to K^{(*)}\nu\bar{\nu}$, $B^+ \to K^+\nu\bar{\nu}$, physique des particules, méson B, change de saveur par courant neutre, nouvelle physique, modèle standard

Résumé en anglais

This thesis details the strategy for measuring the branching fraction of $B \to K^{(*)}\nu\bar{\nu}$ decays, focusing on $B^+ \to K^+\nu\bar{\nu}$, in the Belle II experiment using 487.7 fb^{-1} of data at the $\Upsilon(4S)$ resonance. These decays, via the $b \to s\nu\bar{\nu}$ transition, are a sensitive probe of new physics due to the Standard Model's precise prediction. The presence of two neutrinos requires the reconstruction of the associated B (B_{tag}) in the $\Upsilon(4S) \to B_{\text{tag}}\, B_{\text{sig}}$ reaction. The B_{tag} semi-leptonic events are reconstructed using the FEI algorithm. The analysis relies on simulation; control samples are defined to correct for modeling errors.
With 10 million $B^+ \to K^+\nu\bar{\nu}$ events and samples $B\bar{B}/q\bar{q}$ samples corresponding to 1948 fb^{-1}, the total expected uncertainty is estimated at 5×10^{-6} (under the SM assumption) and 7×10^{-6} (current state of the art). This measurement will be as precise as previous ones from Belle II, confirming its importance as an independent measurement and for combining the results.

Mots-clés: Belle II, SuperKEKB, $B \to K^{(*)}\nu\bar{\nu}$, $B^+ \to K^+\nu\bar{\nu}$, particle physics, B meson, Flavour Changing Neutral Current, New Physics, Standard Model