Search for $B \to K\nu\bar{\nu}$ decays with a machine learning method at the Belle II experiment

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Cyrille Praz
aus Sion (Schweiz)

Hamburg
2022

Gutachter/innen der Dissertation:	Dr. Alexander Glazov Prof. Dr. Kerstin Tackmann
Zusammensetzung der Prüfungskommission:	Dr. Alexander Glazov Prof. Dr. Johannes Haller Prof. Dr. Jochen Liske Dr. Kai Schmidt-Hoberg Prof. Dr. Kerstin Tackmann
Vorsitzender der Prüfungskommission:	Prof. Dr. Jochen Liske
Datum der Disputation:	23.08.2022
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

This thesis documents a search for the rare decay of a B meson into a K meson and a pair of neutrinos at the Belle II experiment, which is located along the SuperKEKB energy-asymmetric electron-positron collider. This decay has never been observed, its branching fraction is predicted with accuracy in the standard model of particle physics, and is a good probe of physics beyond the standard model. A novel method to search for this decay, the inclusive tagging, is developed on a data sample corresponding to an integrated luminosity of $189\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance, and a complementary sample of $18\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance. For this integrated luminosity, the expected upper limits on the branching fraction of $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^0_\mathrm{S} \nu\bar{\nu}$ are determined from simulation to be 1.0×10^{-5} and 1.8×10^{-5} at the 90% confidence level, respectively. When the method is applied to data samples of $63\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and $9\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance, no significant signal is observed, and an upper limit on the branching fraction of $B^+ \to K^+ \nu\bar{\nu}$ is determined to be 4.1×10^{-5} at the 90% confidence level.

Zusammenfassung

Diese Arbeit dokumentiert die Suche nach dem seltenen Zerfall eines B-Mesons in ein K-Meson und ein Paar Neutrinos mit dem Belle II-Experiment, am Elektron-Positron-Beschleuniger SuperKEKB. Dieser Zerfall eines B-Mesons wurde bis dato noch nie direkt beobachtet, sein Verzweigungsverhältnis ist im Standardmodell der Teilchenphysik jedoch genau vorhergesagt. Er stellt somit einen exzellenten Kandidaten dafür dar, mögliche Physik jenseits des Standardmodells zu untersuchen. Eine neue Methode zur Suche nach diesem Zerfall, das „Inclusive Tagging“, wird entwickelt auf einen Datensatz von $189\,\mathrm{fb}^{-1}$ integrierter Luminosität, aufgezeichnet bei einer Schwerpunktsenergie, die der $\Upsilon(4S)$-Resonanz entsoricht, ergänzt durch einen Datensatz von $18\,\mathrm{fb}^{-1}$, aufgezeichnet $60\,\mathrm{MeV}$ unterhalb der Resonanz. Basierend auf der Menge dieser Daten ist eine obere Grenze auf das Verzweigungsverhältnis von bis zu 1.0×10^{-5} für $B^+ \to K^+\nu\bar{\nu}$ und 1.8×10^{-5} für $B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$ zu erwarten bei einem Konfidenzniveau von 90%. Angewendet auf einem Datensatz von $63\,\mathrm{fb}^{-1}$, genommen an der $\Upsilon(4S)$-Resonanz und $9\,\mathrm{fb}^{-1}$ unterhalb der Resonanz wird kein signifikantes Signal beobachtet. Damit lässt sich ein Verzweigungsverhältnis von bis zu 4.1×10^{-5} für $B^+ \to K^+\nu\bar{\nu}$ Zerfälle ausschließen bei einem Konfidenzniveau von 90%.

Contents

1

Introduction

The standard model of particle physics (SM) is a theoretical framework describing the known elementary particles and their interactions, and was mainly developed in the second half of the last century [1, 2, 3, 4]. Since then, most of the predictions of the SM were successfully verified by a large number of experiments around the world. One of the most famous results is the discovery in 2012 of the Higgs boson by the ATLAS and the CMS experiments [5, 6], nearly 50 years after the existence of this particle had been predicted from theoretical considerations [7, 8, 9].

Despite its experimental success, the SM is not the theory of everything. In particular, the SM does not include a theory of gravity and is currently unable to explain the origin of dark matter, an unknown type of matter whose presence is deduced from gravitational effects [10, 11, 12]. Moreover, several experimental results appear to contradict some predictions of the SM, like the measured anomalous magnetic moment of the muon [13, 14], or the evidence for lepton flavour non-universality [15, 16, 17, 18].

These experimental tensions with the SM point to the presence of new physics (NP), not described by the SM. A large number of extensions of the SM were proposed in the last decades, predicting the existence of new particles or new interactions, and an important role of experimental particle physics is to exclude or at least constrain the proposed models, by measuring observables that are sensitive to NP.

The goal of this thesis is to search for the decay of a B meson into a K meson and a pair of neutrinos, noted $B \to K\nu\bar{\nu}$. This decay has never been observed, it is predicted with accuracy in the SM, and, as will be shown, is an excellent probe of physics beyond the SM. To achieve this goal, data collected by the Belle II detector [19] are analysed. This detector is located along the SuperKEKB accelerator [20], an electron-positron collider in Tsukuba, Japan, working at an energy just above the threshold to produce pairs of B mesons.

This thesis is organised as follows:

- Chapter 2 presents the theoretical motivation to search for $B \to K\nu\bar{\nu}$ decays. It starts with an overview of elements of the SM that are important to describe these decays, followed by an explanation of why the decays of B mesons are important tests of the SM, a theoretical computation of the $B \to K\nu\bar{\nu}$ decay probability, and a summary of previous experimental results.

- Chapter 3 describes the experimental setup, namely the SuperKEKB accelerator, needed to produce B mesons, and the Belle II detector, designed to observe the decay products of B mesons.

- Chapter 4 introduces several important data analysis techniques and tools that are used and referred to in Chapter 5 to select $B \to K\nu\bar{\nu}$ decay candidates, and to measure the decay probability.

- Chapter 5 presents every step of the method that is developed to select $B \to K\nu\bar{\nu}$ decay candidates, how data compares to simulation, what are the sources of systematic uncertainties, and finally the obtained results.

If you read this document in a portable document format (pdf), note that even if not highlighted, the references to chapters, sections, figures, tables, bibliography, pages, as well as most acronyms (including pdf), are clickable links, following an idea proposed by Tim Berners-Lee [21].

2

Theoretical motivation

This chapter presents the theoretical motivation to search for $B \to K\nu\bar{\nu}$ decays. Section 2.1 gives an overview of elements of the standard model (SM) that are important to describe these decays. Section 2.2 defines what is a B factory. Section 2.3 shows how the $B \to K\nu\bar{\nu}$ decay probability is computed in the SM, explains how new physics (NP) may affect this probability, and finally summarises previous experimental results.

2.1. The Standard Model

As stated in the introduction, the SM describes the known elementary particles and their interactions. The three interactions of the SM are the electromagnetic interaction, the strong interaction, and the weak interaction. These interactions are mediated by elementary bosons of spin-1. The photon (γ) is the mediator of the electromagnetic interaction, the gluons ($g_i : i = 1, ..., 8$) are the mediators of the strong interaction, and the weak bosons (W^+, W^-, Z^0) are the mediators of the weak interaction. The strong interaction is described by a theory called quantum chromodynamics (QCD), and the electromagnetic and the weak interactions are described by the electroweak theory (EW) [1, 2, 3]. In the

Table 2.1.: $SU(2)_L$ doublets in the SM made of left-handed leptons and quarks. The prime notation designs the weak eigenstate (see text for details).

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

SM, the bosons presented above emerge from the requirement for the theory to be invariant under local phase transformations [22]. These local phase transformations are called gauge transformations, and the bosons are therefore called gauge bosons.

Besides the gauge bosons, matter is built from elementary fermions of spin-$\frac{1}{2}$, which come in two kinds, the leptons and the quarks:

- There are three charged leptons (e^-, μ^-, τ^-), three neutral leptons, called neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$, and six corresponding antileptons: (e^+, μ^+, τ^+) and $(\overline{\nu}_e, \overline{\nu}_\mu, \overline{\nu}_\tau)$. The neutrinos are coupled only to the weak interaction, making them hard to detect. The charged leptons are coupled to the weak and the electromagnetic interactions. The leptons are organised in three generations $(e^-\ \nu_e)$, $(\mu^-\ \nu_\mu)$, $(\tau^-\ \nu_\tau)$ with similar properties, but different masses.

- There are six quarks (u, d, c, s, t, b) and six corresponding anti-quarks $(\overline{u}, \overline{d}, \overline{c}, \overline{s}, \overline{t}, \overline{b})$. They all are coupled to the weak, the strong and the electromagnetic interactions. Quarks are not observed free, but are confined within particles called hadrons. A meson is a hadron formed by a quark-antiquark pair, and a baryon is a hadron made of three quarks. A specificity of the t quark is that it decays before hadronising, because of the shortness of its lifetime. Since the quarks are coupled to the strong interaction, they also carry a strong charge c called color $c \in \{r, g, b\}$. Similarly to the leptons, the quarks are organised in three generations (ud), (cs), (tb) with similar properties, but different masses.

The weak interaction plays a central role in this thesis and is now examined further. An important fact is that the weak charged bosons $W^\pm$ couple only to fermions of left chirality. The left-handed fermions are grouped in doublets of the weak isospin group $SU(2)_L$. These $SU(2)_L$ doublets are listed in Table 2.1. The lepton-flavour universality principle states that the coupling strength to the $W^\pm$ bosons is identical among the lepton doublets.

In the SM with massless neutrinos, no transition is possible between two lepton generations.

In the quark sector, transitions between the generations are possible, because the mass eigenstates d, s, b are different from the weak eigenstates d', s', b' of Table 2.1. The mass and the weak eigenstates are related through the Cabibbo-Kobayashi-Maskawa (CKM) matrix [23]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.1}$$

where $V_{ij} \in \mathbb{C}$. Originally, the CKM matrix was introduced to explain the violation of the charge-parity (CP) symmetry by the weak interaction. The CP violation is possible only if some of the terms $V_{ij} \notin \mathbb{R}$, or equivalently, $V_{ij} \neq V_{ij}^*$. Kobayashi and Maskawa show in [23] that while it is possible to have a fully real 2×2 matrix by a suitable definition of the quark phases, this is not possible anymore for a 3×3 matrix. This prediction of a third generation of quarks came before the discovery of the b and t quarks.

In the SM, the CKM matrix, noted V_{CKM}, is unitary, meaning that $V_{\text{CKM}}^{-1} = V_{\text{CKM}}^{\dagger}$, which is equivalent to the unitary conditions

$$\sum_{i\in\{u,d,s\}} V_{ij} V_{ik}^* = \delta_{jk}, \qquad \sum_{j\in\{u,d,s\}} V_{ij} V_{kj}^* = \delta_{ik}, \tag{2.2}$$

where δ_{ij} is unity if $i = j$, zero otherwise. The cancellation of the off-diagonal terms in the sum is at the basis of the Glashow–Iliopoulos–Maiani (GIM) mechanism [24], which forbids flavour-changing neutral currents (FCNC) at the lowest order and also suppresses them at higher order. The GIM mechanism is discussed more concretely in Section 2.3.2 (see also Section 20.7 in [22]).

The vanishing sums of complex numbers in Equation (2.2) can be represented as triangles in the complex plane. Out of the six sums, the following one is conventionally chosen:

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0. \tag{2.3}$$

To draw the conventional unitary triangle, Equation (2.3) is reordered and divided by $V_{cd} V_{cb}^*$ to give

$$1 + \frac{V_{td} V_{tb}^*}{V_{cd} V_{cb}^*} + \frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} = 0, \tag{2.4}$$

so that the unitary triangle is represented as the null sum of three complex numbers $1 + z_1 + z_2$ (Figure 2.1). One of the main goals of flavour physics is to measure the parameters of this triangle.

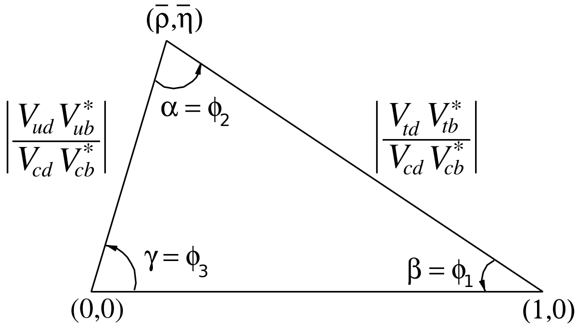


Figure 2.1.: Unitary triangle in the complex plane defined as the sum of the three complex numbers in Equation (2.4). The two real parameters $\bar\rho$ and $\bar\eta$ are defined as $\bar\rho + i\bar\eta = -V_{ud}V_{ub}^*/V_{cd}V_{cb}^*$. Credits to [25].

The experimental amplitudes of the CKM matrix components are approximately [25]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.226 & 0.004 \\ 0.226 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}. \tag{2.5}$$

When getting away from the diagonal, the magnitudes are getting smaller, meaning that the transition probability between two generations of quarks, which is proportional to the square of the magnitude, is also getting lower.

To conclude this first section, Table 2.2 lists the SM leptons, bosons, and a selection of mesons and baryons that are referred to in this thesis.

Table 2.2.: List of SM leptons, bosons, and of a selection of mesons and baryons. The experimental values of the particle spin, parity and mass are taken from [25]. Only rounded values of the particle masses are given, even if they are measured with greater precision.

Particle name	Symbol	Quark content	Spin$^{\text{Parity}}$	Mass [eV/c^2]
Electron	e^-	-	$\frac{1}{2}$	5.1×10^5
Electron neutrino	ν_e	-	$\frac{1}{2}$	< 1.1
Muon	μ^-	-	$\frac{1}{2}$	1.1×10^8
Muon neutrino	ν_μ	-	$\frac{1}{2}$	$< 1.9 \times 10^5$
Tau	τ^-	-	$\frac{1}{2}$	1.8×10^9
Tau neutrino	ν_τ	-	$\frac{1}{2}$	$< 1.8 \times 10^7$
Photon	γ	-	1^-	$< 10^{-18}$
Z boson	Z^0	-	1	9.1×10^{10}
W boson	W^+	-	1	8.0×10^{10}
Gluon	g	-	1^-	$< 10^7$
Higgs boson	H^0	-	0^+	1.3×10^{11}
Charged π meson (pion)	π^+	$u\bar{d}$	0^-	1.4×10^8
Neutral pion	π^0	$\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$	0^-	1.3×10^8
Charged K meson (kaon)	K^+	$u\bar{s}$	0^-	4.9×10^8
Neutral kaon	K^0	$d\bar{s}$	0^-	5.0×10^8
Short-lived kaon	K^0_{S}	$\frac{d\bar{s}-s\bar{d}}{\sqrt{2}}$	0^-	5.0×10^8
Long-lived kaon	K^0_{L}	$\frac{d\bar{s}+s\bar{d}}{\sqrt{2}}$	0^-	5.0×10^8
Charged D meson	D^+	$c\bar{d}$	0^-	1.9×10^9
Neutral D meson	D^0	$c\bar{u}$	0^-	1.9×10^9
J/Psi meson	$J/\psi(1S)$	$c\bar{c}$	1^-	3.1×10^9
Charged B meson	B^+	$u\bar{b}$	0^-	5.3×10^9
Neutral B meson	B^0	$d\bar{b}$	0^-	5.3×10^9
Upsilon meson	$\Upsilon(4S)$	$b\bar{b}$	1^-	1.1×10^{10}
Proton	p	uud	$\frac{1}{2}^+$	0.9×10^9
Neutron	n	udd	$\frac{1}{2}^+$	0.9×10^9

2.2. Physics of B factories

A B factory is an experimental setup designed to produce B mesons and to detect the product of their decays. The basic principle is to collide electrons and positrons at the energy of a $b\bar{b}$ resonance, the $\Upsilon(4S)$ boson, whose mass is $10.58\,\mathrm{GeV}/c^2$. When a $\Upsilon(4S)$ is produced, it decays into a pair of B mesons with a probability higher than 96% [25] .

The first generation of B factories, Babar [26, 27] and Belle [28, 29], collected data in the first decade of the third millennium. Babar was located along the PEP-II accelerator in Stanford, it has collected $433\,\mathrm{fb}^{-1}$ of data at the $\Upsilon(4S)$-resonance energy between 1999 and 2008, and the analysis of the data is still ongoing [30]. Belle was located along the KEKB accelerator in Tsukuba, Japan, it has collected $711\,\mathrm{fb}^{-1}$ of data at the $\Upsilon(4S)$-resonance energy between 1999 and 2010, and the analysis of the data is also ongoing [30]. The only representative of the second generation of B factories is the direct successor of Belle, Belle II, which is presented in more details in Chapter 3. Belle II started to record data in 2018 and has collected so far a dataset of approximately $400\,\mathrm{fb}^{-1}$.

The decays of B mesons provide a rich set of observables that are both tests of the SM and probes of potential physics beyond the SM. One reason for this is that the B^+ and B^0 mesons are the lightest hadrons containing a b quark, implying that their decays are necessarily based on a flavour-changing process. And the only mechanism in the SM allowing for this kind of process is the weak interaction, known to have interesting properties such as the violation of the P and the CP symmetries. The objectives of the B factories include:

- The study of the violation of the CP symmetry. Belle and Babar independently observed for the first time the presence of CP violation in a system of neutral B mesons [31, 32].
- The determination of the CKM quark-mixing matrix parameters, in particular $|V_{cb}|$ and $|V_{ub}|$, by studying B meson decays involving the transition of a b quark into a c quark or a u quark, a charged lepton and a neutrino [33, 34, 35, 36, 37].
- The search of rare decays forbidden at the tree level in the SM such as $B \to K^{(*)}\ell^+\ell^-$ ($\ell = e, \mu, \tau$) [38, 39, 40, 41] (more on that in Section 2.3.3).
- The measurement of the properties of other particles than B mesons. A B factory is also a D factory and a τ factory (see Section 3.1). For example, the mass of the τ lepton was measured by both Belle and Babar with a precision of the order of 0.02% [42, 43], and the Belle II measurement of the D^0 meson and D^+ meson lifetimes is the most precise to date [44].

B factories do not hold a monopoly on the study of B mesons. Another experiment whose attention is focused on B mesons is the LHCb experiment [45, 46], located along the Large Hadron Collider (LHC) [47], a proton-proton collider operated by the European organisation for nuclear research (CERN) across the Franco-Swiss border near Geneva. The LHC is working at an energy of 13 TeV in centre-of-mass system, far higher than the energy at B factories, and the current world record. This energy allows the experiments working with the LHC to explore the high-energy frontier, where potential new heavy particles are directly produced. By contrast, the B factories explore the high-precision frontier, where heavy particles contribute virtually to the Feynman diagrams.

An important advantage of a B factory is that the initial state, i.e. the four-momentum of the e^+e^- system, is fully known, allowing to efficiently reject background and also infer information about the final-state particles that are not detected (neutrinos or particles outside of the detector acceptance for example), thanks to the energy-momentum conservation. This is more difficult at LHCb, where the B mesons are produced from interactions of partons, whose initial energy and momentum are unknown (for example, a gluon issued from one proton beam may interact with a quark from the other proton beam).

2.3. The $B \to K\nu\bar{\nu}$ decay

This section focuses on the $B \to K\nu\bar{\nu}$ decay, which is based on the weak transition of a b quark into a s quark and a pair of neutrinos ($b \to s\nu\bar{\nu}$). This decay has never been observed and is rare in the SM, because there is no possibility of flavour-changing neutral current at the tree level. The lowest-order Feynman diagrams that contribute to the decay amplitude are either of the loop type or the box type, and involve virtual Z^0 and $W^\pm$ bosons (Figure 2.2).

Section 2.3.1 presents a technique called operator product expansion, which is employed when computing the amplitude of the weak decays of a hadron. Section 2.3.2 summarises how the branching fraction of $B \to K\nu\bar{\nu}$, noted $\mathrm{Br}(B \to K\nu\bar{\nu})$, is calculated. Section 2.3.3 explains why an experimental measurement of $\mathrm{Br}(B \to K\nu\bar{\nu})$ is a good probe of physics beyond the SM. Section 2.3.4 gives an overview of previous experimental searches for $B \to K\nu\bar{\nu}$ decays.

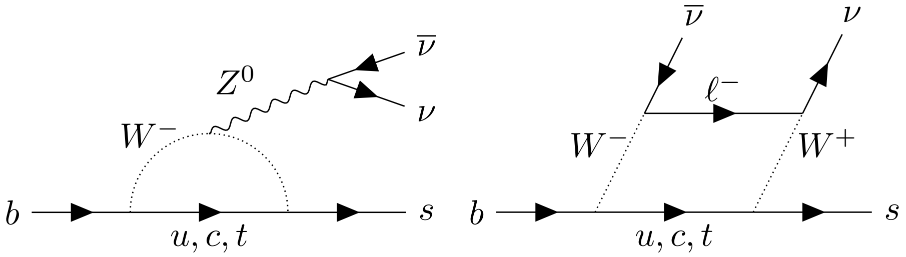


Figure 2.2.: Lowest-order Feynman diagrams of the transition of a b quark into an s quark and a pair of neutrinos. The diagrams are of the loop type (left), or of the box type (right).

2.3.1. Operator product expansion

In this section and the next one (Section 2.3.2), the natural system of units is used, in which $c = \hbar = 1$, where c is the speed of light, and $\hbar$ is the reduced Planck constant. The operator product expansion [48, 49] is a technique that simplifies the computation of the amplitude of the weak decays of a hadron. The core of the method is to work with an effective Hamiltonian written as the sum of local operators $\mathcal{O}_i$ representing effective point-like vertices depending on the weak decay under consideration [50, 51, 52]. Each operator $\mathcal{O}_i$ is multiplied by an effective coupling constant C_i called a Wilson coefficient, which captures the short-distance contributions from the virtual $W^\pm$ bosons, Z^0 boson, and t quark in particular.

More explicitly, in an operator product expansion, the effective Hamiltonian of the weak decay of a hadron is

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \sum_{i=1}^{N} \lambda_i^{\text{CKM}} \, C_i \, \mathcal{O}_i + \text{h.c.}, \tag{2.6}$$

where

- G_F is the Fermi constant, which is a function of the universal coupling constant of the weak interaction and the mass of W boson;
- λ_i^{CKM} are factors depending on the components of the CKM matrix;
- h.c. stands for Hermitian conjugate.

The Wilson coefficients C_i are derived from the evaluation of box and loop diagrams in perturbation theory. By contrast, the evaluation of the matrix elements of the local operators $\mathcal{O}_i$ involves long-distance contributions and requires non-pertubative techniques, such as lattice computation and light-cone sum rules.

The presence of the Fermi constant in Equation (2.6) is not accidental. This development is similar to the modern understanding of the Fermi description of β-decays [53, 54], in

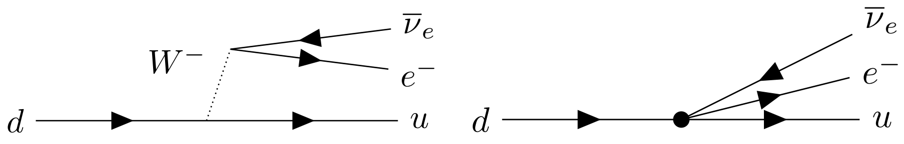


Figure 2.3.: Lowest-order Feynman diagram of the β-decay of a neutron (ddu) into a proton (udu), an electron and an anti-neutrino (left), and 4-fermion effective vertex of the same process in the modern understanding of the Fermi description (right). The spectator quarks (du) are not shown.

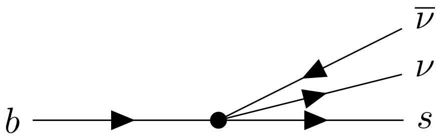


Figure 2.4.: In the operator product expansion, the $b \to s\nu\bar{\nu}$ transition is described with a single effective vertex.

which the full Feynman diagram that includes a W^- boson is replaced by a four-fermion effective vertex (Figure 2.3).

2.3.2. Branching fraction

This section explains how the SM predicts the value of the branching fraction of $B \to K\nu\bar{\nu}$, following mainly [55]. In the operator product expansion, the effective Hamiltonian for the $B \to K\nu\bar{\nu}$ decay is given as a function of a single operator $\mathcal{O}_L$ corresponding to the effective vertex shown in Figure 2.4. The effective Hamiltonian is

$$\mathcal{H}_{\text{eff}}^{\text{SM}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* C_L^{\text{SM}} \mathcal{O}_L + \text{h.c.}, \tag{2.7}$$

where

- V_{tb} and V_{ts} are elements of the CKM matrix defined in Section 2.1;
- C_L^{SM} is a dimensionless Wilson coefficient defined as $C_L^{\text{SM}} = -X(x_t)/\sin^2(\theta_W)$, where θ_W is the electroweak mixing angle [1, 2], and $X(x_t)$ is an Inami-Lim function [56] that describes the short-distance t-quark contribution (x_t is the ratio between the t-quark mass and the W-boson mass) [57, 58, 59];
- $\mathcal{O}_L$ is an operator defined as

$$\mathcal{O}_L = \frac{e^2}{16\pi^2} (\bar{s}_L \gamma_\mu b_L) (\bar{\nu}_L \gamma^\mu \nu_L), \tag{2.8}$$

where e is the electric charge of the positron, γ_μ are the Dirac matrices, and s_L, b_L, ν_L are the left-handed spinors describing the fermion fields.

In principle, contributions from the u quark and the c quark should also appear in Equation (2.7), resulting in the following sum:

$$\sum_{q\in\{u,c,t\}} V_{qb}V_{qs}^{*}X(x_q), \tag{2.9}$$

where $x_q \equiv m_q/m_W$. If the three quarks had an equal mass, this sum would cancel because of the unitarity of the CKM matrix, and FCNC processes would not be possible. However, the quarks have very different masses at this energy scale ($m_t > m_W \gg m_c, m_u$), and the t-quark term fully dominates in the result [57]. This is an example of breakdown of the GIM mechanism, due to the dispersion of the quark masses [51].

The total branching fraction of the $B \to K\nu\bar{\nu}$ decay is derived from Fermi's golden rule:

$$\mathrm{Br}(B \to K\nu\bar{\nu}) = N\tau_B \left|\langle K\nu\bar{\nu}|\, \mathcal{H}_{\mathrm{eff}}^{\mathrm{SM}}\, |B\rangle\right|^2 \rho, \tag{2.10}$$

where N is a normalisation factor, τ_B stands for the lifetime of the B meson, and ρ is a phase-space factor.

However, for reasons that will become clear later, it is more convenient to work with a differential branching fraction, which can be evaluated for different regions of the squared invariant mass of the two-neutrino system (q^2). The result of the calculation gives [55]

$$\frac{\mathrm{dBr}(B \to K\nu\bar{\nu})_{\mathrm{SM}}}{\mathrm{d}q^2} = 3\,\tau_B \left|\frac{G_F\alpha}{16\pi^2}\sqrt{\frac{m_B^3}{3\pi}}\, V_{tb}V_{ts}^{*}\, C_L^{\mathrm{SM}}\, f_+(q^2)\right|^2 \left(\frac{\lambda_K(q^2)}{m_B^4}\right)^{\frac{3}{2}}, \tag{2.11}$$

where, in addition to the quantities already defined above,

- the factor of 3 in front is the number of neutrino flavours, each contributing equally;
- α is the electromagnetic coupling (see section 10.2.2 of [25]);
- m_B is the mass of the B meson;
- $f_+(q^2)$ is a hadronic form factor, which is parametrised below, and that captures the q^2-dependence of the matrix element $\langle K|\, \bar{s}\gamma_\mu b\, |B\rangle$;
- $\lambda_K(q^2)$ is a phase-space factor, defined as $\lambda_K(q^2) \equiv \lambda(m_B^2, m_K^2, q^2)$, with $\lambda(a,b,c) \equiv a^2 + b^2 + c^2 - 2(ab + bc + ac)$. The normalisation by m_B^4 makes it a dimensionless factor.

The only factors that have a dimension on the right-hand side of Equation (2.11) are τ_B (GeV^{-1}), G_F^2 (GeV^{-4}), and m_B^3 (GeV^3). Combined together, they result in a quantity with dimension GeV^{-2}, as expected.

Integrated over the full q^2 range, the branching fractions of the decay in the charged case and the neutral case are [60]

$$\begin{aligned} \mathrm{Br}(B^+ \to K^+\nu\bar{\nu})_{\mathrm{SM}} &= (4.6 \pm 0.5) \times 10^{-6}, \\ \mathrm{Br}(B^0 \to K^0\nu\bar{\nu})_{\mathrm{SM}} &= (4.3 \pm 0.5) \times 10^{-6}, \end{aligned} \tag{2.12}$$

where the difference comes from the lifetime ratio $\tau_{B^+}/\tau_{B^0} = 1.076$ [25]. Experimentally, K_{S}^0 is easier to identify than K_{L}^0. For this reason, the branching fraction of the $B^0 \to K_{\mathrm{S}}^0\nu\bar{\nu}$ decay proves more useful:

$$\mathrm{Br}(B^0 \to K_{\mathrm{S}}^0\nu\bar{\nu})_{\mathrm{SM}} = \frac{\mathrm{Br}(B^0 \to K^0\nu\bar{\nu})_{\mathrm{SM}}}{2}. \tag{2.13}$$

The main source of theoretical uncertainty in Equation (2.12) comes from the hadronic form factor $f_+(q^2)$, which is now examined.

Form factor parametrisation

In [55], the form factor $f_+(q^2)$ is parametrised with three numbers $\alpha_1, \alpha_2, \alpha_3$ as

$$f_+(q^2) = \frac{1}{1 - q^2/m_+^2}\left[\alpha_1 + \alpha_2 z(q^2) + \alpha_3 z^2(q^2) + \frac{z^3(q^8)}{3}(-\alpha_2 + 2\alpha_3)\right], \tag{2.14}$$

where

$$z(t) = \frac{\sqrt{t_+ - t} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - t} + \sqrt{t_+ - t_0}}, \tag{2.15}$$

with $t_\pm = (m_B \pm m_K)^2$, $t_0 = t_+(1 - \sqrt{1 - t_-/t_+})$, and $m_+ = m_B + 0.046\,\mathrm{GeV}$. Combining the results from lattice computation valid at high q^2 and from light-cone sum rules valid at low q^2, a fit gives the following result [55]:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0.432 \\ -0.664 \\ -1.20 \end{pmatrix}. \tag{2.16}$$

Moreover, the covariance matrix of $\boldsymbol{\alpha}$, noted $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, which is necessary for propagating the uncertainties, is determined from the correlation matrix and uncertainties given in [55]. The computation gives

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \begin{pmatrix} 1.2 \times 10^{-4} & 3.4 \times 10^{-4} & -2.8 \times 10^{-3} \\ 3.4 \times 10^{-4} & 9.2 \times 10^{-3} & 1.7 \times 10^{-2} \\ -2.8 \times 10^{-3} & 1.7 \times 10^{-2} & 4.8 \times 10^{-1} \end{pmatrix}. \tag{2.17}$$

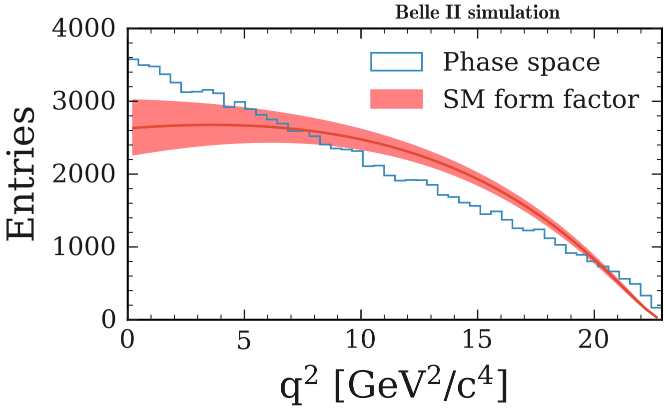


Figure 2.5.: Number of simulated $B^+ \to K^+\nu\bar{\nu}$ decays in bins of q^2 when taking into account only the variation coming from the phase space factor (blue histogram), and comparison with the expectation when including the q^2-dependence of the form factor (red line) and the theoretical uncertainty (red band).

Figure 2.5 gives a visualisation of the above result by comparing the q^2-dependence of simulated $B^+ \to K^+\nu\bar{\nu}$ decays when taking into account or not the form factor $f_+(q^2)$ present in Equation (2.11). The uncertainty band in Figure 2.5 is obtained as follows:

1. The three orthogonal unit eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ of $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ are extracted together with the respective eigenvalues σ_1^2, σ_2^2, σ_2^3;
2. The varied form factors $f_+(q^2, \boldsymbol{\alpha} \pm \boldsymbol{\sigma}_i)$ are computed for $i = 1, 2, 3$, with the variation vectors given by $\boldsymbol{\sigma}_i \equiv \sigma_i \, \boldsymbol{v}_i$;
3. The uncertainty band is defined as the region covering the varied form factors.

2.3.3. Search for New Physics

In case of presence of physics beyond the SM, other operators contribute to the Hamiltonian of the effective field theory (Equation (2.7)). If a violation of the lepton-flavour universality exists, there could be one left-handed operator per neutrino flavour. Moreover, a right-handed operator per neutrino flavour is also possible, giving a total of six operators (section 9.5 of [61]):

$$\mathcal{O}_L^\ell = \frac{e^2}{16\pi^2} \, (\bar{s}_L \gamma_\mu b_L) \, (\bar{\nu}_{\ell L} \gamma^\mu \nu_{\ell L}), \tag{2.18}$$

$$\mathcal{O}_R^\ell = \frac{e^2}{16\pi^2} \, (\bar{s}_R \gamma_\mu b_R) \, (\bar{\nu}_{\ell L} \gamma^\mu \nu_{\ell L}), \tag{2.19}$$

where $\ell = e, \mu, \tau$. One advantage of the operator product expansion is that it is possible to parametrise the NP contributions in a model-independent way, relying only on a set of six complex Wilson coefficients associated to the six operators above. In [55], the Wilson coefficients are combined to define the following real parameters that parametrise the NP effects:

$$\epsilon_\ell = \frac{\sqrt{\left|C_L^\ell\right|^2 + \left|C_R^\ell\right|^2}}{\left|C_L^{\mathrm{SM}}\right|}, \qquad \eta_\ell = \frac{-\mathrm{Re}\left(C_L^\ell C_R^{\ell *}\right)}{\left|C_L^\ell\right|^2 + \left|C_R^\ell\right|^2}. \tag{2.20}$$

In the SM, $\epsilon_e = \epsilon_\mu = \epsilon_\tau = 1$ and $\eta_e = \eta_\mu = \eta_\tau = 0$.

There are reasons to believe that NP effects may show up in $b \to s\nu\bar{\nu}$ transitions. The next paragraphs give a summary of recent experimental results that deviate from the SM predictions in the sector of the $b \to s\ell^+\ell^-$ transitions, and the end of this section gives three examples of NP models that have implications for $b \to s\nu\bar{\nu}$ transitions.

Experimental anomalies

In the last years, several experimental results were in tensions with the predictions of the SM for the family of decays $B \to K^{(*)}\ell^+\ell^-$ ($\ell = e, \mu$), which are based on the transition $b \to s\ell^+\ell^-$. One important observable in these experimental studies is the ratio

$$R_H \equiv \frac{\mathrm{Br}(B \to H\mu^+\mu^-)}{\mathrm{Br}(B \to He^+e^-)}, \tag{2.21}$$

where $H \in \{K^+, K^0, K^{*+}, K^{*0}\}$. The ratio of branching fractions is predicted with higher accuracy than the branching fractions themselves, because the form factor terms cancel out in the division. In the SM, the ratio R_H is close to unity because of the universality of the weak coupling.

Below are listed a selection of recent published results showing tensions with the SM. The tensions are quantified as a certain number of standard deviations (σ).

- In 2017 [15], LHCb reported a value of $R_{K^{*0}}$ $2.3\,\sigma$ smaller than the SM prediction;
- In 2020 [16], LHCb reported the result of a fit to the angular variables of $B^0 \to K^{*0}\mu^+\mu^-$ decays showing a $3.3\,\sigma$ deviation from the predicted SM value of a Wilson coefficient specific to $B \to K^{(*)}\ell^+\ell^-$ decays called C_9;
- In 2021 [17], LHCb reported the result of a fit to the angular variables of $B^+ \to K^{*+}\mu^+\mu^-$ decays showing a $3.1\,\sigma$ deviation from the predicted SM value of the same Wilson coefficient C_9;
- In 2022 [18], LHCb reported a value of R_{K^+} $3.1\,\sigma$ smaller than the SM prediction.

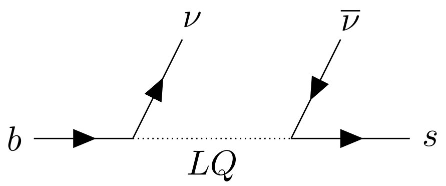


Figure 2.6.: Feynman diagram of the transition of a b quark into an s quark and a pair of neutrinos mediated by a hypothetical leptoquark (LQ).

The experimental picture is not completely clear. For example, Belle reported in 2021 [62, 41] values of R_H with $H \in \{K^+, K^0, K^{*+}, K^{*0}\}$ compatible with the SM, but with larger uncertainties than LHCb. And recently, LHCb also reported values compatible with the SM for $R_{K^0_\mathrm{S}}$ and $R_{K^{*+}}$ [63].

These tensions further motivate the study of $B \to K\nu\bar{\nu}$ decays to clarify the experimental picture, and to also impose constraints on NP models, some of them trying to explain the experimental anomalies.

Leptoquark

Leptoquarks are hypothetical particles that couple to quark-lepton pairs [64]. Recently, a family of scenarii of scalar and vector leptoquarks with a mass of the order of $1\,\mathrm{TeV}/c^2$ were proposed to explain the experimental anomalies observed on the $R_{K^{(*)}}$ observables [65, 66]. The existence of leptoquarks would have consequences on the branching fraction of $B \to K\nu\bar{\nu}$ decays by contributing to the Feynman diagrams, one example of scenario being shown in Figure 2.6. In particular, this would imply a shift in the Wilson coefficient C_L with respect to the SM prediction [65].

Axion

Axions (A^0) are hypothetical bosons than could explain why the strong interaction conserves the CP symmetry [67, 68, 69, 70]. In [71], the focus is given to invisible axions with a mass $m_{A^0} \ll \mathrm{eV}/c^2$ and a lifetime larger than the age of the universe. While it is possible to search specifically for the two-body decay $B \to KA^0$, an experimental search for $B \to K\nu\bar{\nu}$ decays is also imposing bounds on the A^0 coupling if $q^2 \approx 0\,\mathrm{GeV}^2/\mathrm{c}^4$ is included in the search region [71].

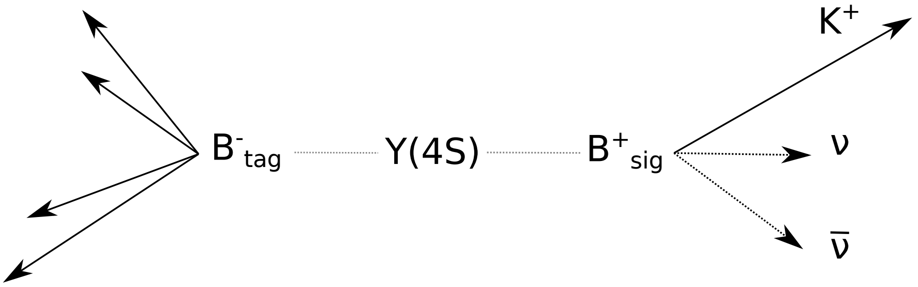


Figure 2.7.: Tagging method to search for $B^+ \to K^+\nu\bar{\nu}$ decays. The accompanying B meson (B^-_{tag}), is reconstructed either in a hadronic decay or in a semileptonic decay. The signal kaon is selected from the remaining particles in the event.

Dark matter

In [72], a new hypothetical scalar S with a mass in the GeV/c^2 range is proposed as the mediator of a new force between the SM particles and dark matter. In particular, the decay $B \to KS(\to \chi\bar{\chi})$, with χ an invisible dark matter fermion, would be possible. Noting that the final state of a $B \to KS(\to \chi\bar{\chi})$ decay is experimentally identical to the final state of a $B \to K\nu\bar{\nu}$ decay, a study of $\text{Br}(B \to K\nu\bar{\nu})$ in bins of q^2 can be reinterpreted to impose constraints on this model [72].

2.3.4. Overview of previous searches

While the search for $B \to K\nu\bar{\nu}$ decays has a strong theoretical motivation, it is experimentally challenging, because the two neutrinos escape the detection. The previous generation of B factories allowed the Belle experiment and the Babar experiment to search for this decay, but it has never been observed so far, so only experimental upper limits on its branching fraction are available.

All the previous searches for $B \to K\nu\bar{\nu}$ decays where relying on either the hadronic tagging method or the semileptonic tagging method (Figure 2.7). The first step of both methods is to reconstruct the accompanying B meson in the $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ event either in a hadronic decay or in a semileptonic decay. The second step is select the signal kaon from the remaining particles in the event.

Below, a brief chronological review of the most recent published results is given, and numerical results are reported in Table 2.3. The quoted integrated luminosities correspond to data collected at the $\Upsilon(4S)$ resonance along a e^+e^- collider.

- In 2010 [73], the Babar experiment used the semileptonic tagging method on a data sample corresponding to $418\,\mathrm{fb}^{-1}$ of integrated luminosity. First, $B_{\mathrm{tag}} \to D^{(*)}\ell\nu_{\ell}$ ($\ell = e, \mu$) decays were reconstructed, with the $D^{(*)}$ meson decaying hadronically. Then, signal kaon candidates (K^{+} or $K^{0}_{\mathrm{S}} \to \pi^{+}\pi^{-}$) were selected in the rest of the event, and the background was reduced with the use of boosted decision trees (see Section 4.1).

- In 2013 [74], the Babar experiment used the hadronic tagging method on a data sample corresponding to $429\,\mathrm{fb}^{-1}$ of integrated luminosity. First, $B_{\mathrm{tag}} \to SX_{\mathrm{had}}$ decays were reconstructed, were S stands for $D^{(*)}$, $D^{(*)}_{s}$, or J/ψ decaying hadronically, and X_{had} is a set of at most five mesons, each being either a kaon or a pion. Then, signal kaon candidates (K^{+} or $K^{0}_{\mathrm{S}} \to \pi^{+}\pi^{-}$) were selected in the rest of the event. In addition to the upper limits on the branching fraction obtained from the hadronic tagging method, upper limits combining the two tagging methods were also derived. For the K^{+} mode, the combined upper limit of $\mathrm{Br}(B^{+} \to K^{+}\nu\bar{\nu}) < 1.6 \times 10^{-5}$ at the 90% confidence level is the value currently reported by the Particle Data Group [25].

- In 2013 [75], the Belle experiment used the hadronic tagging method on a data sample corresponding to $711\,\mathrm{fb}^{-1}$ of integrated luminosity. First, B_{tag} candidates were reconstructed with a neural network-based algorithm [76] capable of recognizing > 1000 exclusive hadronic decays of the B meson. Then, signal kaon candidates (K^{+} or $K^{0}_{\mathrm{S}} \to \pi^{+}\pi^{-}$) were selected in the rest of the event.

- In 2017 [77], the Belle experiment used the semileptonic tagging method on a data sample corresponding to $711\,\mathrm{fb}^{-1}$ of integrated luminosity. First, $B_{\mathrm{tag}} \to D^{(*)}\ell\nu_{\ell}$ ($\ell = e, \mu$) decays were reconstructed with a neural network-based tagging algorithm described in [76, 78] and capable of recognising $\mathcal{O}(100)$ different semileptonic decays. Then, signal kaon candidates (K^{+} or $K^{0}_{\mathrm{S}} \to \pi^{+}\pi^{-}$) were selected in the rest of the event, and the background was reduced with the use of another neural network algorithm. In particular, the determined upper limit $\mathrm{Br}(B^{0} \to K^{0}\nu\bar{\nu}) < 2.6 \times 10^{-5}$ at the 90% confidence level is the best to date and is the value currently reported by the Particle Data Group [25].

Table 2.3.: Experimental results of previous searches for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0\nu\bar{\nu}$ decays. Are given the name of the experiment, the year of publication, the employed method (SL stands for semileptonic tagging, HAD stands for hadronic tagging, and COM stands for the combination of the two), the integrated luminosity of the data sample (L), the decay mode (either K^+ or K^0), the total signal selection efficiency (ε_{sig}), and the observed upper limit on the branching fraction at the 90% confidence level. The results of [73] are shown with an asterisk (*), because the q^2-dependence of the predicted branching fraction is not properly taken into account in the method, causing the signal selection efficiency to be approximately 10% greater than its correct value according to [74].

Experiment	Year	L [fb^{-1}]	Method	Mode	ε_{sig} [%]	Limit at 90% CL	Ref.
Babar	2010	418	SL	K^+	0.16*	$< 1.3 \times 10^{-5}*$	[73]
				K^0	0.03*	$< 5.6 \times 10^{-5}*$	[73]
Babar	2013	429	HAD	K^+	0.04	$< 3.7 \times 10^{-5}$	[74]
				K^0	0.01	$< 8.1 \times 10^{-5}$	[74]
Babar	2013	429	COM	K^+	-	$< 1.6 \times 10^{-5}$	[74]
				K^0	-	$< 4.9 \times 10^{-5}$	[74]
Belle	2013	711	HAD	K^+	0.06	$< 5.5 \times 10^{-5}$	[75]
				K^0	0.004	$< 19.4 \times 10^{-5}$	[75]
Belle	2017	711	SL	K^+	0.2	$< 1.9 \times 10^{-5}$	[77]
				K^0	0.05	$< 2.6 \times 10^{-5}$	[77]

3

Experimental setup

This chapter introduces the experimental setup employed to search for $B \to K\nu\bar{\nu}$ decays. The two main experimental components are the SuperKEKB accelerator (Section 3.1), which accelerates and collides electrons and positrons at an energy sufficient to produce B mesons, and the Belle II detector (Section 3.2), which is designed to detect the products of B meson decays.

3.1. The SuperKEKB accelerator

The SuperKEKB accelerator [20], whose schematic view is shown in Figure 3.1, is a double-ring electron-positron collider located in Tsukuba, Japan. Electrons and positrons are accelerated in an injector linear accelerator to an energy of $E_{e^-} = 7.0\,\mathrm{GeV}$ and $E_{e^+} = 4.0\,\mathrm{GeV}$, respectively, and stored in two rings. The two beams collide at an interaction point located within the Belle II detector (Section 3.2).

The invariant mass available in an electron-positron interaction, denoted as $\sqrt{s}$, is given by

$$\sqrt{s} \equiv \sqrt{\left(\frac{E_{e^-} + E_{e^+}}{c^2}\right)^2 - \left(\frac{\mathbf{p}_{e^-} + \mathbf{p}_{e^+}}{c}\right)^2} \approx \sqrt{(7+4)^2 - (7-4)^2} = 10.58\,\mathrm{GeV}/c^2, \quad (3.1)$$

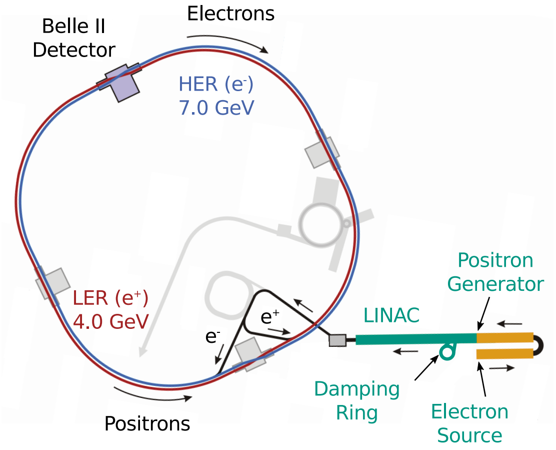


Figure 3.1.: Schematic of the SuperKEKB accelerator. Electron bunches are produced via the photoelectric effect with a photocathode, and positron bunches are produced by sending electrons on a tungsten target. A damping ring is employed to reduce the emittance of the positron beam. After an acceleration in an injector linear accelerator (LINAC), the electrons are stored at an energy of 7 GeV in a high-energy ring (HER), and the positrons at an energy of 4 GeV in a low-energy ring (LER). The two beams collide at an interaction point located within the Belle II detector. Credits to [79].

where $\mathbf{p}_{e^+}$ ($\mathbf{p}_{e^-}$) is the three-momentum of the positron (electron), and where the lepton mass is neglected in the second equality. As already mentioned in Section 2.2, this invariant mass of 10.58 GeV/c^2 corresponds to the mass of the $\Upsilon(4S)$ resonance, which, if produced in the interaction, decays into a pair of B mesons with a probability higher than 96% [25], making SuperKEKB a member of the B-factory family.

Table 3.1 lists the main processes expected in an electron-positron collision at this energy. In addition to $e^+e^- \to \Upsilon(4S)$, there are five background processes, referred to as continuum background, that will play an important role during the selection of $B \to K\nu\bar{\nu}$ decays: $e^+e^- \to q\bar{q}$ ($q = u, d, c, s$) and $e^+e^- \to \tau^+\tau^-$. The five other background processes listed at the bottom of Table 3.1 have relatively large cross-sections, but they are easy to suppress, because they largely differ from typical $e^+e^- \to \Upsilon(4S)$ events (Section 3.2.6).

When a pair of B mesons is produced, they are emitted nearly at rest in the $\Upsilon(4S)$ rest frame, since each B meson has a mass of 5.28 GeV/c^2, which is smaller but close to $\sqrt{s}/2$. In the laboratory frame, each B meson has a momentum of approximately $|\mathbf{p}_{e^+} + \mathbf{p}_{e^-}|/2 \approx$ 1.5 GeV/c, corresponding to a Lorentz boost factor $\beta\gamma \approx 0.284$, where $\beta = v/c$ is the ratio between the B meson velocity v and the speed of light, and $\gamma \equiv 1/\sqrt{1-\beta^2}$. This Lorentz boost is of particular interest for CP-violation studies, where the decay time difference between the two B mesons needs to be measured. If Δt is the decay time difference in

Table 3.1.: Cross-section of the main processes resulting from electron-positron collisions at an energy of 10.58 GeV in the centre-of-mass system. See chapter 4 of [61] for details on the quoted numbers.

Process	Cross-section [nb]
$e^+e^- \to \Upsilon(4S)$	1.10
$e^+e^- \to u\bar{u}(\gamma)$	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
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.92
$e^+e^- \to e^+e^-(\gamma)$	300
$e^+e^- \to e^+e^-e^+e^-$	39.7
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.9
$e^+e^- \to \gamma\gamma(\gamma)$	4.99
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.15

the centre-of-mass system, the distance Δz between the two decay vertices is given by $\Delta z = c\beta\gamma\Delta t$ and is of the order of 100 μm.

The designed instantaneous luminosity of SuperKEKB is $8 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, a value 40 times higher than what was achieved by the predecessor of SuperKEKB, KEKB. The main steps to reach such an instantaneous luminosity are to increase the beam currents by a factor of approximately two with respect to KEKB, and to squeeze the beams at the interaction point following a method called the nano-beam scheme, explained below, after the introduction of the coordinate system.

The origin of the Belle II coordinate system is the nominal interaction point. The x axis is in the horizontal plane and points towards the outside of the accelerator circular tunnel, the y axis is vertical and points upwards; the z axis is the symmetry axis of the Belle II solenoid and points in a direction close to the electron-beam direction.

In the nano-beam scheme [80], which is applied at SuperKEKB and depicted in Figure 3.2, the nominal vertical width σ_y of the bunches is squeezed to a value of the order of 50 nm. An important challenge is an effect called the hourglass effect, which makes the value of σ_y to reach its minimum only in a small z region (Figure 3.2). To overcome this, the nano-beam scheme works with a large half crossing angle $\phi_x \approx 40$ mrad and a small bunch width

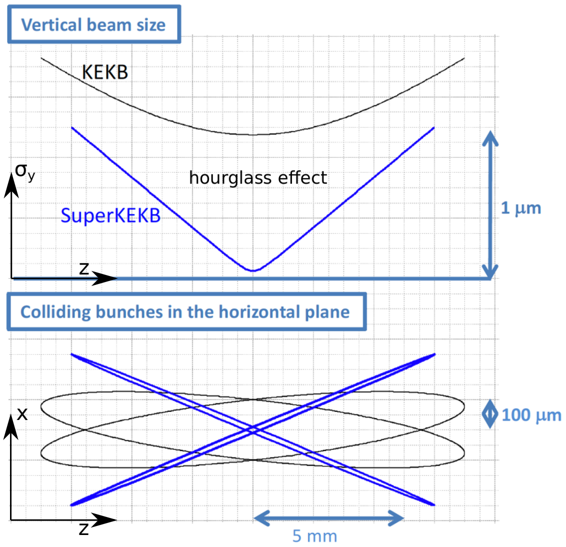


Figure 3.2.: Vertical beam size at KEKB and SuperKEKB illustrating the hourglass effect (top) and view from above of an electron bunch and a positron bunch crossing each other (bottom). In the nano-beam scheme used at SuperKEKB, only a small part of each bunch is overlapping at any given time, resulting in a small effective bunch length. Adapted from [20].

in the x direction $\sigma_x \approx 10\,\mu\text{m}$. By doing so, even if the actual length σ_z of the bunches in the z direction is of the order of 10 mm, only a small part of each bunch is overlapping at any given time, resulting in an effective bunch length of $\sigma_z^{\text{eff}} = \sigma_x / \sin\phi_x \approx 0.25\,\text{mm} \ll \sigma_z$.

From the beginning of the current operation phase of SuperKEKB in early 2019 until mid-2022, the Belle II detector has recorded a total integrated luminosity of the order of $400\,\text{fb}^{-1}$. So far, the maximum instantaneous luminosity achieved by SuperKEKB is approximately $3.8 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, which is the current world record, but still far from the design value.

3.2. The Belle II detector

A view of the Belle II detector is shown in Figure 3.3. Several subsystems work together to detect the particles produced in the electron-positron collisions and infer their momentum, mass and energy. A detailed description of the Belle II detector can be found in [19]. The rest of this section introduces the Belle II subsystems.

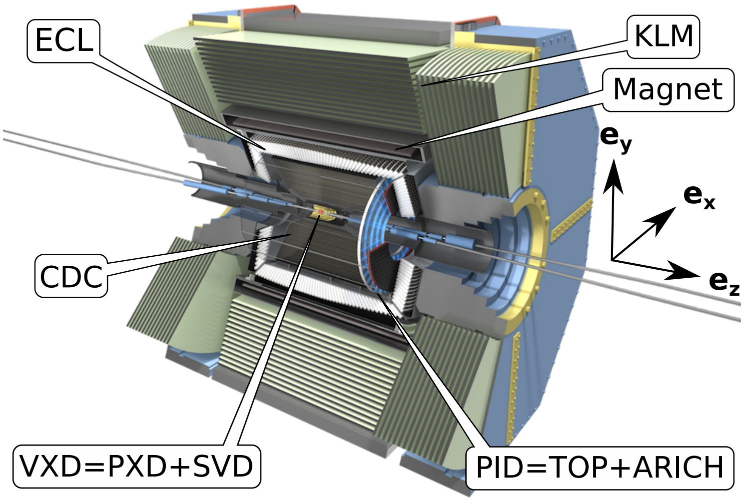


Figure 3.3.: Schematic of the Belle II detector, and of its subdetectors: the vertex detector (VXD), composed of the pixel vertex detector (PXD) and the silicon vertex detector (SVD), the central drift chamber (CDC), the time-of-propagation detector (TOP), the aerogel ring-imaging Cherenkov detector (ARICH), the K_{L}^0 and μ detector (KLM). The TOP and ARICH detectors are used in particular for particle identification (PID). The origin of the Belle II coordinate system is the nominal interaction point. The x axis is in the horizontal plane and points towards the outside of the accelerator circular tunnel, the y axis is vertical and points upwards; the z axis is the symmetry axis of the Belle II solenoid and points in a direction close to the electron-beam direction. The polar angle θ is defined with respect to the z axis, and the azimuthal angle ϕ is defined in the xy plane. Adapted from [19].

3.2.1. Tracking

The Belle II tracking system relies on three detectors: the pixel detector (PXD), the silicon vertex detector (SVD), and the central drift chamber (CDC). A comprehensive description of the track-finding system is given in [81]. In the following, each detector is briefly presented.

PXD

The PXD is the innermost part of Belle II and a novel detector that was not present in Belle. It detects the passage of charged particles by exploiting the production of electron-hole pairs in depleted p-channel field-effect transistors (DEPFETs) [82]. The choice of this detection technology is based on the need to cope with a large density of particles coming from the beams and the interaction region, while keeping thin sensors to reduce multiple scattering effects.

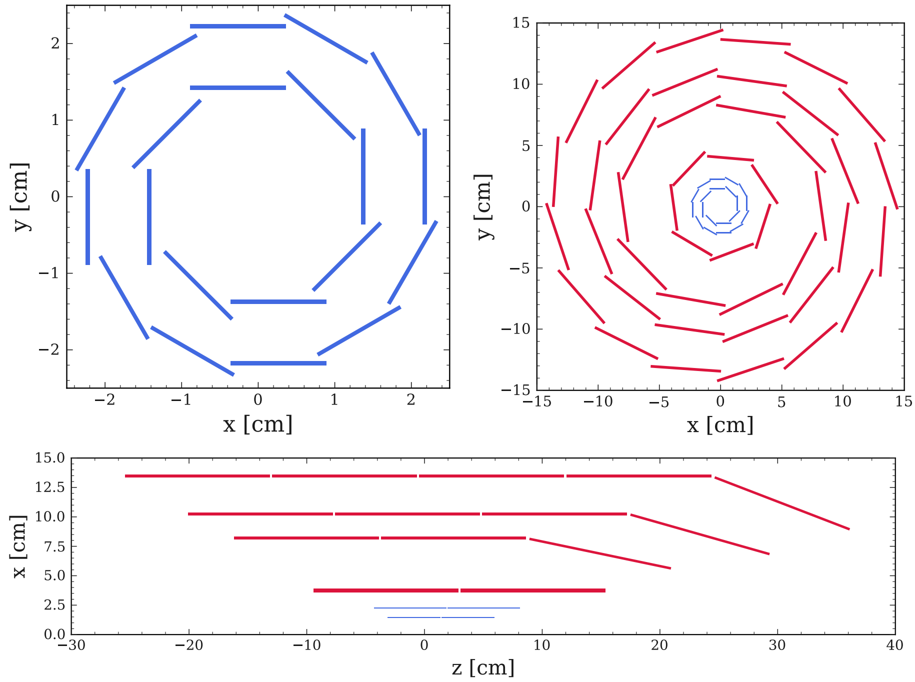


Figure 3.4.: PXD sensors in the transverse plane at $z = 0\,\mathrm{cm}$ (upper left), PXD and SVD sensors in the transverse plane at $z = 0\,\mathrm{cm}$ (upper right), and PXD and SVD sensors in the xz-plane at $y = 0\,\mathrm{cm}$ (bottom). Adapted from [79].

The PXD has a collection of 40 detection modules, each possessing a matrix of 768 × 250 DEPFET pixels. Pairs of modules are glued together to form ladders and the PXD is made of two layers of ladders around the beam pipe (Figure 3.4, upper left). The first layer is 14 mm away from the beam line and has 8 ladders; the second layer is 22 mm away from the beam line and will have 12 ladders. At the time of writing, the first layer of the PXD is fully installed, but only two ladders are installed in the second layer. The full installation of the PXD is foreseen by the end of 2022.

When combined with the SVD presented below, the presence of the PXD improves the impact-parameter and decay-time resolution by a factor of approximately two with respect to Belle, resulting in an impact-parameter resolution of approximately $12\,\mu\mathrm{m}$ and a D meson decay-time resolution of the order of 70 fs [44].

SVD

Going away from the beam pipe, the detection surface to cover increases as the square of the distance and a pixel-based detector is not an option beyond the PXD because of the cost and number of channels it would imply. For this reason, the SVD detects the passage

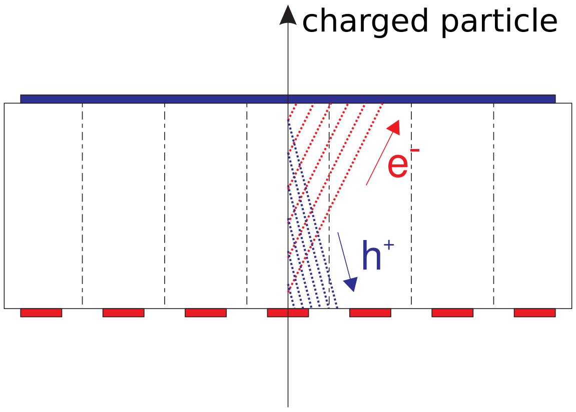


Figure 3.5.: Working principle of an SVD sensor. The n-side strips (blue rectangle at the top) are perpendicular to the p-side strips (red rectangles at the bottom). Adapted from [19].

of charged particles from the production of electron-hole pairs in double-sided silicon strip detectors (DSSDs), a technology that was already used in Belle.

The working principle of the SVD is illustrated in Figure 3.5. When a charged particle passes through an SVD sensor, the electrons and holes resulting from ionisation are collected by p-side and n-side strips located on each side of the sensor. The p-side strips are perpendicular to the n-side strips, so that the coordinates of the charged particle can be determined.

The SVD is organised in four layers of sensors whose radii with respect to the beam line range from 39 mm to 135 mm (Figure 3.4). Most sensors have a rectangular shape, with the exception of the sensors located in the most forward region of the three outer layers, which have a trapezoidal shape and are slanted in order to reduce the amount of needed material.

CDC

The CDC completes the tracking system of Belle II and measures the momentum and energy loss of charged particles, and provides trigger signals. The CDC detects the passage of charged particles from the ionisation of a gas mixture ($\mathrm{He} - \mathrm{C_2H_6}$) contained in its detection volume. The electrons that are released during the ionisation cause avalanches of electrons that are collected by a set of wires whose electric potential is tuned to create a strong electric field.

The CDC has a collection of 56 layers of wires, whose radial location with respect to the beam line ranges from 168 mm to 1111 mm. Axial wires are parallel to the beam axis

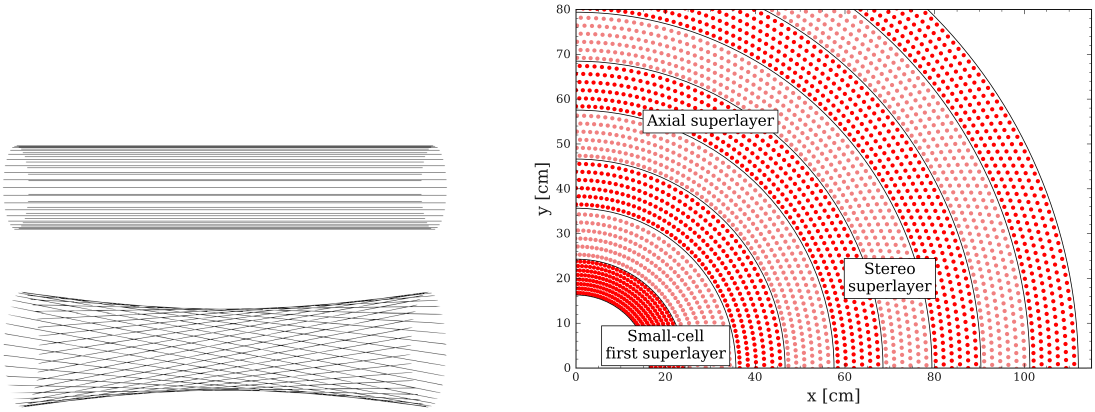


Figure 3.6.: Example of axial wires (top left), of stereo wires (bottom left), and CDC wires in the transverse plane at $z = 0\,\text{cm}$ (right). Layers of wires are grouped into axial and stereo superlayers. The first superlayer has a higher wire density. Adapted from [79, 81].

and stereo wires have a skewed orientation with respect to the axial wires (Figure 3.6). Layers of axial and stereo wires are grouped into axial and stereo superlayers. The first superlayer contains 8 layers of wires, and the other superlayers contain 6 layers of wires. By alternating stereo superlayers and axial superlayers (Figure 3.6), the three-dimensional trajectory of charged particles can be reconstructed in the CDC.

The momentum of a charged particle can be deduced from its reconstructed trajectory. At Belle II, a solenoid produces a magnetic field parallel to the beam axis, bending the charged particle trajectories in the transverse plane (Section 3.2.4). The transverse momentum p_T of the particle follows from

$$p_T = |q|B\rho, \tag{3.2}$$

where q is the electric charge of the particle, B is the magnitude of the magnetic field, and ρ is the bending radius of the particle trajectory. The relative uncertainty on p_T achieved by the CDC is of the order of 0.1% [83]. Combining the information about the momentum and the energy loss $\mathrm{d}E/\mathrm{d}X$ helps identifying charged particles[1], especially those whose momentum is smaller than $1\,\text{GeV}/c$ (section 34.2 of [25]).

3.2.2. Particle identification

The particle identification system is composed of two Cherenkov detectors: the aerogel ring-imaging Cherenkov detector (ARICH) and the time-of-propagation detector (TOP).

---

[1] In the expression $\mathrm{d}E/\mathrm{d}X$, E denotes the energy of the charged particle and X is a distance measured along the trajectory of the charged particle.

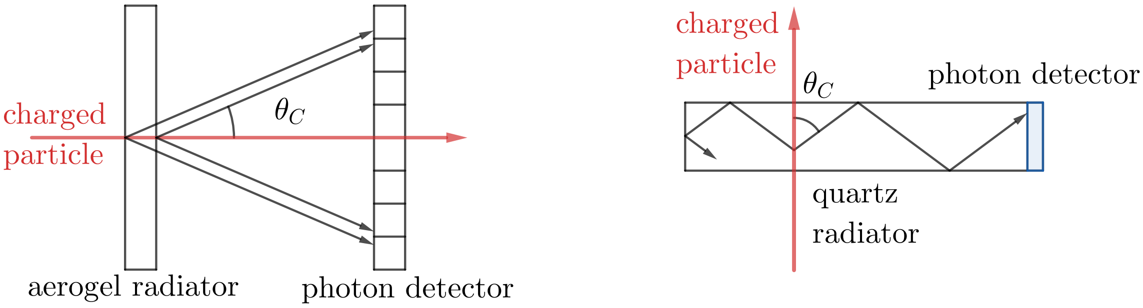


Figure 3.7.: Working principle of the ARICH (left) and TOP (right) detectors.

ARICH

The ARICH detector, located at the forward end-cap outside the CDC, is designed to distinguish between kaons and pions in the momentum range [0.4,4] GeV/c, as well as between muons, electrons and pions whose momentum is below 1 GeV/c. Its working principle is illustrated in Figure 3.7. When a charged particle passes through an aerogel radiator with a velocity larger than the speed of light in the radiator, the particle emits Cherenkov photons. The emission angle θ_C of the photons with respect to the velocity of the charged particle follows

$$\beta = \frac{1}{n \cos \theta_C}, \tag{3.3}$$

where β is the velocity of the charged particle divided by the speed of light in vacuum and n the refraction index of the radiator. By combining information about the momentum p and the velocity β of a particle, the mass of the particle (i.e. its identity) is given by $m = p/(\gamma\beta)$, where $\gamma = 1/\sqrt{1-\beta^2}$.

TOP

The TOP detector is installed in the barrel region outside the CDC and its primary purpose is to identify kaons and pions. Its working principle is illustrated in Figure 3.7. The Cherenkov photons emitted by a charged particle propagate through a quartz radiator. The photons stay within the radiator thanks to total internal reflections before reaching a photon detection plate located at the edge of the radiator. The propagation time of the Cherenkov photons is a function of the Cherenkov angle θ_C, from which the velocity of the particle can be derived with Equation (3.3). A detailed description of the TOP detector can be found in [84]. In particular, one detail that is not mentioned here is the need to take into account chromaticity, i.e. the dependence of θ_C on the photon wavelength.

3.2.3. Calorimetry

The electromagnetic crystal calorimeter (ECL) occupies the remaining volume inside the Belle II magnet. Its main role is to detect photons and measure their energy, and to identify electrons. The ECL also participates in the detection of K^0_L, provides trigger signals and is used to measure the luminosity received by Belle II.

The ECL is a collection of thousands of scintillation crystals, made of thallium-doped Cesium Iodide (CsI(Tl)), and covering a polar angle range of $12.4° < \theta < 155.1°$, i.e. approximately 90% of the solid angle in the centre-of-mass frame. Each crystal is coupled to two photodiodes for readout. When a photon or an electron passes through an ECL crystal, it may initiate an electromagnetic shower. The final products of the electromagnetic shower, low energy photons, are collected by the photodiodes and converted to an electronic signal.

In terms of performance, the ECL provides a mass resolution of approximately $5\,\mathrm{MeV}/c^2$ for a neutral pion decaying into a pair of photons. In addition, the ECL provides a relative energy resolution of approximately 4% at an energy of 100 MeV and 2% at 8 GeV [61, 85].

3.2.4. Solenoid

A superconducting solenoid radially surrounds the ECL. It generates a magnetic field of 1.5 T parallel to the beam axis in a cylindrical volume of 1.7 m in radius and 4.4 m in length. The superconductor is an allow of niobium and titanium and is cooled with a cryogenic system based on liquid helium. The generated magnetic field bends the trajectory of the charged particles and allows for a measurement of their transverse momentum (Equation (3.2)).

3.2.5. K^0_L and μ detection

The K^0_L and μ detection system (KLM) forms the outermost part of the Belle II detector. Its primary function is to detect muons and showers initiated by K^0_L mesons. In addition, it serves as a magnetic return circuit for the solenoid.

The KLM is an alternating series of iron plates and detection layers located outside of the solenoid. The iron plates provide a total of 3.9 nuclear interaction lengths of material for K^0_L mesons passing through with a normal incidence, one nuclear interaction length being defined as the mean distance travelled by a K^0_L meson before undergoing an inelastic nuclear interaction. By comparison, the ECL provides 0.8 nuclear interaction lengths of material for K^0_{L} mesons.

A muon with a momentum higher than $\sim 0.6\,\mathrm{GeV}/c$ passes through the KLM with a nearly straight trajectory and escapes the detector. On the other hand, a K^0_{L} meson likely interacts with a nucleus of the ECL or the KLM material and initiates a hadronic shower detected by the ECL, the KLM or both.

3.2.6. The trigger system

The Belle II trigger system selects events of interest resulting from electron-positron interactions. The trigger system is divided in two layers:

- The online trigger (L1) combines information from several trigger subsystems and takes the decision to keep or to exclude a recorded event with a fixed latency of $5\,\mu$s. Since this decision needs to come very fast, L1 is based on configurable hardware and only a partial reconstruction of the event is made. The main L1 trigger subsystems are the CDC and the ECL, which provide information about the tracks and energy clusters present in the event.
- The high-level trigger (HLT) further selects events of interest based on a full event reconstruction made at the software level.

In particular, the trigger system is employed to suppress low multiplicity background sources such as $e^+e^- \to \ell^+\ell^-$, $e^+e^- \to e^+e^-\ell^+\ell^-$ ($\ell = e, \mu$), and $e^+e^- \to \gamma\gamma$. These background sources, that were already mentioned at the beginning of this chapter (Table 3.1), are easy to identify, because they usually do not produce more than two tracks in the CDC nor more than two clusters in the ECL. By contrast, typical $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ events produce at least three tracks in the CDC, and for those events the trigger efficiency is close to 100%.

3.2.7. Beam-induced backgrounds

In addition to the low-multiplicity backgrounds mentioned in the previous section, below are listed three beam-induced backgrounds that are present at Belle II:

- The Touschek scattering, which is a Coulomb scattering of two electrons or positrons in the same bunch. This scattering causes the particles to deviate from their nominal energy and to induce showers by interacting with the beam pipe. The Touschek scattering rate is inversely proportional to the beam size and thus enhanced by the nano-beam scheme presented in Section 3.1. The Touschek scattering is mitigated by the use of collimators and metal shields to prevent the scattered particles from reaching the Belle II detector.

- The beam–gas scattering, which is a scattering (either Coulomb or Bremsstrahlung) of beam particles by residual gas molecules present in the beam pipe. Similarly to the Touschek scattering, the beam-gas scattering causes the particles to deviate from their nominal energy and to hit the inner side of the beam pipe. This source of background is also mitigated with collimators and metal shields.
- The synchrotron radiation that is emitted by the beams. The power of the synchrotron radiation varies as the square of the beam energy, implying that this radiation comes more from the electron beam (7 GeV) than from the positron beam (4 GeV). To protect the Belle II detector from the synchrotron photons, the inner surface of the beam pipe is coated with an absorbing layer of gold. Moreover, the shape of the pipe is designed to avoid direct hits on the detector from synchrotron photons.

3.2.8. Software

The Belle II analysis software framework (basf2) [86] is a large collection of open-source tools and algorithms for event simulation, reconstruction and analysis. In particular, this framework includes tools for unpacking raw data, track finding and reconstruction, ECL clustering, vertex fitting, and the computation of high-level variables.

In addition to basf2, the following packages are employed for event generation at Belle II:

- EvtGen [87] simulates the decays of B mesons.
- PYTHIA8 [88] simulates the hadronisation of the quarks issued from continuum events $e^+e^- \to q\bar{q}$ $(q = u, d, c, s)$.
- KKMC [89, 90] generates $e^+e^- \to \tau^+\tau^-$ events.
- TAUOLA [91, 92] handles the simulation of τ-lepton decays.
- GEANT4 [93] simulates the detector response.

4

Data analysis techniques

This chapter presents a collection of data analysis techniques and tools that will be used in the next chapter to search for $B \to K\nu\bar{\nu}$ decays. Since the content of this chapter is rather technical, the reader may skip it and come back to it when the tools are effectively employed in the next chapter.

In Section 4.1, binary classification algorithms, decision trees and boosted decision trees, are introduced. In Section 4.2, a set of discriminative variables are defined. In Section 4.3, the definition of a covariance matrix and of a correlation matrix are recalled, as well as a method to estimate them. Section 4.4 introduces a tool to measure a branching fraction from a binned maximum-likelihood fit. Section 4.5 presents a statistical method to determine an upper limit on a branching fraction. Section 4.6 defines a non-parametric probability density estimator called the kernel density estimator.

4.1. Binary classification

Given N_{variable} explanatory variables, two classes, signal ($y = 1$) and background ($y = 0$), and a set of N_{train} examples

$$\{(\mathbf{x}_1, y_1),\ ...,\ (\mathbf{x}_{N_{\text{train}}},\ y_{N_{\text{train}}})\}, \tag{4.1}$$

where $\mathbf{x}_i \in \mathbb{R}^{N_{\text{variable}}}$ is an observation of the explanatory variables associated with class $y_i \in \{0, 1\}$, the goal of binary classification is to define, based on the set of examples, an algorithm that outputs a signal probability prediction $\hat{y} \in (0, 1)$ for any input $\mathbf{x} \in \mathbb{R}^{N_{\text{variable}}}$. The construction of the algorithm, learning from the set of examples, is called training.

A loss function $\ell(\cdot, \cdot)$, which measures the distance between a class example y_i and a corresponding prediction $\hat{y}_i$, is defined and a naïve training would minimise the sum

$$\sum_{i=1}^{N_{\text{train}}} \ell\left(y_i, \, \hat{y}_i\right). \tag{4.2}$$

The usual loss function used in binary classification is called the negative binomial log-likelihood (or cross-entropy), and is defined as [94]

$$\ell\left(y, \, \hat{y}\right) = -\left[y \log \hat{y} + (1 - y) \log(1 - \hat{y})\right]. \tag{4.3}$$

Below are presented two binary classification algorithms: the decision tree, and the gradient-boosted decision tree. An important difference between the two algorithms is their behaviour regarding the bias-variance trade-off: if no limit is imposed on the model complexity, it is possible to train a model that separates perfectly the signal from the background when this model is applied on its own training set (low bias), but that makes poor predictions when applied to an independent dataset (high variance); this regime is called overfitting. On the other hand, if the model is too simple, then its performance does not depend on the dataset (low variance), but its predictions are inaccurate (high bias); this regime is called underfitting.

4.1.1. Decision tree

A decision tree is a recursive partitioning of the explanatory variable space [94]. The first node of a decision tree divides the space into two subspaces (branches) according to the value of a single variable with respect to a certain threshold (Figure 4.1). This procedure is repeated by the following nodes, until reaching the terminal nodes, called leaves. Each leaf corresponds to a certain region of the variable space and is associated with a weight, a negative weight corresponding to a background prediction, and a positive weight corresponding to a signal prediction (Figure 4.1).

For a given observation $\mathbf{x} \in \mathbb{R}^{N_{\text{variable}}}$, a decision tree classifier t assigns a weight $f_t(\mathbf{x}) \in (-\infty, +\infty)$ to the observation $\mathbf{x}$. The corresponding predicted signal probability $\hat{y}(\mathbf{x}) \in (0, 1)$ is computed as

$$\hat{y}(\mathbf{x}) = P\left(f_t(\mathbf{x})\right) = \frac{1}{1 + \exp(-f_t(\mathbf{x}))}, \tag{4.4}$$

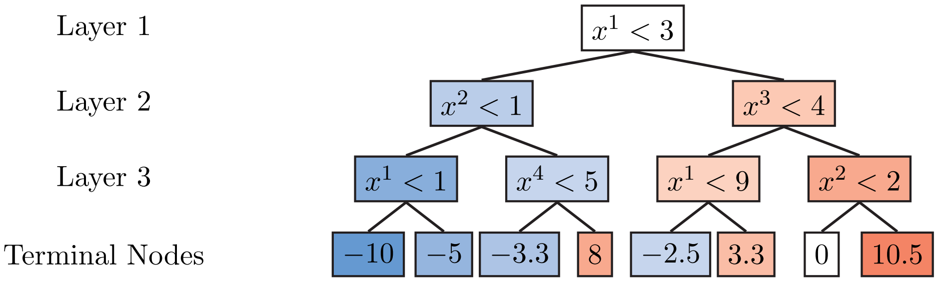


Figure 4.1.: Example of binary decision tree with three layers and four explanatory variables $\mathbf{x} \equiv (x^1, x^2, x^3, x^4)$, where the superscripts are indices, not exponents. The decision tree is organised in layers of nodes, and each node divides the remaining space in two, depending on the value of an explanatory variable x^i with respect to a certain threshold. Each terminal node is associated with a weight $f_t(\mathbf{x}) \in (-\infty, +\infty)$, a large weight corresponding to a high predicted signal probability. Adapted from [95].

where $P(\alpha) \equiv 1/(1 + \exp(-\alpha))$ maps the raw weight $\alpha \in (-\infty, +\infty)$ into a predicted signal probability $P(\alpha) \in (0, 1)$.

The main disadvantage of a simple decision tree is that it falls rapidly in an overfitting regime when the depth of the tree increases. For example, an arbitrary set of 10^6 data points can be exactly classified by a decision tree with a depth of 20, because $2^{20} > 10^6$.

4.1.2. Gradient-boosted decision tree

By contrast with relying on a single high-complexity decision tree to make predictions, the idea behind boosted decision trees is to base the prediction on a set of $N_{\text{tree}} \gg 1$ low-complexity tree classifiers [96, 97].

For a given observation $\mathbf{x} \in \mathbb{R}^{N_{\text{variable}}}$, and $t \in \{1, 2, ..., N_{\text{tree}}\}$, the low-complexity tree classifier t assigns a weight $f_t(\mathbf{x}) \in (-\infty, +\infty)$ to the observation $\mathbf{x}$. The weights of the trees are added together to obtain a global prediction weight

$$f^{(N_{\text{tree}})}(\mathbf{x}) = \sum_{t=1}^{N_{\text{tree}}} f_t(\mathbf{x}), \tag{4.5}$$

and the predicted signal probability $\hat{y}^{(N_{\text{tree}})}(\mathbf{x})$ is given by

$$\hat{y}^{(N_{\text{tree}})}(\mathbf{x}) = P\left(f^{(N_{\text{tree}})}(\mathbf{x})\right), \tag{4.6}$$

where $P : \mathbb{R} \to (0, 1)$ is defined in Equation (4.4).

The training of a boosted decision tree is done in a forward-stagewise additive manner. An initial uniform weight $f^{(0)}(\mathbf{x}) = 0 \implies P(f^{(0)}(\mathbf{x})) = 0.5$ is set for all $\mathbf{x}$. Then, each decision tree classifier $t = 1, 2, ..., N_{\text{tree}}$ is trained by iteratively solving [97]

$$f_t(\mathbf{x}) = \arg\min_{f(\mathbf{x})} \left\{ \sum_{i=1}^{N_{\text{train}}} \ell\left[y_i,\, P\left(f^{(t-1)}(\mathbf{x}_i) + f(\mathbf{x})\right)\right] + \Omega(f_t) \right\}, \tag{4.7}$$

where $f^{(t-1)}(\mathbf{x}_i)$ is the sum of the weights up to the previous iteration (Equation (4.5)), $l(\cdot,\cdot)$ is the loss function (Equation (4.3)), and $\Omega(f_t)$ is a term that penalises the complexity of the updated model f_t.

The minimisation of the objective function defined in Equation (4.7) is done by computing the gradient of the loss function. For this reason, this family of models are called gradient-boosted decision trees.

It is shown in [98] that the forward-stagewise additive modelling approach (Equation (4.7)) with an exponential loss function is equivalent to the AdaBoost algorithm [99]. In AdaBoost, a sequence of low-complexity classifiers is trained on iteratively reweighted versions of the training sample. At each iteration, the weight of each element of the training sample is increased if it is not correctly classified. This procedure allows the later low-complexity classifiers to focus on the elements of the training sample that are more difficult to classify.

4.1.3. Classifier parameters

Two implementations of the gradient-boosted decision tree algorithm, FastBDT [95] and XGBoost [97], are employed later in this thesis. In both implementations, the following parameters are tunable during the training:

- The total number of trees in the model;
- The maximal depth of each tree;
- The learning rate $0 < \eta \leq 1$, which is a factor that scales down the newly added weights at each iteration of the boosting, in order to reduce the influence of each individual tree;
- The sampling rate $0 < s \leq 1$, which is the fraction of the training sample that is randomly sampled for each iteration of the training;
- The number of equal-frequency bins in which each training variable is divided into (see [95]);

Increasing the value of any of the above parameters increases the complexity of the model, making it less robust against overfitting.

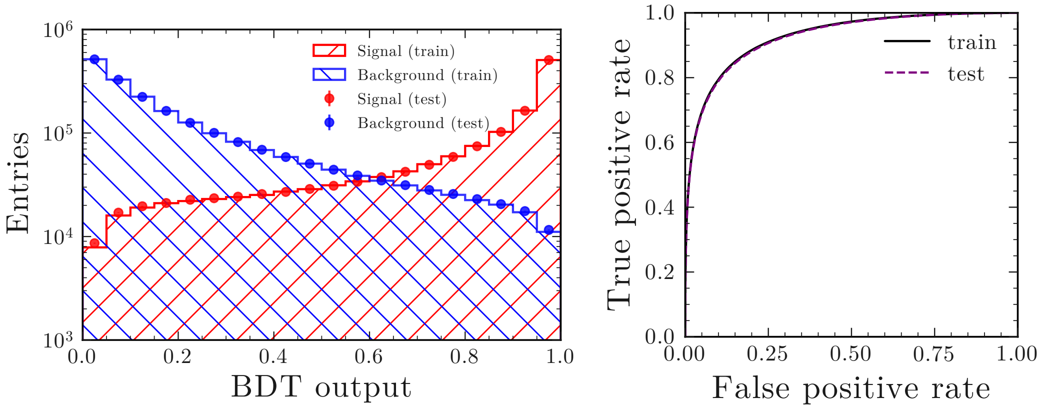


Figure 4.2.: On the left, example of boosted decision tree output for signal and background inputs issued from the training sample (train) and an independent sample (test) of same size. On the right, corresponding ROC curves.

4.1.4. Classifier overfitting and performance

As mentioned above, a high-complexity model may fall in an overfitting regime, in which the classification performance of the model is significantly better when the model is applied on its own training sample than when applied on an independent sample called a test sample. A method to check whether a model is overfitting is to compare the output of the classifier for the training and the test samples (Figure 4.2, left). If the comparison does not show an important difference, the model is not overfitting.

An equivalent method is to examine the receiver operating characteristic (ROC) curve [100], which represents the evolution of the true positive rate and the false positive rate when scanning all possible lower thresholds on the classifier output (Figure 4.2, right). If the ROC curve for the training sample is close to the ROC curve for the test sample, as it is the case in Figure 4.2, the model is not overfitting.

In addition, the area under the ROC curve, noted AUC, is a measure of the classification performance. A perfect classification is characterised by AUC $= 1$, meaning that there exists a lower threshold on the classifier output that selects a sample with a true positive rate of unity and a false positive rate of zero. The opposite case is a purely random classification, where the classifier output is uniform for both signal and background inputs. In that case, the ROC curve would be a straight line going from $(0, 0)$ to $(1, 1)$ in Figure 4.2 (right), and AUC $= 0.5$.

4.1.5. Input variable importance

In order to quantify the relative importance of an input variable[1] in the classification, one starts by defining the gain provided by a tree node (Figure 4.1). The gain of a particular node is the quantity by which the objective function (Equation (4.7)) is improved with the introduction of this node. From this, the importance of a variable v in the classification is defined as the sum of the gains across all the nodes in which the variable v is used, normalised by the total gain:

$$\mathrm{Importance}(v) \equiv \frac{\sum\limits_{s\in\{\mathrm{splits\,on}\,v\}} \mathrm{Gain}(s)}{\sum\limits_{s\in\{\mathrm{splits}\}} \mathrm{Gain}(s)}. \tag{4.8}$$

Other methods to quantify the importance of an input variable are possible. For example, one can examine how the classification performance varies when including or not the variable in the training.

4.2. Topological discrimination variables

In this section, several variables characterising the distribution of the momenta in an event resulting from a electron-positron collision are defined. These variables will prove useful to distinguish between signal and background events, and are employed in the next chapter as input variables of boosted decision trees. All of the variables presented in this section are defined for a set of N three-momenta $\mathbf{p}_1$, $\mathbf{p}_2$, ..., $\mathbf{p}_N$ measured in the centre-of-mass system. The respective norms are noted p_1, p_2, ..., p_N.

4.2.1. Sphericity

The sphericity matrix $\mathbf{S}$ is the symmetric 3×3 matrix whose elements $S^{\alpha\beta}$ $(\alpha, \beta = x, y, z)$ are defined as (chapter 9 of [30]):

$$S^{\alpha\beta} = \frac{\sum\limits_{i=1}^{N} p_i^\alpha p_i^\beta}{\sum\limits_{i=1}^{N} p_i^2}. \tag{4.9}$$

The event sphericity S is the real number

$$S = \frac{3}{2}(\lambda_2 + \lambda_3), \tag{4.10}$$

---

[1] Input variable is another name for explanatory variable.

where λ_2 and λ_3 are the two smallest eigenvalues of $S^{\alpha\beta}$.

The event sphericity S is a number between 0 and 1. An event that has a spherical topology is characterised by a sphericity close to unity. This is understood intuitively from the fact that a perfectly spherical distribution is described by a diagonal matrix with the single eigenvalue $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, and that an event where all momenta are collinear is characterised by $\lambda_1 = 1 > \lambda_2 = \lambda_3 = 0$.

4.2.2. Thrust

The thrust axis $\mathbf{T}$ of an event is defined as the unit vector along which the sum of the projections of the momenta $\sum_{i=1}^{N} |\mathbf{T} \cdot \mathbf{p}_i|$ is maximal (chapter 9 of [30]).

The thrust magnitude of an event is the positive number

$$T = \frac{\sum\limits_{i=1}^{N} |\mathbf{T} \cdot \mathbf{p}_i|}{\sum\limits_{i=1}^{N} |\mathbf{p}_i|}. \tag{4.11}$$

A spherical event has a thrust close to zero, and a jet-like event has a thrust close to unity.

4.2.3. Fox-Wolfram moments

The normalised Fox-Wolfram moments were introduced by Geoffrey C. Fox and Stephen Wolfram in [101, 102]. They are noted R_ℓ ($\ell = 1,\ 2,\ ...$) and defined recursively as follows:

$$H_0 = \sum_{i,j=1}^{N} p_i\, p_j, \tag{4.12}$$

$$H_1 = \sum_{i,j=1}^{N} p_i\, p_j\, \cos\alpha_{ij}, \tag{4.13}$$

$$H_\ell = \sum_{i,j=1}^{N} p_i\, p_j\, P_\ell\left(\cos\alpha_{ij}\right), \tag{4.14}$$

$$R_\ell = \frac{H_\ell}{H_0}, \tag{4.15}$$

where $P_\ell(\cdot)$ is the Legendre polynomial of ℓ-th order and $\alpha_{i,j}$ is the angle between $\mathbf{p}_i$ and $\mathbf{p}_j$. The first Legendre polynomials are

$$P_l(x) = \begin{cases} l = 0: & 1, \\ l = 1: & x, \\ l = 2: & \frac{1}{2}\,(3x^2 - 1), \\ l = 3: & \frac{1}{2}\,(5x^3 - 3x), \\ l = 4: & \frac{1}{8}\,(35x^4 - 30x^2 + 3). \end{cases} \tag{4.16}$$

4.2.4. Harmonic moments

The harmonic moments B_ℓ ($\ell = 0, 1, ...$) with respect to an axis $\mathbf{A}$ are defined as [101]

$$B_0 = \sum_{i=1}^{N} \frac{p_i}{\sqrt{s}}, \tag{4.17}$$

$$B_1 = \sum_{i=1}^{N} \frac{p_i}{\sqrt{s}} \cos\alpha_i, \tag{4.18}$$

$$B_\ell = \sum_{i=1}^{N} \frac{p_i}{\sqrt{s}} P_\ell\left(\cos\alpha_i\right), \tag{4.19}$$

where $\sqrt{s}$ is the available energy in the centre-of-mass frame, and the angle α_i is measured between $\mathbf{p}_i$ and the axis $\mathbf{A}$.

4.2.5. Modified Fox-Wolfram moments

The modified Fox-Wolfram moments were introduced by the Belle collaboration in [103] in a search for $B^0 \to \pi^0\pi^0$ decays. Below, the definitions and notations are close to what is found in chapter 9 of [30].

Before defining these modified Fox-Wolfram moments, a few quantities are introduced, assuming as before an event composed of N particle candidates. First, a signal B meson candidate is formed from a collection of N_B particle candidates. The rest of the event (ROE) is defined as the set of the $N_{\text{ROE}} = N - N_B$ particle candidates that are not associated with the B meson candidate. The number of charged particle candidates in the ROE is noted N_c, and the number of neutral particle candidates in the ROE is noted N_n, meaning that $N_c + N_n = N_{\text{ROE}}$. In addition, the missing momentum in the event $\mathbf{p}_{\text{miss}}$ is defined as the momentum needed to cancel the sum of all the other momenta in the event in the centre-of-mass frame:

$$\mathbf{p}_{\mathrm{miss}} = -\sum_{i=1}^{N} \mathbf{p}_i. \tag{4.20}$$

The signal-ROE (so) modified Fox-Wolfram moment of degree $\ell \in \mathbb{N}$ and of category $\xi \in \{\text{charged}\,(n), \text{neutral}\,(n), \text{missing}\,(m)\}$ is defined as

$$H^{so}_{\xi,\ell} = \frac{1}{Z} \sum_{i=1}^{N_B} \sum_{j_\xi=1}^{N_\xi} C^{\ell}_{ij_\xi}\, p_{j_\xi}\, P_\ell(\cos\alpha_{ij_\xi}), \tag{4.21}$$

where from left to right:

- the normalisation factor $Z = 2(\sqrt{s} - E^*_B)$ depends on the available energy in the centre-of-mass frame ($\sqrt{s}$) and the energy of the signal B meson candidate in the centre-of-mass frame (E^*_B).
- the first sum runs over the N_B children of the signal B meson candidate;
- the second sum runs over the N_c charged candidates in the ROE, or the N_n neutral candidates in the ROE, or the unique missing momentum ($N_m = 1$);
- the factor $C^{\ell}_{ij_\xi} \in \{-1, 0, +1\}$ is the product of the charges of candidate i and candidate j_ξ if ℓ is odd (the missing momentum is assumed to correspond to a neutral charge), and $C^{\ell}_{ij_\xi} = 1$ if ℓ is even;
- $P_\ell(\cdot)$ is the Legendre polynomial of ℓ-th order (Equation (4.16));
- α_{i,j_ξ} is the angle between $\mathbf{p}_i$ and $\mathbf{p}_{j_\xi}$.

The ROE-ROE (oo) modified Fox-Wolfram moment of degree $\ell \in \mathbb{N}$ is

$$R^{oo}_{\ell} = \frac{1}{Z^2} \sum_{i=1}^{N_{\mathrm{ROE}}} \sum_{j=1}^{N_{\mathrm{ROE}}} C^{\ell}_{ij}\, p_i\, p_j\, P_\ell(\cos\alpha_{ij}), \tag{4.22}$$

with the same notations as in Equation (4.21).

4.3. Covariance matrix estimate

This section recalls the definition of a covariance matrix and of a correlation matrix, and presents a method to estimate them.

Given a p-dimensional real random variable

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \tag{4.23}$$

the corresponding covariance matrix $\mathbf{\Sigma}$ is the $p \times p$ symmetric matrix defined as

$$\mathbf{\Sigma} = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]. \tag{4.24}$$

The correlation matrix of $\mathbf{x}$ is the covariance matrix of the normalised random variable $\mathbf{x}'$ defined as

$$\mathbf{x}' = \begin{pmatrix} x_1/\sigma(x_1) \\ \vdots \\ x_p/\sigma(x_p) \end{pmatrix}, \tag{4.25}$$

where σ stands for the square root of the variance.

Typically, the covariance and the correlation matrices are unknown and need to be estimated from a set of N observations $\mathbf{x}^1$, ..., $\mathbf{x}^N$. The (i,j)-th element of an estimate of the covariance matrix $\hat{\mathbf{\Sigma}}$ is given by

$$\hat{\Sigma}_{ij} = \sum_{r=1}^{N} \frac{\left(x_i^r - \sum_{r'=1}^{N} \frac{x_i^{r'}}{N} \right) \left(x_j^r - \sum_{r'=1}^{N} \frac{x_j^{r'}}{N} \right)}{N}. \tag{4.26}$$

Since the covariance matrix is real, symmetric and positive semi-definite, it exists p orthogonal unit eigenvectors $\mathbf{v}_1$, ..., $\mathbf{v}_p$, and corresponding eigenvalues $\sigma_1^2 \geq ... \geq \sigma_p^2 \geq 0$ such that

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^{p} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T, \tag{4.27}$$

where Q is the $p \times p$ matrix whose columns are the eigenvectors, and Λ is the diagonal matrix whose diagonal elements are the eigenvalues.

If some eigenvalues are significantly larger than others, a reasonable approximation is given by

$$\mathbf{\Sigma} \approx \sum_{i=1}^{t} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T + \mathrm{diag}\left(\sum_{i=t+1}^{p} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T \right), \tag{4.28}$$

where only the first $t < p$ terms associated with the t largest eigenvalues are fully considered, while only the diagonal elements of the $p - t$ smallest terms are added.

The variation vectors $\boldsymbol{\sigma}_i \equiv \sigma_i \, \boldsymbol{v}_i$ for $i = 1, ..., t$ are used to propagate correlated uncertainties, in the same manner as what was done for the form factor variations (Equation (2.17) and following lines in Section 2.3.2), and the remaining terms in Equation (4.28) are treated as uncorrelated uncertainties on each element of the random variable $\mathbf{x}$.

4.4. Binned maximum-likelihood fit

The main goal of this thesis is to measure the branching fraction of $B \to K\nu\bar{\nu}$, or equivalently, the signal strength μ defined as a factor with respect to the standard model prediction for $\mathrm{Br}(B \to K\nu\bar{\nu})$:

$$\mu = \frac{\mathrm{Br}(B \to K\nu\bar{\nu})}{\mathrm{Br}(B \to K\nu\bar{\nu})_{\mathrm{SM}}}. \tag{4.29}$$

To extract μ from data, a binned maximum-likelihood fit is applied. The procedure, explained in details in [104, 105], is summarised below, starting by introducing each element used to define the likelihood model. Since this presentation is rather abstract, the reader may prefer to come back to this section when the method is concretely used from Section 5.10.

Given a set of N_b bins counting events after a certain selection, the expected numbers of events in each bin, noted ν_1, ..., ν_{N_b}, are derived from simulation and modelled as the sum of several contributions, coming from one signal sample and $n \geq 1$ background samples:

$$\nu_b(\mu, \boldsymbol{\theta}) = \sum_{s \in \{\mathrm{samples}\}} \nu_{bs}(\mu, \boldsymbol{\theta}), \tag{4.30}$$

where ν_{bs} is the number of expected events in bin b for the sample s, μ is the signal strength defined above, and $\boldsymbol{\theta}$ is a vector of N nuisance parameters that apply variations to the nominal expectations. Assuming $n \geq 1$ background samples, $\boldsymbol{\theta}$ contains n normalisation parameters μ_1, ..., μ_n, one for each background sample, and $N - n$ other nuisance parameters:

$$\boldsymbol{\theta} = \Big(\mu_1, \, ..., \, \mu_n, \, \theta_{N-n+1}, \, ..., \, \theta_N\Big)^T. \tag{4.31}$$

The notation μ_i ($i = 1$, ..., n) is not accidental: similarly to the signal strength $\mu \equiv \mu_{\mathrm{signal}}$, each μ_i corresponds to a certain background strength. From this, Equation (4.30) can be developed as

$$\nu_b(\mu, \boldsymbol{\theta}) = \sum_{s \in \{\mathrm{samples}\}} \mu_s \left(\nu_{bs}^0 + \Delta_{bs}(\boldsymbol{\theta})\right), \tag{4.32}$$

where ν^0_{bs} is the nominal number of expected events in bin b for sample s, μ_s is a normalisation variation for sample s (it is the same for all the bins), and $\Delta_{bs}(\boldsymbol{\theta})$ is an additive variation in bin b for sample s. More explicitly,

$$\Delta_{bs}(\boldsymbol{\theta}) = \sum_{i=N-n+1}^{N} \theta_i \, \delta^i_{bs}, \tag{4.33}$$

with δ^i_{bs} an additive variation for bin b and sample s. This additive variation is scaled up or down by the nuisance parameter θ_i. The set of numbers δ^i_{bs} is an input of the model and describes the systematic uncertainties beyond the background normalisation uncertainties. Note that $\delta^i_{bs} = 0$ is possible (meaning that the nuisance parameter θ_i has no influence in bin b for sample s). If for a given nuisance parameter θ_i, $\delta^i_{bs} \neq 0$ for multiple bins b or samples s, then the numbers δ^i_{bs} are describing correlated uncertainties among the bins or the samples, and are interpreted as the components of a variation vector of correlated uncertainties.

Given now the same set of N_b bins counting events after a certain selection, in which n_1, ..., n_{N_b} data events are observed, the likelihood of the observations is modelled as

$$\mathcal{L}(\mu, \boldsymbol{\theta} | n_1, \, ..., \, n_{N_b}) = \frac{1}{Z} \prod_{b \in \{\text{bins}\}} \text{Pois}(n_b | \nu_b(\mu, \boldsymbol{\theta})) \, p(\boldsymbol{\theta}), \tag{4.34}$$

where Z is a normalisation parameter that has no influence on the fit, and $\text{Pois}(n_b|\nu_b(\mu,\boldsymbol{\theta}))$ denotes the Poisson density function with expectation $\nu_b(\mu,\boldsymbol{\theta})$ evaluated at the point n_b, and $p(\boldsymbol{\theta})$ is the prior probability given to the nuisance parameters.

This prior probability term contains information on how the systematic uncertainties are modelled. It is the product of Gaussian densities centered at unity for the normalisation variations and at zero for the additive variations:

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{n} \text{Gauss}(\theta_i \,|\, 1, \, \sigma^2_{\text{norm},\, i}) \prod_{j=N-n+1}^{N} \text{Gauss}(\theta_j \,|\, 0, \, 1), \tag{4.35}$$

where $\text{Gauss}(x \,|\, m, \, \sigma^2)$ is the Gaussian density with expectation m and variance σ^2. The background normalisation uncertainties $\sigma_{\text{norm},\, i}$ are inputs of the model, similarly to the δ^i_{bs} factors of Equation (4.33). Note that the parameter of interest μ does not appear in Equation (4.35), because this parameter is unconstrained, or, in other words, it has a uniform prior distribution.

The signal strength μ is extracted from data by maximising the likelihood defined in Equation (4.34). The employed software package that implements this method and the statistical model is called pure-python HistFactory (`pyhf`) [105]. Since this tool is relatively new, a simplified Gaussian model (`sghf`) is implemented to validate the results from `pyhf`.

The simplified Gaussian model is identical to the model described by Equation (4.34), except that the Poisson density $\mathrm{Pois}(n_b|\nu_b(\mu, \boldsymbol{\theta}))$ in Equation (4.34) is replaced by a Gaussian density centered at $\nu_b(\mu, \boldsymbol{\theta})$ and with a standard deviation corresponding to the square root of the expected yield, $\sqrt{\nu_b^0}$, for each $b \in \{\text{bins}\}$.

4.5. Upper-limit determination

If no significant signal is observed, as it was the case for the previous searches for $B \to K\nu\bar{\nu}$ decays (see Section 2.3.4), a method called $\mathrm{CL_s}$ [106] is employed to determine an upper limit on the signal strength μ defined in Equation (4.29).

Given a likelihood model like the one defined in Equation (4.34) and a hypothesised value of μ, one defines the likelihood ratio [107]

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}} \,|\, n_1,\, ...,\, n_{N_b})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}} \,|\, n_1,\, ...,\, n_{N_b})}, \tag{4.36}$$

where $(\hat{\mu}, \hat{\boldsymbol{\theta}})$ are the parameters that maximise the likelihood for the set of observations n_1, ..., n_{N_b} (the same notations as in Section 4.4 are used here), and where $\hat{\hat{\boldsymbol{\theta}}}$ maximises the likelihood for a given value of μ.

The property $0 \leq \lambda(\mu) \leq 1$ follows from the definition above. When $\lambda(\mu)$ is close to unity, the data and the hypothesised value of μ are in good agreement.

The next step is to define a test statistic q_μ as [107]

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \text{if } \mu \geq \hat{\mu}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.37}$$

From this, the level of agreement between the data and the hypothesised value of μ is quantified with the p-value

$$p_{s+b} = P(q_\mu > q_{\mu,\mathrm{obs}} \,|\, \mu) = \int_{q_{\mu,\mathrm{obs}}}^{\infty} p(q_\mu|\mu)\, \mathrm{d}q_\mu, \tag{4.38}$$

where $q_{\mu,\mathrm{obs}}$ is the observed value of q_μ, and $p(q_\mu|\mu)$ denotes the probability density function of q_μ under the assumption of a signal strength of μ. An approximation of this density is computed in [107].

From Equation (4.38), it is already possible to determine an upper limit. For example, a 90% confidence level (CL) upper limit on μ is the largest value of μ such that p_{s+b} stays above 0.1.

In the $\mathrm{CL_s}$ method, one extra step is needed: a p-value for the background-only hypothesis is defined as

$$p_b = P(q_\mu > q_{\mu,\mathrm{obs}} \,|\, 0) = \int_{q_{\mu,\mathrm{obs}}}^{\infty} p(q_\mu|0)\,\mathrm{d}q_\mu, \tag{4.39}$$

where $p(q_\mu|0)$ denotes the probability density function of q_μ under the assumption of a signal strength of zero (background-only hypothesis). As before, an approximation of this density is computed in [107].

Finally, the ratio

$$\mathrm{CL_s} = \frac{p_{s+b}}{p_b} \tag{4.40}$$

is computed, and the 90% CL upper limit on μ is the largest value of μ such that $\mathrm{CL_s}$ stays above 0.1. An advantage of the $\mathrm{CL_s}$ method is that it gives more robust upper limits in situations where there is little information to distinguish between the background-only hypothesis and the signal+background hypothesis [106].

As for the likelihood maximisation, the `pyhf` package [105] is employed for the upper-limit determination.

4.6. Kernel density estimator

This section presents a tool called kernel density estimator, which is used in Section 5.10 in the study of the systematic uncertainties.

A kernel density estimator [108, 109] is a function that approximates the probability density of a random variable. This estimator has the particularity of being non-parametric, by contrast for example with the likelihood model defined in Equation (4.34), which depends on a potentially large number of (nuisance) parameters.

Given N observations x^1, x^2, ..., x^N of a one-dimensional random variable x with a variance $0 < \sigma^2 < \infty$, the gaussian-kernel density estimator is defined as

$$f_h(x) = \frac{1}{Z} \sum_{i=1}^{N} \exp\left[-\frac{1}{2}\left(\frac{x - x^i}{h\sigma}\right)^2\right], \tag{4.41}$$

where $Z > 0$ is a normalisation factor and $h > 0$ a smoothing factor.

This construction is easily generalised for multi-dimensional random variables. Given N observations $\mathbf{x}^1$, $\mathbf{x}^2$, ..., $\mathbf{x}^N$ of a multi-dimensional random variable $\mathbf{x}$ with an invertible covariance matrix $\mathbf{\Sigma}$, the gaussian-kernel density estimator is

$$f_h(\mathbf{x}) = \frac{1}{Z'} \sum_{i=1}^{N} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}^i)^T (h^{-2} \mathbf{\Sigma}^{-1}) (\mathbf{x} - \mathbf{x}^i) \right], \tag{4.42}$$

where $Z' > 0$ is a normalisation factor and $h > 0$ a smoothing factor.

The SciPy package [110] provides a software implementation of the gaussian-kernel density estimator.

5

Search for $B \to K\nu\bar{\nu}$ decays

The previous chapters prepared the ground for the main objective of this thesis, the search for $B \to K\nu\bar{\nu}$ decays. More specifically, the goal of this chapter is to measure two observables: the branching fraction of the $B^+ \to K^+\nu\bar{\nu}$ and the $B^0 \to K^0_S\nu\bar{\nu}$ modes, noted as $\mathrm{Br}(B^+ \to K^+\nu\bar{\nu})$ and $\mathrm{Br}(B^0 \to K^0_S\nu\bar{\nu})$, respectively[1]. If, at the end, the amount of observed signal is not large enough for a precise measurement of the branching fractions, upper limits are determined.

This chapter begins with the list of the input data and simulated samples (Section 5.1) and a presentation of the event selection strategy (Section 5.2). Sections 5.3 to 5.8 detail each step of the selection and are concluded by the definition of the signal search region. Section 5.9 compares data and simulation in different regions of the phase space to validate the selection procedure. Section 5.10 studies multiple sources of systematic uncertainties that affect the measurement of the branching fractions. Section 5.11 presents the likelihood model that is used to determine the branching fractions, and expected upper limits are derived from simulation. Sections 5.12 and 5.13 conclude this chapter by presenting and discussing the obtained results.

---

[1] Charge-conjugate modes are implied throughout this chapter.

5.1. Input samples

The search for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$ decays is conducted with a sample of data collected by Belle II between 2019 and 2021. Most of the data sample ($189\,\text{fb}^{-1}$) was collected at the energy of the $\Upsilon(4S)$ resonance. This sample is referred to as the on-resonance data. The remaining data ($18\,\text{fb}^{-1}$) were collected at an energy $60\,\text{MeV}$ below the $\Upsilon(4S)$ resonance and are referred to as the off-resonance data. The later sample does not contain decays of B mesons, because the energy is not sufficient to produce them, but this data sample is useful to study the continuum background ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$, and $e^+e^- \to \tau^+\tau^-$).

In addition, the following samples of simulated events are used:

- 8×10^6 simulated events containing a $B^+ \to K^+\nu\bar{\nu}$ decay.
- 8×10^6 simulated events containing a $B^0 \to K^0_S\nu\bar{\nu}$ decay.
- Simulated background events corresponding to $1000\,\text{fb}^{-1}$ of equivalent integrated luminosity. Seven dominant background categories are simulated: the non-signal B meson decays ($e^+e^- \to B^+B^-$ and $e^+e^- \to B^0\overline{B}^0$), and five continuum contributions ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$, and $e^+e^- \to \tau^+\tau^-$).

The simulated samples are taken from the official Belle II simulation production, that made use of the software libraries listed in Section 3.2.8. Table 5.1 gives an overview of the data and simulated samples.

5.2. Event selection strategy

As already detailed in Section 2.3.4, the previous searches for $B \to K\nu\bar{\nu}$ decays were using a hadronic tagging method, a semileptonic tagging method, or a combination of both. A disadvantage of these methods is their small signal selection efficiency, well below 1%, caused by the explicit reconstruction of the accompanying B meson in the $e^+e^- \to B\overline{B}$ event.

In the following, another method is developed and followed: the *inclusive tagging*. In this method, the accompanying B meson is not explicitly reconstructed. Instead, the event selection relies on a set of variables that distinguish an event containing a $B \to K\nu\bar{\nu}$ decay from the more common $e^+e^- \to B\overline{B}$, $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$), and $e^+e^- \to \tau^+\tau^-$ events. A similar method that inspires this work was recently developed by the Belle collaboration in a search for $B^+ \to \mu^+\nu_\mu$ decays [111].

Table 5.1.: Data and simulated samples, and corresponding integrated luminosity (L) or number of events (# events), and energy in the centre-of-mass frame ($\sqrt{s}$).

Data sample	L [fb^{-1}]	$\sqrt{s}$
On-resonance	189	$\mathrm{M}(\Upsilon(4S))$
Off-resonance	18	$\mathrm{M}(\Upsilon(4S)) - 60\,\mathrm{MeV}/c^2$

Simulated background	L [fb^{-1}]	$\sqrt{s}$
$e^+e^- \to B^+B^-$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to B^0\overline{B}^0$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to u\overline{u}$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to d\overline{d}$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to c\overline{c}$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to s\overline{s}$	1000	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to \tau^+\tau^-$	1000	$\mathrm{M}(\Upsilon(4S))$

Simulated signal	# events	$\sqrt{s}$
$e^+e^- \to B^+(\to K^+\nu\bar{\nu})B^-$	8×10^6	$\mathrm{M}(\Upsilon(4S))$
$e^+e^- \to B^0(\to K^0_S\nu\bar{\nu})\overline{B}^0$	8×10^6	$\mathrm{M}(\Upsilon(4S))$

The steps of the event selection method are enumerated below and each is linked to a corresponding section:

1. Particle candidate lists (Section 5.3): lists of particle candidates are defined in each event, based on the information collected by the Belle II subdetectors.
2. Signal candidate selection (Section 5.4): in each event, one signal kaon candidate is chosen.
3. Basic event selection (Section 5.5): simple criteria are applied to reject background events.
4. Input variable computation (Section 5.6): a set of discriminative variables are defined and computed for each event.
5. Binary classification (Section 5.7): the variables defined in the previous step are computed on simulated events, and used to train binary classifiers that are employed to finalise the event selection.

Table 5.2.: Fraction of charged particles expected from B meson decays. The numbers are derived from simulated $e^+e^- \to B\overline{B}$ events [81].

Charged particle	Fraction [%]
$\pi^\pm$	72.8
$K^\pm$	14.9
$e^\pm$	5.8
$\mu^\pm$	4.7
$p^\pm$	1.8

5.3. Particle candidate lists

This section defines general lists of particle candidates reconstructed in each event. The charged particle candidates are all the reconstructed tracks[2] satisfying the following conditions:

- The transverse momentum p_T of the track is between 0.1 and 5.5 GeV/c. The lower threshold suppresses the impact from the beam background and the upper threshold is a security threshold suppressing potentially mis-reconstructed tracks.
- The transverse impact parameter dr of the track with respect to the average interaction point (IP) is smaller than 0.5 cm and the longitudinal impact parameter $|dz|$ of the track with respect to the IP is smaller than 3 cm (see Figure 3.3 in Section 3.2 for a definition of the Belle II coordinate system). This condition filters out charged particles that come from a region away from the IP.
- The polar angle θ of the track is within the CDC acceptance. This condition suppresses charged particles due to the beam background whose trajectory is close to the beam line and that are only detected by the VXD.

At Belle II, tracks are mainly produced by pions, kaons, electrons, muons, and protons. The likelihood of each mass hypothesis is computed by combining PID information from the Belle II subdetectors. Each track is associated with its most-likely mass hypothesis, with a prior probability derived from simulated $e^+e^- \to B\overline{B}$ events (Table 5.2).

The photon candidates are built from ECL clusters fulfilling the following requirements:

---

[2] Information about how tracks are reconstructed at Belle II can be found in [81].

- The cluster energy E is between 0.1 and 5.5 GeV. Similarly to the p_{T} requirement for the tracks, the lower threshold suppresses the impact from the beam background, and the upper threshold is a security threshold suppressing potentially mis-reconstructed ECL clusters.
- The cluster is within the CDC acceptance. This condition suppresses clusters produced by charged particles potentially missed by the tracking system.

The K_{S}^0 candidates are reconstructed from two tracks of opposite charges that are fit to a common vertex, meaning that only the decay $K_{\mathrm{S}}^0 \to \pi^+\pi^-$, which has a branching fraction of approximately 70% [25], is reconstructed. The fully-neutral decay $K_{\mathrm{S}}^0 \to \pi^0\pi^0 \to 4\gamma$, which has a branching fraction of approximately 30% [25], is not considered, because of a large number of potential wrong combinations of photon candidates and the presence of neutral hadrons in the photon candidate list (mis-identified photons). The following selection is applied on the reconstructed K_{S}^0 candidates:

- As for the charged particle candidates, each track satisfies $0.1 < p_T < 5.5\,\mathrm{GeV}/c$ and $\theta \in$ CDC acceptance.
- The mass of the K_{S}^0 candidate is in the interval $[0.485, 0.510]\,\mathrm{GeV}/c^2$ defined around the nominal K_{S}^0 mass, which is $0.498\,\mathrm{GeV}/c^2$ [25].
- The cosine of the angle between the momentum of the K_{S}^0 candidate and the vector from the average interaction point to the K_{S}^0 candidate vertex is greater than 0.98. This requirement ensures that the K_{S}^0 candidate comes from a region close to the average interaction point.

No explicit list of K_{L}^0 candidates is defined, but since the requirements on the photon candidates are loose, it is expected that the K_{L}^0 contribution to the visible energy is taken into account for K_{L}^0 producing ECL clusters.

5.4. Signal candidate selection

In a $B \to K\nu\bar{\nu}$ decay, the kaon is the only particle that can be detected, since the two neutrinos escape the detector without any interaction. An important variable that is used during the selection of the signal candidate is the invariant mass squared of the two-neutrino system, noted q^2, which was already introduced when computing the branching fraction of the $B \to K\nu\bar{\nu}$ decay in Section 2.3.2.

If $\mathbf{P}_B^* \equiv (E_B^*, \mathbf{p}_B^*)$ and $\mathbf{P}_K^* \equiv (E_K^*, \mathbf{p}_K^*)$ are the 4-momenta of the B meson and the kaon in the centre-of-mass system, then the invariant mass squared of the two-neutrino system

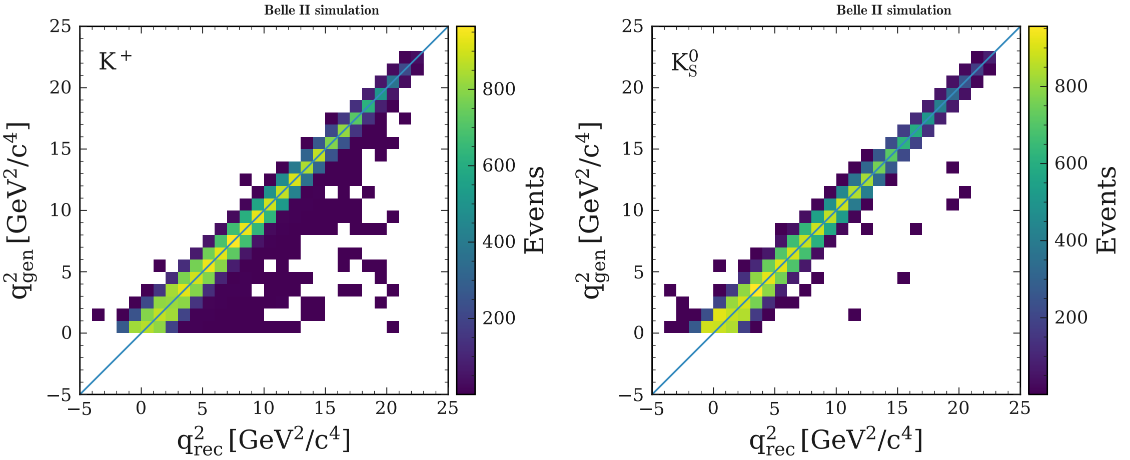


Figure 5.1.: Generated (q^2_{gen}) and reconstructed (q^2_{rec}) invariant mass squared of the two-neutrino system of 10^5 simulated $B^+ \to K^+\nu\bar{\nu}$ decays (left) and $B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$ decays (right).

is

$$q^2 = (\mathbf{P}^*_B - \mathbf{P}^*_K)^2 = m^2_B + m^2_K - 2\,E^*_B E^*_K + 2\,\mathbf{p}^*_B \cdot \mathbf{p}^*_K, \tag{5.1}$$

where m_B and m_K are the mass of the B meson and the kaon, respectively. Experimentally, the 4-momentum of the B meson is not accessible. For this reason, an approximated q^2, called reconstructed q^2, and noted q^2_{rec}, is defined as

$$q^2_{\mathrm{rec}} = m^2_B + m^2_K - 2\,m_B E^*_K. \tag{5.2}$$

Equation (5.2) is equivalent to Equation (5.1) if the 3-momentum of the B meson in the centre-of-mass frame is neglected. This is a valid approximation since the B mesons are nearly at rest in this frame, as already mentioned in Section 3.1.

Figure 5.1 presents the correlation between q^2 and q^2_{rec} in 10^5 simulated $B \to K\nu\bar{\nu}$ decays, and shows that q^2_{rec} approximates q^2 with a resolution of the order of $1\,\mathrm{GeV}^2/\mathrm{c}^4$. The tail observed in the $B^+ \to K^+\nu\bar{\nu}$ case that is absent in $B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$ is due to radiative $B^+ \to K^+\nu\bar{\nu}(\gamma)$ events causing an underestimation of E^*_K.

Now that this variable is defined, the signal candidate selection is done as follows.

Signal K^+ candidate

For the charged mode ($B^+ \to K^+\nu\bar{\nu}$), the signal K^+ candidate is a track that satisfies the criteria for charged particle candidates listed in Section 5.3. In addition, the following conditions are imposed:

- The track contains at least one PXD hit. This ensures an optimal resolution on the candidate trajectory.
- The track contains at least 20 CDC hits. This requirement is recommended by the Belle II performance group to ensure that enough information is available for the particle identification, which is, in particular, a function of the energy loss dE/dX within the CDC, where E is the energy of the particle candidate, and X is a distance measured along the trajectory of the particle candidate.
- The candidate has a kaon-hypothesis likelihood (PID score) of at least 0.9. This requirement retains approximately 60% of true kaons and rejects approximately 95% of mis-identified kaons (mainly pions).
- Finally, in each event, all the candidates are ranked according to q^2_{rec}, and the one with the smallest q^2_{rec} is chosen as the signal candidate. This choice is motivated by the signal simulation, that shows that if a list of multiple candidates contains the true signal K^+, then the true signal K^+ has the smallest q^2_{rec} in more than 90% of the cases.

Figure 5.2 (left) shows simulated $B^+ \to K^+\nu\bar{\nu}$ and background events in bins of q^2_{rec}. In the background histogram, seven background categories are stacked: the non-signal B meson decays ($e^+e^- \to B^+B^-$ and $e^+e^- \to B^0\overline{B}^0$), and the five continuum contributions ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$, and $e^+e^- \to \tau^+\tau^-$).

In Figure 5.2, a correction weight is applied to the simulated signal events in order to obtain a realistic q^2-dependence of the signal branching fraction. This correction is necessary, because simulation does not take into account the q^2-dependence of the form factor entering in the computation of the $B \to K\nu\bar{\nu}$ branching fraction (Section 2.3.2). The correction weight is computed as the ratio between the two distributions shown in Figure 2.5 (Section 2.3.2). Unless specified otherwise, this correction weight is applied to all the simulated signal events in the following.

Signal K^0_{S} candidate

For the neutral mode ($B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$), the signal K^0_{S} candidate is taken from the list of K^0_{S} candidates defined in Section 5.3. Similarly to the charged mode, the signal K^0_{S} candidate with the smallest q^2_{rec} is chosen in each event. Signal simulation shows that if a list of multiple candidates contains the true signal K^0_{S}, then the true signal K^0_{S} has the smallest q^2_{rec} in more than 90% of the cases.

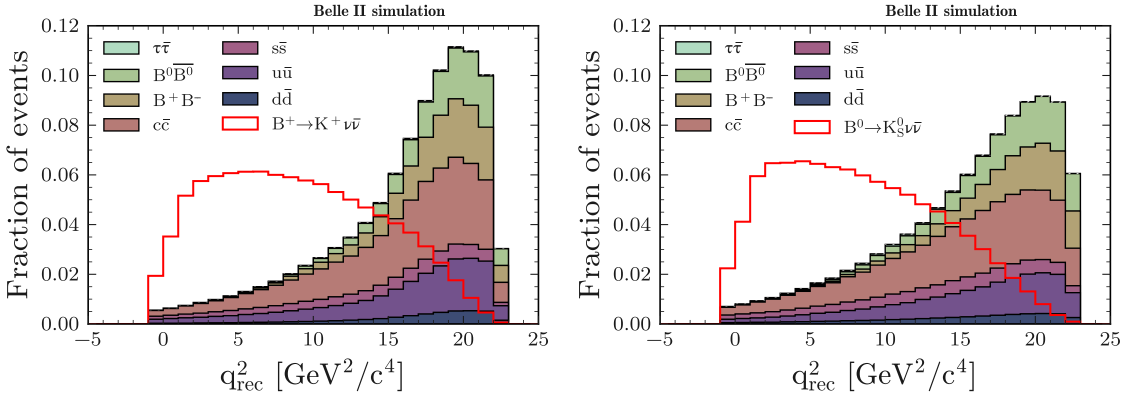


Figure 5.2.: Simulated signal and background events in bins of the reconstructed invariant mass squared of the two-neutrino system for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

Figure 5.2 (right) shows simulated $B^0 \to K^0_S\nu\bar{\nu}$ and background events in bins of q^2_{rec}, and Figure 5.3 (top left) shows that the mass of the selected signal K^0_S candidates is peaking at the nominal K^0_S mass, which is 0.498 GeV/c^2 [25].

5.5. Basic event selection

Now that a single signal kaon candidate is selected in each event, a set of basic requirements, which are listed below, are imposed on each event. Signal and background distributions of the variables mentioned in this section are shown in Figure 5.3.

- An event containing a $B \to K\nu\bar{\nu}$ decay tends to have a smaller number of tracks than a background event, because the decay of a signal B meson produces only one track in the $B^+ \to K^+\nu\bar{\nu}$ case, and two tracks in the $B^0 \to K^0_S\nu\bar{\nu}$ case. For this reason, an event must have at most 10 (11) charged particle candidates in total when selecting $B^+ \to K^+\nu\bar{\nu}$ ($B^0 \to K^0_S\nu\bar{\nu}$) events. This requirement causes a small signal loss, of the order 0.4%, while suppressing background events that have a large number of tracks.
- In addition, an event must have at least four charged particle candidates in total. The reason for this requirement is that at least one track is missing in an event with three tracks, because of charge conservation, and the remaining information in such

an event is too limited. The same argument of information limitation applies to events with less than three tracks.

- The visible energy in an event in the centre-of-mass frame must be greater than 4 GeV. This requirement is necessary to suppress non-simulated low-multiplicity background. An example of such background is the process $\mathrm{e}^+e^- \to \mathrm{e}^+e^-\pi^+\pi^-\pi^+\pi^-$, mediated by a pair of virtual photons, where the electron and the positron in the final state are outside of the detector acceptance, because their trajectories are too close to the beam line (see [112, 113, 114] for experimental results regarding this class of processes).
- The polar angle θ of the missing momentum in an event must be within the CDC acceptance (see Equation (4.20) for a definition of the missing momentum). This requirement also suppresses the low-multiplicity background and more generally ensures that the missing momentum is not due to particles travelling outside of the detector acceptance.
- Finally, only events whose signal kaon candidate is such that $q^2_{\mathrm{rec}} \geq -1\,\mathrm{GeV}^2/\mathrm{c}^4$ are retained, because the events with $q^2_{\mathrm{rec}} < -1\,\mathrm{GeV}^2/\mathrm{c}^4$ are mainly background events (Figures 5.1 and 5.2).

Table 5.3 summarises the evolution of the signal selection efficiency when applying the criteria listed in Sections 5.3 to 5.5. For the charged mode ($B^+ \to K^+\nu\bar{\nu}$), the drop of the signal selection efficiency from 89.2% to 52.1% after the definition of the signal candidate list is mainly due to the PID requirement used to reject the pion background (Section 5.4). After the first step of the selection (Section 5.3), the signal selection efficiency is smaller for the neutral mode (43.3%) than for the charged mode (89.2%), because only the channel $K^0_{\mathrm{S}} \to \pi^+\pi^-$ is considered to define K^0_{S} candidates, and both pions need to be in the detector acceptance.

At this stage of the event selection, background largely dominates, and the role of the next section is to define a set of variables that will be used in binary classification algorithms for background suppression.

Table 5.3.: Evolution of the signal selection efficiency for the charged mode ($B^+ \to K^+\nu\bar{\nu}$) and the neutral mode ($B^0 \to K^0_S\nu\bar{\nu}$) when applying the criteria listed in Sections 5.3 to 5.5. The efficiency is computed by reconstructing a total of 10^5 simulated signal events.

Selection stage	$B^+ \to K^+\nu\bar{\nu}$ [%]	$B^0 \to K^0_S\nu\bar{\nu}$ [%]
Particle candidate lists (Section 5.3)	89.2	43.3
Signal candidate list (Section 5.4)	52.1	43.3
Best signal candidate selection (Section 5.4)	50.2	42.7
Basic event selection (Section 5.5)	41.3	37.3

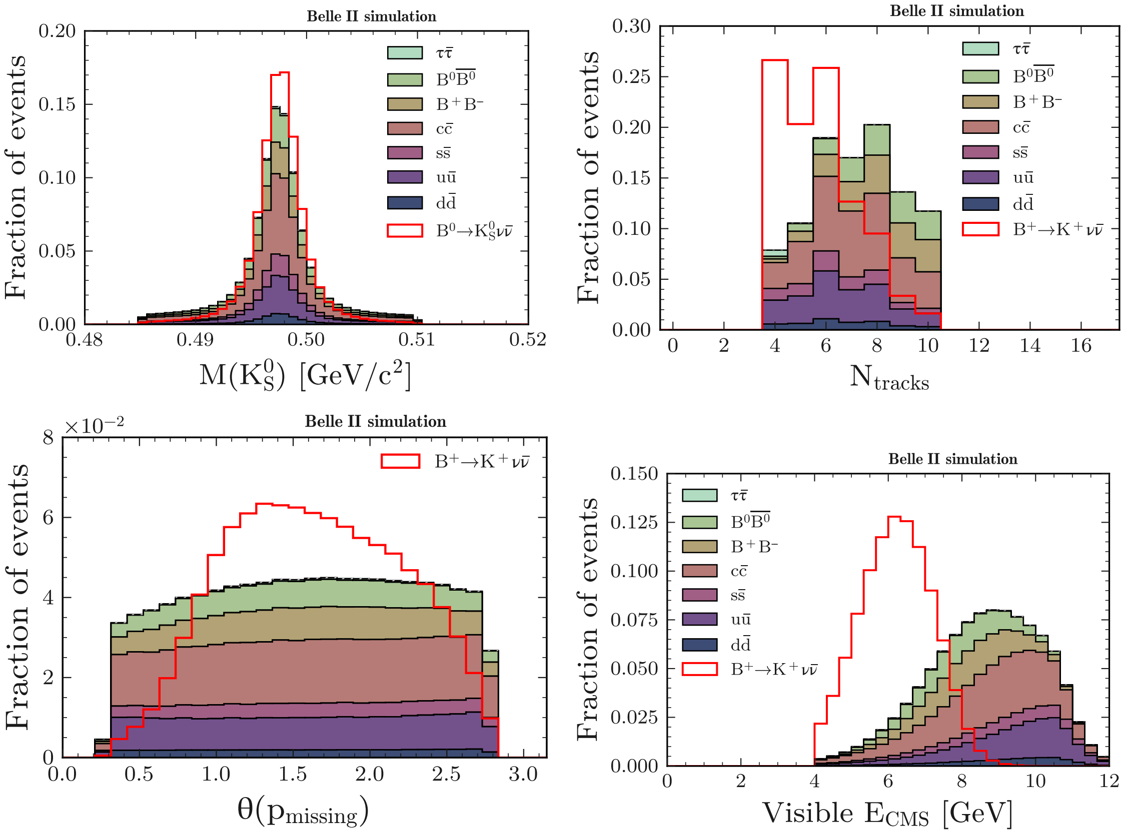


Figure 5.3.: Simulated signal and background events in bins of the mass of the signal K_S^0 candidate (top left), the number of charged candidates in the event (top right), the polar angle of the missing momentum (bottom left), and the visible energy in the centre-of-mass frame (bottom right). The top-left plot corresponds to the neutral mode ($B^0 \to K_S^0 \nu\bar{\nu}$), the other three plots correspond to the charged mode ($B^+ \to K^+ \nu\bar{\nu}$). The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

5.6. Input variables

After the basic event selection presented in Section 5.5, a set of variables are computed for each event. These variables are used to classify events in two categories: signal or background. This section gives a list of variables that are candidates to serve as discriminative variables used during the event classification. The classification itself, and the set of variables that are retained for the final selection are discussed later.

The variables are organised in four categories and defined in dedicated subsections.

- Variables related to all the particle candidates in an event (Section 5.6.1).
- Variables related to the properties of the signal kaon candidate (Section 5.6.2).
- Variables related to the rest of the event (ROE), which refers to all the particle candidates in an event that are not the signal kaon candidate (Section 5.6.3).
- Variables designed to suppress the background from D meson decays (Section 5.6.4).

The reference frame is the laboratory, unless specified otherwise. Some variables are computed in the centre-of-mass system (CMS).

Multiple overlays of signal and background distributions are shown in this section. Similarly to Figures 5.2 and 5.3, the histograms are obtained by selecting with the criteria listed in Sections 5.3 to 5.5 a total of approximately 10^6 simulated signal events and 10^6 simulated background events.

Appendix A.1 contains summary tables with the list of all variables defined in this section and complementary figures.

5.6.1. Entire event

The variables listed below are defined as a function of the momentum distribution in an event, or as a function of observables that depend on the entire event (signal+ROE). The definition of most of these variables is given in Section 4.2. Signal and background distributions are shown in Figures 5.5 and 5.6 for the $B^+ \to K^+\nu\bar{\nu}$ mode. They are similar for the $B^0 \to K^0_S\nu\bar{\nu}$ mode (Appendix A.1).

The variables shown in Figure 5.5 are (from left to right and top to bottom):

- The normalised Fox-Wolfram moments R_1, R_2, and R_3, defined in Section 4.2.3, and computed in the CMS.

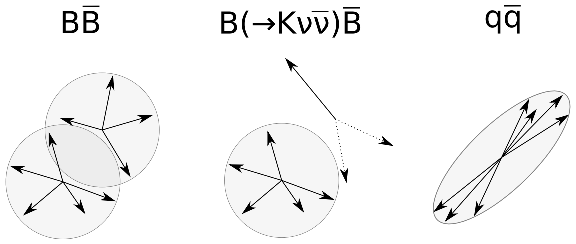


Figure 5.4.: Momentum distribution of a typical $e^+e^- \to B\overline{B}$ event, which tends to be spherical (left), an $e^+e^- \to B(\to K\nu\overline{\nu})\overline{B}$ event (middle), and an $e^+e^- \to q\overline{q}$ event, which tends to be jet-like (right).

- The signal-ROE modified Fox-Wolfram moments $H^{so}_{c,2}$, $H^{so}_{n,2}$, $H^{so}_{m,0}$, $H^{so}_{m,2}$, and $H^{so}_{m,4}$, defined in Section 4.2.5, and computed in the CMS. The moments that depend on the missing momentum (i.e. those with a m subscript) are particularly discriminative, because a signal event has a large missing momentum due to the two undetected neutrinos.
- The harmonic moments B_0 and B_2, defined in Section 4.2.4, with respect to the thrust axis in the event, defined in Section 4.2.2, and computed in the CMS. The B_0 moment, which is a normalised sum of the momenta of the detected particles, is smaller for signal events because of the two undetected neutrinos.

In the following, the tracks in the ROE are fit to a common vertex called the tag vertex. In addition, for a given track T and a given point P, the point of closest approach of T with respect to P, noted POCA, is defined as the point on the track T whose distance to the point P is minimal in the transverse plane [81].

The variables shown in Figure 5.6 are (from left to right and top to bottom):

- The radial and the longitudinal distances between the POCA of the K^+ candidate track and the tag vertex, noted $dr(K^+, \text{Tag Vertex})$ and $dz(K^+, \text{Tag Vertex})$, respectively.
- The event sphericity, defined in Section 4.2.1, and computed in the CMS. The momentum distribution of a typical $e^+e^- \to B\overline{B}$ event tends to be spherical, and the momentum distribution of an $e^+e^- \to q\overline{q}$ event tends to be jet-like (Figure 5.4).
- The missing mass squared in the event ${M_{\text{missing}}}^2$, defined as

  $${M_{\text{missing}}}^2 = E^2_{\text{missing}} - p^2_{\text{missing}}, \tag{5.3}$$

  where E_{missing} and p_{missing} are the missing energy and the missing momentum in the event, respectively, defined in Section 4.2.5, and computed in the CMS.

- The polar angle of the missing momentum, noted in $\theta(p_{\text{missing}})$, and computed in the CMS.
- The square of the sum of the electric charges in the event.
- The number of tracks N_{tracks}, of photons N_γ, and of leptons N_{lepton} in the event that pass the criteria listed in Section 5.3. Lepton candidates are obtained by selecting tracks with an electron-hypothesis likelihood or a muon-hypothesis likelihood (PID score) of at least 0.9.
- The magnitude of the event thrust, defined in Section 4.2.2, and computed in the CMS.
- The cosine of the polar angle of the event thrust axis, noted $\cos(\theta(\text{thrust}))$, and computed in the CMS.
- The cosine of the angle between the momentum line of the signal kaon candidate and the thrust axis of the ROE, noted $\cos(\text{thrust}_B, \text{thrust}_{\text{ROE}})$, and computed in the CMS. In a signal event, the momentum of the signal kaon is not correlated to the momentum of the ROE particles, implying that the distribution of this variable is approximately uniform for signal events (bottom right of Figure 5.6).

5.6.2. Signal kaon candidate

The variables listed below depend on the properties of the signal kaon candidate. Signal and background distributions are shown in Figure 5.7 for the $B^+ \to K^+\nu\bar{\nu}$ mode and the $B^0 \to K^0_S\nu\bar{\nu}$ mode, some variables being specific to a certain mode.

In the following, the average interaction point is noted IP.

The variables shown in Figure 5.7 are (from left to right and top to bottom):

- The radial and the longitudinal distances between the POCA of the K^+ candidate track and the IP, noted $dr(K^+)$ and $dz(K^+)$, respectively.
- The azimuthal angle of the K^+ candidate momentum at the POCA with respect to the IP, noted $\phi(K^+)$.
- The mass of the K^0_S candidate, noted $M(K^0_S)$.
- The cosine of the angle between the K^0_S candidate momentum line and the line from the IP to the K^0_S candidate vertex, noted $\cos(p_{K^0_S}, \text{line}(\text{IP}, K^0_S\,\text{vertex}))$.

- The cosine of the angle between the thrust axis of the signal kaon candidate and the z axis, noted $\cos(\mathrm{thrust}_B,\, z)$.

- The radial and the longitudinal distances between the K_S^0 candidate momentum line and the IP, noted $dr(p_{K_S^0})$ and $dz(p_{K_S^0})$, respectively.

5.6.3. Rest of the event

The variables listed below are functions of the properties of the ROE. Signal and background distributions are shown in Figure 5.8 for the $B^+ \to K^+\nu\bar{\nu}$ mode. They are similar for the $B^0 \to K_S^0\nu\bar{\nu}$ mode (Appendix A.1).

The variables shown in Figure 5.8 are (from left to right and top to bottom):

- The difference between the ROE energy in the CMS and the energy of one beam in the CMS ($\sqrt{s}/2$), noted ΔE_{ROE}. This variable is one of the most discriminative ones. The ROE with respect to a mis-identified signal kaon typically has an energy larger than $\sqrt{s}/2$ (i.e. $\Delta E_{\mathrm{ROE}} > 0\,\mathrm{GeV}$), because in this case, the ROE is a combination of decay products of both B mesons in an $e^+e^- \to B\overline{B}$ event. A similar argument holds for continuum background events ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$, and $e^+e^- \to \tau^+\tau^-$).

- The ROE-ROE modified Fox-Wolfram moments R_0^{oo} and R_2^{oo}, defined in Section 4.2.5, and computed in the CMS.

- The invariant mass of the ROE, noted $M(\mathrm{ROE})$.

- The polar angle of the ROE momentum, noted $\theta(p_{\mathrm{ROE}})$.

- The magnitude of the ROE momentum, noted p_{ROE}.

- The magnitude of the ROE thrust, noted $\mathrm{thrust}_{\mathrm{ROE}}$, and computed in the CMS.

- The quality of the tag vertex fit, quantified with a p-value, noted p-value(Tag Vertex). The p-value is close to unity when the quality of the vertex fit is good.

- The components of the vector from the IP to the tag vertex, noted $dx(\mathrm{Tag\,Vertex})$, $dy(\mathrm{Tag\,Vertex})$ and $dz(\mathrm{Tag\,Vertex})$. The Belle II coordinate system is defined in Section 3.2.

- The variance of the transverse momentum of the ROE tracks, noted $\mathrm{Variance}_{\mathrm{ROE}}(p_T)$.

5.6.4. D meson suppression

The variables defined below serve to suppress the background coming from D mesons decaying into a kaon and one or two pions, and where the kaon is selected as the signal kaon candidate. In order to suppress this background, D meson candidates are defined by combining the signal kaon candidate with tracks in the ROE, and by fitting the tracks to a common vertex. The D meson candidates are ranked according to the quality of the vertex fit, and the best D meson candidate refers to the one with the best vertex fit.

For the $B^+ \to K^+\nu\bar{\nu}$ mode, D^0 candidates are obtained by combining the signal K^+ candidate with each track of opposite charge in the ROE, and D^+ candidates are obtained by combining the signal K^+ candidate with two tracks of opposite charge in the ROE.

For the $B^0 \to K^0_S\nu\bar{\nu}$ mode, D^+ candidates are obtained by combining the signal K^0_S candidate with each track in the ROE, and D^0 candidates are obtained by combining the signal K^0_S candidate with two tracks of opposite charges in the ROE.

Signal and background distributions are shown in Figure 5.9 for the $B^+ \to K^+\nu\bar{\nu}$ mode. Figures for the $B^0 \to K^0_S\nu\bar{\nu}$ mode are shown in Appendix A.1.

The variables shown in Figure 5.9 are (from left to right and top to bottom):

- The radial and the longitudinal distances between the best D^+ candidate vertex and the IP, noted $dr(D^+)$ and $dr(D^+)$, respectively.
- The quality of the vertex fit of the best D^+ candidate, quantified with a p-value, noted p-value(D^+).
- The radial and the longitudinal distances between the best D^0 candidate vertex and the IP, noted $dr(D^0)$ and $dr(D^0)$, respectively.
- The quality of the vertex fit of the best D^0 candidate, quantified with a p-value, noted p-value(D^0).
- The median of the fit quality of the D^0 candidate vertices, noted Median(p-value(D^0)).
- The mass of the best D^0 candidate, noted $M(D^0)$. The distribution of this variable for background events has a peak close to $1.87\,\mathrm{GeV}/c^2$, which is the nominal mass of the D^0 meson [25] (bottom right of Figure 5.9). The other peaks in the distribution at lower masses are likely due to the $K^{*0}(892)$ and $\phi(1020)$ resonances [25].

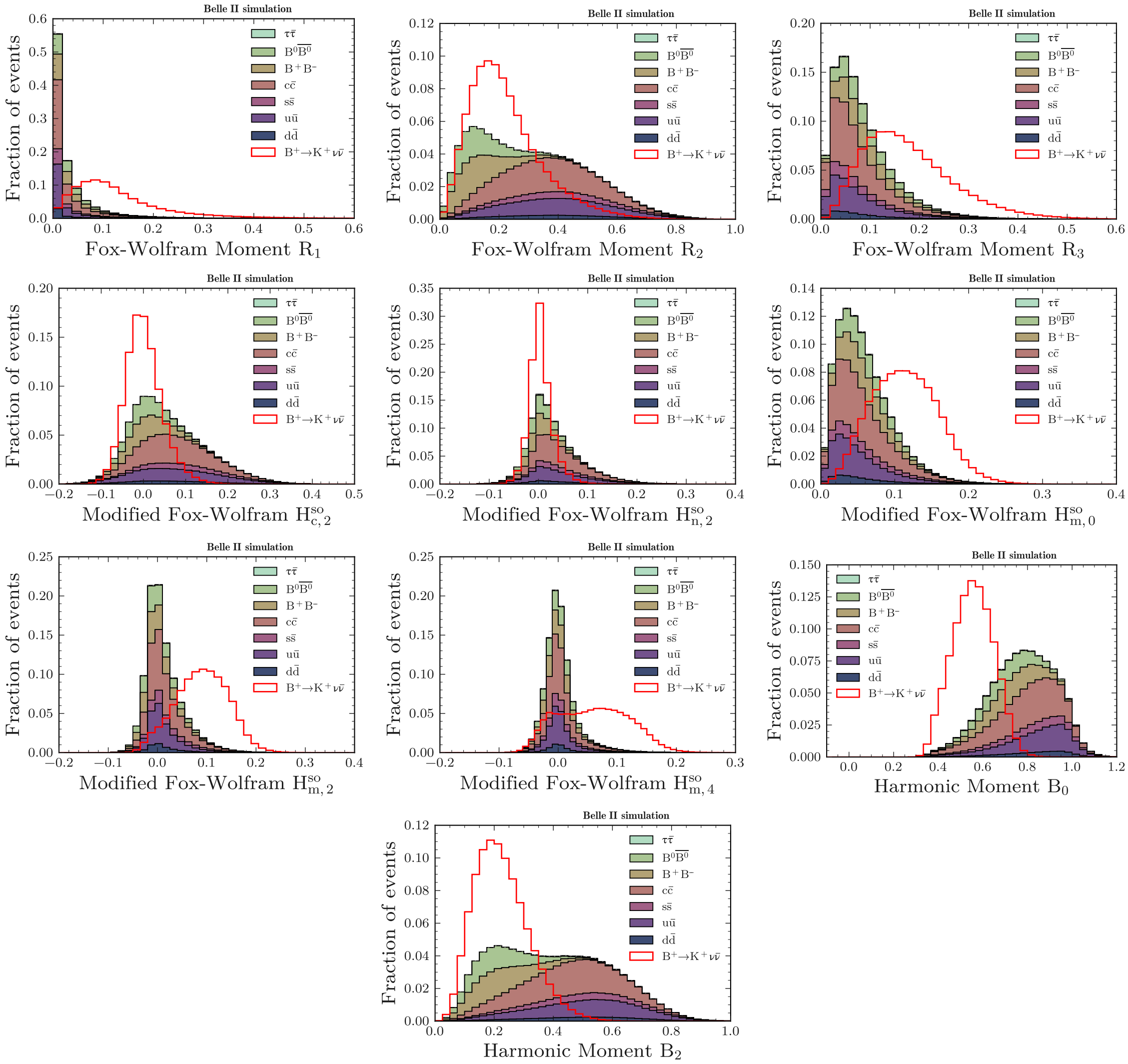


Figure 5.5.: Simulated signal and background events in bins of variables related to the momentum distribution in the entire event for the $B^+ \to K^+\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

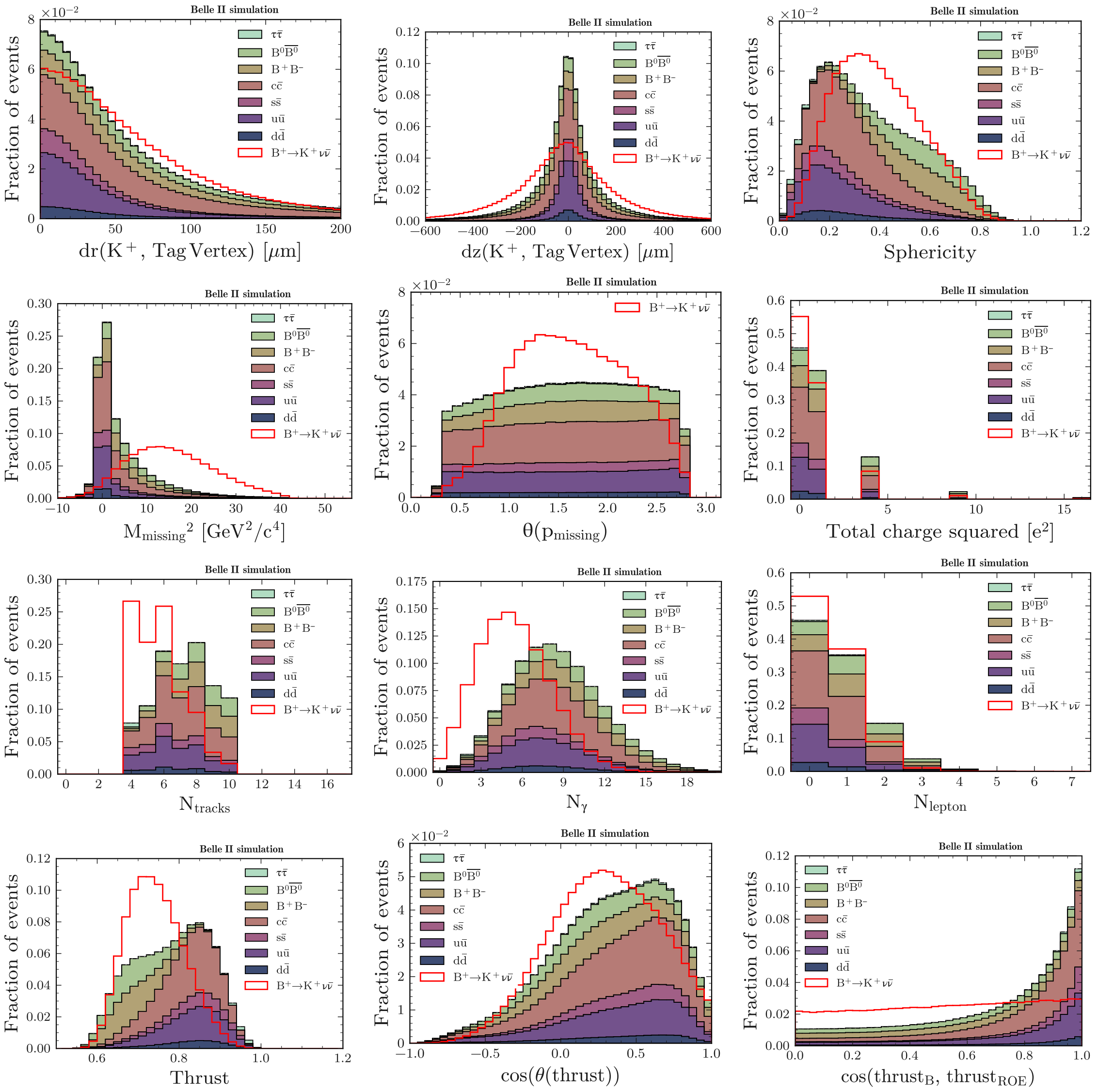


Figure 5.6.: Simulated signal and background events in bins of variables related to the entire event for the $B^+ \to K^+\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

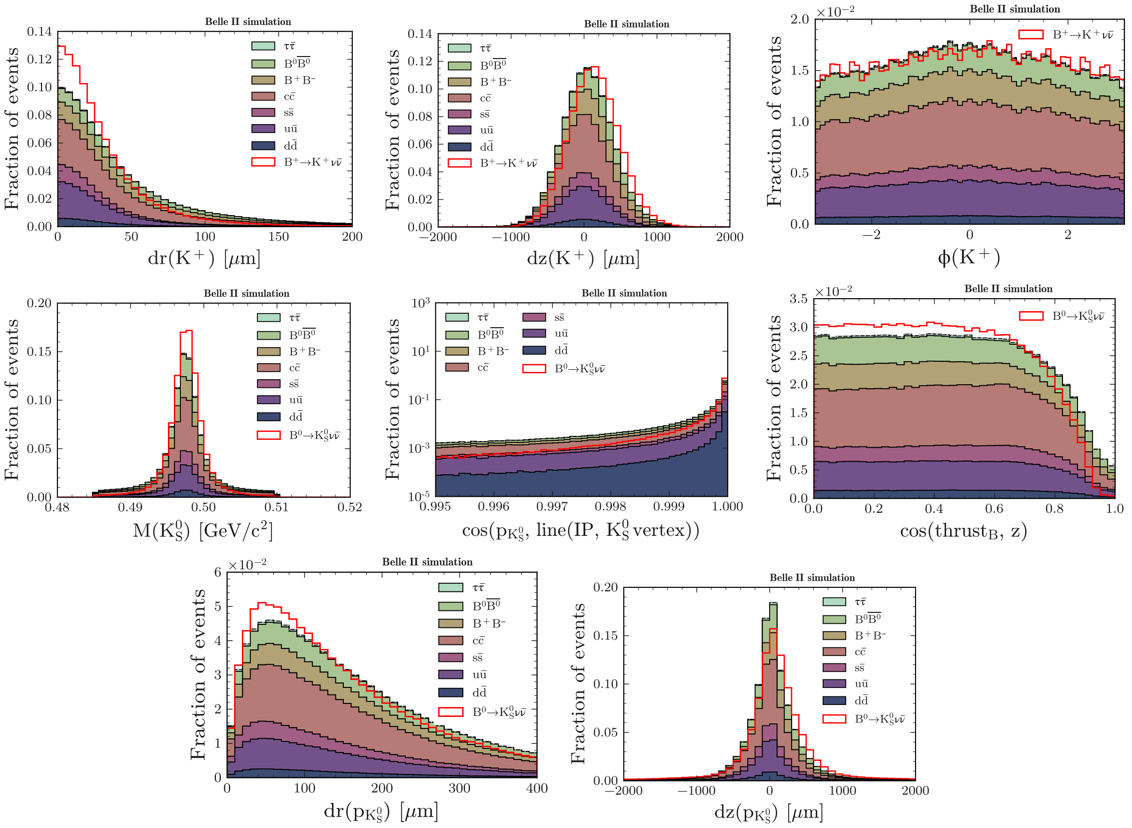


Figure 5.7.: Simulated signal and background events in bins of variables related to the properties of the signal kaon candidate. The first line of plots corresponds to the $B^+ \to K^+\nu\bar{\nu}$ mode, the second and the third lines of plots correspond to the $B^0 \to K^0_S\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

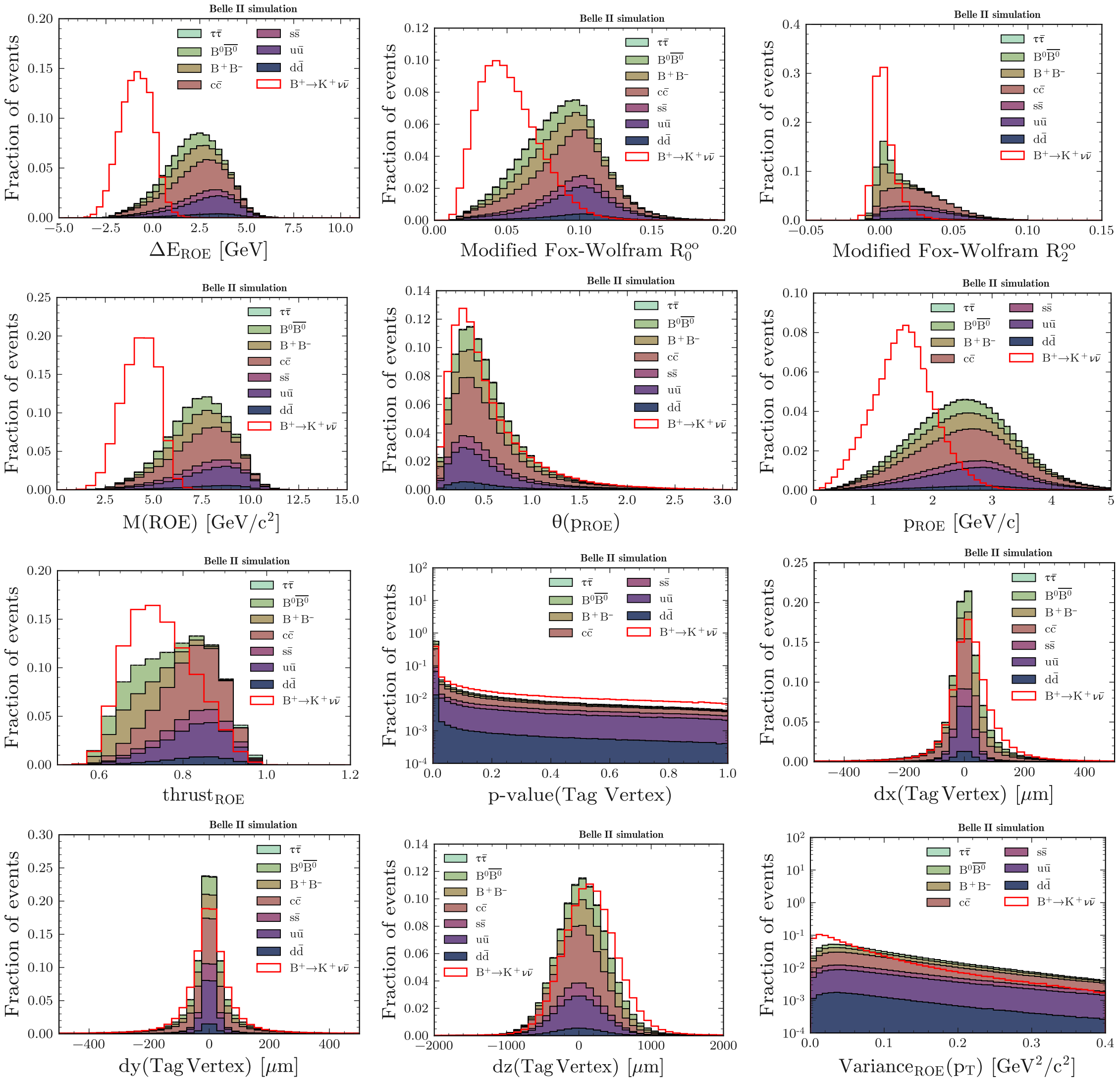


Figure 5.8.: Simulated signal and background events in bins of variables related to the ROE for the $B^+ \to K^+\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

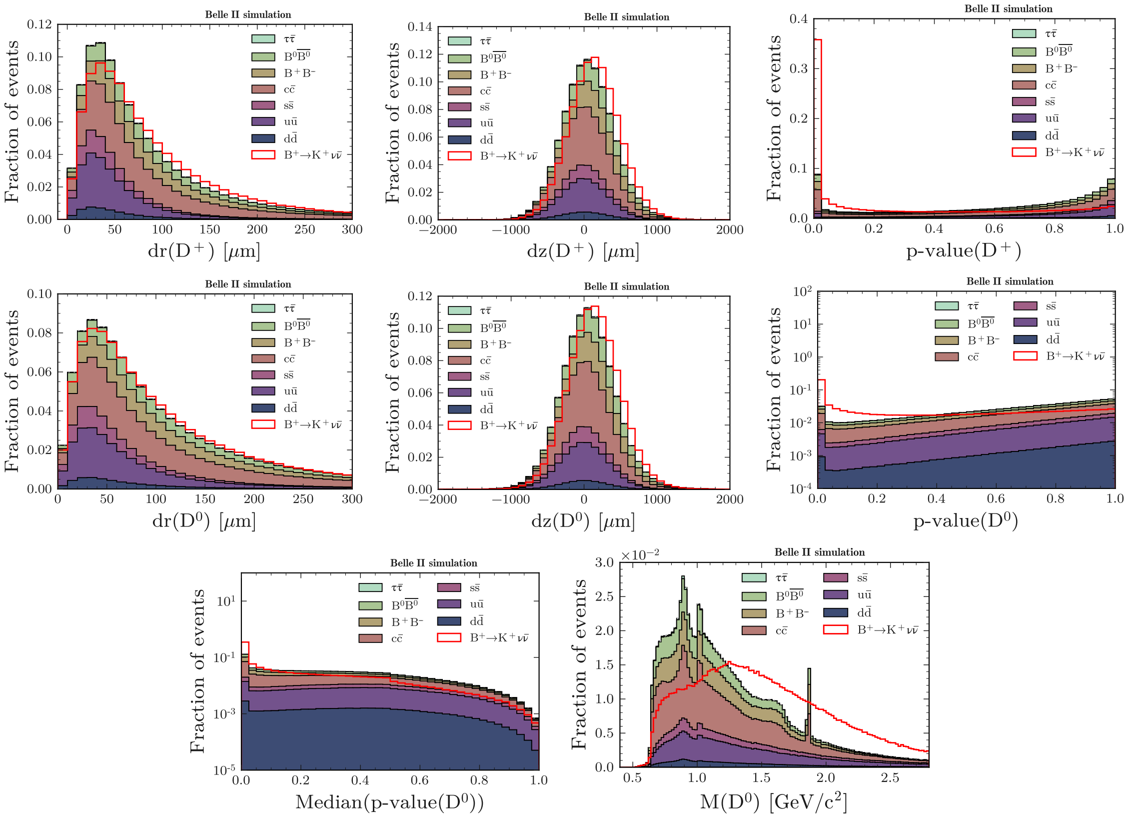


Figure 5.9.: Simulated signal and background events in bins of variables related to the D meson suppression for the $B^+ \to K^+\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain.

5.7. Binary classification

This section presents the central part of the event selection: binary classification. The definition of binary classification and of the algorithms that are used to achieve it are detailed in Section 4.1.

In order to separate signal and background events, two boosted decision trees, BDT_1 and BDT_2, are trained for each mode ($B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$):

- BDT_1 serves as a first layer of background suppression. It uses a limited set of discriminative variables, common for both $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$ modes, to filter out the most obvious background.
- BDT_2 is trained on events that pass the first layer of background suppression and uses additional discriminative variables, including mode-specific variables, to reject background events that are similar to signal events. The signal search region is defined from the output of BDT_2 (Section 5.8).

The reason why two classification layers are needed is the limited memory size of the computers on which the classifiers are trained. As an illustration, if one background event out of $L \gg 1$ is similar to a signal event, and if $M \gg 1$ such background events are required for a proper training, then the total number of background events that need to fit into memory during the training a single classification layer is LM. By contrast, if BDT_1 filters out the most obvious background and retains only one background event out of $N \gg 1$, then BDT_2 can be trained with only LM/N background events selected from a large sample of LM background events.

In the next paragraphs (Section 5.7.1 and Section 5.7.2), the classifiers BDT_1 and BDT_2 are defined, and in Section 5.7.3, the expected classification performance is measured.

For each classifier, the same steps are followed:

- selection of the input variables;
- choice of the classifier parameters;
- check for potential overfitting.

5.7.1. First classifier (BDT_1)

The first classifier, BDT_1, relies on the FastBDT algorithm [95], presented in Section 4.1. BDT_1 is trained on simulated signal and background events that pass the basic selection criteria listed in Section 5.5. The seven background categories are present in the

Table 5.4.: Number of simulated events employed in the training of the BDT_1 and BDT_2 classifiers, depending on the targeted mode.

Classifier	Mode	Number of signal events	Number of background events
BDT_1	$B^+ \to K^+\nu\bar{\nu}$	6.2×10^5	3.6×10^6
BDT_1	$B^0 \to K^0_S\nu\bar{\nu}$	5.7×10^5	1.6×10^6
BDT_2	$B^+ \to K^+\nu\bar{\nu}$	1.3×10^6	2.0×10^6
BDT_2	$B^0 \to K^0_S\nu\bar{\nu}$	1.2×10^6	9.2×10^5

background sample: $e^+e^- \to B^+B^-$, $e^+e^- \to B^0\overline{B}^0$, $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$), and $e^+e^- \to \tau^+\tau^-$. The total number of selected events for the training of BDT_1 are listed in Table 5.4. An independent sample with the same number of events is used to test the performance of the classifier.

During the training, the events are assigned the following weights:

- The signal events are weighted to reproduce the q^2-dependence of the signal branching fraction (see Figure 2.5 in Section 2.3.2).
- A weight is assigned to the background events to balance the signal class and the background class, such that the sum of the background weights is equal to the sum of the signal weights. With this global weight, the output of BDT_1 is interpreted as a predicted signal probability, assuming a prior of 0.5 for both the signal and the background classes.

Input variables

The BDT_1 input variables are selected from the list of variable candidates presented in Section 5.6. The variable selection relies on the variable importance, defined by Equation (4.8) in Section 4.1.5. An initial training is done with the full list of variable candidates to measure the importance of each variable in the classification, and, for simplicity, the 12 most important variables that are common to both modes are retained for the final training of BDT_1 (Table 5.5). The simple variable selection strategy of taking the 12 most important variables is sufficient for BDT_1, whose role is to reject the most obvious background. For the training of BDT_2, which serves as the final step of the event selection, a more elaborate variable selection strategy is developed in Section 5.7.2.

Table 5.5.: Importance of the 12 input variables entering the training of BDT_1. The variable importance is defined by Equation (4.8) in Section 4.1.5. The variables are defined in Section 5.6 and the used notations are summarised in Appendix A.1.

Variable	$B^+ \to K^+\nu\bar{\nu}$	$B^0 \to K^0_S\nu\bar{\nu}$
ΔE_{ROE}	0.62682	0.59926
Modified Fox-Wolfram $H^{so}_{m,2}$	0.08731	0.09382
p_{ROE}	0.03517	0.03869
Modified Fox-Wolfram $H^{so}_{m,4}$	0.02372	0.01397
Modified Fox-Wolfram R^{oo}_0	0.02003	0.00876
Modified Fox-Wolfram R^{oo}_2	0.01988	0.02958
$\theta(p_{\mathrm{ROE}})$	0.01813	0.01584
Harmonic Moment B_0	0.01732	0.02434
$\cos(\mathrm{thrust}_B, \mathrm{thrust}_{\mathrm{ROE}})$	0.01371	0.00954
Fox-Wolfram Moment R_1	0.01072	0.03725
$\cos(\theta(\mathrm{thrust}))$	0.00953	0.00992
Harmonic Moment B_2	0.00648	0.01390

Table 5.5 shows that the two most important variables are ΔE_{ROE} and the modified Fox-Wolfram moment $H^{so}_{m,2}$, which totalise approximately 70% of the separation power. These two variables strongly depend on the missing energy and momentum in the event, and are thus expected to provide a good discrimination between signal and background events, because in a signal event, the energy and momentum of the two neutrinos is not detected.

Choice of the classifier parameters

The classifier parameters are tuned to maximise the area under the receiver-operating-characteristic (ROC) curve evaluated on the test sample ($\mathrm{AUC}_{\mathrm{test}}$), while avoiding to fall in an overfitting regime. The concept of ROC curve is introduced in Section 4.1.4. Only the $B^+ \to K^+\nu\bar{\nu}$ mode is considered for the parameter tuning, assuming the optimal parameters to be similar for the $B^0 \to K^0_S\nu\bar{\nu}$ mode.

A grid search in the parameter space is conducted: for each of the 900 combinations of parameter values listed in Table 5.6, a classifier is trained and the classification performance is reported in Figure 5.10 (left), where the results are ranked according to the AUC

Table 5.6.: Tested and chosen values of the BDT_1 and BDT_2 parameters. In total, $4 \cdot 5 \cdot 3 \cdot 3 \cdot 5 = 900$ combinations of parameters are tested. The definition of the parameters is given in Section 4.1.3.

Parameter	Tested values	BDT_1	BDT_2
Number of trees	[200, 500, 1000, 2000]	2000	2000
Tree depth	[2, 3, 4, 5, 6]	2	3
Learning rate	[0.05, 0.1, 0.2]	0.2	0.2
Sampling rate	[0.5, 0.8, 1.0]	0.5	0.5
Number of equal-frequency bins	[2^4, 2^6, 2^8, 2^{10}, 2^{12}]	2^8	2^8

measured on the training sample (AUC_train).

On the left part of the Figure 5.10 (left), AUC_train becomes significantly larger than AUC_test, meaning that the classifiers are in an overfitting regime. On the right part of Figure 5.10 (left), AUC_train and AUC_test are closer, but the classification performance is lower, meaning that the classifiers are in an underfitting regime.

A trade-off is found by plotting the parameter to maximise, AUC_test, as a function of the overfitting ratio

$$\frac{\text{AUC}_\text{train} - \text{AUC}_\text{test}}{\text{AUC}_\text{test}} \geq 0. \tag{5.4}$$

An overfitting regime is characterised by a large overfitting ratio. The result is shown in Figure 5.10 (right), where the retained combination of classifier parameters is highlighted. The chosen parameters are listed in Table 5.6.

Overfitting check

As an additional check, the output of BDT_1 is compared for the training and the test samples (Figure 5.11, top). As expected from the parameter selection procedure, the overfitting is under control. An equivalent check is to show that the ROC curves computed on the training and the test samples nearly overlap (Figure 5.11, bottom).

5.7.2. Second classifier (BDT_2)

The second classifier, BDT_2, relies on the XGBoost implementation of the tree boosting algorithm [97], presented in Section 4.1.2. The advantage of XGBoost is that it offers the

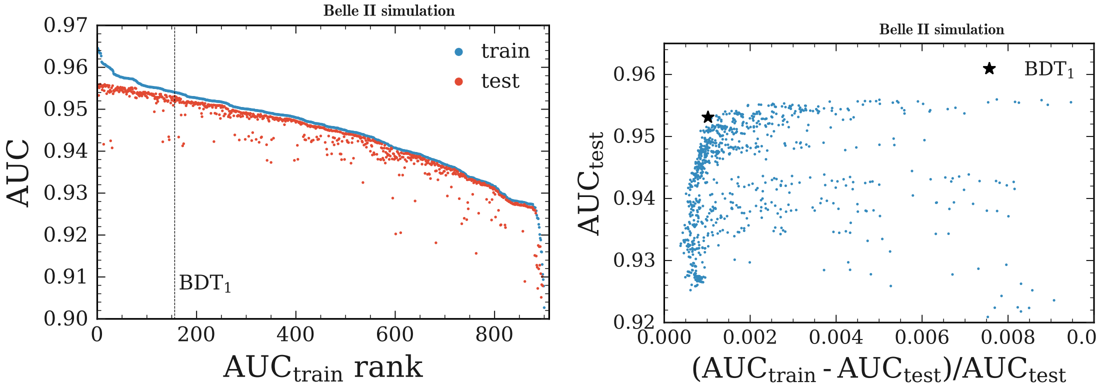


Figure 5.10.: Comparison of 900 classifiers, each trained with a certain combination of input parameters. On the left, the classifiers are ranked according to AUC$_{\text{train}}$, and, for each, AUC$_{\text{train}}$ and AUC$_{\text{test}}$ are shown. On the right, AUC$_{\text{test}}$ as a function of the overfitting ratio. The selected combination of input parameters for BDT$_1$ is highlighted on both plots.

possibility of using a graphics processing unit (GPU) during the training, which drastically reduces the training time compared to FastBDT[3].

BDT$_2$ is trained on simulated background and signal events that pass the selection criteria listed in Section 5.5 and that also satisfy BDT$_1 > 0.9$. For both modes ($B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$), the selection BDT$_1 > 0.9$ retains approximately 85% of the signal events and reject 98% of the background events in the test sample of BDT$_1$. The total number of selected events for the training of BDT$_2$ are listed in Table 5.4. An independent sample with the same number of events is used to test the performance of the classifier. As for BDT$_1$, a global weight is used to balance the signal class and the background class.

Input variables

The input variable selection strategy for BDT$_2$ is more elaborate than the strategy used for BDT$_1$, because BDT$_2$ is the final stage of the event selection, so it is important to not reject an input variable that could contribute to the classification performance.

For the $B^+ \to K^+\nu\bar{\nu}$ mode and the $B^0 \to K^0_S\nu\bar{\nu}$ mode, a total of 47 and 46 input variable candidates are considered, respectively. For each mode, a first classifier is trained with the entire set of input variable candidates. The next step is to remove the least important variable and to train a new classifier on the restricted set of input variable candidates.

---

[3] FastBDT is chosen for BDT$_1$, because unlike XGBoost, it is fully integrated into the software used for the distributed computing needed for the first steps of the selection.

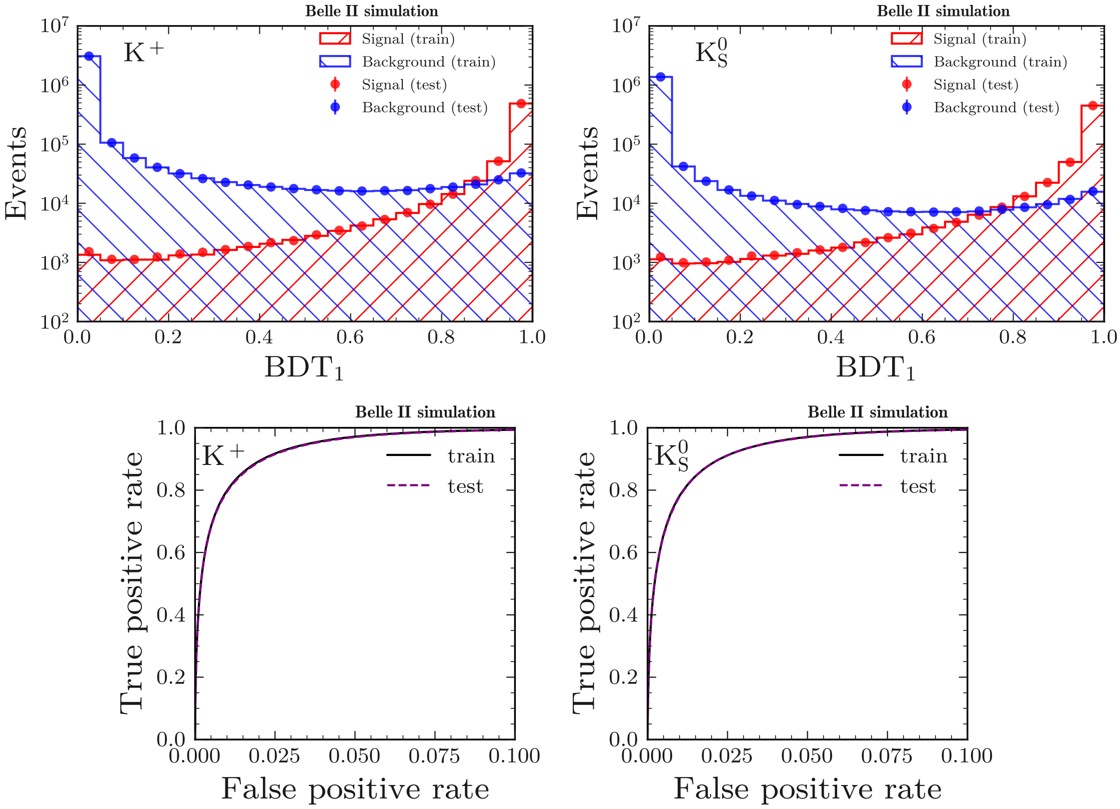


Figure 5.11.: Simulated signal and background events taken from the training and the test samples in bins of the BDT$_1$ output for the $B^+ \to K^+\nu\bar{\nu}$ mode (upper left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (upper right), and corresponding ROC curves (bottom left and bottom right), showing how the true positive rate and the false positive rate evolve when scanning all possible lower thresholds on the BDT$_1$ output.

The loss in AUC$_\text{test}$ caused by the removal of the variable, noted $\Delta(\text{AUC}_\text{test})$, is stored. This process is repeated until all the variables but one are removed from the training.

Figure 5.12 shows the result of this procedure by plotting the cumulative sum of the AUC$_\text{test}$ loss as a function of the number of removed variables. The removal of approximately 10 variables has no impact on the classification performance. For each mode, the 35 variables with the largest loss are retained for the final training of BDT$_2$ (Tables 5.7 and 5.8).

Choice of the classifier parameters

For the choice of the BDT$_2$ parameters, the method used for BDT$_1$ is replicated. Again, only the $B^+ \to K^+\nu\bar{\nu}$ mode is considered for the parameter tuning, assuming the optimal

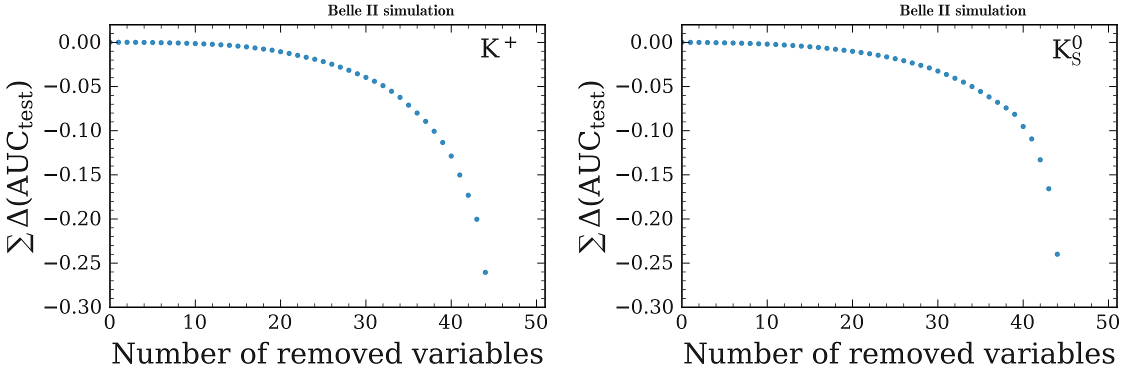


Figure 5.12.: Cumulative sum of the $\mathrm{AUC_{test}}$ loss caused by the iterative removal of input variables from the training of $\mathrm{BDT_2}$ for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K_S^0\nu\bar{\nu}$ mode (right).

parameters to be similar for the $B^0 \to K_S^0\nu\bar{\nu}$ mode.

A grid search in the parameter space is conducted: for each of the 900 combinations of parameter values listed in Table 5.6, a classifier is trained and the classification performance is reported in Figure 5.13.

The chosen parameters result from a trade-off between the parameter to maximise, $\mathrm{AUC_{test}}$, and the overfitting ratio $(\mathrm{AUC_{train}}-\mathrm{AUC_{test}})/\mathrm{AUC_{test}}$. These parameters are highlighted in Figure 5.13 and listed in Table 5.6.

Overfitting check

The check for potential $\mathrm{BDT_2}$ overfitting is the same as the one used for $\mathrm{BDT_1}$. The $\mathrm{BDT_2}$ output and the ROC curves computed on the training and the test samples are shown in Figure 5.14. As for $\mathrm{BDT_1}$, the overfitting is under control.

5.7.3. Expected classification performance

A more concrete idea of the classification performance is given by comparing the expected significance $S/\sqrt{S+B}$ of the selection and the signal efficiency when scanning lower thresholds on the output of $\mathrm{BDT_1}$ and $\mathrm{BDT_2}$, where S and B are the expected number of signal and background events, respectively.

In a data sample corresponding to an integrated luminosity of L, and for a given signal selection efficiency $\varepsilon_{\mathrm{sig}}$, the expected number of selected events containing a $B^+ \to K^+\nu\bar{\nu}$

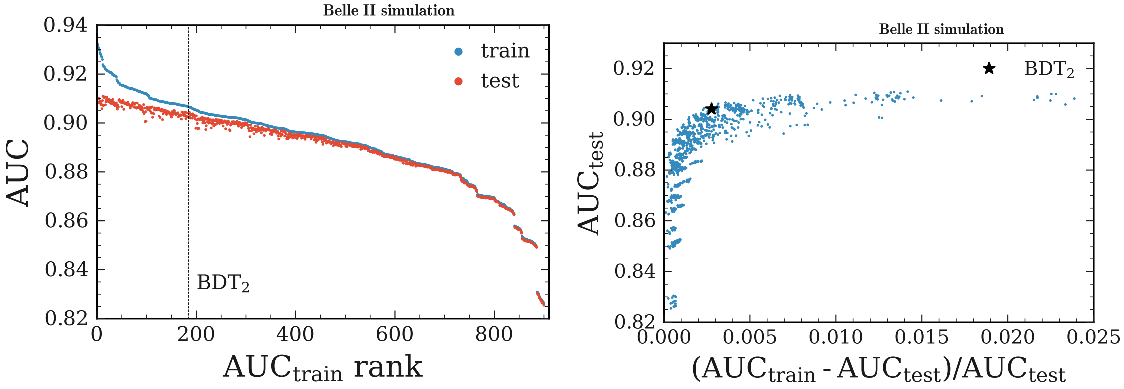


Figure 5.13.: Comparison of 900 classifiers, each trained with a certain combination of input parameters. On the left, the classifiers are ranked according to $\mathrm{AUC}_{\mathrm{train}}$, and, for each, $\mathrm{AUC}_{\mathrm{train}}$ and $\mathrm{AUC}_{\mathrm{test}}$ are shown. On the right, $\mathrm{AUC}_{\mathrm{test}}$ as a function of the overfitting ratio. The selected combination of input parameters for BDT_2 is highlighted on both plots.

decay is

$$S = \varepsilon_{\mathrm{sig}} \cdot L \cdot \sigma(e^+e^- \to \Upsilon(4S)) \cdot 2 \cdot \mathrm{Br}(\Upsilon(4S) \to B^+B^-) \cdot \mathrm{Br}(B^+ \to K^+\nu\bar{\nu}), \tag{5.5}$$

where $\sigma(e^+e^- \to \Upsilon(4S))$ is the cross-section of the $\Upsilon(4S)$ production, and the factor of two is due to the two B mesons present in each $\Upsilon(4S)$ decay. For the $B^0 \to K^0_S\nu\bar{\nu}$ mode, the equation becomes

$$S = \varepsilon_{\mathrm{sig}} \cdot L \cdot \sigma(e^+e^- \to \Upsilon(4S)) \cdot 2 \cdot \mathrm{Br}(\Upsilon(4S) \to B^0\overline{B}^0) \cdot \mathrm{Br}(B^0 \to K^0_S\nu\bar{\nu}). \tag{5.6}$$

The predicted branching fraction of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$ is given by Equation (2.12) in Section 2.3.2. The values of the other factors are [25, 115]:

$$\begin{aligned} \sigma(e^+e^- \to \Upsilon(4S)) &= (1.10 \pm 0.02) \times 10^6 \text{ fb}, \\ \mathrm{Br}(\Upsilon(4S) \to B^+B^-) &= 0.514 \pm 0.06, \\ \mathrm{Br}(\Upsilon(4S) \to B^0\overline{B}^0) &= 0.486 \pm 0.06. \end{aligned} \tag{5.7}$$

To correct for multiple sources of mis-modelling, and to obtain a more realistic value of the significance, each simulated event is given a correction weight in the calculation of S and B:

- The selection efficiency of the PID requirement imposed on signal K^+ candidates (Section 5.4) differs between data and simulation. By comparing data and simulation, the Belle II performance group provides weights that correct for the efficiency difference. These weights are functions of the transverse momentum p_T and the

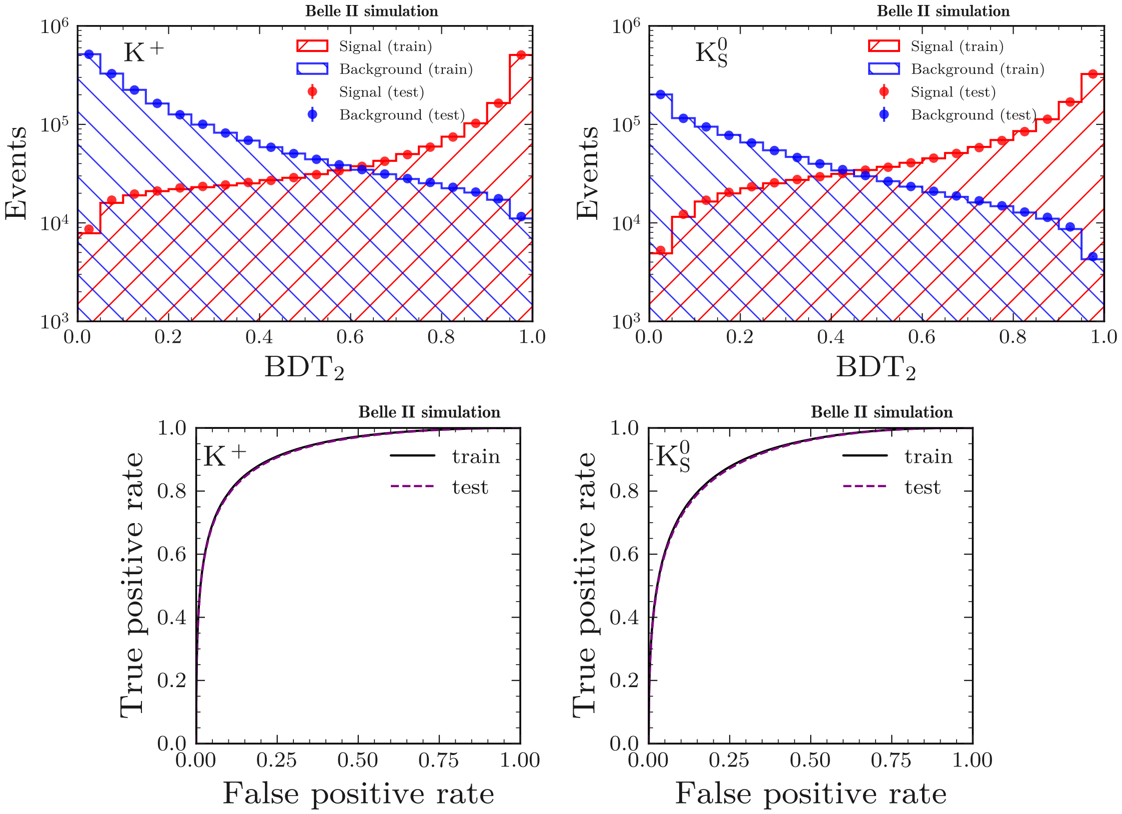


Figure 5.14.: Simulated signal and background events taken from the training and the test samples in bins of the BDT$_2$ output for the $B^+ \to K^+\nu\bar{\nu}$ mode (upper left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (upper right), and corresponding ROC curves (bottom left and bottom right), showing how the true positive rate and the false positive rate evolve when scanning all possible lower thresholds on the BDT$_2$ output.

polar angle θ of the signal K^+ candidate. They apply to all signal and background samples, but only when searching for the charged mode ($B^+ \to K^+\nu\bar{\nu}$), because there is no PID requirement for the neutral mode ($B^0 \to K^0_S\nu\bar{\nu}$).

- As already mentioned in Section 5.4, simulation does not take into account the q^2-dependence of the form factor entering in the computation of the $B \to K\nu\bar{\nu}$ branching fraction (Section 2.3.2). A correction weight is applied to the simulated signal events to produce a realistic q^2-dependence of the signal branching fraction. This weight is computed as the ratio between the two distributions shown in Figure 2.5 (Section 2.3.2).

The total correction weight for a simulated event is the product of all the weights that apply to this events. Appendix A.3 gives an overview of all the correction weights that are used at the end of the analysis.

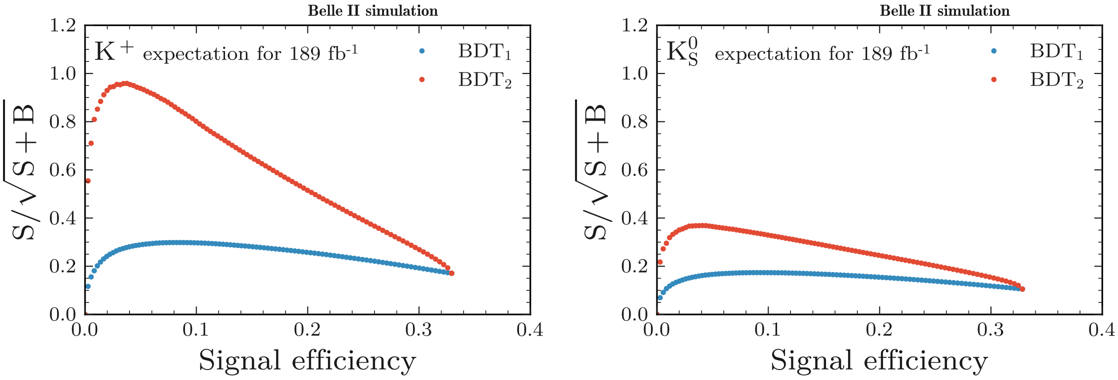


Figure 5.15.: Evolution of the signal selection efficiency and expected significance for an integrated luminosity of 189 fb^{-1} when scanning lower thresholds on the BDT$_1$ and BDT$_2$ outputs, for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The selection BDT$_1$ > 0.9 is already applied at this stage.

Figure 5.15 presents the expected significance for an integrated luminosity of 189 fb^{-1} as a function of the signal selection efficiency when scanning lower thresholds on the BDT$_1$ and BDT$_2$ outputs. The classification performance is evaluated on test samples where the requirement BDT$_1$ > 0.9 is already applied. As expected, BDT$_2$ has a better performance than BDT$_1$, since BDT$_2$ is trained with more input variables and more events in the region BDT$_1$ > 0.9.

The maximum of significance is obtained for signal selection efficiencies of the order of 4%, far greater than the efficiencies obtained with the hadronic or the semileptonic tagging method (see Table 2.3 in Section 2.3.4).

Table 5.7.: Input variable candidates ranked according to the AUC_{test} loss caused by their removal from the BDT_2 training for the $B^+ \to K^+\nu\bar{\nu}$ mode. The 35 variables with the largest loss (above the dotted line) are retained for the final training of BDT_2. The variables are defined in Section 5.6 and the used notations are summarised in Appendix A.1.

Variable	Δ(AUC)
Modified Fox-Wolfram $H^{so}_{m,2}$	−0.53286
cos(thrust$_B$, thrust$_{\text{ROE}}$)	−0.11077
Median(p-value(D^0))	−0.06013
p-value(Tag Vertex)	−0.02718
$M(D^0)$	−0.02297
$\theta(p_{\text{missing}})$	−0.02144
ΔE_{ROE}	−0.01536
Modified Fox-Wolfram R^{oo}_2	−0.01272
p-value(D^+)	−0.01123
N_{lepton}	−0.00946
Total charge squared	−0.00889
$dz(K^+$, Tag Vertex)	−0.00881
$dr(K^+)$	−0.00688
$\theta(p_{\text{ROE}})$	−0.00641
Modified Fox-Wolfram $H^{so}_{n,2}$	−0.00497
Modified Fox-Wolfram $H^{so}_{c,2}$	−0.00436
Modified Fox-Wolfram $H^{so}_{m,4}$	−0.00409
p_{ROE}	−0.00398
${M_{\text{missing}}}^2$	−0.00354
Sphericity	−0.00333
Fox-Wolfram Moment R_2	−0.00299
N_{tracks}	−0.00263
N_γ	−0.00223
$dr(D^+)$	−0.00220
dx(Tag Vertex)	−0.00212
p-value(D^0)	−0.00198
dz(Tag Vertex)	−0.00176
Fox-Wolfram Moment R_1	−0.00126
cos(thrust$_B$, z)	−0.00125
cos(θ(thrust))	−0.00121
Modified Fox-Wolfram R^{oo}_0	−0.00088
Variance$_{\text{ROE}}(p_T)$	−0.00088
Fox-Wolfram Moment R_3	−0.00058
$dr(K^+$, Tag Vertex)	−0.00057
Harmonic Moment B_0	−0.00045
dy(Tag Vertex)	−0.00036
$dr(D^0)$	−0.00033
M(ROE)	−0.00032
thrust$_{\text{ROE}}$	−0.00019
$dz(D^0)$	−0.00019
$dz(K^+)$	−0.00018
$dz(D^+)$	−0.00012
$N_{\text{tracks}} + N_\gamma$	−0.00010
Harmonic Moment B_2	−0.00004
$\phi(K^+)$	−0.00003
Modified Fox-Wolfram $H^{so}_{m,0}$	+0.00003
Thrust	+0.00013

Table 5.8.: Input variable candidates ranked according to the $\mathrm{AUC}_{\mathrm{test}}$ loss caused by their removal from the BDT_2 training for the $B^0 \to K^0_S \nu\bar{\nu}$ mode. The 35 variables with the largest loss (above the dotted line) are retained for the final training of BDT_2. The variables are defined in Section 5.6 and the used notations are summarised in Appendix A.1.

Variable	Δ(AUC)	Variable	Δ(AUC)
Modified Fox-Wolfram $H^{so}_{m,2}$	−0.68504	$M(K^0_S)$	−0.00137
cos(thrust$_B$, thrust$_{\mathrm{ROE}}$)	−0.07418	cos(thrust$_B$, z)	−0.00135
$\theta(p_{\mathrm{missing}})$	−0.03280	$M_{\mathrm{missing}}{}^2$	−0.00121
cos($p_{K^0_S}$, line(IP, K^0_S vertex))	−0.02363	$dr(p_{K^0_S})$	−0.00113
Fox-Wolfram Moment R_2	−0.01403	dz(Tag Vertex)	−0.00106
Modified Fox-Wolfram R^{oo}_2	−0.01384	cos(θ(thrust))	−0.00095
ΔE_{ROE}	−0.00724	Fox-Wolfram Moment R_3	−0.00080
N_{lepton}	−0.00636	Modified Fox-Wolfram $H^{so}_{m,0}$	−0.00071
Modified Fox-Wolfram $H^{so}_{m,4}$	−0.00613	$M(D^0)$	−0.00068
$\theta(p_{\mathrm{ROE}})$	−0.00612	Variance$_{\mathrm{ROE}}(p_T)$	−0.00061
$dz(p_{D^0})$	−0.00560	$dz(p_{K^0_S})$	−0.00054
Modified Fox-Wolfram $H^{so}_{c,2}$	−0.00503	Fox-Wolfram Moment R_1	−0.00042
p_{ROE}	−0.00436	$M(D^+)$	−0.00040
p-value(Tag Vertex)	−0.00428	$dr(p_{K^0_S}$, Tag Vertex)	−0.00034
Modified Fox-Wolfram $H^{so}_{n,2}$	−0.00386	Modified Fox-Wolfram R^{oo}_0	−0.00030
N_{tracks}	−0.00360	M(ROE)	−0.00022
N_γ	−0.00290	$N_{\mathrm{tracks}} + N_\gamma$	−0.00021
p-value(D^+)	−0.00268	$dr(p_{D^0})$	−0.00019
Harmonic Moment B_0	−0.00257	dy(Tag Vertex)	−0.00018
$dz(p_{K^0_S}$, Tag Vertex)	−0.00228	$dz(p_{D^+})$	−0.00011
Total charge squared	−0.00203	Thrust	−0.00007
dx(Tag Vertex)	−0.00188	Harmonic Moment B_2	−0.00002
Sphericity	−0.00171	thrust$_{\mathrm{ROE}}$	+0.00008

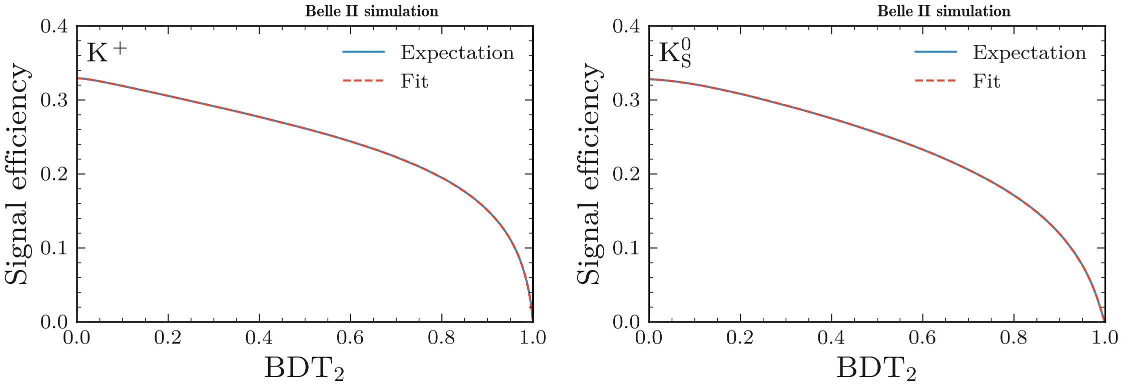


Figure 5.16.: Signal selection efficiency as a function of the lower threshold on the BDT$_2$ output for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The result of a fit with a piecewise-linear function is overlaid.

5.8. Signal search region

Now that the event classifiers are trained, the role of this section is to finalise the event selection strategy by defining a region of the phase space where a binned-likelihood model will be fit to data in order to determine $\mathrm{Br}(B \to K\nu\bar{\nu})$. Section 5.8.1 defines this region, called the signal search region. Section 5.8.2 discusses the expected background events that survive the selection.

5.8.1. Definition

The signal search region is defined by a lower threshold on the BDT$_2$ output. However, the BDT$_2$ output has no intuitive physical interpretation. For this reason, before defining the signal search region, lower thresholds on the BDT$_2$ output are first translated into corresponding signal selection efficiencies. This is done in Figure 5.16, which shows the signal selection efficiency $\varepsilon_{\mathrm{sig}}$ as a function of the lower threshold on the BDT$_2$ output. A fit with a piecewise-linear function allows to define a bijection between $\varepsilon_{\mathrm{sig}}$ and BDT$_2$. The next step is to define the signal selection efficiency quantile $\tilde{\varepsilon}_{\mathrm{sig}}$ as

$$\tilde{\varepsilon}_{\mathrm{sig}} \equiv 1 - \varepsilon_{\mathrm{sig}}(\mathrm{BDT}_2), \tag{5.8}$$

where $\varepsilon_{\mathrm{sig}}(\mathrm{BDT}_2)$ is the signal selection efficiency resulting from the fit shown in Figure 5.16. For example, imposing the condition $\tilde{\varepsilon}_{\mathrm{sig}} > 0.9$ is equivalent to selecting events with the BDT$_2$ classifier such that 10% of the signal events survive the total selection.

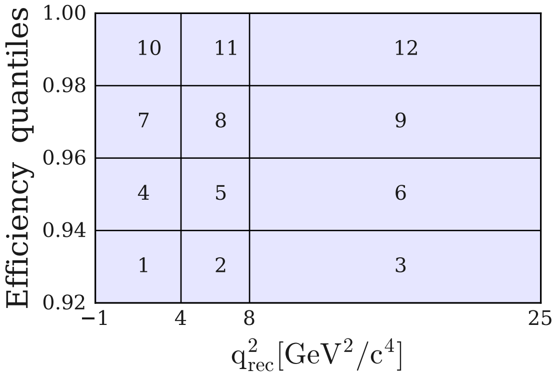


Figure 5.17.: Bin numbering scheme of the signal search region, defined as a function of the signal selection efficiency quantiles, defined in Equation (5.8), and the reconstructed invariant mass squared of the two-neutrino system q^2_{rec}, defined in Equation (5.2) in Section 5.4.

Building on this, the signal search region is defined as $\tilde{\varepsilon}_{\mathrm{sig}} > 0.92$, which for both modes includes the region of maximum of expected significance (see Section 5.7.3). The signal search region is divided in 4×3 bins in the $\tilde{\varepsilon}_{\mathrm{sig}} \times q^2_{\mathrm{rec}}$ space: $[0.92, 0.94, 0.96, 0.98, 1.00] \times [-1, 4, 8, 25]\,\mathrm{GeV}^2/\mathrm{c}^4$. The boundaries of the q^2_{rec} bins are chosen to match the bins in the theoretical reference [55]. Figure 5.17 defines the bin numbering scheme of the signal search region that is used in the following.

By definition, the total signal selection efficiency in the signal search region is 8%, but this efficiency is not flat in bins of the invariant mass squared of the two-neutrino system q^2 (Figure 5.18). For $q^2 \approx 0\,\mathrm{GeV}^2/\mathrm{c}^4$, the signal selection efficiency is of the order of 15% for the $B^+ \to K^+\nu\bar{\nu}$ mode, and of the order of 18% for the $B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$ mode. The signal selection efficiency drops as q^2 increases, meaning that lower-momentum signal kaons are harder to distinguish from background. A peaking signal selection efficiency in the low-q^2 region is a interesting property to constrain models that predicts a contribution from new light particles (see Section 2.3.3). Note that a direct comparison between the two modes is not fair, because in the neutral mode, only K^0_{S}, not K^0_{L}, are selected, meaning that the selection efficiency of $B^0 \to K^0\nu\bar{\nu}$ decays at $q^2 \approx 0\,\mathrm{GeV}^2/\mathrm{c}^4$ is $18/2 = 9\%$.

5.8.2. Expectation

Figure 5.19 shows the expected signal and background yields in the signal search region for an integrated luminosity of $189\,\mathrm{fb}^{-1}$. At the top of the figure, four bins of the signal quantiles $\tilde{\varepsilon}_{\mathrm{sig}}$ are displayed, each containing 2% of the total signal sample. By construction, the signal distribution is flat in bins of $\tilde{\varepsilon}_{\mathrm{sig}}$. At the bottom of the same figure, these four

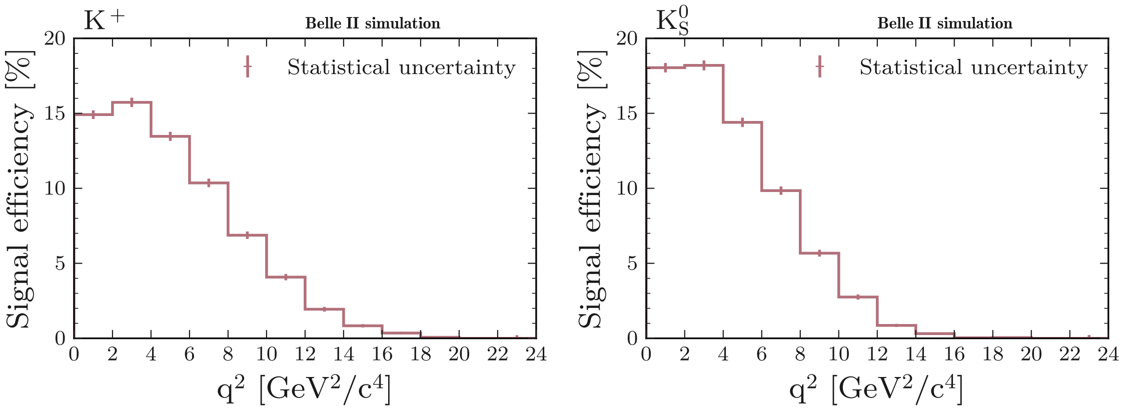


Figure 5.18.: Signal selection efficiency in the signal search region in bins of the invariant mass squared of the two-neutrino system for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). Note that this is the simulated q^2, not q^2_{rec}.

$\tilde{\varepsilon}_{\text{sig}}$ bins are further divided in three q^2_{rec} bins. The fact that the lowest-q^2_{rec} bin contains more signal events in the final $\tilde{\varepsilon}_{\text{sig}}$ bin indicates a negative correlation between the BDT$_2$ output, on which $\tilde{\varepsilon}_{\text{sig}}$ is defined, and q^2_{rec}. In other words, BDT$_2$ tends to select low-q^2_{rec} events, as already anticipated above by the signal selection efficiency drop for large q^2 (Figure 5.18).

To obtain the expected yields in Figure 5.19, simulated background samples corresponding to 800 fb^{-1} of equivalent integrated luminosity are weighted to match the integrated luminosity of the data sample (189 fb^{-1}). In addition, 4×10^6 simulated signal events are weighted before the selection to match the total number of expected signal events in 189 fb^{-1} of data (Section 5.7.3). Note that in Figure 5.19, the signal expectation is upscaled by a factor of 100 to help the visualisation.

As explained in Section 5.4, a single signal kaon candidate is chosen in each event. In the signal search region, the signal kaon candidate is the correct one in 99.9% of the selected simulated $B^+ \to K^+\nu\bar{\nu}$ events, and in 99.7% of the selected simulated $B^0 \to K^0_S\nu\bar{\nu}$ events.

For the charged mode ($B^+ \to K^+\nu\bar{\nu}$), the background contributing the most in the signal search region is the charged B meson background. For the neutral mode ($B^0 \to K^0_S\nu\bar{\nu}$), the neutral B meson background contributes more than the charged B meson background, as expected, but the $e^+e^- \to c\bar{c}$ background remains also a relatively large source of background in the last bins of the signal search region. The reason for the large $e^+e^- \to c\bar{c}$ contribution to the neutral mode is unclear, but may be due to a less effective continuum-background suppression in the neutral case, because several discriminative variables are based on impact parameters, and the precision on the direction of the kaon momentum is

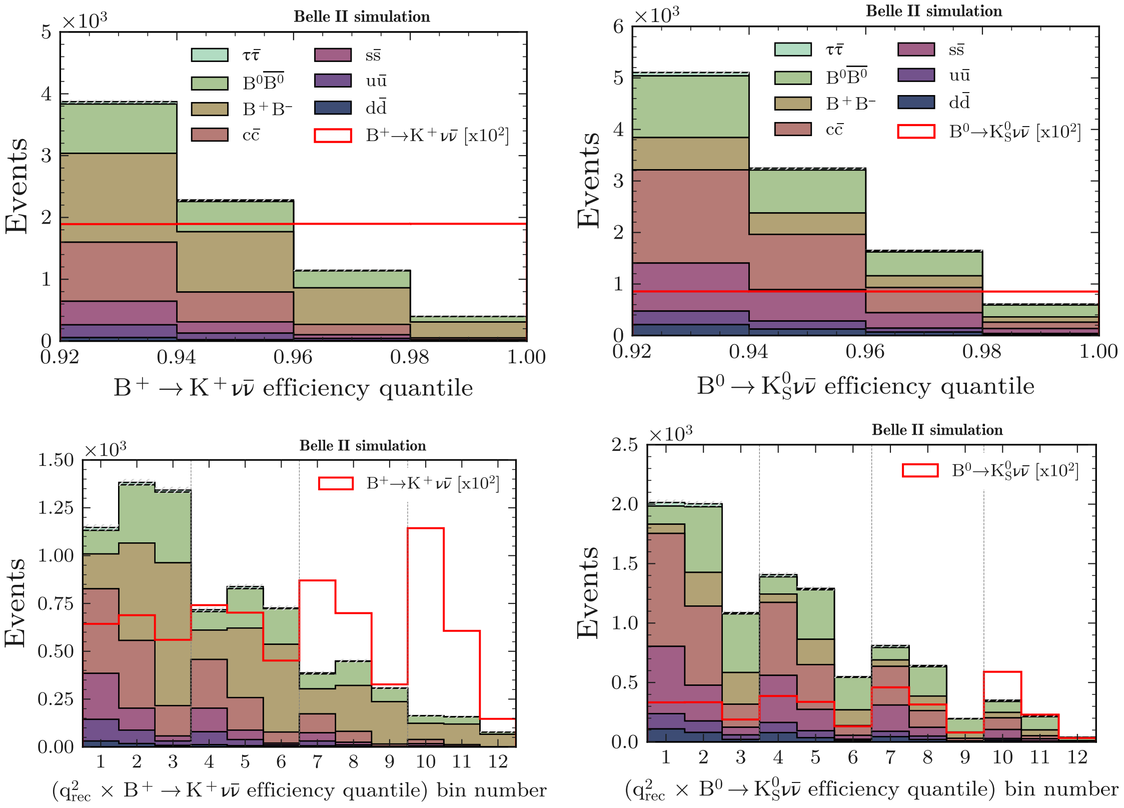


Figure 5.19.: Expected number of signal and background events for 189 fb^{-1} of data in bins of efficiency quantiles (top), and in bins of the signal search region (bottom), for the $B^+ \to K^+\nu\bar{\nu}$ mode (left), and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The signal expectation is upscaled by a factor of 100. The bins of the signal search region are defined in Figure 5.17.

lower in the K^0_S case, where two tracks are fit to a common vertex, than in the K^+ case.

The simulated B meson decays contributing the most in the signal search region are the semileptonic $B \to D^{(*)}\ell\nu$ decays ($\ell = e, \mu$), where a kaon from the D meson decay is selected as the signal kaon candidate. These decays represent more than 50% of the B meson background in the signal search region for both the charged and the neutral modes, and their presence motivates the use of discriminative variables focusing on the D meson suppression during the classification. Details on the composition of the B meson background in the signal search region are given in Appendix A.2.

Another source of B meson background comes from the cross-feed of non-signal decays that are also of the $B \to K^{(*)}\nu\bar{\nu}$ type. For example, the signal search region of the $B^+ \to K^+\nu\bar{\nu}$ mode may be polluted by events containing a $B^0 \to K^0\nu\bar{\nu}$ decay, a $B \to K^{*0}\nu\bar{\nu}$ decay, or a $B \to K^{*+}\nu\bar{\nu}$ decay. To study this background, simulated events of each decay category are

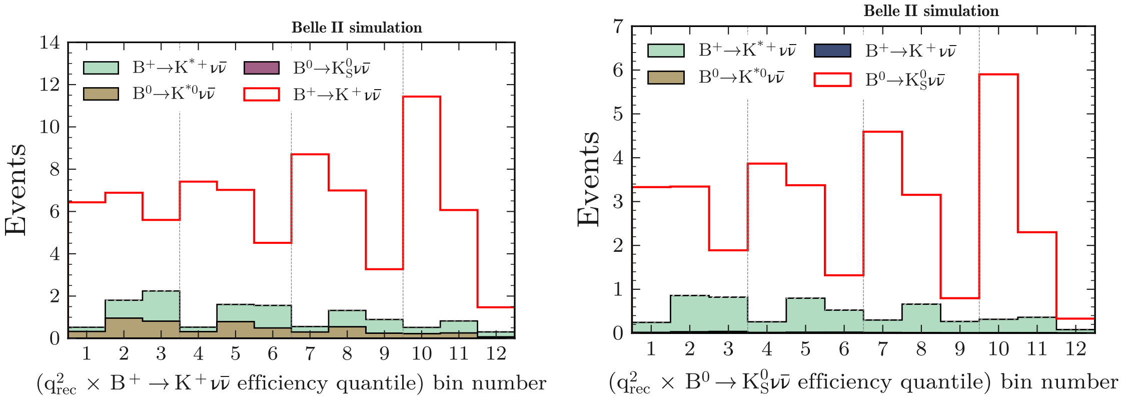


Figure 5.20.: Expected number of signal events for 189 fb^{-1} of data in bins of the signal search region, for the $B^+ \to K^+\nu\bar{\nu}$ mode (left), and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right), and expected number of events from other $B \to K^{(*)}\nu\bar{\nu}$ modes. The bins of the signal search region are defined in Figure 5.17.

reconstructed, selected, and weighted to match the number of events expected in 189 fb^{-1} of data. Figure 5.20 shows that the expected cross-feed from the other $B \to K^{(*)}\nu\bar{\nu}$ modes is at the order of 10% with respect to the expected signal yield in the signal search region. In simulation, these background decays are included in the $e^+e^- \to B^+B^-$ and $e^+e^- \to B^0\overline{B}^0$ background samples and do not receive a special treatment.

Intermission

In the previous sections (Sections 5.3 to 5.8), only simulated samples are used to define the selection of $B \to K\nu\bar{\nu}$ decays and to determine the expected significance and efficiency of the selection.

In principle, since the bins of the signal search region are defined in Section 5.8, it would be possible to apply the same selection on simulated and data events, to fit the binned-likelihood model introduced in Section 4.4, and to determine $\mathrm{Br}(B \to K\nu\bar{\nu})$ with no further considerations.

However, this naïve approach would neglect any kind of discrepancy between data and simulation, and several preparatory steps are still needed:

- In Section 5.9, data and simulation are compared in different regions of the phase space to validate the selection procedure.

- In Section 5.10, the impact of mis-modelling on the determination of $\mathrm{Br}(B \to K\nu\bar{\nu})$ is quantified via the introduction of multiple systematic uncertainties.

Finally, Sections 5.11 and 5.12 focus on the fit of the model to the data and the determination of $\mathrm{Br}(B \to K\nu\bar{\nu})$, or a least of an upper limit on $\mathrm{Br}(B \to K\nu\bar{\nu})$.

5.9. Comparison between data and simulation

As mentioned above, the role of this section is to validate the selection procedure by comparing data and simulation in several regions where the signal is absent or negligible compared to the background. This absence or negligible amount of signal is important to avoid biasing the signal selection strategy (closed-box principle).

- In Section 5.9.1, a validation channel, $B \to KJ/\psi$, is employed to verify that the classifiers behave similarly for events taken from simulation and from data.
- In Section 5.9.2, the agreement between continuum background simulation and off-resonance data is examined.
- In Section 5.9.3, a method to correct the mis-modelling of the continuum background is presented.
- In Section 5.9.4, simulation and on-resonance data events that are close to the signal search region are compared.

5.9.1. Validation channel

The $B \to KJ/\psi$ decay, with J/ψ decaying into a pair of muons, is employed to verify that the classifiers BDT_1 and BDT_2 behave similarly for events taken from simulation and from data. A first advantage of the $B \to KJ/\psi$ decay is a relatively large branching fraction, of the order of 10^{-3} [25]. Another advantage is that this decay can be selected with high purity with requirements on the invariant mass of the J/ψ candidate and the energy of the B candidate. The following procedure is applied:

1. $B \to KJ/\psi(\to \mu^+\mu^-)$ decays are selected in data and in simulated samples. The data sample is the full on-resonance data sample corresponding to $189\,\mathrm{fb}^{-1}$ of integrated luminosity, and a dedicated simulated sample of $B \to KJ/\psi$ decays is used in addition. The selection of the kaon candidate in $B \to KJ/\psi(\to \mu^+\mu^-)$ is identical as the one used for $B \to K\nu\bar{\nu}$ (Section 5.4). The muon candidates are tracks of

opposite charges that satisfy the basic criteria listed in Section 5.3. In addition, the muon candidates are required to have a muon-hypothesis likelihood (PID score) larger than 0.5. Finally, the following conditions are imposed to ensure a high-purity sample of $B \to KJ/\psi$ decays:

$$\begin{aligned} \left| M_{\mu\mu} - M_{J/\psi}^{\mathrm{PDG}} \right| &< 0.05\,\mathrm{GeV}/c^2, \\ |\Delta E| \equiv \left| E_B^* - \frac{\sqrt{s}}{2} \right| &< 0.1\,\mathrm{GeV}, \\ M_{\mathrm{bc}} \equiv \sqrt{\left(\frac{\sqrt{s}}{2c^2}\right)^2 - \left(\frac{p_B^*}{c}\right)^2} &> 5.25\,\mathrm{GeV}/c^2, \end{aligned} \tag{5.9}$$

where $M_{\mu\mu}$ is the invariant mass of the two-muon system, $M_{J/\psi}^{\mathrm{PDG}}$ is the known mass of J/ψ [25], ΔE is the energy difference in the centre-of-mass system between the energy of the B meson candidate E_B^* and half of the collision energy $\sqrt{s}$, and where the beam-constrained mass M_{bc} is defined as a function of the momentum of the B meson candidate in the centre-of-mass system p_B^*. Figure 5.21 shows $M_{\mu\mu}$, ΔE and M_{bc} for data and simulated events after this selection. As expected, the sample has a high purity.

2. For each selected event in step 1, the two muon candidates are removed and the kaon candidate is replaced with a kaon sampled from simulated $B \to K\nu\bar{\nu}$ events. By doing so, the $B \to KJ/\psi(\to \mu^+\mu^-)$ events (both data and simulated events) are modified in order to mimic $B \to K\nu\bar{\nu}$ events. This modification is called signal embedding, because the $B \to KJ/\psi$ decay is removed and replaced by a $B \to K\nu\bar{\nu}$ decay, while the rest of the event is left unmodified.

3. All the classifier input variables are recalculated for the modified events resulting from step 2, and the outputs of BDT_1 and BDT_2 are examined.

Figure 5.22 presents the result of this procedure. Without the muon removal and the kaon replacement, the $B \to KJ/\psi(\to \mu^+\mu^-)$ events are categorised as background, as they should be. By contrast, after the signal embedding procedure, the events are categorised as signal, and there is a good agreement between data and simulation.

In Figure 5.23, data and simulated $B \to KJ/\psi$ events after the signal embedding procedure are compared in bins of the signal search region (i.e. $0.92 < \tilde{\varepsilon}_{\mathrm{sig}} \leq 1$). There is a reasonable agreement between data and simulation for the shape of the distribution. If the selection efficiency in the signal search region is noted $\varepsilon_{\mathrm{simulation}}$ for simulation and $\varepsilon_{\mathrm{data}}$ for data, the ratio between the two indicates a normalisation discrepancy of 8%, consistent for the

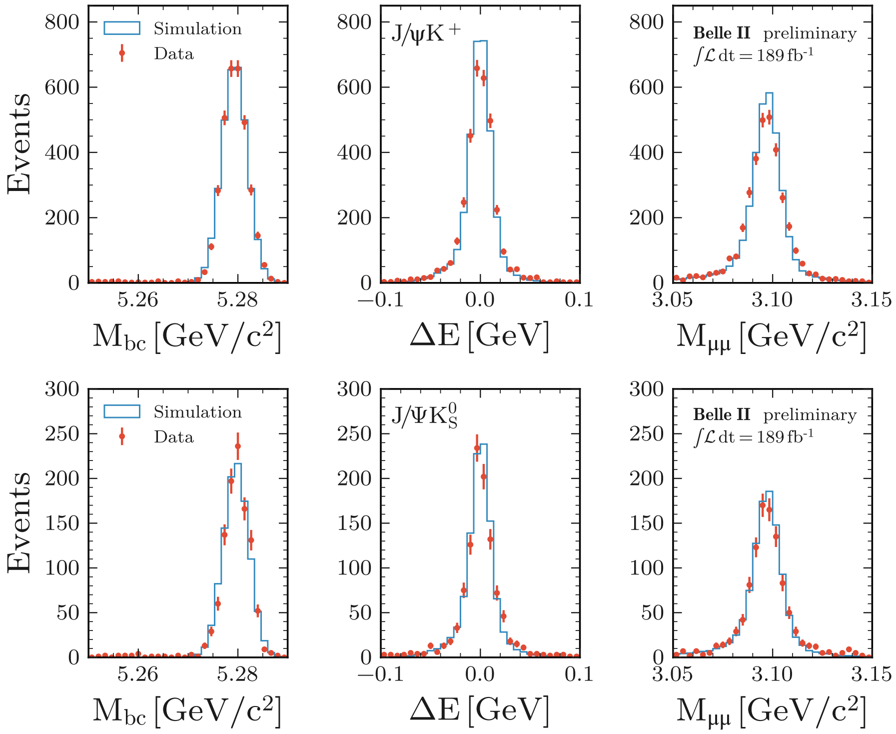


Figure 5.21.: Simulated and data events satisfying the $B \to KJ/\psi(\to \mu^+\mu^-)$ selection criteria in bins of the beam-constrained mass of the B meson candidate (left), the ΔE of the B meson candidate (middle), the mass of the two-muon system (right), for the $B^+ \to K^+J/\psi$ mode (top) and the $B^0 \to K^0_S J/\psi$ mode (bottom).

charged and the neutral modes:

$$\begin{aligned} \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{simulation}}} &= 1.08 \pm 0.04 \qquad \text{for } B^+ \to K^+J/\psi \text{ events,} \\ \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{simulation}}} &= 1.08 \pm 0.07 \qquad \text{for } B^0 \to K^0_S J/\psi \text{ events.} \end{aligned} \tag{5.10}$$

The normalisation discrepancy is not visible in Figure 5.23, because only the shape of the distributions is shown. This 8% normalisation discrepancy will be treated as a source of systematic uncertainty in Section 5.10.

5.9.2. Off-resonance data and simulation

The data sample collected at an energy 60 MeV below the $\Upsilon(4S)$ resonance, corresponding to an integrated luminosity of 18 fb^{-1}, does not contain $e^+e^- \to B^+B^-$ nor $e^+e^- \to B^0\overline{B}^0$ events and is used for two purposes:

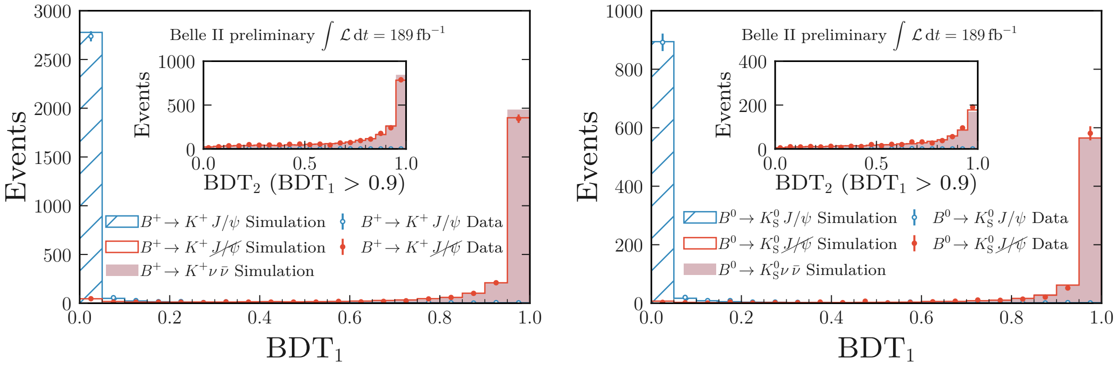


Figure 5.22.: Simulated and data events in bins of BDT$_1$ (main figure) and BDT$_2$ when restricting to the region BDT$_1$ $>$ 0.9 (inset), for the charged mode (left) and for the neutral mode (right). Five samples are compared: simulated $B \to K\nu\bar{\nu}$ events (filled red histogram), unmodified data and simulated $B \to KJ/\psi$ events (blue dots and blue histogram peaking at BDT$_1$ $=$ 0), and data and simulated $B \to KJ/\psi$ events after the signal embedding procedure, noted $B \to K\cancel{J/\psi}$ (red dots and red histogram peaking at BDT$_2$ $=$ 1). The simulation histograms are scaled to the total number of $B \to KJ/\psi$ events selected in data.

- To validate and correct the modelling of the continuum background simulation (this section and Section 5.9.3).
- To constrain the normalisation of the continuum background samples in the binned-likelihood model (Sections 5.11 and 5.12).

Figure 5.24 compares the off-resonance data and the continuum background simulation in bins of ΔE_{ROE} in the region $0.75 < \tilde{\varepsilon}_{\mathrm{sig}} \leq 1.0$ for the charged mode selection ($B^+ \to K^+\nu\bar{\nu}$). In this region, a global normalisation discrepancy factor $\frac{\text{Data}}{\text{Simulation}}$ of 1.33 is observed with respect to the expectation from the measured integrated luminosity. In Figure 5.24, the expectations are scaled up by this factor of 1.33 in order to facilitate the comparison of the distribution shapes. In the signal search region (i.e. $0.92 < \tilde{\varepsilon}_{\mathrm{sig}} \leq 1.0$), the global normalisation discrepancy factor increases to a value of 1.5, which will be taken into account in the systematic uncertainties (Section 5.10).

The reason for this normalisation discrepancy is unclear, but may be due to mis-modelling of the quark fragmentation in the PYTHIA library, mentioned in Section 3.2.8. In [116], the Belle collaboration compares several PYTHIA tunings that cause a spread in the simulated cross-sections. In [77], the Belle collaboration is also mentioning a normalisation discrepancy for the continuum background simulation.

In addition to the normalisation discrepancy, Figure 5.24 shows that the continuum background suffers from mis-modelling that affects the shape of ΔE_{ROE}. This specific variable

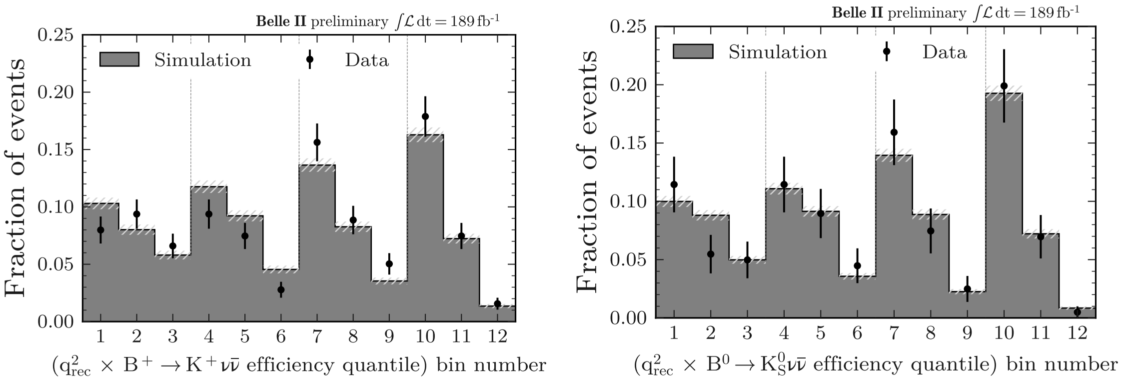


Figure 5.23.: Data and simulated $B \to KJ/\psi$ events after the signal embedding procedure in bins of the signal search region for the charged mode (left) and the neutral mode (right). The data and simulation histograms are divided by the total number of events that they contain. The bins of the signal search region are defined in Section 5.8.

is chosen as an example, because the effect of the mis-modelling on the distribution is particularly visible. This data-simulation disagreement is mitigated by applying correction weights to the simulated events, and these correction weights are defined below.

5.9.3. Correction of the continuum mis-modelling

Following [117], the continuum background simulation is corrected by applying the method presented below:

1. A new binary classifier, $\mathrm{BDT_c}$, is trained to separate off-resonance *data* events (considered here as signal) and *simulated* continuum events (considered here as background). If the modelling of the continuum background was perfect, the classifier would not be able to distinguish between data and simulation, and $\mathrm{BDT_c}$ would peak at 0.5 for both data and simulated events. In case of mis-modelling, some events are more data-like ($\mathrm{BDT_c} > 0.5$) and others are more simulation-like ($\mathrm{BDT_c} < 0.5$).

2. For each simulated continuum event, $\mathrm{BDT_c}$ is interpreted as an estimate of the likelihood for this event to come from data, and $1 - \mathrm{BDT_c}$ is an estimate of the likelihood for this event to come from simulation. Thus, the continuum background simulation is corrected by applying to each simulated event a weight w_c corresponding to the likelihood ratio $w_c \equiv \mathrm{BDT_c}/(1 - \mathrm{BDT_c})$. Since w_c becomes arbitrarily large if $\mathrm{BDT_c} \to 1$, a security ensures that weights larger than 5 are set to 5. Moreover, to avoid affecting the background normalisation, w_c is divided by its average value.

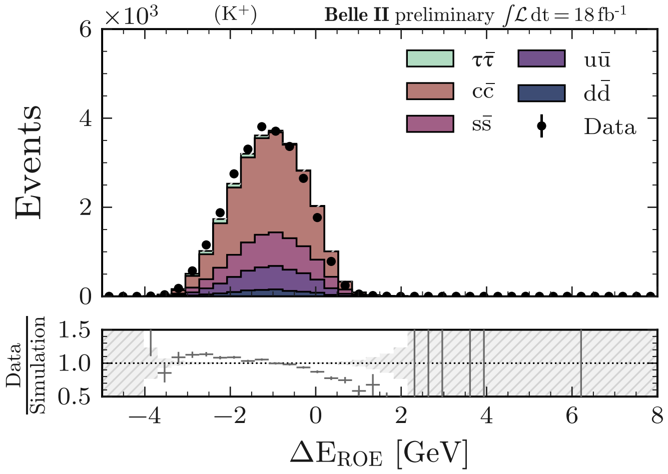


Figure 5.24.: Simulated continuum background events and off-resonance data events in bins of ΔE_{ROE} in the region $0.75 < \tilde{\varepsilon}_{\text{sig}} \leq 1.0$. The simulation histograms are scaled up by 33% with respect to the expectation to correct for the normalisation discrepancy observed in this region. The statistical uncertainty due to simulation is indicated by the hashed area.

Table 5.9.: Number of events employed in the training of the BDT_{c} classifier, depending on the mode. The signal events are off-resonance *data* events, and the background events are *simulated* continuum events.

Mode	Number of signal events	Number of background events
$B^+ \to K^+\nu\bar{\nu}$	2.5×10^4	2.3×10^5
$B^0 \to K^0_S\nu\bar{\nu}$	2.5×10^4	2.0×10^5

The input variables of BDT_{c} are the same as the input variables of BDT_2 (Section 5.7.2), plus q^2_{rec} and the output of BDT_2. The data events used for the BDT_{c} training are selected from the off-resonance sample corresponding to $18\,\text{fb}^{-1}$ of integrated luminosity. The simulated continuum background events are selected from a sample of $100\,\text{fb}^{-1}$ of equivalent integrated luminosity. BDT_{c} is trained in the region $0.75 < \tilde{\varepsilon}_{\text{sig}} \leq 1.0$, and the number of events used for the training are summarised in Table 5.9. Since the number of data events used in the training is rather small ($\mathcal{O}(10^4)$), the classifier parameters, listed in Table 5.10, are chosen to strongly suppress the overfitting.

Figure 5.25 (left) diplays the output of BDT_{c}. In this figure, only a half of the data and simulated samples are used for the training of BDT_{c}, and the other half are used as independent test samples, to check that the overfitting is under control. For the final training

Table 5.10.: Parameters of the $\mathrm{BDT_c}$ classifier.

Parameter	Value
Number of trees	2000
Tree depth	1
Learning rate	0.01
Sampling rate	0.01
Number of equal-frequency bins	2^8

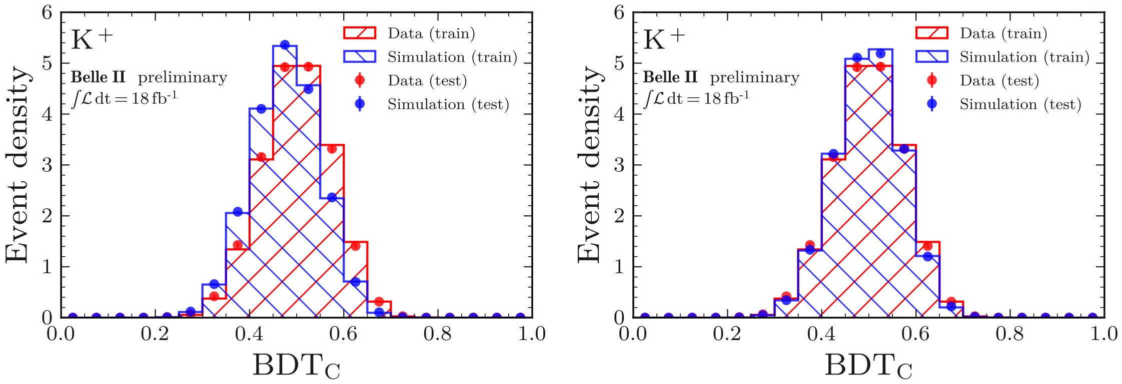


Figure 5.25.: Simulated continuum background events and off-resonance data events from the training and the test samples of $\mathrm{BDT_c}$ in bins of the $\mathrm{BDT_c}$ output for the $B^+ \to K^+\nu\bar{\nu}$ mode. On the left, no weight is applied. On the right, each simulated continuum background event is assigned a weight $w_c \equiv \mathrm{BDT_c}/(1-\mathrm{BDT_c})$.

of $\mathrm{BDT_c}$, the entire data and simulated samples are used. On the right of Figure 5.25, the simulated continuum background events are rescaled with weights $w_c \equiv \mathrm{BDT_c}/(1-\mathrm{BDT_c})$. As anticipated, the weights make the simulation distribution more similar to the data distribution.

Figure 5.26 shows how the data-simulation agreement for ΔE_{ROE} and q^2_{rec} is improved when applying the correction weights w_c to each event of the continuum background simulation in the region $0.75 < \tilde{\varepsilon}_{\mathrm{sig}} \leq 1.0$ for the charged mode selection ($B^+ \to K^+\nu\bar{\nu}$). Unless specified otherwise, the w_c correction weights are always applied to the simulated continuum background events in the following. Appendix A.3 gives an overview of all the correction weights that are applied to simulated events.

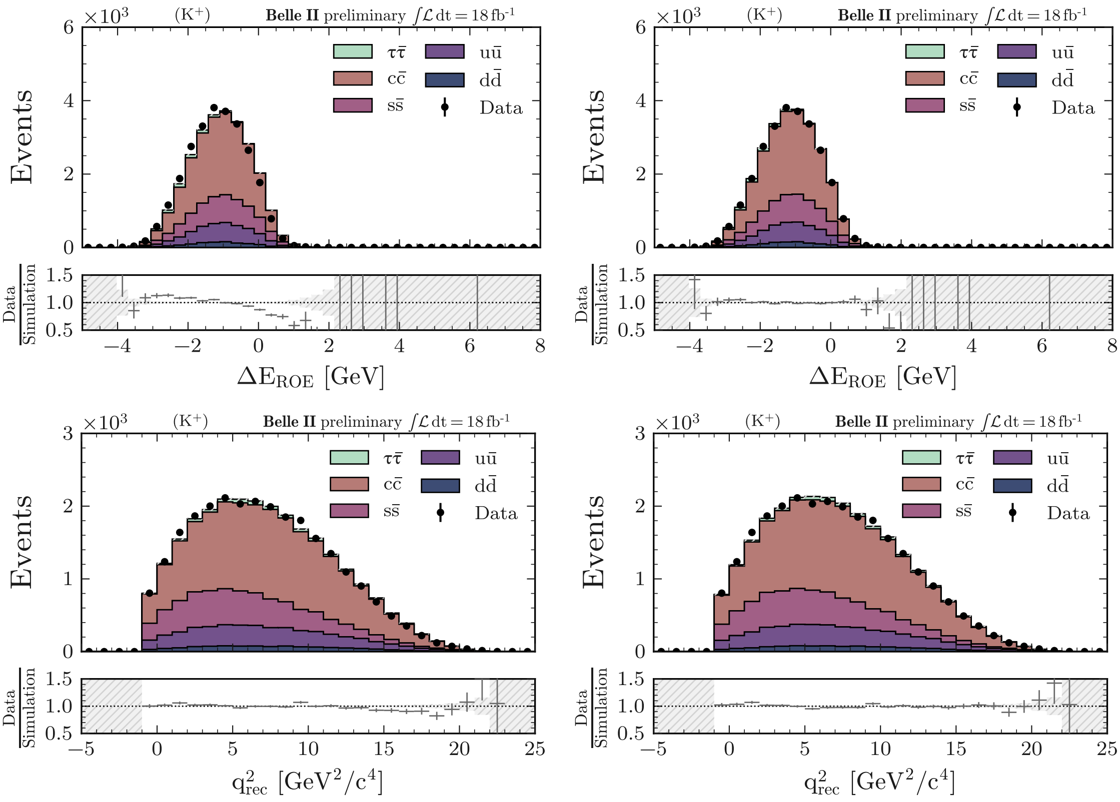


Figure 5.26.: Simulated continuum background events and off-resonance data events in bins of ΔE_{ROE} (top) and q^2_{rec} (bottom) in the region $0.75 < \tilde{\varepsilon}_{\mathrm{sig}} \leq 1.0$. On the left, no weight is applied. On the right, each simulated continuum background event is assigned a weight $w_c \equiv \mathrm{BDT_c}/(1 - \mathrm{BDT_c})$. The simulation histograms are scaled up by 33% with respect to the expectation to correct for the global normalisation discrepancy observed in this region.

Figure 5.27 compares the off-resonance data and simulated continuum background events in the bins of the signal search region after the application of the w_c weights for the charged and neutral mode selections. Data and simulation are in good agreement if the simulation is upscaled by a factor of 1.5 to correct for the normalisation discrepancy already mentioned in the previous section.

5.9.4. On-resonance data and simulation

In this section, the on-resonance data and simulation are compared in the region $0.75 < \tilde{\varepsilon}_{\mathrm{sig}} < 0.90$, where the amount of expected signal with respect to background is negligible. The data events are selected from the on-resonance sample corresponding to $189\,\mathrm{fb}^{-1}$ of

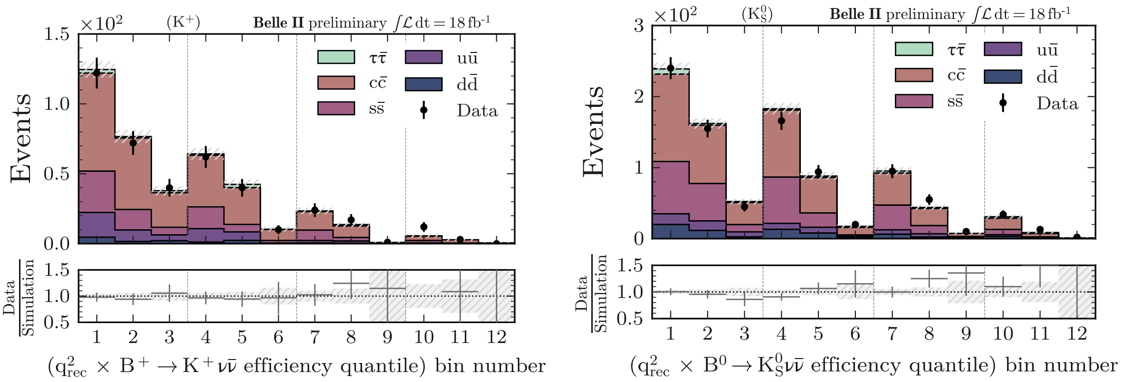


Figure 5.27.: Simulated continuum background events and off-resonance data events in bins of the signal search region for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The simulation histograms are scaled up by 50% with respect to the expectation to correct for the global normalisation discrepancy observed in this region. The bins of the signal search region are defined in Section 5.8.

integrated luminosity. The simulated background events are selected from a sample of $800\,\mathrm{fb}^{-1}$ of equivalent integrated luminosity, and the events are weighted to match the integrated luminosity of data. The simulated signal events are selected from a sample of 4×10^6 simulated $B \to K\nu\bar{\nu}$ events, and are also weighted to match the expected number of signal events in $189\,\mathrm{fb}^{-1}$ of data (Equations (5.5) and (5.6) in Section 5.7.3).

Figure 5.28 compares on-resonance data and simulation for q^2_{rec} and $\tilde{\varepsilon}_{\mathrm{sig}}$ in the region $0.75 < \tilde{\varepsilon}_{\mathrm{sig}} < 0.90$. The simulated continuum background events are weighted following the procedure described in Section 5.9.3. In addition, the continuum background normalisation discrepancy factor of 1.5 observed in Section 5.9.2 is applied to the simulated continuum background events. After this upscaling of the continuum background, the normalisation discrepancy between data and simulation is of 2% for the charged mode ($B^+ \to K^+\nu\bar{\nu}$), and 3% for the neutral mode ($B^0 \to K^0_S\nu\bar{\nu}$). This shows that the normalisation discrepancy mainly comes from the continuum background, not from the B meson background. This residual discrepancy will be treated as a source of systematic uncertainty in Section 5.10. If the simulation yields are upscaled to correct for this residual normalisation discrepancy, the agreement between data and simulation is excellent.

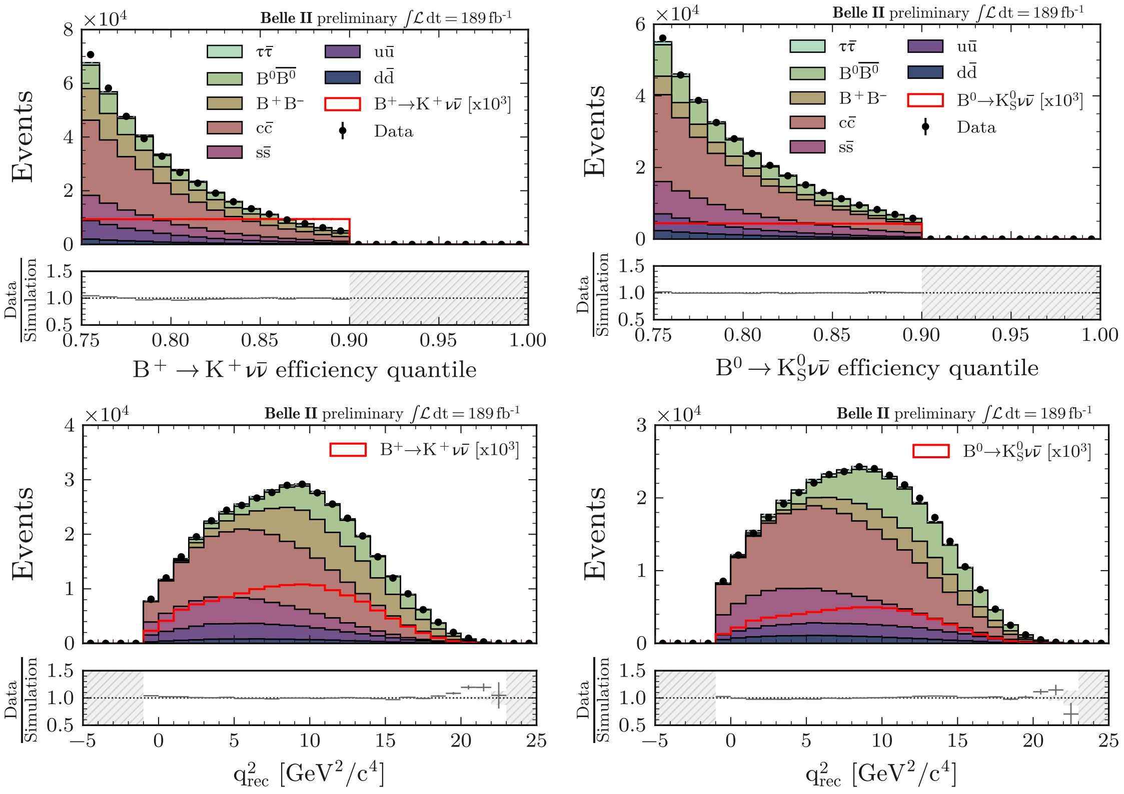


Figure 5.28.: Simulated signal and background events and on-resonance data events in bins of $\tilde{\varepsilon}_{\text{sig}}$ (top) and in bins of q^2_{rec} (bottom), for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right), in the region $0.75 < \tilde{\varepsilon}_{\text{sig}} < 0.9$. The simulated continuum background is scaled up by 33% with respect to the expectation to correct for the normalisation discrepancy observed in Section 5.9.3. After this continuum background scaling, the normalisation discrepancy between data and simulation is of 2% for the charged mode, and 3% for the neutral mode. This residual normalisation discrepancy is corrected for by upscaling all the simulated samples to match data. The signal expectation is independently upscaled by a factor of 10^3 to help the visualisation.

5.10. Systematic uncertainties

This section is dedicated to the evaluation of the systematic uncertainties that affect the measurement of the signal branching fraction. Several tools that are used in this section are introduced in the previous chapter. In particular, the reader can refer to Sections 4.3 and 4.4, which explain in particular how the binned-likelihood model is defined, and how variation vectors are derived from a covariance matrix.

To summarise briefly, a source of systematic uncertainty enters the likelihood model via a set of nuisance parameters that apply variations to the nominal expectations in the bins of the signal search region.

On one hand, nuisance parameters that apply a normalisation variation are noted μ_s, with the index s corresponding to a certain background or signal sample. In the likelihood model, the prior distribution of a normalisation nuisance parameter is a Gaussian density centered at unity (no variation) and whose width is an input to the model (Equation (4.35)).

On the other hand, nuisance parameters that apply an additive variation are noted $\theta_i^{(\mathrm{sys})}$, with (sys) designating a certain source of systematic uncertainty, and i an index used if multiple nuisance parameters are associated to a single source of systematic uncertainty. Each additive nuisance parameter enters the likelihood model together with a variation vector that applies variations that are potentially correlated among the bins of the signal search region and among the samples (Equation (4.33)). In the likelihood model, the prior distribution of an additive nuisance parameter $\theta_i^{(\mathrm{sys})}$ is a Gaussian density centered at zero (no variation) and whose width is unity (variation by one variation vector).

The considered systematic uncertainties are listed below and divided into two categories: the ones arising from physics modelling, and the ones arising from detector modelling. The uncertainties whose propagation requires a more elaborated treatment are simply named in the list and detailed in dedicated subsections.

- Physics modelling:
    - Normalisation of each background source (Section 5.10.1).
    - Modelling of the signal form factor (Section 5.10.2).
    - Uncertainties on the branching fractions of B meson decays (Section 5.10.4).
    - Modelling of the continuum background. Instead of applying the correction presented in Section 5.9.3, one could use the difference between the uncorrected and

corrected continuum background samples as a variation vector in the likelihood model with an associated nuisance parameter $\theta^{(\mathrm{BDT_c})}$. However, no impact on the result is observed. Thus no nuisance parameter is added to the model for this source of systematic uncertainty.

- Detector modelling:
    - Modelling of the PID selection efficiency (Section 5.10.3).
    - Modelling of the K^0_{S} reconstruction efficiency (Section 5.10.5).
    - Modelling of the BDT$_2$ selection efficiency for signal events. Section 5.9.1 shows an efficiency difference of 8% between data and simulation in the signal search region. For this reason, a nuisance parameter, noted $\theta^{(J/\psi)}$, is added to the likelihood model together with a variation vector that varies the signal yield in every bin of the signal search region by 8%.
    - Modelling of the track-finding efficiency (Section 5.10.6).
    - Modelling of the energy of neutral particles (Section 5.10.7).

In addition, Section 5.10.8 defines nuisance parameters used to take into account the statistical uncertainty of the simulated samples, and Section 5.10.9 gives a summary of the nuisance parameters used in the binned-likelihood model.

5.10.1. Background normalisation

In Section 5.9.2, a normalisation discrepancy of 50% is observed in the signal search region when comparing off-resonance data and simulated continuum background. For this reason, a systematic uncertainty of 50% is assigned to each continuum background sample by adding to the likelihood model five normalisation nuisance parameters $\mu_{u\bar{u}}$, $\mu_{d\bar{d}}$, $\mu_{c\bar{c}}$, $\mu_{s\bar{s}}$, $\mu_{\tau^+\tau^-}$ with a prior Gaussian distribution centered at 1.0 and with a width of 0.5.

A normalisation uncertainty of 50% is also assigned to the B meson background by adding two nuisance parameters $\mu_{B^+B^-}$ and $\mu_{B^0\overline{B}^0}$. This normalisation uncertainty for the B meson background is conservative, given that Section 5.9.4 shows that the normalisation discrepancy mainly comes from the continuum background.

The uncertainty on the measured integrated luminosity of the data samples is neglected, because it is fully covered by the background normalisation uncertainty.

5.10.2. Signal form factor

A source of systematic uncertainty comes from the theoretical uncertainty on the parametrisation of the form factor $f^+(q^2)$, that was introduced in Section 2.3.2 to compute the signal branching fraction as a function of q^2. The form factor is parametrised with a triplet of real numbers $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and an associated covariance matrix $\boldsymbol{\Sigma_\alpha}$ (Equations (2.11), (2.14), (2.16) and (2.17)).

The uncertainties on $\alpha_1, \alpha_2, \alpha_3$ are propagated to the likelihood model as follows:

1. The three orthogonal unit eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ of $\boldsymbol{\Sigma_\alpha}$ are extracted together with their respective eigenvalues σ_1^2, σ_2^2, σ_3^2 ;
2. The varied form factors $f_+(q^2, \boldsymbol{\alpha}+\boldsymbol{\sigma}_i)$ are computed for $i = 1, 2, 3$, with the variation vectors given by $\boldsymbol{\sigma}_i \equiv \sigma_i \, \boldsymbol{v}_i$;
3. The effect of the form factor variations on the shape of the signal is checked in the 12 bins of the signal search region (Figure 5.29). If the form factor is computed for a certain $\boldsymbol{\alpha}$, then the number of expected signal events in the i-th bin of the signal search region ($i = 1, ..., 12$) is noted $N_i(\boldsymbol{\alpha})$.
4. Three form-factor variation vectors, noted $\boldsymbol{\delta}_1^{(ff)}, \boldsymbol{\delta}_2^{(ff)}$, and $\boldsymbol{\delta}_3^{(ff)}$, are defined as

$$\boldsymbol{\delta}_i^{(ff)} = \begin{pmatrix} N_1(\boldsymbol{\alpha}+\boldsymbol{\sigma}_i) - N_1(\boldsymbol{\alpha}) \\ \vdots \\ N_{12}(\boldsymbol{\alpha}+\boldsymbol{\sigma}_i) - N_{12}(\boldsymbol{\alpha}) \end{pmatrix} \qquad i = 1, 2, 3. \tag{5.11}$$

5. The three variation vectors are added to the likelihood model with three respective nuisance parameters $\theta_1^{(ff)}, \theta_2^{(ff)}$, and $\theta_3^{(ff)}$.

As shown in Figure 5.29, the variations due to this source of systematic uncertainty are small, of the order of a few percents.

5.10.3. Particle identification

For the $B^+ \to K^+\nu\bar{\nu}$ mode only, a source of systematic uncertainty comes from the PID requirement that is imposed on the K^+ candidate. As explained in Section 5.7.3, each simulated event is given a weight that corrects for the efficiency difference between data and simulation of the PID selection (see also Appendix A.3). These PID weights are provided by the Belle II performance group in bins of $(p_T, \cos\theta)$, where p_T is the transverse momentum of the K^+ candidate, and θ the polar angle of the corresponding track.

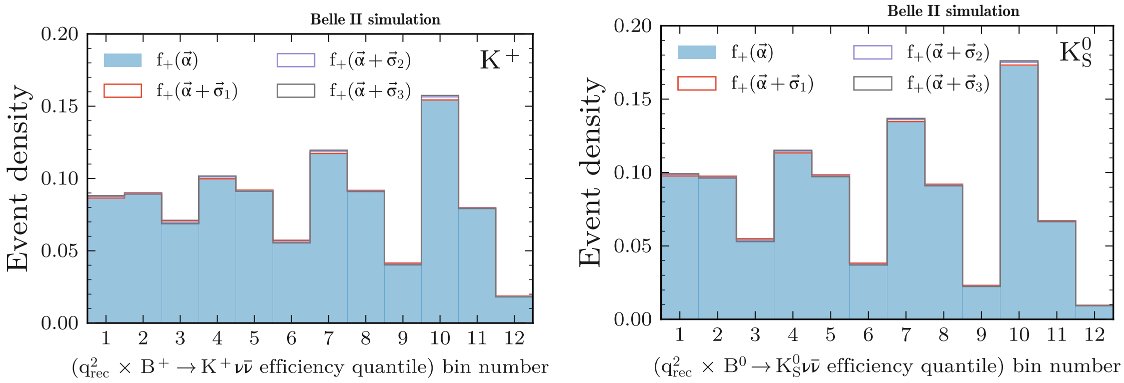


Figure 5.29.: Density of simulated signal events in bins of the signal search region for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The twelve bins of the signal search region are defined in Section 5.8. The filled histogram is weighted according to the standard model form factor, which is parametrised in Equation (2.14) as a function of a vector $\boldsymbol{\alpha}$. The three empty histograms are weighted according to varied versions of the form factor. The variations are obtained by shifting $\boldsymbol{\alpha}$ with three orthogonal vectors $\boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}, \boldsymbol{\sigma_3}$, derived from the covariance matrix of $\boldsymbol{\alpha}$ (see Equation (2.17) and following lines).

The Belle II performance group provides also uncertainties on the PID weights, and these uncertainties are propagated to the likelihood model as follows:

1. A sample of 4×10^6 simulated signal events and background samples of $800\,\mathrm{fb}^{-1}$ of equivalent integrated luminosity are reconstructed. The $e^+e^- \to \tau^+\tau^-$ sample is not considered, because it does not contribute significantly in the signal search region.

2. For each simulated event present in the signal search region, a set of 1000 varied PID weights are sampled from a probability density function whose standard deviation corresponds to the PID weight uncertainty, creating 1000 replicas of each simulated sample.

3. For each replica, the varied weights are summed up in each bin of the signal search region in order to generate 1000 pseudo-observations.

4. From the pseudo-observations obtained at the previous step, a covariance matrix Σ_{ij} is computed with Equation (4.26), where the (i, j) indices both run over the samples and the bins of the signal search region. The corresponding correlation matrix is shown on the left of Figure 5.30.

5. Seven eigenvectors of the covariance matrix corresponding to the seven largest eigenvalues are used to define variation vectors (see Equation (4.28) and following lines), and each variation vector is added to the likelihood model with a dedicated nuisance

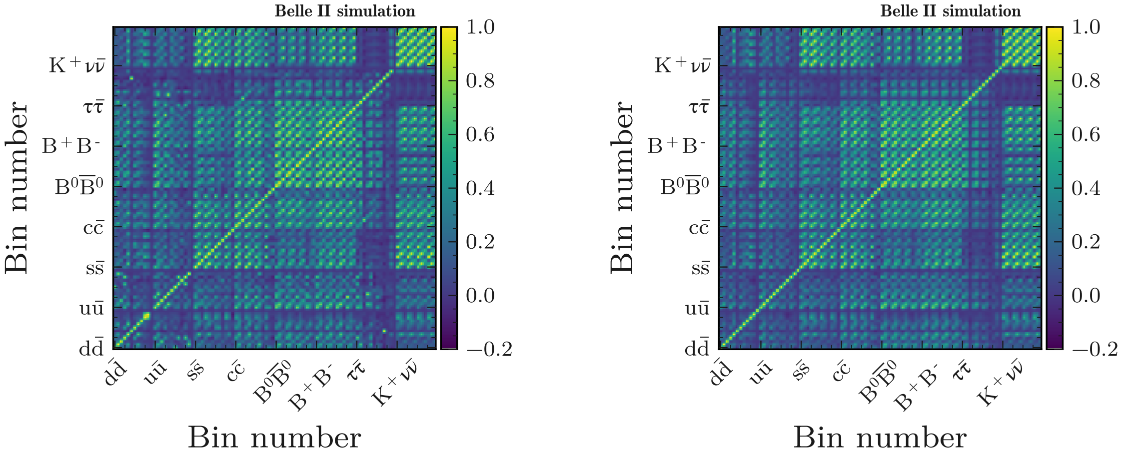


Figure 5.30.: Matrix of correlation between the expected numbers of events in the bins of the signal search region. The correlations are derived by varying the PID weight of each simulated event present in the signal search region. There are 8 samples of simulated signal or background events, and 12 bins in the signal search region, making the correlation matrix to have a size of $(8{\cdot}12)\times(8{\cdot}12) = 96\times96$. On the left, the original correlation matrix. On the right, an approximation of the correlation matrix obtained with a truncation of the covariance eigen-decomposition (Equation (4.28)). In this approximation, only the 7 eigenvectors corresponding to the 7 largest eigenvalues are considered (i.e. $t = 7$ in Equation (4.28)). The bins of the signal search region are defined in Section 5.8.

parameter, for a total of seven nuisance parameters $\theta_1^{(\mathrm{PID})}$, ..., $\theta_7^{(\mathrm{PID})}$. The number seven is chosen because it is observed that seven eigenvectors are sufficient to reasonably approximate the correlations (Figure 5.30).

6. The diagonal of the residual terms in the covariance eigen-decomposition (Equation (4.28) and following lines) are added in quadrature to the uncorrelated systematic uncertainties presented in Section 5.10.8;

Relative PID uncertainties are obtained by normalising each variation-vector component by the nominal expectation for the corresponding bin and sample. Figure 5.31 shows that the relative PID uncertainties are of the order of 0.5%.

5.10.4. Branching fraction of B meson decays

A source of systematic uncertainty comes from the limited knowledge of the branching fractions of the B meson decays contributing to the background present in the signal search region (Section 5.8.2). A list of the B meson decays contributing the most to the background is presented in Appendix A.2.

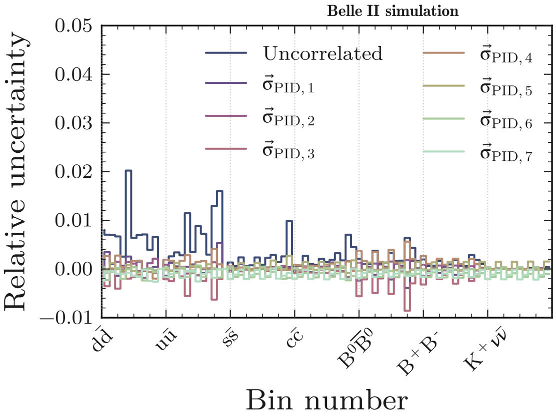


Figure 5.31.: Components of 7 vectors of correlated relative PID uncertainties and 1 vector of uncorrelated relative PID uncertainties. Each vector component corresponds to a certain signal or background sample and a certain bin of the signal search region. There are 7 samples of simulated signal or background events, and 12 bins in the signal search region, making each vector to have a dimension of $7 \cdot 12 = 84$. The bins of the signal search region are defined in Section 5.8.

The branching fraction uncertainties are propagated to the likelihood model as follows:

1. A collection $\mathcal{C}$ of branching fractions and uncertainties taken from [25] is created, starting from the decays that appear more often in the signal search region. For the $B^+ \to K^+\nu\bar{\nu}$ mode, 82% of the charged B meson decays appearing in the signal search region are included in $\mathcal{C}$, and 63% of the neutral B meson decays are included in $\mathcal{C}$. For the $B^0 \to K^0_S\nu\bar{\nu}$ mode, 78% of the charged B meson decays are included in $\mathcal{C}$, and 68% of the neutral B meson decays are included in $\mathcal{C}$.

2. Two simulated background samples, one of $e^+e^- \to B^0\bar{B}^0$ events, and one of $e^+e^- \to B^+B^-$ events, each corresponding to $800\,\mathrm{fb}^{-1}$ of equivalent integrated luminosity, are reconstructed.

3. For each simulated event e present in the signal search region, a set of 1000 weights, noted w_1^e, ..., w_{1000}^e, are computed as

$$w_i^e = \max\left(1 + \frac{s_i^e}{\mathrm{Br}(e)}, 0\right), \tag{5.12}$$

   where the maximum function ensures the positivity of the weights, $\mathrm{Br}(e)$ is the branching fraction of the B meson decay that is mis-identified as a signal decay in the event e, and s_i^e is a number sampled from a Gaussian density function centered at zero and whose width is the uncertainty on $\mathrm{Br}(e)$. This procedure creates 1000 replicas of each simulated sample.

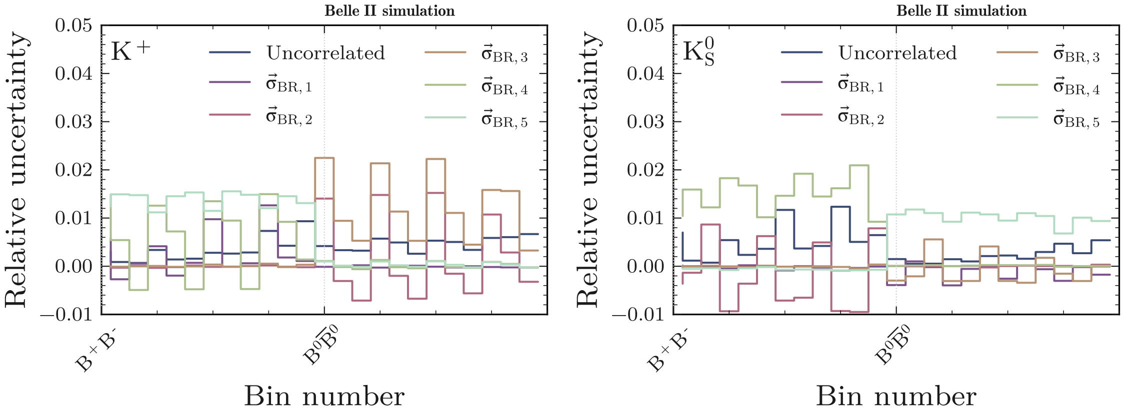


Figure 5.32.: Components of 5 vectors of correlated relative uncertainties and 1 vector of uncorrelated relative uncertainties for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The uncertainties are derived from the uncertainties on the branching fractions of the B meson decays present in the signal search region. Each vector component corresponds to a certain background sample and a certain bin of the signal search region. There are 2 samples of background events, and 12 bins in the signal search region, making each vector to have a dimension of $2 \cdot 12 = 24$. The bins of the signal search region are defined in Section 5.8.

4. For each replica, the weights are summed up in each bin of the signal search region in order to generate 1000 pseudo-observations.

5. From the pseudo-observations obtained at the previous step, a covariance matrix Σ_{ij} is computed with Equation (4.26), where the (i, j) indices both run over the samples and the bins of the signal search region.

6. Five eigenvectors of the covariance matrix corresponding to the five largest eigenvalues are used to define variation vectors (see Equation (4.28) and following lines), and each variation vector is added to the likelihood model with a dedicated nuisance parameter, for a total of five nuisance parameters $\theta_1^{(\mathrm{Br})}$, ..., $\theta_5^{(\mathrm{Br})}$. Similarly to what is done for the PID uncertainties in the previous section, it is checked that five eigenvectors are sufficient in this case to reasonably approximate the correlations.

7. The diagonal of the residual terms in the covariance eigen-decomposition (Equation (4.28) and following lines) are added in quadrature to the uncorrelated systematic uncertainties presented in Section 5.10.8;

Corresponding relative uncertainties are obtained by normalising each variation-vector component by the nominal expectation for the corresponding bin and sample. Figure 5.32 shows that the relative uncertainties are of the order of 1%.

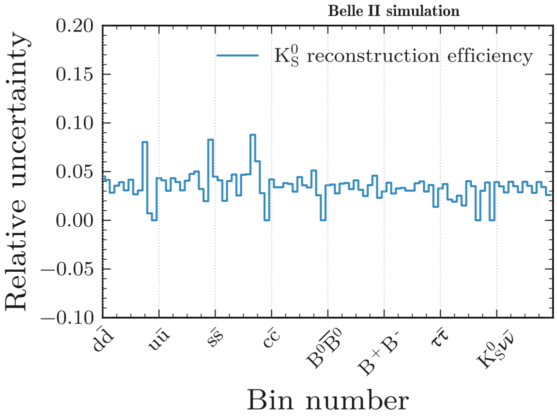


Figure 5.33.: Components of a vector of correlated relative uncertainties. The uncertainties are derived from Equation (5.13). There are 8 samples of simulated signal or background events, and 12 bins in the signal search region, making the vector to have a dimension of $8 \cdot 12 = 96$. The bins of the signal search region are defined in Section 5.8.

5.10.5. K^0_S reconstruction efficiency

A source of systematic uncertainty that is specific to the neutral mode $B^0 \to K^0_S\nu\bar{\nu}$ comes from the mis-modelling of the K^0_S reconstruction efficiency. The Belle II performance group compared data and simulation, and recommends to assign a relative uncertainty on the K^0_S reconstruction efficiency as a linear function of the K^0_S-candidate flight distance, noted distance(K^0_S):

$$\text{Relative uncertainty} = 0.004\,\text{cm}^{-1} \cdot \text{distance}(K^0_S). \tag{5.13}$$

A vector of relative uncertainties is computed on simulated events by multiplying, in each bin of the signal search region, the average value of distance(K^0_S) by $0.004\,\text{cm}^{-1}$. Figure 5.33 shows that the relative uncertainties are of the order of 5%. One nuisance parameter, noted $\theta^{(K^0_S)}$, is added to the likelihood model to take into account this source of systematic uncertainty.

5.10.6. Track-finding efficiency

A source of systematic uncertainty comes from a potential overestimation of the track-finding efficiency in simulation. The Belle II tracking group compared data and simulation and recommends to assign a systematic uncertainty of 0.9% on the track-finding efficiency. A recent estimation points to an effect of the order of 0.5%, but the more conservative value of 0.9% is assumed here.

The track-finding efficiency uncertainty is propagated to the likelihood model as follows:

1. A sample of 4×10^6 simulated signal events and background samples of $100\,\mathrm{fb}^{-1}$ of equivalent integrated luminosity pass through two different versions of the event reconstruction:

    - the standard event reconstruction;
    - an event reconstruction where each track in the event is dropped with a probability of 0.9%.

    The later modified reconstruction simulates the effect of track-finding inefficiency. Note that the $e^+e^- \to \tau^+\tau^-$ and $e^+e^- \to d\bar{d}$ samples are not considered here, because they do not contribute much in the signal search region.

2. The number of events in each bin of the signal search region and for each sample is compared between the two versions of the reconstructed samples. If, for a given bin b and a given sample s, ν_{bs} events are found after the standard reconstruction, and ν^*_{bs} events after the modified reconstruction, then the relative uncertainty is estimated to be

$$\text{Relative uncertainty} = (\nu^*_{bs} - \nu_{bs})/\nu_{bs}. \tag{5.14}$$

3. The uncertainties are the components of a vector of correlated uncertainties, and this vector is added to the likelihood model with an associated nuisance parameter $\theta^{(\text{tracking})}$.

An additional complexity arises because of the limited size of the simulated background samples corresponding to $100\,\mathrm{fb}^{-1}$ of equivalent integrated luminosity. The procedure above would lead to a significant overestimation of the systematic uncertainty, because the statistical uncertainty of the simulated background samples contributes to the estimate given in Equation (5.14). To overcome this problem, the distributions obtained with the standard and modified reconstructions are both smoothed before their difference is computed.

The smoothing method consists in replacing the distributions with their respective Gaussian kernel density estimators, which are introduced in Section 4.6. Since the signal search region is defined in bins of q^2_{rec} and $\tilde{\varepsilon}_{\text{sig}}$, a two-dimensional kernel density estimator $f_h(q^2_{\text{rec}}, \tilde{\varepsilon}_{\text{sig}})$ is computed for each distribution, where h is the smoothing factor of Equation (4.42). Several values of the smoothing factor between 0.1 and 1.0 are tested. If the smoothing is too weak, the effect of statistical uncertainty is not reduced; if the smoothing is too strong, the systematic uncertainty may be underestimated. A reasonable

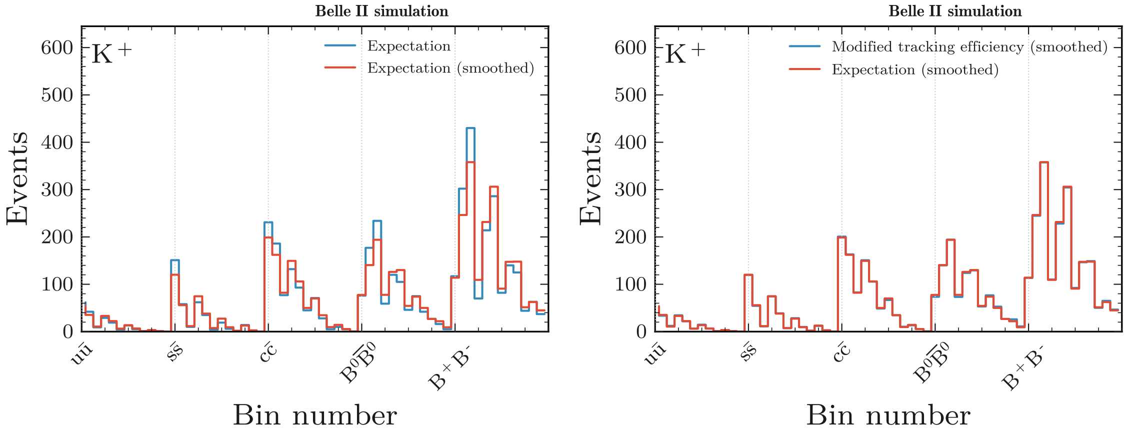


Figure 5.34.: Number of events in each bin of the $B^+ \to K^+\nu\bar{\nu}$ signal search region for 5 simulated background samples of 100 fb^{-1} of equivalent integrated luminosity. The bins of the signal search region are defined in Section 5.8. On the left, the effect of the smoothing procedure is illustrated by comparing the expected number of events before and after the smoothing procedure. On the right, simulated background samples reconstructed with a modified track-finding efficiency are compared to the nominal expectations; both distributions are smoothed and visually nearly overlap.

value for h is found to be 0.5. No smoothing of the simulated signal sample is necessary, because the statistical uncertainty of this sample is small in the signal search region.

The left part of Figure 5.34 illustrates the smoothing procedure, and the right part of Figure 5.34 compares the smoothed distributions resulting from the standard and modified reconstructions. The top part of Figure 5.35 shows the relative systematic uncertainties obtained with the method presented above. Uncertainties obtained with and without smoothing are compared to the statistical uncertainties of the simulated samples. For the signal sample, the statistical uncertainty is small, and the relative systematic uncertainty due to the mis-modelling of the track-finding efficiency is stable and of the order of a few percents. For the background samples, the systematic uncertainties fluctuate more in regions of large statistical uncertainty. These fluctuations, which motivate the smoothing procedure, are likely due to migrations of individual events between bins, which cause a large effect when the total number of events in a bin is small.

5.10.7. Neutral particle energy

A source of systematic uncertainty arises from the uncertainty on the energy measured by the Belle II calorimeter (ECL), namely the energy of photons and neutral hadrons.

The Belle II performance group recommends to assign an uncertainty of the order of 0.5% to the photon energy. For the neutral hadron energy, no recommendation is given. A comparison between data and simulation of variables related to the energy of the ECL clusters indicates that the uncertainty on the neutral hadron energy is of the order of 10%.

The same procedure as for the track-finding efficiency (Section 5.10.6) is independently applied for these sources of systematic uncertainties. This time, instead of reconstructing the samples with a modified track-finding efficiency, the modified parameter is the energy of the simulated photons (modified by 0.5%) in one case, and the energy of the simulated neutral hadrons (modified by 10%) in the other case. After following the steps of the procedure, two vectors of correlated uncertainties are added to the likelihood model with two respective nuisance parameters $\theta^{(\gamma)}$ and $\theta^{(\text{neutral hadrons})}$.

The middle and bottom parts of Figure 5.35 show the obtained relative systematic uncertainties. Uncertainties obtained with and without the smoothing procedure presented in Section 5.10.6 are compared to the statistical uncertainties due to the limited size of the simulated background samples.

5.10.8. Simulation statistical uncertainty

In addition to the systematic uncertainties listed above, a set of nuisance parameters are used in the binned-likelihood model to take into account the statistical uncertainty of the simulated samples. These statistical uncertainties are considered uncorrelated among the simulated samples and the bins of the signal search region, implying that one nuisance parameter is needed for each simulated sample and for each of the 12 bins of the signal search region.

For the signal sample, the $e^+e^- \to B^+B^-$ sample, and the $e^+e^- \to B^0\overline{B}^0$ sample, a total of $3 \cdot 12 = 36$ nuisance parameters are needed.

For each continuum background category, two independent simulated samples are compared to the data in the binned-likelihood model:

- The first simulated sample contributes to the expectations that are compared to the on-resonance data in the 12 bins of the signal search region.
- The second simulated sample contributes to the expectations that are compared to the off-resonance data in 12 bins that replicate the bins of the signal search region.

Table 5.11.: Nuisance parameters defined in Section 5.10. The particle identification uncertainty applies only to the charged mode ($B^+ \to K^+\nu\bar{\nu}$), and the K_S^0 reconstruction efficiency applies only to the neutral mode ($B^0 \to K_S^0\nu\bar{\nu}$).

Source of systematic uncertainty	Nuisance parameters	Count
Background normalisation	$\mu_{u\bar{u}}$, $\mu_{d\bar{d}}$, $\mu_{c\bar{c}}$, $\mu_{s\bar{s}}$, $\mu_{\tau^+\tau^-}$, $\mu_{B^+B^-}$, $\mu_{B^0\bar{B}^0}$	7
Signal form factor	$\theta_i^{(ff)}$ $i = 1, 2, 3$	3
Particle identification	$\theta_i^{(\mathrm{PID})}$ $i = 1, ..., 7$	7
B meson branching fractions	$\theta_i^{(\mathrm{Br})}$ $i = 1, ..., 5$	5
K_S^0 reconstruction efficiency	$\theta^{(K_S^0)}$	1
Track-finding efficiency	$\theta^{(\mathrm{tracking})}$	1
Photon energy	$\theta^{(\gamma)}$	1
Neutral hadron energy	$\theta^{(\mathrm{neutral\ hadrons})}$	1
Signal selection efficiency	$\theta^{(J/\psi)}$	1
Simulation statistical uncertainty	$\theta_i^{(\mathrm{sim.\,stat.})}$ $i = 1, ..., 156$	156
Total ($B^+ \to K^+\nu\bar{\nu}$ mode)		182
Total ($B^0 \to K_S^0\nu\bar{\nu}$ mode)		176

Since there are five continuum background categories ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$, and $e^+e^- \to \tau^+\tau^-$), a total of $5 \cdot (12 + 12) = 120$ nuisance parameters are needed.

In total, $36 + 120 = 156$ nuisance parameters of uncorrelated uncertainties $\theta_i^{(\mathrm{sim.\,stat.})}$ are used in the binned-likelihood model.

5.10.9. Summary

Table 5.11 lists all the nuisance parameters defined in the above sections. In total, the binned-likelihood model of the charged mode ($B^+ \to K^+\nu\bar{\nu}$) has 182 nuisance parameters, and the binned-likelihood model of the neutral mode ($B^0 \to K_S^0\nu\bar{\nu}$) has 176 nuisance parameters.

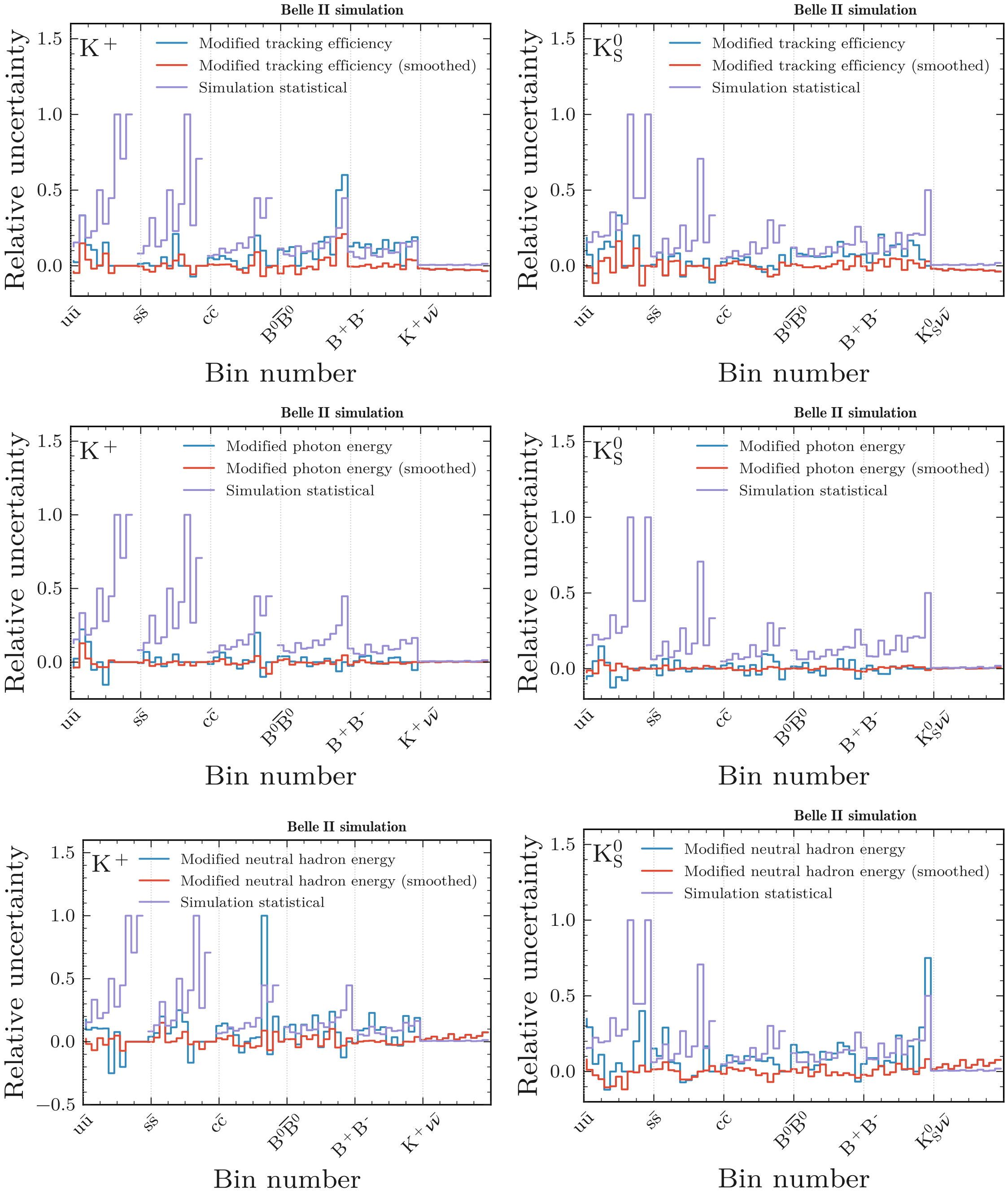


Figure 5.35.: Components of vectors of relative uncertainties for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The uncertainties are derived from a modification of the track-finding efficiency (top), of the photon energy (middle), or of the neutral hadron energy (bottom). There are 6 samples of simulated signal or background events, and 12 bins in the signal search region (Section 5.8), implying that each vector has a dimension of $6 \cdot 12 = 72$. In each plot, the vectors of relative uncertainties are shown with and without the smoothing procedure, and compared to the relative statistical uncertainty coming from the limited size of the simulated samples. Without the smoothing procedure, the systematic uncertainties are overestimated, because they include effects from statistical uncertainties of the simulated samples.

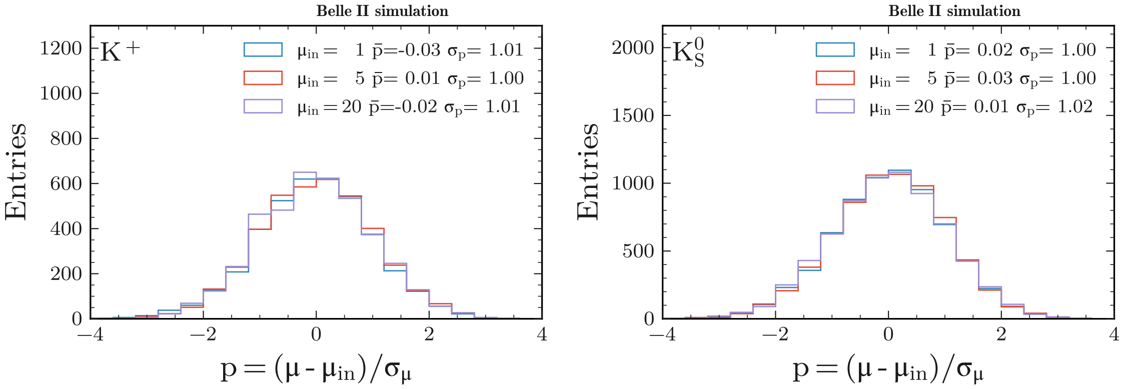


Figure 5.36.: Pull distributions of pseudo-observation studies using the `sghf` fit, for the $B^+ \to K^+\nu\bar{\nu}$ mode (left) and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). Pseudo-observations are generated with both Poisson statistical and Gaussian systematic fluctuations around the expectations. An injected signal strength is varied between $\mu_{\text{in}} = 1$ (SM expectation), $\mu_{\text{in}} = 5$, and $\mu_{\text{in}} = 20$. The pull is defined as $p = (\mu - \mu_{\text{in}})/\sigma_\mu$, where μ and σ_μ are the fit signal strength and its uncertainty. No significant bias is observed for the mean value of the pull $\bar{p}$, and the standard deviation σ_p is consistent with unity.

5.11. Model fitting and expected upper limit

At this point, all the ingredients that enter the binned-likelihood model, whose functional expression is given by Equation (4.34) in Section 4.4, are ready:

- The signal search region consists of $4 \times 3 = 12$ bins in the $\tilde{\varepsilon}_{\text{sig}} \times q^2_{\text{rec}}$ space.
- Simulated background events, corresponding to an equivalent integrated luminosity of $800\,\text{fb}^{-1}$, and 4×10^6 signal events, provide the expected yields in the 12 bins of the signal search region, once properly weighted to match the integrated luminosity of the on-resonance data ($189\,\text{fb}^{-1}$).
- Simulated continuum background events, corresponding to an equivalent integrated luminosity of $100\,\text{fb}^{-1}$, provide the expected yields in 12 bins that replicate those of the signal search region, once properly weighted to match the integrated luminosity of the off-resonance data ($18\,\text{fb}^{-1}$).
- The binned-likelihood model compares data and simulation in the 24 bins defined above.

The degrees of freedom of the binned-likelihood model are the signal strength μ, which is the parameter of interest (defined by Equation (4.29) in Section 4.4), and nuisance parameters that modify the expected yields according to the modelling of the systematic

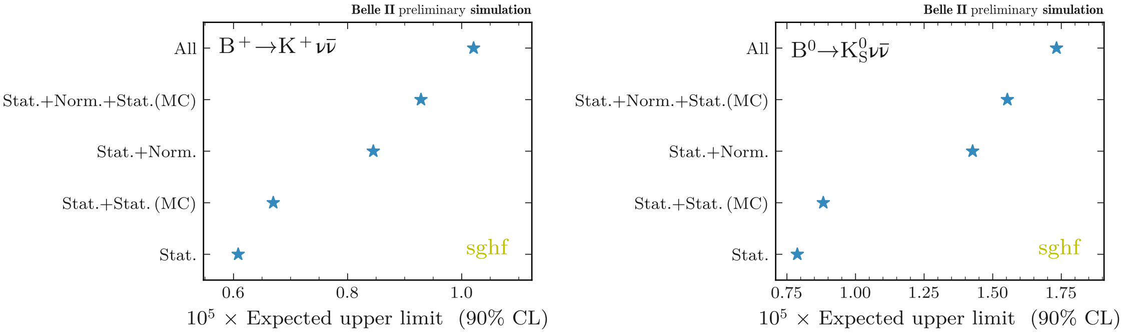


Figure 5.37.: Expected upper limits on Br($B \to K\nu\bar{\nu}$) at the 90% confidence level for the background-only hypothesis, obtained using the `sghf` fit, for the $B^+ \to K^+\nu\bar{\nu}$ mode (left), and the $B^0 \to K^0_S\nu\bar{\nu}$ mode (right). The limits are computed as $\mathrm{Br}(B \to K\nu\bar{\nu})_{\mathrm{SM}} \cdot \sigma_\mu \cdot 1.645$, where σ_μ is the uncertainty on the signal strength μ, and 1.645 is the conversion factor from the 68% confidence level to the 90% confidence level. The limits are obtained using only expected data statistical uncertainty (Stat.), data and simulation statistical uncertainties (Stat.+Stat.(MC)), data statistical and background normalisation uncertainties (Stat.+Norm.), data statistical, background normalisation and simulation statistical uncertainties (Stat+Norm.+Stat.(MC)), and total uncertainties (All).

uncertainties (Section 5.10). A unique set of continuum normalisation nuisance parameters $\{\mu_{u\bar{u}}, \mu_{d\bar{d}}, \mu_{c\bar{c}}, \mu_{s\bar{s}}, \mu_{\tau^+\tau^-}\}$ is shared by the 12 on-resonance bins and the 12 off-resonance bins, so that the off-resonance data sample helps to constrain the normalisation of the continuum background.

Two implementations of the binned-likelihood model, that are introduced at the end of Section 4.4 and give similar results, are used: the `pyhf` model, which is a product of Poisson densities, and the `sghf` model, which approximates the Poisson densities by Gaussian densities. Even though the `pyhf` model is used to determine the final result, fits with the `sghf` models take less computing time. For this reason, the `sghf` model is preferred for tests that require a large number of fits.

Before opening the box, meaning looking at the on-resonance data in the signal search region, a few more tests are carried out:

- Several thousands of pseudo-observations are generated to verify that the maximum likelihood fit delivers an unbiased value of the signal strength μ and of its uncertainty. The pseudo-observations are generated with Poisson statistical fluctuations and Gaussian systematic fluctuations around the expectations. In addition, pseudo-observations are generated for three injected values of μ: $\mu_{\mathrm{in}} = 1, 5, 20$. Figure 5.36 shows that a fit of the `sghf` model to the pseudo-observations lead to an unbiased

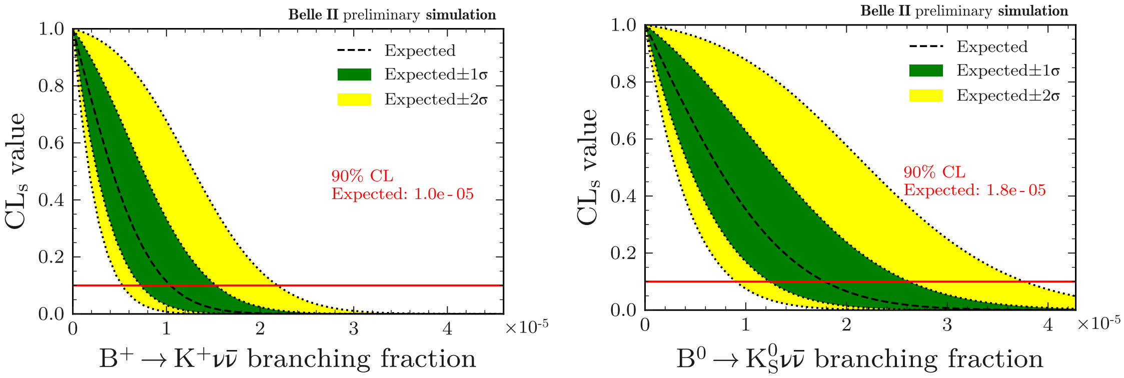


Figure 5.38.: Expected $\mathrm{CL_s}$ value as a function of the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ (left) and $B^0 \to K^0_S\nu\bar{\nu}$ (right) using the `pyhf` fit, and corresponding upper limits at the 90% confidence level (CL). The expected upper limits are derived for the background-only hypothesis.

value of μ and of its uncertainty. It is shown in Appendix A.5 that the `pyhf` model and the `sghf` model deliver similar results.

- The impact of the systematic and statistical uncertainties on the expected upper limit on $\mathrm{Br}(B \to K\nu\bar{\nu})$ is studied by fits of the `sghf` model to the expectations, for multiple combinations of uncertainties included or not (Figure 5.37). The systematic uncertainty contributing the most is the 50% normalisation uncertainty on each background source (Section 5.10.1).
- The $\mathrm{CL_s}$ method, presented in Section 4.5, is followed to determine an expected upper limit on $\mathrm{Br}(B \to K\nu\bar{\nu})$ with the `pyhf` model. The result is shown in Figure 5.38. The expected upper limits on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$ are determined to be 1.0×10^{-5} and 1.8×10^{-5} at the 90% confidence level, respectively. For $\mathrm{Br}(B^+ \to K^+\nu\bar{\nu})$, the expected upper limit is better than the current world-best observed upper limit, which is 1.6×10^{-5} at the 90% confidence level [25].
- An additional check is the study of the stability of the data yield per unit of integrated luminosity in the signal search region as a function of the data taking period. This specific study is presented in Appendix A.4 and shows that the data yield per unit of integrated luminosity is stable.

Time did not allow to look at the data in the signal search region when selecting events from the full data sample of $189\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance. For this reason, the result of a first iteration of the inclusive tagging method is presented in the next section, with data samples of $63\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and $9\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance.

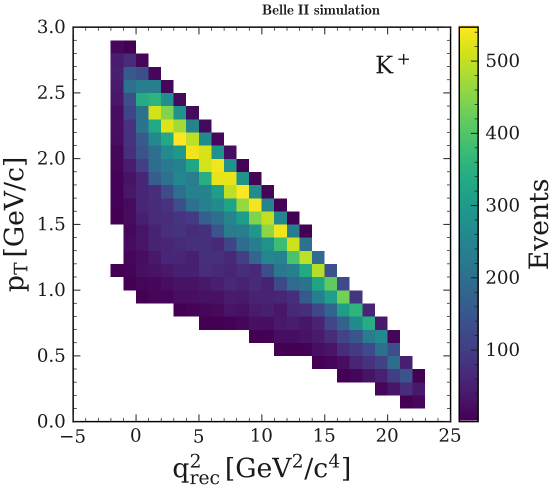


Figure 5.39.: Transverse momentum (p_T) and reconstructed invariant mass squared of the two-neutrino system (q^2_{rec}) in 10^5 simulated $B^+ \to K^+\nu\bar{\nu}$ decays.

5.12. Results of a first iteration of the method

This section presents the result of a first iteration of the method, on data samples of $63\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and $9\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance. This first iteration was published in [118] and has a few differences with respect to what is developed in the previous sections. In particular,

- Only $B^+ \to K^+\nu\bar{\nu}$ decays are selected, not $B^0 \to K^0_S\nu\bar{\nu}$ decays.
- Instead of selecting the signal candidate as the one with the smallest q^2_{rec}, the one with the highest transverse momentum (p_T) is selected. As illustrated by Figure 5.39, this difference is minor, since the two variables are strongly anti-correlated.
- The signal search region is defined by 12 bins in the $\mathrm{BDT}_2 \times p_T$ space, instead of 12 bins in the $\tilde{\varepsilon}_{\mathrm{sig}} \times q^2_{\mathrm{rec}}$ space. The exact definition of the signal search region bins is $\mathrm{BDT}_2 \in [0.93, 0.95, 0.97, 0.99, 1.0]$ and $p_T(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$.
- In the first iteration of the method, BDT_1, BDT_2, and BDT_c share a common set of input variables, which include all the variables that are considered in Section 5.7.2. No significant impact is expected from this difference, because Section 5.7.2 shows than the removal of the least discriminative variables does not impact the classification performance.

Before looking at the on-resonance data in the signal search region, the compatibility between the data and the `pyhf` model is quantified as follows:

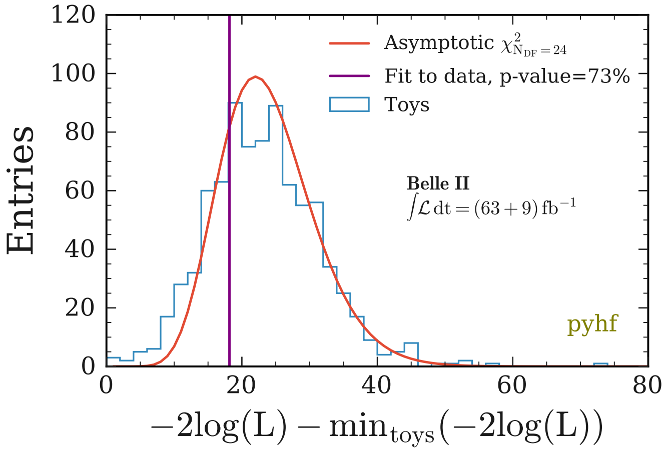


Figure 5.40.: Distributions of the fit negative log-likelihood for pseudo-observation studies using the `pyhf` fit. Pseudo-observations (toys) are generated with Poisson statistical and Gaussian systematic fluctuations around the expectations. The statistical fluctuations are based on the observed data yield. The p-value of the fit to the data (73%) is the fraction of pseudo-observations that yield fits with a larger negative log-likelihood value than that of the data fit. Since the fit region consists of $2 \cdot 12 = 24$ bins, the asymptotic χ^2 distribution for 24 degrees of freedom is overlaid.

1. Several thousands of pseudo-observations are generated with Poisson statistical fluctuations and Gaussian systematic fluctuations around the expectations in the signal search region, using the data statistical uncertainties for the Poisson statistical fluctuations.
2. The `pyhf` model is fit to each pseudo-observation, and the found minimum of the negative log-likelihood is stored.
3. The `pyhf` model is fit to data, and the found minimum of the negative log-likelihood is stored. The fit value of the signal strength is kept hidden.
4. Comparing the minimum of the negative log-likelihood found with the data and the pseudo-observations allows to define a p-value that quantifies the data-model compatibility (Figure 5.40).

The above study shows an excellent compatibility between the model and the data ($p =$ 73%), allowing to look at the on-resonance data in the signal search region.

The result of the fit to the 12+12 bins of on-resonance and off-resonance data is shown in Figure 5.41. The observed signal strength μ is

$$\mu = 4.2^{+3.4}_{-3.2} = 4.2^{+2.9}_{-2.8}(\mathrm{stat})^{+1.8}_{-1.6}(\mathrm{syst}), \tag{5.15}$$

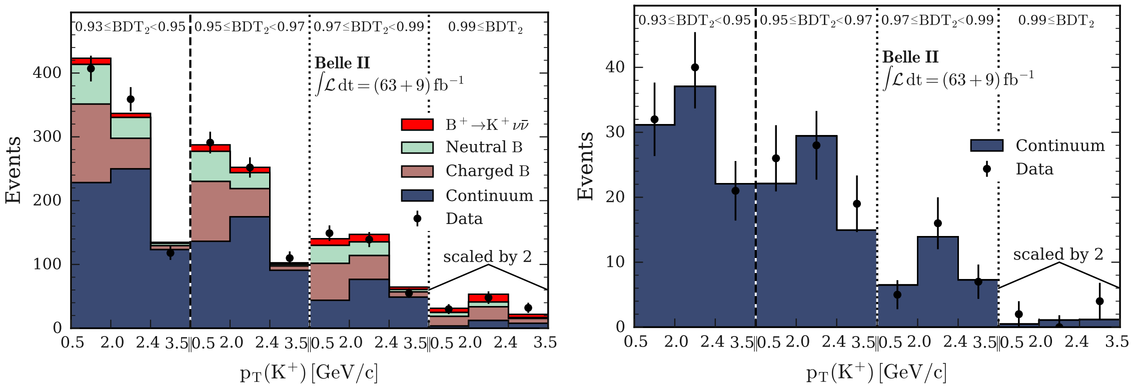


Figure 5.41.: Yields in data and as predicted by the simultaneous fit to the on-resonance data (left) and off-resonance data (right), corresponding to an integrated luminosity of 63 fb^{-1} and 9 fb^{-1}, respectively. The predicted yields are shown individually for charged and neutral B meson decays and the sum of the five continuum background categories. The signal search region consists of $4 \times 3 = 12$ bins: 4 bins defined on the classifier output, $\mathrm{BDT}_2 \in [0.93, 0.95, 0.97, 0.99, 1.0]$, and 3 bins on the transverse momentum of the signal kaon candidate, $p_T(K^+) \in [0.5, 2.0, 2.4, 3.5]$ GeV/c. Yields in the rightmost three bins are upscaled by a factor of 2. Figure published in [118].

meaning that the observed signal is 4.2 times larger than the SM expectation, but with a large uncertainty that makes this measurement compatible with the SM within one standard deviation. If translated into a measured branching fraction, the result is

$$\mathrm{Br}(B^+ \to K^+\nu\bar{\nu}) = \left[1.9^{+1.6}_{-1.5}\right] \times 10^{-5} = \left[1.9^{+1.3}_{-1.3}(\mathrm{stat})^{+0.8}_{-0.7}(\mathrm{syst})\right] \times 10^{-5}. \tag{5.16}$$

Information about the post-fit normalisation nuisance parameters can be found in Appendix A.6.

In Equation (5.15), the total uncertainty on μ is obtained with a profile likelihood scan, which checks how the negative log-likelihood varies when the model is fit to the data with fixed values of μ around the value that minimises the negative log-likelihood (Figure 5.42, left). The statistical uncertainty on μ is determined by taking the standard deviation of a sample of signal strengths determined from fits on pseudo-observations that are generated as Poisson statistical fluctuations around the observed data yields. The systematic uncertainty $\sigma_{\mathrm{syst}}(\mu)$ is determined by subtracting in quadrature the total uncertainty $\sigma(\mu)$ and the statistical uncertainty $\sigma_{\mathrm{stat}}(\mu)$:

$$\sigma_{\mathrm{syst}}(\mu) = \sqrt{\sigma^2(\mu) - \sigma^2_{\mathrm{stat}}(\mu)}. \tag{5.17}$$

The precision of the obtained measurement is comparable to previous results (Figure 5.42, right), even if the size of the used data sample is much smaller. In order to compare

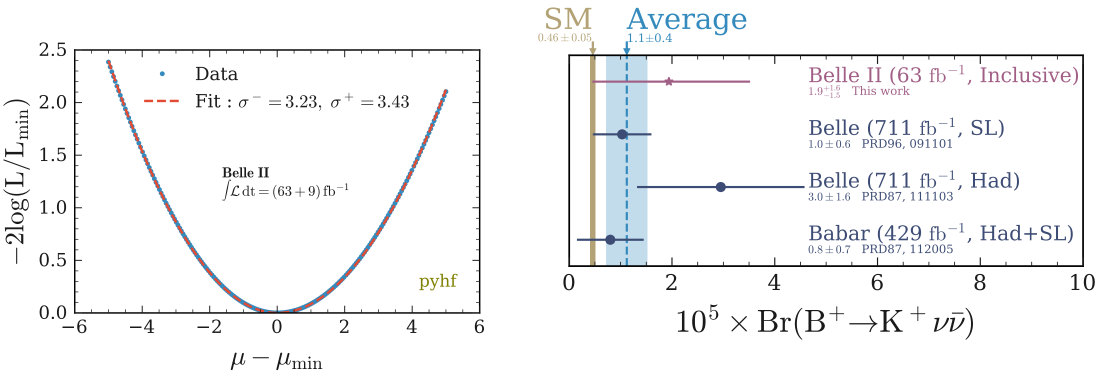


Figure 5.42.: On the left, scan of the signal strength μ. Each point is obtained by fixing μ and minimising the negative log-likelihood L of the `pyhf` model with respect to all the other model parameters. The asymmetric uncertainty on μ is estimated by fitting the collection of points with an asymmetric parabola $f(x) = (x/\sigma^-)^2$ for $x < 0$, and $f(x) = (x/\sigma^+)^2$ for $x \geq 0$. The fit yields $\sigma^- = 3.23$ and $\sigma^+ = 3.43$. On the right, comparison of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction measured by Belle II and previous experiments [74, 75, 77] with the SM prediction [60]. The values reported for Belle are computed based on the quoted observed number of events and efficiency. The weighted average is computed assuming that uncertainties are uncorrelated. Figure published in [118].

the sensitivity of the inclusive tagging method with the sensitivity of other methods, the product $\sigma_{\mathrm{Br}} \cdot \sqrt{L}$ is computed, where σ_{Br} is the uncertainty on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$, and L is the integrated luminosity of the data sample used to make the measurement. This metric assumes that the total uncertainty scales as the inverse of the square root of the integrated luminosity, and is smaller for a more sensitive method. Table 5.12 compares $\sigma_{\mathrm{Br}} \cdot \sqrt{L}$ for the obtained result and previous published results. With respect to this metric, the inclusive tagging method is a factor of 3.5 better than the hadronic tagging of [75], approximately 20% better than the semileptonic tagging of [77], and approximately 10% better than the combined hadronic and semileptonic tagging of [74].

Since no significant signal is observed, the $\mathrm{CL_s}$ method, presented in Section 4.5, is followed to determine an observed upper limit on $\mathrm{Br}(B^+ \to K^+\nu\bar{\nu})$ with the `pyhf` model (Figure 5.43). The observed upper limit is

$$\mathrm{Br}(B^+ \to K^+\nu\bar{\nu}) < 4.1 \times 10^{-5} \qquad \text{at the 90\% confidence level.} \tag{5.18}$$

Table 5.12.: Experimental results of searches for $B^+ \to K^+\nu\bar{\nu}$ decays. Are given the name of the experiment, the integrated luminosity of the on-resonance data sample (L), the employed tagging method (combined stands for the combination of the inclusive and semileptonic tagging methods), the symmetrised uncertainty on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction (σ_{Br}), and the product $\sigma_{\mathrm{Br}} \cdot \sqrt{L}$. The final line is the result of this work.

Experiment	L [fb^{-1}]	Method	σ_{Br} [10^{-5}]	$\sigma_{\mathrm{Br}} \cdot \sqrt{L}$ [$10^{-4}\sqrt{\mathrm{fb}^{-1}}$]	Ref.
Belle	711	Hadronic	1.63	4.36	[75]
Belle	711	Semileptonic	0.57	1.51	[77]
Babar	429	Combined	0.65	1.35	[74]
Belle II	63	Inclusive	1.55	1.23	[118]

5.13. Discussion

The obtained result shows that the inclusive tagging method is particularly suitable to search for rare decays if the size of the data sample is limited. The main reason for this is the high efficiency of the signal selection: 4% at the point of maximum significance (Figure 5.15 in Section 5.7.3), to be compared with the signal selection efficiencies of the hadronic and the semileptonic tagging methods, which are well below 1% (Table 2.3 in Section 2.3.4).

The success of the inclusive tagging method opens new opportunities in the search for rare decays with neutrinos in the final state or any kind of invisible energy. As already mentioned, the inclusive tagging method presented in this thesis is inspired by the method used by the Belle collaboration in a search for $B^+ \to \mu^+\nu_\mu$ decays [111]. Examples of other decays that could benefit from the method include $B \to K^*\nu\bar{\nu}$, $B \to K^{(*)}\tau^+\tau^-$, and $B \to KJ/\psi(\to \nu\bar{\nu})$.

A disadvantage of the developed inclusive tagging method is that it is channel-specific: a dedicated reconstruction and selection chain is defined for each mode ($B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$). In the future, it will be interesting to define a first classifier (BDT$_1$) trained with a mixture of simulated signal decays, so that the first step of the background rejection could be shared by multiple targeted signal decays. After a common event selection with BDT$_1$, a second classifier (BDT$_2$) could be trained once per targeted signal decay to finalise the event selection. This updated strategy would be similar to the one used by the Belle collaboration in a search for $B \to H\nu\bar{\nu}$ decays with a semileptonic tagging method,

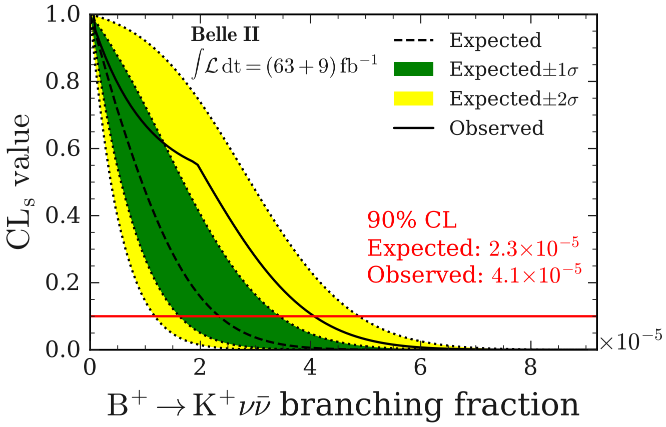


Figure 5.43.: $\mathrm{CL_s}$ value as a function of the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ for expected and observed signal yields and corresponding upper limits at the 90% confidence level (CL). The expected upper limit is derived for the background-only hypothesis. The observed upper limit is derived from a simultaneous fit to the on-resonance and off-resonance data, corresponding to an integrated luminosity of $63\,\mathrm{fb}^{-1}$ and $9\,\mathrm{fb}^{-1}$, respectively. The change of slope of the observed $\mathrm{CL_s}$ curve around 1.9×10^{-5} is an expected feature of the $\mathrm{CL_s}$ method due to the presence of a threshold in the definition of the test statistic q_μ (Equation (4.37) in Section 4.5). Figure published in [118].

where $H \in \{K^+, K^0_S, K^{*+}, K^{*0}, \pi^+, \pi^0, \rho^+, \rho^0\}$ [77]. In this case, the selection of the accompanying B meson was common among the channels, and channel-specific classifiers were trained for the final stage of the event selection.

Currently, the main source of systematic uncertainty is the 50% normalisation uncertainty on each of the background categories. As already mentioned in Section 5.10.1, this value of 50% is too conservative for the B meson background. For the continuum background, more studies are needed to understand the cause of the normalisation discrepancy, which could be a mis-modelling of the quark fragmentation (Section 5.9.2).

To improve the inclusive tagging method, it will also be interesting in the future to test other types of classification algorithms than boosted decision trees, such as neural networks [94], or a combination of multiple classifiers. In addition, a combination of different tagging methods (hadronic, semileptonic and inclusive) may increase the sensitivity of the measurement.

6

Conclusion

This thesis has documented a search for $B \to K\nu\bar{\nu}$ decays at the Belle II experiment with an inclusive tagging method.

Chapter 2 has shown how the standard model of particle physics predicts the branching fraction of this decay via an effective Hamiltonian. An experimental determination of this branching fraction is not only an important test of the theory. It is also sensitive to a potential violation of the lepton flavour universality, because all the neutrino flavours contribute to $B \to K\nu\bar{\nu}$, and it imposes constraints on several new physics models predicting the existence of invisible particles, since a search for $B \to K\nu\bar{\nu}$ is also a search for $B \to K +$ invisible.

Chapter 3 has given an overview of the employed experimental setup, namely the SuperKEKB accelerator, which produces pairs of B mesons by colliding electrons and positrons at the energy of the $\Upsilon(4S)$ resonance, and the Belle II detector, which is designed to detect the decay products of B mesons.

Chapter 4 has introduced several analysis tools and techniques, and in particular, boosted decision tree classifiers and input variables that are function of the momentum distribution in an event. These variables are central in the selection of $B \to K\nu\bar{\nu}$ decays, because an event containing a $B \to K\nu\bar{\nu}$ decay tends to exhibit a large missing momentum due to the undetected neutrino pair.

Chapter 5 has detailed each step of the selection of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$ decays in a data sample of $189\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and a complementary sample of $18\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance. A selection strategy that had never been employed to search for these decays, the inclusive tagging method, is defined. The procedure is validated by comparing data and simulation in several regions of the phase space, and by checking that the maximum-likelihood fit procedure provides an unbiased measurement of the branching fraction of $B \to K\nu\bar{\nu}$. For this amount of data, the expected upper limits on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_S\nu\bar{\nu}$ are determined from simulation to be 1.0×10^{-5} and 1.8×10^{-5} at the 90% confidence level, respectively.

In a first iteration of the method, restricted to data samples of $63\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and $9\,\mathrm{fb}^{-1}$ collected 60 MeV below the resonance, no signal is observed, and an upper limit on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ is determined to be 4.1×10^{-5} at the 90% confidence level. This result is published in [118]. If translated into an uncertainty on the branching fraction, the inclusive tagging method is a factor of 3.5 better per integrated luminosity than the hadronic tagging of [75], approximately 20% better than the semileptonic tagging of [77], and approximately 10% better than the combined hadronic and semileptonic tagging of [74].

The success of the inclusive tagging method opens new opportunities for the study of rare decays involving neutrinos in the final state. Examples of such decays include $B \to K^*\nu\bar{\nu}$, $B \to K\tau^+\tau^-$, and $B \to KJ/\psi(\to \nu\bar{\nu})$. Method improvements are also possible. By combining different types of classifiers (decision trees, neutral networks) or different types of B meson tagging (hadronic, semileptonic, inclusive), the overall sensitivity may increase. With the new data that the Belle II and the LHCb experiments are planning to collect in the next decade, long-awaited results in flavour physics are expected in a near future.

On a broader perspective, the success of the standard model in predicting a rich variety of experimental results is undeniable. Yet, by looking at the history of science, it is very likely that most of our current understanding of the universe in general, and of particle physics in particular, is wrong. When feeling satisfied to see symmetries in the theory, are we so different from Kepler, who was postulating that the planet orbits were contained in an ideal succession of platonic solids? If we are not to trust too much our interpretations, we can at least believe in the results of experiments. To quote Cicero in *De natura deorum*, "Time erases the fictions of opinion, but confirms the judgments of nature [119]."

A

Appendix

A.1. Definition of variables

Tables A.1 to A.4 list discriminative variables that are introduced in Section 5.6, and referred to multiple times in the main text. In complement to the figures presented in Section 5.6, Figures A.1 to A.4 compare simulated signal and background events in bins of these variables for the $B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$ mode. The histograms are obtained by selecting with the criteria listed in Sections 5.3 to 5.5 a total of approximately 10^6 simulated signal events and 10^6 simulated background events.

In the tables, the following notations are used:

- The average interaction point is noted IP.
- For a given track T and a given point P, the point of closest approach of T with respect to P, noted POCA, is defined as the point on the track T whose distance to the point P is minimal in the transverse plane [81].
- Some variables are computed in the centre-of-mass-system (CMS), but unless specified otherwise, the reference frame is the laboratory.

- The tracks in the rest of the event (ROE) are fit to a common vertex called the tag vertex.
- The notations related to the Belle II coordinate system are presented in Figure 3.3 in Section 3.2.

Table A.1.: Variables related to the entire event (see also Section 5.6.1). The definition of the signal-ROE modified Fox-Wolfram moments is given in Section 4.2.5, and, more generally, the definitions of variables depending on a momentum distribution are given in Section 4.2.

Variable	Unit	Definition
Modified Fox-Wolfram $H^{so}_{\xi,\ell}$		Signal-ROE modified Fox-Wolfram moment calculated in the CMS
$dr(K^+, \text{Tag Vertex})$	[μm]	Radial distance between the POCA of the K^+ candidate track and the tag vertex
$dr(p_{K^0_S}, \text{Tag Vertex})$	[μm]	Radial distance between the K^0_S candidate momentum line and the tag vertex
$dz(K^+, \text{Tag Vertex})$	[μm]	Longitudinal distance between the POCA of the K^+ candidate track and the tag vertex
$dz(p_{K^0_S}, \text{Tag Vertex})$	[μm]	Longitudinal distance between the K^0_S candidate momentum line and the tag vertex
$\cos(\text{thrust}_B, \text{thrust}_{\text{ROE}})$		Cosine of the angle between the thrust axis of the signal kaon candidate and the thrust axis of the ROE computed in the CMS
${M_{\text{missing}}}^2$	[GeV^2/c^4]	Square of the missing invariant mass
$\theta(p_{\text{missing}})$		Polar angle of the missing three-momentum in the CMS
Fox-Wolfram Moment R_ℓ		Normalised Fox-Wolfram moment in the CMS
Harmonic Moment B_ℓ		Harmonic moment with respect to the thrust axis in the CMS
N_γ		Number of photon candidates
N_{lepton}		Number of charged lepton candidates ($e^\pm$ or $\mu^\pm$)
N_{tracks}		Number of charged particle candidates
Sphericity		Event sphericity in the CMS
Thrust		Magnitude of the event thrust in the CMS
$\cos(\theta(\text{thrust}))$		Cosine of the polar angle of the thrust axis in the CMS
Total charge squared	[e^2]	Square of the total charge of tracks in the event

Table A.2.: Variables related to the properties of the signal kaon candidate (see also Section 5.6.2).

Variable	Unit	Definition
$M(K_S^0)$	[GeV/c^2]	Mass of the K_S^0 candidate
$\cos(p_{K_S^0}, \text{line(IP, } K_S^0 \text{ vertex))}$		Cosine of the angle between the K_S^0 candidate momentum line and the line from the IP to the K_S^0 candidate vertex
$dr(K^+)$	[μm]	Radial distance between the POCA of the K^+ candidate track and the IP
$dr(p_{K_S^0})$	[μm]	Radial distance between the K_S^0 candidate momentum line and the IP
$dz(K^+)$	[μm]	Longitudinal distance between the POCA of the K^+ candidate track and the IP
$dz(p_{K_S^0})$	[μm]	Longitudinal distance between the K_S^0 candidate momentum line and the IP
$\phi(K^+)$		Azimuthal angle of the K^+ candidate momentum at the POCA with respect to the IP
$p_T(K)$	[GeV/c]	Magnitude of the transverse momentum of the kaon candidate
$\cos(\text{thrust}_B, z)$		Cosine of the angle between the thrust axis of the signal kaon candidate and the z axis
q^2_{rec}	[GeV^2/c^4]	Reconstructed invariant mass squared of the two-neutrino system (Equation (5.2))
$\tilde{\varepsilon}_{\text{sig}}$		Signal selection efficiency quantile (Equation (5.8))

Table A.3.: Variables related to the ROE (see also Section 5.6.3). The definition of the ROE-ROE modified Fox-Wolfram moments is given in Section 4.2.5.

Variable	Unit	Definition
Modified Fox-Wolfram R_ℓ^{oo}		ROE-ROE (oo) modified Fox-Wolfram moment calculated in the CMS
ΔE_{ROE}	[GeV]	Difference between the ROE energy in the CMS and the energy of one beam in the CMS ($\sqrt{s}/2$)
M(ROE)	[GeV/c^2]	Invariant mass of the ROE
$\theta(p_{\mathrm{ROE}})$		Polar angle of the ROE momentum
p_{ROE}	[GeV/c]	Magnitude of the ROE momentum
thrust$_{\mathrm{ROE}}$		Magnitude of the ROE thrust in the CMS
p-value(Tag Vertex)		p-value of the tag vertex fit
dx(Tag Vertex)	[μm]	x component of the vector from the IP to the tag vertex
dy(Tag Vertex)	[μm]	y component of the vector from the IP to the tag vertex
dz(Tag Vertex)	[μm]	z component of the vector from the IP to the tag vertex
Variance$_{\mathrm{ROE}}(p_T)$	[GeV^2/c^2]	Variance of the transverse momentum of the ROE tracks

Table A.4.: Variables related to the D meson suppression. Section 5.6.4 gives information about how the D meson candidates are defined.

Variable	Unit	Definition
Median(p-value(D^0))		Median p-value of the vertex fits of the D^0 candidates
$M(D^0)$	[GeV/c^2]	Mass of the best D^0 candidate
p-value(D^0)		p-value of the best D^0 candidate vertex fit
$dr(D^0)$	[μm]	Radial distance between the best D^0 candidate vertex and the IP
$dz(D^0)$	[μm]	Longitudinal distance between the best D^0 candidate vertex and the IP
p-value(D^+)		p-value of the best D^+ candidate vertex fit
$dr(D^+)$	[μm]	Radial distance between the best D^+ candidate vertex and the IP
$dz(D^+)$	[μm]	Longitudinal distance between the best D^+ candidate vertex and the IP
$dr(p_{D^0})$	[μm]	Radial distance between the best D^0 candidate momentum line and the IP
$dz(p_{D^0})$	[μm]	Longitudinal distance between the best D^0 candidate momentum line and the IP
$M(D^+)$	[GeV/c^2]	Mass of the best D^+ candidate
p-value(D^+)		p-value of the best D^+ candidate vertex fit
$dz(p_{D^+})$	[μm]	Longitudinal distance between the best D^+ candidate momentum line and the IP

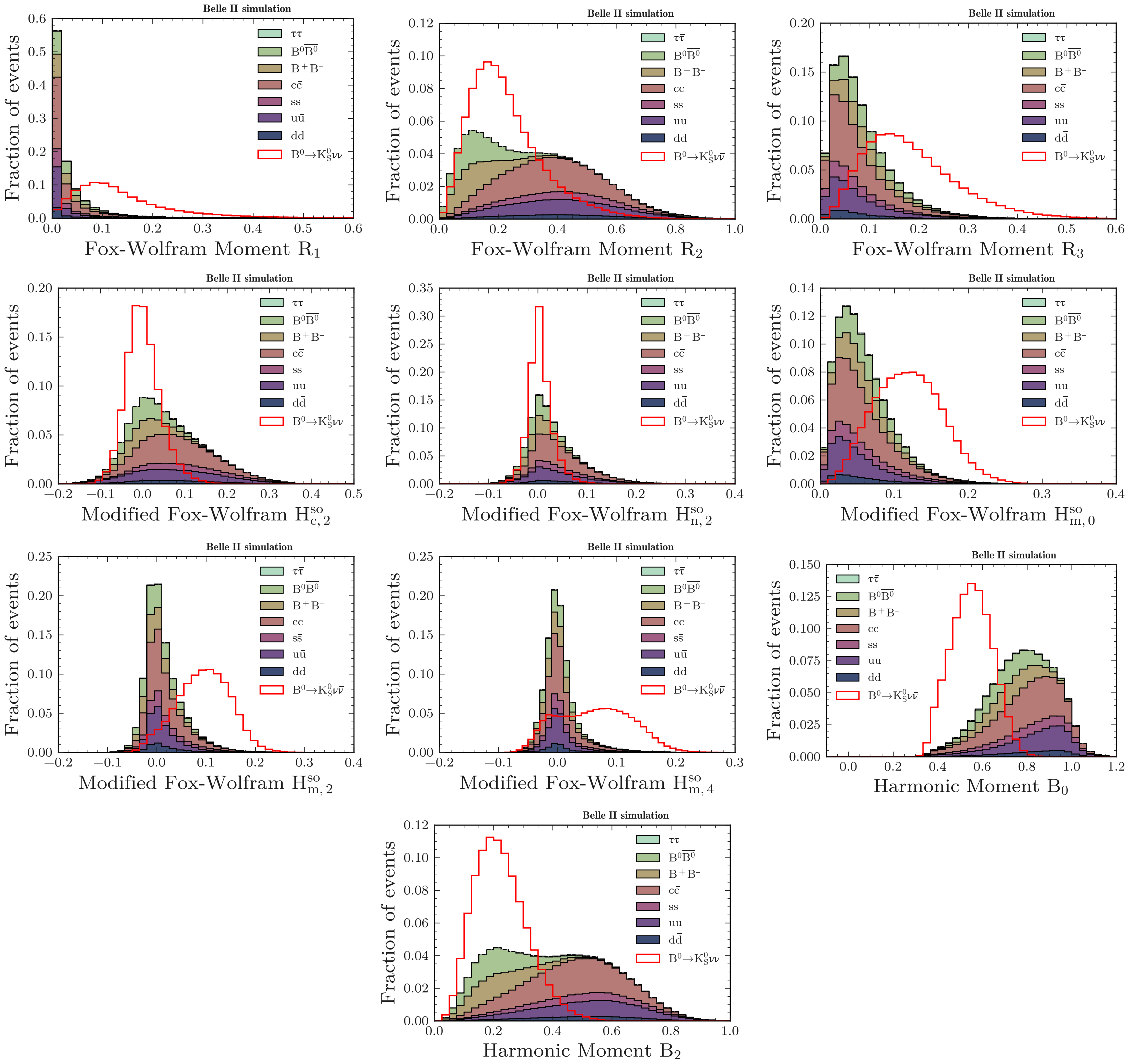


Figure A.1.: Simulated signal and background events in bins of variables related to the momentum distribution in the entire event for the $B^0 \to K^0_S\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain. The variable definitions are given in Table A.1.

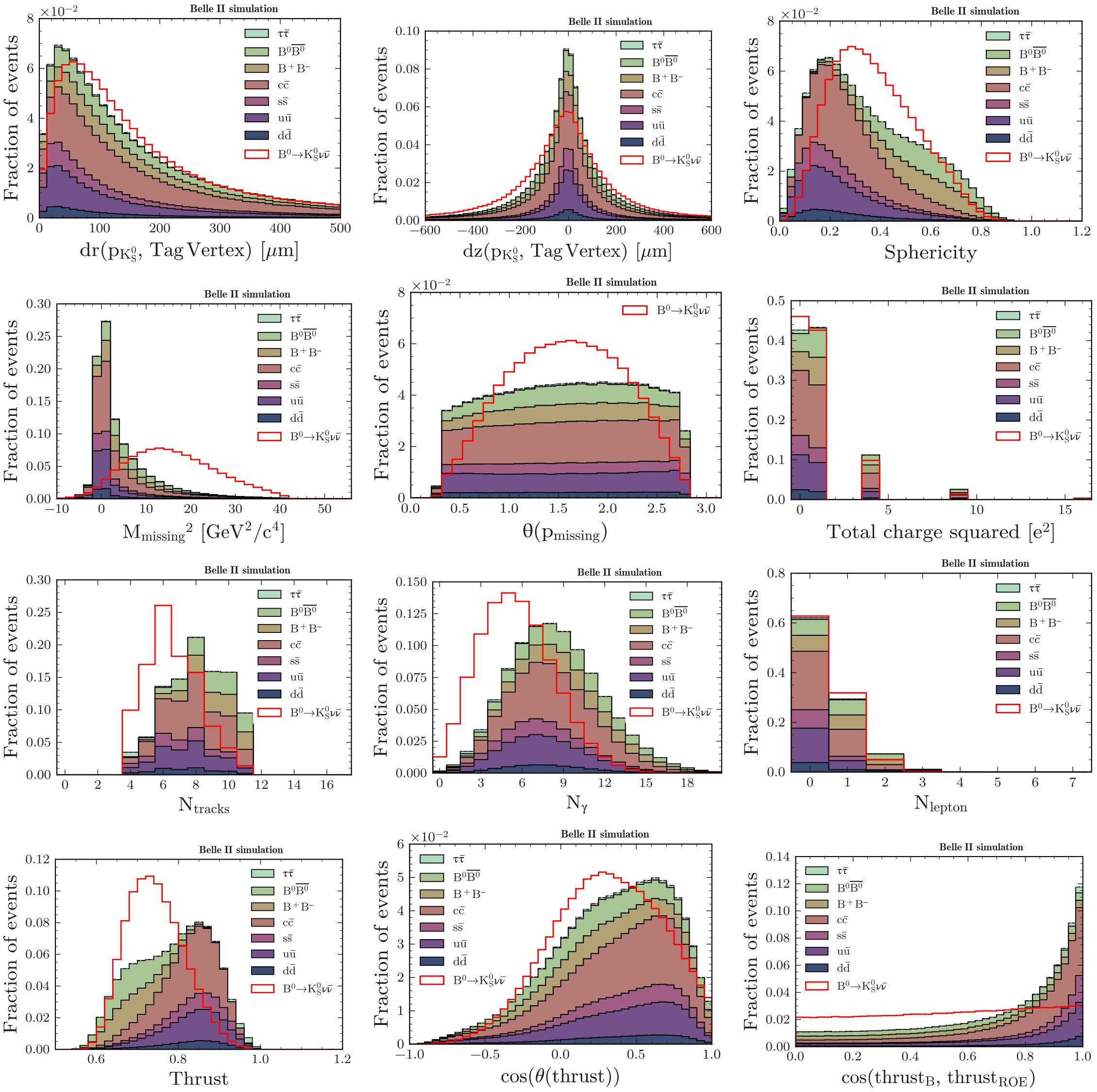


Figure A.2.: Simulated signal and background events in bins of variables related to the entire event for the $B^0 \rightarrow K^0_S \nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain. The variable definitions are given in Table A.1.

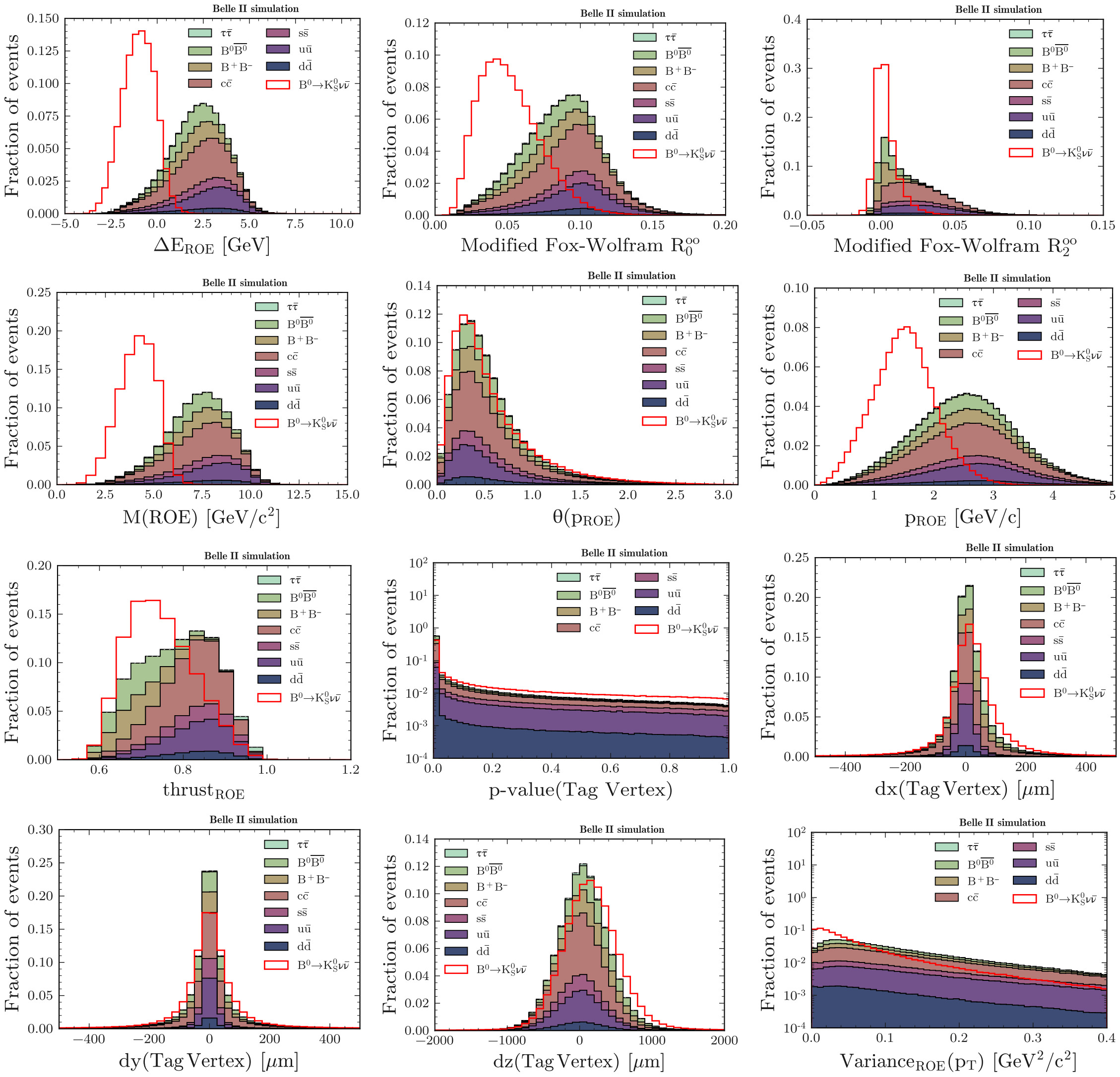


Figure A.3.: Simulated signal and background events in bins of variables related to the ROE for the $B^0 \to K^0_S \nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain. The variable definitions are given in Table A.3.

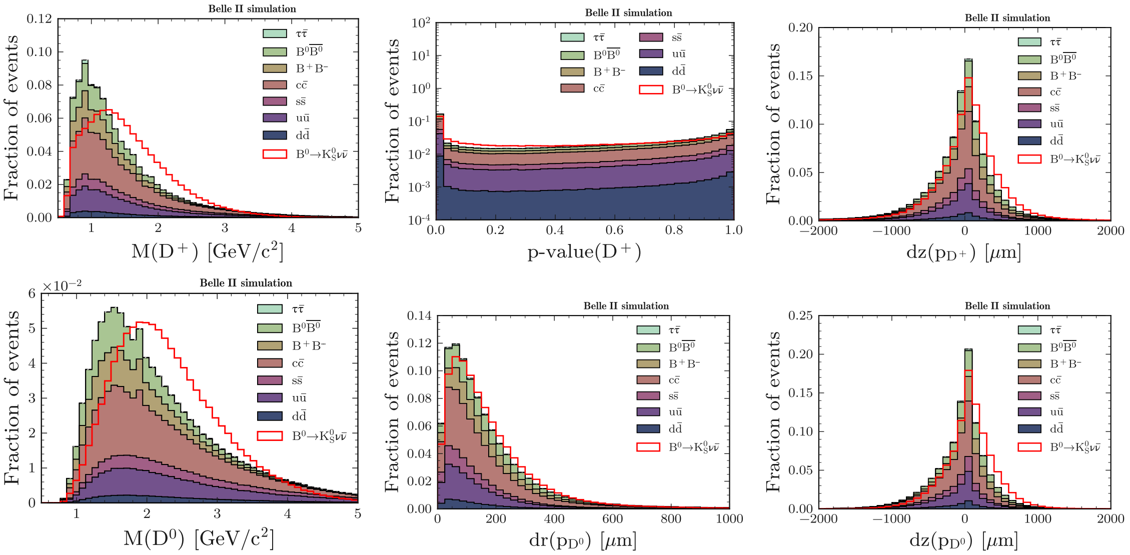


Figure A.4.: Simulated signal and background events in bins of variables related to the D meson suppression for the $B^0 \to K^0_S\nu\bar{\nu}$ mode. The histograms are obtained by selecting a total of approximately 10^6 simulated signal events and 10^6 simulated background events with the criteria listed in Sections 5.3 to 5.5. The signal histogram and the stacked background histogram are divided by the total number of events that they contain. The variable definitions are given in Table A.4.

A.2. Background coming from B meson decays

In complement to the discussion of Section 5.8.2, Tables A.5 to A.8 present the composition of the simulated background coming from decays of charged and neutral B mesons for the $B^+ \to K^+\nu\bar{\nu}$ mode and the $B^0 \to K^0_S\nu\bar{\nu}$ mode in their respective signal search region.

Table A.5.: Most important sources of background coming from charged B meson decays for the $B^+ \to K^+\nu\bar{\nu}$ mode and for two regions defined as a function of the signal selection efficiency quantile $\tilde{\epsilon}_{\text{sig}}$. The region $\tilde{\epsilon}_{\text{sig}} > 0.92$ corresponds to the twelve bins of the signal search region, and the region $\tilde{\epsilon}_{\text{sig}} > 0.98$ corresponds to the three final bins of the signal search region. The bins of the signal search region are defined in Section 5.8. Each entry in the table combines two charge-conjugate modes.

$\tilde{\varepsilon}_{\text{sig}}(B^+ \to K^+\nu\bar{\nu}) > 0.92$	Fraction [%]	$\tilde{\varepsilon}_{\text{sig}}(B^+ \to K^+\nu\bar{\nu}) > 0.98$	Fraction [%]
$B^\pm \to D^0\mu^\pm\nu_\mu$	13.4	$B^\pm \to D^0\mu^\pm\nu_\mu$	13.0
$B^\pm \to D^*(2007)^0\mu^\pm\nu_\mu$	10.0	$B^\pm \to D^0e^\pm\nu_e$	8.2
$B^\pm \to D^0e^\pm\nu_e$	10.0	$B^\pm \to K^\pm K^0 K^0$	6.7
$B^\pm \to D^*(2007)^0e^\pm\nu_e$	7.2	$B^\pm \to D^*(2007)^0\mu^\pm\nu_\mu$	6.3
$B^\pm \to D^0e^\pm\nu_e\gamma$	4.6	$B^\pm \to D^0K^\pm$	4.4
$B^\pm \to K^\pm K^0 K^0$	3.6	$B^\pm \to D^0e^\pm\nu_e\gamma$	4.1
$B^\pm \to D^*(2007)^0e^\pm\nu_e\gamma$	3.4	$B^\pm \to D^*(2007)^0e^\pm\nu_e$	3.9
$B^\pm \to D^0K^\pm$	2.9	$B^\pm \to D^*(2007)^0K^\pm$	3.7
$B^\pm \to D^*(2007)^0K^\pm$	2.6	$B^\pm \to D^*(2007)^0e^\pm\nu_e\gamma$	3.6
$B^\pm \to \eta_c(1S)K^\pm$	2.4	$B^\pm \to \eta_c(1S)K^\pm$	3.3
$B^\pm \to D^0K^\pm K^0$	1.7	$B^\pm \to D^0\pi^\pm$	3.2
$B^\pm \to D^0\pi^\pm$	1.6	$B^\pm \to \tau^\pm\nu_\tau$	2.0
$B^\pm \to \rho(770)^\pm D^0$	1.5	$B^\pm \to K^0K^\pm$	1.7
$B^\pm \to D^0\tau^\pm\nu_\tau$	1.5	$B^\pm \to \rho(770)^\pm D^0$	1.6
$B^\pm \to J/\psi(1S)K^\pm$	1.4	$B^\pm \to nnK^\pm$	1.4

Table A.6.: Most important sources of background coming from neutral B meson decays for the $B^+ \to K^+\nu\bar{\nu}$ mode and for two regions defined as a function of the signal selection efficiency quantile $\tilde{\epsilon}_{\text{sig}}$. The region $\tilde{\epsilon}_{\text{sig}} > 0.92$ corresponds to the twelve bins of the signal search region, and the region $\tilde{\epsilon}_{\text{sig}} > 0.98$ corresponds to the three final bins of the signal search region. The bins of the signal search region are defined in Section 5.8. Each entry in the table combines two charge-conjugate modes.

$\tilde{\varepsilon}_{\text{sig}}(B^+ \to K^+\nu\bar{\nu}) > 0.92$	Fraction [%]	$\tilde{\varepsilon}_{\text{sig}}(B^+ \to K^+\nu\bar{\nu}) > 0.98$	Fraction [%]
$B^0 \to D^*(2010)^\pm\mu^\pm\nu_\mu$	7.7	$B^0 \to D^\pm K^\pm$	12.5
$B^0 \to D^\pm K^\pm$	6.9	$B^0 \to D^\pm\pi^\pm$	6.8
$B^0 \to D^*(2010)^\pm e^\pm\nu_e$	6.0	$B^0 \to D^*(2010)^\pm K^\pm$	3.8
$B^0 \to D^\pm K^\pm K^0$	5.8	$B^0 \to D^*(2010)^\pm\mu^\pm\nu_\mu$	3.8
$B^0 \to D^\pm\mu^\pm\nu_\mu$	5.1	$B^0 \to D^\pm K^\pm K^0$	3.5
$B^0 \to D^\pm e^\pm\nu_e$	3.4	$B^0 \to D^*(2010)^\pm e^\pm\nu_e$	2.8
$B^0 \to D^*(2010)^\pm e^\pm\nu_e\gamma$	2.7	$B^0 \to D^\pm\mu^\pm\nu_\mu$	2.6
$B^0 \to D^*(2010)^\pm K^\pm$	2.6	$B^0 \to D^*(2010)^\pm e^\pm\nu_e\gamma$	2.4
$B^0 \to D^\pm\pi^\pm$	2.4	$B^0 \to \rho(770)^\pm D^\pm$	2.4
$B^0 \to D^*(2010)^\pm K^\pm K^0$	2.4	$B^0 \to K^\pm K^\pm K^0$	2.1
$B^0 \to D^\pm e^\pm\nu_e\gamma$	1.7	$B^0 \to K^*(892)^\pm D^\pm$	2.1
$B^0 \to \rho(770)^\pm D^\pm$	1.6	$B^0 \to D^\pm e^\pm\nu_e$	1.9
$B^0 \to K^\pm K^\pm K^0$	1.5	$B^0 \to D^*(2010)^\pm K^\pm K^0$	1.6
$B^0 \to K^*(892)^\pm D^\pm$	1.4	$B^0 \to D^\pm\pi^\pm\gamma$	1.6
$B^0 \to D^*(2010)^\pm\mu^\pm\nu_\mu\gamma$	1.3	$B^0 \to D^*(2010)^\pm\pi^\pm$	1.6

Table A.7.: Most important sources of background coming from charged B meson decays for the $B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}$ mode and for two regions defined as a function of the signal selection efficiency quantile $\tilde{\epsilon}_{\mathrm{sig}}$. The region $\tilde{\epsilon}_{\mathrm{sig}} > 0.92$ corresponds to the twelve bins of the signal search region, and the region $\tilde{\epsilon}_{\mathrm{sig}} > 0.98$ corresponds to the three final bins of the signal search region. The bins of the signal search region are defined in Section 5.8. Each entry in the table combines two charge-conjugate modes.

$\tilde{\varepsilon}_{\mathrm{sig}}(B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}) > 0.92$	Fraction [%]	$\tilde{\varepsilon}_{\mathrm{sig}}(B^0 \to K^0_{\mathrm{S}}\nu\bar{\nu}) > 0.98$	Fraction [%]
$B^\pm \to D^0\mu^\pm\nu_\mu$	17.6	$B^\pm \to D^0\mu^\pm\nu_\mu$	20.5
$B^\pm \to D^0e^\pm\nu_e$	12.9	$B^\pm \to D^0e^\pm\nu_e$	11.5
$B^\pm \to D^*(2007)^0\mu^\pm\nu_\mu$	10.4	$B^\pm \to D^*(2007)^0\mu^\pm\nu_\mu$	10.1
$B^\pm \to D^*(2007)^0e^\pm\nu_e$	7.7	$B^\pm \to K^\pm K^0 K^0$	6.7
$B^\pm \to D^0e^\pm\nu_e\gamma$	5.5	$B^\pm \to D^*(2007)^0e^\pm\nu_e$	5.5
$B^\pm \to D^*(2007)^0e^\pm\nu_e\gamma$	3.6	$B^\pm \to D^0e^\pm\nu_e\gamma$	4.1
$B^\pm \to K^\pm K^0 K^0$	2.5	$B^\pm \to K^0_S K^0_L \pi^\pm$	2.8
$B^\pm \to K^*(892)^\pm D^0$	1.8	$B^\pm \to K^*(892)^\pm D^0$	2.5
$B^\pm \to D^0\mu^\pm\nu_\mu\gamma$	1.7	$B^\pm \to D^0e^\pm\nu_e\gamma\gamma$	2.1
$B^\pm \to D^0e^\pm\nu_e\gamma\gamma$	1.4	$B^\pm \to D^*(2007)^0e^\pm\nu_e\gamma$	2.1
$B^\pm \to D^0D_s^\pm$	1.4	$B^\pm \to D^0\mu^\pm\nu_\mu\gamma$	1.8
$B^\pm \to D^0\tau^\pm\nu_\tau$	1.3	$B^\pm \to D^*(2007)^0K^*(892)^\pm$	1.6
$B^\pm \to D^*(2007)^0K^*(892)^\pm$	1.2	$B^\pm \to K^*(892)^\pm K^0$	1.6
$B^\pm \to D^*(2007)^0\mu^\pm\nu_\mu\gamma$	1.2	$B^\pm \to \rho(770)^\pm D^0$	1.4
$B^\pm \to D_0^*(2300)^0e^\pm\nu_e$	1.1	$B^\pm \to K_0^*(1430)^\pm K^0$	1.4

Table A.8.: Most important sources of background coming from neutral B meson decays for the $B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$ mode and for two regions defined as a function of the signal selection efficiency quantile $\tilde{\epsilon}_\mathrm{sig}$. The region $\tilde{\epsilon}_\mathrm{sig} > 0.92$ corresponds to the twelve bins of the signal search region, and the region $\tilde{\epsilon}_\mathrm{sig} > 0.98$ corresponds to the three final bins of the signal search region. The bins of the signal search region are defined in Section 5.8. Each entry in the table combines two charge-conjugate modes.

$\tilde{\epsilon}_\mathrm{sig}(B^0 \to K^0_\mathrm{S}\nu\bar{\nu}) > 0.92$	Fraction [%]	$\tilde{\epsilon}_\mathrm{sig}(B^0 \to K^0_\mathrm{S}\nu\bar{\nu}) > 0.98$	Fraction [%]
$B^0 \to D^\pm\mu^\pm\nu_\mu$	16.7	$B^0 \to D^\pm\mu^\pm\nu_\mu$	16.1
$B^0 \to D^\pm e^\pm\nu_e$	12.2	$B^0 \to K^0K^0K^0$	12.5
$B^0 \to D^*(2010)^\pm\mu^\pm\nu_\mu$	6.9	$B^0 \to D^\pm e^\pm\nu_e$	9.4
$B^0 \to D^\pm e^\pm\nu_e\gamma$	6.1	$B^0 \to D^\pm e^\pm\nu_e\gamma$	5.1
$B^0 \to K^0K^0K^0$	5.7	$B^0 \to D^*(2010)^\pm\mu^\pm\nu_\mu$	4.1
$B^0 \to D^*(2010)^\pm e^\pm\nu_e$	4.9	$B^0 \to D^*(2010)^\pm e^\pm\nu_e$	3.8
$B^0 \to D^*(2010)^\pm e^\pm\nu_e\gamma$	2.3	$B^0 \to K^0_SK^0_LK^0_L$	3.2
$B^0 \to D^\pm\mu^\pm\nu_\mu\gamma$	2.3	$B^0 \to K^0_SK^0_SK^0_L$	2.5
$B^0 \to D^\pm e^\pm\nu_e\gamma\gamma$	1.9	$B^0 \to K^0K^0K^*(892)^0$	1.9
$B^0 \to D^\pm\tau^\pm\nu_\tau$	1.6	$B^0 \to K^*(892)^\pm D^\pm$	1.9
$B^0 \to K^*(892)^\pm D^\pm$	1.5	$B^0 \to D^\pm\mu^\pm\nu_\mu\gamma$	1.5
$B^0 \to K^0_SK^0_LK^0_L$	1.3	$B^0 \to D^\pm e^\pm\nu_e\gamma\gamma$	1.4
$B^0 \to D^\pm K^\pm K^0$	1.2	$B^0 \to D^*(2010)^\pm e^\pm\nu_e\gamma$	1.3
$B^0 \to D^\pm D^\pm_s$	1.2	$B^0 \to K^0K^0$	1.2
$B^0 \to K^0_SK^0_SK^0_L$	1.1	$B^0 \to D^0K^0$	1.2

A.3. Correction weights for simulated events

Table A.9 summarises the weights that are applied to simulated events to correct for multiple sources of mis-modelling:

- The selection efficiency of the PID requirement imposed on signal K^+ candidates (Section 5.4) differs between data and simulation. The Belle II performance group provides weights, noted w_{PID}, that correct for this efficiency difference. These weights are functions of the transverse momentum p_T and the polar angle θ of the signal K^+ candidate.
- By default, simulation does not take into account the q^2-dependence of the form factor entering in the computation of the $B \to K\nu\bar{\nu}$ branching fraction (Section 2.3.2). A correction weight, noted w_{ff}, is applied to the simulated signal events to produce a realistic q^2-dependence of the signal branching fraction. This weight is computed as the ratio between the two distributions shown in Figure 2.5 (Section 2.3.2).
- In Section 5.9.3, a correction weight is introduced to correct for the mis-modelling of the continuum background simulation. This weight is noted w_c and is function of the BDT_c output.

The total correction weight applied to a simulated event is the product of the individual correction weights.

Table A.9.: The product of a set of correction weights is applied to each simulated event. The weights in the table are defined for the selection of $B^+ \to K^+\nu\bar{\nu}$ decays. For the neutral mode ($B^0 \to K^0_S\nu\bar{\nu}$), the PID weight w_{PID} is absent, because there is no PID requirement (Section 5.4), and the other weights are the same as for the charged mode.

Simulated sample	Event correction weight
$e^+e^- \to B^+(\to K^+\nu\bar{\nu})B^-$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_{ff}(q^2)$
$e^+e^- \to B^+B^-$	$w_{\mathrm{PID}}(p_T, \cos\theta)$
$e^+e^- \to B^0\overline{B}^0$	$w_{\mathrm{PID}}(p_T, \cos\theta)$
$e^+e^- \to u\overline{u}$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_c(\mathrm{BDT_c})$
$e^+e^- \to d\overline{d}$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_c(\mathrm{BDT_c})$
$e^+e^- \to s\overline{s}$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_c(\mathrm{BDT_c})$
$e^+e^- \to c\overline{c}$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_c(\mathrm{BDT_c})$
$e^+e^- \to \tau^+\tau^-$	$w_{\mathrm{PID}}(p_T, \cos\theta) \cdot w_c(\mathrm{BDT_c})$

A.4. Yield stability in the signal search region

The Belle II dataset is segmented into several experiments, each corresponding to a certain data collection period. Moreover, each experiment is subdivided into runs, each corresponding to an uninterrupted period of data taking. In this way, each event at Belle II is unambiguously defined by the triplet (experiment number, run number, event number). The amount of on-resonance data studied here corresponds to a total of $189\,\mathrm{fb}^{-1}$ of integrated luminosity, with experiment numbers ranging from 7 to 18.

Figure A.5 shows the number of on-resonance data events in the signal search region per unit of integrated luminosity as a function of the data taking period for the charged mode ($B^+ \to K^+\nu\bar{\nu}$) and the neutral mode ($B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$). In order to follow the closed-box principle mentioned at the beginning of Section 5.9, the average number of events is subtracted to hide the actual number of selected data events in the signal search region. The data yield per unit of integrated luminosity is reasonably stable, with fluctuations that may be due to the fact that the simulated events are run-independent, meaning that they are calibrated globally for the entire data sample.

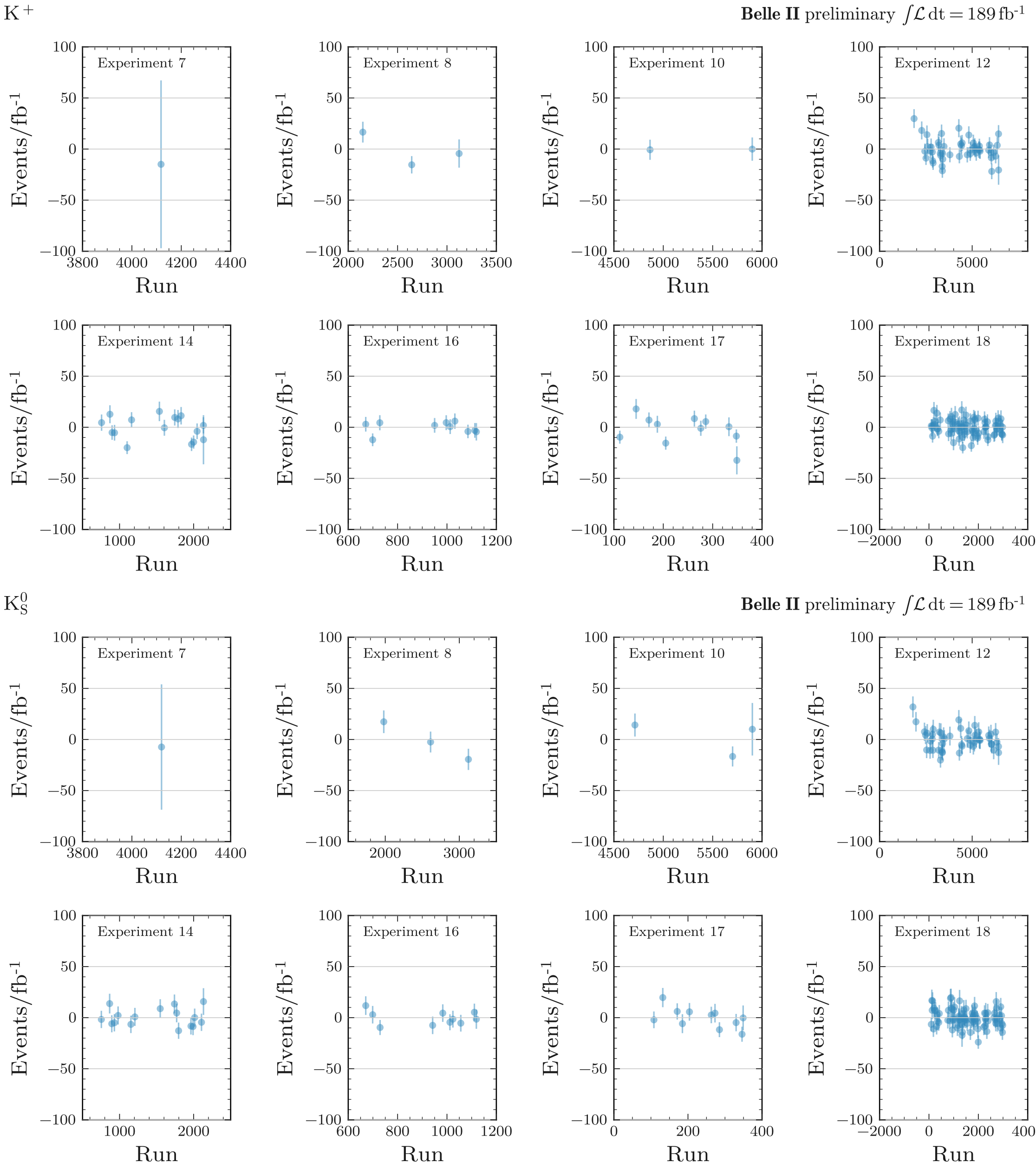


Figure A.5.: Number of on-resonance data events in the signal search region in bunches of $1\,\mathrm{fb}^{-1}$ as a function of the experiment number and the run number. The average number of events is subtracted to hide the actual number of selected data events. The eight top plots correspond to the $B^+ \to K^+\nu\bar{\nu}$ mode, and the eight bottom plots correspond to the $B^0 \to K^0_S\nu\bar{\nu}$ mode.

A.5. Comparison between the Poisson-based and the Gaussian-based models

In complement to Figure 5.36 in Section 5.11, Figure A.6 compares the signal strength μ obtained by fitting the `pyhf` model and the `sghf` model to pseudo-observations generated with several input signal strengths $\mu_{\text{sig}} = 1, 5, 20$. The results are the same for the two models in most cases.

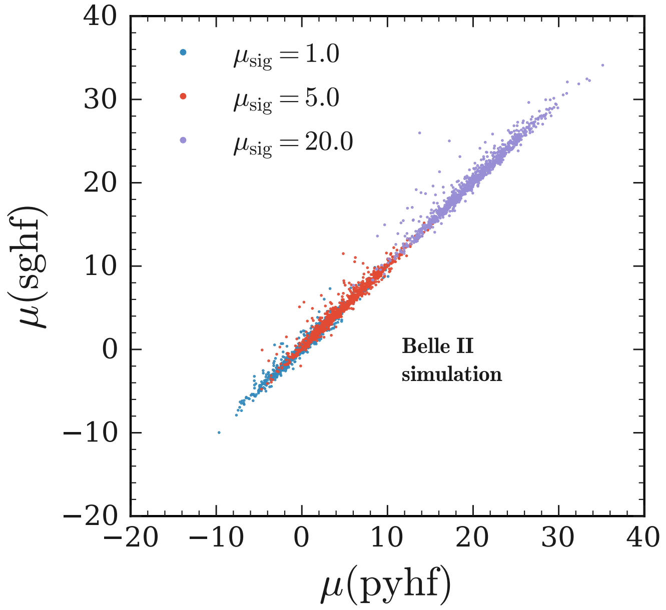


Figure A.6.: Signal strength determined by fitting the `sghf` model and the `pyhf` model to pseudo-observations. Pseudo-observations are generated with both Poisson statistical and Gaussian systematic fluctuations around the expectations. An injected signal strength is varied between $\mu_{\text{sig}} = 1$ (standard model expectation), $\mu_{\text{sig}} = 5$, and $\mu_{\text{sig}} = 20$.

A.6. Post-fit normalisation parameters

This appendix gives information about the normalisation nuisance parameters obtained after a fit of the binned-likelihood model to data samples of $63\,\mathrm{fb}^{-1}$ collected at the $\Upsilon(4S)$ resonance and $9\,\mathrm{fb}^{-1}$ collected $60\,\mathrm{MeV}$ below the resonance (Section 5.12).

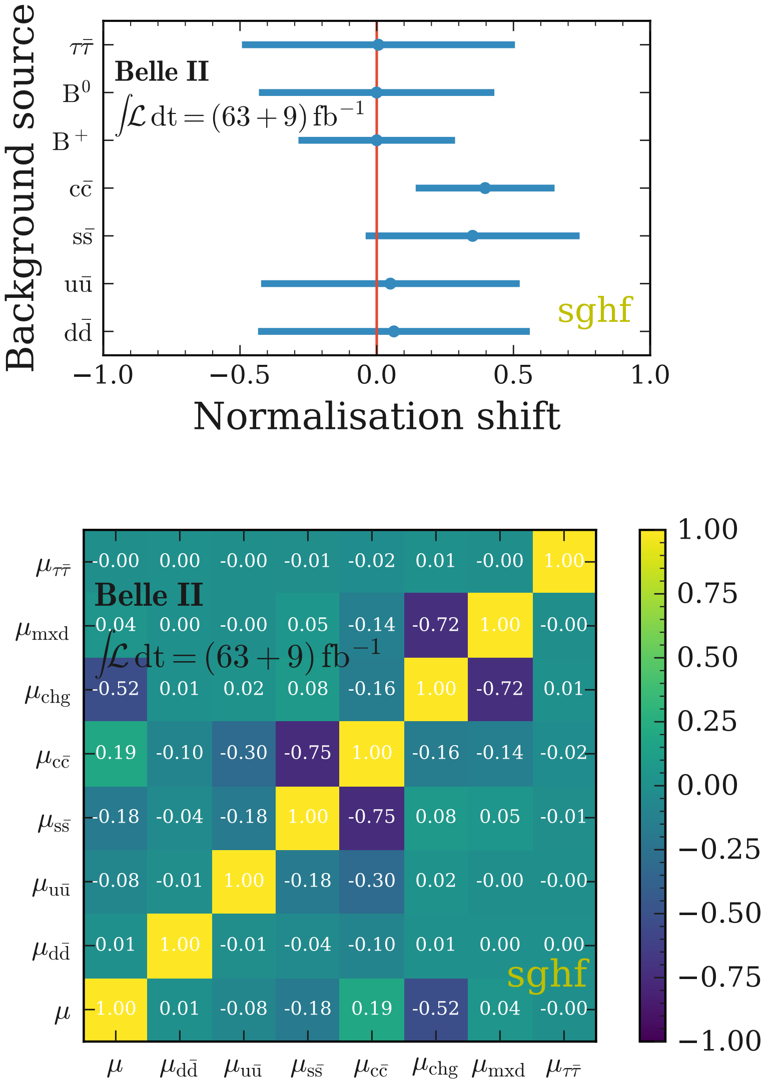
If s is a background sample taken from the set of considered background sources,

$$s \in \left\{e^+e^- \to B^+B^-, e^+e^- \to B^0\overline{B}^0\right\} \cup \left\{e^+e^- \to q\overline{q} \,:\, q = u, d, c, s\right\} \cup \left\{e^+e^- \to \tau^+\tau^-\right\},$$

then the background normalisation shift of s is defined as $\mu_\mathrm{s}-1$, where μ_s is the background strength (i.e. the background normalisation nuisance parameter) of the background sample s (Section 5.10.1). For example, a normalisation shift of zero corresponds to no variation with respect to the expectation, and a shift of $+0.5$ corresponds to an upscale of the expected yield by a factor of 1.5.

Figure A.7 (top) presents the post-fit normalisation shifts of the background sources. The normalisation shifts of the B meson background are close to zero (no variation). By contrast, the normalisation shifts of the $e^+e^- \to c\overline{c}$ background and the $e^+e^- \to s\overline{s}$ background are close to 0.4, meaning that the corresponding post-fit yields are upscaled by a factor of 1.4 with respect to the pre-fit expectations. This upscaling of the continuum background was anticipated by the normalisation discrepancy observed when comparing simulated continuum background and off-resonance data in the bins of the signal search region at the end of Section 5.9.3.

Figure A.7 (bottom) shows the correlation between the signal strength μ, defined at the beginning of Section 4.4, and the strengths μ_s of the seven background sources. The signal strength is anti-correlated to the strength of the charged B meson background, because this background is similar to signal in the most sensitive bins of the signal search region (Figure 5.41 in Section 5.12). In addition, the strength of the charged B meson background is strongly anti-correlated to the strength of the neutral B meson background, meaning that these two background sources are hard to distinguish from each other. Similarly, the strengths of the two main sources of continuum background, $e^+e^- \to c\overline{c}$ and $e^+e^- \to s\overline{s}$, are strongly anti-correlated.



	μ	$\mu_{d\bar{d}}$	$\mu_{u\bar{u}}$	$\mu_{s\bar{s}}$	$\mu_{c\bar{c}}$	μ_{chg}	μ_{mxd}	$\mu_{\tau\bar{\tau}}$
$\mu_{\tau\bar{\tau}}$	-0.00	0.00	-0.00	-0.01	-0.02	0.01	-0.00	1.00
μ_{mxd}	0.04	0.00	-0.00	0.05	-0.14	-0.72	1.00	-0.00
μ_{chg}	-0.52	0.01	0.02	0.08	-0.16	1.00	-0.72	0.01
$\mu_{c\bar{c}}$	0.19	-0.10	-0.30	-0.75	1.00	-0.16	-0.14	-0.02
$\mu_{s\bar{s}}$	-0.18	-0.04	-0.18	1.00	-0.75	0.08	0.05	-0.01
$\mu_{u\bar{u}}$	-0.08	-0.01	1.00	-0.18	-0.30	0.02	-0.00	-0.00
$\mu_{d\bar{d}}$	0.01	1.00	-0.01	-0.04	-0.10	0.01	0.00	0.00
μ	1.00	0.01	-0.08	-0.18	0.19	-0.52	0.04	-0.00

Figure A.7.: At the top, post-fit normalisation shifts and uncertainties of the background sources as obtained by the `sghf` fit to the data. The normalisation shift is defined as $\mu_{\mathrm{bkg}} - 1$, where μ_{bkg} is the background strength (i.e. the background normalisation nuisance parameter). At the bottom, correlation between the signal strength μ and the strengths of the seven background sources. The strength of the neutral B meson background is noted μ_{mxd}, and the strength of the charged B meson background is noted μ_{chg}.

Glossary

ARICH Aerogel ring-imaging Cherenkov detector. 33, 36, 37

AUC Area under the receiver-operating-characteristic curve. 45, 80, 81, 82, 83, 84, 85, 88, 89

basf2 Belle II analysis software framework. 40

CDC Central drift chamber. 33, 35, 36, 37, 39, 60, 61, 63, 65

CKM Cabibbo-Kobayashi-Maskawa (matrix). 13, 14, 16, 18

CMS Centre-of-mass system. 68, 69, 70, 71, 129, 131, 133

CP Charge-parity (symmetry). 13, 16, 24

ECL Electromagnetic calorimeter. 38, 39, 40, 60, 61, 114, 115

EW Electroweak (theory). 11

FCNC Flavour-changing neutral current. 13, 20

GIM Glashow–Iliopoulos–Maiani (mechanism). 13, 20

HLT High-level trigger. 39

IP Interaction point. 60, 70, 71, 72, 129, 132, 133, 134

KLM K_{L}^{0} and μ detection system. 33, 38, 39

L1 Online trigger. 39

LHC Large hadron collider. 17

NP New physics. 9, 11, 23, 24

pdf Portable document format. 10

PID Particle identification. 33, 60, 63, 65, 70, 85, 86, 96, 106, 107, 108, 109, 110, 111, 116, 144, 145

POCA Point of closest approach. 69, 70, 129, 131, 132

PXD Pixel detector. 33, 34, 63

pyhf Pure-python HistFactory. 53, 54, 119, 120, 121, 122, 124, 148

QCD Quantum chromodynamics. 11

ROC Receiver-operating-characteristic curve. 45, 80, 81, 83, 84, 86

ROE Rest of the event. 48, 68, 69, 70, 71, 72, 76, 98, 130, 131, 133, 137

sghf Simplified Gaussian model. 53, 118, 119, 120, 148, 150

SM Standard model. 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 22, 23, 24, 25, 118, 123, 124

SVD Silicon vertex detector. 33, 34, 35

TOP Time-of-propagation detector. 33, 36, 37

VXD Vertex detector. 33, 60

Bibliography

[1] S.L. Glashow, *Partial symmetries of weak interactions*, *Nucl. Phys.* **22** (1961) 579. (cited on pages 9, 11, and 19)

[2] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (1967) 1264. (cited on pages 9, 11, and 19)

[3] A. Salam, *Weak and electromagnetic interactions*, *Conf. Proc. C* **680519** (1968) 367. (cited on pages 9 and 11)

[4] G. 't Hooft and M.J.G. Veltman, *Regularization and renormalization of gauge fields*, *Nucl. Phys. B* **44** (1972) 189. (cited on page 9)

[5] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [`1207.7214`]. (cited on page 9)

[6] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [`1207.7235`]. (cited on page 9)

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (1964) 321. (cited on page 9)

[8] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (1964) 508. (cited on page 9)

[9] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (1964) 585. (cited on page 9)

[10] F. Zwicky, *On the masses of nebulae and of clusters of nebulae*, *Astrophys. J.* **86** (1937) 217. (cited on page 9)

[11] D. Clowe, M. Bradac, A.H. Gonzalez, M. Markevitch, S.W. Randall, C. Jones *et al.*, *A direct empirical proof of the existence of dark matter*, *Astrophys. J.* **648** (2006) L109 [`astro-ph/0608407`]. (cited on page 9)

[12] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [`1807.06209`]. (cited on page 9)

[13] Muon g-2 collaboration, *Measurement of the positive muon anomalous magnetic moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (2021) 141801 [`2104.03281`]. (cited on page 9)

[14] Muon g-2 collaboration, *Final report of the muon E821 anomalous magnetic moment measurement at BNL*, *Phys. Rev. D* **73** (2006) 072003 [`hep-ex/0602035`]. (cited on page 9)

[15] LHCb collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, *JHEP* **08** (2017) 055 [`1705.05802`]. (cited on pages 9 and 23)

[16] LHCb collaboration, *Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay*, *Phys. Rev. Lett.* **125** (2020) 011802 [`2003.04831`]. (cited on pages 9 and 23)

[17] LHCb collaboration, *Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay*, *Phys. Rev. Lett.* **126** (2021) 161802 [`2012.13241`]. (cited on pages 9 and 23)

[18] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, *Nature Phys.* **18** (2022) 277 [`2103.11769`]. (cited on pages 9 and 23)

[19] Belle II collaboration, *Belle II technical design report*, `1011.0352`. (cited on pages 10, 32, 33, and 35)

[20] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*, *Nucl. Instrum. Meth. A* **907** (2018) 188. (cited on pages 10, 29, and 32)

[21] T. Berners-Lee, *Information management: a proposal*, CERN (1989). (cited on page 10)

[22] I.J.R. Aitchison and A.J.G. Hey, *Gauge theories in particle physics: a practical introduction, volume 2: non-Abelian gauge theories : QCD and the electroweak*

theory, fourth edition, Taylor & Francis (2013), 10.1201/9781466513105. (cited on pages 12 and 13)

[23] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak interaction*, *Prog. Theor. Phys.* **49** (1973) 652. (cited on page 13)

[24] S.L. Glashow, J. Iliopoulos and L. Maiani, *Weak interactions with lepton-hadron symmetry*, *Phys. Rev. D* **2** (1970) 1285. (cited on page 13)

[25] Particle Data Group, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01. (cited on pages 14, 15, 16, 20, 21, 26, 30, 36, 61, 64, 72, 85, 95, 96, 110, and 120)

[26] BaBar collaboration, *The BaBar detector*, *Nucl. Instrum. Meth. A* **479** (2002) 1 [`hep-ex/0105044`]. (cited on page 16)

[27] BaBar collaboration, *The BaBaR detector: upgrades, operation and performance*, *Nucl. Instrum. Meth. A* **729** (2013) 615 [`1305.3560`]. (cited on page 16)

[28] Belle collaboration, *The Belle detector*, *Nucl. Instrum. Meth. A* **479** (2002) 117. (cited on page 16)

[29] Belle collaboration, *Physics achievements from the Belle experiment*, *PTEP* **2012** (2012) 04D001 [`1212.5342`]. (cited on page 16)

[30] BaBar and Belle collaboration, *The physics of the B factories*, *Eur. Phys. J. C* **74** (2014) 3026 [`1406.6311`]. (cited on pages 16, 46, 47, and 48)

[31] Belle collaboration, *Observation of large CP violation in the neutral B meson system*, *Phys. Rev. Lett.* **87** (2001) 091802 [`hep-ex/0107061`]. (cited on page 16)

[32] BaBar collaboration, *Observation of CP violation in the B^0 meson system*, *Phys. Rev. Lett.* **87** (2001) 091801 [`hep-ex/0107013`]. (cited on page 16)

[33] BaBar collaboration, *Measurement of $|V_{cb}|$ and the form-factor slope in $\overline{B} \to D\ell^-\overline{\nu}_\ell$ decays in events tagged by a fully reconstructed B Meson*, *Phys. Rev. Lett.* **104** (2010) 011802 [`0904.4063`]. (cited on page 16)

[34] Belle collaboration, *Measurement of the decay $B \to D\ell\nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$*, *Phys. Rev. D* **93** (2016) 032006 [`1510.03657`]. (cited on page 16)

[35] BaBar collaboration, *Measurement of the inclusive electron spectrum from B meson decays and determination of $|V_{ub}|$*, *Phys. Rev. D* **95** (2017) 072001 [`1611.05624`]. (cited on page 16)

[36] Belle collaboration, *Measurements of partial branching fractions of inclusive $B \to X_u \ell^+ \nu_\ell$ decays with hadronic tagging*, *Phys. Rev. D* **104** (2021) 012008 [`2102.00020`]. (cited on page 16)

[37] Belle collaboration, *Measurement of the CKM matrix element $|V_{cb}|$ from $B^0 \to D^{*-}\ell^+\nu_\ell$ at Belle*, *Phys. Rev. D* **100** (2019) 052007 [`1809.03290`]. (cited on page 16)

[38] BaBar collaboration, *Direct CP, lepton flavor and isospin asymmetries in the decays $B \to K^{(*)}\ell^+\ell^-$*, *Phys. Rev. Lett.* **102** (2009) 091803 [`0807.4119`]. (cited on page 16)

[39] Belle collaboration, *Measurement of the differential branching fraction and forward-backward asymmetry for $B \to K^{(*)}\ell^+\ell^-$*, *Phys. Rev. Lett.* **103** (2009) 171801 [`0904.0770`]. (cited on page 16)

[40] BaBar collaboration, *Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment*, *Phys. Rev. Lett.* **118** (2017) 031802 [`1605.09637`]. (cited on page 16)

[41] Belle collaboration, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, *JHEP* **03** (2021) 105 [`1908.01848`]. (cited on pages 16 and 24)

[42] Belle collaboration, *Measurement of the mass of the τ-lepton and an upper limit on the mass difference between τ^+ and τ^-*, *Phys. Rev. Lett.* **99** (2007) 011801 [`hep-ex/0608046`]. (cited on page 16)

[43] BaBar collaboration, *Measurements of the τ mass and the mass difference of the τ^+ and τ^- at BaBaR*, *Phys. Rev. D* **80** (2009) 092005 [`0909.3562`]. (cited on page 16)

[44] Belle II collaboration, *Precise measurement of the D^0 and D^+ lifetimes at Belle II*, *Phys. Rev. Lett.* **127** (2021) 211801 [`2108.03216`]. (cited on pages 16, 34, and 163)

[45] LHCb collaboration, *The LHCb detector at the LHC*, *JINST* **3** (2008) S08005. (cited on page 17)

[46] LHCb collaboration, *LHCb detector performance*, *Int. J. Mod. Phys. A* **30** (2015) 1530022 [`1412.6352`]. (cited on page 17)

[47] L. Evans and P. Bryant, eds., *LHC machine*, *JINST* **3** (2008) S08001. (cited on page 17)

[48] K.G. Wilson, *Non-Lagrangian models of current algebra*, *Phys. Rev.* **179** (1969) 1499. (cited on page 18)

[49] K.G. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, *Commun. Math. Phys.* **24** (1972) 87. (cited on page 18)

[50] A.J. Buras and M. Lindner, eds., *Heavy flavours II*, WSP, Singapore (1998), 10.1142/2706. (cited on page 18)

[51] A.J. Buras, *Weak Hamiltonian, CP violation and rare decays*, in *Les Houches summer school in theoretical physics*, 1998 [`hep-ph/9806471`]. (cited on pages 18 and 20)

[52] A.J. Buras, *Flavor physics and CP violation*, in *European school of high-energy physics*, 2004 [`hep-ph/0505175`]. (cited on page 18)

[53] E. Fermi, *An attempt of a theory of beta radiation*, *Z. Phys.* **88** (1934) 161. (cited on page 18)

[54] R.P. Feynman and M. Gell-Mann, *Theory of Fermi interaction*, *Phys. Rev.* **109** (1958) 193. (cited on page 18)

[55] A.J. Buras, J. Girrbach-Noe, C. Niehoff and D.M. Straub, *$B \to K^{(*)}\nu\bar{\nu}$ decays in the Standard Model and beyond*, *JHEP* **02** (2015) 184 [`1409.4557`]. (cited on pages 19, 20, 21, 23, and 91)

[56] T. Inami and C.S. Lim, *Effects of superheavy quarks and leptons in low-energy weak processes $K_L \to \mu\bar{\mu}$, $K^+ \to \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \overline{K}^0$*, *Prog. Theor. Phys.* **65** (1981) 297. (cited on page 19)

[57] G. Buchalla and A.J. Buras, *The rare decays $K^+ \to \pi^+\nu\bar{\nu}$ and $K_L \to \mu^+\mu^-$ beyond leading logarithms*, *Nucl. Phys. B* **412** (1994) 106 [`hep-ph/9308272`]. (cited on pages 19 and 20)

[58] G. Buchalla and A.J. Buras, *The rare decays $K \to \pi\nu\bar{\nu}$, $B \to X\nu\bar{\nu}$ and $B \to l^+l^-$: an update*, *Nucl. Phys. B* **548** (1999) 309 [`hep-ph/9901288`]. (cited on page 19)

[59] J. Brod, M. Gorbahn and E. Stamou, *Two-Loop electroweak corrections for the $K \to \pi\nu\bar{\nu}$ decays*, *Phys. Rev. D* **83** (2011) 034030 [`1009.0947`]. (cited on page 19)

[60] T. Blake, G. Lanfranchi and D.M. Straub, *Rare B decays as tests of the Standard Model*, *Prog. Part. Nucl. Phys.* **92** (2017) 50. (cited on pages 21 and 124)

[61] Belle II collaboration, *The Belle II physics book*, *PTEP* **2019** (2019) 123C01 [`1808.10567`]. (cited on pages 22, 31, and 38)

[62] Belle collaboration, *Test of lepton-flavor universality in $B \to K^{*}\ell^{+}\ell^{-}$ decays at Belle*, *Phys. Rev. Lett.* **126** (2021) 161801 [`1904.02440`]. (cited on page 24)

[63] LHCb collaboration, *Tests of lepton universality using $B^{0} \to K_{S}^{0}\ell^{+}\ell^{-}$ and $B^{+} \to K^{*+}\ell^{+}\ell^{-}$ decays*, *Phys. Rev. Lett.* **128** (2022) 191802 [`2110.09501`]. (cited on page 24)

[64] W. Buchmuller, R. Ruckl and D. Wyler, *Leptoquarks in lepton-quark collisions*, *Phys. Lett. B* **191** (1987) 442. (cited on page 24)

[65] D. Bečirević, I. Doršner, S. Fajfer, N. Košnik, D.A. Faroughy and O. Sumensari, *Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies*, *Phys. Rev. D* **98** (2018) 055003 [`1806.05689`]. (cited on page 24)

[66] A. Angelescu, D. Bečirević, D.A. Faroughy, F. Jaffredo and O. Sumensari, *Single leptoquark solutions to the B-physics anomalies*, *Phys. Rev. D* **104** (2021) 055017 [`2103.12504`]. (cited on page 24)

[67] R.D. Peccei and H.R. Quinn, *CP conservation in the presence of instantons*, *Phys. Rev. Lett.* **38** (1977) 1440. (cited on page 24)

[68] R.D. Peccei and H.R. Quinn, *Constraints imposed by CP conservation in the presence of instantons*, *Phys. Rev. D* **16** (1977) 1791. (cited on page 24)

[69] S. Weinberg, *A new light boson?*, *Phys. Rev. Lett.* **40** (1978) 223. (cited on page 24)

[70] F. Wilczek, *Problem of strong P and T invariance in the presence of instantons*, *Phys. Rev. Lett.* **40** (1978) 279. (cited on page 24)

[71] J. Martin Camalich, M. Pospelov, P.N.H. Vuong, R. Ziegler and J. Zupan, *Quark flavor phenomenology of the QCD axion*, *Phys. Rev. D* **102** (2020) 015023 [`2002.04623`]. (cited on page 24)

[72] A. Filimonova, R. Schäfer and S. Westhoff, *Probing dark sectors with long-lived particles at Belle II*, *Phys. Rev. D* **101** (2020) 095006 [`1911.03490`]. (cited on page 25)

[73] BaBar collaboration, *Search for the rare decay $B \to K\nu\bar{\nu}$*, *Phys. Rev. D* **82** (2010) 112002 [`1009.1529`]. (cited on pages 26 and 27)

[74] BaBar collaboration, *Search for $B \to K^{(*)}\nu\bar{\nu}$ and invisible quarkonium decays*, *Phys. Rev. D* **87** (2013) 112005 [`1303.7465`]. (cited on pages 26, 27, 124, 125, and 128)

[75] Belle collaboration, *Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample*, *Phys. Rev. D* **87** (2013) 111103 [`1303.3719`]. (cited on pages 26, 27, 124, 125, and 128)

[76] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander *et al.*, *A hierarchical NeuroBayes-based algorithm for full reconstruction of B Mesons at B Factories*, *Nucl. Instrum. Meth. A* **654** (2011) 432 [`1102.3876`]. (cited on page 26)

[77] Belle collaboration, *Search for $B \to h\nu\bar{\nu}$ decays with semileptonic tagging at Belle*, *Phys. Rev. D* **96** (2017) 091101 [`1702.03224`]. (cited on pages 26, 27, 98, 124, 125, 126, and 128)

[78] Belle collaboration, *Measurement of the branching fraction of $B^+ \to \tau^+\nu_\tau$ decays with the semileptonic tagging method*, *Phys. Rev. D* **92** (2015) 051102 [`1503.05613`]. (cited on page 26)

[79] N.L. Braun, *Combinatorial Kalman filter and high level trigger reconstruction for the Belle II experiment*, Ph.D. thesis, KIT, Karlsruhe, 2018. 10.1007/978-3-030-24997-7. (cited on pages 30, 34, and 36)

[80] SuperB collaboration, *SuperB: a high-luminosity asymmetric e^+e^- super flavor factory. Conceptual design report*, `0709.0451`. (cited on page 31)

[81] Belle II tracking group, *Track finding at Belle II*, *Comput. Phys. Commun.* **259** (2021) 107610. (cited on pages 33, 36, 60, 69, 129, and 163)

[82] J. Kemmer, E. Belau, U. Prechtel, W. Welser and G. Lutz, *Low capacity drift diode*, *Nucl. Instrum. Meth. A* **253** (1987) 378. (cited on page 33)

[83] T.V. Dong, S. Uno, M. Uchida, H. Ozaki, N. Taniguchi, E. Nakano *et al.*, *Calibration and alignment of the Belle II central drift chamber*, *Nucl. Instrum. Meth. A* **930** (2019) 132. (cited on page 36)

[84] K. Inami, *Development of a TOP counter for the super B factory*, *Nucl. Instrum. Meth. A* **595** (2008) 96. (cited on page 37)

[85] K. Miyabayashi, *Belle II electromagnetic calorimeter and its performance during early SuperKEKB operation*, *JINST* **15** (2020) C10016. (cited on page 38)

[86] Belle II framework software group, *The Belle II core software*, *Comput. Software Big Sci.* **3** (2019) 1. (cited on page 40)

[87] D.J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152. (cited on page 40)

[88] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten *et al.*, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159. (cited on page 40)

[89] S. Jadach, B.F.L. Ward and Z. Was, *The precision Monte Carlo event generator KK for two fermion final states in* e^+e^- *collisions*, *Comput. Phys. Commun.* **130** (2000) 260 [`hep-ph/9912214`]. (cited on page 40)

[90] S. Jadach, B.F.L. Ward and Z. Was, *Coherent exclusive exponentiation for precision Monte Carlo calculations*, *Phys. Rev. D* **63** (2001) 113009 [`hep-ph/0006359`]. (cited on page 40)

[91] S. Jadach, B.F.L. Ward and Z. Was, *TAUOLA: a library of Monte Carlo programs to simulate decays of polarized* τ *leptons*, *Comput. Phys. Commun.* **64** (1991) 275. (cited on page 40)

[92] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was and Z. Was, *Universal interface of TAUOLA technical and physics documentation*, *Comput. Phys. Commun.* **183** (2012) 821 [`1002.0543`]. (cited on page 40)

[93] GEANT4 collaboration, *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250. (cited on page 40)

[94] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning*, Springer, New York (2001). (cited on pages 42 and 126)

[95] T. Keck, *FastBDT: a speed-optimized multivariate classification algorithm for the Belle II experiment*, *Comput. Softw. Big Sci.* **1** (2017) 2. (cited on pages 43, 44, and 78)

[96] J.H. Friedman, *Stochastic gradient boosting*, *Comput. Stat. Data Anal.* **38** (2002) 367. (cited on page 43)

[97] T. Chen and C. Guestrin, *XGBoost: a scalable tree boosting system*, in *Proceedings of the 22nd international conference on knowledge discovery and data mining*, (New York), ACM, 2016 [`1603.02754`]. (cited on pages 43, 44, and 81)

[98] J.H. Friedman, T. Hastie and R. Tibshirani, *Additive logistic regression: a statistical view of boosting*, *Ann. Statist.* **28** (2000) 337. (cited on page 44)

[99] Y. Freund and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, *Jour. Comp. Syst. Sci.* **55** (1997) 119. (cited on page 44)

[100] W. Peterson, T. Birdsall and W. Fox, *The theory of signal detectability*, *Trans. IRE Prof. Group Inf. Theory* **4** (1954) 171. (cited on page 45)

[101] G.C. Fox and S. Wolfram, *Observables for the analysis of event shapes in* e^+e^- *annihilation and other processes*, *Phys. Rev. Lett.* **41** (1978) 1581. (cited on pages 47 and 48)

[102] G.C. Fox and S. Wolfram, *Event shapes in* e^+e^- *annihilation*, *Nucl. Phys. B* **149** (1979) 413. (cited on page 47)

[103] Belle collaboration, *Evidence for* $B^0 \to \pi^0\pi^0$, *Phys. Rev. Lett.* **91** (2003) 261801. (cited on page 48)

[104] ROOT collaboration, *HistFactory: a tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016 (2012). (cited on page 51)

[105] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, *Journal of Open Source Software* **6** (2021) 2823. (cited on pages 51, 53, and 54)

[106] A.L. Read, *Presentation of search results: the* CL_s *technique*, *J. Phys. G* **28** (2002) 2693. (cited on pages 53 and 54)

[107] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554. (cited on pages 53 and 54)

[108] E. Parzen, *On estimation of a probability density function and mode*, *Ann. Math. Stat.* **33** (1962) 1065. (cited on page 54)

[109] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, *Ann. Math. Stat.* **27** (1956) 832. (cited on page 54)

[110] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau *et al.*, *SciPy 1.0: fundamental algorithms for scientific computing in Python*, *Nature Methods* **17** (2020) 261. (cited on page 55)

[111] Belle collaboration, *Search for* $B^+ \to \mu^+\nu_\mu$ *and* $B^+ \to \mu^+ N$ *with inclusive tagging*, *Phys. Rev. D* **101** (2020) 032007 [`1911.03186`]. (cited on pages 58 and 125)

[112] L3 collaboration, *Analysis of the* $\pi^+\pi^-\pi^+\pi^-$ *and* $\pi^+\pi^0\pi^-\pi^0$ *final states in quasi-real two-photon collisions at LEP*, *Phys. Lett. B* **638** (2006) 128 [`hep-ex/0605021`]. (cited on page 65)

[113] Belle collaboration, *Study of charmonia in four-meson final states produced in two-photon collisions*, Eur. Phys. J. C **53** (2008) 1 [`0706.3955`]. (cited on page 65)

[114] Belle collaboration, *High-statistics study of K_S^0 pair production in two-photon collisions*, PTEP **2013** (2013) 123C01 [`1307.7457`]. (cited on page 65)

[115] BaBar collaboration, *A measurement of the total width, the electronic width, and the mass of the $\Upsilon(10580)$ resonance*, Phys. Rev. D **72** (2005) 032005 [`hep-ex/0405025`]. (cited on page 85)

[116] Belle collaboration, *Update of inclusive cross sections of single and pairs of identified light charged hadrons*, Phys. Rev. D **101** (2020) 092004 [`2001.10194`]. (cited on page 98)

[117] D. Martschei, M. Feindt, S. Honc and J. Wagner-Kuhr, *Advanced event reweighting using multivariate analysis*, J. Phys. Conf. Ser. **368** (2012) 012028. (cited on page 99)

[118] Belle II collaboration, *Search for $B^+ \to K^+\nu\overline{\nu}$ decays using an inclusive tagging method at Belle II*, Phys. Rev. Lett. **127** (2021) 181802 [`2104.12624`]. (cited on pages 121, 123, 124, 125, 126, 128, and 163)

[119] Cicero, *De natura deorum* 2,5. (cited on page 128)

[120] Belle II collaboration, *Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment*, Chin. Phys. C **44** (2020) 021001 [`1910.05365`]. (cited on page 163)

[121] Belle II collaboration, *Search for an invisibly decaying Z' boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ plus missing energy final states*, Phys. Rev. Lett. **124** (2020) 141801 [`1912.11276`]. (cited on page 163)

[122] Belle II collaboration, *Search for axion-like particles produced in e^+e^- collisions at Belle II*, Phys. Rev. Lett. **125** (2020) 161806 [`2007.13071`]. (cited on page 163)

[123] Belle and Belle II collaboration, *Combined analysis of Belle and Belle II data to determine the CKM angle ϕ_3 using $B^+ \to D(K_S^0 h^- h^+)h^+$ decays*, JHEP **02** (2022) 063 [`2110.12125`]. (cited on page 163)

Disclaimer

The method used to search for $B \to K\nu\bar{\nu}$ decays presented in this thesis was co-designed by my supervisor Dr. Alexander Glazov (AG), my colleagues Filippo Dattola (FD), Dr. Simon Kurz (SK), Dr. Slavomira Stefkova (S^2), and myself.

I was involved in every step of the analysis, and I carried out the entirety of the search for $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0_\mathrm{S}\nu\bar{\nu}$ decays presented in this thesis with the dataset of $189\,\mathrm{fb}^{-1}$. All the code developed for this analysis was either written or reviewed by me, and I am the author of the shown figures in this thesis, unless specified otherwise. In particular:

- The code employed for the validation channel (Section 5.9.1) was co-written by SK, AG, FD, and myself.
- The code to produce Figures 2.5, 5.42 and A.5 (right) was written by AG.
- The code of the simplified Gaussian model (Figures 5.37 and A.7) was implemented by AG, and validated by AG, S^2, and myself.
- The code to classify the background sources (Appendix A.2) and to plot the binned expected signal selection efficiency (Figure 5.18) was written by FD.
- The code to compute the systematic uncertainties (Section 5.10) was co-written by AG, FD, and myself.
- The code to apply and validate the maximum-likelihood fit, and to produce Figures 5.41 and 5.43, was co-written by S^2, SK, FD, and myself.

The result obtained with a first iteration of the method on a restricted dataset of $63\,\mathrm{fb}^{-1}$ (Section 5.12) is shared between this thesis and the thesis of FD, and was published in the Physical Review Letters in October 2021 [118]. The published article was co-written by FD, AG, SK, S^2, and myself.

My contribution to other publications that I co-authored with the Belle II tracking group [81] or the Belle II collaboration [120, 121, 122, 44, 123] during my doctoral studies are not covered by this thesis.

Acknowledgments

I would like to express my gratitude to the SuperKEKB group and the Belle II collaboration without which this project would not have been possible; to my thesis supervisor Sasha Glazov for his continuous support, his enthusiasm, and the hours that he did not count to share his knowledge and insights; to my working group colleagues Filippo Dattola, Eldar Ganiev, Simon Kurz, and Slavomira Stefkova, for the time spent together in Europe and in Japan, the fruitful discussions, and their helpful comments on the draft of my thesis; to Carsten Niebuhr, for his support and for offering me the opportunity to work in a team that I will miss; to my co-supervisor Kerstin Tackmann, who found the time to follow my thesis; to Adrien Florio, Michael de Nuccio, Navid Rad, Clara Flintrop, and Vlad Popescu for their helpful comments on the draft of my thesis; to my other colleagues at DESY, in particular Henrikas Svidras, Abtin Narimani, Sascha Dreyer, Anselm Baur, Tommy Martinov, Lorenzo Quintavalle, and Philipp Tontsch, for making me feel less alone on the PhD path; to my family and my parents, for their continuous support; to Dolma, for most of the reasons above, and too many others.

Declaration

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Cyrille Praz
Hamburg, the 29th of August 2022