Searches for the rare $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^{*0} \nu\bar{\nu}$ decays using an inclusive tagging 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
Filippo Dattola
aus Reggio Calabria (Italien)

Hamburg
2022

Gutachter/innen der Dissertation:	Dr. Alexander Glazov Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Dr. Alexander Glazov Prof. Dr. Elisabetta Gallo Prof. Dr. Bernd Andreas Kniehl Dr. Georg Steinbrück Prof. Dr. Kerstin Tackmann
Vorsitzender der Prüfungskommission:	Prof. Dr. Bernd Andreas Kniehl
Datum der Disputation:	26.09.2022
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. Heinrich Graener

Abstract

This thesis presents two searches for the rare $B \to K^{(*)}\nu\bar{\nu}$ decays at the Belle II experiment, operating at the SuperKEKB asymmetric-energy electron-positron collider, at the KEK laboratories in Tsukuba, Japan.
The $B \to K^{(*)}\nu\bar{\nu}$ decays are based on the $b \to s\nu\bar{\nu}$ flavour-changing neutral-current transition. This process is forbidden at the Born level in the Standard Model (SM) and can only occur at higher orders in SM perturbation theory, at the cost of large suppression. Nevertheless, the $B \to K^{(*)}\nu\bar{\nu}$ decays are predicted with high theoretical accuracy. Therefore, precise measurements of their branching fractions offer a unique opportunity to test the SM and its extensions.
Searches for the $B \to K^{(*)}\nu\bar{\nu}$ decays are affected by major experimental limitations related to the presence of two neutrinos in the final state and are only possible at B factories. Here, millions of B meson decays, produced at the $\Upsilon(4\mathrm{S})$ resonance, can be recorded in kinematically-constrained and low-background conditions. The searches presented in this thesis are performed with a novel measurement approach based on the implementation of an *inclusive tagging* method. In the inclusive tagging, both the characteristic features of the signal decay and the inclusive properties of the accompanying B meson in the $\Upsilon(4\mathrm{S}) \to B\bar{B}$ event are exploited to suppress backgrounds. Such a technique provides higher signal efficiency and improved sensitivity compared to the other existing tagging methods, in which the second B meson is explicitly reconstructed in hadronic or semileptonic decays.
The first data analysis presented in this thesis is a search for the $B^+ \to K^+\nu\bar{\nu}$ decay with the inclusive tagging performed in 2021. This search uses a data sample corresponding to an integrated luminosity of $63\,\mathrm{fb}^{-1}$ collected by the Belle II detector at the $\Upsilon(4\mathrm{S})$ resonance and an off-resonance sample of $9\,\mathrm{fb}^{-1}$ collected at an energy $60\,\mathrm{MeV}$ below the resonance. No statistically significant signal is observed and an upper limit of 4.1×10^{-5} on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction is set at a 90% confidence level. This measurement is competitive with the results of previous searches obtained using significantly larger data samples.
In the second analysis of the thesis, which is not yet finalised, the inclusive tagging is used to search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay. The analysis is optimised for larger data samples corresponding to on-resonance and off-resonance integrated luminosities of $189\,\mathrm{fb}^{-1}$ and $18\,\mathrm{fb}^{-1}$, respectively. Using simulated events and collision data reconstructed in dedicated control channels, the measurement sets an expected upper limit of 4.4×10^{-5} on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction at a 90% confidence level. The expected limit is in the ballpark of the previous results, despite the smaller integrated luminosity.

Zusammenfassung

In dieser Arbeit werden zwei Suchen nach den seltenen $B \to K^{(*)}\nu\bar{\nu}$-Zerfällen am Belle II-Experiment vorgestellt, das am asymmetrischen Elektron-Positron-Beschleuniger SuperKEKB in den KEK-Laboratorien in Tsukuba, Japan, betrieben wird.
Die $B \to K^{(*)}\nu\bar{\nu}$-Zerfälle basieren auf dem $b \to s\nu\bar{\nu}$ Übergang von Flavour verändernden neutralen Strömen. Dieser Prozess ist auf der Born-Ebene des Standardmodells (SM) verboten und kann in der SM-Störungstheorie nur bei höheren Ordnungen auftreten, was mit einer starken Unterdrückung verbunden ist. Dennoch werden die $B \to K^{(*)}\nu\bar{\nu}$-Zerfälle mit hoher theoretischer Genauigkeit vorhergesagt. Daher bieten präzise Messungen ihrer Verzweigungsverhältnisse eine einzigartige Gelegenheit, das SM und seine Erweiterungen zu testen.
Die Suche nach den $B \to K^{(*)}\nu\bar{\nu}$-Zerfällen unterliegt erheblichen experimentellen Einschränkungen, die mit der Anwesenheit von zwei Neutrinos in den Endzuständen zusammenhängen, und ist nur an B-Fabriken möglich. Hier können Millionen von B-Meson-Zerfällen, die an der $\Upsilon(4\mathrm{S})$-Resonanz erzeugt werden, unter kinematisch eingeschränkten und hintergrundarmen Bedingungen aufgezeichnet werden. Die in dieser Arbeit vorgestellten Suchvorgänge werden mit einem neuartigen Messansatz durchgeführt, der auf der Implementierung einer *inklusiven Tagging*-Methode basiert. Beim inklusiven Tagging werden sowohl die charakteristischen Merkmale des Zerfalls der Signalseite als auch die inklusiven Eigenschaften des begleitenden B-Mesons im $\Upsilon(4\mathrm{S}) \to B\bar{B}$-Ereignis ausgenutzt, um Hintergründe zu unterdrücken. Diese Technik bietet eine höhere Signaleffizienz und eine verbesserte Empfindlichkeit im Vergleich zu den anderen existierenden Tagging-Methoden, bei denen das zweite B-Meson in hadronischen oder semileptonischen Zerfällen explizit rekonstruiert wird.
Die erste in dieser Arbeit vorgestellte Datenanalyse ist eine Suche nach dem $B^+ \to K^+\nu\bar{\nu}$-Zerfall mit dem im Jahr 2021 durchgeführten inklusiven Tagging. Bei dieser Suche wird eine Datenprobe verwendet, die einer integrierten Luminosität von $63\,\mathrm{fb}^{-1}$ entspricht, die vom Belle II-Detektor bei der $\Upsilon(4\mathrm{S})$-Resonanz gesammelt wurde, und eine Off-Resonanz-Probe von $9\,\mathrm{fb}^{-1}$, die bei einer Energie von $60\,\mathrm{MeV}$ unterhalb der Resonanz gesammelt wurde. Es wird kein statistisch signifikantes Signal beobachtet und ein oberes Limit von 4.1×10^{-5} für das $B^+ \to K^+\nu\bar{\nu}$ Verzweigungsverhältnis bei einem Konfidenzintervall von 90% festgelegt. Diese Messung ist konkurrenzfähig mit den Ergebnissen früherer Suchen, die mit wesentlich größeren Datenproben durchgeführt wurden.
In der zweiten Analyse dieser Arbeit, die noch nicht abgeschlossen ist, wird das inklusiven Tagging für die Suche nach dem $B^0 \to K^{*0}\nu\bar{\nu}$-Zerfall benutzt. Die Analyse wird für größere Datenproben optimiert, die integrierten On-Resonanz- und Off-Resonanz-Luminositäten von $189\,\mathrm{fb}^{-1}$ bzw. $18\,\mathrm{fb}^{-1}$ entsprechen. Unter Verwendung von simulierten Ereignissen und Kollisionsdaten, die in speziellen Kontrollkanälen rekonstruiert wurden, setzt die Messung ein zu erwartendes oberes Limit von 4.4×10^{-5} für das $B^0 \to K^{*0}\nu\bar{\nu}$ Verzweigungsverhältnis bei einem Konfidenzintervall von 90%. Das zu erwartende Limit liegt trotz der geringeren integrierten Luminosität in der Größenordnung der bisherigen Ergebnisse.

Disclaimer

The data analyses presented in this thesis are the result of a collaborative effort.
The novel *inclusive tagging* method used in the searches for the $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^{*0} \nu\bar{\nu}$ decays has been designed and implemented by a team of researchers including myself, my supervisor Dr. Alexander Glazov (AG), and my colleagues: Cyrille Praz (CP), Dr. Simon Kurz (SK), Dr. Slavomira Stefkova (SS).
I was directly involved in all the stages of the two analyses: setup, validation, statistical interpretation of the results. I participated in the development and review of the analysis code. Moreover, I contributed to the documentation of the studies. I am co-author, alongside AG, SK, CP, SS, of the publication (Phys. Rev. Lett. 127, 181802 [1]) based on the search for the $B^+ \to K^+ \nu\bar{\nu}$ decay presented in Part II of this thesis.
Unless otherwise specified, I am the author of the figures shown in the manuscript. Figures not produced by myself are always accompanied by a reference to the authors. In particular, figures of the analyses produced by one among AG, SK, CP, SS, are identified by a reference to the analysis note [2] or the published paper [1].
A summary of the specific contributions of my supervisor and colleagues to the studies of the thesis is given in the following:

- The style and colour scheme of the majority of the plots are designed by CP and SK.
- The `pyhf` library [3], introduced in Section 3.2.2 was validated by SS.
- The author of the `sghf` framework, introduced in Section 3.2.3 is AG. The code was also validated by CP and SS.
- The signal reweighting procedure described in Sections 4.2.1 and 12.2.1 were implemented and validated by AG and myself.
- The code for the reconstruction procedure described in Chapter 5 is the result of a collaborative work involving me, AG, SK, CP, SS.
- The main developer of the classifiers described in Chapters 6, 8 (Section 8.2), 14,15 (Section 15.2) is CP. Myself, AG, SK, SS contributed to the validation.
- The implementation of the statistical model described in Chapter 9, including the estimation of the systematic uncertainties, is the result of a collaborative work involving me, AG, SK, CP, SS.
- The pseudo-experiments (toys) used in Section 9.3 for the fit validation were produced by AG.
- The *signal embedding* technique presented in Section 15.1 is based on an idea of AG.
- The code used to evaluate the systematic uncertainties in Section 16.5 is the result of a collaborative work involving me, AG, CP.

In the years of this PhD project, I have been co-author of the Belle II publications in Refs. [4–9]. These studies are not presented in the thesis.

Alle mie nonne.

"My father was a tough guy [a plumber]. He said, 'What do you want to be, a ballet dancer?' I said, 'No, I want to be a physicist.' He said, 'You ain't going to work in no drugstore.' I said, 'No, not a pharmacist! A physicist!' He said, 'What's a physicist?' I told him, 'Like what Einstein does.' And my father lit up and said, 'Einstein?' 'Yeah, I want to do the kind of thing that Einstein did.' And he looked at me and he said, 'Are you any good at this?'
My mother was crying and saying, 'We're going to be broke!' "

(L. Susskind)

Contents

Part I

Introduction

Chapter 1

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

This chapter introduces the reader to the theoretical framework of the studies presented in this thesis. Section 1.1 defines flavour and provides a brief historical overview of flavour physics. Section 1.2 describes the theory of flavour in the Standard Model. The formalism used to study the weak decays of heavy mesons is introduced in Section 1.3, and Section 1.4 provides a description of the $B \to K^{(*)}\nu\bar{\nu}$ decays within such framework. In Section 1.5 new physics scenarios are presented, and their possible impact on $b \to s\nu\bar{\nu}$ transitions is discussed. Section 1.6 summarises the current status of the experimental searches.

1.1 Flavour physics

In 1971, at a Baskin-Robbins ice-cream store in Pasadena, California, Murray Gell-Mann and his student Harald Fritzsch coined the term *flavour* to describe the different types of quarks known at the time [10]. Since then, the word flavour has been used in particle physics to indicate different copies of fermionic fields having the same spin and gauge quantum numbers.
In the Standard Model (SM) there are six flavours of leptons and six flavours of quarks, which are parametrised with flavour quantum numbers. *Flavour physics* is the study of the processes in which flavour quantum numbers change through the agency of the weak force.

1.1.1 Historical overview

Over the decades, flavour physics has given essential contributions to the building and the development of the SM. The first steps towards a consistent theory of flavour were made between the 1950s and 1960s. In 1958, when only three quarks, u, d, s, were theorised, Feynman and Gell-Mann proposed the universality of the weak interactions after correctly predicting the close equality of the effective Fermi coupling constants in the muon decay $\mu^- \to e^-\bar{\nu}_e\nu_\mu$ and in the neutron β-decay $n \to pe^-\bar{\nu}_e$ [11]. However, their theory was not able to explain why the effective Fermi constant in the decays of strange particles, such as the $\Lambda^0 \to pe^-\bar{\nu}_e$ decay, was observed to be smaller by a factor of 4-5. In 1963, Cabibbo preserved the universality of the weak couplings by introducing the angle θ_C, known as Cabibbo angle, which contributes as $\cos\theta_C$ in $d \to u$ transitions and as $\sin\theta_C$ in $s \to u$ transitions. A value $\theta_C \simeq 13^\circ$ correctly described the experimental data with the same Fermi constant [12]. Still, Cabibbo's theory could not explain the suppression of strangeness-changing neutral-current processes, such as the $K^0 \to \mu^+\mu^-$ decay, which the theory predicted to occur at a measurable rate, but was not observed at the time. In 1970, the Glashow-Iliopoulos-Maiani (GIM) mechanism elegantly solved the puzzle with the introduction of the fourth quark, the charm quark c [13]. Its existence had already been theorised by Bjorken and Glashow in 1964, who predicted its mass to be $\sim 1.5\,\mathrm{GeV}/c^2$ [14]. In 1974, the experimental discovery of the J/ψ meson, a bound state of a charm and an anti-charm quark, confirmed the predictions of the GIM mechanism [15, 16]. Moreover, the GIM mechanism and the discovery of the charm quark led to the interpretation of the Cabibbo angle θ_C as a rotation angle

in the flavour space between the weak eigenstates d', s' and the mass eigenstates d, s:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{1.1}$$

However, the theory was still lacking a consistent explanation for the violation of the CP symmetry observed in 1964 by Cronin and Fitch in the decays of neutral kaons [17]. An explanation of such a phenomenon only came in 1973. The Japanese physicists Kobayashi and Maskawa postulated the existence of a third generation of quarks and extended the Cabibbo matrix to a 3×3 matrix, including a complex phase, δ, responsible for CP violation [18]. Their theory found a first experimental confirmation in the discovery of the bottom quark, b, in 1977 by the E288 experiment, at Fermilab (USA), led by Lederman [19]. The discovery of the top quark, t, only came in 1994, with the measurements of the D0 and CDF collaborations at Fermilab [20, 21]. However, indirect hints for the existence of a heavy top quark were found earlier in measurements of the $B^0\bar{B}^0$ mixing, which was first observed in 1987 by the ARGUS experiment at the DORIS II electron-positron collider (DESY, Germany) [22].
In the 1990s, experiments like CLEO at Cornell (USA) made important advances in the study of B physics [23,24] but did not observe CP violation in B decays. A revolution in the field is represented by the advent of *B-factories*, which led to the BaBar experiment (1999-2010) at the PEP-II accelerator (SLAC, USA) and the Belle experiment (1999-2010) at the KEKB collider (KEK, Japan). In 2001, these experiments observed for the first time CP violation in B decays [25, 26], showing that CP violation was genuinely a feature of weak interactions and not a peculiarity of kaon mixing. The two experiments, in addition, performed many other measurements of observables related to the decays of B mesons, which helped to constrain the Kobayashi-Maskawa theory of CP violation and provided a better understanding of flavour physics [27]. After the shutdown of Belle and BaBar, the concept of *super B-factory* was introduced to achieve record instantaneous luminosity, thus leading to the upgrade of KEKB into SuperKEKB and to the setup of the Belle II experiment, which has started to collect data in March 2018. Since the early 2000s, flavour physics has been successfully studied also in hadronic collisions by experiments like D0 and CDF at the Tevatron (Fermilab, USA) and LHCb, CMS, ATLAS, at the LHC (CERN, Switzerland).
The last few years, in particular, have been thrilling for flavour physics. Precise comparisons between experimental measurements and SM predictions have revealed a coherent pattern of *flavour anomalies*. Tensions with the SM have been observed in $b \to sll$ [28–34] and $b \to c\tau\nu$ [35–39] transitions, in the muon anomalous magnetic moment [40,41] and more, as illustrated in Fig. 1.1. Such results are of great interest since flavour observables can be sensitive to energy scales beyond the direct reach of collider searches, and the current scenario seems to indicate the possible contribution of new phenomena. In this context, Belle II can measure $b \to s\nu\bar{\nu}$ transitions and other rare decays with unprecedented precision, helping to shed light on the puzzle of the flavour anomalies and giving a unique contribution to the quest for new physics.

1.2 Flavour in the Standard Model

The current theory of flavour is part of the Standard Model (SM) of particle physics, which was formulated in the mid-1970s as a result of decades of measurements, discoveries and successful predictions. At the current level of experimental precision and at the energies reached so far, the SM still provides the best description of the interplay between elementary particles and fundamental forces (except for gravity).

1.2.1 Particle content

The particle content of the SM is illustrated in Fig. 1.2. Each elementary particle is characterised by a mass, a spin, and by additional quantum numbers, also called charges, responsible for their interactions. The particles carrying spin $1/2$, the fermions, are arranged in three generations of four fermions each. Two of them are quarks, charged under the strong interaction, and two of them are leptons, which are not. The electromagnetic charges of the quarks are $+2/3$ (up quarks) and $-1/3$ (down quarks), while those of the leptons are -1 and 0 (neutrinos), in units where the electromagnetic charge of the electron is -1. The

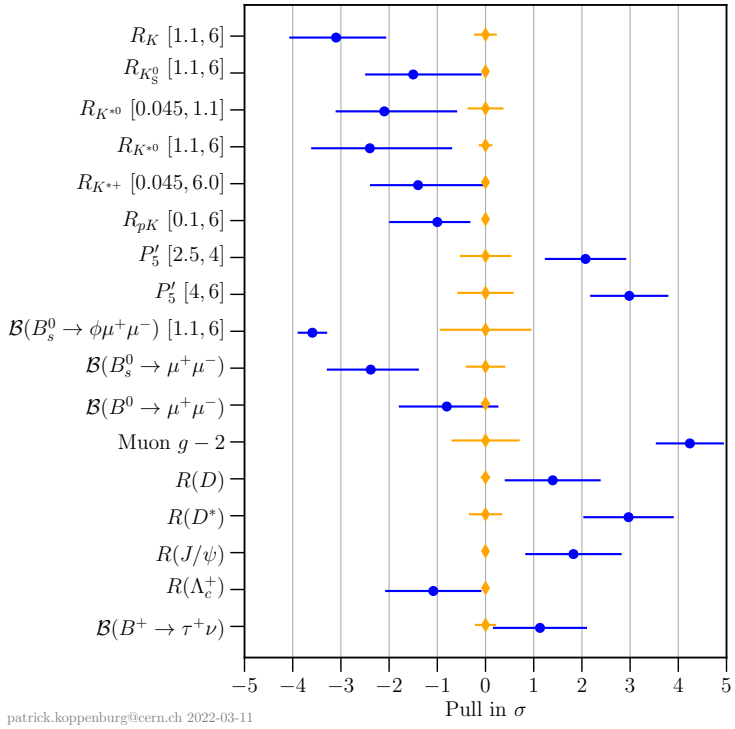


Figure 1.1: A set of cherry-picked flavour anomalies updated to March 2022. The SM expectation (orange diamond) is set to zero. Credit: P. Koppenburg, Ref. [42].

masses of the SM fermions range from values below $1\,\mathrm{eV}/c^2$, for the neutrinos, to $173\,\mathrm{GeV}/c^2$ [43] for the top quark, with a hierarchical separation among the masses of the three generations. Chirality is also a fundamental quantum number of the SM fermions. Charged fermions are described by the combination of two eigenstate of opposite chirality, the left-handed and the right-handed chiral states, while only neutrinos in the left-handed chiral state have been observed so far. The SM interactions are associated with the exchange of vector (spin-1) gauge bosons: the photon for the electromagnetic interaction; the gluons for the strong interaction; the Z and W for the weak interaction. The photon and the gluons are massless, while the Z and the W are massive, with $m_Z = 91\,\mathrm{GeV}/c^2$ and $m_W = 80\,\mathrm{GeV}/c^2$, respectively [43]. The last fundamental component of the SM is the Higgs boson, a scalar (spin-0) particle of mass $m_H = 125\,\mathrm{GeV}/c^2$ [43], responsible for the masses of the other fundamental particles of the SM. The existence of the Higgs boson is related to the breaking of the symmetry that nature exhibits at energies above $\sim 100\,\mathrm{GeV}$, the so-called electroweak scale, where the electromagnetic and weak interactions become indistinguishable and the left-handed chirality components of up and down fermions are unified in electroweak doublets. All this is described and successfully predicted within the framework of the SM.

1.2.2 Gauge structure

The SM is a non-abelian gauge theory [45] with local gauge symmetry group

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{1.2}$$

$SU(3)_C$ is the non-abelian gauge group of the strong interaction [46], with generators corresponding to the eight Gell-Mann matrices t^a [47, 48]. The quantum number related to the gauge symmetry is the *colour charge*, and the vector-boson gauge fields are the eight gluons G_μ^a. The field strength tensor of $SU(3)_C$ is

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \tag{1.3}$$

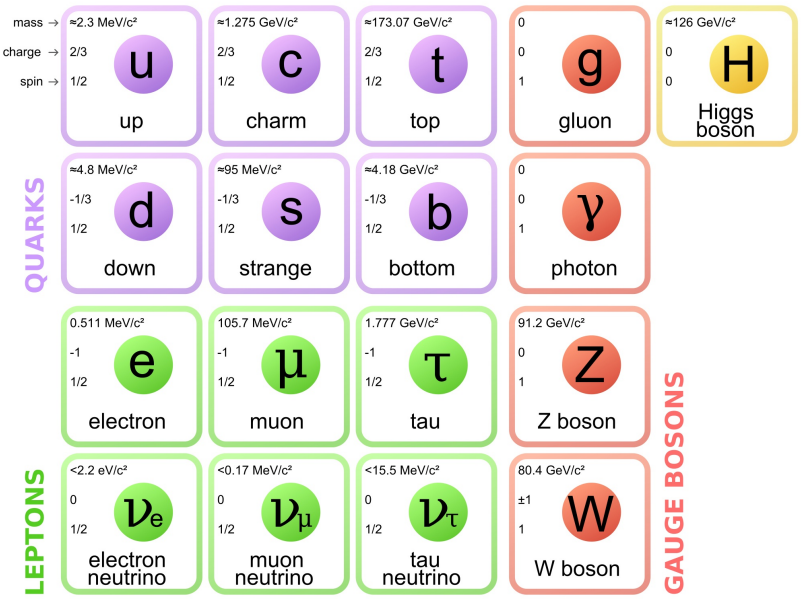


Figure 1.2: Fundamental components of the Standard Model of particle physics. Credit: Ref. [44].

with $a = 1, ..., 8$, where g_s is the coupling constant of the strong interaction and f^{abc} are the antisymmetric structure constants of $SU(3)_c$.
$SU(2)_L \times U(1)_Y$ is the gauge group of the Glashow-Weinberg-Salam model [49–51], which unifies the electromagnetic and weak interactions in the electroweak interaction. The $SU(2)_L$ gauge group has three generators, $t^a = \sigma^a/2$, where σ^a correspond to the Pauli matrices. The quantum number associated with the $SU(2)_L$ symmetry is the *weak isospin*, T, which can also be indicated with the third component T_3. Three vector-boson gauge fields, W^a_μ, are required and the field strength tensor corresponds to

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu, \tag{1.4}$$

with $a = 1, 2, 3$, where ϵ^{abc} is the totally antisymmetric three-index tensor and g is the coupling constant of the weak interaction. The gauge group $U(1)$ is generated by the *weak hypercharge*, $Y = 2(Q - T_3)$ [52, 53] and implies the existence of one gauge vector-boson B_μ. The coupling constant is g' and the gauge field tensor can be written as

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.5}$$

The self-interactions of the gauge fields are described by the following gauge-invariant Lagrangian

$$\mathcal{L}_{gauge} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.6}$$

while the interactions of the gauge fields with the SM fermions involve the covariant derivative

$$\mathcal{D}_\mu = \partial_\mu - ig' B_\mu Y - igW^a_\mu T^a - ig_s G^a_\mu t^a. \tag{1.7}$$

Strong interactions conserve all flavours, but flavour can change in electroweak interactions. Hence, in the following, the description will focus on the $SU(2)_L \times U(1)_Y$ model.

1.2.3 Fermions

The electroweak gauge group distinguishes between left-handed and right-handed quarks and leptons. The left-handed, ψ_L, and right-handed, ψ_R, chiral states of a Dirac fermion, ψ, are defined as

$$P_L\psi \equiv \psi_L, \qquad P_R\psi \equiv \psi_R, \tag{1.8}$$

Family	Left-chiral fermions				Right-chiral fermions			
		Q	(T, T_3)	Y		Q	(T, T_3)	Y
Leptons	$\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$	0	$(1/2, +1/2)$	-1	$\nu_{eR}, \nu_{\mu R}, \nu_{\tau R}$	Not part of the SM		
	e_L, μ_L, τ_L	-1	$(1/2, -1/2)$	-1	e_R, μ_R, τ_R	-1	$(0, 0)$	-2
Quarks	u_L, c_L, t_L	$+2/3$	$(1/2, +1/2)$	$+1/3$	u_R, c_R, t_R	$+2/3$	$(0, 0)$	$+4/3$
	d_L, s_L, b_L	$-1/3$	$(1/2, -1/2)$	$+1/3$	d_R, s_R, b_R	$-1/3$	$(0, 0)$	$-2/3$

Table 1.1: Quantum numbers of the electroweak gauge group $SU(2)_L \times U(1)_Y$ for left and right chiral fermions: Q is the electric charge; T and T_3 are the weak isospin and its third component; Y is the weak hypercharge.

where

$$P_L = \frac{1 - \gamma^5}{2}, \qquad P_R = \frac{1 + \gamma^5}{2} \tag{1.9}$$

are the projection operators of chirality, expressed through the Dirac matrix γ^5.
Based on the $SU(2)_L \times U(1)_Y$ quantum numbers, quarks and leptons are grouped into three *generations* of left-handed weak-isospin doublets

$$\mathbf{q}_L = \left(\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right), \tag{1.10}$$

$$\mathbf{l}_L = \left(\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \right), \tag{1.11}$$

and three generations of right-handed weak isospin singlets,

$$\mathbf{u}_R = (u_R, c_R, t_R), \tag{1.12}$$

$$\mathbf{d}_R = (d_R, s_R, b_R), \tag{1.13}$$

$$\mathbf{e}_R = (e_R, \mu_R, \tau_R). \tag{1.14}$$

The values of the quantum numbers are summarised in Table 1.1. Anti-fermions carry hypercharges opposite to those of the corresponding fermions, given the reversed sign of electric charge and T_3 under charge conjugation.

1.2.4 Symmetry breaking via Higgs mechanism

The electroweak gauge symmetry $SU(2)_L \times U(1)_Y$ is broken by the Higgs field $\mathbf{\Phi}$, a $SU(2)$ doublet of complex scalars, having hypercharge $Y = 1$. The Higgs field develops a $U(1)_{em}$ symmetric vacuum expectation value $\langle \mathbf{\Phi} \rangle_0$, causing the Spontaneous Symmetry Breaking (SSB) of the electroweak group into the electromagnetic subgroup [54, 55]:

$$SU(2)_L \times U(1)_Y \xrightarrow{\langle \mathbf{\Phi} \rangle_0} U(1)_{em}\,. \tag{1.15}$$

A $U(1)_{em}$ symmetric $\langle \mathbf{\Phi} \rangle_0$ implies that one scalar of the $SU(2)$ doublet must be neutral, so that the Higgs field can be written as

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{1.16}$$

where ϕ_1, ϕ_2, ϕ_3, ϕ_4 are real scalar fields.
The dynamics of the Higgs field is described in the SM by the gauge-invariant Lagrangian

$$\mathcal{L} = (\mathcal{D}_\mu \mathbf{\Phi})^\dagger (\mathcal{D}^\mu \mathbf{\Phi}) - V(\mathbf{\Phi}) + \mathcal{L}_{\text{Yukawa}}, \tag{1.17}$$

with the first term describing the kinetic and gauge interactions of the Higgs field, the second term representing a potential energy, and the third term referring to the Yukawa interactions between the Higgs field and pairs of fermions. The most generic gauge-invariant potential energy, involving $\mathbf{\Phi}$, corresponds to

$$V(\mathbf{\Phi}) = -\mu^2 \mathbf{\Phi}^\dagger \mathbf{\Phi} + \lambda(\mathbf{\Phi}^\dagger \mathbf{\Phi})^2, \tag{1.18}$$

and when $-\mu^2 < 0$ and $\lambda > 0$, $V(\mathbf{\Phi})$ has a minimum away from $|\mathbf{\Phi}| = 0$, specifically at $|\mathbf{\Phi}| = \sqrt{\mu^2/\lambda}$, as illustrated in Fig. 1.3. The minimisation only fixes the modulus of $\mathbf{\Phi}$, meaning that the vacuum is not a singlet of the symmetry group. It is possible to set a specific basis of states ϕ_1, ϕ_2, ϕ_3, ϕ_4 as the physical vacuum, but the choice of one state, out of an infinite number of degenerate ground states, breaks the symmetry. Choosing the physical vacuum as the one for which

$$\langle \phi_3 \rangle \equiv v = \sqrt{\frac{\mu^2}{\lambda}}, \qquad \langle \phi_1 \rangle = \langle \phi_2 \rangle = \langle \phi_4 \rangle = 0, \tag{1.19}$$

the vacuum expectation value of the Higgs is

$$\langle \mathbf{\Phi} \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.20}$$

The field ϕ_3 can be defined as $\phi_3 = v + h$, where $\langle h \rangle = 0$, and from Eq. 1.18 it follows that h is a massive field, corresponding to the actual Higgs boson with $m_h = \sqrt{v^2\lambda}$, while ϕ_1, ϕ_2, ϕ_4 are massless fields. In particular, the scalar fields ϕ_1, ϕ_2, ϕ_4 do not represent physical degrees of freedom, as their contribution can be entirely removed from the Lagrangian in Eq. 1.17 by means of a gauge transformation.

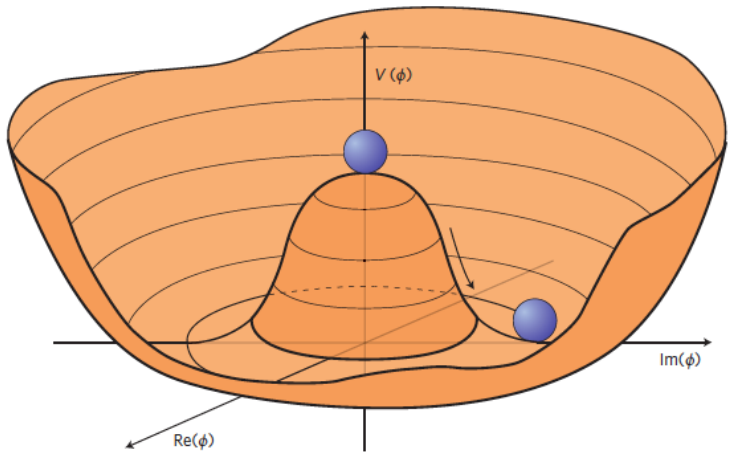


Figure 1.3: An illustration of the Higgs potential for $-\mu^2 < 0$. In this case, the minimum is at $|\mathbf{\Phi}|^2 = -\mu^2/(2\lambda)$. Credit: Ref. [56].

1.2.5 Interactions of the Higgs with gauge bosons

The gauge-kinetic term in the Higgs Lagrangian of Eq. 1.17 corresponds to $(\mathcal{D}_\mu \mathbf{\Phi})^\dagger (\mathcal{D}^\mu \mathbf{\Phi})$, with the covariant derivative 1.7 assuming the specific form

$$\mathcal{D}_\mu = \partial_\mu - i\frac{g'}{2} B_\mu - i\frac{g}{2} W_\mu^a \sigma^a, \tag{1.21}$$

such that its application to the Higgs doublet in the vacuum state gives

$$(\mathcal{D}_\mu \mathbf{\Phi})^\dagger (\mathcal{D}^\mu \mathbf{\Phi}) = \frac{1}{8} g^2 v^2 (W_\mu^1 - iW_\mu^2)(W^{1\mu} + iW^{2\mu}) + \frac{1}{8} v^2 (-g' B_\mu + g W_\mu^3)^2. \tag{1.22}$$

The combinations $(W^1 \pm iW^2)$ in Eq. 1.22 correspond to the charged W bosons

$$W_\mu^+ \equiv \frac{W_\mu^1 - iW_\mu^2}{\sqrt{2}}, \qquad W_\mu^- \equiv \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}}, \tag{1.23}$$

therefore Eq. 1.22 becomes

$$(\mathcal{D}_\mu \mathbf{\Phi})^\dagger (\mathcal{D}^\mu \mathbf{\Phi}) = \frac{g^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (-g' B_\mu + g W_\mu^3)^2, \tag{1.24}$$

and the squared mass of the W boson corresponds to

$$m_W^2 = \frac{g^2 v^2}{4}. \tag{1.25}$$

The Higgs vacuum expectation value is responsible for the mass of the W boson, and from the measured values of m_W and g it is possible to determine $v \simeq 246\,\text{GeV}$.
The linear combination of W_μ^3 and B_μ appearing in the second term of Eq. 1.24 can be written as

$$\begin{aligned} (gW_\mu^3 - g'B_\mu) &= \sqrt{g^2+g'^2}\left(\frac{g}{\sqrt{g^2+g'^2}} W_\mu^3 - \frac{g'}{\sqrt{g^2+g'^2}} B_\mu\right) \\ &\equiv \sqrt{g^2+g'^2}(c_W W_\mu^3 - s_W B_\mu) \\ &\equiv \sqrt{g^2+g'^2} Z_\mu, \end{aligned} \tag{1.26}$$

with $s_W = \sin\theta_W$, $c_W = \cos\theta_W$, where θ_W is the weak mixing angle, also called the Weinberg angle [50].
The field combination Z_μ represents the Z boson, whose mass is given by

$$m_Z^2 = \frac{(g^2+g'^2)v^2}{4} \tag{1.27}$$

as a direct consequence of the Higgs vacuum expectation value.
The state orthogonal to Z_μ is

$$A_\mu \equiv (s_W W_\mu^3 + c_W B_\mu). \tag{1.28}$$

It does not couple to the Higgs field, meaning that it does not acquire a mass via the Higgs mechanism, and it can be identified with the photon.

1.2.6 Fermion masses and CKM matrix

The Lagrangian in Eq. 1.17, describing the Yukawa interaction between the Higgs field and the SM fermions, is

$$\mathcal{L}_{\text{Yukawa}} = -\left[\widetilde{\mathbf{\Phi}} \bar{\mathbf{q}}_L \lambda_U \mathbf{u}_R + \mathbf{\Phi} \bar{\mathbf{q}}_L \lambda_D \mathbf{d}_R + \mathbf{\Phi} \bar{\mathbf{l}}_L \lambda_E \mathbf{e}_R + \text{h.c}\right], \tag{1.29}$$

where

$$\bar{\mathbf{q}}_L \equiv q_{Lj}, \qquad \mathbf{u}_R \equiv u_{Rj}, \qquad \mathbf{d}_R \equiv d_{Rj}, \qquad j = 1,2,3 \tag{1.30}$$

$$\bar{\mathbf{l}}_L \equiv l_{Lj}, \qquad \mathbf{e}_R \equiv e_{Rj}, \qquad j = 1,2,3 \tag{1.31}$$

are the flavour eigenstates (quark and lepton fields with a generation index j), $\widetilde{\mathbf{\Phi}} = i\sigma^2 \mathbf{\Phi}^*$ represents the conjugate Higgs doublet, and λ_U, λ_D, λ_E are the complex 3×3 Yukawa matrices.
Replacing the Higgs doublet with its vacuum expectation value of Eq. 1.20, the Yukawa Lagrangian becomes

$$\mathcal{L}_{\text{Yukawa}} = -(\bar{u}_1, \bar{u}_2, \bar{u}_3)_R \mathcal{M}_U \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}_L - (\bar{d}_1, \bar{d}_2, \bar{d}_3)_R \mathcal{M}_D \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_L - (\bar{e}_1, \bar{e}_2, \bar{e}_3)_R \mathcal{M}_E \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}_L + \text{h.c.}, \tag{1.32}$$

where

$$\mathcal{M}_{Uij} = \frac{v}{\sqrt{2}}\lambda_{Uij}, \qquad \mathcal{M}_{Dij} = \frac{v}{\sqrt{2}}\lambda_{Dij}, \qquad \mathcal{M}_{Eij} = \frac{v}{\sqrt{2}}\lambda_{Eij}, \tag{1.33}$$

are the quark and lepton mass matrices in the generation space, each made of 9 complex entries.
To find the quark mass eigenstates, the matrices $\mathcal{M}_U$, $\mathcal{M}_D$ have to be diagonalised. They can be transformed into real diagonal matrices by the application of appropriate unitary matrices U_L, U_R, D_L, D_R, such that

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}_{L,R} = U_{L,R} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R}, \qquad \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_{L} = D_{L,R} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R}, \tag{1.34}$$

where u, c, t, d, s, b are the quark mass eigenstates, and

$$U_R^{-1}\mathcal{M}_U U_L = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}, \qquad D_R^{-1}\mathcal{M}_D D_L = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}. \tag{1.35}$$

Transforming the up-type quarks by U_L and the down-type quarks by D_L, as in Eq. 1.34, the $SU(2)$ left-handed quark doublets appear to be *broken*, and generation-changing weak interactions are allowed.
The gauge-invariant Lagrangian describing the interactions among gauge bosons and quark flavour eigenstates ψ has the form

$$\mathcal{L}_{\text{fermion}} = \bar{\psi} i\gamma^\mu \mathcal{D}_\mu \psi, \tag{1.36}$$

with the covariant derivative $\mathcal{D}_\mu$ acting on left-handed quark doublets as

$$\mathcal{D}_\mu \mathbf{q}_L = \left(\partial_\mu - i\frac{g'}{6}B_\mu - ig\frac{\sigma^a}{2}W_\mu^a\right)\mathbf{q}_L, \tag{1.37}$$

and on right-handed quark singlets as

$$\mathcal{D}_\mu \mathbf{u}_R = \left(\partial_\mu - ig'\frac{2}{3}B_\mu\right)\mathbf{u}_R, \tag{1.38}$$

$$\mathcal{D}_\mu \mathbf{d}_R = \left(\partial_\mu + i\frac{g'}{3}B_\mu\right)\mathbf{d}_R. \tag{1.39}$$

Hence, by means of Eqs. 1.23, 1.26 and 1.28, the Lagrangian in Eq. 1.36 can be written as

$$\begin{aligned} \mathcal{L}_{\text{fermion}} &= \frac{g}{\sqrt{2}}(J_\mu^+ W^{+\mu} + J_\mu^- W^{-\mu}) + eQJ_\mu^{em}A^\mu + \frac{g}{\cos\theta_W}J_\mu^0 Z^\mu \\ &\equiv \mathcal{L}_{\text{CC}} + \mathcal{L}_{\text{NC}} \end{aligned} \tag{1.40}$$

In Eq. 1.40 the neutral-current interactions are described by $\mathcal{L}_{\text{NC}}$, in which the photon A^μ couples with the electromagnetic current $J_\mu^{em} \equiv \bar{\psi}\gamma^\mu\psi$ and the Z^μ boson with the fermionic neutral-current

$$J_\mu^0 = g_L\left(\bar{\mathbf{u}}_L\gamma_\mu\mathbf{u}_L + \bar{\mathbf{d}}_L\gamma_\mu\mathbf{d}_L\right) + g_R\left(\bar{\mathbf{u}}_R\gamma_\mu\mathbf{u}_R + \bar{\mathbf{d}}_R\gamma_\mu\mathbf{d}_R\right), \tag{1.41}$$

where

$$g_L = T_3 - Q\sin^2\theta_W, \tag{1.42}$$

$$g_R = -Q\sin^2\theta_W. \tag{1.43}$$

The charged-current interactions are described by $\mathcal{L}_{CC}$, where the charged $W_\mu^\pm$ boson is coupled with the fermionic charged-current $J_\mu^\pm$, which can be written as

$$J_\mu^+ = (\bar{u}_1, \bar{u}_2, \bar{u}_3)_L\gamma_\mu \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_L = (\bar{u}, \bar{c}, \bar{t})_L U_L^\dagger \gamma_\mu D_L \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L = (\bar{u}, \bar{c}, \bar{t})_L \gamma_\mu V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L, \tag{1.44}$$

where $V_{\rm CKM} \equiv U_L^\dagger D_L$ is the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [12,18]

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.45}$$

The matrix can be parametrised by means of three mixing angles and one CP-violating phase, also known as the Kobayashi-Maskawa phase. The standard parameterisation [57] is

$$\begin{aligned} V_{\rm CKM} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{!3}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \end{aligned} \tag{1.46}$$

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$, with angles θ_{ij} that can be chosen such that $s_{ij}, c_{ij} \geq 0$, and δ is the phase responsible for CP-violation in flavour-changing processes in the SM.
From experimental measurements it is found that $s_{13} \ll s_{23} \ll s_{12} \ll 1$, and it is convenient to exhibit this hierarchy adopting the Wolfenstein parametrisation, with the following substitutions [58–60]:

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}; \tag{1.47}$$

$$s_{23} = A\lambda^2 = \lambda\frac{|V_{cb}|}{|V_{us}|}; \tag{1.48}$$

$$s_{13}e^{i\delta} = V_{ub}^* = A\lambda^3(\rho + i\eta) = \frac{A\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1-\lambda^2}[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})]}. \tag{1.49}$$

Using these relations, $\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$ is independent of the phase convention, and the CKM matrix in terms of λ, A, $\bar{\rho}$, $\bar{\eta}$ is unitary to all orders in λ. Traditionally the CKM matrix approximated to $\mathcal{O}(\lambda^3)$,

$$V_{\rm CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{1.50}$$

The unitarity of the CKM matrix, $V_{\rm CKM}V_{\rm CKM}^\dagger = V_{\rm CKM}^\dagger V_{\rm CKM} = 1$, implies that $\sum_k V_{ik}V_{jk}^* = 0$ for $j \neq i$. This sum can be represented in the complex plane by a triangle for each combination of rows or columns identified by i and j. The triangles obtained by the product of neighbouring rows or columns are nearly degenerate, thus the unitarity triangle that is conventionally used is the one obtained from the ratio of

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.51}$$

by its middle term $V_{cd}V_{cb}^*$. It corresponds to a triangle with unit base in the $(\bar{\rho}, \bar{\eta})$ plane, as illustrated in Fig. 1.4. Its vertices are $(0, 0)$, $(1, 0)$ and $(\bar{\rho}, \bar{\eta})$ in agreement with Eqs. 1.47, 1.48, 1.49, and its angles are:

$$\beta = \phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \tag{1.52}$$

$$\alpha = \phi_2 = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \tag{1.53}$$

$$\gamma = \phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.54}$$

The area of the triangle is equal to half of the Jarlskog invariant J [61], which is a measure of CP violation, invariant under phase redefinitions of the quark fields, defined by

$$\mathrm{Im}[V_{ij}V_{kl}V_{il}^*V_{kj}^*] = J(\delta_{ij}\delta_{kl} - \delta_{il}\delta_{kj}). \tag{1.55}$$

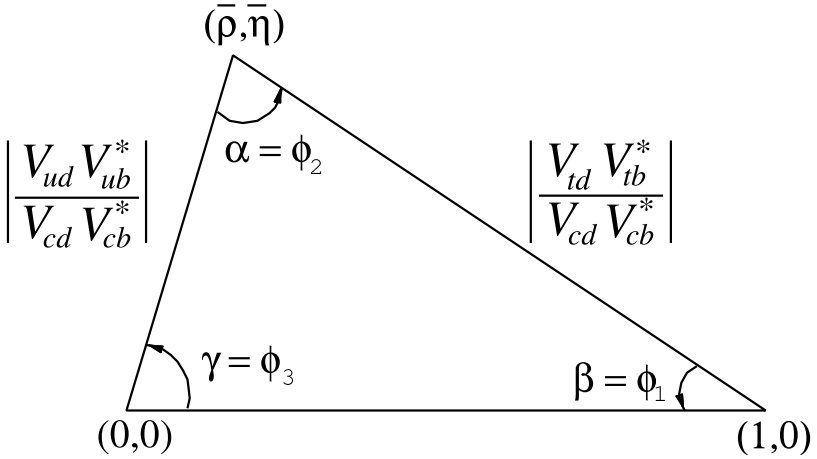


Figure 1.4: Sketch of the conventional unitarity triangle in the $(\bar{\rho}, \bar{\eta})$ plane. Credit: Ref. [43].

1.2.6.1 Experimental measurement of CKM elements and global fit

The magnitudes of the CKM-matrix elements can be independently measured in specific processes involving flavour transitions. $|V_{ud}|$ is measured in allowed nuclear transitions; $|V_{us}|$, $|V_{cd}|$, $|V_{cs}|$, $|V_{ub}|$, $|V_{cb}|$ are probed in measurements of semileptonic decays of mesons; $|V_{tq}|$, with $q = d, s, b$, can be inferred either from processes occurring beyond the tree level with the contribution of a virtual top quark, or from single top production or decay. Using such measurements, the unitarity of the CKM matrix can be tested. As reported in Ref. [43], currently $|V_{ud}|^2+|V_{us}|^2+|V_{ub}|^2 = 0.9985\pm0.005$ (first row), $|V_{cd}|^2+|V_{cs}|^2+|V_{cb}|^2 = 1.025\pm0.022$ (second row), and $|V_{ud}|^2+|V_{cd}|^2+|V_{td}|^2 = 0.9970\pm0.0018$ (first column), $|V_{us}|^2+|V_{cs}|^2+|V_{ts}|^2 = 1.026\pm0.022$ (second column). In particular, at the moment there is a 3σ tension with unitarity in the first row. The sum of the angles α, β, γ, determined from independent measurements, is $\alpha+\beta+\gamma = (179^{+7}_{-6})^\circ$, which is consistent with the SM expectation.
An alternative determination of the CKM-matrix elements is obtained with a global fit to all the available measurements, imposing SM constraints like the three-generation unitarity. The experimental data are combined by CKMfitter [60,62,63] with frequentist statistics and by UTfit [64,65] with a Bayesian approach. The most recent fit values for the magnitudes of the CKM elements, provided by the Particle Data Group in Ref. [43], are

$$V_{\mathrm{CKM}} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix}. \tag{1.56}$$

A graphical illustration of the constraints on the $(\bar{\rho}, \bar{\eta})$ plane, and of the global fit result in terms of the Wolfenstein parameters, is provided in Fig. 1.5.

1.2.7 Detour into neutrino masses

As of today, neutrinos are the only particles in the SM that have been observed only with left-handed chirality. The absence of right-handed neutrinos implies that no Yukawa-like interaction with the Higgs field is possible and they cannot acquire mass in the SM.
If neutrinos are Dirac fermions, then three right-handed neutrino fields, ν_{Rj}, should exist and can be introduced with a corresponding additional term in the Yukawa Lagrangian of Eq. 1.29. In this way, the part of Lagrangian describing charged and neutral leptons would become

$$\mathcal{L}_{\mathrm{Yukawa}} \supset -\mathbf{\Phi}\bar{\mathbf{l}}_L\lambda_e\mathbf{e}_R - \widetilde{\mathbf{\Phi}}\bar{\mathbf{l}}_L\lambda_\nu\nu_R. \tag{1.57}$$

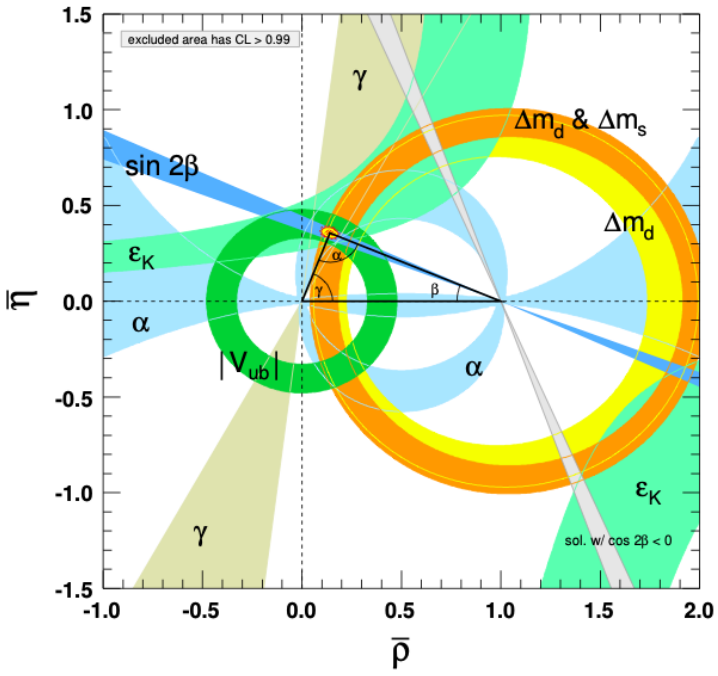


Figure 1.5: Constraints on the $(\bar{\rho}, \bar{\eta})$ plane. The shaded areas have all 99% CL. Credit: Ref. [43].

As demonstrated in Eqs. 1.33, 1.34, 1.35, these terms would generate two real and diagonal mass matrices, one for the lepton and one for the neutrino mass eigenstates. The mass eigenstates of the neutrinos ν_1, ν_2, ν_3, conventionally identified by the generation index, are related to the flavour eigenstates ν_e, ν_μ, ν_τ in the Lagrangian via the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix U [66, 67],

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}_L = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}_L . \tag{1.58}$$

However, under the hypothesis of Dirac neutrinos, the Yukawa couplings needed to generate their masses are extremely small: for a neutrino mass $m_\nu \sim 0.1\,\mathrm{eV}/c^2$ the coupling strength must be

$$\frac{\lambda_\nu}{\sqrt{2}} = \frac{m_\nu}{v} \approx 4 \times 10^{-13}. \tag{1.59}$$

The alternative is that neutrinos are Majorana fermions [68], which correspond to their own antiparticles. In this case, the mass term for neutrinos in the SM would be $m\nu_L\nu_L$, which is not gauge invariant under $SU(2)_L \times U(1)_Y$. However, it can be generated after the SSB by a Lagrangian involving the Higgs field $\mathbf{\Phi}$,

$$\mathcal{L}_{\mathrm{Majorana}} = -\frac{(\tilde{\mathbf{\Phi}} \mathbf{l}_L)^2}{\Lambda}, \tag{1.60}$$

which generates a neutrino mass $m_\nu = v^2/2\Lambda$. This Lagrangian describes a non-renormalizable interaction with cutoff scale Λ, beyond which a more complete theory should be adopted. In such theory, the Majorana mass term involves the presence of a very heavy Majorana right-handed neutrino ν_R with mass of the order of $\Lambda \sim 100\,\mathrm{TeV}$ for each existing light, left-handed, neutrino. This mechanism is known as “Type-I Seesaw" [69].

1.2.8 Neutral current interactions and GIM mechanism

In the neutral current interactions, described by the Lagrangian $\mathcal{L}_{\rm NC}$ in Eq. 1.40, the couplings with the photon and the Z boson are the same for the three generations. Moreover, the terms making up the neutral current operator J^0_μ, defined in Eq. 1.41, correspond to

$$(\bar{u}_1, \bar{u}_2, \bar{u}_3)_{L,R}\gamma_\mu \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}_{L,R} = (\bar{u}, \bar{c}, \bar{t})_{L,R} U^\dagger_{L,R}\gamma_\mu U_{L,R} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} = (\bar{u}, \bar{c}, \bar{t})_{L,R}\gamma_\mu \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \tag{1.61}$$

and

$$(\bar{d}_1, \bar{d}_2, \bar{d}_3)_{L,R}\gamma_\mu \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_{L,R} = (\bar{d}, \bar{s}, \bar{b})_{L,R} D^\dagger_{L,R}\gamma_\mu D_{L,R} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} = (\bar{d}, \bar{s}, \bar{b})_{L,R}\gamma_\mu \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R}, \tag{1.62}$$

where $U_{L,R}(D_{L,R})$ represents the unitary matrix transforming the flavour eigenstates u_1, u_2, u_3 (d_1, d_2, d_3) into the mass eigenstates u, c, t (d, s, b). Equations 1.61, 1.62 imply that neutral-current interactions are *flavour diagonal*. This explains why flavour-changing neutral-currents (FCNCs) are forbidden at the tree level in the SM.
Transitions like $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ ($l = e, \mu, \tau$), represent typical examples of FCNC processes. In the SM, they can only occur at the loop level, as illustrated in Fig. 1.6, and they are largely suppressed.

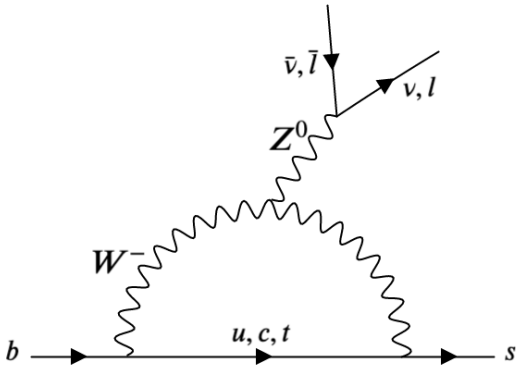


Figure 1.6: One-loop *penguin* Feynman diagram for a $b \to s\nu\bar{\nu}$, or a $b \to sl^+l^-$, transition.

The $b \to s$ transition amplitude is proportional to

$$\mathcal{M} \sim \frac{g^2}{16\pi^2} \sum_{i=u,c,t} V_{ib}V^*_{is} F\left(\frac{m_i^2}{m_W^2}\right), \tag{1.63}$$

where $g^2/16\pi^2$ is the 1-loop suppression factor relative to the flavour-changing current and F is a Inami-Lim function [70]. For quark masses $m_i \ll m_W$, F can be expanded in a power series. By choosing $F(0) = 0$, it is possible to write

$$\sum_{i=u,c,t} V_{ib}V^*_{is} F\left(\frac{m_i^2}{m_W^2}\right) = F\left(\frac{m_t^2}{m_W^2}\right) V_{tb}V^*_{ts} + F'(0) \sum_{i=u,c} V_{ib}V^*_{is}\frac{m_i^2}{m_W^2} + \text{h.c.} \tag{1.64}$$

The contribution from the u and c quarks in Eq. 1.64 is completely negligible, therefore

$$\mathcal{M} \sim \frac{g^2}{16\pi^2} V_{tb}V^*_{ts} F\left(\frac{m_t^2}{m_W^2}\right). \tag{1.65}$$

The function $F(x)$ is slowly increasing and it is $\mathcal{O}(1)$ at the top-quark mass [71]. Hence, the transition is dominated by the virtual top-quark exchange, which carries a CKM suppression factor $V_{tb}V_{ts}^* \sim 10^{-2}$.
The mechanism through which FCNC transitions are suppressed in loop diagrams was uncovered in 1964 by Glashow, Iliopoulos and Maiani [13]. At that time, only the three quarks u, d, s were known, and large discrepancies were observed between the predicted and measured rates of the $K_L^0 \to \mu^+\mu^-$ decay and $K^0\bar{K}^0$ mixing. The observed suppression of these FCNC processes was explained by the GIM mechanism predicting a one-loop transition in which a new quark, the c quark, is involved. In this way, the FCNC transition is suppressed by the product of a loop factor $g^2/16\pi^2$ and a factor m_c^2/m_W^2 related to the mass of the hypothesised c-quark. Based on the experimental results, they were able to set an upper limit on the charm mass, which turned out to be close to its actual value. The charm was actually discovered a few years later in 1974 at SLAC [15] and Brookhaven [16].

1.3 Effective Weak Hamiltonian formalism

The theoretical description of meson decays involving quark-flavour transitions is affected by the confinement of quarks inside hadrons. The strong dynamics in QCD bound states is characterised by interactions taking place at energy scales between $\mathcal{O}(100\,\mathrm{MeV})$ and $\mathcal{O}(1\,\mathrm{GeV})$. Therefore, an effective theory that allows combining together contributions from processes occurring at different energy scales is needed.
The effective field theory described in this section is developed as a generalisation of the Fermi theory for β-decays in the formulation given by Feynman and Gell-Mann in the late 1950s [11]. In order to provide a theoretical description of β-decays, like the one illustrated in Fig. 1.7 (a), they proposed an effective Hamiltonian of the form

$$\mathcal{H}_{eff}^{\beta} = \frac{G_F}{\sqrt{2}} \cos\theta_C \left[\bar{u}\gamma_\mu(1-\gamma_5)d\right]\left[\bar{e}\gamma^\mu(1-\gamma_5)\nu_e\right], \tag{1.66}$$

where the Fermi constant, G_F, and the cosine of the Cabibbo angle, $\cos\theta_C$, multiply an effective local operator, which is the product of two vector−axial $(V-A)$ currents.

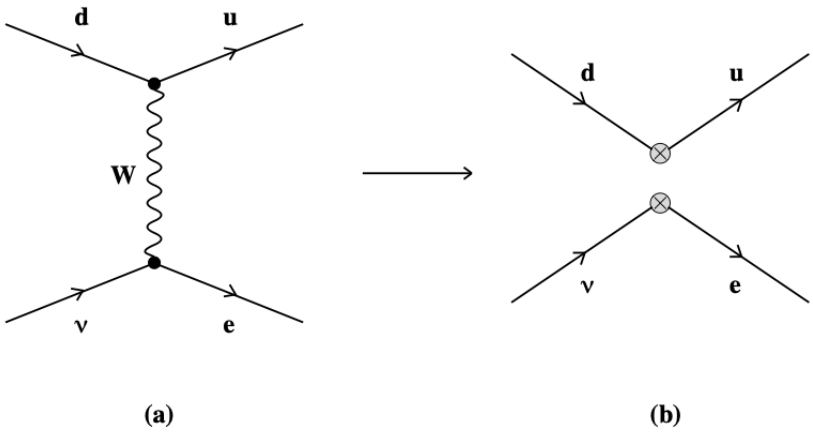


Figure 1.7: β-decay at the quark level in the complete (a) and effective (b) theory.

In analogy with the Fermi theory, the weak decays of hadrons with masses $\mathcal{O}(m_b, m_c, m_K)$ – where m_b, m_c, m_K are the masses of the b quark, c quark and kaon, respectively – can be described by effective point-like vertices, like those sketched in Fig. 1.7 (b). These vertices correspond to local operators and replace Feynman diagrams with full W and Z propagators. Using an Operator Product Expansion

(OPE) [72,73], the effective Hamiltonian of a generic weak decay can be written as

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V_i C_i(\mu) Q_i, \tag{1.67}$$

where the effective operators Q_i are multiplied by the CKM matrix elements V_i and by effective coupling constants C_i, the so-called Wilson Coefficients [74]. The coefficients do not depend on the specific decay, but they are function of the energy scale, μ, and they quantify the physics contributions originating at energies higher than μ. Because of the asymptotic freedom of QCD [75], the Wilson Coefficients can be calculated in perturbation theory if μ is not too small. They generically depend on the mass of the top quark, m_t, and also on the masses of new particles if extensions of the SM are taken into account. This dependence is determined by evaluating higher order Feynman diagrams, where the full propagators and vertices are computed [76].
Starting from $\mathcal{H}_{eff}$ in Eq. 1.67, it is possible to evaluate the transition amplitude for the weak decay of a hadron H:

$$A(H \to X) = \langle X|\mathcal{H}_{eff}|H\rangle = \frac{G_F}{\sqrt{2}} \sum_i V_i C_i(\mu) \langle X|Q_i(\mu)|H\rangle, \tag{1.68}$$

where X is the final state of the decay, and $\langle X|Q_i(\mu)|H\rangle$ are the corresponding matrix elements of Q_i. The matrix elements also depend on the energy scale μ, and they model the physics contributions from energies lower than μ. A great advantage of the OPE is that it allows factorising the decay amplitude into the sum of short-distance (perturbative) terms, represented by the couplings $C_i(\mu)$, multiplied by long-distance (generally non-perturbative) terms, represented by the matrix elements $\langle Q_i(\mu)\rangle$. Form factors of hadrons, characterising their momentum-dependent interactions with gauge bosons, are encoded in the matrix elements $\langle Q_i(\mu)\rangle$. Hadronic form factors are evaluated by means of non-perturbative methods, like Lattice QCD [77] and QCD Sum Rules [78], but also using the Heavy Quark Effective Theory [79,80] in the specific case of semileptonic B decays.
Computations are easier for inclusive decays of heavy mesons, like B mesons. The decay amplitude, in this case, is obtained by summing Eq. 1.68 over all the possible final states of the quark flavour transition of interest,

$$A(B \to X) = \frac{G_F}{\sqrt{2}} \sum_{(f \in X)} V_i C_i(\mu) \langle f|Q_i(\mu)|B\rangle, \tag{1.69}$$

where X identifies a specific collection of final states and B is the decaying B meson. In this case, the corresponding branching fraction can be computed by means of an expansion in powers of m_b^{-1}, and its leading term is described by the spectator model [81] describing the B-meson decay via the b-quark decay:

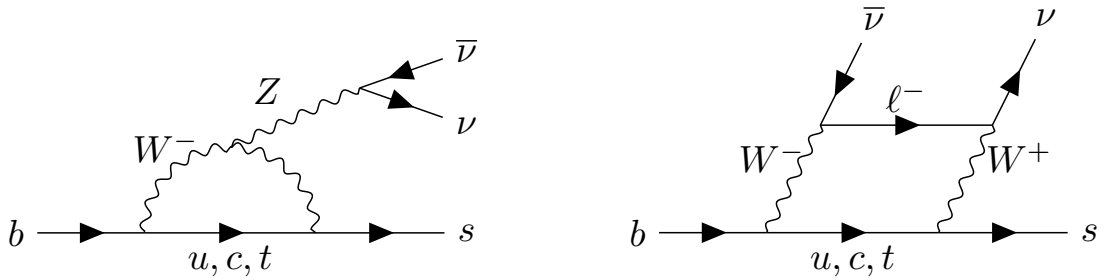
$$\mathrm{Br}(B \to X) = \mathrm{Br}(b \to q) + \mathcal{O}\left(\frac{1}{m_b^2}\right). \tag{1.70}$$

The formula in 1.70 is known as Heavy Quark Expansion [82,83]. Here, given that the leading term in the expansion represents the decay of the b quark, the matrix elements $\langle Q_i(\mu)\rangle$ can be computed in perturbation theory. For this reason, the current theoretical predictions for inclusive decays are more precise.

1.4 The $B \to K^{(*)}\nu\bar{\nu}$ decays

Weak decays of heavy mesons involving $b \to s$ FCNC transitions are of particular interest, since they are expected to play an important role in tests of the SM and in the search for new physics. In this thesis, a specific class of such decays is investigated: the $B \to K^{(*)}\nu\bar{\nu}$ decays. These decays are characterised by the $b \to s\nu\bar{\nu}$ transition, which occurs at the loop level in the SM with the exchange of at least two gauge bosons, as illustrated in Fig. 1.8. The absence of charged leptons in the final state makes the $B \to K^{(*)}\nu\bar{\nu}$ decays theoretically cleaner than the similar $B \to K^{(*)}l^+l^-$ decays. These two processes are governed by the same $B \to K^{(*)}$ form factors, but the photon exchange in $B \to K^{(*)}l^+l^-$ produces *non-factorizable*

effects, which are not yet well described by theorists and represent a major source of uncertainty in the theoretical predictions [84, 85]. Conversely, the $B \to K^{(*)}\nu\bar{\nu}$ decays do not suffer from hadronic uncertainties beyond the form factors, which are reliably predicted, thus they represent optimal processes to probe the SM.



(a) Penguin diagram (b) Box diagram

Figure 1.8: Lowest-order quark-level diagrams for the $b \to s\nu\bar{\nu}$ transition in the SM.

In the SM, the $B \to K^{(*)}\nu\bar{\nu}$ decays can be described by means of the effective weak Hamiltonian

$$\mathcal{H}_{eff}^{SM} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* C_L^{SM} O_L + \text{h.c}, \tag{1.71}$$

where

$$O_L = \frac{e^2}{16\pi^2} \left[\bar{s}\gamma_\mu P_L b\right] \left[\bar{\nu}\gamma^\mu(1-\gamma_5)\nu\right] \tag{1.72}$$

is the local operator representing the point-like vertex interaction involving the b and s quarks and the two neutrinos, with P_L being the chirality projector $(1-\gamma_5)/2$ [85]. The Wilson Coefficient coupled to O_L is

$$C_L^{SM} = -X_t/s_W^2, \qquad X_t = 1.469 \pm 0.017, \tag{1.73}$$

computed in the SM with accuracy up to next-to-leading-order QCD corrections [86–88] and to two-loop electroweak contributions [89].

1.4.1 $B \to K\nu\bar{\nu}$ in the Standard Model

In the $B \to K\nu\bar{\nu}$ decays, the pseudo-scalar ($J^P = 0^-$) $B^+(B^0)$ meson produces the pseudo-scalar $K^+(K^0)$ meson and two neutrinos. The SM differential branching fraction of the decay can be totally characterised in terms of the measurable dineutrino invariant mass squared $q^2 = (p_\nu + p_{\bar{\nu}})^2$ as done in Ref. [85],

$$\frac{d\text{BR}(B \to K\nu\bar{\nu})_{\text{SM}}}{dq^2} = \tau_B 3|N|^2 \frac{X_t^2}{s_W^4} \rho_K(q^2). \tag{1.74}$$

In Eq. 1.74, the factor 3 indicates the inclusive sum over the three neutrino flavours and τ_B is the lifetime of the B meson. The normalisation factor N is defined as

$$N = V_{tb}V_{ts}^* \frac{G_F \alpha}{16\pi^2} \sqrt{\frac{m_B}{3\pi}}, \tag{1.75}$$

with α corresponding to the fine-structure constant and m_B to the mass of the B meson. The term parametrising the q^2 dependence is the rescaled form factor

$$\rho_K(q^2) = \frac{\lambda_K^{3/2}(q^2)}{m_B^4} \left[f_+^K(q^2)\right]^2, \tag{1.76}$$

where $\lambda_{K^{(*)}} \equiv \lambda(m_B^2, m_{K^{(*)}}^2, q^2)$ is the polynomial

$$\lambda(a,b,c) = a^2 + b^2 + c^2 - 2(ab + bc + ac), \tag{1.77}$$

and $f_+^K(q^2)$ is the actual SM form factor for $B \to K$ transitions. The form factor is encoded in the hadronic matrix element $\langle K|\bar{s}\gamma_\mu(1-\gamma_5)b|B\rangle$ originating from Eq. 1.72. It can be represented by means of the series expansion

$$f_+^K(q^2) = \frac{1}{1 - q^2/m_+^2} \sum_k \alpha_k \left[z(q^2)\right]^k, \tag{1.78}$$

depending on the real parameters α_k. In Eq. 1.78, $\frac{1}{1-q^2/m_+^2}$, with $m_+ = m_B + 0.046\,\mathrm{GeV}/c^2$, is the pole corresponding to the first resonance in the q^2 spectrum, and

$$z(t) = \frac{\sqrt{t_+ - t} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - t} + \sqrt{t_+ - t_0}}, \tag{1.79}$$

with $t_\pm = (m_B \pm m_K)^2$ and $t_0 = t_+(1 - \sqrt{1 - t_-/t_+})$, is a variable for which the series rapidly converges [85]. The series in Eq. 1.78 is usually truncated after the quadratic term in z. To characterise the form factor over the entire q^2 range of the decay, the series is fitted to Lattice QCD and Light Cone Sum Rules (LCSR) [90–92] and the parameters α_0, α_1, α_2 are estimated. The Lattice QCD approach is valid at high q^2, while the LCSR can be used in the low q^2 region of the kinematic range.
The most recent SM predictions are based on the fit performed in Ref. [93]. The predicted branching fractions for the charged and neutral modes are

$$\mathrm{BR}(B^+ \to K^+\nu\bar{\nu})_{\mathrm{SM}} = (4.6 \pm 0.5) \times 10^{-6}, \tag{1.80}$$

$$\mathrm{BR}(B^0 \to K^0\nu\bar{\nu})_{\mathrm{SM}} = (4.3 \pm 0.5) \times 10^{-6}. \tag{1.81}$$

The dominant contribution to the theoretical uncertainty is related to the error due to the $B \to K$ form factor. This error is more than twice as large as the error due to parametric uncertainties, which in turn are dominated by uncertainties on CKM matrix elements [85].

1.4.2 $B \to K^*\nu\bar{\nu}$ in the Standard Model

In the $B \to K^*\nu\bar{\nu}$ decays, the pseudo-scalar ($\mathrm{J^P} = 0^-$) $B^+(B^0)$ meson decays into the vector ($\mathrm{J^P} = 1^-$) $K^{*+}(K^{*0})$ meson and two neutrinos. In this case, due to the angular momentum conservation, the dynamics of the decay depends on the kaon polarisation, which defines the spin orientation of the K^*. In particular, the angular distribution of the K^* decay products is correlated with the K^* polarisation.
The observables that can be measured to provide a complete description of the decay are the squared invariant mass of the neutrino-antineutrino pair, q^2, and θ, the angle between the K^* flight direction in the B rest frame and the K flight direction in the $K\pi$ rest frame. Using these observables, the SM differential branching fraction of the decay can be written as done in Ref. [85]:

$$\frac{d^2\mathrm{BR}(B \to K^*\nu\bar{\nu})_{\mathrm{SM}}}{dq^2 d\cos\theta} = \frac{3}{4}\frac{d\mathrm{BR}_T}{dq^2}\sin^2\theta + \frac{3}{2}\frac{d\mathrm{BR}_L}{dq^2}\cos^2\theta, \tag{1.82}$$

where

$$\frac{d\mathrm{BR}_L}{dq^2} = \tau_B 3|N|^2 \frac{X_t^2}{s_W^4}\rho_{A_1}(q^2), \tag{1.83}$$

is the branching fraction for a longitudinally polarised K^*, and

$$\frac{d\mathrm{BR}_T}{dq^2} = \tau_B 3|N|^2 \frac{X_t^2}{s_W^4}[\rho_{A_{12}}(q^2) + \rho_V(q^2)], \tag{1.84}$$

is the branching fraction for a transversely polarised K^*. The equations 1.83, 1.84 depend on the rescaled form factors

$$\rho_V(q^2) = \frac{2q^2\lambda_{K^*}^{3/2}(q^2)}{(m_B + m_{K^*})^2 m_B^4} \left[V(q^2)\right]^2, \tag{1.85}$$

$$\rho_{A_1}(q^2) = \frac{2q^2\lambda_{K^*}^{1/2}(q^2)(m_B + m_{K^*})^2}{m_B^4} \left[A_1(q^2)\right]^2, \tag{1.86}$$

$$\rho_{A_{12}}(q^2) = \frac{64 m_{K^*}^2 \lambda_{K^*}^{1/2}(q^2)}{m_B^2} \left[A_{12}(q^2)\right]^2, \tag{1.87}$$

in which $V(q^2)$, $A_1(q^2)$, $A_{12}(q^2)$ are the actual SM form factors used to parametrise the q^2 dependence of the hadronic matrix element $\langle K^*|\bar{s}\gamma_\mu(1-\gamma_5)b|B\rangle$. As with the $B \to K$ form factor, $V(q^2)$, $A_1(q^2)$, $A_{12}(q^2)$ can individually be represented by rapidly converging series in the variable z, like the one given in Eq. 1.78. The series expansion is truncated after the quadratic term, and the parameters $(\alpha_0, \alpha_1, \alpha_2)_i$, with $i = V, A_1, A_{12}$, are estimated in fits of the form factors to Lattice QCD and LCSR results, performed over the kinematic range of the decay.
The partial branching fractions $d\mathrm{BR}_L/dq^2$ in Eq. 1.83 and $d\mathrm{BR}_T/dq^2$ in Eq. 1.84 can be measured in an angular analysis of the K^{*0} decay products. Otherwise, the decay can be completely characterised by measuring the total differential branching fraction

$$\begin{aligned} \frac{d\mathrm{BR}}{dq^2} &= \int_{-1}^{1} d\cos\theta \frac{d^2\mathrm{BR}}{dq^2 d\cos\theta} = \frac{d\mathrm{BR}_L}{dq^2} + \frac{d\mathrm{BR}_T}{dq^2} \\ &= \tau_B 3|N|^2 \frac{X_t^2}{s_W^4}[\rho_{A_1}(q^2) + \rho_{A_{12}}(q^2) + \rho_V(q^2)], \end{aligned} \tag{1.88}$$

and either the longitudinal or the transverse K^{*0} polarisation fractions

$$F_{L,T} = \frac{d\mathrm{BR}_{L,T}/dq^2}{d\mathrm{BR}/dq^2}, \qquad F_L = 1 - F_T. \tag{1.89}$$

In particular, $F_{L,T}$ represent very clean observables to probe the SM, because some of the uncertainties related to the form factors and to the CKM matrix elements factorise in the ratios.
The most recent SM predictions for the $B \to K^*\nu\bar{\nu}$ decays come from Ref. [93]. The predicted branching fractions for the charged and neutral modes correspond to

$$\mathrm{BR}(B^+ \to K^{*+}\nu\bar{\nu})_{\mathrm{SM}} = (8.4 \pm 1.5) \times 10^{-6}, \tag{1.90}$$

$$\mathrm{BR}(B^0 \to K^{*0}\nu\bar{\nu})_{\mathrm{SM}} = (7.8 \pm 1.4) \times 10^{-6}. \tag{1.91}$$

Also for these decays, the major contribution to the theoretical uncertainty is related to the error due to the $B \to K^*$ form factors. In this case, the form factor uncertainty is about 1.7 times larger than the combined parametric uncertainties [85].

1.5 $B \to K^{(*)}\nu\bar{\nu}$ beyond the Standard Model

Despite still representing the best theory of fundamental interactions, the SM fails to explain some deep unsolved problems. For example, it does not include gravity and it is not valid at energies up to the Planck scale $\sim 10^{19}$ GeV; it does not provide a solution to the hierarchy problem [94]; it does not describe dark matter and dark energy, nor it explains the matter–antimatter asymmetry in the universe [95]. In the light of such open puzzles, the SM is clearly not complete, and it should rather be considered as an effective theory successfully working at the energies probed so far.
In recent years, studies in the flavour sector have revealed potential breaches in the SM. In particular, measurements of decays involving $b \to s$ transitions have been characterised by persistent tensions with the

SM predictions. These tensions include the results of the lepton-flavour universality tests in $B \to K^{(*)}l^+l^-$ decays performed by the LHCb experiment [28–33]. Their latest measurement of the $R_k = \frac{\mathrm{BR}(B^+\to K^+\mu^+\mu^-)}{\mathrm{BR}(B^+\to K^+e^+e^-)}$ ratio reveals evidence for the breaking of lepton universality with a significance of 3.1 standard deviations [34]. This result might be a specific indication of the possible existence of physics beyond the SM.
The relation between $b \to sl^+l^-$ and $b \to s\nu\bar{\nu}$ processes, governed by the same form factors and connected by the $SU(2)_L$ gauge symmetry, implies that possible new physics in $B \to K^{(*)}l^+l^-$ must have an imprint also on $B \to K^{(*)}\nu\bar{\nu}$. Therefore, measurements of the $B \to K^{(*)}\nu\bar{\nu}$ decays offer a complementary probe of potential non-SM scenarios proposed to explain the $b \to sl^+l^-$ anomalies. Moreover, as discussed in Sec. 1.4, the advantage in the study of the $b \to s\nu\bar{\nu}$ transition consists in the much smaller theoretical uncertainty. Given the level of accuracy of the SM prediction for the $B \to K^{(*)}\nu\bar{\nu}$ decays, deviations in the observed decay rates can provide an unambiguous sign of new physics.
In the following, physics beyond the SM in the $b \to s\nu\bar{\nu}$ transition is first examined in a model-independent framework. In this context, contributions of specific new mediators are also evaluated based on the values of the Wilson Coefficients estimated in global fits to the experimental measurements [96]. The characterisation of $B \to K^{(*)}\nu\bar{\nu}$ decays with new sources of missing energy, such as dark matter, is also discussed.

1.5.1 Model-independent analysis of non-SM effects

In the simplest extension of the SM with a model-independent effective theory [85], lepton-flavour universality is preserved, and possible new physics, at least as heavy as the B mesons, is incorporated into the low-energy effective Hamiltonian of Eq. 1.71 by means of an operator O_R,

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \left(C_L O_L + C_R O_R + \mathrm{h.c.}\right), \tag{1.92}$$

where

$$O_R = \frac{e^2}{16\pi^2} \left[\bar{s}\gamma_\mu P_R b\right] \left[\bar{\nu}\gamma^\mu(1-\gamma_5)\nu\right] \tag{1.93}$$

accounts for new right-handed interactions that are absent in the SM. The Wilson Coefficients C_L and C_R in Eq. 1.92 can be used to define two real variables,

$$\epsilon = \frac{\sqrt{|C_L|^2 + |C_R|^2}}{|C_L^{SM}|}, \qquad \eta = \frac{-\mathrm{Re}(C_L C_R^*)}{|C_L|^2 + |C_R|^2}, \tag{1.94}$$

with $\epsilon > 0$ and $\eta \in [-1/2, 1/2]$ [85]. If only SM interactions are involved, then $\epsilon = 1$, $\eta = 0$. Conversely, $\eta \neq 0$ would indicate the presence of right-handed currents. In tests of the SM, the measurable observables of the $B \to K^{(*)}\nu\bar{\nu}$ decays, normalised to the SM predictions, can be expressed in terms of ϵ and η as

$$\mathcal{R}_K^\nu = \frac{\mathrm{BR}(B \to K\nu\bar{\nu})}{\mathrm{BR}(B \to K\nu\bar{\nu})_{\mathrm{SM}}} = (1-2\eta)\epsilon^2, \tag{1.95}$$

$$\mathcal{R}_{K^*}^\nu = \frac{\mathrm{BR}(B \to K^*\nu\bar{\nu})}{\mathrm{BR}(B \to K^*\nu\bar{\nu})_{\mathrm{SM}}} = (1+k_\eta\eta)\epsilon^2, \tag{1.96}$$

$$\mathcal{R}_{F_L} = \frac{F_L}{F_L^{\mathrm{SM}}} = \frac{1+2\eta}{1+k_\eta\eta}, \tag{1.97}$$

where k_η is a parameter depending on the form factors [85]. Ratios different from unity would indicate the presence of non-SM interactions, and the size of new physics contributions is captured by the parameters ϵ and η, which are related to the Wilson Coefficients.
In an extensive model-independent effective theory, the correlation between the $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ transitions can be taken into account by choosing an operator product expansion with dimension-six operators invariant under the full SM gauge symmetry [97, 98]. This effective theory is usually referred to as

SM Effective Field Theory (SM-EFT), since only new physics arising at an energy scale Λ, larger than the electroweak scale, is integrated out. The corresponding dimension-six Lagrangian can be written as

$$\mathcal{L}^{(6)} = \sum_i \frac{c_i}{\Lambda^2} Q_i. \tag{1.98}$$

The dimension-six operators Q_i, defined in Refs. [97] [98], which contribute to the $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ transitions are

$$Q_{\Phi q}^{(1)} = i(\bar{\mathbf{q}}_L \gamma_\mu \mathbf{q}_L)\mathbf{\Phi}^\dagger \mathcal{D}^\mu \mathbf{\Phi}, \qquad Q_{ql}^{(1)} = (\bar{\mathbf{q}}_L \gamma_\mu \mathbf{q}_L)(\bar{\mathbf{l}}_L \gamma^\mu \mathbf{l}_L), \tag{1.99}$$

$$Q_{\Phi q}^{(3)} = i(\bar{\mathbf{q}}_L \gamma_\mu \sigma^a \mathbf{q}_L)\mathbf{\Phi}^\dagger \mathcal{D}^\mu \sigma_a \mathbf{\Phi}, \qquad Q_{ql}^{(3)} = (\bar{\mathbf{q}}_L \gamma_\mu \sigma^a \mathbf{q}_L)(\bar{\mathbf{l}}_L \gamma^\mu \sigma_a \mathbf{l}_L), \tag{1.100}$$

$$Q_{\Phi d} = i(\bar{\mathbf{d}}_R \gamma_\mu \mathbf{d}_R)\mathbf{\Phi}^\dagger \mathcal{D}^\mu \mathbf{\Phi}, \qquad Q_{dl} = (\bar{\mathbf{d}}_L \gamma_\mu \mathbf{d}_L)(\bar{\mathbf{l}}_L \gamma^\mu \mathbf{l}_L). \tag{1.101}$$

At low energies, the Wilson Coefficients c_i coupled to the dimension-six operators in Eq. 1.98 can be mapped into the coefficients [85]

$$C_L^{\rm NP} = C_L^{\rm SM} + \tilde{c}_{ql}^{(1)} - \tilde{c}_{ql}^{(3)} + \tilde{c}_Z, \qquad C_R = \tilde{c}_{dl} + \tilde{c}'_Z, \tag{1.102}$$

$$C_9^{\rm NP} = C_9^{\rm SM} + \tilde{c}_{qe} + \tilde{c}_{ql}^{(1)} + \tilde{c}_{ql}^{3} - \xi\tilde{c}_Z, \qquad C'_9 = \tilde{c}_{de} + \tilde{c}_{dl} - \xi\tilde{c}'_Z \tag{1.103}$$

$$C_{10}^{\rm NP} = C_{10}^{\rm SM} + \tilde{c}_{qe} - \tilde{c}_{ql}^{(1)} - \tilde{c}_{ql}^{(3)} + \tilde{c}_Z, \qquad C'_{10} = \tilde{c}_{de} - \tilde{c}_{dl} + \tilde{c}'_Z, \tag{1.104}$$

of the effective Hamiltonian

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \frac{e^2}{16\pi^2} \sum_i C_i O_i, \tag{1.105}$$

which includes the operators $O_{L,R}$ defined in Eqs. 1.72, 1.93 and also the operators

$$O_9^{(\prime)} = (\bar{s}\gamma_\mu P_{L(R)} b)(\bar{\mathrm{l}}\gamma^\mu \mathrm{l}), \qquad O_{10}^{(\prime)} = (\bar{s}\gamma_\mu P_{L(R)} b)(\bar{\mathrm{l}}\gamma^\mu \gamma_5 \mathrm{l}) \tag{1.106}$$

relevant for $b \to sl^+l^-$ transitions [85].
The possible contribution of physics beyond the SM is captured by the Wilson Coefficients in Eqs.1.102, 1.103, 1.104, which are estimated in global fits to the observables measured in $b \to sl^+l^-$ and $b \to s\nu\bar{\nu}$ processes, such as branching fractions, angular distributions and polarisation fractions.
The recent analysis in Ref. [99] studies the implication of the $b \to sl^+l^-$ anomalies on the $B \to K^{(*)}\nu\bar{\nu}$ decays within the SM-EFT framework. A global fit to the $b \to sl^+l^-$ data is performed, and the effect of the fit results on the $B \to K^{(*)}\nu\bar{\nu}$ decays is evaluated in various NP scenarios. Each scenario is characterised by the best fit values of the SM-EFT coefficients. In particular, four scenarios show good compatibility with the $b \to sl^+l^-$ data. However, all of them produce deviations at more than 3σ significance from the SM prediction for the $B \to K^{(*)}\nu\bar{\nu}$ decays. The results in these four scenarios reported in Ref. [99] are listed in Table 1.2.

SM-EFT couplings	$\mathrm{BR}(B \to K\nu\bar{\nu}) \times 10^{-6}$	$\mathrm{BR}(B \to K^*\nu\bar{\nu}) \times 10^{-6}$
SM	4.006 ± 0.261	9.331 ± 0.744
$(\tilde{c}_{ql}^{(3)}, \tilde{c}'_Z)$	5.485 ± 0.358	3.197 ± 0.223
$(\tilde{c}_Z, \tilde{c}'_Z)$	3.260 ± 0.213	2.141 ± 0.146
$(\tilde{c}_{ql}^{(1)} + \tilde{c}_{ql}^{(}3), \tilde{c}_Z)$	7.419 ± 0.484	17.284 ± 1.378
$(\tilde{c}_{ql}^{(1)} + \tilde{c}_{ql}^{(3)}, \tilde{c}'_Z)$	2.319 ± 0.151	12.857 ± 1.111

Table 1.2: Branching fractions of the $B \to K\nu\bar{\nu}$ and $B \to K^*\nu\bar{\nu}$ decays in the SM and in four model-independent new-physics scenarios showing good compatibility with the $b \to sl^+l^-$ anomalies. Each model is characterised by the best fit values of the SM-EFT coefficients estimated by the global fit in Ref. [99].

1.5.2 Models with new mediators

Specific new-physics models predict new mediators and modified couplings in $b \to s\nu\bar{\nu}$ transitions. Such models can be characterised in the framework of effective field theories. In the following, the discussion is centred around new mediators that can be described by four-fermion operators. This is the case of the Z' gauge boson [100–102] and of leptoquarks [103–105].

1.5.2.1 Generic Z'

A possible extension of the SM is based on the addition of an extra $U(1)$ symmetry to the SM gauge group. The corresponding gauge vector-boson is the Z' boson, which transforms as a singlet under $SU(2)_L$. Recent direct searches are constraining the mass of the Z' to be $M_{Z'} > 1.5\,\text{TeV}/c^2$ [106]. This heavy gauge boson might mediate FCNC transitions in the quark sector at the tree level, as illustrated in Fig. 1.9, resulting in deviations from the SM expectations depending only on the couplings of the Z' with fermions and on its mass.

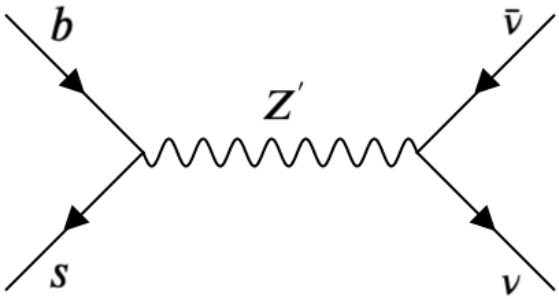


Figure 1.9: Feynman diagram for a tree level $b \to s\nu\bar{\nu}$ transition mediated by a Z' boson.

The Lagrangian describing the interactions of the Z' with the SM fermions $\mathbf{f}_i$ can be written as

$$\mathcal{L}_{Z'} = \bar{\mathbf{f}}_i \gamma^\mu \left[\Delta_L^{ij}(Z')P_L + \Delta_R^{ij}(Z')P_R\right] \mathbf{f}_j Z'_\mu, \tag{1.107}$$

where the sum over the generation indices is implied and $\Delta_L^{\nu\bar{\nu}}(Z') = \Delta_L^{ll}(Z')$ because of the $SU(2)_L$ symmetry [85]. The Wilson Coefficients

$$\tilde{c}_{ql}^{(1)} = -\frac{\Delta_L^{sb}\Delta_L^{ll}}{V_{tb}V_{ts}^*}\left[\frac{5\,\text{TeV}}{M_{Z'}}\right]^2, \qquad \tilde{c}_{dl} = -\frac{\Delta_R^{sb}\Delta_L^{ll}}{V_{tb}V_{ts}^*}\left[\frac{5\,\text{TeV}}{M_{Z'}}\right]^2, \tag{1.108}$$

$$\tilde{c}_{qe} = -\frac{\Delta_L^{sb}\Delta_R^{ll}}{V_{tb}V_{ts}^*}\left[\frac{5\,\text{TeV}}{M_{Z'}}\right]^2, \qquad \tilde{c}_{de} = -\frac{\Delta_R^{sb}\Delta_R^{ll}}{V_{tb}V_{ts}^*}\left[\frac{5\,\text{TeV}}{M_{Z'}}\right]^2, \tag{1.109}$$

are generated in the SM-EFT to account for tree-level contributions of a Z' boson to $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ transitions [85]. At low energies, a scenario with only left-handed couplings, Δ_L^{sb} and $\Delta_L^{ll}(= \Delta_L^{\nu\bar{\nu}})$, would result in

$$C_9^{\text{NP}} = -C_{10}^{\text{NP}} = \tilde{c}_{ql}^{(1)}. \tag{1.110}$$

If the right-handed quark coupling Δ_R^{sb} is also turned on, the additional low-energy Wilson Coefficient

$$C'_9 = -C'_{10} = \tilde{c}_{dl} \tag{1.111}$$

is generated [85].
The analysis in Ref. [107], which includes the result of the $B^+ \to K^+\nu\bar{\nu}$ measurement in this thesis and the latest LHCb lepton-universality test [34], shows that a good fit to the $b \to sl^+l^-$ anomalies can be

obtained via $C_9^{\rm NP} = -C_{10}^{\rm NP} = -0.41^{+0.07}_{-0.07}$ [108] for $\Delta_L^{sb} = (8.5 \pm 6.4) \times 10^{-3}$ and $\Delta_L^{\mu\mu} = 2.00 \pm 0.95$. The constrained parameter space is compatible with the current $R_K^\nu = 2.4 \pm 0.9$ [107], but it can give at most $R_K^\nu = 1.05 \pm 0.03$, which is too close to the SM to be probed with the data that will be available in the next few years. In the same analysis, it is found that the scenario in which the right-handed coupling Δ_R^{sb} is also turned on accommodates the global fit to all $b \to sl^+l^-$ data [109] better, but it produces an effect on $b \to s\nu\bar{\nu}$ of the same size as the one obtained in the previous scenario [107].

1.5.2.2 Leptoquarks

Leptoquarks represent another class of new mediators in extensions of the SM. The Pati-Salam model [103], theorised in 1974, predicts their existence, and they also appear in grand unification theories, like the $SU(5)$ gauge theory by Georgi and Glashow [104]. They are hypothetical bosons, carrying both baryon number and lepton number, which can have either spin 1 (vector leptoquark) or spin 0 (scalar leptoquark). The predicted mass of leptoquarks is $\mathcal{O}(\mathrm{TeV})$, and stringent constraints on it have been recently set in direct searches at the LHC [110].
Leptoquarks can couple to a quark and a lepton at the same time, thus leading to interactions in which the four-fermion operators relevant for $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ processes can be generated by the exchange of a leptoquark at the tree level, or by leptoquark interactions occurring at higher orders in perturbation theory [111–113]. A Feynman diagram illustrating a tree-level $b \to s\nu\bar{\nu}$ transition mediated by a generic leptoquark is shown in Figure 1.10. Incorporating also a SM gauge singlet with right-handed chirality, ν_R, for hypothetical right-handed neutrinos, the allowed leptoquark couplings in $b \to s\nu\bar{\nu}$ transitions at the tree level are listed in Table 1.3. Here the notation $f^c \equiv \mathcal{C}\bar{f}$ indicates the charge-conjugated field of the fermion f.

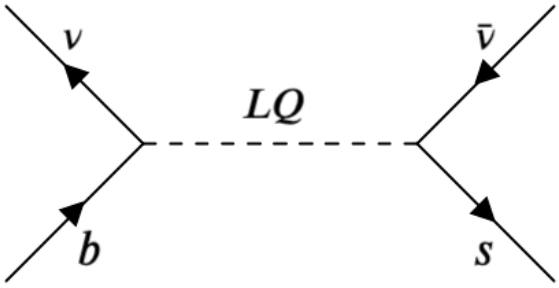


Figure 1.10: Feynman diagram for a tree level $b \to s\nu\bar{\nu}$ transition mediated by a generic leptoquark.

In the analysis of Ref. [107], the triplet scalar leptoquark $S_3(\bar{3}, 3, 1/3)$ is found to be a good candidate mediator, given the combined contribution at the tree level to the $b \to sl^+l^-$ and $b \to s\nu\bar{\nu}$ transitions. In this case, the leptoquark interaction term has the form

$$+ \bar{\mathbf{q}}_L^c Y_{S_3} i\sigma_2 \boldsymbol{\sigma} \cdot \mathbf{S}_3 \mathbf{l}_L, \tag{1.112}$$

where Y_{S_3} is a Yukawa matrix in the flavour space. The corresponding low-energy Wilson Coefficients are

$$C_L^{\rm NP} = \tilde{c}_{ql}^{(1)} - \tilde{c}_{ql}^{(3)}, \tag{1.113}$$

$$C_9^{\rm NP} = -C_{10}^{\rm NP} = \tilde{c}_{ql}^{(1)} + \tilde{c}_{ql}^{(3)}, \tag{1.114}$$

where

$$\frac{[\tilde{c}_{ql}^{(1)}]^{ij\alpha\beta}}{\Lambda^2} = 3\frac{[\tilde{c}_{ql}^{(3)}]^{ij\alpha\beta}}{\Lambda^2} = \frac{3Y_{S_3}^{j\beta}Y_{S_3}^{*i\alpha}}{4M_{S_3}^2}. \tag{1.115}$$

are the coefficients generated in the SM-EFT, in terms of the leptoquark couplings and mass M_{S_3} [107].
The analysis in Ref. [107] finds that a S_3 leptoquark with mass $M_{S_3} = 2\,\mathrm{TeV}/c^2$ could explain the $b \to sl^+l^-$

Leptoquark	Spin	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$	Allowed coupling
S_3	0	$\bar{3}$	3	1/3	$\bar{\mathbf{q}}_L^c \mathbf{l}_L$
$\tilde{R}_2$	0	3	2	1/6	$\bar{\mathbf{d}}_R \mathbf{l}_L,\ \bar{\mathbf{q}}_L \nu_R$
S_1	0	$\bar{3}$	1	1/3	$\bar{\mathbf{q}}_L^c \mathbf{l}_L, \bar{\mathbf{d}}_R^c \nu_R$
U_3^μ	1	3	3	2/3	$\bar{\mathbf{q}}_L \mathbf{l}_L$
V_2^μ	1	$\bar{3}$	2	5/6	$\bar{\mathbf{d}}_R^c \mathbf{l}_L,\ \bar{\mathbf{q}}_L^c \mathbf{e}_R$
$\bar{U}_1^\mu$	1	3	1	−1/3	$\bar{\mathbf{d}}_R \nu_r$

Table 1.3: Quantum numbers and allowed couplings for scalar and vector leptoquarks mediating $b \to s\nu\bar{\nu}$ transitions at the tree level involving SM neutrinos as well as the gauge singlet with right-handed chirality ν_R.

tensions, and at the same time accommodate the current R_K^ν central value, by turning on the couplings $Y_{S_3}^{22}$, $Y_{S_3}^{23}$, $Y_{S_3}^{32}$, $Y_{S_3}^{33}$. The best fit values found in the analysis are

$$Y_{S_3}^{22} = 0.0025 \pm 0.0024, \quad Y_{S_3}^{23} = 1.0 \pm 0.83, \quad Y_{S_3}^{32} = 0.109 \pm 0.085, \quad Y_{S_3}^{33} = -0.041 \pm 0.024. \tag{1.116}$$

They produce $R_K^\nu = 2.4 \pm 3.6$, which is consistent with the current experimental value $R_K^\nu = 2.4 \pm 0.9$ including also the measurement of this thesis [107].
In the same analysis, another good candidate is identified with the vector leptoquark $U_3^\mu(3,3,2/3)$. In this case, the new interaction term is

$$+\,\bar{\mathbf{q}}_L \gamma^\mu \sigma^a Y_{U_3} \mathbf{l}_L U_{3\mu}^a. \tag{1.117}$$

Given that U_3^μ is a $SU(2)_L$ triplet like the scalar leptoquark S_3, the relations between the generated SM-EFT Wilson Coefficients

$$\frac{[\tilde{c}_{ql}^{(1)}]^{ij\alpha\beta}}{\Lambda^2} = -3\frac{[\tilde{c}_{ql}^{(3)}]^{ij\alpha\beta}}{\Lambda^2} = -\frac{3Y_{U_3}^{j\beta} Y_{U_3}^{*i\alpha}}{2M_{U_3}^2} \tag{1.118}$$

and the low-energy Wilson Coefficients are the same as in Eqs. 1.113, 1.114. The results in Ref. [107] show that a U_3^μ leptoquark of mass $M_{U_3} = 2\,\mathrm{TeV}/c^2$ can significantly enhance the $B^+ \to K^+\nu\bar{\nu}$ decay rate and also explain the $b \to s l^+ l^-$ anomalies within their 1σ uncertainties.

1.5.3 Dark matter as a source of missing energy

The $B \to K^{(*)}\nu\bar{\nu}$ decays have the same experimental signature of generic $B \to K^{(*)}\not{E}$ processes, where the final state is characterised by the presence of a kaon and missing energy $\not{E}$. A specific source of missing energy could be dark matter. Dark matter is a hypothetical form of matter which is supposed to represent approximately 85% of the matter in the Universe [43]. Its presence is inferred as a consequence of a variety of astrophysical observations and gravitational effects. Dark matter barely interacts with ordinary baryonic matter and radiation, except through gravity. This implies that possible dark-matter particles would escape the Belle II detector without leaving any experimental signature.
Many are the possible dark matter candidates that can be involved in $B \to K^{(*)}\not{E}$ decays: heavy QCD axions [114], weakly interacting massive particles (WIMPs) [115], sterile massive neutrinos [116], light scalars [117, 118]. The last scenario is illustrated in detail in the next paragraph since dedicated searches for light scalars in $b \to s$ transitions are currently ongoing at Belle II [119].

1.5.3.1 Dark scalars in $B \to K^{(*)}\not{E}$ decays

The study in Ref. [117] explores the resonant production of a dark scalar in $B \to K^{(*)}\not{E}$ decays. This scalar, with mass in the GeV range, can decay into a pair of dark fermions, producing experimental signatures with

missing energy. Such processes are theorised in a simple extension of the SM, obtained by introducing a real scalar field ϕ and a Dirac fermion χ, which are both singlets with respect to the strong and the electroweak interactions. The fermion does not mix with the SM neutrinos and is a stable dark matter candidate. The scalar is the mediator of a new interaction between the Higgs field and the dark fermion χ. In this model, after the electroweak symmetry breaking, the scalar mixes with the neutral component of the SM Higgs doublet producing a dark scalar S and the observed Higgs boson. S can decay in SM particles and dark fermions with a total decay width

$$\Gamma_S = s_\theta^2\,\Gamma_{\rm SM} + c_\theta^2\,\Gamma_{\chi\bar{\chi}}, \tag{1.119}$$

where s_θ and c_θ are the sine and cosine of a mixing angle θ. The branching fraction into SM leptons and dark fermions are [117]

$$\mathrm{BR}(S \to l\bar{l}) = \frac{s_\theta^2\Gamma_{l\bar{l}}}{\Gamma_S} = \frac{m_l^2 s_\theta^2}{8\pi v^2}\frac{m_S}{\Gamma_S}\left(1-\frac{4m_l^2}{m_S^2}\right)^{3/2}, \tag{1.120}$$

$$\mathrm{BR}(S \to \chi\bar{\chi}) = \frac{c_\theta^2\Gamma_{\chi\bar{\chi}}}{\Gamma_S} = \frac{y_\chi^2 c_\theta^2}{8\pi}\frac{m_S}{\Gamma_S}\left(1-\frac{4m_\chi^2}{m_S^2}\right)^{3/2}, \tag{1.121}$$

where y_χ is the Yukawa coupling constant of the dark fermion. Invisible decays become dominant for $m_S > 2m_\chi$ and $y_\chi c_\theta > m_l s_\theta / v$.
The scalar couplings with SM quarks are flavour diagonal, hence $b \to sS$ transitions can occur only at the loop level, through the large top-quark coupling, as illustrated by the Feynman diagram in Fig. 1.11. The

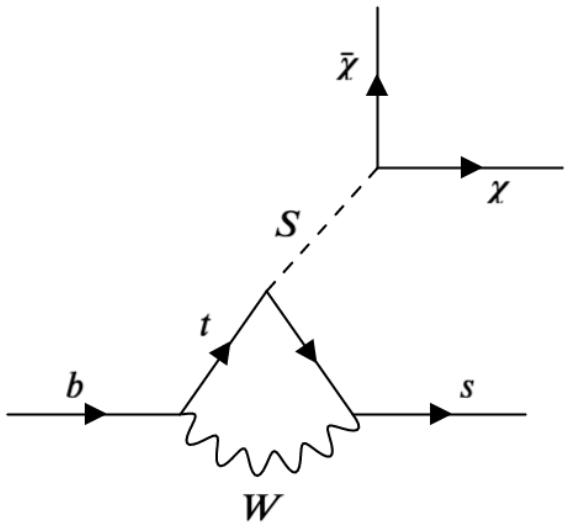


Figure 1.11: Feynman diagram of a $b \to sS$ transition where the dark scalar S decays into a pair of dark fermions $\chi\bar{\chi}$.

branching fraction of the $B^+ \to K^+S$ decay is [117]

$$\begin{aligned}\mathrm{BR}(B^+ \to K^+S) = {} & \frac{\sqrt{2}G_F|C_{bs}|^2}{64\pi\Gamma_B m_B^3}\frac{(m_b+m_s)^2}{(m_b-m_s)^2}f_0^2(m_S^2) \\ & \times (m_B^2-m_K^2)\left[(m_B^2-m_K^2-m_S^2)^2-4m_K^2m_S^2\right]^{1/2},\end{aligned} \tag{1.122}$$

where Γ_B is the total decay width of the B^+ meson, $f_0(m_S^2)$ is the scalar hadronic form factor computed at $q^2 = m_S^2$, and C_{bs} is the Wilson Coefficient coupled with the interaction term for the dark scalar. The branching fractions of the other $B \to K^{(*)}S$ decays can be found in Refs. [120, 121].

The scalar S decays into the pair of dark fermions $\chi\bar{\chi}$ through a narrow resonance, and the observable branching fraction of the B decay into the $K^+\chi\bar{\chi}$ final state is given by [117]

$$\mathrm{BR}(B^+ \to K^+\chi\bar{\chi}) = \mathrm{BR}(B^+ \to K^+S)\,\mathrm{BR}(S \to \chi\bar{\chi}) \propto s_\theta^2 \frac{c_\theta^2 \Gamma_{\chi\bar{\chi}}}{\Gamma_S}. \tag{1.123}$$

For $\mathrm{BR}(S \to \chi\bar{\chi}) \approx 1$ the scalar decays invisibly and creates signatures with missing energy. Experimentally the final state is the same as in $B \to K^{(*)}\nu\bar{\nu}$ decays with SM neutrinos. However, in model-dependent searches, like the analyses in this thesis, which rely on the three-body kinematics of the SM process, the results cannot be directly reinterpreted for the two-body decays $B \to K^{(*)}S(\to \chi\bar{\chi})$, in which the q^2 distribution is very peculiar, as it sharply peaks around the scalar resonance $q^2 = m_S^2$. To optimise the sensitivity to $B \to K^{(*)}\not{E}$ events with dark scalar resonances and distinguish them from $B \to K^{(*)}\nu\bar{\nu}$ events a dedicated search for two-body $B \to K^{(*)}S(\to \chi\bar{\chi})$ decays is needed. In particular, the analysis in Ref. [117] predicts that, with an integrated luminosity of $50\,\mathrm{ab}^{-1}$, Belle II can probe dark scalars with mixing angles down to $\theta \approx 10^{-3}$ and masses $m_S > m_B - m_K$.

1.6 Experimental status of $B \to K^{(*)}\nu\bar{\nu}$ searches

Precise measurements of the $B \to K^{(*)}\nu\bar{\nu}$ decays are experimentally challenging. The major limitations are related to the presence of two neutrinos in the final state, which escape the detector without leaving any measurable signature that can be used to characterise the decay of the signal B meson. Searches for the $B \to K^{(*)}\nu\bar{\nu}$ decays can only be performed by experiments at B factories, exploiting the well-defined initial state and the clean event environment. An illustration of the experimental features characterising a $B^+ \to K^+\nu\bar{\nu}$ event in the Belle II detector is shown in the simulated event display in Fig. 1.12.
Searches for the $B \to K^{(*)}\nu\bar{\nu}$ decays were performed in the past by the CLEO, Belle and BaBar experiments. In all the analyses reported to date, no statistically significant signal was observed, and upper limits on the branching fractions were set at 90% confidence level (CL). A summary of the observed upper limits in searches performed over the last decade is provided in Table 1.4. The measurements all adopted *tagged* approaches, where the second B meson, also called *tag* B meson, produced in the $e^+e^- \to \Upsilon(4\mathrm{S}) \to B\bar{B}$ event, is fully reconstructed in a hadronic decay ($B \to \mathrm{hadrons}$, *hadronic tagging* [122–124]) or in a semileptonic decay ($B \to Dl\nu$, *semileptonic tagging* [125, 126]). Using the hadronic tagging, the kinematics of the tag B decay is fully reconstructed, allowing to precisely infer the kinematics of the signal-B decay. The main disadvantage is the low tagging efficiency, which is $\mathcal{O}(0.1\%)$. The semileptonic tagging offers larger tagging efficiency, $\mathcal{O}(1\%)$. The drawback is an incomplete reconstruction of the tag-side kinematics, resulting in less precise information on the signal-B decay.

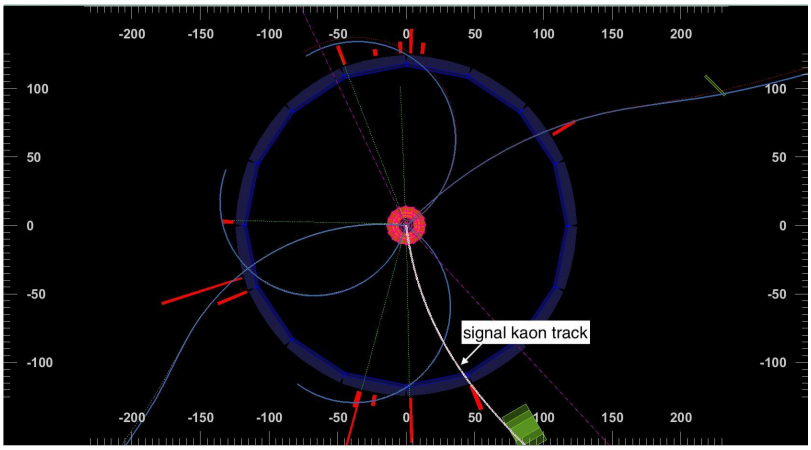


Figure 1.12: Belle II event display of a simulated $\Upsilon(4S) \to B^+B^-$ event with a $B^+ \to K^+\nu\bar{\nu}$ decay.

Decay	Experiment	Year	Observed limit	Tagging	Data $[\text{fb}^{-1}]$
$B^+ \to K^+\nu\bar{\nu}$	BaBar	2013	$< 1.6 \times 10^{-5}$ [124]	semileptonic + hadronic	429
	Belle	2013	$< 5.5 \times 10^{-5}$ [123]	hadronic	711
	Belle	2017	$< 1.9 \times 10^{-5}$ [126]	semileptonic	711
	Belle II	2021	$< 4.1 \times 10^{-5}$ [1]	inclusive	63
$B^0 \to K^0\nu\bar{\nu}$	BaBar	2013	$< 4.9 \times 10^{-5}$ [124]	semileptonic + hadronic	429
	Belle	2013	$< 9.7 \times 10^{-5}$ [123]	hadronic	711
	Belle	2017	$< 1.3 \times 10^{-5}$ [126]	semileptonic	711
$B^+ \to K^{*+}\nu\bar{\nu}$	BaBar	2013	$< 6.4 \times 10^{-5}$ [124]	semileptonic + hadronic	429
	Belle	2013	$< 4.0 \times 10^{-5}$ [123]	hadronic	711
	Belle	2017	$< 6.1 \times 10^{-5}$ [126]	semileptonic	711
$B^0 \to K^{*0}\nu\bar{\nu}$	BaBar	2013	$< 12 \times 10^{-5}$ [124]	semileptonic + hadronic	429
	Belle	2013	$< 5.5 \times 10^{-5}$ [123]	hadronic	711
	Belle	2017	$< 1.8 \times 10^{-5}$ [126]	semileptonic	711

Table 1.4: Observed upper limits at a 90% CL on the $B \to K^{(*)}\nu\bar{\nu}$ branching fractions in searches performed over the last decade.

The current B-tagging algorithms have been developed by BaBar, Belle, and Belle II. BaBar implemented the Semi-Exclusive B Reconstruction (SER) algorithm, in which a D and a D^* meson are first reconstructed and then combined with up to five charmless hadrons (in the hadronic tagging) or with a lepton (in the semileptonic tagging) to form a tag B meson candidate [27]. The Belle experiment developed the Full Reconstruction (FR) algorithm, which employs a hierarchical approach starting from reconstructed tracks and clusters [127]. These objects are used to identify final-state particles $(e^+, \mu^+, K^+, \pi^+, K_L^0, \gamma)$, which are combined in six distinct stages to form intermediate particles $(J/\psi, \pi^0, K_S^0, D, D^*)$ until a tag B meson candidate is reconstructed. Multiple combinations are tried and the best are selected by means of multivariate classifiers trained on simulated $\Upsilon(4\text{S})$ events. The Full Event Interpretation (FEI) algorithm implemented at Belle II is similar to the FR [128]. The FEI extends the list of B decays that can be tagged and uses an improved multivariate classifier. Searches for the $B \to K^{(*)}\nu\bar{\nu}$ decays at Belle II with the FEI have not yet been published but are in progress.

To overcome the limitation of tagged analyses due to low efficiency, the measurements presented in this thesis are performed without any explicit B tagging. A novel *inclusive tagging* technique is implemented, which exploits the characteristic features of the signal decay and the inclusive properties of the tag B meson to identify signal events. The inclusive tagging provides higher signal efficiency compared to the previous tagged searches and allows to obtain competitive results even with small data samples.

Chapter 2

The Belle II experiment

This chapter gives a broad overview of the Belle II experiment. Section 2.1 introduces the physics motivations of Belle II. The concept of B-factory is presented in Section 2.2, and Section 2.3 describes the SuperKEKB collider. Section 2.4 provides an extensive description of the Belle II detector, with individual paragraphs dedicated to each subdetector and system. The Belle II analysis software framework is discussed in Section 2.6. The main features of the Monte Carlo simulation in Belle II are explained in Section 2.7. Finally, Section 2.8 describes in detail the relevant aspects of the reconstruction procedure.

2.1 Introduction

The Belle II experiment at the SuperKEKB electron-positron collider is a state-of-the-art flavour-factory experiment, superseding the previous generation of B-factory experiments, BaBar and Belle [129].
The primary goal of Belle II is the search for physics beyond the SM in the flavour sector at the *intensity frontier*. Here large statistics is exploited to find signatures of new particles or processes by measuring rare or forbidden flavour reactions, or observing deviations from the SM predictions. An approach towards new physics that is complementary to the so-called *energy frontier* pursued at the LHC, which relies on the direct production of possible new particles in proton-proton collisions at center-of-mass energies up to $14\,\mathrm{TeV}$.
The Belle II physics program includes also the improvement of existing precision measurements of SM parameters, some of which are not competitively accessible by experiments at the LHC.
To accomplish such tasks, Belle II plans to collect a data sample equivalent to more than 30 times the combined integrated luminosities of Belle and BaBar.

2.2 B-factories

The B-factories are colliders designed in the 1990s to observe CP-asymmetries in decays of B mesons [27]. The measurement of time-dependent CP-violating asymmetries, related to the interplay of decay and mixing of B mesons, imposes stringent requirements on the design of colliders and detectors. Exclusive final states originating from B decays with suppressed branching fractions, like $B^0 \to J/\psi K^0_S$, the "golden mode" to look for CP violation, have to be reconstructed. Therefore, it is necessary to produce large amounts of B mesons and ensure large trigger and reconstruction efficiencies, keeping contamination from backgrounds low. Clean and abundant samples of B mesons can be produced in electron-positron collisions at a center-of-mass energy $\sqrt{s} = 10.58\,\mathrm{GeV}$, corresponding to the $\Upsilon(4\mathrm{S})$ resonance. This state is a vector meson made of a $b\bar{b}$ quark pair, and it decays roughly in the same proportion into neutral and charged $B\bar{B}$ mesons [43]. Electron-positron colliders can be designed to produce more than one million B mesons per day at the $\Upsilon(4\mathrm{S})$ resonance, reaching instantaneous luminosities larger than $10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Hence, they are called B-factories. The production of B mesons at B-factories offers multiple advantages. Some of them are presented in the list that follows.

- The $e^+e^- \to \Upsilon(4S)$ cross section is sizeable, $\sigma_{\Upsilon(4S)} \approx 1.1\,\text{nb}$ [27], compared to the total production cross section, $\sigma_{q\bar{q}}$, for continuum light-quark pair production: $\sigma_{\Upsilon(4S)}/\sigma_{q\bar{q}} \approx 30\%$. The comparison is graphically illustrated in Fig. 2.1.
- $\Upsilon(4S) \to B\bar{B}$ events are characterised by smaller mean charged-particle multiplicity than events with B-decays at hadronic colliders: ~ 11 charged tracks per event are produced on average, and the tracks are spread over the full solid angle [4]. Low track multiplicity and detector occupancy ensure high B-reconstruction efficiency with low trigger bias.
- The low-background environment allows to reconstruct final states containing photons from decays of π^0, $\rho^\pm$, η, η'.
- At B-factories the initial state and the kinematics of the $\Upsilon(4S)$ decay are precisely known, up to initial-state radiation, hence studies of B-meson decays with missing particles can exploit the conservation of energy and momentum.
- Off-resonance data samples can be collected at a center-of-mass energy $60\,\text{MeV}$ below the $\Upsilon(4S)$ resonance, allowing to study the continuum background from light-quark pair production.

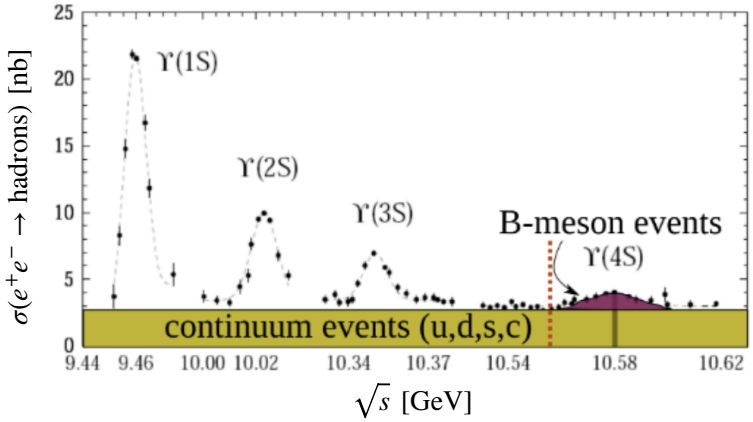


Figure 2.1: $e^+e^- \to \text{hadrons}$ cross section as a function of the center-of-mass energy. The red dashed line marks the kinematic threshold for the production of $B\bar{B}$ pairs.

Another characteristic feature of B-factories is the energy asymmetry between the electron and positron beams. Given that the B mass is $5.279\,\text{GeV}/c^2$ [43], B mesons are produced almost at rest in the center-of-mass system of the incoming beams, and the average spatial separation between the decay vertices of the two B mesons (a necessary quantity for measurements of time-dependent CP-violation) is only about $\sim 60\,\mu m$, which is hardly measurable by vertex detectors. To solve this issue, asymmetric beam energies were adopted so that the $\Upsilon(4S)$ meson is Lorentz-boosted in the laboratory frame along the direction of the high-energy beam.
It must also be mentioned that the data sets that can be collected at B-factories are not made only of B mesons. As illustrated by the chart in Fig. 2.2, at $\sqrt{s} = 10.58\,\text{GeV}$ the production cross sections of $e^+e^- \to \tau^+\tau^-$ and $e^+e^- \to c\bar{c}$ events are comparable to $e^+e^- \to \Upsilon(4S)$. Hence, B-factories are also τ-charm factories allowing to study with high precision the charm and τ production and decays.

2.3 The SuperKEKB collider

The SuperKEKB collider is the upgrade of KEKB [131], which operated from 1998 until June 2010, at the KEK laboratories in Tsukuba, Japan. The design luminosity is $8 \times 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ [132], 40 times higher than KEKB.

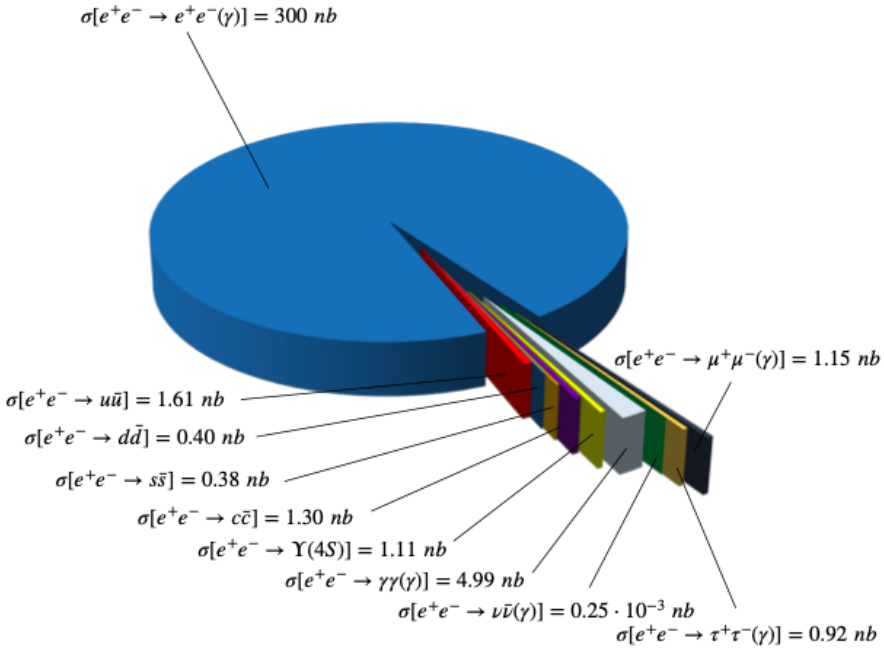


Figure 2.2: Major production cross sections in e^+e^- collisions at $\sqrt{s} = 10.58\,\mathrm{GeV}$. Credit: Ref. [130]

Such increased luminosity is obtained by significantly decreasing the beam sizes at the interaction point (IP), maximising the crossing angle, as prescribed in the nanobeam collision scheme originally theorised by P. Raimondi [133]. A comparison of the crossing angle and beam size at the IP between KEKB and SuperKEKB is shown in Fig. 2.3.

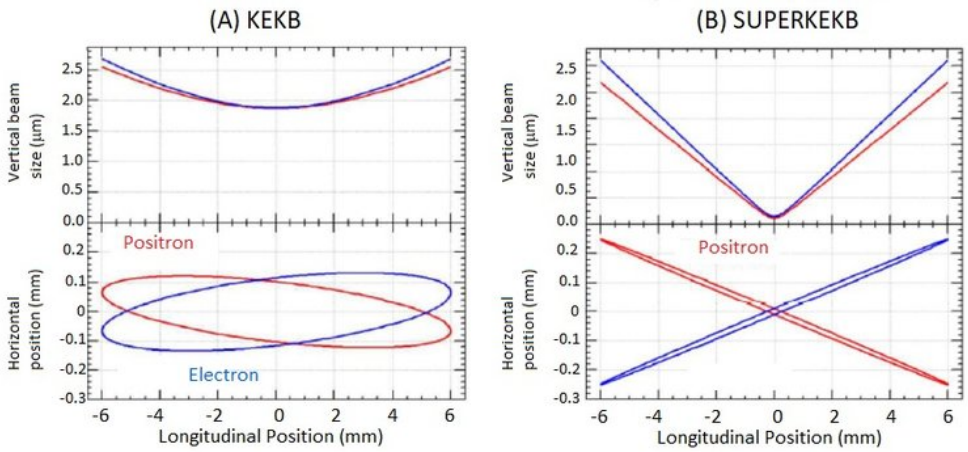


Figure 2.3: The vertical beam size (top) and overlap in the xz plane (bottom) at KEKB (left) and SuperKEKB (right). Credit: Ref. [134].

The double-ring circular structure of SuperKEKB is illustrated in Fig. 2.4, where the principal components of the accelerator system are indicated. A photo-cathode radio-frequency gun is used to produce electrons, which are accelerated to 7 GeV by a linear accelerator (linac) and then injected into the high-energy ring (HER). Positrons are generated by the primary electron beam impinging on a tungsten target coupled with a flux concentrator, a device providing an external magnetic field after the target to increase the positron yield.

The positron beam is accelerated up to $1.1\,\mathrm{GeV}$ by the linear accelerator and injected into the damping ring to reduce the beam emittance. The positrons are then accelerated by the linac up to $4\,\mathrm{GeV}$ and injected into the low-energy ring (LER). The Lorentz boost due to the beam-energy asymmetry is equal to $\beta\gamma \approx 0.28$, which corresponds to an average flight distance of $\sim 130\,\mu\mathrm{m}$ for the B mesons. The boost is slightly reduced with respect to KEKB, in order to mitigate the beam losses due to Touschek scattering in the lower energy beam [135]. An advantage of this modification is the improved solid angle acceptance for missing-energy decays. More technical details about SuperKEKB and its operation can be found in Ref. [132].

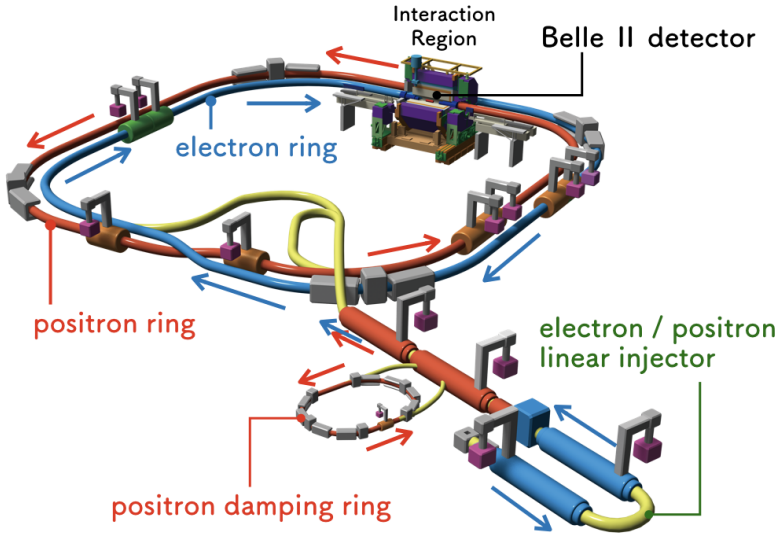


Figure 2.4: Schematic representation of SuperKEKB: the relevant accelerator components producing and delivering electrons and positrons to the interaction point are shown. Credit: SuperKEKB accelerator team.

2.4 The Belle II detector

The Belle II detector [136] consists of a system of subdetectors arranged in a cylindrical structure around the beam pipe, as shown in Fig. 2.5. Three tracking subdetectors are located in the center of Belle II. The pixel and silicon vertex detector (PXD, SVD) are used to reconstruct decay vertices and low-momentum tracks that do not reach the central drift chamber (CDC). The PXD, a new detector that was absent in Belle, is placed at a minimum radius of $14\,\mathrm{mm}$ and improves the resolution of the impact parameter for high-momentum tracks. The SVD is made of four layers of double-sided silicon strips, it was present in Belle and has been upgraded. It has larger radial coverage, new electronics and sensor design. The CDC surrounds the vertex detectors and samples charged-particles tracks to determine charge, momentum and energy loss by ionisation. It is the upgrade of the central drift chamber used in Belle. Compared to its predecessor, it presents additional layers, extending the detector to a larger radius, smaller drift cells and new readout electronics. A time-of-propagation counter (TOP) and an aerogel ring-imaging Cherenkov counter (ARICH) cover the barrel and forward end cap regions of the detector, respectively, and are used for charged-particle identification (PID). The electromagnetic calorimeter (ECL) makes up the remaining volume inside a superconducting solenoid, which generates a magnetic field of $1.5\,\mathrm{T}$. It measures the energy of electromagnetically-interacting particles, photons and electrons in particular, and corresponds to the one used in Belle. An upgraded detector to identify K_L^0 mesons and muons (KLM) is installed in the outermost part of the detector.

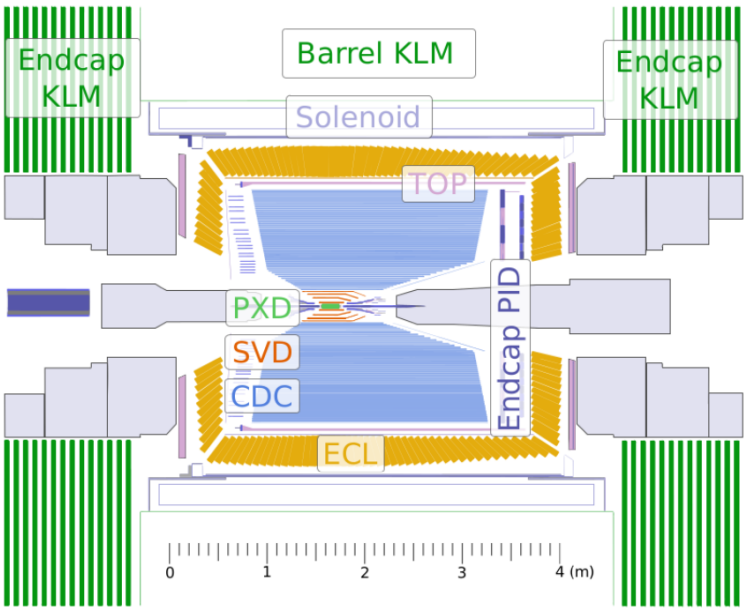


Figure 2.5: Longitudinal cross section of the Belle II detector. The polar asymmetry of the apparatus mirrors the asymmetric acceptance due to energy-asymmetric collisions. Credit: Belle II Collaboration.

2.4.1 The coordinate system

Belle II adopts a Cartesian, right-handed, coordinate system, Fig. 2.6, with origin at the nominal interaction point and axes as follows: the z axis is defined as the symmetry axis of the solenoid, and the positive direction is approximately given by the incoming electron beam; the y axis points upwards, in the direction of the detector-hall roof; the x axis points along the radial direction towards the outside of the accelerator ring. With respect to the z axis, ϕ is the azimuthal angle and θ is the zenith angle, with $\phi = 0$ defined for $(x, y, z) = (1, 0, 0)$ and $\theta = 0$ for $(x, y, z) = (0, 0, 1)$. In addition, ρ is defined as the radius in the $x - y$ plane ($\rho = \sqrt{x^2 + y^2}$) and r in three dimensions ($r = \sqrt{x^2 + y^2 + z^2}$), such that $\sin\theta = \rho/r$, $\cos\theta = z/r$ and $\cos\phi = x/\rho$.

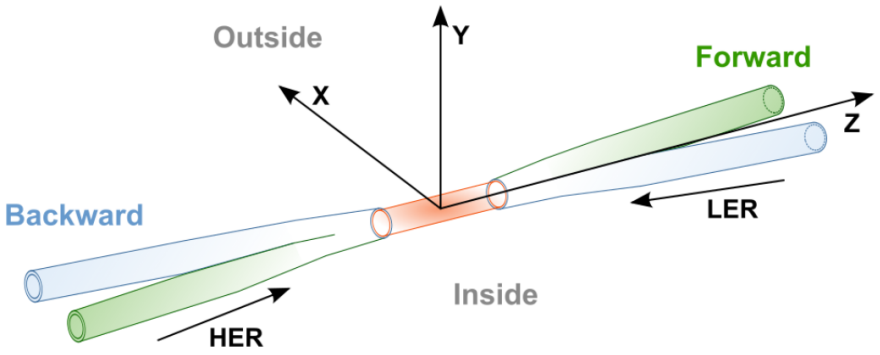


Figure 2.6: The Belle II coordinate system. Credit: Belle II Collaboration.

2.4.2 The Pixel Detector

The PXD [137] represents the two innermost layers of the vertex detector. It has been developed using the DEpleted P-channel Field Effect Transistor (DEPFET) technology [138], which combines low power consumption in the active pixel area, and low intrinsic noise, with a small material budget. The setup and the operating principle of a DEPFET pixel are schematically illustrated in Fig. 2.8. The PXD layers are situated directly outside of the $20\,\mathrm{mm}$ beam pipe, at radial distances $r_1 = 14\,\mathrm{mm}$ and $r_2 = 22\,\mathrm{mm}$, and they consist of 8 and 12 ladders, respectively. An illustration of this structure is provided in Fig. 2.7. Due to a production delay, the outer layer consists only of two ladders in the current setup, and the full installation is foreseen for the Belle II long shutdown starting at the end of summer 2022. The structure is extremely light with the thickness of a PXD layer equivalent to less than 0.2% radiation length per layer. The active area of each module is segmented into 250×768 pixels with sizes ranging from $(50 \times 55)\,\mu\mathrm{m}^2$ in the center of the inner layer to $(50 \times 85)\,\mu\mathrm{m}^2$ in the outer layer.
The PXD position makes it exposed to large QED background, hence the modules are designed to resist to $20\,\mathrm{Mrad}$ radiation dose, which is expected to be accumulated in 10 years of Belle II operation. The PXD performance in data taking observed since 2019 is optimal. The hit-efficiency of most of the modules is measured to be greater than 98%, and the detector ensures vertex reconstruction with an average spatial resolution of approximately $14\,\mu\mathrm{m}$ [137].

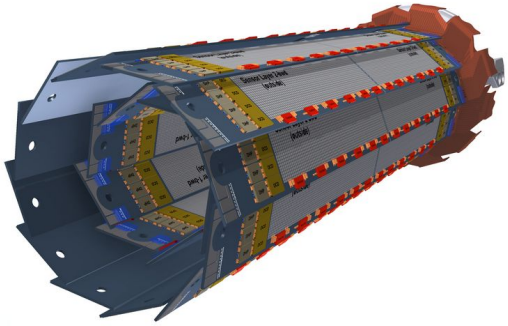


Figure 2.7: Schematic view of the geometrical arrangement of the PXD. Only two ladders of the second layer are installed in the current setup. Credit: Belle II PXD Group.

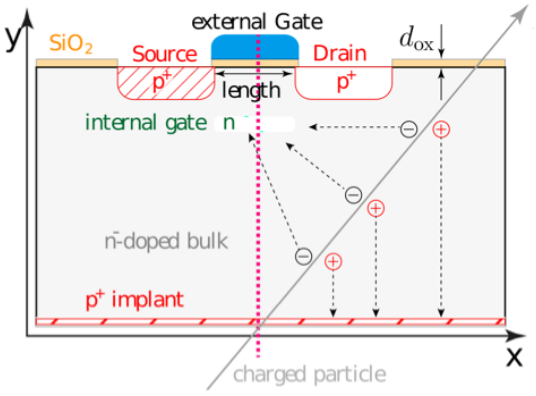


Figure 2.8: A DEPFET pixel and its operating principle. Credit: Belle II PXD Group.

2.4.3 The Silicon Vertex Detector

The SVD [139] represents a major upgrade of the Belle vertex detector. It is composed of four layers of double-sided silicon-strip sensors arranged in the structure illustrated in Fig. 2.9. The layers radii range from $39\,\mathrm{mm}$ to $135\,\mathrm{mm}$ and the acceptance of the detector covers polar angles between $17°$ and $150°$ and the full azimuthal angle. The layers have 7 to 16 ladders, with 2 to 5 sensors per ladder: rectangular sensors in the barrel-shaped part of the layers and trapezoidal in the forward section of the outermost three layers. The sensors have 768 read-out strips per side: long strips on the p-side parallel to the beam axis (z direction) and short strips on the n-side along $r - \phi$, as illustrated in Fig. 2.10. When a charged particle passes through a sensor, electron-hole pairs are created along its path by ionisation. Electrons are collected by n^- strips and holes by the p^+ strips, so that two coordinates of the particle position can be read. In total there are 172 SVD sensors and about 220 thousand strips. All the sensors are connected to fast APV25 front-end chips providing good signal-to-noise ratio for all the layers. The optimised ladder design and support structures result in a material budget per ladder equivalent to only 0.7% radiation length at normal incidence. The

SVD performance observed in data taking since 2019 is excellent. The average hit efficiency is greater than 99.5% and the spatial resolution ranges from $18\,\mu\mathrm{m}$ to $35\,\mu\mathrm{m}$ [140].

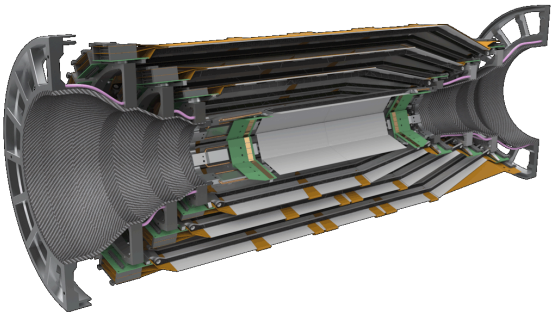


Figure 2.9: Schematic view of the geometrical arrangement of the SVD. Credit: Belle II SVD Group.

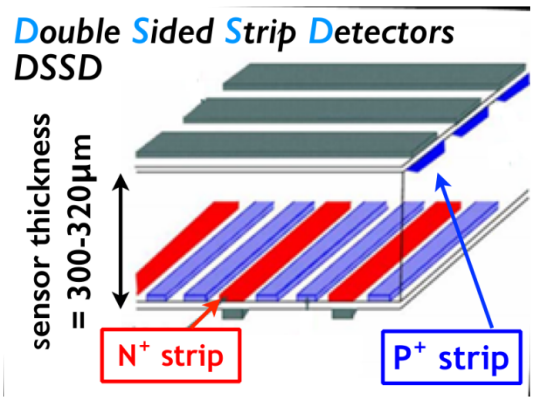


Figure 2.10: Setup of a double-sided silicon-strip sensor in the SVD. Credit: Belle II SVD Group.

2.4.4 The Central Drift Chamber

The Belle II CDC [141] is an upgrade of the Belle central drift chamber. The detector is used for multiple purposes. Primarily, it is a tracking detector designed to reconstruct charged tracks and to measure their momenta. It provides particle identification information by measuring the charged-particle energy loss within its gas volume, which is particularly useful for low momentum tracks not reaching the outer detectors dedicated to particle identification. In addition, the CDC generates reliable trigger signals for charged particles. The volume of the chamber is contained between an inner cylinder having radius of $160\,\mathrm{mm}$ and an outer cylinder with radius of $1130\,\mathrm{mm}$, and it is occupied by about $50\,000$ sense and field wires. The wires define drift cells with a size of about $2\,\mathrm{cm}$, where the electric field is approximately cylindrical. The sense wires are arranged in layers, and six or eight adjacent layers correspond to a so-called superlayer. The outer eight superlayers consist of six layers of 160 to 384 wires. The innermost superlayer is made of eight layers having 160 wires, defining smaller (half-size) drift cells, in order to cope with the increase of background at smaller radial distances from the the IP. The superlayers alternate between axial (A) orientation, aligned with the solenoidal magnetic field, and stereo (U, V) orientation, skewed with respect to the axial wires. The direction changes sign between U and V superlayers, producing an AUAVAUAVA configuration. An illustration of the CDC wire configuration is provided in Fig. 2.11. Combining information from axial and stereo superlayers it is possible to reconstruct a full three-dimensional helix track. The gas mixture used in the chamber is He-C_2H_6 50:50. The average drift velocity is $3.3\,\mathrm{cm}/\mu\mathrm{s}$ and the maximum drift time is about $350\,\mathrm{ns}$ [129]. The spatial resolution is approximately $120\,\mu\mathrm{m}$ [4].

2.4.5 Particle Identification (TOP and ARICH)

In the barrel region of the Belle II detector, charged particle identification, and in particular K/π discrimination, relies on the TOP counter [142]. This subdetector is installed on the CDC outer cover, and it consists of 16 modules composed of four parts glued together: two fused quartz bars (refractive index $n = 1.44$ for photons of $405\,\mathrm{nm}$ wavelength) of dimensions $(125 \times 45 \times 2)\,\mathrm{cm}$, a mirror located at the forward end of the bars, and a $10\,\mathrm{cm}$ long prism that couples the bar with an array of micro-channel-plate photomultiplier (MCP-PMT) tubes [143]. An illustration of a TOP module is provided in Fig. 2.12. The operating principle of the counter is shown in Fig. 2.13. A charged particle traversing a quartz radiator emits Cherenkov photons

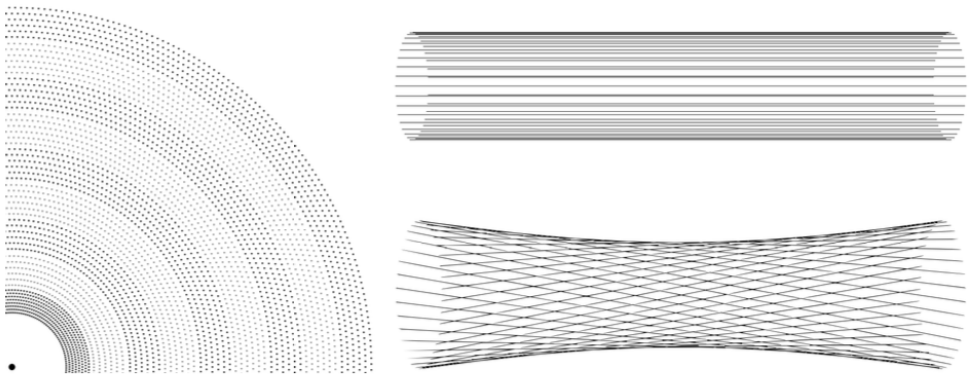


Figure 2.11: Left: a quadrant of a slice of the CDC r-ϕ projection. Right: a visualisation of stereo wires (bottom) relative to axial wires (top). The skew is exaggerated. Credit: Belle II Tracking Group.

in a cone with an opening angle θ_C, depending on the particle velocity. Part of the photons is internally reflected inside the quartz bar. The time of propagation of each photon from the emission point to the photomultiplier surface is related to θ_C.

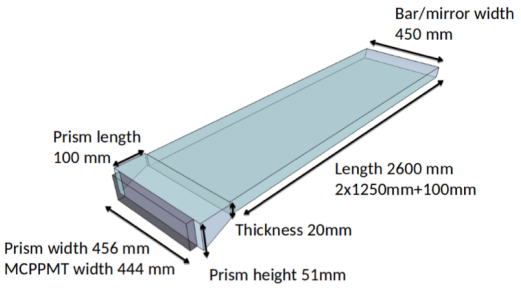


Figure 2.12: Sketch of one module of the TOP detector. Credit: Belle II TOP Group.

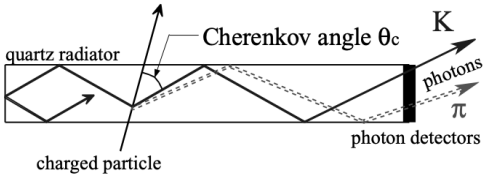


Figure 2.13: Illustration of the operating principle of the TOP detector. Credit: Belle II TOP Group.

Moreover, photon arrival times are measured using the time of the collision as a reference, thus photons are further separated in time based on the time of flight of the corresponding charged particle. Arrival-time differences between photons from kaons and pions are $\sim 100\,\mathrm{ps}$ at $2\,\mathrm{GeV}/c$ in momentum, and the time resolution of the TOP is smaller than this separation. The PID information is extracted by comparing the time distribution of the photons in one of the modules with the probability density functions (PDFs) describing the expected distributions for six particle hypotheses: e, μ, π, K, p, d. The performance of the TOP detector evaluated in terms of K/π separation is good: 85% kaon identification efficiency at a 10% pion misidentification rate [144].
In the forward endcap, the ARICH is used to separate kaons from pions over most of their momentum spectrum and to provide discrimination between pions, muons and electrons below $1\,\mathrm{GeV}/c$ [145]. The ARICH is a proximity-focusing ring-imaging Cherenkov (RICH) detector with aerogel as a radiator. An illustration of this subdetector is shown in Fig. 2.14. It has a toroidal shape with an outside radius of $1145\,\mathrm{mm}$, an inside radius of $420\,\mathrm{mm}$ and a length of $280\,\mathrm{mm}$. The detector components along the beamline, starting with the closest to the collision point, are a $40\,\mathrm{mm}$ thick radiator, made of 248 silica aerogel tiles arranged in 2 layers, an expansion space of $160\,\mathrm{mm}$, and an $80\,\mathrm{mm}$ thick photon detection system consisting

of 420 Hybrid Avalanche Photo Detectors (HAPD) [146]. The operating principle of the ARICH is shown in Fig. 2.15. The type of charged particle traversing the detector can be determined by measuring the emission angle of the Cherenkov photons. The particle identification information is based on the comparison between the observed pattern of photons and the PDFs describing the expected distribution of Cherenkov photons on the photodetector plane for different particle hypotheses. The ARICH counter is capable of separating kaon from pions with an identification efficiency of 93% at a pion misidentification rate of 10% [147].

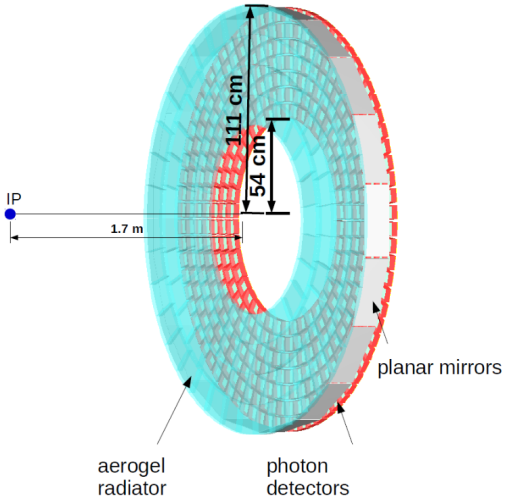


Figure 2.14: Illustration of the ARICH detector with its main components. Credit: Belle II ARICH Group.

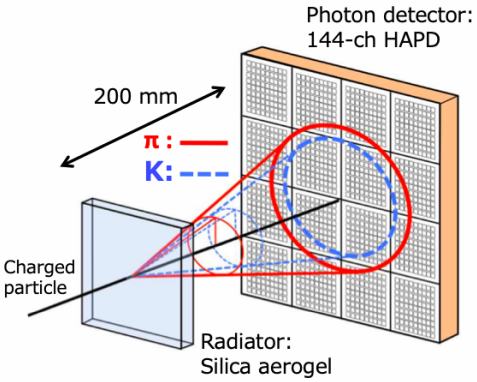


Figure 2.15: Operating principle of the ARICH counter. Credit: Belle II ARICH Group.

2.4.6 The Electromagnetic Calorimeter

The ECL is the same calorimeter used in Belle, with upgraded electronics. It allows performing various tasks, starting with energy measurements of electromagnetic particles. The other tasks include efficient photon detection, electron identification and separation from hadrons, detection of K_L^0 mesons together with the KLM, generation of signal for the trigger and also luminosity measurements.
The geometry of the electromagnetic calorimeter is shown in Fig. 2.16. It consists of a $3\,\mathrm{m}$ long barrel section with an inner radius of $1.25\,\mathrm{m}$ and forward and backward end caps at $z = 2.0\,\mathrm{m}$ and $z = -1.0\,\mathrm{m}$ from the interaction point, respectively. The detector covers a polar angle region $12.4° < \theta < 155.1°$, except for two gaps of $\sim 1°$ between the barrel section and the two end caps. The barrel section is made of 6624 thallium-doped caesium iodide CsI(Tl) crystals. Each crystal is a truncated pyramid with average size of $6 \times 6\,\mathrm{cm}^2$ in cross section and $30\,\mathrm{cm}$ in length (16.2 radiation lengths) [148]. The end caps contain 2112 caesium iodide CsI crystals. Each crystal is wrapped with a layer of $200\,\mu\mathrm{m}$ thick Teflon and covered by $50\,\mu\mathrm{m}$ thick aluminised polyethylene. The light readout is performed by $10 \times 200\,\mathrm{mm}^2$ photodiodes glued to the rear surface of the crystal. A preamplifier is attached to each photodiode providing two independent output lines from each crystal. The two pulses produced are summed in the shaper board. The average output signal from the crystals measured in calibrations with cosmic rays is ~ 5000 photoelectrons per $1\,\mathrm{MeV}$ and the noise level is $\sim 200\,\mathrm{keV}$. The energy resolution for photons, σ_E/E ranges from $\sim 2.5\%$ at $100\,\mathrm{MeV}$ to $\sim 1.7\%$ at $5\,\mathrm{GeV}$ [149].

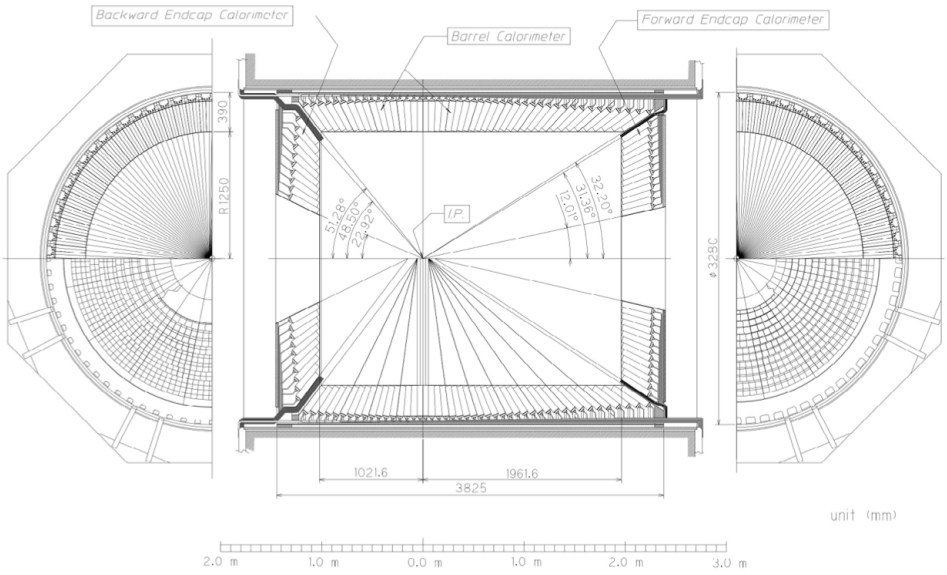


Figure 2.16: Geometrical configuration of the ECL. Credit: Belle II ECL Group.

2.4.7 Superconducting magnet

A superconducting solenoid generates a $1.5\,\mathrm{T}$ magnetic field parallel to the beam axis in a cylindrical volume of $3.4\,\mathrm{m}$ in diameter and $4.4\,\mathrm{m}$ in length. The superconductor used to build the coil is NbTi/Cu. It is supplied with a $4400\,\mathrm{A}$ current and it operates with a liquid helium cryogenic system. The magnetic field has been mapped with Hall sensors with a precision of 0.1% [148]. Inhomogeneities of the field are related to edge effects and to the presence of the final focus magnets system of SuperKEKB.

2.4.8 K_L^0 and μ detection

The KLM is an upgrade of the Belle KLM [136,150]. It is a detector dedicated to K_L^0 and μ identification, consisting of an alternating sandwich of $4.7\,\mathrm{cm}$ thick iron plates and active detector elements, located outside of the superconducting solenoid. The iron plates serve as the magnetic flux return for the solenoid and provide 3.9 interaction lengths where the K_L^0 mesons can shower hadronically [148]. Muons and non-showering charged hadrons with momentum larger than $\sim 0.6\,\mathrm{GeV}/c$ traverse the KLM until they escape the detector or they are stopped because of energy loss due to electromagnetic interaction with the detector material. K_L^0 mesons that interact in the ECL or in the iron plates of the KLM create a hadronic shower which allows their detection using both detectors or the individual detector information only. The KLM has an octagonal shape in the barrel section, covering the polar angle region $45° < \theta < 125°$. Two end caps (forward and backward) extend the angular acceptance of the detector to $20° < \theta < 155°$. The outermost layers of the barrel section are made of modules consisting of two coupled Resistive Plate Chambers (RPC) (proportional gas chambers used in streamer mode), with independent power supplies, as illustrated in Fig. 2.17. In the end caps and in the two innermost barrel layers of the KLM, where the background flux is higher, the RPCs are replaced by two layers of scintillators strips coupled with a silicon photomultiplier.
The muon detection efficiency reaches 89% for tracks with momentum above $1\,\mathrm{GeV}/c$, with hadron fake rate of $\sim 1.3\%$. The K_L^0 detection efficiency increases almost linearly with momentum to a plateau of 80% at $3\,\mathrm{GeV}/c$ [136].

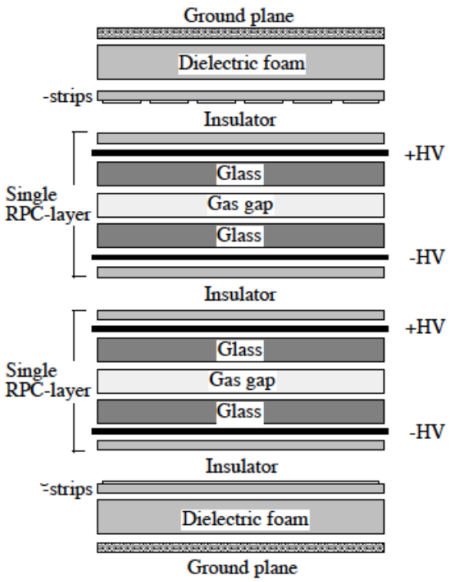


Figure 2.17: Schematic depiction of a module in the barrel section of the KLM. Credit: Belle II KLM Group.

2.4.9 The Belle II trigger system

The Belle II trigger system enables data acquisition for physics events of interest based on the information received from an appropriate set of subdetectors. Given the high luminosity, the increased background rates, and the different characteristics of the physics processes of interest, the trigger needs to be robust and flexible. The main trigger is based on the information coming from the CDC and from the ECL. It is required to have an efficiency of about 100% for hadronic events from $\Upsilon(4\mathrm{S}) \to B\bar{B}$ decays and from continuum, and a high efficiency for $e^+e^- \to \tau^+\tau^-$ and low multiplicity events. The expected event rate at the goal luminosity of $8 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is $15\,\mathrm{kHz}$, with rate of $e^+e^- \to e^+e^-$ and $e^+e^- \to \gamma\gamma$ events pre-scaled by a factor of 1/100. The upper limit on the trigger rate is imposed by the maximum acquisition frequency of the data acquisition system (DAQ), which is $30\,\mathrm{kHz}$. Additional requirements are a timing precision of $\sim 10\,\mathrm{ns}$ and a minimum time separation between two events lower than $200\,\mathrm{ns}$ [151]. The implementation of the Belle II trigger system is based on the same architecture adopted at Belle, but with totally new technologies able to support the increased event rates. The system consists of the hardware-based Level 1 (L1) trigger and the software-based High Level Trigger (HLT). The former removes most of the background events using the raw information from a set of fast subdetectors, the latter is developed with event reconstruction algorithms and is part of the DAQ.

2.4.9.1 Level 1 trigger

The L1 trigger is used to select the events of interest, and it rejects the background using sub-triggers corresponding to signals generated by specific subdetectors. All the components of the Level 1 trigger exploit the Field Programmable Gate Array (FPGA) technology, which replaces hard-wired trigger logics with configurable logics. The CDC sub-trigger, based on the information from the charged tracks detected in the drift chamber, and the ECL sub-trigger, related to the energy clusters in the calorimeter, provide the primary information to the L1 trigger. In particular, the CDC allows the reconstruction of the z coordinate of the primary vertex of an event in a few microseconds [136], suppressing a large fraction of background produced far from the nominal interaction region. The ECL generates fast trigger signals for events involving

both neutral and charged particles based on the total energy released in the calorimeter and on the number of isolated clusters. As illustrated in Fig. 2.18, the reconstructed information and the trigger signals from the CDC and the ECL are merged with the hit information from the TOP and KLM by the Global Reconstruction Logic (GRL), where a low-level reconstruction is performed. The results from the GRL are sent to the Global decision Logic (GDL), which issues a trigger based on the input information. The L1 output is then sent to the High Level Trigger for further background rejection.

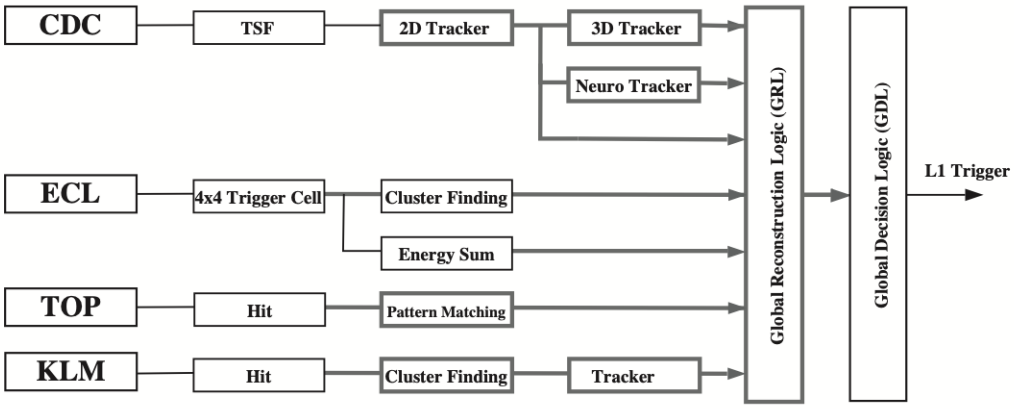


Figure 2.18: Schematic representation of the Level 1 trigger system. Credit: Belle II Trigger Group.

2.4.9.2 High Level Trigger

The HLT is a software trigger running on a server farm developed to further reject backgrounds reducing the $30\,\mathrm{kHz}$ event rate from the L1 trigger to a maximum event rate of $10\,\mathrm{kHz}$ for offline storage. Using the HLT an event can be stored or discarded based on the computed values of variables describing the event topology and other properties of the detected particles. The software trigger uses the full information of all subdetectors, except the PXD, to reconstruct the events from the L1 trigger using the same software as in the offline reconstruction. In addition, the physics trigger classifies the event category according to the occurring physics process, and this information is later used to restrict the recorded data to the specific processes of interest. The track parameters computed in the online reconstruction performed by the HLT are extrapolated to the PXD layers, where a region of Interest is defined. Only the pixels inside this region are read, and the corresponding PXD hits are used to complete the event reconstruction. For runs at the nominal luminosity, approximately 20 HLT units are planned (in the current setup only 10 HLT units are activated), each of them hosting 16 HLT workers with 20 CPUs. The online trigger software runs on each CPU of the workers, resulting in about 6400 parallel processes calculating the trigger decision. With an expected output rate of $20\,\mathrm{kHz}$ (maximum $30\,\mathrm{kHz}$) for the L1 trigger, each process has on average $320\,\mathrm{ms}$ ($213\,\mathrm{ms}$) to compute the decision [151].

2.5 SuperKEKB and Belle II program

The commissioning and operation of Belle II and SuperKEKB were divided into three phases.

- Phase 1 took place in 2016 to commission SuperKEKB. Tests of the accelerator rings, vacuum scrubbing and hardware checks were performed, but no collisions were produced. The Belle II detector was not installed, instead a system of detectors to measure the beam background was placed in the interaction region [135].

- Phase 2 took place in 2018. The Belle II detector, with just an octant of the vertex detector, was moved into its operating location. The first collision events were recorded on April 26 2018. In this phase, extensive background studies were performed, and a data sample corresponding to an integrated luminosity of $500\,\mathrm{pb}^{-1}$ was collected for early physics analyses.
- Phase 3 started in 2019 and will continue until the end of summer 2022. During this period, the PXD and SVD were installed and the full data-taking operations of Belle II started. As of today, June 2022, the total recorded integrated luminosity is $426\,\mathrm{fb}^{-1}$, and the record peak luminosity is $3.81 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. A summary of the data taking in Phase 3 is illustrated in Fig. 2.19.

The plan for the near future targets the increase of the instantaneous luminosity up to $2 \times 10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which should be achievable with the existing accelerator complex, and a consolidation of the Belle II detector. Luminosity projections until 2026 are illustrated in Fig. 2.20.
An upgrade program is under development to reach the target peak luminosity of $6.5 \times 10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The schedule covers the next decade and is presented in Ref. [152]. It is organised in three major time scales:

- 2022: the Long Shutdown 1 will start in July 2022 and last for approximately 15 months. The complete PXD will be installed, together with new TOP PMTs.
- 2026-27: the Long Shutdown 2 is in program for an upgrade of the interaction region and also of many subdetectors to improve the robustness of Belle II against larger backgrounds.
- > 2032: upgrades are under evaluation for beam polarisation and ultra-high peak luminosity to exceed $1 \times 10^{36}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

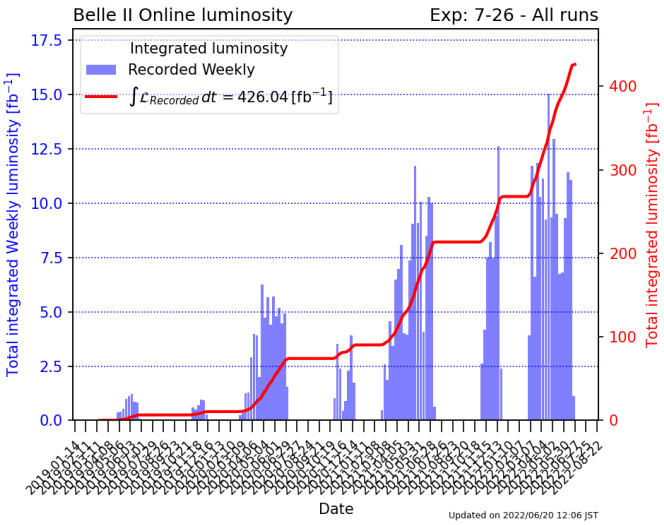


Figure 2.19: Integrated luminosity per week in Phase 3 until June 2022. The current total integrated luminosity corresponds to $426\,\mathrm{fb}^{-1}$. Credit: Belle II Collaboration.

2.6 The Belle II analysis software framework

The Belle II analysis software framework (`basf2`) [129, 153] is developed to accomplish various tasks within the experiment. It is used during the data taking, for offline reprocessing and reconstruction, and also in the data analysis. It is based on independent processing blocks called *modules*, written in C++ [154] or Python [155]. The modules are assembled in a Python *steering* script, where they are executed linearly within a defined path. Modules communicate with a common object store, called *DataStore*, which also keeps track

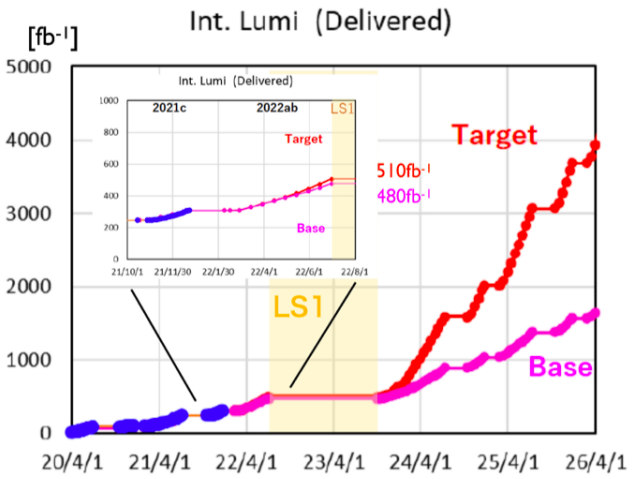


Figure 2.20: Projections of the integrated luminosity until 2026, when the Long Shutdown 2 is scheduled. Two scenarios are illustrated: the target scenario (red) based on extrapolations including expected improvements of Belle II and SuperKEKB; the base scenario (magenta) based on a conservative extrapolation. Credit: Belle II Collaboration.

of the relations between objects in each event. For example, it relates particle objects to primitive objects like tracks, or particles to vertex information and to generator level information in simulation. The data are stored using a reduced format based on ROOT `TTree` [156], which contains a collection of `basf2` objects intended for analysis and their fundamental relations. A schematic illustration of the typical data processing workflow in basf2 is shown in Fig. 2.21.

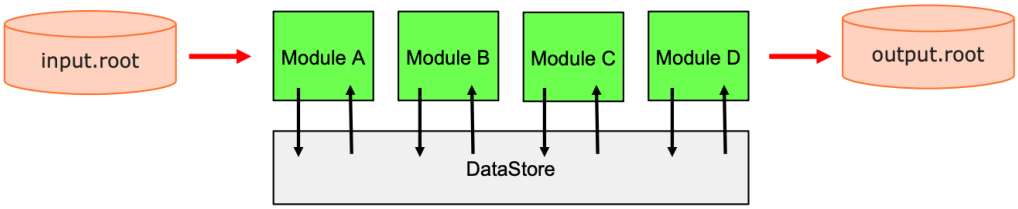


Figure 2.21: Schematic illustration of the data processing flow in basf2. Credit: Belle II Collaboration.

2.7 Simulation

The Monte Carlo (MC) simulation generates collision events and simulates the response of the Belle II detector. It is a fundamental resource for an accurate and reliable measurement, as the analysis is developed based on signal and background MC samples to which the real collision data are eventually compared for the extraction of the physics results. The possible physics processes occurring in e^+e^- collisions at $\sqrt{s} = 10.58\,\text{GeV}$, with their corresponding cross sections within the acceptance, are listed in Table 2.1. These processes are simulated in Belle II by a set of fundamental event generators. `EvtGen 1.3` [157] models the decays of B and D mesons into exclusive final states. `PYTHIA 8.2` [158] generates final states of inclusive decays and models the hadronisation of light-quark pairs. `KKMC 4.15` [159] generates τ pairs and the τ decays are modelled by `TAUOLA` [160]. QED background processes, like $e^+e^- \to e^+e^-(\gamma)$ and

$e^+e^- \to \gamma\gamma(\gamma)$ are simulated by `BABAYAGA.NLO` [161–166]. `AAFH` [167–169] simulates $e^+e^- \to e^+e^-e^+e^-$ and $e^+e^- \to e^+e^-\mu^+\mu^-$. For all the processes, given the initial conditions of the collision, particles positions and four-momenta are generated according to the specific physics model implemented in the event generator.

Physics process	Cross section [nb]	Reference
$\Upsilon(4S)$	1.110 ± 0.008	[27]
$u\bar{u}(\gamma)$	1.61	[159]
$d\bar{d}(\gamma)$	0.40	[159]
$s\bar{s}(\gamma)$	0.38	[159]
$c\bar{c}(\gamma)$	1.30	[159]
$e^+e^-(\gamma)$	300 ± 3 (MC stat.)	[165]
$\gamma\gamma(\gamma)$	4.99 ± 0.005 (MC stat.)	[165]
$\mu^+\mu^-(\gamma)$	1.148	[159]
$\tau^+\tau^-(\gamma)$	0.919	[159]
$\nu\bar{\nu}(\gamma)$	0.25×10^{-3}	[159]
$e^+e^-e^+e^-$	39.7 ± 0.1 (MC stat.)	[167]
$e^+e^-\mu^+\mu^-$	18.9 ± 0.1 (MC stat.)	[168, 169]

Table 2.1: Production cross sections for the main physics processes resulting from e^+e^- collisions at $\sqrt{s} = 10.58\,\text{GeV}$.

All event generators use the same beam parameters, which are stored in a central database. The beam energy smearing is modelled as a single Gaussian both for the HER and the LER beams. The default position of the primary vertex is the origin of the coordinate system, and the vertex smearing is introduced using the covariance matrix computed from the horizontal (x) and vertical (y) beam sizes at the IP. The response of the detector to the interaction of the generated particles with the detector material is simulated with the `Geant4` software [170]. The results are used as inputs for `basf2` to produce the corresponding signals from the Belle II detector. In the MC samples used in this thesis, simulated beam backgrounds are produced separately and added to the physics simulation. In the case of run-dependent MC, the beam background from delayed $e^+e^- \to e^+e^-$ events is overlayed to the simulation.

2.8 Reconstruction

The reconstruction represents the intermediate process necessary to turn raw detector responses into objects that can be used at the analysis level. The same reconstruction is performed on simulated and real data, but in the simulation also the true generated information is accessible. The output information from each subdetector consists of raw objects storing digitised signals, which are processed to produce low-level objects like detector hits and digits. In the reconstruction, dedicated algorithms are used to produce higher-level information from the raw objects. Typical variables produced in this step are ECL clusters, resulting from the application of clustering algorithms on sets of ECL raw objects; tracks, obtained by running tracking algorithms over collections of CDC, SVD, and PXD hits; PID variables. The final goals of the reconstruction procedure are the identification of the particles produced in the interaction and the determination of their energy and momentum. However, not all the produced particles can be correctly identified. Some particles decay before reaching the detector, some can escape the detector without interacting with the material. Moreover, every detector has an intrinsic noise that can mimic a fake interaction and additional background polluting the event can originate from the beam background. Hence, the reconstruction procedure can only transform the detector responses into a set of most likely particles on which the analyst has to perform a statistical interpretation.

2.8.1 Clustering

The clustering is one of the first steps in the reconstruction procedure and it is based on the combination of the detector responses. The clustering is adopted by different detectors in Belle II. The basic idea is to gather neighbouring pixels, strips, or crystals (depending on the detector) with a readout signal above threshold and combine them into clusters of which it is possible to measure the position and other characteristic properties. A simple example of clustering in the ECL is shown in Fig. 2.22.

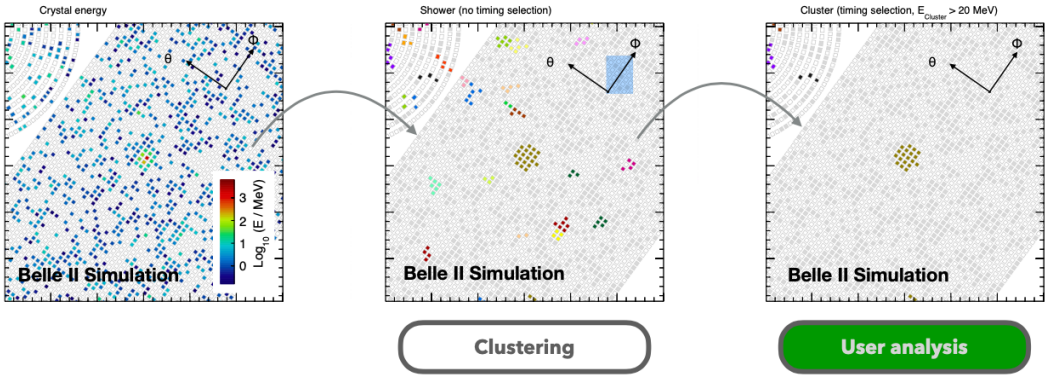


Figure 2.22: Energy deposits in single ECL crystals (left). Result of clustering (centre). Remaining clusters after time and energy selections (right). Credit: Belle II Collaboration.

2.8.2 Tracking

The Belle II track-reconstruction workflow is illustrated in Fig. 2.23. Track finding represents the first step of the track reconstruction and its goal is to find collections of hits or clusters in the tracking detectors (PXD, SVD, and CDC) which belong to the same charged-particle trajectory. Due to the different properties of the tracking detectors, different algorithms are used for each of them.
As a first step, the measured signals in the CDC are filtered and reconstructed by two independent algorithms: a global track-finding based on the Legendre algorithm [171] and a local algorithm employing a cellular automaton. The results of the two algorithms are merged, and the CDC-only tracks are fitted by means of a deterministic annealing filter (DAF) [172]. A combinatorial Kalman filter (CKF) [173] is used to enrich the CDC tracks with SVD clusters. High-curvature tracks that did not produce enough hits in the CDC are reconstructed with a standalone SVD track-finder using an advanced filter concept called Sector Map [4] and a cellular automaton. The results are combined, fitted again with a DAF, and extrapolated to the PXD with a second CKF. The final step, after the track-finding, is the track fit with the DAF provided by the GENFIT2 package [172]. For the fit, a specific particle hypothesis is assumed to calculate the energy loss and the material effects correctly. All tracks are fitted with the pion, kaon, and proton mass hypotheses. The results of the fit are stored to be used in physics analyses.
The track trajectories are represented locally using the helix parametrisation, as illustrated in Fig. 2.24, and the parameters are estimated in the track fit. The three helix parameters in the $x-y$ plane are: the signed distance of the point of closest approach (POCA) to the z axis, d_0; the angle defined by the x axis and the track transverse momentum at the POCA, ϕ_0; the track curvature signed according to the particle charge, ω. The helix can be represented by a straight line in the $s-z$ space, with s being the path length along the circular trajectory in the $x-y$ projection. The two corresponding parameters are: z_0, the z coordinate at d_0; and the tangent of the dip angle $\tan\lambda$. When the POCA is evaluated with respect to the measured IP, instead of the origin of the Belle II coordinate system, the following variables are used to define the distance of the POCA from the IP: the unsigned distance dr in the $x-y$ plane, and dz, the signed distance along z. Such

variables are preferred in physics analyses. In that case, it is also customary to use the polar angle θ of the track at the POCA, instead of λ ($\theta = \pi/2 - \lambda$).

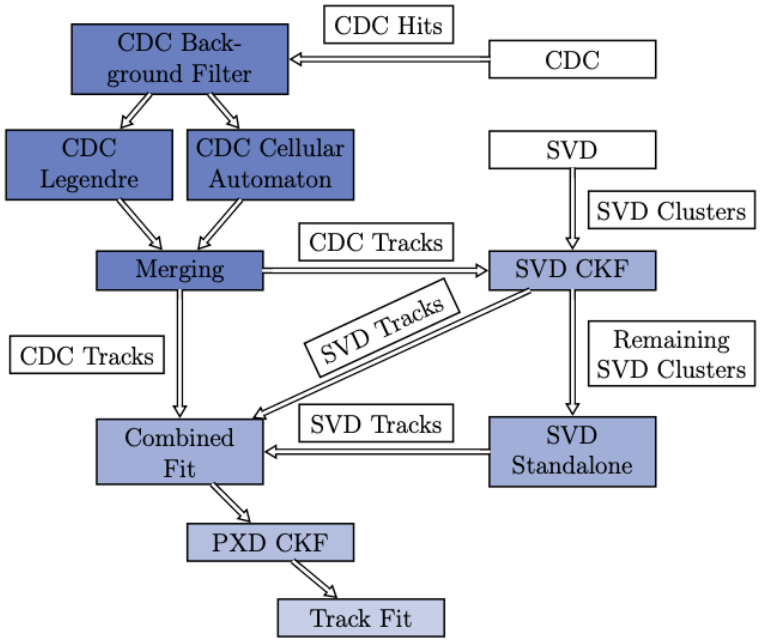


Figure 2.23: Overview of the steps in the track reconstruction workflow at Belle II. Credit: Belle II Tracking Group.

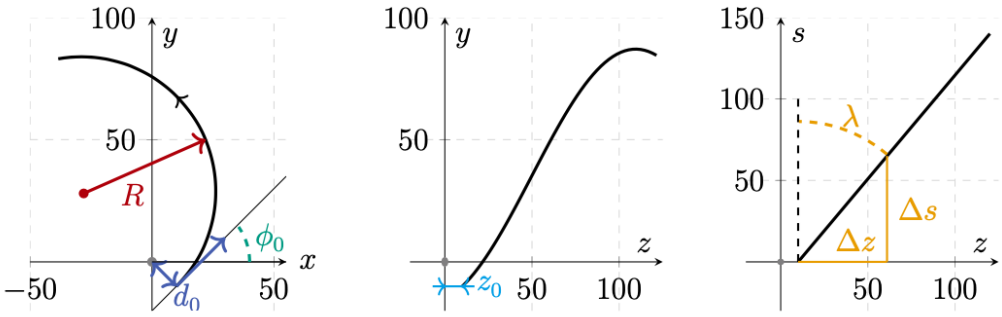


Figure 2.24: A schematic representation of the track's trajectory in the $x - y$ (left), $z - y$ (middle) and $z - s$ projections. Credit: Belle II Tracking Group.

2.8.3 Vertex fitting

Decaying particles are reconstructed from the available final state particles, and the measurable features of the decay are used to distinguish the true signal from the combinatorial background. A common technique in particle physics consists of fitting tracks together to determine if they originate from a common vertex, produced in the decay of a composite particle. In `basf2` three different algorithms can be used for vertex fits: `RAVE` [174], which is a standalone package originating from CMS vertexing libraries; `TreeFitter` [175],

recently developed in Belle II to perform a progressive fit of the whole decay chain; `KFit` [176], a vertex fitter inherited from the Belle software, which is based on least square minimisation and is well suited to fitting single vertices.

2.8.3.1 Tag Vertex Fitting

Since the signal B meson is always produced together with a tag B meson, it is often useful to fit the decay vertex of this second B. In particular, in the analyses presented in this thesis, the tag-vertex fit has played a crucial role in the background suppression procedure. For this reason, a brief description of the algorithm implemented for tag-vertex fitting is given in the following.
Figure 2.25 shows a sketch of the topology of a generic $B\bar{B}$ event. All the tracks that are not associated with the signal B meson are included in the so-called *rest of the event* (ROE). In the tag-vertex fitting algorithm, an optimal set of ROE tracks is chosen, based on the number of hits in the vertex detector and on the distance to the interaction region. Tracks from photon conversions and V^0 decays (decays of a neutral particle into two charged tracks) are removed or replaced with the mother particle. The tracks are then constrained to originate from the interaction region and fitted to a common vertex. If the χ^2 probability of the vertex fit is too large, the worst track is discarded and the vertex fit is performed again. This process is repeated iteratively until the χ^2 probability converges to a satisfactory value or no tracks are left.

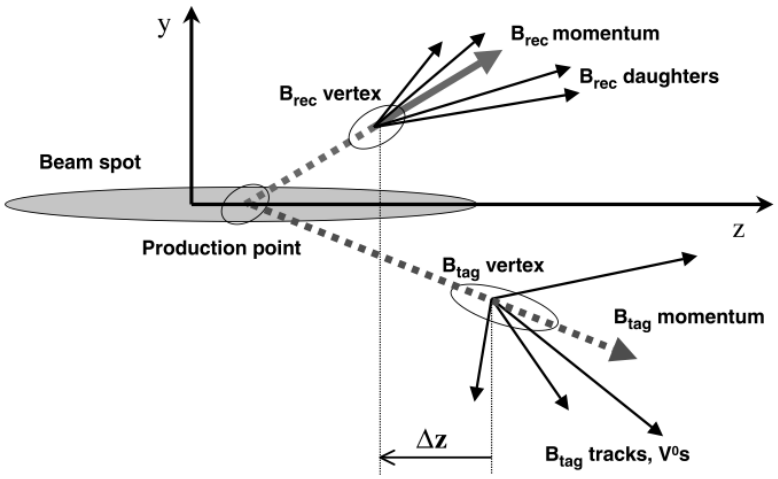


Figure 2.25: Schematic representation of a generic $\Upsilon(4\mathrm{S}) \to B\bar{B}$ decay in the yz plane. Credit: Ref. [27].

2.8.4 Charged particle identification

Final-state charged particles are identified by the way they interact with the detector. For charged hadrons, the information collected by the TOP and ARICH is combined with that from specific ionisation (dE/dx) measurements from the SVD and CDC. Similarly, the ECL provides the primary information for electron identification, and the KLM for muon identification. The PID information from each subdetector is analysed independently to determine an individual detector-based likelihood, $\mathcal{L}_i^{det}$, for each long-lived charged-particle hypothesis: electron, muon, pion, kaon, proton and deuteron ($i = e^+, \mu^+, \pi^+, K^+, p^+, d^+,$). The likelihoods are combined to create a global likelihood for each hypothesis i,

$$\mathcal{L}_i = \prod_{det} \mathcal{L}_i^{det}. \tag{2.1}$$

The final variables provided to the analyst as PID indicators are likelihood ratios:

- global ratios,

$$\mathrm{PID}_i = \frac{\mathcal{L}_i}{\sum_{j=e,\mu,\pi,K,p,d} \mathcal{L}_j}, \tag{2.2}$$

- and binary ratios,

$$\mathrm{PID}(i|j) = \frac{\mathcal{L}_i}{\mathcal{L}_i + \mathcal{L}_j}. \tag{2.3}$$

Selections based on the PID variables are characterised by an efficiency, defined as

$$\text{efficiency} = \frac{\text{Number of } i \text{ particles identified as } i}{\text{Number of } i \text{ particles}}, \tag{2.4}$$

and a misidentification (misID) rate between a particle i and a particle j,

$$\text{misID rate} = \frac{\text{Number of } i \text{ particles identified as } j}{\text{Number of } i \text{ particles}}. \tag{2.5}$$

PID efficiencies and misID rates are often different in the MC simulation and in the experimental data, and the discrepancy has a characteristic dependence on momentum and polar angle. Correction weights, corresponding to data/MC efficiency ratios are then introduced. The ratios, in bins of polar angle and momentum, are determined in control samples in which the type of charged particle can be determined from the event features without biasing the PID performance.

2.8.4.1 kaonID

The application of a PID selection on candidate signal kaons is particularly relevant in the analyses presented in this thesis. A global kaon-ID variable,

$$\mathrm{kaonID} \equiv \mathrm{PID}_K = \frac{\mathcal{L}_K}{\sum_{j=e,\mu,\pi,K,p,d} \mathcal{L}_j}, \tag{2.6}$$

has been used. The correction weights for the discrepancies in efficiency and misID rate (π^+ misidentified as K^+) between data and MC are evaluated in bins of momentum and polar angle in the control channel $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$. Slow pions in the decay are used to tag the D^0 meson, which in turn is used to identify the kaons and pions without using any true MC information. The K^+ identification efficiencies (and π^+ misID rates) are calculated in various bins of the phase space by applying a selection on the mass difference between the D^{*+} and the D^0 and fitting the mass distribution of the D^0.

2.8.5 Neutral particle identification

The identification of neutral particles like π^0, γ, K_L^0 relies on different features. Photons are identified in the ECL by means of parameters that describe the shower shape of the clusters that are not matched to a reconstructed track. The identification of photons exploits the fact that the electromagnetic shower produced by an incident photon has cylindrical symmetry in the lateral direction and the energy deposition decreases exponentially with the distance from the incident axis. The main background polluting photon clusters comes from neutral or charged hadrons interactions. These interactions result in asymmetric shower shapes and produce more than one ECL cluster not matched to charged tracks. Photon likelihoods, based on kinematics, shower shape and timing information, can be used to build specific variables to identify photon candidates with different efficiency and purity.
Neutral pions are reconstructed by combining two-photon candidates. For π^0 energies below $\sim 1\,\mathrm{GeV}$ the angular separation between the two photons results in two non-overlapping ECL clusters. For π^0 energies between $1\,\mathrm{GeV}$ and $2.5\,\mathrm{GeV}$, the two clusters overlap but can still be reconstructed as two separate photon candidates in the ECL [129]. The π^0 energy can be directly reconstructed from the four-momenta, and the energy resolution is improved by a mass-constrained fit of the two-photon candidates to the nominal π^0

mass. Likelihoods for π^0s are built based on the energy information and other characteristic features.
The identification of K_L^0 mesons is based on the information collected by the KLM and ECL. Multivariate classifiers are used to determine the probability for ECL and KLM clusters to originate from a K_L^0. The classifiers are trained with variables related to cluster shapes, kinematics and other specific interaction features. For the KLM, characteristic variables include the distance between a neutral cluster and the closest track, the cluster timing measured with respect to the time of the primary collision, the number of hits in the innermost layers of the detector; for the ECL, the distance between a neutral cluster and the closest track, the shape of the cluster, the energy deposition in the cluster.

Chapter 3

Analysis techniques

This chapter focuses on the statistical tools and techniques that are used in this thesis. Section 3.1 defines the concept of statistical inference, and Section 3.2 describes the basics of the maximum likelihood method, contextualising the tools used for the fit to data. Sections 3.3 defines confidence intervals in the frequentist framework. The concept of hypothesis test is introduced is Sec. 3.4 and is later used in Sec. 3.5 to explain the procedures employed to estimate confidence intervals and to set upper limits. Section 3.6 presents multivariate classification techniques, with particular focus on the algorithms used in the analyses of this thesis. Finally, in Section 3.7 the blind analysis technique is discussed.

3.1 Statistical inference

The goal of experimental particle physics is to make precision measurements and discover natural phenomena through the analysis of experimental data. At experiments like Belle II, millions of collision events are collected, each containing large amounts of data. These data are all different from each other due to the intrinsic randomness of the physics processes and of the detector response. Theory predicts the distributions of the observables measured in data, and the predictions depend on parameters, like particle couplings, branching fractions, etc. Experimentalists aim at extracting information on these parameters from the distribution of the data. This process is called *statistical inference*.

3.2 Maximum likelihood estimation

The maximum likelihood (ML) method is the most popular procedure for statistical inference used in particle physics. The method consists in the estimation of the parameters of an assumed probability distribution, given some observed data.
Consider the outcome of an experiment, represented by a vector of data values $\boldsymbol{x}$, and a set of parameters $\boldsymbol{\theta}$, characterising a hypothesis statement about the probability of the data. The statistical model is the probability model $f(\boldsymbol{x}|\boldsymbol{\theta})$ depending both on the data $\boldsymbol{x}$ and on the parameters $\boldsymbol{\theta}$. The function $\mathcal{L}(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})$, obtained for a specific data sample $\boldsymbol{x}$ and depending only on $\boldsymbol{\theta}$, is called the *likelihood function*. The maximum likelihood estimators for the parameters $\boldsymbol{\theta}$ are defined as the values that maximise $\mathcal{L}(\boldsymbol{\theta})$. It is usually easier to work with the logarithm of the likelihood, $\ln\mathcal{L}(\boldsymbol{\theta})$, which has the maximum for the same parameter values $\boldsymbol{\theta}$ of $\mathcal{L}(\boldsymbol{\theta})$. The estimators are found by solving the equations

$$\frac{\partial \ln\mathcal{L}}{\partial \theta_i} = 0, \qquad i = 1, ..., N, \tag{3.1}$$

where N is the total number of parameters in the model. A special case is when the data consists of a set of n statistically independent quantities, $\boldsymbol{x} = (x_1, ..., x_n)$, all following the same probability density function

(PDF) $f(x|\boldsymbol{\theta})$. In this case, the joint PDF of the data factorises and the likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\,. \tag{3.2}$$

If also the probability to observe n events depends on the parameters $\boldsymbol{\theta}$, the dependence should be included in the likelihood. For n following a Poisson distribution with mean ν, the *extended* likelihood corresponds to

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\,. \tag{3.3}$$

In the limit of large samples, the likelihood $\mathcal{L}$ is Gaussian and $\ln\mathcal{L}$ is (hyper)parabolic. In this case, the values of s times the standard deviations, σ_i, of the parameter estimators are obtained from the hypersurface defined by the values of $\boldsymbol{\theta}$ satisfying the equation

$$-2\ln\mathcal{L}(\boldsymbol{\theta}) = -2\ln\mathcal{L}_{max} + s^2\,. \tag{3.4}$$

In the analyses of this thesis, the maximum likelihood method is applied to binned data. Consider a histogram with N bins, corresponding to a vector of observed data counts $\boldsymbol{n} = (n_1, ..., n_N)$ with expectation values $\boldsymbol{\nu} = E[\boldsymbol{n}]$. If the total number of events $n_{tot} = \sum_i n_i$ is fixed, then the histogram follows a multinomial distribution

$$f(\boldsymbol{n}|\boldsymbol{\theta}) = \frac{n_{tot}!}{n_1! \cdots n_N!} p_1^{n_1} \cdots p_N^{n_N}\,, \tag{3.5}$$

where the bin probabilities p_i are functions of the parameters $\boldsymbol{\theta}$. Otherwise, if n_i are independent and Poisson distributed, the data are described by the product of Poisson probabilities

$$f(\boldsymbol{n}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}, \tag{3.6}$$

where ν_i are functions of the parameters $\boldsymbol{\theta}$.
For a histogram made of one signal and one background contribution, with expected number of events S and B, it is common to introduce also a signal strength parameter μ, such that $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ to the nominal signal + background hypothesis. In this case, the statistical model can be written as

$$f(\boldsymbol{n}|\mu, \boldsymbol{\theta}) = \mathrm{Pois}(n_{\mathrm{tot}}|\mu S + B) \prod_{i=1}^{N} \frac{\mu\nu_i^{\mathrm{sig}} + \nu_i^{\mathrm{bkg}}}{\mu S + B} = \mathcal{N}_{\mathrm{comb}} \prod_{i=1}^{N} \mathrm{Pois}(n_i|\mu\nu_i^{\mathrm{sig}} + \nu_i^{\mathrm{bkg}}), \tag{3.7}$$

where Pois indicates a Poisson distribution, ν_i^{sig} and ν_i^{bkg} the expected number of signal and background events in the bins of the histogram, and $\mathcal{N}_{\mathrm{comb}}$ a constant combinatorial factor. The model can be generalised to describe a more complex scenario. This can be done using the `HistFactory` framework [177].

3.2.1 Parameters of interest and nuisance parameters

Consider a set of measurements $\boldsymbol{x}$ described by the statistical model $f(\boldsymbol{x}|\boldsymbol{\mu})$, from which the values of the parameters of interest $\boldsymbol{\mu}$ have to be determined. In the reality of every statistical analysis, the model is not perfect in any point of the parameter space, and it is always affected to some extent by a systematic bias. The bias should be taken into account in the model by means of *nuisance* parameters $\boldsymbol{\theta}$. The introduction of nuisance parameters helps to reduce the effect of systematic uncertainties, but at the same time it increases the statistical uncertainty on the parameters of interest, due to the intrinsic correlation with the latter. For this reason, the values of the nuisance parameters are usually constrained using control measurements $\boldsymbol{y}$. In the assumption that $\boldsymbol{x}$ and $\boldsymbol{y}$ are statistically independent and $\boldsymbol{y}$ are described by a model $c(\boldsymbol{y}|\boldsymbol{\theta})$, the likelihood function becomes

$$\mathcal{L}(\boldsymbol{\mu}|\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\theta}) \cdot c(\boldsymbol{y}|\boldsymbol{\theta})\,. \tag{3.8}$$

A common choice for $c(\boldsymbol{y}|\boldsymbol{\chi})$ is a Gaussian distribution with mean equal to the expected value of the parameter and width equal to its uncertainty. In a complex model, characterised by parameters of interest and multiple nuisance parameters, the statistical uncertainty on the parameters of interest $\boldsymbol{\mu}$ can be determined from Eq. 3.4 using the *profile likelihood* function $\mathcal{L}(\boldsymbol{\mu}|\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu}))$, which depends only on the parameters of interest $\boldsymbol{\mu}$. In the profile likelihood definition, $\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu})$ indicates the values of the nuisance parameters $\boldsymbol{\theta}$ maximising the full likelihood $\mathcal{L}$ for the specified values of $\boldsymbol{\mu}$.

3.2.2 `HistFactory` and `pyhf`

In the analyses of this thesis, the `HistFactory` [177] formalism, implemented in the python library `pyhf` [3], has been used. The `HistFactory` constructs parametrised PDFs based on template histograms. Statistical models implemented with the `HistFactory` framework are built on the simultaneous measurement of disjoint binned distributions (*channels*) of event counts $\boldsymbol{n}$. In each channel, the total number of expected events (*event rate*) is obtained by summing over all the physics processes (*samples*) involved. Sample rates change according to parametrised variations, which reproduce the effect of free parameters $\boldsymbol{\eta}$ and account for systematic uncertainties as a function of constrained parameters $\boldsymbol{\chi}$. These latter parameters produce a deviation of the nominal event rate that is limited by constraint terms. Free and constrained parameters $(\boldsymbol{\eta}, \boldsymbol{\chi})$ can be further separated into parameters declared as parameters of interest $\boldsymbol{\mu}$ and nuisance parameters $\boldsymbol{\theta}$. Based on this characterisation, a generic `HistFactory` probability model consists of the product

$$f(\boldsymbol{n}, \boldsymbol{a}|\boldsymbol{\eta}, \boldsymbol{\chi}) = \prod_{c\in\text{channels}} \prod_{b\in\text{bins}} \text{Pois}(n_{cb}|\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})) \prod_{\chi\in\boldsymbol{\chi}} c_\chi(a_\chi|\chi), \tag{3.9}$$

where Pois indicates a Poisson distribution, n_{cb} the number of observed events, and $\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})$ the expected rate of events as a function of the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\chi}$. The terms $c_\chi(a_\chi|\chi)$ are one-dimensional functions of the auxiliary data a_χ constraining the specific parameter χ. The total expected event rates are defined as

$$\nu_{cb} = \sum_{s\in\text{samples}} \nu_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s\in\text{samples}} \left(\prod_{\kappa\in\boldsymbol{\kappa}} \kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right) \left(\nu^0_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) + \sum_{\Delta\in\boldsymbol{\Delta}} \Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right), \tag{3.10}$$

where ν^0_{scb} are the nominal rates, and $\kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})$, $\Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})$ are respectively multiplicative and additive rate modifiers. Each modifier is represented by a parameter. Individual bin parameters are denoted with γ and interpolation parameters with α, while λ and μ represent the luminosity and scale factor parameters and they affect all bins equally. A complete summary of the `HistFactory` modifiers and constraints is provided in Table 3.1.

Normalisation-uncertainty modifiers, $\kappa_{scb}(\alpha)$, are responsible for changes of the sample normalisations, changing the total number of events in a sample of a given channel, but keeping the shape of the event distribution invariant across the bins of the channel. Correlated-shape modifiers, $\Delta_{scb}(\alpha)$, are used to account for a source of systematic uncertainty producing a different shape modification in the various samples. In this case, a single nuisance parameter affects the expected sample rates within all the bins of a given channel. Uncorrelated-shape modifiers and MC statistical-uncertainty modifiers correspond to one nuisance parameter, $\kappa_{scb} = \gamma_b$, per bin, each coupled with its own constraint.
Constraint terms are constructed using the input data indicated in Table 3.1. Here, σ_b represent the relative uncertainties of the event rates and δ_b the event-rate uncertainties in the bins of the sample divided by the total nominal event rate $\nu^0_b = \sum_s \nu^0_{sb}$. To account for correlated-shape and normalisation uncertainties, the `HistFactory` exploits interpolating functions $f_p(\alpha|\Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=+1})$ and $g_p(\alpha|\kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=+1})$. A detailed description of the interpolation procedure is given in Ref. [177]. The quantities $\kappa_{scb,\alpha=\pm1}$ and $\Delta_{scb,\alpha=\pm1}$ represent $\pm1\sigma$ variations around the nominal expectations. In the specific case of a modifier implementing correlated shape modifications, $\Delta_{scb,\alpha=\pm1}$ are two vectors of bin-by-bin variations produced by the systematic source.
Once the statistical model is completely characterised and the parameters of interest $\boldsymbol{\mu}$ are declared, the

Description	Modifier	Constraint term c_χ	Input
Uncorrelated shape	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \mathrm{Pois}(r_b = \sigma_b^{-2} \vert \rho_b = \sigma_b^{-2}\gamma_b)$	σ_b
Correlated shape	$\Delta_{scb}(\alpha) = f_p(\alpha \vert \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=+1})$	$\mathrm{Gaus}(a = 0 \vert \alpha, \sigma)$	$\Delta_{scb,\alpha=\pm1}$
Normalisation uncertainty	$\kappa_{scb}(\alpha) = g_p(\alpha \vert \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=+1})$	$\mathrm{Gaus}(a = 0 \vert \alpha, \sigma)$	$\kappa_{scb,\alpha=\pm1}$
MC stat. uncertainty	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \mathrm{Gaus}(a_{\gamma_b} = 1 \vert \gamma_b, \delta_b)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
Luminosity	$\kappa_{scb}(\lambda) = \lambda$	$\mathrm{Gaus}(l = \lambda_0 \vert \lambda, \sigma_\lambda)$	$\lambda_0, \sigma_\lambda$
Normalisation	$\kappa_{scb}(\mu_b) = \mu_b$		
Data-driven shape	$\kappa_{scb}(\gamma_b) = \gamma_b$		

Table 3.1: Modifiers and corresponding constraints in the `HistFactory` framework.

likelihood $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\eta}, \boldsymbol{\chi})$ is defined by evaluating the statistical model $f(\boldsymbol{n}, \boldsymbol{a} \vert \boldsymbol{\eta}, \boldsymbol{\chi})$ on the observed and auxiliary data. The maximum likelihood fit is performed by minimising the profile likelihood ratio

$$-2\ln\lambda_p(\mu) = -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}\,. \tag{3.11}$$

The main properties of this function are described in Sec. 3.4.1.

3.2.2.1 Correlated shape modifications

As discussed in the previous section, the `HistFactory` can account for systematic uncertainties producing correlated shape modifications. These correlated effects are often conveniently captured by a bin covariance matrix. The matrix can be decomposed into the orthogonal $\Delta_{scb,\alpha=\pm1}$ variations that the `HistFactory` interpolates to implement a correlated shape modifier depending on a continuous nuisance parameter. A decomposition procedure based on the Singular Value Decomposition (SVD) and the Principal Component Analysis (PCA), described in Appendix D, is presented in the following.
The bin covariance matrix $\mathbf{S}$ is diagonalised as

$$\mathbf{S} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}, \tag{3.12}$$

where $\mathbf{V}$ is the matrix of the eigenvectors of $\mathbf{S}$, and $\mathbf{D}$ the diagonal matrix constructed with the corresponding eigenvalues λ_i, where i runs over the bins.
The matrix of the principal components is computed as

$$\mathbf{U} = \mathbf{\Lambda}\mathbf{V}^T, \tag{3.13}$$

with $\mathbf{\Lambda}$ corresponding to the diagonal matrix of the singular values $\sqrt{\lambda_i}$.
The main features of the matrix $\mathbf{U}$ can be well reproduced by the principal components having the largest singular values. These vectors of variations are provided to the `pyhf`, each coupled with an individual nuisance parameter responsible for correlated-shape modification. The remaining principal components are treated as uncorrelated sources of uncertainty. For each bin, their entries are summed in quadrature with the MC statistical uncertainty. Hence, after the decomposition, a $n \times n$ bin covariance matrix $\mathbf{S}$ can be approximated as

$$\mathbf{S} \approx \mathbf{u}^T\mathbf{u} + \mathrm{diag}(\sigma_1^2, ..., \sigma_n^2)\,, \tag{3.14}$$

where $\mathbf{u}$ is the matrix obtained by reducing $\mathbf{U}$ to the major principal components and σ_i^2 are the uncorrelated bin-by-bin variations.

3.2.3 Simplified Gaussian model

A statistical tool alternative to the `pyhf (HistFactory)` has been developed and used in the analyses of this thesis. The tool implements a simplified Gaussian model for the measurement. It is named `sghf`. The main assumption in the framework is to have independent bin counts following Gaussian distributions with means equal to the expected rates and Gaussian nuisance parameters. In this case, the parameter values that maximise the likelihood function $\mathcal{L}_c(\boldsymbol{\eta}, \boldsymbol{\chi})$ of a channel c are the same as those that minimise the equivalent χ^2,

$$\chi^2_c(\boldsymbol{\eta}, \boldsymbol{\chi}) = -2\ln\mathcal{L}_c(\boldsymbol{\eta}, \boldsymbol{\chi}) + \text{constant}\,. \tag{3.15}$$

In `sghf`, the χ^2 is constructed as

$$\chi^2(\boldsymbol{\mu}, \boldsymbol{\theta}) = \sum_{b\in\text{bins}} \left(\frac{n_b - \sum_{s\in\text{samples}} M^s_b(\mu_s + \sum_{e\in\text{unc.}} \Gamma^e_{sb}\theta_e))}{\sigma_b} \right)^2 + \sum_{s\in\text{samples}} p^\mu_s(\mu_s - 1)^2 + \sum_{e\in\text{unc.}} p^\theta_e \theta^2_e\,. \tag{3.16}$$

The definitions of the terms in the `sghf` χ^2 are given below:

- n_b is the total observation in a bin b;
- M^s_B are the nominal expectations for each sample s in a bin b;
- Γ^e_{sb} quantifies the influence of an uncertainty source e on a sample s in a bin b;
- μ_s is the normalisation parameter for a sample s, which for the signal sample corresponds to the unconstrained signal strength;
- θ_e is the nuisance parameter for a source of uncertainty e;
- p^μ, p^β_e are the priors for the normalisation parameters and the systematic sources, by default set to unity, with the exception of the signal strength for which the prior is $\mathcal{O}(10^{-4})$;
- σ_b is the total uncertainty in a bin b, defined as $\sqrt{n_b}$.

The minimisation of the χ^2 in Eq. 3.16 with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ leads to a system of linear equations, which can be rapidly solved using the `numpy` functions for linear algebra [178]. These functions rely on libraries such as `BLAS` and `LAPACK` [179], which provide efficient low-level implementations of standard linear algebra algorithms. `sghf` represents an efficient and reliable tool and it is used in the analyses of this thesis to validate the results obtained with the `pyhf` and also as a standalone fitting algorithm.

3.3 Confidence intervals

The result of a measurement is usually expressed by quoting the estimated value of the parameter of interest, μ, together with an interval quantifying the statistical precision of the measurement. In the simplest case, this information is provided by the estimated value $\hat{\mu}$ plus or minus the estimated standard deviation $\sigma_{\hat{\mu}}$. When this is not possible, an interval or a limit is determined such that it covers the true value of the parameter of interest with a specified probability.
In the context of frequentist inference, *confidence intervals* are estimated. These intervals are obtained with the Neyman construction [180], in which the boundary of the interval is given as a function of the data and it would fluctuate in many repetitions of the experiment. The *coverage probability* is the fraction of the intervals that would contain the true value of the parameter. Confidence intervals are constructed such that the coverage probability is greater than or equal to the chosen *confidence level* (CL). The choice of the boundary is not uniquely defined. One option is to use *central intervals*, otherwise it is possible to quote an *upper limit* or a *lower limit*. For central or symmetric confidence intervals, the CL is usually chosen equal to 68.3%. Upper or lower limits have often CL equal to 90% or 95%.

3.4 Hypothesis tests

Hypothesis testing is a method of statistical inference used to determine whether to accept or reject hypotheses, based on the observed data. In the usual practice two hypotheses are considered: the null hypothesis, H_0, and the alternative hypothesis, H_1.
A statistical test is characterised in terms of the following quantities:

- the *significance level*, α, which is the probability to reject H_0, if H_0 is assumed to be true;
- the probability β to reject the alternative hypothesis H_1, if H_1 is assumed to be true ($1-\beta$ is the so-called *power of the test*);

To maximise the power of a test of H_0 with respect to H_1, the Neyman-Pearson lemma states that, for a fixed significance level, α, the selection that minimises the probability β is based on a likelihood ratio:

$$\lambda(x) = \frac{f(\boldsymbol{x}|H_1)}{f(\boldsymbol{x}|H_0)} \geq c_\alpha, \tag{3.17}$$

where c_α is a given constant. Usually, the test is defined using a scalar function of the data $\boldsymbol{x}$ called *test statistic*. In this case, the Neyman-Pearson lemma is equivalent to the statement that the likelihood ratio $\lambda(\boldsymbol{x})$ is the optimal test statistic. In practice it can be difficult to determine $\lambda(\boldsymbol{x})$, since the PDFs $f(\boldsymbol{x}|H_0)$ and $f(\boldsymbol{x}|H_1)$ are not always exactly known. In this case, Monte Carlo models are used to generate instances of $\boldsymbol{x}$ following the PDFs, and multivariate methods are used to construct a test statistic that may approach the performance of the likelihood ratio. These methods are based on machine-learning algorithms, such as Neural Networks and Boosted Decision Trees. A description of multivariate methods is provided in Sec. 3.6.

3.4.1 Tests in searches for new phenomena

Searches for new phenomena rely on frequentist statistical tests. The goal is to observe a signal process whose existence has not been established yet. If the signal does not exist, then the analysed sample will consist only of background events. Otherwise, if the signal process does exist, both signal and background events will be present. Hence, for discovery purposes, the null hypothesis

$$H_0 : \text{ only background exists} \tag{3.18}$$

is tested against the alternative hypothesis

$$H_1 : \text{ both signal and background exist.} \tag{3.19}$$

The rejection of the background-only hypothesis H_0 implies the discovery of a new phenomenon. In this context, the level of agreement between the data and the hypothesis H_0 needs to be quantified. To do this, the analyst chooses a test statistic, t, and computes a p-value under the assumption of the hypothesis in question. If t is designed to result in large values for poor agreement with H_0, the p-value will correspond to

$$p = \int_{t_{obs}}^{\infty} f(t|H_0)dt, \tag{3.20}$$

where t_{obs} is the specific value of the test statistic obtained for the observed data. The hypothesis H_0 can be excluded if a sufficiently low p-value is observed. The p-value is often converted into an equivalent significance, Z, defined as

$$Z = \Phi^{-1}(1-p), \tag{3.21}$$

where Φ is the cumulative distribution of the standard Gaussian PDF. In particle physics, a discovery is claimed when the p-value of the background-only hypothesis is found below 2.87×10^{-7}, which corresponds to a level of significance $Z = 5$ (5σ level).
In experimental measurements, the model corresponding to a specific hypothesis always contains at least

one parameter of interest μ and some nuisance parameters $\boldsymbol{\theta}$. Hence, the p-value for a hypothesised value of μ,

$$p_\mu(\boldsymbol{\theta}) = \int_{t_{obs}}^{\infty} f(t|\mu, \boldsymbol{\theta})dt, \tag{3.22}$$

depends on the nuisance parameters $\boldsymbol{\theta}$. In the frequentist approach, the hypothesis corresponding to a certain μ is rejected only if the p-value is smaller than α for all the possible values of the nuisance parameters $\boldsymbol{\theta}$. The requirement can be satisfied by finding a test statistic t having a PDF $f(t|\mu)$, which does not depend on $\boldsymbol{\theta}$. This can always be approximately achieved with the *profile likelihood ratio*, which corresponds to the ratio between the profile likelihood, defined in Sec. 3.2.1, and the value of the likelihood at its maximum:

$$\lambda_p(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \tag{3.23}$$

The profile likelihood ratio is a powerful tool. For example, the Wilk's theorem [181] proves that, in the limit of large data samples, the distribution of $-2\ln\lambda_p(\mu)$ approaches a χ^2 distribution, dependent only on the parameter of interest μ.

3.5 Upper limits

If the discovery of the signal process by rejection of the background-only hypothesis fails, it is still possible to test various signal models and exclude those that are not compatible with the data. In this case, the model with signal plus background plays the role of the null hypothesis, and an upper limit on the parameter of interest is set. The best practice to compute upper limits is to perform a test on μ using a statistic based on the profile likelihood ratio introduced in Eq. 3.23. Hypothesised values of μ are scanned and for each of them a p-value, p_μ, is computed. In a test of size α, the set of μ values corresponding to $p_\mu \geq \alpha$ will define a confidence interval with CL equal to $1 - \alpha$. The upper extreme, $\mu_{\rm up}$, of this confidence interval is the upper limit on the signal strength μ.
An important distinction should be made between observed and expected limits. To derive observed limits, the observed data are used. Conversely, expected limits are determined using the so-called Asimov data set [182], which is made of the expected background, with no fluctuations.

3.5.1 The $\mathrm{CL_s}$ method

The $\mathrm{CL_s}$ method [183] is frequently used in high energy physics as a conservative frequentist procedure to set upper limits in searches for new phenomena. The method was specifically developed to mitigate the exclusion of signal models to which the experimental sensitivity is very low. The new quantity

$$\mathrm{CL_s} = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL_b}} = \frac{p_{s+b}}{1 - p_b}, \tag{3.24}$$

consisting of a ratio of p-values, is introduced. For a test statistic t, p_{s+b} is equivalent to p_μ, the p-value obtained in the assumption of the signal-plus-background hypothesis with signal characterised by a hypothesised value of μ, and

$$\mathrm{CL}_{s+b} = p_{s+b} = P(t \geq t_{obs}|s+b) = \int_{t_{obs}}^{\infty} f(t|s+b)dt = 1 - F(t_{obs}|\mu)\,, \tag{3.25}$$

where F is the cumulative distribution function of the test statistic t.
Conversely, p_b is the p-value of the background-only hypothesis, for which $\mu = 0$ is assumed. The corresponding CL_b is defined as

$$\mathrm{CL}_b = 1 - p_b = P(t \geq t_{obs}|b) = 1 - \int_{-\infty}^{t_{obs}} f(t|b)dt = 1 - F(t_{obs}|0)\,. \tag{3.26}$$

In the CL_s method, a signal model is rejected if $\mathrm{CL}_s \leq \alpha$. Since $1 - p_b \leq 1$, CL_s is always larger than the simple p-value of the signal-plus-background hypothesis, $\mathrm{CL}_s \geq p_{s+b}$. This is why upper limits computed with the CL_s method can be considered more "conservative". An illustration of the CL_s application is provided in Fig. 3.1. When $f(t|s+b)$ and $f(t|b)$ are well separated, if the signal-plus-background hypothesis is true, p_b will likely be very small and $\mathrm{CL}_s \simeq p_{s+b}$. If the experimental sensitivity to a specific μ is very low, then $f(t|s+b)$ and $f(t|b)$ are largely overlapping. In this case, p_{s+b} and $1 - p_b$ are both small, and the condition $\mathrm{CL}_s \leq \alpha$, for which the model is rejected, cannot be satisfied. In this way, the CL_s method protects against the exclusion of models affected by low experimental sensitivity. As discussed in Sec. 3.4.1, for statistical models depending on the signal strength μ and also on nuisance parameters $\boldsymbol{\theta}$, the appropriate test statistic to use in the CL_s method should be based on the profile likelihood ratio

$$\lambda_p(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \,. \tag{3.27}$$

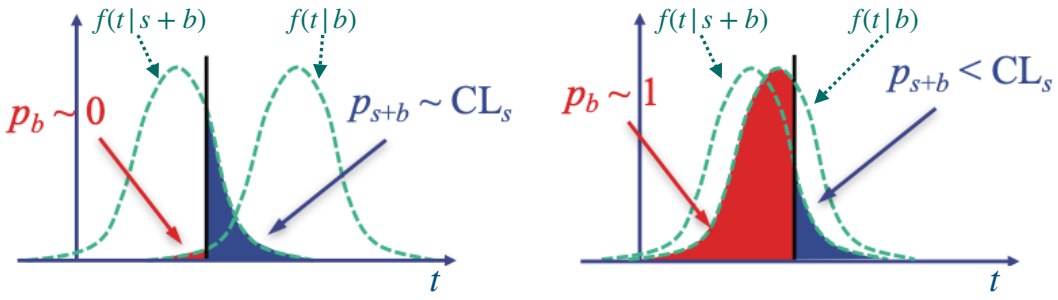


Figure 3.1: Application of the CL_s method in case of well separated distributions, $f(t|s+b)$ and $f(t|b)$, of the test statistic (left plot) and in case of large overlap of the distributions (right plot). Credit: Ref. [184].

3.5.2 Limits with `pyhf`

Upper limits can be computed with the `pyhf`. In this framework, the CL_s method is used and two possible test statistics, closely related, can be chosen. One of them is

$$q_\mu = \begin{cases} -2\ln\lambda_p(\mu) & \hat{\mu} \leq \mu \,, \\ 0 & \hat{\mu} > \mu \,. \end{cases} \tag{3.28}$$

The other is

$$\tilde{q}(\mu) = \begin{cases} -2\ln\tilde{\lambda}_p(\mu) & \hat{\mu} \leq \mu \,, \\ 0 & \hat{\mu} > \mu \,. \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \,, \\ -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu \,, \\ 0 & \hat{\mu} > \mu \,. \end{cases} \tag{3.29}$$

The latter can be preferred when considering only models for which $\mu \geq 0$. It can be noticed that $\tilde{q}(\mu)$ corresponds to $-2\ln\lambda_p(\mu)$ when $0 \leq \hat{\mu} \leq \mu$. The reason for choosing $\tilde{q}(\mu) = 0$ when $\hat{\mu} > \mu$ is that, in an upper limit, data with $\hat{\mu} > \mu$ should not be considered as less compatible with μ than the data obtained.
In principle, the CL_s based either on q_μ or on $\tilde{q}_\mu$ can be computed using simulated pseudo-experiments (toys) as

$$\mathrm{CL}_s = \frac{N(q \geq q_{obs}|s+b)}{N(q \geq q_{obs}|b)} \,, \tag{3.30}$$

where N is the fraction of pseudo-experiments satisfying the specified condition. However, the use of toys is computationally expensive. For this reason, `pyhf` computes the CL_s using asymptotic formulae derived by assuming the validity of the Wilk's theorem and of the Wald approximation [185]. In this framework, q_μ and $\tilde{q}_\mu$ are equivalent, since q_μ can be expressed as a monotonic function of $\tilde{q}_\mu$ and thus they lead to the same results [182]. The two test statistics can be approximated as

$$q_\mu = \begin{cases} \frac{(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} < \mu\,, \\ 0 & \hat{\mu} > \mu\,, \end{cases} \tag{3.31}$$

and

$$\tilde{q}(\mu) = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0\,, \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & 0 \leq \hat{\mu} \leq \mu\,, \\ 0 & \hat{\mu} > \mu\,, \end{cases} \tag{3.32}$$

where $\hat{\mu}$ follows a Gaussian distribution with mean μ and standard deviation σ. The corresponding cumulative distributions are

$$F(q_\mu|\mu) = \Phi\left(\sqrt{q_\mu}\right)\,, \tag{3.33}$$

and

$$F(\tilde{q}_\mu|\mu) = \begin{cases} \Phi\left(\sqrt{\tilde{q}_\mu}\right) & 0 < \tilde{q}(\mu) \leq \mu^2/\sigma^2\,, \\ \Phi\left(\frac{\tilde{q}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) & \tilde{q}(\mu) > \mu^2/\sigma^2\,, \end{cases} \tag{3.34}$$

where Φ is the Gaussian cumulative distribution. They can be easily computed and plugged into Eqs. 3.25, 3.26 to determine upper limits with the $\mathrm{CL_s}$ method. Moreover, for expected limits, the asymptotic formulae allow estimating the boundaries of the $\mathrm{CL_s}$ uncertainty bands, sometimes referred to as the *Brazil bands*. These bands are made of the expected $\mathrm{CL_s}$ values corresponding to the median significance of $\pm N\sigma$ variations of the signal strength from the background-only hypothesis ($\mu = 0$),

$$\mathrm{band}_{N\sigma} = \sigma\Phi^{-1}(1-\alpha) \pm N\sigma\,, \tag{3.35}$$

where $1-\alpha$ corresponds to the value of the confidence level.
The asymptotic formulae, which are discussed in detail in Ref. [182], become exact in the large sample limit, but they provide accurate results also in smaller samples. This has been validated in the first application of `pyhf` in this thesis, by comparing the asymptotic calculation with the toy-based calculation.

3.5.3 Limits with `sghf`

Estimators determined with the `sghf` have Gaussian sampling distribution in the asymptotic limit. Hence, the upper limit on the signal strength μ, at $\mathrm{CL} = 1-\alpha$, can be determined with good approximation in terms of the Gaussian significance. To do this, p_μ is set to α in Eq. 3.21, thus

$$\mu_{\mathrm{up}} = Z \cdot \sigma_\mu = \Phi^{-1}(1-\alpha) \cdot \sigma_\mu, \tag{3.36}$$

where σ_μ is the estimated uncertainty on the parameter of interest. For a 90% CL,

$$Z = \Phi^{-1}(0.90) = 1.645, \tag{3.37}$$

therefore the corresponding upper limit in `sghf` is computed as

$$\mu_{\mathrm{up}} = 1.645 \cdot \sigma_\mu\,. \tag{3.38}$$

The limits estimated in this framework are found in good agreement with the results from the `pyhf`, based on the $\mathrm{CL_s}$ method.

3.6 Multivariate classification

Over the last decades, the use of multivariate classification techniques has played a crucial role in the analysis of data from colliders. Machine learning algorithms relying on multivariate classification offer a greater insight into the collected data, outperforming traditional cut-and-count methods in background-dominated measurements where the signal is well hidden.
Multivariate classification estimates the probability of an event to correspond to signal given a dataset $\mathcal{D}$ of characteristic features $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ and targets $\boldsymbol{y} = (y_1, ..., y_n)$ (e.g., $y_i = 1$ for signal and $y_i = 0$ for background). In supervised machine learning, the existence of a map relating the features and the targets to be predicted is assumed. The goal is to approximate the map and then apply it to input events of unknown targets. The mathematical structure chosen to approximate the map is usually called *classification model*. The simplest classification model is a *linear model*, in which the target predictions are determined by means of a linear combination of weighted input features, $\hat{y}_i = \sum_j \theta_j \boldsymbol{x}_{ij}$, where the coefficients $\boldsymbol{\theta}$ are the model *parameters* that need to be *learned* from data. This task is performed in the *training* stage, in which the best model is determined by finding the values of the parameters $\boldsymbol{\theta}$ that best fit the data in the training sample. To measure how well the model fits the data, an *objective function*,

$$\mathrm{obj}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \Omega(\boldsymbol{\theta}), \tag{3.39}$$

is defined. Here, L and Ω correspond to the so-called loss function and regularisation term. The loss function tells how predictive the model is with respect to the training data. The most common loss function is the cross-entropy, defined as

$$L(\boldsymbol{\theta}) = L(\hat{\boldsymbol{y}}, \boldsymbol{y}) = -\sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - y_i) \right]. \tag{3.40}$$

The regularisation term controls the complexity of the model, helping to prevent possible overfitting. This occurs when the model is too complex and tends to learn the peculiarities of the training sample up to the point that it may fail to classify independent observations, thus producing artificial features. The opposite scenario consists of an underfitted model, in which some of the parameters that would appear in a correctly specified model are missing, and the model is not able to adequately capture the underlying structure of the data. In general, overfitting or underfitting of a model can be avoided by optimising the so-called *hyperparameters*. These are configuration parameters (e.g., the learning rate, the number of layers of a neural network or the tree depth of a boosted decision tree, etc.) whose value cannot be estimated from the training data. They are set before the training and allow controlling the learning process.

3.6.1 Decision trees

Decision trees (DTs) are non-parametric supervised-learning methods largely used for classification. A DT implements a tree-like model of decisions based on consecutive cuts, as illustrated in Fig. 3.2. The maximum number of consecutive cuts is the so-called *depth* of the tree and corresponds to one of the hyperparameters of the model. The features and the cuts used in the nodes are determined in the training. At each node, only the training data surviving the previous cuts are classified, and the resulting fractions of signal and background are computed. These quantities are used to evaluate an objective function estimating the separation power of a feature for different cuts at the node. The feature and the cut value maximising the separation between signal and background are chosen.
A single decision tree is considered a weak classifier, or *weak learner*. The predictions of a DT are often dominated by statistical fluctuations in the training sample, thus the classifier is often overfitted and the performance on an independent sample is not optimal.

3.6.2 Boosted Decision Trees

Boosted Decision Trees (BDTs) represent a class of more robust classification models built on ensembles of sequential DTs. The basic idea behind BDTs is to exploit many weak learners that are built iteratively to

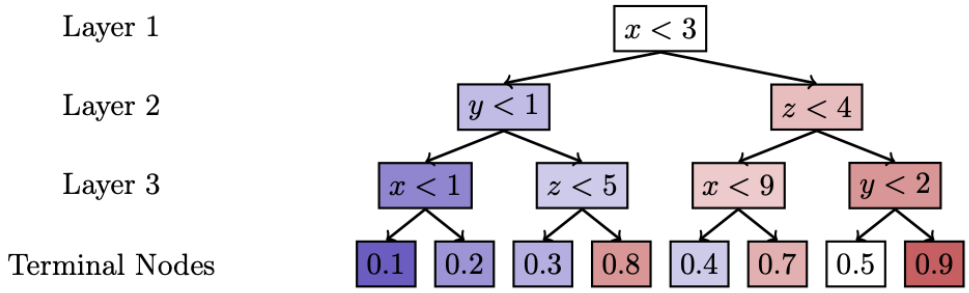


Figure 3.2: A given event, of unknown target and characterised by specific values of the features x, y, z, goes through the decision tree from top to bottom. At each node of the tree, a binary decision is made until a terminal node is reached. Credit: Ref. [186]

improve on the failures of the predecessors. At each step t, a new tree f_t is added and a weight w_t is assigned to the tree output, $f_t(\boldsymbol{x}_i)$, based on its accuracy to fit the training data. Hence, the ensemble prediction at the step t consists of the weighted sum

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} w_k f_k(\boldsymbol{x}_i). \tag{3.41}$$

To decide which tree is added at each step, the objective function

$$\text{obj}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \omega(f_i) \tag{3.42}$$

$$= \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\boldsymbol{x}_i)) + \omega(f_t) \tag{3.43}$$

is optimised in the training stage. Here l is a specific loss function and ω is the regularisation function. Many algorithms are used to iteratively add weak learners for the minimisation of a loss function. Gradient BDTs, for example, implement the gradient descent [187]. In this approach to boosting, the weak learners that best approximate the gradient of the loss function are iteratively added. In practice, the goal is to optimise a Taylor expansion of the objective function,

$$\text{obj}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(\boldsymbol{x}_i) + \frac{1}{2} h_i f_t^2(\boldsymbol{x}_i) \right] + \omega(f_t), \tag{3.44}$$

where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ are the first and second order gradients of the loss function. The `XGBoost` algorithm [188], used in one of the analyses presented in this thesis, implements the so-called *extreme gradient boosting*, an efficient version of the gradient boosting algorithm, achieving state-of-the-art performance. The other BDT algorithm used in this thesis is `FastBDT` [186], which was specifically developed within the Belle II collaboration. `FastBDT` implements a stochastic gradient BDT, in which a randomly drawn sub-sample of the training sample is used at each boosting step. This approach increases the robustness against overfitting since statistical fluctuations in the training samples are averaged out in the sum over all the trees in the ensemble.

3.7 Blind analysis

The data analyses presented in this thesis are performed using the technique known as *blind analysis*. A blind analysis protects the result of a measurement from being biased toward possible preconceptions and prior knowledge of the experimenter. In data analyses performed in particle physics, there are many potential sources of bias. The common procedure, when such biases cannot be directly eliminated, is to estimate their size and include the estimates as systematic uncertainties on the measurement. However, the experimenter's bias requires a different treatment, since it can be caused by the person performing the measurement. The bias may be unconsciously oriented towards previous measurements or prior theoretical expectations, and its size cannot be estimated. Hence, the only valid approach to prevent it consists in the use of an appropriate analysis methodology.
Searches for rare processes or decays, such as $B \to K^{(*)}\nu\bar{\nu}$, where only a few events are expected, can be seriously biased by the experimenter's decisions. These measurements are especially sensitive to the exact values of the selection requirements in data. If the selection is tuned on the knowledge of which events are included or excluded, the results are biased towards either the observation or the elimination of a signal. To protect against this potential bias, the analysis is developed using an MC simulation. However, the simulation does not always guarantee a good representation of the data, and the analyser is forced to look at the real data as much as possible without being able to infer the actual result. In the blind-analysis method adopted for the measurements in this thesis, this is done by keeping the data in the signal region hidden until the full analysis procedure is complete and the necessary sanity checks are performed. The method is well established in the field of experimental particle physics and it is largely used in the Belle II experiment.

Part II

Search for the $B^+ \to K^+ \nu\bar{\nu}$ decay with an inclusive tagging method

Chapter 4

Data samples

This chapter describes the data samples used in this search for the $B^+ \to K^+\nu\bar{\nu}$ decay. In section 4.1, the collision data samples are defined. Section 4.2 describes the simulated data samples and presents the procedure implemented to reweight the generated signal according to the SM prediction.

4.1 Collision data

This search uses data from e^+e^- collisions produced by SuperKEKB in 2019 and 2020. The on-resonance sample is made of collision data recorded by the Belle II detector at a centre-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, corresponding to the $\Upsilon(4\text{S})$ resonance. The size of the sample is equivalent to an integrated luminosity of $63\,\text{fb}^{-1}$ [5]. The estimated number of $B\bar{B}$ pairs in the on-resonance sample is:

$$N_{B\bar{B}} = [68.21 \pm 0.06(\text{stat}) \pm 0.78(\text{syst})] \times 10^6\,. \tag{4.1}$$

This value is determined using the procedure based on continuum subtraction described in Ref. [189].
An off-resonance data sample is also used in the analysis as it allows us to study and characterise the background from light-quark pair production and $e^+e^- \to \tau^+\tau^-$ processes. This sample is equivalent to an integrated luminosity of $9\,\text{fb}^{-1}$ and is collected at an energy $60\,\text{MeV}$ below the $\Upsilon(4\text{S})$ resonance.

4.2 Simulated data

The MC samples used in this analysis come from the 13th production campaign of the official Belle II Monte Carlo simulation. The samples are made of events from seven background categories. Five *continuum* categories correspond to non-resonant events involving light-quark pair production, $e^+e^- \to q\bar{q}$ ($q = u,\, d,\, c,\, s$), and tau pair production, $e^+e^- \to \tau^+\tau^-$. Two categories correspond to generic decays of $B^0\bar{B}^0$ and B^+B^- meson pairs produced on resonance.
Samples from each background category can be combined according to the effective relative cross sections of the underlying physics processes. The relative cross sections are provided in Table 4.1. In the official MC campaign, samples are grouped into *batches*, each corresponding to an integrated luminosity of $100\,\text{fb}^{-1}$, where the realistic mixture of events recorded by Belle II is reproduced.

4.2.1 Simulated signal

Official signal samples simulating $\Upsilon(4\text{S}) \to B^+B^-$ events, in which one of the two B mesons decays as $B^+ \to K^+\nu\bar{\nu}$, are produced with `EVTGEN` [157]. The `EVTGEN` model used to generate $B^+ \to K^+\nu\bar{\nu}$ decays in the official Belle II MC is `PHSP`. In this model, the kinematics of the decay products is described by a generic Lorentz-invariant three-body phase space of the form

MC category	Description	Relative cross section	Generators
Neutral B	$\Upsilon(4S) \to B^0\bar{B}^0$	9%	EVTGEN [157]
Charged B	$\Upsilon(4S) \to B^+B^-$	10%	EVTGEN [157]
$u\bar{u}$	continuum $e^+e^- \to u\bar{u}$	28%	KKMC [159], PYTHIA8 [158], EVTGEN [157]
$d\bar{d}$	continuum $e^+e^- \to d\bar{d}$	7%	KKMC [159], PYTHIA8 [158], EVTGEN [157]
$c\bar{c}$	continuum $e^+e^- \to c\bar{c}$	23%	KKMC [159], PYTHIA8 [158], EVTGEN [157]
$s\bar{s}$	continuum $e^+e^- \to s\bar{s}$	7%	KKMC [159], PYTHIA8 [158], EVTGEN [157]
$\tau^+\tau^-$	continuum $e^+e^- \to \tau^+\tau^-$	16%	KKMC [159], TAUOLA [160].

Table 4.1: The table lists the seven main background categories in the analysis, providing a description of the underlying physics processes, the relative cross sections and the generators used to produce the corresponding MC samples.

$$\prod_{j=K,\nu,\bar{\nu}} \frac{d^3p_j}{(2\pi)^3(2E_j)}\delta\left(\vec{p}_B - \sum_{j=K,\nu,\bar{\nu}} \vec{p}_j\right)\delta\left(E_B - \sum_{j=K,\nu,\bar{\nu}} E_j\right). \tag{4.2}$$

Given that the search is addressed to the discovery of the SM process, the simulated $B^+ \to K^+\nu\bar{\nu}$ decays are always reweighted according to the SM expectation presented in Sec. 1.4.1. In the SM, the decay kinematics is governed by the rescaled form factor $\rho_K(q^2)$ in Eq. 1.76, which can be parametrised in terms of three real parameters α_1, α_2, α_3, by truncating the series in Eq. 1.78 after the quadratic term in the z variable (Eq. 1.79). In Ref. [85], the combined fit to the Lattice QCD and LCSR results, in the allowed q^2 range, gives

$$\alpha_0 = 0.432 \pm 0.0011, \qquad \alpha_1 = -0.664 \pm 0.096, \qquad \alpha_2 = -1.20 \pm 0.69 \tag{4.3}$$

with the following correlation matrix

$$\mathrm{corr}(\alpha_i, \alpha_j) = \begin{pmatrix} 1 & +0.32 & -0.37 \\ +0.32 & 1 & +0.26 \\ -0.37 & +0.26 & 1 \end{pmatrix}. \tag{4.4}$$

Starting from these estimates, q^2-dependent weights are computed to scale the simple phase-space simulation to the SM expectation. As a first step, the differential branching fraction in Eq.1.74 is integrated in bins of q^2, and it is checked that the computations are consistent with the values from Ref. [85] listed in Table 4.2. The comparison is summarised in Fig. 4.1. The uncertainties on the computed values, corresponding to the widths of the rectangles, are derived from the decomposition of the covariance matrix obtained from Eq. 4.4.

$q^2[\mathrm{GeV}^2/c^4]$	$10^6 \times \mathrm{BR}(B^+ \to K^+\nu\bar{\nu})_{\mathrm{SM}}$
$0-4$	$0.93 \pm 0.14 \pm 0.05$
$4-8$	$0.92 \pm 0.11 \pm 0.04$
$8-12$	$0.86 \pm 0.09 \pm 0.04$
$12-16$	$0.71 \pm 0.07 \pm 0.03$
$16-q^2_{\mathrm{max}}$	$0.55 \pm 0.05 \pm 0.04$

Table 4.2: SM branching fractions of the $B^+ \to K^+\nu\bar{\nu}$ decay in five q^2 bins, from 0 to the kinematic limit $q^2_{\mathrm{max}} = 22.9\,\mathrm{GeV}^2/c^4$, computed in Ref. [85] The first error corresponds to the uncertainty related to the form factor and the second error quantifies the parametric uncertainties.

The computed values of the branching fraction illustrated in Figure 4.1 are in good agreement with the predictions from Ref. [85]. In particular, the expected shape in q^2, which is the fundamental feature for the reweighting, is described within 2%. Residual differences are covered by the form factor uncertainty.
The q^2 distribution expected in the SM is compared in Fig. 4.2 to the distribution computed in the basf2

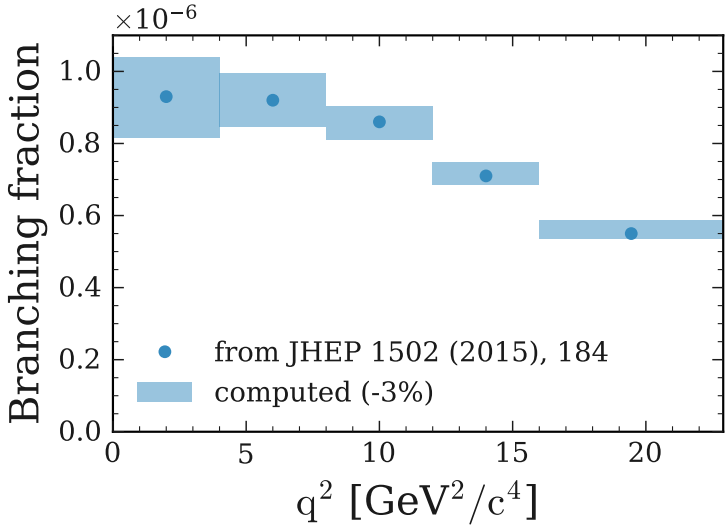


Figure 4.1: SM branching fractions of the $B^+ \to K^+\nu\bar{\nu}$ decay in five q^2 bins, from 0 to the kinematic limit $q^2_{\text{max}} = 22.9\,\text{GeV}^2/c^4$. The values from Table 4.2 are shown as blue dots and are compared to the values computed for the reweighting procedure, which are shown as shaded blue rectangles. The computations are based on the estimations in Eqs. 4.3, 4.4 and on other parameters provided in Ref. [85]. The central values are shifted by -3% for illustration purposes. The rectangle widths correspond to the form factor uncertainty.

phase-space simulation. A sizeable discrepancy between the two distributions can be observed, especially at low q^2.

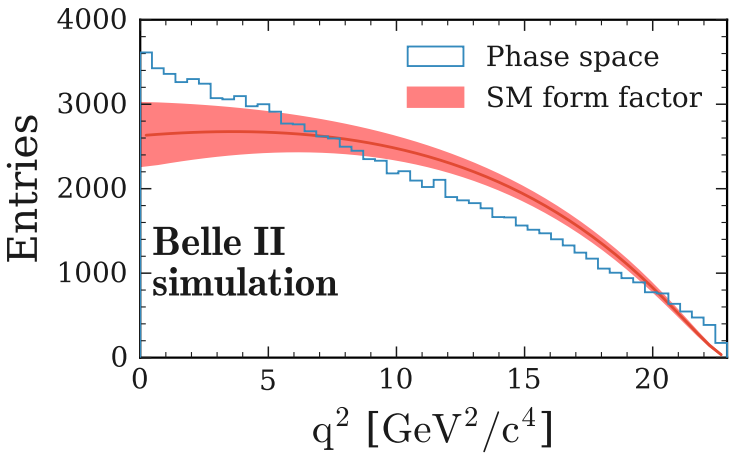


Figure 4.2: Number of $B^+ \to K^+\nu\bar{\nu}$ decays as a function of q^2 in the phase-space simulation (blue histogram) and according to the SM predictions in Ref. [85] (red line). The predictions are scaled to the total number of generated events (10^5). The red band corresponds to the form factor uncertainty. [2]

To compute the weights, the SM distribution is divided by the phase-space distribution. The resulting ratios are interpolated using a spline function from the `SciPy` library, with smoothness parameter set to 0.05 [190,191], as shown in the left plot of Fig. 4.3. The smoothed ratios obtained by the spline interpolation are used as q^2-dependent event weights for the phase-space signal distributions. In the right plot of Fig. 4.3

the transverse-momentum distribution of the charged kaon, K^+, in simulated signal events is shown before and after the reweighting. The application of the weights results in a softer p_{T} spectrum for the signal K^+.

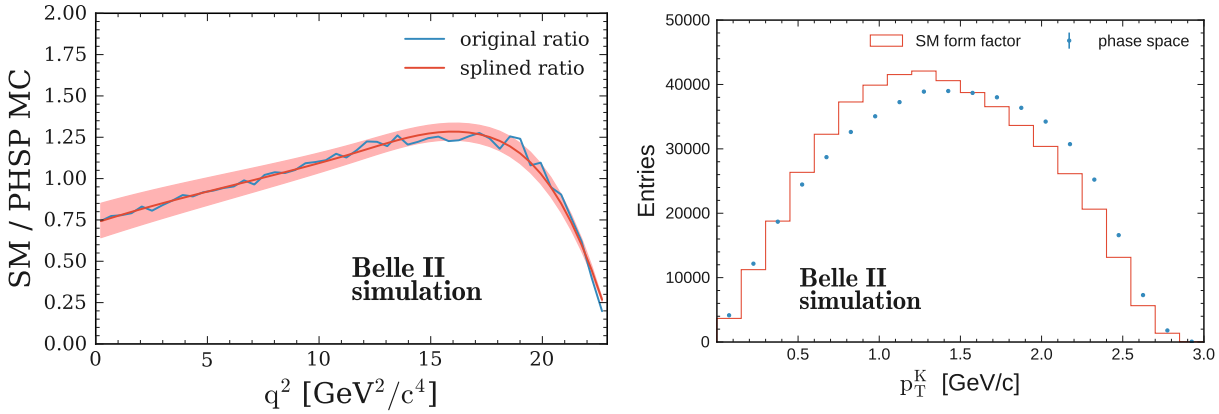


Figure 4.3: Left plot: ratios between the SM and the phase-space q^2 distributions of $B^+ \to K^+\nu\bar{\nu}$ decays in blue and splined ratios in red. The shaded red band is derived from the uncertainty on the SM form factor. Right plot: distributions of the K^+ transverse momentum in the phase-space simulation (blue points) and after the reweighting according to the SM predictions (red histogram).

Chapter 5

$B^+ \to K^+\nu\bar{\nu}$ reconstruction

This chapter introduces the inclusive tagging method implemented in this search for the $B^+ \to K^+\nu\bar{\nu}$ decay. The reconstruction procedure is described, and the discriminating variables, selected to separate signal from background, are presented. Section 5.1 gives an overview of the inclusive tagging. In section 5.2, the selection of reconstructed tracks and photon clusters in the event is presented. Sections 5.3 and 5.4 respectively describe the signal selection and the construction of the rest of the event, which collects all the reconstructed objects not associated to the candidate signal decay. In section 5.5, the studies performed to optimise the resolution of the rest of the event are summarised. Section 5.6 discusses the event selection. Section 5.7 describes the discriminating variables, grouping them by category.

5.1 The inclusive tagging method

The search for the $B^+ \to K^+\nu\bar{\nu}$ decay presented in this thesis relies on an innovative method inspired by Ref. [192]. This novel measurement approach overcomes the limitation represented by the low signal efficiency affecting the analyses performed with the hadronic and semileptonic taggings.
Given that in $B^+ \to K^+\nu\bar{\nu}$ decays the only measurable decay signature corresponds to a single charged track, the technique implemented in this analysis exploits both the properties of the signal decay and the inclusive properties of the second B-meson decay to suppress the dominant background processes. For this reason, the method is called *inclusive tagging*. A schematic representation of the inclusive tagging is provided in Fig. 5.1.

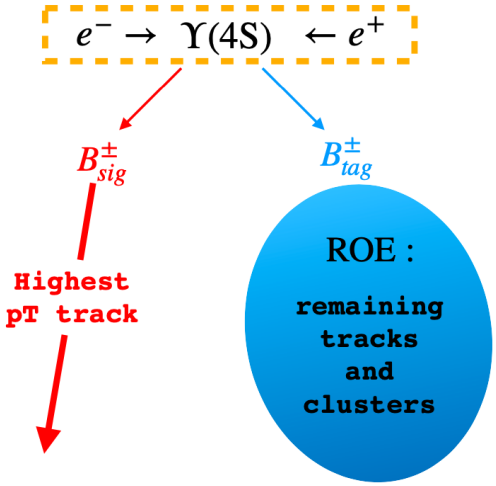


Figure 5.1: Schematic representation of the inclusive tagging.

Since the charged track generated by the signal kaon typically carries larger momentum than the background particles, the signal candidate is reconstructed as the track with the highest transverse momentum in the event. The remaining tracks and energy deposits make up the so-called *rest of the event* (ROE), which corresponds to the decay of the accompanying B meson in a signal event, while it is made of the products of two or more decays in background events. Specific features capturing the characteristic topology and kinematics of the signal and of the ROE are used to separate signal from background, with the leading backgrounds targeted individually. A set of discriminating variables, well described in the MC simulation, are chosen to identify signal events and distinguish them from events of the main background sources.

5.2 Object selection

Charged tracks and photon clusters are final-state objects reconstructed directly in the detector. Their selection represents the so-called *object selection*.
The reconstructed tracks are required to satisfy the following kinematic selections:

$$p_{\mathrm{T}} > 0.1\,\mathrm{GeV}/c, \tag{5.1}$$

$$E < 5.5\,\mathrm{GeV}, \tag{5.2}$$

$$|dz| < 3\,\mathrm{cm}, \tag{5.3}$$

$$dr < 0.5\,\mathrm{cm}, \tag{5.4}$$

$$17^\circ < \theta < 150^\circ, \tag{5.5}$$

where p_{T} is the transverse momentum of the track, dz and dr are the longitudinal and radial distances between the track's POCA and the IP, and θ is the polar angle at the POCA. In particular, the latter condition restricts the set of tracks only to those that are in the CDC geometrical acceptance. This set of selections is referred to as the *track cleanup*. The p_{T} selection is chosen in order to preserve high signal efficiency and good resolution of the ROE, as described in detail in Sec. 5.5. In studies performed in simulated signal events, it is found that the signal selection efficiency of the track cleanup is

$$(89.6 \pm 0.3)\%. \tag{5.6}$$

The photon candidates are ECL clusters without an attached track. They are required to match the conditions of the so-called *photon cleanup*,

$$0.1\,\mathrm{GeV} < E_{\mathrm{ECL}} < 5.5\,\mathrm{GeV}, \tag{5.7}$$

$$17^\circ < \theta < 150^\circ, \tag{5.8}$$

where E_{ECL} corresponds to the energy deposit in the ECL cluster. The selection in Eq. 5.7 is also introduced as a result of the optimisation studies described in Sec. 5.5.

5.3 Signal selection

The only experimental signature in the $B^+ \to K^+\nu\bar{\nu}$ decay corresponds to a kaon track. Hence, signal decays can be reconstructed by simply identifying the track of the signal K^+. The signal kaon carries on average larger momentum than the background kaons. For this reason, the signal track is selected as the highest p_{T} track in the event having at least one PXD hit attached. This latter constraint improves the resolution of the signal track impact parameter and reduces the background contamination without affecting the signal selection efficiency. This is possible thanks to the high PXD hit-reconstruction efficiency, which is measured to exceed 98% [193]. The p_{T} ranking of the signal kaon candidates is studied in simulated signal events. It is found that a true signal kaon in the acceptance defined by the track cleanup corresponds to the candidate

with the highest p_T in 79% of the cases. A summary of the results as a function of the p_T rank is provided in Fig. 5.3. The efficiency of the signal selection is also evaluated and it is found to be

$$(67.4 \pm 0.5)\%. \tag{5.9}$$

The distributions of the signal kaon candidate transverse momentum in simulated signal and background events, as well as in data from a single collision run, are shown in Fig. 5.3.

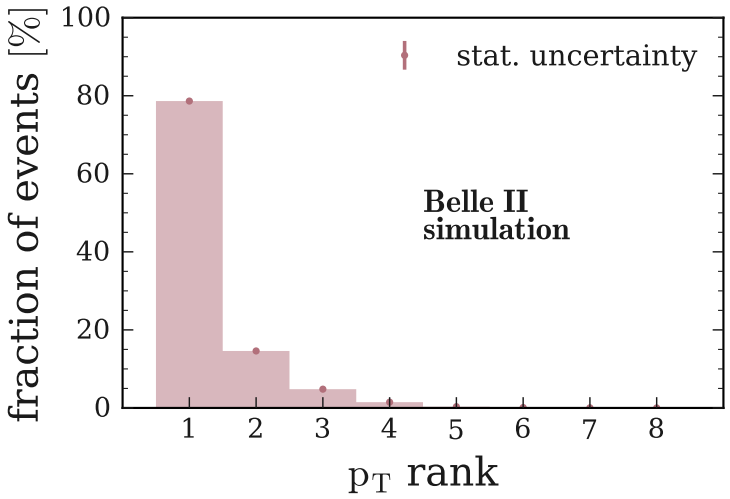


Figure 5.2: Fraction of simulated signal events in which a true signal kaon in the acceptance defined by the track cleanup corresponds to the candidate with p_T rank i. The adopted p_T ranking orders the signal kaon candidates from highest to lowest p_T, so that the rank $i = 1$ identifies the candidate with the highest p_T.

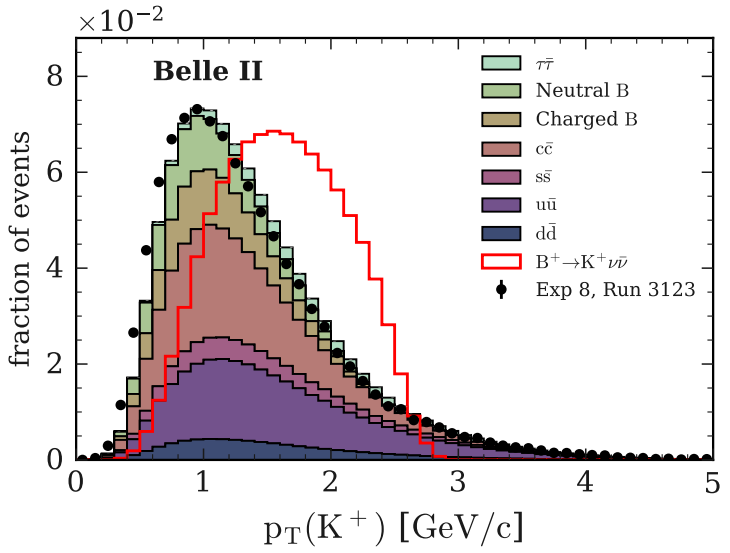


Figure 5.3: Distributions of the transverse momentum p_T of the signal kaon candidate in events from simulated signal and backgrounds, and from data collected in a single collision run. The distributions are normalised to unit areas.

5.4 Rest of the event

In the inclusive tagging, all the remaining tracks and photon clusters in the event, satisfying the track and photon cleanups, are associated to an inclusive object called *rest of the event* (ROE).
Charged final-state particles in the ROE can be electrons, muons, pions, kaons, protons. By default, charged tracks are fitted in Belle II assuming a pion-mass hypothesis. The track fit measures the magnitude of the particle's momentum. The energy of the charged particle is indirectly estimated using the measured momentum and the mass of the *most likely* particle hypothesis, which is assumed on the basis of the PID response. Every charged track reconstructed in the Belle II detector is associated to a set of global PID values (PID_e, PID_μ, PID_π, PID_K, PID_p, PID_d) computed by means of the global likelihoods defined in Eq. 2.2 of Sec. 2.8.4. Taking advantage of this information, it is possible to determine the most likely particle hypothesis for every track in the ROE and compute the energy carried by each track. To do this, the PID values assigned to a charged track are compared with the set of PID prior probabilities listed in Table 5.1, which are estimated in simulated generic $\Upsilon(4\mathrm{S})$ events [4]. If the PID_i, for the particle hypothesis i, is larger than the corresponding prior probability, then the particle hypothesis i is assumed.

Particle hypothesis	PID prior probability
electron e^+	0.058
muon μ^+	0.047
pion π^+	0.728
kaon K^+	0.149
proton p^+	0.0018
deuteron d^+	0

Table 5.1: Prior PID probabilities, for the possible charged-particle hypotheses, estimated in simulated generic $\Upsilon(4\mathrm{S})$ events. The values of the priors are taken from Ref. [4].

This setup allows to compute a set of kinematic observables that characterise the ROE differently in signal and background events. These variables are described in Sec. 5.7.4. Moreover the ROE is combined to the signal-candidate track to evaluate additional variables describing the kinematics and the shape of the whole event.
A common vertex fit is performed to all the tracks of the ROE, by means of an algorithm implementing the procedure described in Sec. 2.8.3.1. In case of signal events, the ROE tracks all originate from the decay vertex of the accompanying B meson. In generic background events, the tracks of the ROE are produced in two, or more, decays corresponding to different and displaced vertices. Therefore, in a background event, the vertex fit of all the ROE tracks to a common vertex will be characterised by low quality. Features of the ROE vertex can be exploited to separate signal from background, as illustrated in Sec. 5.7.5.

5.4.0.1 D mesons

Semileptonic B decays $B \to Dl\nu$, where D corresponds to a D^0 or D^+ meson, or to the excited states $D^*(2007)^0, D^*(2010)^+$, are a sizeable background source in the analysis, as discussed in detail in Chap. 7. The final states of these processes present missing energy, carried by the neutrino, and a charged kaon track, originating from the D^0 or D^+ meson, which can be wrongly selected as the signal-kaon candidate. To identify and characterise such background decays and suppress their contribution, possible D^0 and D^+ decays in an event are explicitly reconstructed.
To reconstruct D^0 candidates, the signal-candidate track is paired with each pion track of opposite charge in the ROE, and the two tracks are fitted to a common vertex. For D^+ candidates, the signal track and two pion tracks in the ROE, having opposite charges, are used in the vertex fit. In both cases, the ROE tracks must have at least one PXD hit. Moreover, the D^0 candidates are required to have a reconstructed invariant mass $M < 2.1\,\mathrm{GeV/c^2}$. The vertex fits are performed using `KFit` [176], and the best D^0 and D^+ candidates are chosen as the vertex fits with the highest χ^2 probabilities.

Candidate D^0 mesons are also reconstructed through vertex fits combining the track of the signal-kaon candidate with a ROE track of opposite charge, assuming its most-likely particle hypothesis. Only ROE tracks with at least one PXD hit attached are used, but no mass constraint is applied to the D^0 candidates. The best D^0 candidate corresponds to the vertex fit with the highest χ^2 probability. In this case, the idea is to target not only D^0 mesons but also other backgrounds with neutral mesons decaying into a pair of charged tracks. If correctly reconstructed, such particles are visible in the mass distribution of the best D^0 candidate as localised resonances.
Characteristic variables chosen to suppress D^0 and D^+ meson decays are described in Sec. 5.7.6.

5.5 Resolution studies

Resolution studies are performed for the optimisation of the momentum and energy selections in the track and photon cleanups. In these studies, the accuracy of the ROE reconstruction is evaluated in signal events using different selections. To do this, the reconstructed kinematics of the ROE is compared to the true kinematics of the corresponding tag B meson generated in the event. The comparison relies on two variables:

$$\Delta E = \sum_i E_i^* - \sqrt{s}/2, \tag{5.10}$$

and

$$M_{\rm bc} = \sqrt{\frac{1}{c^4}\left(\frac{\sqrt{s}}{2}\right)^2 - \frac{1}{c^2}\left(\sum_i p_i^*\right)^2}. \tag{5.11}$$

In Eqs. 5.10, 5.11, $\sqrt{s}/2$ is half of the center-of-mass energy, E_i^* and p_i^* are energies and momenta of the final-state particles, evaluated in the center-of-mass system of the incoming beams (CMS). The reconstructed quantities are indicated as $\Delta E_{\rm ROE}$ and $M_{\rm bc}^{\rm ROE}$. The corresponding quantities computed at the generator level are indicated as $\Delta E_{\rm tagB}$ and $M_{\rm bc}^{\rm tagB}$. Energies and momenta used to compute $\Delta E_{\rm tagB}$ and $M_{\rm bc}^{\rm tagB}$ are required to satisfy the selections in the track and photon cleanups, and they do not include contributions from neutrinos and K_L^0 mesons generated in the decay of the tag B meson. Using these quantities, the ROE resolution is evaluated by the estimator σ_{68} (half of the symmetric range around the median, containing 68% of the values) of the $\Delta E_{\rm ROE} - \Delta E_{\rm tagB}$ and $M_{\rm bc}^{\rm ROE} - M_{\rm bc}^{\rm tagB}$ distributions.
In the analysis, the chosen track and photon cleanups require $p_{\rm T} > 0.1\,{\rm GeV}/c$ (Eq. 5.1) and $E_{\rm ECL} > 0.1\,{\rm GeV}$ (Eq. 5.7), respectively. This allows to have a satisfactory ROE resolution, keeping a good overall data to MC agreement. The results in this setup are shown in Fig. 5.4. The scatter plots for $\Delta E_{\rm ROE}$ as a function of $\Delta E_{\rm tagB}$ and $M_{\rm bc}^{\rm ROE}$ as a function of $M_{\rm bc}^{\rm tagB}$ reveal a diluted linear correlation, as a result of the partial discrepancies between the distributions. The disagreement is enhanced in the comparison between $\Delta E_{\rm ROE}$ and $\Delta E_{\rm tagB}$ (in $M_{\rm bc}$ the leading contribution comes from $s/4$, and discrepancies related to the momenta are less visible). The $\Delta E_{\rm ROE}$ distribution is shifted towards positive values with respect to $\Delta E_{\rm tagB}$, likely as a consequence of possible wrong particle-hypothesis assignments in the ROE, resulting in mis-reconstructed energies of the charged final-state particles. Moreover, part of the $\Delta E_{\rm ROE}$ distribution leaks into the interval of positive values $0 < \Delta E_{\rm ROE} < 1\,{\rm GeV}$. A deeper investigation reveals that positive $\Delta E_{\rm ROE}$ values correspond to events with clone tracks, which cannot be removed even by using tight dz selections in the track cleanup (Eq. 5.3). Another source of discrepancy corresponds to the peak in $\Delta E_{\rm tagB}$ at values close to 0, which is not present in $\Delta E_{\rm ROE}$. This peak is produced by hadronic decays of the tag B meson, involving primarily a D^0 meson and a low-momentum pion. Figure 5.4 also shows the $\Delta E_{\rm ROE} - \Delta E_{\rm tagB}$ and $M_{\rm bc}^{\rm ROE} - M_{\rm bc}^{\rm tagB}$ distributions. Their resolutions are:

$$\sigma_{68}\left(\Delta E_{\rm ROE} - \Delta E_{\rm tagB}\right) = 0.54\,{\rm GeV} \tag{5.12}$$

and

$$\sigma_{68}\left(M_{\rm bc}^{\rm ROE} - M_{\rm bc}^{\rm tagB}\right) = 0.04\,{\rm GeV}/c^2. \tag{5.13}$$

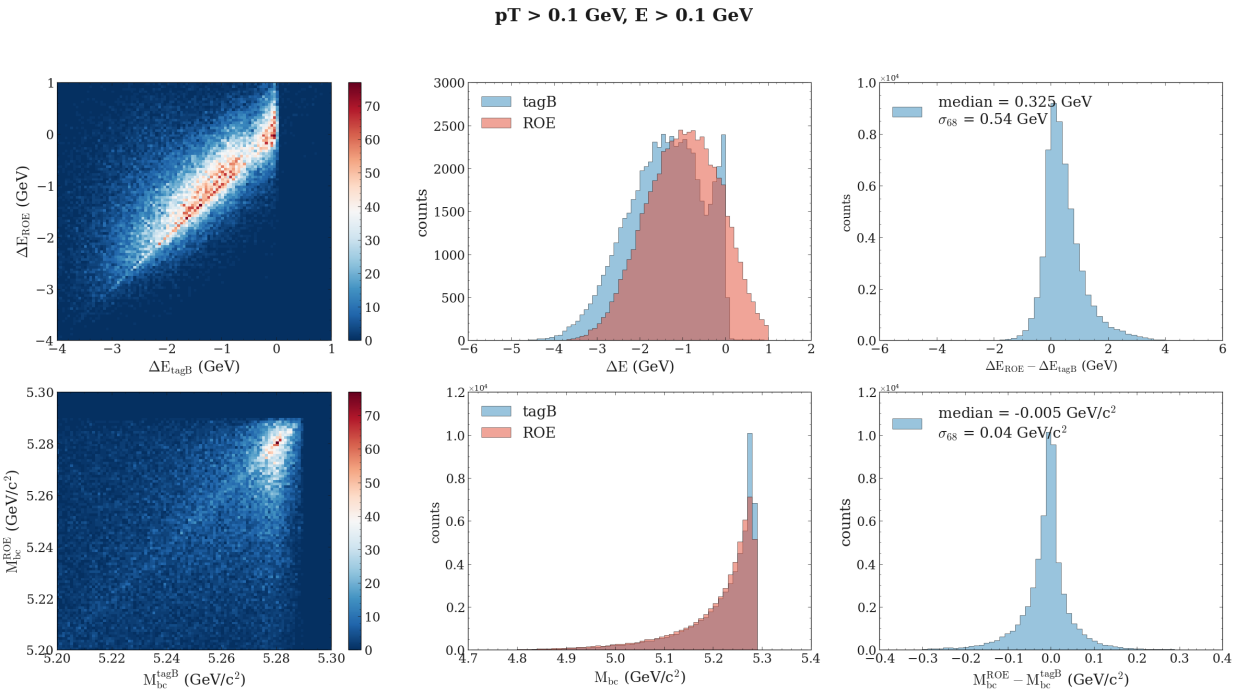


Figure 5.4: ROE resolution for the selections $p_\mathrm{T} > 0.1\,\mathrm{GeV}/c$ and $E_\mathrm{ECL} > 0.1\,\mathrm{GeV}$ in the track and photon cleanups used in the analysis. Left: scatter plots of ΔE_ROE vs ΔE_tagB (top) and $M_\mathrm{bc}^\mathrm{ROE}$ vs $M_\mathrm{bc}^\mathrm{tagB}$ (bottom). Centre: ΔE_ROE distribution in red and ΔE_tagB distribution in blue (top); $M_\mathrm{bc}^\mathrm{ROE}$ distribution in red and $M_\mathrm{bc}^\mathrm{tagB}$ distribution in blue (bottom). Right: distributions of $\Delta E_\mathrm{ROE} - \Delta E_\mathrm{tagB}$ (top) and of $M_\mathrm{bc}^\mathrm{ROE} - M_\mathrm{bc}^\mathrm{tagB}$ (bottom), with median and σ_{68} provided in the legend.

The selections $p_\mathrm{T} > 0.1\,\mathrm{GeV}/c$ and $E_\mathrm{ECL} > 0.1\,\mathrm{GeV}$ are chosen after the evaluation of the ROE resolution for different lower bounds on p_T and E_ECL. Pairs of lower bounds are defined in a 10×10-bins grid characterised by equally spaced values in the ranges $0.05\,\mathrm{GeV} \leq E_\mathrm{ECL} \leq 0.5\,\mathrm{GeV}$ and $0.05\,\mathrm{GeV}/c \leq p_\mathrm{T} \leq 0.5\,\mathrm{GeV}/c$. For each pair, the values of $\sigma_{68}\left(\Delta E_\mathrm{ROE} - \Delta E_\mathrm{tagB}\right)$ and $\sigma_{68}\left(M_\mathrm{bc}^\mathrm{ROE} - M_\mathrm{bc}^\mathrm{tagB}\right)$ are computed, as shown in Fig. 5.5. The best ROE resolution is obtained for the pair of highest lower bounds, corresponding to the selections $E_\mathrm{ECL} > 0.5\,\mathrm{GeV}$ and $p_\mathrm{T} > 0.5\,\mathrm{GeV}/c$. Moreover, the effect of raising the cut on E_ECL seems to impact more on the improvement of the ROE resolution. This is an expected feature, since the minimal energy of an ECL cluster for physics studies is around $100\,\mathrm{MeV}$ with the current level of beam background, and the increase of the lower bound on E_ECL helps to suppress low-energy QED background. Consistently, the resolution is worst when the selections are looser, $E_\mathrm{ECL} > 0.05\,\mathrm{GeV}$ and $p_\mathrm{T} > 0.05\,\mathrm{GeV}/c$.
Eventually, the selections $p_\mathrm{T} > 0.1\,\mathrm{GeV}/c$ and $E_\mathrm{ECL} > 0.1\,\mathrm{GeV}$ are chosen for the analysis, since they provide a satisfactory ROE resolution, without worsening the data to MC agreement. Moreover, these selections provide high signal efficiency in the initial stage of the analysis, which is a fundamental requirement for the inclusive tagging method.

5.6 Event selection

A set of additional selections is applied to specific variables characterising the event topology and kinematics. The total number of tracks in the event, $\mathrm{nTracks}$, is required to be

$$4 < \mathrm{nTracks} < 10. \tag{5.14}$$

The upper bound of 10 tracks is a loose condition, considering the presence of a single track in the final

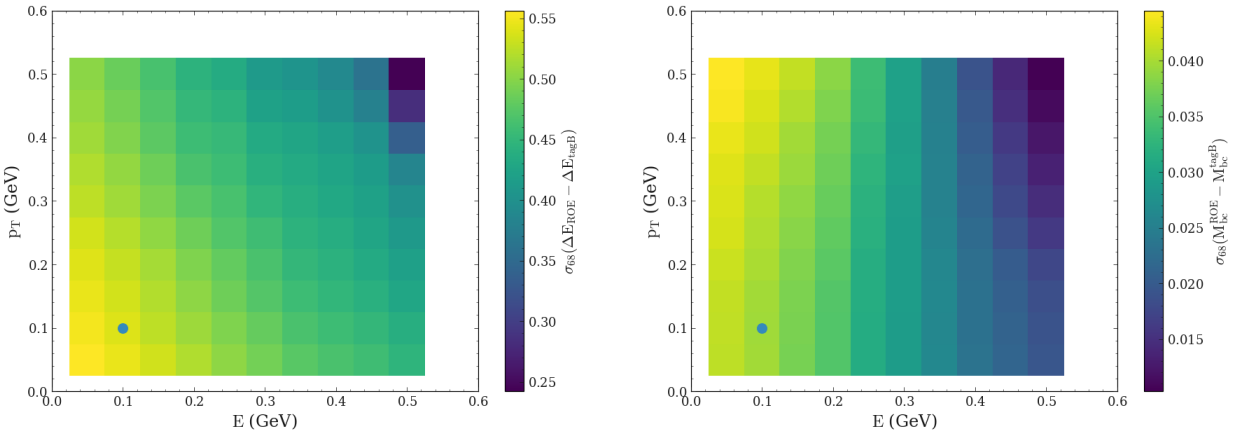


Figure 5.5: $\sigma_{68}\left(\Delta E_{\mathrm{ROE}} - \Delta E_{\mathrm{tagB}}\right)$ (left) and $\sigma_{68}\left(M_{\mathrm{bc}}^{\mathrm{ROE}} - M_{\mathrm{bc}}^{\mathrm{tagB}}\right)$ (right) for 100 pairs of lower bounds on E_{ECL} and p_{T} in the photon and track cleanups. The bin centres in the grid have coordinates corresponding to 10 equally spaced values in the intervals $0.05\,\mathrm{GeV} \leq E_{\mathrm{ECL}} \leq 0.5\,\mathrm{GeV}$ and $0.05\,\mathrm{GeV}/c \leq p_{\mathrm{T}} \leq 0.5\,\mathrm{GeV}/c$. The green circles identify the pair of selections, $E_{\mathrm{ECL}} > 0.1\,\mathrm{GeV}$, $p_{\mathrm{T}} > 0.1\,\mathrm{GeV}/c$, which are used in the analysis.

state of the signal decay. In generic B^+B^- events, where the final state of the signal B decay can have more than a track, the average number of tracks is approximately 11 [4]. The lower bound of 4 tracks helps to suppress background events with final states characterised by low track-multiplicity. To further reduce this background, the polar angle of the missing three-dimensional momentum $\theta(\vec{p}_{miss.})$ is constrained by the condition

$$17^\circ < \theta(\vec{p}_{miss.}) < 160^\circ. \tag{5.15}$$

The vector $\vec{p}_{miss.}$ is the three-momentum component of the momentum-energy four-vector

$$\begin{pmatrix} E_{miss.} \\ \vec{p}_{miss} \end{pmatrix} = \begin{pmatrix} \sqrt{s} - \sum_i E_i^* \\ \vec{P}_{\mathrm{CMS}} - \sum_i \vec{p}_i^* \end{pmatrix}, \tag{5.16}$$

where $\vec{P}_{\mathrm{CMS}}$ is the CMS three-dimensional momentum and the index i runs over the reconstructed objects in the event. Moreover, since background events with low track-multiplicity are typically characterised by a small amount of energy deposited in the detector, the total CMS energy reconstructed in the event is required to exceed $4.0\,\mathrm{GeV}$,

$$E_{\mathrm{CMS}} > 4.0\,\mathrm{GeV}. \tag{5.17}$$

A summary of the signal efficiency variations in the reconstruction algorithm is given in Table 5.2, which reports the signal efficiency computed at each selection stage.

Selection stage	Signal efficiency (%)
Object selection	89.6 ± 0.3
Signal selection	67.4 ± 0.5
Event selection	57.4 ± 0.5

Table 5.2: Signal efficiency computed at each selection stage implemented in the reconstruction algorithm of the inclusive tagging.

5.7 Discriminating variables

In the inclusive tagging, signal identification and background suppression rely on specific discriminating variables. In this analysis, a set of 50 variables is chosen. These variables are well modelled in the MC simulation, ensuring good data to MC agreement, and are selected to form a collection of features where the internal correlation is minimised as much as possible. The variables can be grouped into four main categories:

- variables describing the entire event, related to the multiplicity of the reconstructed objects (7 variables, Sec. 5.7.1) and to the event shape (16 variables, Sec. 5.7.2);
- variables describing the properties of the signal-kaon candidate (4 variables, Sec. 5.7.3);
- variables describing the ROE, related to the tracks and the energy deposits in the ROE (6 variables, Sec. 5.7.4) and to the fit of the ROE vertex (7 variables, Sec. 5.7.5);
- variables used to suppress the D^0 and D^+ meson background (10 variables, Sec. 5.7.6).

All the variables are described in the following sections. They are computed in a MC sample made of 1.6×10^6 reconstructed signal events and 1.6×10^6 reconstructed events of each of the main background categories. For each variable, the distribution obtained by stacking the backgrounds according to the relative cross sections in Table 4.1 is compared to the signal distribution and to data from a single collision run (run 3123 of experiment 8). This preliminary investigation of the data to MC agreement allows to discard variables that are not well simulated. The distributions are always normalised so that the areas under the histograms integrate to unity, and the divergence between the signal and background histograms is measured by means of the Jensen-Shannon (J-S) distance [194] (see Appendix A).

5.7.1 Event-based variables

Seven event-based variables are used in the analysis. Two of them are the number of tracks and the number of photon candidates per event. Their distributions are shown in Fig. 5.6. Since the only experimental signature of a signal decay is a kaon track, signal events are characterised by lower track and photon multiplicities. The sum of the number of tracks and photons is also used. The distributions of this variable in simulated signal and background events are shown in Fig. 5.6. As observed for the individual variables, also the combined multiplicity of tracks and photons is lower in signal than in background events.
Another event-based variable is the square of the total charge of the tracks in the event, shown in Fig. 5.7. The fraction of reconstructed events with zero total charge is larger in signal than in background, while background events are more likely to have squared total charge greater than one.
The total number of lepton candidates, $e^\pm$, $\mu^\pm$, per event is also computed. This variable targets semileptonic decays, which represent a sizeable source of background in the analysis, as discussed in Chap. 7. These decays are characterised by final states with a charged lepton and a neutrino. Candidate charged leptons are identified as candidate electrons and muons in the ROE with $\mathrm{electronID} > 0.9$ and $\mathrm{muonID} > 0.9$, respectively. These PID selections are based on electron and muon global PID likelihood ratios, defined in Eq. 2.2. The distributions of the total number of lepton candidates in reconstructed MC events from signal and background are shown in Fig. 5.7. Events with zero leptons can indicate hadronic final states, in particular hadronic tag B decays in signal events, but can also occur when the leptons are out of the ROE acceptance. In reconstructed signal events, only one lepton can identify a semileptonic tag B decay, and more than one lepton can indicate multiple semileptonic decays in the tag side, where often a D^0 is present ($\mathrm{BR}(D^0 \to h l\nu) > 10\%$ [43]).
The polar angle of the reconstructed missing three-dimensional momentum $\theta(\vec{p}_{miss.})$ is also used to separate signal events from background. The distributions of $\theta(\vec{p}_{miss.})$ are shown in Fig. 5.8. For all the background categories, the distribution is flat over almost the entire allowed range of values, while it has a characteristic shape for signal events. In this case, the distribution is unimodal, and the most frequent direction of the missing momentum is almost perpendicular to the beam axis ($\theta(\vec{p}_{miss.}) \sim 1.6\,\mathrm{rad} \approx 92^\circ$), as a consequence of

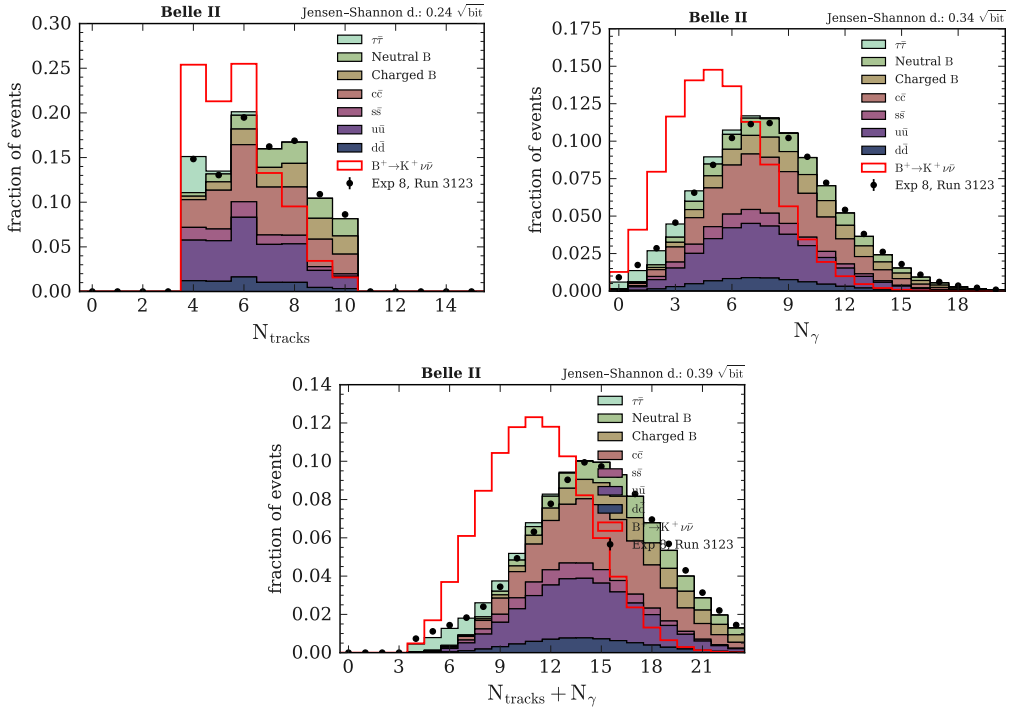


Figure 5.6: Distributions of: number of tracks (left), photons (right), tracks plus photons (bottom) per event. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

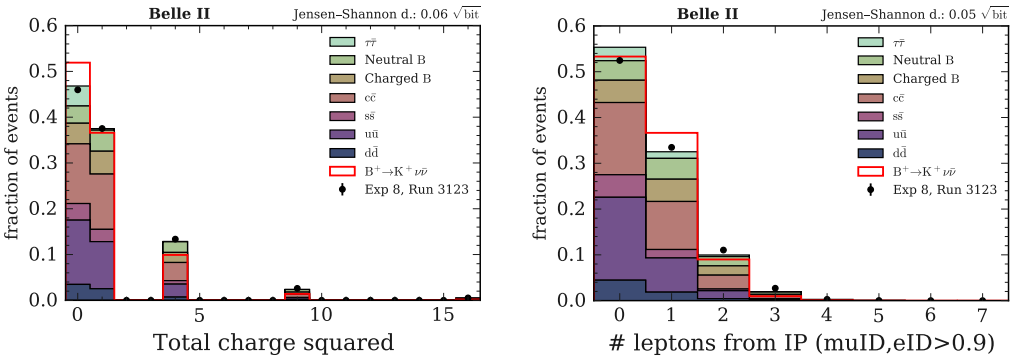


Figure 5.7: Distributions of: squared total charge (left) and number of lepton candidates (right) per event. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

the specific kinematics of the signal decay. Another variable built with the information from the reconstructed

missing four-momentum is the squared missing invariant mass, also shown in Fig. 5.8. Background events have a distribution sharply peaking around zero with a long right tail. The signal distribution is characterised by larger values of the missing invariant mass squared, and it lies, almost entirely, over values greater than zero. The shape is unimodal, peaking at $\sim 10\,\mathrm{GeV}^2/c^4$, and skewed to the right.

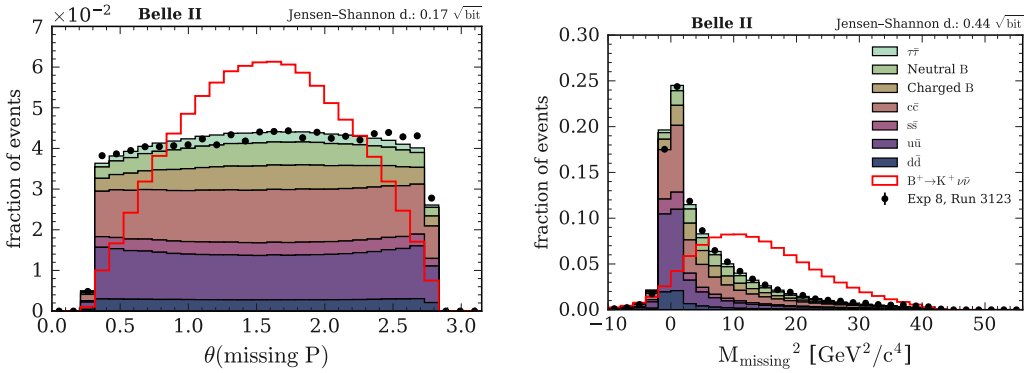


Figure 5.8: Distributions of: polar angle of the reconstructed missing three-dimensional momentum (left) and squared missing invariant mass (right) per event. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

5.7.2 Event-shape variables

Sixteen event-shape variables are used in this analysis. These observables, computed in the CMS, are designed to describe the global shape of the event. Figure 5.9 provides schematic representations of the possible event topologies. Events with two generic B decays have spherical symmetry. The B mesons are produced almost at rest in the CMS, therefore their decay products are isotropically distributed in space. The final state of continuum events consists of two back-to-back jets of particles produced along a natural symmetry axis. In signal events, the final state of the $B^+ \to K^+\nu\bar{\nu}$ decay is made of a single high-momentum track, while the tag B meson decays generically, thus the event shape has a mixed symmetry.
The symmetry axis is well reproduced by the so-called *thrust* axis, $\vec{v}_T$, which is the axis maximising the longitudinal projections of the particles momenta, $\vec{p}_i$. A related observable is the event thrust, defined as

$$T(\vec{v}_T) = \max_{|\vec{v}_T=1|} \frac{\sum_i |\vec{v}_T \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|}. \tag{5.18}$$

The event thrust and the cosine of the thrust-axis polar angle are both used. The distributions of these two variables are shown in Fig. 5.10. The event thrust provides a sizeable distinction of signal events from continuum events. The thrust distribution in the continuum background peaks at larger values than in signal. The separation is slightly reduced for the cosine of the thrust-axis polar angle. The signal peaks at a cosine of approximately 0.3, corresponding to a polar angle $\theta \sim 70°$. The background peak is shifted to values closer to one. This is due in particular to continuum events. For the charged and neutral B backgrounds, the cosine of the thrust-axis polar angle is more uniformly distributed.
Another event-shape variable is the cosine of the angle between the ROE thrust axis and the thrust axis of the signal B meson, which corresponds to the direction of the signal-kaon momentum. The distributions of this variable in signal and background events are shown in Fig. 5.10. The variable has a large separation power. The background distribution presents a long left tail and a prominent peak at values of cosine around one. This feature is primarily produced by continuum events, in which the signal axis is almost parallel to

the ROE thrust axis, given the existence of a common symmetry axis in the event. In signal and generic B events, the variable is more uniformly distributed.

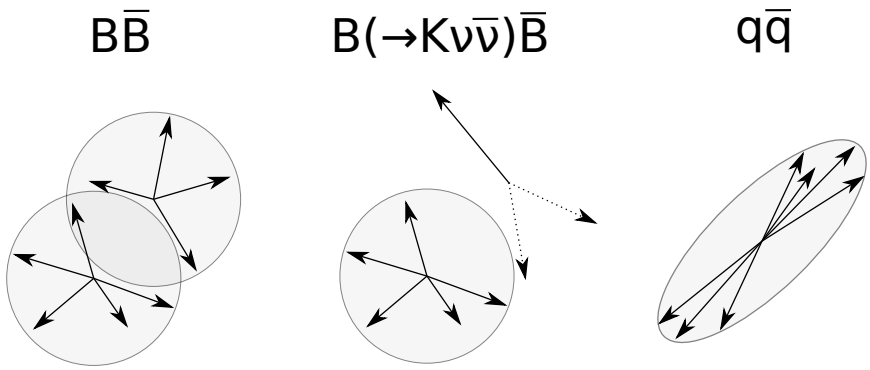


Figure 5.9: Schematic representations of the event topologies: a generic event with two B meson decays (left), an event with one of the B mesons decaying into one kaon and two neutrinos (middle), and a $e^+e^- \to q\bar{q}$ event with $q = u, d, s, c$ (right).

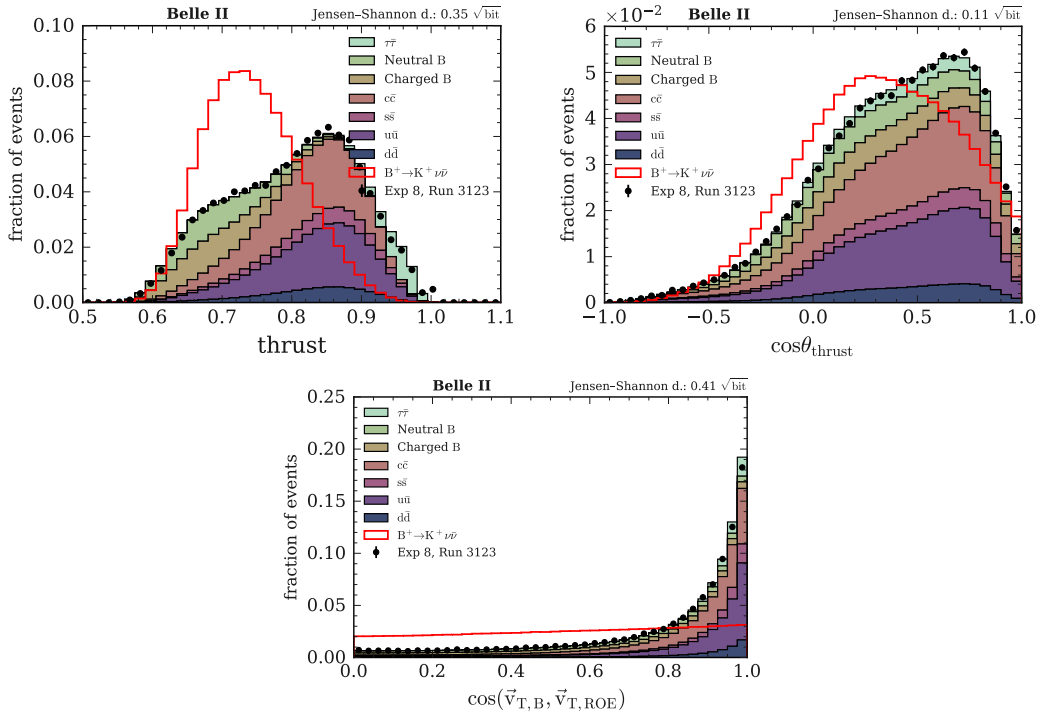


Figure 5.10: Distributions of: event thrust (top left), cosine of the thrust-axis polar angle (top right) and cosine of the angle between the signal-B and ROE thrust axes (bottom). The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

Sphericity is an observable commonly used together with thrust and it is exploited also in this analysis. The concept of sphericity in e^+e^- annihilation processes was introduced by J.D. Bjorken and S.J. Brodsky in

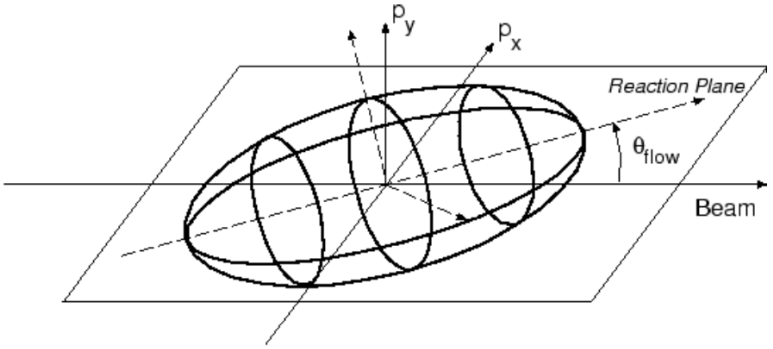


Figure 5.11: Three-dimensional representation of the sphericity ellipsoid.

the 1970s [195]. Given the momenta, $\vec{p}_i$, of the final state particles in the event, the sphericity tensor S is defined as

$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}, \tag{5.19}$$

with $\alpha, \beta = x, y, z$, and it provides the best three-dimensional representation of the spatial distribution of the momenta $\vec{p}_i$. An illustration of the sphericity ellipsoid is shown in Fig. 5.11. For momenta isotropically distributed in space, as in the case of $B\bar{B}$ events, the three eigenvalues λ_k of the tensor have similar magnitudes. For a planar distribution, one of the eigenvalues is much smaller than the others and its eigenvector is orthogonal to the event plane. In the case of momenta collimated along a specific direction, as in continuum events, the eigenvector oriented in that direction has the largest eigenvalue.
The specific variable used in this analysis is the sphericity scalar, defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3), \tag{5.20}$$

where λ_2, λ_3 are the smallest eigenvalues. Values of S close to one characterise events with final-state momenta isotropically distributed, while jet-like events correspond to sphericity values close to zero. The distributions of this variable in signal and background events are shown in Fig. 5.12. Continuum events peak at values of sphericity around 0.15, while events with B^+B^- and $B^0\bar{B}^0$ decays are characterised by larger values of sphericity, on average greater than 0.5. The sphericity distribution computed in signal events is less skewed and presents a peak at intermediate sphericity values, around 0.3, as a consequence of the mixed spatial symmetry of such events.
Another class of event-shape variables are the so-called *harmonic moments*, which correspond to the coefficients of the event expansion in spherical harmonics around an arbitrary axis. The harmonic moments are defined as

$$B_l = \sum_i \frac{|\vec{p}_i|}{\sqrt{s}} P_l\left(\cos\alpha_i\right), \tag{5.21}$$

where α_i is the angle between the momentum of the particle i and the expansion axis, and

$$P_l(x) = \begin{cases} l=0: & 1 \\ l=1: & x \\ l=2: & \frac{1}{2}(3x^2-1) \\ l=3: & \frac{1}{2}(5x^3-3x) \\ l=4: & \frac{1}{8}(35x^4-30x^2+3) \\ l=5: & \frac{1}{8}(63x^5-70x^3+15x) \\ l=6: & \frac{1}{16}(231x^6-315x^4+105x^2-5) \\ \ldots & \end{cases} \tag{5.22}$$

corresponds to the Legendre polynomial of order l. This analysis makes use of the harmonic moments B_0 and B_2 computed with respect to the event thrust axis, and their distributions are provided in Fig. 5.13. B_0 is extremely powerful to separate signal events from events of all the background categories. The signal and background distributions are both unimodal, but the peaks are well distanced and the histograms overlap only partially. In particular, the signal distribution is almost symmetric and peaks at $B_0 \sim 0.6$, while the background distribution is left-skewed and peaks at $B_0 \sim 0.8$. The asymmetry in background is mostly driven by events from the light hadronic continuum, characterised on average by larger B_0 values.

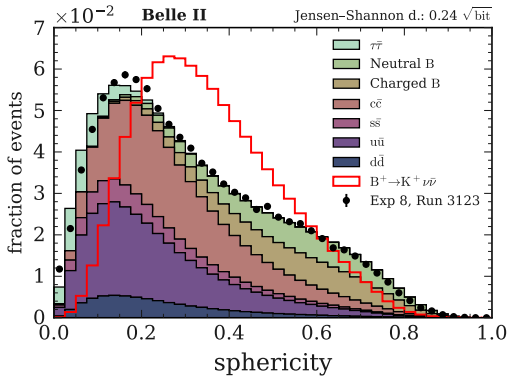


Figure 5.12: Distributions of the event sphericity. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

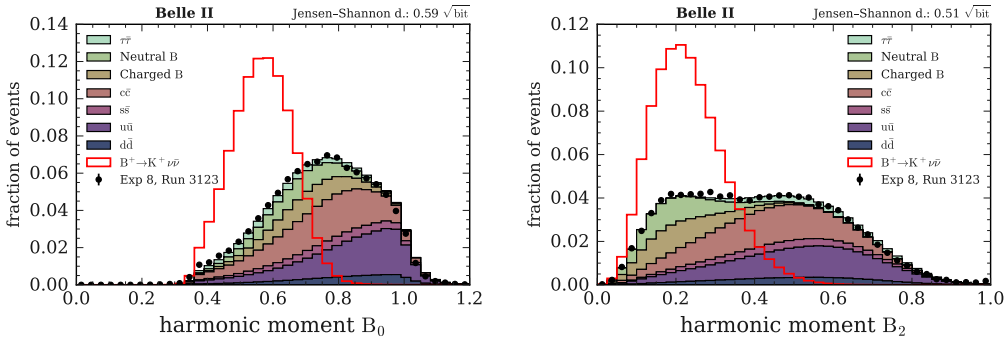


Figure 5.13: Distributions of the harmonic moment B_0 (left) and B_2 (right) with expansion axis corresponding to the event thrust axis. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

The second harmonic moment, B_2, also provides large separation. In this case, the signal distribution is again unimodal with peak at $B_2 \sim 0.2$, while the background distribution is bimodal. In particular, one peak overlaps with the signal distribution due to the low values of B_2 for B^+B^- and $B^0\bar{B}^0$ background events, while the second peak is generated by hadronic continuum events at $B_2 \sim 0.5$. Hence, the second harmonic moment primarily distinguishes signal events from events of the continuum background.

An additional class of powerful variables for the event shape description in e^+e^- annihilation processes are the so-called *Fox-Wolfram moments*, theorised by G. C. Fox and S. Wolfram in the late 1970s [196, 197]. The idea is to construct specific observables, not dependent on the axis choice, which can characterise the shapes of final states by capturing the distributions of momentum and energy flows. To do this, Fox and Wolfram imagined to enclose an e^+e^- event in a sphere, as in Fig. 5.14, and to mark the point at which every produced particle traverses the sphere surface, assigning to it a weight equal to the magnitude of the particle momentum divided by the total center-of-mass energy, s. Using this framework, they defined the set of rotationally invariant observables

$$H_l = \frac{4\pi}{2l+1} \sum_{m=-l}^{+l} \left| \sum_i Y_l^m(\Omega_i) \frac{|\vec{p}_i|}{\sqrt{s}} \right|^2 = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} P_l(\cos\phi_{ij}), \tag{5.23}$$

where p_i indicates the final-state momentum of the particle i, $Y_l^m(\Omega)$ are the usual spherical harmonics, and ϕ_{ij} is the angle between the momenta of the particles i and j. Instead of the simple moments, the event

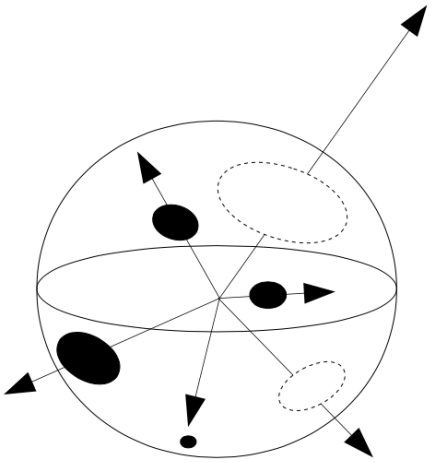


Figure 5.14: Sketch of the Fox-Wolfram parametrisation of the spatial distribution of momenta in the final state of an e^+e^- collision.

shape is always studied using the normalised moments

$$R_k = \frac{H_k}{H_0}, \tag{5.24}$$

defined between 0 and 1. In this study, the first, second, and third normalised Fox-Wolfram moments are used, and their distributions are illustrated in Fig. 5.15. The first normalised Fox-Wolfram moment, R_1, has a large separation power. For all the background categories, the largest faction of the events is characterised by R_1 values smaller than 0.1 and the overall background distribution presents a sharp peak close to zero. Conversely, the signal distribution has a broader peak at $R_1 \sim 0.1$ and a long right tail. For R_2, the background shape changes significantly. The background distribution is bimodal: $B^0\bar{B}^0$ and B^+B^- events are almost entirely concentrated in the interval $0 < R_2 < 0.2$, while events from the hadronic continuum cover R_2 values up to $R_2 \sim 0.9$, with the corresponding distribution peaking at $R_2 \sim 0.4$. $\tau^+\tau^-$ events have the highest R_2 values, reaching approximately $R_2 = 1$. The R_2 distribution of signal events completely overlaps with the $B\bar{B}$ distribution and only partially with the distribution computed in continuum events. Its peak is almost in between the two background peaks. For R_3, the signal and background distributions have similar shapes to R_1, but the separation is reduced. In background events, R_3 is larger than R_1, while it is only slightly larger than R_1 in signal events. Therefore, the overlap is increased.

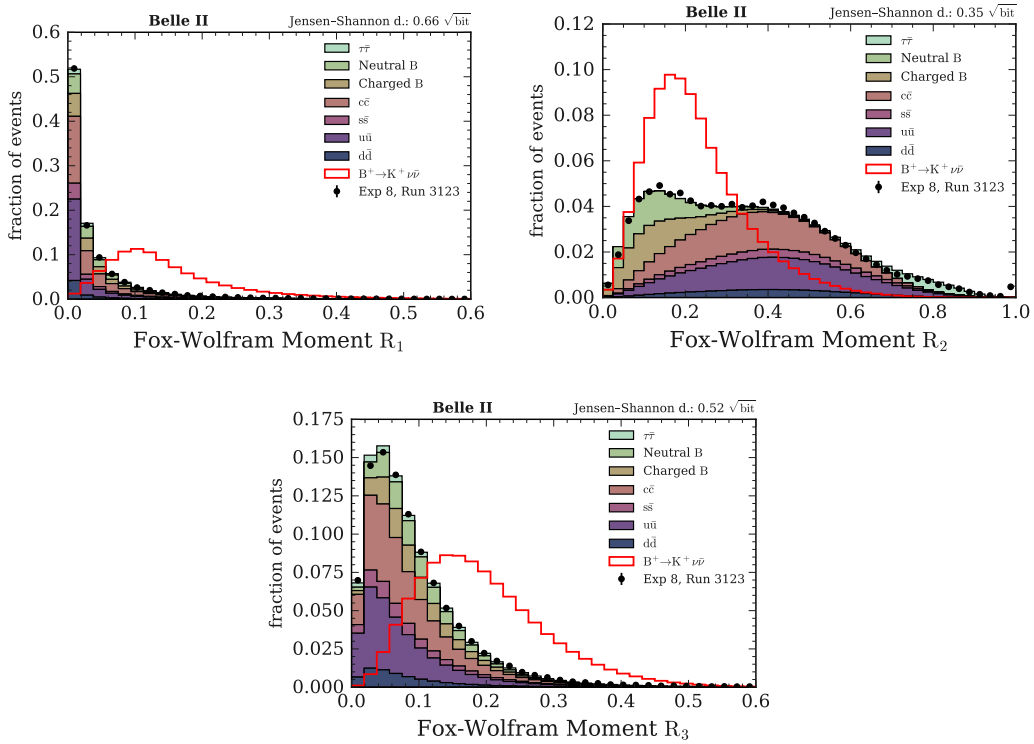


Figure 5.15: Distributions of the first (top left), second (top right), and third (bottom) normalised Fox-Wolfram moments. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

Another class of event-shape variables are the so-called *Kakuno-Super-Fox-Wolfram (KSFW) moments* [198]. They represent modified Fox-Wolfram moments in which the particles involved in the decay of the signal B candidate (s), the particles in the ROE (o), and the missing momentum in the event (m) are distinguished. The first type of KSFW moments are the so-called *linear* moments. Given that, in this analysis, the only reconstructed particle in the signal decay is a charged kaon, K^+, the linear moments can be written in the CMS as

$$H^{so}_{i,l} = \sum_j C_{lj} \frac{|\vec{p}_j| P_l\left(cos\theta_{Kj}\right)}{2\left(\sqrt{s} - E^*_K\right)^2}, \tag{5.25}$$

where θ_{Kj} is the angle between the momentum of the signal-kaon candidate and the momentum of the ROE particle j. Depending on the value of the index i, the sum runs over charged ($i = c$) or neutral ($i = n$) particles in the ROE, or it refers to the missing momentum ($i = m$). For even values of l, $C_{lj} = 1$; for odd values of l, $C_{lj} = 0$ if the neutral particles or the missing momentum are considered, otherwise it is equal to the product of the charges of the particle j and of the signal-kaon candidate. In this analysis, five linear KSFW moments have been used: $H^{so}_{c,2}$, $H^{so}_{n,2}$, $H^{so}_{m,0}$, $H^{so}_{m,2}$, $H^{so}_{m,4}$. The distributions of $H^{so}_{c,2}$ in signal and background events are shown in Fig. 5.16. The distributions are both unimodal with peaks close to zero. The signal histogram is more symmetric, while the background histogram presents a long right tail. Similar features are also visible in the distributions of $H^{so}_{n,2}$, also illustrated in Fig. 5.16, which restricts the particles involved in the computation only to the neutrals in the ROE.
The distributions of $H^{so}_{m,0}$, $H^{so}_{m,2}$, $H^{so}_{m,4}$ are shown in the plots of Fig. 5.17. These variables are sensitive to the

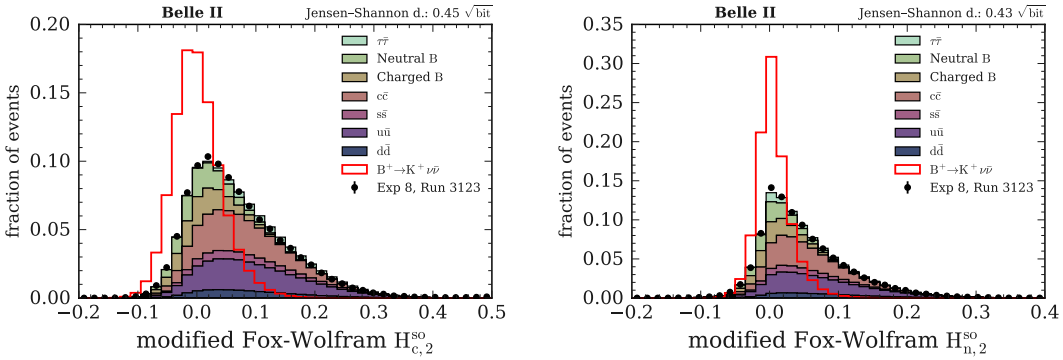


Figure 5.16: Distributions of the linear KSFW moments $H^{so}_{c,2}$ (left) and $H^{so}_{n,2}$ (right). The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

contribution of the missing momentum in the event. For $H^{so}_{m,0}$, $H^{so}_{m,2}$ the signal and background distributions are unimodal. In both cases, the background distributions peak at values close to zero (around zero for $H^{so}_{m,2}$), while the signal distributions have peaks at approximately 0.15. The signal and background peaks can be resolved, providing sizeable separation. In $H^{so}_{m,4}$, the divergence between signal and background is accentuated. The background distribution is unimodal, peaking around zero also in this case, but the signal distribution can be considered bimodal, with a major peak at $H^{so}_{m,4} \sim 0.1$ and a prominent shoulder at negative values of $H^{so}_{m,4}$ close to zero.
The second type of KSFW moments are called *quadratic* moments, and in this analysis they correspond to the variables

$$R^{oo}_l = \sum_i \sum_j C_{lij} \frac{|\vec{p}_i||\vec{p}_j| P_l\,(cos\theta_{ij})}{2\,(\sqrt{s} - E^*_K)^2}, \tag{5.26}$$

which are computed in the CMS using all the ROE particles. For even values of l, $C_{lij} = 1$; for odd l, C_{lij} is zero if either i or j identifies a neutral ROE particle, otherwise it corresponds to the product of the particle charges. R^{oo}_o and R^{oo}_2 are the quadratic KSFW moments used in the analysis, and their distributions are shown in Fig. 5.18. The signal and background R^{oo}_o distributions are unimodal, with distanced peaks. The signal histogram peaks at $R^{oo}_o \sim 0.05$ and has a moderate right tail. The background histogram peaks at $R^{oo}_o \sim 0.1$. It presents a characteristic shape, resulting from the combination of the different features of the R^{oo}_0 distributions in the $B\bar{B}$ backgrounds and in the hadronic continuum. For R^{oo}_2, the peaks of the signal and background distributions overlap around $R^{oo}_2 \sim 0$. Both the distributions have a right tail, but the background tail is longer than the signal tail, which rapidly drops to zero. This feature helps to separate signal from background.

5.7.3 Variables related to the signal-kaon candidate

Four variables related to the properties of the signal-kaon candidate are used in the analysis. The reconstructed final state of the $B^+ \to K^+\nu\bar{\nu}$ decay consists only of a track. Therefore, the characterisation of the signal decay relies on observables describing the track kinematics. The longitudinal and radial distances, dr and dz, from the track POCA to the average IP position are exploited for the purpose. The corresponding distributions are shown in Fig. 5.19. The distributions in dr are similar for signal and background events. Both histograms peak at $dr < 10\,\mu$m and are skewed to the right. Examining the features of the individual background categories, it can be noticed that candidate signal tracks from the light hadronic continuum, $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, are on average closer to the IP, since these tracks are produced promptly in the interaction region in the e^+e^- annihilation. This is also visible in the dz background histograms. For this variable, the

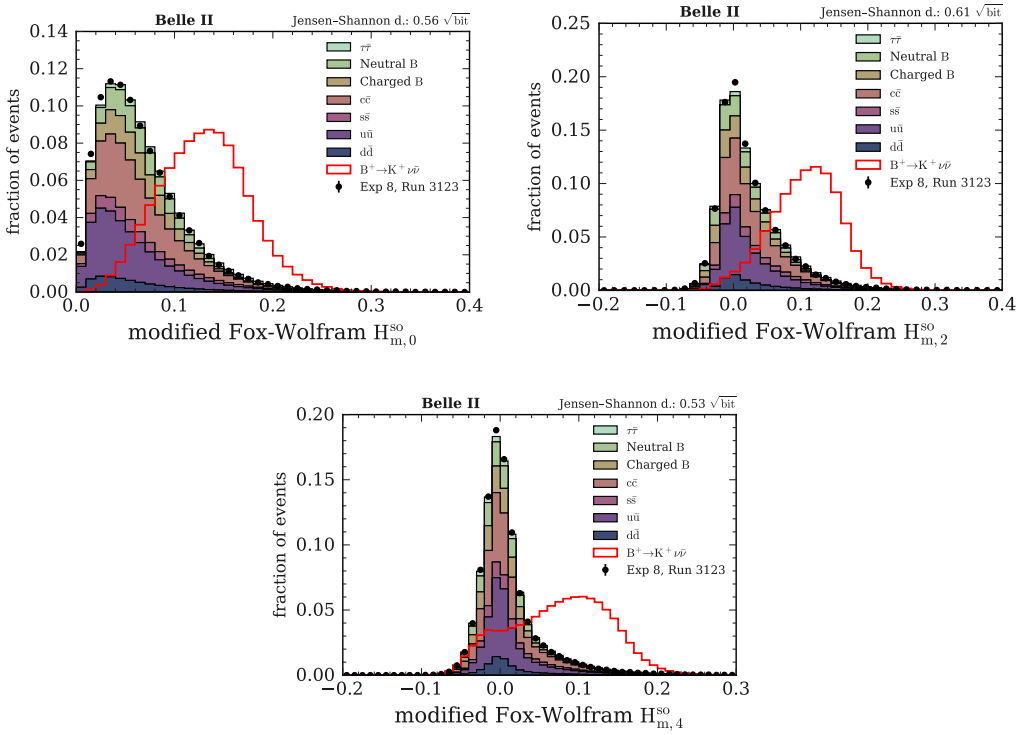


Figure 5.17: Distributions of the linear KSFW moments $H^{so}_{m,0}$ (top left), $H^{so}_{m,2}$ (top right), and $H^{so}_{m,4}$ (bottom). The first index, m, indicates that only the missing momentum is considered. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

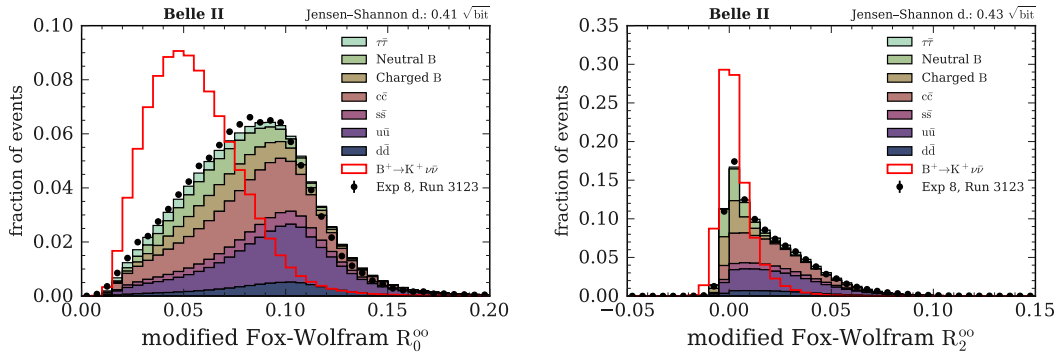


Figure 5.18: Distributions of the quadratic KSFW moments R^{oo}_0 (left) and R^{oo}_2 (right). The superscript oo indicates that all the ROE particles are considered. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

signal and the overall background distributions are both unimodal and symmetric, with background peaking around zero and signal slightly shifted towards the positive dz values, along the direction of the Lorentz boost. Two angular variables are also used to characterise the signal-kaon candidate: the azimuthal angle ϕ_0 and the cosine of the polar angle θ evaluated at the POCA. Their distributions are illustrated in the plots of Fig. 5.20.
The variables related to the properties of the signal-kaon candidate have larger separation power in the second stage of the multivariate classification. More details are given in Chap. 6.

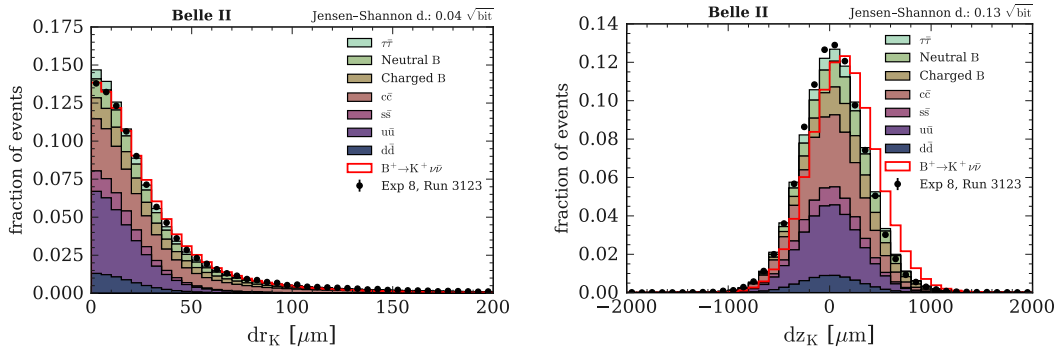


Figure 5.19: Distributions of dr (left) and dz (right) of the signal-kaon candidate track. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

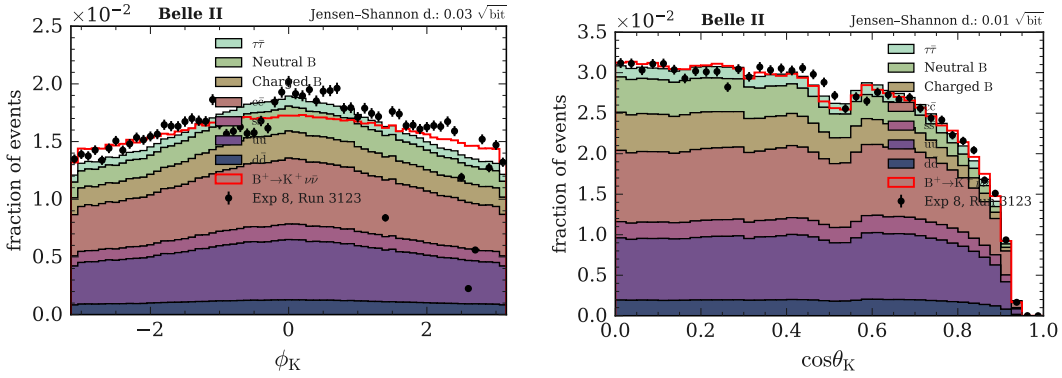


Figure 5.20: Distributions of the azimuthal angle ϕ_0(left) and of the cosine of the polar angle θ (right) at the POCA of the signal-kaon candidate track. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

5.7.4 Variables related to tracks and energy deposits in the ROE

Six variables related to tracks and energy deposits in the ROE are used in the analysis. The invariant mass M of the ROE and the ROE ΔE – where ΔE is the observable defined in Eq. 5.10 – are two fundamental

observables characterising the ROE kinematics. Their distributions are illustrated in Fig. 5.21. In signal events, the ROE is made only of tracks and energy clusters produced in the decay of the tag B meson. If all the products of the tag B-meson decay are reconstructed and are in the ROE acceptance, and if the particle hypotheses in the ROE are correctly assigned, the ROE ΔE in signal events is expected to peak at zero. Similarly, the invariant mass of the ROE is expected to average to the mass of the charged B meson, $5.279\,\mathrm{GeV}/c^2$ [43]. This only happens for a fraction of the events, and the resulting signal distributions of the reconstructed M and ΔE are shifted with respect to the expected mean values. The shifts are mostly caused by the cleanups defining the ROE acceptance, but also by the contribution of semileptonic tag B decays producing neutrinos and hadronic decays with K_L^0 in the final state. In the case of background, the ROE can be made of the combination of tracks and energy clusters of the two B-meson decays or of the several decay products produced in continuum events. This results in an average ROE invariant mass larger than in signal and in an average ROE ΔE greater than zero. Hence, the background peaks in M and ΔE are well distanced and separated from the respective signal peaks, with background and signal distributions only partially overlapping. Both variables have large separation power.

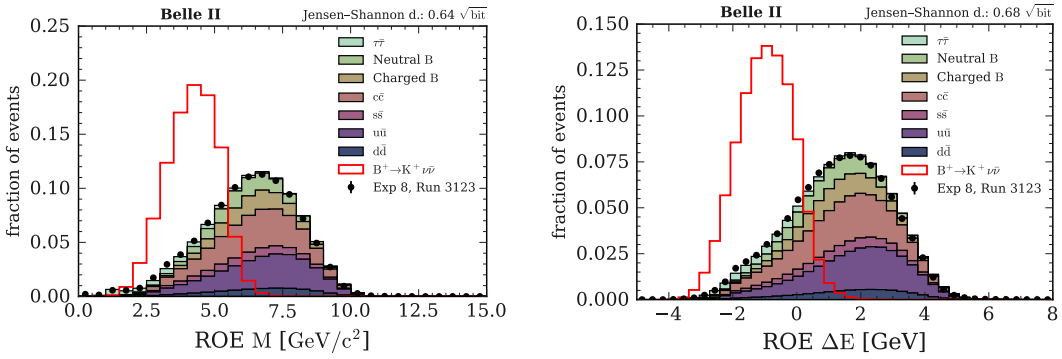


Figure 5.21: Distributions of the ROE invariant mass M (left) and of the ROE ΔE (right). The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

Another powerful ROE variable is the magnitude of the total ROE momentum, $\vec{P}$, shown in the left plot of Fig. 5.22. Both the signal and background distributions are unimodal and approximately symmetric. The signal histogram peaks around $P \sim 1.5\,\mathrm{GeV}/c$ and its full width at half maximum (FWHM) is approximately $1\,\mathrm{GeV}/c$. The background histogram peaks at $P \sim 2.5\,\mathrm{GeV}/c$, meaning that the ROE in background events has on average an extra $1\,\mathrm{GeV}/c$ momentum due to the higher track-multiplicity. The background distribution is also wider and the FWHM is approximately $2\,\mathrm{GeV}/c$. The discrepancy between the signal and background distributions is sizeable. The polar angle $\theta(\vec{P})$ is also used and the corresponding signal and background distributions are illustrated in the right plot in Fig. 5.22. The signal and background histograms have a similar shape, unimodal with a moderate right tail and peaks at values of $\theta(\vec{P})$ smaller than $0.5\,\mathrm{rad}$. The signal distribution is slightly shifted to smaller polar angles with respect to background, and the two histograms do not entirely overlap.
Moreover, the magnitude of the ROE thrust computed in the CMS is used, as it allows characterising the spatial symmetry of the ROE tracks momenta. The signal and background distributions of the ROE thrust are shown in Fig. 5.23. The signal histogram is unimodal with a peak at approximately 0.7 and a short right tail, which makes the distribution asymmetric. It has FWHM ~ 0.2. The background distribution peaks at ~ 0.8, but it is wider than signal. The FWHM is approximately equal to 0.3. This is the result of the different topologies of $B\bar{B}$ and continuum backgrounds. The ROE in B^+B^- and $B^0\bar{B}^0$ events is made of tracks from the two B decays, which decay almost at rest in the CMS, thus producing tracks isotropically distributed

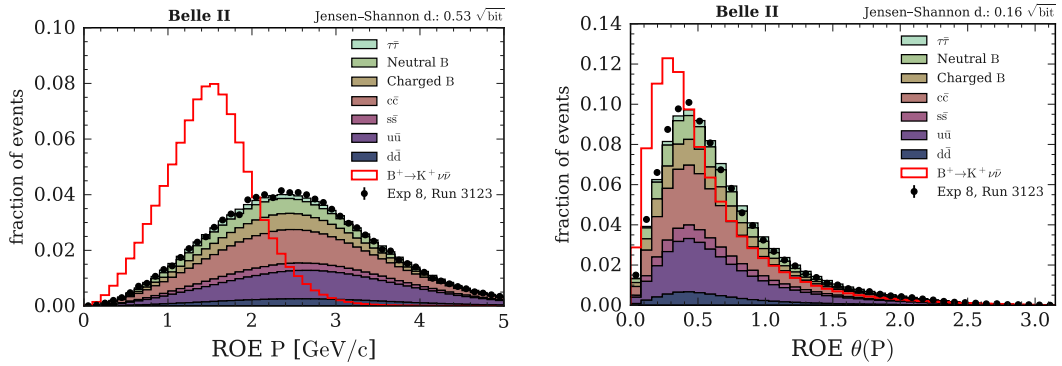


Figure 5.22: Distributions of the magnitude of the ROE total momentum $\vec{P}$ (left) and of its polar angle $\theta(\vec{P})$ (right). The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

in space. Hence, there is not a natural symmetry axis in those events, and the related ROE thrust is low. The $B\bar{B}$ events mostly populate the lower range of the background distribution, below thrust values of 0.8. Conversely, the collimated di-jet symmetry of continuum background events defines a preferred symmetry axis, resulting in higher values of the ROE thrust, as revealed by the stacked continuum histograms peaking around ~ 0.84.
Another variable chosen to characterise the kinematics of the ROE is the variance of the transverse momenta of the ROE tracks, defined as $\sum_{i=1}^{n}(p_{\mathrm{T},i} - \overline{p}_{\mathrm{T}})^2/n$, where i runs over the n ROE tracks and $\overline{p}_{\mathrm{T}}$ is the average ROE transverse momentum. The distributions of the variable computed in reconstructed signal and background events are also shown in Fig 5.23.

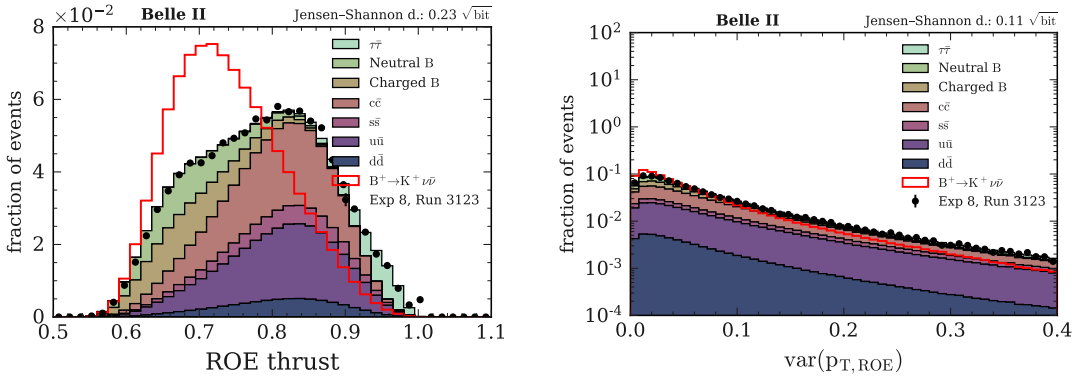


Figure 5.23: Distributions of the ROE thrust computed in the CMS (left) and of the variance of the ROE tracks transverse momenta (right). The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

5.7.5 Variables related to the ROE vertex

Seven variables related to the ROE vertex are used in the analysis. The variables characterise the fit quality and the position of the ROE vertex. In the analysis, the ROE vertex-fit implements the algorithm introduced in Sec. 2.8.3.1, and it uses `KFit` [176] for the kinematic fit. The variables chosen as indicators of the fit quality are the χ^2 and the p-value of the vertex fit, where the latter corresponds to a χ^2 probability. The distributions of the two variables are provided in the plots in Fig. 5.24. In signal events the ROE vertex is made of tracks originating from a real B vertex, corresponding to the decay vertex of the tag B meson, thus the fit quality is higher than in background events, where the ROE tracks originate from two or more decay vertices.

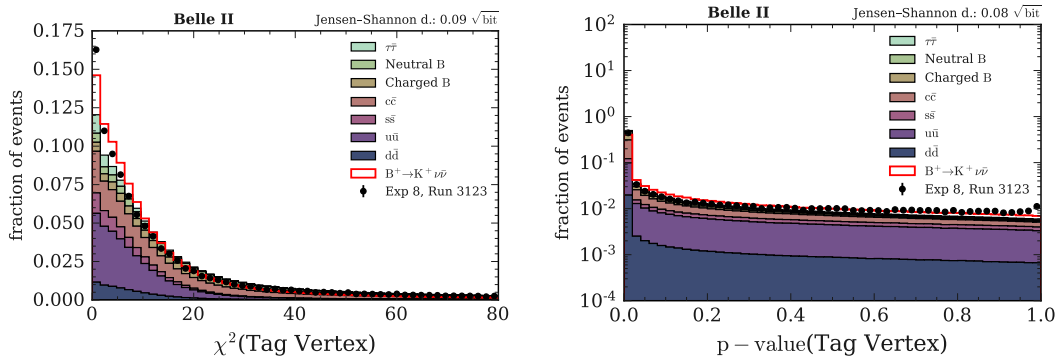


Figure 5.24: Distributions of the χ^2 (left) and p-value (χ^2 probability, right) of the ROE vertex fit. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

The estimated x, y, z coordinates of the ROE vertex, measured with respect to the average IP position, are shown in Fig. 5.25. For all the coordinates, the distributions are unimodal and approximately symmetric both in signal and background. The x, y, z distributions in background peak at values close to zero. The signal distributions for the x and z coordinates are shifted towards positive values. The shift is larger for the z distribution, which has its peak at $\sim 200\,\mu$m. For the y coordinate, the signal peak stays around zero as in background, but it is wider.
Two additional variables are the radial and longitudinal distances between the position of the fitted ROE vertex and the POCA of the signal-candidate track: $dr_K(\text{Tag Vertex})$ and $dz_K(\text{Tag Vertex})$. The distributions of these two variables are also illustrated in Fig. 5.25. The background distributions are characterised by distances on average smaller than in signal. The peaks of the background distributions are produced predominantly by events from the hadronic continuum backgrounds, in which the signal track and the ROE tracks are all produced promptly in the proximity of the interaction region, thus the longitudinal and radial distances are smaller. The signal distributions are broader and present longer tails. In signal events, the ROE and the signal track originate from two separate B vertices, well distanced in space, $\mathcal{O}(100\,\mu\text{m})$, in the laboratory frame.

5.7.6 Variables related to D^0/D^+ suppression

Ten variables for D^0/D^+ background suppression are used in the analysis. Vertices of two and three charged tracks, including the track of the signal-kaon candidate, are reconstructed to identify potential kaons from D^0 and D^+ meson decays, and variables describing the fit quality and the kinematic properties of the vertices are derived. In Sec. 5.4.0.1, a distinction is made between two possible types of D^0 meson candidates: D^0s built with an ROE track for which the most-likely particle hypothesis is assumed, and *simple* D^0 candidates,

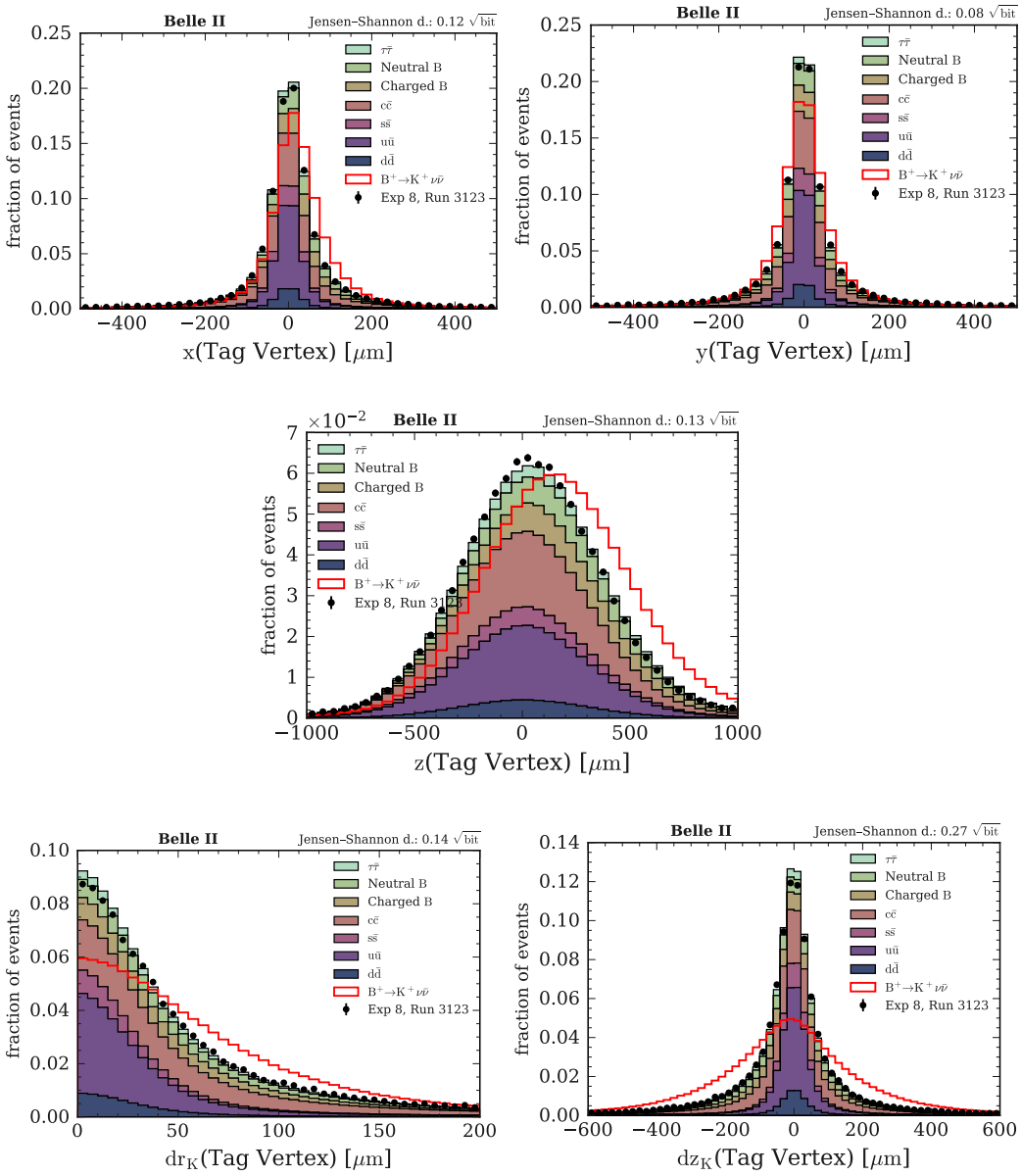


Figure 5.25: Distributions of: estimated x (top left), y (top right), z (centre) coordinates of the ROE vertex position with respect to the average IP, and radial (bottom left) and longitudinal (bottom right) distances between the position of the fitted ROE vertex and the POCA of the signal track. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

where the ROE track is assumed to be a pion. In the following, the candidate D^0s of the latter type, are indicated by the notation D^0_{simple}. Three variables characterising the quality of the D vertex fits are selected to discriminate between signal and background events. These variables are the χ^2 probabilities of the best D^0_{simple} and D^+ vertex fits in an event, and the median χ^2 probability of the vertex fits to all the D^0

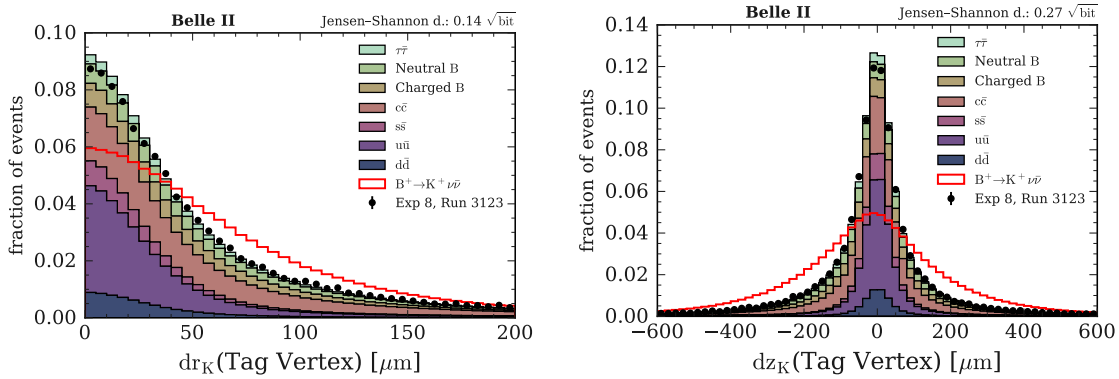


Figure 5.26: Distributions of the radial (left) and longitudinal (right) distances between the position of the fitted ROE vertex and the POCA of the signal track. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

candidates. Their distributions in signal and background events are shown in the plots in Fig. 5.27. The fraction of best D^0_{simple} and D^+ candidates fitted with χ^2 probability close to zero is larger in signal than in background. Conversely, the fraction of best D^0_{simple} and D^+ candidates corresponding to fits with χ^2 probability > 0.5 is larger in background than in signal. Such features are also reproduced in the distributions of the median χ^2 probability of all the D^0 candidates reconstructed per event. It can be noticed, that the fraction of events with a median p-value below 0.1 in signal is significantly larger than in background, revealing the predominance of fake D^0 vertices in signal events. What is observed confirms the intuition behind the reconstruction of D vertices in the inclusive tagging. As expected, more vertices with low fit quality, thus likely fake, are found in signal, where the signal track and the ROE track are expected to originate from two different and displaced vertices. On the other hand, a significant amount of background events is characterised by D^0 and D^+ vertex fits with high χ^2 probability, indicating a fake signal-kaon candidate that likely originates from a true D vertex. These features make the three variables very useful for signal identification and background suppression.
Another powerful variable of this category is the invariant mass of the best D^0 candidate in the event. The distributions of the variable in signal and background events are illustrated in Fig. 5.28. The mass distribution from the combined background sources appears to be significantly different with respect to the one from signal events. It is characterised by localised peaks corresponding to true vertices of two charged tracks. It is possible to identify peaks for: the $\rho(770)^0$ meson, between $0.7\,\text{GeV/c}^2$ and $0.8\,\text{GeV/c}^2$; the $K^*(892)^0$ meson, at $\sim 0.9\,\text{GeV}/c^2$; the $\phi(1020)^0$ meson, at $\sim 1\,\text{GeV}/c^2$; the D^0 meson, at approximately $1.86\,\text{GeV}/c^2$; the $D^*(2007)^0$ meson, between $1.9\,\text{GeV/c}^2$ and $2.0\,\text{GeV/c}^2$. Conversely, the signal distribution is smoother and does not present resonances, since the majority of the vertex candidates are only combinations of oppositely charged tracks that do not originate from the decay of the same meson.
The distances of the D vertices from the average IP position are also used to separate signal and background events. The distributions of the radial, dr, and longitudinal, dz, distances of the best D^0, D^0_{simple} and D^+ vertices are shown in the plots in Fig. 5.29. The features characterising dr and dz are common to the D^0, D^0_{simple}, D^+ candidate vertices. In general, it can be noticed that the signal and background dr distributions are unimodal and skewed to the right. The signal peaks are in the interval approximately between $30\,\mu\text{m}$ and $40\,\mu\text{m}$. In all three cases, the right tail of the distribution is more populated in signal, meaning that the best D^0, D^0_{simple}, D^+ vertices reconstructed in signal events have on average an estimated position that is radially more distant from the IP than in background. The signal and background distributions in dz have the same shape, unimodal and symmetric. The signal histograms are always shifted towards positive

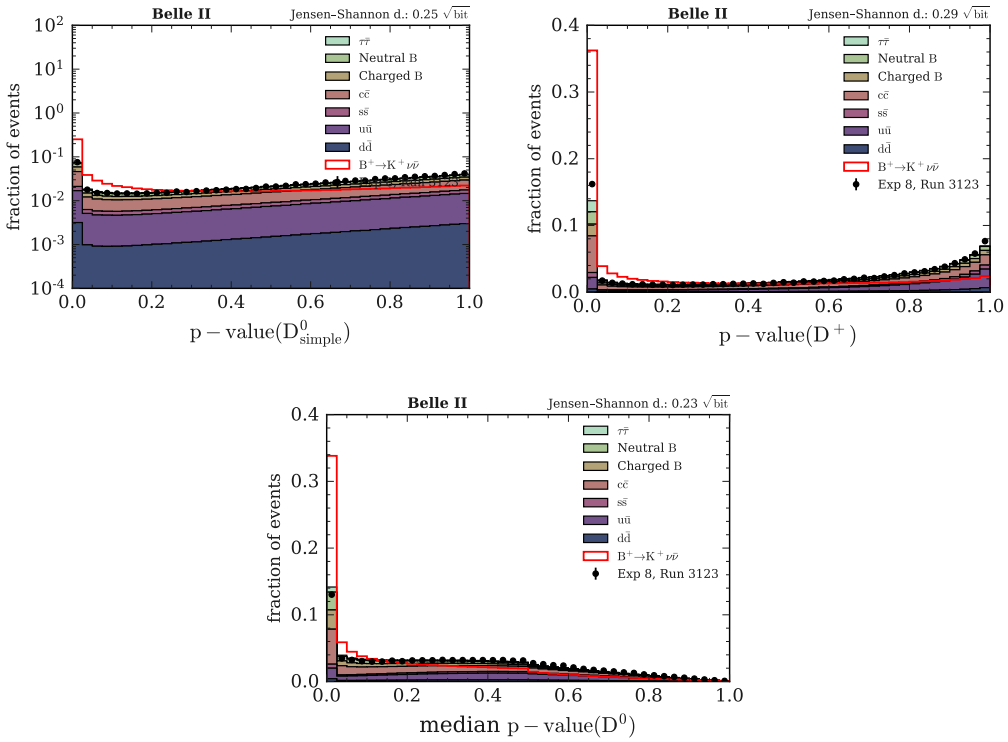


Figure 5.27: Top: distributions of the χ^2 probabilities of the vertex fits to the best D^0_{simple} (left) and D^+ (right) candidates in the event. Bottom: distributions of the median χ^2 probability of the vertex fits to all the D^0 candidates in the event. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

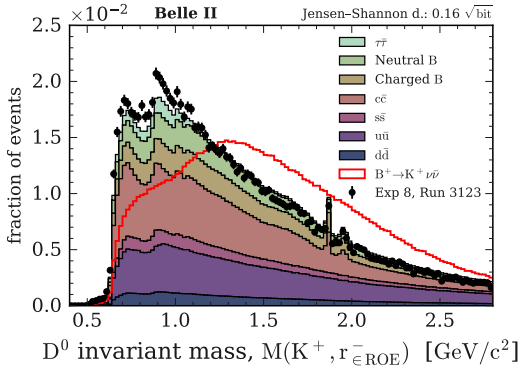


Figure 5.28: Distributions of the invariant mass of the best D^0 candidate in the event. The red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

values of dz with respect to background, thus reducing the overlap between the distributions and making the two peaks partially resolvable. This feature increases the separation power of the dz variables.

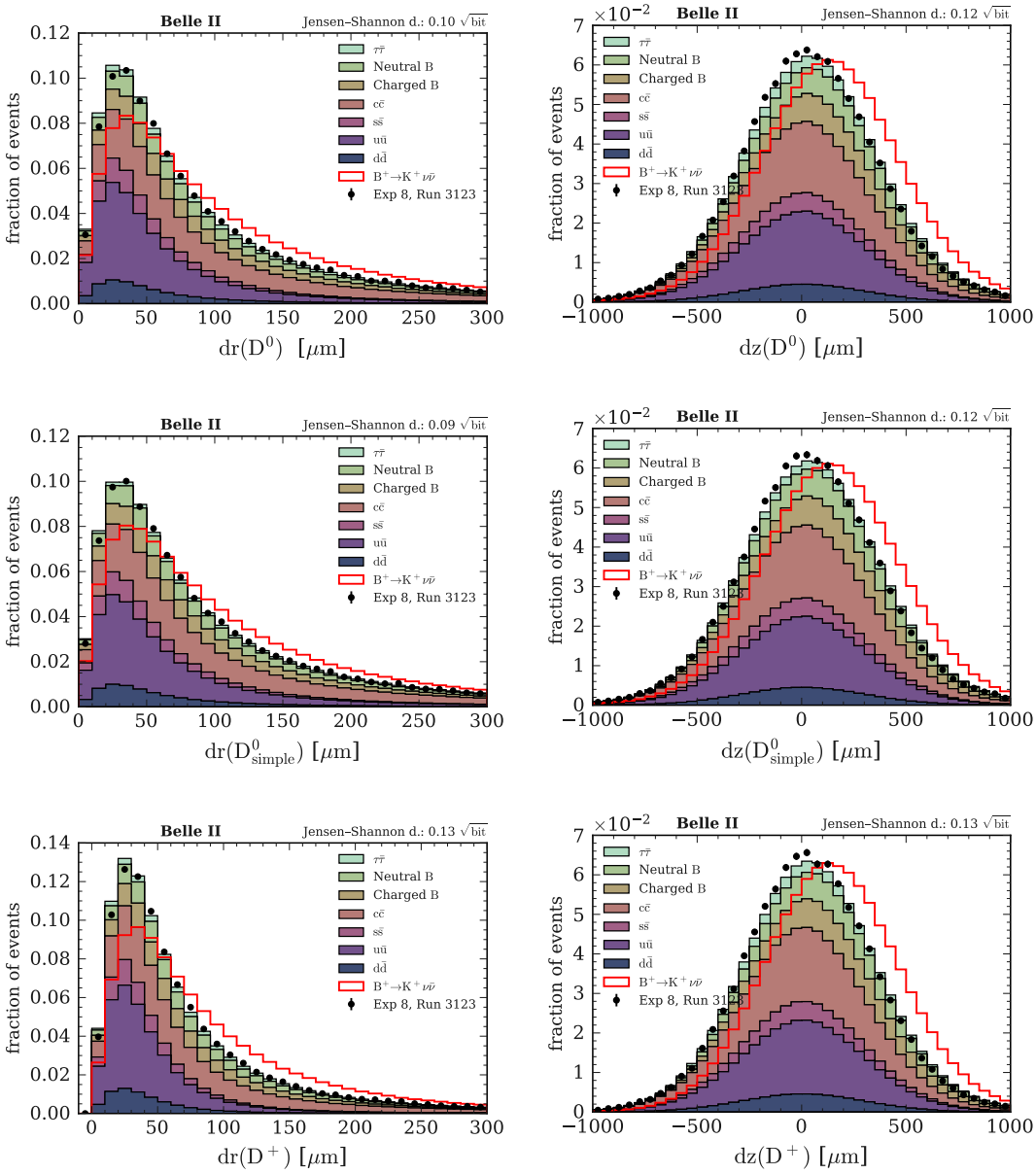


Figure 5.29: Top: distributions of the radial (left) and longitudinal (right) distances from the best D^0 vertex in the event to the average IP. Centre: distributions of dr (left) and dz of the best D^0_{simple} vertex in the event. Bottom: distributions of dr and dz of the best D^+ vertex in the event. In all the plots, the red step histogram originates from true $B^+ \to K^+\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histogram is provided.

Chapter 6

Multivariate classification

This chapter describes the multivariate binary classification procedure implemented in the analysis. Section 6.1 provides an overview of the method. A detailed description of the training samples is given in Section 6.2. The classification model is presented in Section 6.3. Section 6.4 illustrates the importance of the features used to train the classifiers. The classification performance is discussed in Section 6.5.

6.1 Overview

In the analysis, signal identification relies on supervised multivariate classification. Two binary classifiers, used in series, are trained to separate signal from background using reconstructed events from simulation. Both the classifiers implement the `FastBDT` algorithm [186] introduced in Sec. 3.6.2. The first classifier, BDT_1, is trained on 1.6×10^6 simulated events of each of the seven background categories and on the same number of signal events. The set of features used to train BDT_1 consists of the 51 variables described in detail in Sec. 5.7. The first classifier exploits variables characterising the event shape and kinematics, and it allows to identify and discard a large fraction of background events that are not signal-like. To improve the separation performance in the region where it is most likely to find the signal, the second classifier, BDT_2, is trained with the same set of input variables as BDT_1 using reconstructed signal and background events having BDT_1 outputs greater than 0.90. The application of BDT_2 on top of BDT_1 leads to an increase of 35% in signal purity compared to the classification performance of BDT_1 only.

6.2 Training samples

The first classifier, BDT_1, is trained on a sample made of 1.6×10^6 reconstructed events from simulated signal and 1.6×10^6 reconstructed events from each of the seven simulated background categories. Each background event in the training sample is weighted according to the cross section (see Table 2.1) and the reconstruction efficiency of the corresponding background process. The weight assigned to an event of the background category c is

$$w_c = \frac{s}{b_c} \cdot \frac{\sigma_c \cdot \frac{b_c}{B_c}}{\sum_c \left(\sigma_c \cdot \frac{b_c}{B_c} \right)}, \tag{6.1}$$

where B_c and b_c are the number of simulated background events before and after the reconstruction, respectively; s is the number of reconstructed true signal events; σ_c is the cross section of the simulated process. The only weights, $w_{\mathrm{signal}}(q^2)$, applied to signal events (always used in this analysis) are those that allow scaling the initial phase space simulation to the SM expectation, as described in detail in Sec. 4.2.1. To keep signal and background equally balanced in the training sample, such that the total number of signal

and background events are the same, the weights are normalised by requiring that

$$\sum_{c}\sum_{i=1}^{b_c} w_{i,c} = \sum_{i=1}^{s} w_{i,\text{signal}}(q^2). \tag{6.2}$$

The composition of the sample used to train the second classifier, BDT_2, is the following:

- 5×10^6 background events with BDT_1 output > 0.9, reconstructed in a background simulation sample corresponding to an equivalent integrated luminosity of $100\,\text{fb}^{-1}$;
- true $B^+ \to K^+\nu\bar{\nu}$ events with $\text{BDT}_1 > 0.9$ in a sample of 1.6×10^6 reconstructed signal events.

In this case, the background events in the training sample are not weighted using Eq. 6.1, since they originate from a common MC batch in which backgrounds are simulated in the correct proportions according to the cross sections of the seven dominant background processes. Signal events are scaled to the SM expectation by applying the weights $w_{\text{signal}}(q^2)$. Moreover, the signal-to-background ratio in the BDT_2 training sample is required to be equal to one.

6.3 Classification model

A `FastBDT` classification model is defined in terms of the following set of hyperparameters:

- the number of trees (default value $= 100$);
- the depth of each tree (default value $= 3$);
- the number of equal-frequency bins per feature (default value $= 8$);
- the shrinkage or learning rate of the model (default value $= 0.1$);
- the sampling rate, corresponding to the fraction of events randomly drawn from the training sample at each boosting iteration (default value $= 0.5$).

In the analysis, the same classification model is used for BDT_1 and BDT_2. The hyperparameters of this model are tuned by means of a grid-search optimisation performed in the space of the hyperparameters. The chosen hyperparameters values are reported in Table 6.1.

Hyperparameter	Value
Number of trees	2000
Tree depth	4
Shrinkage	0.1
Sampling rate	0.8
Number of equal-frequency bins	16

Table 6.1: Optimised hyperparameters of the classification model

6.4 Feature importances

In a binary classification model, the *importance* of a feature quantifies the influence of that feature on the classification. The `FastBDT` algorithm evaluates the feature importances by summing up the separation gain of each feature over all the trees and nodes in the model. The separation gain corresponds to the entropy reduction in a feature, and details about its computation in `FastBDT` are given in Ref. [186].
The ten most important features for BDT_1 and BDT_2 are listed in Table 6.2 and Table 6.3, together with the

Training variable	Importance score
First normalised Fox-Wolfram moment R_1 computed in the CMS	0.323
ΔE of the ROE	0.167
Linear Fox-Wolfram moment $H^{so}_{m,2}$ computed in the CMS	0.071
Zeroth-order harmonic moment B_0 with respect to the thrust axis in the CMS	0.061
Normalised Fox-Wolfram moment R_1 computed in the CMS	0.051
Magnitude of the ROE momentum	0.038
Normalised Fox-Wolfram moment R_2 computed in the CMS	0.028
Squared missing invariant mass in the event	0.024
Invariant mass of the ROE	0.024
Linear Fox-Wolfram $H^{so}_{m,4}$ moment computed in the CMS	0.023

Table 6.2: List of the ten most important training variables for the BDT_1 classifier ranked by importance score.

Training variable	Importance score
Linear Fox-Wolfram moment $H^{so}_{m,2}$ computed in the CMS	0.118
ΔE of the ROE	0.053
p-value of the ROE vertex fit	0.050
Event sphericity computed in the CMS	0.039
Linear Fox-Wolfram moment $H^{so}_{c,2}$ computed in the CMS	0.039
Linear Fox-Wolfram moment $H^{so}_{m,4}$ computed in the CMS	0.035
Number of charged lepton candidates in the event	0.034
dz of the signal-kaon candidate track from the ROE vertex	0.034
χ^2-probability of the vertex fit to the best D^+candidate	0.033
Median χ^2-probability of the vertex fits to all the D^0 candidates	0.033

Table 6.3: List of the ten most important training variables for the BDT_2 classifier ranked by importance score.

scores quantifying their importances. It is interesting to discuss the composition of the two sets of features reported in the tables. Of the ten most important variables for BDT_1, seven are related to the event shape and kinematics, and three to the kinematics of the ROE. In particular, the two most important features for BDT_1 are the normalised Fox-Wolfram moment R_1 and the ΔE of the ROE. The importances of these two variables are approximately 0.32 and 0.17. The importance scores of the remaining seven features range from approximately 0.07 to 0.02. This means that R_1 and the ROE ΔE provide a major contribution to the classification performed by BDT_1. The list of the ten most important features for BDT_2 is more diversified. There are variables describing the event kinematics and shape, like the linear Fox-Wolfram moments and the event sphericity, but also variables from other categories: variables related to the ROE vertex, variables related to the suppression of D^0/D^+ mesons, and the number of charged lepton candidates in the event, which targets events with semileptonic decays. In particular, Table 6.3 shows that the most important feature is the linear Fox-Wolfram moment $H^{so}_{m,2}$, which depends on the missing momentum in the event and has an importance of approximately 0.12. The remaining nine features have importances going from 0.5 to 0.3. This set of variables suggests that BDT_2 tries to separate signal from background events like those with semileptonic $B \to Dl\nu$ decays. These events represent a sizeable source of background at high BDT_2 outputs, as illustrated in Chapter. 7.

6.5 Performance of the classifiers

The classification performance of BDT_1 and BDT_2 is evaluated in the training sample and compared to the performance in an independent test sample equivalent in size and composition. The results are summarised

in the upper panel of Fig. 6.1. The plots show that the overfitting is small and under control, and it occurs only at the lowest BDT_1 and BDT_2 outputs. The overlap of the signal and background distributions is larger in BDT_2 than in BDT_1, meaning that BDT_1 has a larger separation power. This is an expected feature. BDT_2 is trained to distinguish signal from background events at $\mathrm{BDT}_1 > 0.9$, where the features of background events are similar to signal.
The signal identification performance of the classifiers is evaluated in terms of signal significance using the figure of merit $S/\sqrt{S+B}$, in which S and B are the expected number of events from signal and background (more details on the estimation of the signal significance are given in Appendix B). The significance is computed together with the signal efficiency as a function of increasing lower bounds on the classifier outputs. This sensitivity study uses a simulated background sample corresponding to an integrated luminosity of $500\,\mathrm{fb}^{-1}$ and 4×10^6 simulated signal events. The reconstructed signal-kaon candidates are required to satisfy the PID selection $\mathrm{kaonID} > 0.9$. The results shown in Fig. 6.1 are scaled to the integrated luminosity of $63\,\mathrm{fb}^{-1}$ of the on-resonance collision data sample used in the analysis. The plots show that the introduction of BDT_2 on top of BDT_1 leads to a $\sim 10\%$ increase of the maximum signal significance, which is obtained at a signal efficiency of $\sim 4\%$. In particular, the significance goes up to 50% and the signal purity is increased by 35%. This result proves that the classifier cascade, set up with BDT_2 trained on the classification results of BDT_1, produces an actual boost of the sensitivity to the signal and is beneficial to the search.

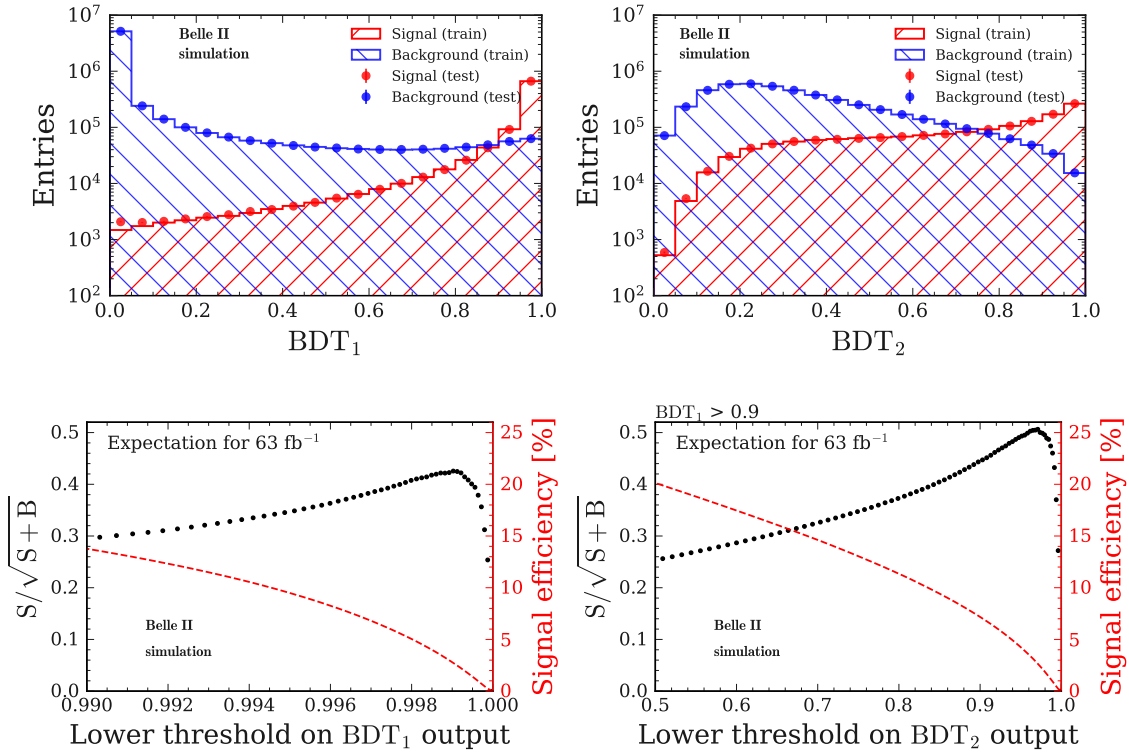


Figure 6.1: Top: distributions of BDT_1 (left) and BDT_2 (right) outputs for signal in the training (red step histogram) and test (red dots) samples, and for background in the training (blue step histogram) and test (blue dots) samples. Bottom: statistical significance of signal (black dots) as a function of the lower bound on the BDT_1 (left) and BDT_2 (right) outputs. The significance is quantified by the figure of merit $S/\sqrt{S+B}$, in which S and B are the expected number of signal and background events. It is computed by reconstructing a simulated background sample equivalent to an integrated luminosity of $500\,\mathrm{fb}^{-1}$ and 4×10^6 simulated signal events. The results are scaled to the integrated luminosity of $63\,\mathrm{fb}^{-1}$. The red dashed line in the plots shows the signal efficiency as a function of the selection on the classifier output. [2]

Chapter 7

Background composition

This chapter presents the study of the background composition in a region characterised by high sensitivity to the signal. Section 7.1 illustrates the tools and the technique developed to identify the decays occurring in the reconstructed events. In Section 7.2 and Section 7.3, the main decays in the charged and neutral B backgrounds are discussed. Section 7.4 describes the composition and the main kinematic properties of the events in the continuum background.

7.1 Investigation of the background

The goal of the study is to investigate the reconstructed background in a region where the sensitivity to the signal is high. The decays occurring in reconstructed events with $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$ are identified and characterised. This can be done by accessing the generator-level information of the simulated physics processes in the reconstructed events. In particular, it is possible to collect the numeric codes assigned by the Particle Data Group (PDG) [43] to the generated particles. These codes are provided by the event generators for all the stages of the generated decays. The procedure implemented in the analysis to identify the background decays relies on this information and is described in the following.
During the reconstruction of the simulated background, the first particle generated in the event is accessed. This particle corresponds to a $\Upsilon(4\mathrm{S})$ meson in case of $B\bar{B}$ events and to a virtual Z^0 meson in continuum events. For the $B\bar{B}$ events, the PDG codes of the two generated B mesons are stored together with the PDG codes of their decay products. In continuum events only the PDG codes of the particles produced in the simulated hadronisation of the quark pair, or in the prompt decay of the τ leptons, are collected. In the reconstruction of simulated $B\bar{B}$ background events, specific variables defined in the `basf2` software allow to determine whether the reconstructed signal-kaon candidate corresponds to a MC particle originating from the decay of a $B^+(B^0)$ or $B^-(\bar{B}^0)$ meson. This information is used to distinguish the decay chain of the signal B meson from that of the tag B meson.
The PDG codes of the particles involved in a decay are translated into the corresponding literal symbols by means of the Particle package [199], and the strings are then merged to build the decay string identifying the whole process. In the following sections the results of the categorisation of the events reconstructed in $300\,\mathrm{fb}^{-1}$ of simulated background are presented.

7.2 Composition of the charged B background

The signal and tag decays, occurring with a frequency larger than 1% in the reconstructed charged-B background events investigated in the study, are illustrated by the charts in Fig. 7.1. Almost all of the most frequent decays in the charged background present a D^0 or a $D^*(2007)^0$ meson. The large contribution of D^0s is explained by the fact that $B^+ \to D^0 X$ is the major inclusive process for the decay of a charged B meson. Moreover, the inclusive mode $D^0 \to K^+ X$ has a measured branching fraction larger than 50% [43]. The contribution of this major background source is suppressed in the analysis by explicitly reconstructing

candidate D^0 mesons as described in Sec. 5.4.0.1 and Sec. 5.7.6.
Semileptonic decays like $B^+ \to \bar{D}^0\, l^+\nu_l\,(\gamma)$, and $B^+ \to \bar{D}^*(2007)^0\, l^+\nu_l\,(\gamma)$, with $l = e,\ \mu$, represent approximately 44% of the signal decays in the investigated charged background. Such decays are characterised by large branching fractions, 2.3% for $B^+ \to \bar{D}^0 l^+\nu_l$ and 5.6% for $B^+ \to \bar{D}^*(2007)^0 l^+\nu_l$ respectively [43], and the neutrino in their final state is a source of missing energy in the event, which additionally contributes to fake the presence of a signal decay. Such processes also represent $\sim 18\%$ of the decays of the tag B meson. In the signal side, the second most frequent decay is $B^+ \to K^+K^0\bar{K}^0$, contributing to 7% of the investigated

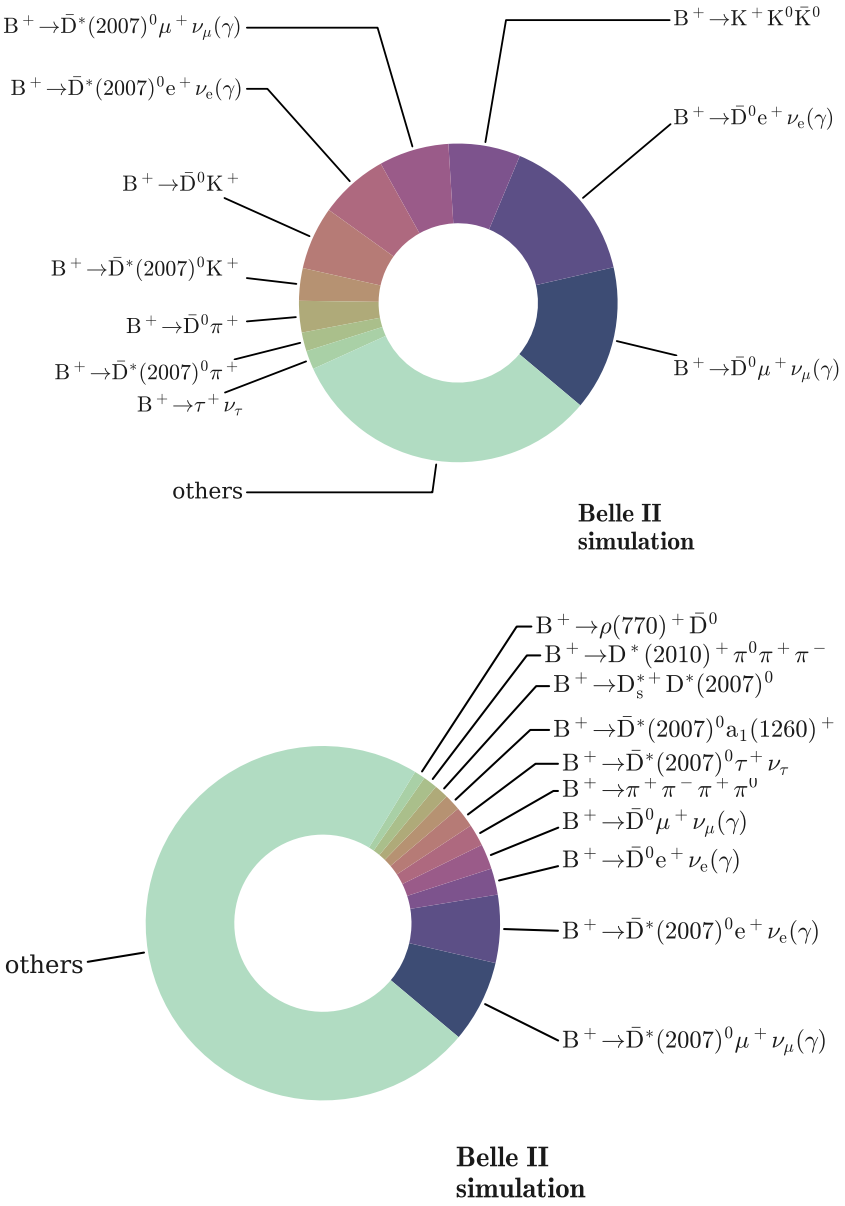


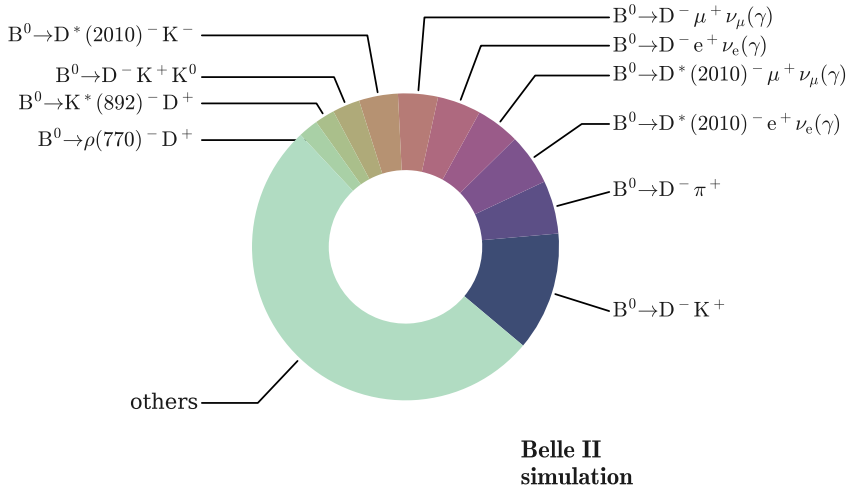
Figure 7.1: Composition of the reconstructed charged B background at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$. The top chart shows the decays occurring in the selected signal side. The bottom chart shows the decays in the tag side. Only contributions above 1% are explicitly reported in the diagrams.

events. The K^+ in this decay is wrongly chosen as a signal kaon, and the presence of the $K^0\bar{K}^0$ pair is a potential source of missing energy in the event. The K^0 meson half of the times corresponds to a K^0_L, which

can escape the detector without being detected, or can easily be misreconstructed. Other hadronic decays, with a D^0 or a $D^*(2007)^0$ meson, amount to approximately 15% of the decays in the signal side. They are also identified in the tag side, where they are present in $\sim 4\%$ of the selected events.
It is interesting to notice that the rare decay $B^+ \to \tau^+\nu_\tau$ ($\mathrm{BR}(\mathrm{B}^+ \to \tau^+\nu_\tau) = (1.09 \pm 0.24) \times 10^{-4}$ [43] occurs in the signal side of $\sim 2\%$ of the charged B events at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$. This decay has been measured at Belle and BaBar with the hadronic and the semi-leptonic taggings [200–202], but this finding suggests the possibility to measure it using the inclusive tagging.

7.3 Composition of the neutral B background

The ten most frequent signal and tag decays in the reconstructed events of the neutral B background at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$ are presented in Fig. 7.2.



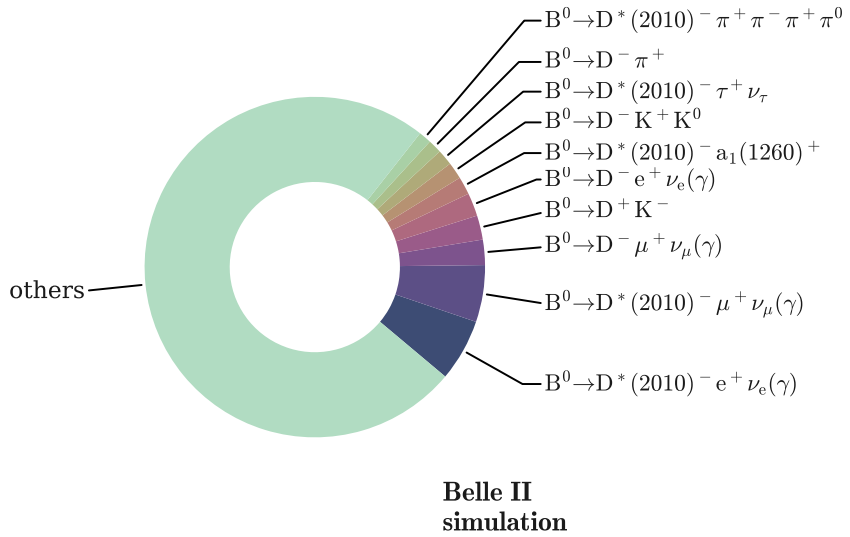


Figure 7.2: Composition of the reconstructed neutral B background at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$. The top chart shows the decays occurring in the selected signal side. The bottom chart shows the decays in the tag side. Only contributions above 1% are explicitly reported in the diagrams.

Both in the signal and tag sides, all the main background decays are characterised by the presence either of a D^+ or a $D^*(2010)^+$ meson. The D^+ meson decays inclusively into a charged kaon with a branching fraction larger than 30% [43], and the decay of a $D^*(2010)^+$ meson produces either a D^0 or a D^+ meson. In the signal side, the dominant decay is $B^0 \to D^- K^+$, with a contribution of 12.5%. Other hadronic decays represent approximately 29% of the signal decays in the selected events. The semileptonic decays $B^0 \to D^- \, l^+ \nu_l \, (\gamma)$, and $B^0 \to D^*(2010)^- \, l^+ \nu_l \, (\gamma)$, with $l = e, \, \mu$, amount to $\sim 19\%$. In the tag side, the same semileptonic decays represent the dominant processes, contributing to approximately 17% of the investigated events. The semileptonic $B^0 \to D^*(2010)^- \tau^+ \nu_\tau$ is present in 1.5% of the selected events. Hadronic processes with a D^+ or a $D^*(2010)^+$ meson sum up to approximately 8% of the tag decays, and the $B^0 \to D^- K^+$ decay, which is the dominant individual contribution in the signal side, contributes to 2.3% of the tag decays.

7.4 Continuum background

The processes occurring in the reconstructed continuum events at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$ are also identified and ranked by frequency. An exemplary summary for the $c\bar{c}$ background is given in Table 7.1.

	continuum: ccbar	frequency (%)
0	$D^*(2010)^+ \pi^0 \pi^- \bar{D}^*(2007)^0$	1.639344
1	$D^*(2007)^0 p \bar{p} \bar{D}^*(2007)^0$	1.092896
2	$D_0^*(2300)^+ \pi^- \bar{D}^*(2007)^0$	1.092896
3	$D^*(2010)^+ \pi^- \bar{D}^*(2007)^0$	1.092896
4	$D^+ \pi^- \omega(782) \bar{D}^*(2007)^0$	1.092896
5	$D_s^{*+} K^- \pi^+ K^0 D_s^{*-}$	1.092896
6	$D^*(2010)^+ \rho(770)^- \pi^+ D^*(2010)^-$	1.092896
7	$D^*(2010)^+ K^0 K^- \eta \pi^0 \bar{D}^*(2007)^0$	0.546448
8	$D^*(2010)^+ \pi^- K^*(892)^+ K^- K^+ D_s^{*-}$	0.546448
9	$D^*(2010)^+ \pi^- \omega(782) \pi^+ D_0^*(2300)^-$	0.546448
10	$D^*(2010)^+ K^*(892)^0 \bar{K}^0 D^*(2010)^-$	0.546448
11	$D^*(2010)^+ \pi^- K^+ \phi(1020) K^- \bar{D}^*(2007)^0$	0.546448
12	$D^*(2010)^+ \pi^- K^+ \bar{K}^*(892)^0 D^*(2010)^-$	0.546448
13	$D^*(2010)^+ \pi^- K^+ K^*(892)^- \pi^+ D^-$	0.546448
14	$D^*(2010)^+ \pi^- K^+ D_s^-$	0.546448
15	$D^*(2010)^+ K^0 K^*(892)^- \bar{D}^0$	0.546448
16	$D^*(2010)^+ K^0 K^*(892)^- \eta \rho(770)^0 \bar{D}_2^*(2460)^0$	0.546448
17	$D^*(2010)^+ \omega(782) \pi^- \bar{D}^*(2007)^0$	0.546448
18	$D^*(2010)^+ \omega(782) K^0 K^- \bar{D}_0^*(2300)^0$	0.546448
19	$D^*(2010)^+ \pi^- \pi^+ \rho(770)^0 D^*(2010)^-$	0.546448

Table 7.1: Summary of the 20 most frequent decays in the reconstructed $c\bar{c}$ continuum background at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$.

In this case, a simple categorisation is not sufficient to infer the main characteristics of the final states in the continuum-background events. Therefore, specific variables have been studied at the generator level in order to characterise the properties of the continuum processes. One of the variables is the true missing energy, which is illustrated in Fig. 7.3. The distributions of the missing energy are all unimodal and peaking at $\sim 4\,\mathrm{GeV}$. The histograms show that the selected events from the hadronic continuum are characterised by an amount of missing energy which approximates the missing energy carried by the neutrino pair in signal decays.

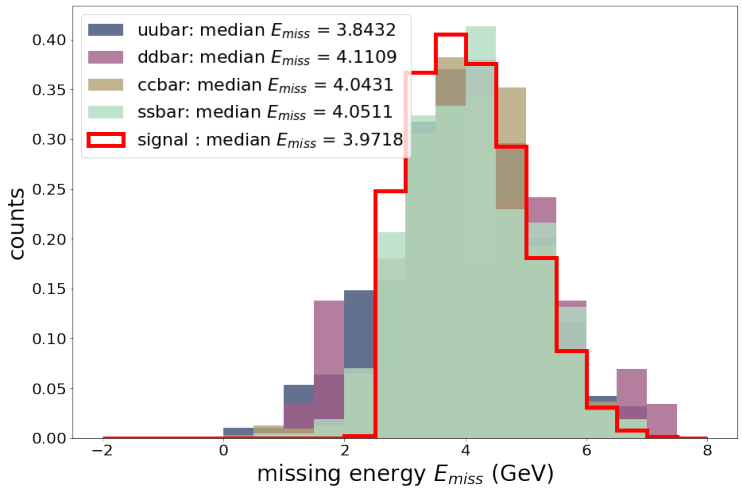


Figure 7.3: Distributions of the missing energy E_{miss} per event for signal (red step histogram) and different sources of hadronic continuum background (filled histograms) at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.93$. Each of the histograms is normalised so that the covered area integrates to unity.

Chapter 8

Validation studies

This chapter presents the validation of the inclusive tagging method. The goal is to test the robustness of the analysis by comparing its performance in simulation and in data with no signal contribution. In particular, the modelling of the variables used in the multivariate classification is checked, and the classification performance is evaluated. Section 8.1 describes the validation with the $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ decay used as an independent control channel. The validation with off-resonance data is discussed in Section 8.2. Finally, Section 8.3 presents the validation studies performed in a sideband region characterised by moderate values of the BDT outputs.

8.1 Validation with $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$

The classification performance of the inclusive tagging relies on the accurate modelling of the variables used in the analysis, therefore an investigation of the data-MC agreement is necessary. The validation with on-resonance data is performed in an independent control channel.
The decay mode chosen for the study is $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$. The $B^+ \to K^+ J/\psi$ decay has a sizeable branching fraction, $\mathrm{BR}(B^+ \to K^+ J/\psi) = (1.020 \pm 0.019) \times 10^{-3}$, and the branching fraction of the $J/\psi \to \mu^+\mu^-$ decay is $\mathrm{BR}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$ [43]. Moreover, the $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ decay can be reconstructed with high efficiency, given its clean experimental signature characterised by the presence of a single kaon track and two muons originating from a common decay vertex.
However, the kinematics and the particle content of $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ are different from the signal decay. In the control mode, two charged leptons are present in the final state instead of two neutrinos, and their contribution can affect the event and ROE-based variables exploited by the inclusive tagging. Moreover, the kinematics of the charged kaon in $B^+ \to K^+ \nu\bar{\nu}$ is different from that of the K^+ in $B^+ \to K^+ J/\psi$. In the first case, the kaon is involved in a three-body decay, while the second is a two-body decay, which is characterised by a different kinematic phase space. Because of these irreducible differences between the signal and control modes, an innovative technique is implemented for the validation. The main steps of the procedure are described below:

- $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events are identified in data and MC by means of tight selection criteria, allowing to collect a high purity sample;
- the tracks of the two muon candidates from $J/\psi \to \mu^+\mu^-$, and the corresponding ECL and KLM clusters, are removed from the event before the reconstruction;
- in the event reconstruction, the four-momentum of the K^+ in an identified $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ decay is replaced with the generator-level four-momentum of the K^+ in a $B^+ \to K^+ \nu\bar{\nu}$ decay randomly selected from simulation;
- the signal kaon candidate is selected among the tracks in the event and the full reconstruction procedure is carried out.

8.1.1 Identification of $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events

The fundamental step in the procedure to identify $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events consists in the selection of $J/\psi \to \mu^+\mu^-$ candidate decays. Muon tracks are required to satisfy the *track cleanup* defined in Sec. 5.2 and the PID requirement $\mathrm{muonID} > 0.5$, where muonID is the PID variable introduced in Sec. 2.8.4. Candidate J/ψ mesons are reconstructed by combining pairs of selected muons, requiring $|M^{\mathrm{PDG}}_{\mathrm{J}/\psi} - M_{\mu^+\mu^-}| < 50\,\mathrm{MeV}/c^2$, where $M^{\mathrm{PDG}}_{\mathrm{J}/\psi} = 3.096\,\mathrm{GeV}/c^2$ is the average J/ψ mass from the PDG [43] and $M_{\mu^+\mu^-}$ is the invariant mass of the muon pair. This invariant-mass selection is imposed to reduce the contamination from the radiative tail of the J/ψ resonance.
The candidate K^+ is selected as the track with highest transverse momentum in the event, satisfying the track cleanup and the PID selection $\mathrm{kaonID} > 0.1$ and $\mathrm{muonID} < 0.5$.
The K^+ and J/ψ candidates are then combined to form a candidate B^+ meson, which is required to have $M_{bc} > 5.25\,\mathrm{GeV}/c^2$ and $|\Delta E| < 100\,\mathrm{MeV}$. The results of the $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ identification in data and simulation are illustrated in Fig. 8.1.

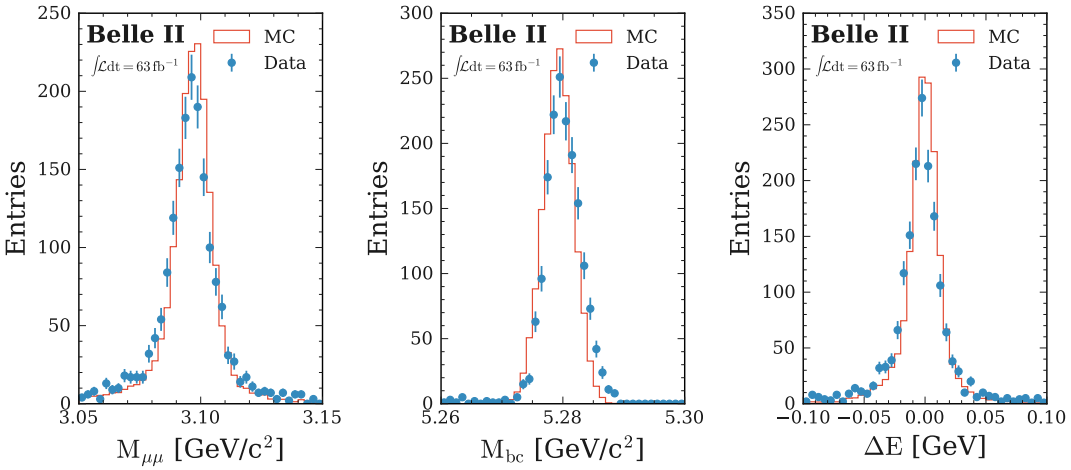


Figure 8.1: $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ candidates selected for the validation study. The distributions of the dimuon invariant mass $M_{\mu^+\mu^-}$ (left plot), beam-constrained mass M_{bc} (central plot) and ΔE (right plot) are compared between data (blue points) and MC (red step histograms). [2]

With this selection, 1720 candidates in 1722 events are reconstructed in the on-resonance data sample. The selection efficiency, evaluated in the MC simulating the decay, is about 46%. Figure 8.1 shows that the overall data-MC agreement is good, with only a slight shift between data and MC distributions at the sub-permille level. Moreover, the selection keeps the background level below 5%. This can be checked by computing the fraction of candidate $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ decays with $M_{bc} > 5.27\,\mathrm{GeV}/c^2$. The low background contribution is ignored in the following tests.

8.1.2 Comparison between $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ and $B^+ \to K^+ \nu\bar{\nu}$

The $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events identified in data and MC are reconstructed with the inclusive tagging algorithm, which computes all the discriminating variables described in Sec. 5.7. The distributions of the variables in $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ and $B^+ \to K^+ \nu\bar{\nu}$ simulated events can be compared to identify the main differences between the signal and control channels. The results are summarised in Figs. 8.2, 8.3, where the distributions of three key observables are illustrated: the charged kaon p_{T}, the event sphericity, and the ROE ΔE. Figure 8.2 shows that the kinematics of the K^+ in $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ is very different from that of the K^+ in $B^+ \to K^+ \nu\bar{\nu}$. In the first case, the charged kaon participates in a two-body decay and the p_{T} (K^+)

distribution has a narrow peak at $\sim 1.5\,\mathrm{GeV}/c$, with a full-width at half maximum smaller than $1\,\mathrm{GeV}/c$ and a long left tail. In contrast, the $p_\mathrm{T}\,(K^+)$ distribution in $B^+ \to K^+\nu\bar{\nu}$, which is a three-body decay, is centred around $p_\mathrm{T} \sim 1.5\,\mathrm{GeV}/c$, but it is more symmetric and much wider, with a FWHM of approximately $2\,\mathrm{GeV}/c$. This is an irreducible difference between the signal and control channels, specifically addressed in the validation study with a modification of the four-momentum of the K^+ in $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$.
The additional difference between the signal and the control channel is the presence of the muon pair instead of the neutrino pair. The muon tracks contribute to the observed discrepancies in variables like sphericity, which depends on the momenta of the tracks in the event. The MC distributions of sphericity in the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events are shown in the left plot of Fig. 8.3. Differences are minor for variables like the ROE ΔE, which has similar distributions in signal and in the control channel, as illustrated in the right plot of Fig. 8.3.
A dedicated procedure to eliminate the irreducible differences between the two modes is presented in the next section.

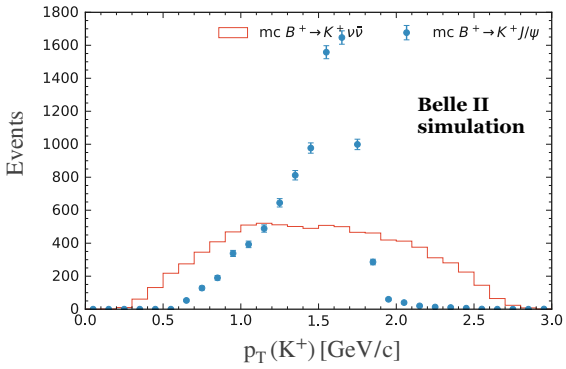


Figure 8.2: Distributions of the transverse momentum of the signal K^+ candidate in simulated $B^+ \to K^+\nu\bar{\nu}$ events (red step histogram) and $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events (blue points).

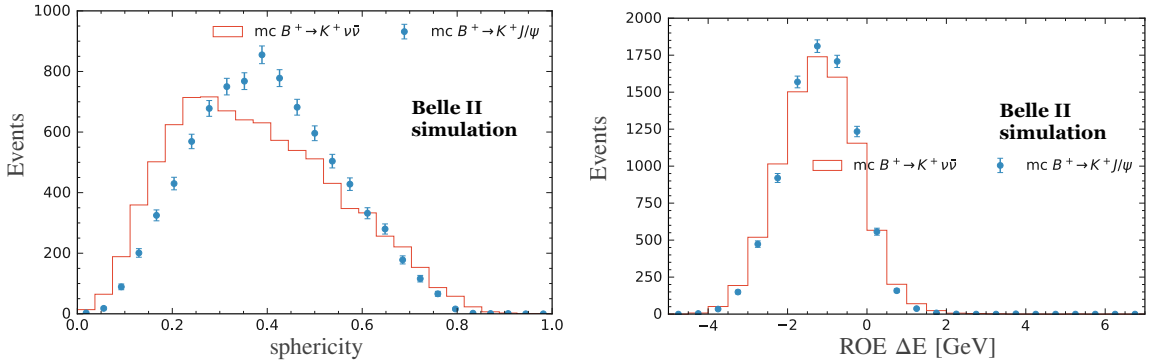


Figure 8.3: Distributions of event sphericity (left) and ROE ΔE (right) computed in simulated $B^+ \to K^+\nu\bar{\nu}$ events (red step histogram) and $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ events (blue points).

8.1.3 Modification of the control channel

To reproduce the experimental and kinematic features of $B^+ \to K^+\nu\bar{\nu}$, the selected $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ decays are modified and reconstructed following the steps summarised in Sec. 8.1.
The modification procedure is implemented in a validation study using:

- 5×10^4 simulated $B^+ \to K^+\nu\bar{\nu}$ events;

- 5×10^4 simulated $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events;
- $63\,\text{fb}^{-1}$ of on-resonance data.

An independent sample of 2×10^5 simulated $B^+ \to K^+\nu\bar{\nu}$ events is used for the four-momentum replacement. The first 10^5 generated K^+ four-momenta are used to correct the K^+ kinematics in $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ decays identified in MC; the 10^5 four-momenta of the second half are used for the four-momentum replacement in $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ decays identified in data.
The fraction of identified $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events passing the selection criteria of the inclusive tagging is $(81.5 \pm 0.9)\%$ in data and $(83.8 \pm 0.2)\%$ in simulation. The reconstructed events in data and MC are classified with BDT_1 and BDT_2 and the results are summarised in Fig. 8.4. The main figure shows the distributions of the BDT_1 outputs computed in data and MC before ($B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$) and after ($B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$) the di-muon removal and the update of the kaon four-momentum. The BDT_1 distributions in data and MC peak at 0 for the unmodified events and at 1 for the modified events. As a reference, also the distribution of the BDT_1 outputs computed in simulated $B^+ \to K^+\nu\bar{\nu}$ events is plotted, and it agrees well with the data and MC distributions of $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events. In the control channel, the overall agreement between the outputs of BDT_1 in simulation and data is good before and after the modifications. In the latter case, a p-value of 7% is estimated by means of the two-sample Kolmogorov-Smirnov (K-S) test [203] (a brief description of the K-S test is given in Appendix C).

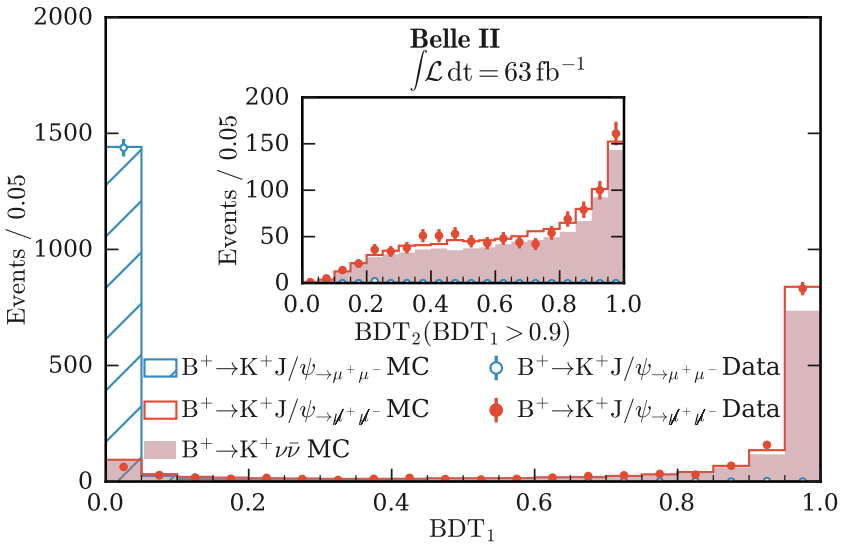


Figure 8.4: Distributions of the classifier outputs BDT_1 (main figure) and BDT_2 for $\text{BDT}_1 > 0.9$ (inset). The distributions are shown before ($B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$) and after ($B^+ \to K^+ J/\psi_{\to e^+e^-}$) the dimuon removal and the update of the kaon-candidate four-momentum in the selected events in MC and data. As a reference, the classifier outputs computed in simulated $B^+ \to K^+\nu\bar{\nu}$ signal events are overlaid. The simulation histograms are scaled to the total number of $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events selected in data.

The inner plot in Fig. 8.4 shows the distributions of the BDT_2 outputs for $\text{BDT}_1 > 0.9$. None of the unmodified $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events from data and MC survive this selection. Conversely, in the region of highest signal sensitivity, $\text{BDT}_1 > 0.9$ and $\text{BDT}_2 > 0.95$, $(9.36 \pm 0.71)\%$ of the $B^+ \to K^+ J/\psi_{\to e^+e^-}$ events pass the selection in data and $(8.85 \pm 0.13)\%$ in MC. For the modified events, the agreement between the BDT_2 distributions in data and MC is excellent, as demonstrated by the p-value of the corresponding K-S test, which is equal to 23%. These distributions agree very well also with the BDT_2 outputs computed in simulated $B^+ \to K^+\nu\bar{\nu}$ events.
The validation is extended to all the variables exploited by the inclusive tagging. In general, excellent data-MC agreement is observed for the modified events, $B^+ \to K^+ J/\psi_{\to e^+e^-}$, especially in the region of

highest signal sensitivity, $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.95$. The distributions of two important variables like the K^+ transverse momentum and the event sphericity are shown in Figs. 8.5, 8.6. The data-MC agreement is excellent before and after the application of the BDT_1 and BDT_2 selections. In particular, for events at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.95$, the K-S test comparing data and simulation estimates p-values of 57% for $p_\mathrm{T}(K^+)$ and 36% for sphericity. The agreement with the distributions computed in simulated $B^+ \to K^+\nu\bar{\nu}$ events is also good. Slight discrepancies are observed for $p_\mathrm{T}(K^+)$ before the application of the classifier selections, but the agreement is excellent at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.95$.
Given the overall excellent data-MC agreement in $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events, the validation with the on-resonance data can be considered successful. Moreover, the modification procedure provides satisfying results.

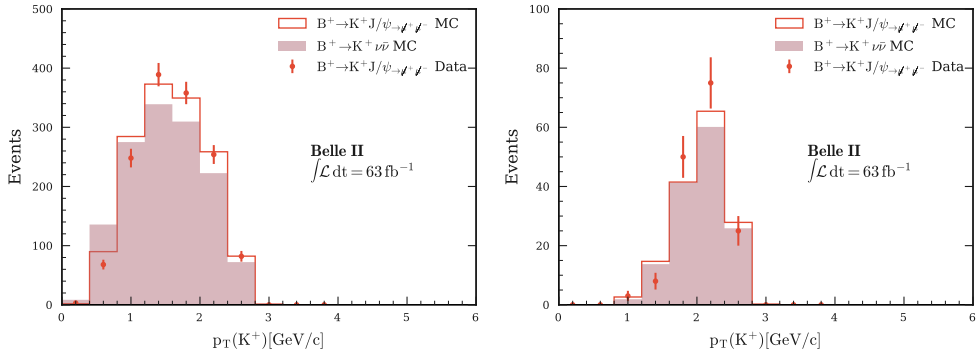


Figure 8.5: Transverse momentum p_T of the signal kaon candidate. The left plot is obtained without any selection on the classifiers, while the right plot is obtained with the requirement $\mathrm{BDT}_1 > 0.90$ and $\mathrm{BDT}_2 > 0.95$. The light-red filled histogram corresponds to events from the $B^+ \to K^+\nu\bar{\nu}$ MC sample, and the bright-red unfilled histogram (data points) corresponds to $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ MC (Data) events after the removal of the dimuon tracks and the update of the kaon kinematics. The simulation histograms are scaled to the total number of $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events selected in data.

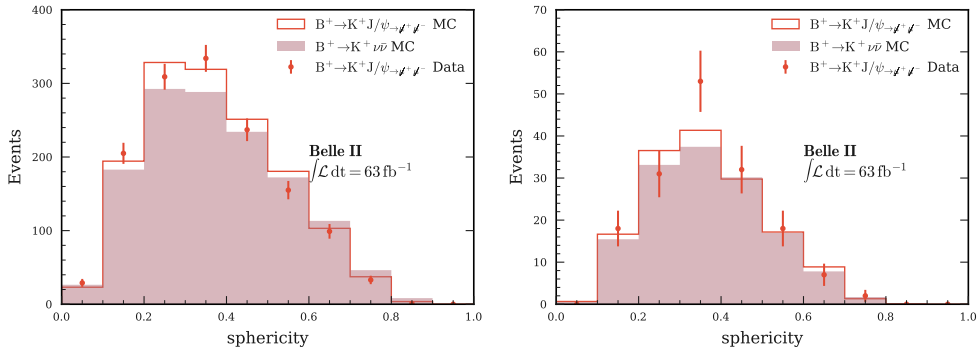


Figure 8.6: Event sphericity. The left plot is obtained without any selection on the classifiers, while the right plot is obtained with the requirement $\mathrm{BDT}_1 > 0.90$ and $\mathrm{BDT}_2 > 0.95$. The light-red filled histogram corresponds to events from the $B^+ \to K^+\nu\bar{\nu}$ MC sample, and the bright-red unfilled histogram (data points) corresponds to $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ MC (Data) events after the removal of the dimuon tracks and the update of the kaon kinematics. The simulation histograms are scaled to the total number of $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$ events selected in data.

8.2 Validation in off-resonance data

The off-resonance data sample consists of $9\,\mathrm{fb}^{-1}$ of collision data collected at an energy $60\,\mathrm{MeV}$ below the $\Upsilon(4S)$ resonance. This sample can be used to check the modelling of the continuum background simulation. The studies are performed using reconstructed events in simulation and data with $\mathrm{BDT}_1 > 0.9$. Events from the continuum backgrounds are reconstructed in a MC sample of $100\,\mathrm{fb}^{-1}$, and the yields in simulation are scaled to the integrated luminosity of the off-resonance data sample. The signal kaon candidate is required to satisfy the PID selection $\mathrm{kaonID} > 0.9$. Correction weights for data-MC discrepancies in the efficiency and misidentification rate of the kaonID selection are applied to the simulated events (see Sec. 2.8.4 for more details). All the variables used in the analysis are validated.

To summarise the results, the distributions in off-resonance data and continuum MC of three relevant variables are shown in Fig. 8.7. The variables are the transverse momentum of the signal kaon candidate, the ROE ΔE and the linear KSFW moment $H^{so}_{m,2}$. An overall 23% normalisation discrepancy between off-resonance data and continuum MC is observed. For this reason, in the plots, the yields in simulation are scaled to data by a factor of 1.23. For variables like ROE ΔE and $H^{so}_{m,2}$, the data and MC distributions appear slightly shifted away from each other. The discrepancies are confirmed by the data/MC ratios per bin, which reproduce the characteristic trends. These observations suggest some underlying mismodelling in the simulation of the continuum background, which may be related to the tuning of PYTHIA8 used at Belle II. A recent measurement of the hadron fragmentation at Belle [204] also supports this argument.

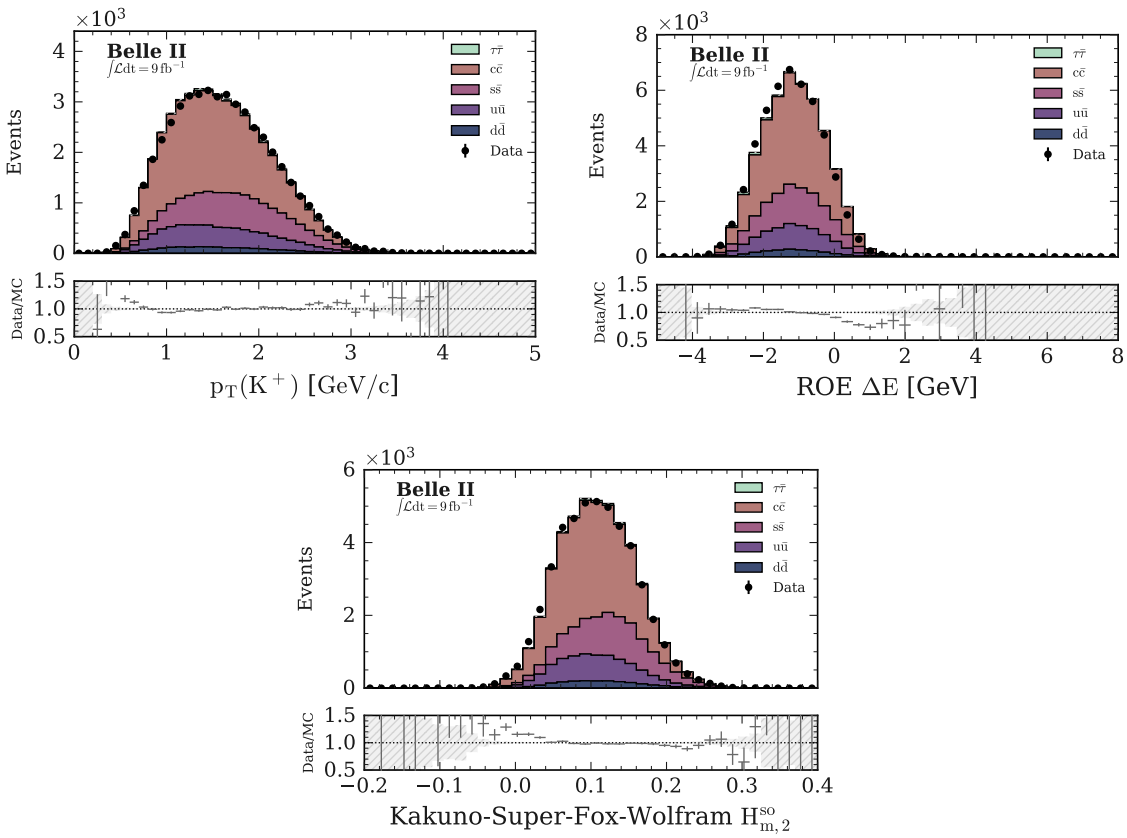


Figure 8.7: Distributions of the signal-kaon candidate transverse momentum (top left), ROE ΔE (top right) and linear KSFW moment $H^{so}_{m,2}$ (bottom) in off-resonance data (black points) and simulated continuum (stacked histogram). Yields in simulation are scaled to data by the ratio data/MC $= 1.23$. Data/MC ratios of events in the bins of the histograms are also illustrated for each variable. [2]

8.2.1 Reweighting of the continuum simulation

To correct the mismodelling observed in the continuum MC, a data-driven technique relying on a binary classifier is implemented. This strategy is inspired by Ref. [205], where a technique to enhance the data-MC agreement by event reweighting of the MC is presented. The implementation of the procedure follows the steps listed below:

- a MC sample equivalent to an integrated luminosity of $100\,\mathrm{fb}^{-1}$ is reconstructed, and the events from the continuum background with $\mathrm{BDT}_1 > 0.9$ are selected;
- events with $\mathrm{BDT}_1 > 0.9$ are also selected in the off-resonance data sample of $9\,\mathrm{fb}^{-1}$;
- a binary classifier, BDT_c, implementing the model defined in Sec. 6.3 and exploiting all the 50 variables described Sec. 5.7, is trained using the selected simulated events as background class and the selected off-resonance data as signal class;
- given the output p of the classifier, the continuum events from simulation are assigned a weight equal to $p/(1-p)$.

The ratio $p/(1-p)$ can be interpreted as an estimate of the likelihood ratio $\mathcal{L}(\mathrm{data})/\mathcal{L}(\mathrm{MC})$, where $\mathcal{L}(\mathrm{data})$ ($\mathcal{L}(\mathrm{MC})$) is the likelihood of the continuum event to be from data (simulation).
The BDT_c performance in continuum simulation and off-resonance data is shown in Fig. 8.8. To validate the classifier, BDT_c is applied to the continuum MC and off-resonance data samples used for the training and to a continuum MC test sample. No overfitting is observed. The discrepancy between the BDT_c distributions in data and simulation is a direct result of the observed mismodelling in the continuum MC. A perfect modelling of the simulated background would result in data and MC distributions completely overlapping and centred around $\mathrm{BDT}_\mathrm{c} = 0.5$.
The results of the application of the event weight $\mathrm{BDT}_\mathrm{c}/(1-\mathrm{BDT}_\mathrm{c})$ are presented in Fig. 8.9. The shape discrepancies visible in Fig. 8.7 are corrected. The data and MC distributions agree well within the statistical uncertainties. However, the 23% normalisation discrepancy between off-resonance data and continuum MC persists. A systematic uncertainty is introduced in the statistical model to take it into account. More details are given in Sec. 9.2.1.
The classification performance is also validated. The distributions of the BDT_1 outputs computed in data and simulation, with BDT_c weights applied, are shown in Fig. 8.10. Good data-MC agreement is observed up to the highest BDT_1 output values.

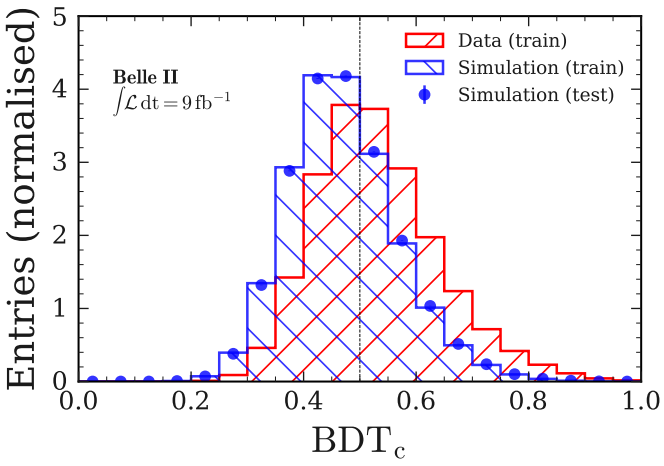


Figure 8.8: BDT_c outputs for off-resonance data (red) and continuum simulation (blue) events of the training (step histograms) and test (data points) samples. Each histogram is normalised to unit area. [2]

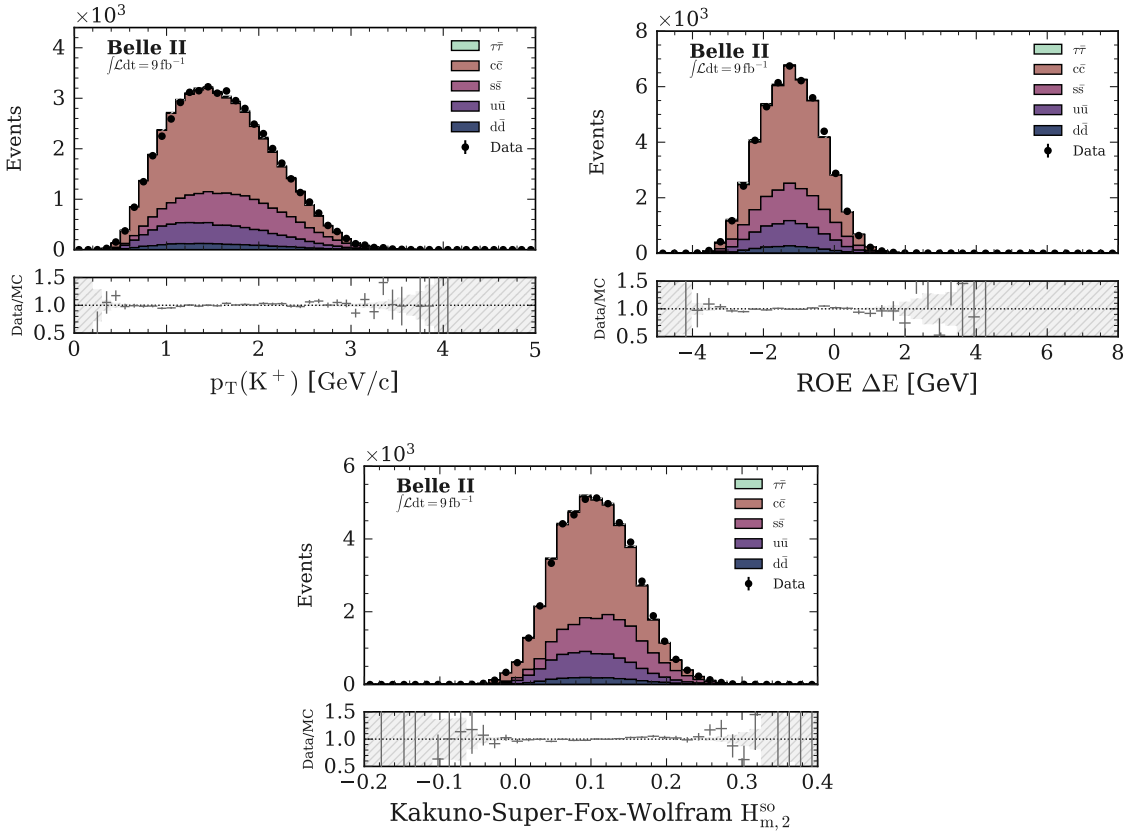


Figure 8.9: Distributions of the signal-kaon candidate transverse momentum (top left), ROE ΔE (top right) and linear KSFW moment $H^{so}_{m,2}$ (bottom) in off-resonance data (black points) and simulated continuum background (stacked filled histograms) after the application of the event weight $\mathrm{BDT_c}/(1-\mathrm{BDT_c})$. Yields in simulation are scaled to data by the ratio data/MC $= 1.23$. Data/MC ratios of events in the bins of the histograms are also illustrated for each variable. [2]

8.3 Validation in a moderate BDT sideband

A validation study with on-resonance data is performed by selecting events with moderate BDT output: $0.9 < \mathrm{BDT_1} < 0.99$ and $\mathrm{BDT_2} < 0.7$. This sideband region allows us to validate the analysis with events that are almost signal-like, but retaining a negligible amount of potential signal.

The first check in the sideband consists in the monitoring of the event yield in data. For every major data-taking period, called *experiment*, runs are grouped into batches of approximately the same integrated luminosity and the event yield in the sideband is evaluated for each batch. The idea is to test the possible presence of run periods affected by misreconstruction of some of the training variables, which would result in visible deviations from the average event yield of the experiment. The analysed data are collected in four experiments: experiment 7 ($0.5\,\mathrm{fb}^{-1}$), experiment 8 ($4.5\,\mathrm{fb}^{-1}$), experiment 10 ($3.6\,\mathrm{fb}^{-1}$), and experiment 12 ($54.6\,\mathrm{fb}^{-1}$). The yield stability in the sideband region is checked for each of them, and the results are presented in Fig. 8.11. Satisfactory yield stability is observed in all experiments.

To evaluate the data-MC agreement in the sideband, six MC batches of $100\,\mathrm{fb}^{-1}$ are reconstructed and compared to the reconstructed events from the on-resonance data sample. In both data and MC, the signal kaon candidate is required to have $\mathrm{kaonID} > 0.9$. The corresponding PID correction weights are applied to the simulation. The $\mathrm{BDT_c}$-based event weights are also assigned to the continuum background. The yields of the reconstructed events in simulation are scaled to the integrated luminosity of the on-resonance data

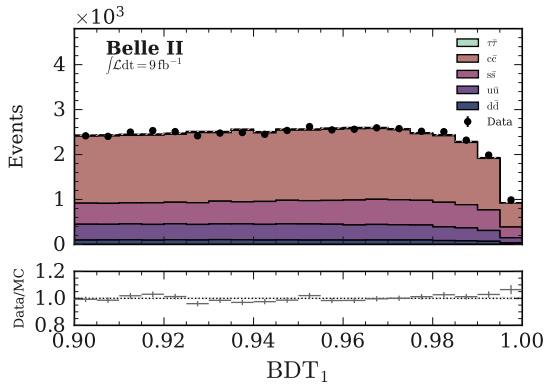


Figure 8.10: Distributions of the BDT_1 outputs computed in off-resonance data (black points) and simulated continuum background (stacked filled histograms) after the application of the event weight $\mathrm{BDT_c}/(1-\mathrm{BDT_c})$. Yields in simulation are scaled to data by the ratio data/MC $= 1.23$. Data/MC ratios of events in the BDT_1 bins are also illustrated. [2]

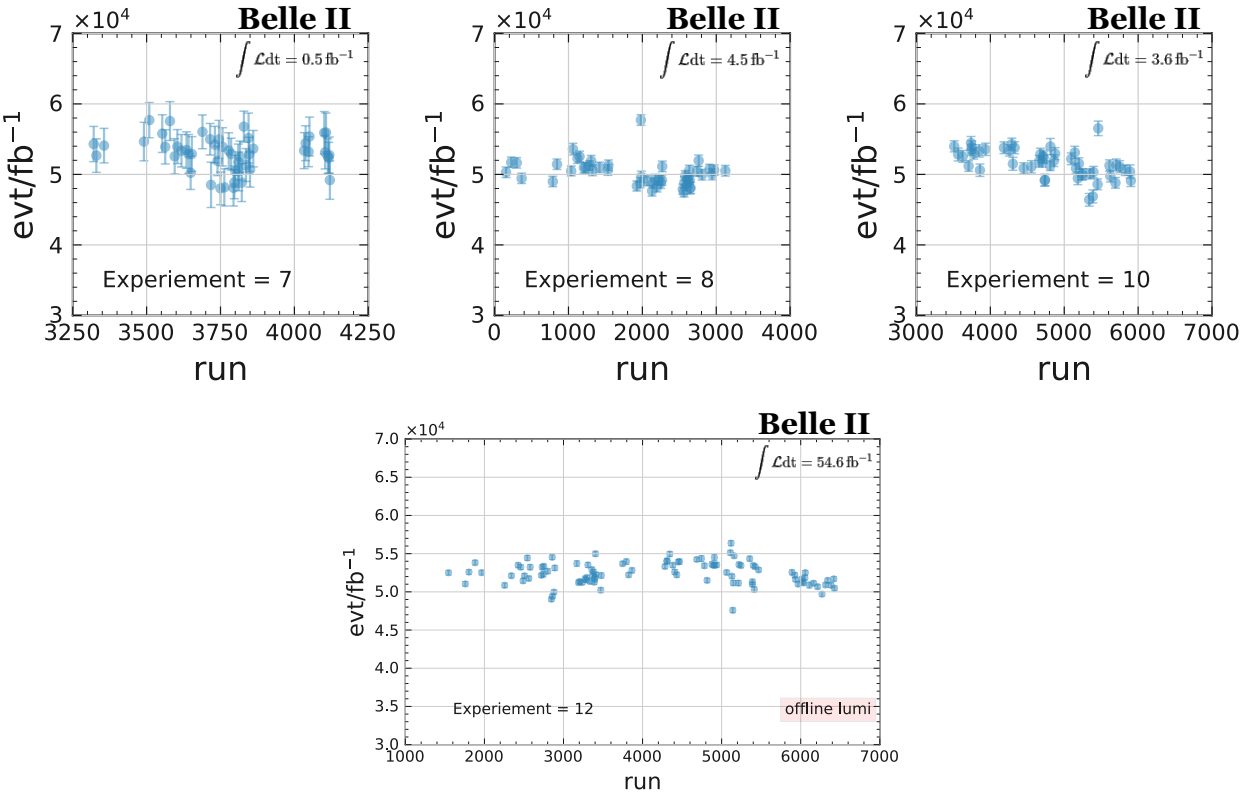


Figure 8.11: Monitoring of the event-yield stability in the sideband region corresponding to $0.9 < \mathrm{BDT}_1 < 0.99$ and $\mathrm{BDT}_2 < 0.7$. All the experiments of the data sample are checked: experiment 7 ($0.5\,\mathrm{fb}^{-1}$), experiment 8 ($4.5\,\mathrm{fb}^{-1}$), experiment 10 ($3.6\,\mathrm{fb}^{-1}$), and experiment 12 ($54.6\,\mathrm{fb}^{-1}$). The yield is computed in run batches defined such that the total integrated luminosity of the experiment is divided into roughly equal parts. [2]

sample, and an additional scale factor of 1.23 is applied to the continuum background to account for the data/MC normalisation discrepancy discussed in Sec. 8.2. The results of the validation are summarised in

Fig. 8.12, where three representative variables are shown: the normalised Fox-Wolfram moment R_1, the ROE ΔE and the BDT_2 output. The simulation describes data reasonably well, and the agreement is within 10% unless limited by the statistical precision. Moreover, the scaling of the continuum background by a factor of 1.23 results in a global data/MC yield ratio close to unity. This observation confirms that the normalisation discrepancy between data and MC is related to some mismodelling of the continuum simulation.

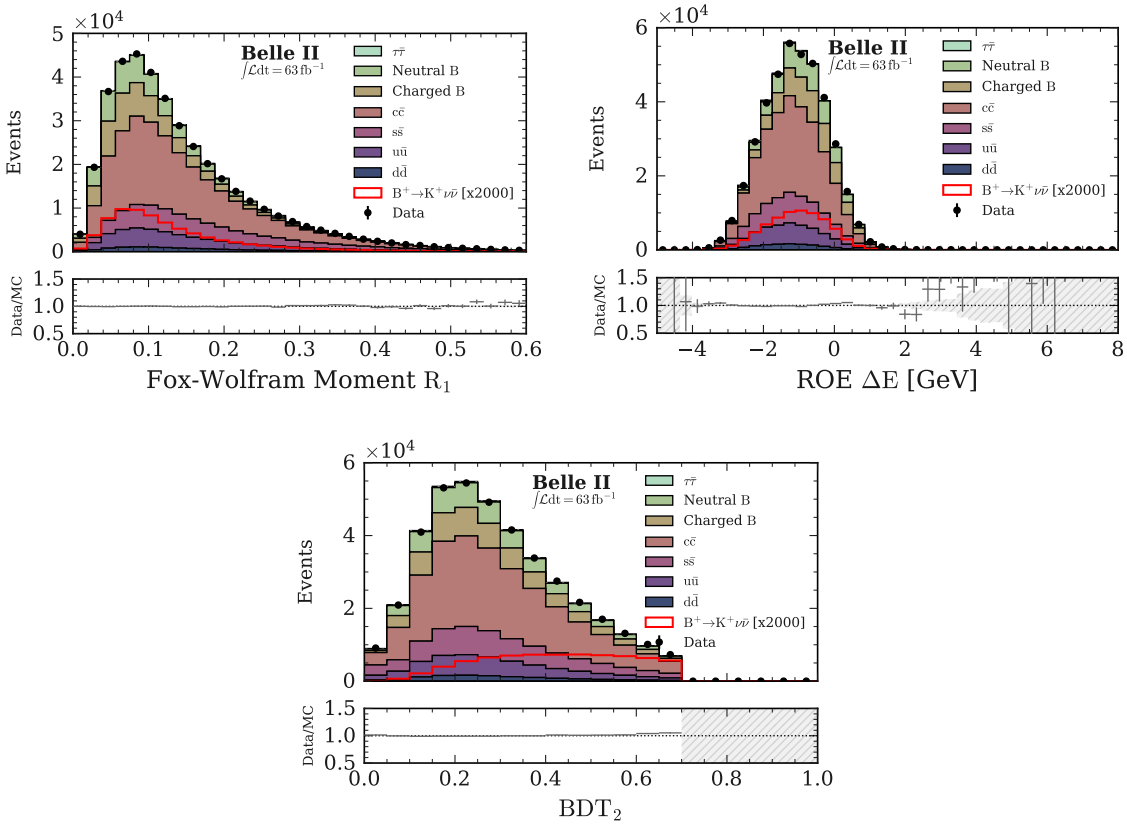


Figure 8.12: Distributions of the normalised Fox-Wolfram moment R_1 (top left), ROE ΔE (top right) and BDT_2 (bottom), from on-resonance data (black points) and simulation (stacked filled histogram) in a moderate BDT sideband ($0.9 < \mathrm{BDT}_1 < 0.99$ and $\mathrm{BDT}_2 < 0.7$). The continuum background is scaled by a factor of 1.23 to account for the 23% data/MC normalisation discrepancy observed in off-resonance data. Data/MC ratios of events in the bins of the histograms are also illustrated for each variable. [2]

Chapter 9

Fitting procedure

This chapter presents the fitting procedure implemented in the analysis. The formalism, and the tools, used to construct the statistical model are those presented in Chapter 3. Section 9.1 describes the statistical model constructed for the fit to the data. Section 9.2 provides a detailed description of the main sources of systematic uncertainty. In Section 9.3, the tests performed to validate the fit are discussed.

9.1 Statistical model

In this search, the reconstructed data from the on-resonance and off-resonance samples are fitted simultaneously. The statistical model is constructed with the `pyhf` library [3] and it is defined in terms of the signal-kaon transverse momentum, $p_{\mathrm{T}}(K^+)$, and of the BDT_2 output. Signal and background templates for the expected event yields in the two-dimensional $p_{\mathrm{T}}(K^+) \times \mathrm{BDT}_2$ distributions are derived from simulation. To model the on-resonance data, six MC batches, each corresponding to $100\,\mathrm{fb}^{-1}$, are reconstructed for the background. A sample of 4×10^6 simulated events is used for the signal. For the off-resonance data, fitted to constrain the yields of continuum events, templates are derived by reconstructing an independent MC batch of $100\,\mathrm{fb}^{-1}$. The signal kaon candidate in all the reconstructed events is required to satisfy the selection $\mathrm{kaonID} > 0.9$. In the signal region, defined in the next section, this kaonID selection retains 62% of kaons while removing 97% of pions. PID correction weights are applied to all the MC events, and the $\mathrm{BDT_c}$ weights defined in Sec. 8.2.1 are assigned to the MC events from the continuum backgrounds. As everywhere else in the analysis, signal events are reweighted to the SM prediction. In the channels of the model used to fit on-resonance data, the yields of reconstructed events from simulation are scaled to the integrated luminosity of $63\,\mathrm{fb}^{-1}$. Otherwise, where only off-resonance data is fitted, the yields of MC events from the continuum backgrounds are scaled to the off-resonance integrated luminosity. The signal yields are determined by scaling the reconstructed $B^+ \to K^+\nu\bar{\nu}$ events to the signal events expected in the on-resonance data sample.

9.1.1 Channels

The statistical model consists of disjoint binned distributions from four channels: a signal region and three control regions. In each channel, the total expected yields are obtained by summing over the contributing samples. A summary of the channel definitions and of the corresponding samples is provided in Table 9.1. The signal region (SR) is the region with the highest sensitivity to the signal. It ranges in p_{T} from $0.5\,\mathrm{GeV}/c$ to $3.5\,\mathrm{GeV}/c$ and in BDT_2 from 0.95 to 1. The boundaries of the p_{T} interval are chosen based on the overlay of the signal and background distributions shown in Fig. 5.3. The BDT_2 lower bound at 0.95 restricts to the peak of the signal significance illustrated in the bottom right plot of Fig. 6.1. In order to maximise the separation of signal from background, the SR is divided into 9 bins. The first optimisation of the bins is performed visually. The distributions in p_{T} and BDT_2 of signal and reconstructed background in a MC sample equivalent to $100\,\mathrm{fb}^{-1}$ are overlaid as shown in Fig. 9.1. The contours drawn in the plot allow us to

isolate 5%, 40%, 65%, 75% and 90% of the events in the 2D distributions, facilitating the definition of the bin boundaries. The chosen bins are $[0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$ in $p_\mathrm{T}(K^+)$ and $[0.95, 0.97, 0.99, 1.0]$ in BDT_2. This setup is validated by estimating the expected upper limit on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction, at 90% CL, in 15 different bin configurations, including the one previously chosen. In the evaluation of the limit the MC statistical uncertainties and the background normalisation uncertainties are included (a detailed discussion of the uncertainties in the statistical model is presented later in the chapter). The results are shown in Fig. 9.2. The configuration with $p_\mathrm{T}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$ and $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$ is the one providing the best upper limit on the signal branching fraction.

Channel	Bin numbers	Bin boundaries	Samples
Signal region (SR)	4 to 12	$p_\mathrm{T}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$, $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$	signal, neutral B, charged B, $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 1 (CR1)	1 to 3	$p_\mathrm{T}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$, $\mathrm{BDT}_2 \in [0.93, 0.95]$	signal, neutral B, charged B, $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 2 (CR2)	4 to 12	$p_\mathrm{T}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$, $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$	$u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 3 (CR3)	1 to 3	$p_\mathrm{T}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$, $\mathrm{BDT}_2 \in [0.93, 0.95]$	$u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$

Table 9.1: Channels and samples of the `HistFactory` [177] model implemented for the fit.

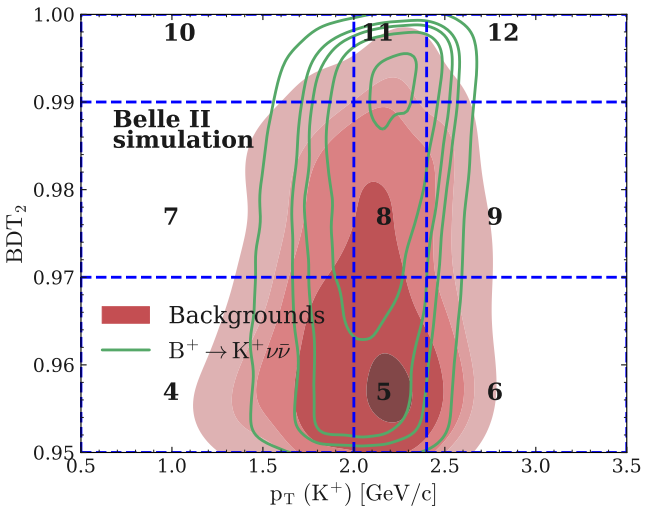


Figure 9.1: Two-dimensional $p_\mathrm{T}(K^+) \times \mathrm{BDT}_2$ distribution of signal (green contours) and background (red filled contours) from simulation after all selections in the SR. From the innermost to the outermost, the contours contain 5, 40, 65, 75, 90% of the events. The optimised bin boundaries are illustrated with blue dashed lines. [2]

The first control region (CR1) corresponds to a background-dominated region of the on-resonance data. The inclusion of this channel in the fit helps to constrain the normalisations of the yields from the different background categories in the SR. The CR1 shares the same bin boundaries of the SR in $p_\mathrm{T}(K^+)$ and consists of the single bin $[0.93, 0.95]$ in BDT_2 output. Fig. 9.3 shows that the chosen $p_\mathrm{T}(K^+)$ bins provide a good separation between the B backgrounds and the continuum backgrounds in on-resonance data.
The second and third control regions, CR2 and CR3, have the same bins in $p_\mathrm{T}(K^+)$ and BDT_2 as the SR and

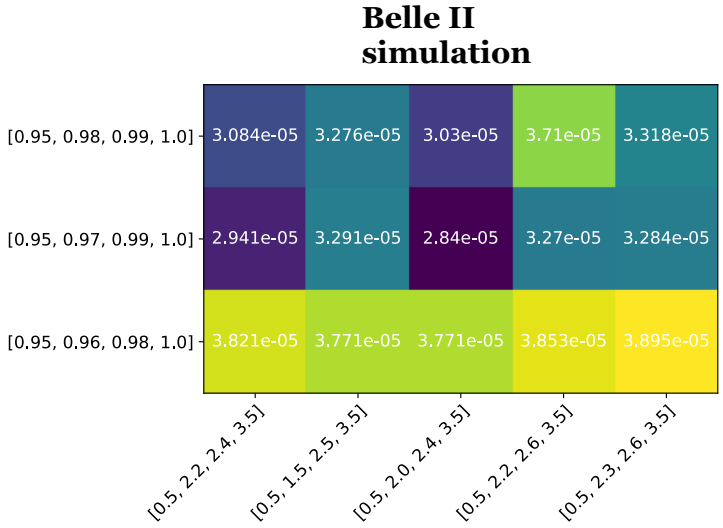


Figure 9.2: Sensitivity scan based on the evaluation of the expected upper limit on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction in 15 configurations of $p_{\mathrm{T}}(K^+)$ and BDT_2 bin boundaries, using a MC sample equivalent to $100\,\mathrm{fb}^{-1}$. Each square of the grid in the plot is identified by a specific set of $p_{\mathrm{T}}(K^+)$ bins (horizontal axis) and BDT_2 bins (vertical axis). The best upper limit is found for $p_{\mathrm{T}}(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$ and $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$. [2]

Sample	Channels	Expected yields
signal	SR, CR1	14, 4
charged B	SR, CR1	263, 174
neutral B	SR, CR1	142, 102
$c\bar{c}$	SR, CR1, CR2, CR3	228, 249, 34, 38
dd	SR, CR1, CR2, CR3	15, 16, 2, 3
$s\bar{s}$	SR, CR1, CR2, CR3	129, 121, 23, 17
$u\bar{u}$	SR, CR1, CR2, CR3	62, 57, 11, 8
$\tau^+\tau^-$	SR, CR1, CR2, CR3	3, 5, 1, 0

Table 9.2: Expected signal and background yields in the channels of the statistical model.

CR1, respectively, but they are populated with off-resonance data only. The inclusion of these two channels in the fit allows to put further constraints on the properties, and in particular on the normalisations, of the individual continuum background categories.
The expected signal and background yields in the SR, CR1, CR2, and CR3 are reported in Table 9.2. The distributions of the expected sample yields in the 12 bins of the SR and CR1 (CR2 and CR3) are shown in Fig. 9.4 (Fig. 9.5). In the SR the dominant background category is the charged B background, while in the CR1 the combined continuum background dominates over the B background. Among the continuum background samples, the largest contribution both in the CR1 (CR3) and SR (CR2) originates from the $e^+e^- \to c\bar{c}$ process.

9.1.2 Model parameters

As discussed in detail in Sec. 3.2.2, the `pyhf` allows the definition of the parameters of interest and the nuisance parameters in the statistical model. The first category groups the unconstrained parameters used to extract information about the signal. The second category consists of all the constrained parameters used

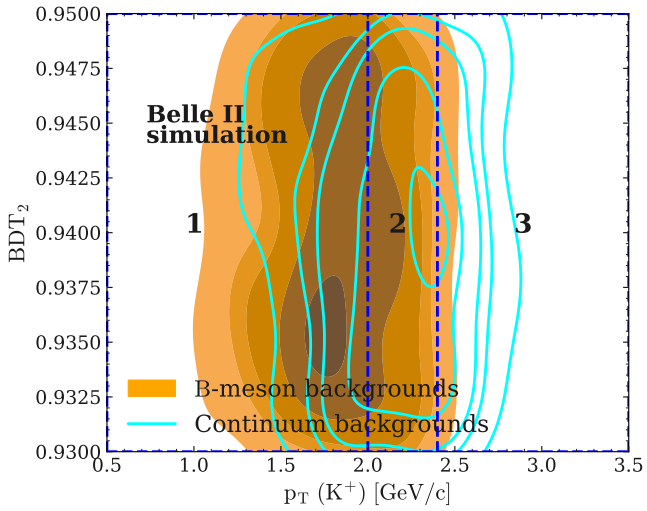


Figure 9.3: Two-dimensional $p_T(K^+) \times \mathrm{BDT}_2$ distribution of continuum background (cyan contours) and B background (orange filled contours) from simulation after all selections in the SR. From the innermost to the outermost, the contours contain $5,\ 40,\ 65,\ 75,\ 90\%$ of the events. The optimised bin boundaries are illustrated with blue dashed lines. [2]

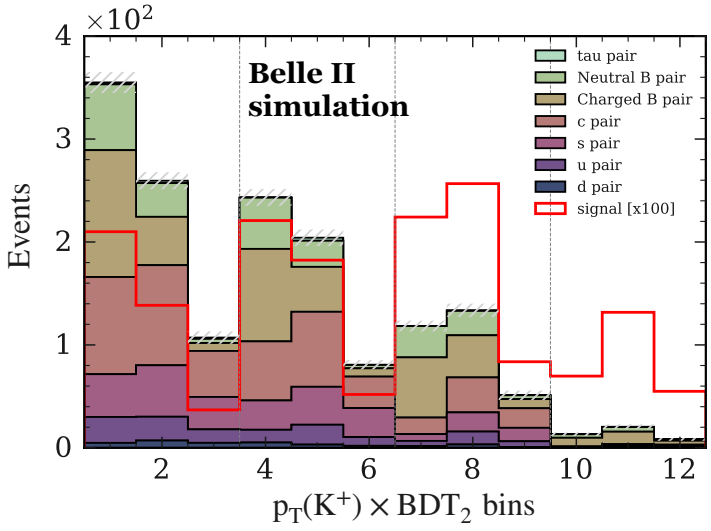


Figure 9.4: Expected sample yields in CR1 (bins 1 to 3) and SR (bins 4 to 12) for an integrated luminosity of $63\,\mathrm{fb}^{-1}$. The expected signal yield is scaled by a factor equal to 100.

to implement systematic uncertainties in the model.

Parameter of interest

The parameter of interest in the measurement is the signal strength μ. It is a multiplicative factor of the SM branching fraction for the $B^+ \to K^+\nu\bar{\nu}$ decay: $\mu = 1$ corresponds to $\mathrm{BR}(B^+ \to K^+\nu\bar{\nu})_{\mathrm{SM}} = 4.6 \times 10^{-6}$ [93]. In the `HistFactory` model, this parameter represents an unconstrained normalisation modifier of the signal sample.

Nuisance parameters

The nuisance parameters introduce constrained variations of the expectations due to various systematic uncertainties. The following list describes the parametrisation of the sources of systematic uncertainty included in the model:

- 7 parameters for the normalisation uncertainties of the seven background categories;
- 3 parameters for the correlated shape variation in the B samples due to the systematic uncertainty on the branching fractions of the leading B background decays;
- 3 parameters for the correlated shape variation in the signal sample due to the uncertainty on the SM form factors;
- 3 parameters for the correlated shape variation due to the uncertainty on the PID correction weights;
- 1 parameter for the correlated shape variation due to the uncertainty on the tracking efficiency;
- 1 parameter for the correlated shape variation due to the uncertainty on the energy calibration of photon clusters;
- 1 parameter for the correlated shape variation due to the uncertainty on the energy calibration of ECL clusters not matched to photons;
- 1 parameter per bin (in all the bins of all the channels, for all the samples) for the sum in quadrature of MC statistical uncertainty and uncorrelated part of the shape systematics.

The total number of nuisance parameters is 175. All the nuisance parameters have a Gaussian constraint. A detailed description of each source of systematic uncertainty is provided in the next section.

9.2 Systematic uncertainties

The evaluation of the systematic uncertainties is presented in this section. Some systematic biases originate from the detector response and the performance of the reconstruction algorithms, others are related to the modelling of the physics processes in the MC simulation.

9.2.1 Normalisation of the background yields

The leading systematic uncertainty is the normalisation uncertainty on the background yields. The yields of the seven individual background categories are allowed to float independently in the fit. A parameter corresponding to a normalisation uncertainty modifier is introduced in the statistical model for each background sample. Each parameter is coupled to a Gaussian constraint centred at the expected background yield from simulation, with a standard deviation corresponding to 50% of the central value. This value is chosen as a result of the global normalisation discrepancy of $(40 \pm 12)\%$ between the off-resonance data and the simulated continuum background in the control regions CR2 and CR3. Restricting to these regions, the data/MC yield ratio increases by approximately 20% compared to the value observed at $\mathrm{BDT}_1 > 0.9$ (see Sec. 8.2). The distributions, in CR2 and CR3, of events from off-resonance data and continuum MC, with yields in simulation scaled to data by a factor 1.4, are shown in Fig. 9.5.
The normalisation discrepancy is not observed in the control channel $B^+ \to K^+ J/\psi_{\to \mu^+\mu^-}$, but, as a conservative choice, a 50% normalisation uncertainty is assigned in the fit model also to the charged and neutral B backgrounds.

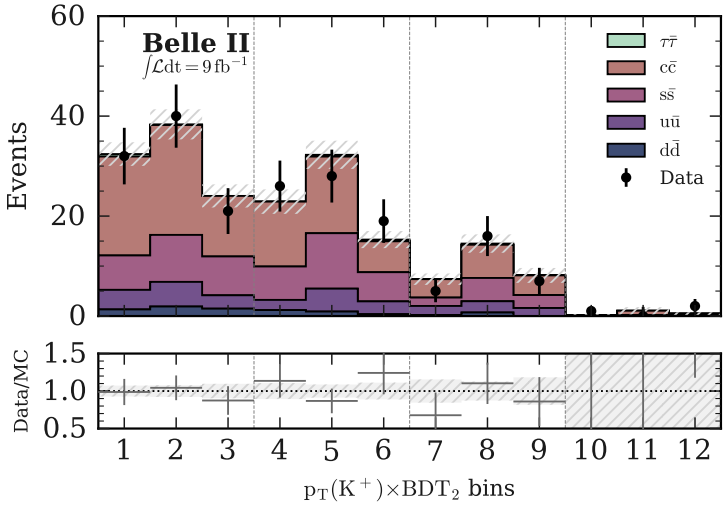


Figure 9.5: Distributions of the simulated continuum backgrounds and off-resonance data in the bins of the CR3 (bins 1 to 3) and CR2 (bins 4 to 12). Yields in simulation are scaled to data, and a correction factor is introduced to account for the observed yield discrepancy data/MC $= 1.40 \pm 0.12$. [2]

9.2.2 Branching fractions of the leading B-background decays

A source of systematic uncertainty is related to the fact that, in the official MC simulation, only the central values of the branching fractions are used to generate the charged and neutral B-meson decays. Hence, the idea is to take into account possible shape variations of the B backgrounds by varying the contribution of the leading decays according to the branching-fraction uncertainty provided by the PDG [43]. This is possible thanks to the background identification procedure presented in Chapter 7, which, for every reconstructed B-background event, allows to determine the generated B decay from which the selected signal-kaon candidate originates.
This systematic is evaluated using the charged and neutral B-background events in the 12 bins of the fit regions, reconstructed in a MC sample corresponding to an integrated luminosity of $300\,\mathrm{fb}^{-1}$. Central value and uncertainty of the branching fraction are stored for $\sim 80\%$ of the charged B decays and for $\sim 62\%$ of the neutral B decays in the signal side of the selected events, starting from the processes contributing the most. A thousand pseudo-experiments (also called MC toys) based on replicas of the charged and neutral B samples in the fit regions are produced. In each replica, an event is weighted if the branching fraction, $\mathrm{BR} \pm \sigma_{\mathrm{BR}}$, of the signal decay is one of those that are stored. The assigned weight is $w_{\mathrm{BR}} = (\mathrm{BR} + \Delta)/\mathrm{BR}$, where Δ is a variation of the central value, BR, drawn from a normal distribution $\mathrm{N}(0, \sigma_{\mathrm{BR}})$. All the other events remain unweighted. Histograms for the charged and neutral backgrounds are computed in the 12 bins of the fit regions for every MC toy. The samples of weighted bin counts $S_i = \{n_i\}_{j=1}^{1000}$, with bin numbers $i = 1, .., 12$ for the charged B background and $i = 13, ..., 24$ for the neutral B background, are determined and the bin covariance matrix $\mathrm{cov}(S_i, S_k)$ is estimated.
The covariance matrix is decomposed using the procedure presented in Sec. 3.2.2.1. The three principal components with the largest singular values are split into the contributions from the charged and neutral B backgrounds, so that three vectors of systematic variations are obtained per sample. In the statistical model, they are coupled to three nuisance parameters accounting for correlated shape variation. The decomposition of the equivalent correlation matrix, $\mathrm{corr}(S_i, S_k) = \mathrm{cov}(S_i, S_k)/(\sigma_i \sigma_k)$, is shown for illustration purpose in Fig. 9.6. Here, the right plot corresponds to the approximation of the original matrix using Eq. 3.14. The main features of the original matrix are captured well by the approximated matrix obtained after the decomposition.

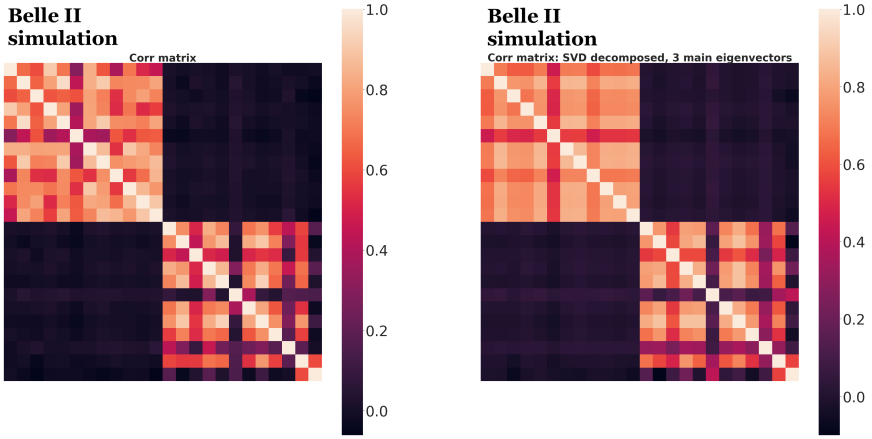


Figure 9.6: Left plot: original correlation matrix for the branching-fraction systematic uncertainty. Right plot: approximated correlation matrix after the decomposition procedure. From top to bottom, left to right, the first 12 bins are populated with events from the charged B background and the last 12 bins with events from the neutral B background.

9.2.3 SM form factor

A specific source of systematic uncertainty affecting only the signal sample arises from the uncertainty on the SM form factor used to reweight the signal (see Sec. 4.2.1). The form factor depends on three real parameters, $(\alpha_0, \alpha_1, \alpha_2)$, which are estimated with a fit to Lattice QCD and LCSR results. The estimated central values are provided with their standard deviations in Eq. 4.3 and the correlation matrix, $\mathrm{corr}(\alpha_i, \alpha_j)$, is given in Eq. 4.4. Using this information, the corresponding covariance matrix is derived and decomposed into its three principal components $\vec{\sigma}_i$, $i = 1, 2, 3$. These three vectors are used to produce three varied $\vec{\alpha}$ vectors, $\vec{\alpha} + \vec{\sigma}_i$, from which three modified form factors are constructed. For each modified form factor, a spline for the signal reweighting is computed as in Sec. 4.2.1. The splines are used to weight the signal events in the 12 bins of the CR1 and SR. Three modified signal distributions are obtained, as shown in Fig 9.7, where 4×10^6 simulated signal events are reconstructed. These distributions are compared to the nominal signal distribution in the 12 bins, and three independent vectors of systematic variations of the bin counts are determined. In the statistical model, each of them is coupled to an individual nuisance parameter for correlated shape variation.

9.2.4 PID correction

PID correction weights for the selection $\mathrm{kaonID} > 0.9$ (see Sec. 2.8.4 for more details) are evaluated in specific $(p, \cos\theta)$ bins. The p bins are not equivalent to the p_T bins of the fit regions. Therefore, the uncertainties on the weights may introduce a correlated shape variation. To evaluate it, 4×10^6 simulated signal events and background equivalent to $300\,\mathrm{fb}^{-1}$ are reconstructed. Based on replicas of the reconstructed events, 500 MC toys are generated. In every toy, the weight for each $(p, \cos\theta)$ bin is varied by a quantity drawn from a lognormal distribution having width equal to the weight uncertainty. The new weights are used to compute the histograms of the signal and background events in the 12 bins of the fit regions. The weighted bin counts $S_{ij} = \{n_{ij}\}_{k=1}^{500}$, with i running over the samples, j over the bins, and k over the toys, are determined and the bin covariance matrix, $\mathrm{cov}(S_{ij}, S_{mn})$, is computed. The covariance matrix is decomposed as in Sec. 3.2.2.1, and the principal components corresponding to the largest three singular values are split among the samples. Three vectors of independent systematic variations are determined per sample. In the statistical model, the vectors are coupled to three nuisance parameters for correlated shape variation. For illustration purpose, the decomposition of the equivalent correlation matrix is shown in Fig. 9.8, where the right plot corresponds to the approximation of the correlation matrix with Eq. 3.14. The approximated matrix reproduces well the main characteristics of the original matrix.

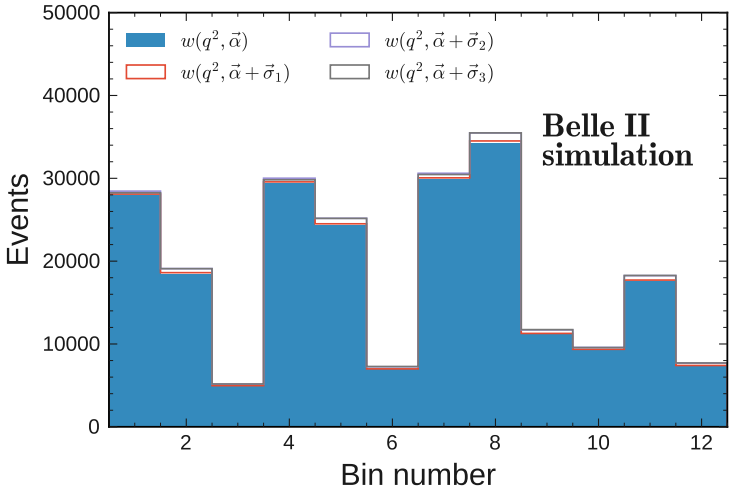


Figure 9.7: Signal yields in the 12 bins of the CR1 (bins 1 to 3) and SR (bins 4 to 12) from the reconstruction of 4×10^6 signal events. The blue filled histogram is obtained by reweighting the signal according to the nominal SM form factor, characterised by the central values of $\vec{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$ given in Eq. 4.3. The other three step histograms are derived using the modified form factors constructed with the varied parameters $\vec{\alpha} + \vec{\sigma}_i$, with $i = 1, 2, 3$.

Figure 9.9 shows the comparison between the relative PID systematic uncertainties, derived from the main diagonal of the covariance matrix, and the relative MC statistical uncertainties in the 12 bins of the fit regions for all the samples. The statistical uncertainty prevails over the PID-correction systematic uncertainty.

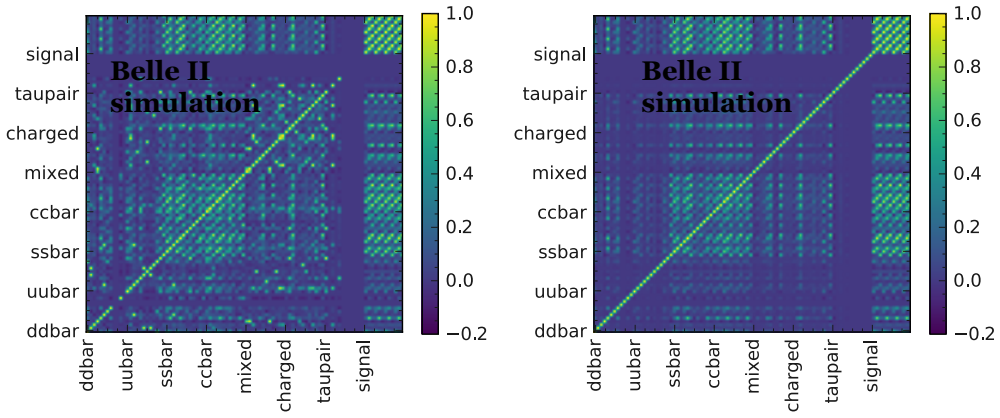


Figure 9.8: Left plot: original correlation matrix for the PID-correction systematic uncertainty [1]. Right plot: approximation of the correlation matrix after the decomposition procedure. [2]

9.2.5 Tracking efficiency

The current tracking-efficiency uncertainty in Belle II is estimated to be 0.9% [206]. The corresponding systematic bias in the analysis is evaluated by simulating a 0.9% probability that one of the tracks in the event is not reconstructed. This feature is introduced in the reconstruction of 4×10^6 simulated signal events and in the reconstruction of $100\,\text{fb}^{-1}$ of simulated background. The reconstructed events are then processed

[1] In some plots in this chapter the term *mixed* is used to indicate the neutral B background. This is a common jargon at B factories.

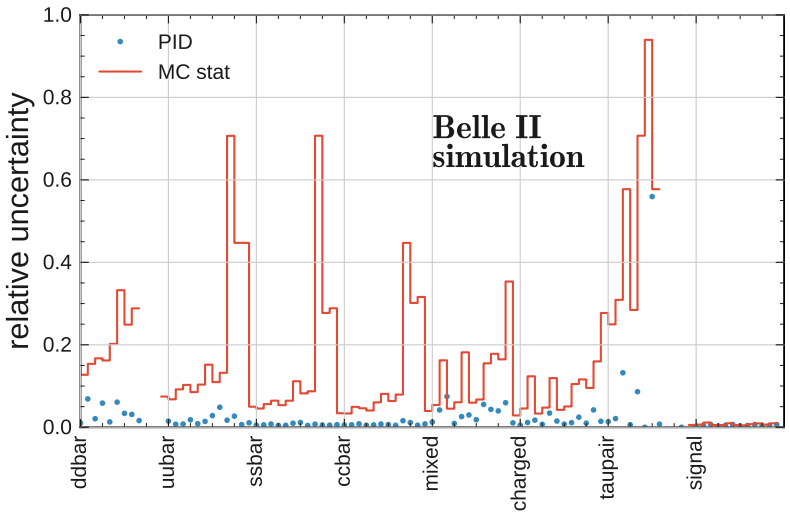


Figure 9.9: Relative PID systematic uncertainties (blue dots) compared to the relative MC statistical uncertainties (red step histogram) in the 12 bins of the fit regions for all the samples. [2]

with the inclusive tagging algorithm. The signal and background histograms in the 12 bins of the fit regions, obtained by simulating the tracking inefficiency, are compared to the nominal histograms to estimate the systematic variations. For the background samples, this simple procedure would result in an overestimation of the systematic uncertainty, due to the large statistical component due to the limited size of the samples. Therefore, the background MC expectations from the reference and modified samples are smoothed before the comparison. The smoothing is performed in p_T and BDT_2 output by means of a Gaussian kernel density estimation (KDE, see Appendix E), with kernel bandwidth $h = 0.5$. The value of the parameter is optimised to avoid oversmoothing, which can introduce additional bias, and undersmoothing, which does not help to reduce the statistical component of the uncertainties. In Fig. 9.10, the background histograms in the 12 bins of the fit regions are shown for the reference and modified samples before and after the KDE smoothing. The smoothed distributions reproduce the main features of the original histograms.
The relative statistical and systematic uncertainties in the 12 bins of the fit regions are illustrated for signal and backgrounds in Fig. 9.11 (the $\tau^+\tau^-$ sample is dropped given its negligible contribution). The relative systematic uncertainty in each bin is evaluated as the modified/reference ratio of the bin counts minus one. For the signal sample, the systematic uncertainty behaves smoothly from bin to bin and is below 5%. For the backgrounds, the smoothed estimates are more reliable than the original estimates, which are characterised by large fluctuations caused by migrations of events between bins. Overall, the figure shows that the systematic uncertainty due to the tracking-efficiency uncertainty is minor with respect to the MC statistical uncertainty.
In the statistical model, the vectors made of the systematic variations are coupled to a common nuisance parameter accounting for correlated shape variation.

9.2.6 Energy calibration of photon clusters

A systematic bias in the measurement can be introduced by the uncertainty on the photon energy calibration. The uncertainty on the energy of photon clusters in the ECL is estimated to be 0.5%. To evaluate the impact of this systematic uncertainty in the analysis, the energy of the ECL clusters matched to photons in simulation is scaled down by 0.5%. The same procedure described in Sec. 9.2.5 is used to estimate the systematic variations in the signal and background samples.
The relative statistical and systematic uncertainties in the 12 bins of the fit regions are illustrated for signal and backgrounds in Fig. 9.12. The relative systematic uncertainty in each bin is evaluated as the

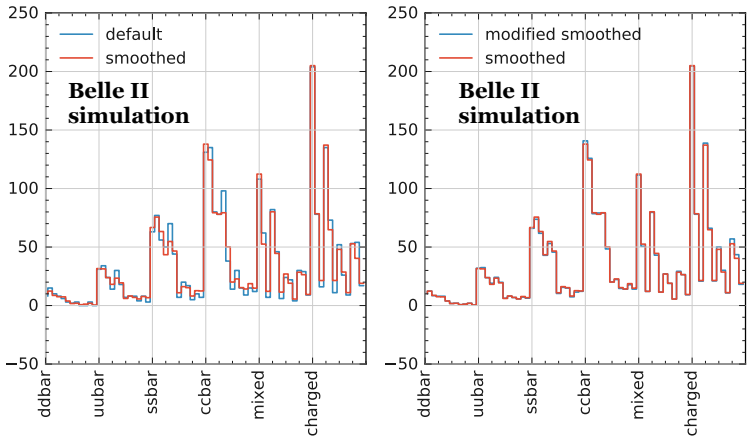


Figure 9.10: Left plot: background histograms from the reference sample obtained before (blue histogram) and after (red histogram) the Gaussian KDE smoothing in $p_{\mathrm{T}}(K^+)$ and BDT_2 output with bandwidth $h = 0.5$. Right plot: smoothed background histograms from the reference sample (red histogram) and the modified sample (blue histogram) derived by simulating the tracking inefficiency. For each background sample, the counts in the 12 bins of the fit regions are illustrated. [2]

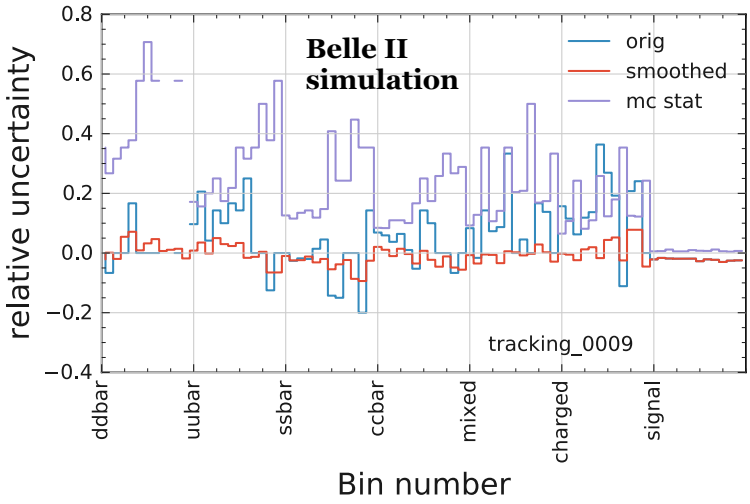


Figure 9.11: Systematic uncertainty related to the tracking efficiency: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The violet histogram represents the relative MC statistical uncertainty. The blue (red) histogram corresponds to the relative systematic uncertainty obtained without (with) KDE smoothing. The $\tau^+\tau^-$ background is excluded given its negligible contribution. [2]

modified/reference ratio of bin counts minus one. It is minor compared to the MC statistical uncertainty. In the statistical model, the vectors of systematic variations obtained for signal and backgrounds are coupled to an individual nuisance parameter implementing a correlated shape modification.

9.2.7 Energy calibration of ECL clusters not matched to photons

Another source of systematic bias is related to the uncertainty on the energy calibration of ECL clusters not matched to photons. The uncertainty on the energy deposits of neutral hadrons, and neutral particles from beam background, is potentially much larger than 0.5%. This could significantly affect variables on which

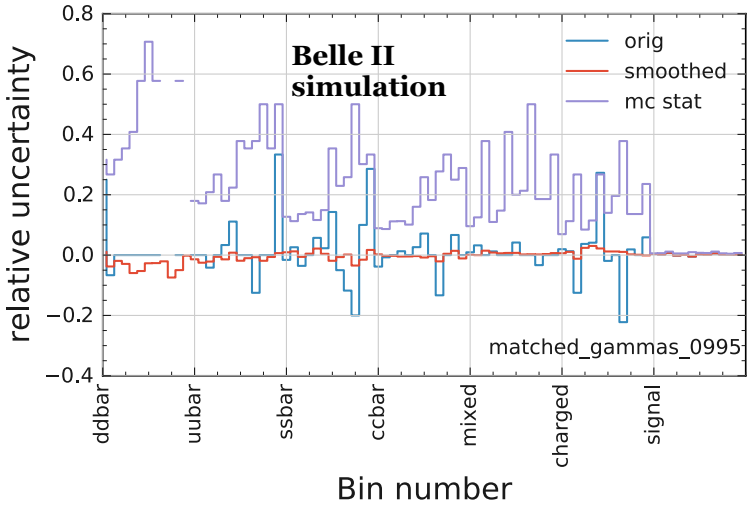


Figure 9.12: Systematic uncertainty related to the energy calibration of ECL clusters matched to photons: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The violet histogram represents the relative MC statistical uncertainty. The blue (red) histogram corresponds to the relative systematic uncertainty obtained without (with) KDE smoothing. The $\tau^+\tau^-$ background is excluded given its negligible contribution. [2]

the inclusive tagging heavily relies, such as the ROE ΔE.
A dedicated study is performed in the control channel $B^+ \to K^+ J/\psi_{\to\mu^+\mu^-}$ to further investigate this source of systematic bias. The fundamental assumption is that the energy of ECL clusters matched to photons is known to a much better accuracy than that of the other neutral ECL clusters, such that the total neutral energy of the ROE can be decomposed as

$$E^n_{\rm ROE}(f_h) = \sum_i E^\gamma_i + f_h \sum_j E^n_j, \tag{9.1}$$

where i runs over all the ECL clusters matched to photons and the j over all the ECL clusters not matched to photons, most of which belong to neutral hadrons. In Eq. 9.1, f_h is a free parameter, a scale factor quantifying the accuracy of the energy calibration for the unmatched clusters. To find the optimal f_h, a range of f_h values is scanned and for each value the distribution of $E^n_{\rm ROE}(f_h)$ computed in simulation is compared with the distribution of $E^n_{\rm ROE}$ in data, by means of a two-sample K-S test (described in Appendix C). The results of the optimisation scan are illustrated in Fig. 9.13. The K-S test p-value is maximised at $f_h \approx 1.05$, but the peak is broad. Its full-width at half maximum is $\sim 10\%$. This value is used as an estimate of the systematic uncertainty on the energy of the unmatched ECL clusters. To estimate the effect of this systematic uncertainty in the analysis, the energy of the reconstructed clusters not matched to photons is scaled down by 10% in simulation and the same procedure described in Sec. 9.2.5 is implemented. The resulting relative systematic uncertainties determined for the signal and background samples in the 12 bins of the fit regions are shown in Fig. 9.14. The KDE smoothing reduces the statistical fluctuations of the systematic uncertainty. Larger fluctuations are still present in the $d\bar{d}$ background sample, due to the very small size of the sample in the fit regions. It is the second smallest background, after $\tau^+\tau^-$, which is excluded from the study. Overall, the MC statistical uncertainty is larger than the systematic uncertainty.
The systematic variations derived in the signal and background samples are coupled in the statistical model to a common nuisance parameter for correlated shape variation.

9.2.8 Impact of systematic uncertainties

The impact of the systematic uncertainties is evaluated in terms of the expected upper limit on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction computed at a 90% CL. The values of the upper limit are illustrated in Fig. 9.15 as a

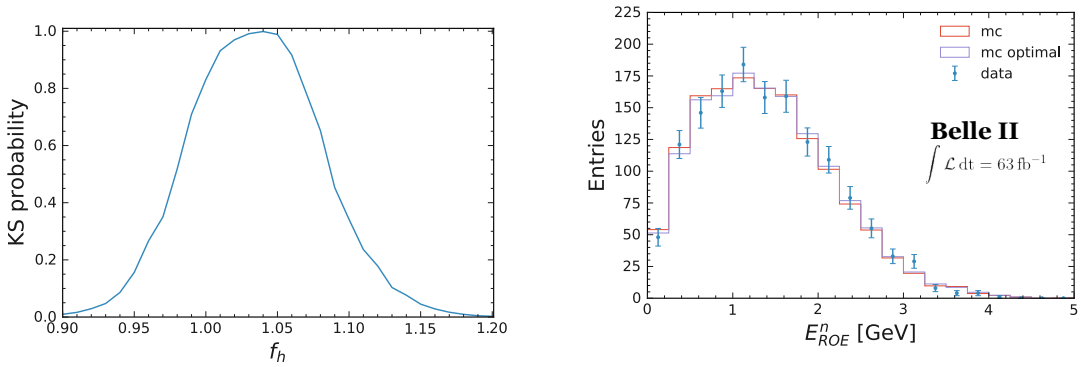


Figure 9.13: Left plot: p-value of the two-sample K-S test between the distribution of E^n_{ROE} in data and the distribution of $E^n_{\mathrm{ROE}}(f_h)$ (see Eq. 9.1) in simulation, as a function of the f_h values of the scan. Right plot: comparison of the E^n_{ROE} distribution in data (blue dots with error bars), default simulation (red histogram) and simulation with energy of the neutrals not matched to photons scaled by the optimal f_h value (violet histogram). [2]

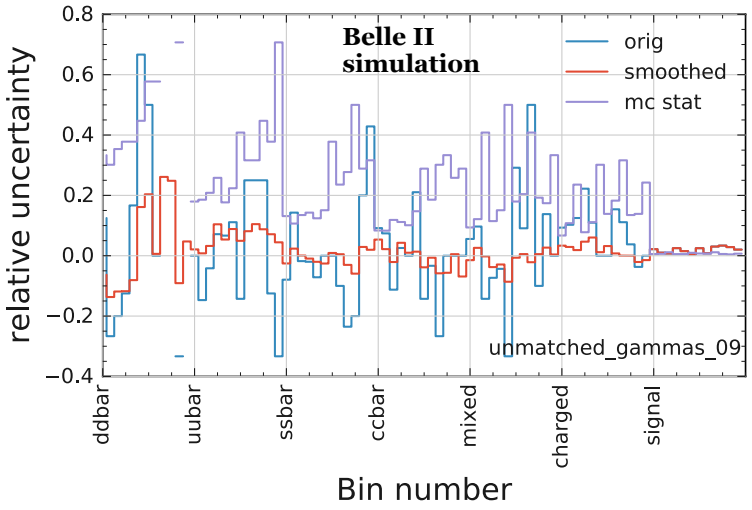


Figure 9.14: Systematic uncertainty related to the energy calibration of ECL clusters not matched to photons: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The violet histogram represents the relative MC statistical uncertainty. The blue (red) histogram corresponds to the relative systematic uncertainty obtained without (with) KDE smoothing. The $\tau^+\tau^-$ background is excluded given its negligible contribution. [2]

function of the sources of systematic uncertainty included in the statistical model. The plot shows that the major increase of the expected upper limit corresponds to the introduction of the 50% systematic uncertainty on the normalisations of the background yields, which produces a 74% increase of the upper limit. The progressive inclusion of the other sources described in the previous sections leads to a final relative increase of only 2.6%.
All the systematic uncertainties per bin in the signal and background samples are summarised in Table F.1 shown in Appendix F.

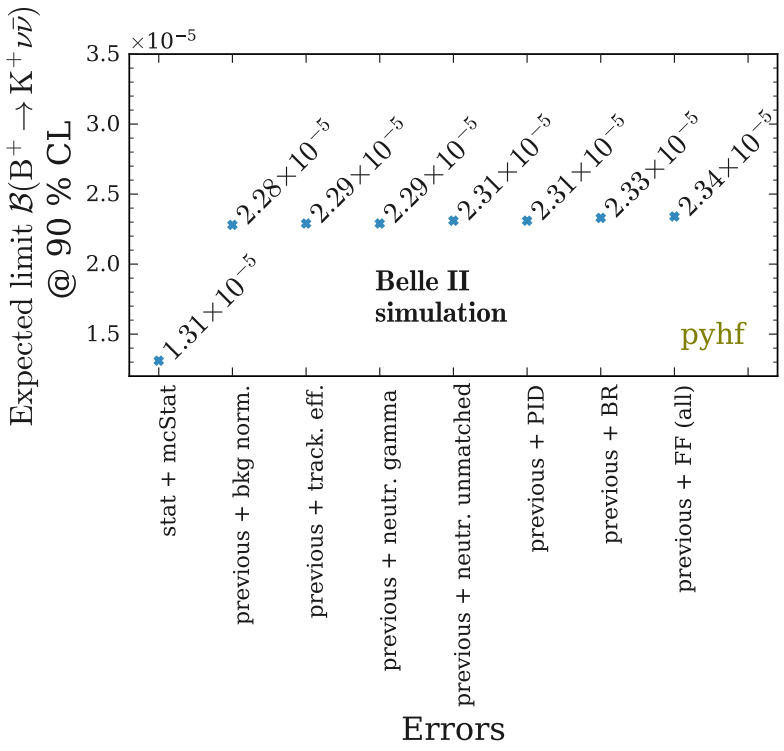


Figure 9.15: Expected upper limit on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction at 90% CL as a function of the sources of systematic uncertainty included in the statistical model. The limit is computed using the `pyhf` library. [2]

9.3 Fit validation

The fit validation is a fundamental step preceding the final signal extraction. In particular, the accuracy of the statistical model and the fit quality are evaluated in two independent tests. Section 9.3.1 describes a signal-injection test, which primarily allows verifying that the statistical model does not introduce any bias in the signal extraction, for example by absorbing part of the signal in the background model. In the second test, discussed in Section 9.3.2, the p-value of the fit to the observed data is estimated. This test gives an important indication about the model description of the data.

9.3.1 Signal injection study

Signal injection studies allow us to evaluate whether the statistical model defined in Sec. 9.1 introduces any bias in the signal extraction. Tests of the fit bias are performed for three signal strength hypotheses: $\mu_{\text{sig}} = 1.0, 5.0, 20.0$. In the three scenarios, enhanced signal is injected and ensembles of MC toys are generated. Every toy consists of fluctuations of the expected bin counts in the fit regions, SR, CR1, CR2, CR3, generated following the Poisson distribution. In order to test the full model, also variations due to the sources of systematic uncertainty are included by generating Gaussian systematic fluctuations.
A thousand MC toys are simulated for $\mu_{\text{sig}} = 1.0, 5.0, 20.0$ and fitted with `pyhf` and `sghf`. The first test performed using the pseudo-experiments is a simple cross-validation of the fit tools. The signal strength μ is estimated with `pyhf` and `sghf` for every toy of the three ensembles. The values estimated with the two fit tools are then compared in Fig. 9.16. The results obtained with `pyhf` and `sghf` are observed to be in good

agreement.

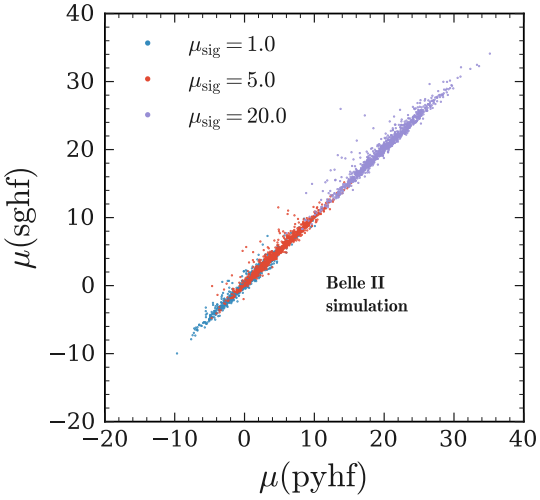


Figure 9.16: `pyhf` vs `sghf` estimates of the signal strength μ from the fits to the 1000 MC toys in the ensembles with injected signal of $\mu_{\rm sig} = 1.0,\ 5.0,\ 20.0$. [2]

To check the accuracy of the statistical model and ensure that no bias is introduced in the signal extraction, the distributions of the pulls

$$\mathrm{p} = \frac{(\mu - \mu_{\rm sig})}{\sigma_\mu} \tag{9.2}$$

are computed in the ensembles of MC toys with injected signal of $\mu_{\rm sig} = 1.0,\ 5.0,\ 20.0$. In Eq. 9.2, μ is the estimated value of the signal strength and σ_μ the post-fit uncertainty on the estimated μ. In particular, σ_μ is computed in a scan of the profile likelihood ratio $-2\ln\lambda_p(\mu)$ as a function of the signal strength μ.
In case of unbiased signal extraction, the pulls in each ensemble should follow a standard normal distribution, $N(0,1)$. The results from the fits with the `pyhf` are shown in Fig. 9.17. The pull distributions from the fits to MC toys with injected signal of strength $\mu_{\rm sig} = 1.0,\ 5.0,\ 20.0$ all reproduce standard normal distributions to a good approximation. This result proves that the model does not introduce any bias in the signal extraction.

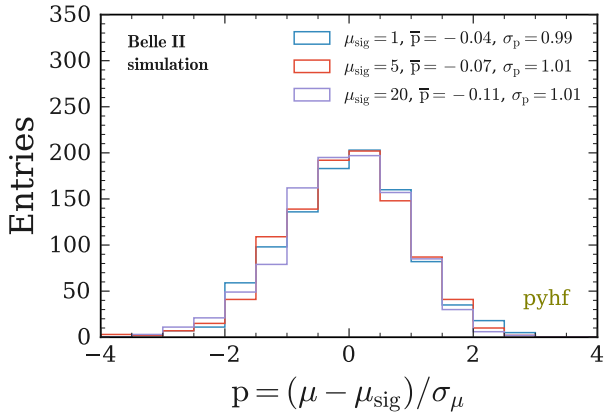


Figure 9.17: Distributions of the pulls, $\mathrm{p} = (\mu - \mu_{\rm sig})/\sigma_\mu$, from fits to the MC toys in the ensembles with injected signal of $\mu_{\rm sig} = 1.0,\ 5.0,\ 20.0$. The legend provides the mean value, $\bar{\rm p}$, and the standard deviation, $\sigma_{\rm p}$, computed from the set of pulls for each ensemble. [2]

9.3.2 Test of fit quality

The compatibility of the statistical model with the observed data is evaluated in a blinded test. In this test, the p-value of the fit to the data in the SR, CR1, CR2 and CR3 is estimated. To determine this p-value, fits are performed to an ensemble of toys simulating pseudo-observations centred on expectations. The observed bin counts $\boldsymbol{n}_{\text{obs}}$ in data are fluctuated into $\boldsymbol{n}'_{\text{obs}}$ following the Poisson distribution. The fluctuations $\Delta\boldsymbol{n}_{\text{obs}} = \boldsymbol{n}_{\text{obs}} - \boldsymbol{n}'_{\text{obs}}$ are derived and added to the expected bin counts $\boldsymbol{n}_{\text{exp}}$ to obtain the toy $\boldsymbol{n}_{\text{toy}} = \boldsymbol{n}_{\text{exp}} + \Delta\boldsymbol{n}_{\text{obs}}$. Moreover, Gaussian systematic variations are also generated and included in the toys. The fits to toys are performed with both `pyhf` and `sghf`. The p-value of the fit to data is estimated in both cases and the values are compared between the two methods. The results are illustrated in Fig. 9.18. In the figure, the left plot shows the distribution of the profile likelihood ratio in the fits to the toys with `pyhf`, represented by the blue histogram, and the result for the fit to data, which corresponds to the violet vertical line. The estimated p-value of the fit to the data is equal to 73%. Such a high value proves that the statistical model provides a good description of the observed data. Moreover, the blue histogram is in good agreement with a χ^2 distribution with 24 degrees of freedom (d.o.f.), one per bin of the statistical model. The right plot in Fig. 9.18 illustrates the results of the fit-quality test performed with the `sghf`. The estimated p-value is equal to 68% and it is in good agreement with the result from `pyhf`.

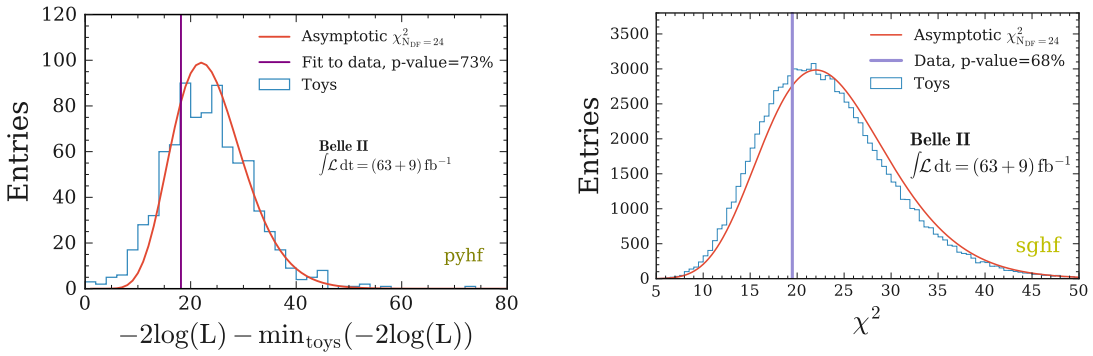


Figure 9.18: Test of the fit quality in the SR, CR1, CR2, CR3. Left plot: distribution of the profile likelihood ratios in fits to toys performed with the `pyhf` (blue step histogram). The histogram reproduces a χ^2 distribution (red solid line) with 24 d.o.f. (12 bins in SR-CR1 and 12 bins in CR2-CR3). The vertical violet line indicates the value of the profile likelihood ratio in the fit to the observed data, for which a p-value = 73% is estimated. Right plot: distribution of the χ^2 in fits to toys with `sghf` (blue step histogram). The histogram agrees to a good approximation with a χ^2 distribution with 24 d.o.f. (red solid line). The χ^2 value from the fit to the observed data is represented by the vertical violet line, and the corresponding p-value is equal to 68%. [2]

Chapter 10

Fit to the data and results

This chapter illustrates the results of the fit to the data. Section 10.1 describes the pre-unblinding results. The post-fit estimates of the nuisance parameters are provided, with particular focus on the normalisations of the background samples. The total uncertainty on the estimated signal strength μ is also determined. Section 10.2 focuses on the signal region unblinding. Post-fit yields are shown and the estimated $B^+ \to K^+\nu\bar{\nu}$ branching fraction is revealed. Moreover expected and observed upper limits are set with the CL_s method.

10.1 Pre-unblinding results

Post-fit shifts of the nuisance parameters in the statistical model are investigated before unblinding the value of the estimated signal strength μ. The shifts are evaluated with respect to the MC expectations, and a complete summary is provided in Table G.1 of Appendix G. The post-fit shifts of the normalisation uncertainty modifiers for the seven background samples are illustrated in Fig. 10.1. It is interesting to discuss these results, since the background normalisations are the largest sources of systematic uncertainty in the model. A conservative 50% pre-fit uncertainty is assigned to the normalisations of all the background yields, based on the observed 40% normalisation discrepancy reported in Sec. 9.2.1. Figure 10.1 shows that the fit to the data produces positive shifts of the parameters coupled to the continuum backgrounds. In particular, larger shifts are observed for the $c\bar{c}$ and $s\bar{s}$ background samples, which are the major sources of continuum background in the fit regions (see Fig. 9.5). No shift is observed for the parameters related to the yields of the charged and neutral B backgrounds, which represent the dominant background contributions in the bins of the SR more sensitive to signal (see Fig. 9.4). These results further support the argument of a mismodelling in the continuum simulation.
The uncertainty σ_μ on the estimated signal strength μ is also determined before the final unblinding. To estimate σ_μ, keeping μ blinded, a scan of the profile likelihood ratio $-2\ln\lambda_p(\mu)$ as a function of the signal strength variation $\mu - \mu_{\min}$ is performed. The result of the scan is illustrated in Fig. 10.2. The parabolic shape of the profile likelihood ratio scan is in good agreement with the parabola predicted by the Wilks' theorem [181]:

$$-2\ln\lambda_p(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma_\mu^2} + \mathcal{O}(1/\sqrt{N}), \tag{10.1}$$

where N is the size of the data sample. Hence, the points of the scan are fitted with a parabola to determine σ_μ. In particular, to account for the slight asymmetry visible in Fig. 10.2, the following fit function is used:

$$f(x) = \begin{cases} (x/\sigma_\mu^-)^2 & \text{if } x < 0\,, \\ (x/\sigma_\mu^+)^2 & \text{if } x \geq 0\,. \end{cases} \tag{10.2}$$

The estimated asymmetric uncertainties are

$$\sigma_\mu^- = 3.23, \quad \sigma_\mu^+ = 3.43\,. \tag{10.3}$$

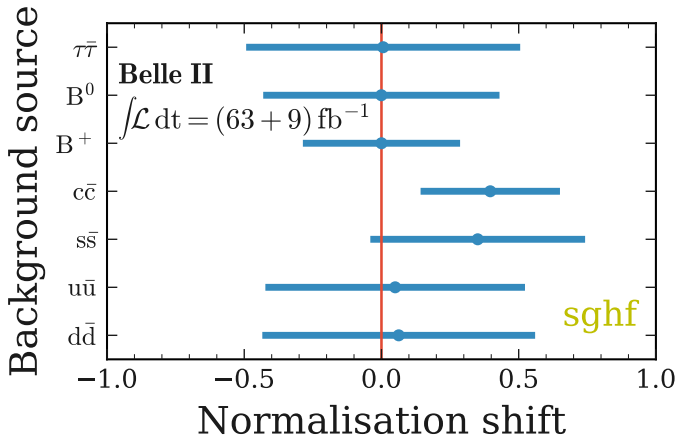


Figure 10.1: Post-fit shifts and uncertainties of the background-yield normalisations from the `sghf` fit to the observed data. The pre-fit uncertainty is 50% for each background sample. [2]

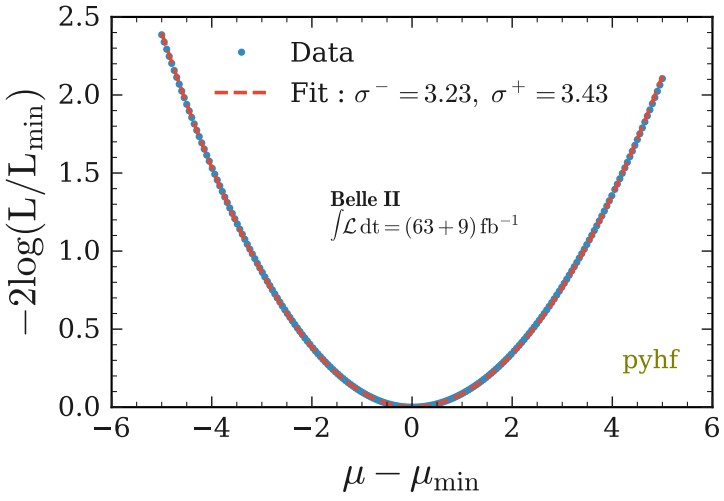


Figure 10.2: Profile likelihood ratio scan as a function of the signal strength variation $\mu - \mu_{\text{min}}$. Each point is obtained by fixing μ and minimising the likelihood with respect to all the other parameters of the statistical model. The asymmetric uncertainty on μ is estimated by fitting the collection of points with the asymmetric parabola in Eq. 10.2. The estimated uncertainties are $\sigma_\mu^- = 3.23$ and $\sigma_\mu^+ = 3.43$. [2]

10.2 Signal region unblinding

The first result of the signal region unblinding is given by the comparison of the post-fit yields with the yields in data. Figure 10.3 shows the observed yields in data and the yields estimated by the fit in the SR and CR1. The yields are shown individually for the charged and neutral B backgrounds and for the sum of the five continuum background samples. The estimated $B^+ \to K^+\nu\bar{\nu}$ yields are also plotted. The combined post-fit yields agree well with the observations in data. The signal purity in the SR is estimated to be equal to 6% and it increases to 22% in the region corresponding to three bins where $\text{BDT}_2 > 0.99$. Events from the continuum samples represent 59% of the background in the SR and 28% of the events with $\text{BDT}_2 > 0.99$. Figure 10.4 shows excellent agreement between observations in off-resonance data and post-fit yields in the CR2 and CR3.

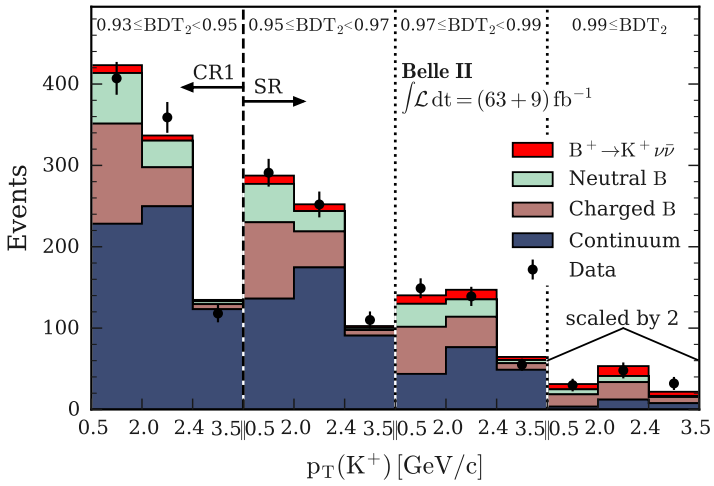


Figure 10.3: Yields in on-resonance data and as predicted by the simultaneous fit to the on- and off-resonance data, corresponding to an integrated luminosity of $63\,\mathrm{fb}^{-1}$ and $9\,\mathrm{fb}^{-1}$, respectively. The predicted yields are shown individually for charged and neutral B meson decays and the sum of the five continuum categories. The leftmost three bins belong to CR1 with $\mathrm{BDT}_2 \in [0.93, 0.95]$ and the other nine bins correspond to the SR, three for each range of $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$. Each set of three bins is defined by $p_T(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$. All yields in the rightmost three bins are scaled by a factor of 2. [1]

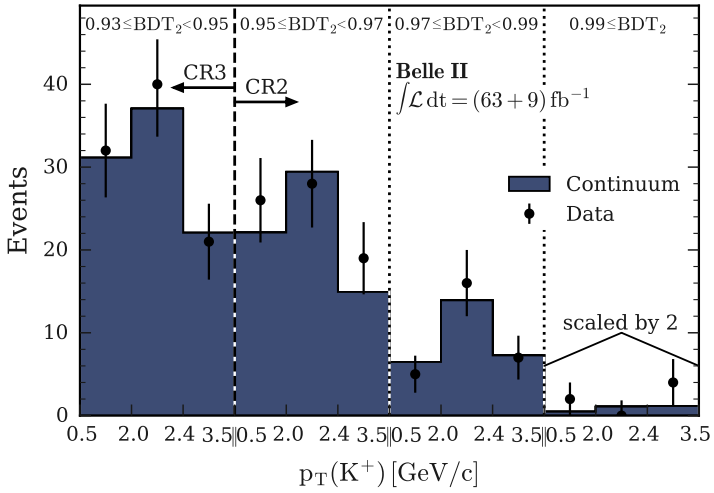


Figure 10.4: Yields in off-resonance data and as predicted by the simultaneous fit to the on- and off-resonance data, corresponding to an integrated luminosity of $63\,\mathrm{fb}^{-1}$ and $9\,\mathrm{fb}^{-1}$, respectively. The predicted yields of the five continuum categories are summed. The leftmost three bins belong to CR3 with $\mathrm{BDT}_2 \in [0.93, 0.95]$ and the other nine bins correspond to the CR2, three for each range of $\mathrm{BDT}_2 \in [0.95, 0.97, 0.99, 1.0]$. Each set of three bins is defined by $p_T(K^+) \in [0.5, 2.0, 2.4, 3.5]\,\mathrm{GeV}/c$. All yields in the rightmost three bins are scaled by a factor of two. [1]

The estimated value of the signal strength μ is then revealed. It is found to be

$$\mu = 4.2^{+3.4}_{-3.2} = 4.2^{+2.9}_{-2.8}(\mathrm{stat})^{+1.8}_{-1.6}(\mathrm{syst})\,. \tag{10.4}$$

The statistical uncertainty is evaluated from the estimates of μ in fits to an ensemble of 10^4 toys, in which fluctuations of the observed data are generated following the Poisson distribution. The uncertainties $\sigma^+_\mu(\mathrm{stat}) = 2.9$ and $\sigma^-_\mu(\mathrm{stat}) = 2.8$ are determined by defining an asymmetric 68% confidence interval

around the median μ. The asymmetric systematic uncertainties $\sigma_{\mu}^{+}(\text{syst}) = 1.8$ and $\sigma_{\mu}^{-}(\text{syst}) = 1.6$ are calculated by subtracting the statistical uncertainty in quadrature from the total uncertainty. An additional 10% theoretical uncertainty arising from the knowledge of the branching fraction in the SM is not included. The result in Eq. 10.4 can be translated into the corresponding estimated branching fraction:

$$\text{BR}(B^+ \to K^+\nu\bar{\nu}) = \left[1.9^{+1.6}_{-1.5}\right] \times 10^{-5} = \left[1.9^{+1.3}_{-1.3}(\text{stat})^{+0.8}_{-0.7}(\text{syst})\right] \times 10^{-5}\,. \tag{10.5}$$

The measured signal strength in Eq. 10.4 is compatible with the SM expectation, $\mu = 1$, at a CL of one standard deviation, and it is also compatible with the background-only hypothesis, $\mu = 0$, at a CL of 1.3 standard deviations. Hence, no evidence of signal can be claimed and expected and observed upper limits on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction are set. The limits are computed with the CL_s method, introduced in Sec. 3.5, using the the statistic $\tilde{q}_\mu$ and the asymptotic formulae presented in Sec. 3.5.2. The scan of the observed and expected CL_s as a function of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction is illustrated in Fig. 10.5. At 90% CL, the observed upper limit is found to be 4.1×10^{-5}, and the expected upper limit is 2.3×10^{-5}. The plot shows also the expected CL_s uncertainty bands, corresponding to $\pm 1\sigma$ (green band) and $\pm 2\sigma$ (yellow band) variations of the signal strength from the background-only hypothesis. The solid black line, defined by the observed CL_s values for the hypothesised values of the branching faction, lies within the 2σ band, indicating no statistically significant departure of the data from the expectation. The kink around a branching fraction of $\sim 2 \times 10^{-5}$ is a computational artifact of CL_s.

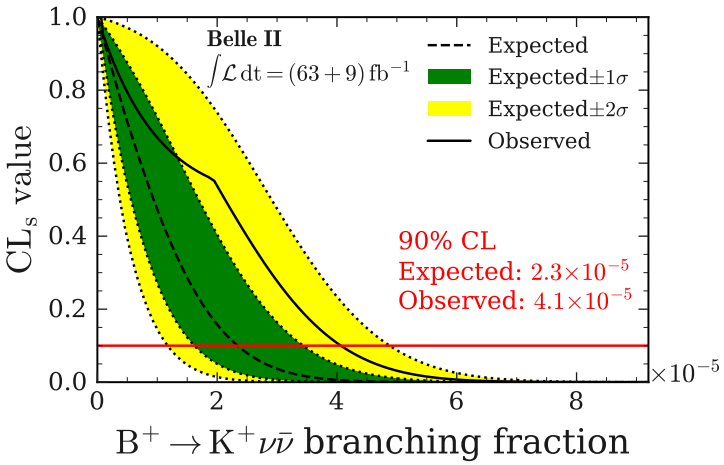


Figure 10.5: Expected and observed CL_s value as a function of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction and corresponding upper limits at 90% CL. The expected limit is derived for the background-only hypothesis. The observed limit is derived from the simultaneous fit to the on-resonance and off-resonance data. [1]

Chapter 11

Discussion of the results

This chapter discusses the results of the search for the $B^+ \to K^+\nu\bar{\nu}$ decay with an inclusive tagging method. In Section 11.1, the estimated value of the branching fraction is compared to the results obtained in previous searches at Belle and BaBar, and an updated average of the measurements is computed. Section 11.2 focuses on the performance of the inclusive tagging and presents a quantitative comparison with the explicit tagging techniques. In Section 11.3, the measurement outlook is discussed.

11.1 Comparison with previous measurements

The estimated value of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction reported in Eq. 10.5 is statistically compatible with the results of the measurements performed by previous experiments. A visual comparison is illustrated in Fig. 11.1.

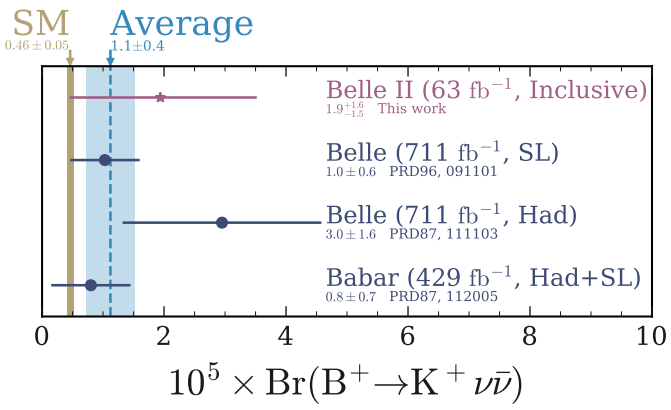


Figure 11.1: Values of the estimated $B^+ \to K^+\nu\bar{\nu}$ branching fraction in this search at Belle II and in previous measurements at Belle [123, 126] and BaBar [124] compared with the current Standard Model expectation [93]. The values reported for Belle are computed based on the quoted observed number of events and efficiency. The weighted average is computed assuming that uncertainties are uncorrelated. [1]

The plot shows that the most precise result remains the one obtained by Belle in 2017: $(1.03 \pm 0.57) \times 10^{-5}$. This value is determined from the observed signal yield and efficiency quoted in Ref. [126]. The measurement was performed using a data sample corresponding to an integrated luminosity of $711\,\mathrm{fb}^{-1}$, approximately 11 times larger than the sample used in the search presented in this thesis. The analysis adopted a semileptonic

tagging.
The second best result corresponds to the BaBar measurement of 2013: $(0.8 \pm 0.65) \times 10^{-5}$ [124]. The quoted value is the combination of the estimated $B^+ \to K^+\nu\bar{\nu}$ and $B^0 \to K^0\nu\bar{\nu}$ branching fractions. Moreover, this result is obtained by combining two statistically independent measurements: one performed with a hadronic tagging in $429\,\mathrm{fb}^{-1}$ of data, the other with a semileptonic tagging in $\sim 418\,\mathrm{fb}^{-1}$ of data [125].
The branching fraction estimated in the Belle measurement of 2013 is equal to $(2.95 \pm 1.63) \times 10^{-5}$. This value is determined from the number of observed signal events and signal efficiency reported in Ref. [123]. The analysis was performed with a hadronic tagging in $711\,\mathrm{fb}^{-1}$ of data. This result represents the largest estimated central value, and it is also $\sim 5\%$ less precise than the result obtained with the inclusive tagging using only $63\,\mathrm{fb}^{-1}$ of data.
The estimated values of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction can be averaged by means of a weighted least-squares procedure. Assuming uncorrelated uncertainties, weighted average and uncertainty combining the four measurements are computed as

$$\frac{\sum_i w_i x_i}{\sum_i w_i} \pm \frac{1}{(\sum_i w_i)^{1/2}}, \tag{11.1}$$

where

$$w_i = 1/(\sigma_{x_i})^2\,. \tag{11.2}$$

Here x_i and σ_{x_i} are the central value and uncertainty in the measurement i. Using Eq. 11.1, the average is found to be

$$\mathrm{BR}(B^+ \to K^+\nu\bar{\nu})|_{\mathrm{average}} = \overline{\mathrm{BR}} \pm \sigma(\overline{\mathrm{BR}}) = (1.1 \pm 0.4) \times 10^{-5}\,. \tag{11.3}$$

The average central value is found to be above the current SM expectation by a factor of 1.4,

$$\frac{\overline{\mathrm{BR}} - \mathrm{BR}_{\mathrm{SM}}}{\mathrm{BR}_{\mathrm{SM}}} = 1.4\,. \tag{11.4}$$

The sizeable average uncertainty, $\sigma(\overline{\mathrm{BR}}) = 0.4$, makes the average compatible with the SM hypothesis at a CL of 1.6 standard deviations and with the background-only hypothesis (null branching fraction) at a CL of 2.8 standard deviations.

11.2 Performance of the inclusive tagging

The main advantage of the inclusive tagging is the much larger signal efficiency compared to the techniques based on the full reconstruction of the tag B meson. As discussed in Sec. 1.6, the current hadronic and semileptonic taggings have typical efficiencies of $\mathcal{O}(0.1\%)$ and $\mathcal{O}(1\%)$, with slight variations depending on the level of purity. In searches relying on those taggings, the efficiency of the signal reconstruction is convolved with the low tagging efficiency, resulting in a total efficiency that can drop well below 1%. This limitation ruled the hadronic and semileptonic taggings out of this search with only $63\,\mathrm{fb}^{-1}$ of data. Instead, the use of the inclusive tagging provides a signal efficiency of 4.3% in the signal region (SR). A summary of the efficiencies in this work and in previous searches is given in Table 11.1.
It is also interesting to evaluate the signal efficiency of the inclusive tagging as a function of the generated dineutrino invariant mass squared q^2. Figure 11.2 shows the signal efficiency in the SR computed in q^2 intervals of $2\,\mathrm{GeV}^2/c^4$, going from $q^2 = 0$ up to the kinematic limit $q^2_{\mathrm{max}} = 22.9\,\mathrm{GeV}^2/c^4$. The efficiency is monotone decreasing as a function of q^2. A maximum signal efficiency of $\sim 13\%$ is observed for $0 \leq q^2 < 2\,\mathrm{GeV}^2/c^4$. The efficiency drops to zero for $q^2 > 16\,\mathrm{GeV}^2/c^4$. The plot in Fig. 11.2 provides a model-independent information about the measurement, which can be used in studies of new signal models. In particular, the inclusive tagging is highly efficient in the low q^2 region, where a possible dark-scalar narrow resonance is predicted by Ref. [117], as discussed in Sec. 1.5.3.1.
The performance of the inclusive tagging can be also evaluated in terms of measurement precision. Under

Measurement	Tagging	Signal efficiency
BaBar (2013), Ref. [124]	hadronic + semileptonic	$4.4 \times 10^{-4} + 1.6 \times 10^{-3}$
Belle (2013), Ref. [123]	hadronic	5.7×10^{-4}
Belle (2017), Ref. [126]	semileptonic	2.2×10^{-3}
Belle II (2021), this work	inclusive	4.3×10^{-2}

Table 11.1: Comparison between the signal efficiency in the signal region (SR) obtained in this analysis with the inclusive tagging and the efficiencies in the signal regions of previous searches with hadronic and semileptonic taggings. Precise definitions of the signal regions can be found in the cited papers.

the assumption that σ_{BR} scales as $1/\sqrt{L}$ [1], the uncertainties estimated in Refs. [123, 124, 126] can be scaled to $63\,\mathrm{fb}^{-1}$ to compare the inclusive tagging with the hadronic and semileptonic taggings adopted in the previous measurements. A summary of the comparison is given in Table 17.3. The inclusive tagging performs a factor of 2.5 better than the hadronic tagging used by Belle in 2013, $\sim 20\%$ better than the semileptonic tagging used by Belle in 2017 and $\sim 10\%$ better than the combination of hadronic and semileptonic taggings used by BaBar in 2013.

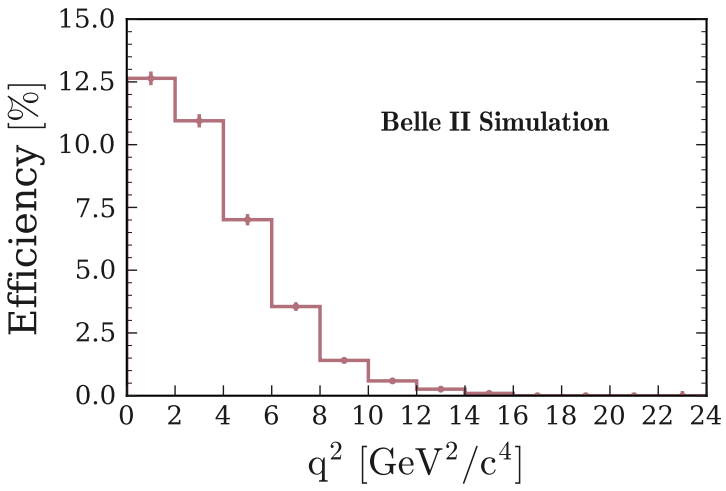


Figure 11.2: Signal efficiency evaluated in bins of generated dineutrino invariant mass squared, q^2, for simulated $B^+ \to K^+ \nu\bar{\nu}$ events in the signal region (SR). Each bin corresponds to an interval of $2\,\mathrm{GeV}^2/c^4$, and the range extends up to the kinematic limit $q^2_{\mathrm{max}} = 22.9\,\mathrm{GeV}^2/c^4$. The error bars indicate the statistical uncertainty.

Measurement	$L\,(\mathrm{fb}^{-1})$	$\sigma_{\mathrm{BR}} \times 10^5$	$\sigma_{\mathrm{BR}}^{\mathrm{scaled}} \times 10^5$
BaBar, hadronic and semileptonic taggings [124]	429	0.65	1.70
Belle, hadronic tagging [123]	711	1.63	5.47
Belle, semileptonic tagging [126]	711	0.57	1.91
Belle II, inclusive tagging (this work)	63	1.55	1.55

Table 11.2: Uncertainties estimated with different tagging techniques in samples of integrated luminosity L and corresponding values scaled to $63\,\mathrm{fb}^{-1}$.

[1] In this assumption, given the uncertainty $\sigma_{\mathrm{BR,i}}$ estimated for an integrated luminosity L_i, the extrapolated uncertainty in a data sample of integrated luminosity L_f is $\sigma_{\mathrm{BR,f}} = \sigma_{\mathrm{BR,i}} \cdot \sqrt{\frac{L_i}{L_f}}$.

The application of the inclusive tagging can be extended to the search of other rare B decays. It is interesting to notice that in the last three bins of the SR, namely those where $\mathrm{BDT}_2 \geq 0.99$, approximately 7.5% of the decays in the signal side of the charged B background are $B^+ \to \tau^+\nu$ decays, as illustrated by the chart in Fig. 11.3. This process represents another *golden mode* of the physics program at B factories and its precise measurement allows to probe the SM and to constrain specific BSM models, such as those predicting the existence of a charged Higgs boson [207, 208]. The current world average of the $B^+ \to \tau^+\nu$ branching fraction is equal to $(1.09 \times 0.24) \times 10^{-4}$ [43]. The experimental signature of this decay is often characterised by a single charged track, making it similar to the $B^+ \to K^+\nu\bar{\nu}$ final state.

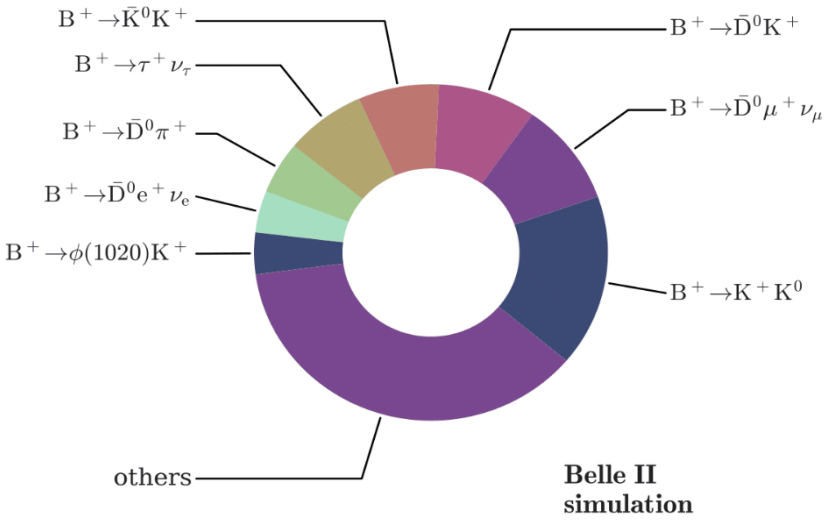


Figure 11.3: Major background decays in the signal side of the reconstructed charged B background in the last three bins of the SR ($\mathrm{BDT}_2 \geq 0.99$).

11.3 Outlook

In the light of the recent 3.1σ evidence of lepton universality violation in $b \to s l^+ l^-$ transitions reported by the LHCb collaboration [34], the discovery of the $B \to K^{(*)}\nu\bar{\nu}$ decays becomes even more compelling. Given the relation between $b \to s\nu\bar{\nu}$ and $b \to s l^+ l^-$ transitions, if new physics is the source of the anomaly observed by LHCb, then it should have an imprint also in $B \to K^{(*)}\nu\bar{\nu}$ decays. Moreover, if the central value of the $B^+ \to K^+\nu\bar{\nu}$ decay remains in agreement with the current value of the average in Eq. 11.3, a discovery of the decay would provide clean evidence of the existence of physics beyond the SM.
To evaluate the future prospects, sensitivity projections can be made for $B^+ \to K^+\nu\bar{\nu}$ searches with the inclusive tagging. Two scenarios are explored: a *baseline* scenario, which assumes no improvements of the method used in this measurement; and an *improved* scenario, corresponding to a 50% increase in signal efficiency at the same background level. Such improvement could be achieved, for example, with the combination of the inclusive tagging with the hadronic and semileptonic tagging methods. In the two scenarios, the uncertainty on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction measured in this work, $\sigma_{\mathrm{BR}} = 1.55 \times 10^{-5}$, is scaled to a set of luminosity milestones for the Belle II experiment, assuming that the main source of uncertainty continues to be of statistical origin. The projections are summarised in Table 11.3, where the scaled uncertainties are normalised to the central value of the SM branching fraction.
Based on the extrapolations in Table 11.3, for a data sample of $10\,\mathrm{ab}^{-1}$ the $B^+ \to K^+\nu\bar{\nu}$ analysis with the inclusive tagging is sensitive to the SM signal rate at 3σ level in the baseline scenario and at 5σ level in the improved scenario. In both cases, the accuracy of the SM prediction would be reached with the complete Belle II data sample of $50\,\mathrm{ab}^{-1}$. The projections are even more optimistic if further improvements of the

$L(\mathrm{ab}^{-1})$	Baseline scenario	Improved scenario
1	0.85	0.69
5	0.38	0.31
10	0.27	0.22
50	0.12	0.10

Table 11.3: Projections, in the baseline and improved scenarios, of the uncertainty on the $B^+ \to K^+\nu\bar{\nu}$ branching fraction (relative to the SM central value) as a function of the integrated luminosity L. The values are obtained by assuming that the observed uncertainty in this work, $\sigma_{\mathrm{BR}} = 1.55 \times 10^{-5}$, scales as $1/\sqrt{L}$.

analysis are taken into account [209]. Preliminary investigations reveal that a $\mathcal{O}(20-30\%)$ increase in sensitivity may be gained by replacing BDT_2 with a neural network consisting of multiple fully connected layers. Further improvement depends also on the reduction of the uncertainty on the background yield normalisation, which is currently the dominant systematic uncertainty in the analysis. Moreover, upgrades are foreseen in the construction of the ROE and in the identification and suppression of K_L^0 mesons, which can improve the reconstruction of the missing energy from neutrinos.
In conclusion, the prospects of the search for the $B^+ \to K^+\nu\bar{\nu}$ decay with the inclusive tagging are encouraging and a discovery might be possible in the near future.

Part III

Search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay with an inclusive tagging method

Chapter 12

Data samples

This chapter presents the data samples used in this search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay. Section 12.1 describes the collision data sample for which this search is optimised. Section 12.2 introduces the simulated data with particular focus on the signal sample.

12.1 Collision data

This analysis is developed and optimised for a search of the $B^0 \to K^{*0}\nu\bar{\nu}$ decay in the on-resonance data sample of $189\,\text{fb}^{-1}$ [5] recorded by Belle II between 2019 and 2021. The estimated number of $B\bar{B}$ pairs in this sample is

$$N_{B\bar{B}} = [197.2 \pm 5.7] \times 10^6\,, \tag{12.1}$$

determined with the procedure based on continuum subtraction described in Ref. [189].
In the analysis, an off-resonance data sample is also used to characterise the continuum background from light-quark pair production and $e^+e^- \to \tau^+\tau^-$ processes. This sample is equivalent to an integrated luminosity of $18\,\text{fb}^{-1}$ and is collected at an energy $60\,\text{MeV}$ below the $\Upsilon(4\text{S})$ resonance.

12.2 Simulated data

Samples produced in the 14th official MC simulation campaign at Belle II are used in this search. Simulated samples include events from seven background categories: the five continuum categories of $e^+e^- \to q\bar{q}$ ($q = u,\, d,\, c,\, s$) processes, and tau pair production, $e^+e^- \to \tau^+\tau^-$, plus two categories corresponding to generic $B^0\bar{B}^0$ and B^+B^- decays produced on resonance. Details about the event generators used to simulate the background processes are given in Table 4.1. The simulated samples are grouped into batches, each equivalent to an integrated luminosity of $100\,\text{fb}^{-1}$, where the realistic mixture of events recorded by Belle II is reproduced.

12.2.1 Simulated signal

The signal MC simulates $\Upsilon(4\text{S}) \to B^0\bar{B}^0$ events in which one of the two B-meson decays corresponds to $B^0 \to K^{*0}(\to K^+\pi^-)\nu\bar{\nu}$ (the branching fraction of $K^{*0} \to K^+\pi^-$ is 66.6% [43]).
In the official simulation, the `EVTGEN` [157] model used to generate $B^0 \to K^{*0}\nu\bar{\nu}$ decays is `PHSP`, which describes the kinematics of the decay products by means of a generic Lorentz-invariant three-body phase space of the form in Eq. 4.2. Given that this analysis targets the discovery of the SM process, simulated $B^0 \to K^{*0}\nu\bar{\nu}$ decays are reweighted according to the SM prediction. In the SM the kinematics of the decay is described by the double differential branching fraction in $q^2 = (p_\nu + p_{\bar{\nu}})^2$ and θ given in Eq. 1.82. The angular dependence splits the branching fraction in the sum of two separate contributions: the longitudinal branching fraction, $d\text{BR}_L/dq^2$, depending on the SM form factor $A_1(q^2)$ and the transverse branching fraction,

$d\mathrm{BR}_T/dq^2$, which depends on the sum of form factors $A_{12}(q^2) + V(q^2)$. The form factors are parametrised by series expansions, as in Eq. 1.78, truncated after the quadratic term and fitted over the allowed q^2 range to Lattice QCD and LCSR results. The values of the nine real parameters $(\alpha_0, \alpha_1, \alpha_2)_i$, $i = A_1, A_{12}, V$, and the corresponding 9×9 covariance matrix are extracted from the fit. In this analysis, the fit results are taken from Ref. [210].
The signal reweighting is implemented following the same procedure described in Sec. 4.2.1. As a first step, the computation of the branching fraction based on the estimated form factors is validated. The double differential branching fraction is integrated in bins of q^2 and the results are compared with the values reported in Table 12.1, which are taken from Ref. [85]. An illustration of this validation is provided in Fig. 12.1. The uncertainties on the computed values of the branching fraction correspond to the width of the rectangles and are derived from the decomposition of the covariance matrix with the technique described in Sec. 3.2.2.1. The computations agree within the uncertainties with the predictions of Ref. [85] and the expected shape in q^2 is well reproduced.
The q^2 and $\cos\theta$ distributions for $B^0 \to K^{*0}\nu\bar{\nu}$ decays in the official phase-space simulation and computed according to the SM expectation are shown in Fig. 12.2. Significant shape differences are visible. The phase-space distribution in q^2 is monotonic decreasing, with the majority of the events characterised by low q^2. The SM predicts an increasing distribution up to $q^2 \sim 13\,\mathrm{GeV}^2/c^4$, which drops to zero towards the kinematic limit $q^2_{\mathrm{max}} = 19.2\,\mathrm{GeV}^2/c^4$. The $\cos\theta$ distribution in the PHSP simulation is uniform, instead the expected SM distribution reproduces the parabolic shape described by the differential branching fraction in Eq. 1.82.

$q^2[\mathrm{GeV}^2/c^4]$	$10^6 \times \mathrm{BR}(B^0 \to K^{*0}\nu\bar{\nu})_{\mathrm{SM}}$
$0 - 4$	$1.38 \pm 0.21 \pm 0.07$
$4 - 8$	$1.88 \pm 0.22 \pm 0.10$
$8 - 12$	$2.27 \pm 0.22 \pm 0.12$
$12 - 16$	$2.36 \pm 0.18 \pm 0.13$
$16 - q^2_{\mathrm{max}}$	$1.30 \pm 0.10 \pm 0.07$

Table 12.1: SM branching fractions of the $B^0 \to K^{*0}\nu\bar{\nu}$ decay in five q^2 bins, from 0 to the kinematic limit $q^2_{\mathrm{max}} = 19.2\,\mathrm{GeV}^2/c^4$, as computed in Ref. [85] The first error is the uncertainty related to the form factors and the second error quantifies the parametric uncertainties.

The plots in Fig. 12.3 illustrate the expected distributions in q^2 for the longitudinal and transverse polarisations of the K^{*0} meson. The ratios to the phase-space distribution are computed and interpolated using a SciPy spline with smoothness parameter equal to 5×10^{-3} [190, 191]. The results are shown in Fig. 12.4. The splines, $\mathrm{spl}_{\mathrm{T}}(q^2)$ for the longitudinal polarisation and $\mathrm{spl}_{\mathrm{L}}(q^2)$ for the transverse polarisation, are used to compute the weight

$$w(q^2, \cos\theta) = \frac{3}{4}\mathrm{spl}_{\mathrm{T}}(q^2)(1 - \cos^2\theta) + \frac{3}{2}\mathrm{spl}_{\mathrm{L}}(q^2)\cos^2\theta, \tag{12.2}$$

derived from Eq. 1.82, which is applied to the simulated signal events in the analysis.
For a closure test, the weight in Eq. 12.2 is computed for each event in the signal MC sample used to implement the reweighting procedure. The application of the weights to the phase-space simulation is illustrated in Fig. 12.5. The reweighted phase-space distributions in q^2 and $\cos\theta$ accurately reproduce the expected SM distributions shown in Fig. 12.2.

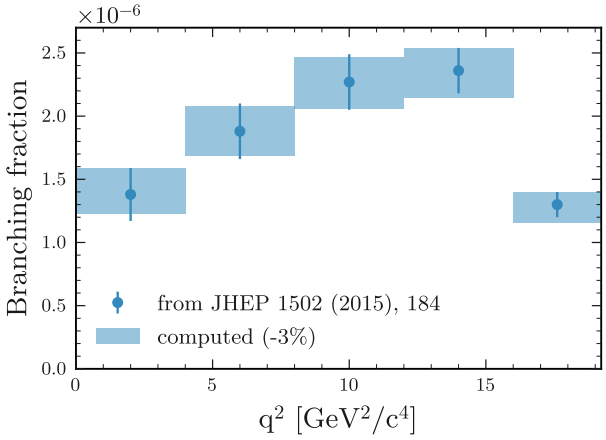


Figure 12.1: SM branching fractions of the $B^0 \to K^{*0}\nu\bar{\nu}$ decay in five q^2 bins, from 0 to the kinematic limit $q^2_{\text{max}} = 19.2\,\text{GeV}^2/c^4$. The values from Table 12.1 are shown as blue dots and are compared to the values computed for the reweighting procedure, shown as shaded blue rectangles. The computations are based on the values of $(\alpha_0, \alpha_1, \alpha_2)_i$, $i = A_1, A_{12}, V$, extracted in the fit presented in Ref. [210], and on the values of the SM parameters provided in Ref. [85]. The central values are shifted by -3% for illustration purposes. The rectangle widths correspond to the form factor uncertainty.

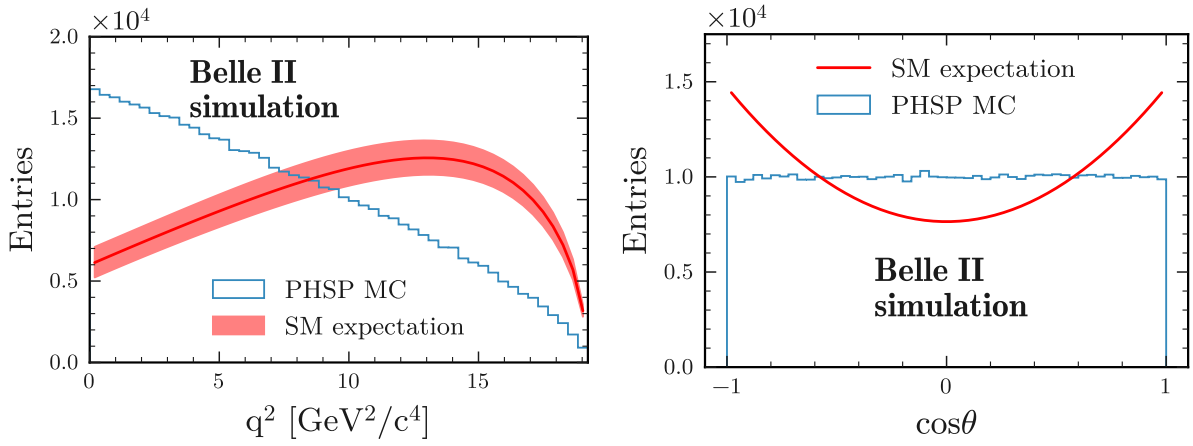


Figure 12.2: Number of $B^0 \to K^{*0}\nu\bar{\nu}$ decays as a function of q^2 (left plot) and $\cos\theta$ (right plot) in the phase-space simulation (blue histogram) and according to the SM expectation derived from Refs. [85, 210] (red line). The predictions are scaled to the total number of generated events (5×10^5). The red band corresponds to uncertainty related to the form factors.

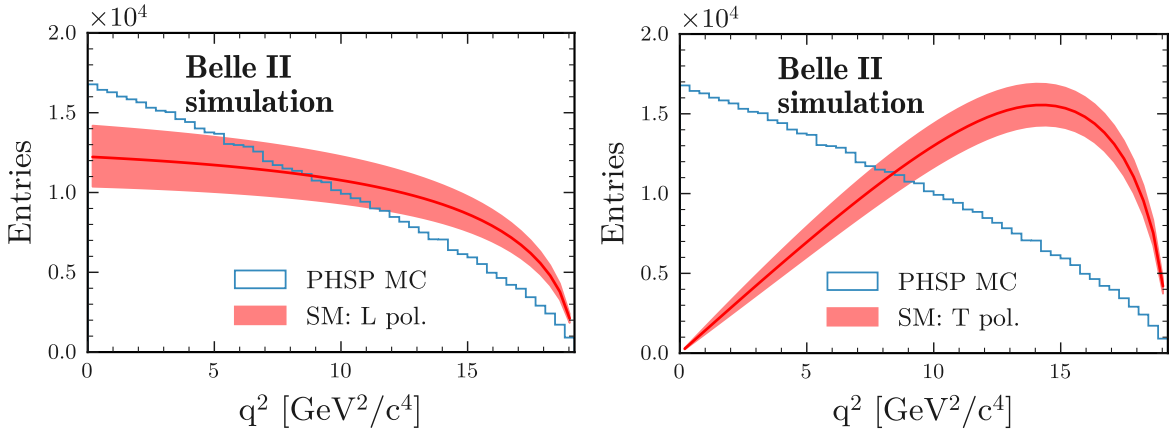


Figure 12.3: Number of $B^0 \to K^{*0}\nu\bar{\nu}$ decays as a function of q^2 in the phase-space simulation (blue histogram) and and according to the SM expectation derived from Refs. [85, 210] (red line) for longitudinally polarised K^{*0} mesons (left) and transversely polarised K^{*0} mesons (right). The predictions are scaled to the total number of generated events (5×10^5). The red band in the plots corresponds to the uncertainty from the form factors evaluation.

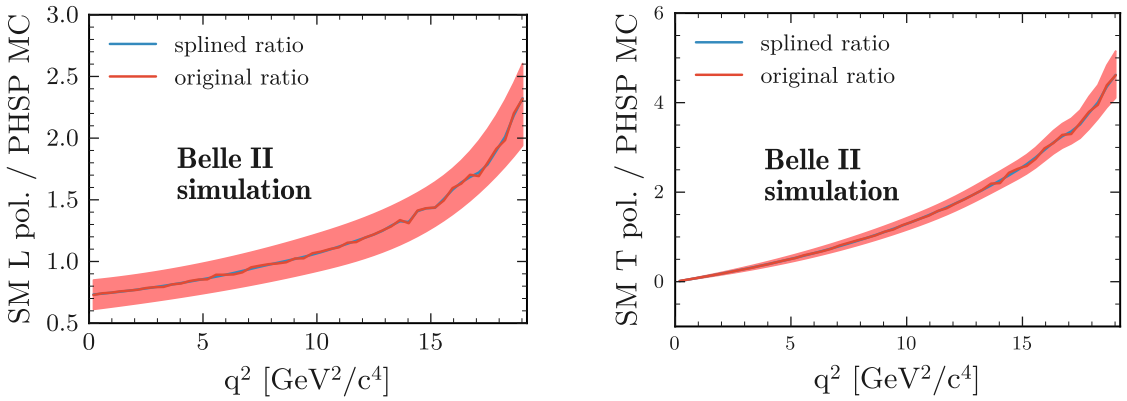


Figure 12.4: Ratios between the SM and the phase-space q^2 distributions (red line) and splined ratios (blue line). The ratios obtained for the longitudinal (transverse) polarisation of the K^{*0} are shown in the left (right) plot. The red band in the plots corresponds to the theoretical uncertainty from the form factors evaluation.

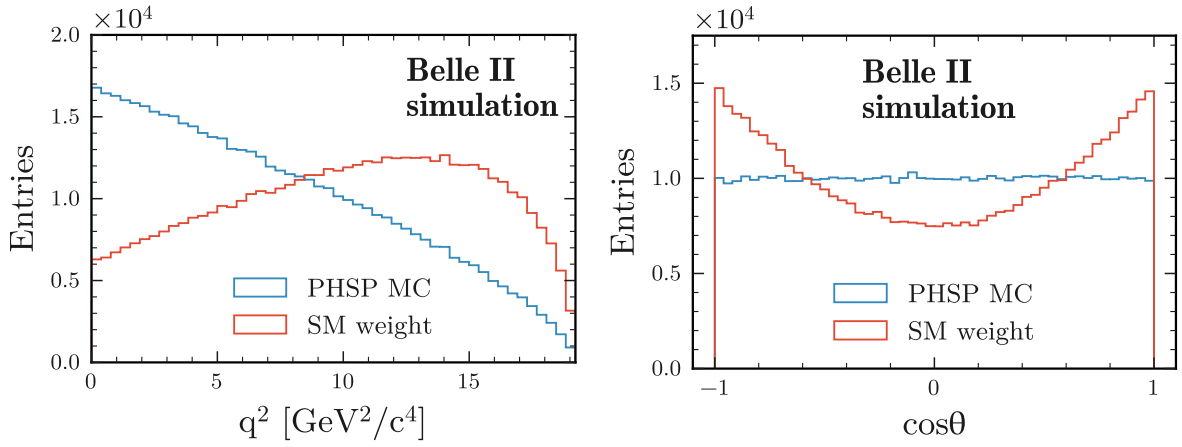


Figure 12.5: Distributions of the generated q^2 (left) and $\cos\theta$ (right) for (5×10^5) $B^0 \to K^{*0}\nu\bar{\nu}$ events in the phase-space simulation (blue step histogram) and after the application of the weights from Eq. 12.2 (red step histogram).

Chapter 13

$B^0 \to K^{*0}\nu\bar{\nu}$ reconstruction

This chapter presents the reconstruction procedure implemented in this search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay with the inclusive tagging. The chapter also describes in detail the variables selected to characterise the signal and distinguish it from the background. Section 13.1 introduces the selection of reconstructed tracks and photon clusters in the event. Section 13.2 describes the signal selection. The construction of the ROE is discussed in Section 13.3. The event selection is defined in Section 13.4. Finally, Section 13.5 presents the discriminating variables grouped by category.

13.1 Object selection

In the object-selection stage charged tracks and photon clusters are selected. The *track and photon cleanups* implemented in the search for the $B^+ \to K^+\nu\bar{\nu}$ decay presented in the previous part of the thesis are used also in this analysis. The selections are listed again in this section to facilitate the reading. Definitions of the observables used in the selections are provided in Sec. 5.2.
The reconstructed tracks are required to satisfy the conditions of the so-called *track cleanup*:

$$p_T > 0.1\,\mathrm{GeV}/c, \tag{13.1}$$

$$E < 5.5\,\mathrm{GeV}, \tag{13.2}$$

$$|dz| < 3\,\mathrm{cm}, \tag{13.3}$$

$$dr < 0.5\,\mathrm{cm}, \tag{13.4}$$

$$17^\circ < \theta < 150^\circ. \tag{13.5}$$

The photon candidates are ECL clusters without an attached track, and their selection is based on the so-called *photon cleanup*:

$$0.1\,\mathrm{GeV} < E_{\mathrm{ECL}} < 5.5\,\mathrm{GeV}, \tag{13.6}$$

$$17^\circ < \theta < 150^\circ. \tag{13.7}$$

After the introduction of the object selection, the signal efficiency, evaluated in the reconstruction of simulated $B^0 \to K^{*0}\nu\bar{\nu}$ events, is found to be

$$(43.0 \pm 0.4)\%. \tag{13.8}$$

Here and in the following, the signal efficiency is computed with respect to the complete $B^0 \to K^{*0}\nu\bar{\nu}$ decay.

13.2 Signal selection

Signal K^{*0} candidates are constructed with two charged tracks corresponding to the candidate signal K^+ and π^- mesons. Both the tracks are required to have at least one PXD hit, $\mathrm{nPXDHits} > 0$. This constraint helps to reduce the background contamination, without worsening the signal selection efficiency (the PXD hit reconstruction efficiency is $> 98\%$ [193]). The track of the signal K^+ candidate is also required to satisfy the PID selection $\mathrm{kaonID} > 0.9$. Following the guidelines of the Belle II PID performance group, the track is also required to have at least 20 CDC hits, $\mathrm{nCDCHits} > 20$. The signal K^+ and π^- candidates are fitted to a common K^{*0} vertex using the `TreeFitter` algorithm [175]. The candidate K^{*0} vertices are subject to the following selection criteria:

$$0.8\,\mathrm{GeV}/c^2 < M(K^{*0}) < 1.0\,\mathrm{GeV}/c^2, \tag{13.9}$$

$$p_{\mathrm{fit}} > 0.1\%, \tag{13.10}$$

$$dr_{K^{*0}} < 0.1\,\mathrm{cm}, \tag{13.11}$$

where $M(K^{*0})$ is the invariant mass resulting from the kinematic fit, p_{fit} is the χ^2 probability of the vertex fit, and $dr_{K^{*0}}$ is the radial distance of the vertex from the average IP position. With this selection, the signal efficiency drops to

$$(25.1 \pm 0.3)\%. \tag{13.12}$$

A summary of the signal efficiency evaluated in simulated signal events as a function of the cuts implemented in the signal candidate selection is given in Table 13.1.

Cut in the signal candidate selection	Signal efficiency (%)
$K^+(\mathrm{nPXDHits} > 0 \text{ and } \mathrm{nCDCHits} > 20)$ and $\pi^-(\mathrm{nPXDHits} > 0)$	39.4 ± 0.4
$K^+(\mathrm{kaonID} > 0.9)$	28.4 ± 0.4
$0.8\,\mathrm{GeV}/c^2 < M(K^{*0}) < 1.0\,\mathrm{GeV}/c^2$	25.2 ± 0.3
$p_{\mathrm{fit}} > 0.1\%$	25.2 ± 0.3
$dr_{K^{*0}} < 0.1\,\mathrm{cm}$	25.1 ± 0.3

Table 13.1: Signal efficiency as a function of the cuts implemented in the signal candidate selection.

The remaining K^{*0} candidates in the event are ranked based on the reconstructed di-neutrino invariant mass squared, which is computed as

$$q^2_{\mathrm{rec}} = M^2_{B^0} + M^2_{K^{*0}} - 2M_B E^*_{K^{*0}}\,, \tag{13.13}$$

where M_{B^0} and $M_{K^{*0}}$ are the PDG values of the B^0 and K^{*0} masses [43], and $E^*_{K^{*0}}$ is the reconstructed energy of the signal K^{*0} candidate, evaluated in the CMS. The relation in Eq. 13.13 is derived in the assumption that the signal B^0 meson can be considered, to a good approximation, at rest in the CMS. The q^2_{rec} variable provides a good estimate of the generated q^2, as shown by Fig. 13.1, and it is inversely proportional to the reconstructed transverse momentum, which was previously used in the signal selection (see Sec. 5.3). The replacement simplifies the interpretation of the analysis results and the comparison with theoretical predictions, which are usually provided in terms of q^2.
After the ranking, only the K^{*0} candidate corresponding to the lowest q^2_{rec} is kept. Studies in simulated signal events show that, with this selection, a true signal K^{*0} in the signal acceptance is chosen in 87% of the cases. A complete summary as a function of the q^2_{rec} rank is illustrated in Fig. 13.2. The signal efficiency corresponding to the selection of only one signal K^{*0} candidate per event is found to be

$$(21.8 \pm 0.3)\%. \tag{13.14}$$

The q^2_{rec} distributions computed for the signal K^{*0} candidate in simulated signal and background events, as well as in data from a single collision run, are shown in Fig. 13.3.

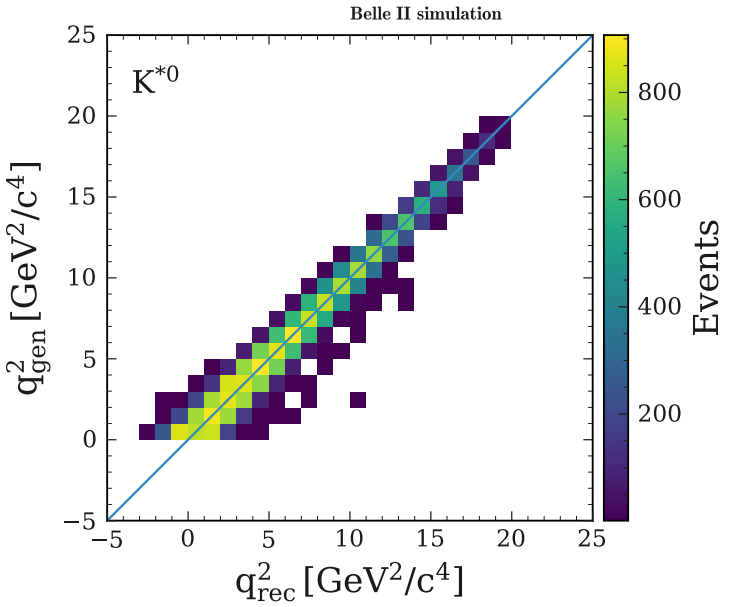


Figure 13.1: Generated di-neutrino invariant mass squared, q^2_{gen}, as a function of the reconstructed di-neutrino invariant mass squared, q^2_{rec}, computed in simulated $B^0 \to K^{*0}\nu\bar{\nu}$ events.

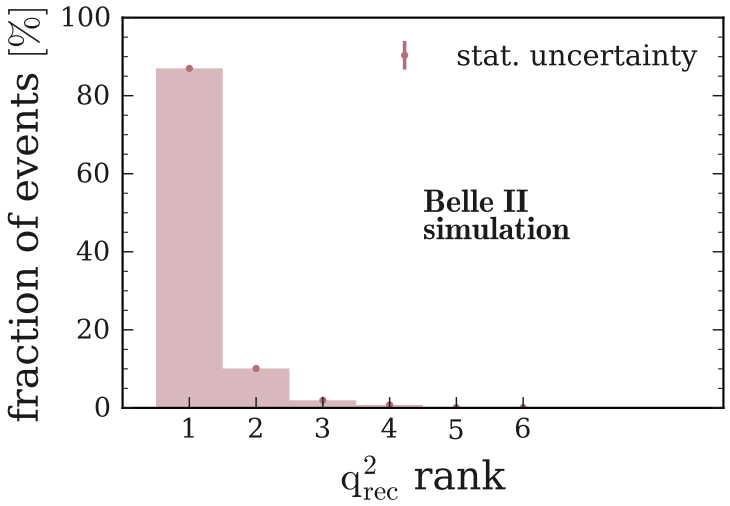


Figure 13.2: Fraction of simulated signal events in which a true signal K^{*0} in the signal acceptance corresponds to the candidate with q^2_{rec} rank i. The adopted q^2_{rec} ranking orders the signal kaon candidates from lowest to highest q^2_{rec}, so that the rank $i = 1$ identifies the candidate with lowest q^2_{rec}.

13.3 Rest of the event

All the reconstructed tracks not associated to the signal K^{*0} candidate, and all the photon clusters in the event, make up the so-called rest of the event (ROE). By default, reconstructed tracks are fitted using the pion-mass hypothesis, but all the ROE tracks have also a *most likely* particle hypothesis determined by the comparison between the global PID values assigned to the track and the values of the PID prior probabilities listed in Tab. 5.1 (see Sec. 5.4 for more details). In this setup, variables characterising the ROE kinematics

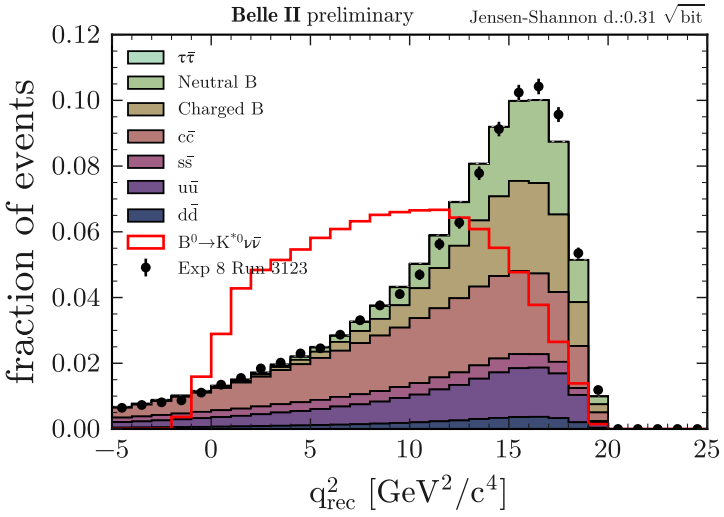


Figure 13.3: Distributions of q^2_{rec} for the signal K^{*0} candidate computed in simulated signal and background events and in data collected in a single collision run. The distributions are normalised to unit areas.

are computed. These variables are presented in Sec. 13.5.5.
The ROE is used together with the signal K^{*0} candidate to compute observables describing the kinematics and the shape of the whole event. Moreover, as done in the $B^+ \to K^+\nu\bar{\nu}$ analysis, the tracks of the ROE are fitted to a common vertex. This allows us to compute specific variables, which are very powerful to separate signal from background. A complete description of these variables is given in Sec. 13.5.6.

13.3.0.1 D mesons

Semileptonic B decays, $B \to Dl\nu$, where D represents a D^0 or a D^+ meson, or one of the excited $D^*(2007)^0, D^*(2010)^+$ mesons, are a major source of background also in this analysis. The final state of D^0 and D^+ hadronic decays is often characterised by an experimental signature similar to that of the signal K^{*0} decay. The K^+ and π^- tracks produced by D mesons can be wrongly selected to form a signal K^{*0} candidate. Therefore, the idea is to explicitly reconstruct and characterise D-meson candidates in the event, using a procedure similar to the one described in Sec. 5.4.0.1. Also in this analysis, candidate D vertices are fitted using the `KFit` algorithm [176].
To reconstruct D^0 candidates, the candidate K^+ track used to reconstruct the signal K^{*0} is paired with an ROE track of opposite charge, and the two tracks are fitted to a common vertex. The most-likely particle hypothesis is assumed for the ROE track, and it is required to have at least one PXD hit. The best D^0 candidate is chosen as the vertex fit with the highest χ^2 probability. D^0 candidates are also reconstructed by fitting to the same vertex the signal K^{*0} candidate and two pion tracks in the ROE having opposite charges. The ROE tracks are required to have at least one PXD hit. Only the D^0 candidate corresponding to the best vertex fit is kept.
Candidate D^+ mesons are reconstructed in vertex fits combining the signal K^{*0} candidate with an ROE track. The most-likely particle hypothesis is assumed for the ROE track, and it must have at least one PXD hit. The best D^+ candidate is selected as the one for which the χ^2 probability of the vertex fit is the highest. Specific observables characterising the properties of the reconstructed D^0 and D^+ candidates in signal and background events are presented in Sec. 13.5.7.

13.4 Event selection

Additional requirements constrain the track multiplicity and the event kinematics. The total number of tracks in the event, $\mathrm{nTracks}$, accounting for the ROE tracks and the pair of tracks used to construct the signal K^{*0} candidate, must satisfy the cut

$$4 < \mathrm{nTracks} < 11. \tag{13.15}$$

The upper bound of 11 tracks is a loose requirement. In signal events the final state of the signal B decay presents only two tracks, and in generic events about 11 tracks are expected on average [4]. The lower bound of 4 tracks is chosen to suppress background events with final states characterised by low track multiplicity. The contribution of the low-multiplicity background is also reduced by requiring that

$$17^\circ < \theta(\vec{p}_{miss.}) < 160^\circ \tag{13.16}$$

and

$$E_{\mathrm{CMS}} > 4.0\,\mathrm{GeV}, \tag{13.17}$$

where $\theta(\vec{p}_{miss.})$ is the polar angle of the missing three-dimensional momentum $\vec{p}_{miss.}$, defined in Eq. 5.16, and E_{CMS} is the total CMS energy.
Following the application of the requirements presented in this section, the computed signal efficiency is $(19.3 \pm 0.3)\%$. A summary providing the values of the signal efficiency at each selection stage in the reconstruction algorithm is given in Table 13.2.

Selection stage	Signal efficiency (%)
Object selection	43.0 ± 0.4
Signal selection	21.8 ± 0.3
Event selection	19.3 ± 0.3

Table 13.2: Signal efficiency computed at each major selection stage implemented in the reconstruction algorithm of the inclusive tagging.

13.5 Discriminating variables

In the inclusive tagging, the separation of signal from background relies on a set of discriminating variables computed during the event reconstruction. They can be grouped into categories collecting variables of the same type or used for the same purpose:

- variables used for the initial event filtering (12 variables, Sec. 13.5.1);
- event-based variables (6 variables, Sec. 13.5.2);
- event-shape variables (6 variables, Sec. 13.5.3);
- variables related to the signal K^{*0} candidate (6 variables, Sec. 13.5.4);
- variables related to the tracks and energy deposits in the ROE (2 variables, Sec. 13.5.5) and to the fit of the ROE vertex (3 variables, Sec. 13.5.6);
- variables related to the suppression of the D^0 and D^+ meson background (14 variables, Sec. 13.5.7.)

The variables are computed in a simulation sample made of 1.6×10^6 reconstructed signal events and 1.6×10^6 reconstructed events of each of the background categories. For each variable, the distribution obtained by stacking the background histograms according to the relative cross sections of Table 4.1 is compared to the signal distribution and also to the distribution from data collected during a single collision run (run 3123 of experiment 8). This preliminary investigation of the data to MC agreement allows discarding variables that are not well simulated. All the variables used in the analysis are presented in the following sections. In the plots, the distributions are always normalised so that the areas under the histograms integrate to unity. The divergence between the histograms is measured by means of the J-S distance [194].

13.5.1 Event-filtering variables

This category consists of variables related to the event shape and the ROE kinematics. The variables allow filtering a significant fraction of background events, mostly from continuum processes, and present features that are common to all the $B \to K^{(*)}\nu\bar{\nu}$ decays. Therefore, only a general overview is given in this section, and more details about the specific features of the signal and background distributions can be found in Sec. 5.7.2 and Sec. 5.7.4 of the $B^+ \to K^+\nu\bar{\nu}$ analysis.
Twelve variables are used for the initial event filtering.

- Three variables describe the ROE kinematics: the ROE ΔE, the magnitude of the ROE momentum $\vec{P}$ and its polar angle $\theta(\vec{P})$.
- Two variables are: the cosine of the event thrust-axis polar angle and the cosine of the angle between the thrust axis of the signal B candidate and the thrust axis of the ROE. The definitions of thrust and thrust axis are given in Eq. 5.18.
- Two variables correspond to the harmonic moments B_0 and B_2 computed in the CMS with respect to the event thrust axis. The harmonic moments are defined in Eq. 5.21.
- One variable is the first normalised Fox-Wolfram moment R_1 computed in the CMS, defined in Eq. 5.24.
- Two variables are the linear KSFW moments $H^{so}_{m,2}$ and $H^{so}_{m,4}$ computed in the CMS, defined in Eq. 5.25.
- Two variables are the quadratic KSFW moments R^{oo}_0 and R^{oo}_2 computed in the CMS, defined in Eq. 5.26.

The distributions of these twelve variables in simulated signal and background events are shown in Fig. 13.4.

13.5.2 Event-based variables

Six additional event-based variables are used in the analysis. Three of them correspond to the total number of tracks and photon candidates in the event, and to their sum. The distributions of these variables are shown in Fig. 13.5. In general, the signal is characterised by lower track and photon multiplicities than the background, since the $B^0 \to K^{*0}\nu\bar{\nu}$ reconstructed final state consists only of a track pair.
Another event-based variable is the number of lepton candidates per event. It helps to identify and suppress semileptonic background decays. These decays represent a major source of background in the analysis, as discussed in Sec. 16.3. Candidate charged leptons are selected as candidate electrons and muons in the ROE satisfying the PID requirements $\mathrm{electronID} > 0.9$ and $\mathrm{muonID} > 0.9$ (see Sec. 2.8.4 for more details). The distributions in signal and background events shown in Fig. 13.5 present similar characteristics to the corresponding distributions computed in the analysis of the $B^+ \to K^+\nu\bar{\nu}$ decay, described in Sec. 5.7.1.
Two more event-based variables are the polar angle of the missing three-momentum $\vec{p}_{miss.}$ and the missing invariant mass squared per event. These variables are extremely powerful to separate signal from background. The characteristic features of the signal and background distributions are similar to those observed in the $B^+ \to K^+\nu\bar{\nu}$ analysis (see Fig. 5.8).

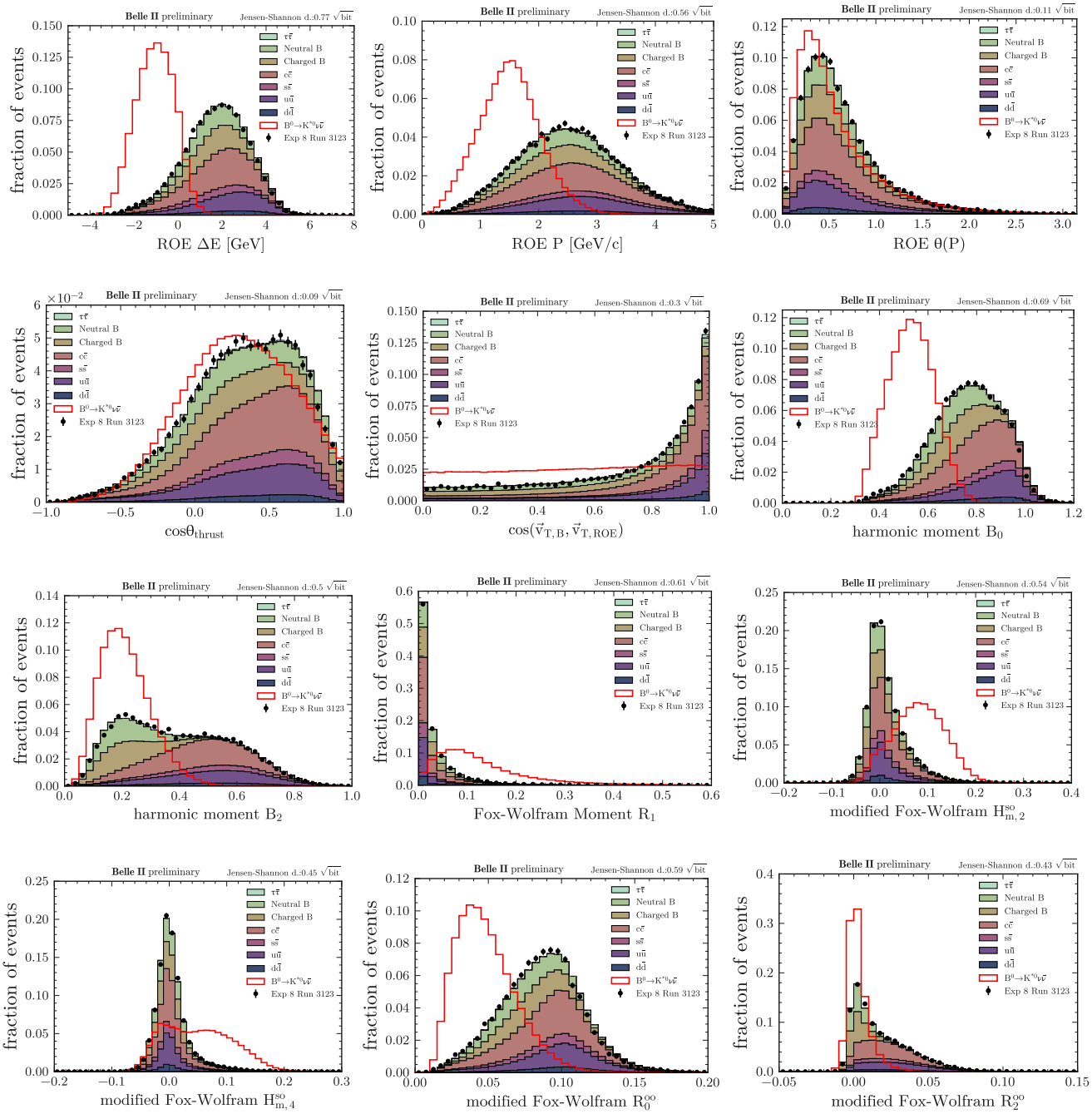


Figure 13.4: Distributions of the 12 variables used for the initial event filtering. From top to bottom, left to right: ROE ΔE; magnitude of the ROE total momentum $\vec{P}$; polar angle $\theta(\vec{P})$; cosine of the event thrust-axis polar angle; cosine of the angle between the signal-B and the ROE thrust axes; harmonic moments B_0 and B_2; first normalised Fox-Wolfram moment R_1; linear KSFW moments $H^{so}_{m,2}$ and $H^{so}_{m,4}$; quadratic KSFW moments R^{oo}_0 and R^{oo}_2. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

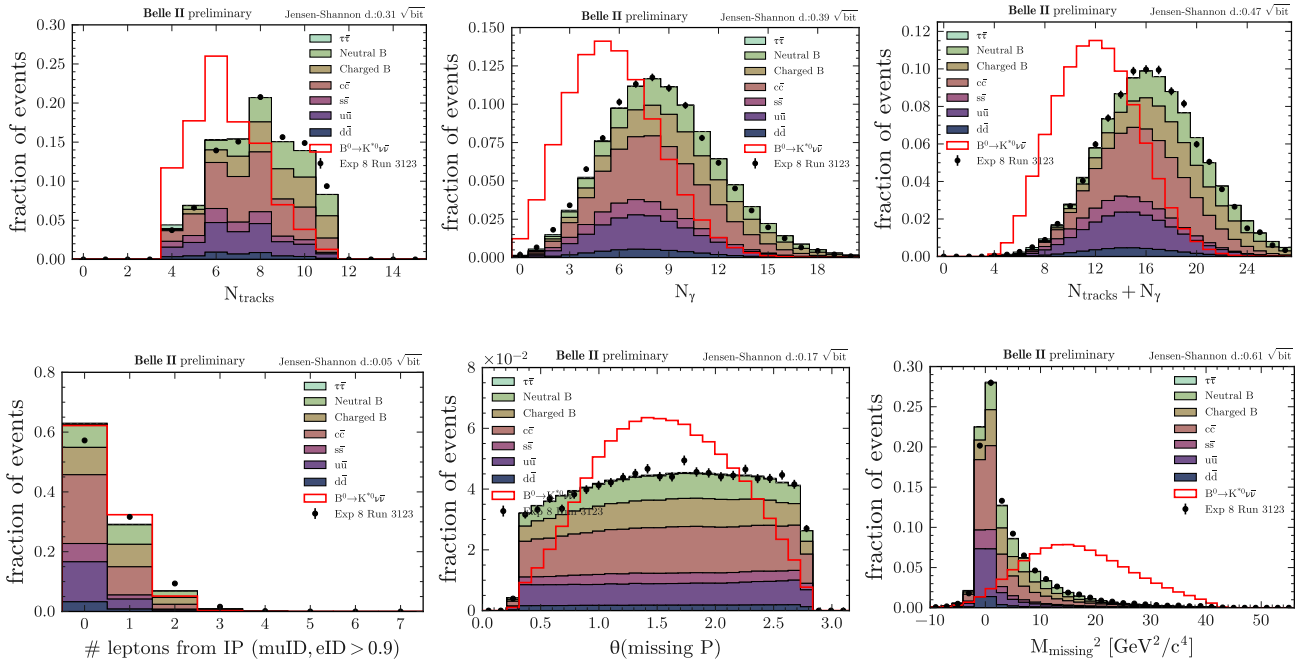


Figure 13.5: Distributions of event-based variables. From top to bottom, left to right: number of tracks, number of photon candidates, and number of tracks plus photon candidates per event; number of lepton candidates ($e^{\pm}$, $\mu^{\pm}$) per event; polar angle of the missing-momentum $\vec{p}_{miss.}$ and missing invariant mass squared per event. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

13.5.3 Event-shape variables

Six more event-shape variables are used to separate signal from background in the analysis:

- the event sphericity, defined in Eq. 5.19;
- the second and third normalised Fox-Wolfram moments, R_2 and R_3, defined in Eq. 5.24;
- the linear KSFW moments $H^{so}_{c,2}$, $H^{so}_{n,2}$, $H^{so}_{m,0}$, defined in Eq. 5.25.

All are computed in the CMS. The distributions of the variables in simulated signal and background events are shown in Fig. 13.6. Also the distributions of these variables present common features to all the $B \to K^{(*)}\nu\bar{\nu}$ modes, and a detailed discussion of the main characteristics of the distributions can be found in Sec. 5.7.2.

13.5.4 Variables related to the signal K^{*0} candidate

Six variables describing the kinematic properties of the signal K^{*0} candidate are computed in reconstructed signal and background MC events. The radial distance of the K^{*0} vertex from the average IP, $dr_{K^{*0}}$, and the reconstructed invariant mass of the K^{*0}, $M(K^{*0})$, are shown in Fig. 13.7. The signal and background distributions of $dr_{K^{*0}}$ have similar shapes. Both histograms peak at $dr_{K^{*0}} \sim 50\,\mu$m and are skewed to the right. However, the background distribution is characterised on average by larger values of $dr_{K^{*0}}$ compared

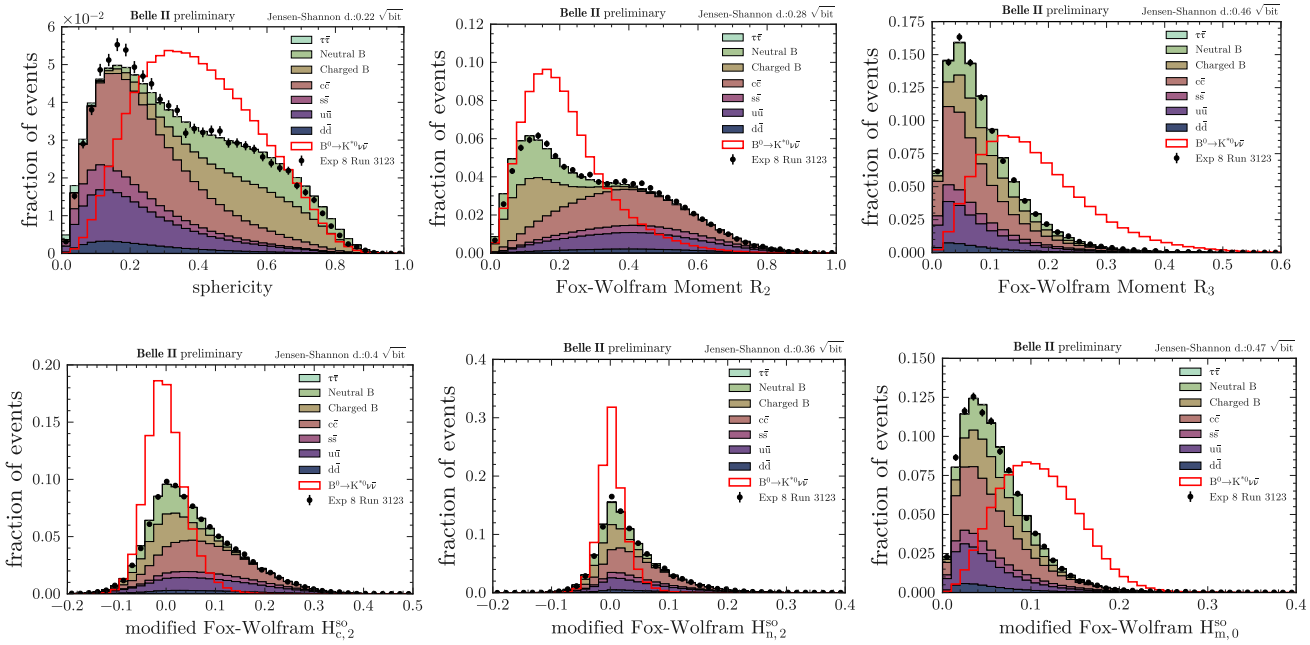


Figure 13.6: Distributions of event-shape variables. From top to bottom, left to right: event sphericity; second, R_2, and third, R_3, normalised Fox-Wolfram moments; linear KSFW moments $H^{so}_{c,2}$, $H^{so}_{n,2}$, $H^{so}_{m,0}$. All the variables are computed in the CMS. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

to the signal, and the relative divergence between the two histograms is sizeable. The invariant mass of the signal K^{*0} candidate is an optimal observable to distinguish signal from background events. The signal distribution presents a characteristic resonance peak around the average $K^{*0} \to K^+\pi^-$ mass value of $0.896\,\mathrm{GeV}/c^2$ [43]. In background events, the peak is smeared out by the large contribution of fake K^{*0} candidates originating from combinatorial background.

Two additional variables used in the analysis are designed to characterise the $K^{*0} \to K^+\pi^-$ decay in the Armenteros-Podolanski space [211]. A sketch of the K^{*0} decay in the laboratory reference frame is drawn in Fig. 13.8. Using the notation of the figure, $p^{\mathrm{T}}_{K^+}$, $p^{\mathrm{T}}_{\pi^-}$ are transverse momenta and $p^{\mathrm{L}}_{K^+}$, $p^{\mathrm{L}}_{\pi^-}$ longitudinal momenta with respect to the direction of flight of the parent particle. Due to the momentum conservation, $p^{\mathrm{T}}_{\mathrm{Arm.}} = p^{\mathrm{T}}_{K^+} = p^{\mathrm{T}}_{\pi^-}$, and the longitudinal momentum asymmetry is defined as

$$\alpha_{\mathrm{Arm.}} = \frac{p^{\mathrm{L}}_{K^+} - p^{\mathrm{L}}_{\pi^-}}{p^{\mathrm{L}}_{K^+} + p^{\mathrm{L}}_{\pi^-}} \left(\text{or } \frac{\mathrm{p}^{\mathrm{L}}_{\pi^+} - \mathrm{p}^{\mathrm{L}}_{\mathrm{K}^-}}{\mathrm{p}^{\mathrm{L}}_{\pi^+} + \mathrm{p}^{\mathrm{L}}_{\mathrm{K}^-}} \right), \tag{13.18}$$

depending on the charge sign of the daughter particles. The distributions of $p^{\mathrm{T}}_{\mathrm{Arm.}}$ and $\alpha_{\mathrm{Arm.}}$ in simulated signal and background events are illustrated in Fig. 13.9. Both the variables provide good separation power and capture the characteristic kinematic features of the $K^{*0} \to K^+\pi^-$ decay in $B^0 \to K^{*0}\nu\bar{\nu}$.

Another variable used in the analysis is the longitudinal distance $\delta z(K^+, \pi^-)$ between the POCAs of the tracks combined to form the signal K^{*0} candidate. The signal and background distributions of this variable are shown in Fig. 13.10. Both distributions are unimodal and peaking around zero. The signal peak is narrower, meaning that signal events are on average characterised by smaller distances between the two tracks. This feature produces a fairly large separation between the signal and background histograms.

Also the cosine

$$\cos\alpha_{K\pi} = \frac{(\vec{p}_{K^+} \cdot \vec{p}_{\pi^-})}{|\vec{p}_{K^+}||\vec{p}_{\pi^-}|}, \tag{13.19}$$

of the angle between the three-momentum vectors $\vec{p}_{K^+}$, $\vec{p}_{\pi^-}$ of the tracks forming the signal K^{*0} candidate is exploited. The distributions of this variable in simulated signal and background events are illustrated in Fig. 13.10. The signal and background histograms are both skewed to the left, but the tail of the signal distribution drops more rapidly to zero. The peaks in signal and background are both at cosine values close to one. It is also interesting to notice the features of the various background contributions. For the continuum events the peak at values close to one is more prominent, meaning that in such events the two tracks more often have momenta almost parallel and pointing in the same direction. This is a consequence of tracks momenta being very collimated in continuum events. Conversely, the histograms of generic $B^0\bar{B}^0$ and B^+B^- events are flatter. The combination of these features results in a sizeable divergence between the background and signal distributions.

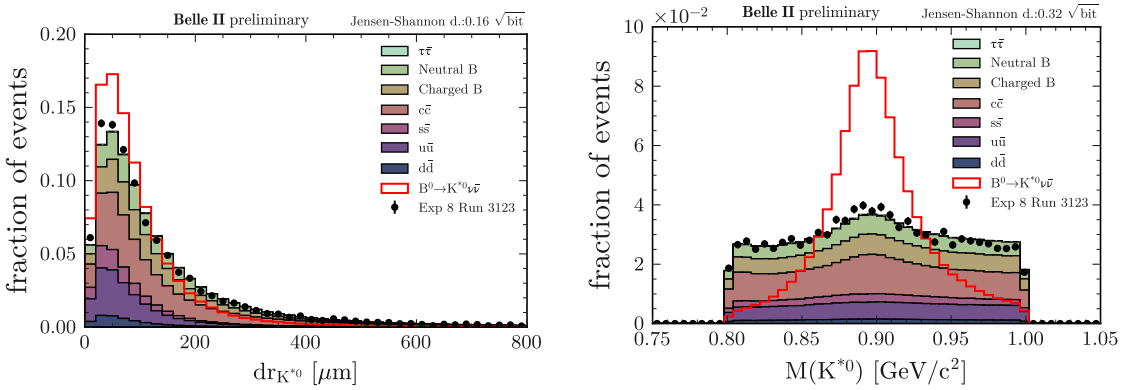


Figure 13.7: Distributions of $dr_{K^{*0}}$ (left) and $M(K^{*0})$ (right). The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

13.5.5 Variables related to tracks and energy deposits in the ROE

Some variables depending on the tracks momenta and on the energy deposits in the ROE are exploited in the initial event filtering and are described in Sec. 13.5.1. Two additional variables of this category are used in the second stage of the signal identification procedure. One of them is the invariant mass of the ROE. The distributions of this variable computed in reconstructed signal and background MC events are shown in Fig. 13.11. They are very similar to the corresponding distributions in Fig. 5.21 of the $B^+ \to K^+\nu\bar{\nu}$ analysis. The average ROE invariant mass in signal is smaller than the mass of the B^0 meson, $5.279\,\mathrm{GeV}/c^2$ [43]. This shift can have different origins. It is mostly caused by events in which some final state particles are out of the ROE acceptance (defined by the track and photon cleanups described in Sec. 13.1). In part, it is the result of events with neutrinos, or K^0_L mesons in the final state of the tag B decay. It can also be partly caused by wrong particle hypothesis assignments in the ROE. Conversely, in background events, the ROE consists of the decay products from the two B mesons in generic $B\bar{B}$ events, and from multiple decays in the continuum, thus the invariant mass is on average larger than in signal, and also larger than the B^0 mass. As a consequence, the background and signal peaks are well separated one with respect to the other, and the separation power provided by the ROE invariant mass is very large.
The second variable of this category is the variance of the ROE tracks transverse momenta, defined as

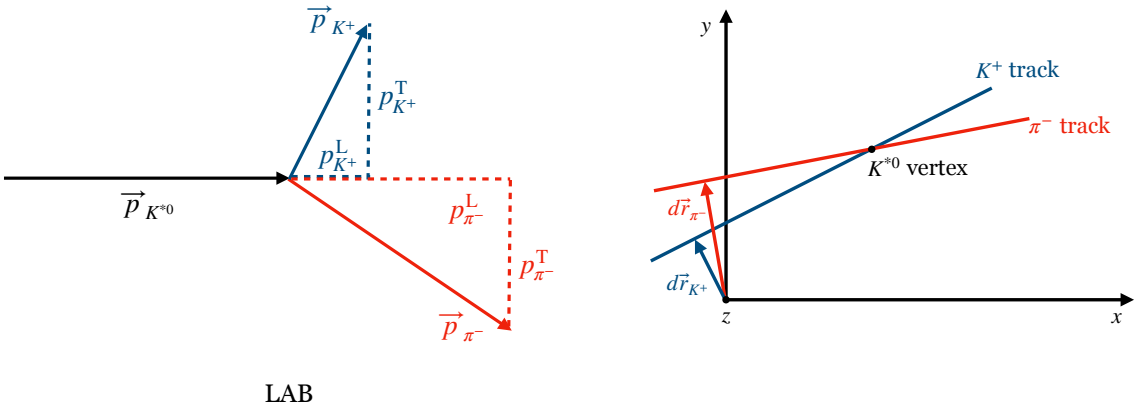


Figure 13.8: Left: sketch representing the kinematics of the $K^{*0} \to K^+\pi^-$ decay in the laboratory (LAB) reference frame. Right: sketch illustrating the positions $d\vec{r}_{K^+}$, $d\vec{r}_{\pi^-}$ of the POCAs of the K^+, π^- tracks and the vertex of the signal K^{*0} candidate; for simplicity, the positions are evaluated with respect to the origin of the coordinate system.

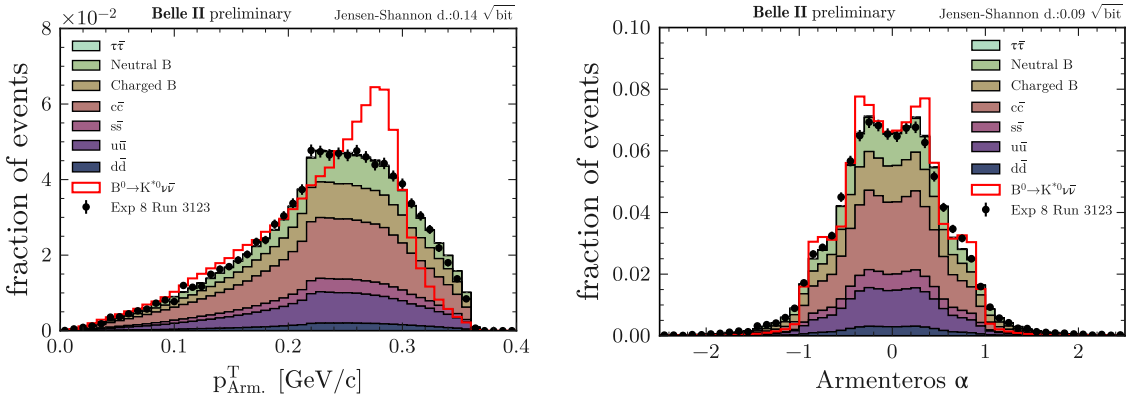


Figure 13.9: Distributions of the Armenteros-Podolanski variables $p^{\mathrm{T}}_{\mathrm{Arm.}}$ (left) and α (right) defined in Eq. 13.18. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

$\sum_{i=1}^{n}(p_{\mathrm{T},i} - \overline{p}_{\mathrm{T}})^2/n$, where i runs over the total number of tracks in the ROE, and $\overline{p}_{\mathrm{T}}$ is the mean value of the ROE transverse momentum. The signal and background distributions of the variable are illustrated in Fig. 13.11. Both are skewed to the right, but the signal histogram peaks more prominently at values close to zero, and its tail drops more rapidly than in background. Also this variable is fairly powerful in separating signal from background events.

13.5.6 Variables related to the ROE vertex

The ROE vertex is fitted using the algorithm described in Sec. 2.8.3.1, and the `KFit` algorithm [176] is adopted for the kinematic fit. Three variables related to the ROE vertex are used in the analysis. Their

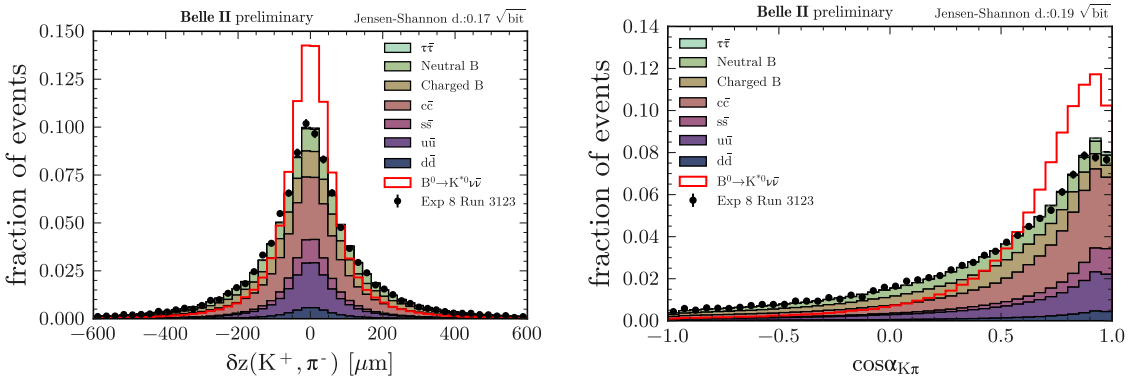


Figure 13.10: Distributions of the longitudinal distance $\delta z(K^+, \pi^-)$ between the POCAs of the K^+, π^- candidate tracks (left) and of the cosine $(\vec{p}_{K^+} \cdot \vec{p}_{\pi^-})/|\vec{p}_{K^+}||\vec{p}_{\pi^-}|$ (right). The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

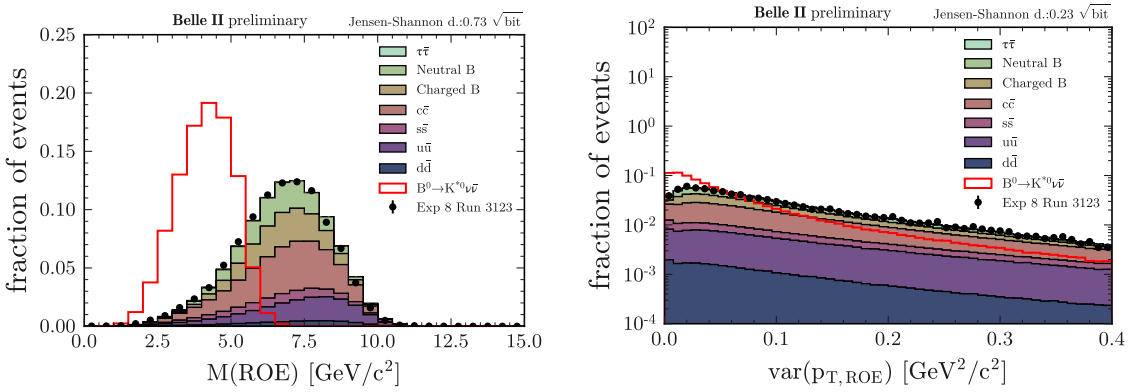


Figure 13.11: Distributions of the ROE invariant mass (left) and variance of the transverse momentum of the ROE tracks (right). The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

distributions in reconstructed signal and background MC events, and in data from a single collision run, are shown in Fig. 13.12. The first variable is the χ^2 of the vertex fit. The fraction of events with $\chi^2 < 5$ is larger in signal than in background. In true signal events, the ROE vertex is fitted using tracks all originating from the decay vertex of the tag B meson, thus vertex fits are expected to have small χ^2. The fit quality is worse in background events, where the ROE is made of tracks produced by two, or more, decays. Therefore, the χ^2 of the ROE vertex fit provides a good separation of signal from background.
The other two variables are the radial and longitudinal distances, $dr_{K^{*0}}(\text{Tag Vertex})$ and $dz_{K^{*0}}(\text{Tag Vertex})$, between the position of the fitted ROE vertex and the vertex of the signal K^{*0} candidate. For $dr_{K^{*0}}(\text{Tag Vertex})$, the signal and background histograms have similar asymmetric shapes. The signal peak is at $\sim 80\,\mu\text{m}$, while the background peak is at $\sim 50\,\mu\text{m}$. Both histograms are skewed to the right, but the tail in signal drops

to zero slightly faster than in background. For $dz_{K^{*0}}(\text{Tag Vertex})$, the signal and background histograms are both symmetric with peaks around zero. The background distribution has a FWHM of $\sim 200\,\mu\text{m}$, while the signal distribution is much broader, with a FWHM of $\sim 300\,\mu\text{m}$. Overall, the background distributions are characterised by radial and longitudinal distances on average smaller than in signal. Moreover, the background peaks of the two variables are mainly produced by events from the light hadronic continuum ($e^+e^- \to q\bar{q}$, where $q = u, d, s$) in which the tracks are produced promptly in the proximity of the interaction region.

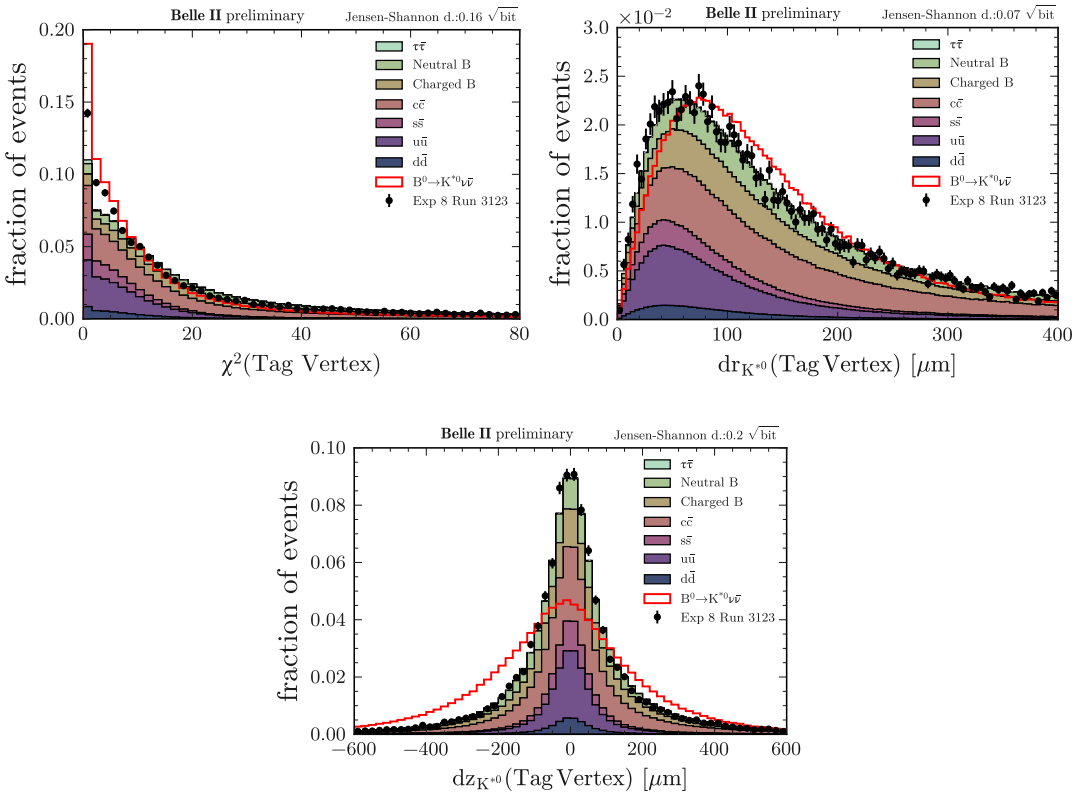


Figure 13.12: Distributions of variables related to the ROE vertex: χ^2 of the ROE vertex fit (top left); radial (top right) and longitudinal (bottom) distances between the position of the fitted ROE vertex and the signal K^{*0} vertex. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

13.5.7 Variables related to D^0/D^+ suppression

Fourteen variables are used for the suppression of D^0 and D^+ background decays. Variables describing the properties of the best D^0 candidate, formed by combining the track of the signal K^+ candidate and an ROE track of opposite charge, are shown in Fig. 13.13. One of the variables is the invariant mass of the D^0 candidate. The background distribution presents three visible resonances. It is possible to identify: a peak for true $K^*(892)^0$ mesons at approximately $0.89\,\text{GeV}/c^2$; a sharp peak for the $\phi(1020)^0$ meson at invariant mass of approximately $1\,\text{GeV}/c^2$; another sharp peak for true D^0 mesons at approximately $1.86\,\text{GeV}/c^2$. The signal distribution is smoother, since the majority of the vertex candidates in signal are

only combinations of oppositely charged tracks originating from decays of distinct mesons. Another variable is the longitudinal distance, dz, from the fitted position of the D^0 vertex to the average IP. The signal and background distributions are both symmetric and unimodal, but the signal distribution is shifted towards larger dz values. The background events peak at $dz \sim 50\,\mu$m, while the signal peak is at $dz \sim 150\,\mu$m. Two additional variables are the longitudinal and radial distances, $\delta z(K^+, r^-_{\in \mathrm{ROE}})$ and $\delta r(K^+, r^-_{\in \mathrm{ROE}})$, between the POCAs of the two tracks forming the D^0 candidate. The signal and background δz distributions are both symmetric and peaking around zero, but the signal peak is spread out and the distribution is broader. Also in δr, the signal and background distributions have similar shapes, unimodal and skewed to the right, but the signal is characterised on average by slightly larger δr values.

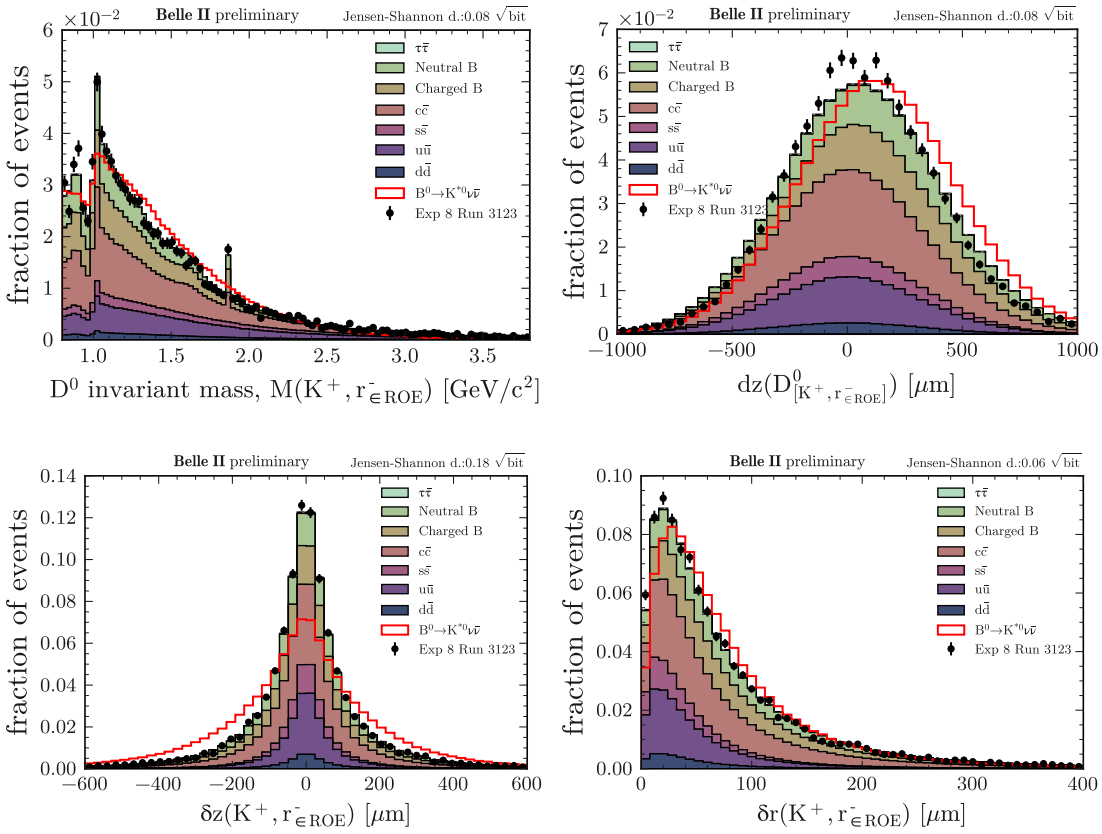


Figure 13.13: Distributions of variables related to the best D^0 candidate constructed with the signal-candidate K^+ track and an ROE track of opposite charge for which the most-likely particle hypothesis is assumed. From top to bottom, left to right: invariant mass of the D^0 candidate; longitudinal distance from the D^0 vertex to the average IP position; longitudinal and radial distances between the POCAs of the tracks forming the D^0 candidate. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

As described in Sec. 13.3.0.1, D^0 candidates are also reconstructed by fitting to the same vertex the signal K^{*0} candidate and two pion tracks in the ROE. A set of characteristic variables are computed for the best D^0 candidate in signal and background events and their distributions are illustrated in Fig. 13.14. The invariant mass of the best D^0 candidate is one of the variables. The distribution in signal events is unimodal, peaking at approximately $2\,\mathrm{GeV}/c^2$, and has a moderate right tail. The background distribution has a longer tail, and

it presents a sharp resonance, corresponding to true D^0 vertices, at the average D^0 mass of $1.86\,\mathrm{GeV}/c^2$. The longitudinal distance between the D^0 vertex and the average IP is also exploited. The distributions in signal and background events are similar to the corresponding distributions observed for D^0 candidates constructed with only two tracks. The radial distance from the IP is also used and it provides a good separation power. The signal and background distributions have similar shapes, but candidate D^0 vertices in signal events have on average larger radial distances from the IP. Two additional variables are the means of the longitudinal and radial distances between the POCAs of the pion tracks and the position of the signal K^{*0} vertex, defined as

$$\overline{dz}_\pi(K^{*0}) = \frac{dz_{\pi^+}(K^{*0}) + dz_{\pi^-}(K^{*0})}{2} \tag{13.20}$$

and

$$\overline{dr}_\pi(K^{*0}) = \frac{dr_{\pi^+}(K^{*0}) + dr_{\pi^-}(K^{*0})}{2}\,. \tag{13.21}$$

The signal and background $\overline{dz}_\pi(K^{*0})$ distributions are unimodal and symmetric, centred around zero. The background histogram has a narrow peak, while the signal peak is smeared out and much broader. The distributions of $\overline{dr}_\pi(K^{*0})$ in signal and background events have similar asymmetric shapes, both unimodal and skewed to the right, but the signal peak is broader and signal events are on average characterised by larger values of the variable.

The signal and background distributions of the variables related to the best D^+ candidate in an event, reconstructed by pairing the signal K^{*0} candidate and an ROE track, are shown in Fig. 13.15. One of these variables is the χ^2-probability of the best vertex fit, which has a very large separation power. The fraction of best fits in signal with χ^2-probability $< 20\%$ is much larger than in background, while in the interval of χ^2-probability $> 80\%$, corresponding to the high quality fits, the signal/background ratio is almost 0.5. Two other variables are the longitudinal and radial distances from the position of the fitted D^+ vertex to the average IP. The distributions of these variables in signal and background events present characteristic features similar to those observed in the corresponding distributions for D^0 candidates. Moreover, the longitudinal distance between the POCA of the ROE track and the vertex of the signal K^{*0} candidate is exploited. The signal and background histograms both peak around zero, but the signal peak is broader. One more variable is the cosine of the angle between the momentum of the signal K^{*0} candidate and the momentum of the ROE track used to form the best D^+ candidate,

$$\cos(\vec{p}_{K^{*0}}, \vec{p}_{r^-\in\mathrm{ROE}}) = \frac{\vec{p}_{K^{*0}} \cdot \vec{p}_{r^-\in\mathrm{ROE}}}{|\vec{p}_{K^{*0}}||\vec{p}_{r^-\in\mathrm{ROE}}|}. \tag{13.22}$$

In signal events the distribution of this variable is monotonic increasing from -1 to +1. The shape is different in background, where it presents two peaks, one at approximately -0.9 and the other close to +1.0, and is depleted in between.

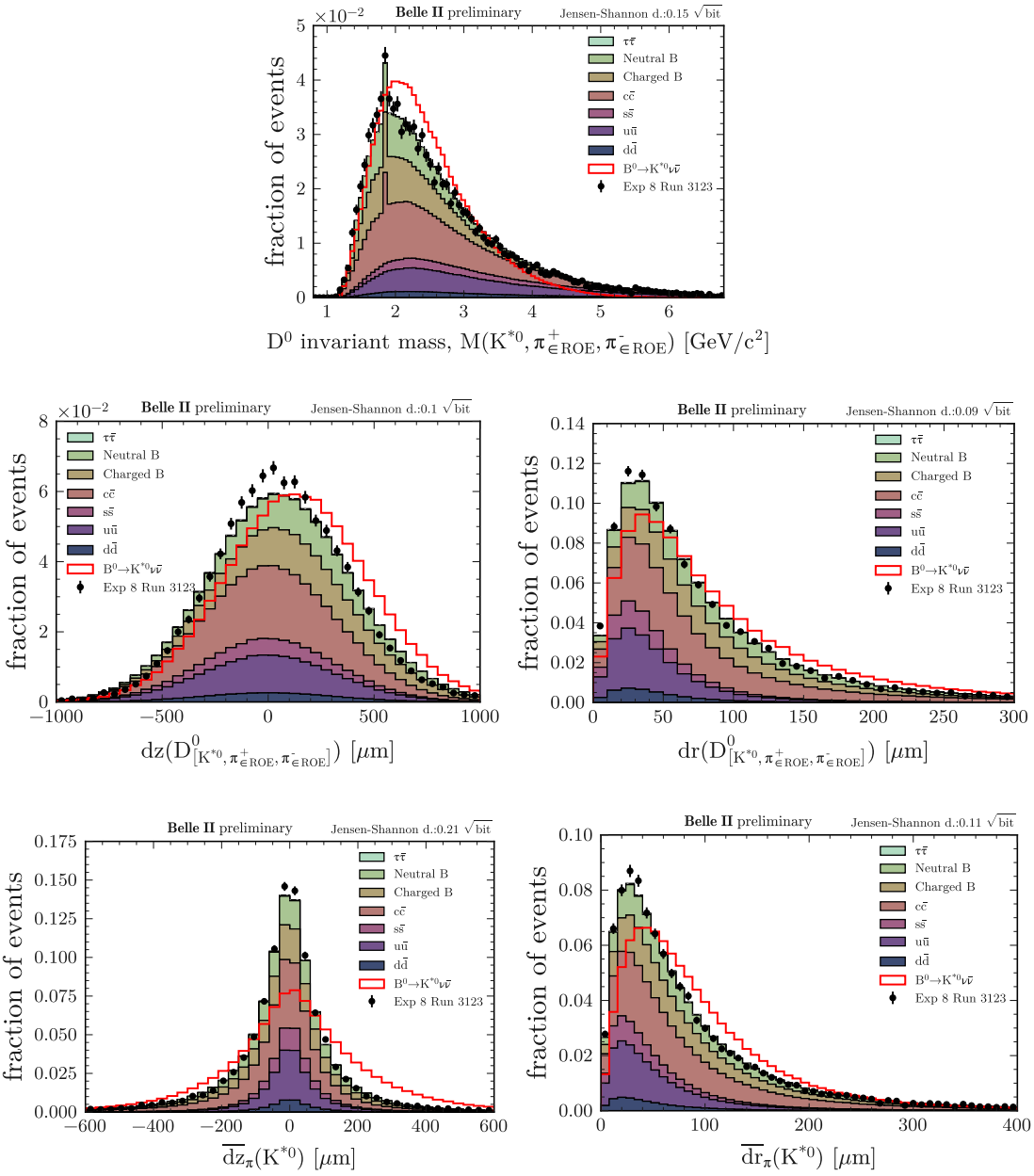


Figure 13.14: Distributions of variables related to the best D^0 candidate constructed with the signal K^{*0} candidate and two pion tracks in the ROE. From top to bottom, left to right: invariant mass of the D^0 candidate; longitudinal and radial distances from the D^0 vertex to the average IP; means of the longitudinal and radial distances between the POCAs of the pion tracks and the K^{*0} vertex. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

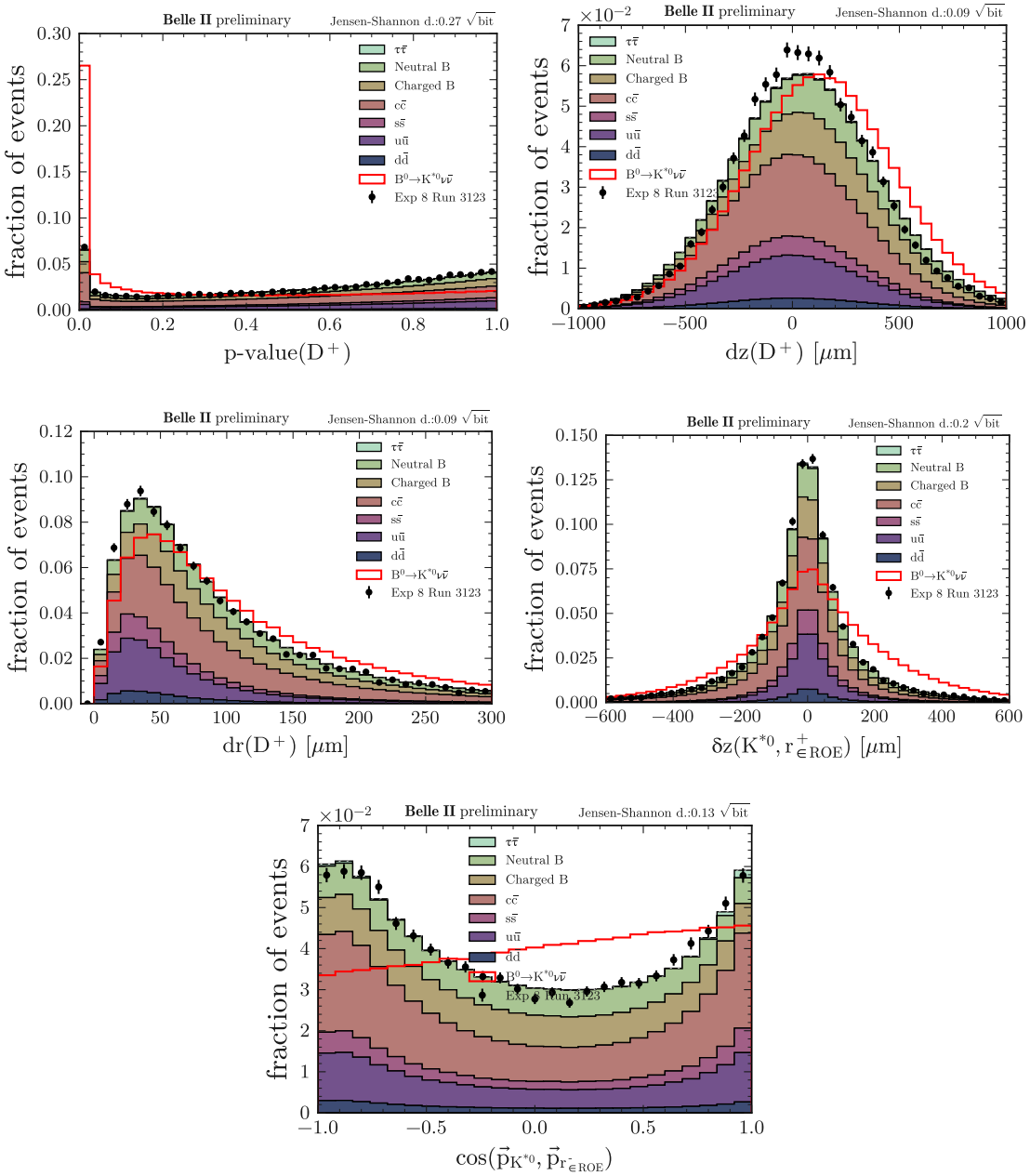


Figure 13.15: Distributions of variables related to the best D^+ candidate constructed with the signal K^{*0} candidate and an ROE track for which the most-likely particle hypothesis is assumed. From top to bottom, left to right: χ^2-probability of the vertex fit; longitudinal and radial distances from the D^+ vertex to the average IP position; longitudinal distance between the POCA of the ROE track and the K^{*0} vertex; cosine of the angle between the K^{*0} momentum and the momentum of the ROE track. The red step histogram originates from true $B^0 \to K^{*0}\nu\bar{\nu}$ signal events; the stacked filled histograms correspond to simulated events of the seven main background categories; the black dots reproduce the distribution from data collected in a single collision run. The distributions are normalised to unit areas. At the top right corner, the estimated J-S distance between the signal and the global background histograms is provided.

Chapter 14

Multivariate classification

In the analysis, signal identification relies on multivariate binary classification. This chapter presents the classification procedure and discusses the performance of the classifiers. Section 14.1 gives an overview of the workflow and defines the training samples and features. The classification models are briefly described in Section 14.2. Section 14.3 focuses on the importance of the training features. Finally, Section 14.4 illustrates the classification performance.

14.1 Overview

The signal identification relies on supervised multivariate classification. Two binary classifiers are used in series for the task. The first classifier, BDT_1, is based on the `FastBDT` algorithm [186] (see Sec. 3.6.2 for more details). It is trained on 1.6×10^6 simulated events of each of the background categories and on the same number of signal events. Background and signal events are weighted using the procedure presented in Sec. 6.2. The training features correspond to 12 variables, described in Sec. 13.5.1, characterising the event shape and the main kinematic properties of the ROE. The first classifier is used as a *filter* to discard a large fraction of background events. The second classifier, BDT_2, improves the background suppression in the region where the signal is most likely to be found. BDT_2 uses a GPU-accelerated `XGBoost` algorithm [188]. This classifier is trained with 37 variables, described in the sections 13.5.2–13.5.7. The training is performed using the events with BDT_1 outputs greater than 0.90, which are reconstructed in 1.6×10^6 simulated signal events and in a simulated background sample of $100\,\mathrm{fb}^{-1}$.

14.2 Classification models

The classification models of BDT_1 and BDT_2 are characterised by the hyperparameter values listed in Table 14.1, which are determined in a dedicated optimisation study. The two models differ from each other only by the tree depth, which corresponds to the maximum number of consecutive cuts in a decision tree. Better performance is observed for BDT_2 by increasing the tree depth from 2 to 3.

Hyperparameter	Value
Number of trees	2000
Tree depth	2 for BDT_1, 3 for BDT_2
Shrinkage	0.2
Sampling rate	0.5
Number of equal-frequency bins	256

Table 14.1: Optimised hyperparameters of the classification models.

14.3 Feature importances

The feature importances quantify the improvements in the prediction accuracy brought by each of the training features. Both `FastBDT` and `XGBoost` evaluate the feature importances by summing up the separation gain of each feature over all trees and nodes of the model. More details can be found in Refs. [186, 188]. The 12 training features of BDT_1 are ranked by importance in Table 14.2. It is interesting to notice that, in this small set, the combined importance of the top three variables, ΔE of the ROE (score $= 0.350$), first normalised Fox-Wolfram moment R_1 (score $= 0.210$), linear KSFW moment $H^{so}_{m,2}$ (score $= 0.143$), sums up to 70%. This is not surprising, since the ROE ΔE and $H^{so}_{m,2}$ are respectively sensitive to the missing energy and momentum in the event, and R_1 powerfully distinguishes the specific symmetry of signal events from that of generic background events.

Training variable	Importance score
ΔE of the ROE	0.350
First normalised Fox-Wolfram moment R_1 computed in the CMS	0.210
Linear KSFW moment $H^{so}_{m,2}$ computed in the CMS	0.143
Harmonic moment B_0 computed in the CMS	0.060
Magnitude of the total ROE momentum	0.052
Quadratic KSFW moment R^{oo}_0 computed in the CMS	0.046
Linear KSFW moment $H^{so}_{m,4}$ computed in the CMS	0.039
Harmonic moment B_2 computed in the CMS	0.034
Quadratic KSFW moment R^{oo}_2 computed in the CMS	0.027
Cosine of the angle between the signal-B thrust axis and the ROE thrust axis	0.018
Cosine of the polar-angle thrust-axis	0.010
Polar angle of the total ROE momentum	0.009

Table 14.2: List of the training variables of BDT_1 ranked by importance score.

The ten most important features for BDT_2 are ranked by importance score in Table 14.3. The variable at the top of the table is the invariant mass of the signal K^{*0} candidate, which has an importance score equal to 0.236. This variable provides the key information to suppress the contribution of purely combinatorial background. The other nine features consist of:

- 2 variables related to D^0/D^+ suppression: the χ^2-probability of the vertex fit to the best D^+ candidate (score $= 0.083$) and the longitudinal distance between the POCAs of the track pair forming the best D^0 candidate (score $= 0.037$);
- 1 variable related to the signal K^{*0} candidate: dr of the signal K^{*0} (score $= 0.067$);
- 3 event-based variables: the number of photon candidates in the event (score $= 0.049$), the polar angle of the missing momentum vector (score $= 0.043$), and the number of charged lepton candidates in the event (score $= 0.034$);
- 2 event-shape variables: the linear Fox-Wolfram moments $H^{so}_{n,2}$ (score $= 0.046$) and $H^{so}_{c,2}$ (score $= 0.035$);
- 1 variable related to the ROE vertex: the longitudinal distance between the position of the fitted ROE vertex and the signal K^{*0} vertex (score $= 0.033$).

These numbers further quantify the importance of the K^{*0} invariant mass in the second classification stage.

Training variable	Importance score
Invariant mass of the signal K^{*0} candidate	0.236
χ^2-probability of the vertex fit to the best D^+ candidate	0.083
dr of the signal K^{*0} candidate	0.067
Number of photon candidates in the event	0.049
Linear Fox-Wolfram moment $H^{so}_{n,2}$ computed in the CMS	0.046
Polar angle of the missing momentum vector $\vec{p}_{miss.}$	0.043
Longitudinal distance between the POCAs of the track pair of the best D^0	0.037
Linear Fox-Wolfram moment $H^{so}_{c,2}$ computed in the CMS	0.035
Number of charged lepton candidates in the event	0.034
Longitudinal distance between the fitted ROE vertex and the signal K^{*0} vertex	0.033

Table 14.3: List of the ten most important training variables for the BDT_2 classifier ranked by importance score.

14.4 Performance of the classifiers

The classification performance of BDT_1 and BDT_2 is evaluated in the training sample and in an independent test sample equivalent in size and composition. The results are summarised in Fig. 6.1. The classification models are characterised by a very good accuracy, and no evident overfitting is visible. As observed also in the $B^+ \to K^+\nu\bar{\nu}$ analysis, the signal and background distributions of the classifier outputs are more overlapped for BDT_2 than BDT_1, meaning that BDT_1 has a larger separation power than BDT_2.
The signal identification performance of BDT_1 and BDT_2 is evaluated in terms of the statistical significance of signal using the figure of merit $S/\sqrt{S+B}$ (see Appendix B). S and B are the expected number of signal and background events, respectively. The significance is computed by reconstructing a simulated background sample of $800\,\mathrm{fb}^{-1}$ and 4×10^6 simulated signal events. The results are scaled to the integrated luminosity of $189\,\mathrm{fb}^{-1}$, corresponding to the size of the on-resonance collision-data sample for which this analysis is optimised. The figure of merit is evaluated as a function of the lower bounds on the BDT_1 and BDT_2 outputs. To simplify the interpretation of the results, the selection on the classifier output is translated into the equivalent signal efficiency. Figure 14.2 illustrates the performance of BDT_1 and BDT_2 evaluated at $\mathrm{BDT}_1 > 0.90$ as a function of the signal efficiency. The significance for BDT_1 is almost flat for selections corresponding to signal efficiencies larger than 2%. In the approximate plateau the significance is $\sim 15\%$. The plot shows that the application of BDT_2 provides a maximum increase of $\sim 30\%$ in significance, which goes up to $\sim 45\%$, boosting the signal identification performance.

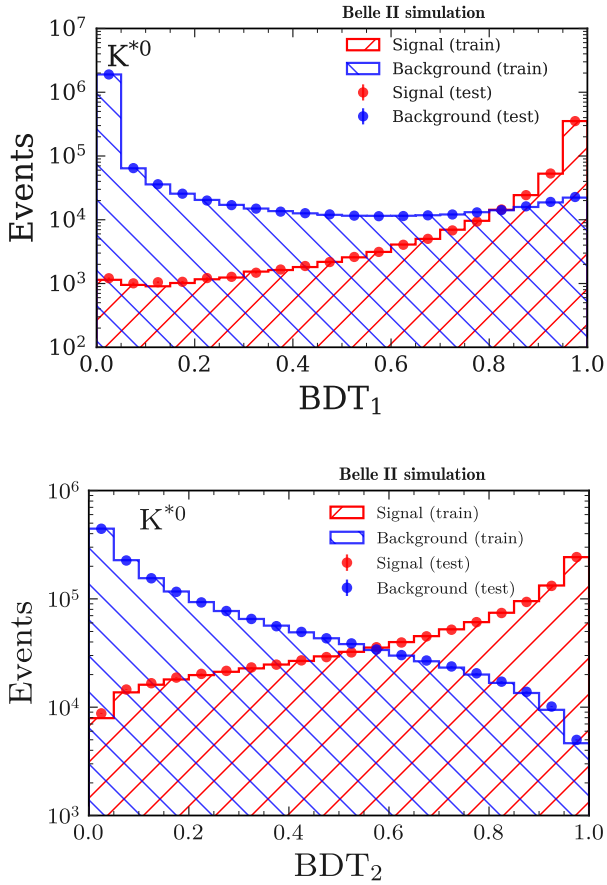


Figure 14.1: Distributions of BDT_1 (top) and BDT_2 (bottom) outputs, for signal in the training (red step histogram) and test (red dots) samples, and for background in the training (blue step histogram) and test (blue dots) samples.

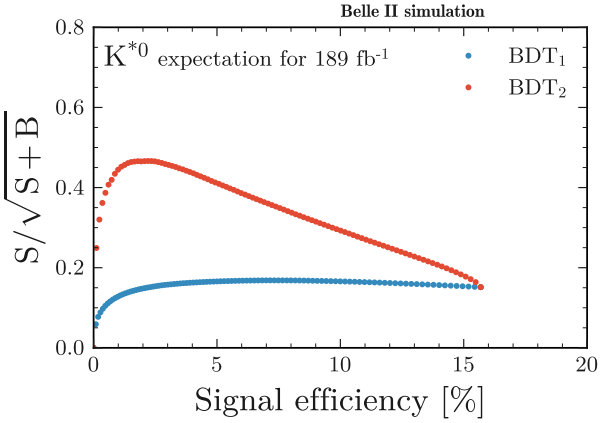


Figure 14.2: Statistical significance of signal at $\mathrm{BDT}_1 > 0.90$ estimated with the figure of merit $S/\sqrt{S+B}$, where S and B are the expected number of signal and background events. It is computed using a simulated background sample equivalent to $800\,\mathrm{fb}^{-1}$ and 4×10^6 simulated signal events. The results are scaled to the integrated luminosity of $189\,\mathrm{fb}^{-1}$. The significance is evaluated as a function of the lower bounds on the BDT_1 (blue dots) and BDT_2 (red dots) outputs translated into the equivalent signal efficiencies.

Chapter 15

Validation studies

This chapter presents the studies performed to validate the analysis with collision data in which no signal contribution is present. Section 15.1 describes the validation with the $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ decay used as an independent control channel. In this section, a new method generalising the technique described in Section 8.1 is presented. Section 15.2 discusses the validation with the off-resonance data sample.

15.1 Validation with $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$

The classification performance of the inclusive tagging is validated with on-resonance data using an independent control channel. By analogy with the $B^+ \to K^+ \nu\bar{\nu}$ analysis, the chosen decay mode for the validation study is $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$. The decay has a sizeable branching fraction, $\mathrm{BR}(B^0 \to K^{*0} J/\psi) = (1.27 \pm 0.05) \times 10^{-3}$ and $\mathrm{BR}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$ [43]. Moreover, it is characterised by a clean experimental signature, which guarantees high reconstruction efficiency.
The $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ events selected in data and simulation are modified. The modification of the kaon kinematics and the removal of the muon tracks and clusters from $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ decays are essential for the validation. In this analysis, the procedure implemented for this task represents an upgrade of the strategy presented in Sec. 8.1. The new technique is called *signal embedding*. Its workflow is summarised by the chart in Fig. 15.1. First, $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ events are identified in the MC simulation and in data. All the final-state tracks and clusters associated with the $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ decay are removed from the events and only the ROEs are kept. In parallel, true signal $B^0 \to K^{*0} \nu\bar{\nu}$ events are identified in the signal simulation and all the tracks and clusters not related to the $B^0 \to K^{*0} \nu\bar{\nu}$ decay are discarded. The signal decay in each event is then combined with one of the stored ROEs from the selected $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ events, and *embedded* events are created.

15.1.1 Identification of $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ events

To identify $B^0 \to K^{*0} J/\psi_{\to\mu^+\mu^-}$ events in simulation and data, $K^{*0} \to K^+\pi^-$ and $J/\psi \to \mu^+\mu^-$ candidate decays are first reconstructed.
Candidate muon tracks are required to satisfy the track cleanup defined in Sec. 13.1 and the PID selection $\mathrm{muonID} > 0.5$, where muonID is the PID variable defined in Sec. 2.8.4. Moreover, the muon tracks must have at least one PXD hit attached. The J/ψ candidates are reconstructed by pairing the selected muons and applying the condition $|M^{\mathrm{PDG}}_{\mathrm{J}/\psi} - M_{\mu^+\mu^-}| < 50\,\mathrm{MeV}/c^2$, where $M^{\mathrm{PDG}}_{\mathrm{J}/\psi} = 3.096\,\mathrm{GeV}/c^2$ is the average J/ψ mass provided by the PDG [43] and $M_{\mu^+\mu^-}$ is the invariant mass of the muon pair. This selection reduces the contamination from the radiative tail of the J/ψ resonance.
The reconstruction and selection of the signal K^{*0} candidates rely on the same procedure adopted in the $B^0 \to K^{*0}\nu\bar{\nu}$ reconstruction algorithm, described in detail in Sec. 13.2. The only extra requirement is the PID selection $\mathrm{muonID} < 0.5$, orthogonal to the one used to identify muon candidates, which is applied to the tracks used to form a K^{*0} candidate.

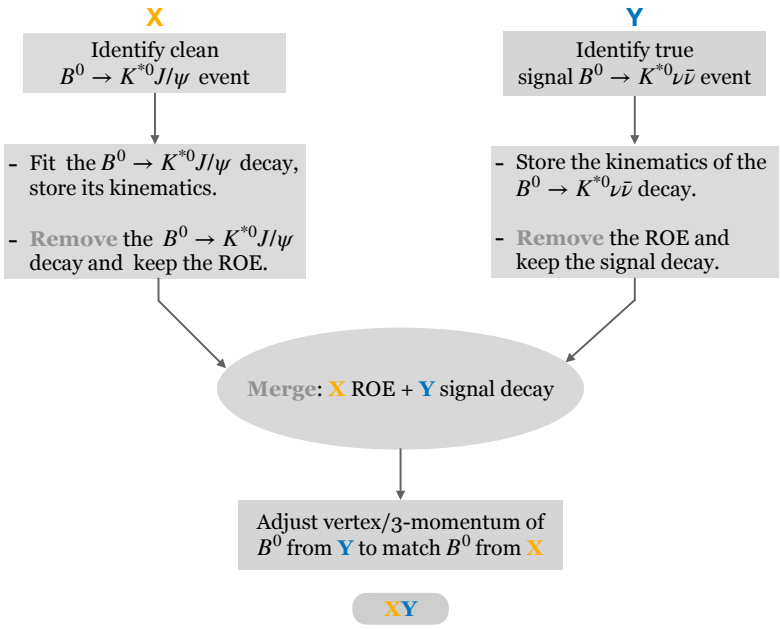


Figure 15.1: Flowchart summarising the main steps of the *signal embedding*. The same procedure is used with $B^0 \to K^{*0} J/\psi$ events identified in data and MC. Credit: S. Glazov.

The K^{*0} and and J/ψ candidates are then combined to form a B^0 meson, which is required to have $M_{bc} > 5.25\,\mathrm{GeV}/c^2$ and $|\Delta E| < 100\,\mathrm{MeV}$. In total, 1855 $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ events are identified in the on-resonance data sample. The selection efficiency is evaluated in a MC sample simulating $B^0 \to K^{*0}(\to K^+\pi^-) J/\psi(\to l^+l^-)$ events in which both the e^+e^- and $\mu^+\mu^-$ leptonic final states of the J/ψ decay are generated. It is found to be approximately 11%. The results of the $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ selection are illustrated for data and simulation in Fig. 15.2. The overall data-MC agreement is good. Minor discrepancies in ΔE and $M_{\mu^+\mu^-}$ can be reduced with the application of muonID and kaonID correction weights, which are not used to produce the plots in Fig. 15.2. It is also estimated that the selection keeps the background contamination at $\mathcal{O}(5\%)$.

15.1.2 Results

The identified $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ events, with and without signal *embedded*, are reconstructed using the algorithm implemented for the analysis, which is described in detail in Chap. 13. The fraction of $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ events surviving the selection of the inclusive tagging algorithm is $(80.9 \pm 1.0)\%$ in data and $(80.4 \pm 0.3)\%$ in simulation.
The validation of the classifiers is illustrated in Fig. 15.3. The main plot shows the distributions of the BDT_1 outputs for events of the control channel, without ($B^0 \to K^{*0} J/\psi$) and with ($B^0 \to K^{*0} \cancel{J/\psi}$) signal embedded, reconstructed in data and MC simulation. The data and MC distributions of the BDT_1 outputs peak at 0 for the unmodified events and at 1 for the events of the control channel where the signal is embedded. Good data-MC agreement is observed in both cases. A p-value equal to 89% is estimated by means of a two-sample K-S test [203] performed on the distributions from embedded events in simulation and data. The distribution of the BDT_1 outputs for simulated signal $B^0 \to K^{*0} \nu\bar{\nu}$ events is also illustrated. It is in good agreement with the corresponding distributions from $B^0 \to K^{*0} \cancel{J/\psi}$ events reconstructed in data and MC simulation.

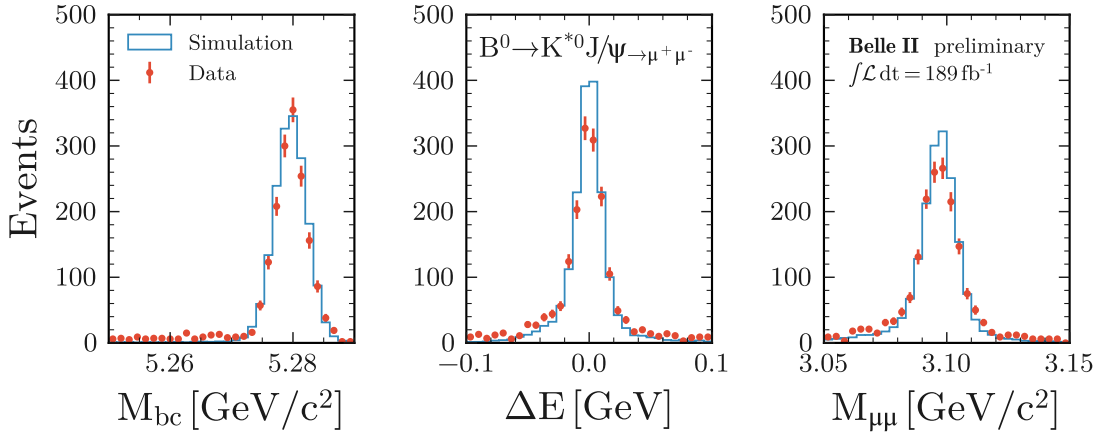


Figure 15.2: Selected $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ candidates for the validation study. The distributions of the beam-constraint mass M_{bc} (left plot), ΔE (central plot) and dimuon invariant mass $M_{\mu^+\mu^-}$ (right plot) in data (red points) and simulation (blue step histograms) are compared.

The inner plot in Fig. 15.3 shows the distributions of the BDT_2 outputs computed for events with $\mathrm{BDT}_1 > 0.9$. All the unmodified events from data and simulation are suppressed by this selection. Conversely, approximately 78% of the $B^0 \to K^{*0} J\!\!\!/\psi$ events reconstructed in data and simulation satisfy the $\mathrm{BDT}_1 > 0.9$ selection. The data-MC agreement of the corresponding BDT_2 distributions is excellent. A K-S test is performed and a p-value equal to 51% is estimated. The distributions agree very well also with the BDT_2 outputs for simulated $B^0 \to K^{*0}\nu\bar{\nu}$ events.
In the analysis, the region of highest signal sensitivity is characterised by signal efficiency $< 3\%$ (see Fig. 14.2), which is equivalent to the selection $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.94$. This selection corresponds approximately to the last bin of the histograms in the inset of Fig. 15.3. Here, it is interesting to compare the classification efficiencies in data and simulation for events of the control channel with embedded signal. The results are summarised in Table 15.1. The data/MC efficiency ratio is found to be 1.09 ± 0.06. To account for this slight discrepancy, a systematic uncertainty is introduced in the statistical model constructed for the fit.

Sample	Classification efficiency
Data	0.209 ± 0.011
MC	0.192 ± 0.003
Efficiency ratio	
1.09 ± 0.06	

Table 15.1: Classification efficiencies, in data and MC, for events of the control channel with *embedded* signal ($B^0 \to K^{*0} J\!\!\!/\psi$) at $\mathrm{BDT}_1 > 0.9$ and $\mathrm{BDT}_2 > 0.94$.

15.2 Validation in off-resonance data

The off-resonance data sample is used to validate the modelling of the continuum simulation. The studies are carried out using reconstructed events in simulation and data with $\mathrm{BDT}_1 > 0.9$. MC events from the continuum background are reconstructed in a sample equivalent to $100\,\mathrm{fb}^{-1}$ integrated luminosity, and the yields in simulation are then scaled to $18\,\mathrm{fb}^{-1}$, which is the integrated luminosity of the off-resonance data sample. The PID correction weights for the selection $\mathrm{kaonID} > 0.9$ (see Sec. 13.2) are applied to the simulated events. All the variables used in the analysis are validated. To summarise the results, the distributions of the ROE ΔE and of the signal K^{*0} candidate invariant mass, respectively the most important

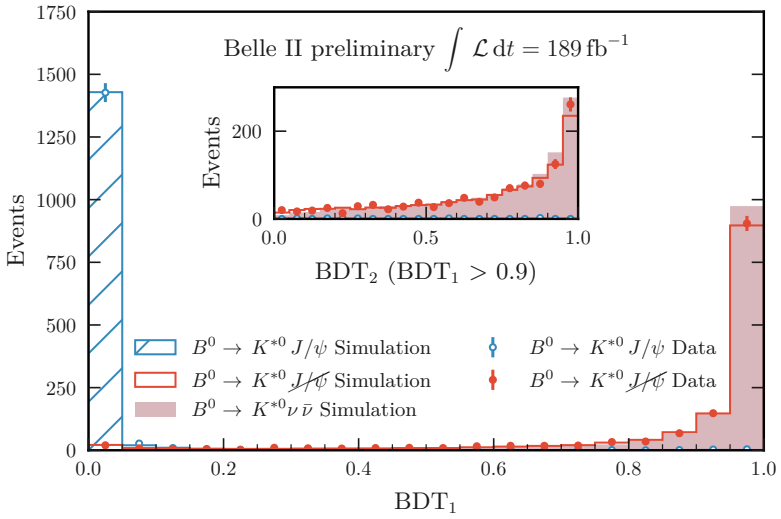


Figure 15.3: Distributions of the classifier outputs BDT_1 (main figure) and BDT_2 for $\mathrm{BDT}_1 > 0.9$ (inset). The distributions are shown before ($B^0 \to K^{*0} J/\psi$) and after ($B^0 \to K^{*0} \not{J/\psi}$) the *signal embedding*. As a reference, the classifier outputs directly obtained for simulated $B^+ \to K^+ \nu\bar{\nu}$ signal events are overlaid. The simulation histograms are scaled to the total number of $B^0 \to K^{*0} J/\psi_{\to \mu^+\mu^-}$ events selected in data.

variables for BDT_1 and BDT_2, are shown in Fig. 15.4. Here yields in simulation are scaled to data by a factor 1.24. This value corresponds to the observed data/MC yield ratio. A normalisation discrepancy of approximately the same size was observed in the $B^+ \to K^+ \nu\bar{\nu}$ analysis, and it is documented in Sec. 8.2. Other minor data-MC discrepancies are present. Among the others, such discrepancies can be noticed by comparing the ROE ΔE distributions in data and simulation. The right plot in Fig. 15.4 shows that they are slightly shifted one with respect to the other. To correct the mismodelling, the same technique described in Sec. 8.2.1, is adopted. A binary classifier, BDT_c, is trained taking the off-resonance data as the positive class. In this analysis, BDT_c is based on the `XGBoost` algorithm [188], and the classification model is presented in Table 15.2. The values of the hyperparameters listed in the table are optimised against overfitting. The classifier is trained on the events with $\mathrm{BDT}_1 > 0.9$ reconstructed in a MC sample of $100\,\mathrm{fb}^{-1}$ and in the off-resonance data sample. The BDT_c training variables consist of all the BDT_2 training variables, plus the BDT_2 output and q^2_rec.

Hyperparameter	Value
Number of trees	2000
Tree depth	1
Shrinkage	0.01
Sampling rate	0.01
Number of equal-frequency bins	256

Table 15.2: Optimised hyperparameters of the binary classification model of BDT_c.

The BDT_c performance in continuum simulation and off-resonance data is shown in Fig. 15.5. BDT_c is applied to the events from continuum MC and off-resonance data used for the training and to the events from an independent test sample in order to check for possible overfitting of the model. No overfitting is observed. The BDT_c distributions in data and simulation do not completely overlap due to the mismodelling in the continuum MC. An ideal modelling of the continuum background would produce data and MC distributions completely overlapping and centred around $\mathrm{BDT}_\mathrm{c} = 0.5$.
The BDT_c outputs are computed for the events of the continuum simulation and the corresponding weights

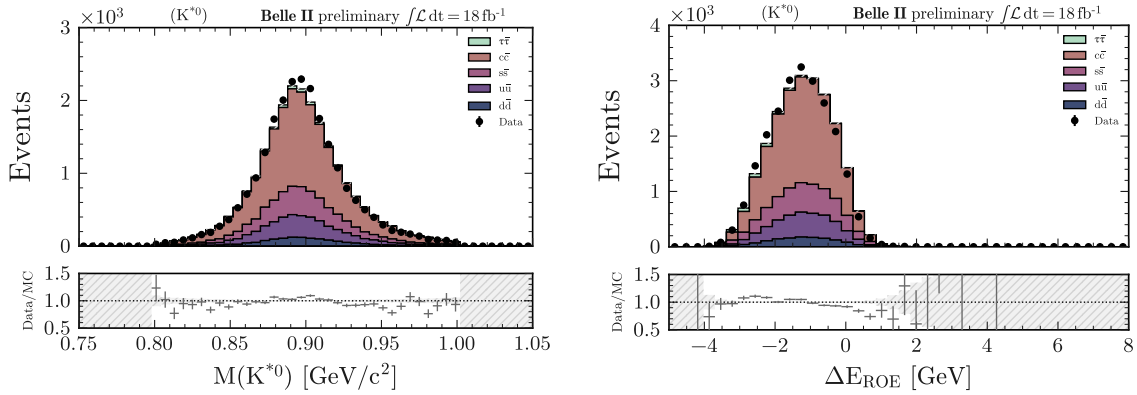


Figure 15.4: Distributions of the signal K^{*0} candidate invariant mass (left) and ROE ΔE (right) in off-resonance data (black points) and simulated continuum background (stacked filled histograms). Yields in simulation are scaled to data by the ratio data/MC $= 1.24$. Data/MC ratios of events in the bins of the variables are also illustrated.

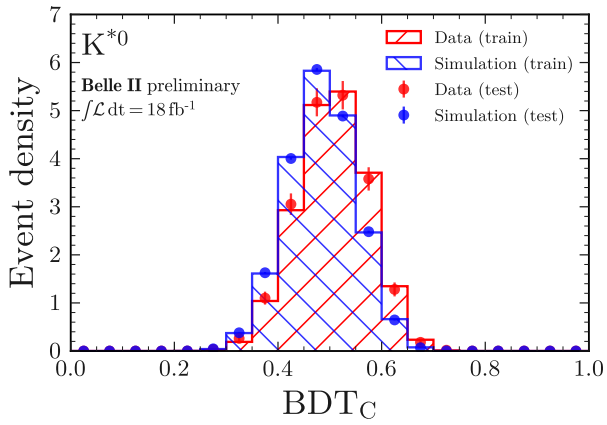


Figure 15.5: BDT_c outputs for off-resonance data (red) and continuum simulation (blue) events of the training (step histograms) and test (data points) samples. Each histogram is normalised to unit area.

$BDT_c/(1 - BDT_c)$ are applied. The results of the reweighting are shown, for the ROE ΔE and the signal K^{*0} candidate invariant mass, in the upper panel of Fig. 15.6. The shape discrepancies are corrected and the data and MC distributions agree well within the statistical uncertainties. Nevertheless, the 24% normalisation discrepancy between off-resonance data and continuum simulation is still present and a systematic uncertainty is introduced to account for it. More details are given in Sec. 16.5.1.
The classification performance in the off-resonance data is also validated. The bottom plot in Fig. 15.6 shows the distributions of the efficiency quantile in events from data and simulation with BDT_c weights applied. The efficiency quantile is one of the two variables used to build the statistical model for the fit. It is obtained by mapping the BDT_2 output into the corresponding signal efficiency ϵ, and it is defined as $1 - \epsilon$. The data-MC agreement is good also in the region of large efficiency quantile, where the sensitivity to the signal is highest (see Fig. 14.2).

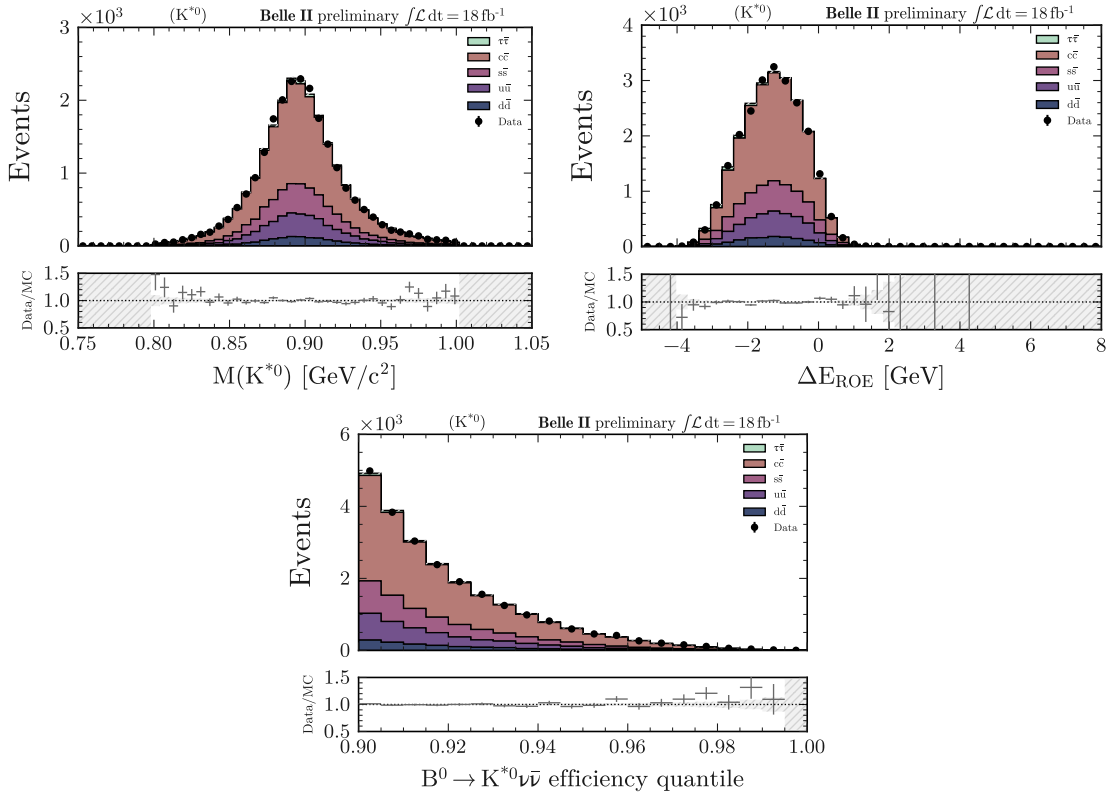


Figure 15.6: Distributions of the signal K^{*0} candidate invariant mass (left), ROE ΔE (right), and efficiency quantile (bottom) in off-resonance data (black points) and simulated continuum background (stacked filled histograms) after the application of the event weight $\mathrm{BDT_c}/(1-\mathrm{BDT_c})$. Yields in simulation are scaled to data by the ratio data/MC$=1.24$. Data/MC ratios of events in the bins of the variables are also illustrated.

Chapter 16

Statistical model

This chapter presents the statistical model implemented in the analysis. Section 16.1 introduces the observables and the MC samples used to build the model. In section 16.2 the channels of the model are defined. In section 16.3 the expectations are presented, focusing on the expected yields and on the background composition. Section 16.4 summarises the model parametrisation. Section 16.5 lists the various sources of systematic uncertainty and describes each of them in a dedicated subsection. The final section 16.8 discusses the impact of the systematic uncertainties on the measurement.

16.1 Overview

Data from the on-resonance and off-resonance samples are fitted simultaneously. The statistical model is constructed using the `pyhf` library [3], and it is defined in the two-dimensional space of the reconstructed di-neutrino invariant-mass squared, q^2_{rec}, and of the signal efficiency quantile, $\bar{\epsilon} = 1 - \epsilon$ (where ϵ represents the signal efficiency, as defined in Sec. 15.2). Signal and background templates for the expected event yields are derived from the MC simulation. To model the on-resonance data, eight MC batches, each equivalent to an integrated luminosity of $100\,\text{fb}^{-1}$, are reconstructed. A sample of 4×10^6 simulated signal events is used for the signal. An independent MC sample of $100\,\text{fb}^{-1}$ is reconstructed to derive templates for the off-resonance data, included in the fit to constrain the continuum yields. The PID correction weights for the selection $\text{kaonID} > 0.9$ (see Sec. 13.2) are applied to all the events from simulation. The BDT_c weights defined in Sec. 15.2 are assigned to the simulated events from the continuum backgrounds. As everywhere else in the analysis, signal events are reweighted to the SM prediction. In the channels of the model used to fit on-resonance data, the yields of reconstructed events from simulation are scaled to the integrated luminosity of $189\,\text{fb}^{-1}$. Otherwise, where only off-resonance data is fitted, the yields of simulated events from the continuum backgrounds are scaled to the off-resonance integrated luminosity, which is equal to $18\,\text{fb}^{-1}$. The expected signal yields are determined by scaling the reconstructed $B^+ \to K^+\nu\bar{\nu}$ events to the expected number of signal events in $189\,\text{fb}^{-1}$ of on-resonance data.

16.2 Channels

The statistical model consists of disjoint binned distributions from four channels. The channels are defined in the region of the $q^2_{\text{rec}} \times \bar{\epsilon}$ space where the signal significance is maximal. For consistency with the description given in Sec. 9.1, the same nomenclature is adopted: the channel where the sensitivity to signal is highest is called signal region (SR) and the remaining others are called control regions (CR1, CR2, CR3). Their complete characterisation is provided in Table 16.1. Figure 16.1 illustrates the bin boundaries in the $q^2_{\text{rec}} \times \bar{\epsilon}$ space and the adopted numbering scheme. The chosen boundaries of efficiency quantile, $[0.9700, 0.9775, 0.9850, 0.9925, 1.000]$, keep the signal-efficiency integral over q^2_{rec} equal to 0.75% in all the four $\bar{\epsilon}$ bins. The q^2_{rec} bins, $[-1, 4, 8, 25]\ \text{GeV}^2/c^4$, are inspired by the SM analysis in Ref. [85]. The lower

bound of q^2_{rec} is set to -1 to account for the $\sim 1\,\text{GeV}^2/c^4$ resolution of q^2_{rec} at generated q^2 equal to zero, as illustrated in Fig. 13.1.

Channel	Bin numbers	Bin boundaries	Samples
Signal region (SR)	4 to 12	$q^2_{\text{rec}} \in [-1.0, 4.0, 8.0, 25.0]\,\text{GeV}^2/c^4$, $\bar{\epsilon} \in [0.9775, 0.9850, 0.9925, 1.0000]$	signal, neutral B, charged B, $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 1 (CR1)	1 to 3	$q^2_{\text{rec}} \in [-1.0, 4.0, 8.0, 25.0]\,\text{GeV}^2/c^4$, $\bar{\epsilon} \in [0.9700, 0.9775]$	signal, neutral B, charged B, $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 2 (CR2)	4 to 12	$q^2_{\text{rec}} \in [-1.0, 4.0, 8.0, 25.0]\,\text{GeV}^2/c^4$, $\bar{\epsilon} \in [0.9775, 0.9850, 0.9925, 1.0000]$	$u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$
Control region 3 (CR3)	1 to 3	$q^2_{\text{rec}} \in [-1.0, 4.0, 8.0, 25.0]\,\text{GeV}^2/c^4$, $\bar{\epsilon} \in [0.9700, 0.9775]$	$u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$, $\tau^+\tau^-$

Table 16.1: Channels and samples of the `HistFactory` [177] model constructed in the two-dimensional space of the reconstructed di-neutrino invariant-mass squared, q^2_{rec}, and of the signal efficiency quantile, $\bar{\epsilon} = 1 - \epsilon$.

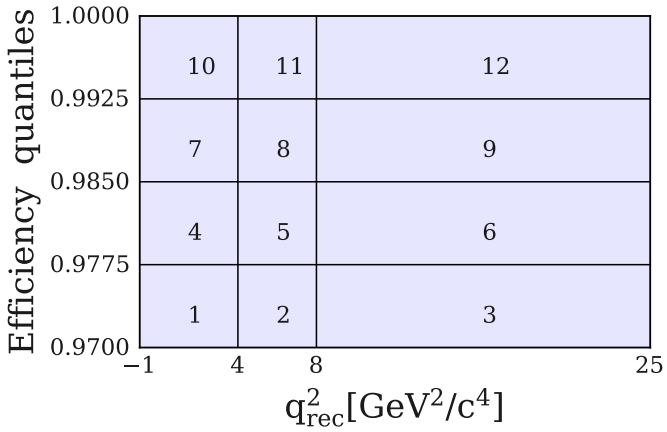


Figure 16.1: Illustration of the bin boundaries and numbering scheme.

16.3 Expectations

The expected signal and background yields in the SR, CR1, CR2 and CR3 are provided in Table 16.2. The distributions of the yields in the twelve bins of the SR and CR1 (CR2 and CR3) are shown in Fig. 16.2 (Fig. 16.6). The sum of the expected signal yields in each bin of efficiency quantile is constant, as required by construction. However, the signal distribution is not exactly flat due to correlation between q^2_{rec} and the signal efficiency. The background distribution presents a characteristic pattern, with features repeated in each bin of efficiency quantile. In all the four bins, the background yields are monotonic increasing as a function of q^2_{rec}. The dominant backgrounds correspond to events with generic neutral and charged B decays. The continuum background contribution decreases as a function of the $\bar{\epsilon}$ bin and is almost negligible in the last three bins (10, 11, 12), in which the sensitivity to the signal is highest.
The *crossfeed* among the $B \to K^{(*)}\nu\bar{\nu}$ channels in the twelve $q^2_{\text{rec}} \times \bar{\epsilon}$ bins is evaluated. The four channels have similar characteristics and experimental features. Therefore, even if optimised for $B^0 \to K^{*0}\nu\bar{\nu}$ decays, the inclusive tagging may retain contributions from the other channels. The distributions of the expected

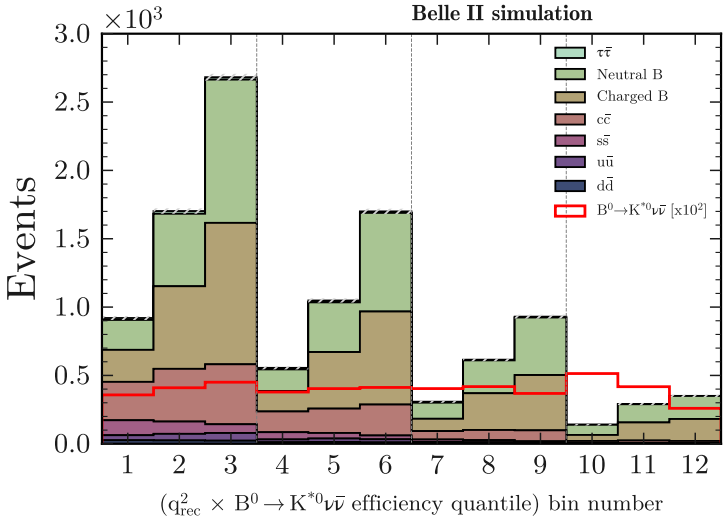


Figure 16.2: Expected sample yields in CR1 (bins 1 to 3) and SR (bins 4 to 12) for an integrated luminosity of $189\,\mathrm{fb}^{-1}$. The expected signal yield is scaled by a factor equal to 100.

Sample	Channels	Expected yields
signal	SR, CR1	36, 12
charged B	SR, CR1	2345, 1873
neutral B	SR, CR1	2397, 1797
$c\bar{c}$	SR, CR1, CR2, CR3	824, 1105, 78, 102
$d\bar{d}$	SR, CR1, CR2, CR3	46, 73, 4, 6
$s\bar{s}$	SR, CR1, CR2, CR3	170, 264, 17, 25
$u\bar{u}$	SR, CR1, CR2, CR3	98, 141, 11, 19
$\tau^+\tau^-$	SR, CR1, CR2, CR3	58, 52, 5, 5

Table 16.2: Expected signal and background yields in the channels of the statistical model.

yields for each channel are illustrated in Fig. 16.3. It is found that the combined contribution from the $B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, $B^+ \to K^{*+}\nu\bar{\nu}$ modes corresponds to approximately 10% of the expected yield for $B^0 \to K^{*0}\nu\bar{\nu}$. This contribution is covered by the normalisation systematic discussed in Sec. 16.5.1.
The generic $B\bar{B}$ background is also explored more in depth, using the procedure presented in Chap. 7. The ten most frequent signal and tag decays of the B^+B^- background in the SR and CR1 are illustrated in Fig. 16.4. In all the listed decays, either a D^0 or a $D^*(2007)^0$ meson is present. In the signal side, the semileptonic decays $B^+ \to D^0 l^+ \nu_l(\gamma)$, $B^+ \to D^*(2007)^0 l^+ \nu_l(\gamma)$ with $l = e, \mu$, and $B^+ \to D^0\tau^+\nu_\tau$, $B^+ \to D^*(2007)^0\tau^+\nu_\tau$ are the dominant processes. Their combined contribution is equal to 66% of the total. The other processes reported in the chart are hadronic decays, and they are present in 4.3% of the candidate signal decays. In the tag side of the selected events, the same semileptonic decays are the major processes, occurring with a frequency of approximately 34%. The remaining decays listed in the tag-side chart are hadronic decays, with a combined contribution of approximately 5%.
The ten major decays in the signal and tag sides of the neutral B background are listed in the charts of Fig. 16.5. The composition is similar to that of the charged B background. In the signal side of the selected events, the semileptonic processes $B^0 \to D^+ l^- \nu_l(\gamma)$, $B^0 \to D^*(2010)^+ l^- \nu_l(\gamma)$, and $B^0 \to D^+\tau^-\nu_\tau$, $B^0 \to D^*(2010)^+\tau^-\nu_\tau$ are dominant, representing approximately 45% of the decays. The others are hadronic decays, whose contribution amounts to $\sim 5\%$. Among them, it is interesting to notice the presence of $B^0 \to J/\psi K^{*0}$. One of the possible modes is $B^0 \to J/\psi(\to n\bar{n})K^{*0}$, which is characterised by a final state with experimental signature similar to $B^0 \to K^{*0}\nu\bar{\nu}$ (the branching fraction of $J/\psi \to n\bar{n}$

is $(2.09 \pm 0.16) \times 10^{-3}$ [43]). In the tag side of the selected $B^0\bar{B}^0$ background events, semileptonic $B^0 \to D^+(D^*(2010)^+)l^-\nu_l(\gamma)$, and $B^0 \to D^+(D^*(2010)^+)\tau^-\nu_\tau$ decays sum up to 38% of the total and the major hadronic decays to $\sim 3\%$.

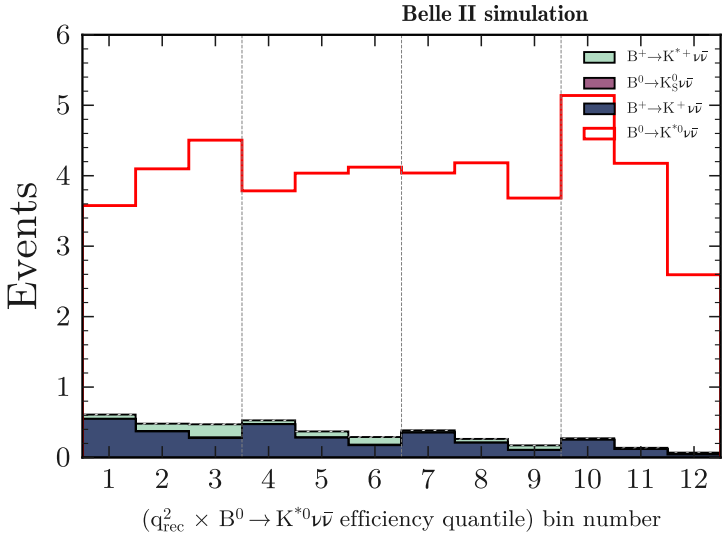


Figure 16.3: Crossfeed among the $B \to K^{(*)}\nu\bar{\nu}$ decays in 12 bins of the fit regions. Expected sample yields are determined for the integrated luminosity of $189\,\mathrm{fb}^{-1}$.

16.4 Model parameters

The statistical model depends on one parameter of interest and on a set of nuisance parameters accounting for various sources of systematic uncertainty. The former is the signal strength μ, representing the unconstrained normalisation parameter coupled to the SM branching fraction of the $B^0 \to K^{*0}\nu\bar{\nu}$ decay. The signal strength is defined such that $\mu = 1$ corresponds to $\mathrm{BR}(B^0 \to K^{*0}\nu\bar{\nu})_{\mathrm{SM}} = 7.8 \times 10^{-6}$ [93]. The nuisance parameters of the model are:

- 7 parameters for the normalisation uncertainties of the seven background categories;
- 5 parameters for the correlated shape variation in the B samples due to the systematic uncertainty on the branching fractions of the leading B background decays;
- 9 parameters for the correlated shape variation in the signal sample due to the uncertainty on the SM form factors;
- 7 parameters for the correlated shape variation due to the uncertainty on the PID correction weights;
- 1 parameter for the correlated shape variation due to the uncertainty on the tracking efficiency;
- 1 parameter for the correlated shape variation due to the uncertainty on the energy calibration of photon clusters;
- 1 parameter for the correlated shape variation due to the uncertainty on the energy calibration of ECL clusters not matched to photons;
- 1 parameter for the correlated shape variation in the signal sample due to the BDT classification efficiency;

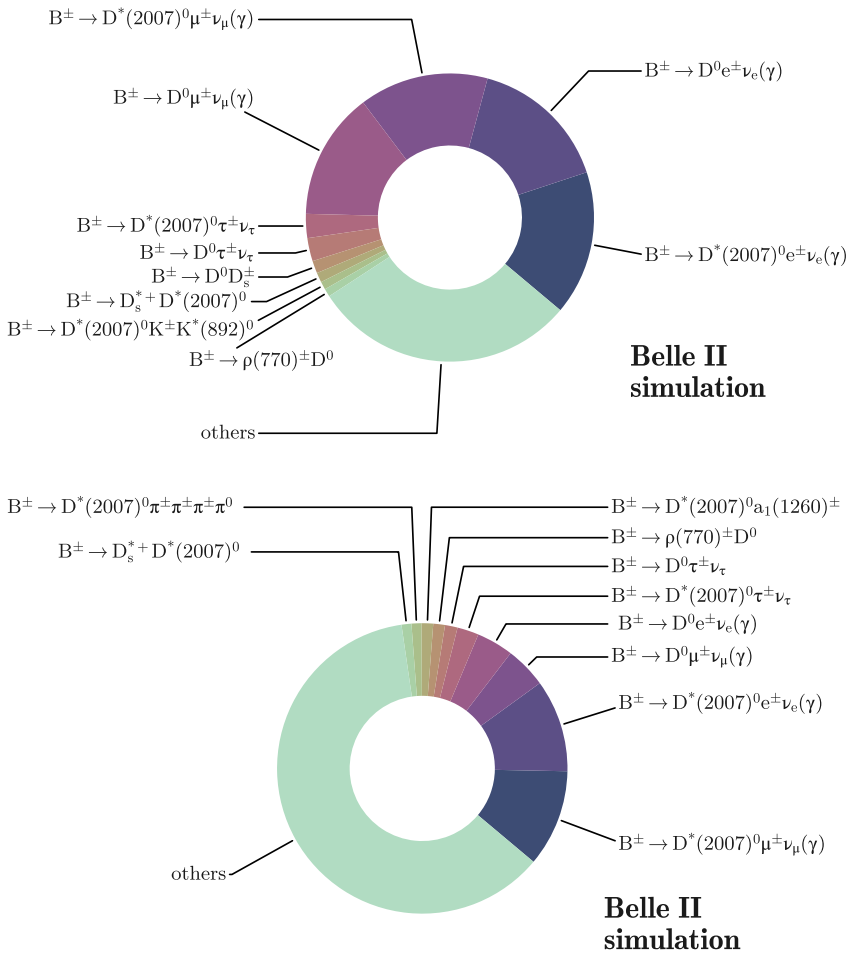


Figure 16.4: Composition of the reconstructed charged B background in the 12 bins of the SR and CR1. The top chart shows the decays occurring in the selected signal side. The bottom chart shows the decays in the tag side. Only the 10 most frequent decay modes are explicitly reported in the diagrams.

- 1 parameter for the correlated shape variation due to a partial mismodelling of the hadronic resonances in the MC simulation;
- 1 parameter per bin (in all the bins of all the channels, for all the samples) for the sum in quadrature of MC statistical uncertainty and the uncorrelated part of the shape systematics.

In total, there are 190 nuisance parameters. All the nuisance parameters have a Gaussian constraint. Each source of systematic uncertainty is described in detail in the following section.

16.5 Systematic uncertainties

The estimation of the systematic uncertainties is presented in this section. Two distinct categories of systematic biases can be identified: those originating from the detector response and from the performance of the reconstruction algorithms, and those related to the modelling of the physics processes in the MC simulation.

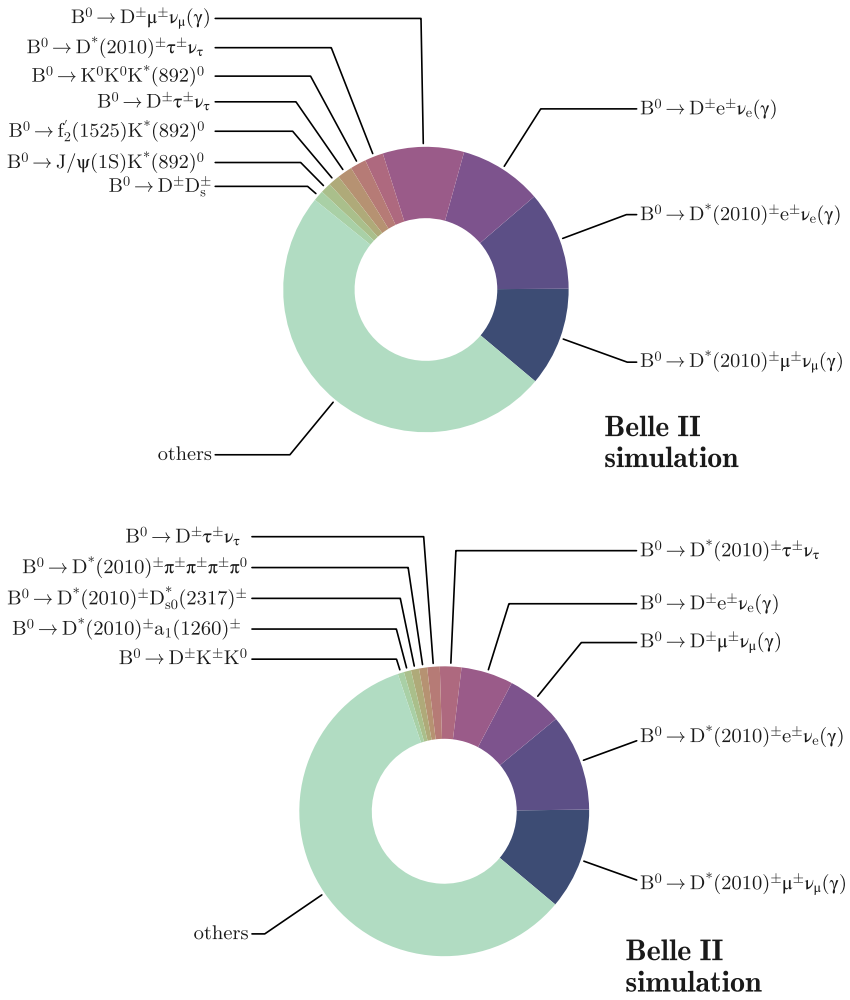


Figure 16.5: Composition of the reconstructed neutral B background in the 12 bins of the SR and CR1. The top chart shows the decays occurring in the selected signal side. The bottom chart shows the decays in the tag side. Only the 10 most frequent decay modes are explicitly reported in the diagrams.

16.5.1 Normalisation of the background yields

The leading systematic uncertainty is the normalisation uncertainty on the background yields. The yields of the seven individual background categories are allowed to float independently in the fit. A parameter corresponding to a normalisation-uncertainty modifier is introduced in the statistical model for each background sample. Each parameter is coupled to a Gaussian constraint centred at the expected background yield from simulation and with a standard deviation corresponding to 50% of the central value. This value is chosen to cover the global normalisation discrepancy of $(41 \pm 11)\%$ between the off-resonance data and the simulated continuum background observed in the control regions CR2 and CR3. The discrepancy is consistent with the value observed in the $B^+ \to K^+\nu\bar{\nu}$ analysis, discussed in Sec. 9.2.1. Also in this case the data/MC yield ratio increases in the fit region. The normalisation discrepancy is only 23% at $\mathrm{BDT}_1 > 0.9$, as reported in Sec. 15.2. The distributions of the off-resonance data and simulated continuum, with yields in simulation scaled to data by a factor 1.41, are shown in Fig. 16.6. A good shape agreement between data and MC is observed.

A data/MC normalisation discrepancy does not emerge in the control channel $B^0 \to K^{*0}J/\psi_{\to\mu^+\mu^-}$. Nevertheless, as in the previous analysis, a conservative approach is adopted, and a 50% normalisation uncertainty

is assigned in the fit model also to the charged and neutral B backgrounds.

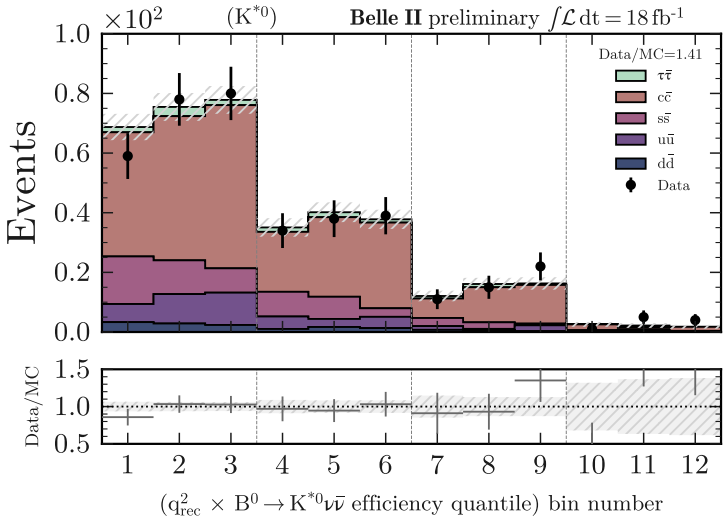


Figure 16.6: Distributions of the simulated continuum backgrounds and off-resonance data in the bins of CR3 (bins 1 to 3) and CR2 (bins 4 to 12). Yields in simulation are scaled to data, and a correction factor is introduced to account for the observed yield discrepancy data/simulation $= 1.41 \pm 0.11$.

16.5.2 Branching fractions of the leading B-background decays

A source of systematic bias is related to the fact that, in the MC simulation at Belle II, B decays are generated using only the central values of the branching fractions provided by the PDG. Systematic shape variations of the B backgrounds, due to the branching-fraction uncertainties of the leading decays in the signal side, are evaluated by means of the procedure described in detail in Sec. 9.2.2. Eight MC batches, each equivalent to an integrated luminosity of $100\,\text{fb}^{-1}$, are used for the purpose. In this analysis, central value and uncertainty of the branching fraction are covered for $\sim 76\%$ of the charged B decays and for $\sim 58\%$ of the neutral B decays in the signal side of the selected events (the ten major decays in the signal side are presented in Sec. 16.3). The estimated covariance matrix is decomposed using the procedure described in Sec. 3.2.2.1. The five principal components with largest singular values are split into the contributions from the charged and neutral B backgrounds, so that five vectors of systematic variations are obtained per sample. In the statistical model, the vectors are coupled to five nuisance parameters for correlated shape variation. The decomposition of the equivalent correlation matrix is shown for illustration purposes in Fig. 16.7. Here, the right plot corresponds to the approximation of the original matrix using Eq. 3.14. The main features of the original matrix are well reproduced.
Figure 16.8 shows the comparison between the relative branching-fraction systematic uncertainty, obtained from the diagonal of the covariance matrix, and the relative MC statistical uncertainty in the bins of the fit regions. The statistical uncertainty is larger in all the bins.

16.5.3 SM form factors

A source of systematic bias in the signal sample is related to the uncertainty on the SM form factors used to reweight the signal (see Sec. 4.2.1). To treat this systematic, the procedure presented in Sec. 9.2.3 is adapted to the SM description of the $B^0 \to K^{*0}\nu\bar{\nu}$ decay. In this case, three form factors, $V(q^2)$, $A_1(q^2)$, $A_{12}(q^2)$, are used to model the decay. The form factors depend on the parameters $\vec{\alpha}_i = (\alpha_0, \alpha_1, \alpha_2)_i$, $i = A_1, A_{12}, V$, which are estimated with a fit to Lattice QCD and LCSR results in Ref. [210], where the 9×9 covariance matrix of the fit is also provided. Using this information and the procedure in Sec. 9.2.3, the nine modified

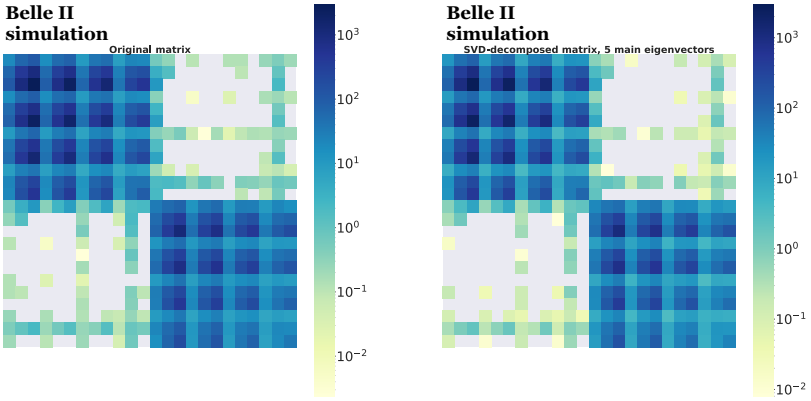


Figure 16.7: Left plot: original correlation matrix for the branching-fraction systematic uncertainty. Right plot: approximated correlation matrix after the decomposition procedure. From top to bottom, left to right, the first 12 bins are populated with events from the charged B background and the last 12 bins with events from the neutral B background.

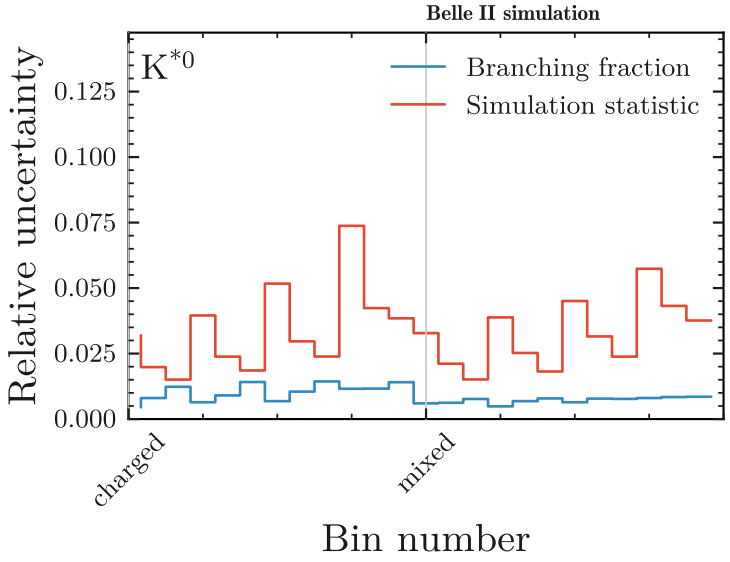


Figure 16.8: Bin-by-bin relative branching-fraction systematic uncertainty (blue step histogram) compared to the relative MC statistical uncertainty (red step histogram) in the charged and neutral (mixed) samples [1].

signal distributions shown in Fig. 16.9 are obtained. These distributions are compared to the nominal signal distribution in the 12 bins of the CR1 and SR, and nine independent vectors of systematic variations of the bin counts are determined. In the statistical model, each of them is coupled to an individual nuisance parameter for correlated shape variation.

16.5.4 PID correction

PID correction weights are applied for the $\mathrm{kaonID} > 0.9$ selection introduced in Sec. 13.2. Systematic shape variations in the samples due to the uncertainty on the weights are estimated using $800\,\mathrm{fb}^{-1}$ of simulated background and 4×10^6 signal events, with the procedure described in Sec. 9.2.4. The covariance

[1] In some plots in this chapter the term *mixed* is used to indicate the neutral B background. This is a common jargon at B factories.

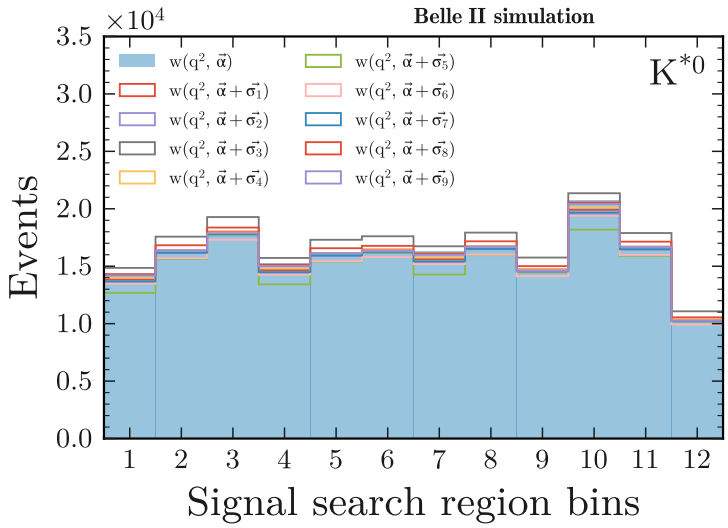


Figure 16.9: Signal yields in the 12 bins of the CR1 (bins 1 to 3) and SR (bins 4 to 12) from the reconstruction of 4×10^6 signal MC events. The blue filled histogram is obtained in the nominal SM reweighting, using the central values of $\vec{\alpha}_i = (\alpha_0, \alpha_1, \alpha_2)_i$, $i = A_1, A_{12}, V$, from Ref. [210]. The other 9 step histograms correspond to signal yields obtained with SM reweighting based on modified form factors, constructed with the 9 variations of $\vec{\alpha}_i$ derived from the decomposition of the fit covariance matrix.

matrix is decomposed to the seven principal components corresponding to the largest singular values. Seven independent vectors of systematic variations are determined per sample. The vectors are coupled to seven parameters for correlated shape variation.
For illustration, the equivalent correlation matrix is shown in Fig. 16.10 before and after the decomposition. The approximated matrix in the right plot, obtained after the decomposition, is able to reproduce the characteristic features of the original matrix. Figure 16.11 shows the relative PID-correction systematic uncertainty per bin, compared to the relative MC statistical uncertainty. The statistical uncertainty dominates over the PID systematic uncertainty.

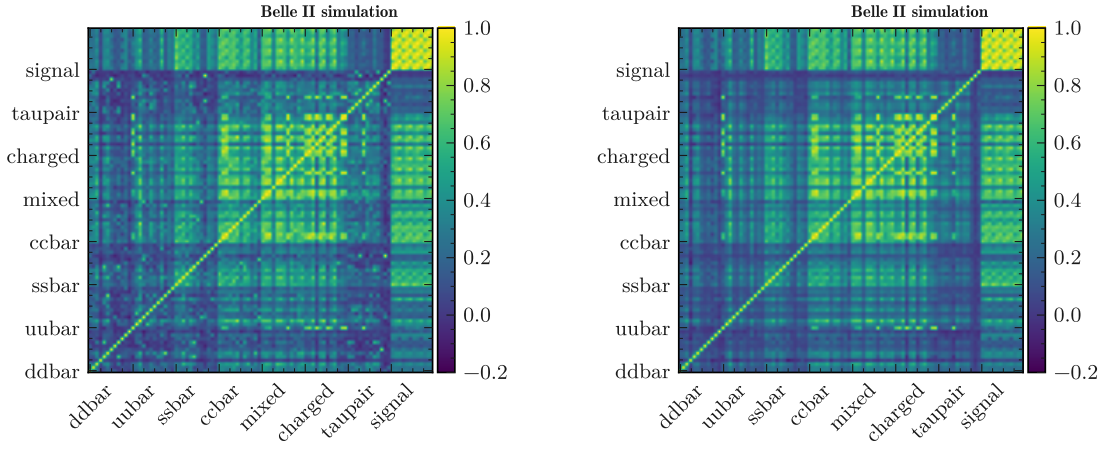


Figure 16.10: Left plot: original correlation matrix for the PID-correction systematic uncertainty. Right plot: approximation of the correlation matrix after the decomposition procedure.

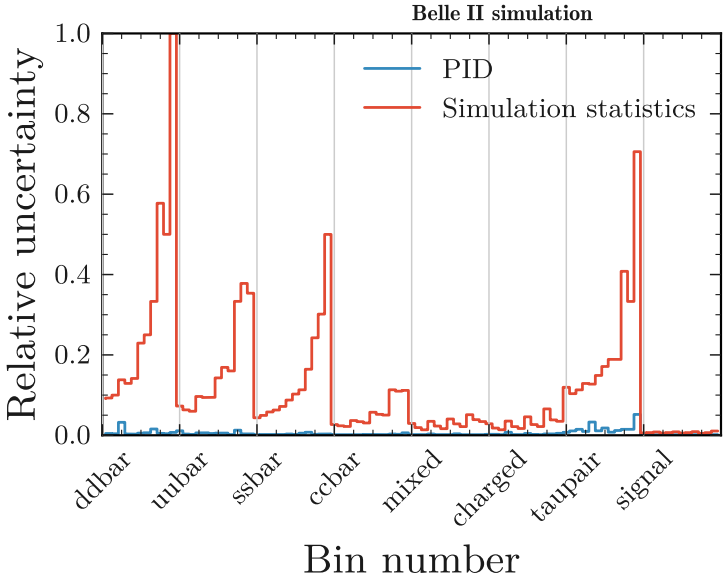


Figure 16.11: Bin-by-bin relative PID-correction systematic uncertainty (blue step histogram) compared to the relative MC statistical uncertainty (red step histogram) in all the samples.

16.5.5 Tracking efficiency

The systematic bias due to the uncertainty on the tracking efficiency is evaluated with the same technique used in the previous analysis (see Sec. 9.2.5). Also in this case, a 0.9% uncertainty per track is assumed, in agreement with the official Belle II results in Ref. [206]. This uncertainty is propagated to the analysis by simulating a 0.9% probability that one of the tracks in the event is not reconstructed. A simulated background sample equivalent to an integrated luminosity of $100\,\text{fb}^{-1}$ and 4×10^6 simulated signal events are used for this study. The background MC expectations from the reference and modified samples are smoothed before the comparison. The smoothing relies on a Gaussian KDE with kernel bandwidth $h = 1.0$. Histograms before and after the smoothing are shown in Fig. 16.12. The $\tau^+\tau^-$ and $d\bar{d}$ background samples are excluded from the study, given their minor contribution to the background in the statistical model. The relative statistical and tracking-efficiency systematic uncertainties in the 12 bins of the fit regions are illustrated for signal and backgrounds in Fig. 16.13. In the background samples, the MC statistical uncertainty is fairly larger than the systematic uncertainty. In the signal sample, the tracking systematic uncertainty prevails, but it is below 5%. In the statistical model, the vectors of systematic variations, estimated in the 12 bins of the fit regions for each sample, are coupled to a common nuisance parameter accounting for correlated shape variation.

16.5.6 Energy calibration of photon clusters

As in the previous analysis, the uncertainty on the energy of photon clusters in the ECL is assumed to be 0.5%. To evaluate the effect of this systematic uncertainty in the analysis, the energy of ECL clusters matched to photons is scaled down by 0.5%. As for the tracking efficiency, the expectations obtained with the modified samples are compared to the original expectations to determine a vector of systematic variations for each sample. In the statistical model, the vectors of variations are coupled to an individual nuisance parameter for correlated shape modification. The relative statistical and systematic uncertainties in the 12 bins of the fit regions are illustrated for signal and backgrounds in Fig. 16.14. The systematic uncertainty is minor compared to the MC statistical uncertainty.

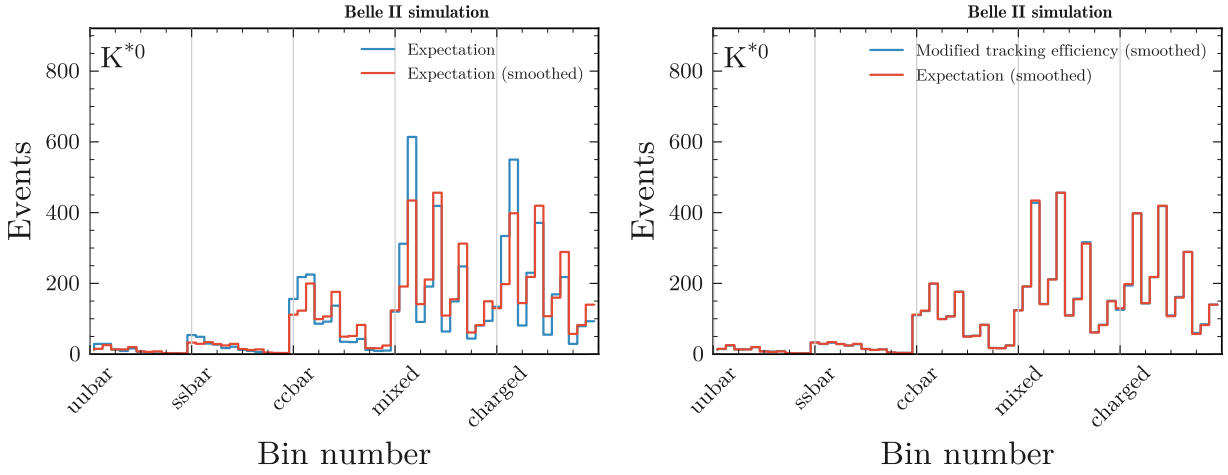


Figure 16.12: Left plot: background histograms from the reference sample obtained before (blue histogram) and after (red histogram) the Gaussian KDE smoothing in $\bar{\epsilon}$ and q^2_{rec} with bandwidth $h = 1.0$. Right plot: smoothed background histogram from the reference sample (red histogram) and smoothed background histogram from the modified sample (blue histogram), where the tracking inefficiency is simulated. For each background sample, the counts in the 12 bins of the fit regions are illustrated. The $\tau^+\tau^-$ and $d\bar{d}$ backgrounds are excluded from the study, since they represent minor background sources in the statistical model.

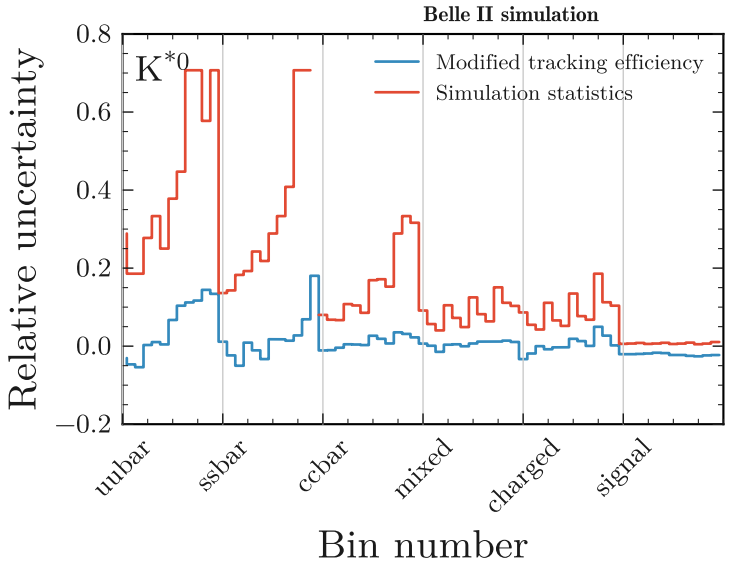


Figure 16.13: Systematic uncertainty related to the tracking efficiency: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The blue histogram corresponds to the relative systematic uncertainty. The red histogram represents the relative MC statistical uncertainty. The $\tau^+\tau^-$ and $d\bar{d}$ backgrounds are excluded from the study, since they represent minor background sources in the statistical model.

16.5.7 Energy calibration of ECL clusters not matched to photons

The uncertainty on the energy calibration of ECL clusters not matched to photons can produce a sizeable systematic bias. In the $B^+ \to K^+\nu\bar{\nu}$ analysis, a 10% uncertainty on the energy deposits of neutral hadrons,

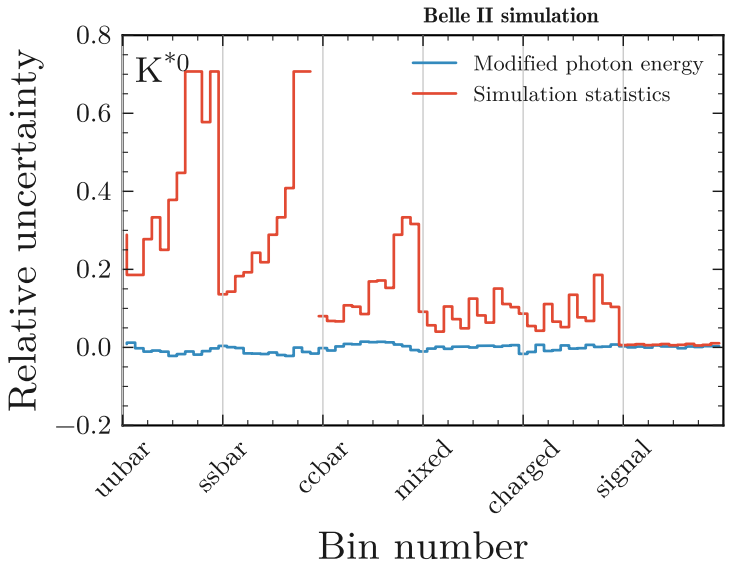


Figure 16.14: Systematic uncertainty related to the energy calibration of ECL clusters matched to photons: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The blue histogram corresponds to the relative systematic uncertainty. The red histogram represents the relative MC statistical uncertainty. The $\tau^+\tau^-$ and $d\bar{d}$ backgrounds are excluded from the study, since they represent minor background sources in the statistical model.

and neutral particles from beam background, is assumed based on the findings of the study presented in Sec. 9.2.7. This value was approved by the neutrals performance group in the Belle II Collaboration. The study is repeated with the data sample for which this analysis is optimised, and the results suggest to maintain the 10% assumption. The bias due to this systematic uncertainty is estimated by scaling down by 10% the energy of the reconstructed clusters not matched to photons. The expectations obtained with the modified simulation are compared with the original expectations to determine vectors of systematic variations for each sample, as done for the tracking efficiency and the energy of photon clusters. The vectors of variations determined for each sample are coupled to a nuisance parameter for correlated shape modification in the statistical model.
In Fig. 16.15, the relative systematic uncertainties are compared to the relative MC statistical uncertainties in the bins of the fit regions for all the samples in the statistical model. Overall, the MC statistical uncertainty dominates over the systematic uncertainty in all the background samples. The systematic uncertainty is slightly larger than the statistical uncertainty in the signal sample, but it stays below 5%.

16.6 BDT classification efficiency

A systematic uncertainty of 8% on the BDT classification efficiency is assumed in the signal sample. This systematic uncertainty is assigned given the measured data/MC classification-efficiency ratio equal to 1.09 ± 0.06 at $\bar{\epsilon} > 0.97$, which is reported in Sec. 15.1.2. In the statistical model, 8% variations of the signal expectations in the 12 bins of the SR and CR1 are coupled to a dedicated nuisance parameter implementing a correlated shape variation.

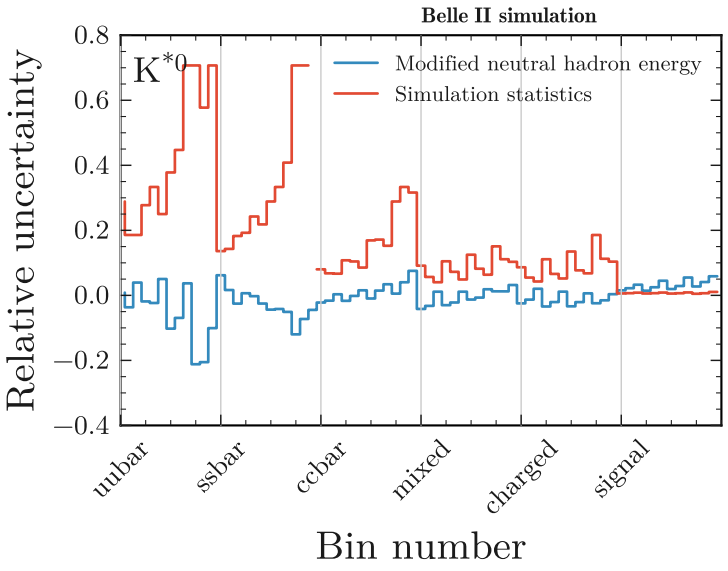


Figure 16.15: Systematic uncertainty related to the energy calibration of ECL clusters not matched to photons: comparison of relative systematic and statistical uncertainties in the 12 bins of the fit regions for signal and backgrounds. The blue histogram corresponds to the relative systematic uncertainty. The red histogram represents the relative MC statistical uncertainty. The $\tau^+\tau^-$ and $d\bar{d}$ backgrounds are excluded from the study, since they represent minor background sources in the statistical model.

16.7 Mismodelling of hadronic resonances

The simulated background samples used in the analysis are affected by a partial mismodelling of the wide hadronic resonances generated by fragmentation in the MC simulation. Correction weights are used to fix the mismodelling. This systematic bias is taken into account in the statistical model by means of a nuisance parameter for correlated shape modification. The parameter is coupled to the systematic variations determined for each background sample by comparing the expectations obtained without and with the application of the correction weights.

16.8 Impact of the systematic uncertainties

It is interesting to evaluate the impact of the various systematic uncertainties on the measurement. To do this, the expected upper limit on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction is computed with `sghf` at 90% CL as a function of the sources of systematic uncertainty included in the statistical model. The results are illustrated in Fig. 16.16. The largest variation corresponds to the introduction of the 50% systematic uncertainty on the normalisations of the background yields, which produces a 55% increase of the limit. Sizeable increments are observed also with the inclusion of the systematic uncertainties related to the tracking efficiency and to the energy calibration of the ECL clusters not matched to photons: $+10\%$ and $+8\%$ increments of the upper limit, respectively. Another source producing a variation $> 1\%$ is the systematic uncertainty due to the mismodelling of hadronic resonances. Based on these results, the improvement of the continuum simulation is strongly needed for a better measurement.

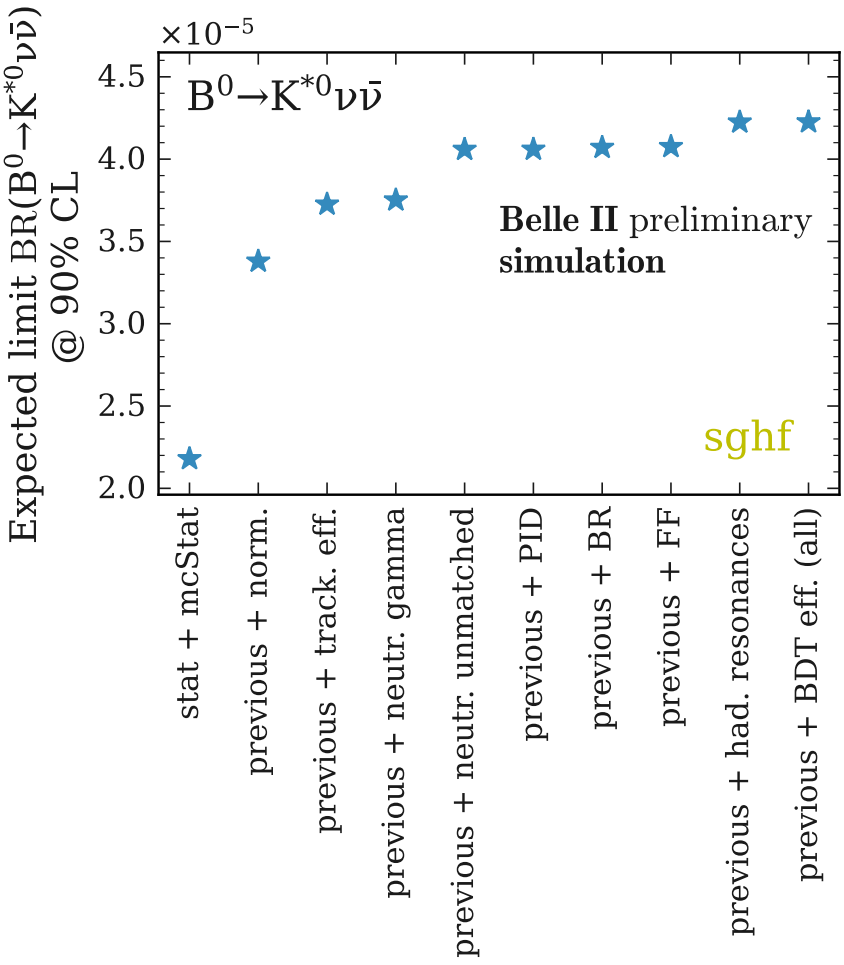


Figure 16.16: Expected upper limit (`sghf`) on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction computed at 90% CL as a function of the sources of systematic uncertainty included in the statistical model.

Chapter 17

Results and discussion

This chapter presents the preliminary results of this search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay with an inclusive tagging method. What is shown and discussed in this chapter is based only on MC expectations. Section 17.1 summarises the current status of the analysis. Section 17.2 presents the expected upper limit on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction. In Section 17.3 this search is compared to the previous measurements. The performance of the inclusive tagging is discussed in Section 17.4. In conclusion, Section 17.5 gives an outlook of the measurement.

17.1 Preamble

As in the $B^+ \to K^+\nu\bar{\nu}$ analysis, the fitting procedure is validated. The same tests described in Sec. 9.3 are performed. The results of these tests, and the pre-unblinding results of the checks presented in Sec. 10.1, suggest that the statistical model describes the data well. The analysis is also validated in a signal-region sideband and no issue is found. However, the analysis is still under internal review, and the full unblinding of the signal region has not been performed yet. For this reason, the fit to the data is not presented and only the results based on MC expectations are described and discussed in this chapter.

17.2 Results

The expected upper limit on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction, for an integrated luminosity of $189\,\mathrm{fb}^{-1}$, is estimated with the `pyhf` using the $\mathrm{CL_s}$ method. The test statistic $\tilde{q}_\mu$, defined in Eq. 3.29, is chosen and the $\mathrm{CL_s}$ computation is performed by means of the asymptotic formulae presented in Sec. 3.5.2. The scan of the expected $\mathrm{CL_s}$ as a function of hypothesised values of the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction is shown in Fig. 17.1. An expected upper limit of 4.4×10^{-5} is evaluated at 90% CL. The figure illustrates also the expected $\mathrm{CL_s}$ uncertainty bands, defined in Sec. 3.5, which correspond to $\pm 1\sigma$ (green band) and $\pm 2\sigma$ (yellow band) variations of the signal strength from the background-only hypothesis.
The expected upper limit estimated with the `sghf` is 4.2×10^{-5}. This value is computed using the Gaussian approximation described in Sec. 3.5.3 and it is in good agreement with the $\mathrm{CL_s}$ estimate.

17.3 Comparison with previous measurements

The expected upper limit set in this analysis is compared in Table 17.1 to the expected and observed limits from previous searches for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay. The limit estimated in this work is in the ballpark of the published results, despite the smaller integrated luminosity. The most stringent upper limit on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction at 90% CL remains the one set by Belle in Ref. [126]. That measurement is performed with a semileptonic tagging using a data sample of $711\,\mathrm{fb}^{-1}$. However, the search is optimised with signal modelled by a generic three-body phase-space kinematics, while in the analysis of this thesis the signal is reweighted following the SM prediction. This difference has to be taken into account in the

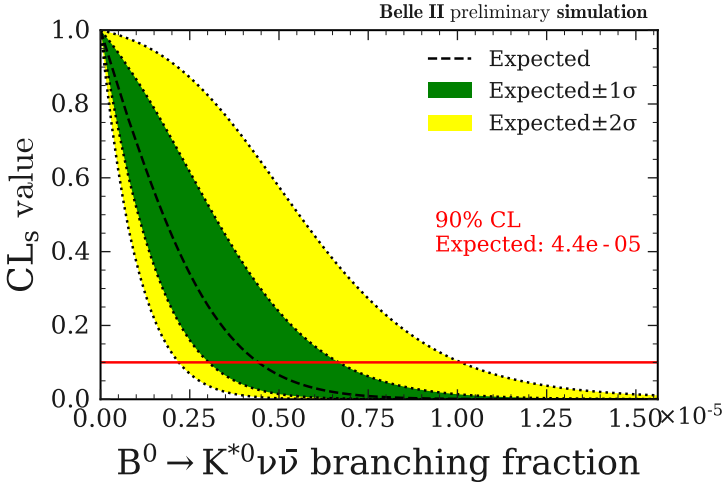


Figure 17.1: Expected $\mathrm{CL_s}$ value as a function of the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction, and corresponding expected upper limit at 90% CL, computed for an integrated luminosity of $189\,\mathrm{fb}^{-1}$.

comparison of the results.
The expected limit found with the inclusive tagging is already competitive with the limits estimated in the Belle SM search presented in Ref. [123]. That search is performed with a hadronic tagging using a data sample of $711\,\mathrm{fb}^{-1}$, an integrated luminosity approximately four times larger than the one for which this analysis is optimised.

Measurement	Expected UL	Observed UL	Tagging	$L\,[\mathrm{fb}^{-1}]$
Babar (2013), Ref. [124]	–	$< 12 \times 10^{-5}$	semileptonic + hadronic	429
Belle (2013), Ref. [123]	$< 4.6 \times 10^{-5}$	$< 5.5 \times 10^{-5}$	hadronic	711
Belle (2017), Ref. [126]	$< 2.4 \times 10^{-5}$	$< 1.8 \times 10^{-5}$	semileptonic	711
Belle II (2022), preliminary	$< 4.4 \times 10^{-5}$	–	inclusive	189

Table 17.1: Expected and observed upper limits on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction estimated at 90% CL in previous measurements, compared to the preliminary result of this analysis.

17.4 Performance of the inclusive tagging

In this analysis, the efficiency in the signal region (SR) is by construction equal to 2.25% (3% in the extended CR1 + SR). Table 17.2 gives an overview of the efficiencies in the signal regions of searches performed with the hadronic and semileptonic taggings. A rigorous comparison should be made at the same signal purity, but the values in the table give a measure of the irreducible drop in signal efficiency caused by the low semileptonic and hadronic tagging efficiencies.
The signal efficiency of the inclusive tagging is evaluated in the SR also as a function of the generated di-neutrino invariant mass squared, q^2, as illustrated in Fig. 17.2. The q^2 range goes from zero to the kinematic limit $q^2_{\mathrm{max}} = 19.2\,\mathrm{GeV}^2/c^4$, and it is divided in bins of $2\,\mathrm{GeV}^2/c^4$. The efficiency is highest in the second bin, $2\,\mathrm{GeV}^2/c^4 < q^2 < 4\,\mathrm{GeV}^2/c^4$, reaching a maximum of 5.6%. It progressively decreases for $q^2 > 4\,\mathrm{GeV}^2/c^4$ and drops to zero for $q^2 > 16\,\mathrm{GeV}^2/c^4$. Also in this analysis, the inclusive tagging is highly efficient at low q^2, where a possible invisibly-decaying dark-scalar resonance is predicted by dark-matter models [117].
The performance of the inclusive tagging is compared to the hadronic and semileptonic taggings adopted in the previous searches using the precision of the results as a measure. In the comparison, the uncertainty

on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction reported for this work is the expected uncertainty estimated with the `sghf` for $189\,\mathrm{fb}^{-1}$: $\sigma_{\mathrm{BR}} = 2.6 \times 10^{-5}$. Assuming that σ_{BR} scales as $1/\sqrt{L}$, the uncertainties estimated in Refs. [123, 124, 126] are scaled to the integrated luminosity of $189\,\mathrm{fb}^{-1}$ as summarised in Table 17.3. Under this assumption, the inclusive tagging performs a factor 3.6 better than the hadronic tagging used by Belle in 2013 [123] and approximately 60% better than the combination of hadronic and semileptonic taggings used by BaBar in 2013 [124]. The performance of the inclusive tagging is worse only compared to the semileptonic tagging adopted by Belle in 2017, which gives a result 20% more precise.

Measurement	Tagging	Signal efficiency
BaBar (2013), Ref. [124]	hadronic + semileptonic	$1.3 \times 10^{-4} + 6.9 \times 10^{-4}$
Belle (2013), Ref. [123]	hadronic	1.44×10^{-4}
Belle (2017), Ref. [126]	semileptonic	0.51×10^{-3}
Belle II (2022), this work	inclusive	2.25×10^{-2}

Table 17.2: Comparison between the signal efficiency in the signal region (SR), obtained in this search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay with the inclusive tagging, and the efficiencies in the signal regions of the previous searches with hadronic and semileptonic taggings. Precise definitions of the signal regions can be found in the cited papers.

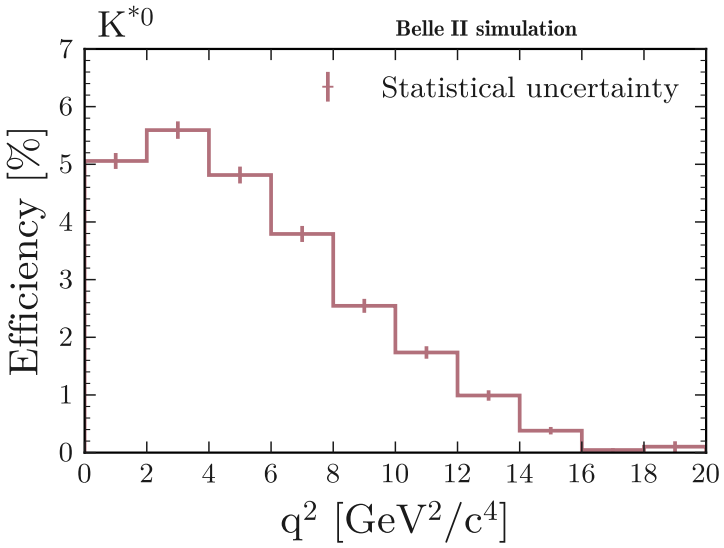


Figure 17.2: Signal efficiency evaluated in bins of generated di-neutrino invariant mass squared, q^2, for simulated $B^0 \to K^{*0}\nu\bar{\nu}$ events in the signal region (SR). Each bin corresponds to an interval of $2\,\mathrm{GeV}^2/c^4$, and the range extends up to the kinematic limit $\mathrm{q}^2_{\mathrm{max}} = 19.2\,\mathrm{GeV}^2/c^4$. The error bars indicate the statistical uncertainty.

17.5 Outlook

The discovery of $B^0 \to K^{*0}\nu\bar{\nu}$ would complement that of $B^+ \to K^+\nu\bar{\nu}$. The two decays probe largely different regions in the parameter space of the Wilson Coefficients [116]. Moreover, polarisation measurements in $B^0 \to K^{*0}\nu\bar{\nu}$ provide further tests of weak- and strong-interaction dynamics [212].
The future prospects of the search for the $B^0 \to K^{*0}\nu\bar{\nu}$ decay can be explored based on the expected branching-fraction uncertainty estimated in this analysis. As done in Sec. 11.3, sensitivity projections are

Measurement	$L\,(\mathrm{fb}^{-1})$	$\sigma_{\mathrm{BR}} \times 10^5$	$\sigma_{\mathrm{BR}}^{\mathrm{scaled}} \times 10^5$
BaBar, hadronic and semileptonic taggings [124]	429	2.8	4.2
Belle, hadronic tagging [123]	711	6.2	12.0
Belle, semileptonic tagging [126]	711	1.1	2.1
Belle II, inclusive tagging (preliminary)	189	2.6	2.6

Table 17.3: Uncertainties on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction, σ_{BR}, estimated in samples of integrated luminosity L with different tagging techniques, and values scaled to $189\,\mathrm{fb}^{-1}$. The uncertainty quoted for this analysis is the expected σ_{BR} estimated with the `sghf` for $189\,\mathrm{fb}^{-1}$.

made for a set of benchmark luminosities, under the assumption that σ_{BR} scales as the inverse square root of the integrated luminosity. The baseline and improved scenarios are considered. In the latter, a 50% increase in signal efficiency at the same background level is assumed.
The projections are summarised in Table 17.4. Based on these extrapolations, sensitivity to the SM signal rate at 3σ level should be achieved in the improved scenario with a data sample of $\sim 10\,\mathrm{ab}^{-1}$. The complete Belle II sample of $50\,\mathrm{ab}^{-1}$ would be needed to reach the SM accuracy. More rigorous projections, taking into account also various improvements of the systematic uncertainties, can be found in Ref. [209]. The leading systematic uncertainty in this work corresponds to the normalisation of the background yields, thus a major improvement of the analysis depends on its reduction. Increased sensitivity can also be achieved by exploiting alternative classifiers, such as neural networks, and by further optimising the signal candidate selection and the ROE reconstruction.

$L(\mathrm{ab}^{-1})$	Baseline scenario	Improved scenario
1	1.45	1.18
5	0.65	0.53
10	0.46	0.37
50	0.20	0.17

Table 17.4: Projections, in the baseline and improved scenarios, of the expected uncertainty on the $B^0 \to K^{*0}\nu\bar{\nu}$ branching fraction (relative to the SM central value) as a function of integrated luminosity L. The values are obtained by assuming that the expected uncertainty estimated in this work for $189\,\mathrm{fb}^{-1}$, $\sigma_{\mathrm{BR}} = 2.6 \times 10^{-5}$, scales as $1/\sqrt{L}$.

Chapter 18

Bibliography

[1] F. Abudinén et al. (Belle II Collaboration). Search for $B^+ \to K^+\nu\bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II. *Phys. Rev. Lett.*, 127:181802, Oct. 2021.

[2] F. Dattola, S. Glazov, S. Kurz, C. Praz, and S. Stefkova. Search for $B^+ \to K^+\nu\bar{\nu}$ decays using an inclusive tagging method at Belle II. *BELLE2-NOTE-PH-2020-057*.

[3] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer. pyhf: pure-Python implementation of HistFactory statistical models. *J. Open Source Softw.*, 6:2823, Feb. 2021.

[4] V. Bertacchi et al. Track finding at Belle II. *Comput. Phys. Commun.*, 259:107610, Feb. 2021.

[5] F. Abudinén et. al (Belle II Collaboration). Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment. *Chin. Phys. C*, 44:021001, Feb. 2020.

[6] F. Abudinén et al. (Belle II Collaboration). Search for Axionlike Particles Produced in e^+e^- Collisions at Belle II. *Phys. Rev. Lett.*, 125:161806, Oct. 2020.

[7] F. Abudinén et al. (Belle II Collaboration). Precise Measurement of the D^0 and D^+ Lifetimes at Belle II. *Phys. Rev. Lett.*, 127:211801, Nov. 2021.

[8] F. Abudinén et al. (Belle II Collaboration). Combined analysis of Belle and Belle II data to determine the CKM angle ϕ_3 using $B^+ \to D(K^0_S h^- h^+)h^+$ decays. *J. High Energ. Phys.*, 02:063, Feb. 2022.

[9] I. Adachi et al. (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.*, 124:141801, Apr. 2020.

[10] T. E. Browder et al. New physics at a Super Flavor Factory. *Rev. Mod. Phys.*, 81:1887–1941, Dec. 2009.

[11] R. P. Feynman and M. Gell-Mann. Theory of the Fermi Interaction. *Phys. Rev.*, 109:193–198, Jan. 1958.

[12] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, Jun. 1963.

[13] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2:1285–1292, Oct. 1970.

[14] J. D. Bjorken and S. L. Glashow. Elementary Particles and SU(4). *Phys. Lett.*, 11:255–257, Aug. 1964.

[15] J. E. Augustin et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, 33:1406–1408, Dec. 1974.

[16] J. J. et al. Aubert. Experimental Observation of a Heavy Particle J. *Phys. Rev. Lett.*, 33:1404–1406, Dec. 1974.

[17] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.*, 13:138–140, Jul. 1964.

[18] M. Kobayashi and T. Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, Feb. 1973.

[19] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, Aug. 1977.

[20] F. Abe et al. (CDF Collaboration). Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr. 1995.

[21] S. Abachi et al. (D0 Collaboration). Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632–2637, Apr. 1995.

[22] H. Albrecht et al. (ARGUS Collaboration). Observation of B^0-B^0 Mixing. *Phys. Lett. B*, 192:245–252, Jun. 1987.

[23] M. S. Alam et al. (CLEO Collaboration). First Measurement of the Rate for the Inclusive Radiative Penguin Decay $b \to s\gamma$.

[24] R. Ammar et al. (CLEO Collaboration). Evidence for penguin-diagram decays: First observation of B$\to K^*(892)\gamma$. *Phys. Rev. Lett.*, 71:674–678, Aug. 1993.

[25] K. Abe et al. (Belle Collaboration). Observation of Large CP Violation in the Neutral B Meson System. *Phys. Rev. Lett.*, 87:091802, Aug. 2001.

[26] B. Aubert et al. (BaBar Collaboration). Observation of CP violation in the B^0 meson system. *Phys. Rev. Lett.*, 87:091801, Aug. 2001.

[27] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J. C*, 74(11), Nov. 2014.

[28] R. Aaij et al. (LHCb Collaboration). Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$. *Phys. Rev. Lett.*, 111:191801, Nov. 2013.

[29] R. Aaij et al. (LHCb Collaboration). Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays. *J. High Energy Phys.*, 06, Jun. 2014.

[30] R. Aaij et al. (LHCb Collaboration). Angular analysis of the $B^0 \to K^{(*0)}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity. *J. High Energy Phys.*, 02:104, Feb. 2016.

[31] R. Aaij et al. (LHCb Collaboration). Test of lepton universality with $B^0 \to K^{*0}l^+l^-$ decays. *J. High Energy Phys.*, 08:055, Aug. 2017.

[32] R. Aaij et al. (LHCb Collaboration). Search for Lepton-Universality Violation in $B^+ \to K^+\ell^+\ell^-$ Decays. *Phys. Rev. Lett.*, 122:191801, May 2019.

[33] R. Aaij et al. (LHCb Collaboration). Measurement of CP-averaged observables in the $B^0 \to K^{(*0)}\mu^+\mu^-$ decay. *Phys. Rev. Lett.*, 125:011802, Jul. 2020.

[34] R. Aaij et al. (LHCb Collaboration). Test of lepton universality in beauty-quark decays. *Nature Phys.*, 18:277–282, Mar. 2022.

[35] S. Hirose et al. (Belle Collaboration). Measurement of the τ Lepton Polarization and $R(D^*)$ in the Decay $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$. *Phys. Rev. Lett.*, 118:211801, May 2017.

[36] M. Huschle et al. (Belle Collaboration). Measurement of the branching ratio of $B \to D^*\tau^-\bar{\nu}_\tau$ relative to $B \to D^*\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. *Phys. Rev. D*, 92(7):072014, Oct. 2015.

[37] J. P. Lees et al. (BaBar Collaboration). Evidence for an Excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ Decays. *Phys. Rev. Lett.*, 109:101802, Sep. 2012.

[38] J. P. Lees et al. (BaBar Collaboration). Measurement of an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays and implications for charged Higgs bosons. *Phys. Rev. D*, 88:072012, Oct. 2013.

[39] R. Aaij et al. (LHCb Collaboration). Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$. *Phys. Rev. Lett.*, 115:111803, Sep. 2015.

[40] G. W. Bennett et al. (Muon $g2$ Collaboration). Final report of the E821 muon anomalous magnetic moment measurement at BNL. *Phys. Rev. D*, 73:072003, Apr. 2006.

[41] B. Abi et al. (Muon $g2$ Collaboration). Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr. 2021.

[42] P. Koppenburg. Flavour Anomalies, Mar. 2022.

[43] P. A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Prog. Theor. Exp. Phys.*, 2020:083C01, Aug. 2020.

[44] Wikimedia Commons. The Standard Model of elementary particles, 2019.

[45] C. N. Yang and R. L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys. Rev.*, 96:191–195, Oct. 1954.

[46] M. Y. Han and Y. Nambu. Three-Triplet Model with Double $\mathrm{SU}(3)$ Symmetry. *Phys. Rev.*, 139:B1006–B1010, Aug. 1965.

[47] M. Gell-Mann. Symmetries of Baryons and Mesons. *Phys. Rev.*, 125:1067–1084, Feb. 1962.

[48] S. Weinberg. Non-Abelian Gauge Theories of the Strong Interactions. *Phys. Rev. Lett.*, 31:494–497, Aug. 1973.

[49] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, Feb. 1961.

[50] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266", Nov. 1967.

[51] A. Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, May 1968.

[52] T. Nakano and K. Nishijima. Charge Independence for V-particles. *Prog. Theor. Exp. Phys.*, 10:581–582, Nov. 1953.

[53] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Nuovo Cim.*, 4:848–866, Apr. 1956.

[54] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, Aug. 1964.

[55] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct. 1964.

[56] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos. *A Historical Profile of the Higgs Boson*, chapter 14, pages 255–274. Oct. 2016.

[57] L. Chau and W. Keung. Comments on the Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 53:1802–1805, Nov. 1984.

[58] L. Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945–1947, Nov. 1983.

[59] A. J. Buras, M. E. Lautenbacher, and G. Ostermaier. Waiting for the top quark mass, $K^+ \to \pi^+\nu\bar{\nu}$, $B_s^0 - B_s^0$ mixing, and CP asymmetries in B decays. *Phys. Rev. D*, 50:3433–3446, Sep 1994.

[60] J. Charles et al. CP violation and the CKM matrix: assessing the impact of the asymmetric B factories. *Eur. Phys. J. C*, 41(1):1–131, May 2005.

[61] C. Jarlskog. Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation. *Phys. Rev. Lett.*, 55:1039–1042, Sep. 1985.

[62] A. Höcker, H. Lacker, S. Laplace, and F. Le Diberder. A new approach to a global fit of the CKM matrix. *Eur. Phys. J. C*, 21:225–259, Jun. 2001.

[63] G. Eigen, G. Dubois-Felsmann, D. G. Hitlin, and F. C. Porter. Global CKM fits with the scan method. *Phys. Rev. D*, 89, Feb. 2014.

[64] M. Bona et al. The 2004 UTfit collaboration report on the status of the unitarity triangle in the standard model. *J. High Energy Phys.*, 2005:028–028, Jul. 2005.

[65] M. Bona et al. Model-independent constraints on $\Delta F = 2$ operators and the scale of new physics. *J. High Energy Phys.*, 2008:049–049, Mar. 2008.

[66] B. Pontecorvo. Inverse beta processes and nonconservation of lepton charge. *Zh. Eksp. Teor. Fiz.*, 34:247, Jan. 1958.

[67] Z. Maki, M. Nakagawa, and S. Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, Nov. 1962.

[68] E. Majorana and L. Maiani. A symmetric theory of electrons and positrons. *Il Nuovo Cimento*, 14:171–84, 1937.

[69] T. Yanagida. Horizontal Symmetry and Masses of Neutrinos. *Prog. Theor. Phys.*, 64:1103–1105, Sep. 1980.

[70] 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 $K^0 - K^0$. *Prog. Theor. Phys.*, 65:297–314, Jan. 1981.

[71] B. Grinstein. Lectures on Flavor Physics and CP Violation, Jan. 2017.

[72] K. G. Wilson. Non-Lagrangian Models of Current Algebra. *Phys. Rev.*, 179:1499–1512, Mar. 1969.

[73] K. G. Wilson and W. Zimmermann. Operator product expansions and composite field operators in the general framework of quantum field theory. *Commun. Math. Phys.*, 24:87–106, Jun. 1972.

[74] K. G. Wilson. The renormalization group and critical phenomena. *Rev. Mod. Phys.*, 55:583–600, Jul. 1983.

[75] D. J. Gross and F. Wilczek. Ultraviolet Behavior of Non-Abelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun. 1973.

[76] G. Buchalla, A. J. Buras, and M. E. Lautenbacher. Weak decays beyond leading logarithms. *Rev. Mod. Phys.*, 68:1125–1244, Oct. 1996.

[77] K. G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct. 1974.

[78] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov. QCD and Resonance Physics. Theoretical Foundations. *Nucl. Phys. B*, 147:385–447, Feb. 1979.

[79] H. Georgi. An Effective Field Theory for Heavy Quarks at Low-energies. *Phys. Lett. B*, 240:447–450, Apr. 1990.

[80] E. Eichten and B. R. Hill. An Effective Field Theory for the Calculation of Matrix Elements Involving Heavy Quarks. *Phys. Lett. B*, 234:511–516, Jan. 1990.

[81] G. Altarelli, N. Cabibbo, G. Corbo, L. Maiani, and G. Martinelli. Leptonic Decay of Heavy Flavors: A Theoretical Update. *Nucl. Phys. B*, 208:365–380, Dec. 1982.

[82] J. Chay, H. Georgi, and B. Grinstein. Lepton energy distributions in heavy meson decays from QCD. *Phys. Lett. B*, 247:399–405, Sep. 1990.

[83] A. V. Manohar and M. B. Wise. Inclusive semileptonic B and polarized Λ_b decays from QCD. *Phys. Rev. D*, 49:1310–1329, Feb. 1994.

[84] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto. Global analysis of $b \to s\ell\ell$ anomalies. *J. High Energy Phys.*, 06:092, Jun. 2016.

[85] 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. *J. High Energy Phys.*, 02:184, Feb. 2015.

[86] G. Buchalla and A. J. Buras. QCD corrections to rare K- and B-decays for arbitrary top quark mass. *Nucl. Phys. B*, 400:225–239, Jul. 1993.

[87] M. Misiak and J. Urban. QCD corrections to FCNC decays mediated by Z-penguins and W-boxes. *Phys. Lett. B*, 451:161–169, Apr. 1999.

[88] 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*, 548:309–327, May 1999.

[89] J. Brod, M. Gorbahn, and E. Stamou. Two-loop electroweak corrections for the $K \to \pi\nu\overline{\nu}$ decays. *Phys. Rev. D*, 83:034030, Feb. 2011.

[90] I.I. Balitsky, V.M. Braun, and A.V. Kolesnichenko. Radiative decay $\sigma^+ \to p\gamma$ in quantum chromodynamics. *Nucl. Phys. B*, 312:509–550, Jan. 1989.

[91] V. M. Braun and I. E. Filyanov. QCD sum rules in exclusive kinematics and pion wave function. *Z. Phys. C*, 44:157–166, Mar. 1989.

[92] V. L. Chernyak. *Vacuum Structure and QCD Sum Rules*, volume 10. Elsevier, Jan. 1992.

[93] T. Blake, G. Lanfranchi, and D. M. Straub. Rare B decays as tests of the Standard Model. *Prog. Part. Nucl. Phys.*, 92:50–91, Jan. 2017.

[94] G. 't Hooft et al. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B*, 59:135–157, 1980.

[95] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967.

[96] J. Aebischer, J. Kumar, P. Stangl, and D. M. Straub. A Global Likelihood for Precision Constraints and Flavour Anomalies. *Eur. Phys. J. C*, 79:509, Jun. 2019.

[97] W. Buchmüller and D. Wyler. Effective Lagrangian analysis of new interactions and flavour conservation. *Nuclear Physics B*, 268:621–653, May 1986.

[98] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *J. High Energy Phys.*, 10:085, Oct. 2010.

[99] N. Rajeev and R. Dutta. Consequences of $b \to s\mu^+\mu^-$ anomalies on $B \to K^{(*)} \to \nu\bar{\nu}$, $B_s \to (\eta, \eta') \to \nu\bar{\nu}$ and $B_s \to \phi\nu\bar{\nu}$ decay observables. *Phys. Rev. D*, 105(11):115028, Dec. 2022.

[100] P. Langacker. The physics of heavy Z' gauge bosons. *Rev. Mod. Phys.*, 81:1199–1228, Aug. 2009.

[101] A. Leike. The phenomenology of extra neutral gauge bosons. *Phys. Rep.*, 317:143–250, Aug. 1999.

[102] L. J. Hewett and T. G. Rizzo. Low-energy phenomenology of superstring-inspired E6 models. *Phys. Rep.*, 183:193–381, Nov. 1989.

[103] J. C. Pati and A. Salam. Lepton number as the fourth color. *Phys. Rev. D*, 10:275–289, Jul. 1974.

[104] H. Georgi and S. L. Glashow. Unity of All Elementary-Particle Forces. *Phys. Rev. Lett.*, 32:438–441, Feb. 1974.

[105] W. Buchmüller, R. Rückl, and D. Wyler. Leptoquarks in lepton-quark collisions. *Phys. Lett. B*, 191:442–448, Jun. 1987.

[106] B. C. Allanach, J. M. Butterworth, and Tyler Corbett. Collider constraints on $Z^{'}$ models for neutral current B-anomalies. *J. High Energ. Phys.*, 2019(8), Aug. 2019.

[107] T. E. Browder, N. G. Deshpande, R. Mandal, and R. Sinha. Impact of $B \to K\nu\bar{\nu}$ measurements on beyond the Standard Model theories. *Phys. Rev. D*, 104:053007, Sep. 2021.

[108] W. Altmannshofer and P. Stangl. New physics in rare B decays after Moriond 2021. *Eur. Phys. J. C*, 81, oct 2021.

[109] M. Alguero et al. $b \to s\ell^+\ell^-$ global fits after R_{K_S} and $R_{K^{*+}}$. *Eur. Phys. J. C*, 82:326, Apr. 2022.

[110] A.M. Sirunyan et al. (CMS Collaboration). Search for singly and pair-produced leptoquarks coupling to third-generation fermions in proton-proton collisions at s=13 TeV. *Phys. Lett. B*, 819:136446, Aug. 2021.

[111] A. Crivellin, C. Greub, F. Saturnino, and D. Müller. Importance of Loop Effects in Explaining the Accumulated Evidence for New Physics in B Decays with a Vector Leptoquark. *Phys. Rev. Lett.*, 122, Jan. 2019.

[112] J. Fuentes-Martín, G. Isidori, M. König, and N. Selimović. Vector leptoquarks beyond tree level. III. Vectorlike fermions and flavor-changing transitions. *Phys. Rev. D*, 102, Dec. 2020.

[113] C. Cornella et al. Reading the footprints of the B-meson flavor anomalies. *J. High Energy Phys.*, 2021, Aug. 2021.

[114] S. Chakraborty et al. Heavy QCD axion in $b \to s$ transition: Enhanced limits and projections. *Phys. Rev. D*, 104(5), Sep. 2021.

[115] C. Bird, P. Jackson, R. Kowalewski, and M. Pospelov. Dark Matter Particle Production in $b \to s$ Transitions with Missing Energy. *Phys. Rev. Lett.*, 93(20), Nov. 2004.

[116] T. Felkl, S. L. Li, and M. A. Schmidt. A Tale of Invisibility: Constraints on New Physics in $b \to s\nu\nu$. *J. High Energ. Phys.*, 2021(118), Dec. 2021.

[117] A. Filimonova, R. Schäfer, and S. Westhoff. Probing dark sectors with long-lived particles at Belle II. *Phys. Rev. D*, 101:095006, May 2020.

[118] A. Kachanovich, U. Nierste, and I. Nišandžić. Higgs portal to dark matter and $B \to K^{(*)}$ decays. *Eur. Phys. J. C*, 80(7), Jul. 2020.

[119] T. Ferber. Prospects for long-lived particle searches at Belle II. Jul. 2021.

[120] J. F. Kamenik and C. Smith. FCNC portals to the dark sector. *J. High Energ. Phys.*, 03:090, 2012.

[121] I. Boiarska et al. Phenomenology of GeV-scale scalar portal. *J. High Energ. Phys.*, 11:162, Nov. 2019.

[122] T. E. Browder et al. (CLEO Collaboration). Search for $B \to \tau\nu$ and $B \to K\nu\bar{\nu}$. *Phys. Rev. Lett.*, 86:2950–2954, Apr. 2001.

[123] O. Lutz et al. (Belle Collaboration). Search for $B \to h^{(*)}\nu\overline{\nu}$ with the full Belle $\Upsilon(4S)$ data sample. *Phys. Rev. D*, 87:111103, Jun. 2013.

[124] J. P. Lees et al. (BaBar Collaboration). Search for $B \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays. *Phys. Rev. D*, 87:112005, Jun. 2013.

[125] P. del Amo Sanchez et al. (BaBar Collaboration). Search for the rare decay $B \to K\nu\overline{\nu}$. *Phys. Rev. D*, 82:112002, Dec. 2010.

[126] J. Grygier et al. (Belle Collaboration). Search for $B \to h\nu\overline{\nu}$ decays with semileptonic tagging at Belle. *Phys. Rev. D*, 96:091101, Nov. 2017.

[127] M. Feindt et al. A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories. *Nucl. Instrum. Methods Phys. Res. A*, 654:432–440, Oct. 2011.

[128] T. Keck et al. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. *Comput. Software Big Sci.*, 3:6, Feb. 2019.

[129] E. Kou et al. The Belle II Physics Book. *Prog. Theor. Exp. Phys.*, 2019, Dec. 2019.

[130] E. Ganiev. Measurement of the branching fraction, longitudinal polarization fraction, and charge-parity violating asymmetry in $B^+ \to \rho^+\rho^0$ decays at Belle. Sep. 2021.

[131] N. Toge et al. KEKB B-factory Design Report, Aug. 1995.

[132] K. Akai, K. Furukawa, and H. Koiso. SuperKEKB collider. *Nucl. Instrum. Methods Phys. Res. A*, 907:188–199, Nov. 2018.

[133] SuperB Collaboration. SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report, Mar. 2007.

[134] P. Branchini (Belle II Collaboration). The Belle II Experiment: Status and Prospects. *Universe*, 4:101, Oct. 2018.

[135] P. M. Lewis et al. First Measurements of Beam Backgrounds at SuperKEKB. *Nucl. Instrum. Meth. A*, 914:69–144, Jan. 2019.

[136] T. Abe et al. Belle II Technical Design Report, Nov. 2010.

[137] H. Ye et al. Commissioning and performance of the Belle II pixel detector. *Nucl. Instrum. Methods Phys. Res. A*, 987:164875, Jan. 2021.

[138] J. Kemmer and G. Lutz. New detector concepts. *Nucl. Instrum. Methods Phys. Res. A*, 253:365–377, Jan. 1987.

[139] K. Adamczyk et al. (Belle II SVD). The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector. Jan. 2022.

[140] G. Rizzo et al. The Belle II Silicon Vertex Detector: Performance and Operational Experience in the First Year of Data Taking. *JPS Conf. Proc.*, 34:010003, Dec. 2021.

[141] N. Taniguchi (Belle II Collaboration). Central Drift Chamber for Belle-II. *J. Instrum.*, 2:C06014, Jun. 2017.

[142] J. Fast (Belle II TOP Group). The Belle II imaging Time-of-Propagation (iTOP) detector. *Nucl. Instrum. Meth. A*, 876:145–148, Dec. 2017.

[143] K. Inami. MCP-PMT development for Belle-II TOP counter. *Phys. Procedia*, 37:683–690, Oct. 2012.

[144] K. Kojima (Belle II TOP Group). The operation and performance of the TOP detector at the Belle II experiment. *Proc. Sci.*, EPS-HEP202:803, May 2022.

[145] S. Nishida et al. Aerogel RICH for the Belle II forward PID. *Nucl. Instrum. Meth. A*, 766:28–31, Dec. 2014.

[146] S. Korpar (Belle II ARICH Group). A 144-channel HAPD for the Aerogel RICH at Belle II. *Nucl. Instrum. Meth. A*, 766:145–147, 2014.

[147] M Yonenaga et al. Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data. *Prog. Theor. Exp. Phys.*, 2020(9), Aug. 2020. 093H01.

[148] I. Adachi et al. Detectors for extreme luminosity: Belle II. *Nucl. Instrum. Meth. A*, 907:46–59, Nov. 2018.

[149] V. Aulchenko et al. Electromagnetic calorimeter for Belle II. *J. Phys. Conf. Ser*, 587:012045, Apr. 2015.

[150] T. Aushev et al. A scintillator based endcap K_L and muon detector for the Belle II experiment. *Nucl. Instrum. Meth. A*, 789:134–142, Jul. 2015.

[151] S. Yamada et al. Data Acquisition System for the Belle II Experiment. *IEEE Trans. Nucl. Sci.*, 62(3):1175–1180, Jun. 2015.

[152] F. Forti et al. (Belle II Collaboration). Snowmass Whitepaper: The Belle II Detector Upgrade Program. In *2022 Snowmass Summer Study*, Mar. 2022.

[153] T. Kuhr et al. The Belle II Core Software. *Comput. Software Big Sci.*, 3:1, Nov. 2019.

[154] Standard C++ Foundation. The Standard : Standard C++.

[155] G. Van Rossum and F. L. Drake Jr. *Python reference manual*. 1995.

[156] R. Brun and F. Rademakers. ROOT — An object oriented data analysis framework. *Nucl. Instrum. Methods Phys. Res.*, 389(1):81–86, Apr. 1997.

[157] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, Apr. 2001.

[158] T. Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, Jun. 2015.

[159] 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.*, 130:260–325, Aug. 2000.

[160] S. Jadach, B. F. L. Ward, and Z. Was. TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons. *Comput. Phys. Commun.*, 64:275–299, May 1990.

[161] G. Balossini et al. Matching perturbative and parton shower corrections to Bhabha process at flavour factories. *Nucl. Phys. B*, 758:227–253, Dec. 2006.

[162] G. Balossini et al. Photon pair production at flavour factories with per mille accuracy. *Phys. Lett. B*, 663:209–213, May. 2008.

[163] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini. The BABAYAGA event generator. *Nucl. Phys. B Proc. Suppl.*, 131:48–55, Apr. 2004.

[164] C. M. Carloni Calame. An Improved parton shower algorithm in QED. *Phys. Lett. B*, 520:16–24, Nov. 2001.

[165] C. Carloni Calame et al. Large angle Bhabha scattering and luminosity at flavor factories. *Nucl. Phys. B*, 584:459–479, Sep. 2000.

[166] L. Barze et al. Radiative Events as a Probe of Dark Forces at GeV-Scale e^+e^- Colliders. *Eur. Phys. J. C*, 71:1680, Jun. 2011.

[167] F. A. Berends, P. H. Daverveldt, and R. Kleiss. Complete Lowest Order Calculations for Four Lepton Final States in electron-Positron Collisions. *Nucl. Phys. B*, 253:441–463, 1985.

[168] F. A. Berends, P. H. Daverveldt, and R. Kleiss. Radiative Corrections to the Process $e^+e^- \to e^+e^-\mu^+\mu^-$. *Nucl. Phys. B*, 253:421, 1985.

[169] F. A. Berends, P. H. Daverveldt, and R. Kleiss. Monte Carlo Simulation of Two Photon Processes. 1. Radiative Corrections to Multiperipheral $e^+e^-\mu^+\mu^-$ Production. *Comput. Phys. Commun.*, 40:271–284, Jan. 1986.

[170] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, Jul. 2003.

[171] T. Alexopoulos, M. Bachtis, E. Gazis, and G. Tsipolitis. Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers. *Nucl. Instrum. Methods Phys. Res. A*, 592(3):456–462, Jul. 2008.

[172] T. Bilka et al. Implementation of GENFIT2 as an experiment independent track-fitting framework, Feb. 2019.

[173] N. Braun. *Combinatorial Kalman Filter*, pages 117–174. Springer, Cham, Aug. 2019.

[174] W. Waltenberger. RAVE—A Detector-Independent Toolkit to Reconstruct Vertices. *Nucl. Instrum. Methods Phys. Res. A*, 581:549–552, Oct. 2007.

[175] J. F. Krohn et al. Global decay chain vertex fitting at Belle II. *Nucl. Instrum. Methods Phys. Res. A*, 976:164269, 2020.

[176] J. Tanaka. *Precise measurements of charm meson lifetimes and search for D^0-$\overline{D}^0$ Mixing*. PhD thesis, Tokyo U., 2001.

[177] K. Cranmer et al. (ROOT Collaboration). HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, Jan. 2012.

[178] C. R. Harris et al. Array programming with NumPy. *Nature*, 585:357–362, Sep. 2020.

[179] L. S. Blackford et al. An updated set of basic linear algebra subprograms (BLAS). *ACM Trans. Math. Softw.*, 28:135–151, Jun. 2002.

[180] J. Neyman. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philos. Trans. R. Soc. A*, 236:333–380, Aug. 1937.

[181] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Stat.*, 9:60 – 62, Mar. 1938.

[182] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, Feb. 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[183] A. L. Read. Presentation of search results: The CL(s) technique. *J. Phys. G: Nucl. Part. Phys.*, 28:2693–2704, Sep. 2002.

[184] L. Lista. Practical Statistics for Particle Physicists. In *2016 European School of High-Energy Physics*, pages 213–258, Sep. 2016.

[185] A. Wald. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Trans. Am. Math. Soc.*, 54:426–482, Nov. 1943.

[186] T. Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, page 2, Sep. 2017.

[187] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 29:1189–1232, Oct. 2001.

[188] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016.

[189] F. Abudinen et al. (Belle II Collaboration). Approved B counting analysis results with ICHEP2020 dataset. *BELLE2-NOTE-PL-2020-006*, Jul. 2020.

[190] G. Birkhoff. The Theory of Splines and Their Applications (J. H. Ahlberg, E. N. Nilson and J. L. Walsh). *SIAM Rev.*, 11:91–92, Jan. 1969.

[191] P. Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, 17:261–272, Feb. 2020.

[192] T. M. Prim et al. (Belle Collaboration). Search for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging. *Phys. Rev. D*, 101:032007, Feb. 2020.

[193] B. Spruck et al. (Belle II PXD). Belle II Pixel Detector Commissioning and Operational Experience. *PoS Proc. Sci.*, page 015, Sep. 2020.

[194] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory*, 37:145–151, Jan. 1991.

[195] J. D. Bjorken and S. J. Brodsky. Statistical Model for Electron-Positron Annihilation into Hadrons. *Phys. Rev. D*, 1:1416–1420, Mar. 1970.

[196] G. C. Fox and S. Wolfram. Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes. *Phys. Rev. Lett.*, 41:1581–1585, Dec. 1978.

[197] G. C. Fox and S. Wolfram. Event Shapes in e^+e^- Annihilation. *Nucl. Phys. B*, 157, Oct. 1979.

[198] S. H. Lee et al. (Belle Collaboration). Evidence for $B^0 \to \pi^0\pi^0$. *Phys. Rev. Lett.*, 91:261801, Dec. 2003.

[199] E. Rodrigues, H. Schreiner. scikit-hep/particle: PDG particle data and identification codes.

[200] B. Kronenbitter et al. (Belle Collaboration). Measurement of the branching fraction of $B^+ \to \tau^+\nu_\tau$ decays with the semileptonic tagging method. *Phys. Rev. D*, 92:051102, Sep. 2015.

[201] I. Adachi et al. (Belle Collaboration). Evidence for $B^- \to \tau^-\bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle. *Phys. Rev. Lett.*, 110:131801, Mar. 2013.

[202] J. P. Lees et al. (BaBar Collaboration). Evidence of $B^+ \to \tau^+\nu$ decays with hadronic B tags. *Phys. Rev. D*, 88:031102, Aug. 2013.

[203] J. L. Hodges. The significance probability of the Smirnov two-sample test. *Ark. Mat.*, 3:469–486, Jan. 1958.

[204] R. Seidl et al. (Belle Collaboration). Update of inclusive cross sections of single and pairs of identified light charged hadrons. *Phys. Rev. D*, 101:092004, May 2020.

[205] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr. Advanced event reweighting using multivariate analysis. *J. Phys. Conf. Ser.*, 368:012028, Jun. 2012.

[206] Belle II Collaboration. Measurement of the tracking efficiency and fake rate with $e^+e^- \to \tau^+\tau^-$ events. *BELLE2-NOTE-PL-2020-014*, Jul. 2020.

[207] W. Hou. Enhanced charged Higgs boson effects in $B^- \to \tau\bar{\nu}, \mu\bar{\nu}$ and $b \to \tau\bar{\nu} + X$. *Phys. Rev. D*, 48:2342–2344, Sep. 1993.

[208] A. Crivellin, C.H Greub, and A. Kokulu. Explaining $B \to D\tau\nu$, $B \to D^*\tau\nu$ and $B \to \tau\nu$ in a two Higgs doublet model of type III. *Phys. Rev. D*, 86:054014, Sep. 2012.

[209] Belle II Collaboration. Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond. *Submissions to the Snowmass 2021 Proceedings - Rare Processes and Precision Measurements*, Mar. 2022.

[210] A. Bharucha, D. M. Straub, and R. Zwicky. $B \to Vl^+l^-$ in the Standard Model from light-cone sum rules. *J. High Energ. Phys.*, 8:098, Aug. 2016.

[211] J. Podolanski and R. Armenteros. Analysis of V-events. *Philos. Mag.*, 45:13–30, Apr. 1954.

[212] D. Das, G. Hiller, and I. Nišandžić. Revisiting $B \to K^*(\to K\pi)\nu\bar{\nu}$ decays. *Phys. Rev. D*, 95:073001, Apr. 2017.

[213] S. Kullback and R. A. Leibler. On Information and Sufficiency. *Ann. Math. Stat.*, 22:79 – 86, Mar. 1951.

[214] G. Punzi. Sensitivity of searches for new signals and its optimization. *eConf*, C030908:MODT002, Aug. 2003.

[215] The Pennsylvania State University, Eberly College of Science. Applied data mining and statistical learning.

[216] Wikimedia Commons. Comparison of a histogram and a kernel density estimate, 2010.

[217] Wikimedia Commons. Comparison of 1D bandwidth selectors, 2010.

Appendix A

Jensen-Shannon distance

The Jensen-Shannon divergence (JSD) between two discrete probability distributions P and Q is a statistical distance, which measures how much the distribution P differs from the reference distribution Q. The JSD is based on the Kullback-Leibler divergence [213],

$$D_{\mathrm{KL}}(P|Q) = \sum_{x\in\chi} P(x)\ln\left(\frac{P(x)}{Q(x)}\right), \tag{A.1}$$

where χ is the sample space of the random variable x. The JSD symmetrizes the Kullback-Leibler divergence and is defined as

$$\mathrm{JSD}(P|Q) = \frac{1}{2}D_{\mathrm{KL}}(P|M) + \frac{1}{2}D_{\mathrm{KL}}(Q|M), \tag{A.2}$$

where $M = (P+Q)/2$ is the pointwise mean of P and Q. Besides being symmetric, the JSD always has a finite value and it is bounded by unity:

$$0 \leq \mathrm{JSD}(P|Q) \leq 1. \tag{A.3}$$

In the usual practice, the metric used as a distance is the square root of the JSD and it is called Jensen-Shannon distance.

Appendix B

Signal significance

In searches for physics signals, it is often useful to optimise the analysis based on the experimental sensitivity to the signal. This quantity corresponds to the expected signal significance, defined as the median of the distribution $f(Z_\mu|\mu)$, where Z_μ is the significance introduced in Eq. 3.21 and μ is the assumed value of the signal strength. In particular, the significance for a discovery is usually evaluated as the expected Z_0 (significance for $\mu = 0$) in the presence of signal ($\mu = 1$).
Figures of merit used to compute the signal significance can be derived in the context of a Poisson counting experiment. In an experiment where the observations are number of events, n, they can be modelled by a Poisson distribution with mean $\mu S + B$. Here S and B are the expected number of signal and background events. In the asymptotic limit of large counts, n can be approximated as Gaussian variable with mean $\mu S + B$ and $\sigma = \sqrt{\mu S + B}$. The p-value of the hypothesis $\mu = 0$ is the probability

$$p_0 = \Phi^{-1}\left(\frac{n - B}{\sqrt{B}}\right), \tag{B.1}$$

which corresponds to the significance

$$Z_0 = \Phi^{-1}(1 - p_0) = \frac{n - B}{\sqrt{B}}. \tag{B.2}$$

In this case, the median of n under the assumption $\mu = 1$ is equal to the mean $S + B$. Hence, the median discovery significance is

$$\text{med}\left[Z_0|\mu = 1\right] = \frac{S}{\sqrt{B}}, \tag{B.3}$$

which is a common figure of merit used in particle physics as a measure of the signal significance.
Some analyses, like those presented in this thesis, are optimised based on the accuracy of the signal measurement. The best estimator of the signal strength μ in a counting experiment can be determined using the ML method. In this case, the likelihood function is

$$\mathcal{L}(\mu) = \frac{(\mu S + B)^n}{n!} e^{-(\mu S + B)}, \tag{B.4}$$

from which it is possible to derive the estimator of the signal strength:

$$\hat{\mu} = \frac{n - B}{S}. \tag{B.5}$$

Assuming $\mu = 1$, the variance of $\hat{\mu}$ is

$$\text{var}\left[\hat{\mu}\right] = \text{var}\left[\frac{n - B}{S}\right] = \frac{1}{S^2}\text{var}\left[n\right] = \frac{S + B}{S^2}, \tag{B.6}$$

and the corresponding standard deviation is $\sigma_{\hat{\mu}} = \sqrt{S+B}/S$. Therefore the quantity

$$\frac{S}{\sqrt{S+B}} \tag{B.7}$$

can be maximised to obtain the most accurate measurement of the signal strength μ. This quantity is also the expected significance for the rejection of S under the assumption of no signal. The figure of merit $S/\sqrt{S+B}$ has been used in the analyses of this thesis to estimate the signal significance.
Other figures of merit used in particle physics to quantify the experimental sensitivity to the signal can be found in Refs. [182, 214]

Appendix C

Two-sample Kolmogorov-Smirnov test

In validation studies with data, the two-sample Kolmogorov-Smirnov test (K-S test) [203] has been used to evaluate the data-MC agreement. It is a nonparametric test that computes a distance between the empirical cumulative distributions of two samples and estimates the probability that the two samples originate from the same, but unknown, probability distribution.
The empirical cumulative distribution function $F_n(x)$ for n independent and identically distributed observations X_i is defined as

$$F_n(x) = \frac{\text{number of (elements in the sample} \leq \text{x)}}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{[-\inf,x]}(X_i), \tag{C.1}$$

where

$$1_{[-\inf,x]}(X_i) = \begin{cases} 1 & \text{for } X_i \leq x, \\ 0 & \text{for } X_i > x. \end{cases} \tag{C.2}$$

Given two independent samples of observations $X_1, ..., X_m$ and $X_1, ..., X_n$ with empirical distribution functions $F_m(x)$ and $G_n(x)$, the two sample K-S test is used to test the hypotheses

$$H_0 \,:\, F = G \qquad vs \qquad H_1 \,:\, F \neq G, \tag{C.3}$$

where F, G are the true cumulative distribution functions.
The K-S test statistic is

$$D_{n,m} = \sup_x |G_n(x) - F_m(x)|, \tag{C.4}$$

where $\sup$ is the supremum function. For large samples, the null hypothesis H_0 is rejected at a significance level α if

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n \cdot m}}. \tag{C.5}$$

The value of the parameter $c(\alpha)$ is given in Table C.1 for the most common values of α

α	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.073	1.138	1.224	1.358	1.48	1.628	1.731	1.949

Table C.1: Values of the parameter $c(\alpha)$ for the most common values of the significance level α used in hypothesis testing.

and in general it corresponds to

$$c(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1}{2}}. \tag{C.6}$$

Therefore, the condition to reject the null hypothesis can be written as

$$D_{n,m} > \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1+\frac{m}{n}}{2m}}, \tag{C.7}$$

which implies that for samples of the same size, $m = n$, the minimal bound scales in the size of the samples according to the inverse square root.

Appendix D

Singular Value Decomposition

The Singular Value Decomposition (SVD) is one of the most popular matrix-decomposition methods, and it is the fundamental tool to implement the Principal Component Analysis (PCA) technique for dimensionality reduction.
Consider the real matrix $\mathbf{X}$, of dimension $N \times p$,

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & ... & x_{1,p} \\ x_{2,1} & x_{2,2} & ... & x_{2,p} \\ ... & ... & ... & ... \\ x_{N,1} & x_{N,2} & ... & x_{N,p} \end{pmatrix} \tag{D.1}$$

where the rows represent observations in a single measurement and the columns identify different component variables of the individual measurements. Let's suppose also that the columns of $\mathbf{X}$ are centred, meaning that the estimated column mean is subtracted from each column.
The SVD of the matrix $\mathbf{X}$ is $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.

- $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N)$ is a $N \times N$ orthogonal matrix, with $\mathbf{u}_j$, $j = 1, .., N$ forming an orthonormal basis for the space spanned by the columns of $\mathbf{X}$.
- $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p)$ is a $p \times p$ orthogonal matrix, with $\mathbf{v}_j$, $j = 1, ..., p$ forming an orthonormal basis for the space spanned by the rows of $\mathbf{X}$.
- $\mathbf{D}$ is a $N \times p$ rectangular matrix with nonzero elements along the diagonal of the first $p \times p$ submatrix, $\mathrm{diag}(\mathrm{d}_1, \mathrm{d}_2, ..., \mathrm{d}_\mathrm{p})$, with $\mathrm{d}_1 \geq \mathrm{d}_2 \geq \cdots \geq \mathrm{d}_\mathrm{p} \geq 0$.

The columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and the *principal component directions* of $\mathbf{X}$. The diagonal values in $\mathbf{D}$ are the square roots of the $\mathbf{X}^T\mathbf{X}$ eigenvalues and are called the *singular values* of $\mathbf{X}$.
The covariance matrix of $\mathbf{X}$ is $\mathbf{S} = \mathbf{X}^T\mathbf{X}/N$, a square $p \times p$ matrix. The diagonalisation of $\mathbf{X}^T\mathbf{X}$ is

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}\mathbf{D}\mathbf{V}^T)^T(\mathbf{U}\mathbf{D}\mathbf{V}^T)\,. \tag{D.2}$$

This means that the diagonalisation of the matrix $\mathbf{X}^T\mathbf{X}$ is equivalent to the SVD decomposition of $\mathbf{X}$. Hence the eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{v}_j$, can be obtained by diagonalising $\mathbf{X}^T\mathbf{X}$ or by doing the SVD of $\mathbf{X}$.
The SVD and PCA allow identifying the directions that capture most of the variation in the data sample represented by $\mathbf{X}$. The k-th principal component direction of $\mathbf{X}$, $\mathbf{v}_k$, has the following properties:

- $\mathbf{v}_k$ is the eigenvector associated with the k-th largest eigenvalue, $\mathbf{d}_k^2$, of $\mathbf{X}^T\mathbf{X}$;
- $\mathbf{z}_k = \mathbf{X}\mathbf{v}_1$ has the k-th largest sample variance among all the normalised linear combinations of the columns of $\mathbf{X}$;
- $\mathbf{z}_k$ is called the k-th principal component of $\mathbf{X}$, and $\mathrm{var}(\mathbf{z}_k) = d_k^2/N$.

These features are beneficial for dimension reduction. In an extreme case, like the one illustrated in Fig. D.1, where the data lie all along one direction but two features represent the data, the dimension of the dataset can be reduced to one using the linear combination of the features given by the first principal component.

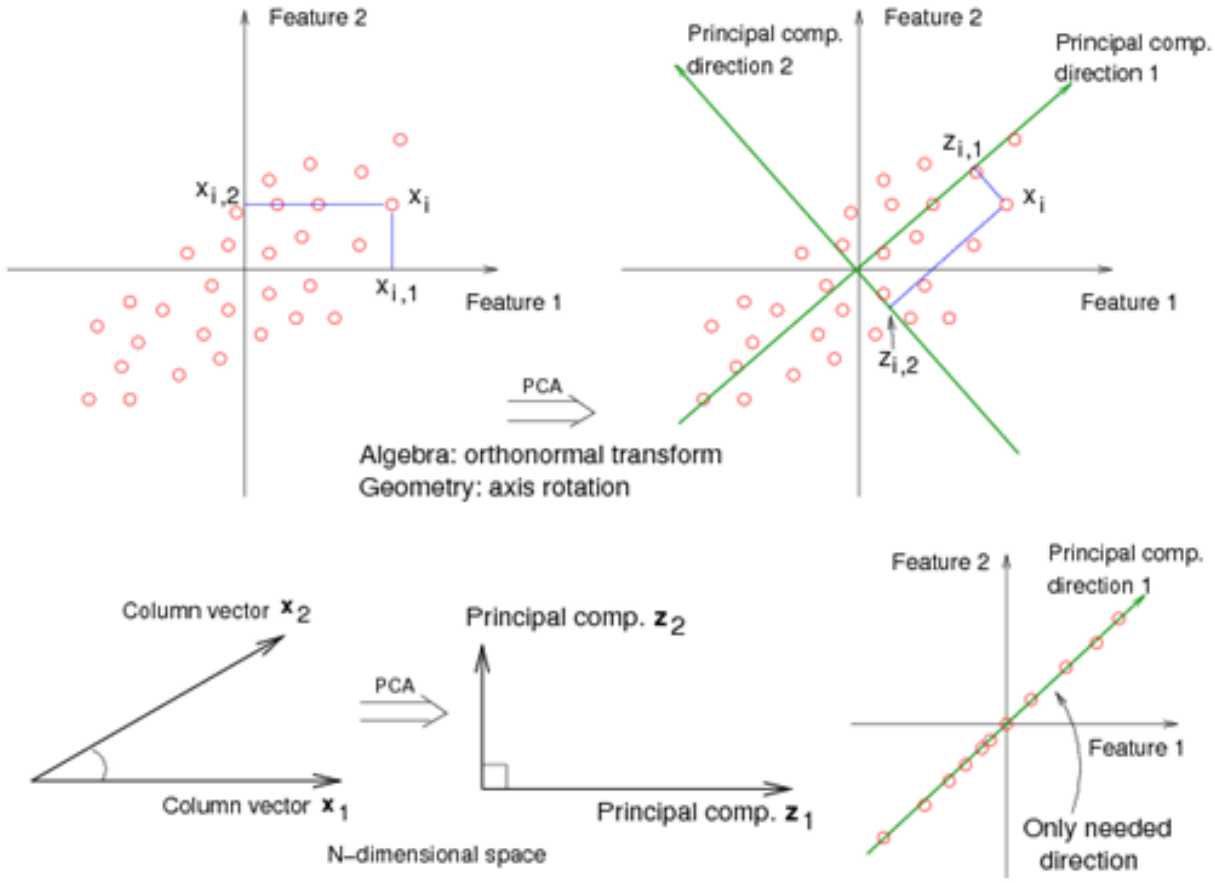


Figure D.1: Geometric illustration of the PCA method. Credit: Ref. [215].

Appendix E

Kernel density estimation

In statistics, kernel density estimation (KDE) is a non-parametric method used for data smoothing. Consider the sample of observations $(x_1, x_2, ..., x_n)$ corresponding to independent and identically distributed random variables following an unknown PDF $f(x)$. The goal of KDE is to estimate the shape of f. The kernel density estimator of f is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{E.1}$$

where K is a non-negative function called kernel, and $h > 0$ is a smoothing parameter called bandwidth. The most used kernel function is the standard normal $N(0, 1)$.
An application of KDE to a histogram of six observations is illustrated in Fig. E.1.

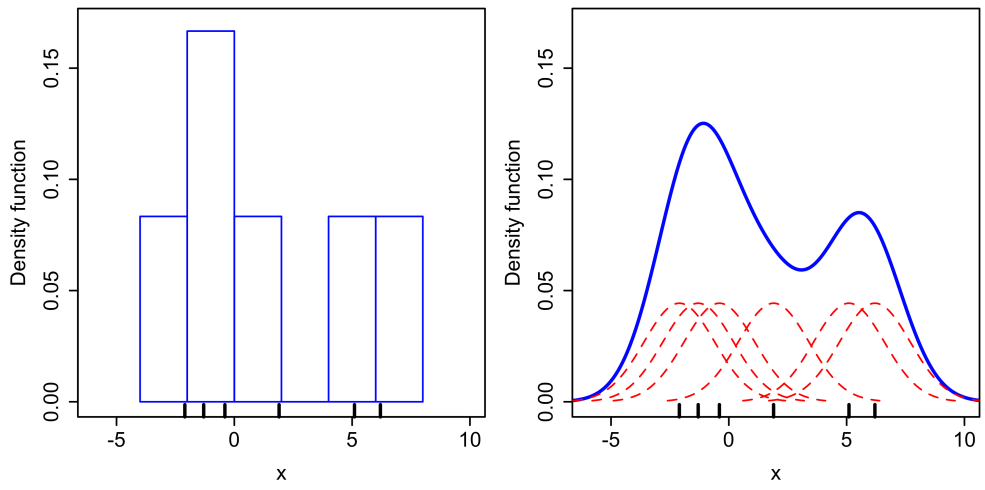


Figure E.1: Comparison of an histogram of six observations (left) and kernel density estimate (right) constructed with the same data. The six individual kernels are the red dashed curves, while the kernel density estimate corresponds to the blue curve. The observations are marked on the horizontal axis with vertical ticks. Credit: Ref. [216].

The bandwidth h of the kernel is a free parameter and it strongly influences the result of the estimate. An illustrative example is provided in Fig. E.2, where different KDEs are performed on a sample of 100 random numbers drawn from a normal PDF, $N(0, 1)$. The red curve ($h = 0.05$) is *undersmoothed* and presents many spurious features due to the small bandwidth. The green curve ($h = 2$) is *oversmoothed*, because the bandwidth is too large and it hides the underlying structure of the data. The black curve ($h = 0.337$) is optimally smoothed since its density estimate is close to the true density. One extreme case corresponds to

the limit $h \to 0$ (no smoothing), in which the estimate is a sum of n delta functions centred at the values of the observations. The other limit case occurs with $h \gg 1$ (completely smooth), in which the estimate keeps the shape of the kernel and is centred at the mean of the observations.

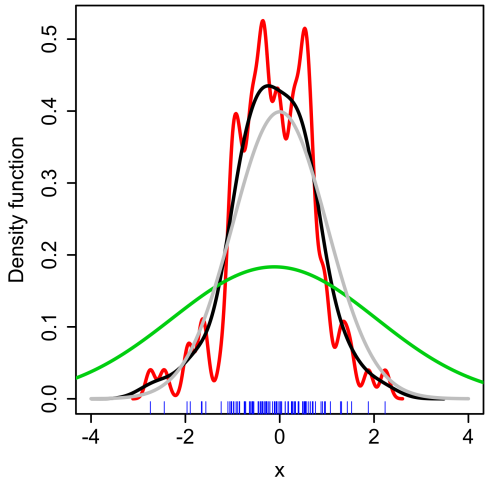


Figure E.2: KDEs with different bandwidths performed on a random sample of 100 points from a standard normal distribution. Grey curve: true PDF. Red curve: KDE with $h = 0.05$. Black curve: KDE with $h = 0.337$. Green curve: KDE with $h = 2$. Credit: Ref. [217].

Appendix F

Table of systematic uncertainties

Channel	Bin	BrB	FF	NeutG	NeutH	pid	tr	Channel	BrB	FF	NeutG	NeutH	pid	tr
signal	0	0.0	1.9	0.4	2.2	0.3	2.1	ssbar	0.0	0.0	0.9	2.6	0.4	1.0
	1	0.0	4.9	0.2	1.0	0.4	1.6		0.0	0.0	1.2	2.4	0.5	2.5
	2	0.0	7.1	0.7	0.7	0.7	1.8		0.0	0.0	0.6	0.1	1.1	2.4
	3	0.0	2.3	0.1	2.5	0.3	1.9		0.0	0.0	0.5	0.9	0.5	1.2
	4	0.0	5.0	0.1	1.5	0.4	1.9		0.0	0.0	2.2	2.0	0.5	3.1
	5	0.0	7.2	0.5	0.4	0.7	1.9		0.0	0.0	0.4	2.6	1.7	2.0
	6	0.0	3.1	0.2	2.8	0.3	2.8		0.0	0.0	2.4	0.9	0.6	6.3
	7	0.0	5.2	0.2	1.8	0.4	2.2		0.0	0.0	1.0	1.3	0.5	4.9
	8	0.0	7.2	0.4	1.9	0.7	2.1		0.0	0.0	0.5	1.6	0.7	2.4
	9	0.0	4.3	0.2	3.7	0.4	3.3		0.0	0.0	3.7	9.8	0.0	10.0
	10	0.0	5.7	0.1	3.1	0.4	2.9		0.0	0.0	1.2	3.5	0.6	14.8
	11	0.0	7.3	0.0	2.9	0.7	2.8		0.0	0.0	3.0	4.4	1.1	2.6
charged	0	1.3	0.0	0.6	3.3	0.9	0.2	uubar	0.0	0.0	1.4	2.1	2.4	0.8
	1	1.9	0.0	1.1	2.9	1.5	0.5		0.0	0.0	2.5	0.8	1.0	3.5
	2	2.1	0.0	1.2	1.9	2.1	2.4		0.0	0.0	2.1	3.2	1.2	0.2
	3	1.5	0.0	2.4	4.7	0.7	1.2		0.0	0.0	0.6	10.4	1.7	5.0
	4	2.0	0.0	3.0	6.0	4.2	1.9		0.0	0.0	1.4	5.4	1.0	3.0
	5	1.9	0.0	2.2	1.2	2.2	1.7		0.0	0.0	0.8	8.8	1.3	2.4
	6	2.1	0.0	0.9	2.3	0.6	5.5		0.0	0.0	2.0	5.6	3.5	2.5
	7	3.3	0.0	0.8	0.4	1.4	6.1		0.0	0.0	0.9	8.1	4.9	3.0
	8	3.4	0.0	0.6	0.6	1.5	3.4		0.0	0.0	0.5	8.8	3.2	2.9
	9	3.0	0.0	0.0	1.5	0.9	9.2		0.0	0.0	1.4	13.1	2.9	1.3
	10	3.7	0.0	0.1	2.4	4.1	8.8		0.0	0.0	0.3	5.3	0.0	9.0
	11	5.6	0.0	0.4	1.6	1.4	4.7		0.0	0.0	2.4	0.8	1.3	8.7
mixed	0	0.9	0.0	0.1	1.5	1.4	0.7	ddbar	0.0	0.0	1.1	6.3	1.1	2.1
	1	2.1	0.0	0.0	2.6	1.6	3.5		0.0	0.0	3.8	13.7	7.1	0.1
	2	2.0	0.0	1.5	0.2	10.8	0.5		0.0	0.0	1.9	11.9	2.4	2.0
	3	1.3	0.0	0.8	3.8	1.0	0.6		0.0	0.0	2.9	11.8	5.6	5.4
	4	2.9	0.0	0.6	2.9	1.9	3.4		0.0	0.0	6.0	8.1	0.4	7.1
	5	3.8	0.0	0.7	8.6	2.0	0.6		0.0	0.0	5.3	16.1	5.5	0.9
	6	1.8	0.0	0.6	0.0	1.8	0.9		0.0	0.0	3.5	24.1	1.2	2.6
	7	2.4	0.0	0.2	1.6	6.4	0.6		0.0	0.0	3.1	1.7	2.6	3.5
	8	2.2	0.0	0.3	0.9	3.7	1.2		0.0	0.0	1.6	21.3	3.3	0.1
	9	2.3	0.0	0.1	0.1	6.4	4.7		0.0	0.0	12.5	9.0	0.0	0.6
	10	2.9	0.0	0.6	2.5	3.6	1.2		0.0	0.0	6.9	15.8	0.0	0.8
	11	3.0	0.0	1.1	0.4	1.0	3.7		0.0	0.0	0.7	9.0	0.0	3.3
ccbar	0	0.0	0.0	0.2	5.4	0.5	2.1	taupair	0.0	0.0	0.0	0.0	2.0	0.0
	1	0.0	0.0	0.2	2.2	0.7	1.1		0.0	0.0	0.0	0.0	3.7	0.0
	2	0.0	0.0	0.2	2.1	0.7	1.0		0.0	0.0	0.0	0.0	15.7	0.0
	3	0.0	0.0	0.5	4.3	0.5	1.4		0.0	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.0	0.4	1.0	0.6	0.3		0.0	0.0	0.0	0.0	10.4	0.0
	5	0.0	0.0	0.3	1.3	0.8	3.5		0.0	0.0	0.0	0.0	0.0	0.0
	6	0.0	0.0	0.5	4.6	0.4	0.4		0.0	0.0	0.0	0.0	0.0	0.0
	7	0.0	0.0	0.1	0.3	0.5	2.9		0.0	0.0	0.0	0.0	0.8	0.0
	8	0.0	0.0	2.1	6.5	0.9	4.7		0.0	0.0	0.0	0.0	0.0	0.0
	9	0.0	0.0	1.0	4.4	0.9	3.0		0.0	0.0	0.0	0.0	0.0	0.0
	10	0.0	0.0	2.6	1.1	0.6	6.9		0.0	0.0	0.0	0.0	0.0	0.0
	11	0.0	0.0	0.6	5.8	0.7	6.9		0.0	0.0	0.0	0.0	0.0	0.0

Table F.1: Relative systematic uncertainties per bin in the signal and background samples of the statistical model constructed for the $B^+ \to K^+\nu\bar{\nu}$ analysis. BrB: branching-fraction systematics. FF: form-factor systematic. pid: PID systematic tr: tracking systematic. NeutG: systematic due to the energy calibration of photon clusters. NeutH: systematic due to the energy calibration of ECL clusters not matched to photons.

Appendix G

Post-fit shifts

Nuisance parameters	Samples	Channels	Shift $= (\mathrm{par}_{fit} - \mathrm{par}_{nom})/\sigma_{nom}$
7 normalisation-uncertainty modifiers	charged B	SR,CR1,CR2,CR3	0.0
	neutral B	SR,CR, CR2,CR3	-0.05
	$c\bar{c}$	SR,CR1,CR2,CR3	0.86
	$d\bar{d}$	SR,CR1,CR2,CR3	0.07
	$s\bar{s}$	SR,CR1,CR2,CR3	0.76
	$u\bar{u}$	SR,CR1,CR2,CR3	0.04
	$\tau^+\tau^-$	SR,CR1,CR2,CR3	0.0
3 correlated-shape modifiers for BR systematic	neutral B, charged B	SR,CR1,CR2,CR3	-0.02
	neutral B, charged B	SR,CR1,CR2,CR3	0.02
	neutral B, charged B	SR,CR1,CR2,CR3	0.01
3 correlated shape modifiers for FF systematic	signal	SR,CR1	0.0
	signal	SR,CR1	0.0
	signal	SR,CR1	0.0
3 correlated shape modifiers for PID correction systematic	all	SR,CR1,CR2,CR3	0.06
	all	SR,CR1,CR2,CR3	0.03
	all	SR,CR1,CR2,CR3	-0.03
1 correlated shape modifier for photon energy systematic	all	SR,CR1,CR2,CR3	0.17
1 correlated shape modifier for hadron energy systematic	all	SR,CR1,CR2,CR3	0.68
1 correlated shape modifier for tracking efficiency systematic	all	SR,CR1,CR2,CR3	0.13
144 modifiers for MC statistical uncertainties (1 modifier per bin)	charged B	CR1	-0.14, 0.15, -0.1
	charged B	SR	0.04, 0.0, 0.06, 0.16, -0.11, -0.14, -0.04, -0.12, 0.28
	neutral B	CR1	-0.1, 0.13, -0.09
	neutral B	SR	0.03, -0.0, 0.04, 0.11, -0.11, -0.09, -0.02, -0.07, 0.14
	$c\bar{c}$	CR1	-0.22, 0.39, -0.48
	$c\bar{c}$	CR2	0.34, -0.09, 0.32, -0.18, 0.2, -0.04, 0.0, -0.71, 0.24
	$c\bar{c}$	CR3	0.06, 0.2, -0.08
	$c\bar{c}$	SR	0.06, -0.0, 0.22, 0.15, -0.18, -0.37, -0.02, -0.1, 0.24
	$d\bar{d}$	CR1	-0.03, 0.07, -0.08
	$d\bar{d}$	CR2	0.06, -0.01, 0.05, -0.04, 0.04, -0.0, 0.0, -0.0, 0.0
	$d\bar{d}$	CR3	0.01, 0.03, -0.02
	$d\bar{d}$	SR	0.01, -0.0, 0.06, 0.04, -0.04, -0.06, 0.0, 0.0, 0.0
	$s\bar{s}$	CR1	-0.13, 0.24, -0.37
	$s\bar{s}$	CR2	0.18, -0.08, 0.26, -0.12, 0.14, -0.02, 0.51, -0.14, 0.43
	$s\bar{s}$	CR3	0.03, 0.1, -0.06
	$s\bar{s}$	SR	0.04, -0.0, 0.18, 0.08, -0.12, -0.26, -0.01, -0.13, 0.23
	$u\bar{u}$	CR1	-0.1, 0.12, -0.19
	$u\bar{u}$	CR2	0.09, -0.03, 0.13, -0.1, 0.07, -0.02, 0.0, -0.0, 0.27
	$u\bar{u}$	CR3	0.02, 0.06, -0.02
	$u\bar{u}$	SR	0.02, -0.0, 0.09, 0.06, -0.09, -0.16, -0.01, -0.02, 0.12
	$\tau^+\tau^-$	CR1	-0.02, 0.03, -0.06
	$\tau^+\tau^-$	CR2	0.0, -0.01, 0.07, -0.0, 0.03, 0.0, 0.0, -0.0, 0.0
	$\tau^+\tau^-$	CR3	0.01, 0.01, 0.0
	$\tau^+\tau^-$	SR	0.0, -0.0, 0.02, 0.01, -0.01, -0.0, 0.0, -0.01, 0.0

Table G.1: Post-fit shifts, $(\mathrm{par}_{fit} - \mathrm{par}_{nom})/\sigma_{nom}$, of the nuisance parameters in the statistical model constructed for the $B^+ \to K^+\nu\bar{\nu}$ analysis. A detailed characterisation of the nuisance parameters is provided in Sec. 9.1.2.

Acknowledgements

This thesis is dedicated to my grandmas, they are always in my thoughts.
I am a lucky guy. Whenever I'm struggling, I know my mom, my dad, and my sister are there for me. *Vi voglio bene*. The same goes for the rest of my family. We may be far from each other, but we're always close.
I want to thank the people in the Belle II group at DESY, it's been a pleasure to spend my days with you. Especially, I would like to write a few words on two people that I really admire. One of them is my supervisor, Sasha Glazov. It's been a honour for me to work day by day with him. Sasha is a brilliant mind. You never stop learning from him. He's been a great supervisor. Most important, he is a good man. The other person is Carsten Niebuhr. Carsten is exceptional. Tireless, just. He's really one of a kind.
Let me also express my gratitude to the colleagues and friends that read this thesis. A special thank goes to Simon Kurz, Alberto Martini, and Eldar Ganiev: *grazie mille!* They did a lot. I also want thank Sally Stefkova, Michel Hernández Villanueva, and Navid Rad. Thanks to Elisa Manoni, who gave precious feedback and suggestion (it's been a pleasure to have you as EWP convener in these years!). A special thanks also to Ale Gaz, another person that I admire. He's a friend. He's always ready to help. I appreciate it a lot.
Thanks to my office and PhD mate Cyrille Praz. He's a great person and I've learned a lot from him. Thanks to Munira Khan and Tommy Martinov, who were the first to read the pages of this thesis, and to Maria Konstantinova, who helped me with the *Zusammenfassung*.
It's always been nice to work at DESY in the Belle II group, I've made some good friends: Henrikas Svidras, Sascha Dreyer, Abtin Narimani, Anselm Baur, Varghese Babu.
Physics is something special in my life, I cannot imagine myself doing something else. I know it's hard to make it, but I'll do my best. I'm a real *testa dura* from Calabria, I don't give up.

Declaration of Authorship

I, the undersigned, hereby declare that I have not submitted a dissertation with the same research topic to an academic higher education institution that had been already accepted or evaluated as insufficient in an earlier doctoral procedure.
Ich versichere, dass ich keine Dissertation mit dem gleichen Forschungsthema schon einmal in einem fruheren Promotionsverfahren an einer wissenschaftlichen Hochschule eingereicht habe, die angenommen oder als ungenugend beurteilt worden ist.

I, the undersigned, declare upon oath that I have written my present thesis independently and that I have not used anything other than the specified aids - in particular any sources not mentioned in the bibliography.
Ich versichere hiermit ehrenwörtlich, dass ich meine vorliegende Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbe- sondere keine im Quellenverzeichnis nicht benannten Quellen – benutzt habe.

I, the undersigned, declare upon oath that the dissertation submitted in electronic form and the printed bound copy of the dissertation submitted for archiving are identical.
Ich versichere an Eides statt, dass die in elektronischer Form eingereichte Dissertation und das zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertation identisch sind.

Filippo Dattola
Hamburg, 10 October 2022