Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE PHYSIQUE
Institut Pluridisciplinaire Hubert Curien (IPHC), UMR 7178

THÈSE

présentée par:

Lucas Martel

soutenue le : **20 Septembre 2023**

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

Search for the $B^+ \to K^+\nu\bar{\nu}$ decay in the Belle II experiment

THÈSE dirigée par:

Dr. Isabelle Ripp-Baudot	Institut Pluridisciplinaire Hubert Curien

---

RAPPORTEURS:

Dr. Vincent TISSERAND	Laboratoire de Physique de Clermont
Dr. Diego GUADAGNOLI	Laboratoire d'Annecy-le-Vieux de Physique Théorique

EXAMINATEURS:

Dr. Anne-Catherine LE-BIHAN	Institut Pluridisciplinaire Hubert Curien
Dr. Elisa MANONI	INFN Sezione di Perugia
Dr. Giulio DUJANY	Insitut Pluridisciplinaire Hubert Curien

Search for the $B^+ \to K^+\nu\bar{\nu}$ decay in the Belle II experiment

Abstract: This thesis describes the first search for the decay of a charged B-meson into a charged kaon and a pair of neutrinos using a hadronic tagging method at the Belle II experiment, operating at the asymmetric electron-positron collider SuperKEKB located at KEK, Tsukuba, Japan. The $B^+ \to K^+\nu\bar{\nu}$ decay operates, at the quark level, through a $b \to s\nu\bar{\nu}$ flavour changing neutral-current transition. This decay has never been observed due to the experimental challenge posed by the undetected pair of neutrinos in its final state. However its branching fraction is predicted with accuracy in the Standard Model of particle physics, thus, a precise measurement of this branching fraction offers a unique opportunity to probe beyond Standard Model contributions.

The analysis described therein makes use of the Full Event Interpretation algorithm (FEI), developed by the Belle II collaboration to sequentially reconstruct the most probable decay of the B_{tag} meson accompanying the signal meson B_{sig} in $\Upsilon(4S) \to B_{sig}B_{tag}$ events. The analysis exploits a data sample corresponding to an integrated luminosity of 362 fb^{-1} collected at the $\Upsilon(4S)$ resonance mass, completed by a sample of 42 fb^{-1} collected 60 MeV below said resonance.

Given this dataset, the expected upper limit on the branching fraction of $B^+ \to K^+\nu\bar{\nu}$ is determined to be 2.3×10^{-5} at 90% confidence level, using simulated events and data collected in specific control channels. This measurement is expected to be competitive with previous measurements performed by the BaBar and Belle experiments with on-resonance datasets of 421 fb^{-1} and 711 fb^{-1} respectively.

Furthermore, the development of an algorithmic method to improve the Belle II Silicon Vertex Detector (SVD) resolution on position is presented. This method corrects charge sharing effects between silicon strips in the detector, allowing to improve the spatial resolution for specific sensors by 5 to 15%.

Recherche de la désintégration $B^+ \to K^+\nu\bar{\nu}$
au sein de l'expérience Belle II

Résumé: Cette thèse décrit la première recherche de la désintégration d'un méson B en un kaon chargé et une paire de neutrinos en utilisant une méthode de reconstruction hadronique du B compagnon au sein de l'expérience Belle II, auprès du collisionneur électron-positon asymétrique SuperKEKB situé à KEK, Tsukuba au Japon. La désintégration $B^+ \to K^+\nu\bar{\nu}$ opère, au niveau des quarks, à travers une transition de courant neutre à changement de saveur $b \to s\nu\bar{\nu}$.
Cette désintégration n'a jamais été observée en raison du défi expérimental posé par la paire de neutrinos non détectée dans son état final. Cependant son rapport d'embranchement est prédit avec précision dans le modèle standard de la physique des particules, la mesure de ce rapport d'embranchement offre donc une opportunité unique de sonder les limites du Modèle Standard. L'analyse décrite ici tire partie de l'algorithme de *Full Event Interpretation* (FEI), développé par la collaboration Belle II pour reconstruire séquentiellement la désintégration la plus probable du méson B_{tag} accompagnant le méson signal B_{sig} dans les évènements de type $\Upsilon(4S) \to B_{sig}B_{tag}$. L'analyse exploite un échantillon de données correspondant à une luminosité de 362 fb^{-1} collectée à l'énergie de la résonance $\Upsilon(4S)$, complétée par un échantillon de 42 fb^{-1} collecté 60 MeV en dessous de ladite résonance.
Compte tenu de cet ensemble de données, la limite supérieure attendue du rapport d'embranchement de $B^+ \to K^+\nu\bar{\nu}$ est déterminé comme étant 2.3×10^{-5} à un niveau de confiance de 90 %, en utilisant des échantillons d'évènements simulés ainsi que des données collectées pour des canaux de contrôle spécifiques. Cette mesure attendue est compétitive avec les mesures précédentes effectuées par les expériences BaBar Belle avec des ensembles de données de 421 fb^{-1} et 711 fb^{-1} respectivement.
Par ailleurs, le développement d'une méthode algorithmique pour améliorer la résolution spatiale du détecteur de vertex à pistes de silicium (SVD) de Belle II est présentée. Cette méthode corrige les effets de partage de charge entre les pistes de silicium dans le détecteur, permettant d'améliorer la résolution spatiale des modules de détection de 5 à 15 %.

Contents

Introduction

In the second half of the XX^{th} century, a succession of theoretical works [1–7] trying to make sense of numerous experimental observations [8, 9] ultimately resulted in what is now the Standard Model (SM) of particle physics. The SM is a theoretical framework used to describe elementary particles and their interactions and has been extensively tested since its inception. It proved to be extremely accurate as well as capable to predict experimental results [10–13] culminating in the discovery of the Higgs boson by the ATLAS and CMS experiments at CERN in 2012 [14, 15].

However, despite this success, the SM fails to be a theory of *eveything*. While describing 3 of the 4 fundamental interactions observed in the Universe, it does not include a description of gravity and fails to explain the observed asymmetry between matter and antimatter, as well as the origin and constituents of Dark Matter (DM), an unknown type of matter which existence can be infered by their gravitational effect in astronomical observations [16–19]. In addition, recent experimental results seem to exhibit tensions with expected SM values [20–22].

Amongst the physical processes showing tensions with the standard model, the Flavour Changing Neutral Current (FCNC) $b \to s$ quark transitions are of particular interest. Indeed, several models proposed to expand the SM expect modifications of these processes from New Physics (NP), which are new particles or interactions not described by the SM.

Observations in $b \to s\ell^+\ell^-$ seemed to hint towards significant deviations with the SM and have (at least partially) motivated and justified numerous studies of B-mesons decays. However, recent re-evaluations of these $b \to s\ell^+\ell^-$ observations [23] have seen said deviation vanish.

However, of these $b \to s$ transitions, the case of $B \to K^{(*)}\nu\bar{\nu}$ decays is particularly interesting. These decays of B-mesons into a $K^{(*)}$ meson, a neutrino and an anti-neutrino happen through $b \to s\nu\bar{\nu}$ quark transitions and have never been observed to this day. This is due to the fact that neutrinos are not directly detected in collider experiments as well as to the low probability of such $b \to s\nu\bar{\nu}$ transitions to happen via SM processes. These decay could still be sensitive to NP effects while being compatibles with the recent $B \to s\ell^+\ell^-$ observations.

Because the probability of $B \to K^{(*)}$ decays is precisely known in the SM, a precise measurement of these processes would allow to identify possible NP contributions and constraint most NP models, advancing towards the goal of a complete SM.

In this thesis, we develop a full analysis aimed at the first measurement of the $B^+ \to K^+\nu\bar{\nu}$ decay, using the strengths of the Belle II detector [24] at the SuperKEKB accelerator [25] which specifically aims at studying such processes. This document is split into chapters, expanding on the motvations to measure $b \to s$ processes, the analysis devised to perform such a measurement as well as additional work performed as part of the Belle II collaboration. These chapters are organised as follows:

- Chapter 1 introduces the main concepts of the Standard Model, and how the measurement of the $B^+ \to K^+\nu\bar{\nu}$ decay can help to better understand it and constrain contributions from processes beyond the Standard Model.
- Chapter 2 describes the experimental apparatus used in this work, namely the Belle II detector and the SuperKEKB collider.
- Chapter 3 presents work performed in order to improve the performances of the Silicon Vertex Detector of the Belle II experiment.
- Chapter 4 describes the analysis techniques and tools used in the search for the $B^+ \to K^+\nu\bar{\nu}$ decay.
- Chapter 5 presents the strategy aiming at the analysis of data recorded by the Belle II experiment to measure the branching ratio of the $B^+ \to K^+\nu\bar{\nu}$ decay.

Chapter 1

Theoretical motivation

Contents

As stated in the introduction, the SM succesfully explains most of the current experimental observations and has allowed to predict numerous discoveries [10–15]. However it falls short in a theoretical point of view, as it fails to incorporate gravity and neutrino masses as well as providing an explanation for the matter/antimatter asymmetry in the unvierse. Some recent observations seem to diverge from SM predictions in the $b \to sl^+l^-$ [22] and $b \to c\tau\nu$ [26–29] transitions. In this chapter we will briefly introduce the SM (Section 1.1) as well as an effective formalism (Section 1.2) which allows to describe the SM as an approximation of a broader theory valid at a specific energy scale. This allows to study $B \to K^{(*)}\nu\bar{\nu}$ decays in the SM (Section 1.3) as well as describe several NP scenarios which could affect these decays (Section 1.4). Finally, Section 1.5 will present previous experimental results.

1.1 The Standard Model of particle physics

The SM is a theory describing how half-odd spin *fermions* interact with each other through the exchange of integer spin *gauge bosons* that mediate the three fundamental strong, weak and electromagnetic interactions. The 12 fermions (and their 12 corresponding anti-particles) form multiplets of the $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$ group components of the local gauge symmetry of the SM:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

Where $SU(3)_C$ corresponds to quantum chromodynamics (QCD) describing the strong interaction and $SU(2)_L \otimes U(1)_Y$ to the electroweak interaction. All objects

transforming under $SU(3)_c$ carry a colour charge C which can take one of three colour values (red (r), green (g), blue(b)) and/or one of three anti-colour values (anti-red $(\bar{r})$, anti-green $(\bar{g})$, anti-blue $(\bar{b})$). The gluons $g_i, i \in [1,8]$ are the gauge bosons mediating the strong interaction, coupling to colour charge while carrying a colour/anti-colour mixture. The gauge bosons W^i_μ, $i = 1,2,3$ and B_μ are associated to the $SU(2)_L$ and $U(1)_Y$ factors respectively, coupling to the generator of the associated group (weak isospin T for $SU(2)_L$ and weak hypercharge Y for $U(1)_Y$) with coupling constants g and g'.
Through the Higgs mechanism, part of the electroweak gauge symmetry breaks, giving rise to 4 physical boson fields:

$$W^\pm = \frac{(W^1_\mu \mp iW^2_\mu)}{\sqrt{(2)}}, \quad Z = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W, \quad A = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W. \tag{1.2}$$

Where $\theta_W = \tan^{-1}(g/g')$ is the weak angle, $W^\pm$ are the charged weak boson field, Z the neutral weak boson field and A is the photon (γ) field. The photon couples to the electric charge $Q = T_3 + \frac{1}{2}Y$. This symmetry breaking also gives rise to a neutral scalar boson field: the Higgs boson H.
Fermions can then be divided into two classes depending on their behaviour under $SU(3)_c$:

- **leptons** form a $SU(3)_c$ singlet, meaning they do not interact strongly. There are 3 charged leptons (e^-, μ^-, τ^-), 3 neutral leptons called neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$ and 6 corresponding anti-leptons $(e^+, \mu^+, \tau^+, \bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau)$. Neutrinos only couple to the weak interaction while charged leptons also couple to the electromagnetic interaction.

- **quarks** are fermions that transform under $SU(3)_c$, they couple to the three fundamental interactions of the standard model. There are 6 quarks (u, d, c, s, b, t) and 6 anti-quarks $(\bar{u}, \bar{d}, \bar{c}, \bar{s}, \bar{b}, \bar{t})$. Because of the long distance behaviour of QCD, free quarks cannot be observed and spontaneously bind into hadrons (with the exception of the top quark, which spontaneously decays without forming hadrons). The two most common types of hadrons are **mesons** which are formed by a quark and an anti-quark, and **baryons** which are formed by three quarks.

Both leptons and quarks are organised in 3 generations each:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

These come from the way fermions transform under $SU(2)_L$: in order to accurately describe the weak interaction, fermions are arranged in weak isospin doublets (*left-handed fermions*, L) and singlets (*right-handed fermions*, R) which are neutral under the weak interaction. These multiplets are:

$$\boldsymbol{\mathcal{F}_L} = \left\{ \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \begin{pmatrix} u' \\ d' \end{pmatrix}_L \begin{pmatrix} c' \\ s' \end{pmatrix}_L \begin{pmatrix} t' \\ b' \end{pmatrix}_L \right\}$$

$$\mathcal{F}_{\boldsymbol{R}} = \{e_R, \mu_R, \tau_R, u'_R, d'_R, c'_R, s'_R, b'_R, t'_R\}$$

Because the quark electroweak eigenstates are not the same as the mass eigenstates, they are labeled here with primed symbols. In addition, right handed neutrinos are not mentionned because they are neutral to all the interactions of the SM and so are not SM particles. Table 1.1 lists the particles discussed here as well as their properties.

The Higgs mechanism introduces Yukawa couplings between the Higgs boson and fermion doublets, producing fermion mass terms. The quark weak eigenstates can be linked to the mass eigenstates by the Cabibo-Kobayashi-Maskawa (CKM) unitary matrix:

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

The coefficients of V_{CKM} are linked to the transitions between different flavours of quarks. The transition from a quark flavour i to a quark flavour j being proportional to $|V_{ij}|^2$. Being unitary, V_{CKM} needs to verify:

$$\sum_{i\in\{u,d,s\}}^{n} V_{ij}V_{ik}^{*} = \delta_{jk}, \sum_{j\in\{u,d,s\}}^{n} V_{ij}V_{kj}^{*} = \delta_{ik}. \tag{1.3}$$

Where $\delta_{ij} = 1$ if $i = j$ and 0 for off-diagonal terms. This is the basis of the Glashow-Iopoulos-Maiani (GIM) mechanism, which forbids transitions between quark flavours of same electric charge, called flavour-changing neutral currents (FCNC), at tree level in the SM and suppresses them at higher order. On the other hand, leptons weak eigenstates are also mass eigenstates due to the fact that right-handed neutrinos do not exist in the SM, meaning that no transition between lepton flavours can occur in the SM.
In addition, the 6 off-diagonal relations of Equation 1.3 can each be interpreted as triangles in the complex plane. Amongst them, the relation:

$$V_{ud}V_{ub}^{*} + V_{cd}V_{cb}^{*} + V_{td}V_{tb}^{*} = 0, \tag{1.4}$$

is conventionally chosen to draw what is referred to as a unitarity triangle. This sum is furthermore reordered as:

$$1 + \frac{V_{td}V_{tb}^{*}}{V_{cd}V_{cb}^{*}} + \frac{V_{ud}V_{ub}^{*}}{V_{cd}V_{cb}^{*}} = 0, \tag{1.5}$$

in order to place the vertices of the unitarity triangle of Figure 1.1 at $(0,0)$, $(1,0)$ and $(\bar{\rho},\bar{\eta})$ in the complex plane, with $\bar{\rho} + i\bar{\eta} = -V_{ud}V_{ub}^{*}/V_{cd}V_{cb}^{*}$. The lengths of the triangle sides can then be expressed with the CKM matrix elements as:

$$\overline{AB} = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right|, \tag{1.6}$$

$$\overline{AC} = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right|, \tag{1.7}$$

$$\overline{CB} = 1, \tag{1.8}$$

as well as the three angles :

$$\alpha = \arg\left(\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \tag{1.9}$$

$$\beta = \arg\left(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right) \tag{1.10}$$

$$\gamma = \arg\left(\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.11}$$

The measurement of the triangle parameters is a long standing goal of particle physics, as their values allow to constrain the 4 free parameters of the SM related to the CKM matrix.

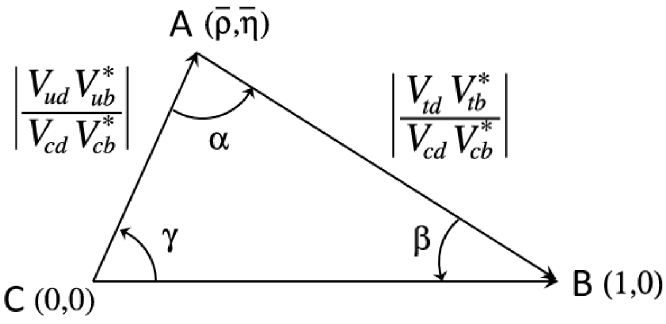


Figure 1.1: Representation of Equation 1.5 in the complex plane.

Finally, Table 1.1 lists the different particles of the SM, as well as their physical properties. In addition, several composite particles (mesons and baryons) relevant to this work are listed as well.

Particle	**Symbol**	**quark content**	**Spin$^{\text{parity}}$**	**Mass [eV/c^2]**	**Electric Charge**
		Gauge Bosons			
Photon	γ	-	1	0	0
W boson	$W^{\pm}$	-	1	$8 \times 10^{1}0$	± 1
Z boson	Z^0	-	1	$9.1 \times 10^{1}0$	0
Gluon	g	-	1	0	0
		Higgs Boson			
Higgs boson	H	-	0	$1.3 \times 10^{1}1$	0
		Leptons			
Electron	e^-	-	1/2	5.1×10^5	-1
Muon	μ^-	-	1/2	1.1×10^8	-1
Tau	τ^-	-	1/2	1.8×10^9	-1
Neutrino	ν	-	1/2	< 1.1	0
		Quarks			
Up quark		u	1/2	2.2×10^6	$\frac{2}{3}$
Down quark		d	1/2	4.7×10^6	$-\frac{1}{3}$
Charm quark		c	1/2	1.3×10^9	$\frac{2}{3}$
Strange quark		s	1/2	9.3×10^7	$-\frac{1}{3}$
Top quark		t	1/2	$1.7 \times 10^{1}1$	$\frac{2}{3}$
Bottom quark		b	1/2	4.2×10^9	$-\frac{1}{3}$
		Mesons			
Charged pi meson (pion)	π^+	$u\bar{d}$	0^-	1.4×10^8	$+1$
neutral pi meson	π^0	$\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$	0^-	1.3×10^8	0
Charged K meson (kaon)	K^+	$u\bar{s}$	0^-	4.9×10^8	$+1$
neutral K meson	K^0	$\bar{d}s$	0^-	5.0×10^8	0
short-lived K meson	K^+	$\frac{d\bar{s}-s\bar{d}}{\sqrt{2}}$	0^-	5.0×10^8	0
short-lived K meson	K^+	$\frac{d\bar{s}+s\bar{d}}{\sqrt{2}}$	0^-	5.0×10^8	0
Charged D meson	D^+	$c\bar{d}$	0^-	1.9×10^9	$+1$
Neutral D meson	D^0	$c\bar{u}$	0^-	1.9×10^9	0
J/Psi meson	$J/\Psi(1S)$	$c\bar{c}$	1^-	3.1×10^9	0
Charged B meson	B^+	$u\bar{b}$	0^-	5.3×10^9	$+1$
Neutral B meson	B^0	$d\bar{b}$	0^-	5.3×10^9	0
Upsilon meson	$\Upsilon(4S)$	$b\bar{b}$	1^-	$1.1 \times 10^{1}0$	0
		Baryons			
proton	p	uud	$1/2^+$	0.9×10^9	$+1$
neutron	n	udd	$1/2^+$	0.9×10^9	0

Table 1.1: Properties of the elementary particles of the SM

1.2 Effective Field Theory formalism

As stated in the introduction, the SM does remarkably well to describe most processes involving elementary particle and has even proven succesful at predicting several experimental observations. However, we know it to be incomplete. The first phase of the LHC [14, 15] showed that the Higgs boson seems to be SM-like and "light", and that there is a mass gap above the current SM spectrum. Indeed, were there particles in the range $[m_t, \text{TeV}]$ they should have been observed at the LHC.

The limits of the SM described in Section 1.1 lend to believe that the SM is in fact an *effective field theory* (EFT), low-energy limit of a broader theory valid at a higher scale Λ. In that case, working at an energy $E << \Lambda$ does not require to precisely know of the physics at the Λ scale but only to describe it with a set of effective parameters (whose number depends on the wanted accuracy). This in turn allows to work out physics at different energy scales, which is needed in the case of B meson decays where different scales are involved: the b quark mass $m_b \simeq 4$ GeV, the W boson mass $M_W \simeq 80$ GeV corresponding to the scale of electroweak processes and $\Lambda_{QCD} \simeq 1$ GeV the scale at which QCD becomes non perturbative.
An effective Hamiltonian can then be built in the form:

$$\mathcal{H}_{eff} = \sum_i C_i(\mu) O_i(\mu). \tag{1.12}$$

Where the coefficients C_i describing the physics at high energy are called *Wilson Coefficients* and O_i are all the operators compatible with the symmetries of the system. Here, μ is an intermediate scale between the high energy and low energy limits. Specifically, in the case of the weak decay of a hadron, the effective hamiltonian can be expressed as:

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i^N V^i_{CKM} C_i(\mu) O_i + h.c, \tag{1.13}$$

where G_F is the Fermi constant such that $G_F/\sqrt{2} = g^2/8M_W^2$ and h.c stands for Hermitian conjuguate.

1.3 The $B \to K^{(*)}\nu\overline{\nu}$ decays in the Standard Model

Following the framework described in Section 1.2 the effective Hamiltonian describing $b \to sll$ transitions (including $b \to s\nu\bar{\nu}$) is:

$$\mathcal{H}_{eff} = -\frac{G_F}{\sqrt{2}} V_{ts}^* V_{tb} \times [\sum_{i=1}^{6} C_i O_i + c_{7\gamma} O_{7\gamma} + c_{8G} O_{8G} + c_{9V} O_{9V} + c_{10A} O_{10A} + C_L^\nu O_L^\nu + C_L^\mu O_L^\mu] + h.c \tag{1.14}$$

Where the $|V_{us}^* V_{ub}|$ term is omitted as the t quark related term is $\simeq 50$ times greater. This is the origin of the breakdown of the GIM mechanism at the one-loop level which causes FCNCs to appear at one-loop level [30]. The operators $\mathcal{O}_i$ are as described in [31]:

$$
\begin{aligned}
\mathcal{O}_1 &= (\bar{s}_i c_j)_{V-A}(\bar{c}_j b_i)_{V-A} \\
\mathcal{O}_2 &= (\bar{s}c)_{V-A}(\bar{c}b)_{V-A} \\
\mathcal{O}_3 &= (\bar{s}b)_{V-A}\sum_q(\bar{q}q)_{V-A} \\
\mathcal{O}_4 &= (\bar{s}_i b_j)_{V-A}\sum_q(\bar{q}_j q i)_{V-A} \\
\mathcal{O}_5 &= (\bar{s}b)_{V-A}\sum_q(\bar{q}q)_{V+A} \\
\mathcal{O}_6 &= (\bar{s}_i b_j)_{V-A}\sum_q(\bar{q}_j q_i)_{V+A} \\
\mathcal{O}_{7\gamma} &= \frac{e}{8\pi^2} m_b \bar{s}_i \sigma^{\mu\nu}(1+\gamma_5) b_i F_{\mu\nu} \\
\mathcal{O}_{8G} &= \frac{g}{8\pi^2} m_b \bar{s}_i \sigma^{\mu\nu}(1+\gamma_5) T^a_{ij} b_j G^a_{\mu\nu} \\
\mathcal{O}_{9V} &= (\bar{s}b)_{V-A}(\bar{\ell}\ell)_V \\
\mathcal{O}_{10A} &= (\bar{s}b)_{V-A}(\bar{\ell}\ell)_A \\
\mathcal{O}^{\nu}_L &= (\bar{b}s)_{V-A}(\bar{\nu}\nu)_{V-A} \\
\mathcal{O}^{\ell}_L &= (\bar{b}s)_{V-A}(\bar{\ell}\ell)_{V-A}
\end{aligned}
\tag{1.15}
$$

In the case of $b \to s\nu\bar{\nu}$ transitions, $\mathcal{O}^{\nu}_L$ is the sole contributing operator. The corresponding dimensionless Wilson coefficient C^{SM}_L is defined as:

$$C^{SM}_L = -X_t/s^2_w, \tag{1.16}$$

where $X_t = 1.468(17)$ the two-loop electroweak corrections to the top-quark contribution to the decay and $s^2_w = \sin^2\theta_w = 0.23126(5)$, with θ_w the electroweak mixing angle [32]. Thus, C^{SM}_L is known to a precision of $\mathcal{O}(1\%)$.
From there, the total branching fraction of the $B \to K\nu\bar{\nu}$ decay can be derived from Fermi's golden rule:

$$\mathcal{B}(B \to K\nu\bar{\nu}) = N\tau_B |\langle K\nu\bar{\nu}|\mathcal{H}_{eff}|B\rangle|^2 \rho, \tag{1.17}$$

with N a normalization factor, τ_B the lifetime of the B meson and ρ a phasespace factor.

However, it is more convenient to study the differential $B \to K\nu\bar{\nu}$ branching ratio with respect to the squared invariant mass of the neutrino system (q^2, defined in subsubsection 5.4.1.5):

$$\frac{d\mathcal{B}(B \to K\nu\bar{\nu})}{dq^2} = \frac{(\eta_{EW} G_F)^2 \alpha^2_{EW} X^2_t}{32\pi^5 \sin^4\theta_W} \times \tau_B |V_{tb}V^*_{ts}|^2 |\vec{p}_K|^3 f^2_+(q^2) \tag{1.18}$$

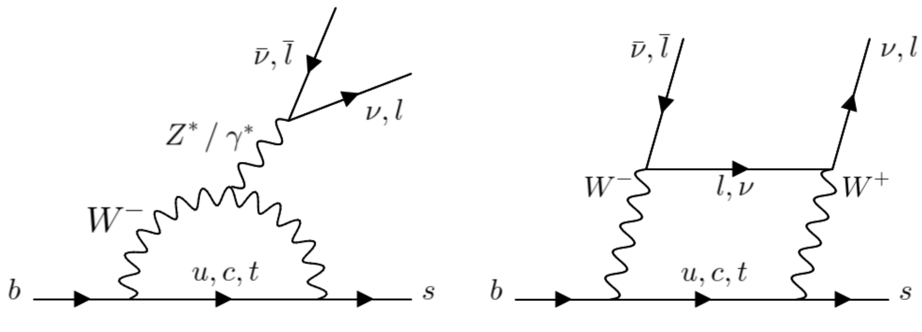


Figure 1.2: One-loop (left) and box (right) Feynman diagrams for $b \to s l^+ l^-$ and $b \to s \nu \bar{\nu}$ processes

Where α_{EW} is the fine structure constant evaluated at the Z boson mass, η_{EW} is a short-distance correction factor to G_F and $f_+(q^2)$ is a vector form factor described in [33–35]. When integrating over the full q^2 range, this gives [36]:

$$\mathcal{B}(B^+ \to K^+ \nu\bar{\nu}) = (4.43 \pm 0.42) \times 10^{-6}. \tag{1.19}$$

From Equation 1.18, one can also derive:

$$\mathcal{B}(B^0 \to K^0 \nu\bar{\nu}) = (4.10 \pm 0.38) \times 10^{-6}. \tag{1.20}$$

With the ratio of the two branching ratios being equal to τ_{B^+}/τ_{B^0}. The value in Equation 1.19 does not take into account the long-distance contribution [37] from the intermediate tau state ($B^+ \to \tau^+ \bar{\nu_\tau}$ and $\tau^+ \to K^+ \nu_\tau$), which is treated in Chapter 5 as an irreducible background:

$$\mathcal{B}(B^+ \to K^+ \nu_\tau \bar{\nu}_\tau)_{LD} = \frac{|(\eta_{EW} G_F)^2 V_{ub} V_{us}^* f_{K+} f_{B+}|^2}{128\pi^2 M_{B+}^3} \times \frac{m_\tau (M_{B+}^2 - m_\tau^2)^2 (M_{K+}^2 - m_\tau^2)^2}{\Gamma_\tau \Gamma_{B+}} \tag{1.21}$$

Where f_{K+} and f_{B+} are the kaon and B-meson decay constants respectively. This gives:

$$\mathcal{B}(B^+ \to K^+ \nu_\tau \bar{\nu_\tau})_{LD} = (6.28 \pm 0.06) \times 10^{-7}. \tag{1.22}$$

Finally, taking into account additional form factors [36], one finds:

$$\mathcal{B}(B^+ \to K^{*+} \nu\bar{\nu}) = (10.86 \pm 1.89) \times 10^{-6}, \tag{1.23}$$
$$\mathcal{B}(B^0 \to K^{*0} \nu\bar{\nu}) = (9.05 \pm 1.80) \times 10^{-6}. \tag{1.24}$$

The different branching fraction values are subjected to change based on the CKM parameters and form factors used in computation (*e.g.* see [35]). The main source of uncertainty in Equation 1.19 comes from the form factor $f_+(q^2)$. Thus, we further develop on how this form factor is computed, which will allow us to accurately estimate its effect on the total uncertainty of our measurement down the line. From [38], the form factor can be parametrized using three parameters $\alpha_0, \alpha_1, \alpha_2$, such as:

$$f_+(q^2) = \frac{1}{1 - q^2/m_+^2}[\alpha_0 + \alpha_1 z(q^2) + \alpha_2 z^2(q^2) + \frac{z^3(q^8)}{3}(-\alpha_1 + 2\alpha_2)], \tag{1.25}$$

with

$$z(t) = \frac{\sqrt{t_+ - t} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - t} + \sqrt{t_+ - t_0}}, \tag{1.26}$$

where $t_\pm = (m_B \pm m_K)^2$, $t_0 = t_+(1 - \sqrt{1 - t_-/t_+})$ and $m_+ = m_B + 0.046$ GeV. Using lattice computation valid at high q^2 as well as the light cone sum rules to cover the full kinematical region, a fit performed in [35] gives:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0.2545 \\ -0.71 \\ 0.32 \end{pmatrix}, \tag{1.27}$$

with the associated uncertainty vector σ:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_0 \\ \sigma_1 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} 0.0090 \\ 0.14 \\ 0.59 \end{pmatrix}. \tag{1.28}$$

In addition, to propagate the uncertainties on the value of the $\boldsymbol{\alpha}$ parameters, we compute the covariance matrix C_α of α from the correlation matrix given in [35] as:

$$C_\alpha = \begin{pmatrix} 1.0 & 0.32 & -0.37 \\ 0.32 & 1.0 & 0.26 \\ -0.37 & 0.26 & 1.0 \end{pmatrix} \tag{1.29}$$

1.4 Search for New Physics in $b \to s\nu\bar{\nu}$ transitions

This section explains general corrections from NP to the effective treatment of the $B \to K^{(*)}\nu\bar{\nu}$ decays. We then briefly introduce several NP models impacting to these decays and show how the measurement of $\mathcal{B}(B \to K\nu\bar{\nu})$ allows to constrain these models.
Considering NP (at energies larger than the B-meson mass), two additional operators C_L and C_R appear in the effective low-energy Hamiltonian of Equation 1.14:

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^*(C_L\mathcal{O}_L + C_R\mathcal{O}_R) + h.c \tag{1.30}$$

With:

$$\mathcal{O}_R = \frac{e^2}{16\pi^2}(\bar{s}\gamma_\mu P_R b)(\bar{\nu}\gamma^\mu(1-\gamma_5)\nu) \tag{1.31}$$

It is important to note that LFU of NP is assumed here. It is then possible to define two real parameters $\varepsilon > 0$ and $\eta \in [-\frac{1}{2}, \frac{1}{2}]$, defined from the Wilson coefficients:

$$\varepsilon = \frac{\sqrt{|C_L|^2 + |C_R|^2}}{C_L^{SM}}, \quad \eta = \frac{-Re(C_L C_R^*)}{|C_L|^2 + |C_R|^2} \tag{1.32}$$

Thus, $\varepsilon = 1$ and $\eta = 0$ in the SM. Deviations would signal the presence of right-handed currents.
The branching ratios of $B \to K\nu\bar{\nu}$ and $B \to K^*\nu\bar{\nu}$ can then be linked to ε and η:

$$\begin{aligned} R_K^\nu &\equiv \frac{\mathcal{B}(B \to K\nu\bar{\nu})}{\mathcal{B}(B \to K\nu\bar{\nu})_{SM}} = (1 - 2\eta)\varepsilon^2, \\ R_{K^*}^\nu &\equiv \frac{\mathcal{B}(B \to K^*\nu\bar{\nu})}{\mathcal{B}(B \to K^*\nu\bar{\nu})_{SM}} = (1 + \kappa\eta)\varepsilon^2 \end{aligned} \tag{1.33}$$

Where κ is a ratio of binned form factors [34]. Thus, the measurement of different $\mathcal{B}(B \to K^{(*)}\nu\bar{\nu})$ allows to constrain C_L and C_R and quantify hypothetical NP effects (Figure 1.3).
Even though in principle no general constraint on the size of NP effects in $B \to K\nu\bar{\nu}$ decays can be gauged from other processes, several models draw a link between $b \to s\nu\bar{\nu}$ and $b \to sl^+l^-$ transitions, as left-handed neutrinos and charged leptons are grouped in doublets under the $SU(2)_L$ gauge symmetry. Thus the disparition of tensions seen in $b \to sl^+l^-$ transitions mentionned in the introduction limits the size of possible NP effects in $B \to K^{(*)}\nu\bar{\nu}$. However, there are still models in which NP effects in $B \to K^{(*)}\nu\bar{\nu}$ arise without constraints from $b \to sl^+l^-$. These different cases will be briefly discussed in the next sections.

In addition, lepton flavour has been thus far neglected because all three neutrino flavours contribute to $B \to K^{(*)}\nu\bar{\nu}$ and they cannot be distinguished experimentally. However in the case of $b \to sl^+l^-$ transitions, measurements have only been performed for $l = e, \mu$, with the muon modes providing the most precise results and the electron modes being less constrained. In addition, $b \to s\tau^+\tau^-$ modes have not been observed at all because of the experimental challenge posed by the tau-leptons reconstruction. However if NP couples mostly to the third generation of leptons, large modifications in $B \to K\nu\bar{\nu}$ could be seen while being compatible with $b \to se^+e^-$ and $b \to s\mu^+\mu^-$ observations.

Finally, as mentionned before, the experimental apparatuses of current collider experiments do not allow to detect neutrinos. Thus, the measurement of $B \to K^{(*)}\nu\bar{\nu}$ decays actually includes all $B \to K^{(*)}+$ invisible modes, with the additional particles being potential dark matter or SUSY candidates. If such particles were contributing here, the measured value of the $B \to K^{(*)}+$ invisible could be enhanced while being compatible with $b \to sl^+l^-$ observations.

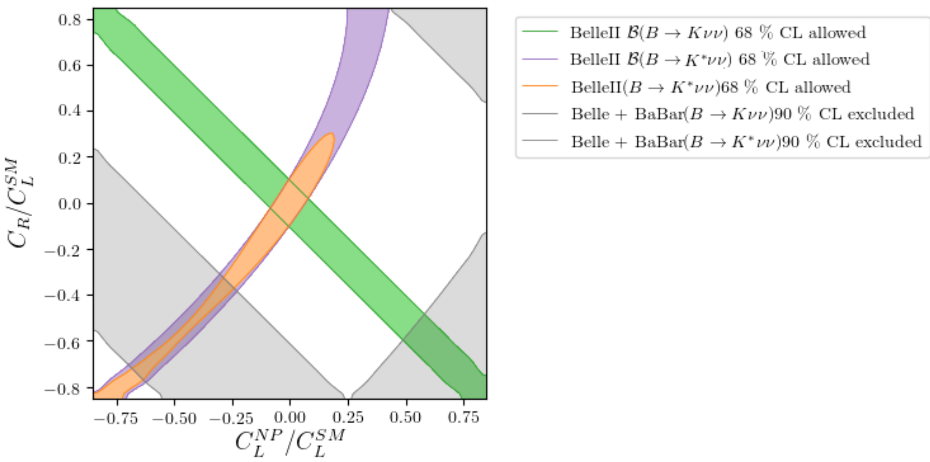


Figure 1.3: Constraints put on the C_R and C_L^{NP} Wilson coefficients with combined Belle, BaBar and Belle II measurements (expected at target luminosity $\mathcal{L} = 50\ \mathrm{ab}^{-1}$) of $B \to K^{(*)}\nu\bar{\nu}$ observables. The grey areas correspond to 90% confidence level exclusion regions from published measurements of $\mathcal{B}(B \to K^{(*)}\nu\bar{\nu})$. The colored zones correspond to 68% confidence level allowed regions from expected Belle II measurements of $\mathcal{B}(B \to K\nu\bar{\nu})$ (green), $\mathcal{B}(B \to K^*\nu\bar{\nu})$ (purple) and the longitudinal polarization fraction F_L of $B \to K^*\nu\bar{\nu}$ defined in [34] (orange). Produced using `Flavio` [39].

1.4.1 Flavour changing massive neutral boson (Z')

Modifications to the $B \to K^{(*)}\nu\bar{\nu}$ decays can occur through the introduction of an additional massive neutral gauge boson, *i.e.* Z'. Such an addition could significantly enhance the decay rate of $B^+ \to K^+\nu\bar{\nu}$ by allowing tree-level $b \to s$ transitions (Subsection 1.4.1). Several Z' models have been described (see [40] and references within), however, the SM-like behavior observed in $b \to sl^+l^-$ tends to constrain some of them.
Still, it is possible to accomodate a $B^+ \to K^+\nu\bar{\nu}$ enhancement from Z' contributions with $b \to sl^+l^-$ observations, for example by having a light Z' decaying primarily invisibly, or by requiring a third-generation coupling preference for said boson.

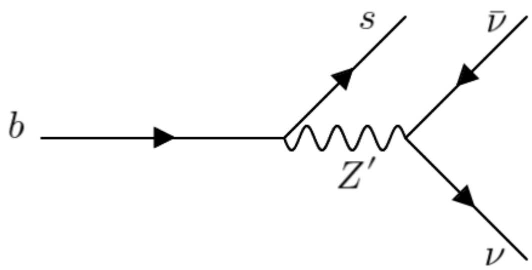


Figure 1.4: Tree-level contribution to $b \to s\nu\bar{\nu}$ transitions mediated by a Z' boson.

An additional case combines an hypothetical new light neutrino coupling to a Z' boson, described in [41]. However, the number of neutrino flavours $N_\nu = 2.9840 \pm 0.0082$ [42] is severly constrained by measurements of the invisible Z boson decay width at LEP [43] and cosmological constraints. A light sterile neutrino interacting with the SM through a Z' could however exist while contributing only marginally to the Z decay width and N_ν, while modifying the values of $R_\nu^{(*)}$ (see Figure 1.5). In addition, this model has the benefit of being unconstrained by $b \to sl^+l^-$ observations.

1.4.2 Leptoquarks

Several models introduce leptoquarks (LQ), heavy scalar or vector particles interacting with both quarks and leptons allowing tree-level FCNC transitions. Numerous LQ scenarios have been explored [44–50], while some have been designed to accomodate the previously seen $b \to sl^+l^-$ tensions with the SM, numerous others do not required such tensions or are even incompatible with them and could thus be now reconsidered. These LQ could imply a significant increase of $\mathcal{B}(B \to K\nu\bar{\nu})$, as can be seen in Figure 1.6.

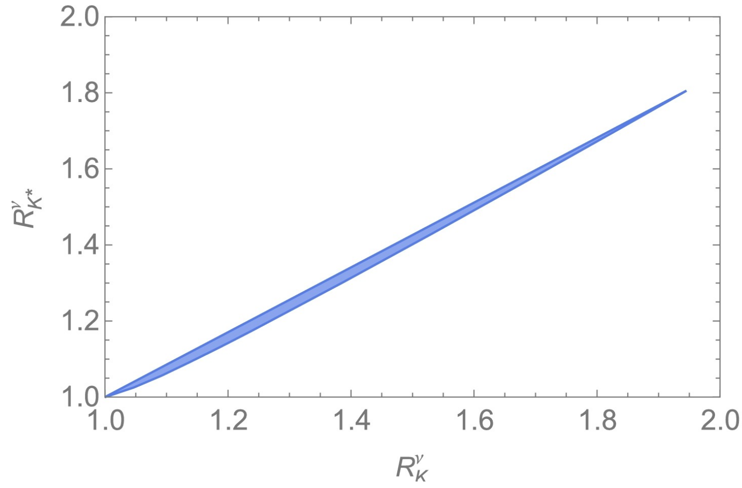


Figure 1.5: Correlation between R^{ν}_{K} and $R^{\nu}_{K^*}$ with in blue the allowed region from the model described in [41] showing the increase of the $B \to K^{(*)}\nu\bar{\nu}$ decay rates with regard to their SM values.

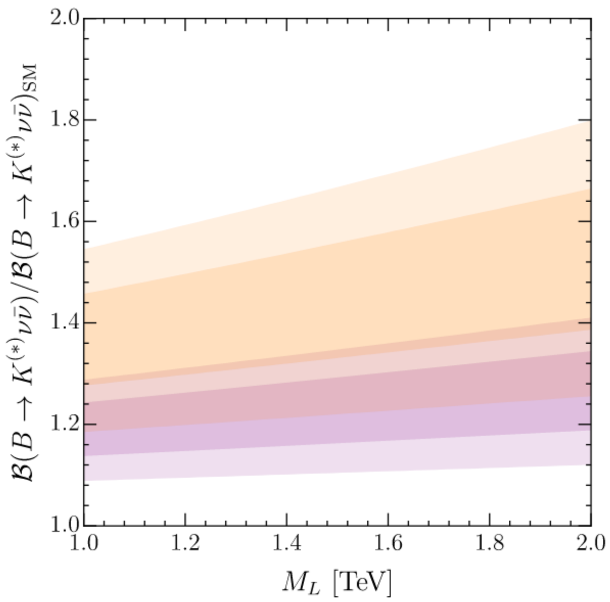


Figure 1.6: Prediction (best 1σ and 2σ fit regions) for the predicted $B \to K^{(*)}\nu\bar{\nu}$ branching ratio as a function of M_L, the mass of the vector-like leptons involved the LQ couplings. Orange and purple bands correspond to different coupling values. From [50].

1.4.3 $B \to K^{(*)}$ + invisible

In addition to adding constraints to models having an effect on $b \to s\nu\bar{\nu}$ transitions, the measurement of $\mathcal{B}(B \to K\nu\bar{\nu})$ allows to indirectly study any NP invisible (= weakly or non-interacting) particles. Inded, as neutrinos are not seen in most particle colliders experiments, measuring $\mathcal{B}(B \to K\nu\bar{\nu})$ actually boils down to measuring $\mathcal{B}(B \to K + I)$, with I being any number of non-detectable particles, including neutrinos. In this section, we briefly describe two NP invisible particle candidates.
QCD axions (A^0) are hypothetical bosons introduced to solve the strong CP problem [51–54]. They are expected to be very-weakly interacting and light ($\mu eV < c^2 \times m_{A^0} < eV$). Measurements of $B \to K\nu\bar{\nu}$ decays allow to impose bounds on $B \to KA^0$ [55].
Other pseudoscalar particles sharing similarities with the QCD axion, Axion-Like Particles (ALPs), noted a' are also described, with masses $m_{a'}$ varying greatly between a few MeV and GeV. Searches for ALPs in $b \to s$ transitions have already been performed in the cases where a' decay visibly [56, 57].
Both axions and ALPs could couple to $W^{\pm}$ bosons (Figure 1.7) and their invisible decays could enhance the $B \to K^{(*)}\nu\bar{\nu}$ decay rates.

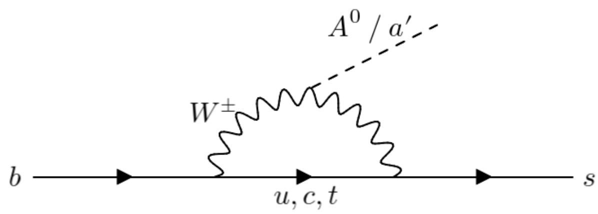


Figure 1.7: Loop-level contribution to $b \to s + I$ transitions from QCD axions A^0 and ALPs a'.

Dark matter (DM), the hypothetical weakly interacting matter expected to contribute $\sim 25\%$ of the energy density of the universe (to be compared to the $\sim 5\%$ of ordinary matter) [58], can also be constrained by the study of $B \to K + I$ decays. In particular, a scalar S with $m_S \simeq 1$ GeV/c^2 [59, 60] could play into a $B \to KS$ decay, with S decaying into either a pair of invisible DM fermions or a visible final state leaving no signature in the detector (as S would be long-lived at detector scale).

1.5 Previous $B \to K^{(*)}\nu\bar{\nu}$ decay searches

As shown in previous sections, the search for $B \to K\nu\bar{\nu}$ is strongly motivated and has thus been performed several times in the past. However, because of the particles escaping detection in the final state of the decays coming from the neutrino pair and the SM-expected low branching ratio, such a study proves to be experimentally

Experiment	Year	$\mathcal{L}[\text{fb}^{-1}]$	Method	Mode	Limit at 90% CL	Ref
BaBar	2010	418	SL	K^+	$< 1.3 \times 10^{-5}$	[61]
				K^0	$< 5.6 \times 10^{-5}$	
BaBar	2013	429	HAD	K^+	$< 3.7 \times 10^{-5}$	[62]
				K^0	$< 8.1 \times 10^{-5}$	
				K^{*+}	$< 11.6 \times 10^{-5}$	
				K^{*0}	$< 9.3 \times 10^{-5}$	
			COM	K^+	$< 1.6 \times 10^{-5}$	
				K^0	$< 4.9 \times 10^{-5}$	
				K^{*+}	$< 6.4 \times 10^{-5}$	
				K^{*0}	$< 12 \times 10^{-5}$	
Belle	2013	711	HAD	K^+	$< 5.5 \times 10^{-5}$	[63]
				K^0	$< 19.4 \times 10^{-5}$	
				K^{*+}	$< 4.0 \times 10^{-5}$	
				K^{*0}	$< 5.5 \times 10^{-5}$	
Belle	2017	711	SL	K^+	$< 1.9 \times 10^{-5}$	[64]
				K^0	$< 2.6 \times 10^{-5}$	
				K^{*+}	$< 6.1 \times 10^{-5}$	
				K^{*0}	$< 1.8 \times 10^{-5}$	
Belle II	2021	63	INC	K^+	$< 4.1 \times 10^{-5}$	[65]

Table 1.2: Results of previous searches for $B \to K^{(*)}\nu\bar{\nu}$ decays, given with the experiment name, year of publication, integrated luminosity of the data sample and method used (SL stands for semileptonic tagging, HAD for hadronic tagging, COM for a combination of the two and INC for an inclusive method).

challenging and requires specific instrumentation.
To this day, three experiments have attempted to observe $B \to K\nu\bar{\nu}$ decays: Belle, Belle II and BaBar. All three experiments belong to a type of particle-collider experiments called B-factories, which will be described in Chapter 2. Belle II is the most recent B-factory while BaBar and Belle belong to the previous generation of such experiments.
Because of the experimental challenge, the previous searches have only allowed to set upper limits on the branching ratios of $B \to K\nu\bar{\nu}$ decays.

Studies performed at Belle and BaBar relied on hadronic or semileptonic *tagging* described in Section 4.1 while the Belle II search was based on an inclusive method, identifying the kaon in the final state of the $B^+ \to K^+\nu\bar{\nu}$ decay with the highest-momentum track in the event and associating all the remaining information in the event to reconstruct the second B-meson of a $\Upsilon(4S) \to B^+B^-$ decay. Table 1.2 and Figure 1.8 summarise the results of the previous $B \to K^{(*)}\nu\bar{\nu}$ searches.

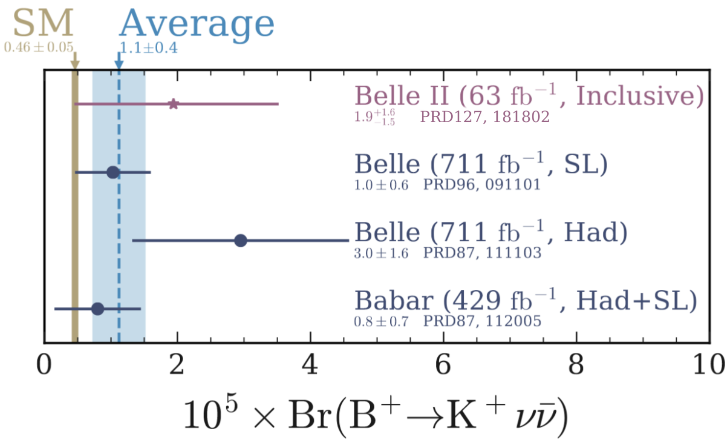


Figure 1.8: Results of previous measurements of the $B^+ \to K^+\nu\bar{\nu}$ decay by the BaBar, Belle and Belle II collaborations, with the different tagging methods specified (SL stands for semileptonic tagging, Had for hadronic tagging and Inclusive for inclusive tagging).

CHAPTER 2

Experimental setup

Contents

This chapter presents the experimental setup used in this thesis comprised of the SuperKEKB accelerator and the Belle II detector. Section 2.1 gives a brief description and history of *B-factories*, of which Belle II is the latest iteration. Section 2.2 presents the SuperKEKB accelerator while Section 2.3 describes the Belle II detector. In particular, Subsection 2.3.2 describes the Belle II Silicon Vertex Detector on which the study shown in Chapter 3 has been performed. Finally, Section 2.6 and 2.7 present the experiment-specific software tools used in the simulation and reconstruction of collision events.

2.1 On B-factories

B-factories are collider particle physics experiments designed to specifically study B-mesons (and to some extent τ-leptons and D-mesons) physics. To produce a large number of B-mesons, these experiments rely on collisions between electrons and positrons at the energy of a $b\bar{b}$ resonnance, the $\Upsilon(4S)$ meson of mass 10.58 GeV/c^2. The $\Upsilon(4S)$ has around 100% chance of decaying into a pair of B-mesons, with about the same probability of decaying into B^+B^- and $B^0\bar{B}^0$ pairs [66]. This setup allows for several experimental perks: compared to hadron-hadron collisions (p-p, Pb-Pb), electron-positron collisions produce few particles which eases event reconstruction. In addition, the four momentum of the e^+e^- system is known which allows to reject background and infer the presence of undetected particles (neutrinos, DM candidates, particles outside the detector acceptance) in the final state of the event. This proves especially useful in this analysis, where most of the signal consists of undetected neutrinos.

The production and study of a large number of B-mesons is motivated by the precise measurement of SM processes with the goal of discovering NP. Indeed, B-meson decays operate through the weak interaction which possesses interesting properties (flavour change, CP symmetry violation). In addition, NP might couple more heavily to third generation fermions, such as b-quark and τ-lepton which further motivates B-factories physics programs.

To this day, three specimen of B-factories have been built. The first generation of B-factories, BaBar and Belle, started collecting data at the end of the 1990s. BaBar was based in Stanford, USA and has collected 433 fb^{-1} of data at the $\Upsilon(4S)$ resonnance provided by the PEP-II accelerator between 1999 and 2008 [67]. Belle was based in Tsukuba, Japan and has collected 711 fb^{-1} of data at the energy of the $\Upsilon(4S)$ resonance between 1999 and 2010 using the KEKB accelerator [68]. The analysis of the data from both experiments is still ongoing [69]. The second generation of B-factories (*Super* B-factories) consists solely of the Belle II experiment, direct successor of Belle described in more detail in this section. Belle II started collecting data in 2019, accumulating until the first half of 2022 a dataset of 424 fb^{-1} (see Figure 2.1), out of which 362 fb^{-1} have been collected at the $\Upsilon(4S)$ mass.

However, B-factories are not the only experiments focused on the study of B-meson physics. The LHCb experiment, located at the France-Switzerland border along the Large Hadron Collider (LHC) studies B-mesons produced by proton-proton collisions at an energy of several TeV. This experimental setup makes use of the large production rate of B-mesons at high energy at the expense of lower luminosity and the loss of information on the four momentum of the collision event.

2.2 The SuperKEKB accelerator

SuperKEKB is an asymmetric circular electron-positron collider, 3 kilometers in diameter, operating with an energy around the $\Upsilon(4S)$ mass. The electron beam is generated in a pre-injector at the beginning of a linear accelerator (LINAC) and

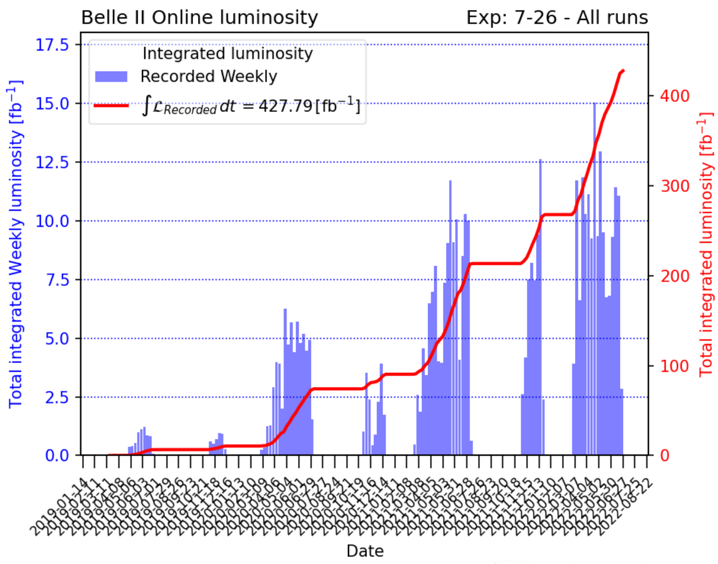


Figure 2.1: Evolution of the total integrated luminosity recorded by the Belle II experiment before the first long shutdown (LS1).

is accelerated to an energy $E_{e^-} = 7.007$ GeV. The positron beam is obtained by irradiating a tungsten target with electrons produced in the pre-injector. Produced positrons are then accelerated up to $E_{e^+} = 4.0$ GeV. Beams are stored in two storage rings, the High Energy Ring (HER) for e^- and Low Energy Ring (LER) for e^+ and collided at the interaction point (IP) of the Belle II detector (see Figure 2.2). The energy of the collision in the center of mass (CM) is given by:

$$\sqrt{s} = \sqrt{\left(\frac{E_{e^-} + E_{e^+}}{c^2}\right)^2 - \left(\frac{\mathbf{p}_{e^-} + \mathbf{p}_{e^+}}{c}\right)^2} \approx 10.58\ \mathrm{GeV/c^2} \tag{2.1}$$

Where $\mathbf{p}_{e^-}, \mathbf{p}_{e^+}$ are the three-momenta of the leptons. Because of the asymmetric energy of the positron and electron beams, the products of the collision undergo a Lorentz boost defined as:

$$\beta\gamma = \frac{\mathbf{p}_{e^-} - \mathbf{p}_{e^+}}{\sqrt{s}} \simeq 0.28 \tag{2.2}$$

The energy asymmetry values of the beams are voluntarily set to produce such a boost, as it helps identifying the decay vertices of the B mesons, wich is especially useful in the case of time-dependent CP violation analyses. Table 2.1 shows the different physics processes producible with this configuration.

Even though SuperKEKB uses the same tunnel as KEKB and shares similarities in beam energies (8 GeV electron and 3.5 GeV positron beams in the case of KEKB), it is expected to reach a luminosity 40 times higher than its predecessor. To reach

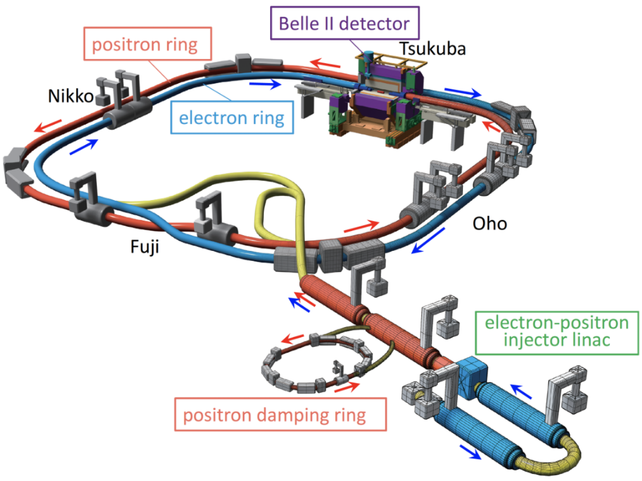


Figure 2.2: Schematic view of the SuperKEKB accelerator showing the LINAC, the positron damping ring used to reduce the emittance of positrons, the electron ring (HER) and positron ring (LER). Collision events happend at the Interaction Point located at the heart of the Belle II detector. Taken from [25].

this goal, the main improvements to SuperKEKB consist of a twofold increase to the HER/LER currents as well as the "nano-beam" scheme which was intially invented for the SuperB project [70]. The concept behind this scheme is to reduce the beam size at the collision point by a factor of 20 compared to KEKB. The vertical width σ_y of the lepton bunches is squeezed to a minimal value of $\simeq 50$ nm, which results in the "hourglass effect" where the minimal value is only reached in a small region along the z axis corresponding to the beam direction. To counter this, the horizontal half crossing angle is set to 41.5 mrad (compared to 11.5 mrad at KEKB). This allows to drive the instantaneous luminosity L which depends on x as $L \sim 1/\sigma_x^2$.

However, the higher currents and reduced beam size give rise to more machine-induced background (beam-gas and Touschek scattering, synchrotron radiation, two photon QED pair production and radiative Bhabha) in the Belle II detector. This poses a challenge as the detector design needs to accomomdate such harsh conditions. So far and since the beginning of the run operation (from early 2019 to mid 2022), the Belle II experiment has recorded 424 fb^{-1} of integrated luminosity delivered at the $\Upsilon(4S)$ energy by the SuperKEKB accelerator, reaching a maximum instantaneous luminosity of $4.7 \times 10^{34}\ \text{cm}^{-2}\ \text{s}^{-1}$. This value consitutes the current world record, while the targeted nominal value is $6 \times 10^{35}\ \text{cm}^{-2}\ \text{s}^{-1}$ (see Figure 2.3). SuperKEKB also allows to vary the beam energies, which gives access to collisions between the $\Upsilon(1S)$ and $\Upsilon(6S)$ energies ($9.46 - 11.24$ GeV). The Belle II experiment thus performs "energy scans" for physics or background characterization studies.

Process	Cross-section [nb]
$e^+e^- \to \Upsilon(4S)$	1.11
$e^+e^- \to u\bar{u}(\gamma)$	1.61
$e^+e^- \to d\bar{d}(\gamma)$	0.40
$e^+e^- \to s\bar{s}(\gamma)$	0.38
$e^+e^- \to c\bar{c}(\gamma)$	1.30
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.92
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.15
$e^+e^- \to e^+e^-(\gamma)$	300.0
$e^+e^- \to e^+e^-e^+e^-$	39.7
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.9
$e^+e^- \to \gamma\gamma(\gamma)$	4.99

Table 2.1: Cross-sections of the main e^+e^- collision processes at $\sqrt{s} = 10.58$ GeV, taken from chapter 4 of [71].

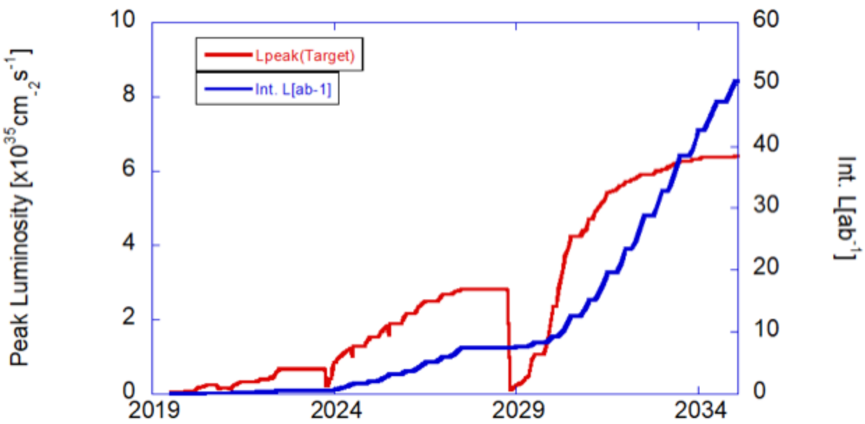


Figure 2.3: Expected evolution of the instantaneous luminosity delivered by SuperKEKB (red) and integrated luminosity (blue), reaching 50 ab^{-1} by 2035. From the Belle II collaboration.

2.3 The Belle II detector

The Belle II detector (Figure 2.4) follows the typical pattern of a modern particle collider detector: it consists of a slightly asymmetric (to account for the Lorentz boost) barrel-shaped series of sub-detectors completed with backward and forward endcaps. Its specificity lies in the different types of detectors designed specifically for Belle II physics program as well as the need to maintain high performances in spite of the high background levels from the SuperKEKB accelerator. The main Belle II subdetectors (described in the following sections) are:

- The PiXel Detector (PXD), closest to the beam pipe. Consisting of one layer of DEPFET silicon pixel sensors (a second layer is currently - first half of 2023 - being installed). Its excellent spatial resolution assists in the vertex localisation.
- The Silicon Vertex Detector (SVD). 4 layers of double-sided silicon strip sensors are used for tracking, vertex reconstruction and particle identification.
- The Central Drift Chamber (CDC), which occupies a larger volume and has a higher granularity compared to Belle's CDC, used for tracking and particle identification.
- A particle identification (PID) system split in barrel and endcap regions. The barrel region consists of a Time Of Propagation (TOP) detector while the forward endcap region is equipped with the Aerogel Ring Imaging CHerenkov (ARICH) detector. These mainly allow to well distinguish between pions and kaons.
- An electromagnetic calorimeter (ECAL) based on the CsI(Ti) crystals of Belle's calorimeter. These are put under much pressure from SuperKEKB's background and thus faster readout electronics have been chosen to reduce pileup.
- A supraconductive magnet producing a 1.5 T magnetic field to bend the trajectories of charged particles within the detector volume.
- A K_L^0 and muon detector (KLM) made of a sandwich of thick iron plates and resistive plate chambers making up the outermost layer of the Belle II detector.

The full detector is described at length in [24].

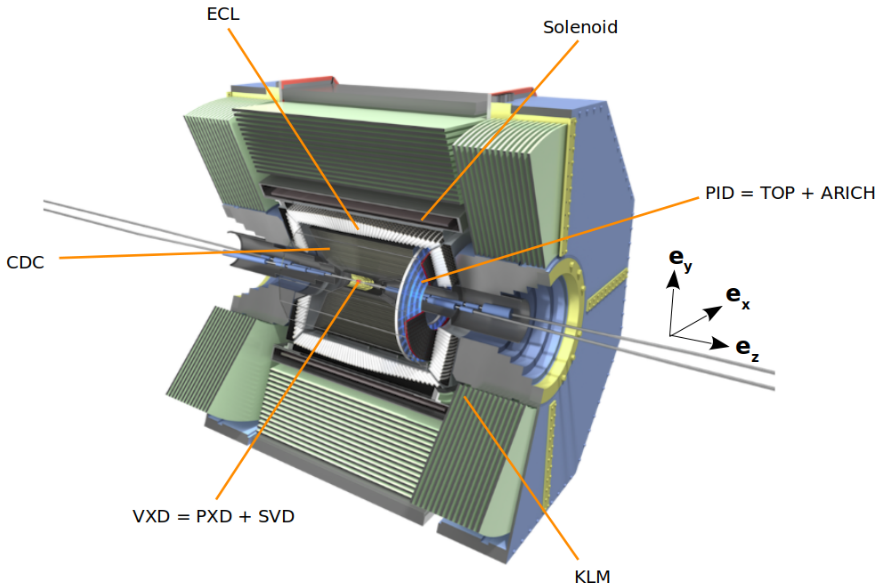


Figure 2.4: Schematic view of the Belle II detector. The origin of the Belle II coordinate system is taken as the nominal interaction point. The z-axis corresponds to the symmetry axis of the solenoid and has a direction close to the electron-beam. The x-axis is in the horizontal plane and points towards the outside of the accelerator ring while the y-axis is vertical and points upwards. The polar angle θ is defined with regards to the z-axis and covers the $[-\pi, \pi]$ interval while the azimuthal ϕ angle is defined in the xy plane, in the range $[0, 2\pi]$. The additional radial coordinate r supplements the polar angles (θ, ϕ) to form a spherical coordinate system. Adapted from Belle II collaboration resources.

2.3.1 The Pixel Detector

Because of the higher machine-inducend background faced by the Belle II detector, the choice has been made to add additional layers to the vertex detection system of Belle II. The Belle detector used to rely solely on a silicon strip vertex detector close to the interaction point, however, background conditions in this region in the nano-beam scheme forbid from using silicon strips. Inded the detector occupancy (the fraction of channel hit in each triggered event) would get too high in the Belle II scenario, which prompted the use of a pixelated detector with a higher number of channels for the innermost layers of the vertex detection system.

The PiXel Detector (PXD) consists of two layers of sensors (numbered L1 and L2) with radii of 14 mm and 22 mm centered around the beam pipe (Figure 2.5). Detection modules, each possessing a matrix of 768×250 pixels are glued by pairs to build ladders. The innermost PXD layer is made of 8 ladders and the second layer is expected to have 12 ladders. At the time of writing only two ladders are installed in the second layer, the full installation of the PXD is expected to take place in the near future. The acceptance covered by the sensor is in the range $17° < \theta < 155°$.

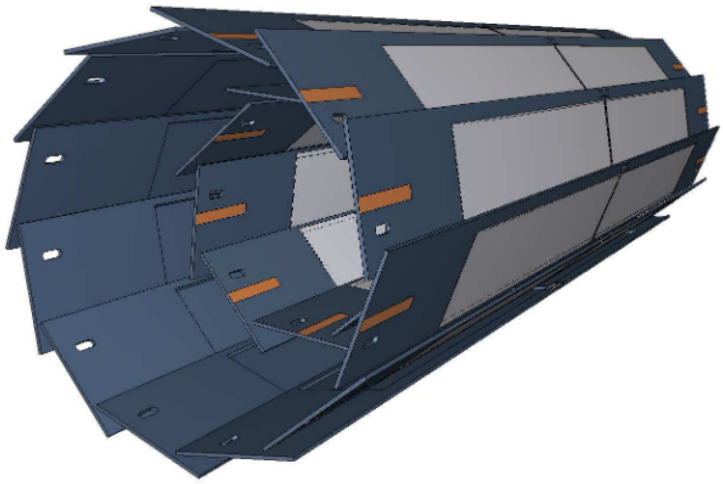


Figure 2.5: Schematic view of the two-layered pixel detector. The grey areas correspond to the DEPFET pixel sensors. The dark blue areas correspond to the sensor mounts. From Belle II PXD group.

The PXD sensors are based on the DEPleted Field Effect Transistor (DEPFET) technology [72] in which a semiconductor detector combines detection and amplification of signal. Figure 2.6 shows the cross section of a DEPFET sensor. Here, a high negative voltage to a p^+ contact on the back side of the device induces the full depletion of a n-type substrate. This creates a potential minimum ("Internal Gate") where the electrons created by a charged particle passing through the fully depleted bulk, while holes drift to the back contact. When the transistor is on, accumulated electrons modulate the channel current. To reset charges in the sensor, a n^+ contact is put to a positive voltage to empty the internal gate.
The readout of the sensor takes 20 μs for a full cycle, with 100 ns of downtime per cycle.

2.3.2 The Silicon Vertex Detector

Futher away from the beam pipe is the Silicon Vertex Detector (SVD). It is arranged in the same geometry as the PXD (concentric layers made of ladders, barrel geometry) and together they make up the Belle II VerteX Detector (VXD). Because of the larger surface area to cover, and because it is less close to the beam pipe, the SVD is equipped with 172 Double Sided Silicon Strip Detectors (DSSD). The number of sensors, their sizes, and the number of strips per ladders vary depending on the layer (see Table 2.2). In addition layers L4, L5 and L6 possess trapezoidal sensors in the forward region of the detectors to cope with the Lorentz boost induced by the asymmetry of the collisions. These are slanted in order to improve the angular acceptance and optimize the incident angle of particles coming from the IP. Trapezoidal sensors are thinner than the rectangular sensors making up the rest of

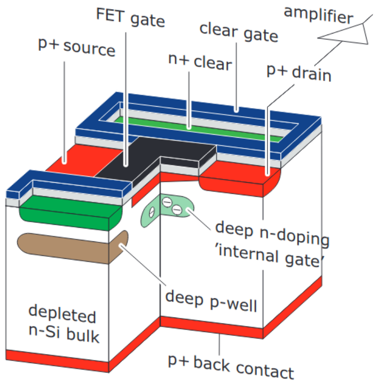


Figure 2.6: Cross-section of a DEPFET sensor. Taken from [24]

Layer number	Ladders/layer	Sensors/layer	Trapezoidal Sensor angle (°)
3	7	2 (2 smalls)	n/a
4	10	3 (2 larges + 1 trapezoidal)	11.9°
5	12	4 (3 larges + 1 trapezoidal)	17.2°
6	16	5 (4 larges + 1 trapezoidal)	21.1°

	Small Sensors	Large Sensors	Trapezoidal Sensors
Readout strips P-side	768	768	768
Readout strips N-side	768	512	512
Readout pitch P-side	50 μm	75 μm	50 - 75 μm
Readout pitch N-side	160 μm	240 μm	240 μm
Sensor active area (mm^2)	122.90×38.55	122.90×57.72	$122.76 \times (38.42 - 57.59)$
Sensor thickness	320 μm	320 μm	300 μm
Manufacturer	Hamamatsu	Hamamatsu	Micron

Table 2.2: Features of the SVD setup. Information taken from [73].

the detector (300 μm versus 320 μm). The total geometric acceptance of the SVD is also $17° < \theta < 150°$.

Because of the relatively low energy of the collisions, particles produced are subject to deflection from multiple scattering, thus their tracks cannot be used for precision alignment. To do so, high-energy cosmic muons are rather used, but as their rate is limited, the SVD is built with an overlap between adjacent sensors in the range $8 \sim 10\%$ (depending on layer) to facilitate alignment, at the cost of a slightly increased material budget.

The DSSDs are made of an N-type bulk with high resistivity on which sensing strips are implanted with either acceptors or donors depending on the sensor side. The side implanted with acceptors is called "u/P-side" while the other side is called "v/N-side". The readout strips on the v/N-side are arranged perpendicularly with regards to the ones on the u/P-side, allowing to measure the z and ϕ direction re-

	Belle	Belle II
Radius of inner cylinder (mm)	77	160
Radius of outer cylinder (mm)	880	1130
Radius of innermost sense wire (mm)	88	168
Radius of outermost sense wire (mm)	863	1111.4
Number of layers	50	56
Number of sense wires	8400	14336
Gas mixture	$He - C_2H_6$	$He - C_2H_6$
Diameter of sense wire (μm)	30	30

Table 2.3: Main parameters of Belle and Belle II drift chambers. Information from [24].

spectively.

Because of the strong constraints brought by high machine background, readout electronics with a fast shaping time is required. APV25 [74] chips, which were initially used in the CMS experiment, were chosen for the SVD. The chips consist of 128 identical channels of low-noise preamplifiers followed by a 50 ns (tunable) shaper stage. APV25 are also sufficiently resistant to radiation and can tolerate an ionising dose in excess of 30 MRad (10 MRad would suffice for the experimental conditions). For each ladder, APV25 chips are installed directly on the sensors, connected by flexible printed circuits with a thermal isolation foam in between. All APV25 chips are installed on the same side of the sensors and are connected to the strips on the other side by flex circuits wraped around the edge in a scheme called origami, referencing the folding action. This design allows to cool all chips using only one cooling pipe, thus reducing material budget (Figure 2.7).

The first data taking period of Belle II confirms the excellent behaviour of the SVD. The strip noise, dominated by APV25 capacitive input load, leads to a satisfactory signal-to-noise ratio, which further validates the choice of the origami chip-on-sensor scheme. The spatial resolution of the detector is $10 \sim 15$ μm for the P-side and $15 \sim 30$ μm for the N-side, with some room for improvement in the reconstruction (see Chapter 3). The hit-time resolution is also good, with 2.4 ns on the N-side and 2.9 ns on the P-side. The hit-time resolution will become crucial when running at the nominal SuperKEKB luminosity to reject off-time beam background hits in the SVD and maintain good tracking efficiency.

2.3.3 The Central Drift Chamber

The role of the Central Drift Chamber (CDC) is threefold: to reconstruct charged tracks and allow to measure their momentum precisely, to provide 3D trigger information for charged particles and to allow to perform particle identification based on energy loss within its gas volume. Because de Belle CDC showed great performance and reliability for over ten years, the Belle II CDC design mainly follows the structure of its predecessor, with Table 2.3 showing the main parameters of both

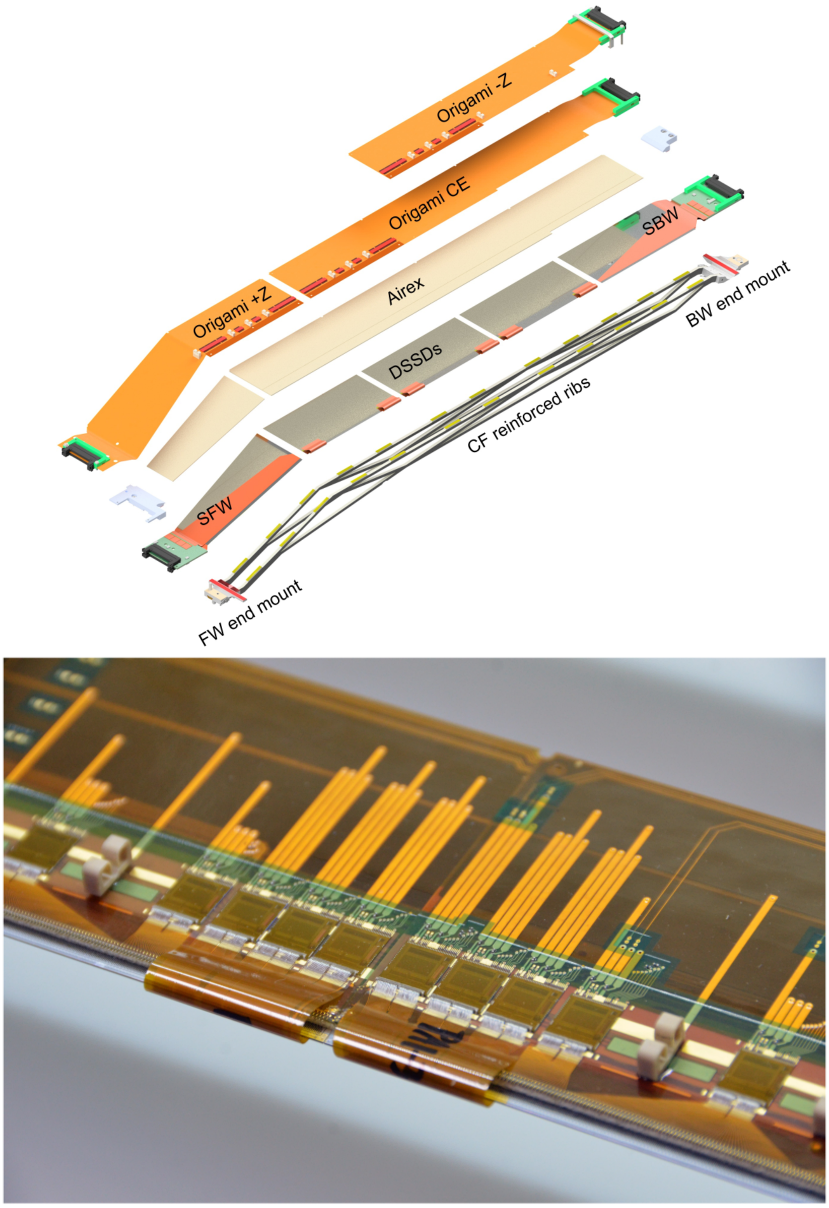


Figure 2.7: Top: Layout of an SVD ladder (layer 6). Bottom: APV 25 chips installed on a sensor and flexible pitch adapters (origami scheme). Images from [73].

detectors.
The CDC consists of a barrel-shaped structure made of an inner and outer carbon-fiber reinforced plastic cylinders and two aluminum endplates (Figure 2.8).

The structure is filled with a 50% helium - 50% ethane mixture, chosen for its

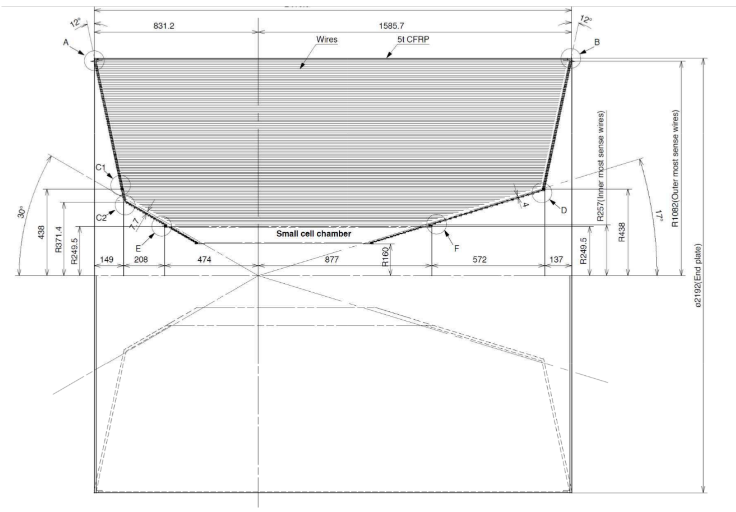


Figure 2.8: Structure of the CDC detector. Image from [24].

adequate drift velocity, low radiation length, good energy loss resolution, good position resolution and low cross section for synchrotron radiation X-rays. Inside the structure, more than 14000 wires are arranged in 56 layers, further divided into 9 superlayers with axial-stereo readout. The wire configuration is shown in Figure 2.9. The geometric acceptance of the CDC is in the range $17° < \theta < 150°$, while the spatial resolution on individual hits is around 100 μm.
The main differences with regards to the Belle CDC are the readout electronics, which need to be able to cope with the higher background.

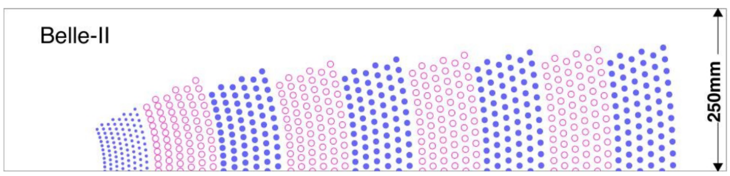


Figure 2.9: Layout of the CDC 9 wire superlayers. The innermost superlayer is made of two layers while the other are composed of six layers each. Image from [75].

2.3.4 Particle Identification (TOP, ARICH)

In Belle II the particle identification (PID) system is composed of two separate Cherenkov detectors. The Time Of Propagation (TOP) detector is located in the barrel region and the Aerogel Ring Imaging CHerenkov (ARICH) detector is installed at the forward endcap of the Belle II detector.

TOP

The TOP consists of quartz radiators (Figure 2.10) arranged in 16 modules around the CDC at a radius of 1.24 m. Charged particles crossing the radiators with enough velocity produce Cherenkov photons that totally reflect at the interface of the quartz. Cherenkov photons are then focused and directed towards micro-channel plate photo multipliers (MCP-PMTs) located at the end of the quartz bar. It is possible to relate the Cherenkov photon emission angle θ_C, to the velocity β of the particle and the refraction index of the radiator n by:

$$\beta = \frac{1}{n \cos\theta_C} \tag{2.3}$$

Here $n = 1.44$ for photons of 405 nm wavelength. It is possible to measure θ_C using information of the time of propagation of the photons in the radiator. The time resolution of the detector is lower than 50 ps, which allows to distinguish between kaons and pions, for which the difference of photon arrival time is $\sim$ 100 ps at 2 GeV/c^2.
To determine the efficiency of the particle identification, the detected photons distributions are tested against probabiliy distribution functions (PDFs) for each particle hypotheses (K, π, e, μ, p, d). For the specific case of K/π separation, the TOP performs well with an efficiency of 85% for a 10% pion misidentification rate [76].

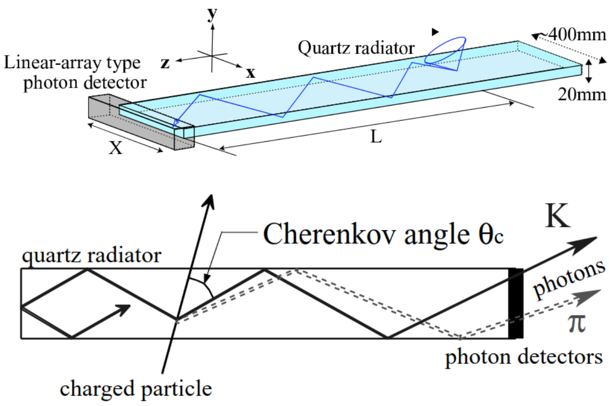


Figure 2.10: Top: Schematic view of a TOP radiator element. Bottom: Side-view showing the working principle of the TOP counter. Image from [24].

ARICH

In the forward endcap region of the Belle II detector, the ARICH is used to provide separation between kaons and pions over most of the momentum range, as well as discrimination between muons, electrons and pions below 1 GeV/c^2. The working principle of the ARICH is also based on a measurement of Cherenkov light. Here, the radiator is made of a silica aerogel, chosen to be highly transparent in order to limit photon loss via Rayleigh scattering or absorption. Two 20 mm thick layers of aerogels are used, with refractive indices of 1.055 and 1.065 (these values are chosen in order for the Cherenkov rings produced in each layer to overlap on the detection plane). After propagating through a 20 cm expansion volume, the produced photons are detected by an array of position sensitive photon detectors, Hybrid Avalanche Photo Detectors (HAPD), read by integrated circuit chips. The sensors and readout electronics were chosen because of their ability to detect single photons in a high magnetic field with a good 2D resolution and high efficiency. Figure 2.11 shows a schematic view of the detector working principle.

The ARICH covers a geometric acceptance in the range $15° < \theta < 30°$ and performs adequately, with a separation efficiency between kaons and pion of 93% with a pion misidentification rate of 10% [77].

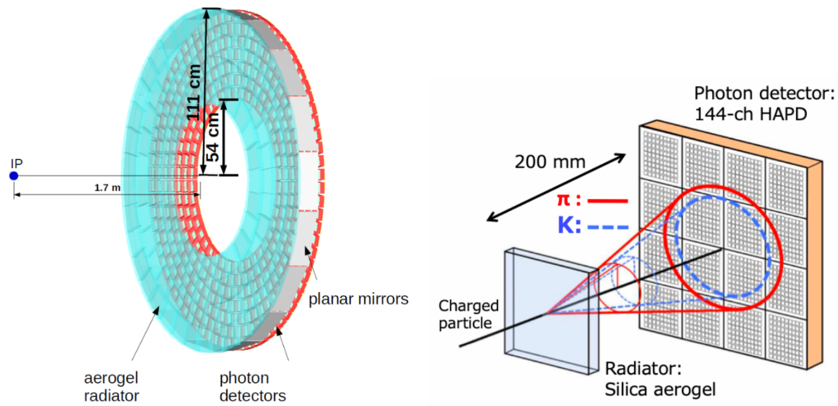


Figure 2.11: Left: Schematic view of the ARICH detector showing its main characteristics. Right: Working principle of the ARICH. From Belle II ARICH group.

2.3.5 The Electromagnetic Calorimeter

The Electromagnetic CaLorimeter (ECL) is used to detect photons, which is crucial in Belle II since one third of B-meson decays produce π^0 and other neutral particles that decay into photons, in a wide energy range ($2 \times 10^{-2} \sim 4$ GeV).

The main calorimeter region consists of 6624 CsI(Tl) pyramidal crystals arranged in a 3 m long barrel shape of inner radius 1.25 m. These crystals have an average cross section of 6×6 cm^2 and an average length of 30 cm (corresponding to 16.1 radiation lengths). This barrel is completed by two endcaps regions, consisting of 2112 CsI crystals, at $z_1 = 2.0$ m and $z_2 = -1.0$ m from the IP (Figure 2.12).

This layout provides a geometric acceptance in the range $12.4° < \theta < 155.1°$, except

for two 1° gaps at the junction of the barrel and endcap regions.
Each crystal is wrapped in a 200 μm thick Teflon layer and covered by a sheet of 25 μm thick aluminium and 25 μm thick mylar.
For each crystal, two 10×20 mm^2 glued-on photodiodes are used for scintillation light readout. A preamplifier associated to each photodiode produces two independ signal outputs for each crystal, these two outputs are then summed in a shaper board.
From performance measurements using cosmic muons, the average output signal for the crystals is estimated at $\sim$ 5000 photoelectrons per MeV for a noise level of $\sim$ 200 keV. The intrinsic energy resolution of the detector can be approximated as:

$$\frac{\sigma_E}{E} = \sqrt{(\frac{0.066\%}{E})^2 + (\frac{0.81\%}{\sqrt{E}})^2 + (1.34\%)^2} \tag{2.4}$$

With E in GeV.

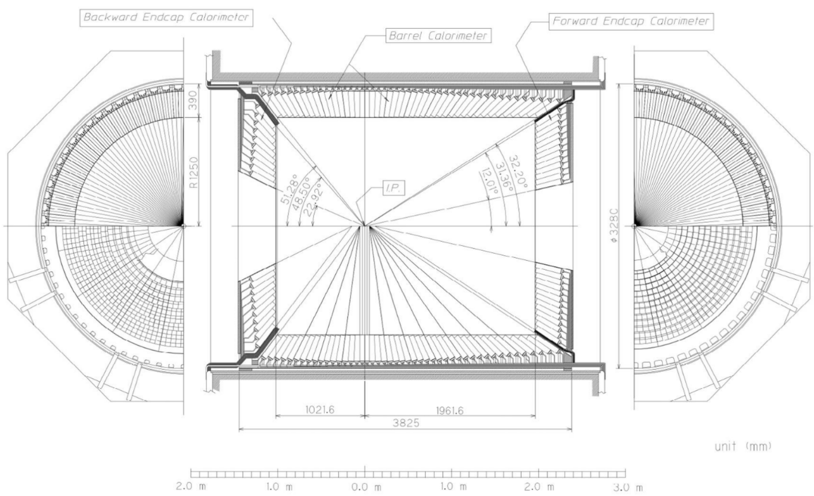


Figure 2.12: Schematic view of the ECL detector showing the three regions (barrel, and both endcaps). Image from [75].

2.3.6 Solenoid

Around the ECL, a superconducting solenoid provides a 1.5 T magnetic field in a cylindrical volume 4.4 m in length and 3.4 m in diameter. The main coil of the solenoid is made out of a NbTi/Cu superconducting alloy, powered with a 4400 A current and cooled with a liquid helium cryogenic system. It is used to bend the charged particles trajectories to allow the measurement of their momentum. In addition, the iron structure of the Belle II detector is used as a return path for the magnetic flux.

2.3.7 The K Long and Muon Detector

The outermost part of the Belle II detector is the K Long and Muon detector (KLM), which consists of alternating layers of 4.7 cm iron and active detector. The iron plates serve as both magnetic flux return for the solenoid and 3.9 radiation lengths of material to allow the K_L^0 to shower hadronically. The KLM is composed of an octogonal barrel region using Resistive Plates Chambers (RPCs) as detection elements and covering a polar angle $45° < \theta < 125°$. Two endcap structures carrying scintillator strips coupled with silicon photomultipliers (SiPM), extend the acceptance to $20° < \theta < 155°$.
The barrel region consists of 15 layers of detectors and 14 iron layers, while the endcap regions use 14 layers of detectors and 14 iron layers [78].
The RPCs are made of two 2 mm glass electrodes planes separated by a 2 mm thick plane filled with a 62% HFC-134a (freon 134a), 30% argon and 8% butane-silver. High-voltage is distributed along the electrodes using a thin layer of carbon-doped paint. Particles going through the gas volume ionize it, generated electrons are then collected by metal strips located at the end of the RPCs. These strips are separated from a ground plane by dielectric foam, working as a transmission line with a characteristic impedance of 50 Ω. In order to improve detection efficiency, two RPCs are coupled to form a superlayer, with Figure 2.13 showing a structure of a superlayer.
The endcap regions suffer more from machine background hit rate as they are not shielded against neutrons. The use of scintillator detectors in these regions is driven by the long dead time of RPCs. The scintillator strips measure up to 2.8 m in length and have a cross section of 7 to 10 mm $\times$40 mm. In total, the endcaps carry 16800 of these scintillator strips. Scintillation light is measured by the SiPMs, the whole detection system has the advantage of having a good time resolution (around 0.7 ns) and high output rate.
For tracks with a momentum above 1 GeV/c, muon detection efficiency reaches 89% for a hadron contamination of 1.3%. The K_L^0 detection efficiency reaches 80% for momenta over 3 GeV/c and decreases linearily for lower momentum values.

2.4 Trigger System

The Belle II trigger system permits the collection of data for physics events of interest. The system is designed to perform adequately at the nominal SuperKEKB luminosity and must thus satisfy several requirements. Its efficiency for hadronic events from $\Upsilon(4S) \to B\bar{B}$ and $e^+e^- \to q\bar{q}$ must be $\sim$ 100% and it should have a maximum average trigger rate of 30 kHz to accommodate the expected collision rate at nominal luminosity. In addition, the trigger fixed latency should be $\sim$ 5 ns and its timing precision be better than 10 ns. The minimum separation power between two events should be at least 200 ns.
As with much of the detector, Belle II trigger system follows the Belle trigger scheme

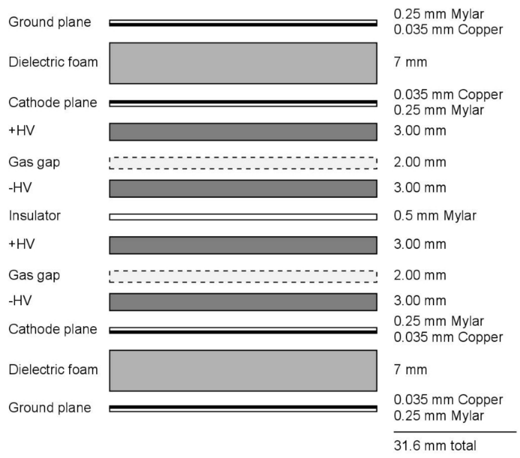


Figure 2.13: Cross-section of an RPC superlayer. A KLM module is composed of two superlayers on each side of an iron plate. Image taken from [24].

with all components replaced to follow the increased event rate. The trigger scheme consists of two tiers: the hardware based Level 1 (L1) trigger uses detector information to remove most of the background while the software-based High Level Trigger (HLT), uses reconstructed event information to reduce data as part of the Data Acquisition System (DAQ).

Level 1 trigger

The L1 trigger is used to reject background events and select events of interest. To do so, it harvests raw information from the Belle II subdetectors thanks to sub-trigger systems. The information is fed to a Global Reconstruction Logic (GRL) which performs a low level reconstruction and sends its output to a centralized Global Decision Logic (GDL) which makes the final decision. All the components of the L1 possess a Field Programmable Gate Array (FPGA) which allows to configure trigger logic.

The CDC sub-trigger, which provides information on charged tracks and the ECL sub-trigger linked to energy clusters in the calorimeter, are at the root of the L1 trigger system. The CDC sub-trigger consists of a 2D trigger based on track reconstruction in the (x, y) plane, followed by a 3D trigger which allows to estimate the z coordinate of the primary vertex of the event. This allows to reject machine background contributions coming away from the IP. The ECL sub-trigger generates fast signals based on the total energy deposited in the calorimeter and number of clusters for events with both charged and neutral particles. The trigger signals from CDC and ECL are then merged with information form the KLM and TOP (Figure 2.14) by the GRL and transmitted to the GDL which performs a trigger decision based on the output of the different sub-systems.

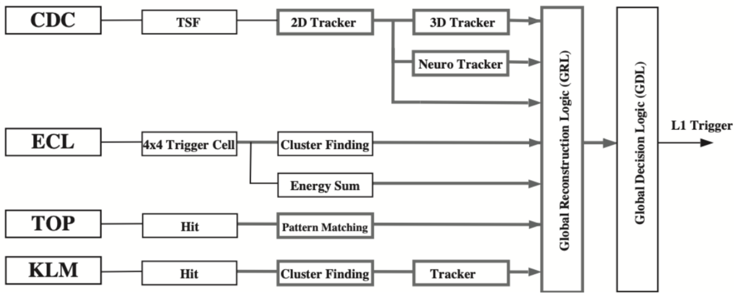


Figure 2.14: Overview of the L1 trigger. Output from the different sub-systems are sent to the GDL which makes the final trigger decision. Image from Belle II trigger group.

The L1 output is then fed to the HLT to further refine the selection.

High Level Trigger

The HLT relies on a full, real time reconstruction of the event based on information from all detectors but PXD. In order to avoid additional systematic uncertainties, the reconstruction software is the one described in Section 2.7, used for offline reconstruction as well.
The software trigger runs on a dedicated server farm and makes the final decision of storing or discarding events based on event topology. Furthermore, the physics trigger allows to classify events by category (hadronic events, low multiplicity...) which is used to restrict the collected data to the processes of interest.
In addition, because the PXD possesses a large ammount of pixels, it is impossible to perform its full readout for each event. A reduction of PXD data by a factor ten is needed before it is combined with other sub-systems. To do so, the HLT extrapolates information from the CDC and SVD to define regions of interest (ROIs) of the PXD, for which particle hits are read.

2.5 The Belle II Analysis Software Framework

The Belle II Analysis Software Framework (basf2) [71, 79] is developed and maintained by the Belle II collaboration to provide for the experiment software needs: online data processing (as with the HLT), offline reconstruction, physics analysis or detector studies. The framework consists of independant *modules* written in `C++` [80] or `python` [81], which are handled using python steering scripts where they are integrated sequencially in *paths*. The number, type and order of the modules used in such scripts depend on the task performed. All modules have access to the studied data through a common container: the *DataStore*. Additional data that are not event-based (calibration, specifics of sub-detectors, etc..) are stored in *conditions* and are accessed in a similar container called the *DBStore*.

Input and output data analysed with basf2 are usually stored using the ROOT `TTree` [82] format.

2.6 Simulation

Monte Carlo (MC) simulation is used to generate physics processes as well as the interaction between generated particles and the Belle II detector. Cross sections for the most important physics processes that can occur in e^+e^- at $\sqrt{s} = 10.58$ GeV are given in Table 2.1. Different *generators* are used to simulate base physics processes. `EvtGen 1.3` is used to generate B and D mesons decays into exlusive final states [83]. `PYTHIA 8.2` [84] models inclusive meson decay final states as well as continuum $q\bar{q}$ production. `KKMC 4.15` generates τ pair production while `TAUOLA` [85] is used to model τ decays. In addition, several generators are used specifically to simulate QED processes with high cross sections: `BABAYAGA` [86–90] for $e^+e^- \to e^+e^-(\gamma)$ and $e^+e^- \to \gamma\gamma(\gamma)$ and `AAFH` [91–93] for $e^+e^- \to e^+e^-e^+e^-$ and $e^+e^- \to e^+e^-\mu^+\mu^-$.
In addition, a specific generator, `SAD` [94] is used to generate beam background and produce simulated background hit files.
Finally, the Belle II detector and its interaction with generated particles are simulated using `Geant4` [95, 96].

2.7 Reconstruction

Reconstruction is the process through which the enormous amount of raw data collected independantly by the detectors is transformed into manageable physics information, in terms of quantity, quality and meaningfulness. After reconstruction, data are still at a very fundamental stage and can be studied for the benefits of specific physics analyses, though it does not require an expert knowledge of each subdetector to make sense.
Several algorithms are developed within the basf2 framework by groups working on each detector, these allow to use low-level objects (detector signal) to produce higher-level objects (ECL clusters, tracks, etc..).
The same reconstruction is applied to both collected raw data and simulation digitized data. For the latter, "true" generated information can be obtained to test the performance of reconstruction, although this is dependent on how well the process of interest is simulated.

2.7.1 Tracking

Tracking mostly consists in reconstructing the path taken by charged particles through the detector. The basic idea is to identify hits from the CDC and VXD generated by particles of interest amidst background hits and to establish a possible trajectory from a fit to the hit positions within the magnetic field.
Different *track finder* algorithms are used for the tracking detectors as they do not operate on the same principles nor scales. The common purpose of these algorithms

is to identify patterns in detector hits to create track candidates.
Firstly two track finders are used in conjugation to produce CDC-only track candidates. CDC track candidates are then linked to SVD clusters using a Combinatorial Kalman Filter (CKF). In parralel, tracks that did not reach the CDC due to their low momentum (and thus curvature) are reconstructed with the SVD track finder using a series of filters of increasing sophistication to avoid high combinatorics [97]. CDC and SVD track candidates are then combined and extrapolated to the PXD with another CKF. Finally, the track is fitted with the GENFIT2 package [98].
All tracks are fitted with different particle mass hypotheses (pion, kaon and proton) to estimate energy loss.

2.7.2 Charged particle identification

Efficient particle identification (PID) is crucial for physics analysis, which is why the Belle II detector has benefited from a signifcant upgrade to its PID system with regard to Belle. In addition to the designated detectors (TOP, ARICH), information from ionisation (dE/dx) measured in the CDC and SVD is used to identify charged particles. In addition, energy deposits in the ECL are used to identify electrons while the KLM helps to identify muons. Each detector provides a PID likelihood $\mathcal{L}_i^{det}$ for each charged particle hypothesis, which is computed independently. These likelihoods are then combined to produce an overall likelihood for each hypothesis i or j:

$$\mathcal{L}_i = \prod_{det} \mathcal{L}_i^{det} \tag{2.5}$$

This overall likelihood can then be used to compute global PID ratios:

$$PID_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j} \tag{2.6}$$

or binary PID ratios:

$$PID(i|j) = \frac{\mathcal{L}_i}{\mathcal{L}_i + \mathcal{L}_j} \tag{2.7}$$

These PID indicators can then be used in physics analyses.

2.7.3 Neutral particle identification

Neutral particles do not ionise materials they pass through, which means that the CDC and SVD cannot assist in their identification. Photons are identified using the ECL by designing a parameter describing the shower shape of ECL clusters that are not matched to any track. Neutral or charged hadron interactions with the ECL sometime create hadronic splitoffs, which can mimic photon signatures.
Neutral pions are reconstructed in the $\pi^0 \to \gamma\gamma$ channel using two photon candidates. For low energy (< 1 GeV) π^0, the two photons are usually separated enough for the ECL showers to not overlap. For pions with energies in the range [1 GeV, 2.5 GeV], the ECL showers overlap but can still be reconstructed as two separate

photons. For π^0 with higher energies, the two showers are usually reconstructed as a single photon candidate, however, the pion's energy can be estimated from the shower's second moment shape variable.

K_L^0 identification is done using information from the KLM and ECL. Several multivariate methods are used to determine if ECL or KLM clusters originate from a K_L^0. The variables used for this classification are related to kinematics and cluster shapes as well as the distance between clusters and the closest track and timing information.

Chapter 3

Improvement of the SVD cluster position resolution

Contents

In this chapter we describe how *clusters* are reconstructed from the information collected by silicon strips in the Silicon Vertex Detector (SVD) introduced in Subsection 2.3.2. Furthermore, the resolution on the cluster position is defined in Section 3.1 and the performances of the detector are estimated. Following observations of discrepancies between simulated and measured SVD spatial resolution, and in a general effort to better detector performances, we present an algorithmic method destined to refine the computation of the cluster position resolution. The novel method of *cluster unfolding* is devised to correct for a strip-charge sharing effect seen in recorded data and its effect on spatial resolution performances is estimated in Section 3.3.

3.1 Definition of the cluster position resolution

As described in Subsection 2.3.2, the SVD collects information from charged particles crossing detector sensitive volume. The objects used to estimate particle hit-position are called *clusters* and are built from strip information.
In order to be retained to build a cluster, strip signals need to verify:

$$SNR = \frac{S_i}{N_i} > 3 \tag{3.1}$$

Where S_i is the maximal signal height collected by the strip i and N_i is the strip electronic noise.

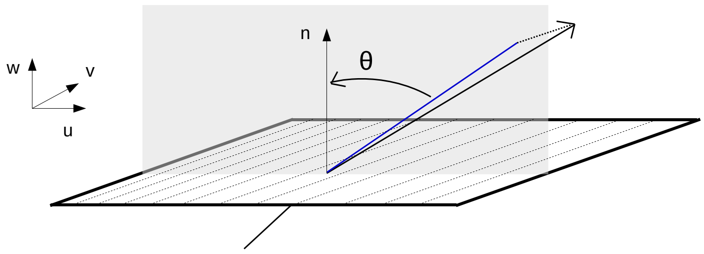


Figure 3.1: Schematic view of a track crossing a u/p-side SVD sensor. The strips of the sensors (dashed lines) are parallel to the v local direction. The blue line corresponds to the projection of the track on the ($\hat{\boldsymbol{u}}$,$\hat{\boldsymbol{w}}$) plane orthogonal to the strips. The incident track angle θ is the angle between the track projection in the ($\hat{\boldsymbol{u}}$,$\hat{\boldsymbol{w}}$) plane and the normal vector $\hat{\boldsymbol{n}}$, orthogonal to the ($\hat{\boldsymbol{u}}$,$\hat{\boldsymbol{v}}$) plane (sensor plane). The local v direction is parallel to the global Belle II ϕ coordinate, while the local coordinate u measures the global z direction.

.

A cluster can be constructed as a collection of any number of contiguous strips meeting this requirement in addition to requiring one strip (called *seed* strip) with $SNR_{seed} > 5$.
Basic cluster information can be further computed to be used in tracking. The cluster charge S_{CL} is defined as the sum of the individual charges of the strips making up the cluster:

$$S_{CL} = \sum_{i=0}^{i<size} S_i \tag{3.2}$$

The size of clusters depends mainly on the incident angle θ of particle trakcs (Figure 3.1). The cluster time t_{CL} is computed as the weighted average (center of gravity or CoG) of the strip times:

$$t_{CL} = \frac{\sum_{i=0} t_i \times S_i}{\sum_{i=0} S_i} \tag{3.3}$$

With t_i the time of the strip i. Finally, the cluster position x_{CL} is computed from the position of the individual strips with the same CoG method:

$$x_{CL} = \frac{\sum_{i=0} x_i \times S_i}{\sum_{i=0} S_i} \tag{3.4}$$

With x_i the local position of the strip i.

The cluster position is used by the tracking algorithm described in Subsection 2.7.1, making it a key component of Belle II physics performances. Because of that, performance studies on the cluster position resolution need to be performed regularly by the collaboration to ensure the quality of tracking.
In order to estimate the spatial resolution of the detector, the reconstructed cluster position should be compared to the true position of the particle crossing the detector. Of course this true position is not known, but it can be estimated by reconstructing the particle track and extrapolating it on the SVD sensor surface.
To this end, it is possible to compute, for each reconstructed track, the *unbiased track intercept position*, further used as the estimator of the true position of the studied cluster. Here, the track reconstructed by the track fitting algorithm using clusters from all SVD layers is re-fitted while excluding the cluster of interest. The position x_t at which this track crosses the studied cluster plane is the unbiased track intercept position, to which an error σ_t is associated. The distance between the measured cluster position x_{CL} and x_t is the *residual* ε_t. The cluster resolution σ_{CL} is given by:

$$\sigma_{CL} = \sqrt{< \varepsilon_t^2 - \sigma_t^2 >} \tag{3.5}$$

The cluster position resolution study is performed both on data and simulation. In the case of the latter, the *true* position x of clusters is also known, as well as the *true* cluster position residual $\varepsilon_m = x_{CL} - x$, the track true position t and true residual $\varepsilon_{true} = t - x$. The definition of these variables is shown in Figure 3.2.

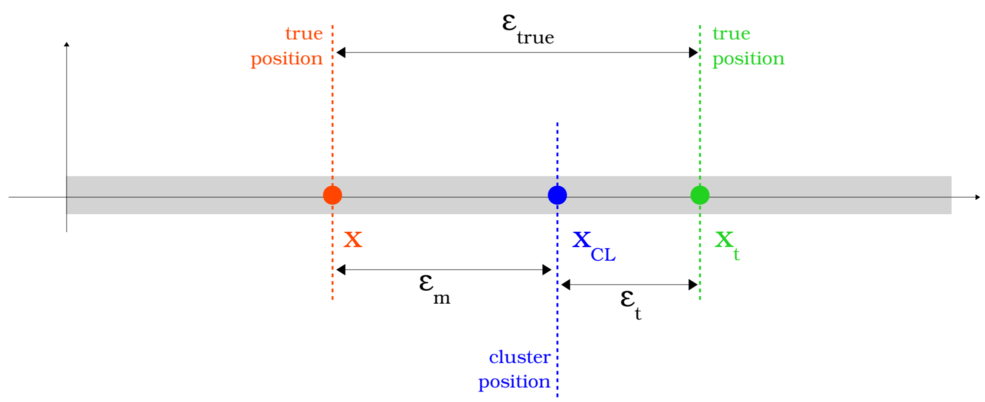


Figure 3.2: Schematic view of the main quantities used in the estimation of the spatial resolution of the SVD.

Example distributions of the measured residuals, true cluster and track residual, and track extrapolation error are shown in figure Figure 3.3 for layer 4 u/P clusters.

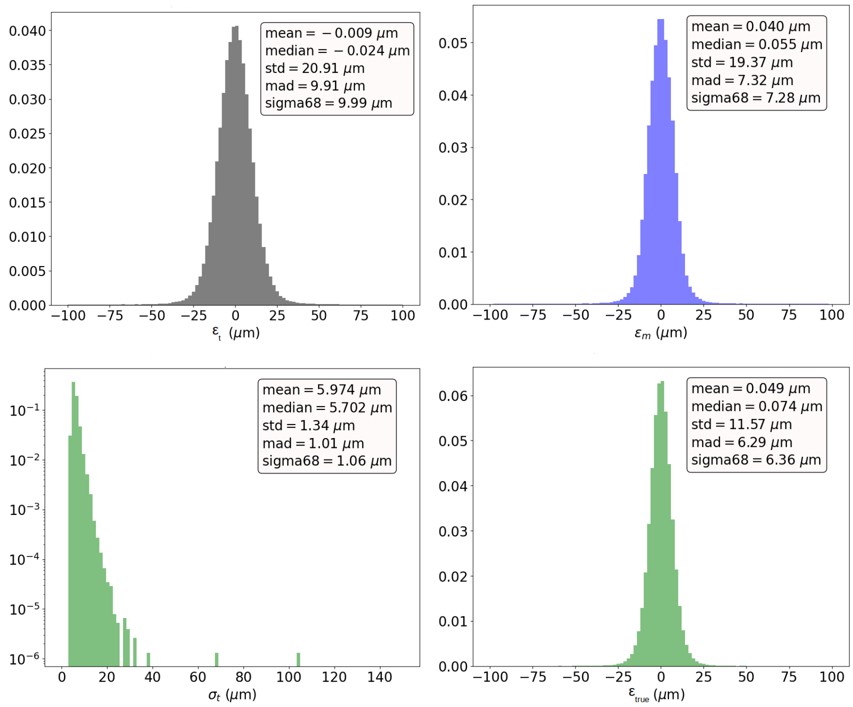


Figure 3.3: Distributions of the measured residuals ε_t (top left), true cluster residual ε_m (top right), track extrapolation error σ_t (bottom left) and true track residual ε_{true} (bottom right), for L4U clusters from simulated di-muon events. Adapted from Belle II's SVD group.

.

3.2 Data/simulation comparison

Figure 3.4 shows the resolution for the layer 3 for both detector sides as well as the sum of layer 4, 5 and 6 for both sides. The resolutions for data and simulation are computed as described in Section 3.1. We see that discrepancies in resolution appear between data and simulation. This trend is more pronounced for u/P sides than for v/N sides and is clearly noticeable for layer 3 in u/P side.
These discrepancies can be caused by several mechanisms. Firstly, the Belle II SVD simulation uses a simplified model of data collection and, for example, does not take into account effects described by the Shockley-Ramo theorem [99]. In addition, electronic effects within the detector may have not been identified during detector calibration and may thus not be simulated. This results in an optimistic simulation with regards to the estimation of SVD performances on position resolution.

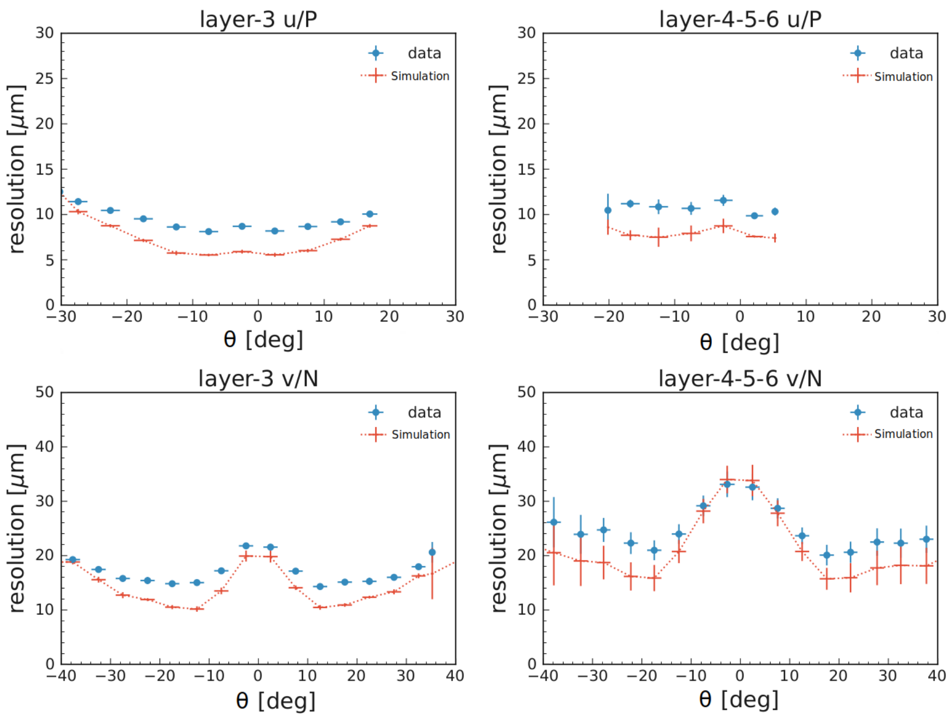


Figure 3.4: Comparison between data and simulation of the cluster position resolution as a function of the track incident angle θ.

However, the actual detector performances observed in data are extremely satisfactory and close to the expected digital resolution of the detector. It then seems relevant to point out that the discrepancies observed are expected to be the result of a combination of small mechanisms, which are not obvious to identify. Nevertheless, we try here to identify and correct for these effects, in order to deepen our knowledge of the detector and to try and reach optimal performances.

3.3 The Unfolding Method

When, during calibration runs, a charge is injected in one of the APV channels, a small signal $\simeq 5$ ADC count (here, ADC count refers to the output of an Analog to Digital Converter and is proportional to the deposited charge) is seen on the adjacent channel with a lower peaking time (by 7/8 APV clock $\simeq$ 27ns), showing a coupling between the two channels (Figure 3.5). This effect modifies the observed strip charge. Preliminary studies show that the observed adjacent strip charge could be underestimated by $\simeq 6\%$ of the seed strip charge.

Because the strip charge is used in the computation of the cluster position x_{CL} this might degrade the position resolution.

In order to correct for this effect, we propose a method aimed at unfolding the strip charges in a cluster by extending the coupling effect observed on APV channels to

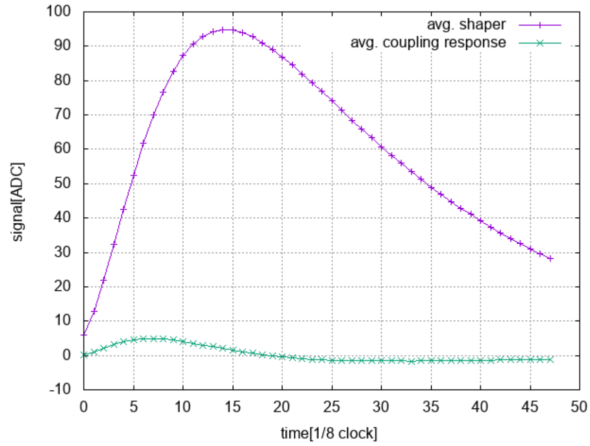


Figure 3.5: Response curve of an APV channel (purple) to the injection of a MIP-equivalent signal and from its adjacent channel (green) showing a coupling response. Taken from the Belle II SVD software group

a whole cluster.

3.3.1 Design of the Unfolding method

In order to model the impact of the APV coupling effect on the charge distribution in a cluster, we make the following hypotheses, also schematically explained in Figure 3.6:

1. Each strip in the cluster gives away $c \simeq 6\%$ of its collected charge to one neighbour on each side (for a total loss of 12% of the initial charge).

2. Edge strips lose $c \simeq 6\%$ of their charge by exchanging it with strips that do not pass the charge threshold to be included in the cluster. This charge is lost from the reconstructed cluster.

3. Edge strips do not gain charge from strips that do not pass the charge threshold.

4. These charge exchanges happen simultaneously.

To correct these effects and estimate their impact on resolution, the *true* strip charges have to be computed from the *observed* strip charges, then, the cluster position has to be computed and compared for both sets of charges using the CoG algorithm.

Because our hypothesis on the behavior of the edge strips, the total charge is not expected to be conserved between the *true* and *observed* clusters. In addition, both clusters are expected to have the same size.

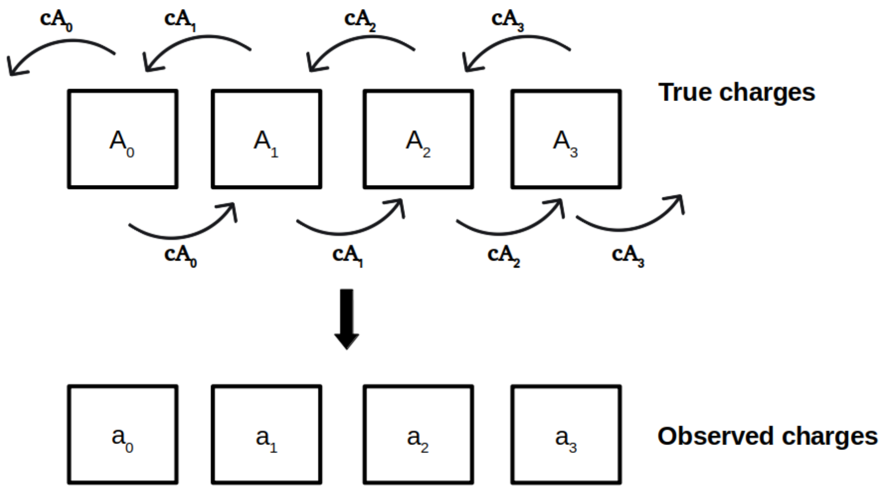


Figure 3.6: Relationship between the *real* strip charges A_i and the *observed* charges a_i, depending on the *unfolding coefficient* c. For the edge real charges (here A_0 and A_3), the outermost arrows represent lost charge.

To each *observed* cluster of size n, composed by the strips with charges a_i, $i \in (0; n-1)$, we want to associate the corresponding *true* cluster composed by the strips with charges A_i. We define the Unfolding Matrix M of size $n \times n$ such as:

$$\begin{cases} M_{ij} = 1 - 2c & \text{if } i = j; \\ M_{ij} = c & \text{if } \mid i - j \mid = 1; \\ M_{ij} = 0 & \text{for all others } (i, j); \end{cases} \tag{3.6}$$

With $i, j \in (0, n-1)$ and the *unfolding coefficient* $c = 0.06$ (corresponding to the expected 6% loss of charge for a given strip).
The *true* strip charges A_i are then computed as:

$$\begin{pmatrix} A_0 \\ A_1 \\ \dots \\ A_{n-1} \end{pmatrix} = M^{-1} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_{n-1} \end{pmatrix} \tag{3.7}$$

3.3.2 Implementation in the Belle II analysis software

In order to evaluate the effect of the unfolding method on the cluster position resolution, the method has been implemented in the Belle II analysis software. While strip charges are used in the reconstruction process at different stages (*e.g.* evaluation of the cluster time, see Equation 3.3), these processes give good results. Because the developement of the unfolding method is performance-motivated, we prefer not to

alter the computation methods giving satisfactory results. Thus, the scope of this implementation is to correct strip charges with the unfolding method and to use these corrected charges only in the computation of the cluster position.
The unfolding method is implemented by defining a new reconstruction function which takes a reconstructed cluster as argument and returns a cluster with the same attributes, except for the strip charges which are corrected as seen in Equation 3.7. The unfolding coefficient c is defined with a different value for u/P and v/N-side strips. The corrected strip charge is compared to a threshold T and set to 0 if its value is lower than T (so that the strip will not affect the CoG computation of the cluster position).
A threshold is already defined in `basf2` to discard noisy strips: a given strip S_i with strip noise N_i is discarded if its charge is below $3 \times N_i$, as seen in Equation 3.1, with the average noise being:

$$\begin{cases} L_3u: & 1100\ ADC; \\ L_{456}u: & 900\ ADC; \\ L_3v: & 900\ ADC; \\ L_{456}v: & 600\ ADC; \end{cases} \tag{3.8}$$

Ideally the *unfolding* threshold T should also be defined strip by strip. Here, two T values have been implemented: $T = 0\ ADC$ in order to discard negative (non physical) corrected strip charges, and $T = 3000\ ADC$ as a general value corresponding to $\simeq 3 \times N_i$ for any given strip.

3.3.3 Datasets

Several datasets have been used in the developement of the unfolding method. Two event topologies are studied:

- **di-muon samples**: these samples correspond to $e^+e^- \to \mu^+\mu^-$ events. These events are selected so that the two muon tracks have a transverse momentum $p_T > 1.0$ GeV/c, come from a region close to the interaction point and are of good quality with regards to the tracking (more than one hit in the PXD, 8 in the SVD and 30 in the CDC). Finally, only muon pairs with an invariant mass between 10 and 11 GeV/c^2.
These events consist solely of two clean and well-separated tracks, which allow to gauge the performances of the detector in an optimal scenario.

- **hadronic events**: these samples are selected so that at least three tracks come from the IP and verify $p_T > 0.2$ GeV/c^2 are kept. This loose selection allows to discard several high cross-section processes (bhabha scattering, 4-electrons production...) while retaining most hadronic events ($e^+e^- \to B\bar{B}/q\bar{q}$).
These events allow to estimate the detector performances in the physics analysis regime, where the conditions are less than ideal because of varying track

quality, higher impact of multiple scattering due to a broader particle momentum distribution, etc..

Furthermore, the samples used in the estimation of the unfolding method performances are splitted between recorded data and simulation.
For Data:

- The preliminary tests and the optimization of the (c, T) values have been performed on $\simeq 0.035\ \text{fb}^{-1}$ of data with both di-muon events and hadronic events. These have been selected amongst a sample of *good runs* for the SVD, corresponding to data taking periods for which the SVD data quality is known to be excellent.
  Datasets using each possible (c, T) couples have been produced for both sample types.
- Final results have been extracted from $\simeq 1\ \text{fb}^{-1}$ of data, using the same reconstruction on dimuon events.

For Simulation:

- Sets of 500k dimuon events (corresponding to $\sim 0.043\ \text{fb}^{-1}$) have been generated and reconstructed. Because the unfolding method is solely applied on recorded data, these samples have been used as a baseline to which the corrected datasets have been compared.

3.3.4 Effects on the position resolution

A full performance study has been performed, in order to assess the scale of the correction. The position resolution is first estimated for each (c, T) couple in dimuon events, as seen in Figure 3.7 and 3.8. A threshold value of 3000 ADC slightly worsens the resolution for every values of c in most cases. The same study has been performed on hadronic events with the same effect being observed (*cf.* Appendix A). Taking $T = 0$ thus seems a reasonable choice motivated both by detector performances and physical consideration. Indeed, further inspection validates that the majority of strip charges that would end up below 3000 ADC after the unfolding end up with negative (non-physical) charges and are cut away by a 0 ADC threshold.

Furthermore, several sensor types are used in the SVD as described in Table 2.2. In order to assess if a sensor-dependent c value is needed, the impact of the unfolding method on the spatial resolution has been studied for all sensor types for each c values (Figure 3.9). Finally, the effect of the correction has also been studied based on the angle between the tracks considered and the sensors (Appendix A).
The correction does not have a clear positive effect on V-side sensors. However, an improvement is seen on U-side sensors, for which the optimal c value varies between 0.05 and 0.15 depending on the incident angle between the track and the sensor. However, the ranking of performance gained from the different c values is not clear,

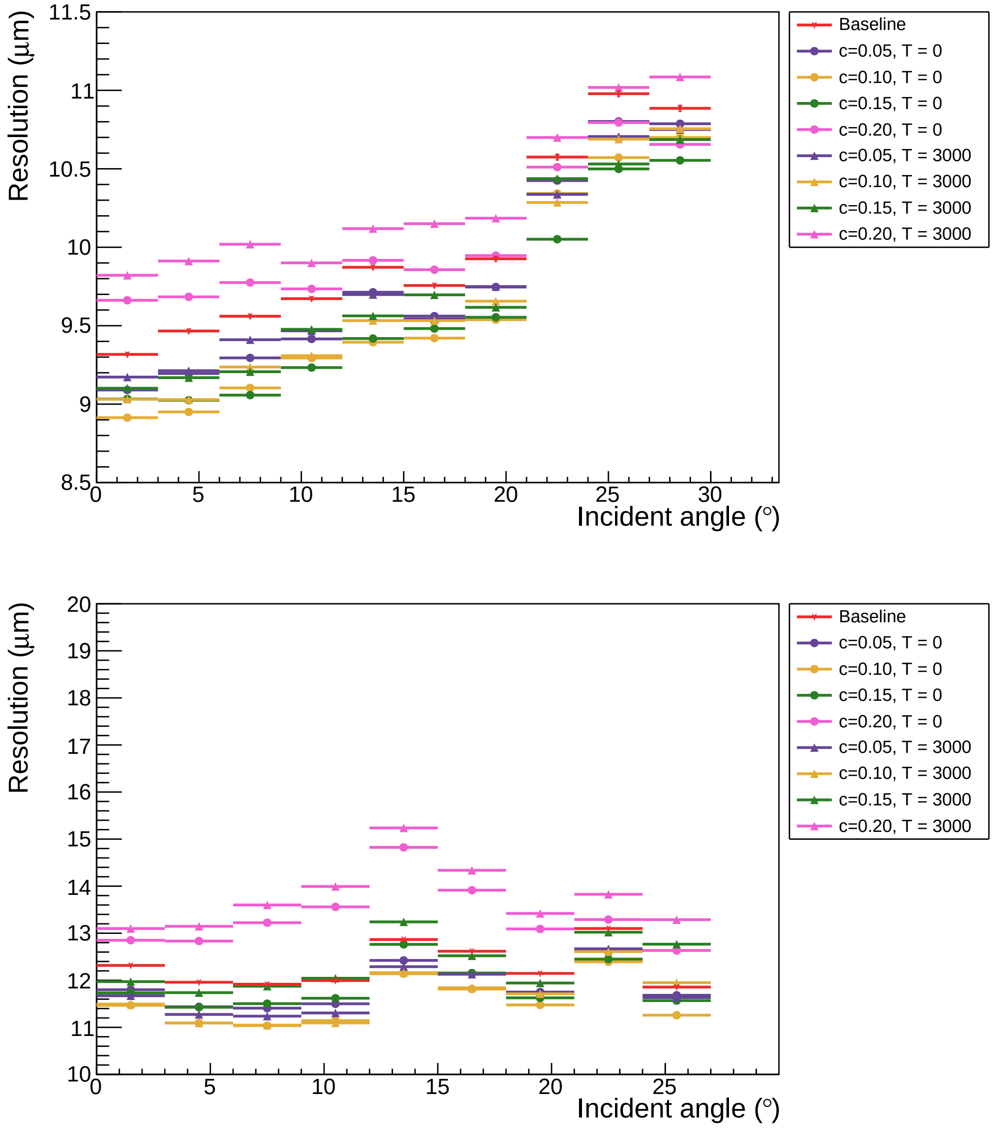
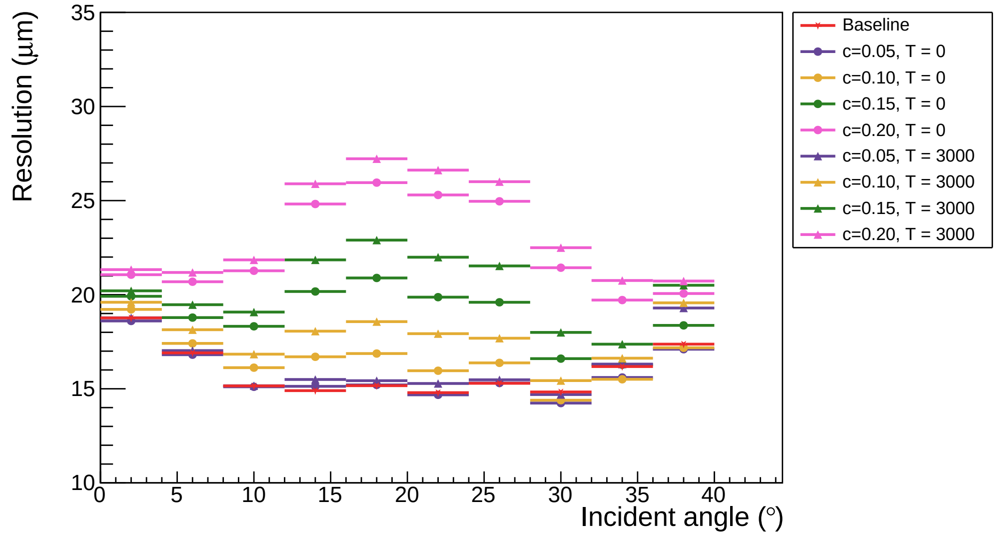


Figure 3.7: Cluster position resolution as a function of the incident angle of the track for all (c,T) couples. Each color corresponds to a given c value, circle markers correspond to $T = 0$ ADC and triangle markers correspond to $T = 3000$ ADC. The red points correspond to the baseline (*i.e* no correction applied). For the Layer 3 u/P-side (top) and Layer 4,5 and 6 u/P-side (bottom).



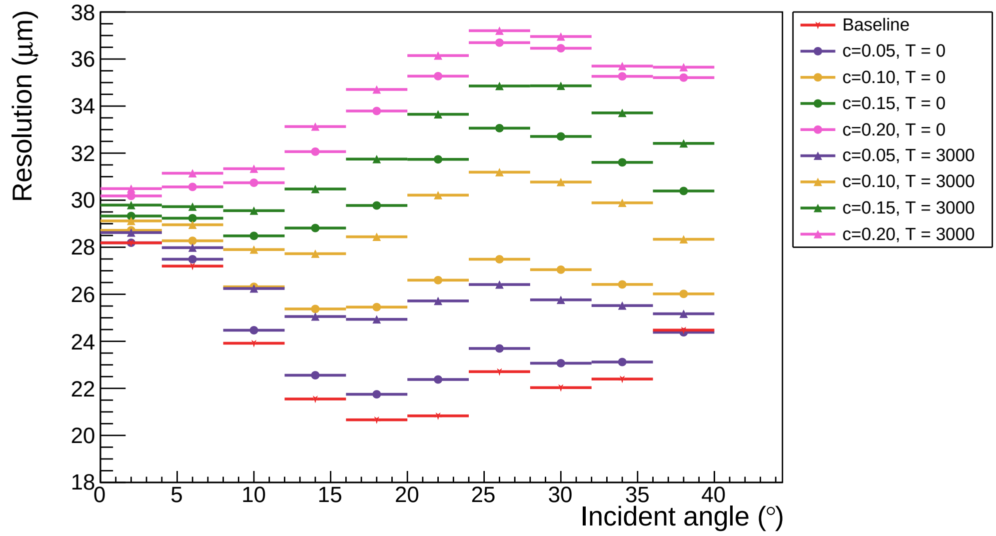


Figure 3.8: Cluster position resolution as a function of the incident angle of the track for all (c,T) couples. Each color corresponds to a given c value, circle markers correspond to $T = 0$ ADC and triangle markers correspond to $T = 3000$ ADC. The red points correspond to the baseline (*i.e* no correction applied). For the Layer 3 v/N-side (top) and Layer 4,5 and 6 v/N-side (bottom).

due to statistical fluctuations. When taking the cluster position resolution for a given sensor type averaged over all incident track angles (Table 3.1), $c = 0.1$ always leads to the best results.

This value is close to the estimated effect ($\simeq 6\%$ of the seed strip charge) of the observed APV channels cross talk. The fact that the optimal value observed is slightly higher than the expected one could be explained by other processes that have yet to be identified but end up being (partially) corrected by the unfolding method.

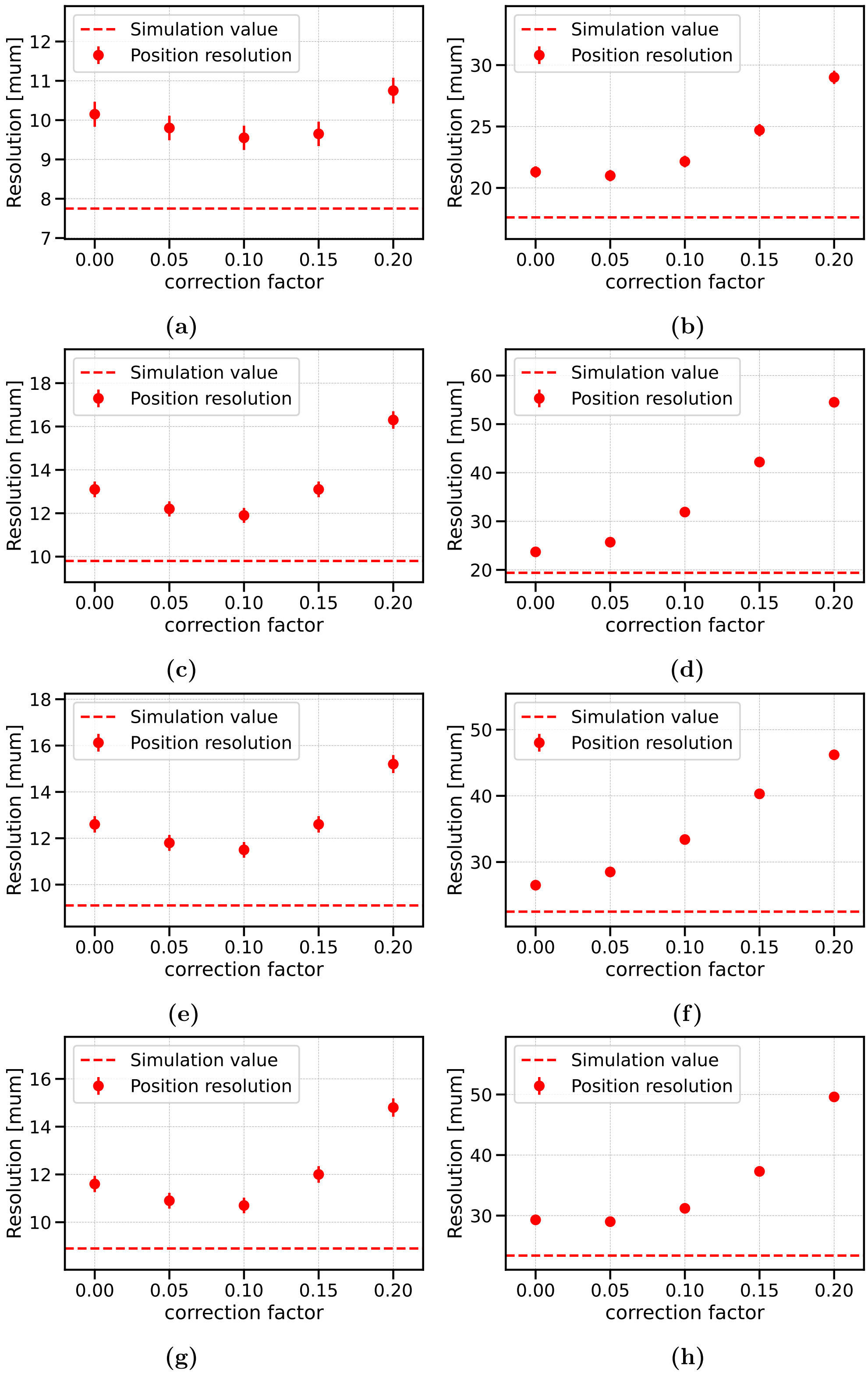


Figure 3.9: Averaged cluster position resolution depending on the value of the unfolding coefficient c for the Layer 3 u/P-side (a) and v/N-side (b) and Layer 4, 5 and 6 u/P-side backward sensors (e), v/N-side backward sensors (f), Layer 4, 5 and 6 u/P-side backward sensors (g), v/N-side backward sensors (h). The dashed red line corresponds to the position resolution computed in the simulation.

3.4 Conclusion

All things considered, applying the unfolding method on all the clusters with an unfolding coefficient of 0.1 and a threshold of 0 ADC allows to improve the overall cluster position resolution of u-side sensors by 5% to 15%, depending on the sensor type. Because this effect is not simulated, only collected data is corrected by the method, which subsequently reduces the disagreement on cluster position resolution seen between data and simulation (Figure 3.10).

Sensors - u-side	c = 0	c = 0.05	c = 0.1	c = 0.15	c = 0.20
L3.1	10.7	10.2	10	10.4	12.3
L3.2	11.8	11.4	11.1	11.1	12.3
L456 backward	14.9	14	13.2	14.9	18.6
L456 origami	15.7	14.9	14.5	15.3	18.6
l456 slanted	12.7	12.2	12	13.3	16.2
Sensors - v-side	**c = 0**	**c = 0.05**	**c = 0.1**	**c = 0.15**	**c = 0.20**
L3.1	25.1	24.5	24.8	25.6	27.8
L3.2	17.5	17.5	19.5	23.8	30.2
L456 backward	23.7	25.7	31.9	42.2	54.5
L456 origami	26.5	28.5	33.4	40.3	46.2
l456 slanted	29.3	29	31.2	37.3	49.6

Table 3.1: Averaged cluster position resolution (in μm) estimated for each type of senor for different values of c.

The evolution with time and instantaneous luminosity conditions of the effect studied here and its correction is not yet known, thus this study will need to be conducted again in the future in order to ensure an optimal correction to the cluster position resolution.

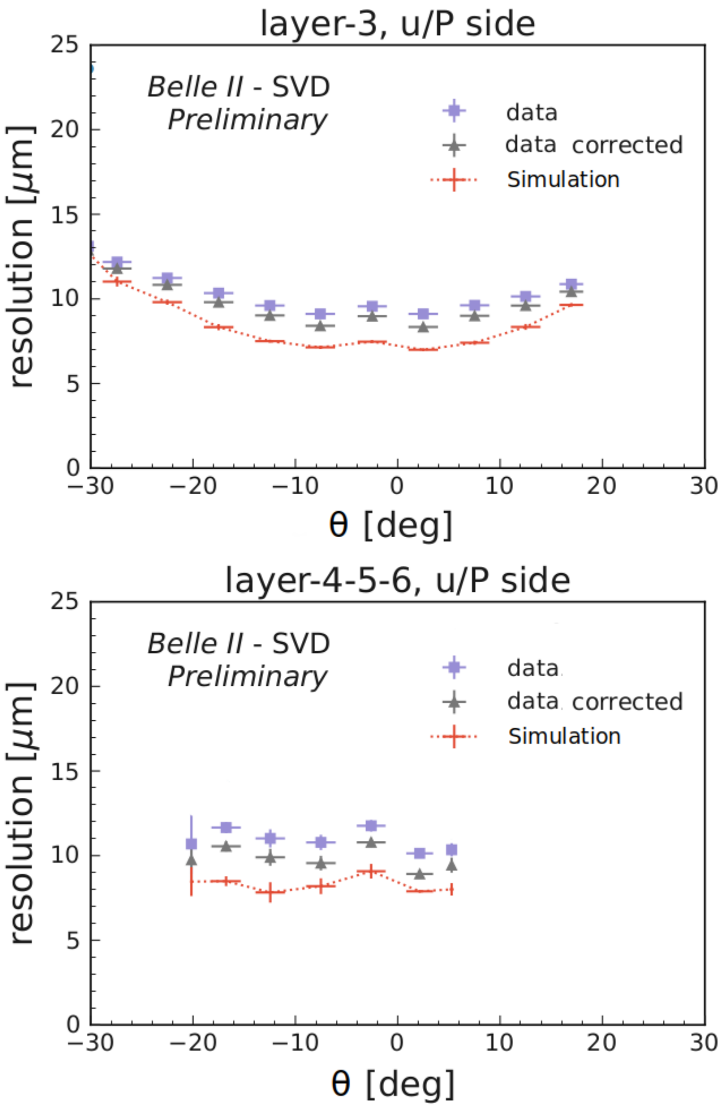


Figure 3.10: Cluster position resolution as a function of the track incident angle showing the effect of the unfolding method on recorded data. For layer 3 u-side (top) and layer 4, 5 and 6 u-side (bottom).

Chapter 4

Analysis tools and methods

Contents

The different analysis techniques and tools used in this work are reported in this chapter. The next sections are rather independent as they treat of various subjects. Section 4.1 provides a description of the algorithm used to perform B-meson tagging in the Belle II experiment while Section 4.2 consists in a brief overview of binary classification. Section 4.3 presents a figure of merit used in our search for the $B^+ \to K^+\nu\bar{\nu}$ decay, adapted from the work of G. Punzi [100]. Section 4.4 and 4.5 describe the statistical tools used to extract $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ from observations, as well as the way experimental uncertainties are propagated to the final measurement. In the absence of clear signal observation, Section 4.6 shows how an upper limit on the value of the branching fraction can be computed. Finally, Section 4.7 introduces the concept of *blind* analyses and the reasons to proceed in such a manner.

Because of the technicality and variety of subjects found in this chapter, the reader may skip it and come back to it when specific topics are referenced in Chapter 5.

4.1 The Full Event Interpretation algorithm

This analysis makes use of the Belle II-developed Full Event Interpretation (FEI) algorithm [101]. The FEI is a hierarchical reconstruction algorithm estimating the

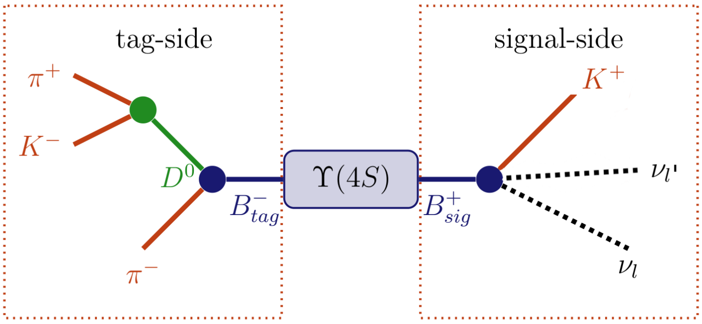


Figure 4.1: Schematic view of the $\Upsilon(4S)$ decay showing (left) a generic *tag-side* and (right) the *signal-side* $B^+ \to K^+\nu\bar{\nu}$ decay. It is important to note that this separation is only conceptual and that the tracks coming from both sides overlap spatially in the detector. Adapted from [101].

most probable decays of B mesons in $\Upsilon(4S) \to B\bar{B}$ events based on detector information.

This algorithm has been specifically developed to help the study of B meson decays with indetectable final state particles, such as $B \to D^*\ell\nu$ and $B^+ \to K^+\nu\bar{\nu}$. The $\Upsilon(4S)$ decay can be split into two conceptual sides. The *signal-side* corresponds to the tracks and calorimeter clusters compatible with the decay of interest. The *tag-side* contains the remaining objects in the event, compatible with any decay of the B-meson. The B-meson associated to each side are labeled B_{sig} and B_{tag} respectively. Figure 4.1 illustrates this concept.

First, tracks, displaced vertices (*i.e.* sets of tracks not originating from the interaction point) and calorimeter clusters of an event are identified. These objects are combined to reconstruct the final state particles of the event ($e^\pm, \mu^\pm, \pi^\pm, K^\pm, p^\pm, n, \gamma$ and K^0_L). Afterwards, these final states particles are combined to form intermediate particles ($\pi^0, D^{\pm/0}, J/_\psi, K^0_S$, $D^{*\pm/0}$ and baryons). Latter stages of the reconstruction allow to combine previously reconstructed particles to form heavier intermediate particles. The last stage of the reconstruction combines intermediate and final state particles into B-mesons.

For each step of this procedure, the probability of the reconstructed particle (and its associated decay chain) is estimated using a multivariate classifier trained on simulated events using several features (vertex position, particle four-momentum, etc..). The output of said classifier is called $\mathcal{P}_{FEI}$ and can be interpreted as a probability of correct identification. This reconstruction process is illustrated in Figure 4.2.

The FEI is an *exclusive* tagging algorithm, meaning that it reconstructs particles (in this case B_{tag}) through explicit decay channels. Taking into account all intermediate particle decays implemented in the FEI, the algorithm can reconstruct $\mathcal{O}(10000)$ different decay chains. For our analysis, it provides B^+ mesons in 36 hadronic modes. The different modes are shown in Table 4.1.

The FEI tag-side efficiency for fully hadronic B^+ reconstruction is $\simeq 0.66\%$, includ-

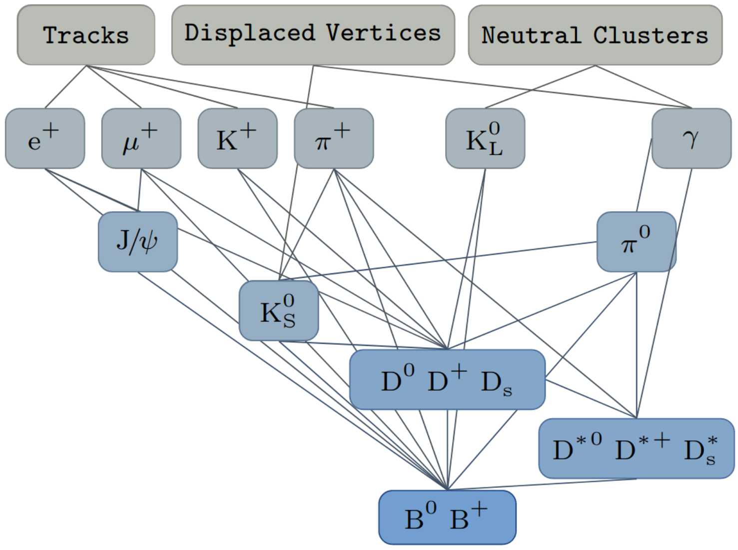


Figure 4.2: Conceptual overview of the FEI algorithm reconstruction steps. The objects in gray boxes correspond to objects built by the Belle II reconstruction software. Taken from [101].

ing branching fractions and reconstruction efficiency.

4.2 Binary classification

We discuss techniques used in this work to classify the events studied. Our goal is to separate **signal** (events where a $B^+ \to K^+\nu\bar{\nu}$ decay is present) from **background** (all other events). Several approaches can be adopted to do so, resulting in different efficiencies in the classification and purities.
We present here two algorithms used to perform this task: the decision tree and the boosted decision tree. Similarly to a cut-based selection, these methods extract information from a set of discriminative variables to classify events in the defined classes. However, many events do not exhibit *all* characteristics of either classes. These methods allow to keep events rejected by a criterion and check if other criteria allow to classify them properly.

4.2.1 Decision tree

Considering a set of N_v explanatory variables, a binary decision tree recursively splits the N_v-dimensional variable space based on binary selections. A first node divides the space into two subspaces based on a splitting value defined for a specific variable. The splitting value is chosen to maximize the separation (*i.e.* keeps mostly signal in one branch, mostly background in the other). This optimization is done

	B^+ decay modes
1	$B^+ \to \bar{D}^0\pi^+$
2	$B^+ \to \bar{D}^0\pi^+\pi^0$
3	$B^+ \to \bar{D}^0\pi^+\pi^0\pi^0$
4	$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-$
5	$B^+ \to \bar{D}^0\pi^+\pi^+\pi^-\pi^0$
6	$B^+ \to \bar{D}^0 D^+$
7	$B^+ \to \bar{D}^0 D^+ K_S^0$
8	$B^+ \to \bar{D}^{0*} D^+$
9	$B^+ \to \bar{D}^0 D^{+*} K_S^0$
10	$B^+ \to \bar{D}^{0*} D^{+*} K_S^0$
11	$B^+ \to \bar{D}^0 D^0 K^+$
12	$B^+ \to \bar{D}^{0*} D^0 K^+$
13	$B^+ \to \bar{D}^0 D^{0*} K^+$
14	$B^+ \to \bar{D}^{0*} D^{0*} K^+$
15	$B^+ \to \bar{D}_s^+ \bar{D}^0$
16	$B^+ \to \bar{D}^{0*}\pi^+$
17	$B^+ \to \bar{D}^{0*}\pi^+\pi^0$
18	$B^+ \to \bar{D}^{0*}\pi^+\pi^0\pi^0$
19	$B^+ \to \bar{D}^{0*}\pi^+\pi^+\pi^-$
20	$B^+ \to \bar{D}^{0*}\pi^+\pi^+\pi^-\pi^0$
21	$B^+ \to \bar{D}_s^{+*} \bar{D}^0$
22	$B^+ \to \bar{D}_s^+ \bar{D}^{0*}$
23	$B^+ \to \bar{D}^0 K^+$
24	$B^+ \to D^-\pi^+\pi^+$
25	$B^+ \to D^-\pi^+\pi^+\pi^0$
26	$B^+ \to J/_\psi K^+$
27	$B^+ \to J/_\psi K^+\pi^+\pi^-$
28	$B^+ \to J/_\psi K^+\pi^0$
29	$B^+ \to J/_\psi K_S^0\pi^+$
30	$B^+ \to \Lambda_c^- p\pi^+\pi^0$
31	$B^+ \to \Lambda_c^- p\pi^+\pi^+\pi^-$
32	$B^+ \to \bar{D}^0 p\bar{p}\pi^+$
33	$B^+ \to \bar{D}^{0*} p\bar{p}\pi^+$
34	$B^+ \to D^+ p\bar{p}\pi^+\pi^-$
35	$B^+ \to D^{+*} p\bar{p}\pi^+\pi^-$
36	$B^+ \to \Lambda_c^- p\pi^+$

Table 4.1: List of the hadronic B^+ meson decay modes reconstructed by the FEI algorithm and used in our analysis.

by evaluating a loss function, here the cross-entropy:

$$\mathcal{L}(y, \hat{y}) = -\left[y\, log\, \hat{y} + (1-y)\, log(1-\hat{y})\right], \tag{4.1}$$

Where $y \in \{0, 1\}$ is the target class (background = 0, signal = 1) and $\hat{y} \in (0, 1)$ is a prediction probability. This is repeated for following nodes, until reaching the final nodes, called leaves. Leaves correspond to a specific region of the variable space (defined by a succession of nodes, called branches) and are assigned weights. A negative weight corresponds to a background favoured prediction while a positive weight corresponds to a signal favoured prediction.
To a given observation $x \in \mathbb{R}^{N_v}$, a decision tree m assigns a weight $w(x) \in \mathbb{R}$. The corresponding prediction probability $\hat{y}(x)$ is then computed as:

$$\hat{y}(x) = P(w_m(x)) = \frac{1}{1+e^{-w_m(x)}}, \tag{4.2}$$

Decision trees prove to be useful tools to devise finer classifications (compared to cut-based techniques, of which they are a sequential generalization) and have the advantage of being easily interpreted as a set of boolean (here physics-based) decisions. However, they show high variance, as small changes in sample can greatly influence the output. Usually the classification power of a single decision tree can only marginaly surpass that of random guesses.

4.2.2 Gradient-boosted decision tree

The issues linked to the use of a single decision tree can be addressed by employing *Boosted Decision Trees* (BDTs). BDTs are ensembles of decision trees, allowing to combine the output of the different trees to enhance the overall classification performances. For a given observation $x \in \mathbb{R}^{N_v}$ and a set of N_t decision trees, a given tree assigns a weight $w_i(x) \in \mathbb{R}$ to x. The weights of all trees in the ensemble can then be summed to define a global weight $W(x)$:

$$W(x) = \sum_{i=1}^{N_t} w_i(x), \tag{4.3}$$

with an associated global prediction probability $\hat{y}_g$ given by:

$$\hat{y}_g = P(W(x)), \tag{4.4}$$

where P is defined in Equation 4.2. To train a BDT, an initial weight $w^0(x) = 0$ is applied to all x. Each decision tree in the ensemble is then trained, iteratively solving:

$$w_m(x) = \underset{w(x)}{\arg\min}\{\sum_{i=1}^{N_t} \mathcal{L}[y_i, P\left(w_{m-1}(x_i) + w(x)\right)] + \Omega(w_m)\}, \tag{4.5}$$

where w_{m-1} corresponds to the sum of the weights up to the previous iteration, $\mathcal{L}$ is the loss function defined in Equation 4.1 and $\Omega(w_m)$ is a regularization term

penalizing complexity in the model, which helps prevent overfitting.
One way to solve Equation 4.5 is by computing the gradient of the loss function. Thus, this variety of models are called gradient-boosted decision trees.
The analysis presented in Chapter 5 makes use of a gradient-boosted decision tree algorithm, `XGBoost` [102].

4.2.3 Variable importance

Boosted decision trees are usually resistant to correlations amongst the explanatory variables. They are also insensitive to variable duplicates and noise coming from irrelevant variables. However, it is usually best to use as few features as possible, in order to save computing time, mitigate the risk of variable simulation issues (since it is trained on simulated data) and to facilitate the interpretation of the models.
In order to identify a reasonable set of input features, it is possible to rely on the relative importance of the variables. To quantify this, we can define the gain provided by a tree node as the quantity by which the objective function (Equation 4.5) is modified by said node. The importance of a given variable v can then be defined as the sum of the gains across all nodes featuring v, normalised by the total gain:

$$I(v) = \frac{\sum\limits_{i \in S^0} Gain(i)}{\sum\limits_{j \in S} Gain(j)}, \tag{4.6}$$

with $I(v)$ the relative importance of v, S^0 the set of nodes featuring v and S the set of all nodes present in the tree.
Still, the relative importance of variables is difficult to assess. A potential shortcoming comes from *variable masking* [103]: considering two variables v_1 and v_2, the way $I(v_2)$ is estimated in Equation 4.6 depends on the number of nodes featuring v_2. However, if v_2 is only slightly less discriminative than v_1, it ends up featured in fewer nodes and is then considered as irrelevant. However, removing v_1 from the features set renders v_2 very relevant.
A possible way to identify an optimal set of variables is to start with a set of n variables, train the model with all $n-1$ combinations and pick the combination with the best performances and repeat it. This allows to identify which variables have the largest effect on the classifier performance.

4.2.4 k-folding

K-folding is a form of cross-validation used to evaluate classifier's ability to adapt to new data. In the case of particle physics analyses using classifiers, it can prove useful to make the most out of a limited dataset. K-folding and other cross-validation methods allow to gauge the overfitting of a classifier. Overfitting corresponds to the dependence of the classifier on the data on which it is trained and is illustrated in Figure 4.3.
Considering a dataset $\mathcal{L}$ on which to train a classifier, k-folding validation consists

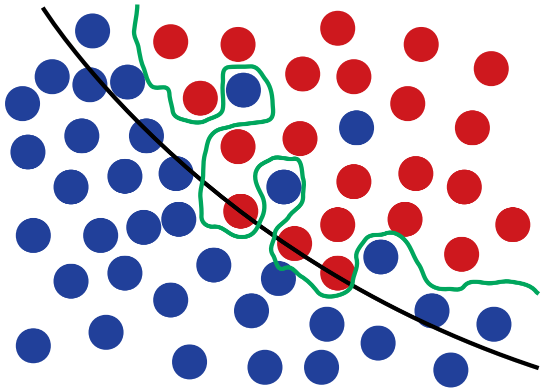


Figure 4.3: Schematic view of the effects of overfitting. Suppose a classifier trained to separate two classes (red/blue dots), the black line corresponds to a generalized model, which would perform adequatly on a different dataset. The green line corresponds to an overfitted model, which, even though giving a better separation power on the training data, is too reliant on that dataset and would likely show a worse separation power on a new dataset.

in splitting said dataset in k equal sized subsamples $\mathcal{L}_i$ such that $\mathcal{L} = \bigcup_{i=1}^{k} \mathcal{L}_i$. Of these subsamples, $k-1$ are used to train the model while the remaining one is used for testing. This is done k times, changing the training sample each time. In the end, the k training results can be averaged.

4.3 Modified Punzi figure of merit

In [100], the computation of a figure of merit for optimizing a Poisson distributed event counting experiment is described. A *sensitivity region* is defined for a given confidence level CL:

$$1 - \beta_\alpha(\mu_{sens}) > CL, \tag{4.7}$$

as the region of parameters for which the experiment is sensitive, with α the significance of the test and β the probability of rejecting the signal strength μ_{sens} with the given confidence level. The definition of this sensitivity region means that the experiment is expected to lead to a discovery with a probability greater than CL with significance α and can at least exclude the entire region in case the observed number is the maximum that does not allow to observe the signal with significance α.

In the original case of a counting event, the sensitivity region can be defined by the number of signal events:

$$S_{sens} = a\sqrt{B} + b\sqrt{B + S_{sens}}. \tag{4.8}$$

With B the number of background events and a and b the number of standard deviation corresponding to one-sided Gaussian tests at significance α and β respectively. Solving for S_{sens} gives a figure of merit which one can minimize to find the best selection for a counting experiment.
Here we propose a modified version of this figure of merit applicable to our analysis. Considering the histograms B_i and S_i with the background and signal event distribution, we expect in each bin:

$$N_i = B_i + \mu' S_i. \tag{4.9}$$

Where μ' is the true value of the signal strength μ. In each bin we can estimate $\hat{\mu}_i = N_i - B_i/S_i$ with an uncertainty $\sigma_{\hat{\mu}_i} = \sqrt{N_i}/S_i$ (Gaussian-Poisson approximation). By averaging the $\hat{\mu}_i$ using as weights the inverse of their squared uncertainties, we get

$$\hat{\mu} = \mu', \quad \sigma_{\hat{\mu}} = \frac{1}{\sqrt{\Sigma \frac{1}{\sigma^2_{\hat{\mu}_i}}}} = \frac{1}{\sqrt{\Sigma \frac{S_i^2}{B_i + \mu' S_i}}}. \tag{4.10}$$

This allows us to define the sensitivity region as:

$$\mu_{sens} = a\sigma_0 + b\sigma_{\mu_{sens}} \quad \text{with} \;\; \sigma_0 = \frac{1}{\sqrt{\Sigma \frac{S_i^2}{B_i}}}, \quad \sigma_{sens} = \frac{1}{\sqrt{\Sigma \frac{S_i^2}{B_i + \mu_{sens} S_i}}}. \tag{4.11}$$

It is then possible to numerically solve for μ_{sens} and minimize it to optimize our selection.

4.4 Binned maximum-likelihood fit

We aim at measuring the value of $\mathcal{B}(B^+ \to K^+ \nu\bar{\nu})$, being motivated in part by the search for beyond Standard Model physics, as mentionned in Section 1.4. We define the *signal strength* μ as:

$$\mu = \frac{\mathcal{B}(B^+ \to K^+ \nu\bar{\nu})}{\mathcal{B}(B^+ \to K^+ \nu\bar{\nu})_{SM}}, \tag{4.12}$$

which is the ratio of the measured branching fraction and the value predicted in the SM.
As described in Section 5.5, we base our measurement on the observed binned distribution of a classifier of data events. In order to estimate μ, we propose to perform a binned maximum-likelihood fit to this distribution. The method is discussed at length in [104, 105] and summarized below.
For a set of N_b bins counting events after a given selection, the expected number of events $\nu_1, ..., \nu_{Nb}$ in each bin is estimated from simulation for each type of contributions from several event types, one signal sample and $n \geq 1$ background sources:

$$\nu_b(\mu, \boldsymbol{\theta}) = \sum_{t \in \{\text{event types}\}} \nu_{b,t}(\mu, \boldsymbol{\theta}), \tag{4.13}$$

where $\nu_{b,t}$ is the expected number of events in bin b for the event category sample t and $\boldsymbol{\theta}$ is a vector of N nuisance parameters which may impact the base expectations. Assuming $n \geq 1$ background sources, $\boldsymbol{\theta}$ contains n nuisance parameters $\mu_1, ..., \mu_n$, and $N - n$ additional nuisance parameters such that:

$$\boldsymbol{\theta} = \left(\mu_1, ..., \mu_n, \theta_{N-n}, ..., \theta_N\right)^T , \tag{4.14}$$

the normalisation parameters $\mu_i, i \in \{1, ..., n\}$ are voluntarily named similarly to the signal strength μ, as each μ_i corresponds to a given background strength. We can then develop Equation 4.13 as:

$$\nu_b(\mu, \boldsymbol{\theta}) = \sum_{t \in \{\text{event types}\}} \mu_t \left(\nu^0_{b,t}, \Delta_{b,t}(\boldsymbol{\theta})\right) , \tag{4.15}$$

with $\nu^0_{b,t}$ the nominal number of expected events in bin b for the event type t and μ_t is the normalisation parameter associated to the event type t (kept at the same value for all bins). $\Delta_{b,t}(\boldsymbol{\theta})$ is an additive variation in the bin b for the sample t such as:

$$\Delta_{b,t}(\boldsymbol{\theta}) = \sum_{i=N-n+1}^{N} \theta_i \delta^i_{b,t}, \tag{4.16}$$

where $\delta^i_{b,t}$ is an additive variation for the bin b and the sample of event type t. This variation is modulated by the nuisance parameter θ_i. The set of $\delta^i_{b,t}$ is an input of the model, describing the systematic uncertainties. If for a given θ_i one has $\delta^i_{b,t} \neq 0$ for multiples bins b or background samples t, then the $\delta^i_{b,t}$ describe uncertainties correlated among the bins or the samples and are then interpreted as components of a variation vector of correlated uncertainties. The following cases arise:

- Uncertainties are uncorrelated: one θ_i is associated to each bin and sample,
- Uncertainties are bin-correlated: one θ_i is associated to each sample,
- Uncertainties are sample-correlated: only one θ_i is defined for all contributions.

From these cases, we define the following uncertainty categories:

- **Normalization**: the parameters cause a global scale variation on all bins. The effect is different and uncorrelated for the different components;
- **Normalization-correlated**: the parameters induce a global scale variation on all bins, with correlation among components;
- **Bin-correlated**: the parameters cause correlated bin-by-bin variations for each component, with no correlation among components;
- **Component-correlated**: the parameters create correlated bin-by-bin variation on each component, with correlation among components;

- **Uncorrelated**: the parameters cause totally uncorrelated bin-by-bin variation on each component.

Given the same set of N_b bins in which $n_1, ..., n_{Nb}$ data events are observed, we can now model the likelihood of the observation as:

$$\mathcal{L}(\mu, \boldsymbol{\theta}|n_1, ..., n_{Nb}) = \frac{1}{Z} \prod_{b \in \{bins\}} Pois\left(n_b|\nu_b(\mu, \boldsymbol{\theta})\right) p(\boldsymbol{\theta}), \tag{4.17}$$

where Z is a simple normalization parameter (having no impact on the fit), $Pois\left(n_b|\nu_b(\mu, \boldsymbol{\theta})\right)$ corresponds to the Poisson density function with expectation $\nu_b(\mu, \boldsymbol{\theta})$ evaluated at n_b and $p(\boldsymbol{\theta})$ is the prior probability given to the different nuisance parameters.
Said prior probability contains information on how the systematic uncertainties are modelled. It is the product of several Gaussian densities centered at unity for the normalisation variations and at zero for the additive variations:

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{n} Gauss\left(\theta_i|1, \sigma^2_{norm,i}\right) \prod_{j=N-n+1}^{N} Gauss\left(\theta_j|0, 1\right), \tag{4.18}$$

where $Gauss(x|m, \sigma^2)$ is the Gaussian density with expectation m and variance σ^2. The background normalization uncertainties $\sigma_{norm,i}$ are inputs of the model, similarly to the $\delta^i_{b,t}$ factors seen in Equation 4.16. We see that the parameter of interest μ is not present in Equation 4.18. This is because μ is unconstrained, meaning that its prior distribution is uniform.
The parameter of interest μ is finally extracted from data by maximizing the likelihood function defined in Equation 4.17. In our analysis, a software package called pure-python HistFactory (`pyhf` [105]) is used to implement this method as well as the statistical model.

4.5 Propagation of uncertainties

As with any measurement, the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ measured here is expected to be given with associated uncertainties. Several systematic uncertainty contributions in our analysis come from pre-existing measurements (*e.g* branching ratio values of B mesons decays used for the simulation) whose uncertainties need to be propagated to the satistical model described in Section 4.4. We describe here a general method to do so, varying input values based on their respective uncertainties and transforming this information to feed it to our statistical model.

4.5.1 Toy simulation

Toy simulations are used to estimate the propagation of uncertainties on an event-by-event basis. It consists of building a set of replicas created for each event considered. For each replica, a weight associated to the considered uncertainty source is computed. Considering a quantity of interest θ with an associated uncertainty σ, we create for each event e a set of N replicas. To each replica $r \in \{1, ..., N\}$ we associate a modified value θ_r and a weight $w_r^e(\theta)$ such that:

$$\theta_r^e = \theta^e + \mathcal{N}(0, \sigma), \tag{4.19}$$

$$w_r^e = \frac{\theta_r^e}{\theta^e}. \tag{4.20}$$

Here, we make the hypothesis that the uncertainty follows a gaussian distribution $\mathcal{N}(0, \sigma)$.

The bins of the statistical model are then filled appropriately with the replica, based on the bin value and category associated with the event e. Sums of weights S_i^r are computed for the different replica, with i corresponding to the fit contribution category.

4.5.2 Estimation of the covariance matrix

Using the sums of weights S_i^r, we can define a covariance matrix as:

$$C_{ij} = \sum_r^{N_r} \frac{\left(S_i^r - \bar{S}_i\right)\left(S_j^r - \bar{S}_j\right)}{N_r}, \tag{4.21}$$

with the corresponding correlations:

$$\rho_{ij} = \frac{C_{ij}}{C_{ii} \cdot C_{jj}}, \tag{4.22}$$

where $\bar{S}_i$ is the average over all replicas of the sums of weights for a given bin i. The covariance matrix is an $m \times m$ matrix, with $m = n_{bins} \times n_{cat}$. n_{bins} being the number of bins and n_{cat} the number of contribution categories used in the statistical model.

The `pyhf` software package used for the implementation of the statistical model described in Section 4.4 requires systematic uncertainties to be described as nuisance parameters. A possible approach is to use singular value decomposition (SVD). This allows to identify the most significant eigenvectors of the covariance matrix and add the remaining ones in quadrature, allowing to simplify the treatment of minor uncertainty sources.

Because the covariance matrix $\mathbf{C}$ is real, symmetric and positive semi-definite, there exists m orthogonal unit eigenvectors $\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_m$ with associated eigenvalues $\lambda_1 \geq ... \geq \lambda_m \geq 0$ such that:

$$\mathbf{C} = \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T = \sum_{i=1}^{m} \lambda_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T, \tag{4.23}$$

where $\mathbf{Q}$ is the $m \times m$ matrix whose columns are the eigenvectors and $\mathbf{\Sigma}$ is the diagonal matrix whose non-zero elements are the corresponding eigenvalues. Ordering said eigenvalues, if the first $t < m$ eigenvalues are significantly larger than the rest, we can assume:

$$\mathbf{C} \approx \sum_{i=1}^{t} \lambda_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T + diag\left(\sum_{j=t+1}^{m} \lambda_j \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j^T\right), \tag{4.24}$$

so that only the diagonal elements of the p through t minor terms are considered. Nuisance vectors $\lambda_{\mathbf{i}} = \lambda_i \hat{\mathbf{u}}_i$ for $i \in (1, t)$ can then be used to propagate correlated uncertainties to the statistical model while the remaining terms in Equation 4.24 are treated as uncorrelated uncertainties.

4.6 Upper limit determination

Previous searches for the $B^+ \to K^+ \nu \bar{\nu}$ decay have seen no significant signal (see Section 1.5). Thus, we propose to determine an upper limit on the signal strength μ defined in Equation 4.12.
From the likelihood model defined in Equation 4.17 and an assumed μ value, we can define the likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}} | n_1, ..., n_{N_b})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}} | n_1, ..., n_{N_b})}, \tag{4.25}$$

where the parameters $(\hat{\mu}, \hat{\boldsymbol{\theta}})$ maximize the likelihood for the set of observations $\{n_1, ..., n_{N_b}\}$ when the value $\hat{\mu}$ is allowed to fluctuate. In addition, $\hat{\hat{\boldsymbol{\theta}}}$ maximizes the likelihood for the same set of observations and a fixed μ value [106].
We can then define a likelihood-ratio test Λ_μ:

$$\Lambda_\mu = -2 \ln \lambda(\mu), \tag{4.26}$$

the -2 factor ensures that Λ_μ approaches asymptotically the χ^2 distribution [107]. It is then possible to evaluate an upper limit on μ for a given confidence level (CL) by finding the value μ verifying:

$$\Lambda_\mu = CDF_{\chi^2}^{-1}(\mathcal{C}), \tag{4.27}$$

where $\mathcal{C}$ corresponds to the required CL (ex: 0.9 for a 90% CL) and $CDF_{\chi^2}^{-1}$ is the cumulative distribution function of the χ^2 distribution.
The `pyhf` package is used for the upper limit determination.

4.7 Blind analysis

The analysis described in Chapter 5 is performed as a *blind analysis*. This allows to protect the analysis' result from potential biases. Some biases coming from the experimental apparatus have an effect on the result that can be gauged and are

usually treated with systematic uncertainties associated to the result. Other biases, coming from the person performing the measurement, are impossible to precisely estimate. Blind analyses are performed to limit the effect of the latter.
In this work, the analysis is developed using simulated physics samples. Which allows to gauge the behavior of the different analysis parts, such as the reconstruction, event classification and expected result. However, doing so exposes the analysis to mis-modeling in the simulation. Thus, the analysis process is then cross-checked using measured data, using specific selection criteria to identify independant data samples containing as few signal as possible (*cf.* Subsection 5.6.2 and 5.6.3). Potential discrepancies between data and simulation can, for example, be included in the result as associated systematic uncertainties.
Finally, once the sanity of the analysis has been duly checked, the analysis procedure is applied on the full data sample (*unblinding*).

Chapter 5

Search for the $B^+ \to K^+\nu\bar{\nu}$ decay

Contents

After describing the main tools and methods used in the different stages of this analysis in the previous chapter, we now aim at describing the steps devised to measure the branching ratio $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ using data collected by the Belle II experiment. This chapter first presents the overall selection method:

- Data samples used in this analysis are described in Section 5.1
- Event pre-selection (Section 5.2): Low-level objects are defined, before a broad selection is performed when reconstructing B_{tag} candidates using the FEI algorithm (described in Section 4.1). A tighter selection is then applied to create manageable datasets, based on the physical properties of the signal studied.
- Signal candidate selection (Section 5.3): In each event, one signal K^+ is identified and associated to a B_{tag} candidate.
- Event classification (Section 5.4): A set of variables is defined to differentiate between signal events and events from background processes. These variables are then studied on simulated events. Afterwards, a multivariate classifier is built and trained on simulated samples to classify events based on their signal-likeness.

The method is then validated using data, as a way to identify potential detector issues or mismodelling in the simulation (Section 5.6):

- Validation using embedded signal (Subsection 5.6.1): Using $B^+ \to K^+J/\Psi(\mu^+\mu^-)$ events identified in data, we swap the K^+ and $J/\Psi(\mu^+\mu^-)$ in the event with simulated $K + \nu\bar{\nu}$ and match the kinematics to mimic our signal. We use this sample to control the behavior of signal events during the selection process.
- Validation using off-resonance data (Subsection 5.6.2): Using data collected at an energy in the centre of mass frame 60 MeV below the mass of the $\Upsilon(4S)$ resonance, we control the behavior of $e^+e^- \to q\bar{q}$ events where $q \in (u, d, s, c)$.
- Validation using on-resonance data (Subsection 5.6.3): We further validate the selection by defining two orthogonal samples in the signal region of data, with the requirements that these samples be dominated by background and only marginaly populated by actual signal. This allows to study signal-like data events without introducing a bias by fine tuning parts of the analysis on data signal events.

Finally, we develop a statistical model (described in Section 4.4) to measure the value $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ (if not enough signal events are selected, we set an upper limit on this value). We also describe in Section 5.7 the different sources of systematic uncertainty on our measurement, as well as the methods used to evaluate them. The final result of our measurement is presented in Section 5.8.

5.1 Input datasets

The Belle II experiment aims at collecting 50 ab^{-1} of data at a collision energy corresponding to the mass of the $\Upsilon(4S)$ resonance. The analysis described thereafter makes use of a data sample corresponding to 362 fb^{-1} collected at the energy of the $\Upsilon(4S)$ resonance between 2019 and the summer of 2022 when the first Belle II long shutdown was started, which corresponds to 387.1×10^6 $B\bar{B}$ pairs. This sample is referred to as the *on-resonance data*.
In addition, a sample of 42 fb^{-1} is collected at an energy 60 MeV below the $\Upsilon(4S)$ resonance and is used for validation. The interest in this sample comes from the fact that it does not contain any B meson decays, as its associated energy is not sufficient to produce them. We refer to this sample as *off-resonance data*.
Finally, the following samples, simulated using the tools described in Section 2.6 are used to develop the analysis:

- 50×10^6 $B^+ \to K^+\nu\bar{\nu}$ events, refered to as *signal sample*,
- A sample corresponding to 1 ab^{-1} of equivalent integrated luminosity of $e^+e^- \to q\bar{q}$ events, with $q \in \{u, d, s, c\}$, refered to as *continuum background*,
- A sample corresponding to 3 ab^{-1} of equivalent integrated luminosity of $e^+e^- \to B\bar{B}$ events, refered to as *$B\bar{B}$ background*,

The simulated samples are taken from the official Belle II simulation production, produced with the tools described in Section 2.6.

5.2 Object selection

The first step of the reconstruction in this analysis is the identification of B_{tag} candidates using the FEI algorithm. This allows to fully reconstruct one of the two B mesons coming from the decay of the $\Upsilon(4S)$ in the hadronic modes listed in Table 4.1. Several B_{tag} candidates might be reconstructed for each event, with an associated probability $\mathcal{P}_{FEI}$. We then search for the signal signature ($B_{sig} \to K^+\nu\bar{\nu}$) in their recoil, reconstructed with remnant tracks.

To save computing time, reconstructed events are required to have at least 3 tracks (see [97]), complying with the following requirements in order to be able to reconstruct a B_{tag}:

- The transverse impact parameter of the track, $|d_0|$ is lower than 0.5 cm and $|z_0|$, its the longitudinal impact parameter, is lower than 2 cm (cf. Figure 2.4 for a description of Belle II's coordinate system). This allows to discard events without enough charged particles originating from the interaction point.
- The transverse momentum of the track, p_T must be greater that 0.1 GeV. This allows to discard a large portion of beam background tracks.

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

Table 5.1: Expected fractions of charged particles in B-meson decays. These are estimated from $e^+e^- \to B\bar{B}$ events [108].

These tracks are used to build charged particle candidates, identified amongst pions, kaons, electrons, muons or protons using PID information from the different Belle II subdetectors (cf. Subsection 2.7.2 and 2.7.3). An additional identification probability is derived from simulated $e^+e^- \to B\bar{B}$ events (Table 5.1).

Furthermore, considered events are required to contain at least 3 calorimeter clusters such that:

- The cluster energy E is greater than 0.1 GeV/c, this allows to suppress a large portion of beam background.
- The cluster polar angle θ verifies $0.297 < \theta < 2.618$ rad. This angular region corresponds to the CDC acceptance and so this requirement suppresses clusters potentially produced by charged particles that have not been tracked.

These ECL (see Subsection 2.3.5) clusters are used to build photon candidates. Finally, we require that the total visible energy in the event be greater than 4 GeV and that the total energy deposited in the calorimeter be in the range [2, 7] GeV. The last two quantities are computed considering the tracks and clusters previously defined.

Only B_{tag} candidates with a beam-constrained mass $M^*_{bc} > 5.27\,\mathrm{GeV}/c^2$ and $|\Delta E| < 0.3\,\mathrm{GeV}/c$ are retained, with:

$$M^*_{bc} = \sqrt{\left(\frac{\sqrt{s}}{2c^2}\right)^2 - \left(\frac{p^*_B}{c}\right)^2}, \tag{5.1}$$

$$\Delta E = \sqrt{E_B - \frac{\sqrt{s}}{2}} \tag{5.2}$$

Where $\sqrt{s}$ is the collision energy and p^*_B is the momentum of the B_{tag} candidate computed in the CMS, while E_B is the energy of the considered B-meson.

To each B_{tag}, we assign a signal probability ($\mathcal{P}_{FEI}$). $\mathcal{P}_{FEI}$ is the output of the final FEI multivariate classifier that ranges from 0 (misreconstructed) to 1 (correctly reconstructed). For each B_{tag} candidate, $\mathcal{P}_{FEI}$ is required to be greater than 0.001. Finally, events with more than 12 tracks with $|z_0| < 4$ cm, $|d_0| < 2$ cm are further rejected. With This requirement is due to the low multiplicity expected in signal events of the type $\Upsilon(4S) \to B_{tag}+B_{sig}$, with $B_{tag} \to$ hadronic modes, $B_{sig} \to K^+\nu\bar{\nu}$.

5.3 Signal candidate selection

As described in the previous section, several B_{tag} candidates might be reconstructed in each event. We then search a B_{sig} for each of them. Because B_{sig} decays as $B_{sig} \to K^+\nu\bar{\nu}$ this comes down to pair each B_{tag} candidate to a K^+ candidate. Down the line, only one set of $B_{tag} + B_{sig}$ is retained.
Signal kaon candidates are selected from tracks verifying:

- Basic IP constraint: $d_0 < 0.5$ cm, $|z_0| < 2$ cm,
- CDC acceptance requirement: $0.297 < \theta < 2.618$,
- Tracking quality: at least 20 hits in the CDC and 1 hit in the PXD,
- Particle identification: kaonID > 0.9.

This retains around 60% of true kaons and rejects around 95% of mis-identified kaons.
Once a B_{tag} and signal side kaon have been paired together, the number of extra-tracks not associated to either B_{tag} nor to the K^+ candidate is required to be zero. Such counting is done on objects with $d_0 < 2$ cm, $|z_0| < 4$ cm, reconstructed in CDC acceptance and with at least 20 CDC hits. In addition, we require that no additional reconstructed π^0, K_S^0 and Λ^0 be left in the event. Afterwards, we define the rest-of-event (ROE), which consists of remaining tracks and ECl clusters not associated with either B_{tag} nor with B_{sig}. For perfectly reconstructed signal events, the ROE contains no particles. For mis-reconstructed events, given the aforementioned cut on extra-tracks, the ROE is formed by neutral deposits not associated with charged particles.
In addition, we require that the B_{tag} and B_{sig} be of opposite electric charge. Finally, we compute the missing momentum vector $\mathbf{p}_{miss}$ as:

$$\mathbf{p}_{miss} = -\sum_{i=1}^{N} \mathbf{p}_i \tag{5.3}$$

Where N is the number of particle candidates in the event. The polar angle of the missing momentum, θ_{miss} is required to verify $0.3 < \theta_{\text{miss}} < 2.8$ rad, in order to make sure that the missing momentum is not due to particles escaping the detector acceptance.

In order to retain a single $B_{tag} + B_{sig}$ pair per collision, the B_{tag} candidate with the highest FEI probability is identified. This is done after the classifier selection (see Section 5.5).

5.4 Background suppression

The main challenge in observing the $B^+ \to K^+\nu\bar{\nu}$ signal is the large background contamination. Therefore, powerful background suppression is needed. After the selection described in the previous section, we identify a set of discriminating variables used to train a multivariate classifier to separate signal and background. To achieve optimal separation, we explore several categories of variables to extract distinctive signal feature information. The variables used are sensitive to the event topology and kinematic properties of the ROE and B_{tag}, or characterize the signal candidate. In addition, we consider variables obtained by reconstructing vertices and invariant masses of two and three charged particles including the signal K^+ candidate to identify and veto potential contributions from D^0 and D^+ meson decays. Numerous variables are considered, though only a minimal set of variables that are well described in the simulation are kept. The data-simulation agreement is confirmed with control-sample studies, as described in Section 5.6.

5.4.1 Variables of interest

A set of variables is built with the intent of using said variables as features for the training of a multivariate classifier tasked with estimating the signal-likeness of the event studied. The choice of variables is motivated by:

- Number of features: In order to avoid correlations between variables and over-complication of the classifier (see Section 4.2), we choose to select as few features as possible, discarding variables showing a discriminative power under a certain threshold.
- Discriminative power: Features kept in the classification process should show adequate discrimination between signal and background events. This is evaluated on the simulated samples described in Section 5.1. The estimation of this discriminative power is described in Subsection 5.4.4.
- Adequate modeling: The computation and testing of the variables of interest being performed on simulated samples, it is important to check that they are well modelled. Indeed, physical processes not taken into account during simulation, or inefficiencies of the detectors can bias the distribution of the computed features, compared with what is seen in recorded data. To avoid these issues, the data/simulation agreement for the features is studied in several control channels (see Section 5.6).

The variables are split into different categories described as below. The distributions shown in the different figures are based on the simulated samples mentionned in Section 5.1, after the selection steps described in Section 5.3. The variables are computed in the laboratory reference frame unless otherwise specified (some are computed in the centre of mass frame, noted CMS). Distributions are normalized to unitarity area.

5.4.1.1 General event properties

Several variables used in the classification are related to the geometrical distribution of reconstructed particles in the event or their multiplicity. These features are mainly computed using the momenta of the particles in the event.
The event shape variables retained in the classification are:

- The modified Fox-Wolfram moments H_{22}^{so}, H_{02}^{so} and H_0^{oo}, as described below, are computed in the CMS and provide good discrimination between signal and $q\bar{q}$ events. This is due to the difference in event shapes expected between the different event types.

- The number of remaining tracks in the event. As mentionned in Section 5.3, we require that no *clean* tracks remain in the event after reconstructing a $\Upsilon(4S)$ from a B_{tag} and B_{sig} pair. The feature computed here then corresponds to the number of tracks left in the event that do not meet the requirements to be classified as *clean* tracks. This variable proves to be extremely discriminative as signal events are expected to show exactly zero extra track, while the missing component of the signal can be mimicked in background events by low quality tracks not used in the reconstruction of the B_{tag} candidates.

- The extra energy in the event associated to ECL clusters from neutral particles, NE_{ECL}^{Extra}. This feature is defined as the sum of the energy from calorimeter clusters that are not associated to any track in the event. This extra energy in the event proves to be the most discriminative feature and is further detailed below.

The distributions of these variables for simulated signal and background samples can be found in Figure 5.1.

Modified Fox-Wolfram moments

Fox-Wolfram moments were first introduced by G. C. Fox and S. Wolfram to provide variables to describe event shapes in e^+e^- annihilation [109, 110]. Modified Fox-Wolfram moments were later developed by the Belle collaboration [69].
These variables are developed specifically within the framework of B-factories, dividing particles produced in events into two conceptual classes: B-meson candidate daughters (labeled s) and particles coming from the rest of the event (ROE), denoted as o. For a given event, the total number of particles N verifies $N = N_s + N_o$, with N_s and N_o corresponding to the number of particles in the s and o classes respectively. In addition, particles are further classified in 3 subsets labeled with integers: charged particles (label 0), neutral particles (label 1) and missing particles (label 2). It is worth noting that the entirety of the missing momentum in the event (defined in subsubsection 5.4.1.4) is treated as one missing particle.
The signal-ROE (so) modified Fox-Wolfram moment of degree $l \in \mathbb{N}$ for the particle

category $x \in \{0, 1, 2\}$ is defined as:

$$H_{xl}^{so} = \frac{1}{Z} \sum_{i=1}^{N_s} \sum_{j_x=1}^{N_x} C_{ij_x}^{l} p_{j_x} P_l(\cos \alpha_{ij_x}), \tag{5.4}$$

with:

- Z a normalization factor verifying $Z = 2\left(\sqrt{s} - E_B^*\right)$, with $\sqrt{s}$ the available energy in the center-of-mass frame and E_B^* the signal B-meson candidate energy in the center-of-mass frame.
- $C_{ij_x}^l \in \{-1, 0, 1\}$ the product of the charges for the candidates i and j_x if l is odd; $C_{ij_x}^l = 1$ if l is even.
- P_l the Legendre polynomial of l-th order.
- α_{ij_x} the angle between the momenta $\mathbf{p}_i$ and $\mathbf{p}_{j_x}$.

The ROE-ROE (oo) modified Fox-Wolfram moment of degree l can then be described as:

$$H_l^{oo} = \frac{1}{Z^2} \sum_{i=1}^{N_o} \sum_{j=1}^{N_o} C_{ij}^{l} p_i p_j P_l(\cos \alpha_{ij}), \tag{5.5}$$

with the same notations as in Equation 5.4.

Extra energy in the calorimeter

The extra energy from neutral sources in the event, NE_{ECL}^{Extra}, is computed from energy deposits in the ECL subdetector associated to photons in the ROE defined in Section 5.3. These photon candidates must verify the following requirements:

- The photon candidate associated cluster energy must be greater than (0.100, 0.060, 0.150) GeV, for clusters in the (forward, barrel, backward) regions of the ECL.
- The distance between the photon candidate and the closest track in the event must be greater than 50 cm.
- The photon candidate must be within the CDC acceptance.

NE_{ECL}^{Extra} corresponds to the sum of the energy deposited in the ECL for each retained photon candidate.

5.4.1.2 B meson kinematic variables

The kinematics of the signal kaon candidate are expected to vary between signal and background events. The relationship between the B_{tag} and B_{sig} momenta is also expected to provide discriminative power. Additional variables have been considered (e.g. signal kaon candidate momentum) but have not been retained because of the correlations they show with other variables.
These kinematic variables are:

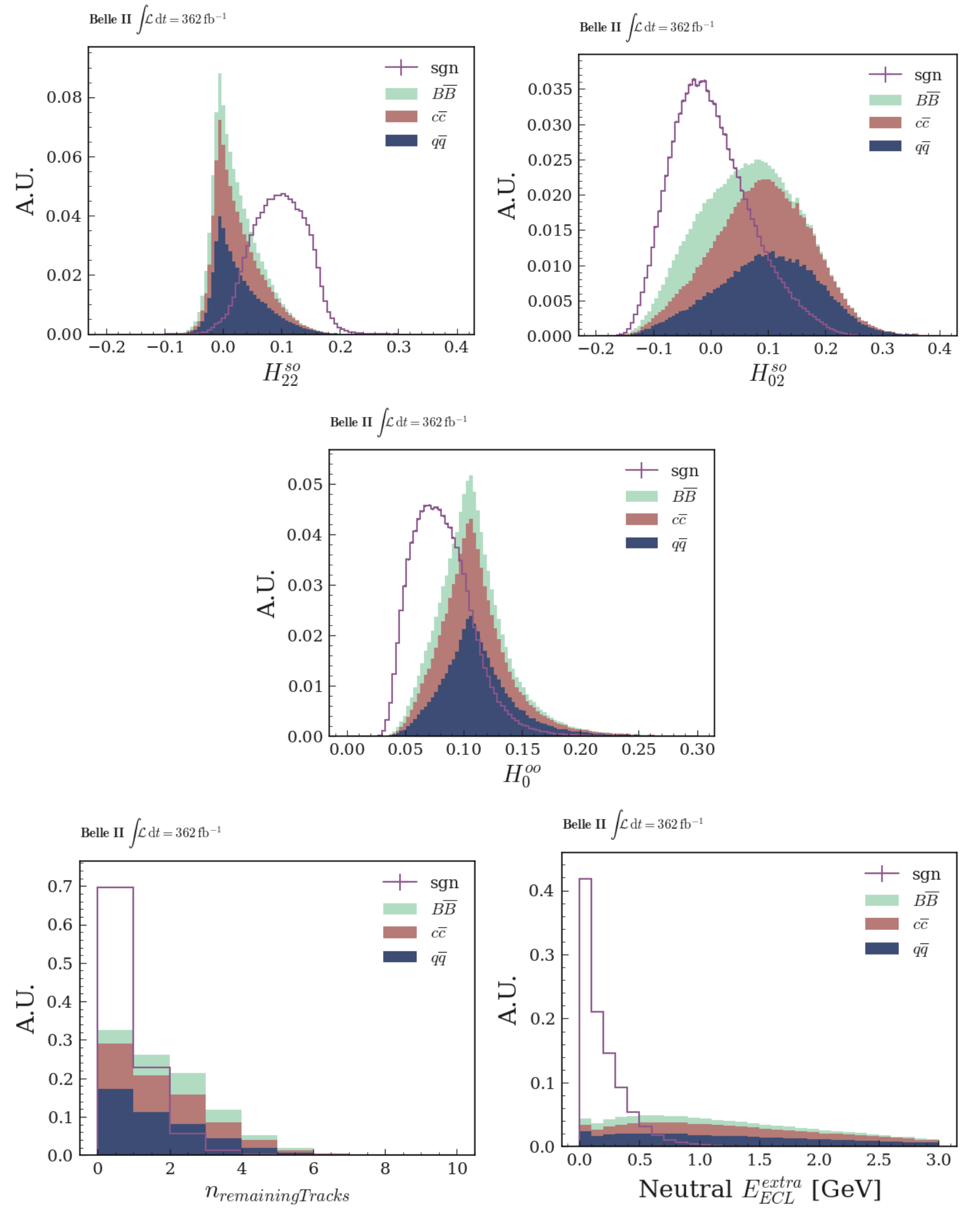


Figure 5.1: Distribution of the general event variables used in the classification. The Kakuno-Super-Fox-Wolfram moments (first two rows), the number of tracks remaining in the event after the $\Upsilon(4S)$ reconstruction (bottom left) and the extra energy in the event NE_{ECL}^{Extra} (bottom right), for the different simulated samples.

- The cosine of the angle between the kaon candidate three-momentum and the thrust axis of the ROE, $cos(\theta_{Bthr})$, computed in the CMS. We see in Figure 5.2 that the distribution of this variable is mostly uniform in signal events. This is due to the fact that, in signal events, the momentum of the signal kaon is not correlated to the momentum of the ROE.

- The recoil mass of the kaon associated to the signal B-meson candidate.

5.4.1.3 D meson identification variables

D-mesons decaying into a kaon and one or two pions contribute to the background when said kaon is selected as the signal kaon candidate.
To suppress such background, we reconstruct D-meson candidates using the signal kaon candidate and ROE tracks, fitting them to a common vertex. Several D mesons candidates are reocnstructed in this manner and are ranked based on the p-value of their vertex fit.
Two hypotheses are retained for D-meson candidates: D^0 candidates reconstructed using the signal kaon candidate and one ROE track, and D^+ candidates reconstructed using the signal kaon candidate and two ROE tracks. The ROE tracks are constructed using a pion hypothesis.
The p-values of the best D-meson candidate in both categories are used as input variables for the classifier, the correpsonding distributions are shown in Figure 5.3.

5.4.1.4 Variables related to missing quantities

Finally, because a large fraction of the event 4-momentum is carried by the neutrino pair in signal events, we expect variables related to the event missing observables (missing energy or momentum) to be strongly discriminative. We also expect some

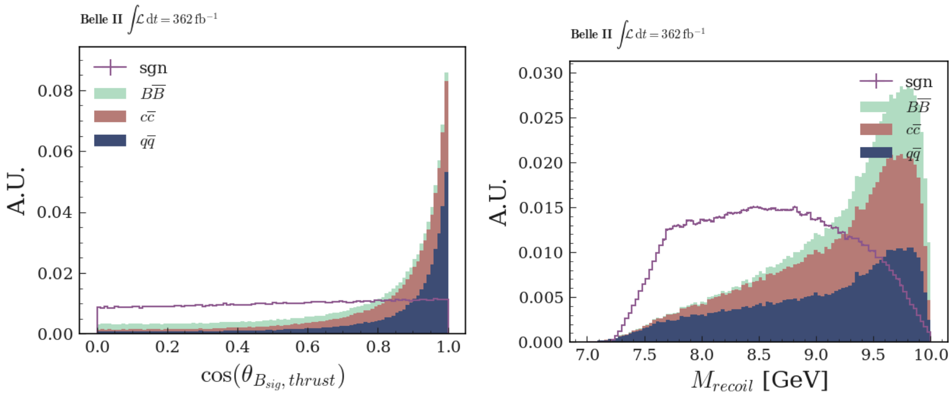


Figure 5.2: Distributions of the kinematic variables used in the training of the classifier. The cosine of the angle between the kaon candidate three-momentum and the thrust axis of the ROE (left) and the recoil mass of the B_{sig} candidate (right).

background events to display similar missing quantities as a result of particles travelling outside the detector acceptance or being ineffectively detected, as well as long-lived neutral particles leaving the detector without interacting before eventually decaying.
The features computed using missing quantities in the event are:

- The angle between the missing momentum and the signal kaon candidate momentum computed in the CMS, $\phi^*(K, p_{miss})$, computed in the CMS frame and defined as:

$$cos(\phi^*(K^+, p_{miss})) = \frac{\mathbf{p}_K \cdot \mathbf{p}_{miss}}{|\mathbf{p}_K||\mathbf{p}_{miss}|} \tag{5.6}$$

  With $\mathbf{p}_K$ the signal kaon candidate momentum.

- The sum of the missing energy and momentum in the event, $E_{miss} + c\mathbf{p}_{miss}$, computed in the CMS. Signal events are expected to have significantly higher missing energy and momentum than background events.
  The distributions of these variables are shown in Figure 5.4.

5.4.1.5 Features left out of the classifier training

The following features prove important for controls as well as for the interpretation of the measurement but are not used in the training of the classifier:

- The invariant mass of the neutrino pair, computed as:

$$q^2 = m_B^2 + m_K^2 - 2E_B E_K + 2\mathbf{p}_B \cdot \mathbf{p}_K, \tag{5.7}$$

  where m_B and m_K correspond to the masses of the B^+ and K^+ mesons respectively, while $E_B/\mathbf{p}_B$ and $E_K/\mathbf{p}_K$ correspond to their energies/momenta.

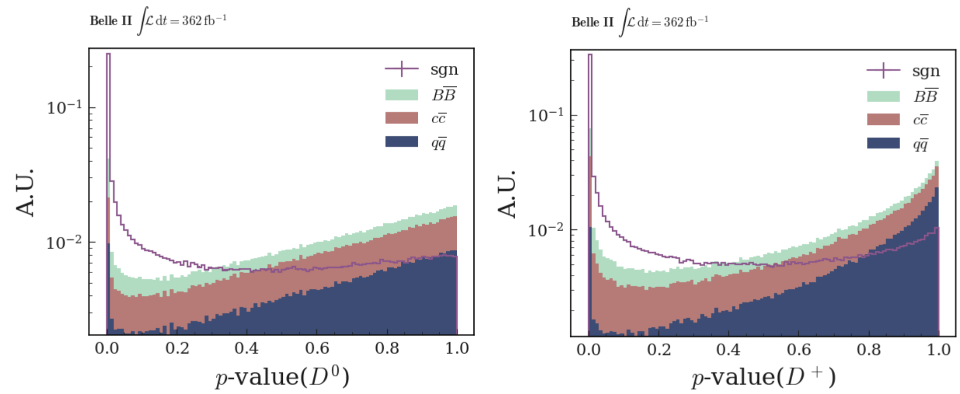


Figure 5.3: Variables related to the D-meson identification: the p-value of the fit for D^0 candidates (left) and D^+ candidates (right).

This is an important quantity, as $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ exhibits a q^2-dependence. We can furthermore express q^2 using reconstructed quantities:

$$q^2 \approx \frac{s}{4} + m_K^2 - \sqrt{s}E_K^* - 2\mathbf{p}_{tag} \cdot \mathbf{p}_K, \tag{5.8}$$

where $\sqrt{s}$ is the available energy in the collision event defined in Equation 2.1, $\mathbf{p}_{tag}$ is the momentum of the B_{tag} meson and E_K^* is the energy of the reconstructed signal candidate in the center-of-mass frame. This approximation assumes that the $\Upsilon(4S)$ meson is approximately at rest in the center-of-mass frame, then $\mathbf{p}_B = -\mathbf{p}_{tag}$ follows. In addition, using $\sqrt{s}/2$ instead of m_B allows to better reflect the variations of $\sqrt{s}$ dependent on the experimental condiditions.

- The number of extra photons in the event N_γ corresponds to the number of photon candidates in the *ROE* of the event satisfying the requirements described in subsubsection 5.4.1.1. This variable is used to derive a correction to the most discriminative variable, NE_{ECL}^{extra}, detailed in Subsection 5.7.8.

5.4.2 Event classification

In this section, we describe the main selection step in this analysis. We classify events based on their signal-likeness using a gradient-boosted decision tree (BDT) based on XGBoost [102]. The working principle of binary classification as well as the way it is implemented in this analysis are described in Section 4.2.
We detail in the following the way the classifier is built, trained and we measure its classification performance.

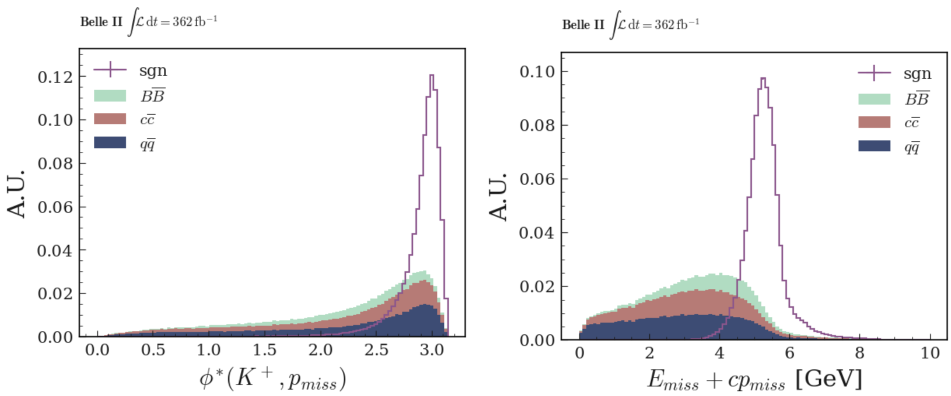


Figure 5.4: Distribution of the variables related to the missing 4-momentum in the event: $E^*_{miss} + p^*_{miss}$ (left), and ϕ angle between the signal kaon and missing tri-momentum (right).

5.4.3 Classifier training

The classifier is built with the 13 variables described in Subsection 5.4.1 and trained using the full simulated samples described in Section 5.1. In order to keep overtraining under control, the simulation sample is randomly split in halves, the classifier is then trained on both sub-samples simultaneously, using the other subsample to test the training (this corresponds to a 2-fold validation, which is discussed in Subsection 4.2.4). The training sample is further split into signal (what the classifier has to identify), and background (containing the three types of background events: $B\bar{B}$, $c\bar{c}$, $q\bar{q}$). In the case of the testing sample, we conserve the information on the type of event, while the classifier is kept blind to it.
As can be seen in Figure 5.5 there is a good agreement between the output of the two trainings.
It is possible, after performing the training, to estimate the gain brought by each feature. Figure 5.7 shows the importance of each feature in the classification of the events. We see that some features bear a larger importance than others. Even though BDTs are typically good at handling correlations, we want to retain the minimum number of features needed to achieve good performance. This reduces correlations as well as the potential masking between variables. Because the feature importance can be tricky to interpret, it is useful to proceed by backwards elimination to identify the best set of features to use.
To do so, we train the classifier using n features, then train $n-1$ classifiers using as features the full set of variables to which a random variable is substracted and pick the best set (that is, the one giving the lowest μ_{sens} value evaluated on the testing sample, see Section 4.3 for the definition of μ_{sens}), and so on and so forth.
Finally, we transform the features to follow a uniform distribution which helps with shielding against outliers. The variables kept after this procedure are the ones described in Subsection 5.4.1.

5.4.4 Classifier parameters

Several parameters of the classifier impact its training:

- The number of trees (n_T);
- The maximum depth of each tree (d_T);
- The learning rate ($0 < \eta < 1$);
- The sampling rate ($0 < \sigma < 1$);
- The positive/negative weights balance S_w.

The η parameter shrinks feature weights after each boosting round in order to prevent overfitting, while σ corresponds to the fraction of the training sample used in each boosting round: for each round, the training procedure randomly samples

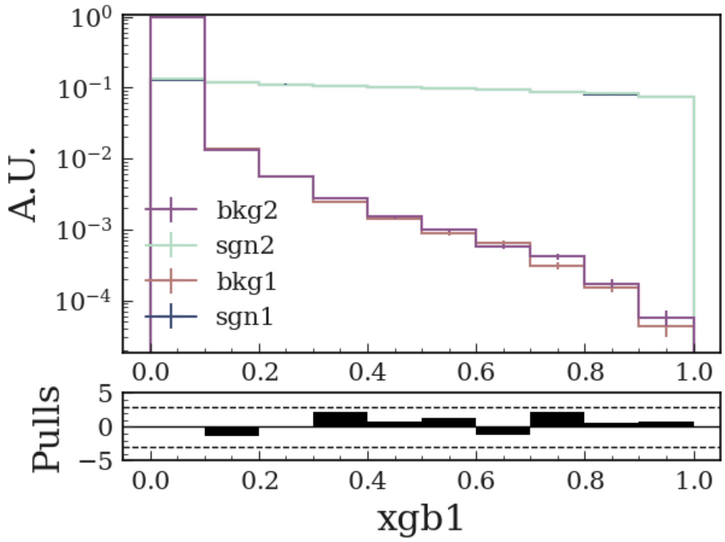


Figure 5.5: Classifier output for the two BDTs trained by splitting simulation samples in 2 and swapping training and testing sample.

$\sigma \times n_{train}$ to use in the training, with the aim of reducing overfitting. S_w controls the balance of positive and negative weights for unbalanced classes.

To optimally parameterize the classifier, we investigate different values for the parameters n_T, η and σ. The tree depth is kept at a constant value $d_T = 3$, we also fix $S_w = 10 \times n_{bkg}/n_{sgn}$.

We then aim at finding a $(n_T; \eta; \sigma)$ set offering a good trade-off between classifier performance and overfitting. To do so, we make use of the `Optuna` package [111] to perform an optimization in the parameter space. `Optuna` allows one to automatically search for a given parameter space with the goal of minimizing a user-defined *objective* function. Here, the *objective* is defined as the μ_{sens} defined in Section 4.3 evaluated on the testing sample. In order to monitor overfitting we compare this value to the μ_{sens} computed for the validation sample. Figure 5.6 shows the result of this optimization. We find an adequate set of parameters to be $(n_T = 1300, \eta = 0.03, \sigma = 0.8)$.

Figure 5.8 shows a good trade-off between classifier output performance and overfitting. The values chosen for each parameter of the classifier cam be found in Table 5.2.

5.5 Signal search region

After training and optimizing the classifier, we now aim at defining a region, based on the classifier output, on which the binned-likelihood model defined in Section 4.4 will be applied to data to measure the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$. In Subsection 5.5.1 with describe how this signal region (SR) is defined. In Subsection 5.5.2 we study

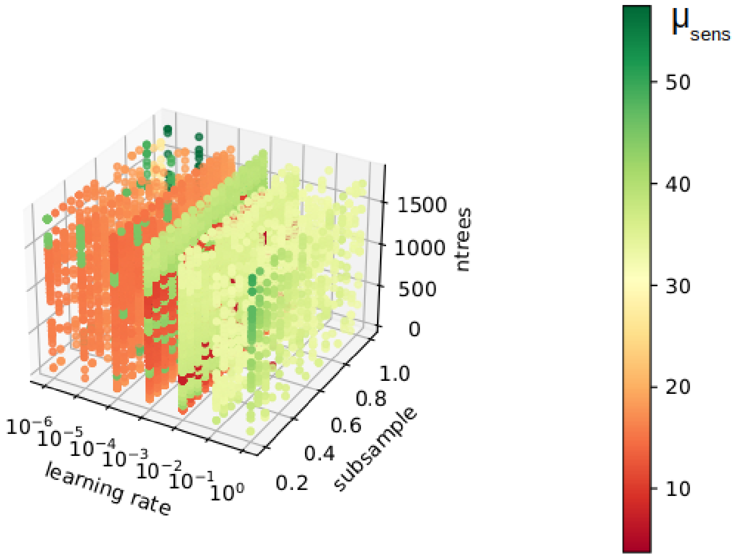


Figure 5.6: Distribution of the estimated μ_{sens} for each $(n_T; \eta; \sigma)$ combination.

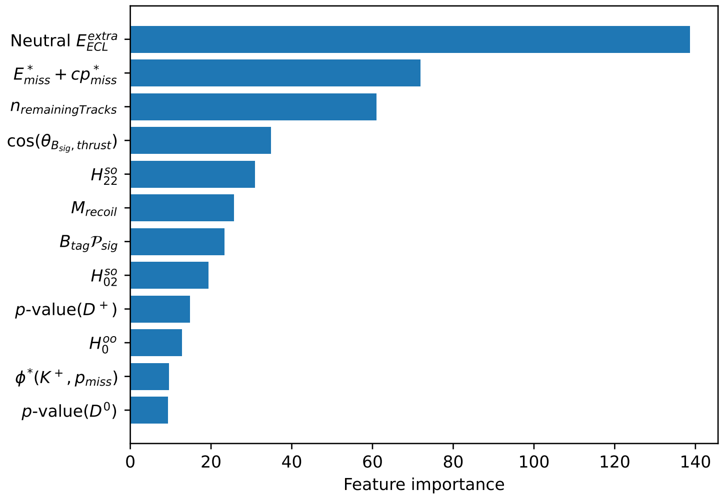


Figure 5.7: Importance of the 13 features used in the training of the classifier.

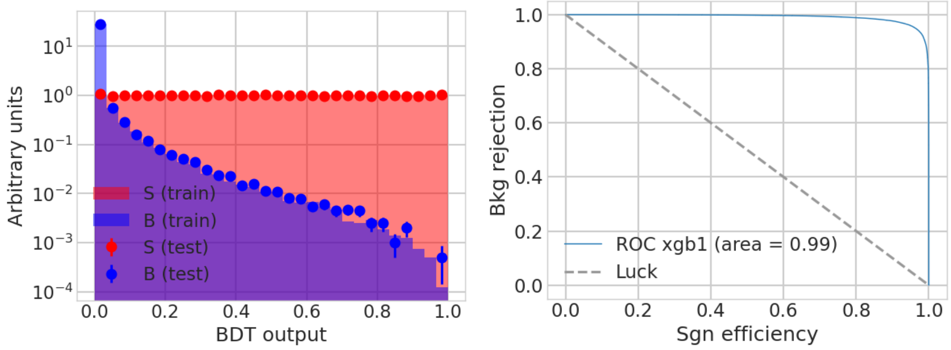


Figure 5.8: Training-testing agreement for signal (red) and background (blue) samples and Area Under the Curve (AUC) for our choice of classifier parameters.

Parameter	Value
Number of trees (n_T)	1300
Tree depth (d_T)	3
Shrinkage (η)	0.03
Sampling rate σ	0.8
Positive/negative weights balance (S_w)	1

Table 5.2: Hyperparameters of the classification model used in the analys.

the contribution of each event types to the SR using simulated samples and we characterize the leading sources of background contributions.

5.5.1 Definition

We define the signal search region based on a requirement on the classifier output value. This value is taken to correspond to about 60% signal selection efficiency after the pre-selection described in previous sections. In the end, in the SR the signal selection efficiency is $\sim$ 0.40%. This selection corresponds to a lower threshold requirement on the classifier output value $BDT > 0.4$. The region is divided in 6 equal bins of classifier output value. The comparison between data yields and expected yields from simulation in these bins will be the primary input in the binned-likelihood model to measure $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$.
Table 5.3 shows the signal selection efficiency at different stages of the selection.

5.5.2 Simulation study

We use the simulated samples described in Section 5.1 to study the expected behavior of the SR. Figure 5.9 shows the expected signal and background yields in the SR for an integrated luminosity of 360 fb^{-1}. The classifier output distribution is, by construction, flat for the signal contribution. This allows to easily treat classifier

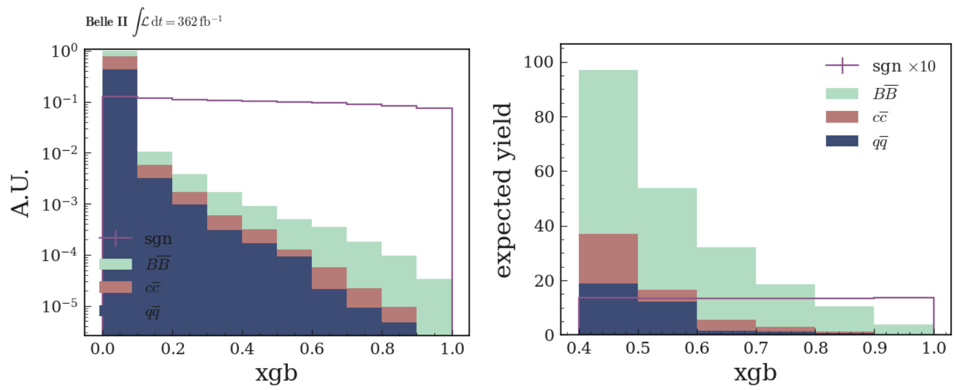


Figure 5.9: Distribution of $B^+ \to K^+\nu\bar{\nu}$ candidates in the whole classifier output range (left) and signal search region (right) obtained in simulated (filled histograms) generic background and (grey line) corresponding signal samples. The expectations are provided for L = 362 fb^{-1}. The signal expectation is magnified by a factor of 10 for better visibility.

Selection stage	ε_{sig} $(\times 10^{-2})$
Hadronic FEI selection	2.482 ± 0.002
Basic event selection	0.6598 ± 0.0011
Signal search region	0.3996 ± 0.0009

Table 5.3: Signal selection efficiency at various stages of the selection. The uncertainties quoted are statistical only.

output bins as signal efficiency quantile regions.

We expect the three background contributions ($B\bar{B}$ pairs, $c\bar{c}$ and light $q\bar{q}$) to populate the lower classifier output bins, with $q\bar{q}$ events only populating the first SR bins. In addition, we see that the lower threshold defining the SR allows to discard most of the $q\bar{q}$ contribution.

Almost all of the background contamination in the last SR bins comes from $B\bar{B}$ pair events. Subsection 5.5.3 describes the study and classification of these events in simulation. Here, the simulated $B\bar{B}$ events are classified according to the generated decays of both B mesons, as several factors can fake the signal signature.

5.5.3 Background composition in the signal region

The $B\bar{B}$ events populating the signal region are classified and counted in order to assess the main contributions to the $B\bar{B}$ sample yields.

Because the selection is based on the tagging method described in Section 4.1, signal events are of the type $\Upsilon(4S) \to B^+(K^+\nu\bar{\nu})B^-(X)$, where X corresponds to one of the decays listed in Table 4.1. Several issues can lead to a $B\bar{B}$ event being wrongfully selected as signal (misidentification of the signal K^+, wrong reconstruction of the

B-meson decay category	Requirements
$Dn\pi$	One B daughter is in the $\boldsymbol{D}$ class, the other daughters are in the $\boldsymbol{n\pi}$ class.
$D\ell\nu$	B has 3 daughters. One is in the $\boldsymbol{D}$ class, one is in the $\boldsymbol{\ell}$ class and one is in the $\boldsymbol{\nu}$ class.
$D\tau\nu$	B has 3 daughters. One is in the $\boldsymbol{D}$ class, one is in the $\boldsymbol{\tau}$ class and one is in the $\boldsymbol{\nu}$ class.
D Hadrons	One B daughter is in the $\boldsymbol{D}$ class, the other daughters are in the **Hadrons** class.
DD	B has 2 daughters. Both are in the $\boldsymbol{D}$ class.
$n\pi\ell\nu$	One B daughter is in the $\boldsymbol{\ell}$ class, one is in the $\boldsymbol{\nu}$ class and the others are in the $\boldsymbol{n\pi}$ class.
$K^+K^0K^0$	B has 3 daughters. One is a K^+, the others are $K^0/\bar{K}^0$.
$c\bar{c}$	At least one B daughter is in the $\boldsymbol{c\bar{c}}$ class.
Hadrons	All B daughters are in the **Hadrons** class.

Table 5.4: B-meson decay categories used to classify the $B\bar{B}$ background events. The categories are mutually exclusive (a given $B\bar{B}$ event cannot be present in different categories). The different classes, written in bold, are defined in Appendix D.

B_{tag}). Because of this, both B-mesons in $e^+e^- \to Upsilon(4S)$ events need to be studied to understand the composition of the background in the SR. We decide to classify B-mesons decays in several categories described in Table 5.4, the prevalence of $B\bar{B}$ background in the SR is then studied in simulated samples, based on these categories (see Table 5.5 and 5.6).

Around 90% of the $B\bar{B}$ contribution to the SR comes from charged B^+B^- pairs. The main overall background contribution($\simeq 50\%$ of all charged $B\bar{B}$ yields) comes from events where one B meson decays semileptonically as $B \to D^{(*)}\ell\nu$, with $(\ell = e, \mu)$ and the other B meson decays into a final state composed of several pions and a D-meson. In these cases, a kaon from the D meson decay is selected as the signal kaon, while the undetected neutrino in the event, potentially associated to an additional particle travelling outside the detector acceptance, mimics the missing energy expected in the signal.

Because of their prevalence, these decays motivate the development of the D meson suppression variables described in Subsection 5.4.1.

In addition, several decays are expected to populate the signal region because they inherently show the same experimental signature as the signal. This includes the $B^+ \to K^+n\bar{n}$ and $B^+ \to K^+K^0_L\bar{K}^0_L$ decays. We further discuss these in Subsection 5.7.5 and Subsection 5.7.6.

B^+B^- event type	occurence (%)
misidentified K_{sig}	3.42%
$Dn\pi + D\ell\nu$	50.34%
$Dn\pi + Hadrons$	4.97%
$Dn\pi + c\bar{c}$	3.84%
$D\ell\nu + D\ell\nu$	3.77%
$Dn\pi + K^+K^0K^0$	3.69%
$D\ell\nu + DHadrons$	3.54%
$D\ell\nu + DD$	2.94%
$Dn\pi + D\tau\nu$	2.86%
$Dn\pi + DHadrons$	2.86%
$D\ell\nu + c\bar{c}$	2.64%
$Dn\pi + Dn\pi$	2.03%
$Dn\pi + DD$	0.98%
$D\ell\nu + D\tau\nu$	0.90%
$D\ell\nu + Hadrons$	0.60%
$c\bar{c} + DD$	0.45%
$c\bar{c} + Hadrons$	0.45%
$DHadrons + DHadrons$	0.45%
$DHadrons + Hadrons$	0.45%
$D\tau\nu + c\bar{c}$	0.30%
$K^+K^0K^0 + c\bar{c}$	0.23%
$DD + DHadrons$	0.23%
$D\tau\nu + DHadrons$	0.15%
$K^+K^0K^0 + DD$	0.15%
$DD + Hadrons$	0.15%
$D\ell\nu + K^+K^0K^0$	0.08%
$n\pi\ell\nu + c\bar{c}$	0.08%
$D\tau\nu + DD$	0.08%
$K^+K^0K^0 + DHadrons$	0.08%
$c\bar{c} + c\bar{c}$	0.08%
$c\bar{c} + DHadrons$	0.08%
$Hadrons + Hadrons$	0.08%
other	10.12%

Table 5.5: Prevalence of simulated B^+B^- decays in the signal region of the analysis. Precisions on the naming scheme can be found in Appendix D. The "misidentified K_{sig}" category corresponds to the percentage of events where the identified signal K^+ is not a generated K^+.

$B^0\bar{B}^0$ event type	occurence (%)
misidentified K_{sig}	10.14%
$Dn\pi + D\ell\nu$	41.13%
$Dn\pi + DHadrons$	10.48%
$D\ell\nu + D\ell\nu$	6.45%
$D\ell\nu + cc$	4.03%
$D\ell\nu + DD$	4.03%
$D\ell\nu + Hadrons$	3.23%
$Dn\pi + Hadrons$	2.42%
$D\ell\nu + DHadrons$	2.42%
$DHadrons + DHadrons$	2.42%
$Dn\pi + Dn\pi$	1.61%
$Dn\pi + D\tau\nu$	1.61%
$Dn\pi + cc$	1.61%
$Dn\pi + DD$	1.61%
$DHadrons + Hadrons$	1.61%
$D\ell\nu + D\tau\nu$	0.81%
$D\tau\nu + DHadrons$	0.81%
$D\tau\nu + Hadrons$	0.81%
$cc + Hadrons$	0.81%
$DD + Hadrons$	0.81%
other	10.14%

Table 5.6: Prevalence of simulated $B^0\bar{B}^0$ decays in the signal region of the analysis. Precisions on the naming scheme can be found in Appendix D. The "misidentified K_{sig}" category corresponds to the percentage of events where the identified signal K^+ is not a generated K^+.

5.6 Simulation validation using control channels

Every step of the analysis described up to this point has been developed using simulated samples. Considerable efforts have been put into the development of the different tools described in Section 2.6, with the ultimate goal of accurately describing the physical processes and detector interactions in the Belle II experiment. However, small but potentially harmful discrepancies might exist between measured data and simulation. In order to ensure a reliable estimation of the desired parameters, it is essential to identify and correct such discrepancies.
In this section, we investigate the agreement between data and simulation throughout the analysis process. However, we cannot measure and correct potential effects directly on events that populate the analysis' signal region, as we could introduce bias to the result. To avoid biases, we need to define several *control samples* to be studied in both simulation and data on which to gauge the robustness of the selection process without unblinding our signal sample:

- We check the efficiency of signal selection using modified $B^+ \to K^+ J/\psi$ events reconstructed in data and simulation. We describe in Subsection 5.6.1 the process through which these events are modified to mimic our signal signature.

- We check the agreement between off-resonance data and $q\bar{q}$ simulation of the distributions of the classifier features in Subsection 5.6.2. Off-resonance data are expected to behave similarly to $q\bar{q}$ continuum. In addition, the off-resonance data sample size (42 fb^{-1} of integrated luminosity) allows to shield this study against too much statistical fluctuations (which is a limitation in the study of the other control samples). We also describe how we improve data-simulation agreement for $q\bar{q}$ events by building an additional classifier trained on off-resonance data.

- Finally, we check data/simulation agreement for the entire background contribution (continuum $q\bar{q}$ and $B\bar{B}$ coming from $\Upsilon(4S)$ production) in signal sidebands. We define several signal sidebands, described in Subsection 5.6.3. These samples all consist in on-resonance data passing the signal selection with some requirements being inverted to assure that contamination from actual signal is kept to a minimum. It is optimal to construct several sideband samples, fully orthogonal to each other, to identify and decouple potential simulation issues.

5.6.1 Signal efficiency validation in embedded $B \to K^+J/\Psi$ events

We want to validate the behavior of signal events in the analysis using data events, without unblinding actual signal candidates.
To do so, we use three different samples: simulated signal events, simulated $B^+ \to K^+J/\psi$ events and reconstructed $B^+ \to K^+J/_{\psi}$ events. Specifically, we restrict the selection to events with $J/_{\psi} \to \mu^+\mu^-$. This decay is considered because it is rather easily reconstructed and shares kinematic similarities with our signal. The steps of the method are enumerated below.

1. Events containing a $B^+ \to K^+J/\psi$ decay in data and simulated samples are identified and selected.
2. All objects associated with the selected $B^+ \to K^+J/\psi$ decay are removed, keeping only the ROE, which contains the decay product of the accompanying B^- meson when the $B^+ \to K^+J/\psi$ decay is correctly identified.
3. Events containing a $B^+ \to K^+\nu\bar{\nu}$ decay are selected in signal simulated samples, and the same procedure is used to remove all objects *not* associated with the $B^+ \to K^+\nu\bar{\nu}$ decay in the events.
4. The signal decay of step 3 is combined with the ROE of step 2 to form an "embedded" event.
5. Finally, the signal decay kinematics is adjusted to match the kinematics of the original $B^+ \to K^+J/\psi$ decay. The reconstructed signal K^+ is shifted and rotated so that the position of the decay vertex and the direction of the B^+ meson for the simulated signal B^+ agree with those determined for the reconstructed $B^+ \to K^+J/\psi$.

The *signal embedding* procedure is applied to both data and simulation:

- A sample of 73651 events is used in simulation,
- A sample of 7214 events is used in data.

These events are then subjected to the reconstruction and selection described in Section 5.2. A sample of 112 (1709) candidates on data (simulation) are retained at this stage. Figure 5.10 shows the distributions of some of the BDT input variables for the embedded simulated and data samples along with signal simulated events. The distributions of all input variables are reported in Appendix C.
The embedded simulated sample reproduces the simulated signal well. We also see an overall good agreement between the embedded data and simulation. The classifier optimized for the signal search is run on the embedded samples, the output distribution is reported in Figure 5.11.

Table 5.7 presents the selection efficiencies for embedded samples after preselection and after final selection. The efficiencies are normalized to the number of events passing the embedding procedure (7214 for data and 73651 for simulation).

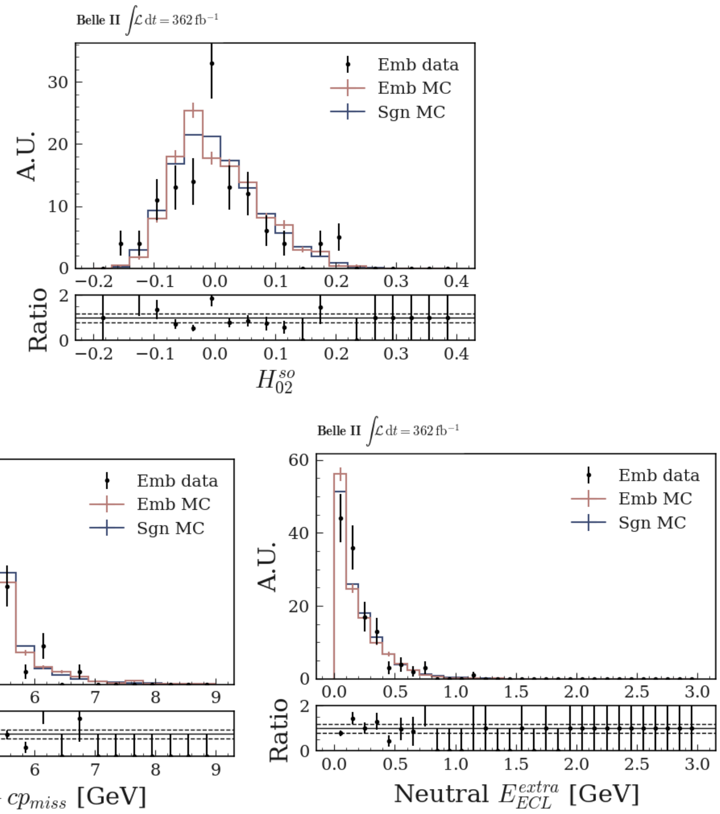


Figure 5.10: Distributions of the Kakuno-Super-Fox-Wolfram moment H_{o2}^{s0} (top right), sum of missing energy and momentum computed in the CMS (bottom left) and sum of the extra energy in the calorimeter (bottom right) for simulated signal (light blue histogram), simulated embedded sample (red histogram), and embedded data (points). The distributions are normalized to the number of events in data. No best candidates selection is applied, distributions appear as they are inputed to the classifier.

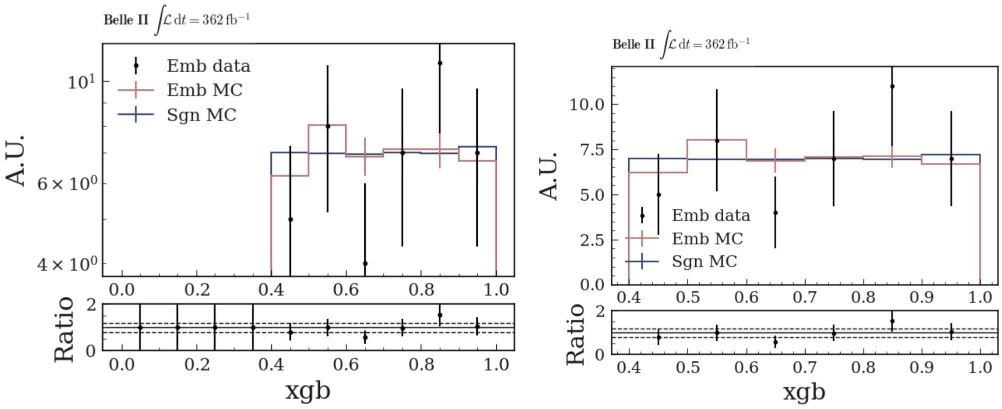


Figure 5.11: Classifier output distribution in the full range (left) and in signal region (right) for simulated signal (light blue histogram), simulated embedded sample (red histogram), and embedded data (points). The distributions are normalized to the number of events in data. Best candidate selection is applied.

1877 As shown inTable 5.7, the data-simulation ratio at pre-selection level is around 0.67
1878 and is consistent with the ratio found at the end of the selection. As a consequence,
1879 in the next steps of the analysis, 0.67 is used as calibration factor for the signal effi-
1880 ciency and an uncertainty of 16% (from the efficiency ratio in the BDT signal region
1881 after best candidate selection selection) will be considered as systematic uncertainty.

Sample	pre-selection	Signal search region
Data	$1.71 \pm 0.15\%$	$0.58 \pm 0.09\%$
Simulation	$2.51 \pm 0.06\%$	$0.96 \pm 0.04\%$
Ratio	0.68 ± 0.06	0.60 ± 0.10

Table 5.7: Selection efficiency in the signal region for the embedded data and simulated samples at different stages of the reconstruction and selection.

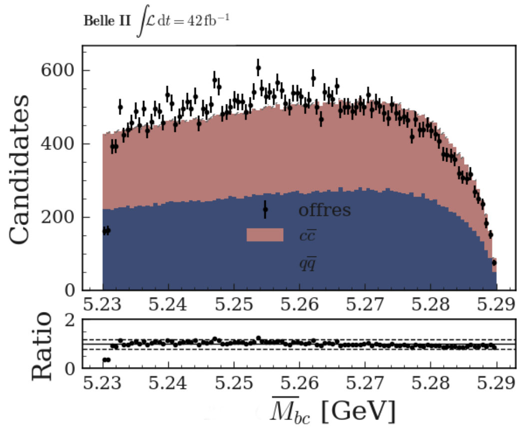


Figure 5.12: Distribution of modified M_{bc}. The off-resonance data is altered to mimic the on-resonance continuum. Distributions are normalized to the same number of events.

5.6.2 $q\bar{q}$ background validation using off-resonance data

After validating the behavior of signal events throughout the selection process, we want to verify whether the continuum simulation provides a good description of the off-resonance data. We use off-resonance data corresponding to 42 fb^{-1} of integrated luminosity.

The background yield in the signal region is evaluated by using continuum simulation. Indeed, a large part of the background contributions is continuum light $q\bar{q}$ and $c\bar{c}$. Generic simulated continuum samples can be corrected by comparing them to off-resonance data.

This comparison relies on the assumption that the kinematic features of the continuum events do not appreciably depend on the beam energy. Instead some variables directly related to the beam energy should be modified accordingly to allow comparisons. For this reason, the beam constrained mass of the B_{tag} candidate, M_{bc}, is modified in the off-resonance sample to mimic the on-resonance distribution:

$$\tilde{M}_{bc} = \sqrt{\left(\frac{E^*_{ON}}{2}\right)^2 - \left(\frac{E^*_{ON}}{E^*} \cdot p^*_{B_{tag}}\right)^2}, \tag{5.9}$$

where E^*_{ON} is the nominal beam energy in the on-resonance data (10.58 GeV) in the CMS, E^* is the beam energy of the considered event in the CMS and $p^*_{B_{tag}}$ is the momentum of the B_{tag} in the CMS. After this , the data-simulation comparison for M_{bc} is shown in Figure 5.12. We use the total off-resonance data sample, as well as the simulated continuum sample corresponding to 1 ab^{-1} of integrated luminosity.

The data-simulation agreement for the classifier input variables distributions is quite satisfactory. The distribution of the most discriminative variable related to missing quantities, $E^*_{miss} + cp^*_{miss}$, is shown in Figure 5.13 (top). Nevertheless, an event-by-event correction is further applied: a classifier (noted BDT$_c$) is trained after the preselection (with a relaxed selection on the modified M_{bc}: $\tilde{M}_{bc}{>}5.23$ GeV/c^2)

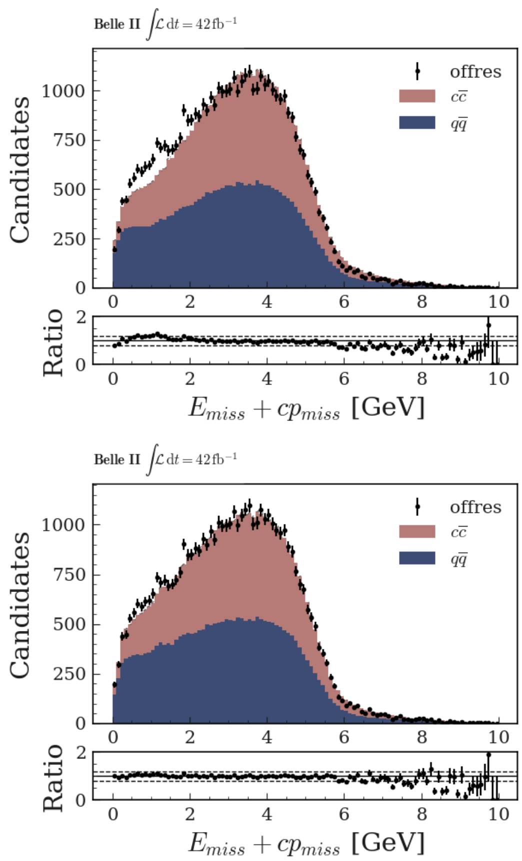


Figure 5.13: Distributions of $E^*_{miss} + p^*_{miss}$ for off-resonance data and continuum simulation before (top) and after (bottom) BDT$_c$ reweighting. Distributions are normalized to the same number of events.

using the off-resonance data as signal and the continuum simulation as background. The input variables are the same as in the main classifier described in Subsection 5.4.2, except for M_{bc}, which is removed. The BDT$_c$ classifier provides as output a factor p per candidate and a correction weight $p/(1-p)$ is applied on a candidate-by-candidate basis to the simulated continuum events. The result is an improved agreement, as shown in Figure 5.13 (bottom). The data/simulation comparison of all the other variables used as input for the main classifier are shown in Appendix B).

After the reweighting and the tighter selection cut $\tilde{M}_{\rm bc} > 5.27$ GeV$/c^2$, the overall data-simulation ratio is equal to 0.82 ± 0.01. This value is used to reweight $c\bar{c}$ and light $q\bar{q}$ events before the main classifier training (see Section 5.4). The same ratio, computed after the classifier output selection described in Section 5.5 and best candidate selection, is equal to 1.5 ± 0.5. This is consistent with the correction factor obtained at pre-selection level. For this reason, 0.82 is kept as a normalization factor for the continuum component in the rest of the analysis and a 50% uncertainty, coming from the data-simulation ratio computed in the classifier signal region, is assigned to this correction.

5.6.3 Background validation using on-resonance data

Finally, the data-simulation agreement for the input variables for both continuum and $\Upsilon(4S)$ samples is performed on on-resonance data. To be sure to comply with the blinding procedure described in Section 4.7, we aim at defining control samples with as few pollution from our signal as possible, we identify:

- A wrong B-meson charge sideband: the signal kaon and B_{tag} are requested to have the same charge.
- A particle ID sideband: the signal kaon is requested to have kaonID>0.1 and pionID>0.5.

These sideband samples are built using the particle identification methods described in Subsection 2.7.2. In these sidebands, the $B\bar{B}$ simulated samples are corrected in normalization with the overall factor extracted from the embedding procedure (0.67, see Subsection 5.6.1). The $q\bar{q}$ and $c\bar{c}$ simulated samples are corrected by using the off-resonance data, both in the normalization, with a factor 0.82, and in the shape of the distributions with the candidate by candidate weights obtained with the use of the BDT$_c$ (see Subsection 5.6.2).

The sideband data and simulation samples are processed through the nominal classifier of the analysis. The classifier output restricted to the signal region is shown in Figure 5.14. On the top panel, a comparison of the simulation between sideband and nominal samples is shown. On the bottom panel, data-simulation comparison in each sideband is reported.

From these samples, data/simulation ratios are computed to correct potential remaining discrepancies: ratios of 1.6 ± 0.6 for the wrong B-meson charge sideband and 1.24 ± 0.27 for the particle ID sideband are found. These ratios agree with each other and are compatible with unity, meaning that the corrections already applied on $B\bar{B}$ and $q\bar{q}$ cover data-simulation differences. Therefore, no further correction is applied on the general background normalization. The relative uncertainties on the ratios are 38% for the wrong B-meson charge and 22% for the particle ID control samples. Finally, a 30% uncertainty on the $B\bar{B}$ component normalization is assigned as systematic uncertainty.

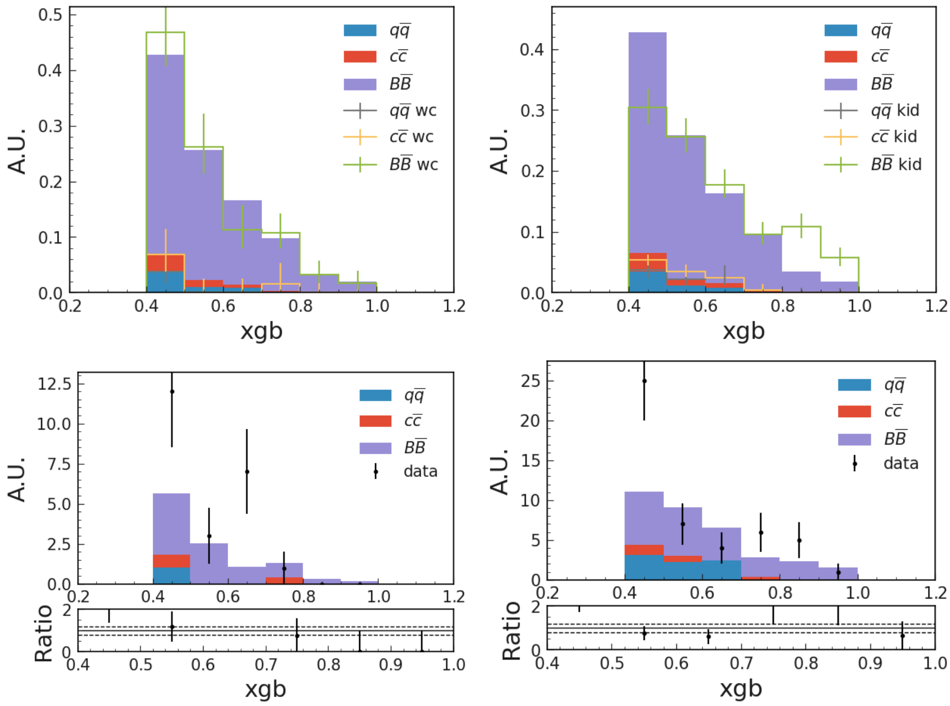


Figure 5.14: (top) Classifier output in the BDT signal region for nominal and sideband simulation and (bottom) data-simulation comparison in the BDT signal region. The distributions on top are normalized to unity. While the distributions on bottom are normalized to the same number of events. Wrong charge sideband is on the left, kaonID sideband is on the right.

5.7 Systematic uncertainties

In this section, we describe how we evaluate the diffrent systematic uncertainties that enter the measurement of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$. These systematic uncertainties affect the likelihood model described in Section 4.4 through a set of nuisance parameters, which cause variations with respect to expectations in the bins of the signal region.

Systematic uncertainties come from physical processes mismodelling and detector interaction mismodelling. The uncertainties considered in the statistical model are:

- Particle ID selection modeling uncertainty for the signal K^+, described in Subsection 5.7.1.
- Tracking efficiency modeling. Detailed in Subsection 5.7.2, this is only relevant for the signal kaon track, as tracking modeling is already taken into account for the FEI reconstructed B_{tag}.
- Branching fractions of the leading B^0 and B^+ background decays, which are varied according to their PDG uncertainties. Described in Subsection 5.7.3, furthermore Subsection 5.5.3 provides a detailed categorisation of these decays.
- Form factor uncertainties derived from Ref. [34] (Detailed in Subsection 5.7.4).
- Modeling of the low-multiplicity decay $B^+ \to K^+n\bar{n}$ involving neutrons and kaons in the final state. A study on this background is described in Subsection 5.7.5.
- Modeling of the signal-like $B^+ \to K^+K^0\bar{K}^0$ decay, described in Subsection 5.7.6.
- Branching fractions of B-mesons decays to excitations of D-mesons (D^{**}), as discussed in Subsection 5.7.7.
- Correction on the number of photon in the event to mitigate data/simulation discrepancies. This is described in Subsection 5.7.8.
- Difference between simulation and data embedded samples for the signal selection efficiency study. The correction factor derived in Subsection 5.6.1 is applied. Due to a small sample size it is not possible to derive a normalization variation from the control sample but an uncorrelated bin-by-bin variation on the efficiency correction, according to its error, is allowed.
- The number of $B\overline{B}$ events used as input in the measurement of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$, which corresponds to 387.1×10^6 pairs with an uncertainty of 1.5%. For the continuum normalization, cross section and luminosity are needed. The uncertainty on the latter is computed centrally for the whole collaboration and is of the order of 1%, we consider this to be included in the overall continuum normalization factors.

- Background contributions from B-meson decays involving a direct K^+ production. These become prevalent in the high sensitivity area of the signal search region. Out of these decays, $B^+ \to K^+ D^{(*)0/-}$ are of particular interest due to a relevant and less-known fraction of charmed mesons decays involving K^0_L mesons [112]. These decays are studied in the particle ID sideband described in Subsection 5.6.3, and are scaled by 30%(±10%).

- Uncertainty on the estimated background yield and background shape: for $c\bar{c}$ and light $q\bar{q}$ a 45% uncertainty in the normalization is considered, coming from the BDT reweighting described in Subsection 5.6.2. In addition, for the $B\bar{B}$ component, a normalization uncertainty of 30% is applied (details in Subsection 5.6.3).

- K^0_L reconstruction efficiency, studied centrally by the Belle II collaboration. From these studies, we derive a 17% uncertainty all signal and background components.

5.7.1 Particle identification

One source of systematic uncertainty comes from the particle identification requirement to select the signal kaon candidate. Simulated events (signal and backround alike) are given a weight correcting for discrepancies between data and simulation particle identification. These weights are provided by the Belle II performance group for a collaboration-wide use. The PID weights are defined in bins of p_T (transverse momentum) and $cos(\theta)$ (cosine of the polar angle of the associated track) of the K^+ candidate. In addition, uncertainties on the weights values are also provided, they are then propagated to our statistical model:

- For each event e present in the signal region, a series of 500 replicas i are produced following the method described in Subsection 4.5.1, computing modified PID weights values based on the associated PID weight uncertainty.

- From sums of the PID weights for each event category and signal region bin, the covariance matrix C_{PID} is computed as described in Subsection 4.5.2. A representation of C_{PID} can be seen in Figure 5.15.

- The Single Value Decomposition method described in Subsection 4.5.2 is used to identify the three eigenvectors associated to the three largest eigenvalues of C_{PID}. Each eigenvector is then added to the likelihood model with an associated nuisance parameter θ_i^{PID}, $i = 1, 2, 3$.

- The remaining elements of C_{PID} decomposition (see Equation 4.24) are added in quadrature to the uncorrelated systematic uncertainties shown in Subsection 5.8.1.

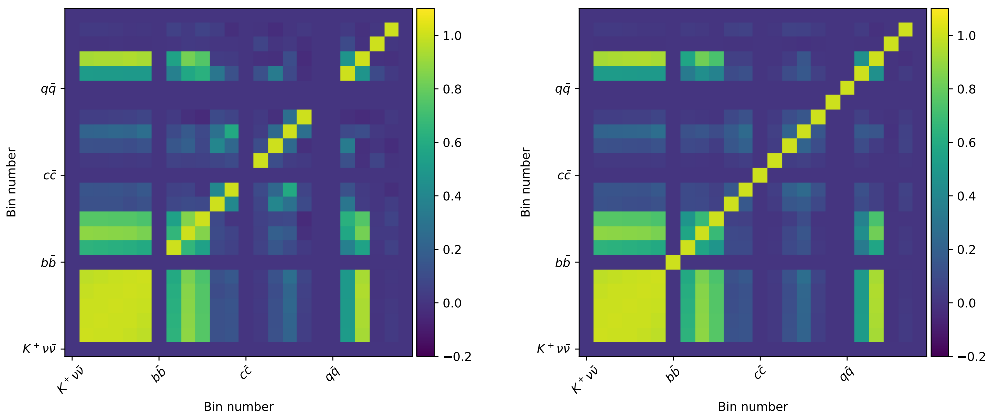


Figure 5.15: Correlation matrix between the expected yields in the different signal region bins. The signal search region is composed of 6 bins, the expected yields are observed in 4 simulated samples (3 background samples and 1 signal sample), thus the matrix is of size 24×24. The left figure shows the original correlation matrix, while the right figure shows an approximation of said matrix obtained by a truncation of the covariance eigen-decomposition described in Equation 4.24. Here, the 3 eigenvectors associated to the 3 largest eigenvalues are used for the decomposition.

5.7.2 Tracking efficiency

A systematic uncertainty comes from a possible inacurate modeling of the track finding efficiency in simulation. As mentionned before, this effect only needs to be estimated for the reconstructed tracks taken as the signal K^+ candidate, as the tracking efficiency uncertainty is already taken into account for the other tracks in the event through the FEI algorithm. Following guidelines from dedicated studies performed by the Belle II tracking group [113], we assign an uncertainty of 0.9% on the track-finding efficiency which translates to a 0.9% uncertainty on the signal normailization introduced in the model.

5.7.3 Branching fraction of leading backgrounds

One source of uncertainty comes from the measurement of the B meson decays making up the $B\bar{B}$ background. The generalities about the signal region background composition have been described in Section 5.5. The study described here is based on the full $B\bar{B}$ sample described in Section 5.1.
The associated uncertainty arises from the values of the branching ratios used to generate such decays in the simulation. To account for this uncertainty, we derive nuisance parameters in the likelihood model by varying the branching ratios values of the decays populating the signal region, based on their nominal values and associated uncertainties taken from [66]. The uncertainties on the branching ratios values are then propagated to the likelihood model as follows:

1. A set of branching ratios and associated uncertainties corresponding to the

leading $B\bar{B}$ decays in the signal region is created. $\sim 80\%$ of $B^{\pm}$ decays and $\sim 60\%$ of $B^0/\bar{B}^0$ decays appear in this set.

2. For each event e present in the signal region, a series of 1000 replicas i are created. For each replica, a modified branching ratio value $Br_{i,n}(e)$ is computed from $Br_n(e)$ with associated weights $w^e_{i,n}$ as described in Subsection 4.5.1, where $n \in \{0,1\}$ corresponds to the index of the B meson considered in the pair and $Br_n(e)$ is the nominal value of the branching ratio for the decay of the B meson considered. Decays not present in the set of decays studied are assigned a weight of 1.
   Finally, for each replica, a single weight w^e_i is computed as : $w^e_i = w^e_{i,0} * w^e_{i,1}$.

3. The bins of the likelihood fit (Subsection 5.8.1) are filled with the replicas according to the bin value of e. The end result is an array of 6 elements. Each element contains an array $S_j, j \in \{1,..,6\}$ of 1000 values, corresponding to the number of counts to the associated bin observed in a replica.

4. The bin-by-bin covariance is computed over the $N = 1000$ replicas as described in Subsection 4.5.2

5. Three eigenvectors of the covariance matrix corresponding to the three largest eigenvalues are used to define variation vectors (see Subsection 4.5.2 and 4.4), each variation vector is incorporated in the likelihood model with an associated nuisance parameter.

5.7.4 Signal form factors

We described in Section 1.3 how the Standard Model form factor $f_+(q^2)$ is needed to compute the signal branching fraction as a function of q^2. However, the simulated signal events are generated based on a uniform phase space for the decay products. Thus, we introduce a correction to properly take the form factor contribution into account. This correction is then treated as an additional source of systematic uncertainty.
The form factor $f_+(q^2)$ has been parametrised using three real values $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ with corresponding uncertainties $\sigma = (\sigma_0, \sigma_1, \sigma_2)$, for which the associated covariance matrix C_α has been computed (*cf.* Equation 1.25, 1.27, 1.28 and 1.29). The uncertainties σ_i are then propagated to the statistical model:

- C_α is decomposed using the Single Value Decomposition method described in Subsection 4.5.2 to extract the three unit uncertainty eigenvectors v_1, v_2, v_3 as well as their respective eigenvalues e_1^2, e_2^2, e_3^2.

- Modified form factors are then computed as $f_+(q^2, \alpha + e_i\mathbf{v}_i)$

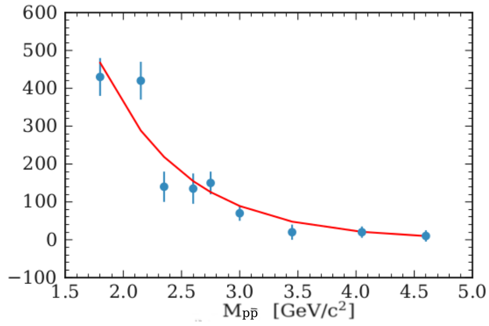


Figure 5.16: Result of an exponential fit of $M(p\bar{p})$ obtained in $B^+ \to K^+ p\bar{p}$ data from Ref [114].

- The expected number of signal events in the i-th bin of the signal search region associated to a given α, $N_i(\alpha)$ is computed.
- The three form factor modified vectors, $\delta_1, \delta_2, \delta_3$ are defined as:

$$\delta_i^{ff} = \begin{pmatrix} N_1(\alpha + \sigma_\mathbf{i}) - N_1(\alpha) \\ N_2(\alpha + \sigma_\mathbf{i}) - N_2(\alpha) \\ N_3(\alpha + \sigma_\mathbf{i}) - N_3(\alpha) \\ N_4(\alpha + \sigma_\mathbf{i}) - N_4(\alpha) \\ N_5(\alpha + \sigma_\mathbf{i}) - N_5(\alpha) \\ N_6(\alpha + \sigma_\mathbf{i}) - N_6(\alpha) \end{pmatrix}, \; i = 1, 2, 3 \tag{5.10}$$

  with each coefficient corresponding to a bin of the signal search region.

- The three modified vectors computed are added to the statistical model as described in Section 4.4 with their respective nuisance parameters θ_i^{ff}.

The variations due to this source of uncertainty are of the order of the percent.

5.7.5 Modeling of $B^+ \to K^+ n\bar{n}$

The decay $B^+ \to K^+ n\bar{n}$ is of particular concern in this analysis. Because neutrons are stable and do not interact with the detector, they can easily mimic the experimental signature of the neutrimo pair present in the signal. In addition, this decay has never been observed, even though its branching ratio can be predicted from isospin symmetry using $B^+ \to K^+ p\bar{p}$, which has been measured to be $\mathcal{B}(B^+ \to K^+ p\bar{p}) = 6.7(\pm 0.5 \pm 0.4) \times 10^{-6}$. The $B^+ \to K^+ n\bar{n}$ decay is modelled according to the 3-body phase-space in the standard Belle II simulation. However, [114] shows that this decay is expected to be enhanced at the $n\bar{n}$ threshold.
In order to model this enhancement, the data taken from [114] are fitted as shown in Figure 5.16. Afterwards, a dedicated 100.000 events $B^+ \to K^+ n\bar{n}$ sample is

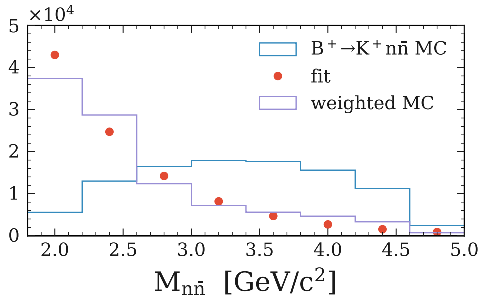


Figure 5.17: $M_{n\bar{n}}$ distribution in simulated $B^+ \to K^+n\bar{n}$ events. Red points indicate fit results from Figure 5.16, blue histogram corresponds to phase-space MC and the magenta histogram is obtained after applying threshold enhancement.

produced and reweighted (from the original phase space modeling to the aforementionned fit), as seen in Figure 5.17. This enhancement has a significant impact on the background rejection.
This modification of the modeling for $B^+ \to K^+n\bar{n}$ events is propagated through the computation of the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ by reweighting the $B^+ \to K^+n\bar{n}$ events in the $B\bar{B}$ background sample.
Furthermore, a systematic uncertainty corresponding to 100% of the correction is applied to cover potential additional mismodeling.
The uncertainty is treated using a single correlated systematic uncertainty source that affects the B^+B^- background. The way in which systematic sources are accounted for in the fit is summarised in Table 5.8, dominant sources are due to the uncertainty on the $B\bar{B}$ normalization and the signal efficiency.

5.7.6 Modeling of $B^+ \to K^+K^0\overline{K^0}$

Similarily to $B^+ \to K^+n\bar{n}$, the $B^+ \to K^+K^0\bar{K}^0$ can also pollute the signal search region. Three final states from this decay need to be considered: $B^+ \to K^+K^0_LK^0_L$, $B^+ \to K^+K^0_SK^0_L$, and $B^+ \to K^+K^0_SK^0_L$.
K^0_L are a general issue in the search for $B^+ \to K^+\nu\bar{\nu}$ as they easily go undected and create sources of missing energy. Decay modes with K^0_S also contribute to the background composition of the signal region for a similar reason, albeit to a lesser extent. In the Belle II simulation, $B^+ \to K^+K^0\bar{K}^0$ are generated using the phase-space dependence of their branching ratios. An additional set of resonant modes are considered and treated independently. Hower, a more accurate prediciton of the differential branching ratio for the $B^+ \to K^+K^0_SK^0_S$ decay mode can be found in [115]. Assuming isospin asymmetry, we expect the same behavior for the $B^+ \to K^+K^0_LK^0_L$ decay mode. We proceed to assign weights to the relevant events following the prescriptions from [115].
Finally the $B^+ \to K^+K^0_SK^0_L$ final state is treated separately since intermediate scalar resonances cannot decay to the CP odd $K^0_SK^0_L$ pair. In this case, weights are

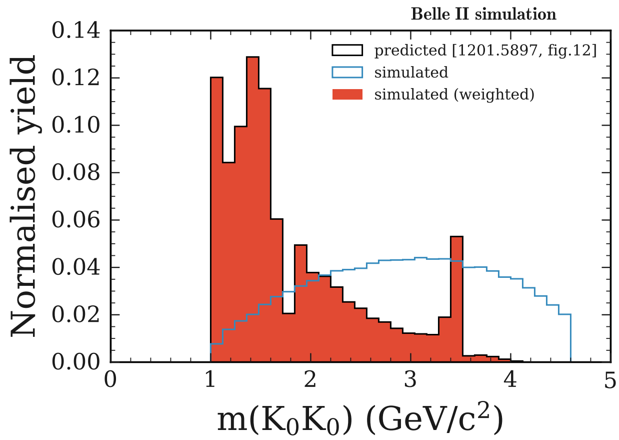


Figure 5.18: Density of simulated $B^+ \to K^+K^0\overline{K}^0$ events (without any selection) in bins of the invariant mass of the $K^0\overline{K}^0$ system. The blue histogram corresponds to events where the decay is simulated according to the phase space. The black histogram corresponds to the predictions of [115]. The red histogram shows the result of the reweighting of the simulated events. By construction, the red and black histograms exactly overlap.

derived from the amplitude analysis described in [115].
In the case of $B^+ \to K^+K^0_SK^0_S$ and $B^+ \to K^+K^0_LK^0_L$, a binned reweighting procedure is used, based on the distribution of the invariant mass of the two-kaon system, in order to match expectations, as seen in Figure 5.18.

For the $B^+ \to K^+K^0_SK^0_L$ final state, the decays are modelled as a sum of $B^+ \to K^+\ \phi^0$ resonances and a non-resonant p-wave contribution described in [115]. The resonant contribution is taken directly from the Belle II simulation, checking the branching ratio value against the world average from [66]. The p-wave contribution is taken into account by applying weights (Figure 5.19) to the phase space simulation.

Following this correction, the total expected simulated $B\bar{B}$ sample yield in the signal region defined in Section 5.5 goes up by 0.81%
The uncertainty associated to the correction is then estimated. For each *BDT* output bin i of the signal region, the relative uncertainty $u_r(i)$ is computed as:

$$u_r(i) = \frac{\nu_B(i) - \nu^*_B(i)}{\nu_B(i)}, \tag{5.11}$$

where $\nu_B(i)$ is the expected $B\bar{B}$ yield in the bin i before the correction and $\nu^*_B(i)$ is the expected $B\bar{B}$ yield for the same bin after the correction. The uncertainty is then propagated to the satistical model through a vector $\boldsymbol{U}$ containing the correlated $u_r(i)$ as described in Section 4.4 with an associated nuisance parameter θ^{3K}.

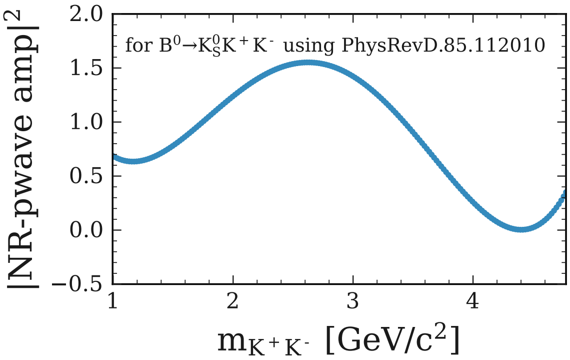


Figure 5.19: Amplitude squared for the p-wave contribution as a function of K^+K^- invariant mass. Based on [115].

5.7.7 Modeling of $B \to D^{**} + X$ decays

As shown in Subsection 5.5.2, the main background contribution in the signal search region comes from $B\bar{B}$ pairs where at least one B-meson decays as $B \to D^{(*)/(**)+X}$. The cases including D^{**} mesons are especially problematic, as they are less known experimentally and are handled by `PYTHIA` [84] (here, D^{**} refers to one of the following excited states: $D_0^*(2300)^+$, $D_0^*(2300)^0$, $D_1^*(2420)^+$, $D_1^*(2420)^0$, $D_1(H)^+$, $D_1^*(2430)^0$, $D_2^*(2460)^+$, $D_2^*(2460)^0$, $D_{s0}^*(2317)^+$, $D_{s1}^*(2536)^+$, $D_{s1}^*(2460)^+$ and $D_{s2}^*(2573)^+$).
These events represent 3% and 5% of the simulated B^+B^- and $B^0\bar{B}^0$ background samples respectively. We apply a 50% systematic uncertainty on the value of the branching ratios of the relevant decays to account for potential mismodeling.

5.7.8 Photon multiplicity correction

Even though the selection on the ECL clusters used to compute the neutral ECL extra energy NE_{ECL}^{extra} and the photon multiplicity N_γ (subsubsection 5.4.1.1) is devised to minimize data-simulation disagreement, discrepancies are observed in these distributions. These discrepancies are expected to come mainly from background simulation, as they are seen in the sidebands described in Subsection 5.6.3 and are only minimal in the embedded samples (Subsection 5.6.1). Figure 5.20 shows the distributions of interest for the different samples.

In order to derive a correction, the sideband that best describes the background distribution in the signal search sample (on-resonance events passing the selection described up to Section 5.3, before the selection on the BDT output is performed) for both NE_{ECL} and N_γ is identified (Figure 5.21). The wrong B-meson charge sideband is chosen to derive the correction applied to the right B-meson charge sample while the particle ID sideband is used for validation.

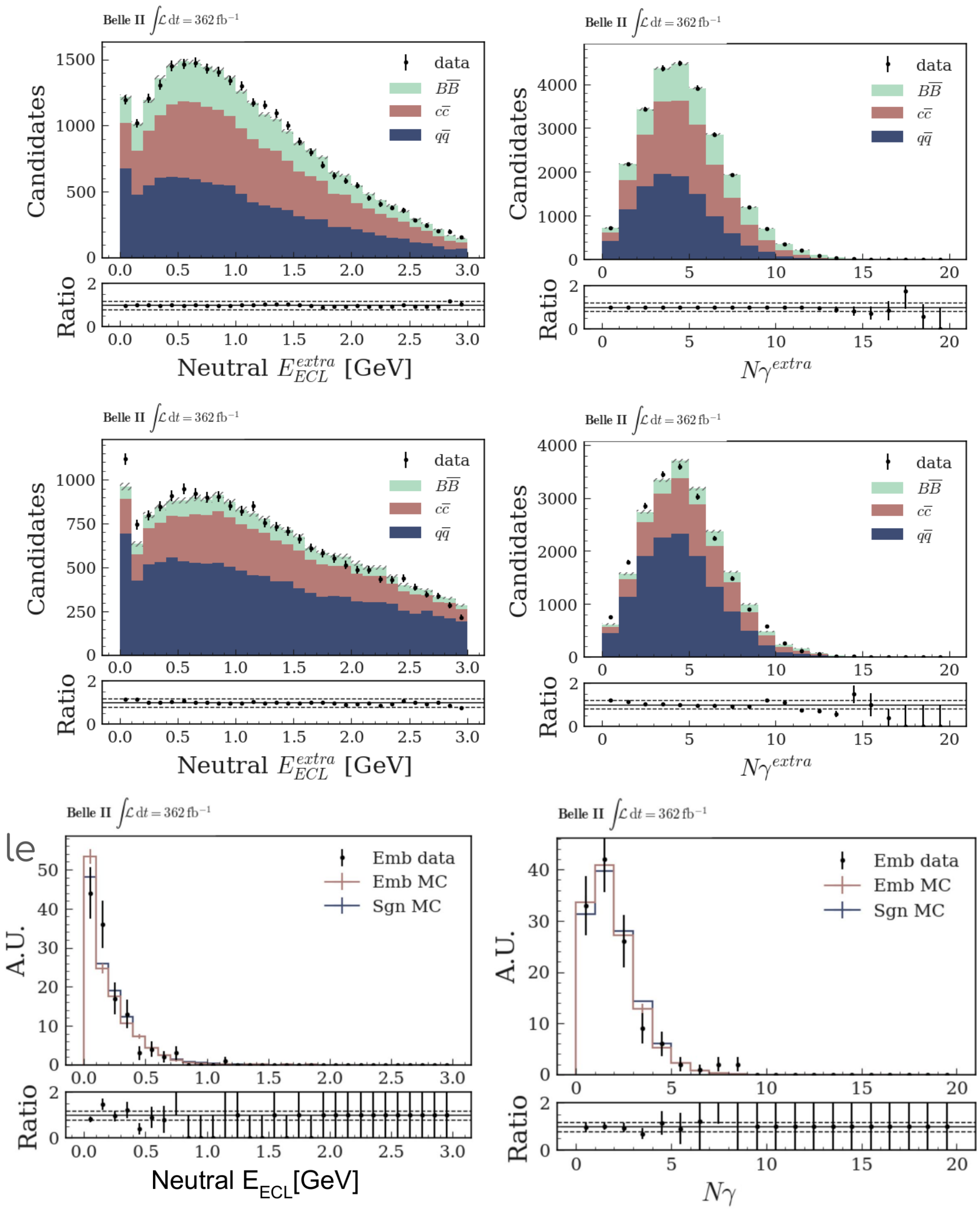


Figure 5.20: Distribution of the neutral ECL extra energy NE_{ECL}^{extra} and the photon multiplicity N_γ for the simulated (filled) and data (points) wrong B-meson charge sideband samples (top) and for the particle ID sideband samples (middle). The distribution for the same variables are shown for the embedded $B^+ \to J/\Psi K^+$ simulated (red) and data (points) samples as well as for the signal (blue) simulated sample (bottom).

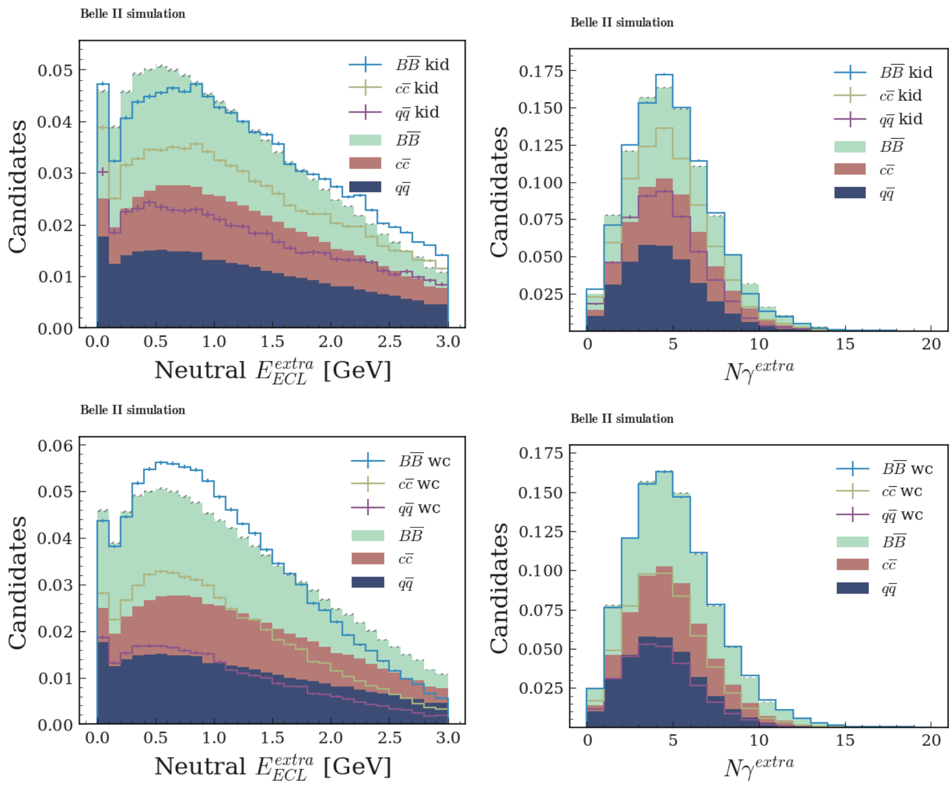


Figure 5.21: Distribution of the neutral ECL extra energy (left) and N_γ (right) for the particle ID (top, line histograms) and wrong B-meson charge (bottom, line histograms) sidebands. The filled histograms show the relevant distributions for the simulated background samples passing the signal selection, before the *BDT* output cut.

For each N_γ value in the wrong B-meson charge sample, we compute the weight:

$$w_{N_\gamma} = \frac{n_D(N_\gamma)}{n_S(N_\gamma)}, \tag{5.12}$$

where $n_D(N_\gamma)$ and $n_S(N_\gamma)$ correspond to the number of expected background events with N_γ extra photon candidates, in data and simulation respectively. The events in the signal region of the right B-meson charge sample are then weighted based on the associated N_γ value.

The correction is then validated using the particle ID sideband sample. The sample is further divided into wrong B-meson charge and right B-meson charge. Then, the correction process is repeated as described before with the weight defined in Equation 5.12 computed using the wrong B-meson charge subsample of the particle ID sideband. The events in the right B-meson charge subsample are then reweighted accordingly. Figure 5.22 shows the effect of the correction in this sample.

Although an improvement is seen in the control sample after applying the correction, some residual discrepancies persist. This indicates that the data-simulation

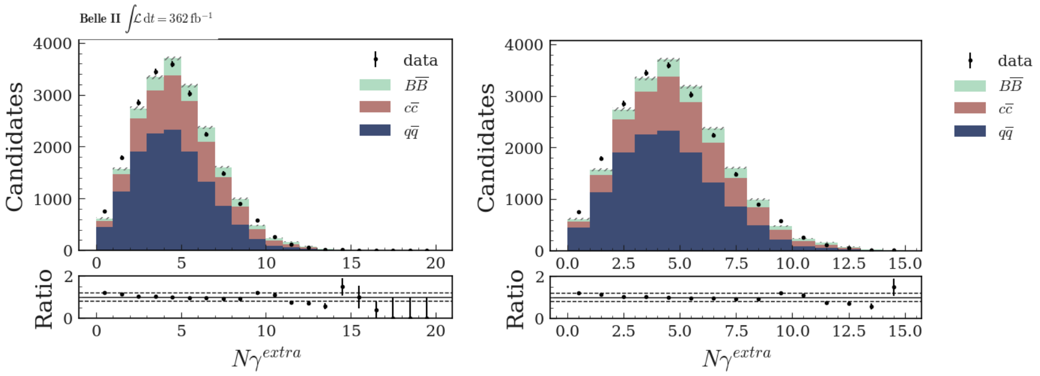


Figure 5.22: Distribution of the photon multiplicity N_γ for the right B-meson charge subsample of the particle ID sideband before (left) and after (right) correction. Distributions are shown for simulated background events (filled) and data (points).

disagreement in the wrong B-meson charge sample might be slightly different with regards to the right B-meson charge sample. In order to account for this effect, we choose to assign an associated systematic uncertainty corresponding to $\pm 100\%$ of the correction.
Finally, even though the data-simulation agreement in the embedded samples seem acceptable, the size of the data sample is low which limits the comparison. To cover for a potential discrepancy, the simulated signal sample is also corrected using the method described previously. And the associated systematic is assigned.

5.7.9 Summary

Table 5.8 lists the different systematic uncertainty crontributions to the statistical model.

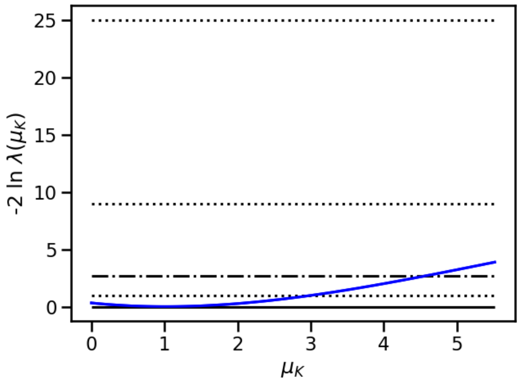


Figure 5.23: Distribution of the likelihood-ratio test Λ_μ, expecting a SM value for the $B^+ \to B^+\nu\bar{\nu}$ branching ration. The different horizontal lines correspond (from top to bottom), to the $5\sigma, 3\sigma, 90\%$ CL and 1σ levels.

5.8 Results

We now possess all the ingredients needed to measure the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$. We have defined the signal search region in Section 5.5 as well as the different sources of systematic uncertainties in Section 5.7. The expected yields in the signal region are computed for the four different event classes (signal, $B\bar{B}$, $c\bar{c}$ and $q\bar{q}$) using the full simulated samples described in Section 5.1, weighted to match the on-resonance data sample integrated luminosity.

At this point in time, the analysis is still kept blind (Section 4.7). Thus, we describe in Subsection 5.8.1 how the branching fraction value for $B^+ \to K^+\nu\bar{\nu}$ is computed, providing an expected measurement based on simulated samples. In Subsection 5.8.2, the expected measurement is compared to previous results.

5.8.1 Signal extraction setup

From the likelihood $\mathcal{L}(\mu, \boldsymbol{\theta})$, the expected upper limit on the value of $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ is computed as described in Section 4.6. Expecting SM value for $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$, Figure 5.23 shows the distribution of the likelihood-ratio test Λ_μ defined in Equation 4.26. From this, we extract the expected upper limit:

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 2.3 \times 10^{-5} \tag{5.13}$$

at 90% confidence level. The significance level α_0 of the associated signal strength μ is computed as:

$$\alpha_0 = \sqrt{2 \ln \mathcal{L}(\mu = 0) - 2 \ln \mathcal{L}(\mu = \mu_{min})}, \tag{5.14}$$

we extract the significance $\alpha_0 = 0.55$ to reject the null hypothesis.

Source	Affected category	Treatment	Size
Kaon-ID	sig, $B\bar{B}$, cont	3 component-correlated + 1 uncorrelated	
tracking	sig	normalization	0.9%
signal efficiency	sig	uncorrelated	16%
signal BF form factors	sig	3 bin-correlated	
dominant $B\bar{B}$ background BF	$B\bar{B}$	bin-correlated	
$q\bar{q}$ normalization	cont	normalization	50%
$B\bar{B}$ normalization	$B\bar{B}$	normalization	30%
$q\bar{q}$ shape	cont	component-correlated	100% of correction
Extra photon multiplicity correction	sig, $B\bar{B}$, cont	component-correlated	100% of remaining discrepancy in kaon ID sideband
K_L efficiency	sig, $B\bar{B}$, cont	component-correlated	17%
Threshold enhancement for $B^+ \to K^+ n\bar{n}$	$B\bar{B}$	bin-correlated	100%
Branching fraction for $D \to K_L X$	$B\bar{B}$	bin-correlated	10%
Branching fraction for $B^+ \to K^+ K_L K_L$	$B\bar{B}$	bin-correlated	20%
Branching fraction for $B \to D^{**}$	$B\bar{B}$	bin-correlated	50%
number of $B\bar{B}$ pairs	sig, $B\bar{B}$	normalization-correlated	1.5%

Table 5.8: Summary of the systematic uncertainties treatment. The "cont" category includes both $c\bar{c}$ and $q\bar{q}$ samples.

Experiment	Uncertainty on $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ ($\times 10^{-6}$)
Belle semileptonic	5.7
Belle hadronic	16
BaBar semileptonic	8.0
BaBar hadronic	13.5
BaBar combined	6.5
Belle II inclusive	16
Belle II hadronic expected	9.5

Table 5.9: Measured uncertainties on the branching fraction for this and published results.

Experiment	Uncertainty on $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ ($\times 10^{-6}$)
Belle semileptonic	8.0
Belle hadronic	23
BaBar semileptonic	8.9
BaBar hadronic	15.0
BaBar combined	7.2
Belle II inclusive	6.4
Belle II hadronic expected	9.5

Table 5.10: Measured uncertainties on the branching fraction for this and published results scaled to the luminosity of 362 fb^{-1} assuming $1/\sqrt{\mathcal{L}}$ dependence.

5.8.2 Comparison with previous measurements

An uncertainty on the value of the signal strength μ can be derived from the statistical model. This allows to compute an uncertainty on the value of the branching fraction of the $B^+ \to K^+\nu\bar{\nu}$ decay. Table 5.9 presents a comparison of the branching fraction uncertainty from the previous analyses of Belle [63, 64], BaBar [62] and Belle II [65] with the expected uncertainty for this analysis. Table 5.10 provides similar information, with uncertainties from the previous experiments scaled as $\sqrt{\mathcal{L}/362\ \text{fb}^{-1}}$ to the luminosity of this analysis. For Belle, the uncertainties on the branching fraction are obtained using published information on the signal yield and signal selection efficiency.

The expected preliminary results for this analysis are very competitive with previous publications. The main improvements compared to the previous Belle hadronically-tagged result come from a higher tagging efficiency and better performance of the final BDT selection.

Chapter 6

Conclusion

This thesis has described the first search for the $B^+ \to K^+ \nu\bar{\nu}$ decay using a hadronic tagging method in the Belle II experiment, as well the development of an algorithmic method to improve the spatial resolution of the experiment's Silicon Vertex Detector.

Chapter 1 discussed how the $B^+ \to K^+ \nu\bar{\nu}$ decay is predicted in the Standard Model of particle physics, operating through a suppressed flavour changing neutral current quark transition, as well as how its branching fraction can be computed in said model. In addition, possible beyond Standard Model conntribution to this decay have been described, showing how an experimental determination of the branching fraction value can help to constrain new physics models.

Chapter 2 showed an overview of the experimental apparatus used to perform the works presented. This apparatus consists in SuperKEKB accelerator, colliding electron/positron at the $\Upsilon(4S)$ resonance in order to produce pairs of B-mesons, as well as the Belle II detector used to study said collisions.

Chapter 3 described the way the spatial resolution of Belle II's vertex detector is estimated, as well as an algorithmic method, the *cluster unfolding*, designed to correct for charge sharing between silicon strips. This method has been introduced to improve the performances of the detector, as well as to reduce the discrepancies seen in the spatial resolution estimation between data and simulation. This method allows to improve the detector's spatial resolution by 5 to 15% for specific sensors.

Chapter 5 presented the full analysis developed to perform the search for the $B^+ \to K^+ \nu\bar{\nu}$ decay using a data sample of 362 fb^{-1} equivalent integrated luminosity at the $\Upsilon(4S)$ resonance and 42 fb^{-1} collected 60 MeV below. The selection of events of interest has been described, as well as the hadronic method employed to reconstruct B-mesons in said events. The sanity of the analysis has also been tested on several control samples, and the different systematic uncertainty contributions to the expected measurement have been thoroughly evaluated.
Given the available datasets, the analysis is expected to put an upper limit on the value of the branching ratio $\mathcal{B}(B^+ \to K^+ \nu\bar{\nu})$ at 2.3×10^{-5} at 90% confidence level. The measurement is expected to be $\sim 30\%$ more precise that the world leading measurement for hadronically tagged searches for the $B^+ \to K^+ \nu\bar{\nu}$ decay published by the BaBar collaboration [62], and $\sim 40\%$ more than the previous Belle collaboration measurement [63].

Still, there remain many opportunities for the study of $B \to K^{(*)}\nu\bar{\nu}$ decays. New experimental methods are currently developed which would benefit the searches for these decays, such as the inclusive tagging method used by the Belle II collaboration in the search for the $B^+ \to K^+\nu\bar{\nu}$ decay using a reduced dataset of 63 fb^{-1} collected at the $\Upsilon(4S)$ resonance [65] .
The use of machine learning in tagging algorithms is also being studied, which could yield higher efficiencies in the studies of such decays. In addition, the search for the other $B \to K^{(*)}\nu\bar{\nu}$ decay modes $B^+ \to K^{*+}\nu\bar{\nu}$, $B^0 \to K^0\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ is also underway, using hadronic, semileptonic and inclusive tagging methods. Combining future results will allow to better understand the Standard Model of particle physics, as well as to constrain numerous new physics models.

Finally, in addition to improvement in methods, the new data planed to be collected by the Belle II and LHCb experiments in the future will surely allow to provide exciting flavour physics results.

Chapter 7

Résumé en Français

Mon travail de thèse s'est déroulé au sein de la collboration Belle II, regroupant plus de 1000 membres issus de 27 pays.

Au cours de ma thèse, j'ai développé une méthode complète visant à analyser les données collectées par la collaboration Belle II, afin de mettre en évidence un processus physique jamais observé. De plus, j'ai également prit part au fonctionnement et à l'amélioration de l'expérience Belle II en participant à l'amélioration des performances de l'un des détecteurs utilisé au cours de la prise de données. Ces deux axes de recherche originaux sont détaillés ci-après.

Recherche de la désintégration $B^+ \to K^+\nu\bar{\nu}$ au sein de l'expérience Belle II

La majeure partie de cette thèse est consacrée à la mesure de la désintégration $B^+ \to K^+\nu\bar{\nu}$. Ce procesus est décrit par le Modèle Standard (MS) de la physique des particules avec un rapport d'embranchement $\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) = (4.43 \pm 0.38) \times 10^{-6}$ [35]. Cette mesure est motivée d'une part par le fait que ce processus n'a jamais été mesuré et, d'autre part, car plusieurs modèles de Nouvelle Physique (NP, décrivant l'ensemble des théories non incluses dans le MS), prédisent des modifications du rapport d'embranchement du canal de désintégration $B^+ \to K^+\nu\bar{\nu}$ [40, 41, 44–50].

Le fait que la désintégration $B^+ \to K^+\nu\bar{\nu}$ n'ait à ce jour pas été observée peut être expliqué par son faible rapport d'embranchement ainsi que par les difficultés expérimentales liées à son observation. En effet, les deux neutrinos présents dans l'état final intéragissent très faiblement avec la matière, ils sont donc dans les faits "invisibles" pour nos détecteurs. Afin de mesurer un processus physique rare et partiellement invisible, il est nécessaire de tirer avantage d'un dispositif expérimental spécifique: dans notre cas, l'expérience Belle II.

L'expérience Belle II est composée de l'accélerateur SuperKEKB, permettant la production de nombreuses collisions e^+e^- à une énergie de 10.58 GeV. SuperKEKB détient actuellement le record du monde de luminosité instantanée ($4.7 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$) et a permis de collecter un échantillon de données correspondant à 424 $\text{fb}^{-}1$ entre 2019 et 2022. L'expérience Belle II est complétée par le détecteur Belle II construit autour du point de collision de SuperKEKB. Ce détecteur de forme cylindrique est formé de plusieurs couches de sous-détecteurs spécialisés, permettant de réaliser des mesures complètes des collisions produites (voir Figure 7.1). Ces sous-détecteurs sont, par ordre croissant de distance au point de collision:

- Le détecteur à pixels (PXD), utilisé pour reconstruire les vertexs produits par les collisions, est composé d'une couche de senseurs DEPFET. Une seconde couche a été installée courant 2023).

- Le détecteur de vertex à pistes de silicium (SVD). Ce détecteur est également utilisé pour la reconstruction de vertexs ainsi que pour la trajectographie et l'identification de particules et est composé de 4 couches de détecteurs à piste de silicium.

- La chambre à dérive (CDC) participe à la trajectographie et à l'identification des particules. Elle consiste en un volume gazeux complété par de nombreux fils métalliques servant à la détection de particules chargées éléctriquement.

- Le détecteur de temps de propagation (TOP) situé sur la partie cyllindrique du détecteur Belle II et le détecteur de radiation Cherenkov à aerogel (ARICH) situé aux extrémités axiales du détecteur forment le système d'identification des particules.

- Le calorimètre électromagnétique (ECAL), composé de cristaux de CsI(Ti) permet de reconstruire les particules électriquement neutres.

- Un aimant supraconducteur générant un champs magnétique de 1.5 T permet de modifier la trajéctoire des particules chargées afin de mesurer leurs impulsions.

- Le détecteur de K_L^0 et de muons, composé d'un sandwich d'épaisses couches de fer et de chambres RPC, fini de copléter le détecteur Belle II.

L'alliance du grand nombre de collisions e^+e^- produites par SuperKEKB, des performances du detecteur Belle II et de son herméticité font de Belle II la seule expérience de physique des particules sur collisioneur de sa génération capable d'observer le canal de désintégration $B^+ \to K^+\nu\bar{\nu}$. J'ai donc développé une chaîne d'analyse complète en utilisant les outils de Belle II afin d'observer la désintégration $B^+ \to K^+\nu\bar{\nu}$ pour la première fois en utilisant les données collectées avant l'été 2022.

Cette analyse tire profit de l'algorithme de *Full Event Interpretation* (FEI) développé par la collaboration Belle II [101]. L'énergie de collision de SuperKEKB étant fixée à la valeur nécessaire à la production de la resonance $\Upsilon(4S)$, se désintégrant selon le canal $\Upsilon(4S) \to B\bar{B}$, les évènements de collisions de signal sont composés de deux mésons B, l'un (B_{sig}) se désintégrant dans le canal $B^+ \to K^+\nu\bar{\nu}$ tandis que l'autre (B_{tag}) se désintègre de façon aléatoire. Le but de l'algorithme FEI est de reconstruire le B_{tag} selon la chaîne de désintégration la plus probable tirée d'une liste de plus de 10000 chaînes possibles. Cet algorithme a été développé spécifiquement pour l'étude de canaux de désintégrations impliquant des neutrinos. En effet, les détails de la collision étant précisément connus, la reconsutruction du B_{tag} permet, au travers de lois de conservations, d'accéder aux propriétés des neutrinos produits (et échappant à la détection).

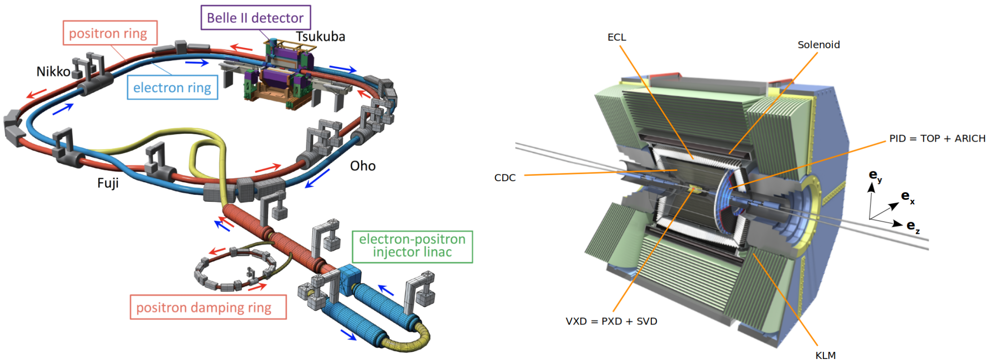


Figure 7.1: Vue schématique de l'accélérateur SuperKEKB (gauche) et du détécteur Belle II (droite). Adapté de [25].

Mon analyse utilise un arbre de décision boosté (BDT) entraîné à différencier entre les évènements de signal et de bruits de fond. Cet entraînement est effectué sur un échantillon simulé. Cette simulation reproduit les processus physiques issus des collisions ainsi que les performances du détecteur. Les évènements de signal et de bruits de fond sont différenciés en se basant sur 12 variables. Les variables présentant le plus fort pouvoir de discrimination étant l'énergie mesurée dans l'ECAL et non associée au B_{tag} reconstruit ou au kaon issu de la désintégration du B_{sig}, et la somme de l'énergie et de l'impulsion manquantes dans l'évènement. Les autres variables utilisées rendent compte de la distribution dans l'espace des différentes particules produites dans l'évènement, ainsi que de la cinématique du B_{sig}.
La structure du BDT est elle aussi optimisée, à l'aide du logiciel `optuna` [111], afin d'obtenir la classification la plus performante possible. Le BDT ainsi entraîné permet une bonne séparation entre évènements de signal et de bruits de fond, et le sur-entraînement du BDT est gardé à un niveau raisonnablement bas.
Il est alors possible de définir une région de signal basée sur la distribution de la variable de sortie du BDT: pour chaque évènement cette variable prend une valeur comprise entre 0 et 1: une valeur élevée traduit une forte probabilité que l'évènement en question soit un évènement de signal. La région de signal est alors définie comme l'ensemble des évènements ayant une valeur de sortie de BDT supérieure à 0.4. On s'attend dans cette région à trouver un maximum d'évènements de signal, tout en limitant la contamination des bruits de fond (Figure 7.2).

À ce stade le comportement de l'analyse n'est estimé que sur des échantillons simulés, il est alors nécessaire de s'assurer que la simulation décrit les résultats mesurés de manière satisfaisante, tout en évitant d'étudier la région de signal dans les données, afin de ne pas introduire de biais dans la mesure. Pour ce faire, différents canaux de contrôles sont définis:

- L'efficacité de la sélection pour le signal est difficile à estimer dans les données sans introduire de biais de mesure. Des évènements partiellement simulés sont

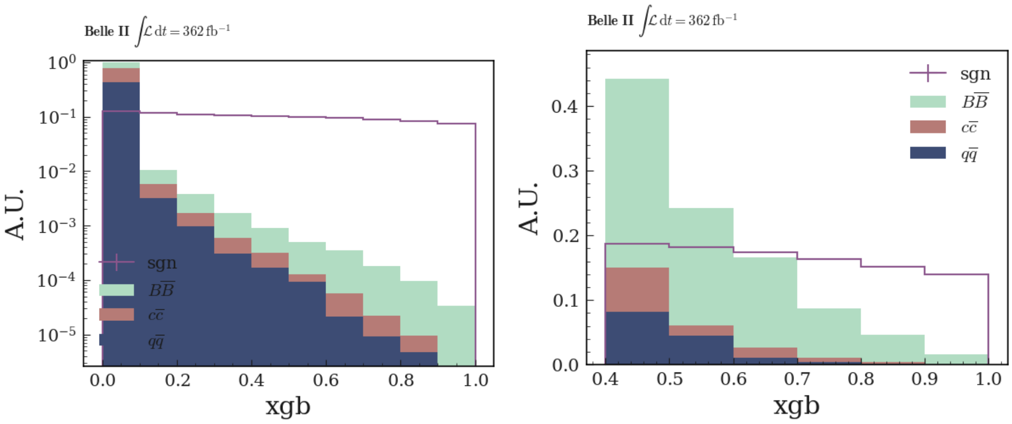


Figure 7.2: Distributions de la valeur de sortie du BDT pour différents échantillons simulés. La figure de gauche montre l'ensemble de l'interval de définition de la variable tandis que la figure de droite correspond à la région de signal. Les distributions présentées correspondent au signal (violet) et aux évènements $e^+e^- \to B\bar{B}$ (vert), $e^+e^- \to c\bar{c}$ (rouge) et $e^+e^- \to q\bar{q}$ (bleu).

donc étudiés: des évènements de la forme $e^+e^- \to \Upsilon(4S) \to B_{tag}B'_{sig}$, avec $B'_{sig} \to K^+J/\Psi(\mu^+\mu^-)$ sont séléctionnés, la contribution du B'_{sig} est ensuite remplacée par la contribution de $B_{sig} \to K^+\nu\bar{\nu}$ extraite d'un évènement de signal simulé. Cette procédure est appellée *embedding*.
La procédure d'*embedding* est également appliquée à des évènements simulés comme vérification supplémentaire. Figure 7.3 montre la distribution de la valeur de sortie du BDT pour les différents échantillons considérés. Malgrès des limitations liées à la taille de l'échantillon de données, la simulation du signal et l'efficacité de sa sélection semblent bien reproduire ce qui est observé dans les données.

- La qualité de la simulation pour les évènements de type $e^+e^- \to q\bar{q}$, avec q un quark u, d, s ou c est étudiée grâce à des données collectées avec une énergie de collision 60 MeV en dessous de l'énergie nécessaire à la production de la resonance $\Upsilon(4S)$ (données non-resonnantes). Cet échantillon de données à l'avantage d'être totalement dépourvu de contributions de signal. Les distributions des différentes variablers utilisées dans la selection sont comparées entre les données et les échantillons simulés. Un accord correct est observé. Une correction de la simulation est cependant développée afin d'améliorer la description des données. Pour ce faire, une pondération est calculée pour chaque évènement simulé à partir de la valeur de sortie d'un BDT (nommé BDT$_c$), entrainé sur les données non-résonnantes et sur les évènements $e^+e^- \to q\bar{q}$ simulés. Figure 7.4 montre l'impact de cette correction sur l'accord entre données et simulation.

- Enfin, la qualité de la simulation pour les évènements de bruits de fond de type $e^+e^- \to q\bar{q}$ et $e^+e^- \to B\bar{B}$ est étudiée dans les données collectées à l'énergie de la résonnance $\Upsilon(4S)$. Afin d'éviter une observation des évènements de la region de signal, des échantillons sont définis en inversant certains critères de la selection nominale de l'analyse (les autres critères sont gardés tels quels). Les deux échantillons ainsi construits sont: un échantillon pour lequel les B_{tag} et B_{sig} reconstruits sont requis de possèder la même charge électrique (échantillon WC) et un échantillon pour lequel la particule identifiée comme le kaon provenant du B_{sig} a une forte probabilité d'être un pion (échantillon KID). Figure 7.5 montre un désaccord entre données et simulation pour ces échantillons (tempéré par de conséquentes incertitudes statistiques liées à la taille des échantillons de données). Une incertitude systématique est alors éstimée à partir de cette étude afin de couvrir de potentiels problèmes de simulation.

Par la suite, plusieurs sources d'incertitudes systématiques sont identifiées et leur impact sur la mesure est éstimé. Ces incertitudes sont d'origines diverses: inéfficacités du détecteur, incertitudes théoriques liées aux prédictions du MS ou simulation des canaux de bruits de fond.
Toutes les étapes nécessaires à la mesure de $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ sont alors développées. Cependant, comme décrit ci-avant, cette analyse doit être validée par l'ensemble de la collaboration Belle II avant d'autoriser l'étude de la région de signal, afin d'éviter tout biais. Cette analyse est à ce jour en attente de cette validation. Il est cependant possbile d'estimer sa sensibilité en se basant sur l'étude d'échantillons simulés. En supposant une valeur de $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ égale à la valeur attendue dans le SM, il est possible d'extraire la limite supérieure:

$$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 2.3 \times 10^{-5} \tag{7.1}$$

pour un niveau de confiance de 90%.
Ce résultat attendu est compétitif avec les tentatives de mesures de $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$ menées par le passé par les collaboration Belle et BaBar [61–64], pourtant basées sur des échantillons jusqu'à deux fois plus conséquents que celui étudié ici. Ceci est expliqué par les performances du detecteur Belle II et de la sélection développée ici, comparées aux performances d'expériences plus anciennes.

Amélioration de la résolution spatiale du détecteur de vertex de l'expérience Belle II

Le detecteur de vertex à pistes de silicium (SVD) de l'expérience Belle II est un élément crucial du détecteur, contribuant à la trajectographie, à l'identification de particules et permettant la reconstruction des vertex de désintgration des particules produites au sein de l'expérience. Le principe de mesure du SVD se base sur l'ionisation de pistes de silicium réparties sur quatre couches concentriques induite par le passage de particules chargées. Sur une couche et pour une particule donnée,

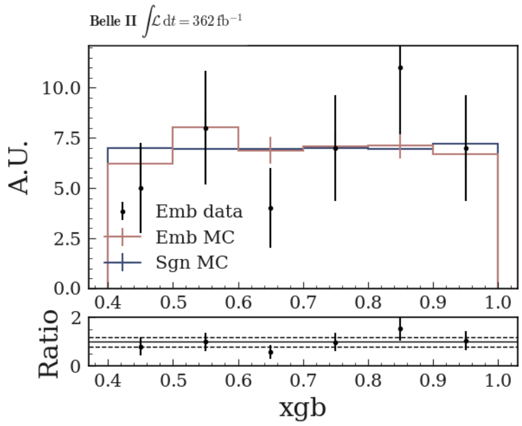


Figure 7.3: Distributions de la valeur de sortie du BDT dans la région de signal pour l'échantillon de signal simulé (bleu) ainsi que pour les échantillons ayant subi la procédure d'*embedding* (simulation en rouge et données en points).

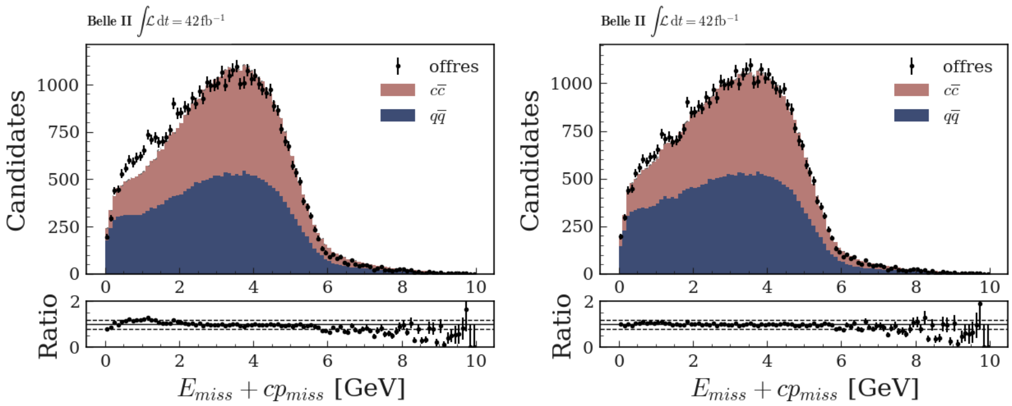


Figure 7.4: Distributions de la somme de l'énergie manquante pour les échantillons $e^+e^- \to q\bar{q}$ simulés ainsi que pour les données non-résonnantes (points), avant (gauche) et après correction (droite).

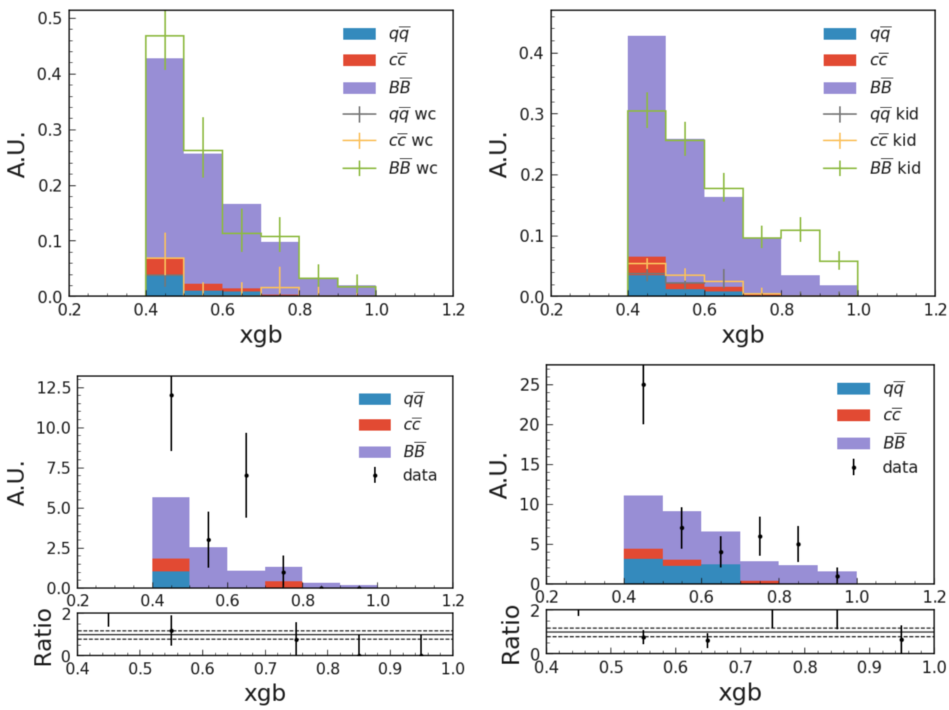


Figure 7.5: (haut) valeur de sortie du BDT dans la région de signal pour les échantillons simulés après application de la séléction nominale (histogrammes pleins) et après les séléctions modifiées pour éviter une contamination pour le signal (histogrammes en lignes). (bas) Comparaison entre données et simulations dans la région de signal après application de ces mêmes séléctions modifiées. Les figures de gauche correspondent à l'échantillon *WC* tandis que les figures de droites correspondent à l'échantillon *KID*.

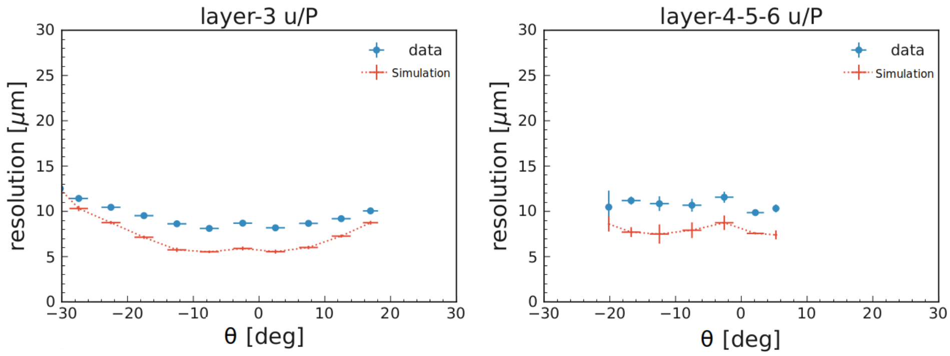


Figure 7.6: Comparison entre données et simulation de l'estimation de la résolution spatiale en fonction de l'angle de la trajectoire de la particule incidente θ.

plusieurs pistes subissent cette ionisation autour du point de passage de la particule. Ces pistes sont alors regroupées en un amas servant à estimer la position de l;intersection entre la trajectoire de la particule et la couche. Plusieurs caractéristiques de l'amas peuvent être calculées (temps écoulé entre l'évènement de collision et la traversée de la couche par la particule, charge totale collectée par les pistes, position de l'amas).

La position x_A de l'amas est calculée comme:

$$x_A = \frac{\sum_{i=0} x_i \times S_i}{\sum_{i=0} S_i} \tag{7.2}$$

avec x_i la position de la piste i et S_i la charge collectée par la même piste.

La position des différents amas est utilisée par les algorithmes de trajectographie, c'est pourquoi la mesure de cette position se doit d'être la plus précise possible. Cette précision est estimée par la résolution spatiale du détecteur, prenant en compte l'écart entre la position mesurée d'un amas et la position attendue de l'intersection entre la trajectoire de la particule et la couche portant l'amas (éstimée grâce à la position des amas reconstruits sur les autres couches du détecteur) ainsi que l'erreur associée à la mesure de cet écart.

La résolution spatiale du détecteur est estimée en utilisant les données collectées ainsi que des échantillons simulés (Figure 7.6). On observe alors un désaccord entre données et simulation, la résolution éstimée à l'aide d'échantillons simulés étant systématiquement plus basse (simulation optimiste). J'ai alors mené un travail de recherche visant à déterminer de possibles causes expliquant ce désaccord.

Des mesures effectuées sur le détecteur semblent indiquer un effet électronique menant à un biais dans la mesure du signal collecté par les différentes pistes. À cause de cet effet, la mesure de la charge collectée par une piste se voit biaisée à hauteur de 6% de la charge collectée par les pistes voisines. Figure 7.7 illustre cet effet qui semble systématique et affecte l'ensemble des pistes du détecteur.

Afin de corriger ce biais j'ai développé une méthode algorithmique visant à découpler cet échange de charge apparent. Pour ce faire, une matrice M de taille $n \times n$

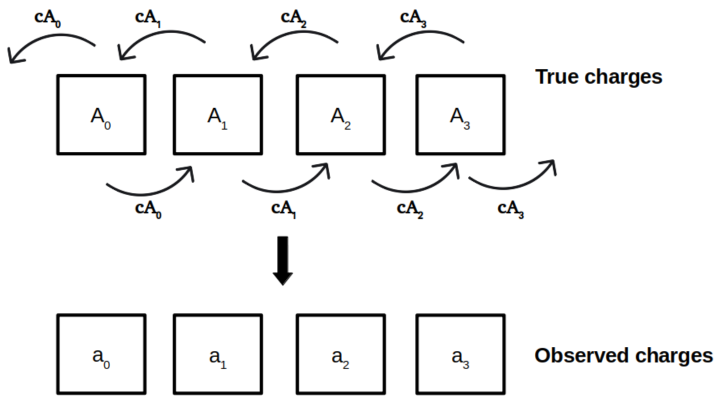


Figure 7.7: Représentation schématique de l'effet éléctronique biaisant la mesure de la charge colléctée par les pistes du détecteur. La relation entre la charge réelle A_i collectée par la piste i et a_i la valeur observée biaisée par l'effet est illustrée.

(n correspondant au nombre de pistes formant l'amas considéré) est définie comme:

$$\begin{cases} M_{ij} & = & 1-2c & \text{if } i=j; \\ M_{ij} & = & c & \text{if } \mid i-j \mid = 1; \\ M_{ij} & = & 0 & \text{for all others } (i,j); \end{cases} \tag{7.3}$$

où $c = 0.06$ correspond à la fraction de charge collectée apparemment échangée entre deux pistes adjacentes. Il devient alors possible d'estimer la valeur A_i réélement collectée par la piste i à partir des charges collectées observée a_i, a_{i-1} et a_{i+1}:

$$\begin{pmatrix} A_0 \\ A_1 \\ \ldots \\ A_{n-1} \end{pmatrix} = M^{-1} \begin{pmatrix} a_0 \\ a_1 \\ \ldots \\ a_{n-1} \end{pmatrix}. \tag{7.4}$$

En appliquant cette correction, on observe une réduction du désaccord entre données et simulation dans l'estimation de la résolution spatiale FIGURE****. Cette correction est implémentée dans le système d'analyse central de la collaboration Belle II.
On s'attend à ce que le désaccord restant entre données et simulation soit dû à une combinaison de plusieurs effets de faibles amplitude, rendant leur identification complexe.

Conclusion

La mesure du rapport d'embranchement de la désintégration $B^+ \to K^+\nu\bar{\nu}$ est au centre du programme de physique de l'expérience Belle II. En effet, Belle II est la seule expérience de sa génération à pouvoir mesurer cette observable, de plus, ce résultat est attendu par l'ensemble de la communauté de la physique des saveurs, car de nombreux modèles d'exetension du Modèle Standard de la physique des particules prédisent des modifications de $\mathcal{B}(B^+ \to K^+\nu\bar{\nu})$. Au cours de ma thèse j'ai développé une analyse complète visant à permettre la première observation de cette désintégration, et j'ai eu l'occasion de valider cette méthode sur des évènements simulés.

De plus, j'ai développé et implémenté dans le système d'analyse central de la collaboration une méthode permettant de corriger l'estimation de la resolution spatiale du détecteur de vertex de Belle II.

Appendices

Unfolding method

A.1 Hadronic events study

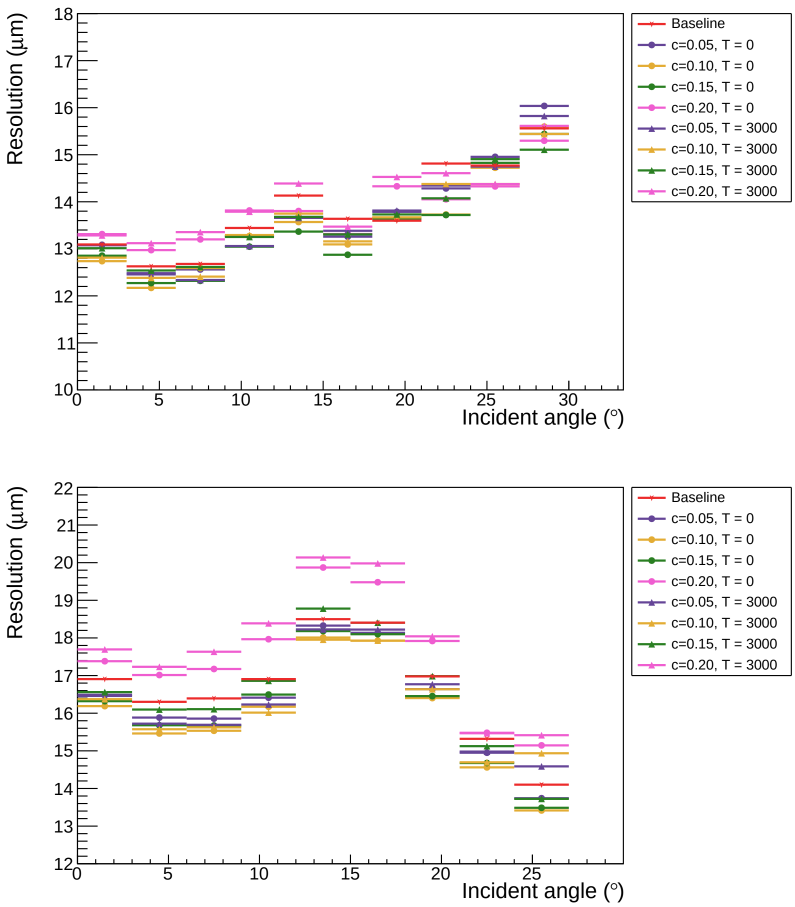


Figure A.1: Cluster position resolution as a function of the incident angle of the track for all (c,T) couples. Each color corresponds to a c value, circle markers correspond to $T = 0$ ADC and triangle markers correspond to $T = 3000$ ADC. The red points correspond to the baseline (*i.e* no correction applied). For the Layer 3 u/P-side (top) and Layer 4,5 and 6 u/P-side (bottom).

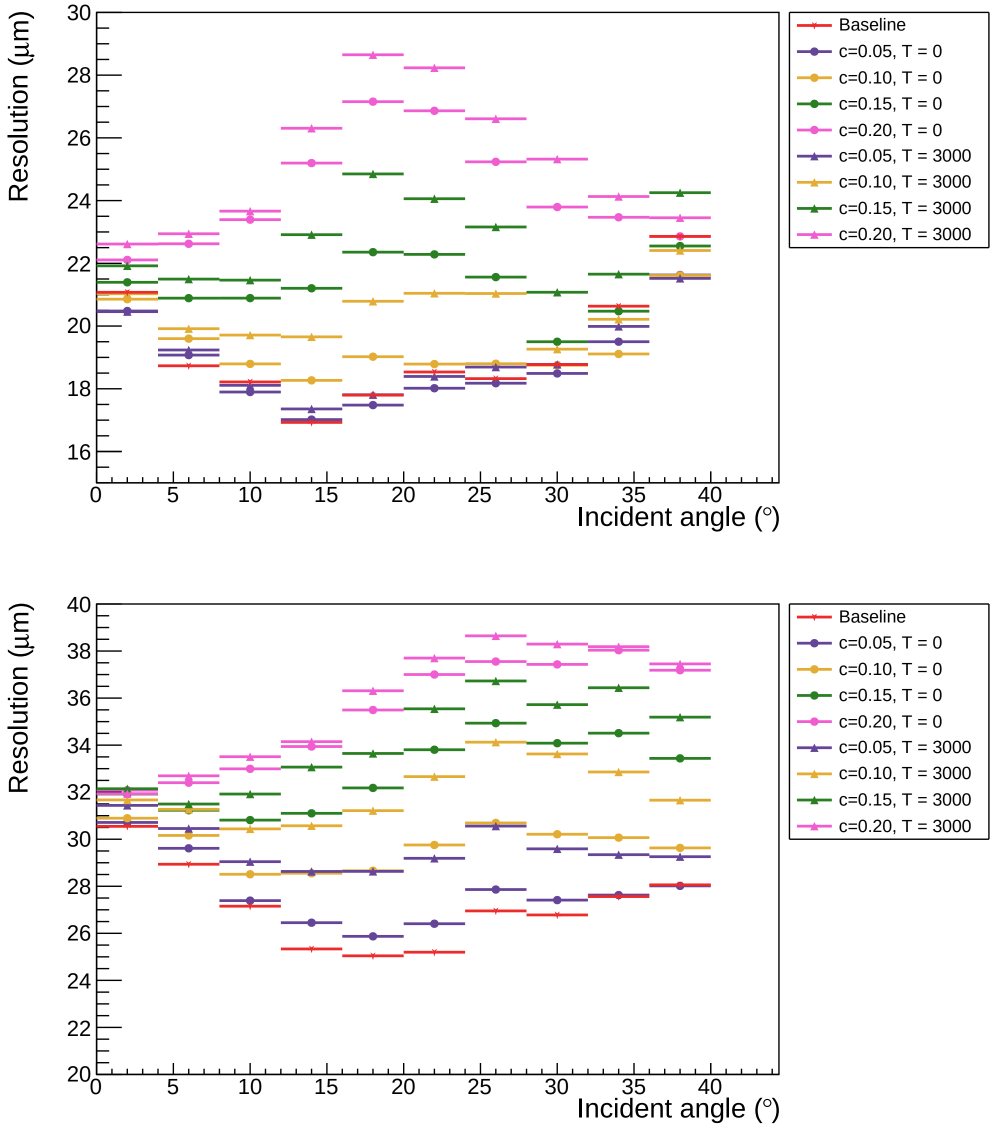


Figure A.2: Cluster position resolution as a function of the incident angle of the track for all (c,T) couples. Each color corresponds to a c value, circle markers correspond to $T = 0$ ADC and triangle markers correspond to $T = 3000$ ADC. The red points correspond to the baseline (*i.e* no correction applied). For the Layer 3 v/N-side (top) and Layer 4,5 and 6 v/N-side (bottom).

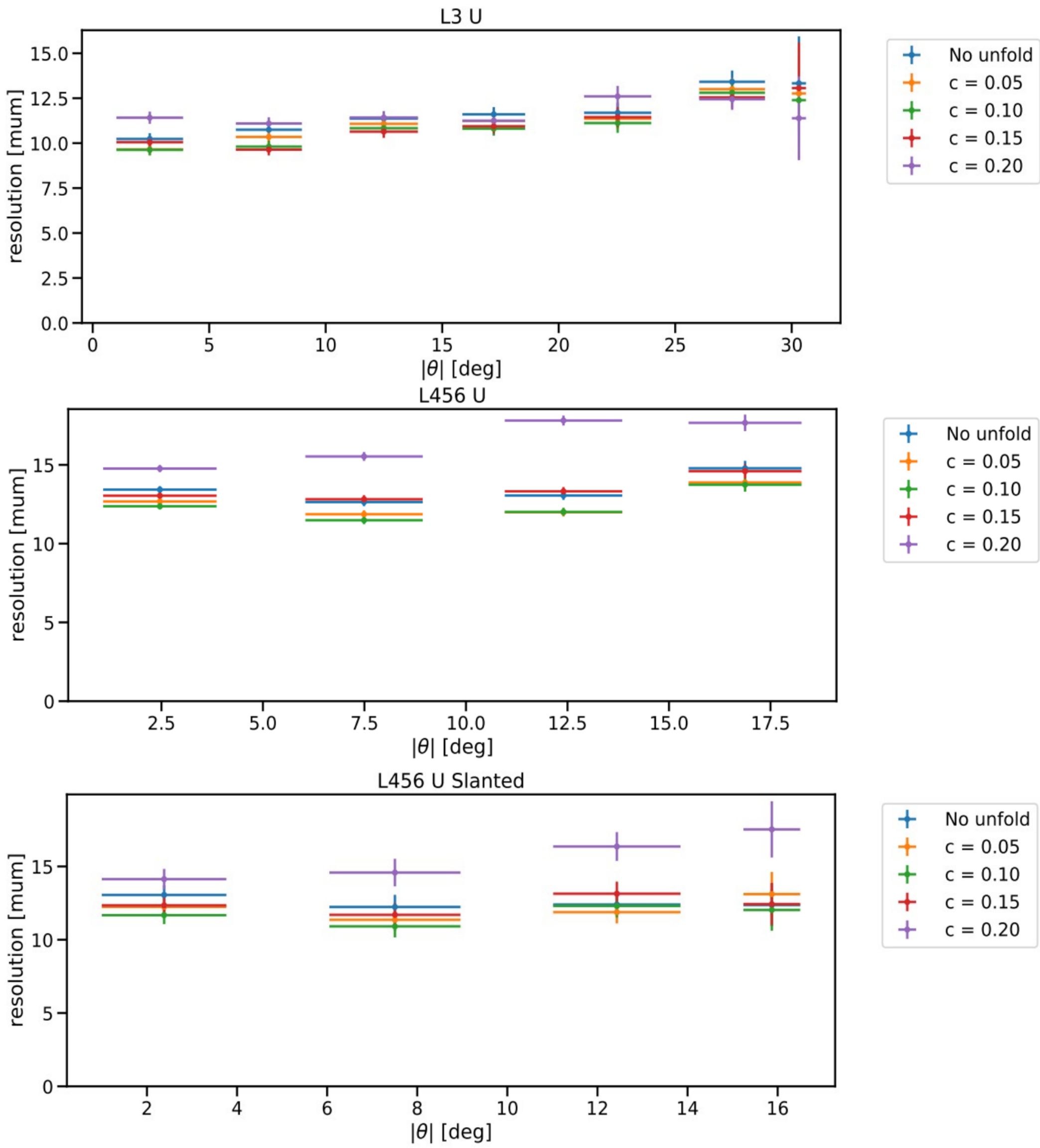


Figure A.3: Cluster position resolution as a function of the track incident angle θ computed for several values of the unfolding parameter c compared to the nominal resolution (blue). For the u/P side layer 3 sensors (top), layer 4, 5 and 6 barrel sensors (middle) and slanted sensors (bottom).

A.2 Track incident angle

We show here a comparison between the cluster position resolution computed using different values for c for each sensor type as a function of the track incident angle.

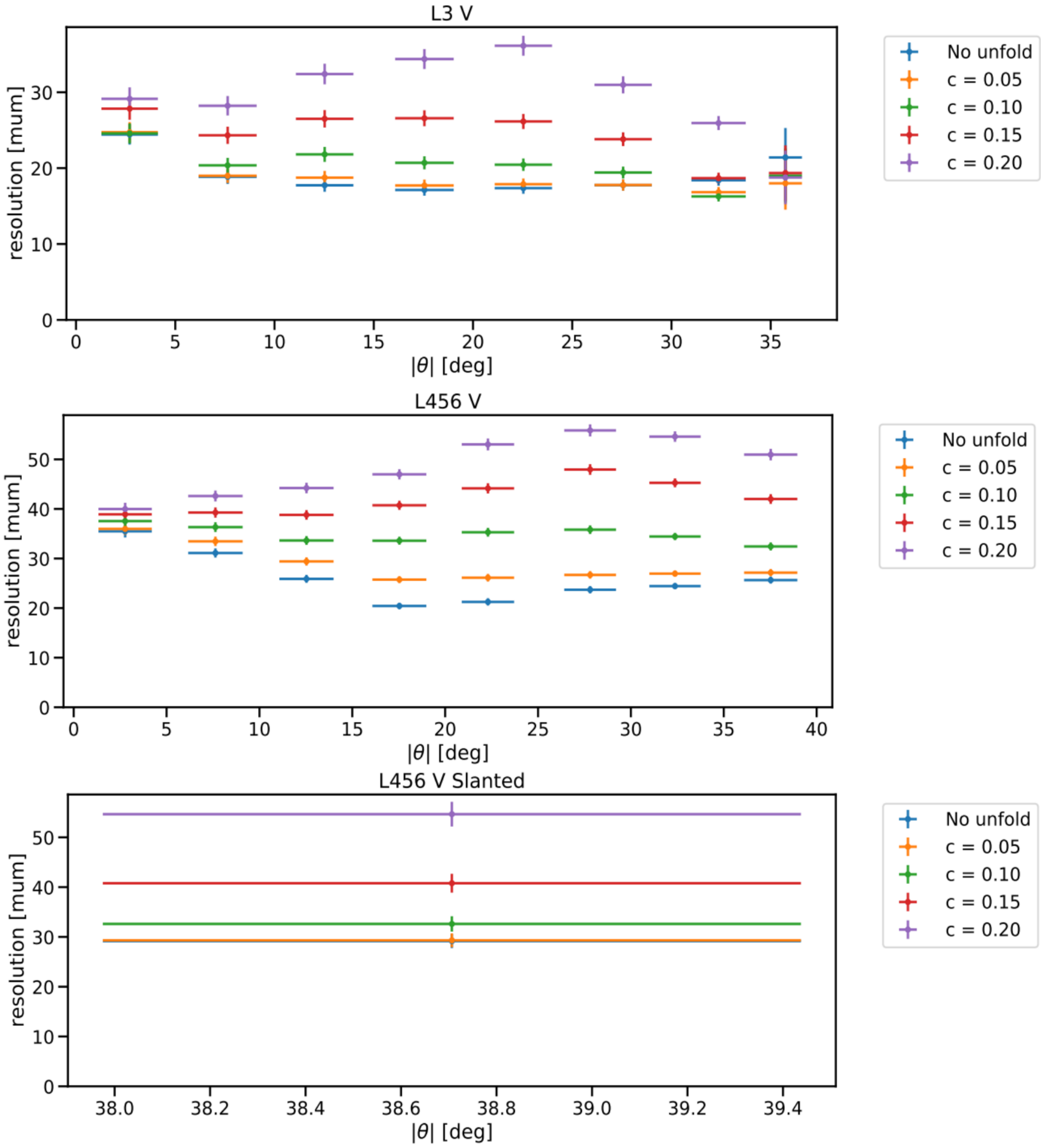


Figure A.4: Cluster position resolution as a function of the track incident angle θ computed for several values of the unfolding parameter c compared to the nominal resolution (blue). For the v/N side layer 3 sensors (top), layer 4, 5 and 6 barrel sensors (middle) and slanted sensors (bottom).

Appendix B

Variable validation using off-resonance data

We show here the distributions for all the variables listed in Subsection 5.4.1 for the continuum simulated samples and the off-resonance data sample. All the distributions are shown after the BDT_c reweighting.

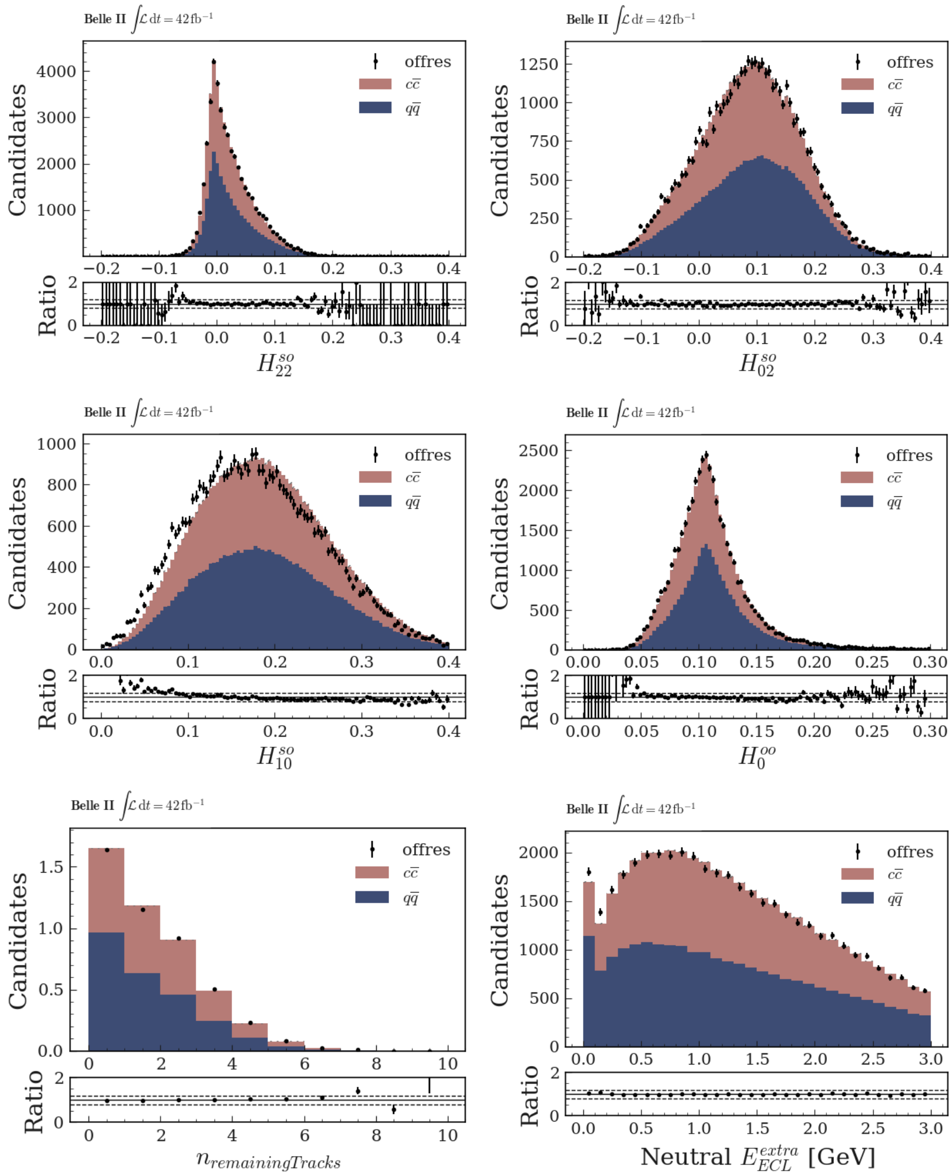


Figure B.1: Distribution of the discriminative features used in the training of the BDT for the light-$q\bar{q}$ (blue) and $c\bar{c}$ (red) simulated sample and off-resonance data (dots). The definition of each variable can be found in Subsection 5.4.1

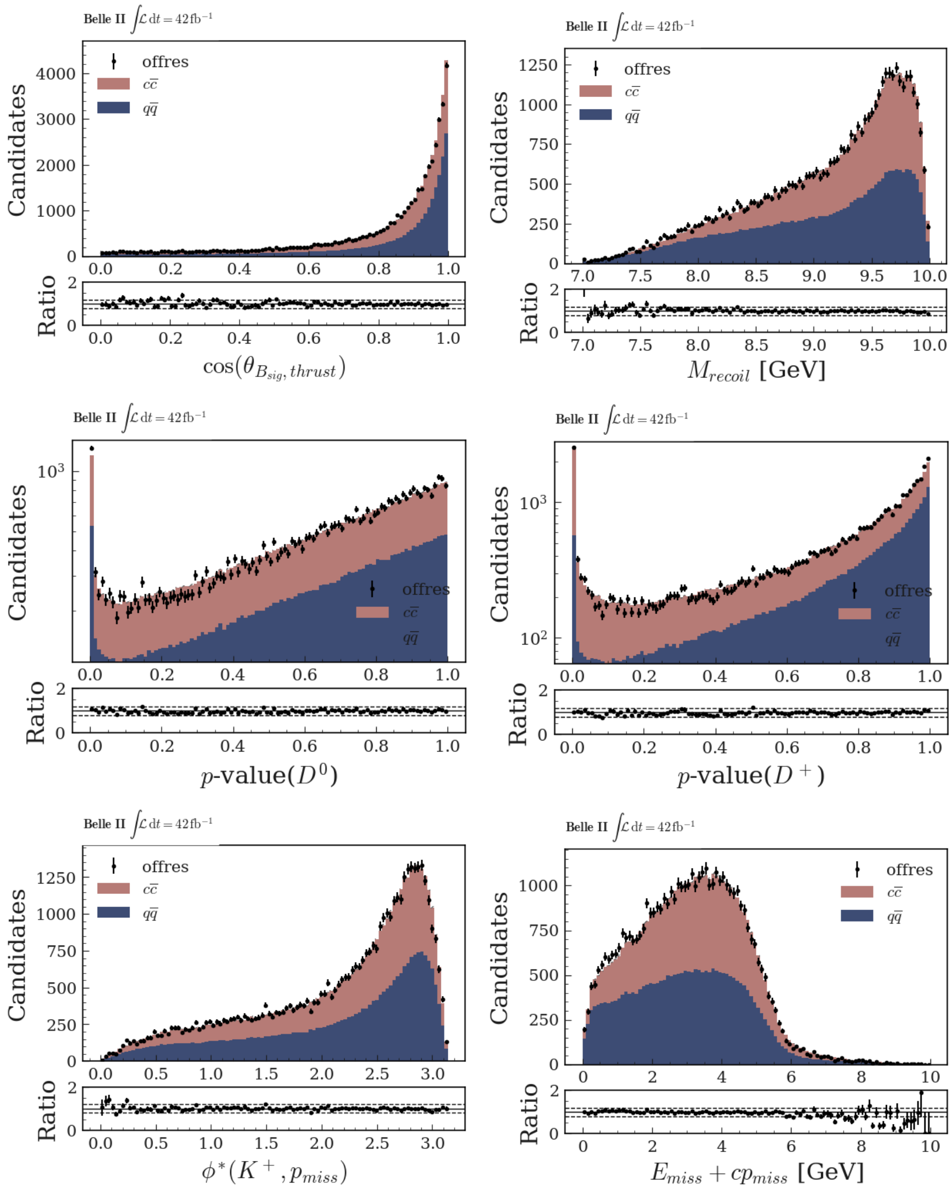


Figure B.2: Distribution of the discriminative features used in the training of the BDT for the light-$q\bar{q}$ (blue) and $c\bar{c}$ (red) simulated sample and off-resonance data (dots). The definition of each variable can be found in Subsection 5.4.1

Appendix C

Variable validation using embedded data

We show here the distributions for all the variables listed in Subsection 5.4.1 for the signal and embedded $B^+ \to J/Psi(\mu^+\mu^-)K^+$ simulated sample and the embedded $B^+ \to J/Psi(\mu^+\mu^-)K^+$ data sample.

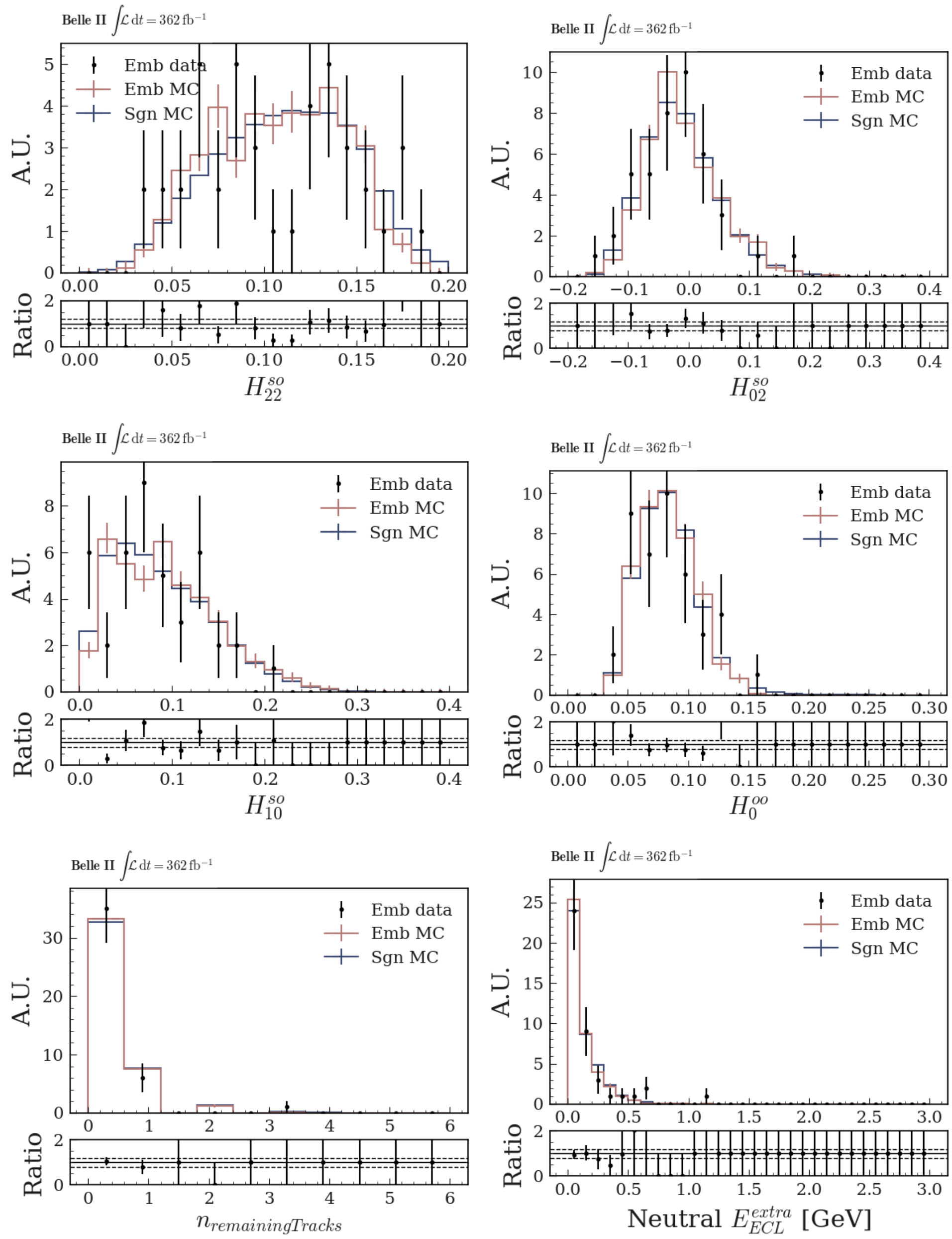


Figure C.1: Distribution of the discriminative features used in the training of the BDT for the signal (blue) and embedded (red) simulated sample and embedded data (dots). The definition of each variable can be found in Subsection 5.4.1

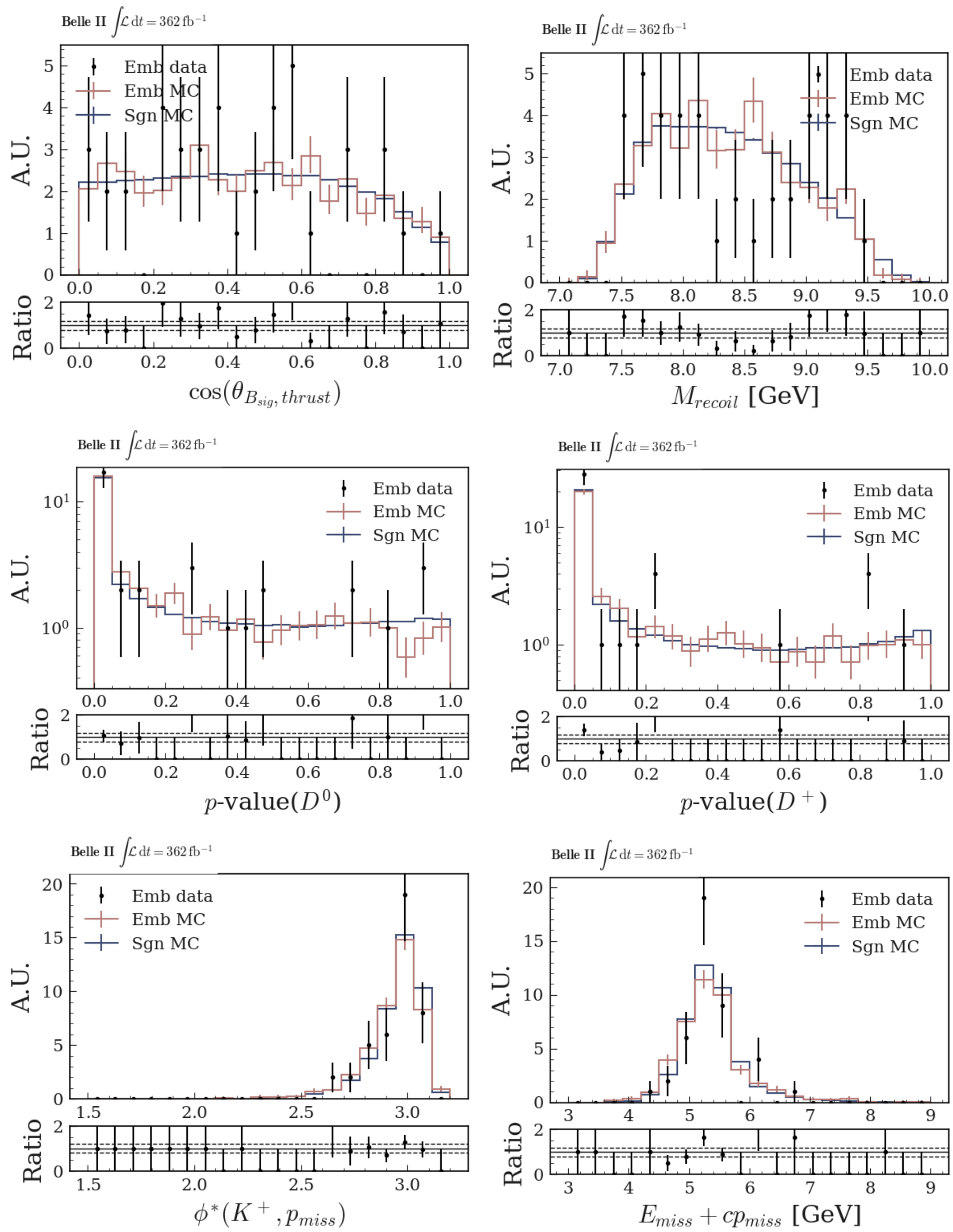


Figure C.2: Distribution of the discriminative features used in the training of the BDT for the signal (blue) and embedded (red) simulated sample and embedded data (dots). The definition of each variable can be found in Subsection 5.4.1

Appendix D

Background composition in the signal region

Several particle classes are defined to categorize B-meson decays, in order to better study the $B\bar{B}$ contribution to the analysis. Here, we show how these classes are built.

Class	Particles
D	D^+, D^0, $D_0^*(2300)^+$, $D_0^*(2300)^0$, $D^*(2010)^+$, $D^*(2007)^0$, $D_1(2420)^+$, $D_1(2420)^0$, 20413, $D_1(2430)^0$, $D_2^*(2460)^+$, $D_2^*(2460)^0$, D_s^+, $D_{s0}^*(2317)^+$, D_s^{*+}, $D_{s1}(2536)^+$, $D_{s1}(2460)^+$, $D_{s2}^*(2573)^+$
ℓ	e^-, μ^-
τ	τ^-
ν	ν_e, ν_μ, ν_τ
$n\pi$	π^0, π^+, $a_0(980)^0$, $a_0(980)^+$, $\pi(1300)^0$, $\pi(1300)^+$, $a_0(1450)^0$, $a_0(1450)^+$, $\pi(1800)^0$, $\pi(1800)^+$, $\rho(770)^0$, $\rho(770)^+$, $b_1(1235)^0$, $b_1(1235)^+$, $a_1(1260)^0$, $a_1(1260)^+$, $\pi_1(1400)^0$, $\pi_1(1400)^+$, $\rho(1450)^0$, $\rho(1450)^+$, $\pi_1(1600)^0$, $\pi_1(1600)^+$, $a_1(1640)^0$, $a_1(1640)^+$, $\rho(1700)^0$, $\rho(1700)^+$, $a_2(1320)^0$, $a_2(1320)^+$, $\pi_2(1670)^0$, $\pi_2(1670)^+$, $a_2(1700)^0$, $a_2(1700)^+$, $\rho_3(1690)^0$, $\rho_3(1690)^+$, $a_4(1970)^0$, $a_4(1970)^+$, η, $\eta'(958)$, $f_0(500)$, $f_0(980)$, $\eta(1295)$, $f_0(1370)$, $\eta(1405)$, $\eta(1475)$, $f_0(1500)$, $f_0(1710)$, $\omega(782)$, $\phi(1020)$, $h_1(1170)$, $f_1(1285)$, $h_1(1415)$, $f_1(1420)$, $\omega(1650)$, $\phi(1680)$, $f_2(1270)$, $f_2'(1525)$, $\eta_2(1645)$, $f_2(1950)$, $f_2(2010)$, $f_2(2300)$, $f_2(2340)$, $\omega_3(1670)$, $\phi_3(1850)$, $f_4(2050)$
$c\bar{c}$	$\eta_c(1S)$, $\chi_{c0}(1P)$, $\eta_c(2S)$, $J/\psi(1S)$, $h_c(1P)$, $\chi_{c1}(1P)$, $\psi(2S)$, $\psi(3770)$, $\psi(4040)$, $\psi(4160)$, $\psi(4415)$, $\chi_{c2}(1P)$, $\chi_{c2}(2P)$
Hadrons	K_L^0, K_S^0, K^0, K^+, $K_0^*(700)^0$, $K_0^*(700)^+$, $K_0^*(1430)^0$, $K_0^*(1430)^+$, $K(1460)^+$, $K(1830)^+$, $K_0^*(1950)^+$, $K^*(892)^0$, $K^*(892)^+$, $K_1(1270)^0$, $K_1(1270)^+$, $K_1(1400)^0$, $K_1(1400)^+$, $K^*(1410)^0$, $K^*(1410)^+$, $K_1(1650)^0$, $K_1(1650)^+$, $K^*(1680)^0$, $K^*(1680)^+$, $K_2^*(1430)^0$, $K_2^*(1430)^+$, $K_2(1580)^+$, $K_2(1770)^0$, $K_2(1770)^+$, $K_2(1820)^0$, $K_2(1820)^+$, $K_2^*(1980)^0$, $K_2^*(1980)^+$, $K_2(2250)^+$, $K_3^*(1780)^0$, $K_3^*(1780)^+$, $K_3(2320)^+$, $K_4^*(2045)^0$, $K_4^*(2045)^+$, $K_4(2500)^+$, π^0, π^+, $a_0(980)^0$, $a_0(980)^+$, $\pi(1300)^0$, $\pi(1300)^+$, $a_0(1450)^0$, $a_0(1450)^+$, $\pi(1800)^0$, $\pi(1800)^+$, $\rho(770)^0$, $\rho(770)^+$, $b_1(1235)^0$, $b_1(1235)^+$, $a_1(1260)^0$, $a_1(1260)^+$, $\pi_1(1400)^0$, $\pi_1(1400)^+$, $\rho(1450)^0$, $\rho(1450)^+$, $\pi_1(1600)^0$, $\pi_1(1600)^+$, $a_1(1640)^0$, $a_1(1640)^+$, $\rho(1700)^0$, $\rho(1700)^+$, $a_2(1320)^0$, $a_2(1320)^+$, $\pi_2(1670)^0$, $\pi_2(1670)^+$, $a_2(1700)^0$, $a_2(1700)^+$, $\rho_3(1690)^0$, $\rho_3(1690)^+$, $a_4(1970)^0$, $a_4(1970)^+$, η, $\eta'(958)$, $f_0(500)$, $f_0(980)$, $\eta(1295)$, $f_0(1370)$, $\eta(1405)$, $\eta(1475)$, $f_0(1500)$, $f_0(1710)$, $\omega(782)$, $\phi(1020)$, $h_1(1170)$, $f_1(1285)$, $h_1(1415)$, $f_1(1420)$, $\omega(1650)$, $\phi(1680)$, $f_2(1270)$, $f_2'(1525)$, $\eta_2(1645)$, $f_2(1950)$, $f_2(2010)$, $f_2(2300)$, $f_2(2340)$, $\omega_3(1670)$, $\phi_3(1850)$, $f_4(2050)$, p, n, $\Delta(1232)^{++}$, $\Delta(1232)^+$, $\Delta(1232)^0$, $\Delta(1232)^-$, Λ, Σ^+, Σ^0, Σ^-, $\Sigma(1385)^0$, $\Sigma(1385)^-$, Ξ^0, Ξ^-, $\Xi(1530)^0$, $\Xi(1530)^-$, Ω^-

Table D.1: Particle classes used to categorize B-meson decays in simulated $B\bar{B}$ samples.

Bibliography

[1] C. N. Yang and R. L. Mills, “Conservation of isotopic spin and isotopic gauge invariance,” *Phys. Rev.*, vol. 96, pp. 191–195, Oct 1954. (Cited on page vii.)

[2] M. Gell-Mann, “Symmetries of baryons and mesons,” *Phys. Rev.*, vol. 125, pp. 1067–1084, Feb 1962. (Cited on page vii.)

[3] S. Weinberg, “A model of leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967. (Cited on page vii.)

[4] A. Salam, “Weak and electromagnetic interactions.” (Cited on page vii.)

[5] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961. (Cited on page vii.)

[6] G. ’t Hooft and M. Veltman, “Regularization and renormalization of gauge fields,” *Nuclear Physics B*, vol. 44, no. 1, pp. 189–213, 1972. (Cited on page vii.)

[7] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964. (Cited on page vii.)

[8] T. D. Lee and C. N. Yang, “Question of parity conservation in weak interactions,” *Phys. Rev.*, vol. 104, pp. 254–258, Oct 1956. (Cited on page vii.)

[9] R. Brown *et al.*, “Observations with electron-sensitive plates exposed to cosmic radiation,” *Nature*, vol. 163, pp. 47 – 51, Jan 1949. (Cited on page vii.)

[10] V. E. Barnes *et al.*, “Observation of a hyperon with strangeness minus three,” *Phys. Rev. Lett.*, vol. 12, pp. 204–206, Feb 1964. (Cited on pages vii and 1.)

[11] F. Hasert *et al.*, “Search for elastic muon-neutrino electron scattering,” *Physics Letters B*, vol. 46, no. 1, pp. 121–124, 1973. (Cited on pages vii and 1.)

[12] F. Abe *et al.*, “Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab,” *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, Apr 1995. (Cited on pages vii and 1.)

[13] S. Abachi *et al.*, “Observation of the top quark,” *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, Apr 1995. (Cited on pages vii and 1.)

[14] G. Aad *et al.*, “Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc,” *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012. (Cited on pages vii, 1 and 6.)

[15] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 gev with the CMS experiment at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012. (Cited on pages vii, 1 and 6.)

[16] V. C. Rubin and J. Ford, W. Kent, "Rotation of the andromeda nebula from a spectroscopic survey of emission regions," *apj*, vol. 159, p. 379, feb 1970. (Cited on page vii.)

[17] F. Zwicky, "On the masses of nebulae and of clusters of nebulae," *apj*, vol. 86, p. 217, Oct. 1937. (Cited on page vii.)

[18] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter," *The Astrophysical Journal*, vol. 648, p. L109, aug 2006. (Cited on page vii.)

[19] P. Collaboration, "Planck 2018 results - vi. cosmological parameters," *A&A*, vol. 641, p. A6, 2020. (Cited on page vii.)

[20] B. Abi *et al.*, "Measurement of the positive muon anomalous magnetic moment to 0.46 ppm," *Phys. Rev. Lett.*, vol. 126, p. 141801, Apr 2021. (Cited on page vii.)

[21] S. Mertens, "Direct neutrino mass experiments," *Journal of Physics: Conference Series*, vol. 718, p. 022013, may 2016. (Cited on page vii.)

[22] R. Aaij *et al.*, "Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay," *Phys. Rev. Lett.*, vol. 126, p. 161802, Apr 2021. (Cited on pages vii and 1.)

[23] L. collaboration, "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays," 2022. (Cited on page vii.)

[24] T. Abe *et al.*, "Belle ii technical design report," 2010. (Cited on pages viii, 22, 26, 27, 29, 30 and 34.)

[25] K. Akai, K. Furukawa, and H. Koiso, "SuperKEKB collider," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 188–199, nov 2018. (Cited on pages viii, 20 and 115.)

[26] S. Hirose *et al.*, "Measurement of the τ lepton polarization and $r(D^*)$ in the decay $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$," *Phys. Rev. Lett.*, vol. 118, p. 211801, May 2017. (Cited on page 1.)

[27] M. Huschle *et al.*, "Measurement of the branching ratio of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ relative to $\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell$ decays with hadronic tagging at belle," *Phys. Rev. D*, vol. 92, p. 072014, Oct 2015. (Cited on page 1.)

[28] J. P. Lees *et al.*, "Evidence for an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays," *Phys. Rev. Lett.*, vol. 109, p. 101802, Sep 2012. (Cited on page 1.)

[29] R. Aaij *et al.*, "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.*, vol. 115, p. 111803, Sep 2015. (Cited on page 1.)

[30] A. J. Buras, "Weak hamiltonian, cp violation and rare decays," 1998. (Cited on page 7.)

[31] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, "Weak decays beyond leading logarithms," *Reviews of Modern Physics*, vol. 68, pp. 1125–1244, oct 1996. (Cited on page 7.)

[32] J. Brod, M. Gorbahn, and E. Stamou, "Two-loop electroweak corrections for the $k \to \pi\nu\bar{\nu}$ decays," *Physical Review D*, vol. 83, feb 2011. (Cited on page 7.)

[33] W. G. Parrott, C. Bouchard, and C. T. H. Davies, "Standard model predictions for $B \to K\ell^+\ell^-$, $B \to K\ell_1^-\ell_2^+$ and $B \to K\nu\bar{\nu}$ using form factors from $n_f = 2+1+1$ lattice qcd," 2022. (Cited on page 8.)

[34] 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," 2014. (Cited on pages 8, 10, 11 and 97.)

[35] W. Parrott, C. Bouchard, and C. D. and, "Standard model predictions for $B \to K\ell^+\ell^-$, $B \to K\ell_1^-\ell_2^+$ and $B \to K\nu\bar{\nu}$ using form factors from $n_f = 2+1+1$ lattice qcd.," *Physical Review D*, vol. 107, jan 2023. (Cited on pages 8, 9 and 113.)

[36] D. Bečirević, G. Piazza, and O. Sumensari, "Revisiting $B \to K^{(*)}\nu\bar{\nu}$ decays in the standard model and beyond," *The European Physical Journal C*, vol. 83, mar 2023. (Cited on page 8.)

[37] J. F. Kamenik and C. Smith, "Tree-level contributions to the rare decays $B^+ \to \pi^+\nu\bar{\nu}$, $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$ in the standard model," *Physics Letters B*, vol. 680, pp. 471–475, oct 2009. (Cited on page 8.)

[38] C. Bourrely, L. Lellouch, and I. Caprini, "Erratum: Model-independent description of $B \to \pi l\nu$ decays and a determination of $|V_{ub}|$ [phys. rev. dprvdaq1550-7998 79, 013008 (2009)]," *Phys. Rev. D*, vol. 82, p. 099902, Nov 2010. (Cited on page 9.)

[39] D. M. Straub, "flavio: a python package for flavour and precision phenomenology in the standard model and beyond," 2018. (Cited on page 11.)

[40] A. Crivellin, C. A. Manzari, W. Altmannshofer, G. Inguglia, P. Feichtinger, and J. M. Camalich, "Toward excluding a light z' explanation of $b \to sl^+l^-$," *Physical Review D*, vol. 106, aug 2022. (Cited on pages 12 and 113.)

[41] X. G. He and G. Valencia, "$R^{\nu}_{K^{(*)}}$ and non-standard neutrino interactions," *Physics Letters B*, vol. 821, p. 136607, oct 2021. (Cited on pages 12, 13 and 113.)

[42] The ALEPH Collaboration, "Precision electroweak measurements on the z resonance," *Physics Reports*, vol. 427, no. 5, pp. 257–454, 2006. (Cited on page 12.)

[43] D. B. et. al., "A direct measurement of the invisible width of the z from single photon counting," *Physics Letters B*, vol. 313, no. 3, pp. 520–534, 1993. (Cited on page 12.)

[44] A. P. M. Bordone, G. sidori, "On the standard model predictions for R_K and R_{K^*}," *The European Physical Journal C*, vol. 76, no. 8, p. 440, 2016. (Cited on pages 12 and 113.)

[45] D. Buttazzo, A. Greljo, G. Isidori, and D. Marzocca, "B-physics anomalies: a guide to combined explanations," *Journal of High Energy Physics*, vol. 2017, nov 2017. (Cited on pages 12 and 113.)

[46] B. Bhattacharya, A. Datta, D. London, and S. Shivashankara, "Simultaneous explanation of the rk and r(d(*)) puzzles," *Physics Letters B*, vol. 742, pp. 370–374, 2015. (Cited on pages 12 and 113.)

[47] F. Feruglio, P. Paradisi, and A. Pattori, "Revisiting lepton flavor universality in B decays," *Phys. Rev. Lett.*, vol. 118, p. 011801, Jan 2017. (Cited on pages 12 and 113.)

[48] D. Beč irević, I. Doršner, S. Fajfer, D. A. Faroughy, N. Košnik, and O. Sumensari, "Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies," *Physical Review D*, vol. 98, sep 2018. (Cited on pages 12 and 113.)

[49] A. Angelescu, D. Beč irević, D. A. Faroughy, F. Jaffredo, and O. Sumensari, "Single leptoquark solutions to the B-physics anomalies," *Physical Review D*, vol. 104, sep 2021. (Cited on pages 12 and 113.)

[50] C. Cornella, D. A. Faroughy, J. Fuentes-Martín, G. Isidori, and M. Neubert, "Reading the footprints of the b-meson flavor anomalies," *Journal of High Energy Physics*, vol. 2021, aug 2021. (Cited on pages 12, 13 and 113.)

[51] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles," *Phys. Rev. Lett.*, vol. 38, pp. 1440–1443, Jun 1977. (Cited on page 14.)

[52] R. D. Peccei and H. R. Quinn, "Constraints imposed by CP conservation in the presence of pseudoparticles," *Phys. Rev. D*, vol. 16, pp. 1791–1797, Sep 1977. (Cited on page 14.)

[53] F. Wilczek, "Problem of strong p and t invariance in the presence of instantons," *Phys. Rev. Lett.*, vol. 40, pp. 279–282, Jan 1978. (Cited on page 14.)

[54] S. Weinberg, "A new light boson?," *Phys. Rev. Lett.*, vol. 40, pp. 223–226, Jan 1978. (Cited on page 14.)

[55] J. M. Camalich, M. Pospelov, P. N. H. Vuong, R. Ziegler, and J. Zupan, "Quark flavor phenomenology of the QCD axion," *Physical Review D*, vol. 102, jul 2020. (Cited on page 14.)

[56] E. Izaguirre, T. Lin, and B. Shuve, "Searching for axionlike particles in flavor-changing neutral current processes," *Phys. Rev. Lett.*, vol. 118, p. 111802, Mar 2017. (Cited on page 14.)

[57] S. Chakraborty, M. Kraus, V. Loladze, T. Okui, and K. Tobioka, "Heavy qcd axion in $b \to s$ transition: Enhanced limits and projections," *Phys. Rev. D*, vol. 104, p. 055036, Sep 2021. (Cited on page 14.)

[58] Planck Collaboration, "Planck 2013 results. i. overview of products and scientific results," *A&A*, vol. 571, p. A1, 2014. (Cited on page 14.)

[59] C. Bird, P. Jackson, R. Kowalewski, and M. Pospelov, "Dark matter particle production in $B \to s$ transitions with missing energy," *Phys. Rev. Lett.*, vol. 93, p. 201803, Nov 2004. (Cited on page 14.)

[60] A. Filimonova, R. Schäfer, and S. Westhoff, "Probing dark sectors with long-lived particles at belle II," *Physical Review D*, vol. 101, may 2020. (Cited on page 14.)

[61] P. del Amo Sanchez *et al.*, "Search for the rare decay $B \to K\nu\bar{\nu}$," *Physical Review D*, vol. 82, dec 2010. (Cited on pages 15 and 117.)

[62] J. P. Lees *et al.*, "Search for $B \to K^{(*)}\nu n\bar{u}$ and invisible quarkonium decays," *Physical Review D*, vol. 87, jun 2013. (Cited on pages 15, 110, 111 and 117.)

[63] O. Lutz *et al.*, "Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full belle $v(4s)$ data sample," *Physical Review D*, vol. 87, jun 2013. (Cited on pages 15, 110, 111 and 117.)

[64] J. Grygier *et al.*, "Search for $B \to h\nu\bar{\nu}$ decays with semileptonic tagging at belle," *Physical Review D*, vol. 96, nov 2017. (Cited on pages 15, 110 and 117.)

[65] F. Abudiné n *et al.*, "Search for $B^+ \to K^+\nu\bar{\nu}$ decays using an inclusive tagging method at belle ii," *Physical Review Letters*, vol. 127, oct 2021. (Cited on pages 15, 110 and 112.)

[66] P. D. Group, "Review of Particle Physics," *Progress of Theoretical and Experimental Physics*, vol. 2020, 08 2020. 083C01. (Cited on pages 18, 99 and 103.)

[67] D. Boutigny *et al.*, "BaBar technical design report," *aa*, 3 1995. (Cited on page 18.)

[68] A. Abashian *et al.*, “The Belle Detector,” *Nucl. Instrum. Meth. A*, vol. 479, pp. 117–232, 2002. (Cited on page 18.)

[69] A. J. Bevan *et al.*, “The physics of the b factories,” *The European Physical Journal C*, vol. 74, nov 2014. (Cited on pages 18 and 75.)

[70] SuperB Collaboration, “Superb: A high-luminosity asymmetric e+ e- super flavor factory. conceptual design report,” 2007. (Cited on page 20.)

[71] E. Kou *et al.*, “The belle II physics book,” *Progress of Theoretical and Experimental Physics*, vol. 2019, dec 2019. (Cited on pages 21 and 35.)

[72] J. Kemmer, G. Lutz, E. Belau, U. Prechtel, and W. Welser, “Low capacity drift diode,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 253, no. 3, pp. 378–381, 1987. (Cited on page 25.)

[73] K. Adamczyk *et al.*, “The design, construction, operation and performance of the belle ii silicon vertex detector,” *Journal of Instrumentation*, vol. 17, p. P11042, nov 2022. (Cited on pages 26 and 28.)

[74] M. French *et al.*, “Design and results from the APV25, a deep sub-micron cmos front-end chip for the CMS tracker,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 466, no. 2, pp. 359–365, 2001. 4th Int. Symp. on Development and Application of Semiconductor Tracking Detectors. (Cited on page 27.)

[75] I. Adachi, T. Browder, P. Križan, S. Tanaka, and Y. Ushiroda, “Detectors for extreme luminosity: Belle ii,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 46–59, 2018. Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn). (Cited on pages 29 and 32.)

[76] K. Kojima, “The operation and performance of the TOP detector at the Belle II experiment,” *PoS*, vol. EPS-HEP2021, p. 803, 2022. (Cited on page 30.)

[77] M. Yonenaga *et al.*, “Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data,” *Progress of Theoretical and Experimental Physics*, vol. 2020, 08 2020. 093H01. (Cited on page 31.)

[78] T. Aushev *et al.*, “A scintillator based endcap kl and muon detector for the belle ii experiment,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 789, pp. 134–142, 2015. (Cited on page 33.)

[79] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, “The belle II core software,” *Computing and Software for Big Science*, vol. 3, nov 2018. (Cited on page 35.)

[80] “Standard C++ foundation.” (Cited on page 35.)

[81] G. V. Rossum and F. L. D. Jr., “Python reference manul,” 1995. (Cited on page 35.)

[82] R. Brun and F. Rademakers, “Root — an object oriented data analysis framework,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1, pp. 81–86, 1997. New Computing Techniques in Physics Research V. (Cited on page 36.)

[83] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth. A*, vol. 462, pp. 152–155, 2001. (Cited on page 36.)

[84] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An introduction to PYTHIA 8.2,” *Computer Physics Communications*, vol. 191, pp. 159–177, jun 2015. (Cited on pages 36 and 104.)

[85] N. Davidson, G. Nanava, T. Przedziński, E. Richter-Was, and Z. Was, “Universal interface of tauola: Technical and physics documentation,” *Computer Physics Communications*, vol. 183, pp. 821–843, mar 2012. (Cited on page 36.)

[86] G. Balossini, C. Bignamini, C. C. Calame, G. Montagna, O. Nicrosini, and F. Piccinini, “Photon pair production at flavour factories with per mille accuracy,” *Physics Letters B*, vol. 663, pp. 209–213, may 2008. (Cited on page 36.)

[87] G. Balossini, C. M. C. Calame, G. Montagna, O. Nicrosini, and F. Piccinini, “Matching perturbative and parton shower corrections to bhabha process at flavour factories,” *Nuclear Physics B*, vol. 758, pp. 227–253, dec 2006. (Cited on page 36.)

[88] C. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini, “The babayaga event generator,” *Nuclear Physics B - Proceedings Supplements*, vol. 131, pp. 48–55, 2004. SIGHADO3. (Cited on page 36.)

[89] C. M. C. Calame, “An improved parton shower algorithm in QED,” *Physics Letters B*, vol. 520, pp. 16–24, nov 2001. (Cited on page 36.)

[90] C. C. Calame, C. Lunardini, G. Montagna, O. Nicrosini, and F. Piccinini, “Large-angle bhabha scattering and luminosity at flavour factories,” *Nuclear Physics B*, vol. 584, pp. 459–479, sep 2000. (Cited on page 36.)

[91] F. A. Berends, P. H. Daverveldt, and R. Kleiss, "Monte Carlo Simulation of Two Photon Processes. 2. Complete Lowest Order Calculations for Four Lepton Production Processes in electron Positron Collisions," *Comput. Phys. Commun.*, vol. 40, pp. 285–307, 1986. (Cited on page 36.)

[92] F. Berends, P. Daverveldt, and R. Kleiss, "Monte carlo simulation of two-photon processes: Ii: Complete lowest order calculations for four-lepton production processes in electron-positron collisions," *Computer Physics Communications*, vol. 40, no. 2, pp. 285–307, 1986. (Cited on page 36.)

[93] F. Berends, P. Daverveldt, and R. Kleiss, "Radiative corrections to the process $e+e- \to e^+e^-\mu^+\mu^-$," *Nuclear Physics B*, vol. 253, pp. 421–440, 1985. (Cited on page 36.)

[94] "Strategic accelerator design(a)." (Cited on page 36.)

[95] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. (Cited on page 36.)

[96] J. Allison *et al.*, "Geant4 developments and applications," *IEEE Transactions on Nuclear Science*, vol. 53, no. 1, pp. 270–278, 2006. (Cited on page 36.)

[97] V. Bertacchi *et al.*, "Track finding at belle ii," *Computer Physics Communications*, vol. 259, p. 107610, 2021. (Cited on pages 37 and 71.)

[98] T. Bilka, N. Braun, T. Hauth, T. Kuhr, L. Lavezzi, F. Metzner, S. Paul, E. Prencipe, M. Prim, J. Rauch, J. Ritman, T. Schlueter, and S. Spataro, "Implementation of genfit2 as an experiment independent track-fitting framework," 2019. (Cited on page 37.)

[99] S. Ramo, "Currents induced by electron motion," *Proceedings of the IRE*, vol. 27, no. 9, pp. 584–585, 1939. (Cited on page 42.)

[100] G. Punzi, "Sensitivity of searches for new signals and its optimization," 2003. (Cited on pages 55 and 61.)

[101] e. a. Keck, T, "The full event interpretation," *Computing and Software for Big Science*, vol. 3, 2019. (Cited on pages 55, 56, 57 and 114.)

[102] 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*, KDD '16, (New York, NY, USA), pp. 785–794, Association for Computing Machinery, 2016. (Cited on pages 60 and 80.)

[103] Y. Coadou, "Boosted decision trees," in *Artificial Intelligence for High Energy Physics*, pp. 9–58, WORLD SCIENTIFIC, feb 2022. (Cited on page 60.)

[104] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, "HistFactory: A tool for creating statistical models for use with RooFit and RooStats," tech. rep., New York U., New York, 2012. (Cited on page 62.)

[105] M. Feickert, L. Heinrich, and G. Stark, "pyhf: pure-python implementation of histfactory with tensors and automatic differentiation," 2022. (Cited on pages 62 and 64.)

[106] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, p. 1554, 2011. (Cited on page 66.)

[107] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60 – 62, 1938. (Cited on page 66.)

[108] J. Kemmer, G. Lutz, E. Belau, U. Prechtel, and W. Welser, "Low capacity drift diode," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 253, no. 3, pp. 378–381, 1987. (Cited on page 72.)

[109] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in e^+e^- annihilation and other processes," *Phys. Rev. Lett.*, vol. 41, pp. 1581–1585, Dec 1978. (Cited on page 75.)

[110] G. Fox and S. Wolfram, "Event shapes in e^+e^- annihilation," *Nuclear Physics B*, vol. 157, no. 3, pp. 543–544, 1979. (Cited on page 75.)

[111] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019. (Cited on pages 82 and 115.)

[112] M. Ablikim *et al.*, "Measurements of branching fractions for inclusive $\bar{K}^0/k^0$ and $k^{*\pm}(892)$ decays of neutral and charged d mesons," *Physics Letters B*, vol. 643, no. 5, pp. 246–250, 2006. (Cited on page 98.)

[113] The Belle II Collaboration, "Measurement of the tracking efficiency and fake rate with $e^+e^- \to \tau^+\tau^-$ events," Jul 2020. (Cited on page 99.)

[114] B. Aubert *et al.*, "Measurement of the $B^+ \to p\overline{p}K^+$ branching fraction and study of the decay dynamics," *Phys. Rev. D*, vol. 72, p. 051101, Sep 2005. (Cited on page 101.)

[115] J. P. Lees *et al.*, "Study of cp violation in dalitz-plot analyses of $B^0 \to K^+K^-K^0_S$, $B^+ \to K^+K^-K^+$, and $B^+ \to K^0_S K^0_S K^+$," *Phys. Rev. D*, vol. 85, p. 112010, Jun 2012. (Cited on pages 102, 103 and 104.)

Acknowledgments