

Experimental study of inclusive charmless semileptonic beauty meson decays at Belle II

Dissertation zur Erlangung des Doktorgrades

an der

Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Tommy Borislavov MARTINOV

Hamburg
2025



Gutachter/innen der Dissertation:

Prof. Dr. Peter Schleper
Prof. Dr. Kerstin Tackmann

Zusammensetzung der Prüfungskommission:

Dr. Lydia Beresford
Prof. Dr. Sven-Olaf Moch
Dr. Krisztian Peters
Prof. Dr. Peter Schleper
Prof. Dr. Kerstin Tackmann

Vorsitzender der Prüfungskommission:

Prof. Dr. Sven-Olaf Moch

Datum der Disputation:

11 September 2025

Vorsitzender Fach-Promotionsausschusses PHYSIK:

Prof. Dr. Wolfgang J. Parak

Leiter des Fachbereichs PHYSIK:

Prof. Dr. Markus Drescher

Dekan der Fakultät MIN:

Prof. Dr.-Ing. Norbert Ritter

Abstract

A measurement of partial branching fractions of charmless semileptonic beauty meson decays using an inclusively reconstructed hadronic system is presented. These decays are denoted as $B \rightarrow X_u \ell \nu$, where X_u represents a hadronic system containing a u quark produced in the underlying quark-level $b \rightarrow u$ transition. The measurement uses the full Run 1 Belle II sample corresponding to 365 fb^{-1} of $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ data. One of the two B mesons is reconstructed in a hadronic decay channel to constrain the signal kinematics. Events containing a signal electron or muon ℓ are selected, while the rest of the event defines the hadronic system X_u associated with the signal. To discriminate the signal from the 50-times larger background originating from Cabibbo-favoured semileptonic $B \rightarrow X_c \ell \nu$ decays, a template fit is performed in both signal and control regions after applying an optimised selection. The control regions are used to constrain the kinematic properties of $B \rightarrow X_c \ell \nu$ decays based on data and therefore reduce the dependence of the measurement on their modelling. The partial branching fraction measured for lepton energies greater than 1 GeV in the signal B meson rest frame is

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3},$$

where the uncertainties are statistical and systematic, respectively. From this measurement, using the Gambino, Giordano, Ossola and Uraltsev theoretical framework, the strength of the beauty-to-up quark coupling is determined to be

$$|V_{ub}| = (4.01 \pm 0.19_{-0.08}^{+0.07}) \times 10^{-3},$$

where the uncertainties are experimental and theoretical, respectively. This value is in good agreement with the world average of $|V_{ub}| = (4.06 \pm 0.16) \times 10^{-3}$. Different theoretical predictions and partial branching fractions measured in other phase-space regions, defined by additional selections on the X_u and leptonic system masses, are also used to determine $|V_{ub}|$.

Zusammenfassung

In dieser Arbeit wird eine Messung der partiellen Verzweungsverhältnisse charmloser semileptonischer Zerfälle von Beauty-Mesonen vorgestellt, bei der ein hadronisches System inklusiv rekonstruiert wird. Diese Zerfälle werden als $B \rightarrow X_u \ell \nu$ bezeichnet, wobei X_u ein hadronisches System darstellt, das ein u -Quark enthält, das im zugrunde liegenden Quark-Übergang $b \rightarrow u$ erzeugt wird. In dieser Arbeit wird eine Messung partieller Verzweungsverhältnisse charmloser semileptonischer $B \rightarrow X_u \ell \nu$ Zerfälle unter Verwendung eines inklusiv rekonstruierten hadronischen Systems vorgestellt. Die Messung basiert auf dem vollständigen Run 1-Datensatz von Belle II, der 365 fb^{-1} an $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ -Daten enthält. Einer der beiden B -Mesonen wird in einem hadronischen Zerfallskanal rekonstruiert, um die Kinematik des Signals zu bestimmen. Es werden Ereignisse ausgewählt, die ein Signal-Elektron oder -Myon ℓ enthalten, während der restliche Teil des Ereignisses das hadronische System X_u definiert, das mit dem Signal assoziiert ist. Um das Signal von dem 50-mal größeren Hintergrund aus Cabibbo-bevorzugten semileptonischen $B \rightarrow X_c \ell \nu$ Zerfällen zu unterscheiden, wird ein Template-Fit sowohl in Signal- als auch in Kontrollregionen nach Anwendung einer optimierten Selektion durchgeführt. Die Kontrollregionen werden verwendet, um die kinematischen Eigenschaften der $B \rightarrow X_c \ell \nu$ Zerfälle basierend auf Daten zu bestimmen und somit die Abhängigkeit der Messung von der Modellierung dieser Zerfälle verringern. Das partielle Verzweungsverhältnis für Leptonenergien größer als 1 GeV im Ruhesystem des Signal- B -Mesons beträgt

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3},$$

wobei die angegebenen Unsicherheiten statistisch bzw. systematisch sind. Aus dieser Messung ergibt sich unter Verwendung des theoretischen Modells von Gambino, Giordano, Ossola und Uraltsev die Kopplungsstärke zwischen Beauty- und Up-Quark zu

$$|V_{ub}| = (4.01 \pm 0.19^{+0.07}_{-0.08}) \times 10^{-3},$$

wobei die Unsicherheiten experimentell bzw. theoretisch sind. Dieser Wert steht in guter Übereinstimmung mit dem Weltmittelwert von $|V_{ub}| = (4.06 \pm 0.16) \times 10^{-3}$. Zur Bestimmung von $|V_{ub}|$ werden ebenfalls verschiedene theoretische Vorhersagen sowie in anderen Phasenraumbereichen gemessene partielle Verzweungsverhältnisse, die durch zusätzliche Selektionen auf die Massen des X_u - und des leptonischen Systems definiert sind, herangezogen.

Disclaimer

The results presented in this thesis are the outcome of collaborative efforts.

The subject of this thesis was proposed by my PhD advisor, Prof. Dr. Kerstin Tackmann.

This work is based on data collected with the Belle II detector which was designed, installed and maintained by members of the Belle II collaboration. The Belle II detector collects data from electron-positron collisions provided by the SuperKEKB accelerator which is operated by the SuperKEKB collaboration. The work presented in this thesis partially relied on the Belle II software framework which is maintained by members of the Belle II collaboration. Whenever internal results from the Belle II collaboration are used, they are explicitly cited. Part of this work was performed on the National Analysis Facility hosted at DESY.

Dr. Marcel Hohmann contributed to the measurement presented in this thesis by developing the code for event reconstruction, scaling of events (including branching fractions, continuum calibration, hadronic tagging, particle identification and K_S^0 efficiency) and parts of the code for computing systematic uncertainties in these areas. He also developed utility tools for plotting and implementing systematic uncertainties in the signal extraction framework.

The measurement received feedback from Prof. Dr. Florian Bernlochner and Dr. Lu Cao as part of the analysis team; from Dr. Mirco Dorigo, Prof. Dr. Robert Kowalewski and Dr. Alessandro Gaz as members of the Review Committee; from Dr. Kazuki Kojima, Dr. Chia-Ling Hsu and Dr. Markus Prim as part of the Working Group Review Committee; and from Dr. Keri Vos regarding theoretical aspects.

All remaining work, including the writing of the text and the preparation of plots (unless otherwise stated), was carried out by myself.

The results and figures presented here are preliminary.

This work was supported by the Helmholtz-Gemeinschaft W2/W3-116 grant.

Declaration on oath

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Tommy Martinov
Hamburg, 16 September 2025

A handwritten signature in black ink, appearing to read 'Tommy Martinov', is written on a light gray rectangular background.

Acknowledgments

Many people contributed directly or indirectly to the success of this PhD and I would like to thank them all.

First of all, I would like to thank my advisor, Kerstin Tackmann, who accepted me as one of her PhD students and was always present to answer my questions and propose solutions to the difficulties I faced.

Then I would like to thank Marcel Hohmann without whom the work presented here would've never reached the stage it did. I used for my measurement several pieces of code written by Marcel who's computing skills far surpass mine. His experience with semileptonic B decays made him the perfect ideas man and he was always available to help me find an answer to the problems the analysis was throwing at me. Apart from our common work, I am happy to call him a friend with whom I spent heaps of time arguing and complaining.

Many people contributed to my work by giving valuable feedback. I would like to thank Lu Cao who guided me through the early parts of the measurement and whose experience with the subject was particularly helpful; Florian Bernlochner who was always able to understand very quickly what I was talking about and gave extremely useful ideas; Mirco Dorigo, Bob Kowalewski and Ale Gaz who pushed me to explore the smallest details of my analysis and allowed me to find hidden issues; Kojima Kazuki-san, Chia-Ling Hsu and Markus Prim for giving useful comments during the first stage of review; and finally Keri Vos for taking the time to answer my questions about the very complex theoretical description of inclusive semileptonic B meson decays.

I am thankful to the SuperKEKB and Belle II collaborations for providing me with the data, software framework and performance studies necessary to conduct my measurement.

I would like to thank all the people who took the time to read parts of this thesis and helped me to improve its content: Ivan Novikov and Guilherme Guedes who read the theory chapter; Christian Wessel who read the Belle II and SuperKEKB descriptions; Kerstin, Luísa Carvalho and Riccardo Manfredi who read the measurement chapter; Merle Graf-Schreiber who helped me write the Zusammenfassung and finally Marcel who read almost every chapter.

I want to thank all the members of my PhD committee for taking the time to read this thesis and attend the defence.

I would also like to thank Filippo Dattola, a physics nerd like me, for the many useful discussions, not only about our work but about life in general. During my time at DESY, my free time in Hamburg and the conferences and schools I attended, I made many friends. Since I would feel bad about forgetting anyone of them, I will not try to make an exhaustive list but they will all recognise themselves.

Finally, I would like to thank my family for their unconditional love and support. I am grateful to them for helping me pursue my dream of working in fundamental physics and always being there for me, as much in the worst times as in the best.

*To my loving family, who will understand none of what follows and
simply wishes I had become an engineer.*

Contents

1	Introduction	13
2	Theoretical introduction to semileptonic beauty meson decays	17
2.1	Introduction: the Standard Model Lagrangian	17
2.1.1	Quantum electrodynamics	18
2.1.2	Quantum chromodynamics	19
2.1.3	The weak interaction and electroweak unification	21
2.1.4	Electroweak symmetry breaking: the Brout-Englert-Higgs mechanism	23
2.2	The CKM matrix and flavour physics	26
2.3	Semileptonic B decays	29
2.3.1	Theoretical methods	31
2.3.2	Exclusive semileptonic $B \rightarrow X_c \ell \nu$ decays	32
2.3.3	Exclusive semileptonic $B \rightarrow X_u \ell \nu$ decays	35
2.3.4	Inclusive semileptonic $B \rightarrow X_c \ell \nu$ decays	37
2.3.5	Inclusive semileptonic $B \rightarrow X_u \ell \nu$ decays	38
2.4	Experimental overview	46
2.4.1	Exclusive $B \rightarrow X_u \ell \nu$ measurements	46
2.4.2	Inclusive $B \rightarrow X_u \ell \nu$ measurements	47
2.4.3	Summary of $B \rightarrow X_u \ell \nu$ measurements	48
2.5	Additional remarks	48
2.6	Conclusion and role of the measurement	50
3	Experimental setup	53
3.1	B factories	53
3.2	The SuperKEKB collider	54
3.3	The Belle II detector	56
3.3.1	The Pixel Detector (PXD)	57
3.3.2	The Silicon Vertex Detector (SVD)	57
3.3.3	The Central Drift Chamber (CDC)	58
3.3.4	Particle identification (ARICH and TOP)	59
3.3.5	The Electromagnetic Calorimeter (ECL)	60
3.3.6	K_L^0 and muon detection system (KLM)	60
3.3.7	Trigger	61
3.3.8	Particle identification	61
3.4	The Belle II software	62
3.4.1	Simulation	62
4	Analysis techniques and tools	65
4.1	Maximum likelihood estimation	65
4.1.1	The HistFactory/Pyhf model	66
4.2	Multivariate classification	68
4.2.1	Boosted Decision Trees	69
4.2.2	Neural Networks	70

4.2.3	Hyperparameter optimisation	71
4.2.4	Multivariate classifier metrics	72
4.3	Full event interpretation	73
4.4	Form factor reweighting	74
4.4.1	eFFORT	74
4.4.2	Hammer	75
5	Measurement of inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions and V_{ub}	77
5.1	Analysis overview	77
5.2	Data and simulated data samples	79
5.2.1	The Belle II dataset	79
5.2.2	Simulated samples	79
5.3	Event reconstruction	84
5.3.1	Kinematic variables	85
5.3.2	Tag Side	85
5.3.3	Signal Side	86
5.3.4	Additional selections	90
5.4	Sample corrections	90
5.4.1	Branching fractions	90
5.4.2	Form factors	91
5.4.3	Broad D^{**} modelling	91
5.4.4	Continuum calibration	92
5.4.5	B tagging efficiency	96
5.4.6	Charged particle identification	96
5.4.7	K_S^0 reconstruction efficiency	97
5.4.8	Slow pion reconstruction efficiency	97
5.5	Machine learning selection	98
5.5.1	Continuum suppression	98
5.5.2	$B \rightarrow X_c \ell \nu$ suppression	103
5.6	Kinematic selections	106
5.7	Signal region	107
5.8	Agreement between data and simulated data	112
5.8.1	Control regions	112
5.8.2	$B \rightarrow X_c \ell \nu$ normalisation correction	115
5.8.3	$B \rightarrow X_c \ell \nu$ shape correction	116
5.8.4	$B \rightarrow X_u \ell \nu$ modelling validation	120
5.9	Signal extraction	121
5.9.1	Fitted regions	121
5.9.2	Fit templates	121
5.9.3	Fitted variables	122
5.9.4	Systematic uncertainties	122
5.9.5	Signal extraction validation	128
5.10	Pre-unblinding tests	137
5.10.1	Signal region distributions	137
5.10.2	Split fits	138
5.11	Results	141
5.11.1	Prefit distributions	141
5.11.2	Fit outcome	142
5.11.3	Partial branching fractions	146
5.11.4	$ V_{ub} $ determination	151

6	Conclusions and outlook	155
A	Simulation of inclusive $B \rightarrow X_u \ell \nu$ decays	159
B	Input features of the continuum suppression classifier	163
B.1	Thrust-axis variables	163
B.2	Kakuno-Super-Fox-Wolfram moments	164
B.3	CLEO cones	166
B.4	Other variables	167
C	Input features of the $B \rightarrow X_c \ell \nu$ suppression classifier	169
D	Comparison between data and simulated data in control regions	171
D.1	$\text{CR}_{K,\text{low}}$	171
D.2	$\text{CR}_{K,\text{high}}$	172
E	$D \rightarrow K_L^0 X$ modelling	173
F	Fits 2 and 3 figures	175
F.1	Fit 2	175
F.2	Fit 3	181
	Bibliography	187

Chapter 1

Introduction

“If you haven’t found something strange during the day, it hasn’t been much of a day.”

John A. Wheeler

With the discovery of a plethora of elementary and composite particles over the course of the past 130 years, the field of particle physics has rapidly developed. Our knowledge of particles and their fundamental interactions can be summarised in the Standard Model (SM). While being able to predict a wide range of phenomena with excellent accuracy, there are several questions which the SM is unable to address, such as the origin of neutrino masses, the nature of dark matter, the matter-antimatter asymmetry and the description of gravity as a quantum theory. Furthermore, several characteristics of the SM appear to have no satisfactory justification. For instance, there is no explanation for the cancellation of large corrections to the Higgs boson mass giving it an unexpectedly small value, a problem known as the *hierarchy problem*, or for the hierarchical structure of fermion masses, a problem known as the *flavour puzzle*. All these open questions are clear indications of the incompleteness of the SM and call for the development of a deeper theory.

While there exist many different areas of particle physics which are able to challenge the SM as an attempt to answer these questions, the present manuscript focuses on the properties of quarks and leptons, an area known as the *flavour sector*. Flavour physics offers a large variety of tests to probe the SM. For instance, the SM predicts that the three generations of charged leptons (electrons, muons and taus) have equal electroweak coupling strengths, a characteristic known as *lepton flavour universality* (LFU). The only distinction in their electroweak interactions originates from their different masses and an observed deviation from LFU would be a clear indication of non-SM effects. Furthermore, such effects could be uncovered in rare hadron decays (for example, the famous penguin decays), which are expected to be sensitive to interactions with particles predicted by non-SM theories. In this thesis, however, we investigate more specifically changes between quark flavours. Within the SM, a quark from the up generation (up, charm, top) can transition to a quark from the down generation (down, strange, beauty) and vice versa. Such interactions are governed by the 3×3 unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix, which is near diagonal, another feature of the SM which has no apparent explanation. The matrix has nine complex elements, resulting in a total of 18 parameters which, because of the matrix unitarity and invariance under phase transformations, can be reduced to four, three rotation angles and one complex phase. There exist different ways to parametrise the CKM matrix and measure its elements or combinations of them and this wide variety of tests represents a powerful probe of the CKM unitarity and therefore of the SM.

Of the nine CKM matrix elements, V_{ub} has the smallest magnitude and one of the largest relative uncertainties. The magnitude of this element can be measured via tree-level transitions, which are *a priori* insensitive to non-SM effects but the large uncertainty on its value limits the

precision of the CKM unitarity test. It is therefore crucial to improve our understanding of $|V_{ub}|$ to better constrain the CKM matrix and, in turn, enhance the sensitivity of other CKM tests to non-SM interactions.

There exist different ways of measuring $|V_{ub}|$ but the two methods which currently yield the best precision rely on decays of beauty mesons to a charmless hadronic system, a lepton and a neutrino (*semileptonic* decays) which are denoted $B \rightarrow X_u \ell \nu$. First, in exclusive measurements, the hadronic system is reconstructed as a specific final state, such as a pion in $B \rightarrow \pi \ell \nu$ transitions. The theoretical description of exclusive semileptonic decays is based on functions known as *form factors*, which encapsulate the non-perturbative dynamics of the transition. On the other hand, inclusive charmless semileptonic decays can be described by the *heavy quark expansion* (HQE), an expansion in powers of the strong coupling constant, α_s , and inverse powers of the b quark mass, m_b . However, selections on kinematic variables such as the lepton energy or the X_u mass are necessary in experimental analyses to suppress backgrounds, in particular from Cabibbo-favoured semileptonic decays to a hadronic system containing a charm quark denoted $B \rightarrow X_c \ell \nu$. The HQE breaks down because of these kinematic selections and instead, predictions of differential decay rates are necessary. These predictions are sensitive to the *shape function* - a non-perturbative object describing the motion of the b quark inside the meson - in particular in restricted regions of phase space. Careful treatment of such non-perturbative effects has led to the development of various theoretical frameworks yielding different predictions of the partial decay rate over a given region of phase space, $\Delta\Gamma(B \rightarrow X_u \ell \nu)$ (excluding the $|V_{ub}|^2$ term). A value of $|V_{ub}|$ can be extracted from the measurement of an inclusive partial branching fraction over the same phase space, $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$, via

$$|V_{ub}| = \sqrt{\frac{\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)}{\tau_B \Delta\Gamma(B \rightarrow X_u \ell \nu)}},$$

where τ_B is the B meson lifetime. The two most recent averages of $|V_{ub}|$ extracted from exclusive and inclusive decays read

$$\begin{aligned} |V_{ub}|_{\text{excl}} &= (3.43 \pm 0.12) \times 10^{-3}, \\ |V_{ub}|_{\text{incl}} &= (4.06 \pm 0.16) \times 10^{-3}, \end{aligned}$$

where the total uncertainty is given. The two values of $|V_{ub}|$ disagree by about three standard deviations, thereby limiting the precision with which this CKM matrix parameter can be determined.

In this thesis, we present the measurement of partial branching fractions of $B \rightarrow X_u \ell \nu$ decays in three different regions of phase-space, defined by selections on the lepton energy, the X_u mass and the lepton-neutrino system mass, using a dataset of 365 fb^{-1} collected by the Belle II experiment. The overwhelming background from $B \rightarrow X_c \ell \nu$ decays is suppressed using a multivariate classifier which allows us to consider relatively broad phase-space regions by applying looser kinematic selections. The branching fractions are extracted with a binned template fit to a signal region and a control region used to correct the modelling of the $B \rightarrow X_c \ell \nu$ background. From each of these branching fractions, $|V_{ub}|$ is determined based on three different theoretical predictions of the partial decay rate. The partial branching fraction measured in a region of lepton energy greater than 1 GeV in the signal B meson rest frame is

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3},$$

where the uncertainties are statistical and systematic, respectively. From this measurement,

using the Gambino, Giordano, Ossola and Uraltsev theoretical framework, the magnitude of V_{ub} is determined to be

$$|V_{ub}| = (4.01 \pm 0.19^{+0.07}_{-0.08}) \times 10^{-3},$$

where the uncertainties are experimental and theoretical, respectively.

The remainder of this thesis is organised as follows: we introduce the main elements of the Standard Model and the theoretical frameworks necessary to describe semileptonic B meson decays in Chapter 2; we outline the SuperKEKB accelerator and the Belle II experiment in Chapter 3; we discuss the analysis techniques used to perform the measurement in Chapter 4; we present the measurement in detail in Chapter 5; and finally, we conclude in Chapter 6.

Throughout this thesis, natural units $\hbar = c = 1$ are used. Unless explicitly stated, charge conjugation is implied and when no charge is added to a particle label, all possible charges are considered (for example, B is used to label both neutral and charged beauty mesons). Unless explicitly stated, the term lepton includes electrons (e) and muons (μ).

Chapter 2

Theoretical introduction to semileptonic beauty meson decays

“Aristotle said a bunch of stuff that was wrong. Galileo and Newton fixed things up. Then Einstein broke everything again. Now, we’ve basically got it all worked out, except for small stuff, big stuff, hot stuff, cold stuff, fast stuff, heavy stuff, dark stuff, turbulence, and the concept of time.”

Zach Weinersmith

In this chapter, we present the theoretical frameworks needed to study inclusive semileptonic B decays. We start by introducing the Standard Model of particle physics and outline its fundamental components: quantum electrodynamics, quantum chromodynamics and the weak interaction. We then present the Higgs mechanism, which leads to electroweak symmetry breaking and provides an explanation for particle masses through the Yukawa couplings. From that point, we introduce the CKM matrix and, more generally, flavour physics. We then focus specifically on semileptonic decays of B mesons and finally, we give an overview of recent experimental results in that field.

2.1 Introduction: the Standard Model Lagrangian

The quest to understand the most fundamental elements of our universe started more than a century ago with the discovery of the electron by Sir J. J. Thomson [1]. Since then, many Nobel Prize winners and many different experiments have paved the way to the current deep (and yet incomplete) understanding of particle physics. The Standard Model (SM) of particle physics is usually summarised by its particle content, which is illustrated in Figure 2.1. Fundamental particles are characterised by various properties such as their masses, their spin as well as other quantum numbers referred to as charges. Subsequently, these particles can be divided into different categories. Fermions and bosons can be separated by their spin, the former carrying a spin of $\frac{1}{2}$ while the latter carry an integer spin. Among fermions, those that do not interact via the strong force are called leptons. The three electrically charged leptons are, from lightest to heaviest: the electron e , the muon μ and the tau τ . Their electrical charge allows them to interact via the electromagnetic force. Similarly, there are three generations of neutral leptons which only interact via the weak force, called neutrinos, each one associated to a charged lepton: the electron neutrino ν_e , the muon neutrino ν_μ and the tau neutrino ν_τ . The SM predicts the neutrinos to be massless but from the phenomenon of neutrino oscillations, it is now known that they do have masses which are extremely small compared to the masses of other SM particles. The quarks, on the other hand, carry fractional electric charges; $+\frac{2}{3}$ for the three up-like quarks: the up u , the charm c and the top t (also called truth) from lightest to

heaviest; and $-\frac{1}{3}$ for the three down-like quarks: the down d , the strange s and the beauty b (also called bottom). Quarks also carry a charge called colour, which makes them the only SM fermions which can interact via the strong force. The spin-1 bosons are, within the SM, the force carriers: the massless neutral photon γ for the electromagnetic interaction, the massless neutral gluons g for the strong interaction and the massive charged $W^\pm$ and neutral Z^0 for the weak interaction. The electromagnetic and weak interactions can be in fact unified in a single force: the electroweak interaction. Finally, the massive neutral spin-0 Higgs boson is introduced in the SM by the Brout-Englert-Higgs mechanism, which describes the origin of particle masses.

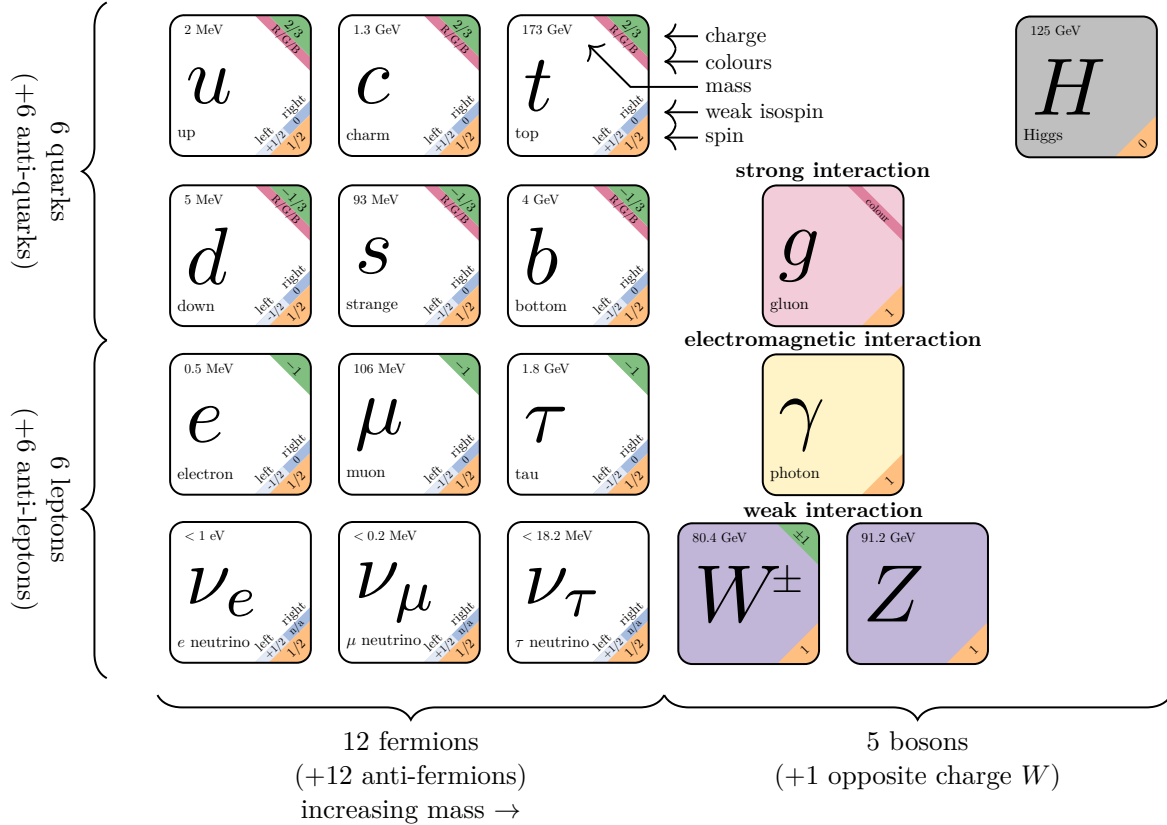


FIGURE 2.1: Summary of the fundamental particle content of the Standard Model. Adapted from Ref. [2].

The measurement presented in this thesis focuses particularly on the flavour¹ sector of particle physics, a research area which investigates the properties of quarks and leptons. In the following sections, we briefly describe the different pieces of the SM mentioned above before presenting flavour physics and semileptonic decays more specifically.

This chapter is based on the author's knowledge as well as information from Refs. [4–6].

2.1.1 Quantum electrodynamics

Quantum electrodynamics (QED) was the first quantum field theory (QFT) developed to describe fundamental particles. Its initial formulation, in the 1920s, is credited to Paul Dirac.

¹A term coined by Murray Gell-Mann and his student Harald Fritzsch in 1971 while eating ice-cream in Pasadena, California [3].

Starting from the Lagrangian for a free electron field ψ with mass m [4, 5]

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - m\bar{\psi}\psi, \quad (2.1)$$

where γ_μ is the set of γ matrices, one can introduce an additional (gauge) field A_μ representing the photon and impose to ψ and A_μ invariance under the local $U(1)_{\text{EM}}$ gauge transformations

$$\psi \rightarrow e^{ie\lambda(x)}\psi, \quad A^\mu \rightarrow A^\mu - \partial^\mu\lambda, \quad (2.2)$$

with e the elementary charge and λ some phase depending on the space-time coordinate x . By adding a free-field part for A_μ , the full QED Lagrangian, invariant under the local gauge transformations specified in Equation 2.2, therefore reads:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma_\mu\partial^\mu - m)\psi - e\bar{\psi}\gamma_\mu\psi A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.3)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. It is important to note that no mass term for the gauge field can be added to the Lagrangian as such term of the form $m_\gamma^2 A_\mu A^\mu$ would break the required gauge invariance. QED describes all interactions between charged fermions and photons as illustrated in Figure 2.2. QED was developed as a perturbative theory, expanded in its coupling

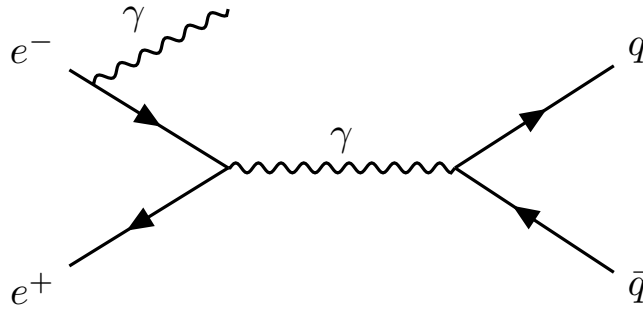


FIGURE 2.2: Example of an electron-positron annihilation producing a pair of quarks together with a photon radiated from the initial state.

constant $\alpha_{\text{em}} \equiv \frac{e^2}{4\pi} \approx \frac{1}{137} \ll 1$. It was however noticed during the 1930s that the theory was only valid at the lowest order and that at higher orders, computations yield infinities. After much effort, this problem was solved by understanding that these infinities could be absorbed by other infinite corrections to QED parameters such that the fundamental parameters remain finite and in fact close to their experimental values. This process is known as renormalisation. It is beyond the scope of this thesis to describe renormalisation of QED and the SM but two important consequences of renormalisation should be mentioned. First, the renormalisation procedure defines bare parameters, which can be infinite and renormalised parameters, which are finite, therefore determining the physical meaning of each observable in the theory. Second, through the renormalisation group equations, the behaviour of the theory at different energy scales can be related. A particular consequence is that coupling constants are no longer constant but depend on the energy scale of the considered processes, which is referred to as *running* of the coupling constants. For instance, at the Z mass scale α_{em} becomes $\alpha_{\text{em}} \approx \frac{1}{128}$.

2.1.2 Quantum chromodynamics

The theory of quantum chromodynamics (QCD) was developed later than QED, in the 1960s, following bubble chamber experiments where a large number of particles, known as hadrons,

were discovered. QCD, as opposed to QED, is a non-Abelian gauge theory, meaning that two successive gauge transformations do not commute. The strong interaction is described by an $SU(3)$ gauge theory with eight gauge bosons (the gluons whose fields are denoted A^a , $a = 1, \dots, 8$) and the QCD Lagrangian including the six types of quarks is written as

$$\mathcal{L}_{\text{QCD}} = \sum_{f \in \{u, d, s, c, b, t\}} \bar{\psi} \left(i \not{\partial} - g_s \not{A}^a \frac{\lambda^a}{2} - m_f \right) \psi - \frac{1}{4} G_{\mu\nu}^a G^{a\mu\nu}, \quad (2.4)$$

where for any element p , $\gamma^\mu p_\mu = \not{p}$, g_s is the strong coupling constant, λ^a are the eight 3×3 Gell-Mann matrices and $G_{\mu\nu}^a$ is the gluon field strength. The QCD charge, called colour, can take three values, labelled red, green and blue. In contrast with QED, the gluons also carry colour charge and they can therefore interact with themselves, giving rise to triple and quartic gluon vertices as illustrated in Figure 2.3. An important feature of QCD is that no free quark

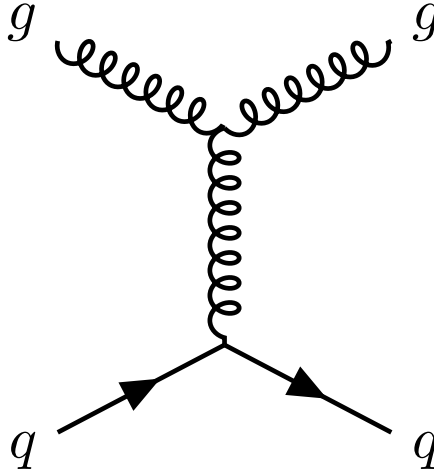


FIGURE 2.3: Example of a QCD Feynman diagram with a triple gauge coupling

has ever been observed. This phenomenon is explained by the hypothesis of colour confinement stating that particles carrying colour are always confined in colour singlet (*i.e.* colourless) states. Even though there is no analytic proof of colour confinement, it can be understood by thinking about the (virtual) gluon-gluon interactions between two quarks that are pulled away from each other. Because of all these interactions, the colour field is confined to a tube which has a constant energy density. The energy of the colour field is therefore proportional to the distance between the quarks and it would require an infinite amount of energy to separate them. The only colour singlet states containing two and three quarks are of the type $q\bar{q}$ (mesons) and qqq (baryons). Exotic states with four, five or even more quarks are also allowed and have been observed in nature [4, 7].

The running of the strong force coupling constant $\alpha_s \equiv \frac{g_s^2}{4\pi}$ at an energy scale q^2 is expressed as²

$$\alpha_s(q^2) = \frac{\alpha_s(\mu^2)}{1 + b_{\text{QCD}}\alpha_s(\mu^2)\log\left(\frac{q^2}{\mu^2}\right)}, \quad (2.5)$$

where α_s is known at some energy scale $q^2 = \mu^2$ and

$$b_{\text{QCD}} = \frac{11N_C - 2N_f}{12\pi}, \quad (2.6)$$

where $N_C = 3$ is the number of SM colours and $N_f \leq 6$ the number of active quarks, *i.e.* the number of quarks not neglected in the theory, so six in the full SM. As b_{QCD} is greater than zero, α_s decreases with the energy scale, meaning that at very large energies, quarks can be considered as quasi-free particles. With a small value of α_s , perturbative calculations can be performed but at low energies, the large value of α_s leads to important non-perturbative effects which require specific computation techniques that will be discussed further below.

2.1.3 The weak interaction and electroweak unification

In the 1950s, Chen-Ning Yang and Tsung-Dao Lee theorised the violation of parity symmetry in weak interactions [8]. Until then, it was considered that all physical laws should remain identical under a flip of the spatial coordinates, which is true for the gravitational, electromagnetic and strong forces. In 1956, Chien-Shiung Wu conducted a famous experiment in which she studied the β decay of polarised cobalt-60 nuclei. The observation that electrons were primarily emitted in a direction opposite to that of the nuclear spin confirmed Yang and Lee's prediction, which awarded them the Nobel Prize a year later³.

We now know that the weak interaction currents $j_{\pm}^{\mu}$, corresponding to the $W^{\pm}$ bosons, have a vector – axial-vector form:

$$j_{\pm}^{\mu} \propto \gamma^{\mu} \frac{1}{2}(1 - \gamma^5) = \gamma^{\mu} P_L \quad (2.7)$$

where γ^5 is the product of γ -matrices $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and $P_L = \frac{1}{2}(1 - \gamma^5)$ is called the left-handed chiral projection operator (equivalently $P_R = \frac{1}{2}(1 + \gamma^5)$ is defined as the right-handed projection). Therefore, the $W^{\pm}$ bosons can only couple to left-handed particles and right-handed anti-particles, which corresponds to a maximal parity violation. The weak interaction is the only interaction that couples particles with different flavours, such as leptons with neutrinos or up-quarks with down-quarks. These flavour changing charged currents play a crucial role in B physics as will be described later. The propagator for a W boson with four-momentum q is written:

$$\frac{-i\left(g_{\mu\nu} - \frac{q_{\mu}q_{\nu}}{m_W^2}\right)}{q^2 - m_W^2}, \quad (2.8)$$

where m_W is the W boson mass. In the limit where $q^2 \ll m_W^2$, the propagator becomes $\frac{ig_{\mu\nu}}{m_W^2}$ which results in a point-like interaction. This propagator can be related to the theory proposed by Fermi in 1933 to explain β decays as four-point vector currents with the Fermi constant G_F as coupling strength. This description is still relevant today as an effective field theory for B decays.

In the 1960s, Glashow, Salam and Weinberg independently developed a theory unifying the electromagnetic and weak interactions [9–11]. Their work awarded them the Nobel Prize in

²This formula corresponds to the leading order running of α_s , at higher orders it would be more complicated.

³The choice to exclude Wu from the Nobel Prize has been criticised numerous times.

1979, six years after the discovery of weak neutral currents proving the existence of the Z^0 boson predicted by their theory. Since W bosons couple only to left-handed particles and they change particle flavour, left-handed fermions can be placed in doublets of the non-Abelian group $SU(2)$:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L.$$

The local gauge transformation of such a group can be written as:

$$\psi(x) \rightarrow \psi'(x) = e^{i\vec{\alpha}(x) \cdot \frac{\vec{\sigma}}{2}} \psi(x), \quad (2.9)$$

where $\vec{\alpha}(x)$ are three functions specifying the phase at each space-time point x and $\vec{\sigma}$ are the three Pauli matrices. The wave functions, therefore, have two components represented by the weak isospin $I_W = \frac{1}{2}$. Right-handed particles are placed in weak isospin singlets with $I_W = 0$. The local gauge invariance requirement gives rise to three bosons W_μ^i , $i = 1, 2, 3$. The first two can be combined to define the complex gauge fields $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2)$. However, the third field cannot be identified with the Z boson as the latter also couples to right-handed fermions. Therefore, a new field B^μ is introduced such that the photon and Z boson fields are expressed as a mixture of W_μ^3 and B^μ :

$$A_\mu = +\cos\theta_W B_\mu + \sin\theta_W W_\mu^3, \quad (2.10)$$

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3, \quad (2.11)$$

where θ_W is called the weak mixing (or Weinberg) angle. Fermions are given an additional charge, called *weak hypercharge* Y , under a $U(1)$ transformation gauged by B^μ with coupling constant g' . Fermionic fields therefore transform under a unified electroweak $SU(2)_L \times U(1)_Y$ transformation. The weak hypercharge Y , electromagnetic charge Q and third component of the weak isospin $I_W^3 = 0, \pm\frac{1}{2}$ are related via the Gell-Mann–Nishijima formula:

$$Y = 2(Q - I_W^3). \quad (2.12)$$

The values of these three parameters for each SM fermion are summarised in Table 2.1.

Fermion	L			R		
	Q	(I_W, I_W^3)	Y	Q	(I_W, I_W^3)	Y
$\nu_{e,\mu,\tau}$	0	$(1/2, +1/2)$	−1	−	−	−
e, μ, τ	−1	$(1/2, −1/2)$	−1	−1	$(0, 0)$	−2
u, c, t	+2/3	$(1/2, +1/2)$	+1/3	+2/3	$(0, 0)$	+4/3
d, s, b	−1/3	$(1/2, −1/2)$	+1/3	−1/3	$(0, 0)$	−2/3

TABLE 2.1: The electroweak quantum numbers of fermions. Quantum numbers of right-handed neutrinos can be deduced in the same way as for left-handed neutrinos but they are forbidden in the SM.

The electromagnetic coupling e , weak coupling g_W and $U(1)_Y$ coupling g' are also related such that:

$$e = g_W \sin\theta_W = g' \cos\theta_W. \quad (2.13)$$

The weak mixing angle's value is not predicted by the SM and needs to be measured. It is found to be $\sin^2 \theta_W \approx 0.231$. The currents associated to W_μ^3 and B_μ are respectively expressed as:

$$j_3^\mu = I_W^3 g_W \bar{f} \gamma^\mu P_L f, \quad (2.14)$$

$$j_Y^\mu = g' \frac{Y_L}{2} \bar{f} \gamma^\mu P_L f + g' \frac{Y_R}{2} \bar{f} \gamma^\mu P_R f, \quad (2.15)$$

where $Y_{L/R}$ are the hypercharges of the right/left fields. Using Equation 2.10 and defining the Z coupling as $g_Z = \frac{g_W}{\cos \theta_W}$, the Z current can then be expressed as

$$j_Z^\mu = g_Z (I_W^3 - Q \sin^2 \theta_W) \bar{f} \gamma^\mu P_L f - g_Z (Q \sin^2 \theta_W) \bar{f} \gamma^\mu P_R f \quad (2.16)$$

indicating that the Z boson couples to both right and left-handed fermions but with different strengths.

2.1.4 Electroweak symmetry breaking: the Brout-Englert-Higgs mechanism

The unification of the electromagnetic and weak interactions poses three important conceptual problems. First, because left and right-handed fermions are put in different representations of the $SU(2)_L \times U(1)_Y$ group, no fermion mass term can be written. Second, as was explained earlier, gauge invariance forbids a gauge boson to have a mass but, even though the photon is known to be massless, the W and Z bosons have masses $\mathcal{O}(10^2)$ GeV. Finally, the electromagnetic and weak interactions are known to be very different. The former is a long-range force, whereas the latter is mediated over a very short range (due to the associated boson masses). The Brout-Englert-Higgs (BEH) mechanism solves these issues through the process of *spontaneous symmetry breaking* [12–14] which awarded Peter Higgs and François Englert the 2013 Nobel Prize⁴. The BEH mechanism postulates the existence of an additional field, the Higgs field H . This field transforms as a doublet of $SU(2)$ with hypercharge $Y = 1$. The Higgs Lagrangian is written as:

$$\mathcal{L}_H = (D_\mu H)^\dagger (D_\mu H) - V(H), \quad (2.17)$$

with the Higgs potential V taking the form:

$$V(H) = -\mu^2 H^\dagger H + \lambda (H^\dagger H)^2, \quad (2.18)$$

where λ and μ are two real positive free parameters. The form of this potential in the Higgs field complex plane is illustrated in Figure 2.4. The potential gives rise to a non-trivial structure of the vacuum – the state of minimum energy in the universe – corresponding to the state minimising the potential. The classical vanishing vacuum $\langle H \rangle = 0$ yields a non-stable minimum and therefore the Higgs potential is minimised when

$$\sqrt{H^\dagger H} \equiv \frac{v}{\sqrt{2}} = \sqrt{\frac{\mu^2}{2\lambda}}, \quad (2.19)$$

which yields a non-zero *vacuum expectation value* (VEV) $\langle H \rangle = \frac{v}{\sqrt{2}}$ with v some real parameter. There are infinitely many degenerate states which can minimise the potential and any of them breaks the $SU(2) \times U(1)$ gauge symmetry. During the electroweak phase transition, the early universe cooled down enough to give to the Higgs potential the form defined in Equation 2.18, which leads to the spontaneous symmetry breaking of the electroweak interaction down to the remaining $U(1)_{\text{EM}}$ of QED. Without loss of generality, the ground state can be expressed in

⁴The spontaneous symmetry breaking mechanism was in fact independently predicted by several people but only two of them were awarded the prize and only one of them gave his name to the famous boson.

$SU(2)$ doublet components as:

$$H_G = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \quad (2.20)$$

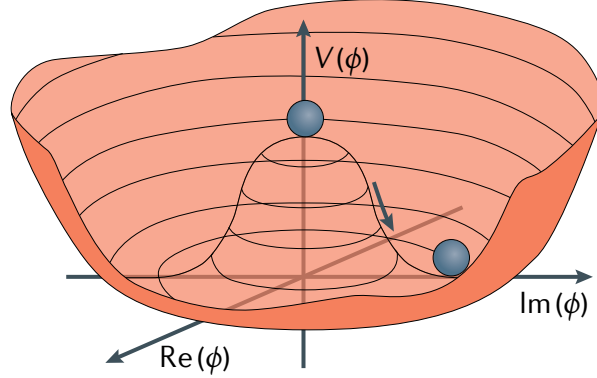


FIGURE 2.4: Illustration of the Higgs potential in the complex plane. The degeneracy of the vacuum states is made apparent as a circle around the origin. If the reader is in a festive mood, the Higgs potential can be referred to as the Mexican hat or wine bottle potential.

The general form of the Higgs field is written as:

$$H(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} \quad (2.21)$$

which can be expanded around the ground state in terms of fields with vanishing VEV as

$$H(x) = \begin{pmatrix} \phi^+(x) \\ \frac{1}{\sqrt{2}}[v + h(x) + i\chi(x)] \end{pmatrix}, \quad \phi^-(x) = [\phi^+]^\dagger. \quad (2.22)$$

The physical field $h(x)$ corresponds to the well-known Higgs boson field. The other three components are unphysical degrees of freedom. According to Goldstone's theorem, for a spontaneously broken global symmetry, each one of these fields would give rise to a physical massless boson⁵. However, since the BEH mechanism breaks a gauge symmetry, the (would-be) Goldstone fields are unphysical. By performing an appropriate gauge rotation (under which the symmetry is by definition invariant), the three additional degrees of freedom are "eaten" by the gauge bosons, W^+ , W^- and Z^0 , providing them with longitudinal polarisation components. It is important to note that the SM Higgs field is the simplest realisation of the BEH mechanism. For instance, a Higgs sector consisting of two $SU(2)$ doublets would yield another four physical Higgs fields. By expanding the covariant derivative term in the Higgs Lagrangian defined in Equation 2.17,

$$D_\mu = \partial_\mu - ig_W I_W^a W_\mu^a + ig' Y_W B_\mu, \quad (2.23)$$

⁵If the symmetry breaking is not exact, the bosons can be massive but light as is the case for the spontaneously broken chiral symmetry giving rise to the π which, with masses around 135-140 MeV, are much lighter than other mesons.

two mass terms appear for the weak interaction gauge bosons:

$$\frac{1}{4}(g_W v)^2 W_\mu^+ W^{-\mu} + \frac{1}{8}(g_W^2 + g'^2)v^2 Z_\mu Z^\mu. \quad (2.24)$$

Therefore the W and Z boson masses are

$$m_W = \frac{vg_W}{2}, \quad m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2} \quad (2.25)$$

and the photon remains massless. The values of m_W and m_Z having been measured very precisely, one can infer the value of the VEV to be $\frac{v}{\sqrt{2}} \sim 174$ GeV. Finally, the Higgs boson self-interactions read

$$\frac{m_h^2}{2}h^2 + \frac{m_h^2}{2v}h^3 + \frac{m_h^2}{8v^2}h^4. \quad (2.26)$$

The first term gives the Higgs mass m_h and the last two allow for triple and quartic Higgs vertices.

Fermion masses

The fermion interaction terms with the Higgs field are given by the Yukawa Lagrangian

$$\mathcal{L}_Y = -\bar{Q}_L Y_u H^c u_R - \bar{Q}_L Y_d H d_R - \bar{L}_L Y_\ell H \ell_R + \text{h.c.}, \quad (2.27)$$

where Q_L and L_L are respectively the three-vectors of quark and lepton $SU(2)$ doublets as defined above; u_R , d_R and ℓ_R are respectively the three-vectors of up-quark, down-quark and lepton singlets; $Y_{u,d,\ell}$ are the 3×3 complex Yukawa matrices; $H^c \equiv i\sigma_2 H^*$ and h.c. denotes the Hermitian conjugate. The Yukawa matrices can be diagonalised, which rewrites the Lagrangian as

$$\mathcal{L}_Y = -\bar{Q}_L U_u Y_u^{\text{diag}} H^c W_u^\dagger u_R - \bar{Q}_L U_d Y_d^{\text{diag}} H W_d^\dagger d_R - \bar{L}_L U_\ell Y_\ell H W_\ell^\dagger \ell_R + \text{h.c.}, \quad (2.28)$$

where $U_{u,d,\ell}$ and $W_{u,d,\ell}$ are unitary matrices. By rotating the weak eigenstates to mass eigenstates, the fermion fields are redefined as

$$Q_L \rightarrow U_d Q_L, \quad d_R \rightarrow W_d d_R, \quad u_R \rightarrow W_u u_R, \quad L_L \rightarrow U_\ell L_L, \quad \ell_R \rightarrow W_\ell \ell_R. \quad (2.29)$$

The Yukawa Lagrangian, therefore, becomes⁶

$$\mathcal{L}_Y = -\bar{Q}_L (U_d^\dagger U_u) Y_u^{\text{diag}} H^c u_R - \bar{Q}_L Y_d^{\text{diag}} H d_R - \bar{L}_L Y_\ell H \ell_R + \text{h.c.} \quad (2.30)$$

The second and third terms are now diagonal but not the first. The 3×3 unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix $V_{\text{CKM}} = U_d^\dagger U_u$ appearing in this term is the source of all flavour-violating interactions in the SM. The CKM matrix was introduced by Makoto Kobayashi and Toshihide Maskawa [15] as an extension of the two-dimensional matrix introduced by Nicola Cabibbo [16] in order to explain the experimentally observed CP -violation effects⁷. The properties of the CKM matrix will be discussed in the next section. When inserting the Higgs VEV in place of the Higgs field in the Yukawa Lagrangian, the fermion mass matrices can be expressed

⁶This is a choice of basis. The fields could also be transformed such that the CKM matrix appears in the down quark term.

⁷Which lead the two Japanese to a Nobel Prize but, in yet another Nobel controversy, not Cabibbo.

as

$$M_u = \frac{v}{\sqrt{2}} Y_u^{\text{diag}} = \text{diag}(m_u, m_c, m_t), \quad (2.31)$$

$$M_d = \frac{v}{\sqrt{2}} Y_d^{\text{diag}} = \text{diag}(m_d, m_s, m_b), \quad (2.32)$$

$$M_\ell = \frac{v}{\sqrt{2}} Y_\ell^{\text{diag}} = \text{diag}(m_e, m_\mu, m_\tau), \quad (2.33)$$

where m_f denotes the mass of fermion f . Therefore, similarly to gauge bosons, the interaction strength of each fermion species with the Higgs boson depends on the mass of that fermion. Neutrino masses could also be included in the Yukawa Lagrangian in the same way as quarks and charged leptons if right-handed neutrino fields are added to the Lagrangian. The Yukawa term would therefore be of the form

$$\mathcal{L}_Y \supset \bar{\nu}_L (U_\ell^\dagger U_\nu) Y_\nu^{\text{diag}} H^c \nu_R \quad (2.34)$$

where Y_ν^{diag} is the neutrino mass matrix and $V_{\text{PMNS}} = U_\ell^\dagger U_\nu$ is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. However, neutrinos with right-handed chirality have never been observed and this term is therefore forbidden in the SM Lagrangian. This type of *sterile neutrinos* has been theorised as extremely heavy particles compared to their left-handed counterparts. Another alternative would be to add to the Lagrangian a Majorana mass term involving only left-handed neutrinos, which would therefore be their own anti-particles.

2.2 The CKM matrix and flavour physics

We presented in the previous section the SM and we showed how the CKM matrix enters its Lagrangian. In this section, we will focus on the implications and properties of the CKM matrix. First, it can be noticed that diagonalising the Yukawa mass matrices breaks up the quark $SU(2)_L$ doublets since the up-quark doublets transform as $u_L \rightarrow V_{\text{CKM}}^\dagger u_L$ independently of the down-quark doublets. In the weak interaction Lagrangian, the components of each $SU(2)_L$ doublet mix up in the term

$$\mathcal{L}_{\text{weak}} \supset \frac{g_W}{\sqrt{2}} W_\mu^- \bar{d}_L \gamma^\mu V_{\text{CKM}}^\dagger u_L + \text{h.c.} \quad (2.35)$$

The CKM matrix, therefore, allows mixing between different generations of quarks. In its most general representation, the CKM matrix can be written as

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (2.36)$$

A complex 3×3 matrix has a total of $2 \times 3^2 = 18$ degrees of freedom. Since the CKM matrix is unitary and since it is invariant under certain phase transformations, the number of degrees of freedom can be reduced to four, three real angles and one complex phase. The CKM matrix can be parametrised using the so-called standard parametrisation:

$$\begin{aligned} V_{\text{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \end{aligned} \quad (2.37)$$

where $s_{ij} = \sin \theta_{ij}$, $c_{ij} = \cos \theta_{ij}$, θ_{ij} being the three real angles and δ the complex phase which is responsible for all CP violation effects in the SM. The structure of the CKM matrix is highly hierarchical as $s_{13} \ll s_{23} \ll s_{12} \ll 1$. Based on this observation, it can also be parametrised according to the Wolfenstein parametrisation:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad (2.38)$$

where

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}} \quad (2.39)$$

$$s_{23} = A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right| \quad (2.40)$$

$$s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) \quad (2.41)$$

and in addition $\bar{\rho}$ and $\bar{\eta}$ are defined such that $\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$. The magnitude of each CKM matrix element can be measured from decays involving the corresponding quark transition. A global fit combining information from all measurements constrained by the matrix unitarity are regularly performed by the CKMfitter [17] and UTfit [18] groups. The latest global fit made by CKMfitter gives⁸

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.974358_{-0.000054}^{+0.000049} & 0.22498_{-0.00022}^{+0.00023} & 0.003730_{-0.000048}^{+0.000044} \\ 0.22484_{-0.00021}^{+0.00023} & 0.973509_{-0.00059}^{+0.00054} & 0.04160_{-0.00058}^{+0.00020} \\ 0.008573_{-0.000158}^{+0.000046} & 0.04088_{-0.00066}^{+0.00020} & 0.9991248_{-0.000074}^{+0.0000268} \end{pmatrix}. \quad (2.42)$$

The unitarity of V_{CKM} yields the relations

$$\sum_k V_{ik}V_{jk}^* = \delta_{ij}, \quad \sum_k V_{ki}V_{kj}^* = \delta_{ij}, \quad (2.43)$$

which can be used to define a triangle containing the matrix information. For example, since

$$\begin{aligned} V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* &= 0 \\ \Rightarrow 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} &= 0, \end{aligned} \quad (2.44)$$

the unitarity triangle can be defined with vertices $(0, 0)$, $(1, 0)$ and $(\bar{\rho}, \bar{\eta})$ and angles

$$\alpha = \arg \left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right) \quad (2.45)$$

$$\beta = \arg \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right) \quad (2.46)$$

$$\gamma = \arg \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right). \quad (2.47)$$

⁸Note that these are the nominal values quoted by CKMfitter. The collaboration also quotes values obtained from fits which exclude part of the information used in the global fit. It is for instance interesting to obtain values of CKM matrix elements by excluding inputs from measurements of this specific parameter. Values obtained this way are only constrained by orthogonal measurements.

The unitarity triangle in the $(\bar{\rho}, \bar{\eta})$ plane is illustrated in Figure 2.5. The area of the triangle is equal to half of the Jarlskog invariant J such that $\text{Im}(V_{ij}V_{kl}V_{il}^*V_{kj}^*) = J \sum_{m,n} \epsilon_{ikm}\epsilon_{jln}$ where ϵ_{ijk} is the Levi-Civita symbol. The mixing of quark generations allows a wide variety of decays such

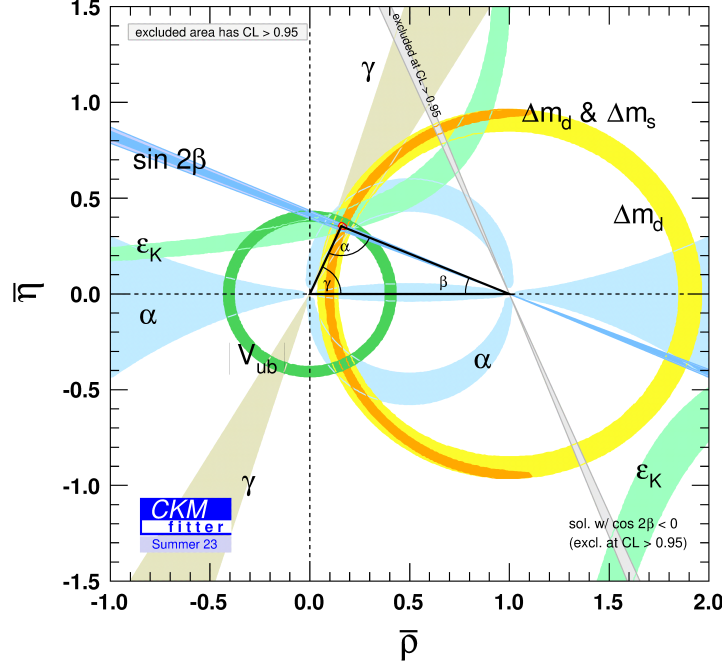


FIGURE 2.5: Illustration of the CKM unitarity triangle as obtained from a global fit performed by CKMfitter. The coloured bands show constraints of CKM parameters coming from various flavour measurements. Taken from Ref. [17]

as semi-leptonic decays which are the main subject of this thesis as well as so-called *box diagrams* and *penguin diagrams*⁹. The latter two are illustrated in Figure 2.6. We quickly outline how every CKM parameter, as illustrated in Figure 2.5, can be measured:

- V_{ub} : measured through semi-leptonic transitions $b \rightarrow u\ell\nu$; this parameter is the main subject of this thesis and will therefore be discussed in detail in subsequent sections and chapters
- α : time-dependent CP -violation measurements in $b \rightarrow u\bar{u}d$ decays, such as $B \rightarrow \pi\pi$, $B \rightarrow \rho\rho$ and $B \rightarrow \pi\rho$
- γ : time-dependent CP -violation measurements in $B \rightarrow DK$ and $B \rightarrow D\pi$ decays
- $\sin 2\beta$: measurements of the interference of mixing and decay in neutral B mesons into CP eigenstates; the golden channel for measuring the β angle is $B \rightarrow J/\psi K_S$
- ϵ_K : CP -violation measurements in K mixing; the exact expression of ϵ_K is not shown here in its entirety but it is dominated by the term $\text{Im}[(V_{ts}V_{td}^*)^2]$
- $\Delta m_{d/s}$: measurements of $B_{d,s} - \bar{B}_{d,s}$ mixing

⁹If the reader is not familiar with the origin of this name, they are invited to ask Professor John Ellis to tell the story himself.

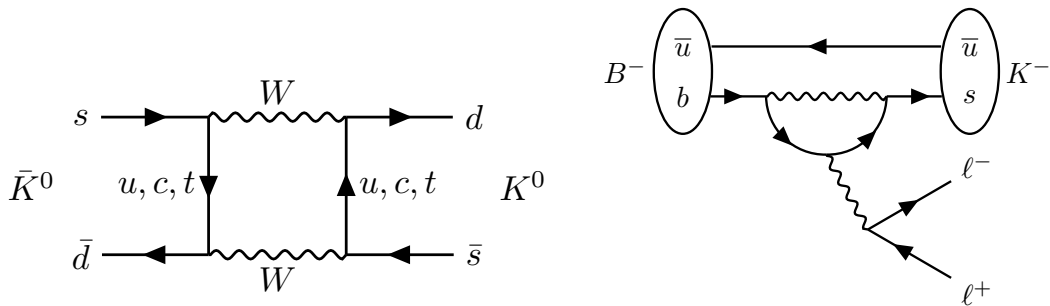


FIGURE 2.6: Examples of Feynman diagrams allowed by quark generation transitions. The left diagram represents neutral kaon mixing from a box diagram and the right diagram represents a $B \rightarrow K\ell\ell$ penguin diagram.

2.3 Semileptonic B decays

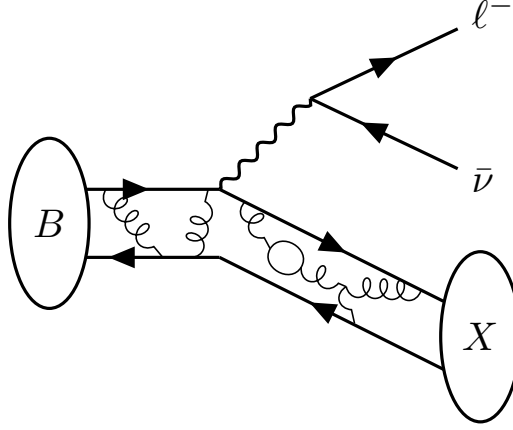
Semileptonic decays are ideally suited to measure CKM matrix elements [19, 20]. They are tree decays, are therefore not suppressed by the GIM mechanism¹⁰ and they involve a single CKM element. Strange (K), charm (D) and beauty (B) mesons can all decay semileptonically. However, this thesis focuses on B meson decays, which allow the extraction of two CKM parameters: $|V_{cb}|$ through $b \rightarrow c$ transitions and $|V_{ub}|$ through $b \rightarrow u$ transitions. Semileptonic B decays will, in general, be denoted as $B \rightarrow X\ell\nu$, where X represents a hadronic system. Electromagnetic charges will be omitted and charge conjugation implied unless explicitly stated. An example of a Feynman diagram of such a decay is shown in Figure 2.7. The subscript u (c) added to X will indicate a hadronic system containing a u (c) quark. X can represent either a specific resonance such as π for a $b \rightarrow u$ transition or D^* for a $b \rightarrow c$ transition or a non-resonant hadronic system. Two types of $B \rightarrow X\ell\nu$ measurements can therefore be performed: exclusive, where the hadronic system is reconstructed as a specific meson or inclusive, in which no channel-specific requirements are placed on the reconstruction of X , which effectively corresponds to summing over all possible final states. This thesis discusses the measurement of $|V_{ub}|$ via inclusive $B \rightarrow X_u\ell\nu$ decays. This type of decays is modelled as a combination of resonant and non-resonant modes, as will be explained later and therefore both types are described in this section. Since V_{cb} is about an order of magnitude larger than V_{ub} , $B \rightarrow X_c\ell\nu$ decays represent the largest background to $B \rightarrow X_u\ell\nu$ decays, so they will also be discussed here. We will first make a few general remarks regarding the theoretical description of semileptonic decays before focusing on each different type in dedicated sections.

In the following, we will denote p_B , p_X , p_ℓ and p_ν the four-momenta of the B meson, the X system, the lepton and the neutrino, respectively. They are related such that

$$p_B = p_X + p_\ell + p_\nu, \quad p_B^2 = m_B^2, \quad p_X^2 = M_X^2, \quad p_\ell^2 = m_\ell^2, \quad p_\nu^2 = 0, \quad (2.48)$$

where m_B , M_X and m_ℓ are the B , X and lepton masses, respectively. The neutrino is considered to be massless. When the lepton is an electron or a muon, it can be approximated to be massless as well. In that case, the lepton energy E_ℓ is equal to its three-momentum magnitude $|\vec{p}_\ell|$. When taken in the B meson rest frame, the lepton energy is labelled E_ℓ^B . Decays to tau leptons will not be discussed in detail here as only light leptons were considered in the measurement. The

¹⁰The Glashow-Iliopoulos-Maiani mechanism [21] strongly reduces the contribution of loops involving light quarks, which in turn reduces the overall rate decays involving such loops.

FIGURE 2.7: An example of a semileptonic $B \rightarrow X \ell \nu$ Feynman diagram.

four-momentum transfer squared or leptonic system invariant mass squared q^2 is defined as:

$$q^2 = (p_B - p_X)^2 = (p_\ell + p_\nu)^2, \quad m_\ell^2 \leq q^2 \leq (m_B - M_X)^2. \quad (2.49)$$

As explained in Section 2.1.3, at energies much below 80 GeV, the W boson can be integrated out, simplifying the associated propagator. Therefore, in semileptonic B decays, the effective weak Hamiltonian can be expressed as

$$\mathcal{H}_{\text{eff}} = \frac{4G_F}{\sqrt{2}} V_{qb} (\bar{q}_L \gamma_\mu b_L) (\bar{\ell}_L \gamma_\mu \nu_L^\ell), \quad (2.50)$$

where the last two terms represent the hadronic and leptonic currents, respectively. The hadronic current is non-trivial to compute since, at the energy scale of a B meson decay, α_s is $\mathcal{O}(1)$ and perturbative QCD calculations cannot be applied. For specific resonant decay modes, using the fact that the hadronic current must be Lorentz covariant, it can be expressed as a function of the Lorentz vectors and scalars entering the hadronic matrix element and unknown functions called *form factors* which encode the non-perturbative dynamics of the transition. For example, for a B transition to a pseudo-scalar P (such as D , π , η or η') with four-momentum p_P , the hadronic current can be expressed as

$$H_\mu = \langle P | \bar{q}_L \gamma_\mu b_L | B \rangle = f_+(q^2) (p_B + p_P)_\mu + f_-(q^2) (p_B - p_P)_\mu, \quad (2.51)$$

where f_+ and f_- are two form factors. Details about each resonant decay type are given below. The hadronic current in inclusive semileptonic decays requires specific techniques and models to be computed, as will also be outlined below. The leptonic current on the other hand, is easier to compute as it doesn't involve any QCD effects.

After having briefly introduced semileptonic decays and their kinematics, we will now discuss the main theoretical tools used to describe them. We will then present in more detail the theoretical frameworks describing exclusive and inclusive decays.

2.3.1 Theoretical methods

Heavy quark methods

Since the 1980s, an extensive effort has been dedicated to methods relying on the fact that the b quark mass $m_b \sim 4.2$ GeV¹¹ [24] is much larger than the QCD scale $\Lambda_{\text{QCD}} = \mathcal{O}(100)$ MeV which allows one to expand the QCD matrix elements in powers of Λ_{QCD}/m_b . This principle is particularly suited to B decays but is also applicable to D decays even though the charm quark is lighter than the beauty quark.

Based on this idea, an effective field theory, known as the *heavy quark effective theory* (HQET) [25] can be used to simplify the QCD Lagrangian given in equation 2.4. For a heavy field Q with mass m_Q , its kinematic part can be written as

$$\mathcal{L}_{\text{QCD}} = \bar{Q} i \not{D} Q - m_Q \bar{Q} Q, \quad (2.52)$$

where the covariant derivative is $D_\mu = \partial_\mu + ig_s A_\mu^a \frac{\lambda^a}{2}$. The heavy quark momentum p_Q can be decomposed as

$$p_Q = m_Q v + k, \quad (2.53)$$

where v is the four-velocity of the quark and $k = \mathcal{O}(\Lambda_{\text{QCD}})$ is its residual momentum, which can be changed by interactions of the quark with soft partons in the hadron, also referred to as *light degrees of freedom*. The heavy field Q is decomposed as

$$h_v(x) = e^{im_Q v \cdot x} \frac{1 + \not{v}}{2} Q(x), \quad (2.54)$$

$$H_v(x) = e^{im_Q v \cdot x} \frac{1 - \not{v}}{2} Q(x) \quad (2.55)$$

and thus

$$Q(x) = e^{-im_Q v \cdot x} (h_v(x) + H_v(x)). \quad (2.56)$$

In addition, we define

$$D_T^\mu = D^\mu - v^\mu v \cdot D \Rightarrow \not{D} = \not{v} v \cdot D + \not{D}_T. \quad (2.57)$$

By plugging in equations 2.56 and 2.57 in 2.52, the QCD Lagrangian can be rewritten

$$\mathcal{L}_{\text{HQET}} = \bar{h}_v i v \cdot D h_v - \bar{H}_v (i v \cdot D + 2m_Q) H_v + \bar{h}_v i \not{D}_T H_v + \bar{H}_v i \not{D}_T h_v. \quad (2.58)$$

Since the field H_v is given a mass $2m_Q$ in this Lagrangian, in the limit where m_Q is taken as infinite, H_v can be integrated out in a similar way to the W boson in the weak effective Hamiltonian. Therefore, at leading order in the HQET

$$\mathcal{L}_{\text{HQET}} = \bar{h}_v i v \cdot D h_v. \quad (2.59)$$

In HQET, the heavy-quark field can be seen as a static source colour field surrounded by a brown muck of soft partons which induce small corrections. Since the HQET Lagrangian at leading order doesn't depend on any mass term, a new symmetry is exhibited, the heavy-flavour symmetry, in which all heavy-quark fields are equivalent, implying that the b and c quarks have exactly the same dynamics. Furthermore, as no γ -matrices appear in $v \cdot D$, the leading order term of the HQET Lagrangian is purely scalar. Thus, the heavy-quark spin decouples which exhibits

¹¹In the $\overline{\text{MS}}$ renormalisation scheme. Renormalisation schemes are not discussed in this thesis. They are defined as sets of prescriptions used to fix the counterterms in renormalisation. The $\overline{\text{MS}}$ scheme is one of the most popular schemes, especially for QCD. It was introduced by Gerard 't Hooft and Steven Weinberg [22, 23]. The b quark mass and other parameters are defined in specific renormalisation schemes but for the sake of simplicity, the scheme will not be made explicit every time a parameter is used.

another symmetry, the heavy-quark spin symmetry. These symmetries were first formulated in a famous paper by Isgur and Wise [26]. Additional power suppressed terms of $\mathcal{L}_{\text{HQET}}$ break these symmetries. HQET principles are particularly useful when computing matrix elements for resonant semileptonic decays.

The *heavy quark expansion* (HQE) is a related method, developed mainly for inclusive decays [27, 28]. Starting from the optical theorem, the semileptonic decay rate of a B (or D to a lesser extent) can be related to the forward matrix element of a scattering amplitude:

$$\Gamma = \text{Im} \int d^4x \langle B(p_b) | T[\mathcal{H}_{\text{eff}}(x)\mathcal{H}_{\text{eff}}(0)] | B(p_B) \rangle, \quad (2.60)$$

where T represents the time-ordered product which can be written as an *operator product expansion* (OPE) such that

$$\int d^4x T[\mathcal{H}_{\text{eff}}(x)\mathcal{H}_{\text{eff}}(0)] = \sum_{n,i} \frac{1}{m_b^n} C_{n,i} \mathcal{O}_{n+3,i}, \quad (2.61)$$

where $\mathcal{O}_{n+3,i}$ are a set of operators which encode the non-perturbative QCD effects and the $C_{n,i}$ are Wilson coefficients which are calculable in perturbation theory at different orders in α_s . At $n = 0$, the HQE decay rate represents the decay rate of a free b quark. At $n = 1$, all operators can be related to $n \geq 2$ operators and thus the contribution of order $1/m_b$ can be eliminated. At $n \geq 2$, a plethora of non-perturbative operators contribute. These are known at the first few orders in $1/m_b$ and α_s for inclusive $B \rightarrow X_c \ell \nu$ decays as summarised in Table 2.2. Because of the non-perturbative character of the OPE, experimental input is crucial to obtain precise values of its parameters.

Lattice QCD

Lattice QCD (LQCD) is a framework developed to perform non-perturbative quantum field theory calculations [29]. The idea of LQCD resides in formulating correlation functions of QCD operators on a discrete Euclidean space-time. This formulation of QCD on a grid can then be simulated using computer algorithms. To obtain precise values of QCD observables via this method, various LQCD parameters can be tuned, such as the grid size and spacing, as well as the quark bare masses. However, when reaching high precision levels, LQCD becomes extremely expensive in terms of computing resources and its success has been made possible by progress in computing power. In the context of semileptonic B decays, LQCD proves particularly useful for the calculation of form factors for resonant modes, as will be discussed later in this section.

2.3.2 Exclusive semileptonic $B \rightarrow X_c \ell \nu$ decays

We will now present the theoretical details of the most common $B \rightarrow X_c \ell \nu$ resonant channels. The masses, widths and strong decays of the D , D^* and D^{**} mesons are summarised in Figure 2.8. For the $B \rightarrow X_c \ell \nu$ resonant decays, it is useful to define the recoil parameter w such that

$$w = \frac{m_B^2 + m_D^2 - q^2}{2m_B m_D}, \quad (2.62)$$

where m_D is the mass of the charm meson (D , D^* or D^{**}).

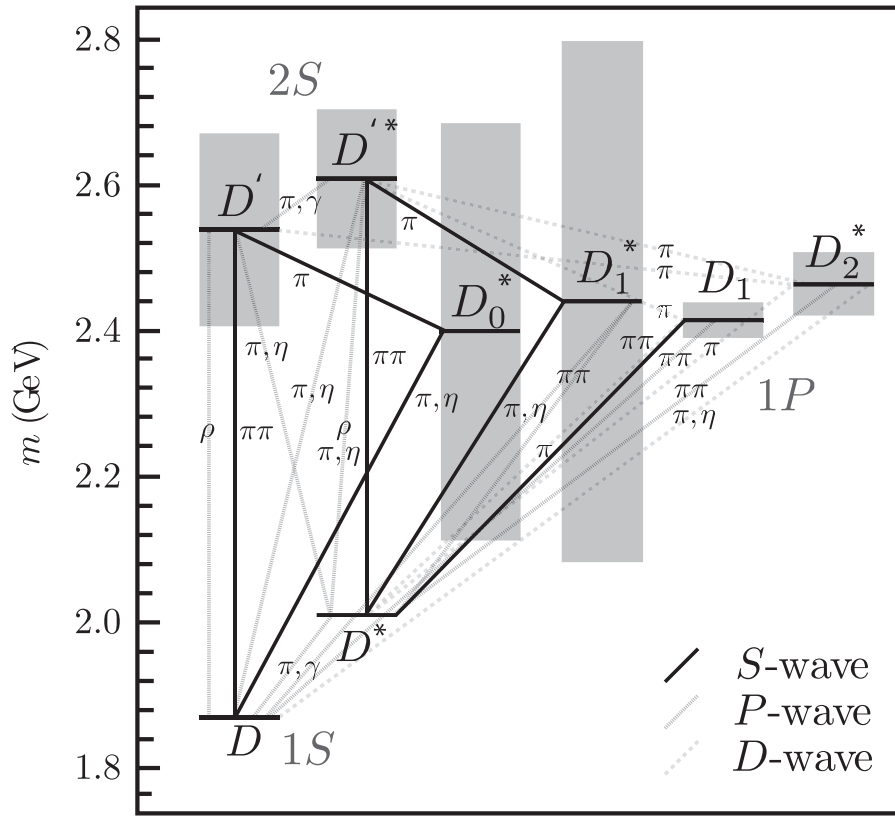


FIGURE 2.8: Strong decays of the radially excited charm mesons. The four $1P$ states are referred to as D^{**} . D and D^* are the two $1S$ states. The style of the decay lines indicates the orbital angular momentum of the partial wave. The grey bands indicate the measured widths of each excited state. The y -axis represents the mass of the meson.

$B \rightarrow D^{(*)}\ell\nu$ decays

In their most general form, the $B \rightarrow D^{(*)}$ transition matrix elements can be written as functions of the B and $D^{(*)}$ meson velocities, v and v' respectively, such that [20, 30]

$$\frac{\langle D(v') | \bar{c}\gamma_\mu b | \bar{B}(v) \rangle}{\sqrt{m_B m_D}} = h_+(v + v')_\mu + h_-(v - v')_\mu, \quad (2.63)$$

$$\frac{\langle D^*(v', \epsilon) | \bar{c}\gamma_\mu b | \bar{B}(v) \rangle}{\sqrt{m_B m_{D^*}}} = i h_V \epsilon_{\mu\nu\alpha\beta} \epsilon^{*\nu} v'^\alpha v^\beta, \quad (2.64)$$

$$\frac{\langle D^*(v', \epsilon) | \bar{c}\gamma_\mu \gamma_5 b | \bar{B}(v) \rangle}{\sqrt{m_B m_{D^*}}} = h_{A_1}(w + 1)\epsilon^{*\mu} - (h_{A_2}v_\mu + h_{A_3}v'_\mu)\epsilon^* \cdot v, \quad (2.65)$$

where ϵ is the D^* polarisation, h_+ and h_- are two $B \rightarrow D$ form factors and h_V , h_{A_1} , h_{A_2} and h_{A_3} are four $B \rightarrow D^*$ form factors. All six form factors are functions of w . For decays to pseudoscalar mesons such as D , only the vector current contributes to the matrix element. There exist alternative parametrisations of $B \rightarrow D^{(*)}$ form factors which can be related to each other. For instance, the $B \rightarrow D$ matrix element is often expressed as a function of q^2 in which case it depends on two form factors labelled f_+ and f_0 and the $B \rightarrow D^*$ matrix elements can be expressed in the helicity basis in which case they depend on three form factors, f , g and $\mathcal{F}_1$. When relevant, we will collectively refer to these five functions as f_i . The differential decay

rates as a function of w read

$$\frac{d\Gamma(B \rightarrow D\ell\nu)}{dw} = \frac{G_F^2}{48\pi^3} |V_{cb}|^2 (m_B + m_D)^2 m_D^3 (w^2 - 1)^{3/2} \eta_{EW}^2 |\mathcal{G}(w)|^2, \quad (2.66)$$

$$\frac{d\Gamma(B \rightarrow D^*\ell\nu)}{dw} = \frac{G_F^2 m_B^5}{48\pi^3} |V_{cb}|^2 (w^2 - 1)^{1/2} P(w) \eta_{EW}^2 |\mathcal{F}(w)|^2, \quad (2.67)$$

where $\mathcal{G}$ and $\mathcal{F}$ are expressed in terms of w , m_{D^*} , m_B and the form factors defined in Equations 2.63 to 2.65. P is a phase-space factor expressed in terms of w , m_{D^*} and m_B . The form factors are non-perturbative functions and there exist various ways to compute them. A common way to derive the $B \rightarrow D^{(*)}\ell\nu$ form factors is to use the model-independent parametrisation introduced by Boyd, Grinstein and Lebed (BGL)¹² [30] in which the five functions f_i are expressed in an expansion of the form

$$f_i(z) = \frac{1}{P_i(z)\phi_i(z)} \sum_{n=0}^{\infty} a_{i,n} z^n \quad (2.68)$$

where z corresponds to a mapping of w such that

$$z = \frac{\sqrt{w+1} - \sqrt{2}}{\sqrt{w+1} + \sqrt{2}} \quad (2.69)$$

and P_i and ϕ_i are the so-called Blaschke factors and outer functions, $a_{i,n}$ are numerical coefficients which can be determined from fits to data and/or LQCD constraints. See for example Refs. [32] and [33] for recent determinations of these coefficients using Belle and Belle II data. Each expansion in Equation 2.68 is truncated at a fixed order, usually between order two and four. Higher truncation orders ensure a better convergence of the series but a large number of free parameters can complicate fits to data. The normalisation of the form factors *i.e.*, their value at zero recoil $w = 0$, needs to be determined separately using HQET or LQCD. The BGL parametrisation yields a constraint on the form factor parameters such that

$$\sum_{n=0}^{\infty} a_{i,n}^2 < 1. \quad (2.70)$$

Because of this additional constraint and because of unknown pole terms introduced by the BGL definition of the Blaschke factors and outer functions, the BCL parametrisation presented in Section 2.3.3 is being preferred for current studies of $B \rightarrow D\ell\nu$ decays [34]. The form factors can also be derived using the HQET directly. In this case, the hadronic matrix elements can be expanded in powers of Λ_{QCD}/m_b , Λ_{QCD}/m_c and α_s and the form factors are expressed in terms of Isgur-Wise functions [35]. The expansion introduced by Bernlochner, Ligeti, Papucci, Robinson, Xiong and Prim (BLPRXP) [36] adds a supplemental power counting based on a chiral symmetry breaking parameter. The total set of Isgur-Wise functions needed up to second order in HQET is reduced and constrained by this additional expansion. The BLPRXP framework is particularly useful as it allows a simultaneous fit of $B \rightarrow D\ell\nu$ and $B \rightarrow D^*\ell\nu$ data.

$B \rightarrow D^{**}\ell\nu$ decays

While the D and D^* mesons have orbital angular momentum $L = 0$, there exist four orbitally excited states with $L = 1$, collectively referred to as D^{**} . Two of them, D_0^* with total angular momentum $J = 0$ and D_1^* (or D_1') with $J = 1$, are relatively broad resonances, whereas the other two, D_1 with $J = 1$ and D_2^* with $J = 2$, have much narrower widths. There also exist two states

¹²The parametrisation developed by Caprini, Lellouch, and Neubert [31] (CLN) uses a similar method together with relations in the HQE. This method will not be discussed here as it is considered to have been superseded.

with $L = 2$ but the semileptonic decays to these states are not very well understood and they are therefore not discussed here. The masses, widths and decays of radially excited charm mesons are summarised in Figure 2.8. Similarly to the $B \rightarrow D^{(*)}$ case, the $B \rightarrow D^{**}\ell\nu$ matrix elements are expressed in terms of form factors which can be expanded in HQET as functions of Isgur-Wise functions [37]. The total number of Isgur-Wise functions increases significantly beyond the leading order in HQET, each one being a function of the recoil parameter w . Because the resulting form factor expressions cannot be constrained by the available data, approximations have to be made. For instance, the Isgur-Wise functions can be expanded in powers of w or treated as constant parameters rather than functions. These simplifications allow one to express the broad (narrow) $B \rightarrow D^{**}$ form factors in terms of three (four) parameters, which were determined by fits to Belle and BaBar data in Ref. [37]. This model is referred to as the Bernlochner, Ligeti, Robinson (BLR) model. In the measurement presented in this manuscript, the parameters defined by approximation C of the BLR model are used to simulate $B \rightarrow D^{**}\ell\nu$ decays.

2.3.3 Exclusive semileptonic $B \rightarrow X_u\ell\nu$ decays

$B \rightarrow \pi\ell\nu$ decays

Semileptonic B decays to a pion allow nowadays one of the most precise determinations of $|V_{ub}|$. Denoting the pion momentum p and the leptonic system momentum q , the $B \rightarrow \pi$ matrix element can be written as [38]

$$\langle \pi | \bar{u}\gamma_\mu b | B \rangle = \left(2p_\mu + q_\mu - q_\mu \frac{m_B^2 - m_\pi^2}{q^2} \right) f_+(q^2) + q_\mu \frac{m_B^2 - m_\pi^2}{q^2} f_0(q^2), \quad (2.71)$$

where f_+ and f_0 are the vector and scalar form factors, respectively. For light leptons, the f_0 contribution can be neglected and the differential decay rate is therefore expressed as

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{192\pi^3 m_B^3} \lambda^{3/2}(q^2) |f_+(q^2)|^2, \quad (2.72)$$

where $\lambda(q^2) = (m_B^2 + m_\pi^2 - q^2)^2 - 4m_B^2 m_\pi^2$. Using the Bourrely, Caprini, Lellouch (BCL) parametrisation, the two form factors can be expanded in a similar way as the $B \rightarrow D^{(*)}$ form factors as shown in Equation 2.68. With respective truncation orders N_+ and N_0 , the two form factors are therefore expressed as

$$f_+(q^2) = \frac{1}{1 - q^2/m_{B^*}^2} \sum_{n=0}^{N_+-1} a_n^+ \left(z^n - (-1)^{n-N_+} \frac{n}{N_+} z^{N_+} \right), \quad (2.73)$$

$$f_0(q^2) = \sum_{n=0}^{N_0-1} a_n^0 z^n, \quad (2.74)$$

where m_{B^*} represents the mass of the B^* pole in the expansion and a_n are non-perturbative numerical coefficients. The latter can be determined by fits to data or LQCD predictions. The most recent average of these coefficients was calculated by the Flavour Lattice Averaging Group (FLAG) in an expansion $N_+ = N_0 = 3$ [39]. This results in five coefficients, the parameter a_2^0 being fixed by the $f_+(q^2 = 0) = f_0(q^2 = 0)$ constraint.

$B \rightarrow \rho\ell\nu$ and $B \rightarrow \omega\ell\nu$ decays

$|V_{ub}|$ can also be measured via decays to vector mesons, such as ρ and ω . In that case, both the axial and vector currents contribute to the matrix elements such that, for vector V with

momentum p_V , mass m_V and polarisation ε [40]

$$\langle V | \bar{u} \gamma^\mu b | B \rangle = \frac{ic_V V_1 \epsilon^{\mu\nu\rho\sigma} \varepsilon_\nu^* (p_B + p_V)_\rho q_\sigma}{m_B + m_V}, \quad (2.75)$$

$$\begin{aligned} \langle V | \bar{u} \gamma^\mu \gamma^5 b | B \rangle = c_V \Bigg(& A_1(m_B + m_V) \varepsilon^{*\mu} - A_2 \frac{(p_B + p_V)^\mu \varepsilon^* \cdot q}{m_B + m_V} \\ & + \frac{\varepsilon^* \cdot q q^\mu}{q^2} [A_2(m_B - m_V) - A_1(m_B + m_V) + 2m_V A_0] \Bigg), \end{aligned} \quad (2.76)$$

where $c_V = 1/\sqrt{2}$ for ρ^0 and ω^0 and $c_V = 1$ for $\rho^\pm$ and V , A_0 , A_1 and A_2 are form factors parametrising the matrix element and depending on q^2 . It is usually convenient to redefine the form factors in the helicity basis $\{H_+, H_-, H_0, H_s\}$ such that

$$H_\pm(q^2) = \frac{2m_B |\vec{p}_V| V_1}{m_B + m_V} \pm (m_B + m_V) A_1, \quad (2.77)$$

$$H_0(q^2) = \frac{8m_B m_V}{\sqrt{q^2}} \left(\frac{|\vec{p}_V|^2 m_B}{4m_V(m_B + m_V)} A_2 - \frac{(m_B^2 - m_V^2 - q^2)(m_B + m_V)}{16m_B m_V} A_1 \right), \quad (2.78)$$

$$H_s(q^2) = \frac{2m_B m_V}{\sqrt{q^2}} A_0. \quad (2.79)$$

In this basis, the differential decay rate for $B \rightarrow V \ell \nu$ transitions is

$$\begin{aligned} \frac{d\Gamma}{dq^2} = & \frac{G_F^2 |V_{ub}|^2 c_V^2}{96\pi^3} |\vec{p}_V| \frac{q^2}{m_B^2} \left(1 - \frac{m_\ell^2}{q^2} \right)^2 \\ & \times \left[\left(1 + \frac{m_\ell^2}{2q^2} \right) (H_+^2(q^2) + H_-^2(q^2) + H_0^2(q^2)) + \frac{3m_\ell^2}{2q^2} H_s^2(q^2) \right]. \end{aligned} \quad (2.80)$$

Similarly to the $B \rightarrow \pi$ case, the form factors can be expanded in the z parameter using the BCL parametrisation. In Ref. [40], the Bharucha, Straub, Zwicky (BSZ) parametrisation [41], a modified variation of BCL, is used to fit the decay rates to combined Belle and BaBar data and LCSR constraints to extract the 11 BSZ parameters. The ρ meson decays almost exclusively to pairs of pions, whereas ω mostly decays to 3π but can also decay to $\pi^+\pi^-$. The ρ – ω interference plays an important role in the study of the $B \rightarrow \pi^+\pi^-\ell\nu$ channel. In addition, because of its relatively large width, the ρ lineshape must be taken into account in measurements of the $B \rightarrow \rho\ell\nu$ branching fraction. In the inclusive measurement discussed in this thesis, these effects play only a negligible role.

$B \rightarrow \eta\ell\nu$ and $B \rightarrow \eta'\ell\nu$ decays

The differential decay rate of $B \rightarrow \eta\ell\nu$ and $B \rightarrow \eta'\ell\nu$ decays can be expressed in a similar way to the decay rate of $B \rightarrow \pi\ell\nu$ (Equation 2.72):

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{192\pi^3 m_B^3} \lambda^{3/2}(q^2) |f_{B\eta^{(\prime)}}^+(q^2)|^2, \quad (2.81)$$

where the λ function is defined for the η and η' masses. The simplest parametrisation of the f_+ form factor gives

$$f_{B\eta^{(\prime)}}^+(q^2) = \frac{f_{B\eta^{(\prime)}}^+(0)}{(1 - q^2/m_B^2)(1 - \alpha_{B\eta^{(\prime)}}^+ q^2/m_B^2)}, \quad (2.82)$$

where α^+ is a parameter which can be determined from a fit to external data. In Ref. [42], the form factor value at $q^2 = 0$ and its q^2 -dependence for η and η' were calculated using QCD light-cone sum rules [43]. A fit of Equation 2.82 to the resulting form factors fixes the values of $\alpha_{B\eta}^+$ and $\alpha_{B\eta'}^+$.

2.3.4 Inclusive semileptonic $B \rightarrow X_c \ell \nu$ decays

Inclusive semileptonic decays to charm mesons are described in the heavy quark expansion (HQE) framework, which was presented in Section 2.3.1. It is important to note that predictions of inclusive $B \rightarrow X_c \ell \nu$ decay rates rely on the quark-hadron duality assumption which states that predictions made at the parton level over sufficiently large portions of the phase-space should yield reliable predictions of hadronic observables within the uncertainties related to the limitations of the theory, such as the finite order expansion [44]. Following Equations 2.60 and 2.61, the $B \rightarrow X_c \ell \nu$ decay rate can schematically be expressed as:

$$d\Gamma_n = \sum_i C_{n,i} \langle B | \mathcal{O}_{n+3,i} | B \rangle = d\Gamma_0 + \frac{1}{m_b^2} d\Gamma_2 + \frac{1}{m_b^3} d\Gamma_3 + \dots \quad (2.83)$$

where $\langle B | \mathcal{O}_{n+3,i} | B \rangle$ are non-perturbative matrix elements which need to be extracted from data. For instance, at second order, two matrix elements contribute. They can be expressed using non-perturbative parameters μ_π^2 and μ_G^2 such that [45, 46]

$$2m_B \mu_\pi^2 = - \langle B | \bar{b}_v (iD)^2 b_v | B \rangle, \quad (2.84)$$

$$2m_B \mu_G^2 = \langle B | \bar{b}_v (-i\sigma^{\alpha\beta}) (iD_\alpha) (iD_\beta) b_v | B \rangle, \quad (2.85)$$

where D is the covariant derivative and the b quark field is redefined in a similar way to Equation 2.56, $b_v(x) = \exp(im_b v \cdot x) b(x)$. At third order in the operator product expansion (OPE), two more matrix elements contribute. At higher orders in $(1/m_b)^n$, the number of matrix elements increases drastically, with nine at $n = 4$ and 18 at $n = 5$. However, since the parameter v is introduced artificially, quantities invariant under the infinitesimal transformation $v_\mu \rightarrow v_\mu + \delta v_\mu$ can be introduced [47]. This reparametrisation invariance allows one to relate different orders in the OPE and therefore reduce the number of independent operators. For example, at $1/m_b^5$, only 10 independent operators are found. The HQE parameters can be determined from measurements of the kinematic moments of $B \rightarrow X_c \ell \nu$ decays. The most general definition of the k^{th} moment of inclusive decays is given by

$$\langle M^k[w] \rangle = \int d\Phi f_w^k(v, p_\ell, p_\nu) H^{\mu\nu} L_{\mu\nu}, \quad (2.86)$$

where the hadronic and leptonic tensors, H and L are integrated over the decay phase-space $d\Phi$ and f_w is a weight function. The decay rate prediction can be recovered by choosing $f_w = 1$, whereas the q^2 moments can be computed by choosing $f_w = q^2$. The normalised q^2 moments are therefore defined for an experimental selection q_{cut}^2 such that

$$\langle (q^2)^k \rangle_{q^2 \geq q_{\text{cut}}^2} \equiv \int_{q_{\text{cut}}^2} dq^2 (q^2)^n \frac{d\Gamma}{dq^2} \bigg/ \int_{q_{\text{cut}}^2} dq^2 \frac{d\Gamma}{dq^2}. \quad (2.87)$$

The q^2 moments were measured by the Belle and Belle II experiments [48, 49] which allows for the simultaneous determination of the HQE non-perturbative parameters and V_{cb} [50]. Predictions for the decay rate and the q^2 moments are available up to $(1/m_b^4)$ at tree level in QCD. However, higher order corrections in α_s are sizeable and have therefore been computed. The current status of known perturbative corrections is summarised in Table 2.2.

$\Gamma / \langle (q^2)^k \rangle$	tree	α_s	α_s^2	α_s^3
Partonic	✓	✓	✓	✓
$1/m_b^2$	✓	✓		
$1/m_b^3$	✓	✓		
$1/m_b^4$	✓			

TABLE 2.2: Summary of currently known HQE parameters for the B semileptonic decay rate at different orders in $1/m_b$ and α_s . The green check marks indicate the corrections known for the decay rate and the q^2 moments, whereas the blue check mark indicates the correction known only for the decay rate. Taken from Ref. [51].

2.3.5 Inclusive semileptonic $B \rightarrow X_u \ell \nu$ decays

It is, in principle, possible to use the OPE presented above to describe semileptonic charmless B decays. However, since $|V_{cb}|^2$ is about two orders of magnitude above $|V_{ub}|^2$, these types of processes are overwhelmed by $B \rightarrow X_c \ell \nu$ decays in experimental measurements. Historically, this has required the application of tight kinematic selections using the fact that charmless decays are expected to dominate in the low hadronic mass region and in the high E_ℓ and q^2 regions where $B \rightarrow X_c \ell \nu$ decays are kinematically forbidden. For instance, because the lightest charm meson has a mass close to 1.8 GeV, selecting decays with $M_X < 1.7$ GeV rejects $B \rightarrow X_c \ell \nu$ decays very efficiently. The differences between $B \rightarrow X_c \ell \nu$ and $B \rightarrow X_u \ell \nu$ decays are illustrated in Figure 2.9 by comparing predicted distributions of E_ℓ^B , q^2 and M_X .

The downside of such selections is that they break the inclusivity requirement of the OPE, which can therefore not be used [52]. Instead, predictions of differential decay rates are necessary. The main difficulty in building a consistent theoretical description of the $B \rightarrow X_u \ell \nu$ decay rate is that it is sensitive to the *shape function* F . The shape function is a non-perturbative object which can be interpreted as a description of the b quark motion inside the B meson and is therefore a function of the residual momentum k of the b quark. Decay rate predictions are particularly sensitive to the shape function in restricted regions of phase space. The key difference with $B \rightarrow X_c \ell \nu$ decays is therefore that instead of a description based on a limited number of power-suppressed non-perturbative matrix elements, the $B \rightarrow X_u \ell \nu$ decay description includes an unknown function already at leading order in the HQE. Even though the shape function cannot be expressed analytically, it is known to be universal to B decays and can therefore be extracted from $B \rightarrow X_s \gamma$ decays [53]. In fact, at tree level and leading order in the HQE, the photon energy spectrum is directly proportional to the shape function [54]. With this idea in mind, the SIMBA collaboration proposed to perform a global fit of $B \rightarrow X_s \gamma$ and $B \rightarrow X_u \ell \nu$ data to simultaneously extract the shape function and $|V_{ub}|$ using a model-independent approach [55, 56]. Alternatively, Ref. [57] describes a parametrisation of the shape function using neural networks. In both of these approaches, the shape function is constrained by its moments M_n , which are related to the heavy-quark parameter expansion such that [58]

$$M_n \equiv \int dk k^n F(k), \quad (2.88)$$

$$M_0 = 1, \quad (2.89)$$

$$M_1 = m_B - m_b = \bar{\Lambda}, \quad (2.90)$$

$$M_2 = \bar{\Lambda}^2 - \frac{\lambda_1}{3}, \quad (2.91)$$

$$M_3 = \bar{\Lambda}^3 - \lambda_1 \bar{\Lambda} + \frac{\rho_1}{3}, \dots \quad (2.92)$$

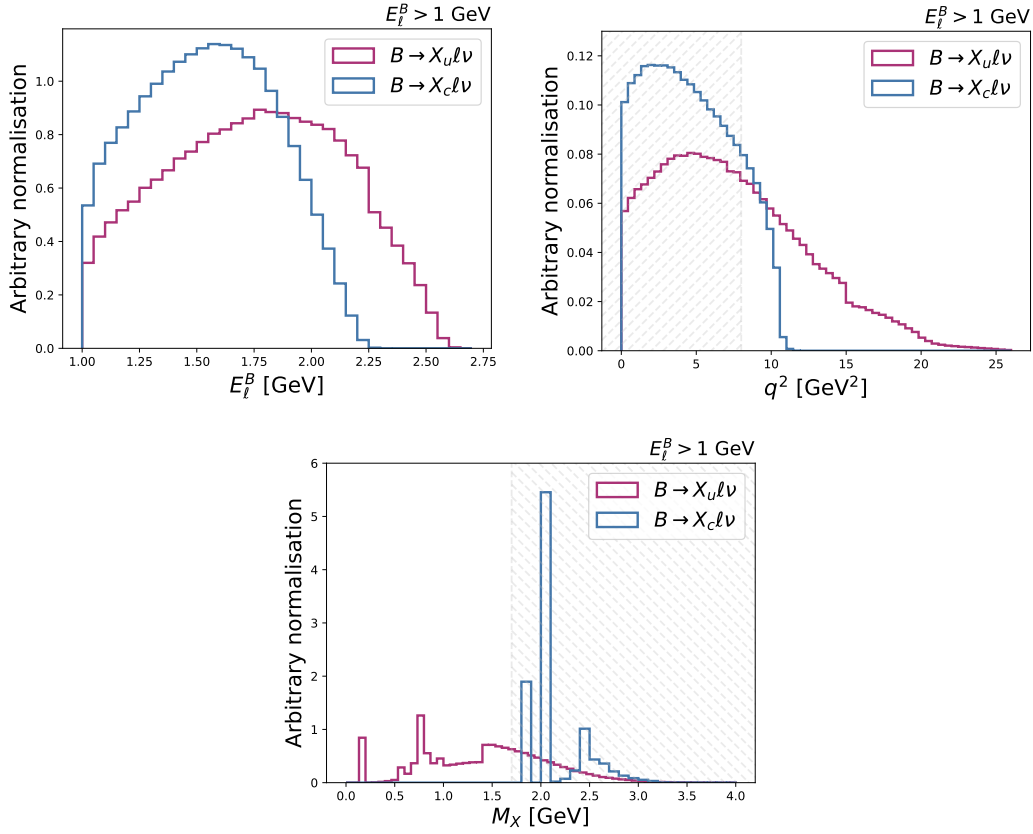


FIGURE 2.9: Histograms of E_ℓ^B (top left), q^2 (top right) and M_X (bottom) distributions of simulated $B \rightarrow X_u \ell \nu$ (purple) and $B \rightarrow X_c \ell \nu$ (blue) decays. The two histograms are normalised to the same area. The shaded regions are excluded in the experimental studies described in Chapter 5. A requirement that $E_\ell^B > 1 \text{ GeV}$ is applied to all decays.

where λ_1 and ρ_1 are HQE matrix elements¹³. At order $1/m_b$ in the HQE, additional subleading shape functions appear but little is known about them except their first few moments.

Having introduced the most important features of $B \rightarrow X_u \ell \nu$ theoretical models, we present in the following a few of the most common frameworks. For this discussion, it is useful to define the light-cone momenta $P^\pm = E_X \mp |\vec{p}_X|$, with E_X and $\vec{p}_X$ the hadronic system energy and three-momentum. The different regions discussed below can be illustrated in the (P_+, P_-) plane as shown in Figure 2.10. The OPE region is characterised by the kinematic hierarchy $\Lambda_{\text{QCD}} \ll P_+ \lesssim P_- \lesssim m_B$, whereas in the shape-function region, $\Lambda_{\text{QCD}} \sim P_+ \ll P_- \lesssim m_B$. In the resonance region, where $\Lambda_{\text{QCD}} \sim P_+ \lesssim P_- \ll m_B$, the inclusive approach is no longer valid because of the presence of individual resonances. Nevertheless, when the inclusive decay rate is computed over sufficiently large phase-space regions, the quark-hadron duality holds and the calculation including the resonance region can be performed.

¹³ λ_1 is related to μ_π^2 by $\lambda_1 = -2m_B\mu_\pi^2$ and ρ_1 is one of the third order HQE matrix elements $\rho_1 = \langle B | \bar{b}_v (iD_\mu)(v \cdot iD)(iD^\mu)b_v | B \rangle$.

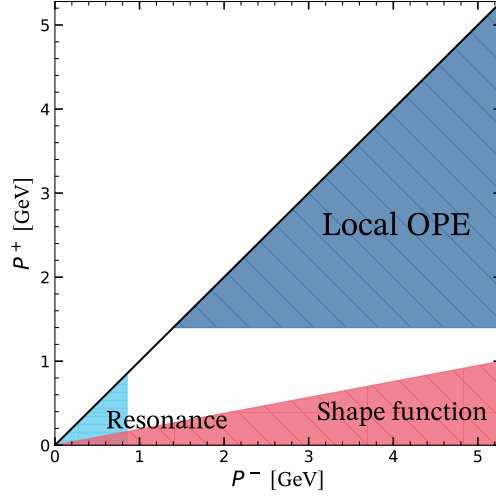


FIGURE 2.10: Illustration of different relevant regions in the $P^- : P^+$ plane: the OPE region in blue, the shape-function region in pink and the resonant region in cyan.

De Fazio, Neubert

In the De Fazio, Neubert (DFN) model [59], the triple differential decay rate is expressed as a function of the three parameters

$$x = \frac{2E_\ell}{m_b}, \quad \hat{p}^2 = \frac{p_p^2}{m_b^2}, \quad z = \frac{2v \cdot p_p}{m_b}, \quad (2.93)$$

where p_p is the total parton momentum. The decay rate is defined as

$$\begin{aligned} \frac{d^3\Gamma}{dx dz d\hat{p}^2} = 12\Gamma_0 & \left[(1 + \bar{x} - z)(z - \bar{x} - \hat{p}^2) \frac{m_b^2}{2} W_1 + (1 - z + \hat{p}^2) \frac{m_b}{2} W_2 \right. \\ & \left. + [\bar{x}(z - \bar{x}) - \hat{p}^2] \frac{m_b}{4} (W_3 + 2m_b W_4 + m_b^2 W_5) \right], \end{aligned} \quad (2.94)$$

where $\bar{x} = 1 - x$ and Γ_0 is the lowest order total decay rate

$$\Gamma_0 = \frac{G_F^2 |V_{ub}|^2 m_b^5}{192\pi^3}. \quad (2.95)$$

The five W_i are structure functions necessary to describe the hadronic tensor, which encodes the QCD dynamics of semileptonic decays. The structure functions are expressed up to $\mathcal{O}(\alpha_s)$ such that

$$W_i(z, \hat{p}^2) = W_i^{(0)}(z, \hat{p}^2) + \frac{C_F \alpha_s}{4\pi} W_i^{(1)}(z, \hat{p}^2) + \mathcal{O}(\alpha_s^2), \quad (2.96)$$

where at leading order all functions vanish except W_1 which reads $\frac{m_b^2}{2} W_1^{(0)} = \delta(\hat{p}^2)$. The shape function is defined as

$$F(k) = N(1 - k/\bar{\Lambda})^a e^{(1+a)k/\bar{\Lambda}} \quad (2.97)$$

where the parameter a is defined such that $\mu_\pi^2/3 = \bar{\Lambda}^2/(1+a)$. The shape function can therefore be defined for the set of parameters (m_b, a) or $(\bar{\Lambda}, \mu_\pi^2)$. Incorporating all ingredients described here, the DFN model yields predictions for triple, double and single differential spectra. The lepton energy spectrum as predicted in the DFN framework is illustrated in Figure 2.11.

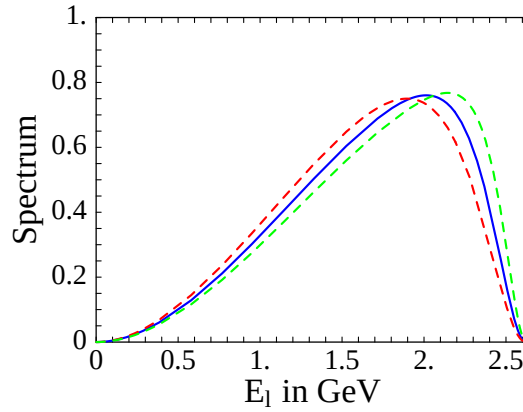


FIGURE 2.11: The lepton energy spectrum as predicted by the DFN model. Each dashed line corresponds to a prediction based on a different b quark mass value: 4.65 GeV (red), 4.8 GeV (blue) and 4.95 GeV (green) [59].

Bosch, Lange, Neubert, Paz

The model developed by Bosch, Lange, Neubert and Paz (BLNP) [54, 60] relies on complex theoretical technology to optimise the description of inclusive $B \rightarrow X_u \ell \nu$ decays in the shape-function region where the hadronic system energy is large while its mass is small *i.e.* $P_+ \ll P_-$. The BLNP model smoothly interpolates between the shape-function region and the OPE region where hadronic kinematic variables are close to the B meson mass. In the OPE region, the decay rate can be expressed in terms of local matrix elements, whereas in the shape-function region, the non-perturbative physics is summed up in the shape function. The two regimes are connected because, as outlined earlier, the shape function can be expressed in terms of HQE operators. In both regions, an expansion in inverse powers of m_b is therefore used but since the power-counting is different in the two regimes, the connection is not straightforward. In the shape-function region, different energy scales are separated through QCD factorisation theorems at each order in $1/m_b$. In the BLNP approach, the HQE expansion is effectively reorganised in terms of hadronic kinematic variables instead of partonic variables which allows one to eliminate any explicit dependency on m_b ¹⁴. One of the key elements of the BLNP model is that it also yields a prediction for the $B \rightarrow X_s \gamma$ photon energy spectrum. A fit of the predicted spectrum to precise data would therefore allow the shape function to be determined, yielding high-accuracy predictions of the partial $B \rightarrow X_u \ell \nu$ decay rate over any arbitrary phase-space region. We outline in the following the main technical elements of the BLNP approach.

The triple differential rate is expressed as a function of P_+ , P_- and $P_\ell = m_B - 2E_\ell$ as

$$\begin{aligned} \frac{d^3\Gamma}{dP_+ dP_- dP_\ell} = \frac{G_F^2 |V_{ub}|^2}{16\pi^3} (m_B - P_+) [(P_- - P_\ell)(m_B - P_- + P_\ell - P_+) \mathcal{F}_1 \\ + (m_B - P_-)(P_- - P_+) \mathcal{F}_2 + (P_- - P_\ell)(P_\ell - P_+) \mathcal{F}_3], \end{aligned} \quad (2.98)$$

where the function $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ depend directly on the structure functions defined in Equation 2.94. At leading power and tree-level, $\mathcal{F}_2$ and $\mathcal{F}_3$ vanish, whereas $\mathcal{F}_1$ is equal to the leading order shape function. At one-loop order, $\mathcal{F}_2$ doesn't contribute either. $\mathcal{F}_1$ and $\mathcal{F}_3$ on the other hand can be expressed by separating effects contributing at different energy scales: the hard

¹⁴Note that the decay rate expressions are not free of m_b terms since the first moment of the shape function depends on this parameter.

scale $\sim m_b$, the jet scale $\sim \sqrt{m_b \Lambda_{\text{QCD}}}$ and the soft scale $\sim \Lambda_{\text{QCD}}$. In this context

$$\mathcal{F}_{1,3} \approx H_{1,3} J \otimes S, \quad (2.99)$$

where H and J and respectively the hard and jet function and S is the shape function. This factorisation of hard, collinear and soft degrees of freedom is particularly useful as it allows the application of the *soft-collinear effective theory* (SCET) [61] framework. By matching SCET predictions of the structure functions to full QCD predictions and HQET predictions, expressions of H and J up to first order in α_s are obtained. Additional subleading corrections contribute to $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$. The first are kinematic power corrections, which are organised in powers of $1/(m_B - P_+)$. In the BLNP model, these corrections are included up to second order. As mentioned earlier, beyond the leading HQE order, additional subleading shape functions enter the $B \rightarrow X_u \ell \nu$ decay rate. These functions contribute to the expressions of $\mathcal{F}_1$ and $\mathcal{F}_3$ but since they cannot be extracted from the $B \rightarrow X_s \gamma$ photon energy spectrum, they yield additional theoretical uncertainties. At order $(1/m_b)^2$, another set of sub-subleading shape functions contributes. The associated expressions are obtained by subtracting the leading and subleading $\mathcal{F}_i$ contributions expanded in terms of their moments from the $(1/m_b)^2$ correction to the pure OPE triple-differential $B \rightarrow X_u \ell \nu$ decay rate. These residual hadronic power corrections are therefore suppressed by $1/(m_B - P_+)^2$ and depend on the leading order shape function and OPE parameters. The last contribution that must be taken into account stems from the weak annihilation contribution to $B \rightarrow X_u \ell \nu$ decays. We discuss this effect in more detail in Section 2.3.5 below. Since little is known about the weak annihilation contribution on both the experimental and theoretical level, BLNP suggests to either add a related uncertainty on the total rate prediction (and similarly on experimental determinations of $|V_{ub}|$) or to simply reject the high- q^2 region where weak annihilation is expected to contribute. In the BLNP model, three functional forms of the shape function are proposed: an exponential-type, a Gaussian-type and a hyperbolic-type function. They are motivated by the shape of the $B \rightarrow X_s \gamma$ photon energy spectrum and they each depend on two parameters which can be tuned. However, the only known constraints on the subleading shape functions are the expressions of their first moment and the fact that their total normalisation vanishes. Based on this information, there is an infinite number of possible functions. BLNP chooses to express the subleading functions as the sum of the leading shape function's first derivative and an arbitrary set of functions. The choices made for the leading and subleading shape functions are varied to estimate related uncertainties. The P_+ spectrum as predicted in the BLNP framework is shown in Figure 2.12.

Gambino, Giordano, Ossola, Uraltsev

The model developed by Gambino, Giordano, Ossola and Uraltsev (GGOU) [62] takes a different approach compared to the BLNP model as it avoids focusing on the region where $P_+ \ll P_-$. The differential decay rate is first defined as

$$\begin{aligned} \frac{d^3\Gamma}{dq^2 dq_0 dE_\ell} &= \frac{G_F^2 |V_{ub}|^2}{8\pi^3} \left\{ q^2 W_1 - \left[2E_\ell^2 - 2q_0 E_\ell + \frac{q^2}{2} \right] W_2 + q^2 (2E_\ell - q_0) W_3 \right\} \\ &\times \theta \left(q_0 - E_\ell - \frac{q^2}{4E_\ell} \right) \theta(E_\ell) \theta(q^2) \theta \left(q_0 - \sqrt{q^2} \right), \end{aligned} \quad (2.100)$$

where q_0 is the total leptonic energy and θ are Heaviside functions. In this case, two of the five structure functions are neglected since the charged lepton is considered to be massless. In the following, q_0 and q^2 will be normalised such that $\hat{q}_0 = q_0/m_b$ and $\hat{q}^2 = q^2/m_b^2$. The key initial idea of the GGOU model is to split each structure function into a perturbative part and

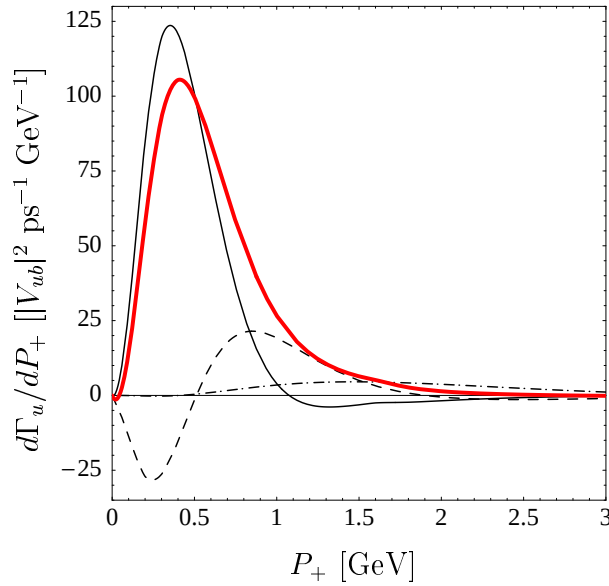


FIGURE 2.12: The P_+ spectrum as predicted by the BLNP model. The black solid line represents the leading-power prediction, the dashed line the first-order power corrections, the dash-dotted line the second-order power corrections and the red line the sum of all these contributions [60].

a power-suppressed non-perturbative part based on a cut-off μ ¹⁵ such that

$$W_i(q_0, q^2, \mu) = m_b(\mu)^{n_i} \left[W_i^{\text{pert}}(\hat{q}_0, \hat{q}^2, \mu) + W_i^{\text{pow}}(\hat{q}_0, \hat{q}^2, \mu) \right] \quad (2.101)$$

with $n_{1,2} = -1$ and $n_3 = -2$. The W_i^{pow} correspond to the OPE power corrections up to third order. To be able to take the HQE parameters directly from fits of $B \rightarrow X_c \ell \nu$ moments, the kinetic renormalisation scheme is used. The cut-off scale dependence cancels out when computing inclusive quantities such as moments of W_i . The perturbative terms W_i^{pert} are expanded up to order $\mathcal{O}(\alpha_s^2 \beta_0)$ with $\beta_0 = 11 - \frac{2}{3}N_f$ where the number of active flavours is set to $N_f = 3$ (see Equation 2.6). Furthermore, the structure functions can be expressed as

$$W_i(q_0, q^2) = m_b^{n_i}(\mu) \int dk F_i(k, q^2, \mu) W_i^{\text{pert}} \left[q_0 - \frac{k}{2} \left(1 - \frac{q^2}{m_b m_B} \right), q^2, \mu \right]. \quad (2.102)$$

Here, the three distribution functions F_i include the leading and subleading functions, which makes them non-universal and therefore different for each structure function. Since they are defined to absorb all infrared effects below the scale μ , they depend on this parameter. The distribution functions can be constrained by matching Equation 2.102 to the OPE description of the structure functions as defined in Equation 2.101. This results in an expression of each structure function up to orders $\mathcal{O}(1/m_b^3)$ and $\mathcal{O}(\alpha_s^2 \beta_0)$ which doesn't depend on μ and where the only arbitrary choice corresponds to the functional form of the distribution functions F_i . Ref. [62] considers four two-parameter functional forms of the distribution functions – an exponential form, a Gaussian form, a hyperbolic form and a Roman form – and multiplies them by a distortion factor such that the distribution functions always match their moment constraints. An uncertainty related to the choice of functional form is extracted by using the different proposed functions. Finally, the GGOU model carefully takes into account effects in the high- q^2 region

¹⁵In Ref. [62], $\mu = 1$ GeV is chosen for numerical results.

where the OPE breaks down and where weak annihilation plays a non-negligible role. The M_X^2 spectrum as predicted by the GGOU model is shown in Figure 2.13.

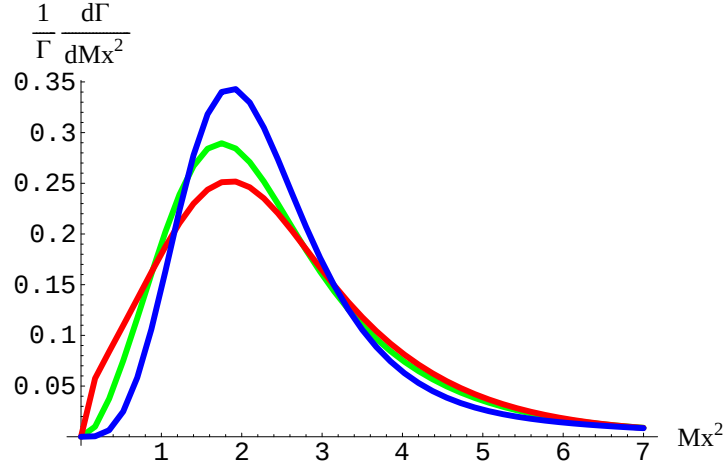


FIGURE 2.13: The M_X^2 spectrum as predicted in the GGOU framework using three different functional forms for the distribution functions F_i [62].

Other models and additional considerations

In the $|V_{ub}|$ measurement presented in this thesis, the DFN and BLNP frameworks are chosen to model the $B \rightarrow X_u \ell \nu$ events. The GGOU calculation is not implemented in any event generator but it is chosen by the Heavy Flavour Averaging Group (HFLAV) to quote a reference value of $|V_{ub}|$ [63]. Besides these three frameworks, other models have been developed to describe inclusive $B \rightarrow X_u \ell \nu$ decays. Most notably, the Dressed Gluon Exponentiation (DGE) computation [64] relies on a resummation of the perturbation theory. The DGE model utilises the momentum scale hierarchy in the semileptonic decays without introducing a momentum cut-off. The model developed by Aglietti, Di Lodovico, Ferrera and Ricciardi (ADFR) [65] also relies on soft-gluon resummation and introduces an effective QCD coupling free from Landau singularities. The resummation methodology makes it possible to calculate the elements of the differential decay rate, rather than parametrising them as in the models described above. However, since no uncertainties are associated to the assumptions on which the more model-dependent DGE and ADFR frameworks are based on, the BLNP and GGOU descriptions are deemed to be more reliable. The BLNP, GGOU, DGE and ADFR frameworks are compared in Figure 2.14. Finally, the framework presented in Ref. [58] computes the $B \rightarrow X_u \ell \nu$ differential decay rate at NNLO and N³LL accuracy using SCET.

Two additional contributions to $B \rightarrow X_u \ell \nu$ decays must be mentioned: multi-hadron final states and weak annihilation. The first one corresponds to non-resonant decays where the hadronic final state is comprised of more than one hadron, such as $B \rightarrow \pi \pi \ell \nu$. In principle, this type of decays are indirectly described by non-resonant models and do not require a separate treatment for an inclusive measurement. However, they can represent important contributions for exclusive measurements and therefore need to be studied precisely. One multi-hadron channel, namely $B \rightarrow K K \ell \nu$, impacts inclusive charmless measurements directly. Since kaons are characteristic of charm meson decays, they are usually rejected, thus removing $B \rightarrow K K \ell \nu$ decays, which contribute to the $B \rightarrow X_u \ell \nu$ rate through $s\bar{s}$ fragmentation. In the measurement presented here, an uncertainty related to this contribution is derived using hadronisation modelling parameters but with the uncertainty on $|V_{ub}|$ expected to decrease in the near future, it will become crucial

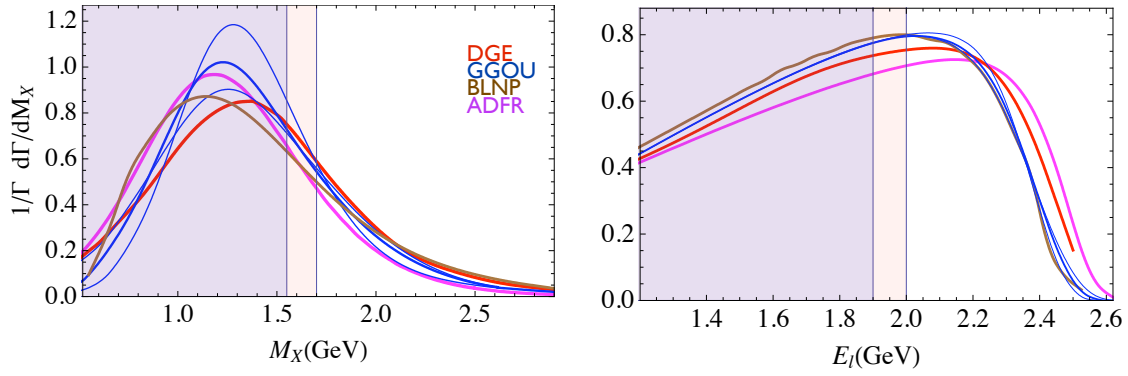


FIGURE 2.14: Comparison of the hadronic mass (left) and lepton energy (right) spectra as predicted by the DGE (red), GGOU (blue), BLNP (brown) and ADFR (pink) models of inclusive $B \rightarrow X_u \ell \nu$ decays. All four spectra are computed with a fixed set of inputs. Similarly to Figure 2.13, three GGOU predictions are shown based on different functional forms of the distribution functions. The lavender and beige areas correspond to the regions $M_X < 1.55$ GeV, $E_\ell < 1.9$ GeV and $M_X < 1.7$ GeV, $E_\ell < 2.0$ GeV respectively. The figures are taken from Ref. [66].

to estimate precisely the $B \rightarrow KK\ell\nu$ decay rate. On the other hand, the weak annihilation contribution to $B \rightarrow X_u \ell \nu$ decays arises at third order in the HQE in the form of four-quark operators of the type [67–69]

$$\begin{aligned}\mathcal{O}_{V-A} &= (\bar{b}\gamma_\mu u_L)(\bar{u}_L\gamma_\mu b), \\ \mathcal{O}_{S-P} &= (\bar{b}u_L)(\bar{u}_L b).\end{aligned}\tag{2.103}$$

The weak annihilation contribution to the total decay rate can therefore be expressed as

$$\delta\Gamma = \frac{1}{12\pi} G_F^2 |V_{ub}|^2 f_B^2 m_b^2 m_B (B_2 - B_1),\tag{2.104}$$

where f_B is the B meson decay constant and B_1 and B_2 are defined such that

$$\begin{aligned}\langle B | \mathcal{O}_{V-A} | B \rangle &= \frac{f_B^2 m_B}{8} B_1, \\ \langle B | \mathcal{O}_{S-P} | B \rangle &= \frac{f_B^2 m_B}{8} B_2.\end{aligned}\tag{2.105}$$

Compared to other $\mathcal{O}(1/m_b^3)$ contributions, weak annihilation is enhanced by a factor $16\pi^2$. Weak annihilation processes do not necessarily occur as an interaction between the b quark and the valent u quark but could also be due to annihilation with a sea quark, as illustrated in Figure 2.15. The neutral B channel could therefore yield a non-vanishing contribution, which could even be larger than for the charged B channel. If the matrix elements in Equation 2.105 are factorised and the vacuum $|0\rangle\langle 0|$ is inserted, the two constants are expected to be equal – $B_1 = B_2 = 1$ for $B^\pm$, $B_1 = B_2 = 0$ for $\bar{B}^0$. The weak annihilation contribution would therefore vanish but these assumptions are expected to break at roughly the 10% level. Furthermore, the hadronic system is expected to be very soft, which would lead to the observation of a peaking structure at high leptonic energies. It is, however, not straightforward to quantify precisely the effect that weak annihilation has on the total decay rate and it is therefore usually included as a source of theoretical uncertainty (except in the GGOU model, which accounts for the contribution).

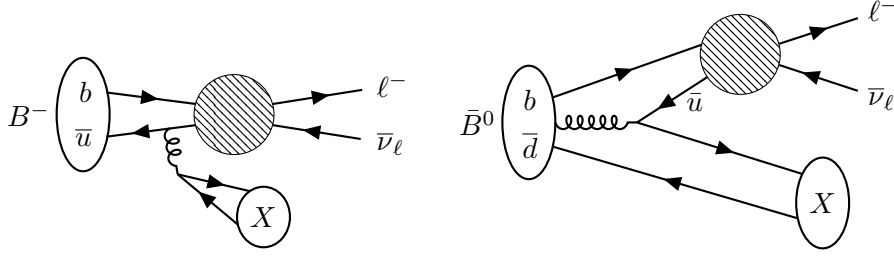


FIGURE 2.15: Feynman diagrams of the weak annihilation contributions to $B \rightarrow X_u \ell \nu$ decays.

We motivated the modelling of non-resonant $B \rightarrow X_u \ell \nu$ decays by referring to the phase-space cuts necessary to reduce the large background contamination. However, with machine learning algorithms, it is now possible to isolate regions with high signal purity without tight selections on kinematic variables. In fact, in the measurement presented in this manuscript, the only kinematic selection applied to define the nominal signal region is the requirement that $E_\ell^B > 1$ GeV. One could therefore imagine a measurement which doesn't rely on a specific model of the inclusive tripe-differential decay rate but rather on the OPE directly, similarly to what is done for inclusive $B \rightarrow X_c \ell \nu$ studies. Two main caveats, however, still make it necessary to use the inclusive models. First, the non-resonant contribution to the inclusive $B \rightarrow X_u \ell \nu$ decay rate is much larger than for the $B \rightarrow X_c \ell \nu$ decay rate. The relatively small gap between sum of exclusive rates and inclusive rate for $B \rightarrow X_c \ell \nu$ decays is usually filled with a set of resonant modes¹⁶ which is not feasible for $B \rightarrow X_u \ell \nu$ decays. Second, the relatively large $B \rightarrow X_c \ell \nu$ decay rate and the clean data samples allow for a simultaneous extraction of OPE parameters and $|V_{cb}|$ through a measurement of kinematic moments which, because of the limited event yields and large backgrounds, could only be done with relatively low precision with $B \rightarrow X_u \ell \nu$ decays. Directly measuring the inclusive $B \rightarrow X_u \ell \nu$ branching fraction and extracting $|V_{ub}|$ from it, therefore, remains the optimal strategy.

2.4 Experimental overview

Over the past 25 years, the value of $|V_{ub}|$ has been measured several times using semileptonic B decays, starting with measurements by the CLEO collaboration in the early 2000s. Since then, $b \rightarrow u$ transitions have been studied mostly at B factories even though measurements at hadronic collider experiments such as LHCb were also performed (see next chapter for a more detailed description of flavour physics experiments). The main advantage of B factories is the possibility to reconstruct the partner B meson, which helps determining precisely the kinematics of semileptonic events despite the presence of an undetected neutrino and an inclusive hadronic system. We give in this section an overview of existing measurements of $|V_{ub}|$ and discuss the existing tension between the value extracted via exclusive and inclusive decays.

2.4.1 Exclusive $B \rightarrow X_u \ell \nu$ measurements

The value of $|V_{ub}|$ has been measured through decays to π , ρ and ω with fits to the $B \rightarrow \pi \ell \nu$ q^2 spectrum yielding the most precise values. The four most precise measurements were performed with data collected by the BaBar and Belle experiments. The two BaBar measurements were untagged, meaning that the partner B meson was not reconstructed, thereby maximising the available statistics for the measurements whereas Belle performed an untagged measurement and

¹⁶Because these *gap modes* are poorly known, the methodology is seriously questioned though this is not the main focus of this thesis.

a tagged measurement where the tag-side B meson was reconstructed in its hadronic decays. Using these four measurements, the HFLAV averages the extracted q^2 spectrum and fits it to the BCL form factor parametrisation with additional lattice QCD constraints [63]. The obtained $|V_{ub}|$ average reads

$$|V_{ub}|_{B \rightarrow \pi \ell \nu} = (3.75 \pm 0.06 \pm 0.19) \times 10^{-3}, \quad (2.106)$$

where the first uncertainty is experimental and the second is theoretical. An untagged measurement of the $B \rightarrow \pi \ell \nu$ q^2 spectrum was recently performed by the Belle II collaboration [70] but its results have not been included in the HFLAV averages. Measurements of $|V_{ub}|$ via decays to vector mesons have also been performed by Belle and BaBar, though they yield less precise measurements since ρ mesons decay to pairs of pions and ω mesons mostly decay to three pions. Besides the experimental challenges added by these multi-body decays, theoretical uncertainties arise from the modelling of the ρ and ω lineshapes and interference effects between the two modes [71]. When combined and averaged, the most precise measurements of the $B \rightarrow \rho/\omega \ell \nu$ q^2 spectra yield [40]

$$\begin{aligned} |V_{ub}|_{B \rightarrow \rho \ell \nu} &= (2.96 \pm 0.29) \times 10^{-3}, \\ |V_{ub}|_{B \rightarrow \omega \ell \nu} &= (2.99 \pm 0.35) \times 10^{-3}. \end{aligned} \quad (2.107)$$

The two values are below the $B \rightarrow \pi \ell \nu$ -extracted average by $1 - 2\sigma$. Measurements of $|V_{ub}|$ through $B \rightarrow \eta/\eta' \ell \nu$ decays yield far too large uncertainties to be competitive with $B \rightarrow \pi \ell \nu$ measurements. The LHCb collaboration reports two measurements of the $|V_{ub}|/|V_{cb}|$ ratio via exclusive semileptonic decays of Λ_b baryons and B_s^0 mesons [72, 73]. Using the latest Particle Data Group (PDG) [24] average value of $|V_{cb}| = (41.1 \pm 1.2) \times 10^{-3}$ and the average value of $|V_{ub}|/|V_{cb}|$ from the two LHCb measurements, a value of $|V_{ub}| = (3.43 \pm 0.32) \times 10^{-3}$ is obtained. The HFLAV includes in its $|V_{ub}| - |V_{cb}|$ combined average the two LHCb results in addition to $B \rightarrow \pi \ell \nu$ measurements, whereas the PDG doesn't use the LHCb information, thus yielding slightly different averages:

$$\begin{aligned} |V_{ub}|_{\text{HFLAV}}^{\text{excl}} &= (3.43 \pm 0.12) \times 10^{-3}, \\ |V_{ub}|_{\text{PDG}}^{\text{excl}} &= (3.70 \pm 0.16) \times 10^{-3}. \end{aligned} \quad (2.108)$$

2.4.2 Inclusive $B \rightarrow X_u \ell \nu$ measurements

The Belle and BaBar collaborations have performed several tagged and untagged inclusive measurements in different phase-space regions. Each measurement is carried out assuming a specific inclusive model and assigning uncertainties associated to this assumption. The resulting partial branching fraction can then be extrapolated to a value of $|V_{ub}|$ valid over the full phase-space for each available model. The latest BaBar analysis [74] measures a value of the inclusive branching fraction for four different models, highlighting large discrepancies in the lepton momentum modelling, especially in the endpoint region. The three theoretical frameworks chosen by the HFLAV to extract $|V_{ub}|$ are BLNP, DGE and GGOU, with the latter one used to quote the nominal average. The values of $|V_{ub}|$ for each of these three models as measured by CLEO, Babar and Belle studies are summarised in Table 50 in Ref. [63]. The averages of all these measurements read

$$\begin{aligned} |V_{ub}|_{\text{BLNP}}^{\text{incl}} &= (4.25 \pm 0.13 \pm 0.19) \times 10^{-3}, \\ |V_{ub}|_{\text{DGE}}^{\text{incl}} &= (3.87 \pm 0.10 \pm 0.10) \times 10^{-3}, \\ |V_{ub}|_{\text{GGOU}}^{\text{incl}} &= (4.06 \pm 0.12 \pm 0.11) \times 10^{-3}, \end{aligned} \quad (2.109)$$

where the first uncertainty is experimental and the second is theoretical. The PDG quotes the arithmetic average of the three values from the 2021 HFLAV report [75] and increases the uncertainty based on the spread of values from different models: $|V_{ub}|_{\text{PDG}}^{\text{incl}} = (4.13 \pm 0.26) \times 10^{-3}$, which differs only slightly from the latest nominal HFLAV average. It is interesting to note that, even though the three values agree with each other within uncertainties, the theoretical uncertainties largely differ between models. The value of $|V_{ub}|_{\text{GGOU}}^{\text{incl}}$ differs from the HFLAV and PDG exclusive averages from Equation 2.108 by about 3.0 and 1.4σ respectively. The PDG also quotes an averaged value of $|V_{ub}|$, which reads

$$|V_{ub}|_{\text{PDG}}^{\text{avg}} = (3.82 \pm 0.20) \times 10^{-3}, \quad (2.110)$$

where the uncertainty has been increased based on the χ^2 value of the combined average.

2.4.3 Summary of $B \rightarrow X_u \ell \nu$ measurements

The CKMfitter group extracts an average value of $|V_{ub}|$ from a global fit to the CKM unitarity triangle using inputs from measurements of other CKM parameters but excluding inputs from measurements of $|V_{ub}|$ [17]. The value obtained from this global fit reads

$$|V_{ub}|_{\text{CKMf}}^{\text{avg}} = (3.67^{+0.07}_{-0.08}) \times 10^{-3}, \quad (2.111)$$

which is closer to the values measured in exclusive measurements than in inclusive measurements. We compare in Figure 2.16 constraints obtained by CKMfitter from tree-level quantities using only input on $|V_{ub}|$ from inclusive or exclusive decays. This comparison confirms that the global fit appears to favour the value of $|V_{ub}|$ obtained from exclusive measurements.

The evolution over the past two decades of the exclusive and inclusive PDG averages is summarised in Figure 2.17.

2.5 Additional remarks

A possible way to ease the tension between the inclusive and exclusive values of $|V_{ub}|$ would be to introduce a small right-handed contribution in the hadronic current, such that [76]

$$H_\mu = \bar{u} \gamma_\mu P_L b \rightarrow \bar{u} \gamma_\mu P_L b + \varepsilon_R \bar{u} \gamma_\mu P_R b, \quad (2.112)$$

where ε_R quantifies the strength of the right-handed current. Such a current appears to be favoured by inclusive and $B \rightarrow \pi \ell \nu$ measurements but disfavoured by the $\Lambda_b^0 \rightarrow p \mu \nu$ measurement (see Figure 4 in Ref. [72]). Measurements of angular variables in $B \rightarrow \rho \ell \nu$ decays could help shed light on a potential right-handed contribution.

Even though the focus of the study presented here is not on a measurement of $|V_{cb}|$, $B \rightarrow X_u \ell \nu$ decays are largely dominated by background from $B \rightarrow X_c \ell \nu$ decays. A precise understanding of $B \rightarrow X_c \ell \nu$ decays is therefore essential for an accurate measurement of $|V_{ub}|$. The inclusive and exclusive values of $|V_{cb}|$ disagree at the level of about 3σ (see Figure 25 in Ref. [63] for a combined average of $|V_{ub}|$ and $|V_{cb}|$). As of now, it is not yet clear why the OPE-extracted inclusive value differs from the exclusive value, which relies on relatively well-known form factors. One possible explanation could come from B meson decays to radially excited D^{**} states and multiple hadrons, which are not as well modelled as decays to D and D^* mesons.

Given the experimental and theoretical challenges posed by $B \rightarrow X_u \ell \nu$ decays to light leptons, decays to tau leptons have only marginally been studied. For the inclusive measurement

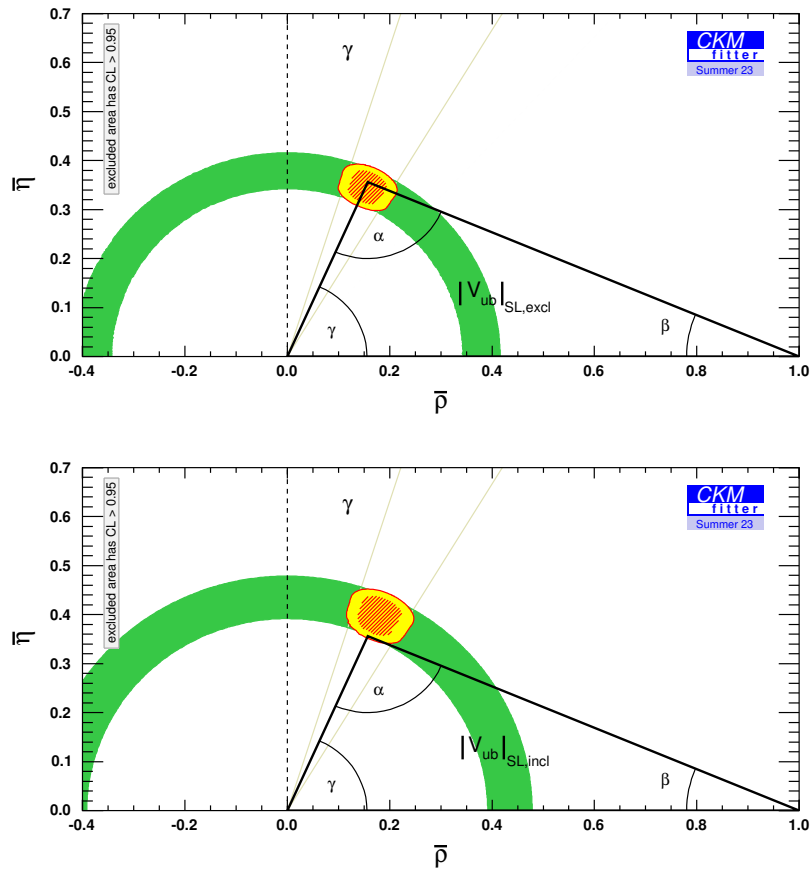


FIGURE 2.16: Comparison of constraints on $|V_{ub}|$ (green band) obtained from exclusive (top) or inclusive (bottom) semileptonic decays. The black triangle is the best CKM triangle obtained from the global fit. The grey band corresponds to the constraint obtained on the angle γ and the red and yellow circle marks the overlap between the $|V_{ub}|$ and γ constraints. Taken from Ref. [17].

presented here, they would add a relatively small amount of data but would greatly complicate the event reconstruction and treatment of systematic uncertainties. However, $B \rightarrow X_c \tau \nu$ modes are currently under great scrutiny because of the 3σ discrepancy between the SM prediction and the measurements of the ratios $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)} \tau \nu) / \mathcal{B}(B \rightarrow D^{(*)} \ell \nu)$ with $\ell = e, \mu$ [63].

In this chapter, we only presented semileptonic decays as a means of measuring $|V_{ub}|$ but purely leptonic B decays represent the cleanest channel for such measurements. The only hadronic input needed to measure the $B^\pm \rightarrow \ell^\pm \nu$ decay rate is the B meson decay constant f_B which has been determined by lattice QCD calculations to a precision level of about 0.7%. The downside of leptonic decays is that they are helicity suppressed. Therefore, extremely large amounts of data must be collected by experiments to reduce the uncertainty of $|V_{ub}|$ measured through leptonic decays to the current level of precision obtained with semileptonic decays. As of now, no measurement of $\mathcal{B}(B^\pm \rightarrow \ell^\pm \nu)$ has ever crossed the 5σ threshold. If the Belle II experiment or proposed future experiments such as FCC-ee or CEPC manage to collect sufficiently large datasets, measurements of $|V_{ub}|$ via $B \rightarrow \tau \nu$ or even $B \rightarrow \mu \nu$ appear to be feasible.

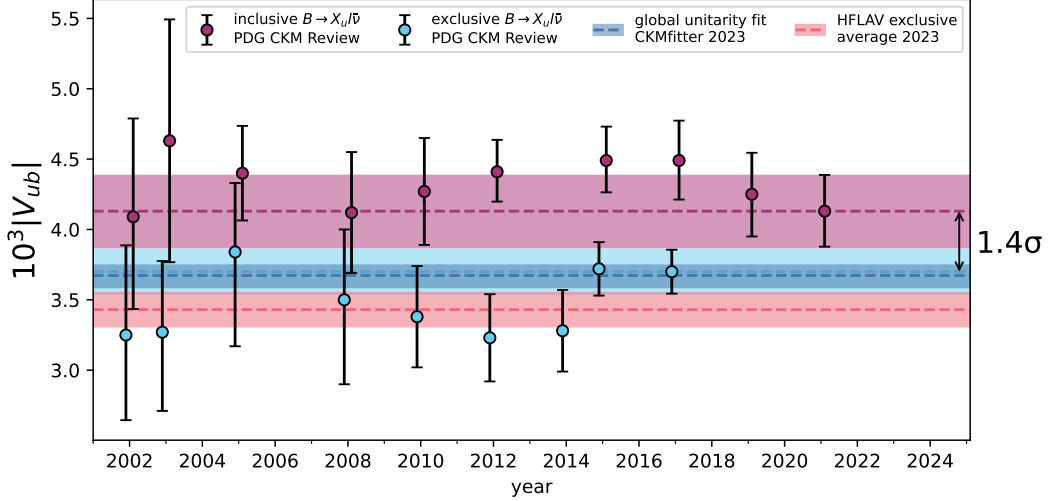


FIGURE 2.17: PDG averages of inclusive (purple points) and exclusive (cyan points) $|V_{ub}|$ values over the past 23 years. The latest averages are also shown with purple and cyan bands. The latest CKMfitter extracted value is shown as a blue band and the latest exclusive HFLAV average is shown as a pink band. The latest HFLAV inclusive average differs only slightly from the PDG average and is therefore not shown. The difference between the exclusive and inclusive PDG averages is calculated by the author of this thesis. Adapted from Ref. [58].

2.6 Conclusion and role of the measurement

Measuring the nine CKM matrix element magnitudes to test the unitarity of the matrix represents a crucial precision test of the SM. As was highlighted in the previous sections, inclusive and exclusive measurements of $|V_{ub}|$ yield significantly different values, resulting in one of the largest relative uncertainty of all nine elements, which limits the precision of the CKM unitarity test. In the context of the $|V_{ub}|$ puzzle, the study presented in this manuscript therefore serves one main purpose: to improve the precision of the measured value of $|V_{ub}|$. To extract the value of $|V_{ub}|$, we measure the partial branching fraction of inclusive $B \rightarrow X_u \ell \nu$ decays. As highlighted throughout this chapter, there remain numerous open questions regarding the modelling of inclusive $B \rightarrow X_u \ell \nu$ decays. Different models employ various techniques to describe these decays, resulting in different determinations of $|V_{ub}|$, as shown in Equation 2.109. Moreover, the choice of how to parametrise the shape function introduces additional uncertainty in predictions. Experimental selections designed to suppress background contributions can further affect the extracted value of $|V_{ub}|$ as the applied cuts alter the kinematic phase space in which the measurement is performed. The measurement described in this thesis aims to provide insights into the determination of $|V_{ub}|$ and the $|V_{ub}|$ puzzle. We present a measurement of inclusive $B \rightarrow X_u \ell \nu$ branching fractions in three phase-space regions defined by applying the following sets of kinematic selections:

- $E_\ell^B > 1 \text{ GeV}$,
- $E_\ell^B > 1 \text{ GeV}$ and $M_X > 1.7 \text{ GeV}$
- $E_\ell^B > 1 \text{ GeV}$, $M_X > 1.7 \text{ GeV}$ and $q^2 > 8 \text{ GeV}^2$.

From each partial branching fraction $\Delta\mathcal{B}$ measured over a given phase-space region, we extract $|V_{ub}|$ via

$$|V_{ub}| = \sqrt{\frac{\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)}{\tau_B \Delta\Gamma(B \rightarrow X_u \ell \nu)}}, \quad (2.113)$$

where τ_B is the B meson lifetime and $\Delta\Gamma$ is the partial decay rate predicted by a theoretical calculation (excluding the $|V_{ub}|^2$ factor) over the same phase-space region. We use partial decay rates predicted by the BLNP, DGE and GGOU frameworks.

Chapter 3

Experimental setup

“Never send a human to do a machine’s job.”

Agent Smith

We give in this chapter an overview of the experimental apparatus which produced the dataset of B meson decays used for the measurement presented in this thesis. We start by introducing the concept of B factory with a quick historical overview of such experiments and we highlight their main advantages. We then present SuperKEKB, a complex accelerator facility which produces and accelerates the electron and positron beams that are collided at the Belle II interaction point. Finally, we detail the properties of the various components of the Belle II detector placed around this interaction point and discuss the most important aspects of the Belle II software framework.

3.1 B factories

In the 1980’s, the discovery of a long B meson lifetime and $B^0 - \bar{B}^0$ mixing indicated that the idea developed by Kobayashi and Maskawa [77], namely that three generations of quarks are needed to introduce CP -violation, could be experimentally tested via B meson decays. Technological developments during that decade made possible the study of CP -violation by producing a large number of B mesons from asymmetric e^+e^- collisions at the $\Upsilon(4S)$ mass resonance as was proposed in Ref. [78]. This concept led to the approval of the BaBar experiment¹ placed at the PEP-II electron-positron storage ring and the Belle experiment² placed at the KEKB ring in the early 1990s. Before these two B factories were built, the CLEO detector, placed at the Cornell Electron Storage Ring (CESR) and operational between 1979 and 2008, was redesigned several times to optimise the early study of B mesons. B meson decays have also been studied at hadron colliders in the HERA-B experiment, for example and more recently at LHCb. However, B factories present important advantages in the study of flavour and B physics:

- The $e^+e^- \rightarrow \Upsilon(4S)$ cross-section $\sigma_{\Upsilon(4S)} = 1.1$ nb [79] is sizeable compared to the cross-section of $e^+e^- \rightarrow q\bar{q}$ with $q \in \{u, d, s, c\}$, $\sigma_{q\bar{q}} = 3.7$ nb. In the study of B decays, background processes coming from such events (called continuum events) are therefore kept to a reasonable level. By collecting data at a centre-of-mass energy slightly below the $\Upsilon(4S)$ mass, light-quark pair events can also be studied. The $e^+e^- \rightarrow \text{hadrons}$ cross-section as a function of the centre-of-mass energy is summarised in Figure 3.1.
- The environment is extremely clean as opposed to hadron colliders which allows the precise reconstruction of both charged and neutral intermediate resonances. The low detector occupancy allows for a high reconstruction efficiency of B meson decays.

¹Which obviously used the famous elephant as its mascot.

²Which sadly did not use the famous Disney princess as its mascot.

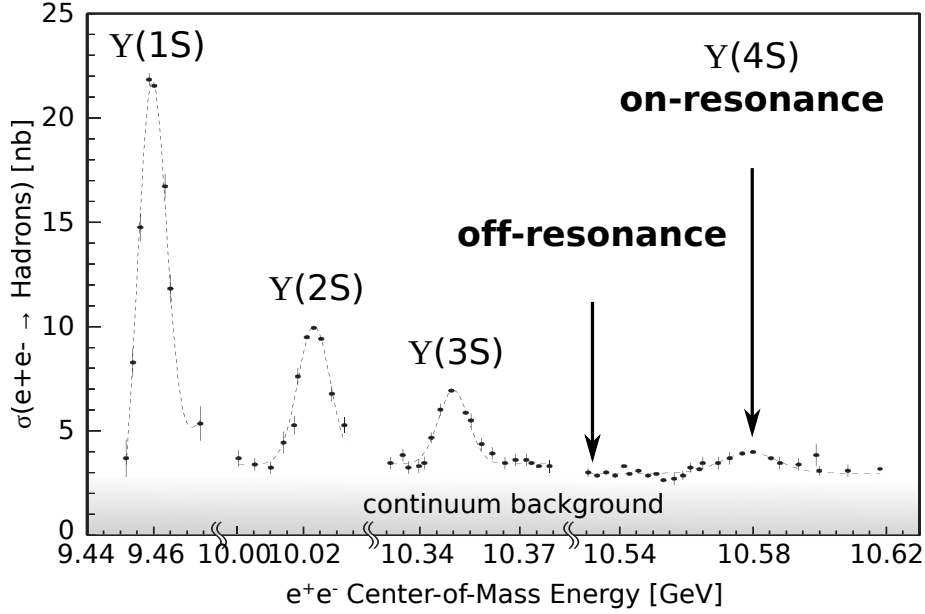


FIGURE 3.1: $e^+e^- \rightarrow \text{hadrons}$ cross-section as a function of the centre-of-mass energy as measured at the Cornell Electron Storage Ring by the Columbia University-Stony Brook detector [81]. At B factories, the centre-of-mass energy is most commonly chosen to be the $\Upsilon(4S)$ resonance mass. Collisions around 60 MeV below the $\Upsilon(4S)$ resonance provide data on $e^+e^- \rightarrow q\bar{q}$ processes ($q \in \{u, d, s, c\}$).

- The precisely known initial state kinematics can be exploited to study decays with missing particles as well as inclusive decays where one of the final state objects is not reconstructed.
- One of the fundamental requirements of a B factory is the asymmetric energy of the electron and positron beams. The boost of the subsequent neutral B mesons allows the measurement of their time sequence and therefore of time-dependent CP -violation.
- Finally, B factories allow the study of various types of heavy and light flavour decays such as tau lepton decays or charm meson decays. The $e^+e^- \rightarrow \tau^+\tau^-$ and $e^+e^- \rightarrow c\bar{c}$ cross-sections are 0.9 nb and 1.3 nb respectively at the $\Upsilon(4S)$ centre-of-mass energy. Besides, the measurement of the $e^+e^- \rightarrow \pi^+\pi^-$ cross-section leads to a precise determination of the hadronic vacuum polarisation contribution to the muon anomalous magnetic moment [80].

The cross-sections of the most common products of an e^+e^- collision at the $\Upsilon(4S)$ centre-of-mass energy are summarised in Figure 3.2.

3.2 The SuperKEKB collider

The SuperKEKB accelerator [83], located in Tsukuba, Japan, is an upgrade of the KEKB project, which operated between 1998 and 2010 and was used to collect the full Belle experiment dataset of more than 1 ab^{-1} . SuperKEKB is an electron-positron collider with asymmetric beam energies of $E_{e^+} = 4 \text{ GeV}$ for positrons and $E_{e^-} = 7 \text{ GeV}$ for electrons. The two beams are sent in two separate rings, which only cross at a specific interaction point, around which the Belle II detector is placed. The centre-of-mass energy of the e^+e^- collisions is given by:

$$\sqrt{s} = \sqrt{(E_{e^+} + E_{e^-})^2 - (\vec{p}_{e^+} + \vec{p}_{e^-})^2} \approx 10.58 \text{ GeV}, \quad (3.1)$$

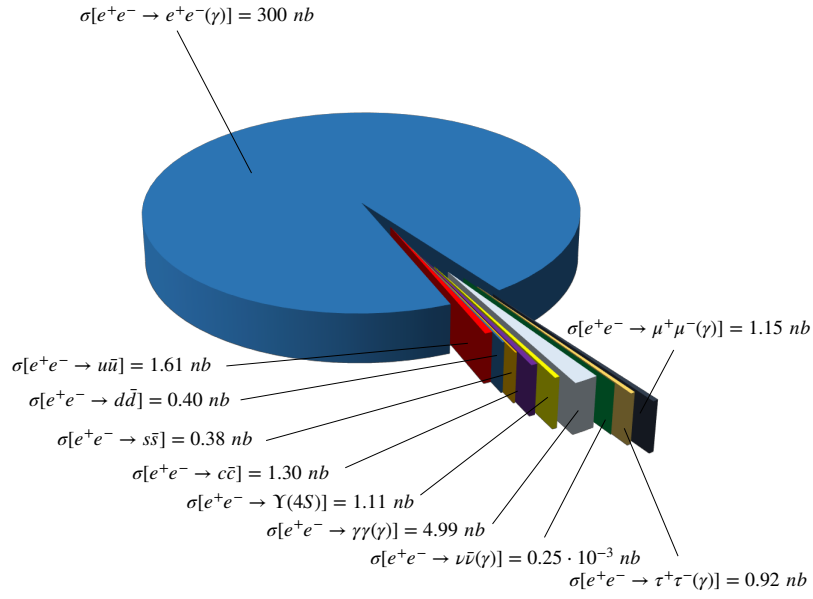


FIGURE 3.2: Cross-section of the most common final states produced by an e^+e^- collision at the $\Upsilon(4S)$ centre-of-mass energy [82].

where $\vec{p}_{e^+}$ ($\vec{p}_{e^-}$) is the three-momentum of the positron (electron). A schematic of SuperKEKB is shown in Figure 3.3.

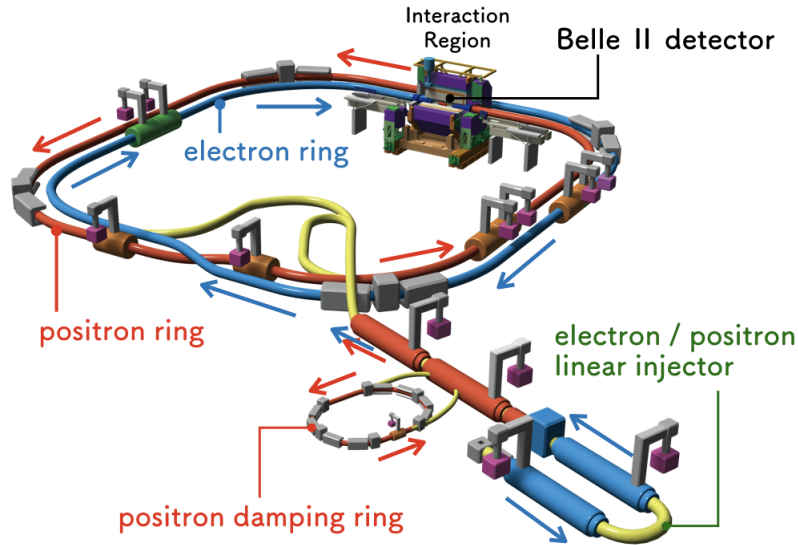


FIGURE 3.3: Schematic of the SuperKEKB accelerating facility. The Belle II detector is placed at the interaction point. Taken from Ref. [83].

In order to improve on the then world record instantaneous luminosity of $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ achieved by KEKB, the *nano-beam scheme* is applied at SuperKEKB. Let us define the luminosity L as [83]

$$L = \frac{\gamma_{\pm}}{2e r_e} \left(1 + \frac{\sigma_y^*}{\sigma_x^*} \right) \left(\frac{I_{\pm} \xi_{y\pm}}{\beta_y^*} \right) \left(\frac{R_L}{R_{\xi_y}} \right), \quad (3.2)$$

where the suffix $\pm$ specifies the positron (+) and electron (−) beams; $\gamma_{\pm}$, e and r_e are respectively the Lorentz factors, the elementary charge and the electron classical radius; R_L and R_{ξ_y} are geometrical correction factors; $\sigma_{x,y}^*$ are the horizontal and vertical beam sizes at the interaction point; $I_{\pm}$ are the beam currents; $\xi_{y\pm}$ are the vertical beam-beam tune-shift parameters; and β_y^* is the vertical beta function at the interaction point. To increase the luminosity, several of these beam parameters can be tuned but the nano-beam scheme focuses specifically on decreasing the beta function β_y^* . In head-on collisions, the size of β_y^* is limited by the bunch length σ_z due to the *hourglass effect* [24]. To mitigate this effect, the two beams collide at a large horizontal crossing angle $2\phi = 2 \times 41.5$ mrad in the nano-beam scheme. In that case, β_y^* is limited by the effective bunch length $d \approx \sigma_x^*/\phi$ and a larger crossing angle therefore leads to a higher instantaneous luminosity. The hourglass effect and the nano-beam scheme are illustrated in Figure 3.4.

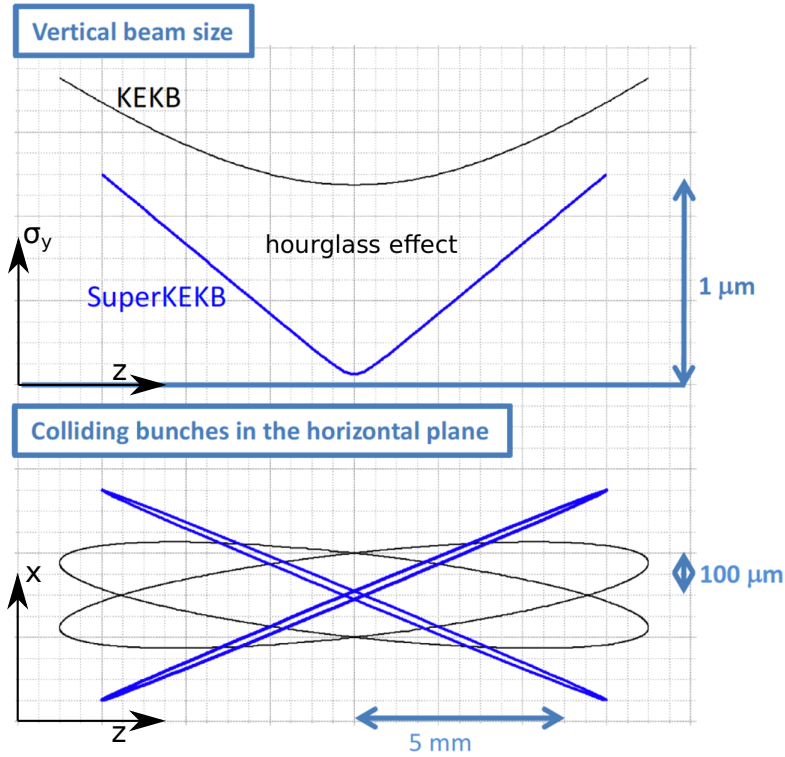


FIGURE 3.4: Vertical beam size comparison between KEKB and SuperKEKB (top) and bunch collision illustration in the horizontal plane (bottom).

3.3 The Belle II detector

The Belle II detector is a cylindrical detector consisting of various sub-detectors, each with a specific role in the reconstruction of the e^+e^- collision products. The three-dimensional Cartesian coordinate system used in the following is defined as follows: the z axis corresponds to the symmetry axis of the solenoid pointing in the direction of the electron beam, the y axis corresponds to the vertical axis, pointing upwards and the x axis, perpendicular to the y and z axes, points towards the outside of the accelerator ring. The $x - y$ plane is referred to as the *transverse plane*. Belle II has a large solid angle acceptance, close to 4π , and has the ability to reconstruct and identify both charged and neutral particles with a momentum spanning between 50 MeV and a few GeV. In the following, a *track* is defined as a charged particle trajectory as reconstructed from the successive energy deposits left by the particle in a region of the detector and a *cluster* is defined as a collection of responses from a group of neighbouring detector units

which can be produced by both charged and neutral particles. Charged particles which have a long enough lifetime to travel through large sections of the detector consist of $e^\pm$, $\mu^\pm$, $\pi^\pm$, $K^\pm$, $p^\pm$ and 2H (deuterons). Final state neutral particles consist of γ , n and K_L^0 . All other particles are reconstructed as intermediate resonances. Starting from the innermost layer, Belle II consists of the *pixel detector* (PXD), the *silicon vertex detector* (SVD), followed by the *central drift chamber* (CDC), the *time-of-propagation detector* in the barrel region (TOP), the *aerogel ring imaging Cherenkov counter* in the barrel and forward regions (ARICH), the *electromagnetic calorimeter* (ECL), and finally, the $K_L^0 - \mu$ system (KLM). The detector is illustrated in Figure 3.5. The properties of all sub-detectors are detailed in this section.

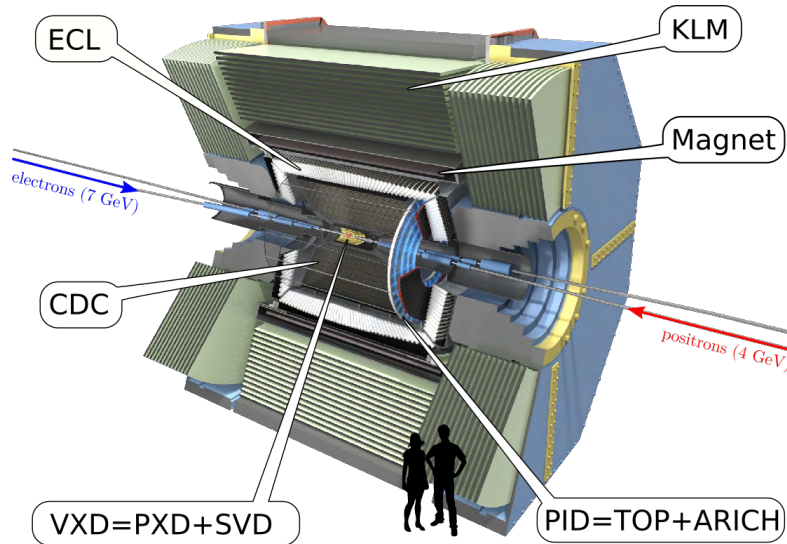


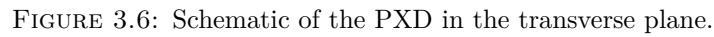
FIGURE 3.5: Illustration of the Belle II detector

3.3.1 The Pixel Detector (PXD)

The PXD [84] consists of two layers of silicon pixel *depleted p-type field-effect* (DEPFET) sensors positioned at 14 mm and 22 mm from the interaction point, respectively. This type of sensors is used in order to cope with the large levels of QED background produced by the beams. The sensors are kept thin in order to reduce multiple scattering effects. The two layers are designed to contain 8 and 12 ladders, respectively, with each ladder made up of two modules. Each module comprises 250×768 pixels. For the data used for the analysis presented in this thesis, only two ladders from the outer layer were employed to compensate for a dead module in the inner layer. The hit efficiency of most modules exceeds 98%, allowing for accurate vertex reconstruction with an average spatial resolution of approximately $14 \mu\text{m}$. A view of the PXD in the transverse plane is shown in Figure 3.6.

3.3.2 The Silicon Vertex Detector (SVD)

The SVD [85] is equipped with four layers of double-sided silicon-strip sensors positioned between 39 mm and 135 mm from the interaction point. These sensors operate by collecting the electron-hole pairs generated through ionisation as a charged particle passes through the sensor. Strips on the p-side and n-side collect the charges, and their perpendicular orientation allows for a precise determination of the particle's coordinates. The detector provides full azimuthal coverage and polar angle acceptance between 17° and 150° . In the forward direction, the sensors on the three outermost layers are trapezoidal and slanted towards the beam line. This helps improving the acceptance for particles travelling in the forward direction and at the same time reduces the



A detailed 3D cutaway rendering of the Large Hadron Collider (LHC) tunnel. The image shows the internal structure of the tunnel, including the superconducting magnets, the beam pipe, and the various support structures. The tunnel is shown in a perspective view, highlighting its length and the complex internal components. The rendering is in a light blue and grey color scheme, with some internal components highlighted in yellow and red. The tunnel is shown in a cutaway view, revealing the internal structure and the beam pipe. The beam pipe is a long, thin tube that runs through the center of the tunnel. It is surrounded by a series of superconducting magnets that are used to accelerate and steer the particle beams. The magnets are arranged in a series of dipole and quadrupole configurations. The tunnel is supported by a series of concrete pillars and a network of steel beams. The overall design is highly complex and precise, reflecting the advanced technology of the LHC.

3.3.3 The Central Drift Chamber (CDC)

The CDC [86] is the third component of the tracking system in Belle II. It operates by detecting charged particles through the ionisation of a gas mixture of helium and ethane ($\text{He-C}_2\text{H}_6$). The resulting electron avalanches are collected by 56 layers of wires, positioned between 168 mm and 1111 mm from the interaction point. The wires are arranged into two types: axial wires, which are parallel to the beam axis, and stereo wires, which are oriented at a skewed angle relative to the axial wires. These wires are grouped into axial and stereo superlayers, each containing either 6 or 8 wire layers. The alternating arrangement of superlayers enables the reconstruction of three-dimensional trajectories of particles. Charged tracks are bent in the $x - y$ plane by a magnetic field parallel to the beam axis and produced by a superconducting solenoid which

surrounds the ECL and generates a magnetic field of 1.5T. The superconductor is made of niobium and titanium (NbTi) and is cooled down by a liquid helium cryogenic system. The curvature R of a track is related to its transverse momentum p_t by [4]

$$p_t \cos \lambda = 0.3BR,$$

where B is the magnetic flux density and λ is the angle of the track with respect to the y -axis in the $y-z$ plane. The momentum of a charged particle can therefore be precisely determined from a three-dimensional reconstruction of the track it leaves in the CDC. The momentum relative uncertainty achieved by the detector is at the per mille level. The CDC can also measure the energy loss of a particle within the gas volume. A combination of the momentum and energy loss information collected by the CDC is used to reliably identify particles as further discussed below. The CDC detector and its wires are illustrated in Figure 3.8.

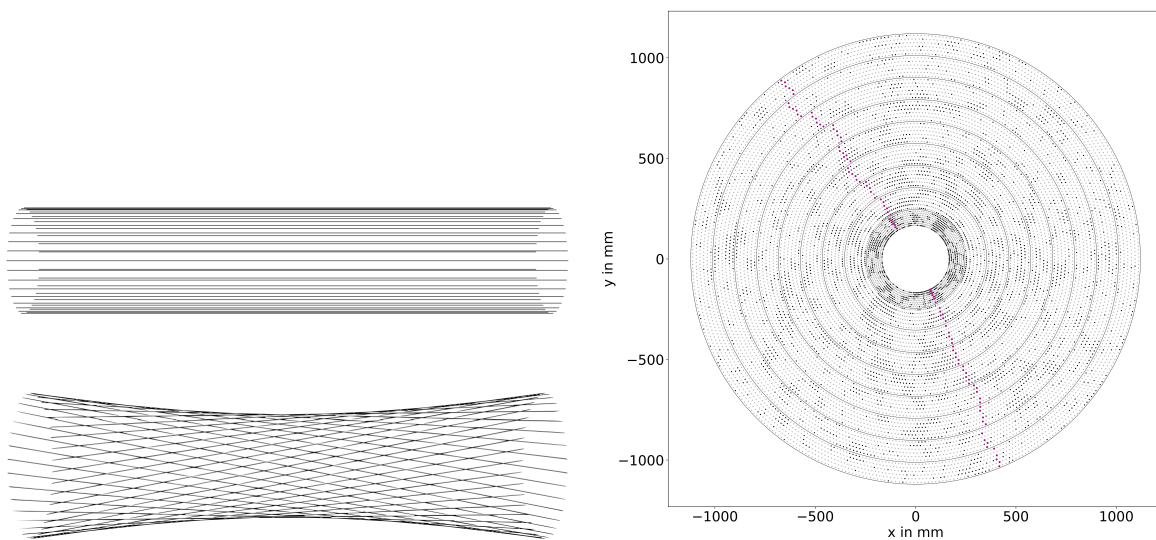


FIGURE 3.8: Illustration of the axial (top left) and stereo wires with an exaggerated skew (lower left). Illustration of the CDC in the transverse plane with hits from a simulated cosmic ray muon in purple and background hits in black (right).

3.3.4 Particle identification (ARICH and TOP)

The particle identification system of Belle II consists of two sub-detectors: the ARICH and the TOP.

Aerogel Ring Imaging Cherenkov Counter (ARICH)

The ARICH detector [87] is located at the forward end-cap, adjacent to the CDC. It is designed to differentiate between kaons and pions with momenta ranging from 400 MeV to 4 GeV, and between electrons, muons and pions with momenta below 1 GeV. The ARICH operates by utilising the Cherenkov effect: when a charged particle passes through its aerogel radiator with a velocity exceeding the speed of light in the material, it emits Cherenkov photons. The particle's velocity is inversely proportional to the cosine of the emission angle of these photons and by combining this with its momentum, the mass of the particle can be determined. The resulting Cherenkov photon patterns are also used to identify the particle type. The ARICH has a toroidal shape with a thickness of 280 mm, consisting of a 40 mm thick radiator made up of 248 silica aerogel tiles, a 160 mm expansion space, and an 80 mm photon detection system. It achieves a

kaon identification efficiency of 94%, with a pion misidentification rate of 11% [88]. The working principle of the ARICH is illustrated in Figure 3.9.

Time-of-Propagation detector (TOP)

The TOP detector [89] is positioned in the barrel region outside of the CDC and is designed to identify kaons and pions. It consists of 16 modules, each made up of two fused quartz bars. When a charged particle passes through, it emits photons via the Cherenkov effect described above. The photons are internally reflected within the bar, eventually reaching a photon detection plate at the bar's end. The time it takes for the photons to propagate in the material is related to their emission angle, which in turn is linked to the particle type. For kaons and pions with a momentum of 2 GeV, the typical time difference is 200 ps, while the TOP detector achieves a time resolution of about 35 ps. This results in a kaon identification efficiency of 85%, with a pion misidentification rate of 10% [90]. The working principle of the TOP is illustrated in Figure 3.9.

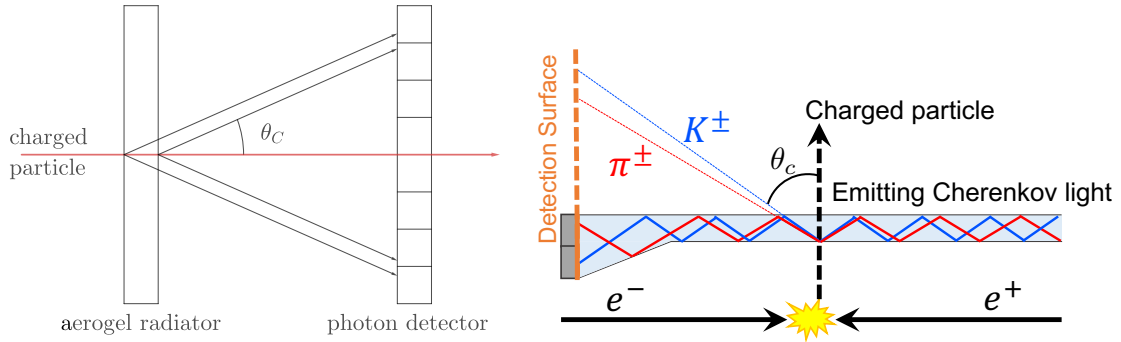


FIGURE 3.9: Illustration of the working principle of the ARICH (left) and TOP (right), the two sub-detectors of the Belle II particle identification system.

3.3.5 The Electromagnetic Calorimeter (ECL)

The ECL [91] is a key component of the detector system, responsible for energy measurement of particles, photon detection, electron identification, providing signals for the trigger, luminosity measurements, and K_L^0 detection in conjunction with the KLM. The ECL is divided into three sections: the barrel and two end caps (forward and backward), with a polar angle acceptance between 12.4° and 155.1° , and two 1° gaps between the barrel and the end caps. This setup covers 90% of the solid angle in the centre-of-mass frame. The barrel section contains 6,624 thallium-doped caesium iodide (CsI(Tl)) crystals, while the end caps consist of 2,112 pure caesium iodide (CsI) crystals. Each crystal is read out by two photodiodes that collect the low-energy photons generated by electromagnetic showers, converting them into an electronic signal. The ECL achieves a photon energy resolution ranging from 4% at 100 MeV to 1.6% at 8 GeV, and a π^0 mass resolution of 4.5 MeV [79]. The ECL configuration is illustrated in Figure 3.10.

3.3.6 K_L^0 and muon detection system (KLM)

The KLM detector [92] is designed to detect two types of long-lived particles, K_L^0 mesons and muons, which pass through the inner detectors with minimal interaction. Similarly to the ECL, the KLM is divided into the barrel and two end cap regions. It consists of alternating detector and iron layers: 14 iron layers, each 4.7 cm thick, and 15 detector layers (14 in the end caps). The dense iron layers provide 3.9 interaction lengths, enabling particles to interact within the detector, compared to the 0.8 interaction lengths provided by the ECL. The iron also serves as a

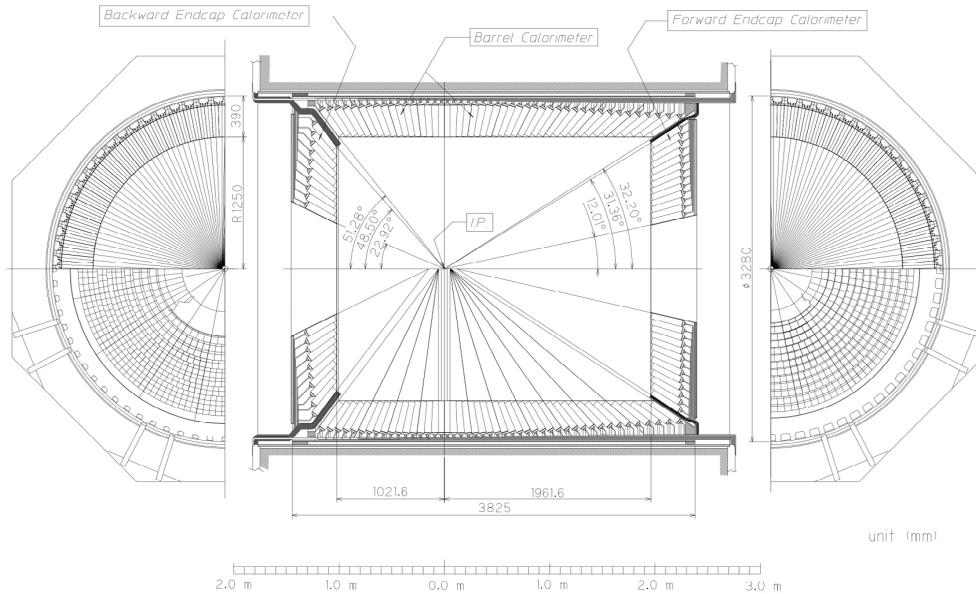


FIGURE 3.10: Schematic of the ECL configuration

return yoke for the solenoid. The KLM covers a polar angle acceptance from 20° to 155° . Muons with momentum above 600 MeV either pass through the KLM, leaving a nearly straight track, or stop due to electromagnetic interactions within the detector. K_L^0 mesons interact with either the ECL or KLM, producing hadronic showers that can be detected by one or both detectors. The KLM achieves a muon detection efficiency of 89% for momenta above 1 GeV, with a hadron fake rate of 1.3%, and the K_L^0 detection efficiency, which depends on the momentum, reaches a plateau of 80% at 3 GeV [93].

3.3.7 Trigger

The Belle II trigger system is designed to filter out background processes, particularly Bhabha scattering (see Figure 3.2), which typically produces only two tracks and/or clusters, whereas $B\bar{B}$ events exhibit higher particle multiplicities. For the data rate to stay below the maximum data acquisition rate of 30 kHz, the trigger system must efficiently reject background events while maintaining a high efficiency for $B\bar{B}$ events and low-multiplicity events of interest such as $e^+e^- \rightarrow \tau^+\tau^-$. The trigger system comprises two subsystems: the hardware-based *level 1* (L1) trigger and the software-based *high level trigger* (HLT). The L1 trigger utilises information from the CDC, TOP, ECL, and KLM to perform a fast partial reconstruction of events based on tracks and clusters, making decisions with a latency of $5 \mu\text{s}$. The HLT further refines event selection through full event reconstruction, using the Belle II reconstruction software and data from all detectors except the PXD. The specific trigger selection and performance will not be discussed further in this thesis as $B\bar{B}$ events have a trigger efficiency close to 100%.

3.3.8 Particle identification

To distinguish between different types of charged final-state particles, the information collected by all the subdetectors (except the PXD) can be used to assign each one of them an identification score [94]. Using the specific information of each subdetector, a likelihood $\mathcal{L}$ is computed for each of the six particle types e , μ , π , K , p and 2H . The individual subdetector likelihoods are then multiplied together in a global likelihood which is normalised to the sum of likelihoods of

each particle hypothesis, thus defining the particle identification score for particle i as

$$\mathcal{P}_i = \frac{\mathcal{L}_i}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_{2H}}. \quad (3.3)$$

The following information is used to compute the particle identification score:

- optic response in the TOP and ARICH
- particle ionisation energy loss and three-momentum magnitude in the CDC and SVD
- cluster energy from the ECL; the cluster energy over momentum magnitude ratio proves to be particularly powerful for distinguishing leptons from pions
- penetration range and scattering in the KLM to distinguish muons from K_L^0 ; the presence of a track associated with a KLM cluster also helps distinguish muons from neutral hadrons

In addition to the identification scores defined above, the measured particle variables can also be used as input features to multivariate classifiers such as boosted decision trees or neural networks (BDT and NN, *cf* section 4.2). A BDT was trained to separate between leptons and pions using their subdetector-specific likelihood differences and additional track and cluster information, such as track penetration depth and shower shapes, for instance. The output score of this BDT is used as the electron identification criteria in the analysis described in this manuscript whereas the likelihood-based score is preferred to identify muons (see section 5.3 for a detailed description of the reconstruction and selection procedure). Furthermore, a NN was trained to identify kaons and pions based on the information used to define their likelihood-based scores but the latter were preferred in the $B \rightarrow X_u \ell \nu$ measurement as it was noticed that they lead to lower misidentification rates relatively to the NN-based scores.

The only final state neutral particles of interest in our analysis are photons since K_L^0 are not considered, K_S^0 are reconstructed from pairs of $\pi^\pm$, and π^0 are reconstructed from pairs of γ . Photons are identified from clusters to which no track can be associated. In addition, photon clusters are expected to have a relatively symmetrical shape whereas clusters formed by charged and neutral hadrons tend to produce less confined showers and secondary split-off clusters. Photons from beam background radiation are likely to leave much lower energy deposits than photons resulting from hadron decays in $B\bar{B}$ events.

3.4 The Belle II software

The Belle II analysis software [95] (**basf2**) is a collection of algorithms used for various purposes such as the simulation of large numbers of events or the offline processing of simulated and collected data. In addition to the tools available in **basf2**, we also use in this thesis various external packages which will be mentioned explicitly.

3.4.1 Simulation

The simulation of large samples of events relies on Monte Carlo methods (*i.e.* repeated random sampling). As the scope of physics processes occurring at Belle II is wide, dedicated generator packages are used for specific types of decays. The main generators used in Belle II are as follows:

- **EvtGen** [96] is used to simulate $\Upsilon(4S) \rightarrow B\bar{B}$ events and the subsequent decays of B mesons as well as D mesons and other hadrons; for each possible B decay, a specific model implemented in **EvtGen** is chosen and the values of its input parameters are fixed,

- PYTHIA 8 [97] is used for quark hadronisation processes,
- KKMC [98] is used to simulate $\Upsilon(4S) \rightarrow q\bar{q}$ decays as well as $\Upsilon(4S) \rightarrow \tau^+\tau^-$ decays,
- TAUOLA [99] takes care of simulating τ decays,
- PHOTOS [100] is used to simulate photon final state radiation for all decays,
- BABAYAGA.NLO [101] simulates QED processes such as $e^+e^- \rightarrow e^+e^-(\gamma)$ and $e^+e^- \rightarrow \gamma\gamma$,
- PHOKHARA [102] is used to simulate specific $e^+e^- \rightarrow$ hadrons processes such as $e^+e^- \rightarrow \pi^+\pi^-(\pi^0)$,
- Finally, the detector response of all generated particles is simulated with GEANT4 [103].

Chapter 4

Analysis techniques and tools

“The main thing to master is the technique of work, and the understanding of subtleties will come later on its own.”

Lev Landau

To perform a measurement with the data collected by a high-energy physics experiment, different mathematical and computational tools are necessary. The tools used to perform the inclusive $B \rightarrow X_u \ell \nu$ measurement are described in this chapter. We start by presenting the maximum likelihood method used to perform fits to data before describing multivariate classifiers used to calibrate simulated data and reject backgrounds. We then present the Belle II Full Event Interpretation algorithm and finally, we introduce two software tools used to correct the modelling of resonant semileptonic B decays.

4.1 Maximum likelihood estimation

High-energy physics measurements rely on the comparison of samples of simulated and collected data. These discrete collections of events are described by distributions of measured observables. When comparing a distribution predicted by a theoretical model to the corresponding histogram of collected events to measure a given quantity, a statistical analysis of the samples is performed. The extraction of the chosen quantity from the observed data is commonly performed via the *maximum likelihood* (ML) method.

Let us define $f(x|\boldsymbol{\theta}) dx$ as the probability to find a random variable X in the interval $[x, x + dx]$ for a set $\boldsymbol{\theta}^1$ of m parameters. If an assumption is made on the functional form of $f(x|\boldsymbol{\theta})$, the *probability density function* (PDF), the values of the parameters $\boldsymbol{\theta}$ can be estimated from the *likelihood function* $\mathcal{L}$ defined for a set of n measured data points, $\{x_i\}$, $i = 1, \dots, n$. The likelihood quantifies the probability of observing this specific dataset under the assumed statistical model. If all x_i measurements are statistically independent and uncorrelated, the probability of measuring the dataset $\boldsymbol{x} = (x_1, \dots, x_n)$ is then simply the product of the individual probabilities

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

If $f(\boldsymbol{x}|\boldsymbol{\theta})$ is a good statistical model of the data, the total probability is expected to be close to unity. Similarly, well-chosen values of the parameters $\boldsymbol{\theta}$ would lead to a high likelihood. Therefore, the product of all $f(x_i|\boldsymbol{\theta})$ is maximised for the best set of values of $\boldsymbol{\theta}$, *i.e.* the set for which the measured data points $\boldsymbol{x}$ are the most probable. The ML method consists of finding

¹Vectors are indicated by bold symbols in this section.

the maximum of the likelihood function with respect to the m parameters $\boldsymbol{\theta}$ by solving the set of equations.

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_i} = 0, \quad i = 1, \dots, m. \quad (4.1)$$

Note that the logarithm of the likelihood, $\ln \mathcal{L}$, will often be considered instead of the likelihood itself. An important caveat in the ML method, especially for high-dimensional problems, comes from the fact that several solutions to Equation 4.1 could exist, each one corresponding to a local maximum. In any fit to collected data, it is crucial to avoid local maxima, which can be achieved by optimising the fitting strategy or by tuning the parameters of the fitting setup, such as the number of fit iterations or the convergence tolerance. So far, the set of measured points $\boldsymbol{x}$ was considered to have a fixed length but in a dataset collected by a particle physics experiment, the number of observations n is intrinsically random. If the probability to measure n points is governed by a Poisson² distribution with mean ν , an *extended likelihood function* can be defined as:

$$\mathcal{L}(\nu; \boldsymbol{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i | \boldsymbol{\theta}). \quad (4.2)$$

The dataset considered in the measurement presented in this manuscript is binned, meaning that the data points are grouped in intervals of a given variable (hence the use of histograms). In this case, the dataset reduces to a vector of N observed data counts $\boldsymbol{n} = (n_1, \dots, n_N)$ with expectation values $\boldsymbol{\nu} = E[\boldsymbol{n}]$. The n_i counts are usually assumed to be Poisson distributed and therefore the PDF can be described by a product of Poisson probability distributions:

$$f(\boldsymbol{n} | \boldsymbol{\theta}) = \prod_{i=1}^N \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} = \prod_{i=1}^N \text{Poiss}(n_i | \nu_i), \quad (4.3)$$

where the set of expected yields ν_i depends on the parameters $\boldsymbol{\theta}$.

In the context of a physics measurement, the parameters $\boldsymbol{\theta}$ of a statistical model can be separated into *parameters of interest* and *nuisance parameters*. The former define the measured parameters, whereas the latter represents the intrinsic uncertainties of the measurement. When extracting a physical observable from a simulated dataset by fitting it to collected data, the nuisance parameters can be seen as additional degrees of freedom of the simulated data based on predefined sources of uncertainty (theoretical modelling of a specific type of decays, limited detector performance, etc.). The parameters $\boldsymbol{\theta}$ can be constrained, labelled $\boldsymbol{\chi}$, based on prior³ knowledge of the measured dataset or left as *free parameters*, labelled $\boldsymbol{\eta}$. It is common to constrain nuisance parameters but not parameters of interest. For instance, if a theoretical input value is known up to some limited precision, the associated statistical model parameter is constrained to reflect the prior knowledge of the theoretical input. In the following, we will refer to prior knowledge as auxiliary data labelled $\boldsymbol{a}$.

4.1.1 The HistFactory/Pyhf model

The HistFactory package [104] was developed at CERN with the aim to perform binned template fits with C++ code. Its Python counterpart, called Pyhf [105], is the tool used to extract $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ and $|V_{ub}|$ in the analysis presented in this thesis. Based on the measurement of c disjoint distributions called channels, each with b bins, the HistFactory/Pyhf PDF is expressed as

$$f(\boldsymbol{n}, \boldsymbol{a} | \boldsymbol{\eta}, \boldsymbol{\chi}) = \prod_c \prod_b \text{Pois}(n_{cb} | \nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})) \prod_{\chi \in \boldsymbol{\chi}} c_{\chi}(a_{\chi} | \chi), \quad (4.4)$$

²Named after the French mathematician Siméon Denis Poisson and not after the French word for "fish".

³The Bayesian reader will appreciate such an expression.

where n_{cb} and ν_{cb} are respectively the measured and expected event counts in a specific channel and bin and c_χ are one-dimensional constraint terms with auxiliary data a_χ constraining the parameter χ . Each distribution is split into different samples, usually referred to as templates, which represent different types of events within a dataset ($B \rightarrow X_c \ell \nu$ and $B \rightarrow X_u \ell \nu$ events for example). In the fit procedure, the event rates are defined as

$$\nu_{cb} = \sum_{s \in \text{samples}} \nu_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s \in \text{samples}} \left(\prod_{\kappa \in \boldsymbol{\kappa}} \kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) \right) \cdot \left(\nu_{scb}^0(\boldsymbol{\eta}, \boldsymbol{\chi}) + \sum_{\Delta \in \boldsymbol{\Delta}} \Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) \right), \quad (4.5)$$

where ν_{scb}^0 is the nominal rate in a given bin and κ_{scb} and Δ_{scb} are multiplicative and additive modifiers respectively. Each source of uncertainty identified in a measurement is implemented in the statistical model as a nuisance parameter, which can modify the normalisation and/or shape of a template of the distribution. There exist different types of **Pyhf** modifiers, of which five are used for our measurement.

Global normalisation uncertainties of a template are implemented as multiplicative modifiers whose effect is computed by an interpolation between an up variation and a down variation given as input (*i.e.* the auxiliary data, for example normalisations of 0.8 and 1.1 for a given nuisance parameter). Details about the interpolation functions are given in Ref. [104].

Uncertainties on the shape of a template distribution are implemented via additive modifiers, which, similarly to normalisation modifiers, interpolate between up and down variations. In this case, the auxiliary data takes the form of binned distributions; for example, [55, 60] and [45, 52] for variations around a nominal bin count of [49, 54]. Normalisation and shape nuisance parameters are constrained by Gaussian functions centred on the nominal event rate with a width equal to the up/down variation. Effectively, such a constraint limits the variation of a template that is due to the nuisance parameter. In the case that the auxiliary data is not distributed symmetrically around the nominal event rate, the Gaussian is asymmetric which can lead to biases in the fit. One solution is to simply symmetrise the variations about the nominal rate by picking the largest of the two variations or by taking their average, for instance. Such a symmetrisation procedure was implemented in the fit setup of our measurement. More details about the specificities of our fit setup are given in Section 5.9.

Uncertainties due to the limited size of simulated samples are implemented via a separate type of modifiers, a set of bin-wise scale factors γ_b each constrained by a Gaussian with central value equal to unity and width equal to $\sqrt{\sum_s \delta_{csb}^2} / \sum_s \nu_{csb}^0$ where δ_{csb} is the absolute yield uncertainty in each bin.

Finally, overall and binned free-floating normalisation parameters can be implemented for each template. As they are not constrained, no auxiliary data is input in the statistical model. The different type of modifiers described here are summarised in Table 4.1.

In this thesis, **Pyhf** was used together with the **Cabinetry** package [106], which relies on the statistical inference methods implemented in **Pyhf** and provides additional tools to study the outcome of a fit. For instance, the agreement between simulated and experimental data yields after performing a fit, can be quantified by the *p-value* or *goodness-of-fit* calculated by **Cabinetry**. First, **Cabinetry** computes the difference between the logarithmic likelihood achieved in the fit and the logarithmic saturated likelihood *i.e.*, the maximum possible likelihood that can be achieved when the statistical model perfectly fits the data. Second, the number of degrees of freedom of the fit statistical model is calculated as the difference between the total number of

Description	Modification	Constraint term
Correlated shape (histosys)	$\Delta_{scb}(\alpha) = f_p(\alpha \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=+1})$	Gauss($a = 0 \alpha, \sigma = 1$)
Normalisation uncertainty (normsys)	$\kappa_{scb}(\alpha) = g_p(\alpha \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=+1})$	Gauss($a = 0 \alpha, \sigma = 1$)
MC statistical uncertainty (staterror)	$\kappa_{scb} = \gamma_b$	$\prod_b \text{Gauss}(a_{\gamma_b} = 1 \gamma_b, \delta_b)$
Normalisation (normfactor)	$\kappa_{scb} = \mu$	Free-floating
Data-driven shape (shapefactor)	$\kappa_{scb} = \gamma_b$	Free-floating

TABLE 4.1: Modifiers and constraints implemented in the **Pyhf** statistical model. The functions f_p and g_p are interpolation functions which depend on the interpolation parameter α .

bins and the number of free parameters⁴. The p -value is then computed from a χ^2 statistic test performed using the logarithmic likelihood difference and the number of degrees of freedom. We will use the p -value to assess the performance of each fit.

4.2 Multivariate classification

Machine learning is nowadays an integral part of high energy physics experiments. Various types of supervised and unsupervised machine learning algorithms are used for accelerator and detector tuning, particle detection, background rejection... In the measurement presented in this manuscript, machine learning is used to build multivariate classifiers. The goal of multivariate classifiers is to learn to distinguish between two or more classes of events by relying on input distributions referred to as features. The ability of multivariate classifiers to exploit complex high-dimensional connections between input features allows them to identify classes of events much more efficiently than more traditional sets of one-dimensional selections. Supervised multivariate classifiers are first trained with a (simulated) dataset in which the different classes of events are known. They are controlled by different hyperparameters which can be tuned to optimise the classifier performance and prevent it from learning too closely the particular properties of the training dataset, a phenomenon referred to as *overtraining* or *overfitting*. An overtrained model can perform poorly on a sample independent of the training sample and therefore, to reach the optimal performance, overtraining must be limited⁵. This is particularly important in the context of a physics measurement in which the classifier performance is evaluated on simulated samples and then applied on collected data samples. A different performance of the classifier on different samples would make the expectations built on simulated samples unreliable for data samples. After training, the performance of the classifier can be tested on an orthogonal set of events using various metrics. We present in this section the two types of multivariate classifiers used in the $B \rightarrow X_u \ell \nu$ analysis, boosted decision trees and neural networks. In the following, we will consider a set of n events, each associated with a set $\mathbf{x}_i$, $i = 1, \dots, n$ of input feature values and labelled $y_i \in \{0, 1, \dots\}$ based on the class in which the event falls. Effectively, the multivariate classifiers developed specifically for our measurement are binary classifiers and we therefore focus the conversation on these types of classifiers.

⁴The number of degrees of freedom should not be negative. As explained in detail in Section 5.9, in the fits performed to measure the $B \rightarrow X_u \ell \nu$ partial branching fractions, we implement N data-driven shape factors and one normalisation factor for $2N$ bins. The number of degrees of freedom of our fits is therefore always equal to $N - 1 > 0$.

⁵The importance of overtraining in high energy physics measurements has lead to numerous coffee break debates between the author and one of his colleagues whose opinion has been left out of this manuscript.

4.2.1 Boosted Decision Trees

A seemingly simple approach to perform a classification task is to use a decision tree (DT) in which the training dataset is iteratively split in two by a binary selection, typically the value of an input feature being larger or smaller than a fixed threshold. Each splitting point is called a node and starting from a single node, the number of nodes doubles with each layer. After the last set of nodes (called leaves), each event is assigned a weight $f(\mathbf{x}) \in (-\infty, +\infty)$ which can be mapped to a score $\hat{y}(\mathbf{x}) \in (0, 1)$ using, for example

$$\hat{y}(\mathbf{x}) = \frac{1}{1 + \exp(-f(\mathbf{x}))}.$$

In general, this score doesn't correspond to a probability but for a well-calibrated classifier, the output score will be close to the true probability of an event being in a specific class. A simple example of a DT is shown in Figure 4.1.

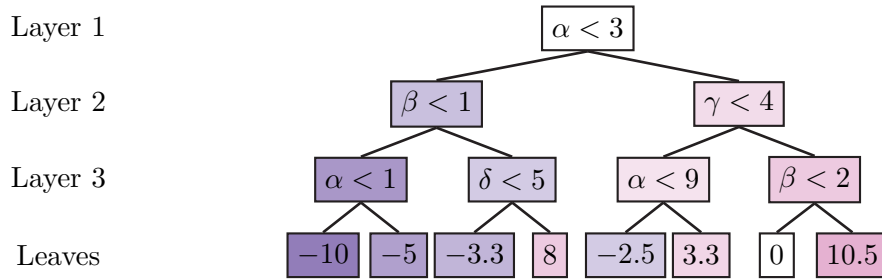


FIGURE 4.1: Example of a decision tree with three layers and four input features $\mathbf{x} = (\alpha, \beta, \gamma, \delta)$. Events in each leaf are assigned a weight $f(\mathbf{x})$. Adapted from Refs. [107] and [108].

Individual DTs suffer from a major drawback as they can easily learn on peculiar effects in the training sample which are not necessarily representative of the global properties of events in each class. In a single DT, a sufficiently large number of layers ensures an exact classification of any dataset with an arbitrary size. Therefore, to limit overtraining, a large number of shallow (typically 3 to 6 layers) DTs are usually grouped together in ensembles called *gradient-boosted decision trees* (BDT). The term *boosted* refers to the procedure of adding trees to the classifier as explained below. As mentioned above, for a set of input feature values $\mathbf{x}_i$ and a label y_i associated with a single event i among n events, each tree t in an ensemble of N trees is assigned a weight $f_t(\mathbf{x}_i)$ during training. The weights of all trees are summed in a global prediction weight

$$f_N(\mathbf{x}_i) = \sum_{t=1}^N f_t(\mathbf{x}_i)$$

which can be mapped to a score $\hat{y}_N(\mathbf{x}_i)$. The first tree in the ensemble is assigned a weight of 0 for all events. The classifier is then trained iteratively by minimising at each step t the function

$$\sum_{i=1}^n \ell(y_i, \hat{y}_t(\mathbf{x}_i)) = \sum_{i=1}^n \ell(y_i, \hat{y}_{t-1}(\mathbf{x}_i) + f_t(\mathbf{x}_i)), \quad (4.6)$$

where ℓ is the loss function, which is usually defined as the binary cross-entropy (BCE) function

$$\ell(y, \hat{y}) = -[y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})].$$

and where the contribution strength of each tree is quantified by the *learning rate*. The minimisation of Equation 4.6 is performed by computing the gradient of the loss function (hence the term gradient-boosted). In addition to the loss function, a regularisation term can be added in Equation 4.6 whose purpose is to penalise complex trees with large differences between the leaf weights and therefore reduce overtraining. Other strategies, often controlled by hyperparameters, can be added to a BDT algorithm to limit the overtraining. In our measurement, the BDT package `XGBoost` [109] is used for the continuum calibration procedure described in Section 5.4.4. The most important BDT hyperparameters are:

- the maximum depth of each decision tree,
- the initial learning rate of the model,
- the decay of the learning rate,
- α , the minimum reduction in loss required to further split a node of the tree; prevents overtraining,
- λ , γ , multiplicative factors (strength) of the regularisation terms added to the loss to prevent overtraining,
- the fraction of training events sampled to train each tree; prevents overtraining,
- the fraction of variables sampled in training each tree; prevents overtraining by forcing trees to rely on different sets of variables,
- the minimum number of training events required in each node after splitting a sample; prevents overtraining by avoiding extremely targeted selections,
- the grow policy, either *depth wise* in which case new nodes are added by dividing each node of a layer in two, or *loss guide*, in which case new nodes are added to a tree in order of maximum reduction in loss (which can result in branches of different lengths),
- the number of bins used to bin continuous features; binning such variables and training on the bin number speeds up the process and prevents overtraining

4.2.2 Neural Networks

Neural networks (NN) are multipurpose machine learning algorithms which can be very efficient for classification tasks. Their operating principle is complex⁶ and is thus presented shortly in the following. NN are simulations of biological webs of neurons, also called *perceptrons*. They represent a series of interconnected layers, each with a fixed number of neurons. In the context of classification, and using the same notation as for BDT, the input layer has as many nodes as there are features in the input vector $\mathbf{x}$ and, since the total output score sums to 1, the output layer has a single node for binary classification⁷. The number of layers and nodes per layer are hyperparameters of the NN and need to be tuned for optimal performance. An illustration of a simple two-layer NN is shown in Figure 4.2. Each neuron takes as input all or part of all values u_j from the previous layer and transforms them into an output value v_k such that

$$v_k = \sigma \left(\sum_j w_{jk} u_j + b_j \right), \quad (4.7)$$

⁶Physicists often refer to neural networks as "dark magic".

⁷There are n output nodes for n classes when $n \geq 3$

where the sum runs over a subset of nodes from the previous layer and w_{jk} are a set of weights for the layer k , b_j is a bias and σ is a non-linear activation function. The logistic function, the rectified linear unit or the hyperbolic tangent, for instance, can be chosen as activation functions. The process of computing an output value v_k at each node following Equation 4.7 (called *forward propagation*), is repeated for all events for each layer up to the last, outputting a prediction $\hat{y}_i$ of each event true class y_i . Performing the entire process only once would yield unreliable predictions of event classes and the NN is therefore trained by repeating the forward propagation using the entire available dataset, each training step being called an *epoch*. The training dataset is split into *batches* of fixed numbers of events and after each batch training, the learnable parameters of the NN are modified to minimise a predefined loss function $\ell(y_i, \hat{y}_i)$ in a similar way to BDT (this part of the process is called the *backward propagation*). The loss function is typically a BCE but more complex loss functions can be defined with additional terms whose purpose is, for example, to reduce overtraining or reduce the correlation between the output score and a physical variable. The loss is minimised by gradient descent with a step size quantified by the *learning rate*. In the analysis presented in this manuscript, NN are used to reject the two main types of background, continuum and $B \rightarrow X_{\ell}\nu$ events (*c.f.* Sections 5.5.1 and 5.5.2 for more details about these specific NN). The most important NN hyperparameters are:

- the number of training epochs,
- the number of layers in the network,
- the number of neurons in each layer of the network, which can be chosen to be different for each layer,
- the batch size,
- the activation function,
- the optimisation algorithm, which modifies the weights and biases of the NN,
- the learning rate of the model,
- the multiplicative factor (strength) of the weight decay regularisation term added to the loss - helps to prevent overtraining by penalising overly complex models,
- the dropout rate, which governs the fraction of randomly dropped input connections in dropout layers placed in between training layers - helps to prevent overtraining,
- the label smoothing factor s , which modifies the integer class labels (0 or 1 for binary classifiers) by adding or subtracting s ($0 + s$ and $1 - s$) - helps to prevent overtraining.

4.2.3 Hyperparameter optimisation

Learnable parameters of each multivariate classifier, such as the weights of a NN, are modified internally during the multivariate training. However, external parameters of the model - the hyperparameters - are set by the user before the training and their values must be tuned to optimise the performance of the multivariate. A fraction of the training dataset is often used as *validation sample* to find the optimal set of hyperparameters. The metric used for the optimisation is typically the output value of the model loss. The simplest approach for optimising is to use a grid search in which different sets of hyperparameter values are probed one by one. These types of optimisation processes become computationally expensive when the hyperparameter space is large and alternative techniques have therefore been developed. The **hyperopt** [110] and

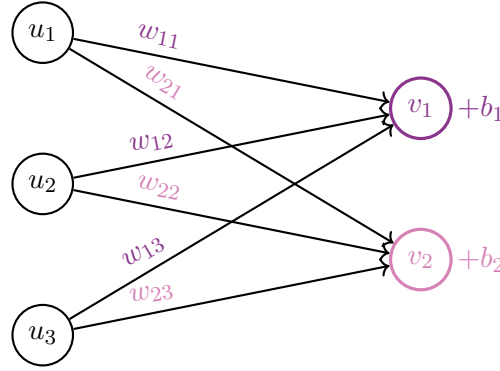


FIGURE 4.2: Example of a simple two-layer neural network with respectively three and two nodes per layer. The values u at each input node are modified by a weight w and a bias b , summed together and transformed to result in the value v at each output node. Refer to the text for details. By that point, the reader might have noticed that the author of this manuscript particularly likes pink and purple colours.

`optuna` [111] packages, which we respectively use to optimise the BDT and NN described in the next chapter, rely on Bayesian optimisation [112] (also known as sequential model-based optimisation). These optimisation algorithms are convenient for complex multivariate classifiers with a large number of hyperparameters as they implement parallelisation and pruning techniques to accelerate the hyperparameter tuning. Optimisation can be balanced against overtraining by predefining fixed ranges of values for the optimised hyperparameters or by adding terms in the loss which penalise specific behaviours.

4.2.4 Multivariate classifier metrics

Depending on the exact goal of the classifier, various metrics can be used to evaluate its performance. We outline in the following the metrics which are used for the classifiers developed for our $B \rightarrow X_u \ell \nu$ analysis.

Receiver operating characteristic curve

The receiver operating characteristic (ROC) curve is a plot of the false positive rate against the true positive rate for different classifier score thresholds. In the context of binary classifiers used to reject a specific type of background events, the ROC curve is equivalently defined as the background rejection rate against the signal efficiency. Examples of ROC curves are shown in Figures 5.11 and 5.15. For a random binary classification, the ROC curve is linear which indicates that, for any score threshold, a fixed fraction of events in each class pass the threshold. The area between the ROC curve and the linear curve can be used to quantify the classifier performance and is sometimes chosen as the criteria in hyperparameter optimisation. In addition, overtraining of the classifier can be detected by comparing the ROC curve of the training and test samples. The two curves are expected to match but the performance on the training sample can exceed that on the test sample in case of overtraining.

Variable importance

Once a multivariate classifier has been trained, the information extracted from all input features is combined in an output. It is nevertheless interesting to evaluate the individual impact of each input feature on the classifier outcome. For that purpose, we use the Shapley additive explanation (SHAP) values as implemented in the `Shap` package [113]. The SHAP values, based

on game theory concepts, quantify the importance of each input feature during training by decomposing the model output of each event as a sum of the contributions of each feature. For trees, SHAP values can be computed exactly [114], whereas for NN, the decomposed model must be approximated. Once the SHAP values of each event have been computed, they can be averaged over the full sample which results in a single value for the impact of each input feature. The SHAP values of the multivariate classifiers used in our measurement are shown in Figures 5.6, 5.12 and 5.16.

4.3 Full event interpretation

In B factories, the production of coherent pairs of charged and neutral B mesons from $\Upsilon(4S)$ decays, the clean environment and the well-known initial state of the decay chain allow for a precise reconstruction of both B mesons of each $B\bar{B}$ pair. The B decaying in the channel under study is labelled as signal B , B_{sig} , whereas the partner B is referred to as tag B , B_{tag} . The kinematic properties of the B_{sig} decay products can be constrained by the reconstruction of the B_{tag} via a tagging method. At Belle II, two such methods exist, the *semileptonic* and *hadronic* tagging strategies.

Reconstructing the B_{tag} via its hadronic decays allows one to achieve relatively large signal purities at the cost of a low overall efficiency $\mathcal{O}(0.1\%)$. Since the B_{tag} can be fully reconstructed, the signal side kinematics can be inferred, which is particularly useful for measurements of channels involving undetected neutrinos and/or systems on which no explicit requirements are placed (called *inclusive* systems). The $B \rightarrow X_u \ell \nu$ measurement presented in this manuscript is an example of a channel including both a neutrino and an inclusive object. The semileptonic tagging approach yields a higher reconstruction efficiency $\mathcal{O}(1\%)$ but in that case, the available kinematic information on the tag-side is limited because of the presence of a missing neutrino. A third method, the *inclusive* tagging strategy in which the B_{tag} is in fact not reconstructed and the collected information on the tag side of the event is simply required to be consistent with that of a B meson decay, can also be used. While this method suffers from a limited knowledge of the event kinematics, it offers a reconstruction efficiency close to 100%. Inclusive tagging is therefore the preferred method for measurements of very rare B decays. As highlighted in this paragraph, all three tagging procedures have advantages and disadvantages and it is often useful to investigate which one offers the highest sensitivity for a specific signal channel.

The hadronic tagging method is performed by the full event interpretation (FEI) algorithm [115] implemented in the Belle II software. The FEI algorithm reconstructs B mesons following a hierarchical six-stage approach as illustrated in Figure 4.3. First, reconstructed objects such as tracks, vertices and clusters are used to identify e , μ , $K^\pm$, $\pi^\pm$, p and γ ⁸ which are then combined to form intermediate mesons (π^0 , J/ψ , K_S^0 , D and D^*) and baryons (Λ , Λ_c and Σ) before finally reconstructing the B meson. A total of $\mathcal{O}(10,000)$ distinct decay channels are included in the FEI algorithm. At each stage of the FEI reconstruction, the kinematic and vertex information of particles and their daughters is used by a BDT to determine the reconstruction quality of the particle. The resulting network of BDT, trained beforehand on simulated data, assigns to each event an overall output score $\mathcal{P}_{\text{FEI}}$. An algorithm similar to FEI was used in the Belle experiment. The Belle II FEI algorithm achieves higher reconstruction efficiencies than the Belle Full Reconstruction (FR) algorithm [116], in particular at low purity, with a maximum efficiency about 2.5 times larger. The improvement can mainly be explained by the larger number of channels reconstructed by the FEI compared to the FR and by the updated selection criteria. As detailed in the next chapter, a selection on $\mathcal{P}_{\text{FEI}}$ is applied on signal event candidates during the

⁸The algorithm also reconstructs K_L^0 but these are not used in this analysis.

$B \rightarrow X_u \ell \nu$ reconstruction procedure and FEI efficiency calibration factors are used to correct the simulated data.

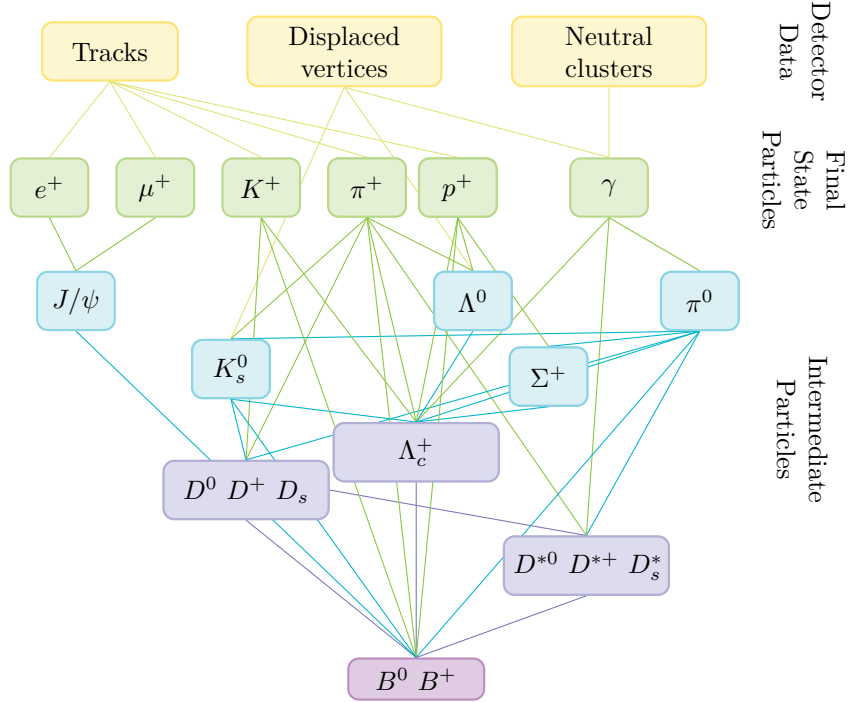


FIGURE 4.3: Schematic of the hierarchical reconstruction of the FEI algorithm. Only the channels used for the hadronic tagging approach are illustrated in this diagram. The six-stage reconstruction is illustrated in the neutral cluster $\rightarrow \gamma \rightarrow \pi^0 \rightarrow K_s^0 \rightarrow \Lambda_c^+ \rightarrow B$ reconstruction chain for example. Adapted from Ref. [115].

4.4 Form factor reweighting

Resonant semileptonic B meson decays are described by form factors as discussed in length in Section 2.3. These theoretical models can be implemented in event generators such as **EvtGen** [96] to simulate large numbers of decays. The form factor models depend on input parameters which cannot be derived from first principles and must therefore be extracted from lattice QCD or experimental data. Since the simulated event samples and distributions were generated with a fixed set of form factor parameters, two important questions arise. First, without having to generate a new set of events, how can the model parameters or the model itself be updated? Second, how can model parameter uncertainties be taken into account when performing a measurement? Both of these questions are answered by *form factor reweighting*⁹ tools. Based on differential decay rate predictions and event kinematics, these algorithms compute weights which are used to scale simulated distributions such that they match an updated or varied model. We present in the following two form factor reweighting tools, **eFFORT** [117] and **Hammer** [118].

4.4.1 eFFORT

eFFORT (from **Form Factor Reweighting Tool**¹⁰) is a flexible package which allows the computation of integrated and differential decay rates based on specific form factor models. Given

⁹Even though some native English speakers would disagree, the word "reweighting" will be used throughout this thesis instead of "reweighing".

¹⁰Not the greatest acronym physicists have come up with but certainly not the worst.

a decay type and a momentum transfer squared q^2 (or w) value, the differential decay rate, normalised to the total decay rate, is computed for the model used to generate an event and for the updated model or varied set of parameters. The ratio between the two predictions, therefore, gives the factor by which the event is reweighted. In the measurement presented in this manuscript, **eFFORT** was used to reweight all resonant $B \rightarrow X_u \ell \nu$ modes *i.e.* decays to π , ρ , ω , η and η' . Even though ρ and ω are vector mesons, the package only considers one-dimensional differential decay rates. An updated version of the package, **eFFORT2** [119], allows the computation of four-dimensional decay rates for $B \rightarrow \rho \ell \nu$ and $B \rightarrow \omega \ell \nu$ by factoring in the angular information of the leptonic system and the ρ/ω decay. However, since only a small difference is expected between these two approaches, only **eFFORT** was used in the inclusive $B \rightarrow X_u \ell \nu$ analysis. An example of the differential decay rate as a function of q^2 computed by **eFFORT** for $B^+ \rightarrow \pi^0 \ell^+ \nu$ decays is shown in Figure 4.4 for 50 reference points.

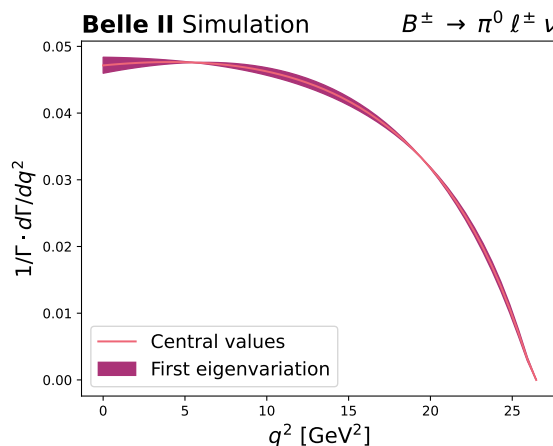


FIGURE 4.4: Normalised differential decay rate as a function of q^2 for $B^+ \rightarrow \pi^0 \ell^+ \nu$ decays with the BCL model and form factor parameter values from Ref. [120] as calculated with **eFFORT**. The variation band due to the first up/down eigenvariations is also shown.

4.4.2 Hammer

Hammer (for **H**elicity **A**mplitude **M**odule for **M**atrix **E**lement **R**eweighting¹¹) works on a similar principle as **eFFORT**: the total and differential decay rates are computed for each event and the ratio between two model predictions represents the weight assigned to that event. In the $B \rightarrow X_u \ell \nu$ measurement, **Hammer** is used to reweight all resonant $B \rightarrow X_c \ell \nu$ decays. One of the key features of **Hammer** is that it computes the amplitudes of the charm meson decays, therefore making use of the entire kinematic information of every event. For instance, in a $B \rightarrow D_1 \ell \nu$ decay, the $D_1 \rightarrow D^*(\rightarrow D\pi)\pi$ decay is computed using a partial-wave parametrisation. **Hammer** is computationally extremely efficient as it expresses decay rates as functions of form-factor-generalised objects, which need only be defined once for every event and can then be contracted with form factor parameters. This procedure is particularly useful when calculating weights for parameter variations. More details about **Hammer**'s design are given in its manual [121].

One flaw of both **eFFORT** and **Hammer** comes from the fact that they compute decay rates assuming fixed meson masses. This doesn't affect the calculation for narrow resonances such as π or D^* but it could potentially play a non-negligible role for broader resonances, in particular D_0^* and D_1' . This is however a wider theoretical problem as fixed mass values are already assumed

¹¹Probably one of the greatest acronyms physicists have come up with.

in form factor models. See, for example, Ref. [122] for an investigation of $B \rightarrow D^{**} \ell \nu$ lineshape modelling.

Chapter 5

Measurement of inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions and $|V_{ub}|$

“If I could remember the names of all these particles, I’d be a botanist.”

Enrico Fermi

In this chapter, we describe in detail the measurement of inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ and $|V_{ub}|$ in different regions of phase-space using 365 fb^{-1} of data collected by the Belle II detector. After giving an overview of the analysis, we present the data and simulated data samples used for the measurement, the event reconstruction and selection and the corrections applied to simulated samples. We then discuss the agreement between data and simulated data and the signal extraction procedure before presenting the results of the measurement.

5.1 Analysis overview

The measurement of inclusive $B \rightarrow X_u \ell \nu$ branching fractions is challenging for several reasons. First of all, as was described in Section 2.3.5, the theoretical description of inclusive $B \rightarrow X_u \ell \nu$ decays necessitates the introduction of specific models, each leading to different predictions of the $B \rightarrow X_u \ell \nu$ decay rate and $|V_{ub}|$ and yielding different uncertainties. To add to the complexity of the signal modelling, the non-resonant decays must be combined with resonant $B \rightarrow X_u \ell \nu$ decays because inclusive predictions cannot model exclusive decays. The combination is done via the *hybrid* approach described later in this chapter. Second, the presence of an undetected neutrino in the final state and the requirement to reconstruct the hadronic system inclusively limit the precision of the event reconstruction. Finally, the overwhelming background coming mainly from $B \rightarrow X_c \ell \nu$ decays makes necessary the application of tight kinematic selections. Such selections reduce the phase-space available for the measurement and we therefore measure partial branching fractions $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ in various phase-space regions. With the usage of machine learning techniques, however, it is possible to achieve relatively high signal purities even in broader phase-space regions.

We illustrate in Figure 5.1 the differences between $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ decays in the three main kinematic variables used to describe semileptonic decays: the lepton energy in the B rest frame, E_ℓ^B , the hadronic system mass, M_X , and the momentum transfer, $q^2 = (p_\ell + p_\nu)^2$ where p_ℓ and p_ν are the lepton and neutrino four-momenta, respectively. A requirement that E_ℓ^B is greater than 1 GeV is always applied as it allows us to suppress various sources of backgrounds such as $e^+e^- \rightarrow q\bar{q}$ events ($q \in \{u, d, s, c\}$). We also measure partial branching fractions in phase-space regions where M_X is required to be less than 1.7 GeV and q^2 is greater than 8 GeV^2 . As

shown in Figure 5.1, these selections are particularly effective at suppressing $B \rightarrow X_c \ell \nu$ events while retaining large fractions of $B \rightarrow X_u \ell \nu$ events. Decays used for Figure 5.1 have undergone the full detector response simulation. See Figure 2.9 for decays simulated without the detector response. However, in tighter phase-space regions, the models used to simulate $B \rightarrow X_u \ell \nu$ decays become more sensitive to a non-perturbative function called the *shape function*, which describes the b quark motion inside the meson (see Section 2.3.5 for a detailed discussion of these models) and to the b -quark mass, m_b . Measurements of $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ and $|V_{ub}|$ in tight regions are therefore more dependent on the particular treatment of the shape function of a given model than measurements over broad regions.

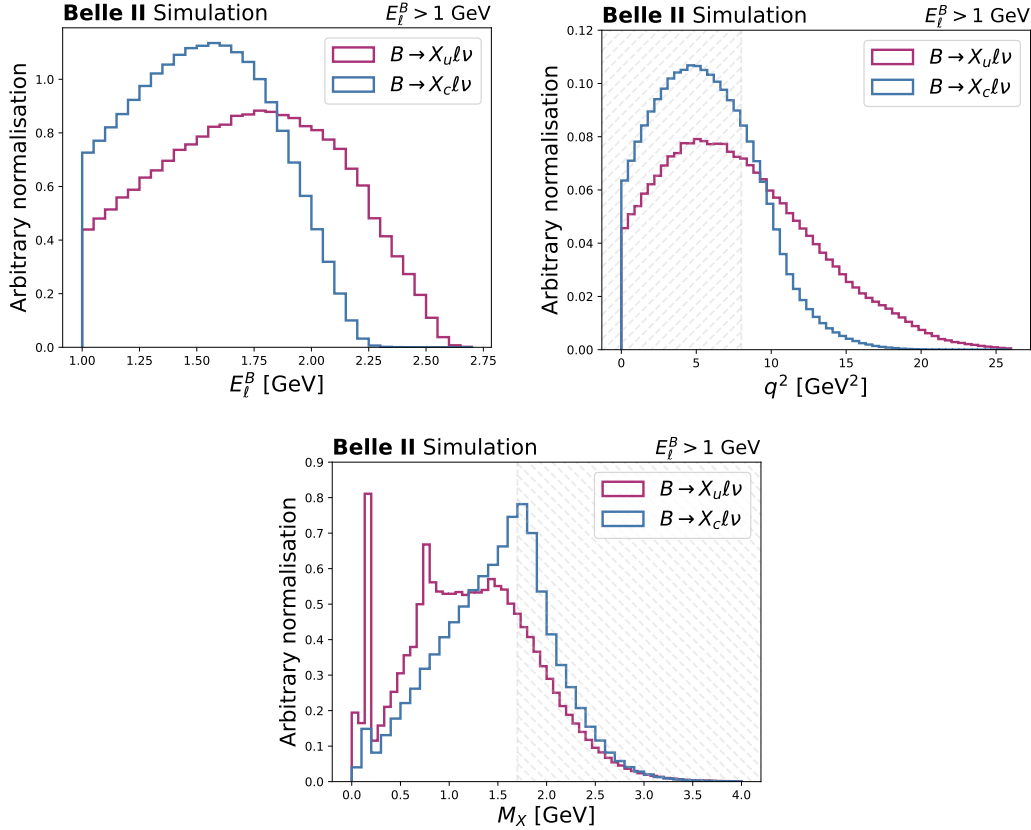


FIGURE 5.1: Comparison between $B \rightarrow X_c \ell \nu$ (blue) and $B \rightarrow X_u \ell \nu$ (purple) decays in reconstruction-level distributions of E_ℓ^B (top left), q^2 (top right) and M_X (bottom). The shaded regions are excluded in some of the measurements. The two histograms are normalised to the same area. The simulated samples used to plot the distributions are described in Section 5.2. Events are reconstructed as described in Section 5.3. A selection of $E_\ell^B > 1$ GeV is applied.

Since in e^+e^- collisions at Belle II, B mesons are produced in pairs, when one of the two B mesons is reconstructed in the signal decay channel, the partner B can also be reconstructed. This meson is referred to as *tag-side* B or B_{tag} . The reconstruction of the B_{tag} helps constrain the kinematics of the event and reduce background contributions. In the analysis described here, the B_{tag} is reconstructed in its hadronic decay channels using the Full Event Interpretation (FEI) algorithm [115] described in Section 4.3. On the signal side, the lepton is reconstructed as a single track identified as an electron or a muon. $B \rightarrow X_u \tau \nu$ events are not considered as their reconstruction wouldn't be as precise as for the e and μ channels. The neutrino is identified with the missing energy of the event and finally, all tracks and clusters not associated with the

B_{tag} or the signal lepton are grouped to define an object called the *rest-of-event* (ROE), which is identified as the inclusive hadronic system X_u . A typical decay of interest is illustrated in Figure 5.2.

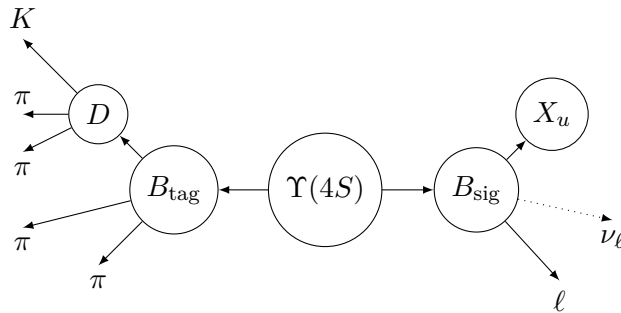


FIGURE 5.2: Illustration of a $B \rightarrow X_u \ell \nu$ decay on the signal side and a B hadronic decay on the tag side, which is reconstructed by the FEI.

5.2 Data and simulated data samples

5.2.1 The Belle II dataset

The measurement uses the Belle II dataset collected between March 2019 and June 2022 at a centre-of-mass energy of $\sqrt{s} = 10.58$ GeV, which corresponds to the mass of the $\Upsilon(4S)$ resonance. This sample of $365.37 \pm 1.70 \text{ fb}^{-1}$ [123] is referred to as the *on-resonance* sample. In addition, we use a dataset collected at a centre-of-mass energy 60 MeV below the $\Upsilon(4S)$ resonance mass for background studies and calibration. This dataset is referred to as *off-resonance* sample and corresponds to $42.74 \pm 0.20 \text{ fb}^{-1}$.

5.2.2 Simulated samples

Monte Carlo (MC) simulated samples¹ are used to analyse the behaviour of different types of decays and the response of the detectors. The expectations obtained from the simulated samples are compared to the Belle II dataset to extract the $B \rightarrow X_u \ell \nu$ branching fraction. The generated physical processes can be studied via *generator level* quantities, which are smeared by the detector response simulation, resulting in *reconstruction level* quantities. The latter can therefore be compared to data distributions. The measurement described in this thesis is performed via a *blind* or *closed box* analysis technique, meaning that the signal extraction procedure is first prepared on simulated samples before using the collected dataset. With this strategy, the introduction of unconscious biases orienting the measurement towards prior expectations is avoided. We use the equivalent of 2.8 ab^{-1} of $e^+e^- \rightarrow B\bar{B}$ events where the B mesons decay generically. In addition, for specific signal and background decays, we generate dedicated samples based on the theoretical models described in Section 2.3.

$B \rightarrow X_u \ell \nu$ samples

The sample of $B \rightarrow X_u \ell \nu$ events is a mixture of non-resonant and resonant modes. Non-resonant models give predictions for the total inclusive decay rate but cannot describe resonant decays. The kinematics of inclusive decays are very different from the kinematics of resonant ones, where

¹Monte Carlo event generation relies on random sampling to model the probabilistic character of particle interactions. For collider experiments, all physical processes from the initial collision to the detector interaction are simulated via MC generators. The name Monte Carlo originates from the famous casino in Monaco. In the following, simulated data will sometimes be simply referred to as MC.

the hadronic system mass is expected to peak at the known meson mass. The two types of events are therefore combined using the *hybrid* approach first proposed in Ref. [124]. After presenting the samples of $B \rightarrow X_u \ell \nu$ decays, the hybrid model is described.

The simulation of non-resonant and resonant decays is performed using the **EvtGen** generator [96]. For the production of the nominal sample of non-resonant events, the De Fazio, Neubert (DFN) model is used with input parameters $m_b^{\text{KN}} = 4.66$ GeV and $a^{\text{KN}} = 1.3$ [125]. In addition, we produce a sample of non-resonant events with the Bosch, Lange, Neubert, Paz (BLNP) framework to extract an uncertainty related to the modelling of such events. See Appendix A for a discussion about the choice of the non-resonant $B \rightarrow X_u \ell \nu$ theoretical model. Samples of resonant decays are produced using specific form factor models. The sample of $B \rightarrow \pi \ell \nu$ events is modelled following the expansion of Bourrely, Caprini, Lellouch (BCL), truncated at the fifth term from Table 57 in Ref. [39]. The samples of $B \rightarrow \rho \ell \nu$ and $B \rightarrow \omega \ell \nu$ events are modelled following the Bharucha, Straub, Zwicky (BSZ) expansion with input parameter values from Table 4 in Ref. [40]. The samples of $B \rightarrow \eta \ell \nu$ and $B \rightarrow \eta' \ell \nu$ events are modelled via the QCD light cone sum rule (LCSR) approach with input parameters taken from Ref. [42] (Equations 5.2 and 6.3). As discussed in more detail in Section 5.4.2, the form factor parameters are updated to the values given here using the **eFFORT** package (see Section 4.4.1). The branching fractions used for the modelling of resonant $B \rightarrow X_u \ell \nu$ decays are taken from the most up-to-date Heavy Flavour Averaging Group (HFLAV) report [63]. The non-resonant $B \rightarrow X_u \ell \nu$ branching fractions are taken separately for B^0 and B^+ as quoted in the latest Particle Data Group (PDG) review [24]. The assumed branching fractions of all $B \rightarrow X_u \ell \nu$ decays are summarised in Table 5.1.

Decay mode	$\mathcal{B} [\times 10^{-4}]$	
	B^+	B^0
$B \rightarrow \pi \ell \nu$	0.78 ± 0.03	1.50 ± 0.06
$B \rightarrow \rho \ell \nu$	1.58 ± 0.11	2.94 ± 0.21
$B \rightarrow \omega \ell \nu$	1.19 ± 0.09	-
$B \rightarrow \eta \ell \nu$	0.35 ± 0.04	-
$B \rightarrow \eta' \ell \nu$	0.24 ± 0.07	-
$B \rightarrow X_u \ell \nu$	19.2 ± 2.4	17.6 ± 2.2

TABLE 5.1: Assumed branching fractions for all $B \rightarrow X_u \ell \nu$ channels.

Hybrid modelling

The idea of the hybrid model is to sum the resonant and non-resonant channels, which are scaled down such that the total inclusive $B \rightarrow X_u \ell \nu$ branching fraction matches the non-resonant prediction. This rescaling is done by applying weights to non-resonant events in three-dimensional bins of q^2 , E_ℓ^B and M_X such that:

$$w_{i,j,k} = \frac{\Delta \mathcal{B}_{ijk}^{\text{incl}} - \Delta \mathcal{B}_{ijk}^{\text{res}}}{\Delta \mathcal{B}_{ijk}^{\text{incl}}}, \quad (5.1)$$

where $\Delta \mathcal{B}_{ijk}^{\text{incl}}$ and $\Delta \mathcal{B}_{ijk}^{\text{res}}$ are respectively the inclusive and resonant partial branching fractions in a given hybrid bin ijk of the three-dimensional phase-space. Effectively, since the hybrid weights are computed from simulated samples of events, the partial branching fractions correspond to event yields in each bin of the generator level distributions of q^2 , E_ℓ^B and M_X . The binning

used for the hybrid reweighting is defined by the combinations of the one-dimensional bins:

$$\begin{aligned} q^2 &: [0, 2.5, 5, 7.5, 10, 12.5, 15, 20, 30] \text{ GeV}^2, \\ E_\ell^B &: [0, 0.5, 1, 1.25, 1.5, 1.75, 2, 2.25, 3] \text{ GeV}, \text{ and} \\ M_X &: [0, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5] \text{ GeV}. \end{aligned}$$

The hybrid binning is tuned to avoid large negative weights. For this reason, in the low M_X region, which is populated by most X_u resonances, a single broad bin is defined. The resulting q^2 , E_ℓ^B and M_X distributions are illustrated in Figure 5.3 for the charged B channel. The kinks that can be seen in the hybrid model distribution appear at the edges of bins with a relatively large fraction of resonant events. These kinks are smeared out by the limited detector resolution, as can be seen from Figure 5.38 for instance.

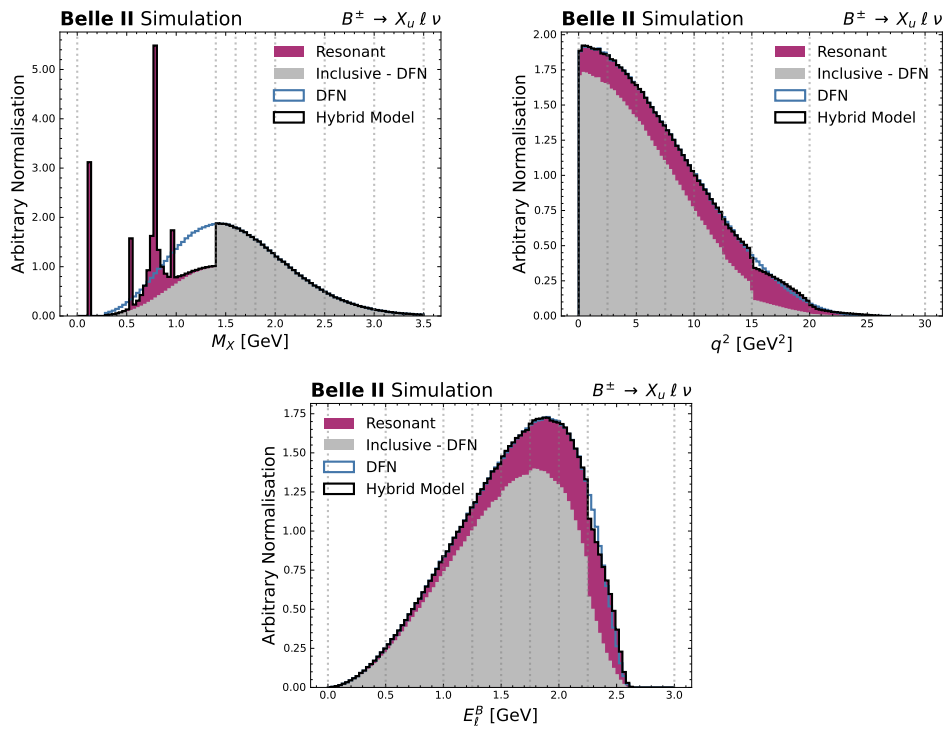


FIGURE 5.3: Projections of the hybrid MC distributions M_X , q^2 and E_ℓ^B in the $B^+ \rightarrow X_u \ell \nu$ channel. The grey vertical dotted lines show the boundaries of the bins used for hybrid weights calculations from Equation 5.2.2. A finer binning is used to illustrate the hybrid procedure.

$B \rightarrow X_c \ell \nu$ samples

Three types of resonant $B \rightarrow X_c \ell \nu$ decay modes are considered: $B \rightarrow D \ell \nu$, $B \rightarrow D^* \ell \nu$ and $B \rightarrow D^{**} \ell \nu$. These decays are produced in the generic $e^+e^- \rightarrow B\bar{B}$ sample in which the $B \rightarrow D \ell \nu$ and $B \rightarrow D^* \ell \nu$ decays are simulated using the Boyd, Grinstein, Lebed (BGL) model. Using the **Hammer** package described in Section 4.4.2, we reweight these events to match the more up-to-date Bernlochner, Ligeti, Papucci, Robinson, Xiong, Prim (BLPRXP) theoretical model with input parameter values taken from the first column in Table VII in Ref. [36]. The $B \rightarrow D^{**} \ell \nu$ decays are modelled with the Bernlochner, Ligeti, Robinson (BLR) model. **Hammer** is used to update the input parameters from their default values to those from Table V in Ref. [37].

The $B \rightarrow D \ell \nu$, $B \rightarrow D^* \ell \nu$ and $B \rightarrow D^{**} \ell \nu$ decay rates respectively represent about 21%,

48% and 12% of the inclusive $B \rightarrow X_c \ell \nu$ rate and therefore cover only 81% of the rate. Hence, other known multi-hadronic decay channels must also be considered. The $B^+ \rightarrow D_s^{(*)} K \ell \nu$ decays have been measured by the BaBar and Belle collaborations [126, 127]. They are also produced in the generic $e^+e^- \rightarrow B\bar{B}$ sample and account for 0.5% of the inclusive decay rate. The $B \rightarrow D^{(*)}\pi\pi\ell\nu$ channels were also observed by BaBar and more recently by Belle [128, 129]. However, the $B \rightarrow D\pi\pi\ell\nu$ channel is almost entirely saturated by the resonant decay $B \rightarrow D_1(\rightarrow D\pi\pi)\ell\nu$. In total, the non-resonant $B \rightarrow D^{(*)}\pi\pi\ell\nu$ channels account for about 2.5% of the inclusive decay rate. Dedicated samples of $B \rightarrow D^{(*)}\pi\pi\ell\nu$ events are produced. No specific model exists for these types of decays so they are simulated using an intermediate D^{**} resonance with the BLR model and the same input parameters as for the nominal $B \rightarrow D^{**}\ell\nu$ modes. They are equally split between the $B \rightarrow D_1'(\rightarrow D^{(*)}\pi\pi)\ell\nu$ and $B \rightarrow D_0^*(\rightarrow D^{(*)}\pi\pi)\ell\nu$ channels. It is not exactly clear what kind of decays fill the remainder of the gap between the sum of exclusive decay rates and the inclusive rate. From past measurements, it is known that $B \rightarrow D^{(*)}\pi\ell\nu$ decay modes are saturated by decays via D^{**} resonances. Various ideas were proposed to solve the *gap puzzle* [130] but the method employed in semileptonic analyses in Belle II consists in adding dedicated $B \rightarrow D^{(*)}\eta\ell\nu$ decay samples. The branching fractions of these decay modes are chosen such that they fill the gap. They are equally split between the D_0^* and D_1^* channels and the events are simulated with the BLR model. Since these decays have never been observed and since it is in fact not known whether they can fill the entire gap, a relative uncertainty of 100% is assigned to their branching fractions. The $B \rightarrow D^{(*)}\pi\pi\ell\nu$ and $B \rightarrow D^{(*)}\eta\ell\nu$ channels will collectively be referred to as *gap modes* or $B \rightarrow D_{\text{Gap}}^{**}\ell\nu$.

The branching fractions of the $B \rightarrow X_c \ell \nu$ resonant modes are taken from the most recent HFLAV review [63], assuming isospin symmetry between the neutral and charged channels. In $B \rightarrow D^* \ell \nu$ branching fraction measurements, the D^* meson is reconstructed via one of the $D^* \rightarrow D\pi$ channels which, because of the difficulty to reconstruct low-momentum neutral pions, favours the study of the $B^0 \rightarrow D^{*-}(\rightarrow D^0\pi^-)\ell\nu$ channel. For this reason, the $B^0 \rightarrow D^{*-}\ell\nu$ branching fraction has been measured more often than that of $B^- \rightarrow D^{*0}\ell\nu$. Assuming isospin symmetry, the branching fraction of the $B^- \rightarrow D^{*0}\ell\nu$ channel can therefore be obtained from

$$\mathcal{B}(B^- \rightarrow D^{*0}\ell\nu) = \mathcal{B}(B^0 \rightarrow D^{*-}\ell\nu) \cdot \frac{\tau_{B^-}}{\tau_{B^0}}$$

where τ_{B^-}/τ_{B^0} is the ratio of B^- and B^0 lifetimes and reads $\tau_{B^-}/\tau_{B^0} = 1.076 \pm 0.004$. Similarly, the $B^- \rightarrow D^0\ell\nu$ branching fraction is obtained from that of the $B^0 \rightarrow D^-\ell\nu$ channel. The computation is slightly more challenging for decays to D^{**} resonances because their decay products overlap and only partial branching fractions are measured to facilitate the final state reconstruction. Following the prescription of Ref. [131], the total branching fractions are computed from these partial branching fractions, which directly assumes isospin symmetry. The isospin factor of the two-body pion mode is

$$f_\pi = \frac{\mathcal{B}(D^{**} \rightarrow D^{(*)}\pi^+)}{\mathcal{B}(D^{**} \rightarrow D^{(*)}\pi)} = \frac{2}{3}.$$

From the average of $\mathcal{B}(B^- \rightarrow D_0^{*0}(\rightarrow D^+\pi^-)\ell\nu)$, the $B^- \rightarrow D_0^*\ell\nu$ branching fraction is therefore calculated as

$$\mathcal{B}(B^- \rightarrow D_0^*\ell\nu) = \mathcal{B}(B^- \rightarrow D_0^{*0}(\rightarrow D^+\pi^-)\ell\nu)/f_\pi$$

and the B^0 channel branching fraction is calculated by dividing by τ_{B^-}/τ_{B^0} . The $B \rightarrow D_1'\ell\nu$ branching fractions are computed in the same way from $\mathcal{B}(B^- \rightarrow D_1'^0(\rightarrow D^{*+}\pi^-)\ell\nu)$. For the $B \rightarrow D_2^*\ell\nu$ channel, $D_2^* \rightarrow D^{(*)}\pi\pi$ decays are not considered but they are expected to yield only a very small branching fraction. These decays are usually measured via $D_2^{*0} \rightarrow D^{*+}\pi^-$ decays

and therefore, to also take into account the $D\pi$ contribution, we use [24]

$$f_{D_2^*} = \frac{\mathcal{B}(D_2^* \rightarrow D^- \pi^+)}{\mathcal{B}(D_2^* \rightarrow D^{*-} \pi^+)} = 1.52 \pm 0.14.$$

The $B^- \rightarrow D_2^* \ell \nu$ branching fraction is therefore calculated from

$$\mathcal{B}(B^- \rightarrow D_2^* \ell \nu) = \mathcal{B}(B^- \rightarrow D_2^{*0} (\rightarrow D^{*+} \pi^-) \ell \nu) \cdot (1 + f_{D_2^*}) / f_\pi$$

and the B^0 decay is obtained by dividing by τ_{B^-} / τ_{B^0} . For the D_1 channel, both the $D_1 \rightarrow D^* \pi$ and $D_1 \rightarrow D \pi \pi$ channels must be considered. Since the $B \rightarrow D_1 \ell \nu$ branching fraction has been measured through $D_1 \rightarrow D^* \pi^+$ decays, the former contribution can be computed using f_π whereas for the latter one we need to use [132]

$$f_{D_1} = \frac{\mathcal{B}(D_1 \rightarrow D^0 \pi^- \pi^+)}{\mathcal{B}(D_1 \rightarrow D^{*-} \pi^+)} = 0.43 \pm 0.11$$

and the three-body pion mode isospin factor

$$f_{\pi\pi} = \frac{\mathcal{B}(D^{**} \rightarrow D^{(*)} \pi^+ \pi^-)}{\mathcal{B}(D^{**} \rightarrow D^{(*)} \pi \pi)} = \frac{1}{2} \pm \frac{1}{6}$$

and the $B^- \rightarrow D_1 \ell \nu$ branching fraction is therefore calculated as

$$\mathcal{B}(B^- \rightarrow D_1 \ell \nu) = \mathcal{B}(B^- \rightarrow D_1^0 (\rightarrow D^{*-} \pi^+) \ell \nu) \cdot \left(\frac{1}{f_\pi} + \frac{f_{D_1}}{f_{\pi\pi}} \right)$$

and the B^0 channel is again computed assuming isospin symmetry. Based on the previously determined branching fractions of $B \rightarrow D^{**} \ell \nu$ resonant modes, it appears that there is very little phase-space left for non-resonant $B \rightarrow D^{(*)} \pi \ell \nu$ decays, whose contribution is therefore neglected, but $B \rightarrow D^{(*)} \pi \pi \ell \nu$ decays must be explicitly factored in. Finally, since the combined inclusive $B \rightarrow X_c \ell \nu$ branching fraction is $\mathcal{B}(B \rightarrow X_c \ell \nu) = (10.66 \pm 0.15)\%$, the inclusive B^- and B^0 branching fractions are

$$\mathcal{B}(B^0 \rightarrow X_c \ell \nu) = \frac{2}{1 + \frac{\tau_{B^-}}{\tau_{B^0}}} \cdot \mathcal{B}(B \rightarrow X_c \ell \nu) = (10.27 \pm 0.15)\%,$$

$$\mathcal{B}(B^- \rightarrow X_c \ell \nu) = \frac{2}{1 + \frac{\tau_{B^0}}{\tau_{B^-}}} \cdot \mathcal{B}(B \rightarrow X_c \ell \nu) = (11.05 \pm 0.16)\%$$

from which all previously calculated branching fractions can be subtracted to find the $B \rightarrow D^{(*)} \eta \ell \nu$ contribution. The resulting branching fractions used to simulate all $B \rightarrow X_c \ell \nu$ modes are summarised in Table 5.2.

Other backgrounds

Three additional types of decays contribute as sources of background: $e^+ e^- \rightarrow q \bar{q}$ with $q \in \{u, d, s, c\}$ events, which are referred to as *continuum*, *secondary lepton* decays and *fake lepton* decays. A dataset of 1 ab^{-1} of continuum events at the $\Upsilon(4S)$ resonance and a dataset of 50 fb^{-1} of events 60 MeV below the $\Upsilon(4S)$ resonance are generated. Secondary lepton decays consist of decays where the lepton doesn't originate from the B decay vertex but from another vertex further down the decay chain. These events are predominantly of the type $B \rightarrow D(\rightarrow X \ell \nu) X'$, where X and X' represent any system. Fake lepton decays represent decays where a hadron in

Decay mode	$\mathcal{B} [\times 10^{-2}]$	
	B^+	B^0
$B \rightarrow D \ell \nu$	2.27 ± 0.06	2.11 ± 0.05
$B \rightarrow D^* \ell \nu$	5.27 ± 0.12	4.90 ± 0.11
$B \rightarrow D^1 \ell \nu$	0.64 ± 0.10	0.59 ± 0.10
$B \rightarrow D_0^* \ell \nu$	0.13 ± 0.19	0.12 ± 0.18
$B \rightarrow D_1' \ell \nu$	0.28 ± 0.04	0.26 ± 0.04
$B \rightarrow D_2^* \ell \nu$	0.32 ± 0.03	0.30 ± 0.03
$B \rightarrow D_s K \ell \nu$	0.03 ± 0.01	-
$B \rightarrow D_s^* K \ell \nu$	0.03 ± 0.02	-
$B \rightarrow D_0^*(\rightarrow D\eta) \ell \nu$	0.90 ± 0.90	0.86 ± 0.86
$B \rightarrow D_1'(\rightarrow D^* \eta) \ell \nu$	0.90 ± 0.90	0.86 ± 0.86
$B \rightarrow D_1'(\rightarrow D\pi\pi) \ell \nu$	$0.5 \times (0.07 \pm 0.09)$	$0.5 \times (0.07 \pm 0.08)$
$B \rightarrow D_0^*(\rightarrow D\pi\pi) \ell \nu$	$0.5 \times (0.07 \pm 0.09)$	$0.5 \times (0.07 \pm 0.08)$
$B \rightarrow D_1'(\rightarrow D^* \pi\pi) \ell \nu$	$0.5 \times (0.22 \pm 0.10)$	$0.5 \times (0.20 \pm 0.10)$
$B \rightarrow D_0^*(\rightarrow D^* \pi\pi) \ell \nu$	$0.5 \times (0.22 \pm 0.10)$	$0.5 \times (0.20 \pm 0.10)$

TABLE 5.2: Assumed branching fractions for all $B \rightarrow X_c \ell \nu$ channels. The gap mode branching fractions are given in the lower block.

the decay chain has been misidentified as a lepton. The latter two types of decays are produced in the generic $e^+e^- \rightarrow B\bar{B}$ sample.

5.3 Event reconstruction

As already mentioned, we use the FEI algorithm to reconstruct the partner B meson or B_{tag} in its hadronic decays, which is then paired with a signal lepton candidate. All remaining tracks and clusters on the signal side are grouped in a single object to form the rest-of-event (ROE). The ROE is used to represent the hadronic system X of a semileptonic decay. For the discussion below, it is useful to first define several variables which are used to select events. For any system X , with the $*$ superscript indicating that the variable is computed in the centre-of-mass frame, we define:

- the beam energy: E_{beam} ,
- the transverse momentum: p_t ,
- the polar angle: θ ,
- the difference between the energy of X in the centre-of-mass frame and half the beam energy: $\Delta E = E_{B_{\text{tag}}}^* - E_{\text{beam}}^*/2$,
- the beam constrained mass: $M_{\text{bc}} = \sqrt{\left(\frac{E_{\text{beam}}^*}{2}\right)^2 - p_B^{*2}}$, with p_B the B momentum magnitude,
- the distance between the interaction point and the track point-of-closest-approach in the $x-y$ plane: dr ,
- the distance between the interaction point and the track point-of-closest-approach along the z -axis: dz ,

- the time of the event relative to the trigger time: t_0 .

5.3.1 Kinematic variables

The rest frame of the signal B is inferred from the B_{tag} as in the centre-of-mass frame, the two B mesons are produced back-to-back. We will therefore usually consider the lepton energy in the signal B rest frame, E_ℓ^B . In the following, the mass of the lepton in a semileptonic decay will be neglected, therefore fixing $E_\ell^B = |\vec{p}_\ell^B|$. This approximation is valid as we only consider decays to electrons and muons with energies above 1 GeV. The hadronic system four-momentum p_X is computed from the energies and three-momenta of all tracks and clusters in the ROE, with all tracks being assigned a pion mass hypothesis. As described below, when kaons and leptons are found in the ROE, an event is rejected hence leaving only pion tracks in the ROE of selected events and rarely protons. From p_X , the hadronic system mass M_X is computed. Using the four-momenta of the initial e^+e^- state, the B_{tag} , the lepton and the hadronic system, the neutrino four-momentum can be inferred from the missing four-momentum as:

$$p_\nu = p_{\text{miss}} = p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_X. \quad (5.2)$$

Furthermore, the momentum transfer squared (or leptonic system invariant mass squared) q^2 is defined as $(p_\ell + p_\nu)^2$, where the missing energy is constrained to be equal to the magnitude of the missing three-momentum, hence assuming a massless neutrino. q^2 is therefore computed as

$$q^2 = (|\vec{p}_{\text{miss}}|, \vec{p}_{\text{miss}} + p_\ell)^2, \quad (5.3)$$

where the missing momentum magnitude is used instead of the missing energy, as its distribution has a better resolution, and such a definition of q^2 doesn't depend on the mass hypothesis assigned to tracks. Finally, the missing mass squared M_{miss}^2 is computed as $M_{\text{miss}}^2 = p_{\text{miss}}^2$.

5.3.2 Tag Side

The B_{tag} candidate is reconstructed by the FEI in one of $\mathcal{O}(10^4)$ hadronic B channels. As recommended by the Belle II performance group, we use the same tag-side selections as in the FEI calibration study performed with semileptonic decays on the signal side [133]. The selections are as follows:

- ΔE is required to be close to 0: $-0.15 < \Delta E < 0.10$ GeV,
- M_{bc} is required to be greater than 5.24 GeV (this is a first loose selection on M_{bc} which allows us to reject continuum background or very poorly reconstructed events; the selection is tightened at a later stage),
- the FEI signal probability $\mathcal{P}_{\text{FEI}}$ is required to be greater than 0.01,
- to suppress continuum background, the variable $\cos \theta_{\mathcal{T}_B^{\text{FEI}} \mathcal{T}_O}$ (the magnitude of the cosine of the angle between the thrust axes of the tag B meson and the ROE) is required to be lower than 0.9. This variable is defined with the ROE selections used for the FEI calibration. These selections are as follows:

– Tracks are required to:

- * originate from close to the interaction point with $dr < 2$ cm and $|dz| < 4$ cm,
- * have a minimum transverse momentum of $p_t > 0.2$ GeV,
- * have momentum polar angle within the CDC acceptance *i.e.* $17^\circ < \theta < 150^\circ$.

– Clusters are required to:

- * have a minimum energy of 80, 30, and 60 MeV respectively in the forward, barrel, and backwards regions of the ECL,
- * have $17^\circ < \theta < 150^\circ$ to reject clusters from the edges of the ECL,
- * be recorded within 200 ns of t_0 , consistent with originating from photons created in the event.

These selections do not necessarily match the ROE selections used in our measurement, which we define below. The $\cos \theta_{\mathcal{T}_B \mathcal{T}_O}$ variable is further discussed in Section 5.5.1 and illustrated in Appendix B.

At this stage, we select the B_{tag} candidate with the highest $\mathcal{P}_{\text{FEI}}$. In poorly reconstructed events, both a charged and a neutral candidate can be selected, which happens for about 1.5% of events (from simulation) but one of the two candidates is always rejected by other requirements such as the $\cos \theta_{\mathcal{T}_B \mathcal{T}_O}^{\text{FEI}}$ selection. Since B mesons are produced nearly at rest in the centre-of-mass frame, the M_{bc} distribution of B meson decays is expected to peak around 5.28 GeV and the M_{bc} requirement is thus tightened to $M_{\text{bc}} > 5.27$ GeV. For off-resonance events, we apply the selection to a scaled beam-constrained mass to account for the energy difference between on-resonance and off-resonance events,

$$M_{\text{bc}}^{\text{corr.}} = \sqrt{\frac{E_{\text{on}}^2}{4} - \left(\frac{E_{\text{on}}}{E_{\text{off}}}\right)^2 p_{B_{\text{tag}}}^{*2}}, \quad (5.4)$$

where E_{on} and E_{off} are the nominal beam energies of the on and off-resonance samples: $E_{\text{on}} = 10.58$ GeV and $E_{\text{off}} = 10.52$ GeV.

5.3.3 Signal Side

Signal Lepton

The signal lepton candidate is reconstructed as an electron or a muon. Tracks identified as signal electron or muon candidates are required to:

- have an origin which is consistent with the interaction point with $dr < 1$ cm and $|dz| < 3$ cm,
- fall in the phase-space region studied in Belle II lepton identification performance studies [134–136] by having a polar angle *i.e.* $17^\circ < \theta < 150^\circ$ for electrons and $23^\circ < \theta < 149^\circ$ for muons and a three-momentum magnitude in the laboratory frame $|\vec{p}_{\text{lab}}| > 0.3$ GeV,
- have a BDT-based electron identification score greater than 0.9 or a likelihood-ratio-based muon identification score greater than 0.9 as recommended by the Belle II performance group (see Section 3.3.8 for a description of particle identification in Belle II). The electron identification requirement selects roughly 40% of events and the muon requirement roughly 60% of events. About 10% of electrons and 35% of muons are misidentified particles (mostly pions).
- Since the electron identification BDT was not trained to reject protons, we additionally require that electron candidates have a proton identification score lower than 0.99. This selection retains close to 100% of electrons while rejecting about 75% of protons misidentified as electrons.
- Finally, the four-momentum of electron candidates is corrected for bremsstrahlung energy loss caused by the emission of photons when the electron decelerates in the electric field of a nucleus in the detector material [24]. Bremsstrahlung photons are identified within a

0.05 rad ($\sim 2.9^\circ$) radius cone around the electron initial momentum direction. They are required to have an energy below 1 GeV and meet the criteria applied to ROE photons as described in the next section.

After having been preselected, the signal lepton candidates are combined with the selected B_{tag} . If the lepton candidate can be combined with both the charged and neutral B_{tag} candidate, we select the B_{tag} with the highest $\mathcal{P}_{\text{FEI}}$. Events with multiple signal lepton candidates are rejected and the lepton charge is required to be consistent with the signal B flavour inferred from the B_{tag} flavour, which effectively rejects $B^0 - \bar{B}^0$ mixing. Even though this requirement vetoes about a fifth of neutral B signal events [24], it also rejects events in which a secondary lepton has been reconstructed as the signal lepton. Moreover, the FEI calibration factors were only extracted for events with a lepton charge consistent with the B_{tag} flavour.

Signal lepton candidates may also arise from the decay $B \rightarrow J/\psi(\rightarrow \ell\ell)X$ where X represents any system and, for electrons, from photon conversion. These two-body decays can nevertheless be rejected rather efficiently. Each lepton candidate is combined with all oppositely charged tracks found in the ROE and each of these tracks is assigned the same mass hypothesis as the signal lepton candidate (no lepton identification criteria is applied to this track). Of all possible pairs, the one with an invariant mass closest to the nominal mass of the J/ψ (3.097 GeV) for the J/ψ rejection or to 0 GeV for the photon rejection is selected. For the J/ψ rejection, the second track is required to pass the ROE selections defined in the next section whereas for the photon conversion rejection, these selections are not applied since these events are expected to occur in the detector material and therefore further away from the interaction point. The bremsstrahlung correction mentioned earlier is applied to tracks which are assigned an electron mass hypothesis. Signal lepton candidates are rejected if:

- an electron pair has a mass below 36 MeV,
- an electron pair has a mass between 3.035 and 3.132 GeV,
- a muon pair has a mass between 3.078 and 3.116 GeV.

These selections are optimised to reject about 95% of true photon conversion tracks and J/ψ , except for electrons below the J/ψ mass where a broad tail caused by bremsstrahlung can be observed. In this case, the threshold rejects 68% of true J/ψ .

Rest-of-Event

The ROE is used to characterise the hadronic system without explicitly reconstructing its constituents. Various sets of selections can be used for the ROE reconstruction. The nominal ROE selections applied to neutral clusters were optimised to reject phase-space regions populated by beam background and misidentified photons, which tend to have lower energies than true photons originating from collision events and be reconstructed at later times. In addition, the polar angle selections applied to clusters and tracks suppress particles hitting the edges of the detector, whose response is poorly modelled. A fit is performed using all tracks and clusters found in the ROE to match them to a single decay vertex. The ROE selections are as follows:

- Tracks are required to:
 - originate from close to the interaction point with $dr < 1$ cm and $|dz| < 3$ cm,
 - have a polar angle within $18^\circ < \theta < 149^\circ$,
 - have left at least two energy deposits in the CDC.

- In addition, low-momentum tracks which curl inside the detector magnetic field are rejected as they can be reconstructed as multiple tracks and bias the event net charge. The rejection is done using a boosted decision tree (BDT), which was developed for Belle analyses [137] and retrained with Belle II data. The BDT uses as input variables the five track helix parameters, the product of the charges of the tracks, the angle between the tracks and their transverse and longitudinal momenta.
- Clusters are required to:
 - have a minimum energy of 70, 60 and 70 MeV in the forward, barrel, and backward regions of the ECL, respectively,
 - have a polar angle within the CDC acceptance, *i.e.* $17^\circ < \theta < 150^\circ$,
 - be recorded within 70 ns of t_0 , consistent with photons created in the event.
 - In addition, misidentified photons are rejected using a BDT classifier trained to distinguish them from real photons. The BDT uses various cluster properties and was developed for Belle II analyses [138].

Shifts of the total energy of about 5 – 15 MeV in various decays involving final-state photons have been observed in Belle II data when compared to simulated samples [139]. These shifts are attributed to a poor energy calibration of the photons and experimental data is therefore corrected using a set of scaling factors provided by the Belle II performance group.

As mentioned in the previous section, events with lepton candidates in the ROE (based on the criteria defined above for leptons), in addition to the signal lepton candidate, are rejected. Their presence is indicative of secondary semileptonic decays, typically $B \rightarrow X_c(\rightarrow X \ell \nu) \ell \nu$. Requiring that a single lepton candidate is found rejects about 1.5% of signal events and 3.6% of background events.

Slow pions

Most decays of D^* mesons produce a D meson and a pion. Approximately two-thirds of $D^{*\pm}$ decay to $D^0 \pi^\pm$ and the remaining third to $D^\pm \pi^0$. Likewise D^{*0} mesons decay to $D^0 \pi^0$ with a branching fraction of about two-thirds and to $D^0 \gamma$ with a branching fraction of one-third². Since the mass of D is close to that of D^* , pions from $D^* \rightarrow D \pi$ decays are produced with a relatively low momentum and are boosted in the flight direction of the D^* . We will refer to these pions as slow pions³. Using slow pions, we perform an inclusive reconstruction of D^* mesons in the ROE. The properties of the D^* mesons and the subsequent slow pions can be used to reject $B \rightarrow D^* \ell \nu$ events as discussed in Section 5.5.2⁴. Each D^* candidate is assumed to have a momentum along the slow pion flight direction in the laboratory frame and their energy is calculated from:

$$E_{D^*} = \frac{M_{D^*}}{M_{D^*} - M_D} E_\pi, \quad (5.5)$$

where M_{D/D^*} are the masses of the D/D^* mesons and E_π is the energy of the slow pion. Having computed the D^* four-momentum and knowing the B_{tag} and signal lepton candidate four-momenta, the missing mass of each D^* event can be computed as:

$$M_{\text{miss}}^2(D^* \rightarrow D \pi) = (p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_{D^*})^2. \quad (5.6)$$

² $D^{*0} \rightarrow D^\pm \pi^\mp$ is not possible as the sum of the $D^\pm$ and $\pi^\pm$ masses exceeds the D^{*0} mass

³For non-physicists, the word "slow" might not seem particularly adapted to a particle moving at a speed somewhere between 100,000 and 250,000 km/s.

⁴The properties of the reconstructed D^* mesons are only used in this context and are not used to compute the kinematic characteristics of the hadronic system.

When multiple slow pion candidates are reconstructed, the one with the lowest absolute $M_{\text{miss}}^2(D^* \rightarrow D\pi)$ is selected. The reconstruction efficiency of slow $\pi^\pm$ and π^0 was calibrated in independent studies by the Belle II performance group (see Section 5.4.8). We define the momentum range of slow pions to be between 50 and 200 MeV in the laboratory frame. Slow $\pi^\pm$ are identified as single tracks with a charge opposite to that of the signal lepton and curl tracks are rejected using the classifier mentioned above. π^0 are reconstructed from pairs of photons. We align our slow pion selections with those used in the calibration studies [140, 141]. The clusters identified as photons originating from the slow π^0 are required to:

- have a polar angle within $17^\circ < \theta < 150^\circ$,
- have deposited energy in at least two crystals in the ECL,
- have a minimum energy of 80, 30 and 60 MeV respectively in the forward, barrel, and backwards regions of the ECL, respectively,
- must be recorded within 200 ns of t_0 .
- Misidentified photons are rejected using the fake photon BDT classifier mentioned in the previous section.

Using the two photons, a vertex fit is performed to identify the $\pi^0 \rightarrow \gamma\gamma$ vertex. The slow π^0 candidate is required to have a mass within 15 MeV of the nominal π^0 mass, 135 MeV.

Kaons

Kaons are representative of charm meson decays and we therefore reject events where a $K^\pm$ or K_S^0 is found. K_L^0 are not considered in our analysis as they are poorly reconstructed by the KLM. We discuss further the implications of poorly detected K_L^0 in Appendix E. The kaon veto can also be inverted to define control regions orthogonal to the signal region, as highlighted in Section 5.8. Charged kaons are reconstructed as charged tracks with a likelihood-based kaon identification probability greater than 0.6, which ensures that the particle is likely a $K^\pm$ and preserves a relatively high reconstruction efficiency (*c.f.* Section 3.3.8 for a detailed description of particle identification at Belle II). $K^\pm$ candidates are required to have at least 20 energy deposits in the CDC and have their momentum p and polar angle θ in the laboratory frame within the phase-space region considered in Belle II hadron identification performance studies [142] *i.e.* $0.3 < p < 4.5$ GeV and $17^\circ < \theta < 150^\circ$. K_S^0 candidates are reconstructed from two oppositely charged tracks defined under the pion mass hypothesis. As K_S^0 have a relatively long lifetime, their decay vertex (identified by performing a vertex fit) is expected to be far away from the interaction point and we therefore don't apply on K_S^0 candidates the ROE selections described above. K_S^0 candidates are required to have an invariant mass close to the nominal K_S^0 mass of 497.6 MeV *i.e.* between 470 and 530 MeV. Finally, we ensure that K_S^0 candidates have a momentum vector consistent with the vector connecting the interaction point and the fitted decay vertex by requiring the cosine of the angle between these two vectors to be greater than 0.998. The $K^\pm$ fake rate and the K_S^0 and $K^\pm$ detection efficiency calibrations are discussed in Sections 5.4.6 and 5.4.7 respectively.

High-momentum $\pi^\pm$

Charged pions are used to derive an uncertainty on the X_u hadronisation modelling as described in Section 5.9.4. Similarly to $K^\pm$, $\pi^\pm$ candidates are required to have a likelihood-based pion identification probability greater than 0.6, have recorded at least 20 energy deposits in the CDC and have their momentum and polar angle within the phase-space region considered in the Belle II hadron identification performance study.

5.3.4 Additional selections

All selections discussed in this section – except the kaon veto – are referred to as *preselections*. Applying the preselections defines a phase-space region that will be referred to as control region 1 (CR1). After the preselection, we apply a continuum and $B \rightarrow X_c \ell \nu$ event rejection using two separate neural network classifiers, which are described in detail in dedicated sections below. The kaon veto and the $B \rightarrow X_c \ell \nu$ suppression selection are used at a later stage to define control regions orthogonal to the signal region. The importance of these control regions will also be highlighted further in the next sections. Furthermore, applying selections on the main variables of interest, E_ℓ^B , M_X and q^2 , is very effective to reject sources of background. A requirement of $E_\ell^B > 1$ GeV is applied during the reconstruction and preselection procedure, which rejects about 13% of the accessible $B \rightarrow X_u \ell \nu$ phase-space (see Table 5.7). Looser selections were explored but the low lepton energy region is highly populated by continuum and secondary and fake lepton events. Additional kinematic selections are further discussed in Section 5.6.

5.4 Sample corrections

Various corrections are applied to the simulated samples by scaling each event with dedicated weights. We describe all applied corrections in the following, starting with corrections related to the modelling of events and then discussing corrections related to detector effects. A detailed presentation of the implementation of associated sources of uncertainty in the signal extraction setup is given in Section 5.9.4. Based on the number of simulated events and the size of the experimental dataset, events are scaled such that their total yields match the expected number of collected events.

In the following sections, all corrections described below are applied to simulated samples used in comparisons between simulated and experimental samples unless otherwise indicated or unless the correction is not relevant for a given sample, as explained below for each type of correction. In addition, all uncertainties described in Section 5.9.4, except those applied to the $B \rightarrow X_u \ell \nu$ branching fractions and $|V_{ub}|$ after the signal extraction, are included in the simulated sample uncertainty band labelled *MC Uncert.* or *Uncertainty* in figures. In figures where one source of uncertainty is not included in the error band, this source is mentioned explicitly.

5.4.1 Branching fractions

The branching fractions used during sample production for each decay type are not necessarily up-to-date with the latest measurements or averages performed by averaging groups. Semileptonic events are scaled such that their branching fraction matches those listed above in Table 5.1 and 5.2 for $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$, respectively. In addition, we also update the D decay branching fractions. These types of decays, especially three and four-body decays, are particularly complicated to model but most of them are simulated by simple phase-space models in the samples used for our measurement. This poor modelling is thought to be one of the causes of observed discrepancies between simulated and experimental data distributions (see Section 5.8). While a better modelling would only be achieved by replacing the simple phase-space D decay models by dedicated models which can be extracted from Dalitz analyses, for example, such an effort is still work in progress in the Belle II collaboration. Nevertheless, the modelling can already be improved by updating the D decay branching fractions used in the simulation. For this purpose, we use the PDG 2024 list of branching fractions [24]⁵. To mitigate any potential impact of this procedure on the tag-side, and therefore on the FEI calibration weights, and on

⁵In recent versions of the PDG list, some of these branching fractions are not listed and we therefore refer to the latest PDG list where they are mentioned.

the continuum, which we rescale separately, we only correct the branching fraction of a charm meson decay when it is associated with the same B meson as the signal lepton. With this rescaling method, decay channels which are listed in the PDG but not included in the simulation cannot be incorporated. In channels which have only been measured through the decay of an intermediate resonance, the D decay branching fraction is corrected assuming the branching fraction of the intermediate decay to be the one listed in the simulation and no uncertainty is assigned to this branching fraction. For instance, $\mathcal{B}(D^0 \rightarrow (\phi \rightarrow K^- K^+) \pi^0)$ is corrected assuming $\mathcal{B}(\phi \rightarrow K^- K^+) = 48.97\%$. When only a limit is quoted in the PDG, the branching fraction is set to this limit and it is assigned a relative uncertainty of 100%. After correcting the branching fractions, the total D^0 rate slightly exceeds 100%, whereas the total D^+ rate is slightly below. Therefore, the branching fraction of the poorly known $D^0 \rightarrow K^- a_1^+$ channel is decreased from 14% to 11.94% and the difference is added as an uncertainty. The branching fractions of the unmeasured $D^+ \rightarrow \eta 5\pi$ channels are increased by a factor of about 2.5 and assigned a relative uncertainty of 100%. A total of 227 charm decay branching fractions are thus corrected.

5.4.2 Form factors

Similarly to the branching fractions discussed above, the samples of semileptonic events used for our measurement are simulated assuming specific form factor models and parameter values. However, more up-to-date models and input values are available for each decay type and we therefore rescale the semileptonic samples such that their modelling matches the form factor models mentioned in Section 5.2. The scale factors are computed with **eFFORT** for $B \rightarrow X_u \ell \nu$ decays and **Hammer** for $B \rightarrow X_c \ell \nu$ decays (*c.f.* section 4.4.1 and 4.4.2 for a detailed description of these software packages). We summarise the form factor corrections for all semileptonic decays in Table 5.3. Figure 5.4 illustrates the differences in the modelling of semileptonic decays between two separate theoretical descriptions.

Decay mode	Generation model	Updated model/inputs
$B \rightarrow \pi \ell \nu$	BCL [38]	BCL [39]
$B \rightarrow \rho \ell \nu$	BCL	BSZ [40]
$B \rightarrow \omega \ell \nu$	BCL	BSZ
$B \rightarrow \eta \ell \nu$	ISGW2 [143]	LCSR [42]
$B \rightarrow \eta' \ell \nu$	ISGW2	LCSR
$B \rightarrow D \ell \nu$	BGL [30]	BLPRXP [36]
$B \rightarrow D^* \ell \nu$	BGL	BLPRXP
$B \rightarrow D^{**} \ell \nu$	LLSW [144]	BLR [37]
$B \rightarrow D_{\text{Gap}}^{**} \ell \nu$	LLSW	BLR

TABLE 5.3: Summary of the form factor models used during event generation and those that we update to. In the second column, the references where these models are described are given. In the third column, the references where the input parameter values are taken from are given.

5.4.3 Broad D^{**} modelling

The BLR parametrisation used to model $B \rightarrow D^{**} \ell \nu$ decays relies on the Breit-Wigner distribution and narrow-width approximations to describe the D^{**} resonance lineshapes (recent work has focused on developing theoretical frameworks avoiding these approximations [122, 145]). A consequence of the imperfect modelling is that decays to the two broad D^{**} resonances – D_0^*

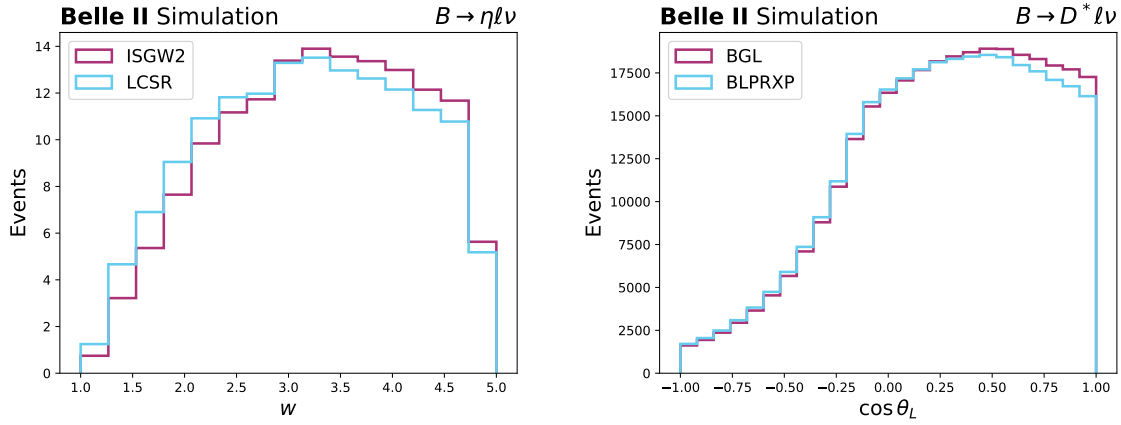


FIGURE 5.4: Left: histogram of the generated recoil parameter w distribution in the preselected sample of $B \rightarrow \eta \ell \nu$ events for the ISGW2 and LCSR models. Right: histogram of the generated $\cos \theta_L$ distribution in the preselected sample of $B \rightarrow D^* \ell \nu$ events for the BGL and BLPRXP models. The θ_L variable is the angle between the lepton direction and the direction of the W boson in the B meson rest frame. The distributions are shown in purple for the model used when generating the samples and in cyan for the updated model.

and D_1^* – occur at higher rates than expected in the high M_X region with many simulated D^{**} mesons approaching the B meson mass (the masses and widths of charm mesons are shown in Figure 2.8). The generator-level M_X spectra of the $B^- \rightarrow D_0^{*0} \ell^- \nu$ and $B^- \rightarrow D_1^{*0} \ell^- \nu$ decays are shown in Figure 5.5. The impact of the unexpectedly high hadronic mass on event kinematics becomes apparent in the w spectrum, where a spike at $w \approx 1$ can be observed as shown in Figure 5.5. These unphysical events are rejected by applying a selection on the broad D^{**} mass in simulated events. D_0^* (D_1^*) events are required to yield a hadronic mass that does not exceed the nominal resonance mass by more than 3 (2.5) times its width. The effect of this selection is illustrated in Figure 5.5. Events which pass the selection are upscaled to compensate for those that are rejected and maintain the same integrated luminosity.

5.4.4 Continuum calibration

Continuum events ($e^+e^- \rightarrow q\bar{q}$, $q \in \{u, d, s, c\}$) are simulated with the KKMC and Pythia 8.3 generators, the former modelling the $e^+e^- \rightarrow q\bar{q}$ transition and the latter modelling the hadronisation of quarks. A large number of parameters are available in Pythia to tune the fragmentation modelling and an ongoing data-driven tuning effort is taking place for Belle II simulated datasets. Without this dedicated tuning and since the FEI calibration factors are not applied to continuum events, discrepancies are observed between experimental and simulated data distributions of the continuum. To correct these differences, we perform a data-driven calibration using the off-resonance data and simulated data samples, which are expected to be populated only by $e^+e^- \rightarrow q\bar{q}$ events. After preselection, we respectively find 11,965 events in the data sample and 13,500 in the simulated data sample.

Using 40% of the samples (which are the training samples in the next step described below), we extract overall normalisation calibration factors separately for the charged and neutral B channels. We respectively find calibration factors of 0.874 and 0.938. To improve the overall modelling of the continuum component, we then perform a method originally proposed in Ref. [146] and used in other Belle II measurements such as the recent search for $B \rightarrow K \nu \nu$ decays [147]. To extract the calibration factors, we train a boosted decision tree (BDT) with the

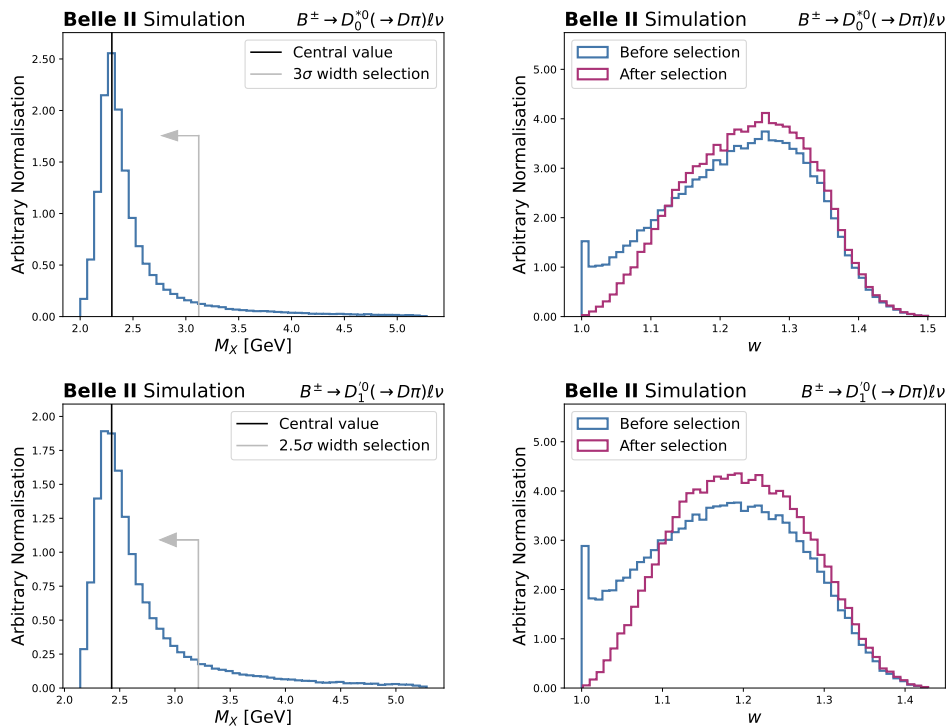


FIGURE 5.5: On the left-hand side, the generator level hadronic mass distributions of $B \rightarrow D^{**} \ell \nu$ events are shown together with the corresponding nominal mass values (black line) and the maximum allowed mass (grey arrow). On the right-hand side, the generator level w distributions are shown before (blue) and after (purple) the mass selection. Distributions are normalised to arbitrary values.

Xgboost package [109] to distinguish between off-resonance data and off-resonance simulated data. From the classifier output score $\mathcal{S}_{\text{cont}}$, a calibration factor for each event can be derived as:

$$w_{\text{cont}} = \frac{\mathcal{S}_{\text{cont}}}{1 - \mathcal{S}_{\text{cont}}}. \quad (5.7)$$

Once the classifier has been trained, it can be applied to continuum events in the on-resonance simulated data sample. The closer $\mathcal{S}_{\text{cont}}$ is to 1, the more data-like the simulated event is and the more weight it is given. As input features, we use the 30 event shape variables, the transverse energy $E_{t,\text{CS}}$ and the missing mass squared $M_{\text{miss,CS}}^2$ which are described in detail in Section 5.5.1. The event shape variables allow us to distinguish jet-like light $q\bar{q}$ events from more spherical $B\bar{B}$ events. The BDT is trained using the sample from which the normalisation factors are extracted as mentioned above. The remainder of the sample is split into a validation sample (10% of the full sample) and a test sample (50%). The validation sample is used to avoid the development of overly complex models, as the training is terminated if the performance of the classifier on this independent sample remains stagnant for 20 consecutive iterations. The BDT is optimised using the **Optuna** [111] package, which is discussed in Section 4.2.3. The optimised values of hyperparameters are summarised in Table 5.4. Finally, the test sample is used to check the performance of the BDT after training and optimisation.

The importance of each input feature as computed by the **Shap** package [113] is shown in Figure 5.6 (see Section 4.2.4 for a description of the **Shap** values). The larger is the importance of a variable in the BDT training, the larger is the discrepancy between the data and simulated data distributions of this variable. The BDT output score $\mathcal{S}_{\text{cont}}$ is shown in Figure 5.7. After applying the calibration, the agreement between the simulated and data distributions of $\mathcal{S}_{\text{cont}}$

Parameter	Value	Range	Description
Max Depth (d)	5	[1, 11]	Maximum depth of each decision tree. Note that the maximum number of trees permitted in the model is dependent on the depth of each tree. If <i>lossguide</i> is adopted as the grow policy, we instead limit the tree to 2^d nodes.
Initial Learning Rate (l_0)	0.124	[0.01, 1.0]	Initial learning rate of model.
Minimum Learning Rate ($l_{\min}$)	0.034	[0.001, 0.1]	Minimum learning rate of model.
Learning Rate Decay (r)	0.003	$[10^{-6}, 0.05]$	Decay of the learning rate.
Alpha (α)	0.078	$[10^{-5}, 10^4]$	Minimum reduction in loss required to further split a node of the tree. Helps prevent overfitting.
Lambda (λ)	53	[0.05, 1000]	L1 regularisation term on the weights. Helps prevent overfitting.
Gamma (γ)	0.007	[0.001, 2]	L2 regularisation term on the weights. Helps prevent overfitting.
Subsample	0.895	[0.5, 1.0]	Fraction of training events sampled to train each tree. Helps prevent overfitting.
Column Subsample	0.559	[0.5, 1.0]	Fraction of variables sampled when training each tree. Helps prevent overfitting by forcing the model to rely on a broader range of variables.
Minimum Child Weight	5	-	Minimum number of training events required in each node after splitting a sample. Helps prevent overfitting by preventing extremely targeted selections.
Grow Policy	depthwise	[depthwise, lossguide]	Whether new nodes are added to a tree in order of closest to the root node or in order of maximum reduction in loss. If <i>lossguide</i> is selected, rather than limiting the maximum depth of the tree, we limit the maximum number of nodes of each tree.
Tree Method	gpu_hist	-	The method used by XGBoost to train the trees. The <i>hist</i> methods pre-bin continuous features into histograms and base their node splitting on the bin indices of the histograms.
nBins	512	-	Number of bins used to bin continuous features.

TABLE 5.4: Hyperparameters of the continuum calibration BDT. All parameters not listed are left at their default value as per version 1.7.6 of **Xgboost**. More details on BDT are given in Section 4.2.

is much improved as the value of χ^2 over the number of degrees of freedom decreases from $\chi^2/\text{d.o.f} = 135/21$ to $\chi^2/\text{d.o.f} = 54/22$. The improved modelling is illustrated in Figure 5.8 for

the most important variable used in the training, $\mathcal{C}_1$, and one of the main kinematic variables, q^2 . Similar improvement is observed in all variables used in the training and E_ℓ^B and M_X .

Belle II Simulation

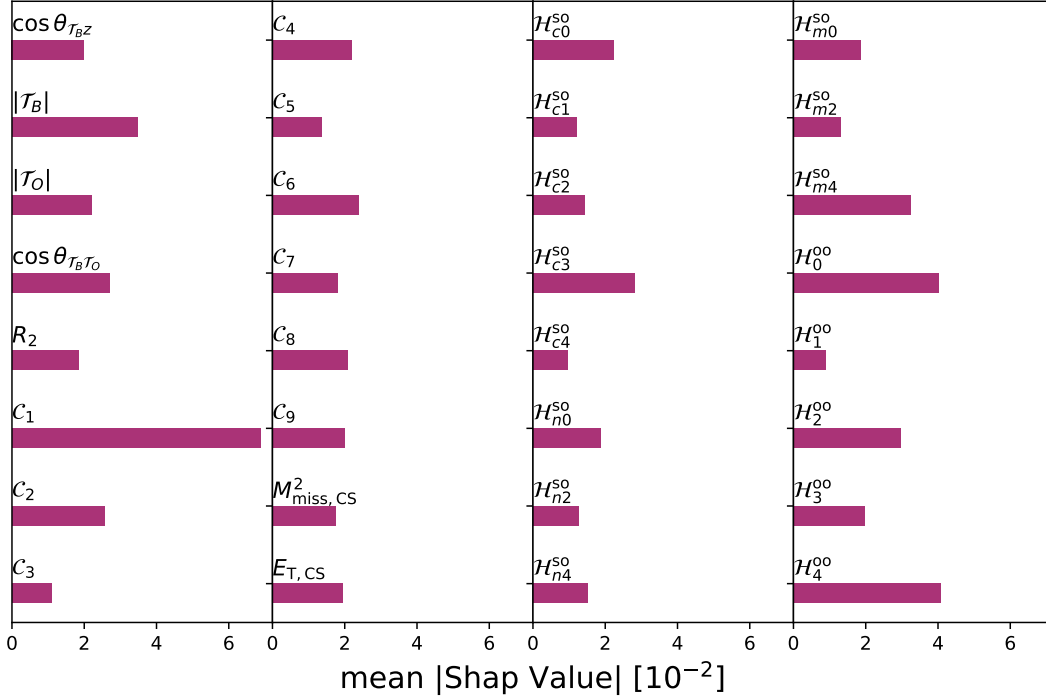


FIGURE 5.6: Importances of the continuum calibration BDT input features.

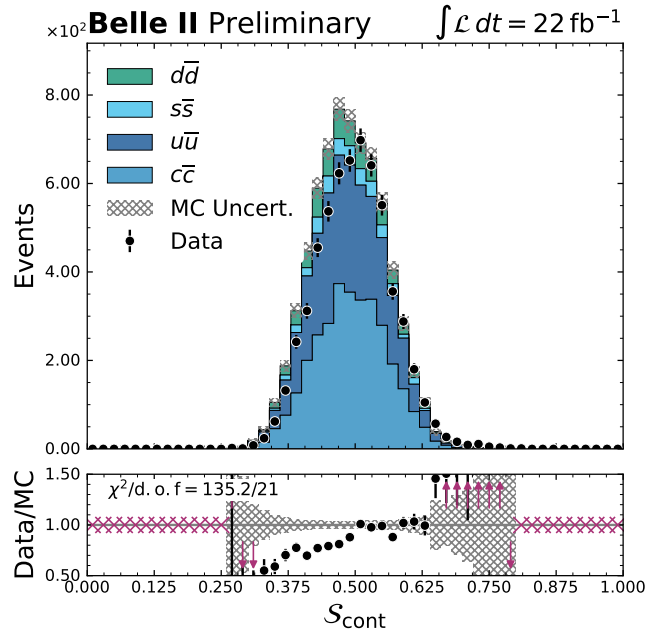


FIGURE 5.7: Continuum calibration BDT output score. All systematic sources of uncertainty except the continuum calibration uncertainties are included in the MC error band. The uncertainty on the data points is purely statistical.

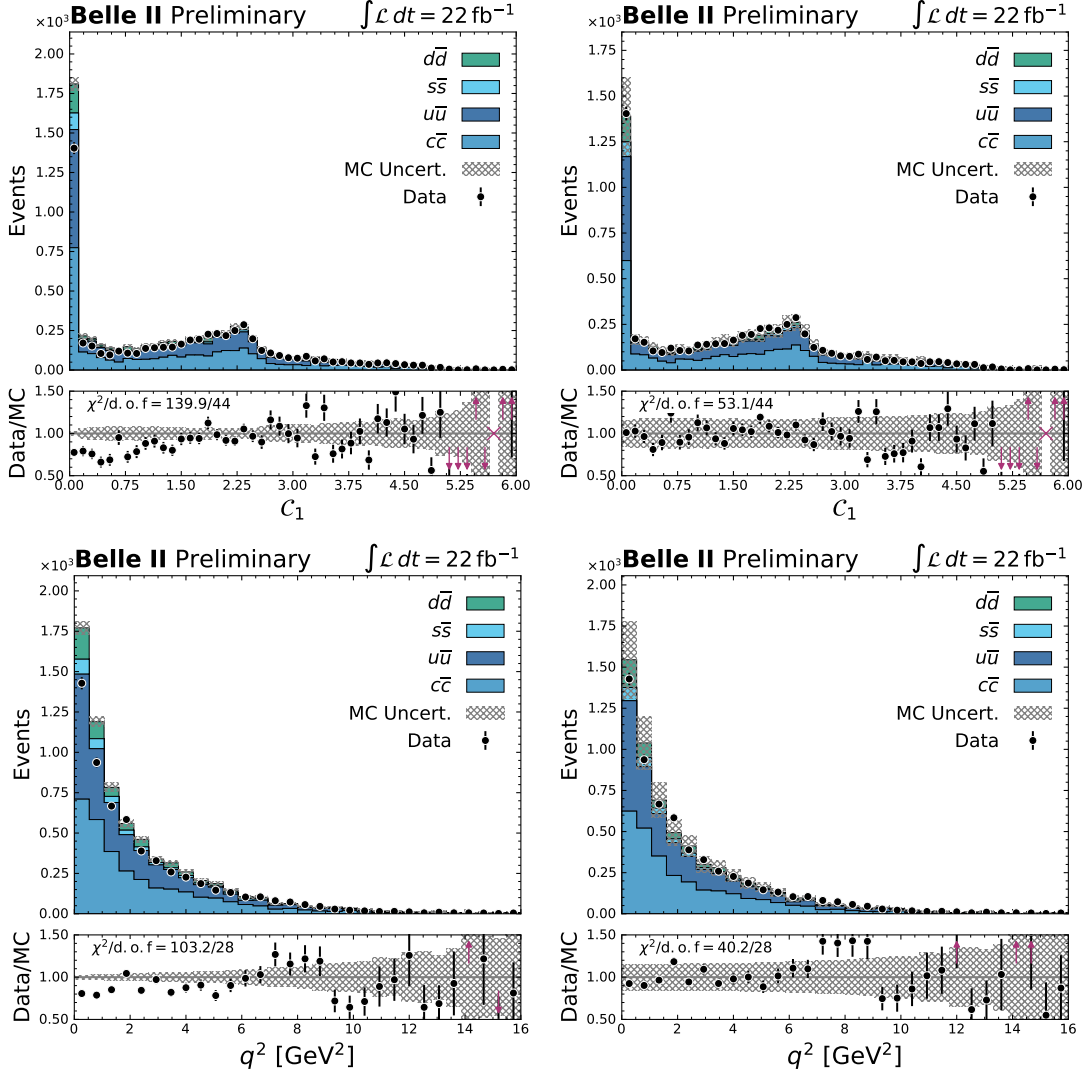


FIGURE 5.8: Distributions of C_1 (see exact definition in section 5.5.1) and q^2 in the test off-resonance sample before (left) and after (right) the continuum calibration is applied. The continuum calibration uncertainties are only included in the MC error band after applying the continuum calibration.

5.4.5 B tagging efficiency

Due to the large number and complexity of decay chains reconstructed by the FEI algorithm, the efficiency of the FEI selection varies between data and simulated data. Calibration factors for the FEI efficiency were determined in two separate studies, focusing on the signal-side decay channels $B \rightarrow X \ell \nu$ and $B \rightarrow D^{(*)} \pi$, which were then combined with a χ^2 fit [148]. The factors extracted from the $B \rightarrow X \ell \nu$ channel dominate the average, as this channel has a significantly larger branching fraction. The factors were computed for two standard FEI score selections, $\mathcal{P}_{\text{FEI}} > 10^{-3}$ and $\mathcal{P}_{\text{FEI}} > 10^{-2}$, which is the one we apply in the preselection. The correction factors were calculated separately for the 11 (10) most common B^+ (B^0) decay channels, with an additional factor derived for all other possible decays for each B charge.

5.4.6 Charged particle identification

The efficiency and misidentification rate of the identification selections applied to charged particles ($e^\pm$, $\mu^\pm$, $K^\pm$, $\pi^\pm$, $p^\pm$) are corrected using calibration factors extracted from independent

dedicated studies performed by the Belle II performance group [142]. The calibration factors are computed from the ratios of identification efficiencies in experimental and simulated data in different bins of track momentum, polar angle and particle charge. The lepton identification efficiency is studied using the $e^+e^- \rightarrow \ell^+\ell^-(\gamma)$, $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ decay channels [134, 135]. The $\pi^\pm$ to lepton misidentification fake rate is extracted from $K_S^0 \rightarrow \pi^+\pi^-$ [135] and $e^+e^- \rightarrow \tau^+\tau^-$ studies and the $K^\pm$ to lepton misidentification fake rate is studied in the $D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+)\pi^+$ channel [136]. The $\pi^\pm/K^\pm$ identification efficiencies and $\pi^\pm \rightarrow K^\pm$ and $K^\pm \rightarrow \pi^\pm$ misidentification rates were studied in the $D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+)\pi^+$ and $K_S^0 \rightarrow \pi^+\pi^-$ channels. The rate of protons misidentified as electrons is also corrected using factors extracted from a study of $\Lambda^0 \rightarrow p^+\pi^-$ decays [149]. The correction factor tables for lepton identification efficiency and misidentification rate are defined for a fixed set of selections whereas the hadron identification and $p \rightarrow e$ fake rate correction tables can be computed automatically for a custom set of constraints and momentum and polar angle bin edges. The correction factors, which combine efficiency and misidentification correction, are typically in the range 0.95 to 1.02 for kaons and 0.90 to 1.03 for leptons. In the rare case of an ROE lepton track not covered by the momentum and polar angle range, its correction factor is set to unity and the uncertainty is computed from the average correction factor uncertainty and difference from unity. This occurs in less than 0.01% of electron tracks and in about 1% of muon tracks, mostly with momentum below 200 MeV and polar angle below 23° . The efficiency and misidentification correction factors of $\pi^\pm$ with momenta larger than 200 MeV are only applied in the $\pi^\pm$ multiplicity distribution used to extract an uncertainty on the X_u fragmentation as described in Section 5.9.4.

The $K^\pm$ veto used in the signal region definition in simulation is not applied as a binary selection but instead using the corresponding efficiency calibration factors. If no $K^\pm$ is identified in an event, it is assigned a weight of 1 but if a $K^\pm$ is identified and the event is assigned an identification calibration factor w , then the event is given a weight of $1 - w$. This way, events without an identified $K^\pm$ are not rejected but $K^\pm$ events are kept with a small weight (since calibration factors are typically close to unity). This method can be illustrated with a simple example: if, for a given $K^\pm$ identification selection, the $K^\pm$ identification efficiency is equal to 80% in simulated data and 60% in data, the associated calibration factor is 0.75 and the veto weight is 0.25, indicating a probability of 25% for $K^\pm$ to not be properly identified as such and therefore for events to not be rejected by the $K^\pm$ veto. If events where a $K^\pm$ is found were simply rejected, then no $K^\pm$ identification efficiency would be applied. In the figures shown throughout this chapter, the $K^\pm$ calibration factors are only applied when a selection on the $K^\pm$ multiplicity is applied. Events with lepton candidates in the ROE are also rejected with the same veto method using the lepton identification calibration factors.

5.4.7 K_S^0 reconstruction efficiency

The K_S^0 reconstruction efficiency was studied in an independent dedicated analysis of the $D^{*+} \rightarrow (D^0 \rightarrow K_S^0\pi^+\pi^-)\pi^+$ decay channel [150]. A correction factor of 0.55% per cm of flight distance from the interaction point was extracted and it is used to downscale events where a K_S^0 is found. The K_S^0 veto is applied in the same way as the $K^\pm$ veto described in the previous section and the calibration factors are only applied when a selection on the K_S^0 multiplicity is applied.

5.4.8 Slow pion reconstruction efficiency

The low momentum pion reconstruction efficiency mismatch between data and simulated data is calibrated using factors extracted from independent Belle II studies for $\pi^\pm$ and π^0 separately [140, 141]. The calibration factors are extracted in three momentum bins between 50 and 200 MeV. Charged slow pions were studied in the $B^0 \rightarrow (D^{*-} \rightarrow D^0\pi_{\text{slow}}^+)\pi^+$ decay channel

and the neutral slow pions in the $B^+ \rightarrow (D^{*0} \rightarrow D^0 \pi_{\text{slow}}^0) \pi^+$ decay channel. The slow $\pi^\pm$ calibration factors are applied to all signal-side tracks with a momentum lower than 200 MeV as no dedicated track finding efficiency calibration has been extracted for this energy range⁶. The slow pion efficiency variables are only used in the $B \rightarrow X_c \ell \nu$ suppression classifier (*c.f.* section 5.5.2) and the associated calibration factors are therefore only applied if a selection on the output classifier score is applied.

5.5 Machine learning selection

We train two separate neural networks (NN) using the `TensorFlow` package [151] with hyperparameters optimised using the `Optuna` package. The NN are trained to distinguish signal events from continuum ($e^+e^- \rightarrow q\bar{q}$, $q \in \{u, d, s, c\}$) and $B \rightarrow X_c \ell \nu$ events respectively.

5.5.1 Continuum suppression

As illustrated in Figure 3.2, the total cross-section of $e^+e^- \rightarrow q\bar{q}$ events is more than three times higher than the $e^+e^- \rightarrow \Upsilon(4S)$ cross-section but the preselections described in the previous section significantly reduce the background coming from such events. However, as shown in Figure 5.13, the continuum component remains an important source of background, in particular at low lepton momentum. Due to the lower mass of hadrons composed of u, d, s and c quarks compared to B mesons, the properties of continuum events differ largely from those of $B\bar{B}$ events. In continuum events, light mesons are boosted, thus producing two back-to-back jet-like structures. On the other hand, B mesons from a $\Upsilon(4S)$ decay are produced nearly at rest in the centre-of-mass frame, resulting in more uniformly distributed decay products, which is illustrated in Figure 5.9. The event shape differences between continuum and $B\bar{B}$ can therefore be exploited to distinguish them and therefore suppress the continuum background. The variables used for that purpose are described in detail in Ref. [152]. We give a short outline of these variables in the following:

- **Thrust-axis:** for a group of N particles with momenta $\vec{p}_i$ ($i = 1, \dots, N$), the thrust axis $\vec{\mathcal{T}}$ is defined as the unit vector along which their total projection is maximised. The thrust (scalar) $\mathcal{T}$ is defined as

$$\mathcal{T} = \frac{\sum_{i=1}^N |\vec{\mathcal{T}} \cdot \vec{p}_i|}{\sum_{i=1}^N |\vec{p}_i|}. \quad (5.8)$$

We define $\cos \theta_{\mathcal{T}_B \mathcal{T}_O}$ and $\cos \theta_{\mathcal{T}_B z}$ as the cosines of the angle between the thrust axis of the tag-side B and the thrust axis of the ROE or the z -axis, respectively. In addition, we define $\mathcal{T}_B$ and $\mathcal{T}_O$ as the thrust scalars of the B_{tag} and the system recoiling against the B_{tag} (so the B_{sig} which in our case is composed of the ROE and the signal lepton), respectively.

- **Fox-Wolfram moments:** the k -th order Fox-Wolfram (FW) moment H_k is defined as [153]:

$$H_k = \sum_{i,j}^N |\vec{p}_i| |\vec{p}_j| P_k(\cos \theta_{ij}), \quad (5.9)$$

where θ_{ij} is the angle between the momenta $\vec{p}_i$ and $\vec{p}_j$ and P_k is the k -th order Legendre polynomial. The FW moments can be normalised to the first moment H_0 , thereby defining ratios R_k of which R_2 is the most effective at separating continuum from $B\bar{B}$ events. For our analysis, we use a set of modified FW moments, originally developed for Belle analyses

⁶In the signal region, since events with kaons and leptons found in the ROE are rejected, all surviving tracks are *a priori* pion tracks.

(see Ref. [154] for example) and called the Kakuno-Super-Fox-Wolfram (KSWF) moments $\mathcal{H}_k$. The KSWF moments are denoted with the superscript *so* if one of the two particles comes from the B_{tag} and the other one from the B_{sig} and *oo* if both particles come from the B_{sig} . In the *so* KSWF moments calculation, charged particles, neutral particles and missing momentum are treated separately. The moments will be denoted with the subscripts *c*, *n* and *m* for each one of these three categories, respectively. The KSWF moments considered in Belle II analyses are: $\mathcal{H}_{c0}^{so}$, $\mathcal{H}_{c1}^{so}$, $\mathcal{H}_{c2}^{so}$, $\mathcal{H}_{c3}^{so}$, $\mathcal{H}_{c4}^{so}$, $\mathcal{H}_{n0}^{so}$, $\mathcal{H}_{n2}^{so}$, $\mathcal{H}_{n4}^{so}$, $\mathcal{H}_{m0}^{so}$, $\mathcal{H}_{m2}^{so}$, $\mathcal{H}_{m4}^{so}$, $\mathcal{H}_0^{oo}$, $\mathcal{H}_1^{oo}$, $\mathcal{H}_2^{oo}$, $\mathcal{H}_3^{oo}$ and $\mathcal{H}_4^{oo}$.

- **CLEO cones:** the CLEO cones are a set of variables introduced by the CLEO collaboration in Ref. [155]. Starting from the thrust axis of the B_{tag} , the momenta of all particles in intervals of 10° each are summed up, thus defining nine CLEO cone variables denoted $\mathcal{C}_1, \dots, \mathcal{C}_9$.

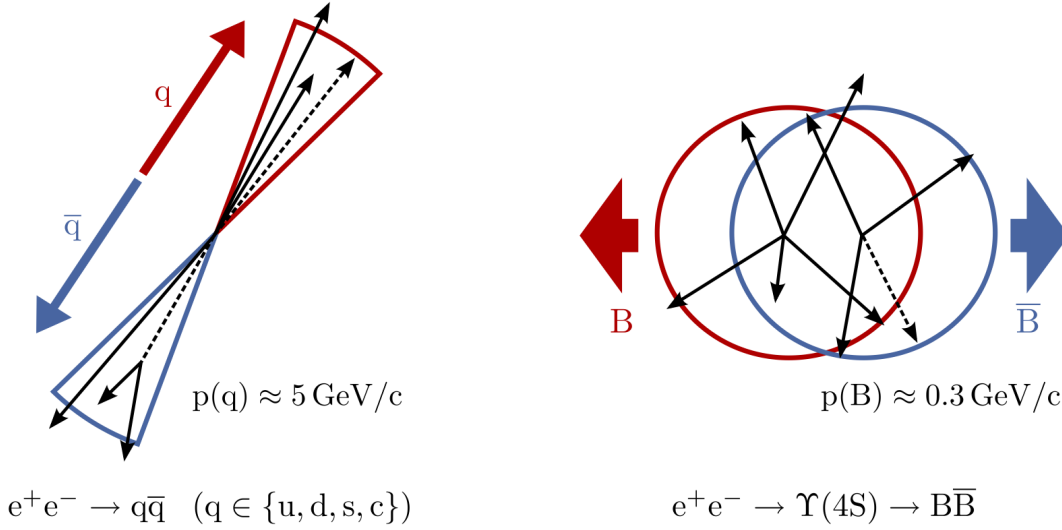


FIGURE 5.9: Illustration of the difference in event shape between continuum events (left) and $B\bar{B}$ events (right). Schematic taken from Ref. [156].

In addition, the transverse energy $E_{t,\text{CS}}$, the missing mass squared $M_{\text{miss,CS}}^2$ and the number of charged tracks N_{tracks} and neutral clusters $N_{\text{neu.clusters}}$ offer a good separation power between continuum and signal events. The average track and cluster multiplicities of signal events before applying the continuum suppression are about 7.6 and 8.5, respectively, whereas, for continuum events, they are about 8.3 and 11.1, respectively. The distributions of the 34 variables introduced here are illustrated in Appendix B. The ROE selections used in the FEI calibration study are applied to the distributions used for the continuum suppression, which, in particular, differentiates $M_{\text{miss,CS}}^2$ from M_{miss}^2 (see Section 5.3.2 for the exact definitions of these selections). Based on the agreement between data and simulated data distributions, a subset of 14 variables is kept for the NN training: $\mathcal{C}_2$, $\cos\theta_{\mathcal{T}_B\mathcal{T}_O}$, $\cos\theta_{\mathcal{T}_Bz}$, $\mathcal{H}_{c1}^{so}$, $\mathcal{H}_{c2}^{so}$, $\mathcal{H}_{c3}^{so}$, $\mathcal{H}_{c4}^{so}$, $\mathcal{H}_{n2}^{so}$, $\mathcal{H}_{n4}^{so}$, $\mathcal{H}_{m0}^{so}$, $\mathcal{H}_1^{oo}$, $\mathcal{H}_3^{oo}$, $\mathcal{H}_4^{oo}$ and N_{tracks} . To train the NN, a small fraction of the continuum and $B \rightarrow X_u \ell \nu$ samples described in section 5.2 is used. The optimised hyperparameter values used for the trained NN are summarised in Table 5.5 and the output score is shown in Figure 5.10. The performance of the classifier can be evaluated using various metrics. For instance, it is crucial to assess whether the NN has been overtrained. This can be done using the receiver operating characteristic (ROC) curve. From the ROC curve shown in Figure 5.11, it can be seen that the classifier performance is similar when applied to the training and test samples. Using the **Shap** mean absolute value,

the importance of each of the 14 variables can be quantified. As illustrated in Figure 5.12, $\mathcal{H}_{m0}^{so}$, $\mathcal{H}_{c2}^{so}$, $\cos \theta_{\mathcal{T}_B \mathcal{T}_O}$ and N_{tracks} are the variables offering the largest separation power. The optimal continuum suppression classifier score is chosen by scanning a set of scores over the full possible range and choosing the score which maximises the significance defined as $S/\sqrt{S+B}$ where S is the number of $B \rightarrow X_u \ell \nu$ events and B the number of continuum events surviving the selection. The optimal score is found to be 0.80, which rejects about 95% of continuum events and 32% of signal events. Applying this selection defines a phase-space region which will be referred to as $B \rightarrow X \ell \nu$ control region ($\text{CR}_{X\ell\nu}$). The rejection of the continuum using the classifier is illustrated in the q^2 distribution in Figure 5.13.

Parameter	Value	Range	Description
N Layers	2	[2, 4]	Number of layers in the feed forward network.
Layer Size	[256, 128]	{8, 16, 32, 64, 128, 256}	Number of neurons in each layer of the network. The number is independently selected for each layer and is restricted to a power of two.
Activation Function	Swish	-	We use the swish [157] activation function of the form $f(x) = \frac{x}{1+\exp(-x)}$. Tests with ReLU showed similar performance, indicating the network is largely agnostic to the specific activation function.
Optimiser	AdamW	-	We adopt the AdamW [158] optimiser, which computes adaptive learning rates for each parameter and has demonstrated rapid convergence in few epochs. Tests using the SGD optimiser converged on similar models.
Learning Rate	$9.3 \cdot 10^{-3}$	$[10^{-5}, 10^{-2}]$	Learning rate of the model. Governs the rate of the gradient descent.
Weight Decay	$9.6 \cdot 10^{-4}$	$[10^{-8}, 10^{-3}]$	Strength of the weight decay regularisation. Helps prevent overfitting by penalising overly complex models.
Dropout Rate	0.25	-	We add a dropout layer after each dense layer. This parameter governs the rate at which the dropout layers randomly drop a fraction of input connections. The remaining connections are rescaled to compensate. Dropout is used to reduce overtraining of the classifier, so its value is fixed.
Label Smoothing	$8.8 \cdot 10^{-5}$	$[10^{-5}, 0.05]$	Slightly smooths the labels by changing them from binary, 1 or 0 for signal and background, to $1-s$ and s . This Helps prevent overfitting and avoiding issues with many events being assigned a score of exactly 1 or 0 due to floating point precision.
Batch Size	4096	{1024, 2048, 4096}	Size of a minibatch for each epoch of training.

TABLE 5.5: Hyperparameters of the continuum suppression NN. All hyperparameters not mentioned are left at their default values as of version 2.17.0 of **Tensorflow**. The classifier is trained over 200 epochs.

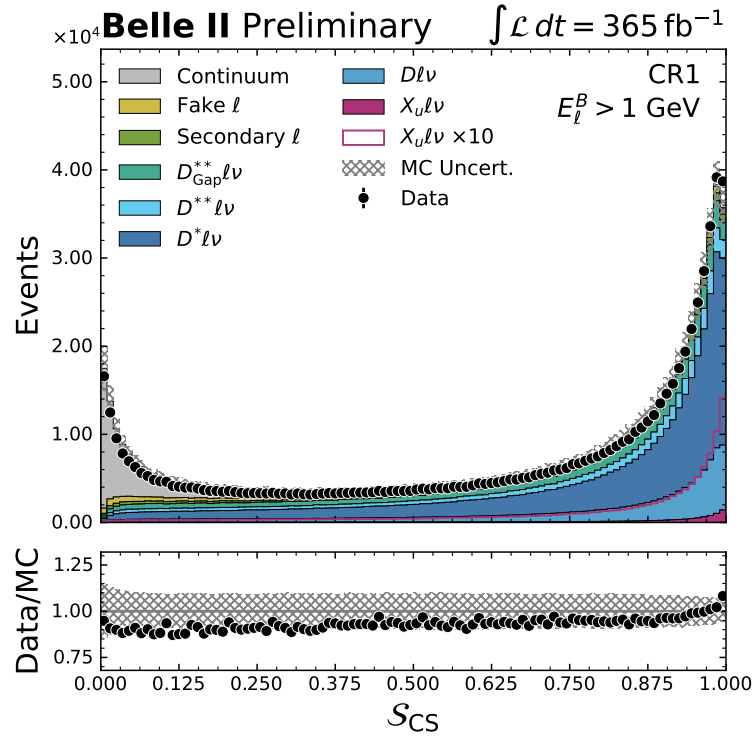


FIGURE 5.10: Output score of the continuum suppression classifier.

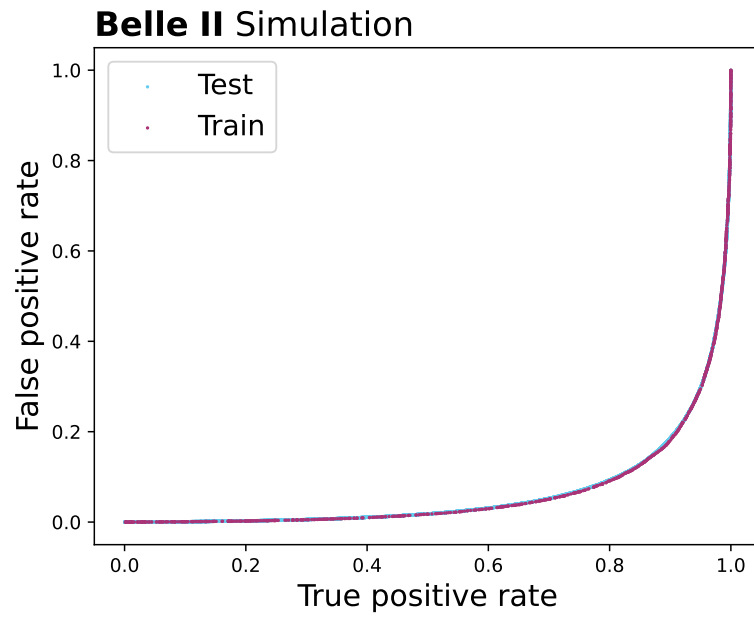


FIGURE 5.11: Receiver operating characteristic curve of the continuum suppression classifier score.

Belle II Simulation

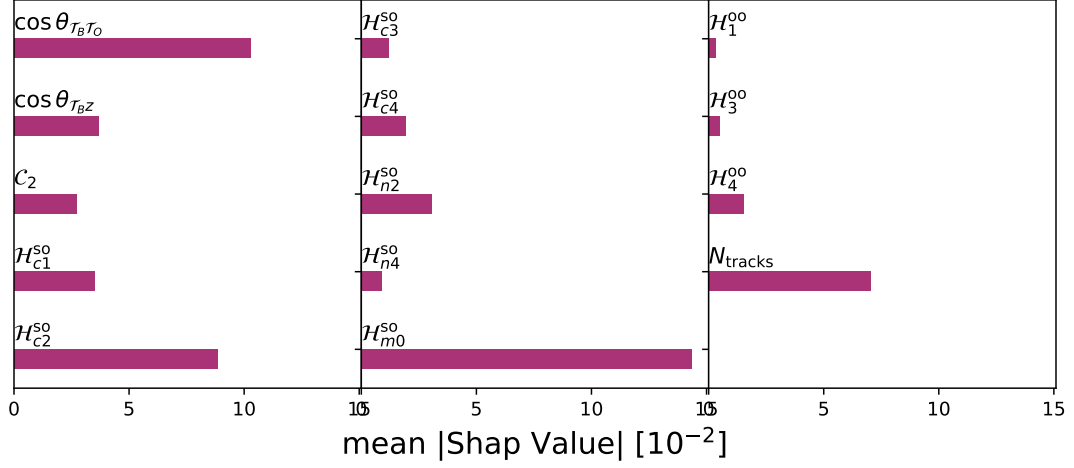


FIGURE 5.12: Importances of input features in the continuum suppression classifier training measured by Shap.

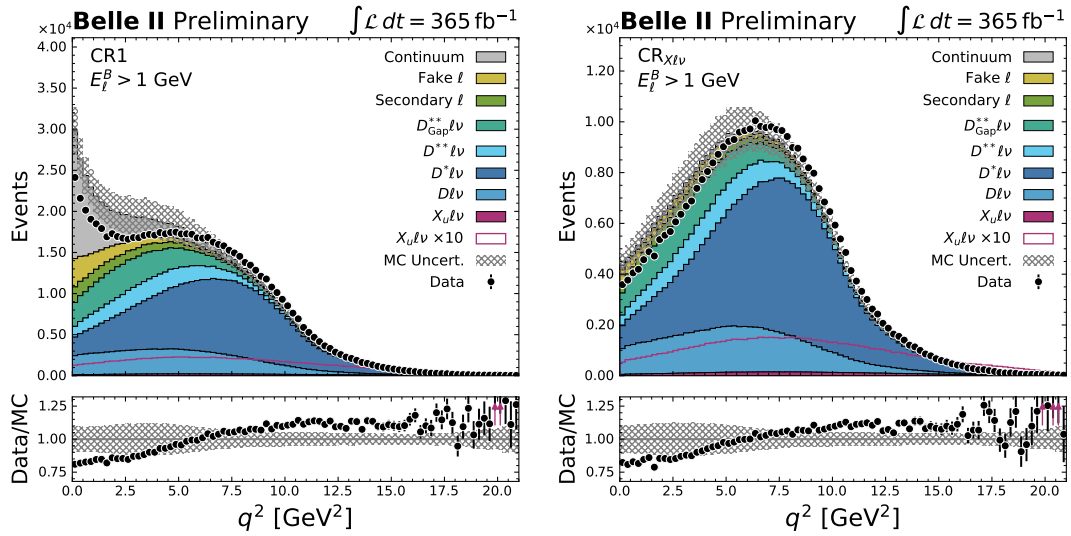


FIGURE 5.13: Comparison of the q^2 distribution before (left) and after (right) applying the continuum suppression on the preselected sample. Only the MC statistical uncertainty is included in the error band.

5.5.2 $B \rightarrow X_c \ell \nu$ suppression

To suppress the overwhelming $B \rightarrow X_c \ell \nu$ background, we train a separate NN using a fraction of the $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ samples on which the preselection, the continuum suppression and the kaon veto have been applied. The following nine variables are used to distinguish between signal and $B \rightarrow X_c \ell \nu$ events:

- the event missing mass squared, M_{miss}^2 ,
- the total event charge, Q_{tot} ,
- the p -value of the ROE vertex fit, $p\text{-val}(\text{ROE})$,
- when a D^{*+} or D^{*0} has been inclusively reconstructed from a charged or neutral slow pion:
 - missing mass squared $M_{\text{miss}}^2(D^* \rightarrow D\pi)$ as defined in equation 5.6,
 - cosine of the angle between the flight directions of the reconstructed D^* and the signal lepton in the centre-of-mass frame, $\cos \theta_c(D^* \rightarrow D\pi)$,
 - cosine of the angle between the momenta in the centre-of-mass frame of a B candidate reconstructed from the D^* and that of a nominal B (defined as a particle with a B meson mass and energy equal to half of the beam energy), $\cos \theta_{BY}(D^* \rightarrow D\pi)$ ⁷.

The average track and cluster multiplicities of $B \rightarrow X_c \ell \nu$ events are about 7.6 and 9.8, respectively, whereas for signal events, they are about 7.4 and 8.3. Since the average particle multiplicity of $B \rightarrow X_c \ell \nu$ events is larger than for $B \rightarrow X_u \ell \nu$ events, the event reconstruction quality of charmed decays is expected to be poorer because of missed particles or because of the limited detector resolution. The first three variables therefore help reject poorly reconstructed $B \rightarrow X_c \ell \nu$ events which tend to populate the high M_{miss}^2 tail, have a $p\text{-val}(\text{ROE})$ close to zero and a Q_{tot} different than zero.

As mentioned in Section 5.3.3, the low-momentum pion variables allow us to identify events with a D^* decay. For well reconstructed $B \rightarrow D^*(\rightarrow D\pi)\ell\nu$ events, $M_{\text{miss}}^2(D^* \rightarrow D\pi)$ is expected to be consistent with zero, $\cos \theta_c(D^* \rightarrow D\pi)$ is expected to be close to -1 and $\cos \theta_{BY}(D^* \rightarrow D\pi)$ is expected to be close to 1. The NN input variables are illustrated in Appendix C and the NN hyperparameters are summarised in Table 5.6. The output score, the ROC curve and the input variable importances are shown in Figures 5.14 to 5.16. The latter shows that M_{miss}^2 offers the highest separation power between $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ events.

The irregularities in the classifier score distribution are due to peculiarities of the input features, in particular the slow pion variables. When no slow pion is found in the event, the variable is assigned a fixed value far away from the bulk of the distribution. Moreover, Q_{tot} is a discrete variable and $p\text{-val}(\text{ROE})$ exhibits three spikes at -1 (failed vertex fits), 0 (very poor fits) and 1 (very good fits). The combination of these factors explains the irregular distribution of $\mathcal{S}_{X_c}$ at low scores and scores around $0.71 - 0.75$. The optimal score is chosen by scanning a set of scores over the full range and by selecting the score which minimises the total uncertainty of the $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ extracted from a fit to the two-dimensional variable $E_\ell^B:q^2$ with Asimov data (this fit is referred to as fit 1; see Section 5.9 for a detailed description of the signal extraction setup). The optimal score is found to be 0.87 , which rejects about 98% of $B \rightarrow X_c \ell \nu$ and 74% of $B \rightarrow X_u \ell \nu$ events from the sample on which the preselections, the continuum suppression and the kaon veto are applied. The classifier requirement tends to reject $B \rightarrow D \ell \nu$ events at a slightly lower rate ($\sim 97\%$) than other $B \rightarrow X_c \ell \nu$ decays ($\sim 99\%$), as $B \rightarrow D^* \ell \nu$ events are well

⁷This variable is expected to be between -1 and 1 only if a massless particle such as a neutrino is missing in the reconstructed event.

suppressed because of the slow pion variables and $B \rightarrow D^{**} \ell \nu$ events are reconstructed with lower accuracy. The classifier also tends to select a higher fraction of $B \rightarrow X_u \ell \nu$ events with low multiplicities, such as $B \rightarrow \pi \ell \nu$ events. These are rejected at a rate of about 40%, whereas other resonant and non-resonant events are rejected at rates of about 60% to 80%. See Figure 5.38 for kinematic distributions on which the $B \rightarrow X_c \ell \nu$ suppression classifier has been applied.

Parameter	Value	Range	Description
N Layers	4	[2, 4]	Number of layers in the feed forward network.
Layer Size	[64, 32, 128, 128]	{8, 16, 32, 64, 128, 256}	Number of neurons in each layer of the network. The number is independently selected for each layer and is restricted to a power of two.
Activation Function	Swish	-	We use the swish [157] activation function of the form $f(x) = \frac{x}{1+\exp(-x)}$. Tests with ReLU showed similar performance, indicating the network is largely agnostic to the specific activation function.
Optimiser	AdamW	-	We adopt the AdamW [158] optimiser, which computes adaptive learning rates for each parameter and has demonstrated rapid convergence in a few epochs. Tests using the SGD optimiser converged on similar models.
Learning Rate	$3.3 \cdot 10^{-3}$	$[10^{-5}, 10^{-2}]$	Learning rate of the model. Governs the rate of the gradient descent.
Weight Decay	$1.2 \cdot 10^{-5}$	$[10^{-8}, 10^{-3}]$	Strength of the weight decay regularisation. Helps prevent overfitting by penalising overly complex models.
Dropout Rate	0.40	-	We add a dropout layer after each dense layer. This parameter governs the rate at which the dropout layers randomly drop a fraction of input connections. The remaining connections are rescaled to compensate. Dropout is used to reduce overtraining of the classifier, so its value is fixed.
Label Smoothing (s)	$7.1 \cdot 10^{-3}$	$[10^{-5}, 0.05]$	Slightly smooths the labels by changing them from binary, 1 or 0 for signal and background, to $1 - s$ and s . This Helps prevent overfitting and avoiding issues with many events being assigned a score of exactly 1 or 0 due to floating point precision.
Batch Size	4096	{1024, 2048, 4096}	Size of a minibatch for each epoch of training.
L1/L2 Regularisers	0.01, 0.01	-	L1 and L2 regularisations are penalty terms to the loss function. They reduce the complexity of the model, which helps avoid overtraining.

TABLE 5.6: Hyperparameters of the $B \rightarrow X_c \ell \nu$ suppression neural network. All hyperparameters not mentioned are left at their default values as of version 2.17.0 of **Tensorflow**. The classifier is trained over 128 epochs.

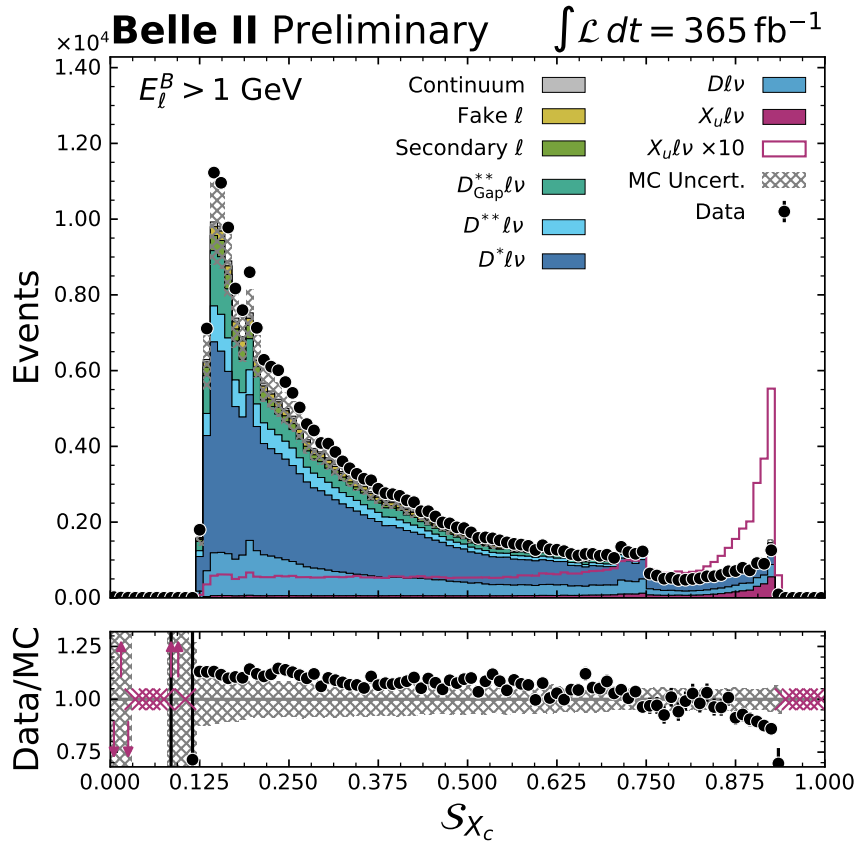


FIGURE 5.14: $B \rightarrow X_c \ell \nu$ suppression classifier output score in the preselected region on which the continuum suppression and kaon veto have been applied.

Classifier induced model dependence

The selections based on the two classifier output scores can introduce undesired modelling dependence in the variables used in the signal extraction if these variables are strongly correlated with the output scores. The first order correlation between E_ℓ^B , q^2 and M_X on one hand and the classifier scores on the other is at most ~ 0.3 in absolute value which indicates that the correlation is relatively weak. We nevertheless attempted to reduce the correlation by introducing in the NN losses of each classifier an additional decorrelation term called DisCo which was introduced in Ref. [159]. The strength of the decorrelation can be tuned by a single parameter λ_{DisCo} . Each classifier was trained with different values of λ_{DisCo} from 1 to 50 and a significant decrease in the correlation was only observed for large values of the parameter. However, the performance of classifiers trained with $\lambda_{\text{DisCo}} > 0$ worsens: the signal efficiency and background rejection decrease, the classifiers are slightly overtrained despite optimising the hyperparameters and large peaking structures whose origin is not understood appear in the output score distributions. For instance, when adding the DisCo loss term in the $B \rightarrow X_c \ell \nu$ suppression classifier with a strength of $\lambda_{\text{DisCo}} = 10$, the correlation between the output score and E_ℓ^B shifts from 0.30 to 0.20 and the correlation between the output score and q^2 shifts from 0.15 to 0.05 but the background rejection decreases by about 15% when keeping the signal efficiency fixed. In summary, the application of the DisCo algorithm likely reduces the model dependence introduced by the classifiers but it also results in a less stable performance and poorer background rejection. Consequently, we opted for a more conventional implementation of the neural network classifier by setting λ_{DisCo} to zero.

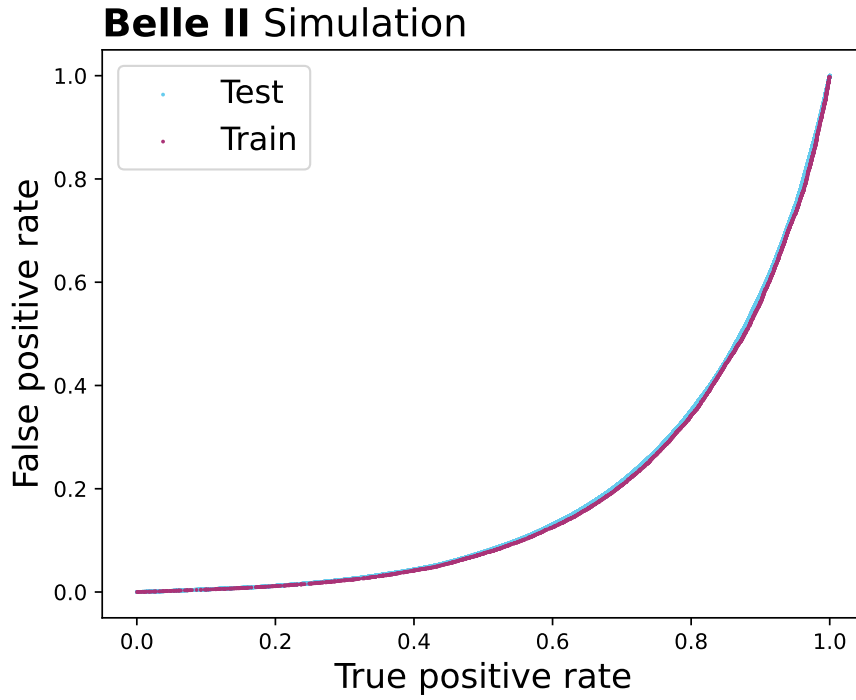


FIGURE 5.15: Receiver operating characteristic curve of the $B \rightarrow X_c \ell \nu$ suppression classifier score.

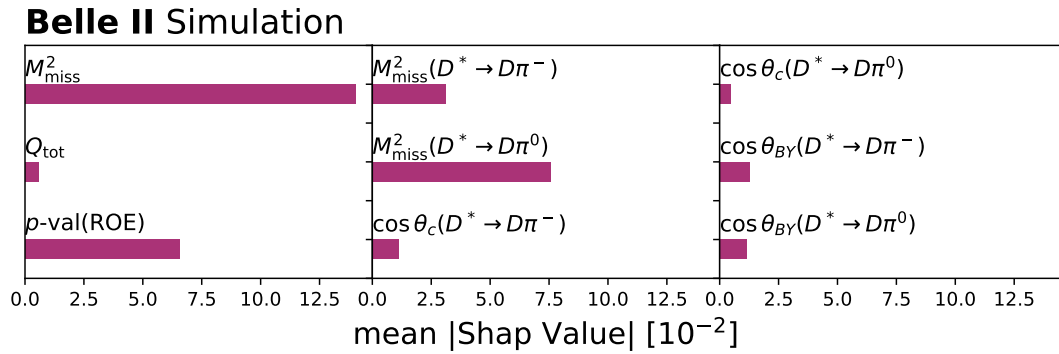


FIGURE 5.16: Importances of input features in the $B \rightarrow X_c \ell \nu$ suppression classifier training measured by Shap.

5.6 Kinematic selections

In addition to all the selections discussed so far, selections on the three kinematic variables E_ℓ^B , M_X and q^2 can be used to further reject background sources. In the following, a selection of $E_\ell^B > 1$ GeV is applied to all samples as it rejects a portion of phase-space mainly populated by continuum and fake and secondary lepton events. Furthermore, a region where an $M_X < 1.7$ GeV selection is applied is also explored. The mass of the lightest charm resonance – the D meson – being around 1.8 GeV, this selection rejects very efficiently $B \rightarrow X_c \ell \nu$ decays. Finally, a region where a $q^2 > 8$ GeV² selection is applied, in addition to the E_ℓ^B and M_X selections, is also considered. The three signal regions used to measure $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ are defined by applying all preselections, the continuum suppression, the $B \rightarrow X_c \ell \nu$ suppression and one of the three sets of kinematic selections. The signal phase-space acceptance, *i.e.* the covered fraction of accessible $B \rightarrow X_u \ell \nu$ phase-space, for the three sets of kinematic selections is given in Table 5.7. Inclusive

$B \rightarrow X_u \ell \nu$ models are most reliable in the broadest phase-space region but the additional selections on M_X and q^2 increase the signal purity, to further test the theoretical frameworks and to compare their respective predictions.

Kinematic selections	Acceptance
$E_\ell^B > 1 \text{ GeV}$	87%
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$	57%
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$ $q^2 > 8 \text{ GeV}^2$	31%

TABLE 5.7: $B \rightarrow X_u \ell \nu$ phase-space acceptance for the three sets of kinematic selections used to define the signal regions.

5.7 Signal region

We discuss here the most important features of the signal region as defined by the selections presented in the previous sections. The discussion is based exclusively on simulated samples. Distributions of experimental data in the signal region are shown in Section 5.10 and 5.11.

The number of events found in the signal region and the fractions of signal and background when applying the three sets of kinematic selections are given in Table 5.8.

Kinematic selections	Number of events	$B \rightarrow X_u \ell \nu$	$B \rightarrow X_c \ell \nu$	Other backgrounds
$E_\ell^B > 1 \text{ GeV}$	6082	$(30.0 \pm 0.4)\%$	$(60.9 \pm 0.5)\%$	$(9.1 \pm 0.4)\%$
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$	2418	$(61.0 \pm 0.6)\%$	$(33.6 \pm 0.5)\%$	$(5.4 \pm 0.4)\%$
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$ $q^2 > 8 \text{ GeV}^2$	1411	$(67.7 \pm 0.7)\%$	$(28.0 \pm 0.6)\%$	$(4.3 \pm 0.6)\%$

TABLE 5.8: Total number of simulated events found in the signal region when applying one of the three sets of kinematic selections. In addition, the fraction of $B \rightarrow X_u \ell \nu$, $B \rightarrow X_c \ell \nu$ and other background events is given. The statistical uncertainty is given for the fractions.

It is instructive to study the composition of the signal region sample in terms of B meson charge, lepton flavour and $B \rightarrow X_u \ell \nu$ modes. Due to the higher efficiency of the muon identification requirement compared to the requirement for electrons (see Section 5.3.3), there is a slight excess of events where the signal lepton candidate has been reconstructed as a muon. In the signal region, $(56.0 \pm 0.4)\%$ of events are muon events and $(44.0 \pm 0.3)\%$ are electron events, where the uncertainties are statistical. However, because a larger fraction of muons than electrons are misidentified particles (mainly pions), about 80% of non-semileptonic background events fall in the muon channel. This difference is illustrated in Figure 5.17 where the q^2 distribution in the signal region is compared for muon and electron events. Due to the relatively

small non-semileptonic background component in the electron channel, the signal extraction setup described in Section 5.9 is unstable when the fit is split by lepton flavour. For this reason, we do not measure the $B \rightarrow X_u e \nu$ and $B \rightarrow X_u \mu \nu$ partial branching fractions separately.

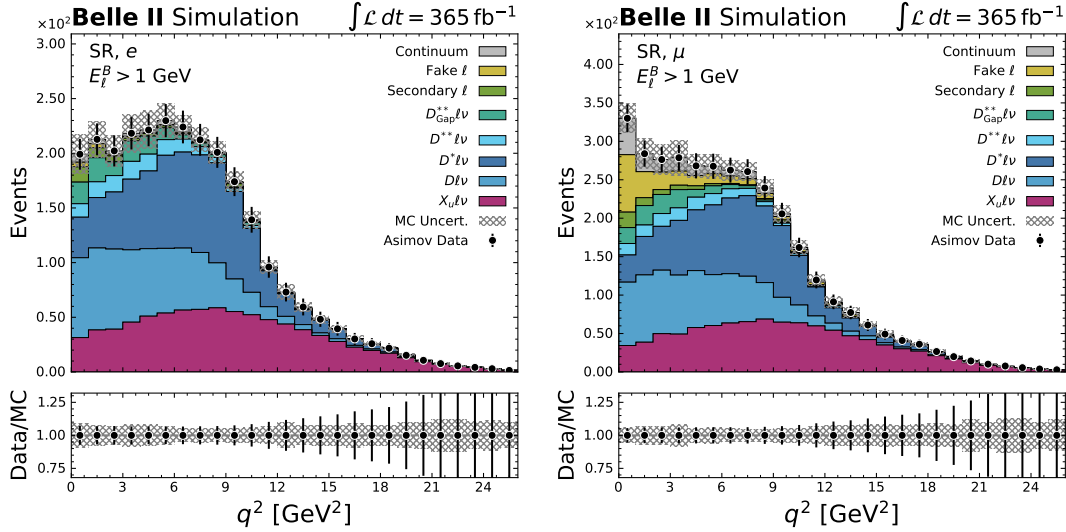


FIGURE 5.17: Simulated q^2 distribution in the signal region where the signal lepton candidate has been identified as an electron (left) or a muon (right). The $E_\ell^B > 1$ GeV requirement is applied.

In the signal region, the fraction of events with a charged B candidate is significantly higher than that of events with a neutral B candidate. The former represent about $(73.4 \pm 0.4)\%$ of events and the latter $(26.6 \pm 0.3)\%$ of events. This difference is caused by several effects. First of all, the branching fractions of charmed and charmless B^+ semileptonic decays are slightly higher than those of B^0 semileptonic events (see Tables 5.1 and 5.2). Second, as mentioned in Section 5.3.3, the signal lepton requirements reject $B^0 - \bar{B}^0$ events where the signal B flavour has changed. These mixed- B^0 events represent about 20% of $B^0 - \bar{B}^0$ events. Third, the average multiplicity of charged tracks is larger in B^0 events than B^+ events [160]. The reconstruction quality of B^0 events is therefore expected to be slightly worse than that of B^+ events because of missed particles or misreconstructed particle trajectories. This effect can be observed in the $p\text{-val}(\text{ROE})$ and Q_{tot} variables used to train the $B \rightarrow X_c \ell \nu$ suppression classifier. The effect is illustrated in Figure 5.18 for the $p\text{-val}(\text{ROE})$ distribution. Finally, since the decay chain $D^{*0} \rightarrow D^+ \pi^-$ is kinematically forbidden, the identification of $B^+ \rightarrow D^{*0} \ell \nu$ decays can occur only via neutral pions whereas $B^0 \rightarrow D^{*+} (\rightarrow D \pi) \ell \nu$ events can be identified via either charged or neutral pions. Therefore, the input features of the $B \rightarrow X_c \ell \nu$ suppression classifier which rely on the reconstruction of a slow π^+ help rejecting $B^0 \rightarrow D^{*+} \ell \nu$ events but not $B^+ \rightarrow D^{*0} \ell \nu$ events as illustrated in Figure 5.18. Since the suppression of $B \rightarrow X_c \ell \nu$ events is based on the identification of poorly reconstructed events and the reconstruction of D^* mesons from low-momentum pions (see Section 5.5.2), the $B \rightarrow X_c \ell \nu$ suppression classifier tends to reject a larger fraction of B^0 events than B^+ events leading to the unbalance observed in the signal region.

In the simulated mix of $B \rightarrow X_u \ell \nu$ events, non-resonant decays represent roughly three quarters of events and it is essential to keep the relative fractions of $B \rightarrow X_u \ell \nu$ decay modes close to values which are representative of this mix. It is important to note that, since the hadronic mass and momentum transfer of non-resonant events are spread out across the full accessible range, the M_X and q^2 selections reject relatively larger fractions of such events. Moreover, as was highlighted in Section 5.5.2, the $B \rightarrow X_c \ell \nu$ suppression classifier tends to select events with

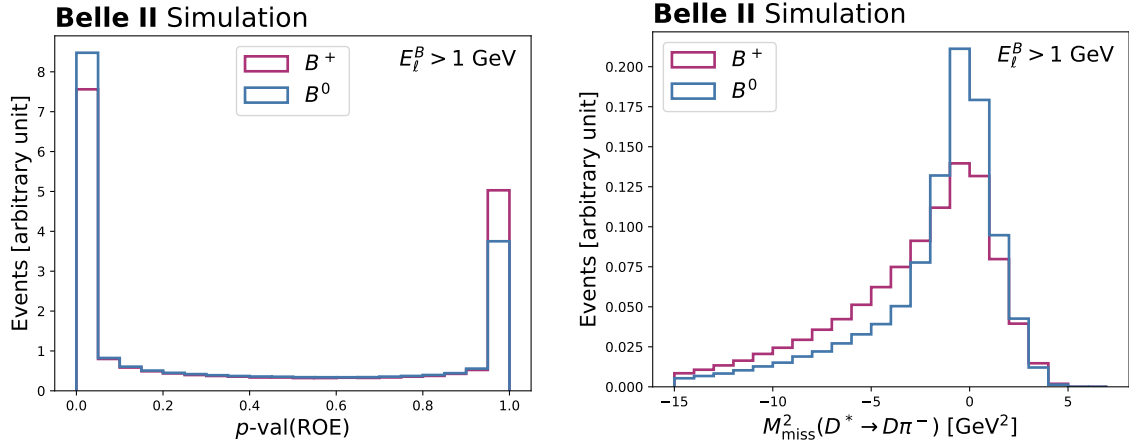


FIGURE 5.18: Distribution of two distributions, $p\text{-val}(\text{ROE})$ (left) and $M_{\text{miss}}^2(D^* \rightarrow D\pi^-)$ (right), used as input features in the $B \rightarrow X_c \ell \nu$ suppression classifier. The two distributions are shown separately for B^+ (purple) and B^0 (blue) events. The distributions can be compared to the histograms shown in Appendix C.

lower multiplicities and the signal region is therefore enriched in resonant decays, in particular $B \rightarrow \pi \ell \nu$ decays. The reconstruction efficiency of simulated $B \rightarrow X_u \ell \nu$ events for different components in the three different signal regions is shown in Table 5.9. The efficiency of $B \rightarrow X_u \ell \nu$ events with $E_\ell^B > 1$ GeV is also shown as a function of the three kinematic variables of interest at the generator level in Figure 5.19. We show in Table 5.10 the fractions of different types of $B \rightarrow X_u \ell \nu$ events in the three signal regions and we confirm that non-resonant events correspond to a representative fraction of all $B \rightarrow X_u \ell \nu$ events. In addition, the E_ℓ^B , M_X and q^2 distributions in the broadest signal region with the $B \rightarrow X_u \ell \nu$ component broken down in different decay modes are shown in Figure 5.20.

Kinematic selections	$B \rightarrow \pi \ell \nu$	$B \rightarrow \rho \ell \nu$	$B \rightarrow \omega \ell \nu$	$B \rightarrow \eta/\eta' \ell \nu$	Res.	Non-res.	All $B \rightarrow X_u \ell \nu$
$E_\ell^B > 1$ GeV	0.67%	0.41%	0.32%	0.36%	0.46%	0.17%	0.23%
$E_\ell^B > 1$ GeV $M_X < 1.7$ GeV	0.67%	0.40%	0.31%	0.35%	0.45%	0.11%	0.19%
$E_\ell^B > 1$ GeV $M_X < 1.7$ GeV $q^2 > 8$ GeV ²	0.47%	0.28%	0.22%	0.20%	0.32%	0.06%	0.12%

TABLE 5.9: Signal region reconstruction and selection efficiency of the simulated $B \rightarrow X_u \ell \nu$ sample split by component for the three sets of kinematic selections. The statistical uncertainty is negligible.

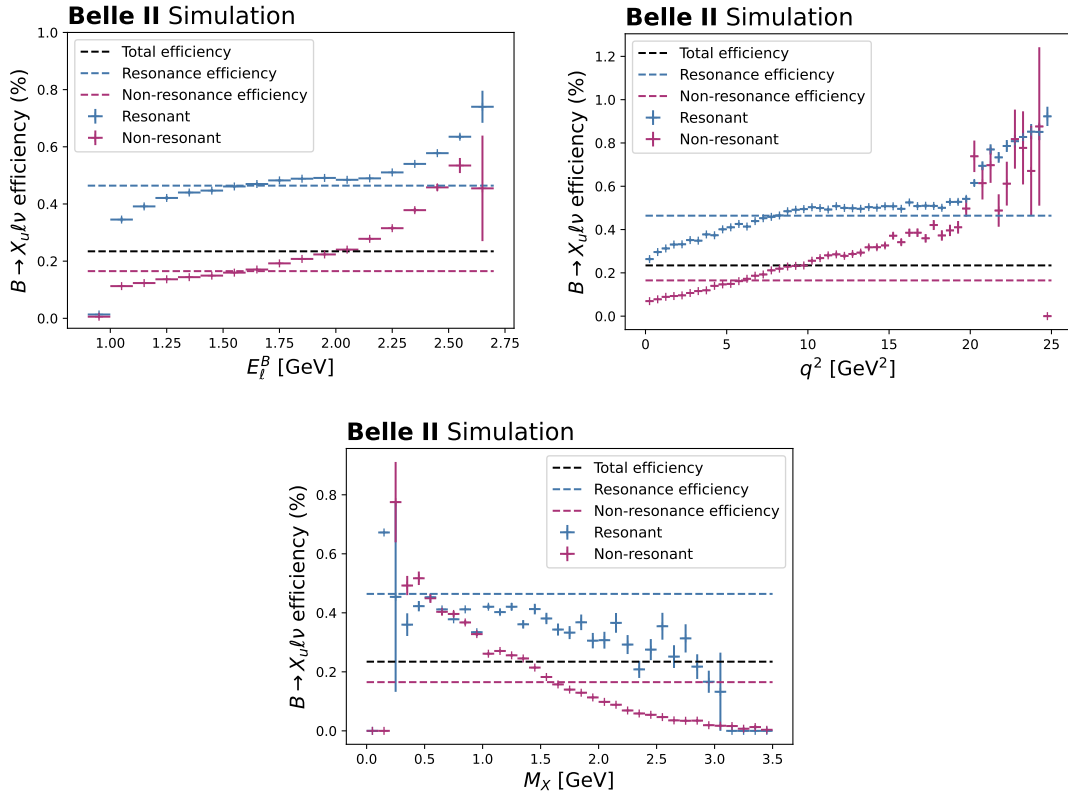


FIGURE 5.19: Signal region reconstruction and selection efficiency of resonant (blue) and non-resonant (purple) $B \rightarrow X_u \ell \nu$ events as a function of E_ℓ^B (top left), q^2 (top right) and M_X (bottom). Generator-level quantities are used and the $E_\ell^B > 1$ GeV requirement is applied at reconstruction level on the signal region sample. In addition, we show the overall efficiency of all $B \rightarrow X_u \ell \nu$ events (black dotted line), resonant events (blue dotted line) and non-resonant events (purple dotted line) whose exact values are given in Table 5.9.

Kinematic selections	$B \rightarrow \pi \ell \nu$	$B \rightarrow \rho \ell \nu$	$B \rightarrow \omega \ell \nu$	$B \rightarrow \eta/\eta' \ell \nu$	Non-res.
$E_\ell^B > 1 \text{ GeV}$	$(17.6 \pm 0.1)\%$	$(21.3 \pm 0.4)\%$	$(4.5 \pm 0.1)\%$	$(2.5 \pm 0.1)\%$	$(54.1 \pm 0.4)\%$
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$	$(21.5 \pm 0.1)\%$	$(25.3 \pm 0.1)\%$	$(5.4 \pm 0.1)\%$	$(3.0 \pm 0.1)\%$	$(44.8 \pm 0.3)\%$
$E_\ell^B > 1 \text{ GeV}$ $M_X < 1.7 \text{ GeV}$ $q^2 > 8 \text{ GeV}^2$	$(23.3 \pm 0.2)\%$	$(27.7 \pm 0.2)\%$	$(6.1 \pm 0.1)\%$	$(2.7 \pm 0.1)\%$	$(40.2 \pm 0.4)\%$

TABLE 5.10: Fraction of different types of $B \rightarrow X_u \ell \nu$ events in the three signal regions. The statistical uncertainty is given for the fractions.

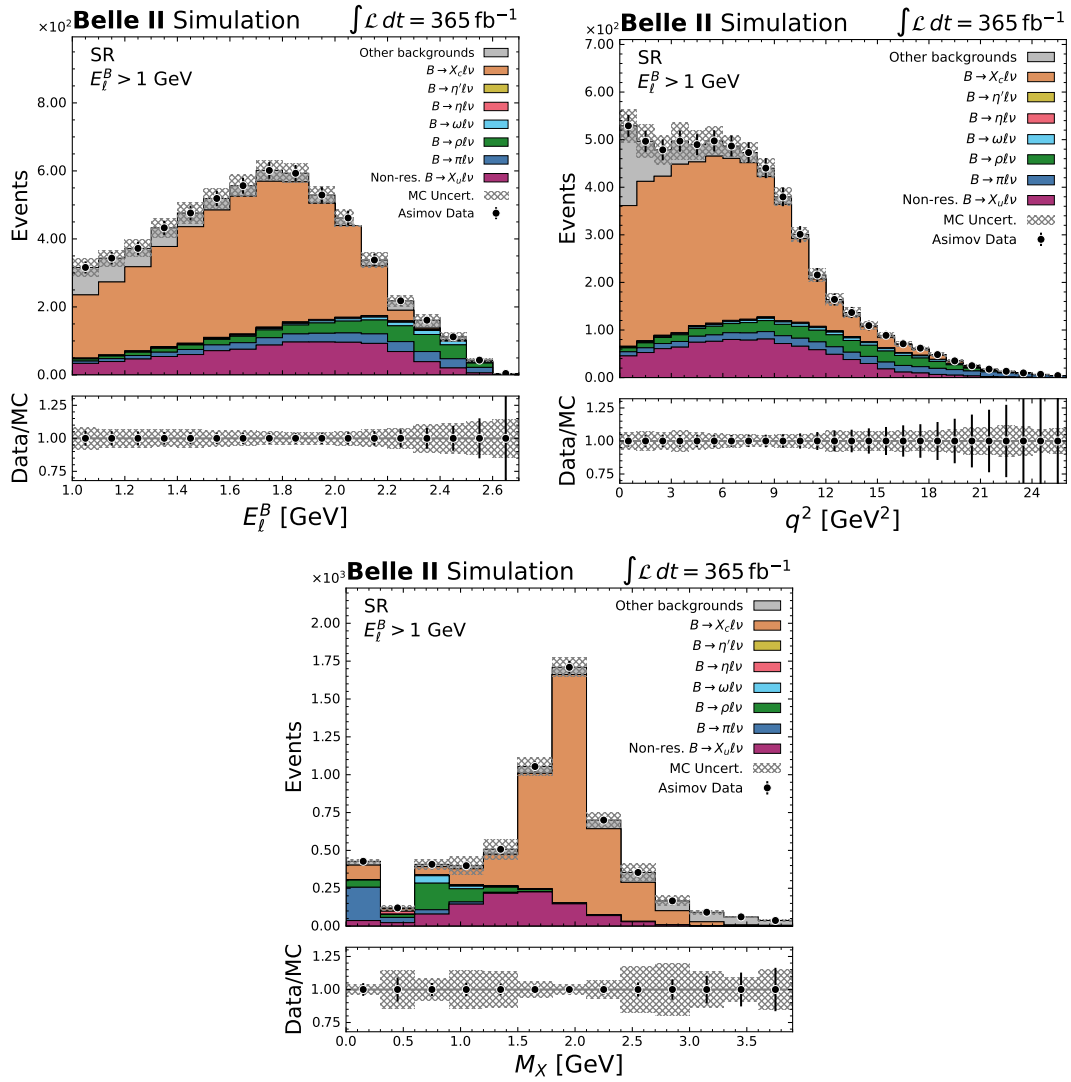


FIGURE 5.20: Simulated E_ℓ^B (top left), q^2 (top right) and M_X (bottom) distributions in the signal region with the signal component broken down in subcomponents. The $E_\ell^B > 1 \text{ GeV}$ requirement is applied.

5.8 Agreement between data and simulated data

5.8.1 Control regions

Before performing the signal extraction, we first define control regions to evaluate the agreement between experimental and simulated data distributions. By applying the preselections, the continuum suppression and by inverting the kaon veto and/or requiring a low $B \rightarrow X_c \ell \nu$ classifier suppression score, three such control regions are defined:

- $\text{CR}_{0,\text{low}}$: the kaon veto is applied but events are required to have a $B \rightarrow X_c \ell \nu$ suppression classifier score below 0.60
- $\text{CR}_{K,\text{high}}$: the presence of at least one kaon and a $B \rightarrow X_c \ell \nu$ classifier score higher than 0.87 are required
- $\text{CR}_{K,\text{low}}$: the presence of at least one kaon and a $B \rightarrow X_c \ell \nu$ classifier score lower than 0.60 are required.

In addition, two validation regions, $\text{VR}_{K,\text{mid}}$ and $\text{VR}_{0,\text{mid}}$, with classifier scores between 0.60 and 0.87, are defined. The same kinematic selections are applied to all regions and the $E_\ell^B > 1$ GeV is always applied. When additional selections on M_X and q^2 are applied, they are mentioned explicitly. The resulting subdivided phase-space is illustrated in Figure 5.21.

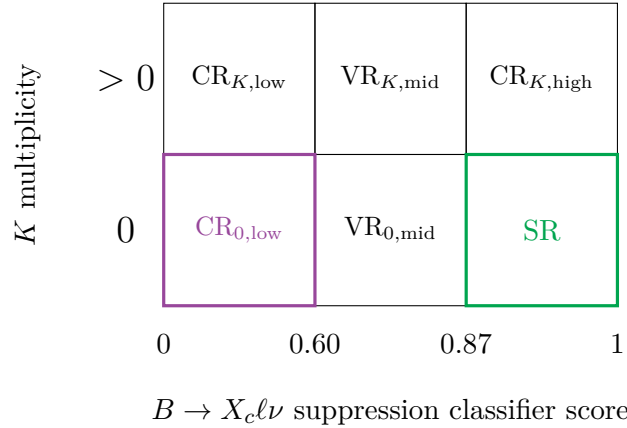


FIGURE 5.21: The data samples after reconstruction, preselection and continuum suppression are subdivided in six regions as illustrated here. The phase-space is subdivided based on the kaon multiplicity (vertical axis) and the $B \rightarrow X_c \ell \nu$ suppression classifier score (horizontal axis). The signal and control regions used for the signal extraction are highlighted in green and purple, respectively.

The E_ℓ^B , q^2 and M_X distributions in $\text{CR}_{0,\text{low}}$ are shown in Figure 5.22. Similar distributions in $\text{CR}_{K,\text{low}}$ and $\text{CR}_{K,\text{high}}$ are shown in Appendix D.

From the kinematic distributions in control regions, a disagreement between experimental and simulated data becomes apparent. The background originating from fake and secondary leptons is known to be poorly modelled but it represents a relatively small fraction of events in the control regions and it cannot explain the observed mismodelling, especially in high E_ℓ^B/q^2 and low M_X regions where it is almost absent. Furthermore, the continuum is also known to be poorly modelled but this component was calibrated using off-resonance data as described in Section 5.4.4. Small shape differences have been observed between off-resonance and on-resonance distributions so it is possible that the calibration factors extracted from off-resonance samples

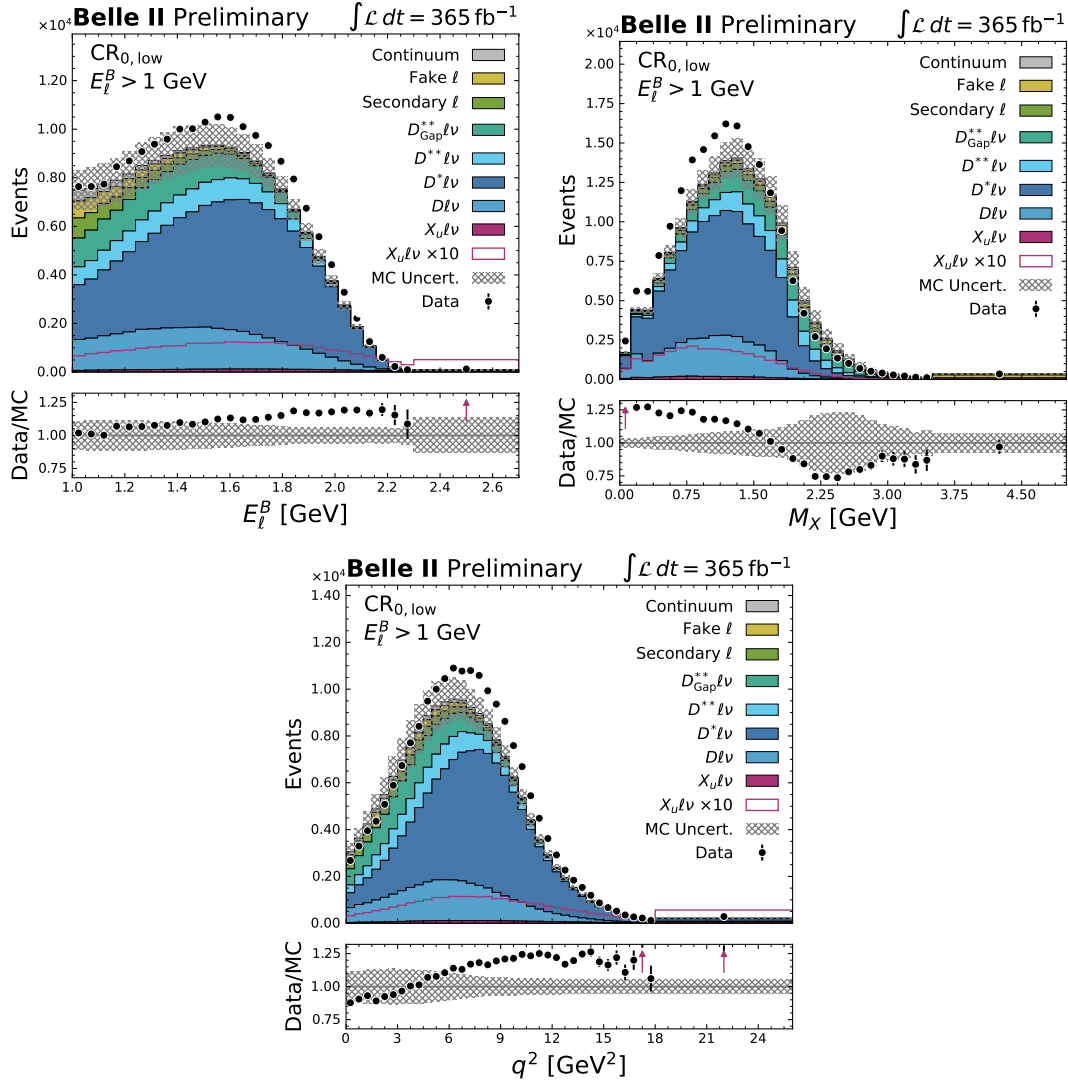


FIGURE 5.22: E_ℓ^B (top left), M_X (top right) and q^2 (bottom) distributions in experimental and simulated data in the $\text{CR}_{0,\text{low}}$ region defined in the text. The $E_\ell^B > 1 \text{ GeV}$ kinematic selection is applied.

do not translate perfectly to on-resonance data. However, large systematic uncertainties related to the continuum correction are added and, similarly to fake/secondary lepton backgrounds, the relatively low contribution of continuum events can hardly explain the mismodelling observed in control regions. The data-MC disagreement is therefore expected to stem from $B \rightarrow X_c \ell \nu$ decays. Multibody semileptonic decays such as $B \rightarrow D_s^{(*)} K \ell \nu$ or $B \rightarrow D^{(*)} \pi \pi \ell \nu$ are challenging to model and have not been studied extensively. In the samples used for our measurement, the former are simulated via a simple phase-space model and the latter via an intermediate D^{**} resonance relying on the BLR form factor parametrisation. However, these contributions are assigned large uncertainties and they also represent a small fraction of $B \rightarrow X_c \ell \nu$ events. Similarly, since $B \rightarrow D^{(*)} \eta \ell \nu$ decays used as gap modes have never been observed, there exist no good theoretical description of these types of processes but a relative uncertainty of 100% is assigned to their branching fractions to factor in this imperfect modelling.

A possible explanation for the observed mismodelling comes from D meson decays to three

or more particles since most of them are simulated via phase-space modelling instead of decay-specific models. Since D mesons are part of the decay chains of all charmed semileptonic B decays, such poor modelling can bias the kinematics of $B \rightarrow X_c \ell \nu$ decays. This hypothesis was first suggested and studied in Ref. [161]. In this study, an attempt was made to correct simulated samples using generator-level information by adjusting the relative rates of various $D \rightarrow KX$ decays (where X represents any system of particles). Such corrections of the simulated physical processes seem to reduce the observed discrepancies, therefore making a strong argument in favour of the D decay hypothesis. Besides, the discrepancies observed in inclusive analyses such as the one presented in Ref. [161] and our analysis tend to be significantly larger than those observed in exclusive $B \rightarrow X_c \ell \nu$ measurements (see for instance, Ref. [162] for a recent example). This difference would likely be caused by the fact that inclusive measurements are more sensitive to the poor modelling of D decays than measurements in which the D meson is reconstructed in a limited number of specific decays. To test further the D decay modelling hypothesis, we compare simulated and experimental data in a sideband composed of events in which the signal lepton candidate charge doesn't match the signal B candidate charge and on which all preselections have been applied. In this sideband, called *wrong-sign lepton* sideband, about 80% of events are continuum and fake and secondary lepton events, while $B \rightarrow X_c \ell \nu$ events fill the remaining 20%. The comparison is shown in Figure 5.23 for E_ℓ^B , q^2 and M_X . A similar disagreement can be observed as in region $\text{CR}_{0,\text{low}}$ (see Figure 5.22) but it appears to be less pronounced. While observing similar discrepancies in non-semileptonic backgrounds as in $B \rightarrow X_c \ell \nu$ decays could be evidence of additional sources of discrepancy (poor detector response modelling for instance), it is important to note that non-semileptonic backgrounds are also sensitive to D decays. Charm mesons can be formed in $e^+e^- \rightarrow c\bar{c}$ production and B mesons decay hadronically to D mesons at relatively high rates. However, in the context of our semileptonic event reconstruction, such background events are not as well reconstructed as $B \rightarrow X_c \ell \nu$ events. Besides, the average particle multiplicity of non-semileptonic backgrounds (~ 19.5) is expected to be larger than that of semileptonic events (~ 17.7). The poorer reconstruction and higher average multiplicity of non-semileptonic events could diffuse effects caused by the D decay mismodelling, therefore explaining the differences between $\text{CR}_{0,\text{low}}$ and the wrong-sign lepton sideband.

In the measurement presented here, the branching fractions of all D decays are corrected to match the latest PDG average and the errors related to the branching fractions are added as systematic uncertainties to partially cover discrepancies originating from the imperfect D decay modelling. Moreover, the impact of the modelling of $B \rightarrow X_c \ell \nu$ events with a D meson decaying to a K_L^0 was studied in detail by scaling their contribution by arbitrary factors similarly to what was done in Ref. [163]. Upscaling this component by 30% only marginally improves the agreement between data and simulated data, indicating that it is not possible to explain the observed discrepancies only from the poor modelling of $D \rightarrow K_L^0 X$ decays. We discuss further the modelling of $D \rightarrow K_L^0 X$ decays in Appendix E.

To conclude, there is strong evidence that the D decay modelling causes the observed discrepancies, but some of the elements mentioned in this discussion are subject to interpretation. Therefore, D decay modelling cannot be unequivocally identified as the source of discrepancies between experimental and simulated data. Besides, correcting consistently the modelling is extremely complex as it would necessitate the implementation of dedicated models for the simulation of multibody D decays. This effort is being undertaken within the Belle II collaboration. In our measurement, the signal region $B \rightarrow X_c \ell \nu$ component is corrected as a whole using the control regions defined above. The normalisation and shape mismodelling are corrected separately.

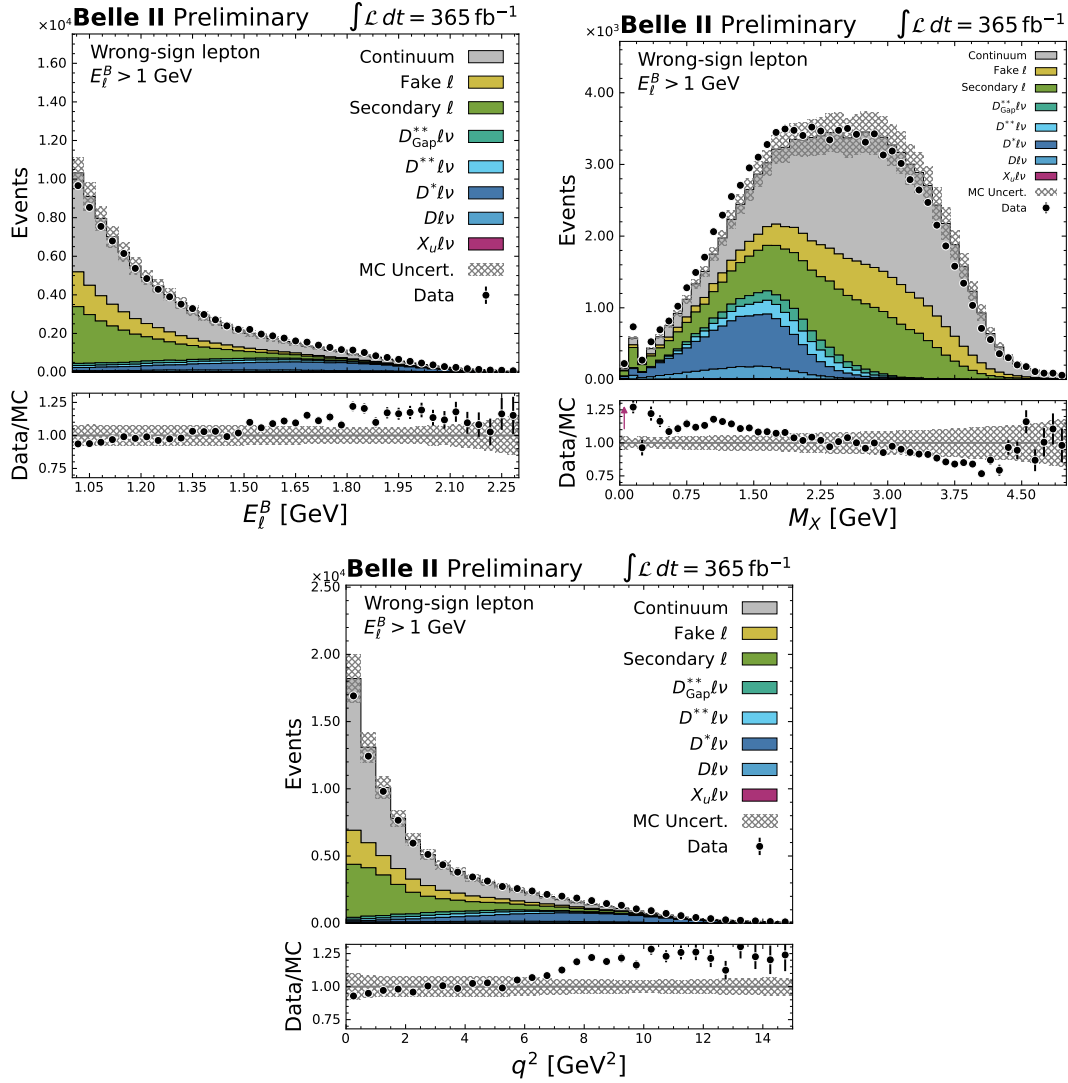


FIGURE 5.23: E_ℓ^B (top left), M_X (top right) and q^2 (bottom) distributions in experimental and simulated data in the wrong-sign lepton sideband. The $E_\ell^B > 1$ GeV kinematic selection is applied.

5.8.2 $B \rightarrow X_c \ell \nu$ normalisation correction

Before measuring the $B \rightarrow X_u \ell \nu$ branching fraction, the normalisation of the $B \rightarrow X_c \ell \nu$ component in the signal region is corrected by a factor extracted via an ABCD method relying on the observed normalisations in the three control regions. The ABCD method has been used in a wide variety of particle physics analyses and was already applied in measurements by the CDF collaboration in the early 1990s [164]. Since the kaon multiplicity and classifier score are weakly correlated – the Pearson correlation coefficient for these two variables is -0.04 as computed on the preselected sample on which the continuum suppression has been applied – it is assumed that the normalisation correction factor of the $B \rightarrow X_c \ell \nu$ component in the signal region, R_{SR} , can be determined from the normalisation correction factors in the control regions such that

$$R_{\text{SR}} = \frac{R_{0,\text{low}} \cdot R_{K,\text{high}}}{R_{K,\text{low}}}. \quad (5.10)$$

The $B \rightarrow X_c \ell \nu$ normalisation correction factors $R_{0,\text{low}}$, $R_{K,\text{low}}$ and $R_{K,\text{high}}$ are calculated by subtracting from data and simulated data all non- $B \rightarrow X_c \ell \nu$ events. This component represents between 2%–10% of all events in $\text{CR}_{0,\text{low}}$ and $\text{CR}_{K,\text{low}}$ and between 8%–25% of events in $\text{CR}_{K,\text{high}}$ with the three sets of kinematic selections. Since this calculation relies on various sources of background and signal to be normalised to unity, an uncertainty on each normalisation correction factor is derived by varying the non- $B \rightarrow X_c \ell \nu$ yields by 50% and computing the four normalisation correction factors for each variation. When applying the kinematic selection $E_\ell^B > 1$ GeV for example, the values obtained for $R_{0,\text{low}}$, $R_{K,\text{high}}$ and $R_{K,\text{low}}$ are respectively 1.11 ± 0.01 , 0.95 ± 0.02 and 1.04 ± 0.01 where the uncertainty is statistical, therefore yielding $R_{\text{SR}} = 1.01 \pm 0.02$. The values obtained by upscaling and downscaling the non- $B \rightarrow X_c \ell \nu$ component by 50% are $R_{0,\text{low}} = 1.07$, $R_{K,\text{high}} = 0.90$, $R_{K,\text{low}} = 1.00$, $R_{\text{SR}} = 0.96$ and $R_{0,\text{low}} = 1.16$, $R_{K,\text{high}} = 0.99$, $R_{K,\text{low}} = 1.07$, $R_{\text{SR}} = 1.07$, respectively, therefore adding uncertainties ± 0.05 , ± 0.05 , ± 0.04 and ± 0.06 to the four factors. Moreover, the calculation defined in Equation 5.10 is tested by replacing $\text{CR}_{K,\text{high}}$ and the signal region by $\text{VR}_{K,\text{mid}}$ and $\text{VR}_{0,\text{mid}}$, respectively. Even though the signal purity is larger in $\text{VR}_{0,\text{mid}}$ (about 10 – 20% for the three kinematic selections) than in all other control regions, it is still much smaller than in the signal region (30 – 70%). The $B \rightarrow X_c \ell \nu$ normalisation correction factor obtained for $\text{VR}_{0,\text{mid}}$ via the ABCD method and the factor obtained by directly taking the $B \rightarrow X_c \ell \nu$ data over simulated data yield ratio in that region are compared and the difference between these two factors is added as an uncertainty on R_{SR} . With the $E_\ell^B > 1$ GeV kinematic selection, they respectively read $R_{0,\text{mid}}^{\text{ABCD}} = 1.07$ and $R_{0,\text{mid}}^{\text{Data/MC}} = 1.03$, indicating that the ABCD method overestimates the true $B \rightarrow X_c \ell \nu$ normalisation correction factor by about 0.04. For each of the four factors, all uncertainties are added in quadrature therefore giving the values summarised in Table 5.11 below⁸.

Kinematic selection	$\text{CR}_{K,\text{low}}$	$\text{CR}_{K,\text{high}}$	$\text{CR}_{0,\text{low}}$	SR
$E_\ell^B > 1$ GeV	1.04 ± 0.04	0.95 ± 0.05	1.11 ± 0.05	1.01 ± 0.08
$E_\ell^B > 1$ GeV, $M_X < 1.7$ GeV	1.14 ± 0.03	1.03 ± 0.10	1.18 ± 0.04	1.07 ± 0.12
$E_\ell^B > 1$ GeV, $M_X < 1.7$ GeV, $q^2 > 8$ GeV ²	1.19 ± 0.02	1.04 ± 0.19	1.24 ± 0.03	1.09 ± 0.21

TABLE 5.11: Values of $R_{K,\text{low}}$, $R_{K,\text{high}}$, $R_{0,\text{low}}$ and R_{SR} in the three phase-space regions considered for the signal extraction. The errors shown correspond to the total uncertainties computed as described in the text. The larger uncertainties for $R_{K,\text{high}}$ when applying the M_X and q^2 selections originate from the relatively low yields and high fraction of non- $B \rightarrow X_c \ell \nu$ events in $\text{CR}_{K,\text{high}}$.

In the following sections, the $B \rightarrow X_c \ell \nu$ normalisation correction factors are applied to all prefit simulated distributions, unless explicitly stated.

5.8.3 $B \rightarrow X_c \ell \nu$ shape correction

After having corrected the $B \rightarrow X_c \ell \nu$ normalisation mismodelling, the shape correction is implemented in the signal extraction procedure by simultaneously fitting the signal region and $\text{CR}_{0,\text{low}}$ as described in detail in Section 5.9 below. The normalisation of the $B \rightarrow X_c \ell \nu$ component in the control region is also corrected before the fit by the factor given in Table 5.11. The uncertainty associated to the normalisation correction factors is added as a nuisance parameter in the fit to give the freedom to the $B \rightarrow X_c \ell \nu$ template normalisation to float independently in the control and signal regions. Implementing the shape correction directly in the fits not only allows us to automatically correct the $B \rightarrow X_c \ell \nu$ shape in the signal region but it also allows

⁸If a particularly curious reader wants to repeat the calculations with the numerical factors given as example in this paragraph, they might not be able to retrieve the exact values because of small truncation effects.

us to factor in all correlations between correction factors and other fit parameters and yield a $B \rightarrow X_u \ell \nu$ branching fraction value which incorporates the uncertainty related to the $B \rightarrow X_c \ell \nu$ background correction.

Using any of the three control regions to correct the shape of the $B \rightarrow X_c \ell \nu$ component relies on the assumption that the disagreement between data and simulated data is similar in the signal region and the chosen control region. This assumption is supported by the fact that the $B \rightarrow X_c \ell \nu$ data over simulated data yield ratios follow a similar trend in various kinematic variables in all three control regions, as illustrated in Figure 5.24. To correct the signal region mismodelling, a region as similar as possible to the signal region must be used, which favours $\text{CR}_{K,\text{high}}$ over $\text{CR}_{0,\text{low}}$ and $\text{CR}_{K,\text{low}}$. In particular, both the shape of the $B \rightarrow X_c \ell \nu$ component in the signal region and its composition resemble those in $\text{CR}_{K,\text{high}}$. A comparison of the $B \rightarrow X_c \ell \nu$ component shape in the three control regions and the signal region is shown in Figure 5.25 and the $B \rightarrow X_c \ell \nu$ composition in these four regions is shown in Table 5.12. However, as shown in Table 5.13, the $B \rightarrow X_c \ell \nu$ yields in $\text{CR}_{K,\text{high}}$ are of the same order as those in the signal region and by using this control region in the signal extraction, the $B \rightarrow X_c \ell \nu$ shape correction would be strongly influenced by the signal region distribution. As there doesn't seem to be any strong argument in favour of using either $\text{CR}_{K,\text{low}}$ or $\text{CR}_{0,\text{low}}$ to correct the mismodelling, we choose the latter which has larger $B \rightarrow X_c \ell \nu$ yields and can be defined in our signal extraction setup together with the signal region by simply requiring a low $B \rightarrow X_c \ell \nu$ suppression score.

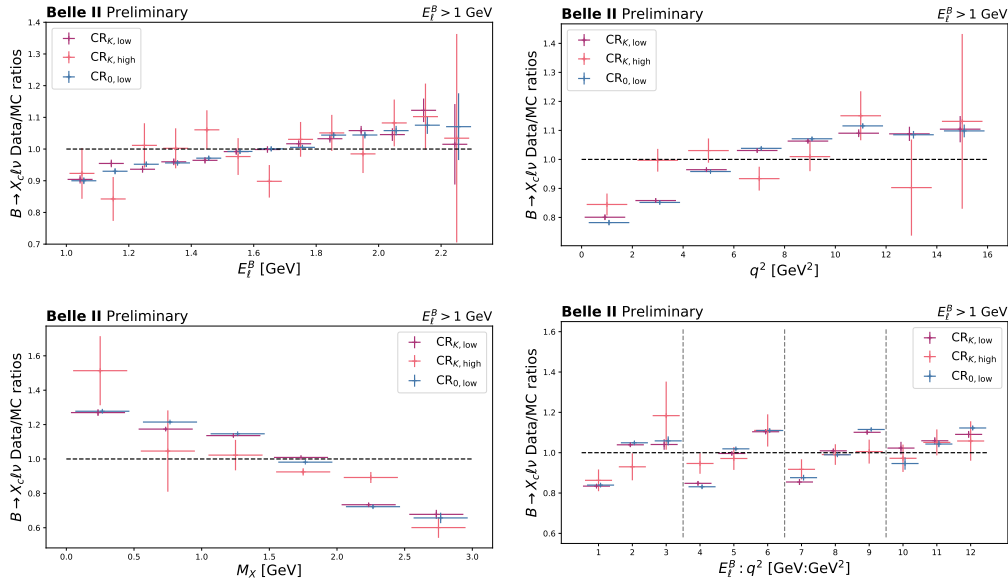


FIGURE 5.24: The normalised $B \rightarrow X_c \ell \nu$ data over simulated data yield ratios are shown for the three control regions $\text{CR}_{K,\text{low}}$ (purple), $\text{CR}_{K,\text{high}}$ (pink) and $\text{CR}_{0,\text{low}}$ (blue) in E_ℓ^B (top left), q^2 (top right), M_X (bottom left) and 2D $E_\ell^B : q^2$ (bottom right). Only the $E_\ell^B > 1$ GeV kinematic selection is applied. The labelling of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins. Similar figures are shown for the kinematic selections of fits 2 and 3 in Appendix F.

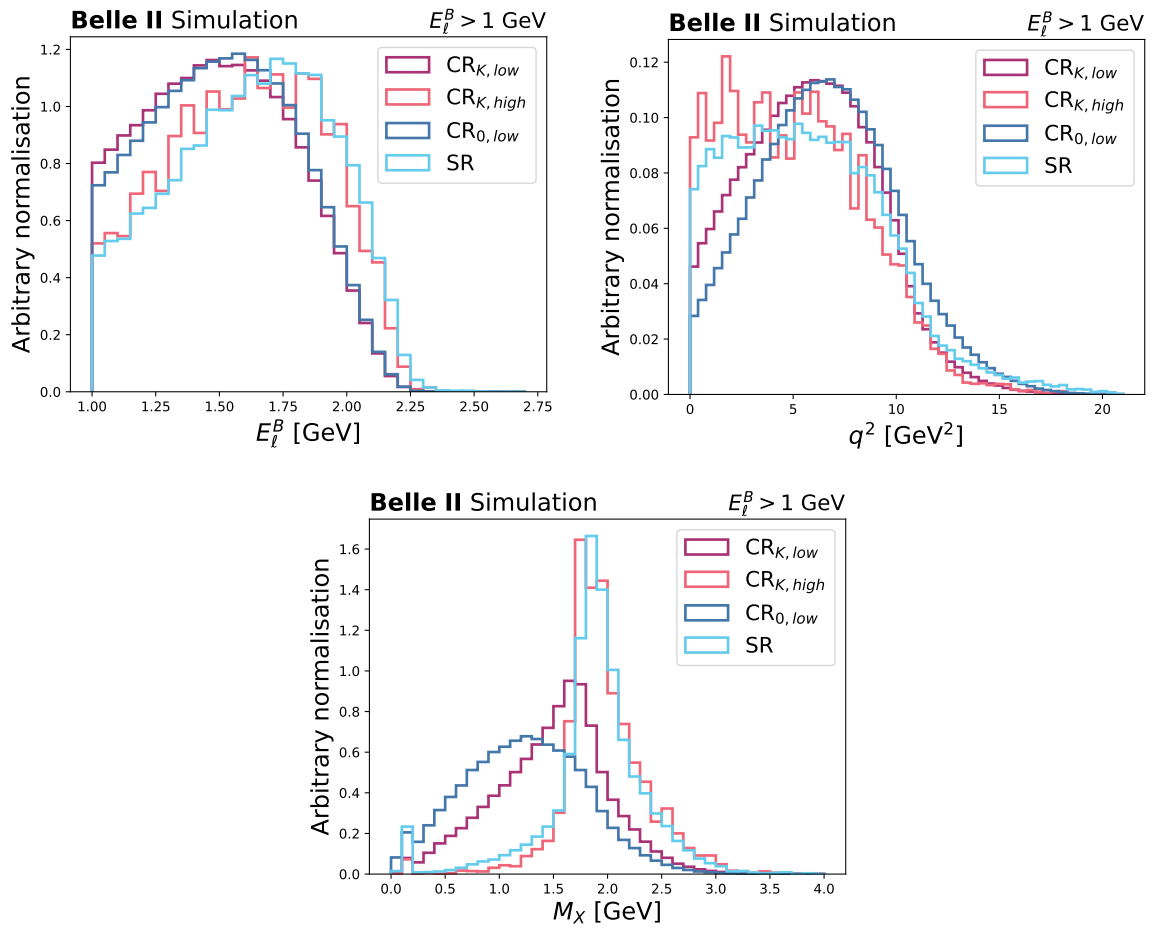


FIGURE 5.25: Comparison of the shape of the $B \rightarrow X_c \ell \nu$ component in the signal region (cyan), $CR_{K,low}$ (purple), $CR_{K,high}$ (pink) and $CR_{0,low}$ (blue) in the E_ℓ^B (top left), q^2 (top right) and M_X (bottom) distributions.

	CR _{K,low}	CR _{K,high}	CR _{0,low}	SR
$E_\ell^B > 1 \text{ GeV}$				
$B \rightarrow D\ell\nu$	17.3%	40.4%	19.1%	37.3%
$B \rightarrow D^*\ell\nu$	61.2%	47.3%	58.7%	49.2%
$B \rightarrow D^{**}\ell\nu$	9.7%	6.2%	10.0%	6.9%
$B \rightarrow D_{\text{Gap}}^{**}\ell\nu$	11.8%	6.1%	12.2%	6.6%
$E_\ell^B > 1 \text{ GeV}, M_X < 1.7 \text{ GeV}$				
$B \rightarrow D\ell\nu$	22.0%	64.3%	21.5%	47.7%
$B \rightarrow D^*\ell\nu$	65.3%	33.6%	61.9%	47.3%
$B \rightarrow D^{**}\ell\nu$	6.4%	1.5%	8.0%	2.7%
$B \rightarrow D_{\text{Gap}}^{**}\ell\nu$	6.3%	0.6%	8.6%	2.3%
$E_\ell^B > 1 \text{ GeV}, M_X < 1.7 \text{ GeV}, q^2 > 8 \text{ GeV}^2$				
$B \rightarrow D\ell\nu$	14.6%	35.9%	16.5%	32.1%
$B \rightarrow D^*\ell\nu$	79.9%	61.4%	75.2%	63.2%
$B \rightarrow D^{**}\ell\nu$	3.5%	2.0%	5.1%	3.2%
$B \rightarrow D_{\text{Gap}}^{**}\ell\nu$	1.9%	0.8%	3.3%	1.5%

TABLE 5.12: $B \rightarrow X_c\ell\nu$ composition in terms of the four types of subcomponents in each control and signal region for the three sets of kinematic selections.

Kinematic selection	CR _{K,low}	CR _{K,high}	CR _{0,low}	SR
$E_\ell^B > 1 \text{ GeV}$	101,139 $\pm$ 143	3,678 $\pm$ 24	173,680 $\pm$ 187	3,739 $\pm$ 24
$E_\ell^B > 1 \text{ GeV}, M_X < 1.7 \text{ GeV}$	65,285 $\pm$ 105	659 $\pm$ 9	137,851 $\pm$ 157	820 $\pm$ 10
$E_\ell^B > 1 \text{ GeV}, M_X < 1.7 \text{ GeV}, q^2 > 8 \text{ GeV}^2$	21,139 $\pm$ 55	178 $\pm$ 5	54,514 $\pm$ 88	399 $\pm$ 7

TABLE 5.13: Expected $B \rightarrow X_c\ell\nu$ events yields in the three control regions and the signal region when applying the three sets of kinematic selections. The statistical uncertainty of the simulated sample is given.

5.8.4 $B \rightarrow X_u \ell \nu$ modelling validation

Given the large mismodelling observed in $B \rightarrow X_c \ell \nu$ decays, it is crucial to confirm that the normalisation of the simulated $B \rightarrow X_u \ell \nu$ component is not significantly biased. Commonly used techniques, such as validating the analysis selections on a decay channel similar to the channel under study (called a control mode), are not viable in our case. It is not possible to find a decay mode similar enough to inclusive $B \rightarrow X_u \ell \nu$ decays, which populate a specific corner of the B decay phase-space. We considered, for instance, the $B \rightarrow X J/\psi (\rightarrow \ell \ell)$ decay channel as a control mode where one of the two leptons would be treated as a neutrino. These decays, however, occur via penguin diagrams and therefore result in very different kinematics compared to semileptonic decays, which already arise at tree-level. In fact, the decay channel most similar to our signal channel is the $B \rightarrow X_c \ell \nu$ decay mode. The control regions enriched in semileptonic decays could therefore be used to validate the efficiency of the main selections defining our signal region. For that purpose, we use the nominal preselected experimental and simulated data samples on which we apply the following selections one by one: the continuum suppression, the kaon veto and the $B \rightarrow X_c \ell \nu$ suppression. At each step, the efficiency of the selections in experimental and simulated data is compared, taking into account the statistical uncertainty in data and the total uncertainty in simulated samples. We confirm that the difference between the two efficiencies is always within two standard deviations, indicating a relatively good agreement. It is important to emphasize that these tests are rather qualitative because of the differences between the modelling of inclusive $B \rightarrow X_c \ell \nu$ and $B \rightarrow X_u \ell \nu$ decays. Besides, the (prefit) uncertainty of the signal component is expected to be larger than that of the $B \rightarrow X_c \ell \nu$ background because of the modelling uncertainties considered for the signal mode (see a detailed discussion in Section 5.9.4). As illustrated later in Figure 5.38, the modelling of $B \rightarrow X_u \ell \nu$ events appears to be relatively good compared to that of $B \rightarrow X_c \ell \nu$ events.

To validate the FEI selection efficiency, we use both simulated and experimental control samples reconstructed similarly to the nominal analysis samples, with the only difference being a looser hadronic tagging requirement of $\mathcal{P}_{\text{FEI}} > 0.001$ instead of $\mathcal{P}_{\text{FEI}} > 0.01$. We apply the preselection and continuum suppression to these samples to isolate a region enriched in semileptonic decays. All calibration factors, including the FEI factors, are then applied to both the nominal and low- $\mathcal{P}_{\text{FEI}}$ simulated samples. We compute the total event yields in data and simulation for both selections, and from these yields, extract the efficiency of the $\mathcal{P}_{\text{FEI}} > 0.01$ selection in data and MC. The difference in efficiency between data and simulation is less than 0.1 standard deviations, indicating excellent agreement and validating the FEI selection efficiency. Besides, a possible limitation of the hadronic tagging calibration factors comes from the fact that they were extracted by reconstructing specific types of decays on the signal side. To mitigate a potential dependence on the signal side topology, the calibration factors were extracted by reconstructing $B \rightarrow X \ell \nu$ or $B \rightarrow D^{(*)} \pi$ decays on the signal side and the two sets of factors were then combined in a χ^2 fit. Because of the much larger branching fraction, the calibration factors extracted from the semileptonic study dominate the average. It was however demonstrated in these studies that the calibration factors vary only slightly with the signal-side track multiplicity and we therefore assume that, within uncertainties, the factors are valid for any signal-side decay, including $B \rightarrow X_u \ell \nu$ decays.

Despite the difficulty in validating the modelling of signal decays, the tests discussed in this section indicate that it is reasonably good and that potential efficiency discrepancies between experimental and simulated data are covered by the uncertainties that we assign.

5.9 Signal extraction

The $B \rightarrow X_u \ell \nu$ partial branching fraction is extracted from a binned template fit with the `Pyhf` package [105]. Within the `Pyhf` statistical model, modifiers are implemented to include uncertainties related to specific sources by varying the normalisation and shape of the different components. The `Pyhf` statistical model and the different types of modifiers are described in detail in Section 4.1.1. The various modifiers included in our `Pyhf` setup are discussed in the following sections.

5.9.1 Fitted regions

In order to constrain the shape of the $B \rightarrow X_c \ell \nu$ background in the signal region, the $\text{CR}_{0,\text{low}}$ distribution is fitted together with the signal region distribution. As discussed in previous sections, the partial branching fraction of $B \rightarrow X_u \ell \nu$ is measured in three different phase-space regions defined by applying kinematic selections on E_ℓ^B , M_X and q^2 both on the signal and control region. The fraction of accessible $B \rightarrow X_u \ell \nu$ phase-space that is covered by each set of selections is given in Table 5.7. Each set of selections, therefore, characterises a separate fit setup whose specificities are highlighted in the following. The setups will be referred to as fits 1, 2 and 3 from broadest phase-space region to tightest. The fit 1 phase-space region is used to extract our nominal value of $|V_{ub}|$ as inclusive $B \rightarrow X_u \ell \nu$ models are most reliable in broader regions. The additional selections on M_X and q^2 increase the signal purity and allow us to further test the different theoretical frameworks by comparing the values of $|V_{ub}|$ obtained from their respective decay rate predictions. In the following, the signal extraction procedure is illustrated for the fit 1 kinematic selection but all figures relating to fit 2 and 3 are shown in Appendix F.

5.9.2 Fit templates

Event distributions used for the signal extraction procedure are split into the following templates:

- Signal-in ($B \rightarrow X_u \ell \nu$ in): signal events that fall in the chosen phase-space region; the template is assigned an unconstrained normalisation factor, referred to as *signal strength*, in the fit statistical model which is the parameter of interest of our signal extraction fits (`normfactor`, see Section 4.1.1); its postfit value corresponds to the multiplicative scaling of the simulated signal template and can therefore be directly translated in a measurement of the branching fraction; this template is fixed in $\text{CR}_{0,\text{low}}$ where its contribution is negligible.
- Signal-out ($B \rightarrow X_u \ell \nu$ out): a small fraction of signal events pass the reconstruction-level selections given in Table 5.7 but do not actually fall in the chosen phase-space region (based on the generator-level selections); they represent only a small fraction of signal events; the resolution of the E_ℓ^B distribution is extremely good which allows the $B \rightarrow X_u \ell \nu$ out template to be neglected in fit 1; for fits 2 and 3, this component is kept as it represents 2% and 5% of all signal events in their respective signal regions; the template is assigned the same unconstrained normalisation as the signal-in template.
- $B \rightarrow X_c \ell \nu$ background: this template represents the largest source of background; to capture the shape of this template in $\text{CR}_{0,\text{low}}$, it is assigned one free-floating normalisation factor per bin (`shapefactor`); in each bin of the signal and control region distributions, the same factor is assigned; the normalisation of this template is corrected before fitting following the procedure described in Section 5.8.2; the $B \rightarrow X_c \ell \nu$ normalisation is allowed to float independently in each region and is constrained by a Gaussian function with width equal to the 1σ window given in Table 5.11 (`normsys`). It is important to note that, since

the shape factors represent normalisation factors of the $B \rightarrow X_c \ell \nu$ template in each bin, they can also shift the overall normalisation of this template.

- *Other backgrounds:* all other backgrounds are summed up in a single template; they include continuum events as well as fake and secondary lepton events; after applying the continuum suppression classifier, the yields of these components are relatively small compared to semileptonic event yields. In fit 1, this template is assigned a normalisation modifier which is constrained by a Gaussian function with a width of 50% of the template yields to conservatively vary this component (**normsys**); in fits 2 and 3, the contribution from other backgrounds is small (3% – 5% of all events in the signal and control regions) and this modifier is therefore not added.

5.9.3 Fitted variables

Different variables were tested in the signal extraction setup described in this section. To use more event information, two and three-dimensional variables can be fitted. However, it was noticed that fitting the M_X distribution yields a much larger total uncertainty compared to fits of leptonic variables. Therefore, the 2D $E_\ell^B : q^2$ variable is fitted for the first and second setups, splitting E_ℓ^B into four bins and q^2 into three bins. When applying both the tight M_X and q^2 selections, only the E_ℓ^B variable covers a range large enough to extract the signal branching fraction. In the third phase-space region, the lepton energy is therefore fitted. The binning for each variable is chosen to yield a statistically significant number of events in each bin, both in the control and the signal region. Moreover, high- q^2 and high- E_ℓ^B events are grouped in a single broad bin to avoid relying on the poor modelling of the $B \rightarrow X_u \ell \nu$ shape in that region. The variable and binning for each signal extraction setup are given in Table 5.14. Different binnings similar to those eventually used were tested for each fitted variable but no significant decrease of the total uncertainty was observed. In figures shown in the following, the $E_\ell^B : q^2$ variable is flattened in 12 bins, which are labelled from 1 to 12 as shown in Figure 5.26.

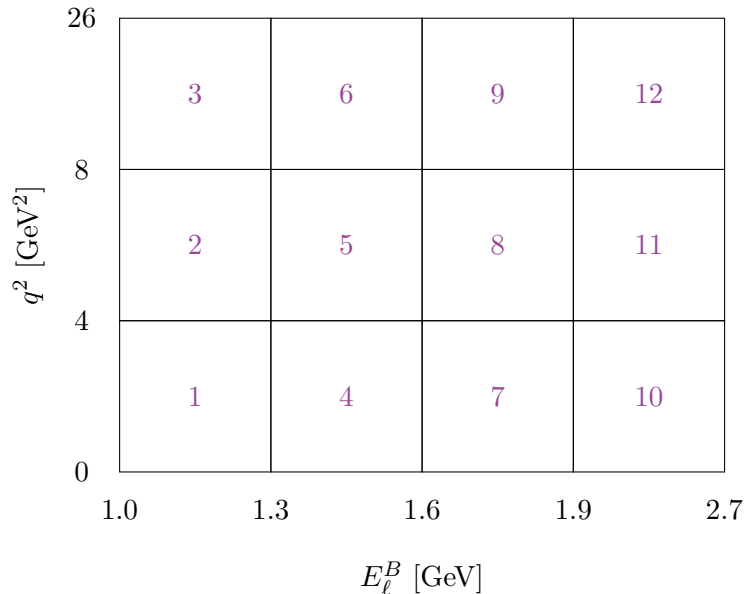
	Variable	Binning
Fit 1	$E_\ell^B : q^2$	$[1.0, 1.3, 1.6, 1.9, 2.7] \text{ GeV} \times [0, 4, 8, 26] \text{ GeV}^2$
Fit 2	$E_\ell^B : q^2$	$[1.0, 1.3, 1.6, 1.9, 2.7] \text{ GeV} \times [0, 4, 8, 26] \text{ GeV}^2$
Fit 3	E_ℓ^B	$[1.0, 1.4, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.7] \text{ GeV}$

TABLE 5.14: Variable and binning chosen for each one of the three signal extraction setups.

5.9.4 Systematic uncertainties

Sources of systematic uncertainty are directly implemented in the fit as nuisance parameters. Nuisance parameters can modify the shape and the normalisation of the template(s) to which they are assigned via **histosys** (additive) and **normsys** (multiplicative) **Pyhf** modifiers respectively. Most of the systematic uncertainties included in the signal extraction setup can modify a template via an up and a down variation – called *two-sided* variations – which is the case of branching fraction uncertainties, for instance. If a group of parameters are known to be correlated (such as form factor parameters, for example), their covariance matrix is computed and diagonalised to obtain its eigenvalues λ_i and eigenvectors $\vec{e}_i$. The parameters are then varied in the eigendirections⁹ by $\pm\sqrt{\lambda_i}\vec{e}_i$ yielding the up/down template variations entering the fit.

⁹The direction of an eigenvector is called the eigendirection and a variation in this direction is called an eigenvariation.

FIGURE 5.26: Labelling of the flattened $E_\ell^B : q^2$ bins.

Some of the uncertainties that we consider are implemented as *one-sided* variations (switching the non-resonant $B \rightarrow X_u \ell \nu$ model, for instance), which we symmetrise about the nominal distribution to make them two-sided variations.

Several sources of uncertainty are propagated by generating a large set of toy variations (for the continuum shape corrections, for example). From the set of variations, a global covariance matrix is constructed in a flattened binning which assigns a unique index to each template bin in the fit (with two channels, three templates and 12 variable bins, for example, the resulting covariance matrix would be of dimension 72). Similarly to correlated parameters, the covariance matrix is diagonalised and its eigenvalues and eigenvectors are used to compute a set of variations for each template. If the source of uncertainty is not related to a particular template, no variations are yielded for that template.

In several cases of two-sided and toy variations, the dimension of the covariance matrix is relatively large. To avoid implementing a considerable number of nuisance parameters with a negligible impact, the number of parameters is truncated. To do so, only the largest n eigenvariations are kept such that their eigenvalues sum up to 99% of the sum of all eigenvalues. This method reduces the total number of fit parameters from about 1,000 to about 150 while covering more than 99% of the total variance.

Most systematic uncertainties are symmetric about the nominal template in the upward and downward varied templates. However, to avoid introducing biases in the fit, the variations that are known to be asymmetric, such as form factor variations or inclusive $B \rightarrow X_u \ell \nu$ model parameter variations, are symmetrised. The asymmetry A of an uncertainty source is computed as

$$A = \max_i \frac{(H_i - H_i^-) - (H_i^+ - H_i)}{H_i}, \quad (5.11)$$

where i runs over the flattened global bin index, H_i is the nominal yield in the bin, and $H_i^\pm$ are the yields of the varied templates in the bin. If the asymmetry is found to be small, $A < 0.005$, the distance between the nominal and varied templates is averaged, whereas if it is found to be large, the maximum envelope of the varied templates is taken.

An overview of systematic uncertainties considered in the signal extraction is given in the following. Uncertainties can either stem from modelling or experimental effects. The relative uncertainty contributions obtained for all three fits using Asimov data are summarised in Table 5.16. Some of the uncertainties are directly related to sample corrections described in Section 5.4.

B and D meson branching fractions

All semileptonic B and all D decay branching fractions are allowed to vary within their respective uncertainties (given in Tables 5.1 and 5.2 for semileptonic decays), therefore yielding one nuisance parameter per decay type. The branching fraction uncertainty of the unmeasured $B \rightarrow D^{(*)} \eta \ell \nu$ modes is set to 100%. The branching fraction uncertainty of $B \rightarrow D_0^* \ell \nu$ decays is increased to partially cover the large difference between the two measurements entering the HFLAV average [63]. The number of parameters related to D decay branching fractions is truncated from 227 to one following the method described above.

The fit parameter related to the inclusive $B \rightarrow X_u \ell \nu$ branching fraction is only allowed to modify the template shape but not its normalisation since this effect is related to the parameter of interest that is extracted from the fit. For each branching fraction variation of a charmless decay, the hybrid weights are recalculated, therefore keeping the generated inclusive $B \rightarrow X_u \ell \nu$ branching fraction fixed to its nominal value but changing the relative rates of resonant to non-resonant decays. At the reconstruction level, efficiency effects still lead to normalisation variations of the signal template.

Semileptonic B decay form factors

The modelling of semileptonic resonant modes relies on assumed form factor models which depend on a set of parameters extracted from dedicated measurements and lattice QCD. The form factor parameter eigenvariations are computed with the **eFFORT** and **Hammer** packages for $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ decays, respectively (see Section 4.4 for a detailed description of these packages). The parameter values, uncertainties and correlation matrices are taken from the references given for each model in Section 5.2.2. An example of a differential decay rate variation for $B \rightarrow \pi \ell \nu$ decays is shown in Figure 4.4. Similarly to the branching fractions, the hybrid weights are recalculated for each form factor variation of charmless decays.

Inclusive $B \rightarrow X_u \ell \nu$ model uncertainties

The two DFN parameters computed in the Kagan-Neubert renormalisation scheme, $m_b^{\text{KN}} = 4.66 \pm 0.04$ GeV and $a^{\text{KN}} = 1.3 \pm 0.5$, are correlated, therefore yielding two up/down eigenvariations of the non-resonant $B \rightarrow X_u \ell \nu$ component which are included as nuisance parameters. For each of the four sets of $(m_b^{\text{KN}}, a^{\text{KN}})$, a sample is generated to compute event weights in fine three-dimensional bins of M_X , q^2 and E_ℓ^B such that each weight in a bin ijk is $w_{ijk} = \Delta \mathcal{B}^{\text{varied}} / \Delta \mathcal{B}^{\text{nominal}}$. For each variation, the hybrid weights are recalculated. In addition, to factor in differences between inclusive $B \rightarrow X_u \ell \nu$ models, a set of events is simulated via the Bosch, Lange, Neubert, Paz (BLNP) framework (see Section 2.3.5) whose predictions are compared to those of the DFN model in Figure 5.27. The difference between the nominal fitted distribution and the BLNP distribution is symmetrised and added as an uncertainty. For each variation, the hybrid weights are recalculated.

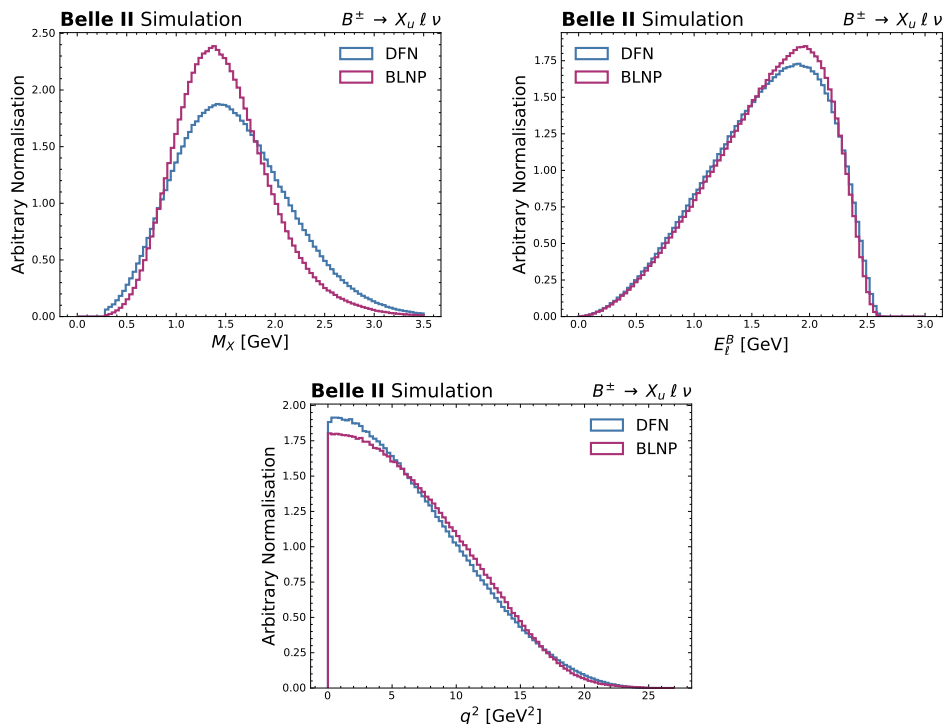


FIGURE 5.27: Comparison between the inclusive $B^\pm \rightarrow X_u \ell \nu$ M_X (top left), E_ℓ^B (top right) and q^2 (bottom) spectra as predicted by the DFN (blue) and BLNP (purple) theoretical frameworks.

Strange quark fragmentation

The production of $s\bar{s}$ quark pairs in simulated samples is controlled by the **Pythia** 8 parameter γ_s whose default value is set to 0.217 as determined from the Monash tune [165] dominated by LEP data. From Figure 25 in Ref. [165], it can be seen that at low energies, the kaon multiplicity as predicted by **Pythia** is underestimated compared to data. However, the same comparison with $\gamma_s = 0.35$ shows that the kaon multiplicity is overestimated [166], indicating that the true value of γ_s is most likely in the range 0.25 – 0.30 for centre-of-mass energies around 10 – 40 GeV. Since the hadronic system in the semileptonic B decays considered in our measurement have energies below 5 GeV and mostly around 1 – 3 GeV, these considerations should be interpreted cautiously¹⁰. Based on the available information, we decide to use a relatively broad range of values by averaging the measurements of the TASSO [167] ($\gamma_s = 0.35 \pm 0.05$) and JADE [168] ($\gamma_s = 0.27 \pm 0.06$) collaborations to obtain $\gamma_s = 0.30 \pm 0.09$ where the uncertainty covers the full spread of 1σ variations from the two values entering the average. We produce three sets of events where γ_s is set to 0.21, 0.30 and 0.39. The sample with γ_s set to its central value is used as nominal and the variations with respect to the other two distributions are added as systematic uncertainties and for each variation, the hybrid weights are recalculated. The kaon multiplicity in $B \rightarrow X_u \ell \nu$ events in the three samples after preselection is shown in Figure 5.28.

¹⁰The validity of the **Pythia** model at energy scales below 10 GeV is questionable even though this model remains the best available to simulate hadronisation.

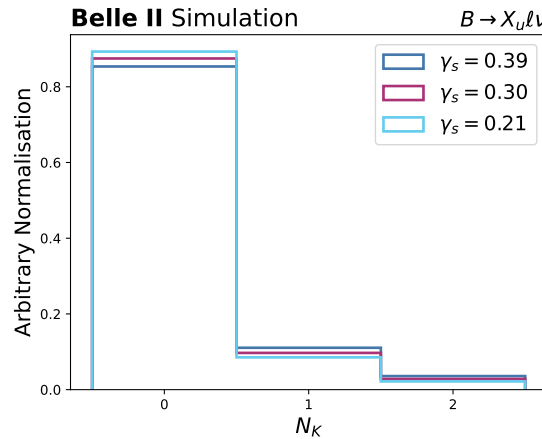


FIGURE 5.28: Comparison between the kaon multiplicity in $B \rightarrow X_u \ell \nu$ events for different values of the *Pythia* parameter γ_s .

f^{+-}/f^{00}

The ratio of production fractions of charged and neutral pairs of B mesons in $\Upsilon(4S)$ decays is taken from Ref. [63] and reads $f^{+-}/f^{00} = 1.052 \pm 0.031$ ¹¹. The relative fractions of B^+ and B^0 decays are allowed to vary in the fit within this uncertainty.

X_u hadronisation modelling

To derive an uncertainty on the fragmentation modelling of the X_u system in non-resonant signal events, we attempted to modify a subsets of all *Pythia* parameters and identify those which impact the most this aspect of the modelling. However, unlike for the $s\bar{s}$ fragmentation described above, it is not possible to identify a single parameter or a small number of parameters which control the hadronisation of the X_u system. An uncertainty is therefore derived using the signal region $\pi^\pm$ multiplicity. The fit results are projected onto the $\pi^\pm$ multiplicity distribution in each of the three signal regions and it is assumed that the remaining discrepancy between experimental and simulated data is entirely due to the imperfect X_u hadronisation modelling (fit projections are discussed in Section 5.9.5). For each signal region, three scale factors are extracted in bins of the generator-level $\pi^\pm$ multiplicity of non-resonant $B \rightarrow X_u \ell \nu$ events ($0\pi^\pm$, $1\pi^\pm$ or $\geq 2\pi^\pm$) such that the simulated postfit $\pi^\pm$ multiplicity distribution matches perfectly the data distribution. The scale factors are applied to the prefit signal region distributions and the three signal extraction fits are repeated. The difference between the measured $B \rightarrow X_u \ell \nu$ branching fraction and the branching fraction extracted from the rescaled distribution is added as an uncertainty. Part of the uncertainty could originate from mismodelling of backgrounds in the $\pi^\pm$ multiplicity distribution. However, because of the relatively large signal purity in the three signal regions, the conservative approach described here is adopted.

Slow pion reconstruction efficiency

Efficiency correction factors for low momentum pions originating from D^* decays are extracted from independent studies by the Belle II performance group [140, 141]. The correction factors are extracted separately for charged and neutral pions in three laboratory frame momentum bins between 50 and 200 MeV. The correction factors being correlated, a covariance matrix is

¹¹A small fraction of $\Upsilon(4S)$ resonances do not decay to B mesons. Since this fraction is estimated to be about 0.3% of all decays, an order of magnitude lower than the relative uncertainty on f^{+-}/f^{00} , we neglect the effect and we do not add any related uncertainty.

computed for π^+ and π^0 from which eigenvariations are calculated. This results in a total of six additional nuisance parameters.

Tracking efficiency

The efficiency of each reconstructed track with a momentum higher than 200 MeV is varied by 0.24% in simulated events and the full difference between the nominal and the corrected template is added as an uncertainty [169].

Charged particle identification

The correction factors for identification efficiency and misidentification rate of $e^\pm$, $\mu^\pm$ and $K^\pm$ are extracted from dedicated studies by the Belle II performance group [134–136, 142, 149] in bins of the track charge, laboratory frame momentum and polar angle. Each correction factor is given with associated statistical and systematic uncertainties. A total of 200 toy variations about the central value of each correction factor are drawn for lepton and hadron identification separately. From the set of variations, a covariance matrix is calculated to extract a truncated number of variations (from 400 to six in each fit).

K_S^0 reconstruction efficiency

An efficiency correction factor of 0.55% per cm of flight distance from the interaction point is assigned to each K_S^0 candidate to downscale simulated events [150]. The full difference between the nominal and the corrected template is added as a symmetric uncertainty.

B tagging efficiency

Hadronic tagging efficiency correction factors are extracted from dedicated studies by the Belle II performance group [148] in 12 (11) separate B^+ (B^0) channels covering all decay modes used by the FEI algorithm. From these correction factors and their uncertainties, two covariance matrices are derived to obtain a truncated number of eigenvariations (from 23 to two in each fit).

Continuum calibration

The continuum sample is corrected using a data-driven approach described in Section 5.4.4. The full difference between the nominal distribution and the distribution corrected by the normalisation factors is added as an uncertainty. Besides, uncertainties on the continuum component shape arise because of the limited number of experimental and simulated data events used to extract the correction factors. A total of 200 bootstrapped variations of the off-resonance experimental and simulated data samples are trained separately, yielding as many sets of correction factors. From these sets, a covariance matrix is calculated to extract a truncated number of eigenvariations (from 200 to five in each fit).

Finite size of simulated data samples

The simulated data statistical uncertainty is computed from the number of simulated events and propagated in the signal extraction setup. This systematic effect is included in the fit setup via a dedicated `staterror` modifier described in Section 4.1.1.

Fit template normalisations

In fit 1, a normalisation parameter is assigned to the *other backgrounds* template, which contributes to the total signal branching fraction uncertainty. To vary the component normalisation within a conservative range, the nuisance parameter is constrained by a Gaussian function with a width of 50% of the total component yields. In fits 2 and 3, the yields of non-semileptonic events being negligible, this parameters is not added but other nuisance parameters related to this template can vary both its normalisation and shape. Furthermore, the control region and signal region $B \rightarrow X_c \ell \nu$ template normalisations are corrected before the fit and they are assigned a nuisance parameter constrained by a Gaussian function with a width equal to the correction factor uncertainties given in Table 5.11.

Number of $\Upsilon(4S)$ decays

To extract a $B \rightarrow X_u \ell \nu$ branching fraction from the fit result, the output signal normalisation is multiplied by the assumed input branching fraction. Since charged and neutral $B \rightarrow X_u \ell \nu$ decays are simulated separately but the measurement is performed simultaneously on all decays, the combined inclusive $B \rightarrow X_u \ell \nu$ input branching fraction is computed from the number of simulated signal events N_{X_u} and the number of collected $\Upsilon(4S)$ events such that $\mathcal{B}(B \rightarrow X_u \ell \nu) = N_{X_u} / (2N_{\Upsilon(4S)})$. The number of $\Upsilon(4S)$ events collected by Belle II has been measured to be $N_{\Upsilon(4S)} = (387 \pm 6) \times 10^6$ and its uncertainty is therefore added to the total uncertainty of the measured branching fraction.

5.9.5 Signal extraction validation

Before using signal region data to measure the $B \rightarrow X_u \ell \nu$ partial branching fractions, various tests are performed to validate the stability of the fit and confirm that the setup behaves as expected. First, we perform each fit using Asimov data¹², *i.e.* using the simulated template yields as dataset, and toy datasets. We then perform each fit using control and validation regions instead of $\text{CR}_{0,\text{low}}$ and the signal region.

Asimov and toy data validation

The inclusive $B \rightarrow X_u \ell \nu$ branching fraction used to generate signal events over the full available phase-space is $\mathcal{B}(B \rightarrow X_u \ell \nu) = 1.82 \times 10^{-3}$, which can be translated to an expectation value for each fitted phase-space region by multiplying by the phase-space acceptance of Table 5.7. The signal normalisation extracted from the fit can be used to scale the input branching fraction, thus yielding a measured value for each partial branching fraction. We first perform the three nominal fits with Asimov data, *i.e.* using the simulated yields in place of experimental data. The prefit and postfit distributions are shown in Figures 5.29 and 5.30, respectively. When performing the three nominal fits with Asimov data, the output normalisation is equal to one as expected, resulting in the three branching fractions and associated total uncertainties given in Table 5.15. In each Asimov fit, the binned free-floating $B \rightarrow X_c \ell \nu$ factors are also equal to one.

In the correlation matrices obtained from the three fits with Asimov data, we can see that the $B \rightarrow X_c \ell \nu$ shape factors are strongly correlated with each other, reflecting the fact that they collectively float to vary the shape (and normalisation) of the $B \rightarrow X_c \ell \nu$ component. Furthermore, the shape factors are anti-correlated with the background normalisations — both the signal and control region $B \rightarrow X_c \ell \nu$ normalisations and the *other backgrounds* normalisation —

¹²The name Asimov comes from Isaac Asimov's short story *Franchise* where one person is selected to represent the entire electorate during the United States presidential election.

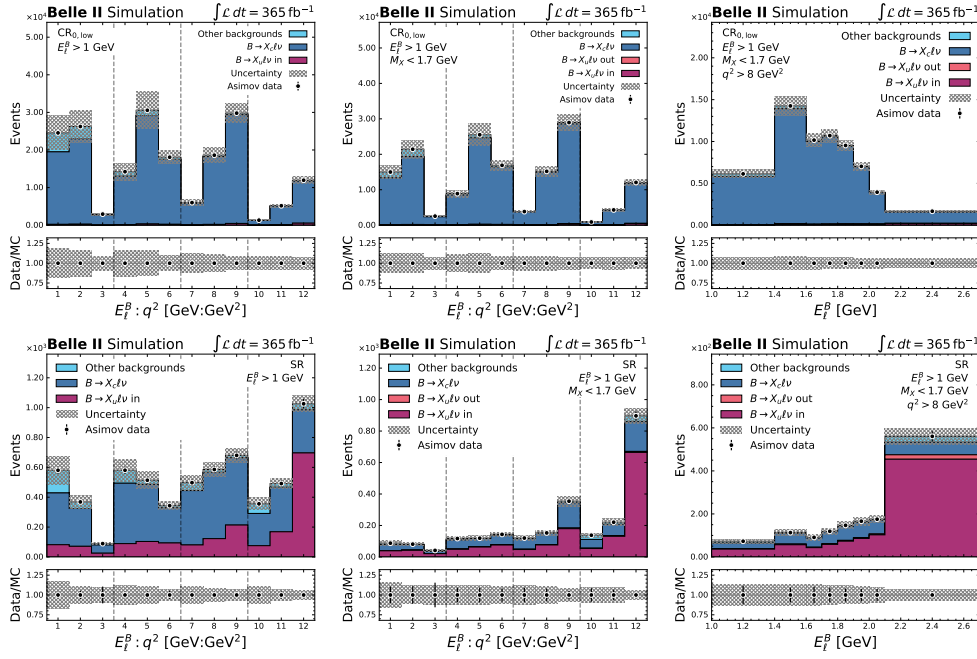


FIGURE 5.29: Prefit control region (top) and signal region (bottom) distributions for fit 1 (left), 2 (middle) and 3 (right) with Asimov data. The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

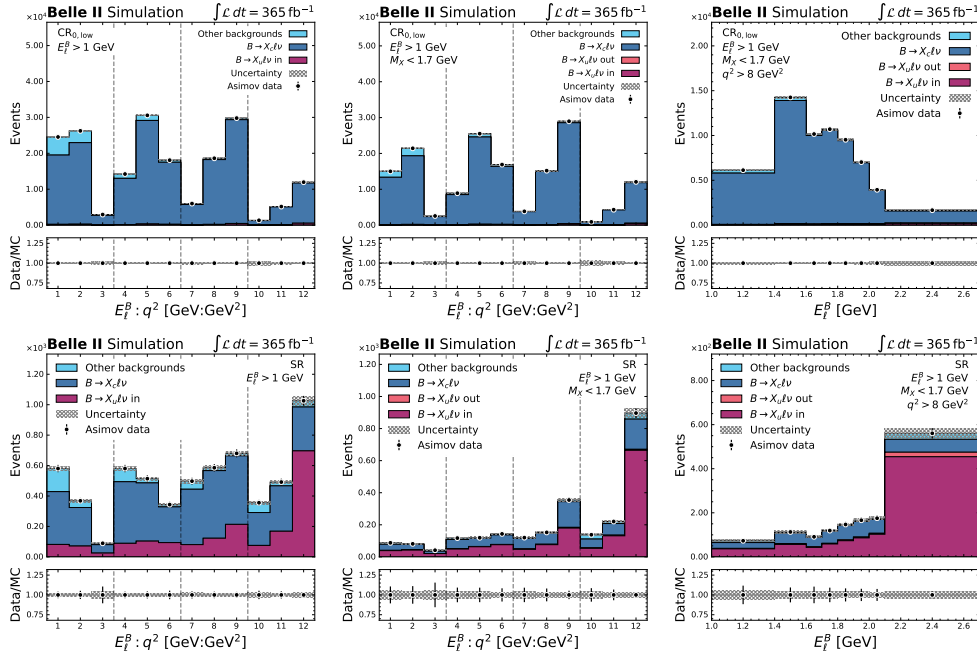


FIGURE 5.30: Postfit control region (top) and signal region (bottom) distributions for fit 1 (left), 2 (middle) and 3 (right) with Asimov data. The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

since variations in the shape factors can also alter the overall normalisation of the $B \rightarrow X_c \ell \nu$ template. This anti-correlation is particularly strong with the $B \rightarrow X_c \ell \nu$ normalisation in $\text{CR}_{0,\text{low}}$ from which the shape factors are effectively extracted because of the much larger number of $B \rightarrow X_c \ell \nu$ events compared to the signal region. The shape factors also display correlations

	$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) \times 10^3$
Fit 1	1.58 ± 0.13
Fit 2	1.03 ± 0.08
Fit 3	0.57 ± 0.05

TABLE 5.15: Expected partial branching fractions extracted from the three nominal fits.

with other $B \rightarrow X_c \ell \nu$ parameters such as the D decay branching fractions, the efficiency calibration factors of low-momentum pions and the $B \rightarrow D^{**} \ell \nu$ form factors and branching fractions which is expected as all these parameters can collectively modify the $B \rightarrow X_c \ell \nu$ component. The signal strength itself is anti-correlated with the $B \rightarrow X_c \ell \nu$ normalisation in the signal region, since both parameters impact the total signal yield in that region. This effect is more pronounced in the regions employed for fits 2 and 3, although the background contribution there is significantly reduced. Moreover, the signal strength is correlated with the DFN model variation and γ_s parameters, which appear to have the largest impact on the measured signal normalisation (see Figure 5.32). In fit 3, the signal strength is also correlated to the variation between the DFN and BLNP models, as the signal purity is particularly high in the fit 3 kinematic region.

It is instructive to evaluate the pulls of fit parameters in Asimov and toy data tests. The pull of a parameter is defined as:

$$\begin{cases} \frac{\mu_{\text{in}} - \mu}{\sigma_+}, & \text{if } \mu \leq \mu_{\text{in}}, \\ \frac{\mu - \mu_{\text{in}}}{\sigma_-}, & \text{if } \mu > \mu_{\text{in}}, \end{cases} \quad (5.12)$$

where μ_{in} is the initial value of the parameter, μ its postfit value and σ_+/σ_- are the amplitudes of the associated positive/negative uncertainties. The pull is expected to be equal to zero for every constrained parameter in an Asimov fit, which is confirmed for all three fits. Furthermore the fit setup is tested on a large number of randomly generated toy datasets. Each pseudo-experiment is generated by varying the Asimov yield in each bin about its expectation value following a Poisson distribution and the auxiliary measurement constraining a nuisance parameter following a Gaussian distribution. A total of 5,000 toy datasets are produced this way and each one of them is fitted. The resulting set of output signal normalisation pulls is expected to be distributed as a normal probability density function, which is illustrated in Figure 5.31 for fit 1. The same test is repeated for the $B \rightarrow X_c \ell \nu$ shape factors. Furthermore, to further assess the stability of the fit, large numbers of toy datasets with different input signal normalisations are produced. We test 11 values of the input signal normalisation, equally distributed in the range 0.5 to 1.5 and for each value, 1,000 pseudo-experiments are generated. All toy datasets are fitted and for each set, the mean value of the signal strength is extracted. The resulting linear response, shown in Figure 5.31 for fit 1, confirms that the signal extraction setups can accurately determine the $B \rightarrow X_u \ell \nu$ normalisation.

Finally, we estimate the impact of each nuisance parameter on the extracted signal normalisation. To do so, the fit is repeated by keeping one nuisance parameter fixed to its postfit central value plus/minus its associated uncertainty and letting all other fit parameters float. The impact of a nuisance parameter on the signal normalisation is then measured as the difference between the nominal normalisation postfit value and the value obtained from the fit where the nuisance parameter has been varied. The expected impact of the 20 most important nuisance parameters is shown in Figure 5.32 for fit 1. As expected, the inclusive modelling (via the DFN parameter variations) represents the largest source of uncertainty for the $B \rightarrow X_u \ell \nu$

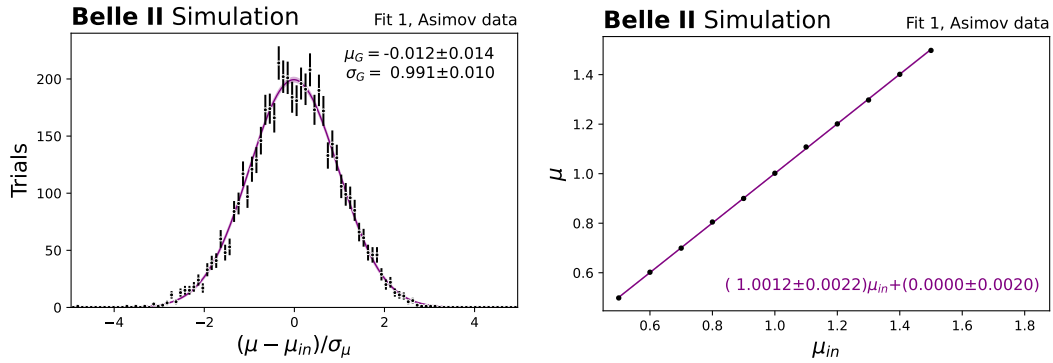


FIGURE 5.31: Left: pull distribution of the signal normalisation extracted by fit 1 from 5,000 toy datasets (black points) and the fitted normal distribution (purple line). μ_G and σ_G are the Gaussian curve mean and variance, respectively. Right: linearity test for fit 1, where the black points are the average value of 1,000 toy fits and the purple line is the linear curve fitted through the points. The curve fitted parametrisation is shown in purple on the plot.

branching fraction measurement. However, it is interesting to note that the $s\bar{s}$ fragmentation uncertainty represents the second most impactful nuisance parameter, whereas using the BLNP framework to model the fitted distribution appears to have only a limited impact on the signal extraction. As anticipated, the normalisation of the $B \rightarrow X_c \ell \nu$ background in the signal region significantly shifts the extracted $B \rightarrow X_u \ell \nu$ branching fraction. To better understand the impact of various sources of uncertainty on the signal extraction, we estimate their contribution to the total uncertainty. For each group of parameters relating to the same uncertainty source, we produce 1,000 pseudo-experiments with fixed data yields and fixed auxiliary data for all parameters except those belonging to that group. Each of these pseudo-experiments is fitted, resulting in 1,000 measured signal normalisations which are Gaussian-distributed. The standard deviation of this distribution represents the uncertainty related to the studied group of parameters. This technique is approximate as the correlations between fit parameters, which in the fits presented here are sizeable, are ignored in the toy dataset production. Hence, the sum in quadrature of all individual sources of uncertainty does not necessarily match the total uncertainty as obtained from the fit. However, the approximation is good enough to yield representative values for the uncertainties. The uncertainty breakdown for fits 1, 2 and 3 using Asimov data is given in Table 5.16.

$\text{CR}_{K,\text{low}} - \text{CR}_{K,\text{high}}$ fit

The $B \rightarrow X_c \ell \nu$ background shape correction strategy presented in Section 5.8.3 relies on the assumption that correction factors can be extracted from the $B \rightarrow X_c \ell \nu$ component in a low classifier score region ($\text{CR}_{0,\text{low}}$) and applied to a high classifier score region (the signal region). To confirm this assumption, each fit is repeated by using $\text{CR}_{K,\text{low}}$ as control region and $\text{CR}_{K,\text{high}}$ as signal region. The normalisation of the $B \rightarrow X_c \ell \nu$ component is corrected by the factors given in Table 5.11. The main kinematic distributions in these two control regions are shown in Appendix D. The signal yields in $\text{CR}_{K,\text{high}}$ (between 3% and 17% of all events when applying the three sets of kinematic selections) are small compared to those in the signal region (between 30% and 68%) and the $\text{CR}_{K,\text{low}} - \text{CR}_{K,\text{high}}$ fit setup is therefore not expected to yield a reliable measurement of the $B \rightarrow X_u \ell \nu$ branching fraction. However, because the $B \rightarrow X_c \ell \nu$ yields are much larger in $\text{CR}_{K,\text{low}}$ than in $\text{CR}_{K,\text{high}}$, the extracted $B \rightarrow X_c \ell \nu$ shape factors are expected to be relatively similar to the $\text{CR}_{K,\text{low}}$ $B \rightarrow X_c \ell \nu$ data/MC ratios computed by subtracting

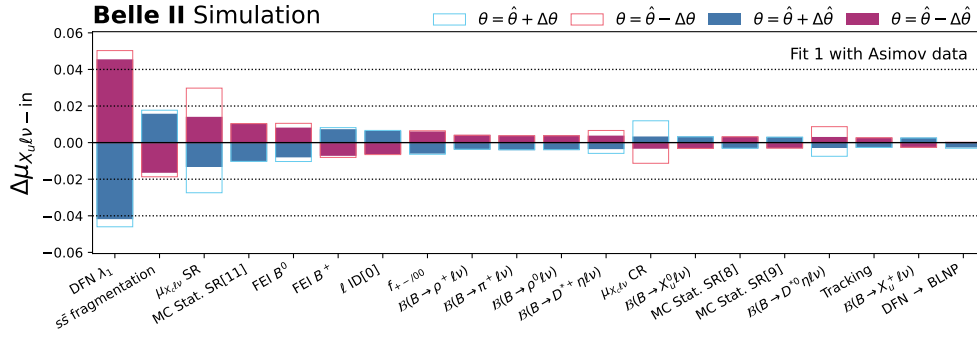


FIGURE 5.32: Impact of the 20 most important nuisance parameters as obtained from fit 1 with Asimov data. Each nuisance parameter θ is shifted from its central value $\hat{\theta}$. The prefit impact, shown by the cyan and pink lines, is computed by varying each nuisance parameter by the uncertainty given by its constraint term ($\Delta\theta$). The postfit impact, shown by the blue and purple boxes, is computed by varying each nuisance parameter by the uncertainty estimated from the fit ($\Delta\hat{\theta}$).

the non- $B \rightarrow X_c \ell \nu$ simulated data yields from experimental and simulated data. The $\text{CR}_{K,\text{low}} - \text{CR}_{K,\text{high}}$ fit repeated with the three nominal kinematic selections yields a good p -value (> 0.50) as illustrated in Figure 5.33 for the fit 1 kinematic selection. We compare in Figure 5.34 the $B \rightarrow X_c \ell \nu$ data/MC ratios in each of the two control regions to the $B \rightarrow X_c \ell \nu$ shape factors extracted from the fit. These tests confirm that the shape of the $B \rightarrow X_c \ell \nu$ component in high classifier score regions can reliably be corrected by simultaneously fitting this region and a low classifier score region.

Other fits using control regions and validation regions were also performed. Two of these setups are worth discussing. First, we repeated the $\text{CR}_{K,\text{low}} - \text{CR}_{K,\text{high}}$ fit with the three kinematic selections by replacing $\text{CR}_{K,\text{low}}$ by the control region used in the signal extraction fits, $\text{CR}_{0,\text{low}}$. Similarly to the setup described above, the extraction of the $B \rightarrow X_c \ell \nu$ shape factors mostly relies on $\text{CR}_{0,\text{low}}$ and they significantly improve the agreement between experimental and simulated data in $\text{CR}_{K,\text{high}}$. Second, we performed fits with $\text{CR}_{0,\text{low}}$ as control region and $\text{VR}_{0,\text{mid}}$ replacing the signal region to test a setup where the (fake) signal region has a larger signal purity than in $\text{CR}_{K,\text{high}}$ (the signal purity is between 10% and 20% in $\text{VR}_{0,\text{mid}}$ for the three sets of kinematic selections). For these tests, the prefit normalisation of the $B \rightarrow X_c \ell \nu$ template in $\text{VR}_{0,\text{mid}}$ is corrected by a factor estimated from the ABCD method described in Section 5.8.2 using $B \rightarrow X_c \ell \nu$ yields in $\text{CR}_{K,\text{low}}$, $\text{CR}_{0,\text{low}}$ and $\text{VR}_{K,\text{mid}}$. Similarly to the two setups described above in this section, the postfit good is excellent with p -values larger than 0.3. However, when comparing the data over simulated data ratios in $\text{CR}_{0,\text{low}}$ and $\text{VR}_{0,\text{mid}}$ to the $B \rightarrow X_c \ell \nu$ shape factors extracted from each fit, we notice that the shape factors underestimate the ratios by about $1 - 2\sigma$, in particular in low E_ℓ^B bins. It appears that other fit parameters, most notably the $B \rightarrow D^{**} \ell \nu$ and $B \rightarrow D_{\text{Gap}}^{**} \ell \nu$ form factor parameters, pull significantly to compensate for the lower shape factors. Such effects are likely due to the fact that the prefit $B \rightarrow X_c \ell \nu$ normalisation correction is slightly overestimated compared to the true correction factor (1.07 and 1.03, respectively, see Section 5.8.2) but differences between the shape of the $B \rightarrow X_c \ell \nu$ component in $\text{CR}_{0,\text{low}}$ and $\text{VR}_{0,\text{mid}}$ could also impact the fits. The consequences of such effects in the nominal signal extraction setups are evaluated as discussed in the next section and in Section 5.10.1.

Uncertainty source	Relative uncertainty (%)		
	Fit 1	Fit 2	Fit 3
DFN parameters	4.7	4.2	5.7
DFN $\rightarrow$ BLNP	0.1	0.4	1.3
γ_s	1.6	1.2	2.3
$B \rightarrow \pi \ell \nu$ form factors	0.3	0.3	0.3
$B \rightarrow \rho \ell \nu$ form factors	0.3	0.3	0.2
$B \rightarrow \omega \ell \nu$ form factors	0.1	0.1	0.1
$B \rightarrow \eta/\eta' \ell \nu$ form factors	< 0.1	< 0.1	< 0.1
$B^\pm \rightarrow X_u \ell \nu$ branching fractions	1.0	0.7	0.5
$B^0 \rightarrow X_u \ell \nu$ branching fractions	0.6	0.5	0.5
$B \rightarrow D_{\text{Broad}}^{**}$ form factors	0.4	0.2	0.2
$B \rightarrow D_{\text{Narrow}}^{**}$ form factors	0.1	< 0.1	< 0.1
$B \rightarrow D/D^* \ell \nu$ form factors	< 0.1	< 0.1	< 0.1
$B^\pm \rightarrow X_c \ell \nu$ branching fractions	0.7	0.6	0.2
$B^0 \rightarrow X_c \ell \nu$ branching fractions	0.7	0.2	0.1
D decay branching fractions	0.2	0.3	0.1
Signal region $X_c \ell \nu$ normalisation	1.5	2.3	3.5
CR _{0,low} $X_c \ell \nu$ normalisation	1.0	0.8	0.4
Other background normalisation	0.2	N/A	N/A
$N_{\Upsilon(4S)}$	1.4	1.4	1.4
FEI	1.3	1.4	1.4
Slow pion efficiency	0.4	0.2	0.3
ℓ ID	0.7	0.7	0.6
$f^{\pm/00}$	0.6	0.7	0.6
Continuum reweighting	0.1	0.1	0.2
Tracking	0.3	0.2	0.3
K_S^0 efficiency	0.1	0.1	< 0.1
$K^\pm$ ID	< 0.1	< 0.1	< 0.1
Simulated data statistics	1.2	1.1	0.8
Total systematic	6.1	6.4	8.4
Statistical	5.2	5.0	4.6
Total	8.0	8.1	9.6

TABLE 5.16: Systematic uncertainty breakdown for all fits using Asimov data. The implementation of systematic uncertainties in the signal extraction setup is described in detail in Section 5.9.4. The X_u fragmentation uncertainty is not included as it requires the fits to be performed using signal region data to be estimated.

CR_{K,low}–CR_{K,high} fit with injected signal

The setup described in the previous section can be modified to mimic the nominal CR_{0,low}–SR setup by injecting signal in experimental and simulated data in CR_{K,high}. This setup is created by adding to the data and MC in CR_{K,high}, the simulated $B \rightarrow X_u \ell \nu$ template from the signal region. The first test performed with this setup is to vary the normalisation of the signal

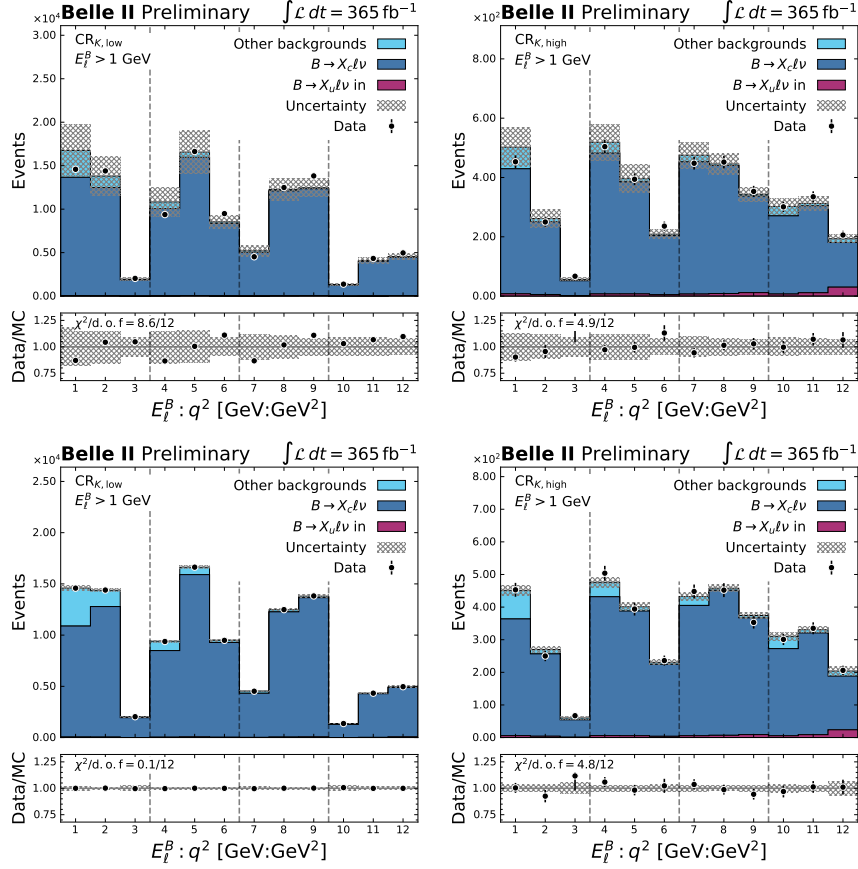


FIGURE 5.33: Prefit (top) and postfit (bottom) $E_\ell^B : q^2$ distributions in $\text{CR}_{K,\text{low}}$ (left) and $\text{CR}_{K,\text{high}}$ (right) for the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit. The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

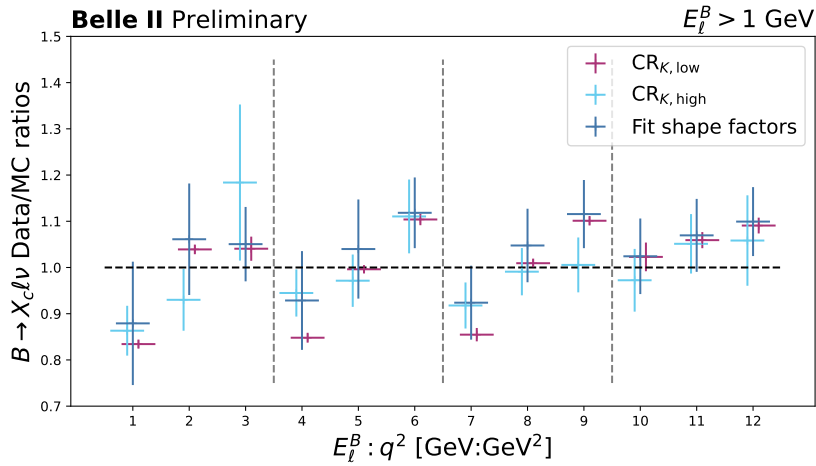


FIGURE 5.34: The normalised data over simulated data yield ratios in $\text{CR}_{K,\text{low}}$ (purple) and $\text{CR}_{K,\text{high}}$ (cyan) in bins of $E_\ell^B : q^2$ with the fit 1 kinematic selection applied are shown. The ratios are compared to the $B \rightarrow X_c \ell \nu$ shape factors extracted from the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit performed with the same kinematic selection (blue). The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins. Similar plots are shown for fits 2 and 3 in Appendix F.

component injected in data in the range $[0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15]$ while keeping the normalisation of the signal in simulated data fixed and compare the input normalisation to the measured one. The $B \rightarrow X_c \ell \nu$ normalisation in the two regions is corrected before the fit. The observed response in all three fits points to a slight bias in the branching fraction measurement as shown in Figure 5.35 for the test performed using the fit 1 kinematic selection. This bias arises because of differences in shape and composition of the $B \rightarrow X_c \ell \nu$ component in the control and signal regions used in the fit. Because of these differences, the $B \rightarrow X_c \ell \nu$ shape correction extracted from the control region doesn't adjust perfectly the $B \rightarrow X_c \ell \nu$ shape in the signal region, thus forcing the fit to pull on various fit parameters, including the free-floating $B \rightarrow X_u \ell \nu$ normalisation, to match the simulated templates to data. The bias is corrected by using the response obtained in the validation setup. For fit 1 for example, the measured signal normalisation μ_{out} is corrected such that $\mu_{\text{out}} \rightarrow (\mu_{\text{out}} - 0.052)/0.920$ which shifts its value by about 2–3%. An uncertainty related to the bias correction procedure is estimated by repeating the injected signal fit over a large number of toy datasets. Each pseudo-experiment is created by varying the $\text{CR}_{K,\text{high}}$ data yields within their statistical uncertainties while keeping the injected signal yields, the $\text{CR}_{K,\text{low}}$ yields and the auxiliary data fixed. This method allows us to test a large number of $B \rightarrow X_c \ell \nu$ data configurations in the pseudo signal region without changing the control region data or the signal data injected in $\text{CR}_{K,\text{high}}$. With this test, we investigate the limits of the assumption that the $B \rightarrow X_c \ell \nu$ data over simulated data ratios are similar in different regions (see Figure 5.24). For each of the seven injected signal normalisations that are tested, 1,000 toys are produced and fitted. The distribution of the 1,000 output normalisations is Gaussian as shown in Figure 5.36 for the fit 1 validation setup and an injected signal normalisation of 1. The standard deviations of the seven Gaussian distributions are nearly equal to each other and the largest of the seven values is added as an uncertainty on the measured $B \rightarrow X_u \ell \nu$ branching fraction. The uncertainty estimation is repeated for each set of kinematic selections and the obtained errors represent an additional 2.6%, 2.3% and 1.8% relative uncertainty for fits 1, 2 and 3, respectively.

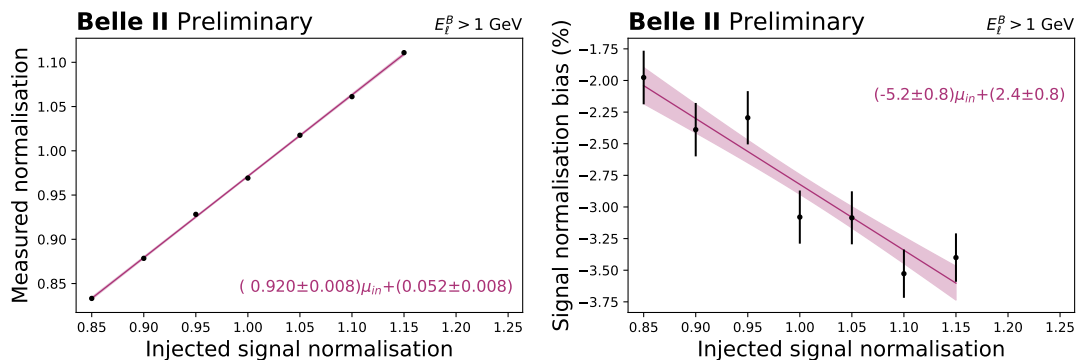


FIGURE 5.35: Comparison between the normalisation of the $B \rightarrow X_u \ell \nu$ component injected in data in the fit 1 validation setup and the measured signal normalisation (left). The observed bias (defined as the relative difference between the injected and measured normalisations) is shown explicitly on the right. The purple line corresponds to a linear curve fitted through the points and its parametrisation is shown in purple on the plots. Equivalent figures are shown for fits 2 and 3 in Appendix F.

Since the correction and uncertainty are extracted by relying on the $\text{CR}_{K,\text{high}}$ and $\text{CR}_{K,\text{low}}$ regions instead of $\text{CR}_{0,\text{low}}$ and the signal region, an additional uncertainty related to the differences in $B \rightarrow X_c \ell \nu$ composition in each region is extracted. The $B \rightarrow X_c \ell \nu$ composition in the signal and control regions for the three sets of kinematic selections is shown in Table 5.12. The

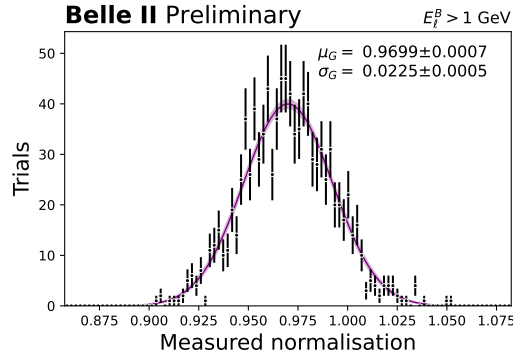


FIGURE 5.36: Distribution of the output normalisation obtained from fitting 1,000 toy datasets in the validation setup described in the text. The normalisation of the signal injected in data is fixed to 1 and the fit 1 kinematic selection are applied. The purple line is the result of the Gaussian fit and the curve parameters are shown in the top right-hand corner.

injected signal fit setup is repeated by fixing the normalisation of the signal injected in data to unity and by rescaling the $B \rightarrow X_c \ell \nu$ composition of $\text{CR}_{K,\text{high}}$ ($\text{CR}_{K,\text{low}}$) such that it matches the composition of the signal region ($\text{CR}_{0,\text{low}}$). The difference in measured normalisations between the rescaled fit and the nominal injected signal fit is taken as an additional uncertainty. For fits 1, 2 and 3, an uncertainty of 1.4%, 5.7% and 0.2%, respectively, is therefore added to the total branching fraction error. The uncertainty is much larger for fit 2 than for the other fits because, as can be seen in Table 5.12, the $B \rightarrow X_c \ell \nu$ composition is very different between $\text{CR}_{K,\text{high}}$ and the signal region after applying the $M_X < 1.7$ GeV selection. On the other hand, the composition uncertainty is almost negligible for fit 3 because the E_ℓ^B variable is not particularly sensitive to the hadronic composition.

In addition, to validate our expectations about the injected signal fit, its results can be projected onto different variables. Using **Cabinetry**, such projections can conveniently be performed by matching fit parameters of the initial statistical model onto a new model built from another distribution. This way, the shape and normalisation of each template in the new distribution are modified according to the template variations extracted from the initial fit. To properly transfer the $B \rightarrow X_c \ell \nu$ binned shape factors, the $B \rightarrow X_c \ell \nu$ template is split into sub-components, each one matching a bin in the distribution used in the initial fit and each newly defined $B \rightarrow X_c \ell \nu$ template is assigned a free-floating normalisation corresponding to one of the shape factors. A caveat from fit projections performed with this method is that the statistical uncertainty of simulated data cannot be transferred, as the associated **Pyhf** modifier is defined separately for every bin of a fitted distribution. In the fits discussed here, this source of uncertainty has a relatively small impact and the fit projections shown in the following remain valid as cross-checks of the fit outcome. When projecting the results of the nominal injected signal $E_\ell^B:q^2$ fit (normalisation of the $B \rightarrow X_u \ell \nu$ component injected in data set to 1) with the fit 1 kinematic selections, the agreement between the simulated and experimental E_ℓ^B and q^2 distributions improves as shown in Figure 5.37. However, the agreement in variables which rely on information from the hadronic system, M_X and M_{miss}^2 , improves only marginally which is also illustrated in Figure 5.37. It is nevertheless important to confirm that the agreement doesn't worsen in these distributions as this would otherwise invalidate the $B \rightarrow X_c \ell \nu$ correction implemented in the fit.

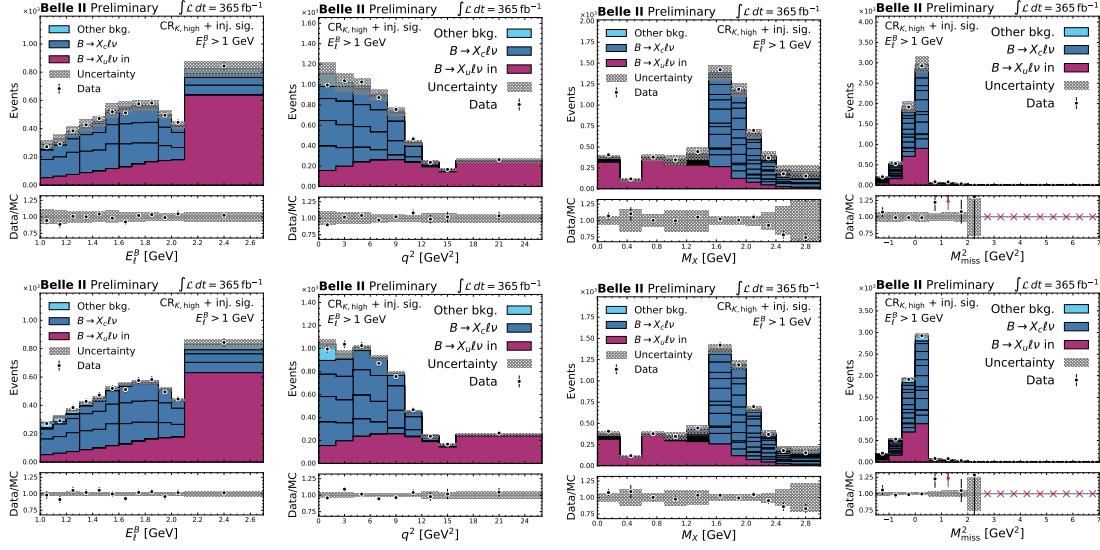


FIGURE 5.37: Top: distributions of E_ℓ^B (left), q^2 (left centre), M_X (right centre) and M_{miss}^2 (right) in $\text{CR}_{K,\text{high}}$ with the $B \rightarrow X_u \ell \nu$ template from the signal region injected in simulated and experimental data. Bottom: projections of the nominal injected signal $E_\ell^B:q^2$ fit (normalisation of the $B \rightarrow X_u \ell \nu$ component injected in data set to 1) with the fit 1 kinematic selections onto the E_ℓ^B , q^2 , M_X and M_{miss}^2 distributions. The $B \rightarrow X_c \ell \nu$ template is split into 12 templates, one for each bin of $E_\ell^B:q^2$.

5.10 Pre-unblinding tests

In the context of our blinded analysis, we have so far avoided using signal region data. Before eventually performing the signal extraction, we further validate the fit setup and signal modelling with pre-unblinding tests presented in this section.

5.10.1 Signal region distributions

The signal region distributions of the most important kinematic variables, E_ℓ^B , q^2 , M_X and the two-dimensional $E_\ell^B:q^2$ variable used for the signal extraction are shown in Figure 5.38 for the fit 1 kinematic selection and Appendix F for fits 2 and 3.

While the agreement between data and simulated data appears to be relatively good in signal-enriched regions, it is not the case in regions of the phase-space where $B \rightarrow X_c \ell \nu$ decays dominate. The shape mismodelling of this component is expected to be corrected by simultaneously fitting each distribution in the signal region and $\text{CR}_{0,\text{low}}$. However, the normalisation of the charm background in the signal regions appears to be overestimated after applying the ABCD-extracted normalisation correction factors (*c.f.* Table 5.11). To evaluate the impact of the overestimated $B \rightarrow X_c \ell \nu$ normalisations, the injected signal setup described in Section 5.9.5 is used. The normalisation of the signal injected in data is fixed to unity and the $B \rightarrow X_c \ell \nu$ template in $\text{CR}_{K,\text{high}}$ is scaled up by an *overestimation factor* in the range [1.05, 1.10, 1.15, 1.20]. After correcting for the bias described in Section 5.9.5, a slight bias increasing linearly with the $B \rightarrow X_c \ell \nu$ overestimation factor is observed as shown in Figure 5.39. In the three considered signal regions, it is, however, not possible to extract precisely the true $B \rightarrow X_c \ell \nu$ normalisation without making an assumption on the $B \rightarrow X_u \ell \nu$ normalisation which is extracted from the fit. By simply checking the ratio of experimental data yields over simulated data yields and assuming that the signal normalisation is close to one, it is estimated that the $B \rightarrow X_c \ell \nu$ normalisation is overestimated by a factor roughly in the range 1.10 – 1.15 in all three regions. A

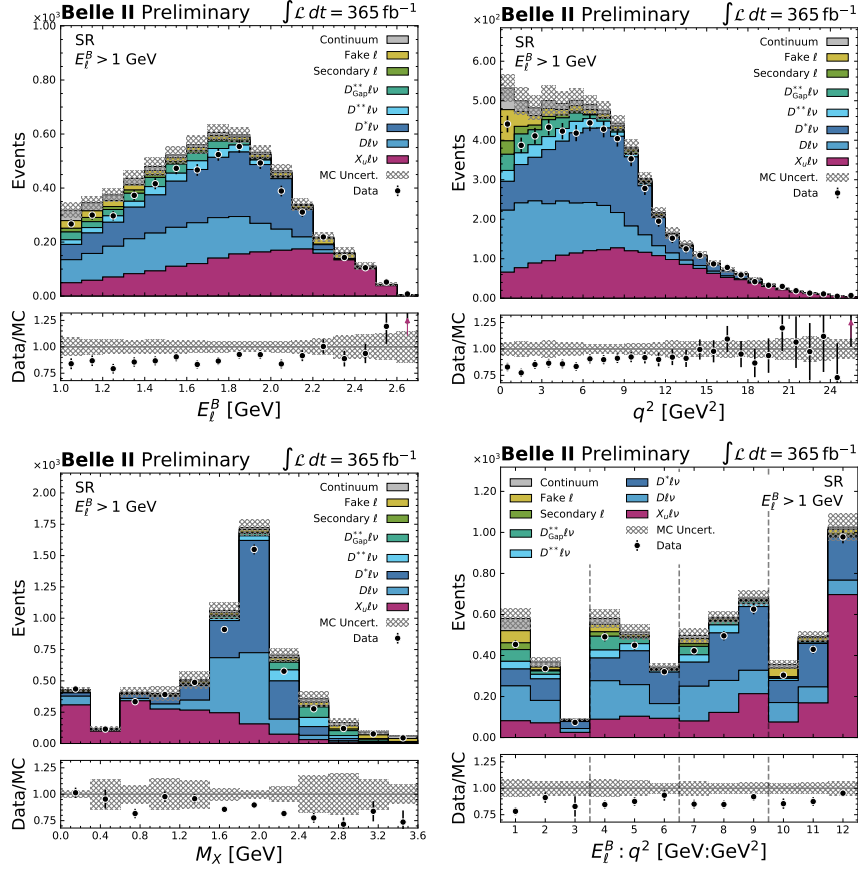


FIGURE 5.38: Distribution of E_ℓ^B (top left), q^2 (top right), M_X (bottom left) and $E_\ell^B : q^2$ (bottom right) in the signal region with the fit 1 kinematic selection. The $B \rightarrow X_c \ell \nu$ component has been scaled by the normalisation correction factor extracted via the ABCD method. The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

flat correction factor is therefore extracted by assuming that the overestimation factor is equal to 1.15. This correction factor is equal to 1.026, 1.004 and 1.013 for fits 1, 2 and 3, respectively. The full difference between the measured branching fraction (after applying the bias correction described in Section 5.9.5) and the branching fraction corrected by this factor is added as an uncertainty. The bias correction related to the overestimation of the $B \rightarrow X_c \ell \nu$ normalisation is tested assuming various overestimation factors in the range $[1.00 - 1.30]$. In all cases, the uncertainty assigned to the corrected output signal normalisation is large enough to cover the true injected signal normalisation within one standard deviation as illustrated in Figure 5.40 for the fit 1 setup.

5.10.2 Split fits

The final pre-unblinding test consists in splitting the nominal signal and control regions into two similar samples each, perform the signal extraction for each of the two subsamples and check that the two measured $B \rightarrow X_u \ell \nu$ branching fractions agree within uncertainties. In this test, the bias corrections discussed in the previous section and this one are applied and a hidden random offset is added equally to both measured branching fractions, keeping their absolute values blinded and only the difference between the resulting values is checked. The test is repeated for three sample splits based on the missing momentum polar angle ($\theta_{\text{miss}} < \pi/2$, $\theta_{\text{miss}} \geq \pi/2$), the

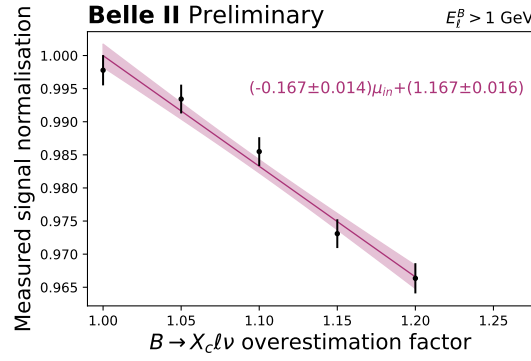


FIGURE 5.39: Measured signal normalisation as a function of the $B \rightarrow X_c \ell \nu$ overestimation factor in the injected signal setup with the fit 1 kinematic selection. The normalisation of the injected signal is set to 1 and the bias described in Section 5.9.5 is corrected for. The black points are the normalisations obtained from the test and the purple line is the linear curve fitted through the points. Its fitted parametrisation is shown in purple on the figure.

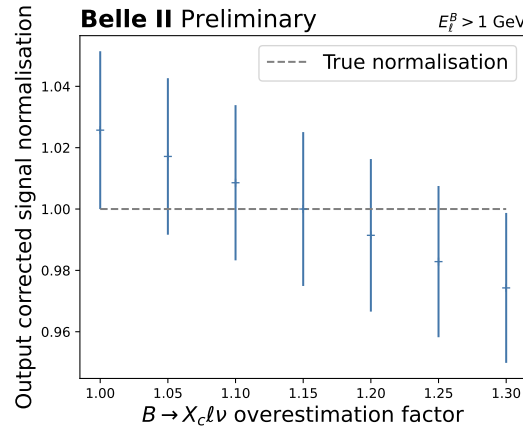


FIGURE 5.40: Corrected output signal normalisation as a function of various $B \rightarrow X_c \ell \nu$ normalisation overestimation factors in the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit 1 setup with injected signal. The output normalisation is already corrected for the source of bias described in Section 5.9.5. It is then scaled by a factor of 1.026 to correct for the source of bias described in this section and the difference between the scaled and non-scaled value is added as an uncertainty. The normalisation of the injected signal is shown as a grey dotted line.

lepton flavour (e, μ)¹³ and the lepton charge (positive, negative). To illustrate the procedure, the θ_{miss} distribution and the $E_\ell^B:q^2$ distribution split in low/high- θ_{miss} in the signal region used for fit 1 are shown in Figure 5.41 and the same $E_\ell^B:q^2$ distribution split in electron/muon channel is shown in Figure 5.42.

In each case, the two split samples are statistically independent but it is reasonable to assume that the two fits are systematically correlated to some extent. As it is not straightforward to quantify numerically this correlation, we compute the uncertainty of the difference between

¹³A small caveat is worth mentioning for the e/μ split fit. As discussed in Section 5.7, most fake lepton events occur in the μ channel which results in significantly different yields of the *other backgrounds* template in the e and μ split fits. This difference can slightly impact the result, especially in fit 1, where that template is assigned an additional constrained normalisation nuisance parameter.

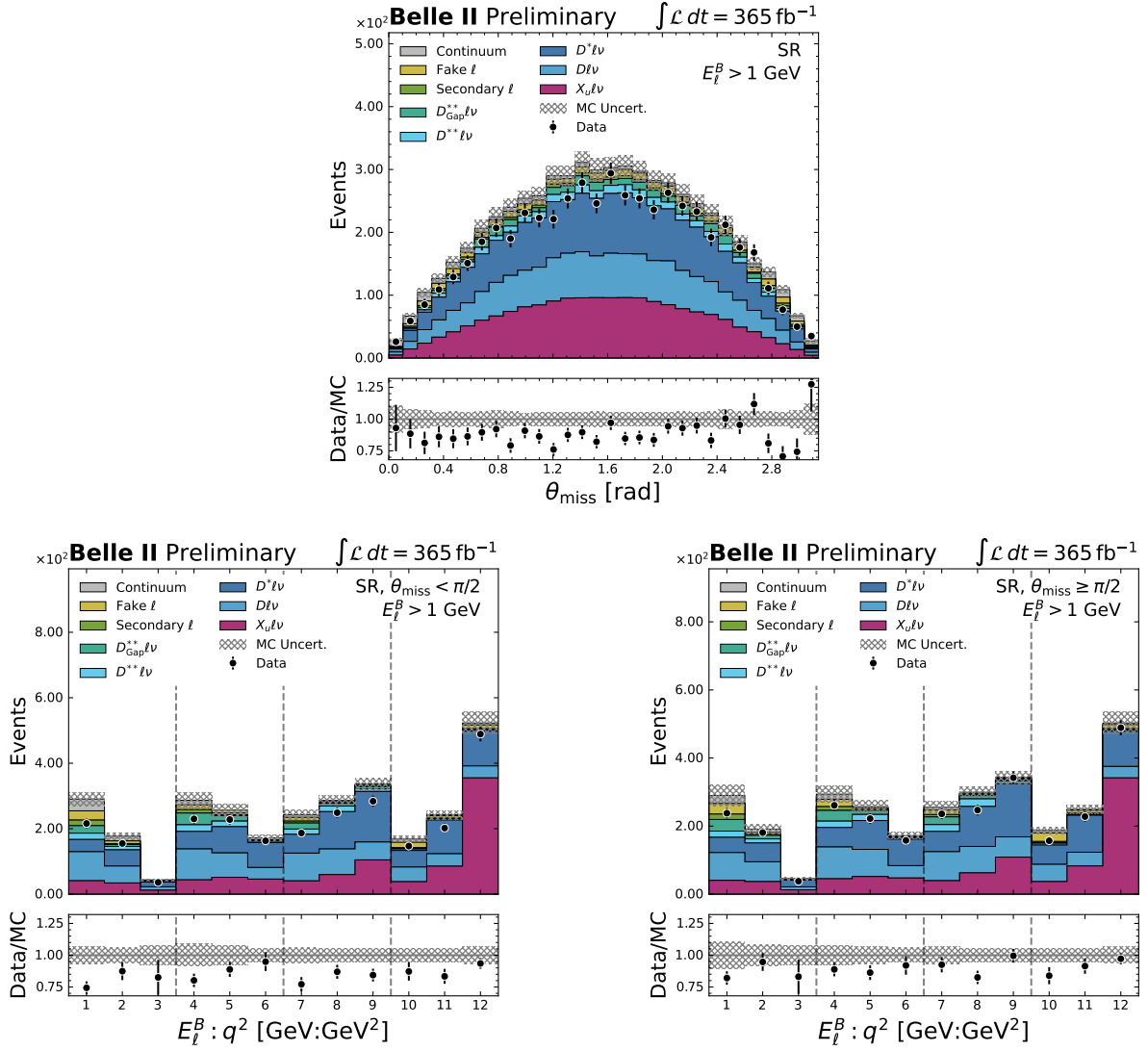


FIGURE 5.41: θ_{miss} distribution (top) and fit 1 $E_\ell^B : q^2$ distribution split in low ($< \pi/2$, bottom left) and high ($\geq \pi/2$, bottom right) θ_{miss} regions. The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

the two measured branching fractions, assuming that they are either not correlated or fully correlated. The difference in standard deviations for each split and each fit is given in Table 5.17. The largest observed difference is about one standard deviation, which further confirms the fit stability.

Split	Fit 1	Fit 2	Fit 3
$\theta_{\text{miss}} < / \geq \pi/2$	$0.4 - 0.6\sigma$	$0.2 - 0.2\sigma$	$0.4 - 0.6\sigma$
e / μ	$1.0 - 1.3\sigma$	$1.0 - 1.4\sigma$	$0.8 - 1.3\sigma$
ℓ^+ / ℓ^-	$1.0 - 1.3\sigma$	$0.8 - 1.1\sigma$	$0.4 - 0.6\sigma$

TABLE 5.17: Difference in standard deviations between the branching fractions measured in each split fit. The two quoted number are an upper (100% systematic correlation) and lower bound (0% systematic correlation) on the difference.

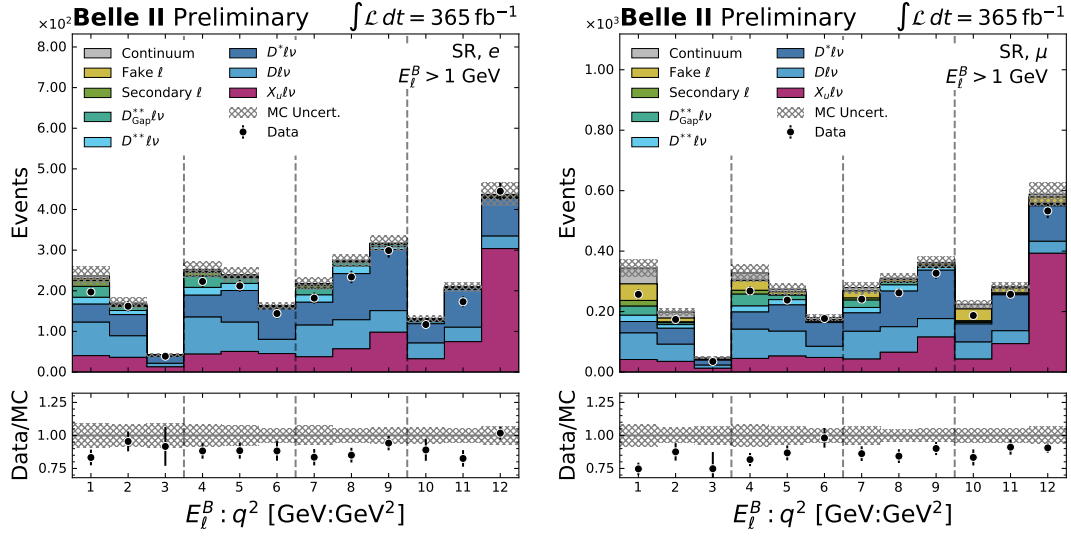


FIGURE 5.42: Fit 1 $E_\ell^B : q^2$ distribution split in events where the signal lepton candidate is reconstructed as an electron (left) or a muon (right). The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

5.11 Results

5.11.1 Prefit distributions

The CR_{0,low} and signal region distributions fitted with the three sets of kinematics selections are shown in Figure 5.43.

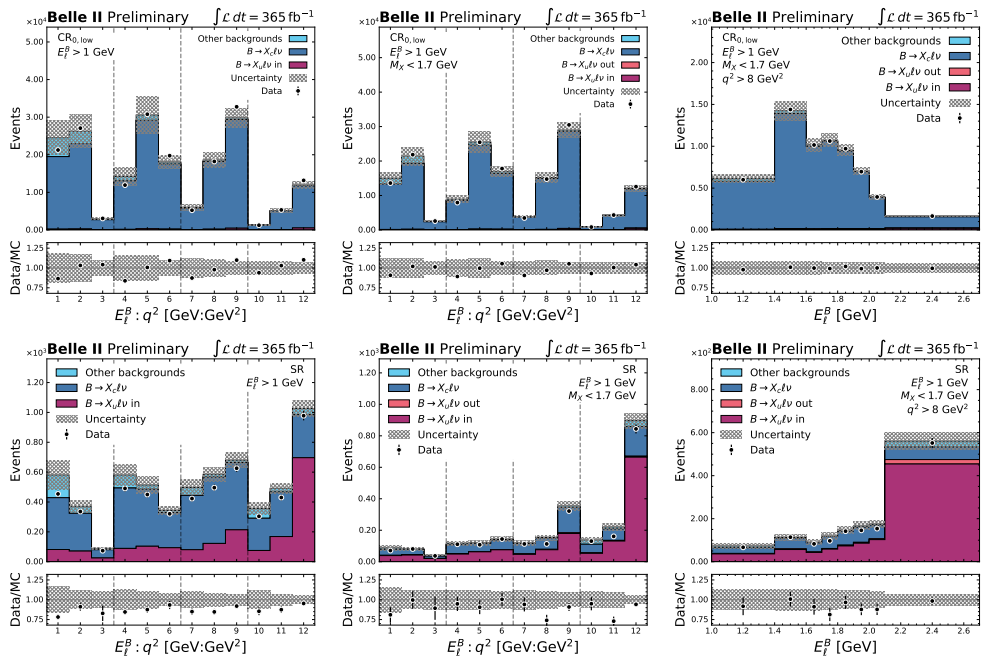


FIGURE 5.43: Prefit control region (top) and signal region (bottom) distributions for fit 1 (left), 2 (middle) and 3 (right). The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

5.11.2 Fit outcome

We perform all three fits with collected data in the signal and control regions and we obtain p -values of 0.79, 0.20 and 0.84 for fits 1, 2 and 3, respectively. The postfit distributions in both regions for each of the three fits are shown in Figure 5.44. Similarly to what is shown in Figure 5.34 for the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit, we compare in Figure 5.45 the prefit $B \rightarrow X_c \ell \nu$ data over simulated data ratios in $\text{CR}_{0,\text{low}}$ and the signal region to the $B \rightarrow X_c \ell \nu$ shape factors obtained from each fit.

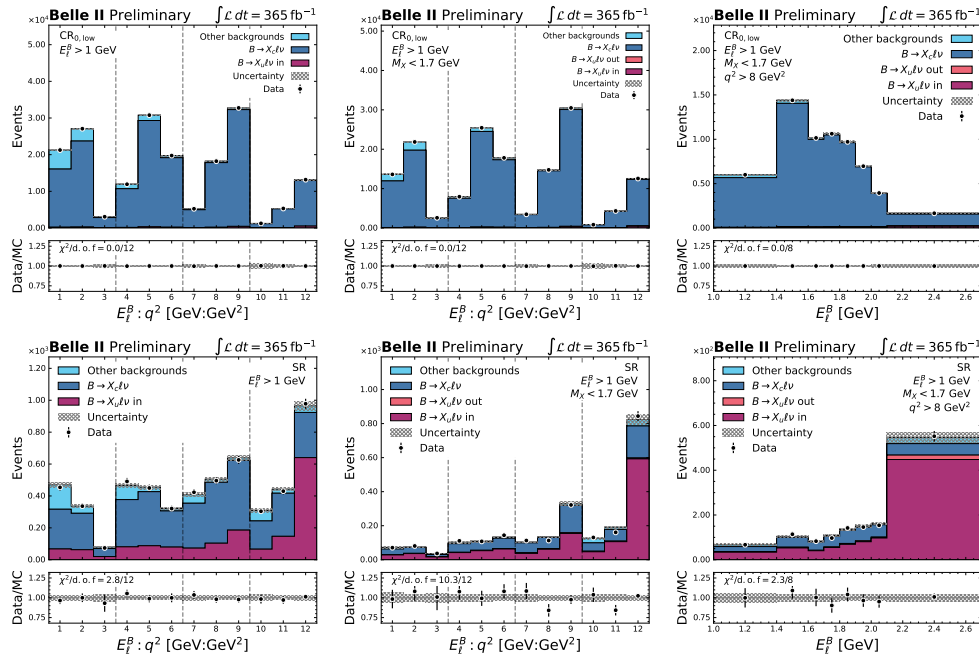


FIGURE 5.44: Postfit control region (top) and signal region (bottom) distributions for fit 1 (left), 2 (middle) and 3 (right). The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

The event yields as extracted from fits 1, 2 and 3 are given in Table 5.18. In the two regions considered for fits 1 and 2, the signal yields are similar to those obtained in Ref. [170] using a 711 fb^{-1} dataset collected by the Belle detector but the background yields are about 30% – 45% lower. The increased signal purity compared to the Belle measurement can be explained by a significantly higher hadronic tagging efficiency. The Belle II FEI (described in Section 4.3) includes more B decay chains compared to the Belle-equivalent Full Reconstruction algorithm [116], and applies new selection criteria for B meson candidates. In addition, we used two classifiers to suppress continuum and $B \rightarrow X_c \ell \nu$ backgrounds separately instead of a single classifier for both background sources. The employment of a more modern hyperparameter optimisation via **Optuna** [111] and neural networks instead of boosted decision trees can also explain the higher signal purity obtained. Finally, the increased reconstruction efficiency of slow pions helps reject a larger fraction of $B \rightarrow D^* \ell \nu$ events. The pulls of all nuisance parameters are shown in Figure 5.46 for fit 1 and in Appendix F for fits 2 and 3. In each fit, all pulls are within one standard deviation, indicating that the fits are stable and yield reliable results. As expected from the tests discussed in Section 5.10, the $B \rightarrow X_c \ell \nu$ normalisation parameter in the signal region pulls significantly to adjust the overestimated normalisation of this component.

Similarly to the injected signal fits described in Section 5.9.5, we project the results from fits 1 to 3 onto other variables. The projections of fit 1 onto E_ℓ^B , q^2 , M_X and M_{miss}^2 in the signal

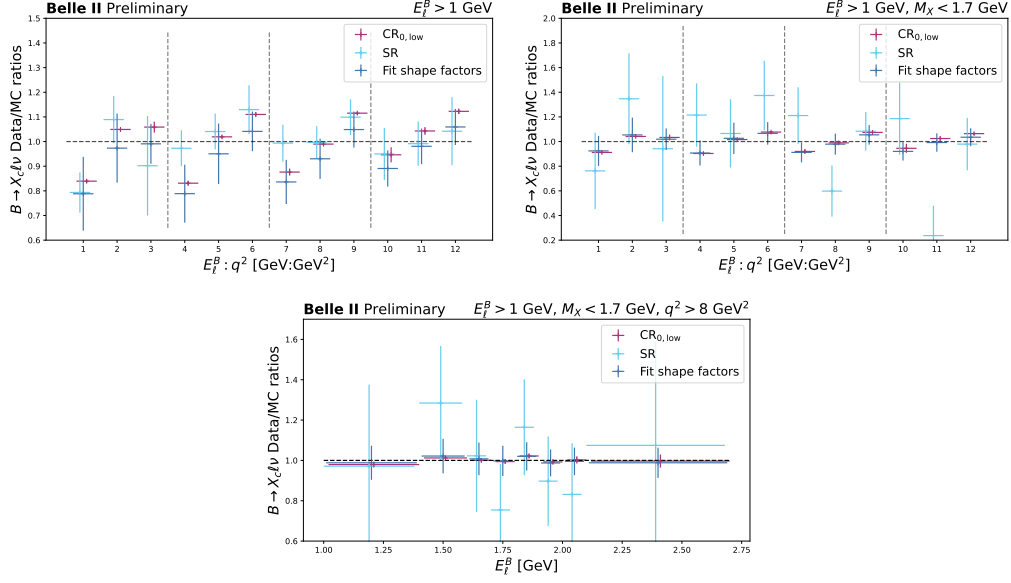


FIGURE 5.45: The data over simulated data yield ratios in $\text{CR}_{0,\text{low}}$ (purple) and the signal region (cyan) in fits 1 (upper left), 2 (upper right) and 3 (bottom) are shown. These ratios are compared to the $B \rightarrow X_c \ell \nu$ shape factors extracted from each fit (blue). The numbering of $E_\ell^B : q^2$ bins is given in Figure 5.26. The vertical dotted lines delimit the four E_ℓ^B bins.

Fit	$B \rightarrow X_u \ell \nu$ in	$B \rightarrow X_u \ell \nu$ out	Backgrounds	Data
Fit 1	1615 ± 122	—	3782 ± 217	5383 ± 73
Fit 2	1240 ± 105	33 ± 5	965 ± 95	2236 ± 47
Fit 3	876 ± 68	42 ± 4	439 ± 62	1355 ± 37

TABLE 5.18: Signal, background and data yields in the signal region as extracted from each fit. The error given for $B \rightarrow X_u \ell \nu$ and background yields is the total uncertainty as obtained from each fit and the uncertainty given for data yields is statistical.

region are shown in Figure 5.47 from which we can see that the agreement between simulated and data distributions improves for all variables and in particular the leptonic ones since fit 1 uses the $E_\ell^B : q^2$ distribution. The agreement also improves in the control region for each variable.

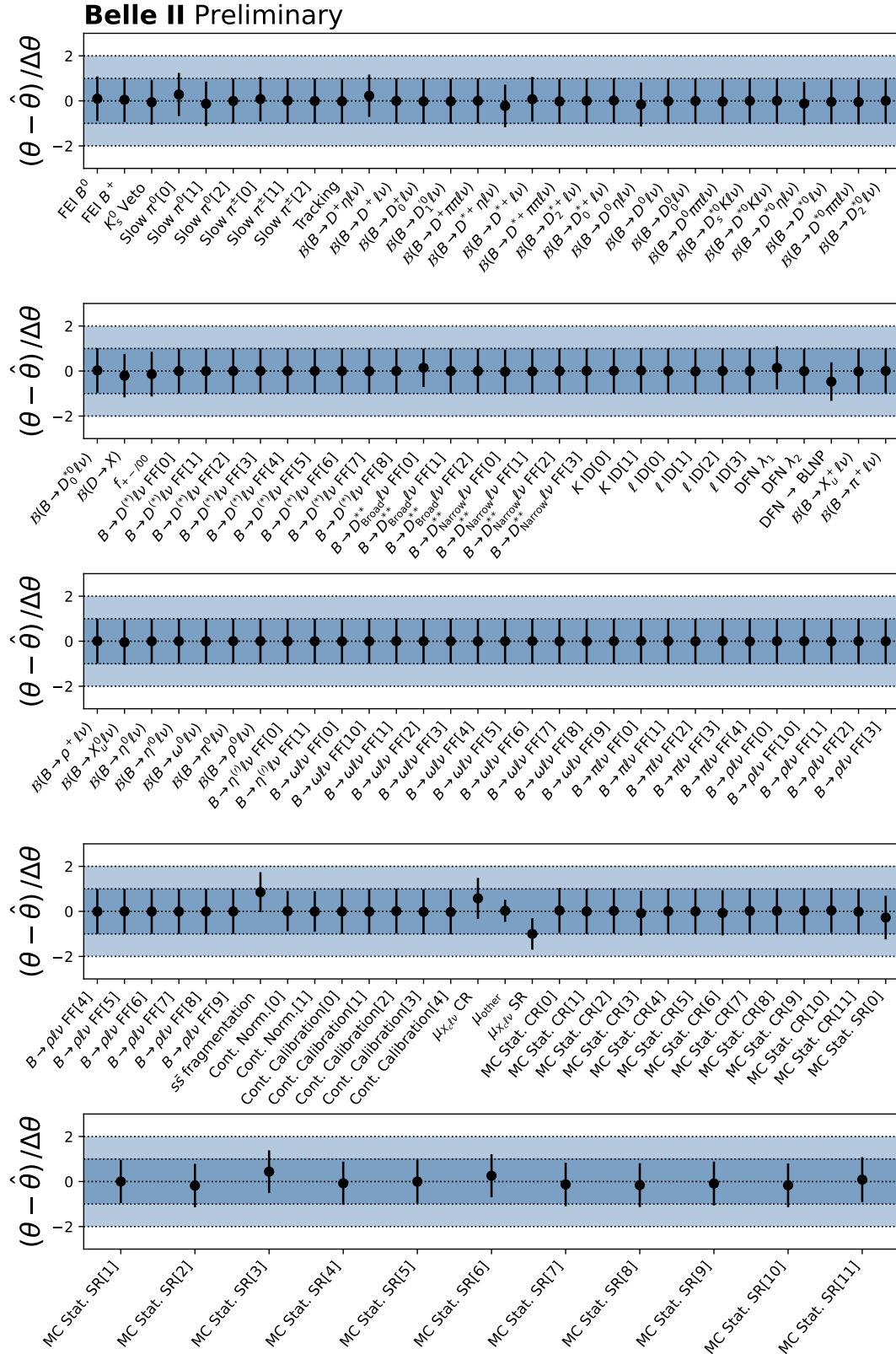


FIGURE 5.46: Nuisance parameter pulls in fit 1. The acronyms *FF*, *ID*, *Cont.* and *Norm.* are used for form factors, identification, continuum and normalisation, respectively. Sets of parameters related to the same source of uncertainty are indexed by the numbers in brackets.

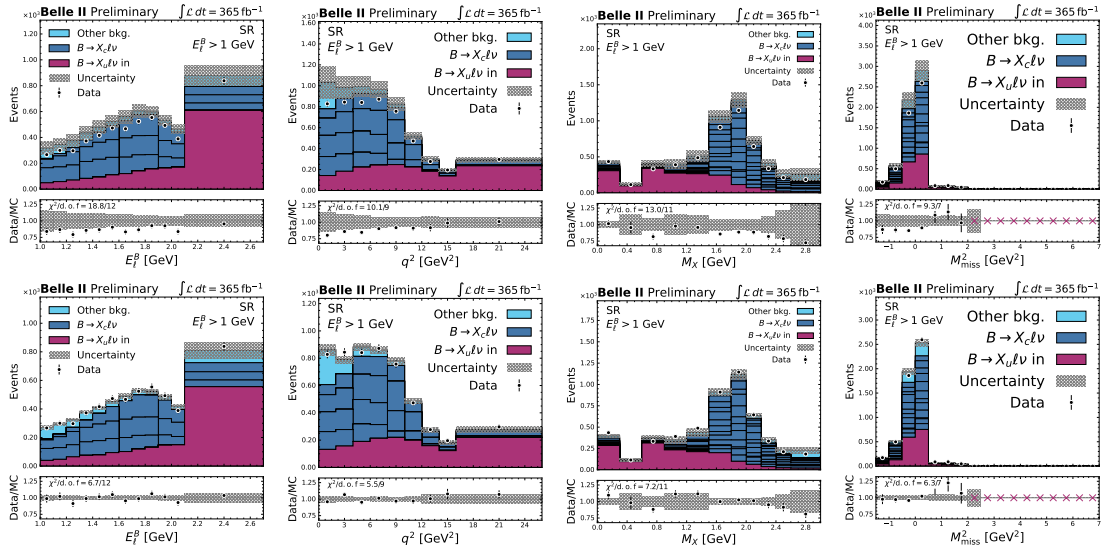


FIGURE 5.47: Top: distributions of E_ℓ^B (left), q^2 (left centre), M_X (right centre) and M_{miss}^2 (right) in the signal region with $E_\ell^B > 1$ GeV. Bottom: projections of fit 1 results onto the E_ℓ^B , q^2 , M_X and M_{miss}^2 distributions. The $B \rightarrow X_c \ell \nu$ template is split into 12 templates, one for each bin of $E_\ell^B:q^2$. Similar figures are shown for fits 2 and 3 in Appendix F.

5.11.3 Partial branching fractions

The $B \rightarrow X_u \ell \nu$ normalisations obtained from fits 1 to 3 respectively read 0.926 ± 0.076 , 0.928 ± 0.081 and 0.965 ± 0.088 , where the total fit uncertainty is given. However, two sources of bias in the fit were identified and the procedure to correct these biases and factor in related uncertainties was described in Sections 5.9.5 and 5.10.1. We summarise the entire procedure by giving a numerical example for fit 1:

- Correct the output normalisation by $(\mu_{\text{out}} - 0.052)/0.920$, which gives 0.950 ± 0.083 (this first step indirectly increases the relative uncertainty because of the subtraction)
- add a 2.6% uncertainty on the corrected normalisation based on the tests performed by varying the datasets used to extract the correction factor: 0.950 ± 0.087
- add a 1.4% uncertainty based on the composition difference between $\text{CR}_{K,\text{high}}/\text{SR}$ and $\text{CR}_{K,\text{low}}/\text{CR}_{0,\text{low}}$: 0.950 ± 0.088
- correct for the bias coming from the overestimation of the $X_c \ell \nu$ normalisation in the SR using a flat correction factor of 1.027 and add the difference as an uncertainty: 0.975 ± 0.094

Finally, we add the X_u fragmentation uncertainty derived postfit as explained in Section 5.9.4. This source yields a relative uncertainty of 0.3%, 4.4% and 3.9% for fits 1, 2 and 3, respectively. The uncertainties are relatively large for fits 2 and 3 as they are sensitive to the signal modelling because of the larger signal purity. In addition, there appears to be some mismodelling of the charged pion multiplicity distribution in the signal region. This mismodelling is unlikely to stem only from the X_u hadronisation modelling but because of the large signal purity in these two regions, we adopt these conservative uncertainties. The X_u fragmentation uncertainty is smaller in fit 1 because the scale factors extracted from the pion multiplicity mostly shift the number of signal events in low-signal-yield bins. Because of this effect, most of the signal normalisation shift is absorbed in other fit parameters.

The same correction procedure as described above is applied for the signal normalisations obtained from fits 2 and 3. The numerical values of correction factors and uncertainties used for these two values can be found in Sections 5.9.5 and 5.10 and Appendix F. Therefore, the inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions obtained for fits 1, 2 and 3 read, respectively

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)_1 = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3}, \quad (5.13)$$

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)_2 = (0.95 \pm 0.05 \pm 0.10) \times 10^{-3}, \quad (5.14)$$

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)_3 = (0.55 \pm 0.03 \pm 0.05) \times 10^{-3}, \quad (5.15)$$

where the uncertainties are statistical and systematic, respectively. With a relative uncertainty of 9.6%, the measurement of the branching fraction over a phase-space region defined by a selection $E_\ell^B > 1$ GeV is competitive with the most precise measurement performed by the BaBar collaboration (8.1% relative uncertainty) using an inclusive tagging method [74]. It is also more precise than similar measurements performed by the Belle (11.4%) and BaBar (14.5%) collaborations using a hadronic tagging method [170, 171]. With relative uncertainties of 11.9% and 10.7%, the measurements of the branching fraction over the two tighter phase-space regions are competitive with the most precise determinations over the same regions. Part of the systematic uncertainty of the branching fractions that we measure is correlated with the uncertainty from other measurements, mainly because of uncertainties related to the theoretical description of $B \rightarrow X_u \ell \nu$ decays (the DFN model input parameter uncertainties for instance). As it is not straightforward to derive correlations between different measurements, especially when systematic effects are propagated through a fit, we do not attempt to quantify the agreement between

our results and results obtained in the past. Comments about the agreement or disagreement between values are therefore purely qualitative.

When extrapolated to the full phase-space region, the three measured branching fractions in fits 1, 2 and 3 are $(1.78 \pm 0.17) \times 10^{-3}$, $(1.67 \pm 0.20) \times 10^{-3}$ and $(1.76 \pm 0.19) \times 10^{-3}$, respectively. They can be compared to the HFLAV average value¹⁴ of $(1.92 \pm 0.21) \times 10^{-3}$ [63]. In Figure 5.48, the branching fractions extrapolated to the full phase-space region are compared to the HFLAV average and the most recent Belle and BaBar measurements as listed in the PDG [24].

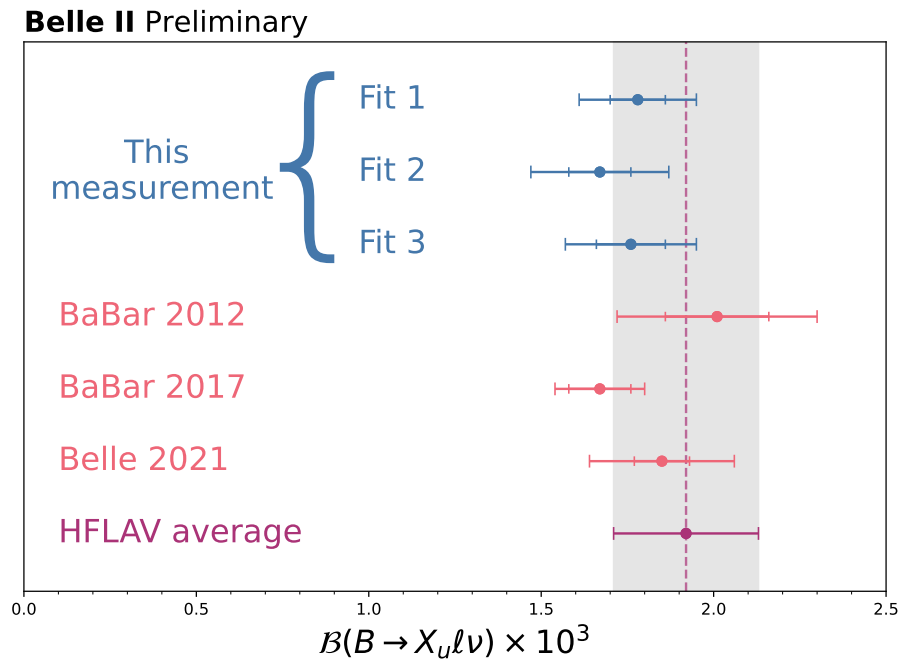


FIGURE 5.48: Comparison between the most recent branching fractions measured by BaBar (2012 [171], 2017 [74]) and Belle [170] in pink, the HFLAV average [63] in purple and the three branching fractions obtained from our measurement in blue. The inner error bar of individual measurements corresponds to the statistical uncertainty and the outer error bar represents the total uncertainty. The uncertainty of the branching fraction quoted by HFLAV is inflated based on the different results entering the average so it cannot be compared to total uncertainties of individual measurements.

Since the three branching fractions that we measure are extracted from subsets of the same data set, we check that they agree with each other within uncertainties. For this test, we use the output signal normalisations given above (or, equivalently, the branching fractions extrapolated to the full phase-space). Estimating precisely the systematic correlations between our fits is not possible, as any existing method relies on varying independently the fit parameters, which would ignore the (sizeable) correlations between parameters. It is, however, possible to estimate the correlations between the statistical uncertainties of our results via a Poisson bootstrapping [172]. The correlation matrix of the statistical uncertainties of the three fits is given in Table 5.19. The statistical correlation is similar between each pair of fits because there is a large overlap in signal events used in each one of them. In fact, about 80% (50%) of signal events passing the fit

¹⁴The uncertainty on the value quoted by HFLAV is inflated based on the different results entering the average. The branching fractions we quote were simply scaled by the phase-space acceptances given in Table 5.7 but they lack theoretical uncertainties related to this procedure.

1 kinematic selection also pass the fit 2 (3) kinematic selections. Furthermore, in all three fits, the measurement is sensitive to the region with best signal purity *i.e.* the region with high E_ℓ^B and q^2 (see Figure 5.38). Since the kinematic requirements tend to select exactly these types of events, there is a large overlap in events in this region, hence the large statistical correlation between branching fraction uncertainties. These events are grouped into a single broad bin to ensure that a meaningful number of events fall in the corresponding bin in the control region. Moreover, this binning reduces the dependence on the modelling of the kinematic endpoint, which is known to vary significantly from one theoretical framework to another. Assuming a systematic correlation of 100% and using the statistical correlations given in Table 5.19, we confirm that the measured branching fractions agree with each other within about one standard deviation. The statistical correlations that we quote must be factored in when calculating world averages like is done by the HFLAV collaboration [63].

	Fit 1	Fit 2	Fit 3
Fit 1	1	0.86	0.81
Fit 2	0.86	1	0.83
Fit 3	0.81	0.83	1

TABLE 5.19: Correlations between statistical uncertainties of the branching fractions measured in fits 1, 2 and 3.

The impact of the 20 most important nuisance parameters on the signal normalisation extracted from fits 1, 2 and 3 is shown in Figure 5.49. The systematic uncertainty breakdown in terms of individual sources of uncertainty is shown in Table 5.20. The method used to obtain values of the uncertainty related to individual sources is described in Section 5.9.5. As already noted, the method is approximate but yields values of these uncertainties which are representative of the impact of each source. For all three measured branching fractions, the DFN model parameter uncertainties represent one of the largest sources of error. It is larger in fit 3 than in the other two fits because the modelling of the E_ℓ^B distribution is known to be imperfect [74, 170], especially in the endpoint region, and because of the large signal purity in the fit 3 phase-space region. Fitting q^2 in addition to E_ℓ^B helps mitigate uncertainties related to the inclusive $B \rightarrow X_u \ell \nu$ modelling. Furthermore, all three fits are impacted by the normalisation of the $B \rightarrow X_c \ell \nu$ component in the signal region, whose effect is magnified by the large difference in normalisation between the $B \rightarrow X_c \ell \nu$ component in experimental and simulated data. In addition, the branching fraction extracted from fits 2 and 3 is also impacted by switching the inclusive $B \rightarrow X_u \ell \nu$ modelling to the BLNP framework. Inclusive $B \rightarrow X_u \ell \nu$ models are known to be more reliable for predictions made over more inclusive regions of phase-space and the signal distributions in the tighter phase-space regions used in fits 2 and 3 are therefore more sensitive to the particular choice of modelling, especially with the high signal purity in these regions. Moreover, one of the leading uncertainties of the branching fractions measured with fits 2 and 3 is the hadronic system fragmentation modelling uncertainty. As discussed in the previous section, part of the uncertainty could originate from mismodelling of backgrounds in the $\pi^\pm$ multiplicity distribution but given the large mismodelling of the $B \rightarrow X_c \ell \nu$ background and since the modelling of signal decays cannot be corrected, this methodology is employed. Finally, the postfit correlations between fit parameters look very similar to the correlations expected from Asimov fits as described in Section 5.9.5.

The uncertainties related to the bias corrections are summarised in Table 5.21 together with the systematic, statistical and total uncertainties obtained for the corrected branching fractions.

The result obtained from fit 2 is significantly impacted by the $B \rightarrow X_c \ell \nu$ composition uncertainty described in Section 5.9.5 because the relative fractions of $B \rightarrow X_c \ell \nu$ sub-components are considerably different between the signal region and $\text{CR}_{K,\text{high}}$ when applying the E_ℓ^B and M_X selections (see Table 5.12).



FIGURE 5.49: Impact of the 20 most important nuisance parameters in fits 1 (top), 2 (middle) and 3 (bottom). Each nuisance parameter θ is shifted from its central value $\hat{\theta}$. The prefit impact, shown by the cyan and pink lines, is computed by varying each nuisance parameter by the uncertainty given by its constraint term ($\Delta\theta$). The postfit impact, shown by the blue and purple boxes, is computed by varying each nuisance parameter by the uncertainty estimated from the fit ($\Delta\hat{\theta}$).

Uncertainty source	Relative uncertainty (%)		
	Fit 1	Fit 2	Fit 3
DFN parameters	4.4	4.5	5.7
DFN $\rightarrow$ BLNP	0.2	0.8	1.3
γ_s	1.7	2.1	2.1
$B \rightarrow \pi \ell \nu$ form factors	0.3	0.3	0.3
$B \rightarrow \rho \ell \nu$ form factors	0.3	0.3	0.2
$B \rightarrow \omega \ell \nu$ form factors	0.1	0.1	0.1
$B \rightarrow \eta/\eta' \ell \nu$ form factors	< 0.1	< 0.1	< 0.1
$B^\pm \rightarrow X_u \ell \nu$ branching fractions	0.9	0.6	0.5
$B^0 \rightarrow X_u \ell \nu$ branching fractions	0.6	0.5	0.5
$B \rightarrow D_{\text{Broad}}^{**}$ form factors	0.5	0.1	0.2
$B \rightarrow D_{\text{Narrow}}^{**}$ form factors	0.1	< 0.1	< 0.1
$B \rightarrow D/D^* \ell \nu$ form factors	< 0.1	< 0.1	< 0.1
$B^\pm \rightarrow X_c \ell \nu$ branching fractions	0.7	0.5	0.2
$B^0 \rightarrow X_c \ell \nu$ branching fractions	0.6	0.2	0.1
D decay branching fractions	0.1	0.3	0.1
SR $X_c \ell \nu$ normalisation	1.6	3.5	3.4
CR $X_c \ell \nu$ normalisation	0.9	1.1	0.4
Other backgrounds normalisation	0.3	N/A	N/A
X_u fragmentation	0.3	4.4	3.9
$N_{Y(4S)}$	1.4	1.4	1.4
FEI	1.3	1.3	1.4
Slow pion efficiency	0.4	0.2	0.3
ℓ identification	0.7	0.7	0.6
$f^{\pm/00}$	0.6	0.7	0.6
Continuum calibration	0.2	0.2	0.2
Tracking	0.3	0.3	0.3
K_S^0 efficiency	0.1	0.1	< 0.1
$K^\pm$ ID	< 0.1	< 0.1	< 0.1
Simulated data statistics	1.1	1.1	0.8

TABLE 5.20: Breakdown of the systematic uncertainties of the partial branching fraction obtained from each fit.

	Relative uncertainty (%)		
Fit bias	2.6	2.3	1.8
Composition uncertainty	1.3	5.7	0.2
$B \rightarrow X_c \ell \nu$ overestimation correction	2.6	0.4	1.3
Total systematic	7.9	10.5	9.7
Statistical	5.4	5.6	4.5
Total	9.6	11.9	10.7

TABLE 5.21: Breakdown of uncertainties added following the bias corrections described in Sections 5.9.5 and 5.10. In addition, we show the systematic, statistical and total uncertainties of the corrected branching fractions.

5.11.4 $|V_{ub}|$ determination

A value of $|V_{ub}|$ can be extracted from a measurement of the inclusive partial branching fraction $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ over a given phase-space via

$$|V_{ub}| = \sqrt{\frac{\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)}{\tau_B \Delta\Gamma(B \rightarrow X_u \ell \nu)}}, \quad (5.16)$$

where $\Delta\Gamma(B \rightarrow X_u \ell \nu)$ is a partial decay rate prediction over the same phase-space region (excluding the $|V_{ub}|^2$ factor) and $\tau_B = 1.578 \pm 0.003$ ps is the B meson lifetime, averaged for neutral and charged B decays [63]. Following the methodology of the HFLAV collaboration [63], three different theoretical frameworks are used to extract values of $|V_{ub}|$ (see Section 2.3.5 for a detailed discussion about these models):

- the BLNP model [54, 60] with input values in the shape-function (SF) renormalisation scheme [173, 174]: $m_b^{\text{SF}} = 4.600 \pm 0.022$ GeV and $\mu_\pi^{2\text{SF}} = 0.184^{+0.051}_{-0.062}$ GeV²
- the DGE model [64] with input values in the $\overline{\text{MS}}$ scheme: $m_b^{\overline{\text{MS}}} = 4.206 \pm 0.040$ GeV
- the GGOU model [62] with input values in the kinetic scheme [175]: $m_b^{\text{kin}} = 4.573 \pm 0.012$ GeV and $\mu_\pi^{2\text{kin}} = 0.454 \pm 0.043$ GeV²

We obtained the values of $\Delta\Gamma$ from Ref. [176]. They are based on the most up-to-date determinations of the heavy-quark parameters and are therefore considered to be the best available values. When computing the values of $|V_{ub}|$, we do not account for the mismatch between the non-perturbative parameter values used in our signal simulations and those employed in the calculation of the decay rates¹⁵. Nevertheless, the uncertainties assigned to the $B \rightarrow X_u \ell \nu$ modelling are intended to account for differences between theoretical models, thereby yielding measurements that are, *a priori*, model-independent within their uncertainties.

Furthermore, since the uncertainties on the inclusive $B \rightarrow X_u \ell \nu$ modelling enter both the measured branching fractions and the decay rate predictions, they are correlated. In our case, these uncertainties are the values of m_b and a used to model $B \rightarrow X_u \ell \nu$ decays with the DFN model and the difference between the distributions predicted by the DFN and BLNP frameworks (see Section 5.9.4). However, because the modelling uncertainties are constrained in our fits (see Figure 5.49), it is not straightforward to estimate the correlations between branching fraction and decay rate uncertainties. Besides, there exist no direct equivalent of the DFN/BLNP-difference uncertainty and it is not clear how correlated to the decay rate uncertainties it would be. The best way to properly account for correlations between an individual measurement and a decay rate prediction would be to perform the signal extraction with samples simulated with different models (or reweighted to match different models) and extract a value of $|V_{ub}|$ directly instead of a partial branching fraction. This way, the values of the input model parameters used in sample simulation and decay rate prediction would match, all correlations between related modelling uncertainties would be factored in the $|V_{ub}|$ extraction and a value of $|V_{ub}|$ would be obtained for every model¹⁶. Correlations between uncertainties represent an even more significant challenge for the computation of averages, as the correlations also arise between branching fraction uncertainties. Part of the correlations are taken into account in the HFLAV averages. For instance, statistical correlations between measurements using similar data sets are factored in and reconstruction-related uncertainties between measurements performed by the same experiment are assumed to be fully correlated. However, correlations between theory-related uncertainties

¹⁵Values in different renormalisation schemes are not identical but can be converted from one scheme to another.

¹⁶The BaBar collaboration performed a measurement of $|V_{ub}|$ which used samples based on four different models therefore extracting four different values of the $B \rightarrow X_u \ell \nu$ partial branching fraction [74].

are ignored in the average calculations. For simplicity, to extract values of $|V_{ub}|$, we follow the HFLAV procedure by ignoring correlations between branching fraction and decay rate modelling uncertainties but we stress that the HFLAV method should be improved in the future to factor in these non-negligible effects.

The values of $|V_{ub}|$ obtained from each theoretical prediction and each signal extraction fit are summarised in Table 5.22. The values of $|V_{ub}|$ extracted from each phase-space region are compared to the latest HFLAV [63] and CKMfitter averages [17] in Figure 5.50.

Inclusive $B \rightarrow X_u \ell \nu$ model		Fit 1	Fit 2	Fit 3
$\Delta\Gamma$ (ps ⁻¹)	BLNP	$62.9^{+5.3}_{-5.9}$	$47.5^{+4.5}_{-4.8}$	$24.4^{+2.6}_{-2.9}$
	DGE	$59.4^{+2.9}_{-3.2}$	$43.1^{+4.5}_{-4.2}$	$25.0^{+1.8}_{-2.0}$
	GGOU	$60.8^{+2.2}_{-2.6}$	$47.6^{+3.0}_{-3.2}$	$25.0^{+2.2}_{-2.6}$
$ V_{ub} \times 10^3$	BLNP	$3.94 \pm 0.11 \pm 0.16^{+0.17}_{-0.18}$	$3.55 \pm 0.10 \pm 0.19^{+0.17}_{-0.18}$	$3.79 \pm 0.09 \pm 0.18^{+0.20}_{-0.22}$
	DGE	$4.05 \pm 0.11 \pm 0.16^{+0.10}_{-0.11}$	$3.73 \pm 0.10 \pm 0.20^{+0.19}_{-0.18}$	$3.75 \pm 0.08 \pm 0.18^{+0.13}_{-0.15}$
	GGOU	$4.01 \pm 0.11 \pm 0.16^{+0.07}_{-0.08}$	$3.55 \pm 0.10 \pm 0.19^{+0.11}_{-0.12}$	$3.75 \pm 0.08 \pm 0.18^{+0.17}_{-0.20}$

TABLE 5.22: Theoretical $B \rightarrow X_u \ell \nu$ decay rates $\Delta\Gamma$ and $|V_{ub}|$ obtained from three different theoretical predictions in three separate phase-space regions. The values of $\Delta\Gamma$ are used to compute the values of $|V_{ub}|$ following Equation 5.16. The decay rates exclude the $|V_{ub}|^2$ factor and the uncertainty given is their total theoretical uncertainty. The uncertainties on $|V_{ub}|$ are statistical, systematic and theoretical, respectively.

The values of $|V_{ub}|$ obtained from our measurement are all slightly lower than the HFLAV inclusive average but they appear to be compatible with it within uncertainties. The values extracted from the broadest phase-space region appear to be in slight tension with the exclusive HFLAV average but in agreement with the $B \rightarrow \pi \ell \nu$ HFLAV average. On the other hand, the values obtained from the two tighter phase-space regions appear to be compatible with the exclusive HFLAV averages and the value extracted by the CKMfitter collaboration. However, given the relatively large uncertainties of the measured values and the averages, definitive conclusions can hardly be made. Since we quote the correlations between the statistical uncertainties of the three branching fractions that we measure, all three values could be included in the next iterations of average calculations performed by different collaborations. Other recent measurements of $|V_{ub}|$, such as a measurement via $B \rightarrow \pi \ell \nu$ and $B \rightarrow \rho \ell \nu$ decays [70] and a measurement of the $|V_{ub}|/|V_{cb}|$ ratio via inclusive decays [177], could also be included in these averages. The determination of $|V_{ub}|$ still suffers from relatively large discrepancies but, with different studies being performed regularly, it will be interesting to see how the $|V_{ub}|$ landscape evolves in the years ahead.

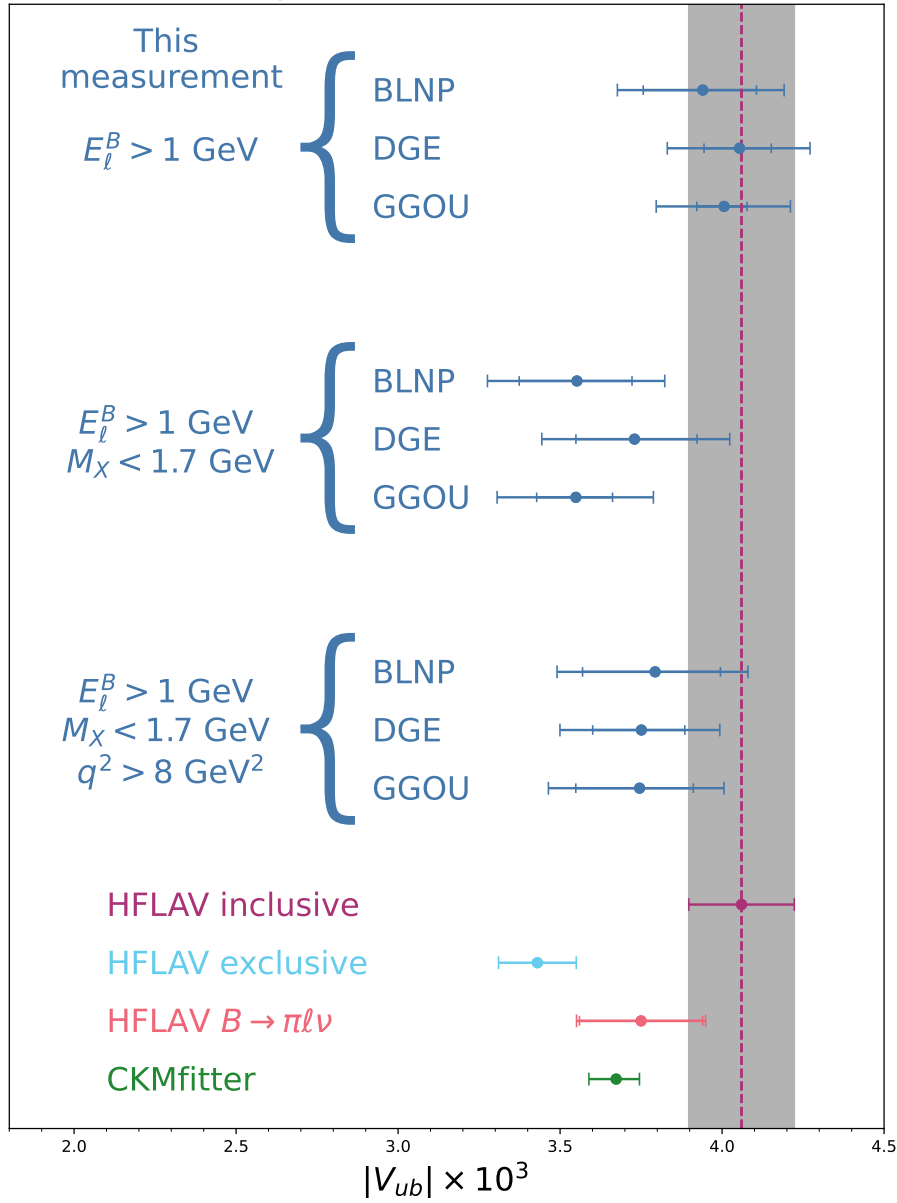
Belle II Preliminary

FIGURE 5.50: Comparison between the values of $|V_{ub}|$ obtained from fits 1, 2 and 3 (blue), the inclusive (purple), exclusive (cyan) and $B \rightarrow \pi \ell \nu$ (pink) averages quoted in the latest HFLAV report [63] and the value obtained by CKMfitter from a global unitary triangle fit using as input all CKM parameter measurements except $|V_{ub}|$ measurements [17] (green). The outer error bars represent the total uncertainty and for the nine measured values and the $B \rightarrow \pi \ell \nu$ average, the inner error bars represent the theoretical uncertainty.

Chapter 6

Conclusions and outlook

*“The best that most of us can hope to achieve in physics
is simply to misunderstand at a deeper level.”*

Wolfgang Pauli

We presented in this thesis the measurement of partial branching fractions of inclusive charmless semileptonic beauty meson decays over three different phase-space regions using 365 fb^{-1} of data collected by the Belle II detector. From each one of these branching fractions, we extract several values of $|V_{ub}|$, based on different theoretical inputs. Precise measurements of $|V_{ub}|$ are crucial to challenge the unitarity of the CKM matrix. The $|V_{ub}|$ element is currently one of the CKM parameters that is known with the lowest precision [17, 63]. Apart from the inclusive method, the value of $|V_{ub}|$ can be determined from semileptonic exclusive studies where the hadronic system is reconstructed as a specific final state such as a pion. Exclusive studies tend to measure values of $|V_{ub}|$ which are lower than those measured from inclusive studies leading to a discrepancy between the two world averages of about three standard deviations [63].

The theoretical description of inclusive $B \rightarrow X_u \ell \nu$ decays is challenging because it necessitates predictions of the $B \rightarrow X_u \ell \nu$ differential decay rate [52]. Such predictions are sensitive to a non-perturbative object called the *shape function*, which is interpreted as the distribution of the b quark momentum in the B meson. Various theoretical predictions of $B \rightarrow X_u \ell \nu$ differential decay rates have been developed, each using a different strategy to incorporate the shape function. These calculations become particularly sensitive to the shape function in tighter phase-space regions defined by kinematic selections that help suppress the overwhelming background from charmed semileptonic decays.

In our analysis, the $B \rightarrow X_c \ell \nu$ background was suppressed using a neural network classifier which allowed us to perform the measurement over a relatively broad phase-space region, defined by signal lepton energies greater than 1 GeV in the B meson rest frame. Over this region, which covers about 87% of the accessible $B \rightarrow X_u \ell \nu$ phase-space, we measured the partial branching fraction to be

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3},$$

where the uncertainties are statistical and systematic, respectively. From this measurement, using the Gambino, Giordano, Ossola and Uraltsev theoretical framework [62], the value of $|V_{ub}|$ is determined to be

$$|V_{ub}| = (4.01 \pm 0.11 \pm 0.16_{-0.08}^{+0.07}) \times 10^{-3},$$

where the uncertainties are statistical, systematic and theoretical, respectively. We performed the measurement over two tighter phase-space regions defined by hadronic system energies lower

than 1.7 GeV and lepton momentum transfer squared greater than 8 GeV². Our study yields uncertainties on the partial branching fractions and $|V_{ub}|$ which are lower than or competitive with similar measurements performed by other B factory experiments, most recently Belle and BaBar. The values of $|V_{ub}|$ that we measure are slightly lower than the inclusive $|V_{ub}|$ world average computed by the Heavy Flavour Averaging Group (HFLAV) [63] collaboration but compatible with this value within uncertainties. The values obtained from the two tighter phase-space regions, in particular, appear to agree with the exclusive $|V_{ub}|$ world average but their large uncertainties do not allow us to make any definitive conclusions. Our results could be included in inclusive $|V_{ub}|$ average calculations performed by various collaborations such as HFLAV, the Particle Data Group [24] and the CKMfitter collaboration [17].

Since our measurement uses a dataset which is about the same size of BaBar’s dataset and about half the size of Belle’s dataset, it is interesting to compare those measurements to identify the elements of our analysis which are improved and those that can be improved. First of all, we use a tagging algorithm, the Full Event Interpretation (FEI) [115], to reconstruct the partner B meson in $\Upsilon(4S) \rightarrow B\bar{B}$ decays. The FEI relies on modern boosted decision tree classifiers which were trained to identify more than 10,000 different B decay chains and offers a reconstruction efficiency which is higher than that achieved by similar algorithms used in previous measurements. Second, we used two separate neural networks to identify and suppress background originating from $e^+e^- \rightarrow q\bar{q}$ events ($q \in \{u, d, s, c\}$), on the one hand, and $B \rightarrow X_c\ell\nu$ decays, on the other hand. The second classifier relies on the poorer reconstruction of $B \rightarrow X_c\ell\nu$ events compared to $B \rightarrow X_u\ell\nu$ events because of their higher average particle multiplicity. In addition, the classifier utilises the properties of low-momentum pions to identify $B \rightarrow D^*(\rightarrow D\pi)\ell\nu$ events. The Belle II reconstruction efficiency of low-momentum pions is higher than that achieved by previous experiments. The improved tagging algorithm and background suppression classifiers allow us to achieve higher signal purities in the different signal regions relatively to other measurements. The second major improvement of this analysis resides in the signal extraction. We use a modern tool to incorporate in our fits systematic effects, correlate them to each other and constrain systematic uncertainties. We designed the fits to simultaneously extract the shape of the $B \rightarrow X_c\ell\nu$ component from data in a control region and measure the $B \rightarrow X_u\ell\nu$ branching fraction in a signal region, thus reducing the dependence of our results on the modelling of $B \rightarrow X_c\ell\nu$ decays.

Even though our nominal measurement is performed in a region covering a relatively large fraction of the $B \rightarrow X_u\ell\nu$ accessible phase-space, the results are still sensitive to the high signal purity region. The modelling of $B \rightarrow X_u\ell\nu$ decays in this region, corresponding to decays with high E_ℓ^B/q^2 and low M_X , strongly varies from one theoretical framework to another and depends in particular on the shape function implementation within these frameworks. In all three of our partial branching fraction measurements, events falling in the kinematic endpoint region are grouped in a single broad bin to reduce the model-dependence of our results. In the future, with the improvement of tagging algorithms and machine-learning-based background suppression, a high signal purity over the entire spectrum would help reduce even further the model-dependence of inclusive $B \rightarrow X_u\ell\nu$ branching fraction measurements.

It is important to note that if the modelling of the charmed semileptonic background described correctly the observed data, no data-driven extraction would be necessary. The discrepancies between simulated and experimental data induced additional uncertainties which can be avoided by improving the $B \rightarrow X_c\ell\nu$ modelling. The primary suspected cause of the observed discrepancies is the poor modelling of multibody D meson decays. An ongoing effort within the Belle II collaboration aims to replace the simple phase-space models currently used to simulate D meson decays by dedicated models based on Dalitz analyses. Furthermore, a common problem of inclusive $B \rightarrow X_u\ell\nu$ and $B \rightarrow X_c\ell\nu$ analyses lies in the definition of the decays filling the gap

between the total $B \rightarrow X_c \ell \nu$ inclusive branching fraction and the sum of exclusive branching fractions. As of now, Belle II analyses use $B \rightarrow D^{(*)} \eta \ell \nu$ decays modelled via an intermediate broad D^{**} resonance to fill this gap. However, it is currently not known whether this is a reliable description of the non-resonant processes which cause the difference between inclusive and sum of exclusive. To verify this, dedicated studies of $B \rightarrow D^{(*)} \eta \ell \nu$ and other non-resonant decays such as $B \rightarrow D \pi \ell \nu$ are necessary. There is a sustained effort on the theoretical side to develop descriptions of such decays and on the experimental side to characterise their properties.

The relatively low uncertainty of our measurement is encouraging for future determinations of $|V_{ub}|$ with inclusive decays at Belle II. However, as the total uncertainty continues to decrease, it is important to emphasize that inclusive $B \rightarrow X_u \ell \nu$ decays, while better understood theoretically than processes like pure hadronic decays, still present significant theoretical challenges. Different models predict different kinematic variable spectra and yield different values of $|V_{ub}|$. Uncertainties are added in experimental studies to cover known deviations between these models and between different shape function ansätze. The inclusive $B \rightarrow X_u \ell \nu$ modelling could nevertheless be responsible for the long standing discrepancy between exclusive and inclusive determinations of $|V_{ub}|$. Inclusive models are currently being updated with higher order contributions in the strong coupling constant and b quark mass but such models remain limited by our incomplete knowledge of the shape function. Various ideas were proposed to constrain the shape function using input from the photon energy spectrum in $B \rightarrow X_s \gamma$ decays [56, 57] and several studies of such decays using Belle II data are ongoing.

Recent inclusive $|V_{ub}|$ analyses, including the one presented here, have investigated relatively broad phase-space regions with the help of modern machine learning techniques to suppress the $B \rightarrow X_c \ell \nu$ background. Over regions which nearly cover the full phase-space, a value of $|V_{ub}|$ could be extracted by relying on the heavy quark expansion (HQE) [27, 28] rather than a specific partial decay rate prediction, similarly to the way $|V_{cb}|$ can be extracted. Such a measurement would still rely on an inclusive $B \rightarrow X_u \ell \nu$ framework for the modelling of simulated events. The high background yields would also limit the precision of the results and render the measurement dependent on high signal purity regions. In this context, a simultaneous measurement of $|V_{ub}|$ and $|V_{cb}|$ could be designed to extract an inclusive value of the $|V_{ub}|/|V_{cb}|$ ratio where theoretical uncertainties related to the HQE, in addition to experimental uncertainties, would cancel and thus yield a strong constraint on the determinations of these two CKM matrix elements.

On the other hand, the exclusive average of $|V_{ub}|$ has mainly been driven by $B \rightarrow \pi \ell \nu$ measurements [63]. The $B \rightarrow \pi \ell \nu$ decay reconstruction is clean relatively to other exclusive channels but the values of $|V_{ub}|$ extracted from such measurements are limited by the form factor theoretical uncertainty. Measurements of $|V_{ub}|$ using $B \rightarrow \rho \ell \nu$ and $B \rightarrow \omega \ell \nu$ decays have also been performed but the associated uncertainties are larger than those obtained in $B \rightarrow \pi \ell \nu$ studies. A third input on $|V_{ub}|$ is given by exclusive measurements of $|V_{ub}|/|V_{cb}|$ performed by the LHCb collaboration [72, 73]. The measured values of this ratio tend to pull the exclusive $|V_{ub}|$ average further away from the inclusive average. Interestingly, the exclusive values of $|V_{ub}|$ or $|V_{ub}|/|V_{cb}|$ extracted by relying on theoretical input from QCD light-cone sum rules are significantly lower than those obtained using lattice QCD input. Finally, a value of $|V_{ub}|$ is obtained by a global fit of the unitary triangle using all CKM matrix inputs except direct $|V_{ub}|$ measurements [17]. Such a determination of $|V_{ub}|$ is therefore orthogonal to all measurements mentioned previously but the obtained value favours the exclusive average.

With datasets much larger than those collected so far by B factories, measurements of $|V_{ub}|$ via purely leptonic decays would yield the most precise results. The only theoretical input needed to measure $|V_{ub}|$ from $B^+ \rightarrow \ell^+ \nu$ branching fractions is the B meson decay constant

which is currently known at the level of 0.7% [178]. However, because of helicity suppression, the electron, muon and tau channel branching fractions are expected to be of the order of 10^{-11} , 10^{-7} and 10^{-4} , respectively. Since the $B^+ \rightarrow \tau^+ \nu$ branching fraction is close to that of $B \rightarrow \pi \ell \nu$, it represents a promising channel for studying $|V_{ub}|$ on shorter time scales. The reconstruction of a tau is however challenging and measurements of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu)$ suffer from large backgrounds. From a combination of all measurements of the $B^+ \rightarrow \tau^+ \nu$ branching fraction, the decay has only been observed at the level of about 4.5σ [24].

Despite numerous measurements using a wide variety of experimental techniques and theoretical inputs, the $|V_{ub}|$ puzzle remains unsolved. Several additional inputs will be included in the next HFLAV report, so it remains to be seen how the overall picture evolves. The current uncertainty on the value of $|V_{ub}|$ limits the precision of the CKM unitarity test. To be able to challenge the Standard Model of particle physics, a more precise value of $|V_{ub}|$ is therefore needed.

Final words from the author

Studies of inclusive $B \rightarrow X_u \ell \nu$ decays are extremely challenging and the poor modelling of $B \rightarrow X_c \ell \nu$ decays that we observed made it even more challenging. I am pleased to see that we succeeded in reaching uncertainties which are competitive with previous measurements despite these difficulties. As highlighted above, with a larger dataset and a better modelling, the inclusive value of $|V_{ub}|$ could be measured at Belle II with world record precision in the near future. Over the time necessary to complete this PhD, I have come to understand the importance of measuring $|V_{ub}|$ to improve the overall precision of the unitarity triangle tests. I hope that based on the ideas mentioned above, we can resolve the long-standing tension between the exclusive and inclusive values of $|V_{ub}|$ and that we can reach the precision levels of other CKM matrix parameters. There is still much work to be done in this exciting subfield of flavour physics and I hope that I can contribute to it further in the future. Apart from the end result, I am also happy that, while doing the measurement, I learned a lot about particle physics, both on the theoretical and experimental levels. And finally, to paraphrase Pauli, I probably misunderstand inclusive semileptonic decays now as much as I did when I started working on the measurement. But at a much deeper level.

Appendix A

Simulation of inclusive $B \rightarrow X_u \ell \nu$ decays

Two inclusive $B \rightarrow X_u \ell \nu$ models have been implemented in the `EvtGen` generator and can therefore be used to simulate such events: DFN and BLNP (see Section 2.3.5 for a detailed description of both of these theoretical frameworks). For the study presented in this manuscript, it was initially planned to use the BLNP model as nominal since the DFN model is considered to be outdated whereas BLNP relies on complex theoretical techniques to yield predictions of the triple differential decay rate over arbitrary regions of phase-space at second order in the heavy quark expansion. However, this model suffers from several drawbacks which we discuss here. First of all, it was observed in a BaBar study of inclusive charmless decays that the lepton energy endpoint region is poorly modelled by BLNP compared to other theoretical predictions [74]. Second, the number of negative hybrid weights obtained with BLNP is much larger than with DFN. The hybrid weights are calculated from Equation 5.1 and if the number of resonant events in a given bin exceeds the number of non-resonant events, the weight for this bin is negative and is simply set to 0 to avoid yielding negative numbers of events. If this happens for a limited number of bins, the impact of negative hybrid weights can be neglected. This is the case when using DFN but the amount of negative hybrid weights becomes problematic with BLNP. Finally, we observed in two-dimensional distributions of kinematic variables, unphysical spikes of events at the edges of the $B \rightarrow X_u \ell \nu$ phase-space as shown in Figure A.1. From this peculiar behaviour of BLNP, we concluded that the model, as implemented in `EvtGen`, is not viable for event simulation.

A possible explanation for the unexpected spikes originates from the way BLNP implements the soft-collinear effective theory expansion. The model is optimised for the shape function region where $P_+ \ll P_-$ and it then interpolates to the full phase-space. However, in regions where $P_+ \sim P_-$, the BLNP differential decay rate calculation yields contributions $\mathcal{O}(1)$, therefore enhancing the rate at the edges of the phase-space as illustrated in Figure A.2.

This hypothesis is further confirmed if the $P_+ \sim P_-$ region is selected by requiring that $\frac{1}{\sqrt{2}}(P_- - P_+) < 0.4$ GeV. As illustrated in Figure A.3, this selection isolates the problematic events in the two-dimensional distributions.

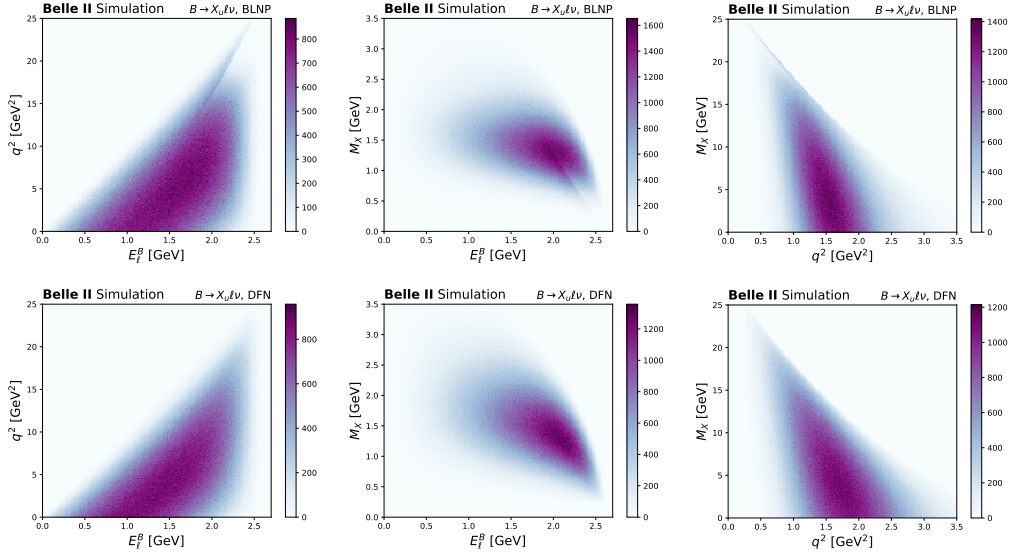


FIGURE A.1: Comparison of generated two-dimensional distributions of $E_\ell^B:q^2$ (left), $E_\ell^B:M_X$ (middle) and $q^2:M_X$ (right) of $B \rightarrow X_u \ell \nu$ events with the BLNP (top) and DFN (bottom) models. Clear unphysical spikes outside of the bulk of events can be observed for the BLNP distributions but not the DFN ones. The distributions use 4M charged and 4M neutral B decays.

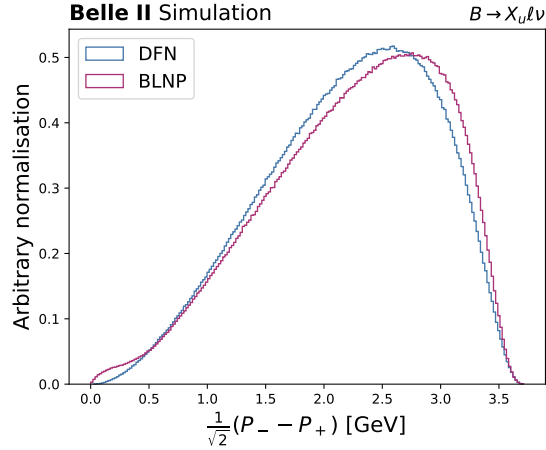


FIGURE A.2: Comparison of generated one-dimensional $\frac{1}{\sqrt{2}}(P_- - P_+)$ distribution of $B \rightarrow X_u \ell \nu$ events with the BLNP (purple) and DFN (blue) models. An excess of events can be observed in the BLNP distribution in the region where $P_+ \sim P_-$.

The distribution uses 4M charged and 4M neutral B decays.

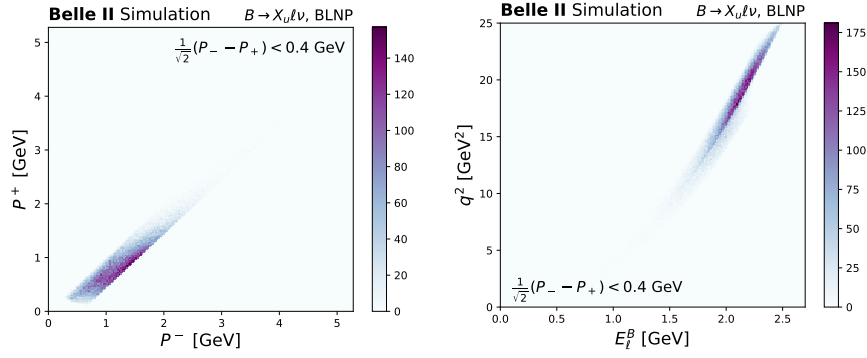


FIGURE A.3: Generated two-dimensional distributions of $P_- : P_+$ (left) and $E_\ell^B : q^2$ (right) of $B \rightarrow X_u \ell \nu$ events in a region where $\frac{1}{\sqrt{2}}(P_- - P_+) < 0.4 \text{ GeV}$. The spike of events observed in the $q^2 : E_\ell^B$ distribution clearly corresponds to events with $P_+ \sim P_-$. The distributions use 4M charged and 4M neutral B decays.

Appendix B

Input features of the continuum suppression classifier

The distributions of the 34 variables that can be used as input features for the continuum background suppression classifier are shown below after reconstruction and preselection (control region 1). We select a subset of well-modelled variables among the 34 available. All of these variables were introduced in Section 5.5.1. The 30 event shapes, the transverse energy and the missing mass squared are used to calibrate the continuum component using off-resonance samples as described in Section 5.4.4. Only the simulated data statistical uncertainty is shown here.

B.1 Thrust-axis variables

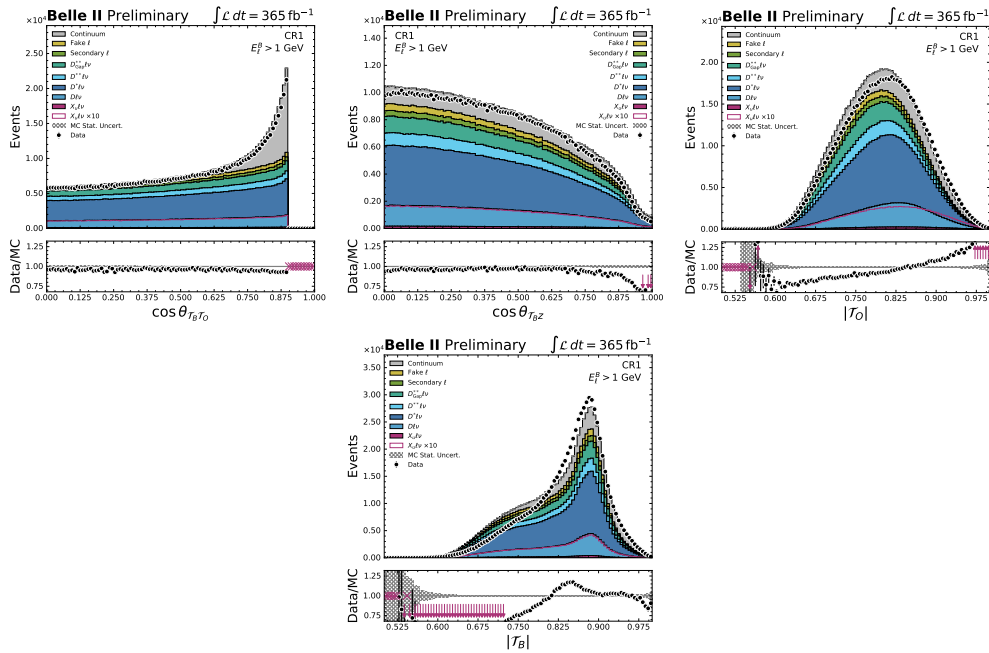


FIGURE B.1: The four thrust-axis variables in CR1

B.2 Kakuno-Super-Fox-Wolfram moments

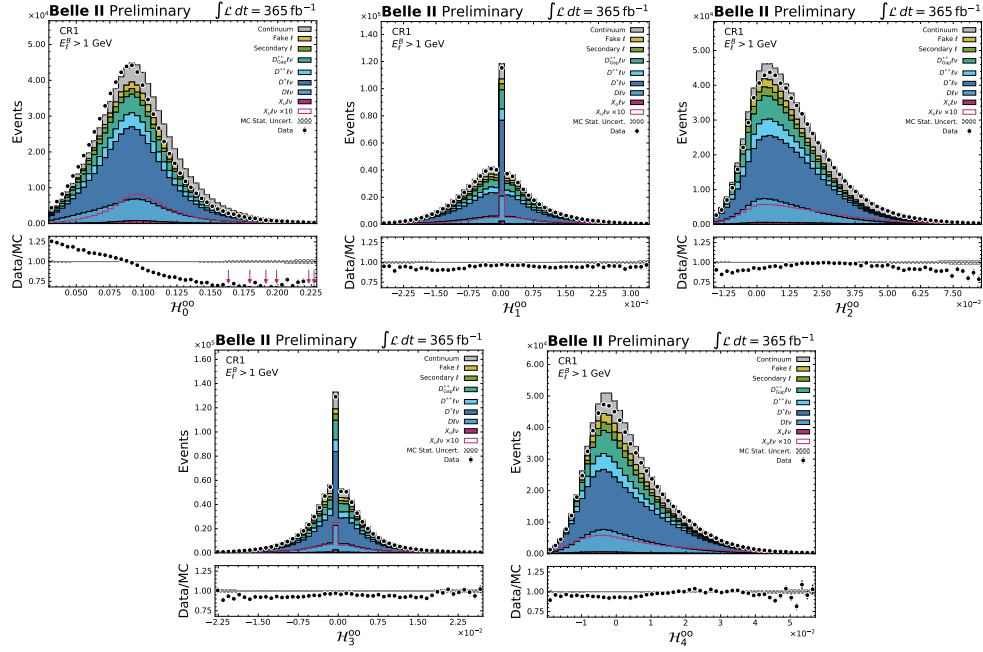


FIGURE B.2: The five H^{oo} variables in CR1

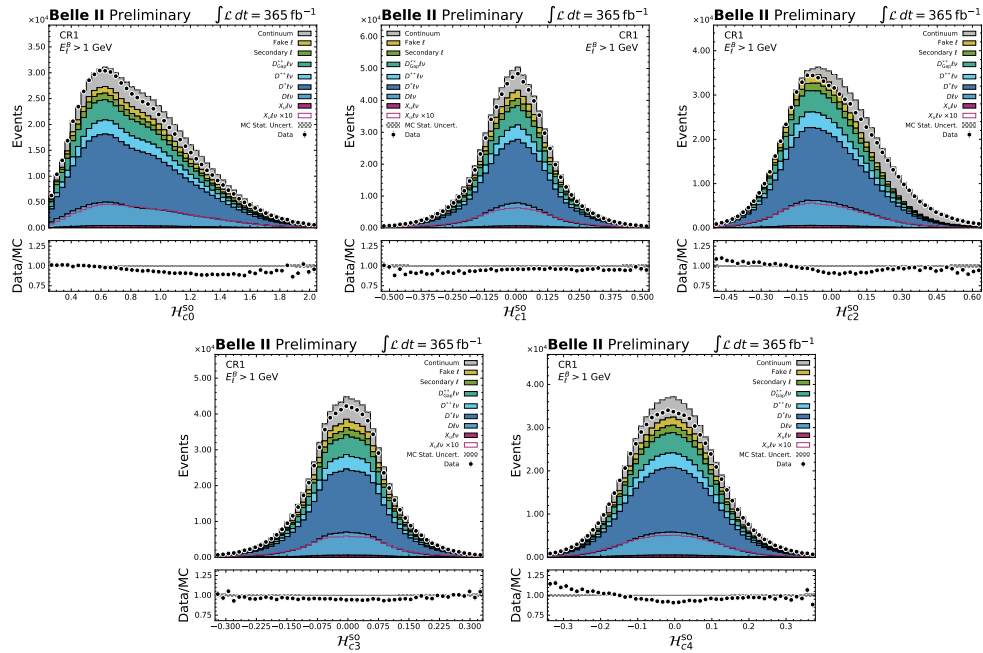
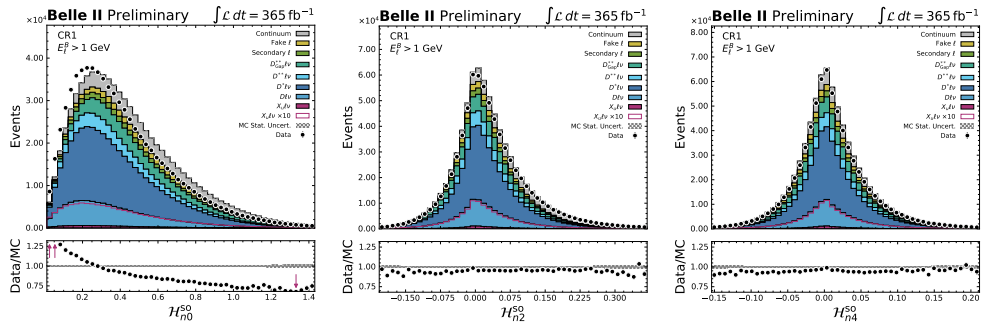
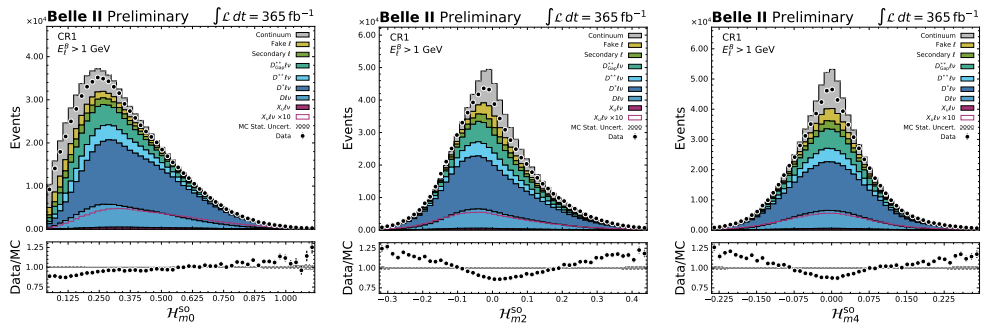
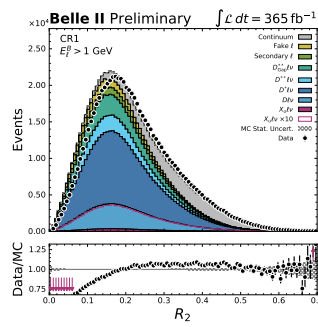


FIGURE B.3: The five H_c^{so} variables in CR1

FIGURE B.4: The three $\mathcal{H}_n^{so}$ variables in CR1FIGURE B.5: The three $\mathcal{H}_m^{so}$ variables in CR1FIGURE B.6: The Fox-Wolfram ratio $R_2 = H_2/H_0$ in CR1

B.3 CLEO cones

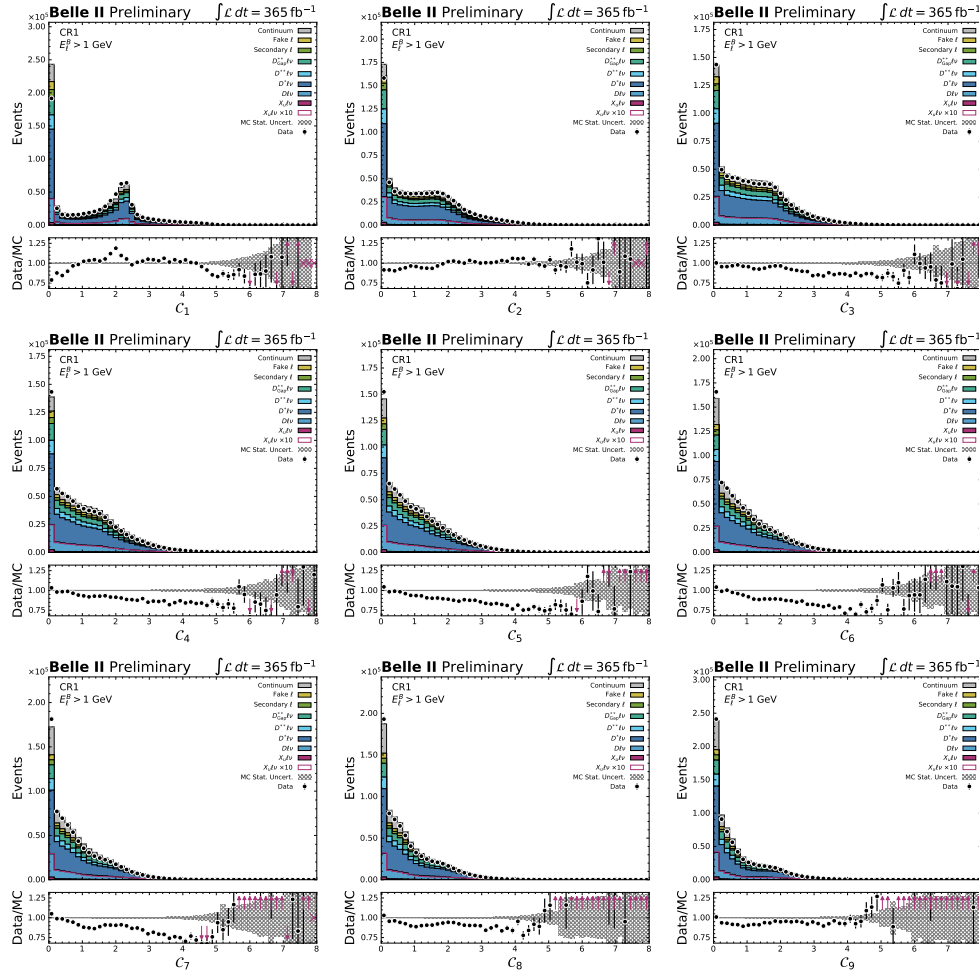


FIGURE B.7: The nine CLEO cone variables in CR1

B.4 Other variables

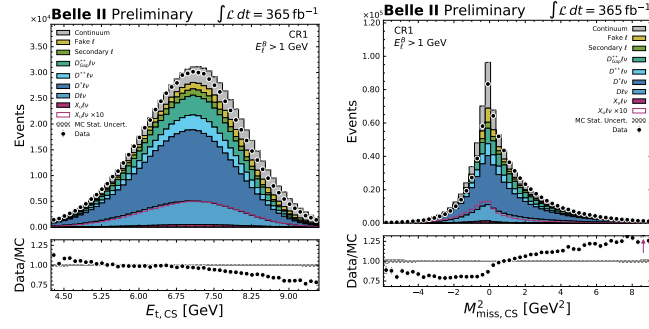


FIGURE B.8: The transverse energy and the missing mass squared in CR1

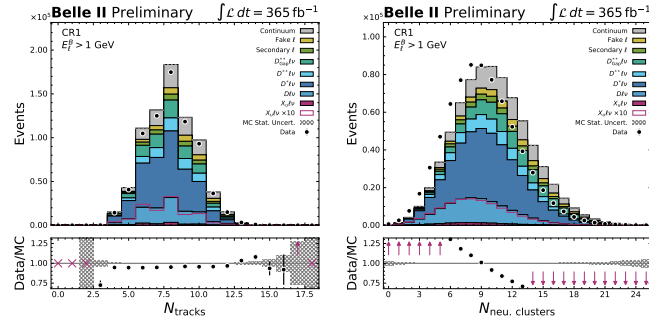


FIGURE B.9: The total track and cluster multiplicities in CR1

Appendix C

Input features of the $B \rightarrow X_c \ell \nu$ suppression classifier

The nine input features used to train the $B \rightarrow X_c \ell \nu$ suppression classifier described in Section 5.5.2 are displayed here: in Figure C.1, the shapes of $B \rightarrow X_c \ell \nu$ and $B \rightarrow X_u \ell \nu$ distributions are compared and in Figure C.2, the simulated and experimental distributions are compared. The continuum suppression and kaon veto are applied on all distributions as the $B \rightarrow X_c \ell \nu$ classifier is optimised for the signal region selections. We do not show the default values chosen for $p\text{-val}(\text{ROE})$ when the vertex fit fails and for the six $D^* \rightarrow D\pi$ variables when no low momentum π is found in the event. These values are -1 , -20 , -10 and 1 for $p\text{-val}(\text{ROE})$, M_{miss}^2 , $\cos \theta_{BY}$ and $\cos \theta_c$, respectively. Only the simulated data statistical uncertainty is shown here in Figure C.2.

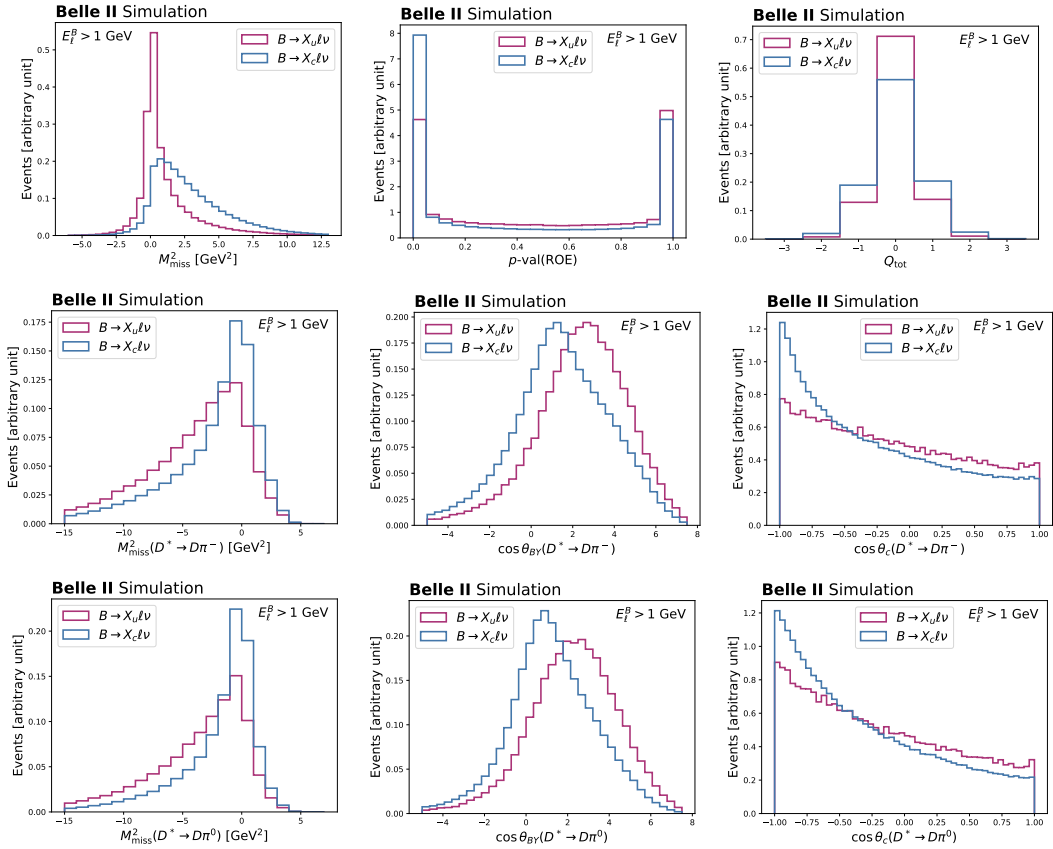


FIGURE C.1: Shape comparison of the nine $B \rightarrow X_c \ell \nu$ suppression classifier input features between $B \rightarrow X_c \ell \nu$ and $B \rightarrow X_u \ell \nu$ events. The histograms are normalised to an arbitrary scale.

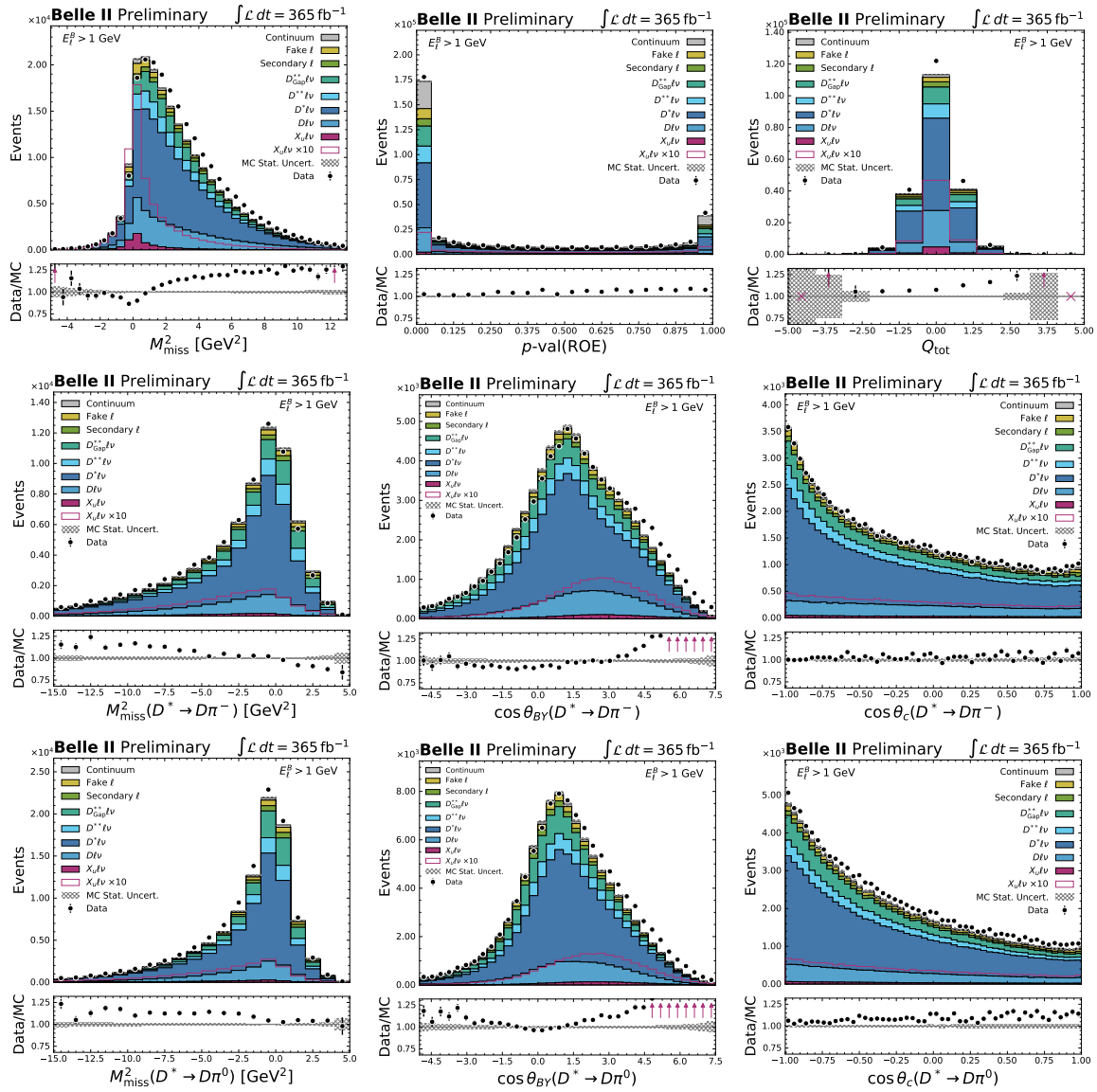


FIGURE C.2: Comparison of the nine $B \rightarrow X_c \ell \nu$ suppression classifier input features between simulated and experimental data distributions.

Appendix D

Comparison between data and simulated data in control regions

We show below the distributions of E_ℓ^B , M_X and q^2 in $\text{CR}_{K,\text{low}}$ and $\text{CR}_{K,\text{high}}$ with a kinematic selection $E_\ell^B > 1$ GeV applied. Refer to Section 5.8 and 5.10 for similar figures in $\text{CR}_{0,\text{low}}$ and the signal region.

D.1 $\text{CR}_{K,\text{low}}$

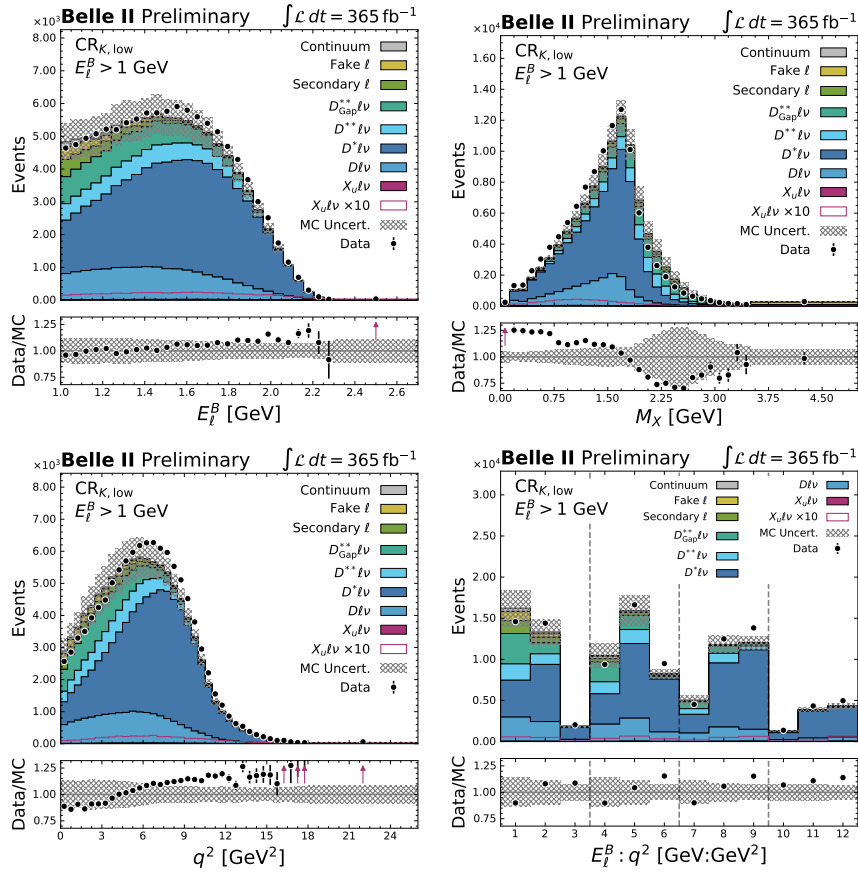


FIGURE D.1: E_ℓ^B (top left), M_X (top right), q^2 (bottom left) and 2D $E_\ell^B:q^2$ (bottom right) distributions in data and simulated data in the $\text{CR}_{K,\text{low}}$ region defined in the main text. The $E_\ell^B > 1$ GeV kinematic selection is applied.

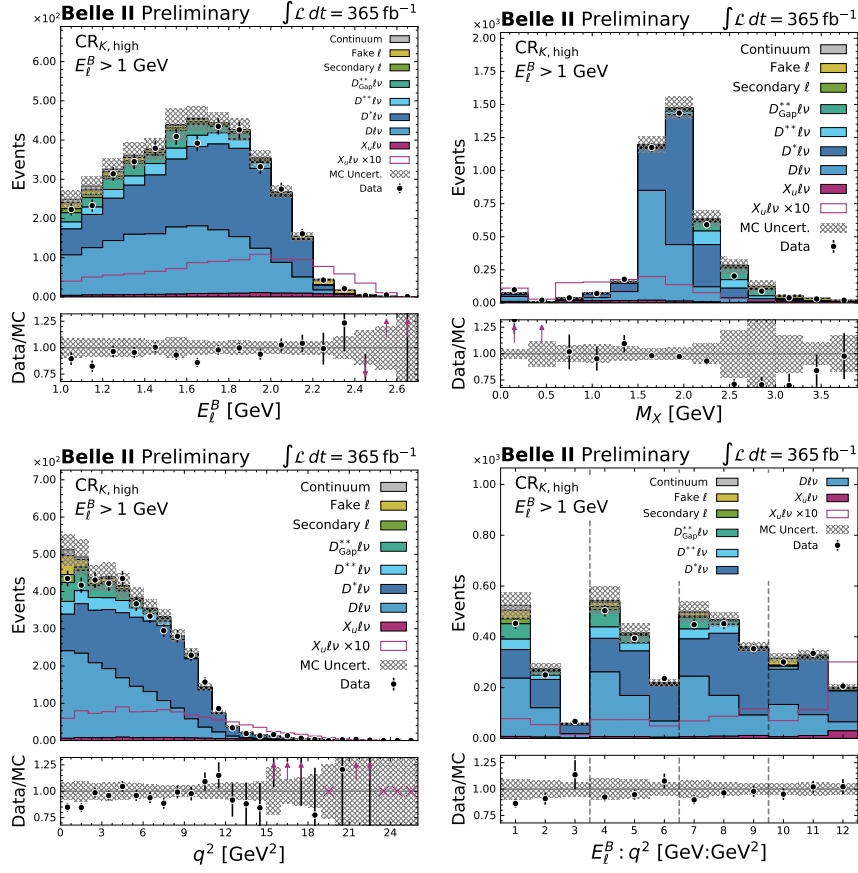
D.2 $\text{CR}_{K,\text{high}}$ 

FIGURE D.2: E_ℓ^B (top left), M_X (top right), q^2 (bottom left) and 2D $E_\ell^B:q^2$ (bottom right) distributions in data and simulated data in the $\text{CR}_{K,\text{high}}$ region defined in the main text. The $E_\ell^B > 1 \text{ GeV}$ kinematic selection is applied.

Appendix E

$D \rightarrow K_L^0 X$ modelling

As mentioned in Section 5.8, K_L^0 mesons are poorly detected by the KLM. As they are produced in charm decays, they can therefore lead to mismodelling of the $B \rightarrow X_c \ell \nu$ component. We study the impact of potential mismodelling of $D \rightarrow K_L^0 X$ decays by upscaling their contribution by 30% as was proposed in the Belle II study of $B^+ \rightarrow K^+ \nu \nu$ decays [163]. As illustrated in Figure E.1, this upscaling has little effect on the agreement between data and simulated data. However, since the $D \rightarrow K_L^0 X$ rate is different in $\text{CR}_{0,\text{low}}$ and the signal region, any large mismodelling could bias the signal extraction (see Section 5.9).

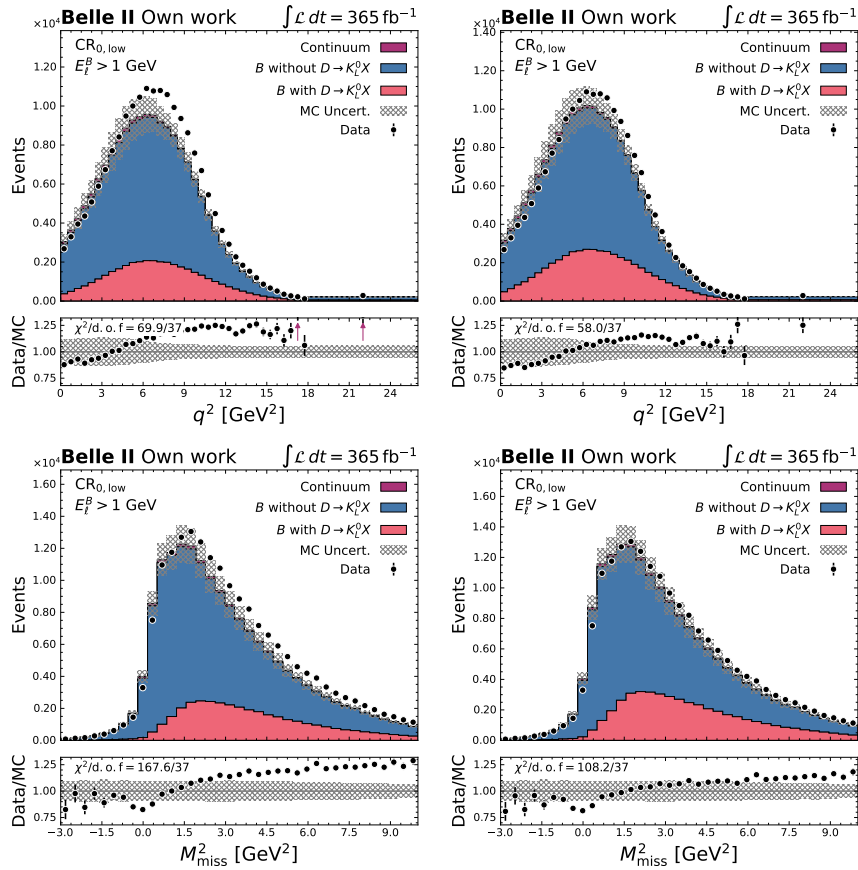


FIGURE E.1: The q^2 (top) and M_{miss}^2 (bottom) distributions in $\text{CR}_{0,\text{low}}$ before (left) and after (right) the upscaling of the $D \rightarrow K_L^0 X$ component rate (pink).

By comparing the signal region distributions in Figure E.2 to the $\text{CR}_{0,\text{low}}$ distributions in Figure E.1, it can be noticed that $D \rightarrow K_L^0 X$ events seem to be well rejected by the $B \rightarrow X_c \ell \nu$ classifier. However, since we can only reject K^+ and K_S^0 , the K veto used to define the signal

region doesn't reject events where a K_L^0 is found. Being able to reject K_L^0 could therefore help reduce the $X_c \ell \nu$ background even further. However, the relative rate of $D \rightarrow K_L^0 X$ events is about three times higher in $\text{CR}_{0,\text{low}}$ than in the signal region. Moreover, the rate of $D \rightarrow K_L^0 X$ events in the signal region is about 20% of that of signal events and the $D \rightarrow K_L^0 X$ modelling could therefore impact the $B \rightarrow X_c \ell \nu$ correction and $B \rightarrow X_u \ell \nu$ branching fraction measurement. As discussed in Section 5.9.5, the differences between $\text{CR}_{0,\text{low}}$ and the signal region slightly bias the signal extraction. This bias is corrected and two sources of uncertainties related to the correction procedure are taken into account. These uncertainties therefore cover differences between $\text{CR}_{0,\text{low}}$ and the signal region such as $D \rightarrow K_L^0 X$ rates. To complete the discussion about $D \rightarrow K_L^0 X$ events, we repeat the nominal signal injected fit described in Section 5.9.5 (signal injected normalisation and $B \rightarrow X_c \ell \nu$ overestimation factor set to 1) by scaling up the $D \rightarrow K_L^0 X$ rate by 30%. The scaling shifts the measured $B \rightarrow X_u \ell \nu$ branching fraction by about 0.2% indicating that any discrepancy caused by the mismodelling of $D \rightarrow K_L^0 X$ decays is well covered by the uncertainties already assigned to the measured signal branching fraction.

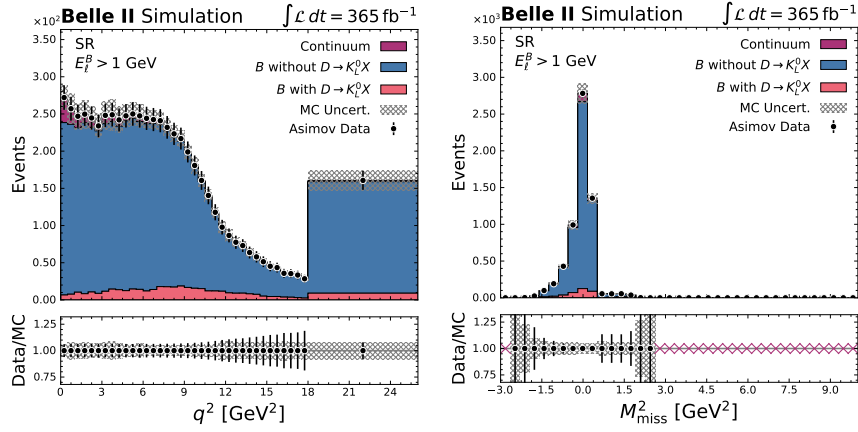


FIGURE E.2: The q^2 (left) and M_{miss}^2 (right) distributions in the signal region.

Appendix F

Fits 2 and 3 figures

We show here all figures which are relevant to illustrate fits 2 and 3. They are the equivalent to all figures shown only for fit 1 in Sections 5.9 to 5.11.

F.1 Fit 2

The fit 2 signal and control regions are defined by applying the following kinematic selections: $E_\ell^B > 1$ GeV, $M_X < 1.7$ GeV; and by fitting the $E_\ell^B:q^2$ distribution.

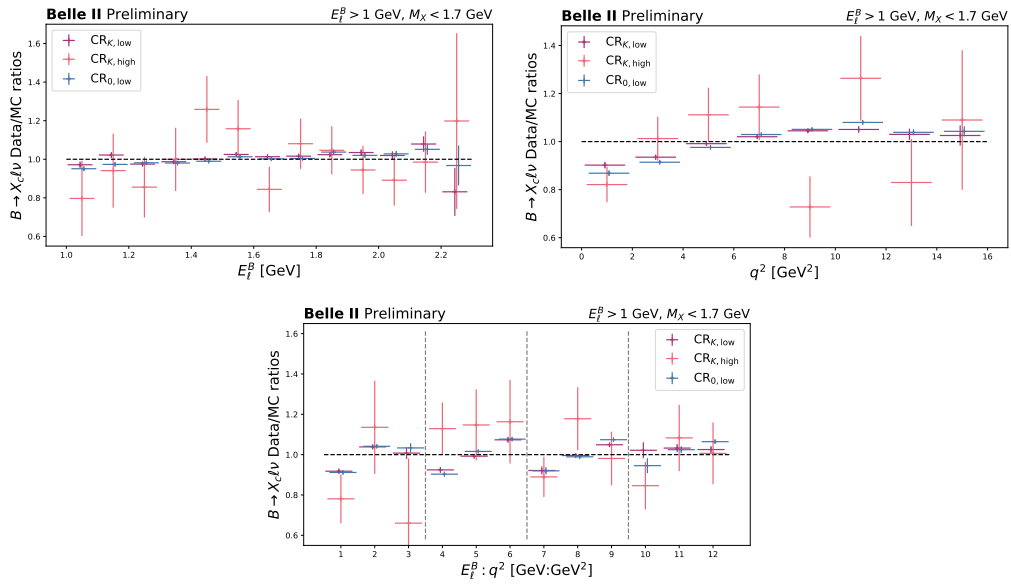


FIGURE F.1: The normalised $B \rightarrow X_c \ell \nu$ data over simulated data yield ratios are shown for the three control regions $CR_{K,low}$ (purple), $CR_{K,high}$ (pink) and $CR_{0,low}$ (blue) in E_ℓ^B (top left), q^2 (top right) and 2D $E_\ell^B:q^2$ (bottom). The $E_\ell^B:q^2$ numbering is given in Figure 5.26.

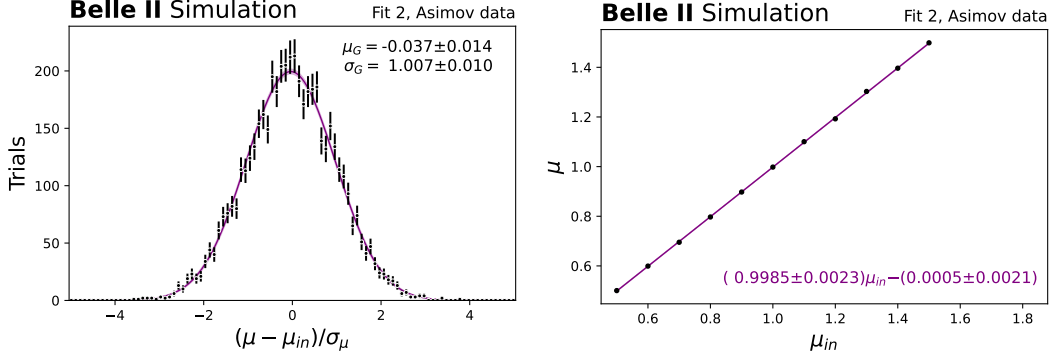


FIGURE F.2: Left: pull distribution of the signal normalisation extracted by fit 2 from 5,000 toy datasets. Right: linearity test for fit 2. These tests are performed using Asimov data.

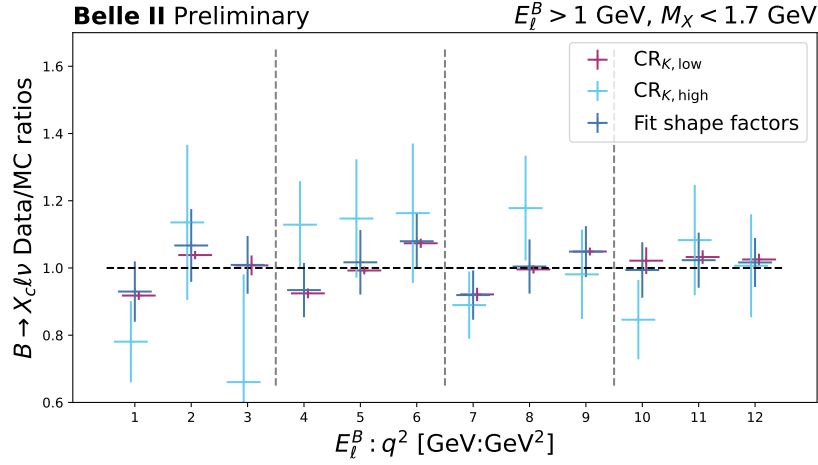


FIGURE F.3: The data over MC yield ratios in $\text{CR}_{K,\text{low}}$ (purple) and $\text{CR}_{K,\text{high}}$ (cyan) in bins of $E_\ell^B : q^2$ with the fit 2 kinematic selections applied are shown. These ratios are compared to the $B \rightarrow X_c \ell \nu$ shape factors extracted from the $\text{CR}_{K,\text{low}}\text{-}\text{CR}_{K,\text{high}}$ fit performed with the same kinematic selections (blue).

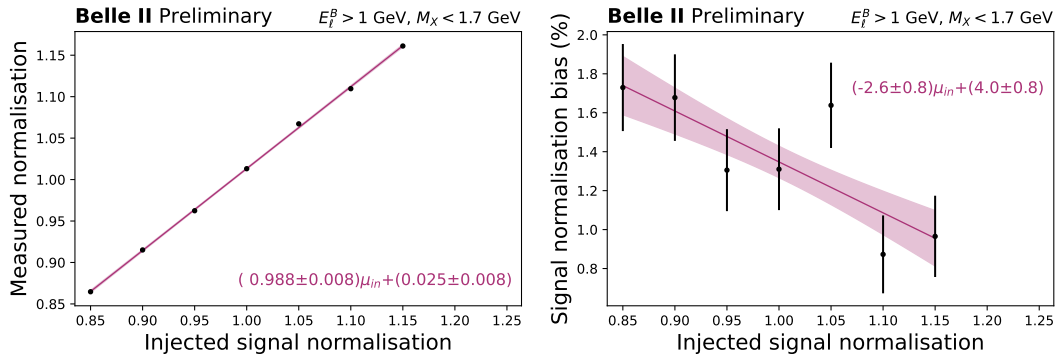


FIGURE F.4: Comparison between the normalisation of the $B \rightarrow X_u \ell \nu$ component injected in data in the fit 2 validation setup and the measured signal normalisation (left). The observed bias is shown explicitly on the right. The purple line corresponds to a linear curve fitted through the points and its parametrisation is shown in purple on the plots.

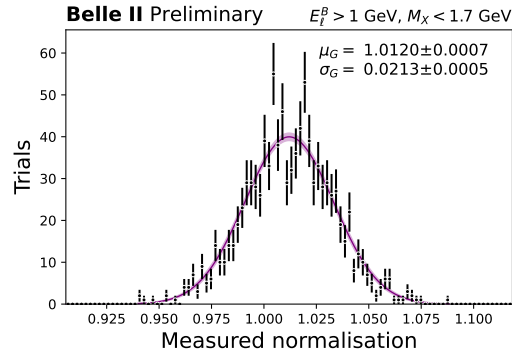


FIGURE F.5: Distribution of the output normalisation obtained from fitting 1,000 toys in the validation setup described in the text. The normalisation of the signal injected in data is fixed to 1 and the fit 2 kinematic selections are applied.

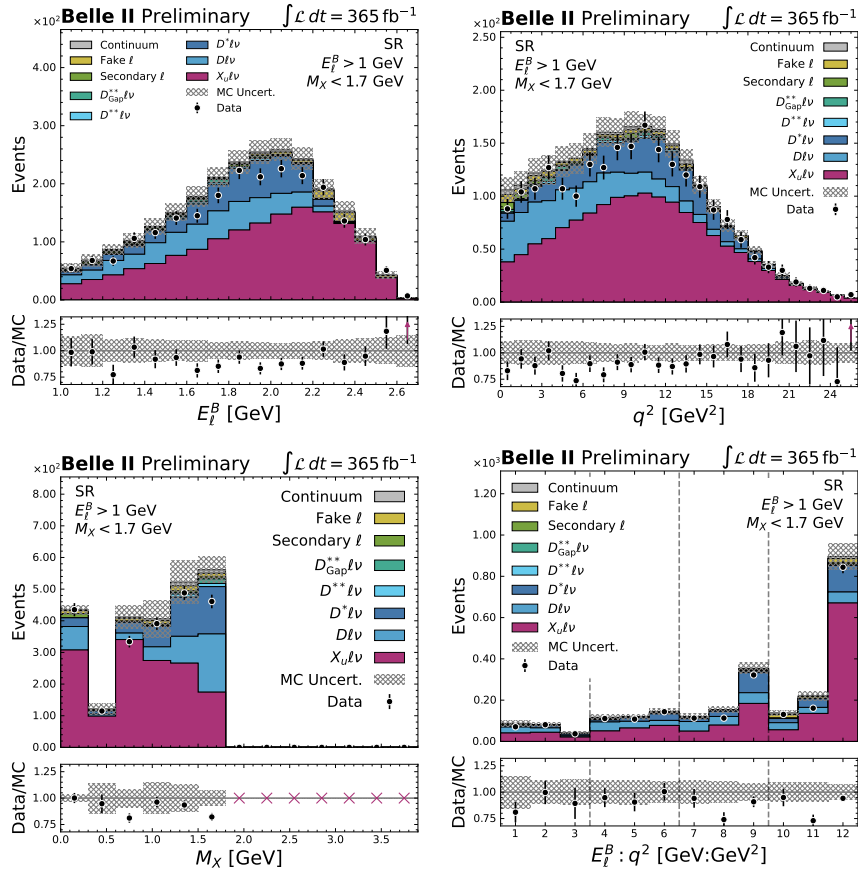


FIGURE F.6: Distribution of E_ℓ^B (top left), q^2 (top right), M_X (bottom left) and $E_\ell^B : q^2$ (bottom right) in the signal region with the fit 2 kinematic selections. The $B \rightarrow X_c \ell \nu$ component has been scaled by the normalisation correction factor extracted via the ABCD method.

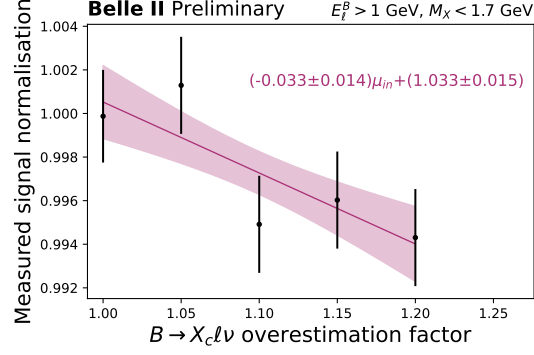


FIGURE F.7: Measured signal normalisation as a function of the $B \rightarrow X_c \ell \nu$ overestimation factor in the signal injected setup with the fit 2 kinematic selections. The normalisation of the injected signal is set to 1 and the bias described in Section 5.9.5 is corrected for.

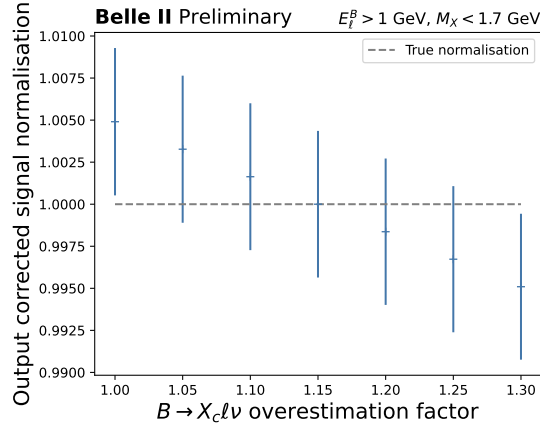


FIGURE F.8: Corrected output signal normalisation as a function of various $B \rightarrow X_c \ell \nu$ normalisation overestimation factors in the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit 2 setup with injected signal. The output normalisation is already corrected for the source of bias described in Section 5.9.5. It is then scaled by a factor of 1.004 to correct for the source of bias described in Section 5.10.1 and the difference between the scaled and non-scaled value is added as an uncertainty. The normalisation of the injected signal is shown as a grey dotted line.

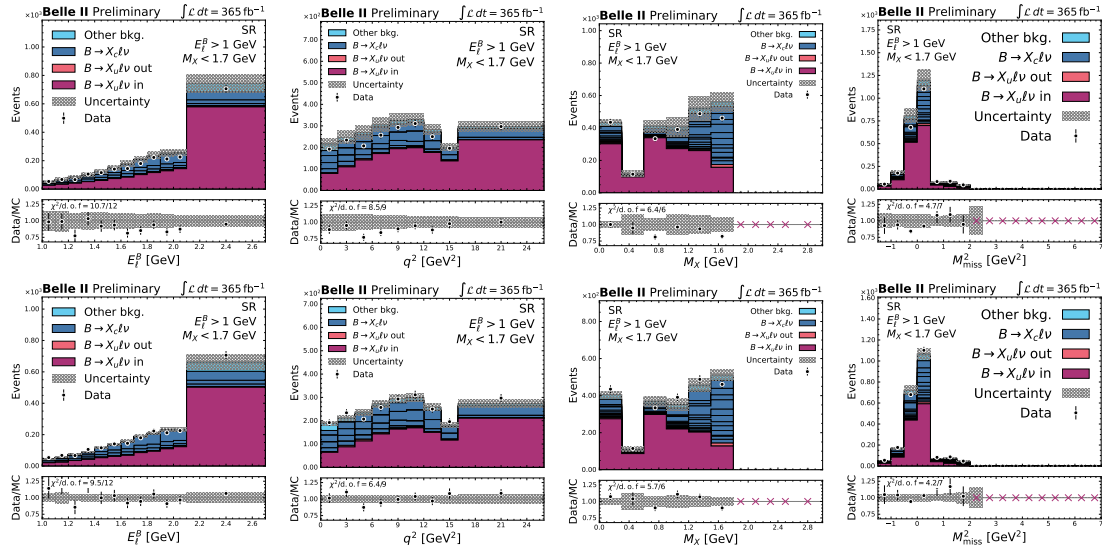


FIGURE F.10: Top: distributions of E_ℓ^B (left), q^2 (left centre), M_X (right centre) and M_{miss}^2 (right) in the signal region with $E_\ell^B > 1 \text{ GeV}$ and $M_X < 1.7 \text{ GeV}$. Bottom: projections of fit 2 results onto the E_ℓ^B , q^2 , M_X and M_{miss}^2 distributions. The $B \rightarrow X_c \ell \nu$ template is split in 12 templates, one for each bin of $E_\ell^B : q^2$.

F.2 Fit 3

The fit 3 signal and control regions are defined by applying the following kinematic selections: $E_\ell^B > 1$ GeV, $M_X < 1.7$ GeV, $q^2 > 8$ GeV²; and by fitting the E_ℓ^B distribution.

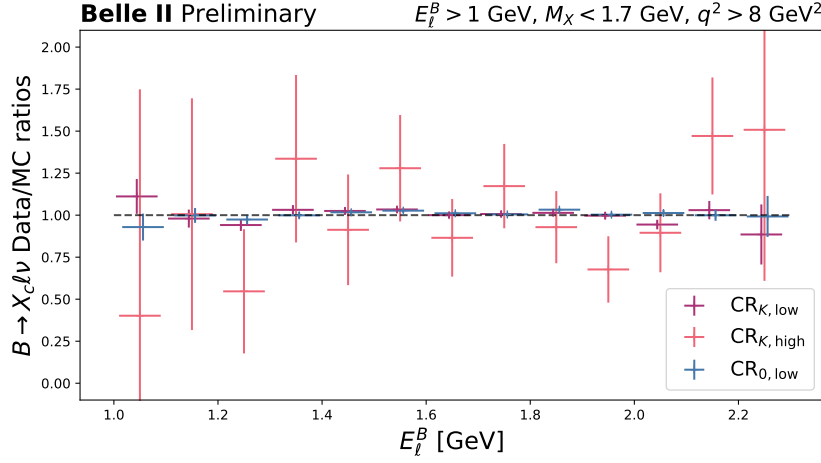


FIGURE F.11: The normalised $B \rightarrow X_c \ell \nu$ data over simulated data yield ratios are shown for the three control regions $\text{CR}_{K,\text{low}}$ (purple), $\text{CR}_{K,\text{high}}$ (pink) and $\text{CR}_{0,\text{low}}$ (blue) in E_ℓ^B .

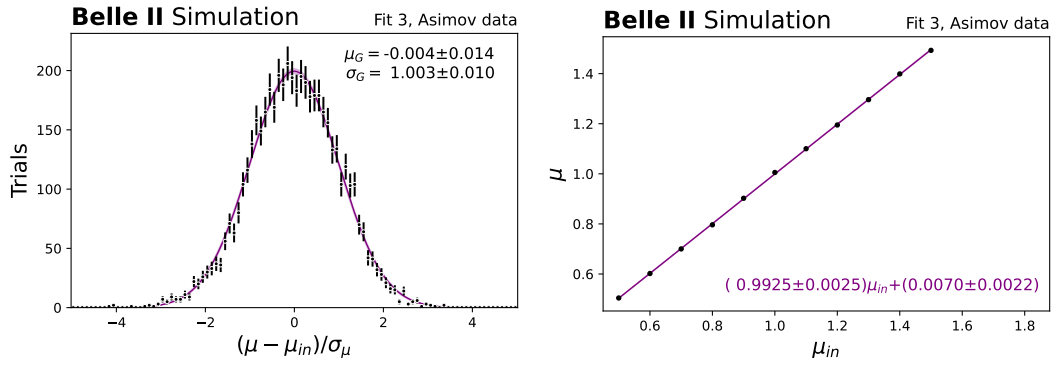


FIGURE F.12: Left: pull distribution of the signal normalisation extracted by fit 3 from 5,000 toy datasets. Right: linearity test for fit 3. These tests are performed using Asimov data.

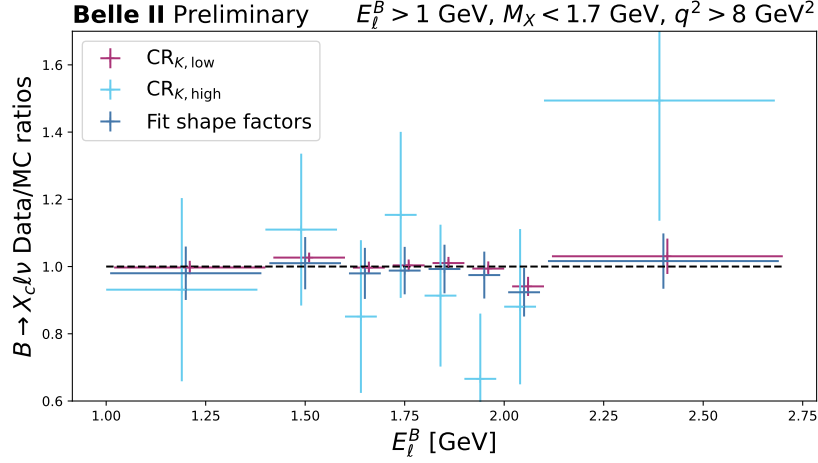


FIGURE F.13: The data over MC yield ratios in $\text{CR}_{K,\text{low}}$ (purple) and $\text{CR}_{K,\text{high}}$ (cyan) in bins of E_ℓ^B with the fit 3 kinematic selection applied are shown. These ratios are compared to the $B \rightarrow X_c \ell \nu$ shape factors extracted from the $\text{CR}_{K,\text{low}} - \text{CR}_{K,\text{high}}$ fit performed with the same kinematic selections (blue).

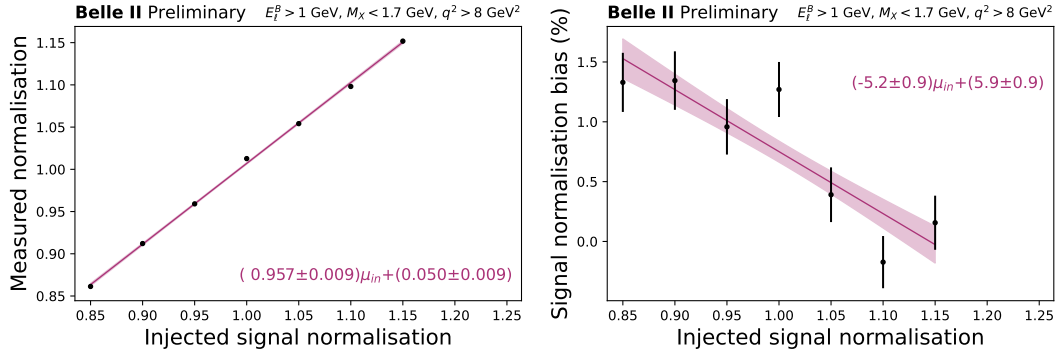


FIGURE F.14: Comparison between the normalisation of the $B \rightarrow X_u \ell \nu$ component injected in data in the fit 3 validation setup and the measured signal normalisation (left). The observed bias is shown explicitly on the right. The purple line corresponds to a linear curve fitted through the points and its parametrisation is shown in purple on the plots.

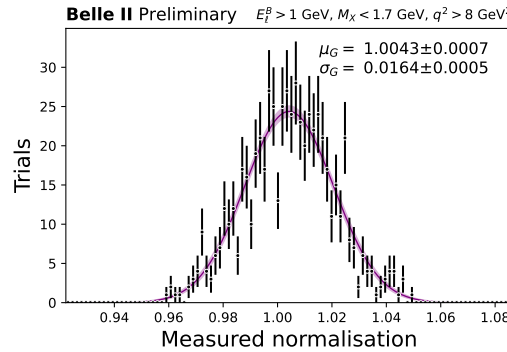


FIGURE F.15: Distribution of the output normalisation obtained from fitting 1,000 toys in the validation setup described in the text. The normalisation of the signal injected in data is fixed to 1 and the fit 3 kinematic selections are applied.

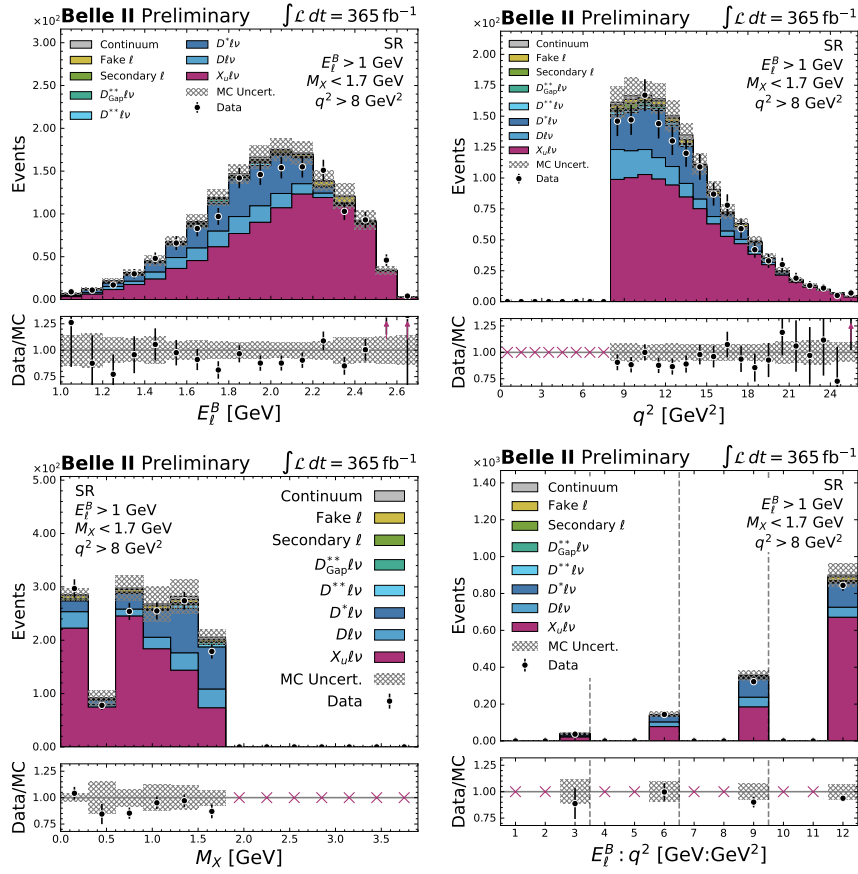


FIGURE F.16: Distribution of E_ℓ^B (top left), q^2 (top right), M_X (bottom left) and $E_\ell^B : q^2$ (bottom right) in the signal region with the fit 3 kinematic selections. The $B \rightarrow X_c \ell \nu$ component has been scaled by the normalisation correction factor extracted via the ABCD method.

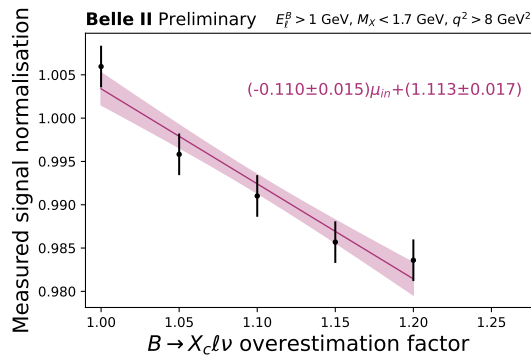


FIGURE F.17: Measured signal normalisation as a function of the $B \rightarrow X_c \ell \nu$ overestimation factor in the signal injected setup with the fit 3 kinematic selections. The normalisation of the injected signal is set to 1 and the bias described in Section 5.9.5 is corrected for.

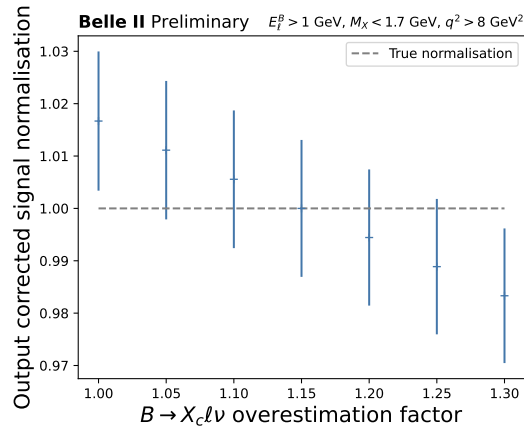


FIGURE F.18: Corrected output signal normalisation as a function of various $B \rightarrow X_c \ell \nu$ normalisation overestimation factors in the $\text{CR}_{K,\text{low}}\text{--}\text{CR}_{K,\text{high}}$ fit 3 setup with injected signal. The output normalisation is already corrected for the source of bias described in Section 5.9.5. It is then scaled by a factor of 1.013 to correct for the source of bias described in Section 5.10.1 and the difference between the scaled and non-scaled value is added as an uncertainty. The normalisation of the injected signal is shown as a grey dotted line.

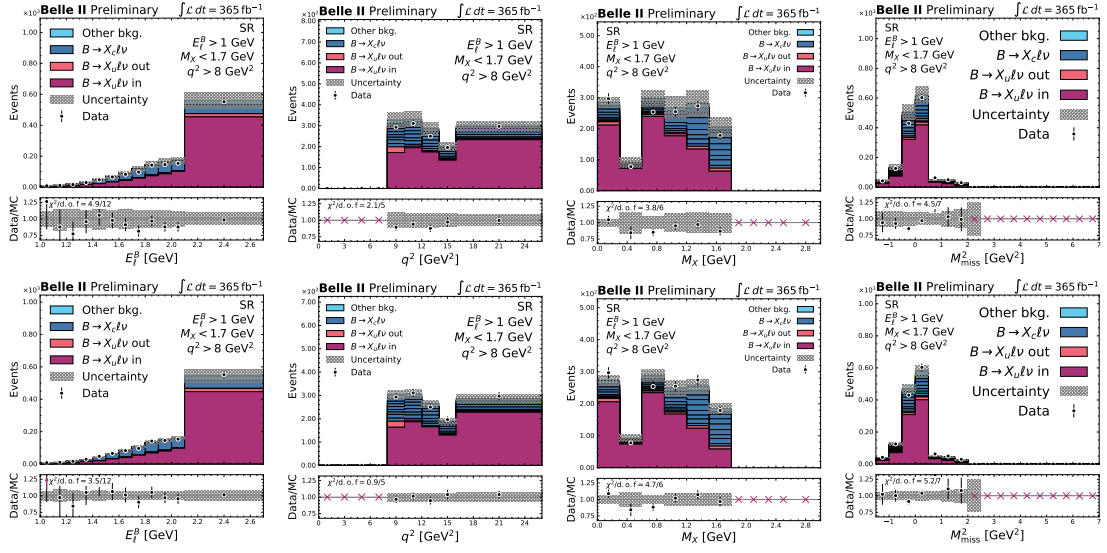


FIGURE F.20: Top: distributions of E_ℓ^B (left), q^2 (left centre), M_X (right centre) and M_{miss}^2 (right) in the signal region with $E_\ell^B > 1$ GeV, $M_X < 1.7$ GeV and $q^2 > 8$ GeV². Bottom: projections of fit 3 results onto the E_ℓ^B , q^2 , M_X and M_{miss}^2 distributions. The $B \rightarrow X_c \ell \nu$ template is split in eight templates, one for each bin of E_ℓ^B .

Bibliography

- [1] J. J. Thomson, *Xl. Cathode Rays*, [The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science](#) **44**, 293–316 (1897) (cit. on p. 17).
- [2] H. Svidras, *Measurement of the photon energy spectrum in inclusive radiative B meson decays using the hadronic-tagging method*, PhD thesis (U. Hamburg (main), DESY, Hamburg U., Hamburg, 2023) (cit. on p. 18).
- [3] J. Zupan, *Introduction to flavour physics*, [arXiv:1903.05062](#) (2019) (cit. on p. 18).
- [4] M. Thomson, *Modern Particle Physics* (Cambridge University Press, New York, Oct. 2013) (cit. on pp. 18–20, 59).
- [5] M. D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, 2014) (cit. on pp. 18, 19).
- [6] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995) (cit. on p. 18).
- [7] N. Hüsken, E. S. Norella, and I. Polyakov, *A brief guide to exotic hadrons*, [arXiv:2410.06923](#) (2025) (cit. on p. 20).
- [8] T. D. Lee and C.-N. Yang, *Question of Parity Conservation in Weak Interactions*, [Phys. Rev.](#) **104**, 254–258 (1956) (cit. on p. 21).
- [9] S. L. Glashow, *Partial-symmetries of weak interactions*, [Nuclear Physics](#) **22**, 579–588 (1961) (cit. on p. 21).
- [10] A. Salam, *Weak and Electromagnetic Interactions*, [Conf. Proc. C](#) **680519**, 367–377 (1968) (cit. on p. 21).
- [11] S. Weinberg, *A model of leptons*, [Phys. Rev. Lett.](#) **19**, 1264–1266 (1967) (cit. on p. 21).
- [12] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, [Phys. Rev. Lett.](#) **13**, 321–323 (1964) (cit. on p. 23).
- [13] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, [Phys. Rev. Lett.](#) **13**, 508–509 (1964) (cit. on p. 23).
- [14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, [Phys. Rev. Lett.](#) **13**, 585–587 (1964) (cit. on p. 23).
- [15] M. Kobayashi and T. Maskawa, *CP -Violation in the Renormalizable Theory of Weak Interaction*, [Progress of Theoretical Physics](#) **49**, 652–657 (1973) (cit. on p. 25).
- [16] N. Cabibbo, *Unitary symmetry and leptonic decays*, [Phys. Rev. Lett.](#) **10**, 531–533 (1963) (cit. on p. 25).
- [17] J. Charles et al., *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, [Eur. Phys. J. C](#) **41**, updated results and plots available at: <http://ckmfitter.in2p3.fr>, 1–131 (2005) (cit. on pp. 27, 28, 48, 49, 152, 153, 155–157).
- [18] M. Bona et al., *New $utfit$ analysis of the unitarity triangle in the cabibbo–kobayashi–maskawa scheme*, [Rendiconti Lincei. Scienze Fisiche e Naturali](#) **34**, updated results and plots available at: <http://www.utfit.org>, 37–57 (2023) (cit. on p. 27).

- [19] J. D. Richman and P. R. Burchat, *Leptonic and semileptonic decays of charm and bottom hadrons*, [Rev. Mod. Phys.](#) **67**, 893–976 (1995) (cit. on p. 29).
- [20] J. Dingfelder and T. Mannel, *Leptonic and semileptonic decays of B mesons*, [Rev. Mod. Phys.](#) **88**, 035008 (2016) (cit. on pp. 29, 33).
- [21] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, [Phys. Rev. D](#) **2**, 1285–1292 (1970) (cit. on p. 29).
- [22] G. 't Hooft, *Dimensional regularization and the renormalization group*, [Nuclear Physics B](#) **61**, 455–468 (1973) (cit. on p. 31).
- [23] S. Weinberg, *New approach to the renormalization group*, [Phys. Rev. D](#) **8**, 3497–3509 (1973) (cit. on p. 31).
- [24] S. Navas et al., *Review of particle physics*, [Phys. Rev. D](#) **110**, 030001 (2024) (cit. on pp. 31, 47, 56, 80, 83, 86, 87, 90, 147, 156, 158).
- [25] G. Buchalla, *Heavy quark theory*, [arXiv:hep-ph/0202092](#) (2002) (cit. on p. 31).
- [26] N. Isgur and M. B. Wise, *Weak decays of heavy mesons in the static quark approximation*, [Physics Letters B](#) **232**, 113–117 (1989) (cit. on p. 32).
- [27] T. Mannel, *Operator product expansion for inclusive semileptonic decays in heavy quark effective field theory*, [Nuclear Physics B](#) **413**, 396–410 (1994) (cit. on pp. 32, 157).
- [28] J. Chay, H. Georgi, and B. Grinstein, *Lepton energy distributions in heavy meson decays from qcd*, [Physics Letters B](#) **247**, 399–405 (1990) (cit. on pp. 32, 157).
- [29] R. Gupta, *Introduction to Lattice QCD*, [arXiv:hep-lat/9807028](#) (1998) (cit. on p. 32).
- [30] C. Glenn Boyd, B. Grinstein, and R. F. Lebed, *Model-independent determinations of $B \rightarrow D\ell\nu$, $B \rightarrow D^*\ell\nu$ form factors*, [Nuclear Physics B](#) **461**, 493–511 (1996) (cit. on pp. 33, 34, 91).
- [31] I. Caprini, L. Lellouch, and M. Neubert, *Dispersive bounds on the shape of $B \rightarrow D^{(*)}\ell\nu$ form factors*, [Nuclear Physics B](#) **530**, 153–181 (1998) (cit. on p. 34).
- [32] R. Glattauer, C. Schwanda, A. Abdesselam, et al., *Measurement of the decay $B \rightarrow D\ell\nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$* , [Phys. Rev. D](#) **93**, 032006 (2016) (cit. on p. 34).
- [33] I. Adachi et al., *Determination of $|V_{cb}|$ using $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays with Belle II*, [Phys. Rev. D](#) **108**, 092013 (2023) (cit. on p. 34).
- [34] P. Horak, *Determination of $|V_{cb}|$ using $B \rightarrow D\ell\nu$ decays at Belle II*, PhD thesis (Vienna, Tech. U., 2024) (cit. on p. 34).
- [35] N. Isgur and M. B. Wise, *Weak transition form factors between heavy mesons*, [Physics Letters B](#) **237**, 527–530 (1990) (cit. on p. 34).
- [36] F. U. Bernlochner et al., *Constrained second-order power corrections in HQET: $R(D^{(*)})$, $|V_{cb}|$, and new physics*, [Phys. Rev. D](#) **106**, 096015 (2022) (cit. on pp. 34, 81, 91).
- [37] F. U. Bernlochner, Z. Ligeti, and D. J. Robinson, *Model independent analysis of semileptonic B decays to D^{**} for arbitrary new physics*, [Phys. Rev. D](#) **97**, 075011 (2018) (cit. on pp. 35, 81, 91).
- [38] C. Bourrely, L. Lellouch, and I. Caprini, *Model-independent description of $B \rightarrow \pi\ell\nu$ decays and a determination of $|V_{ub}|$* , [Phys. Rev. D](#) **79**, 013008 (2009) (cit. on pp. 35, 91).
- [39] Y. Aoki, T. Blum, G. Colangelo, et al., *FLAG Review 2021*, [Eur. Phys. J. C](#) **82**, 869 (2022) (cit. on pp. 35, 80, 91).

- [40] F. U. Bernlochner, M. T. Prim, and D. J. Robinson, $B \rightarrow \rho \ell \bar{\nu}$ and $\omega \ell \bar{\nu}$ in and beyond the Standard Model: Improved predictions and $|V_{ub}|$, *Phys. Rev. D* **104**, 034032 (2021) (cit. on pp. 36, 47, 80, 91).
- [41] A. Bharucha, D. M. Straub, and R. Zwicky, $B \rightarrow V \ell^+ \ell^-$ in the Standard Model from light-cone sum rules, *JHEP* **2016**, 98 (2016) (cit. on p. 36).
- [42] G. Duplancic and B. Melic, Form factors of $B, B_s \rightarrow \eta^{(\prime)}$ and $D, D_s \rightarrow \eta^{(\prime)}$ transitions from QCD light-cone sum rules, *JHEP* **11**, 138 (2015) (cit. on pp. 37, 80, 91).
- [43] P. Colangelo and A. Khodjamirian, “QCD Sum Rules, a Modern Perspective”, in *At the frontier of particle physics* (World Scientific, 2001), pp. 1495–1576 (cit. on p. 37).
- [44] M. Shifman, *The quark hadron duality*, eConf **C030614**, edited by P. Colangelo et al., 001 (2003) (cit. on p. 37).
- [45] K. Olschewsky, *Pushing the precision in B physics*, PhD thesis (Siegen U., June 2022) (cit. on p. 37).
- [46] T. Mannel, I. S. Milutin, and K. K. Vos, Inclusive semileptonic $b \rightarrow c \ell \bar{\nu}$ decays to order $1/m_b^5$, *Journal of High Energy Physics* **2024**, 226 (2024) (cit. on p. 37).
- [47] T. Mannel and K. K. Vos, Reparametrization invariance and partial re-summations of the heavy quark expansion, *Journal of High Energy Physics* **2018**, 115 (2018) (cit. on p. 37).
- [48] R. van Tonder et al., Measurements of q^2 moments of inclusive $B \rightarrow X_c \ell^+ \nu_\ell$ decays with hadronic tagging, *Phys. Rev. D* **104**, 112011 (2021) (cit. on p. 37).
- [49] F. Abudinén et al., Measurement of Lepton Mass Squared Moments in $B \rightarrow X_c \ell \bar{\nu}_\ell$ Decays with the Belle II experiment, *arXiv:2205.06372* (2022) (cit. on p. 37).
- [50] F. Bernlochner et al., First extraction of inclusive V_{cb} from q^2 moments, *arXiv:2205.10274* (2022) (cit. on p. 37).
- [51] I. Milutin, *Intrinsic Charm and Higher Order $1/m_b$ corrections in inclusive $B \rightarrow X_c \ell \bar{\nu}$* , *Challenges in Semileptonic B Decays*, Österreichische Akademie der Wissenschaften, Vienna (2024) (cit. on p. 38).
- [52] E. Gardi, *On the determination of $|V_{ub}|$ from inclusive semileptonic B decays*, *arXiv:0806.4524* (2008) (cit. on pp. 38, 155).
- [53] I. Bigi et al., *On the motion of heavy quarks inside hadrons: universal distributions and inclusive decays*, *International Journal of Modern Physics A* **09**, 2467–2504 (1994) (cit. on p. 38).
- [54] B. O. Lange, M. Neubert, and G. Paz, *Theory of charmless inclusive B decays and the extraction of V_{ub}* , *Phys. Rev. D* **72**, 073006 (2005) (cit. on pp. 38, 41, 151).
- [55] F. U. Bernlochner et al., *Status of SIMBA*, *arXiv:1101.3310 [hep-ph]* (2011) (cit. on p. 38).
- [56] F. U. Bernlochner et al., Precision global determination of the $B \rightarrow X_s \gamma$ decay rate, *Phys. Rev. Lett.* **127**, 102001 (2021) (cit. on pp. 38, 157).
- [57] P. Gambino, K. J. Healey, and C. Mondino, *Neural network approach to $B \rightarrow X_u \ell \nu$* , *Phys. Rev. D* **94**, 014031 (2016) (cit. on pp. 38, 157).
- [58] I. Novikov, *Precision theory predictions and interpretation of differential distributions in inclusive decays at Belle II*, PhD thesis (U. Hamburg (main), Hamburg U., 2023) (cit. on pp. 38, 44, 50).
- [59] F. De Fazio and M. Neubert, $B \rightarrow X_u \ell \bar{\nu}_\ell$ decay distributions to order α_s , *JHEP* **06**, 017 (1999) (cit. on pp. 40, 41).
- [60] S. Bosch et al., *Factorization and shape-function effects in inclusive b-meson decays*, *Nuclear Physics B* **699**, 335–386 (2004) (cit. on pp. 41, 43, 151).

- [61] T. Becher, A. Broggio, and A. Ferroglia, *Introduction to Soft-Collinear Effective Theory* (Springer International Publishing, 2015) (cit. on p. 42).
- [62] P. Gambino et al., *Inclusive semileptonic B decays and the determination of $|V_{ub}|$* , [JHEP **10**, 058 \(2007\)](#) (cit. on pp. 42–44, 151, 155).
- [63] S. Banerjee et al., *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2023*, [arXiv:2411.18639 \(2024\)](#) (cit. on pp. 44, 47–49, 80, 82, 124, 126, 147, 148, 151–153, 155–157).
- [64] J. R. Andersen and E. Gardi, *Inclusive spectra in charmless semileptonic B decays by dressed gluon exponentiation*, [JHEP **01**, 097 \(2006\)](#) (cit. on pp. 44, 151).
- [65] U. Aglietti et al., *Inclusive measure of $|V_{ub}|$ with the analytic coupling model*, [Eur. Phys. J. C **59**, 831–840 \(2009\)](#) (cit. on p. 44).
- [66] P. Gambino, *Status and prospects of inclusive $b \rightarrow ul\nu$ theory*, [Sixth Meeting on B Physics, Istituto Universitario di Studi Superiori, Ferrara \(2009\)](#) (cit. on p. 45).
- [67] M. Voloshin, *Nonfactorization effects in heavy mesons and determination of $|V_{ub}|$ from inclusive semileptonic B decays*, [Phys. Lett. B **515**, 74–80 \(2001\)](#) (cit. on p. 45).
- [68] I. I. Bigi and N. Uraltsev, *Weak annihilation and the endpoint spectrum in semileptonic B decays*, [Nucl. Phys. B **423**, 33–55 \(1994\)](#) (cit. on p. 45).
- [69] Z. Ligeti, M. Luke, and A. V. Manohar, *Constraining weak annihilation using semileptonic D decays*, [Phys. Rev. D **82**, 033003 \(2010\)](#) (cit. on p. 45).
- [70] I. Adachi et al., *Determination of $|V_{ub}|$ from simultaneous measurements of untagged $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ and $B^+ \rightarrow \rho^0 \ell^+ \nu_\ell$ decays*, [arXiv:2407.17403 \(2024\)](#) (cit. on pp. 47, 152).
- [71] F. U. Bernlochner and S. Wallner, *Role of ρ – ω interference in semileptonic $B \rightarrow \pi^+ \pi^- \ell \bar{\nu}_\ell$ decays*, [Phys. Rev. D **109**, 074040 \(2024\)](#) (cit. on p. 47).
- [72] R. Aaij et al., *Determination of the quark coupling strength $|v_{ub}|$ using baryonic decays*, [Nature Physics **11**, 743–747 \(2015\)](#) (cit. on pp. 47, 48, 157).
- [73] R. Aaij, C. A. Beteta, T. Ackernley, et al., *First observation of the decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and Measurement of $|V_{ub}|/|V_{cb}|$* , [Phys. Rev. Lett. **126**, 081804 \(2021\)](#) (cit. on pp. 47, 157).
- [74] J. Lees, V. Poireau, V. Tisserand, et al., *Measurement of the inclusive electron spectrum from B meson decays and determination of $|V_{ub}|$* , [Phys. Rev. D **95**, 072001 \(2017\)](#) (cit. on pp. 47, 146–148, 151, 159).
- [75] Y. S. Amhis, S. Banerjee, E. Ben-Haim, et al., *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2021*, [Phys. Rev. D **107**, 052008 \(2023\)](#) (cit. on p. 48).
- [76] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, *New ways to search for right-handed current in $B \rightarrow \rho \ell \bar{\nu}$ decay*, [Phys. Rev. D **90**, 094003 \(2014\)](#) (cit. on p. 48).
- [77] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, [Prog. Theor. Phys. **49**, 652–657 \(1973\)](#) (cit. on p. 53).
- [78] P. Oddone, *Detector Considerations*, [eConf **C870126**, 423–446 \(1987\)](#) (cit. on p. 53).
- [79] W. Altmannshofer et al., *The Belle II Physics Book*, [PTEP **2019**, edited by E. Kou and P. Urquijo, \[Erratum: PTEP 2020, 029201 \(2020\)\], 123C01 \(2019\)](#) (cit. on pp. 53, 60).
- [80] S. J. Brodsky and E. de Rafael, *Suggested boson-lepton pair couplings and the anomalous magnetic moment of the muon*, [Phys. Rev. **168**, 1620–1622 \(1968\)](#) (cit. on p. 54).
- [81] D. M. J. Lovelock et al., *Masses, Widths, and Leptonic Widths of the Higher Upsilon Resonances*, [Phys. Rev. Lett. **54**, 377–380 \(1985\)](#) (cit. on p. 54).

- [82] E. Ganiev, *Measurement of the branching fraction, longitudinal polarization fraction, and charge-parity violating asymmetry in $B^+ \rightarrow \rho^+ \rho^0$ decays at Belle*, PhD thesis (Trieste U., 2021) (cit. on p. 55).
- [83] K. Akai, K. Furukawa, and H. Koiso, *SuperKEKB Collider*, [*Nucl. Instrum. Meth. A* **907**, 188–199 \(2018\)](#) (cit. on pp. 54, 55).
- [84] H. Ye et al., *Commissioning and performance of the Belle II pixel detector*, [*Nucl. Instrum. Meth. A* **987**, 164875 \(2021\)](#) (cit. on p. 57).
- [85] K. Adamczyk et al., *The design, construction, operation and performance of the Belle II silicon vertex detector*, [*Journal of Instrumentation* **17**, P11042 \(2022\)](#) (cit. on p. 57).
- [86] N. Taniguchi, *Central Drift Chamber for Belle-II*, [*Journal of Instrumentation* **12**, C06014 \(2017\)](#) (cit. on p. 58).
- [87] S. Nishida et al., *Aerogel RICH for the Belle II forward PID*, [*Nucl. Instrum. Meth. A* **766**, 28–31 \(2014\)](#) (cit. on p. 59).
- [88] 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* **2020**, 093H01 \(2020\)](#) (cit. on p. 60).
- [89] J. Fast, *The Belle II imaging Time-of-Propagation (iTOP) detector*, [*Nucl. Instrum. Meth. A* **876**, 145–148 \(2017\)](#) (cit. on p. 60).
- [90] K. Kojima, *The operation and performance of the TOP detector at the Belle II experiment*, [*PoS EPS-HEP2021*, 803 \(2022\)](#) (cit. on p. 60).
- [91] V. Aulchenko et al., *Electromagnetic calorimeter for Belle II*, [*J. Phys. Conf. Ser.* **587**, 012045 \(2015\)](#) (cit. on p. 60).
- [92] T. Aushev et al., *A scintillator based endcap K_L and muon detector for the Belle II experiment*, [*Nucl. Instrum. Meth. A* **789**, 134–142 \(2015\)](#) (cit. on p. 60).
- [93] T. Abe, I. Adachi, K. Adamczyk, et al., *Belle II Technical Design Report*, [arXiv:1011.0352 \[physics.ins-det\]](#), [10.48550/arXiv.1011.0352](#) (2010) (cit. on p. 61).
- [94] M. Milesi, J. Tan, and P. Urquijo, *Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers*, [*EPJ Web Conf.* **245**, 06023 \(2020\)](#) (cit. on p. 61).
- [95] T. Kuhr et al., *The Belle II Core Software*, [*Computing and Software for Big Science* **3**, 1 \(2018\)](#) (cit. on p. 62).
- [96] D. Lange, *The EvtGen particle decay simulation package*, [*Nucl. Instrum. Meth. A* **462**, 152–155 \(2001\)](#) (cit. on pp. 62, 74, 80).
- [97] C. Bierlich, S. Chakraborty, N. Desai, et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, [*SciPost Phys. Codebases* **8**, 10.21468/SciPostPhysCodeb.8 \(2022\)](#) (cit. on p. 63).
- [98] S. Jadach et al., *Multi-photon Monte Carlo event generator KKMCEE for lepton and quark pair production in lepton colliders*, [*Comput. Phys. Commun.* **283**, 108556 \(2023\)](#) (cit. on p. 63).
- [99] N. Davidson et al., *Universal interface of TAUOLA: Technical and physics documentation*, [*Comput. Phys. Commun.* **183**, 821–843 \(2012\)](#) (cit. on p. 63).
- [100] E. Barberio, B. van Eijk, and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays*, [*Comput. Phys. Commun.* **66**, 115–128 \(1991\)](#) (cit. on p. 63).

- [101] C. Carloni Calame et al., *The BABAYAGA event generator*, *Nucl. Phys. B Proc. Suppl., SIGHADO3* **131**, 48–55 (2004) (cit. on p. 63).
- [102] G. Rodrigo et al., *Radiative return at nlo and the measurement of the hadronic cross-section in electron-positron annihilation*, *Eur. Phys. J. C* **24**, 71–82 (2002) (cit. on p. 63).
- [103] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth. A* **506**, 250–303 (2003) (cit. on p. 63).
- [104] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, *CERN-OPEN-2012-016*, 10.17181/CERN-OPEN-2012-016 (2012) (cit. on pp. 66, 67).
- [105] L. Heinrich et al., *Pyhf: pure-Python implementation of HistFactory statistical models*, *Journal of Open Source Software* **6**, 2823 (2021) (cit. on pp. 66, 121).
- [106] Cranmer, Kyle and Held, Alexander, *Building and steering binned template fits with cabinetry*, *EPJ Web Conf.* **251**, 03067 (2021) (cit. on p. 67).
- [107] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, *Comput. Softw. Big Sci.* **1**, 2 (2017) (cit. on p. 69).
- [108] C. Praz, *Search for $B \rightarrow K\nu\nu$ decays with a machine learning method at the Belle II experiment*, PhD thesis (U. Hamburg (main), Hamburg U., Hamburg U., Dept. Math., Hamburg, 2022) (cit. on p. 69).
- [109] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (Aug. 2016), pp. 785–794 (cit. on pp. 70, 93).
- [110] J. Bergstra et al., *Hyperopt: a Python library for model selection and hyperparameter optimization*, *Comput. Sci. Discov.* **8**, 014008 (2015) (cit. on p. 71).
- [111] T. Akiba et al., *Optuna: a next-generation hyperparameter optimization framework*, *arXiv:1907.10902* (2019) (cit. on pp. 72, 93, 142).
- [112] J. Mockus, *Bayesian approach to global optimization*, Vol. 37 (Jan. 2006), pp. 473–481 (cit. on p. 72).
- [113] S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, in *Adv. Neural Inf. Process. Syst.* Vol. 30 (2017) (cit. on pp. 72, 93).
- [114] S. M. Lundberg et al., *From local explanations to global understanding with explainable AI for trees*, *Nat Mach Intell* **2**, 56–67 (2020) (cit. on p. 73).
- [115] T. Keck et al., *The full event interpretation*, *Computing and Software for Big Science* **3**, 6 (2019) (cit. on pp. 73, 74, 78, 156).
- [116] M. Feindt et al., *A hierarchical neurobayes-based algorithm for full reconstruction of B mesons at B factories*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654**, 432–440 (2011) (cit. on pp. 73, 142).
- [117] *eFFORT*, <https://github.com/b2-hive/eFFORT> (2022) (cit. on p. 74).
- [118] F. U. Bernlochner et al., *Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays*, *Eur. Phys. J. C* **80**, 883 (2020) (cit. on p. 74).
- [119] *eFFORT2*, <https://github.com/MarkusPrim/eFFORT2> (2023) (cit. on p. 75).
- [120] C. G. Boyd, B. Grinstein, and R. F. Lebed, *Precision corrections to dispersive bounds on form factors*, *Phys. Rev. D* **56**, 6895–6911 (1997) (cit. on p. 75).

- [121] F. U. Bernlochner et al., *An Introduction to Hammer: Helicity Amplitude Module for Matrix Element Reweighting*, <https://hammer.physics.lbl.gov/HammerManual.pdf> (2024) (cit. on p. 75).
- [122] C. A. Manzari and D. J. Robinson, *On-shell recursion and holomorphic HQET for heavy quark hadronic resonances*, [arXiv:2402.12460](https://arxiv.org/abs/2402.12460) (2024) (cit. on pp. 76, 91).
- [123] J. Zhou and C. Shen, *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*, *Chinese Physics C*, **10.1088/1674-1137/ad806c** (2024) (cit. on p. 79).
- [124] C. Ramirez, J. F. Donoghue, and G. Burdman, *Semileptonic $b \rightarrow u$ decay*, *Phys. Rev. D* **41**, 1496 (1990) (cit. on p. 80).
- [125] O. L. Buchmüller and H. U. Flächer, *Fit to moments of inclusive $B \rightarrow X_c \ell \bar{\nu}$ and $B \rightarrow X_s \gamma$ decay distributions using heavy quark expansions in the kinetic scheme*, *Phys. Rev. D* **73**, 073008 (2006) (cit. on p. 80).
- [126] P. del Amo Sanchez, J. Lees, V. Poireau, et al., *Observation of the Decay $B^- \rightarrow D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell$* , *Phys. Rev. Lett.* **107**, 041804 (2011) (cit. on p. 82).
- [127] J. Stypula, M. Rozanska, I. Adachi, et al., *Evidence for $B^- \rightarrow D_s^+ K^- \ell^- \bar{\nu}_\ell$ and search for $B^- \rightarrow D_s^{*+} K^- \ell^- \bar{\nu}_\ell$* , *Phys. Rev. D* **86**, 072007 (2012) (cit. on p. 82).
- [128] J. Lees, V. Poireau, V. Tisserand, et al., *Observation of $\bar{B} \rightarrow D^{(*)} \pi^+ \pi^- \ell^- \bar{\nu}$ decays in $e^+ e^-$ collisions at the $\Upsilon(4S)$ resonance*, *Phys. Rev. Lett.* **116**, 041801 (2016) (cit. on p. 82).
- [129] F. Meier et al., *First observation of $B \rightarrow \bar{D}_1(\rightarrow \bar{D} \pi^+ \pi^-) \ell^+ \nu_\ell$ and measurement of the $B \rightarrow \bar{D}^{(*)} \pi \ell^+ \nu_\ell$ and $B \rightarrow \bar{D}^{(*)} \pi^+ \pi^- \ell^+ \nu_\ell$ branching fractions with hadronic tagging at belle*, *Phys. Rev. D* **107**, 092003 (2023) (cit. on p. 82).
- [130] F. U. Bernlochner et al., *Constraints on exclusive branching fractions $\mathcal{B}_i(B^+ \rightarrow X_c^i \ell^+ \nu)$ from moment measurements in inclusive $B \rightarrow X_c \ell \nu$ decays*, *Eur. Phys. J. C* **74**, 2914 (2014) (cit. on p. 82).
- [131] F. U. Bernlochner and Z. Ligeti, *Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$* , *Phys. Rev. D* **95**, 014022 (2017) (cit. on p. 82).
- [132] R. Aaij et al., *Measurements of the branching fractions for $B_{(s)} \rightarrow D_{(s)} \pi \pi \pi$ and $\Lambda_b^0 \rightarrow \Lambda_c^+ \pi \pi \pi$* , *Phys. Rev. D* **84**, 092001 (2011) (cit. on p. 83).
- [133] W. Sutcliffe and F. Bernlochner, *Performance and calibration of hadronic tagging with the FEI*, *Belle II Intern. Note* (2023) (cit. on p. 85).
- [134] M. Milesi, *Muon and electron identification performance with 189 fb¹ of Belle II data*, *Belle II Intern. Note* (2022) (cit. on pp. 86, 97, 127).
- [135] D. Ferlewicz et al., *Study of the dependency on track-based isolation variables of lepton identification efficiencies from $J/\psi \rightarrow \ell^+ \ell^-$ candidates and pion-lepton mis-identification rates from $K_S^0 \rightarrow \pi^+ \pi^-$ candidates*, *Belle II Intern. Note* (2021) (cit. on pp. 86, 97, 127).
- [136] S. Sandilya and A. Schwartz, *Kaon and pion identification performances in Phase III data*, *Belle II Intern. Note* (2020) (cit. on pp. 86, 97, 127).
- [137] M. Hohmann, *An inclusive measurement of $|V_{ub}|$ at the Belle experiment*, MA thesis (The University of Melbourne, 2019) (cit. on p. 88).
- [138] P. Cheema et al., *Development of FastBDT classifiers to suppress beam background clusters and fake photons*, *Belle II Intern. Note* (2023) (cit. on p. 88).

- [139] R. M. S. R. F. Abudinen E. Ganiev and D. Tonelli, *A novel sensitive probe for photon-energy calibration biases*, [Belle II Intern. Note \(2020\)](#) (cit. on p. 88).
- [140] D. Dorner, P. Rados, and C. Schwanda, *Relative tracking efficiency study of charged, low momentum pions*, [BELLE2-NOTE-PH-2023-035](#) (cit. on pp. 89, 97, 126).
- [141] T. Koga, *Measurement of momentum dependent π^0 reconstruction efficiency with D decays*, [BELLE2-NOTE-PH-2020-061](#) (cit. on pp. 89, 97, 126).
- [142] I. Adachi et al., *Charged-hadron identification at Belle II*, [arXiv:2506.04355 \(2025\)](#) (cit. on pp. 89, 97, 127).
- [143] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: an update*, [Phys. Rev. D](#) **52**, 2783–2812 (1995) (cit. on p. 91).
- [144] A. K. Leibovich et al., *Semileptonic B decays to excited charmed mesons*, [Phys. Rev. D](#) **57**, 308–330 (1998) (cit. on p. 91).
- [145] E. J. Gustafson et al., *A model independent description of $B \rightarrow D\pi\ell\nu$ decays*, [arXiv:2311.00864 \(2024\)](#) (cit. on p. 91).
- [146] D. Martschei et al., *Advanced event reweighting using multivariate analysis*, [J. Phys. Conf. Ser.](#) **368**, 012028 (2012) (cit. on p. 92).
- [147] F. Abudinén et al., *Search for $B^+ \rightarrow K^+\nu\bar{\nu}$ decays using an inclusive tagging method at Belle II*, [Phys. Rev. Lett.](#) **127**, 181802 (2021) (cit. on p. 92).
- [148] W. Sutcliffe et al., *Combination of calibration factors for hadronic tagging*, [BELLE2-NOTE-PH-2023-029](#) (cit. on pp. 96, 127).
- [149] B. Scavino, *Lambda reconstruction at Belle II*, [Belle II Intern. Note \(2020\)](#) (cit. on pp. 97, 127).
- [150] S. Dreyer et al., *Displaced vertex reconstruction at Belle II*, [Belle II Intern. Note \(2021\)](#) (cit. on pp. 97, 127).
- [151] M. Abadi et al., *Tensorflow: large-scale machine learning on heterogeneous distributed systems*, [arXiv:1603.04467 \(2016\)](#) (cit. on p. 98).
- [152] A. Bevan, B. Golob, Th. Mannel, et al., *The Physics of the B Factories*, [Eur. Phys. J. C](#) **74**, 3026 (2014) (cit. on p. 98).
- [153] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, [Phys. Rev. Lett.](#) **41**, 1581 (1978) (cit. on p. 98).
- [154] S. Lee, K. Suzuki, K. Abe, et al., *Evidence for $B^0 \rightarrow \pi^0\pi^0$* , [Phys. Rev. Lett.](#) **91**, 261801 (2003) (cit. on p. 99).
- [155] D. Asner, M. Athanas, D. Bliss, et al., *Search for exclusive charmless hadronic B decays*, [Phys. Rev. D](#) **53**, 1039–1050 (1996) (cit. on p. 99).
- [156] M. Roehrken, *Continuum suppression*, [Belle II Software](#) (cit. on p. 99).
- [157] P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for Activation Functions*, Oct. 2017 (cit. on pp. 100, 104).
- [158] I. Loshchilov and F. Hutter, *Decoupled Weight Decay Regularization*, in [7th Int. Conf. Learn. Represent. ICLR 2019 New Orleans USA May 6-9 2019](#) (2019) (cit. on pp. 100, 104).
- [159] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, [Phys. Rev. Lett.](#) **125**, 122001 (2020) (cit. on p. 105).
- [160] F. Bernlochner et al., *Novel approaches to determine $B^\pm$ and B^0 meson production fractions*, [Phys. Rev. D](#) **110**, 014007 (2024) (cit. on p. 108).

- [161] H. Junkerkalefeld, *Tests of lepton universality in the rates of inclusive semileptonic B-meson decays at Belle II*, PhD thesis (Physikalisches Institut, U. Bonn (main), 2024) (cit. on p. 114).
- [162] I. Adachi et al., *Determination of $|V_{cb}|$ using $B \rightarrow D\ell\nu_\ell$ Decays at Belle II*, [arXiv:2506.15256 \(2025\)](#) (cit. on p. 114).
- [163] I. Adachi et al., *Evidence for $B^+ \rightarrow K^+\nu\bar{\nu}$ decays*, [Phys. Rev. D **109**, 112006 \(2024\)](#) (cit. on pp. 114, 173).
- [164] F. Abe et al., *Measurement of $\sigma B(W \rightarrow e\nu)$ and $\sigma B(Z^0 \rightarrow e^+e^-)$ in $\bar{p}p$ collisions at $\sqrt{s} = 1800$ gev*, [Phys. Rev. D **44**, 29–52 \(1991\)](#) (cit. on p. 115).
- [165] P. Skands, S. Carrazza, and J. Rojo, *Tuning pythia 8.1: the monash 2013 tune*, [The European Physical Journal C **74**, 3024 \(2014\)](#) (cit. on p. 125).
- [166] Private communication with Peter Skands (cit. on p. 125).
- [167] M. Althoff, W. Braunschweig, F. Kirschfink, et al., *A Detailed Study of Strange Particle Production in e^+e^- Annihilation at High-energy*, [Z. Phys. C **27**, 27 \(1985\)](#) (cit. on p. 125).
- [168] W. Bartel, L. Becker, C. Bowdery, et al., *Charged particle and neutral kaon production in e^+e^- annihilation at PETRA*, [Z. Phys. C - Particles and Fields **20**, 187–206 \(1983\)](#) (cit. on p. 125).
- [169] P. Rados et al., *Measurement of the track reconstruction efficiency and fake rate with $e^+e^- \rightarrow \tau^+\tau^-$ events*, [Belle II Intern. Note \(2020\)](#) (cit. on p. 127).
- [170] L. Cao, W. Sutcliffe, R. Van Tonder, et al., *Measurements of Partial Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays with Hadronic Tagging*, [Phys. Rev. D **104**, 012008 \(2021\)](#) (cit. on pp. 142, 146–148).
- [171] J. Lees, V. Poireau, V. Tisserand, et al., *Study of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays in $B\bar{B}$ events tagged by a fully reconstructed B-meson decay and determination of $|V_{ub}|$* , [Phys. Rev. D **86**, 032004 \(2012\)](#) (cit. on pp. 146, 147).
- [172] B. Efron and R. Tibshirani, *An introduction to the bootstrap (1st ed.)* (Chapman and Hall/CRC, 1994) (cit. on p. 147).
- [173] M. Neubert, *Two-loop relations for heavy-quark parameters in the shape-function scheme*, [Physics Letters B **612**, 13–20 \(2005\)](#) (cit. on p. 151).
- [174] M. Neubert, *Advanced predictions for moments of the $\bar{B} \rightarrow X_s \gamma$ photon spectrum*, [Phys. Rev. D **72**, 074025 \(2005\)](#) (cit. on p. 151).
- [175] G. Finauri and P. Gambino, *The q^2 moments in inclusive semileptonic B decays*, [Journal of High Energy Physics **2024**, 206 \(2024\)](#) (cit. on p. 151).
- [176] C. Bozzi, member of the HFLAV collaboration; private communication (cit. on p. 151).
- [177] M. Hohmann et al., *Measurement of the ratio of partial branching fractions of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and the ratio of their spectra with hadronic tagging*, [Phys. Rev. D **111**, 092016 \(2025\)](#) (cit. on p. 152).
- [178] Y. Aoki et al., *FLAG Review 2024*, [arXiv:2411.04268 \(2025\)](#) (cit. on p. 158).