technische universität
dortmund

PhD Thesis

Measurement of Isospin Asymmetry in $B \to K\mu^+\mu^-$ decays at the LHCb experiment

Fabio De Vellis

First Reviewer
Prof. Dr. Johannes Albrecht

Supervisor
Biljana Mitreska

Second Reviewer
PD Dr. Dominik Elsässer

Technische Universität Dortmund
Fakultät Physik
AG Albrecht

Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegte Dissertation.

Gutachter:
Prof. Dr. Johannes Albrecht
PD Dr. Dominik Elsässer

Vorsitzender der Prüfungskommission:
Prof. Dr. Zhe Wang

Vertreter der wissenschaftlichen Mitarbeiter:
Dr. Moritz Cygorek

Datum der Einreichung der Dissertation:
10.02.2026

Datum der mündlichen Prüfung:
17.04.2026

"Chiedo tempo, son della razza mia,
per quanto grande sia, il primo che ha studiato."

— Francesco Guccini

Abstract

Symmetries in the Standard Model of particle physics stem from fundamental aspects of the theory and play a determinant role in shaping our understanding of elementary particles. Isospin symmetry predicts nearly identical decay rates for processes differing only by a spectator quark, such as the rare decays $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$. The isospin asymmetry, defined as the difference in their partial widths, is a theoretically clean observable: dominant hadronic uncertainties cancel out, offering a precise test of the theory.

Previous measurements have hinted at coherent deviations from the Standard Model expectations, motivating a more precise determination using a larger dataset. This thesis details the complete strategy for an updated measurement of the differential branching fractions and isospin asymmetry of these decays, using the full proton-proton collision dataset collected by the LHCb experiment during Run 1 and Run 2, corresponding to $9\,\mathrm{fb}^{-1}$. The analysis is performed in intervals of the dimuon invariant mass squared, q^2, using the resonant decays $B^{0,+} \to J/\psi K^{0,+}$ as normalisation to reduce systematic uncertainties.

A blinded approach is adopted to prevent experimenter bias. Consequently, while the final numerical results are not included, this thesis details the comprehensive analysis strategy. All key aspects are presented, including the optimisation of the event selection, the calibration of the simulation, and the development of the simultaneous fitting framework. The robustness of the method is demonstrated through extensive validation using control modes and pseudo-experiments. Finally, the expected sensitivity of the measurement is evaluated, highlighting the potential of the full dataset to significantly improve the precision compared to previous results.

Zusammenfassung

Symmetrien im Standardmodell der Teilchenphysik entspringen fundamentalen Aspekten der Theorie und spielen eine bestimmende Rolle für unser Verständnis der Elementarteilchen. Die Isospin-Symmetrie sagt nahezu identische Zerfallsraten für Prozesse voraus, die sich nur durch ein Zuschauerquark unterscheiden, wie etwa die seltenen Zerfälle $B^0 \to K^0\mu^+\mu^-$ und $B^+ \to K^+\mu^+\mu^-$. Die Isospin-Asymmetrie, definiert als die Differenz ihrer Partialbreiten, ist eine theoretisch saubere Observable: Dominierende hadronische Unsicherheiten heben sich auf, was einen präzisen Test der Theorie ermöglicht.

Frühere Messungen lieferten Hinweise auf kohärente Abweichungen von den Erwartungen des Standardmodells, was eine präzisere Bestimmung unter Verwendung eines größeren Datensatzes motiviert. Diese Arbeit beschreibt die vollständige Strategie für eine aktualisierte Messung der differenziellen Verzweigungsverhältnisse und der Isospin-Asymmetrie dieser Zerfälle unter Verwendung des gesamten Proton-Proton-Kollisionsdatensatzes, der vom LHCb-Experiment während Run 1 und Run 2 aufgezeichnet wurde und einer integrierten Luminosität von $9\,\mathrm{fb}^{-1}$ entspricht. Die Analyse wird in Intervallen der quadrierten invarianten Dimuon-Masse, q^2, durchgeführt, wobei die resonanten Zerfälle $B^{0,+} \to J/\psi K^{0,+}$ als Normalisierung verwendet werden, um systematische Unsicherheiten zu reduzieren.

Um eine unbewusste Beeinflussung zu vermeiden, wird ein verblindeter Analyseansatz (blinded analysis) gewählt. Folglich sind die endgültigen numerischen Ergebnisse nicht enthalten, jedoch beschreibt diese Arbeit die umfassende Analysestrategie. Alle wesentlichen Aspekte werden vorgestellt, einschließlich der Optimierung der Ereignisselektion, der Kalibrierung der Simulation und der Entwicklung des Frameworks für den simultanen Fit. Die Robustheit der Methode wird durch umfangreiche Validierung unter Verwendung von Kontrollkanälen und Pseudo-Experimenten demonstriert. Schließlich wird die erwartete Sensitivität der Messung evaluiert, wobei das Potenzial des vollen Datensatzes hervorgehoben wird, die Präzision im Vergleich zu früheren Ergebnissen signifikant zu verbessern.

Acknowledgements

Writing a PhD thesis is often described as the solitary struggle of a poor student against his project, but looking back at these past years, I realise how far from the truth that is. This work is the culmination of a long and challenging journey. And while it directly involved scientific research, I would like to stress how far beyond that it went in shaping my personal growth. I could never have reached this milestone without the guidance, support and patience of the many people who walked alongside me.

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Johannes Albrecht, for giving me the opportunity to work in his research group and to be part of the LHCb collaboration. I truly admire Johannes for his talent in shaping such a fantastic, stimulating, and self-reliant team. He always showed an extraordinary empathy, sensing exactly what a specific moment required. He often understood my struggles before I fully did, giving me the perspective I needed to understand myself. Fortunately, this kind of support is not a rare exception in academia, as demonstrated by my second referee, PD Dr. Dominik Elsässer. I want to thank Dominik for his willingness to evaluate this thesis, not treating it as a mere duty, but showing genuine interest in my work and my experience. I am especially grateful for his extreme flexibility, which made the final steps of this process as smooth as possible during a very stressful time for me.

A very special thank you goes to my daily supervisors, Vitalii Lisovskyi and Biljana Mitreska. I want to thank Vitalii for teaching me the foundations of data analysis during the early stages of my PhD. I genuinely consider him one of the most brilliant scientists I have ever met, and each discussion with him was a stimulating challenge. I am equally grateful to Biljana, whose exceptional management and mentoring skills were the true anchor of my daily work. She has been a formidable guide, constantly pushing me to improve myself and providing structure and unwavering support exactly when I needed it most. If there is clarity and quality in this manuscript, it is largely to her credit.

I am very grateful to Thomas Oeser and Christoph Langenbruch for the excellent teamwork. Developing our analyses in parallel alongside the $B \to K^{*}\mu^{+}\mu^{-}$ counterpart was essential for refining our strategies, and sharing this effort made the daily work much more enjoyable.

I also want to thank all the colleagues and friends at TU Dortmund. The welcoming atmosphere and the shared time both in and out of the office made this journey much more pleasant. A special mention goes to my office mates Jan, Louis, and Jannis, who helped me integrate into German culture and made me a real Schlager fan. Furthermore, I gratefully acknowledge the support of the entire LHCb collaboration; even though it is impossible to know everyone personally in such a large experiment, I have always felt like part of a warm community.

Finally, none of this would have been possible without the unconditional love and support of my family and friends outside of physics. Thank you to my mother and father for asking and listening, even though this world is miles away from you. Thank you to my longtime friends, Forte and Lucia, for reminding me that there is much more to life than work. Thank you to the new friends I found in Dortmund—Filippo, Elena, Mahnaz, and Emad—who helped me feel at home in Germany. And a final, special mention goes to Gaia for always standing by my side, even when I am my own worst enemy.

This thesis was defended on April 17.
On this day in 1944, the Quadraro neighborhood in Rome
suffered a devastating deportation.
May the pursuit of knowledge always walk hand in hand
with the preservation of memory and freedom.

Contents

1. Introduction

The progress of physics has been historically driven by the ambition to find unity in diversity. Moving beyond the idea of distinct laws for different domains of reality, the scientific enquiry has focused on distilling the complexity of the Universe into its most fundamental components. This pursuit of the fundamental, applied to both visible and invisible matter, has culminated in the formulation of the Standard Model of particle physics. This theoretical framework represents one of the greatest achievements of modern science, describing with remarkable precision the behaviour of elementary particles over a vast range of energies.

Despite its experimental success, the Standard Model is widely acknowledged to be an incomplete theory, as it fails to address several critical phenomena observed in nature. Traditionally, the limitations are summarised in four major points. Firstly, the Standard Model offers no explanation for the nature of dark matter and dark energy, which constitute the majority of the Universe's energy budget. Secondly, it cannot account for the observed prevalence of matter over antimatter. Thirdly, it does not incorporate gravity into a quantum framework. Finally, it provides no mechanism for the origin of neutrino masses.

These open questions point to the existence of Physics Beyond the Standard Model, driving the current experimental programme in high-energy physics. At high-energy colliders, such as the Large Hadron Collider (LHC), this search proceeds along two complementary frontiers. The *direct approach* seeks to produce new, heavy particles on-shell by smashing protons at the highest possible energies. Conversely, the *indirect approach* focuses on precision measurements of low-energy processes. Here, new heavy particles, too massive to be produced directly, can manifest as virtual contributions in quantum loops, altering decay rates or angular distributions predicted by the Standard Model.

In this context, rare decays of B mesons mediated by flavour-changing neutral currents, such as the $b \to s\ell^+\ell^-$ transitions, are particularly sensitive probes. Being forbidden at tree level in the SM, these processes are suppressed and thus highly susceptible to New Physics contributions, which could compete with the Standard Model effects. Among the observables accessible in these decays, the isospin asymmetry, defined as the difference in the partial widths between charged and neutral B meson decays, plays a unique role. It is a theoretically clean quantity where the dominant uncertainties related to hadronic form factors cancel out at leading order, offering a precise test of the Standard Model.

This thesis presents an updated measurement of the differential branching fractions of the rare decays $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0\mu^+\mu^-$, and of their isospin asymmetry, using the full dataset collected by the LHCb experiment during Run 1 and Run 2. After the definition of a theoretical framework of the Standard Model and the phenomenology of rare B decays in Ch. 2, and the description of the LHCb detector in Ch. 3, the analysis strategy is discussed in Ch. 4. Proceeding to the core of the analysis, the selection of the data from the large LHCb dataset is detailed in Ch. 5, followed by the description of the corrections applied to the simulated samples in Ch. 6. Ch. 7 illustrates the fit models used to extract the signal yields from the invariant mass spectra. The simultaneous fit procedure to determine the branching fractions and the isospin asymmetry is presented in Ch. 8. Finally, the evaluation of systematic uncertainties is discussed in Ch. 9, leading to the final remarks and conclusions.

Author’s contribution

The work presented in this thesis is based on data collected by the LHCb experiment during Run 1 and Run 2. While the data acquisition and the development of the general experimental software are the result of the collective effort of the LHCb collaboration, the analysis presented in this document is the result of my original work, under the supervision of Vitalii Lisovskyi first, and Biljana Mitreska later.

I personally optimised the selection criteria, including the definition of the multivariate selection strategies and the specific vetoes for peaking backgrounds. I developed the full correction chain for the simulated samples and designed the fit framework used to extract the observables of interest, implementing the simultaneous maximum likelihood fit to the resonant and rare modes. Furthermore, the validation of the fitting strategy and the evaluation of the systematic uncertainties are the result of my own work.

The measurement of the isospin asymmetry in $B \to K\mu^+\mu^-$ decays is part of a larger analysis comprising the vector counterpart represented by the $B \to K^*\mu^+\mu^-$ decays. The complete analysis is a joint effort with my colleagues Thomas Oeser and Christoph Langenbruch, who developed a parallel framework for the K^* mode. While the code implementation and the data analysis for the $B \to K\mu^+\mu^-$ channel were performed independently by me, the overall analysis strategy was defined jointly to ensure coherence between the scalar and vector modes in view of a single publication.

The results presented here are currently being prepared for publication within the collaboration.

2. Theoretical Foundations

The Standard Model of particle physics stands as one of the most successful achievements of modern science, providing a unified and elegant description of the fundamental constituents of matter and their interactions. Precise experimental tests of the theory, such as the one presented in this thesis, rely on the interplay between the fundamental Lagrangian and effective approximations capable of describing the dynamics at the energy scale of the B mesons.

This chapter outlines the theoretical framework necessary to interpret the measurement of the isospin asymmetry in $B \to K\mu^+\mu^-$ decays. The discussion moves from the foundations of the Standard Model, its particle content and the gauge principle (Sec. 2.1–Sec. 2.3), to the mathematical formalism of perturbative Quantum Field Theory (Sec. 2.5). Particular attention is dedicated to the Effective Field Theory approach and the Operator Product Expansion (Sec. 2.6), which are essential to disentangle short-distance physics from hadronic effects. Finally, the focus is shifted to the phenomenology of $b \to s\ell\ell$ transitions (Sec. 2.7) and the definition of the primary observable of this analysis, the isospin asymmetry (Sec. 2.8).

2.1. Fundamental particles

The Standard Model of particle physics represents currently the most accurate picture of the subatomic world. According to the model, the known matter (and antimatter) can be explained in terms of fundamental particles and their interactions. The entire complexity of our universe arises from a limited number of fundamental particles, shown in Fig. 2.1.

The elementary fermions are the constituents of matter. They are twelve spin-1/2 particles further classified into quarks and leptons. Each has a corresponding antiparticle, characterised by the same quantum numbers, but opposite charge.

The six quarks, namely up (u), down (d), charm (c), strange (s), top (t) and bottom (b), are grouped in pairs into three generations, (u, d), (c, s), (t, b), largely differentiated in mass. Within each generation an up-type and a down-type quark can be distinguished, the former carrying an electric charge of 2/3 e, the latter with a charge of -1/3 e. The electric charge regulates the electromagnetic interaction, but quarks also experience other fundamental forces that differ in range, strength, and characteristic timescales, collectively known as the *strong interaction*. Analogously, this is governed by a strong charge which can assume three values traditionally named after colours, namely *red*, *green* and *blue* together with their anti-colours *anti-red*, *anti-green*, *anti-blue* for antiquarks. Free quarks have never been observed. Instead, they are found in nature only in *colourless* bound states, where a colour is combined with its respective anti-colour in a two-quark state (*mesons*), or all three different colours are combined together in a three-quark state (*baryons*). More complex combinations are also possible, as demonstrated by the observations of *tetraquark* [1] and *pentaquark* [2] states, formed by four or five quarks respectively. This phenomenon is explained theoretically by the concept of *confinement* [3].

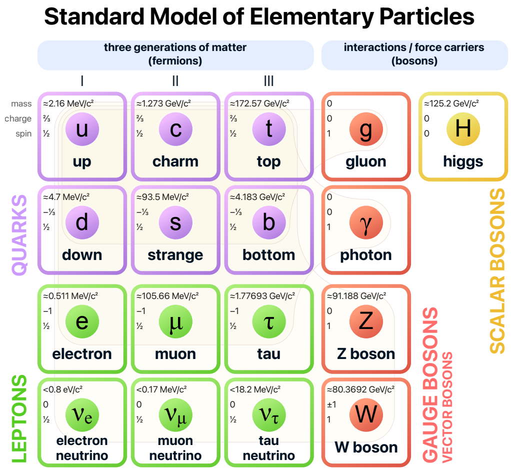


Figure 2.1.: The elementary particles of the Standard Model (adapted from Wikimedia Commons).

The six leptons can be collected into three families. They are the electron (e^-), the muon (μ^-) and the tau (τ^-), with electric charge of -e, all with their respective electrically neutral neutrino (ν_e, ν_μ, ν_τ). In contrast to quarks, leptons do not carry any colour charge and are therefore not subject to the strong interaction. They interact only via electroweak forces, which are insufficient to form bound states.

In the Standard Model the interactions are mediated by so-called *gauge bosons*. These are spin-1 particles, whose exchange determines the nature of the force involved. The photon (γ) is the mediator of the electromagnetic interaction and couples to electric charge. It is massless, a feature with a deep theoretical interpretation, which will be covered in the next section. Gluons (g) are also massless and mediate the strong interaction. There are distinct gluons, each corresponding to a coloured combination of a colour and an anti-colour. Because such combinations are not colourless, free gluons, like free quarks, are not observable. Finally, the three massive W^+, W^- and Z bosons are the mediators of the weak interaction, and the only ones that couple to all the 12 fermions.

Completing the Standard Model is the recently discovered Higgs boson (H). Unlike the gauge bosons, the Higgs boson has spin 0 and is associated with a scalar field, namely the *Higgs field*. This field is however quite unique: rather than mediating an observed interaction, it is introduced to spontaneously break the symmetry of the model and to allow particle masses to arise through a perturbative expansion around the local minimum [4]. Within this framework, the masses turn out to be the coupling constants of the various particles to the Higgs. This interpretation provided a formal solution to the problem of masses within a gauge theory, and for this reason the observation of the Higgs boson in 2012 by the ATLAS and CMS collaborations at CERN [5, 6] represented a major success for the Standard Model which predicted its existence almost 50 years earlier.

2.2. In search of lost symmetry

The path that led to the compact picture depicted in the last section was not linear and extended through half a century amidst right and wrong hypotheses. The efforts in unifying the principles of quantum mechanics and special relativity resulted in the formulation of quantum field theories [7]. The Standard Model is a quantum field theory, where any of the particles listed in Fig. 2.1 is associated with a quantised field, i.e. an operator defined over spacetime. However, while the field interpretation is sufficient to remove ambiguities in the works of Klein and Gordon [8, 9] and Dirac [10], and successfully describes free particles, considerations are required to incorporate interactions consistently in the theory.

The concept of *gauge invariance* was already known from classical electromagnetism. When the theory was formulated in terms of scalar and vector potentials ϕ and A, it was noted that the derived physical fields

$$\mathrm{E} = -\nabla\phi - \frac{\partial \mathrm{A}}{\partial t} \quad , \qquad \mathrm{B} = \nabla \times \mathrm{A} \tag{2.1}$$

remain unchanged under a whole class of transformations of the potentials themselves, defined by any differentiable function $\lambda(\mathrm{x}, t)$

$$\begin{cases} \mathrm{A} & \longrightarrow \mathrm{A}' = \mathrm{A} + \nabla\lambda \\ \phi & \longrightarrow \phi' = \phi - \frac{\partial\lambda}{\partial t} \end{cases} . \tag{2.2}$$

Following parallel developments in mathematics, and in particular in group theory, it became natural to identify this class of transformations as a symmetry group. Pioneering work by Yang and Mills [11] extended the concept of gauge invariance to local and non-abelian symmetry groups. This effort was justified by the aim of finding symmetries that could lead, through Noether's theorem [12], to conserved currents and hence interactions. The correct symmetry underlying electroweak interactions was proposed by Glashow [13]

$$SU(2)_L \otimes U(1)_Y \, ,$$

where the subscripts denote that the two groups act independently and are associated with distinct quantum numbers, *weak isospin* (L) and *hypercharge* (Y).

Interpreted within the framework of group theory, Noether's theorem implies that whenever a continuous symmetry group is introduced, it gives rise to a number of massless gauge bosons equal to the number of generators of that group. In the case just presented it would be three for $SU(2)_L$ and one for $U(1)_Y$, giving a total of four. However, such massless gauge bosons could not explain the short range of weak interactions. Another concept had to be incorporated into the model, the *spontaneous symmetry breaking* mechanism proposed by Higgs [4]. Through the previously introduced Higgs field, the four-dimensional $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ symmetry is broken down to the one-dimensional $U(1)_\mathrm{EM}$, whose associated gauge boson can be interpreted as the photon without any ambiguity. The fact that the photon is massless ensures that this residual symmetry remains exact and unbroken.

In its formulation completed by Weinberg [14] and Salam [15], the Standard Model successfully described lepton interactions. Meanwhile, parallel experiments by McAllister and Hofstadter [16], and theoretical developments by Feynman [17], Bjorken [18] and Gell-Mann [19] supported the hypothesis that matter itself is composed of elementary constituents. This resulted in the quark model after the inputs from Glashow, Iliopoulos and Maiani [20], and Kobayashi and Maskawa [21]. The resulting framework is now included in the general theory of particle physics by introducing the eight-dimensional symmetry group $SU(3)_\mathrm{C}$, where the subscript C stands for *colour*, the charge associated with the strong interaction. This symmetry is exact, giving rise to eight massless gauge bosons responsible for the strong force, the gluons.

Finally the complete symmetry of the Standard Model is defined as

$$SU(2)_{\mathrm{L}} \otimes U(1)_{\mathrm{Y}} \otimes SU(3)_{\mathrm{C}} \,,$$

with the caveat of the breaking mechanism

$$SU(2)_{\mathrm{L}} \otimes U(1)_{\mathrm{Y}} \to U(1)_{\mathrm{EM}} \,.$$

Once the symmetry is fixed, the next step is to select the appropriate mathematical objects representing the different particles, and to combine them, while preserving the invariance, into a Lagrangian that encapsulates the full dynamics. As noted earlier, these objects are fields; moreover they require additional properties to satisfy the imposed symmetry. In order for the Lagrangian to be invariant under the chosen group, all fields must transform according to a *representation* of the group itself. Left-handed leptons, for example, belong to the one-dimensional representations of both $U(1)_{\mathrm{Y}}$ and $SU(3)_{\mathrm{C}}$, but to the two-dimensional representation of $SU(2)_{\mathrm{L}}$, forming doublets

$$\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix} . \tag{2.3}$$

In other words, they follow defined transformation rules under the different groups, transforming as scalars under $U(1)_{\mathrm{Y}}$ and $SU(3)_{\mathrm{C}}$ and as spinors with respect to $SU(2)_{\mathrm{L}}$. With these building blocks, it is always possible to construct combinations that are invariant under symmetry transformations, thus constituting the valid terms of the Lagrangian.

2.3. The Standard Model Lagrangian

Since the advent of quantum mechanics it was clear that the Lagrangian formalism, originally developed to give a solid mathematical formulation for classical mechanics, offers the level of abstraction required to describe the non-intuitive nature of the wave function. The same central role is played by the Lagrangian in quantum field theory, where it provides the most natural and compact framework for expressing the model and for making quantitative predictions. As discussed in Sec. 2.2, the Lagrangian must be constructed in accordance with the imposed symmetry, that is, it must remain invariant under the corresponding gauge transformations. When this requirement is fulfilled, gauge bosons and interaction terms naturally emerge from the formalism.

A basic illustration of this mechanism can be given in the simplified case of quantum electrodynamics. A first attempt toward a relativistic quantum interpretation of electromagnetism could be made by combining Dirac's theory for a generic fermion ψ and Maxwell's dynamics for the electromagnetic field A^μ, rewritten in terms of the electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,. \tag{2.4}$$

This results in the free fields Lagrangian (2.4) which contains no interaction. Moreover, it does not exhibit any symmetry of physical interest.

According to the aforementioned Yang-Mills theory, Eq. (2.4) can be made invariant under a local gauge transformation for an arbitrary $\theta(x)$

$$\begin{cases} \psi' = e^{ie\theta(x)}\psi \\ A_\mu{}' = A_\mu + \partial_\mu\theta(x) \end{cases} \tag{2.5}$$

by replacing the partial derivative with a *covariant* one

$$\partial_\mu \longrightarrow D_\mu = \partial_\mu + ieA_\mu \,. \tag{2.6}$$

Substituting Eq. (2.6) into the starting definition of Eq. (2.4)

$$\begin{aligned} \mathcal{L}_{\mathrm{QED}} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &= \underbrace{\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi}_{\text{free fermion field}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Maxwell dynamics}} - \underbrace{e\bar{\psi}\gamma^\mu\psi A_\mu}_{\text{interaction term}} \end{aligned} \tag{2.7}$$

not only makes the Lagrangian invariant under (2.5), as required, but also automatically gives rise to a mixed term that describes the interaction between the fermion and the photon fields.

The construction outlined above holds for the $U(1)_{\mathrm{EM}}$ group generated by the photon field A^μ, but the same machinery can be applied with some modification to the complete Standard Model symmetry. In particular it should be noted that the fermion mass appears explicitly in Eq. (2.7), thus breaking gauge invariance in the general case. The extension must be formulated under the hypothesis of massless particles, and with the help of an ad hoc field, the Higgs field ϕ, and its associated potential $V(\phi)$. The interactions between the Higgs field and the matter fields can then be interpreted as mass terms, arising after spontaneous symmetry breaking. These considerations lead to the Standard Model Lagrangian

$$\mathcal{L}_{\mathrm{SM}} = \underbrace{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Gauge sector}} + \underbrace{i\bar{\psi}\not{D}\psi + \mathrm{h.c.}}_{\text{Fermion-gauge boson interaction}} + \underbrace{\psi_{\mathrm{L}}y_{\mathrm{LR}}\psi_{\mathrm{R}}\phi + \mathrm{h.c.}}_{\text{Yukawa coupling}} + \underbrace{\left|D_\mu\phi\right|^2 - V(\phi)}_{\text{Higgs sector}} \tag{2.8}$$

presented here in its compact formulation, where ψ can now be a left-handed doublet in the form of Eq (2.3) or a right-handed singlet, and runs over all the fermions, the covariant derivatives D_μ include all the gauge bosons, and the $\not{D} = \gamma^\mu D_\mu$ notation is introduced. The matrix y_{LR} regulates the mixing of left- and right-handed fermion fields in the *Yukawa term*, which can be interpreted as a mass term for fermions after symmetry breaking. The mass terms for the massive bosons are implicitly contained in the covariant derivative acting on the Higgs field.

2.4. Quark mixing

The Standard Model Lagrangian derived in the previous section encapsulates the complete dynamics of particle interactions. However, a fundamental feature governing the phenomenology of flavour physics is embedded within the fermion mass generation mechanism. Before proceeding further, it is necessary to distinguish between the fields appearing in the gauge couplings and the states with definite mass. The physical fields are defined as the mass eigenstates. Since the matrix y_{LR} is generally not diagonal, these do not coincide with the weak interaction eigenstates ψ described so far.

Focusing on the quark sector, to work with physical fields, the interaction eigenstates must be rotated in flavour space to diagonalise the mass terms. Due to electric charge conservation, this will happen independently for up- and down-type quarks through the unitary transformations

$$\begin{pmatrix} u' \\ c' \\ t' \end{pmatrix} = U_u \begin{pmatrix} u \\ c \\ t \end{pmatrix} \ , \quad \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = U_d \begin{pmatrix} d \\ s \\ b \end{pmatrix} \ . \tag{2.9}$$

Crucially, these rotations affect what was called the fermion-gauge boson interaction term differently.

For neutral currents (mediated by γ and Z), the interactions involve terms like $\bar{q}'\gamma^\mu q'$. Under the rotation, these become

$$\bar{q}U^\dagger\gamma^\mu U q = \bar{q}\gamma^\mu q \tag{2.10}$$

due to the unitarity condition $U^\dagger U = \mathbb{1}$, regardless of whether q is an up- or down-type quark. This cancellation ensures the absence of tree-level flavour-changing neutral currents in the Standard Model.

Conversely, the charged-current interaction mediated by $W^\pm$ couples up-type to down-type quarks. In the mass basis, this assumes the form

$$\begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix} U_u^\dagger \gamma^\mu U_d \begin{pmatrix} d \\ s \\ b \end{pmatrix} \ . \tag{2.11}$$

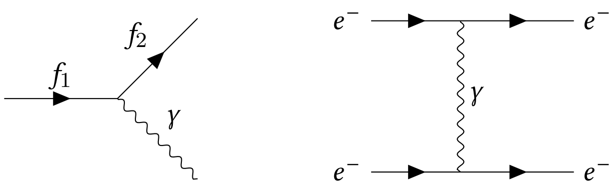
In this case the product $U_u^\dagger U_d$ does not reduce to the identity but generates the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix [21]. The elements V_{ij} of this unitary matrix quantify the strength of the charged-current transition between an up-type quark i and a down-type quark j:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \ . \tag{2.12}$$

A complex 3×3 matrix has 18 free parameters, but its unitarity imposes nine constraints. Together with the unphysical nature of five relative phases between the six quarks, this reduces the number of free parameters to four in the CKM matrix. A common parametrisation was introduced by Wolfenstein [22] in terms of the scale $\lambda = V_{us} \approx 0.22$ and the auxiliary parameters A, ρ and η

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A(\rho - i\eta)\lambda^3 \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A(1 - \rho - i\eta)\lambda^3 & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4) \ . \tag{2.13}$$

This representation has the advantage of exhibiting the hierarchy of terms. Elements on the diagonal are nearly unity, showing that couplings within the same flavour family are favoured. In contrast, transitions between quarks of different families are suppressed by increasing powers of the scale factor λ.



(a) (b)

Figure 2.2.: Quantum electrodynamics vertex coupling two fermions f_1, f_2 to a photon γ (a) and leading-order Feynman diagram of the elastic scattering $e^-e^- \longrightarrow e^-e^-$ (b).

2.5. Feynman diagrams

In classical mechanics the Lagrangian is used to determine the equations of motion of the system under study. In principle this also applies to quantum field theory, where the evolution of the fields can be computed using the *action* formalism. However, while the classical trajectory is something that can be intuitively understood, fields are not physical observables and cannot be measured.

In particle physics it is typical to have scattering experiments, with two colliding particles. Given an initial setup for the experiment, the interest is on the resulting products that due to the probabilistic nature of Quantum Mechanics cannot be determined a priori. Formally, this corresponds to the projection of the initial state $|a\rangle$ onto the final state $|b\rangle$ via an evolution operator S that encodes the scattering process. The *scattering matrix* is then defined as

$$S_{ba} = \langle b| S |a\rangle \tag{2.14}$$

where the squared modulus of these elements represents the probability of transitioning from the initial to the final state.

The operator S is indeed an evolution operator that can be derived from the Lagrangian in Eq. (2.8). However, while an analytical expression can always be written, calculating exactly the value on the state and the matrix element is usually unfeasible. What can be done instead, if the element in Eq. (2.14) can be expressed as a series of terms of decreasing magnitude, is approximating the result with a perturbative expansion. Despite its formal nature, this procedure admits a graphical representation introduced by Feynman [23], which provides a physical interpretation of the computation and allows the interaction vertices to be inferred directly from the Lagrangian terms. Each field appearing in the expansion is associated with a line, whose type indicates the nature of the particle, while each product of fields corresponds to a vertex. In quantum electrodynamics, for example, the interaction term in Eq. (2.7) gives rise to the diagram in Fig. 2.2a, where two fermionic fields, $\bar{\psi}$ and ψ, are connected to the photon, A_μ, via a vectorial vertex, $e\gamma^\mu$. These diagrams can be combined in various ways to connect initial and final states. The graphical correspondence is completed by the rule that any internal line represents a *propagator*, a function of the carried momentum related to the probability for a mediator particle to travel between two spacetime points.

The Feynman diagram in Fig. 2.2b shows the simplest way to connect initial and final state for an elastic electron-electron scattering, representing the leading-order contribution to the process. Looking at the figure, the physical interpretation of the diagram is clear: when approaching, two electrons are exchanging a photon and scatter off each other. This "exchange" interpretation is in general useful to

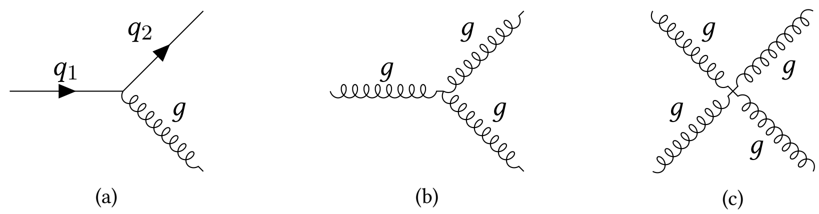


Figure 2.3.: Quantum chromodynamics vertices coupling two quarks q_1 and q_2 with a gluon g (a), three gluons (b) and four gluons (c).

reconstruct the diagrams that are contributing to the process of interest, however it must not be mistaken for a classical causal sequence, which cannot be meaningfully defined within a quantum theory.

2.6. Effective field theory

Feynman diagrams are a powerful tool to obtain the probability of a process directly from the interaction Lagrangian. It is worth stressing, however, that a perturbative expansion is only possible if the quantity in Eq. (2.14) can be written as a series of infinitesimal terms, typically in successive powers of the coupling constant. This condition is not always satisfied. Quantum chromodynamics becomes *non-perturbative* at low energies, meaning that such a series can no longer be defined. Gluon vertices can still be drawn, as shown in Fig. 2.3, and they remain useful at high energies, where the coupling is small and perturbation theory is valid. But when going to low-energy scales ambiguities arise, since higher-order diagrams become more significant than lower-order ones.

This makes the method unsuitable for studying certain processes, such as meson decays where the initial state consists of two quarks bound by low-energy gluons.

To overcome this difficulty, non-perturbative methods like *lattice QCD computations* [3] and *light-cone sum rules* [24] were developed. At present, these represent the only tools to compute non-perturbative contributions and, even with technical complexities that can sometimes be prohibitive, they are sufficient to handle any hadronic matrix element in Eq. (2.14). In this section, the theoretical framework in which flavour-physics measurements are commonly interpreted is presented, following the approach outlined in Ref. [25] to disentangle perturbative and non-perturbative scales.

The concept finds historical foundation in Fermi's theory for β-decay [26] which reduces the weak force in the neutron decay to a point-like chiral interaction. The simplification, possible because the mass of the neutron is much smaller than the mass of the W boson, is depicted in Fig. 2.4: the propagator of the full Feynman diagram is contracted to a point, resulting in a product of two left-handed currents. This translates mathematically into the effective Hamiltonian

$$\mathcal{H}_{\mathrm{F}} = \frac{G_{\mathrm{F}}}{\sqrt{2}} \cos\theta_{\mathrm{C}} \left[\bar{u}\gamma_{\mu}\left(1-\gamma_5\right) d \cdot \bar{e}\gamma^{\mu}\left(1-\gamma_5\right)\nu_e\right] \tag{2.15}$$

where $\cos\theta_{\mathrm{C}}$ fixes the quark coupling and G_{F} is the *Fermi constant.* In this equation and throughout the chapter, particle labels denote the corresponding Dirac field operators (e.g., d represents the down quark field generally denoted as ψ before).

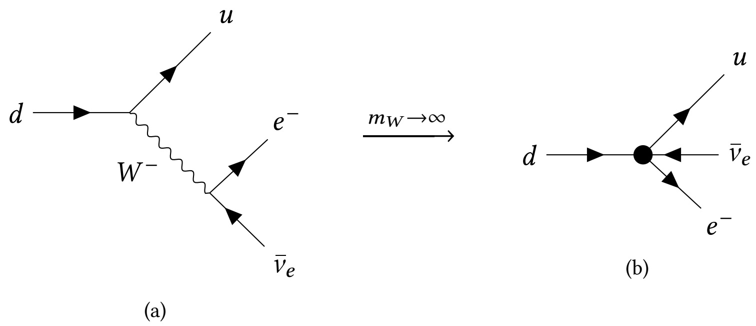


Figure 2.4.: Diagrams for neutron β-decay at quark level in the full (a) and in Fermi's (b) theory.

By analogy, one can introduce several point-like interactions and express a generic effective weak Hamiltonian as

$$\mathcal{H}_{\text{eff}} = \frac{G_{\text{F}}}{\sqrt{2}} \sum_i C_i(\lambda) O_i(\lambda) \ , \tag{2.16}$$

where the Fermi constant is kept for historical reasons. Here the point-like nature of the interactions is encoded in the local operators $O_i(\lambda)$, while the weights $C_i(\lambda)$ of the linear combination are named *Wilson coefficients*. The effective Hamiltonian plays the same role as the S operator in Eq. (2.14) and the decay amplitude of a meson M into a final state F is given by

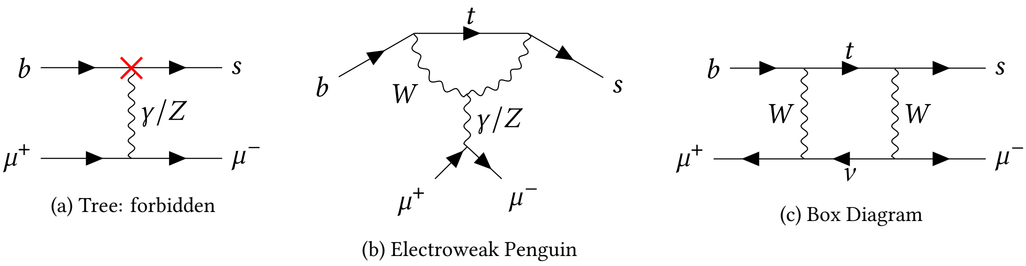
$$\mathcal{A}(M \to F) = \langle F| \mathcal{H}_{\text{eff}} |M\rangle = \frac{G_{\text{F}}}{\sqrt{2}} \sum_i C_i(\lambda) \langle F| O_i(\lambda) |M\rangle \ . \tag{2.17}$$

In both equations, the parameter λ is a factorisation scale that allows the decoupling of *long-distance* contributions, non-perturbative and described by the operators $O_i(\lambda)$, and *short-distance* contributions, included in Wilson coefficients and computed in perturbation theory.

The formalism introduced in this section is known as *Operator Product Expansion* [27]. It is particularly convenient for weak decays since the mass of the decaying particle sets the scale. At the same time, Wilson coefficients are by construction independent of the underlying process and sensitive to physics beyond the Standard Model. This makes the framework a powerful tool to combine results from different measurements and to test extensions of the Standard Model.

2.7. $B^{0,+} \to K^{0,+}\mu^+\mu^-$ decays

The analysis presented in this thesis focuses on the rare decays $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0\mu^+\mu^-$. They are part of the large class of $b \to s\ell^+\ell^-$ processes, both theoretically and experimentally extensively studied. As these are *flavour changing neutral current* processes, the coupling between b and s quarks is not possible at tree level, but happens at leading order via "penguin" or "box" diagrams, as shown in Fig. 2.5. This makes the transition highly suppressed and therefore particularly sensitive to precise tests of the Standard Model. Since the associated amplitudes are dependent on the momentum transferred to the lepton pair, it is usual to perform these tests as a function of the dilepton invariant mass squared, denoted as q^2.



(a) Tree: forbidden (b) Electroweak Penguin (c) Box Diagram

Figure 2.5.: Leading-order Feynman diagrams for $b \longrightarrow s\ell\ell$ processes.

The general class manifests as the decays of interest through the addition of a light spectator quark, d in the case of $B^0 \longrightarrow K^0_\mathrm{S}\mu^+\mu^-$, u for $B^+ \longrightarrow K^+\mu^+\mu^-$. The framework developed in the previous section provides the theoretical basis for their description. The Wilson coefficients C_7, C_9 and C_{10} associated with the operators

$$O_7 = \frac{e}{16\pi^2} m_b \bar{s}\sigma^{\mu\nu} P_\mathrm{R} b F_{\mu\nu}\,, \quad O_9 = (\bar{s}\gamma_\mu P_\mathrm{L} b)(\bar{\mu}\gamma^\mu\mu)\,, \quad O_{10} = (\bar{s}\gamma_\mu P_\mathrm{L} b)(\bar{\mu}\gamma^\mu\gamma_5\mu)\,, \tag{2.18}$$

where $P_\mathrm{R} = (1+\gamma_5)/2$ and $P_\mathrm{L} = (1-\gamma_5)/2$ are the chiral projectors, are responsible for the largest contributions to the branching fraction. The first operator, O_7, also called magnetic operator, represents the coupling of the two quarks with a photon and reflects the electromagnetic penguin diagram of Fig. 2.5b, with a photon emitted from the loop. Explicitly, the antistrange and bottom quark fields, $\bar{s}$ and b, are coupled to the electromagnetic field $F_{\mu\nu}$ via the $\sigma^{\mu\nu}$ tensor, which describes the spin structure of the interaction. The remaining two operators, O_9 and O_{10}, describe the coupling to a vector and axial current respectively, and originate from weak penguin and box diagrams of Figs. 2.5b and 2.5c. This interpretation helps in visualising the q^2 dependence of the Wilson coefficients. The emission of a massive boson is more likely at high energies, consequently C_9 and C_{10} become increasingly relevant at large q^2. In contrast, at low q^2, photon emission dominates. C_7 enhancement, however, is limited by polarization effects.

It is worth adding that to any operator in (2.18) there corresponds a primed one O_7', O_9' and O_{10}', where the projectors P_R and P_L are interchanged. These also contribute to the branching fraction, although their Wilson coefficients are suppressed by the factor m_s/m_b in the Standard Model. A complete computation can be found in Ref. [28], and the resulting Standard Model prediction for the differential branching fraction of $\bar{B}^0 \longrightarrow \bar{K}^0\mu^+\mu^-$ is shown in Fig. 2.6. A similar shape is obtained for $B^+ \longrightarrow K^+\mu^+\mu^-$, however the different spectator quark is introducing differences that will be the subject of the next section.

2.8. Isospin asymmetry

The only difference between the $B^0 \longrightarrow K^0_\mathrm{S}\mu^+\mu^-$ and $B^+ \longrightarrow K^+\mu^+\mu^-$ decays lies in the spectator quark. Given the absence of interactions involving the spectator quark in the initial or final state, the leading-order diagrams of Fig. 2.5 are insensitive to this difference. Consequently, the main contributions to the branching fractions of both decay channels are identical. When considering couplings between the spectator quark and the flavour-changing diagrams, however, the different charges of the spectator quarks

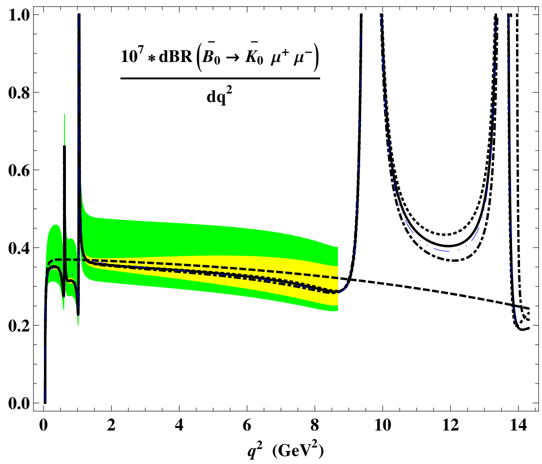
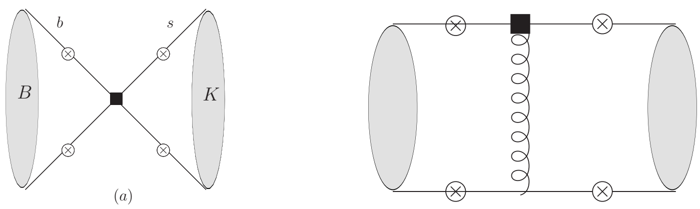


Figure 2.6.: Differential branching fraction of $\bar{B}^0 \longrightarrow \bar{K}^0\mu^+\mu^-$ as reported in Ref. [28].



(a) $O^q_{1,2}$ (b) O_8

Figure 2.7.: Diagrams for $b \longrightarrow s\ell\ell$ involving the Wilson coefficients $C_{1,2}$ (a) and C_8 (b), reproduced from Ref. [28]. The crosses denote where a photon can be radiated to produce the dilepton pair.

become relevant, and electromagnetic effects determine a difference between the two channels. To account for such corrections, the main contributing operators of (2.18) are no longer sufficient, but the quark-quark operators O_1-O_6 and O_8

$$
\begin{aligned}
O_1^q &= (\bar{s}_i\gamma^\mu P_\mathrm{R} q_j)(\bar{q}_j\gamma_\mu P_\mathrm{R} b_i)\,, \quad O_2^q = (\bar{s}_i\gamma^\mu P_\mathrm{R} q_i)(\bar{q}_j\gamma_\mu P_\mathrm{R} b_j)\,,\\
O_3 &= (\bar{s}_i\gamma^\mu P_\mathrm{R} b_i)\sum_q(\bar{q}_j\gamma_\mu P_\mathrm{R} q_j)\,, \quad O_4 = (\bar{s}_i\gamma^\mu P_\mathrm{R} b_j)\sum_q(\bar{q}_j\gamma_\mu P_\mathrm{R} q_i)\,,\\
O_5 &= (\bar{s}_i\gamma^\mu P_\mathrm{R} b_i)\sum_q(\bar{q}_j\gamma_\mu P_\mathrm{L} q_j)\,, \quad O_6 = (\bar{s}_i\gamma^\mu P_\mathrm{R} b_j)\sum_q(\bar{q}_j\gamma_\mu P_\mathrm{L} q_i)\,,\\
O_8 &= -\frac{g_s}{16\pi^2} m_b \bar{s}_i \sigma^{\mu\nu} P_R T^a_{ij} b_j G^a_{\mu\nu}
\end{aligned}
\tag{2.19}
$$

must be added, with $q = (u, d)$ depending on the specific decay. Unlike penguin and box diagrams, in the diagrams arising from Eq. (2.19) the lepton pair is not produced directly from a flavour-changing loop, and a photon must be radiated from a quark line, as illustrated in Fig. 2.7.

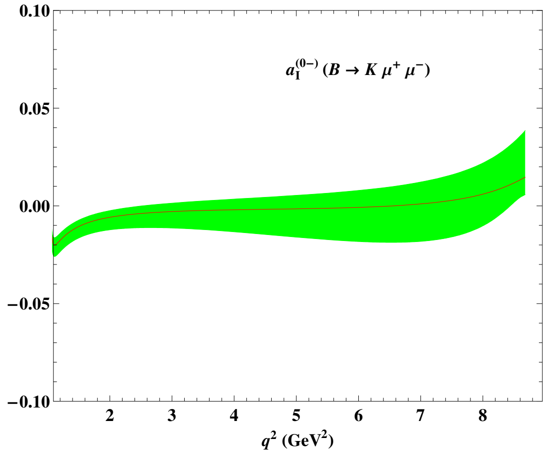


Figure 2.8.: Isospin asymmetry A_{I} in $B \longrightarrow K\mu^+\mu^-$ decays as reported in Ref. [28].

A full computation yields a plot analogous to Fig. 2.6 for $B^+ \longrightarrow K^+\mu^+\mu^-$. Both channels suffer from large uncertainties due to the limited precision of the form factor estimates. A powerful strategy to reduce this effect is shifting the focus from the absolute branching fractions to the differences between the two channels, building an observable where the leading form factor uncertainties cancel out through subtractions and ratios. One such observable is the *isospin asymmetry*

$$\begin{aligned} A_{\mathrm{I}} &= \frac{\Gamma(B^0 \longrightarrow K^0\mu^+\mu^-) - \Gamma(B^+ \longrightarrow K^+\mu^+\mu^-)}{\Gamma(B^0 \longrightarrow K^0\mu^+\mu^-) + \Gamma(B^+ \longrightarrow K^+\mu^+\mu^-)} \\ &= \frac{\mathcal{B}(B^0 \longrightarrow K^0\mu^+\mu^-) - \frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}(B^+ \longrightarrow K^+\mu^+\mu^-)}{\mathcal{B}(B^0 \longrightarrow K^0\mu^+\mu^-) + \frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}(B^+ \longrightarrow K^+\mu^+\mu^-)} \end{aligned} \tag{2.20}$$

which embodies the differences in decay widths Γ, and hence in branching fractions $\mathcal{B}$, once B meson lifetimes τ_B are accounted for. As it is shown in Fig. 2.8 the Standard Model expectation for A_{I} is of the order of 10^{-2} in the q^2 region below the J/ψ resonance [28, 29]. At larger q^2 there is no precise prediction, although the asymmetry is expected to be smaller there, since the photon emission in the diagrams of Fig. 2.7 becomes less probable at higher momentum transfer.

This asymmetry and the differential branching fractions of the $B^0 \longrightarrow K^0\mu^+\mu^-$ and $B^+ \longrightarrow K^+\mu^+\mu^-$ decays represent the main observables measured in this thesis.

3. The LHCb experiment

Current particle physics experiments are performed using particle accelerators, large machines where particles are collided at high energy to study the resulting products. The *Large Hadron Collider* (LHC) is currently the world's largest and most powerful particle accelerator, hosting four main experiments, including LHCb. The LHCb experiment is designed to study flavour physics in beauty and charm hadron decays. The work described in this thesis is based on data collected from proton-proton collisions at LHCb. A brief overview of the CERN accelerator complex and a more detailed description of the LHCb detector are given in this chapter, focusing on the aspects of major interest for the presented analysis.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [30] is a synchrotron located at CERN (Conseil Européen pour la Recherche Nucléaire) near Geneva. It is designed to collide protons or heavy ions using two counter-rotating beams circulating in separate vacuum pipes within a $27\,\mathrm{km}$ underground tunnel, crossing at four interaction points. The beams are organised in bunches of particles accelerated by radiofrequency cavities, bent by superconducting dipole magnets and focused by quadrupole magnets. The frequency of the cavities fixes the time spacing between the bunches to $25\,\mathrm{ns}$. Conversely, the bending magnets, producing a field of $8\,\mathrm{T}$, limit the proton beam energy to $7\,\mathrm{TeV}$, resulting in a centre-of-mass energy of $14\,\mathrm{TeV}$. These beam parameters are designed to achieve a peak instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Due to the limited dynamic range of the magnets, it would be impossible to directly pass from rest to such a large energy. For this reason, a series of increasingly larger machines precedes the LHC. Protons are initially produced at the linear accelerator Linac2 and injected into the Proton Synchrotron Booster (PSB) with an energy of $50\,\mathrm{MeV}$. Here, protons are held until they reach the energy of $1.4\,\mathrm{GeV}$, before being transferred to the Proton Synchrotron (PS), where they are accelerated to an energy of $26\,\mathrm{GeV}$. At this stage, protons are collected into bunches of 10^{11} particles and prepared for the Super Proton Synchrotron (SPS), which is the last step before entering the collider. The energy is further increased to $450\,\mathrm{GeV}$, and they are finally injected into the LHC in opposite directions via two transfer lines to collide head-on at the four interaction points. Collectively, the machines described above form the CERN accelerator complex, which serves various experiments, as shown in Fig. 3.1.

The four main experiments are installed at the interaction points of the LHC. ATLAS [32] and CMS [33] (Compact Muon Solenoid) are general-purpose detectors designed for Higgs physics and direct searches beyond the Standard Model. ALICE [34] (A Large Ion Collider Experiment) is specifically designed to study heavy ion collisions, and in particular to characterise the so-called *quark-gluon plasma*, the very dense state of the universe few moments after its creation. In contrast, LHCb [35] (Large Hadron Collider beauty) is dedicated to the study of flavour physics in beauty and charm hadron decays and to precision tests of the Standard Model. This will be described in detail in Sec. 3.2.

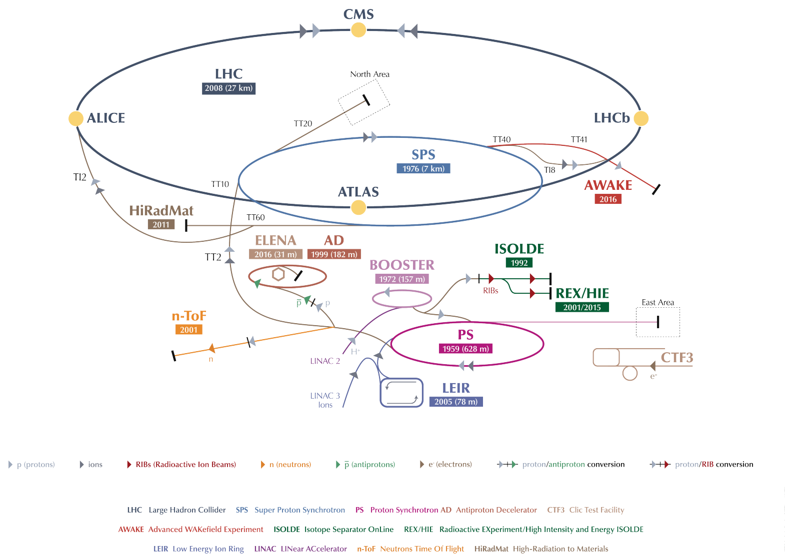


Figure 3.1.: The CERN accelerator complex as in 2018 at the end of the so-called Run 2 period [31].

Data-taking campaigns at the LHC are organised into operational periods known as *Runs*, separated by Long Shutdowns (LS) for maintenance and upgrades. The first period, Run 1, lasted from 2010[1] to 2012. During the years 2010 and 2011 the centre-of-mass energy was $7\,\mathrm{TeV}$, while in 2012 it was increased to $8\,\mathrm{TeV}$. Following the first Long Shutdown (LS1), Run 2 covered the years 2015–2018 with a collision energy of $13\,\mathrm{TeV}$. The subsequent LS2 (2019–2022) was used for major upgrades, particularly for the LHCb detector, marking the transition to a new experimental phase. The analysis presented in this thesis is based on the dataset collected during Run 1 and Run 2.

3.2. The LHCb detector

The LHCb detector [35] is a single-arm forward spectrometer originally designed to study decays of b and c hadrons produced in proton-proton collisions at the LHC. LHCb is unique among the LHC experiments as it operates in the forward region, a choice motivated by the kinematics of heavy-flavour production. At high energies, $b\bar{b}$ pairs are predominantly produced at low angles with respect to the beam axis, driven by the significant longitudinal boost of the partonic system, as shown in Fig. 3.2. With a coverage of $10\,\mathrm{mrad}$ to $300\,\mathrm{mrad}$ in the bending plane, and to $250\,\mathrm{mrad}$ in the non-bending plane, which translates into a pseudorapidity acceptance of $2 < \eta < 5$, the LHCb detector intercepts around 25% of all produced B mesons. This design choice maximises the efficiency in terms of occupied volume and material costs. A schematic view of the detector is presented in Fig. 3.3, where the coordinate system in use at LHCb is introduced. The z-axis is along the beam pipe, the y-axis is perpendicular to the ground and the x-axis, horizontal, lies in the bending plane of the magnet. To accommodate the forward particle flux, the detector adopts a conical geometry: it is compact near the proton-proton interaction point and expands

[1] Although formally part of Run 1, 2010 is considered as a commissioning year, and data collected during this period are not used for physics analyses.

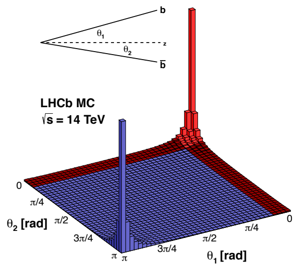


Figure 3.2.: Polar angular distribution of the $b\bar{b}$ pairs produced at LHC for $\sqrt{s} = 14\,\mathrm{TeV}$ as simulated with the PYTHIA8 event generator [36]. In red the region covered by the LHCb detector.

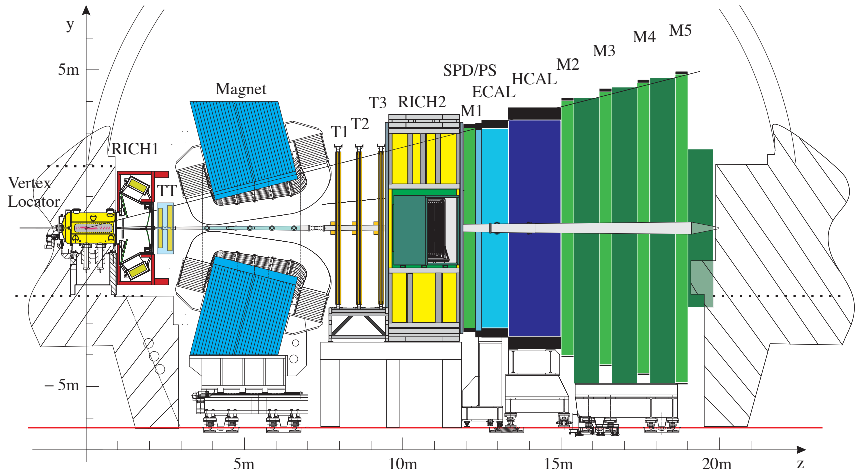


Figure 3.3.: The LHCb detector as in 2018 at the end of the Run 2 period [35].

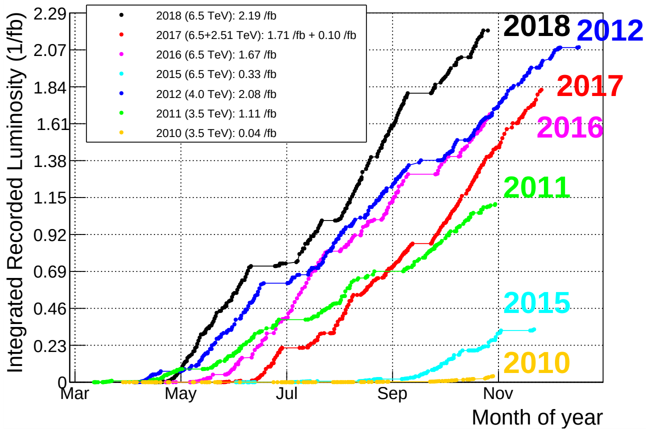


Figure 3.4.: Integrated luminosity recorded at LHCb.

transversely as the distance from the vertex increases. Similarly, the detector granularity varies with the distance from the beam axis to maintain a uniform occupancy across the active regions. The subdetectors can be broadly categorised based on their primary function. The *tracking system* provides precise measurements of the trajectories and momenta of charged particles traversing the detector. Complementary to this, the *particle identification system* is dedicated to determining the identity of each particle type.

It is worth noting that LHCb faces different challenges compared to B-factories, electron-positron colliders operating at the $\Upsilon(4S)$ resonance to maximise the production of B mesons. While lepton collisions offer a cleaner environment and constrained kinematics, the hadronic environment at the LHC is dominated by the QCD background. This makes event reconstruction more complicated, resulting in a smaller b-tagging efficiency at LHCb. However, these difficulties are counterbalanced by the large $b\bar{b}$ production cross section and the high collision energy, which allows for the production of b-hadrons heavier than the B_s^0. The production rate is sufficiently high that handling the full instantaneous luminosity provided by the LHC would be challenging for the detector occupancy and trigger bandwidth. To mitigate this, the beams are shifted with respect to each other so that the active overlap region is reduced and kept constant throughout the run. The procedure is known as *luminosity levelling*. The luminosity recorded at LHCb during each year of the Run 1 and Run 2 periods is reported in Fig. 3.4.

3.2.1. Tracking System

As introduced above, the tracking system is responsible for reconstructing the trajectories of charged particles. By combining high-precision position measurements upstream and downstream of the dipole magnet, it determines the track curvature, which yields the particle momentum. The system consists of the Vertex Locator (VELO), the Tracker Turicensis (TT) and the three Tracking Stations (T1–T3). The VELO surrounds the proton-proton interaction point and is aimed at reconstructing primary and secondary vertices, where long-lived particles like B mesons decay. The tracking continues with the TT and the T1–T3 subsystem, four planar tracking stations of different technologies placed upstream and downstream of the magnet, respectively.

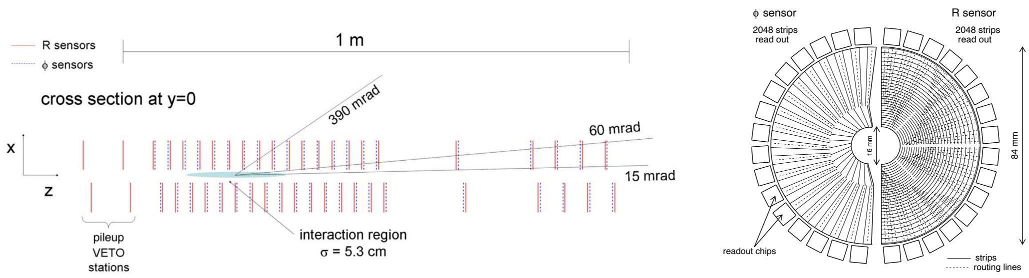


Figure 3.5.: Arrangement of the VELO stations along the beam axis [35] (left) and one half-disk of the VELO [37] (right).

3.2.1.1. Vertex Locator

The Vertex Locator (VELO) [37] is the closest detector to the beam and is designed to locate and reconstruct primary vertices, defined as the points where the two protons collide, and secondary vertices where long-lived particles decay. It is composed of a series of 42 silicon modules located perpendicular to the beam axis according to the scheme given in Fig. 3.5. The modules are shaped as semi-circles with an outer radius of $42\,\mathrm{mm}$, an inner radius of $7\,\mathrm{mm}$, and thickness of $300\,\mathrm{\mu m}$. Each of them provides a measurement of the polar coordinates, radius (r) and azimuthal angle (ϕ), on the module plane. The three-dimensional reconstruction is then completed by knowing the z-position of the module itself. The measurement is provided by single-sided strip sensors arranged circularly (R-sensors) or radially (ϕ-sensors) as shown in Fig. 3.5. The silicon strip technology enables a spatial resolution of up to $4\,\mathrm{\mu m}$. In order to have a good primary vertex resolution, the sensors should be as close as possible to the beam. For this reason, the sensitive area of the VELO sensors begins at only $7\,\mathrm{mm}$ from the beam, and the whole detector operates inside a vacuum vessel separated from the LHC vacuum by a corrugated aluminium foil, known as the *RF-foil*, of $300\,\mathrm{\mu m}$ thickness. However, such a short distance would be hazardous during the injection of the LHC, when the beam is unstable and its aperture is significantly larger. To avoid damage, the VELO is installed in a movable system which can retract to a safety distance of $3\,\mathrm{cm}$ during injection, and move back to the operational distance when the beam is stable.

3.2.1.2. Tracker Turicensis

The Tracker Turicensis [38] (TT) is a tracking detector located immediately before the magnet. Its presence is crucial to detect neutral long-lived particles (e.g. K^0_S, Λ) decaying downstream of the VELO, and to reconstruct low-momentum charged particles that are bent out of the detector acceptance by the magnetic field before reaching the subsequent stations. The TT is made up of four rectangular layers of silicon strips, about $150\,\mathrm{cm}$ wide and $130\,\mathrm{cm}$ high, arranged into two stations, TTa and TTb, as described in Fig. 3.6. While strips in the first and fourth layers are oriented vertically, those in the second and third layers are rotated by $\pm 5^\circ$ to allow a complete reconstruction in the xy-plane. The single sensors are rectangular silicon tiles of dimension $9.64\,\mathrm{cm} \times 9.44\,\mathrm{cm}$ and $500\,\mathrm{\mu m}$ thickness, carrying 512 readout strips with a pitch of $183\,\mathrm{\mu m}$. This results in a nominal spatial resolution of $50\,\mathrm{\mu m}$.

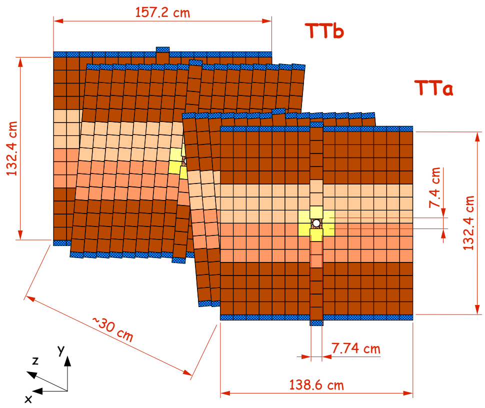


Figure 3.6.: Layout of the four TT detection layers [39].

3.2.1.3. Inner Tracker

A technology similar to that of the Tracker Turicensis is employed downstream of the magnet for the inner region of the tracking stations T1–T3, covering the high-occupancy area closest to the beam pipe. The Inner Tracker [38] (IT) is a silicon detector consisting of four layers with the same stereo angle arrangement as described for the TT in Fig. 3.6. This four-layer structure is repeated in each of the three tracking stations. Unlike the TT, the sensors in the IT are arranged in a cross-shaped layout around the beam pipe, as shown in Fig. 3.7. Each layer consists of four separate boxes of silicon sensors to accommodate the beam pipe passing through the centre. Each box houses seven silicon modules using the same technology as the TT: silicon microstrip sensors with a strip pitch of $200\,\mu\text{m}$. Similar to the TT, the resulting spatial resolution is about $50\,\mu\text{m}$.

3.2.1.4. Outer Tracker

While silicon technology is required for the regions close to the beam where the high particle flux density imposes tight constraints on occupancy, for the external parts of the tracking stations T1–T3 the smaller flux allows for relaxed requirements. Consequently, the area surrounding the IT, approximately $5\,\text{m} \times 6\,\text{m}$, is covered by a straw drift tube detector called Outer Tracker [40] (OT). The standard arrangement of four layers per tracking station is implemented using modules containing two staggered rows of straw tubes. For each module there are 64 tubes, consisting of cylinders with an inner diameter of $4.9\,\text{mm}$ acting as cathodes, enclosing a central anode wire and filled with a gas mixture of Ar, CO_2 and O_2. The mixture is chosen to guarantee a maximum drift time of $50\,\text{ns}$. Although this value exceeds the LHC bunch spacing of $25\,\text{ns}$, it is sufficiently short to limit the signal spillover to acceptable levels for reconstruction. The resulting spatial hit position resolution is approximately $200\,\mu\text{m}$.

3.2.1.5. The magnet

The LHCb dipole magnet [42] is located between the TT and the tracking stations T1–T3, and provides the magnetic field necessary to bend the trajectories of charged particles traversing the detector, enabling

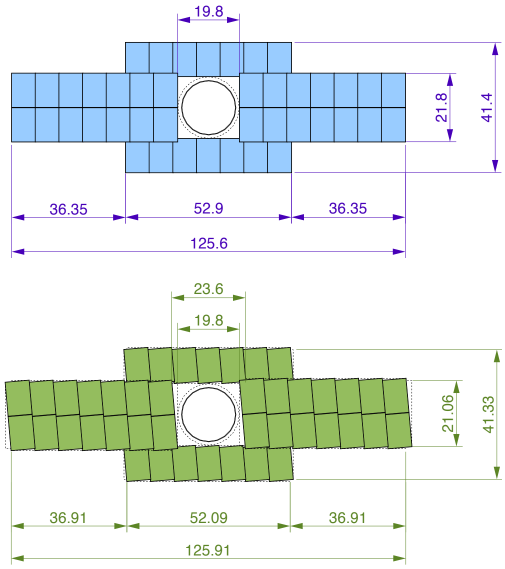


Figure 3.7.: Layout of two layers of the Inner Tracker, one vertical and one rotated as described in the text [39].

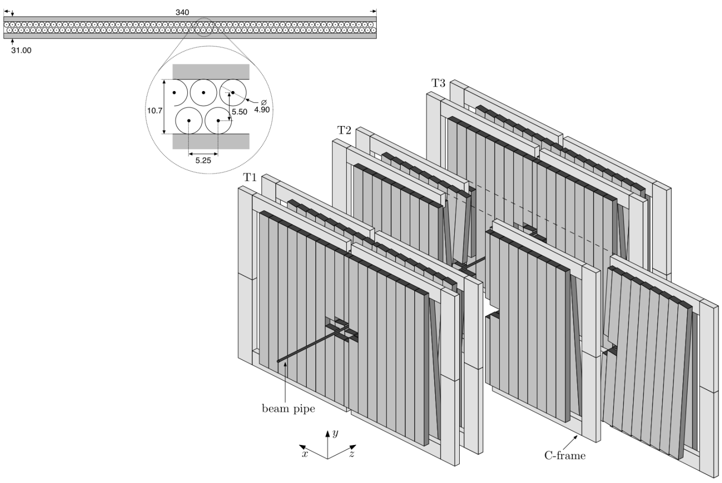


Figure 3.8.: Layout of the Outer Tracker with a view of the module cross section [41].

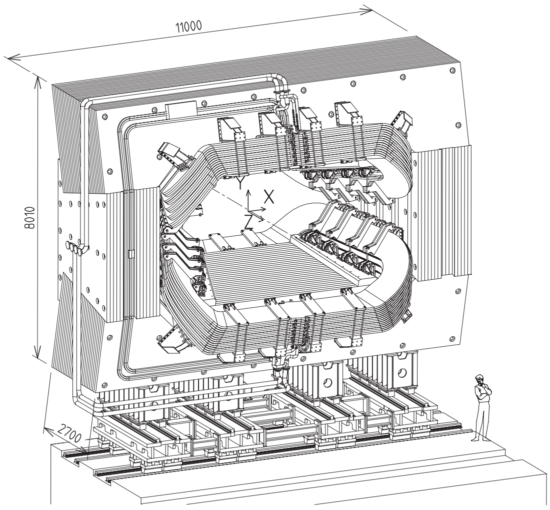


Figure 3.9.: Perspective view of the LHCb magnet [42].

the measurement of their momentum. The magnet, depicted in Fig. 3.9, consists of two warm (non-superconducting) coils inside a window-frame iron yoke. The saddle shape produces a magnetic field primarily aligned along the vertical y-axis. The coils are designed to minimise inhomogeneities across the acceptance region while ensuring the field drops rapidly to zero outside.

Momentum is determined by measuring the angular deflection of tracks between the detectors upstream (VELO, TT) and downstream (T1–T3) of the magnet. Consequently, momentum resolution is proportional to the angular resolution, which in turn is directly related to the spatial resolution of the tracking detectors described previously. Moreover, a strong magnetic field is essential to ensure a measurable deflection even for high-momentum particles. Therefore, the precision of the measurement also depends on the intensity of the field. The bending power of the magnet is characterised by the field integral $\int B\,\mathrm{d}l$. With a field integral of $4\,\mathrm{T\,m}$ and the tracking performance reported previously, the resulting momentum resolution at LHCb varies from $\delta p/p = 0.5\,\%$ for low-momentum particles up to $20\,\mathrm{GeV/c}$, to $\delta p/p = 0.8\,\%$ for momenta of about $100\,\mathrm{GeV/c}$ [43].

Particles with opposite charges are bent in opposite directions. To control systematic uncertainties related to detection asymmetries, the polarity of the magnet is periodically reversed during data taking.

3.2.2. Particle Identification System

Once the kinematics is reconstructed, determining the nature of the particles is crucial for background rejection and signal selection. The particle identification system relies on specific technologies to separate charged hadrons (specifically kaons from pions), identify muons, and measure the energy of the particles. It is composed of two Ring Imaging Cherenkov (RICH) detectors, the calorimeter system (ECAL and HCAL), and the Muon System.

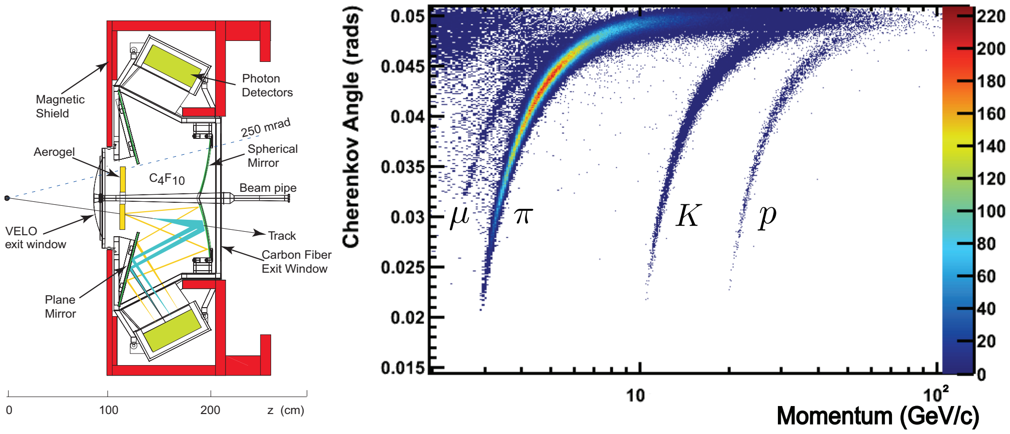


Figure 3.10.: Side view of the RICH1 detector [46] (left) and reconstructed Cherenkov angle as a function of track momentum in RICH1 [47] (right).

3.2.2.1. Ring Imaging Cherenkov detectors

When a charged particle traverses a medium with a velocity higher than the local phase velocity of light, it emits Cherenkov radiation [44]. The photons are emitted in a cone around the direction of the particle, and the angle of emission θ_C is related to the particle velocity βc and the refractive index of the medium n via the formula

$$\cos\theta_C = \frac{1}{n\beta} \ .$$

A measurement of the Cherenkov angle translates then into a measurement of the velocity of the particle. When combining this information with the momentum reconstructed by the tracking system, the mass of the particle can be determined, allowing for particle type discrimination. At LHCb, this is done with the two Ring Imaging Cherenkov detectors RICH1 and RICH2 [45] which are able to reconstruct the emission cone of a particle traversing gaseous C_4F_{10} (RICH1) or CF_4 from their projection on the detecting surface. More specifically, emitted photons are focused by spherical mirrors and are reflected out of the interaction area through flat mirrors. In an area safe from radiation and residual magnetic field, Hybrid Photon Detectors are located to detect the projected rings, as shown in Fig. 3.10.

The choice of two distinct detectors is driven by the need to cover a wide momentum range using gas radiators with different refractive indices. RICH1, located upstream of the magnet, is designed to cover a lower momentum range, namely 1–$60\,\mathrm{GeV}/c$, and it needs a radiator with a refractive index larger than the one used for RICH2, which on the other hand covers the range 15–$100\,\mathrm{GeV}/c$. Particle identification relies on two physical regimes. First, Cherenkov emission is a threshold effect: only particles with $\beta > 1/n$ emit light. This provides excellent discrimination at low momentum. Second, for particles well above the threshold, the angle θ_C approaches an asymptotic value ($\arccos(1/n)$). In this regime, the bands for different particle hypotheses (as shown in Fig. 3.10) become closer, making discrimination increasingly challenging at very high momenta.

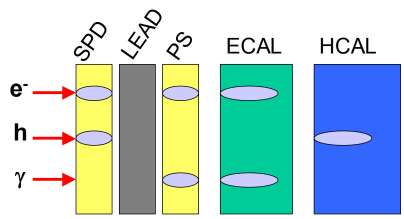


Figure 3.11.: Signal deposited on the different parts of the calorimeter by an electron, a hadron, and a photon [49]

3.2.2.2. The Calorimeters

The LHCb calorimeter system [48] determines the energy of particles and complements the particle identification of the RICH detectors. Furthermore, it provides fast information used by the hardware-level trigger. It is composed of a Scintillating Pad Detector (SPD), a lead converter, a Preshower (PS), an Electromagnetic Calorimeter (ECAL) and a Hadronic Calorimeter (HCAL), arranged longitudinally in sequence. Despite specific differences in design, all components operate on the same principle: scintillation light produced by particles traversing the active material is collected by wavelength-shifting fibres and transmitted to photomultipliers. The longitudinal segmentation is key to particle identification, relying on the different interactions of particles with matter as illustrated in Fig. 3.11.

The SPD identifies charged particles, allowing for the initial discrimination between neutral particles, which leave no signal, and charged ones. However, since charged hadrons also generate a signal in the scintillator, they are indistinguishable from electrons at this stage. To resolve this, a lead converter placed between the SPD and the PS initiates electromagnetic showers for electrons and photons. Hadrons, in contrast, traverse this layer without showering, as its thickness is negligible compared to the nuclear interaction length. Consequently, the PS, detecting the showers, can distinguish electrons and photons from hadrons. The combined response of the two detectors defines the truth table for particle identification depicted in Fig. 3.11: a signal in both the SPD and the PS identifies electrons, a shower in the PS without a signal in the SPD indicates photons, while charged hadrons induce a signal in the SPD but do not shower in the PS.

The SPD and the PS share an identical layout of rectangular scintillator pads and are often collectively referred to as the SPD/PS detector. To match the decreasing hit density at larger distance from the beam, both planes are divided into three regions of cells of increasing size, as shown in Fig. 3.12. Each cell consists of a $15\,\mathrm{mm}$ thick polystyrene-based scintillator pad, which incorporates a coiled fibre to collect the light.

Downstream of the PS, the ECAL provides a precise energy measurement for photons and electrons. It is a sampling calorimeter based on *shashlik* technology, constructed from stacks of $4\,\mathrm{mm}$ thick polystyrene-based scintillator tiles alternated with $2\,\mathrm{mm}$ sampling lead sheets. The scintillating layers are organised with a layout analogous to that of the SPD/PS shown in Fig. 3.12. The total depth corresponds

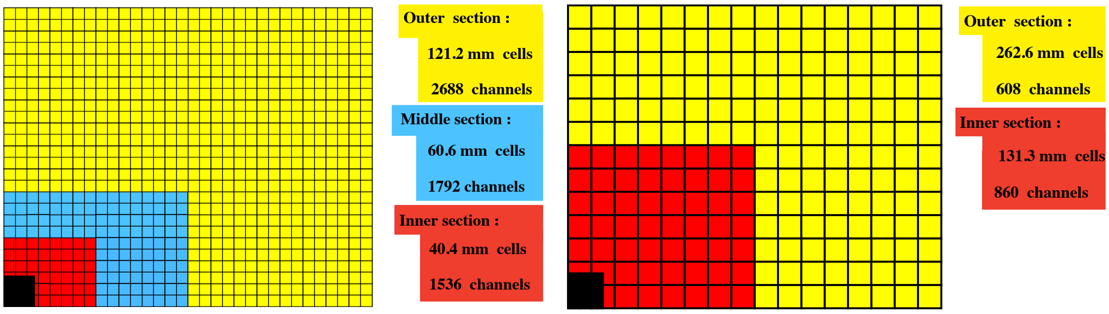


Figure 3.12.: Layout of the planes of SPD/PS, ECAL (left) and HCAL (right) [48].

to 25 radiation lengths, ensuring the full containment of electromagnetic showers for optimal energy resolution. The resulting energy resolution is

$$\frac{\sigma_E}{E} = \frac{10\,\%}{\sqrt{E/\mathrm{GeV}}} \oplus 1\,\% \ .$$

Finally, the HCAL measures the energy of hadrons. Reflecting the larger dimension of hadronic showers compared to electromagnetic ones, its granularity is coarser than that of the ECAL and is divided into only two regions, as shown in Fig. 3.12. The module structure consists of $4\,\mathrm{mm}$ scintillator tiles alternated with $16\,\mathrm{mm}$ iron absorbers. The total thickness of $1.2\,\mathrm{m}$ corresponds to 5.6 interaction lengths. Since this does not guarantee full containment of high-energy hadronic showers, the energy resolution is limited to

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E/\mathrm{GeV}}} \oplus (9 \pm 2)\% \ .$$

3.2.2.3. The Muon System

The last subdetector of LHCb is the muon system [50]. Since muons are the only charged particles likely to traverse the entire calorimeter system without being absorbed, hits in this detector provide a robust signature for particle identification. The system consists of 5 tracking stations called M1–M5. The first station, M1, is located upstream of the calorimeters to assist in the momentum measurement for the hardware trigger. This configuration is chosen to mitigate the resolution degradation caused by multiple scattering in the calorimeter system. The remaining stations, M2–M5, are located downstream of the HCAL and serve primarily for muon identification. To absorb hadronic particles that might punch through the calorimeters, $80\,\mathrm{cm}$ thick iron absorbers are interleaved between stations M2–M5. The five detection planes consist of rectangular chambers arranged in a projective geometry, i.e. their transverse dimensions scale with z, increasing from $3.7\,\mathrm{m} \times 3\,\mathrm{m}$ for M1 to $10\,\mathrm{m} \times 9\,\mathrm{m}$ for M5. Each plane is divided into four regions, R1–R4, with increasing granularity towards the beam axis, similar to the calorimeters. With the exception of the innermost region (R1) of station M1, which employs triple Gaseous Electron Multipliers (GEMs) due to the high radiation environment, all regions are composed of Multi-Wire Proportional Chambers (MWPCs).

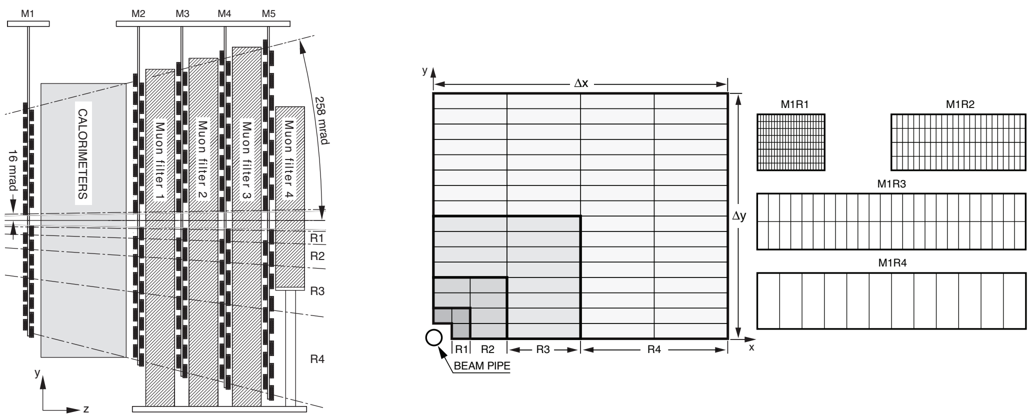


Figure 3.13.: Side view of the muon system (left) and granularity of a single station (right) [50].

3.3. The LHCb trigger system

LHCb operates in a high-rate environment dominated by hadronic background processes. Not all events contain interesting information to be kept for analysis. Therefore, the LHCb experiment employs a trigger system designed to lower the LHC rate and save important data for offline use. The LHCb trigger [51] consists of three stages, a first hardware level and two subsequent software levels, which progressively reconstruct the event and reduce the rate to storage.

The Level-0 trigger (`L0`) is a hardware trigger implemented on custom electronics. It uses information from the calorimeter and muon systems to reduce the bunch crossing rate of $40\,\mathrm{MHz}$, provided by the LHC, to $1\,\mathrm{MHz}$. At these rates, a decision must be made so quickly that a full track reconstruction is not feasible. Consequently, the momentum of muons is estimated using only the hits in the five muon stations, with a resolution of roughly $25\,\%$. The trigger selects events containing muon candidates with large transverse momentum (p_T), which are in general associated with the decays of b-hadrons. For other particles, since tracking information is not available, momentum cannot be determined. Instead, the calorimeter system selects high transverse energy (E_T) deposits in the ECAL and HCAL, indicating the presence of electrons, photons and hadrons. The `L0` output is processed by the High Level Trigger (`HLT`), which, unlike `L0`, is software-based. The first stage, `HLT1`, performs a partial reconstruction to reduce the rate further by an order of magnitude, from $1\,\mathrm{MHz}$ to $110\,\mathrm{kHz}$. At this stage, only tracks with significant transverse momentum (p_T $>500\,\mathrm{MeV/c}$) are reconstructed. Hits in the VELO are fitted to determine the primary vertex and tracks are extrapolated to the tracking stations. Finally, a full event reconstruction is performed at the second software stage `HLT2`. The reduced input rate allows for a detailed analysis including low-momentum and long-lived particle tracks, and full particle identification using RICH information.

An overview of the three decision levels is given in Fig. 3.14 for Run 2. The scheme includes the additional alignment and calibration steps, which were not present in Run 1. During Run 2, data selected from HLT1 were buffered and used to align and calibrate the detector in real-time. Alignment determines the relative positions of the subdetectors, while the calibration refers to the timing of the OT, the refractive index in the RICH detectors and the energy deposition in the calorimeters. Applying these corrections

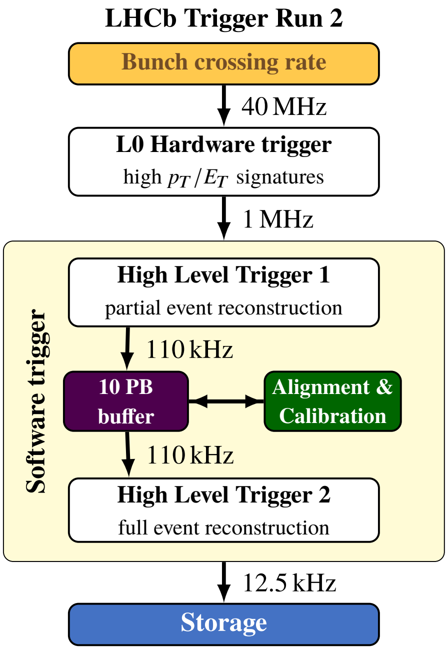


Figure 3.14.: Overview of the LHCb trigger system in use during Run 2 [52]. During Run 1 the $10\,\mathrm{PB}$ buffer was not present and alignment and calibration were not performed online in between HLT1 and HLT2; in that case the rate to storage was $5\,\mathrm{kHz}$.

during the data taking maximises the selection efficiency and ensures the quality of the reconstruction already at the trigger level, thereby improving the quality of the data for the offline analysis.

Fully reconstructed events are finally saved to storage at a rate of $5\,\mathrm{kHz}$ in Run 1 and $12.5\,\mathrm{kHz}$ in Run 2. The selection is performed by specific algorithms known as `HLT2` *lines*, each targeting a specific decay topology or physics process. Depending on the information retained, these lines are grouped into output streams characterised by different degrees of persistence. In the *Full Stream* the full event is stored, including raw data banks used for offline reconstruction. In contrast, the *Turbo Stream* saves only the high-level information of the reconstructed decay candidates. This allows storing data in a reduced format and saving space in those cases where no additional reconstruction is needed. Finally, the *Calib Stream* represents a hybrid approach, storing high-level information for the reconstructed candidates along with specific raw data banks required for offline calibration.

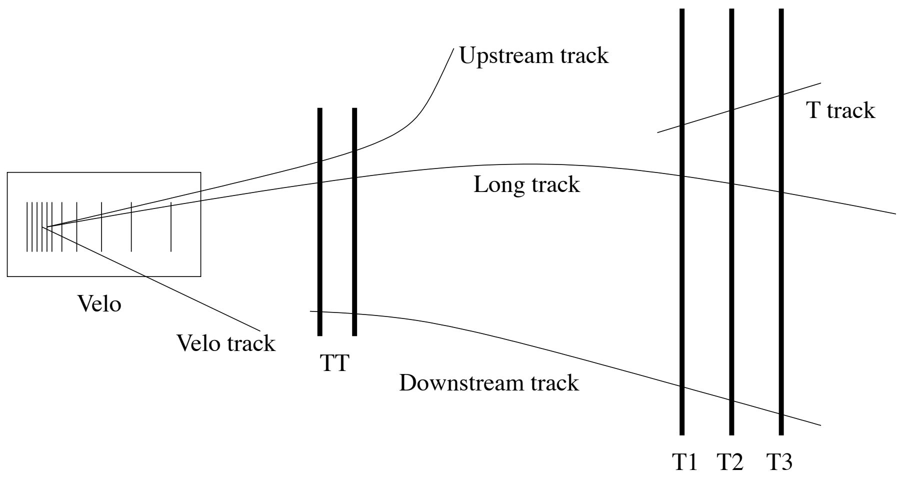


Figure 3.15.: Different types of tracks reconstructed at LHCb [35].

3.4. Track reconstruction at LHCb

As discussed in the previous section, tracks are reconstructed by identifying and fitting hits in the subdetectors into trajectories and connecting the results from the various components. Ideally, a track is completely reconstructed from the VELO to the final tracking stations. However, due to geometrical acceptance, in-flight decays or neutral particles in the decay chain, this is not always possible. Depending on the subdetectors involved, different track types are defined:

- **Long tracks** traverse the full tracking system, providing the most precise momentum determination.
- **Downstream tracks** are formed with hits in the TT and T1-T3 stations. They are typically associated with long-lived particles (such as K^0_S and Λ) decaying outside of the VELO. Although tracking information before and after the magnet guarantees a good momentum resolution, this is lower than that of long tracks due to the lack of precise information from the VELO.
- **VELO tracks** consist of hits in the VELO modules only. They usually correspond to particles that fall outside the detector acceptance, i.e. at large angles with respect to the beam axis or backward.
- **Upstream tracks** are formed with hits in the VELO and TT stations. These usually correspond to low-momentum particles bent out of the detector acceptance by the magnetic field before reaching the tracking stations.
- **T tracks** are formed with hits in the downstream tracking stations T1-T3 only. Due to the lack of information upstream of the magnet, momentum reconstruction is not possible.

A visual description of the track types is given in Fig. 3.15. The analysis presented in this thesis is based on long and downstream tracks only. They will be referred to in the following as L- and D-type respectively.

3.5. LHCb simulation

Alongside data, modern physics analyses make extensive use of simulated samples for background estimation, signal selection optimisation and detailed detector studies. At LHCb, Monte Carlo (MC) simulations of physical processes and detector interactions are centrally handled by the GAUSS project [53]. Initially, proton-proton collisions are simulated using PYTHIA [36], which includes parton showering and hadronisation models to describe the evolution from the partonic collision to the primary hadrons. Subsequently, hadronic particles are decayed according to known branching fractions within the EvtGen package [54], while final-state radiation is simulated with PHOTOS [55]. The interaction of the generated particles with the detector material is simulated using Geant4 [56]. Once the detector response has been digitised, the simulated data are processed through the same reconstruction chain as real data: trigger algorithms are applied, tracks and particles are reconstructed, and the resulting samples are made available for analysis.

4. Motivation and Analysis Strategy

The aim of the presented analysis is to measure the *isospin asymmetry* in $B \to K\mu^+\mu^-$ decays, as well as the differential branching fractions of the decays $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$, in intervals of the dimuon invariant mass squared (q^2). The study is based on the full dataset collected by the LHCb experiment in the years 2011 and 2012, referred to as Run 1, and between 2015 and 2018, referred to as Run 2. In this chapter, the physics motivation underlying this measurement is discussed in the context of the current landscape of flavour physics. An overview of previous experimental results on isospin asymmetry is provided, highlighting the need for improved precision. Finally, the strategy adopted to perform the analysis is outlined.

4.1. Status of $b \to s\ell^+\ell^-$ measurements

As discussed in Ch. 2, $b \to s\ell^+\ell^-$ transitions are flavour-changing neutral currents, forbidden at tree level in the Standard Model. Their suppression makes them ideal for searching for New Physics effects, which could manifest as deviations from Standard Model predictions in branching fractions, angular observables, or ratios of these quantities. Over the last decade, several tensions between experimental measurements and Standard Model predictions have emerged in this sector, collectively referred to as *flavour anomalies*.

Experimentally, branching fractions are the most straightforward observable to measure. Persistent tensions have been observed for both the decay modes considered for this thesis. Fig. 4.1 shows that distinct measurements from LHCb [57], CDF [58] and Belle [59] of the differential branching fractions of $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ lie systematically below the Standard Model expectations across the entire q^2 range. Similar results are obtained for other $b \to s\ell^+\ell^-$ processes like $B^0 \to K^{*0}\mu^+\mu^-$ [61, 62], $B_s^0 \to \phi\mu^+\mu^-$ [63], and $\Lambda_b \to \Lambda\mu^+\mu^-$ [64], the last one attesting a coherent behaviour in the baryonic

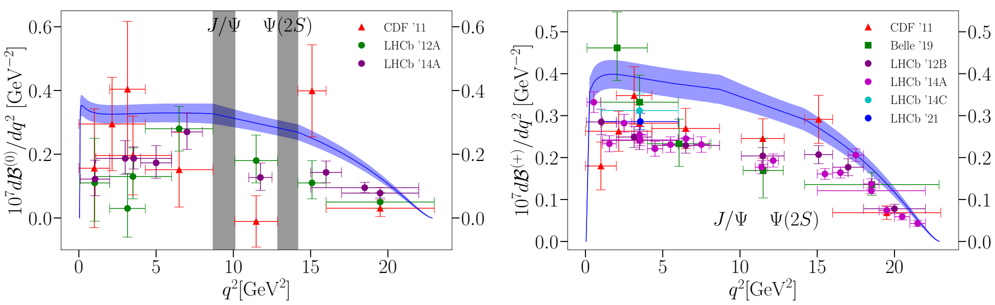


Figure 4.1.: Differential branching fractions for $B^0 \to K^0\mu^+\mu^-$ (left) and $B^+ \to K^+\mu^+\mu^-$ (right) as a function of q^2. Measurements from LHCb, CDF and Belle are compared with the Standard Model prediction (blue bands) from Ref. [60].

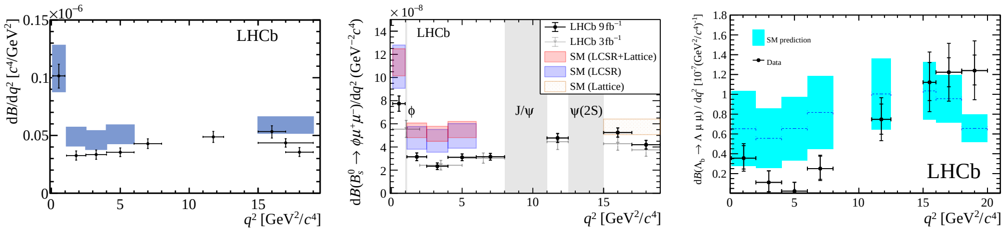


Figure 4.2.: Differential branching fractions for $B^0 \to K^{*0}\mu^+\mu^-$ (left), $B_s^0 \to \phi\mu^+\mu^-$ (centre), and $\Lambda_b \to \Lambda\mu^+\mu^-$ (right) as a function of q^2 measured at LHCb. Standard Model predictions are indicated with coloured bands [61, 63, 64].

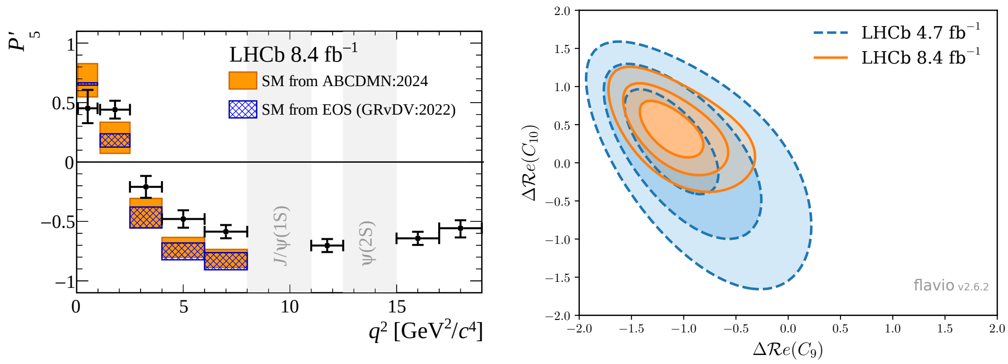


Figure 4.3.: Angular observable P_5' in $B^0 \to K^{*0}\mu^+\mu^-$ as a function of q^2, measured at LHCb and compared with the Standard Model predictions (left), and interpretation in terms of differences with respect to the expected values of the Wilson Coefficients C_9 and C_{10} (right) [65].

sector as well. Measurements from LHCb of the differential branching fractions for these channels are reported in Fig. 4.2. In these cases as well, data are systematically below the Standard Model predictions indicated in the plots, showing tensions of the order of one to three standard deviations, especially in the region $1\,\mathrm{GeV}^2/\mathrm{c}^4 < q^2 < 8\,\mathrm{GeV}^2/\mathrm{c}^4$.

Although these results return a consistent picture for $b \to s\ell^+\ell^-$ processes from an experimental point of view, branching fraction computations are affected by significant hadronic uncertainties related to the form factors. To mitigate these, other observables were introduced with the explicit aim of reducing uncertainties through cancellation of terms. Particularly studied are angular observables, which can be defined for decays into three or more final states. The $B^0 \to K^{*0}\mu^+\mu^-$ decay provides a rich set of angular observables, among which is the P_5' quantity. This variable is related to the interference between the longitudinal and transverse polarisation amplitudes of the K^{*0} meson, and is constructed from ratios of angular coefficients to be largely insensitive to hadronic uncertainties, thus providing a theoretically clean probe of the Wilson coefficient C_9. Measurements of P_5' have shown persistent discrepancies from the Standard Model prediction, as visible in the left side of Fig. 4.3, reporting results from LHCb [65] together with the theoretical expectations. Global fits incorporating these tensions consistently favour a shift in the Wilson coefficient C_9 while constraining C_{10}, pointing towards potential New Physics contributions, as indicated in the right plot of Fig. 4.3. However, interpretations are complicated by non-

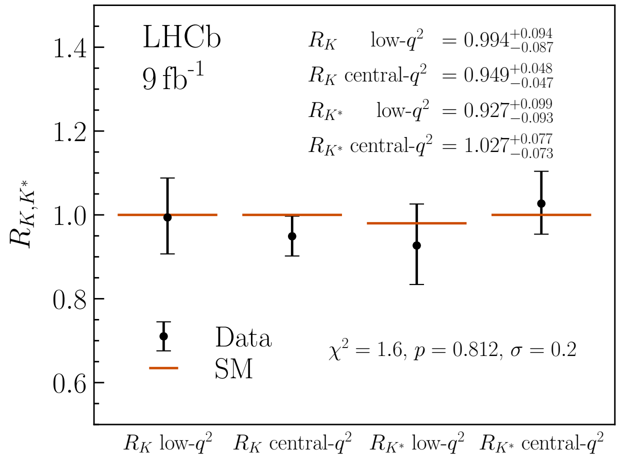


Figure 4.4.: Measurements of R_K and R_{K^*} compared to the Standard Model prediction [67].

local hadronic effects, specifically arising from charm-quark loops, which are challenging to calculate.

Finally, the theoretically cleanest category of observables involves tests of *lepton flavour universality*. In the Standard Model, the coupling of the electroweak gauge bosons to leptons is identical for the three generations. Therefore, the ratios of branching fractions

$$R_X = \frac{\mathcal{B}(B \to X\mu^+\mu^-)}{\mathcal{B}(B \to Xe^+e^-)} \ ,$$

with X being any possible hadronic system, are expected to be close to unity with high precision, in the kinematic region where the mass difference between electrons and muons can be neglected. An important result in this class of measurements has been recently published by LHCb [66]. A simultaneous analysis of R_K and R_{K^*}, summarised in Fig. 4.4, found values which are in perfect agreement with the Standard Model. Superseding previous tensions in lepton flavour universality observables, this announcement contributed to a complex scenario where anomalies remain unsolved for specific quantities. This has made $b \to s\ell^+\ell^-$ transitions an exciting laboratory for complementary tests, such as the isospin asymmetry in $B \to K\mu^+\mu^-$, the subject of this thesis.

4.2. Past measurements of isospin asymmetry

This work aims to contribute to the general efforts for a better comprehension of elementary interactions. Specifically, the focus is on observables that were previously measured, but not with sufficient precision to draw a definitive conclusion. Furthermore, the isospin asymmetry offers a unique diagnostic power regarding the structure of the effective Hamiltonian. While branching fractions and angular observables are primarily sensitive to the semileptonic operators O_9 and O_{10}, A_I provides access to the chromomagnetic dipole coefficient C_8.

The first observation of $B^{0,+} \to K^{0,+}\mu^+\mu^-$ was made in 2001 [68] by the Belle collaboration, a flavour physics experiment operating at the b-factory KEKB in Japan [69]. At that time the neutral and charged

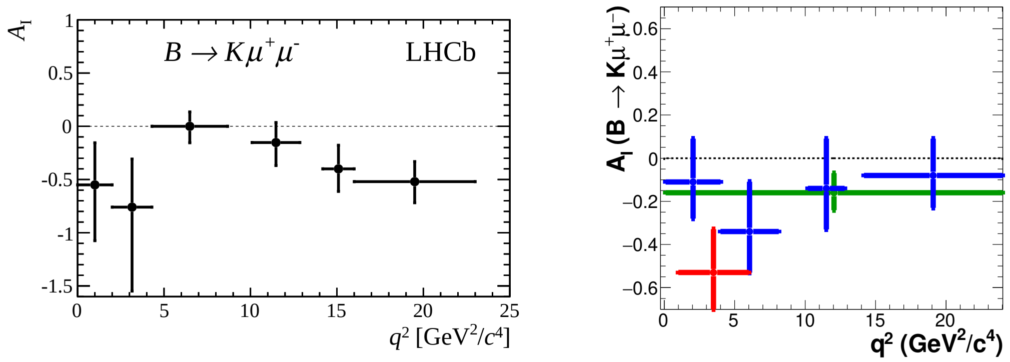


Figure 4.5.: Isospin asymmetry as a function of q^2 measured at LHCb in 2014 (left), and Belle in 2021 (right) as reported in Refs. [57, 59].

cases were treated together, ignoring possible differences between the two. A first experimental study explicitly distinguishing the two channels was performed by the LHCb collaboration in 2014 [57], using data collected during LHC Run 1 and corresponding to $3\,\mathrm{fb}^{-1}$. An analogous measurement was then performed by Belle in 2021 [59]. The results of the measurements are shown in Fig. 4.5 and present an isospin asymmetry of $\mathcal{O}(20\,\%)$. Given the large uncertainties, the measurements are in agreement with the theoretical expectations reported in Ch. 2, which on the other hand are of $\mathcal{O}(1\,\%)$. However, from the plots it is evident that measurements consistently favour negative values, leaving open the possibility for deviations from the Standard Model hiding in plain sight. The analysis presented here will supersede the LHCb measurement from Ref. [57], with a dataset three times larger than the one analysed in 2014, benefiting from additional improvements in track reconstruction and particle identification during the Run 2 data-taking period.

4.3. Analysis overview

This analysis aims to deliver an updated measurement of the branching fractions of the $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0\mu^+\mu^-$ decays in light of the current observed tensions. To the same end, an updated measurement of the isospin asymmetry defined in Ch. 2, which is a clean quantity not affected by theoretical uncertainties due to form factors, is performed.

The analysis is structured into different parts: selection, corrections to simulations, fit to the invariant mass and systematic uncertainty evaluation. Data collected by the LHCb experiment during Run 1 and Run 2, corresponding to an integrated luminosity of $9\,\mathrm{fb}^{-1}$, are selected to remove as many background candidates as possible. In doing so, the signal efficiency should be kept reasonable and introducing biases between the two modes should be avoided. This is obtained with cuts at several stages which are as similar as possible for the two different channels. The full selection is described in Ch. 5, which also comprises the training on simulated samples of a *Boosted Decision Tree* to discriminate between signal and combinatorial background. Since simulated samples are used to define the selection chain and to compute efficiencies, it is of fundamental importance that they correctly reproduce the data. For this reason, a set of correction weights is computed and presented in Ch. 6.

To prevent unintentional biases during the optimisation of the selection and the definition of the fit model, the analysis is performed using a "blind" procedure. The data in the invariant mass signal region for the rare decays are masked and inspected only after the full analysis chain, including the evaluation of systematic uncertainties, is defined and frozen. The resonant modes, serving as control channels and normalisation, are instead treated as an open dataset to validate the procedure.

The isospin asymmetry and the branching fractions are measured via extracting the number of events (yields) from an invariant mass fit, expressing the yields in terms of the parameters of interest. The mass spectra are individually studied for each mode, as shown in Ch. 7. Then, the asymmetry and the $B^0 \to K^0\mu^+\mu^-$ branching fraction are extracted via a simultaneous fit to both channels, using the resonant decays $B^{0,+} \to (J/\psi \to \mu^+\mu^-)K^{0,+}$ as normalisation. The $B^+ \to K^+\mu^+\mu^-$ branching fraction is determined with a fit to the $B^+ \to K^+\mu^+\mu^-$ dataset only, again normalising the yields to those of the resonance. This allows accounting automatically for uncertainties within the fit procedure instead of propagating them back through the combination, giving the possibility of a finer binning as well. All the measurements are performed in intervals (bins) of the dimuon invariant mass squared (q^2). A comprehensive discussion of the fit strategy is given in Ch. 8, together with a validation of the method. Finally, Ch. 9 is dedicated to the study of systematic uncertainties.

5. Data extraction

The precise measurement of rare decay branching fractions requires the extraction of a high-purity signal sample from the vast amount of data collected by the LHCb detector. This chapter details the selection strategy adopted to isolate the $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ candidates, suppressing the dominating background contributions while maximising signal retention. The discussion follows the hierarchical structure of the data processing: starting from the definition of the event topology and the online trigger selection, it proceeds to the centralised stripping and the offline refinement. Finally, the definition of the q^2 binning scheme and the evaluation of the selection efficiencies, which are essential ingredients for the measurement, are presented.

5.1. Event topology

Event reconstruction at particle detectors works backward, tracing back the decay chain starting from the final products. The so-called *signature* is in general represented by a set of tracks that can be extrapolated back through intermediate vertices and states to the primary vertex. To reconstruct $B^+ \to K^+\mu^+\mu^-$ decays, three tracks are required in the final state, namely two muons reaching the muon stations and one kaon producing a shower in the hadronic calorimeter. The tracks are required to have a common vertex in the VELO, corresponding to the decaying B meson, displaced from the primary vertex, i.e. the collision point between the protons. In the case of $B^0 \to K^0\mu^+\mu^-$, the reconstruction is similar but more complex. The K^0 in fact is not directly visible but has to be detected through its decay. Furthermore, the strongly produced K^0 propagates as a superposition of the short-lived K^0_S and the long-lived K^0_L, weakly decaying mostly into two and three pions respectively. Given its long lifetime, the K^0_L is effectively undetectable at LHCb. Therefore, the detection has to go through the $K^0_S \to \pi^+\pi^-$ decay. This replacement, from $B^0 \to K^0\mu^+\mu^-$ decay to $B^0 \to K^0_S\mu^+\mu^-$, implies a factor of $1/2$ arising from the decomposition

$$|K^0\rangle = \frac{1}{\sqrt{2}}\left(|K^0_S\rangle + |K^0_L\rangle\right)$$

when considering the branching fraction. Apart from the K^0_S decay, the reconstruction strategy is the same as for the charged case. The four tracks are still combined in intermediate vertices, this time looking for one more distinct vertex for the K^0_S meson. In this case it is worth mentioning that the reconstruction quality depends on whether the K^0_S decays inside or outside the VELO region. If it decays inside, the two pions will be both *long tracks*, such that the vertex as well as the momenta can be determined with higher precision, resulting in an overall better reconstruction quality. If it decays outside, on the contrary, the two pions will be *downstream tracks* and the measurement cannot benefit from the vertex locator. The two cases, namely *long-long* (LL) and *down-down* (DD) are summarised in Fig. 5.1.

It is clear that the signal cannot be identified based on the final state alone. Different processes could result in the same signature in the detector. Moreover a wrong combination of tracks or a missed particle

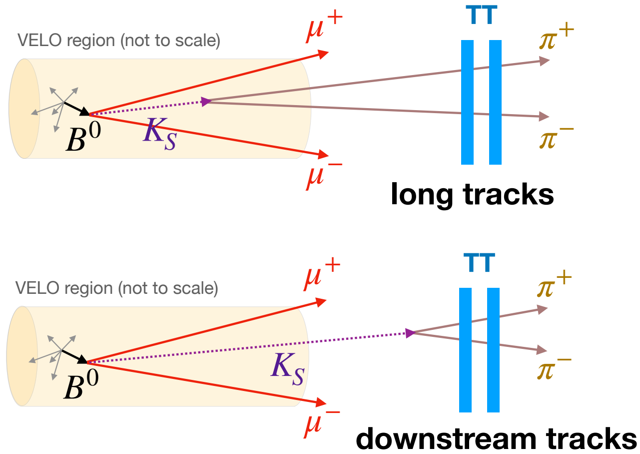


Figure 5.1.: Definition of the long-long (top) and down-down (bottom) track types depending on whether the K^0_S decays inside or outside the VELO region.

could mimic the signal. In general there are three main classes of backgrounds: misidentified background, partially reconstructed background and combinatorial background. The first two are due to actual decays, where one or more tracks are misidentified, meaning assigned to a wrong particle, or completely missed, respectively. Combinatorial backgrounds are on the contrary events made up of random combinations of unrelated tracks, distributed in space like the decay of interest purely by chance. For rare processes like the ones studied in this analysis it is critical to reduce the contamination from these backgrounds as much as possible. At the same time, the signal efficiency needs to be maximised where possible. In the following the selection criteria are described in detail.

5.2. General selection

The selection starts with some general cuts that define the dataset of study. These cuts are all aimed at translating the considerations from the previous section into quantitative requirements, although acting on different levels. Data are collected after an online selection, the *trigger*, which determines which events are worth storing. The *stripping selection* is the collaboration-wide set of requirements that allows filtering events with the desired topology. Finally, some additional cuts are added, specific to the decays of interest.

5.2.1. Trigger selection

The *trigger* is the system that, after an online reconstruction, activates data acquisition, storing the information only for those events that are judged as interesting. As explained in Sec. 3.3 the trigger is organised in three levels of increasing precision, and for each level several *trigger lines* are available. A trigger line is a batch of requirements imposed on hits or track properties, compatible with the information available at that stage. At L0 level, for example, the decision has to be made looking at hits in the calorimeters and in the muon chambers, having only a rough estimate of the momentum or energy. Concerning $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ decays, the `Muon` and `DiMuon` lines check for the presence of

Table 5.1.: `isMuon` algorithm hit requirements in the muon stations based on the momentum of the track.

p_{track} [GeV/c]	Hits in muon stations
3–6	M2 and M3
6–10	M2 and M3 and (M4 or M5)
> 10	M2 and M3 and M4 and M5

Table 5.2.: Trigger requirements for the three separate data taking periods (Run 1, 2016, 2017+2018) considered in this analysis. On each trigger stage at least one of the triggers is required to have fired. All trigger decisions are evaluated `TOS`.

	Run 1	2015 – 2016	2017 – 2018
`L0`	`Muon`	`Muon`	`Muon`
	`DiMuon`	`DiMuon`	`DiMuon`
`HLT1`	`TrackAllL0`	`TrackMVA`	`TrackMVA`
	`TrackMuon`	`TrackMuon`	`TrackMuon`
		`TwoTrackMVA`	`TwoTrackMVA`
`HLT2`	`Topo[2,3,4]BodyBBDT`	`Topo[2,3,4]Body`	`Topo[2,3,4]Body`
	`TopoMu[2,3,4]BodyBBDT`	`TopoMu[2,3,4]Body`	`TopoMu[2,3,4]Body`
	`DiMuonDetached`		`DiMuonDetached`
	`SingleMuon`		

a muon or a pair of muons, respectively. This is determined mainly with the `isMuon` algorithm, a boolean method that identifies muons based on the number of muon stations hit by the track and its momentum. Low-momentum tracks will be absorbed before the end of the detector, while high-momentum tracks are expected to reach the last stations. A summary of the criteria is shown in Tab. 5.1. At HLT1, a partial reconstruction is available and displaced vertices for high p_{T} tracks are selected. These are the object of the lines `TrackAllL0` for Run 1 and `TrackMVA` and `TwoTrackMVA` for Run 2. The same is imposed on muon tracks with the line `TrackMuon`. Finally, a complete reconstruction is performed at HLT2 and the whole topology of the decay can be exploited. The `Topo[N]Body` lines are specifically designed to select `N`-body B-meson decays with at least two charged child particles. In addition to that, the `SingleMuon` and `DiMuonDetached` lines select displaced muon and di-muon events. For each level at least one of the criteria described above has to be fulfilled, although not all of the lines are available for all the years. The trigger selection chain per data-taking period is given in Tab. 5.2. It should be noted that the decision is *Trigger On Signal* (TOS), meaning that it is evaluated on the signal candidate part of the event only. Conversely, it could happen that other elements among the products of the proton-proton collision cause the trigger to fire. Such events are known as *Trigger Independent Signal* (TIS) and they will be used to determine correction weights to simulation samples in Ch. 6.

5.2.2. Stripping selection

The LHCb detector serves many purposes of a very different nature. The amount of data collected is large and not all relevant when focusing on specific decays. On the contrary, an extreme excess of unnecessary data is detrimental to the measurement. The reduction of the dataset to a manageable size, selecting the interesting events among all the available ones passing the trigger stage, is therefore a priority. At LHCb this step is called *stripping* and gives a dataset of signal candidates with the correct displacement in space, where the reconstruction quality is ensured by studying ad hoc quantities. The goodness of the combination of two or more tracks into a vertex is evaluated with the *Distance Of Closest Approach* (`DOCA`), that is, the minimum distance between the tracks. Of course, if the tracks are originating from the same point this distance should be close to zero, considering the limited resolution. The *Impact Parameter* (`IP`) is the minimum distance of the track, or its extrapolation, from the primary vertex. It is useful when looking for displaced vertices, since a large value is expected in those cases. For a more performant test on secondary vertices, however, the χ^2_{IP} is defined. This is the difference between the χ^2 of the primary vertex reconstructed with and without the track under consideration. Analogously to the impact parameter, a large χ^2_{IP} indicates a secondary vertex, while a small value is compatible with the hypothesis that the track comes directly from the primary vertex. Once the displaced vertex is identified, the hypothesis that this is due to a long-lived particle is exploited with the *Direction Angle* (`DIRA`) corresponding to the angle between the line connecting primary and secondary vertices and the sum of the four-momenta of the decay products. Finally, a multivariate classifier discriminates against tracks made up of random hits in the detector, giving a *ghost probability*, and the overall occupancy of the detector is limited by requiring a maximum number of hits in the SPD.

Information from the RICH detectors is processed to extract likelihoods for each track of being a specific particle. Typically they are given relatively to some standard as $\texttt{DLL}_X$, the difference between the logarithm of the likelihoods of being a specific particle X and of being a pion. For the muons, besides `isMuon`, a cut on $\texttt{DLL}_\mu$ is applied. No further cut on particle identification variables is required for hadrons. Only for $B \longrightarrow K^0_{\text{S}}\mu^+\mu^-$, fiducial cuts are imposed on K^0_{S} reconstruction through its lifetime τ, the χ^2_{Vtx} of the decay vertex and the flight distance. A complete overview of the stripping selection is given in Tab. 5.3.

5.2.3. Additional cuts

While the stripping is a collaboration-wide procedure to select general classes of processes, additional selection criteria specific to the decays of interest are applied. First, a reasonable angular separation is required between all tracks. Indeed, if the angle between two tracks is extremely small, they could be just a reconstruction artefact of a single track being reconstructed twice. A more advanced particle identification requirement is then applied to muons and charged kaons. This is based on the `ProbNN` variable, the output of a multivariate classifier combining not only information from the RICH detectors, but also from the tracking and calorimeter systems. In contrast, no particle identification is applied to the pions originating from the K^0_{S} decay. In this case, a tight range for the reconstructed mass $m(K^0_{\text{S}})$ is sufficient, different for LL and DD-type tracks. Details on the additional selection are given in Tab. 5.4.

5.3. Vetoes against exclusive backgrounds

The selection described in the previous section provides candidates with the correct topology, but as already mentioned it does not uniquely identify the processes of study. Specific decays can be included in the dataset when tracks are misidentified or not correctly combined. Tighter cuts on particle identification variables would help for the first case, but this approach comes at the cost of signal efficiency

Table 5.3.: Stripping requirements used to select $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ events.

B parent cuts	
Candidate	Requirement
B^0, B^+	Invariant mass $\in [4600.0, 7100.0]\,\mathrm{MeV/c^2}$
B^0, B^+	Decay vertex χ^2 per DOF $<$ 8.0
B^0, B^+	$\mathrm{IP}_{\chi^2} < 16.0$
B^0, B^+	$\mathrm{DIRA} > 0.9999$
B^0, B^+	Flight distance $\chi^2 > 64.0$
B^0, B^+	Maximum child $\mathrm{IP}_{\chi^2} > 9.0$

K^0_S cuts	
Candidate	Requirement
K^0_S	Invariant mass $\in [467.0, 527.0]\,\mathrm{MeV/c^2}$
K^0_S	Lifetime $\tau_{K^0_S} > 2\,\mathrm{ps}$
π^+, π^-	Momentum $p > 2.0\,\mathrm{GeV/c}$
DD type-specific (child) cuts	
K^0_S	Decay vertex $\chi^2 < 25$
π^+, π^-	Minimum $\mathrm{IP}_{\chi^2} > 4.0$
π^+, π^-	χ^2 DOCA < 25.0
LL type-specific (child) cuts	
K^0_S	Decay vertex $\chi^2 < 30$
K^0_S	Flight distance $\chi^2 > 4.0$
π^+, π^-	Minimum $\mathrm{IP}_{\chi^2} > 9.0$
π^+, π^-	χ^2 DOCA < 30.0

Dimuon system cuts	
Candidate	Requirement
$\mu^+\mu^-$	Invariant mass $< 7100.0\,\mathrm{MeV/c^2}$
$\mu^+\mu^-$	Decay vertex χ^2 per DOF $<$ 12.0
$\mu^+\mu^-$	DIRA > -0.9
$\mu^+\mu^-$	Flight distance $\chi^2 > 9.0$
$\mu^+\mu^-$	Maximum IP_{χ^2} of either μ^+ or μ^- > 6.0
Muon cuts	
μ^+, μ^-	Minimum $\mathrm{IP}_{\chi^2} > 6.0$
μ^+, μ^-	$\mathrm{DLL}_\mu > -3.0$
μ^+, μ^-	Positive `isMuon` decision, based on hits in muon chambers

General candidate requirements	
Candidate	Requirement
Tracks	Ghost probability < 0.5
Event	Maximum hits in SPD subdetector < 600

Table 5.4.: Additional selection specific to $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ events.

Candidate	Requirement
Track pairs	Angular separation $\theta > 1\,\mathrm{mrad}$
μ^+, μ^-	$\texttt{ProbNN}_\mu(\mu) > 0.1$
$K^\pm$	$\texttt{ProbNN}_K(K) > 0.1$
K^0_S reconstructed mass	$m_{K^0_S} \in [487.6, 507.6]\,\mathrm{MeV/c^2}$ (LL-type)
	$m_{K^0_S} \in [482.6, 512.6]\,\mathrm{MeV/c^2}$ (DD-type)

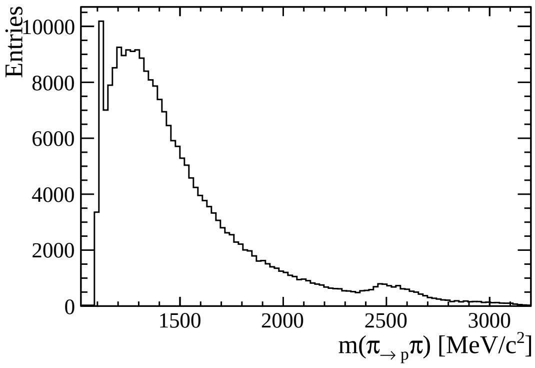


Figure 5.2.: Invariant mass of the $\pi\pi$ pair when the proton mass is assigned to the first track. A clear peak is evident at the mass of the Λ, removed with the veto defined in the text.

without addressing the issue of false reconstruction. A more effective method is to recombine tracks exploiting different mass hypotheses. When recombining tracks with the right hypotheses, contaminating backgrounds manifest as peaks in the recomputed mass spectrum. Excluding data in the corresponding region with a veto has the advantage of removing almost all the background while having a small effect on actual signal.

The decay $\Lambda_b \to \Lambda(\to p\pi^-)\,\mu^+\mu^-$ enters the $B^0 \to K^0_S(\to \pi^+\pi^-)\,\mu^+\mu^-$ dataset if the proton is misidentified as a pion. It can be revealed by considering the $\pi^+\pi^-$ invariant mass, assigning the proton mass to one of the two tracks, which will be indicated as $m(\pi_{\to p}\pi)$. As shown in Fig. 5.2, a peak can be seen at the mass of the Λ, being a proof for the presence of the background. Studying its shape, the region $1086\,\mathrm{MeV/c^2} < m(\pi_{\to p}\pi) < 1146\,\mathrm{MeV/c^2}$ was excluded. Something similar happens in the $B^+ \to K^+\mu^+\mu^-$ case with the $B^+ \to \bar{D}^0(\to K^+\pi^-)\,\pi^+$ decay, when both the pions are misidentified as muons. Here, the $K^+\mu^-$ invariant mass, recomputed assuming the pion mass for the muon candidate, is considered and a veto centred on the D^0 mass, namely in the range $1835\,\mathrm{MeV/c^2} < m(K\mu_{\to\pi}) < 1895\,\mathrm{MeV/c^2}$, is imposed. Finally, the resonant decay $B^+ \to J/\psi(\to \mu^+\mu^-)\,K^+$ deserves special attention. When correctly reconstructed, it populates the dataset of the charged mode in the q^2 range around the J/ψ mass and is used as a control and normalisation channel. However, it can contaminate the rest of the range if the kaon is misidentified as a muon and the muon is misidentified as a kaon. Combining then the kaon track with the opposite-sign muon candidate, while assigning the muon mass to the first, the contamination appears. To remove it, the veto $3037\,\mathrm{MeV/c^2} < m(K_{\to\mu}\mu) < 3157\,\mathrm{MeV/c^2}$ can

Table 5.5.: Vetoes against exclusive backgrounds. The notation $m(AB_{\to X})$, defined in the text, indicates the recomputed invariant mass of the AB combination when assigning the mass of X to the B track.

Dataset	Vetoed region
$B^0 \to K^0_S\mu^+\mu^-$	$1086\,\mathrm{MeV/c^2} < m(\pi_{\to p}\pi) < 1146\,\mathrm{MeV/c^2}$
$B^+ \to K^+\mu^+\mu^-$	$1835\,\mathrm{MeV/c^2} < m(K\mu_{\to\pi}) < 1895\,\mathrm{MeV/c^2}$
	$3037\,\mathrm{MeV/c^2} < m(K_{\to\mu}\mu) < 3157\,\mathrm{MeV/c^2}$

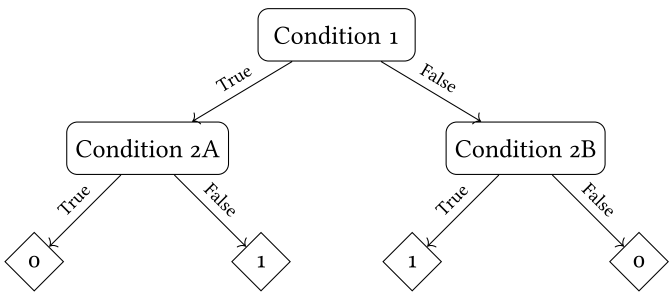


Figure 5.3.: Sketch of a decision tree with depth 3. Starting from the top the first condition is evaluated and one of the two branches is followed accordingly to the result. This is repeated until the final decision is reached in one of the leaves.

be applied. The three vetoes are summarised in Tab. 5.5. Other combinations were considered but no additional background was found.

5.4. Multivariate selection

Another background affecting the dataset is composed of independent tracks, correctly or wrongly identified, coming from different decays but displaced from the primary vertices like the signal and passing the previous selection purely by chance. Given its very general nature there is no obvious feature to recognise this type of event, which, conversely, shares only an overall different behaviour from signal. If one property is not sufficient to spot the differences, a classification is possible when looking at the distribution of many variables. This is the idea of multivariate analysis: considering the overall event to determine a measure of the likelihood of being signal or background. Given the tabular nature of the data, *Boosted Decision Trees* (BDT) are particularly suitable for this aim.

A decision tree is a binary classifier organised as a chain of condition cases. At each condition, corresponding to a *node*, the chain splits into two *branches* and this is repeated until the final *leaves* which provide the decision. The number of nodes that the decision process has to go through is called *depth*. A graphical description of a depth-3 decision tree classifying between two categories, 0 and 1, is shown in Fig. 5.3. Each condition corresponds to a cut on one or more quantities and machine learning algorithms allow configuring all the nodes with a *training* procedure, fitting the output to reference samples of both signal and background. A single decision tree is however usually not sufficient to exploit the full variable space and combination of more trees, weighting the single outputs and summing them to obtain a

Table 5.6.: Input features for the multivariate algorithms.

Input variables		
B^0 LL-type	B^0 DD-type	B^+
$p_{\mathrm{T}}(B^0)$	$p_{\mathrm{T}}(B^0)$	$p_{\mathrm{T}}(B^+)$
$\eta(B^0)$	$\eta(B^0)$	$\eta(B^+)$
$\mathrm{IP}(B^0)$	$\mathrm{IP}(B^0)$	$\mathrm{IP}(B^+)$
$\min \chi^2_{\mathrm{IP}}(B^0)$	$\min \chi^2_{\mathrm{IP}}(B^0)$	$\min \chi^2_{\mathrm{IP}}(B^+)$
Proper lifetime τ (B^0)	Proper lifetime τ (B^0)	Proper lifetime τ (B^+)
$\mathrm{DOCA}(B^0)$	$\mathrm{DOCA}(B^0)$	$\mathrm{DOCA}(B^+)$
$\log_{10}(\mathrm{DIRA})$ (B^0)	$\log_{10}(\mathrm{DIRA})$ (B^0)	$\log_{10}(\mathrm{DIRA})$ (B^+)
$\log_{10}(\mathrm{MTDOCA})$ (B^0)	$\log_{10}(\mathrm{MTDOCA})$ (B^0)	$\log_{10}(\mathrm{MTDOCA})$ (B^+)
$\log_{10}(\chi^2_{\mathrm{ENDVERTEX}})$ (B^0)	$\log_{10}(\chi^2_{\mathrm{ENDVERTEX}})$ (B^0)	χ^2_{VTX} (B^+)
$p_{\mathrm{T}}(K^0_{\mathrm{S}})$	$p_{\mathrm{T}}(K^0_{\mathrm{S}})$	$p_{\mathrm{T}}(K^+)$
$\min \chi^2_{\mathrm{IP}}(\mu)$	$\min \chi^2_{\mathrm{IP}}(\mu)$	$\min \chi^2_{\mathrm{IP}}(\mu)$
$\beta = \frac{p(\mu^+\mu^-)-p(K^0_{\mathrm{S}})}{p(\mu^+\mu^-)+p(K^0_{\mathrm{S}})}$	$\beta = \frac{p(\mu^+\mu^-)-p(K^0_{\mathrm{S}})}{p(\mu^+\mu^-)+p(K^0_{\mathrm{S}})}$	$\min \chi^2_{\mathrm{IP}}(K^+)$

score. In boosted decision trees weights are sequentially assigned, again making use of machine learning, focusing on classification errors and trying to recover from them [70]. In particular for this analysis the XGBoost algorithm [71] was used. Before the training, the package allows defining the model setting several *hyperparameters*. Besides the aforementioned depth, the complexity of the model is driven by the number of trees. The more trees are present, the better the classification can reproduce the training dataset. This remains true until the limit of *overfitting*, when the reproduction is too specifically tailored to the training dataset and statistical fluctuations are recognised as properties of the class. This would bring a loss of efficiency when applying the classifier to the real sample and has to be carefully avoided. A possible way to keep the effect under control consists of training the BDT on a reduced dataset while comparing performances with the rest of the sample. Consistent performance ensures the absence of overfitting, at the cost of smaller datasets available for the training. A more efficient approach to perform the same check is dividing the dataset into k subsets, iteratively using $k-1$ of them for the training and the remaining one for testing. The average among the k instances exploits the whole sample while giving a controlled output. This technique is called *k-folding* and was followed for this study, choosing the value $k = 5$.

A different model was trained for $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$ LL-type, $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$ DD-type and $B^+ \to K^+\mu^+\mu^-$ data, for Run 1 and Run 2 separately. It results in six different classifiers sharing the number of trees, set to 600, the depth, set to 3, and the learning rate, an additional hyperparameter that regulates the convergence speed balancing with the accuracy, that was set to 0.01. For the training, corrected simulated samples of the respective decays were used as proxies of the signal, while every real event falling in the region $m(B) > 5400\,\mathrm{MeV}/c^2$, the upper mass sideband, was assumed as background. The models were fed with the topological and kinematic observables listed in Tab. 5.6, most of them already introduced in Sec. 5.2.2. The choice of these variables, called *features* in a classification context, is motivated by the physical nature of the combinatorial background. Combinatorial background comes from wrong combination or reconstruction of the tracks and this affects the quality of vertices and impact

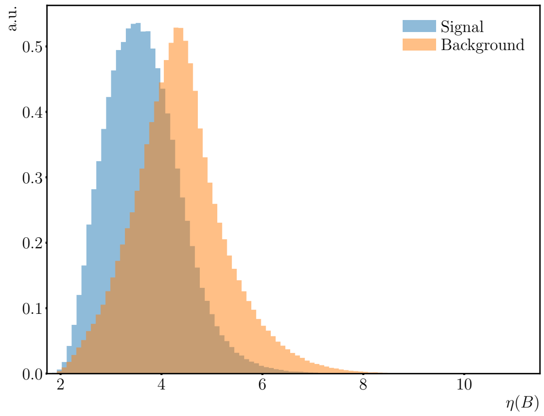


Figure 5.4.: Distribution of the pseudorapidity of signal and background $B^+ \to K^+\mu^+\mu^-$ events. The signal distribution is determined from Run 1 $B^+ \to K^+\mu^+\mu^-$ simulated samples, the background distribution is determined from data mass sidebands for the same period of data-taking.

parameters. Moreover, putting together independent tracks, momentum distributions are not constrained by the properties of a B meson decay and can be used as discriminant. The latter, in particular, provide significant discrimination power even if some care is needed when including them. Since in the end a fit to the reconstructed invariant mass is performed, features used for the BDT should not be correlated with the B mass, to avoid the risk of shaping the spectrum and creating artificial structures. All momenta are naturally correlated with the B mass, reconstructed from the momenta themselves, but these correlations are considered acceptable for secondary tracks, given the benefits they provide. The actual separation power of the features is checked comparing the respective distributions on signal and background samples. To be of use for the classification, there has to be a sensible difference between the two, as visible in Fig. 5.4. At the same time since the training relies on simulated samples, it is important that in those cases distributions are well reproduced. For this aim correction to simulations are considered, the procedure being the subject of the next chapter. To be sure of the accuracy of the corrections the control channels $B^0 \to J/\psi K^0_S$ and $B^+ \to J/\psi K^+$, whose kinematics are similar to the signal and for which a statistical background subtraction is possible, are used. For the chosen variables the distributions in simulated signal samples are compared with the ones in background subtracted control channel data, to test their compatibility. Finally correlations between features are computed and excluded if too large, avoiding redundant use of the same information.

The trained model returns as output a value between 0 and 1, reflecting the likelihood of the event of being background or signal: the closer to 1 the value the more likely to be signal. The results evaluated on training and testing datasets for one example are shown in Fig. 5.5. Similar results are obtained for all the models and all the training folds. The clear separation of the peaks between signal and background samples proves the power of the classification. The fact that the behaviour between training and test samples is the same demonstrates that the risk of overfitting is well prevented. When applied to real data the BDT output can be used for a further selection to remove combinatorial background. In principle, cutting below any threshold would reduce the background, but a compromise has to be found between background rejection and signal efficiency. In the end the aim should be a significance that is as large as

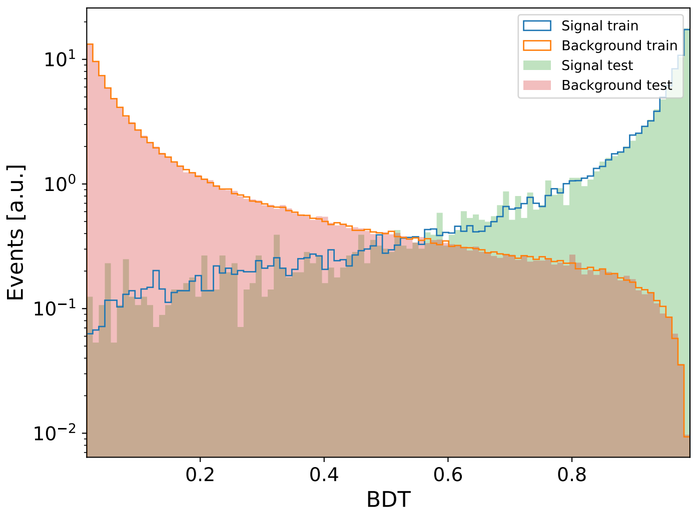


Figure 5.5.: BDT distributions after the training on $B^0 \to K^0_S \mu^+\mu^-$ DD-type samples for Run 2 evaluated on training and test samples. An example fold is shown, but similar results are obtained on the other folds.

Table 5.7.: Optimal BDT cut points as determined maximising the figure of merit $S/\sqrt{S+B}$.

Period	B^0 LL	B^0 DD	B^+
Run 1	0.775	0.88	0.725
Run 2	0.89	0.91	0.72

possible, but at this step, with a blinded dataset, it is not possible to predict it. A good proxy is represented by the *figure of merit*

$$\frac{S}{\sqrt{S+B}}, \tag{5.1}$$

where S is the number of signal events and B is the number of background events. The figure of merit defined in Eq. (5.1) is dependent on the cut and maximising it means maximising the expected significance. The procedure is depicted in Fig. 5.6 for Run 2: for each BDT cut the function is computed. The number of background events is determined by performing a fit of an exponential function to the data sample, blinded in the region $[5220, 5330]$ MeV/c^2. The number of signal events, not directly accessible because of the blinding, is determined by scaling the number of control channel events obtained from a mass fit to data, in the resonant q^2 window. For the scaling relative branching fractions as reported in Ref. [72] and relative efficiencies of rare and resonant channels are used. Maximum values of the figures of merit, for each model and each period, determine the optimal BDT cuts reported in Tab. 5.7.

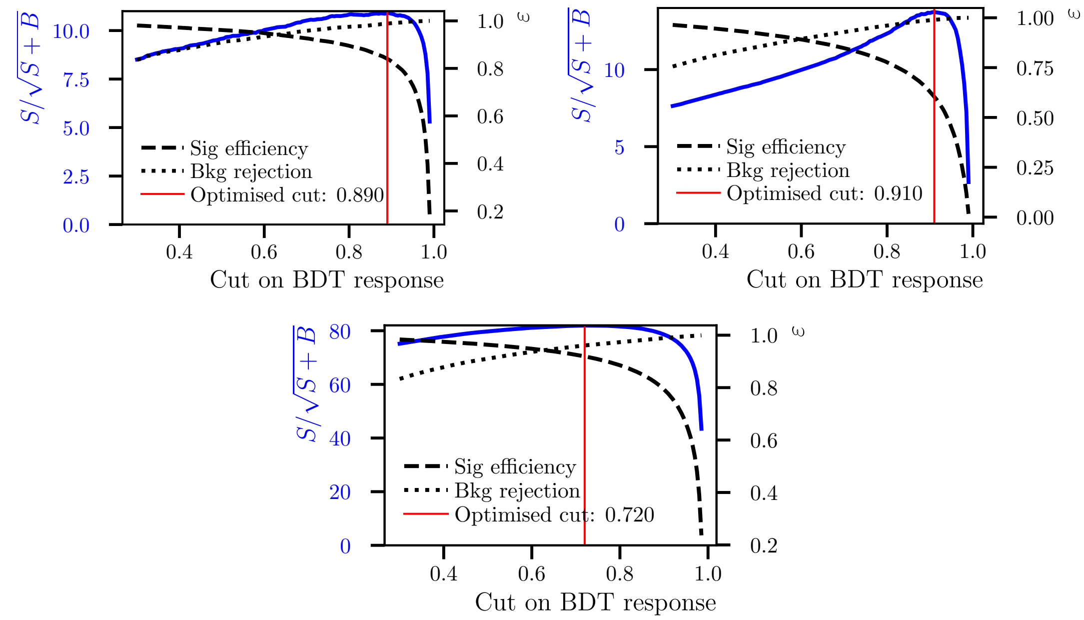


Figure 5.6.: Figures of merit for the modes $B^0 \to K^0_S\mu^+\mu^-$ LL-type (top-left), $B^0 \to K^0_S\mu^+\mu^-$ DD-type, and $B^+ \to K^+\mu^+\mu^-$ in Run 2. The red line represents the optimal BDT cut. For reference, the signal efficiency and the background rejection are also reported.

5.5. q^2 binning

As mentioned in Ch. 2, for processes involving $b \to s\ell\ell$ transitions the dependency on the exchanged energy is translated into the squared dilepton invariant mass (q^2). All relevant quantities are dependent on q^2, but more important effects beyond the Standard Model could give different contributions to the same quantities in different q^2 ranges. For this reason, the theoretical community provides differential versions of the physical observables. To perform a comparison with these results, from the experimental side a procedure referred to as *binning* is used: the dataset is divided into several q^2 intervals and the same analysis is repeated for the different subsets, leading to different results dependent on the corresponding q^2 range.

The number of intervals, or *bins*, depends on the available statistics. Of course it would be ideal to have as many intervals as possible, to fully analyse the dependence approaching the continuum limit. However, more intervals means also fewer events per single bin, and consequently lower statistical precision. A compromise should be found, and this analysis follows two strategies. Due to K^0_S reconstruction, the decay $B^0 \to K^0_S\mu^+\mu^-$ presents much fewer candidates than the other considered decays. In the context of a simultaneous measurement the small statistics of this channel drives the uncertainty on the measured quantities, and the binning scheme has to be designed on the basis of it. Repeating what was done in the Run 1 analysis, the scheme described in Tab. 5.8 is adopted for the measurement of the $B^0 \to K^0\mu^+\mu^-$ branching fraction, and the isospin asymmetry. In contrast, a denser bin scheme, reported in Tab. 5.9, is used when separately measuring the branching fraction of the decay $B^+ \to K^+\mu^+\mu^-$, which has much higher statistics.

Table 5.8.: q^2 binning schemes used for the measurement of $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and A_I.

q^2 binning scheme $[\mathrm{GeV}^2/\mathrm{c}^4]$
0.1, 0.98, 1.1, 2.5, 4.0, 6.0, 8.0, 11.0, 12.50, 15.0, 17.0, 19.0

Table 5.9.: q^2 binning scheme used for the measurement $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$.

q^2 binning scheme $[\mathrm{GeV}^2/\mathrm{c}^4]$
0.1, 0.98, 1.1, 2.0–8.0 in 1 $\mathrm{GeV}^2/\mathrm{c}^4$ steps, 11.0, 11.75, 12.50, 15.0–22.0 in 1 $\mathrm{GeV}^2/\mathrm{c}^4$ steps

5.6. Selection efficiencies

While removing background, the selection acts on signal events too, modifying the number of candidates. For an analysis which relates the yields to the measurement of the branching fraction, it is critical to determine with precision the efficiencies on signal of the applied cuts. Each step of the selection chain contributes to the total efficiency, which then factorises as

$$\varepsilon_{\mathrm{tot}} = \varepsilon_{\mathrm{geo}} \times \varepsilon_{\mathrm{strip|geo}} \times \varepsilon_{\mathrm{trig|strip}} \times \varepsilon_{\mathrm{presel|trig}} \times \varepsilon_{\mathrm{PID|presel}} \times \varepsilon_{\mathrm{BDT|PID}} \tag{5.2}$$

where all indices are self-explanatory abbreviations for the levels introduced before, except for the term $\varepsilon_{\mathrm{geo}}$ which indicates the geometrical acceptance of the detector. Selection efficiencies are computed from corrected simulation using a Bayesian approach, as described in [73]. Corrections to the simulated samples will be discussed in the next chapter. For now, it is sufficient to mention that a set of per-event weights is introduced to modify the distribution of simulated events. This means that any candidate counts not as one, but as the weight w associated with it. In each q^2 interval and for each selection step the efficiency is then calculated using the estimator

$$E(\varepsilon) = \frac{\sum\limits_{\mathrm{passed}} w_i \; + \frac{1}{2}}{\sum\limits_{\mathrm{all}} w_i \; + 1} \equiv \frac{M + \frac{1}{2}}{N+1} \tag{5.3}$$

where $M = \sum\limits_{\mathrm{passed}} w_i$ indicates the sum of the weights for events that passed the selection step, while $N = \sum\limits_{\mathrm{all}} w_i$ is the sum of weights of all events, prior to the applied cut. The corresponding variance is obtained from

$$V(\varepsilon) = \frac{(M + \frac{1}{2})\;(N - M + \frac{1}{2})}{(N+1)^2\;(N+2)} \quad . \tag{5.4}$$

Eq. (5.3) and Eq. (5.4) require a study of the candidates before the selection is applied. While this is in general straightforward for most of the steps, from regular simulations it is not possible to extract events which were not reconstructed because they fall outside of the detector acceptance, or are rejected from the loose preselection defined in the stripping. To save execution time and storage space, in fact, events are usually completely simulated only if all products lie inside the geometrical acceptance of the detector. For this reason, a so-called *generator level* simulation, on a lower scale but including unfiltered information from the production, was privately produced. In this case a complete reconstruction of all the variables is not necessary, and a study of dilepton invariant mass and track angles with respect to the beam axis is sufficient to evaluate the effect of the pseudorapidity cut, $2 < \eta < 5$, and the fiducial cuts applied at LHCb, $5\,\mathrm{mrad} < \theta < 400\,\mathrm{mrad}$ for neutral particles, $10\,\mathrm{mrad} < \theta < 400\,\mathrm{mrad}$ for charged particles. On the other hand, simulated samples are distributed within the collaboration after the stripping is applied. In order to compute stripping efficiencies, a so-called *restripping* was performed, removing cuts at stripping level from the samples.

Given the observations regarding the first steps of the selection chain, the efficiencies are computed separately for the resonant and the rare modes. However, in the context of a relative measurement, the crucial quantity is the ratio between the two. The ratio between the efficiency computed on the rare channel and the one computed on the corresponding resonant channel is indeed what directly enters the formulae for branching fractions and isospin asymmetry, as will be clear in Ch. 8. The ratios between the efficiencies computed on $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ LL-type samples, $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ DD-type samples, and $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi K^+$, are reported in Fig. 5.7 for all the years. These values result from the product of the efficiencies of all selection steps, following Eq. (5.2), and are computed as a function of q^2 adopting the scheme of Tab. 5.8 valid for the simultaneous measurement. Since a single measurement will be performed for $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, efficiency ratios are reported also with the alternative scheme of Tab. 5.9 for the charged case in Fig. 5.8.

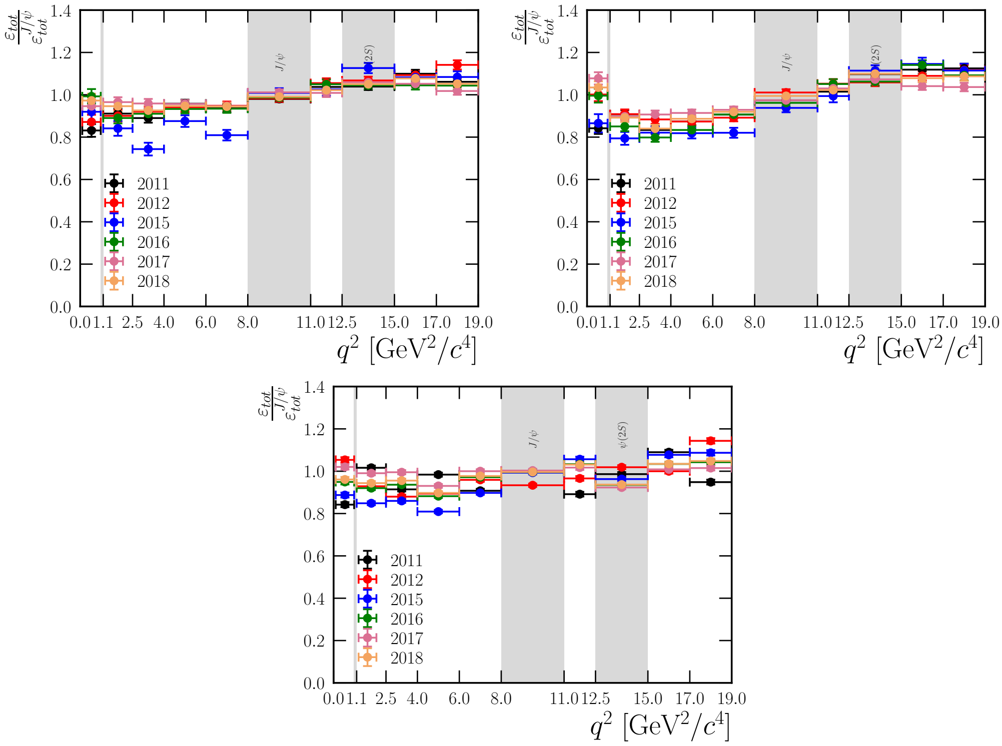


Figure 5.7.: Ratios between the total efficiencies computed on $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ *LL*-type samples (top-left), $B^0 \to K^0_S\mu^+\mu^-$ and $B^0 \to J/\psi K^0_S$ *DD*-type samples (top-right), and $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi K^+$ (bottom), for all the years of data-taking as a function of q^2. These values are input for the measurement of the $B^0 \to K^0\mu^+\mu^-$ branching fractions and the isospin asymmetry.

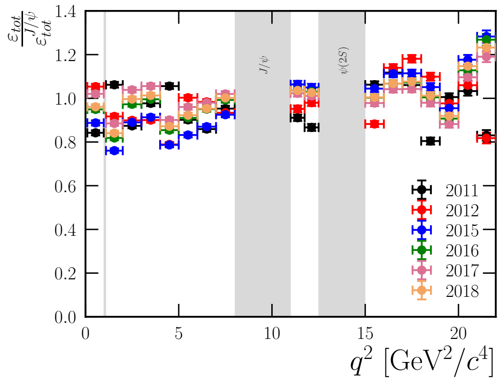


Figure 5.8.: Ratios between the total efficiencies computed on $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to J/\psi K^+$ for all the years of data-taking, adopting the dense q^2 binning scheme of Tab. 5.9. These values are input for the measurement of the $B^+ \to K^+\mu^+\mu^-$ branching fraction.

6. Corrections to simulation

As discussed in the previous chapters, this analysis relies on simulated data. Monte Carlo simulations of both rare ($B^{0,+} \to K^{0,+}_{(\mathrm{S})}\mu^+\mu^-$) and resonant ($B^{0,+} \to J/\psi\, K^{0,+}_{(\mathrm{S})}$) channels are used to train the BDT models for combinatorial background rejection and to compute efficiencies. As will be described in Ch. 7 and Ch. 8, they are employed for modelling the invariant mass shapes for the fit procedure. Since the simulated samples play an important role in the analysis, it is crucial that they correctly reproduce the behaviour of real events. However, due to the complexity of the experimental environment, this is not always the case. In the following, the correction strategy based on comparison with calibration data, already applied for the presented results, is discussed in detail.

6.1. ${}_s\mathcal{P}lot$ technique

Monte Carlo simulations provide pure samples accounting only for the signal of interest, free from any contamination by other processes. This represents a unique feature not replicable in experimental data, which are generally affected by background. For this reason, when considering distributions, a direct comparison between the two types of datasets is not possible, at least not before performing a background subtraction. To achieve a statistical subtraction without discarding events, the ${}_s\mathcal{P}lot$ technique [74] is employed.

In data, in fact, two or more irreducible components of different nature contribute to the observed shape. Further cuts to eliminate the undesired contamination would also affect the signal, leading to an inconvenient loss of efficiency. Therefore, statistical methods are needed to isolate the distribution of signal events. In a sense, the problem is the inverse of a yield measurement, where, knowing all the probability density functions, the counts for different processes are extracted from a total number of events. Usually, this is done with a fit to the invariant mass, since mass shapes can be well modelled. The strategy relies on using the information on the dataset population, derived from the invariant mass fit, to disentangle the components and to unfold the background in variables other than the fit variable.

Intuitively, if the reconstructed invariant mass of an event lies in a region where the signal is dominant, it is more likely for that event to be signal. Conversely, if the signal component is suppressed in a particular invariant mass range, it is natural to consider the events falling in such a range as coming from background. These concepts can be translated into mathematical terms by weighting candidates according to the signal/background population at the corresponding invariant mass. A first approach was attempted in Ref. [75], where a weight for each event e with invariant mass m_e,

$$w_e = \frac{N_s f_s(m_e)}{\sum\limits_{k=1}^{M} N_k f_k(m_e)} \tag{6.1}$$

was proposed in terms of the invariant mass PDFs $f_k(m)$ and yields N_k for the signal (s) and the other $M-1$ components. It is shown in Ref. [74] that such weights reproduce the correct distribution for the signal component s only for those variables which can be expressed as a function of the invariant mass, or for which, in other words, an expression $y = y(m)$ holds exactly. This is generally not relevant for the purpose of background subtraction, as the reweighting procedure would be an unnecessary complication for a simple change of variables. Nevertheless, Eq. (6.1) is a good depiction of the concept: each event gets a weight according to the probability of being signal with respect to the probability of being something else. The concept is generalised and extended to quantities independent of the mass with the ${}_s\mathcal{P}lot$ technique [74]. In this extension the weights, called ${}_s\mathcal{W}eights$, take the form

$$
{}_s w_e = \frac{\sum\limits_{j=1}^{M} \mathrm{V}_{sj} f_j(m_e)}{\sum\limits_{k=1}^{M} N_k f_k(m_e)} \tag{6.2}
$$

which includes an additional term due to the covariance matrix V of the maximum likelihood fit performed to extract the yields N_k, defined by the inverse of the Hessian

$$
\mathrm{V}_{ij}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_i \partial N_j} \ . \tag{6.3}
$$

It is important to emphasise that the technique reproduces the correct distribution only for those observables independent of the mass, which is the variable on which the fit is performed. If, on the contrary, the considered observable is correlated with the invariant mass, Eq. (6.2) is strictly not applicable. This limitation must be carefully considered when applying the ${}_s\mathcal{P}lot$ method, keeping in mind that, as with BDT features, small correlations may still be acceptable without invalidating the result at a given precision.

6.2. Particle identification correction

Particle identification observables are complex combinations of information from different subdetectors. Small local differences between data and simulations can have a large impact on the final result. This is particularly true for the ProbNN variables, whose output comes from a neural network which could amplify potential discrepancies in the input. Moreover, the response of the RICH detectors, which are the main contributors to particle identification, cannot be completely modelled in simulation. As discussed in Sec. 3.2.2.1, the Cherenkov cones directly depend on the refractive index of the radiator gas, sensitive to pressure and temperature fluctuations, and the complex transport of the photons is hard to reproduce. It is well known that these difficulties result in a mismatch between the distributions of particle identification variables. To overcome this issue, the simulated single-track output is regenerated according to distributions determined from calibration data samples. The samples are chosen in order to be clean, easy to detect and abundant: for kaon tracks the decay $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\,\pi^+$ is considered, while the decay $J/\psi \rightarrow \mu^+\mu^-$ is used for muon tracks [76]. The ${}_s\mathcal{P}lot$ technique allows the determination of the real distributions of $\texttt{ProbNN}_K$ and $\texttt{ProbNN}_\mu$ for kaons and muons, respectively. Since these quantities are dependent on the track kinematics and the detector occupancy, the distributions are studied separately in different ranges of track transverse momentum (p_T) and pseudorapidity (η), and number of tracks for the total event (nTracks). These distributions are smoothed with a Gaussian filter within the pidgen2 package [77], which provides the *probability density functions* (PDFs). Two approaches are possible at this

Table 6.1.: Kinematic range covered by calibration samples

Track type	Kinematic region
μ tracks	$1.9 < \eta < 5.0, 800\,\mathrm{MeV/c} < p_\mathrm{T} < 2.5 \cdot 10^4\,\mathrm{MeV/c}$, $\texttt{nTracks} \in [5, 1260]$
K tracks	$1.9 < \eta < 5.0, 250\,\mathrm{MeV/c} < p_\mathrm{T} < 2.5 \cdot 10^4\,\mathrm{MeV/c}$, $\texttt{nTracks} \in [5, 1260]$

point. One of them consists of randomly generating the output from scratch, according to the given PDF $p(x|p_\mathrm{T}, \eta, \texttt{nTracks})$. This is done in the `pidgen2` package by inverting the cumulative function

$$P(x|p_\mathrm{T}, \eta, \texttt{nTracks}) = \int_{-\infty}^{x} p(y|p_\mathrm{T}, \eta, \texttt{nTracks})\,\mathrm{d}y \quad . \tag{6.4}$$

If ξ is uniformly distributed in the interval $[0, 1]$, the new value

$$x = P^{-1}(\xi|p_\mathrm{T}, \eta, \texttt{nTracks}) \tag{6.5}$$

is distributed according to the right PDF $p(x|p_\mathrm{T}, \eta, \texttt{nTracks})$. New values have no relation with the old ones, and since the generation is independent for the different variables, nor do they relate to the other quantities. In contrast, in the second approach, relations between old and new particle identification output are preserved by the substitution of the randomly generated ξ with the cumulative function of the simulated distribution $P_\mathrm{MC}(x_\mathrm{MC}|p_\mathrm{T}, \eta, \texttt{nTracks})$, computed at the simulated instance x_MC

$$x = P^{-1}(\,P_\mathrm{MC}(x_\mathrm{MC}|p_\mathrm{T}, \eta, \texttt{nTracks})\,|\,p_\mathrm{T}, \eta, \texttt{nTracks}\,) \quad . \tag{6.6}$$

By construction, this gives again the correct distribution, since the cumulative function $P_\mathrm{MC}(x_\mathrm{MC}|p_\mathrm{T}, \eta,$ $\texttt{nTracks})$ is uniformly distributed in $[0, 1]$. Moreover, if the cumulative functions $P(x|p_\mathrm{T}, \eta, \texttt{nTracks})$ and $P_\mathrm{MC}(x_\mathrm{MC}|p_\mathrm{T}, \eta, \texttt{nTracks})$, or analogously the PDFs $p(x|p_\mathrm{T}, \eta, \texttt{nTracks})$ and $p_\mathrm{MC}(x_\mathrm{MC}|p_\mathrm{T}, \eta,$ $\texttt{nTracks})$, are not too different, the variables x and x_MC are correlated, hence all old correlations between particle identification quantities are conserved. This is only possible if simulated samples of the calibration channels are available. Specifically for this analysis, Monte Carlo simulations are available for $D^{*+} \to D^0(\to K^-\pi^+)\,\pi^+$, for the Run 2 period only, but not for $J/\psi \to \mu^+\mu^-$. The second approach is then used to correct $\texttt{ProbNN}_K$ on Run 2 kaon tracks, while for Run 1 kaon and all muon tracks $\texttt{ProbNN}_K$ and $\texttt{ProbNN}_\mu$ are resampled using the first method. This forced choice does not cause any limitation, since particle identification observables are not taken into consideration as BDT features, and no product of them is considered when defining the selection. Both procedures described above are performed event by event, when tracks are contained in a kinematic region covered by the calibration sample. If, conversely, the event is outside the window described in Tab. 6.1, the uncorrected value is retained.

6.3. Kinematic reweighting

The distributions of the B meson transverse momentum (p_T) and pseudorapidity (η) are known to be poorly modelled in simulation when compared to real data. Since these kinematic properties of the B parent particle are crucial inputs to the multivariate algorithm trained against combinatorial background, they are corrected using a data-driven reweighting approach. The kinematic reweighting aims to correct

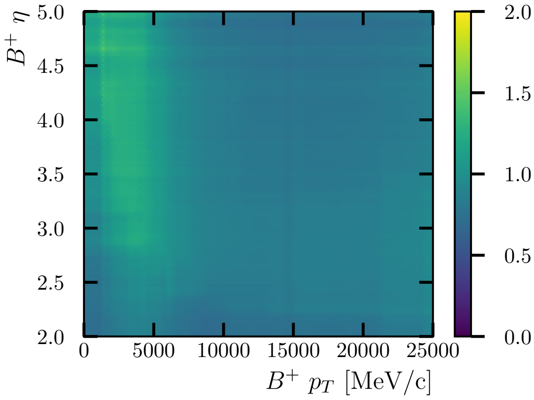


Figure 6.1.: Response of the kinematic reweighter for 2016 samples.

for the residual differences between distributions of kinematic variables by assigning a weight to each simulated event. While it is not possible to use the signal channels $B^{0,+} \longrightarrow K^{0,+}\mu^+\mu^-$ for this, due to the analysis being blinded, a proxy decay is used. A convenient choice is the resonant channel $B^+ \longrightarrow J/\psi K^+$ which shares the B kinematics with the rare counterpart. No particular difference is expected in the kinematics of B^+ and B^0 mesons. Therefore, the same calibration mode can be used for both charged and neutral cases. Again, the *sPlot* technique allows the determination of the distributions in data, selected in the resonant q^2 range $[8.0, 11.0]\ \mathrm{GeV}^2/\mathrm{c}^4$, of the two considered quantities p_T and η. These have to be compared with the ones obtained from simulated $B^+ \longrightarrow J/\psi\, K^+$ decays. Having two histograms, $h_D(p_\mathrm{T}, \eta)$ for data and $h_\mathrm{MC}(p_\mathrm{T}, \eta)$ for simulations, the correction can be obtained from the division of the two histograms

$$w = \frac{h_\mathrm{D}(p_\mathrm{T}, \eta)}{h_\mathrm{MC}(p_\mathrm{T}, \eta)} \tag{6.7}$$

in two-dimensional intervals of p_T and η. This would however depend on the definition of the intervals and it is difficult to define an optimal one a priori. Another approach to avoid this issue would be to consider a set of weights, one for each event, and adjust them until the two distributions are the same. This is possible by employing machine learning algorithms, and in particular by a *Boosted Decision Tree Reweighter* [78]. As the name suggests, a BDT like the one already described in Sec. 5.4 looks for differences between the two samples. Specifically, in this case, a region which maximises the separation is found in the two-dimensional space of the features. Events are counted and the ratio of the two numbers N_D/N_MC is assigned as an initial weight to the simulated events falling in the region. Then, the procedure is repeated keeping all the weights previously assigned. The region determination could be seen as a generalisation of the interval definition for the two-dimensional histograms. In this interpretation, the event ratio performed by the reweighter reduces to Eq. (6.7), iteratively applied. Looking for the region with the largest separation, where larger weights are expected, ensures proceeding by successive approximations, converging to the correct set. For this analysis, a BDT reweighter is implemented using the `hep_ml` Python package [79], separately for each year of data-taking. An example result for the two-dimensional response of the reweighter is given in Fig. 6.1 for 2016. Similar results are obtained for the other years.

6.4. L0 trigger correction

The trigger thresholds for the `L0Muon` and `L0DiMuon` trigger lines are usually not aligned across data and simulation for a given data-taking period. The resulting differences in trigger efficiency can have an influence on the calculation of selection efficiencies using simulated samples and therefore they need to be corrected on the basis of results in data. The trigger defines what is recorded and what is not, and there is no way to recover information that was discarded at this stage. In particular, it is not possible to know the number of total events entering the acceptance of the detector, hence the denominator in Eq. (5.3).

At LHCb, a strategy called TISTOS method [80] was developed to reconstruct trigger efficiencies which can be applied both to real and simulated data. It makes use of trigger categories introduced in Sec. 5.2.1, namely TOS, indicating those events triggered on signal tracks, and TIS, where another part of the event, different from the signal, was deemed relevant, most likely for other purposes. For the analysis, TOS events are considered, but the two categories are not exclusive. A single event in its different parts can turn out to be interesting for independent studies. If the two categories are independent, it is possible to switch to the conditioned efficiency

$$\frac{N_{\mathrm{Trig}}}{N} \equiv \varepsilon_{\mathrm{Trig}} = \varepsilon_{\mathrm{TOS|TIS}} = \frac{N_{\mathrm{TISTOS}}}{N_{\mathrm{TIS}}} \tag{6.8}$$

removing the dependency on the number of total events N. Here N_{TISTOS} is indeed the number of events passing both TIS and TOS lines and is measurable in data like the number of TIS events, the aforementioned difficulty being completely resolved. It is worth noting that the hypothesis of independence of the two categories is in general wrong. Even if belonging to different parts, tracks originate from the same event with a fixed energy and at least their kinematics will be related. Nevertheless, this relation is usually lost when applying TIS and TOS, if no particular effect occurs. It is then crucial to properly choose the lines in order to avoid such issues.

Given a method to compute efficiencies in data, a calibration sample is needed. As in the case of kinematic reweighting, the resonant decay $B^+ \to J/\psi K^+$, selected in the q^2 region $[8.0, 11.0]\ \mathrm{GeV}^2/\mathrm{c}^4$, can be used. Trigger efficiencies are computed both in data and simulations, and a correction is assigned to the latter consisting of the ratio

$$w_{\mathrm{Trig}} = \frac{\varepsilon^{\mathrm{Data}}_{\mathrm{TOS|TIS}}}{\varepsilon^{\mathrm{MC}}_{\mathrm{TOS|TIS}}} . \tag{6.9}$$

The `L0Muon` and `L0DiMuon` lines depend on muon track momenta. Therefore, the efficiencies are studied in ranges of muon p_{T}. However, this cannot be done separately for the two muons, `L0DiMuon` introducing a coupling between the tracks. This can be taken into account by considering two-dimensional intervals, but another complexity in the $p_{\mathrm{T}}(\mu^+)$–$p_{\mathrm{T}}(\mu^-)$ plane is due to the J/ψ. The resonance, in fact, fixes the energy of the dimuon pair, in such a way that if a muon has a large momentum the other will have a small one and vice versa. The distribution of events in the plane, visible in the scatter plot in Fig. 6.2, is not uniform and rectangular ranges would result in few highly populated regions with small uncertainties and several empty regions with large fluctuations. To avoid this, equally populated intervals are found with a recursive algorithm, in each step splitting the whole range into two parts along one dimension, and alternating the axis along which the split is applied. Correction weights are computed according to Eq. (6.9), separately for each year of data-taking, and after applying particle identification variable

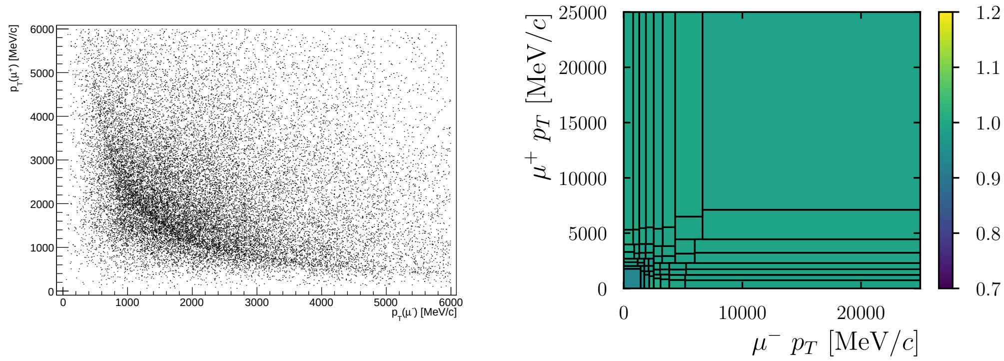


Figure 6.2.: Scatter plot of simulated $B^+ \to J/\psi K^+$ sample in the plane $p_{\mathrm{T}}(\mu^+) - p_{\mathrm{T}}(\mu^-)$ (left) and distribution of trigger correction weights for the year 2017 (right).

resampling and kinematic reweighting. An example result is given in Fig. 6.2 for 2017. Similar weights are obtained for the other years.

6.5. Multiplicity reweighting

Quantities related to the detector occupancy are poorly modelled in simulation. Among them, the number of reconstructed tracks in the detector, `nTracks`, is important for the correction of the particle identification variables and requires correction itself. Weights w_{Mult} are then calculated by comparing the distribution of `nTracks` in ${}_s\mathcal{W}eighted$ data and simulation in a histogram-based approach, i.e. applying Eq. (6.7), in this case in one dimension only

$$w_{\mathrm{Mult}} = \frac{h_{\mathrm{D}}(\texttt{nTracks})}{h_{\mathrm{MC}}(\texttt{nTracks})} \quad . \tag{6.10}$$

In principle, this is exactly the same problem as the kinematic reweighting, and another BDT reweighter could be introduced here as well. However, while in higher dimensions the improvement is significant, the benefits for a one-dimensional correction are considered insufficient to justify the use of such an advanced tool. Also in this case, calibration data consist of samples of resonant decays. However, since the number of tracks depends on the specific decay topology, the $B^+ \to J/\psi K^+$ channel, while suitable for studying the corrections for the $B^+ \to K^+\mu^+\mu^-$ mode, cannot be used for the $B^0 \to K^0_{\mathrm{S}}\mu^+\mu^-$ mode. The mere presence of the K^0_{S} instead of the K^+ introduces an additional track. Moreover, the remainder of the event, associated with a neutral particle instead of a charged one, is generally different. For this reason, the $B^0 \to J/\psi K^0_{\mathrm{S}}$ decay is taken as a reference for the neutral mode, studying separately LL and DD-type tracks. Selection is specific to the channel and described in Ch. 5, but in general the resonance is selected in the q^2 region $[8.0, 11.0]$ GeV^2/c^4. Particle identification resampling, kinematic correction weights, as well as the L0 correction weights are applied to the simulations before calculating the multiplicity weights w_{Mult}. Fifty `nTracks` ranges, which define the histograms in Eq. (6.10), are chosen in order to be equally populated. As for the other steps, the corrections are studied separately for each year of data-taking. An example of the obtained weights for 2012 is shown in Fig. 6.3. Similar weights are obtained for the other years.

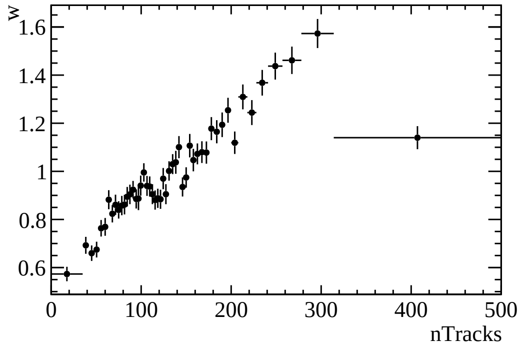


Figure 6.3.: Example set of multiplicity weights valid for B^0 as obtained in 2012.

6.6. Agreement between data and simulation

In the previous sections, several methods to correct different aspects of the simulated data have been described. They are all necessary to obtain a good agreement between data and simulations for the relevant features in this analysis. All the corrections are applied as efficiency weights, except for those related to particle identification variables, for which corrected values are directly available through the `pidgen` package. This means that to apply the full correction chain, simulated samples must first be selected following the instructions given in Ch. 5, but using the corrected variables when considering the particle identification cuts. Then, when studying a distribution or when training the multivariate algorithm on simulated samples, the histograms are weighted with the product of all the weights, i.e. each event has to be counted a number of times equal to the product of the weights. To validate the effectiveness of the chain, the distributions of uncorrected and corrected simulated data can be studied. The calibration samples again provide a robust reference. Comparisons of $B^0 \to J/\psi K^0_S$ and $B^+ \to J/\psi K^+$ simulations with the corresponding ${}_s\mathcal{W}eighted$ data samples are shown in Fig. 6.4 and Fig. 6.5, respectively, for several observables. Good agreement is observed when applying the full correction chain, not only for the quantities directly corrected, such as the B kinematics, the particle identification variables and the multiplicity, but also for related quantities like the secondary track kinematics. Residual irreducible differences are present but they are evaluated as acceptable, appearing mostly in regions with limited statistics. Moreover, residual differences cannot be excluded in distributions not considered in this check. However, they would not affect the analysis since the variables they refer to are not used in any part.

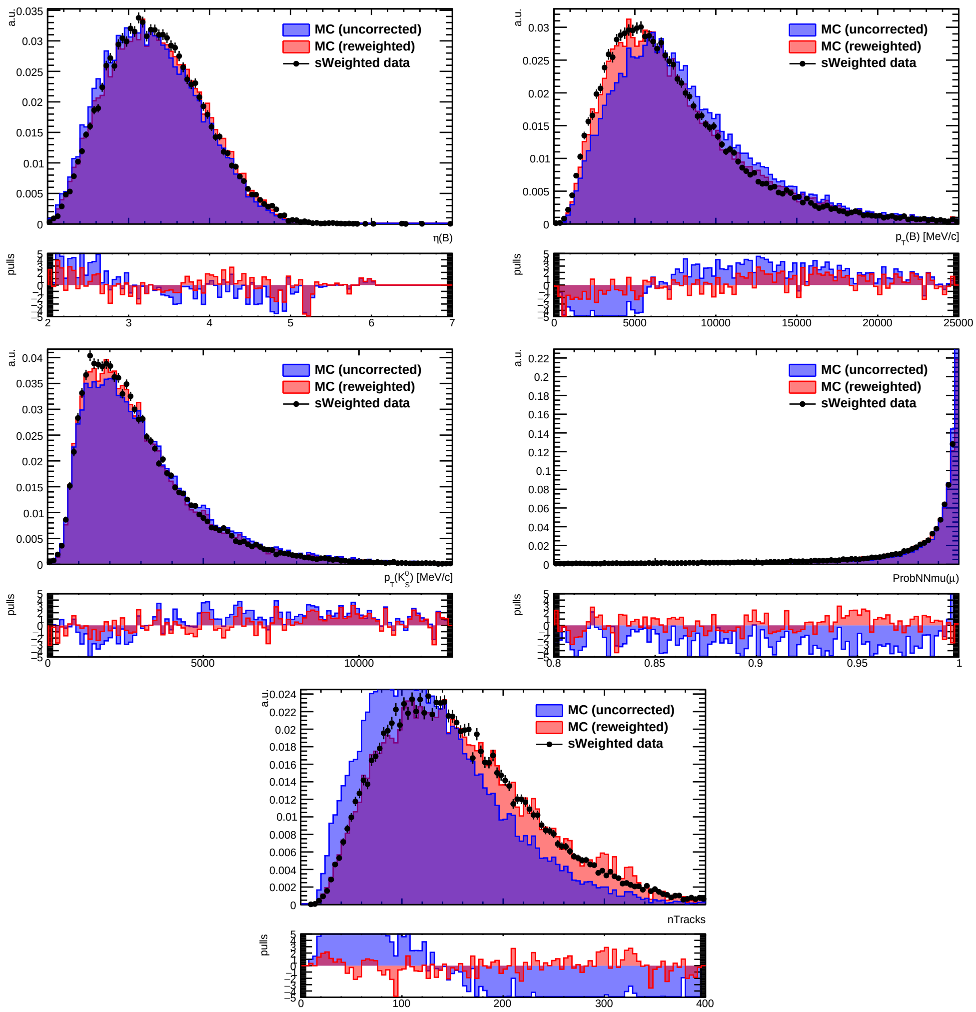


Figure 6.4.: Comparison between $B^0 \to J/\psi K^0_\mathrm{S}$ uncorrected and corrected simulations and background-subtracted data for B^0 pseudorapidity (top-left), B^0 transverse momentum (top-middle), K^0_S transverse momentum (top-right), μ $\texttt{ProbNN}_\mu$ (bottom-left) and event $\texttt{nTracks}$ (bottom-right) distributions. For a better visualisation, the pulls, defined as the difference between the simulation and data histograms normalised by the uncertainty, are shown for both uncorrected and corrected simulations at the bottom of each plot. The agreement between data and simulations improves upon applying the full correction chain.

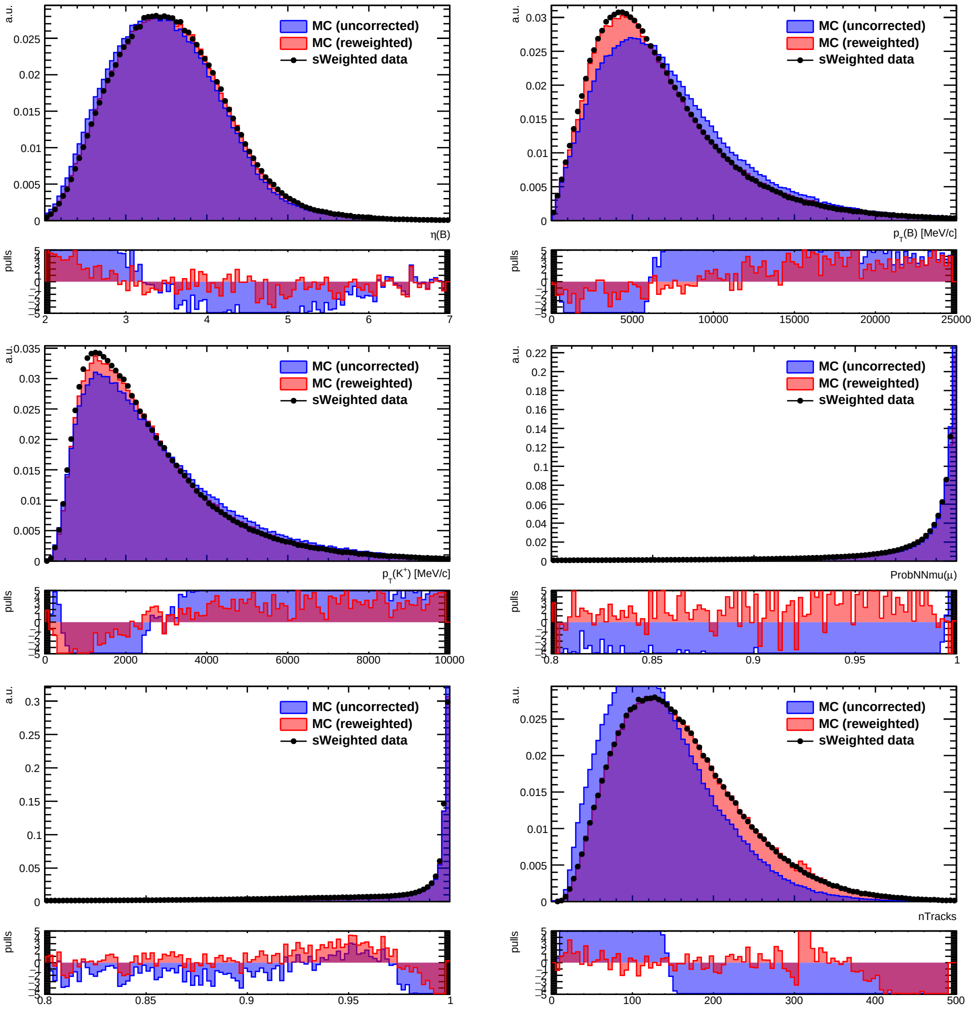


Figure 6.5.: Comparison between $B^+ \to J/\psi K^+$ uncorrected and corrected simulations and background-subtracted data for B^+ pseudorapidity (top-left), B^+ transverse momentum (top-middle), K^+ transverse momentum (top-right), μ $\texttt{ProbNN}_\mu$ (bottom-left), K^+ $\texttt{ProbNN}_K$ (bottom-middle) and event `nTracks` (bottom-right) distributions. For a better visualisation, the pulls, defined as the difference between the simulation and data histograms normalised by the uncertainty, are shown for both uncorrected and corrected simulations at the bottom of each plot. The agreement between data and simulations improves upon applying the full correction chain.

7. Invariant mass fit

In order to extract the signal yields, which are inputs for the computation of both the branching fraction and, according to Eq. (2.20), the isospin asymmetry, an unbinned maximum likelihood fit to the reconstructed B mass is performed. While the quantities of interest are directly related to the rare channels $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$, the resonant decays $B^0 \to J/\psi K^0_S$ and $B^+ \to J/\psi K^+$ are also analysed to design the signal shape models and constrain the fit parameters for the rare modes, as well as for normalisation purposes. For all modes, the common B-mass range $[5170, 5700]\ \mathrm{MeV/c^2}$ is used. This range is specifically chosen to exclude the lower mass region where partially reconstructed backgrounds, most notably from $B \to K^*\mu^+\mu^-$ decays, would otherwise dominate, allowing for a simplified background model. In this chapter, the models for the different mass fits are discussed in detail.

7.1. $B^0 \to J/\psi K^0_S$

The fit to the $B^0 \to J/\psi K^0_S$ decay serves multiple purposes. Firstly, the resonant decay is a normalisation mode for this analysis, and its yield is required to convert the signal yield of the rare decay into a branching fraction, which is the fundamental quantity entering the isospin asymmetry calculation. Secondly, the invariant mass shape parameters for $B^0 \to K^0_S\mu^+\mu^-$ are constrained using this mode, after considering differences between resonant and rare decays on simulated samples. Additionally, the same mass fit is used to compute ${}_s\mathcal{W}eights$ for multiplicity corrections already discussed in Ch. 6. Data are selected according to the criteria in Ch. 5, with the exception of the q^2 veto, which is inverted to cover the resonant region $[8.0, 11.0]\ \mathrm{GeV^2/c^4}$. After the selection, the sample is composed of three components: the signal decay $B^0 \to J/\psi K^0_S$, the subdominant contribution from $B^0_s \to J/\psi K^0_S$ and the residual combinatorial background. Each component exhibits a distinct behaviour in the invariant mass spectrum.

The reconstruction of the exclusive $J/\psi K^0_S$ final state results in a prominent signal peak centred at the nominal B^0 mass. The observed width is dominated by the detector momentum resolution, while the selection criteria are designed to preserve the natural shape of the resonance. The chosen model to parametrise this shape is a double-sided Crystal Ball function [81], defined as the sum of two single Crystal Ball functions with a shared mean m_B

$$
\begin{aligned}
f_\mathrm{L}(m|m_B, \sigma_\mathrm{L}, \alpha, n) &= N_\mathrm{L} \cdot \begin{cases} A_\mathrm{L}\left(B_\mathrm{L} - \frac{m-m_B}{\sigma_\mathrm{L}}\right)^{-n} & , \quad \text{for } \frac{m-m_B}{\sigma_\mathrm{L}} \leq \alpha \\ e^{-\frac{(m-m_B)^2}{\sigma_\mathrm{L}^2}} & , \quad \text{for } \alpha < \frac{m-m_B}{\sigma_\mathrm{L}} \end{cases} \\
f_\mathrm{R}(m|m_B, \sigma_\mathrm{R}, \beta, k) &= N_\mathrm{R} \cdot \begin{cases} e^{-\frac{(m-m_B)^2}{\sigma_\mathrm{R}^2}} & , \quad \text{for } \frac{m-m_B}{\sigma_\mathrm{R}} < \beta \\ A_\mathrm{R}\left(B_\mathrm{R} - \frac{m-m_B}{\sigma_\mathrm{R}}\right)^{-k} & , \quad \text{for } \frac{m-m_B}{\sigma_\mathrm{R}} \geq \beta \end{cases}
\end{aligned}
\tag{7.1}
$$

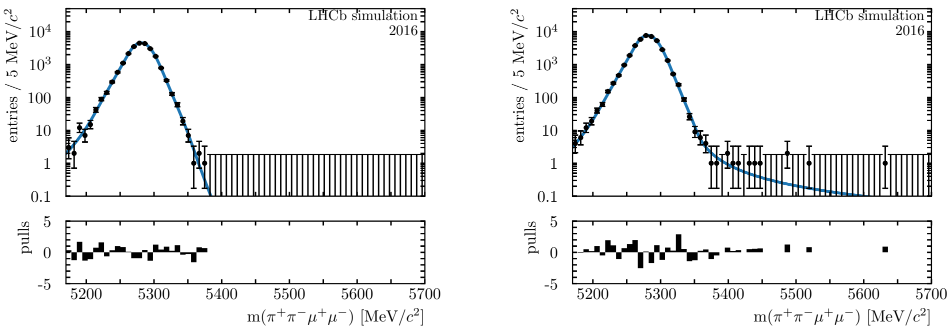


Figure 7.1.: Fit to $B^0 \rightarrow J/\psi K^0_\mathrm{S}$ corrected simulation to determine the shape parameters of the signal peak in 2016 for LL-type (left) and DD-type (right). The data points (black) are overlaid with the best fit result (blue line). The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

where $A_\mathrm{L,R}$ and $B_\mathrm{L,R}$ are constants to ensure continuity and differentiability

$$
\begin{aligned}
A_\mathrm{L} &= \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{\alpha^2}{2}} \quad , \qquad & A_\mathrm{R} &= \left(\frac{k}{|\beta|}\right)^k e^{-\frac{\beta^2}{2}} \\
B_\mathrm{L} &= \frac{n}{|\alpha|} - |\alpha| \quad , \qquad & B_\mathrm{R} &= \frac{k}{|\beta|} - |\beta|
\end{aligned}
\tag{7.2}
$$

and $N_\mathrm{L,R}$ are normalisation factors. Eq. (7.1) demonstrates the versatility of this function. A Gaussian core centred around the mass of the B meson m_B, on a mass range large $\sigma(\beta - \alpha)$, describes the mass peak. Outside the central region, left and right tails are separately modulated, through n and k, to take into account energy loss or other non-Gaussian effects. An additional parameter ρ governs the relative contribution of the two Crystal Ball functions, effectively balancing the left and right tails:

$$
f_\mathrm{DCB} = \rho f_\mathrm{L} + (1 - \rho) f_\mathrm{R} \ . \tag{7.3}
$$

The function fits the $B^0 \rightarrow J/\psi K^0_\mathrm{S}$ simulated data well, as shown in Fig. 7.1 in the case of 2016, for both LL and DD-type tracks. The corresponding plots for the other years are presented in App. A. Besides demonstrating the quality of the model, the fit to simulated data serves to determine the shape parameters for the subsequent fit to data. The values extracted for all data-taking years are summarised in Tab. 7.1 and Tab. 7.2 for LL and DD track categories, respectively. Specifically, the tail parameters α, n, β and k are fixed to the values obtained from simulation. To account for potential discrepancies in mass resolution between data and simulation, the widths are not fixed. Instead, a global scaling factor s_MC is introduced:

$$
\sigma_\mathrm{L,R} = s_\mathrm{MC} \cdot \sigma^\mathrm{(MC)}_\mathrm{L,R} \ . \tag{7.4}
$$

This factor is left floating in the fit to data, allowing the resolution to vary while preserving the ratio σ_L/σ_R determined from simulation. Similarly, possible differences in energy calibration are taken into account treating the mean mass m_B as a free parameter in the fit to data.

The same double-sided Crystal Ball shape is used to model the $B^0_s \rightarrow J/\psi K^0_\mathrm{S}$ background. Given the identical final state and very similar kinematics, the shape parameters are shared with the signal

Table 7.1.: Signal shape parameters determined from the fit to $B^0 \to J/\psi K^0_S$ LL-type simulated samples for different years. These values are used to fix the tail parameters in the fit to data. Uncertainties from the fit are not reported as not considered in the following.

Year	m_B [MeV/c^2]	σ_L [MeV/c^2]	σ_R [MeV/c^2]	α	β	n	k	ρ
2011	5280.7	14.3	23.8	−1.5	3.5	12	2.2	0.78
2012	5280.9	13.7	22.5	−1.4	3.4	10	25	0.71
2015	5281.6	13.3	22.2	−1.3	5.4	16	9	0.69
2016	5281.2	14.4	23.4	−1.5	3.2	9	24	0.80
2017	5281.4	14.0	23.2	−1.3	3.3	16	2.7	0.78
2018	5281.7	14.0	23.1	−1.3	6.2	25	14	0.79

Table 7.2.: Signal shape parameters determined from the fit to $B^0 \to J/\psi K^0_S$ DD-type simulated samples for different years. These values are used to fix the tail parameters in the fit to data. Uncertainties from the fit are not reported as not considered in the following.

Year	m_B [MeV/c^2]	σ_L [MeV/c^2]	σ_R [MeV/c^2]	α	β	n	k	ρ
2011	5280.5	14.1	23.1	−1.5	3.2	8.7	1.2	0.79
2012	5280.7	13.9	22.2	−1.4	3.2	9.0	1.6	0.76
2015	5281.3	13.9	21.9	−1.2	3.3	20.8	1.4	0.77
2016	5281.1	14.5	23.7	−1.5	3.0	10	1.7	0.82
2017	5281.1	14.0	22.2	−1.4	3.0	9.1	1.9	0.77
2018	5281.0	14.3	22.8	−1.4	2.8	9.3	2.6	0.81

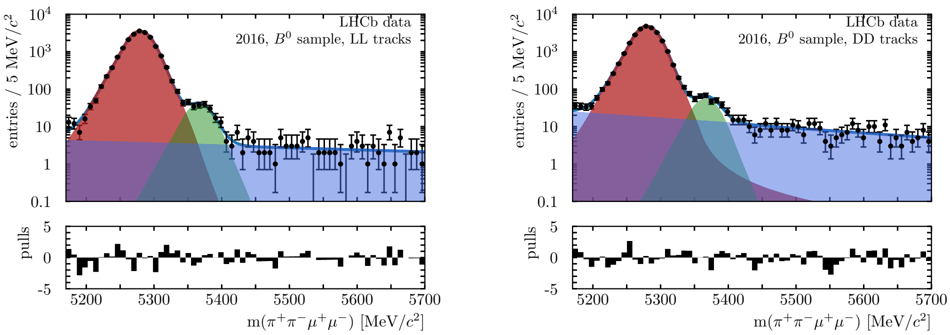


Figure 7.2.: Fit to $B^0 \to J/\psi K^0_S$ data, LL-type (left) and DD-type (right), for 2016. In the plots, the signal component (red), the $B^0_s \to J/\psi K^0_S$ background (green), and the combinatorial background (blue) are shown, together with the total fit function. The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty. The resulting $B^0 \to J/\psi K^0_S$ yields are used for normalisation, while the model as fixed in this fit is used for the rare mode $B^0 \to K^0_S\mu^+\mu^-$.

component. The only exception is the mean of the Gaussian core, which is defined through a fixed shift relative to the B^0 peak. The mass difference $m(B^0_s)-m(B^0)$ is computed from Ref. [72] to be $87\,\mathrm{MeV}/c^2$. Due to the small magnitude of this difference, the energy dependence of the resolution is negligible and therefore the width is not further scaled in the fit.

The combinatorial background, on the contrary, is of a completely different nature. It arises from random combinations of tracks, which are kinematically more probable at lower invariant mass. A high-energy combination, in fact, would need tracks that are all highly energetic. This happens less frequently at LHCb and consequently, the mass spectrum exhibits a decreasing trend. There are no theoretical motivations to prefer one model to another, but the exponential function is often shown to work well in this context, having a good flexibility despite its single parameter. A decreasing exponential is therefore chosen to model the combinatorial background, with the slope parameter constrained to be negative and left free to vary in the fit to data.

The complete model for the $B^0 \to J/\psi K^0_S$ data sample consists of the sum of all the functions described above. The free parameters in the fit to data are the B^0 mass m_B, the scaling factor s_{MC} for the widths of the two double-sided Crystal Balls, and the slope of the combinatorial exponential, as well as the yields for the three components. The fit is performed separately on LL and DD-type tracks, for each year of data-taking. Example results are shown in Fig. 7.2 for the 2016 dataset, for both LL and DD-type samples. The obtained shape for the $B^0 \to J/\psi K^0_S$ channel, summarised in Tab. 7.3 and Tab. 7.4, respectively, is applied to the rare mode $B^0 \to K^0_S\mu^+\mu^-$ after some corrections discussed in the following section. The $B^0 \to J/\psi K^0_S$ yields are used in Ch. 8 as the normalisation input to extract the quantities of interest.

Table 7.3.: Additional parameters determined from the fit to $B^0 \to J/\psi K^0_\mathrm{S}$ LL-type data samples for different years. The mean mass m_B and the scale s_MC, together with the parameters from Tab. 7.1, are used for the fit to the rare channel. Signal yields $N^{B^0}_{J/\psi}$ serve as normalisation input. Uncertainties from the fit are not reported as not considered in the following (with the exception of the ones on $N^{B^0}_{J/\psi,LL}$ which will be properly considered in Ch. 9).

Year	m_B [MeV/c^2]	s_MC	$N^{B^0}_{J/\psi,LL}$	$N_{B^0_s}$	N_comb
2011	5284.5	1.10	$6.65 \cdot 10^3$	70	120
2012	5285.0	1.09	$1.42 \cdot 10^4$	100	420
2015	5281.0	1.14	$3.26 \cdot 10^3$	30	50
2016	5279.6	1.13	$1.98 \cdot 10^4$	200	200
2017	5280.0	1.09	$2.31 \cdot 10^4$	260	340
2018	5280.0	1.10	$2.68 \cdot 10^4$	280	350

Table 7.4.: Additional parameters determined from the fit to $B^0 \to J/\psi K^0_\mathrm{S}$ DD-type data samples for different years. The mean mass m_B and the scale s_MC, together with the parameters from Tab. 7.2, are used for the fit to the rare channel. Signal yields $N^{B^0}_{J/\psi}$ serve as normalisation input. Uncertainties from the fit are not reported as not considered in the following (with the exception of the ones on $N^{B^0}_{J/\psi,DD}$ which will be properly considered in Ch. 9).

Year	m_B [MeV/c^2]	s_MC	$N^{B^0}_{J/\psi,DD}$	$N_{B^0_s}$	N_comb
2011	5284.6	1.11	$8.47 \cdot 10^3$	90	320
2012	5284.8	1.10	$1.86 \cdot 10^4$	200	660
2015	5280.9	1.11	$3.65 \cdot 10^3$	30	150
2016	5279.7	1.11	$2.63 \cdot 10^4$	280	780
2017	5280.0	1.09	$3.32 \cdot 10^4$	360	1000
2018	5280.7	1.09	$4.02 \cdot 10^4$	450	1300

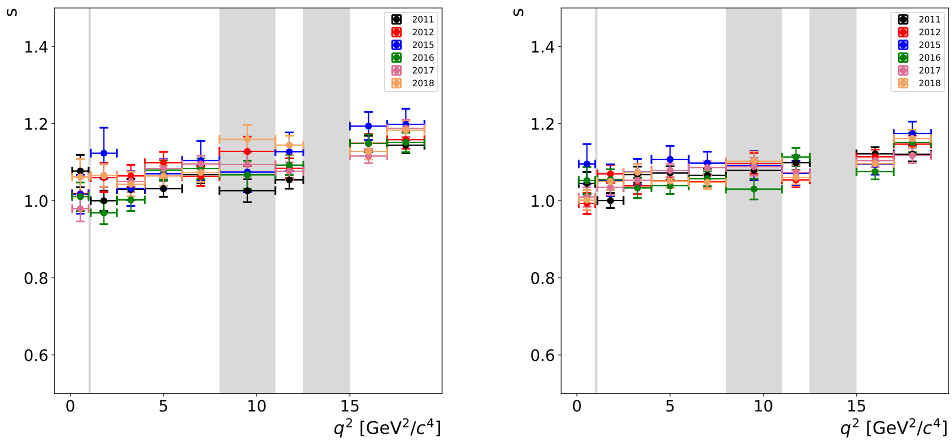


Figure 7.3.: Scale factors as obtained from a fit to $B^0 \to K^0_S\mu^+\mu^-$ simulated data, LL (left) and DD-type (right), as a function of q^2. For reference, scale factors are also given in the resonant region $[8.0, 11.0]\ \mathrm{GeV}^2/\mathrm{c}^4$, but they are not used in the following.

7.2. $B^0 \to K^0_S\mu^+\mu^-$

The yields of the rare channels are the primary input to extract the branching fractions and consequently the isospin asymmetry. A maximum likelihood fit, adapted from the resonant mode ($B^0 \to J/\psi K^0_S$), is performed on the $B^0 \to K^0_S\mu^+\mu^-$ data, separately in different intervals of the dimuon invariant mass squared, q^2. In this case, after the selection, only two components populate the sample, namely the signal decay $B^0 \to K^0_S\mu^+\mu^-$ and the residual combinatorial background.

The signal is very similar to the resonant counterpart $B^0 \to J/\psi K^0_S$, a peaking structure around the B^0 mass which is modelled by the same double-sided Crystal Ball presented in the last section: Eq. (7.1) with the parameters reported in Tabs. 7.1–7.4. Given the kinematic similarity of the two decays, the shape parameters governing the tails are shared over the whole q^2 range. However, special attention is required for the width, which governs the high-statistics region of the signal. Since the mass resolution depends on the energy release, the width is intrinsically dependent on q^2. To account for this, a specific resolution scale factor s_i is introduced for each q^2 bin, denoted by the index i. Defined analogously to Eq. (7.4), this factor scales the width relative to the resonant mode:

$$\sigma_{i\ \mathrm{L,R}} = s_i\ \sigma^{(J/\psi)}_{\mathrm{L,R}}\ . \tag{7.5}$$

The scale factors s_i embody differences between the rare and resonant simulated samples. They are determined from a fit to simulated $B^0 \to K^0_S\mu^+\mu^-$ events where all other parameters are fixed from $B^0 \to J/\psi K^0_S$ simulations. This is done separately for LL and DD-type tracks, for each year of data-taking, and results are given in Fig. 7.3. In the figure, scale factors are also provided for the resonant region $[8.0, 11.0]\ \mathrm{GeV}^2/\mathrm{c}^4$ for reference. However, these are not used in the final model. This q^2-dependent correction is applied in addition to the global data-simulation resolution scaling derived in the previous

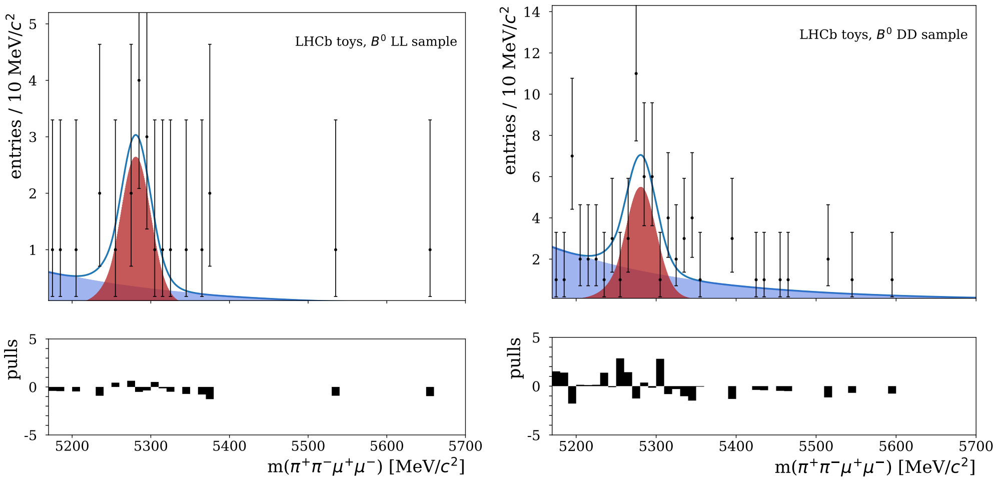


Figure 7.4.: Projections of the invariant mass fit to a $B^0 \to K^0_S\mu^+\mu^-$ toy dataset for an arbitrary chosen q^2 bin ($[0.1, 0.98]$ GeV^2/c^4), for LL-type (left) and DD-type (right) tracks. The signal component is shown in red and the combinatorial background in blue. The pulls at the bottom confirm the validity of the model.

section. Consequently, the actual width used for the $B^0 \to K^0_S\mu^+\mu^-$ data fit in the i-th q^2 interval is given by:

$$\sigma_{i\ \mathrm{L,R}} = s_i \cdot s_{\mathrm{MC}} \cdot \sigma_{\mathrm{L,R}}^{(J/\psi,\,\mathrm{MC})} \quad . \tag{7.6}$$

Finally, the combinatorial background is modelled by an exponential function with a free slope.

While the fit model for the rare mode is fully defined, signal yields are blinded to avoid biases during the analysis preparation. Therefore, it is not possible to show the fit projections in the signal region. However, the fit performance is illustrated using a representative pseudo-experiment, called toy dataset. This sample is generated based on the probability density functions defined above, assuming the Standard Model branching fraction for the signal and a combinatorial background level estimated from the sidebands. The projections of the fit to this toy dataset are shown in Fig. 7.4 for one q^2 bin.

7.3. $B^+ \to J/\psi K^+$

The same strategy used for the neutral case, discussed in the last two sections, can be applied with a few modifications to the charged modes. Also in this case, the normalisation and the shape for the fit to the rare mode are fixed on the resonant counterpart, here $B^+ \to J/\psi K^+$. Furthermore, this channel is crucial for the calibration of simulated samples, entering kinematic, L0 trigger and multiplicity corrections described in Ch. 6, as the fit described in the next lines was implicitly used when computing ${}_s\mathcal{W}eights$. Resonance in data is ensured by inverting the q^2 veto in the selection described in Ch. 5 to cover the region $[8.0, 11.0]$ GeV^2/c^4. Three components can be recognised in the selected sample: the signal decay $B^+ \to J/\psi K^+$, the misidentified $B^+ \to J/\psi\ \pi^+$ background where the pion is reconstructed as a kaon, and the residual combinatorial background.

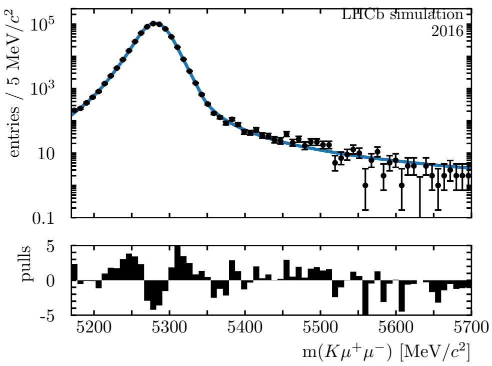


Figure 7.5.: Fit to $B^+ \to J/\psi K^+$ corrected simulation to determine the shape parameters of the signal peak in 2016. The data points (black) are overlaid with the best fit result (blue line). The bottom panel displays the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

Table 7.5.: Signal shape parameters determined from the fit to $B^+ \to J/\psi K^+$ simulated samples for different years. These values are used to fix the tail parameters in the fit to data. Uncertainties from the fit are not reported as not considered in the following.

Year	m_B [MeV/c^2]	σ_L [MeV/c^2]	σ_R [MeV/c^2]	σ_G	α	β	n	k	ρ_1	ρ_2
2011	5280.4	14.2	24.7	14.1	−0.95	2.6	6.5	1.7	0.48	0.52
2012	5280.4	14.3	23.1	14.1	−1.1	2.6	5.1	1.7	0.51	0.59
2015	5281.0	13.6	23.9	14.7	−0.96	2.6	6.6	1.6	0.55	0.55
2016	5281.0	14.6	23.8	14.2	−0.89	2.5	12	1.6	0.52	0.51
2017	5280.9	14.2	22.9	14.0	−0.89	2.5	9.1	1.7	0.49	0.55
2018	5280.9	14.6	23.4	14.0	−1.0	2.5	7.9	1.7	0.58	0.60

Similar to the $B^0 \to J/\psi K^0_S$ decay, the $B^+ \to J/\psi K^+$ signal appears as a peaking structure in the mass spectrum. While the physical effects determining the distribution are analogous, a simple double-sided Crystal Ball (Eq. (7.1)) was found to be insufficient to model the charged decay with the required precision. Based on simulation studies, a sum of a double-sided Crystal Ball function and a Gauss distribution, sharing the mean, was chosen as the final fit model. Two different parameters, ρ_1 and ρ_2, govern the relative contribution of the two Crystal Ball (f_L and f_R) and the Gaussian (f_G) functions:

$$f_{DCB+G} = \rho_2\,(\rho_1 f_L + (1-\rho_1) f_R) + (1-\rho_2) f_G \ . \tag{7.7}$$

The agreement between the fitted model and simulated data is illustrated in Fig. 7.5 for 2016. The corresponding plots for the other years are presented in App. A. The procedure is the same as before, parameters are fixed from a fit to the simulated $B^+ \to J/\psi K^+$ sample with the exception of the mean, which is kept free in the fit to data. The values extracted for all data-taking years are summarised in Tab. 7.5. A single

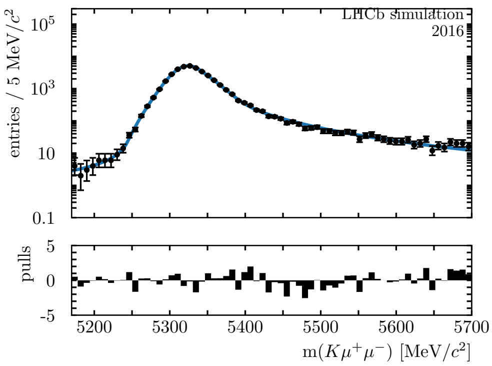


Figure 7.6.: Fit to $B^+ \to J/\psi\ \pi^+$ simulation to determine the shape parameters of the misidentified background in 2016. The data points (black) are overlaid with the best fit result (blue line). The bottom panel displays the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

scale factor for all the three widths, σ_L and σ_R for the Crystal Ball, σ_G for the additional Gaussian, is also introduced such that, analogously to Eq. (7.4), the parameters used for the fit to data are

$$\sigma_{\mathrm{L,R,G}} = s_{\mathrm{MC}}\ \sigma^{(\mathrm{MC})}_{\mathrm{L,R,G}}\ , \tag{7.8}$$

with s_{MC} free parameter in the fit to data.

A distinct peaking structure in the mass spectrum is due to the $B^+ \to J/\psi\ \pi^+$ background, when a pion is misidentified as a kaon before being combined with the two muon tracks. As a larger mass is assigned to the pion, the reconstructed energy of the process is artificially increased, resulting in a peak moved toward higher masses. Moreover the misidentification distorts the distribution making the solution of a shifted model, like the one introduced for $B^0_s \to J/\psi K^0_{\mathrm{S}}$, inapplicable. A new, independent double-sided Crystal Ball function is used to model this background. Its parameters are determined from a specific $B^+ \to J/\psi\pi^+$ simulation, as shown in Fig. 7.6. All parameters are free when fitting to simulated samples and fixed to the obtained results in Tab. 7.6 when fitting to data. Usual exceptions are the misidentified mean $m^{\prime}_B$, for which the difference with the signal mean is fixed, and the widths that are modified by the same free scale factor of Eq. (7.8). Finally, the combinatorial background is modelled by a decreasing exponential with free slope.

The complete model for the $B^+ \to J/\psi K^+$ data sample is the sum of all the functions presented in this section. As already mentioned, the free parameters are the mass of the B^+ (m_B), the scaling factor (s_{MC}) correcting all the widths and the slope of the combinatorial exponential, as well as yields for the three components. The fit is performed separately for each year of data-taking and an example is shown in Fig. 7.7 for 2016. The corresponding plots for the other years are presented in App. A. The obtained shape for the $B^+ \to J/\psi K^+$ channel, summarised in Tab. 7.7 for all the years, is applied to the rare mode $B^+ \to K^+\mu^+\mu^-$ after some corrections discussed in the next section. The $B^+ \to J/\psi K^+$ yields will be used in Ch. 8 as normalisation to extract the branching fractions and the isospin asymmetry.

Table 7.6.: Shape parameters for the misidentified background determined from the fit to $B^+ \to J/\psi\pi^+$ simulated samples for different years. These values are used to fix the tail parameters of the background shape in the fit to data. Uncertainties from the fit are not reported as not considered in the following.

Year	m_B' [MeV/c^2]	σ_L' [MeV/c^2]	σ_R' [MeV/c^2]	α'	β'	n'	k'	ρ'
2011	5324.3	29.0	19.0	−2.1	1.0	4.0	2.1	0.29
2012	5324.1	26.2	17.5	−2.5	0.7	1.4	2.3	0.47
2015	5325.2	26.9	18.5	−2.5	0.9	2.5	2.1	0.41
2016	5324.4	27.3	17.4	−3.0	0.9	0.8	2.3	0.41
2017	5324.3	27.4	17.5	−3.0	0.9	0.8	2.3	0.41
2018	5325.3	29.5	19.2	−2.8	1.2	0.6	2.0	0.31

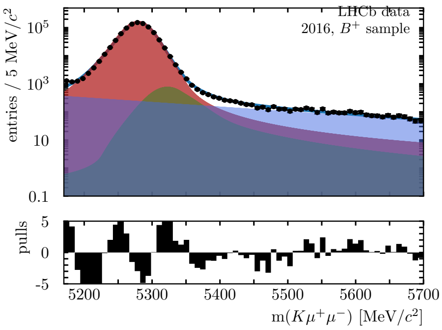


Figure 7.7.: Fit to $B^+ \to J/\psi K^+$ data for 2016. The signal component (red), the $B^+ \to J/\psi\pi^+$ background (green), and the combinatorial background (blue) are shown, together with the total fit function. The bottom panel displays the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty. The $B^+ \to J/\psi K^+$ yields are used for normalisation, while the model as fixed in this fit will be used for the rare mode $B^+ \to K^+\mu^+\mu^-$.

Table 7.7.: Additional parameters determined from the fit to $B^+ \to J/\psi K^+$ data samples for different years. The mean mass m_B and the scale s_{MC}, together with the parameters from Tabs. 7.5–7.6, are used for the fit to the rare channel. Signal yields $N^{B^+}_{J/\psi}$ serve as normalisation input. Uncertainties from the fit are not reported as not considered in the following (with the exception of the ones on $N^{B^+}_{J/\psi}$ which will be properly considered in Ch. 9).

Year	m_B [MeV/c^2]	s_{MC}	$N^{B^+}_{J/\psi}$	N_{misID}	N_{comb}
2011	5283.9	1.14	$3.09 \cdot 10^5$	$1.2 \cdot 10^3$	$5.1 \cdot 10^3$
2012	5284.1	1.14	$6.99 \cdot 10^5$	$2.6 \cdot 10^3$	$1.2 \cdot 10^4$
2015	5279.9	1.15	$1.46 \cdot 10^5$	$1.1 \cdot 10^3$	$1.7 \cdot 10^3$
2016	5279.5	1.15	$9.05 \cdot 10^5$	$6.0 \cdot 10^3$	$9.3 \cdot 10^3$
2017	5279.3	1.12	$9.87 \cdot 10^5$	$8.5 \cdot 10^3$	$1.2 \cdot 10^4$
2018	5280.1	1.12	$1.20 \cdot 10^6$	$9.0 \cdot 10^3$	$1.5 \cdot 10^4$

7.4. $B^+ \to K^+\mu^+\mu^-$

The final step before the measurement of the branching fractions and the isospin asymmetry concerns the extraction of the $B^+ \to K^+\mu^+\mu^-$ yields. The strategy is similar to the previous invariant mass fits: a maximum likelihood fit, adapted from the resonant mode, is performed on selected $B^+ \to K^+\mu^+\mu^-$ data, separately in different intervals of q^2. As with the neutral rare channel, after the selection only two components are present, namely the signal decay and the residual combinatorial background. The model for the signal peak is identical to the resonant counterpart, a sum of a double-sided Crystal Ball and a Gaussian function, where parameters are fixed from the previous fit. Like in the neutral case, however, a new scale factor is introduced for all the occurring widths and is determined on $B^+ \to K^+\mu^+\mu^-$ simulated samples. This is applied in combination with the global data-simulation scaling defined in Eq. (7.8) and the final expression for the widths is analogous to Eq. (7.5). The extracted scale factors are shown in Fig. 7.8.

The model is completed by an exponential function with free slope to describe the residual combinatorial background. Although the signal yields in real data are blinded to avoid unintentional biases, the fit performance is illustrated using a representative pseudo-experiment (toy dataset). This sample is generated based on the probability density functions defined above, assuming the Standard Model branching fraction for the signal and a combinatorial background level estimated from the sidebands. The projections of the fit to this toy dataset are shown in Fig. 7.9 for one q^2 bin.

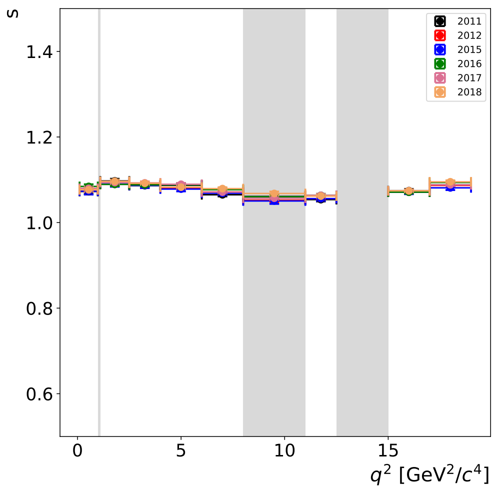


Figure 7.8.: Scale factors as obtained from a fit to $B^+ \to K^+\mu^+\mu^-$ simulated data as a function of q^2. For reference, scale factors are also given in the resonant region $[8.0, 11.0]\ \mathrm{GeV}^2/\mathrm{c}^4$, but they are not used in the following.

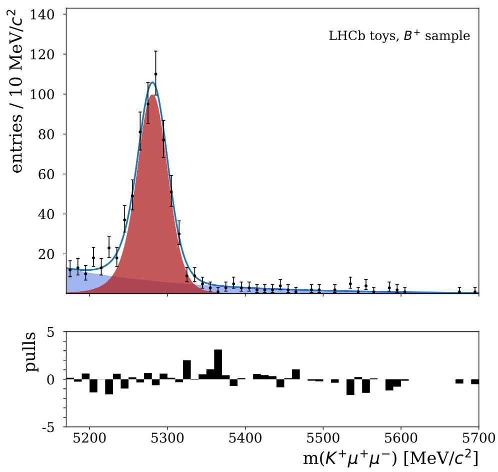


Figure 7.9.: Projections of the invariant mass fit to a $B^+ \to K^+\mu^+\mu^-$ toy dataset for an arbitrary chosen q^2 bin ($[0.1, 0.98]\ \mathrm{GeV}^2/\mathrm{c}^4$). The signal component is shown in red and the combinatorial background in blue. The pulls at the bottom confirm the validity of the model.

8. Isospin asymmetry and branching fraction

The inputs derived in the preceding chapters, specifically the signal yields, efficiencies, and normalisation constants, are now combined to measure the differential branching fractions and the isospin asymmetry of $B \to K\mu^+\mu^-$ decays. The branching fraction of the charged mode, $B^+ \to K^+\mu^+\mu^-$, is determined through a direct calculation relative to the normalisation channel, as detailed in the first part of this chapter. While the neutral branching fraction and the isospin asymmetry could be explicitly computed from the individual yields, such a sequential approach complicates the correct propagation of correlated uncertainties. Therefore, a simultaneous fit to both charged and neutral datasets is employed to extract these quantities directly. This chapter describes the measurement methodology and presents a validation study performed using pseudo-experiments to verify the robustness of the simultaneous fit strategy.

8.1. Branching fraction measurement

The branching fraction of a decay $A \to X$ represents the probability for the parent particle A to decay into the specific multi-body final state X. This has to be determined with respect to the total number of decaying particles N_A and, given N_X observed events of the decay of interest with an efficiency ε_X, can be expressed as:

$$\mathcal{B}(A \to X) = \frac{1}{\varepsilon_X}\frac{N_X}{N_A} \quad . \tag{8.1}$$

For the case of the $B^+ \to K^+\mu^+\mu^-$ decay, in Eq. 8.1 the number of produced B mesons is required. In the environment of the LHCb experiment, this is related to the integrated luminosity $\mathcal{L}$ and the b-production parameters:

$$N_{B^+} = \mathcal{L} \cdot \sigma_{b\bar{b}} \cdot f_u \ , \tag{8.2}$$

where $\sigma_{b\bar{b}}$ is the beauty production cross-section and f_u is the hadronisation fraction for the specific meson, a measure of the probability of producing a B^+ from the initial proton-proton collision. These quantities are accessible at LHCb, but they are affected by large uncertainties which would have a negative impact on the quality of the branching fraction measurement. For this reason the analysis presented here makes use of normalisation channels. By inverting Eq. (8.1) for an independent, well-measured channel involving the initial state A, the normalisation factor N_A can be rewritten as

$$N_A = \frac{N_{\mathrm{norm}}}{\varepsilon_{\mathrm{norm}}\,\mathcal{B}_{\mathrm{norm}}} \quad . \tag{8.3}$$

Substituting this expression into the original equation removes the dependence on the luminosity and production cross-sections. In practice, this means studying the decay frequency with respect to another process, and the method is effective only if the reference process is well known. Specifically, its branching fraction must be known with high precision. Furthermore, if the normalisation channel shares a similar topology and selection with the signal, many systematic uncertainties related to detection efficiencies cancel out in the ratio.

As discussed in Ch. 6, the resonant decay $B^+ \to J/\psi K^+$ is the optimal choice. It shares the same selection as the rare $B^+ \to K^+\mu^+\mu^-$ mode, when the resonance is reconstructed through the muonic $J/\psi \to \mu^+\mu^-$ decay, and has a precise branching fraction available in the literature [72]:

$$\mathcal{B}(B^+ \to J/\psi K^+) = (1.020 \pm 0.019) \cdot 10^{-3} \ , \quad \mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \cdot 10^{-2} \ . \tag{8.4}$$

Normalising to this channel, the formula for the $B^+ \to K^+\mu^+\mu^-$ branching fraction becomes

$$\mathcal{B}(B^+ \to K^+\mu^+\mu^-) = \mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)\frac{\varepsilon^{B^+}_{J/\psi}}{\varepsilon^{B^+}_{\mu\mu}}\frac{N^{B^+}_{\mu\mu}}{N^{B^+}_{J/\psi}} \ , \tag{8.5}$$

where the term $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ accounts for the J/ψ reconstruction.

All inputs for Eq. 8.5 have been defined in previous chapters. The total efficiencies (ε), comprising acceptance, reconstruction, selection, and trigger effects, are determined on corrected simulation as described in Ch. 5. The yields (N) are extracted from the mass fits detailed in Ch. 7. These quantities, all determined separately for each year of data-taking, would lead to six independent measurements of the same branching fraction that need to be averaged. To avoid this and correctly propagate uncertainties, the rare mode yield $N^{B^+}_{\mu\mu}$ is parametrised in the mass fit model directly in terms of the branching fraction. By inverting Eq. 8.5 the yield can be expressed as

$$N^{B^+}_{\mu\mu} = \frac{\varepsilon^{B^+}_{\mu\mu}}{\varepsilon^{B^+}_{J/\psi}} N^{B^+}_{J/\psi} \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)} = S_{B^+} \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \ . \tag{8.6}$$

In this approach, $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$ becomes the single floating parameter of a simultaneous fit performed across all years. In practice, however, the parameter extracted from the fit is the relative branching fraction defined as

$$\mathcal{B}_{\mathrm{rel}}(B^+ \to K^+\mu^+\mu^-) = \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \ , \tag{8.7}$$

such that the expression for the yield simplifies to

$$N^{B^+}_{\mu\mu} = S_{B^+}\, \mathcal{B}_{\mathrm{rel}}(B^+ \to K^+\mu^+\mu^-) \ . \tag{8.8}$$

Other inputs, such as efficiencies and normalisation yields, are treated as fixed constants. Consequently, the factor

$$S_{B^+} = \frac{\varepsilon^{B^+}_{\mu\mu}}{\varepsilon^{B^+}_{J/\psi}} N^{B^+}_{J/\psi} \tag{8.9}$$

is fixed to its central value in the likelihood maximisation. Although all terms in Eq. (8.9) contribute to the total uncertainty, they are not propagated automatically by the fit. Instead, their impact, stemming from finite simulation statistics, normalisation yield errors, and external branching fraction inputs, is evaluated

Table 8.1.: Classification of the parameters for the fit to the B^+ invariant mass to determine the $B^+ \to K^+\mu^+\mu^-$ branching fraction.

Parameter	Sub-parameter	Uncertainty propagated	Strategy
S_{B^+}	$\varepsilon^{B^+}_{\mu\mu}$	syst	fixed
	$\varepsilon^{B^+}_{J/\psi}$	syst	
	$N^{B^+}_{J/\psi}$	syst	
$\mathcal{B}_{\rm rel}(B^+ \to K^+\mu^+\mu^-)$		–	free

Table 8.2.: q^2 binning schemes used for the single fit to measure $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, and the simultaneous fit to measure both $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and $A_{\rm I}$.

Measurement	q^2 binning scheme $[\mathrm{GeV}^2/\mathrm{c}^4]$
$\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$	$0.1, 0.98, 1.1, 2.0$–8.0 in $1\,\mathrm{GeV}^2/\mathrm{c}^4$ steps, $11.0, 11.75, 12.50, 15.0$–22.0 in $1\,\mathrm{GeV}^2/\mathrm{c}^4$ steps
$\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and $A_{\rm I}$	$0.1, 0.98, 1.1, 2.5, 4.0, 6.0, 8.0, 11.0, 12.50, 15.0, 17.0, 19.0$

separately and assigned as a systematic uncertainty. The structure of the fit model, with the classification of parameters, is summarised in Tab. 8.1.

Although the equations above present the branching fraction in its integral form, the physics of the decay varies with the dimuon invariant mass squared, q^2. Therefore, the measurement is performed separately in different ranges of q^2 to determine the differential quantity $d\mathcal{B}/dq^2$. Since signal yields are blinded, it is not possible to show the fit to the complete datasets. However, to illustrate the procedure, the projections of a fit to a toy dataset in the lowest q^2 interval $0.1\,\mathrm{GeV}^2/\mathrm{c}^4 < q^2 < 0.98\,\mathrm{GeV}^2/\mathrm{c}^4$ for all the years are shown in Fig. 8.1. The toy dataset is generated according to the expected model and statistics as detailed in Sec. 8.3.

The procedure described above could also be applied to the neutral case to measure $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and, making use of Eq. 2.20, the isospin asymmetry. However, a simultaneous fit to both charged and neutral datasets is preferred. This strategy allows for the extraction of the neutral branching fraction and the isospin asymmetry at the same time, automatically handling the uncertainty propagation within the fit procedure. Details on this implementation are provided in the next section. This strategy is particularly advantageous because it allows the adoption of different schemes for q^2 intervals for the two modes. Without having to address the $K^0_{\rm S}$ reconstruction, the charged dataset is approximately ten times larger than the neutral one. This larger amount of data allows for a more granular scheme, enabling a more detailed comparison with theoretical predictions. The specific q^2 intervals used for both measurements, already presented in Sec. 5.5, are listed in Tab. 8.2.

8.2. Isospin asymmetry measurement

As stated in the previous section, the isospin asymmetry and the neutral branching fraction are measured through a simultaneous fit to both the $B^0 \to K^0_{\rm S}\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ samples. To achieve this,

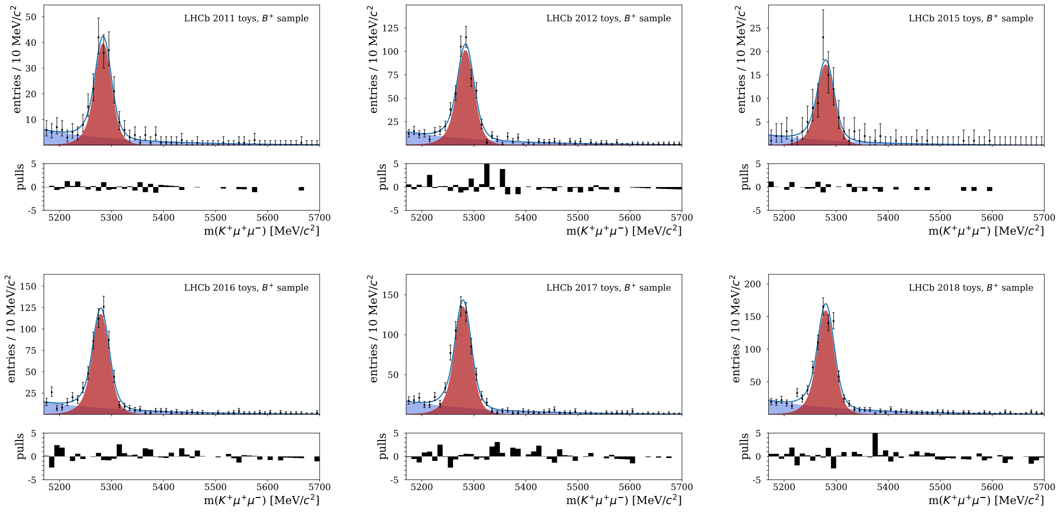


Figure 8.1.: Projections of a simultaneous fit to the B^+ invariant mass of a toy dataset to determine the $B^+ \to K^+\mu^+\mu^-$ branching fraction in the lowest q^2 interval, for 2011 (top left), 2012 (top centre), 2015 (top right), 2016 (bottom left), 2017 (bottom centre), 2018 (bottom right). In the plots, the signal component is shown in red and the combinatorial background in blue. The pulls at the bottom confirm the validity of the model.

analogously to the previous case, the signal yields are expressed in terms of the physical observables of interest, A_{I} and $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$. These quantities can therefore be considered free parameters of the extended fit.

The reparameterisation builds upon Eq. 8.8, which relates the charged yield to its branching fraction. Applying the same logic to the neutral channel, the raw yields can be expressed as:

$$\begin{aligned} N_{\mu\mu}^{B^0} &= \frac{\varepsilon_{\mu\mu}^{B^0}}{\varepsilon_{J/\psi}^{B^0}} \frac{N_{J/\psi}^{B^0}}{\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)\,\langle K_{\mathrm{S}}^0|K^0\rangle^2}\, \mathcal{B}(B^0 \to K^0\mu^+\mu^-)\,\langle K_{\mathrm{S}}^0|K^0\rangle^2 \ , \\ N_{\mu\mu}^{B^+} &= \frac{\varepsilon_{\mu\mu}^{B^+}}{\varepsilon_{J/\psi}^{B^+}} \frac{N_{J/\psi}^{B^+}}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)}\, \mathcal{B}(B^+ \to K^+\mu^+\mu^-) \ , \end{aligned} \tag{8.10}$$

where the factor $|\langle K_{\mathrm{S}}^0|K^0\rangle|^2 = 1/2$ accounts for the probability that a K^0 decays as a K_{S}^0. The isospin asymmetry can be introduced into the fit model by using its definition, repeated here for practical reasons:

$$A_{\mathrm{I}} = \frac{\mathcal{B}(B^0 \to K^0\mu^+\mu^-) - \frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^0\mu^+\mu^-) + \frac{\tau_{B^0}}{\tau_{B^+}}\mathcal{B}(B^+ \to K^+\mu^+\mu^-)} \ . \tag{8.11}$$

By inverting Eq. 8.11, the branching fraction of the charged mode, whose measurement is discussed in the previous section, can be rewritten as

$$\mathcal{B}(B^+ \to K^+\mu^+\mu^-) = \frac{\tau_{B^+}}{\tau_{B^0}} \frac{1 - A_\mathrm{I}}{1 + A_\mathrm{I}} \mathcal{B}(B^0 \to K^0\mu^+\mu^-) \tag{8.12}$$

and substituted in Eq. 8.10. After substitution, the yields for both channels read

$$\begin{aligned} N_{\mu\mu}^{B^0} &= \frac{\varepsilon_{\mu\mu}^{B^0}}{\varepsilon_{J/\psi}^{B^0}} \frac{N_{J/\psi}^{B^0}}{\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \mathcal{B}(B^0 \to K^0\mu^+\mu^-) \ , \\ N_{\mu\mu}^{B^+} &= \frac{\varepsilon_{\mu\mu}^{B^+}}{\varepsilon_{J/\psi}^{B^+}} \frac{N_{J/\psi}^{B^+}}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \frac{\tau_{B^+}}{\tau_{B^0}} \frac{1 - A_\mathrm{I}}{1 + A_\mathrm{I}} \mathcal{B}(B^0 \to K^0\mu^+\mu^-) \ . \end{aligned} \tag{8.13}$$

While Eq. 8.13 is sufficient to define the simultaneous fit, a further algebraic simplification is performed to improve the readability of the result. Since the measurement is normalised to the resonant mode, it is convenient to introduce the relative branching fraction

$$\mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-) = \frac{\mathcal{B}(B^0 \to K^0\mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \ , \tag{8.14}$$

corresponding to the ratio of decay probabilities between the signal $B^0 \to K^0\mu^+\mu^-$ and the normalisation channel $B^0 \to J/\psi(\to \mu^+\mu^-)\, K^0$. Without any substantial difference, the last quantity can be taken as free parameter instead of the absolute branching fraction, and the yields simplify to

$$\begin{aligned} N_{\mu\mu}^{B^0} &= \frac{\varepsilon_{\mu\mu}^{B^0}}{\varepsilon_{J/\psi}^{B^0}} N_{J/\psi}^{B^0} \mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-) = S_0\, \mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-) \ , \\ N_{\mu\mu}^{B^+} &= \frac{\varepsilon_{\mu\mu}^{B^+}}{\varepsilon_{J/\psi}^{B^+}} \frac{\mathcal{B}(B^0 \to J/\psi\, K^0)}{\mathcal{B}(B^+ \to J/\psi\, K^+)} \frac{\tau_{B^+}}{\tau_{B^0}} N_{J/\psi}^{B^+} \frac{1 - A_\mathrm{I}}{1 + A_\mathrm{I}} \mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-) \\ &= S_+ \frac{1 - A_\mathrm{I}}{1 + A_\mathrm{I}} \mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-) \ . \end{aligned} \tag{8.15}$$

The factors

$$S_0 = \frac{\varepsilon_{\mu\mu}^{B^0}}{\varepsilon_{J/\psi}^{B^0}} N_{J/\psi}^{B^0} \ , \quad S_+ = \frac{\varepsilon_{\mu\mu}^{B^+}}{\varepsilon_{J/\psi}^{B^+}} \frac{\mathcal{B}(B^0 \to J/\psi\, K^0)}{\mathcal{B}(B^+ \to J/\psi\, K^+)} \frac{\tau_{B^+}}{\tau_{B^0}} N_{J/\psi}^{B^+} \tag{8.16}$$

are introduced to collect all the inputs previously determined. These are taken from the preceding chapters (efficiencies from Ch. 5, resonant yields from Ch. 7) or from the literature [72]:

$$\begin{aligned} \mathcal{B}(B^0 \to J/\psi K^0) = (8.91 \pm 0.21) \cdot 10^{-4} \ , \quad \mathcal{B}(B^+ \to J/\psi K^+) = (1.020 \pm 0.019) \cdot 10^{-3} \ , \\ \tau_{B^0} = (1.517 \pm 0.004) \cdot 10^{-12}\,\mathrm{s} \ , \quad \tau_{B^+} = (1.638 \pm 0.004) \cdot 10^{-12}\,\mathrm{s} \ . \end{aligned} \tag{8.17}$$

Table 8.3.: Classification of the parameters for the fit to B^0 and B^+ samples to determine the isospin asymmetry and the $B^0 \to K^0\mu^+\mu^-$ branching fraction.

Parameter	Sub-parameter	Uncertainty propagated	Strategy
$S_{LL/DD}$	$\varepsilon^{B^0}_{\mu\mu,LL/DD}$	syst	fixed
	$\varepsilon^{B^0}_{J/\psi,LL/DD}$	syst	
	$N^{B^0}_{J/\psi,LL/DD}$	syst	
S_+	$\varepsilon^{B^+}_{\mu\mu}$	syst	fixed
	$\varepsilon^{B^+}_{J/\psi}$	syst	
	$\mathcal{B}(B^0 \to J/\psi\, K^0)$	syst	
	$\mathcal{B}(B^+ \to J/\psi\, K^+)$	syst	
	τ_{B^+}	syst	
	τ_{B^0}	syst	
	$N^{B^+}_{J/\psi}$	syst	
A_{I}		–	free
$\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$		–	free

Although derived in integral form, this formalism is applied independently in each q^2 interval to determine the differential distributions. Additionally, since the neutral mode analysis distinguishes between LL and DD track categories, the factor S_0 is computed separately for each category:

$$S_{LL} = \frac{\varepsilon^{B^0}_{\mu\mu,LL}}{\varepsilon^{B^0}_{J/\psi,LL}} N^{B^0}_{J/\psi,LL} \ , \quad S_{DD} = \frac{\varepsilon^{B^0}_{\mu\mu,DD}}{\varepsilon^{B^0}_{J/\psi,DD}} N^{B^0}_{J/\psi,DD} \ . \tag{8.18}$$

The constraining strategy follows that defined for the fit to the charged branching fraction. The S factors are fixed to their central values and uncertainties associated with all the terms are treated as systematic uncertainties at a later stage. The structure of the parameters in this simultaneous fit is summarised in Tab. 8.3.

Also in this case signal yields are blinded, and it is not possible to show the fit to the complete datasets. However, the procedure can be illustrated using toy datasets, generated according to the expected model and number of $B^0 \to K^0_S\mu^+\mu^-$ LL- and DD-type and $B^+ \to K^+\mu^+\mu^-$ events, as described in Sec. 8.3. Fig. 8.2 shows the projections of a simultaneous fit to the three components of the toy dataset in the lowest q^2 interval $0.1\,\mathrm{GeV}^2/c^4 < q^2 < 2.0\,\mathrm{GeV}^2/c^4$ for all the years.

8.3. Fit validation

The simultaneous fits described in Sec. 8.1 and Sec. 8.2 offer the aforementioned advantage of extracting the quantities of interest, A_{I}, $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, directly from the likelihood maximisation. On the other hand, due to the complexity of the yield parameterisation, potential biases could be introduced. Therefore, the analysis framework requires validation using a set of pseudo-experiments (toys), generated according to the expected distributions. The objective is to verify that the

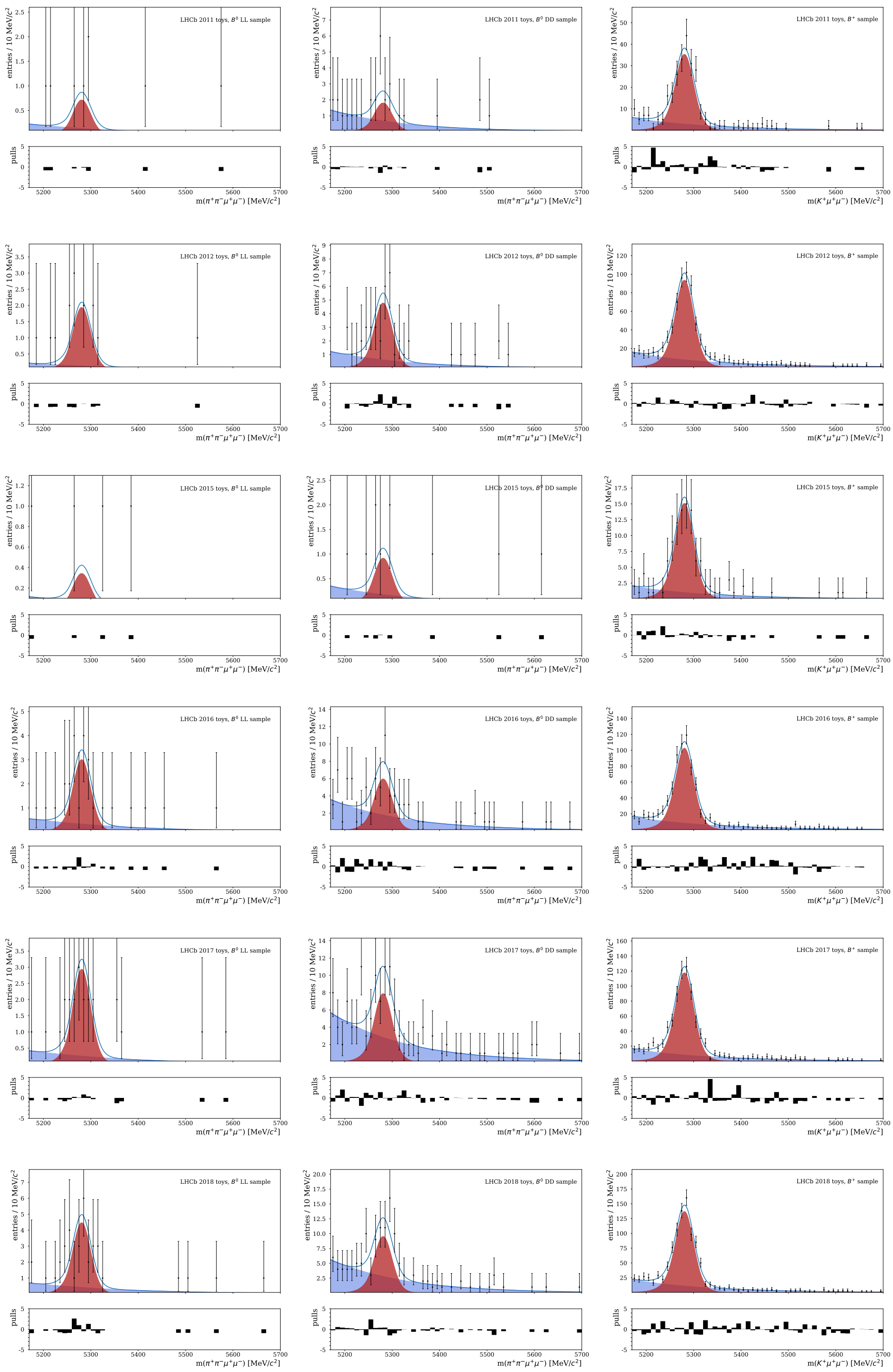


Figure 8.2.: Projections of a simultaneous fit to a LL- (left) and DD-type (centre) $B^0 \longrightarrow K^0_S\mu^+\mu^-$ and $B^+ \longrightarrow K^+\mu^+\mu^-$ (right) toy dataset, for (top to bottom) 2011, 2012, 2015, 2016, 2017 and 2018. In the plots, the signal component is shown in red and the combinatorial background in blue. The pulls at the bottom confirm the validity of the model.

fit provides unbiased estimators and accurate uncertainty estimates. This is assessed by studying the distribution of the *pulls*, defined as the difference between the measured and input values (x) normalised by the statistical uncertainty (σ_{meas}):

$$\text{pull} = \frac{x_{\text{meas}} - x_{\text{true}}}{\sigma_{\text{meas}}} \ . \tag{8.19}$$

For an unbiased estimator with correctly estimated errors, the pull distribution is expected to follow a standard Gaussian with mean $\mu = 0$ and width $\sigma = 1$. Significant deviations from $\mu = 0$ would indicate a bias in the measurement, while $\sigma \neq 1$ would imply an underestimation or overestimation of the statistical uncertainty.

The validation is performed separately for each q^2 interval. For both fit configurations, the background generation follows a common data-driven strategy. Assuming the mass sidebands are populated only by combinatorial background, the expected background yields and slopes are derived from an exponential fit to the blinded data. The yields are then extrapolated to the full mass range $[a, b] = [5170, 5700]\ \text{MeV}/c^2$ using the scaling factor

$$\Gamma = \frac{e^{-\lambda a} - e^{-\lambda b}}{(e^{-\lambda a} - e^{-\lambda A}) + (e^{-\lambda B} - e^{-\lambda b})} \ , \tag{8.20}$$

where λ is the fitted slope, and $[A, B] = [5230, 5330]\ \text{MeV}/c^2$ is the blinded signal region. The specific details for the two fit configurations are provided in the following subsections.

8.3.1. Branching fraction measurement

The validation of the fit to the charged mode, used to determine $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, involves generating 1000 pseudo-datasets for each q^2 bin. Each dataset consists of $B^+ \to K^+\mu^+\mu^-$ candidates distributed according to the probability density functions defined in Sec. 7.4: a sum of a double-sided Crystal Ball and a Gaussian function for the signal, and an exponential function for the background. The number of events is Poisson-fluctuated around the expected yield. The expected signal yield is calculated by fixing the branching fraction in Eq. 8.8 to the Standard Model prediction

$$\mathcal{B}^{\text{SM}}_{\text{rel}}(B^+ \to K^+\mu^+\mu^-)_{[a,b]} = \frac{1}{\mathcal{B}(B^+ \to J/\psi\, K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \int_a^b \left.\frac{d\mathcal{B}}{dq^2}\right|_{\text{SM}} dq^2 \ , \tag{8.21}$$

where the integral is computed using the `flavio` package [82]. Expected background yields and slopes are derived from exponential fits to the data sidebands, extrapolated to the full mass range through the factor defined in Eq. (8.20).

For each generated dataset, the fit is performed treating the branching fraction as a free parameter, while the normalisation factors S_{B^+} are fixed to their central values. The pull distributions for $\mathcal{B}_{\text{rel}}(B^+ \to K^+\mu^+\mu^-)$ are then fitted with a Gaussian function. Figure 8.3 shows an example of the pull distribution for the lowest q^2 interval. The means and widths of the pull distributions are consistent with 0 and 1 respectively for all q^2 bins, confirming that the single-fit procedure for the charged mode is unbiased. A summary of the pull parameters for all intervals is reported in Fig. 8.4.

8.3.2. Isospin asymmetry measurement

The validation of the simultaneous fit to both neutral and charged samples follows a similar strategy. In this case, each of the 1000 pseudo-datasets contains LL- and DD-type B^0 candidates as well as B^+

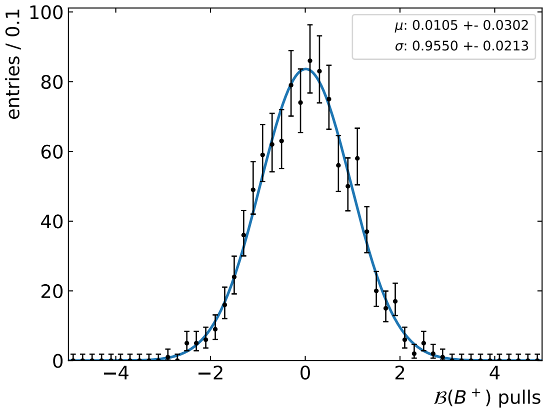


Figure 8.3.: Gaussian fit to the pull distribution for the branching fraction $\mathcal{B}_{\mathrm{rel}}(B^+ \to K^+\mu^+\mu^-)$ parameter in the lowest q^2 interval. The obtained values for mean and standard deviation are reported in the legends, showing the distribution is compatible with a normal distribution.

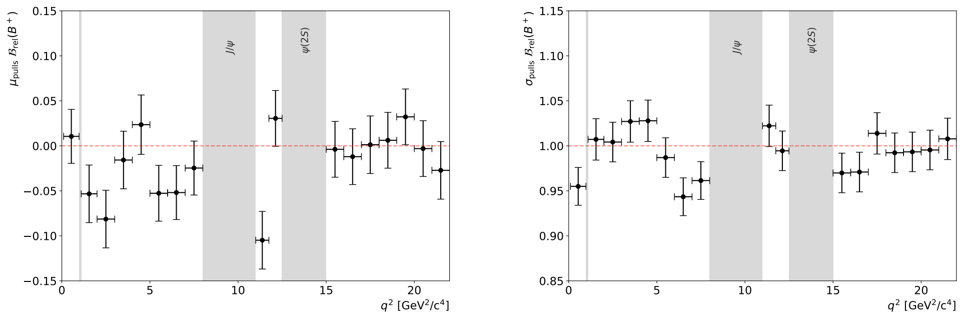


Figure 8.4.: Mean (left) and standard deviation (right) of the pull distribution as obtained from a Gaussian fit for the branching fraction $\mathcal{B}_{\mathrm{rel}}(B^+ \to K^+\mu^+\mu^-)$ as a function of q^2.

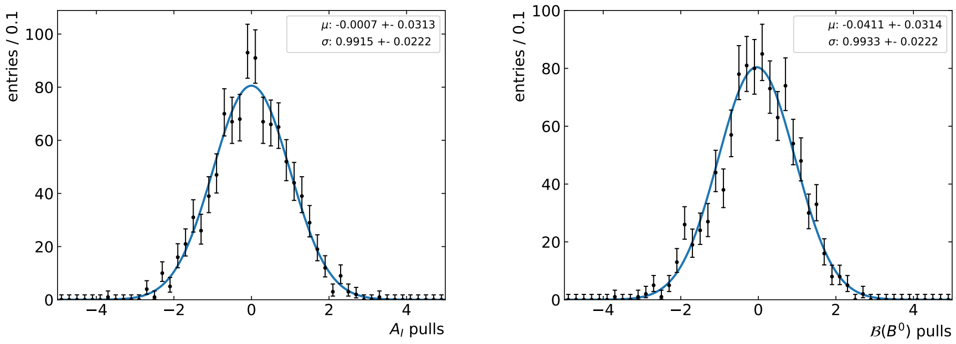


Figure 8.5.: Gaussian fit to the pull distribution for the isospin asymmetry A_I (left) and the relative branching fraction $\mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-)$ (right) in the lowest q^2 interval. The obtained values for mean and standard deviation are reported in the legends.

candidates. Candidates are distributed according to the models described in Sec. 7.2 and Sec. 7.4: double sided Crystal Ball functions, with an additional Gaussian function in the charged case, for the signal, and exponential functions for the background, independent in the three subsets. The expected number of signal events is obtained by fixing the isospin asymmetry to 0.0 in Eq. 8.15, and setting the relative branching fraction to the Standard Model expectation

$$\mathcal{B}_\mathrm{rel}^\mathrm{SM} = \frac{\mathcal{B}^\mathrm{SM}(B^0 \to K^0\mu^+\mu^-)_{[a,b]}}{\mathcal{B}(B^0 \to J/\psi\, K^0)\mathcal{B}(J/\psi \to \mu^+\mu^-)} \quad . \tag{8.22}$$

The background modelling follows the same data-driven sideband extrapolation described previously.

The simultaneous fit is performed on each dataset, floating A_I and $\mathcal{B}_\mathrm{rel}$ while fixing the S_0 and S_+ factors to their central values. Figure 8.5 shows the pull distributions for the two parameters of interest in the lowest q^2 interval. The Gaussian fits to the pulls show standard deviations compatible with unity, indicating a correct estimation of the statistical uncertainty. However, a small bias is observed in the means, which deviate from zero by approximately 5% of the statistical error depending on the q^2 bin. To investigate the origin of this deviation, a second set of pseudo-experiments was generated with signal and background yields scaled by a factor of 100. The pull distributions for this high-statistics scenario exhibit means fully compatible with zero. This demonstrates that the fit procedure is asymptotically unbiased and that the small deviation observed at nominal yields is an effect related to the limited sample size. Consequently, no correction is applied to the central values at this stage. Instead, this residual bias will be evaluated and accounted for as a systematic uncertainty after the unblinding of the data. The complete summary of the pull parameters for the nominal statistics scenario is presented in Fig. 8.6, while only the pull means are reported in Fig. 8.7 for the high-statistics scenario.

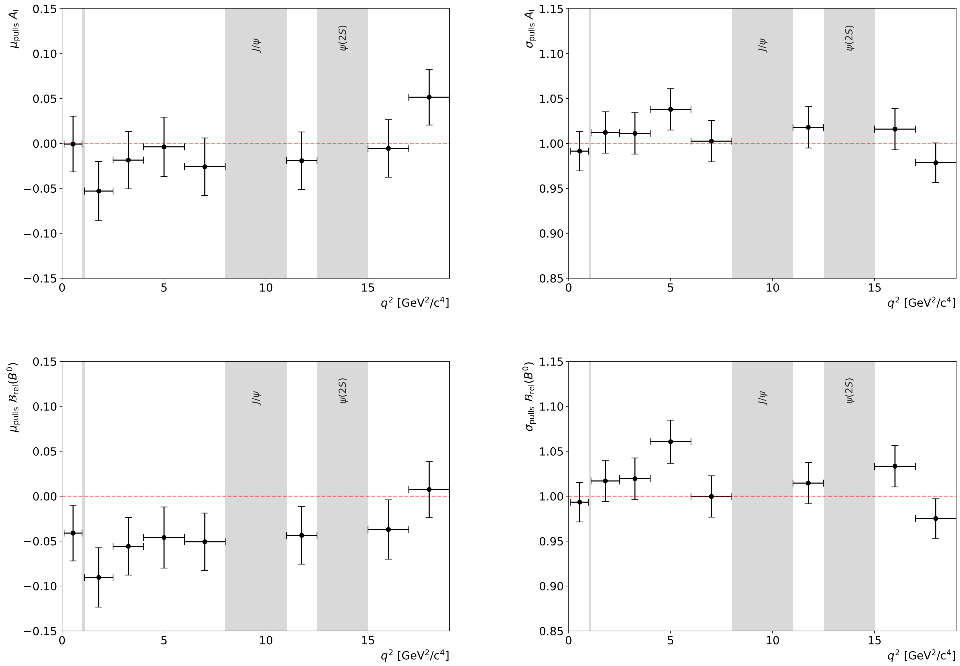


Figure 8.6.: Mean (left) and standard deviation (right) of the pull distribution for the isospin asymmetry A_{I} (top) and the branching fraction $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$ (bottom) as obtained from a Gaussian fit, as a function of q^2.

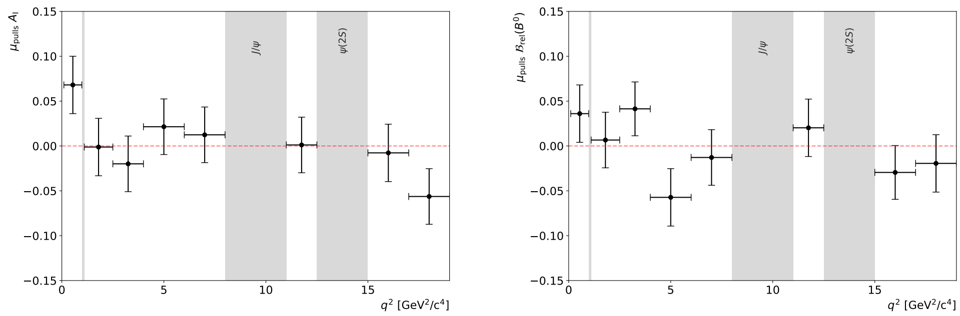


Figure 8.7.: Mean of the pull distribution for the isospin asymmetry A_{I} (left) and the branching fraction $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$ (right) in the high-statistics scenario, as a function of q^2.

9. Systematic uncertainties

The measurements of the isospin asymmetry and branching fractions described in Ch. 8 are affected by systematic uncertainties in addition to the statistical ones derived from the fit procedure. These uncertainties arise from the specific assumptions made in the modelling of signal and background distributions, the finite precision of the calibration samples, and the uncertainties on normalisation factors. This chapter details the evaluation of these systematic effects. The general strategy adopted for the estimation relies on high-statistics pseudo-experiments, following a logic similar to the validation described in Sec. 8.3. By generating datasets under alternative hypotheses or varying specific parameters within their uncertainties, the potential bias on the final results is isolated and quantified. The sources of systematic uncertainty are categorised and discussed in the following order: the choice of the invariant mass fit model, the evaluation of efficiency correction weights, the particle identification calibration, and the precision of the inputs determining the normalisation factors.

9.1. Fit model

The choice of the fit model for the invariant B mass is driven by phenomenological considerations, as the physical processes shaping the distribution are not fully derived from first principles. While the nominal functions show good agreement with the simulated data, there are no theoretical reasons assuring they are the real ones. The choice of the fit model is a source of uncertainty which is not accounted for by the statistical error of the fit. To estimate the systematic uncertainty arising from this choice, a study based on pseudo-experiments (*toys*) is performed. The objective is to evaluate the deviation in the results when an alternative, equally valid model is used to fit the data.

To decouple the systematic effect from statistical fluctuations, these pseudo-experiments are generated with a sample size 100 times larger than the expected data yield. A total of 1000 high-statistics toys per q^2 interval are generated according to the nominal model. Each toy dataset is composed of three subsets, two for LL and DD-type B^0 candidates and one for B^+ candidates, distributed as described in Ch. 7. The number of events is fixed according to Eq. (8.15), setting $A_{\mathrm{I}} = 0$ and computing the integrated differential branching fraction with the `flavio` package. Specifically, for the neutral mode subsets in the q^2 interval $[a, b]$

$$N_{\mu\mu}^{B^0, LL/DD} = 100 \times S_{LL/DD} \frac{\int_a^b \frac{d\mathcal{B}(B^0 \to K^0 \mu^+\mu^-)}{dq^2} dq^2}{\mathcal{B}(B^0 \to J/\psi K^0)\, \mathcal{B}(J/\psi \to \mu^+\mu^-)} \tag{9.1}$$

events are generated according to the nominal double-sided Crystal Ball functions determined from $B^0 \to J/\psi K^0_S$ simulated samples. Similarly, for the charged mode

$$N_{\mu\mu}^{B^+} = 100 \times S_+ \frac{\int_a^b \frac{d\mathcal{B}(B^0 \to K^0 \mu^+\mu^-)}{dq^2} dq^2}{\mathcal{B}(B^0 \to J/\psi K^0)\, \mathcal{B}(J/\psi \to \mu^+\mu^-)} \tag{9.2}$$

events are distributed according to the nominal sum of a double-sided Crystal Ball and a Gaussian, fixed on $B^+ \to J/\psi K^+$ simulations. For each subset the background is simulated independently, generating a decreasing exponential distribution. The number of background events is set to 100 times the expected yield determined in Ch. 8.

The input values for the physical observables, A_I, $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, are reliably recovered when fitting the generated distributions with the nominal model. A valid alternative model must satisfy a similar condition: it must describe the simulated samples with a quality comparable to the nominal one. If this condition is met, there is no strong statistical reason to prefer one parameterisation over the other, and any observed difference in the final result would have to be assigned as a systematic uncertainty. The choice of the alternative model is therefore critical: the test should not be invalidated by an imprecise description of the distribution. To alternatively describe the $B^0 \to K^0_S\mu^+\mu^-$ invariant mass distribution a sum of two Gaussian functions sharing the mean but with different widths is chosen. For the $B^+ \to K^+\mu^+\mu^-$ decay, a sum of three Gaussians is used, accounting for the specific features of the charged mode distribution previously observed.

The parameters of the alternative models are determined by fitting the simulated samples, separately for LL and DD-type B^0 and B^+ candidates. For the neutral mode these are the shared mean and the two widths, while three widths need to be determined in addition to the mean for the charged mode. The last step also serves to validate the quality of the alternative description. Fig. 9.1 shows that the agreement with simulations is good. Despite the reduced complexity of the models, the bulk of the distribution, corresponding to the high-statistic peak region, is well described. This level of agreement is considered adequate for the evaluation of the systematic uncertainties.

The systematic evaluation proceeds by fitting the 1000 generated datasets twice: once using the nominal signal model, and once using the alternative model. This is done for both the fit configurations.

For the single branching fraction measurement, the procedure extracts $\mathcal{B}_\mathrm{rel}(B^+ \to K^+\mu^+\mu^-)$ from the B^+ subsample only. For each q^2 interval the distribution of the difference between the results obtained in the two scenarios is analysed. An example of these distributions is shown in Fig. 9.2. The mean of the distribution represents the average bias introduced by the model choice and is assigned as the systematic uncertainty. The resulting values are listed in Tab. 9.1.

For the simultaneous fit to isospin asymmetry and neutral branching fraction, the full toy dataset, comprising B^0 and B^+ components, is used. The two fit parameters, A_I and $\mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-)$ are determined using the nominal model and the alternative one. The mean difference between the two results is taken as the systematic uncertainty. The results are reported in Tab. 9.2.

9.2. Correction weights

A similar strategy based on pseudo-experiments is used to study systematic uncertainties originating from the reweighting procedure, introduced to correct for the distribution of the B meson transverse momentum p_T and pseudorapidity η, from the distribution of the `nTracks` variable, and from the `L0` trigger efficiencies.

To estimate the systematic uncertainty due to these simulation corrections, a conservative approach is adopted. For each correction type, the impact is evaluated by removing the corresponding weights. The fit is performed using the nominal efficiencies and repeated using alternative efficiencies computed

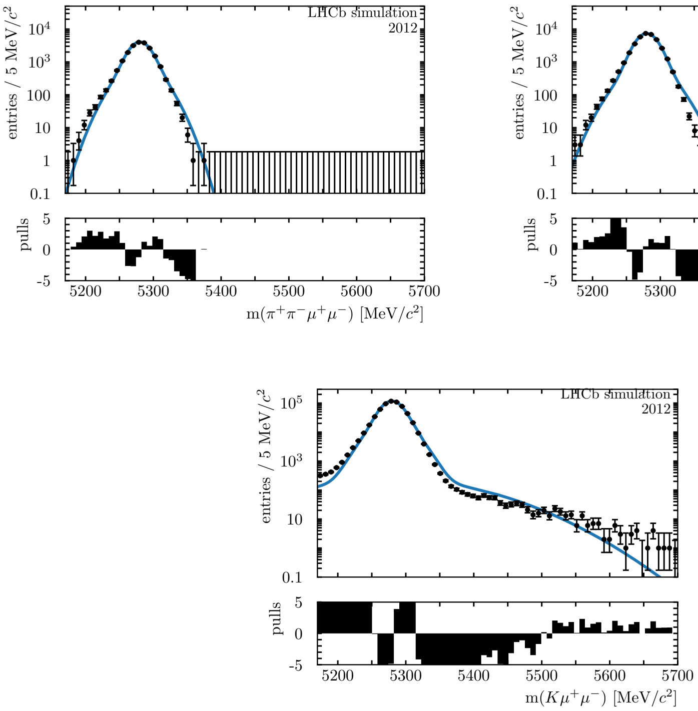


Figure 9.1.: Fit to the B invariant mass using the alternative models as described in the text on LL (top left) and DD-type (top right) $B^0 \longrightarrow J/\psi K^0_S$ and $B^+ \longrightarrow J/\psi K^+$ (bottom) simulated samples.

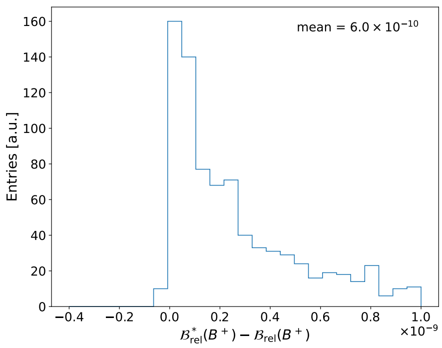


Figure 9.2.: Distribution of the difference between the results obtained when fitting the $B^+ \longrightarrow K^+\mu^+\mu^-$ toy dataset with the nominal and the alternative model, for the lowest q^2 interval $[0.1, 0.98]$ $\mathrm{GeV}^2/\mathrm{c}^4$.

Table 9.1.: Systematic uncertainty on $\mathcal{B}_{\text{rel}}(B^+ \to K^+\mu^+\mu^-)$, arising from the choice of the fit model in the invariant B mass, as a function of q^2.

q^2 region $[\text{GeV}^2/\text{c}^4]$	Systematic uncertainty $\sigma^{\text{syst}}_{\mathcal{B}_{\text{rel}}(B^+)}$
$[0.1, 0.98]$	$6.0 \cdot 10^{-10}$
$[1.1, 2.0]$	$6.7 \cdot 10^{-10}$
$[2.0, 3.0]$	$5.9 \cdot 10^{-10}$
$[3.0, 4.0]$	$5.9 \cdot 10^{-10}$
$[4.0, 5.0]$	$6.5 \cdot 10^{-10}$
$[5.0, 6.0]$	$6.1 \cdot 10^{-10}$
$[6.0, 7.0]$	$6.1 \cdot 10^{-10}$
$[7.0, 8.0]$	$6.0 \cdot 10^{-10}$
$[11.0, 11.75]$	$5.8 \cdot 10^{-10}$
$[11.75, 12.5]$	$5.3 \cdot 10^{-10}$
$[15.0, 16.0]$	$6.7 \cdot 10^{-10}$
$[16.0, 17.0]$	$5.6 \cdot 10^{-10}$
$[17.0, 18.0]$	$5.5 \cdot 10^{-10}$
$[18.0, 19.0]$	$5.2 \cdot 10^{-10}$
$[19.0, 20.0]$	$5.9 \cdot 10^{-10}$
$[20.0, 21.0]$	$6.4 \cdot 10^{-10}$
$[21.0, 22.0]$	$5.2 \cdot 10^{-10}$

Table 9.2.: Systematic uncertainty on A_{I} and $\mathcal{B}_{\text{rel}}(B^0 \to K^0\mu^+\mu^-)$, arising from the choice of the fit model in the invariant B mass, as a function of q^2.

q^2 region	Systematic uncertainty	
$[\text{GeV}^2/\text{c}^4]$	$\sigma^{\text{syst}}_{A_{\text{I}}}$	$\sigma^{\text{syst}}_{\mathcal{B}_{\text{rel}}(B^0)}$
$[0.1, 0.98]$	$7.3 \cdot 10^{-6}$	$9.2 \cdot 10^{-9}$
$[1.1, 2.5]$	$4.5 \cdot 10^{-6}$	$3.7 \cdot 10^{-9}$
$[2.5, 4.0]$	$1.3 \cdot 10^{-5}$	$1.4 \cdot 10^{-8}$
$[4.0, 6.0]$	$2.6 \cdot 10^{-6}$	$4.3 \cdot 10^{-9}$
$[6.0, 8.0]$	$1.6 \cdot 10^{-6}$	$2.6 \cdot 10^{-9}$
$[11.0, 12.5]$	$8.0 \cdot 10^{-6}$	$9.8 \cdot 10^{-9}$
$[15.0, 17.0]$	$6.6 \cdot 10^{-6}$	$4.3 \cdot 10^{-9}$
$[17.0, 19.0]$	$2.2 \cdot 10^{-5}$	$1.1 \cdot 10^{-8}$

without the specific correction weights. The difference between the parameter values obtained in the two cases is assigned as the systematic uncertainty.

For each correction step, 1000 pseudo-experiments per q^2 interval (with 100 times the expected statistics) are generated according to the nominal model, following the procedure described in Sec. 9.1. Similarly, the generated datasets are fitted twice: once using nominal efficiencies to parametrise yields, and once using alternative efficiencies, computed from simulation samples when excluding the correction weights. The efficiencies enter the S factors of Eq. (8.9) and Eq. (8.16), governing the relationship between signal yields and the quantities of interest.

For each q^2 interval, the distribution of the difference between the results obtained in the two scenarios is evaluated. The mean of this distribution is taken as the systematic uncertainty. This evaluation is performed for both the single branching fraction fit and the simultaneous fit to the asymmetry and neutral branching fraction. For the single branching fraction measurement of the charged mode, the systematic uncertainties arising from kinematic (p_T, η), multiplicity (`nTracks`), and `L0` trigger corrections are reported in Tab. 9.3, Tab. 9.4 and Tab. 9.5, respectively. For the isospin asymmetry and neutral branching fraction measurement, the systematic uncertainties on A_{I} and $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$ are listed in Tables 9.6, 9.7, and 9.8.

Table 9.3.: Systematic uncertainty on $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, arising from the reweighting of the B meson transverse momentum p_T and pseudorapidity η, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{\mathcal{B}(B^+)}$
$[0.1, 0.98]$	$2.5 \cdot 10^{-5}$
$[1.1, 2.0]$	$1.7 \cdot 10^{-5}$
$[2.0, 3.0]$	$1.6 \cdot 10^{-5}$
$[3.0, 4.0]$	$1.8 \cdot 10^{-5}$
$[4.0, 5.0]$	$1.6 \cdot 10^{-5}$
$[5.0, 6.0]$	$1.3 \cdot 10^{-5}$
$[6.0, 7.0]$	$8.7 \cdot 10^{-6}$
$[7.0, 8.0]$	$1.8 \cdot 10^{-7}$
$[11.0, 11.75]$	$5.9 \cdot 10^{-6}$
$[11.75, 12.5]$	$7.7 \cdot 10^{-6}$
$[15.0, 16.0]$	$1.3 \cdot 10^{-6}$
$[16.0, 17.0]$	$6.5 \cdot 10^{-6}$
$[17.0, 18.0]$	$8.3 \cdot 10^{-8}$
$[18.0, 19.0]$	$2.4 \cdot 10^{-6}$
$[19.0, 20.0]$	$4.0 \cdot 10^{-6}$
$[20.0, 21.0]$	$3.0 \cdot 10^{-6}$
$[21.0, 22.0]$	$6.4 \cdot 10^{-6}$

Table 9.4.: Systematic uncertainty on $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, arising from the reweighting of the `nTracks` variable, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{\mathcal{B}(B^+)}$
$[0.1, 0.98]$	$3.0 \cdot 10^{-6}$
$[1.1, 2.0]$	$2.4 \cdot 10^{-6}$
$[2.0, 3.0]$	$1.7 \cdot 10^{-6}$
$[3.0, 4.0]$	$1.6 \cdot 10^{-6}$
$[4.0, 5.0]$	$1.2 \cdot 10^{-6}$
$[5.0, 6.0]$	$3.4 \cdot 10^{-7}$
$[6.0, 7.0]$	$6.1 \cdot 10^{-8}$
$[7.0, 8.0]$	$2.3 \cdot 10^{-7}$
$[11.0, 11.75]$	$1.0 \cdot 10^{-7}$
$[11.75, 12.5]$	$3.7 \cdot 10^{-7}$
$[15.0, 16.0]$	$2.5 \cdot 10^{-7}$
$[16.0, 17.0]$	$1.8 \cdot 10^{-7}$
$[17.0, 18.0]$	$2.0 \cdot 10^{-7}$
$[18.0, 19.0]$	$1.8 \cdot 10^{-7}$
$[19.0, 20.0]$	$1.6 \cdot 10^{-7}$
$[20.0, 21.0]$	$6.2 \cdot 10^{-7}$
$[21.0, 22.0]$	$3.1 \cdot 10^{-7}$

Table 9.5.: Systematic uncertainty on $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, arising from the correction of the L0 trigger efficiencies, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{\mathcal{B}(B^+)}$
$[0.1, 0.98]$	$3.0 \cdot 10^{-6}$
$[1.1, 2.0]$	$2.6 \cdot 10^{-6}$
$[2.0, 3.0]$	$2.5 \cdot 10^{-6}$
$[3.0, 4.0]$	$2.2 \cdot 10^{-6}$
$[4.0, 5.0]$	$1.8 \cdot 10^{-6}$
$[5.0, 6.0]$	$1.4 \cdot 10^{-6}$
$[6.0, 7.0]$	$1.1 \cdot 10^{-6}$
$[7.0, 8.0]$	$7.2 \cdot 10^{-7}$
$[11.0, 11.75]$	$4.2 \cdot 10^{-7}$
$[11.75, 12.5]$	$4.9 \cdot 10^{-7}$
$[15.0, 16.0]$	$5.5 \cdot 10^{-7}$
$[16.0, 17.0]$	$6.0 \cdot 10^{-7}$
$[17.0, 18.0]$	$5.6 \cdot 10^{-7}$
$[18.0, 19.0]$	$5.1 \cdot 10^{-7}$
$[19.0, 20.0]$	$4.7 \cdot 10^{-7}$
$[20.0, 21.0]$	$6.2 \cdot 10^{-7}$
$[21.0, 22.0]$	$3.1 \cdot 10^{-7}$

Table 9.6.: Systematic uncertainty on A_I and $\mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-)$, arising from the reweighting of the B meson transverse momentum p_T and pseudorapidity η, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{A_\mathrm{I}}$	$\sigma^{\mathrm{syst}}_{\mathcal{B}_\mathrm{rel}}$
$[0.1, 0.98]$	$3.5 \cdot 10^{-3}$	$2.2 \cdot 10^{-5}$
$[1.1, 2.5]$	$3.7 \cdot 10^{-4}$	$1.4 \cdot 10^{-5}$
$[2.5, 4.0]$	$8.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-5}$
$[4.0, 6.0]$	$6.5 \cdot 10^{-3}$	$7.7 \cdot 10^{-6}$
$[6.0, 8.0]$	$1.9 \cdot 10^{-3}$	$2.1 \cdot 10^{-6}$
$[11.0, 12.5]$	$4.1 \cdot 10^{-3}$	$5.4 \cdot 10^{-6}$
$[15.0, 17.0]$	$9.0 \cdot 10^{-3}$	$4.4 \cdot 10^{-6}$
$[17.0, 19.0]$	$2.7 \cdot 10^{-3}$	$5.0 \cdot 10^{-6}$

Table 9.7.: Systematic uncertainty on A_{I} and $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$, arising from the reweighting of the `nTracks` variable, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{A_{\mathrm{I}}}$	$\sigma^{\mathrm{syst}}_{\mathcal{B}_{\mathrm{rel}}}$
$[0.1, 0.98]$	$1.7 \cdot 10^{-3}$	$5.3 \cdot 10^{-6}$
$[1.1, 2.5]$	$4.6 \cdot 10^{-4}$	$1.8 \cdot 10^{-6}$
$[2.5, 4.0]$	$7.0 \cdot 10^{-4}$	$8.3 \cdot 10^{-7}$
$[4.0, 6.0]$	$8.0 \cdot 10^{-4}$	$1.7 \cdot 10^{-6}$
$[6.0, 8.0]$	$8.2 \cdot 10^{-4}$	$8.8 \cdot 10^{-7}$
$[11.0, 12.5]$	$6.6 \cdot 10^{-4}$	$9.8 \cdot 10^{-7}$
$[15.0, 17.0]$	$1.7 \cdot 10^{-3}$	$1.3 \cdot 10^{-7}$
$[17.0, 19.0]$	$2.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-6}$

Table 9.8.: Systematic uncertainty on A_{I} and $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$, arising from the correction of the L0 trigger efficiencies, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{A_{\mathrm{I}}}$	$\sigma^{\mathrm{syst}}_{\mathcal{B}_{\mathrm{rel}}}$
$[0.1, 0.98]$	$1.3 \cdot 10^{-3}$	$1.7 \cdot 10^{-6}$
$[1.1, 2.5]$	$9.5 \cdot 10^{-6}$	$2.8 \cdot 10^{-6}$
$[2.5, 4.0]$	$5.9 \cdot 10^{-5}$	$2.3 \cdot 10^{-6}$
$[4.0, 6.0]$	$9.5 \cdot 10^{-5}$	$1.8 \cdot 10^{-6}$
$[6.0, 8.0]$	$3.8 \cdot 10^{-4}$	$5.3 \cdot 10^{-7}$
$[11.0, 12.5]$	$1.8 \cdot 10^{-4}$	$3.0 \cdot 10^{-7}$
$[15.0, 17.0]$	$1.9 \cdot 10^{-4}$	$7.2 \cdot 10^{-7}$
$[17.0, 19.0]$	$3.9 \cdot 10^{-4}$	$3.0 \cdot 10^{-7}$

9.3. Particle identification corrections

The particle identification responses are corrected or resampled using high-statistics calibration samples (using the `PIDCorr` package) to determine probability density functions depending on track pseudorapidity η, transverse momentum p_T, and detector occupancy `nTracks`. In this process, two possible sources of systematic uncertainties are introduced. Firstly, the calibration samples have a limited size, but the information is lost when smoothing the high-resolution histograms, and is not propagated to the *PDF* definition. Moreover, the smoothing procedure depends on the choice of a Gaussian kernel, in particular its width, and, like for the fit model, there are no strong motivations to prefer one kernel to another. Since it is non-trivial to include these effects in the standard propagation of the statistical uncertainty, a strategy based on pseudo-experiments is defined.

To account for the limited size, different sets of corrections can be computed by *bootstrapping* the calibration samples, meaning randomly selecting only a subsample. This is repeated three times and three different sets of corrections are therefore obtained. Another set of corrections is derived using a different set of kernel widths for the Gaussian smoothing of the complete samples. As in the case of correction weights, particle identification corrections mainly affect efficiencies. The four alternative sets of corrections lead to four different sets of efficiencies whose effect on yields and finally on the quantities of interest is studied on pseudo-experiments.

The same toy framework already used in Sec. 9.1 and Sec. 9.2 is adopted, generating 1000 toys with 100 times the expected events for each q^2 interval using the nominal model. The datasets are then fitted five times, the first time expressing the yields in terms of the nominal efficiencies, and the other four times using the alternative sets of efficiencies. For each q^2 interval and for each set of alternative efficiencies, the distribution of the difference between the obtained results with respect to the nominal reference is studied. For the four distributions, the mean indicates the average bias introduced by the specific corrections. In this case, the largest mean difference among the four scenarios is taken as the systematic uncertainty. This evaluation is performed for both the single branching fraction measurement and the simultaneous fit to the asymmetry and neutral branching fraction. For the single branching fraction measurement of the charged mode, the systematic uncertainty on $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$ is reported in Tab. 9.9. For the isospin asymmetry and neutral branching fraction measurement, the systematic uncertainties on A_I and $\mathcal{B}_\mathrm{rel}(B^0 \to K^0\mu^+\mu^-)$ are listed in Tab. 9.10.

9.4. Normalisation inputs

The quantities $S_{LL/DD}$ and S_+ defined in Eq. (8.16) and Eq. (8.18), as well as the factor S_{B^+} for the charged mode measurement (Eq. (8.9)), depend on inputs determined within the analysis, like the efficiencies and the yields of normalisation channels, and on external inputs taken from literature, such as the branching fractions of the resonant modes and B meson lifetimes. Although these uncertainties could be propagated directly within the fit procedure by treating the S factors as Gaussian-constrained parameters, they are fixed to their central values to ensure fit stability. Consequently, their contribution is evaluated separately as a systematic uncertainty using pseudo-experiments.

A set of 1000 toys per q^2 interval is generated according to the nominal model, as described in Sec. 9.1. Crucially, during the generation, the S factors are fixed to their central values. The datasets are then fitted twice. In the first fit, the S factors are fixed to their nominal values. In the second fit, the S

Table 9.9.: Systematic uncertainty on $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$, arising from the resampling of the particle identification variables, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{\mathcal{B}(B^+)}$
$[0.1, 0.98]$	$3.1 \cdot 10^{-8}$
$[1.1, 2.0]$	$1.1 \cdot 10^{-7}$
$[2.0, 3.0]$	$3.2 \cdot 10^{-8}$
$[3.0, 4.0]$	$9.6 \cdot 10^{-8}$
$[4.0, 5.0]$	$7.4 \cdot 10^{-8}$
$[5.0, 6.0]$	$3.7 \cdot 10^{-8}$
$[6.0, 7.0]$	$8.0 \cdot 10^{-8}$
$[7.0, 8.0]$	$2.8 \cdot 10^{-8}$
$[11.0, 11.75]$	$2.8 \cdot 10^{-8}$
$[11.75, 12.5]$	$6.7 \cdot 10^{-8}$
$[15.0, 16.0]$	$2.2 \cdot 10^{-7}$
$[16.0, 17.0]$	$3.8 \cdot 10^{-8}$
$[17.0, 18.0]$	$3.9 \cdot 10^{-8}$
$[18.0, 19.0]$	$6.0 \cdot 10^{-8}$
$[19.0, 20.0]$	$1.2 \cdot 10^{-8}$
$[20.0, 21.0]$	$6.2 \cdot 10^{-8}$
$[21.0, 22.0]$	$3.0 \cdot 10^{-8}$

Table 9.10.: Systematic uncertainty on A_{I} and $\mathcal{B}_{\mathrm{rel}}(B^0 \rightarrow K^0\mu^+\mu^-)$, arising from the resampling of the particle identification variables, as a function of q^2.

q^2 region	Systematic uncertainty	
$[\mathrm{GeV}^2/\mathrm{c}^4]$	$\sigma^{\mathrm{syst}}_{A_{\mathrm{I}}}$	$\sigma^{\mathrm{syst}}_{\mathcal{B}_{\mathrm{rel}}}$
$[0.1, 0.98]$	$5.9 \cdot 10^{-5}$	$5.3 \cdot 10^{-8}$
$[1.1, 2.5]$	$4.9 \cdot 10^{-5}$	$4.6 \cdot 10^{-8}$
$[2.5, 4.0]$	$5.0 \cdot 10^{-5}$	$5.2 \cdot 10^{-8}$
$[4.0, 6.0]$	$2.7 \cdot 10^{-5}$	$3.0 \cdot 10^{-8}$
$[6.0, 8.0]$	$3.7 \cdot 10^{-5}$	$3.6 \cdot 10^{-8}$
$[11.0, 12.5]$	$5.9 \cdot 10^{-5}$	$3.7 \cdot 10^{-8}$
$[15.0, 17.0]$	$1.9 \cdot 10^{-4}$	$1.5 \cdot 10^{-8}$
$[17.0, 19.0]$	$1.7 \cdot 10^{-4}$	$7.1 \cdot 10^{-8}$

factors are fixed to randomised values, determined by varying all the inputs according to their uncertainties. The mean difference between the results obtained in the two scenarios is assigned as a systematic uncertainty. This evaluation is performed for both fit configurations. For the single branching fraction measurement of the charged mode, the sources of uncertainty are the branching fractions $\mathcal{B}(B^+ \rightarrow J/\psi K^+)$ and $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)$, and the selection efficiencies. The resulting systematic uncertainty on $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$ is reported in Tab. 9.11. For the isospin asymmetry and neutral branching fraction measurement, the lifetime ratio is considered, in addition to the branching fractions and selection efficiencies. The systematic uncertainties on A_I and $\mathcal{B}_\mathrm{rel}(B^0 \rightarrow K^0\mu^+\mu^-)$ are listed in Tab. 9.12.

Table 9.11.: Systematic uncertainty on $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$, arising from the limited precision of the inputs determining the normalisation factors used in the fit model, as a function of q^2.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Systematic uncertainty $\sigma^{\mathrm{syst}}_{\mathcal{B}(B^+)}$
$[0.1, 0.98]$	$8.0 \cdot 10^{-8}$
$[1.1, 2.0]$	$5.9 \cdot 10^{-8}$
$[2.0, 3.0]$	$1.2 \cdot 10^{-7}$
$[3.0, 4.0]$	$1.2 \cdot 10^{-7}$
$[4.0, 5.0]$	$2.9 \cdot 10^{-8}$
$[5.0, 6.0]$	$2.5 \cdot 10^{-7}$
$[6.0, 7.0]$	$6.5 \cdot 10^{-9}$
$[7.0, 8.0]$	$7.4 \cdot 10^{-8}$
$[11.0, 11.75]$	$2.2 \cdot 10^{-8}$
$[11.75, 12.5]$	$3.8 \cdot 10^{-8}$
$[15.0, 16.0]$	$1.8 \cdot 10^{-7}$
$[16.0, 17.0]$	$3.1 \cdot 10^{-8}$
$[17.0, 18.0]$	$5.7 \cdot 10^{-9}$
$[18.0, 19.0]$	$9.6 \cdot 10^{-9}$
$[19.0, 20.0]$	$9.0 \cdot 10^{-8}$
$[20.0, 21.0]$	$8.0 \cdot 10^{-8}$
$[21.0, 22.0]$	$3.5 \cdot 10^{-8}$

9.5. Total uncertainty

The individual contributions to the systematic uncertainty discussed in the previous sections are combined to estimate the total systematic error affecting the measurements. As the sources of uncertainty are considered independent, the total systematic uncertainty is obtained by summing the partial contributions in quadrature. In this final section, the detailed breakdown of the systematic uncertainties is presented separately for each physical observable. To establish the precision regime of the analysis, the total systematic uncertainty is explicitly compared with the statistical uncertainty obtained from the likelihood fits described in Ch. 8.

Table 9.12.: Systematic uncertainty on A_{I} and $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$, arising from the limited precision of the inputs determining the normalisation factors used in the fit model, as a function of q^2.

q^2 region	Systematic uncertainty	
$[\mathrm{GeV}^2/\mathrm{c}^4]$	$\sigma^{\mathrm{syst}}_{A_{\mathrm{I}}}$	$\sigma^{\mathrm{syst}}_{\mathcal{B}_{\mathrm{rel}}}$
$[0.1, 0.98]$	$3.8 \cdot 10^{-4}$	$2.7 \cdot 10^{-7}$
$[1.1, 2.5]$	$2.0 \cdot 10^{-4}$	$1.5 \cdot 10^{-8}$
$[2.5, 4.0]$	$5.4 \cdot 10^{-5}$	$3.0 \cdot 10^{-8}$
$[4.0, 6.0]$	$1.0 \cdot 10^{-3}$	$2.5 \cdot 10^{-7}$
$[6.0, 8.0]$	$8.1 \cdot 10^{-5}$	$8.7 \cdot 10^{-8}$
$[11.0, 12.5]$	$5.1 \cdot 10^{-4}$	$1.2 \cdot 10^{-7}$
$[15.0, 17.0]$	$2.3 \cdot 10^{-4}$	$1.8 \cdot 10^{-7}$
$[17.0, 19.0]$	$5.4 \cdot 10^{-4}$	$9.8 \cdot 10^{-8}$

Tab. 9.13 reports the error budget for the branching fraction of the charged mode, $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$. Similarly, the breakdowns for the neutral relative branching fraction $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$ and the isospin asymmetry A_{I} are presented in Tab. 9.14 and Tab. 9.15, respectively. Overall, the uncertainty arising from the p_{T} and η reweighting (σ_{kin}) is the dominant source, often reaching values comparable to the statistical error. This large contribution is a direct consequence of the conservative approach adopted for this estimation, where the entire effect of the correction weights is taken as a systematic uncertainty. While this strategy ensures a robust upper bound on the systematic error, it likely overestimates the true bias. The impact of this systematic source is more pronounced for the branching fractions $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$ and $\mathcal{B}_{\mathrm{rel}}(B^0 \to K^0\mu^+\mu^-)$. Conversely, the isospin asymmetry A_{I} benefits from a partial cancellation of the systematic effects in the ratio. In general, while the current evaluation is acceptable for the other systematic sources, the kinematic weighting systematic may require a more refined study (for example, assessing the impact of an alternative histogram-based correction rather than removing the weights entirely) before the final unblinding of the data.

Table 9.13.: Breakdown of systematic uncertainties for $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$. The total systematic uncertainty σ_{syst}, obtained summing in quadrature the uncertainties for all the sources, is compared with the statistical uncertainty σ_{stat}.

q^2 region $[\text{GeV}^2/\text{c}^4]$	Fit Model σ_{model}	Kinematic σ_{kin}	Multiplicity σ_{mult}	L0 σ_{L0}	PID σ_{PID}	Normalisation σ_{norm}	**Total Syst** $\boldsymbol{\sigma_{\text{syst}}}$	**Stat** $\boldsymbol{\sigma_{\text{stat}}}$
$[0.1, 0.98]$	$6.0 \cdot 10^{-10}$	$2.5 \cdot 10^{-5}$	$3.0 \cdot 10^{-6}$	$3.0 \cdot 10^{-6}$	$3.1 \cdot 10^{-8}$	$8.0 \cdot 10^{-8}$	$2.5 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
$[1.1, 2.0]$	$6.7 \cdot 10^{-10}$	$1.7 \cdot 10^{-5}$	$2.4 \cdot 10^{-6}$	$2.6 \cdot 10^{-6}$	$1.1 \cdot 10^{-7}$	$5.9 \cdot 10^{-8}$	$1.7 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
$[2.0, 3.0]$	$5.9 \cdot 10^{-10}$	$1.6 \cdot 10^{-5}$	$1.7 \cdot 10^{-6}$	$2.5 \cdot 10^{-6}$	$3.2 \cdot 10^{-8}$	$1.2 \cdot 10^{-7}$	$1.6 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
$[3.0, 4.0]$	$5.9 \cdot 10^{-10}$	$1.8 \cdot 10^{-5}$	$1.6 \cdot 10^{-6}$	$2.2 \cdot 10^{-6}$	$9.6 \cdot 10^{-8}$	$1.2 \cdot 10^{-7}$	$1.8 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$
$[4.0, 5.0]$	$6.5 \cdot 10^{-10}$	$1.6 \cdot 10^{-5}$	$1.2 \cdot 10^{-6}$	$1.8 \cdot 10^{-6}$	$7.4 \cdot 10^{-8}$	$2.9 \cdot 10^{-8}$	$1.6 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
$[5.0, 6.0]$	$6.1 \cdot 10^{-10}$	$1.3 \cdot 10^{-5}$	$3.4 \cdot 10^{-7}$	$1.4 \cdot 10^{-6}$	$3.7 \cdot 10^{-8}$	$2.5 \cdot 10^{-7}$	$1.3 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
$[6.0, 7.0]$	$6.1 \cdot 10^{-10}$	$8.7 \cdot 10^{-6}$	$6.1 \cdot 10^{-8}$	$1.1 \cdot 10^{-6}$	$8.0 \cdot 10^{-8}$	$6.5 \cdot 10^{-9}$	$8.8 \cdot 10^{-6}$	$1.2 \cdot 10^{-5}$
$[7.0, 8.0]$	$6.0 \cdot 10^{-10}$	$1.8 \cdot 10^{-7}$	$2.3 \cdot 10^{-7}$	$7.2 \cdot 10^{-7}$	$2.8 \cdot 10^{-8}$	$7.4 \cdot 10^{-8}$	$7.8 \cdot 10^{-7}$	$1.2 \cdot 10^{-5}$
$[11.0, 11.75]$	$5.8 \cdot 10^{-10}$	$5.9 \cdot 10^{-6}$	$1.0 \cdot 10^{-7}$	$4.2 \cdot 10^{-7}$	$2.8 \cdot 10^{-8}$	$2.2 \cdot 10^{-8}$	$5.9 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$
$[11.75, 12.5]$	$5.3 \cdot 10^{-10}$	$7.7 \cdot 10^{-6}$	$3.7 \cdot 10^{-7}$	$4.9 \cdot 10^{-7}$	$6.7 \cdot 10^{-8}$	$3.8 \cdot 10^{-8}$	$7.7 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$
$[15.0, 16.0]$	$6.7 \cdot 10^{-10}$	$1.3 \cdot 10^{-6}$	$2.5 \cdot 10^{-7}$	$5.5 \cdot 10^{-7}$	$2.2 \cdot 10^{-7}$	$1.8 \cdot 10^{-7}$	$1.5 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$
$[16.0, 17.0]$	$5.6 \cdot 10^{-10}$	$6.5 \cdot 10^{-6}$	$1.8 \cdot 10^{-7}$	$6.0 \cdot 10^{-7}$	$3.8 \cdot 10^{-8}$	$3.1 \cdot 10^{-8}$	$6.5 \cdot 10^{-6}$	$9.6 \cdot 10^{-6}$
$[17.0, 18.0]$	$5.5 \cdot 10^{-10}$	$8.3 \cdot 10^{-8}$	$2.0 \cdot 10^{-7}$	$5.6 \cdot 10^{-7}$	$3.9 \cdot 10^{-8}$	$5.7 \cdot 10^{-9}$	$6.0 \cdot 10^{-7}$	$8.8 \cdot 10^{-6}$
$[18.0, 19.0]$	$5.2 \cdot 10^{-10}$	$2.4 \cdot 10^{-6}$	$1.8 \cdot 10^{-7}$	$5.1 \cdot 10^{-7}$	$6.0 \cdot 10^{-8}$	$9.6 \cdot 10^{-9}$	$2.5 \cdot 10^{-6}$	$8.2 \cdot 10^{-6}$
$[19.0, 20.0]$	$5.9 \cdot 10^{-10}$	$4.0 \cdot 10^{-6}$	$1.6 \cdot 10^{-7}$	$4.7 \cdot 10^{-7}$	$1.2 \cdot 10^{-8}$	$9.0 \cdot 10^{-8}$	$4.0 \cdot 10^{-6}$	$7.9 \cdot 10^{-6}$
$[20.0, 21.0]$	$6.4 \cdot 10^{-10}$	$3.0 \cdot 10^{-6}$	$6.2 \cdot 10^{-7}$	$6.2 \cdot 10^{-7}$	$6.2 \cdot 10^{-8}$	$8.0 \cdot 10^{-8}$	$3.1 \cdot 10^{-6}$	$7.1 \cdot 10^{-6}$
$[21.0, 22.0]$	$5.2 \cdot 10^{-10}$	$6.4 \cdot 10^{-6}$	$3.1 \cdot 10^{-7}$	$3.1 \cdot 10^{-7}$	$3.0 \cdot 10^{-8}$	$3.5 \cdot 10^{-8}$	$6.4 \cdot 10^{-6}$	$5.5 \cdot 10^{-6}$

Table 9.14.: Breakdown of systematic uncertainties for $\mathcal{B}_{\text{rel}}(B^0 \to K^0\mu^+\mu^-)$. The total systematic uncertainty σ_{syst}, obtained summing in quadrature the uncertainties for all the sources, is compared with the statistical uncertainty σ_{stat}.

q^2 region $[\text{GeV}^2/\text{c}^4]$	Fit Model σ_{model}	Kinematic σ_{kin}	Multiplicity σ_{mult}	L0 σ_{L0}	PID σ_{PID}	Normalisation σ_{norm}	**Total Syst** $\boldsymbol{\sigma_{\text{syst}}}$	**Stat** $\boldsymbol{\sigma_{\text{stat}}}$
$[0.1, 0.98]$	$9.2 \cdot 10^{-9}$	$2.2 \cdot 10^{-5}$	$5.3 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$5.3 \cdot 10^{-8}$	$2.7 \cdot 10^{-7}$	$2.3 \cdot 10^{-5}$	$4.4 \cdot 10^{-5}$
$[1.1, 2.5]$	$3.7 \cdot 10^{-9}$	$1.4 \cdot 10^{-5}$	$1.8 \cdot 10^{-6}$	$2.8 \cdot 10^{-6}$	$4.6 \cdot 10^{-8}$	$1.5 \cdot 10^{-8}$	$1.4 \cdot 10^{-5}$	$4.2 \cdot 10^{-5}$
$[2.5, 4.0]$	$1.4 \cdot 10^{-8}$	$1.1 \cdot 10^{-5}$	$8.3 \cdot 10^{-7}$	$2.3 \cdot 10^{-6}$	$5.2 \cdot 10^{-8}$	$3.0 \cdot 10^{-8}$	$1.1 \cdot 10^{-5}$	$4.6 \cdot 10^{-5}$
$[4.0, 6.0]$	$4.3 \cdot 10^{-9}$	$7.7 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$1.8 \cdot 10^{-6}$	$3.0 \cdot 10^{-8}$	$2.5 \cdot 10^{-7}$	$8.1 \cdot 10^{-6}$	$4.1 \cdot 10^{-5}$
$[6.0, 8.0]$	$2.6 \cdot 10^{-9}$	$2.1 \cdot 10^{-6}$	$8.8 \cdot 10^{-7}$	$5.3 \cdot 10^{-7}$	$3.6 \cdot 10^{-8}$	$8.7 \cdot 10^{-8}$	$2.3 \cdot 10^{-6}$	$4.4 \cdot 10^{-5}$
$[11.0, 12.5]$	$9.8 \cdot 10^{-9}$	$5.4 \cdot 10^{-6}$	$9.8 \cdot 10^{-7}$	$3.0 \cdot 10^{-7}$	$3.7 \cdot 10^{-8}$	$1.2 \cdot 10^{-7}$	$5.5 \cdot 10^{-6}$	$3.8 \cdot 10^{-5}$
$[15.0, 17.0]$	$4.3 \cdot 10^{-9}$	$4.4 \cdot 10^{-6}$	$1.3 \cdot 10^{-7}$	$7.2 \cdot 10^{-7}$	$1.5 \cdot 10^{-8}$	$1.8 \cdot 10^{-7}$	$4.5 \cdot 10^{-6}$	$3.2 \cdot 10^{-5}$
$[17.0, 19.0]$	$1.1 \cdot 10^{-8}$	$5.0 \cdot 10^{-6}$	$1.4 \cdot 10^{-6}$	$3.0 \cdot 10^{-7}$	$7.1 \cdot 10^{-8}$	$9.8 \cdot 10^{-8}$	$5.2 \cdot 10^{-6}$	$2.8 \cdot 10^{-5}$

Table 9.15.: Breakdown of systematic uncertainties for A_{I}. The total systematic uncertainty σ_{syst}, obtained summing in quadrature the uncertainties for all the sources, is compared with the statistical uncertainty σ_{stat}.

q^2 region $[\mathrm{GeV}^2/\mathrm{c}^4]$	Fit Model σ_{model}	Kinematic σ_{kin}	Multiplicity σ_{mult}	`L0` $\sigma_{\mathtt{L0}}$	PID σ_{PID}	Normalisation σ_{norm}	**Total Syst** $\boldsymbol{\sigma}_{\mathrm{syst}}$	**Stat** $\boldsymbol{\sigma}_{\mathrm{stat}}$
$[0.1, 0.98]$	$7.3 \cdot 10^{-6}$	$3.5 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$5.9 \cdot 10^{-5}$	$3.8 \cdot 10^{-4}$	$4.1 \cdot 10^{-3}$	$3.5 \cdot 10^{-2}$
$[1.1, 2.5]$	$4.5 \cdot 10^{-6}$	$3.7 \cdot 10^{-4}$	$4.6 \cdot 10^{-4}$	$9.5 \cdot 10^{-6}$	$4.9 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$	$4.0 \cdot 10^{-2}$
$[2.5, 4.0]$	$1.3 \cdot 10^{-5}$	$8.9 \cdot 10^{-3}$	$7.0 \cdot 10^{-4}$	$5.9 \cdot 10^{-5}$	$5.0 \cdot 10^{-5}$	$5.4 \cdot 10^{-5}$	$8.9 \cdot 10^{-3}$	$3.9 \cdot 10^{-2}$
$[4.0, 6.0]$	$2.6 \cdot 10^{-6}$	$6.5 \cdot 10^{-3}$	$8.0 \cdot 10^{-4}$	$9.5 \cdot 10^{-5}$	$2.7 \cdot 10^{-5}$	$1.0 \cdot 10^{-3}$	$6.6 \cdot 10^{-3}$	$4.1 \cdot 10^{-2}$
$[6.0, 8.0]$	$1.6 \cdot 10^{-6}$	$1.9 \cdot 10^{-3}$	$8.2 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$3.7 \cdot 10^{-5}$	$8.1 \cdot 10^{-5}$	$2.1 \cdot 10^{-3}$	$3.8 \cdot 10^{-2}$
$[11.0, 12.5]$	$8.0 \cdot 10^{-6}$	$4.1 \cdot 10^{-3}$	$6.6 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$5.9 \cdot 10^{-5}$	$5.1 \cdot 10^{-4}$	$4.2 \cdot 10^{-3}$	$3.9 \cdot 10^{-2}$
$[15.0, 17.0]$	$6.6 \cdot 10^{-6}$	$9.0 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$9.2 \cdot 10^{-3}$	$4.5 \cdot 10^{-2}$
$[17.0, 19.0]$	$2.2 \cdot 10^{-5}$	$2.7 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$3.9 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	$5.4 \cdot 10^{-4}$	$3.7 \cdot 10^{-3}$	$5.1 \cdot 10^{-2}$

10. Results and prospects

The analysis presented in this thesis describes the strategy for the measurement of the differential branching fractions and the isospin asymmetry of the $B \to K\mu^+\mu^-$ decays using the full Run 1 and Run 2 dataset collected by the LHCb experiment. The procedure relies on simultaneous unbinned maximum likelihood fits to the charged and neutral resonant and rare modes, designed to extract the physical observables while automatically handling the propagation of statistical uncertainties. As the analysis is performed blindly, to avoid experimenter's bias, the final values of the observables on data are not accessible at this stage. However, the sensitivity of the measurement, the evaluation of the error budget, and the comparison with previous results can be discussed.

10.1. Sensitivity and error budget

A key outcome of this work is the definition of the simultaneous fit strategy to measure the quantities of interest A_I, $\mathcal{B}(B^0 \to K^0\mu^+\mu^-)$ and $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$. The validation studies performed on pseudo-experiments, described in Ch. 8, demonstrate the robustness of the fitting framework. The pulls of the physical parameters are compatible with a standard normal distribution across the majority of the q^2 intervals, ensuring correct statistical coverage. However, a small bias observed in the neutral branching fraction, attributed to finite sample size effects in low-yield bins, will be re-evaluated and accounted for as a systematic uncertainty once the actual data yields are accessible. With this provision, the analysis framework is considered fully validated and ready to be applied to the unblinded dataset to extract the final results.

Complementing the robust fit definition, a comprehensive evaluation of the systematic uncertainties has been presented in Ch. 9. The comparison between the systematic and statistical uncertainties, summarised in Sec. 9.5, reveals distinct regimes depending on the source considered. With the exception of the kinematic reweighting, the systematic uncertainties arising from the fit model, the correction chain, and the normalisation inputs are generally an order of magnitude smaller than the statistical uncertainty derived from the fit to data. This indicates that, for the majority of the effects, the analysis strategy is well-optimised and statistically limited. However, the systematic uncertainty associated with the p_{T} and η weighting is currently comparable to the statistical error. As discussed, this is attributed to the conservative estimation method adopted, which likely sets a loose upper bound on the bias. Excluding this specific source, the total systematic uncertainty would be negligible compared to the statistical precision. Therefore, while the strategy is defined, a refined evaluation of the kinematic systematic is identified as one possible additional step before unblinding, to ensure the total uncertainty reflects the true potential of the dataset.

Table 10.1.: Comparison between the total uncertainty on A_I measured in Run 1 (σ_{Run1}) and the expected total uncertainty of this analysis ($\sigma_{\text{Run1+2}}$). The improvement factor represents the ratio $\sigma_{\text{Run1}}/\sigma_{\text{Run1+2}}$.

q^2 region $[\text{GeV}^2/\text{c}^4]$	Run 1 Total Unc. σ_{Run1}	**Run 1+2 Unc. (this work)** $\boldsymbol{\sigma_{\text{Run1+2}}}$	**Improvement Factor**
$[0.1, 0.98]$	0.20	0.035	5.7
$[1.1, 2.5]$	0.20	0.040	5.0
$[2.5, 4.0]$	0.14	0.040	3.5
$[4.0, 6.0]$	0.15	0.042	3.6
$[6.0, 8.0]$	0.11	0.038	2.9
$[11.0, 12.5]$	0.17	0.039	4.4
$[15.0, 17.0]$	0.12	0.046	2.6
$[17.0, 19.0]$	0.11	0.051	2.2

10.2. Comparison with previous measurements

The current reference measurement for the isospin asymmetry in $B \to K\mu^+\mu^-$ decays was performed by the LHCb collaboration using the Run 1 dataset ($3\,\text{fb}^{-1}$), corresponding to the 2011–2012 data taking period [57]. The analysis presented in this thesis utilises the full Run 1 and Run 2 dataset (2011–2012, 2015, 2016–2018), corresponding to an integrated luminosity of approximately $9\,\text{fb}^{-1}$. Beyond the factor of three increase in luminosity, the Run 2 data benefits from a higher b-production cross-section due to the increased centre-of-mass energy ($13\,\text{TeV}$ vs 7–$8\,\text{TeV}$).

In addition to the increase in sample size, the analysis takes advantage of significant improvements in the experimental software techniques. While the detector hardware remained largely unchanged between Run 1 and Run 2, the track reconstruction algorithms underwent substantial optimisation. Of particular importance for this measurement is the improved reconstruction of K^0_S mesons. Since the sensitivity to the isospin asymmetry is heavily dependent on the statistical precision of the neutral mode $B^0 \to K^0_\text{S}\mu^+\mu^-$, the enhanced efficiency and mass resolution for K^0_S candidates translate directly into a higher signal yield and an improved signal-to-background ratio compared to the previous analysis.

To quantify the progress represented by this work, it is useful to compare the expected precision using the full available dataset with the published uncertainties. Tab. 10.1 shows the comparison between the total uncertainty reported in the Run 1 measurement and the total uncertainty expected for this analysis (obtained by summing in quadrature the expected statistical error and the systematic uncertainties estimated in Ch. 9). The improvement factor, defined as the ratio between the Run 1 and Run 2 uncertainties, is also reported. A significant reduction in uncertainty is expected across all the kinematic intervals. This gain stems from the combination of the increased integrated luminosity, the higher production cross-section, and the refined reconstruction and selection strategies.

The significant improvement in precision is visualised in Fig. 10.1. The figure compares the statistical uncertainty of the reference Run 1 measurement with the statistical uncertainty of this analysis as a function of q^2.

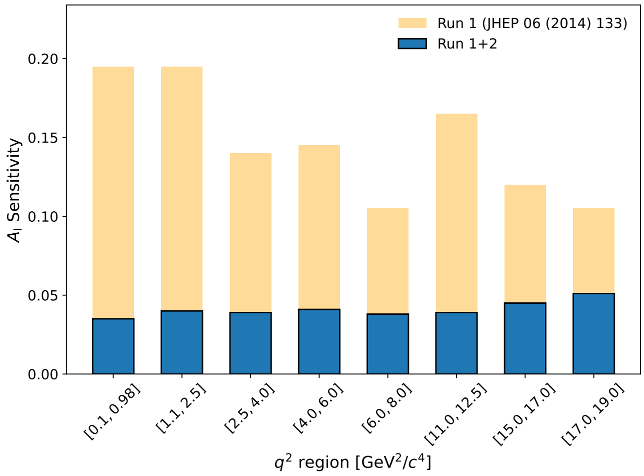


Figure 10.1.: Expected sensitivity for the isospin asymmetry A_I in the different q^2 intervals. The expected statistical uncertainty of this analysis (blue) is compared with the statistical uncertainty obtained in the Run 1 measurement (orange) [57].

10.3. Physics implications

The precise measurement of the branching fractions and isospin asymmetry of $B \to K\mu^+\mu^-$ decays is of particular importance in the context of the current flavour physics landscape. As discussed in Ch. 2, this observable offers a unique probe into the structure of New Physics. While standard analyses of branching fractions and angular observables primarily constrain the dominant Wilson coefficients C_7, C_9, and C_{10}, the isospin asymmetry provides access to "exotic" coefficients or specific four-quark operators that are otherwise difficult to constrain. In particular, A_I is sensitive to the gluonic operator O_8. In terms of physical processes, this implies that the asymmetry is sensitive to non-standard isospin-breaking effects, such as New Physics contributions that might couple differently to the up and down spectator quarks.

Theoretical predictions for A_I in the Standard Model are robust, as the leading form factor uncertainties largely cancel out in the ratio, particularly in the low-q^2 region. Consequently, any significant deviation from the expectation $A_I \approx 0$ would be a clear indication of non-factorisable QCD effects or New Physics signatures.

To illustrate the constraining power of the measurement presented in this thesis, Fig. 10.2 shows the expected sensitivity compared to the previous measurement, and to the Standard Model prediction. The expected data points are centred at the hypothesis of no asymmetry ($A_I = 0$), with error bars corresponding to the total expected precision derived in this work. For comparison, the theoretical uncertainty on the Standard Model prediction [29] is displayed. The comparison highlights that the experimental precision achieved in this analysis approaches the level of theoretical uncertainty, enabling a stringent test of the Standard Model.

Complementarily, the precise measurement of the differential branching fractions provides direct constraints on the Wilson coefficients C_9 and C_{10}, which describe the effective strength of the vector and axial-vector couplings. Global fits to $b \to s$ data have historically shown tensions with the Standard

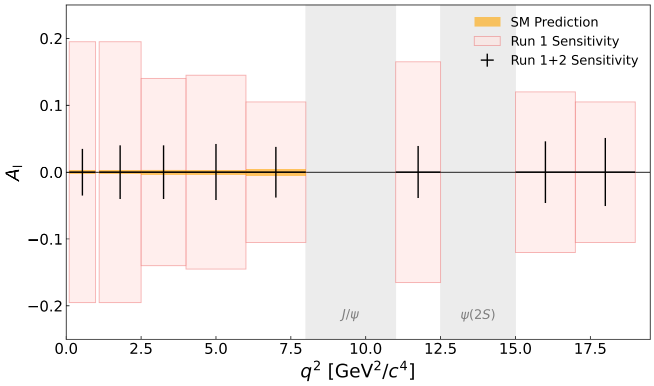


Figure 10.2.: Sensitivity of the A_I measurement assuming the null hypothesis ($A_I = 0$). The Standard Model prediction as a function of q^2 is computed with the `flavio` package [82].

Model predictions, often interpreting deficits in the branching fractions as a shift in C_9. The results of this analysis, with their improved precision and the use of the full Run 1 and Run 2 dataset, will add significant weight to these global fits, helping to clarify the nature of these anomalies.

11. Conclusion

The measurement of rare $b \to s\ell\ell$ transitions represents one of the most sensitive tools to probe the Standard Model of particle physics and search for New Physics effects. In this context, the study of the isospin asymmetry in $B \to K\mu^+\mu^-$ decays offers a unique opportunity. Unlike branching fractions or angular observables, which are often limited by hadronic form-factor uncertainties, the isospin asymmetry A_I benefits from the cancellation of the leading hadronic effects, providing a precise null test of the Standard Model. Any significant deviation from zero would unambiguously signal the presence of non-factorisable QCD effects or new interactions distinguishing between up and down spectator quarks.

This thesis presented a comprehensive strategy for the measurement of the differential branching fractions and the isospin asymmetry of $B \to K\mu^+\mu^-$ decays, utilising the full Run 1 and Run 2 datasets collected by the LHCb experiment, corresponding to an integrated luminosity of approximately $9\,\mathrm{fb}^{-1}$. The analysis leverages the high statistics and the improved experimental techniques of Run 2, particularly the optimised reconstruction of K^0_S mesons, to overcome the statistical limitations of previous measurements in the neutral decay mode. To achieve the required precision, a rigorous selection procedure was optimised to isolate the rare signal transitions from the high-rate background environment, employing multivariate classification techniques to suppress combinatorial noise and specific vetoes to exclude peaking backgrounds. Furthermore, particular attention was devoted to the accurate determination of the efficiencies. Data-driven methods were extensively used to correct the simulation for imperfections in the modelling of the particle identification response, the tracking efficiency, and the kinematic distributions of the B mesons.

A central achievement of this work is the development of a robust analysis framework capable of handling the complex background environment and the differences in reconstruction efficiency between the charged $B^+ \to K^+\mu^+\mu^-$ and neutral $B^0 \to K^0\mu^+\mu^-$ channels. To ensure optimal precision, tailored fitting configurations were defined for the different observables. The branching fraction of the charged decay is determined through a dedicated fit to the B^+ dataset. Conversely, the neutral branching fraction and the isospin asymmetry are extracted via a simultaneous unbinned maximum likelihood fit to both the charged and neutral modes. This simultaneous approach is particularly advantageous for A_I, as it allows the asymmetry to be determined directly as a free parameter of the fit, ensuring the correct propagation of statistical uncertainties and a consistent treatment of the normalisation channels, $B \to J/\psi K$.

As the analysis was performed in a blinded fashion to avoid experimenter's bias, the procedure was extensively validated using high-statistics pseudo-experiments. The validation studies demonstrated that the fitting framework provides unbiased estimates of the physical parameters and correct statistical coverage across the entire kinematic range. Residual effects due to the limited size of the sample are left to be studied after unblinding. Furthermore, a detailed evaluation of the systematic uncertainties was performed, covering effects related to the fit model parameterisation, corrections to simulations, and external inputs. While the conservative estimate of the kinematic reweighting uncertainty is currently comparable to the statistical error, all other sources of systematic uncertainty were found to be negligible. This indicates that, following the planned refinement of the kinematic systematic, the total uncertainty will be significantly smaller than the expected statistical uncertainty in all q^2 intervals. This result confirms that the measurement is projected to be statistically limited and that the analysis strategy is fully optimised for the available dataset.

The sensitivity studies indicate that this analysis will yield the most precise determination of the isospin asymmetry to date. The expected precision represents a significant improvement over the Run 1 reference measurement. This enhancement is driven by the combination of the larger dataset, the higher b-production cross-section at $\sqrt{s} = 13\,\mathrm{TeV}$, and the refined selection strategy.

In conclusion, the analysis framework presented in this thesis is complete, validated, and ready to be applied to the unblinded data. The forthcoming results will provide stringent constraints on the Wilson coefficients C_9 and C_{10}, contributing to the global effort to clarify the nature of the anomalies observed in $b \to s\ell\ell$ transitions. Moreover, the precise measurement of A_I will serve as a critical test of the Standard Model, with the potential to uncover subtle New Physics signatures or to significantly improve our understanding of hadronic effects in loop-induced decays.

Bibliography

[1] **LHCb** Collaboration, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state. Phys. Rev. Lett.* **112** (2014) .

[2] **LHCb** Collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays. Phys. Rev. Lett.* **115** (2015) .

[3] K. G. Wilson, *Confinement of quarks. Phys. Rev. D* **10** (1974) .

[4] P. W. Higgs, *Broken symmetries and the masses of gauge bosons. Phys. Rev. Lett.* **13** (1964) .

[5] **ATLAS** Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B* **716** (2012) .

[6] **CMS** Collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B* **716** (2012) .

[7] S. Weinberg, *The Quantum Theory of Fields*, vol. 1. Cambridge University Press, 1995.

[8] O. Klein, *Quantentheorie und fünfdimensionale Relativitätstheorie. Z. Physik* **37** (1926) .

[9] W. Gordon, *Der Comptoneffekt nach der Schrödingerschen Theorie. Z. Physik* **40** (1926) .

[10] P. A. M. Dirac, *The quantum theory of the electron. Proc. Roy. Soc. Lond. A* **117** (1928) .

[11] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance. Phys. Rev.* **96** (1954) .

[12] E. Noether, *Invariant variation problems. Trans. Theory Stat. Phys.* **1** (1971) .

[13] S. L. Glashow, *Partial-symmetries of weak interactions. Nucl. Phys.* **22** (1961) .

[14] S. Weinberg, *A model of leptons. Phys. Rev. Lett.* **19** (1967) .

[15] A. Salam, *Weak and electromagnetic interactions. Conf. Proc. C* **680519** (1968) .

[16] R. W. McAllister and R. Hofstadter, *Elastic scattering of 188-MeV electrons from the proton and the alpha particle. Phys. Rev.* **102** (1956) .

[17] R. P. Feynman, *Very high-energy collisions of hadrons. Phys. Rev. Lett.* **23** (1969) .

[18] J. D. Bjorken and E. A. Paschos, *Inelastic electron-proton and γ-proton scattering and the structure of the nucleon. Phys. Rev.* **185** (1969) .

[19] M. Gell-Mann, *Symmetries of baryons and mesons. Phys. Rev.* **125** (1962) .

[20] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry. Phys. Rev. D* **2** (1970) .

[21] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction. Prog. Theor. Phys.* **49** (1973) .

[22] L. Wolfenstein, *Parametrization of the kobayashi-maskawa matrix. Phys. Rev. Lett.* **51** (1983) .

[23] R. P. Feynman, *Space-time approach to Quantum Electrodynamics. Phys. Rev.* **76** (1949) .

[24] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *QCD and resonance physics. theoretical foundations. Nucl. Phys. B* **147** (1979) .

[25] A. J. Buras, *Gauge Theory of Weak Decays: The Standard Model and the Expedition to New Physics Summits.* Cambridge University Press, 2020.

[26] F. L. Wilson, *Fermi's theory of beta decay. Am. J. Phys.* **36** (1968) .

[27] S. Hollands and R. M. Wald, *The operator product expansion in Quantum Field Theory.* in *Encyclopedia of Mathematical Physics (Second Edition).* Academic Press, Oxford, second ed., 2025.

[28] A. Khodjamirian, T. Mannel, and Y. M. Wang, $B \to K\ell^+\ell^-$ *decay at large hadronic recoil. JHEP* **2** (2013) .

[29] J. Lyon and R. Zwicky, *Isospin asymmetries in* $B \to (K^*, \rho)\gamma/\ell^+\ell^-$ *and* $B \to K\ell^+\ell^-$ *in and beyond the Standard Model. Phys. Rev. D* **88** (2013) .

[30] L. Evans and P. Bryant, *LHC machine. JINST* **3** (2008) .

[31] E. Mobs, *The CERN accelerator complex. CERN Document Server* (2016) .

[32] **ATLAS** Collaboration, G. Aad *et al.*, *The ATLAS experiment at the CERN Large Hadron Collider. JINST* **3** (2008) .

[33] **CMS** Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC. JINST* **3** (2008) .

[34] **ALICE** Collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC. JINST* **3** (2008) .

[35] **LHCb** Collaboration, A. Augusto Alves Jr. *et al.*, *The LHCb detector at the LHC. JINST* **3** (2008) .

[36] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1. Comp. Phys. Commun.* **178** (2008) .

[37] **LHCb** Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb VELO (VErtex LOcator).* Technical design report. LHCb. CERN, Geneva, 2001.

[38] **LHCb** Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb Inner Tracker.* Technical design report. LHCb. CERN, Geneva, 2002.

[39] **LHCb** Collaboration, J. J. van Hunen, *The LHCb tracking system. Nuclear Instrum. Meth. A* **572** (2007) .

[40] **LHCb** Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb Outer Tracker.* Technical design report. LHCb. CERN, Geneva, 2001.

[41] **LHCb** Collaboration, R. Arink *et al.*, *Performance of the LHCb Outer Tracker. JINST* **9** (2014) .

[42] **LHCb** Collaboration, S. Amato *et al.*, *LHCb magnet.* Technical design report. LHCb. CERN, Geneva, 2000.

[43] **LHCb** Collaboration, R. AAij *et al.*, *LHCb detector performance. Int. J. Mod. Phys. A* **30** (2015) .

[44] J. V. Jelley, *Cerenkov radiation and its applications. Br. J. Appl. Phys.* **6** (1955) .

[45] **LHCb** Collaboration, S. Amato *et al.*, *LHCb RICH.* Technical design report. LHCb. CERN, Geneva, 2000.

[46] **LHCb** Collaboration, R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2. JINST* **17** (2022) .

[47] **LHCb** Collaboration, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC. Eur. Phys. J. C* **73** (2013) .

[48] **LHCb** Collaboration, S. Amato *et al.*, *LHCb calorimeters.* Technical design report. LHCb. CERN, Geneva, 2000.

[49] **LHCb** Collaboration, E. Picatoste Olloqui, *LHCb Preshower(PS) and Scintillating Pad Detector (SPD). J. Phys. Conf. Ser.* **160** (2009) .

[50] **LHCb** Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb muon system.* Technical design report. LHCb. CERN, Geneva, 2001.

[51] **LHCb** Collaboration, R. Antunes-Nobrega *et al.*, *LHCb trigger system.* Technical design report. LHCb. CERN, Geneva, 2003.

[52] **LHCb** Collaboration, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. JINST* **14** (2019) .

[53] **LHCb** Collaboration, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework. J. Phys. Conf. Ser.* **331** (2011) .

[54] D. J. Lange, *The EvtGen particle decay simulation package. Nucl. Instrum. Meth. A* **462** (2001) .

[55] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED correctionsin Z and W decays. Eur. Phys. J. C* **45** (2006) .

[56] **Geant4** Collaboration, S. Agostinelli *et al.*, *Geant4 — a simulation toolkit. Nucl. Instrum. Meth. A* **506** (2003) .

[57] **LHCb** Collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries in $B \to K^{(*)}\mu^+\mu^-$ decays. J. High Energ. Phys.* **2014** (2014) .

[58] **CDF** Collaboration, T. Aaltonen *et al.*, *Observation of the baryonic flavor-changing neutral current decay $\Lambda_b^0 \to \Lambda\mu^+\mu^-$. Phys. Rev. Lett.* **107** (2011) .

[59] **Belle** Collaboration, S. Choudhury *et al.*, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays. J. High Energ. Phys.* **2021** (2021) .

[60] **HPQCD** Collaboration, W. G. Parrott, C. Bouchard, and C. T. H. Davies, *Standard model predictions for $B \to K\ell^+\ell^-$, $B \to K\ell_1^-\ell_2^+$ and $B \to K\nu\bar{\nu}$ using form factors from $N_f = 2 + 1 + 1$ lattice QCD. Phys. Rev. D* **107** (2023) .

[61] **LHCb** Collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction. J. High Energ. Phys.* **2016** (2016) .

[62] **CMS** Collaboration, V. Khachatryan *et al.*, *Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV. Phys. Lett. B* **753** (2016) .

[63] **LHCb** Collaboration, R. Aaij *et al.*, *Branching fraction measurements of the rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. Phys. Rev. Lett.* **127** (2021) .

[64] **LHCb** Collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays. J. High Energ. Phys.* **2015** (2015) .

[65] **LHCb** Collaboration, R. Aaij *et al.*, *A comprehensive analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay. arXiv:2512.18053* (2025) .

[66] **LHCb** Collaboration, R. Aaij *et al.*, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays. Phys. Rev. Lett.* **131** (2023) .

[67] F. Reiss, *Flavour anomalies in $b \to s\ell^+\ell^-$ and $b \to c\ell\nu_\ell$ transitions. PoS* **LHCP2023** (2024) .

[68] **Belle** Collaboration, K. Abe *et al.*, *Observation of the decay $B \to K\ell^+\ell^-$. Phys. Rev. Lett.* **88** (2001) .

[69] F. Takasaki, *Status of KEKB accelerator and detector, Belle. Int. J. Mod. Phys. A* **15** (2000) .

[70] J. H. Friedman, *Greedy function approximation: A gradient boosting machine. Ann. Statist.* **29** (2001) .

[71] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system.* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Association for Computing Machinery, New York, NY, USA, 2016.

[72] **Particle Data Group** Collaboration, S. Navas *et al.*, *Review of particle physics. Phys. Rev. D* **110** no. 3, (2024) 030001.

[73] D. Casadei, *Estimating the selection efficiency. JINST* **7** (2012) .

[74] M. Pivk and F. Le Diberder, *$_s\mathcal{P}lot$: A statistical tool to unfold data distributions. Nucl. Instr. Meth. A* **555** (2005) .

[75] P. E. Condon and P. L. Cowell, *Channel likelihood: An extension of maximum likelihood for multibody final states. Phys. Rev. D* **9** (1974) .

[76] L. Anderlini *et al.*, *The PIDCalib package.* tech. rep., CERN, Geneva, 2016. `https://cds.cern.ch/record/2202412`.

[77] A. Poluektov, "PIDGen2 Project." `https://pypi.org/project/pidgen2/`.

[78] A. Rogozhnikov, *Reweighting with Boosted Decision Trees. J. Phys. Conf. Ser.* **762** (2016) .

[79] A. Rogozhnikov, "`hep_ml` Project." `https://pypi.org/project/hep-ml/`.

[80] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb.* tech. rep., CERN, Geneva, 2016. `https://cds.cern.ch/record/1701134`.

[81] M. Oreglia, *A Study of the Reactions $\psi' \longrightarrow \gamma\gamma\psi$.* Other thesis, Stanford U., Phys. Dept., 12, 1980.

[82] D. M. Straub, *flavio: a python package for flavour and precision phenomenology in the standard model and beyond.* 2018. `https://arxiv.org/abs/1810.08132`.

A. Invariant mass fit

The following figures illustrate the invariant mass fits performed on the $B^0 \to J/\psi K^0_S$, $B^+ \to J/\psi K^+$, and $B^+ \to J/\psi \pi^+$ simulated samples. These fits are used to determine the signal shape parameters, which are subsequently fixed in the fit to data and are tabulated in Ch. 7.

A.1. $B^0 \to J/\psi K^0_\mathrm{S}$

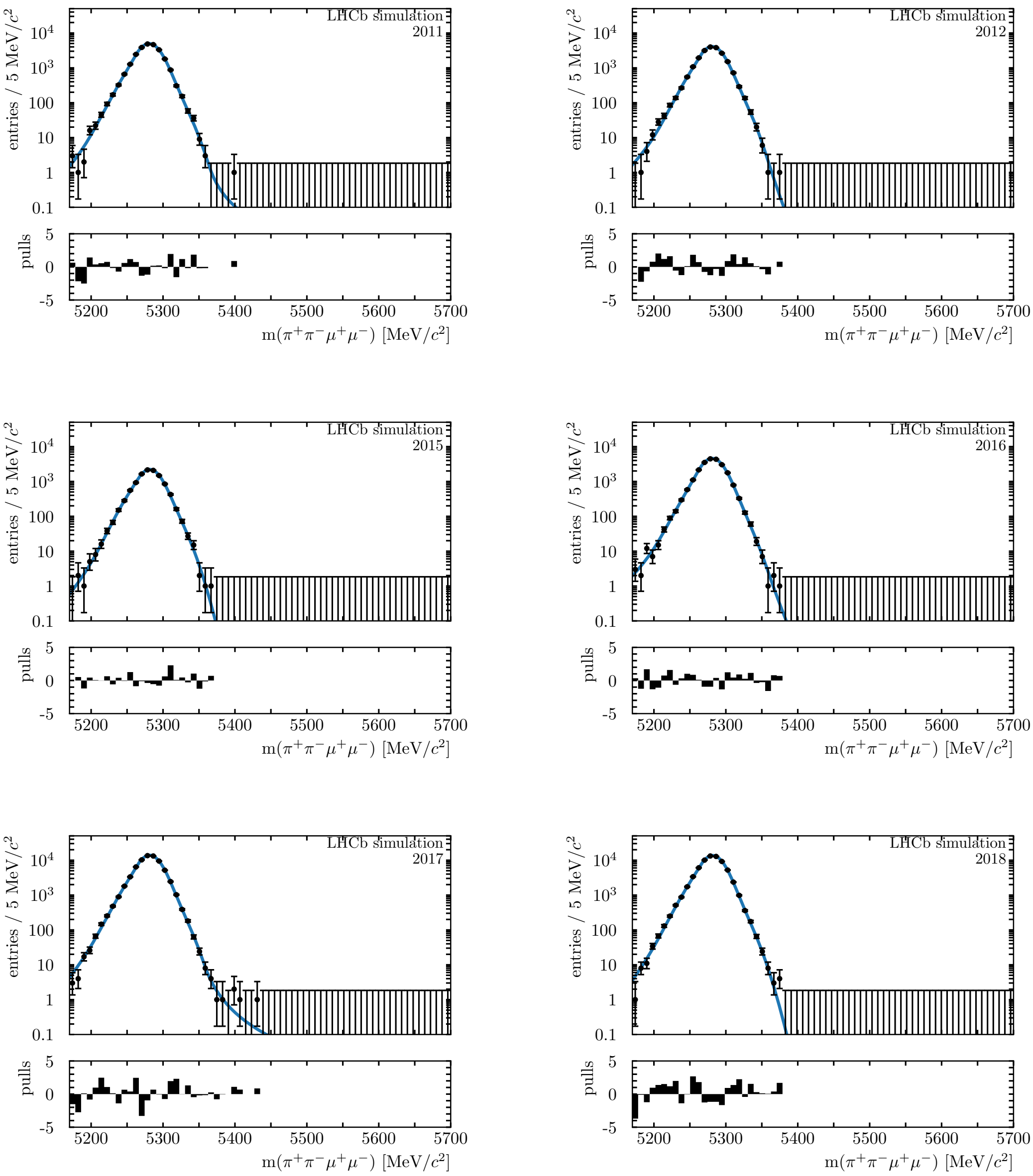


Figure A.1.: Invariant mass fits to $B^0 \to J/\psi K^0_\mathrm{S}$ corrected simulation for the LL reconstruction category. The plots correspond to the 2011–2012, 2015 and 2016–2018 datasets (ordered from top to bottom, left to right). The data points (black) are overlaid with the best fit result (blue line). The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

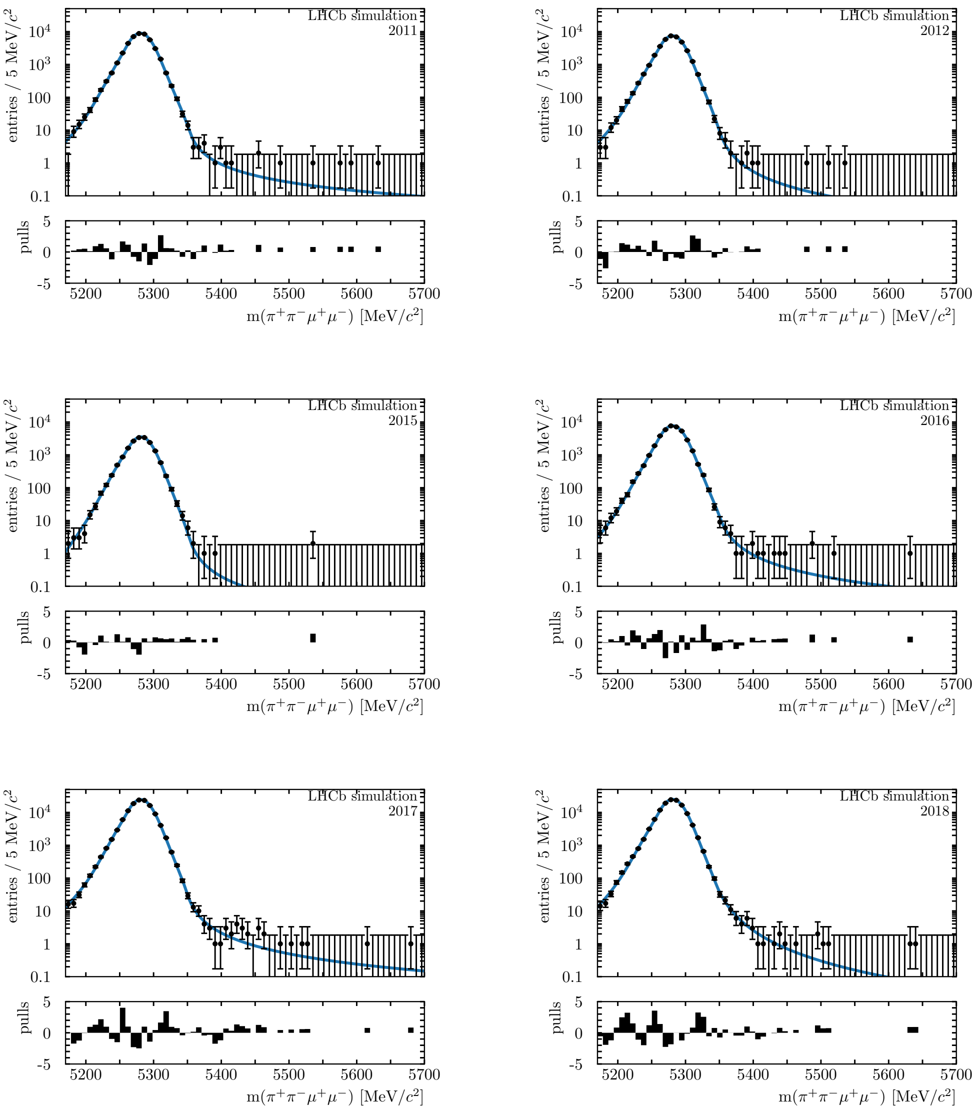


Figure A.2.: Invariant mass fits to $B^0 \to J/\psi K^0_S$ corrected simulation for the DD reconstruction category. The plots correspond to the 2011–2012, 2015 and 2016–2018 datasets (ordered from top to bottom, left to right). The data points (black) are overlaid with the best fit result (blue line). The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

A.2. $B^+ \to J/\psi K^+$

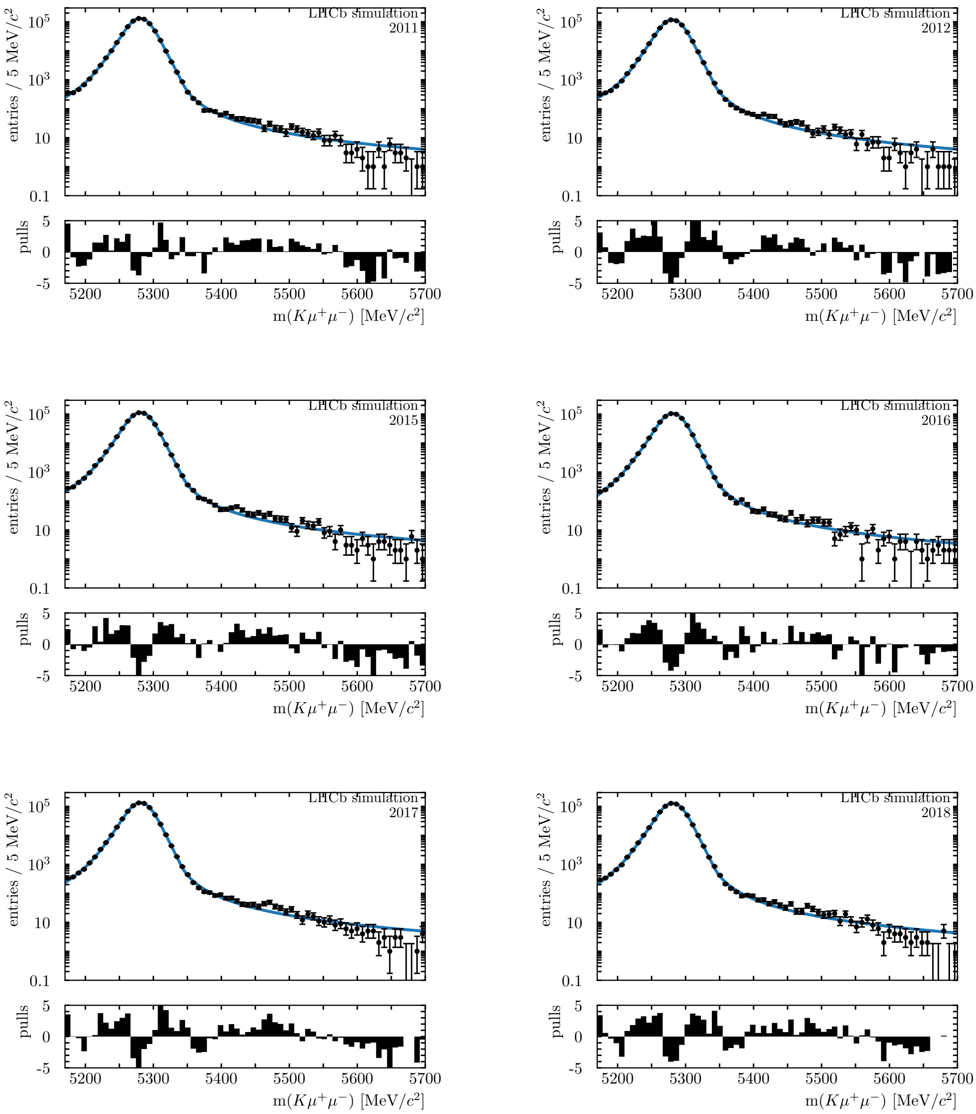


Figure A.3.: Invariant mass fits to $B^+ \to J/\psi K^+$ corrected simulation. The plots correspond to the 2011–2012, 2015 and 2016–2018 datasets (ordered from top to bottom, left to right). The data points (black) are overlaid with the best fit result (blue line). The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.

A.3. $B^+ \to J/\psi\pi^+$

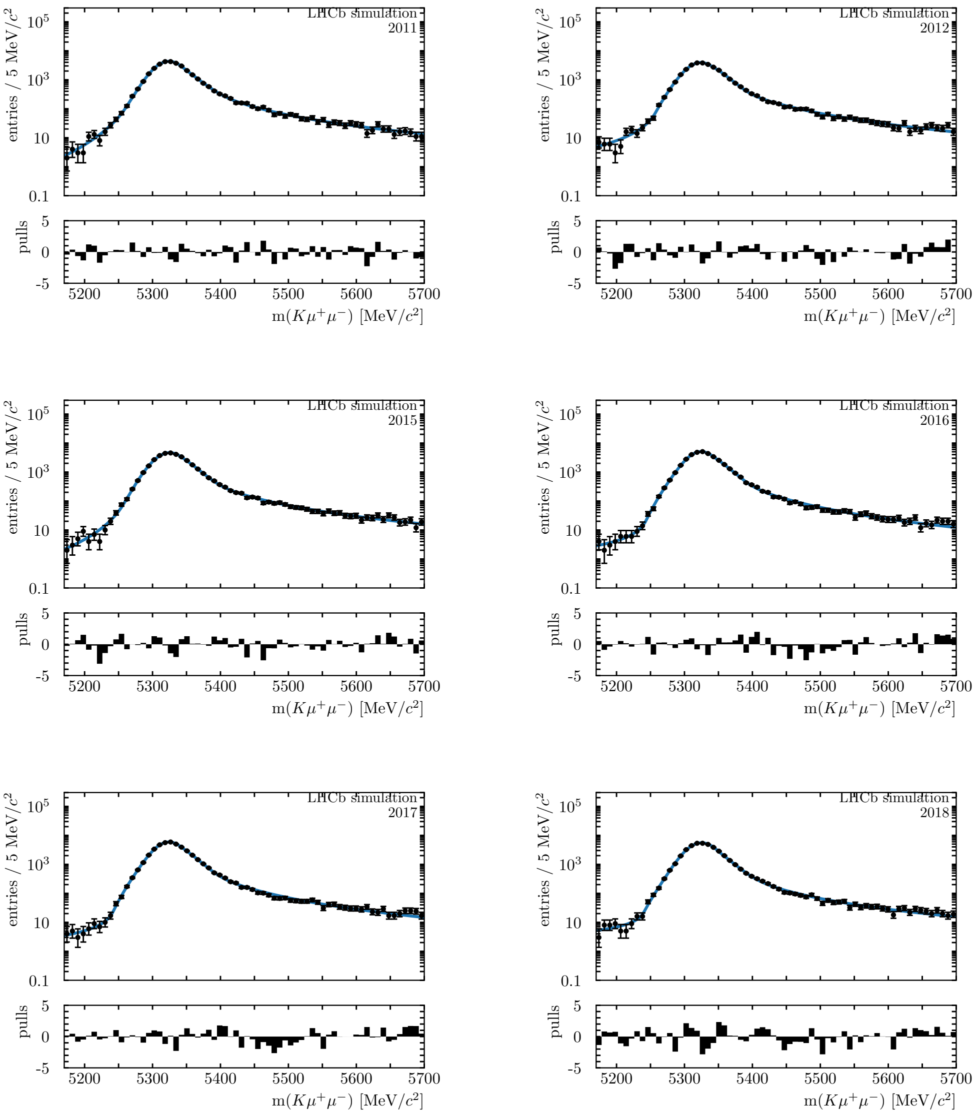


Figure A.4.: Invariant mass fits to $B^+ \to J/\psi\pi^+$ simulation. The plots correspond to the 2011–2012, 2015 and 2016–2018 datasets (ordered from top to bottom, left to right). The data points (black) are overlaid with the best fit result (blue line). The bottom panels display the pulls, defined as the difference between the fitted data and the model normalised by the statistical uncertainty.