Study of the rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation vorgelegt von

Sophie Kretzschmar

aus

Hagen

Berichter: Dr. rer. nat. Christoph Langenbruch
Univ.-Prof. Dr. rer. nat. Stefan Schael

Tag der mündlichen Prüfung: 06. Oktober 2022

Abstract

The Standard Model of particle physics (SM) describes the microcosm of nature with great success. However, observations of phenomena not included in the SM indicate that the SM is not complete and motivate searches for New Physics (NP). An excellent laboratory for NP searches are precision measurements of rare B decays, which gained particular interest with the appearance of the so-called flavour anomalies—a series of deviations between measurements and SM predictions.

One of these flavour anomalies is the branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$, measured with data taken during 2011–2012 with the LHCb detector, which was found to be around 3 standard deviations (σ) lower compared to the SM expectation at small invariant dimuon mass squared, q^2.

This thesis presents the measurement of $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ using the full LHCb data sample collected during 2011–2018. The branching fraction is measured relative to the high-yield $B_s^0 \to J/\psi\phi$ mode and in intervals of q^2, and the data are selected using kinematic and particle identification variables, including a multivariate classifier. The signal decays are extracted from data using an unbinned maximum likelihood fit, and the selection efficiency is evaluated on simulation of the signal decays, to which data-driven corrections are applied to ensure a good description of data. The resulting branching fraction is measured as

$$\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-) = (8.14 \pm 0.21 \pm 0.16 \pm 0.39 \pm 0.03) \times 10^{-7}$$

with (in order) the statistical uncertainty, the systematic uncertainty, the uncertainty from the branching fraction $\mathcal{B}(B_s^0 \to J/\psi\phi)$, and an uncertainty from the extrapolation to the full q^2 range. In the q^2 range $1.1 < q^2 < 6.0\,\text{GeV}^2/c^4$, the branching fraction is found to be $3.6\,\sigma$ below a prediction based on lattice QCD and light-cone sum rule (LCSR) calculations and $1.8\,\sigma$ lower compared to LCSR calculations only.

This thesis further presents the first observation of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decay and a measurement of its branching fraction. The analysis is performed similarly to the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays, however, the higher level of background contamination requires a tighter data selection and a more complex fit model.

The $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decay is observed with a significance of 9σ based on Wilks's theorem, and the branching fraction is measured as

$$\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-) = (1.57 \pm 0.19 \pm 0.06 \pm 0.06 \pm 0.08) \times 10^{-7},$$

with (in order) the statistical uncertainty, the systematic uncertainty, the uncertainty due to the branching fraction of the normalisation mode, and the q^2-range extrapolation. This result is in agreement with the SM predictions.

Contents

Chapter 1

Introduction

"It's not a discovery, but something is going on,"—this quote applies quite well to the circumstances under which this thesis is written.[1] The situation is the following. A single theory, the *Standard Model of particle physics (SM)*, has proven to explain the microcosm of nature with great success, explaining a large range of phenomena: from reactions that keep the sun burning to the existence of polar lights to the harm that is done by radioactivity—all of these phenomena are understood as results of fundamental particles interacting with each other, collectively described by the SM.

Despite its success, there exist phenomena that cannot be explained in the SM, *e.g.* the observations of (1) *dark matter* and of (2) the *matter-antimatter asymmetry* in the universe. The former, dark matter, is a type of matter that is affected only by gravity, potentially also the weak force, but nothing else; it was first observed via unexpected rotation curves of planets, and its presence has been later confirmed by measurements with gravitational lenses and with the cosmic microwave background. However, the SM provides no dark-matter particle candidate. The other issue, the matter-antimatter asymmetry, is based on the everyday observation that matter seems to dominate the universe. Whereas the SM allows for minute differences between matter and anti-matter, these differences are too small to explain the different abundances in the universe.

Such shortcomings of the SM triggered many searches for new fundamental particles and forces not yet included, often referred to as *New Physics (NP)*. Many of these were conducted at the Large Hadron Collider (LHC) where hopes were high to find new particles produced in the proton-proton collisions; so far, no significant signs of directly produced new heavy particles were found, confirming the SM to be a robust theory across all the performed measurements.

Searches for NP particles produced directly in the LHC proton-proton collisions are referred to as *direct searches*. Due to energy conservation, the mass of the NP particles probed via direct searches is limited by the available proton-proton collision energy. As such, direct searches only reach NP mass scales of a few TeV. Higher NP mass-scales can be probed with *indirect searches*, where certain observables are measured precisely to compare them with the SM prediction. These indirect searches exploit Heisenberg's uncertainty principle, which allows for NP particles with mass far heavier than available collision energies to be produced,

[1] Quote from D. Straub, taken from CERN Courier, *Flavour anomalies continue to intrigue*, 7 May 2019.

as long as they only exist for a very short time, *i.e.* they manifest via quantum fluctuations. If NP particles existed, they could affect observables as quantum fluctuations and appear as deviations between measured and predicted observables.

An excellent laboratory for indirect searches are *rare decays of b hadrons*, which proceed via the transition of a b quark to an s quark. These decays are referred to as *flavour-changing neutral current (FCNC)* processes as the *flavour* of the quark is altered but not its electromagnetic charge, and the processes are *rare* because they are suppressed in the SM as they are forbidden at lowest perturbation order. This feature makes them particularly interesting for NP searches since NP particles could contribute significantly, leading to large alterations in observables.

The proton-proton collisions at the LHC result in copious production of b hadrons. One of the experiments at the LHC is the LHCb experiment, which is specifically designed to measure b-hadron decays and therefore best-suited for the above mentioned precision measurements. The LHCb detector enabled the LHCb collaboration to collect the largest sample of b hadrons to date resulting in world-leading precision across a range of B-decay observables.

Measurements of semileptonic b-hadron decays with $b \to s\ell^+\ell^-$ transition are exactly the area where *something is going on*: a set of deviations from the SM predictions, referred to as the *flavour anomalies*. The flavour anomalies appear in various measurements of *branching fractions*, *i.e.* relative decay rates; angular decay distributions; and ratios of decay rates. The measurements differ individually from the SM prediction at the level of 2–3.5 standard deviations (σ). These individual tensions are not significant enough to claim a discovery, however, the similarity between the trends indicate an intriguingly coherent pattern, which generated large interest in the particle-physics community.

From the experimental point of view, the precision of the flavour anomalies is limited by the data sample sizes; so, analysing more data will help clarify the situation. As some of the observables are measured with only a fraction of the currently recorded LHCb data, effort is ongoing to perform the measurements using the complete LHCb data sample taken to-date, namely during the years 2011–2018. This thesis presents one of these efforts, namely the repeated measurement of the branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$, which is documented in detail in an internal analysis note [1].

The $B_s^0 \to \phi\mu^+\mu^-$ branching fraction has been previously measured with data collected during 2011–2012 by the LHCb experiment [2]. The result gained particular attention as it exhibits a local tension with the SM prediction with over $3\,\sigma$ significance, making it one of the most significant flavour anomalies. Until 2018, the LHCb experiment has collected more data so that the data sample to analyse $B_s^0 \to \phi\mu^+\mu^-$ decays is increased roughly by a factor of four— an updated measurement, significantly improving the precision of the result, was therefore eagerly awaited.

Interpreting the flavour anomalies is rather complex. To explain them, different NP models are currently being discussed; however, the possibility that some of the anomalies stem from underestimated hadronic effects, related to the b-quark bound states, has not been ruled out yet. Collecting as much complementary information as possible will help in constraining the interpretations; for this reason, not just more precise measurements are needed but also more new measurements. One of these is also presented in this thesis, namely the measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$

decays. The decay process had not been observed until this work was carried out, and it constitutes the first measurement of a semileptonic FCNC decay with a spin-2 hadron in the final state.

This thesis aims to give the complete picture on both analyses, *i.e.* the measurement of $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$, and it is organised as follows: After this introduction, Chapter 2 explains the theoretical and experimental context, comprising an overview on the SM, where the focus is on the aspects relevant for rare decays of b hadrons; and a discussion of the specific flavour anomaly measurements. The following chapter, Chapter 3, introduces the LHCb experiment, which recorded the data used in this thesis. There, the subdetectors will be explained as well the LHCb software. The following chapters present the specific analyses: Chapter 4 explains the data selection and treatment of simulated samples, and Chapters 5 and 6 detail the analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, respectively. Finally, a summary and an outlook are given in Chapter 7.

Chapter 2

Theoretical and experimental context

The measurements of $b \to s\ell^+\ell^-$ decays, including the measurements presented in this thesis, constitute tests of the Standard Model of particle physics (SM). Section 2.1 gives a qualitative overview on its ingredients, the fundamental particles and forces, and describes briefly the mathematical framework. The description is mainly focused on *quark flavour physics*, which is the relevant sector of the SM for the measurements in this thesis.

Calculating predictions for rare B-meson decay processes is complicated by the different SM-physics types involved—*e.g.* describing bound particles like the B_s^0 or ϕ meson relies on different calculation methods than those needed to describe the $b \to s\ell^+\ell^-$ decay. These different physics scales are disentangled by using *effective field theories (EFT)*, explained generally in Sect. 2.2.1 and specifically for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays in Sect. 2.2.2.

The second part of this chapter concerns the current experimental situation of $b \to s\ell^+\ell^-$ measurements. The focus is set on the *flavour anomalies*, a series of deviations from the SM prediction discussed in Sect. 2.3. These individual deviations are limited in precision, and each measurement on its own is not yet significantly deviating, so is interesting to combine experimental information from individual measurements. These efforts are pursued by the theoretical community in *global fits* to the SM, described in Sect. 2.4. Interestingly, these combinations indicate a strikingly coherent pattern that can be coherently explained in several NP models.

2.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) [3, 4] describes the fundamental particles and their interactions, as summarised in Fig. 2.1. There are different kinds of these fundamental particles: twelve matter particles, the *fermions*, four interaction-mediator particles, the *gauge bosons*, and a spin-0 boson, the *Higgs boson*. For each of the fermions, there is an antiparticle with opposite *charges*, which are explained in the following, but the same mass. The gauge bosons are associated to the fundamental interactions—the strong force, the weak force, and the electromagnetic force—and the Higgs boson is associated to the *Higgs mechanism*, which gives the fundamental particles their mass.

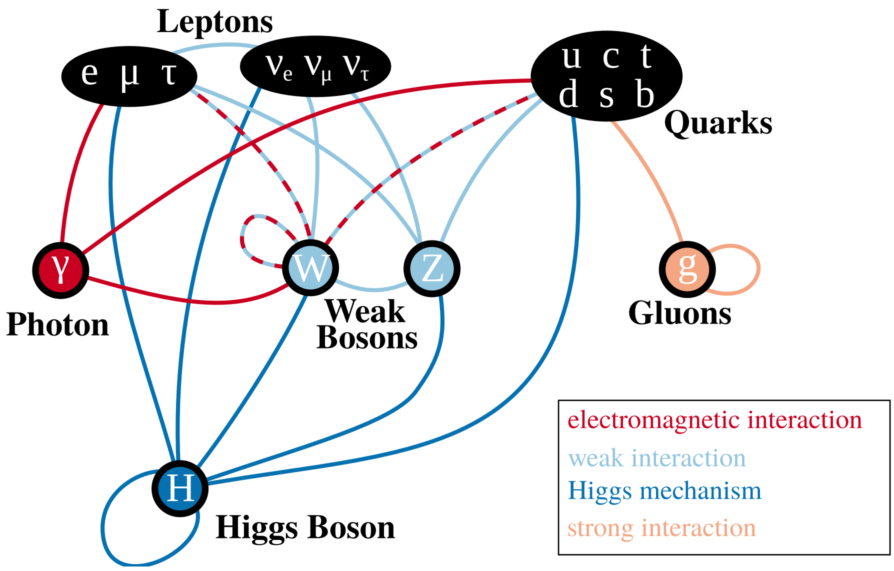

Figure 2.1: Overview on the fundamental particles of the Standard Model of particle physics. The fermions (leptons and quarks), vector bosons (the photon, the weak bosons, and the gluons), and the Higgs boson are shown. The lines connect the particles that directly interact with each other, and the colour indicates the interaction type. The figure is based on the graphic from Ref. [5].

The fermions, carrying spin 1/2, comprise the *leptons* and *quarks*. The leptons can be further divided into the *charged leptons*, which are the electron e, the muon μ, and the tau lepton τ; and the uncharged leptons, referred to as *neutrinos*, which are the electron neutrino ν_e, the muon neutrino ν_μ, and the tau neutrino ν_τ. The six quarks can be categorised into the *up-type* quarks, which are the up quark u, the charm quark c, and the top quark t; and the *down-type* quarks, which are the down quark d, the strange quark s, and the bottom quark b.

This description already suggests a particular structure of the SM, in which the fermions can be grouped into three *generations*, namely

$$\begin{array}{c} \text{up-type quarks:} \\ \text{down-type quarks:} \\ \text{charged leptons:} \\ \text{neutrinos:} \end{array} \qquad \underbrace{\left|\begin{array}{c} u \\ d \\ e \\ \nu_e \end{array}\right|}_{\text{1st gen.}} \qquad \underbrace{\left|\begin{array}{c} c \\ s \\ \mu \\ \nu_\mu \end{array}\right|}_{\text{2nd gen.}} \qquad \underbrace{\left|\begin{array}{c} t \\ b \\ \tau \\ \nu_\tau \end{array}\right|}_{\text{3rd gen.}} ,$$

with each generation containing one particle of each *type*, *i.e.* an *up-type* quark, a *down-type* quark, a charged lepton, and a neutral lepton. Particles of the same type from different generations, *e.g.* the electron and the muon, differ only by their mass—other than that, they are copies of each other with exactly the same quantum numbers and properties. The reason for this three-generation structure of the SM is unknown to this date.

The five bosons complete the particle content of the SM. Four of them are the spin-1 carrying force mediators, and the last is the spin-0 carrying Higgs boson. Each of the four force mediators is associated to an interaction: the photon γ to the electromagnetic interaction, the gluons g to the strong force, and the weak $W^\pm$ and Z^0 bosons to the weak force. The *gauge structure* of the SM, detailed in the next section, determines which interactions are possible among the force carriers and matter particles.

The photon γ couples to particles that carry an *electromagnetic* charge q, which are the charged leptons ($q = -1$), the up-type quarks ($q = +2/3$), the down-type quarks ($q = -1/3$), and the charged weak bosons $W^{\pm}$ ($q = \pm 1$).Examples for electromagnetic interactions are Compton scattering, where a photon looses energy when it interacts with an electron, and ionisation processes, where charged particles interact with electrons bound in atoms.

The gluons g couple to particles that carry *colour charge*, which are only the quarks and the gluons themselves. The quarks carry one of the three colour charges ("red", "green", "blue"), and the gluons carry a colour-anticolour combination, *e.g.* blue-antired. The distinct feature of the strong interaction is its varying coupling strength: it is small in interaction processes with large momentum transfer, and it is large in interaction processes with small momentum transfer, referred to as *asymptotic freedom* and *confinement*, respectively.

The former phenomenon, asymptotic freedom, can be observed at high-energy experiments like the Large Hadron Collider (LHC), where the hard scattering of partons in proton collisions are described well as collisions of quarks while neglecting that they are bound to protons. The latter phenomenon, confinement, explains why single quarks are not found in nature: for these quarks, the strong coupling becomes so large that it is energetically favoured to create new quark-antiquark pairs from the vacuum to form colourless bound states. The notion of "colour" relates directly to the possible colourless quark-bound states: the *baryons*, formed by three quarks ("red + green + blue = colourless"); the *mesons*, formed by a quark and an antiquark ("colour + anticolour = colourless"), or mixtures of these configurations, *e.g.* the pentaquark states, formed by three quarks and a quark-antiquark pair. Examples for mesons are the B^0_s meson, which consists of an $\bar{b}$ and an s quark.

Finally, the weak $W^{\pm}$ and Z^0 bosons couple to all particles except gluons. The corresponding weak force is the only interaction that connects particles from different flavours and generations, allowing processes like the neutron decay, in which a d quark decays into a u quark, a lepton and a neutrino; or the muon decay, in which the muon decays into an electron, an antielectron neutrino and a muon-neutrino. Similarly, the weak force mediates the decays studied in this thesis, the $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays, where the b quark from the B^0_s meson decays to an s-quark and a dimuon pair. These *flavour-changing neutral currents (FCNC)*, which are only allowed via higher-order interaction processes, are discussed in Sect. 2.2.

The last ingredient of the Standard Model is the Higgs mechanism. The Higgs mechanism is needed to include the observed masses of the massive gauge bosons and the fermions in the SM, and it introduces the flavour-changing interactions as explained in more detail in Sect. 2.1.2. Its associated gauge boson, the Higgs boson, has been discovered in 2012 at the Large Hadron Collider at CERN [6, 7], and this discovery concluded the search for the fundamental particles of the SM.

2.1.1 Mathematical framework

The SM is formulated as *relativistic quantum field theory*, which considers the theory of special relativity, needed to convert between mass and energy, *i.e.* particles being destroyed or created, and quantum mechanics, needed to describe the subatomic scales of the fundamental particles and their interactions [4, 8]. The SM is encoded in the Lagrangian $\mathscr{L}_{\text{SM}}$, which is the most general renormalizable Lagrangian that fulfills three requirements [9]: it is (1) SM-gauge symmetric, *i.e.* it stays invariant under SM-gauge transformations; it contains (2) the SM-fermions and -scalars, which are defined by their behaviour under the SM-gauge transformations; and it incorporates a (3) mechanism to spontaneously break the SM gauge-symmetry.

The SM-gauge transformations are defined by the gauge-group $\mathscr{G}_{SM} = SU(3)_c \times SU(2)_L \times U(1)_{\text{Y}}$; $SU(3)_c$ is being referred to as the *colour group*, connected to the strong-interactions, and $SU(2)_L \times U(1)_{\text{Y}}$ is being referred to as the *electroweak group*, connected to the electroweak interactions.

The SM-fermions and -scalars are represented by six field types: the (1) *left-handed quark fields* Q_{L_i}, the (2) *right-handed up-* and (3) *down-type quark fields* U_{R_i} and D_{R_i}, respectively; the (4) *left-* and (5) *right-handed lepton fields* L_{L_i} and E_{R_i}, respectively; and (6) the scalar field ϕ (6). Each of the fermion fields comes in three generations i, similar to what was explained before. These six field-types are defined by how they transform under the $\mathscr{G}_{SM}$ transformation.

The $\mathscr{G}_{SM}$ transformations are defined by the $\mathscr{G}_{SM}$ generators, which are the eight matrices containing the Gell-Mann matrices for the colour-group $SU(3)_C$, the three matrices t^a containing the Pauli-matrices for $SU(2)_L$, and a complex number of $U(1)_Y$. These generators are associated to *gauge boson fields*: the eight generators of the $SU(3)_C$ to eight gluon fields; the three generators of $SU(2)_L$ to three W^i boson fields; and the generator of $U(1)_Y$ to the B boson field. The boson fields mediate the corresponding forces, *i.e.* the gluons mediate the strong interaction, and the weak B and W^i bosons the electroweak interaction.[1]

The particle fields' behaviour under group transformations manifests the interactions between particles in the SM: the quark fields Q_{L_i}, U_{R_i} and D_{R_i} transform as triplets under the colour group $SU(3)_c$, so they couple to the gluons, *i.e.* they interact via strong interaction; the triplet structure corresponds to the three colour charges of the quarks. The other particle fields are invariant under $SU(3)_c$ transformations, so they do not couple to gluons.

The left-handed fields, Q_{L_i} and L_{L_i}, and the scalar field ϕ transform as doublets under $SU(2)_L$, so they couple to the weak bosons W^i; this doublet-structure indicates that the fields come with two charges, referred to as *weak isospin* T, distinguishing the left-handed *up-type* and the *down-type* fields. Correspondingly, the left-handed fields can be written with their up- and down-type component, *e.g.* for the quarks, $Q_{L_i} = (U_{L_i}, D_{L_i})^T$. The right-handed fields are invariant under $SU(2)_L$, so they do not couple to the weak bosons.

Lastly, all particle fields transform as singlets under $U(1)_Y$, so all particles couple to the B boson. The corresponding charge is the *weak hypercharge* Y.

The SM Lagrangian $\mathscr{L}_{SM}$ is defined by these requirements, *i.e.* the (1) invariance under $\mathscr{G}_{SM}$ and the (2) specific fields and their transformation behaviour [9].

---

[1]The B and W^i fields are related to the previously discussed photon field A and the weak Z^0 and $W^{\pm}$ boson fields via the Weinberg angle θ_W.

Furthermore, $\mathscr{L}_{SM}$ is required to be relativistically invariant and to be *renormalisable*, which means that divergences, appearing when using the Lagrangian to predict measurement observables, can be removed by redefining parameters, *e.g.* particle masses. The SM-Lagrangian $\mathscr{L}_{\mathrm{SM}}$ [4, 10] is given by

$$\mathscr{L}_{\mathrm{SM}} = \mathscr{L}_{\mathrm{gauge}} + \mathscr{L}_{\mathrm{fermions}} + \mathscr{L}_{H} + \mathscr{L}_{\mathrm{Yukawa}}, \tag{2.1}$$

with $\mathscr{L}_{\mathrm{gauge}}$ containing all terms with the gauge-boson fields only and $\mathscr{L}_{\mathrm{fermion}}$ containing the fermion fields, *i.e.*

$$\begin{aligned}\mathscr{L}_{\mathrm{fermion}} = i\sum_{i=1}^{3} &\overline{Q}_{L_i} \not{D}^q Q_{L_i} + \overline{U}_{R_i} \not{D}^q U_{R_i} + \overline{D}_{R_i} \not{D}^q D_{R_i} \\ &+ \overline{L}_{L_i} \not{D}^\ell L_{L_i} + \overline{E}_{R_i} \not{D}^\ell E_{R_i},\end{aligned}$$

with the covariant derivative $\not{D}^\ell = \gamma^\mu D_\mu$ defined for left-handed leptons ψ_{L_i} as

$$\not{D}^\ell \psi_{L_i} = \gamma^\mu \left[\partial_\mu + igt^a W^a_\mu + ig'\frac{1}{2} Y B_\mu\right] \psi_{L_i}, \tag{2.2}$$

and for right-handed leptons ψ_{R_i} as

$$\not{D}^\ell \psi_{R_i} = \gamma^\mu \left[\partial_\mu + ig'\frac{1}{2} Y B_\mu\right] \psi_{R_i} \tag{2.3}$$

with the gauge coupling parameters g and g'. For the quarks, the derivatives $\not{D}^q$ are identical besides the additional interaction terms with the gluon fields.

The terms containing the scalar field ϕ are collected in $\mathscr{L}_{\mathrm{H}}$, given by

$$\mathscr{L}_{\mathrm{H}} = D_\mu \phi^\dagger D^\mu \phi + \underbrace{\mu^2 \phi^\dagger \phi - 1/2\lambda(\phi^\dagger\phi)^2}_{-V(\phi^\dagger\phi)}, \tag{2.4}$$

where the derivative D_μ is the same as in Eq. 2.2, and $V(\phi^\dagger\phi)$ is referred to as *Higgs* potential.

Finally, the interaction terms between the fermions and the scalar ϕ are collected in $\mathscr{L}_{\mathrm{Yukawa}}$, given for the quarks by

$$\mathscr{L}^{\mathrm{quarks}}_{\mathrm{Yukawa}} = -\overline{Q}_{L_i} \Gamma^d_{ij} D_{R_j} \phi - \overline{Q}_{L_i} \Gamma^u_{ij} U_{R_j} \widetilde{\phi}, \tag{2.5}$$

with the *Yukawa matrices* Γ^d and Γ^u, and $\widetilde{\phi} = i\sigma_2\phi^*$. The Yukawa part for leptons $\mathscr{L}^{\mathrm{leptons}}_{\mathrm{Yukawa}}$ is constructed similarly.

The missing ingredient to the SM is the mechanism of spontaneous symmetry breaking, often referred to as Higgs mechanism [11, 12]. This mechanism is needed to fix the theory formulated so far, which only contains massless fields; mass terms like $-m^2\overline{\psi}_{L_i}\psi_{R_i}$ would violate the gauge invariance; however, having only massless fields clearly contradicts the observations.

The Higgs-mechanism solves this issue. It spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry by choosing a specific parameter for the Higgs potential $V(\phi^\dagger\phi)$, *i.e.* $\mu^2 < 0$ [8]. This choice sets the $V(\phi^\dagger\phi)$-minimising ground state ϕ_{gr} to a non-zero value; it can be chosen as $\phi_{\mathrm{gr}} \propto (0, v)^T$ with the real parameter v. With this ground state ϕ_{gr}, the SM Lagrangian is not invariant under $SU(2)_L \times U(1)_Y$ anymore. When requiring the photon to be massless, the residual symmetry group is $U(1)_{\mathrm{EM}}$, for which the corresponding boson is the photon γ.

Using $\phi = \phi_{\mathrm{gr}} + h$ in the Yukawa part of the SM Lagrangian introduces the missing mass terms for the massive gauge bosons and for the fermions, *e.g.* for down-type quarks the terms read

$$-\overline{Q}_{L_i}\Gamma^d_{ij}D_{R_j}\phi = -(\overline{U}_{L_i}, \overline{D}_{L_i})\Gamma^d_{ij}D_{R_j}\phi \xrightarrow{\phi \to \phi_{\mathrm{gr}}=(0,v)^T} -\overline{D}_{L_i}\Gamma^d_{ij}D_{R_j}v. \tag{2.6}$$

When referring to quarks, however, one usually means the particles identified by their specific mass, *e.g.* the top quark is the particle with the top-quark mass m_t. But the quark fields in Eq. 2.6 are different objects since the Yukawa matrices Γ^d_{ij} are not necessarily diagonal, so mass-terms like $-m_t^2\overline{t}_{L_i}t_{R_i}$ for top-quark fields t are still missing. The fact that the *weak-eigenstate* fields, *i.e.* the Q_{L_i} fields, differ from the *mass-eigenstate* fields, *e.g.* the top quark field t_{L_i}, is the defining feature for quark flavour physics [9, 10].

2.1.2 Quark flavour physics

To describe the SM in terms of *mass-eigenstate fields*, *i.e.* the objects one refers to normally with the term *quark fields*, one needs to express the quark fields in the *mass basis*, which can be done via the transformations

$$\Gamma^d = V_l\Gamma^d_{\mathrm{diag}}V_r^\dagger \qquad \Gamma^u = T_l^\dagger\Gamma^u_{\mathrm{diag}}T_r \tag{2.7}$$
$$D_L = V_l d_L \qquad U_L = T_l u_L \tag{2.8}$$
$$D_R = V_l d_R \qquad U_R = T_r u_R \tag{2.9}$$

with the unitary basis-transformation matrices $V_{l,r}$ and $T_{l,r}$. Using the basis-transformation in Eq. 2.6 yields the required mass terms, *i.e.*

$$-\overline{D}_{L_i}\Gamma^d_{ij}D_{R_j}v = -\underbrace{v\Gamma^d_{diag}}_{\sim m_d^2}\overline{d}_L d_R. \tag{2.10}$$

In addition to the mass terms, interactions with $W^\pm$ bosons appear from the quark part of the Lagrangian $\mathcal{L}^{\mathrm{quark}}_{\mathrm{SM}}$, *i.e.*

$$i\sum_i \overline{Q}_{L_i}\not{D}Q_{L_i} = -\overline{Q}_{L_i}gt^a\gamma^\mu W^a_\mu Q_{L_i} + ... \tag{2.11}$$
$$\to -\overline{u}_L\underbrace{\left(T_l^\dagger V_l\right)}_{V_{\mathrm{CKM}}}d_L g\gamma^\mu W^-_\mu - \overline{d}_L\underbrace{\left(V_l^\dagger T_l\right)}_{V^\dagger_{\mathrm{CKM}}}d_L g\gamma^\mu W^+_\mu + ... \tag{2.12}$$

with the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM} [13, 14].

The CKM matrix is the key feature of quark flavour physics as its non-diagonalness is the only mechanism allowing differently flavoured quarks to interact. It is a unitary matrix with four parameters—three real parameters and one mixing phase. One parametrisation is the *Wolfenstein parametrisation* [15], given by

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

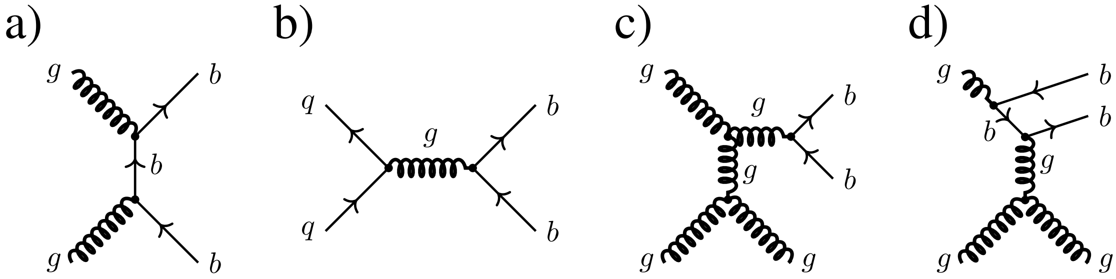

Figure 2.2: Production mechanisms to produce b-quark pairs from quark or gluon interactions. Tree-level (a-b) and higher-order processes (c-d) are shown.

with the expansion parameter λ and parameters A, ρ, and η. This parametrisation is motivated by the fact that the CKM matrix is measured to be almost diagonal with only small corrections.

The $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays studied in this thesis involve a flavour-mixing interaction, *i.e.* a transition from a b quark to an s quark. As discussed above, flavour-changing processes can only be mediated by $W^\pm$ bosons. However, W bosons couple only to combinations of up- and down-type quarks; direct interactions between two down-type quarks are not included in the SM. While these *flavour-changing neutral-current (FCNC)* processes are forbidden at leading-order *tree-level* processes in the SM, they are allowed via suppressed higher-order *loop-level* processes, discussed in the following.

2.2 Production and FCNC decays of *b* hadrons

Hadrons are strongly bound states of quarks and antiquarks, and they are either *mesons*, formed by a quark-antiquark pair, or *baryons*, formed by three quarks. Hadrons with at least one b quark are referred to as *b hadrons*, and their specific type is defined by their companion quarks; examples are B_s^0 mesons, consisting of a b antiquark and an s quark, and B^+ mesons, consisting of a b antiquark and u quark, and Λ_b^0 baryons with b quark and a ud-quark pair.

Hadrons with b quarks are produced copiously at the Large Hadron Collider (LHC) where the asymptotically-free partons of the proton, *i.e.* quarks and gluons, interact and create b-quark pairs in the hard collision processes as shown in Fig. 2.2 [16]. The b quark pairs are produced mostly with flight direction parallel to the proton-beam pipes, which motivates the forward-geometry of the LHCb detector described in Sect. 3. In the area covered by the LHCb detector, b quark pairs are produced with cross-sections $\sigma_{pp\to bbX} = 72.0 \pm 0.3 \pm 6.8$ µb and $\sigma_{pp\to bbX} = 144 \pm 1 \pm 2$ µb at proton-collision centre-of-mass energies of 7 TeV and 13 TeV, respectively [17].

As explained before, the colour-carrying quarks directly hadronise into colourless bound states. Which *specific* bound state is formed is described by the *hadronisation fractions*, quantifying the probability f_X to form a b hadron with b and X quark, *e.g.* $f_s = 0.103(5)$ to form a B_s^0 meson or $f_u = 0.404(6)$ to form a B^+ meson [18].

The ground states of b-flavoured hadrons decay via the weak interaction. The decay processes studied in this thesis, *i.e.* $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$, include a $b \to s\ell^+\ell^-$ transition: the B_s^0 meson, a bound state of a b and an $\overline{s}$ quark,

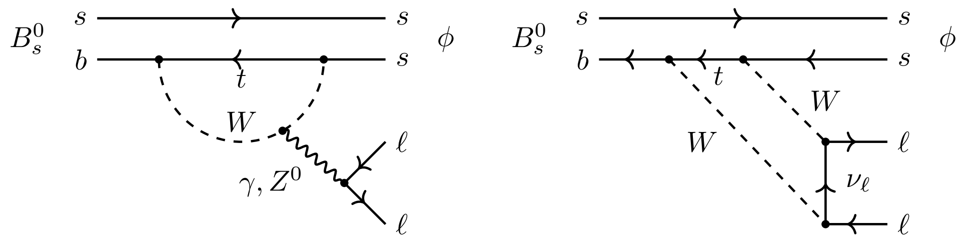

Figure 2.3: Feynman diagrams of $B_s^0 \to \phi\mu^+\mu^-$ decays via (left) electroweak penguin and (right) box diagram.

decays to a ϕ or $f_2'(1525)$ meson, which are $s\bar{s}$-quark bound states, and a muon pair. As explained before, the $b \to s\ell^+\ell^-$ transitions are forbidden at tree-level in the SM; they only proceed via FCNC processes, for which examples are shown in Fig. 2.3 for $B_s^0 \to \phi\mu^+\mu^-$ decays.

The FCNC decays are suppressed in the SM, and their branching fractions are typically at the order of $10^{-7} \sim 10^{-6}$. This suppression, however, makes them particularly interesting because not-yet-observed heavy particles could contribute to these processes and significantly change observables such as decay rates. Performing precision measurements of rare FCNC decays thus constitutes the idea of *indirect searches* for new physics (NP). It has been estimated that NP particles with masses up to $\mathcal{O}(100\,\text{TeV})$ could be in reach with flavour observables in the b sector [19]—this is especially interesting in the current situation where extensive *direct searches*, *i.e.* attempts to produce NP particles directly at particle colliders or detect their collisions with known matter particles, delivered only null-results and excluded NP with masses $\lesssim 10\,\text{TeV}$ (*e.g.* Refs. [20, 21]).

Indirect searches rely not only on precise measurements but also precise predictions of the SM expectation. Predicting observables for processes like $B_s^0 \to \phi\mu^+\mu^-$ decays, however, is not trivial as they involve different interaction-types, which are relevant at different energy-scales and which require different calculation techniques: the b-quark is bound in a B_s^0 meson where colour interactions at energies of order $\Lambda_{QCD} < \mathcal{O}(m_b) \approx 4\,\text{GeV}$ are relevant. At these energies, the quarks are confined and the interactions are governed by complex hadronic interactions. The $b \to s\ell^+\ell^-$ transition, however, includes a W-boson interaction, and at these energies, $m_W \approx 80\,\text{GeV}$, the quarks are asymptotically free and contributions from QCD can be calculated in perturbation theory. Potential NP particles are expected at energies larger $\Lambda_{NP} \gtrsim$ TeV, and their interaction types are unknown. To disentangle these different energy scales and interactions from each other, effective field theories (EFT) have proven to be a useful tool [9, 22–24].

2.2.1 Effective field theories

Decay rates Γ of particle physics processes are calculated via

$$\Gamma \sim \int |\mathcal{M}|^2 \text{d(phase-space)}, \tag{2.14}$$

with the matrix element $\mathcal{M}$ encoding the physics interactions, and the integral takes into account all possible momentum configurations of the process, referred to

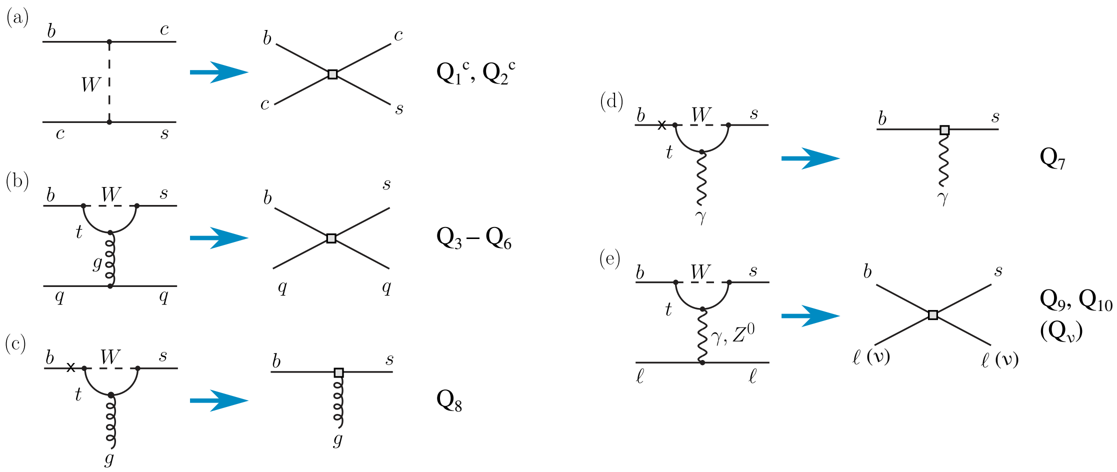

Figure 2.4: Examples for b-quark decays in the full SM and in the EFT. The SM diagrams are taken taken from Ref. [22].

as *phase-space* [9]. For a decay process $i \to f$ the matrix element would be of form

$$\mathcal{M} = \langle f|\mathcal{H}|i\rangle, \tag{2.15}$$

with the Hamiltonian $\mathcal{H}$ that is defined by the Lagrangian of the theory, *e.g.* the SM Lagrangian $\mathcal{L}_{\text{SM}}$. In effective field theories (EFT), the Hamiltonian is written as *effective* Hamiltonian $\mathcal{H}_{\text{eff}}$, defined as

$$\mathcal{H}_{\text{eff}} \sim \sum_i f^i_{\text{CKM}} C_i(\mu) Q_i(\mu), \tag{2.16}$$

with a factor containing the CKM-matrix elements f^i_{CKM}, the *Wilson coefficients* $C_i(\mu)$ and the *local operators* $Q_i(\mu)$ [9, 22, 23]. This formulation separates the interactions into (1) the interactions relevant *below* the energy scale μ, represented by the local operators $Q_i(\mu)$, and (2) the interactions relevant at energy scales *above* μ, represented by the complex Wilson coefficients $C_i(\mu)$.

For b physics, one chooses typically $\mu = m_b$, so the local operators Q_i contain the light quark fields, the gluons, and the photon; correspondingly, the Wilson coefficients C_i contain the effects from the massive gauge bosons and the top quark. Figure 2.4 shows examples for decay processes in the full SM compared to the corresponding local operators in the EFT.

To calculate the Wilson coefficients, the first step is to calculate the matrix elements with the full SM at $\mu = M_W$, which can be done in perturbation theory as the colour interactions are small in that energy regime. The results are *matched* to the EFT-matrix elements to determine the Wilson coefficients $C_i(M_W)$; the second step evolves the Wilson coefficients $C_i(M_W)$ down to the desired energy scale to obtain $C_i(m_b)$ [22, 24].

The matrix elements with the local operators Q_i contain the light fields and thereby complex colour interactions in the non-perturbative regime. These *subleading hadronic effects*, such as *form factors*, depend strongly on the specific decay process and on the momentum transfer to the leptons; the relevant details for exclusive $b \to s\ell^+\ell^-$ decays, like $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f'_2(1525)\mu^+\mu^-$ decays, are discussed in the next section.

In the context of indirect NP searches, one is specifically interested in the Wilson coefficients C_i. As they contain the physics interactions at large energy scales, effects from NP particles, which are typically expected to be heavy, would change the Wilson coefficients. This fact motivates also the specific choice of measurement observables, *e.g.* ratios of decay rates or specifically designed angular observables, in which hadronic effects cancel and increase the sensitivity to the Wilson coefficients; these observables are discussed in the second part of this chapter.

2.2.2 Exclusive semileptonic B decays

As explained before, effective field theories are a powerful tool to describe B decay processes. The Wilson coefficients are universal quantities across all B decays, but the local hadronic matrix elements depend strongly on the decay type.

Processes like $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays are *exclusive semileptonic B decays*, for which the local matrix elements read schematically

$$\mathcal{M} \sim \langle X\ell\ell|Q_i|B\rangle, \tag{2.17}$$

with X denoting the hadron in the final state, in this analysis' case a ϕ or f_2' meson. The matrix elements $\mathcal{M}$ can be separated into the quark- and lepton-factorising part and the *subleading hadronic effects*, *i.e.*

$$\mathcal{M} \sim \langle X|\mathcal{O}_{\mathcal{H}}|B\rangle \times \langle\ell\ell|\mathcal{O}_{b\to s}|0\rangle + \Delta_{\text{corr.}}, \tag{2.18}$$

with the *hadronic form factor* $\langle X|\mathcal{O}_{\mathcal{H}}|B\rangle$, the FCNC part of the interaction $\langle\ell\ell|\mathcal{O}_{b\to s}|0\rangle$, and the non-factorising corrections $\Delta_{\text{corr.}}$.

The form-factor calculations depend on the involved hadrons and the invariant dilepton mass squared q^2, and they are complicated by non-perturbative strong interactions, which are sources of large uncertainties in the predictions. For $B_s^0 \to \phi\mu^+\mu^-$, the current form-factor calculations use light-cone sum rules (LCSR) at low q^2 [25] and lattice QCD at high q^2 [26, 27], and the most precise predictions combine both methods in a fit across the q^2 spectrum [25]. For $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, the situation is rather unclear. Available form-factor predictions are compared in Ref. [28]; however, they differ and are thus topic of current debates.

The leading-order FCNC-part $\langle\ell\ell|\mathcal{O}_{b\to s}|0\rangle$ is affected by different operators depending on q^2; Appendix A provides a comprehensive list of the operators Q_i. The local operators Q_7, Q_9, and Q_{10} contribute with corresponding Wilson coefficients C_7, C_9, and C_{10}, and the operators Q_9 and Q_{10} contribute to the full q^2 range, and Q_7 enhances the decay rate at low q^2. The corresponding EFT diagrams are shown in Figs. 2.5(a–b).

Higher-order processes, shown in Figs. 2.5(c–d), include contributions from the four-quark operators Q_1^q and Q_2^q, which can be calculated with perturbation theory, and they are often merged with the operator Q_9 and associated to the *effective* Wilson coefficient C_9^{eff}. The contribution from Q_1^q and Q_2^q becomes large at q^2 values of $q^2 \approx 9.5\,\text{GeV}^2/c^4$ or $q^2 \approx 11.5\,\text{GeV}^2/c^4$, where the decay becomes a tree-level decay producing a J/ψ or $\psi(2S)$ meson. The q^2 regions that are dominated by these tree-level non-FCNC decays are typically vetoed in measurements of rare $b \to s\ell^+\ell^-$ decays.

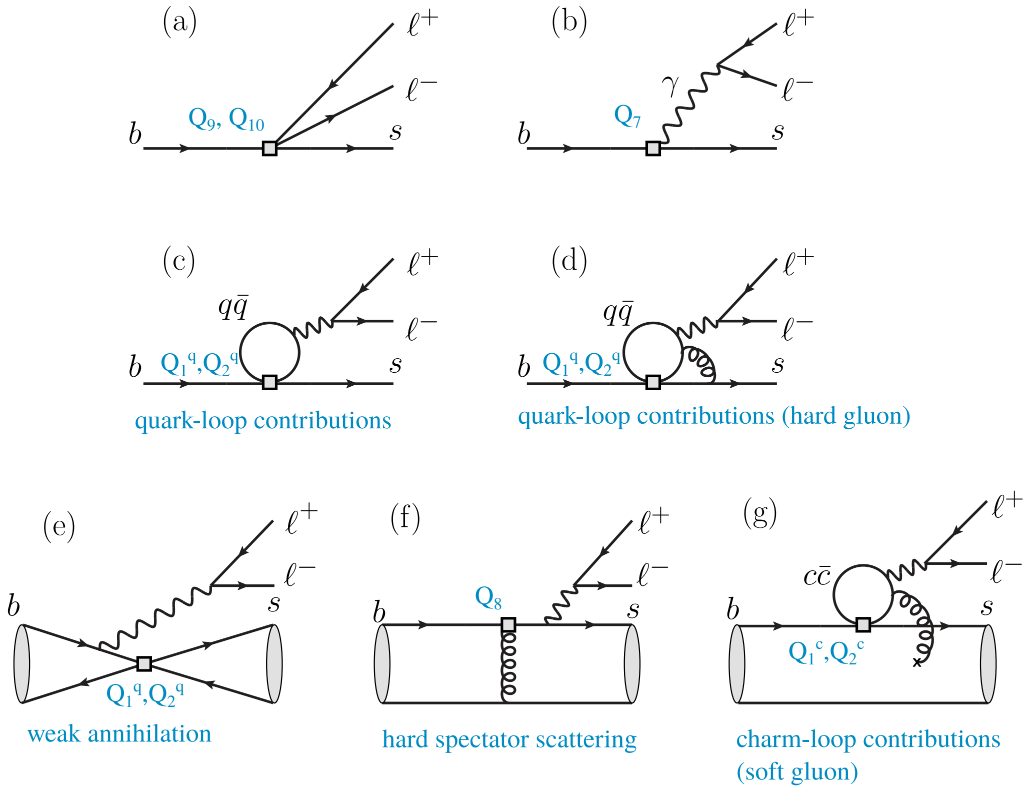

Figure 2.5: Effective field operators contributing to $b \to s\ell^+\ell^-$ decays. The figure is based on diagrams taken from Ref. [22].

Examples for processes where factorisation into the hadronic and the FCNC parts breaks down are processes where the hadronic bound-state is directly connected to the dilepton pair. The weak-annihilation process, shown in Fig. 2.5 (e), and the hard-spectator quark scattering, shown in Fig. 2.5 (f), contribute mostly at low q^2, and they can be computed with LCSRs. The most problematic subleading hadronic effect, however, is from charm-loop contributions with soft gluon, shown in Fig. 2.5 (g). Contributions from these processes are estimated with LCSRs, but they lead to significant systematic uncertainties on the predictions of decay rates.

All of the effects discussed above are included for $B^0_s \to \phi\mu^+\mu^-$ decays in the FLAVIO software package [29], which is used for the SM prediction and the corresponding theoretical uncertainties shown in Fig. 2.6. In addition to the SM prediction, two NP scenarios are also shown, where the Wilson coefficients C_9 and a combination of C_9 and C_{10} have been modified. These NP scenarios are of particular interest in the context of recent measurements of $b \to s\ell^+\ell^-$ processes, discussed in the following.

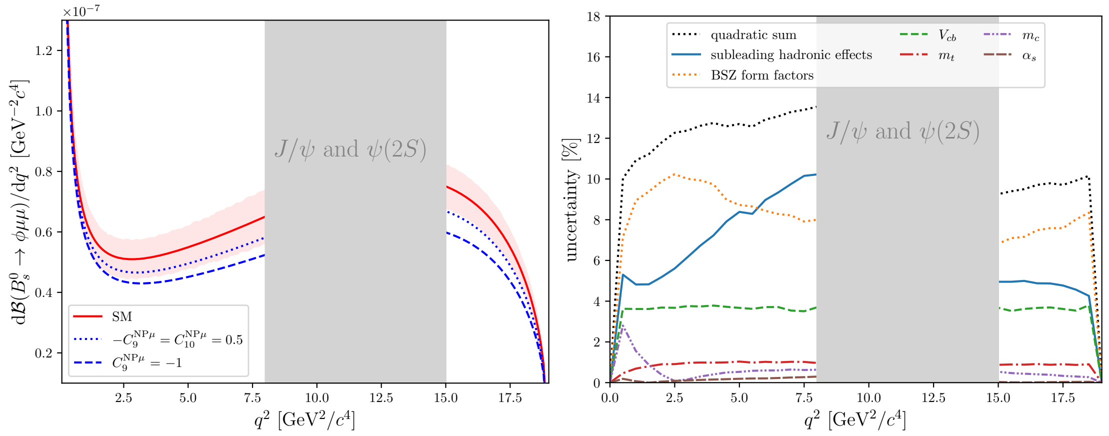

Figure 2.6: Theoretical (left) decay rate prediction for $B_s^0 \to \phi\mu^+\mu^-$ decays as a function of the invariant dimuon mass q^2 and (right) corresponding uncertainty breakdown. The SM prediction and total uncertainty is shown in red and two NP scenarios in blue dashed and dotted. The theoretical SM uncertainties are dominated by the form factors, shown in yellow, and the subleading hadronic effects blue line. The predictions are taken from the FLAVIO software package [29].

2.3 Measurements of rare $b \to s\ell^+\ell^-$ decays

The second part of this chapter focuses on the experimental status of $b \to s\ell^+\ell^-$ decays. Recently, discrepancies between SM predictions and corresponding measurements have generated interest in the flavour-physics community as so-called *flavour anomalies* [22, 23], which appeared in measurements of (1) ratios of decay rates testing *lepton-flavour universality (LFU)*, (2) angular distributions of $b \to s\mu^+\mu^-$ decays, and (3) exclusive $b \to s\mu^+\mu^-$ decay rates. The following sections summarise the experimental and theoretical aspects of the findings.

2.3.1 Lepton universality tests

The most striking deviations from the SM-prediction appeared in ratios of decay rates that test *lepton-flavour universality (LFU)*. The SM predicts that $b \to s\ell^+\ell^-$ decay rates are the same for final states with muons and electrons when neglecting lepton masses, so the ratios R_X

$$R_X = \frac{\Gamma(B \to X\mu^+\mu^-)}{\Gamma(B \to Xe^+e^-)}, \tag{2.19}$$

with X being a hadron, *e.g.* a K^0, a K^{*0}, or a ϕ meson, are predicted to be unity in the SM.

This prediction is precise as all effects related to the complex hadronic structures, *i.e.* the form factors or the non-factorising contributions, cancel in the ratio as these effects are known to be lepton-flavour universal. The leading theoretical uncertainties are from QED effects at 1–2% [30].

The most precise measurements of R_X have been carried out by the LHCb collaboration, measuring R_K [31], $R_{K^{*0}}$ [32], R_{pK} [33], $R_{K_S^0}$ [34], and $R_{K^{*+}}$ [34]. The measurement uncertainties are driven by statistical uncertainties; systematic

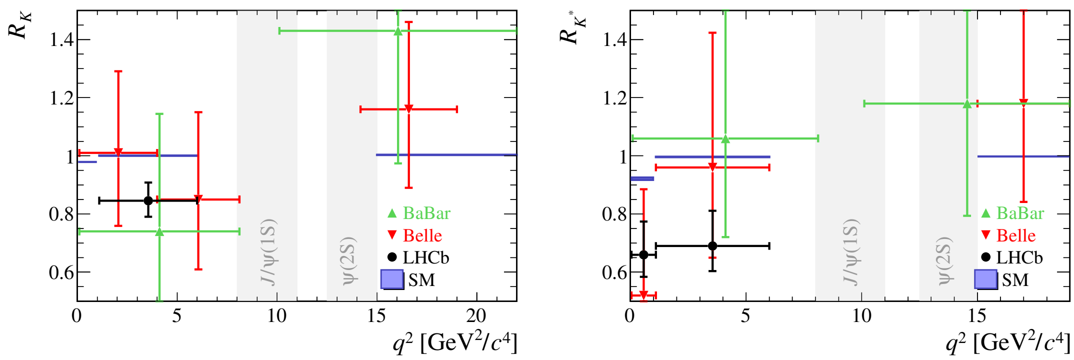

Figure 2.7: Lepton universality tests with the ratios of decay rates (left) R_K and (right) $R_{K^{*0}}$ as function of invariant dilepton mass squared q^2. Measurements from BaBar [37], Belle [35, 36], and LHCb [31, 32] are shown, and the values from LHCb are the most precise; the ratios are lower than expected in the low q^2 region. The figures are taken from Ref. [23].

effects cancel to a large degree because the ratios R_X are measured as *double ratios*, *i.e.*

$$R_X = \frac{\Gamma(B \to X\mu^+\mu^-)}{\Gamma(B \to X(J/\psi \to \mu^+\mu^-))} \Big/ \frac{\Gamma(B \to Xe^+e^-)}{\Gamma(B \to X(J/\psi \to e^+e^-))}, \tag{2.20}$$

with the precisely known flavour-universal high-yield $B \to XJ/\psi$ decay modes. These modes decay to the same final-state particles as the rare $b\to s\ell^+\ell^-$ decays, so the double ratios cancel many effects related to the reconstruction of final-state particles.

The measured R_X ratios seem to favour lower values than the SM prediction, and the largest deviations are found by LHCb in R_K with 3.1σ [31] and ($R_{K^{*0}}$) 2.1–2.5σ [32]. Their values are shown in Fig. 2.7, overlaid with results from Belle [35,36] and BaBar [37]. These and other LFU results from LHCb are compiled in Fig. 2.8 where a less significant, but similar trend is clearly visible.

To this current date, R_K is the only LFU test carried out with the complete data sample collected with the LHCb detector during the data-taking period 2011–2018. The other LHCb measurements so far use only half or quarter of that data; work on updating the results with all data is currently ongoing.

In the coming years, results from Belle II are also expected. The experiment started taking data in 2021, and will give important complementary information from a second experiment.

2.3.2 Angular analyses

The second flavour anomaly appeared in *angular analyses* of $b\to s\ell^+\ell^-$ decays. In angular analyses, the decay rate Γ is measured as differential to the decay angles Ω and invariant dilepton mass squared q^2, *i.e.*

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega\mathrm{d}q^2} = \sum_i I_i(q^2) f_i(\Omega), \tag{2.21}$$

with angular functions $f_i(\Omega)$ and corresponding coefficients $I_i(q^2)$.

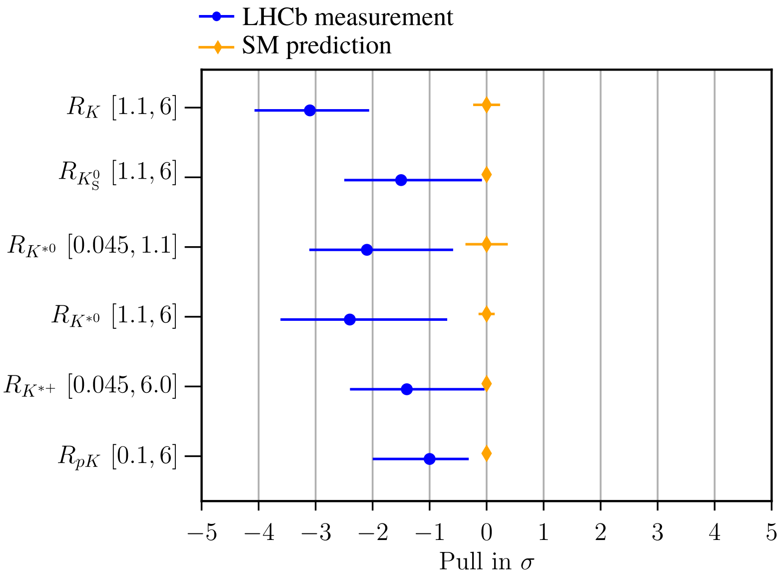

Figure 2.8: Differences between LFU variables measured by LHCb (blue) and predicted from the SM (yellow). All measurements lie below the SM prediction with R_K showing the largest deviation. This figure is based on a plot from Ref. [38].

The number of decay angles needed to describe the decay depends on the specific final state, *e.g.* one angle is needed to describe final states with a pseudoscalar meson, *e.g.* $B \to K\ell\ell$ decays, and three angles are needed for final states with vector mesons, *e.g.* $B_s^0 \to \phi\mu^+\mu^-$ decays. Angular analyses measure the *angular observables*, which are the q^2-dependent symmetries S_i (asymmetries A_i) defined as the normalised sum (difference) between the angular coefficients I_i for the decay rate Γ and the angular coefficients $\bar{I}_i$ for the *CP*-conjugated decay rate $\bar{\Gamma}$.

Of particular interest are theoretically clean observables, which are constructed from the observables to cancel hadronic form factors at leading order; they can be predicted precisely in the SM. One example for these are the P_i' observables [39], which are accessible in $B^0 \to K^{*0}\mu^+\mu^-$ decays. They have been measured by Belle [40], CMS [41], ATLAS [42], and most precisely by LHCb [43]; and in $B^+ \to K^{*+}\mu^+\mu^-$ decays, measured by LHCb [44]. Figure 2.9 shows the results for P_5', which exhibits a tension with the SM prediction in the q^2 range $1.1 < q^2 < 6.0\,\text{GeV}^2/c^4$ and thus constitutes one of the flavour anomalies. In both decay modes, P_5' tends to be above the SM prediction, and the deviation is particularly striking for $B^0 \to K^{*0}\mu^+\mu^-$ decays, where a local tension with the SM is found with 2.5-2.9$\,\sigma$ [43].

Other angular analyses of $B \to K\mu^+\mu^-$ [45] and $B_s^0 \to \phi\mu^+\mu^-$ [46] decays have been carried out by LHCb and other experiments and were found to be in agreement with the SM predictions [23]. It should be noted, however, that even though the $B_s^0 \to \phi\mu^+\mu^-$ decay contains a vector, its angular analysis cannot easily access P_5' as the decay is not *flavour specific*, *i.e.* the final-state particles are identical for decays of B_s^0 and $\overline{B}{}_s^0$ mesons. Without additional information on the initial flavour of the B_s^0 meson, only the *CP*-averaged decay rate can be measured. This means that only a subset of angular observables can be extracted, and measuring P_5' is only possible via an experimentally more challenging analysis that takes the initial B_s^0 meson's flavour and its decay time into account [23].

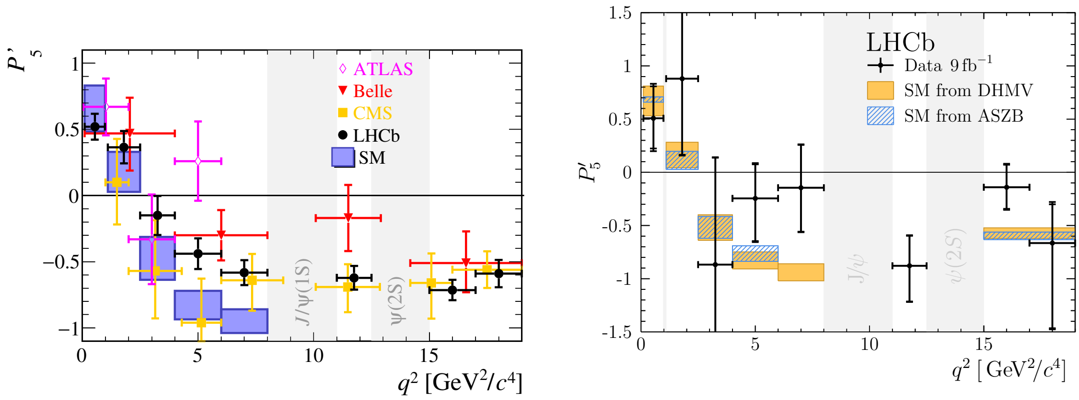

Figure 2.9: Angular observable P'_5 in (left) $B^0 \to K^{*0}\mu^+\mu^-$ decays, measured by ATLAS [42], Belle [40], CMS [41], and LHCb [43], and (right) $B^+ \to K^{*+}\mu^+\mu^-$ decays, measured by LHCb [44]. The figures are taken from Ref. [23] and Ref. [44], respectively. The $B^0 \to K^{*0}\mu^+\mu^-$ measurement is in tension with the SM prediction at low q^2 region, which is compatible with the situation in $B^+ \to K^+\mu^+\mu^-$ decays.

Similar to the LFU tests, the precision of the measurements is statistically limited, so effort is ongoing to update the measurements using the full LHCb data sample. Of particular interest is the $B^0 \to K^{*0}\mu^+\mu^-$ angular flavour anomaly, for which the latest measurement uses only half of the recorded LHCb data. The update of the result that uses the full data is currently ongoing.

2.3.3 Exclusive decay rates

Exclusive decay rates Γ are often measured relative to the total decay rate Γ_{tot}, namely as *branching fractions* $\mathcal{B} = \Gamma/\Gamma_{\text{tot}}$. Figure 2.10 summarises the results for $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$, $B^0_s \to \phi\mu^+\mu^-$ and $\Lambda^0_b \to \Lambda\mu^+\mu^-$ decays as a function of the invariant dimuon mass squared q^2 [23]. The measurements are mostly compatible with the SM predictions, but they tend to favour lower decay rates than expected in the low-q^2 region. A strikingly large deviation has been found for $\mathcal{B}(B^0_s \to \phi\mu^+\mu^-)$, which was around 3σ lower compared to the SM prediction in an analysis using the LHCb data collected during 2011–2012 [2]. The result has been confirmed by the analysis presented in this thesis, which includes the complete LHCb data collected to-date, *i.e.* during 2011–2018 [47].

The statistical precision of the exclusive decay-rate measurements is limited by the available data sample sizes. The systematic uncertainties are reduced by the measurement strategy: the branching fractions are measured *as ratios* to high-yield *normalisation decay modes*, which decay to similar final-state particles as the rare FCNC decay processes. This way, the measurement does not rely on the less well-known luminosity of the data or the hadronisation fraction as detailed in Sect. 4.1.

In addition, effects related to the reconstruction of the final-state particles cancel at leading order in the ratios, which reduces the related systematic uncertainties; however, the branching fractions of the normalisation decay modes are needed as input parameters. They constitute a systematic uncertainty that will be limiting the precision for some decay modes in the future.

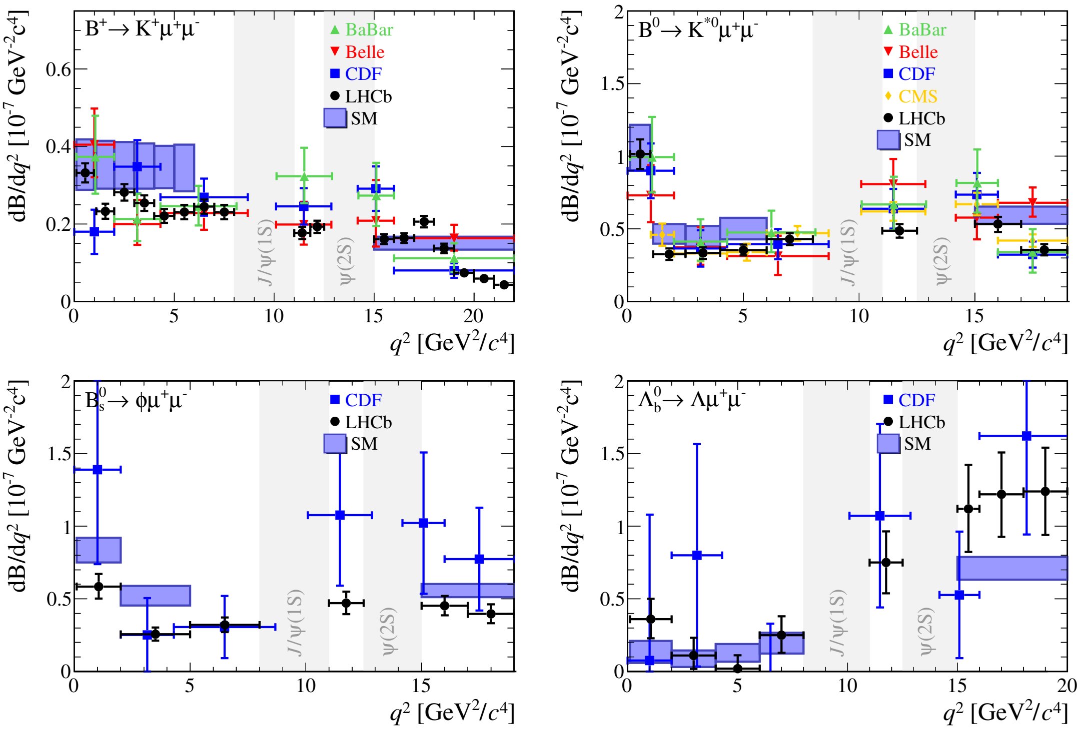

Figure 2.10: Branching fraction measurements of $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$, $B^0_s \to \phi\mu^+\mu^-$ and $\Lambda^0_b \to \Lambda\mu^+\mu^-$. The plots are taken from Ref. [23] and represent the situation before the $B^0_s \to \phi\mu^+\mu^-$ update, presented in this thesis, was published.

The theoretical predictions of decay rates are affected by relatively large uncertainties from the hadronic form factors and non-factorising diagrams, as discussed in Sect. 2.2.2. Especially the contribution from the charm-loop diagrams enters as a large source of uncertainty.

The theoretical uncertainties are reduced for the decay rate of the very rare $B^0_{(s)} \to \mu^+\mu^-$ decay as its final state contains only leptons and no hadronic bound state. For this reason, these decays are of specific interest. Their decay rate, however, is even harder to assess experimentally due to their small decay rates of $\mathcal{O}(10^{-10}\text{-}10^{-9})$. The $B^0 \to \mu^+\mu^-$ mode remains unobserved to-date, but the $B^0_s \to \mu^+\mu^-$ decay rate has been measured by LHCb [48], CMS [49], and ATLAS [50], and a combination of results from Ref. [51] indicates an agreement with the SM prediction at $\sim 2\sigma$.

Similar to the angular analyses and the LFU tests, more data will help to clarify the experimental situation. Several LHCb analyses only use part of the LHCb data sample, so updates of the measurements using the full data sample are currently ongoing to increase the precision of the results. Besides this experimental effort, work is also ongoing from the theory side, *e.g.* in Ref. [52]. Both the hadronic form factors and the charm loops currently pose large uncertainties on the SM predictions, complicating the interpretation of the results. Efforts to understand these effects better have increased with the interest in the flavour anomalies, so more precise predictions are to be expected.

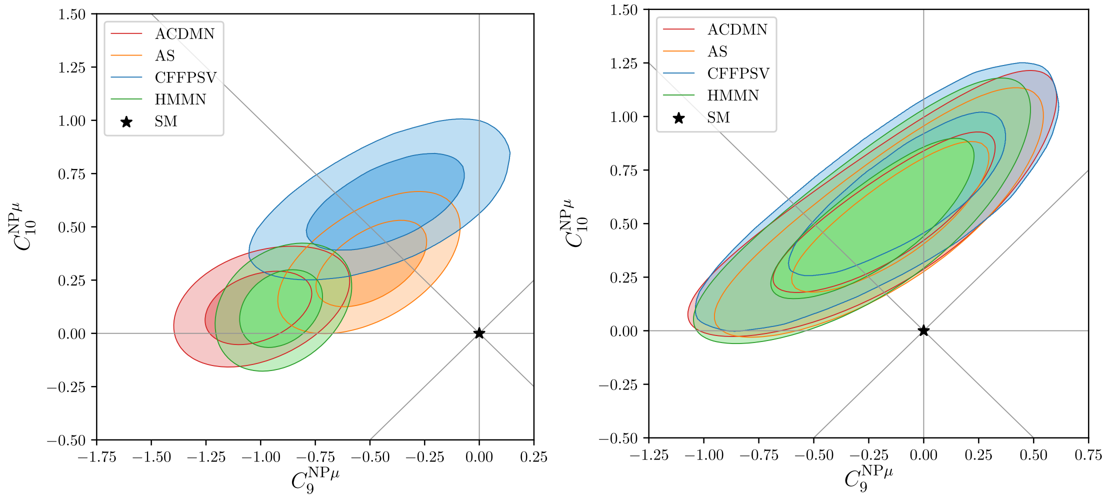

Figure 2.11: Wilson coefficient differences to the SM-values C_9^{NP} and C_{10}^{NP}, determined by taking into account two sets of flavour measurements: the (left) full set of $b \to s\ell^+\ell^-$ observables, or (right) only theoretically clean LFU tests and $B_s^0 \to \mu^+\mu^-$ decays [25, 51, 53, 54]. For both configurations, the best fit values for the Wilson coefficients deviate from the SM-expectation, and the results from all groups are in remarkable agreement when only using the clean observables. The figures are taken from Ref. [55].

2.4 New physics?

As shown in the previous sections, several tensions with the SM predictions have been observed in measurements of decays involving $b \to s\ell^+\ell^-$ transitions. None of these individual measurements is significant enough to claim NP, however, *many* of these individual deviations have appeared: branching fractions favouring lower values at low q^2, clean angular observables pointing towards values slightly larger than expected, and, most stringently, experimentally and theoretically clean LFU tests observing fewer muons than expected compared to electrons. All of the measurements test the SM, so combining them and testing the theory globally is of interest.

In the framework to describe the B decays, which are EFTs discussed Sect. 2.2.1, potential NP is encoded in the Wilson coefficients. So, determining the Wilson coefficients that describe best the complete set of $b \to s\ell^+\ell^-$ measurements tests the SM *globally* for NP. This effort has been carried out by multiple theory groups, *e.g.* [25, 51, 53, 54], and Fig. 2.11 shows their results for the Wilson coefficients C_9 and C_{10} that describe the data best [55]; for which (left) all $\mathcal{O}(100)$ available observables are taken into account, which are branching fractions of inclusive decays $B \to X\gamma$ and $B \to X\ell\ell$, of exclusive leptonic decays $B_{(s)}^0 \to \mu^+\mu^-$, and of exclusive radiative and semileptonic decays; angular observables, and LFU ratios of these decays; or (right) only theoretically-clean observables are considered, namely the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and the LFU ratios R_K and R_{K^*}.

The results are intriguing: both sets of observables clearly favour Wilson coefficients deviating from the SM prediction, even though the results differ slightly across the theory groups when considering the full set of observables. The level of agreement between the groups when limiting the analysis to the clean observables is, however, remarkable.

Limitations on the precision of the fitted Wilson coefficients originate from both the theory and the experimental side. While the LFU observables can be predicted very precisely, hadronic effects and associated theoretical uncertainties play a role for other observables. Efforts are ongoing on the theoretical side to understand whether the tensions are indeed signs for NP or only a misunderstanding of the hadronic effects. From the experimental side, efforts are ongoing to repeat the measurements with the full available data and update the results, which are currently statistically limited.

This thesis presents an analysis directly dedicated to this effort by measuring $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$, the most striking flavour anomaly among the decay rates, using the full LHCb data sample collected during 2011–2018. The results of this work, published as Ref. [47], have already been included in the global fits presented in Ref. [55], and they are consistent with the globally fitted Wilson coefficients C_9 and C_{10}.

Chapter 3

The LHCb experiment

The measurement of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, described in this thesis, uses data collected with the LHCb detector [56]. The LHCb detector is specifically designed for these kinds of heavy flavour decays, *i.e.* decays involving beauty or charm quarks, and allows to precisely measure their decay properties. These precision measurements are of interest as they allow testing the Standard Model (SM) prediction and, at the same time, searching indirectly for New Physics (NP) contributions, which could appear as deviations between the measurements and the Standard Model predictions. Examples for these precisely measured quantities at LHCb are *CP*-violation parameters or rates of rare decay processes, which include the measurement presented in this work.

The LHCb detector is located at *point 8*, which is one of the four interaction points of the Large Hadron Collider (LHC) at CERN. The LHC, located at the French-Swiss border close to Geneva, accelerates protons in a ring with a circumference of 27 km and collides them at four interaction points.[1] During the LHC operations between 2011 and 2018, the protons were collided with centre-of-mass energies between 7 TeV during 2011–2012 to 13 TeV during 2016–2018. The collisions are recorded by several experiments, among which CMS [57], ATLAS [58], ALICE [59], and LHCb [56] are the largest.

The proton collisions produce many particles including heavy quarks, as discussed in Sect. 2.2, and the LHCb detector is specifically designed to record their decays. The geometry of the LHCb detector covers the forward-region with a pseudorapidity range from 2 to 5, in which b quarks are produced as shown in Fig. 3.1. These so-produced b quarks form bound states like B mesons, which are the initial-state particles for the decay processes studied in this thesis.

The LHCb detector consists of several subdetector systems, which enable to measure the particle trajectories precisely and to identify the corresponding particle types. The subdetectors are specifically designed for analyses including heavy quark decays, and their combined information enables to cleanly distinguish the decays of interest from other decay processes, referred to as *background decays*. The detector design enabled to measure many decay processes with unprecedented precision; currently, 602 papers have been published with LHCb measurements [61], and 55 out of the 62 new hadrons discovered at the LHC have been found with the LHCb detector [62]. In addition to fulfilling the anticipated goals of LHCb, the detector has also shown to be useful in surprising sectors, namely by providing

[1] In a small fraction of the time, lead ions are accelerated and collided.

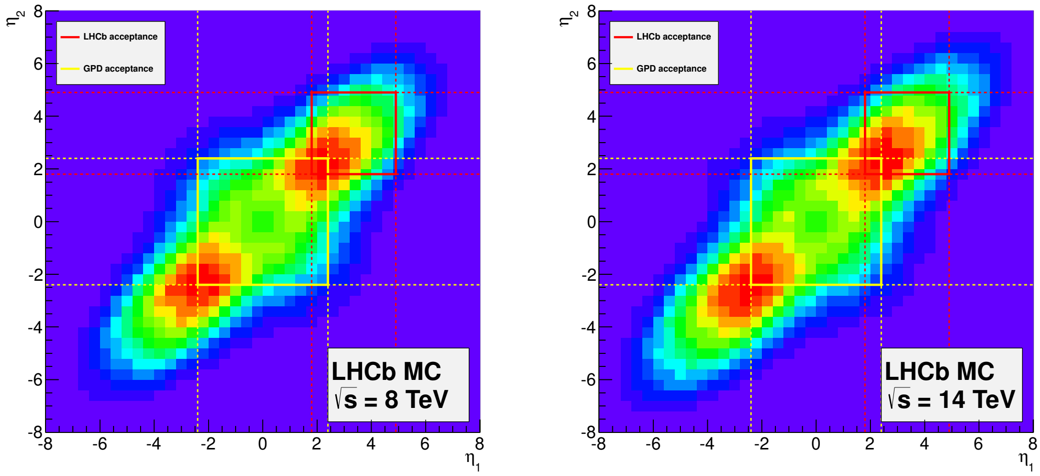

Figure 3.1: Simulated production angles of $b\bar{b}$-quark pairs in proton-proton collisions at centre-of-mass energies of (left) 8 TeV and 14 TeV [60]. The boxes indicate the typical acceptance of general-purpose detectors (GPD) (left) and of the LHCb detector (red).

complementary information on the production of W and Z bosons or on direct searches for NP particles [63].

The following sections describe the detector design including the subdetector systems and explain how the data is recorded and how the decay processes are reconstructed. The next section, Sect. 3.1, starts with sketching the typical B-decay structure, which drives the detector hardware design with its subdetectors, described Sect. 3.2. Then, the data-taking process is detailed in Sect. 3.3 including the LHCb trigger system, the track reconstruction, the particle identification, and the full decay reconstruction. The latter section presents also the author's specific contribution, which was re-optimising the selection criteria for the reconstruction of B_s^0 meson decays with two muons in the final state and implementing two new decay hypotheses in the LHCb data processing. All reconstruction steps rely on software, which is specifically developed for the LHCb experiment, explained in Sect. 3.4. Finally, the future of the LHCb experiment is sketched in Sect. 3.5 where the ongoing and future upgrades of the LHCb detector are being described briefly.

3.1 Decay signature and LHCb terminology

Analysing decays of b hadrons relies on cleanly distinguishing them from background decay processes. For this, specific quantities are employed among LHCb analyses that specifically exploit the typical decay signatures; these decay signatures are explained in the following. As the quantities are widely used in LHCb analyses and throughout the thesis, this section aims also at providing a glossary for the corresponding LHCb terminology.

A distinct feature of b hadrons is their significant lifetime, which makes them typically fly several millimeters before they decay. If they are neutral particles, like the B_s^0 meson, they traverse the tracking detectors without leaving a signal, and they remain undetected until they decay into charged particles. This behaviour results in a clear signature in the detector: charged particle tracks that are incompatible with

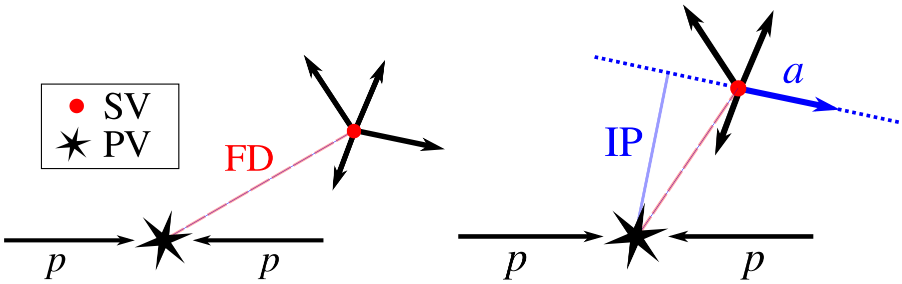

Figure 3.2: Graphic visualising the definition of the (left) impact parameter (IP) and (right) flight distance (FD) for an example B-hadron decay. The FD is defined as length of the connection between the primary vertex (PV) and the secondary vertex (SV), which is typically a few centimeters for b hadrons. The IP corresponds to the shortest distance between the PV and a particle track. It is typically small for tracks originating from the PV, e.g. particle b, and larger for secondary particles, e.g. particle a.

coming from the proton-proton collision, referred to as **primary vertex (PV)**, but that form a displaced vertex, referred to as **secondary vertex (SV)**.

A simple quantity to identify these decays is the **flight distance (FD)** of the potential B hadron, which is defined as the distance between PV and SV and visualised in Fig. 3.2. As both the positions of the PV and the SV are measured quantities with associated uncertainties, one often prefers the **flight-distance significance** ($\chi^2_{\mathbf{FD}}$), which is the FD divided by its uncertainty.

Another quantity denoting whether a particle originates from the PV is the **impact parameter (IP)**, defined as the minimum distance between a particle track and the PV, sketched in Fig. 3.2. Similar as for the FD, one often prefers the **impact-parameter significance** ($\chi^2_{\mathbf{IP}}$), *i.e.* the IP divided by its uncertainty.[2]

Finally, the **direction angle (DIRA)** is often used to check whether an unstable particle is consistent with stemming from the PV. It is defined as the angle between (a) the reconstructed momentum vector of the particle and (b) the connection between PV and decay vertex of that particle. A particle stemming from the PV has a small DIRA.[3]

The quantities described above are typical for general B hadron decays—however, to measure specific B-hadron decays, like $B_s^0 \to \phi\mu^+\mu^-$, it is crucial to identify also the final-state particles. The LHCb experiment provides three types of quantities for this task: the **PID**$_x$ variables, the **ProbNN**$_x$ variables, and binary classifiers, *e.g.* the `isMuon` variable; Sect. 3.3.3 describes them in detail.

The **PID**$_x$ variables, sometimes also referred to as **combined delta log-likelihood (CombDLL**$_x$**)** variables, combine information from several LHCb subdetectors to compare the likelihood of a particle being of species x with the likelihood of the particle being a pion. Using the pion hypothesis as reference for the PID variables is practical because pions are the most abundant particles produced in the proton-proton collisions.

The second type of particle-identification variables are the **ProbNN**$_x$ variables. These variables combine the PID$_x$ variables with more quantities from the LHCb subdetectors using neural nets. Their values are between 0 and 1 where large

[2]Technically, one determines the χ^2_{IP} of a particle x as change in the PV-fit quality (χ^2_{vtx}) with and without considering the track of particle x, but the resulting value is very similar to IP/σ_{IP}.
[3]Most often, the cosine of the DIRA is used.

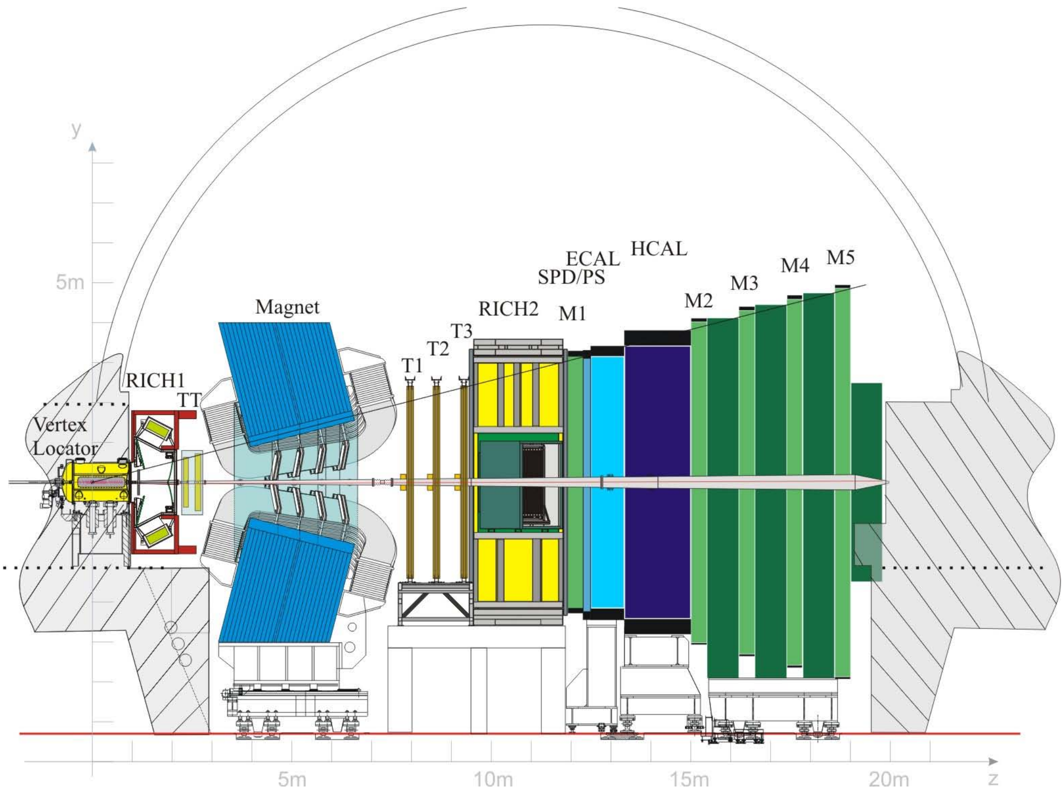

Figure 3.3: Overview of the LHCb detector [65].

values indicate that the corresponding particle is compatible with the particle hypothesis x. The ProbNN$_x$ variables perform often better than the corresponding PID$_x$ variables; in some cases, however, analysts prefer using the less complex PID$_x$ variables as they are easier to understand and to model in simulation.

Finally, the third type of particle identification variables classify particles using a binary quantity. The most relevant for this thesis is the `isMuon` classifier, which identifies particle tracks as muons based on a list of criteria, listed in Ref. [64]. The `isMuon` classifier is determined already while taking the data since it can be evaluated very rapidly; this makes it a crucial ingredient to the LHCb data processing. In the high-level data analysis, the muons are mostly selected based on the PID$_\mu$ or ProbNN$_\mu$ variables as they combine more information and therefore perform better.

3.2 The LHCb detector

An overview of the LHCb detector with its subdetectors is shown in Fig. 3.3. This version of the detector was used to collect the data during 2011–2018, *i.e.* the data that are used for the measurements described in this thesis. Currently, the detector is being upgraded for future data taking; this upgrade and the future upgrade plans are described in Sect. 3.5.

The following sections describe the currently used subdetectors briefly, and a comprehensive description of the detector design can be found in Ref. [56]. The subdetectors provide complementary information on the particles that are produced in the proton-proton collisions, and combining this information allows to reconstruct the complete underlying decay processes.

The tracking system, detailed in Sect. 3.2.1, detects *hits* where charged particles traversed the detector material, and these hits are the basis to reconstruct the full trajectories of the particles in the LHCb detector. The trajectories carry also information on the momentum of the particles as they are bent by the LHCb magnet, which provides an integrated magnetic field of 4 Tm. The bending radius R depends on the particles' momenta, *i.e.* $R = p_{\mathrm{T}}/(cqB)$ with the transverse momentum of a particle p_{T}, its charge q, and the magnetic field strength B. In addition to measuring the individual particle tracks, the tracking system is also used to obtain the position of the initial proton-proton collision, the primary vertex (PV), which is crucial to identify the b hadron decays by their significant lifetime.

The other subdetectors are designed to identify the particle types that correspond to the measured tracks. The most important subdetectors for this task are the ring-imaging Cherenkov detectors (RICH), detailed in Sect. 3.2.2. The RICH detectors measure Cherenkov light emitted from the charged particles traversing them, allowing to infer the track's velocity; combining information on a track's velocity with the track momentum, determined with the tracking system, allows to infer its rest mass and therefore its type.

The RICH detectors also provide PID information for muons, but the most stringent muon identification is provided by the muon detectors, described in Sect. 3.2.3. These detectors form the last part of LHCb, which is only reached by muons. In contrast to other particles, muons traverse the material as *minimally ionizing particles (MIP)*, which means their energy loss in the material is close to the minimum. Therefore, they traverse the full detector and the muon detectors in contrast to other particles, which are stopped before in the detector or by the absorbers interlacing the muon detectors. This fact is used to identify muons: if particle tracks are compatible with hits in the muon chambers, they are considered as muons.

The remaining type of subdetectors are the calorimeters. The calorimeters measure the energy deposit of stopped particles, which is essential for analyses with neutral particles and electrons. For this thesis, information from the calorimeters is less relevant; however, as they are crucial for other analyses, they are briefly described in Sect. 3.2.4.

3.2.1 Tracking detectors

The LHCb tracking system comprises five subdetector systems: the VELO detector [66], the tracker turicensis (TT), and the three tracking stations T1–T3 [67, 68]. The VELO detector consists of 42 half-disk shaped modules, which surround the beam pipe close to the proton-proton interaction point. Each module contains two kinds of silicon-strip sensors, one returning the radial and the other the angular coordinate of a charged particle hit relative to the beam direction. The single-hit resolution depends on the track coordinates, but it is roughly between 5 and 25 μm for the radial and between 7 to 18° for the angular coordinate [69]. This precision enables to determine the primary vertices (PV) with an x and y (z) resolution of better than 15 (100) μm for a typical event that contains more than 15 tracks.

The other tracking detectors, *i.e.* the TT and T1–T3, measure charged particle hits between the two RICH detectors, where the TT is placed in front of the magnet and T1–T3 behind. Each of these tracking detectors comprises four detector layers with the first and fourth layer containing vertical sensors, measuring the magnet-bent x-coordinate of a particle hit, and the second and third layer containing $\pm 5°$-rotated sensors to get additional information on the y coordinate.

Two sensor-technology types are employed: (1) silicon strips for the TT and the inner parts of T1–T3, referred to as *inner tracker (IT)*, and (2) straw drift tubes for the outer part of T1–T3, referred to as *outer trackers (OT)*. Silicon strips provide a good resolution and radiation hardness, which are required features for the high particle fluxes in the regions covered by the TT and IT. In the OT region, straw drift tubes are sufficient due to the smaller particle fluxes there.

The information from all tracking detectors is combined for the reconstruction, detailed in Sect. 3.3.2, where the particle hits are combined to form reconstructed particle tracks with momentum information. The momentum is measured with a resolution of roughly 5 (8) per mill for particles with momentum of $< 20\,\mathrm{GeV}/c$ (around $100\,\mathrm{GeV}/c$) and the impact parameter (IP) with a resolution of roughly $15\,\mu\mathrm{m}$ [70].

3.2.2 Ring-imaging Cherenkov detectors

The ring-imaging Cherenkov (RICH) detectors provide information on the particle type associated to a measured particle track [71]. The LHCb detector uses two RICH detectors: RICH 1, which is placed between VELO and TT, and RICH 2, placed between the tracking detector T3 and the first muon chamber. Both RICH detectors measure the Cherenkov light that is emitted by a traversing particle, and the light yield and emittance angle depend on the velocity of the particle; the emittance angle, also denoted as Cherenkov angle θ_C, is given by $\cos\theta = (n\beta)^{-1}$ with the refractive index of the material n and the speed of the particle relative to the speed of light β. The combination of track momentum, measured with the tracking system, and the emitted Cherenkov light allows to infer the particle's rest mass. Details on the specific algorithms to obtain the particle identification (PID) are given in Sect. 3.3.3.

The RICH-detector material is optimised to identify particles within a wide momentum range: less energetic particles with momenta between 2 and $40\,\mathrm{GeV}/c$ are measured with RICH 1, which is filled with silica aerogel and C_4F_{10} gas,[4] and more energetic particles with momentum between 15 and $100\,\mathrm{GeV}/c$ are measured with RICH 2, which is filled with CF_4 gas. Both detectors use a similar configuration of mirrors and hybrid photon detectors, which were specifically developed for the RICH detectors, to measure the emitted Cherenkov light.

For this thesis, the most relevant PID information from the RICH is to identify kaons. Using RICH information, kaons can be cleanly distinguished from the highly abundant pions: identifying kaons with an efficiency of 95% results in a pion misidentification-rate of only 5% [72].

[4]The aerogel has been removed for 2015–2018 data-taking. Due to the large track multiplicities, it produced too many photons decreasing the RICH 1 efficiency and increasing significantly the computing times.

3.2.3 Muon detectors

The muon system in LHCb consist of the five muon detectors M1–M5 [73]. The first detector, M1, is placed behind RICH 2 and in front of the calorimeters, and the others, M2–M5, behind the calorimeters. The muon detectors are composed of multi-wire proportional chambers except for the inner part of M1, which uses gas electron multiplier detectors to cope with the larger particle fluxes in that area. Between M2 and M5, ion absorber plates are placed to stop other particles and ensure that only muons traverse the muon detectors.

The clean signature of muons in the muon chambers allows using the muon identification information already in the trigger, which is discussed in more detail in Sect. 3.3.1. It further allows to select the muons efficiently with low background-misidentification rates: selecting muons with an efficiency of 97% results in a pion misidentification-rate of only 1–3% [64].[5]

3.2.4 Calorimeters

The calorimeter system of LHCb consists of the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the pre-shower (PS), and the scintillating pad detector (SPD), which sit between the muon chamber M1 and the other muon detectors M2–M5. The calorimeters are essential for identifying and measuring neutral particles and electrons. As the decays studied in this thesis only contain kaons and muons in the final state, information from the calorimeters is less crucial; however, calorimeter information is used in the muon hardware trigger, where events with too many tracks in the SPD are vetoed, and it is used to complement the muon identification, as detailed in Sect. 3.3.3.

The calorimeter detectors in LHCb share a common design: they are composed of dense material layers that are alternated with scintillation detectors. An incoming particle interacts with the dense material and deposits its energy by initiating a *shower* of secondary particles; these secondary particles are then converted to photons in the scintillation detectors and serve as measure for the energy deposited by the incident particle. This detector design is applied to all parts of the calorimeter system, for which only the specific material differs.

The SPD and PS detectors use a lead plate as absorber and plastic scintillator to identify electrons and photons in combination with the ECAL, which consists of several layers of polystyrene scintillator and lead plates. The ECAL stops electrons, photons, and neutral pions, and it is crucial to correct the electron energy when it emitted bremsstrahlung photons. The HCAL consists of several layers of scintillating tiles and lead absorber, and it is optimised for neutral hadrons. The information from all calorimeter systems is combined to identify the particle type corresponding to a measured track as detailed in Sect. 3.3.3.

[5] These numbers are obtained when combining information from the muon chambers, the RICH detectors, and the calorimeters; when using only information from the muon system, the pion misidentification rate is slightly larger.

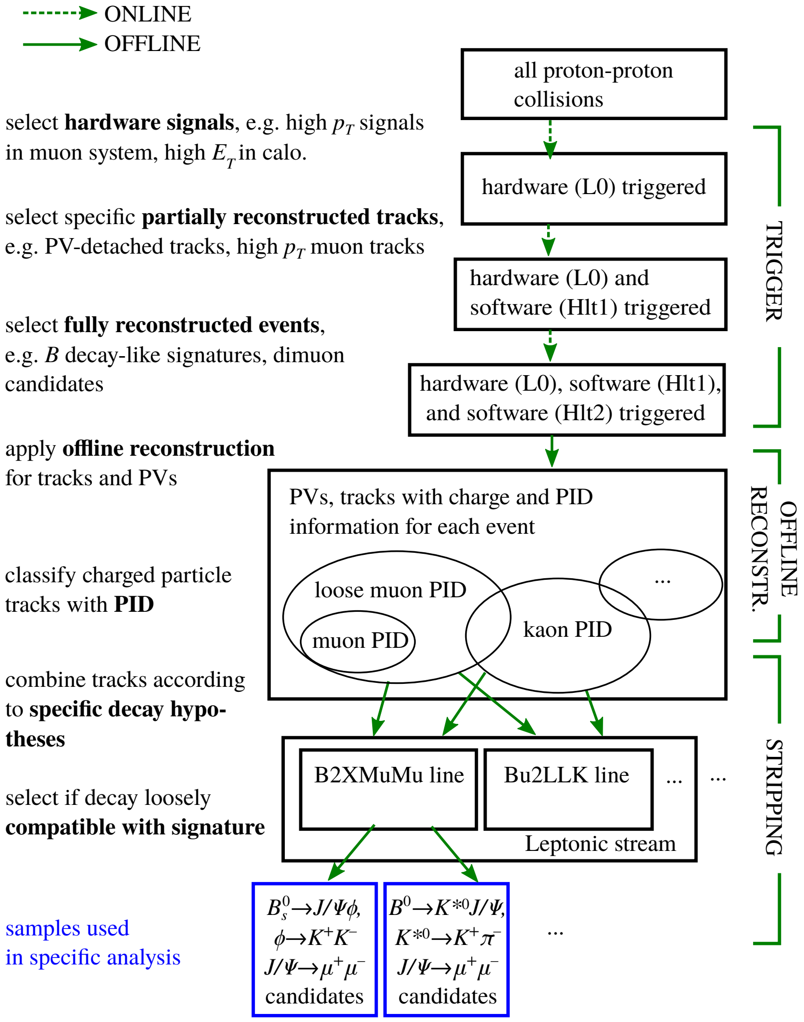

Figure 3.4: Sketch of the data flow for a typical rare-decay analysis in LHCb. The triggers select the events worth recording during the data-taking (*online*), and all other steps, *i.e.* the full track- and PV-reconstruction, the particle identification and the matching with decay hypotheses are performed after the data-taking (*offline*).

3.3 Trigger and reconstruction

Following the description of the detector hardware, this section explains how the detector is run to collect the data that are used later for the measurement of the $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. The schematic in Fig. 3.4 gives an overview on the different data-taking stages, starting with the proton-proton collisions and ending with the data samples containing the loosely prefiltered decay candidates of interest. All of these steps are performed globally for the different analyses; the individual analysts usually start their work with the output of this global data flow.

As already indicated before, recording all particles emerging from the proton-proton collisions would by far exceed the available writing speed and disk storage. So, only interesting events are recorded, and these interesting events are selected by the LHCb trigger system.

The LHCb trigger system is detailed in Sect. 3.3.1. It comprises a hardware

(L0) trigger, which uses large transverse momenta p_{T} in the muon chambers or large transverse energy deposits E_{T} in the calorimeters; and a software trigger, which uses more complex signatures, *e.g.* tracks that are displaced from the primary vertex (PV). The software trigger is organised in two stages, where the first stage (Hlt1) performs a partial and the second stage (Hlt2) a full event-reconstruction. For this, *reconstruction algorithms* are used, which combine the responses from different subdetectors to obtain the particle trajectories with charge and, in Hlt2, particle identification information. As the trigger systems run while the data is taken, they are referred as *online* stages of the data-taking, and events rejected at this stage cannot be restored.

The next step of the data-taking process is the *offline reconstruction*, which runs on the data stored after the trigger decisions. The offline reconstruction consists of the track and PV reconstruction, described in Sect. 3.3.2, and the particle identification, described in Sect. 3.3.3, and it uses similar algorithms as in the software trigger; however in the offline reconstruction, processing time is less critical.

The last stage of the data flow, the *stripping stage*, combines the tracks and PVs to form the *b hadron decay candidates*, as detailed in Sect. 3.3.4. In this stage, the data are already loosely selected to match the signal decay; in this analysis, decay candidates are further processed if they are compatible with a B meson that further decays into two muons and a hadron, which further decays to two kaons.

3.3.1 Trigger system

The LHCb trigger system is designed to identify the particle decays that are interesting for LHCb analyses across the vast amount of other processes emerging from the proton-proton collisions. The protons collide with a frequency of 40MHz—orders of magnitude larger than what is manageable in terms of storage writing speed and memory space. This makes triggering of interesting events a crucial part of the LHCb data-taking. The typical branching fractions of rare decay processes, *e.g.* the decays analysed in this thesis, are of order 10^{-7}, so the trigger decisions need to select the decay signatures efficiently to make these measurements possible.

These constraints, *i.e.* reducing the event rate while maintaining a high selection efficiency, are addressed by the LHCb trigger system. The trigger system combines a fast hardware trigger (L0) stage, which uses signals from individual subdetectors, and two software trigger (Hlt) stages, which combine information from different subdetectors. For the latter, events are first reconstructed partially (Hlt1) and then fully (Hlt2) to decide whether to keep or reject an event under question.

The following paragraphs give a brief overview on the three trigger stages and the corresponding trigger selection criteria referred to as *trigger lines*. The trigger selections specifically chosen for the analysis of $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f'_2(1525)\mu^+\mu^-$ can be found in Sect. 4.3.2.

The L0 stage is the first trigger stage, running synchronously with the proton-proton collisions in the detector. The L0 trigger lines uses information from the muon stations and the calorimeter system. The muon detectors identify events with single (`L0Muon`) or pairs of muons (`L0DiMuon`) that exceed certain transverse momentum thresholds, for which the muons are identified as hits in the muon system, and the momentum is estimated by comparing the hit-positions in the first and second muon chamber M1 and M2 to the position of the interaction point; this estimate results in a momentum resolution of roughly 20% for each track [74]. The calorimeters identify events with large transverse energy deposits that indicate the presence of high-energy electrons (`L0Electron`), photons (`L0Photon`), or hadrons (`L0Hadron`); these are not discussed in this thesis as they are not used in the measurement of $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f_2'(1525)\mu^+\mu^-$. The combined L0 trigger lines reduce the 40 MHz bunch crossing rate to 1 MHz at which the full detector can be read out.

The Hlt1 stage uses primary vertices (PVs) and charged high-momentum long tracks, which are reconstructed similarly to the reconstruction applied offline, described Sect. 3.3.2. Only tracks with transverse momenta larger than 1.2 GeV/c^2 are processed during 2011–2012 data-taking, and this momentum threshold could be lowered to 500 MeV/c^2 in 2015–2018 data-taking [75] by adding a charge estimate to the track, which limits the search window for hits in the tracking system and by that reduces the computing times. For these long tracks, muon identification is available thanks to the clean signatures in the muon chambers; other PID information is not available in the Hlt1 trigger stage since their corresponding processing time is too large.

The PVs and long tracks with muon PID information are used in the Hlt1 trigger lines to identify events with PV-displaced tracks (*e.g.* `Hlt1TrackMVA`), events with two tracks and common PV-displaced vertex (*e.g.* `Hlt1TwoTrackMVA`), and events with muon tracks that are either displaced (*e.g.* `Hlt1MuonTrack`), with very high momentum (*e.g.* `Hlt1SingleMuonHighPT`), or in pairs with constraints on their combined invariant mass (*e.g.* `Hlt1DiMuonHighMass`). The full Hlt1 stage reduces the event rate to roughly 70kHz in 2011–2012 and 110kHz in 2015–2018.

The Hlt2 trigger stage performs a full and offline-like event reconstruction, for which the full PID information is available. This full detector information enables using requirements tailored specifically to the full variety of physics programs, *e.g.* special triggers for charm decays or for lifetime measurements. A commonly used class of trigger lines to identify B-hadron decays are the *topological trigger lines*, which search for PV-displaced 2-, 3-, or 4-track vertices with B decay-like properties (`Hlt2Topo{2,3,4}Body`), which are combined in a multi-variate analysis (MVA) and tuned on simulated signal and data background decays [76, 77]. If one of the tracks is identified as a muon (`Hlt2TopoMu{2,3,4}Body`), the requirements on the MVA can be loosened.

A further trigger line selects displaced single (*e.g.* `Hlt2SingleMuon`) or pairs of (*e.g.* `Hlt2DiMuonDetached`) muon tracks, similar to the Hlt1 triggers but with more fine-tuned parameters based on the full event-reconstruction. After the Hlt2 stage, the data is reduced to roughly 5kHz in 2011–2012 data and to 12.5kHz for 2015–2018 data, which is stored for offline analysis.

The trigger scheme described in this section will be dramatically changed after the LHCb upgrade, briefly described in Sect. 3.5, where the full hardware trigger stage will be removed. The upgrade is needed as the LHC is upgraded toward a high-luminosity LHC (HL-LHC), with the luminosity of colliding proton-beams in LHCb eventually being increased by two orders of magnitude. This larger luminosity requires a re-design of the subdetector parts to enable data taking at this changed environment, and the upgrade will enable taking more data: currently, the hardware trigger reduces the event rate to 1 MHz at which the detector can be read out. The hardware trigger limits especially the efficiency of purely hadronic modes, but these efficiencies will be improved by a factor 2 to 4 with the upgraded detector, which can be read out at the LHC bunch-crossing rate of 40 MHz [78].

3.3.2 Track and primary vertex reconstruction

Reconstructing the PVs and the charged tracks is performed during the offline reconstruction of the collected data, however, parts of these algorithms are also used in the software trigger as explained before in Sect. 3.3.1. As the analysis described in this thesis uses only charged particles in the final state, the following covers only the reconstruction of charged particles; the reconstruction of neutral particles in LHCb is described in Ref. [69].

Most analyses, including this one, use only tracks that are reconstructed as *long tracks* [70], which are defined as tracks with a sufficient number of hits in the full tracking system, *i.e.* the VELO, the TT, and the three tracking chambers T1–T3. Long tracks are reconstructed by first connecting hits in the VELO detector with straight lines, which is a good description due to the absence of the magnetic field in the VELO. These *VELO tracks* are only considered if at least six hits are measured in the VELO.

Then, the VELO tracks are combined with hits in the tracking chambers T1–T3 using two different algorithms: the *forward tracking* and the *track matching*. The forward tracking algorithm starts by combining a VELO track with a single hit in the tracking chambers. If other hits in the tracking detectors confirm that particle track, it is accepted, and the other hits are assigned to it. The track-matching algorithm combines the VELO tracks with tracks in the tracking detectors that are found from a standalone algorithm. Both types of composite tracks are finally combined with compatible hits in the TT to improve the corresponding momentum resolution.

The long tracks are then fitted with a Kalman Filter described in Ref. [79]. The fit returns the x and y positions of the track, the derivatives dx/dz and dy/dz, for which z is the coordinate along the beam; and the charge-momentum ratio q/p at a given positions z, which is derived including information on the magnetic field [69].

The VELO tracks are further used to determine the position of the primary vertices [80]. For this, all tracks with hits in the VELO are used in 2011–2012, and only VELO tracks in 2015–2018 data [75]. The PV reconstruction consists of two steps: first, finding possible PV positions, and second, fitting the tracks to confirm or improve the PV position. The fit minimises the impact-parameter significance χ^2_{IP} of the considered tracks with respect to the PV, and a method is used to remove biases from tracks that originate from secondary vertices [80].

3.3.3 Particle identification

Identifying the particle species of a given track relies on information from the RICH detectors, the calorimeters, and the muon stations. Hadrons are mostly identified with the RICH detectors and muons with the muon detectors; however, information from other subdetectors are added to provide a combined particle identification variable. This section focuses on the identification of charged hadrons and muons as these particles constitute the final states of the $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, investigated in this thesis.

The RICH detectors provide information on whether a track corresponds to an electron, muon, pion, kaon, proton, or deuteron. As explained in the description of the RICH in Sect. 3.2.2, the measurement relies on the Cherenkov light emitted by the particle track under question. As, however, the events are typically heavily crowded with many Cherenkov rings overlapping, it is not possible to extract the separate Cherenkov angles and photon yields for each track. Therefore, the full Cherenkov ring-pattern in an event is evaluated using the following iterative approach [81, 82]:

First, (1) all tracks are assumed to be pions, and (2) the expected Cherenkov pattern in the RICH detectors is calculated using the tracks' measured momenta. This expected pattern is (3) compared to the measured pattern, and a likelihood quantifying the difference between the measured and the expected pattern is calculated. This procedure is repeated with each track's particle hypothesis in (1) altered, and the hypothesis minimising the likelihood in (3) is accepted. This minimised likelihood $\mathcal{L}^{\text{RICH}}$ is the basis for the RICH particle identification variable $\text{DLL}_x^{\text{RICH}}$, which is defined as

$$\text{DLL}_x^{\text{RICH}} = \log \mathcal{L}_x^{\text{RICH}} - \log \mathcal{L}_\pi^{\text{RICH}}, \tag{3.1}$$

with the likelihood (a) considering the respective track as pion $\log \mathcal{L}_\pi^{\text{RICH}}$ and (b) considering the respective track as particle type x, which can be an electron, muon, pion, kaon, proton, or deuteron, $\log \mathcal{L}_x^{\text{RICH}}$. The pion hypothesis is chosen as reference as these are the most abundant particles in the LHCb detector.

The second system providing particle identification (PID) information is the muon system. The muon system is used for two types of PID variables [83]: the binary `isMuon` classifier and the $\text{DLL}_x^{\text{MUON}}$ variable. The former variable classifies a track as muon if a sufficient number of hits is found in the muon system close to the track under study; the details can be found in Ref. [83]. If a track is classified as muon, further information on the particle identity is provided with the $\text{DLL}_x^{\text{MUON}}$ variable, in which the distance between (a) the muon detector hits and (b) the respective track extrapolated to the muon chambers is taken into account. This measured hit-track distance is then compared to the expected distances from simulated muon (proton) samples, and a muon-likelihood (non-muon-likelihood) is constructed from the difference. The muon PID classifier, $\text{DLL}_\mu^{\text{MUON}}$, is defined as the difference between the two likelihoods, *i.e.*

$$\text{DLL}_\mu^{\text{MUON}} = \log \mathcal{L}_\mu^{\text{MUON}} - \log \mathcal{L}_{\text{non-}\mu}^{\text{MUON}}. \tag{3.2}$$

Finally, PID information is provided by the calorimeter measurements. The calorimeters are especially important to identify neutral particles and electrons; however, they also add information for muon identification. Even though the muons are mainly identified with the muon system, they can also be discriminated against other particles based on their energy deposits in the HCAL and ECAL [84]: they are *minimum ionizing particles (MIP)*, *i.e.* they deposit minimal energy in the calorimeters, whereas other particles tend to deposit more energy along their track. This feature is used to create different likelihoods for calorimeter energy deposits to be caused by muons, $\log \mathcal{L}(\mu)^{\text{CALO}}$, or other particles, $\log \mathcal{L}(\text{non-}\mu)^{\text{CALO}}$, which are determined on data samples containing hadrons and electrons traversing the calorimeters. The calorimeter PID for muons, $\mathrm{DLL}_x^{\mathrm{CALO}}$, is then defined as

$$\mathrm{DLL}_\mu^{\mathrm{CALO}} = \log \mathcal{L}(\mu)^{\mathrm{CALO}} - \log \mathcal{L}(\text{non-}\mu)^{\mathrm{CALO}}, \tag{3.3}$$

for which the information from the ECAL and HCAL are added.

All information explained before, *i.e.* information from the RICH detectors, the muon detectors, and the calorimeters, are combined in two global types of particle identification variables: The (1) *PID_x* variables, and the (2) *$ProbNN_x$* variables. The PID_x variables simply add the likelihoods from the different sub-systems, *i.e.* the muon-PID variable PID_μ is defined as

$$\mathrm{PID}_\mu = \mathrm{DLL}_\mu^{\mathrm{MUON}} + \mathrm{DLL}_\mu^{\mathrm{RICH}} + \mathrm{DLL}_\mu^{\mathrm{CALO}}, \tag{3.4}$$

and the kaon-PID variable PID_K as

$$\mathrm{PID}_K = \mathrm{DLL}_\mu^{\mathrm{RICH}}, \tag{3.5}$$

with the variables defined as above.

The more complex *$ProbNN_x$* variables are obtained by combining the PID_x variables and other subdetector variables, *e.g.* track momenta and track-fit qualities, into a classifier based on neural nets.[6] The neural nets are trained on simulated particle decays to reach maximal separation power between the different particle hypotheses, and in many cases they perform better than the simple PID_x variables.

In addition to the standard particle hypotheses, *e.g.* ProbNN_μ and ProbNN_K, a variable named *$ProbNN_{ghost}$* is available. This variable is created similarly to the standard ProbNN_x variables, but it is specifically trained to identify *ghosts*, which are particle signatures created by random combinations of detector hits and noise. An important input parameter to the *$ProbNN_{ghost}$* variable is the *ghost probability* variable `GhostProb`, which uses information from the tracking systems [86].

Figure 3.5 shows example distributions of the PID_μ variable from simulated muons and pions and the distribution of the ProbNN_K variable for simulated kaons and pions; the distributions clearly differ for the signal particles and the background pions. The PID_μ and ProbNN_K variables are used in the selection of decay candidates for this thesis' analysis.

3.3.4 Decay reconstruction in the stripping framework

The tracks with PID information are combined to further reconstruct the full decay process and build the *decay candidate*, *i.e.* a potential signal B meson decay. This stage of the reconstruction is referred to as the *stripping stage*, which combines two tasks, namely (1) reconstructing the full decay process from the individual final-state particle tracks and (2) preselecting the data to make this reconstruction computationally manageable. The corresponding reconstruction algorithms and selection criteria are organised in *stripping lines*, which are specialised for different groups of decay hypotheses. The analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays, *e.g.*, uses the `B2XMuMu` stripping line, which loosely selects and reconstructs decays with B mesons in the initial and two muons and something else ("X") in the final state. The

[6]The full list of input parameters can be found in Ref. [85].

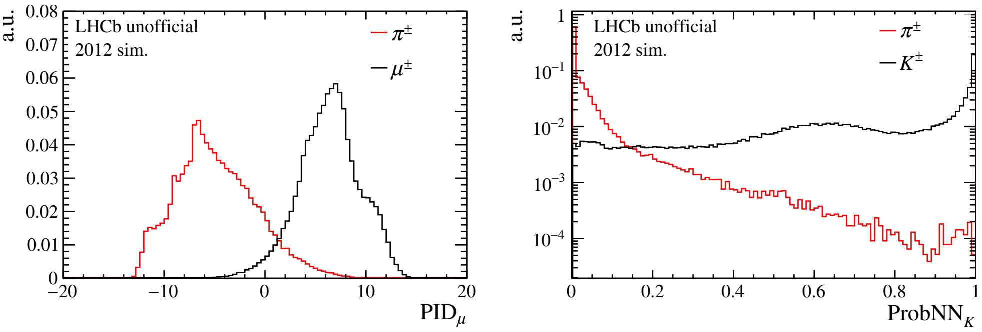

Figure 3.5: Example distribution of the particle identification variables (left) PID$_\mu$ and (right) ProbNN$_K$ for kaons and muons (black), taken from a simulated sample of $B^0_s \to \phi\mu^+\mu^-$ decays, and pions (red), taken from a simulated sample of $B^0 \to K^{*0}\mu^+\mu^-$ decays. The variables separate clearly between the different particle species.

stripping lines are run globally on the collected data, and the selected candidates, which are reconstructed under different decay hypotheses, are saved to files that are accessible for the analysts. As different stripping lines may select similar signatures, these files contain the combined output of the *stripping streams*, which are collections of similar stripping lines. The `B2XMuMu` line, *e.g.*, is part of the *Leptonic stream*, which contains stripping lines selecting decays with leptons in the final state.

A stripping line selects and reconstructs the implemented decay hypothesis step-by-step—this way, decay candidates are rejected early in the process if they are incompatible with the signal, which optimises the use of computational resources. This procedure is explained further in the following, using the reconstruction of $B^0_s \to J/\psi\phi$ decays with the `B2XMuMu` line as an example.

The `B2XMuMu` line contains several decay hypotheses, and the one relevant to select decays via $B^0_s \to \phi\mu^+\mu^-$ is encoded in "$B \to J/\psi\phi$", where B indicates the final state, being either a B^0 or B^0_s meson, and J/ψ and ϕ the bachelor particles.

The stripping line starts by reconstructing the bachelor particles defined by the stripping line. In the above example, the ϕ meson is built from two hadron tracks, and the dimuon candidate ("J/ψ ") is built from two oppositely charged muon tracks. Certain requirements on the tracks themselves are posed to remove incompatible candidates directly at the beginning; examples are requirements on the impact-parameter significance χ^2_{IP} or the track-fit quality.

The track pairs defined above, *i.e.* the two tracks forming the ϕ meson and the two tracks forming the J/ψ, are only accepted if they are consistent with coming from a common vertex that is displaced from the PV. Then, the ϕ and dimuon candidates are combined to reconstruct the initial-state particle, *i.e.* the B meson. This combined candidate is written to file if it passes the last set of selection criteria, which includes a good common-vertex fit quality and other criteria that indicate whether the candidate is compatible with a B meson.[7]

This procedure is performed for all decay hypotheses implemented in the stripping line, and the respective algorithms are optimised to be computationally efficient. This means, *e.g.* in case of the `B2XMuMu` line, that the dimuon candidate

[7]The detailed list of requirements on the tracks and reconstructed particles for this analysis is discussed in Sect. 4.3.1.

is only reconstructed once per event, and all decay hypotheses access this result instead of reconstructing it again.

As explained above, the LHCb stripping lines are run globally on all collected data. This means that if an event in the LHCb data sample is rejected by all stripping lines, it is not directly available for the analysts. So, changes to the stripping lines, such as implementing new decay hypotheses, require a full re-run on the collected LHCb data (referred to as *re-stripping campaigns*), for which computing power is only available during the LHC shutdowns.

This thesis' project was carried out during the second LHC shutdown (LS2) where the LHCb stripping lines were revised, updated, and re-run on the data. In context of designing this analyses' selection, the author of this thesis contributed to the `B2XMuMu` line by (1) improving the selection criteria, and (2) implementing two new decay hypotheses which are potentially interesting to future data analyses; the following paragraphs detail these contributions.

1) Improving the selection criteria for the `B2XMuMu` line The `B2XMuMu` line has been successfully used for a wide range of data analyses, and most of the selection criteria remained unchanged during the years of data-taking. In the context of this work, the selection criteria were optimised for two variables: the (1) impact-parameter significance χ^2_{IP} of the individual muon tracks and the (2) flight-distance significance χ^2_{FD} of the initial-state B meson. Setting both to lower values selected roughly 3% more $B^0 \to K^{*0}\mu^+\mu^-$ decays, and similar improvement is expected for other decays. This change maintained acceptable selection rates and computing times, so it is now implemented in the LHCb stripping framework and used for the latest version of the data samples.

2) Adding the first four-pion final-state decay The other larger change the author contributed to in the `B2XMuMu` stripping line is the selection of two new decays, namely the $B_{(s)} \to f_1(1285)(\to \pi^+\pi^-\pi^+\pi^-)\mu^+\mu^-$ and $B^0 \to \omega(782)(\to \pi^+\pi^-\pi^0)\mu^+\mu^-$ decays. Measuring these decays in combination is of interest as it gives potential access to theoretically clean observables described in Ref. [87].

The modes itself have not been observed yet, only the corresponding J/ψ modes, *i.e.* the $B_{(s)} \to f_1(1285)J/\psi$ and $B^0 \to \omega(782)J/\psi$ modes [88–90]. Their corresponding branching fractions are between roughly 10^{-3} for the $B^0 \to \omega(782)J/\psi$ decay and 10^{-5} for the $B^0 \to f_1(1285)J/\psi$ decay.

Measuring the rare decay modes is not only complicated by the small decay rates but also the structure of the final state: reconstructing a five- or six-track final state is less efficient than a typical four-track final state. How efficient these decays can be selected in the end depends on the level of background pollution and requires future studies; however, the first step towards their measurement is done by implementing them in the stripping framework and making them available for future analyses.

The first step in the implementation is the reconstruction of the intermediate state particles, namely $f_1(1285)$ built from four pions; the $\omega(782) \to \pi^+\pi^-\pi^0$ decay was already included in the `B2XMuMu` stripping line. The newly added reconstruction of $f_1(1285) \to \pi^+\pi^-\pi^+\pi^-$ decays uses the algorithms employed for

$a_1(1260)^+ \to \pi^+\pi^-\pi^+$ decays, and typical selection criteria for an intermediate-state particle of a b hadron decay and its decay products are chosen that result in acceptable data rates.

With this update, four-track final states are now generally accessible, which could also be used for other charged hadron final-states. In the next step, the $\omega(782)$ and the $f_1(1285)$ candidates are combined with the dimuon candidate to build the final B-hadron candidate.

A rough idea on the reconstruction and stripping efficiency is obtained from a simulated sample of $B_s^0 \to (f_1(1285) \to \pi^+\pi^-\pi^+\pi^-)\mu^+\mu^-$ decays, generated taking only the phase space of the decay into account. Roughly 15% of the simulated candidates decay within the LHCb acceptance, and roughly 16% of these are reconstructed with all final-state particles being long tracks. The efficiency combining the acceptance, reconstruction, and stripping selections is roughly 2.8%, which is three times lower than for $B^0 \to K^{*0}\mu^+\mu^-$ decays. When assuming that the further selection steps could be as efficient as for $B^0 \to K^{*0}\mu^+\mu^-$, and using the rough estimate for the branching fraction, $\mathcal{B}(B_s^0 \to f_1(1285)\mu^+\mu^-) \approx 7 \times 10^{-7}$, one expects roughly 50 candidates in the full LHCb data sample after Run 3, so this measurement might be in reach for the near future.

3.4 Software framework

All previously described data-taking stages, starting with the trigger decisions and ending with the files containing the pre-filtered and reconstructed decay candidates ready for analysis, rely on a multitude of software applications, which are implemented in the LHCb software framework. Describing this software framework in great detail would go beyond the scope of this thesis, therefore, this section gives only a brief overview and lists the references to the detailed descriptions.

The core of the LHCb framework is Gaudi [91], which provides the architecture to run the different sub-applications of the LHCb software. The most important sub-applications comprise Moore [92], in which the software trigger algorithms are implemented; Brunel [93], which contains the reconstruction algorithms; and DaVinci [94], which embeds the stripping framework.

In addition to the software needed to process the data, generating simulated samples of the decay processes is crucial for measurements in LHCb, *e.g.* to tune selection criteria or determine selection efficiencies. The LHCb simulation framework is implemented in Gauss [95], and a typical procedure to obtain a simulated decay-sample is the following: in a first step, the inelastic proton-proton collision is simulated with Pythia [96]. Then, the decays of the produced heavy-flavour particles is modelled with EvtGen [97], and final-state radiation is included with Photos [98]. A next stages concerns the particles' interaction with the detector material, simulated with Geant4 [99], and the hardware response, simulated with Boole [100]. From this stage on, candidates are treated similar to data, *i.e.* handed to Moore for the software trigger responses and being reconstructed by Brunel and DaVinci.

3.5 Detector upgrade

To this date, the LHCb detector has recorded data corresponding to a total integrated luminosity of $9\,\mathrm{fb}^{-1}$ during 2011–2012 and 2015–2018. In these years of data-taking, the detector provided excellent performance, eventually exceeding the expectations by allowing analyses that were not even initially anticipated. It was clear from the beginning, however, that the precision of many measurements is limited by the amount of recorded data. Therefore, collecting more data is essential for future measurements carried out at the LHCb experiment.

One current bottleneck in the collection of data is the hardware trigger, which reduces the event rate to 1 MHz at which the current detector can be read out fully. This, however, limits especially the efficiency for hadronic decay modes, so it was already clear in 2011 that this limiting factor needs to be removed for the future.

The upgrade of the LHCb detector that is currently being prepared will measure data with *removed* hardware trigger, relying fully on the software triggers, and new readout electronics which allow a full detector readout-rate of 40MHz—corresponding to the proton bunch-crossing rate of the LHC.

The first upgrade of LHCb, referred to as *Upgrade-I*, is currently being prepared during the second long LHC-shutdown (LS2), which started in 2018 and is planned to end in 2022 with the next data-taking period (*Run 3*), planned for 2022-2023 [101]. In Run 3, the LHC will deliver an increased instantaneous luminosity, and the luminosity inside the LHCb detector will increase roughly by a factor two compared to the previous data taking, $\mathcal{L} = 2 \times 10^{33}\mathrm{cm}^2\mathrm{s}^{-1}$.[8] This enables recording even more data, but it comes at a price: instead of one proton-proton collision per bunch-crossing, roughly five will appear during Run 3. This means that larger track densities inside the LHCb detector need to be disentangled with higher resolution subdetectors, which in addition need to cope with the increased radiation levels.

These new challenges, posed by removing the hardware trigger and running at a higher luminosity, are addressed by an exchange of most of the electronic readout systems and an upgrade to several LHCb subdetector parts: in the tracking system, the VELO silicon strip detectors are being replaced by pixel detectors [103]; the TT is being replaced by the UT; and the straw tubes in the tracking stations T1–T3 are being replaced with new detectors using a completely new tracking technology based on scintillating fibres (SciFi) [104]. In addition to these tracking system changes, the RICH optics and hybrid photomultipliers are being exchanged [105], and the detector parts only used in the hardware trigger stage, *i.e.* the first muon chamber M1, the PS, and the SPD, are removed.

This upgraded LHCb detector is planned to be incrementally improved and further upgraded after the Run 3 data-taking, *i.e.* during the third long shutdown (LS3) from 2024 to 2026. The data-taking for Run 4 starts then in 2026 and is planned to end by 2029. After Run 4, the LHCb detector is expected to have collected approximately $50\,\mathrm{fb}^{-1}$ of data, which is roughly five times larger than the currently collected data sample.

This Upgrade-I LHCb detector serves also as preparation for the *Upgrade-II* detector, which is planned to be set up during the fourth long shutdown (LS4)

[8] The LHC will deliver more than a factor two more luminosity, however, the LHCb detector uses *luminosity levelling* to tune down the luminosity delivered by the LHC [102].

starting in 2030 after the Run 4 data taking [106, 107]. This upgrade will make the LHCb detector a completely new machine: the detector will run at an additional order of magnitude higher instantaneous luminosity, with roughly 50 proton-proton interactions per beam crossing. Maintaining the performance of the LHCb detector in this harsh environment is only possible if timing information is added to disentangle the PVs from each other, which enforces further upgrades of the VELO, the RICH, and the ECAL. A further planned change in the detector design is the removal of the HCAL for the sake of additional shielding in front of the muon chambers to prevent fake muon signals. The inner part of the tracking detectors will be replaced for higher granularity and radiation hardness, and to gain additional information on low momentum tracks, scintillation fibre tracking stations will be added inside the magnet. With this detector, the aim is to collect $300\,\text{fb}^{-1}$ of data, which corresponds to roughly two orders of magnitude more than what the LHCb detector has collected so far—making the term *rare decays* not applicable to the amount of collected $B_s^0 \to \phi\mu^+\mu^-$ decay candidates anymore.

Chapter 4

Branching fraction measurement strategy and preparations

Before presenting the individual analyses of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, this section discusses generally how the decays are measured. As explained in the following section, Sect. 4.1, the measurements can be split into three steps: (1) *selecting* the data, (2) measuring the number of signal decay processes, or *yield*, in the selected data sample, and (3) determining the *efficiency* of the selection.

The first step, selecting the data, is similar for the measurements since both the $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays are measured using the full LHCb data sample, introduced in Sect. 4.2, and comprise the same final-state particles, *i.e.* with $\phi \to K^+K^-$ and $f_2' \to K^+K^-$ decays. Therefore, the data selection is detailed jointly for both modes in Sect. 4.3.

The second step, obtaining the signal yield in the data samples, uses a *maximum likelihood fit* to separate the signal decays from *background decays*, which are non-signal processes present in the selected data sample, and to determine the signal yields. The common setup for the fits is introduced in Sect. 4.4, and details for the two analyses are presented in Sect. 5.3 and Sect. 6.3 for the analysis of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, respectively, as the different background contributions require different fit settings.

The third step, determining the selection efficiencies, is described in Sect. 4.5. This step uses *Monte-Carlo simulated samples* of the decays processes, and to ensure that the simulated samples describe the data well, data-driven corrections are applied. Given the similarities between the modes, the applied corrections are similar and discussed jointly in Sect. 4.6.

4.1 Branching fraction definition and measurement

The branching fraction of a particle decay-process $\mathcal{B}(X \to Y)$ can be defined as the relative decay rate via the specific mode, *i.e.*

$$\mathcal{B}(X \to Y) = \frac{\Gamma(X \to Y)}{\Gamma(X \to \mathit{anything})}. \tag{4.1}$$

with the decay rate of the specific mode $\Gamma(X \to Y)$, normalised to the total decay rate of X, *i.e.* $\Gamma(X \to \mathit{anything})$. In a sample of decaying X particles,

$N(X \to$ anything), the branching fraction could be determined by measuring how many of the X particles decay via an $X \to Y$ decay process, $N(X \to Y)$.

This sample of X particles could, for example, be obtained at the Large Hadron Collider (LHC) if X particles are produced in the proton-proton collisions. Then, the number of unstable particles $N(X \to \textit{anything})$ corresponds to the number of produced particles $N(pp \to X)$, which can be expressed as $N(pp \to X) = \sigma_{pp\to X} \times \mathcal{L}_{\text{int}}$ with the *production cross-section* $\sigma_{pp\to X}$, quantifying the probability to produce X in a proton-proton collision; and the *integrated luminosity* $\mathcal{L}_{\text{int}}$, quantifying the amount of proton-proton collisions.

The number of decay processes $N(X \to Y)$ is not directly accessible as any measurement is affected by detection *efficiencies*. An efficiency quantifies the fact that decays can be overlooked, *e.g.* when particles fly out of the detector acceptance or when decays are rejected because they do not match the data-selection criteria; So, the measured number of decay processes $N(X \to Y)$ needs to be corrected by the *efficiency* $\epsilon(X \to Y)$ to obtain an efficiency-corrected number of decay processes, *i.e.* $N(X \to Y) = N(X \to Y)^{\text{meas.}}/\epsilon(X \to Y)$.

Inserting this and parametrising $N(X \to$ anything) as explained above in Eq. 4.1, the branching fraction $\mathcal{B}(X \to Y)$ is expressed as

$$\mathcal{B}(X \to Y) = \frac{N(X \to Y)^{\text{meas.}}/\epsilon(X \to Y)}{\sigma_{pp\to X} \times \mathcal{L}_{\text{int}}}. \tag{4.2}$$

The integrated luminosity $\mathcal{L}_{\text{int}}$ and the production cross-section $\sigma_{pp\to X}$ are associated with large uncertainties that limit the precision of the measurement. The dependency on these parameters can be removed by measuring the branching fraction of the decay $X \to Y$, referred to *signal mode*, relative to another decay process $X' \to Y'$, referred to as *normalisation mode*. This relative branching fraction, $\frac{\mathcal{B}(X\to Y)}{\mathcal{B}(X'\to Y')}$, is then given by

$$\frac{\mathcal{B}(X \to Y)}{\mathcal{B}(X' \to Y')} = \frac{N(X \to Y)^{\text{meas.}}}{\epsilon(X \to Y) \times \sigma_{pp\to X}} \times \frac{\epsilon(X' \to Y') \times \sigma_{pp\to X'}}{N(X' \to Y')^{\text{meas.}}}. \tag{4.3}$$

This relative branching fraction does not depend on the integrated luminosity anymore, and if the initial-state particles are the same, *i.e.* $X = X'$, the associated production cross-sections cancel as well. The relative branching fraction relies only on the efficiencies and the number of recorded signal and normalisation mode decays.

This method is also advantageous because it reduces systematic uncertainties since many systematic effects are ideally similar for the signal and the normalisation modes—an example are effects on the reconstruction of final-state particles, which can be the same in the signal and normalisation mode; these effects then cancel at leading order in the ratio.

Both branching fraction measurements presented in this thesis make use of this approach. The signal decays, *i.e.* the $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays, are measured relative to the normalisation mode $B^0_s \to J/\psi\phi$ with the intermediate J/ψ meson decaying via the $J/\psi \to \mu^+\mu^-$ mode. These J/ψ-decay modes, referred to as *resonant modes*, have the same final-state particles as the rare rare signal modes, but they are approximately 100 times more abundant than the signal modes as they decay via tree-level decay; a corresponding Feynman

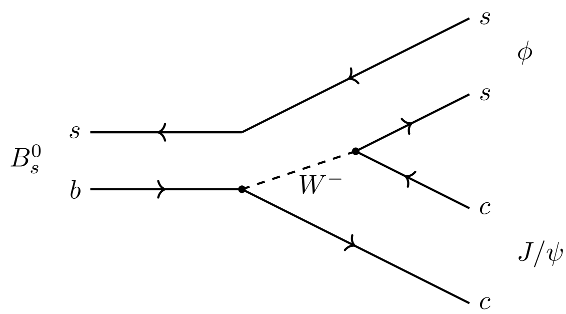

Figure 4.1: Feynman diagram of the normalisation mode $B_s^0 \to J/\psi\phi$.

diagram is shown in Fig. 4.1. The final-state particles in the resonant and rare modes are the same, *i.e.* two oppositely charged kaons and two muons, which further reduces the systematic uncertainties on the measurement: effects associated to the reconstruction of the final-state particles are similar between the modes and therefore cancel at leading order in the ratios.

As discussed in the theoretical discussion in Sect. 2.2.2, the underlying physics as function of the momentum transfer to the dimuon system are of particular interest. The relative branching fraction of $B_s^0 \to \phi\mu^+\mu^-$ decays is measured for different intervals of the invariant dimuon-mass squared, q^2. So, the branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ is therefore measured as

$$\boxed{\frac{\mathrm{d}\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\phi)\mathrm{d}q_i^2} = \frac{\mathcal{B}(J/\psi)}{\Delta q_i^2} \times \frac{N(\phi\mu^+\mu^-)_i}{N(J/\psi\phi)} \times \frac{\epsilon(J/\psi\phi)}{\epsilon(\phi\mu^+\mu^-)_i}} \tag{4.4}$$

with the efficiencies $\epsilon(\phi\mu^+\mu^-)_i$ and $\epsilon(J/\psi\phi)$ and the number of decay candidates $N(\phi\mu^+\mu^-)_i$ and $N(J/\psi\phi)$ of the signal and the normalisation mode, respectively, the q^2-interval width Δq_i^2, and the branching fraction of $J/\psi \to \mu^+\mu^-$ decays, $\mathcal{B}(J/\psi \to \mu^+\mu^-)$. The index i indicates that the variable is determined for the specific q^2 bin under question.

Due to the smaller rate of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, its measurement is limited by the available data sample size, and currently does not allow for a split into different q^2 regions. Therefore, the measurement of this mode is only carried out using the full q^2 spectrum. The relative branching fraction of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ is measured via

$$\boxed{\frac{\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\phi)} = \mathcal{B}(J/\psi) \times \frac{\mathcal{B}(\phi)}{\mathcal{B}(f_2')} \times \frac{N(f_2'\mu^+\mu^-)}{N(J/\psi\phi)} \times \frac{\epsilon(J/\psi\phi)}{\epsilon(f_2'\mu^+\mu^-)}} \tag{4.5}$$

with the efficiencies $\epsilon(f_2'\mu^+\mu^-)$ and $\epsilon(J/\psi\phi)$, and the number of decay processes $N(f_2'\mu^+\mu^-)$ and $N(J/\psi\phi)$ for the signal and normalisation mode, respectively; and the abbreviated branching fractions of the hadrons to kaons, *i.e.* $\mathcal{B}(\phi \to K^+K^-)$ and $\mathcal{B}(f_2' \to K^+K^-)$, and of the J/ψ meson, $\mathcal{B}(J/\psi \to \mu^+\mu^-)$.

A common principle among LHCb analyses is to avoid an experimenter's bias as much as possible. This principle was realised in this analysis by designing and preparing the full analysis procedure without examining any relevant results of the analysis itself, *i.e.* by withholding any information taken from the signal region.

The analysis procedure was developed with the help of *control modes*, which are samples of high-yield decay processes similar to the signal modes. These control modes are used to validate the efficiencies and the extraction of the signal yield from the data. The specific decay modes used in this analysis are the *resonant* J/ψ-partners to the rare signal modes, *i.e.* $B_s^0 \to J/\psi\phi$ decays for the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays, which is also the decay used for normalisation; and $B_s^0 \to J/\psi f_2'(1525)$ decays for the measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. As explained above, the resonant decays provide a precisely known high-statistic sample of decays, which makes them well-suited for designing and validating the analysis.

4.2 Data samples

This work uses data corresponding to an integrated luminosity of approximately $9\,\mathrm{fb}^{-1}$ recorded by the LHCb detector during the years 2011–2018, where the 2011 data corresponding to 1.1 fb^{-1} was taken at proton-beam energies of 3.5 TeV, the 2012 data corresponding to $2.08\,\mathrm{fb}^{-1}$ at 4 TeV, and the 2015–2018 data corresponding to $6\,\mathrm{fb}^{-1}$ at 6.5 TeV. The 2011–2012 data have been previously used to measure $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ in Ref. [2].

The expected improvement in precision gained by adding these data can be estimated by calculating the corresponding number of B_s^0 decays inside the LHCb detector, namely via

$$N_{B_s^0} = f_s \times \mathfrak{L}_{\mathrm{int}} \times \sigma_{pp\to b\bar{b}X}, \tag{4.6}$$

with the hadronisation fraction $f_s = 0.101(8)$ [18], the relevant integrated luminosities $\mathfrak{L}_{\mathrm{int}}$, and production cross-section of $b\bar{b}$-pairs inside the LHCb acceptance at 7 TeV (13 TeV) centre-of-mass energies $\sigma_{pp\to b\bar{b}X} = 72.0 \pm 0.3 \pm 6.8\,\mu\mathrm{b}$ ($144 \pm 1 \pm 21\,\mu\mathrm{b}$) [17]. From this formula, roughly 2.4×10^{10}, 2.8×10^{10} and 5.5×10^{10} B_s^0 mesons are expected in the different data sets. Comparing this to the 2011–2012 sample alone, the number of B_s^0 mesons available in this analysis is increased by about a factor of four, thus reducing the statistical uncertainties by roughly a factor two.

4.3 Signal selection

This work studies *rare decay processes*, and the reason for this label becomes evident with the expected decay rates: for a million b hadrons produced at LHCb, roughly one decays via a specific rare decay mode. So, the data needs to be prefiltered to remove the background decay-processes while keeping as much of the signal as possible. This *selection* of data uses the specific properties of the signal decays that distinguish them from the background decay processes; examples for these properties are typical spatial decay structures, as discussed in Sect. 3.1, invariant masses of the initial or intermediate-state particles, or particle identification (PID) variables.

This section describes the specific criteria used to select the data for the measurement. These criteria were developed based on the previous LHCb measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays [2] and in cooperation with Marcel Materok, whose PhD project was on the angular analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays. The angular analysis uses the same selection criteria as the branching fraction measurement described in this thesis, and it is documented in the internal analysis note in Ref. [108] and published under Ref. [46].

The criteria are applied successively in different *selection stages*, which are the following: the (1) stripping selection, the (2) specific trigger selection, the (3) preselection, the (4) suppression of specific backgrounds, the (5) multi-variate (MVA) classifier selection, and the (5) invariant dimuon mass q^2 division.

The *stripping selection* requires the data to pass the criteria applied in the stripping framework, explained in Sect. 3.3.4. The stripping framework builds the full decay process from the individual tracks, and with many tracks in the detector, the amount of track combinations become large. So, to make this reconstruction step computationally manageable, the amount of data needs to be reduced by applying criteria to remove tracks that are unlikely to originate from the decay process of interest. Section 4.3.1 explains the selection requirements posed by the stripping line used in this analysis, namely the `B2XMuMuLine` line.

The specific *trigger selection* requires the data to be recorded because of a trigger that is connected to the specifics of the signal decay fired. This step ensures a well-defined decay candidate that caused the trigger decision and reduces contribution from processes that are unlikely the signal decays. The corresponding requirements are presented in Sect. 4.3.2.

The *preselection* selects more specifically processes compatible with $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays by posing criteria on the invariant dikaon mass and PID variables, discussed in Sect. 4.3.3. The signal purity in the data sample is further improved by applying *misID vetoes*, which remove contribution from Λ_b^0 and B^0 decays, and an *MVA classifier selection*, which is specifically designed to remove contribution from random track combinations. These selection steps are explained in Sect. 4.3.4 and Sect. 4.3.5, respectively.

The resulting data samples contain both rare $B_s^0 \to \phi\mu^+\mu^-$ or $B_s^0 \to f_2'(1525)\mu^+\mu^-$ and the resonant $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays. To separate these decays, the invariant mass of the two muons is used as explained in Section 4.3.6. Finally, the content of these data samples ready for analysis is visualised and examined in Sect. 4.3.7.

4.3.1 Stripping selection

The stripping line used in this analysis, the `B2XMuMu` line, selects data compatible with the decay signature $B \to X\mu^+\mu^-$, *i.e.* a B meson decaying to two oppositely-charged muons and a dihadron system X in the final-state. The decay hypothesis chosen in this analysis requires the dihadron-system to further decay into two hadrons with opposite charge h^+ and h^-. The comprehensive list of selection criteria are listed in Table 4.1.

The first criteria select decay candidates based on the invariant masses of final-state particle combinations. The invariant mass of the final-state particles combined, $m(B)$, is required to lie between 4900 (4700) and 7000 MeV for the 2011–2012 (2015–2018) data sets, which includes the known B^0 and B_s^0 meson masses. The invariant masses of the dimuon system, $m(\mu^+\mu^-)$, and the mass of the dihadron-system, $m(h^+h^-)$, are restricted towards large masses via $m(\mu^+\mu^-) < 7100\,\text{MeV}$ and $m(h^+h^-) < 6200\,\text{MeV}$, respectively.

Further requirements exploit the candidate's *decay topology*, which is the spacial

Variable	Requirement
$m(B)$	$> 4900^{\text{Run 1}} > 4700^{\text{Run 2}}$ MeV/c^2
$m(B)$	< 7000 MeV/c^2
$m(\mu^+\mu^-)$	< 7100 MeV/c^2
$m(X)$	< 6200 MeV/c^2
$\chi^2_{\text{FD}}(B)$	> 121.0
$\chi^2_{\text{vtx}}(B)/\text{ndof}$	< 8.0
$\chi^2_{\text{vtx}}(X)/\text{ndof}$	< 12.0
$\chi^2_{\text{vtx}}(\mu^+\mu^-)/\text{ndof}$	< 12.0
$\chi^2_{\text{IP}}(B)$	< 16.0
$\chi^2_{\text{IP}}(h^\pm)$	> 6.0
$\max[h^+\chi^2_{\text{IP}},\, h^-\chi^2_{\text{IP}}]$	> 9.0
$\chi^2_{\text{IP}}(\mu^\pm)$	> 9.0
$\cos[\text{DIRA}(B)]$	> 0.9999
$\text{PID}_\mu(\mu^\pm)$	> -3.0
`isMuon`$(\mu^\pm)$	true
`GhostProb`$(\mu^\pm)$	$< 0.35^{\text{Run1}} < 0.5^{\text{Run2}}$
`GhostProb`$(h^\pm)$	$< 0.35^{\text{Run1}} < 0.5^{\text{Run2}}$
HasRICH$(h^\pm)$	true
$\chi^2_{\text{FD}}(X)$	$> 9.0^{\text{Run1}} > 16.0^{\text{Run2}}$
$\chi^2_{\text{FD}}(\mu^+\mu^-)$	> 9.0
DIRA(X)	> -0.9
DIRA$(\mu^+\mu^-)$	> -0.9
hits in SPD	< 600
trigger filters$^{\text{Run 1}}$	(`Hlt1TrackAllL0 | Hlt1TrackMuon`)
	& (`Hlt2Dimuon* | Hlt2Topo* | Hlt2SingleMuon*`)

Table 4.1: Numerical values for the stripping selection used in this analysis to filter the data for $B \to X(\to h^+h^-)\mu^+\mu^-$ candidates.

structure of the final-state particle tracks. A distinct property of B mesons is their significant lifetime. It enables the B meson to fly a few millimetres before it decays, and thus displaces the B meson's decay vertex, the *secondary vertex (SV)*, from its production vertex, the *primary vertex (PV)*. The decay products emerge from the SV, therefore, they do not stem from the PV. Section 3.1 gives more detailed description of this *decay topology* along with the associated variables.

These decay features are utilised in the stripping selection by requiring a large flight distance significance, $\chi^2_{\text{FD}}(B)$, and a small impact-parameter significance for the B hadron candidate, $\chi^2_{\text{IP}}(B)$, but large significances for the hadrons $\chi^2_{\text{IP}}(h^\pm)$ and muons $\chi^2_{\text{IP}}(\mu^\pm)$. Lastly, the cosine of the direction angle DIRA is used, requiring it to be close to one for the B hadron candidates.

Further requirements are placed on the fit qualities of the vertex fits. A good B-decay vertex-fit is required by using the vertex-fit $\chi^2_{\text{vtx}}(X, \mu^+\mu^-)/\text{ndof}$ as well as the fit qualities of the individual dimuon- and dihadron-vertices $\chi^2_{\text{vtx}}(\mu^+, \mu^-)/\text{ndof}$ and $\chi^2_{\text{vtx}}(h^+, h^-)/\text{ndof}$, respectively.

The last set of criteria requires that the particles are correctly identified. For this, PID variables are used, which are explained in detail in Sect. 3.3.3, and requirements are placed on the muons and general tracks. The muons are required to fulfill PID variable requirements posed on PID_μ and `isMuon`, and all charged particle tracks are required to be inconsistent with coming from random detector-hit combinations, also referred to as *ghost tracks*, via the `GhostProb` variable. To allow for a well-performing particle identification in the further selection stages, the hadron tracks are further required to fulfill `HasRICH`, which means that they are measured by the RICH detectors, *i.e.* being in the RICH's acceptance with a sufficient number of hits.

The remaining requirements are applied specifically to reduce the rate of candidates further to improve the computing performance. One example for this is the selection on the flight distance significance of dimuon- and dihadron-system $\chi^2_{\mathrm{FD}}(\mu^+\mu^-)$ and $\chi^2_{\mathrm{FD}}(X)$, respectively, which reduces the number of track combinations before the algorithm builds the b-hadron candidate. Another example is the rejection of *high-multiplicity events*, *i.e.* events with many tracks, that result in large computing times due to the large combinatorics. The high-multiplicity events are identified by a large number of hits in the scintillating pad detector (SPD), and they are removed from the data sample. These events are likely dismissed by the specific selection of software triggers, discussed in Sect. 4.3.2, so no signal is lost at this point; the same argument applies similarly to the trigger filter applied to only the 2011–2012 data.

4.3.2 Specific trigger selection

Signal candidates are only selected for further analysis if they pass the trigger selection; a detailed description of the LHCb trigger system can be found in Sect. 3.3.1. This analysis restricts the data to have a well-defined signal candidate, which means that the trigger decision was caused by the signal candidate.

This specific trigger selection requires positive trigger responses from at least one of the chosen triggers from each trigger stage, *i.e.* the hardware (L0) and the two software (Hlt) stages. The triggers are required to be fired by the signal candidate, so a trigger fired by any other particle in the event that is unrelated to the signal is not accepted.

The different triggers chosen for this analysis are typical choices for analyses of b-hadron decays with muons in the final state, and they are compiled in Table 4.2. Their requirements are discussed in the following.

The hardware triggers (L0) required in this analysis consist of the single muon trigger (`L0Muon`) and the dimuon trigger (`L0DiMuon`). The single muon trigger `L0Muon` is fired by signals in the muon chambers that correspond to a muon with transverse momentum $p_{\mathrm{T}}(\mu^{\pm})$ exceeding a threshold between 1.3–2.8 GeV; the dimuon trigger `L0DiMuon` is fired by two muons with transverse momentum-product $p_{\mathrm{T}}(\mu_1^{\pm}) \times p_{\mathrm{T}}(\mu_2^{\pm})$ exceeding a threshold between 1.7–3.2 GeV^2. The ranges indicate the different settings during the data taking, for which the detailed values can be found in Table B.1 in Appendix B. Both hardware triggers, however, are only fired if the number of signals in the scintillating pad detector (SPD) in the event is below a certain threshold; this requirement is necessary to reject events with high track-multiplicities to reduce the processing times of the further software trigger stages. The processing time could be further reduced for the 2017–2018 data-taking with an additional requirement introduced to `L0Muon`, which prevents

stage	2011–2012	2015–2018
L0	`Muon`	`Muon`
	`DiMuon`	`DiMuon`
Hlt1	`TrackMuon`	`TrackMuon`
	`TrackAllL0`	`TrackMVA`
	-	`TwoTrackMVA`
Hlt2	`Topo2BodyBBDT`	`Topo2Body`
	`Topo3BodyBBDT`	`Topo3Body`
	`Topo4BodyBBDT`	`Topo4Body`
	`TopoMu2BodyBBDT`	`TopoMu2Body`
	`TopoMu3BodyBBDT`	`TopoMu3Body`
	`TopoMu4BodyBBDT`	`TopoMu4Body`
	`DiMuonDetached`	`DiMuonDetached`$^{2017/18}$
	`DiMuonDetachedHeavy`	`DiMuonDetachedHeavy`

Table 4.2: Triggers used in the trigger selection for the hardware (LO) and the two software (Hlt) stages.

firing the trigger if tracks from the previous pp-bunch crossing deposit significant energy transverse to the beam axis $E_{\mathrm{T}}^{\mathrm{prev.}}$ in the tracking system [109].

The first software trigger stage (Hlt1) consists of a muon track trigger (`Hlt1TrackMuon`), a general track trigger (`Hlt1TrackAllL0`/ `Hlt1TrackMVA`), and for 2016–2018, an additional two-track trigger (`Hlt1TwoTrackMVA`). These triggers are fired by individual or pairs of tracks that are displaced from the PV.

The muon track trigger, `Hlt1TrackMuon`, selects well-reconstructed tracks, *i.e.* tracks with small track-fit χ^2/ndof, that are associated with signals in the muon chambers and have a positive response from a hardware muon trigger. In the majority of data, they need to have a minimum momentum $p > 6\,\mathrm{GeV}/c$, a minimum momentum transverse to the beam axis $p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$, a minimum impact parameter (IP), and a minimum impact-parameter significance χ^2_{IP}. In the 2016–2018 data-taking, the tracks are further required to be incompatible with coming from random combinations of detector hits, and this requirement is implemented using the `GhostProb` variable, which combines different information from the tracking system [86]. The detailed numerical values for the `Hlt1TrackMuon` trigger requirements are compiled in Table B.2 in Appendix B.

The general track trigger, `Hlt1TrackAllL0`, used in 2011–2012 data-taking, selects general PV-displaced tracks that are associated with a sufficient number of hits in the tracking system and that are well-reconstructed, *i.e.* have a small track-fit χ^2/ndof. The displacement criterion is fulfilled if the tracks have a minimum IP and χ^2_{IP}, and only tracks with a momentum $p > 10(3)\,\mathrm{GeV}/c$ and transverse momentum to the beam axis $p_{\mathrm{T}} > 1.6(1.7)\,\mathrm{GeV}/c$ fire the trigger in the majority of the 2011 (2012) data-taking.

For 2015–2018 data-taking, the `Hlt1TrackAllL0` trigger has been replaced by the re-optimised general-track trigger `Hlt1TrackMVA`. Similar to the `Hlt1TrackAllL0` trigger used in 2011–2012, it selects general well-reconstructed tracks, *i.e.* tracks with small track-fit χ^2/ndof, that are displaced from the primary vertex and unlikely ghosts. The displacement requirement on χ^2/ndof, however, is now combined with the transverse momentum p_T to a two-dimensional criterion. The details on the criterion can be found in Appendix B along with all numerical values for both general displaced-track triggers, *i.e.* the `Hlt1AllL0` and the `Hlt1TrackMVA` triggers, compiled in Table B.3.

The remaining trigger of the first software stage, `Hlt1TwoTrackMVA`, has been added for 2015–2018 data taking [110], and it selects pairs of tracks forming a common vertex that is separated from any PV. To fire the trigger, the individual tracks are required to be displaced from the PV, have a transverse momentum p_T larger than 500 MeV, have a good track-fit quality, *i.e.* small χ^2/ndof; and are inconsistent with coming from random detector-hit combinations, *i.e.* require a small `GhostProb` value; furthermore, the tracks' common vertex is required to be well-fitted (small χ^2_vtx) and connected to the PV within the pseudorapidity-range covered by the detector. Lastly, requirements are placed on the two-track system, namely having a minimum transverse momentum p_T and a minimum corrected mass m_corr.[1] Candidates that fulfill the above criteria are selected based on a multi-variate (MVA) classifier. The classifier evaluates whether the track pair stems from a b-hadron decay, and it combines the vertex-fit quality, the vertex displacement, the scalar sum of p_T of the two tracks, and the displacement of the individual tracks. The numerical values for the trigger can be found in Appendix B in Table B.4.

The second software stage (Hlt2) comprises two different trigger types: the (1) *topological triggers*, which are fired by combinations of tracks that appear as b-hadron like, and the (2) *detached dimuon triggers*, which are fired by a pair of muon tracks that form a PV-displaced common vertex.

The topological triggers (`Topo{2,3,4}Body(BBDT)`) combine successively two, three, and four tracks with common vertex, and the triggers are fired if the vertices are consistent with typical b-hadron decay properties. The track and vertex requirements are similar to those for the `Hlt1TwoTrack` trigger, and the comprehensive list can be found in Table B.4 in Appendix B. The trigger decision is based on an MVA classifier, which combines the following variables: the track multiplicity $N(\mathrm{tracks})$ with tracks having a significant impact parameter, the corrected mass m_corr, the scalar sum of transverse momenta of the tracks $\sum p_\mathrm{T}$, the vertex-fit χ^2_vtx, the pseudo-rapidity η with respect to the PV, the flight distance significance χ^2_FD, the minimal p_T of the tracks, the impact-parameter significance χ^2_IP, the number of displaced tracks with large momentum $N(\mathrm{displ.\ tracks})$, and the number of $N(\mathrm{tracks,\ small\ }\chi^2_\mathrm{IP})$ with smaller impact-parameter significance χ^2_IP [111]. A further set of topological triggers are specialised to identify B-decay signatures with muon tracks (`TopoMu{2,3,4}Body(BBDT)`). They are similar to the general topological triggers, but they allow to choose a looser MVA selection criterion if at least one of the tracks is positively identified as muon, *i.e.* fulfills the `isMuon` criterion [112].

The detached dimuon triggers (`Hlt2DiMuonDetached(Heavy)`) are fired by pairs of muon tracks that form a common PV-separated vertex with good fit-quality.[2]

---

[1] The corrected mass is defined as $m_\mathrm{corr} = \sqrt{m^2 + |p_\mathrm{T}{}^\mathrm{miss}|^2} + |p_\mathrm{T}{}^\mathrm{miss}|$ with the invariant mass of the two-track system m and the missing momentum $p_\mathrm{T}^\mathrm{miss}$ in flight direction of the two-track system.

[2] For 2015–2016 data-taking, only the `Hlt2DiMuonDetachedHeavy` trigger is used, as the `Hlt2DiMuonDetached` trigger's invariant-mass requirement was mis-configured to be the same

variable	ϕ data set	$f_2'(1525)$ data set
$m(K^+K^-\mu^+\mu^-)$ [MeV/c^2]	$\in [5270, 5700]$	$\in [5270, 5700]$
$m(K^+K^-)$ [MeV/c^2]	$\in [1007.461, 1031.461]$	$\in [1300, 1750]$
$\mathrm{ProbNN}_K(K^\pm)$	> 0.025	> 0.2

Table 4.3: Preselection requirements to select the ϕ and the $f_2'(1525)$ data sets. The wider $f_2'(1525)$ resonance is selected in a larger dikaon mass range, so, the larger background levels are controlled with a stricter requirement on the particle identification variable.

The tracks each are required to be displaced from the PV, have a good track-fit quality, and a minimum transverse momentum. The `Hlt2DiMuonDetachedHeavy` trigger selects only muon-track pairs with a large invariant mass, which in turn enables a looser criterion on the PV-vertex distance. The full list of requirements with the numerical values can be found in Appendix B in Table B.5.

4.3.3 Preselection

The preselection stage comprises requirements tailored to specifically select the decay processes analysed in this work, *i.e.* the $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. Table 4.3 summarises the requirements.

The first selection criteria restrict the invariant masses of particle combinations, *i.e.* the invariant masses of all final-state particles combined $m(K^+K^-\mu^+\mu^-)$ and the invariant mass of the dikaon system $m(K^+K^-)$. The former invariant-mass combination, $m(K^+K^-\mu^+\mu^-)$, is required to lie around the known B_s^0 mass of $m_{B_s^0} = 5366.88(14)\,\mathrm{MeV}$ [89], where the chosen mass range 5270–5700 MeV/c^2 includes the B_s^0-meson peak and high-mass background candidates. The latter, stemming mainly from random-track combinations when excluding the charmonium region, is used to constrain the background contributions in the signal region, as explained in more detail in Sect. 4.4.

To select the $B_s^0 \to \phi\mu^+\mu^-$ or the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, different dikaon mass ranges are used to select the ϕ *data sets* and the $f_2'(1525)$ *data sets*, respectively. For the former, the dikaon mass is selected within a window of 12 MeV/c^2 around the known ϕ meson mass, and for the latter within a window of 225 MeV/c^2 around the known $f_2'(1525)$ meson mass. The ranges are different because the ϕ resonance is a more narrow resonance, *i.e.* with width $\Gamma_\phi = 4.249(13)\,\mathrm{MeV}/c^2$, than the $f_2'(1525)$ resonance with width $\Gamma_{f_2'(1525)} = 86 \pm 5\,\mathrm{MeV}/c^2$ [89], depicted in Fig. 4.2.

as for the `Hlt2DiMuonDetachedHeavy` trigger.

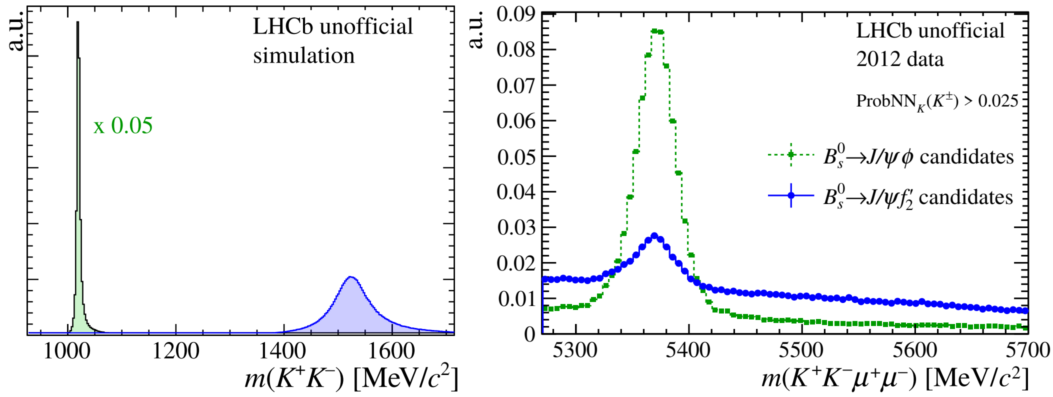

Figure 4.2: Comparison of (left) dikaon mass distributions from simulated $B_s^0 \to \phi\mu^+\mu^-$ (green) and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ (blue) decays and (right) invariant mass distributions from $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ data selected with the PID selection applied for the ϕ data sample. The $B_s^0 \to J/\psi f_2'(1525)$ candidates clearly exhibit larger levels of background contributions due to the larger selected dikaon mass range, which is addressed with a tighter kaon PID requirement.

The last requirement is on the kaon particle identification (PID) variable ProbNN_K, which suppresses contribution from background processes with other particles than kaons. The optimised requirement on the f_2' data set is stricter than for the ϕ selection, *i.e.* in the $f_2'(1525)$ data, the kaons are required to be more *kaon-like* than in the ϕ data. This is necessary as more background processes contribute to the $f_2'(1525)$ sample due to the wider dikaon mass selection. The larger amount of background is clearly visible in Fig. 4.2 where the invariant mass distributions $m(K^+K^-\mu^+\mu^-)$ of resonant $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decay candidates in data are being compared.

4.3.4 Suppressing background from Λ_b^0 and B_s^0 decays

The next selection stage, comprising the *misidentification (misID) vetoes*, removes specific candidates from data that are compatible with stemming from *misidentified* B_s^0 meson or Λ_b^0 baryon decays. Misidentifying a decay process means that one or more of the final-state particle identities are confused in the reconstruction, *e.g.* a proton is falsely identified as a kaon. If these background decay processes are misreconstructed as the signal decays, they can distort the measurement. To prevent this, misID vetoes are applied to remove them from the data.

The misID vetoes designed for this analysis remove two types of decay processes: (1) $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays with the final-state proton being misidentified as kaon ($K^+ \to p$) and (2) $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays with a final-state muon identified as kaon ($\mu^+ \to K^+$) and at the same time a final-state kaon identified as muon ($K^+ \to \mu^+$). The misID veto requirements are listed in Table 4.4 and explained in the following.

Decays via $\Lambda_b^0 \to pK^-\mu^+\mu^-$ mode contribute potentially to the data samples since Λ_b^0 mesons are produced copiously inside the LHCb detector, and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays appear with a similar rate to the signal mode ($\mathcal{B}(\Lambda_b^0 \to pK^-\mu^+\mu^-) \approx 3 \times 10^{-7}$). To suppress them, the Λ_b^0 *veto* is applied, which rejects data candidates if the corresponding final-state kaon is *proton-like*, *i.e.*

decay	veto
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	$\lvert m(K^\pm \to p^\pm, K^\mp, \mu^+, \mu^-) - m_{\Lambda_b^0}\rvert < 50\ \text{MeV}/c^2$ and $\text{ProbNN}_p \times (1-\text{ProbNN}_K)(K^\pm) > 0.35$
$B_s^0 \to J/\psi\phi$ double misID $B^0 \to J/\psi K^{*0}$ double misID	$\lvert m(K^\pm \to \mu^\pm, \mu^\mp) - m_{J/\psi}\rvert < 45.0\ \text{MeV}/c^2$ and ($\texttt{isMuon}(K^\pm)$ or $\text{PID}_\mu(K^\pm) > 5.0$)
$B_s^0 \to \psi(2S)\phi$ double misID	$\lvert m(K^\pm \to \mu^\pm, \mu^\mp) - m_{\psi(2S)}\rvert < 45.0\ \text{MeV}/c^2$ and ($\texttt{isMuon}(K^\pm)$ or $\text{PID}_\mu(K^\pm) > 5.0$)

Table 4.4: Table of background processes and the applied misID vetoes against them.

its PID variables are consistent with the proton-hypothesis, and if the candidate's invariant mass is close to the Λ_b^0 meson mass $m_{\Lambda_b^0} = 5619.60(17)\,\text{MeV}/c^2$ [89] when it is calculated with the *changed particle-hypothesis* $m(K^+ \to p, K^-, \mu^+\mu^-)$. This changed particle-hypothesis is obtained by recalculating the invariant mass with the momenta of the final-state particles p_{K^+}, p_{K^-}, p_{μ^+}, and p_{μ^-}, and the kaon's mass is replaced by the proton's $m_p = 938.272(0)\,\text{MeV}/c^2$ [89].

The decays $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ are normally isolated from the rare mode data samples by rejecting candidates with invariant dimuon mass squared q^2 close to the known J/ψ mass. This requirement, however, can be evaded if two final-state particles are confused in the reconstruction, *i.e.* if one muon is misidentified as a kaon and one kaon as a muon. In this case, the invariant mass of the confused dimuon pair $m(K^\pm \to \mu^\pm, \mu^\mp)$ is not necessarily close to the J/ψ mass anymore. Even though this *double misID* is unlikely, the large abundance of the resonant decays balances them to potentially affect the measurement. To remove these processes, the *double-misID vetoes* are used, which remove candidates with a kaon that is muon-like, *i.e.* whose particle-identification variables are consistent with being a muon, if the reconstructed dimuon mass under the exchanged particle-hypothesis, *i.e.* $m(\mu, K \to \mu)$, is close to the J/ψ mass. Similar vetoes are applied to remove double misID processes including $\psi(2S)$ mesons.

Simulated samples of the signal and the background decays discussed above are used to optimise the vetoes. Figures 4.3 and 4.4 show the PID and invariant mass distributions in those simulated decay samples and visualise the effect of the misID vetoes.

Both vetoes remove less than 1% of the signal decays for both the ϕ and $f_2'(1525)$ data samples. The Λ_b^0 veto removes 45(2) % of the Λ_b^0 decays from the ϕ and 31.4(3) % of the f_2' data sample; the double-misID veto removes 92.8(7)% and 96.6(3)% of the double misID $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays. How the remaining background decays affect the measurements is discussed in the individual analysis chapters, *i.e.* in Sect. 5.2 for the $B_s^0 \to \phi\mu^+\mu^-$ and in Sect. 6.2 for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ analysis.

4.3.5 Suppressing random track combinations

Another source of background decays is of *combinatorial* nature. It comprises candidates that are formed from randomly combined tracks in the event, which by chance fulfill the selection criteria and are selected for analysis. Since these randomly combined tracks can stem from diverse decay processes, there is no single

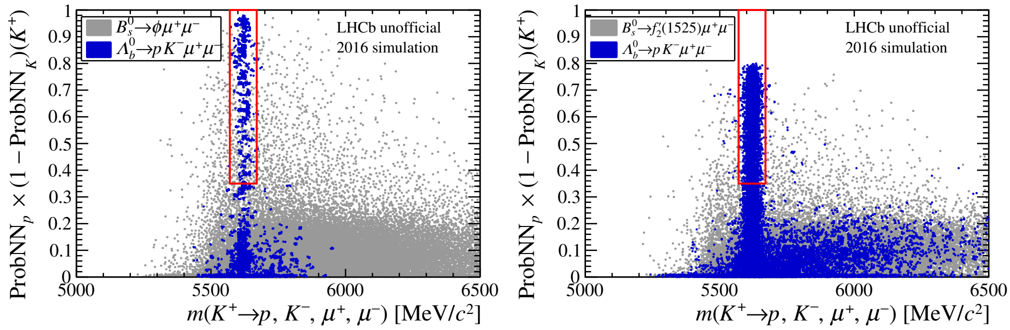

Figure 4.3: Scatter plot of simulated signal (gray) and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ (blue) candidates. The candidates are simulated with 2016 data-taking conditions and have the preselection for the ϕ (left) or $f_2'(1525)$ (right) data sets applied. The distribution of candidates is shown according to the invariant mass under the indicated exchanged particle hypothesis on the x-axis, and the combination of particle identification variables corresponding to the proton-likeness of the kaon on the y-axis. The red box indicates the candidates that are removed by the Λ_b^0-veto.

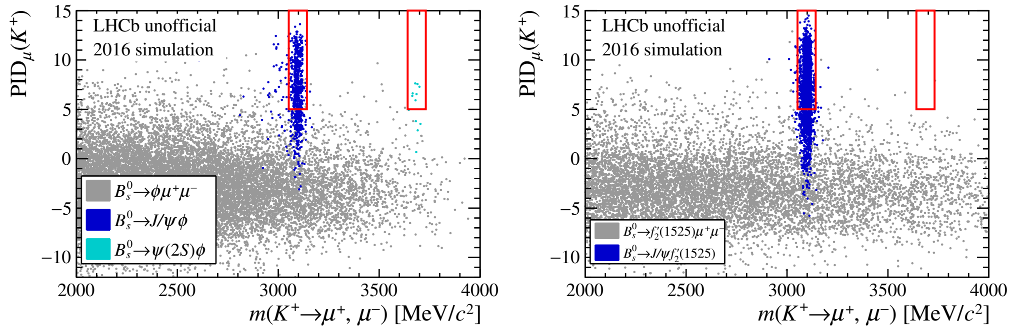

Figure 4.4: Scatter plot of simulated signal (gray) and double misID background (blue, cyan) candidates. The candidates are simulated with 2016 data-taking conditions and have the stripping selection applied. The distribution of candidates is shown according to the invariant mass under the indicated exchanged particle hypothesis on the x-axis, and the combination of particle identification variables corresponding to the proton-likeness of the kaon on the y-axis. The red box indicates the candidates that are removed by the double-misID veto.

distinct feature that can be used to reduce the background type sufficiently.

Therefore, to distinguish the signal from the combinatorial background decays, multiple variables are combined to a *multi-variate (MVA) classifier*. The MVA method in this analysis uses boosted decision trees (BDT) [113, 114] from the TMVA software package [115]. The resulting classifier, the *BDT response*, is then used to further select or reject candidates for the analysis.

The variables that are combined in the BDT reflect the decay topology and kinematics. They comprise (1) PID variables and impact-parameter significances (χ^2_{IP}) of the final-state particles and (2) properties of the formed B_s^0 candidate, *i.e.* the fit quality of the B_s^0-candidate's vertex (χ^2_{vtx}), the transverse momentum p_{T}, impact-parameter and flight-distance significance (χ^2_{FD}) of the B_s^0 candidate, and the angle between flight direction and momentum vector of the B_s^0 (DIRA). The input variables are summarised in Table 4.5.

BDT input variables
χ^2_{vtx} (B^0_s)
$p_{\mathrm{T}}(B^0_s)$
χ^2_{IP} (B^0_s)
χ^2_{FD} (B^0_s)
DIRA(B^0_s)
χ^2_{IP} ($K^\pm$), χ^2_{IP} ($\mu^\pm$)
max[ProbNN$_K(K^+)$,ProbNN$_K(K^-)$]
min[ProbNN$_K(K^+)$,ProbNN$_K(K^-)$]
min[ProbNN$_\mu(\mu^+)$,ProbNN$_\mu(\mu^-)$]

Table 4.5: Variables that are combined in the BDT to remove combinatorial background from the data sets.

The variables are combined to a BDT response by *training* the classifier. In the training of the BDT, the resulting BDT response is optimised to separate the signal and background most efficiently. The training relies on signal and background *proxies*, which are decay samples with typical signal and background candidates. In this work, the signal and background proxies are taken from the preselected ϕ data set:

For the signal proxy, $B^0_s \to J/\psi\phi$ candidates are used. They are selected in an invariant mass of $50\,\mathrm{MeV}/c^2$ around the known B^0_s mass $m_{B^0_s}$ of $5366.88(14)\,\mathrm{MeV}$ [89], and q^2 is required to be around the known J/ψ mass squared, namely $q^2 \in [8, 11]\,\mathrm{GeV}^2/c^4$. The small contamination of combinatorial background is statistically subtracted with the sPlot technique [116].

For the background proxy, candidates with large invariant masses, *i.e.* $m(K^+K^-\mu^+\mu^-) > 5566\,\mathrm{MeV}/c^2$, are used. This high-mass region provides a pure sample of combinatorial background without contamination from low-mass tails from partially reconstructed decays. To increase the sample size for the background proxy, the $m(K^+K^-)$ requirement is loosened, and candidates in a $50\,\mathrm{MeV}/c^2$ window around the known ϕ mass are taken. The regions where the J/ψ and $\psi(2S)$ resonances dominate, *i.e.* $q^2 \in [8, 11]\,\mathrm{GeV}^2/c^4$ and $q^2 \in [12.5, 15]\,\mathrm{GeV}^2/c^4$, are excluded from the background training sample.

The data used for the signal and background proxies are visualised in Fig. 4.5. The BDTs are trained separately for 2011–2012, 2016, 2017, and 2018. For 2015, the data set is too small to train a separate classifier, so for the selection of 2015 data, the 2011–2012 classifier is used.[3]

When training a multi-variate classifier, *overtraining* needs to be avoided. Overtraining denotes the case in which the classifier is tuned on features from statistical fluctuations in the training samples. When such an overtrained classifier is applied to an independent sample without these fluctuation-specific features, its performance is suboptimal. Whether a classifier is affected by overtraining can be checked by comparing the classifier output from the *training sample* with the classifier output applied to an independent sample, which has not been used for training, referred to as *testing sample*. Overtraining of the classifier would appear

[3] Using the 2011–2012 classifier was chosen because the PID variables are similar in 2011–2012 and 2015; the classifier trained on 2016 data resulted in worse performance on 2015 data.

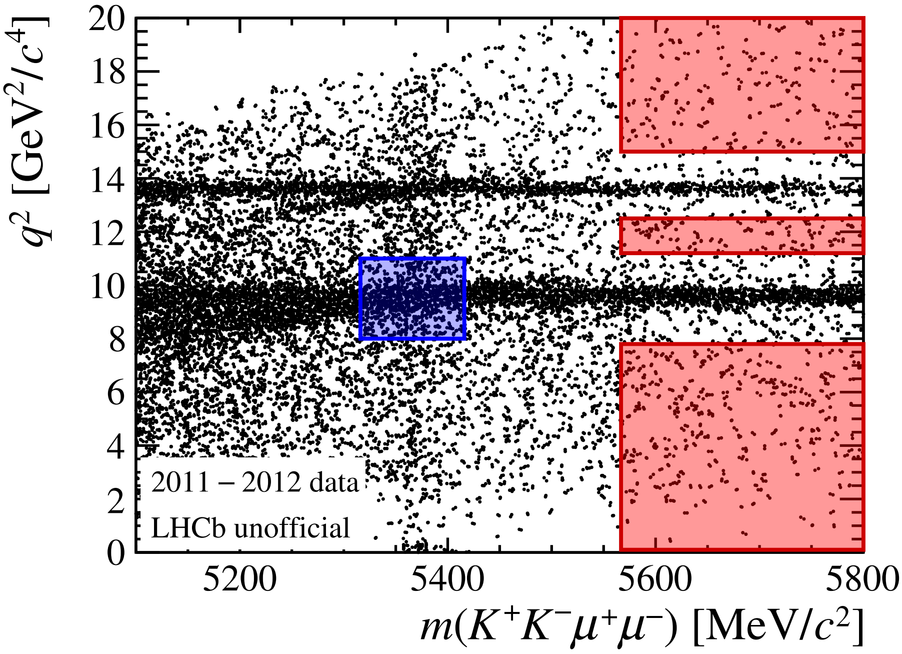

Figure 4.5: Scatter plot of the ϕ data collected during 2011–2012 after the preselection and misID vetoes. As signal proxy for the BDT, background-subtracted $B_s^0 \to J/\psi\phi$ candidates are used (blue box). As background proxy, large invariant-mass candidates are used where the J/ψ and $\psi(2S)$ regions are excluded (red boxes).

as differences between the distributions from the training and the testing samples.

As described later in this section, this testing sample is also used for the *optimisation* of the classifier requirement, *i.e.* finding the optimal selection criterion on the classifier. This, however, could lead to an issue for the measurement: if the selection criteria are optimised on the data used for the measurement, then this could bias the measurement. Therefore, optimising the selection on data directly should be avoided.

This issue is overcome by using a *cross-validation* technique [117]; for this work, the *k-Folding* approach with ten folds was chosen. The method assures that the selection of data candidates is independent from the training or the optimisation of the classifier requirement, and it assures that the training and testing samples are independent from each other. Figure 4.6 illustrates the approach: the data sample is divided into ten independent subsamples.[4] As an example, the first BDT is trained on the eight subsamples (2–9), excluding the first (1), and last (10). This larger fraction to train the data is used to have a large training sample, which also reduces overtraining. This BDT output is tested and optimised on the last sample (10), and used for the selection of data on the first sample (1).

By repeating this method, *i.e.* training, testing, and applying nine other BDTs on permuted combinations of subsamples as shown in Fig. 4.6, no data candidate is selected based on a BDT that was trained or optimised on this candidate.

The resulting outputs of the BDT responses are shown for the testing and the training samples in Fig. 4.7(left) for 2016 data as an example. As explained

[4]Technically, the subsample i is obtained by selecting data with unique candidate identification number N that fulfills $N \mod 10 == i$.

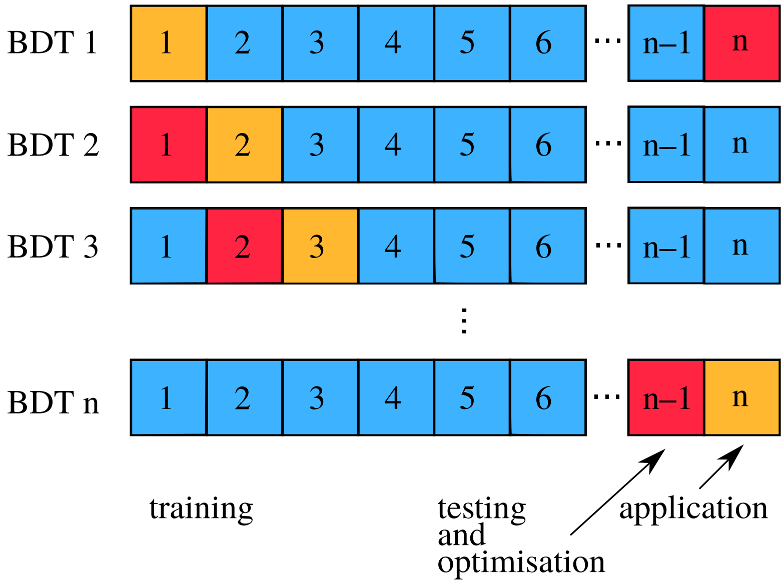

Figure 4.6: Visualisation of k-Folding with n folds. The data is divided into n independent subsamples of equal size, and n BDTs are trained. Each BDT is trained on $n-2$ subsamples of the data (blue), tested and optimised (red) and applied for the final analysis (orange). This way, the training, testing and optimisation and application of the BDT classifier is independent for every data candidate.

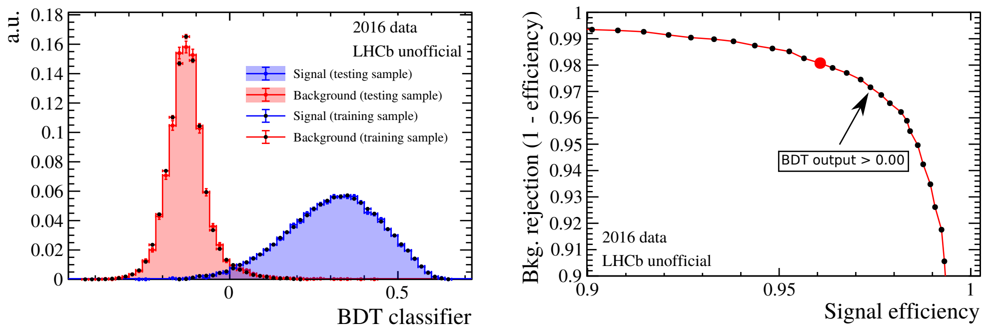

Figure 4.7: Distribution of the (left) BDT classifier output of the background (red) and signal (blue) training and testing samples, and (right) the corresponding ROC curve, which shows the signal efficiency and background rejection for different selection requirements. The BDT requirement chosen for the analysis is marked in red, and the point corresponding to BDT classifier > 0 is indicated as example. Both plots show 2016 data as example, the other years look similar.

above, the BDT outputs are compared between the testing and training samples to check for overtraining. As all distributions are in good agreement, no signs of overtraining are visible in any data set.

Figure 4.7(right) shows the classifier's separation power between the signal and background, visualised in the *receiver operating characteristic* (ROC) curve. This curve results from scanning over different selection requirements on the classifier output, and adding the corresponding signal efficiency and background rejection to the plot for each requirement. The ROC curve shows that the BDT enables to remove the combinatorial background very efficiently, while keeping a large fraction of the signal.

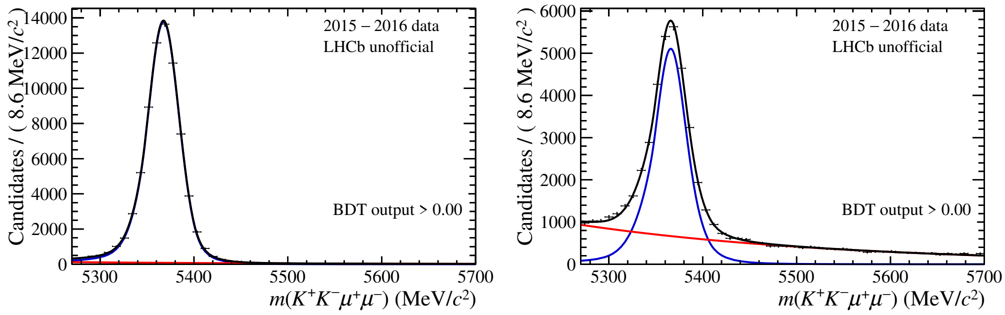

Figure 4.8: Example fits to the (left) $B^0_s \to J/\psi\phi$ and (right) $B^0_s \to J/\psi f_2'(1525)$ candidates after the preselection and after requiring a BDT response > 0. The number of signal candidates extracted from these fits are used to find the optimal BDT requirement for the selection. The 2015–2016 data sets are shown as example.

The BDT classifier requirement that is used to select the data for the measurement is found by *optimising* the requirement on the BDT output. In this work, the optimal BDT requirement is defined as the requirement that maximises the *signal significance* $\sigma_{\text{sig.}}(x) = N_S(x)/\sqrt{N_S(x) + N_B(x)}$, where N_S denotes the number of signal candidates and N_B the number of background candidates in the *signal region* after requiring a BDT response $> x$. The signal region is defined as candidates with invariant mass m that fulfill $|m - m_{B^0_s}| < 50\,\text{MeV}$.

The number of signal candidates N_S after each BDT output requirement x is estimated by

$$N_S(x) = \underbrace{\frac{\mathcal{B}(B^0_s \to J/\psi\phi)}{\mathcal{B}(B^0_s \to \phi\mu^+\mu^-)}}_{=r} \times N(J/\psi), \tag{4.7}$$

with the number of $B^0_s \to J/\psi\phi$ ($B^0_s \to J/\psi f_2'(1525)$) candidates $N(J/\psi)$ and the resonant-rare branching fraction ratio with $r = 0.013$. This branching fraction ratio r is also used to estimate the number of $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays.

The $B^0_s \to J/\psi\phi$ ($B^0_s \to J/\psi f_2'(1525)$) candidates are measured in the J/ψ region $q^2 \in [8, 11]$ of the preselected ϕ ($f_2'(1525)$) data set, and the data is *fitted* to disentangle the signal candidates from the combinatorial background using the invariant mass $m(K^+K^-\mu^+\mu^-)$ distribution. Fitting the data to disentangle different decay processes is described in more detail in Sect. 4.4, and examples for such fits to extract the number of signal candidates is shown in Fig. 4.8.

The number of combinatorial background candidates N_B in the signal region is determined by interpolating the combinatorial background from outside the signal region with invariant masses at least $50\,\text{MeV}/c^2$ away from the B^0_s meson mass $m_{B^0_s}$ to the signal region. For this, an exponential function is fit to the invariant mass distribution of preselected data where the q^2 regions dominated by the J/ψ and $\psi(2S)$ and the signal region have been excluded. This exponential function is then used to calculate the number of candidates that are expected in the signal region. Figure 4.9 shows the data and the fit to interpolate into the signal region after a specific BDT requirement.

Figure 4.10 shows the resulting signal significance as function of the BDT requirement for the 2011–2012 data set as an example. The resulting selection criteria is a BDT response larger than 0.05 (0.2) for the $B^0_s \to \phi\mu^+\mu^-$ ($B^0_s \to f_2'(1525)\mu^+\mu^-$)

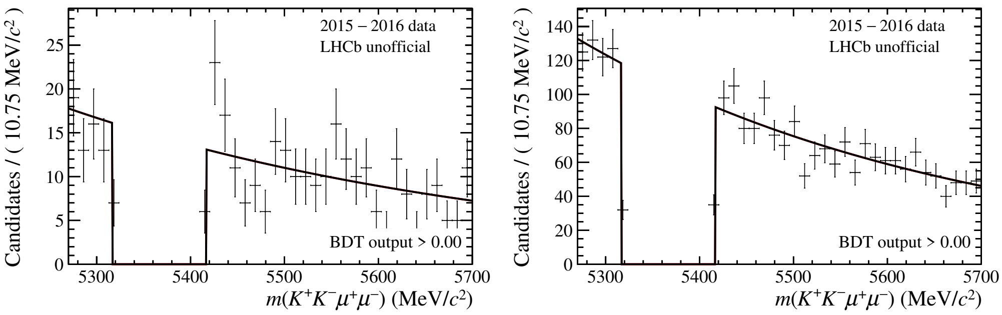

Figure 4.9: Example fits to the upper and lower mass regions of the $B_s^0 \to \phi\mu^+\mu^-$ (left) and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ (right) candidates after the preselection and after requiring a BDT response > 0. The candidates in the signal region are excluded. The number of background candidates are obtained from an extrapolation of the fit function into the signal region, and they are used to find the optimal BDT requirement for the selection. The 2015–2016 data sets are shown as example.

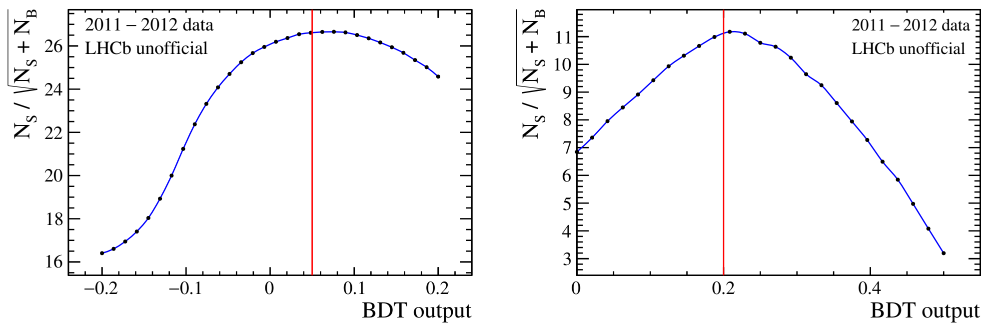

Figure 4.10: Signal significances for different BDT output requirements for the ϕ and $f_2'(1525)$ selection. The red line indicates the BDT requirement chosen for the analysis. These curves are determined on 2011–2012, the curves from other years look similar.

analysis in all data samples, *i.e.* for all data-taking periods.

4.3.6 Invariant dimuon mass squared q^2

The ϕ and the $f_2'(1525)$ data sets contain boththe rare signal $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays and the tree-level $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays. The different modes are separated from each other with the invariant dimuon mass squared q^2, and the tree-level $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays are selected in $q^2 \in [8, 11]\,\mathrm{GeV}^2/c^4$; and the rare mode decays are selected in the full q^2 range that is not dominated by resonances decaying to two muons, *e.g.* $J/\psi \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$ or $\phi \to \mu^+\mu^-$ decays, *i.e.* with $q^2 \notin [0.98, 1.1] \cup [8, 11] \cup [12.5, 15]\,\mathrm{GeV}^2/c^4$. The selection criteria besides those on q^2 are the same for the rare and tree-level modes.

4.3.7 Final data samples

The final ϕ and $f_2'(1525)$ data samples are visualised in Fig. 4.11. These figures show the candidates passing the full selection plotted with their invariant mass

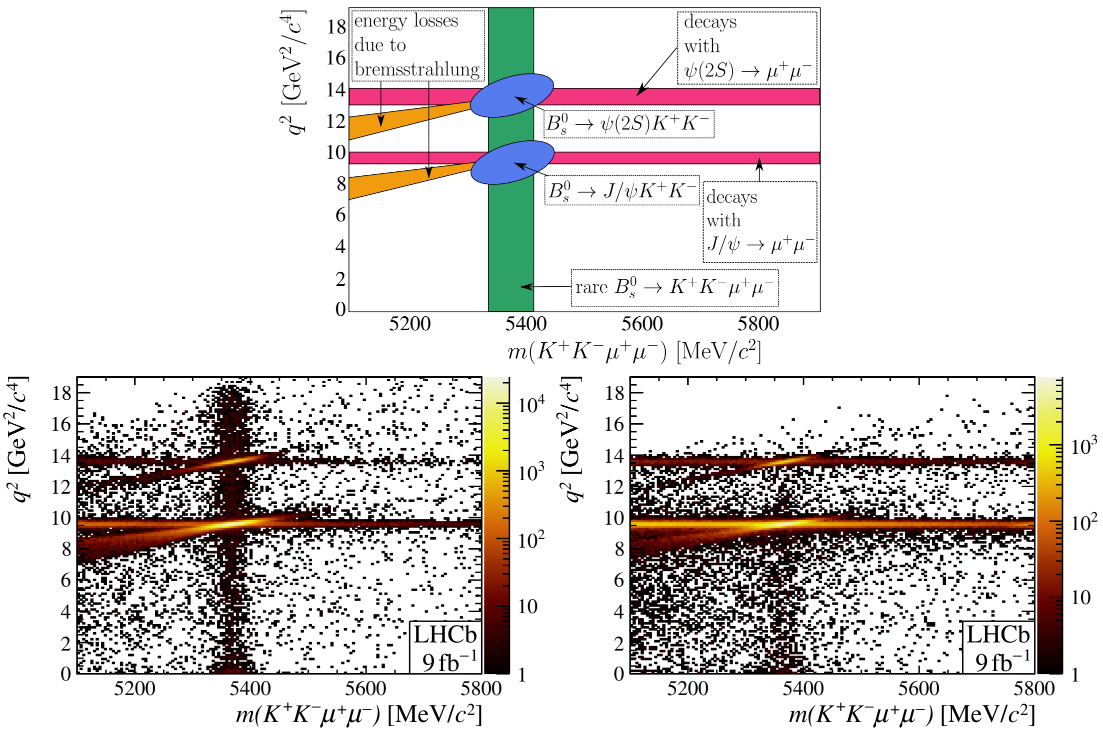

Figure 4.11: Visualisation of the ϕ (bottom left) and $f_2'(1525)$ (bottom right) data sets after the selection. The visible features are labelled in the top plot: The horizontal bands correspond to decays that include $J/\psi \to \mu^+\mu^-$ or $\psi(2S) \to \mu^+\mu^-$ decays, and the accumulation of events at the known B_s^0 mass are from $B_s^0 \to J/\psi K^+K^-$ decays, for which the mass of the dikaon system is compatible with the ϕ or the $f_2'(1525)$ mass. The vertical band at the B_s^0 mass corresponds to the rare $B_s^0 \to K^+K^-\mu^+\mu^-$ decays.

$m(K^+K^-\mu^+\mu^-)$ against their invariant dimuon mass squared q^2. Clear features are visible that correspond to different particle decays contributing to the ϕ and $f_2'(1525)$ data sets: in both data sets, contribution from decays that include charmonium decays, *i.e.* $J/\psi \to \mu^+\mu^-$ and $\psi(2S) \to \mu^+\mu^-$ decays, are clearly visible as accumulation of candidates across a horizontal band at the known J/ψ and $\psi(2S)$ masses. Among these, the $B_s^0 \to J/\psi\phi$ ($B_s^0 \to J/\psi f_2'(1525)$) and $B_s^0 \to \psi(2S)\phi$ ($B_s^0 \to \psi(2S) f_2'(1525)$) decays appear at invariant masses close to the known B_s^0 meson mass in the ϕ ($f_2'(1525)$) data set.

From these regions, diagonal bands of candidates are visible. These diagonal bands correspond to decays with final-state particles that emitted bremsstrahlung, which reduces the invariant masses of the reconstructed B candidate. A small fraction of candidates stretch also toward larger masses; these correspond to decays whose reconstructed masses are smeared by the detector resolution.

Finally, the rare $B_s^0 \to K^+K^-\mu^+\mu^-$ modes are clearly visible as excess of decay candidates that are accumulated at the B_s^0 mass with different values of q^2. As expected from the masses of the initial- and final-state particles, the candidates in the $B_s^0 \to \phi\mu^+\mu^-$ data sample reach a q^2 of up to approximately $19\,\text{GeV}^2/c^4$, whereas the candidates from the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample are limited to roughly $15\,\text{GeV}^2/c^4$.

In addition to these distinct features, background decay candidates with various $\text{m}(K^+K^-\mu^+\mu^-)$ and q^2 values are visible in the plots. The level of background decays is visibly higher in the $f_2'(1525)$ data than in the ϕ data, which is expected from the looser dikaon mass selection for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample. As explained in the beginning of this chapter, the branching fraction measurement relies on determining the number of signal decays in the data sample. How the signal decays are distinguished from the background contribution is explained in the following section.

4.4 Separating signal and background using fits

As shown in the previous section, the final data samples contain not only the signal decays but also contamination from background processes. The branching fraction measurement, however, relies on the number of signal candidates; so, the signal needs to be separated from the background processes. This section describes generally how this is done, namely by *fitting* the data. The details on the fits are presented individually for the two measurements, *i.e.* in Sect. 5.3 and Sect. 6.3.2, as the different background contributions require different fit models.

To separate the signal from the background decays, *discriminating variables* are used, which are differently distributed across the signal and background decays; one example for such a discriminating variable is the invariant mass of a B_s^0 candidate. Since the signal decays have a B_s^0 meson in the initial-state, the corresponding candidates accumulate at the known B_s^0 mass. For background decays without B_s^0 meson in the initial-state, the distribution depends on the specific background type. For combinatorial background contributions as an example, the distribution of invariant masses can be described empirically with an exponential function that decreases toward higher masses. These shape differences are used to separate the different components by using the *unbinned extended maximum likelihood method* implemented in the RooFit framework [118].

The maximum likelihood method is one way for *parameter estimation* from a measurement, *i.e.* to infer the value of a parameter from a data sample [119]: an estimate for the parameter θ can be obtained by maximising the probability $P(x|\theta)$ for the measurements x as function of the parameter θ, which is referred to as the *likelihood* $L(\theta)$ of the parameter θ. If the data consists of independent values x_i that follow the same PDF $f(x_i;\theta)$, then, the likelihood is given by $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$. Maximising this likelihood is done by solving $\partial \ln(L)/\partial\theta_i = 0$; by maximising the logarithm $\ln(L)$ instead of the likelihood L itself,the calculations are simplified by turning the product into a sum.

If a parameter θ affects the number of observed decay processes n, an *extended likelihood* is employed, which includes the distribution of n, *e.g.* a Poisson distribution $P(n;\mu)$ with mean μ. Then, the extended likelihood is given by

$$L(\theta) = e^{-\mu} \times \frac{\mu^n}{n!} \times \prod_{i=1}^{n} f(x_i;\theta). \tag{4.8}$$

One example for such a parameter influencing the number of observed decay processes is the signal fraction f_S in a data sample that is composed of a signal and a background component. In this case, the PDF $f(x; f_S, a, b)$ can be written as

$$f(x; f_S, a, b) = f_S \times f_{sig}(x;a) + (1 - f_S) \times f_{bkg}(x;b) \tag{4.9}$$

with the signal PDF $f_{sig}(x; a)$, the background PDF $f_{bkg}(x; a)$, and their corresponding sets of parameters a and b, respectively. The corresponding likelihood is then given by

$$L(a, b, \mu, f_S) = e^{-\mu} \times \frac{\mu^n}{n!} \times \prod_{i=1}^{n} \left[f_S \times f_{sig}(x_i; a) + (1 - f_S) \times f_{bkg}(x_i; b) \right], \quad (4.10)$$

and can be re-parametrised in terms of the specific signal and background yields $N_{\text{sig}} = f_S \times \mu$ and $N_{\text{bkg}} = (1 - f_S) \times \mu$, respectively, namely by

$$L(a, b, N_{\text{sig}}, N_{\text{bkg}}) = e^{-(N_{\text{sig}} + N_{\text{bkg}})} \times \frac{1}{n!} \quad (4.11)$$

$$\times \prod_{i=1}^{n} \left[N_{\text{sig}} \times f_{sig}(x_i; a) + N_{\text{bkg}} \times f_{bkg}(x_i; b) \right]. \quad (4.12)$$

Minimising this extended likelihood allows extracting the parameters, in this example a and b, and the yields of the different decay processes, N_{sig} and N_{bkg}, contributing to the data.

4.4.1 Fitting the data simultaneously

The ϕ and the $f_2'(1525)$ data sets are split into three data sets, *i.e.* the 2011–2012, the 2015–2016, and the 2017–2018 data set. The branching fraction of the signal, however, is a fundamental constant and therefore the same for the three data samples, so its measured value must be consistent between the data sets. This fact is used to stabilise the fitting procedure: instead of fitting independent PDFs to three data samples to obtain three signal yields, the PDFs are re-parametrised to depend on the branching fraction as fit parameter—then, the re-parametrised PDFs are fit to the three data sets simultaneously, and the branching fraction is shared across the data sets.

This re-parametrisation is based on the branching fraction formula from Sect. 4, Eq. 4.3, and replaces the signal yields in the PDFs for data set i via

$$N(\text{signal})^i = \mathcal{B}(\text{signal}) \times \frac{N(\text{norm.})^i}{\mathcal{B}(\text{norm.})} \times \frac{\epsilon(\text{signal})^i}{\epsilon(\text{norm.})^i}, \quad (4.13)$$

with the signal (normalisation) mode yield $N(\text{signal (norm.)})^i$ and the corresponding efficiencies and branching fractions, $\epsilon(\text{signal (norm.)})^i$ and $\mathcal{B}(\text{signal})$ ($\mathcal{B}(\text{norm.})$), respectively.

This way, the signal branching fraction $\mathcal{B}(\text{signal})$ can be extracted directly from the fits, where all three data sets are directly used. The other parameters, *i.e.* yields of the normalisation mode and the efficiencies, are determined before and used as input parameters for the simultaneous fit: the normalisation mode yields $N(\text{norm.})^i$ in Sect. 5.3.1 and the efficiencies for the analyses of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays in Sect. 5.1 and Sect. 6.1, respectively. The efficiencies are determined on simulated samples of the decay processes, as explained in the following sections.

4.5 Determining the efficiencies using simulation

The measurement of the branching fraction relies on determining the number of signal decay processes in the data samples. This number needs to be corrected for the *signal efficiency*, which quantifies the rate of decay processes that contribute to the selected data samples; signal decays, *e.g.*, can be lost if they decay outside of the LHCb detector or do not fulfill the data selection criteria explained before.

The signal efficiencies are determined on simulated samples of the corresponding decay processes. These simulated samples include all stages from the production of B_s^0 mesons in simulated proton-proton collisions to the corresponding detector responses of the final-state particles; a more detailed description of the simulation framework can be found in Sect. 3.4.

The simulated samples are used to determine the efficiency ϵ for a certain selection requirement via

$$\epsilon_{\text{requirement}} = \frac{N_{\text{after requirement}}}{N_{\text{before requirement}}}, \tag{4.14}$$

where N is the number of decay processes before and after testing whether the requirement under question is fulfilled. These requirements can be the selection criteria, *e.g.* the requirement of the dikaon mass lying within the selection window, or criteria related to the track-reconstruction, *e.g.* whether a decay is successfully identified by the LHCb reconstruction algorithms.

The efficiencies are determined separately for the following selection stages: the detector acceptance (*det.*), the reconstruction and stripping (*reco.+strip.*), the preselection and misID veto (*presel.*), the trigger, and the BDT stage. The efficiency for passing through the stages is determined as *conditional efficiency*, *i.e.* with respect to the previous stage; for example, the trigger efficiency is evaluated on the sample of preselected decay processes. The product of these conditional efficiencies then corresponds to the total efficiency ϵ_{total}, *i.e.* the number of recorded decay processes with respect to the initially produced decay processes, namely

$$\epsilon_{\text{total}} = \frac{N_{\text{passing full selection}}}{N_{\text{simulated decays}}} = \epsilon_{\text{det.}} \times \epsilon_{\text{reco.+strip.}} \times \epsilon_{\text{trigger}} \times \epsilon_{\text{presel.}} \times \epsilon_{\text{BDT}}. \tag{4.15}$$

with the conditional efficiencies listed above. The total efficiency is used as input to the branching fraction measurement. For the measurement of $B_s^0 \to \phi\mu^+\mu^-$, the branching fraction is determined in q^2 intervals; therefore, the efficiencies are determined separately for decay processes with q^2 values of the respective intervals.

The efficiencies are also calculated separately for different data-taking conditions to reflect the different conditions present in data. For this, the simulated samples are generated specifically for the separate years, *i.e.* separately for 2011, 2012, 2015, 2016, 2017, and 2018. The most significant differences between these years are the different centre-of-mass energies in 2011–2012, 2015, and 2016–2018 data, which affect the B_s^0-meson boost and therefore the fraction of decays inside the LHCb detector; and the hardware trigger thresholds, which are varied during the data-taking as detailed in Sect. 4.3.2.

4.6 Simulation corrections

As described in the previous section, the efficiencies are calculated on simulated samples of the decay processes. This approach, therefore, relies on simulated samples that correctly model the decay properties, *e.g.* the kinematics of the decay, and the detector responses, *e.g.* the corresponding PID variables or whether a trigger was fired or not.

Overall, rare decay processes are well described by the simulated samples produced with the LHCb simulation framework described in Sect. 3.4. However, during the past years of analysing LHCb data, some shortcomings of the simulated samples were found; by now, these effects have been studied and common approaches, referred to as *simulation corrections*, have been developed to correct them. These approaches, explained in the following, have become standard practice across analyses of $b \to s\ell^+\ell^-$ decays in LHCb.

Corrections are applied to simulated samples for both analyses, *i.e.* for the $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to f_2'(1525)\mu^+\mu^-$ measurement. The corrections for the former were developed in cooperation with Marcel Materok, who applied those corrections to the angular analysis of $B^0_s \to \phi\mu^+\mu^-$ decays, which is described in the internal analysis note [108] and published under Ref. [46].

Two methods to correct the simulated samples are widely used in LHCb analyses including this one: (1) *generating* new variables that replace the mismodelled variables and (2) assigning *weights* to the simulated decay processes to effectively change the decay-process composition in the samples. Most of the corrections are determined with input from data and therefore referred to as *data-driven corrections*.

The first method, generating new variables, is used to correct the simulated PID variables as detailed in Section 4.6.1. They are corrected based on the PID variables from calibration data samples, which use decay processes that can be cleanly selected in data without any PID requirements [85, 120].

The second method, assigning weights, is employed for all other applied corrections in this thesis. An example for such a correction is the track multiplicity in the simulated decays. The track multiplicity tends to be smaller in the simulated events than what is found in data, *i.e.* the simulated samples contains too few tracks in an event when comparing to data. This tendency is corrected for in simulation by assigning weights to the simulated decay processes that increase the relative weight of those events with large track multiplicities. The weights are specifically determined to align the track-multiplicity distribution between data and simulation.

The general weighting method is detailed in Sect. 4.6.2, and it is used to correct the following distributions: the (1) track multiplicity of the simulated events (Sect. 4.6.3), the (2) transverse momentum of the simulated B^0_s meson (Sect. 4.6.4), the (3) hardware trigger efficiency (Sect. 4.6.5), the (4) B^0_s meson decay time in $B^0_s \to \phi\mu^+\mu^-$ decays (Sect. 4.6.6), and (5) the q^2 distribution of $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays (Sect. 4.6.7).

The first three distributions are corrected with data, namely with the control mode $B^0_s \to J/\psi\phi$ decays. These decays provide a high-yield sample of decays, which have very similar distributions than the signal mode decays, and which can be selected with high purity. The other two distributions are specific to the rare modes and therefore cannot be corrected with the control mode decays; determining them based on data is not possible since the number of signal candidates is very small, so no reasonable correction could be calculated from this sample. Therefore, the distributions are corrected based on SM predictions instead as explained in the respective Sects. 4.6.6 and Sect. 4.6.7.

Finally, after all corrections are applied, the simulated samples are *validated* by comparing distributions from the simulated samples with those from the data samples. The validation is discussed in Sect. 4.6.8.

4.6.1 Particle identification

Particle identification (PID) variables are modelled imperfectly in the simulated samples. As explained in Sect. 3.3.3, the PID variables used to select the data samples combine information from the RICH detectors, the calorimeters, and the muon chambers. For muons, most discrimination power comes from the muon-detector information and for kaons from the RICH detectors.

Simulating the PID information from the RICH detectors is complicated as the RICH detectors are sensitive to experimental conditions, *e.g.* the temperature, the gas pressure, and the detector alignment, which can vary during the data taking [85, 120, 121]. These dependencies hint at why the RICH simulation differs between data and simulation, where more photons than expected are found in data than in the simulation [105]. As a result, the combined PID responses differ between the simulated and the data samples. These differences are corrected for with the LHCb-widely used software package PIDCalib [85, 120], which is used to generate new PID responses based on calibration data.

These calibration data contain samples of the given particle species to study the PID performance on. For each particle species, different decay processes are used that can be selected cleanly from data without imposing PID requirements on the particle type under study; residual backgrounds are removed from the samples using the sPlot technique [116]. Most relevant to this analysis are the calibration samples containing kaons and muons since they are used to correct the signal simulation.

The muon samples are from $J/\psi \to \mu^+\mu^-$ decays. These decays are selected from events with one well-identified muon of certain charge that forms a good-quality vertex with another track of opposite charge, where the invariant mass of the particle-combination is required to be close to the known J/ψ mass. These other tracks that have no PID requirements applied are taken as calibration sample for the muons.

The kaon samples are taken from $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$ decays, for which the soft pion produced together with the D^0 defines the charge of the kaon, *e.g.* if the soft pion has positive charge, the track with negative charge from the D^0 decay corresponds to the kaon. This allows to identify the kaons from the D^0 decays without requirements on their PID responses.

From these data samples, PDFs are extracted that describe the distribution of the PID variable under study. These PDFs are a function of the transverse track momentum p_{T}, its pseudo-rapidity η and the track multiplicity in the event N_{tracks}, and they are modelled using kernel density estimations that have been modified to improve the description of boundary effects and narrow structures [122].

Both the PDF extraction and the generation of corrected PID variables is implemented in the PIDCalib software package [85, 120]. Two methods are provided for the latter: the *resampling* and the *transformation* methods. The resampling method randomly samples for each candidate a PID value that follows the PID probability distribution $\mathrm{PDF}(\mathrm{PID}|p_{\mathrm{T}}, \eta, N_{\mathrm{tracks}})$, extracted from the calibration sample for a given transverse momentum p_{T} and a pseudo-rapidity η of the simulated particle and the track multiplicity N_{tracks}. This way, the corrected PID response is effectively taken from data, and the correlations between the PID variables and the kinematics, *i.e.* p_{T}, η, and N_{tracks}, are preserved. It should be noted that the PID correction therefore relies on correctly modelled kinematics; *i.e.* if the signal simulation is affected by mismodelled kinematics, they are likely also affected by mismodelled PID responses, and the kinematic variables need to be corrected to ensure well modelled PID responses.

A disadvantage of this method is that it does not preserve correlations with other variables; this shortcoming is addressed in the transformation method. Instead of sampling new variables according to the PDF extracted from the calibration samples, the transformation method uses the PID response from the simulated particle and transforms it to the corrected PID response. This transformation is designed such that the cumulative distribution functions of the PID variables $P(\mathrm{PID})$ are the same between the calibration sample and the simulation, and the corrected PID variable $\mathrm{PID}_{\mathrm{corr}}$ is given by [85]

$$\mathrm{PID}_{\mathrm{corr}} = P^{-1}_{\mathrm{calibration}} \left(P_{\mathrm{simulation}}(\mathrm{PID}_{\mathrm{sim.}}|p_{\mathrm{T}}, \eta, N_{\mathrm{tracks}})|p_{\mathrm{T}}, \eta, N_{\mathrm{tracks}}) \right) . \tag{4.16}$$

For given kinematics p_{T}, η and N_{tracks}, the corrected PID value is the value that has the same cumulative probability as the corresponding simulated PID value. In other words: if a track under study has a simulated PID value x, and the probability to get this value or smaller is P' in the simulated sample; then the corrected PID value x_{corr} is the value that corresponds to this cumulative probability P' in the calibration sample. So, a simulated candidate with a PID variable that appears rarely in the simulated sample is corrected with a PID variable value that appears rarely in the calibration sample.

This way, the correlations between the PID variables of the same track are preserved by taking them effectively from the simulated sample. Preserving the correlations is especially important if combinations of PID variables are used for the selection, *e.g.* in the misID vetoes discussed in Sect. 4.3.4, which makes the transformation method the preferred one to correct the PID variables.

4.6.2 Corrections using weights

The other general method to correct the simulated samples is by assigning weights to the simulated candidates. The weights are used to adjust the composition of decay candidates, *e.g.* if a simulated sample lacks candidates with high momentum tracks, this sample can be corrected by assigning larger weights to the underrepresented candidates with high momentum, and lower weights to the overabundant candidates with lower momentum.

Correction by assigning weights can be advantageous since it not only corrects a distribution under question itself, but it also corrects distributions that are correlated with this distribution. This effect is especially visible when correcting

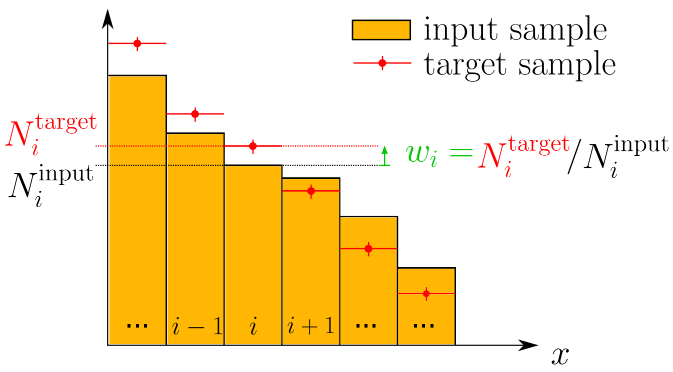

Figure 4.12: Visualisation of correction weight extraction. An input sample is assigned weights w_i such that its content N_i^{input} is aligned to that from the target distribution, N_i^{target} for each histogram bin i.

the track multiplicity of an event, which affects the PID distributions—this is expected as the PID performance depends on the track occupancy in the RICH detectors.

The correction weights are extracted directly from the comparison of two histograms: the *input histogram*, which corresponds to the distribution of a variable of the mismodelled simulated sample, and the *target histogram*, which corresponds to the desired distribution. The weights are then designed such that the weighted input histogram exactly aligns with the target histogram, and Fig. 4.12 visualises the procedure: for each histogram bin i, the weight w_i is defined as

$$w^i = \frac{N_i^{\text{target}}}{N_i^{\text{input}}}, \tag{4.17}$$

with the number of candidates in the ith histogram bin of the input and target histogram, N_i^{target} and N_i^{input}, respectively.

The binning scheme of the histograms is often optimised such that the bins are *isopopulated*, which means that the number of candidates contributing to each bin is approximately the same. This avoids weights with large statistical uncertainties from sparsely populated histogram bins.

These weights are used to correct simulated samples by applying them to each decay candidate in the simulated sample. Each candidate j is assigned the weight w_i^j that corresponds to the weight correcting the amount of simulated processes falling in the histogram bin j. These per-event weights are used whenever candidates are counted, most importantly in the efficiency calculation, where the number of candidates is then replaced by their sum of weights, *i.e.*

$$\epsilon_{\text{requirement}} = \frac{N_{\text{after requirement}}}{N_{\text{before requirement}}} \rightarrow \frac{\sum\limits_{j,\text{ after requirement}} w^j}{\sum\limits_{j,\text{ before requirement}} w^j} \tag{4.18}$$

The weights are only defined up to a global factor f_n, *i.e.* Eq. 4.18 would give the same result with weights $w_j' = f_n \times w_j$. For convenience, the normalisation factor f_n is chosen to fulfill the following condition: the sum of weights is required to be equal to the number of candidates at the point of application, *i.e.* after the selection requirement where the weights are applied first, namely

$$\sum_{j,\text{ after requirement}} w_j = N_{\text{after requirement}}. \tag{4.19}$$

It should be noted, however, that selection stage where the weights are determined is not necessarily the stage where the weights are applied; *e.g.* as explained later in Sect. 4.6.4, the weights to correct the transverse momentum of the B_s^0 are determined after the reconstruction, but applied to the simulation directly after generation from the physics model.

This correction-weight approach can be extended to correct correlated variables simultaneously by using multi-dimensional histograms. Instead of comparing the bin contents between two 1D-histograms, the bin contents between two multi-dimensional histograms are compared. These weights are then applied and normalised similar as described above for the 1D case. The only technical difficulty is the number of candidates needed to fill these multi-dimensional histograms as the number of bins grows quickly with more dimensions.

The specific correction weights used in this analysis are discussed in the following sections, and an overview on them is shown in Fig. 4.13.

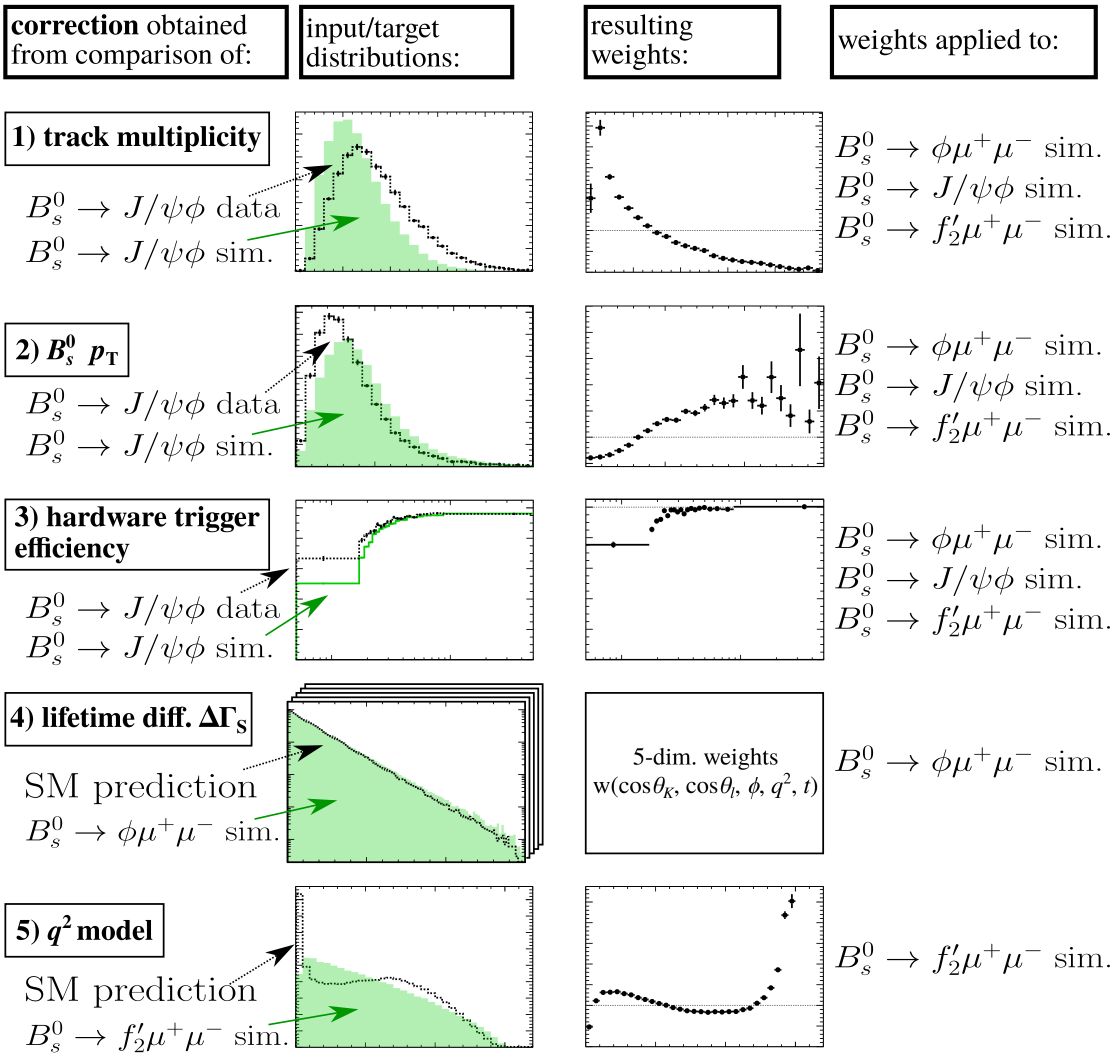

Figure 4.13: Overview on the different correction weights applied to the simulated samples. The (1) track multiplicity, the (2) transverse momentum of the B_s^0 meson, and the (3) hardware trigger efficiency is corrected in all simulated signal samples with weights obtained from comparing $B_s^0 \to J/\psi\phi$ data and simulation. The (4) lifetime difference of the B_s^0 meson $\Delta\Gamma_s$ is corrected for $B_s^0 \to \phi\mu^+\mu^-$ simulation with the 5-dimensional decay rate taken from the SM prediction. Finally, the (5) q^2 model is improved for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation samples using an SM prediction.

4.6.3 Track multiplicity

In simulation, the track multiplicities tend to be smaller than what is observed in data. The track multiplicity requirements used in the stripping selection are very loose, so there, a mismodelling would have only small effect on the efficiency evaluated on the simulated sample. However, the track multiplicity is, *e.g.*, strongly connected to the RICH performance, and therefore with the PID responses, which are used in the selection. Correcting the track multiplicity before applying the PID selection is therefore crucial to obtain the correct selection efficiencies.

The track multiplicity weights are derived by comparing the distributions of track multiplicities in simulation to the distribution in data, following the procedure explained in the previous section. For this comparison, the track multiplicities of

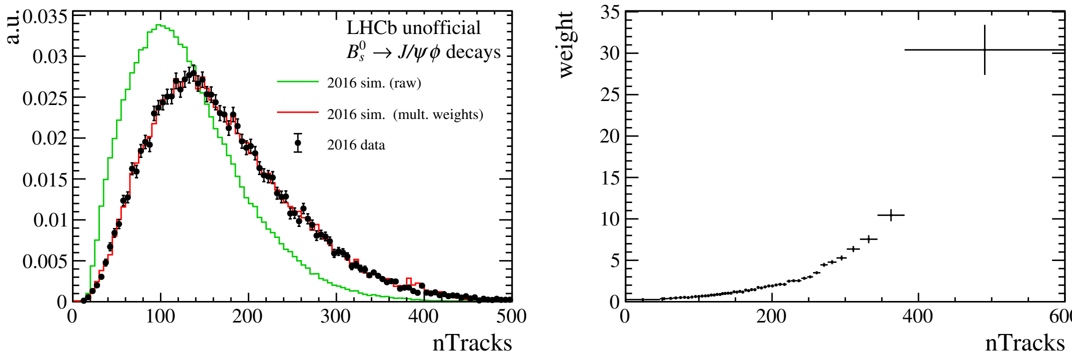

Figure 4.14: Plot of the (left) normalised multiplicity distribution in $B_s^0 \to J/\psi\phi$ statistically background-subtracted ("sPlotted") data (black markers) and simulation, for which the uncorrected (green) and corrected (red) simulation is shown, and (right) the correction weights depending on the event multiplicity value of a decay candidate.

$B_s^0 \to J/\psi\phi$ candidates are used, where both the data and the simulated candidates are required to pass the preselection without PID requirements, *i.e.* without requirement on kaon PID. The background in the data sample is removed with the sPlot technique [116] and by using only the signal mass region, *i.e.* $50\,\mathrm{MeV}/c^2$ around the known B_s^0 mass.

The weights are extracted with isopopulated 50-bin histograms containing the multiplicities from 0 to 500. The weights are determined separately for each year since the track multiplicities are dependent on the beam conditions, which vary during data-taking. However, since the 2015 data sample is too small to derive statistically solid corrections, the weights to correct the 2016 simulation are used. Similar to all other corrections, this choice is validated as explained in Section 4.6.8.

The distributions to extract the weights are shown in Fig. 4.14 together with the corresponding correction weights for 2016 data-taking conditions as an example. These weights, derived with $B_s^0 \to J/\psi\phi$ decays, are used to correct all simulated signal, normalisation and control mode samples, *i.e.* $B_s^0 \to \phi\mu^+\mu^-$, $B_s^0 \to f_2'(1525)\mu^+\mu^-$, $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays. This approach relies on similar track multiplicity distributions in all decay modes, which has been checked on simulated samples as shown in Fig. 4.15 as an example for the comparison on simulation with 2012 data-taking conditions.

4.6.4 Transverse momentum of the B_s^0

The production kinematics of the B_s^0 meson are not described perfectly in the simulated samples, so the transverse momenta p_T of the B_s^0 mesons $p_\mathrm{T}(B_s^0)$ tend to be underestimated in the simulation when comparing it to the distribution in data. The $p_\mathrm{T}(B_s^0)$ variable is not directly used as selection requirement, however, it is used as input parameter of the BDT classifier. So, a good modelling of $p_\mathrm{T}(B_s^0)$ is needed for a good modelling of the BDT response, which is used as selection requirement.

Similar to the multiplicity weights, the B_s^0 p_T weights are determined by comparing the B_s^0 p_T distribution in $B_s^0 \to J/\psi\phi$ data with that in $B_s^0 \to J/\psi\phi$ simulation, and the procedure to extract weights is described in Section 4.6.2. The

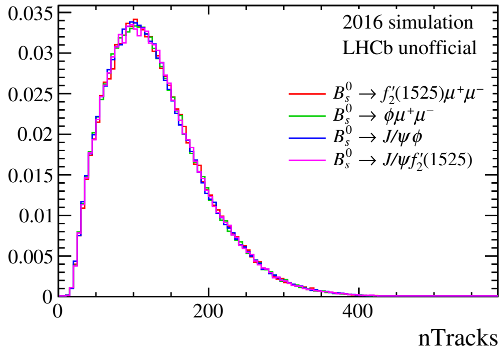

Figure 4.15: Comparison between the multiplicity distributions from $B_s^0 \to J/\psi\phi$ (blue), $B_s^0 \to J/\psi f_2'(1525)$ (magenta), $B_s^0 \to \phi\mu^+\mu^-$ (green) and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ (red) simulation with 2016 conditions. The distributions are very similar for all decay processes.

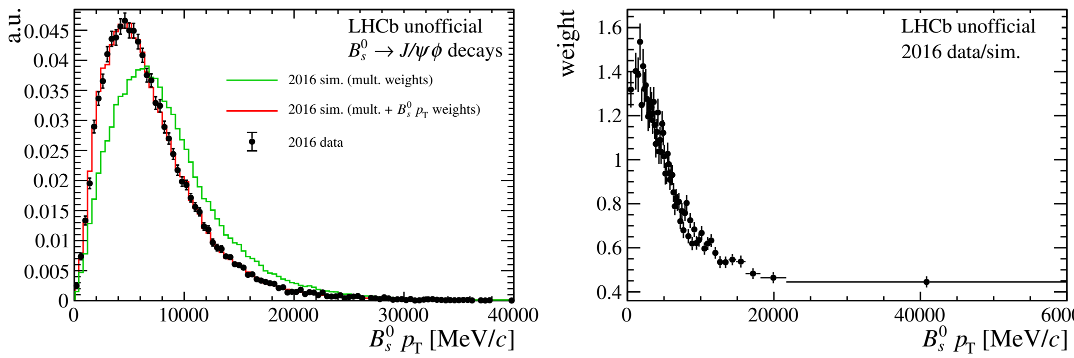

Figure 4.16: Distribution of (left) reconstructed B_s^0 p_{T} in $B_s^0 \to J/\psi\phi$ simulation with multiplicity weights (red), $B_s^0 \to J/\psi\phi$ simulation with multiplicity and B_s^0 p_{T} weights and (red line) and $B_s^0 \to J/\psi\phi$ data (black markers) for the 2016 samples, and (right) the B_s^0 p_{T} weights extracted from the comparison of $B_s^0 \to J/\psi\phi$ data and simulation. The weights are with the corresponding true B_s^0 p_{T} value since the simulated samples are weighted according this true B_s^0 p_{T} value to enable the correction already before the reconstruction.

samples to fill the histograms have the preselection applied without kaon PID and BDT response requirement and are corrected for the track multiplicity distribution as explained before. In data, the background has been removed with the sPlot technique and invariant masses within 50 MeV/c^2 around the known B_s^0 mass. The simulated sample has the multiplicity weights applied.

The B_s^0 p_{T} histograms to extract the weights cover a B_s^0 p_{T} range from 0 to 60000 MeV and contain 50 intervals, and the weights are determined separately for the different years to capture the different underlying B_s^0 p_{T} distributions caused by different centre-of-mass energies. The 2015 sample is corrected using the correction derived on 2016 samples since the 2015 sample is too small to derive a statistically solid correction by itself. The distributions to extract the weights are shown together with the weights themselves in Fig. 4.16 for the 2016 conditions as example.

The B_s^0 p_{T} weights, extracted from $B_s^0 \to J/\psi\phi$ candidates selected in analogy to what is done for the multiplicity weights discussed before, are applied to correct

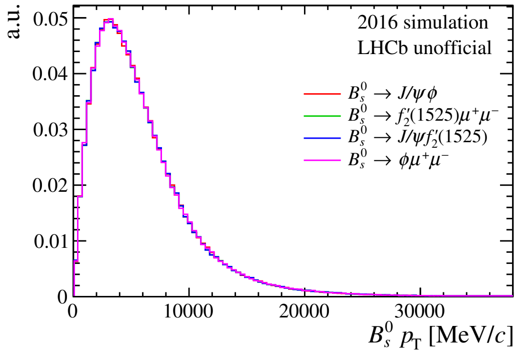

Figure 4.17: Comparison of B_s^0 p_{T} without reconstruction effects for the different decay modes in 2016 simulated sample before any selection criteria are applied.

the B_s^0 production kinematics, so they need to be corrected before any selection requirement. This also means that the B_s^0 p_{T} weights are applied before the reconstruction algorithms, so the correction is applied to the *generated* p_{T} of the simulated B_s^0 meson instead of the *reconstructed* B_s^0 p_{T} value.

The B_s^0 p_{T} weights, determined on $B_s^0 \to J/\psi\phi$ decays, are used to correct all other simulated B_s^0-meson samples. This can be done since the B_s^0 momentum distribution is independent of the decay process, it only depends on the proton-proton collision from which the B_s^0 meson is produced. This is validated by comparing the simulated distributions from the different decay modes. An example for this is shown in Fig. 4.17 for the 2016 simulations.

4.6.5 Hardware trigger

As explained in detail in Sect. 4.3.2, the trigger settings are varied during the data-taking. In contrast to data, the simulated samples are generated with only one specific trigger setting. The trigger settings, set in a trigger configuration key (TCK), directly affect the trigger efficiencies, so differences in the settings leads to differences in the efficiencies. Since the efficiencies are fully determined on simulation, it is crucial that the simulated trigger efficiencies reproduce the trigger efficiencies in data.

The trigger efficiencies, *i.e.* the fraction of decay processes that fire a trigger, cannot be determined *directly* on data since no sample of data without any trigger decision is available—any recorded data event requires the positive response of at least one trigger per trigger stage, *i.e.* hardware, first, and second software stage. The trigger efficiencies, however, can be extracted from data using the *TISTOS method* [123]. The TISTOS method allows to measure the TISTOS trigger efficiencies, which are defined as the efficiency for an event to be *triggered on the signal (TOS)* evaluated on events that are *triggered independently of the signal (TIS)*, *i.e.* which were recorded due to a trigger fired by the rest of the event. Therefore, the TISTOS efficiency $\epsilon_{\mathrm{TISTOS}}$ is a conditional efficiency $\epsilon_{\mathrm{TOS|TIS}}$, and it can be calculated as

$$\epsilon_{\mathrm{TISTOS}} = \epsilon_{\mathrm{TOS|TIS}} = \frac{N_{\mathrm{TIS\ and\ TOS}}}{N_{\mathrm{TIS}}}, \tag{4.20}$$

with the number of decay processes that are triggered independently of the signal

N_{TIS} and the decays that fire the trigger with both the signal and the rest of the event $N_{\mathrm{TIS\ and\ TOS}}$.

The trigger settings are varied during the data taking mostly for the hardware trigger stage, so the data-simulation differences are largest when comparing the hardware trigger efficiencies. A detailed list of the different thresholds can be found in Appendix B. The software trigger settings remain mostly unchanged during the data taking, and, the software trigger efficiencies are in good agreement between data and simulation. Therefore, only hardware trigger efficiency is corrected in the simulation as explained in the following.

Similar to other simulation corrections discussed before, weights are determined that align the hardware trigger efficiency in simulation to that found in data. For this, the TISTOS trigger efficiencies are evaluated on data and simulation $B_s^0 \to J/\psi\phi$ candidates that are preselected and from the B_s^0 mass signal region, *i.e.* with invariant masses $m(K^+K^-\mu^+\mu^-)$ in a $50\,\mathrm{MeV}/c^2$ window around the known B_s^0 mass. The latter requirement is added to provide a pure data sample with negligible background.

The hardware trigger TISTOS efficiency is compared between data and simulation in intervals of the maximum of the two muons' p_{T}, *i.e.* $\epsilon_{\mathrm{TISTOS}}(\max[\mu_1(p_{\mathrm{T}}), \mu_2(p_{\mathrm{T}})])$. Firstly, this variable is chosen since it reflects the underlying kinematic variable most responsible for the trigger decision: the maximum of the two muons' p_{T} is directly linked to the momentum of the particle that most likely fired the single muon trigger, and it is strongly correlated to the product of the momenta of the two muons. Comparing the trigger efficiencies between data and simulation as function of $\max[\mu_1(p_{\mathrm{T}}), \mu_2(p_{\mathrm{T}})])$ allows to directly assess the effect of the different momentum threshold settings.

Secondly, correcting the efficiency in bins of $\max[\mu_1(p_{\mathrm{T}}), \mu_2(p_{\mathrm{T}})]$ removes the dependency on the underlying momentum distribution and therefore enables to port the correction weights to other decay modes with different muon momentum distribution. This, in particular, allows to port the correction weights determined on $B_s^0 \to J/\psi\phi$ decays to the simulated samples of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays.

Figure 4.18 compares the TISTOS efficiencies between $B_s^0 \to J/\psi\phi$ data and simulation for the different years of data-taking before the hardware correction weights are assigned. The differences between data and simulation behave as expected from the different trigger thresholds between data and simulation; *e.g.* in 2016 (2017), the uncorrected simulation underestimates (overestimates) the trigger efficiencies at low muon momenta close to the trigger thresholds. The fact that these differences between data and simulation are indeed caused by the different trigger settings can be seen when repeating the comparison using data that is restricted to the TCK used in the simulated sample—then, the differences vanish and the efficiencies are in good agreement, as shown for 2016 as an example in Fig. 4.19.

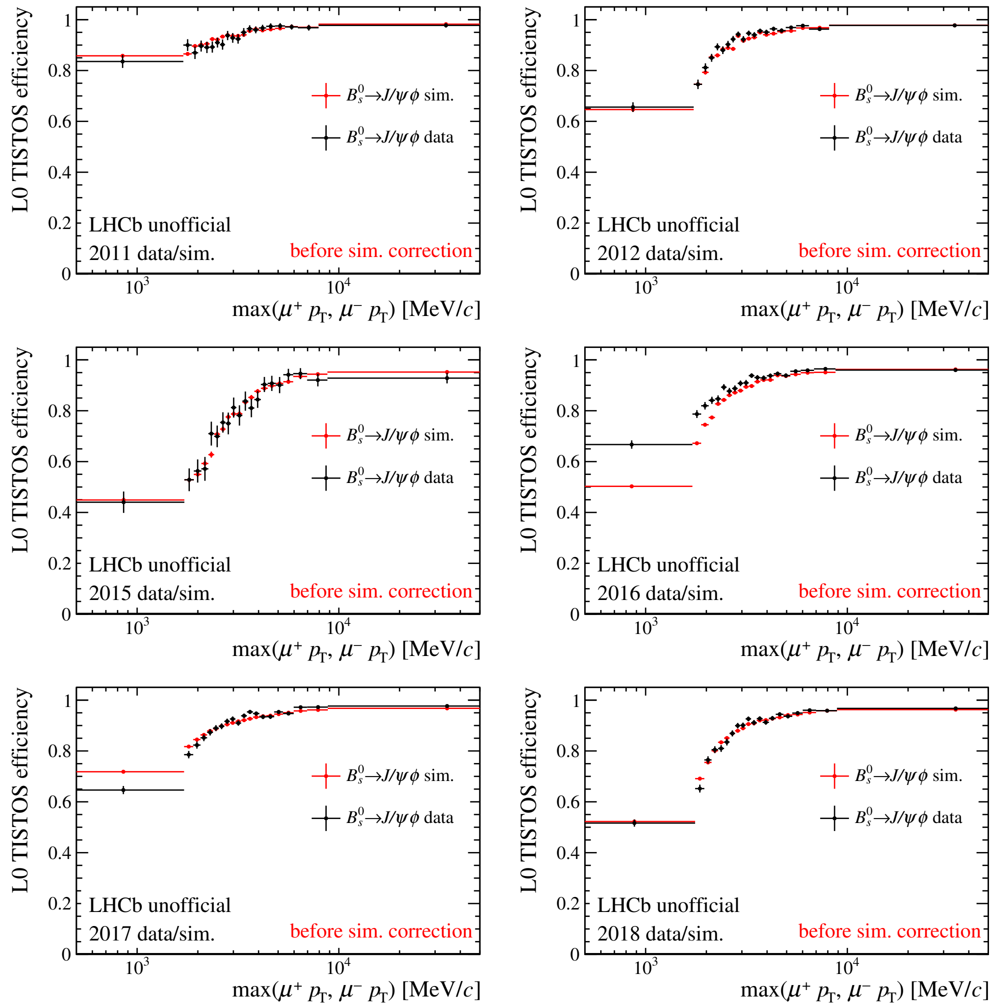

Figure 4.18: TISTOS efficiencies for intervals of $\max[\mu_1(p_T), \mu_2(p_T)]$ for $B_s^0 \to J/\psi\phi$ decays for different years of data-taking. The efficiencies from data are shown (black markers) overlaid on the efficiencies determined on simulation (red markers). The ratio of these efficiency curves is assigned as weight to the simulated samples.

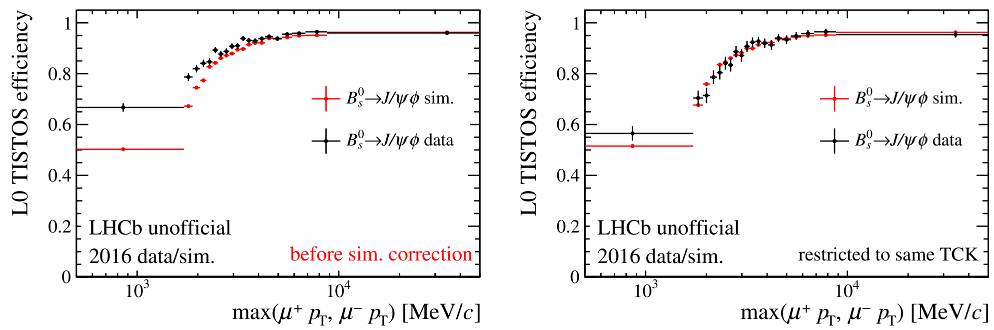

Figure 4.19: Comparison of TISTOS efficiencies determined on $B_s^0 \to J/\psi\phi$ decays in data (black) and simulation (red). The left plot is the same as in Fig. 4.18, and the right shows the efficiencies when restricting the data to the same TCK as used for the simulated sample. In this case, the efficiencies are in good agreement, confirming that the differences are due to the different trigger settings.

4.6.6 Lifetime difference in $B^0_s \to \phi\mu^+\mu^-$ simulation

The available $B^0_s \to \phi\mu^+\mu^-$ simulation is generated with the model from Ref. [124], which neglects the lifetime difference between the heavy and light mass eigenstate of the B^0_s meson. To correct the $B^0_s \to \phi\mu^+\mu^-$ simulation, weights are applied that modify the simulation to take the lifetime difference between the two mass eigenstates into account.

Following the considerations from Ref. [125], the decay rate of $B^0_s \to \phi\mu^+\mu^-$ decays Γ can be parametrised by the decay time t, three decay angles Ω and the invariant dimuon mass squared, q^2, and expressed in a basis of angular functions $f_i(\Omega)$, and angular coefficients $J_i(q^2,t)$, namely

$$\frac{\mathrm{d}^3\Gamma}{\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}q^2} = \sum_i J_i(q^2,t) \times f_i(\Omega), \tag{4.21}$$

where the form factors of the decay, which are explained in Sect. 2.2.2, are included in angular observables $J_i(q^2,t)$.

In this measurement, where the ϕ meson is reconstructed via the $\phi \to K^+K^-$ decay, decays from B^0_s cannot be distinguished from $\overline{B^0_s}$ decays since they would have exactly the same final state particles. Therefore, this measurement assesses only the *untagged CP-averaged decay rate*, which is given by

$$\frac{\mathrm{d}^3(\Gamma+\bar{\Gamma})}{\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}q^2} = \sum_i \underbrace{\left[J_i(q^2,t) + \zeta_i \bar{J}_i(q^2,t)\right]}_{\equiv \mathcal{J}_i(q^2,t)} \times f_i(\Omega), \tag{4.22}$$

with

$$\mathcal{J}_i(q^2,t) = e^{-\Gamma_s t} \times \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) \times \left(I_i(q^2) + \zeta_i \bar{I}_i(q^2)\right)\right. \tag{4.23}$$

$$\left. + \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \times h_i(q^2)\right], \tag{4.24}$$

with the q^2-dependent angular observables I_i, $\bar{I}_i$ and h_i, the average B^0_s lifetime Γ_s, and the lifetime difference $\Delta\Gamma_s$.

Equation (4.22) describes the decay rate of $B^0_s \to \phi\mu^+\mu^-$ decays that includes the effect of the lifetime difference between the light and the heavy B^0_s eigenstate. This improved description of the decay is used to determine the correction weights for the simulated samples of $B^0_s \to \phi\mu^+\mu^-$, taking the angular coefficients I_i, $\bar{I}_i$ and h_i from the FLAVIO software package [126].

The correction weights align the five-dimensional distribution of t, Ω, and q^2 between the simulated $B^0_s \to \phi\mu^+\mu^-$ decay sample and a sample generated according to Eq. 4.22. Figure 4.20 shows the distributions before the weights are applied. The five-dimensional histogram ensures that the weights also correct the correlations between the parameters. This is a different approach than for the other corrections, where only 1D histograms where used, and it is complicated by the sample sizes needed: the chosen weighting configuration uses 20 intervals for the angles and q^2, and 100 intervals for the decay time t, which corresponds to 16 million 5D-intervals. Filling these intervals with a sufficient amount of data points requires the generation of at least 500 million $B^0_s \to \phi\mu^+\mu^-$ candidates.

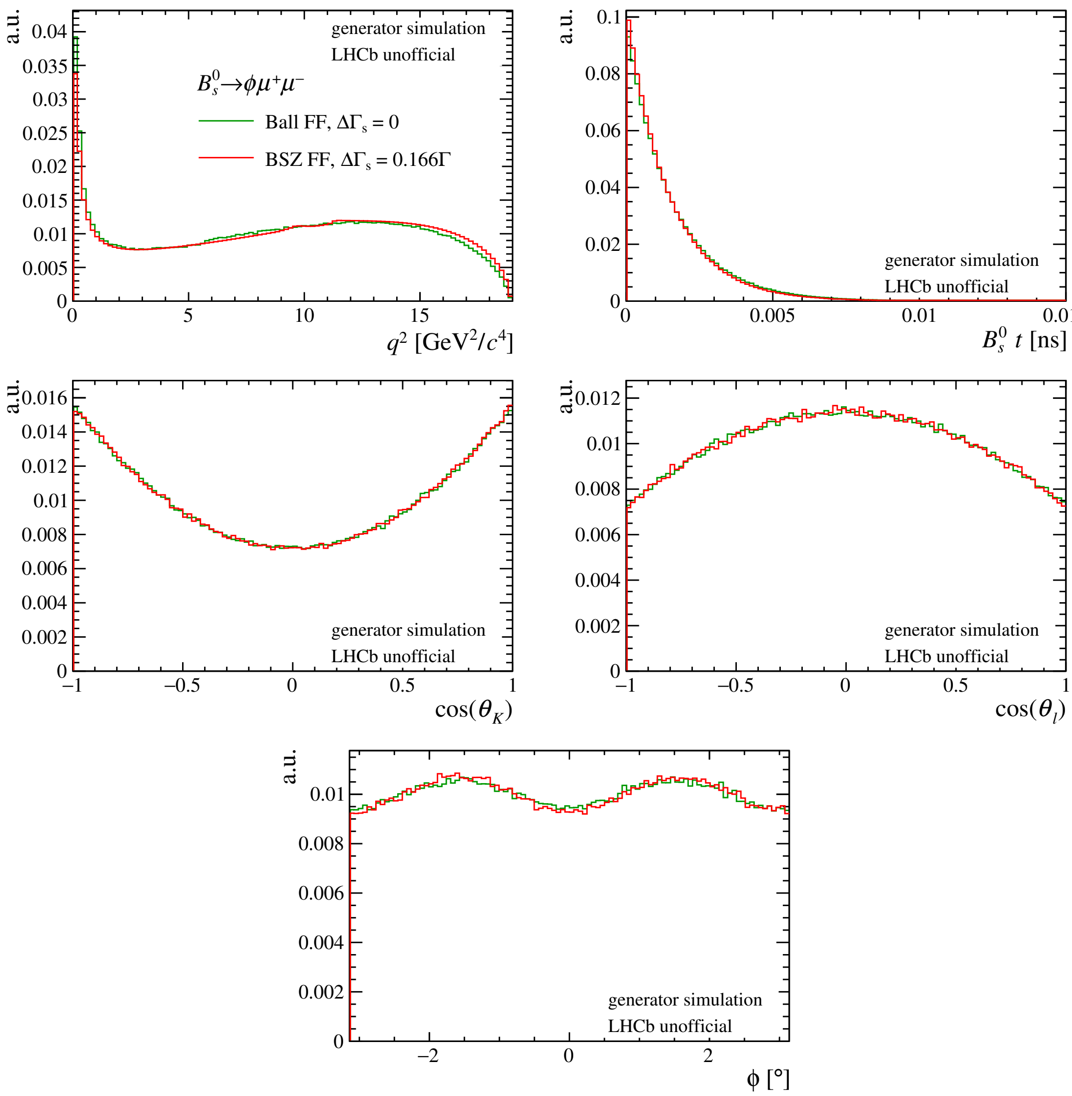

Figure 4.20: Comparison between invariant dimuon mass q^2 (top left), B_s^0 meson decay time t (top right), $\cos\theta_K$ (middle left), $\cos\theta_\ell$ (middle right), and ϕ (bottom) from uncorrected generator simulation (green) and B_s^0 lifetime and form factor corrected generator simulation (red).

As the angular coefficients in Eq. 4.22 are taken from the FLAVIO package, which uses the Ball-Straub-Zwicky (BSZ) form factors from Ref. [127], the weighting also changes the underlying form factors. Therefore, the weighted simulation does not only consider the non-zero lifetime difference between the two mass eigenstates of the B_s^0 meson, but it also updates the older form factors from Ref. [124] to the more recent calculations from Ref. [127].

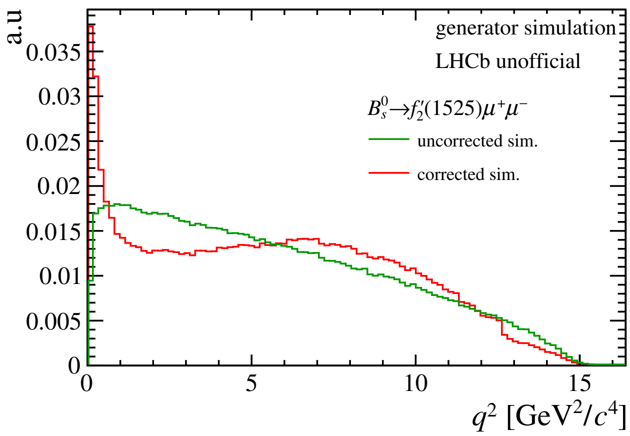

Figure 4.21: Comparison of the q^2 distribution between the uncorrected (green) and corrected (red) $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation before the reconstruction. The simulation uses the simple phase-space approach, whereas weights correct the simulation to resemble the distribution from Ref. [128].

4.6.7 Invariant dimuon mass in $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation

The q^2 distribution of the available $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation is modelled by considering only the available phase space of the decay without any features of a typical $b \to s\ell^+\ell^-$ transition. Since the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode is measured integrated over q^2, a realistic underlying q^2 distribution is important to calculate the correct efficiencies.

To improve the q^2 description in the simulated sample, weights are assigned that align the simulated distribution to the SM-prediction of the q^2 shape from Ref. [128]. The weights are determined in 40 isopopulated bins from $q^2 = 0\,\text{GeV}^2/c^4$ to the phase-space limit at $q^2 = 15\,\text{GeV}^2/c^4$. The q^2 distribution before and after this correction is shown in Fig. 4.21.

4.6.8 Correction validation

The corrections are finally validated by comparing distributions between data and simulated candidates from the control mode $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays after all the corrections are applied.

The validation using the $B_s^0 \to J/\psi f_2'(1525)$ decays is especially important since all weights applied to this mode, *i.e.* the B_s^0 p_{T}, multiplicity, and trigger corrections, were determined using $B_s^0 \to J/\psi\phi$ decays. Even though the distributions related to the muon kinematics are similar between the $B_s^0 \to J/\psi f_2'(1525)$ and $B_s^0 \to J/\psi\phi$ mode, differences can occur from the difference between the distributions related to the dikaon system.

For the comparison, the $B_s^0 \to J/\psi\phi$ candidates are selected without preselection, PID, and BDT response requirements so that the distributions can be compared before these selection requirements. This, however, is not possible for the $B_s^0 \to J/\psi f_2'(1525)$ mode since the PID and BDT requirements are essential to obtain a sufficiently pure sample. Both modes are considered in the B_s^0 signal

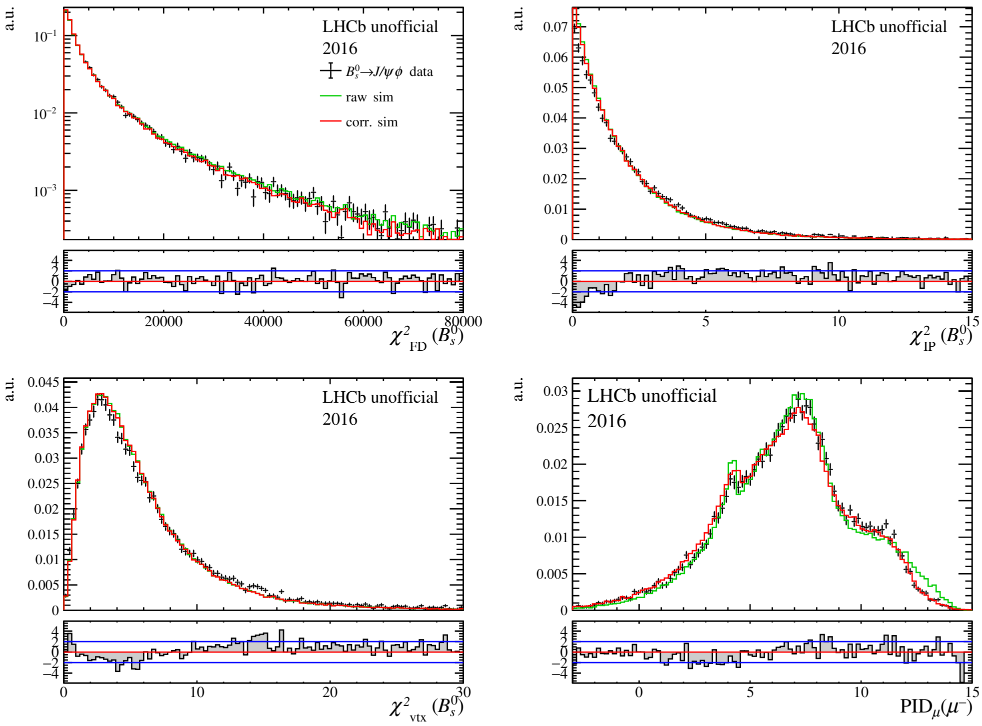

Figure 4.22: Distribution of variables used in the stripping selection from $B_s^0 \to J/\psi\phi$ data (black markers) and simulation before (green) and after (red) correction. The distributions are shown for 2016 conditions as an example.

region, which is defined to be $50\,\mathrm{MeV}/c^2$ around the known B_s^0 mass, and residual background has been statistically subtracted with the sPlot technique [116].

The most relevant variables are those used in the signal selection, which is described in Section 4.3. Examples for the comparison of variables used in the stripping selection of $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi f_2'(1525)$ decays are shown in Figs. 4.22 and 4.23, respectively; variables used in the preselection and misid-vetoes, and the BDT responses in Figs. 4.24 and 4.25, respectively.

After the corrections are applied, good agreement between data and simulation is found throughout the observables for all data-taking periods and both modes. Small residual differences, seen in the BDT response, the $\chi^2_{\rm IP}$, and $\chi^2_{\rm FD}$ of the B_s^0 meson in the 2016–2018 data-taking periods, have been further studied, and their effects are either negligible themselves or sufficiently cancel in the ratio of efficiencies between rare and normalisation mode.

For $B_s^0 \to \phi\mu^+\mu^-$ decays, the specific lifetime correction is checked by comparing the decay time distribution in the data and simulation, shown in Fig. 4.26. In the same figure, also the BDT responses are compared to ensure good agreement across q^2 between data and simulation for different regions of invariant dimuon masses q^2. Good agreement between $B_s^0 \to \phi\mu^+\mu^-$ simulation and data is found, however the comparison is clearly limited by the amount of data candidates. This check has been added after obtaining the permission to reveal the signal region.

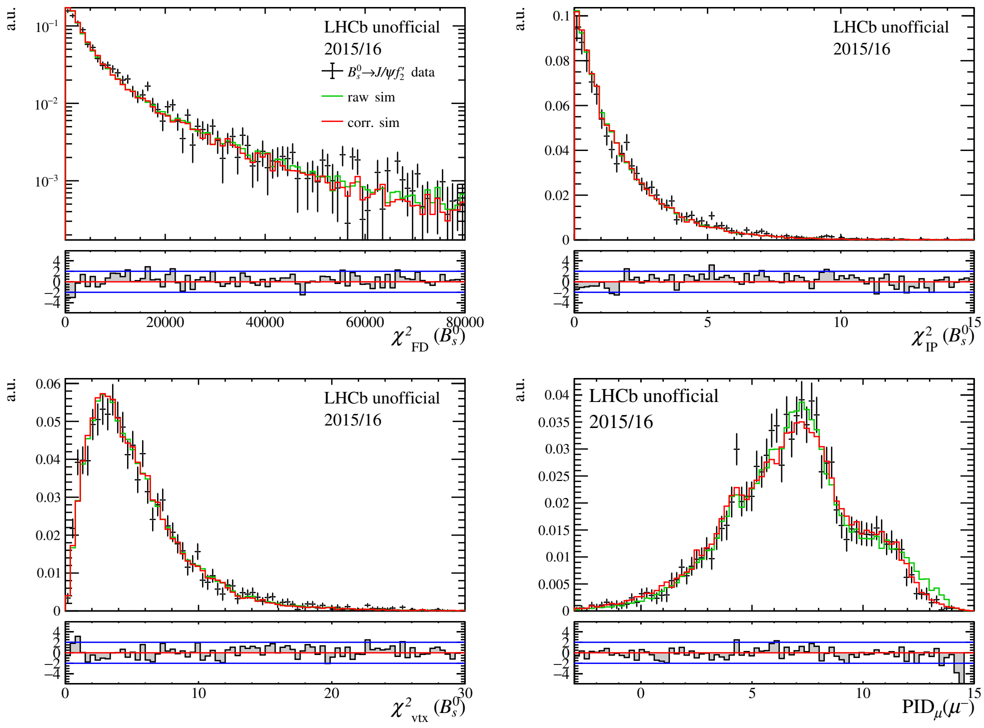

Figure 4.23: Distribution of variables used in the stripping selection from $B_s^0 \to J/\psi f_2'(1525)$ data (black markers) and simulation before (green) and after (red) correction. The distributions are shown for the combined 2015–2016 conditions as an example.

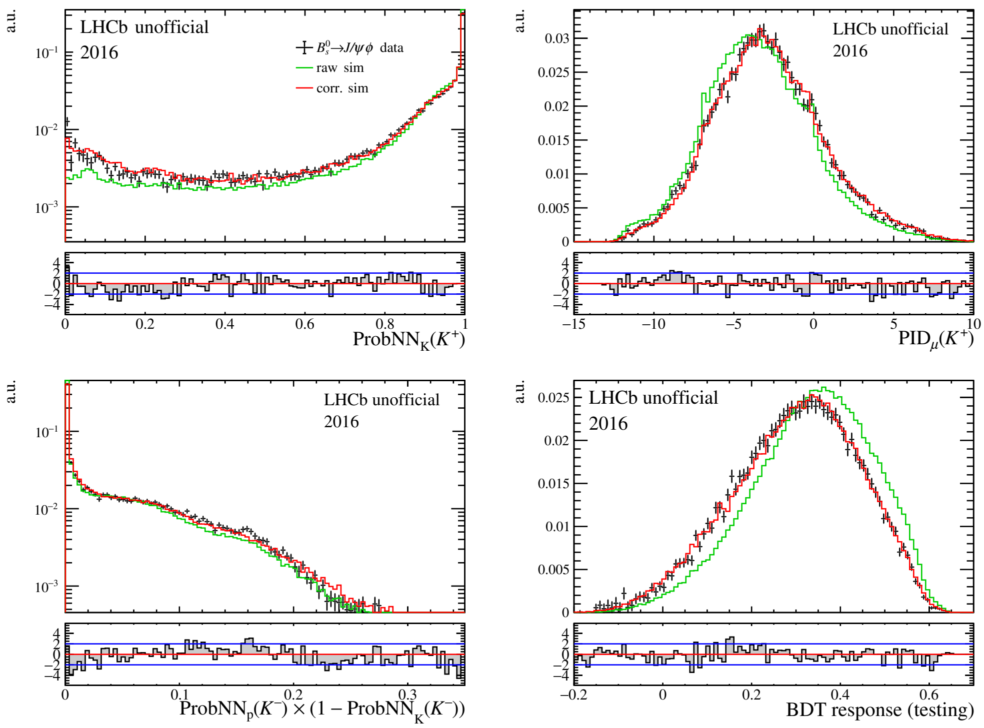

Figure 4.24: Distribution of variables used in the preselection and misID vetoes, and the BDT response from $B_s^0 \to J/\psi\phi$ data (black markers) and simulation before (green) and after (red) correction. The distributions are shown for 2016 conditions as an example.

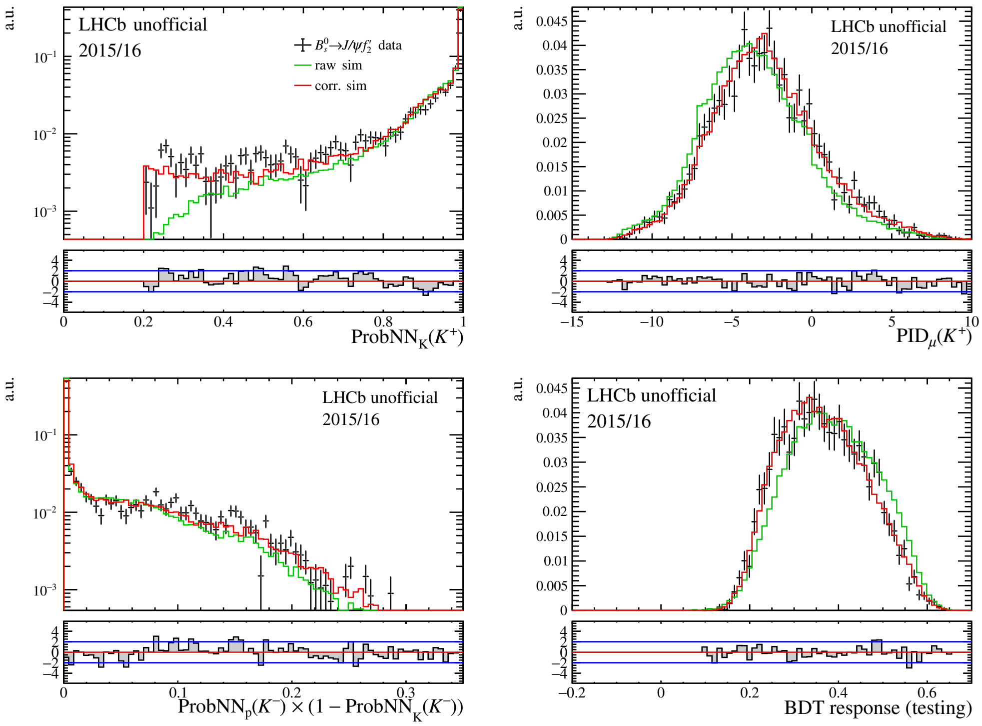

Figure 4.25: Distribution of variables used in the preselection and misID vetoes and the BDT response in $B_s^0 \to J/\psi f_2'(1525)$ 2 data (black markers) and simulation before (green) and after (red) correction. The distributions are shown for the combined 2015–2016 conditions as an example.

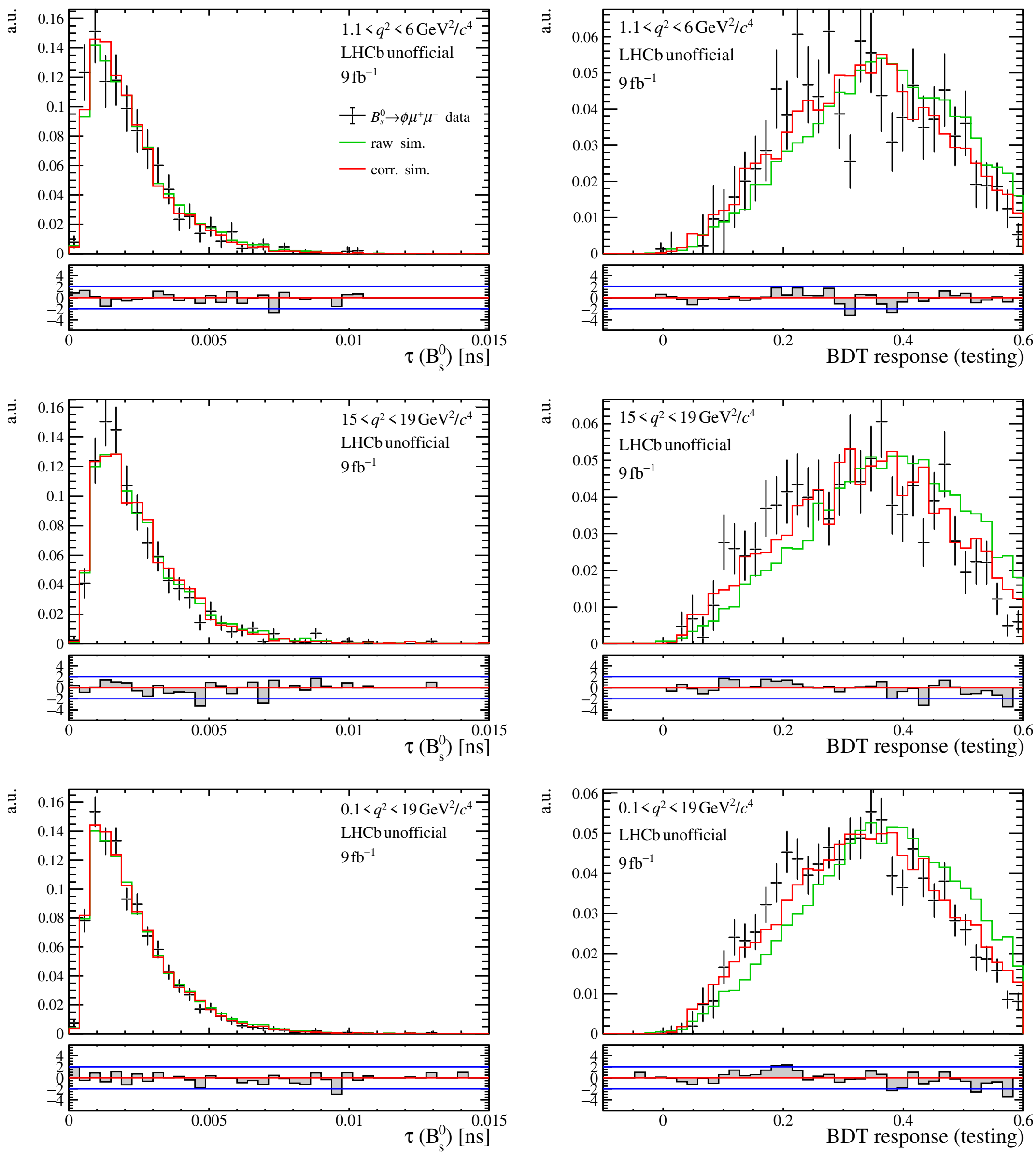

Figure 4.26: Distribution of (left) B_s^0 meson decay time and (right) BDT response from the complete statistically background-subtracted ("sPlotted") $B_s^0 \to \phi\mu^+\mu^-$ data (black markers) and simulation before (green) and after (red) correction.

Chapter 5

Branching fraction measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays

Following the previous chapter discussing the general strategy and preparations for both branching fraction measurements, this chapter now details the specific measurement of $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$. The branching fraction is measured in intervals of the invariant dimuon mass squared q^2, which are visualised in Fig. 5.1. The low q^2 region with $q^2 < 8\,\text{GeV}^2/c^4$ is divided into five q^2 intervals, which are two more than in the previous measurement in Ref. [2]. This finer q^2 division is possible with the now larger data sample compared to Ref. [2], and it allows for a more detailed study of the underlying q^2 distribution. This finer division corresponds further to q^2 intervals chosen for the analysis of $B^0 \to K^{*0}\mu^+\mu^-$ in Ref. [43] and it simplifies the comparison between the two results.

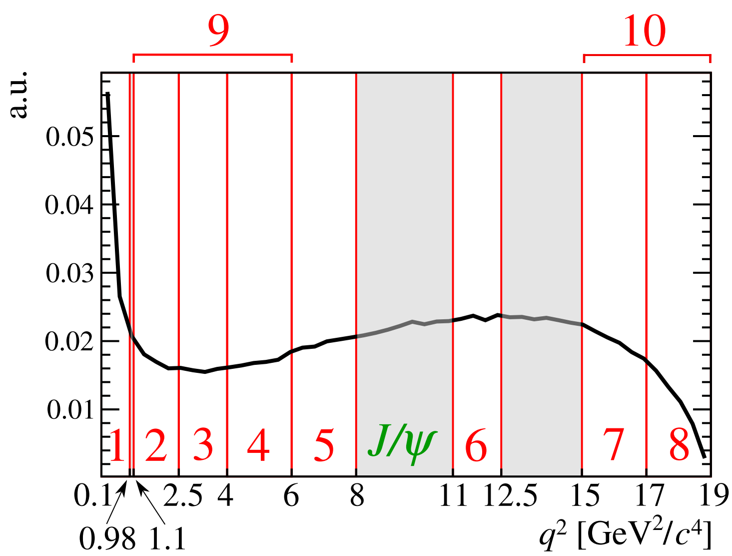

Figure 5.1: Invariant dimuon-mass squared q^2 intervals in which $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ is measured overlaid on the distribution taken from $B_s^0 \to \phi\mu^+\mu^-$ simulation. The measurement is performed in eight narrow (1–8) and two wide (9–10) q^2 intervals. The regions dominated by $B_s^0 \to \phi\phi$, $B_s^0 \to J/\psi\phi$, and $B_s^0 \to \psi(2S)\phi$ decays are indicated in gray, and they are excluded from the branching fraction measurement.

The following sections feature the different steps of the measurement: Section 5.1 presents the signal selection efficiencies, which are determined on simulated samples of the signal decays. The potential background processes that could contribute to the selected data samples are discussed in Sect. 5.2, and the extraction of the signal using data in Sect. 5.3. Both the selection efficiencies and the signal extraction are associated with systematic uncertainties, which are discussed in Sect. 5.4. The analysis is validated in Section 5.5 where the measurement is checked for internal consistency and agreement with Ref. [2]. The final result is shown in Sect. 5.6.

5.1 Signal efficiencies

The efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays are input parameters for the branching fraction measurement as the ratio of efficiencies $\epsilon(B_s^0 \to J/\psi\phi)/\epsilon(B_s^0 \to \phi\mu^+\mu^-)$ enters the branching fraction formula discussed in Sect. 4.1. The efficiencies for both the normalisation and the rare mode decays are determined on corrected simulation of the decays, as described in Sect. 4.5. The efficiencies are calculated separately for the different years of data taking and the different selection stages. For $B_s^0 \to \phi\mu^+\mu^-$ decays, the efficiencies are further determined separately for the different q^2 intervals.

The efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays are shown in Fig. 5.2 for the different selection stages. Roughly 15–17% of the simulated final-state particles of the B_s^0-meson decays are inside the LHCb acceptance, which is defined between $10--400\,\mathrm{mrad}$ with respect to the beam axis. The fraction is smaller in 2011–2012 than for the 2015–2018 data-taking, which is due to the larger centre-of-mass energy $\sqrt{s}$ of the pp-collisions during 2015–2018 data-taking. The larger $\sqrt{s}$ results in B_s^0 mesons with increased boost along the beam axis, and this boost collimates the decays more into the forward-direction covered by the LHCb detector.

The combined reconstruction and stripping efficiency is 10–11% at low q^2 and falls down to roughly 6–7% at high q^2. This q^2-trend is explained by the kinematics of the decay: decays via $B_s^0 \to \phi\mu^+\mu^-$ mode with large q^2 tend to have a low-momentum kaon in the final state, and this kaon is often bent out of the detector acceptance. In this case, it cannot be reconstructed as long-track, *i.e.* with hits in the full tracking system, and the candidate is rejected.

The preselection efficiency is relatively flat at roughly 85%. For the $B_s^0 \to J/\psi\phi$ mode, the efficiencies are slightly larger than expected from a naive interpolation of the $B_s^0 \to \phi\mu^+\mu^-$ efficiencies; this efficiency difference comes from the invariant-mass selection $m(K^+K^-\mu^+\mu^-) \in [5270, 5700]\,\mathrm{MeV}$, which removes decay candidates in the low-mass tails. These mass tails are slightly different for $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ decays due to the different underlying q^2 distributions—for $B_s^0 \to J/\psi\phi$, the decay candidates accumulate at the known J/ψ mass, and for $B_s^0 \to \phi\mu^+\mu^-$, they are distributed across q^2. This difference affects how the decays migrate out of the $[8, 11]\,\mathrm{GeV}^2/c^4$ bin, which slightly changes the fraction of decay candidates falling in the low invariant-mass tails removed by the preselection.

The trigger efficiencies vary significantly with q^2 and for the different data-taking conditions. The q^2 trend is similar for the different data-taking conditions, and it is explained by the correlation of q^2 and the muon momenta: a larger q^2 value is associated with larger muon momenta, and large muon momenta are associated with a larger probability to fire the muon hardware triggers, explained in Sect. 4.3.2. The trigger-efficiency levels are different between the years, and they

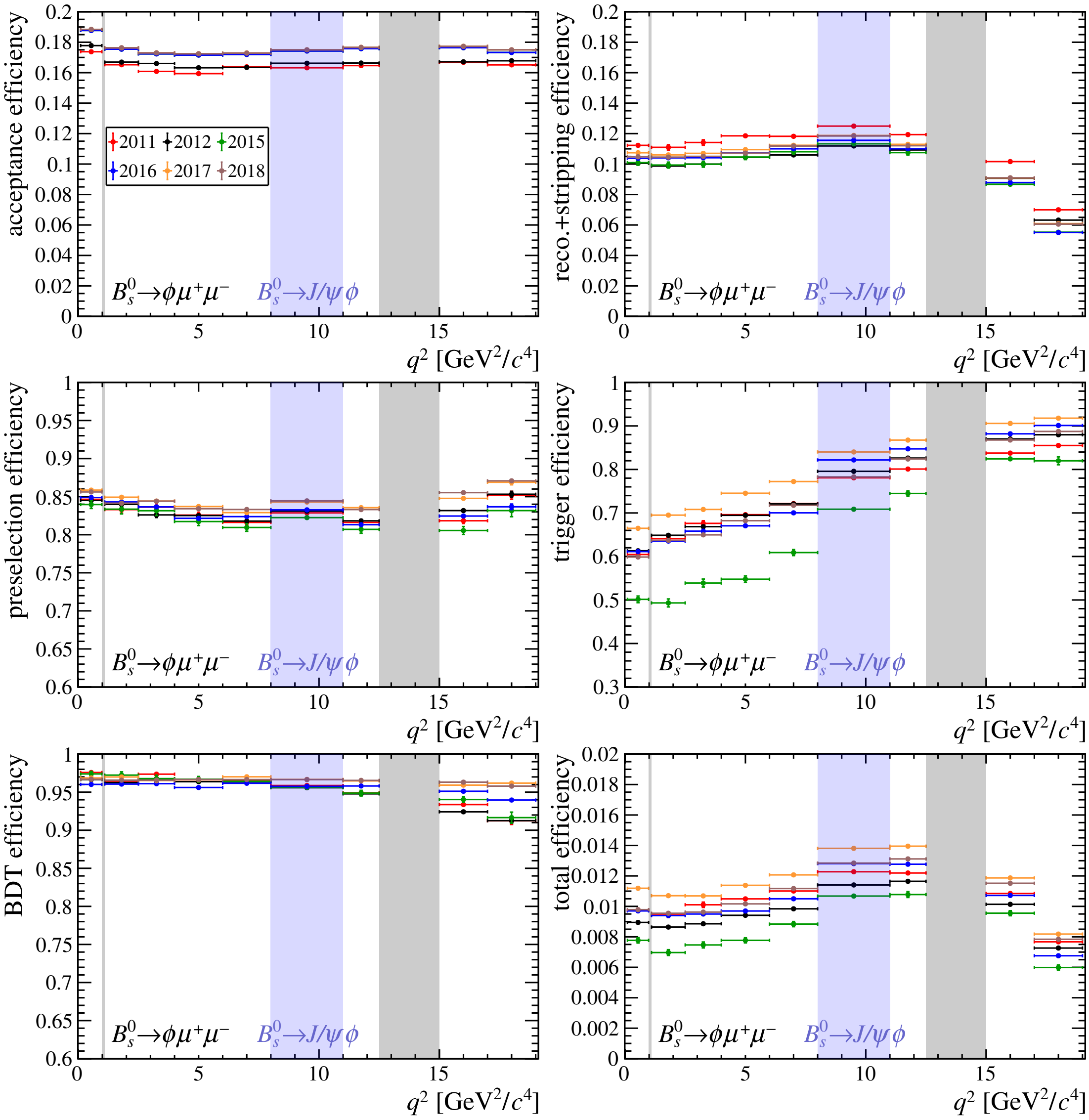

Figure 5.2: efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays for the different data-taking conditions in the different stages: The LHCb acceptance (top left), the combined reconstruction and stripping (top right), the preselection (middle left), the trigger (middle right), the BDT (bottom left), and the total efficiency (bottom right). The efficiencies for $B_s^0 \to J/\psi\phi$ decays are determined on simulated samples of $B_s^0 \to J/\psi\phi$ decays and are indicated in the blue area. The efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ decays are determined on simulated samples of $B_s^0 \to \phi\mu^+\mu^-$ decays.

are largest for 2017 and smallest for 2015. This is directly connected to the muon hardware trigger thresholds, which are set to lowest values for 2017 and highest for 2015 data-taking.

The BDT efficiency is approximately constant at roughly 95% for 2016–2018, but it is decreasing for 2011–2015 from 95% to roughly 92% at large q^2 values. This difference is directly connected to differences in the BDT input parameters: in the 2016–2018 sample, the kaons' ProbNN$_K$ variables are more optimised for soft kaons which results in a higher efficiency for high q^2. Since the BDT efficiency is overall still high, this trend in 2011–2015 data is not problematic to the selection, especially since the BDT response is found to be well modelled in the simulated

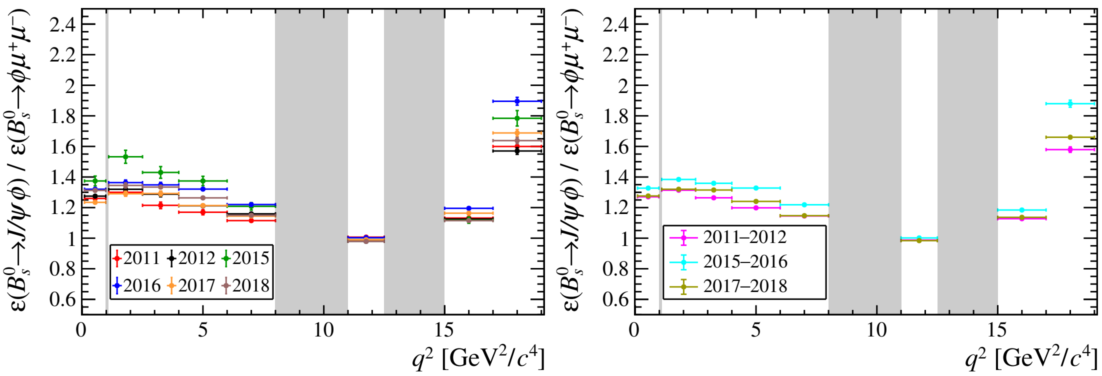

Figure 5.3: Ratios between the efficiencies for the normalisation mode $B_s^0 \to J/\psi\phi$ and the rare mode $B_s^0 \to \phi\mu^+\mu^-$ decays. The efficiency ratios are (left) determined for the separate years of data-taking and then (right) combined for the branching fraction measurement using the combined 2011–2012, 2015–2016 and 2017–2018 data sets.

sample across q^2 as shown in Section 4.6.8.

The total selection efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays are between roughly 0.8% and 1.3%. The efficiencies are largest for $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ decays with q^2 close to the J/ψ mass squared; the efficiencies are largest for 2017 and smallest for 2015 data.

In the branching fraction measurement, the efficiency ratios $\epsilon(J/\psi\phi)/\epsilon(\phi\mu^+\mu^-)$ are used, which are shown in Fig. 5.3. As the signal decays are extracted with simultaneous fits to the three data sets containing the combined 2011–2012, 2015–2016, and 2017–2018 data, the yearly efficiency ratios are combined. This is done by using a weighted sum, *e.g.* for the combination of 2011 and 2012 via

$$\begin{aligned}\left(\frac{\epsilon(\phi\mu^+\mu^-)}{\epsilon(J/\psi\phi)}\right)^{2011\text{–}2012} &= \frac{N(J/\psi\phi)^{2011}}{N(J/\psi\phi)^{2011}+N(J/\psi\phi)^{2012}} \times \left(\frac{\epsilon(\phi\mu^+\mu^-)}{\epsilon(J/\psi\phi)}\right)^{2011} \\ &+ \frac{N(J/\psi\phi)^{2012}}{N(J/\psi\phi)^{2011}+N(J/\psi\phi)^{2012}} \times \left(\frac{\epsilon(\phi\mu^+\mu^-)}{\epsilon(J/\psi\phi)}\right)^{2012},\end{aligned} \tag{5.1}$$

with the number of $B_s^0 \to J/\psi\phi$ candidates $N(J/\psi\phi)$ extracted from fits to the respective data samples, and the efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays, $\epsilon(\phi\mu^+\mu^-)$ and $\epsilon(J/\psi\phi)$ determined on simulation. This way, the yearly efficiencies are weighted according to the amount of data taken in the corresponding year where the amount of data is estimated with the fraction of $B_s^0 \to J/\psi\phi$ candidates in the corresponding data sample. These combined efficiencies are used as input parameters to the branching fraction measurement, and they are also shown in Fig. 5.3.

5.2 Background processes

The branching fraction is measured with $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decay candidates extracted from the data. If other decay processes pass the selection and contribute to the selected data, they can be confused with the signal and thereby affect the measurement. These contributions from other decay processes, referred to as *background* processes, can be taken into account with two approaches:

(1) separating the background contribution from the signal decays with a fit to data as explained in Sect. 4.4 or (2) taking the effect from background decays into account with a systematic uncertainty that covers the potential effect of the background decays on the measurement.

What approach is chosen for an analysis depends on the amount of background contribution that is expected in the data—if many background candidates are expected, they need to be included in the fit and separated from the signal. If few background candidates are expected that affect the measurement insignificantly compared to other effects, they can be neglected in the measurement and added as a systematic uncertainty.

Following these considerations, this section discusses the potential background sources and their effect on the measurement. For this, the number of expected background candidates $N_{\text{bkg.}}$ is determined relative to the number of $B_s^0 \to J/\psi\phi$ decays via

$$N_{\text{bkg.}} = \frac{N^{\text{data}}_{B_s^0 \to J/\psi\phi}}{\epsilon_{B_s^0 \to J/\psi\phi}} \times \frac{f_{\text{bkg.}}}{f_s} \times \epsilon_{\text{bkg.}} \times \frac{\mathcal{B}(\text{bkg.})}{\mathcal{B}(B_s^0 \to J/\psi(\to \mu^+\mu^-)\phi(\to K^+K^-))}, \quad (5.2)$$

with the number of $B_s^0 \to J/\psi\phi$ data candidates $N^{\text{data}}_{B_s^0 \to J/\psi\phi}$, the efficiencies $\epsilon_{B_s^0 \to J/\psi\phi}$ ($\epsilon_{\text{bkg.}}$), the branching fractions $\mathcal{B}(B_s^0 \to J/\psi\phi)$ ($\mathcal{B}(\text{bkg.})$), and the fragmentation fraction f_s ($f_{\text{bkg.}}$) for the $B_s^0 \to J/\psi\phi$ (background) decays. Determining the number of background decays relative to the $B_s^0 \to J/\psi\phi$ decays is more precise than calculating it directly, *i.e.* using the proton-proton collision cross-sections $\sigma_{pp \to b\bar{b}X}$ and the luminosities as the $B_s^0 \to J/\psi\phi$ parameters are more precise than those related to the proton-proton collisions.

The background candidates in the *signal region*, *i.e.* with invariant masses within 50 MeV around the B_s^0 mass $m_{B_s^0}$, have the largest impact on the fit. Therefore, the background candidates specifically falling into this region are determined.

The efficiencies for Eq. 5.2 are determined using corrected simulation samples of $B_s^0 \to J/\psi\phi$ and the respective background decays, where 2012 simulation is taken as benchmark for the combined 2011–2012, 2016 simulation for the combined 2015–2016, and 2017 simulation for the combined 2017–2018 data sample. Even though the efficiencies are slightly different for the different years, this estimate is reasonable given that precision of $N_{\text{bkg.}}$ is limited by the knowledge on the branching fractions and the statistical limitations of the simulation samples.

The effect of a background to the measurement is dependent on its amount relative to the signal, so the expected signal contribution in the data is determined first. The relevant background sources for this analysis, summarised in Figure 5.4, are the following: contributions from (1) Λ_b^0 baryon and (2) B^0 meson decays with one misidentified final-state particle are discussed in Sect. 5.2.3 and Sect. 5.2.2, respectively, and contributions from (3) *double misID* B_s^0 and (4) B^0 decays, *i.e.* decays with J/ψ meson and two misidentified final-state particles, in Sect. 5.2.4. Contributions from (5) non-resonant $B_s^0 \to K^+K^-\mu^+\mu^-$ decays with the dikaon system being in an S-wave configuration is assessed in Sect. 5.2.5 and contributions from (6) random track combinations in Sect. 5.2.6.

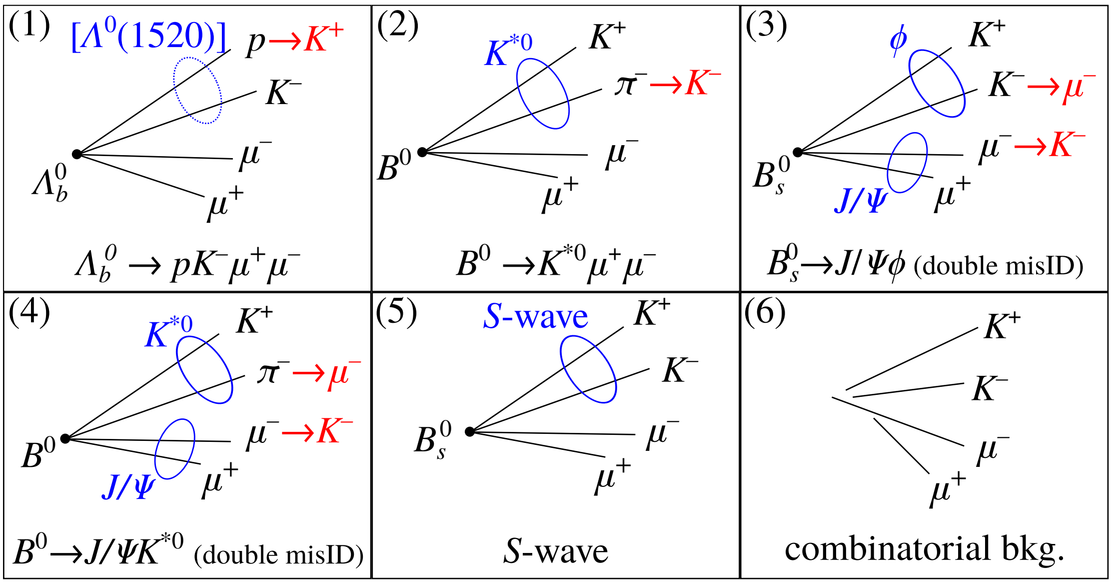

Figure 5.4: Overview of the background sources relevant to the analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays. Expected contribution are from (1) Λ_b^0 decays with the proton being misidentified as a kaon, for which the decays involving a $\Lambda^0(1520)$ resonance are dominant; from (2) B^0 decays with the pion being misidentified as kaon; from (3) B_s^0 and (4) B^0 decays with two confused final-state particles that can contribute to the q^2 region outside the J/ψ veto; from (5) $B_s^0 \to K^+K^-\mu^+\mu^-$ decays with the dikaon system in an S-wave configuration; and from (6) random track combinations. The misidentfied particles are indicated in red.

5.2.1 Signal yield expectation

The expected number of signal decays is calculated as explained before with Eq. 5.2 using the branching fractions from Ref. [89], namely

$$\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-) = (8.2 \pm 1.2) \times 10^{-7} \tag{5.3}$$

and

$$\mathcal{B}(\phi \to K^+K^-) = (49.2 \pm 0.5)\%, \tag{5.4}$$

and the efficiency and q^2 distribution taken from the simulated samples of $B_s^0 \to \phi\mu^+\mu^-$ decays. From that, 420 ± 60, 480 ± 70, and 1070 ± 160 $B_s^0 \to \phi\mu^+\mu^-$ candidates are expected in the 2011–2012, 2015–2016, and 2017–2018 samples, respectively, for the *full q^2 range of the rare mode* defined as $q^2 \in [0.1, 19]\,\text{GeV}^2/c^4$ with $q^2 \notin [0.98, 1.1] \cup [8, 11] \cup [12.5, 15]\,\text{GeV}^2/c^4$. The uncertainties are dominated by the uncertainty on the rare mode branching fraction.

5.2.2 $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and $\Lambda_b^0 \to J/\psi pK^-$

Decays from Λ_b^0 baryons can be confused with the signal decays if the final-state proton is misidentified as a kaon. The selection includes a specific veto against these decays, which is described in Sect. 4.3.4, but as this veto is a compromise between background rejection rate and signal efficiency, it does not remove all Λ_b^0 decays from the selected data. In addition, for each B_s^0 meson, there are roughly two Λ_b^0 baryons produced[1]—so this background needs careful investigation.

[1]The fragmentation fraction of Λ_b^0 relative to B_s^0 mesons is $f_{\Lambda_b^0}/f_s = 1.85(33)$ [18].

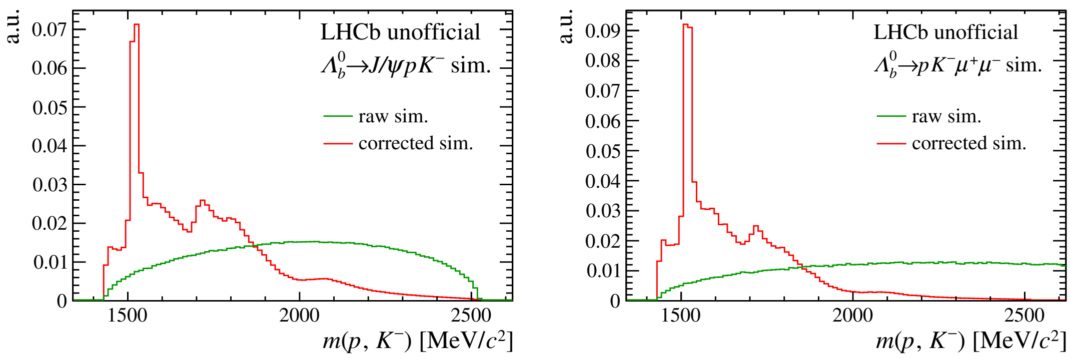

Figure 5.5: Distribution of dihadron masses from simulation of (left) $\Lambda_b^0 \to J/\psi pK^-$ and (right) $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays before any detector effects. The simulated samples are generated with a simple phase-space approach (green) and corrected (red) to resemble the distribution from $\Lambda_b^0 \to J/\psi pK^-$ data analysed in Ref. [129].

Both the rare $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and tree-level $\Lambda_b^0 \to J/\psi pK^-$ decays can contribute to the selected data samples. The former could contribute to the $B_s^0 \to \phi\mu^+\mu^-$ data sets and the latter to the $B_s^0 \to J/\psi\phi$ data sets, so both decays are investigated here.

The expected background yields are determined with Eq. 5.2 as explained before, and the background efficiencies are determined on simulated samples of the decays. These simulated samples are generated based on a simple *phase-space model*, in which the matrix element is set to unity and only the phase space of the decay is considered, so the model neglects the resonance-rich pK-spectrum found in the data. Since the efficiencies are dependent on the underlying pK-structure, the simulated samples are corrected to improve the modelling.

This correction is applied by weighting the simulated samples, so that the angular and $m(pK^-)$ distribution is similar to what is found in an analysis of $\Lambda_b^0 \to J/\psi pK^-$ data [129]. Figure 5.5 shows the $m(pK^-)$ distributions for the simulated $\Lambda_b^0 \to J/\psi pK^-$ and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays before and after this correction. Even though the angular structure and pK-spectrum is not necessarily the same for $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays, using the $\Lambda_b^0 \to J/\psi pK^-$ data to correct the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ simulation is still expected to improve the simulation-data agreement compared to using the phase-space model.

The efficiencies vary also with q^2, so the simulation of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ is further improved by correcting its q^2 distribution, which is also modelled only with the simple phase-space model. For this, the simulated samples are assigned weights that align the q^2 distribution to that predicted for $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays from Ref. [130]. Even though the q^2 range is different between $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays, the corrected distribution is more realistic than that from the phase-space model. Figure 5.6 shows the q^2 distribution in simulated $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays before and after this additional correction.

The branching fraction $\mathcal{B}(\Lambda_b^0 \to J/\psi pK^-) = (3.2 \pm 0.6) \times 10^{-4}$ is taken from Ref. [131]. Unfortunately, the branching fraction for $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays has not been measured yet for the full q^2 range, so the branching fraction is estimated as follows: a prediction for the branching fraction of $\Lambda_b^0 \to \Lambda^0(1520)\mu^+\mu^-$ is available from Ref. [132], namely $\mathcal{B}(\Lambda_b^0 \to \Lambda^0(1520)\mu^+\mu^-) = (2.1 \pm 0.8) \times 10^{-7}$, where this

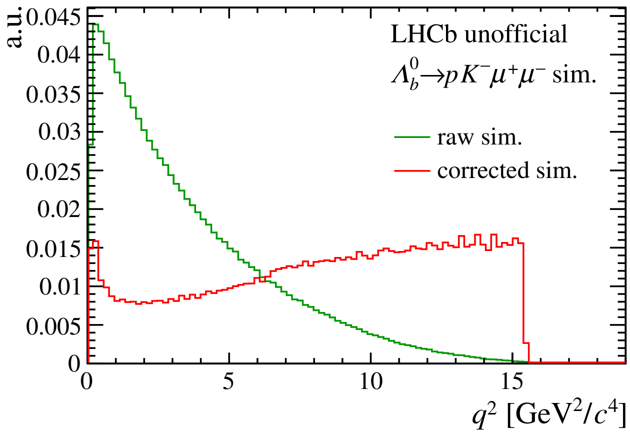

Figure 5.6: Distribution of q^2 from simulated $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays before any detector effects. The simulated sample is generated with a simple q^2 model (green) and corrected (red) to resemble the distribution from a prediction of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays [130]. The cut-off at large q^2 values comes from the model of the latter decay, which restricts the q^2 values with the intermediate Λ resonance.

value is the larger of the two results obtained from different form-factor models; the uncertainty reflects the difference between the two form-factor models. This branching fraction, however, describes only the Λ_b^0 decays that proceed via $\Lambda^0(1520)$ resonance, so when taking into account $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays proceeding via other pK-configurations, the branching fraction is larger, which is taken into account with the following considerations.

The ratio between the $\mathcal{B}(\Lambda_b^0 \to \Lambda^0(1520)\mu^+\mu^-)$ and $\mathcal{B}(\Lambda_b^0 \to pK^-\mu^+\mu^-)$ branching fractions is estimated with the assumption that it is similar to the resonant mode, *i.e.* between $\Lambda_b^0 \to J/\psi\Lambda^0(1520)$ and $\Lambda_b^0 \to J/\psi pK^-$ decays. These tree-level decays were measured in Ref. [129] where their full pK-spectrum was investigated, and the fraction of $\Lambda_b^0 \to J/\psi\Lambda^0(1520)$ decays is provided relative to the full spectrum of $\Lambda_b^0 \to J/\psi pK^-$ decays. This fraction is used as benchmark to extrapolate the prediction from Ref. [132] to the full pK-spectrum, which results in the branching fraction estimate $\mathcal{B}(\Lambda_b^0 \to pK^-\mu^+\mu^-) = (2.9 \pm 1.1) \times 10^{-7}$.

The resulting expected yields of the two Λ_b^0 decay modes are calculated. The expected yield of $\Lambda_b^0 \to J/\psi pK^-$ candidates in the $B_s^0 \to J/\psi\phi$ sample is 255 ± 69, 339 ± 88, and 770 ± 200 in the 2011–2012, 2015–2016, and 2017–2018 data sets, respectively, which corresponds to roughly 0.5% relative to the number of $B_s^0 \to J/\psi\phi$ candidates. The expected yield of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ candidates in the $B_s^0 \to \phi\mu^+\mu^-$ sample is 1.71 ± 0.73, 1.77 ± 0.75 and 5.6 ± 2.4 candidates, which corresponds to roughly 0.5% relative to the number of expected $B_s^0 \to \phi\mu^+\mu^-$ candidates in the full q^2 range. The expected yield resolved in q^2 is shown in Figure 5.7 where it is compared to the expected $B_s^0 \to \phi\mu^+\mu^-$ yield.

Since both $\Lambda_b^0 \to J/\psi pK^-$ and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays are expected to contribute with less than 1% relative to the $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ data samples, respectively, they are neglected when determining the number of signal candidates and accounted for with a small systematic uncertainty on the branching fraction result as detailed in Sect. 5.4.3. Furthermore, since the modes contribute to both the $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ data samples, the effect is expected to be even

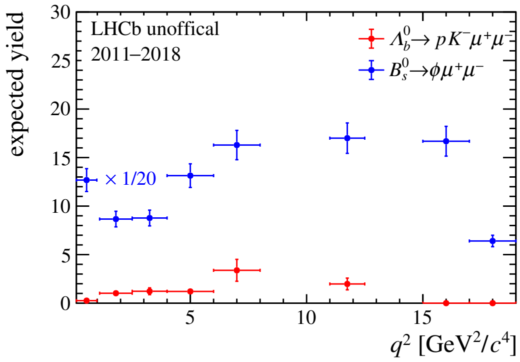

Figure 5.7: Expected number of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays (red) in the full $B_s^0 \to \phi\mu^+\mu^-$ data set. The $B_s^0 \to \phi\mu^+\mu^-$ signal expectation scaled down by a factor of 20 is shown in blue.

further reduced in the ratio of event yields used as input to the branching fraction measurement.

5.2.3 $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi K^{*0}$

Decays from B^0 mesons via $B^0 \to K^+\pi^-\mu^+\mu^-$ can contribute to the data samples if the pion is misidentified as a kaon. For the analysis of $B_s^0 \to \phi\mu^+\mu^-$, the most dominant contribution from B^0 decays is where the $K\pi$-system stems from a K^{*0} meson, *i.e.* from $B^0 \to K^{*0}\mu^+\mu^-$ decays. Similar to what was discussed before, background from B^0 decays can contribute to both the $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ data samples, namely via the rare $B^0 \to K^{*0}\mu^+\mu^-$ mode for the former and via the tree-level $B^0 \to J/\psi K^{*0}$ mode to the latter data sample.

The branching fractions for the decays are taken from Ref. [89], namely $(9.4 \pm 0.5) \times 10^{-7}$ for $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ and $(1.27 \pm 0.05) \times 10^{-3}$ for $\mathcal{B}(B^0 \to J/\psi K^{*0})$. These are at the same order of magnitude as the signal branching fractions, but with fragmentation fractions of $f_d/f_s = 3.86 \pm 0.3$ [18], so B^0 mesons are produced roughly 4 times more often than B_s^0 mesons.

The expected number of $B^0 \to J/\psi K^{*0}$ and rare $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data samples are calculated using Eq. 5.2, and the efficiencies are determined on simulated samples of rare $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi K^{*0}$ decays. From this, 263 ± 29, 210 ± 29, and 580 ± 64 candidates of $B^0 \to J/\psi K^{*0}$ decays are expected in the $B_s^0 \to J/\psi\phi$ sample, and 1.58 ± 0.25, 1.81 ± 0.35 and 4.08 ± 0.54 of rare $B^0 \to K^{*0}\mu^+\mu^-$ decays in the full q^2-range of the $B_s^0 \to \phi\mu^+\mu^-$ samples of 2011–2012, 2015–2016 and 2017–2018, respectively. Figure 5.8 shows the expected yield resolved in q^2. Over the full q^2 range, B^0 decays contribute with roughly 0.4% and 0.5% candidates relative to the normalisation and rare mode, respectively. Since this contribution is small, it is neglected in the measurement and accounted for as systematic uncertainty. Similarly as for the Λ_b^0 decays, the total effect on the branching fraction measurement is expected to be even smaller, since additional cancellation is expected in the ratio of event yields, as discussed in Sect. 5.4.3.

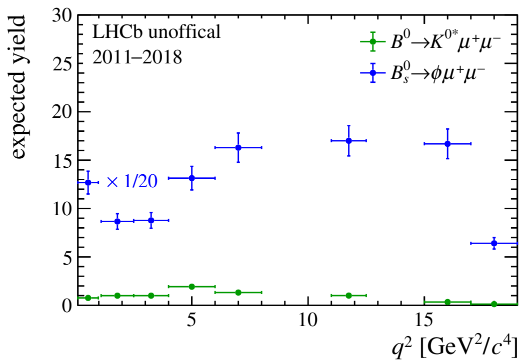

Figure 5.8: Expected number of $B^0 \to K^{*0}\mu^+\mu^-$ decays (green) in the full $B^0_s \to \phi\mu^+\mu^-$ data set. The $B^0_s \to \phi\mu^+\mu^-$ signal expectation scaled down by a factor of 20 is shown in blue.

5.2.4 Doubly misidentified $B^0_s \to J/\psi\phi$ and $B^0 \to J/\psi K^{*0}$

Another potential background source contributing to the $B^0_s \to \phi\mu^+\mu^-$ data sample is from $B^0_s \to J/\psi\phi$ or $B^0 \to J/\psi K^{*0}$ decays with two misidentified final-state particles: a final-state kaon (pion) misidentified as a muon and a final-state muon as kaon (pion). The corresponding q^2 value of the such-reconstructed $B^0_s \to J/\psi\phi$ ($B^0 \to J/\psi K^{*0}$) decays is not necessarily at the J/ψ meson mass squared anymore, so these decays can contribute to the rare mode q^2 region, which is used to select the $B^0_s \to \phi\mu^+\mu^-$ decays.

Even though such a double-misidentification is unlikely, the large branching fractions of the tree-level decays make them a background that potentially affects the measurement. The backgrounds are further suppressed by a dedicated veto explained in Sect. 4.3.4. However, as there is no similar background in the $B^0_s \to J/\psi\phi$ data sample, no partial cancellation of these doubly-misidentified backgrounds is expected in the ratio of event yields.

The number of expected background candidates is calculated as before with Eq. 5.2, and the efficiencies are determined on simulated samples of $B^0_s \to J/\psi\phi$ and $B^0 \to J/\psi K^{*0}$ decays that are reconstructed under the above-explained double-misidentification. However, none of the 160×10^6 (70×10^6) simulated $B^0_s \to J/\psi\phi$ ($B^0 \to J/\psi K^{*0}$) decays are misreconstructed with such a double-misidentification and then pass the full selection. Therefore, upper limits on the efficiencies are determined that correspond to a 90% confidence level. These limits are calculated with the simplifying assumption that the efficiency is constant across the data-taking conditions.

Using the efficiency limits, < 0.03 < 0.04 and < 0.07 doubly-misidentified $B^0_s \to J/\psi\phi$ decays are expected in the $B^0_s \to \phi\mu^+\mu^-$ data sample at a 90% confidence and < 0.74, < 0.75 and < 1.42 doubly-misidentified $B^0 \to J/\psi K^{*0}$ decays for the 2011–2012, 2015–2016, and 2017–2018 data samples, respectively. As these contributions are at per-mill level with respect to the expected $B^0_s \to \phi\mu^+\mu^-$ signal, they are neglected in the measurement and accounted for with a small systematic uncertainty.

The expected number of candidates above are given corresponding to what is expected in the full q^2 region. To further assess the corresponding q^2 distribution, several selection criteria are removed from the selection, *i.e.* the misID vetoes and the kaon PID requirement. For $B^0 \to J/\psi K^{*0}$ candidates, also the muon PID is removed. The comparison between expected yield for $B_s^0 \to \phi\mu^+\mu^-$ and doubly-misidentified $B_s^0 \to J/\psi\phi$ and $B^0 \to J/\psi K^{*0}$ candidates is shown in Fig. 5.9.

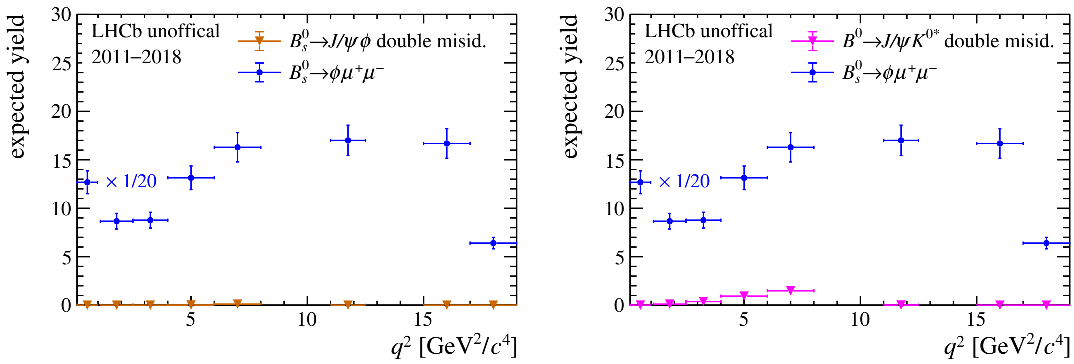

Figure 5.9: Upper limit for the number of doubly-misidentified $B_s^0 \to J/\psi\phi$ (left, orange) and $B^0 \to J/\psi K^{*0}$ decays (right, magenta) in the full $B_s^0 \to \phi\mu^+\mu^-$ data set. The $B_s^0 \to \phi\mu^+\mu^-$ signal expectation scaled down by a factor of 20 is shown in blue. The background is visibly small and accounted for as systematic uncertainty to the measurement.

5.2.5 S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$

Another source of background is from $B_s^0 \to K^+K^-\mu^+\mu^-$ decays where the dikaon system is in a spin-0 configuration, referred to as S-wave. This background is not distinguishable from the signal by its distribution of invariant masses $m(K^+K^-\mu^+\mu^-)$ since it stems from B_s^0-meson decays with very similar kinematics to the signal decays. However, the dikaon mass distribution $m(K^+, K^-)$ is flat in contrast to that of $B_s^0 \to \phi\mu^+\mu^-$ decays, where the candidates peak at the ϕ resonance; therefore, this background is suppressed largely by selecting the $B_s^0 \to \phi\mu^+\mu^-$ candidates with $|m(K^+, K^-) - m_\phi| < 12\,\mathrm{MeV}/c^2$.

The S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$ decays contribute to both data sets, *i.e.* via tree-level $B_s^0 \to J/\psi K^+K^-$ decays to the $B_s^0 \to J/\psi\phi$ data set and via the rare $B_s^0 \to K^+K^-\mu^+\mu^-$ mode to the $B_s^0 \to \phi\mu^+\mu^-$ data set. Since the initial- and final-state particles are the same between the two modes, the efficiencies are the same between the signal and this background when considering the candidates passing the dikaon mass requirement.

The branching fraction for S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$ decays has not been measured directly yet, nor are there reliable branching fraction predictions. However, the amplitude analysis of $B_s^0 \to J/\psi K^+K^-$ data in Ref. [133] measured the fraction of S-wave decays in the data to be $F_S^{J/\psi} = 1.1 \pm 0.1^{+0.2}_{-0.1}\%$, where the dikaon mass selection was identical to the one used in this analysis. This means that the same fraction of S-wave candidates is expected in this analysis' $B_s^0 \to J/\psi\phi$ data sample.

For the rare mode S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$ decays, no such measurement exists to date. Therefore, the number of S-wave candidates in the $B_s^0 \to \phi\mu^+\mu^-$ data sample is estimated to be the same as for the $B_s^0 \to J/\psi\phi$ data sample, *i.e.* $F_S^{\mu^+\mu^-,\,\text{full}\,q^2} = F_S^{J/\psi}$. The S-wave fraction is further estimated to be distributed

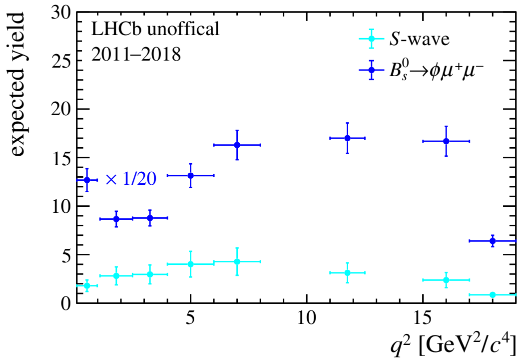

Figure 5.10: Expected number of rare S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$ decays (cyan) in the $B_s^0 \to \phi\mu^+\mu^-$ data sample, compared to the expected $B_s^0 \to \phi\mu^+\mu^-$ signal (blue). The $B_s^0 \to \phi\mu^+\mu^-$ signal expectation scaled down by a factor of 20.

across q^2 similarly to the angular observable F_L, so $F_S^{\mu^+\mu^-}(q^2)$ is given by

$$F_S^{\mu^+\mu^-}(q^2) = \frac{F_L^{\mu^+\mu^-}(q^2)}{F_L^{\mu^+\mu^-}} \times F_S^{\mu^+\mu^-,\,\text{full}\,q^2}, \tag{5.5}$$

with the q^2-dependent angular observable $F_L^{\mu^+\mu^-}(q^2)$ and the q^2-integrated angular observable $F_L^{\mu^+\mu^-}$, and the S-wave fraction in the full q^2 range $F_S^{\mu^+\mu^-,\,\text{full}\,q^2}$. The values for $F_L^{\mu^+\mu^-}(q^2)$ and $F_L^{\mu^+\mu^-}$ are taken from the FLAVIO software package described in Ref. [126].

The resulting number of expected S-wave $B_s^0 \to K^+K^-\mu^+\mu^-$ candidates in the $B_s^0 \to \phi\mu^+\mu^-$ data set are shown in Figure 5.10, where the number of candidates are shown resolved in q^2 and compared to the $B_s^0 \to \phi\mu^+\mu^-$ expected yields. This S-wave background is small with roughly 0.7%–1.7% candidates relative to the respective signal modes in the different q^2 intervals, so it is neglected and assigned for as systematic uncertainty to the measurement. The effect on the measurement is suppressed, since the contribution to $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ data samples cancel largely in the relative branching fraction measurement, as discussed in Sect. 5.4.3.

5.2.6 Combinatorial background

A further source of background contributions to the data samples is from randomly combined tracks, referred to as combinatorial background. This background is already largely suppressed by the dedicated requirement on the BDT classifier discussed in Sect. 4.3.5, but as the BDT's background rejection is below 100%, there are background candidates remaining. Part of these remaining background candidates are clearly visible in the data shown in Sect. 4.3.7, appearing as data candidates displaced from the known B_s^0 mass.

The combinatorial background contributes differently to the $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ data sets, since its origin is different for the two data samples: in the $B_s^0 \to J/\psi\phi$ sample, the combinatorial background decays include mostly muons from J/ψ meson decays that are combined with two random tracks; for $B_s^0 \to \phi\mu^+\mu^-$, the two muons are often from independent decay processes. Therefore, contribution from combinatorial background is not expected to cancel in the ratio of event yields used to measure the branching fraction.

Parametrising the invariant mass distribution of the combinatorial background decays is complicated by the fact that it stems from variety of different underlying decay processes. For this reason, there is also no simulation of combinatorial background available; but, the background can be assessed in data.

A sample of combinatorial background decays in data can be obtained by selecting invariant masses significantly larger than the B_s^0 meson mass; there, neither partially nor fully reconstructed B_s^0 or B^0 meson decays are expected to contribute. With this combinatorial background sample, it has been empirically shown that the invariant mass distribution of combinatorial background decays is in many cases well modelled with an exponential function. There has also been effort to motivate this description analytically, *e.g.* in Ref. [134], but the most convincing argument seems to be that it simply describes the data well.

5.3 Fitting for the normalisation mode yields and the branching fraction

As explained in the previous section, functions are fitted to the data to separate the signal $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ contributions from the combinatorial background, as explained generally in Sect. 4.4. These functions consist of a signal and a combinatorial background component, and they model the distribution of $m(K^+K^-\mu^+\mu^-)$.

The component describing the normalisation mode $B_s^0 \to J/\psi\phi$ and the signal mode $B_s^0 \to \phi\mu^+\mu^-$ decays is parametrised by a sum of two *Crystal Ball (CB)* functions [135] with shared mean value μ and tails to opposite sides, referred to as *double Crystal Ball (DCB)* function. The CB function $\mathrm{CB}(m;\mu,\sigma,\alpha,n)$ corresponds to a Gaussian function with mean μ and width σ with a power-law tail parametrised by α and n. The signal component is modelled with $f^{\mathrm{signal}}(m;\mathbf{p_1},\mathbf{p_2})$ is parametrised as

$$\begin{aligned} f^{\mathrm{signal}}(m;\mathbf{p_1},\mathbf{p_2}) &= \mathrm{DCB}(m;f_{\mathrm{CB}},\mathbf{p_1},\mathbf{p_2}) \\ &= f_{\mathrm{CB}} \times \mathrm{CB}_1(m;\mathbf{p_1}) + (1-f_{\mathrm{CB}}) \times \mathrm{CB}_2(m;\mathbf{p_2}), \end{aligned} \tag{5.6}$$

with the ratio between the two CB functions f_{CB}, the invariant mass of the B_s^0 candidate abbreviated as m, and the set of parameters for CB_i as $\mathbf{p_i}$. This model describes the signal as a Gaussian function with power-law tails [135].

The combinatorial background is described with an exponential function with decay constant a, *i.e.*

$$f^{\mathrm{comb.\ bkg.}}(m;a) = \frac{1}{a} \times \exp(m;a), \tag{5.7}$$

as explained in Sect. 5.2.6.

The above defined functions are used for both the $B_s^0 \to J/\psi\phi$ and the $B_s^0 \to \phi\mu^+\mu^-$ data samples. For the $B_s^0 \to J/\psi\phi$ data sample, the functions are fitted separately to the 2011–2012, 2015–2016, and 2017–2018 data samples, whereas for the $B_s^0 \to \phi\mu^+\mu^-$ data sample, the functions are fitted simultaneously to all data samples as described in Sect. 4.4.1.

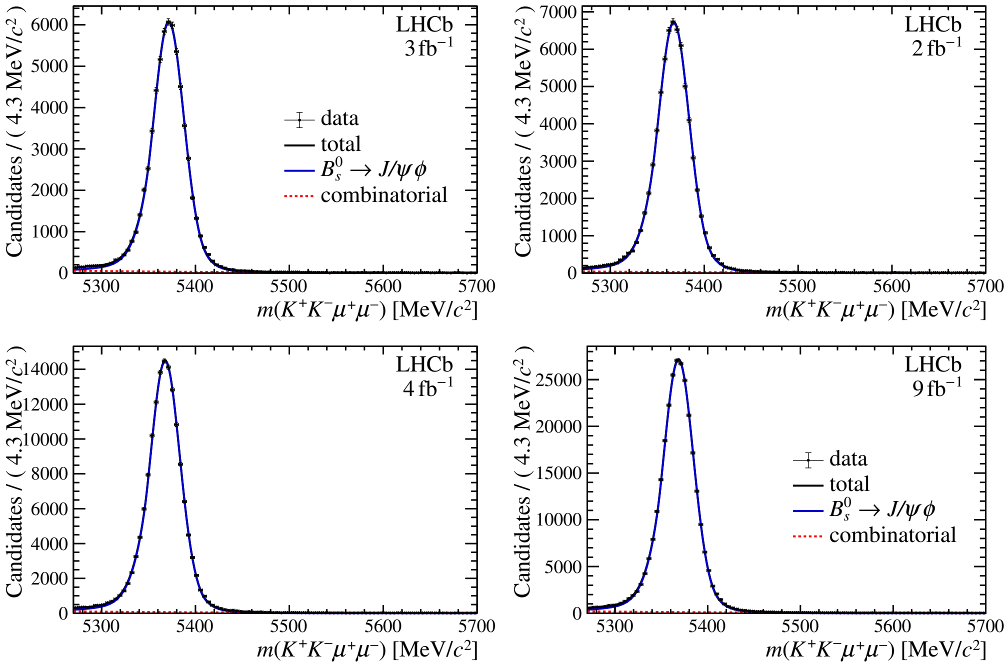

Figure 5.11: Invariant mass distribution of $B_s^0 \to J/\psi\phi$ data overlaid with the fitted model for the 2011–2012 (top left), 2015–2016 (top right), 2017–2018 (bottom left) and combined 2011–2018 data sample (bottom right). The latter plot has been published in Ref. [47]. The signal component of the fit model is shown in blue, and the combinatorial background component in dotted red.

5.3.1 Fits to the normalisation $B_s^0 \to J/\psi\phi$ decays

The full model that is fitted to the invariant mass distribution of the $B_s^0 \to J/\psi\phi$ data samples is given by

$$\begin{aligned} f(m; \mathbf{p_1}, \mathbf{p_2}, a) = N(B_s^0 \to J/\psi\phi) \times \mathrm{DCB}(m; f_{\mathrm{CB}}, \mathbf{p_1}, \mathbf{p_2}) \\ + N(\text{comb.bkg.}) \times \exp(m; a) \end{aligned} \tag{5.8}$$

with the number of $B_s^0 \to J/\psi\phi$ decays $N(B_s^0 \to J/\psi\phi)$ and combinatorial background decays N(comb. bkg.); the technical details with respect to the extended maximum likelihood fits can be found in Sect. 4.4.

To stabilise the fits, the tail parameters α_i, n_i, and the ratio between the two CB functions f_{CB} is determined from fits to simulated samples of $B_s^0 \to J/\psi\phi$, with the parameters for the three data samples are determined separately on the respective simulation samples, *e.g.* the parameters for the 2011–2012 fit are determined from a fit to a combination of 2011 and 2012 simulation. All other parameters are determined in the fit using an unbinned extended maximum likelihood fit implemented in the RooFit software package described in Ref. [118].

The resulting fitted functions are shown in Figure 5.11 overlaid on the distribution in data for all three $B_s^0 \to J/\psi\phi$ data samples. From these fits,

$$N(B_s^0 \to J/\psi\phi) = 62980 \pm 270 \ (2011 - 2012),$$
$$N(B_s^0 \to J/\psi\phi) = 70970 \pm 290 \ (2015 - 2016),$$
and
$$N(B_s^0 \to J/\psi\phi) = 148490 \pm 410 \ (2017 - 2018)$$

decays for the normalisation mode are found.

5.3.2 Simultaneous fit to $B_s^0 \to \phi\mu^+\mu^-$ data

The $B_s^0 \to \phi\mu^+\mu^-$ data is fitted with the same model as for $B_s^0 \to J/\psi\phi$ data, given in Eq. 5.8, but the signal yield is re-parametrised as function of the branching fraction as explained in Sect. 4.4.1. This way, the branching fraction parameter can be shared between the data samples in a simultaneous fit. The data of each q^2 interval is fitted separately, so in total, 11 fits are performed for the $B_s^0 \to \phi\mu^+\mu^-$ data sample: eight for the narrow q^2 intervals, two for the wide q^2 intervals, and one for the full q^2 range for visualisation.

Besides the branching fractions, only the combinatorial background yield N(comb. bkg) and the corresponding slope a is left floating in each fit; all other parameters, *i.e.* the tail parameters, fraction between CB functions f_{CB} and the DCB mean μ, are fixed from the $B_s^0 \to J/\psi\phi$ signal fit. Floating these fit parameters is not possible given the small candidate yields in the $B_s^0 \to \phi\mu^+\mu^-$ data sample. The widths of the DCB, σ_1 and σ_2, are also taken from the $B_s^0 \to J/\psi\phi$ mode, but they are modified with a correction factor to account for the q^2-dependent invariant mass resolution. This correction factor is determined for each q^2 bin separately by comparing the widths from $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ simulation.

The fitted functions are shown overlaid on the full data sample for the narrow q^2 intervals in Fig. 5.12, and for the wide intervals and full q^2 in Fig. 5.13. From these fits, the following number of signal $B_s^0 \to \phi\mu^+\mu^-$ decays are found in the full q^2 range:

$$N(B_s^0 \to \phi\mu^+\mu^-) = 458 \pm 12 \ (2011\text{–}2012),$$
$$N(B_s^0 \to \phi\mu^+\mu^-) = 484 \pm 13 \ (2015\text{–}2016),$$
and
$$N(B_s^0 \to \phi\mu^+\mu^-) = 1064 \pm 28 \ (2017\text{–}2018).$$

As already mentioned in Sect. 4.1, these data distributions, the fitted functions, and the candidate yields were not revealed until the data analysis was fixed completely and validated as presented in Sect. 5.5.

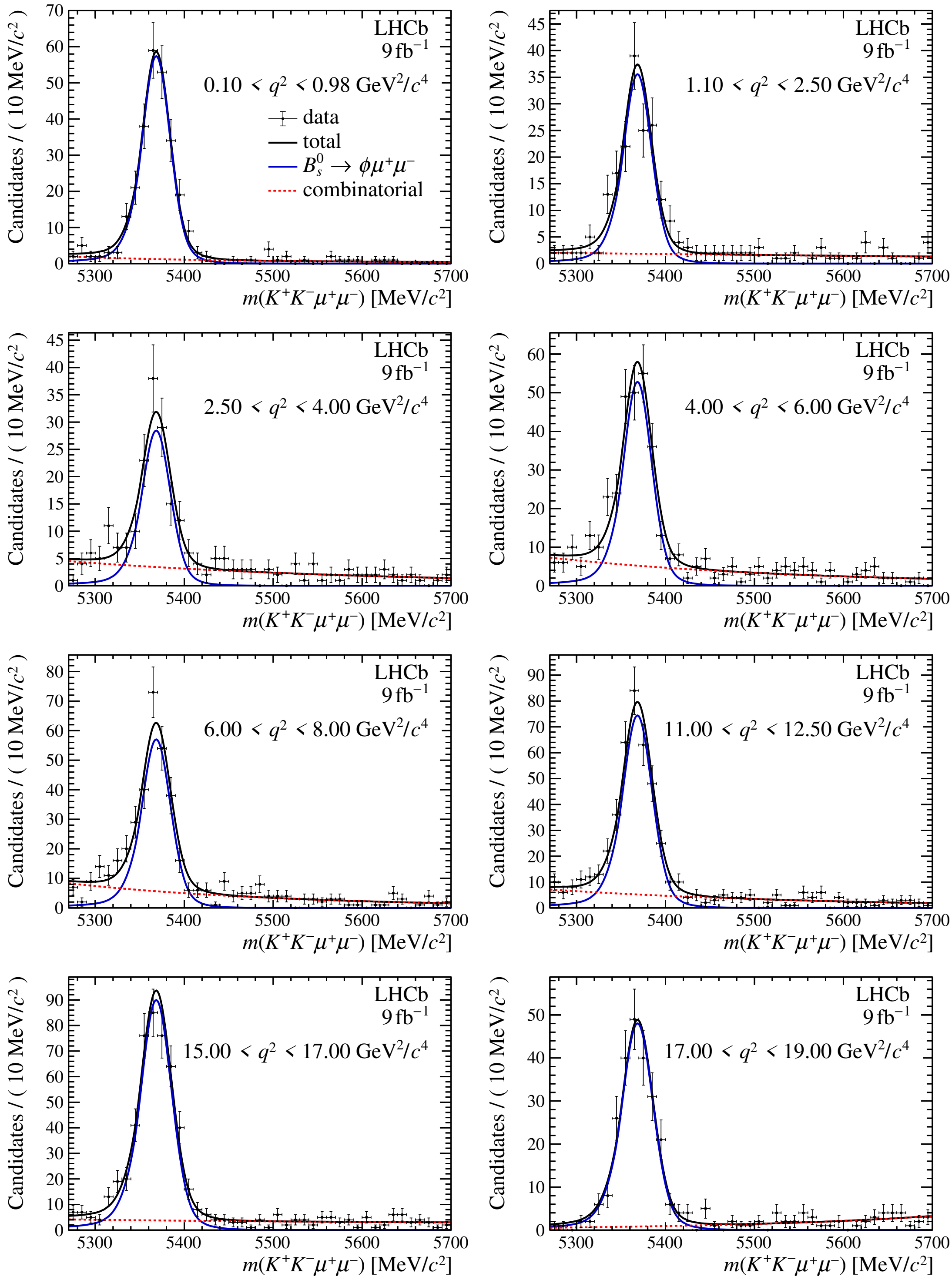

Figure 5.12: Invariant mass distribution of the B_s^0 candidates in the $B_s^0 \to \phi\mu^+\mu^-$ data sets for the different narrow q^2 bins in the combined 2011–2018 data sample. The fitted functions with the signal component in blue and the combinatorial background component in red is overlaid on the data. The plots have been published in Ref. [47].

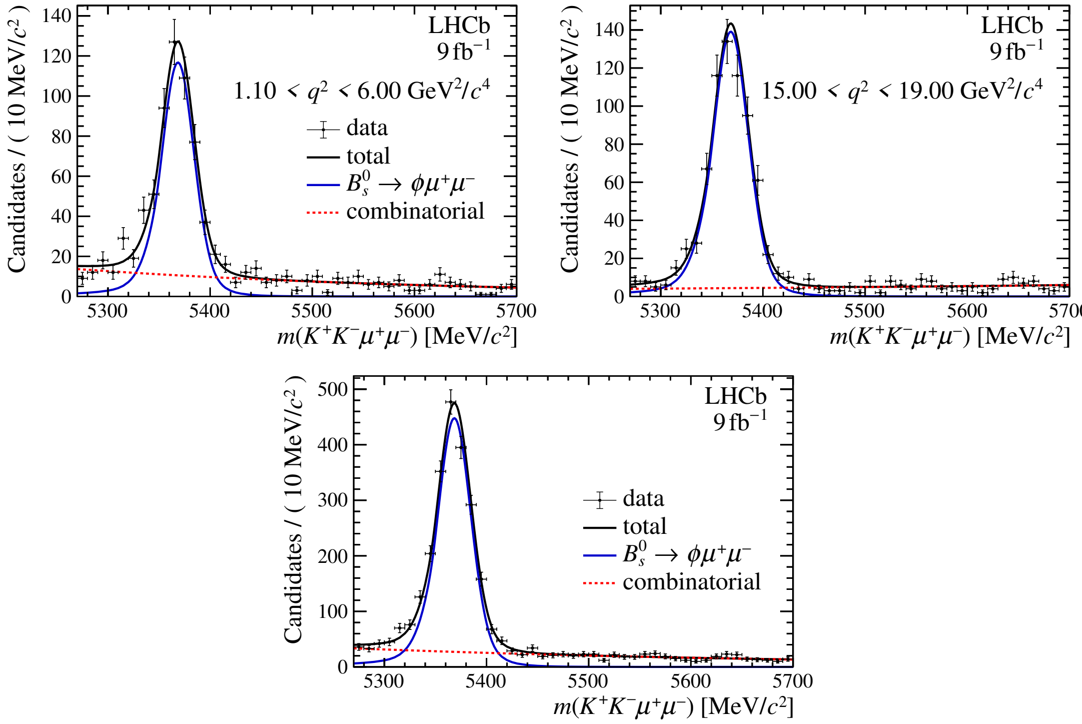

Figure 5.13: Invariant mass distribution of the B_s^0 candidates in the $B_s^0 \to \phi\mu^+\mu^-$ data sets for the different wide q^2 bins (top) and the full q^2 range (bottom) in the combined 2011–2018 data sample. The plots have been published in Ref. [47]. The fitted functions with the signal component in blue and the combinatorial background component in red is overlaid on the data.

5.3.3 Fit validation

As the branching fraction and its uncertainty is directly extracted from the fits described in the previous chapter, it is important to validate that the fitted parameters are unbiased which means that there is no systematic shift between the true and the fitted parameter. This validation is especially important since the maximum likelihood method only produces unbiased parameter estimates in the limit of large data samples [89], which is potentially not fulfilled anymore by the small $B_s^0 \to \phi\mu^+\mu^-$ data samples. In addition, since the uncertainties are also extracted from the fit, it is further important to check that these correctly describe the statistical fluctuations in the data.

The fit validation is performed as follows: many pseudodata samples, referred to as toy data samples, are generated according to the parameters of the fitted PDF. These toy data samples are fitted again, and the resulting fitted parameters p^{fit} are compared to the parameters used for generation p^{gen} in the *pull distribution*. The pull distribution is defined as the distribution of $(p^{\text{fit}} - p^{\text{gen}})/\sigma_{p^{\text{fit}}}$, and it is expected to follow a Gaussian distribution with mean 0 and standard deviation 1. If the fit returned (a) biased parameters or (b) under- or overestimated uncertainties, the pull distribution would indicate it with a systematic shift from zero in the former case, or standard deviations smaller or larger than 1 for the latter case.

A similar study on pseudodata samples has also been used to decide on the fit strategy for $B_s^0 \to \phi\mu^+\mu^-$. This study showed that fitting the three pseudodata

$B_s^0 \to \phi\mu^+\mu^-$ samples separately results in biased fit parameters and high fit failure rates. Fitting the $B_s^0 \to \phi\mu^+\mu^-$ pseudodata samples simultaneously with shared branching fraction parameter produced only small biases and negligible failure rate, so this fitting strategy was chosen for the measurement.

The fit-validation procedure is applied to all fits performed in this analysis, so to the three fitted functions for the $B_s^0 \to J/\psi\phi$ mode and to each of the 11 the simultaneously fitted $B_s^0 \to \phi\mu^+\mu^-$ functions. The toy studies are performed while keeping the fitted branching fractions in the data unrevealed, so only the differences between generated and fitted branching fraction normalised to the fitted uncertainty, referred to as *pull*, is checked. This pull distribution gives no access to the parameters found in the data themselves, so the fit validation could be performed before revealing the branching fraction result.

This study suggests that all fits perform well for the $B_s^0 \to J/\psi\phi$ mode and return unbiased candidate yields with well estimated uncertainties. For the $B_s^0 \to \phi\mu^+\mu^-$ fits, all biases on the branching fraction are at below per-mill level besides the fit to the last q^2 interval, $q^2 \in [17, 19]\,\text{GeV}^2/c^4$, where a small branching fraction bias of 0.7% is found relative to the measured branching fraction. As this bias is small compared to the statistical precision, it is left uncorrected and included as systematic uncertainty to the measurement.

5.4 Systematic uncertainties

This section discusses the systematic uncertainties associated with the measurement of $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$. They originate from uncertainties on the efficiencies, on the signal yields, and on the branching fractions $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to J/\psi\phi)$. The efficiencies are determined on simulated samples, so statistical fluctuations in the simulation samples, the specific applied corrections, and residual shortcomings of the simulated samples can result in systematic uncertainties. The signal yields are determined with fits to data, in which the neglected small background contribution, the specific choice for the signal shape, and the small observed fit biases could affect the branching fraction result. Lastly, the external branching fractions $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to J/\psi\phi)$ limit the branching fraction result by their precision. All the effects are studied in detail in this section and covered by the systematic uncertainty assigned to the measurement.

The systematic uncertainties are determined separately for each q^2 bin and for each of the potential causes. To directly assess the systematic uncertainty on the branching fraction, each systematic uncertainty σ_p on a parameter p, *e.g.* the efficiency ratio, is determined relative to p as this corresponds directly to the relative branching fraction uncertainty, *i.e.* $\sigma^{\mathcal{B}}_{\text{syst.}}/\mathcal{B}$:

$$\frac{\sigma^{\mathcal{B}}_{\text{syst.}}}{\mathcal{B}}(q^2) = \frac{\sigma^{p}_{\text{syst.}}}{p}(q^2). \tag{5.9}$$

In some cases, methods used in the analysis rely on choices, *e.g.* settings for the simulation weights or the signal model used in the fits. To estimate the effect of these choices, an alternative choice is used to derive $p^{\text{alternative}}$, and then compared to the parameter p^{default} obtained with the default approach used in the measurement. Their difference is assigned as systematic uncertainty to the measurement to cover the effect, *i.e.* $\sigma^{p}_{\text{syst}} = p^{\text{default}} - p^{\text{altern.}}$.

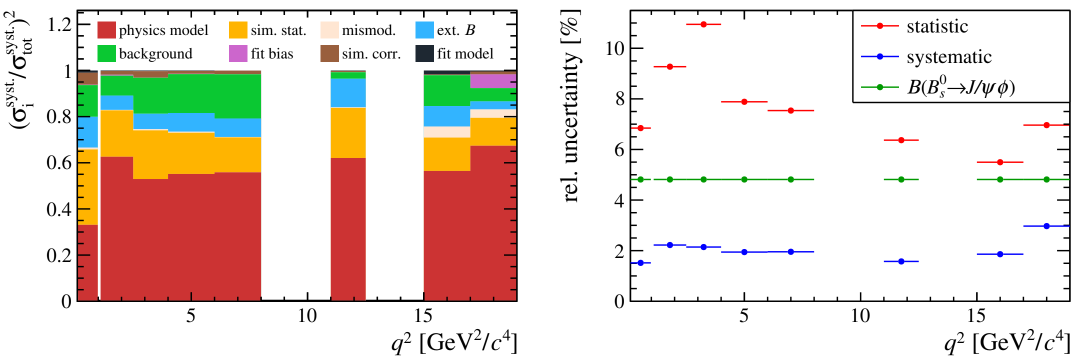

Figure 5.14: Comparison of (left) the different systematic uncertainty sources and (right) the statistical (red), systematic (blue) and normalisation branching fraction (green) uncertainty on the branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ for the different q^2 bins. The different systematic uncertainty sources are the limited knowledge of the physics model (red), the neglected background contribution in the data samples (green), the limited simulation sample size (orange), the fit biases (magenta), the residual mismodelling in the simulated samples (beige), the uncertainties associated to the simulation corrections (brown), the uncertainty on $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ (light blue), and the fit model choice (black).

The systematic effects are evaluated separately for each of the three data samples, *i.e.* 2011–2012, 2015–2016 and 2017–2018. These separate systematic uncertainties $\sigma^{p,i}_{\text{syst.}}$ are combined to the uncertainty on the full result $\sigma^{p}_{\text{syst.}}$ with

$$\sigma^{p}_{\text{syst.}} = \frac{\sum_i \sigma^{p,i}_{\text{syst.}}/(\sigma^{\mathcal{B},i}_{\text{stat.}})^2}{\sum_i 1/(\sigma^{\mathcal{B},i}_{\text{stat.}})^2}, \tag{5.10}$$

with the statistical uncertainty of the branching fraction using a single data sample, $\sigma^{\mathcal{B},i}_{\text{stat.}}$, which is obtained from fits to the separate data samples i. Equation 5.10 is designed to combine the systematic uncertainties $\sigma^{p,i}_{\text{syst.}}$ on the different data samples with full correlation: if the branching fractions $\mathcal{B}^i$ are all shifted by $\sigma^{p,i}_{\text{syst.}}$, the weighted average of the branching fractions $\mathcal{B}_i$ is shifted by $\sigma^{p}_{\text{syst.}}$.

The systematic uncertainties from different sources $\sigma^{p}_{\text{syst.}}$ are assumed as uncorrelated, so the combined uncertainty $\sigma^{\mathcal{B}}_{\text{syst.}}$ is is calculated by adding the individual relative uncertainties $\sigma^{p}_{\text{syst.}}/p$ squared. The uncertainty related to the normalisation branching fraction, $\mathcal{B}(B_s^0 \to J/\psi\phi)$, is quoted separately to simplify a $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ update when a more precise value for $\mathcal{B}(B_s^0 \to J/\psi\phi)$ becomes available. Figure 5.14 gives an overview on the different sources of systematic uncertainties and their relative size as well as a comparison between the statistical and systematic uncertainties. The numerical values for the absolute systematic uncertainties on the final branching fraction can be found in Table 5.1. In the following sections, the specific sources are discussed in detail.

5.4.1 Physics model

The largest systematic uncertainty on $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)/\mathcal{B}(B_s^0 \to J/\psi\phi)$ is from the limited knowledge of the underlying physics model for $B_s^0 \to \phi\mu^+\mu^-$ decays. This model determines the q^2, the angular, and the decay time distributions in $B_s^0 \to \phi\mu^+\mu^-$ simulation, which is used to determine the efficiencies. Therefore,

Source	$\sigma_{\text{syst.}}$ $[10^{-8}\,\text{GeV}^{-2}c^4]$
Physics model	0.04–0.10
Limited simulation sample	0.02–0.07
Residual background	0.01–0.04
$\mathcal{B}(J/\psi \to \mu^+\mu^-)$	0.01–0.04
Simulation corrections	0.00–0.03
Residual mismodelling	0.00–0.02
Fit bias	0.00–0.03
Signal fit model	0.00–0.01
Quadratic sum	0.05–0.12
Normalisation $\mathcal{B}(B_s^0 \to J/\psi\phi)$	0.11–0.37

Table 5.1: Systematic uncertainties on the differential branching fraction $\mathrm{d}\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)/\mathrm{d}q^2$. The ranges indicate the variations across the q^2 intervals.

limited knowledge on the physics model results in limited knowledge of the signal efficiencies.

Section 4.20 describes the correction applied to the $B_s^0 \to \phi\mu^+\mu^-$ simulation, which updates the form factor model and corrects the B_s^0 decay rate. This correction uses the form factors from Ref. [127] and the decay width difference $\Delta\Gamma_s$ from Ref. [136]. To determine the effect of these settings on the measurement, the efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ are re-calculated with changed settings for the correction, *i.e.* with alternative form factors and $\Delta\Gamma_s$ value, and the effect on the efficiency ratio is determined.

First, the effect from the form factors is assessed. For this, the efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ are re-calculated with form factors from Ref. [124], combined to efficiency ratios with $B_s^0 \to J/\psi\phi$ efficiencies, and compared to the default efficiency ratios, *i.e.* those obtained from the $B_s^0 \to \phi\mu^+\mu^-$ simulation corrected with the form factors from Ref. [127]. The resulting difference between those two settings is between 0.2–1.2% depending on the q^2 interval, where the largest difference appears for high q^2.

Then, the effect from the chosen $\Delta\Gamma_s$ value is assessed. For this, the efficiencies for $B_s^0 \to \phi\mu^+\mu^-$ are re-calculated with decay width difference $\Delta\Gamma_s$ varied by the difference between the default value from Ref. [136] and an alternative one from Ref. [89]; this variation changes the branching fraction by about 0.4–2.1% depending on q^2. As $\Delta\Gamma_s$ is only varied for $B_s^0 \to \phi\mu^+\mu^-$ and the default value from Ref. [136] is kept for $B_s^0 \to J/\psi\phi$, this uncertainty is estimated conservatively. Since the uncertainty is still small compared to the statistical uncertainties, the effect is not studied further and this conservative uncertainty is added to the measurement.

5.4.2 Limited simulation sample size

The efficiencies are determined on simulated samples with a finite number of decay candidates, therefore, the efficiencies are limited in statistical precision. This limited knowledge on the efficiencies is accounted for with a systematic uncertainty on the branching fraction result.

As explained in Sect. 4.5, the efficiency is determined as the number of candidates k that pass the selection relative to the total number of candidates N. The related uncertainty on the efficiencies σ_ϵ is determined by

$$\sigma_\epsilon = \frac{1}{N} \times \sqrt{k \times \left(1 - \frac{k}{N}\right)}, \tag{5.11}$$

which is a valid approximation for this analysis where N and k are moderately large [137]. The resulting relative uncertainty lies between 0.7–1.0%.

5.4.3 Residual background

As explained in Sect. 5.2, several small background contributions have been neglected when extracting the signal yields from the data. This simplification is accounted for with a systematic uncertainty.

The systematic uncertainty for the backgrounds $\sigma_{\text{bkg.}}$ is determined by comparing the nominal yield ratios r to the hypothetical background-corrected signal yield ratios r', *i.e.* where the background yields have been subtracted, namely

$$\sigma^{q^2}_{\text{bkg.}} = \underbrace{\frac{N(B^0_s \to \phi\mu^+\mu^-)^{q^2}}{N(B^0_s \to J/\psi\phi)}}_{\equiv\, r} - \underbrace{\frac{N(B^0_s \to \phi\mu^+\mu^-)^{q^2} - N(\mu^+\mu^- \text{ bkg.})^{q^2}}{N(B^0_s \to J/\psi\phi) - N(J/\psi \text{ bkg.})}}_{\equiv\, r'} \tag{5.12}$$

with the rare mode yields $N(B^0_s \to \phi\mu^+\mu^-)$ and background contribution to the rare (normalisation) mode $N(\mu^+\mu^-\ (J/\psi)\text{ bkg.})$ taken from the study on simulated samples explained in Sect. 5.2 and the normalisation mode yields $N(B^0_s \to J/\psi\phi)$ from data explained in Sect. 5.3.1.

This approach acknowledges the fact that if backgrounds contribute to both rare and resonant mode, they cancel partially in the ratio, *e.g.* for decays of Λ^0_b baryons that contribute to the $B^0_s \to \phi\mu^+\mu^-$ yields via loop level $\Lambda^0_b \to pK^-\mu^+\mu^-$ decays and to the $B^0_s \to J/\psi\phi$ yields via tree-level $\Lambda^0_b \to J/\psi pK^-$ decays. An overview on the considered backgrounds and their contributions to the $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to J/\psi\phi$ data sample is given in Table 5.2, along with the corresponding systematic uncertainties. The combined systematic uncertainty for neglecting the background contributions lies between 0.3–0.9%, depending on the q^2 interval.

5.4.4 External branching fractions

To determine the relative branching fraction of $B^0_s \to \phi\mu^+\mu^-$, the value for $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ is taken from Ref. [89], which is given by

$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033)\%. \tag{5.13}$$

The relative uncertainty on $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ of 0.6% directly affects the measurement of the total branching fraction $\mathcal{B}(B^0_s \to \phi\mu^+\mu^-)$.

contrib. to $N_{B_s^0 \to \phi\mu^+\mu^-}$	contrib. to $N_{B_s^0 \to J/\psi\phi}$	syst. [%]
$\Lambda_b^0 \to pK^-\mu^+\mu^-$ FCNC	$\Lambda_b^0 \to J/\psi pK^-$	0.2–0.6
$B^0 \to K^{*0}\mu^+\mu^-$ FCNC	$B^0 \to J/\psi K^{*0}$	0.1–0.4
$B^0 \to J/\psi K^{*0}$ double misID	–	0.0–0.5
$B_s^0 \to J/\psi\phi$ double misID	–	0.0–0.04
$B_s^0 \to K^+K^-\mu^+\mu^-$ S-wave FCNC	$B_s^0 \to J/\psi K^+K^-$	0.2–0.6

Table 5.2: Processes that contribute as background to the rare q^2 region (1st column) and the resonant q^2 region (2nd column) in data. The double-misID backgrounds do not contribute to the $B_s^0 \to J/\psi\phi$ data sample.

When measuring the absolute branching fraction of $B_s^0 \to \phi\mu^+\mu^-$, further input is needed from the normalisation branching fraction $\mathcal{B}(B_s^0 \to J/\psi\phi)$, measured as

$$\mathcal{B}(B_s^0 \to J/\psi\phi) = (1.018 \pm 0.049) \times 10^{-3} \tag{5.14}$$

in Ref. [138]. This relative uncertainty of 4.8% is the largest systematic uncertainty on the absolute branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$, and it is quoted separately to simplify conversion of the result once a more precise value for $\mathcal{B}(B_s^0 \to J/\psi\phi)$ becomes available.

5.4.5 Simulation corrections

As explained in Sect. 4.6, corrections are applied to the simulated samples used to determine the efficiencies. These corrections are limited in precision since they are calculated with finite-size data samples and based on certain choices that could affect the measurement.

This applies to the corrections of the track multiplicity, of B_s^0 p_{T}, and of the hardware trigger efficiencies, which are discussed in Sects. 4.6.3–4.6.5. These corrections rely on a certain binning scheme. The corresponding effect on the measurement is assessed by repeating the efficiency calculations with a changed number of bins for the corrections, *i.e.* with halved and doubled number of bins. The efficiency ratio varies by 0.1-0.3%, which is assigned as relative systematic uncertainty to the measurement.

The PID corrections, discussed in Sect. 4.6.1, use the PID distributions from data calibration samples as input. As these data samples contain only a finite amount of signal candidates, the distributions are associated with statistical uncertainties. To assess the effect of these uncertainties on the PID corrections, alternative corrected PID variables are generated from *bootstrapped* data calibration samples taken from the PIDCalib software package [85, 120]. Using these alternative PID variables in the selection changes the efficiency ratios by 0.2–0.5%, which is accounted for as systematic uncertainty.

Furthermore, as explained in Sect. 4.6.1, the PID correction relies on a parametrisation of the efficiency shapes in the data and simulation samples. The effect of this specific parametrisation choice is assessed by varying it and then using it to determine alternative corrected PID variables. The difference between the efficiency ratios determined using the default PID correction and this alternative PID correction is 0.0–0.3%, which is further assigned as a relative systematic uncertainty.

5.4.6 Residual mismodelling

Small differences between data and simulation are visible when comparing the tracking efficiency and the efficiency of the `isMuon` criterion. To estimate the effect on the branching fraction measurement, these differences are corrected for, and the efficiency ratios are re-calculated and compared to the nominal, uncorrected case. The difference is assigned as systematic uncertainty to the measurement.

The tracking efficiency is corrected by assigning weights to the simulated sample that are determined by comparing the tracking efficiency for the individual simulated tracks with the tracking efficiencies taken from $J/\psi \to \mu^+\mu^-$ decays from data. These tracking efficiencies and the corresponding correction weights are provided by the LHCb tracking group and determined as described in Ref. [70]. When the tracking efficiency weights are applied to the simulated $B^0_s \to J/\psi\phi$ and $B^0_s \to \phi\mu^+\mu^-$ samples, the efficiency ratio is changed by 0.1–0.6% depending on q^2.

Similarly, the efficiency of the `isMuon` criterion is corrected for each final-state muon by assigning weights to the simulated sample that align the efficiencies between data and simulation. The `isMuon` efficiencies in data are calculated on data calibration samples of $J/\psi \to \mu^+\mu^-$ decays as a function of the muon's momentum p and its pseudo-rapidity η. When applying the weights to align the `isMuon` efficiency in $B^0_s \to J/\psi\phi$ and $B^0_s \to \phi\mu^+\mu^-$ samples to that from the data calibration samples, the efficiency ratio is changed by 0.0–0.1% depending on the q^2 interval.

5.4.7 Fit bias

As explained in Sect. 5.3.3, the unbinned maximum log-likelihood fit returns a slightly biased branching fraction parameter. As the biases are small for all q^2 intervals, they are not corrected for, which is accounted for with a systematic uncertainty to the measurement. Depending on the q^2 interval, the effect is 0.0-0.7% relative to the measured branching fraction.

5.4.8 Signal fit model

Using a DCB function to describe the invariant mass distribution of the signal component in the data fit described in Sect. 5.3 is an analysis choice. The effect of this choice on the measurement is checked by using the sum of two Gaussian functions as alternative signal model and then comparing the result with that obtained from the DCB function.

This comparison is not performed directly on data since the resulting difference could be specific to statistical fluctuations in data. Instead, the difference is determined on pseudodata samples that are generated according to the default fit model. These pseudodata samples are fitted with the default and the alternative model, and on each sample, the different fit results are compared. The average difference between the signal yield ratios obtained from the default and the alternative model is 0.1–0.2% depending on the q^2 interval, and this is assigned as systematic uncertainty to the measurement.

5.5 Analysis validation

As described before, no checks are performed using the signal region from the $B_s^0 \to \phi\mu^+\mu^-$ data sample taken during 2015–2018 to avoid an experimenter's bias. Therefore, the final fit projections and results are only revealed when the analysis strategy is fully fixed. So, it is crucial to validate the analysis before revealing the results, and the analysis is validated by checking for internal consistency. In the LHCb collaboration, validation measures have become mandatory for these kind of *blinded* analyses. These measures are often part of the signal-revealing process during the internal peer-review where the analysis is checked for internal consistency or consistency with other measurements; this was also the case for this measurement. As such, validating the analysis was also a precondition in the the internal peer-review process for the publication of the result.

This analysis has been validated with two checks: first, the consistency between the branching fraction determined on the 2011–2012 data only is compared between this analysis and that published in Ref. [2] (Sect. 5.5.1); and second, the internal consistency between the results obtained from the 2011–2012 data is compared to the results from the newly added 2015–2018 data samples (Sect. 5.5.2).

5.5.1 Consistency of the results from 2011–2012 data

The branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ has already been measured by the LHCb collaboration using the 2011–2012 data samples in Ref. [2]. This result is now used to validate the updated analysis of the 2011–2012 data that is part of this analysis.

For this check, the branching fraction is determined on the 2011–2012 data sample only using the q^2-intervals that have been used in Ref. [2]. For simplicity, only the statistical uncertainties are considered in the following, as the statistical uncertainties are dominating.

As the measurements presented in this work and the analysis from Ref. [2] use data collected by LHCb, the data samples are largely identical. They are not completely identical since the selection of this analysis uses a re-optimised BDT and PID selection, but, depending on the q^2-interval, between 83% and 92% of the candidates used in this analysis have also been used in Ref. [2]. This overlap between the data samples is taken into account as detailed in Appendix C when comparing the two branching fraction results.

Both results are shown in Fig. 5.15 where also the overlap-corrected uncertainty on the difference between the results is indicated. The largest difference between the differential branching fractions is $1.5\,\sigma$ in $q^2 \in [15, 19]\,\mathrm{GeV}^2/c^4$, and the agreement of the total branching fractions is at $0.7\,\sigma$. Given that this comparison does not include any systematic uncertainties, the results are deemed to be in excellent agreement with each other.

5.5.2 Consistency between 2011–2012 and 2015–2018

The second check to validate this analysis is to compare the results obtained from the 2011–2012 and from the 2015–2018 data samples. This check was part of the result-revealing process during the internal peer-review of this analysis, and since it comprised the newly added 2015–2018 $B_s^0 \to \phi\mu^+\mu^-$ data sample for the first

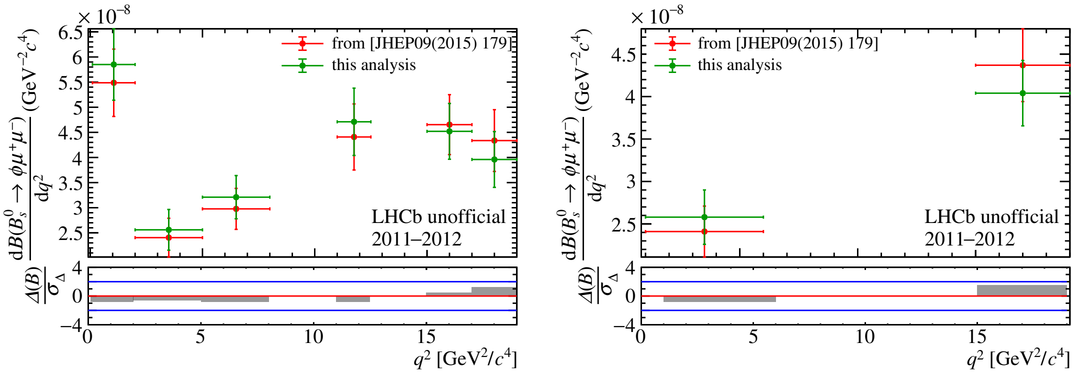

Figure 5.15: Comparison between the $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ result obtained on the 2011–2012 data set using this analysis framework (green) overlaid with the result from Ref. [2] (red). The difference between the two results is indicated below relative to the overlap-corrected uncertainty on the difference.

time, the check was performed in two stages: First, it was checked whether the two results agree at $< 3\sigma$ with each other, and only if this was confirmed, the actual size of the difference was revealed.

The consistency between the two $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ results is checked based on the *test statistics* $T^{\text{differential}}$ and T^{total} where the former is sensitive to the difference between the differential branching fractions, defined as

$$T^{\text{differential}} = \sum_{i\in[q^2\text{ bins}]} \left[\frac{\left(\mathcal{B}_i^{2011\text{–}2012} - \mathcal{B}_{i,j}^{2015\text{–}2018}\right)^2}{\sigma^2_{\mathcal{B}_i^{2011\text{–}2012}} + \sigma^2_{\mathcal{B}_i^{2015\text{–}2018}}} \right], \tag{5.15}$$

and the latter to the total branching fraction, defined as

$$T^{\text{total}} = \frac{\left(\sum\limits_{i\in[q^2\text{ bins}]} \left[\mathcal{B}_i^{2011\text{–}2012}\right] - \sum\limits_{i\in[q^2\text{ bins}]} \left[\mathcal{B}_i^{2015\text{–}2018}\right] \right)^2}{\sum\limits_{i\in[q^2\text{ bins}]} \left[\sigma^2_{\mathcal{B}_i^{2011\text{–}2012}} + \sigma^2_{\mathcal{B}_i^{2015\text{–}2018}} \right]}, \tag{5.16}$$

with the $B_s^0 \to \phi\mu^+\mu^-$ branching fraction $\mathcal{B}_i^j$ for the ith narrow q^2 bin from data set j and the corresponding statistical uncertainty $\sigma_{\mathcal{B}_i^j}$.

To calculate the values for $T^{\text{differential}}$ and T^{total} that correspond to the 3σ threshold, pseudodata sets are generated according to the 2011–2012 and 2015–2018 data samples that reflect the variations expected from the statistical and systematic uncertainties. For each pseudodata sample, $T^{\text{differential}}$ and T^{total} are determined, and their resulting distributions are shown in Fig. 5.16. These distributions are used to define the $3\,\sigma$ threshold, *i.e.* by finding the values of $T^{\text{differential}}$ and T^{total} where only 0.27% of pseudodata samples return a larger value. The resulting values are

$$\begin{gathered} T^{\text{differential}} \leq 25.12 \\ \text{and} \\ T^{\text{total}} \leq 9.02. \end{gathered} \tag{5.17}$$

The distributions of $T^{\text{differential}}$ and T^{total} are well described by χ^2 distributions

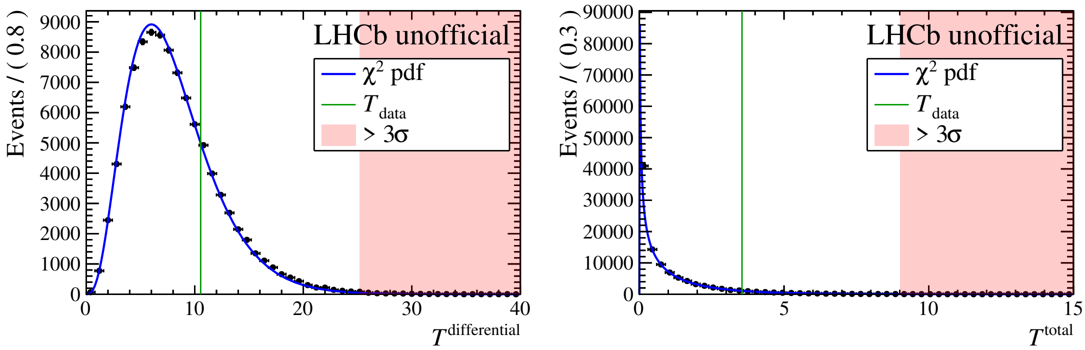

Figure 5.16: Distribution from (left) $T^{\text{differential}}$ and (right) T^{total} obtained from pseudo-data samples with the value obtained from data indicated (green line). The branching fraction results are in better agreement than 3σ, for which the region is marked in red. The distribution of $T^{\text{differential}}$ (T^{total}) is compatible with a χ^2 distributions with eight (one) degrees of freedom (blue line).

with eight and one degree of freedom, respectively. This confirms that the test variables behave as expected, *i.e.* that the branching fraction fits return generally unbiased branching fraction parameters with correct uncertainties.

The test parameters extracted from the data samples are smaller than the thresholds above, so the results pass the internal consistent check. The explicit values found in data are $T^{\text{differential}} = 10.54$ and $T^{\text{total}} = 3.55$, which means that the branching fractions from 2011–2012 and from 2015–2018 are in agreement at the level of $1.17\,\sigma$ and $1.87\,\sigma$, which corresponds to p-values of 0.24 and 0.06, respectively.

5.6 Branching fraction result

The resulting differential branching fractions relative to $\mathcal{B}(B_s^0 \to J/\psi\phi)$, extracted from the simultaneous fit explained in Sect. 5.3.2, are given in Table 5.3 for which the absolute branching fraction is obtained by multiplying the relative branching fraction with $\mathcal{B}(B_s^0 \to J/\psi\phi) = (1.018 \pm 0.049) \times 10^{-3}$ from Ref. [138].

The absolute branching fractions are further visualised in Fig. 5.17 where the measurement is overlaid with SM predictions: at low q^2, predictions based on light-cone sum-rule (LCSR) calculations from Ref. [127] are shown together with a more precise combination of LCSR and lattice calculations from Ref. [127]. At high q^2, predictions from lattice QCD from Ref. [26, 27] are shown.

In the low q^2 interval from 1.1 to $6.0\,\text{GeV}^2/c^4$, the measurement and the prediction agree at the level of $1.8\,\sigma$ when comparing to the LCSR-only result and at the level of $3.6\,\sigma$ when comparing to the combination of LCSR and lattice QCD.

The total branching fraction of $B_s^0 \to \phi\mu^+\mu^-$ is obtained by summing the differential branching fraction measurements across the q^2 range and by correcting their total for the excluded q^2 regions, *i.e.* the ϕ region $q^2 \in [0.98, 1.1]\,\text{GeV}^2/c^4$, the J/ψ region $q^2 \in [8, 11]\,\text{GeV}^2/c^4$, and the $\psi(2S)$ region $q^2 \in [12.5, 15.0]\,\text{GeV}^2/c^4$. The fraction of decays falling in the excluded q^2 regions $f_{q^2\,\text{veto}}$ is determined on the predicted q^2 distribution of $B_s^0 \to \phi\mu^+\mu^-$ decays, and it is determined as $f_{q^2\,\text{veto}} = 65.47 \pm 0.27\%$. The uncertainty reflects the difference between using

q^2 interval $[\mathrm{GeV}^2/c^4]$	$\mathrm{d}\mathcal{B}(B_s^0\to\phi\mu^+\mu^-)/\mathcal{B}(B_s^0\to J/\psi\phi)\mathrm{d}q^2$ $[10^{-5}\,\mathrm{GeV}^{-2}c^4]$	$\mathrm{d}\mathcal{B}(B_s^0\to\phi\mu^+\mu^-)/\mathrm{d}q^2$ $[10^{-8}\,\mathrm{GeV}^{-2}c^4]$
0.1–0.98	$7.61\pm0.52\pm0.12$	$7.74\pm0.53\pm0.12\pm0.37$
1.1–2.5	$3.09\pm0.29\pm0.07$	$3.15\pm0.29\pm0.07\pm0.15$
2.5–4.0	$2.30\pm0.25\pm0.05$	$2.34\pm0.26\pm0.05\pm0.11$
4.0–6.0	$3.05\pm0.24\pm0.06$	$3.11\pm0.24\pm0.06\pm0.15$
6.0–8.0	$3.10\pm0.23\pm0.06$	$3.15\pm0.24\pm0.06\pm0.15$
11.0–12.5	$4.69\pm0.30\pm0.07$	$4.78\pm0.30\pm0.08\pm0.23$
15.0–17.0	$5.15\pm0.28\pm0.10$	$5.25\pm0.29\pm0.10\pm0.25$
17.0–19.0	$4.12\pm0.29\pm0.12$	$4.19\pm0.29\pm0.12\pm0.20$
1.1–6.0	$2.83\pm0.15\pm0.05$	$2.88\pm0.15\pm0.05\pm0.14$
15.0–19.0	$4.55\pm0.20\pm0.11$	$4.63\pm0.20\pm0.11\pm0.22$

Table 5.3: Relative and total differential branching fraction result for $B_s^0\to\phi\mu^+\mu^-$ obtained from the combined 2011–2018 LHCb data sample.

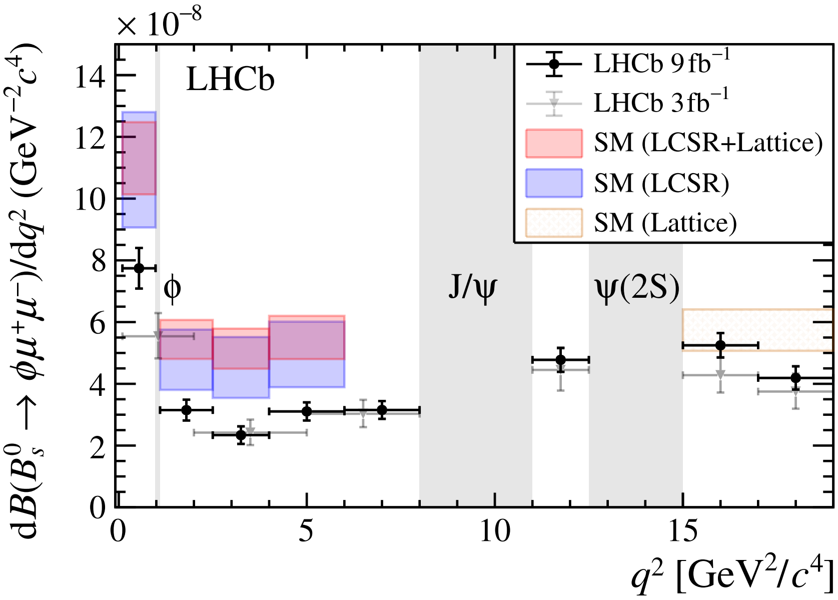

Figure 5.17: Branching fraction of $B_s^0\to\phi\mu^+\mu^-$ decays using the full 2011–2018 LHCb data sample. The result from Ref. [2, 138] is shown as comparison as well as different SM predictions based on LCSR calculations (blue) and a combination of LCSR and lattice calculations (red) for low q^2, and and from lattice calculations at high q^2. This plot has been published in Ref. [47].

form factors from Ref. [127] and from Ref. [124]. The total branching fractions are determined as

$$\frac{\mathcal{B}(B_s^0\to\phi\mu^+\mu^-)}{\mathcal{B}(B_s^0\to J/\psi\phi)} = (8.00\pm0.21\pm0.16\pm0.03)\times10^{-4}\,,$$
$$\mathcal{B}(B_s^0\to\phi\mu^+\mu^-) = (8.14\pm0.21\pm0.16\pm0.03\pm0.39)\times10^{-7},$$

with the statistical uncertainty, systematic uncertainty, the uncertainty from the extrapolation to full q^2, and, for the absolute branching fraction, the uncertainty related to the normalisation mode branching fraction $\mathcal{B}(B_s^0\to J/\psi\phi)$.

Chapter 6

First observation and branching fraction measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays

The strategy to measure $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays is similar to the one used for the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays. One difference is, however, that the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal is measured using the full invariant dimuon mass range, q^2; performing the measurement for different q^2 intervals is not feasible given the small data sample size. The analysis strategy is otherwise very similar to the analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays.

The signal efficiencies are determined on simulated samples of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays (Sect. 6.1), and background processes, which potentially contribute to the data samples, are studied (Sect. 6.2). Compared to the selected $B_s^0 \to \phi\mu^+\mu^-$ data samples, more background processes contribute to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data samples as a larger dikaon-mass window is needed to select the wider $f_2'(1525)$ resonance.

The larger amount of background processes is separated from the signal by performing a two-dimensional fit in the invariant dikaon and invariant B_s^0 mass distribution (Sect. 6.3), which takes into account the different background components.

This fitted PDF is not only used to determine the amount of signal in the data sample, which is needed for the branching fraction measurement; it is also used to calculate the significance of the signal peak using Wilks's theorem (Sect. 6.3.3). Since the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode has not been observed before, one goal of this analysis is also to observe this mode for the very first time.

Another potential method of probing for a significant contribution of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays in the data uses the method of moments (Sect. 6.16). The angular moments are sensitive to the spin configurations of the decay processes contributing to data, and the fact that the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays incorporate a spin-2 resonance can be used to distinguish it from background contributions without a spin-2 component. Since the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode is the only significant spin-2 decay expected in the selected data sample, the presence of such a spin-2 component would imply the presence of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays.

The systematic uncertainties are evaluated similarly to the $B_s^0 \to \phi\mu^+\mu^-$ measurement (Sect. 6.5), for which several systematic effects are the same as for the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays. However, the effects of the fit model are larger since the fitted PDF is more complex, and the underlying q^2 distribution is less well known than for $B_s^0 \to \phi\mu^+\mu^-$ decays.

Finally, the efficiencies and fit model are validated by determining the branching fraction of $B_s^0 \to J/\psi f_2'(1525)$ decays and comparing it to the value from the literature (Sect. 6.6), and the significance and branching fraction of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays is presented (Sect. 6.7).

6.1 Signal efficiencies

The signal efficiencies are determined as the fraction of simulated $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays that pass the selection, which is explained in detail in Sect. 4.5. These simulated signal samples are corrected for simulation-data differences in the track multiplicity, transverse momentum p_T of the B_s^0 meson, hardware trigger efficiency, particle identification variables (PID), and q^2 distributions as detailed in Sect. 4.6. Similarly, the efficiencies are determined for the $B_s^0 \to J/\psi f_2'(1525)$ control mode, which is used to construct the signal fit model and to validate the analysis. The corresponding simulated samples are corrected similarly to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ samples as explained in Sect. 4.6.

Even though the measurement is performed using the integrated q^2 range, the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ efficiencies are determined in the q^2 intervals used for the analysis of $B_s^0 \to \phi\mu^+\mu^-$; this is done to check the efficiency trends for consistency. The efficiencies are also determined separately for each year of data taking such that differences in the data taking, *e.g.* different centre-of-mass energies or trigger thresholds, can be considered. The resulting efficiencies, split by different selection stages, are shown in Fig. 6.1.

The LHCb acceptance, reconstruction and stripping, and trigger efficiencies are very similar between $B_s^0 \to f_2'(1525)\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$ decays. Clear differences, however, are visible in the preselection and the BDT efficiencies, which are significantly lower than for the selection of $B_s^0 \to \phi\mu^+\mu^-$ decays: the preselection efficiency for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays is between 73%–83%, compared to roughly 85% for $B_s^0 \to \phi\mu^+\mu^-$ decays; the BDT efficiency is at 80%–86% for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, compared to 92%–95% for $B_s^0 \to \phi\mu^+\mu^-$ decays. The smaller efficiencies stem from the tighter selection requirements for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, *i.e.* the stricter requirement on the kaon PID and the BDT, which is necessary to control the higher background levels.

The preselection efficiency differs across the years of data taking. This is directly connected to updates of the kaon PID tunings during the data taking periods, which alters the PID distributions. These differences across the years are also visible in $B_s^0 \to \phi\mu^+\mu^-$ simulation, however they do not affect the efficiencies significantly due to the less strict selection criterion.

The overall efficiency for the selection of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays is between 0.5% and 1.2% depending on the year of data-taking and the q^2 interval. This efficiency is lower than for $B_s^0 \to \phi\mu^+\mu^-$ decays with 0.8%–1.3%, which is driven by the stricter requirements in the pre- and BDT selection as explained above.

As the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ branching fraction is measured relative to the $B_s^0 \to J/\psi\phi$ mode, the selection efficiencies for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays

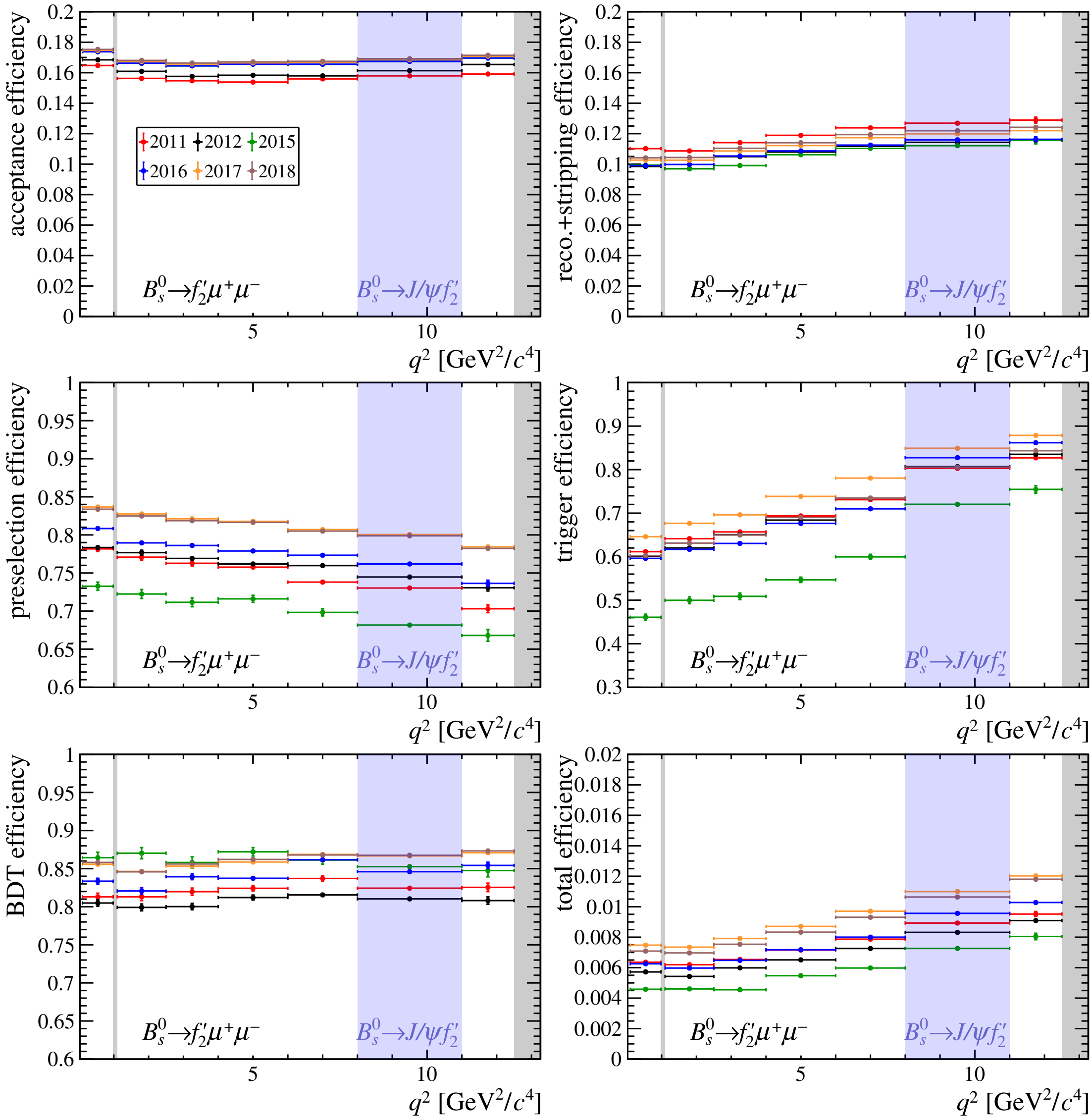

Figure 6.1: Signal efficiencies for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ and $B_s^0 \to J/\psi f_2'(1525)$ decays for the different data-taking conditions in the different stages: The LHCb acceptance efficiency (top left), the combined reconstruction and stripping efficiency (top right), the preselection efficiency (middle left), trigger efficiency (middle right), BDT efficiency (bottom left), and total efficiency (bottom right). The efficiencies for $B_s^0 \to J/\psi f_2'(1525)$ are determined using $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation and are indicated in the blue area. The other efficiencies are determined using $B_s^0 \to f_2'(1525)\mu^+\mu^-$ simulation.

are needed relative to the $B_s^0 \to J/\psi\phi$ efficiencies. For this, the $B_s^0 \to J/\psi\phi$ efficiencies are determined as in Sect. 5.1, and the resulting efficiency ratios are shown in Fig. 6.2 for both the individual years and combined data-taking periods. This combination is performed similarly to what is done for the $B_s^0 \to \phi\mu^+\mu^-$ measurement described in Sect. 5.1.

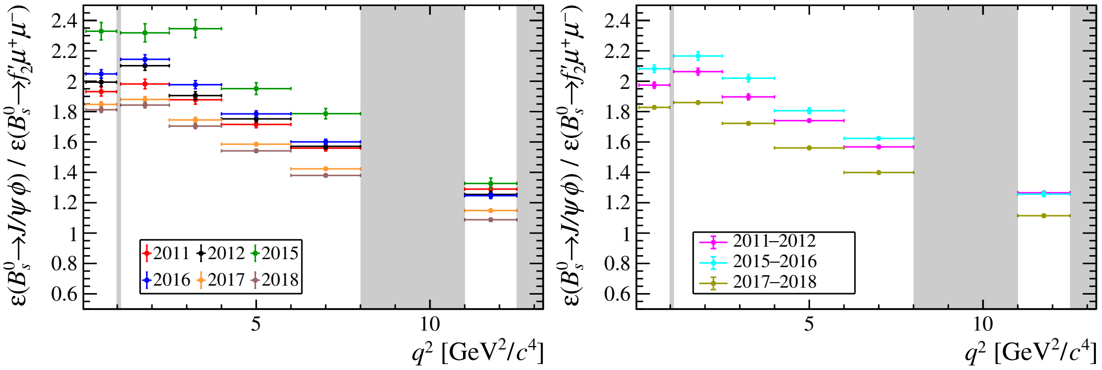

Figure 6.2: Ratios between the efficiencies for the normalisation mode $B_s^0 \to J/\psi\phi$ and the rare mode $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. The efficiency ratios are (left) determined for the separate years of data-taking and then (right) combined for the branching fraction measurement using the combined 2011–2012, 2015–2016 and 2017–2018 data sets.

6.2 Background processes

The amount of expected background decays determines the strategy on how to extract the signal for the measurement: if the background levels are low enough to affect the extracted yields only insignificantly, the background can be neglected for the measurement and taken into account as systematic uncertainty. If the background levels are larger, the background needs to be considered when extracting the signal yields. This can be done by fitting a PDF to the distribution of a discriminating variable and by that distinguish the signal from the background decays as explained in Sect. 4.4.

As mentioned previously, larger background levels are expected in the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ samples compared to the $B_s^0 \to \phi\mu^+\mu^-$ samples due to the larger selected dikaon mass range. Even though the criteria on the PID and BDT response are tightened compared to the $B_s^0 \to \phi\mu^+\mu^-$ selection, several sources of background give significant contributions to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data samples.

Similar to what is done in Sect. 5.2 for the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decay, this section investigates the different decay processes that contribute to the data samples and estimates their contributions relative to the $B_s^0 \to J/\psi\phi$ decays via

$$N_{\text{bkg.}} = \frac{N^{\text{data}}_{B_s^0 \to J/\psi\phi}}{\epsilon_{B_s^0 \to J/\psi\phi}} \times \frac{f_{\text{bkg.}}}{f_s} \times \epsilon_{\text{bkg.}} \times \frac{\mathcal{B}(\text{bkg.})}{\mathcal{B}(B_s^0 \to J/\psi\phi)}, \tag{6.1}$$

with the number of $B_s^0 \to J/\psi\phi$ decays in data $N^{\text{data}}_{B_s^0 \to J/\psi\phi}$, the efficiencies $\epsilon_{B_s^0 \to J/\psi\phi}$ ($\epsilon_{\text{bkg.}}$), branching fractions $\mathcal{B}(B_s^0 \to J/\psi\phi)$ ($\mathcal{B}(\text{bkg.})$) and fragmentation fraction f_s ($f_{\text{bkg.}}$) for the $B_s^0 \to J/\psi\phi$ (background) decays.

This section is structured as follows: first, an estimate of the expected $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal yield is given in Sect. 6.2.1, followed by a discussion of the relevant background contributions summarised in Fig. 6.3: contributions from (1) $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays where the proton is mis-identified as kaon in Sect. 6.2.2; from (2) $B^0 \to K^+\pi^-\mu^+\mu^-$ where the pion is mis-identified as kaon in Sect. 6.2.3; from (3) $B_s^0 \to K^+K^-\mu^+\mu^-$ decays that do not proceed via the $f_2'(1525)$ resonance in Sect. 6.2.4; and from (4) randomly combined tracks, referred to as combinatorial background, in Sect. 6.2.5.

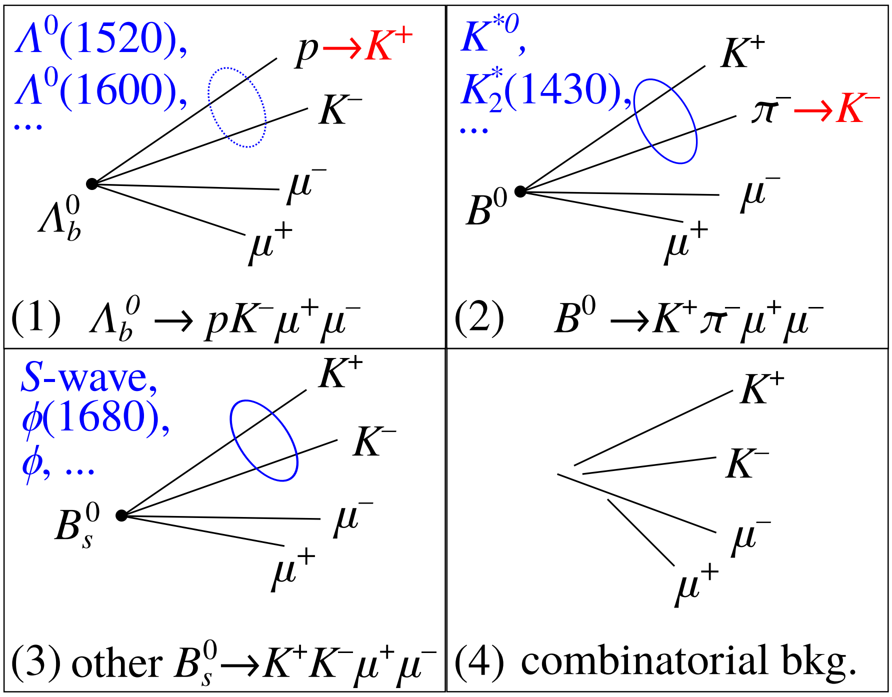

Figure 6.3: Overview of the background sources relevant to the analysis of $B^0_s \to f_2'(1525)\mu^+\mu^-$. Expected contribution are from (1) Λ^0_b decays with the proton being misidentified as a kaon, for which different Λ^* resonances contribute; from (2) B^0 decays with the pion being misidentified as kaon; from (3) $B^0_s \to K^+K^-\mu^+\mu^-$ decays without $f_2'(1525)$ resonance; and from (4) random track combinations.

6.2.1 Signal yield expectation

The branching fraction of the decay $B^0_s \to f_2'(1525)\mu^+\mu^-$ has not been measured yet, so this estimate relies on the branching fraction prediction from Ref. [128], which is

$$\mathcal{B}(B^0_s \to f_2'(1525)\mu^+\mu^-) = (2.13 \pm 0.43) \times 10^{-7}.$$

Combining this value with the efficiencies obtained from $B^0_s \to f_2'(1525)\mu^+\mu^-$ simulated samples and using Eq. 6.1, one obtains 85 ± 15, 92 ± 19, and 220 ± 50 expected $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays in the 2011–2012, 2015–2016 and 2017–2018 data samples, respectively.

6.2.2 $\Lambda^0_b \to pK^-\mu^+\mu^-$

Decays of Λ^0_b baryons via $\Lambda^0_b \to pK^-\mu^+\mu^-$ can contribute to the data sample if the final-state proton is misidentified as a kaon. The amount of $\Lambda^0_b \to pK^-\mu^+\mu^-$ background decays is determined similarly as for the $B^0_s \to \phi\mu^+\mu^-$ mode in Sect. 5.2.2, using the same simulated samples to determine the efficiencies for passing the reconstruction and selection. These simulated samples have been corrected for mismodelling of the distribution of $m(pK^-)$ with $\Lambda^0_b \to J/\psi pK^-$ data and for the distribution of q^2 using a prediction for the q^2 distribution taken from Ref. [130], as detailed in Sect. 5.2.2.

Using Eq. 6.1, the resulting efficiencies are combined with the inclusive branching fraction estimate $\mathcal{B}(\Lambda^0_b \to pK^-\mu^+\mu^-) = (2.9 \pm 1.1) \times 10^{-7}$, which is derived in

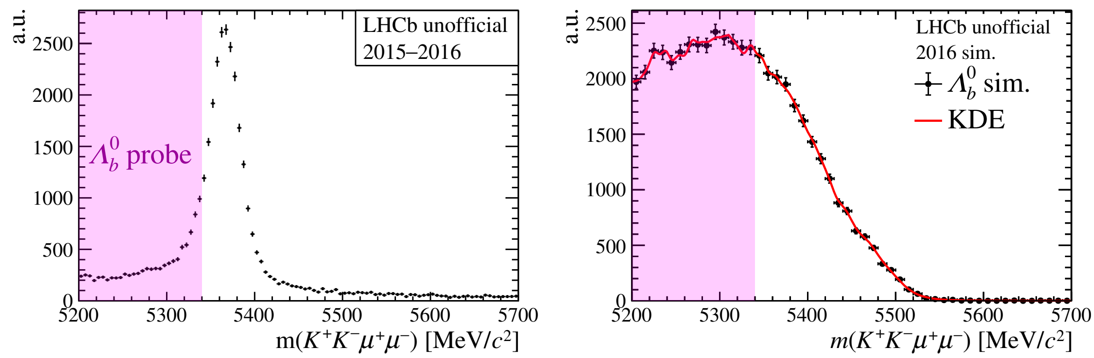

Figure 6.4: Distribution of the reconstructed invariant mass m$(K^+K^-\mu^+\mu^-)$ for (left) the 2015–2016 $B_s^0 \to J/\psi f_2'(1525)$ data sample and (right) $\Lambda_b^0 \to J/\psi pK^-$ simulation with 2016 conditions. The mass region that is used to determine the $\Lambda_b^0 \to J/\psi pK^-$ contamination in data is indicated in magenta, and the simulation is modelled with kernel density estimates (red line).

Sect. 5.7. From that, 19.2 ± 8.0, 22.9 ± 9.6, and 57 ± 24 $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays are expected in the 2011–2012, 2015–2016, and 2017–2018 $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data samples. In the signal region of $\pm 50\,\mathrm{MeV}/c^2$ around the B_s^0-meson mass, these yields correspond to roughly 22–27% relative to the expected $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal. This fraction is too large to neglect it in the following, so this background contribution is modelled in the fit to data.

A study on the expected $B_s^0 \to f_2'(1525)\mu^+\mu^-$ fit stability showed that fitting the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ component without any constraints results in high failure rates, therefore, the yield of the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ component is constrained to its expected value. Since this expected value relies on the efficiencies determined on the simulated samples, it is crucial to check that the simulation agrees with the data.

For this check, the number of Λ_b^0 background decays is determined on data, and this number is compared to the expected background contamination estimated with Eq. 6.1. This check cannot be carried out using the rare mode data as the number of data candidates in the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ sample is very small. Instead, the $B_s^0 \to J/\psi f_2'(1525)$ data is used to extract the number of $\Lambda_b^0 \to J/\psi pK^-$ decays from data, which are then compared to the number of $\Lambda_b^0 \to J/\psi pK^-$ decays expected from Eq. 6.1. Even though this check does not directly validate the calculation for the rare signal modes, it is still considered useful to get an idea about the agreement between data and simulation of the rare mode background contribution, since the decays are very similar as they have an identical final state.

Validation of Λ_b^0 yield with data The Λ_b^0 background yield is extracted from data in a sub-range range of reconstructed invariant masses $m(K^+K^-\mu^+\mu^-)$, referred to as Λ_b^0 probe region and shown in Fig. 6.4. This region provides a background sample with relatively few signal, and it contains 83% of the $\Lambda_b^0 \to J/\psi pK^-$ decays that contribute to the $B_s^0 \to J/\psi f_2'(1525)$ data sample, where this fraction is determined on simulated samples of $\Lambda_b^0 \to J/\psi pK^-$ decays, for which an example invariant mass distribution is shown in Fig. 6.4.

The number of $\Lambda_b^0 \to J/\psi pK^-$ data candidates is determined in the Λ_b^0-

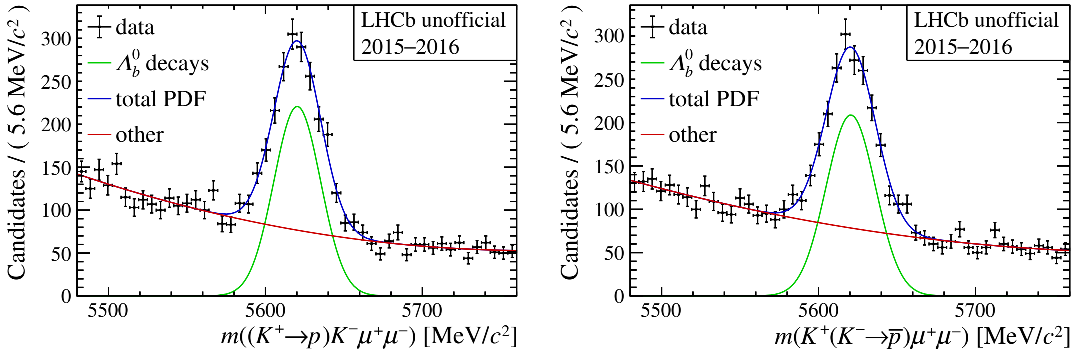

Figure 6.5: Reconstructed invariant masses of $B_s^0 \to J/\psi f_2'(1525)$ candidates with the positively (negatively) charged kaon having assigned the proton mass shown left (right) selected with $m(K^+K^-\mu^+\mu^-)$ in the Λ_b^0 probe region. The data from the 2015–2016 sample is shown as example. The distributions are modelled with a Gaussian function for the peak corresponding to Λ_b^0 decays and with a polynomial function for the other contributions.

baryon probe region by reconstructing the Λ_b^0 baryons directly, which means that the Λ_b^0 invariant masses are calculated with $m(K^+(K^- \to \overline{p})\mu^+\mu^-)$ and $m((K^+ \to p)K^-\mu^+\mu^-)$, which is the invariant mass with one kaon assigned the proton mass. The distribution of these invariant masses is shown in Fig. 6.5. From these distributions, the $\Lambda_b^0 \to J/\psi pK^-$ yields are extracted from a fit that includes a Gaussian function to describe the Λ_b^0-baryon decays and a polynomial function to describe the contribution from other decay processes.

The Λ_b^0 yield is extrapolated to the full range using the fraction of candidates in the Λ_b^0 probe region using the fraction of 83% explained above. This way, 3100 ± 120, 3370 ± 120, and 7710 ± 170 decays via $\Lambda_b^0 \to J/\psi pK^-$ mode are found in the 2011–2012, 2015–2016, and 2017–2018 data samples, respectively. These numbers are in good agreement with 3080 ± 980, 3800 ± 980, and 9200 ± 2400 expected candidates, which are obtained from Eq. 6.1 with branching fraction expectations and efficiencies purely from simulation.

6.2.3 $B^0 \to K^+\pi^-\mu^+\mu^-$

B^0 mesons can contribute via $B^0 \to K^+\pi^-\mu^+\mu^-$ decay to the selected data if the pion is misidentified as a kaon. These background decays were also studied for the $B_s^0 \to \phi\mu^+\mu^-$ data sample; there, the background stems mostly from decays via $B^0 \to K^{*0}\mu^+\mu^-$ modes. For the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ sample, however, the $K^+\pi^-$ system can be from different resonances; *e.g.* in the observation of $B_s^0 \to J/\psi f_2'(1525)$ decays in Ref. [139], significant contribution from $B^0 \to J/\psi K_2^*(1430)^0$ decays were found.

The inclusive background from $B^0 \to K^+\pi^-\mu^+\mu^-$ decays contributing to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample is estimated using the inclusive branching fraction $\mathcal{B}(B^0 \to K^+\pi^-\mu^+\mu^-) = (6.9 \pm 1.4) \times 10^{-7}$, which is obtained from scaling the resonant-mode branching fraction $\mathcal{B}(B^0 \to J/\psi K^+\pi^-) = (1.15 \pm 0.05) \times 10^{-3}$ from Ref. [140] by an estimated resonant-to-rare factor 100 ± 20, which is based on comparisons between other rare and corresponding resonant decay processes.

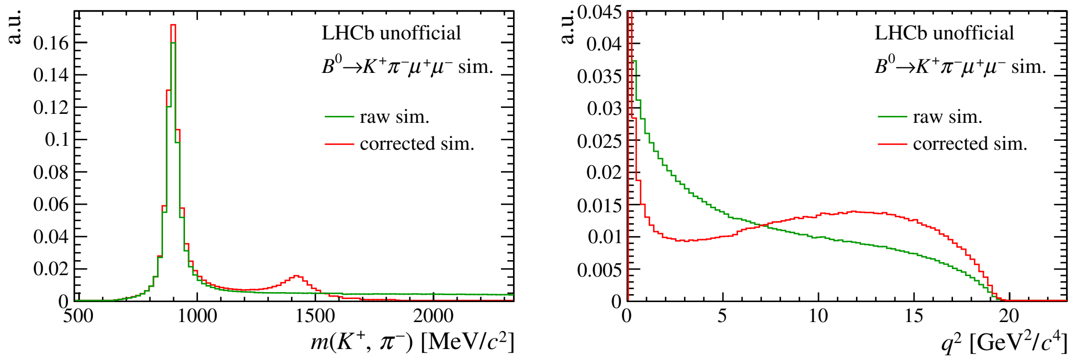

Figure 6.6: Distributions of (left) invariant kaon-pion masses $m(K^+\pi^-)$ and (right) invariant dimuon mass squared q^2 from a mixture of phase-space $B^0 \to K^+\pi^-\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ simulation. The distributions are shown before any correction (green) and after applying weights to correct the distributions (red) to resemble the dikaon masses found in Ref. [141] and the q^2 distribution calculated with form factors from Ref. [124].

The efficiency for selecting the $B^0 \to K^+\pi^-\mu^+\mu^-$ candidates is determined on simulated samples of $B^0 \to K^+\pi^-\mu^+\mu^-$ decays. The simulation used consists of a mixture of simulated $B^0 \to K^{*0}\mu^+\mu^-$ decays, as used in Sect. 5.2.3, and $B^0 \to K^+\pi^-\mu^+\mu^-$ decays that are modelled with a phase-space model, which neglects the resonance spectrum of the $K^+\pi^-$-system. To correct the $K^+\pi^-$ spectrum, the samples are assigned weights such that the corrected $m(K^+\pi^-)$ spectrum is similar to the spectrum observed in $B^0 \to J/\psi K^+\pi^-$ data in Ref. [141].[1] The distribution before and after assigning the weights is shown in Fig. 6.6.

Furthermore, the q^2 distribution is also modelled considering only the available phase space. This distribution is corrected similarly, *i.e.* by assigning weights so that the q^2 distribution is similar to what is expected from a rare FCNC decay. As a benchmark, the q^2 distribution is taken from $B^0 \to K^{*0}\mu^+\mu^-$ decays with form factors from Ref. [124]. The distribution before and after the correction is also shown in Fig. 6.6.

Using the efficiencies determined on these simulated samples and the branching fraction for $B^0 \to K^+\pi^-\mu^+\mu^-$ in Eq. 6.1, one obtains 42.5 ± 9.4, 23.3 ± 5.2, and 72 ± 16 expected $B^0 \to K^+\pi^-\mu^+\mu^-$ decays in the in the 2011–2012, the 2015–2016, and the 2017–2018 data samples, respectively. This corresponds to 14–38% relative to the expected $B^0_s \to f_2'(1525)\mu^+\mu^-$ signal in a window of $\pm 50\,\mathrm{MeV}/c^2$ around the B^0_s mass, so this background component is too large to neglect.

Therefore, the $B^0 \to K^+\pi^-\mu^+\mu^-$ contribution is separated from the signal by modelling it in the fit to data. To stabilise the fit, the $B^0 \to K^+\pi^-\mu^+\mu^-$ yields are Gaussian-constrained to the expected yields in the full fit region. Similar to what was done for the Λ^0_b background decays in Sect. 6.2.2, the B^0 candidate yields are validated as detailed below by comparing the yield of $B^0 \to J/\psi K^+\pi^-$ candidates extracted from $B^0_s \to J/\psi f_2'(1525)$ data and determined using Eq. 6.1, which is purely based on branching fractions and efficiencies from simulation.

[1] These weights were extracted on pseudo-data generated according to Ref. [141], provided by Th. Oeser.

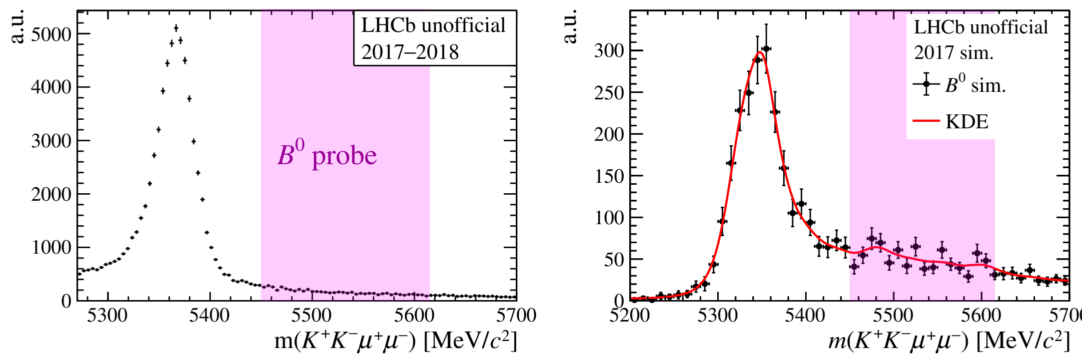

Figure 6.7: Distribution of the reconstructed invariant mass $\mathrm{m}(K^+K^-\mu^+\mu^-)$ for (left) the 2017–2018 $B_s^0 \to J/\psi f_2'(1525)$ data sample and (right) $B^0 \to J/\psi K^+\pi^-$ simulation with 2017 conditions. The mass region that is used to determine the $B^0 \to J/\psi K^+\pi^-$ contamination in data is indicated in magenta, and the distribution in simulation is modelled with a kernel density estimate (red line).

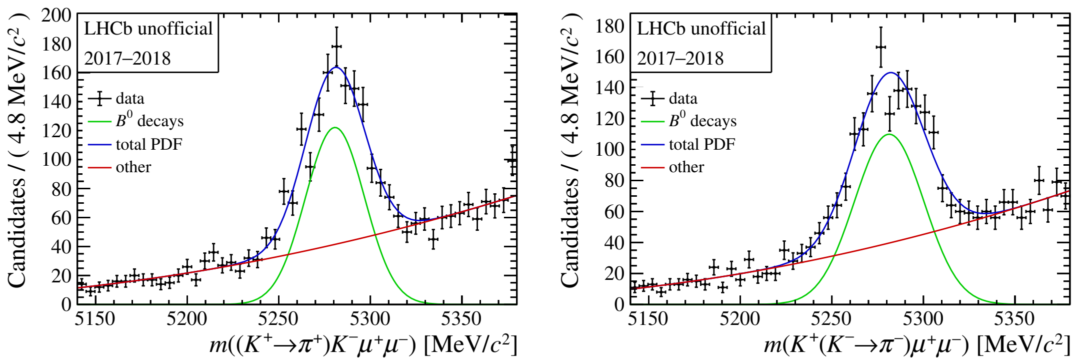

Figure 6.8: Reconstructed invariant masses of $B_s^0 \to J/\psi f_2'(1525)$ candidates with the positively (negatively) charged kaon having assigned the pion mass shown left (right) selected with $m(K^+K^-\mu^+\mu^-)$ in the B^0 probe region. The data from the 2017–2018 sample is shown as example. The distributions are modelled with a Gaussian function for the peak corresponding to B^0 decays and with a polynomial function for the other contributions.

Validation of B^0 yield with data The $B^0 \to J/\psi K^+\pi^-$ background yield is determined from data in a B^0 probe region, shown in Fig. 6.7. This B^0 background yield corresponds to approximately 28% of the total number of candidates expected in the $B_s^0 \to J/\psi f_2'(1525)$ data sample, for which the fraction is determined on $B^0 \to J/\psi K^+\pi^-$ simulation, for which the $m(K\pi)$ distribution has been corrected similar to what was done for simulated $B^0 \to K^+\pi^-\mu^+\mu^-$ decays, as explained above. An example for the invariant mass distribution of $B^0 \to J/\psi K^+\pi^-$ decays is shown in Fig. 6.7.

To determine the background yield in data, the B^0 mesons are directly reconstructed by calculating their invariant mass $m((K^+ \to \pi^+)K^-\mu^+\mu^-)$, *i.e.* the invariant mass where one kaon is assigned the pion mass. The resulting distributions are shown in Fig. 6.8.

From these distribution, the B^0 yields are extracted from a fit with a Gaussian function to describe the B^0 decays and a polynomial function to describe

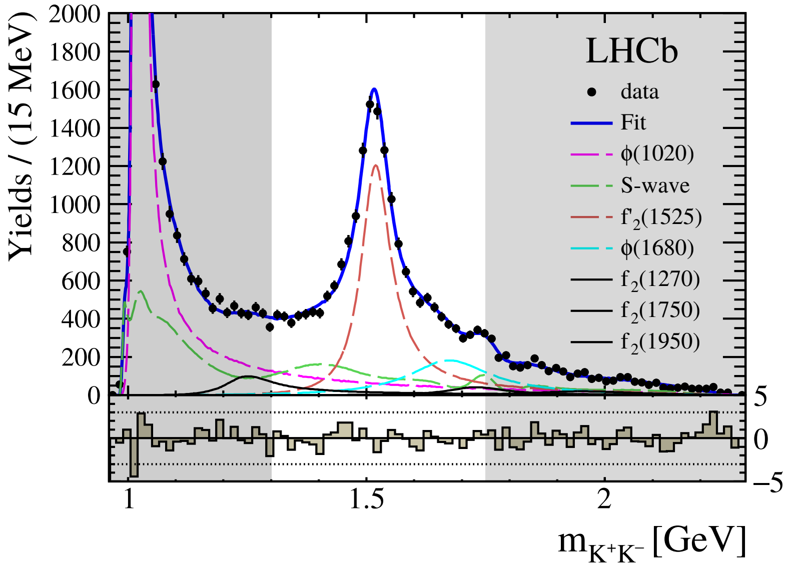

Figure 6.9: Background-subtracted $B_s^0 \to J/\psi K^+K^-$ data overlaid with the fitted PDF projection taken from Ref. [142], for which the dikaon mass regions rejected in this work are marked in gray. Contribution from $B_s^0 \to J/\psi K^+K^-$ decays with the two kaons that originate from S-wave, ϕ and $\phi(1680)$ decays are clearly contributing to the dikaon mass range considered in this analysis.

the remaining contributions. The B^0 yield is extrapolated to the full invariant $\mathrm{m}(K^+K^-\mu^+\mu^-)$ region by using the fraction of events in the B^0 probe region explained above.

From this study, 5250 ± 180, 3300 ± 240, and 8690 ± 330 $B^0 \to J/\psi K^+\pi^-$ candidates are expected in the $B_s^0 \to J/\psi f_2'(1525)$ sample from 2011–2012, 2015–2016, and 2017–2018 data-takin, respectively. These yields are in good agreement with 6350 ± 580, 3130 ± 290 and 10000 ± 910, which are obtained by using Eq. 6.1 with the $B^0 \to J/\psi K^+\pi^-$ branching fraction expectation and efficiencies purely taken from simulation.

6.2.4 Other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays

Other decays of B_s^0 mesons to the $B_s^0 \to K^+K^-\mu^+\mu^-$ final state can in principle proceed with several configurations of the dikaon system, *e.g.* via the ϕ, $f_2'(1525)$, or $\phi(1680)$ resonance, but they can also proceed via non-resonant S-wave decays. The full dikaon spectrum has been analysed for $B_s^0 \to J/\psi K^+K^-$ decays in Ref. [142], and the result is shown in Fig. 6.9 where the background-subtracted data is overlaid with the fitted amplitude model. The dikaon composition in the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ sample is expected to be similar to that in $B_s^0 \to J/\psi f_2'(1525)$ decays, so the findings from Ref. [142] are discussed in the following as benchmark to what is expected for the signal $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode.

From Fig. 6.9 it is clearly visible that other $B_s^0 \to J/\psi K^+K^-$ decays contribute to the $f_2'(1525)$ mass region, *e.g.* decays via $B_s^0 \to J/\psi\phi$, $B_s^0 \to J/\psi\phi(1680)$, or $B_s^0 \to J/\psi K^+K^-$ with dikaon system in S-wave configuration. Using the fit projections from Fig. 6.9 for an estimate of the amount of background relative to the signal, one finds that the $B_s^0 \to J/\psi f_2'(1525)$ signal corresponds to approximately

47% of the full $B_s^0 \to J/\psi K^+K^-$ amplitude. The dikaon resonances are expected to be similar in the rare mode $B_s^0 \to f_2'(1525)\mu^+\mu^-$ sample, so background from other $B_s^0 \to K^+K^-\mu^+\mu^-$ needs to be taken into account in the fit model to data.

A variable that discriminates between the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal decays and the contribution from other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays is the invariant mass of the dikaon pair $m(K^+K^-)$: the contribution from the signal $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays peaks at the known $f_2'(1525)$ mass, and the background from other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays is a composite structure from the different other dikaon contributions.

Since the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample is very small, it is not possible to separately model all the different $B_s^0 \to K^+K^-\mu^+\mu^-$ background contributions in the fit. Instead, the different contributions are combined into one background component, referred to as *other* $B_s^0 \to K^+K^-\mu^+\mu^-$ *decays*, and this component is modelled in the fit.

This combined contribution from other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays is modelled in $m(K^+K^-\mu^+\mu^-)$ with the same function as the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal as both contributions are from decays of B_s^0 mesons. The invariant dikaon mass, however, is described with a linear function. It is not obvious that a linear function describes this complex mixture of dikaon resonances well, so this model and its effect on the measurement is studied in detail in Sect. 6.5.2.

6.2.5 Combinatorial background

Similar to the $B_s^0 \to \phi\mu^+\mu^-$ analysis, background from random track combinations is expected to contribute to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ samples. This background can be separated from the signal due to its different invariant mass distribution: the combinatorial background is distributed according to an exponential function as explained in Sect. 5.2.6, while the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays accumulate at the known B_s^0-meson mass. Therefore, the different distributions are disentangled from each other by fitting the invariant mass distribution $m(K^+K^-\mu^+\mu^-)$ in data; the fitted function contains a component for the B_s^0 decays and a component for the exponential background.

6.3 Fitting for the observation and branching fraction

Summarising the previous section, several backgrounds are expected to contribute significantly to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data samples. These background contributions are from (1) combinatorial background contribution, (2) other $B_s^0 \to K^+K^-\mu^+\mu^-$, (3) $\Lambda_b^0 \to pK^-\mu^+\mu^-$, and (4) $B^0 \to K^+\pi^-\mu^+\mu^-$ decays. They are separated from the signal by performing a two-dimensional fit in which these contributions are modelled; the general approach is explained in Sect. 4.4.

The combinatorial background separates best from the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal decays using the invariant masses $m(K^+K^-\mu^+\mu^-)$, so the fit is performed using $m(K^+K^-\mu^+\mu^-)$, abbreviated as m in the following. The invariant mass m, however, cannot be used to distinguish the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal from the other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays; these components can only be distinguished with the invariant dikaon mass $m(K^+K^-)$, abbreviated as m_{KK} in the following.

To separate the signal from both the combinatorial and other $B_s^0 \to K^+K^-\mu^+\mu^-$ background decays, a two-dimensional fit is performed using a function $f = f(m, m_{KK})$ that assumes that m and m_{KK} factorise, given by by

$$\begin{aligned} f(m, m_{KK}) = &N(\text{signal}) \times f^{\text{signal}}(m, m_{KK}) \\ &+ N(\text{comb. bkg.}) \times f^{\text{comb. bkg.}}(m, m_{KK}) \\ &+ N(\text{other } B_s^0 \to K^+K^-\mu^+\mu^-) \times f^{\text{other } B_s^0 \to K^+K^-\mu^+\mu^-}(m, m_{KK}) \\ &+ N(\Lambda_b^0) \times f^{\Lambda_b^0}(m, m_{KK}) + N(B^0) \times f^{B^0}(m, m_{KK}). \end{aligned} \tag{6.2}$$

The $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal component $f^{\text{signal}}(m, m_{KK})$ consists of a double Crystal Ball (DCB) function to model the invariant mass m and a relativistic spin-2 Breit–Wigner ($\text{BW}_{s=2}$) function for the invariant dikaon mass m_{KK}, *i.e.*

$$f^{\text{signal}}(m, m_{KK}) = \text{DCB}(m; f_{\text{CB}}, \mathbf{p_1}, \mathbf{p_2}) \times \text{BW}_{s=2}(m_{KK}; \mu_{KK}, \Gamma_{KK}), \tag{6.3}$$

with the fraction f_{CB} and parameter sets $\mathbf{p_i}$ for the two Crystal Ball (CB) functions and the mean μ_{KK} and width Γ_{KK} for the relativistic spin-2 Breit-Wigner function $\text{BW}_{s=2}$. The complete parametrisation for $\text{BW}_{s=2}$ is taken from Refs. [139, 142] and can be found in Appendix D. The DCB is the same function as used for the fits to the $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \phi\mu^+\mu^-$ signal decays detailed in Sect. 5.3.

The combinatorial background component $f^{\text{comb. bkg.}}(m, m_{KK})$ is modelled with exponential functions for both invariant masses, *i.e.*

$$f^{\text{comb. bkg.}}(m, m_{KK}) = \exp(m; a_1) \times \exp(m_{KK}; a_2), \tag{6.4}$$

with parameters a_1 and a_2.

Contributions from other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays are modelled with a DCB function for m and a linear function for m_{KK}, namely

$$f^{\text{other } B_s^0 \to K^+K^-\mu^+\mu^-}(m, m_{KK}) = \text{DCB}(m; f, \mathbf{p_1}, \mathbf{p_2}) \times \text{linear}(m_{KK}; c), \tag{6.5}$$

where the DCB function and parameters are the very same as for the signal component in Eq. 6.3.

Finally, background contributions from $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and $B^0 \to K^+\pi^-\mu^+\mu^-$ decays are modelled with two-dimensional kernel density estimates (KDEs), *i.e.*

$$f^{\Lambda_b^0}(m, m_{KK}) = \text{KDE}^{\Lambda_b^0}(m, m_{KK}) \tag{6.6}$$

and

$$f^{B^0}(m, m_{KK}) = \text{KDE}^{B^0}(m, m_{KK}), \tag{6.7}$$

which are determined on simulated samples of the respective decays. The yields of these background contributions are Gaussian constrained to the expected values discussed in Sect. 6.2.3 and 6.2.2.

The fit validation showed that the fitting procedure is most stable if the fit is performed simultaneously to the three data samples, *i.e.* the 2011–2012, 2015–2016, and 2017–2018 samples, for which the signal branching fraction is shared between these data samples. For this, the fit parameters for the signal yields are re-parametrised via

$$N(\text{signal})^i = \mathcal{B}(\text{signal}) \times \frac{N^i_{B_s^0 \to J/\psi\phi}}{\mathcal{B}(B_s^0 \to J/\psi\phi)} \times \frac{\epsilon^i_{\text{signal}}}{\epsilon^i_{B_s^0 \to J/\psi\phi}}, \tag{6.8}$$

so that the branching fraction of the signal $\mathcal{B}(\text{signal})$ enters as fit parameter instead of the yield $N(\text{signal})$. The efficiency ratios $\epsilon^i_{\text{signal}}/\epsilon^i_{B_s^0 \to J/\psi\phi}$ and the normalisation mode yields $N^i_{B_s^0 \to J/\psi\phi}$ are fixed to the values obtained in Sect. 6.1 and Sect. 5.3.1, respectively.

To stabilise the fitting procedure further, the branching fraction of the other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays is shared as well between the data sets, *i.e.* its yield is re-parametrised similar to the signal in Eq. 6.8. The efficiency for the other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays is flat in the considered m_{KK} range, and the differences in efficiency for the different data-taking periods are assumed to be the same as for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays.

As the final signal region in the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample is only examined after the full analysis has been fixed and validated, the model is first tested on the $B_s^0 \to J/\psi f_2'(1525)$ data sample in Sect. 6.3.1. The final fit to the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data is presented in Sect. 6.3.2. Since this measurement aims at observing the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode for the first time, determining the signal significance is of interest. This significance is evaluated using Wilks's theorem, which is explained in Sect. 6.3.3. Finally, the fits are validated on pseudodata samples in Sect. 6.3.4.

6.3.1 Fit to $B_s^0 \to J/\psi f_2'(1525)$ control mode data

The fit model of the m and m_{KK} distributions of the $B_s^0 \to J/\psi f_2'(1525)$ candidates is similar as to model the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data detailed in the previous section: the (1) $B_s^0 \to J/\psi f_2'(1525)$ signal is described by a DCB and relativistic spin-2 BW function; the (2) $B_s^0 \to J/\psi K^+K^-$ background decays are described by a DCB and a linear function; the (3) combinatorial background is described by two exponential functions; and the (4) $\Lambda_b^0 \to J/\psi pK^-$ and (5) $B^0 \to J/\psi K^+\pi^-$ components are described by two two-dimensional kernel density estimates.

The fraction and tail parameters of the DCB functions are determined from fits to $B_s^0 \to J/\psi f_2'(1525)$ simulation, and these parameters are fixed in the fit to data. The Λ_b^0 and B^0 yields are Gaussian-constrained to the expected values from Sect. 6.2.2 and Sect. 6.2.2, respectively. All other parameters are determined from the fit to data, and the data are fit simultaneously with shared $\mathcal{B}$ ($B_s^0 \to J/\psi f_2'(1525)$) and $\mathcal{B}$ (other $B_s^0 \to J/\psi K^+K^-$). Fig. 6.10 shows the resulting projections of the fitted PDF to the invariant mass of the B_s^0 candidate and the dikaon mass, overlaid on the complete data sample. The PDF describes the data well, and 9870 ± 120, 11260 ± 140, and 25980 ± 330 decays of B_s^0 mesons via $B_s^0 \to J/\psi f_2'(1525)$ mode are found in the 2011–2012, 2015–2016, and 2016–2018 data samples.

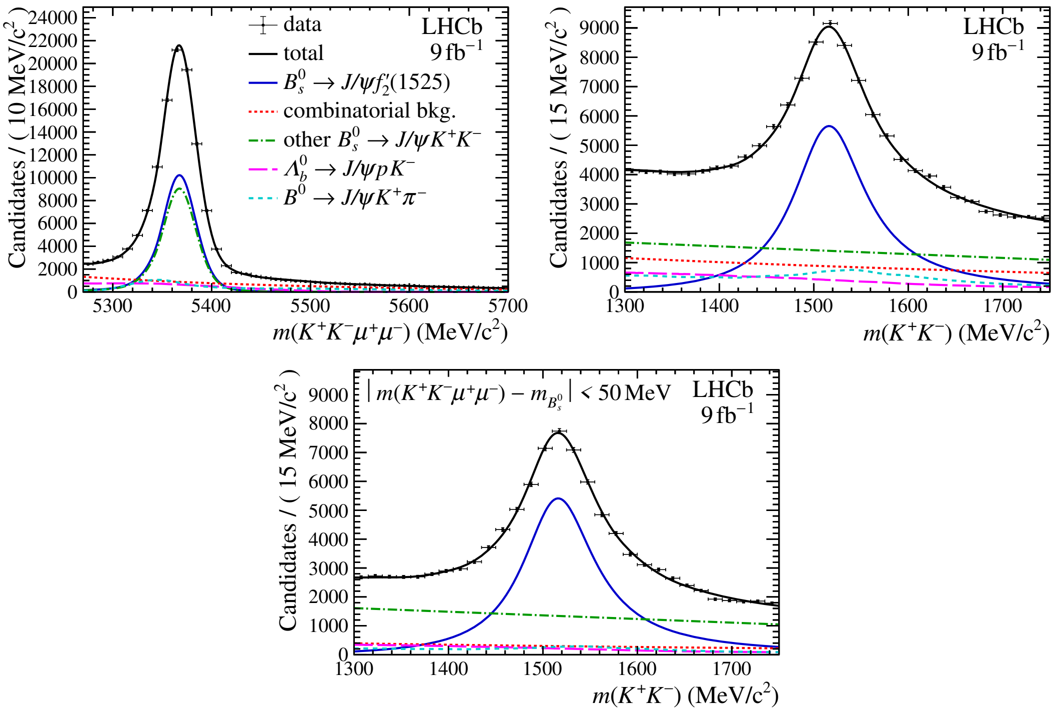

Figure 6.10: Distributions of the (top left) invariant candidate mass $\mathrm{m}(K^+K^-\mu^+\mu^-)$ and the (top right) invariant dikaon mass $\mathrm{m}(K^+K^-)$ from the complete $B_s^0 \to J/\psi f_2'(1525)$ data sample. The projection of the fitted function is overlaid on the data, and the components are shown separately for the signal (blue line), the combinatorial background (red dotted), the other $B_s^0 \to J/\psi K^+K^-$ decays (green dash dot), the $\Lambda_b^0 \to J/\psi pK^-$ decays (magenta dashed), and the $B^0 \to J/\psi K^+\pi^-$ decays (cyan short dashed). The invariant dikaon mass $\mathrm{m}(K^+K^-)$ is also shown for candidates in the signal region (bottom).

6.3.2 Fit to $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal data

The $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample is too small to let all parameters float in the fit; therefore, several parameters are taken from the $B_s^0 \to J/\psi f_2'(1525)$ fit, and only the following are floated in the fit: the (1) signal branching fraction $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$, the (2) branching fraction of the combined other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays $\mathcal{B}(\text{other } B_s^0 \to K^+K^-\mu^+\mu^-)$, the (3–4) combinatorial background slopes a_0 and a_1, the (5) corresponding yield $N(\text{comb. bkg})$, and the (6) Λ_b^0 and (7) B^0 background contributions $N(\Lambda_b^0)$ and $N(B^0)$. Only the two latter, $N(\Lambda_b^0)$ and $N(B^0)$, are Gaussian-constrained to the expected values from Sect. 6.2.2 and Sect. 6.2.3, respectively; all other parameters can float freely.

The final data and fit projections are shown in Fig. 6.11. These distributions, fit projections and resulting parameters have not been examined during the design of the analysis. From this fit,

$$\begin{gathered} N(B_s^0 \to f_2'(1525)\mu^+\mu^-) = 62.6 \pm 7.6 \text{ (2011–2012)}, \\ N(B_s^0 \to f_2'(1525)\mu^+\mu^-) = 66.7 \pm 8.2 \text{ (2015–2016)}, \\ \text{and} \\ N(B_s^0 \to f_2'(1525)\mu^+\mu^-) = 161 \pm 20 \text{ (2017–2018)}. \end{gathered} \tag{6.9}$$

decays for the signal mode are found.

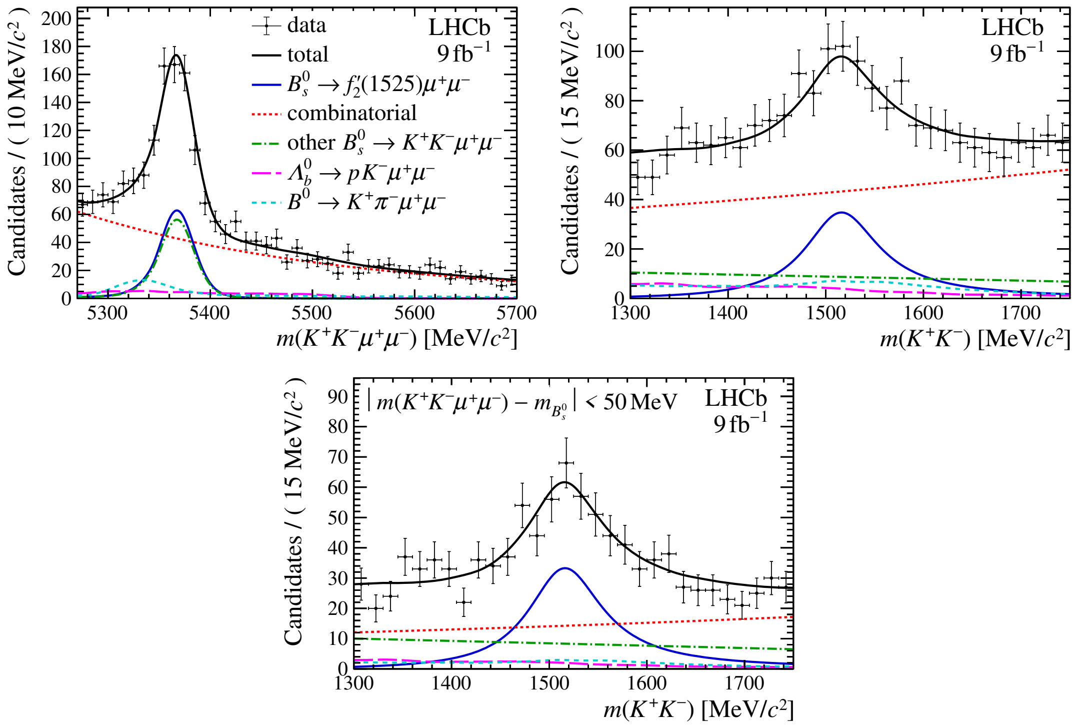

Figure 6.11: Distributions of the (top left) invariant candidate mass $\mathrm{m}(K^+K^-\mu^+\mu^-)$ and the (top right) invariant dikaon mass $\mathrm{m}(K^+K^-)$ from the complete $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample. The projection of the fitted PDF is overlaid on the data, and the components are shown separately for the signal (blue line), the combinatorial background (red dotted), the other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays (green dash dot), the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays (magenta dashed), and the $B^0 \to K^+\pi^-\mu^+\mu^-$ decays (cyan short dashed). The invariant dikaon mass $\mathrm{m}(K^+K^-)$ is also shown for candidates in the signal region close to the B_s^0 mass (bottom).

6.3.3 Signal significance using Wilks's theorem

The $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode has not been observed before, so one goal of this analysis is to observe it for the first time. An observation in particle physics can be claimed if a signal is measured with a significance of larger than 5 standard deviations (σ).

The significance of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal is determined with Wilks's theorem [143], which provides an asymptotic distribution for a likelihood-ratio test, which is a method to test a hypothesis. The two hypotheses that are compared for the likelihood-ratio test are the null hypothesis H_0, which is given by "no signal is present in the data", and the signal hypothesis H_1, given by "the signal component is present in data". The likelihoods for the two hypotheses, $\mathcal{L}_0$ and $\mathcal{L}_1$, are determined with extended maximum-likelihood fits to data, as explained in Sect. 4.4, and the two fit models are the following: the first model, parametrising H_1, is the default fit model used to determine the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal in data, which is described in Sect. 6.3; and the second model, parametrising H_0, is the same fit model as used for H_1 but with signal branching fraction fixed to 0.

The corresponding log-likelihood ratio $\Lambda = \log(\mathcal{L}_1/\mathcal{L}_0)$ is determined, and Wilks's theorem states that $\mathcal{D} = -2 \times \Lambda$ is distributed according to a χ_k^2 distribution with degree-of-freedom difference k. The degree-of-freedom difference k between H_0 and H_1 is $k = 1$ as the only difference between the two is the signal branching fraction. Therefore, $\mathcal{D}$ is distributed asymptotically following a χ^2_1 distribution, and a value of $\mathcal{D} > 25$ is needed for a H_0 incompatibility at $5\,\sigma$, which is the desired threshold to claim an observation of the signal.

The results for the likelihood ratio, and correspondingly the significance of the $B^0_s \to f_2'(1525)\mu^+\mu^-$ signal, was not examined until the analysis design was finished. Before that, the above approach was tested on pseudodata samples generated with the expected number of decay processes. In these studies, more than 99.9% of the pseudodata samples showed a signal significance of greater than $5\,\sigma$, so an observation of the signal was expected. The final signal significance observed in data is presented in Sect. 6.7.

6.3.4 Fit validation

The fitting procedure is validated with a study on pseudodata samples similarly to what is done for the $B^0_s \to \phi\mu^+\mu^-$ measurement in Sect. 5.3.3. This study validates the fits, uncertainties and unbiasedness, and it verifies that the fitting algorithms run without error; from the results of this study, it was also decided to fit the branching fractions simultaneously in data since fitting it separately resulted in large fit failure rates.

The pseudodata samples are generated according to the fit to data, which has not been examined during that stage of the analysis. The pseudodata samples are then fitted again, and the fitted branching fractions are compared to the value used for generating the pseudodata samples. The uncertainties are found to be consistent with the variances expected from the data fluctuations, and only a small bias of 0.2% relative to the central value is found in $\mathcal{B}(B^0_s \to f_2'(1525)\mu^+\mu^-)$. This bias is not corrected for but assigned as systematic uncertainty to the result.

6.4 Method of angular moments

The previous section shows how the $B^0_s \to f_2'(1525)\mu^+\mu^-$ signal decays are distinguished from the background contribution, *i.e.* by performing a fit to the data that discriminates the signal from background decays based on their different distribution of the invariant masses $m(K^+K^-\mu^+\mu^-)$ and $m(K^+K^-)$. Another discriminating feature of $B^0_s \to f_2'(1525)\mu^+\mu^-$ decays is their spin structure: contrary to the significant background processes, the $B^0_s \to f_2'(1525)\mu^+\mu^-$ signal involves a spin-2 dikaon resonance, [2] so finding a significant spin-2 dikaon component in data would therefore be a clear sign for the $B^0_s \to f_2'(1525)\mu^+\mu^-$ signal decay. An analysis method sensitive to the spin structure in data is the *method of angular moments*.

Unfortunately, studies on pseudodata samples indicate that only insignificant results are to be expected from the method of angular moments when using this analysis' data sample, which was collected during 2011–2018. However, more data

[2]Contributions from other decays including a spin-2 dikaon system, *e.g.* with $f_2(1270)$, $f_2(1750)$, or $f_2(1950)$, are expected to be negligibly small in the selected data since their contributions was found to be small in the measurement of $B^0_s \to J/\psi K^+K^-$ decays in Ref. [142].

will be added in the following years, and this data will increase the significance of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ angular moments. This will make the method of angular moments a powerful tool to validate this measurement and further explore the amplitude spectrum contributing to the dikaon regions close to the $f_2'(1525)$ mass. In this sense, the following sections aim to prepare for these future analyses by (1) explaining the method of angular moments and (2) showing how the are extracted from data. The last part demonstrates the approach by extracting the dikaon spin-2 component from $B_s^0 \to J/\psi f_2'(1525)$ data.

Method of angular moments The method of moments is a general statistical approach, and the application to angular distributions of $B \to X\ell^+\ell^-$ decays is discussed in Ref. [144]. It relies on the fact that the decay rate of any decay process Γ can be expressed as differential $d\Gamma$ with respect to the invariant dimuon mass squared q^2 and decay angles, abbreviated as $\vec{\Omega}$. In case of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decay, the decay angles are $\cos\theta_K$, $\cos\theta_\ell$, and ϕ, for which the definitions are given in Appendix E. The differential decay rate $\mathrm{d}^4\Gamma/(\mathrm{dq}^2\mathrm{d}\tilde{\Omega})$ can be parametrised in a basis of orthonormal angular functions $f_i(\vec{\Omega})$, *i.e.* via

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}} = \mathcal{C} \times \left(\sum_{i=1}^{41} \Gamma_i(q^2) \times f_i(\vec{\Omega}) \right), \tag{6.10}$$

with a constant factor $\mathcal{C}$, the ith moment $\Gamma_i(q^2)$ and the angular functions $f_i(\vec{\Omega})$ that consist of the spherical harmonics $Y_l^m(\theta_\ell, \phi)$ and $Y_l^m(\theta_K, 0)$; the explicit form of the $f_i(\vec{\Omega})$ can be found in Ref. [144].

The angular moments Γ_i are the quantities of interest as they are each sensitive to the different combinations of spin-0, spin-1, and spin-2 amplitudes, referred to as S-, P-, and D-wave, and by their interferences. Of particular interest to this analysis are the fifth and tenth moment Γ_5 and Γ_{10}, respectively, since they are sensitive to the D-wave amplitudes. Conveniently, these two angular moments can even be combined so that they correspond to the full D-wave amplitude itself, *i.e.* via

$$|D|^2 = -\frac{7}{18}\left(2 \times \Gamma_5 + 5\sqrt{5} \times \Gamma_{10}\right). \tag{6.11}$$

This is extremely useful as it allows to directly extract the D-wave contribution from data.

Angular moments and D-wave amplitude in data The angular moments Γ_i can be determined by using the orthonormality condition of the angular functions f_i, given by $\int f_i(\vec{\Omega}) f_j(\vec{\Omega})\mathrm{d}\vec{\Omega} = \delta_{ij}$. Applying this condition to Eq. 6.10, the angular moments are given by

$$\Gamma_j(q^2) = \frac{1}{\mathcal{C}} \times \int \left(\frac{\mathrm{d}^4\Gamma}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}} \times f_j(\vec{\Omega}) \right) \mathrm{d}\vec{\Omega}. \tag{6.12}$$

In data, however, the decay rate is not directly accessible as continuous function; instead, individual decay candidates are measured with discrete decay angles $\vec{\Omega}_k$. The decay rate for these individual candidates can then be represented by a delta function $\delta(\vec{\Omega} - \vec{\Omega}_k)$, and the decay rate of the full data sample becomes

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}} = \sum_{k=0}^{N_{\text{data}}} \left(\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} \times \delta(\vec{\Omega}_k - \vec{\Omega}) \right). \tag{6.13}$$

Inserting this discretised version of the decay rate into Eq. 6.12 results in

$$\Gamma_j(q^2) = \frac{1}{\mathcal{C}} \times \int \left(\sum_{k=0}^{N_{\text{data}}} \frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} \times \delta(\vec{\Omega}_k - \vec{\Omega}) \times f_j(\vec{\Omega}) \right) \mathrm{d}\vec{\Omega} \tag{6.14}$$

$$= \frac{1}{\mathcal{C}} \times \sum_{k=0}^{N_{\text{data}}} \left(\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} \times f_j(\vec{\Omega}_k) \right). \tag{6.15}$$

Measuring the angular moments $\Gamma_j(q^2)$ for the integrated q^2 region, Eq. 6.15 becomes

$$\Gamma_j = \mathcal{C}' \times \sum_{k=0}^{N_{\text{data}}} f_j(\vec{\Omega}_k), \tag{6.16}$$

where the constant factor $\mathcal{C}'$ incorporates the q^2-integrated decay rate, and the moment Γ_j is obtained by summing over each decay candidate k weighted by its specific angular function value $f_j(\vec{\Omega}_k)$.

Two other aspects need to be taken into account before calculating the angular moments in data: firstly, the measured angles $\vec{\Omega}_k$ in data might be affected by reconstruction effects. These angle distortions are taken into account by assigning each candidate a weight $\epsilon_{\text{corr.}}(\vec{\Omega}_k)$, which effectively corrects for the detector effects on the angles. This weight $\epsilon_{\text{corr.}}(\vec{\Omega}_k)$ is calculated by comparing the angular distributions from simulated samples of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays before and after the reconstruction and selection;[3] a more detailed explanation of corrections via weights is given in Sect. 4.6.2.

Secondly, the data sample contains a significant amount of background processes after the full selection, which is discussed in detail in Sect. 6.2, and part of these background contributions are from random track combinations or decay processes with confused final-state particles. The angular distributions of these background processes is unknown, so modelling them is not an option; instead, the background contributions are statistically subtracted from data using the sPlot technique [116].

The PDF to extract the sPlot weights is determined using the invariant mass distribution of the B_s^0 candidates, similar to what is described in Sect. 6.3, but using only B_s^0 mass distribution in the the PDF models and therefore combining the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ and $B_s^0 \to K^+K^-\mu^+\mu^-$ component. Fig. 6.12 shows data overlaid on the fitted PDF. This 1D approach was chosen to assess the angular moments also for different bins of the invariant dikaon mass, which would not be possible if information from $m(K^+K^-)$ would have been used in the fit to extract the sPlot weights.

Implementing the angle correcting and the background subtracting weights in Eq. 6.16, one obtains

$$\Gamma_j = \mathcal{C} \times \sum_{k=0}^{N_{\text{data}}} \left(\epsilon_{\text{corr}}(\Omega_k) \times w_S^k \times f_j(\vec{\Omega}_k) \right) \tag{6.17}$$

[3]The effect on the angles is checked to be uncorrelated, so the weight is determined as $\epsilon(\vec{\Omega}_k) = \epsilon(\cos\theta_K) \times \epsilon(\cos\theta_\ell) \times vec\Omega(\phi)$.

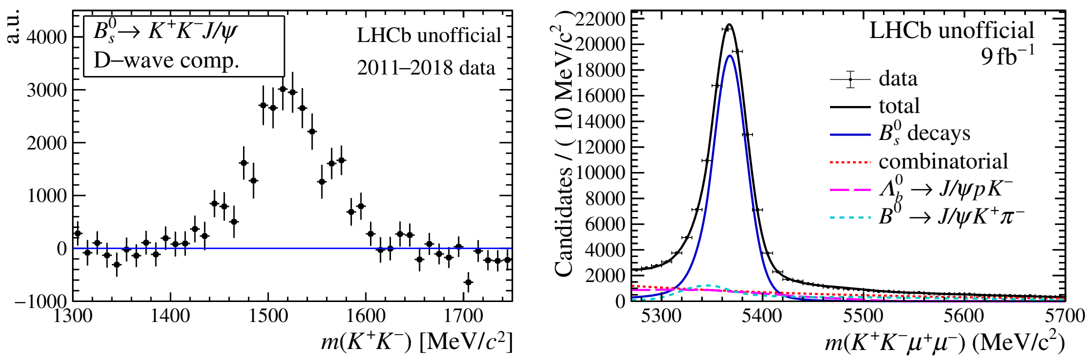

Figure 6.12: Number of (left) D-wave candidates in the $B_s^0 \to J/\psi f_2'(1525)$ data sample in bins of the invariant dikaon mass $m(K^+, K^-)$. The D-wave component in the B_s^0 data is compatible with the $f_2'(1525)$ mass. The non-B_s^0 background decays have been subtracted from the data using the sPlot technique, using a (right) PDF fitted to the invariant B_s^0 mass distribution.

to calculate the jth moment from data with the detector-acceptance correction $\epsilon_{\text{corr}}(\Omega_k)$, the signal weight w_S^k and the angular function-value $f_j(\vec{\Omega}_k)$ for data candidate k.

This method is tested on the control mode $B_s^0 \to J/\psi f_2'(1525)$ data sample, and Fig. 6.12 shows the corresponding D-wave component $|D^2|$ for different bins of the invariant dikaon mass $m(K^+, K^-)$, obtained using Γ_5 and Γ_{10} in Eq. 6.11. A significant contribution at the $f_2'(1525)$ mass is clearly visible. The results for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample D-wave is shown in the result section Sect. 6.7.

6.5 Systematic uncertainties

This section discusses the systematic uncertainties associated to the measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. Similar to the $B_s^0 \to \phi\mu^+\mu^-$ analysis, several sources of systematic effects on the measurement are investigated. Some of these sources are common to both analyses, *i.e.* the effect from the limited simulation sample size, from the choices made in the simulation corrections, from residual mismodelling, and from the mild fit bias observed in the fit validation (Sect. 6.3.4). These effects are briefly summarised in Sect. 6.5.4, and their detailed discussion can be found in Sect. 5.4 for the $B_s^0 \to \phi\mu^+\mu^-$ mode.

Further effects on the branching fraction originate from the specific PDF that is used to describe the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data. This PDF is more complex than for the $B_s^0 \to \phi\mu^+\mu^-$ analysis, and it simplifies the description of other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays as described in Sect. 6.2.4. So, effects from the signal and background modelling are studied for their impact on the branching fraction in Sect. 6.5.2. Since this measurement is performed for the integrated q^2 range, the limited knowledge on the q^2 distribution affects the efficiencies more than for $B_s^0 \to \phi\mu^+\mu^-$ decays, and it is discussed in Sect. 6.5.3. Lastly, uncertainties on the branching fractions $\mathcal{B}(\phi \to K^+K^-)$, $\mathcal{B}(f_2' \to K^+K^-)$, and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ limit the precision of the measurement, as discussed in Sect. 6.5.1.

The systematic uncertainties are all determined separately for the 2011–2012, 2015–2016, and 2017–2018 data samples, and they are combined similarly as for

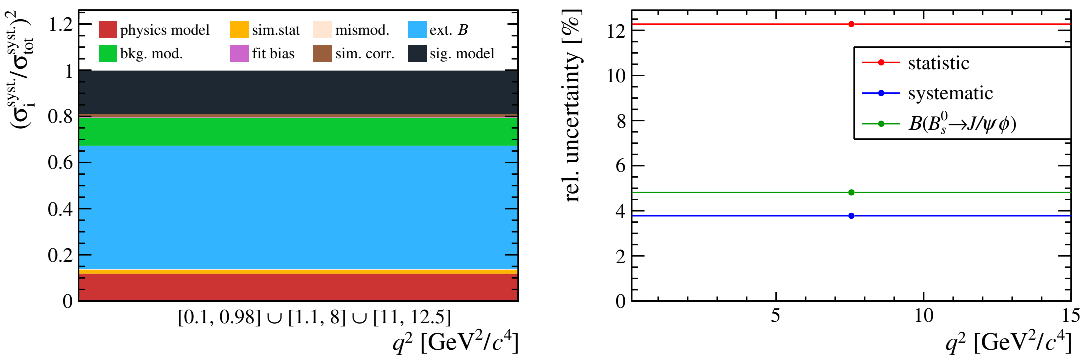

Figure 6.13: Comparison of (left) the different sources of systematic uncertainties and (right) the statistical, combined systematic and normalisation mode uncertainty to the measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$.

Source	$\sigma_{\text{syst.}}$ $[10^{-7}]$
$\mathcal{B}(\phi \to K^+K^-)/\mathcal{B}(f_2'(1525) \to K^+K^-)$	0.04
$\mathcal{B}(J/\psi \to \mu^+\mu^-)$	0.01
Fit model	0.03
Invariant dimuon mass q^2 model	0.02
Limited simulation sample	0.01
Fit bias	< 0.01
Simulation corrections	0.01
Residual mismodelling	< 0.01
Quadratic sum	0.06
Normalisation $\mathcal{B}(B_s^0 \to J/\psi\phi)$	0.07

Table 6.1: Systematic uncertainties on the total branching fraction $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$. The uncertainty from the branching fraction of the normalization mode, $\mathcal{B}(B_s^0 \to J/\psi\phi)$, is quoted separately.

the $B_s^0 \to \phi\mu^+\mu^-$ measurement explained in Sect. 5.4. Fig. 6.13 shows an overview of the systematic uncertainties and their relative effect on $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$ and a comparison between the combined systematic, statistical, and uncertainty from the normalisation branching fraction $B_s^0 \to J/\psi\phi$. The numerical values for the absolute systematic uncertainties on the final branching fraction can be found in Table 6.1.

6.5.1 External branching fractions

The relative branching fraction measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays relies on branching fraction measurements for the light mesons, *i.e.*

$$\mathcal{B}(\phi \to K^+K^-) = (49.2 \pm 0.5)\% \tag{6.18}$$

and

$$\mathcal{B}(f_2' \to K^+K^-) = (43.8 \pm 1.1)\% \tag{6.19}$$

and on the branching fraction of $J/\psi \to \mu^+\mu^-$, which is measured as

$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033)\%. \tag{6.20}$$

The values are all taken from Ref. [89], and their combination affect the measurement by 2.77% relative to the central branching fraction value.

For the total branching fraction, the branching fraction of $B_s^0 \to J/\psi\phi$ decays is needed, which is taken from Ref. [138], namely

$$\mathcal{B}(B_s^0 \to J/\psi\phi) = (1.018 \pm 0.049) \times 10^{-3}. \tag{6.21}$$

The corresponding uncertainty from this normalisation branching fraction of 4.8% is quoted separately to simplify updating the result once a more precise value for $\mathcal{B}(B_s^0 \to J/\psi\phi)$ becomes available.

6.5.2 Fit model

As described in Sect. 6.3, the signal from $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays is extracted from data by performing a fit to the $m(K^+K^-\mu^+\mu^-)$ and $m(K^+K^-)$ distributions. This PDF contains components that describe the signal and each background contribution, and these components rely on chosen shape parameters that can affect the extracted signal.

One of these parameter choices concerns the signal model, where parameters for the spin-2 Breit-Wigner (BW) function are fixed.[4] The effect of this specific choice of parameters is evaluated using pseudodata samples that are generated according to the an alternative fit model where the spin-2 BW parameters, *i.e.* the Blatt-Weisskopf radii and the angular momentum of the B_s^0 meson, have been varied. Two PDFs are then fitted to these pseudodata samples: the default and the altered PDF. The branching fraction parameter from both fitted PDFs is compared for each pseudodata sample, and the average difference of 1.64% relative to the central value is taken as systematic uncertainty for the measurement.

Another choice concerns the Λ_b^0 and B^0 background contributions, which are modelled with kernel-density estimates (KDE's) determined on simulated samples of the respective decays. These KDE's rely on a choice for the smoothing parameters, referred to as *kernel widths*, for which the effect on the measured branching fraction is determined similarly on pseudodata samples as explained above for the spin-2 BW parameters. The average difference between the default and the varied KDE parameters is 0.57%, which is taken as systematic uncertainty.

Lastly, a linear function was chosen to describe the dikaon mass distribution of the ensemble of other $B_s^0 \to K^+K^-\mu^+\mu^-$ decay processes. Choosing a linear function simplifies the distribution that is a complex resonance spectrum as observed in Ref. [142]. However, the following study on pseudodata, generated with the results of Ref. [142], shows that this simplified fit model affects the extracted signal only to a small degree. For this study, high-yield pseudodata samples are generated according to the dikaon mass projection of the fitted PDF from Ref. [142], shown in Fig. 6.14. This fitted PDF describes only the decays from B_s^0 mesons, all other background has been subtracted in Ref. [142].

These pseudodata $m(K^+K^-)$ distributions are then fitted by the $m(K^+K^-)$-part of the default $B_s^0 \to f_2'(1525)\mu^+\mu^-$ component, *i.e.* a spin-2 BW function, and the default component for other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays, *i.e.* linear function. Fig. 6.14 shows an example for such a pseudodata sample and the corresponding fit. The fit fractions optimised from these fitted PDFs are then compared to fit

[4]The relativistic spin-2 BW parametrisation can be found in Appendix D.

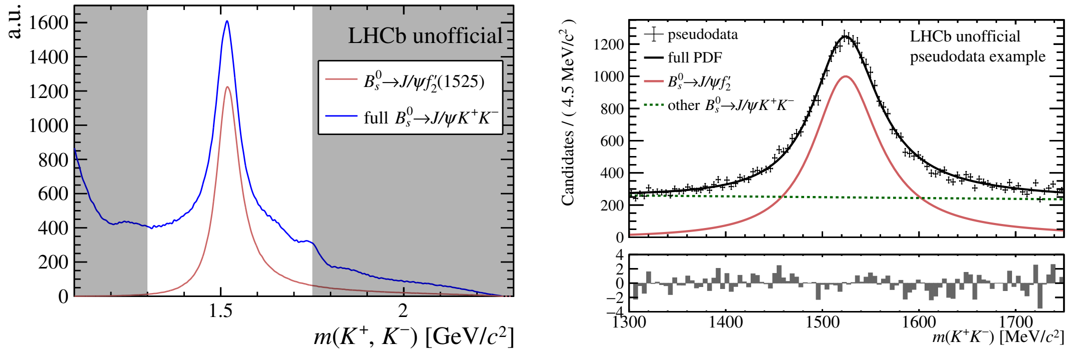

Figure 6.14: Dikaon mass distribution of $B_s^0 \to J/\psi K^+K^-$ decays from (left) a spline describing the PDF projection from Ref. [142] and (right) an example of large-yield pseudodata generated from that spline. The gray area marks the mass regions rejected for this analysis. The pseudodata samples are fit with the default PDF components for the $B_s^0 \to J/\psi f_2'(1525)$ and $B_s^0 \to J/\psi K^+K^-$ decays, and the fit results are compared to the values from Ref. [142] to determine how the chosen fit model affects the measurement.

fractions from Ref. [142], which correspond to the composition of components reflected in the pseudodata samples. On average, the fitted PDF finds 1.18% candidates less than generated, which is assigned as systematic uncertainty to the measurement.

6.5.3 Invariant dimuon mass q^2 model

The branching fraction measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays relies on the efficiencies determined on simulated samples. The efficiencies vary depending on q^2, so the efficiency for the full q^2-sample of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ candidates is dependent on their underlying q^2 distribution. This underlying q^2 distribution, however, is not yet precisely predicted nor measured.

Therefore, the effect of the underlying q^2 distribution is estimated by comparing the efficiencies of two different scenarios: the default efficiencies, determined on simulation with a q^2 distribution from Ref. [128], and alternative efficiencies determined on simulation with a q^2 distribution according to a simple phase-space model. A comparison between these two distributions is shown in Sect. 4.6.7.

The two efficiencies differ by 1.3% relative to the default efficiencies, and this difference is assigned as systematic uncertainty to the measurement.

6.5.4 Limited simulation sample size, fit bias, simulation corrections, and residual mismodelling

The systematic uncertainties on $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$ that originate from the limited simulation sample size, from fit biases observed in studies on pseudodata (Sect. 6.3.4), from the simulation correction settings, and from residual mismodelling in the simulated samples are evaluated likewise as for the analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays, where these are discussed in Sect. 5.4.2, Sect. 5.4.7, Sect. 5.4.5, and Sect. 5.4.6, respectively.

The largest effects among these are from the finite simulation sample size with

0.47% and the simulation correction settings with 0.46%. The residual mismodelling affects the measurement with 0.22%, and the fit bias with 0.16%. All these are added as systematic uncertainties to the branching fraction.

6.6 Analysis validation

Before examining the result of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ significance and branching fraction, the analysis is validated by determining the branching fraction of $B_s^0 \to J/\psi f_2'(1525)$ decays and comparing it to the literature. Similar as for the measurement of $B_s^0 \to \phi\mu^+\mu^-$ decays, validating the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ analysis was a precondition for the publication of the result during the internal LHCb peer-review process.

Since this measurement uses the $B_s^0 \to J/\psi f_2'(1525)$ data to design the fit model, the respective branching fraction is also determined directly from the fit. The comparison of the published and this analysis' result informs about the accuracy of the efficiency determination and the fit model to extract the signal mode.

6.6.1 Branching fraction of $B_s^0 \to J/\psi f_2'(1525)$

The PDF to extract the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ signal from data is designed and tested by fitting a similar PDF to the $B_s^0 \to J/\psi f_2'(1525)$ data sample. The projections of this fitted PDF is shown in Sect. 6.3.1, where it is found to describe the data well. The extracted relative $B_s^0 \to J/\psi f_2'(1525)$ branching fraction is

$$\frac{\mathcal{B}(B_s^0 \to J/\psi f_2'(1525))}{\mathcal{B}(B_s^0 \to J/\psi \phi)} = (24.15 \pm 0.30)\%, \tag{6.22}$$

where the uncertainty comprises only the statistical effects. As this branching fraction measurement is not goal of this analysis, no effort was made to estimate the systematic uncertainties. They are, however, expected to be in the same order of magnitude as for the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ result, which is at the level of 4%, since $\mathcal{B}(B_s^0 \to J/\psi f_2'(1525))$ is similarly determined relative to $B_s^0 \to J/\psi\phi$.

This branching fraction is in good agreement with the branching fraction from Ref. [89] given by

$$\frac{\mathcal{B}(B_s^0 \to J/\psi f_2'(1525))}{\mathcal{B}(B_s^0 \to J/\psi \phi)} = (21 \pm 4)\%. \tag{6.23}$$

6.7 Observation and branching fraction result

The branching fraction is extracted from the fit to data as explained in Sect. 6.3. The signal component of the fitted PDF is found to be significant with 9 standard deviations using Wilks's theorem explained in Sect. 6.3.3, which means that the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode has been observed in this work for the very first time.

The branching fraction is determined on data by using the q^2 region $q^2 \in [0.1, 12.5]$ GeV$^2/c^4$ where $q^2 \in [0.1, 0.98]$ and $[1.1, 8.0]$ GeV$^2/c^4$ have been excluded. To obtain the branching fraction for the full q^2 range, the measured value is corrected for the fraction of candidates measured, *i.e.* outside of the q^2 vetoes, $\epsilon_{q^2} = (73.8 \pm 2.8)\%$. This factor is determined using the q^2 distribution from

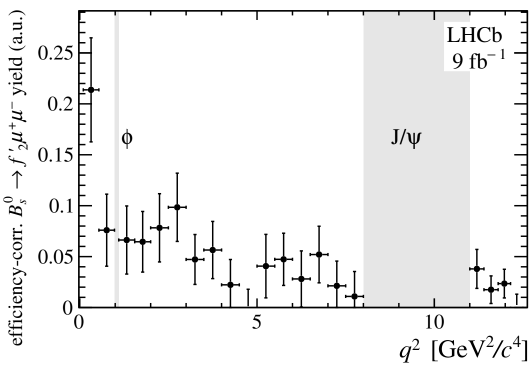

Figure 6.15: Distribution of invariant dimuon mass squared q^2 from efficiency-corrected and background-subtracted $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays.

Ref. [128], and the uncertainty indicates the difference when using the simplistic phase-space model to describe the distribution of q^2. The relative and absolute signal branching fractions are found to be

$$\frac{\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\phi)} = (1.55 \pm 0.19 \pm 0.06 \pm 0.06) \times 10^{-4}, \text{ and}$$
$$\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-) = (1.57 \pm 0.19 \pm 0.06 \pm 0.06 \pm 0.08) \times 10^{-7},$$

where the uncertainties are statistical, systematic, from the extrapolation to the full q^2 region, and, for the absolute branching fraction, from the branching fraction of the normalisation mode $\mathcal{B}(B_s^0 \to J/\psi\phi)$. The measurement of $\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-)$ is in good agreement with the predictions $(2.13 \pm 0.43) \times 10^{-7}$ [128], $(1.8^{+1.1}_{-0.7}) \times 10^{-7}$ [145], and $(2.31^{+0.69}_{-0.50}) \times 10^{-7}$ [146].

This branching fraction measurement is only performed for the full q^2 region since the data sample is too small to measure the branching fraction differential in q^2. To get an idea about the approximate distribution of candidates across q^2, the efficiency-corrected and background-subtracted distribution in data is shown in Fig. 6.15.

As explained in Sect. 6.4, the D-wave component is in principle accessible with the method of moments, and Fig. 6.16 shows the D-wave component for the background-subtracted B_s^0 decays in the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample. Even though the data is compatible with a peak at the known $f_2'(1525)$ mass, no significant D-wave component is found. This is, however, in agreement with what is expected from the sensitivity for a D-wave component extracted with the method of moments, determined on a study using pseudodata samples. More data will be collected in the near future, which will increase the statistical precision of the branching fraction measurement presented in this work and enable to scrutinise the angular moments of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays.

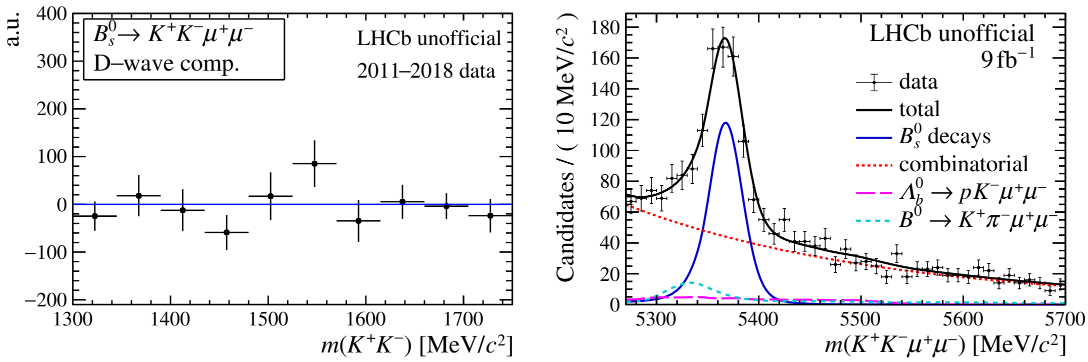

Figure 6.16: Distribution of the (left) dikaon mass of the D-wave component in the combined $B_s^0 \to f_2'(1525)\mu^+\mu^-$ data sample. The non-B_s^0 background decays have been subtracted from the data with the sPlot technique, using a (left) PDF fitted to the invariant B_s^0 candidate mass. The small accumulation of D-wave candidates close to the $f_2'(1525)$ mass is not significant as expected from the statistical limitation of the data sample.

Chapter 7

Summary and conclusion

This thesis presents the updated measurement of the branching fraction $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ and the first observation and branching fraction measurement of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays using data collected at the LHCb experiment. The measurements are documented in detail in the internal analysis note [1], and they have been published as Ref. [47].

The $B_s^0 \to \phi\mu^+\mu^-$ decays have been previously measured using an LHCb data sample collected during 2011–2012, which contains roughly a quarter of decay candidates compared to the data used here. The branching fraction was measured to be more than $3\,\sigma$ below the expectation for low dimuon momenta [2], making this one of the largest deviations in the flavour anomalies.

The $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays have not been observed before this work was carried out. However, estimates based on the already observed high-yield counterparts, the decay $B_s^0 \to J/\psi f_2'(1525)$ [139], indicated that the rare $B_s^0 \to f_2'(1525)\mu^+\mu^-$ mode should be observable with the LHCb data sample collected to-date.

Both decays are measured in this thesis using the complete LHCb data sample collected during the data-taking periods between 2011 and 2018. For the analysis of $B_s^0 \to \phi\mu^+\mu^-$ decays, the data selection criteria have been revised and improved with respect to Ref. [2]. Furthermore, the q^2 intervals have been changed to a finer binning-scheme, which aligns the measurement to other $b \to s\ell^+\ell^-$ analyses and simplifies the theoretical interpretation, so it provides effectively more information on the underlying q^2 distribution. The systematic uncertainties on the measurement have been studied in more detail than in Ref. [2], with the largest improvement coming from the implementation of the lifetime difference between the heavy and the light eigenstate of the B_s^0 meson in the simulated samples; the previous measurement left it uncorrected and assigned a large systematic uncertainty.

For $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, the selection criteria are developed in a similar way to those for $B_s^0 \to \phi\mu^+\mu^-$ decays, however, the larger amount of background contributions demanded a re-optimisation of the criteria leading to stricter requirements. The corresponding selection efficiencies are determined on simulation of the decay processes, and corrections are applied to ensure a good agreement between the simulation and data.

For both modes, the signal decays in data are distinguished from the background contribution by using their differently distributed invariant masses; for this, analytical descriptions of the mass shapes are fitted to data using the unbinned maximum log-likelihood method, and the amount of signal is extracted from these fits. For $B_s^0 \to \phi\mu^+\mu^-$ decays, the B_s^0 candidate mass distribution is used; for $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays, the B_s^0 and $f_2'(1525)$ candidate mass distributions are needed to separate the signal decays from other $B_s^0 \to K^+K^-\mu^+\mu^-$ decays. The branching fractions of the decays are measured normalised to the high-yield $B_s^0 \to J/\psi\phi$ mode, which cancels systematic effects related to the identification and reconstruction of final-state particles at leading order, resulting in small systematic uncertainties related to the measurements.

The branching fraction of $B_s^0 \to \phi\mu^+\mu^-$ decays has been measured as

$$\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-) = (8.14 \pm 0.21 \pm 0.16 \pm 0.39 \pm 0.03) \times 10^{-7} \tag{7.1}$$

with (in order) the statistical uncertainty, the systematic uncertainty, the uncertainty from the normalisation mode branching fraction $\mathcal{B}(B_s^0 \to J/\psi\phi)$, and the uncertainty from the extrapolation to the full q^2 range. The latter accounts for the fact that q^2 regions have been omitted in data, namely the q^2 regions around $q^2 = m_\phi^2$, $q^2 = m_{J/\psi}^2$, and $q^2 = m_{\psi(2S)}^2$, in which contributions from the respective dimuon-resonances dominate. The uncertainty on the extrapolation to the full q^2 range reflects the difference between the form-factor models from Ref. [124] and Ref. [127]. The branching fraction result, improving the statistical precision by a factor two, presents the most precise measurement of $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ to-date.

Of specific interest is also the differential branching fraction of $B_s^0 \to \phi\mu^+\mu^-$ decays, provided now with finer q^2 intervals matching those used for other decays, *e.g.* $B^0 \to K^{*0}\mu^+\mu^-$ decays. Especially the result at low q^2 region is of interest given the result of the 2011–2012 analysis. In the q^2 range $q^2 \in [1.1, 6.0]$, the branching fraction is measured as

$$\frac{\mathrm{d}\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)}{\mathrm{d}q^2\ c^4/\,\mathrm{GeV}^2} = (2.88 \pm 0.21) \times 10^{-8}, \text{ for } q^2 \in [1.1, 6.0]\,\mathrm{GeV}^2/c^4, \tag{7.2}$$

which is below the predicted values at 1.8σ when comparing to a prediction based on light-cone sum-rules (LCSR) [147] and at 3.6σ below the more precise prediction based on a combination of LCSR [147], valid at low q^2, and lattice QCD, valid at high q^2 [26, 27]. The latter deviation is one of the largest tensions with a SM prediction across the flavour anomalies measured in decay rates at LHCb.

The decay $B_s^0 \to f_2'(1525)\mu^+\mu^-$ was observed with a statistical significance of $9\,\sigma$ based on Wilks's theorem. The branching fraction is measured as

$$\mathcal{B}(B_s^0 \to f_2'(1525)\mu^+\mu^-) = (1.57 \pm 0.19 \pm 0.06 \pm 0.06 \pm 0.08) \times 10^{-7}, \tag{7.3}$$

with (in order) the statistical uncertainty, the systematic uncertainty, normalisation mode branching fraction, and the uncertainty from the extrapolation to the full q^2-range. Measuring the differential branching fraction of this mode is not possible to-date due to the limited amount of data. The result is in good agreement with the predictions [128, 145, 146].

The interpretation of the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ measurement in the context of the flavour anomalies is not yet meaningful to-date as it is statistically very limited. This limitation, however, is also accompanied by a rather unclear theoretical situation; so even with more data, the interpretation relies also on advancing the calculation of hadronic effects. As such, this measurement motivates further theoretical studies of this decay mode to continue the effort toward more precise predictions, which started already with recent papers, *e.g.* Ref. [128] and Ref. [146].

For $B_s^0 \to \phi\mu^+\mu^-$ decays, the situation is clearer. The statistical uncertainties have improved by a factor two with respect to the previous iteration, and interestingly, the deviation from the SM prediction at low q^2 persists. This tension is in agreement with the other flavour anomalies, *i.e.* decay rates with muons measured to be lower than expected.

The updated measurement is still limited by both the statistical uncertainty and the uncertainty from the normalisation branching fraction. The statistical uncertainty will become smaller in the future, given the promising efforts from the LHCb upgrade: after the run of the Upgrade-I LHCb detector, one aims to have collected a factor five more data, and after Upgrade-II, an additional factor of six is anticipated. Small data sample sizes will not be the limitation anymore at that point, and the dominating uncertainty on the measured branching fractions will stem from the branching fractions of the normalisation modes. However, the interpretation of the results are also limited by the precision on the predictions, led by hadronic effects, which remain a large source of systematic uncertainty [148]. How these uncertainties evolve in the future cannot be easily foreseen.

The limitation on the prediction is smaller for other observables than branching fractions, namely for angular observables or lepton-universality tests (LFU); so, future developments in these measurements are particularly promising. The work to obtain more measurements from this sector is currently ongoing in three ways: (1) updating the already measured observables, *e.g.* $R_{K^{*0}}$ or the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays, using the full currently-recorded data sample; (2) measuring more observables, like R_ϕ and $R_{K\pi\pi}$; and (3) preparing for the future measurements anticipated with the upgraded LHCb detector; the latter will include, *e.g.* more measurements in the Cabibbo-suppressed $b \to d\ell^+\ell^-$ sector and more angular analyses of $b \to se^+e^-$ modes.

If the LHCb flavour anomalies persist even when more data is being added, it is crucial that these findings will be confirmed by other experiments. And these measurements are underway: CMS and ATLAS already started exploring rare B decays, and they already provided results on the angular analyses of $B^0 \to K^{*0}\mu^+\mu^-$ [41,42] decays and measured the decay rates of $B^0_{(s)} \to \mu^+\mu^-$ [49,50]; in addition, the dedicated heavy flavour experiment Belle-II has started taking data. With their full data sample, the Belle-II collaboration expects determine R_K, $R_{K^{*0}}$ and the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ with a precision at the order of percent [149].

In summary: exciting times are ahead! The near and distant future will certainly deliver more insights on the nature of the flavour anomalies, the area where *something is going on.* And, if the trends persist, maybe indicate that something is *indeed* going on—eventually leading to a hint where the new physics hides, helping to find some answers to the remaining questions in our understanding of nature.

Appendix A

List of local operators for EFT

The following local operators, taken from Ref. [22], are used in the effective field theories to describe the B decays:

- the current-current operators:

$$Q_1^p = (\bar{q}_L \gamma_\mu T^a p_L)(\bar{p}_L \gamma^\mu T^a b_L)\,, \qquad Q_2^p = (\bar{q}_L \gamma_\mu p_L)(\bar{p}_L \gamma^\mu b_L)\,;$$

- the QCD penguin operators:

$$\begin{aligned}
Q_3 &= (\bar{q}_L \gamma_\mu b_L) \sum_p (\bar{p} \gamma^\mu p)\,, \\
Q_4 &= (\bar{q}_L \gamma_\mu T^a b_L) \sum_p (\bar{p} \gamma^\mu T^a p)\,, \\
Q_5 &= (\bar{q}_L \gamma_\mu \gamma_\nu \gamma_\rho b_L) \sum_p (\bar{p} \gamma^\mu \gamma^\nu \gamma^\rho p)\,, \\
Q_6 &= (\bar{q}_L \gamma_\mu \gamma_\nu \gamma_\rho T^a b_L) \sum_p (\bar{p} \gamma^\mu \gamma^\nu \gamma^\rho T^a p)\,,
\end{aligned}$$

- the electromagnetic and chromomagnetic dipole operators

$$\begin{aligned}
Q_7 &= \frac{e}{16\pi^2} m_b (\bar{q}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}\,, \\
Q_8 &= \frac{g_s}{16\pi^2} m_b (\bar{q}_L \sigma^{\mu\nu} T^a b_R) G^a_{\mu\nu}\,;
\end{aligned}$$

- semi-leptonic operators

$$\begin{aligned}
Q_9 &= \frac{e^2}{16\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \ell)\,, \\
Q_{10} &= \frac{e^2}{16\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \gamma_5 \ell)\,, \\
Q_\nu &= \frac{e^2}{8\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\nu}_{\ell L} \gamma^\mu \nu_{\ell L})\,,
\end{aligned}$$

with the sums running over the quarks $p = u, c, d, s, b$; the $SU(3)_C$ group generators T^a, the spin matrices $\sigma^{\mu\nu}$, and the photon and gluon fields $G^a_{\mu\nu}$ and $F_{\mu\nu}$, respectively.

Appendix B

Trigger selection requirements

This section details the requirements to fire the trigger lines used for the measurements. Table B.1 gives the hardware (L0) trigger requirements, and Table B.2, Table B.3, and Table B.4 the first software (Hl1) requirements.

The displacement requirement for `Hlt1TrackMVA` is combines the transverse momentum *pt* of the B_s^0 meson and the impact parameter significance χ^2_{IP} to define the two-dimensional trigger requirement

$$\begin{gathered}\left(p_{\mathrm{T}} > 25\,\mathrm{GeV} \wedge \chi^2_{\mathrm{IP}} > 7.4\right) \\ \vee \\ \Bigg(p_{\mathrm{T}} \in [1\,\mathrm{GeV}, 25\,\mathrm{GeV}] \wedge \\ \log(\chi^2_{\mathrm{IP}}) > \frac{1}{(p_{\mathrm{T}}/\,\mathrm{GeV} - 1)^2} + \frac{x_d}{25 \times 10^3} \times (25 - p_{\mathrm{T}}/\,\mathrm{GeV}) + \log(7.4)\Bigg)\,,\end{gathered} \tag{B.1}$$

with a parameter x_d, which is varied during data-taking.

The second software trigger (Hlt2) stage requirements are given in Table B.4 and Table B.5.

dataset	f	L0Muon: $p_{\mathrm{T}}(\mu^{\pm})$ [GeV]	N(SPD) –	$E_{\mathrm{T}}^{\mathrm{prev.}}$ [GeV]	L0DiMuon: $p_{\mathrm{T}}(\mu_1^{\pm}) \times p_{\mathrm{T}}(\mu_2^{\pm})$ [GeV^2]	N(SPD)
2011	99%	> 1.48	< 600	–	> 1.68	< 900
2012	97%	> 1.76	< 600	–	> 2.56	< 900
2015	77%	> 2.8	< 450	–	> 1.69	< 900
2015	21%	> 2.4	< 450	–	> 1.69	< 900
2016	47%	> 1.3	< 450	–	> 1.69	< 900
2016	34%	> 1.8	< 450	–	> 2.25	< 900
2017	37%	> 1.7	< 450	< 24	> 3.24	< 900
2017	36%	> 1.4	< 450	< 24	> 1.69	< 900
2017	16%	> 1.1	< 450	< 24	> 1.0	< 900
2018	100%	> 1.75	< 450	< 24	> 3.24	< 900

Table B.1: Numerical values for the hardware trigger requirements. The criteria are given separately for the single muon trigger, L0Muon, and the dimuon trigger, L0DiMuon. Trigger settings are only shown if they were used for more than 10% of a dataset, *i.e.* $f > 10\%$.

year	f	IP [mm]	χ^2_{IP}	p_{T} (tr.) [GeV]	p(tr.) [GeV]	tr.-fit χ^2/ndf	GhostProb
2011	100%	> 0.1	> 16	> 1	> 8	< 2	–
2012	70%	> 0.1	> 16	> 1	> 3	< 2.5	–
2012	27%	> 0.1	> 16	> 1	> 8	< 2.5	–
2015	100%	–	> 10	> 0.91	> 6	< 3	–
2016	84%	–	> 35	> 0.1	> 6	< 3	< 0.2
2016	16%	–	> 10	> 0.91	> 6	< 3	–
2017	100%	–	> 35	> 1.1	> 6	< 3	< 0.2
2018	100%	–	> 35	> 1.1	> 6	< 3	< 0.2

Table B.2: Requirements to fire the Hlt1TrackMuon trigger in the different data sets. The variables are explained in the text, and trigger settings are only listed if they were used for more than 10% of a dataset, *i.e.* $f > 10\%$.

`Hlt1TrackAllL0`:

year	f	IP [mm]	χ^2_{IP}	p_{T} (tr.) [GeV]	p(tr.) [GeV]	tr.-fit χ^2/ndf	`GhostProb`
2011	100%	> 0.1	> 16	> 1.6	> 10	< 2.5	–
2012	70%	> 0.1	> 16	> 1.7	> 3	< 2	–
2012	27%	> 0.1	> 16	> 1.6	> 10	< 1.5	–

`Hlt1TrackMVA`:

year	f	IP [mm]	χ^2_{IP} and p_{T} (tr.) in Eq. B.1	p(tr.) [GeV]	tr.-fit χ^2/ndf	`GhostProb`
2015	100%	–	$x_d = 1.1$	> 5	< 2.5	–
2016	44%	–	$x_d = 1.1$	> 5	< 2.5	< 0.2
2016	33%	–	$x_d = 2.3$	> 5	< 2.5	< 0.2
2016	16%	–	$x_d = 1.1$	> 5	< 2.5	–
2017	100%	–	$x_d = 1.1$	> 5	< 2.5	< 0.2
2018	100%	–	$x_d = 1.1$	> 5	< 2.5	< 0.2

Table B.3: Requirements to fire the `Hlt1TrackMVA` and the `Hlt1TrackAllL0` trigger in 2011–2012 and 2015—2018 conditions. The variables are explained in the text, and trigger settings are only listed if they were used for more than 10% of a dataset, *i.e.* $f > 10\%$.

variable	`Hlt1TwoTrackMVA`	`Hlt2Topo2,3,4Body(BBDT)`
tracks:		
p_{T}	> 500	> 200
χ^2_{IP}	> 4	> 4
χ^2/ndof	< 2.5	< 2.5
N(tracks w. $\chi^2_{\text{IP}} < 16$)	–	> 2
vertex:		
p_{T} [GeV]	> 2	–
χ^2_{vtx}	< 10	< 10
$m_{\text{corr.}}$ [GeV]	> 1	$\in [1, 10]$
η	$\in [2, 5]$	$\in [2, 5]$
MVA:		
input variables	$\sum p_{\text{T}}$, χ^2_{vtx}, χ^2_{FD}, N(tracks w. $\chi^2_{\text{IP}} < 16$),	$\sum p_{\text{T}}$, χ^2_{vtx}, χ^2_{FD}, N(tracks w. $\chi^2_{\text{IP}} < 16$) N(tracks), m_{corr} χ^2_{IP}, η, min(p_{T}) N(displ. tracks)

Table B.4: Requirements to fire the `Hlt1TwoTrackMVA` and `Hlt2Topo2,3,4Body(BBDT)` trigger [110, 111]. They first select tracks, combine them to form vertices, and then select these track-combinations based on an multi-variate classifier (MVA). The variables are explained in the text.

`Hlt2DiMuonDetached`:

year	χ^2/ndf	χ^2_{IP}	p_{T}	m($\mu^+\mu^-$)	χ^2_{vtx}	χ^2_{FD}	p_{T} ($\mu^+\mu^-$)
2011	< 5	> 9	> 0.5	$> 1\,\text{GeV}$	< 25	> 49	> 1.5
2012	< 4	> 25	> 0.3	–	< 8	> 49	> 0.6
2015	< 5	> 9	–	$> 2.9\,\text{GeV}$	< 25	> 49	–
2016	< 5	> 9	–	$> 2.9\,\text{GeV}$	< 25	> 49	–
2017	< 5	> 25	> 0.3	–	< 9	> 81	> 0.6
2018	< 5	> 25	> 0.3	–	< 9	> 81	> 0.6

`Hlt2DiMuonDetachedHeavy`:

year	χ^2/ndf	χ^2_{IP}	p_{T}	m($\mu^+\mu^-$)	χ^2_{vtx}	χ^2_{FD}	p_{T} ($\mu^+\mu^-$)
2011	< 5	–	> 0.5	> 2.95	< 25	> 25	–
2012	< 4	–	> 0.3	> 2.95	< 8	> 25	–
2015	< 5	–	> 0.3	> 2.95	< 25	> 25	–
2016	< 5	–	> 0.3	> 2.95	< 25	> 25	–
2017	< 5	–	> 0.3	> 2.95	< 25	> 25	–
2018	< 5	–	> 0.3	> 2.95	< 25	> 25	–

Table B.5: Requirements to fire the `Hlt2DiMuonDetached` and the `Hlt2DiMuonDetachedHeavy` triggers. The variables are explained in the text. The `Hlt2DiMuonDetached` trigger is not used for the 2015–2016 data sets since its requirement on the invariant mass is mis-configured.

Appendix C

Comparing parameters from non-exclusive data samples

Different measurements of the same parameter should provide consistent results within their uncertainties. Assuming a parameter x is measured independently by two experiments A and B as x_A and x_B with Gaussian uncertainties σ_A and σ_B, respectively. Then, the difference between the two measurements $\Delta_{A,B} = x_A - x_B$ should be described with a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma_\Delta = \sqrt{\sigma_A^2 + \sigma_B^2}$. Often, one says two measurements are in agreement if their difference $\Delta_{A,B}$ is not larger than $3 \times \sigma_\Delta$, which corresponds to a probability of 99.7%.

The calculation complicates if the measurements A and B are not independent. The scenario relevant for this work comprises measurements that are performed on *overlapping* data samples, as depicted in Fig. C.1. In this case, a parameter is measured on data sample A, yielding $x_A \pm \sigma_A$, and on data sample B, yielding $x_B \pm \sigma_B$; and, a fraction of data is used in both measurements. Then, the difference between the two measurements is still described by a Gaussian distribution with mean $\mu = 0$, but to calculate the standard deviation σ_Δ, one needs to take the overlap into account.

The parameter measured on sample A can be understood as *weighted average* between the two hypothetical measurements: one (1) carried out on the data that

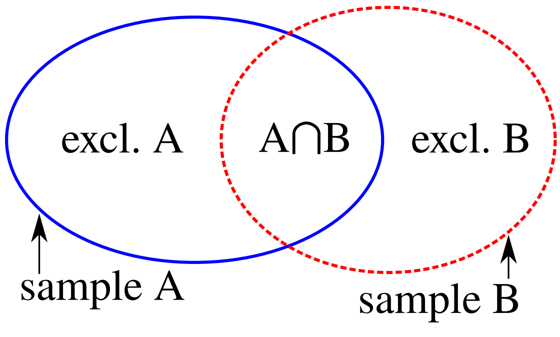

Figure C.1: Schematic to explain the nomenclature used in this section. A parameter is determined on data samples A and B, whereby the samples can be split into the sub samples of data exclusively in the data samples, labelled *excl. A* or *excl. B*, and the data present both in sample A and B, labelled $A \cup B$.

is exclusively used in measurement A, referred to *excl.A*; and the other (2) carried out on the sample used in both measurements A and B, referred to as $A \cap B$. So, the parameter x_A is given by

$$x_A = \frac{\sigma^2_{A\cap B}}{\sigma^2_{excl.A} + \sigma^2_{A\cap B}} x_{excl.A} + \frac{\sigma^2_{excl.A}}{\sigma^2_{excl.A} + \sigma^2_{A\cap B}} x_{A\cap B}, \tag{C.1}$$

with the values x and the uncertainties σ from the hypothetical measurements defined above. The same applied to the measurement x_B, and the corresponding weighted average reads

$$x_B = \frac{\sigma^2_{A\cap B}}{\sigma^2_{excl.B} + \sigma^2_{A\cap B}} x_{excl.B} + \frac{\sigma^2_{excl.B}}{\sigma^2_{excl.B} + \sigma^2_{A\cap B}} x_{A\cap B}. \tag{C.2}$$

From these equations, the difference $\Delta_{A,B}$ is given as

$$\begin{aligned}\Delta_{A,B} = &\left(\frac{\sigma^2_{excl.B}}{\sigma^2_{excl.A} + \sigma^2_{excl.B}} - \frac{\sigma^2_{A\cap B}}{\sigma^2_{excl.A} + \sigma^2_{A\cap B}}\right) \cdot x_{excl.A} \\ &+ \left(\frac{\sigma^2_{excl.A}}{\sigma^2_{excl.A} + \sigma^2_{excl.B}}\right) \cdot x_{excl.B} - \left(\frac{\sigma^2_{excl.A}}{\sigma^2_{excl.A} + \sigma^2_{A\cap B}}\right) \cdot x_{A\cap B},\end{aligned} \tag{C.3}$$

and this expression contains only parameters referring to measurements of *independent* samples, namely *excl.A*, *excl.B*, and $A \cap B$; therefore, one can use standard Gaussian error propagation to calculate σ_Δ.

Calculating σ_Δ requires input for the uncertainties on the hypothetical measurements, *i.e.* on $\sigma_{excl.A}$, $\sigma_{excl.B}$, and $\sigma_{A\cap B}$. To obtain these, the fact that the uncertainties scale with the square-root of the data sample size is used, *e.g.*

$$\sigma_{excl.A} = \sqrt{\frac{N_A}{N_{excl.A}}} \times \sigma_A, \tag{C.4}$$

with the number of data candidates N_A ($N_{excl.A}$) in sample A (*excl.A*).

Using this and performing the error prfopagation on Eq. C.3 yields

$$\sigma^2_\Delta = (1 - f_B) \cdot \sigma^2_A + (1 - f_A) \cdot \sigma^2_B, \tag{C.5}$$

with the fraction of data in sample A (B) f_A (f_B) that is also used in the other measurement, *i.e.* $f_A = N_{A\cap B}/N_A$ ($f_B = N_{A\cap B}/N_B$).

With Eq. C.5, one can now calculate the overlap-corrected uncertainty σ_Δ on the difference between two measurements $\Delta_{A,B}$. With this uncertainty, one can then determine whether the measurements agree, often quantified via $\Delta_{A,B} < 3 \times \sigma_\Delta$.

Appendix D

Relativistic spin-2 Breit-Wigner function

The dikaon mass distribution of $B_s^0 \to f_2'(1525)\mu^+\mu^-$ and $B_s^0 \to J/\psi f_2'(1525)$ decays is modelled with a relativistic spin-2 Breit-Wigner function $\mathrm{BW}(x)$ [129], which is given by

$$\begin{aligned}\mathrm{BW}(x) = {} & N_0 \times m_R \times p_B^{2L_B+1} \left(\frac{p_R}{m_R}\right)^{2L_R+1} \\ & \times \left(F_{L_R}^R F_{L_B}^B\right)^2 \frac{1}{(m_R^2 - x^2)^2 + m_R^2 \Gamma^2(x)},\end{aligned} \tag{D.1}$$

with a normalisation N_0, the f_2' mass m_R, the momentum of the f_2' in the B_s^0 rest frame p_B, the momentum of one of the kaons in the f_2' rest frame p_R, the spin of f_2' meson L_R, the Blatt-Weisskopf coefficients [150] for the f_2' (B) meson F^R (F^B), and the relativistic width of the resonance $\Gamma(x)$. The latter is given by

$$\Gamma(x) = \Gamma_R \frac{m_R}{x} \left(\frac{p_R(x)}{p_R(m_R)}\right)^{2L_R+1} (F_{L_R}^R)^2 \tag{D.2}$$

with the natural width of the f_2' meson Γ_R and the other variables as explained above.

The momenta in the rest-frames, i.e. the above used p_B and p_R, are fixed by the masses of the decaying particles. When considering the two-body decay $M \to m_1 m_2$, the momentum p_1 of the particle with mass m_1 is given by

$$p_1(M, m_1, m_2) = \frac{\sqrt{(M^2 + m_1^2 - m_2^2)^2 - 4M^2 m_1^2}}{2M}. \tag{D.3}$$

Appendix E

Decay angle definitions

The $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ angular decay rates can be parametrised with three decay angles, namely $\cos\theta_K$, $\cos\theta_\ell$, and ϕ. They are depicted in Fig. E.1 and defined as defined as follows:

- $\cos\theta_K$: The angle between the negatively charged kaon and the ϕ (f_2') meson's flight direction.
- $\cos\theta_\ell$: The angle between the negatively charged muon and the B_s^0 meson's flight direction.
- ϕ: The angle between the plane spanned by the two muons and the plane spanned by the two kaons.

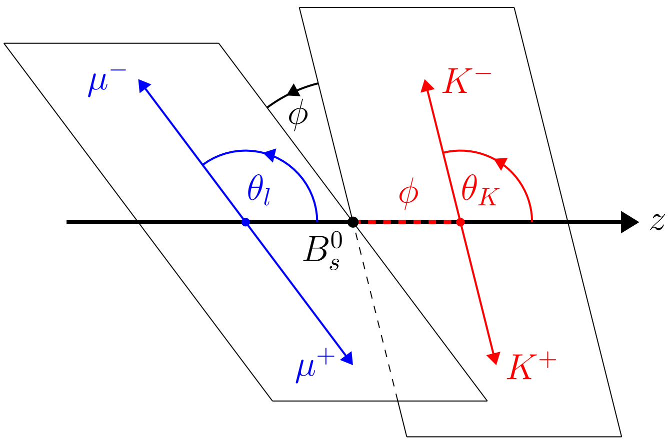

Figure E.1: Scheme depicting the angular distribution of $B_s^0 \to \phi\mu^+\mu^-$ decays.

Bibliography

[1] S. Kretzschmar, E. A. Smith, C. M. Langenbruch, M. Materok, and M. P. Whitehead, *Differential branching fraction of the rare decay $B_s^0 \to \phi\mu^+\mu^-$ and search for $B_s^0 \to f_2'(1525)\mu^+\mu^-$*, CERN Document Server (2020), `https://cds.cern.ch/record/2730622`.

[2] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, `arXiv:1506.08777`.

[3] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics*, Wiley, New York, 1984.

[4] G. Altarelli, *Collider Physics within the Standard Model: a Primer*, `arXiv:1303.2842`.

[5] *Graphic: Elementary particle interactions*, [`https://commons.wikimedia.org/wiki/File:Elementary_particle_interactions.svg`; accessed 08-Feb-2022].

[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) 30, `arXiv:1207.7235`.

[7] 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) 1, `arXiv:1207.7214`.

[8] L. Edelhäuser and A. Knochel, *Tutorium Quantenfeldtheorie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[9] Y. Grossman and P. Tanedo, *Just a Taste: Lectures on Flavor Physics*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics*, 109–295, 2018, `arXiv:1711.03624`.

[10] C. Langenbruch, *Quark flavour physics lecture (summer term 2021)*, 2021.

[11] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.

[13] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.

[14] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[15] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[16] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*, Eur. Phys. J. C **17** (2000) 137, `arXiv:hep-ph/0005110`.

[17] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, Erratum ibid. **119** (2017) 169901, `arXiv:1612.05140`.

[18] HFLAV collaboration, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, Eur. Phys. J. C **77** (2017) 895, `arXiv:1612.07233`.

[19] A. J. Buras, D. Buttazzo, J. Girrbach-Noe, and R. Knegjens, *Can we reach the Zeptouniverse with rare K and $B_{s,d}$ decays?*, JHEP **11** (2014) 121, `arXiv:1408.0728`.

[20] A. Canepa, *Searches for Supersymmetry at the Large Hadron Collider*, Rev. Phys. **4** (2019) 100033.

[21] T. Dorigo, *Hadron collider searches for diboson resonances*, Prog. Part. Nucl. Phys. **100** (2018) 211, `arXiv:1802.00354`.

[22] T. Blake, G. Lanfranchi, and D. M. Straub, *Rare B Decays as Tests of the Standard Model*, Prog. Part. Nucl. Phys. **92** (2017) 50, `arXiv:1606.00916`.

[23] J. Albrecht, D. van Dyk, and C. Langenbruch, *Flavour anomalies in heavy quark decays*, Prog. Part. Nucl. Phys. **120** (2021) 103885, `arXiv:2107.04822`.

[24] H. Georgi, *Effective field theory*, Ann. Rev. Nucl. Part. Sci. **43** (1993) 209.

[25] W. Altmannshofer and P. Stangl, *New physics in rare B decays after Moriond 2021*, Eur. Phys. J. C **81** (2021) 952, `arXiv:2103.13370`.

[26] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Calculation of $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$ observables using form factors from lattice QCD*, Phys. Rev. Lett. **112** (2014) 212003, `arXiv:1310.3887`.

[27] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Rare B decays using lattice QCD form factors*, PoS **LATTICE2014** (2015) 372, `arXiv:1501.00367`.

[28] W. Wang, *B to tensor meson form factors in the perturbative QCD approach*, Phys. Rev. D **83** (2011) 014008, `arXiv:1008.5326`.

[29] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, `arXiv:1810.08132`.

[30] M. Bordone, G. Isidori, and A. Pattori, *On the Standard Model predictions for R_K and R_{K^*}*, Eur. Phys. J. C **76** (2016) 440, arXiv:1605.07633.

[31] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv:2103.11769.

[32] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[33] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays*, JHEP **05** (2020) 040, arXiv:1912.08139.

[34] LHCb collaboration, R. Aaij *et al.*, *Tests of lepton universality using $B^0 \to K_S^0\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays*, Phys. Rev. Lett. **128** (2022) 191802, arXiv:2110.09501.

[35] Belle collaboration, A. Abdesselam *et al.*, *Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle*, Phys. Rev. Lett. **126** (2021) 161801, arXiv:1904.02440.

[36] BELLE collaboration, S. Choudhury *et al.*, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, JHEP **03** (2021) 105, arXiv:1908.01848.

[37] BaBar collaboration, J. P. Lees *et al.*, *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)}l^+l^-$*, Phys. Rev. D **86** (2012) 032012, arXiv:1204.3933.

[38] P. Koppenburg, *Plots listing a cherry–picked sets of flavour anomalies*, https://www.nikhef.nl/~pkoppenb/anomalies.html. [Online; accessed 14-Mar-2022].

[39] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, *Implications from clean observables for the binned analysis of $B \to K^*\mu^+\mu^-$ at large recoil*, JHEP **01** (2013) 048, arXiv:1207.2753.

[40] Belle collaboration, S. Wehle *et al.*, *Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$*, Phys. Rev. Lett. **118** (2017) 111801, arXiv:1612.05014.

[41] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **781** (2018) 517, arXiv:1710.02846.

[42] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **10** (2018) 047, arXiv:1805.04000.

[43] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.

[44] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay*, Phys. Rev. Lett. **126** (2021) 161802, arXiv:2012.13241.

[45] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of charged and neutral $B \to K\mu^+\mu^-$ decays*, JHEP **05** (2014) 082, arXiv:1403.8045.

[46] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the rare decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **11** (2021) 043, arXiv:2107.13428.

[47] 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) 151801, arXiv:2105.14007.

[48] LHCb collaboration, R. Aaij *et al.*, *Analysis of neutral B-meson decays into two muons*, Phys. Rev. Lett. **128** (2022) 041801, arXiv:2108.09284.

[49] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of properties of $B_\mathrm{s}^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*, JHEP **04** (2020) 188, arXiv:1910.12127.

[50] ATLAS collaboration, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, arXiv:1812.03017.

[51] T. Hurth, F. Mahmoudi, D. M. Santos, and S. Neshatpour, *More indications for lepton nonuniversality in $b \to s\ell^+\ell^-$*, Phys. Lett. B **824** (2022) 136838, arXiv:2104.10058.

[52] N. Gubernari, D. van Dyk, and J. Virto, *Non-local matrix elements in $B_{(s)} \to \{K^{(*)}, \phi\}\ell^+\ell^-$*, JHEP **02** (2021) 088, arXiv:2011.09813.

[53] M. Algueró *et al.*, *$b \to s\ell\ell$ Global Fits after R_{K_S} and $R_{K^{*+}}$*, 2021, arXiv:2104.08921.

[54] M. Ciuchini *et al.*, *Lessons from the $B^{0,+} \to K^{*0,+}\mu^+\mu^-$ angular analyses*, Phys. Rev. D **103** (2021) 015030, arXiv:2011.01212.

[55] B. Capdevila, M. Fedele, S. Neshatpour, and P. Stangl, *Effective field theory for $b \to s\ell^+\ell^-$ transitions*, Flavour Anomaly Workshop, 20 October 2021, 2021.

[56] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[57] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[58] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[59] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[60] LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, .

[61] *The LHCb public results – LHCb publications*, `https://lhcbproject.web.cern.ch/lhcbproject/Publications/p/Summary_all.html`. [Online; accessed 19-Jan-2022].

[62] P. Koppenburg, *New particles discovered at the LHC*, `https://www.nikhef.nl/~pkoppenb/particles.html`. [Online; accessed 19-Jan-2022].

[63] I. Belyaev *et al.*, *The history of LHCb*, Eur. Phys. J. H **46** (2021) 3, `arXiv:2101.05331`.

[64] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, `arXiv:1306.0249`.

[65] R. Lindner, *LHCb layout (2)*, LHCb Collection., 2008.

[66] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, 2001.

[67] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, 2002.

[68] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024, 2001.

[69] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, `arXiv:1412.6352`.

[70] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, `arXiv:1408.1251`.

[71] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, 2000.

[72] LHCb RICH Collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II. Performance of the LHCb RICH detectors during the LHC Run II* , CERN, Geneva, 2017. Updated authors' details and DOI, doi: 10.1016/j.nima.2017.03.009.

[73] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, 2001.

[74] T. Head, *The LHCb trigger system*, Journal of Instrumentation **9** (2014) C09015.

[75] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, `arXiv:1812.10790`.

[76] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, `arXiv:1210.6861`.

[77] T. Likhomanenko *et al.*, *LHCb topological trigger reoptimization*, J. Phys. Conf. Ser. **664** (2015) 082025.

[78] A. Piucci, *The LHCb upgrade*, Journal of Physics: Conference Series **878** (2017) 012012.

[79] R. Frühwirth, *Application of kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444.

[80] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044, 2014.

[81] R. W. Forty, *Ring-Imaging Cherenkov Detectors for LHCb*, doi: 10.1016/S0168-9002(96)00951-5.

[82] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, `arXiv:1211.6759`.

[83] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, `arXiv:1306.0249`.

[84] H. Terrier and I. Belyaev, *Particle identification with LHCb calorimeters* , CERN, Geneva, 2003.

[85] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, `arXiv:1803.00824`.

[86] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011, 2017.

[87] J. Gratrex and R. Zwicky, *Parity Doubling as a Tool for Right-handed Current Searches*, JHEP **08** (2018) 178, `arXiv:1804.09006`.

[88] LHCb collaboration, R. Aaij *et al.*, *Study of beauty hadron decays into pairs of charm hadrons*, Phys. Rev. Lett. **112** (2014) 202001, `arXiv:1403.3606`.

[89] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[90] LHCb collaboration, R. Aaij *et al.*, *Observation of $\overline{B}{}^0_s \to J/\psi f_1(1285)$ decays and measurement of the $f_1(1285)$ mixing angle*, Phys. Rev. Lett. **112** (2014) 091802, `arXiv:1310.2145`.

[91] *Documentaton of Gaudi*, `https://gaudi.web.cern.ch/gaudi/index.html`. [Online; accessed 17-Jan-2022].

[92] *Documentaton of Moore*, `https://lhcbdoc.web.cern.ch/lhcbdoc/moore/master/index.html`. [Online; accessed 17-Jan-2022].

[93] *The Brunel project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/`. [Online; accessed 17-Jan-2022].

[94] *The DaVinci project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/`. [Online; accessed 17-Jan-2022].

[95] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[96] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[97] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[98] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937.

[99] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[100] *The Boole project*, http://lhcbdoc.web.cern.ch/lhcbdoc/boole/. [Online; accessed 17-Jan-2022].

[101] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[102] F. Follin and D. Jacquet, *Implementation and experience with luminosity levelling with offset beam*, in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, 183–187, 2014, arXiv:1410.3667.

[103] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-2013-021, 2013.

[104] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, 2014.

[105] LHCb collaboration, *LHCb PID Upgrade Technical Design Report*, CERN-LHCC-2013-022, 2013.

[106] LHCb collaboration, *Expression of Interest for a Phase-II LHCb Upgrade: Opportunities in flavour physics, and beyond, in the HL-LHC era*, CERN-LHCC-2017-003, 2017.

[107] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.

[108] M. Materok, E. A. Smith, C. M. Langenbruch, and S. Kretzschmar, *Updated angular analysis of the rare decay $B_s^0 \to \phi\mu^+\mu$*, CERN Document Server (2020), https://cds.cern.ch/record/2730622.

[109] P. d' Argent *et al.*, *Improved performance of the lhcb outer tracker in lhc run 2*, Journal of Instrumentation **12** (2017) P11016–P11016.

[110] LHCb Collaboration, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2018) P04013. 43 p, arXiv:1812.10790,

46 pages, 35 figures, 1 table. All figures and tables are available at https://cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-DP-2019-001.html.

[111] T. Likhomanenko *et al.*, *LHCb Topological Trigger Reoptimization*, J. Phys. : Conf. Ser. **664** (2015) 082025. 8 p, arXiv:1510.00572, 21st International Conference on Computing in High Energy Physics (CHEP2015).

[112] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, .

[113] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[114] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[115] A. Hoecker *et al.*, *TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide.*, arXiv:physics/0703039.

[116] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[117] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: Bounds for k-fold and progressive cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, COLT '99, (New York, NY, USA), 203–208, ACM, 1999.

[118] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, doi: 10.1142/9781860948985_0039.

[119] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001.

[120] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, 2016.

[121] S. Easo *et al.*, *Simulation of LHCb RICH detectors using GEANT4*, .

[122] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, Journal of Instrumentation **10** (2015) P02011–P02011.

[123] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, 2014.

[124] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited*, Phys. Rev. **D71** (2005) 014029, arXiv:hep-ph/0412079.

[125] S. Descotes-Genon and J. Virto, *Time dependence in $B \to V\ell\ell$ decays*, Journal of High Energy Physics **2015** (2015) .

[126] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, arXiv:1810.08132.

[127] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules*, JHEP **08** (2016) 098, `arXiv:1503.05534`.

[128] N. Rajeev, N. Sahoo, and R. Dutta, *Angular analysis of $B_s \to f_2'(1525)(\to K^+K^-)\,\mu^+\mu^-$ decays as a probe to lepton flavor universality violation*, Phys. Rev. D **103** (2021) 095007, `arXiv:2009.06213`.

[129] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi pK^-$ decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[130] W. Detmold and S. Meinel, *$\Lambda_b \to \Lambda\ell^+\ell^-$ form factors, differential branching fraction, and angular observables from lattice qcd with relativistic b quarks*, Phys. Rev. D **93** (2016) 074501.

[131] LHCb collaboration, R. Aaij *et al.*, *Study of the productions of Λ_b^0 and $\overline{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda_b^0 \to J/\psi pK^-$ branching fraction*, Chin. Phys. **C40** (2016) 011001, `arXiv:1509.00292`.

[132] L. Mott and W. Roberts, *Rare dileptonic decays of Λ_b in a quark model*, Int. J. Mod. Phys. A **27** (2012) 1250016, `arXiv:1108.6129`.

[133] LHCb collaboration, R. Aaij *et al.*, *Amplitude analysis and branching fraction measurement of $\overline{B}_s^0 \to J/\psi K^+K^-$*, Phys. Rev. **D87** (2013) 072004, `arXiv:1302.1213`.

[134] H. Dembinski, *Approximate PDF for the invariant mass distribution of combinatorial background*, `https://github.com/HDembinski/essays/blob/master/invariant%20mass%20combinatorial%20background.ipynb`, 2008. [Online; accessed 13-October-2021].

[135] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[136] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014) 090001.

[137] D. Casadei, *Estimating the selection efficiency*, JINST **7** (2012) P08021, `arXiv:0908.0130`.

[138] LHCb collaboration, R. Aaij *et al.*, *Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B_s^0 decay branching fractions*, Phys. Rev. **D104** (2021) 032005, `arXiv:2103.06810`.

[139] LHCb collaboration, R. Aaij *et al.*, *Observation of $\overline{B}_s^0 \to J/\psi f_2'(1525)$ in $J/\psi K^+K^-$ final states*, Phys. Rev. Lett. **108** (2012) 151801, `arXiv:1112.4695`.

[140] Belle collaboration, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays*, Phys. Rev. D **90** (2014) 112009, `arXiv:1408.6457`.

[141] Belle collaboration, K. Chilikin *et al.*, *Experimental constraints on the spin and parity of the $Z(4430)^+$*, Phys. Rev. D **88** (2013) 074026, arXiv:1306.4894.

[142] LHCb collaboration, R. Aaij *et al.*, *Resonances and CP-violation in B_s^0 and $\overline{B}_s^0 \to J/\psi K^+K^-$ decays in the mass region above the $\phi(1020)$*, JHEP **08** (2017) 037, arXiv:1704.08217.

[143] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[144] B. Dey, *Angular analyses of exclusive $\overline{B} \to X\ell_1\ell_2$ with complex helicity amplitudes*, Phys. Rev. D **92** (2015) 033013, arXiv:1505.02873.

[145] R.-H. Li, C.-D. Lu, and W. Wang, *Branching ratios, forward-backward asymmetries and angular distributions of $B \to K_2^* l^+ l^-$ in the standard model and new physics scenarios*, Phys. Rev. D **83** (2011) 034034, arXiv:1012.2129.

[146] Y.-B. Zuo *et al.*, *$B_{(s)}$ to light tensor meson form factors via LCSR in HQEFT with applications to semileptonic decays*, Eur. Phys. J. C **81** (2021) 30.

[147] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules*, JHEP **08** (2016) 098, arXiv:1503.05534.

[148] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945.

[149] Belle-II collaboration, W. Altmannshofer *et al.*, *The Belle II Physics Book*, PTEP **2019** (2019) 123C01, arXiv:1808.10567, [Erratum: PTEP 2020, 029201 (2020)].

[150] F. Von Hippel and C. Quigg, *Centrifugal-barrier effects in resonance partial decay widths, shapes, and production amplitudes*, Phys. Rev. D **5** (1972) 624.

Danksagung

Diese Doktorarbeit ist alles andere als allein im stillen Kämmerlein entstanden – ganz im Gegenteil. Ohne die Zusammenarbeit mit den vielen Menschen, die diese Arbeit begleitet haben, wäre das Projekt nicht möglich gewesen.

Zuallererst möchte ich meinem Erstbetreuer Christoph Langenbruch danken: Danke für die Möglichkeit in dieser wunderbaren Gruppe mitzuarbeiten und dabei sogar ein Jahr zum CERN gehen zu können; danke für alles, was ich von dir gelernt habe.

Dann möchte ich Herrn Professor Schael danken für seine Bereitschaft, diese Arbeit als Zweitgutachter zu betreuen.

Ein weiterer Dank gilt meiner Aachener Arbeitsgruppe, namentlich Jan-Marc Basels, Leon Carus, Stephan Escher, Andreas Güth, Johannes Heuel, Hendrik Jage, Marcel Materok, Dan Moise, Simon Nieswand, Thomas Oeser, Gedas Sarpis und Sebastian Schmitt. Ich glaube nicht, dass es viele Gruppen gibt, bei denen die Atmosphäre so gut und der Zusammenhalt so stark ist, die sich gegenseitig so unterstützen wie in unserer. Mit euch habe ich gelernt, wie wichtig es ist im Team zusammenzuarbeiten, und dass dabei die besten Ergebnisse rauskommen; und dass es so am meisten Spaß macht.

Ein besonderer Dank gilt Eluned Smith, die diese Arbeit als Postdoktorandin und Freundin begleitet hat. Danke für die viele Zeit, die du dir für mich genommen hast, für die vielen Diskussionen über Physik und die Welt, und für deine unendliche Geduld mit uns.

Many thanks also to Mark Whitehead, whose expertise helped to track down the $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decay, und an Leila Ali Cavasonza, meiner Theoretikerin des Vertrauens; danke an die System-Admins Robin Sonnabend und Thomas Kress, und an Henning Gast, der immer auf der Seite der Studierenden ist.

Nicht vergessen möchte ich auch die, die eher weniger Zeit in Physikzentren verbringen und möglicherweise nie was mit Physik zu tun haben wollten (geschweige denn Teilchenphysik!), mich aber trotzdem bedingungslos unterstützen: Ich danke meinen Eltern, meinem Bruder, meiner Schwester und meinen Freund*innen. Und Florian natürlich. Danke.