

# **Beautiful leptons - setting limits to New Physics with the LHCb experiment**

**Dissertation zur Erlangung des akademischen Grades  
Dr. rer. nat.**

vorgelegt von  
Titus Mombächer  
geboren am 31.05.1993 in Gummersbach

Fakultät Physik  
Technische Universität Dortmund

Dortmund  
September 2020

Gutachter:

Prof. Dr. Johannes Albrecht

Prof. Dr. Kevin Alexander Kröniger

Vorsitzende der Prüfungskommission:

Prof. Dr. Gudrun Hiller

Vertreter der wissenschaftlichen Mitarbeiter/-innen:

Dr. Ilya Akimov

Datum der mündlichen Prüfung:

25. November 2020

“The essential is invisible to the eye.”

*Antoine de Saint-Exupéry*



## Abstract

In this thesis, searches for physics effects beyond the Standard Model are performed by studying purely leptonic and lepton-flavour violating  $B$ -meson decays with the LHCb experiment. Advanced statistical tools to interpret nulltests are implemented and used in the analysis of the very rare decays discussed in this thesis. A search for lepton-flavour violating  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays is performed based on the LHCb Run 1 sample as the first analysis of these transitions with LHCb data. Upper limits on their branching fractions are set below  $10^{-8}$ , which improves previous limits by more than an order of magnitude. Similarly,  $B_{(s)}^0 \rightarrow e^+ e^-$  decays have never before been studied with the LHCb experiment. The upper limits on their branching fractions determined in this thesis with Run 1 and part of Run 2 data improve previous limits by a factor of 30, reaching a sensitivity of  $\mathcal{O}(10^{-8})$ . With this improvement the limits become sensitive to scenarios of new physics effects. In a third analysis, the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays are studied with the full LHCb Run 1 and Run 2 data sample. The analysis is currently being reviewed by the collaboration. Its precision is expected to be similar to that of the combination of previous results of the LHC experiments obtained on a partial data set. The results obtained in this thesis impose strong constraints on popular scenarios of new physics effects.

## Kurzfassung

Diese Arbeit beschäftigt sich mit der Suche nach physikalischen Effekten jenseits des Standardmodells. Dazu werden rein leptonische und die Lepton-Flavourzahl verletzende Zerfälle von  $B$ -Mesonen mit Daten des LHCb Experimentes untersucht. Statistische Methoden zur Interpretierung von Nulltests wurden im Verlauf der Arbeit implementiert und für die Analysen der sehr seltenen Zerfälle verwendet. Eine Suche nach die Lepton-Flavourzahl verletzenden  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  Zerfällen wurde mit dem Run 1 Datensatz des LHCb Experimentes durchgeführt. Dies ist die erste Untersuchung dieser Zerfälle mit LHCb Daten. Obere Ausschlussgrenzen auf ihre Verzweungsverhältnisse werden zu kleiner als  $10^{-8}$  bestimmt, was eine Verbesserung vorheriger Ausschlussgrenzen um mehr als eine Größenordnung bedeutet. Auch  $B_{(s)}^0 \rightarrow e^+ e^-$  Zerfälle wurden in dieser Arbeit zum ersten Mal mit dem LHCb Experiment untersucht. Die oberen Ausschlussgrenzen auf diese Zerfälle wurden mithilfe des LHCb Datensatzes von Run 1 und eines Teils von Run 2 um einen Faktor 30 verbessert worden auf  $\mathcal{O}(10^{-8})$ . Damit stoßen die Ausschlussgrenzen in einen Bereich vor, der sensitiv auf neue physikalische Effekte ist. In einer dritten Studie wurden  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Zerfälle mit dem vollen Run 1 und Run 2 Datensatz des LHCb Experimentes untersucht. Die Analyse wird zurzeit von der Kollaboration begutachtet. Es wird erwartet, dass die Präzision der Analyse der Kombination aller vorherigen LHC Messungen dieser Zerfälle entspricht. Die Ergebnisse dieser Arbeit führen zu starken Einschränkungen von gängigen Szenarien für neue physikalische Effekte.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Rare decays in the Standard Model and beyond</b>	<b>3</b>
2.1	Lepton-flavour violating $B^+ \rightarrow K^+ \mu^\pm e^\mp$ decays . . . . .	6
2.2	$B_{(s)}^0 \rightarrow \ell^+ \ell^-$ decays . . . . .	8
<b>3</b>	<b>Methods to obtain confidence intervals and upper limits</b>	<b>13</b>
3.1	Definition of confidence intervals and limits . . . . .	13
3.2	Statistical tests . . . . .	13
3.3	The $CL_S$ method . . . . .	18
3.4	Treatment of nuisance parameters . . . . .	20
3.5	Sideband subtraction vs. shape-based method . . . . .	21
<b>4</b>	<b>The LHCb experiment at the LHC</b>	<b>23</b>
4.1	The LHCb detector . . . . .	24
4.2	Bremsstrahlung reconstruction at LHCb . . . . .	29
4.3	Simulation at LHCb . . . . .	31
<b>5</b>	<b>Searches for <math>B^+ \rightarrow K^+ \mu^\pm e^\mp</math> and <math>B_{(s)}^0 \rightarrow e^+ e^-</math></b>	<b>33</b>
5.1	Analysis strategies . . . . .	33
5.2	Preselection . . . . .	35
5.3	Data–simulation corrections . . . . .	43
5.4	Multivariate and particle identification selection . . . . .	48
5.5	Efficiency calculations . . . . .	56
5.6	Background estimation . . . . .	68
5.7	Evaluation of systematic uncertainties . . . . .	76
5.8	Results of the searches for $B^+ \rightarrow K^+ \mu^\pm e^\mp$ and $B_{(s)}^0 \rightarrow e^+ e^-$ . . . . .	84
<b>6</b>	<b>Analysis of <math>B_{(s)}^0 \rightarrow \mu^+ \mu^-</math> decays with the full LHCb data set</b>	<b>93</b>
6.1	Analysis strategy and validation of the 2017 and 2018 data sets . . . . .	94
6.2	Corrections for data-simulation differences . . . . .	98
6.3	Normalisation . . . . .	103
6.4	Background estimation . . . . .	106
6.5	Expected results . . . . .	109

<b>7</b>	<b>Extrapolations to the upgrades of the LHCb experiment</b>	<b>117</b>
7.1	Upgrades and future data sets of the LHCb experiment . . . . .	117
7.2	Extrapolation of the sensitivity . . . . .	119
<b>8</b>	<b>Summary</b>	<b>125</b>
	<b>Bibliography</b>	<b>127</b>
<b>A</b>	<b>Appendix: Selection of <math>B_{(s)}^0 \rightarrow \mu^+\mu^-</math> decays</b>	<b>141</b>
	<b>Acknowledgements</b>	<b>143</b>



# 1 Introduction

In particle physics, three of the four fundamental forces of nature can be unified in one model, the Standard Model (SM). It was developed in the 1960's [1–3] and describes electromagnetic, weak and strong interactions between elementary particles. Gravitation, as the fourth interaction between particles, is not included in the SM. Almost all measurements of elementary particle processes have confirmed the SM calculations, including the discovery of the Higgs boson in 2012 [4, 5], the last until then undiscovered particle predicted by the SM. Despite this overwhelming success, there is a number of phenomena that suggest an incompleteness of the SM. Cosmological observations suggest that the matter described in the SM makes up only 5% of the energy density of the universe. The rest, called “dark matter” and “dark energy”, is very little known. Furthermore, all attempts to unify gravitation with the other three interactions have not been successful. Other reasons can be found in the internal structure of the SM itself. The masses of particles span the wide range of  $\mathcal{O}(1 \text{ eV}/c^2)$  to  $\mathcal{O}(100 \text{ GeV}/c^2)$ . They arise in the SM from different couplings of the particles to the Higgs field. However, there is no explanation for the large range of couplings. From these considerations it becomes evident that there must be additional structures that are not modelled by the SM.

In the past decades many searches have been performed for experimental evidence of particles and processes beyond the SM, often referred to as “New Physics” (NP). Direct searches of NP particles produced by or decaying into a set of known particles are limited by the energy provided by the laboratory. At the Large-Hadron-Collider, particles with masses of up to  $\mathcal{O}(1 \text{ TeV})$  can be produced and are therefore be accessible by direct searches. A different approach to search for NP is the precise measurement of SM processes. These could be impacted by NP processes through interference. As the new physics particles can contribute as quantum loop corrections and do not need to be produced as real states, this approach allows to test also energy scales much higher than those accessible by direct searches, up to  $\mathcal{O}(100 \text{ TeV})$  [6]. Particularly interesting for these kinds of measurements are processes that are very rare in the SM such that NP contributions can have a large relative effect.

The LHCb experiment is designed for the study of  $b$ - and  $c$ -hadrons. As such it is ideally suited to perform measurements of  $B$ -meson decay processes that are rare in the SM. And indeed recent measurements of decays involving  $b \rightarrow s\ell^+\ell^-$  processes<sup>1</sup> ( $\ell$  being a lepton) show intriguing hints of deviations from the SM [7–9]. While each of the measurements taken by itself does not differ significantly from the SM, they can be interpreted in a common framework of effective field theories and show strong tensions with SM predictions in

---

<sup>1</sup>Charge conjugated processes are implied throughout this thesis, if not otherwise specified.

combinations [10–14]. In particular the measurements of electroweak  $b$ -hadron decays seem to indicate a different coupling strength of electrons and muons in weak decays, which is in contradiction with the SM. It is therefore crucial to measure these processes more precisely with as many decay modes and as much data as possible.

This thesis extends the study of  $b \rightarrow s\ell^+\ell^-$  transitions to the search for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+e^-$  decays, which have negligible branching fractions in the SM and have not been previously studied with the LHCb experiment. If the  $b \rightarrow s\ell^+\ell^-$  are not lepton-universal as indicated by the measurements referred to above, then it is very likely that lepton-flavour violating decays like  $B^+ \rightarrow K^+\mu^\pm e^\mp$  exist with rates that are accessible with the LHCb experiment [15]. Also some NP scenarios propose new scalar and pseudoscalar mediator particles to explain the tensions. Those would however produce measurable branching fractions of  $B_{(s)}^0 \rightarrow e^+e^-$  decays [16]. Furthermore, the golden rare decay modes  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  are studied in this thesis. Their simplicity in theoretical description as well as in the experimental signature makes them ideal probes of the SM [17]. Indeed a similar small discrepancy between experiment and SM prediction is seen in the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fractions compared to what is found in measurements of other  $b \rightarrow s\ell^+\ell^-$  processes. It is therefore very interesting to measure those decays as precisely as possible to understand whether the discrepancy is a statistical fluctuation or not.

To allow statistical interpretations of the measurements, tools are developed in this thesis that apply common statistical procedures to determine upper limits on physics observables like branching fractions.

The thesis begins with a brief introduction to the SM and the decays analysed in Chapter 2. Chapter 3 reviews the statistical methods implemented to calculate upper limits on branching fractions. After a short description of the LHCb experiment in Chapter 4 the searches for  $B_{(s)}^0 \rightarrow e^+e^-$  and  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays are described in Chapter 5 and the status of the analysis of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays is presented in Chapter 6. An extrapolation of the measurements to future data sets is given in Chapter 7. The thesis then concludes with a summary in Chapter 8.

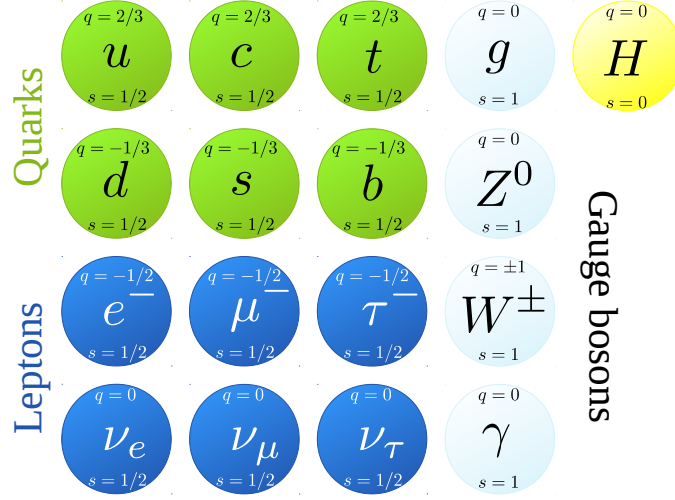
## 2 Rare decays in the Standard Model and beyond

The Standard Model (SM) is a theoretical framework that describes interactions between elementary particles. It is a gauge theory of the form  $SU(3)_C \times SU(2)_L \times U(1)_Y$ , in which the interactions between the particles follow from the invariance under local symmetry transformations [1–3] (for a detailed introduction see e.g. Refs. [18, 19]). The  $SU(3)_C$  corresponds to the strong interaction, while the  $SU(2)_L \times U(1)_Y$  group corresponds to the electroweak interaction. Gravitation, which is negligible at the mass scales of elementary particles, is not described by the SM.

The known elementary particles are of fermionic nature, i.e. have half-integer spin, and can be distinguished by their “flavour”, a property unique for each particle. They can be grouped into six quarks and six leptons. Quarks and leptons can further be organized in three generations as shown in Fig. 2.1. The lepton generations consist of one particle with electrical charge of  $-1e$ , the electron ( $e^-$ ), muon ( $\mu^-$ ) or tauon ( $\tau^-$ ), and a corresponding neutrino, which is electrically neutral. Quarks, carry an additional colour charge, which can take the three different states  $[r, g, b]$ . Furthermore, the up-type quarks up ( $u$ ), charm ( $c$ ) and top ( $t$ ) carry an electrical charge of  $+2/3e$ , while the down-type quarks down ( $d$ ), strange ( $s$ ) and beauty ( $b$ ) carry an electrical charge of  $-1/3e$ . All leptons, however, are neutral under colour charge. Thus there are twelve fundamental particles, of which each has an antiparticle with opposite charges, but otherwise exactly the same properties. Each of the particles receives a mass, the lowest being assigned to neutrinos ( $< 1.1 \text{ eV}/c^2$  [21]) and the highest being assigned to top quarks ( $\approx 173 \text{ GeV}/c^2$  [21]).

The interactions between the fundamental particles are mediated via twelve bosonic particles: a massless photon  $\gamma$  as the mediator of the electromagnetic force, the massive bosons  $W^\pm$  and  $Z^0$  governing weak interactions and eight massless gluons for the strong force. The gluons couple to the colour charge of particles and themselves also carry colour charge. Thus they can self-interact. This introduces the effect that the strong force is small at short distances and high energies and large at long distances and low energies. Therefore quarks always appear confined in colourless bound states like mesons (two-quark systems like the  $B_s^0$  meson with quark content  $[\bar{b}s]$ ) and baryons (three-quark systems like the proton  $p$  with quark-content  $[uud]$ ). Recently even the existence of Pentaquarks [22] and Tetraquarks [23] have been confirmed by the LHCb experiment.

The weak force only acts on particles with left-handed chirality. For massless particles the chirality coincides with the projection of the particle’s spin to its momentum, a property

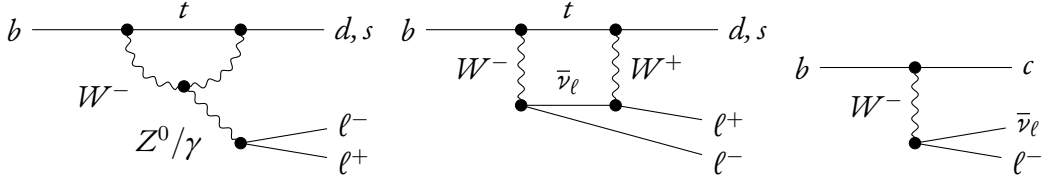


**Figure 2.1:** Fundamental particles in the SM [20]. For each particle the spin  $s$  and electrical charge  $q$  quantum numbers are given.

called helicity, where a positive value is defined as right-handed and a negative value as left-handed. Massive particles always carry a fraction of left-handed and right-handed chirality, where the fraction is governed by the particles' masses. The chirality requirements can lead to strong rate suppressions, as e.g. for  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  decays. Another particularity of the weak force with respect to all other known interactions is that it can change the flavour of particles in charged currents. These transitions occur dominantly within one generation. However, also transitions between different generations of quarks are observed and even transitions between neutrinos are seen [24–26]. These are accommodated in the SM for the quarks by the CKM matrix [27, 28] that transforms the flavour eigenstates ( $d, s, b$ ) of the quarks with respect to the weak interaction into their mass eigenstates ( $d', s', b'$ ) by a linear combination

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \mathbf{V}_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.1)$$

For neutrinos the mixing can be described by a similar procedure with the PMNS matrix [29–31]. No transitions between different generations of charged leptons are observed. This leads to a universal coupling strength of charged leptons as an accidental symmetry of the SM. Particle masses are generated in the SM by the so-called Brout-Englert-Higgs mechanism [32–34]. They arise from the Yukawa-coupling of the particles to the Higgs field. Attached to the Higgs field is the Higgs boson  $H^0$ , which was confirmed only in 2012 by the ATLAS [4] and CMS experiments [5] at the Large Hadron Collider.



**Figure 2.2:** The dominant SM processes contributing to the  $b \rightarrow s\ell^+\ell^-$  and  $b \rightarrow d\ell^+\ell^-$  FCNC transitions on the left and centre can only proceed via loops of charged currents and have therefore suppressed rates with respect to e.g. the charged current “tree” process on the right.

This completes the SM. However, many questions are still left unanswered by the SM, such as the origin and composition of dark matter, the connection of gravity with the other forces, the hierarchy of particle masses and more. Thus there must be new physics (NP) mechanisms beyond the SM that produce the observed theoretical and experimental structures. These mechanisms could become visible in the measurement of processes which are strongly suppressed in the SM.

The decays studied in this thesis mainly proceed via  $b \rightarrow s\ell^+\ell^-$  and  $b \rightarrow d\ell^+\ell^-$  processes, which are flavour-changing neutral currents (FCNC). As there is no flavour-changing neutral interaction in the SM, those can only be induced by loops of multiple charged processes and are thus suppressed with respect to processes that can occur directly via “tree” processes, as illustrated in Fig. 2.2. Therefore they become an ideal place to look for contributions from NP processes, where FCNC interactions might appear as tree level processes and thus have a strong impact on the decay rates. This is even more true for the lepton-flavour violating  $b \rightarrow s\ell^+\ell'^-$ , the core processes for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays, which would require a transitions between charged lepton flavours and are therefore even forbidden in the SM.

In the search for experimental deviations from the SM in  $b$ -hadron decays and their combinations it has become useful to reformulate SM calculations in an effective field theory approach that employs an operator product expansion [35, 36] such that for  $b \rightarrow s\ell^+\ell^-$  and  $b \rightarrow d\ell^+\ell^-$  processes one can use the general form for the transition amplitude  $\mathcal{A}$  of a  $B$ -meson into a final state  $f$

$$\mathcal{A}(B \rightarrow f) = \langle f | \mathcal{H}_{\text{eff}} | B \rangle = \frac{G_F}{\sqrt{2}} \sum_{i=7,9,10} V_{tb} V_{t(d,s)}^* C_i(\mu) \langle f | \mathcal{Q}_i(\mu) | B \rangle. \quad (2.2)$$

Here  $\mathcal{H}_{\text{eff}}$  is the effective Hamiltonian of the transition,  $G_F$  the Fermi constant, and  $C_i$  and  $\mathcal{Q}_i$  the Wilson coefficients and operators, which depend on the energy scale of the mass of the  $b$ -quark  $\mu = m_b$ . In this formulation, the Wilson coefficients comprise perturbatively calculable non-local contributions, while the operators describe the local quark-level contributions. This factorisation allows to determine and overconstrain the quark-level contributions from multiple measurements with different decay modes. The relevant operators contributing to

$b \rightarrow s\ell^+\ell^-$  and  $b \rightarrow d\ell^+\ell^-$  transitions are depicted in the Feynman diagrams in Fig. 2.2. They comprise vector ( $Q_9$ ), axial-vector ( $Q_{10}$ ) transitions and photon radiation ( $Q_7$ ).

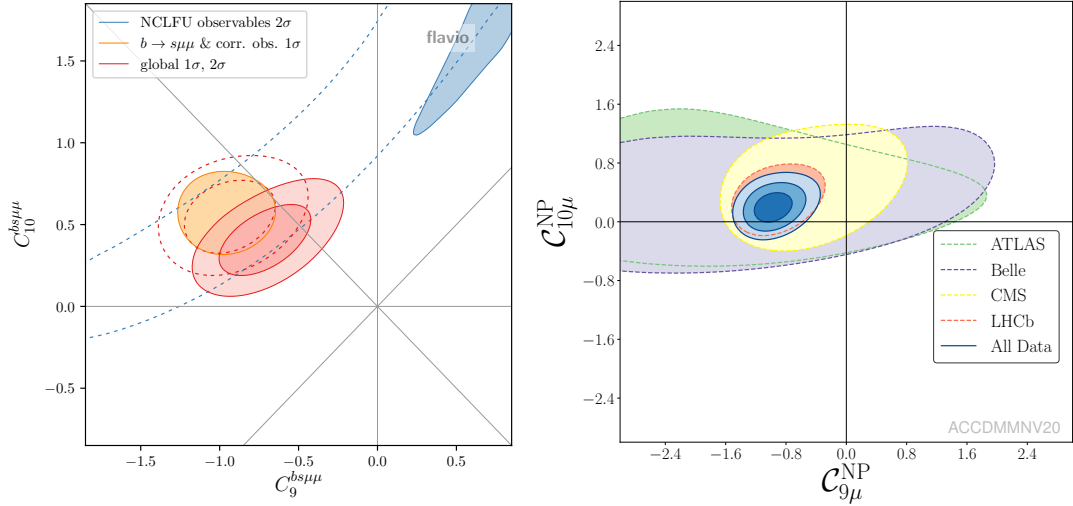
In recent measurements of  $b \rightarrow s\ell^+\ell^-$  processes, in particular differential branching fractions [37–40], angular distributions [41–44] and lepton-flavour universality ratios [7–9, 45–47], deviations from the SM are measured. While each of the deviations is insignificant by itself, in global fits of Wilson coefficients they show a clear shift to lower values of  $C_9^{\mu\mu}$  with respect to the SM [10–14]<sup>1</sup> as shown in Fig. 2.3. Also a small shift in the Wilson coefficient  $C_{10}^{\mu\mu}$  is seen, whose size varies in the global fits depending on how many observables are included and on whether non-SM effects are allowed in the calculation of meson mixing processes. Moreover they suggest a non-universal coupling of leptons, as also indicated by lepton-universality ratio measurements in charged current  $b \rightarrow c\ell\nu$  decays (for a combination of those measurements, see Ref. [49]). Especially important in the context of this thesis, the measurements of the lepton-flavour universality ratio  $R_{K^+} = \frac{B^+ \rightarrow K^+ \mu^+ \mu^-}{B^+ \rightarrow K^+ e^+ e^-}$  and the similar  $R_{K^{*0}} = \frac{B^0 \rightarrow K^{*0} \mu^+ \mu^-}{B^0 \rightarrow K^{*0} e^+ e^-}$  measurements hint at different interactions of electrons and muons in electroweak  $b$ -hadron decays. This contrasts against other measurements of weak interactions, like  $Z^0 \rightarrow \ell^+\ell^-$  decays, where the interactions of electrons and muons have been tested to be similar at great precision [21].

Many possibilities of interactions beyond the SM have been proposed to account for the observed deviations (for a detailed review see e.g. Ref. [36]). Among them the most prominent ones include heavy new mediator particles like  $Z'$  bosons and leptoquarks (those allow the transition from quarks to leptons), as well as additional Higgs-like particles. These models have in common that they introduce on the one hand “tree-level” FCNC interactions and on the other hand they allow interactions that are lepton-flavour violating. Therefore, they imply altered rates in purely leptonic  $b \rightarrow s\ell^+\ell^-$  decays and sizeable rates of decays including lepton-flavour violation. The following sections give an overview of the state-of-the-art SM calculations for the decays studied in this thesis and predictions with proposed NP models, as well as the experimental status without including the results obtained in this thesis.

## 2.1 Lepton-flavour violating $B^+ \rightarrow K^+ \mu^\pm e^\mp$ decays

The decays  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  are not allowed in the SM because they would require a transition from electrons to muons. By means of neutrino mixing lepton-flavour violating decays can be achieved through loop processes, however the rates are  $\mathcal{O}(10^{-40})$  and therefore negligible [50]. Thus the observation of charged lepton-flavour violating decays would be an unambiguous sign for NP processes.

<sup>1</sup>Refs. [10, 48] include also the latest update of the angular analysis of  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  with the LHCb experiment [41]



**Figure 2.3:** Global fits of  $b \rightarrow s\ell^+\ell^-$  processes to Wilson coefficients from (left) Refs. [11] and (right) [10], displayed in the  $C_9^{\mu\mu} - C_{10}^{\mu\mu}$  plane (though the notations are different, they are equivalent in meaning). The coordinate systems are defined such that the SM prediction corresponds to  $(0, 0)$ . The left plot does not contain the latest update of the angular analysis of  $B^0 \rightarrow K^{*0}\mu^+\mu^-$  [41]. In that plot, blue contours correspond to combinations from lepton-universality ratio measurements and the yellow contours to all other observables, while the red contours show the full combination. The dashed lines indicate the results excluding the most recent measurements of lepton-universality ratios [8, 45, 46]. The right plot shows combinations split by several experiments. The total combination is shown in blue and given with  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  contours. While both combinations show a clear shift in  $C_9^{\mu\mu}$  with respect to the SM, in  $C_{10}^{\mu\mu}$  also a small shift is visible in the left plot when requiring that meson-mixing processes are not affected by NP processes.

Indeed, already from a naive point of view it is rather natural that charged lepton-flavour violating processes should exist, since transitions between neutral lepton flavours have also been observed, as have transitions between different quark flavours. Concerning flavour transitions, charged leptons have therefore a unique role, for which there is no explanation. Furthermore, if the measurements indicating different couplings of electrons and muons are confirmed, this is necessarily associated with transitions between charged lepton flavours [15]. In fact, most models proposed to solve the  $b \rightarrow s\ell^+\ell^-$  anomalies, particularly the deviations from the SM in lepton-universality ratios, come with sizeable branching fractions of lepton-flavour violating decays. For example models involving leptoquarks propose branching fractions of  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays in the range of  $\mathcal{O}(10^{-10} - 10^{-8})$  [51–53], while additional heavy neutral bosons could also enhance the branching fractions up to the level of  $\mathcal{O}(10^{-8})$  [54]. More precisely, they connect the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  branching fractions to the anomalies seen in  $b \rightarrow s\ell^+\ell^-$  transitions for example via [51]

$$\mathcal{B}(B^+ \rightarrow K^+\mu^\pm e^\mp) = (31 \pm 4) \times 10^{-9} \left[ \left( \frac{\mathcal{C}_9^{ee}}{\gamma} \right)^2 + (\gamma \mathcal{C}_9^{\mu\mu})^2 \right] \quad (2.3)$$

with a parameter  $\gamma$  that governs the relative strength of leptoquark couplings to electrons and muons or via [52, 53]

$$\mathcal{B}(B^+ \rightarrow K^+\mu^\pm e^\mp) = 3 \times 10^{-8} \kappa^2 \left( \frac{1 - R_{K^+}}{0.23} \right)^2 \quad (2.4)$$

with a strength parameter  $\kappa \approx \mathcal{O}(1)$ .

In a different approach, the  $b \rightarrow s\ell^+\ell^-$  anomalies could be explained by introducing a significant charged-lepton mixing similar to the mixing found in neutrinos, which would enhance the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  branching fractions up to  $1.8 \times 10^{-9}$  [55, 56]. The various NP models are illustrated with possible diagrams of  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays in Fig. 2.4.

The predictions from the various NP scenarios quoted above are close to the experimental upper limit on the branching fractions  $\mathcal{B}$  of these decays determined with the BaBar experiment [57]:

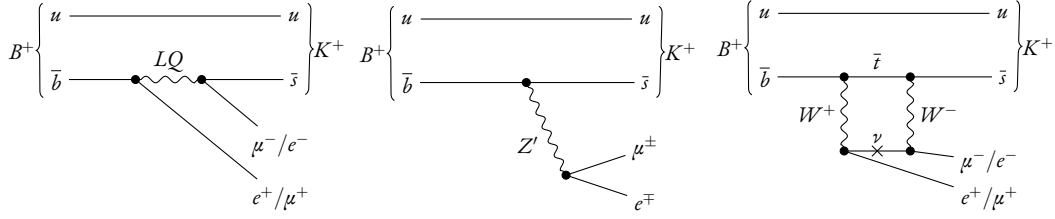
$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+\mu^- e^+) &< 9.1 \times 10^{-8} \text{ at } 90\% \text{ CL and} \\ \mathcal{B}(B^+ \rightarrow K^+\mu^+ e^-) &< 13 \times 10^{-8} \text{ at } 90\% \text{ CL.} \end{aligned}$$

Therefore the search for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays has a high potential to strongly constrain these NP scenarios.

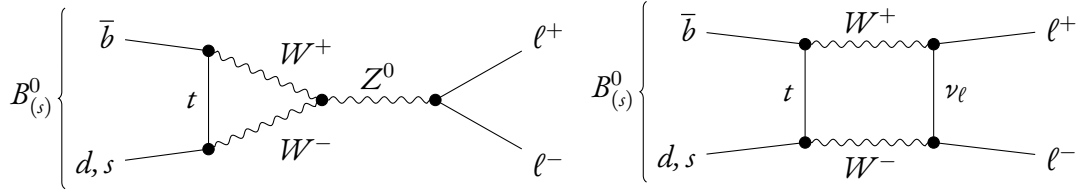
## 2.2 $B_{(s)}^0 \rightarrow \ell^+\ell^-$ decays

The rates of  $B_{(s)}^0 \rightarrow \ell^+\ell^-$  decays are suppressed by  $m_\ell^2$  due to chirality requirements, as they can only proceed via the weak interaction. Additionally these decays are suppressed by the





**Figure 2.4:** Potential processes contributing to  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays according to several NP scenarios. The left diagram describes the decays by introducing a leptoquark  $LQ$  that couples to leptons and quarks at the same time. The central diagram instead proposes an FCNC via a heavy neutral  $Z'$  boson that has lepton-flavour violating couplings. The right diagram accommodates the decays with neutrino mixing, indicated with the  $\times$ . A similar diagram can be obtained by moving the mixing process to one of the charged leptons.



**Figure 2.5:** Dominant SM processes contributing to the branching fractions of  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  decays. The quark  $q$  can be an  $s$ - or a  $d$ -quark.

CKM-factor  $|V_{tb} V_{t(d,s)}^*|^2$ , because they include FCNC's and can therefore only proceed via loop processes. Due to this strong suppression in the SM, changes due to NP processes could have a large impact.

Due to the purely leptonic final state, branching fractions of  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  decays can be very cleanly computed. The dominantly contributing SM processes are depicted in Fig. 2.5.

In the effective field theory formalism, the branching fractions  $\mathcal{B}(B_{(s)}^0 \rightarrow \ell^+ \ell^-)$  with  $\ell = e, \mu$  take the simple form [58, 59] of

$$\mathcal{B} = \tau_{B_{(s)}^0}^* \frac{G_F^2 \alpha_{\text{em}}^2 f_{B_{(s)}^0}^2 m_\ell^2 m_{B_{(s)}^0}}{16\pi^2} \sqrt{1 - \frac{4m_\ell^2}{m_{B_{(s)}^0}^2}} |V_{tb} V_{t(d,s)}^*|^2 |\mathcal{C}_{10}^{\text{SM}}|^2 (|P|^2 + |S|^2), \quad (2.5)$$

where

$$\tau_{B_{(s)}^0}^* = \frac{1 + \mathcal{A}_{\Delta\Gamma} \gamma_{s,d}}{1 - \gamma_{s,d}^2} \tau_{B_{(s)}^0} \quad (2.6)$$

with the electromagnetic coupling constant  $\alpha_{\text{em}}$  and the  $B_{(s)}^0$  meson decay constant  $f_{B_{(s)}^0}$ , the average  $B_{(s)}^0$  lifetime  $\tau_{B_{(s)}^0}$  and the relative decay width difference  $\gamma_{s,d} = \Delta\Gamma_{s,d}/(2\Gamma_{s,d})$  of

the heavy and the light eigenstate of the  $B_{(s)}^0$  meson. In the SM, the  $CP$  asymmetry parameter  $\mathcal{A}_{\Delta\Gamma} = +1$ . The parameters  $P$  and  $S$  describe additional non-SM components to the branching fractions from scalar and pseudoscalar currents with

$$P = \frac{\mathcal{C}_{10} - \mathcal{C}'_{10}}{\mathcal{C}_{10}^{\text{SM}}} + \frac{m_{B_{(s)}^0}^2}{2m_\ell} \frac{m_b}{m_b + m_{d,s}} \left( \frac{\mathcal{C}_P - \mathcal{C}'_P}{\mathcal{C}_{10}^{\text{SM}}} \right), \quad (2.7)$$

and

$$S = \sqrt{1 - \frac{4m_\ell^2}{m_{B_{(s)}^0}^2} \frac{m_b}{m_b + m_{d,s}} \left( \frac{\mathcal{C}_S - \mathcal{C}'_S}{\mathcal{C}_{10}^{\text{SM}}} \right)}. \quad (2.8)$$

Here, quark masses are described with  $m_b$  and  $m_s$  and the Wilson coefficients for scalar and pseudoscalar currents are denoted with  $\mathcal{C}_S$  and  $\mathcal{C}_P$ . Also primed coefficients are introduced, that express the right-handed counterparts of the Wilson coefficients. In the SM,  $\mathcal{C}_S$  and  $\mathcal{C}_P$  and all right-handed currents vanish and therefore  $P = 1$  and  $S = 0$ .

From Eqs. (2.5), (2.7) and (2.8) one sees that in the SM calculations of  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  branching fractions only  $\mathcal{C}_{10}$  coefficients contribute and therefore the measurements of the branching fraction of  $B_s^0 \rightarrow \mu^+ \mu^-$  determine the precision of  $\mathcal{C}_{10}$  in global fits of  $b \rightarrow s \ell^+ \ell^-$  decays [10–14]. The current state-of-the-art predictions of the branching fractions of  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  decays are [16]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.66 \pm 0.14) \times 10^{-9}, \quad (2.9)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.03 \pm 0.05) \times 10^{-10}, \quad (2.10)$$

$$\mathcal{B}(B_s^0 \rightarrow e^+ e^-) = (8.60 \pm 0.36) \times 10^{-14}, \text{ and} \quad (2.11)$$

$$\mathcal{B}(B^0 \rightarrow e^+ e^-) = (2.41 \pm 0.13) \times 10^{-15}. \quad (2.12)$$

The predictions for the  $B_{(s)}^0 \rightarrow e^+ e^-$  decays are obtained here by scaling the predictions for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  [16] by  $m_e^2/m_\mu^2$  [21] and neglecting the electromagnetic corrections discussed in Ref. [16]. These corrections are expected to be of the order  $\mathcal{O}(1\%)$ , which is added as systematic uncertainty to the prediction. However, the uncertainty on the SM predictions is dominated by the knowledge of  $|V_{tb} V_{t(d,s)}^*|$ , which can be experimentally obtained with a precision of 3.1% (4.6% for  $|V_{tb} V_{td}^*|$ ) [60, 61], and the precision of the decay constant  $f_{B_{(s)}^0}$ , which is determined from lattice QCD calculations [62].

An alternative prediction independent of  $f_{B_{(s)}^0}$  and  $|V_{tb} V_{t(d,s)}^*|$  can be obtained by relating the  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  branching fractions to the mass difference of the heavy and light eigenstate of the  $B$  mesons  $\Delta m_{B_{(s)}^0}$  [63, 64]. The alternative prediction is in full agreement with the values quoted above.

In scenarios of NP mechanisms beyond the SM, the branching fractions can be modified by scalar and pseudoscalar operators. As can be seen in Eqs. (2.5), (2.7) and (2.8), in that case

the chirality suppression that leads to the dependence on  $m_\ell^2$  is completely lifted. That allows large rate enhancements in models that introduce scalar and pseudoscalar currents. Typical examples are models with multiple Higgs-like bosons such as the minimal supersymmetric extension of the SM [65–67]. Alternatively also models with Leptoquarks can lead either to an enhancement or a suppression of the branching fractions [17]. It can be shown that by adding scalar or pseudoscalar operators to Eq. (2.5), which account for the anomalies seen in  $b \rightarrow s \ell^+ \ell^-$  measurements, the branching fractions of  $B_{(s)}^0 \rightarrow e^+ e^-$  decays can be sizeably enhanced up to the level of  $\mathcal{O}(10^{-8})$  for  $\mathcal{B}(B_s^0 \rightarrow e^+ e^-)$  and  $\mathcal{O}(10^{-10})$  for  $\mathcal{B}(B^0 \rightarrow e^+ e^-)$ , while the branching fractions of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  do not change much [68]. Furthermore, without adding scalar or pseudoscalar operators, the Wilson coefficient  $\mathcal{C}_{10}$  could be altered with respect to the SM or a right-handed component  $\mathcal{C}'_{10}$  could exist, which is forbidden in the weak interaction of the SM. These variations could be introduced by an additional heavy neutral vector boson  $Z'$  [69,70]. As the space of potential contributions beyond the SM is vast, often the proposed models are constrained by the minimal flavour violation hypothesis [71], which requires that all flavour- and  $CP$ -violating effects are connected to the known structure of quark mixing via the CKM-matrix. Under this hypothesis, the same effects in  $b \rightarrow s \ell^+ \ell^-$  decays have to be present in  $b \rightarrow d \ell^+ \ell^-$  decays as well. Therefore the ratio of the branching fractions  $\mathcal{R}_{\mu^+ \mu^-} = \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) / \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  provides a powerful tool to test this hypothesis. Using Eq. (2.5) and the inputs from Ref. [16], a prediction of

$$\mathcal{R}_{\mu^+ \mu^-} = 0.0281 \pm 0.0016$$

is obtained.

Due to the sizeable decay with differences between the heavy and the light mass eigenstate of the  $B_s^0$  meson, the effective lifetime of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay  $\tau_{B_s^0 \rightarrow \mu^+ \mu^-}$  also yields access to the  $CP$  structure of the decay via

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = \frac{\tau_{B_s^0}}{1 - \gamma_s} \frac{1 + 2\mathcal{A}_{\Delta\Gamma} \gamma_s + \gamma_s^2}{1 + \mathcal{A}_{\Delta\Gamma} \gamma_s}. \quad (2.13)$$

This measurement can support disentangling whether potential contributions beyond the SM come from a variation of  $\mathcal{C}_{10}$  or additional scalar or pseudoscalar contributions [72].

The most precise single experiment measurements of the branching fractions of  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  decays excluding the measurements performed in the course of this thesis have been achieved by the LHCb [73] and CDF [74] experiments and yield

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.0 \pm 0.6_{-0.2}^{+0.3}) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &< 3.4 \times 10^{-10} \text{ at } 95 \% \text{ CL}, \\ \mathcal{B}(B_s^0 \rightarrow e^+ e^-) &< 2.8 \times 10^{-7} \text{ at } 90 \% \text{ CL, and} \\ \mathcal{B}(B^0 \rightarrow e^+ e^-) &< 8.3 \times 10^{-8} \text{ at } 90 \% \text{ CL.} \end{aligned}$$

However, also the CMS [75] and ATLAS [76] experiments have performed similar measurements of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions, leading to combined measurements of [77]

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (2.69_{-0.35}^{+0.37}) \times 10^{-9} \text{ and} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &< 1.6(1.9) \times 10^{-10} \text{ at } 90(95) \% \text{ CL.} \end{aligned}$$

Thus the branching fraction of  $B_s^0 \rightarrow \mu^+ \mu^-$  from the combined measurement is about 25 % below the SM expectation, while being consistent with it at about 2.1 standard deviations ( $\sigma$ ). With combining the results of the several experiments, also the  $\mathcal{R}_{\mu^+ \mu^-}$  is measured to be

$$\mathcal{R}_{\mu^+ \mu^-} = 0.021_{-0.025}^{+0.030},$$

fully compatible with the SM.

The most precise single experiment measurement of the effective  $B_s^0 \rightarrow \mu^+ \mu^-$  lifetime yields [73]

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = (2.04 \pm 0.44 \pm 0.05) \text{ ps,}$$

with the combination of the CMS and LHCb results [77] being

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = (1.91_{-0.35}^{+0.37}) \text{ ps.}$$

Though the results are not precise enough for a clear statement, it is striking to note that they match very well the anomalies seen in other decays with  $b \rightarrow s \ell^+ \ell^-$  transitions that have been discussed above. It is therefore very important to improve the precision of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions and the  $B_s^0 \rightarrow \mu^+ \mu^-$  lifetime with more data to further clarify this picture. The upper limits on  $B_{(s)}^0 \rightarrow e^+ e^-$  decays set with the CDF experiment [74] are close to the region allowed by NP considerations [68]. Therefore an improvement of the upper limits has great potential of constraining those scenarios.

## 3 Methods to obtain confidence intervals and upper limits

This chapter reviews the methods to obtain limits and confidence intervals on parameters of interest that have been implemented in the GammaCombo framework [78, 79] and used in the analyses discussed in this thesis.

### 3.1 Definition of confidence intervals and limits

A confidence interval of a parameter in its frequentist definition is called an interval that contains the true value of the parameter at a given fraction, which is called confidence level (CL). That means for infinite repetitions of similar statistically independent experiments the fraction of confidence intervals containing the true value of the parameter equals the CL. In general this confidence interval can have upper and lower limits. But in cases, where measurements are expected to be close to a physical boundary, typically only upper (or lower) limits are quoted. The upper limit at a given CL is then defined as such that the fraction of limits being larger than the true value of the parameter of interest matches the CL. A typical example in High Energy Physics is the search for rare decays where often no signal is expected to be seen.

Many methods have been proposed to determine confidence intervals. To which extent the fraction of confidence intervals containing the true value of the parameter of interest matches the CL for a given method is called coverage. Methods, where the fraction containing the true value matches the CL have coverage, while methods, where the fraction is lower have under-coverage. On the other hand methods with a fraction higher than the CL have over-coverage. An optimal method has coverage, however, in practise (especially when the experiment contains many parameters) this is often only reached asymptotically with increasing statistics of the measurement, as described in many textbooks, e.g. Ref. [80] and also further discussed in Section 3.4.

### 3.2 Statistical tests

To determine confidence intervals and limits, statistical tests are used to assess the compatibility of a hypothetical value for the parameter of interest  $\mu$  with the data. This section

reviews several test statistics proposed and discussed in [81]. The test statistics designed for upper limits have been implemented in the GammaCombo framework in the course of this thesis.

A powerful test statistic is the Likelihood Ratio

$$t_\mu = -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(\hat{\mu}|x)} \right), \quad (3.1)$$

where the likelihood of the data, given a hypothetical value  $\mu$  of the parameter of interest,  $\mathcal{L}(\mu|x)$ , is normalised against the likelihood of the data given the best fitting value  $\hat{\mu}$  for the parameter of interest,  $\mathcal{L}(\hat{\mu}|x)$ . The data might be distributed according to yet a different (true) value  $\mu'$ . In the context of this thesis, the values  $\mu, \mu', \hat{\mu}$  are branching fractions of an investigated decay. Assuming a large sample and the parameter estimates far from mathematical boundaries (i.e. Gaussian likelihoods) [82, 83], one can derive the distribution of the test statistic  $t_\mu$

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} \left[ \exp \left( -\frac{1}{2} \left( \sqrt{t_\mu} + \frac{\mu - \mu'}{\sigma} \right)^2 \right) + \exp \left( -\frac{1}{2} \left( \sqrt{t_\mu} - \frac{\mu - \mu'}{\sigma} \right)^2 \right) \right], \quad (3.2)$$

where the data are distributed according to the value  $\mu'$ . The symbol  $\sigma$  denotes the standard deviation of the best fitting parameter value  $\hat{\mu}$ . This simplifies for the case  $\mu = \mu'$  to a  $\chi^2$ -distribution with one degree of freedom

$$f(t_\mu|\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t_\mu}} e^{-t_\mu/2}. \quad (3.3)$$

The test statistic  $t_\mu$  provides a natural ordering principle for hypotheses  $\mu$  and is invariant under monotonic transformations of  $\mu$  (see e.g. [80]).

Often (especially in searches for rare decays) the measurement is close to the physical boundary  $\mu > 0$ . To take that into account, it is not desirable to impose that boundary on a likelihood fit, as that biases the measurement. Instead, an alternative test statistic

$$\tilde{t}_\mu = \begin{cases} -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(\hat{\mu}|x)} \right) & \hat{\mu} \geq 0 \\ -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(0|x)} \right) & \hat{\mu} < 0 \end{cases} \quad (3.4)$$

can be proposed, whose distribution can in the large sample limit be derived to the form of

$$\begin{aligned}
 f(\tilde{t}_\mu|\mu') &= \frac{1}{2\sqrt{\tilde{t}_\mu}} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \sqrt{\tilde{t}_\mu} + \frac{\mu - \mu'}{\sigma} \right)^2 \right] \\
 &+ \begin{cases} \frac{1}{2\sqrt{\tilde{t}_\mu}} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \sqrt{\tilde{t}_\mu} - \frac{\mu - \mu'}{\sigma} \right)^2 \right] & \tilde{t}_\mu \leq \frac{\mu^2}{\sigma^2} \\ \frac{1}{(2\mu/\sigma)} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{\left( \tilde{t}_\mu - \frac{\mu^2 - 2\mu\mu'}{\sigma^2} \right)^2}{(2\mu/\sigma)^2} \right] & \tilde{t}_\mu > \frac{\mu^2}{\sigma^2}. \end{cases} \quad (3.5)
 \end{aligned}$$

This test statistic  $\tilde{t}_\mu$  is equivalent to the approach advertised by Feldman and Cousins [84], which was designed to build a transition from confidence intervals to limits at low measured hypotheses  $\hat{\mu}$  that has coverage and avoids the decision, whether to quote a confidence interval or an upper limit based a posteriori on the outcome of the measurement (the so-called “flip-flopping” problem, which introduces under-coverage at the transition point between confidence interval and limit).

In searches for rare decays, however, it is often expected that the measured branching fraction will be compatible with zero and therefore one still a priori decides to determine an upper limit. In these cases the lower limit of the confidence interval is defined to be zero and the two-sided test statistics  $t_\mu$  and  $\tilde{t}_\mu$  can further be simplified to obtain the one-sided test statistics

$$q_\mu = \begin{cases} -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(\hat{\mu}|x)} \right) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad (3.6)$$

and

$$\tilde{q}_\mu = \begin{cases} -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(0|x)} \right) & \hat{\mu} < 0 \\ -2 \ln \left( \frac{\mathcal{L}(\mu|x)}{\mathcal{L}(\hat{\mu}|x)} \right) & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu, \end{cases} \quad (3.7)$$

reasoning that only values for the parameter of interest larger than  $\hat{\mu}$  can contribute to the upper limit. The approximate test statistic distributions in the large sample limit are then

$$f(q_\mu|\mu') = \Phi \left( \frac{\mu' - \mu}{\sigma} \right) \delta(q_\mu) + \frac{1}{2\sqrt{q_\mu}} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma} \right)^2 \right] \quad (3.8)$$

and

$$\begin{aligned}
 f(\tilde{q}_\mu|\mu') &= \Phi\left(\frac{\mu' - \mu}{\sigma}\right) \delta(\tilde{q}_\mu) \\
 &+ \begin{cases} \frac{1}{2\sqrt{\tilde{q}_\mu}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\sqrt{\tilde{q}_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right] & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{2\mu/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(\tilde{q}_\mu + \mu^2/\sigma^2)}{(2\mu/\sigma)^2}\right)^2\right] & \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \quad (3.9)
 \end{aligned}$$

From the test statistic distributions Eqs. (3.2), (3.5), (3.7) and (3.9) the compatibility of the hypothesis and the data can be extracted as the p-value

$$p_\mu = \int_{q_{\text{obs}}}^{\infty} f(q_\mu|\mu') dq_\mu, \text{ or} \quad (3.10)$$

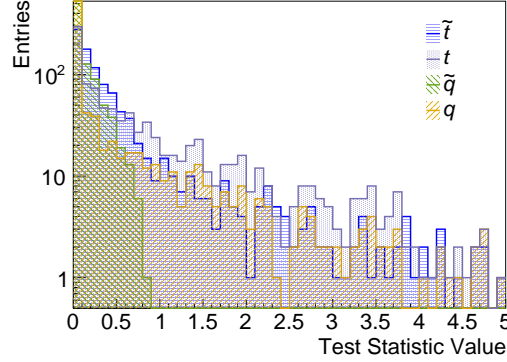
$$p_\mu = \int_{t_{\text{obs}}}^{\infty} f(t_\mu|\mu') dt_\mu \quad (3.11)$$

depending on the test statistics used. If the p-value lies below the threshold  $1 - \text{CL}$ , the test is failed and the hypothesis is rejected. A confidence interval at a given CL then contains all hypotheses with  $p_\mu > 1 - \text{CL}$ .

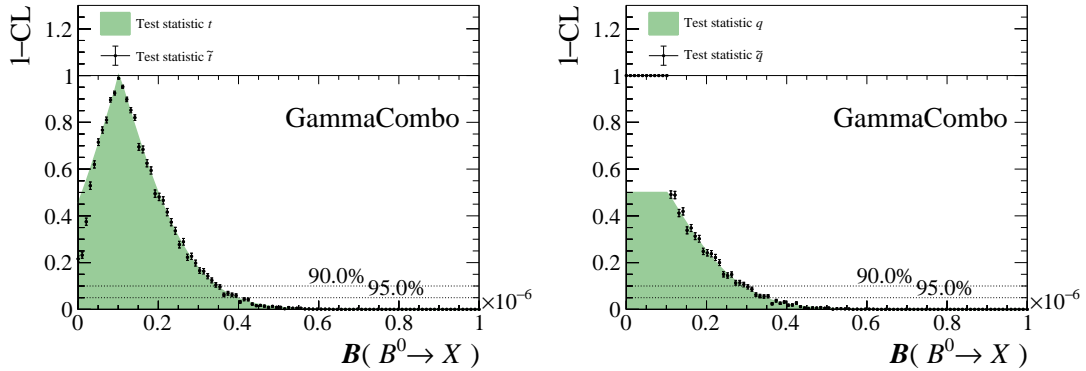
Calculating the p-values with the analytical test statistic distributions derived above is only approximately correct in the limit of large samples. Alternatively, the test statistic distributions can be sampled using pseudo experiments. This allows to remove the reliance on the large sample limit and yields a method that (in the absence of other parameters than the parameter of interest) has coverage by definition. However, often the large sample assumption is already a very good approximation, even with event counts of  $N_{\text{obs}} = 10$  in the measurement, as was e.g. studied in a bachelor thesis [85].

A comparison of the tests statistic distributions is shown in Fig. 3.1. The distributions are obtained with pseudo experiments for an example case, where the investigated value  $\mu$  of the parameter of interest is close to the physical boundary in the region  $0 < \mu < \hat{\mu}$ . In that region a clear difference between the test statistic distributions is visible. The one-sided test statistics  $q_\mu$  and  $\tilde{q}_\mu$  have a delta peak at zero, corresponding to about 50 % of the sample. This leads to the effect that an upper limit at 95 % CL using the one-sided test statistics equals an upper limit at 90 % CL using the two-sided test statistic. The test statistics  $\tilde{t}_\mu$  and  $\tilde{q}_\mu$  have lower test statistic values with respect to  $t_\mu$  and  $q_\mu$ , because the fit for the best fitting parameter value  $\hat{\mu}$  is repeated once  $\hat{\mu}$  is outside the physical boundary fixing  $\hat{\mu}$  to the closest physical boundary, leading to a decrease in the likelihood ratio. Thus also the p-values close to the physical boundary shrinks. These effects are clearly visible in Fig. 3.2, where the p-value distributions of the discussed methods are shown.

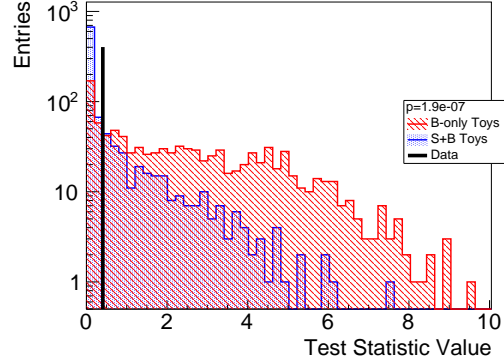




**Figure 3.1:** Distributions of the studied test statistics for an example case with  $\mu = \mu' < \hat{\mu}$  close to the physical boundary  $\hat{\mu} > 0$ . The one-sided test statistic distributions show the expected delta peak at 0. The test statistics  $\tilde{t}$  and  $\tilde{q}$  have lower test statistic value compared to  $t$  and  $q$  due to taking into account the physical boundary.



**Figure 3.2:** P-value distributions of an example case for the one- (right) and two-sided (left) test statistic. In the one-sided test statistic case the p-values have the default value of 0.5 (or 1.0) for  $\mu < \hat{\mu}$  ( $\mu \equiv \mathcal{B}(B^0 \rightarrow X)$ ). The points indicate the values for the test statistics  $\tilde{q}$  and  $\tilde{t}$ , computed with pseudo experiments, and the shaded areas indicate  $q$  and  $t$ , computed from the asymptotic formulae. While  $\tilde{q}$  and  $q$  show no difference, because  $\hat{\mu}$  is away from the physical boundary, close to the physical boundary a strong difference can be seen between  $\tilde{t}$  and  $t$ .



**Figure 3.3:** Test statistic distribution examples for the example case, at a tested value of the parameter of interest far away from the background-only hypothesis (i.e.  $\mu \gg 0$ ), where the  $q_{s+b}$  (blue) and  $q_b$  (red) distributions are different. The  $CL_{s+b}$  ( $CL_b$ ) values are computed by calculating the fraction of the blue (red) distribution right of the measured value in data.

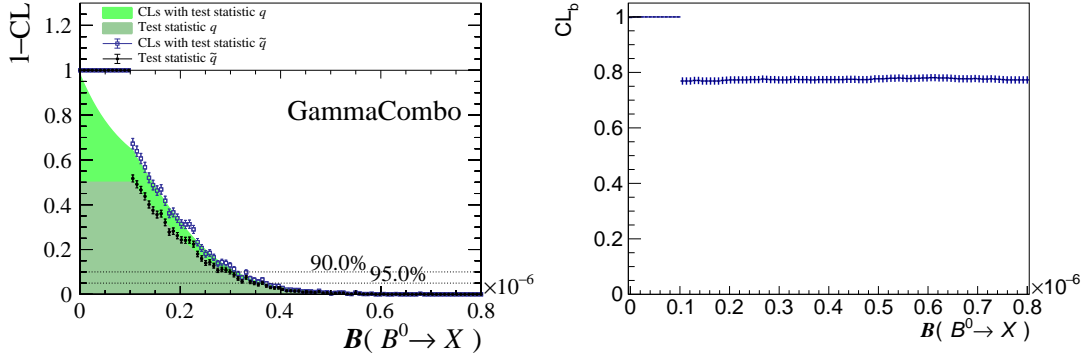
### 3.3 The $CL_s$ method

When determining upper limits on signal rates where also background is present, the methods discussed in the previous section to obtain upper limits can lead to unintuitive behaviours: consider two experiments with the same signal sensitivity, but different expected background level  $N_{\text{bkg}}$ . If now due to fluctuations both experiments observe the same amount of events  $N_{\text{obs}} = N_{\text{sig}} + N_{\text{bkg}}$ , they will set different upper limits on the signal yield  $N_{\text{sig}}$ , where the stronger limit comes from the experiment with weaker background control (higher background level).

While this effect is a valid behaviour from the statistical point of view, it shows that the tests defined in the previous section allow to set limits rather on a signal plus background hypothesis than on the signal rate alone. However, typically the physical quantity of interest is only the signal rate, not the background level. Therefore the  $CL_s$  method was proposed in Ref. [86], which normalises the p-value of the signal plus background hypothesis  $CL_{s+b} = p_\mu(\mu' = \mu)$  from Eq. (3.10) to the compatibility of the tested hypothesis with background-only data  $CL_b = p_\mu(\mu' = 0)$  to

$$CL_s = \frac{CL_{s+b}}{CL_b}. \quad (3.12)$$

An illustration of the computation of the  $CL_{s+b}$  and  $CL_b$  values is given in Fig. 3.3. With the  $CL_s$  technique, the sensitivity of the experiment to differentiate signal plus background hypotheses from the background-only hypothesis is taken into account, removing the non-physical behaviour described above. While from the considerations above the value  $CL_{s+b}$  is designed to have coverage in the parameter of interest,  $CL_s$  will then over-cover, as  $0 \leq CL_b \leq 1$ ,

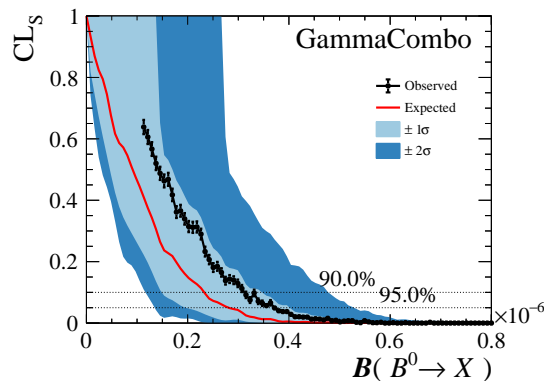


**Figure 3.4:** The left diagram shows p-value distributions for the test statistic  $q$  (dark shaded), the method using the test statistic  $\tilde{q}$  (points) and the CL<sub>S</sub> method for both test statistics (light shaded, boxes). The shaded distributions are computed with asymptotic formulae, while the ones indicated with points are calculated with pseudo experiments. On the right side the  $CL_b$  value distributions for the  $\tilde{q}$  method are shown. The CL<sub>S</sub> method shows only a small correction to the  $CL_{s+b}$  p-values. The plateau in the CL<sub>S</sub> and  $CL_b$  distributions in the left part of the diagrams for the methods with pseudo experiments occurs because the test statistic gives only trivial values ( $\tilde{q}_\mu = 0$ ) there and therefore the p-value is set to 1. However, asymptotically (as shown here) the asymptotic methods and the methods from pseudo experiments match perfectly for all values above the best fit, i.e.  $\mu > \hat{\mu}$ .

and thus set a more conservative limit than described by the confidence level. However, in the limit that the test allows a clear separation between signal plus background and background-only hypotheses, the  $CL_b$  values will be 1, effectively yielding  $CL_S \approx CL_{s+b}$ . The effect of the CL<sub>S</sub> method on the previously used example is shown in Fig. 3.4, where the one-sided test statistic is used. As the data show an upwards fluctuation with respect to the background-only hypothesis  $\mu = 0$ , only a small correction from the CL<sub>S</sub> method compared to the  $CL_{s+b}$  values is found. The stronger the downward fluctuation of the data is, the stronger the corrections from the  $CL_b$  values are. Often the observed CL<sub>S</sub>-value distributions are compared to the background-only expectation, as shown in Fig. 3.5. This is constructed by determining many CL<sub>S</sub>-value distributions with background-only pseudo experiments. The 2.5 %, 16 %, 50 %, 84 %, and 97.5 % quantiles are calculated at each scan point that define the expected CL<sub>S</sub>-value (median) and the  $1\sigma$  and  $2\sigma$  bands.

#### Definition of the CL<sub>S</sub> method for the search for $B^+ \rightarrow K^+ \mu^\pm e^\mp$ :

In intermediate implementations of the CL<sub>S</sub> method in the GammaCombo framework [87] a slightly different definition of the  $CL_b$  value was used with respect to the one given above. Instead of the hypothesis dependent definition  $CL_b = p_\mu(\mu' = 0)$ , the value was set to  $CL_b = p_\mu(\mu = \mu' = 0)$ , being constant over the full scan range. This definition was used in the determination of the upper limits on  $\mathcal{B}(B^+ \rightarrow K^+ \mu^\pm e^\mp)$  [88] with the CL<sub>S</sub> method using the one-sided test statistic  $q_\mu$ , Eq. (3.6). However, it can



**Figure 3.5:**  $CL_s$  value distributions for the example case using the test statistic  $\tilde{q}$ . The expected curve (red line) describes the median of background-only pseudo experiments and the shaded  $1\sigma$  ( $2\sigma$ ) bands contain 68% (95.5%) of background-only pseudo experiments. The observed line is plotted only above the best fitting value for the parameter of interest as the test statistic is defined to be nonzero only in that region.

be shown with the derivations made in Ref. [81] that for the range  $\mu > \hat{\mu}$ , where the upper limit is computed, the approximate  $CL_b$  values are the same as for the test statistic distributions discussed in this chapter.

### 3.4 Treatment of nuisance parameters

When measuring parameters of interest, these can often only be determined in a fit together with other parameters, which are called nuisance parameters. E.g. to obtain branching fractions, the simplest form of a realistic measurement is given as the fit of a measured event count  $N_{\text{obs}}$  with the formula

$$N_{\text{obs}} = N_{\text{sig}} + N_{\text{bkg}} = \frac{\mathcal{B}}{\alpha} + N_{\text{bkg}},$$

where the signal yield  $N_{\text{sig}}$  is expressed by the branching fraction  $\mathcal{B}$  and a normalisation constant  $\alpha$ . In this fit typically  $N_{\text{bkg}}$  and  $\alpha$  are floating parameters that are Gaussian constrained. As the nuisance parameters are usually correlated to the parameters of interest, they have to be considered carefully also in the determination of confidence intervals of the parameter of interest. In general, in the presence of nuisance parameters correlated with the parameter of interest, coverage can only be approximated and cannot be guaranteed anymore. This is because there is no perfect method to project the multidimensional confidence region spanned by all parameters of the fit onto the parameters of interest.

One approach often used and implemented in the GammaCombo framework [79] to project the nuisance parameter space to the parameter of interest is profiling over the likelihood. The

likelihood at a given value of the parameter of interest is evaluated by refitting all nuisance parameters. Correspondingly, when pseudo experiments are generated to determine the test statistics distribution at a given value of the parameter of interest, the nuisance parameters are set to the values obtained from the profiled fit. Often the nuisance parameters are Gaussian constrained to values determined externally. In this case also new constraint values are generated according to their Gaussian probability density function for each pseudo experiment. This approach is usually called the *Plugin* method and can be shown to have coverage approximately [89, 90]. If nuisance parameters are close to the physical boundary, a similar approach can be followed as introduced for the test statistics  $\tilde{t}_\mu$  (3.4) and  $\tilde{q}_\mu$  (3.7): if the free fit goes beyond the physical range of a parameter, it is repeated by fixing the parameter to the boundary value closest to the free fit result. However, in the general case the test statistic distribution for this approach cannot be derived analytically anymore and has to be determined from pseudo experiments.

### 3.5 Sideband subtraction vs. shape-based method

When searching for unobserved decays and setting limits on their branching fractions, two approaches can be chosen to obtain the observed signal yield. One approach is defining a signal window and calculating the signal yield  $N_{\text{sig}}$  as  $N_{\text{sig}} = N_{\text{obs}} - N_{\text{bkg}}$  by counting the number of candidates observed in the signal window  $N_{\text{obs}}$  and subtracting the expected number of background candidates in the signal window  $N_{\text{bkg}}$ . This approach is often referred to as “sideband subtraction”. A likelihood of this approach could be defined as

$$\mathcal{L}(\mathcal{B}, \alpha, N_{\text{bkg}}) = \mathcal{P}(N_{\text{obs}} | \frac{\mathcal{B}}{\alpha} + N_{\text{bkg}}) \times \mathcal{G}(\alpha_{\text{meas}} | \alpha, \sigma_\alpha) \times \mathcal{G}(N_{\text{bkg,meas}} | N_{\text{bkg}}, \sigma_{\text{bkg}}). \quad (3.13)$$

with  $\mathcal{P}$  indicating a Poisson distribution and  $\mathcal{G}$  a Gaussian distribution. The signal yield  $N_{\text{sig}}$  is expressed as  $\mathcal{B}/\alpha$  and the normalisation constant  $\alpha$  and  $N_{\text{bkg}}$  are Gaussian constrained to their measured values, the width of the Gaussian corresponding to the uncertainty of the measured value.

In an alternative approach the signal yield is determined from a fit to a discriminating variable (most often the mass distribution), where the shape of the signal is known. For distinction this method is called “shape-based method” in this thesis. The sideband subtraction has the advantages of being less dependent on the shape of the signal (it only needs the fraction of the signal distribution covered by the signal region) and yielding a more stable fit, as it does not require a full description of the discriminating variable. It is sufficient to describe the background in the control regions such that it can be interpolated into the signal region. However, the shape-based method uses the full distribution of the data (especially in the signal region), thus taking into account more information than the sideband subtraction. Therefore it should in general yield a more precise measurement and therefore a smaller upper limit.

For the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  [88] the two approaches have been compared for the determination of an expected upper limit using the test statistic  $t_\mu$  from (3.1) with the GammaCombo framework [87]. Expected upper limits at 90 % CL have been determined for the sideband subtraction to

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^+ \mu^+ e^-) &< 6.7 \times 10^{-9} \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^- e^+) &< 12.9 \times 10^{-9} \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^\pm e^\mp) &< 11.2 \times 10^{-9}\end{aligned}$$

and for the shape-based method to

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^+ \mu^+ e^-) &< 6.9 \times 10^{-9} \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^- e^+) &< 10.2 \times 10^{-9} \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^\pm e^\mp) &< 9.0 \times 10^{-9}.\end{aligned}$$

While the expected upper limit on the  $B^+ \rightarrow K^+ \mu^+ e^-$  branching fraction remains unchanged, a limit reduction is found with the shape-based method for the  $B^+ \rightarrow K^+ \mu^- e^+$  decay. The reason for this different behaviour lies in the amount of background events expected in the signal region for the two decays (see Section 5.8). While for the  $B^+ \rightarrow K^+ \mu^+ e^-$  decay about 4 events are expected, for the  $B^+ \rightarrow K^+ \mu^- e^+$  decay only one event is predicted. Therefore in the search for the  $B^+ \rightarrow K^+ \mu^- e^+$  decay the mass distribution of the background fluctuations is much more relevant leading to a better sensitivity with the shape-based method.

As the signal mass shapes of the decays depend mainly on the detector resolution and are therefore well known, all analyses described in this thesis use the shape-based method.

## 4 The LHCb experiment at the LHC

The data analysed in this thesis are collected by the Large Hadron Collider Beauty (LHCb) experiment [91], which is one of the four main experiments at the Large Hadron Collider (LHC) [92]. The LHC is a circular accelerator with a circumference of 26.7 km, about 100 m below the surface, operated by the European Organization for Nuclear Research (CERN). It accelerates protons to collide them at centre-of-mass energies of 7 TeV (in 2011), 8 TeV (in 2012) and 13 TeV (2015–2018) with the aid of a system of linear and circular pre-accelerators. Part of its run time each year is also reserved for acceleration and collision of lead-ions and other ions. The protons are organised in 2808 bunches of  $\mathcal{O}(10^{11})$  protons and are accelerated into opposite directions so that they collide at four crossing points at a rate of 40 MHz. The products of the proton-proton ( $pp$ ) collisions at the four crossing points are measured by the ATLAS [93], ALICE [94], CMS [95] and LHCb [91] experiments. The ATLAS and CMS experiments are designed as general purpose detectors to study the mechanism of electroweak symmetry breaking via the Higgs mechanism [32–34], top quark properties and to search for new elementary and mediator particles, while the ALICE experiment focuses on studying strong interaction effects at high energy densities and temperature in lead-lead and other nucleus-nucleus collisions. The LHCb experiment plays a special role in performing precision measurements of  $B$ - and  $D$ -meson properties, which are produced in copious amounts at the LHC. Therefore, while the ATLAS and CMS experiments aimed at peak instantaneous luminosities of  $\mathcal{L} = 1 \times 10^{34} \text{ cm s}^{-1}$  with a large variation throughout the data taking and up to 40 collisions per bunch crossing, the LHCb experiment chose a luminosity of  $\mathcal{L} = 4 \times 10^{32} \text{ cm s}^{-1}$  with only about 1 collision per bunch crossing in favour of more stable data taking conditions and a lower detector occupancy. Indeed, owing to the excellent performance of the trigger and reconstruction system, this is even twice the design luminosity [96].

In the years 2013–2015 the LHC was shut down for maintenance work, which was also used to optimise the data taking strategies of the experiments. Thus the data collected by the LHC experiments are conveniently divided into Run 1 (2011–2012) and Run 2 (2015–2018). Since the end of 2018 the LHC is undergoing a major upgrade with the aim to provide even higher instantaneous luminosities for all experiments, which is supposed to result in  $\mathcal{L} = 2 \times 10^{33} \text{ cm s}^{-1}$  for the LHCb experiment [97]. At the same time the LHCb and ALICE experiments are upgraded [97, 98] to cope with the higher data rates. Resumption of data taking with the upgraded accelerator and experiments is foreseen at the end of 2021 or beginning of 2022.

**Table 4.1:** Data sets used in the analyses reported in this thesis. The total  $b$ -hadron yield is calculated with the measured cross-sections [99–101] and quoted as a cumulative number.

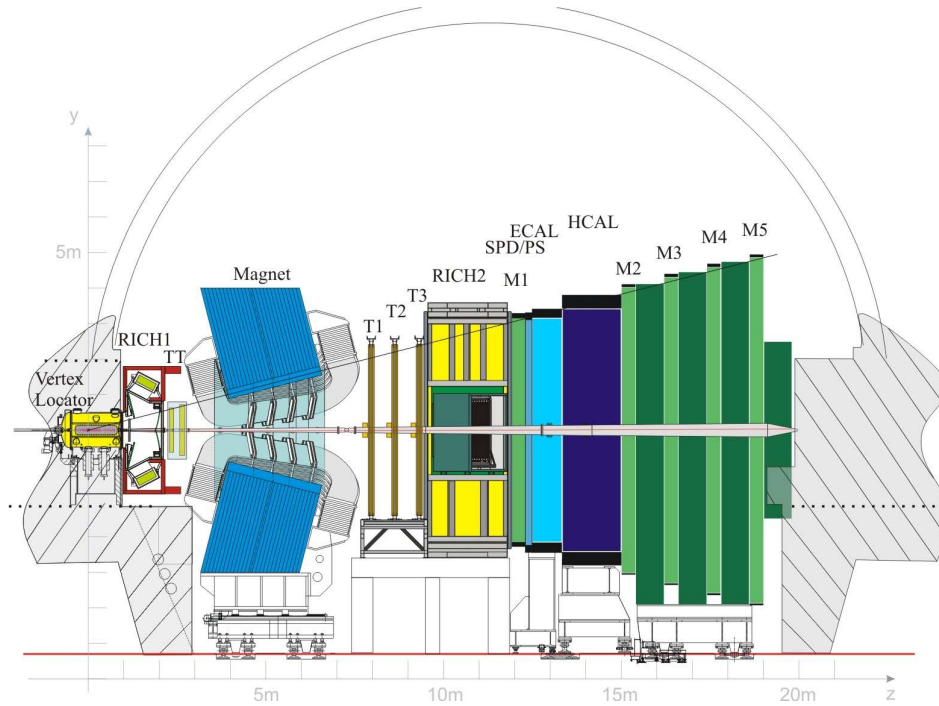
	2011	2012	2015	2016	2017	2018
$\sqrt{s}$ [TeV]	7	8	13	13	13	13
$\mathcal{L}$ [fb $^{-1}$ ]	1.11	2.08	0.33	1.67	1.71	2.19
total $b$ -hadron yield $\times 10^{12}$	0.6	1.9	2.2	3.9	5.6	7.7

The analyses performed during this thesis are based on different subsets of the data collected with the LHCb experiment. The search for the decays  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  has been performed using data from the Run 1 of the LHC, collected in the years 2011 and 2012 collected at centre-of-mass energies of  $\sqrt{s} = 7$  TeV and 8 TeV. For the search for the decays  $B_{(s)}^0 \rightarrow e^+ e^-$  also data from 2015 and 2016 of Run 2 of the LHC, collected at a centre-of-mass energy of  $\sqrt{s} = 13$  TeV are used. An overview of the data sets collected by the LHCb detector and their sizes is given in Table 4.1. The data set size used for the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  is about twice the size used for the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$ , while the full LHCb data set containing also data taken in 2017 and 2018 was used for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analysis and is again twice the size used for the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays. All results obtained in this thesis are also extrapolated to the full LHCb data set in Chapter 7.

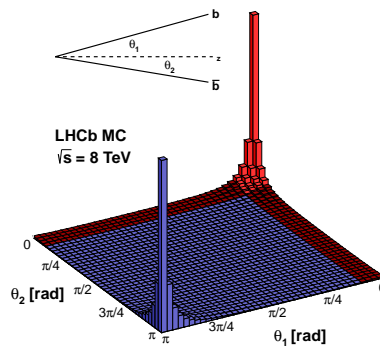
## 4.1 The LHCb detector

The LHCb detector [91, 96] is built as a single-arm forward spectrometer, as shown in Fig. 4.1 covering the pseudorapidity range of  $2 < \eta < 5$ , where  $\eta = -\ln [(\theta/2)]$  and  $\theta$  is the polar angle between a particle's momentum and the beam axis. This geometry is chosen because of the physics goal to study  $B$ -mesons: at the LHC  $b\bar{b}$  pairs in  $pp$  collisions are produced highly boosted into the forward and backward direction with respect to the beam as shown in Fig. 4.2. Thus covering the forward region in one of the two preferred directions yields a high containment of about 27% of the produced  $b\bar{b}$  pairs [102] while allowing a more precise instrumentation than the general purpose detectors. The schematic view of the LHCb detector in Fig. 4.1 shows the various subsystems, which are explained in the following subsections. They can be divided into systems designed for the reconstruction of tracks, as described in Section 4.1.1, and systems to allow distinguishing particle types, as discussed in Section 4.1.2. Furthermore some of the subdetectors can be read out particularly fast to provide information to the trigger system, which is described in Section 4.1.3.





**Figure 4.1:** Schematic view of the LHCb detector with labels of all components as used in the data taking periods Run 1 and Run 2 [91]. The  $z$ -axis is defined in the direction of the clockwise rotating beam and the  $y$ -axis as the vertical axis, the non-bending direction of the magnet. The interaction point of the  $pp$  collisions is displaced by 11.25 m from the centre of the cavern inside the Vertex Locator to make full use of the available space.



**Figure 4.2:** Simulation of the distribution of  $b\bar{b}$ -pairs for  $pp$ -collisions at a centre-of-mass energy of 8 TeV with the LHCb acceptance region displayed in red [102]. The fraction of  $b\bar{b}$ -pairs inside the LHCb acceptance is practically independent of the centre-of-mass energy in the range between 7 TeV and 14 TeV.

### 4.1.1 The Tracking system

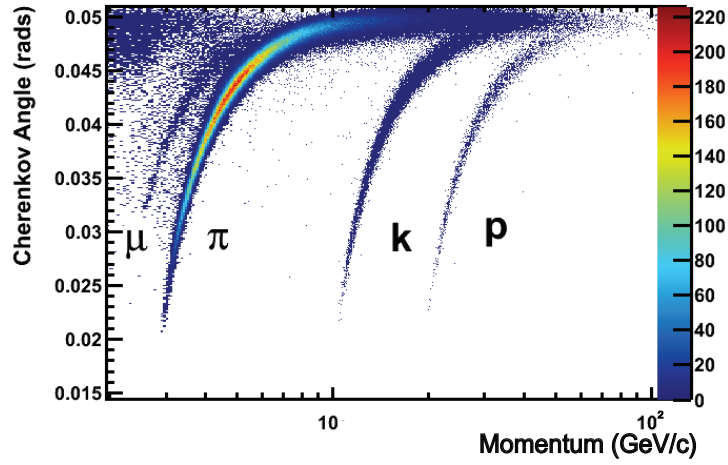
Charged particle tracks are reconstructed using the information from the Vertex Locator (VELO), the Tracker Turicensis (TT) upstream and three tracking stations (T1–T3) downstream of a warm dipole magnet with an integrated field of 4 T m. The VELO [103] is the detector closest to the interaction and is built to precisely measure primary vertices of  $pp$  collisions and secondary vertices (SV) of the  $B$ - and  $D$ -meson decays. Therefore it is made of 21 half-disc silicon modules either side of the beam axis that can be moved as close as 7 mm to the beam. Each module consists of a layer of radial and a layer of angular silicon strip sensors, providing excellent PV resolution of 13  $\mu\text{m}$  in the  $x$  and  $y$  coordinates and 71  $\mu\text{m}$  along the  $z$  axis for a PV with 25 tracks. Another important parameter for the selection of particles that decay displaced from the PV like  $B$  mesons is the impact parameter (IP), which is the distance of closest approach of a track to the PV. The IP resolution of a track depends dominantly on its transverse momentum,  $p_T$ , and is as good as 13  $\mu\text{m}$  for asymptotically high  $p_T$ . The TT is made of silicon microstrip sensors, a technology that is also employed in the inner part of the tracking stations (IT) downstream of the magnet. The outer region of the tracking stations (OT) [104, 105] is instrumented as a drift-time detector by arrays of straw-tube modules, where the straw tubes are filled with a gas mixture of Argon,  $\text{CO}_2$  and  $\text{O}_2$ . Each tracking station including the TT consists of four layers of straw tubes, organized in an  $x - u - v - x$  structure, where the  $x$  layers are vertically aligned and the  $u$  and  $v$  layers are tilted by  $\mp 5^\circ$  to allow a  $y$  resolution of the detector while facilitating the detector readout at the top and bottom of the detector.

Together with the magnet that bends the tracks of charged particles, the tracking system achieves a momentum resolution between 0.5 % and 0.8 %, leading to a mass resolution of about 0.5 %, and a track finding efficiency of about 96 % [96].

### 4.1.2 Particle identification detectors

Particle identification (PID) is pursued with the LHCb detector with three complementary parts: a system of cherenkov detectors, a calorimeter system and a muon system. While the cherenkov system is particularly important to distinguish different types of hadrons, the calorimeters assist especially in separating electrons from hadrons and the muon system provides a very efficient identification of muons.

The cherenkov system provides identification of charged particles with momenta between 2 and 100 GeV/c and is realised with two ring imaging cherenkov detectors (RICH1 and RICH2) [106]. RICH1 is located upstream of the magnet and uses the cherenkov effect in silica aerogel and  $\text{C}_4\text{F}_{10}$  to identify particles in the momentum range of 2 to 60 GeV/c, while RICH2, placed downstream of the magnet uses  $\text{CF}_4$  as a radiator and covers the higher momenta from about 15 to 100 GeV/c. The cherenkov light cones emitted from charged particles with a higher velocity than the speed of light in the radiator material have angles



**Figure 4.3:** Cherenkov angle as a function of the particle’s momentum for the  $C_4F_{10}$  radiator of RICH1 [106]. A clear separation between bands from muons, pions, kaons and protons is visible. However, at large momenta and large cherenkov angles, the bands overlap, indicating that a complementary RICH detector for higher momenta can enhance the separation power significantly.

according to  $\cos(\theta_C) = \frac{c}{nv}$  with the refraction index  $n$ . The cones are reflected out of the detector acceptance by a system of spherical and flat mirrors and are detected by hybrid photon detectors. Together with the momentum information from the tracking system, the cherenkov angle provides a clear separation between particle types, as can be seen in Fig. 4.3. The silica aerogel in RICH1 was intended to provide particle identification for kaons below the threshold of cherenkov light in  $C_4F_{10}$ . However due to the high track multiplicity in  $pp$  collisions this ability was compromised by a low efficiency of the cherenkov angle reconstruction and therefore the aerogel was removed for Run 2 of the LHCb experiment [107].

Additional particle identification as well as energy and position measurement is provided by the calorimeter system [108], which consists of a Scintillating Pad Detector (SPD), a Preshower (PS), an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL) with alternating layers of active and absorbing material. The main purpose of the calorimeter system is to discriminate photons from electrons and both from hadrons and to provide fast transverse energy  $E_T$  measurements of high energetic electrons, photons, and hadrons to the trigger system. The detection approach in the calorimeters is similar for all parts by measuring the scintillation light of charged particle showers in the active layers of scintillator material and producing electromagnetic and hadronic showers in the absorber layers of lead and iron. Electron showers are distinguished from high energy  $\pi^0$  particles by placing the SPD and PS planes of rectangular scintillating pads at the entrance of the ECAL, separated by a layer of lead. The ECAL consists of alternating layers of lead and scintillator material with the aim of covering the full lengths of electromagnetic showers from high energetic

photons and therefore has a total length of 25 radiation lengths. The HCAL however is made of alternating layers of iron as absorption material and scintillator tiles as active material with a total thickness of 5.6 nuclear interaction lengths. It is thinner than the ECAL due to space constraints and because a performant trigger decision on hadrons can be reached already with only partly contained hadronic showers. Due to a strong track density variation over the surface of the calorimeters of about two orders of magnitude, the calorimeter system is laterally segmented with three (two for the HCAL) regions of different cell sizes.

The detector is completed by a system of five rectangular muon stations M1–M5 [109], which are instrumented by multi-wire proportional chambers and interleaved by iron absorbers leading to a thickness of 20 interaction lengths. The first station is placed before the calorimeter system and uses the gas electron multiplier technology in the inner region due to its better performance in radiation environment. While other particles are stopped by the absorption layers in the calorimeters and the muon system, muons with a minimal momentum of 6 GeV/c traverse the five muon stations. Thus the muon system provides excellent separation of muons from other particles. With the lateral segmentation it also provides fast  $p_T$  measurements with an accuracy of about 20 %, sufficient to allow a performant trigger on high energy muons.

The information of the cherenkov detectors, the calorimeter system, and the muon system is combined to provide particle identification variables in several ways [96]. On the one hand analytical variables are created as the sum of log likelihood differences between the hypothesis of a candidate to be of particle type  $X$  and the hypothesis to be a pion for each subsystem:

$$\text{PID}_X = \sum_{\text{system}} \Delta \ln \mathcal{L}^{\text{system}}(X - \pi). \quad (4.1)$$

On the other hand the PID information of all subsystems is combined into a single probability for a particle hypothesis using neural nets [110], which are often more powerful than the likelihood-based approach. These are called  $\text{ProbNN}_X$  throughout this thesis.

### 4.1.3 The trigger system of the LHCb experiment

Not all subdetectors can be read out at the bunch-crossing rate of 40 MHz. Also saving each event on tape would quickly exceed the available storage resources. Since not each bunch-crossing event contains a  $pp$  collision with data relevant for the physics program of LHCb, a trigger system is employed that reduces the data rate to 2–5 kHz in Run 1 [96, 111] ( $\approx 12$  kHz in Run 2 [112]), which can be saved for physics analysis. The focus for triggering events is laid on processes involving  $B$  and  $D$  decays. The trigger system consists of a low level hardware trigger (Level-0, L0) and a subsequent software based trigger (High-Level Trigger, HLT). In the L0 trigger events involving electrons, hadrons, muons, and photons with high  $p_T$  and  $E_T$  are saved at a rate of about 1 MHz using fast information from the

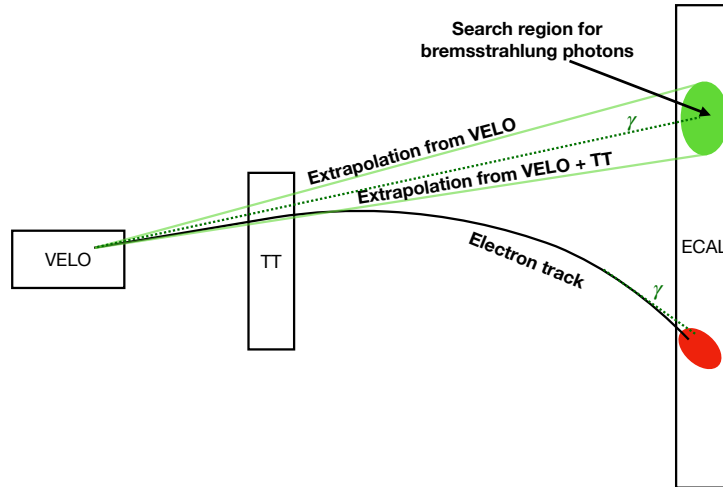
calorimeters and the muon system. A first preliminary reconstruction of tracks is applied only afterwards. The HLT trigger is further divided into the HLT1 stage, which reduces the event rate on the basis of preliminary reconstructed tracks and the HLT2 stage, which selects events using fully reconstructed events. At the HLT1 stage, events are saved with tracks that have a significant  $p_T$  and are displaced from the PV as indicated by the IP, since these properties distinguish particles from heavy flavour hadrons like  $B$  and  $D$  mesons from other particles. The subsequent HLT2 trigger stage selects events based on their topology by searching e.g. for track combinations that form a vertex displaced from the PV. The events that pass the trigger requirements are reprocessed offline using full alignment information.

At the L0 level, the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analyses discussed in this thesis rely dominantly on the single muon and dimuon triggers that require a clear particle signature corresponding to a high  $p_T$  in the muon system. The  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis instead relies on electron and hadron triggers, which require a significant energy deposit in the ECAL and HCAL. At the HLT1 and HLT2 stage the decays are dominantly selected by triggers that identify good quality tracks displaced from the PV with high  $p_T$  or triggers that require good quality two- and three-track vertices. The trigger decisions are evaluated with multivariate methods. The exact requirements of the analyses are discussed in Section 5.2.2 and Appendix A.

## 4.2 Bremsstrahlung reconstruction at LHCb

The  $B_{(s)}^0 \rightarrow e^+ e^-$  and  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analyses performed in this thesis involve the use of electrons as final state particles. Electrons behave different from other charged particles in the detector mainly due to their comparably low mass. They lose part of their energy due to bremsstrahlung, dominantly in the detector material. On the one hand this leads to a worse track resolution and track reconstruction efficiency compared to other particles. On the other hand, if the bremsstrahlung emission occurs before the electron traverses the magnet, only the momentum and energy of the electron after bremsstrahlung emission is measured due to the bending of the electron direction in the magnet, thus biasing the kinematic measurements. If however the electron emits bremsstrahlung after traversing the magnet, the photon direction will be collinear to the electron direction and thus its energy deposit in the calorimeter will be in the same cluster as the electron. Therefore these kinds of bremsstrahlung will not bias the momentum and energy measurement.

Bremsstrahlung photons result in electromagnetic shower clusters in the calorimeter system. If the measured energy deposit exceeds 75 MeV [113], those clusters can be reconstructed, where the threshold is chosen as a balance of signal efficiency over the suppression of fake clusters from random combinations of neighbouring clusters. This can be used to recover the energy lost from bremsstrahlung emitted before traversing the magnet. The reconstructed photon clusters can be reassociated to the electron tracks by searching photon clusters in an



**Figure 4.4:** Schematic view of the bremsstrahlung correction process from the top of the detector adapted from Ref. [113]. The dashed lines indicate bremsstrahlung photons emitted before and after traversing the magnet. The search window for bremsstrahlung photons consistent with being emitted upstream of the magnet is defined by the two tracks extrapolated to the ECAL from using VELO information only and from using VELO and TT information. Photons emitted after the bending of the electron track in the magnet are measured in the same calorimeter cluster as the electron.

area in the calorimeter that is constructed from the track that is extrapolated to the ECAL from VELO information only and the track that is extrapolated from VELO and TT information. This is schematically shown in Fig. 4.4. The  $2\sigma$  interval of the combined uncertainty of the extrapolation variation arising from the two extrapolations and the uncertainty from the position measurement of the photon cluster is defined as the search region for a bremsstrahlung cluster. Adding the energy from the photon cluster and assuming its origin in the PV, the photon momentum can be reconstructed and used to correct the electron 4-momentum. With this procedure about half of the electrons can receive a bremsstrahlung correction, significantly improving the momentum resolution and consequently the invariant masses of combinations involving electrons. At the same time being able to associate a bremsstrahlung photon cluster to a track supports the discrimination of electrons from other particles, as only electrons emit a substantial amount of bremsstrahlung in the detector. Because of these significant differences between the case of having bremsstrahlung correction applied or not, the data in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis are evaluated in categories of bremsstrahlung categories: either no electron has bremsstrahlung correction applied (category 0), one of the electrons has bremsstrahlung correction applied (category 1) or both electrons have the correction applied (category 2).

## 4.3 Simulation at LHCb

The simulation of events in the LHCb detector is realised via a set of Monte-Carlo generators, governed by the GAUSS framework [114]. Full  $pp$  collisions are simulated by the PYTHIA tool [115] with an LHCb-specific configuration. For decays involving  $B_c^+$  mesons that constitute to a relevant background to the  $B_{(s)}^0 \rightarrow \ell^+ \ell^-$  analyses performed in this thesis, the BCVEGPy generator [116, 117] is used to generate the  $gg \rightarrow B_c^+ + X$  process. With the EVTGEN package, decays of unstable particles are implemented [118], where final state radiation is generated using PHOTOS [119]. The interaction of generated particles with the detector is modelled in the GEANT4 toolkit [120, 121]. The detector response of the interactions is then simulated using the BOOLE framework [122] and then reconstructed in the same way as recorded data.

In general, the simulation set up in the described way describes the recorded data very well. However, due to the limited computability of QCD interactions, the momentum spectra of  $B$  mesons and the underlying event separate from the studied signal decay are often not perfectly described. Also the PID variables are often dependent on the data taking conditions, which cannot perfectly be described in simulation. Therefore the simulation is adjusted to data for each analysis, as discussed in Sections 5.3 and 6.2.





## 5 $B^+ \rightarrow K^+ \mu^\pm e^\mp$ and $B_{(s)}^0 \rightarrow e^+ e^-$ – typical searches for rare decays at LHCb

This chapter reviews the searches for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  decays, which lead to the publications [88, 123]<sup>1</sup>. Since the general steps taken for both analyses are very similar and typical examples for rare decays searches at LHCb, they are described together. The analyses make use of the outcome of bachelor’s and master’s theses [124–128]. The search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  has been carried out in a group of several analysts and is covered partially in another PhD thesis [129]. Therefore the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  part is kept brief in this chapter. My main contribution to the analysis was the statistical interpretation of the result with the methods described in Chapter 3. The search for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays was performed jointly with one other PhD thesis [130], where the main focus of my work was on the normalisation, the description of the background and the extraction of the final result. For simplicity, the plots used to illustrate the different analysis steps discussed in this chapter are taken from the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis.

### 5.1 Analysis strategies

Searches for rare decays like the searches for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  typically follow similar schemes. The data are selected with the aim to reject as much background as possible while maintaining a high signal efficiency. The selection and its efficiency are determined using simulated signal decays and data control regions. In searches for rare  $B$  decays the reconstructed invariant  $B$  mass is often used as main discriminating variable, in which the signal, if existing, should show up as a peaking distribution at the nominal  $B$ -meson mass [21]. To avoid experimenter’s bias, the signal region in the invariant  $B$  mass containing most of the potential signal candidates is removed from data while optimising the selection and only added back once the selection is fixed. For the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  this region contains close to 100 % of simulated  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays, while for  $B_{(s)}^0 \rightarrow e^+ e^-$  the fraction is reduced to contain only 90 % of simulated  $B_{(s)}^0 \rightarrow e^+ e^-$  decays due to their much worse mass resolution. The regions outside the exclusion window serve as control regions that contain only background and can therefore be used to tune the selection. In the upper mass sideband

---

<sup>1</sup>In publications of the LHCb collaboration, all  $\approx 1000$  authors are listed alphabetically. Typically 2–3 authors are highlighted as “contact authors”. I am one of the contact authors for both publications described in this chapter.

**Table 5.1:** Exclusion and side band regions used in the analyses discussed in this chapter.

The exclusion region for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays is much broader due to the presence of two electrons, which have a worse momentum resolution than other charged final state particles.

	$m(K^+ e^\pm \mu^\mp)$ [MeV/c <sup>2</sup> ]	$m(e^+ e^-)$ [MeV/c <sup>2</sup> ]
Exclusion region	[4985, 5385]	[4689, 5588]
upper side band	[5385, 6000]	[5588, 6566]
lower side band	[4550, 4985]	[4166, 4689]

of the exclusion region contributions from random track combinations, “combinatorial background”, dominate, while in the lower mass sideband sizeable contributions from partially reconstructed decays and decays involving misidentified particles are present. Thus it is important to have access to both sidebands to control the background description, which drives the choice of the exclusion window for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis. The exclusion and side band regions used in the analyses described in this chapter are quoted in Table 5.1. To largely cancel systematic effects in reconstruction and selection, the branching fractions are measured relative to those of a decay with a large and precisely known branching fraction, which has a similar topology and response in the detector. Thus, the branching fractions can be extracted as

$$\begin{aligned} \mathcal{B}(\text{signal}) &= \frac{N(\text{sig})}{N(\text{norm})} \frac{\varepsilon(\text{norm})}{\varepsilon(\text{sig})} \times \left( \frac{f_{(u,d,s)}}{f_u} \right)^{-1} \times \mathcal{B}(\text{norm}) \\ &\equiv N(\text{sig}) \times \alpha = N(\text{sig}) \times \alpha' \times \left( \frac{f_{(u,d,s)}}{f_u} \right)^{-1} \times \mathcal{B}(\text{norm}), \end{aligned} \quad (5.1)$$

with the observed signal and normalisation yields  $N(\text{sig})$  and  $N(\text{norm})$  and the efficiencies  $\varepsilon(\text{sig})$  and  $\varepsilon(\text{norm})$ , respectively. The normalisation branching fraction  $\mathcal{B}(\text{norm})$  and the fragmentation fractions  $f_{(u,d,s)}$  are external inputs to the analyses. Therefore it is useful to determine additional to the normalisation factor  $\alpha$ , which describes the single event sensitivity of the search, a partial normalisation factor  $\alpha'$ , that excludes all external inputs to the measurement.

The analyses discussed in this chapter use the normalisation mode  $B^+ \rightarrow K^+ J/\psi$  due to its abundance and precisely known branching fraction of  $\mathcal{B}(B^+ \rightarrow K^+ J/\psi) = (1.010 \pm 0.028) \times 10^{-3}$ ,  $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033) \%$  and  $\mathcal{B}(J/\psi \rightarrow e^+ e^-) = (5.971 \pm 0.032) \%$ . The topology matches that of  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and the usage of both  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  and  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  allows to control the different behaviour of electrons and muons in the detector. The decay  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  is suitable as a normalisation mode for  $B_{(s)}^0 \rightarrow e^+ e^-$  because it

contains electrons in a similar kinematic range as the signal mode. The normalisation mode is also used as a control channel to determine and correct for data-simulation differences.

As the data taking conditions differ significantly between Run 1 and Run 2 due to different centre-of-mass energies and refined trigger algorithms, the analyses are performed split by Run. Furthermore in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis the data are evaluated in categories of electrons having bremsstrahlung correction applied, as this correction procedure strongly affects electron kinematics and the invariant dielectron mass resolution.

The selection of signal candidates is performed in several steps: first a loose preselection is applied consisting of an experiment-wide selection when reconstructing candidates and further fiducial requirements and cuts to clean up the data set by removing unphysical candidates. These requirements are described in Section 5.2. After applying the preselection the control channel resonances are clearly visible in the data. Using the  $sPlot$  method [131], the control modes are cleaned of background components in data and thus allow a comparison between simulation and data and consequently a correction of imperfect distributions in simulation. The methods and results are explained in Section 5.3. In a final selection step multivariate classifiers are trained to remove predominantly combinatorial background. In the search for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  also an additional multivariate classifier is used to suppress partially reconstructed decays. Furthermore stringent particle identification criteria are applied in both analyses to suppress decays containing misidentified particles. The approaches are reviewed in Section 5.4. A summary of the efficiencies of the various selection steps applied in the searches for the decays  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+e^-$  is discussed in Section 5.5 and combined with the yield in the normalisation mode to obtain the normalisation factors. The remaining background contributions and their distributions are described in Section 5.6. Systematic uncertainties evaluated for the efficiency and mass distribution estimation are reviewed in Section 5.7 and the final results of the searches for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+e^-$  decays are discussed in Section 5.8.

## 5.2 Preselection

A loose preselection is applied in order to remove unphysical candidates with simple requirements while maintaining a high signal efficiency. This is achieved by applying cuts in the experiment-wide preselection stage, where also signal candidates are formed, as discussed in Section 5.2.1. A subset of trigger decisions is chosen that is well described in simulation as discussed in Section 5.2.2. Furthermore, obvious backgrounds are removed and fiducial requirements are placed to align the selection to calibration data as detailed in Section 5.2.3.

## 5.2.1 Experiment-wide preselection

When combining signal decay candidates, an experiment-wide preselection is applied to reduce the data set to a manageable size and reduce combinatorial background while keeping the signal efficiency as high as possible. These selections are organised in sets of selections, optimised for different types of analyses. The requirements of the different sets used for the searches for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  are listed in Table 5.2. The selections comprise requirements on track and vertex fit qualities, as well as minimal momenta on the final state particles and the displacement of the decay vertex from the primary vertex. Thus they make use of the fact that the  $B$ -meson has a significant lifetime and therefore flies a few centimetres before decaying and is produced with sizeable momenta transverse to the beam axis ( $p_T$ ) due to its high mass.

In detail the requirements on the hits in the SPD detector ( $n_{\text{SPD}}$ ) remove events with high detector occupancy as they tend to have a worse event reconstruction. Requirements on the reconstructed meson mass  $m$  and its difference with respect to the nominal mass [21]  $|m - m_{\text{PDG}}|$  allow to remove partially reconstructed backgrounds. The vertex quality ( $\chi_{\text{vtx}}^2/\text{dof}$ ) and distance of closest approach between tracks (DOCA) is necessary to reduce random track combinations. These can also efficiently be reduced by constraining the angle between the vector from the primary vertex (PV) to the secondary  $B$ -decay vertex (SV) and the  $B$ -momentum vector (DIRA). The flight distance significance  $\chi_{\text{FD}}^2$  ensures that the SV is displaced from the PV. A similar effect is achieved by requiring low  $\chi_{\text{IP}}^2$  values on the  $B$  candidates and high  $\chi_{\text{IP}}^2$  values on the decay products. This variable is defined as the difference in the  $\chi_{\text{vtx}}^2$  of the PV fit with and without taking into account the considered candidate. Thus the requirements constrain the  $B$  candidate to originate from the PV, while the decay products originate from a displaced SV. Particle identification criteria ( $\text{PID}_{e,K}$ ) on electrons and kaons are used to discriminate them against pions, which are produced abundantly in a  $pp$ -collision. For muons it is sufficient to require a clear signature in the muon detectors with the ISMUON criterion. The description of particle identification algorithms is not optimal in simulation, therefore the particle identification criteria are removed from the selection in simulation, calibrated on data as discussed in Section 5.3 and then applied to the samples.

The requirements do not differ much between the investigated decay modes, as the kinematic regions of the final states are similar. However, they vary a bit between the selection sets used for the analyses, as a similar effect can be achieved by either applying stringent vertex requirements at loose kinematic cuts or vice versa.

The simulated samples contain not only the signal decays but also the remainder of the event and are reconstructed in a similar way as recorded data. This leads to the effect that in simulation random track combinations from the remainder of the event can occur and particles can be swapped. Furthermore ghost tracks can be produced from hits in the detector that originate from different tracks. Therefore algorithms are applied to match reconstructed candidates to the true simulated decays and only correctly reconstructed decays (taking into

**Table 5.2:** Requirements on signal and normalisation candidates in the experiment-wide preselection.

Particle or event	Variable	$B^+ \rightarrow K^+ \mu^\pm e^\mp$ ( $B^+ \rightarrow K^+ J/\psi$ )	$B_{(s)}^0 \rightarrow e^+ e^-$ ( $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ )
Event	$n_{\text{SPD}}$	$< 600$	–
$B$	$ m - m_{\text{PDG}} $	$< 1000 \text{ MeV}/c^2$	$< 1200 \text{ MeV}/c^2$ ( $< 600 \text{ MeV}/c^2$ )
	$\chi_{\text{vtx}}^2/\text{dof}$	$< 9$	$< 9 (< 45)$
	$\chi_{\text{FD}}^2$ wrt. PV	$> 100$	$> 225$
	$\chi_{\text{IP}}^2$ wrt. PV	$< 25$	$< 25$
	DIRA wrt. PV	$> 0.9995$	$> 0$
	DOCA	–	0.3 mm
$K$	$p_T$	$> 400 \text{ MeV}/c$	$> 250 \text{ MeV}/c$
	$\chi_{\text{IP}}^2$	$> 9$	$> 25$
	$\text{PID}_K$	$> -5$	–
$e$	$p_T$	$> 300 \text{ MeV}/c$	$> 250 \text{ MeV}/c$
	$\chi_{\text{IP}}^2$ wrt. PV	$> 9$	$> 25$
	$\text{PID}_e$	$> 0$	$> -2(2)$
$\mu$	$p_T$	$> 300 \text{ MeV}$	–
	$\chi_{\text{IP}}^2$ wrt. PV	$> 9$	–
	ISMUON	True	–
$e\mu$ pair	$m(e\mu)$	$> 100 \text{ MeV}$	–
	$\chi_{\text{vtx}}^2(e\mu)/\text{dof}$	$< 9$	–
$J/\psi \rightarrow \mu^+ \mu^-$	$p_T$	$> 0 \text{ MeV}$	–
	$m$	$< 5500 \text{ MeV}$	–
	$\chi_{\text{vtx}}^2/\text{dof}$	$< 9$	–
	$\chi_{\text{FD}}^2$ wrt. PV	$> 256$	–
	$\chi_{\text{IP}}^2$ wrt. PV	$> 0$	–
$J/\psi \rightarrow e^+ e^-$	$p_T$	$> 0 \text{ MeV}$	–
	$m$	$< 5500 \text{ MeV}/c^2$	$\in [2096, 4096] \text{ MeV}/c^2$
	$\chi_{\text{vtx}}^2/\text{dof}$	$< 9$	$< 9$
	$\chi_{\text{FD}}^2$ wrt. PV	$> 256$	$> 169$
	$\chi_{\text{IP}}^2$ wrt. PV	$> 0$	–
	DIRA wrt. PV	–	$> 0$
	DOCA	–	0.3 mm

account that they might have lost a significant amount of unrecovered bremsstrahlung) are retained.

## 5.2.2 Trigger selection

Only a subset of the available trigger decisions is chosen for the analyses. This facilitates the calculation of trigger efficiencies, as some trigger decisions that are made specifically on the remainder of the collision are not well modelled in simulation, thus potentially biasing the efficiency calculation. Those are therefore avoided. Furthermore some trigger decisions allow an event rate that is too high to be stored. Therefore only a fraction of those events is stored, where the fraction is defined by a prescaling factor. As the prescaling factor is not applied in simulation and these trigger decisions have a small impact on the overall trigger efficiency, also prescaled trigger decisions are mostly avoided. Only for the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis two prescaled HLT2 lines, the SingleMuon and SingleMuonLowPT are accepted. The selected trigger algorithms are listed in Table 5.3. The L0 algorithms trigger dominantly on high energy signatures in the muon chambers and the electromagnetic calorimeter, while the HLT 1 algorithms use information on tracks in the event and dominantly trigger on tracks that do not directly come from the PV, but instead a displaced vertex. The HLT 2 algorithms make use of the fully reconstructed event and trigger mostly on the decay topology. The trigger algorithms and their selections have been refined significantly between Run 1, 2015 and 2016-2018 to improve the performance [96, 112]. Therefore the selections are optimised separately for these sets of years. Trigger decisions are made typically either due to the signal decay (triggered on signal, TOS) or due to other decays in the event (triggered independent of signal, TIS). Because the underlying event is not as well described in simulation as the signal decay, the general strategy for the trigger selection is to choose as many TOS decisions as possible, removing the prescaled ones. However, for the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  the use of the L0Global TIS decision, a logical OR of all L0 TIS trigger decisions, adds a large fraction (about 20%) of signal candidates. Therefore it is considered feasible to be also taken into account. The efficiency of the selection is determined from data with a method described in Section 5.5.1 to avoid the above mentioned biases from data-simulation differences.

## 5.2.3 Fiducial requirements and background-specific vetoes

In order to align the selection with calibration data, the kinematic distributions of the final state particles and detector occupancy related variables are reduced to fiducial regions, thus simplifying the particle identification calibration procedure. These requirements are listed in Table 5.4. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis no requirements on the absolute momenta were applied as those were not used for the calibration of the simulation.

**Table 5.3:** Selected trigger algorithms. The algorithms and their selections vary significantly between Run 1, 2015 and 2016–2018 data taking periods. For the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  the trigger algorithms only used for selecting  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  are noted with a \* and the ones used only for  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  are noted with \*\*. If not otherwise specified, the algorithms are selected if they are TOS. In the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  the L0 selections are only applied to the relevant leptons.

$B^+ \rightarrow K^+ \mu^\pm e^\mp$	Run 1		
L0	( $\mu$ ) L0Muon*	(e) L0Electron**	
HLT 1	TrackAllL0	TrackMuon*	
HLT 2	Topo[2,3]BodyBBDT	TopoMu[2,3]BodyBBDT*	
	SingleMuon*	SingleMuonLowPT*	
$B_{(s)}^0 \rightarrow e^+ e^-$	Run 1	2015	2016
L0	L0Electron L0Global(TIS)	L0Electron L0Global(TIS)	L0Electron L0Global(TIS)
HLT 1	TrackAllL0 TrackPhoton	TrackMVA	TrackMVA
HLT 2	Topo2BodyBBDT TopoE2BodyBBDT	Topo2Body	TopoE2Body TopoEE2Body

**Table 5.4:** Fiducial requirements on signal and normalisation candidates to align with calibration data. The  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis did not restrict the absolute particle momenta  $p$ , as those were not used for calibration.

Particle or event	Variable	Run 1	Run 2
Event	$n_{\text{SPD}}$	$< 600$	$< 450$
$K$	$p_T$	$> 250 \text{ MeV}/c$	$> 600 \text{ MeV}/c$
	$p$	$> 2000 \text{ MeV}/c$	$> 2000 \text{ MeV}/c$
$e$	$p_T$	$> 500 \text{ MeV}/c$	$> 500 \text{ MeV}/c$
	$p$	$> 3000 \text{ MeV}/c$	$> 3000 \text{ MeV}/c$
$\mu$	$p_T$	$> 800 \text{ MeV}$	$> 800 \text{ MeV}$

The normalisation candidates are further selected with the requirements given in Table 5.5, providing background-reduced normalisation samples to allow a stable fit of the mass distribution. This fit will then be used to project out the pure signal distribution of the normalisation samples with the  $\mathcal{P}lot$  method [131], thus providing a clean comparison between data and simulation. The selection is made on the  $B^+$  and the  $J/\psi$  mass and analytical ( $\text{PID}_e$ ) and neural net based particle identification variables  $\text{ProbNN}_X$ . A requirement on the  $B^+$  mass  $m_{K+J/\psi}$  using a decay tree fit yields a strong suppression of partially reconstructed background decays like  $B^0 \rightarrow K^{*0} J/\psi$ . In this fit the  $B^+$  mass and momenta are refitted from the final state particle tracks, constraining the  $J/\psi$  to its nominal mass [21].

Some further cuts are applied to the signal and normalisation samples in preparation of the main selection via multivariate classifiers with slightly different intentions: for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the cuts aim at removing nonphysical candidates and ghost candidates from combining hits that belong not to the same track. Therefore loose requirements are applied on the momenta of electron and  $B$ -meson candidates and on the ghost probability of the electron tracks. Also candidates with unphysically high lifetimes are removed. The detailed requirements are listed in Table 5.6.

In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis a lot of background decays are present of which final state particles are misidentified or only a subset of final state particles of the background decay are selected, making the combination similar to the signal decay. One class of these dominant backgrounds are decays involving charmonium resonances, such as  $J/\psi$  and  $\psi(2S)$ , which have a high branching fraction and decay with a high rate into  $\mu^+ \mu^-$  or  $e^+ e^-$  (e.g.  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  or  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ ). These decays can fake the signal decay when at least one of the final state particles is misidentified. They can efficiently be suppressed by vetoing a mass region around the charmonium masses as listed in Table 5.7. The differences in the particle masses are taken into account by assigning a different mass hy-



**Table 5.5:** Additional requirements on the normalisation candidates to allow a clean fit of the normalisation distributions. The values in brackets denote the slightly different selection of  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis.

Particle	Variable	$B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$	$B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$
$B^+$	$m_{K^+ \ell^+ \ell^-}$ [MeV/ $c^2$ ]	[5180, 5700]	[4880(4800), 5700]
	$m_{K^+ J/\psi}$ [MeV/ $c^2$ ]		> 5175(5150)
$K$	ProbNN $_K$	> 0.2	> 0.2
$e$	PID $_e$	–	– (> 3)
$\mu$	ProbNN $_\mu$	> 0.2	–
$J/\psi$	$m_{e^+ e^-}$ [MeV/ $c^2$ ]	[3037, 3157]	[2450, 3176] ([2600, 3300])

**Table 5.6:** Summary of offline selection cuts applied to the  $B$  meson and the electrons prior to the training of the multivariate classifier for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis.

Particle	cut
$e^\pm$	prob $_{\text{ghost}} < 0.3$
	$p_T < 40$ GeV/c
	$p < 500$ GeV/c
$B$	$p_T > 1$ GeV/c
	$\tau(B) < 9 \times \tau_{\text{true}}(B_s^0)$

pothesis for the assumed misidentification. Another class of dominant background processes are decay chains involving semileptonic  $D^0$ -meson decays like  $B^- \rightarrow D^0 (\rightarrow K^- \mu^+ \nu_\mu) e^- \bar{\nu}_e$ , which again have a large rate and can fake the signal, as the neutrinos are not reconstructed. These backgrounds can be effectively reduced by requiring the kaon-lepton mass combinations to be larger than the nominal  $D^0$ -meson mass  $m_{K^\pm \ell^\mp} > 1885 \text{ MeV}/c^2$ . In addition to these mass vetoes, moderate requirements on particle identification variables as listed in Table 5.8 help further reducing backgrounds from decays with misidentified final states. This is applied before the main selection with a multivariate analysis, as topologically very signal-like decays with misidentified final states can appear in the background sample for the training of the multivariate classifier and thus deteriorate its separation power.

**Table 5.7:** List of preselection vetoes against charmonium resonances  $J/\psi$  and  $\psi(2S)$ . All these vetoes are applied on the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  signal channel (charge conjugation is implied). These vetoes are not applied to the normalisation and control modes  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  and  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ .

mass hypothesis	mass region vetoed ( MeV/ $c^2$ )
$K$ with $\mu$ mass	$3000 < m_{K^- \mu^+} < 3200$
	$3630 < m_{K^- \mu^+} < 3740$
$e$ with $\mu$ mass	$2950 < m_{e^- \mu^+} < 3200$
	$3630 < m_{e^- \mu^+} < 3740$
$K$ with $e$ mass	$3000 < m_{K^+ e^-} < 3200$
	$3630 < m_{K^+ e^-} < 3740$
$\mu$ with $e$ mass	$3000 < m_{\mu^+ e^-} < 3200$
	$3630 < m_{\mu^+ e^-} < 3740$

**Table 5.8:** Particle identification criteria applied to the signal and control channels in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis.

Particle	Variable	Cut
$\mu$	$\text{ProbNN}_\mu$	$> 0.2$
$K$	$\text{ProbNN}_K$	$> 0.2$
$e$	$\text{PID}_e$	$> 3.0$

## 5.3 Data–simulation corrections

In general, the simulation used in LHCb analyses is capable to describe the data very well. However, for some variables simulation cannot describe the data distributions perfectly due to imprecise knowledge of all QCD-interactions occurring in a  $pp$ -collision, unclear phase space distributions in decays involving multiple final states and due to computing limitations, which prevent from simulating  $pp$ -collisions in full detail. To study the differences between data and simulation, a clean data sample is necessary that does not contain any background. This is achieved with the  $sPlot$  technique briefly discussed in Section 5.3.1. Afterwards the simulated samples are compared and reweighted to better match the data, as described in Section 5.3.2. Particle identification variables are calibrated separately, as outlined in Section 5.3.3. The calibrations are done at preselection level before strong MVA and PID selections are applied to have the efficiencies as correct as possible. Trigger efficiencies are also determined from data using the TISTOS method which is not discussed in this section, but summarised in Section 5.5.1.

### 5.3.1 The $sPlot$ technique

The  $sPlot$  technique [131] allows to project out single contributions from a data set by the means of discriminating variables, where the contributions can be disentangled. In the examples discussed in this thesis, the reconstructed  $B$ -meson invariant mass  $m(K\ell^+\ell^-)$  of the control modes  $B^+ \rightarrow K^+J/\psi$  is used to describe the signal and background distributions with a fit. This fit is then used to create orthogonal weights, “ $sWeights$ ”, that project out the signal distribution by subtracting all background contributions in control variables that are independent from the discriminating variable  $m(K\ell^+\ell^-)$ . While this procedure is solid for  $B^+ \rightarrow K^+J/\psi (\rightarrow \mu^+\mu^-)$  decays, in  $B^+ \rightarrow K^+J/\psi (\rightarrow e^+e^-)$  decays  $m(Ke^+e^-)$  is correlated to the electron kinematics due to the presence of bremsstrahlung and could thus introduce a bias. However, due to the very low background present in the normalisation samples (see Fig. 5.1), the bias is considered negligible. But in the determination of PID efficiencies for the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis, where the same procedure is used on calibration data, instead of applying the  $sPlot$  method for electron PID, the PID efficiencies are determined by fitting the calibration data in bins of kinematic variables.

The fit function used to describe the mass distribution is a single exponential for the background from random track combinations and a combination of Crystal Ball functions [132] describing the signal distribution. Other contributions are found to be negligible. In the

$B_{(s)}^0 \rightarrow e^+ e^-$  analysis the data are described separately in each bremsstrahlung category with a double sided Crystal Ball (CB) function as probability density function (PDF), defined as

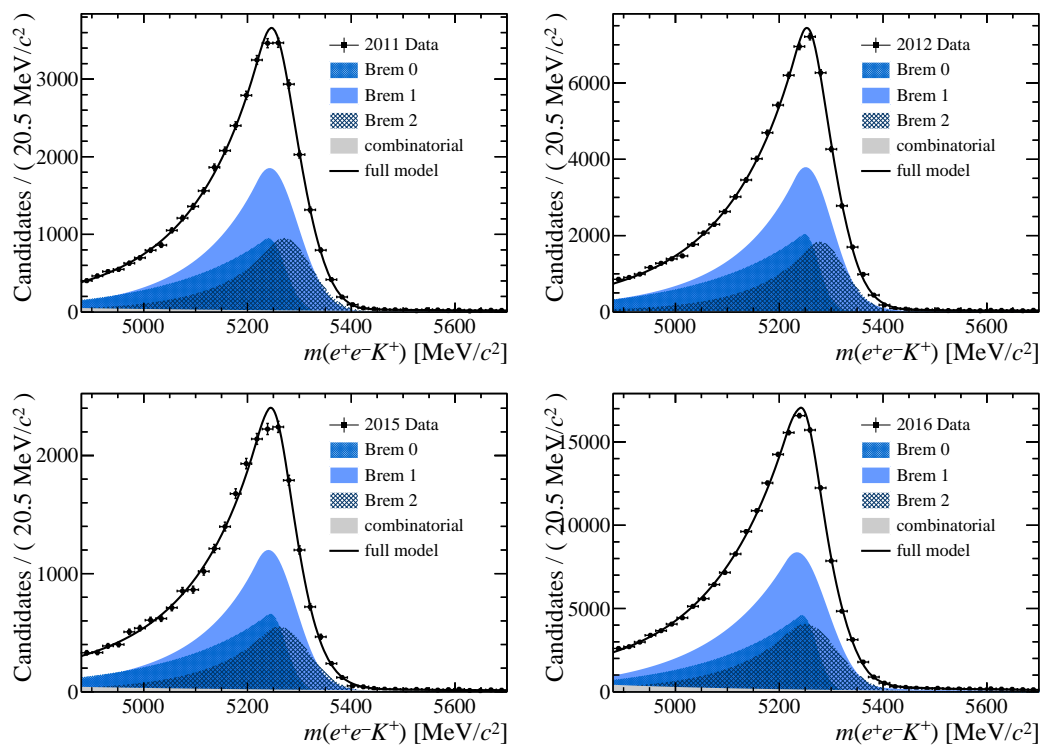
$$\begin{aligned}
 PDF(m) = & \left\{ \frac{N_{\text{sig}}}{N_{\text{sig}} + N_{\text{bkg}}} \times (f_0 \times [CB_0(m; \alpha_0, \alpha'_0, n_0, n'_0, \mu_0, \sigma_0)] \right. \\
 & + f_1 \times [CB_1(m; \alpha_1, \alpha'_1, n_1, n'_1, \mu_1, \sigma_1)] \\
 & + f_2 \times [CB_2(m; \alpha_2, \alpha'_2, n_2, n'_2, \mu_2, \sigma_2)]) \\
 & \left. + \frac{N_{\text{bkg}}}{N_{\text{sig}} + N_{\text{bkg}}} \times (N_{\text{exp}} e^{\lambda m}) \right\} \\
 & \times \mathcal{P}(N_{\text{sig}} + N_{\text{bkg}} | N_{\text{total}}),
 \end{aligned}$$

where  $N_{\text{sig}}$  and  $N_{\text{bkg}}$  are the signal and background yields,  $f_x$  are the relative fractions of the bremsstrahlung categories (required to sum up to unity) and  $CB_x$  are the Crystal Ball functions describing the bremsstrahlung categories ( $x \in \{0; 1; 2\}$ ). The factor  $N_{\text{exp}}$  normalises the integral of the exponential function to unity. The overall yield is constrained by a Poisson term  $\mathcal{P}$  to match the entries in data. As the fits for the extraction of  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ , displayed in Fig. 5.1 for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis, show, the separation in bremsstrahlung categories for  $m(e^+ e^- K^+)$  is necessary as the mass distribution is vastly different depending on whether bremsstrahlung corrections were applied to the electron candidates or not. Each of the double-sided Crystal Ball functions is defined [132] as

$$\begin{aligned}
 CB(m; \alpha, \alpha', n, n', \mu, \sigma) = \\
 N_{CB} \times \begin{cases} \exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right) & , \text{ for } \frac{m-\mu}{\sigma} > -\alpha \\ & \text{ and } \frac{m-\mu}{\sigma} < \alpha' \\ \left(\frac{n}{|\alpha|}\right)^n \times \exp\left(-\frac{\alpha^2}{2}\right) \times \left(\frac{n}{|\alpha|} - |\alpha| - \frac{m-\mu}{\sigma}\right) & , \text{ for } \frac{m-\mu}{\sigma} \leq -\alpha \\ \left(\frac{n'}{|\alpha'}\right)^{n'} \times \exp\left(-\frac{\alpha'^2}{2}\right) \times \left(\frac{n'}{|\alpha'} - |\alpha'| - \frac{m-\mu}{\sigma}\right) & , \text{ for } \frac{m-\mu}{\sigma} \geq \alpha' \end{cases} \quad (5.2)
 \end{aligned}$$

with a normalisation factor  $N_{CB}$ , ensuring that the integral of the Crystal Ball function is equal to unity, the expectation value  $\mu$ , the width of the Gaussian core  $\sigma$ , the constants  $\alpha^{(\prime)}$ , which define the transition point between Gaussian part and power law tails, and the parameters of the power laws  $n^{(\prime)}$ . In the fits the tail parameters are determined from simulation, while the width and mean of the Gaussian core are allowed to float in the fits to data.

For the fit to the  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  mode in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis, a slightly different model is chosen, consisting of the sum of two single-sided Crystal Ball functions (i.e. only one power law tail to the left), where the mean and tail parameters are shared between the two functions.



**Figure 5.1:** Fits to the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  mass distribution split by the four data-taking years used in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis for the application of the  $\mathcal{P}Plot$  technique. The plots for 2011/2012 (left/right) are shown on the top, while the plots for 2015/2016 (left/right) are shown at the bottom. The fit function is composed of an exponential function describing combinatorial background (denoted with “combinatorial”) and a double-sided Crystal Ball function per bremsstrahlung category where bremsstrahlung corrections have been added to none (“Brem 0”), one (“Brem 1”) or both (“Brem 2”) of the electrons.

### 5.3.2 Calibration of the simulation

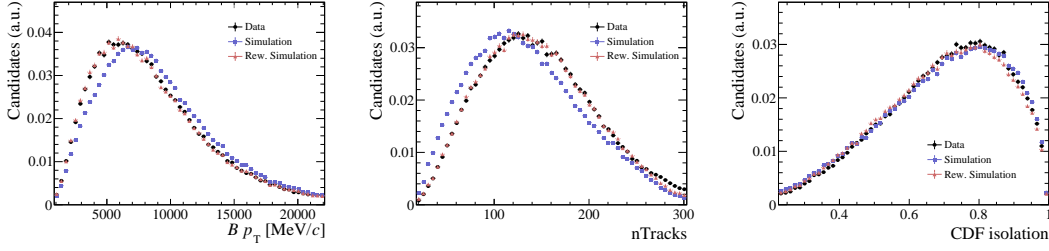
Applying the  $Weights$  to data, the simulation can be compared against a clean background-subtracted data sample, as shown in Fig. 5.2 for the 2016 data set of the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis. There are few distributions, where data and simulation do not match well. Among them are  $B$ -meson kinematic distributions, which occur because the  $b$ -hadronisation is known only to limited precision due to soft-QCD effects. Further differences are found in variables that are related to the underlying event. Those can be described by the number of tracks in the event ( $nTracks$ ) or the number of hits seen in the SPD detector ( $nSPDHits$ ) and arise because due to computing limitations the fragmentation processes in the  $pp$  collision cannot be fully simulated. Also isolation variables like the CDF isolation (described in detail in Section 5.4.2) are affected by the limitations described above, as they rely on the kinematic distributions of the  $B$ -candidate and non-signal tracks in the event. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis also small reconstruction mismodellings are considered by comparing the  $B^+$   $\chi_{\text{vtx}}^2$ , with however a very small impact.

The differences between data and simulation can be mitigated by weighting the simulation to match the background-subtracted data. This is done in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis by training a boosted decision tree with the gradient boosting technique (GBDT) [133] on the differences between background-subtracted data and simulation. The GBDT is trained to discriminate background-subtracted data and simulation, assigning weights according to the largest differences between data and simulation. Afterwards further iterations of the GBDT are trained on background-subtracted data and the weighted simulation of the previous iteration until the GBDT cannot discriminate the two samples anymore.

In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis a different technique is chosen: background-subtracted data and simulation are binned in the variables where they are most different. Then weights are obtained by iterating through the set of reweighting variables and consecutively dividing one-dimensional histograms of background-subtracted data and the weighted simulation with weights from the previous reweighting iteration.

Assuming that the difference between data and simulation is similar between the reweighted control mode and the signal channel that is searched for, the weights obtained in the control mode are applied to the signal mode as well in both analyses.

The choice of variables in which to reweight the simulation follows in general a simple principle: those variables are chosen, which are most different between data and simulation, while the set of variables is restricted to a minimum to keep transferability between signal and normalisation mode. The exact set of reweighting variables for the two investigated analyses is given in Table 5.9. In the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis decays of  $B^0$  and  $B_s^0$  mesons are investigated. Since the hadronisation process of  $b\bar{b}$ -pairs into  $B^0$  mesons and  $B_s^0$  mesons is different due to the different accompanying quark, the quality of the simulation might be different between the two  $b$ -hadron production mechanisms. Therefore a second set of weights are determined, using also the  $B_s^0 \rightarrow J/\psi \phi$  mode with  $J/\psi \rightarrow e^+ e^-$  and  $\phi \rightarrow K^+ K^-$ .



**Figure 5.2:** Distributions of  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  observables in 2016 data, where data and simulation match not perfectly. These are input variables for the reweighting classifier. They are compared for background-subtracted data (black), unweighted simulation (blue) and reweighted simulation (red). The distributions look similar in the other data sets investigated for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis.

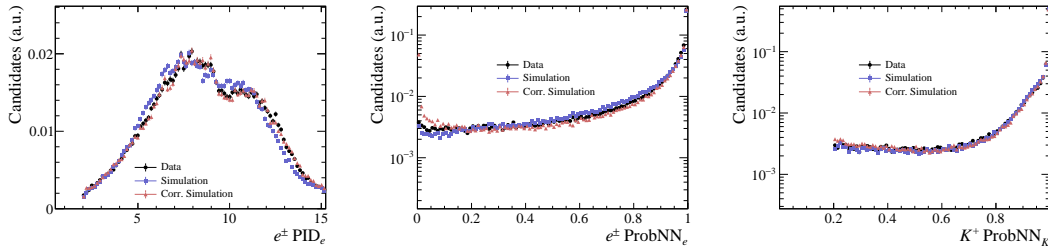
**Table 5.9:** Variables used for reweighting the simulation in the studied analyses.

$B^+ \rightarrow K^+ \mu^\pm e^\mp$ analysis	$B_{(s)}^0 \rightarrow e^+ e^-$ analysis
$B p_T$	$B p_T$
nTracks	nTracks
$B \chi_{\text{vtx}}^2$	CDF isolation

The set of weights from  $B_s^0 \rightarrow J/\psi \phi$  is applied to  $B_s^0 \rightarrow e^+ e^-$  simulation and the set of weights from  $B^+ \rightarrow K^+ J/\psi$  is applied to  $B^0 \rightarrow e^+ e^-$  simulation. However, the difference is found to be small between using either of the reweighting sets. This indicates that the quality of the simulation of  $B^0$  hadronisation and  $B_s^0$  hadronisation is similar. Applying the weights determined in one control mode to the other control mode serves as a cross-check, that the assumption of transferability between the decay modes is valid.

### 5.3.3 Correction of particle identification variables

Another set of variables where simulation does not perfectly describe the data is the set of PID variables. Those are difficult to describe in simulation, as they require the particle kinematics, the per-event detector occupancy and the experimental conditions, which also vary over time to be simulated correctly. Especially the RICH system is sensitive to the experimental conditions. To correctly estimate the efficiency of the requirements on PID variables, they are calibrated with the PIDCalib tool [134] using data with abundant decays that can be cleanly selected. The distributions of  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  decays are used to calibrate electrons,  $B^+ \rightarrow X J/\psi (\rightarrow \mu^+ \mu^-)$  (where  $X$  denotes particularly  $K^+$  and  $K^{*0}$ ) are used to calibrate muons,  $D^0 \rightarrow K^- \pi^+$  from  $D^{*+} \rightarrow D^0 \pi^+$  decays for kaons and pions, and  $\Lambda_c^+ \rightarrow p K^- \pi^+$  decays for protons. The pure decay samples are projected out with the  $sPlot$ -method in a



**Figure 5.3:** Particle identification variables used in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis compared between background-subtracted data (black) with reweighted simulation before (blue) and after resampling (red) in 2016 data. The other data sets studied in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis show a similar behaviour. The  $\text{PID}_e$  distribution shows two peaks corresponding to the two categories whether the electron received bremsstrahlung correction or not. Residual differences between corrected simulation and data are covered by the systematic uncertainties assigned to the simulation correction approach.

similar way as described above. The calibration samples are binned in detector occupancy and particle kinematics to take into account the correlation of those variables to the PID variables. New PID values are then sampled for the simulated candidate from the bin which corresponds to its kinematic and event multiplicity properties. A comparison on the control mode as shown exemplary in Fig. 5.3 for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis validates that the method produces the desired distributions. Some small residual differences are covered by systematic uncertainties that are assigned to this procedure, discussed in Section 5.7.

Different particle identification variables for the same particle, most notably  $\text{ProbNN}_e$  and  $\text{PID}_e$  for the electron candidates, are strongly correlated. Therefore determining the particle identification variable values independently of each other leads to a biased efficiency estimate, as it artificially decorrelates the variables. Thus the efficiencies of requirements on correlated particle identification variables are determined directly in bins of calibration data, where the correlation is correctly taken into account. Another particularity arises with the use of the  $\mathcal{P}lot$  method especially for electron PID variables: as the invariant mass of the calibration sample is correlated to the electron kinematics due to bremsstrahlung, in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the electron PID efficiencies are determined by fitting the calibration samples in bins of the electron kinematics before and after selecting with the PID requirements instead of using the  $\mathcal{P}lot$  method. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis this approach is not pursued, but a corresponding systematic uncertainty is assigned (see Section 5.7).

## 5.4 Multivariate and particle identification selection

The main selection of the performed analyses proceeds via multivariate classifiers and stringent requirements on particle identification variables. This suppresses backgrounds from random



track combinations and partially reconstructed background decays on the one hand and backgrounds from decays with misidentified final state particles on the other hand.

### 5.4.1 Multivariate classifiers

Multivariate classifiers provide powerful tools to discriminate signal and background contributions based on a set of observables that have only weak discriminating power themselves. This is achieved by classifying multidimensional regions in the observable space as corresponding to signal or background. The technique exploited in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  analyses is a Boosted Decision Tree (BDT) [135] implemented in the TMVA [110, 136] ( $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis) and the SCIKIT-LEARN [137] ( $B_{(s)}^0 \rightarrow e^+ e^-$  analysis) packages. In this method, decision trees are created that successively build subsets of the data, which are either dominated by signal or background by choosing a requirement on the input observables based on the variation of a metric like the so-called Gini-Index  $I = p(1 - p)$ , where  $p$  is the fraction of signal events in a subset. The requirement is chosen, if it maximises the Gini-Index difference with respect to the previous step. After completing the tree creation (where the completion is defined by the depth of the tree or perfect separation), the subsets are classified as signal or background depending on the dominating contribution in the subset. On the basis of the classification of one decision tree a new instance of a decision tree is created, where the input data are weighted by the level of misclassification in the previous tree. Subsets with worse classification get higher weights, so that the new decision tree concentrates on classifying those candidates. For the weighting in the discussed analyses the AdaBoost [135] algorithm is used. The total classification of one candidate is then defined as a linear combination of all classifiers, weighted with the AdaBoost algorithm. Thus the algorithm assigns a value to each candidate that represents its signal-likeness. After a maximum number of trees is reached or the addition of more classifiers does not lead to an increase in separation power, as defined by the above mentioned metric, the training is stopped. To avoid training on statistical fluctuations of the input data sets and thus biasing the classification, the  $k$ -Folding technique [138] is applied. The input data are split into  $k = 6$  ( $B_{(s)}^0 \rightarrow e^+ e^-$ ) or  $k = 10$  ( $B^+ \rightarrow K^+ \mu^\pm e^\mp$ ) sub-samples, of which the joined set of  $k - 1$  samples is used for training the classifier, which is then applied on the remaining sub-sample. This procedure is applied for every combination of sub-samples. Data that have not been used for the training of the classifier at all (such as the signal region, which is kept blind during the selection optimisation process) get the average classification value of all  $k$  sub-samples assigned.

### 5.4.2 BDT for the $B_{(s)}^0 \rightarrow e^+ e^-$ analysis

The BDT classifier for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis is trained with the purpose of suppressing background from random track combinations. Therefore the background input sample for the BDT consists of the upper mass sideband in data, where only this combinatorial

background is present. As a signal sample, reweighted  $B_s^0 \rightarrow e^+ e^-$  simulation is used. Due to centre-of-mass differences and new trigger strategies in 2015 and 2016 the data-taking periods 2011–2012 and 2015–2016 are trained separately, leading to different classifiers.

The input observables used for the classifier training are

- the square root of the minimum  $\chi_{\text{IP}}^2$  of the electron tracks,
- the sum of the track isolations (explained below) of the two electrons for the least isolating track (i.e. closest to the signal decay) and next-to-least isolating track,
- the DOCA between the two electron tracks,
- the  $B_s^0 p_{\text{T}}$ ,
- the  $B_s^0$  isolation as defined by CDF [74],
- the  $B_s^0 \chi_{\text{IP}}^2$ ,
- the absolute difference in the electrons pseudorapidity  $\delta\eta$ ,
- the  $B_s^0$  flight distance.

The CDF isolation variable [74] of the  $B_s^0$  candidate is defined as

$$I_{\text{CDF}} = \frac{p_{\text{T}}(B_s^0)}{p_{\text{T}}(B_s^0) + \sum_{\text{tracks}} p_{\text{T}}(tr)}, \quad (5.3)$$

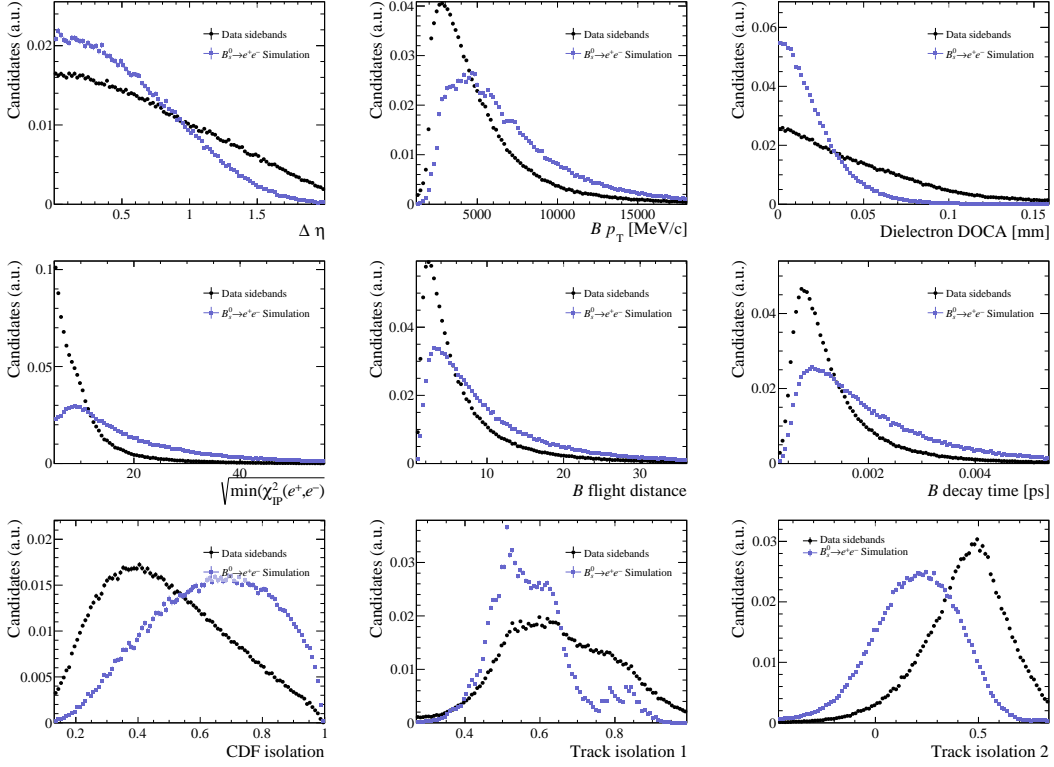
where  $p_{\text{T}}(x)$  are the transverse momenta of the  $B_s^0$  meson and additional tracks in the events. The summation is performed over all tracks inside a cone around the  $B_s^0$  transverse momentum defined by  $\sqrt{\delta\eta^2 + \delta\phi^2} < 1$ , where  $\eta$  is again the pseudorapidity and  $\phi$  the azimuth angle.

The distributions of the classifier input observables in the data sidebands and reweighted  $B_s^0 \rightarrow e^+ e^-$  simulation are shown in Fig. 5.4 exemplary for the combined 2015–2016 data set, emphasising the discrimination power of the individual observables. The isolation variables prove to be the most discriminating observables.

### BDT based track isolation for the $B_{(s)}^0 \rightarrow e^+ e^-$ analysis

The track isolation variable has been developed for the  $\tau^+ \rightarrow \mu^+ \mu^- \mu^+$  and  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analyses [139] but is available for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis also. This track isolation variable is a BDT classifier that is based on the assumption that tracks from the signal decay are isolated from the rest of the event. It calculates the background-likeness of every combination of a signal electron (track) with each other track in the event that traverses all detector stations.

The isolation BDT has been trained using isolating tracks from  $B_s^0 \rightarrow \mu^+ \mu^-$  simulation as signal sample and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  and  $c\bar{c} \rightarrow \mu^+ \mu^- X$  simulation as background sample.



**Figure 5.4:** Distributions of the BDT input observables in the data sidebands and reweighted  $B_s^0 \rightarrow e^+e^-$  simulation for the combined 2015–2016 data. The sideband data and the signal simulation show a clear difference, which is exploited by the BDT to discriminate  $B_s^0 \rightarrow e^+e^-$  signal from combinatorial background. The track isolation 1 distribution shows several peaks because different track types are used in the training of the isolation BDT.

Several track types are considered: whether the tracks have been measured in all tracking detectors, or whether it has not been measured in all detectors but the VELO or whether it has been measured in all detectors but the tracking stations. The input variables for the BDT calculating the track isolation are

- the logarithm of the minimum  $\chi_{\text{IP}}^2$  of the track with respect to any primary vertex in the event,
- the distance between the track primary vertex and the electron primary vertex,
- the distance between the vertex of the track with the electron and the  $B$  vertex,
- the logarithm of the distance of closest approach between the track and the electron,
- the angle between the track momentum and the electron momentum,
- the cone isolation  $f_c = \frac{|P_e + P_{\text{trk}}| \sin \alpha^{e+\text{trk},PV}}{|P_e + P_{\text{trk}}| \sin \alpha^{e+\text{trk},PV} + p_{T,e} + p_{T,\text{trk}}}$ , where  $\alpha^{e+\text{trk},PV}$  is the angle between the sum of the electron and track momenta, and the direction defined by the primary vertex and the track vertex [140].

For each of the signal electron tracks the isolation BDT values for the two most non-isolating (i.e. closest to the signal) tracks in the event are written out and their sum is used as input to the selection classifier. Because several types of tracks are considered for the evaluation of the BDT, the isolation shows several peaks in Fig. 5.4.

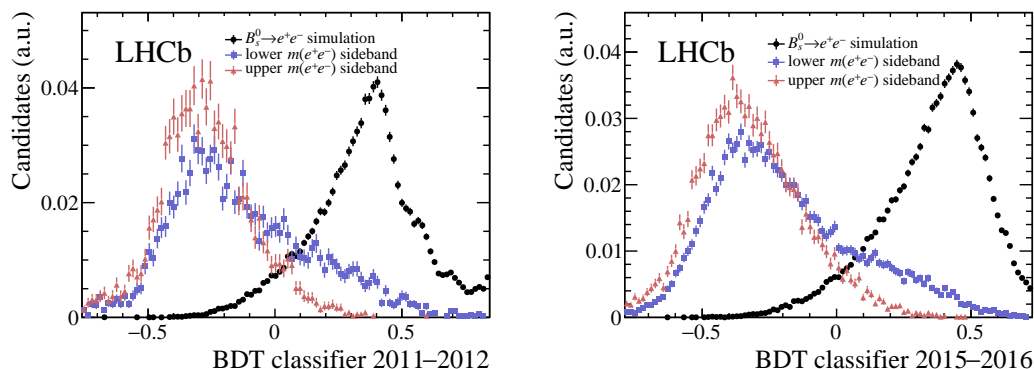
### Results of the BDT classification for $B_{(s)}^0 \rightarrow e^+ e^-$

The resulting classification value distributions for signal simulation and the data sidebands are shown in Fig. 5.5 and proves a strong separation between the signal and background samples. Also a clear difference is seen between the distributions for the lower and the upper mass sideband. It was checked that the classification is independent of the mass of the candidate. Therefore this difference clearly shows that the upper and the lower mass sideband are composed of different contributions. While the upper mass sideband consists purely of backgrounds from random track combinations, the lower mass sideband also contains to a large extent backgrounds from partially reconstructed decays and decays with misidentified final state particles. These have very similar properties to signal decays, as they also come from  $b$ -hadron decays and are modelled in a final fit to obtain the branching fraction.

To find the optimal selection requirement on the classifier value, the Punzi figure of merit [141] is used

$$FoM = \frac{\varepsilon_{\text{sig}}}{N_{\text{bkg}} + \frac{3}{2}}, \quad (5.4)$$

which balances the signal efficiency  $\varepsilon_{\text{sig}}$  against the expected background yield in the signal window  $N_{\text{bkg}}$ . The yield is obtained from fitting a single exponential to the unselected data



**Figure 5.5:** BDT distributions of corrected signal simulation and the dielectron mass sidebands for the (left) Run 1 and (right) Run 2 classifier in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis. They illustrate the good separation power of the classifier. The classifier distribution is checked to not depend on the dielectron invariant-mass. Differences between the upper and lower dielectron mass sideband arise due to specific signal-like background decays present in the lower mass sideband.

sidebands (no BDT requirement) and interpolating into the signal region. The expected background yield after the BDT requirement is then determined as the expected yield before any BDT selection multiplied with the efficiency of the requirement on the data sidebands. This procedure is applied to facilitate the fit to the sidebands, as the requirements on the classifier value change the composition of backgrounds in the lower mass sideband and thus its shape.

The optimal BDT selection points have been found to be  $\text{BDT} > 0.27$  for Run 1 and  $\text{BDT} > 0.30$  for Run 2.

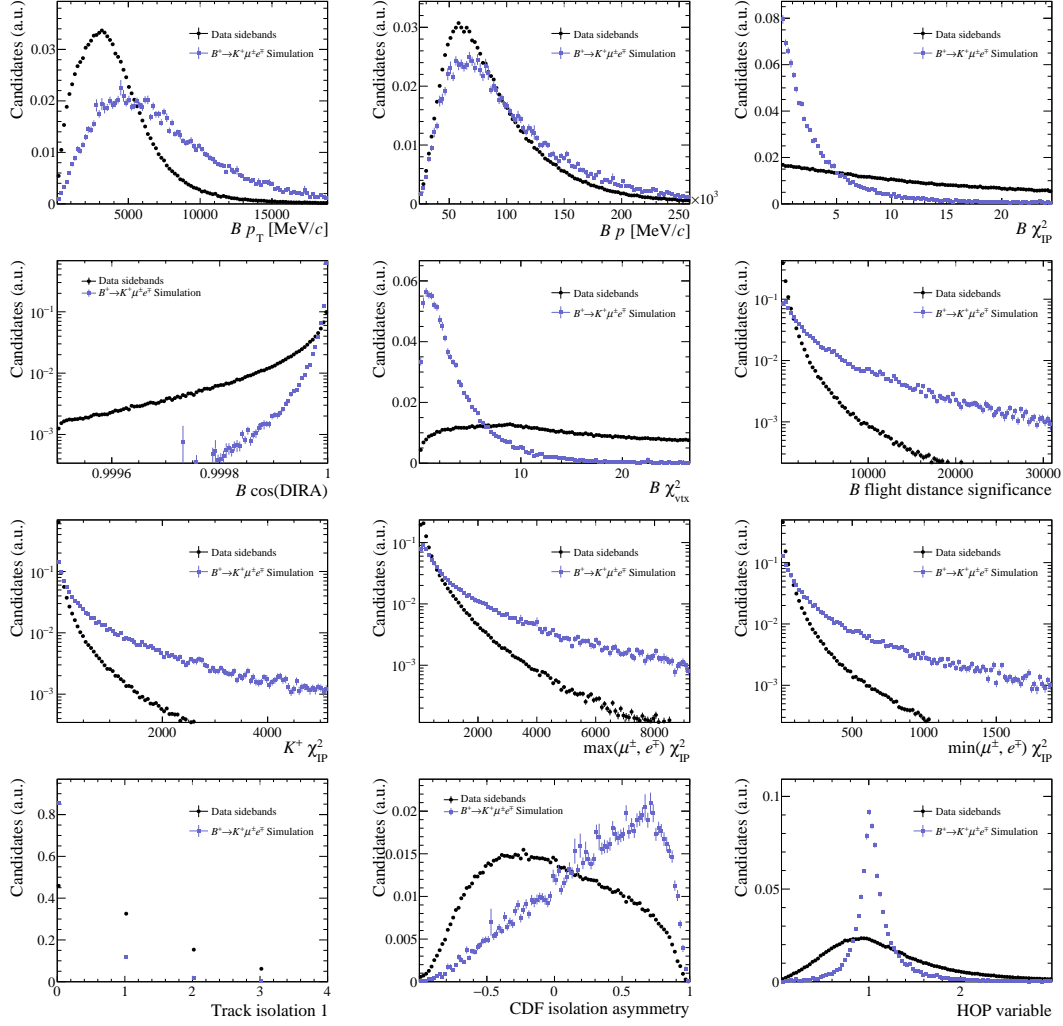
### 5.4.3 BDT for the $B^+ \rightarrow K^+\mu^\pm e^\mp$ analysis

The multivariate selection strategy in the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  analysis is two-fold: first a multivariate classifier is trained against background from random track combinations by training a BDT classifier to discriminate simulated  $B^+ \rightarrow K^+\mu^\pm e^\mp$  samples against upper mass sideband data as done in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis. The requirement on the output of the BDT classifier is determined by defining the expected upper limit from the  $\text{CL}_S$  method as the figure of merit. Then, after applying the BDT classifier cut, another classifier ( $\text{BDT}_{\text{HOP}}$ ) is trained with the aim to remove partially reconstructed backgrounds. These are expected to populate the lower mass sideband. Thus this second classifier is trained to discriminate candidates from the lower mass sideband against  $B^+ \rightarrow K^+\mu^\pm e^\mp$  simulation. The same procedure as in the first BDT is applied to find the optimal requirement on the BDT classifier. For both classifiers the set of input variables is comprised of

- the  $p_T$  of the  $B$  candidate,
- the momentum of the  $B$  candidate,
- the  $\chi_{\text{IP}}^2$ , of the  $B$  candidate,
- the cosine of the DIRA angle of the  $B$  candidate,
- the  $B$  vertex  $\chi^2$ ,
- the  $B$  flight distance  $\chi^2$
- the kaon  $\chi_{\text{IP}}^2$
- the minimum and maximum  $\chi_{\text{IP}}^2$  of the electron and muon candidates
- the number of good quality vertices a lepton track can make with any other track in the event [142] for both leptons as a measure of the “track isolation”,
- the CDF isolation asymmetry [74] for the kaon:  $I_{\text{CDF,asym}} = \frac{p_T(K) - \sum_{\text{tracks}} p_T(tr)}{p_T(K) + \sum_{\text{tracks}} p_T(tr)}$ , and
- the HOP variable as developed for the  $R_{K^*0}$  analysis [7], defined by  $\alpha_{\text{HOP}} = \frac{P_T(e)}{P_T(K\mu)}$  with the momenta  $P_T$  orthogonal to the  $B$  direction of flight,

while the HOP variable is only used in the second BDT. The input observable distributions of the BDT's are displayed in Fig. 5.6 and show a clear difference between the data sidebands and reweighted  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  simulation, which is used by the BDT to discriminate signal  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays from background contributions. Due to the three-prong vertex of the signal decay here a stronger focus is laid on vertex and reconstruction related variables than in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis, as three tracks pointing to the same vertex which is displaced from the primary vertex provide a clear signature. The HOP variable was developed to correct the  $B$ -meson mass for decays involving electrons, since only a part of the electron momenta can be recovered due to bremsstrahlung losses. This loss can be measured by the momentum imbalance  $\alpha_{\text{HOP}}$ . However a similar imbalance appears in the case of partially reconstructed decays, where one or more final state particles are not reconstructed. This imbalance is larger in partially reconstructed decays than in the signal decays, which makes  $\alpha_{\text{HOP}}$  a useful variable to discriminate signal decays from partially reconstructed decays. Thus it is additionally used as input variable for the second BDT.

The optimal requirements on the classifier values have been found to be  $\text{BDT} > 0.5$  and  $\text{BDT}_{\text{HOP}} > 0.98$ . For the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decay two different charge configurations are analysed,  $B^+ \rightarrow K^+ \mu^- e^+$  and  $B^+ \rightarrow K^+ \mu^+ e^-$ . The optimal BDT requirements have been investigated separately for the two charge configurations, yielding consistent optima.



**Figure 5.6:** Distributions of the BDT input observables in the data sidebands and reweighted  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  simulation for the combined Run 1 data. The sideband data and the signal simulation show a clear difference, which is exploited by the BDT to discriminate the two samples. The distribution of the track isolation is only shown for one of the two leptons, as the distribution for the other lepton looks similar.

**Table 5.10:** Final PID requirements for the  $B_{(s)}^0 \rightarrow e^+ e^-$  and  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analyses with the requirements in Run 2 in brackets.

Particle	Variable	Cut $B^+ \rightarrow K^+ \mu^\pm e^\mp$	Cut $B_{(s)}^0 \rightarrow e^+ e^-$ (Run 2)
$\mu$	ProbNN $_\mu$	$> 0.7$	–
$K$	ProbNN $_K$	$> 0.65$	–
$e$	ProbNN $_e$	$> 0.65$	$> 0.9(0.6)$

#### 5.4.4 Particle identification criteria

After the BDT selection to remove backgrounds from random track combinations and partially reconstructed decays, tight PID requirements are applied to all final state particles. The PID variables are output distributions from neural nets, which are based on information from the subdetectors of the LHCb experiment. The PID requirements are chosen by hand to the values reported in Table 5.10 due to various reasons: in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis only few candidates survived the previous selection, making a systematical optimisation impossible due to the large statistical uncertainties. Thus a selection as tight as possible without removing too many signal candidates (as determined from simulation) was chosen. For the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis an attempt was made to find the optimal requirement simultaneously with the requirement on the BDT using the Punzi figure of merit described before. However, it was found that the optimal requirement found this way was expected to leave a significant background contribution from  $B \rightarrow hb'$  decays prominent in the signal region, which cannot be estimated from the sidebands. Therefore the PID requirement was tightened by hand to strongly suppress those decays while maintaining a good signal efficiency. The ProbNN algorithms have been retrained for Run 2 data resulting in significantly improved performance, which allows a higher signal efficiency at similar background rejection.

### 5.5 Efficiency calculations

The efficiencies of reconstruction and selection of signal and normalisation candidates are in general determined using the formulae

$$\varepsilon = \frac{\sum N_{\text{passed}}}{\sum N_{\text{total}}} \text{ and} \quad (5.5)$$

$$\sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{\text{total}}}}, \quad (5.6)$$



which in the case of weighted simulation turns into

$$\varepsilon = \frac{\sum_{\text{passed}} w}{\sum_{\text{all}} w} \text{ and} \quad (5.7)$$

$$\sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{\sum_{\text{all}} w}}. \quad (5.8)$$

The efficiencies are split into different stages of the selection, where the efficiency of each stage is calculated relative to the previous one. Thus the efficiency chains read:

$$\varepsilon_{B_{(s)}^0 \rightarrow e^+e^-} = \varepsilon_{\text{geo}} \times \varepsilon_{\text{rec|geo}} \times \varepsilon_{\text{sel|rec}} \times \varepsilon_{\text{PID|sel}} \times \varepsilon_{\text{BDT|PID}} \times \varepsilon_{\text{trig|BDT}} \quad (5.9)$$

and

$$\begin{aligned} \varepsilon_{B^+ \rightarrow K^+ \mu^\pm e^\mp} &= \varepsilon_{\text{geo}} \times \varepsilon_{\text{rec|geo}} \times \varepsilon_{\text{sel|rec}} \times \varepsilon_{\text{trig|sel}} \times \varepsilon_{\text{PIDs|trig}} \\ &\times \varepsilon_{\text{BDT|PIDs}} \times \varepsilon_{\text{BDTHOP|BDT}} \times \varepsilon_{\text{PID|BDTHOP}}. \end{aligned} \quad (5.10)$$

Wherever possible, efficiencies are determined using simulation, applying weights from Section 5.3.2 to account for data-simulation differences. However, weights cannot be applied to simulation before reconstructing the candidates. Therefore the geometrical efficiency  $\varepsilon_{\text{geo}}$ , which denotes the efficiency of the requirement that all final state particles fall loosely into the acceptance region of the LHCb detector, and the reconstruction efficiency  $\varepsilon_{\text{rec|geo}}$ , describing the efficiency to reconstruct and loosely preselect the candidates in the experiment-wide preselection, are taken from unweighted simulation.

Also trigger efficiencies ( $\varepsilon_{\text{trig}}$ ) and particle identification efficiencies ( $\varepsilon_{\text{PID}}$ ,  $\varepsilon_{\text{PIDs}}$ ) receive particular treatment and can be directly calculated on data as briefly discussed in the following sections.

The fractions of bremsstrahlung categories are taken from simulation and are for the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis cross-checked on normalisation mode data.

### 5.5.1 Trigger efficiencies

The trigger efficiencies of the control and normalisation mode  $B^+ \rightarrow K^+ J/\psi$  are obtained with the help of the TISTOS method [143] on data. As only triggered data are available, Eq. (5.5) and Eq. (5.7) cannot be used to determine efficiencies, because  $N_{\text{total}}$  is not accessible. Instead it is approximated by assuming that the fractions of TOS triggered events and TIS triggered events are independent. Thus the number of untriggered events  $N_{\text{total}}$  can be calculated as

$$N_{\text{total}} = \frac{N_{\text{TOS}}}{\varepsilon_{\text{TOS}}} = \frac{N_{\text{TIS}} N_{\text{TOS}}}{N_{\text{TISTOS}}}, \quad (5.11)$$

with the efficiency of triggering TOS,  $\varepsilon_{\text{TOS}}$ , and the number of events triggered TIS and TOS,  $N_{\text{TISTOS}}$ . The number of triggered signal events are determined by fitting each TOS, TIS or TISTOS triggered data set. However, particularly the momentum spectra of the  $B$  mesons are correlated to the remainder of the event, as they are produced from  $b\bar{b}$ -pairs. Since the trigger selection mainly proceeds via  $p_{\text{T}}$  and IP cuts, the assumption of independence of TOS and TIS can only safely assumed in small regions of phase space. Therefore  $N_{\text{total}}$  is determined in bins of  $B^+$   $p_{\text{T}}$  and  $p_z$ .

To determine the trigger efficiency in the signal mode where no signal events are expected, in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the trigger efficiency from  $B^+ \rightarrow K^+ J/\psi$  is folded with a map in the maximum electron IP and maximum electron  $p_{\text{T}}$  in simulation. Those are variables most relevant to the trigger system and different between the control and signal mode.

The validity of the method is checked on simulation, where the efficiencies can also be determined in a cut-based approach from Eq. (5.7). Differences between cut-based efficiencies and efficiencies from the TISTOS method are assigned as systematic uncertainty.

Since the amount of TIS and TOS triggered events is small (especially for  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ ), the efficiencies determined with this approach come with relatively large uncertainties, which dominate the uncertainty budget of the overall efficiency.

The trigger efficiencies of the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis are directly determined on simulation, as the trigger selection is based on the muon tracks and the decay topology, which are well understood in simulation. This approach was checked with the TISTOS method on the normalisation mode  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  and found to be correct within 1% precision. However, some small data-simulation differences are found. To prove that they do not affect the ratio of trigger efficiencies between signal and normalisation mode, the differences are investigated in bins of the maximum muon  $p_{\text{T}}$ , which is the trigger-related variable most different between the two modes. No significant trend is observed and therefore the data-simulation differences are assumed to be equal between signal and normalisation mode and thus cancel out in the ratio.

## 5.5.2 Particle identification efficiencies

As discussed in Section 5.3.3, the particle identification variables are recomputed individually on simulation from background-subtracted samples in calibration data. However, this neglects the correlation between  $\text{PID}_e$  and  $\text{ProbNN}_e$ . To take the correlation into account, successive requirements on particle identification variables are evaluated on calibration data conditioned to the prior requirement: efficiency maps are evaluated on calibration data in bins of detector occupancy and particle kinematics and are then folded with the distributions in reweighted simulation. This method is checked on data with the normalisation and control modes to give the correct results.

**Table 5.11:** Total selection efficiencies for the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis, split by decay channel, year and bremsstrahlung category. The ratios needed for the normalisation constants are also reported. The selection efficiencies are expected to be similar between  $B_s^0 \rightarrow e^+e^-$  and  $B^0 \rightarrow e^+e^-$ .

Category	$B_s^0 \rightarrow e^+e^-$ [%]	$B^+ \rightarrow K^+J/\psi$ [%]	$B^+ \rightarrow K^+J/\psi / B_s^0 \rightarrow e^+e^-$
2011			
0	$0.109 \pm 0.011$	$0.0173 \pm 0.0016$	$0.160 \pm 0.022$
1	$0.324 \pm 0.029$	$0.0460 \pm 0.0031$	$0.142 \pm 0.016$
2	$0.254 \pm 0.036$	$0.0215 \pm 0.0030$	$0.084 \pm 0.017$
2012			
0	$0.096 \pm 0.007$	$0.0131 \pm 0.0015$	$0.138 \pm 0.015$
1	$0.284 \pm 0.017$	$0.0309 \pm 0.0017$	$0.109 \pm 0.009$
2	$0.187 \pm 0.022$	$0.0162 \pm 0.0013$	$0.087 \pm 0.012$
2015			
0	$0.122 \pm 0.009$	$0.0171 \pm 0.0017$	$0.141 \pm 0.010$
1	$0.350 \pm 0.025$	$0.0392 \pm 0.0035$	$0.112 \pm 0.013$
2	$0.247 \pm 0.022$	$0.0199 \pm 0.0024$	$0.081 \pm 0.012$
2016			
0	$0.135 \pm 0.004$	$0.0202 \pm 0.0007$	$0.150 \pm 0.007$
1	$0.389 \pm 0.011$	$0.0461 \pm 0.0013$	$0.119 \pm 0.005$
2	$0.266 \pm 0.013$	$0.0229 \pm 0.0009$	$0.086 \pm 0.006$

### 5.5.3 Discussion of the total selection efficiencies

The overall efficiencies of the discussed analyses are shown in Table 5.11 and Table 5.12.

The efficiencies for  $B_s^0 \rightarrow e^+e^-$  and  $B^0 \rightarrow e^+e^-$  are assumed to be similar. For the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis a significantly lower efficiency in the normalisation mode is found with respect to the signal mode. This appears due to various reasons, which are connected to the additional track in the  $B^+ \rightarrow K^+J/\psi$  decay. A large efficiency drop appears in reconstructing  $B^+ \rightarrow K^+J/\psi$  decays from the efficiency to reconstruct the additional track and associating three tracks to the same vertex. Due to the additional track also the isolation of the electron tracks is worse, which thus impacts the BDT selection efficiency. Also the trigger selection is optimised for a two-particle final state and therefore has a lower efficiency in the  $B^+ \rightarrow K^+J/\psi$  mode. Finally an additional selection on  $B^+ \rightarrow K^+J/\psi$  is done to allow a clean *sPlot* fit, which further reduces the overall efficiency. The efficiencies are similar between the years across the data taking periods Run 1 and Run 2. However, they are significantly higher in Run 2, as the

**Table 5.12:** Total selection efficiencies of signal, control and normalisation modes for the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis, split by year and charge configuration.

$\varepsilon_{B^+ \rightarrow K^+ \mu^- e^+}$ [%]	$\varepsilon_{B^+ \rightarrow K^+ \mu^+ e^-}$ [%]	$\varepsilon_{B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)}$ [%]	$\varepsilon_{B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)}$ [%]
2011			
$0.0887 \pm 0.0024$	$0.0778 \pm 0.0023$	$0.246 \pm 0.0014$	$0.0805 \pm 0.0011$
2012			
$0.0752 \pm 0.0016$	$0.0672 \pm 0.0015$	$0.213 \pm 0.0009$	$0.0578 \pm 0.0007$

trigger algorithms and particle identification variables have been revised and re-optimised significantly. The different trigger selection and the higher centre-of-mass energy in Run 2 also improve the BDT separation. The ratio of efficiencies is consistent between the years but shows a significant trend over the bremsstrahlung categories. This trend is also induced by the additional track in the normalisation mode. The slightly different reconstruction and the additional preselection have different efficiencies in the bremsstrahlung categories. Also the different impact of the BDT classification supports the trend in the efficiency ratio.

The selection efficiencies in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis are much more precise than in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis due to the different way the trigger efficiencies are estimated. They show some difference between the 2011 and 2012 efficiencies arising from slightly different trigger configurations and a centre-of-mass energy change from 7 TeV to 8 TeV. Differences in the selection efficiencies between the two charge configurations of the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decay arise from the different effect of the veto selection applied. Due to the low resolution of the  $m(K^+ e^-)$  compared to  $m(K^+ \mu^-)$ , the signal efficiency of the veto on this combination is significantly lower. The  $m(K^+ \mu^-)$  veto also improves the trigger efficiency, as it prefers muons with higher momentum, which are more likely triggered on. A strong difference is visible between the efficiency of signal and normalisation mode because the vetoes (partly designed to remove charmonium resonances) are not applied to the normalisation mode. Also the trigger efficiencies and the efficiency of the  $BDT_{\text{HOP}}$  selection is higher in the normalisation mode, as it exchanges one electron in the final state for a muon with respect to the signal mode. This leads to a higher resolution in the HOP variable and is a second particle with signal signature in the L0 trigger selection. These differences also explain the low efficiency of the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  mode. This mode is triggered completely different from the signal and normalisation mode and has a much worse resolution as it contains no muon, but electrons instead, which loose energy due to bremsstrahlung that is only incompletely recovered.

**Table 5.13:** Yields of the normalisation mode as computed from the fits shown in Fig. 5.7. Those yields enter the normalisation factor computation.

$B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ yields for the $B_{(s)}^0 \rightarrow e^+ e^-$ analysis		
	Run 1	Run 2
Category 0	$5050 \pm 70$	$7960 \pm 70$
Category 1	$10430 \pm 110$	$16750 \pm 130$
Category 2	$4950 \pm 70$	$8306 \pm 32$
$B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$ yields for the $B^+ \rightarrow K^+ \mu^\pm e^\mp$ analysis		
	2011	2012
	$26940 \pm 170$	$59220 \pm 250$

### 5.5.4 Normalisation factors

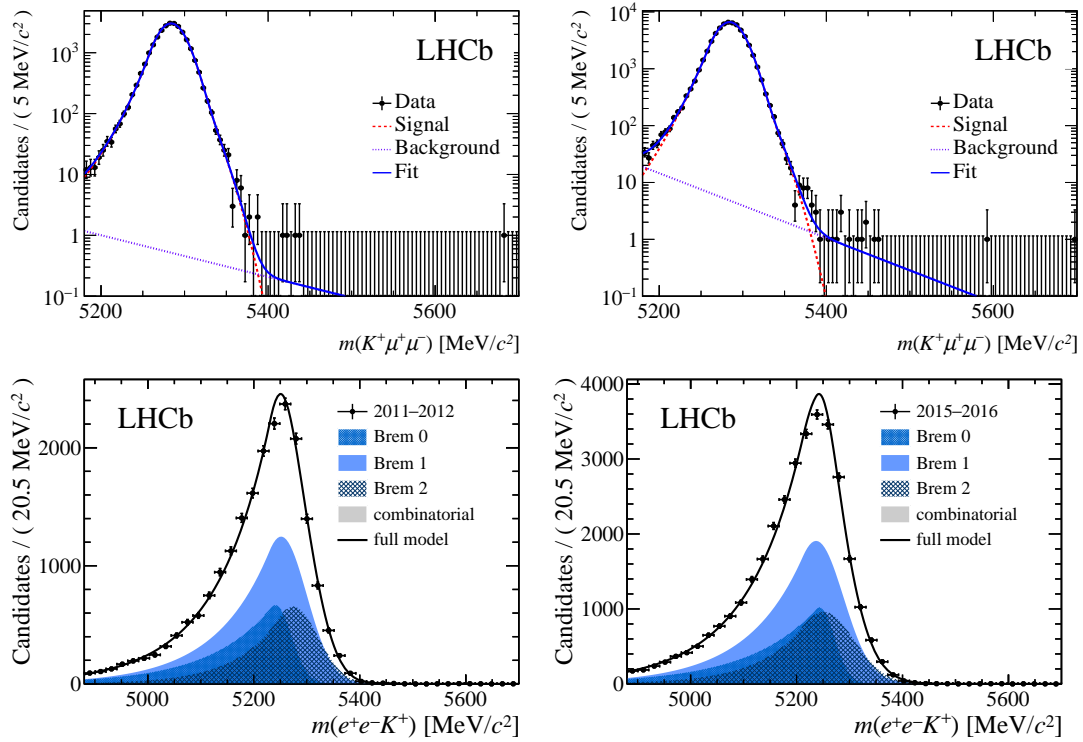
The signal branching fraction is determined in a fit to the mass distribution. To convert the signal yield into the signal branching fraction, the normalisation factor  $\alpha$  and partial normalisation factor  $\alpha'$  have been introduced in Eq. (5.1). They comprise of the efficiencies of signal and normalisation mode and the normalisation yield, as well as external quantities like the normalisation branching fraction and the  $b$ -hadron fragmentation fraction ratio  $f_{u,d,s}/f_u$  for the full normalisation factor. The normalisation yield is calculated from a fit to the mass  $m(K^+ \ell^+ \ell^-)$  distribution of the fully selected normalisation data to the values in Table 5.13 and is displayed in Fig. 5.7.

The values for  $f_{u,d}/f_u$  are assumed to be unity, while  $f_s/f_d$  is measured to be  $0.259 \pm 0.015$  in Run 1 [144] and increase by  $1.068 \pm 0.022$  in Run 2 [145]. In combination, the partial normalisation factors and single event sensitivities result in the numbers given in Tables 5.14 and 5.15. The numbers are combined between the years by the harmonic sum

$$\frac{1}{\alpha^{(l)}} = \sum_{\text{year}} \frac{1}{\alpha_{\text{year}}^{(l)}}. \quad (5.12)$$

The tables also include the systematic uncertainties as discussed in Section 5.7. The normalisation factors of the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis are about one order of magnitude higher than the normalisation factors of the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis, mainly due to the tight selection requirements in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis, which lead to a very small data set after the full selection. This makes the single event sensitivity larger, but reduces the background yield, which the limit is also sensitive to. Also the size of the data set used in the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  is only about half the size of the one used to search for  $B_{(s)}^0 \rightarrow e^+ e^-$ .

While the systematic uncertainties are larger than the statistical uncertainties in these values, this does not mean that the analyses are systematically limited. The normalisation factors



**Figure 5.7:** Fits to obtain the normalisation yield from  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  for the (top)  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis in the (left) 2011 and the (right) 2012 data set. The fits to the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  mode to obtain the normalisation yield for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis are shown on the bottom for (left) Run 1 and (right) Run 2.

**Table 5.14:** Partial normalisation factors and single event sensitivities for the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis. The partial normalisation factors are used in the limit calculation. The first quoted uncertainty stems from the limited size of the simulated samples and normalisation data. The second quoted uncertainty arises from systematic uncertainties as discussed in Section 5.7.

	$\alpha'$ (Run 1) [ $10^{-5}$ ]	$\alpha'$ (Run 2) [ $10^{-5}$ ]
Category 0	$2.85 \pm 0.24 \pm 0.17$	$1.84 \pm 0.08 \pm 0.09$
Category 1	$1.13 \pm 0.08 \pm 0.07$	$0.702 \pm 0.027 \pm 0.030$
Category 2	$1.73 \pm 0.20 \pm 0.11$	$1.04 \pm 0.06 \pm 0.06$
Combined single event sensitivities		
$\alpha_{B^0 \rightarrow e^+e^-}$ [ $10^{-10}$ ]	$1.271 \pm 0.034(\text{stat.}) \pm 0.063(\text{syst.})$	
$\alpha_{B_s^0 \rightarrow e^+e^-}$ [ $10^{-10}$ ]	$4.71 \pm 0.12(\text{stat.}) \pm 0.33(\text{syst.})$	

**Table 5.15:** Partial normalisation factors and single event sensitivities for the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  analysis. The partial normalisation factors are used in the limit calculation. The first quoted uncertainty stems from the limited size of the simulated samples and normalisation data. The second quoted uncertainty arises from systematic uncertainties as discussed in Section 5.7.

	$B^+ \rightarrow K^+\mu^+e^-$	$B^+ \rightarrow K^+\mu^-e^+$
$\alpha'$ [ $10^{-5}$ ]	$3.68 \pm 0.07 \pm 0.21$	$3.27 \pm 0.06 \pm 0.21$
Single event sensitivities		
$\alpha$ [ $10^{-9}$ ]	$2.21 \pm 0.04(\text{stat.}) \pm 0.14(\text{syst.})$	$1.97 \pm 0.04(\text{stat.}) \pm 0.14(\text{syst.})$

do not take into account the size of the final data set, which is fitted to obtain the signal branching fraction. As long as there is no significant signal contribution in the final data set, the uncertainty on the signal yield dominates the measurement.

### 5.5.5 Normalisation cross-check for the $B_{(s)}^0 \rightarrow e^+e^-$ analysis

For a check of the robustness of the analysis the absolute efficiencies of the normalisation mode are investigated. This is a stringent cross-check, as for absolute efficiencies potential systematic effects that do not affect the efficiency ratio between signal and normalisation mode used in the analysis do not cancel. Two checks are performed with different strategies. First a comparison of expected overall  $B^+$  yields is done in Section 5.5.5. Secondly, an alternative strategy is exploited by looking at the branching fraction ratio with the mode  $B^0 \rightarrow K^+\pi^-$  in Section 5.5.6.

### Cross-check of $B^+$ yields

To check the efficiency calculations, the total number of produced  $B^+$  mesons are calculated from the efficiencies and yields found in this analysis and compared to the expected yields from cross-section and luminosity measurements. The measured  $b\bar{b}$  cross-sections at the centre-of-mass energy of  $\sqrt{s} = 7$  TeV (2011) [99],  $\sqrt{s} = 8$  TeV (2012) [100], and  $\sqrt{s} = 13$  TeV (2016) [101] are

$$\begin{aligned}\sigma_{b\bar{b}}(2011) &= (288 \pm 4 \pm 48) \mu\text{b}, \\ \sigma_{b\bar{b}}(2012) &= (298 \pm 2 \pm 36) \mu\text{b}, \\ \sigma_{b\bar{b}}(2015/2016) &= (495 \pm 2 \pm 52) \mu\text{b}.\end{aligned}$$

Together with the hadronisation fraction of  $b$  quarks into  $B^+$  mesons [49],

$$f_u = 0.402 \pm 0.007, \quad (5.13)$$

and the luminosities given in Table 4.1 the number of expected  $B^+$  candidates can be calculated with the formula

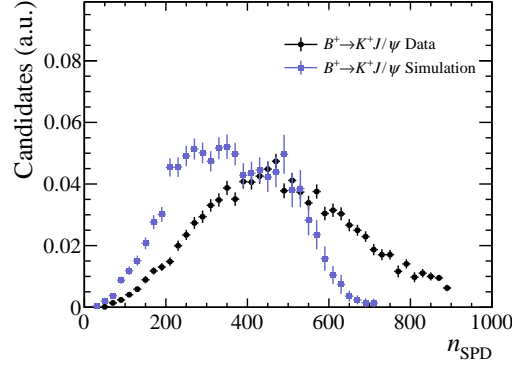
$$N_{B^+, \text{expected}} = \sum_{\text{year}} \mathcal{L} \times 2 \times \sigma_{b\bar{b}}(\text{year}) \times f_u.$$

The same numbers are computed using the calculated efficiencies from Table 5.11 and yields from Table 5.13 with the relation

$$N_{B^+} = \frac{N_{B^+ \rightarrow K^+ J/\psi, \text{category}}}{\varepsilon_{B^+ \rightarrow K^+ J/\psi, \text{category}} \times \mathcal{B}_{B^+ \rightarrow K^+ J/\psi}}. \quad (5.14)$$

During this check a strong deviation between expected and measured  $B^+$  yields was observed, which could be traced to come from an incorrect reweighting of the SPD hit multiplicity ( $n_{\text{SPD}}$ ). Tight requirements are enforced on this quantity with fiducial cuts in Table 5.4 to facilitate the PID calibration, as many L0 triggers have tight thresholds on this observable. However, it is not well modelled in simulation, leading to a wrong efficiency estimation. To estimate the effect, a sample of  $B^+ \rightarrow K^+ J/\psi$  decays is selected with L0DiMuon TIS (where the threshold is at  $n_{\text{SPD}} < 900$ ), applying the nominal reweighting from Section 5.3.2 and removing the requirement on the SPD hit multiplicity in the fiducial cuts. The corresponding distributions for 2016 data are shown in Fig. 5.8. For these samples a one dimensional reweighting in  $n_{\text{SPD}}$  is performed. A correction factor for this effect is obtained as the relative difference  $\Delta\varepsilon_{\text{SPD}}$  between the efficiency of the  $n_{\text{SPD}}$  cut (from Table 5.4), once applied on simulation with the nominal reweighting only and once applied on simulation with nominal





**Figure 5.8:** Comparison between background subtracted  $B^+ \rightarrow K^+J/\psi$  data (black) and simulation with nominal reweighting (blue) in the 2016 data set. The difference is reweighted to obtain the correction factor  $\Delta\varepsilon_{\text{SPD}}$  on the efficiency of the cut on  $n_{\text{SPD}}$ .

reweighting and additionally the one dimensional  $n_{\text{SPD}}$  reweighting. The obtained correction factors are

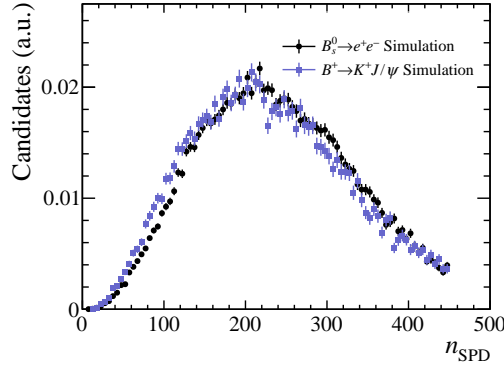
$$\begin{aligned}\Delta\varepsilon_{\text{SPD}}(\text{Run1}) &= 0.86 \pm 0.01(\text{stat.}) \pm 0.01(\text{data-simulation overlap}), \\ \Delta\varepsilon_{\text{SPD}}(2015) &= 0.82 \pm 0.02(\text{stat.}) \pm 0.10(\text{data-simulation overlap}), \\ \Delta\varepsilon_{\text{SPD}}(2016) &= 0.76 \pm 0.01(\text{stat.}) \pm 0.09(\text{data-simulation overlap})\end{aligned}$$

where the first uncertainty is obtained from the sample sizes and the second is obtained from the incomplete overlap between data and simulation. Because the trigger strategy is very similar between 2011 and 2012, the numbers are combined for Run 1.

To check whether the wrong description of the efficiency of the  $n_{\text{SPD}}$  requirement has an impact on the result of the analysis, the simulated data sets after the full selection are investigated, as shown in Fig. 5.9 for the 2015–2016 data set. The distributions of signal and normalisation mode are very similar in the SPD hit multiplicity. The difference is quantified by applying the tighter requirement from Run 2 ( $n_{\text{SPD}} < 450$ ) on the Run 1 samples. This results in a difference of  $< 1\%$ . Thus it is confirmed that the wrong calculation of the efficiency of the  $n_{\text{SPD}}$  requirement does not have an effect on the efficiency ratio and a conservative systematic uncertainty of 1% is assigned as systematic uncertainty on that assumption.

Applying the correction factor calculated above, the expected numbers of produced  $B^+$  mesons are compared to the values expected from luminosity and cross-section measurements in Table 5.16. Since the calculations from efficiencies include the fractions of the bremsstrahlung categories, the yield obtained for every category is expected to match the estimation based on cross-section and integrated luminosity individually.

The yields match well between the categories and also with the expected ones from the cross-section measurements. However, the measurements of  $b\bar{b}$  cross-sections are dominated by



**Figure 5.9:** Comparison of the  $n_{\text{SPD}}$  distribution between signal and normalisation mode on simulation of the 2015–2016 data set after the full selection. The nominal reweighting is applied. The differences are small between the two modes.

**Table 5.16:** Overall  $B^+$  yields as determined from efficiencies and yields found in this analysis compared to the expected  $B^+$  yields from the luminosities. The numbers for the bremsstrahlung categories are corrected by their fractions. The uncertainties on the expected  $B^+$  yields combine statistical and systematic uncertainties, while the experimental ones combine only statistical uncertainties and the data-simulation overlap from the detector occupancy correction factor.

$B^+$ Yields [ $10^{11}$ ]	2011	2012	2015	2016
Category 0	$1.72 \pm 0.33$	$4.3 \pm 0.8$	$0.99 \pm 0.17$	$7.6 \pm 1.0$
Category 1	$1.63 \pm 0.30$	$4.3 \pm 0.7$	$0.91 \pm 0.14$	$7.2 \pm 0.9$
Category 2	$1.8 \pm 0.4$	$4.3 \pm 0.8$	$0.87 \pm 0.17$	$7.1 \pm 0.9$
Expected	$2.28 \pm 0.37$	$4.78 \pm 0.57$	$1.13 \pm 0.14$	$6.55 \pm 0.72$

systematical uncertainties, which might well be underestimated. Therefore a different check is performed additionally, normalising the yields to another external measurement.

### 5.5.6 Branching fraction ratio with $B^0 \rightarrow K^+ \pi^-$

An alternative cross-check of the absolute efficiencies uses the branching fraction ratio  $\mathcal{B}(B^0 \rightarrow K^+ \pi^-)/\mathcal{B}(B^+ \rightarrow K^+ J/\psi)$ , which can be computed as

$$\begin{aligned} \frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi)} &= \frac{N_{B^0 \rightarrow K^+ \pi^-}}{N_{B^+ \rightarrow K^+ J/\psi}} \times \frac{\varepsilon_{B^+ \rightarrow K^+ J/\psi}}{\varepsilon_{B^0 \rightarrow K^+ \pi^-}} \times \frac{f_u}{f_d}, \\ &= 0.325 \pm 0.012 \end{aligned} \quad (5.15)$$

**Table 5.17:** Yields and efficiencies for the mode  $B^0 \rightarrow K^+\pi^-$  taken from the  $B_s^0 \rightarrow \mu^+\mu^-$  analysis [73]. The 2016 yield is scaled by 1.66/1.1 to account for the luminosity difference with respect to this analysis.

	2011	2012	2015	2016
Yield [ $10^4$ ]	$0.68 \pm 0.07$	$1.80 \pm 0.12$	$0.86 \pm 0.08$	$4.3 \pm 0.4$
Efficiency [ $10^{-3}$ ]	$1.500 \pm 0.033$	$1.79 \pm 0.04$	$3.27 \pm 0.11$	$2.97 \pm 0.07$

**Table 5.18:** Measured branching fraction ratios split by year and bremsstrahlung category compared to the nominal one. The uncertainties contain statistical and systematic uncertainties.

	ratio (2011)	ratio (2012)	ratio (2015)	ratio (2016)
Category 0	$0.43 \pm 0.09$	$0.39 \pm 0.08$	$0.43 \pm 0.08$	$0.31 \pm 0.05$
Category 1	$0.46 \pm 0.10$	$0.39 \pm 0.07$	$0.47 \pm 0.09$	$0.33 \pm 0.05$
Category 2	$0.41 \pm 0.10$	$0.38 \pm 0.08$	$0.47 \pm 0.10$	$0.33 \pm 0.06$
Expected	$0.325 \pm 0.012$			

where the expected value is taken from the Particle Data Group [21]. The fraction  $\frac{f_u}{f_d}$  is assumed to be unity. Efficiencies  $\varepsilon_{B^0 \rightarrow K^+\pi^-}$  and yields  $N_{B^0 \rightarrow K^+\pi^-}$  for the mode  $B^0 \rightarrow K^+\pi^-$  are used from the previous  $B_s^0 \rightarrow \mu^+\mu^-$  analysis [73], where a similar check was performed. Since the  $B_s^0 \rightarrow \mu^+\mu^-$  analysis used only part of the data from 2016 ( $\approx 1.1 \text{ fb}^{-1}$ ), the corresponding yield of that year is scaled by the luminosity ratio 1.66/1.1. The numbers for  $B^0 \rightarrow K^+\pi^-$  are listed in Table 5.17. Applying the correction factor calculated in the previous subsection Section 5.5.5, the measured branching fraction ratios are shown in Table 5.18. The numbers are in good agreement between the years and with the expected value taken from the PDG.

## 5.6 Background estimation

To be able to measure the signal branching fraction from a fit to the invariant  $B$  mass distribution, the background contributions to the distribution after the full selection have to be estimated as precisely as possible. The background contributions can be classified into several groups. Firstly, there are combinatorial backgrounds, i.e. random particle combinations from different decays that accidentally fit the requirements imposed in the selection chain. These are strongly suppressed by the BDT classifiers. Furthermore there are background classes, where the reconstructed final state particles originate from the same  $b$ -hadron, thus having much more signal-like signatures in the detector: the second class of backgrounds consists of  $B$  decays that contain the same number of final state particles, but at least one of the particles is misidentified. These decays are the most dangerous category of backgrounds, as they peak in the signal region and can fake a signal peak. These decays are suppressed by stringent particle identification requirements and applying the vetoes. The third category of backgrounds is composed of  $b$ -hadron decays which are only partially reconstructed. Additionally one or more final state particles may be misidentified. The decays of this class can to some extent be suppressed by particle identification requirements. In the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  also the applied  $D$  and  $J/\psi$  vetoes efficiently remove resonant partially reconstructed decays and also the  $\text{BDT}_{\text{HOP}}$  selection suppresses partially reconstructed backgrounds as the momentum imbalance of the final state particles is larger in partially reconstructed decays than in signal decays. However, those partially reconstructed backgrounds are expected to peak (far) below the signal region, because of the not reconstructed particles and therefore only constitute to smooth tails in the mass distribution.

The background contributions are specific to the  $B_{(s)}^0 \rightarrow e^+ e^-$  and  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analyses after the full selection due to the different decay topology. Therefore they are discussed separately in the following subsections.

### 5.6.1 Backgrounds for the search for $B_{(s)}^0 \rightarrow e^+ e^-$

In addition to combinatorial backgrounds, in the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  backgrounds through misidentification could arise from  $B \rightarrow hb^{(\prime)}$  decays with  $h, b'$  being a kaon or pion. Background decays which are only partially reconstructed could be semileptonic decays like  $B_{(s)}^0 \rightarrow b^- e^+ \nu_e$  and  $A_b^0 \rightarrow p e^- \bar{\nu}_e$ , where the neutrino is not reconstructed and a hadron is misidentified as electron,  $B \rightarrow X e^+ e^-$  (where  $X = K^+, K^{*0}, K^0, K^{*+}, \pi^+, \pi^0$ , or  $\gamma$  are not reconstructed),  $B_c^+ \rightarrow (J/\psi \rightarrow e^+ e^-) e^+ \nu_e$ , or  $B^+ \rightarrow (\bar{D}^0 \rightarrow b^+ e^- \bar{\nu}_e) e^+ \nu_e$  (or similar) decays.

While backgrounds from random track combinations can empirically be described by a single exponential function as is tested on same-sign distributions ( $m(e^\pm e^\pm)$ ), the other classes of backgrounds discussed above are evaluated on simulation. The expected background

contributions are evaluated relative to the normalisation mode similar to the procedure used for the signal contribution using the formula

$$N_{\text{bkg}} = N_{\text{control}} \times \frac{\varepsilon_{\text{bkg}}^{\text{tot}}}{\varepsilon_{B^+ \rightarrow K^+ J/\psi}^{\text{tot}}} \times \frac{\mathcal{B}_{\text{bkg}}}{\mathcal{B}_{B^+ \rightarrow K^+ J/\psi}} \times \frac{f_{\text{bkg}}}{f_u}. \quad (5.16)$$

The total selection efficiencies  $\varepsilon^{\text{tot}}$  for the normalisation and background modes are here estimated directly on simulation with the exception of particle identification efficiencies, which are evaluated on calibration data with the PIDCalib package [134], as is done for the signal. The branching fractions  $\mathcal{B}$  of background decays and the normalisation decay are taken from the global averages as performed by the Particle Data Group [21]. For some background decays for the search for  $B_{(s)}^0 \rightarrow e^+e^-$ , however, no branching fraction measurements were existing at the time of publication ( $B_s^0 \rightarrow K^+e^-\bar{\nu}_e$ ,  $B^0 \rightarrow \pi^0e^+e^-$ ,  $B^+ \rightarrow \pi^+e^+e^-$  and  $B_s^0 \rightarrow \gamma e^+e^-$ ) and therefore SM predictions [146–149] are used.

In the search for  $B_{(s)}^0 \rightarrow e^+e^-$ , many of the investigated decay modes have a non-negligible contribution to the final data sets. Therefore they have to be estimated carefully and taken into account in the fit to the data. A few corrections are made to the various background sources because of the way information on them are available. These depend on the studied decay:

- **Backgrounds from partially reconstructed decays:**

The background decays involving more than two charged particles in the final state have been simulated requiring that all charged final state particles are inside the acceptance region of the detector. However, this does not model reality, as only the selected final state particles that mimic the  $B_{(s)}^0 \rightarrow e^+e^-$  candidate are required to fall inside the detector acceptance. Therefore a correction factor is determined from generator level simulation, where requiring only the two electrons inside the detector acceptance is compared to requiring all charged final state particles in the detector acceptance. Correction factors between 16 % and 36 % are found and are applied to the selection efficiency.

- **Backgrounds from  $B^0 \rightarrow \pi^0e^+e^-$ :**

For the decay  $B^0 \rightarrow \pi^0e^+e^-$  no simulation has been produced. However, the decay  $B^0 \rightarrow \pi^0e^+e^-$  is considered to have the same topological and detection properties when treated as a background for  $B_s^0 \rightarrow e^+e^-$  decays as the decay  $B^+ \rightarrow \pi^+e^+e^-$ . Therefore all efficiencies are considered similar. Since the decays  $B^0 \rightarrow \pi^0e^+e^-$  and  $B^+ \rightarrow \pi^+e^+e^-$  have not been observed yet, a theoretical estimate is taken from [148]. The contribution is found to be small compared to the other physical backgrounds (see Table 5.19).

- **Backgrounds from  $B^+ \rightarrow \pi^+e^+e^-$  and  $B_c^+ \rightarrow J/\psi (e^+e^-)e^+\nu_e$ :**

The decays  $B^+ \rightarrow \pi^+e^+e^-$  and  $B_c^+ \rightarrow J/\psi (e^+e^-)e^+\nu_e$  have been simulated using a flat phase space model. This is corrected by reweighting the distributions to proper

physical models. The simulation of  $B^+ \rightarrow \pi^+ e^+ e^-$  is reweighted to a model from Ball and Zwicky [150], while  $B_c^+ \rightarrow J/\psi (e^+ e^-) e^+ \nu_e$  is corrected with to a model from Kiselev [151].

- **Backgrounds from  $B^0 \rightarrow \pi^- e^+ \nu_e$  and  $B_s^0 \rightarrow K^- e^+ \nu_e$ :**

These decay modes yield contributions to the backgrounds due to the non-negligible probability of misidentifying pions and kaons as electrons and the relatively high branching fraction. Unfortunately the simulation for these decays uses the Isgur-Weiss form factor model [152], which is known to be inaccurate. Therefore those decays are reweighted in the momentum transfer  $q^2 = |(p_B^\mu - p_{\pi^-}^\mu)(p_{B,\mu} - p_{\pi^-,\mu})|$  to an accurate description of the decay obtained from recent Light Cone Sum Rules calculations [153] for  $B^0 \rightarrow \pi^- e^+ \nu_e$ . However, the effect of this reweighting is found to be negligible and therefore this is dropped for the even smaller  $B_s^0 \rightarrow K^- e^+ \nu_e$  background. Furthermore, the decay  $B_s^0 \rightarrow K^- e^+ \nu_e$  has not been measured yet. Therefore the SM prediction [148] of this decay is used as estimate for the branching fraction.

- **Backgrounds from semileptonic cascade decays involving  $D$ -mesons:**

Decays of the type  $B \rightarrow D(Xe\nu)e\nu$  (with  $X$  denoting a hadron) are studied using the decay  $B_c^+ \rightarrow J/\psi e^+ \nu_e$  as an efficiency proxy. This type of decay is shifted to masses far below the signal mass region, because a hadron and two neutrinos are not reconstructed. However, these decays have rather high branching fractions and the tail of their mass distribution leads to a sizeable contribution in the studied mass range. The shape for these decays is taken from RapidSim [154] samples for the case that  $X$  is a pion, which yields a conservative estimate for the number of candidates in the signal region, since it is the decay, where the least energy is lost. The RapidSim tool allows fast simulation of decays without running the detector geometry and reconstruction step. Therefore the efficiencies and bremsstrahlung fractions have to be estimated from proxies. However, the contribution of these decays to the signal region is close to zero and thus the modelling of these decays has only an impact on the measurement of the  $B_{(s)}^0 \rightarrow e^+ e^-$  branching fractions through the description of the sidebands.

The expected number of remaining background candidates is documented in Tables 5.19 and 5.20 for the  $B_s^0 \rightarrow e^+ e^-$  and  $B^0 \rightarrow e^+ e^-$  signal regions that contain 68 % of the simulated decays, respectively.

It is found that the strongest exclusive background contributions to the signal regions arise mainly from  $B_c^+ \rightarrow J/\psi e^+ \nu_e$  and  $B^0 \rightarrow \pi^- e^+ \nu_e$ , as well as  $A_b^0 \rightarrow p e^- \bar{\nu}_e$  decays. The background contributions sum up to a total number of expected background candidates of

$$N_{\text{excl}, B_s^0 \rightarrow e^+ e^- \text{ region, Run 1}} = 24.91 \pm 3.97(\text{stat.}) \pm 2.72(\mathcal{B}), \quad (5.17)$$

$$N_{\text{excl}, B_s^0 \rightarrow e^+ e^- \text{ region, Run 2}} = 17.61 \pm 3.15(\text{stat.}) \pm 2.96(\mathcal{B}), \quad (5.18)$$

where the first uncertainty is due to the available amount of simulation (stat.) and the second uncertainty is due to the uncertainty on the branching fractions.

**Table 5.19:** Estimated number of exclusive background candidates in the  $B_s^0$  and  $B^0$  signal regions for the Run 1 data set. In the upper part of the table backgrounds from particle misidentification, that are peaking in the signal region, are reported. The decay  $B^0 \rightarrow \pi^0 e^+ e^-$  is evaluated using the decay  $B^+ \rightarrow \pi^+ e^+ e^-$  as a proxy. The branching fractions for the decays are almost everywhere taken from the world averages [21]. Those that are instead taken from SM predictions are marked with a \* and have an additional reference for the prediction. For each value, the first uncertainty reported is due to the limited amount of available simulation and the second due to the precision of the branching fraction estimate.

Decay mode	$N_{(B_s^0 \text{ signal region})}$	$N_{(B^0 \text{ signal region})}$
$B^0 \rightarrow \pi\pi$	$1.63 \pm 0.08 \pm 0.06$	$1.60 \pm 0.08 \pm 0.06$
$B_s^0 \rightarrow KK$	$0.08 \pm 0.01 \pm 0.01$	$0.08 \pm 0.01 \pm 0.01$
$B^0 \rightarrow K\pi$	$1.30 \pm 0.08 \pm 0.04$	$1.31 \pm 0.08 \pm 0.04$
$B_s^0 \rightarrow \pi K$	$0.09 \pm 0.01 \pm 0.02$	$0.09 \pm 0.01 \pm 0.02$
$B^+ \rightarrow K^+ e^+ e^-$	$0.23 \pm 0.12 \pm 0.07$	$0.50 \pm 0.23 \pm 0.11$
$B^0 \rightarrow K^{*0} e^+ e^-$	$0.06 \pm 0.05 \pm 0.01$	$0.15 \pm 0.07 \pm 0.03$
$B^+ \rightarrow \pi^+ e^+ e^-$ [148]	$0.25 \pm 0.03 \pm 0.06$	$0.82 \pm 0.06 \pm 0.21$
$B^+ \rightarrow K^{*+} e^+ e^-$	$0.06 \pm 0.07 \pm 0.02$	$0.14 \pm 0.15 \pm 0.05$
$B^0 \rightarrow K_S^0 e^+ e^-$	$0.08 \pm 0.02 \pm 0.05$	$0.09 \pm 0.03 \pm 0.06$
$B_c^+ \rightarrow J/\psi e^+ \nu_e$	$4.22 \pm 2.14 \pm 1.95$	$12.29 \pm 6.55 \pm 5.81$
$B^0 \rightarrow \pi^- e^+ \nu_e$	$9.04 \pm 1.49 \pm 0.58$	$16.97 \pm 2.13 \pm 1.09$
$A_b^0 \rightarrow p e^- \bar{\nu}_e$	$7.38 \pm 2.99 \pm 1.80$	$10.96 \pm 4.27 \pm 2.68$
$B_s^0 \rightarrow K^- e^+ \nu_e$ [146, 147]	$0.35 \pm 0.07 \pm 0.09$	$0.69 \pm 0.10 \pm 0.17$
$B^0 \rightarrow \pi^0 e^+ e^-$ [148]	$0.11 \pm 0.01 \pm 0.02$	$0.38 \pm 0.03 \pm 0.10$
$B_s^0 \rightarrow \gamma e^+ e^-$ [149]	$0.03 \pm 0.01 \pm 0.01$	$0.05 \pm 0.01 \pm 0.01$
$B^0 \rightarrow D^0 (\rightarrow \pi e \nu) e \nu$	$< 1$	$< 1$
Total	$24.91 \pm 3.97 \pm 2.72$	$46.12 \pm 8.11 \pm 6.49$

**Table 5.20:** Estimated number of exclusive background candidates in the  $B_s^0$  and  $B^0$  signal regions for the Run 2 data set. In the upper part of the table backgrounds from particle misidentification that are peaking in the signal region are reported. The branching fractions for the decays are almost everywhere taken from the world averages [21]. Those that are instead taken from SM predictions are marked with a \* and have an additional reference for the prediction. For each value, the first uncertainty reported is due to the limited amount of available simulation and the second due to the precision of the branching fraction estimate. Since for some small contributions no simulation has been produced, for completeness the numbers from 2012 are included here, scaled by the mass region yield ratios of  $B^+ \rightarrow K^+ e^+ e^-$  (for  $B^+ / B^0 \rightarrow \pi^+ / \pi^0 e^+ e^-$ ) and  $B^+ \rightarrow \pi^- e^+ \nu_e$  (for  $A_b^0 \rightarrow p e^- \bar{\nu}_e$ ). The decays  $B_s^0 \rightarrow K^- e^+ \nu_e$  are not reported here due to their negligible contribution.

Decay mode	$N_{(B_s^0 \text{ signal region})}$	$N_{(B^0 \text{ signal region})}$
$B^0 \rightarrow \pi\pi$	$1.28 \pm 0.15 \pm 0.05$	$1.28 \pm 0.15 \pm 0.05$
$B_s^0 \rightarrow KK$	$0.02 \pm 0.01 \pm 0.01$	$0.02 \pm 0.01 \pm 0.01$
$B^0 \rightarrow K\pi$	$0.39 \pm 0.04 \pm 0.01$	$0.39 \pm 0.04 \pm 0.01$
$B_s^0 \rightarrow \pi K$	$0.03 \pm 0.01 \pm 0.01$	$0.03 \pm 0.01 \pm 0.01$
$B^+ \rightarrow K^+ e^+ e^-$	$0.65 \pm 0.15 \pm 0.08$	$1.54 \pm 0.35 \pm 0.19$
$B^0 \rightarrow K^{*0} e^+ e^-$	$0.17 \pm 0.11 \pm 0.03$	$0.28 \pm 0.18 \pm 0.05$
$B^+ \rightarrow \pi^+ e^+ e^-$ [148]	$0.11 \pm 0.02 \pm 0.03$	$0.32 \pm 0.06 \pm 0.09$
$B^+ \rightarrow K^{*+} e^+ e^-$	$0.10 \pm 0.07 \pm 0.03$	$0.23 \pm 0.16 \pm 0.07$
$B^0 \rightarrow K_S^0 e^+ e^-$	$0.06 \pm 0.02 \pm 0.04$	$0.07 \pm 0.02 \pm 0.03$
$B_c^+ \rightarrow J/\psi e^+ \nu_e$	$3.54 \pm 1.55 \pm 1.97$	$10.30 \pm 5.99 \pm 5.73$
$B^0 \rightarrow \pi^- e^+ \nu_e$	$7.43 \pm 2.14 \pm 1.91$	$15.17 \pm 4.37 \pm 3.90$
$B_s^0 \rightarrow \gamma e^+ e^-$ [149]	$0.14 \pm 0.01 \pm 0.16$	$0.24 \pm 0.02 \pm 0.27$
$A_b^0 \rightarrow p e^- \bar{\nu}_e$	$3.67 \pm 1.71 \pm 1.11$	$4.30 \pm 2.00 \pm 1.32$
$B^0 \rightarrow \pi^0 e^+ e^-$ [148]	$0.02 \pm 0.01 \pm 0.01$	$0.05 \pm 0.03 \pm 0.03$
$B^0 \rightarrow D^0 (\rightarrow \pi e \nu) e \nu$	$< 1$	$< 1$
Total	$17.61 \pm 3.15 \pm 2.96$	$34.12 \pm 7.69 \pm 7.07$

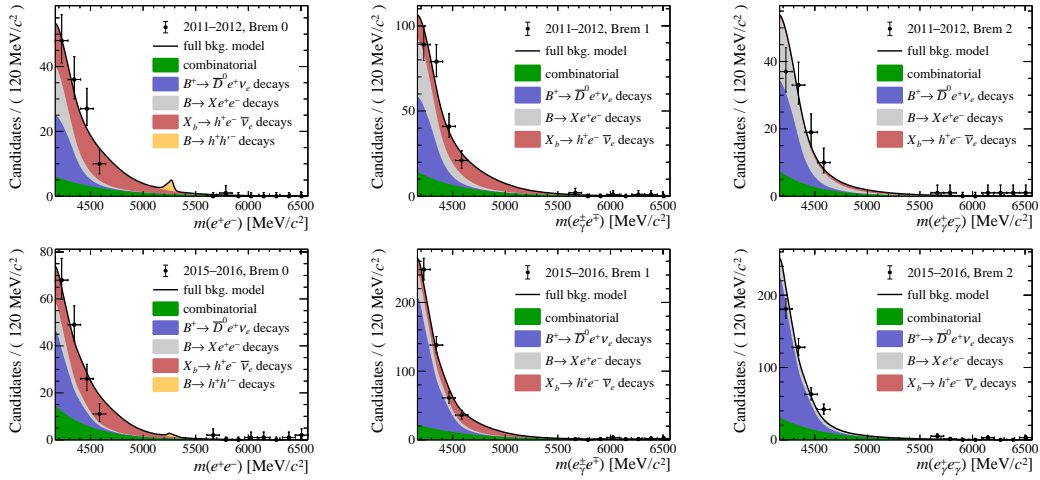


The data sets are split in bremsstrahlung categories (no bremsstrahlung added; one electron received bremsstrahlung correction; both electrons receive bremsstrahlung correction), which are fit independently. This has the advantage that e.g. backgrounds peaking in the signal region like  $B \rightarrow bb'$  only show up in the data set with bremsstrahlung category 0 as their emittance of bremsstrahlung radiation is negligible. The shapes of the background contributions are determined with a kernel density estimator using Gaussian kernels on simulation, separately for each category. Combinatorial background and  $B_c^+ \rightarrow J/\psi(e^+e^-)e^+\nu_e$  are described together from a fit to same-sign data, as same-sign data is expected to be composed of only these two contributions. The yields of the background components are constrained to the estimated yields determined above, where the yield is multiplied by the corresponding bremsstrahlung fraction taken from simulation. A cut-off criterion is chosen that all backgrounds with a yield of  $< 0.1$  in the whole mass region are neglected. For the semileptonic  $D$  cascade decays only RapidSim samples are available, which do not allow to assess the effect of bremsstrahlung. Therefore the same fit function is used for all categories and the bremsstrahlung fractions are obtained from  $B_s^0 \rightarrow e^+e^-$  simulation. The full background PDF ( $\text{PDF}_{\text{Full}}(m, N_{SS}, N_i)$ ) yields

$$\begin{aligned} \text{PDF}_{\text{Full}}(m, N_{SS}, N_i) = & \mathcal{P}(N_{\text{meas}} | N_{SS} + \sum_i N_i) \times \frac{1}{N_{SS} + \sum_i N_i} \\ & \times \left[ N_{SS} \mathcal{G}(N_{SS, \text{meas}}, \sigma_{N_{SS, \text{meas}}} | N_{SS}) \text{PDF}_{SS}(m) \right. \\ & \left. + \sum_i N_i \mathcal{G}(N_{i, \text{meas}}, \sigma_{N_{i, \text{meas}}} | N_i) \text{PDF}_{i, \text{bkg}}(m) \right], \end{aligned}$$

where the exponential function fitted to same sign data is denoted as  $\text{PDF}_{SS}$ , the background PDFs are denoted as  $\text{PDF}_{i, \text{bkg}}$  and the yields, are denoted as  $N_i$ . The extended term to take into account the yields, is a Poisson  $\mathcal{P}$ , while the Gaussian  $\mathcal{G}$  describes the constraint of the yields, where the width is given as the quadrature of the statistical and systematic uncertainties reported in Tables 5.19 and 5.20. The summation is performed over all studied backgrounds. The result of a fit to the mass side bands is displayed in Fig. 5.10 and shows a good agreement with the data. The composition and size of the data sets vary due to different bremsstrahlung fractions of the different background components. The proportions also change from Run 1 to Run 2 due to different performances of particle identification, trigger and BDT algorithms.

It has been studied, whether data-simulation differences have to be taken into account in the background fit. For all shapes of the physical backgrounds that are determined from simulation, the shapes are convoluted with a Gaussian with mean 0 and a variable width. The width is allowed to fluctuate between 0 and the corrected width of the signal as determined in Section 5.7. The difference in the fit between with and without including the resolution convolution for the background components is found to be negligible and therefore this complicated procedure is dropped.



**Figure 5.10:** Fits of the full background model to the mass side bands. The data in the exclusion region are not plotted. The top row corresponds to Run 1 data and the bottom row to the 2015–2016 data set. From left to right, the data sets correspond to the bremsstrahlung correction category with no correction, correcting one electron and correcting both electrons. The small peak in the signal region corresponds to the misidentified  $B \rightarrow hb'$  decays, which emit almost no bremsstrahlung and therefore only show up in the first category. The relative proportions of background contributions change between Run 1 and Run 2 due to different performances of the particle-identification algorithms and BDT selections. Their relative fractions between bremsstrahlung categories follow the expectation from simulation.

The uncertainty on the yields of the various background components in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis is incorporated in the mass description. To understand the impact of those uncertainties, the data side bands are fitted with fixing the yields of the background components to the nominal value reduced and enlarged by one standard deviation. The maximum variation is reported in Table 5.23 together with the other systematic uncertainties and constitute to the major source of systematic uncertainty of the analysis.

### 5.6.2 Backgrounds for the search for $B^+ \rightarrow K^+\mu^\pm e^\mp$

The background composition in the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  analysis is very different from the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis due to the different decay topology of a 3-prong ( $B^+ \rightarrow K^+\mu^\pm e^\mp$ ) and a 2-prong ( $B_{(s)}^0 \rightarrow e^+e^-$ ) vertex. Additional to combinatorial backgrounds, potential background contributions could arise due to misidentification from  $B^+ \rightarrow K^+\ell^+\ell^-$  (where the e.g. the kaon can be identified as electron and a muon as kaon),  $B^+ \rightarrow K^+(J/\psi \rightarrow \ell^+\ell^-)$ , and  $B^+ \rightarrow K^+\pi^+\pi^-$  decays with  $\ell^\pm$  being electrons or muons. The main contributions from decays that are only partially reconstructed comprise of  $B^0 \rightarrow K^{*0}e^+e^-$ ,  $A_b^0 \rightarrow pK^-\ell^+\ell^-$ ,  $A_b^0 \rightarrow pK^-(J/\psi \rightarrow \ell^+\ell^-)$ ,  $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$ , and  $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\ell^+\nu_\ell$  decays.

After evaluating the main potential background modes, in the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  analysis no significant background contribution from a single decay is found compared to the background from random track combinations. Therefore the background is modelled in the invariant  $m(K^+e^\pm\mu^\mp)$  mass by a single exponential function. Since this model is therefore also supposed to describe the very small contributions from other background sources, a systematic uncertainty is estimated on this approach by determining the slope of the exponential at a loosened selection on the BDT and BDT<sub>HOP</sub> values. Refitting the mass side bands of the final data sets yields a difference in the expected background yield in the signal region as shown in Table 5.21, which is assigned as systematic uncertainty to the background yield.

**Table 5.21:** Systematic uncertainty  $\Delta$  for the background contribution estimated as the absolute difference between the nominal fit and a fit fixing the exponential slope from a fit to looser selected data.

Decay mode	Systematic $\Delta$	Nominal fit	Alternative fit
$B^+ \rightarrow K^+\mu^+e^-$	0.60	$3.93 \pm 1.14$	$3.33 \pm 0.69$
$B^+ \rightarrow K^+\mu^-e^+$	0.43	$0.88 \pm 0.63$	$1.30 \pm 0.43$

## 5.7 Evaluation of systematic uncertainties

Several cross-checks are being performed to ensure the robustness of the analysis as reported in this section. Systematic uncertainties on the measurements are evaluated – if not stated differently – separately for each data taking year and for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis also split by bremsstrahlung category.

### Kinematic data-simulation differences

The accuracy of the data-simulation corrections is tested in different ways in the two discussed analyses. Since the efficiencies of the signal mode are not accessible on data, in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the absolute efficiencies of the normalisation mode are compared between data and reweighted simulation for the preselection and BDT selection, where the efficiencies are obtained from reweighted simulation. The differences are evaluated separately for each year and bremsstrahlung category and amount to up to 4 % for the BDT selection, thus providing one of the larger systematic uncertainties. However, the preselection yields with differences lower than 0.6 % good agreement between data and simulation.

In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis a different approach is adopted. The BDT input distributions in the  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  control mode show small discrepancies between data and simulation in the HOP variable and the  $B^+ \chi_{\text{IP}}^2$ . To remove the differences, an additional reweighting is performed in these two variables. The effect is evaluated in the ratio of the BDT and  $\text{BDT}_{\text{HOP}}$  selection efficiencies between signal and normalisation mode. A difference in the ratio between the nominal reweighting strategy and this additional reweighting smaller than 0.6 % is observed, and therefore a conservative systematic uncertainty of 1 % is assigned. Systematic effects could also arise from the electron in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decay due to the reweighting based on the  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  control mode, as electrons behave differently in the detector. This leads to differences in the reconstruction and thus affects the  $B^+ \chi_{\text{vtx}}^2$  variable. Therefore the reweighting is also performed with the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  mode and the difference of the signal-to-normalisation ratio of the BDT and  $\text{BDT}_{\text{HOP}}$  efficiencies is evaluated in a similar manner as before to a value of 1.4 %, which is assigned as additional systematic uncertainty.

### Particle identification efficiencies

The efficiencies of the PID selection are calculated in bins of calibration data. In this procedure two potential sources of systematic effects are identified: the choice of the binning and the  $s$ Plot of the calibration data. In the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the effect of the binning choice was studied by splitting each bin in half in one dimension for all variables that the data are evaluated in separately. The difference with respect to the efficiency with the nominal procedure is evaluated for all binning changes and added in quadrature leading to values of up to 3 %, which are assigned as systematic uncertainty. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis a different approach was followed. Since the BDT distributions might be slightly correlated with the PID variable distributions, the ratio of efficiencies between normalisation and signal mode is investigated for the combination of the efficiencies of the PID and BDT requirements.

Two alternative binning schemes for the PID calibration are investigated by increasing and decreasing the number of bins by a factor of 2. The largest deviation in the efficiency ratio from the nominal procedure is taken as systematic uncertainty and found to be as large as 4.6%. The effect on the efficiencies of the BDT requirements was found to be very small, though.

The  $\mathcal{P}lot$  approach to project out the signal distribution in the calibration data relies on the factorisation of the discriminating variable, the signal mass distribution and the particle's identification variables, its kinematic distributions and the detector occupancy. While this is well justified for the kaon and muon final state particles, where a generous but negligible absolute systematic uncertainty of 0.1% is recommended for the PID efficiency [134], for electrons the assumptions are not valid. The discriminating variable of the calibration data of the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  decay  $m(J/\psi K^+)$  contains strong tails, which are correlated with the momentum of the electrons. This occurs because electrons lose a significant amount of their momentum due to bremsstrahlung in the detector, which is only partly corrected for and sometimes even over-corrected. This is then propagated to the calculation of the invariant mass  $m(J/\psi K^+)$ . Together with a rather high background level this can lead to an imperfect  $\mathcal{P}lot$  projection. Therefore in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the efficiency of the requirement on the electron  $\text{ProbNN}_e$  is also calculated by fitting  $m(J/\psi K^+)$  in the control mode  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  in data before and after the cut in bins of the efficiency calculation procedure. For this study, a tag and probe method is used, where one of the electrons from the  $J/\psi$  (called the tag) has a strong requirement of  $\text{PID}_e > 5$  applied to it, while the other electron (the probe) is used for the efficiency calculation. This procedure is done two times, switching the role of the electrons as tag and probe in between, the resulting efficiencies are averaged to compute the final efficiency in each bin. The efficiency of the selection is then obtained by assigning an efficiency to each event from the control channel simulation corresponding to its kinematic bin and averaging the values to obtain one efficiency. The difference with respect to the nominal efficiency calculation procedure is found to be up to 4.9% and is assigned as systematic uncertainty. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis the estimated systematic uncertainty of 3% was taken from the measurement of  $R_K$  [8], which exploited a similar procedure on the decay mode  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  as signal mode.

### Trigger efficiencies

The trigger efficiencies are obtained with the TISTOS method on data in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis and taken from simulation in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis. In both analyses this approach was cross-checked, as discussed in Section 5.5.1. In the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the TISTOS method is applied to simulation and compared to simple cut-based efficiencies. The difference of up to 5.3% is assigned as systematic uncertainty. This relatively large uncertainty comes from the low number of events per TISTOS bin in the  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  mode. For the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis a similar check is performed, finding that the trigger efficiencies are robust within an uncertainty of 1%, which is assigned as systematic uncertainty. Here

the uncertainty is much smaller, as the  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  decay can be reconstructed much more easily.

### Signal parametrisation

The signal branching fraction is obtained from a fit to the invariant mass distribution of the  $B$  meson. Therefore it is important to know the signal shape in that variable. The line shape is obtained from fitting the signal distribution in simulation with the sum of two Crystal Ball functions for each bremsstrahlung category for the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and a double-sided Crystal Ball function for each data taking period and bremsstrahlung category for the search for  $B_{(s)}^0 \rightarrow e^+ e^-$ , similar to the description of the normalisation modes in Section 5.3.1. However, due to imperfect modelling of the detector response, the mass resolution is slightly different between data and simulation and therefore has to be calibrated on data.

In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis all signal shape parameters  $P$  are predicted using the two control modes,  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  and  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ , by defining

$$P_{e\mu}^{\text{predicted}} = P_{ee}^{\text{data}} + (P_{e\mu}^{\text{MC}} - P_{ee}^{\text{MC}}) \frac{P_{ee}^{\text{data}} - P_{\mu\mu}^{\text{data}}}{P_{ee}^{\text{MC}} - P_{\mu\mu}^{\text{MC}}} \quad (5.19)$$

separately for the categories that no electron has bremsstrahlung correction applied and all electrons have bremsstrahlung correction applied. The results are similar to what was found with a different approach in the analysis of  $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$  decays with the LHCb experiment [155]. Since the mass resolution for final states involving electrons is much different from muonic final states, this approach cannot be followed for the  $B_{(s)}^0 \rightarrow e^+ e^-$  line shape. Therefore in that analysis the tail parameters of the Crystal Ball function are obtained from simulation and only the mean and width of its Gaussian core are calibrated from the dielectron distribution in  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  by defining correction factors of

$$C_\sigma = \frac{\left(\frac{\sigma}{\mu}\right)_{\text{data}}}{\left(\frac{\sigma}{\mu}\right)_{\text{MC}}} \quad \text{and} \quad (5.20)$$

$$C_\mu = \frac{\mu_{\text{data}}}{\mu_{\text{MC}}}. \quad (5.21)$$

The resulting correction factors again yield compatible results as found in the analyses of  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ . To assess the uncertainty on these correction procedures, they are repeated 1000 times with new simulation and data distributions obtained with the bootstrapping method [156]. The parameters from simulation and their corrections are listed in Table 5.22 for both analyses. For each of these signal parametrisations the efficiency of the 90% signal region ( $B^+ \rightarrow K^+ \mu^\pm e^\mp$ ) and 68% signal region are evaluated. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis the efficiency difference between the nominal correction procedure and the mean correction from the bootstrapping is assigned as systematic uncertainty, finding

**Table 5.22:** Signal line shapes as determined on simulation and with the correction procedures applied. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis the parameter  $n$  is fixed from the fit to  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$ . The tail parameter  $\alpha$  is the same for both Crystal Ball functions for the category of no corrected electrons.

Parameter	No correction		One $e$ corrected		All $e$ corrected	
Search for $B^+ \rightarrow K^+ \mu^\pm e^\mp$						
$\mu$ [MeV/ $c^2$ ]	5270.20	$\pm 1.68$			5273.58	$\pm 2.44$
$\mu_{\text{corr}}$ [MeV/ $c^2$ ]	5270.58	$\pm 2.36$			5276.85	$\pm 2.63$
$\sigma$ [MeV/ $c^2$ ]	15.01	$\pm 11.06$			28.54	$\pm 3.35$
$\sigma_{\text{corr}}$ [MeV/ $c^2$ ]	17.30	$\pm 16.73$			34.34	$\pm 4.98$
$\sigma_1$ [MeV/ $c^2$ ]	68.62	$\pm 20.34$			81.38	$\pm 10.94$
$\sigma_{1,\text{corr}}$ [MeV/ $c^2$ ]	90.69	$\pm 31.75$			98.71	$\pm 17.57$
$n$	22.37				8.33	
$\alpha$	0.23	$\pm 0.02$			0.60	$\pm 0.09$
$\alpha_{\text{corr}}$	0.28	$\pm 0.02$			0.66	$\pm 0.10$
$\alpha_1$		–			–2.78	$\pm 0.61$
$\alpha_{1,\text{corr}}$		–			–2.70	$\pm 0.24$
$c$	1.0	$\pm 0.0$			0.681	$\pm 0.071$
$c_{\text{corr}}$	0.998	$\pm 0.007$			0.736	$\pm 0.044$
Search for $B_{(s)}^0 \rightarrow e^+ e^-$ , Run 1						
$\alpha_1$	$0.1516 \pm 0.0008$		$0.4099 \pm 0.0014$		$0.6861 \pm 0.0030$	
$\alpha_2$	$-2.98 \pm 0.22$		$-1.277 \pm 0.005$		$-0.9008 \pm 0.0041$	
$n_1$	$3.69 \pm 0.05$		$3.194 \pm 0.022$		$2.7070 \pm 0.0217$	
$n_2$	$83.28 \pm 124.50$		$2.801 \pm 0.023$		$4.09 \pm 0.05$	
$\mu$ [MeV/ $c^2$ ]	5329.6	$\pm 0.3$	5323.9	$\pm 0.3$	5342.1	$\pm 0.4$
$\sigma$ [MeV/ $c^2$ ]	27.71	$\pm 0.13$	57.43	$\pm 0.18$	68.24	$\pm 0.28$
$\mu$ [MeV/ $c^2$ ]	5329.6	$\pm 0.3$	5323.9	$\pm 0.3$	5342.1	$\pm 0.4$
$\mu_{\text{corr}}$ [MeV/ $c^2$ ]	5328.4	$\pm 2.1$	5328.9	$\pm 1.8$	5352.0	$\pm 2.9$
$\sigma$ [MeV/ $c^2$ ]	27.71	$\pm 0.13$	57.43	$\pm 0.18$	68.24	$\pm 0.28$
$\sigma_{\text{corr}}$ [MeV/ $c^2$ ]	29.6	$\pm 1.1$	57.9	$\pm 1.7$	66	$\pm 6$
Search for $B_{(s)}^0 \rightarrow e^+ e^-$ , Run 2						
$\alpha_1$	$0.1407 \pm 0.0001$		$0.4391 \pm 0.0013$		$0.7608 \pm 0.0003$	
$\alpha_2$	$-2.525 \pm 0.007$		$-1.1944 \pm 0.0004$		$-0.9912 \pm 0.0004$	
$n_1$	$4.351 \pm 0.005$		$2.7285 \pm 0.0014$		$2.7539 \pm 0.0020$	
$n_2$	$9.27 \pm 0.35$		$2.9302 \pm 0.0020$		$3.7232 \pm 0.0037$	
$\mu$ [MeV/ $c^2$ ]	5331.90	$\pm 0.02$	5324.20	$\pm 0.02$	5339.50	$\pm 0.03$
$\mu_{\text{corr}}$ [MeV/ $c^2$ ]	5323.9	$\pm 2.2$	5302.2	$\pm 2.2$	5318.2	$\pm 2.7$
$\sigma$ [MeV/ $c^2$ ]	27.840	$\pm 0.010$	57.696	$\pm 0.016$	76.622	$\pm 0.027$
$\sigma_{\text{corr}}$ [MeV/ $c^2$ ]	32.8	$\pm 1.5$	65	$\pm 4$	91	$\pm 6$

a value of 2.1 %. In the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis instead, the width of the efficiency distribution of the bootstrapped samples is assigned as systematic uncertainty, yielding values between 0.6 % and 1.1 %.

### **Bremsstrahlung fractions**

Another important ingredient to the parametrisation of the signal is the fraction of bremsstrahlung categories. This cannot be measured in the signal mode and therefore is taken from simulation. In the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  fractions of  $f_{\text{no Brem}} = 0.43$  and  $f_{\text{Brem}} = 0.57$  are found and no systematic uncertainty is assigned as it was found to have negligible impact in similar  $B^+ \rightarrow K^+ \ell^+ \ell^-$  analyses, e.g. Ref. [8]. The bremsstrahlung fraction has only a small impact in these kinds of analyses, as it only varies the mass shape slightly. However, in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis the uncertainty on the bremsstrahlung fraction needs to be treated with more care, as the data are split in bremsstrahlung categories. Therefore a check of the impact on the overall normalisation is performed when comparing the bremsstrahlung fractions found in  $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$  data and simulation after the full selection chain. The overall normalisation constants are then compared by combining the values in Table 5.14 across one data-taking period with the same combination, but replacing the bremsstrahlung fractions measured on simulation with the bremsstrahlung fractions measured on data. Differences of 3.6 % (Run 1) and 4 % (Run 2) are found and assigned as systematic uncertainty to the normalisation constant.

### **Track reconstruction efficiencies**

The efficiencies of the very first selection steps are taken from simulation and cannot be cross-checked on data, as the data are not clean enough to find a clear signal in the control modes. While this is not a problem for the selection, as the requirements on the track combinations to form a signal candidate have a very high efficiency, the track reconstruction efficiencies rely on detector conditions, which are not always perfectly modelled in simulation. The track reconstruction efficiencies can be measured on data with a tag and probe method [157] in  $J/\psi \rightarrow \mu^+ \mu^-$  from  $b$ -hadron decays for muons. Since the track reconstruction efficiency depends on the track kinematics, maps of the data-simulation differences in the reconstruction efficiency are created, binning in  $\eta$  and  $p$  of the track. These maps are then convoluted with the kinematic spectrum of the investigated track. As tracks between pions, kaons and muons are similar, the same maps can be used to obtain tracking efficiency corrections for muons and hadrons, adding a systematic uncertainty of 1.1 % to account for hadronic interactions in the detector, which are taken from simulation. A similar method exists for electron tracks [158], which behave significantly different in the detector from other tracks, but only for Run 2 data sets. Since the investigated decay modes are measured relative to a similar normalisation mode, the efficiency corrections mostly cancel. Therefore in the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis no systematic uncertainty is assigned. Only in the  $B_{(s)}^0 \rightarrow e^+ e^-$  the normalisation mode contains an additional kaon track. On this kaon track the reconstruction efficiency correction is determined and found to be unity within the uncertainty of the method and thus only an uncertainty on the hadronic interactions of 1.1 % is assigned as systematic uncertainty.



**Difference between  $B_s^0 \rightarrow e^+e^-$  and  $B^0 \rightarrow e^+e^-$** 

In the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis all efficiencies of the signal mode are measured with  $B_s^0 \rightarrow e^+e^-$  simulation and assumed to be similar between  $B_s^0 \rightarrow e^+e^-$  and  $B^0 \rightarrow e^+e^-$ . The correctness of this assumption is tested by comparing the efficiencies from the 2012 and 2016 data set between  $B^0 \rightarrow e^+e^-$  and  $B_s^0 \rightarrow e^+e^-$  simulation. A small difference of 2.5 % (2012) and 2.1 % (2016) is found and assigned as systematic uncertainty to the  $B^0 \rightarrow e^+e^-$  normalisation for the corresponding data taking period.

Summaries of all systematic uncertainties for the two investigated analyses are given in Tables 5.23 to 5.25, listing also the systematic uncertainties on the background model shape. While all other uncertainties are applied to the normalisation factors, the background uncertainties are assigned to the background yields in the fit to data instead. The uncertainties on the background yields are in both cases the dominant systematic uncertainties. However, due to their nature as searches for yet undiscovered decays, the analyses are dominated by the statistical uncertainties. The expected upper limits have been checked to be robust with increasing the systematic uncertainties even to 20 %

**Table 5.23:** Systematic uncertainties assigned in the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis. Wherever possible, the systematic uncertainty is calculated in each bremsstrahlung category (Brem. cat.) separately, where “0” denotes no bremsstrahlung correction, “1” that one electron receives bremsstrahlung correction and “2” that electron and positron have bremsstrahlung correction applied. The difference between  $B^0$  and  $B_s^0$  only affects the  $B^0$  normalisation.

Procedure	Brem. cat.	Systematic uncertainty [%]			
		2011	2012	2015	2016
Preselection	0	0.34	0.25	0.38	0.35
	1	0.39	0.03	0.30	0.42
	2	0.14	0.37	0.61	0.47
Trigger $B_{(s)}^0 \rightarrow e^+ e^-$	0	2.11	1.93	1.56	2.22
	1	0.70	1.23	0.44	0.29
	2	4.11	5.29	4.15	4.95
Trigger $B^+ \rightarrow K^+ J/\psi$	0	1.42	2.97	5.24	0.06
	1	4.32	4.74	2.52	0.83
	2	3.55	3.42	2.00	0.58
BDT	0	2.22	3.03	4.06	1.37
	1	1.86	1.87	3.23	0.18
	2	2.29	1.30	0.97	0.39
PID (binning)	0	1.37	1.84	0.37	0.41
	1	0.83	3.05	1.62	2.45
	2	0.41	0.69	0.88	1.63
PID ( $_{s} \mathcal{P}lot$ )	0	4.85	3.93	1.97	1.24
	1	4.11	4.31	2.46	0.71
	2	3.05	4.37	1.05	0.18
Mass resolution	0	0.8		0.8	
	1	0.7		0.6	
	2	1.1		1.1	
$n_{\text{SPD}}$ mismodelling	0	1		1	
	1	1		1	
	2	1		1	
Background shape	0	7.91		11.72	
	1	5.56		8.33	
	2	4.61		3.68	
Brem. fraction	all	3.6		4.0	
Kaon tracking	all		1.1		
$B_s^0$ - $B^0$ difference	all	2.5		2.1	

**Table 5.24:** Combined relative systematic uncertainty on the normalisation constants for the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis for the different samples of the final fit. The systematic uncertainties for the difference between  $B^0 \rightarrow e^+e^-$  and  $B_s^0 \rightarrow e^+e^-$  are not included and are added on top of these uncertainties for the  $B^0 \rightarrow e^+e^-$  component.

Brem. category	Run 1 [%]	Run 2 [%]
0	5.82	4.69
1	6.56	4.24
2	6.29	6.23

**Table 5.25:** Summary of systematic uncertainties for the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  analysis. The values are in % unless stated otherwise.

Effect	$B^+ \rightarrow K^+\mu^+e^-$	$B^+ \rightarrow K^+\mu^-e^+$
Data-Simulation corrections	1.0	1.0
Electron-muon differences	1.4	1.4
Fitting model	2.1	2.1
PID corection - binning	3.3	4.6
PID correction - $\mathcal{P}lot$	3	3
Trigger	1.0	1.0
Simulation sample size	1.9	1.8
Background (not in %)	0.60	0.43
Total	5.9	6.7

## 5.8 Results of the searches for $B^+ \rightarrow K^+ \mu^\pm e^\mp$ and $B_{(s)}^0 \rightarrow e^+ e^-$

In an extended maximum likelihood fit to the invariant  $B$ -mass distributions of the fully selected data sets, the  $B_{(s)}^0 \rightarrow e^+ e^-$  and  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  branching fractions are obtained. In these fits, the normalisation factors are Gaussian constrained to the measured values with a width that is calculated from adding up its statistical and systematic uncertainties in quadrature. Since no signal is expected to be found, an upper limit on the branching fractions is prepared with the GAMMACOMBO framework [78, 79], applying the CL<sub>s</sub> method [86], determining the distributions of the one-sided test statistic [81] with pseudo experiments, as discussed in Chapter 3.

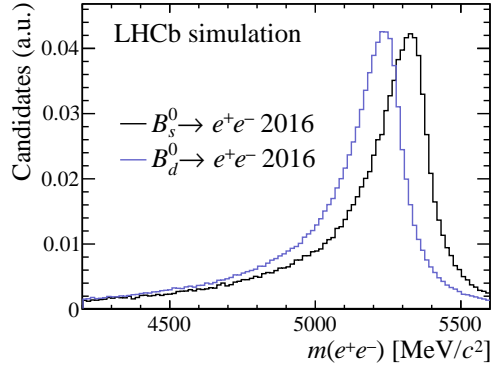
In the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  the number of observed events in the final selected mass region was expected to be very low. With pseudo experiments according to the expected number of observed events, the mass fit procedure was investigated and found to yield a high rate of failed fits, where the signal shape was largely unconstrained. This situation was improved by adding a ghost candidate with a very small weight of 1% at the peak value of the signal shape to the data set. While this ad-hoc approach yields a bias to the measurement of the branching fraction, the procedure is repeated when generating pseudo experiments, thus removing the bias in the limit.

To account for the systematic uncertainty on the background, the background yield is additionally varied in the generation of pseudo experiments, according to the relative systematic uncertainty of the background, thus increasing the spread of the test statistic distributions in the pseudo experiments and increasing the  $p$ -values.

In the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  the branching fractions are obtained from a simultaneous fit to six data sets, as the data are split by data-taking period and bremsstrahlung category. Because the mass resolution of  $B_{(s)}^0 \rightarrow e^+ e^-$  is much larger than the  $B_s^0 - B^0$  mass difference, as demonstrated in Fig. 5.11, in the fit the contributions from  $B_s^0 \rightarrow e^+ e^-$  and from  $B^0 \rightarrow e^+ e^-$  cannot be disentangled. Therefore the mass distribution is fitted assuming either a pure  $B_s^0 \rightarrow e^+ e^-$  contribution or a pure  $B^0 \rightarrow e^+ e^-$  contribution, while the other signal decay is being neglected.

Expected upper limits are calculated by generating pseudo experiments from a fit of the background-only model to the data mass side bands. These result in expected upper limits at 90(95)% CL of

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+ \mu^- e^+) &< 6.9(9.2) \times 10^{-9}, \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^+ e^-) &< 10.2(12.7) \times 10^{-9}, \\ \mathcal{B}(B_s^0 \rightarrow e^+ e^-) &< 7.0(8.6) \times 10^{-9}, \text{ and} \\ \mathcal{B}(B^0 \rightarrow e^+ e^-) &< 2.5(2.5) \times 10^{-9}. \end{aligned}$$



**Figure 5.11:** Invariant-mass distribution  $B_s^0 \rightarrow e^+ e^-$  and  $B_d^0 \rightarrow e^+ e^-$  candidates in simulation of the 2016 data set after the full selection. The distributions overlap strongly. Therefore it is not feasible to estimate the branching fractions of the two contributions at the same time.

As the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis is performed simultaneously on six different data sets, in a step before looking at the final result the consistency of the measurement in all six data sets was investigated: while not printing the resulting branching fraction, each data set was fitted individually and the resulting branching fractions were compared between the data sets. This served as a check to understand whether especially backgrounds from the misidentification of particles are correctly modelled, because those vary strongly between the bremsstrahlung correction categories. It also tested whether the Run 1 and Run 2 measurements are consistent. Because all branching fractions were consistent within 2.5 standard deviations, the final results were calculated.

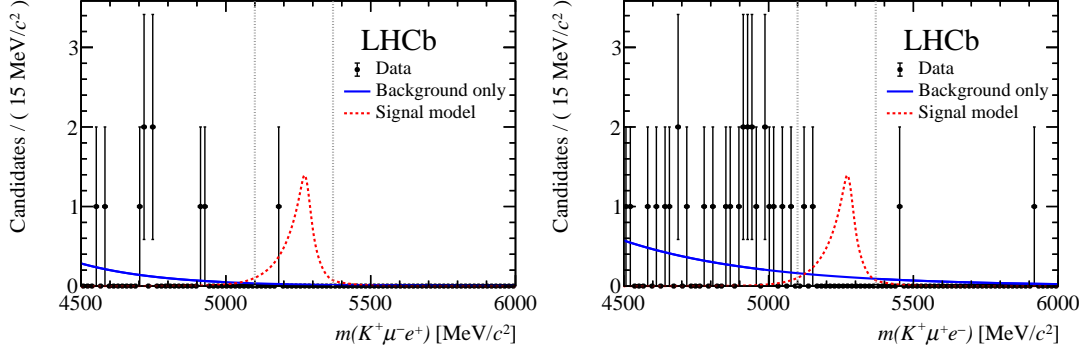
### 5.8.1 Results of the searches for $B^+ \rightarrow K^+ \mu^\pm e^\mp$ decays

In the uncovered signal region, the  $B^+ \rightarrow K^+ \mu^+ e^-$  yields 2 events ( $3.93 \pm 1.14$  expected from background only), while in the  $B^+ \rightarrow K^+ \mu^- e^+$  sample 1 event is observed ( $0.88 \pm 0.63$  expected from background only), well in agreement with background the background-only hypothesis, as is seen in Fig. 5.12. Therefore the observed upper limits are significantly lower (slightly higher) than the expected upper limits in the  $B^+ \rightarrow K^+ \mu^+ e^-$  ( $B^+ \rightarrow K^+ \mu^- e^+$ ) mode and result in

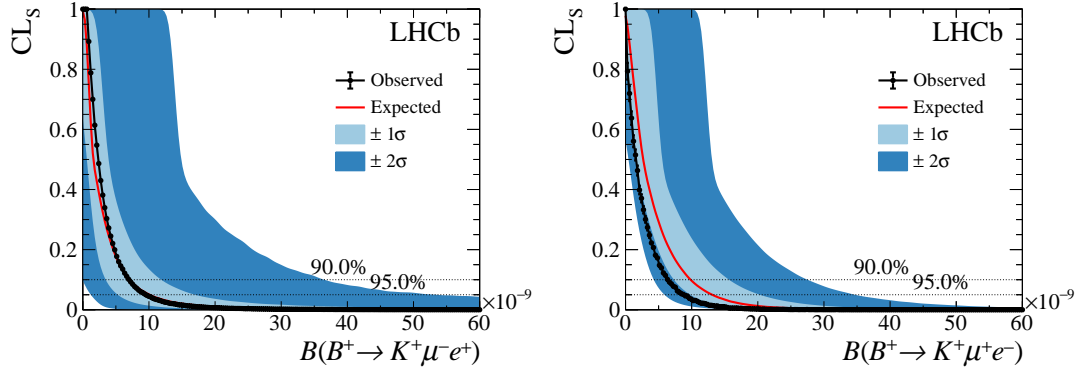
$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+ \mu^- e^+) &< 7.0(9.5) \times 10^{-9} \\ \mathcal{B}(B^+ \rightarrow K^+ \mu^+ e^-) &< 6.4(8.8) \times 10^{-9} \end{aligned}$$

at 90(95) % CL. The corresponding  $\text{CL}_S$ -value curves are shown in Fig. 5.13.

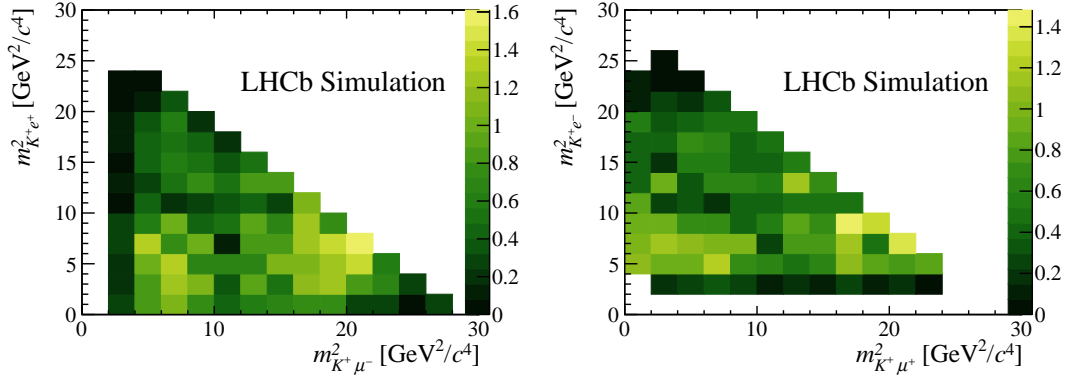
The three-body final state of the decay allows for model-dependent variations in the kinematic distributions of the final state particles, which can have an effect on the overall selection



**Figure 5.12:** Invariant-mass distributions of the (left)  $B^+ \rightarrow K^+ \mu^- e^+$  and (right)  $B^+ \rightarrow K^+ \mu^+ e^-$  candidates obtained on the combined data sets recorded in 2011 and 2012 with background only fit functions (blue continuous line) and the signal model normalised to 10 candidates (red dashed line) superimposed. The signal window is indicated with grey dotted lines. The difference between the two distributions arises dominantly from the effect of the  $m(K^+ \ell^-)$  requirement to veto decays involving  $D$  mesons.



**Figure 5.13:**  $CL_s$  values as a function of the branching fractions of the decays (left)  $B^+ \rightarrow K^+ \mu^- e^+$  and (right)  $B^+ \rightarrow K^+ \mu^+ e^-$ . The red solid line (black solid line with data points) corresponds to the distribution of the expected (observed) upper limits, and the light blue (dark blue) band contains the  $1\sigma$  ( $2\sigma$ ) uncertainties on the expected upper limits. Thresholds corresponding to 90 % and 95 % CL are indicated with dashed lines.



**Figure 5.14:** Efficiency of (left)  $B^+ \rightarrow K^+ \mu^- e^+$  and (right)  $B^+ \rightarrow K^+ \mu^+ e^-$  as function of the squared invariant masses  $m_{K^+e^{-(+)}}^2$  and  $m_{K^+\mu^{+(-)}}^2$ . The variation of efficiency across the Dalitz plane is due to the applied vetoes. The efficiencies are given in per mille.

efficiency, thus potentially lowering or increasing the observed upper limit in the individual decay models. For computing the efficiencies in this analysis, the final state particle kinematics are assumed to be evenly distributed in the Dalitz variables [159]  $m(K\mu)^2$  and  $m(Ke)^2$ . To allow a recasting of the limit, the efficiency distribution is provided across these variables in Fig. 5.14. Convoluting the kinematic distributions of the model with the efficiency distributions, the observed upper limit can be reinterpreted individually for each model. The efficiency maps show also the effect of the vetoes that have been applied.

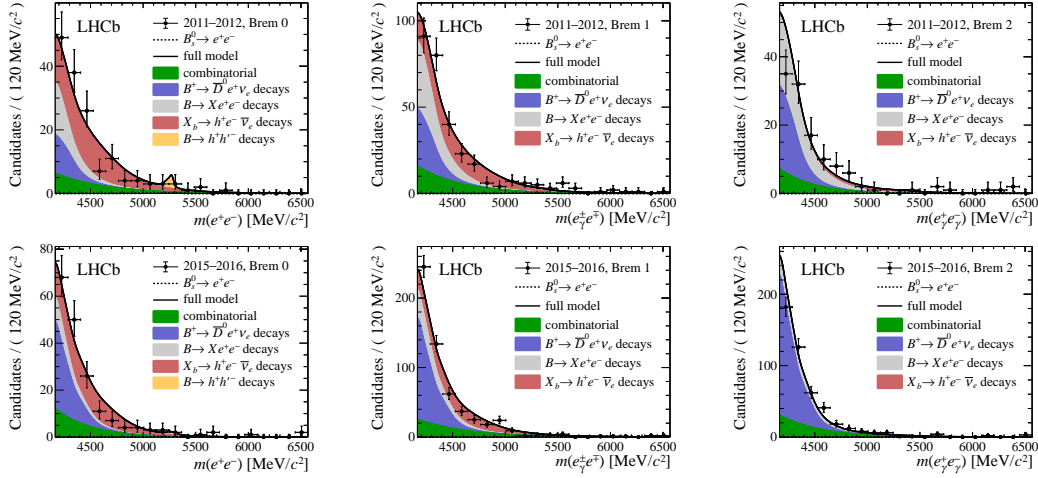
The results obtained in this analysis improve the previous upper limits on the branching fractions of  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays [57] by more than an order of magnitude and impose strong constraints on several scenarios for physics beyond the SM, like models involving leptoquarks and new heavy bosons, discussed in Refs. [51–54]. These scenarios predict branching fractions of up to  $\mathcal{O}(10^{-8})$ . Since the measurements are statistically dominated, with including the data collected by the LHCb in Run 2 and future data taking periods, even stronger constraints will be possible, as discussed in Chapter 7 (if no  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays will be discovered). Although the models from Refs. [51–54] will not completely be ruled out, the possible coupling strength of the proposed interactions will be greatly decreased, which would disfavour those models.

### 5.8.2 Results of the searches for $B_{(s)}^0 \rightarrow e^+ e^-$ decays

The fits to the  $m(e^+e^-)$  mass after uncovering the signal region yield

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow e^+e^-) &= (2.4 \pm 4.4) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow e^+e^-) &= (0.30 \pm 1.29) \times 10^{-9}, \end{aligned}$$

## 5 Searches for $B^+ \rightarrow K^+ \mu^\pm e^\mp$ and $B_{(s)}^0 \rightarrow e^+ e^-$



**Figure 5.15:** Simultaneous fit to the dielectron invariant-mass distribution in all categories, with  $\mathcal{B}(B^0 \rightarrow e^+ e^-)$  fixed to zero. The top figures show the three bremsstrahlung categories in the Run 1 data set and the bottom figures show the Run 2 data set. From left to right, the data sets correspond to the bremsstrahlung correction category with no correction, correcting one electron and correcting both electrons.

where the uncertainties comprise statistical and systematic uncertainties. Thus the results are consistent with zero, following the prediction of the SM. The uncertainties on the  $\mathcal{B}^0 \rightarrow e^+ e^-$  branching fraction are smaller with respect to the  $\mathcal{B}_s^0 \rightarrow e^+ e^-$  measurement by about the factor  $f_s/f_d$ , which is by far the dominant difference between the two measurements. The mass distributions of all investigated data sets are shown in Fig. 5.15 for the fit using the  $B_s^0 \rightarrow e^+ e^-$  hypothesis and in Fig. 5.16 for the fit using the  $B^0 \rightarrow e^+ e^-$  hypothesis.

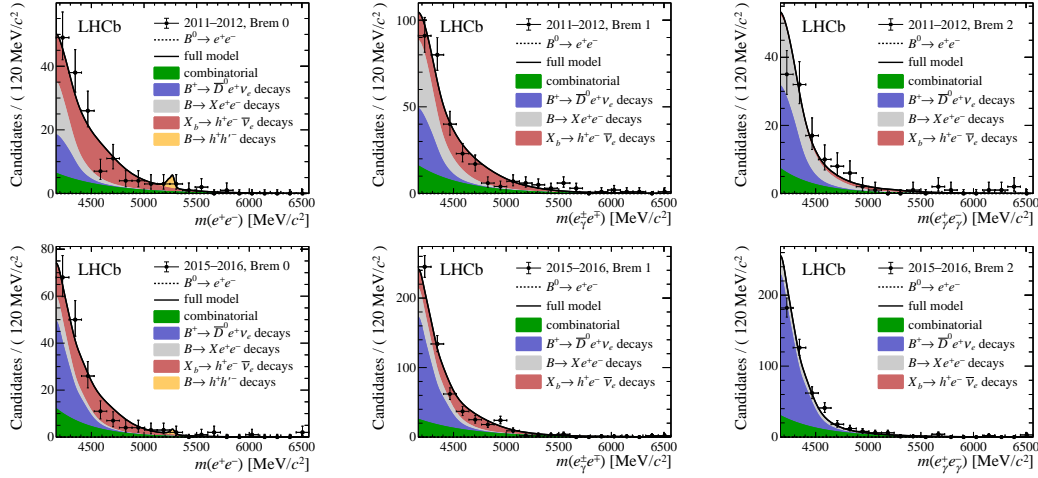
The measurements are translated into upper limits of

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow e^+ e^-) &< 9.4(11.2) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow e^+ e^-) &< 2.5(3.0) \times 10^{-9},\end{aligned}$$

at 90(95) % CL, which are slightly higher than the expected upper limits, as the measured branching fraction is above zero. The corresponding CL<sub>s</sub> curves are shown in Fig. 5.17.

Due to the sizeable decay width difference between the heavy and light mass eigenstate of the  $B_s^0$  meson, its lifetime varies significantly with the  $CP$  parameter  $\mathcal{A}_{\Delta\Gamma}$ . While in the SM the decay proceeds purely through the  $CP$ -odd heavy mass eigenstate, i.e.  $\mathcal{A}_{\Delta\Gamma} = +1$ , in models beyond the SM other  $CP$  admixtures might be realised, up to  $\mathcal{A}_{\Delta\Gamma} = -1$ . Therefore corrections to the result using the values of  $\mathcal{A}_{\Delta\Gamma} = \{-1; 0; 1\}$  are computed, following a strategy described in Ref. [160]. The decay-time-dependent acceptance shape  $\varepsilon(\tau)$  is ex-





**Figure 5.16:** Simultaneous fit to the dielectron invariant-mass distribution in all categories, with  $\mathcal{B}(B_s^0 \rightarrow e^+ e^-)$  fixed to zero. The top figures show the three bremsstrahlung categories in the Run 1 data set and the bottom figures show the Run 2 data set. From left to right, the data sets correspond to the bremsstrahlung correction category with no correction, correcting one electron and correcting both electrons.

tracted by dividing the decay time distribution by the theoretical model. This is then used to determine the correction factor

$$\delta = \frac{\varepsilon^{\mathcal{A}_{\Delta\Gamma}}}{\varepsilon^{MC}} \quad (5.22)$$

$$= \frac{\int_0^\infty (R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t}) \varepsilon(t) dt}{\int_0^\infty (R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t}) dt} \times \frac{\int_0^\infty e^{-\Gamma_{MC} t} dt}{\int_0^\infty e^{-\Gamma_{MC} t} \varepsilon(t) dt}, \quad (5.23)$$

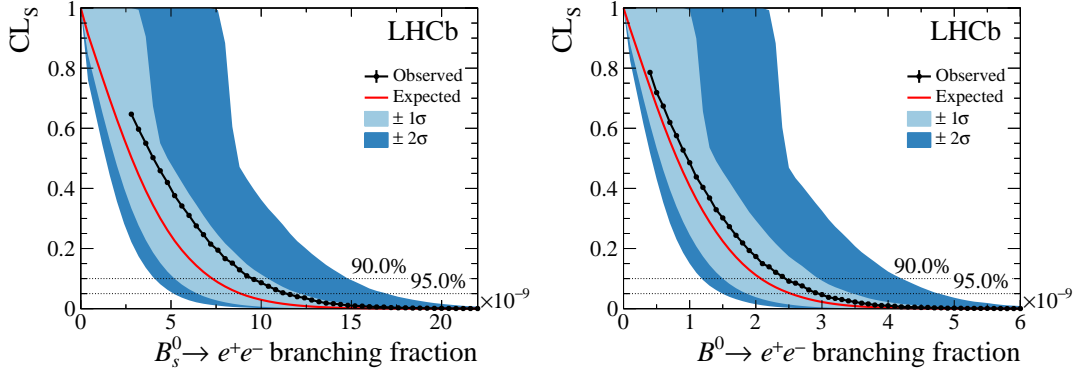
where  $\Gamma_{H/L}$  are the widths and  $R_{H/L}$  the fractional contributions by the heavy/light mass eigenstate, defining  $\mathcal{A}_{\Delta\Gamma}$  as

$$\mathcal{A}_{\Delta\Gamma} = \frac{R_H - R_L}{R_H + R_L}. \quad (5.24)$$

This leads to a correction factor of  $\mp 2.4\%$  for the scenarios  $\mathcal{A}_{\Delta\Gamma} = \pm 1$  with respect to the results quoted above, which correspond to an equal admixture of heavy and light eigenstate  $\mathcal{A}_{\Delta\Gamma} = 0$ . As the decay width differences between light and heavy eigenstate are much smaller for  $B^0$  mesons than for  $B_s^0$  mesons, the effect of different  $CP$  eigenstates can be neglected for  $B^0 \rightarrow e^+ e^-$ .

To investigate the correlation of the measurements of  $\mathcal{B}(B_s^0 \rightarrow e^+ e^-)$  and  $\mathcal{B}(B^0 \rightarrow e^+ e^-)$  further, a simultaneous fit to the data including both components is performed, finding

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow e^+ e^-) &= (5 \pm 7) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow e^+ e^-) &= (-0.85 \pm 2.17) \times 10^{-9}, \end{aligned}$$



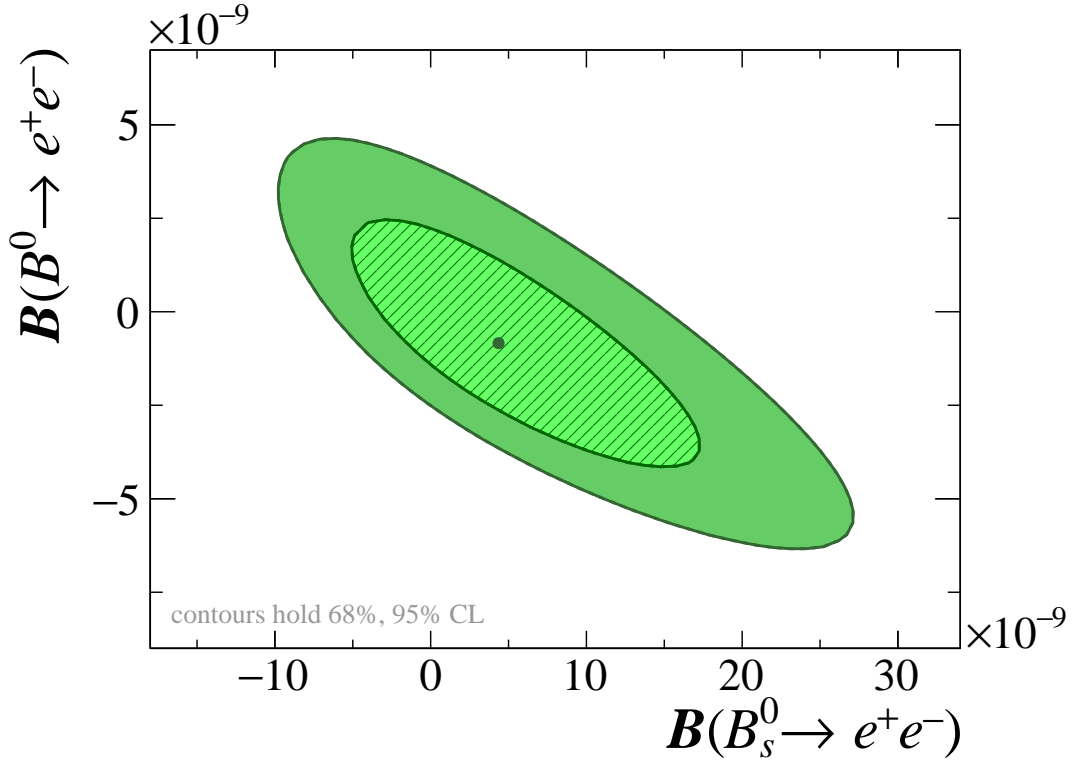
**Figure 5.17:**  $CL_s$  values as a function of the branching fractions of the decays (left)  $B_s^0 \rightarrow e^+ e^-$  and (right)  $B^0 \rightarrow e^+ e^-$ . The red solid line (black solid line with data points) corresponds to the distribution of the expected (observed) upper limits, and the light blue (dark blue) band contains the  $1\sigma$  ( $2\sigma$ ) uncertainties on the expected upper limits. Thresholds corresponding to 90 % and 95 % CL are indicated with dashed lines. The observed values are plotted for branching fractions greater than the measured branching fraction in the data; the test statistic is defined to be nonzero only in that region.

with a correlation of

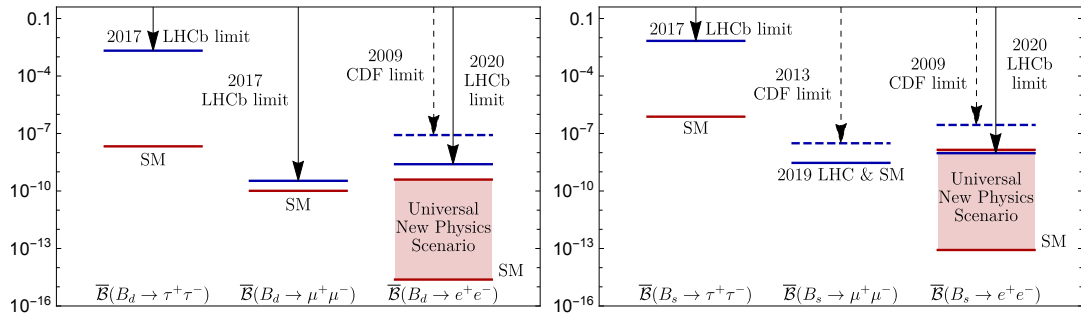
$$\rho[\mathcal{B}(B_s^0 \rightarrow e^+ e^-), \mathcal{B}(B^0 \rightarrow e^+ e^-)] = 79.0\%.$$

This is visualised in the two-dimensional confidence contours in Fig. 5.18, where the p-values are computed with asymptotic formulae [81] using the likelihood ratio test statistic Eq. (3.1). This check proves again that because of the very high correlation, it is not feasible to determine  $\mathcal{B}(B_s^0 \rightarrow e^+ e^-)$  and  $\mathcal{B}(B^0 \rightarrow e^+ e^-)$  at the same time and therefore the upper limits are computed independently.

The limits computed in this analysis are about 30 times smaller than previous limits on these decays and for the first time constrain the allowed region of universal scenarios for physics beyond the SM in the scalar and pseudo-scalar sector [68], as is seen in Fig. 5.19. With analysis improvements and more data, as discussed in Chapter 7, even stronger limits will be possible, further controlling the potential enhancements of scalar and pseudo-scalar currents in  $b \rightarrow s \ell^+ \ell^-$  transitions.



**Figure 5.18:** Two-dimensional profile likelihood contours of the  $B(B_s^0 \rightarrow e^+ e^-)$  and  $B(B^0 \rightarrow e^+ e^-)$  measurement containing the 68 % and 95 % confidence regions. The two branching fraction measurements are strongly correlated.



**Figure 5.19:** Illustration of the (left)  $B^0 \rightarrow \ell^+ \ell^-$  and (right)  $B_s^0 \rightarrow \ell^+ \ell^-$  branching fractions that compares the new measurement discussed in this chapter with the SM predictions and (pseudo-)scalar New Physics scenarios [68]. The new measurement reaches the space allowed by the New Physics scenarios.



## 6 Analysis of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays with the full LHCb data set

The analysis of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays aims at the measurement of their branching fractions, as well as the determination of the effective lifetime of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays. Contrary to the analyses discussed in Chapter 5, which have not been performed with the LHCb experiment before. The last update [73] studied  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  with about half the current data set and investigated the optimal selection for the analysis, the analysis is well established. Therefore the current analysis with the full Run 1+2 data set concentrates on the review of data-simulation corrections as well as the description of background components, as the uncertainties due to those parts were large in the previous analysis. Additionally to the measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions, the analysis now includes a search for the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decay, which could become visible as a shoulder in the  $m(\mu^+ \mu^-)$  mass sideband below the signal region [161]. As the analysis has many contributors, the focus in this thesis is laid only on the parts of the analysis that I mainly contributed to. These comprise the early validation of 2017 and 2018 data, the estimation of background components, in particular backgrounds from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays, as well as the fit to the invariant dimuon mass and the extraction of the branching fraction measurements. Therefore this chapter only briefly discusses the selection strategy in Section 6.1, while focussing on the validation of the newly studied data sets. The most important aspects in the correction of data-simulation differences are highlighted in Section 6.2 and the normalisation is reviewed in Section 6.3. Studies of background components in the extraction of the result are summarised in Section 6.4 with a detailed discussion of the background from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays. Finally, the chapter closes with the description of the fit model and the determination of the expected results of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions in Section 6.5. The measurement of the effective lifetime of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and the search for  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decays, which are performed in the same analysis, are beyond the scope of this thesis and are therefore not discussed. As the analysis is currently still being reviewed by the LHCb collaboration, the final results are not shown in this thesis. An outlook on the remaining parts to finalise the measurement is discussed also in Section 6.5.

## 6.1 Analysis strategy and validation of the 2017 and 2018 data sets

The previous analysis studied the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays with data collected by the LHCb experiment in Run 1, 2015 and part of 2016 [73], amounting to an integrated luminosity of  $4.4 \text{ fb}^{-1}$ . In this analysis round the rest of the data collected in 2016, adding  $0.57 \text{ fb}^{-1}$  to the previously analysed  $1.1 \text{ fb}^{-1}$ , and data collected in 2017 and 2018 are added to the total data set, corresponding to  $9 \text{ fb}^{-1}$ . While the Run 1 data are left untouched, the Run 2 data are reanalysed, meaning that the selection efficiencies and distributions are redetermined and recalibrated with new sets of simulation.

The branching fractions are measured relative to two normalisation modes with precisely known branching fractions,  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  and  $B^0 \rightarrow K^+ \pi^-$ . This has the advantage that  $B^0 \rightarrow K^+ \pi^-$  decays have the same decay topology as  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays, whereas the more abundantly reconstructed  $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$  decay has two muons in the final state and inhibits therefore a very similar behaviour with respect to  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  in the trigger and the PID selection.

Since the selection is almost the same as in the previous analysis [73], it is only summarised here and described in more detail in Appendix A. It consists of an experiment-wide preselection on top of which some loose requirements are added on track and vertex qualities, as well as vertex displacement from the PV and fiducial kinematic requirements with the aim to reduce the data set to a manageable size while keeping a high signal efficiency. A loose requirement on a preselection BDT is applied with 92% signal efficiency to remove a large fraction of background before fine-tuning the selection.

Additionally, a veto against  $B_c^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \mu^+ \nu_\mu$  decays is applied. It requires the invariant-mass combination of a signal muon and any other non-signal track in the event to be outside the  $J/\psi$  mass region, where the non-signal track is identified as a muon. This veto has a very high signal efficiency of 99.7% to 99.8% and reduces the background from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays by 67%.

A most inclusive trigger selection is applied to select  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B^+ \rightarrow K^+ J/\psi$  decays in order to maximise the efficiency for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. This leads to an efficiency of the trigger system of 96% for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays and 90% for  $B^+ \rightarrow K^+ J/\psi$  decays. The difference arises dominantly from kinematic differences between the muons from  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and the muons from  $B^+ \rightarrow K^+ J/\psi$ . To have a signature in the trigger as similar as possible to the signal  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays,  $B^0 \rightarrow K^+ \pi^-$  candidates are selected by requiring an inclusive TIS selection at the L0 and HLT1 stages and requiring TOS on topological two-body trigger lines at the HLT2 stage. Contrary to the previous analysis the inclusive TIS selection is modelled as a combination of exclusive trigger selections to facilitate the efficiency calculation. This strategy yields an efficiency of 4% in Run 1 and 7% in Run

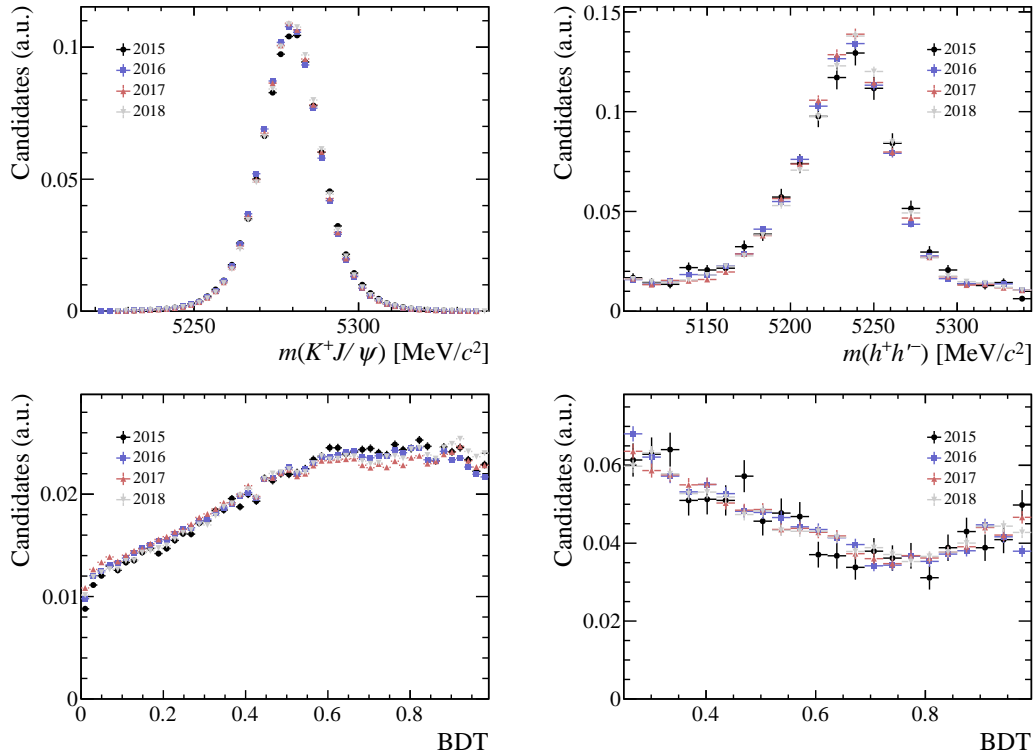
2. The strong difference between the Run 1 and Run 2 efficiencies is due to a thorough reoptimisation of the trigger requirements for Run 2 [96, 112].

To suppress backgrounds from particle misidentification, stringent PID requirements are imposed on the muons with an efficiency of 81 % (Run 1) and 84 % (Run 2). The selection has been chosen to optimise the  $B^0 \rightarrow \mu^+\mu^-$  signal sensitivity [162]. The discrimination of muons from pions and kaons is limited by the fact that pions and kaons can decay into muons while flying through the detector. The muons resulting from the decays in flight are hardly distinguishable from muons produced directly in  $b$ -hadron decays.

The final selection is applied by constructing a selection BDT, which relies on the decay topology of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  and track isolation variables, where the track isolation variables are again BDT classifiers with the aim to provide a measure of how well the signal tracks are separated from other tracks in the event. The isolation variables have been developed specifically for the analysis of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays [162]. No requirements are imposed on the selection BDT. Instead the BDT output distribution is transformed to be flat in a range between 0 and 1 in signal simulation (backgrounds from random muon track combinations accumulate at low BDT values) and then the final branching fraction measurement is obtained by simultaneously fitting the dimuon mass distribution in bins of the flat BDT output distribution. The simultaneous fit of multiple BDT bins allows to exploit the distribution of the BDT and thus enhance the sensitivity with respect to selecting only one region of BDT output distribution.

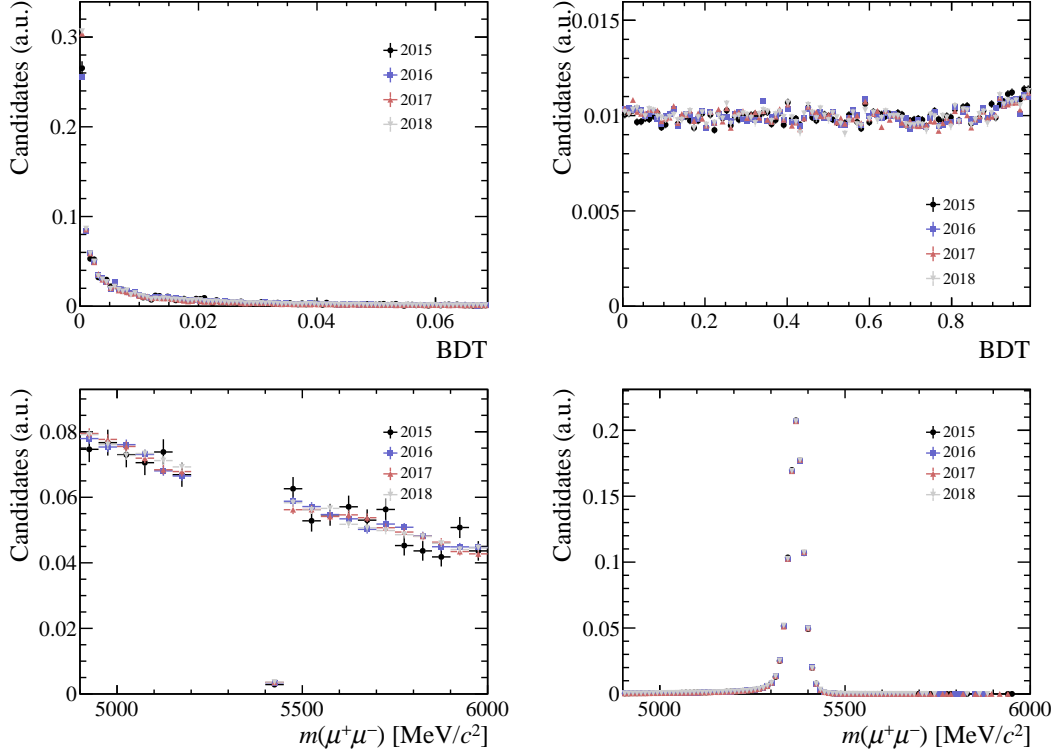
To check whether this strategy, developed for a subset of the data taken in 2011–2016, can be applied to the newly added data taken in 2017 and 2018 as well, the distributions of the selection variables are compared for the Run 2 data sets. The distributions of the invariant mass and the selection BDT are shown in Fig. 6.1 for the normalisation modes and Fig. 6.2 for the signal mode, proving that the data are consistent throughout Run 2. Very small differences are visible in Fig. 6.2 in the 2017 and 2018 data sets with respect to the 2015 and 2016 data sets due to small reconstruction changes and small trigger adjustments. For example the parametrisation of hit uncertainties in the VELO has been changed, leading to a slightly different track reconstruction. However, these differences have negligible effects on the signal and normalisation modes.

Between the previous and the current analysis, the 2015 and 2016 data sets have been reprocessed in an experiment-wide campaign. Comparing the reprocessed samples with the samples used in the previous analysis, some small differences are found. Because the previous analysis was published shortly after taking its data, part of its quality evaluating procedure had not been completed. Comparing the samples after the experiment-wide preselection leads in the comparison with the reprocessed samples to a small loss of about 3 % in 2015 data, because some subsamples had been reevaluated as “bad” very late and subsequently removed. On the contrary some increase of 17 % is found in 2016 data due to later quality evaluation. However, these differences are checked to affect only candidates that are anyhow removed after applying the preselection and therefore do not affect the analysis. The  $B^+ \rightarrow K^+J/\psi$  yields match



**Figure 6.1:** Comparison of the fully selected invariant mass distributions (top) and the BDT distribution (bottom) of  $B^+ \rightarrow K^+ J/\psi$  events (left) and  $B^0 \rightarrow K^+ \pi^-$  events (right) for all Run 2 data taking years. While the BDT distribution of  $B^+ \rightarrow K^+ J/\psi$  is background subtracted with the  $\mathcal{P}lot$  method, the background is suppressed for the  $B^0 \rightarrow K^+ \pi^-$  distributions by requiring  $BDT > 0.25$ . The normalisation modes are very consistent between the years.





**Figure 6.2:** Comparison of the BDT (top) and mass distributions (bottom) for fully selected dimuon sideband data (left) and the  $B_s^0 \rightarrow \mu^+\mu^-$  simulation (right) of all Run 2 data taking years. The data are consistent across Run 2, as reproduced in simulation. Small differences in the BDT distributions of the dimuon sidebands are visible for the 2017 and 2018 data sets with respect to the 2015 and 2016 data sets. The differences arise from small trigger adjustments and small reconstruction changes.

**Table 6.1:** Yields of the normalisation modes compared over the data taking years of Run 2 with statistical and systematic uncertainties combined, assuming an uncertainty of 5 % for the luminosities per year, since the luminosities are uncalibrated. The values in square brackets denote the values from the previous analysis [73], where only a subset of the 2016 data was used. The yields are consistent with the luminosity increase with the exception of 2015, where the trigger and PID algorithms are different. A stricter trigger selection is used for  $B^0 \rightarrow K^+ \pi^-$  with respect to the previous analysis, leading to lower  $B^0 \rightarrow K^+ \pi^-$  yields.

Year	$N_{B^0 \rightarrow K^+ \pi^-}$ [ $10^3$ ]	$N_{B^0 \rightarrow K^+ \pi^-} / \text{fb}^{-1}$ [ $10^3$ ]	$N_{B^+ \rightarrow K^+ J/\psi}$ [ $10^5$ ]	$N_{B^+ \rightarrow K^+ J/\psi} / \text{fb}^{-1}$ [ $10^5$ ]
2015	$4.43 \pm 0.14$	$13.4 \pm 0.8$	$1.676 \pm 0.005$	$5.08 \pm 0.25$
[2015]	$8.6 \pm 0.8$	$26.1 \pm 2.7$	$1.667 \pm 0.004$	$5.05 \pm 0.25$
2016	$23.7 \pm 0.6$	$14.2 \pm 0.8$	$10.369 \pm 0.015$	$6.21 \pm 0.31$
[2016]	$28.4 \pm 2.5$	$25.8 \pm 2.6$	$6.843 \pm 9.000$	$6.22 \pm 0.31$
2017	$24.3 \pm 0.5$	$14.2 \pm 0.8$	$10.820 \pm 0.014$	$6.33 \pm 0.32$
2018	$27.5 \pm 0.6$	$12.6 \pm 0.7$	$13.208 \pm 0.015$	$6.03 \pm 0.30$

very well after the preselection between the previous analysis and the current analysis. The  $B^0 \rightarrow K^+ \pi^-$  and  $B^+ \rightarrow K^+ J/\psi$  yields after the preselection are listed in Table 6.1, normalised to the luminosity of the years and compared to the previous analysis. A good agreement is found between the previous analysis and the current one in the  $B^+ \rightarrow K^+ J/\psi$  mode and also a good agreement throughout Run 2 is found. However, notably the  $B^0 \rightarrow K^+ \pi^-$  yields are smaller in the current analysis, as for this channel the trigger selection has been modified. This improved the fit stability and allowed a more precise efficiency estimate and thus smaller systematic uncertainties. A small decrease in the normalisation yields per  $\text{fb}^{-1}$  is found in 2018 due to an adjustment of trigger thresholds in the L0 trigger. The differences are taken into account in the calculation of the efficiencies.

## 6.2 Corrections for data-simulation differences

This section outlines the various corrections that are applied to simulation to take into account data-simulation differences.

Kinematic and reconstruction effects are corrected by reweighting simulation in  $p_T$ ,  $\eta$  and  $\chi_{\text{IP}}^2$  of the  $B$ -meson candidate with a Gradient Boosting reweighting technique [133] similar to the method described in Section 5.3.2 for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis. In order to account for possible differences in the mismodelling of the  $B_s^0$  and  $B^0$  hadronisation, different reweightings for  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  are applied. To correct  $B^0 \rightarrow \mu^+ \mu^-$  simulation, the  $B^+ \rightarrow K^+ J/\psi$  mode is used (assuming a similar hadronisation of  $B^0$  and  $B^+$ ), while for

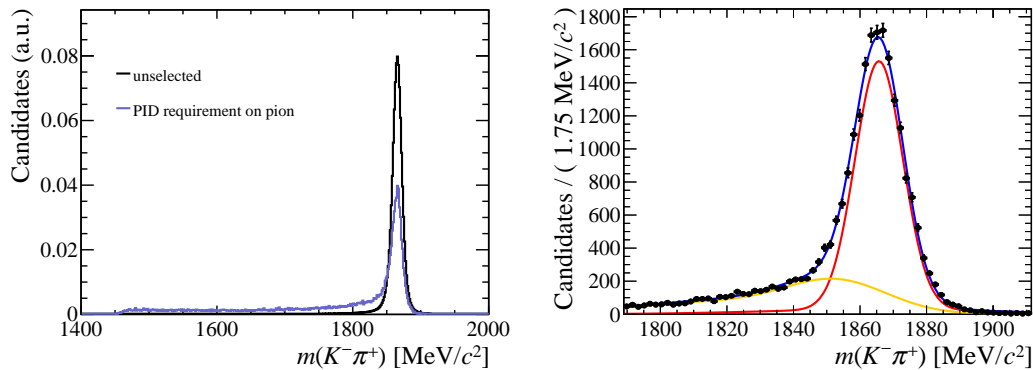
$B_s^0 \rightarrow \mu^+\mu^-$  the mode  $B_s^0 \rightarrow J/\psi \phi$  is exploited, selected as similar as possible to the signal decay.

The track reconstruction efficiency has been measured on data for muons [157] relative to the reconstruction efficiency on MC in maps of the track kinematics and can be adapted to pions and kaons as well, as they have similar track properties. By convoluting the kinematic distributions of the tracks with the correction maps, the total correction factor for each track per sample is obtained. Although many uncertainties cancel out in the normalisation, uncertainties on the correction from the limited knowledge of the amount of material interaction of pions (1.4%) and kaons (1.1%) in the detector do not completely cancel due to the additional kaon in  $B^+ \rightarrow K^+ J/\psi$  and the different particle types with respect to the signal in  $B^0 \rightarrow K^+ \pi^-$  decays.

Trigger efficiencies are computed with the TISTOS method [143] on  $B^+ \rightarrow K^+ J/\psi$  in a similar way as in the  $B_{(s)}^0 \rightarrow e^+e^-$  analysis discussed in Section 5.5.1 and in that way are obtained directly from data. The efficiencies are measured in bins of the maximum muon  $p_T$  and the product of the  $p_T$  of the two muons and are convolved with the distributions of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  simulation to obtain the signal trigger efficiencies. Since the trigger efficiencies are slightly correlated with the selection BDT, the method described above is performed in bins of the BDT classifier. To determine the trigger efficiencies for  $B^0 \rightarrow K^+ \pi^-$ , a similar strategy is adopted for the L0 and HLT1 trigger stages where the trigger relies on signatures in the remainder of the event due to the TIS requirement. Therefore for the application of the TISTOS method, the efficiencies are binned in  $p$  and  $\eta$  of the  $B$  candidate, which are stronger correlated with the remainder of the event. The trigger efficiencies for the HLT 2 stage, which requires TOS for topological trigger lines, are taken directly from simulation, as the decay topology is found to be described well in simulation.

PID efficiencies are determined from data calibration samples with the PIDCalib package [134] in bins of the selection BDT, similar to the approach discussed previously in Section 5.3.3. However, in the course of this analysis the estimate of  $\pi-\mu$  and  $K-\mu$  misidentification has been found to be strongly biased for small misidentification probabilities with this approach. The dominant reason for the misidentification as muons is that pions and kaons can decay into muons while flying through the detector, which has the following effects, as illustrated in Fig. 6.3: on the one hand their reconstructed momentum will be significantly different from the original hadron momentum, thus leading to a large tail in the  $D^0$  invariant-mass in the  $D^0 \rightarrow K^- \pi^+$  calibration sample, which is not described by the fit and correlates the  $D^0$  invariant-mass with the hadron momenta. Therefore the *Plot* method [131], that is used in the PIDCalib package to extract pure signal momentum and  $p_T$  distributions, does not hold anymore. On the other hand a relevant fraction of the large tail is reaching out of selection range of the calibration sample and is thus unaccounted for.

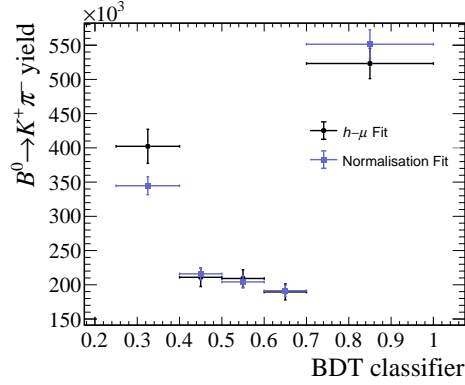
The two effects are corrected for with two different procedures: the correlation between  $D^0$  invariant-mass and momentum is accounted for by determining the misidentification efficiencies in each momentum bin from a fit to the calibration data before and after applying



**Figure 6.3:** Invariant-mass distribution of  $D^0 \rightarrow K^- \pi^+$  decays from high statistics simulation. Calibration data from  $D^0 \rightarrow K^- \pi^+$  decays are used to determine misidentification efficiencies. The left plot shows the distribution without PID requirements (black) and with PID requirements on the pion (blue). A large tail due to decays in flight becomes visible with the PID requirement, which reaches out of the mass window of PIDCalib:  $m(K^- \pi^+) \in [1825, 1910]$   $\text{MeV}/c^2$ . The right plot shows a fit to the mass distribution where the pion track must pass the muon PID selection. The shape of the red component is included in the PIDCalib approach, while the yellow component is the additional tail arising from pion decays in flight and is not modelled in the PIDCalib methods.

the PID selection instead of using the  $\mathcal{P}lot$  method. In these fits the large tail of the signal distribution is taken into account. The amount of decays reaching out of the selection window is determined using large simulated samples. The two procedures lead to an increase of the misidentification efficiency by up to 70 %, while reducing the overall uncertainty of the procedure from more than 50 % in the previous analysis to less than 10 %.

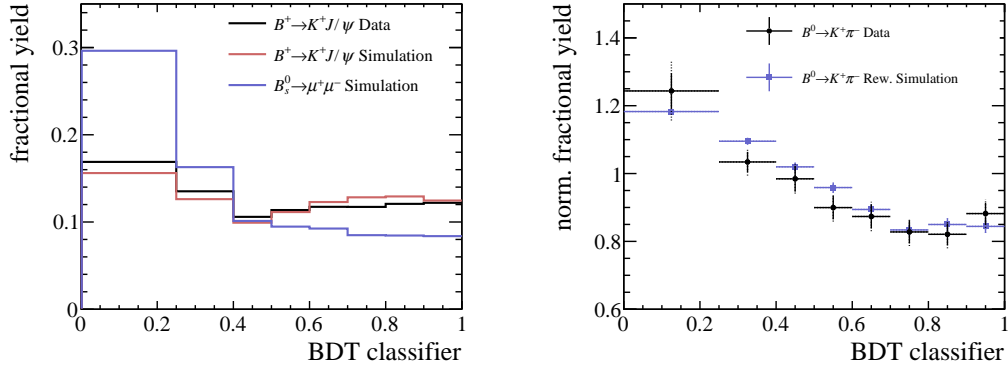
The validity of the corrections is verified with  $B^0 \rightarrow K^+ \pi^-$  decays in an independent check on data: candidates of  $B^0 \rightarrow K^+ \pi^-$  are selected with one of the tracks fulfilling the muon PID criterium of the signal selection and being selected by muon triggers and the other being either identified as a kaon or a pion. Its yield is then corrected by the hadron–muon misidentification estimates with the methods sketched above and compared to  $B^0 \rightarrow K^+ \pi^-$  candidates where one track is identified as a kaon and the other track is identified as a pion, which are used for the normalisation of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. The comparison is given in Fig. 6.4 for Run 2 and shows good agreement between the two approaches, validating the misidentification efficiency determination described above. The uncertainty of this check is assigned as systematic uncertainty to the misidentification efficiencies. Since the PID efficiency is correlated to the selection BDT, the efficiencies and validations are binned in the BDT response. The excellent description of the misidentification probabilities is a major improvement with respect to the previous analysis and a crucial ingredient to the signal sensitivity, as  $B \rightarrow hb^{(\prime)}$  decays with  $b = \pi, K$  are the main source of background peaking close to the signal region.



**Figure 6.4:**  $B^0 \rightarrow K^+ \pi^-$  events of the full Run 2 data set as reconstructed from the  $h - \mu$  fit (black), corrected by misidentification probabilities, and reconstructed from the normalisation selection (blue). The two distributions match very well.

To model the fraction of events falling into each bin of the selection BDT correctly, further data–simulation corrections have to be applied. The BDT uses observables related to event and decay topology and its modelling in simulation can be compared to data via the  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ K^-$  modes, which have the same decay topology as  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ . Again the  $B_s^0$  and  $B^0$  decays are treated separately to correctly take into account potentially different mismodellings of the hadronisation process. It is found that the simulation is not sufficiently corrected using the reweighting discussed in the beginning of this section, mainly because the isolation variables correlate the BDT distribution with the remainder of the event, which is not perfectly described in simulation due to reasons discussed in Section 5.3.2. Therefore an additional two-dimensional binned reweighting is performed in the nTracks variable (describing the track multiplicity in the event) and the BDT distribution itself by comparing data and simulation in the  $B^+ \rightarrow K^+ J/\psi$  mode, as shown exemplary for the 2018 data sample in Fig. 6.5, where the dimuon combination is used as a proxy for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  topology. These weights are then applied to the  $B^0 \rightarrow K^+ \pi^-$ ,  $B_s^0 \rightarrow K^+ K^-$  and  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  modes, proving good agreement between data and corrected simulation in the  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ K^-$  modes, as also demonstrated in Fig. 6.5. Therefore the BDT fractions for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  are taken from reweighted  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  simulation, which reduces the uncertainty to about 1% compared to the previous analysis, where the BDT fractions had been taken from  $B^0 \rightarrow K^+ \pi^-$  recorded data, yielding uncertainties of about 10%.

Finally, for the correct modelling of the  $B_s^0 \rightarrow \mu^+ \mu^-$  efficiencies, the  $CP$  structure has to be taken into account due to the large difference in the decay widths and therefore lifetimes between the heavy and the light eigenstate of the  $B_s^0$  meson. While in simulation an average lifetime of heavy and light eigenstates is assumed, in the SM only the heavy eigenstate can decay into a dimuon pair ( $\mathcal{A}_{\Delta\Gamma} = +1$ ). As the selection, especially the final BDT, is correlated



**Figure 6.5:** BDT distributions on background-subtracted data and simulation for the year 2018. The left plot compares  $B_s^0 \rightarrow \mu^+ \mu^-$  (blue) and  $B^+ \rightarrow K^+ J/\psi$  simulation (red) against background-subtracted  $B^+ \rightarrow K^+ J/\psi$  data (black). Although the distributions are slightly different from  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  due to a different decay topology and muon kinematics, the difference between simulation and data can be reweighted in the  $B^+ \rightarrow K^+ J/\psi$  mode and the weights then applied to  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ K^-$  simulation. The right plot compares the resulting normalised reweighted  $B^0 \rightarrow K^+ \pi^-$  BDT distribution in simulation (blue) against data (black). Statistical uncertainties are marked with solid error bars, while systematic errors are added with dashed bars. Within the uncertainties, the distributions are consistent.

to the decay time of the  $B_s^0$  meson, BDT dependent correction factors are determined in a similar way as discussed in Section 5.8.2 to interpret the result in the SM scenario.

The mass distributions of the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  components are described by double Crystal Ball functions (similar to the model described in Section 5.3.1) with their parameter values calibrated from data: the mean of the Gaussian cores is determined from  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ K^-$  decays and the width is interpolated from fits to  $c\bar{c}$ - and  $b\bar{b}$ -resonances. The tail parameters are determined from simulation, which is smeared with the resolution found with the fit to the  $c\bar{c}$ - and  $b\bar{b}$ -resonances. A separate set of parameters is used for Run 1 and Run 2 and correction factors are applied to take into account small BDT dependences of the signal shape.

## 6.3 Normalisation

The yield of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays,  $N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$ , obtained from a fit to the dimuon invariant-mass, is converted into a measurement of branching fractions via the formula

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{f_d}{f_{d(s)}} \frac{\overbrace{1}^{\alpha_{d(s)}}}{\sum_{y \in \text{Run}} \langle \beta^{(y)} \rangle \varepsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}^{(y)}} \times N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}, \quad (6.1)$$

where  $\alpha_{d(s)}$  denotes the single event sensitivity for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  in each Run. The factor  $\langle \beta^{(y)} \rangle$  is summed for all data taking years in one Run and defined as the uncertainty-weighted average of

$$\begin{aligned} \beta_{B^+ \rightarrow K^+ J/\psi} &= \frac{N_{B^+ \rightarrow K^+ J/\psi}}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi) \times \varepsilon_{B^+ \rightarrow K^+ J/\psi}} \text{ and} \\ \beta_{B^0 \rightarrow K^+ \pi^-} &= \frac{N_{B^0 \rightarrow K^+ \pi^-}}{\mathcal{B}(B^0 \rightarrow K^+ \pi^-) \times \varepsilon_{B^0 \rightarrow K^+ \pi^-}}. \end{aligned} \quad (6.2)$$

The normalisation decay yields  $N_{B^+ \rightarrow K^+ J/\psi}$  and  $N_{B^0 \rightarrow K^+ \pi^-}$  are obtained from fits to selected data, as reported earlier in Table 6.1. Signal and normalisation efficiencies  $\varepsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$ ,  $\varepsilon_{B^+ \rightarrow K^+ J/\psi}$ , and  $\varepsilon_{B^0 \rightarrow K^+ \pi^-}$  are obtained from simulation and calibrated on data as sketched in Section 6.2 and further discussed in this chapter. Another important ingredient and dominant systematic uncertainty for the measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fraction is the hadronisation fraction ratio  $f_s/f_d$  of  $b$ -hadrons, for which a combination of existing measurements by the LHCb collaboration is currently being developed [163] to reduce its uncertainty from 5% to 3.3%. As discussed for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis in Section 5.5.4, also this combination manifests that  $f_s/f_d$  is dependent on the centre-of-mass energy, and thus different values will be used for the Run 1 normalisation than for the Run 2 normalisation. However, as the combination is not ready by the time of handing in this thesis, the results quoted in this chapter use the most precise single measurement, yielding  $f_s/f_d = 0.244 \pm 0.012$  [164]. Since it is known to hold to good accuracy [49],  $f_u = f_d$  is assumed, as in the analyses discussed before. Further relevant inputs to the normalisation are the normalisation mode branching fractions  $\mathcal{B}(B^+ \rightarrow K^+ J/\psi) = (6.00 \pm 0.16) \times 10^{-5}$  (including  $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$ ) and  $\mathcal{B}(B^0 \rightarrow K^+ \pi^-) = (1.96 \pm 0.05) \times 10^{-5}$ , which are taken from the world averages computed by the PDG group [21]. Next to  $f_s/f_d$ , the precision of the normalisation branching fractions constitute also to a relevant systematic uncertainty.

The signal and normalisation efficiencies are calculated split in selection stages as

$$\varepsilon = \varepsilon_{\text{Acc}} \times \varepsilon_{\text{RecSel|Acc}} \times \varepsilon_{\text{Trig|RecSel}}. \quad (6.3)$$

**Table 6.2:** Summary table of the single event sensitivities and expected signal yields for Run 1, Run 2 and combined. The single event sensitivities are also given separately for the two normalisation modes. A good agreement between the two normalisation modes is found.

	$B^0 \rightarrow \mu^+ \mu^-$	$B_s^0 \rightarrow \mu^+ \mu^-$
Run 1		
$\alpha_{B^+ \rightarrow K^+ J/\psi}$	$(2.90 \pm 0.11) \times 10^{-11}$	$(1.06 \pm 0.07) \times 10^{-10}$
$\alpha_{B^0 \rightarrow K^+ \pi^-}$	$(2.80 \pm 0.12) \times 10^{-11}$	$(1.01 \pm 0.08) \times 10^{-10}$
$\alpha_{\text{combined}}$	$(2.87 \pm 0.09) \times 10^{-11}$	$(1.05 \pm 0.07) \times 10^{-10}$
$N_{\text{exp}}$	$3.58 \pm 0.20$	$34.8 \pm 2.6$
Run 2		
$\alpha_{B^+ \rightarrow K^+ J/\psi}$	$(8.9 \pm 0.3) \times 10^{-12}$	$(3.22 \pm 0.21) \times 10^{-11}$
$\alpha_{B^0 \rightarrow K^+ \pi^-}$	$(8.4 \pm 0.4) \times 10^{-12}$	$(3.07 \pm 0.22) \times 10^{-11}$
$\alpha_{\text{combined}}$	$(8.75 \pm 0.19) \times 10^{-12}$	$(3.19 \pm 0.19) \times 10^{-11}$
$N_{\text{exp}}$	$11.8 \pm 0.6$	$114.9 \pm 8.4$
All		
$\alpha_{B^+ \rightarrow K^+ J/\psi}$	$(6.78 \pm 0.22) \times 10^{-12}$	$(2.47 \pm 0.16) \times 10^{-11}$
$\alpha_{B^0 \rightarrow K^+ \pi^-}$	$(6.5 \pm 0.3) \times 10^{-12}$	$(2.36 \pm 0.16) \times 10^{-11}$
$\alpha_{\text{combined}}$	$(6.71 \pm 0.13) \times 10^{-12}$	$(2.45 \pm 0.15) \times 10^{-11}$
$N_{\text{exp}}$	$15.4 \pm 0.8$	$149.7 \pm 10.8$

The efficiency of the detector acceptance  $\varepsilon_{\text{Acc}}$  is evaluated on simulation. Similarly, the reconstruction and selection efficiencies are obtained from simulation, but corrected for data-simulation differences in the track reconstruction efficiencies and the PID efficiencies. Trigger efficiencies are obtained directly from data using the TISTOS method [143] for the  $B^+ \rightarrow K^+ J/\psi$  mode and are transferred to the  $B^0 \rightarrow K^+ \pi^-$  and  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  modes as outlined in Section 6.2. The resulting single event sensitivities are given in Table 6.2 and converted into expected signal yields by multiplying with the SM values of the branching fractions [16].

Several cross-checks of the correctness of the normalisation are performed, the most important one being the branching fraction ratio measurement of

$$\frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi)} = \frac{N_{B^0 \rightarrow K^+ \pi^-}}{N_{B^+ \rightarrow K^+ J/\psi}} \times \frac{\varepsilon_{B^+ \rightarrow K^+ J/\psi}}{\varepsilon_{B^0 \rightarrow K^+ \pi^-}} \times \frac{f_u}{f_d} \quad (6.4)$$

measured independently for each year. This is especially important, as both  $B^0 \rightarrow K^+ \pi^-$  and  $B^+ \rightarrow K^+ J/\psi$  enter the normalisation, but have different trigger and particle identification strategies. The values obtained for each year are listed in Table 6.3 and show excellent consistency over all years of data taking. The combination is found to be consistent with respect to



**Table 6.3:** Ratio of the branching fractions of  $B^+ \rightarrow K^+ J/\psi$  and  $B^0 \rightarrow K^+ \pi^-$  decays (including the sub-decay  $J/\psi \rightarrow \mu^+ \mu^-$ ) obtained from data and compared with the ratio of the world average values (PDG) [21].

Year	$N_{B^+ \rightarrow K^+ J/\psi} [10^5]$	$N_{B^0 \rightarrow K^+ \pi^-} [10^3]$	$\varepsilon_{B^+} / \varepsilon_{B^0}$	$\frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi)}$
PDG				$0.326 \pm 0.012$
2011	$3.479 \pm 0.008$	$3.73 \pm 0.12$	$32.4 \pm 1.2$	$0.347 \pm 0.017$
2012	$7.780 \pm 0.012$	$10.32 \pm 0.20$	$25.4 \pm 0.7$	$0.337 \pm 0.011$
2015	$1.676 \pm 0.005$	$4.43 \pm 0.14$	$12.9 \pm 0.5$	$0.341 \pm 0.016$
2016	$10.369 \pm 0.015$	$23.68 \pm 0.34$	$15.2 \pm 0.5$	$0.347 \pm 0.012$
2017	$10.820 \pm 0.014$	$24.3 \pm 0.4$	$15.2 \pm 0.5$	$0.342 \pm 0.011$
2018	$13.208 \pm 0.015$	$27.78 \pm 0.33$	$16.2 \pm 0.5$	$0.341 \pm 0.011$
All				$0.342 \pm 0.005$

the ratio of the world averages [21] at  $1.1\sigma$ . Although this level of compatibility is already excellent, one point has to be considered in this check: the ratio of hadronisation fractions is assumed to be  $f_u/f_d = 1$  for the world averages of the branching fractions used in this check, which use measurements of the Belle and BaBar experiments. While there is no such measurement of  $f_u/f_d$  in LHCb, the HFLAV average of the Belle and BaBar experiments for the production of two charged over two neutral  $B$  hadrons is  $f^{+/-}/f^{00} = 1.059 \pm 0.027$  [49]. It is expected to deviate from unity, since in those experiments  $B$  mesons are produced from  $\Upsilon(4S)$  decays, which results in small differences in the phase space of  $B^+$  and  $B^0$  mesons. Correcting by this factor would bring the combined value of all data taking years from 0.342 to 0.323, in perfect agreement with the ratio of the world averages. In the normalisation with respect to the  $B_s^0$  mesons however, it is not clear whether the  $f_u$  value should be increased or the  $f_d$  value decreased, so the  $f^{+/-}/f^{00}$  average from Ref. [49] is not used. In that way the full normalisation, being the weighted average of the two channels, covers for any possible difference between LHCb with its hadronic  $B$ -meson production and the Belle and BaBar experiments, where  $B$  mesons are produced from  $\Upsilon(4S)$  decays.

The dominant systematic uncertainties in the normalisation arise from the limited knowledge of the hadronisation fraction ratio  $f_s/f_d$  of about 5% (which will decrease using the ongoing combination) and the knowledge of the normalisation branching fractions of 2.7%, since  $B^+ \rightarrow K^+ J/\psi$  dominates the normalisation. Furthermore the precision of track reconstruction efficiency estimates and PID efficiency estimates are limited to 1% each, yielding subdominant systematic uncertainties.

## 6.4 Background estimation

In the final mass fit several components are left additional to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays, which are described and computed individually:

- $B \rightarrow hb^{(\prime)}$  ( $h, h' = K, \pi$ ) decays, with both kaon and pion misidentified as muons, which peak in the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  signal region;
- $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$  decays, with the pion misidentified as a muon, which contribute to the left mass sideband only;
- $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$  decays, with the kaon misidentified as a muon, also contributing to the left mass sideband only;
- $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  and  $B^0 \rightarrow \pi^0 \mu^+ \mu^-$  decays, with two true muons in the final state, also these only appear in the left mass sideband;
- $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$  decays, with the proton misidentified as a muon, which mainly appears in the left mass sideband but has a tail extending to the  $B^0$  and  $B_s^0$  mass regions;
- $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays, with two true muons in the final state, which dominantly appears in the left mass sideband but has a tail extending almost to the full mass region;
- $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decays, with two muons in the final state and an undetected photon, which contributes in the left part of the mass sideband only with a shape very similar to  $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$  decays. The decay has not been observed yet and corresponds to a very small contribution, which was considered negligible in previous analyses. However, according to SM calculations [149, 165] it might be visible in the data. Therefore it will be included as an extra measurement in the analysis. While the normalisation is not fully calibrated by the time of handing in this thesis, the component is included in the fit to test its influence on the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions.
- Combinatorial background, which comes from random combinations of muon tracks in the event. Its size strongly depends on the BDT bin and is determined directly in the fit. The shape is modelled with an exponential function for each BDT bin.

Backgrounds from  $B \rightarrow hb^{(\prime)}$  decays are the most dangerous class of backgrounds as they peak in the signal dimuon mass region, close to the expected  $B^0 \rightarrow \mu^+ \mu^-$  peak. They are estimated from the selected  $B^0 \rightarrow K^+ \pi^-$  events used for the normalisation. On top of those BDT dependent misidentification and trigger efficiency corrections are applied. As highlighted in Section 6.2 a significant improvement of the misidentification efficiency determination is gained with respect to the previous analysis.

All other backgrounds except combinatorial background are estimated from simulation, where the hadron–muon misidentification probabilities are estimated from calibration data with the PIDCalib package [134] and the improved approaches discussed in Section 6.2. The

mass distributions of all exclusive background decays are modelled by fitting the simulation with Gaussian kernel density estimators.

In particular the background  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  was studied in the course of this thesis leading to significant improvements in the precision of its estimate and is therefore described in more detail here. Partially reconstructed decays of the  $B_c^+$ -meson can be background to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  signal due to the larger mass. However, the hadronisation fraction of a  $b$  quark to a  $B_c^+$ -meson is about two orders of magnitude lower than to a  $B^0$ -meson, though with large uncertainties [166]. The decay  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  (with  $J/\psi \rightarrow \mu^+ \mu^-$ ) could survive the signal selection in case a good vertex is reconstructed between the muon from the semileptonic decay and the oppositely charged muon from the  $J/\psi$ . While the track isolation variables allow the BDT to reject those decays quite well, additionally a simple  $J/\psi$ -veto is applied to further reduce this component. The cut consists of vetoing events in which one of the two candidate muons, coupled to any other oppositely-charged muon in the event (selected with  $\text{ProbNN}_\mu > 0.3$ ), falls in a window of  $|m(\mu^+ \mu^-) - m(J/\psi)| < 30 \text{ MeV}/c^2$ . This veto is expected to reject 68.7% (65.7% in Run 2) of  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  events (in the whole BDT range), with a negligible signal loss of 0.2% in Run 1 and 0.3% in Run 2. All other exclusive backgrounds are unaffected by the veto as well.

The branching fraction of this decay including the hadronisation fraction ratio has recently been measured [166] to be

$$\frac{f_c}{f_u + f_d} \times \mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu) = (7.07 \pm 0.15 \pm 0.24) \times 10^{-5} \text{ for 7 TeV}$$

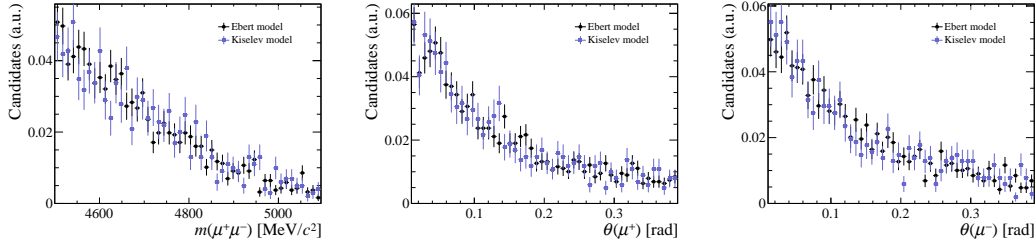
and

$$\frac{f_c}{f_u + f_d} \times \mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu) = (7.36 \pm 0.08 \pm 0.30) \times 10^{-5} \text{ for 13 TeV}$$

in the kinematic region of the  $B_c^+$ -meson  $4 \text{ GeV}/c < p_T < 25 \text{ GeV}/c$  and  $2.5 < \eta < 4.5$ . Assuming  $f_u = f_d$  and  $f_c/f_u$  similar for 7 TeV and 8 TeV, one can then estimate the effective branching fraction of the whole decay chain as

$$\begin{aligned} \mathcal{B}_{\text{eff}} &= 2 \frac{f_c}{f_u + f_d} \times \mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \\ &= (8.43 \pm 0.34) \times 10^{-6} \text{ for Run 1 and} \\ &= (8.8 \pm 0.4) \times 10^{-6} \text{ for Run 2,} \end{aligned}$$

reducing the uncertainty of the estimate from 21% in the previous analysis to below 5%. A correction factor accounting for the ratio of acceptances between  $B^+ \rightarrow K^+ J/\psi$  and  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  under the kinematic selection performed in Ref. [166] is calculated from generator level simulation without acceptance cuts. It is found to be consistent with one ( $\alpha = 1.001 \pm 0.008$ ) and thus dropped.



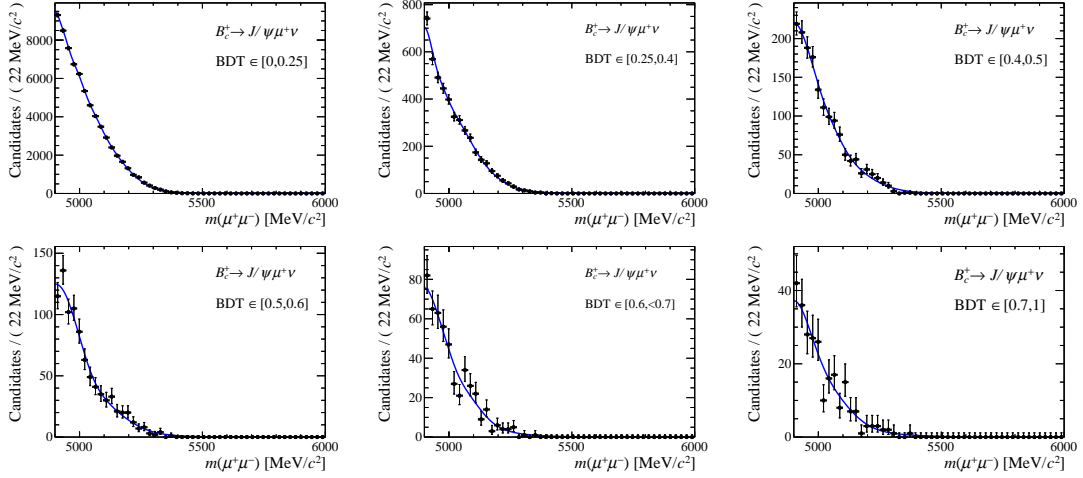
**Figure 6.6:** Generator level distributions of the  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decay on which requirements are made during generation. The distributions between the Kiselev [151] and Ebert [167] form factor model are in full agreement.

The simulation was generated using a form factor model developed by Ebert et al. [167] requiring only the non- $J/\psi$  (opposite-sign) dimuon combination in the LHCb acceptance with a minimal mass of  $4.5 \text{ GeV}/c^2$ . Only this dimuon combination can reach the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  mass range and therefore act as background. Thus this requirement at a very early stage of the simulation makes the generation much more efficient. To investigate the dependence on the theoretical form factor model, the distributions and acceptance efficiencies are compared to an alternative form factor model by Kiselev [151] in Fig. 6.6, with large statistics generator level simulation. No difference between the two theoretical models is found.

Compared to the previous analysis [73] a strong drop in the detector acceptance efficiency computed during simulation generation (previously  $(0.4350 \pm 0.006)\%$  in the 2016 data set, now  $(0.19217 \pm 0.00033)\%$ ) is found. This difference is due to the following effects:

- there was a mistake in the cut tool used in the generation for the previous analysis. This tool accepted all events with more than one  $B_c^+$ . This led to an absolute offset of the efficiency of about 0.2%.
- in the previous analysis all charged final state particles of the decay were required to fly into the LHCb detector to be reconstructed. This mismodels the behaviour on data and was corrected in the previous analysis for the generator level efficiency. However, this also affects the  $J/\psi$  veto. Therefore the acceptance of the LHCb detector is modelled more accurately this way.
- the generation in the previous analysis was done assuming a flat phase space for the muons. While the difference to the Kiselev model was corrected for the generator level efficiencies together with the acceptance mismodelling mentioned above, this correction was probably insufficient due to the mistake in the generator cut tool.

For the simulation used in this analysis the mistakes described above are fixed and the Ebert model and appropriate acceptance requirements are used for a correct estimate of the efficiencies of the detector acceptance. The improvements have a great impact on the expected amount of  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  events with respect to the previous analysis. Due to the reduced



**Figure 6.7:** Invariant mass distribution of the  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  channel and shape fitted with Gaussian kernel density estimators from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom). The mass distributions of the other background decays are modelled in the same manner.

expectation and the largely improved branching fraction accuracy, the correlation with the measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions is now small ( $\approx 1\%$ ) and therefore the background is of minor relevance in the fit. However, for the additional measurement of the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  branching fraction, the accurate determination is important, as the mass distribution has a shoulder as shown in Fig. 6.7 in the place where  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decays could become visible.

## 6.5 Expected results

In this section first the fit model to the mass distribution is described in Section 6.5.1 and sensitivity studies are presented in Section 6.5.2. Furthermore a blind check of the compatibility with the previous result is discussed in Section 6.5.3. A conclusion and an outlook is given on the remaining parts to finalise the analysis in Section 6.5.4.

### 6.5.1 Fit model

The data are fitted simultaneously in several data samples. On the one hand the data are split by Run 1 and Run 2 due to the different data taking environments and on the other hand the data are split in bins of the selection BDT. With the increased data set the binning scheme used in this analysis is finer than in the previous analysis and consists now of 6 bins with the boundaries  $[0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 1.0]$ , while the previous analysis used 5 bins with the

boundaries [0.0, 0.25, 0.4, 0.5, 0.6, 1.0]. Because the description differs significantly from the other BDT bins due to the dominance of background from random track combinations, the lowest BDT bin is excluded from the fit. Correlations in the inputs of the fit are taken into account by sharing common parameters and pdfs in the description of each sample. The distributions of the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  components are described by double Crystal Ball functions (similar to the model described in Section 5.3.1) with their parameters Gaussian constrained to values calibrated from data. Shapes from physical backgrounds are determined on simulation using Gaussian kernel density estimators separate for each BDT bin. The combinatorial background is modeled by a single exponential with a shared but unconstrained slope for all samples of a given Run and free yields.

The  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fractions ( $\mathcal{B}_{\text{sig}}$ ) are free in the fit, shared between all samples, and obtained from the signal yield per BDT bin  $i$  in each Run,

$$N_{\text{sig}}^i = \mathcal{B}_{\text{sig}} \times \frac{f_{\text{BDT},\text{sig}}^i c_{\text{PID},\text{sig}}^i c_{\text{time},\text{sig}}^i}{\sum_i f_{\text{BDT},\text{sig}}^i c_{\text{PID},\text{sig}}^i c_{\text{time},\text{sig}}^i} \times \alpha_{d(s)}, \quad (6.5)$$

where  $f_{\text{BDT},\text{sig}}^i$  (Gaussian constrained),  $c_{\text{PID},\text{sig}}^i$ , and  $c_{\text{time},\text{sig}}^i$  (fixed in the fit due to the negligible uncertainties) denote the estimated fraction of signal candidates falling in each BDT bin and its corrections due to the BDT dependence of PID efficiencies and the BDT dependent decay time acceptance, respectively. The decay time acceptance effect is calculated in a similar way as for the preselection and is negligible for the  $B^0 \rightarrow \mu^+ \mu^-$  component. The BDT fraction in the lowest BDT bin is defined as  $f_{\text{BDT},\text{sig}}^0 = 1 - \sum_{i>0} f_{\text{BDT},\text{sig}}^i$ . The sum over the BDT bins in Eq. (6.5) is performed over all BDT bins, including the lowest BDT bin, thus providing a proper normalisation of the BDT fractions. The formulae for the combined normalisation factors  $\alpha_{d(s)}$  in Eq. (6.1) are implemented into the fit to be able to share  $f_s/f_d$  and the normalisation branching fractions between the data taking periods and the normalisation yields also with the normalisation of background components.

The yields of  $B \rightarrow hb^{(\prime)}$  backgrounds are similarly implemented in the fit as

$$N_{\text{hh}}^i = \frac{N_{B^0 \rightarrow K^+ \pi^-}}{R_{K\pi/bb}} \times \varepsilon_{bb \rightarrow \mu\mu} \times \varepsilon_{\text{trig}, B_{(s)}^0 \rightarrow \mu^+ \mu^-} \times \frac{f_{\text{BDT},\text{hh}}^i c_{\text{PID},\text{hh}}^i}{\sum_i f_{\text{BDT},\text{hh}}^i c_{\text{PID},\text{hh}}^i}, \quad (6.6)$$

with the run-specific normalisation yield  $N_{B^0 \rightarrow K^+ \pi^-}$  (see Table 6.3), corrected by the trigger and PID efficiencies, the ratio of the  $B^0 \rightarrow K^+ \pi^-$  branching fraction over all  $B \rightarrow hb^{(\prime)}$  branching fractions  $R_{K\pi/bb}$ , the trigger efficiencies of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays, and the average misidentification probability  $\varepsilon_{bb \rightarrow \mu\mu}^r$ . The trigger efficiency is assumed to be similar to the one of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays as after the PID selection the remaining  $B \rightarrow hb^{(\prime)}$  decays are expected to be decays where the hadrons have decayed into muons, thus yielding a similar signature in the trigger.

Correspondingly also the other backgrounds are implemented in the fit as

$$N_{\text{bkg}}^i = \mathcal{B}_{\text{bkg}} \times \frac{N_{B^+ \rightarrow K^+ J/\psi}}{\mathcal{B}_{B^+ \rightarrow K^+ J/\psi}} \times \left( \frac{f_{(d,s,c)}}{f_d} \right) \times \varepsilon_{\text{bkg}} \times \frac{f_{\text{BDT,bkg}}^i}{\sum_i f_{\text{BDT,bkg}}^i}, \quad (6.7)$$

with the run-specific efficiency corrected normalisation yield  $N_{B^+ \rightarrow K^+ J/\psi}$ , its branching fraction, and the run-specific hadronisation fraction ratio  $f_{(s,c,d)}/f_d$ . Those values are shared with the normalisation factor computation. All numbers used for the calculation of background yields are Gaussian constrained in the fit.

Projections of the blinded fit to the Run 1 and Run 2 sample in slices of the BDT bins are shown in Fig. 6.8, where the contributions of different background components are also presented. The left sidebands are dominated by backgrounds from  $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$  decays, while the right sidebands contain purely combinatorial background. The strongest background in the still blind signal window is due to misidentified  $B \rightarrow b b^{(\prime)}$  decays. An excellent description of the dimuon sidebands is found, highlighting the accurate estimation of the background contributions.

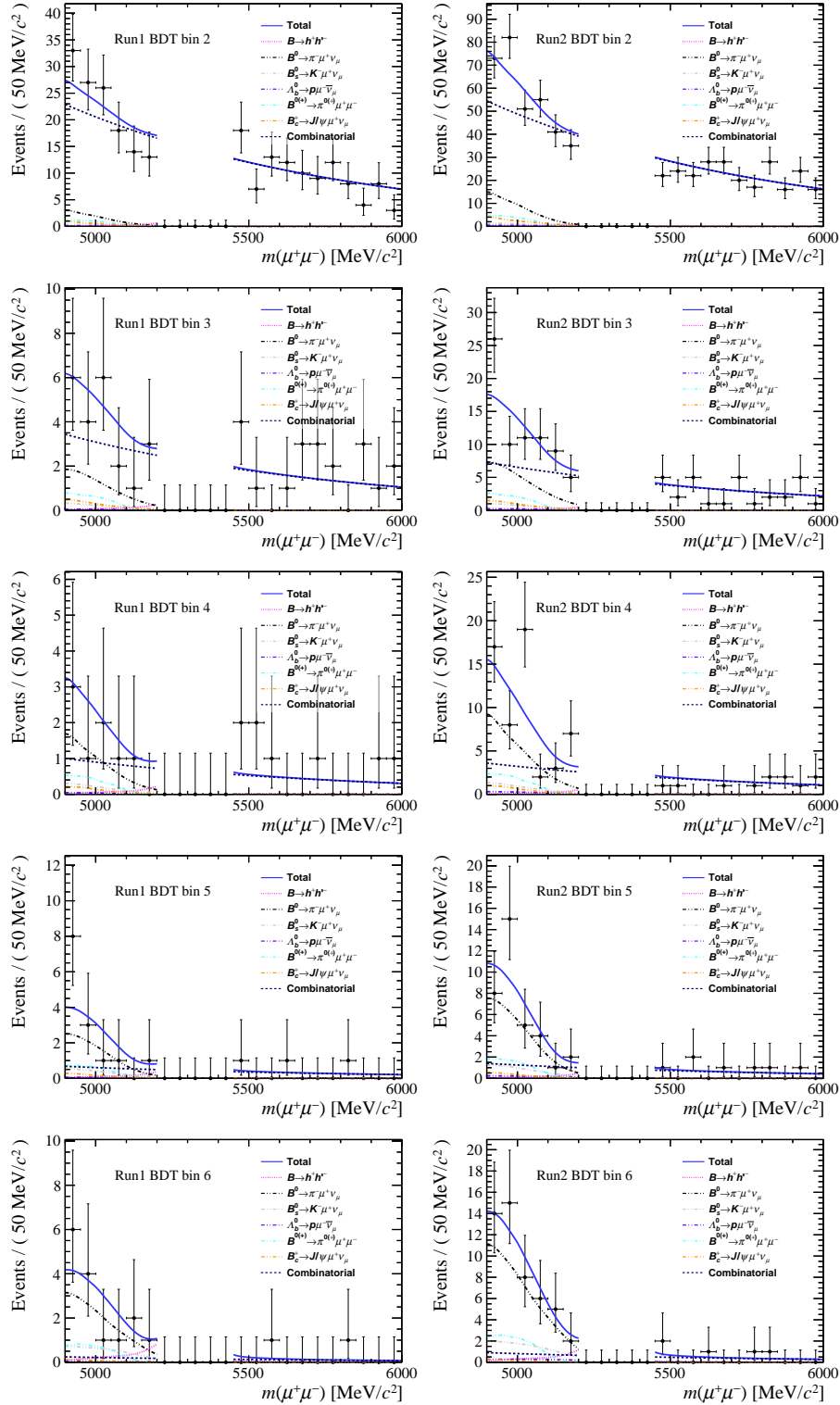
## 6.5.2 Sensitivity studies

Studies with pseudo experiments are performed to estimate the sensitivity of the analysis to the target branching fractions of interest. The pseudo experiments are based on the blinded fit of the full data sample. The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions are set to the SM predictions [16]. From this modified fit function new pseudo data samples are generated and refitted. The global observables that are used in the Gaussian constraints for many fit parameters are generated randomly according to their Gaussian distributions as well, following the *Plugin* approach discussed in Section 3.4. The distributions of the fit results yield the expected sensitivity of the measurement. No bias from the nominal fitting procedure is found as shown in Fig. 6.9 and a sensitivity estimate of

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.67 \pm 0.54) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.0 \pm 0.9) \times 10^{-10} \\ \mathcal{R}_{\mu^+ \mu^-} &= 0.028 \pm 0.025 \end{aligned}$$

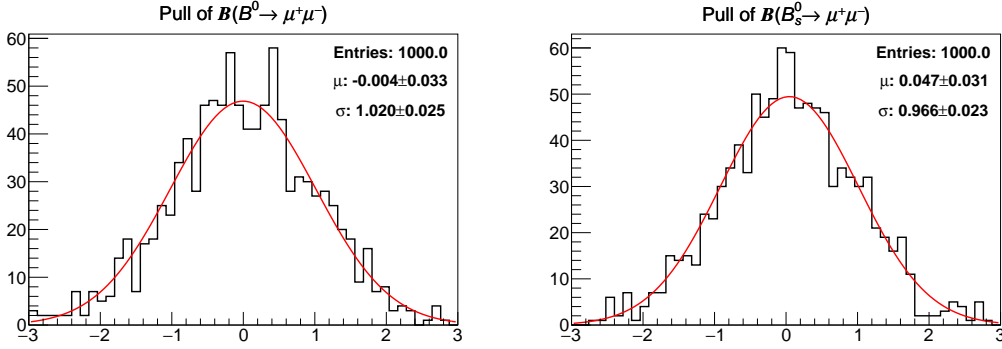
is obtained, including statistical and systematic uncertainties. The relative uncertainties are compatible to the recent combination of all LHC experiments [77]. From these pseudo experiments the probability to find “evidence” for the decay  $B^0 \rightarrow \mu^+ \mu^-$  with more than  $3\sigma$  significance is about 5 %.

Although the absolute normalisation of the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  branching fraction is not fully calibrated yet, preliminary studies with pseudo experiments are performed with including



**Figure 6.8:** Blinded fit to the full Run 1 (left) and Run 2 (right) samples in slices of the BDT bin. The sideband data are described to good accuracy.





**Figure 6.9:** Pull distributions of the signal branching fractions obtained with 1000 pseudo experiments showing that the parameter estimates are unbiased. The pull of a parameter is defined as  $\text{Pull}(x) = \frac{x_{\text{fit}} - x_{\text{generated}}}{\sigma(x_{\text{fit}})}$  with the generation value  $x_{\text{generated}}$  and the fitted value  $x_{\text{fit}}$  and its uncertainty  $\sigma(x_{\text{fit}})$ . This construction produces a normal distribution with mean 0 and width 1 for unbiased fits, while biased fits show deviations in the mean and underestimated (overestimated) uncertainties increase (decrease) the width.

the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  component in the fit to study the impact on the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions. It is found that an unbiased measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  branching fractions can be obtained. However, a correlation of 20% is found between the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  branching fractions, which is incorporated in the uncertainties of the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction value quoted above.

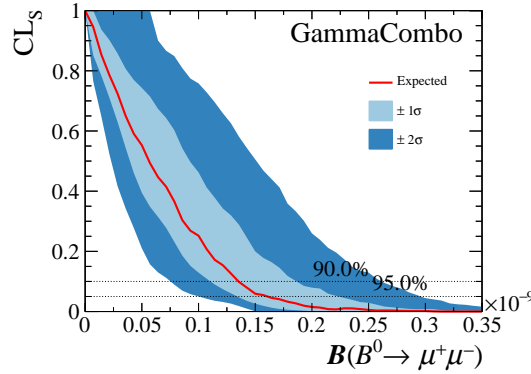
Since finding evidence for  $B^0 \rightarrow \mu^+ \mu^-$  is not likely, an upper limit on the  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  will be set using the CL<sub>S</sub> method [86] with the GAMMA COMBO framework [78, 79], based on pseudo experiments with a one-sided test statistic [81], as discussed in Chapter 3. The expected CL<sub>S</sub> curves assuming negligible  $B^0 \rightarrow \mu^+ \mu^-$  contributions are displayed in Fig. 6.10. An expected upper limit of

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.3(1.5) \times 10^{-10} \text{ at } 90\%(95\%) \text{ CL}$$

is determined.

An estimate of the statistical-only can be obtained by repeating the fit after fixing all the fit parameters, except for the  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^0 \rightarrow \mu^+ \mu^-$  branching fractions and the parameters of the combinatorial background, to their expected values in the sensitivity studies described above. The results are compared to the numbers obtained above that include statistical and systematical uncertainties, such that the results can be expressed as

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.67 \pm 0.49 \pm 0.31) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.0 \pm 0.86 \pm 0.14) \times 10^{-10}, \text{ and} \\ \mathcal{R}_{\mu^+ \mu^-} &= 0.028 \pm 0.025 \pm 0.004. \end{aligned}$$



**Figure 6.10:**  $CL_s$  curves of  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  from pseudo experiments assuming  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = 0$ . The red line corresponds to the expected upper limit curve as the median of the pseudo experiments.

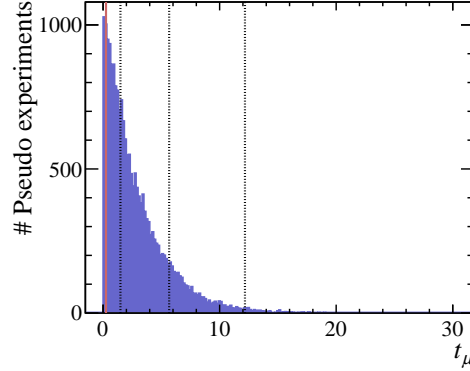
For the comparison it is assumed that the total uncertainty arises from adding systematic and statistical uncertainties in quadrature. The systematic uncertainties of  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  and  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  are small compared to the statistical uncertainty and dominated by the limited knowledge of the exclusive backgrounds, the uncertainty on  $f_s/f_d$  and the uncertainty on the normalisation branching fractions, while subleading systematic uncertainties arise from the limited precision of the PID and track reconstruction efficiency determinations, as discussed in Section 6.3.

### 6.5.3 Compatibility with the previous result

While the analysis is still blind, the compatibility of the data with the previous result is investigated with a Likelihood Ratio test similar to what is described in Section 3.2. Pseudo experiments are generated from the blind fit to the data and the distribution of the Likelihood Ratio  $t_\mu = -2 \ln \left( \frac{\mathcal{L}(\mu|x_\mu)}{\mathcal{L}(\hat{\mu}|x_\mu)} \right)$  is determined, where  $\mathcal{L}(\mu|x_\mu)$  corresponds to the likelihood of the fit to the pseudo experiments with the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions fixed to the current (blind) measurement and  $\mathcal{L}(\hat{\mu}|x_\mu)$  corresponds to the likelihood where the branching fractions are free to vary. This distribution is then compared to the measured test statistic value in data, where the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions are set to the previously measured results [73]:

$$t_{\text{meas}} = -2 \ln \left( \frac{\mathcal{L}(\mu_{\text{prev.}}|x_{\text{data}})}{\mathcal{L}(\hat{\mu}|x_{\text{data}})} \right)$$

The comparison is shown in Fig. 6.11. A good compatibility of the current measurement with the previous result below  $1\sigma$  is found.



**Figure 6.11:** Test statistic  $t_\mu$  distribution obtained from pseudo experiments. The measured test statistic value evaluated at the results of the previous analysis [73] is indicated with a red line. It is well below the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  thresholds indicated by the dotted lines.

#### 6.5.4 Summary and outlook

With expected sensitivities of

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.67 \pm 0.49 \pm 0.31) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.0 \pm 0.86 \pm 0.14) \times 10^{-10}, \text{ and} \\ \mathcal{R}_{\mu^+ \mu^-} &= 0.028 \pm 0.025 \pm 0.004\end{aligned}$$

(assuming SM branching fractions), these results will constitute to the most precise measurements of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  system with a single experiment and narrow down the space allowed for New Physics contributions as e.g. proposed in Ref. [17, 70]. The combination of the measurements by all LHC experiments with the full Run 1+2 data set will allow to clarify whether the branching fraction of  $B_s^0 \rightarrow \mu^+ \mu^-$  follows the SM expectations or the discrepancy seen in previous measurements remains. If the lower experimental value persists, the combination of the LHC experiments is expected to show an evidence of an inconsistency with the SM. The measurement is furthermore a crucial ingredient in the puzzle of the anomalies seen in  $b \rightarrow s \ell^+ \ell^-$  transitions and will continue to play a major role in the question whether there exist NP contributions to axial-vector  $b \rightarrow s \ell^+ \ell^-$  transitions. As discussed in Chapter 7, with future data sets the decay  $B^0 \rightarrow \mu^+ \mu^-$  will be unambiguously observed, which will allow precise tests of the minimal flavour violation hypothesis [71].

The analysis is close to publication and is currently being reviewed by the LHCb collaboration. To finalise the branching fraction measurements, the hadronisation fraction ratio  $f_s/f_d$  needs to be updated with the value of the ongoing combination of LHCb measurements [163]. This will lead to small modifications of the the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fractions reduce its total systematic uncertainty from about 8.4% to 6.8%. Not all efficiencies of the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  channel and the fraction of decays falling in the different BDT regions have been calibrated

on data yet. However, the relevant techniques can be applied from the calibration of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. From the comparison to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decay modes, modifications of the  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  contribution by a few percent are expected. Through its correlation with the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction, this can have an effect on the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction of  $\mathcal{O}(1\%)$ .

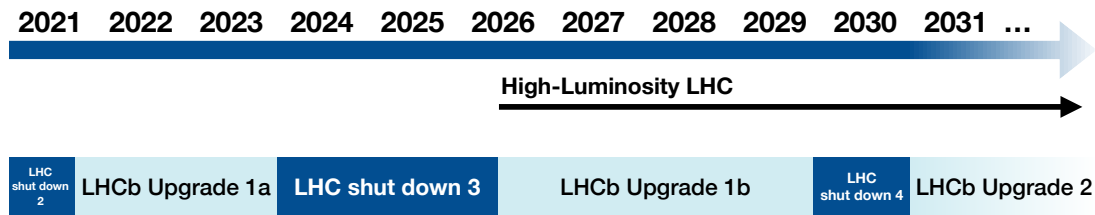
## 7 Extrapolations to the upgrades of the LHCb experiment

The data sets analysed in this thesis constitute only to a part of the data that is expected to be collected by the LHCb experiment. This chapter reviews the ongoing and planned upgrades of the LHCb experiment to collect data until 2035–2040 and their impact on the sensitivities of the analyses discussed in this thesis.

### 7.1 Upgrades and future data sets of the LHCb experiment

In the years 2011–2012 (Run 1) and 2015–2018 (Run 2) the LHCb experiment has collected an integrated luminosity of about  $9 \text{ fb}^{-1}$  of proton-proton collision data at centre-of-mass energies of 7 TeV, 8 TeV and 13 TeV. This was realised by taking data at an instantaneous luminosity of  $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ , which could be achieved with a distance between the proton bunches of 25 ns and about one visible proton-proton collision per bunch crossing (pile-up) [96, 112]. Since the end of 2018, the LHC is being upgraded to achieve even larger luminosities. It is currently scheduled to restart operating in 2021/2022 with a bunch spacing of 25 ns and a pile-up of about 5 [168] for the LHCb experiment, while rising the collision centre-of-mass energy to its design value of 14 TeV. Thus for the LHCb experiment it is aimed to deliver an increased instantaneous luminosity by a factor of five to  $2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ . With this configuration, in the years 2021–2023 (Run 3) and 2026–2029 (Run 4) an integrated luminosity of  $14 \text{ fb}^{-1}$  and  $27 \text{ fb}^{-1}$ , respectively, is expected to be collected. After Run 3 another upgrade of the the LHC to the so-called “High-Luminosity-LHC” is foreseen [169–171], which is aiming to deliver an instantaneous luminosity between  $1 \times 10^{34}$  and  $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  with an average pile-up of up to 50 proton-proton collisions per bunch crossing for the LHCb experiment, but only after Run 4. Thus an integrated luminosity of  $250 \text{ fb}^{-1}$  is expected to be collected by the LHCb experiment beginning in 2031 until the end of the LHC running period 2035–2040 (Run 5). A tentative timeline for the foreseen data taking periods is given in Fig. 7.1. Due to the ongoing COVID-19 pandemic, the restart of data taking in 2021 for Run 3 is considered challenging, though.

In order to operate at the higher data rates and sustain a harsher radiation environment, the LHCb detector itself is planned to be undergoing three upgrades, also increasing its



**Figure 7.1:** Estimated timeline of operations of the LHC accelerator and the LHCb experiment [171]. Due to the ongoing COVID-19 pandemic, the start of Run 3 in 2021 is considered challenging.

performance. In the following, the main modifications expected to impact the sensitivities of the analyses discussed in this thesis are highlighted. As part of the current upgrade “Upgrade 1a”, large parts of the detector are replaced [97]. Most important for the analyses discussed in this thesis are the change of the trigger system to a fully software-based trigger [172] and the removal of the SPD and preshower detectors [173]. While the additional pile-up leads to an increase in the data rate, it also increases the detector occupancy, leading to a worse PID performance especially of RICH and calorimeter detectors. This is not expected to have a strong impact on the muon PID and thus on the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analysis, but the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays and to some extent also the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  will be strongly affected as electron and kaon PID are achieved dominantly through RICH and the calorimeters. Another effect of the increased pile-up is a degradation of the track isolation criteria due to the larger number of tracks in the vicinity of a signal track. These criteria have a high impact on the discussed analyses, especially the analysis of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. The removal of the Preshower and SPD detectors will lead to a slightly worse  $e-\pi^0$  separation. It will also demand higher energy thresholds for photon clusters of bremsstrahlung photons, thus making it harder to apply bremsstrahlung correction (also due to the higher detector occupancy), a key ingredient for the measurements involving electrons. However, the new trigger scheme with a software-only trigger will greatly enhance the selection efficiencies of non-muonic decays. Thus, trigger efficiencies of nearly 100 % will be possible, which will enhance the efficiency of the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays up to a factor of about 1.5 compared to the trigger efficiencies determined for the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis in Section 5.5. Similarly, the search for  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  will profit from this new solution. The impact in this analysis will be smaller, though, since the decays are selected by the triggers via the signature from the muon, which has a high efficiency. As the trigger efficiency for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  modes is already close to 100 % with the current scheme, this measurement will not profit much from the new trigger scheme.

Between Run 3 and Run 4 a minor upgrade “Upgrade 1b” is planned, replacing the ECAL with cells with a better spatial resolution and adding timing information to hits in the ECAL

with a resolution better than 100 ps, as well as timing information with a resolution around 5 ns in the muon system [170]. Furthermore the HCAL will be removed, further improving the performance of the muon system. These modifications are expected to significantly improve the electron PID and bremsstrahlung correction with the help of the improved spatial and time resolution. At the same time the muon reconstruction and PID efficiency will improve the sensitivity of analyses with muons in the final state.

After Run 4 a large scale upgrade, “Upgrade 2”, is being prepared, adding precise timing measurements to all subdetectors and replacing the existing ones, because they will have suffered significant radiation damage [170]. A more fine-grained lateral segmentation of the calorimeters and segmentation along the  $z$ -axis is also being discussed. The timing information is of specific importance, since with Run 5 the pile-up is increased further by a factor of ten. Adding precise time information to the subdetectors will allow to separate the multiple  $pp$ -collisions within an event, thus effectively reducing the pile-up effects on the reconstruction of candidates. It will be of crucial importance to maintain a high track reconstruction and PID efficiency. At the same time it may provide new opportunities to improve the isolation criteria due to the additional time coordinate. For this upgrade the finer segmentation of the calorimeter will also be of great interest to disentangle better the overlapping showers of nearby particles, thus improving electron and photon identification.

The following section gives extrapolations of the sensitivity of the analyses discussed in this thesis based on the expected luminosity increase. The changes of the experimental environment are neglected.

## 7.2 Extrapolation of the sensitivity

To estimate the sensitivity of the measurements described in this thesis at higher luminosities, the data from the fits in Sections 5.8 and 6.5 are scaled. As discussed in Section 5.5.5, the yield of heavy flavour decays – both normalisation and background – increases linearly with luminosity and with centre-of-mass energy. Therefore the statistical uncertainties decrease with the square root of the luminosity, but the systematic uncertainties are assumed to stay constant in this study. Also the efficiencies are assumed to stay constant, which has been studied previously for the data sets in Run 2 [174–176] (and discussed in Chapter 6). The performance of the searches being better in Run 2 than in Run 1, for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  extrapolations, the efficiencies for future data sets are chosen to be similar to the ones found in Run 2. Because the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  selection is expected to be improved [177] by using the previously unavailable track isolation variable developed for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analysis [73], the normalisation constant and the background yield are scaled to match the selection in that thesis ( $N_{\text{bkg}}^* = N_{\text{bkg}} \times 2.26$  and  $\alpha^* = \alpha/2.52$ ). Similarly this new track isolation has the potential to improve also the sensitivity of the search for  $B_{(s)}^0 \rightarrow e^+ e^-$  decays. This has not been studied quantitatively, though.

**Table 7.1:** Sensitivity projections for the analyses discussed in this thesis. For the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  analysis two values are given, once assuming the current systematic uncertainties and once neglecting them. The uncertainties on  $\mathcal{R}_{\mu^+\mu^-}$  are expected similar to  $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ .

	Upper limit at 95 % of $\mathcal{B}(B_s^0 \rightarrow e^+e^-)$ [ $\mathcal{B}(B^0 \rightarrow e^+e^-)$ ]	Upper limit at 95 % of $\mathcal{B}(B^+ \rightarrow K^+\mu^+e^-)$ [ $\mathcal{B}(B^+ \rightarrow K^+\mu^-e^+)$ ]
Run 1–2	$5.4 \times 10^{-9}$ [ $1.6 \times 10^{-9}$ ]	$3.5 \times 10^{-9}$ [ $1.7 \times 10^{-9}$ ]
Run 1–3	$3.2 \times 10^{-9}$ [ $0.94 \times 10^{-9}$ ]	$2.1 \times 10^{-9}$ [ $0.98 \times 10^{-9}$ ]
Run 1–4	$2.1 \times 10^{-9}$ [ $0.61 \times 10^{-9}$ ]	$1.4 \times 10^{-9}$ [ $0.65 \times 10^{-9}$ ]
Run 1–6	$0.82 \times 10^{-9}$ [ $0.24 \times 10^{-9}$ ]	$0.55 \times 10^{-9}$ [ $0.26 \times 10^{-9}$ ]

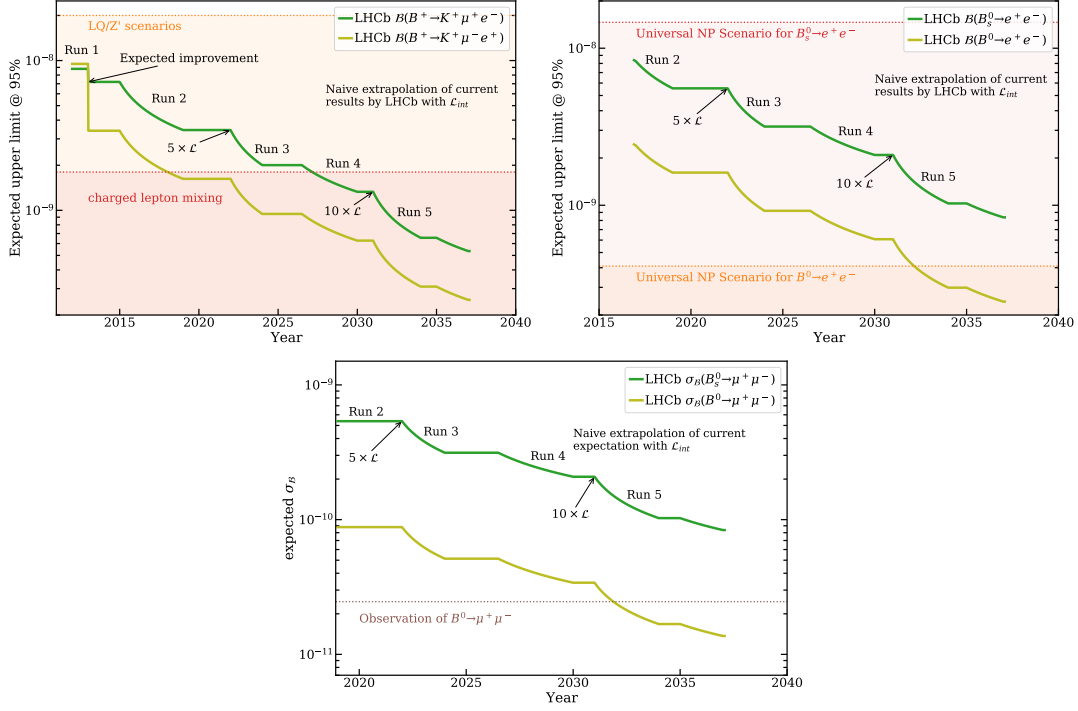
	$\sigma_{\mathcal{B}}(B_s^0 \rightarrow \mu^+\mu^-)$ [ $\sigma_{\mathcal{B}}(B^0 \rightarrow \mu^+\mu^-)$ ] incl. syst.	$\sigma_{\mathcal{B}}(B_s^0 \rightarrow \mu^+\mu^-)$ [ $\sigma_{\mathcal{B}}(B^0 \rightarrow \mu^+\mu^-)$ ] excl. syst.
Run 1–2	$5.4 \times 10^{-10}$ [ $8.8 \times 10^{-11}$ ]	$5.4 \times 10^{-10}$ [ $8.8 \times 10^{-11}$ ]
Run 1–3	$3.6 \times 10^{-10}$ [ $5.1 \times 10^{-11}$ ]	$3.1 \times 10^{-10}$ [ $5.1 \times 10^{-11}$ ]
Run 1–4	$2.8 \times 10^{-10}$ [ $3.4 \times 10^{-11}$ ]	$2.1 \times 10^{-10}$ [ $3.4 \times 10^{-11}$ ]
Run 1–6	$2.5 \times 10^{-10}$ [ $1.3 \times 10^{-11}$ ]	$0.84 \times 10^{-10}$ [ $1.3 \times 10^{-11}$ ]

By scaling the normalisation and the expected background yield from the fits in Sections 5.8 and 6.5 pseudo experiments are generated and fitted to estimate the branching fraction uncertainties at the increased luminosities. The expected upper limit at the increased luminosities is then obtained by multiplying the expected upper limit of the recent measurements by the ratio of the scaled branching fraction uncertainties over the uncertainties discussed in Sections 5.8 and 6.5.

The sensitivity of a signal measurement as well as an upper limit in the presence of background are expected to decrease with the square root of the  $b$ -hadron yield increase, since the signal and background yields follow a Poissonian distribution. The estimated sensitivities of  $B^+ \rightarrow K^+\mu^\pm e^\mp$ ,  $B_{(s)}^0 \rightarrow e^+e^-$  and  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  analyses for Run 1–5 are given in illustrated in Table 7.1 and illustrated in Fig. 7.2.

As can be seen, the expected upper limits of the  $B_{(s)}^0 \rightarrow e^+e^-$  and  $B^+ \rightarrow K^+\mu^\pm e^\mp$  branching fractions follow the expected scaling with the square root of the luminosity increase. However, the precision of the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fraction measurements become dominated by systematic uncertainties and thus saturate. Therefore in Table 7.1 additionally to the expected sensitivities including the systematic uncertainty, the same numbers excluding the systematic uncertainty are given. The uncertainties of  $\mathcal{R}_{\mu^+\mu^-}$  are dominated by the uncertainties on  $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$  and are therefore expected to scale similarly. The systematic uncertainty of the latest  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fraction measurement amounts to 8% (on  $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ )

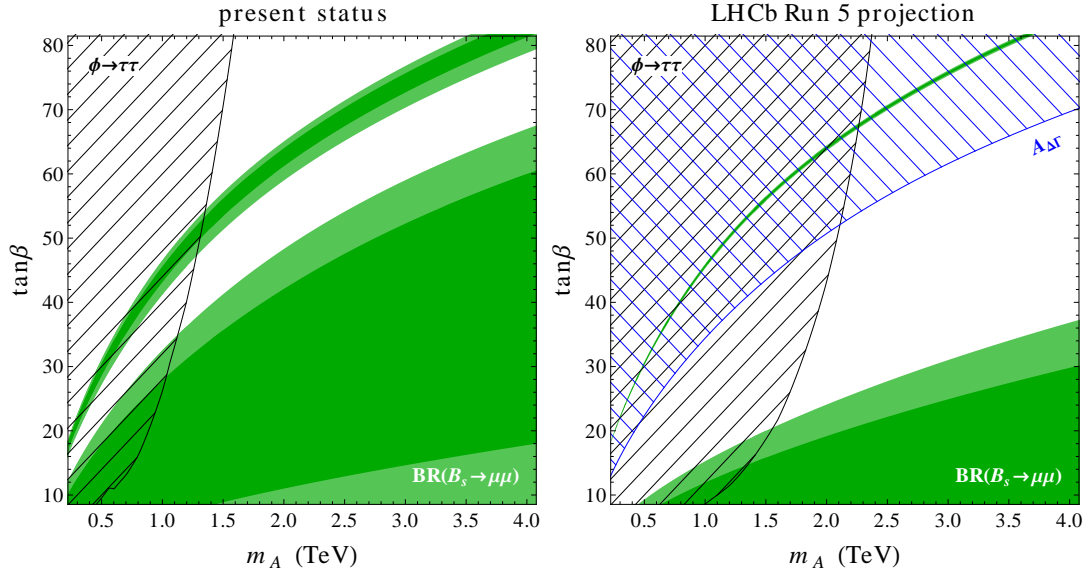




**Figure 7.2:** Sensitivity projections for the analyses discussed in this thesis, starting at the latest measurement and assuming a proportionality to the square root of the integrated luminosity  $\mathcal{L}_{\text{int}}$ . The projections for the  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  analysis (top left) are compared to leptoquark and  $Z'$  scenarios [51–54] and scenarios for charged-lepton mixing [55, 56]. A significant sensitivity improvement is seen by applying the improvements from [177] to the current results [88]. For the projections of the  $B_{(s)}^0 \rightarrow e^+ e^-$  analysis (top right) the range accessible to (pseudo-) scalar new physics scenarios [68] are highlighted. The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  extrapolations (bottom) show the expected branching fraction uncertainty  $\sigma_B$  and assume negligible systematic uncertainties. A dashed line indicates an expected median significance for  $B^0 \rightarrow \mu^+ \mu^-$  of  $5\sigma$ .

and arise from the limited knowledge of the  $b$ -fragmentation fraction ratio  $f_s/f_d$  and the normalisation branching fractions, the limited precision of PID and track reconstruction efficiency estimations, as well as the determination of background contributions. As the uncertainty of these quantities are partly depending on the size of the data set used, one can expect that those systematic uncertainties will decrease as well with increasing luminosity. In Ref. [171] therefore a conservative assumption of a systematic uncertainty of 4% is made. However, since the publication of Ref. [171] the estimation of  $f_s/f_d$  [163] and the estimation of the  $B^+ \rightarrow K^+ J/\psi$  branching fraction [21] have become significantly more precise, as discussed in Section 6.3. The uncertainty of the background contributions is dominated by size of PID calibration data and simulation samples, which are necessary to estimate the  $\pi$ - $\mu$  and  $K$ - $\mu$  misidentification probability. With more data the estimate of this probability will become more precise and thus the systematic uncertainty of the background contributions will decrease. Therefore it is considered a reasonable assumption that the systematic uncertainties will continue to play only a small role as in the current measurement and thus the sensitivity of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions will also scale with the square root of the luminosity increase.

The measurements with increased data sets will play a significant role in constraining and rejecting scenarios for physics beyond the SM motivated by the anomalies seen in  $b \rightarrow s \ell^+ \ell^-$  transitions, as illustrated in Figs. 7.2 and 7.3. Although most models cannot be ruled out completely, the most favoured scenarios can be excluded. Especially Fig. 7.3 highlights also the complementarity of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fraction measurements to direct searches for heavy particles carried out at the ATLAS and CMS experiments like the search for heavy particles decaying into two tauons [178, 179]. As discussed above, there is potential to enhance the sensitivities of the discussed analyses even beyond increasing the data set with the improvement of the selection strategies and the development of new analysis techniques adapted to the evolution of the data taking environment. While the CMS and ATLAS experiments have a slightly lower sensitivity in the measurement of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  than the LHCb experiment, the combination of the measurements with all three experiments with the upgrade data sets will provide an even more precise determination of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions and a clear picture whether the anomalies discussed in Chapter 2 are related to  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays or not.



**Figure 7.3:** Sensitivity projections for a new physics scenario involving an additional pseudoscalar Higgs boson  $A$  [17]. Run 5 includes Run 6 in this notation. The parameter  $m_A$  describes the mass of the additional Higgs boson and  $\tan\beta = v_2/v_1$  denotes the ratio of the two Higgs vacuum expectation values. The light green area is allowed in the model, whereas the black hatched region is excluded by direct searches for a heavy resonance  $\phi$  decaying into a  $\tau^+\tau^-$ -pair with the ATLAS and CMS experiments [178, 179]. For Run 5, those are extrapolated, assuming an integrated luminosity of  $3000 \text{ fb}^{-1}$  for both the ATLAS and CMS experiment, respectively. The blue hatched region is expected to be excluded by measurements of the mass eigenstate rate asymmetry  $A_{\Delta\Gamma}$ , accessible via a measurement of the effective lifetime  $\tau_{B_s^0 \rightarrow \mu^+\mu^-}$ .



## 8 Summary

In the past decades, the SM has proven to be an excellent description of most particle physics processes. Yet it leaves several questions open which motivate searches for NP processes. While direct searches for new particles have been fruitless so far, intriguing deviations from the SM are accumulating in indirect searches via measurements of  $b \rightarrow s\ell^+\ell^-$  processes.

This thesis describes measurements complementary to these anomalies with analyses of purely leptonic and lepton-flavour violating decays that proceed via  $b \rightarrow s\ell^+\ell^{(\prime)-}$  processes. It presents the searches for  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+e^-$  decays and an update of the measurement of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fractions with the LHCb experiment. To allow the interpretation of rare decays searches as limits on branching fractions, statistical methods are implemented in a common framework and applied to the analyses studied in this thesis.

The search for the lepton-flavour violating  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays has never before been carried out with LHCb data and is performed with data corresponding to  $3\text{ fb}^{-1}$  of  $pp$  collisions collected at the centre-of-mass energies of 7 and 8 TeV. These decays are forbidden in the SM. However, in many NP scenarios that explain the anomalies seen in  $b \rightarrow s\ell^+\ell^-$  processes they could have sizeable branching fractions. A major challenge in this analysis is the suppression of partially reconstructed backgrounds to a negligible level with multivariate methods. No excess of  $B^+ \rightarrow K^+\mu^\pm e^\mp$  events has been found and upper limits of

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^+\mu^-e^+) &< 9.5 \times 10^{-9} \text{ and} \\ \mathcal{B}(B^+ \rightarrow K^+\mu^+e^-) &< 8.8 \times 10^{-9}\end{aligned}$$

at 95 % CL are determined on the branching fractions. These limits improve previous results by more than an order of magnitude, which imposes strong constraints on NP scenarios that explain the  $b \rightarrow s\ell^+\ell^-$  anomalies.

Using about twice the data set, corresponding to  $5\text{ fb}^{-1}$  collected at the centre-of-mass energies of 7, 8 and 13 TeV, a search for the purely leptonic  $B_{(s)}^0 \rightarrow e^+e^-$  decays is performed. Also this search is performed with the LHCb experiment for the first time. The  $B_{(s)}^0 \rightarrow e^+e^-$  branching fractions are known to be unmeasurably small, but they could be strongly enhanced by scalar and pseudoscalar currents, while the analogous rates of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays remain relatively unchanged. Despite a significant mass resolution degradation due to bremsstrahlung effects,

a precise test of the SM is performed. The result of the measurement is consistent with the background-only expectation and upper limits of

$$\mathcal{B}(B_s^0 \rightarrow e^+e^-) < 11.2 \times 10^{-9} \text{ and}$$

$$\mathcal{B}(B^0 \rightarrow e^+e^-) < 3.0 \times 10^{-9}$$

at 95 % CL are set, improving previous limits even by a factor of 30. With this improvement, the measurement of  $B_{(s)}^0 \rightarrow e^+e^-$  processes becomes sensitive to universal NP scenarios with scalar and pseudo-scalar  $b \rightarrow s\ell^+\ell^-$  currents.

Again doubling the data set with respect to the search for  $B_{(s)}^0 \rightarrow e^+e^-$  decays, the analogous  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays are studied with the full LHCb data set collected in Run 1 and Run 2, corresponding to  $9 \text{ fb}^{-1}$  collected at the centre-of-mass energies of 7, 8 and 13 TeV. Because of their precise theoretical estimate and experimental simplicity they are key channels in the search for NP processes. Indeed the current combination of LHC experiments shows a tension with the SM that fits to the anomalies seen in other  $b \rightarrow s\ell^+\ell^-$  processes. Due the small decay rates a major challenge in this analysis is the precise description of efficiencies and background components. The analysis is currently being reviewed by the LHCb collaboration. While the result of the analysis is not uncovered yet, a precision of

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.67 \pm 0.49 \pm 0.31) \times 10^{-9},$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (1.0 \pm 0.86 \pm 0.14) \times 10^{-10}, \text{ and}$$

$$\mathcal{R}_{\mu^+\mu^-} = 0.028 \pm 0.025 \pm 0.004$$

is expected (assuming SM branching fraction values), which is close to that of the recent combination of the LHC experiments. If the low values measured by the combination persist, a new combination of the measurement performed in this thesis and the updates from the other LHC experiments is expected to show an evidence of an inconsistency with the SM.

The analyses of  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow e^+e^-$  decays have been performed only with a subset of the available LHCb data set. Analysing the full data collected with the LHCb experiment so far and making use of improved selection techniques will allow to lower the current upper limits already significantly. The new data taking period with an upgraded detector beginning in 2021/2022 yields exciting prospects of even stronger improvement. The analyses of  $B_{(s)}^0 \rightarrow e^+e^-$  and  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays will largely constrain the space of branching fraction values allowed in NP scenarios. If instead one of the decays would be observed, this would be a clear sign for a NP process and would change the understanding of particle physics fundamentally. The measurement of the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction will reach an unprecedented precision and will give a clear answer whether it receives sizeable contributions from NP processes. If the branching fraction of  $B^0 \rightarrow \mu^+\mu^-$  decays is close to the SM value, the increased data set will allow to observe also  $B^0 \rightarrow \mu^+\mu^-$  transitions. This will impose strong constraints on NP contributions in  $b \rightarrow d\ell^+\ell^-$  processes. The precisely measured ratio of the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fractions,  $\mathcal{R}_{\mu^+\mu^-}$ , will provide a clean test of the Minimal Flavour Violation hypothesis.

## Bibliography

- [1] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.
- [2] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [3] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. C **680519** (1968) 367.
- [4] 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.
- [5] 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.
- [6] A. J. Buras and J. Girrbach, *Towards the identification of New Physics through quark-flavour violating processes*, Rept. Prog. Phys. **77** (2014) 086201, arXiv:1306.3775.
- [7] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays*, JHEP **08** (2017) 055, arXiv:1705.05802.
- [8] LHCb collaboration, R. Aaij *et al.*, *Search for lepton-universality violation in  $B^+ \rightarrow K^+ \ell^+ \ell^-$  decays*, Phys. Rev. Lett. **122** (2019) 191801, arXiv:1903.09252.
- [9] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using  $A_b^0 \rightarrow p K^- \ell^+ \ell^-$  decays*, JHEP **05** (2020) 040, arXiv:1912.08139.
- [10] M. Algueró *et al.*, *Emerging patterns of New Physics with and without lepton-flavour-universal contributions*, Eur. Phys. J. C **79** (2019), no. 8 714, arXiv:1903.09578, [Addendum: Eur.Phys.J.C 80, 511 (2020)].
- [11] J. Aebischer *et al.*, *B-decay discrepancies after Moriond 2019*, Eur. Phys. J. C **80** (2020), no. 3 252, arXiv:1903.10434.
- [12] M. Ciuchini *et al.*, *New Physics in  $b \rightarrow s \ell^+ \ell^-$  confronts new data on lepton universality*, Eur. Phys. J. C **79** (2019), no. 8 719, arXiv:1903.09632.
- [13] A. K. Alok, A. Dighe, S. Gangal, and D. Kumar, *Continuing search for New Physics in  $b \rightarrow s \mu^+ \mu^-$  decays: two operators at a time*, JHEP **06** (2019) 089, arXiv:1903.09617.

- [14] A. Arbey *et al.*, *Update on the  $b \rightarrow s$  anomalies*, Phys. Rev. D **100** (2019), no. 1 015045, arXiv:1904.08399.
- [15] S. L. Glashow, D. Guadagnoli, and K. Lane, *Lepton flavor violation in B-decays?*, Phys. Rev. Lett. **114** (2015) 091801, arXiv:1411.0565.
- [16] M. Beneke, C. Bobeth, and R. Szafron, *Power-enhanced leading-logarithmic QED corrections to  $B_q^0 \rightarrow \mu^+ \mu^-$* , JHEP **10** (2019) 232, arXiv:1908.07011.
- [17] W. Altmannshofer, C. Niehoff, and D. M. Straub,  *$B_s^0 \rightarrow \mu^+ \mu^-$  as current and future probe of new physics*, JHEP **05** (2017) 076, arXiv:1702.05498.
- [18] D. Griffiths, *Introduction to elementary particles*, Wiley VCH Weinheim, Germany, 2008.
- [19] C. Berger, *Elementarteilchenphysik: Von den Grundlagen zu den modernen Experimenten*, Springer-Lehrbuch, Springer Berlin-Heidelberg, Germany, 2014.
- [20] T. Mombächer, *Studies for the measurement of the decay width difference  $\Delta\Gamma_d$  in the  $B^0$ -meson system with the LHCb experiment*, Master's thesis, Technische Universität Dortmund, Germany, September, 2016.
- [21] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001, and 2019 update.
- [22] LHCb collaboration, R. Aaij *et al.*, *Observation of  $J/\psi p$  resonances consistent with pentaquark states in  $A_b^0 \rightarrow J/\psi p K^-$  decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv:1507.03414.
- [23] LHCb collaboration, R. Aaij *et al.*, *Observation of exotic  $J/\psi \phi$  structures from amplitude analysis of  $B^+ \rightarrow J/\psi \phi K^+$  decays*, Phys. Rev. Lett. **118** (2017) 022003, arXiv:1606.07895.
- [24] Super-Kamiokande collaboration, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:9807003.
- [25] SNO collaboration, Q. R. Ahmad *et al.*, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, arXiv:nucl-ex/0204008.
- [26] KamLAND collaboration, K. Eguchi *et al.*, *First results from KamLAND: Evidence for reactor anti-neutrino disappearance*, Phys. Rev. Lett. **90** (2003) 021802, arXiv:0212021.
- [27] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.
- [28] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.



- 
- [29] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, Sov. Phys. JETP **7** (1958) 172.
- [30] B. Pontecorvo, *Mesonium and anti-mesonium*, Sov. Phys. JETP **6** (1957) 429.
- [31] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28** (1962) 870.
- [32] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.
- [33] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [34] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.
- [35] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (1996) 1125, arXiv:hep-ph/9512380.
- [36] 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.
- [37] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of  $B \rightarrow K^{(*)} \mu^+ \mu^-$  decays*, JHEP **06** (2014) 133, arXiv:1403.8044.
- [38] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of  $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$  decays*, JHEP **06** (2015) 115, Erratum *ibid.* **09** (2018) 145, arXiv:1503.07138.
- [39] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay  $B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **09** (2015) 179, arXiv:1506.08777.
- [40] LHCb collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in  $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$  decays and the  $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$  differential branching fraction*, JHEP **11** (2016) 047, Erratum *ibid.* **04** (2017) 142, arXiv:1606.04731.
- [41] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.
- [42] Belle collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of  $B^0 \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. Lett. **118** (2017), no. 11 111801, arXiv:1612.05014.
- [43] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of  $B_d^0 \rightarrow K^{*0} \mu^+ \mu^-$  decays in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, JHEP **10** (2018) 047, arXiv:1805.04000.

- [44] CMS collaboration, V. Khachatryan *et al.*, *Angular analysis of the decay  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  from  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, Phys. Lett. B **753** (2016) 424, arXiv:1507.08126.
- [45] Belle collaboration, A. Abdesselam *et al.*, *Test of lepton flavor universality in  $B \rightarrow K^* \ell^+ \ell^-$  decays at Belle*, arXiv:1904.02440.
- [46] Belle collaboration, A. Abdesselam *et al.*, *Test of lepton flavor universality in  $B \rightarrow K \ell^+ \ell^-$  decays*, arXiv:1908.01848.
- [47] BaBar collaboration, J. P. Lees *et al.*, *Measurement of branching fractions and rate asymmetries in the rare decays  $B \rightarrow K^{(*)} \ell^+ \ell^-$* , Phys. Rev. D **86** (2012) 032012, arXiv:1204.3933.
- [48] T. Hurth, F. Mahmoudi, and S. Neshatpour, *On the new LHCb angular analysis of  $B \rightarrow K^* \mu^+ \mu^-$ : hadronic effects or New Physics?*, arXiv:2006.04213.
- [49] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of  $b$ -hadron,  $c$ -hadron, and  $\tau$ -lepton properties as of 2018*, arXiv:1909.12524, updated results and plots available at <https://hflav.web.cern.ch>.
- [50] M. Raidal *et al.*, *Flavour physics of leptons and dipole moments*, Eur. Phys. J. C **57** (2008) 13, arXiv:0801.1826.
- [51] A. Crivellin, D. Müller, A. Signer, and Y. Ulrich, *Correlating lepton-flavor universality violation in  $B$  decays with  $\mu \rightarrow e \gamma$  using leptoquarks*, Phys. Rev. D **97** (2018), no. 1 015019, arXiv:1706.08511.
- [52] I. de Medeiros Varzielas and G. Hiller, *Clues for flavor from rare lepton and quark decays*, JHEP **06** (2015) 072, arXiv:1503.01084.
- [53] G. Hiller, D. Loose, and K. Schönwald, *Leptoquark flavor patterns &  $B$ -decay anomalies*, JHEP **12** (2016) 027, arXiv:1609.08895.
- [54] A. Crivellin *et al.*, *Lepton-flavour violating  $B$ -decays in generic  $Z'$  models*, Phys. Rev. D **92** (2015), no. 5 054013, arXiv:1504.07928.
- [55] S. M. Boucenna, J. W. F. Valle, and A. Vicente, *Are the  $B$ -decay anomalies related to neutrino oscillations?*, Phys. Lett. B **750** (2015) 367, arXiv:1503.07099.
- [56] D. Guadagnoli and K. Lane, *Charged-lepton mixing and lepton-flavor violation*, Phys. Lett. B **751** (2015) 54, arXiv:1507.01412.
- [57] BaBar collaboration, B. Aubert *et al.*, *Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays  $B \rightarrow K \ell^+ \ell^-$  and  $B \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. D **73** (2006) 092001, arXiv:0604007.
- [58] W. Altmannshofer, P. Paradisi, and D. M. Straub, *Model-independent constraints on New Physics in  $b \rightarrow s$  transitions*, JHEP **04** (2012) 008, arXiv:1111.1257.

- 
- [59] F. Beaujean, C. Bobeth, D. van Dyk, and C. Wacker, *Bayesian fit of exclusive  $b \rightarrow s\ell\ell$  Decays: The Standard Model operator basis*, JHEP **08** (2012) 030, arXiv:1205.1838.
- [60] UTfit collaboration, M. Bona *et al.*, *The unitarity triangle fit in the standard model and hadronic parameters from lattice QCD: A reappraisal after the measurements of  $\Delta m_s$  and  $BR(B \rightarrow \tau\nu_\tau)$* , JHEP **10** (2006) 081, arXiv:hep-ph/0606167, updated results and plots available at <http://www.utfit.org/>.
- [61] CKMfitter group, J. Charles *et al.*, *Current status of the standard model CKM fit and constraints on  $\Delta F = 2$  new physics*, Phys. Rev. **D91** (2015) 073007, arXiv:1501.05013, updated results and plots available at <http://ckmfitter.in2p3.fr/>.
- [62] Flavour Lattice Averaging Group, S. Aoki *et al.*, *FLAG Review 2019: Flavour Lattice Averaging Group (FLAG)*, Eur. Phys. J. C **80** (2020), no. 2 113, arXiv:1902.08191.
- [63] A. J. Buras, *Relations between  $\Delta M(s, d)$  and  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  in models with minimal flavor violation*, Phys. Lett. B **566** (2003) 115, arXiv:hep-ph/0303060.
- [64] D. King, A. Lenz, and T. Rauh,  *$B_s^0$ -mixing observables and  $|V_{td}/V_{ts}|$  from sum rules*, JHEP **05** (2019) 034, arXiv:1904.00940.
- [65] C. Hamzaoui, M. Pospelov, and M. Toharia, *Higgs mediated FCNC in supersymmetric models with large  $\tan\beta$* , Phys. Rev. D **59** (1999) 095005, arXiv:hep-ph/9807350.
- [66] S. R. Choudhury and N. Gaur, *Dileptonic decay of  $B_{(s)}^0$  mesons in SUSY models with large  $\tan\beta$* , Phys. Lett. B **451** (1999) 86, arXiv:hep-ph/9810307.
- [67] K. S. Babu and C. F. Kolda, *Higgs mediated  $B^0 \rightarrow \mu^+\mu^-$  in minimal supersymmetry*, Phys. Rev. Lett. **84** (2000) 228, arXiv:hep-ph/9909476.
- [68] R. Fleischer, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi, *In pursuit of New Physics with  $B_{s,d}^0 \rightarrow \ell^+\ell^-$* , JHEP **05** (2017) 156, arXiv:1703.10160.
- [69] A. J. Buras, F. De Fazio, and J. Girrbach, *The anatomy of  $Z'$  and  $Z$  with flavour changing neutral currents in the flavour precision Era*, JHEP **02** (2013) 116, arXiv:1211.1896.
- [70] A. J. Buras, J. Girrbach-Noe, C. Niehoff, and D. M. Straub,  *$B \rightarrow K^{(*)}\nu\bar{\nu}$  decays in the Standard Model and beyond*, JHEP **02** (2015) 184, arXiv:1409.4557.
- [71] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An effective field theory approach*, Nucl. Phys. B **645** (2002) 155, arXiv:hep-ph/0207036.

- [72] K. De Bruyn *et al.*, *Probing New Physics via the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.
- [73] LHCb collaboration, R. Aaij *et al.*, *Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and effective lifetime and search for  $B^0 \rightarrow \mu^+ \mu^-$  decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv:1703.05747.
- [74] CDF collaboration, T. Aaltonen *et al.*, *Search for the decays  $B_{(s)}^0 \rightarrow e^+ \mu^-$  and  $B_{(s)}^0 \rightarrow e^+ e^-$  in CDF Run II*, Phys. Rev. Lett. **102** (2009) 201801, arXiv:0901.3803.
- [75] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of properties of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and search for  $B^0 \rightarrow \mu^+ \mu^-$  with the CMS experiment*, JHEP **04** (2020) 188, arXiv:1910.12127.
- [76] 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.
- [77] ATLAS, CMS and LHCb collaborations, *Combination of the ATLAS, CMS and LHCb results on the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays*, LHCb-CONF-2020-002. Appearing as ATLAS-CONF-2020-049, CMS-PAS-BPH-20-003, LHCb-CONF-2020-002.
- [78] M. Kenzie *et al.*, *GammaCombo: A statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals*, doi: 10.5281/zenodo.3371421.
- [79] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle  $\gamma$  from a combination of LHCb results*, JHEP **12** (2016) 087, arXiv:1611.03076.
- [80] F. James, *Statistical methods in experimental physics*, World Scientific, Singapore, 2006.
- [81] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, Erratum *ibid.* **C73** (2013) 2501, arXiv:1007.1727.
- [82] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943), no. 3 426.
- [83] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.
- [84] G. J. Feldman and R. D. Cousins, *A unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873, arXiv:physics/9711021.

- 
- [85] L. Calefice, *Untersuchung likelihood-basierter Teststatistiken zur Bestimmung von oberen Ausschlussgrenzen*, Bachelor's thesis, Technische Universität Dortmund, June, 2017. English title: "Investigation of likelihood based test statistics for the determination of upper limits".
- [86] A. L. Read, *Presentation of search results: The  $CL_S$  technique*, J. Phys. **G28** (2002) 2693.
- [87] M. Kenzie *et al.*, *GammaCombo framework for combinations of measurements and computation of confidence intervals: Public Release v1.1*, Aug, 2019. doi: 10.5281/zenodo.3373613.
- [88] LHCb collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decays  $B^+ \rightarrow K^+ \mu^\pm e^\mp$* , Phys. Rev. Lett. **123** (2019) 231802, arXiv:1909.01010.
- [89] S. Bodhisattva, M. Walker, and M. Woodroffe, *On the unified method with nuisance parameters*, Statist. Sinica **19** (2009) 301.
- [90] C.-S. Chuang and T. L. Lai, *Hybrid resampling methods for confidence intervals*, Statist. Sinica **10** (2000) 1.
- [91] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008), no. LHCb-DP-2008-001 S08005.
- [92] L. Evans and P. Bryant, *LHC machine*, JINST **3** (2008) S08001.
- [93] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [94] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [95] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [96] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.
- [97] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007.
- [98] ALICE collaboration, B. Abelev *et al.*, *Technical Design Report for the Upgrade of the ALICE Inner Tracking System*, J. Phys. G **41** (2014) 087002.
- [99] LHCb collaboration, R. Aaij *et al.*, *Measurement of  $J/\psi$  production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, Eur. Phys. J. **C71** (2011) 1645, arXiv:1103.0423.
- [100] LHCb collaboration, R. Aaij *et al.*, *Production of  $J/\psi$  and  $\Upsilon$  mesons in  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, JHEP **06** (2013) 064, arXiv:1304.6977.

- [101] LHCb collaboration, R. Aaij *et al.*, *Measurement of forward  $J/\psi$  production cross-sections in  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, JHEP **10** (2015) 172, Erratum *ibid.* **05** (2017) 063, arXiv:1509.00771.
- [102] LHCb collaboration, C. Elsässer,  *$\bar{b}b$  production angle plots*, [https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb\\_ProductionAngles.html](https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html).
- [103] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.
- [104] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.
- [105] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819.
- [106] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.
- [107] LHCb RICH collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Meth. A **876** (2017) 221, arXiv:1703.08152.
- [108] C. Abellán Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, arXiv:2008.11556.
- [109] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.
- [110] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for multivariate data analysis with ROOT*, PoS **ACAT** (2007) 040.
- [111] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.
- [112] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019), no. LHCb-DP-2019-001 P04013, arXiv:1812.10790.
- [113] D. A. Berninghoff, J. Albrecht, and V. Gligorov, *Bremsstrahlung recovery of electrons using multivariate methods*, tech. rep., CERN, Geneva, April, 2016.
- [114] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
- [115] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

- 
- [116] C.-H. Chang, C. Driouichi, P. Eerola, and X. G. Wu, *BCVEGPY: An event generator for hadronic production of the  $B_c^+$ -meson*, *Comput. Phys. Commun.* **159** (2004) 192, arXiv:hep-ph/0309120.
- [117] C.-H. Chang, J.-X. Wang, and X.-G. Wu, *BCVEGPY2.0: An upgrade version of the generator BCVEGPY with an addendum about hadroproduction of the P-wave B(c)-states*, *Comput. Phys. Commun.* **174** (2006) 241, arXiv:hep-ph/0504017.
- [118] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.
- [119] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z- and W-decays*, *Eur. Phys. J.* **C45** (2006) 97, arXiv:hep-ph/0506026.
- [120] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.
- [121] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [122] LHCb collaboration, *The BOOLE project*, June, 2020.
- [123] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays  $B_s^0 \rightarrow e^+e^-$  and  $B^0 \rightarrow e^+e^-$* , *Phys. Rev. Lett.* **124** (2020) 211802, arXiv:2003.03999.
- [124] A. Battig, *Search for the decay  $B^+ \rightarrow K^+\mu^-e^+$  at the LHCb experiment*, Bachelor's thesis, Technische Universität Dortmund, July, 2015. Performed in coordination with the Bachelor's thesis by Niklas S. Nolte.
- [125] N. S. Nolte, *Search for the decay  $B^+ \rightarrow K^+\mu^-e^+$  at the LHCb experiment*, Bachelor's thesis, Technische Universität Dortmund, July, 2015. Performed in coordination with the Bachelor's thesis by Alexander Battig.
- [126] D. Berninghoff, *Search for lepton-flavour violation and optimisation of electron reconstruction at LHCb*, Master's thesis, Technische Universität Dortmund, June, 2016.
- [127] N. Scharmberg, *Search for the rare decay  $B_s^0 \rightarrow e^+e^-$  at the LHCb experiment*, Master's thesis, Technische Universität Dortmund, September, 2016.
- [128] A. Battig, *Search for the rare decay  $B_s^0 \rightarrow e^+e^-$  with the LHCb experiment*, Master's thesis, Technische Universität Dortmund, December, 2017.
- [129] M. M. Pikies, *Searches for New Physics in  $b \rightarrow s\ell^+\ell^-$  transitions at the LHCb experiment*, PhD thesis, Henryk Niewodniczański Institute of Nuclear Physics, Krakow, July, 2018, <https://rifj.ifj.edu.pl/handle/item/304>.
- [130] A. Battig, *Search for the  $B_{(s)}^0 \rightarrow e^+e^-$  and  $B^+ \rightarrow K^+\mu^\pm e^\mp$  decays with the LHCb experiment*, PhD thesis, Technische Universität Dortmund, thesis in progress.

- [131] M. Pivk and F. R. Le Diberder, *Plot : a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.
- [132] 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.
- [133] A. Rogozhnikov, *Reweighting with boosted decision trees*, J. Phys. Conf. Ser. **762** (2016), no. 1, arXiv:1608.05806, [https://github.com/arogozhnikov/hep\\_ml](https://github.com/arogozhnikov/hep_ml).
- [134] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021.
- [135] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of online learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.
- [136] A. Hoecker *et al.*, *TMVA 4 — Toolkit for multivariate data analysis with ROOT. Users guide.*, arXiv:physics/0703039.
- [137] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in python*, J. Machine Learning Res. **12** (2011) 2825, arXiv:1201.0490, and online at <http://scikit-learn.org/stable/>.
- [138] M. Stone, *Cross-validatory choice and assessment of statistical predictions*, J. R. Stat. Soc. B **36** (1974), no. 2 111.
- [139] L. Gavardi, *Search for lepton flavour violation in  $\tau$  decays at the LHCb experiment*, Master's thesis, Università degli studi di Milano-Bicocca, Nov, 2013, presented 28 Nov 2013, CERN-THESIS-2013-259.
- [140] LHCb, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, arXiv:0912.4179.
- [141] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.
- [142] LHCb collaboration, R. Aaij *et al.*, *Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and search for  $B^0 \rightarrow \mu^+ \mu^-$  decays at the LHCb experiment*, Phys. Rev. Lett. **111** (2013) 101805, arXiv:1307.5024.
- [143] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039.
- [144] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio  $f_s/f_d$  and its dependence on B meson kinematics*, JHEP **04** (2013) 001, arXiv:1301.5286.
- [145] LHCb collaboration, R. Aaij *et al.*, *Measurement of  $f_s/f_u$  variation with proton-proton collision energy and B-meson kinematics*, Phys. Rev. Lett. **124** (2020) 122002, arXiv:1910.09934.



- 
- [146] J. M. Flynn *et al.*,  $B \rightarrow \pi \ell \nu$  and  $B_s \rightarrow K \ell \nu$  form factors and  $|V_{ub}|$  from 2+1-flavor lattice QCD with domain-wall light quarks and relativistic heavy quarks, Phys. Rev. **D91** (2015), no. 7 074510, arXiv:1501.05373.
- [147] C. M. Bouchard *et al.*,  $B_s \rightarrow K \ell \nu$  form factors from lattice QCD, Phys. Rev. **D90** (2014) 054506, arXiv:1406.2279.
- [148] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays  $B/B_s \rightarrow (\pi, K)(\ell^+ \ell^-, \ell \nu, \nu \bar{\nu})$  in the perturbative QCD approach beyond the leading-order*, Phys. Rev. **D86** (2012) 114025, arXiv:1207.0265.
- [149] A. Kozachuk, D. Melikhov, and N. Nikitin, *Rare FCNC radiative leptonic  $B_{s,d}^0 \rightarrow \gamma \ell^+ \ell^-$  decays in the Standard Model*, Phys. Rev. **D97** (2018), no. 5 053007, arXiv:1712.07926.
- [150] P. Ball and R. Zwicky,  $B_{d,s}^0 \rightarrow \rho, \omega, K^*, \phi$  decay form-factors from light-cone sum rules revisited, Phys. Rev. **D71** (2005) 014029, arXiv:hep-ph/0412079.
- [151] V. V. Kiselev, *Exclusive decays and lifetime of  $B_c^+$ -mesons in QCD sum rules*, arXiv:hep-ph/0211021.
- [152] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: an update*, Phys. Rev. **D52** (1995) 2783, arXiv:hep-ph/9503486.
- [153] A. Khodjamirian and A. V. Rusov,  $B_s^0 \rightarrow K \ell \nu_\ell$  and  $B_{(s)}^0 \rightarrow \pi(K) \ell^+ \ell^-$  decays at large recoil and CKM matrix elements, JHEP **08** (2017) 112, arXiv:1703.04765.
- [154] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489.
- [155] LHCb collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decays  $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$* , JHEP **03** (2018) 078, arXiv:1710.04111.
- [156] B. Efron, *Bootstrap methods: another look at the jackknife*, Ann. Statist. **7** (1979) 1.
- [157] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.
- [158] R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, arXiv:1909.02957, submitted to JINST.
- [159] R. H. Dalitz, *On the analysis of  $\tau$ -meson data and the nature of the  $\tau$ -meson*, Phil. Mag. Ser. 7 **44** (1953) 1068.
- [160] M. Perrin-Terrin, G. Mancinelli, F. Dettori, and J. Serrano, *Towards a model independent branching ratio measurement of  $B_s^0 \rightarrow \mu^+ \mu^-$* , LHCb-INT-2013-012.

- [161] F. Dettori, D. Guadagnoli, and M. Reboud,  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  from  $B_s^0 \rightarrow \mu^+ \mu^-$ , Phys. Lett. B **768** (2017) 163, arXiv:1610.00629.
- [162] F. Archilli *et al.*, *Background studies for  $B^0 \rightarrow \mu^+ \mu^-$  analysis optimization*, LHCb-INT-2014-047.
- [163] LHCb collaboration, R. Aaij *et al.*, *Combination of  $f_s/f_d$  measurements at LHCb*, Measurement in preparation.
- [164] LHCb collaboration, R. Aaij *et al.*, *Measurement of the  $B^\pm$  production cross-section in  $pp$  collisions at  $\sqrt{s} = 7$  and  $13$  TeV*, JHEP **12** (2017) 026, arXiv:1710.04921.
- [165] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays  $B_{(s)}^0 \rightarrow \ell^+ \ell^- \gamma$* , Phys. Rev. **D70** (2004) 114028, arXiv:hep-ph/0410146.
- [166] LHCb collaboration, R. Aaij *et al.*, *Measurement of the  $B_c^-$  production fraction and asymmetry in 7 and 13 TeV  $pp$  collisions*, Phys. Rev. **D100** (2019) 112006, arXiv:1910.13404.
- [167] D. Ebert, R. N. Faustov, and V. O. Galkin, *Weak decays of the  $B_c^+$  meson to charmonium and  $D$  mesons in the relativistic quark model*, Phys. Rev. **D68** (2003) 094020, arXiv:hep-ph/0306306.
- [168] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001.
- [169] *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, .
- [170] 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.
- [171] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.
- [172] LHCb collaboration, *LHCb Trigger and Online Technical Design Report*, CERN-LHCC-2014-016.
- [173] LHCb collaboration, *LHCb PID Upgrade Technical Design Report*, CERN-LHCC-2013-022.
- [174] J. G. Speer, *Suche nach  $B_s^0 \rightarrow e^+ e^-$  mit LHCb Daten aus den Jahren 2016 bis 2018*, Bachelor's thesis, Technische Universität Dortmund, Germany, July, 2019. English title: "Search for  $B_s^0 \rightarrow e^+ e^-$  with LHCb data of the years 2016 to 2018".
- [175] F. Koch, *Analyse des leptonzahlverletzenden Zerfalls  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  mit Run 2 Daten des LHCb-Experiments*, Bachelor's thesis, Technische Universität Dortmund, Germany, December, 2019. English title: "Analysis of the lepton-flavour violating decays  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  with Run 2 data of the LHCb experiment".

- [176] G. Meier, *Suche nach Verletzung der Leptonflavourzahl mit dem Zerfall  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  am LHCb-Experiment*, Master's thesis, Technische Universität Dortmund, Germany, September, 2018. English title: "Search for lepton-flavour violation with  $B^+ \rightarrow K^+ \mu^\pm e^\mp$  decays at the LHCb experiment".
- [177] J. Wendel, *Suche nach Leptonflavourzahlverletzung in  $B^+ \rightarrow K^+ \mu^\pm e^\mp$ : Studien zur multivariaten Selektion*, Bachelor's thesis, Technische Universität Dortmund, Germany, January, 2020. English title: "Search for lepton-flavour violation in  $B^+ \rightarrow K^+ \mu^\pm e^\mp$ : studies for the multivariate selection".
- [178] CMS collaboration, *Search for a neutral MSSM Higgs boson decaying into  $\tau\tau$  with  $12.9 \text{ fb}^{-1}$  of data at  $\sqrt{s} = 13 \text{ TeV}$* , CMS-PAS-HIG-16-037.
- [179] ATLAS collaboration, *Search for Minimal Supersymmetric Standard Model Higgs Bosons  $H/A$  in the  $\tau\tau$  final state in up to  $13.3 \text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS Detector*, ATLAS-CONF-2016-085.



## A Appendix: Selection of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays

This appendix gives a detailed review of the selection of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  analysis. It consists of an experiment-wide preselection on top of which some loose requirements are added on track and vertex qualities, as well as vertex displacement from the PV and fiducial kinematic requirements with the aim to reduce the data set to a manageable size while maintaining a high signal efficiency. A loose requirement on a preselection BDT “BDTS” is applied to remove large background before fine-tuning the selection. All requirements are listed in Table A.1. The input variables of the BDTS classifier are

- the impact parameter  $\text{IP}(B)$  and impact parameter  $\chi_{\text{IP}}^2(B)$  of the  $B$  candidate;
- the  $\chi_{\text{vtx}}^2$  of the secondary vertex;
- the angle between the direction of the momentum of the  $B$  candidate and the direction defined by the secondary and the primary vertices (DIRA);
- the minimum distance between the two daughter tracks (DOCA);
- the minimum impact parameter of the muons with respect to any primary vertex ( $\text{minIP}(\mu)$ ).

Additionally, a veto of partially reconstructed charmonium decays,  $\Delta m_{J/\psi}$ , is applied, which requires the invariant mass combination of a signal muon and any other non-signal muon track to be outside the  $J/\psi$  mass region, where the non-signal muon track is identified as muon via requiring  $\text{ProbNN}_\mu > 0.3$ .

A most inclusive trigger selection is applied to select  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B^+ \rightarrow K^+ J/\psi$  decays in order to maximise the efficiency for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. To have a signature in the trigger as similar as possible to the signal  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays,  $B^0 \rightarrow K^+ \pi^-$  candidates are selected by requiring an inclusive TIS selection at the L0 and HLT1 stages and requiring TOS on topological two-body trigger lines at the HLT2 stage.

To suppress backgrounds from particle misidentification, stringent PID requirements are imposed on the muons via

$$\text{PID}_{\mu,4} \equiv \text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) > 0.4, \quad (\text{A.1})$$

for Run 1 and 2015, and

$$\text{PID}_{\mu,8} \equiv \text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) > 0.8, \quad (\text{A.2})$$

**Table A.1:** Preselection for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow bb^{(\prime)}$  and  $B^+ \rightarrow K^+ J/\psi$  channels. VDS is the secondary vertex flight distance significance. Cut values in parentheses for track  $\chi^2/\text{ndf}$  and ghost probability show softer cuts used starting from 2015 data taking.

Variable	applied to	Requirement	Requirement
		$B_{(s)}^0 \rightarrow \mu^+ \mu^-, B \rightarrow bb^{(\prime)}$	$B^+ \rightarrow K^+ J/\psi$
track $\chi^2/\text{ndf}$	$\mu/h$	$< 3$ ( $< 4$ )	$< 3$ ( $< 4$ )
ghost prob		$< 0.3$ ( $< 0.4$ )	$< 0.3$ ( $< 0.4$ )
DOCA		$< 0.3$ mm	$< 0.3$ mm
$\chi_{\text{IP}}^2$		$> 25$	$> 25$
$p$		$< 500$ GeV/ $c$	$< 500$ GeV/ $c$
$p_T$		$\in [0.25, 40]$ GeV/ $c$	$\in [0.25, 40]$ GeV/ $c$
ISMUON	$\mu$ only	true	true
$\Delta m_{J/\psi}$		$ m(\mu t) - m(J/\psi) $ $> 30$ MeV/ $c^2$	
$\chi_{\text{vtx}}^2$	$\mu\mu/hh$	$< 9$	$< 9$
VDS		$> 15$	$> 15$
$\Delta m$		$ m(bb, \mu\mu) - m_B $ $< 500$ MeV/ $c^2$	$ m(\mu\mu) - m_{J/\psi} $ $< 60$ MeV/ $c^2$
$\chi_{\text{IP}}^2$	$B$	$< 25$	$< 25$
$t$		$< 9 \times \tau(B_s^0)$	$< 9 \times \tau(B_s^0)$
$p_T$		$> 500$ MeV/ $c$	$> 500$ MeV/ $c$
$m$		$\in [4900, 6000]$ MeV/ $c^2$	$ m(J/\psi K) - m_{B^+} $ $< 100$ MeV/ $c^2$
$\chi_{\text{vtx}}^2$			$< 45$
BDTS		$> 0.05$	$> 0.05$

---

for the rest of Run 2. For Run 1 and 2015 data the MC12TuneV2 tune of ProbNN are used, while for the 2016, 2017 and 2018 data the MC15TuneV1 tune are chosen, which exploits an improved algorithm. Therefore compared to  $\text{PID}_{\mu 4}$  in Run 1, the  $\text{PID}_{\mu 8}$  requirement in Run 2 has the same or lower misID probabilities while keeping a similar muon PID efficiency of about 81 % (Run 1 and 2015) and 84 % (Run 2). The selection has been chosen to optimise the  $B^0 \rightarrow \mu^+ \mu^-$  signal sensitivity [162].

The final selection is applied by constructing a selection BDT with the inputs

- Long track isolation;
- VELO track isolation;
- $\chi_{\text{vtx}}^2$  of the  $B$  candidate;
- $\chi_{\text{IP}}^2$  of the  $B$  candidate with respect to the primary vertex;
- Cosine of the DIRA between the  $B$  direction and the vector joining the primary and secondary vertices;
- $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ , where  $\Delta\phi$  and  $\Delta\eta$  are the azimuthal angle and pseudorapidity differences between the two muons;
- $\min(\chi_{\text{IP}}^2)$  of the two muons with respect to the primary vertex associated to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidate;

where the track isolation variables are again BDT classifiers with the aim to provide a measure of how well the signal tracks are separated from other tracks that either traverse all tracking stations (long tracks) or are only measured in the VELO (VELO tracks). The input variables for these classifiers are listed in Table A.2.

No requirements are imposed on the selection BDT. Instead the BDT output distribution is transformed to be flat between 0 and 1 in signal simulation (background from random muon track combinations peaks at low BDT values) and then the final branching fraction measurement is obtained by simultaneously fitting the dimuon mass distribution in bins of the flat BDT output distribution.

**Table A.2:** Input variables to the track isolation BDTs. The isolation BDTs compute an isolation score for a signal muon track and any other long track (long track isolation) or VELO track (VELO track isolation) in the event. They are trained on isolating tracks in  $B_s^0 \rightarrow \mu^+ \mu^-$  simulation as a signal proxy and non-isolating tracks from inclusive  $b\bar{b} \rightarrow \mu^+ \mu^- X$  simulation as background proxy. Variables only used for the long track isolation are marked with an asterisk \*.

Track isolation inputs
$\min(\chi_{\text{IP}}^2)$ of the track $t$
signed distance between $t\mu$ -Vertex and PV
signed distance between $t\mu$ -Vertex and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ vertex
DOCA( $t, \mu$ )
angle between $t$ and $\mu$
the cone isolation $f_c$ (see Section 5.4.2)
* azimuthal angle difference between track and muon $\Delta\varphi(t, \mu)$
* pseudorapidity angle difference between track and muon $\Delta\eta(t, \mu)$
* track $p_{\text{T}}$



# Acknowledgements

Danke sagen – wo fängt man da an? Wie immer gehört mein erster Dank Gott selbst. Wir sehen seine Spuren, wenn wir die faszinierenden Grundgesetze erforschen, auf denen die von ihm erschaffene Natur basiert. Aber er ist auch derjenige, der mich jeden Tag aufrechterhält und mich stützt, besonders in stressigen Zeiten – und von denen gab es während der Promotion einige.

Ich danke besonders meinem Doktorvater PROF. Dr. Johannes Albrecht (noch einmal herzlichen Glückwunsch zur Professur!) für die doch recht unerwartete Gelegenheit dieser Promotion. Vielen Dank dafür, dass ich nicht nur an einem, sondern gleich an mehreren spannenden Projekten mitarbeiten durfte. Vielen Dank auch für die Geduld und Energie, die Projekte immer wieder auf ihr Ziel auszurichten und in stressigen Zeiten die Arbeit auf die wesentlichen Aspekte zu konzentrieren!

Einen wesentlichen Teil zu einer erfolgreichen Doktorarbeit trägt die Arbeitsgruppe bei. Und ich muss sagen, dass das freundliche und entspannte Miteinander in der E5a-Gruppe mir wirklich gut getan hat. Vielen Dank für die gemeinsame Zeit und die schönen Erinnerungen an Kanutouren (sorry an alle, die vielleicht doch ein wenig zu nass geworden sind...) und andere Erlebnisse. Danke an die Computing Admins, die sich zuverlässig um unsere Rechenkapazitäten gekümmert haben und die es ausbaden mussten, wenn ich in meiner Naivität schon mal das Dateisystem überlastet habe. Ein besonderer Dank geht an Alexander Battig und Maik Becker, mit denen ich viele physikalische (und nicht-physikalische) Diskussionen während der gemeinsamen Arbeit an den  $B_{(s)}^0 \rightarrow e^+e^-$ - und  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ -Analysen führen konnte.

A big thank you to all the great people that I met and collaborated with: Matt Kenzie on the GAMMACOMBO software and the proponents of the  $B^+ \rightarrow K^+\mu^\pm e^\mp$  and  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  analyses. You filled the physics we do with life and a personal atmosphere! Thank you all! Grazie mille a tutti! Dank u wel! Muchas gracias! Dziękuję bardzo!

Abschließend möchte ich mich bei meiner Familie bedanken, die mich ohne Vorbehalt unterstützt hat. Ihr seid das Rückgrat meiner Arbeit und habt mich immer wieder auf den Boden der Tatsachen zurückgebracht. Dafür bin ich euch sehr dankbar! Meine Großeltern möchte ich dabei nicht unerwähnt lassen: bei euch zu wohnen war nicht nur bequem. Nein, ich habe auch viel von den gemeinsamen Gesprächen profitiert und von eurem Vorbild gelernt!