technische universität
dortmund

Search for New Physics in rare decays of B mesons at the LHCb experiment

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

Vorgelegt von
Alexander Battig
geb. in Lüdenscheid

Lehrstuhl für Experimentelle Physik V
Fakultät Physik
Technische Universität Dortmund

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

Erstgutachter:	Prof. Dr. Johannes Albrecht
Zweitgutachter:	Prof. Dr. Kevin Kröninger
Vorsitz der Prüfungskommission:	Prof. Dr. Mark Aßmann
Vertreter der wissenschaftlichen Mitarbeiter*innen:	Dr. Ulf Berges
Datum der mündlichen Prüfung:	23. August 2022

Kurzfassung

Diese Arbeit präsentiert Suchen nach den rein leptonischen Zerfällen $B_s^0 \to e^+e^-$ und $B^0 \to e^+e^-$, ebenso wie den leptonfamilienzahl verletzenden Zerfällen $B^+ \to K^+e^\pm\mu^\mp$. Da diese Zerfälle im Standardmodell selten oder verboten sind ermöglichen sie Nulltests des Standardmodells. Außerdem ermöglichen Messungen dieser Kanäle es Modellen von Physik jenseits des Standardmodells Einschränkungen aufzuerlegen.

Die Suche nach den Zerfällen $B_s^0 \to e^+e^-$ und $B^0 \to e^+e^-$ nutzt Daten, welche vom LHCb Experiment in den Jahren 2011, 2012, 2015 und 2016 aufgezeichnet wurden und stellt die erste Suche am LHC nach diesen Zerfällen dar. Die bestimmten oberen Ausschlussgrenzen auf die Verzweigungsverhältnisse von $B_s^0 \to e^+e^-$ und $B^0 \to e^+e^-$ liegen in der Größenordnung $\mathcal{O}(10^{-8})$, was die vorhergegangenen Ausschlussgrenzen um einen Faktor 30 verbessert.

Die Suche nach den leptonfamilienzahl verletzenden Zerfällen $B^+ \to K^+e^\pm\mu^\mp$ nutzt den kompletten LHCb Datensatz, was obigen Datensätzen die Daten aus den Jahren 2017 und 2018 hinzufügt und so die Menge der B mesonen im analysierten Datensatz effektiv verdoppelt. Da diese Analyse noch nicht veröffentlicht wurde, wird der Status der Analyse zum Zeitpunkt der Abfassung dieses Textes präsentiert.

Abstract

In this thesis searches for the purely leptonic decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$ as well as the lepton-flavour violating decays $B^+ \to K^+e^\pm\mu^\mp$ are presented. As these decays are rare or forbidden in the Standard Model, these searches allow for null tests of the Standard Model. Measurements of these decays also allow to place constraints on scenarios of physics beyond the Standard Model.

The search for the decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$ uses data collected by the LHCb experiment in the years 2011, 2012, 2015 and 2016, it is the first search for these decays at the LHC. Upper limits on the branching fractions of $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$ are set at the order of $\mathcal{O}(10^{-8})$, improving the previous limits by a factor of 30.

The search for the lepton-flavour violating decays $B^+ \to K^+e^\pm\mu^\mp$ uses the full LHCb dataset, adding to the above the data collected in 2017 and 2018, effectively doubling the amount of B mesons in the analysed dataset. As the analysis is not yet published, the status as of the time of writing is presented.

Table of contents

vi

1 Introduction

The Standard Model of particle physics (SM) [1–3] describes three of the four fundamental forces, along with the particles interacting through these forces. The forces described are the electromagnetic interaction, the weak interaction and the strong interaction, while gravitation is not included in the SM. While many measurements have confirmed the predictions made by the SM, some questions remain unanswered. The matter described by the SM only makes up $\approx 5\,\%$ of the energy of the universe, with the rest consisting of dark matter and dark energy, the existence of which is known from cosmological measurements [4, 5], but is not described by the SM. Also the vast difference in masses between SM particles, varying from $\mathcal{O}(1\,\mathrm{eV}/c^2)$ to $\mathcal{O}(100\,\mathrm{GeV}/c^2)$ can not be explained by it, suggesting that the SM is incomplete.

Many searches for possible extensions of the SM, generally summarised as New Physics (NP), have been performed, with no clear observation of NP to date. Direct searches for new particles are limited by the energy available at the collider. At the Large Hadron Collider, particle masses of up to $\mathcal{O}(1\,\mathrm{TeV})$ are accessible. While particles with higher mass can not be produced directly, they could contribute to SM processes through loop contributions or interference, and thus change the observed rate compared to the SM. As NP particles do not need to be produced as on-shell states this way, much higher scales up to $\mathcal{O}(100\,\mathrm{TeV})$ can be accessed [6]. A particularly interesting environment for these searches are rare decays with low rates in the SM, which would allow even small NP contributions to have a visible effect.

Rare decays of B mesons provide one such environment to search for NP contributions. The LHCb experiment is designed for the study of b and c hadrons and so is well suited to search for NP in these processes. Indeed, measurements which study $b \to s\ell^+\ell^-$ transitions [1], like studies of lepton-flavour universality [7–10], show deviations from the SM. No single measurement alone shows significant enough deviations from the SM to claim the observation of NP, however when interpreted in a common framework strong tensions with the SM emerge [11].

---

[1]The inclusion of charge conjugated processes is implied throughout the thesis unless specified otherwise.

In this thesis, searches for the decays $B^0_{(s)} \to e^+e^-$ and $B^+ \to K^+e^\pm\mu^\mp$ are presented. These decays are rare (or in case of $B^+ \to K^+e^\pm\mu^\mp$ forbidden) in the SM, but might be enhanced significantly in NP models. The decays $B^0_{(s)} \to e^+e^-$ are particularly sensitive to new (pseudo-)scalar mediators. If the coupling of these mediators is strong enough to explain the deviations seen in lepton-flavour universality measurements, the branching fraction of $B^0_{(s)} \to e^+e^-$ decays might be enhanced to a level measurable with the currently available data. These observed deviations of lepton-flavour universality in $b \to s\ell^+\ell^-$ processes strongly imply the existence of charged lepton-flavour violation [12]. Such a process can be studied in decays like $B^+ \to K^+e^\pm\mu^\mp$, which are often predicted to have measurable rates when aiming to explain the measured deviations from the SM.

This thesis is structured the following way: at first the SM, as well the processes studied in this thesis, are introduced in more detail in Chapter 2. Chapter 3 then describes the LHCb detector, which was used to record the data the searches presented here were performed on. The searches for $B^0_{(s)} \to e^+e^-$ (in Chapter 4) and $B^+ \to K^+e^\pm\mu^\mp$ (in Chapter 5) decays are presented afterwards. The achieved results are then discussed in Chapter 6.

2 Theory

In the following chapter, the Standard Model of Particle Physics and its building blocks are introduced. The specific processes studied in this thesis are described afterwards in Sections 2.2 and 2.3.

2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) [1–3] contains the theories describing fundamental particles and their interaction. The fundamental particles of the SM are fermions with spin $1/2$, while interactions between particles are mediated by gauge bosons with spin 1. These interactions between particles are the strong interaction, weak interaction and electromagnetic interaction, each having a charge associated with them. For each fermion, an antiparticle of the same mass and opposite charges of the respective particle exists.

The fermions are grouped into two subgroups, differentiated by their electric charge. While leptons have integer charge (± 1 in case of charged leptons, 0 for neutrinos), quarks carry fractional charges of either $2/3$ in the case of up-type quarks or $-1/3$ in the case of down-type quarks. Furthermore, quarks and leptons occur in three generations each, ordered by their mass. The leptons in the SM are

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_{\text{LH}}, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_{\text{LH}}, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}_{\text{LH}},$$

where each generation contains a charged lepton and the corresponding neutrino. The subscript LH indicates, that only left-handed particles (and right-handed anti-particles) are described in the SM. These neutrinos were originally assumed to be massless in the SM, but measurements by the Super-Kamiokande collaboration in 1998 [13] and the SNO collaboration in 2002 [14] show oscillations between the three families, requiring mass differences between the neutrinos, meaning at most one of them can be massless. Current upper limits on the neutrino masses are of the order $\approx 1\text{eV}$ [15].

In the SM, leptons are produced in pairs of one lepton and one anti-lepton of the same generation, meaning that lepton number and lepton flavour are preserved. A similar property of the SM is the assumption of lepton-flavour universality of interactions including charged leptons, meaning that all three generations interact in the same way and their couplings to gauge bosons are equally strong. Recent LHCb measurements [7, 8] hint at a violation of this principle, which in turn motivates the violation of charged lepton-flavour conservation [12], which is studied in this thesis in the form of the decays $B^+ \rightarrow K^+ e^\pm \mu^\mp$.

Similarly to the leptons, quarks also occur in three generations, each consisting of an up-type quark and a down-type quark

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}.$$

In addition to their fractional electric charge, quarks carry the charge of the strong interaction, called colour, which can take one of the three states red, blue and green, where anti-quarks can carry the corresponding anti-charges. Due to the principle called confinement [16], quarks do not occur alone but rather in colour-neutral bound states called hadrons. They can be categorised as mesons, consisting of one quark and one anti-quark, and baryons, consisting of either three quarks or three anti-quarks, with states containing a higher number of quarks also allowed in principle. The existence of bound states containing four quarks (tetraquarks) has been reported by the Belle collaboration in 2008 [17] and has been confirmed by the LHCb collaboration in 2014 [18]. The first observation of states containing five quarks (pentaquarks) has been reported by the LHCb collaboration in 2015 [19], with many more searches performed since [20].

The SM takes the form of a $SU(3)_\mathrm{C} \times SU(2)_\mathrm{L} \times U(1)_\mathrm{Y}$ gauge theory, where interactions between particles result from invariance under local symmetry transformations. The $SU(3)_\mathrm{C}$ corresponds to the strong interaction, called quantum chromo dynamics (QCD), while the $SU(2)_\mathrm{L} \times U(1)_\mathrm{Y}$ describes electroweak interactions. The subscript "C" denotes the gauge symmetry for the colour charge, the subscript "L" stands for the gauge symmetry for the weak interaction of left-handed particles and the subscript "Y" refers to the gauge interactions for the hypercharge. Gravitation is negligible compared to the other forces at mass scales occurring in the SM and is not described by it.

Since QCD follows an SU(3) symmetry there are eight gauge bosons, called gluons. In contrast to the gauge bosons of the other SM interactions, gluons are charged under the interaction they mediate, allowing for gluon-gluon interactions.

The combination of weak and electromagnetic forces is described by the remaining SU(2)xU(1) part of the SM, where the photon γ (coupling to the electric charge of a particle) mediates the electromagnetic interaction (quantum electro dynamics, QED for short), while $W^{\pm}$ and Z bosons (coupling to the weak isospin) mediate the weak interaction. Because neutrinos carry neither electric nor colour charge, the weak interaction is the only force in the SM able to couple to neutrinos.

In the quark sector, changes in a quark's flavour only occur via the weak interaction by exchange of a $W^{\pm}$ boson. The amplitudes of quark transitions are summarised in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [21, 22]. This matrix transforms the flavour eigenstates (d, s, b) of the quarks into their mass eigenstates (d', s', b')

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.1}$$

The mixing of neutrinos is described by a similar mechanism via the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [23–25].

In 2012, the last missing particle of the SM, the Higgs boson, was discovered by the ATLAS and CMS collaborations [26, 27]. It is an excited state of the Higgs field, predicted by the theory describing the Brout-Englert-Higgs mechanism [28–30], which gives mass to the SM particles. The SM particles are summarised in Figure 2.1.

While the SM is complete with the Higgs boson, there remain open questions not answered by it, for example the strong hierarchy observed in the particle masses over the three generations, the origin and composition of dark matter or the observed imbalance of matter and anti-matter, which is not explainable just from the amount of CP violation predicted by the SM. These open questions necessarily imply the existence of New Physics (NP) beyond the SM, which might become visible in the study of processes strongly suppressed in the SM.

One category of processes where NP might be found are flavour-changing neutral current (FCNC) processes of the type $b \rightarrow s\ell^+\ell^-$ and $b \rightarrow d\ell^+\ell^-$, which are studied in the analyses presented in this thesis. These processes can not happen directly on quark level, as no flavour-changing neutral interactions exist in the SM, but instead happen via loops involving multiple intermediate charged processes. This kind of process is illustrated in Figure 2.2. As these processes are suppressed compared to tree-level processes, they are ideal probes for NP, where FCNC processes might be allowed on tree-level, thus changing the decay rate compared to the SM. Even
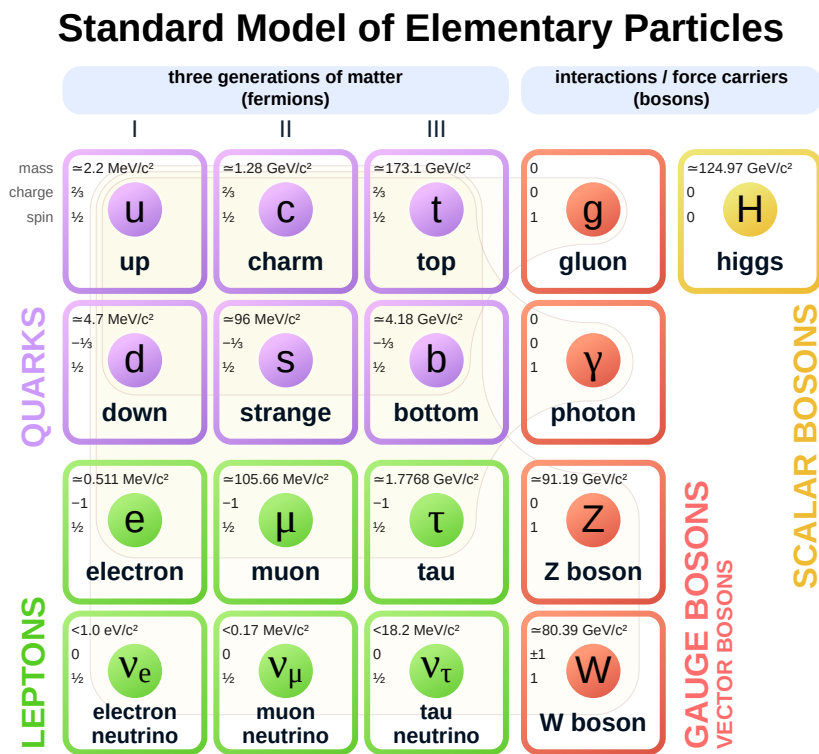
Figure 2.1: Particle content of the Standard Model of Particle Physics. For each particle, the spin, electric charge and mass are given. Figure reproduced from [31].

if NP effects are small compared to the SM, these changes might be visible in the interference between SM and NP.
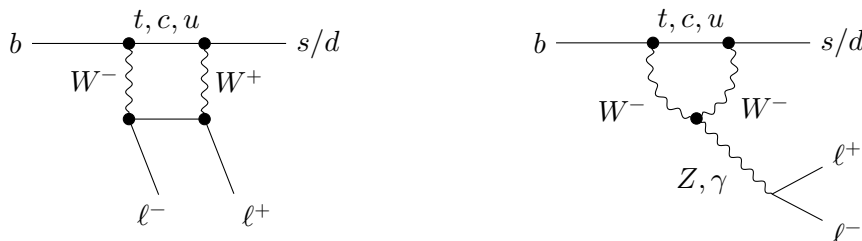

Figure 2.2: Dominant SM processes of $b \to s\ell^+\ell^-$ transitions, with the $W^\pm$ box diagram shown on the left and the penguin diagram shown on the right.

Similarly, processes forbidden in the SM might be allowed by NP, incentivising also searches for these forbidden processes "adjacent" to the SM. One category of processes like this would be the lepton-flavour violating $b \to s\ell^+\ell^{-\prime}$. Because quarks hadronise, this process can not be studied directly but rather in decays of hadrons like $B^+ \to K^+ e^\pm \mu^\mp$. Searches for the violation of charged lepton-flavour are especially well motivated by anomalies observed in studies of lepton-flavour universality, as the violation of lepton-flavour universality implies the violation of charged lepton-flavour in some NP models [12].

In order to easily quantify and combine measurements and possible deviations from the SM, the theoretical description of b hadron decays is formulated as an effective field theory via operator product expansion [32, 33], such that the amplitude of a B meson decaying into a final-state f via a $b \to s\ell^+\ell^-$ transition can be written as

$$\mathcal{A} = \langle f|\mathcal{H}|B\rangle = \frac{G_F}{\sqrt{2}} \sum_{i=7,9,10} \mathrm{V}_{tb}\mathrm{V}_{ts}^* \mathcal{C}_i(\mu)\langle f|\mathcal{O}_i(\mu)|B\rangle, \tag{2.2}$$

with the effective Hamiltonian $\mathcal{H}$, the Fermi constant G_F, the Wilson coefficients $\mathcal{C}_i$ and operators $\mathcal{O}_i$, which depend on the energy scale of the transition, here the B meson mass. The Wilson coefficients contain perturbatively calculable non-local contributions, while the operators describe local quark-level contributions. As the Wilson coefficients do not depend on the final state f, this formulation allows to measure them in various decay modes occurring via the $b \to s\ell^+\ell^-$ transition and thus combine information from these measurements. The relevant operators contributing to $b \to s\ell^+\ell^-$ transitions are $\mathcal{O}_7$ (photon radiation), $\mathcal{O}_9$ (vector transition) and $\mathcal{O}_{10}$ (axial-vector transition).

Recent measurements of $b \to s\ell^+\ell^-$ processes found deviations from the SM predictions in $b \to s\ell^+\ell^-$ branching fractions [34–36], lepton-universality ratios [7–10] and angular analyses [37–39]. None of these deviations is significant enough by itself to claim an observation of New Physics, but combining the measurements in global fits to the Wilson Coefficients show significant deviations from the SM in $\mathcal{C}_9$ and a small shift in $\mathcal{C}_{10}$. One such fit illustrating this shift is shown in Figure 2.3.
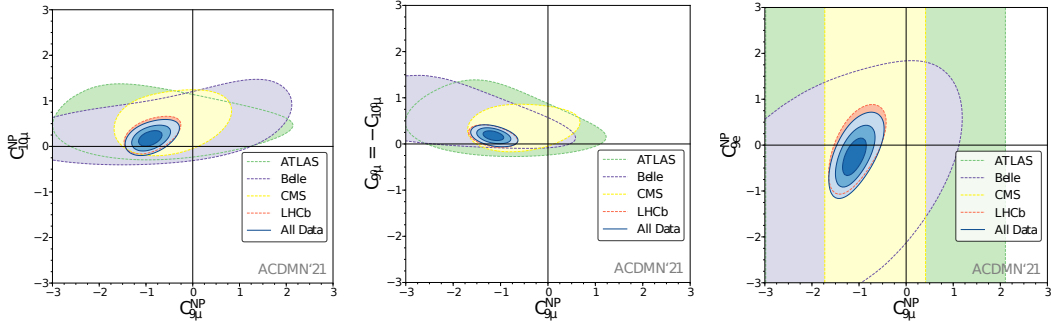
Figure 2.3: Fit to the current experimental status in $b \to s\ell^+\ell^-$ transitions, reproduced from Ref. [11]. The three plots show the allowed shift from the SM in the $(\mathcal{C}_{9,\mu}, \mathcal{C}_{10,\mu})$ plane (left), the $(\mathcal{C}_{9,\mu}, \mathcal{C}_{9,\mu} = -\mathcal{C}_{10,\mu})$ plane (middle) and the $(\mathcal{C}_{9,\mu}, \mathcal{C}_{9,e})$ plane (right) using all available data. The dashed lines represent the 3σ contours for each individual experiment, while the solid lines represent the 1, 2 and 3σ contours of the combination.

As stated before, measurements of lepton-flavour universality ratios show these deviations as well, implying differences in interactions involving muons and electrons in weak decays of b hadrons, contrasting measurements in Z bosons [40–44] and charmonia [45–47], where no such differences are observed.

A multitude of possible NP interactions beyond the SM have been proposed, with the most prominent featuring new heavy mediators like Z′ bosons and leptoquarks or an expanded Higgs sector. These NP scenarios commonly introduce tree-level FCNC processes that are not necessarily required to conserve lepton flavour, making studies of purely leptonic $b \to s\ell^+\ell^-$ transitions and lepton-flavour violating processes ideal places to look for this kind of NP as null-tests of the SM.

The following sections describe the current theoretical status of the decays $B^0_{(s)} \to e^+e^-$ in Section 2.2 and $B^+ \to K^+ e^\pm \mu^\mp$ in Section 2.3, both in the SM and selected NP models. They also summarise the current experimental status not taking into account the results produced in this thesis.

2.2 The decay modes $B^0_{(s)} \to e^+ e^-$

The following sections describe the theoretical basis for decays of the type $B^0_{(s)} \to \ell^+ \ell^-$ in the SM in Section 2.2.1 and possible sources of physics beyond the SM in the decays $B^0_{(s)} \to e^+ e^-$ in Section 2.2.2.

2.2.1 $B^0_{(s)} \to \ell^+ \ell^-$ decays in the Standard Model

In the SM, the decays $B^0_{(s)} \to \ell^+ \ell^-$ happen via flavour-changing neutral current (FCNC) processes, which can only happen at loop level, leading to a suppression of the branching fraction compared to a tree-level process by the CKM factor $|V_{tb}V^*_{t(s,d)}|^2$. In addition, $B^0_{(s)} \to \ell^+ \ell^-$ is helicity-suppressed. The helicity is defined as the projection of a particle's spin onto its momentum direction, where particles that have spin and momentum point in the same direction are called right-handed, while in the case that spin and momentum point in opposite directions they are called left-handed. In the SM, the weak interaction only couples to left-handed particles and right-handed antiparticles.

Since the $B^0_{(s)}$ meson has spin 0, the two leptons are required to have opposite spin. Thus in the centre-of-mass system on the $B^0_{(s)}$, where the lepton momenta have to sum to zero, one of the leptons ends up in the suppressed helicity state. The effect of the helicity suppression can be written as the squared ratio of the masses of initial and final state particles.

As the final state of $B^0_{(s)} \to \ell^+ \ell^-$ processes is purely leptonic, it can be computed very precisely in the SM, the dominant contributions are depicted in Figure 2.4.
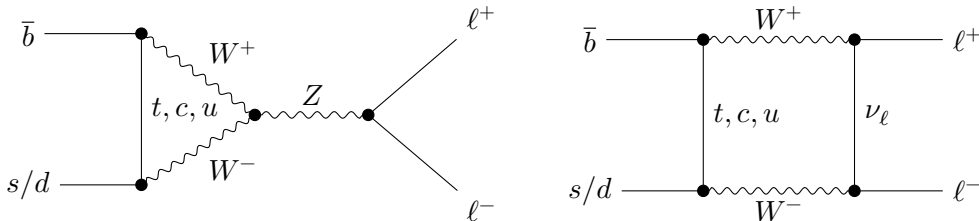

Figure 2.4: Dominant SM processes contributing in $B^0_{(s)} \to \ell^+ \ell^-$ decays. The penguin diagram is shown on the left, the $W^\pm$ box diagram on the right.

In the effective field theory approach described in Chapter 2, the branching fraction $\mathcal{B}$ of $B^0_{(s)} \to \ell^+ \ell^-$ decays takes the form [48, 49]

$$\mathcal{B} = \tau^*_{B^0_{(s)}} \frac{G_F^2 \alpha_{em}^2}{16\pi^2} f^2_{B^0_{(s)}} m_\ell^2 m_{B^0_{(s)}} \sqrt{1 - \frac{4m_\ell^2}{m^2_{B^0_{(s)}}}} |V_{tb}V^*_{t(s,d)}|^2 |\mathcal{C}_{10}^{SM}|^2 (|S|^2 + |P|^2), \quad (2.3)$$

where

$$\tau^*_{B^0_{(s)}} = \frac{1 + \mathcal{A}_{CP} \cdot y_{s,d}}{1 - y^2_{s,d}} \tau_{B^0_{(s)}}, \quad (2.4)$$

with the average $B^0_{(s)}$ lifetime $\tau_{B^0_{(s)}}$, the electromagnetic coupling constant α_{em}, the $B^0_{(s)}$ meson decay constant $f_{B^0_{(s)}}$, the CP asymmetry $\mathcal{A}_{CP}$ and the relative width difference between the heavy and light $B^0_{(s)}$ meson mass eigenstates $y_{s,d}$. As neutral B mesons can transition into their anti-particle, the mass eigenstates do not equate to the flavour eigenstates, but instead are mixed states of the flavour eigenstates $B^0_{(s)}$ and $\bar{B}^0_{(s)}$. In the SM, only the heavy mass eigenstate contributes to the $B^0_{(s)} \to \ell^+\ell^-$ decays, meaning $\mathcal{A}_{CP} = 1$. The parameters S and P expand the Wilson coefficients introduced in Equation (2.2) to describe additional non-SM components from scalar and pseudoscalar currents, as well as their right-handed counterparts. They are defined as

$$S = \sqrt{1 - \frac{4m_\ell^2}{m^2_{B^0_{(s)}}}} \frac{m^2_{B^0_{(s)}}}{2m_\ell} \frac{m_b}{m_b + m_{s,d}} \frac{\mathcal{C}_S - \mathcal{C}'_S}{\mathcal{C}_{10}^{SM}} \quad (2.5)$$

for the scalar case and

$$P = \frac{\mathcal{C}_{10} - \mathcal{C}'_{10}}{\mathcal{C}_{10}^{SM}} + \frac{m^2_{B^0_{(s)}}}{2m_\ell} \frac{m_b}{m_b + m_{s,d}} \frac{\mathcal{C}_P - \mathcal{C}'_P}{\mathcal{C}_{10}^{SM}} \quad (2.6)$$

for the pseudoscalar case, with the quark masses m_b, m_s and m_d and the Wilson coefficients $\mathcal{C}_S$ and $\mathcal{C}_P$ for scalar and pseudoscalar currents. The primed coefficients are the right-handed equivalents to their non-primed counterparts. In the SM $\mathcal{C}_S$, $\mathcal{C}_P$ and all right-handed currents vanish, leading to $P = 1$ and $S = 0$. Notably in these calculations, only $\mathcal{C}_{10}$ enters the SM predictions for $B^0_{(s)} \to \ell^+\ell^-$ decays. The current SM predictions for $B^0_{(s)} \to \mu^+\mu^-$ are taken from [50]. The results for $B^0_{(s)} \to e^+e^-$ are obtained by scaling the $B^0_{(s)} \to \mu^+\mu^-$ results by the expected helicity suppression, so a factor m_e^2/m_μ^2, and neglecting electromagnetic corrections discussed in [50]. They are however included as a systematic uncertainty of $\mathcal{O}(1\,\%)$. This leads to the current SM predictions of

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.66 \pm 0.14) \cdot 10^{-9}, \tag{2.7}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \cdot 10^{-10}, \tag{2.8}$$
$$\mathcal{B}(B^0_s \to e^+e^-) = (8.60 \pm 0.36) \cdot 10^{-14} \text{ and} \tag{2.9}$$
$$\mathcal{B}(B^0 \to e^+e^-) = (2.41 \pm 0.13) \cdot 10^{-15}. \tag{2.10}$$

The predictions for the B^0 decays are smaller than for the B^0_s decays as the transition of a b quark to a d quark depends on the smaller CKM factor V_{td} compared to V_{ts} for the transition from a b quark to a s quark. The dominant contributions to the uncertainty come from the experimental determination of the CKM factors entering the prediction, with uncertainties of 3.1% for $|V_{tb}V^*_{ts}|$ and 4.6% for $|V_{tb}V^*_{td}|$ [51, 52], and the uncertainty on the decay constant $f_{B^0_{(s)}}$, which needs to be obtained from QCD lattice calculations [53].

Alternatively, the $B^0_{(s)} \to \ell^+\ell^-$ branching fractions can be obtained independently from the decay constant and CKM factors by relating them to the mass difference between the heavy and light B meson mass eigenstates, yielding predictions fully consistent with the values reported above [54].

The most precise measurement of $B^0_{(s)} \to e^+e^-$, excluding the measurement performed in the context of this thesis, was performed by the CDF collaboration, where upper limits on the branching fraction of $B^0_{(s)} \to e^+e^-$ were set at 90% confidence level as [55]

$$\mathcal{B}(B^0_s \to e^+e^-) < 2.8 \cdot 10^{-7} \text{ and} \tag{2.11}$$
$$\mathcal{B}(B^0 \to e^+e^-) < 8.3 \cdot 10^{-8}. \tag{2.12}$$

2.2.2 $B^0_{(s)} \to e^+e^-$ in theories beyond the Standard Model

The suppression of $B^0_{(s)} \to e^+e^-$ with respect to $B^0_{(s)} \to \mu^+\mu^-$ makes the SM process inaccessible experimentally with the currently available data. This however does not necessarily hold true in theories beyond the SM, where the branching fraction of $B^0_{(s)} \to e^+e^-$ can be enhanced through various means, particularly an extended (pseudo-)scalar sector compared to the SM, as can be seen in Equations (2.5) and (2.6).

In this case, the helicity suppression can be lifted, allowing for sizeable increases of the $B^0_{(s)} \to e^+e^-$ branching fraction to a similar level of the muonic counterparts.

Typical examples of such models include the introduction of leptoquarks or models with an extended sector of Higgs-like particles, such as a two Higgs doublet model type 2 [56]. This model introduces a minimally enhanced Higgs sector with two Higgs doublets, one neutral and one charged, as well as a neutral pseudoscalar. As this model requires different couplings for the neutral and charged Higgs doublets, the $B^0_{(s)} \to e^+e^-$ branching fraction could be enhanced by the fourth power of the ratio of vacuum expectation values of these doublets, leading to branching fractions of $\mathcal{O}(10^{-10})$.

Also the introduction of light scalar mediators, aimed at explaining the anomalies seen in $b \to s\ell^+\ell^-$ measurements could enhance $B^0_{(s)} \to e^+e^-$ close to their current experimental bounds [57].

Similarly, the introduction of light flavour-universal mediators could lift the helicity suppression and enhance $B^0_{(s)} \to e^+e^-$, while keeping $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to \tau^+\tau^-$ close to the SM [58]. The effect of one such model (denoted below as "Universal New Physics Scenario") on the three $B^0_{(s)} \to \ell^+\ell^-$ decays is illustrated in Figure 2.5.
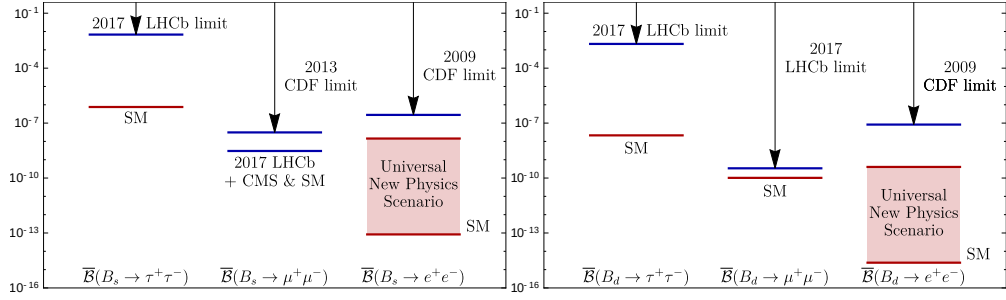
Figure 2.5: Predictions for the three $B^0_{(s)} \to \ell^+\ell^-$ decays for the case of additional light flavour-universal mediators. The predictions for $B^0_{(s)} \to e^+e^-$ are compared to the corresponding SM predictions and current experimental limits. Taken from Ref. [58].

2.3 $B^+ \to K^+ e^\pm \mu^\mp$ decays

While the aforementioned $B^0_{(s)} \to e^+ e^-$ processes are rare, $B^+ \to K^+ e^\pm \mu^\mp$ decays are strictly forbidden in the SM because they violate the conservation of lepton flavour. While these decays can occur when taking neutrino oscillations into account via the process shown in Figure 2.7, the expected branching ratio is unmeasurably small around $\mathcal{O}(10^{-40})$. Thus, $B^+ \to K^+ e^\pm \mu^\mp$ and other lepton-flavour violating decays provide a null-test of the SM, where the observation of a signal would provide an unambiguous sign for New Physics.

These decays have been explored by previous experiments as well as the LHCb experiment. A selection of the experimental status of charged lepton-flavour violation in the B sector at the time of writing is shown in Figure 2.6.
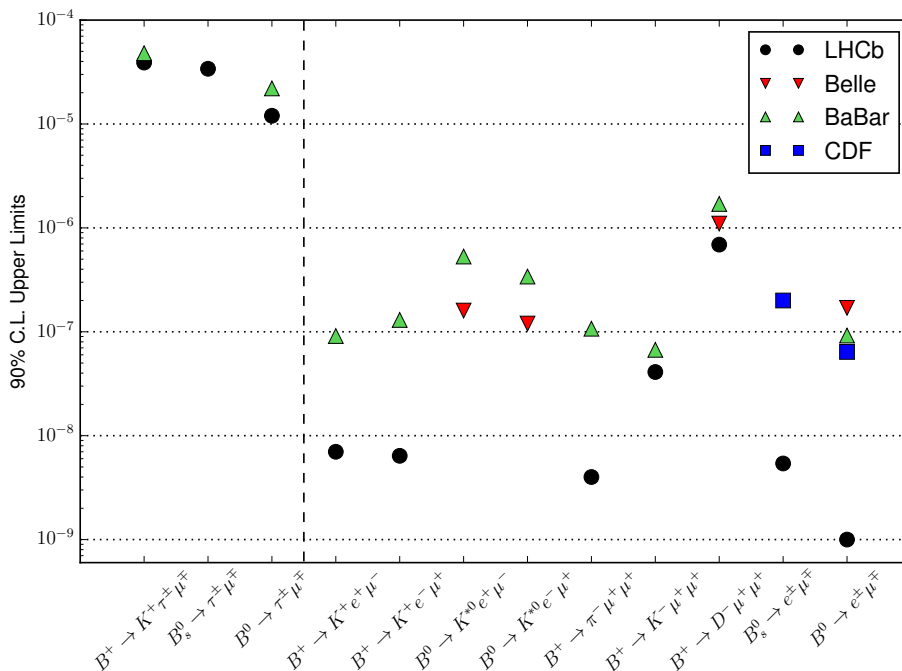
Figure 2.6: Selection of the experimental status of charged lepton-flavour violation in B meson decays. Results published by the LHCb collaboration are compared to previous results from the Belle, BaBar and CDF collaborations. Results from Refs. [59–75].

Searches for charged lepton-flavour violation are motivated by the anomalies seen in $b \to s \ell^+ \ell^-$ transitions, particularly those hinting at the violation of lepton-flavour

universality, the concept of the couplings of gauge bosons to leptons being the same independent of the lepton flavour. This concept is tested in measurements of the ratio of branching fractions of decays into hadrons and a lepton pair, where the leptons are either muons or electrons, given as

$$R(X) = \frac{\mathcal{B}(B \to X\mu^+\mu^-)}{\mathcal{B}(B \to Xe^+e^-)}. \tag{2.13}$$

These lepton-flavour universality ratios can be computed quite cleanly in the SM due to cancellation of hadronic effects, leading to predictions of [76, 77]

$$R(X)_{SM} = 1.00 \pm 0.03, \tag{2.14}$$

where the uncertainty comes from higher order QED corrections to the base process. Measurements of $R(K^+)$ and $R(K^{*0})$ show consistent deviations from unity. Indeed the violation of lepton-flavour universality would also indicate charged lepton-flavour violation [12].

Models aiming to explain the $b \to s\ell^+\ell^-$ anomalies mentioned in Chapter 2 often predict strongly enhanced branching fractions for lepton-flavour violating decays like $B^+ \to K^+ e^\pm \mu^\mp$. One natural addition to the SM would be the inclusion of mixing between charged leptons similarly to the mixing of neutrinos, which could enhance the branching fraction of $B^+ \to K^+ e^\pm \mu^\mp$ up to $\mathcal{O}(10^{-9})$ [78]. Models involving leptoquarks [79] also propose an enhancement of $B^+ \to K^+ e^\pm \mu^\mp$ to the level of $\mathcal{O}(10^{-10} - 10^{-8})$, while the addition of new heavy neutral bosons could enhance $\mathcal{B}(B^+ \to K^+ e^\pm \mu^\mp)$ up to $\mathcal{O}(10^{-8})$ [80]. This enhancement is usually linked to the deviation seen in the $b \to s\ell^+\ell^-$ anomalies via the Wilson coefficients [81]

$$\mathcal{B}(B^+ \to K^+ e^\pm \mu^\mp) = (31 \pm 4) \cdot 10^{-9} \left[\left(\frac{\mathcal{C}_9^{ee}}{\gamma} \right)^2 + (\gamma \mathcal{C}_9^{\mu\mu}) \right], \tag{2.15}$$

with a parameter γ governing the relative strength of the leptoquark coupling to electrons and muons and the Wilson coefficients $\mathcal{C}_9^{\ell\ell}$, determined independently from measurements of electron and muon processes. Another possible relation of the $B^+ \to K^+ e^\pm \mu^\mp$ branching fraction can be directly parametrised by $R(K^+)$ as [79]

$$\mathcal{B}(B^+ \to K^+ e^\pm \mu^\mp) = 3 \cdot 10^{-8} \kappa^2 \left(\frac{1 - R(K^+)}{0.23} \right)^2, \tag{2.16}$$

with a strength parameter κ of $\mathcal{O}(1)$. Possible diagrams to illustrate these New Physics models are shown in Figure 2.7.
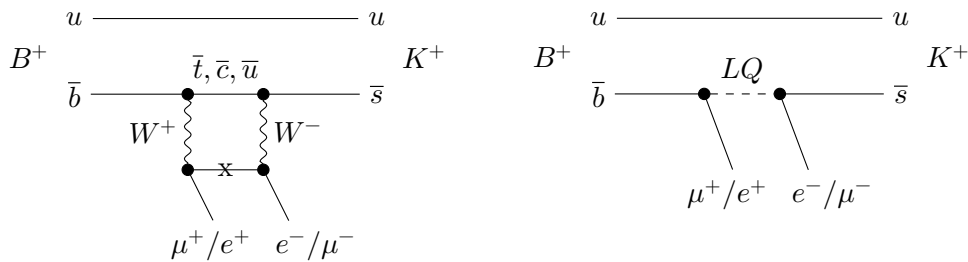
Figure 2.7: Possible Feynman diagrams for the decay $B^+ \to K^+ e^\pm \mu^\mp$. The SM process via a $W^\pm$ box diagram with neutrino oscillation is shown on the left, while a possible tree-level process occurring via a leptoquark is shown on the right.

A search for $B^+ \to K^+ e^\pm \mu^\mp$ has previously been performed by LHCb on data recorded in 2011 and 2012, corresponding to an integrated luminosity $\approx 1 \text{ fb}^{-1}$ recorded at a centre-of-mass energy of 7 TeV and $\approx 2 \text{ fb}^{-1}$ recorded at a centre-of-mass energy of 8 TeV, finding upper limits on the branching fractions of [66]

$$\mathcal{B}(B^+ \to K^+ e^- \mu^+) < 8.8 \cdot 10^{-9} \text{ @95\% CL} \tag{2.17}$$

$$\mathcal{B}(B^+ \to K^+ e^+ \mu^-) < 9.5 \cdot 10^{-9} \text{ @95\% CL}, \tag{2.18}$$

which already imposes strong constraints on possible New Physics contributions. As the search was statistically limited, the additional data available to the LHCb experiment, corresponding to a data sample four times larger due to an increase in $b\bar{b}$ production cross-sections from higher centre-of-mass energy, may allow to either further strengthen the constraints on New Physics or to observe $B^+ \to K^+ e^\pm \mu^\mp$ decays, which would be an unambiguous sign for physics beyond the SM.

3 The LHCb detector

The Large Hadron Collider beauty (LHCb) experiment, where the data analysed in this thesis was recorded, is one of the four large experiments at the Large Hadron Collider (LHC) [82] located at the CERN research centre near Geneva. The LHC is a circular collider with a length of 26.7 km located about 100 m below the surface. Two proton beams are collided at four interaction points, where the experiments are located, with maximum centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ in 2011, $\sqrt{s} = 8\,\text{TeV}$ in 2012 and $\sqrt{s} = 13\,\text{TeV}$ from 2015 onward until the end of Run 2 in 2018.

Besides the LHCb experiment, the ATLAS, CMS and ALICE experiments are located at the LHC. ATLAS and CMS are designed to investigate a wide range of physics, like the search for and studies of the properties of the Higgs boson and direct studies of top quarks, as well as generic searches for New Physics at high masses. The Alice experiment is designed to study the strong interaction in the collisions of heavy ions, like lead, either with protons or other heavy ions. The fourth experiment, on the data of which the analyses presented in this thesis were carried out, is the LHCb experiment. It is designed for precision measurements of the properties of b and c hadrons.

The LHCb detector is a single-arm forward spectrometer with an angular acceptance of 10 mrad to 300 mrad in the horizontal plane and up to 250 mrad in the vertical plane [83]. A cross-section of the detector is shown in Figure 3.1.

The forward-centric architecture of the detector is motivated by the angular distribution of the $b\bar{b}$ pairs originating from the proton-proton interaction. The $b\bar{b}$ pairs are mainly produced with very high forward (or backward) boost, while having low transverse momentum in comparison. The distribution of the b quarks production angles is shown in Figure 3.2. Since space in the cavern, where the detector is placed, is limited, only the forward direction is covered by the LHCb detector, so only about 27 % of the produced b quarks are produced within the LHCb acceptance [85].

The following chapters describe the subsystems of the LHCb detector in greater detail, starting with the Vertex Locator (see Section 3.1), which encompasses the interaction point. The tracking system (see Section 3.2) provides the information necessary to reconstruct the tracks of charged particles in the detector, based on which decay chains are reconstructed. In order to distinguish the particles producing
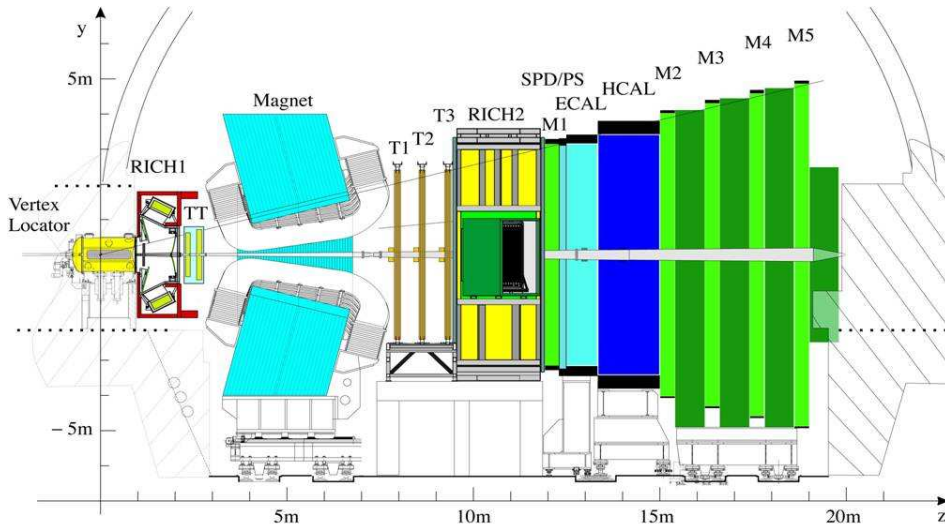
Figure 3.1: Cross-section of the LHCb detector [84]. The coordinate axis parallel to the beam pipe is denoted as z, while in the plane transverse to the beam pipe the horizontal direction is chosen as x, the vertical direction as y.
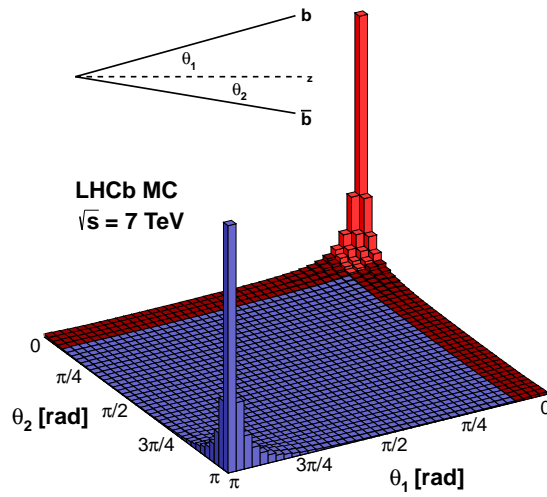
Figure 3.2: Angular distribution of the production angles of a $b\bar{b}$ pair. Most b quarks are produced with angles close to 0 or π radians, which corresponds to the region closest the beam pipe.[86]

17

these tracks, a particle identification system (see Section 3.3) consisting of Ring Imaging Cherenkov detectors, two calorimeters and five muon chambers is used. The latter two systems also provide information necessary for the trigger system (see Section 3.4). The reconstruction of bremsstrahlung, which is essential for searches including electrons in the final state, is described in Section 3.5. Afterwards, the structure of the data flow in LHCb is presented in Section 3.6.

3.1 Vertex Locator

The interaction point of the proton beams is surrounded by the Vertex Locator (VELO). It measures the positions of decay vertices of particles like B and D mesons, which have a large enough lifetime to decay significantly detached from the interaction point. Particles with significantly longer lifetimes like K_S^0 however might not decay inside the VELO. The position of the pp interaction point is also measured using information from the VELO.

The VELO consists of two modules of 21 half-circle silicon elements each, encasing the interaction point. Before a stable beam is achieved in the LHC, the VELO is retracted from the beams to avoid damage to the detector. Once the beams are stable, the VELO is placed in a distance of about 5 mm (8 mm in Run 2) away from the beam.

Due to their high boost and significant lifetime, B mesons fly about 1 cm on average in the detector before decaying. In order to measure the decay time of the B mesons and correctly match their decay products to them, their decay vertices have to be reconstructed with high precision. The resolution of the vertex reconstruction depends strongly on the number of tracks in the events. For example, in an event with 25 tracks the resolution is about 13 μm in transverse direction to the beam pipe and about 71 μm in longitudinal direction [87]. A schematic view of the VELO can be seen in Figure 3.3.
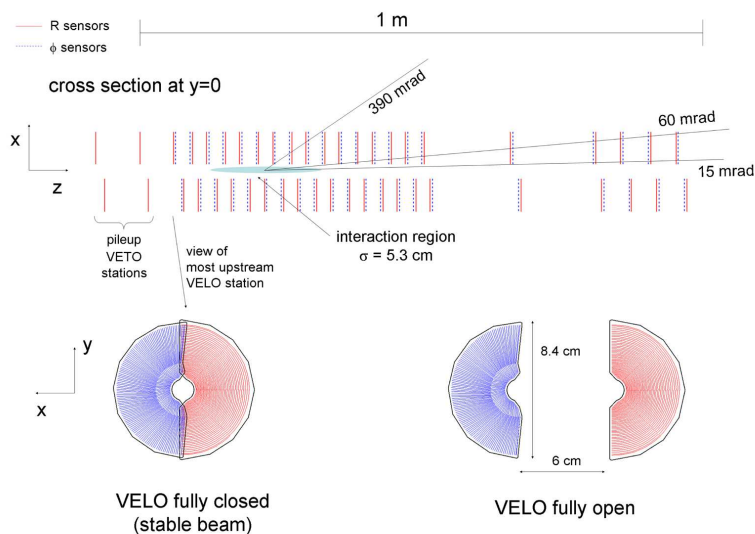
Figure 3.3: Schematic view of the Vertex Locator. On the top a cross section in the x-z plane can be seen, showing the different silicon elements. On the bottom the transverse plane is depicted with the VELO in its closed state (left) and open state (right) [88].

3.2 Tracking system

Tracks in the detector are reconstructed from hits in the tracking system's substations positioned throughout the detector. The tracking system consists of the tracker turicensis (TT) placed upstream of the magnet and the three tracking stations T1-T3 placed downstream of the magnet, but before the second Ring Imaging Cherenkov Detector (RICH2).

The magnet has a bending power of $4\,\mathrm{Tm}$. It bends charged tracks, where the track curvature allows to reconstruct the particle's momentum while the direction in which a particle is diverted allows to determine the particle's charge. The three tracking stations are further divided into the inner tracker (IT), covering a cross-shaped area around the beam pipe, and the outer tracker (OT) covering the rest of the detector acceptance. The TT and the IT are silicon microstrip detectors with four detection layers. The two inner layers are rotated by $\pm 5\,^\circ$, allowing for a better spatial track reconstruction with a resolution of about $50\,\mu\mathrm{m}$. The outer tracker is composed of drift straw tubes filled with a mixture of argon and carbon dioxide. They are ordered in two layers of tubes with a combined resolution of about $200\,\mu\mathrm{m}$, resulting in a momentum uncertainty of $0.5\,\%/p_T$ for momenta below $20\,\mathrm{GeV}/c$ up to $0.8\,\%/p_T$

19

for particles with momenta around $100\,\text{GeV}/c$ [87].

Reconstructed tracks are classified based on the subsystems of the tracking system in which they have hits associated with them. Long tracks are required to have hits in the VELO and the tracking stations associated with them, while downstream tracks do not require hits in the VELO, but instead in the TT and tracking stations. VELO tracks do only have hits in the VELO associated with them, just as T tracks only have tracks in the stations T1-T3 associated with them.

3.3 Particle identification system

The distinction between different kinds of particles is crucial for LHCb analyses. For this purpose, information about particles in the detector is collected in the RICH detectors, calorimeters and muon chambers. This information is then combined to separate different particle species from one another.

3.3.1 RICH

The two Ring Imaging Cherenkov Detectors (RICH1 and RICH2) are located upstream (RICH1) and downstream (RICH2) of the magnet. They utilise Cherenkov radiation for charged particle identification, covering different momentum ranges by using materials of different refractive indices. In Run 1, RICH1 was filled with aerogel and C_4F_{10} as radiators. Since the beginning of Run 2 in 2015, the aerogel has been removed from RICH1, since it caused a significant decrease in performance of the algorithms tagging a particle's production flavour, while only marginally contributing to the particle identification of low momentum tracks. RICH1 covers a lower momentum range of $1\,\text{GeV}/c < p < 60\,\text{GeV}/c$, while RICH2 covers a momentum range of $15\,\text{GeV}/c < p < 100\,\text{GeV}/c$. It is filled with CF_4 as a radiator.

The particle velocity in these detector elements is measured via Cherenkov photons emitted by charged particles travelling faster than the speed of light in a medium with refractive index n. These photons are emitted in a cone around the particle's direction of flight, where the cone's opening angle is defined as

$$\cos\theta_c = \frac{1}{n\beta},$$

(3.1)

where β is the particle velocity in units of the vacuum speed of light. In combination with the momentum measurements done by the tracking system, the opening angle

of the cone allows to test different particle mass hypotheses, separating different particle species as shown in Figure 3.4.
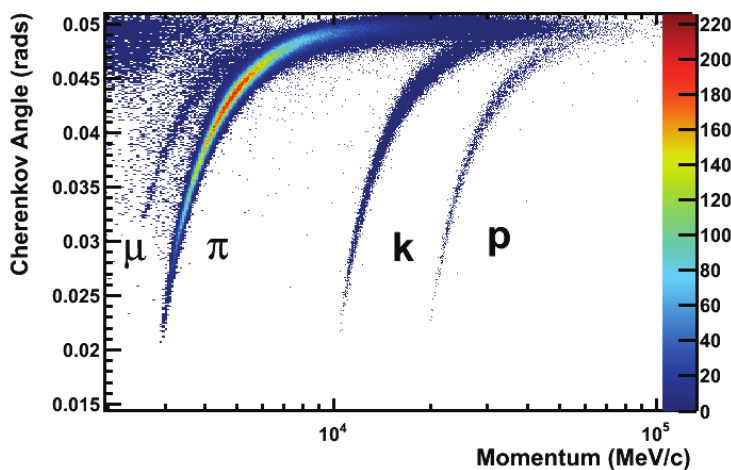
Figure 3.4: Cherenkov angle as a function of particle momentum in RICH1. Different particle species are clearly separated at low momentum, while for high momenta, the distinct bands start to overlap, making the second RICH detector especially necessary for high momentum tracks. Figure reproduced from Ref. [89].

3.3.2 Calorimeters

The calorimeter system of the LHCb detector measures a particle's energy as well as the position of where this energy was deposited. In addition, information about showers in the different calorimeters allows to distinguish between leptons, hadrons and photons. The calorimeter system consists of an electromagnetic (ECAL) and hadronic calorimeter (HCAL), complimented by the Scintillating Pad Detector (SPD) and the Preshower detector (PS).

The SPD is placed before the ECAL to determine the charge of the particles hitting the calorimeter system, as only charged particles cause scintillation. The PS detects showers originating from produced secondary particles, allowing for differentiation of charged pions and electrons. Information from these two detector systems is processed in the trigger (see Section 3.4) in combination with information from the ECAL to indicate electrons, photons or neutral pions being present. The ECAL is placed behind the SPD and PS to measure electron and photon energy. It is composed of 66 layers of lead and scintillators, where electrons and photons produce secondary particles in the lead layers, which deposit energy in the scintillating layers.

Since the event occupancy is about two orders of magnitude higher in regions near the beampipe than it is in the outer regions of the detector, the ECAL is laterally segmented into three regions of different cell size to increase the detection accuracy in the more densely populated area close to the beam pipe.

Hadrons also have a chance to produce showers in the ECAL thus depositing a part of their energy there, but in addition produce showers in the HCAL. Therefore for hadrons, the energy measurement is performed in both calorimeters. The HCAL consists of iron plates and scintillating tiles placed parallel to the beam pipe. The general idea for the detection of hadrons is the same as it is in the ECAL. The iron layers are needed to have interactions where the hadrons produce secondary particles, the energy deposition of these particles is then determined by the scintillating pads. The HCAL is segmented similarly to the ECAL into two regions.

3.3.3 Muon system

Muons are identified by hits in the muon stations M1 - M5, where M1 is placed in front of the calorimeter system and the M2-M5 are placed behind it. These stations each consist of 276 multi-wire proportional chambers. In the M1 station, the inner part consists of triple-GEM (Gas Electron Multiplier) detectors. The stations behind the calorimeter systems are interspaced with 80 cm thick iron absorbers, blocking particles other than muons from reaching the muon system, allowing reliable separation of muons from other particles. At LHCb the detection rate for muons is above 99 % for each muon station, leading to excellent muon identification and reconstruction.

3.4 Trigger system

The LHC bunch crossing rate of about 40 MHz has to be reduced in order to have manageable datasets, as saving all events would require an enormous amount of storage. As many events are not interesting from a physics perspective, the event rate, and thus required storage, can be reduced with only minimal loss of physics, where a focus is placed on B and D mesons. The event rate is reduced to about 5 kHz in Run 1 and 12.5 kHz in Run 2 by deploying one hardware based trigger stage followed by two software based trigger stages.

3.4.1 Level 0 trigger

The first trigger level, referred to as Level 0 (L0) trigger, is a hardware based trigger level, consisting of calorimeter triggers (Hadron, Photon and Electron) and muon triggers (Muon and DiMuon). The muon trigger yields a positive trigger response if an event contains a muon with transverse momentum above a certain threshold. In the dimuon case, a decision is made based on the product of the two highest momentum muon's transverse momenta instead. In the calorimeter triggers, the decision is based on the particle's transverse energy and information about the detecting calorimeter subsystem. In addition, requirements on the number of tracks in the event are applied, where events with very high multiplicity are rejected. The L0 selection reduces the data rate from about 40 MHz to 1 MHz.

3.4.2 High level trigger

The high level trigger (HLT) is the software based trigger stage of the LHCb detector. It is divided into two stages (HLT1 and HLT2). A first reconstruction of the tracks is performed on HLT1, making higher level information available compared to the L0 trigger, like information about the displacement from the interaction point, which is used to trigger displaced signal from B or D mesons. In HLT1 the data rate is further reduced to about 80 kHz.

At the HLT2 level, events are reconstructed completely and selections can be performed on the full event topology. Especially important for the analyses presented in this work are triggers based on bonsai boosted decision trees [90], which allow to trigger on decays of B mesons into two or three body final states.

3.5 Bremsstrahlung reconstruction

Due to their low mass, the behaviour of electrons in the detector differs significantly from hadrons and muons, mainly because of the emission of bremsstrahlung from interactions with the detector material and the bending of the magnet.

This loss of energy to bremsstrahlung worsens the momentum resolution and reconstruction efficiency of electron tracks. Depending on where the bremsstrahlung was emitted, the effect on measurements performed on the track can differ. If the bremsstrahlung is emitted before the magnet, the track is bent away from the resulting bremsstrahlung photons and the photons energy is lost, biasing the reconstructed kinematic information of the electron. In case the bremsstrahlung is emitted after the magnet, the resulting photons reach the calorimeter close to

the emitting electron and end up in the same cluster, so their energy is recovered naturally.

In order to correct for the bias on kinematics from bremsstrahlung emission, algorithms to recover the lost energy are employed during the event reconstruction. As bremsstrahlung photons reach the ECAL, they produce showers similar to photons from other sources. If the bremsstrahlung photon has sufficient energy (above 75 MeV [91]), the cluster produced by it in the ECAL can be reconstructed, where the goal is to maintain high efficiency in the bremsstrahlung recovery while avoiding fake clusters from overlap of multiple neighbouring clusters.

The recovered cluster can then be associated with an electron track by extrapolating the track to the ECAL only from VELO information and from the combined information from VELO and TT combined. An ECAL region corresponding to 2σ, where σ is the combined uncertainty from the extrapolations, around the extrapolated tracks is searched for potential clusters from bremsstrahlung photons. By adding the cluster energy back to the track and assuming the same primary vertex as the electron track, the electron track's four-momentum can be corrected for the emitted bremsstrahlung. A schematic of the process is shown in Figure 3.5.
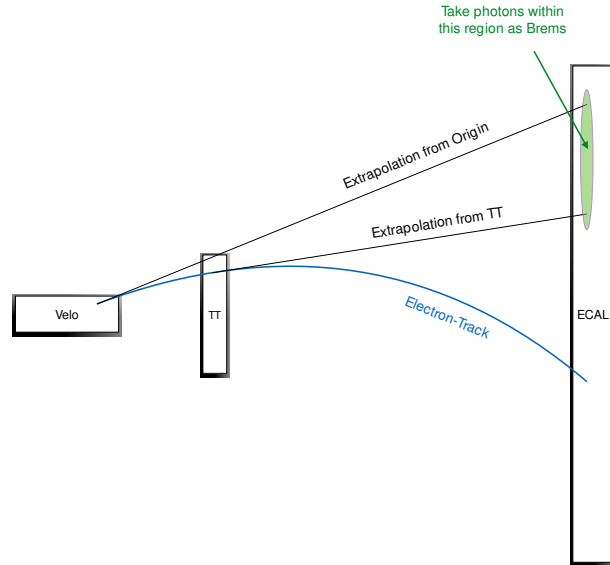
Figure 3.5: Schematic of the bremsstrahlung correction process at LHCb, reproduced from Ref. [91].

Using the procedure described above, about 50 % of electron tracks have bremsstrahlung corrections applied to them, improving their momentum resolution. However, this procedure is not perfect and also introduces overcorrections by

associating the wrong photon to the electron track, which manifest in the invariant mass distribution of combinations of electrons as tails towards higher masses. As a result, the search for $B^0_{(s)} \to e^+e^-$ is carried out separately for three categories of bremsstrahlung, defined by the cases where none, either or both electrons have had bremsstrahlung corrections applied to them.

The application of bremsstrahlung corrections also serves as a powerful tool to distinguish electrons from hadrons, as hadrons do not emit meaningful amounts of bremsstrahlung and thus have no corrections applied to them. The corrections therefore are a vital input to electron particle identification.

3.6 Data processing at LHCb

Since the raw output of the detector consists mainly of hits in different sections of the sub-detectors, it is necessary to translate them into physical values and human understandable information. At LHCb this process is performed in the GAUDI framework [92], which governs all other applications described below.

3.6.1 Track reconstruction

The first step of the data processing is the track reconstruction, taking information from the VELO, the tracking stations TT and T1-T3, as well as the muon chambers into account. This step is performed inside the BRUNEL framework [93], where the found tracks are fitted with a Kalman filter [94], to account for scattering effects and energy loss of the particles.

3.6.2 Event reconstruction

In this step, data samples describing specific decay modes are produced with the DAVINCI framework [95]. In this step also a loose preselection, called Stripping, is applied. It mainly consists of cuts on daughter particles' invariant masses as well as track and vertex qualities and kinematic information. The specific preselections applied to the data in the analyses presented in this thesis are summarised in Sections 4.2.1 and 5.2.1. Possible decay chains are reconstructed by a decay tree fitter algorithm [96], combining the selected tracks to form vertices and calculate information about invariant masses and particle lifetimes.

3.6.3 Simulated events

Since the data samples are dominated by background it is important to describe pure signal decays by generating simulated events. These Monte Carlo (MC) simulations [97] are processed as similarly as possible to real data. The first step in this process is the generation of the events and the simulation of their interactions with the different detector systems. This process is done in the GAUSS framework [98], where the proton-proton collisions are generated by the PYTHIA software package [99]. The B decays in these events are simulated by the EVTGEN [100] package. The propagation of the simulated particles through the detector and their interactions are then simulated in GEANT4 [101]. The last step is the digitisation of energy depositions in the detector systems, done in BOOLE [102]. From this point on, the simulation is required to go through the same steps as real data, including the BRUNEL and DAVINCI steps described above.

4 Search for the rare decays $B^0_{(s)} \to e^+e^-$

In this chapter, the search for the rare decays $B^0_{(s)} \to e^+e^-$ is presented. It was published in [103]. LHCb publications list all authors of the collaboration in alphabetical order, with 2-3 authors highlighted as contact authors. I am one of the contact authors of this publication. The analysis follows the typical structure of a search for a rare decay at LHCb, with the main complication being the emission of bremsstrahlung by electrons. The analysis makes use of inputs from two master's theses on the subject [104, 105] and was performed jointly with another PhD student [106]. The main focus of my work was the definition of the analysis strategy, trigger efficiency calibration and studies of systematic uncertainties as well as cross-checks of absolute efficiencies.

4.1 Analysis strategy

The goal of this analysis was to perform the first LHCb measurement in the channels $B^0_{(s)} \to e^+e^-$. To accomplish this goal an analysis chain was constructed aiming to maintain the highest signal efficiency possible while rejecting a majority of the background present in the data sample. The selection criteria are constructed using simulated $B^0_s \to e^+e^-$ candidates, based on which also the efficiency of the selection is estimated. In searches for rare B decays the invariant mass of the B meson candidate, reconstructed from the kinematics of the final state particles, is often used as main discriminating variable between signal and background, since the signal should show up as a peaking structure at the nominal B-meson mass. Thus, to avoid experimenter's bias, a region in the invariant mass of the electron pair is blinded prior to optimising the selection process. For the purpose of this analysis, two mass regions are defined: the signal region being the smallest region containing 68 % of simulated $B^0_s \to e^+e^-$ candidates and the exclusion region containing 90 % of simulated $B^0_s \to e^+e^-$ candidates. These regions are summarised in Table 4.1 and visualised in Figure 4.1.

Table 4.1: Summary of the signal and exclusion regions for the data samples.

	mass region in MeV/c^2
signal region	[5044, 5428]
exclusion region	[4689, 5588]
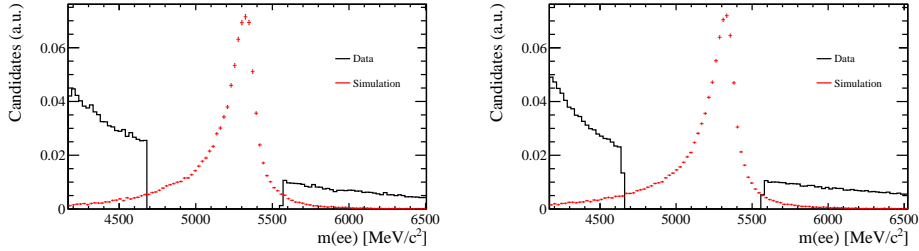
Figure 4.1: Comparison of the dielectron invariant mass distribution in data (black) and $B^0_s \to e^+ e^-$ simulation (red). Run 1 is shown on the left, Run 2 on the right. In data, the exclusion region from Table 4.1 around the B^0_s mass is removed.

The mass regions outside the exclusion region are still of interest for the analysis, where they serve as control regions for different sources of background. The region with masses higher than the exclusion region (upper mass sideband) is free of contributions from physics backgrounds and thus contains only random combinations of electrons, called combinatorial background, while the region with lower masses (lower mass sideband) also contains contributions from partially reconstructed decays. This type of background originates from decays where at least one final state particle is not reconstructed. Thus energy is lost and the invariant mass shifted towards lower masses. Since both kinds of backgrounds can be present in the exclusion region, modelling them based on these control regions is crucial to disentangle the background from a possible signal contribution.

In order to cancel systematic effects from the selection, the branching fraction is measured relative to that of a normalisation mode with high branching fraction and similar topology to the signal decay. In this case, the decay $B^+ \to K^+ J/\psi(\to e^+ e^-)$ is used, since it has a high, well known branching fraction taken from the Particle Data Group (PDG) [107] of

$$\mathcal{B}(B^+ \to K^+ J/\psi) \times \mathcal{B}(J/\psi \to e^+ e^-) = (1.010 \pm 0.028) \cdot 10^{-3} \times (5.971 \pm 0.032)\,\% \quad (4.1)$$

and two electrons in the final state with a similar kinematic range as that expected for $B^0_{(s)} \to e^+ e^-$, allowing for precise control of selection efficiencies. With this, the branching fraction is given as

$$\mathcal{B}(B^0_{(s)} \to e^+e^-) = \frac{N(B^0_{(s)} \to e^+e^-)}{\epsilon(B^0_{(s)} \to e^+e^-)} \frac{\mathcal{B}(B^+ \to K^+ J\!/\!\psi) \cdot \epsilon(B^+ \to K^+ J\!/\!\psi)}{N(B^+ \to K^+ J\!/\!\psi)} \frac{f_u}{f_{s(d)}}$$

$$= N(B^0_{(s)} \to e^+e^-) \cdot \alpha' \cdot \frac{f_u}{f_{s(d)}} \cdot \mathcal{B}(B^+ \to K^+ J\!/\!\psi) \qquad (4.2)$$

$$= N(B^0_{(s)} \to e^+e^-) \cdot \alpha$$

with the signal and normalisation yields after the selection $N(B^0_{(s)} \to e^+e^-)$ and $N(B^+ \to K^+ J\!/\!\psi)$ and their associated selection efficiencies $\epsilon(B^0_{(s)} \to e^+e^-)$ and $\epsilon(B^+ \to K^+ J\!/\!\psi)$. The fragmentation fractions $f_{(u,d,s)}$ and the branching fraction of the control mode $\mathcal{B}(B^+ \to K^+ J\!/\!\psi)$ are external inputs to the analysis. Because of this, in addition to the normalisation factor α, describing the single event sensitivity of the analysis, a partial normalisation factor α' without the external inputs is defined as well, allowing for adjustments should the external inputs change.

Since the data taking conditions differ between Run 1 and Run 2 due to changes in the trigger system and centre-of-mass energy, the normalisation is performed split by Run period. In addition, the analysis is performed in three categories of bremsstrahlung corrections being applied to the final state electrons, namely the cases where none, either or both electrons have bremsstrahlung corrections applied to them, since kinematics and dielectron invariant mass resolution are strongly affected by these corrections.

The selection can be split into a number of steps: first, a loose experiment-wide preselection is applied to the data when reconstructing the candidates as well as an additional cut-based preselection to remove obvious backgrounds and unphysical candidates. This selection is described in Section 4.2.1. Before the training of the multivariate classifier, it needs to be checked that simulation actually describes the signal in data correctly, which unfortunately often times is not the case. While the control mode signal is clearly visible in data after the preselection, also background remains in the sample. In order to allow for comparisons between data and simulation, this background is subtracted using the sPlot method [108]. The procedure is described in Section 4.2.2. Residual differences that might be observed can then be corrected for by reweighting the simulation to match the data. This approach is not feasible for variables used for particle identification, which are resampled from large calibration samples instead. The techniques employed are detailed in Section 4.2.3. After these corrections have been applied, a multivariate selection is performed in the form of a boosted decision tree (BDT) [109] to remove combinatorial background from the sample. As a last step, stringent requirements on particle identification (PID) are imposed to reduce pollution from backgrounds where at

least one final state particle is misidentified. These selection steps are explained in more detail in Section 4.2.4 and Section 4.2.5. After the whole selection is applied the efficiencies for signal and normalisation modes can be calculated. They are combined with the yield of the normalisation mode and external inputs to the normalisation constants. A summary and discussion of the normalisation can be found in Section 4.3, systematic uncertainties are discussed in Section 4.3.6. Remaining physical background contributions are studied in Section 4.4. Lastly, the results of the analysis are shown in Section 4.5.

4.2 Selection

The following section describes the selection applied to $B^0_{(s)} \to e^+ e^-$ data and simulation to separate signal from background candidates. The applied preselection (see Section 4.2.1) is split in three parts: an experiment-wide preselection called stripping, requirements on the LHCb trigger system and fiducial cuts to align data and simulation with calibration data. Afterwards, remaining background in the control mode can be subtracted (see Section 4.2.2) in order to allow comparisons of signal candidates in data and simulation and correct for observed differences (see Section 4.2.3). With these corrections applied, a multivariate analysis is performed (see Section 4.2.4) and requirements on particle identification are applied to remove as much background as possible (see Section 4.2.5).

4.2.1 Preselection

A loose preselection is applied to the data in order to remove unphysical candidates and ensure compatibility with used calibration samples while maintaining high signal efficiency. The candidates are formed at an experiment-wide preselection stage, where basic requirements are applied to the formed candidates. Afterwards, a subset of the LHCb trigger decisions is chosen to ensure similarity between data and simulation and avoid choosing triggers that do not increase the efficiency in data when added to other triggers, as this might introduce additional background without adding to the sensitivity. The selection is described below, the data-driven calibration is described in Section 4.3.2. Fiducial requirements to remove obvious backgrounds and ensure kinematic coverage of the calibration data are applied after the trigger selection.

Experiment-wide preselection

The first selection step is performed in an experiment-wide preselection (called stripping) while combining tracks to form signal candidates. This preselection is organized in stripping lines, summarising the selections to be applied to data and simulation for different decay modes. The selections for $B^0_{(s)} \to e^+e^-$ and $B^+ \to K^+ J/\psi(\to e^+e^-)$ are summarised in Table A.1.

Similar requirements are applied to the particles where the electrons originate from, namely the $B^0_{(s)}$ and J/ψ candidates to minimize differences between signal and control mode. Additional cuts are applied to the B^+ candidate in $B^+ \to K^+ J/\psi(\to e^+e^-)$ to ensure the K^+ and J/ψ originate from the same particle. DIRA is defined as the angle between particle's momentum vector and the vector connecting its primary and secondary vertices. χ^2_{IP} is the significance of the particle's impact parameter, defined as the difference in χ^2 when fitting the primary vertex of the decay with and without the given particle. Here, the $B^0_{(s)}$ candidate is required to have low χ^2_{IP}, meaning it originates directly from the proton-proton collision, while the electrons are required to have high χ^2_{IP}, indicating they originate from a displaced vertex, thus excluding backgrounds from short-lived resonances produced directly in the proton-proton collision. The $\chi^2_{\text{trk(vtx)}}/\text{ndf}$ is a measure for the quality of the track reconstruction (vertex fit). The requirements on transverse momenta ensure that the particles originate from the decay of a massive particle. To discriminate the electrons against pions, which are produced in abundance in the proton-proton collisions, a requirement on particle identification is applied in this stage on $\text{DLL}_{e,\pi}$, which is the difference in the logarithm of the likelihood for the electron and pion hypotheses. In order to reduce both combinatorial and partially reconstructed backgrounds, only select regions around the $B^0_{(s)}$, B^+ and J/ψ masses in the e^+e^-, $K^+e^+e^-$ and e^+e^- invariant mass distributions are studied. The mass window around the J/ψ studied here is not sufficient to reject the decay $B^+ \to K^+\psi(2S)(\to e^+e^-)$, requiring to narrow down the mass window further at a later stage.

Choice of triggers

After the experiment-wide preselection, trigger requirements are applied to the data where the goal is to keep as high a signal efficiency as possible while keeping the overlap between triggers to a minimum to avoid introducing unnecessary background to the data. To choose the triggers for the selection, available trigger decisions are tested on $B^0_s \to e^+e^-$ simulation. The strategy to choose the trigger selection for the analysis is the following: For every trigger level (L0, HLT1 and HLT2) the single most efficient trigger is chosen as a baseline. Additional triggers are then

added to study the increase in efficiency. This is done to reject triggers that have a large overlap with one another for the signal decay but might introduce additional background. In addition, triggers that tend to not be well described in simulation are excluded from these studies, as well as triggers that have had a prescale applied to them. Triggers that have a prescale applied to them do not save every event that passes the trigger decision, but only a fraction according to their prescale in order to reduce the stored rate. As this prescale is not modelled in simulation, the gain from adding these lines is overestimated in simulation, which complicates the calibration of these lines further. The use of TIS (triggered independent of signal) and TOS (triggered on signal) triggers as well as Dec (there was a trigger decision of any kind) triggers have been considered for this analysis.

As a baseline, only TOS triggers are studied, because their behaviour is the easiest to understand. Because of a significant increase in the trigger efficiency, L0Global TIS has been added to the selection. Global TOS is excluded from the selection, because it represents the combination of all TOS triggers on the given trigger level, which would make the choice of other TOS triggers obsolete and introduce unnecessary background, since these additional TOS triggers were found to not increase the signal efficiency by a meaningful amount. The gain from using Dec triggers instead of TOS ones has been found to be rather small ($\mathcal{O}(1\%)$) as well. The use of these triggers is therefore omitted in the selection, because of the difficulty correctly estimating their efficiency, which would lessen the already small gain further. After testing all available triggers the trigger selection shown in Table 4.2 is chosen, where an event needs to pass at least one of the triggers on a given trigger level, but needs to pass all trigger levels.

Table 4.2: Trigger decisions chosen for the selection split by trigger level for Run 1, 2015 and 2016. Unless stated otherwise, all used lines are TOS.

	Run 1	2015	2016
L0	L0Electron L0Global TIS	L0Electron L0Global TIS	L0Electron L0Global TIS
HLT1	HLT1TrackAllL0 HLT1TrackPhoton	HLT1TrackMVA HLT1TwoTrackMVA	HLT1TrackMVA HLT1TwoTrackMVA
HLT2	HLT2Topo2BodyBBDT HLT2TopoE2BodyBBDT	HLT2Topo2Body	HLT2TopoE2Body HLT2TopoEE2Body

Because 2015 data has been added to the analysis later than 2016 data, the triggers for 2015 have been chosen to be as similar as possible to the 2016 trigger selection.

The most notable difference is the absence of TopoE(E)2Body lines on HLT2, because these lines were not available during the 2015 data taking period.

Fiducial requirements

Since calibration data are used for the calibration of particle identification variables, the selection is aligned to ensure agreement with these samples in the kinematics of the decay products and detector occupancy. As the selection of the calibration data changes depending on the run period, this is reflected in the selection here as well. The requirements are listed in Table A.2.

Further selection is necessary for the $B^+ \to K^+ J/\psi (\to e^+ e^-)$ candidates to ensure a stable fit for comparisons between data and simulation via the sPlot method [108]. The selection consists of a requirement on a neural-network based particle identification variable called ProbNN for the kaon to suppress backgrounds from misidentification, as well as a requirement on the B^+ mass with the J/ψ constrained to its nominal mass to strongly suppress partially reconstructed decays like $B^0 \to K^{*0}(\to K^+ \pi^-) J/\psi$, where the pion would not be reconstructed. An additional requirement on the dielectron invariant mass rejects processes proceeding via higher $c\bar{c}$ resonances as well as combinatorial background. These selection steps are summarised in Table A.3.

In order to remove candidates with unphysically high lifetimes as well as tracks without an associated particle (called ghost tracks), which could interfere with the training of the multivariate classifier, a last set of cuts consisting of loose requirements on the electron and B meson momenta as well the probability of a track to be a ghost is applied to both $B^0_{(s)} \to e^+ e^-$ and $B^+ \to K^+ J/\psi (\to e^+ e^-)$. It is summarised in Table A.4.

4.2.2 Background subtraction with the sPlot method

After the aforementioned preselection steps have been applied, the control mode $B^+ \to K^+ J/\psi (\to e^+ e^-)$ still contains a mixture of signal and background. While the remaining background is small, it needs to be subtracted when comparing data and simulation in order be able to correctly identify discrepancies and correct for them. In order to subtract the remaining background, the sPlot method is used to assign weights, called sWeights, to the data, which are calculated from a fit to the B^+ meson invariant mass. The data are described by

- an exponential function for the combinatorial background

- a double-sided Crystal Ball function [110] for the $B^+ \to K^+ J/\psi(\to e^+ e^-)$ peak per bremsstrahlung category to account for differences due to bremsstrahlung corrections of the electrons.

The double-sided Crystal Ball functions are defined as

$$
CB(m; \alpha, \alpha', n, n', \mu, \sigma) =
$$
$$
N \cdot
\begin{cases}
\exp(-\frac{(m-\mu)^2}{2\sigma^2}) & \text{, for } \frac{m-\mu}{\sigma} > -\alpha \\
& \text{and } \frac{m-\mu}{\sigma} < -\alpha' \\
\left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{\alpha^2}{2}\right) \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \frac{m-\mu}{\sigma}\right) & \text{, for } \frac{m-\mu}{\sigma} \leq -\alpha \\
\left(\frac{n'}{|\alpha'|}\right)^{n'} \cdot \exp\left(-\frac{(\alpha')^2}{2}\right) \cdot \left(\frac{n'}{|\alpha'|} - |\alpha'| - \frac{m-\mu}{\sigma}\right) & \text{, for } \frac{m-\mu}{\sigma} \geq -\alpha'
\end{cases}
\tag{4.3}
$$

with a normalisation factor N, the expectation value of the gaussian core μ and its width σ. The constants $\alpha^{(')}$ define the transition point between the gaussian core and the power law tails, while $n^{(')}$ are the free parameters of the power laws. They are chosen for the signal model as they describe the LHCb resolution model quite well and allow for description of the asymmetry in the peak introduced by the loss of bremsstrahlung. The resulting probability density function (PDF) for the fit is then given as

$$
\begin{aligned}
PDF(m) = N_{sig} \cdot (f_0 \cdot (CB_0(m; \alpha_0, \alpha'_0, n_0, n'_0, \mu_0, \sigma_0)) \\
+ f_1 \cdot (CB_1(m; \alpha_1, \alpha'_1, n_1, n'_1, \mu_1, \sigma_1)) \\
+ f_2 \cdot (CB_2(m; \alpha_2, \alpha'_2, n_2, n'_2, \mu_2, \sigma_2))) \\
+ N_{bkg} \cdot (e^{\lambda \cdot m}),
\end{aligned}
$$

where N_{sig} and N_{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 ($x \in \{0; 1; 2\}$), where x denotes the bremsstrahlung category. In order to allow for a more stable fit, the parameters $\alpha^{(')}$ and $n^{(')}$, which describe the tails of the Crystal Ball functions, are obtained from fits to $B^+ \to K^+ J/\psi(\to e^+ e^-)$ simulation and fixed to these values in the fit to data. The other parameters are allowed to float in the fit to data, the results from the fit to simulation are used as starting points. These fits are performed separately for each bremsstrahlung category to account for differences in shape and resolution between the categories. In order to calculate sWeights, the three separate PDFs are fixed and summed up with fractions according to their yields. The resulting PDF is then fit to the whole dataset, where only the overall yields for signal and background are allowed to float. The resulting fits to data are shown in Figure 4.2.
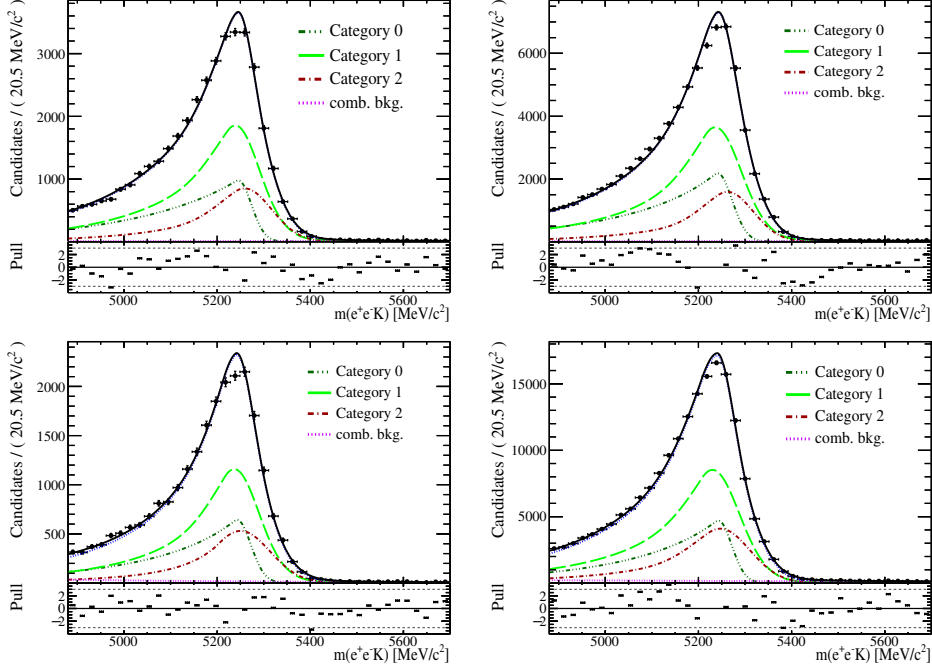
Figure 4.2: Fits to the control channel $K^+e^+e^-$ invariant mass distribution, split per year. 2011 (left) and 2012 (right) are shown on top, while 2015 (left) and 2016 (right) are shown on the bottom. The PDF consists of an exponential function to model the combinatorial background and one double-sided Crystal ball function per bremsstrahlung category, where either none ("0"), either ("1") or both ("2") of the electrons have bremsstrahlung corrections applied.

4.2.3 Data-simulation corrections

After subtracting remaining background from the $B^+ \to K^+J/\psi(\to e^+e^-)$ data, the simulated $B^+ \to K^+J/\psi(\to e^+e^-)$ signal can be compared to real signal in data. While the simulation used in the analysis describes the data quite well, some kinematic and particle identification (PID) variables show discrepancies which need to be corrected for to not bias the training of the multivariate classifier and efficiency determination. The procedures differ for kinematics and PID, thus in the following they are described separately.

Kinematic reweighting

In order to correct mismodelling of kinematics, the simulation is reweighted to match sWeighted data. A multivariate approach called GBReweighter [111] is chosen for this analysis. The GBReweighter uses a gradient boosted [112] decision tree to differentiate between data and simulation and then weight the simulation to reduce the separation power of the trained classifier. This process is iterated until the GBReweighter can not distinguish data from simulation any more. It takes as inputs the B^+ meson transverse momentum (p_T), an isolation variable for the B_s^0 defined by the CDF collaboration [113], in the following referred to as CDFiso and the track multiplicity, nTracks. The CDFiso is defined as

$$I_{\text{CDF}} = \frac{p_T(B_s^0)}{p_T(B_s^0) + \sum_{\text{tracks}} p_T(tr)}, \tag{4.4}$$

where $p_T(x)$ are the transverse momenta of the B_s^0 meson and all additional tracks in the event fulfilling the criterion $\sqrt{\Delta\eta^2 + \Delta\phi^2} < 1$, with the differences in pseudorapidity η and azimuthal angle ϕ between the B_s^0 and the track in question. To account for differences between periods of data taking, independent classifiers are trained per year of data taking. The trained classifier is also applied to $B_{(s)}^0 \rightarrow e^+e^-$ simulation to correct the kinematics.

Since the hadronisation of a b quark into B^+ and B_s^0 mesons differs, the two signal modes $B_s^0 \rightarrow e^+e^-$ and $B^0 \rightarrow e^+e^-$ need to be corrected differently. While for $B^0 \rightarrow e^+e^-$ the reweighting can be trained on $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+e^-)$ (since the hadronisation between B^+ and B^0 does not differ significantly due to isospin symmetry between the two), for $B_s^0 \rightarrow e^+e^-$ the reweighting is done using the decay of a B_s^0 meson, $B_s^0 \rightarrow \phi(\rightarrow K^+K^-) J/\psi (\rightarrow e^+e^-)$. The selection of this mode is kept as similar as possible to the selection of $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+e^-)$, with an additional requirement on the K^+K^- invariant mass to be close to the nominal mass of the ϕ of $1020\,\text{MeV}/c^2$. Since the CDFiso was not available on $B_s^0 \rightarrow \phi(\rightarrow K^+K^-) J/\psi (\rightarrow e^+e^-)$, the pseudorapidity of the B_s^0 is used instead for the reweighting. It was chosen since it shows similar deviations as the CDFiso on $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+e^-)$ and was found to yield similar performance for the reweighting.

The results of the reweighting on $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+e^-)$ are shown in Figure 4.3, the results for $B_s^0 \rightarrow \phi(\rightarrow K^+K^-) J/\psi (\rightarrow e^+e^-)$ are shown in Figure 4.4. The results are shown for 2012 and 2016, the results for 2011 and 2015 are identical to their higher statistics counterparts.
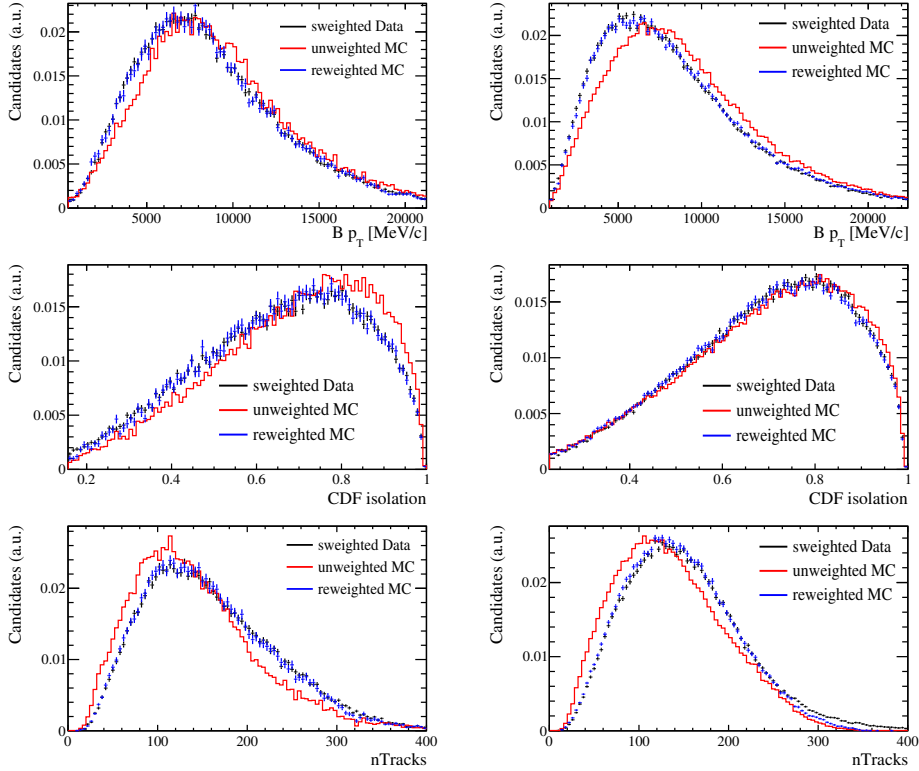
Figure 4.3: The input variables of the reweighting classifier trained on $B^+ \to K^+ J/\psi (\to e^+ e^-)$ compared for sWeighted data (black), unweighted simulation (red) and reweighted simulation (blue). The distributions for 2012 are shown on the left, the distributions for 2016 on the right.
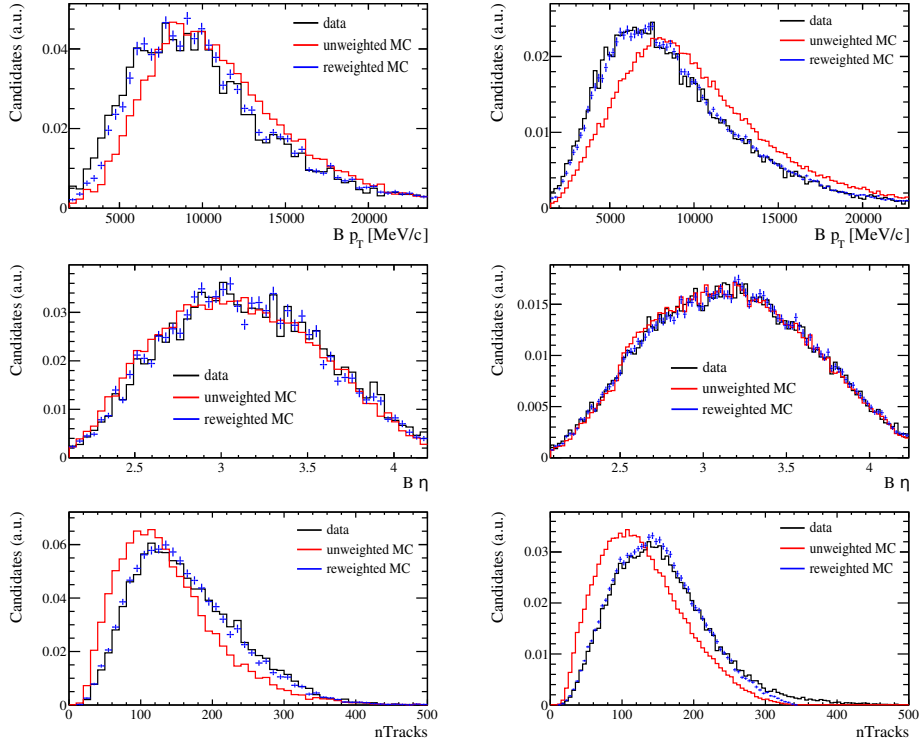
Figure 4.4: The input variables of the reweighting classifier trained on $B^0_s \to \phi(\to K^+ K^-) J/\psi(\to e^+ e^-)$ compared for data (black), unweighted simulation (red) and reweighted simulation (blue). The distributions for 2012 are shown on the left, the distributions for 2016 on the right.

PID resampling

The approach to correct the PID response of the detector, which is known to not be described well in simulation, differs from the correction of kinematics in that the PID is not reweighted but rather resampled from calibration data provided for the PIDCalib package [114]. These are high statistics data samples collected with selections designed to not bias the PID response. In order to achieve this, no PID selections are applied to the probed particle. Electron calibration samples come from the decay $B^+ \to K^+ J/\psi(\to e^+ e^-)$, while the kaon calibration data come from $D^0 \to K^- \pi^+$ decays, with the D^0 candidates originating from the decay $D^{*+} \to D^0 \pi^+$. In order to project out the signal components, the sPlot method is used similar as described above. The resampling is performed in multiple bins in the particle's momentum, pseudorapidity and track multiplicity for 2011 and 2012, while for 2015 and 2016 data, the particle's transverse momentum, pseudorapidity and multiplicity in the silicon pad detector (nSPDHits) are used. The binning has been calculated using a binning optimizer from the URANIA package [115], which aims to provide as fine a binning as possible to properly model the PID response as a function of particle kinematics, while keeping high enough statistics per bin to reduce statistical uncertainties of the result. The binning is summarized in Table A.5 and illustrated in Figure 4.5.

To resample the simulation, for each simulated candidate, the bin corresponding to its kinematics is chosen. Then from all calibration data that falls into this bin, one candidate is chosen at random, the PID response of which is assigned to the simulated candidate. To illustrate the effect of the resampling, the resulting changes to the PID variables of the kaon and one electron from $B^+ \to K^+ J/\psi(\to e^+ e^-)$ used in Section 4.2.5 are shown in Figure 4.6 for 2016.
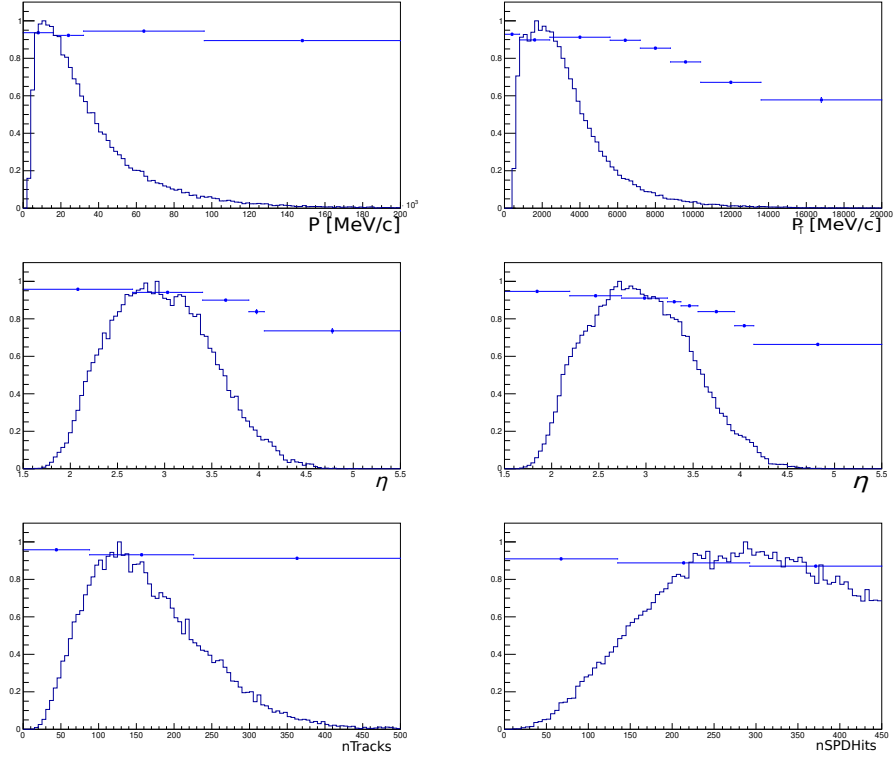
Figure 4.5: Illustration of the binning scheme used in the resampling. The binning scheme for Run 1 is shown on the left, the one for Run 2 on the right. The binning for the particle momentum is shown in the top plot (for Run 2 the transverse momentum is shown), the pseudorapidity in the middle and the multiplicity (nTracks in Run 1 and nSPDHits in Run 2) on the bottom. The distribution of the PIDCalib samples used for the calculation are overlaid.
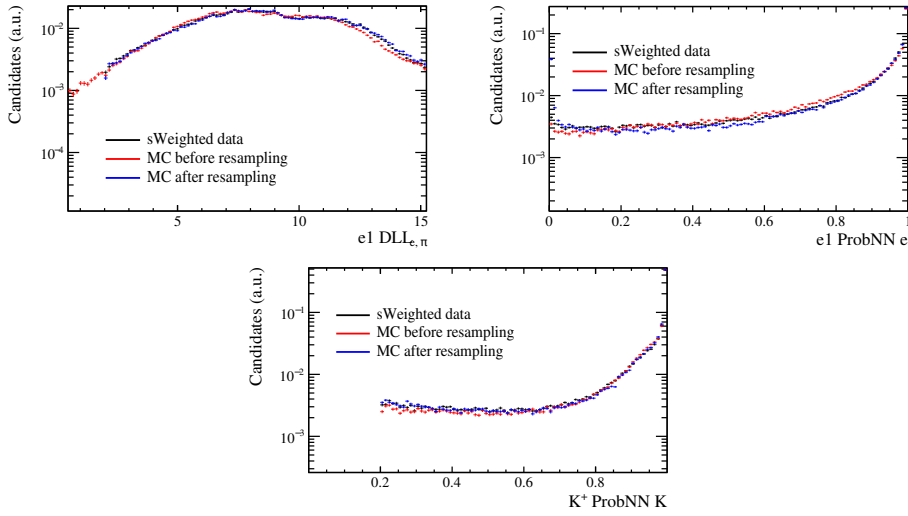
Figure 4.6: Resampled PID variables compared on sWeighted data (black) with reweighted simulation before (red) and after resampling (blue) for 2016 data and simulation. The electron PID shows two peaks, corresponding to the cases of having bremsstrahlung corrections applied or not having any corrections applied.

4.2.4 Multivariate analysis

After the preselection and corrections are applied, a multivariate analysis (MVA) is performed to reduce the combinatorial background dominating the $B_{(s)}^0 \to e^+e^-$ data. The classifier chosen here is a Boosted Decision Tree (BDT) [109]. In this analysis the BDT implemented in SCIKIT-LEARN [116] is used, where the boosting is done by the adaptive boosting algorithm (AdaBoost for short) [117]. For the training of the classifier, the $B_s^0 \to e^+e^-$ simulation is used as a signal proxy, while the upper mass sideband, which is expected to only include contributions from combinatorial background, is used as a background proxy. The lower mass sideband, while still dominated by combinatorial background, is not used in the training due to the presence of physical background, mainly from partial reconstruction of decays of the type $B \to he^+e^-$ with a hadron h as well as semileptonic decays, which are more similar to signal than to combinatorial background, thus worsening the overall separation power of the classifier. The lower mass sideband is still classified after the training to suppress the combinatorial background present.

To distinguish signal from background, the BDT is given a curated set of input variables, which were found to show good separation between signal and background, while showing little to no discrepancies between data and simulation after the

reweighting process. The BDT then in each of its nodes separates the training sample by a cut on one of the input variables, where the goal is to maximise the GINI impurity loss [118] given as

$$\Delta I_{\mathrm{GINI}} = \frac{s_{\mathrm{before}} \cdot b_{\mathrm{before}}}{n^2_{\mathrm{before}}} - \frac{s_1 \cdot b_1}{n^2_1} - \frac{s_2 \cdot b_2}{n^2_2}, \qquad (4.5)$$

where s is the number of signal candidates in the sample reaching the particular node, b is the corresponding number of background candidates and n is the number of total candidates in the sample. The subscript "before" indicates the sample that is present before the split, while 1 and 2 indicate the two subsamples present after the split. This step is repeated for each of the resulting subsamples until the maximum depth of the classifier is reached or no further separation is possible. After each event is classified as either signal or background, the procedure described above is repeated with events misclassified being boosted using AdaBoost, meaning they receive higher weight in the training. Each event is then assigned a classification value based on the average of the trained classifiers. To avoid overtraining a cross-validation [119] is performed to ensure that every event is classified by an independent classifier. In this so called k-folding, the data and simulation are split into k parts, where the BDT is trained on k-1 parts and then applied to the remaining part. This is then repeated until every fold has been classified. Here, $k = 6$ was chosen. Since none of the BDTs have been trained on the lower mass sideband or signal region, the average of all classifiers is assigned as classification value to these regions.

The input variables used in the training are:

- the B^0_s transverse momentum,

- the B^0_s CDFiso,

- the B^0_s χ^2_{IP},

- the B^0_s flight distance

- the square root of the minimum χ^2_{IP} of the electrons,

- the distance of closest approach between the two electrons,

- the absolute difference in the electrons pseudorapidity $\delta\eta$

- the sum of the electron track isolations for the least isolating track (i.e. closest to the signal decay),

- the sum of the electron track isolations for the next to least isolating track (i.e. next-to-closest to the signal decay).

The track isolations are based on BDTs initially developed for the $\tau^+ \to \mu^+ \mu^+ \mu^-$ and $B_s^0 \to \mu^+ \mu^-$ analyses [120] and were made available for $B_{(s)}^0 \to e^+ e^-$ as well. They are of significant importance for the BDT training of $B_{(s)}^0 \to e^+ e^-$, increasing the separation power significantly when added to the training. This variable is used as a measure for the isolation of the signal electron tracks from other tracks in the detector. In the following, these other tracks not associated with the signal decay will be referred to as additional tracks. The input variables for the BDT calculating the track isolation are

- the logarithm of the minimum χ_{IP}^2 of the additional track with respect to any primary vertex in the event,

- the distance between the additional track and the electron track's primary vertex,

- the distance between the electron track's origin vertex and the $B_{(s)}^0 \to e^+ e^-$ decay vertex,

- the logarithm of the distance of closest approach between the additional track and the electron track,

- the angle between the additional track's momentum and the electron track's momentum,

- $\text{f}_c = \frac{|\vec{P}_e + \vec{P_{trk}}| \sin \alpha^{e+trk,PV}}{|\vec{P}_e + \vec{P_{trk}}| \sin \alpha^{e+trk,PV} + p_{T,e} + p_{T,trk}}$, where $\alpha^{e+trk,PV}$ is the angle between the sum of the momenta of electron track and additional track and the direction defined by the electron track's primary vertex and the additional track's vertex.

This variable is calculated with respect to the best isolating and second best isolating tracks in the event, resulting in two classifier values for each electron. In the $B_{(s)}^0 \to e^+ e^-$ BDT the sum of the two electrons values, calculated for the best isolating and second best isolating track separately, enter as inputs, resulting in two input variables for the BDT classifier. Different combinations of these isolation variables have been tested but were found to worsen classifier performance compared to the sum.

A comparison of the input variables for the BDT training between signal simulation and the upper mass sideband in data for the combined Run 2 dataset can be found in Figure 4.7.
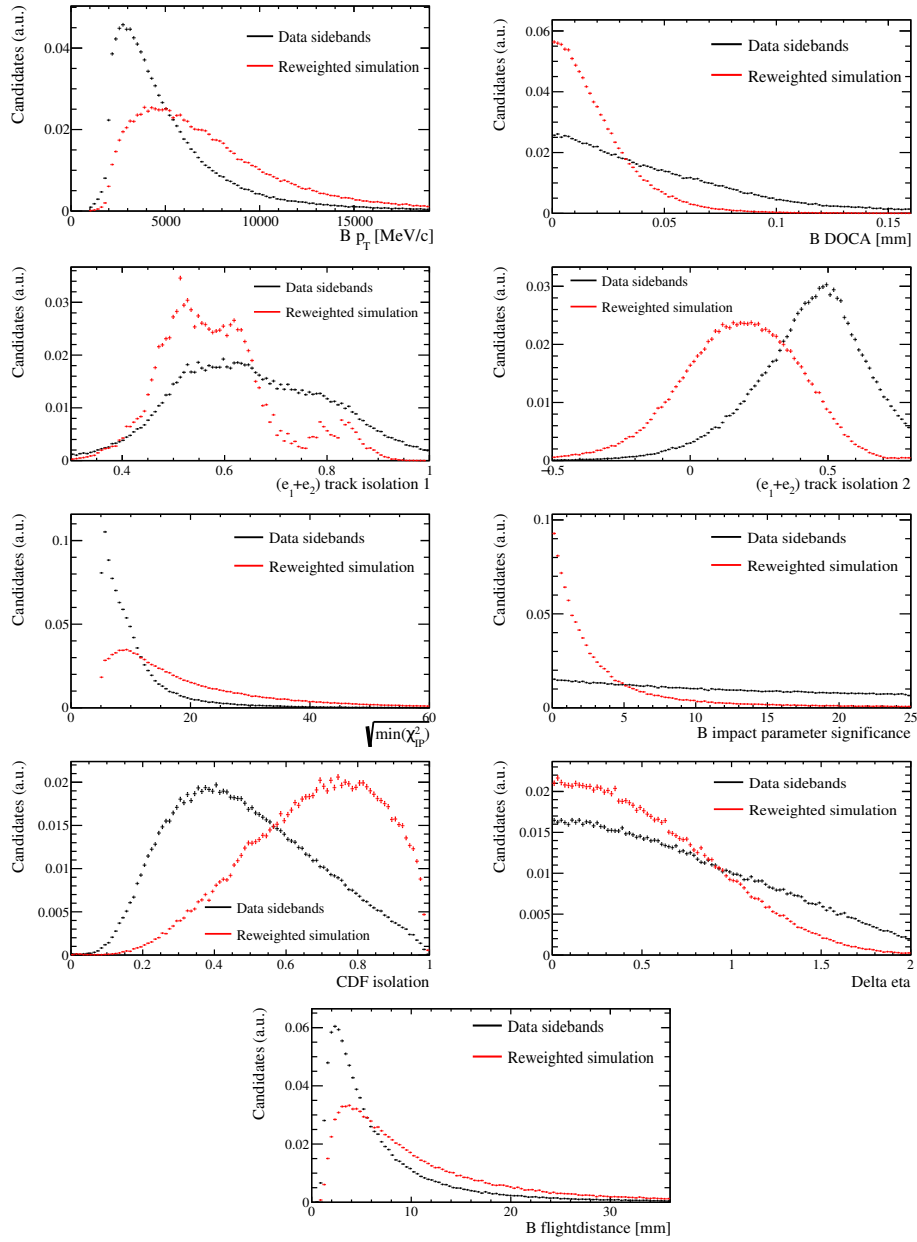
Figure 4.7: Comparison of the BDT classifier input variables between calibrated simulation (black) and the upper mass sideband in data (red) for the combined Run 2 dataset.

Since the 2011 and 2015 data contain far fewer candidates than their 2012 and 2016

counterparts, the BDT is not trained per year but rather per run (thus merging 2011+2012 and 2015+2016) to avoid a decrease in separation power that can come from low statistics in the training sample. The distribution of the trained BDT classifiers are shown in Figure 4.8. As a measure for the BDT performance, the area under the receiver operating characteristic (RoC) curve is calculated, where the RoC curve is defined as the signal efficiency as a function of background efficiency. In case of perfect classification, the area under the RoC curve would be equal to one, while random guessing would return 0.5. The RoC curves are shown in Figure 4.9. The trained BDTs perform quite well on the upper mass sideband, while as expected the performance is significantly lower on the lower mass sideband due to the presence of physical background the BDT can not suppress. This is visible in Figure 4.8 as well, where the classifier on data shows a signal-like component reaching high classification values.
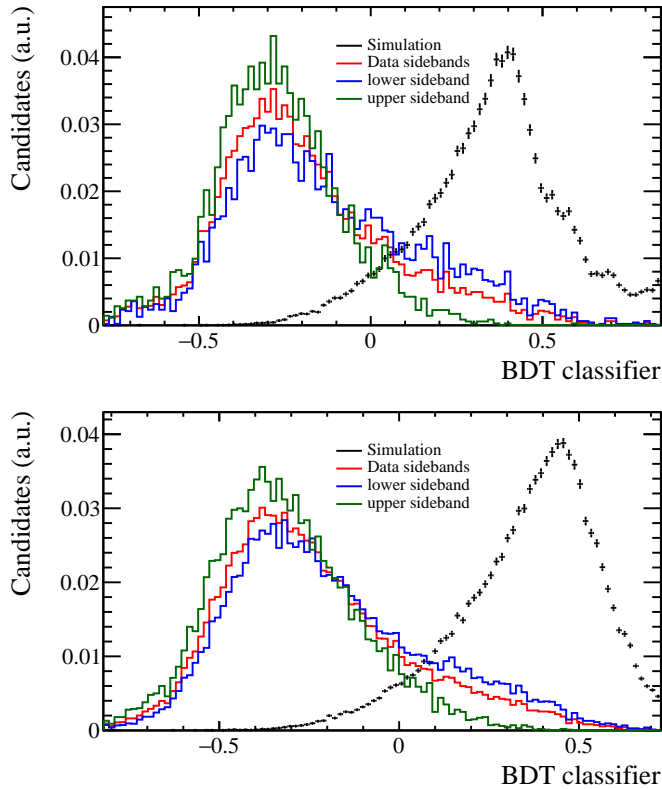
Figure 4.8: Resulting BDT classifier response. Shown are simulated events (black), the complete data (red), the upper mass sideband (green) and the lower mass sideband (blue). Run 1 is shown on top, Run 2 on the bottom.
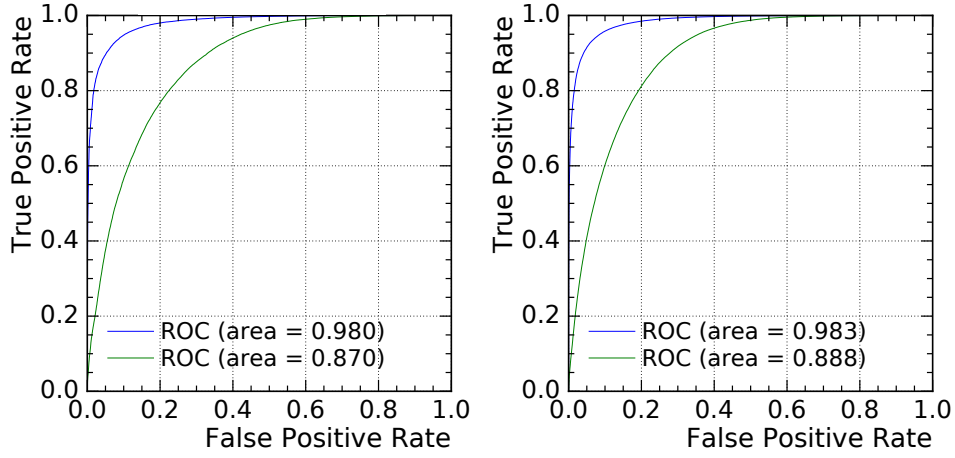
Figure 4.9: ROC curve of the BDT classifier on the upper mass sideband (blue) and the lower mass sideband (green). Run 1 is shown on the left, Run 2 on the right.

The same classifier is also applied to $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ by averaging all folds as was done for the lower mass sideband. While application of the BDT is not necessary in this mode to suppress background, the control mode allows to check for agreement of the BDT classifier between data and simulation as well as cancellation of systematic uncertainties. To ensure similarity to the classifier on $B^0_{(s)} \rightarrow e^+ e^-$, all input variables from the B^0_s are replaced with their counterparts from the B^+ while the distance of closest approach is calculated for the J/ψ to ensure it is calculated between the electrons instead of the K^+ and J/ψ candidates.

Optimisation of the MVA selection

After the training of the multivariate classifier, the requirement applied to it is not chosen by hand but is instead chosen to maximise the Punzi figure of merit (FoM) [121] given as

$$\text{FoM} = \frac{\epsilon_{\text{sig}}}{\sqrt{N_{\text{bkg}}} + {}^{3}\!/\!{}_{2}}, \tag{4.6}$$

with the signal efficiency ϵ_{sig} taken from reweighted simulation and the expected number of background candidates in the signal region $\sqrt{N_{\text{bkg}}}$, which is obtained from extrapolating the mass sidebands into the blinded region. To optimize the selection, the cut on the BDT classifier is scanned in steps of 0.01 and the maximum

of the resulting distribution of the FoM is chosen as cut-point. The FoM is shown in Figure 4.10 separately for the Run 1 and Run 2 samples.
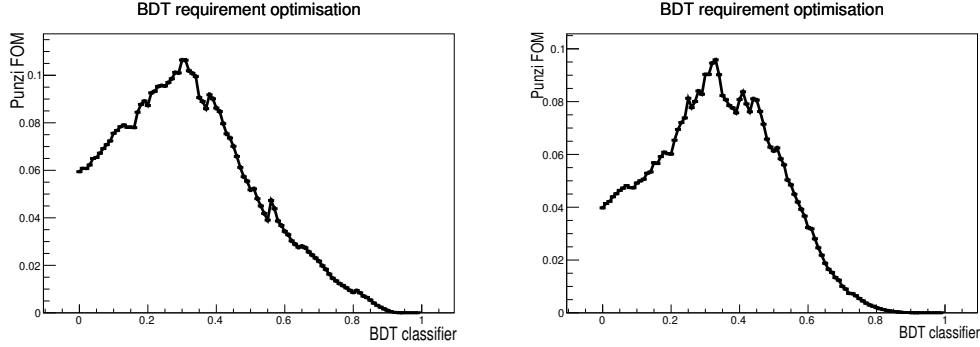
Figure 4.10: Distribution of the Figure of Merit depending on the requirement on the BDT classifier for the Run 1 dataset (left) and for the Run 2 dataset (right).

Some spiking structures can be seen in the distribution of the figure of merit. These spikes can occur at the point where the background is reduced to zero. As they occur in a region separate from the global maximum of the figure of merit, they are not studied further as they can not affect the choice of the cut point.

4.2.5 Requirements on particle identification

While the BDT is capable of suppressing most combinatorial background present in the data, backgrounds originating from misidentification of final-state particles need to be handled as well. To achieve this, cuts are applied on the probability of the particle to be an electron. These probabilities are estimated from a neural network based approach, thus they are called ProbNNe. A selection based on the difference in log-likelihood of the particle being an electron and the particle being a pion as is applied during the stripping (see Section 4.2.1) has also been tested. While this selection would be better at separating electrons from pions, it was found to be significantly worse separating electrons from kaons, worsening the overall background suppression.

The selection of ProbNNe was chosen by hand to reduce the expected number of candidates from backgrounds of the type $B \to hh^{(\prime)}$, where the $h^{(\prime)}$ are either pions or kaons, to a negligible level. These backgrounds are particularly dangerous in $B^0_{(s)} \to e^+e^-$ as they would show a clear peak close to the B^0_s or B^0 mass, where $B^0_{(s)} \to e^+e^-$ would also be expected. The selections are summarised in Table A.6. The cut-point differs as different tunings are used between Run 1 and Run 2 data.

47

4.3 Normalisation

The following section describes the determination of inputs necessary for the calculation of an upper limit on the branching fraction of $B^0_{(s)} \to e^+e^-$. First, efficiencies (see Section 4.3.1) and systematic uncertainties (see Section 4.3.6) need to be determined. Since the determination of the trigger efficiency (see Section 4.3.2) and the calibration of PID efficiencies (see Section 4.3.3) are done using different approaches from the other selection efficiencies, they are discussed separately. The total selection efficiencies are discussed in Section 4.3.4.

4.3.1 Efficiencies

In order to convert a measured number of candidates for a given decay into the decay's branching fraction, it is necessary to determine the efficiency of the selection. In general, the efficiency ϵ of a given selection step is defined as

$$\epsilon = \frac{N_{\text{passed}}}{N_{\text{total}}}, \tag{4.7}$$

with uncertainties given as

$$\sigma(\epsilon) = \sqrt{\frac{\epsilon \cdot (1 - \epsilon)}{N_{\text{total}}}}. \tag{4.8}$$

In the case of weighted simulation, these formulae turn into

$$\epsilon = \frac{\sum_{\text{passed}} \omega}{\sum_{\text{total}} \omega}, \tag{4.9}$$

with uncertainties given as

$$\sigma(\epsilon) = \sqrt{\frac{\epsilon \cdot (1 - \epsilon)}{\sum_{\text{total}} \omega}}, \tag{4.10}$$

where the summation is performed over the weights ω associated to the candidates. The efficiencies are determined separately for each selection step relative to the previous steps. Thus, the whole selection efficiency is given as

$$\epsilon = \epsilon_{\text{geo}} \cdot \epsilon_{\text{stripxreco | geo}} \cdot \epsilon_{\text{sel | stripxreco}} \cdot \epsilon_{\text{PID | sel}} \cdot \epsilon_{\text{BDT | PID}} \cdot \epsilon_{\text{trig | BDT}}. \tag{4.11}$$

Efficiencies are calculated taking into account the kinematic reweighting where possible. However, these weights can not be applied to the simulation before

reconstruction of the candidates, so the efficiency of requiring all final state particles to be within the geometrical acceptance of the detector as well as the efficiency of the stripping and reconstruction (ϵ_{geo} and $\epsilon_{\text{stripxreco | geo}}$ above) do not have weights applied to them.

In this analysis, where possible, selection efficiencies are calculated split into three bremsstrahlung categories. This is again not possible for ϵ_{geo} and $\epsilon_{\text{stripxreco | geo}}$, since information about the emission of bremsstrahlung is not accessible before the reconstruction. The relative fractions of the three categories are taken from simulation in $B^0_{(s)} \to e^+e^-$. The agreement of these fractions between data and simulation is checked using $B^+ \to K^+ J/\psi(\to e^+e^-)$.

While most efficiencies are determined from simulation, the efficiencies of the trigger and PID selections are determined from data instead, since these efficiencies are not well described in simulation and do not get corrected by the kinematic reweighting.

4.3.2 Trigger calibration

The efficiency of the chosen trigger selection is determined from data using the TISTOS method [122], because only triggered events are saved, rendering the amount of candidates before the trigger inaccessible on data. Since the TISTOS method requires access to signal yields in different trigger categories it is not applicable to $B^0_{(s)} \to e^+e^-$ however. It is therefore applied to $B^+ \to K^+ J/\psi(\to e^+e^-)$ and transferred to $B^0_{(s)} \to e^+e^-$ by mapping the efficiency in the maximum electron transverse momentum and impact parameter. While requirements on the trigger are applied in the beginning of the analysis chain, the efficiency is defined after BDT and PID selections have already been applied. This is necessary here to have a sample as clean as possible.

The approach of the TISTOS method is to measure the efficiency of the trigger selection on a TIS sample, it is then given as

$$\epsilon_{trig} = \frac{N_{\text{trig}}}{N_{\text{TIS}}} \cdot \epsilon_{\text{TIS}}. \tag{4.12}$$

This however still requires the knowledge of an inaccessible quantity, the overall TIS efficiency ϵ_{TIS}. While measuring it for the whole sample is impossible, it can be measured for a specific TOS sample from the overlap of TIS and TOS candidates (referred to as TISTOS) as

$$\epsilon_{\text{TIS|TOS}} = \frac{N_{\text{TISTOS}}}{N_{\text{TOS}}}. \tag{4.13}$$

If the TIS efficiency does not depend on the chosen TOS sample, meaning $\epsilon_{\text{TIS|TOS}} = \epsilon_{\text{TIS}}$, this allows to measure the trigger efficiency as

$$\epsilon_{\text{trig}} = \frac{N_{\text{trig}}}{N_{\text{TIS}}} \cdot \frac{N_{\text{TISTOS}}}{N_{\text{TOS}}}, \tag{4.14}$$

where the yields in all four trigger categories can directly be measured from data. This however leaves the TIS efficiency being independent of the sample it is measured on as the main assumption of this approach, implying that the triggered candidate is independent of the rest of the event. This assumption does not hold for B mesons, as the b quarks which form them are always produced in pairs, leading to kinematic correlation with the rest of the event. However in small enough regions of the B meson phase-space the candidate can be assumed to be uncorrelated from the rest of the event. As a consequence, in order to avoid a bias in the trigger efficiency, the efficiency estimation is carried out in several subregions of the B meson phase-space, here defined as three bins each in the B mesons transverse and longitudinal momentum, leading to the formula

$$\epsilon_{\text{trig}} = \frac{N_{\text{trig}}}{\sum\limits_{i} \frac{N^i_{\text{TIS}} \cdot N^i_{\text{TOS}}}{N^i_{\text{TISTOS}}}}, \tag{4.15}$$

where the summation is performed over said bins. The binning is chosen so that the TISTOS sample is uniformly distributed in the 1D binnings. This is done to avoid (almost) empty bins in one category, since the uncertainty is mainly driven by the statistics in the smallest sample, which is the TISTOS sample by definition.

It is important to carry out this estimation independently for each category of bremsstrahlung, since the efficiency differs significantly. This is the case especially for L0, where no bremsstrahlung corrections have been applied yet, thus leading to electrons that emitted bremsstrahlung to be less likely to pass the energy threshold required in the L0 trigger. While the bremsstrahlung categories are defined by the applied corrections (since it is not known which electrons really emitted bremsstrahlung), most electrons affected end up in the higher categories of bremsstrahlung, leading to a clear hierarchy in L0 efficiency shown in Table 4.3 for the control mode $B^+ \to K^+ J/\psi (\to e^+ e^-)$.

Table 4.3: Trigger efficiency of the L0 selection on data for the control mode $B^+ \to K^+ J/\psi (\to e^+ e^-)$ computed using the TISTOS method.

Brem.	Efficiency [%]			
	2011	2012	2015	2016
0	71.14 ± 1.48	69.28 ± 2.96	74.07 ± 2.11	77.98 ± 1.87
1	69.00 ± 0.92	65.40 ± 2.12	68.88 ± 1.63	69.96 ± 1.37
2	66.89 ± 1.26	63.97 ± 2.65	64.74 ± 1.85	68.53 ± 1.68

As mentioned before, the same procedure is not possible for $B^0_{(s)} \to e^+ e^-$, which is still blinded at this point. In order to still estimate the trigger efficiency in $B^0_{(s)} \to e^+ e^-$ from data, the efficiency is transferred from the control mode. For this purpose, the efficiency calculation is done in bins of the maximum electron transverse momentum and impact parameter. The resulting map is then folded with the corresponding distributions from $B^0_{(s)} \to e^+ e^-$ simulation to calculate the trigger efficiency.

4.3.3 PID efficiency calibration

Similarly to the trigger efficiency, the PID efficiencies are not taken from simulation, but rather are estimated from the same calibration data used in the resampling in Section 4.2.3. The efficiencies are not taken from the resampled simulation, since the resampling process does not replicate the correlation between different PID variables. Since cuts both on PID and ProbNN are applied to the electrons (and the kaon from the control mode), the real efficiency would not be reproduced when calculated from resampled simulation. A selection based solely on PID has been investigated, but was found to yield worse performance separating electrons from kaons, increasing the retained background significantly, thus decreasing sensitivity.

The approach to estimate the efficiency is similar to what is done during the resampling. Instead of assigning a new PID response to the simulated candidates, they are assigned the efficiency that a given PID cut has in their corresponding bin, where previous PID selections are taken into account when calculating the efficiency. In order to calculate the efficiency in each bin from calibration data, background is subtracted using sWeights.

4.3.4 Total selection efficiencies

The total selection efficiencies of $B_s^0 \to e^+ e^-$ and $B^+ \to K^+ J/\psi (\to e^+ e^-)$, split by bremsstrahlung category, as well as their ratio, which enters the normalisation constant, are shown in Table 4.4. They include the relative fractions of the three categories of bremsstrahlung, which are taken from simulation and are cross-checked with $B^+ \to K^+ J/\psi (\to e^+ e^-)$ data.

Table 4.4: Total selection efficiencies, split by decay channel, year and bremsstrahlung category. The ratios needed for the normalisation constants are also reported.

Brem.	$\epsilon_{B_{(s)}^0 \to e^+ e^-}$ [%]	$\epsilon_{B^+ \to K^+ J/\psi (\to e^+ e^-)}$ [%]	$\dfrac{\epsilon_{B^+ \to K^+ J/\psi (\to e^+ e^-)}}{\epsilon_{B_{(s)}^0 \to e^+ e^-}}$
2011			
0	0.1087 ± 0.0114	0.0173 ± 0.0016	0.1599 ± 0.0224
1	0.3243 ± 0.0293	0.0460 ± 0.0031	0.1419 ± 0.0159
2	0.2543 ± 0.0356	0.0215 ± 0.0030	0.0843 ± 0.0168
2012			
0	0.0952 ± 0.0073	0.0131 ± 0.0015	0.1377 ± 0.0150
1	0.2842 ± 0.0172	0.0309 ± 0.0017	0.1088 ± 0.0089
2	0.1873 ± 0.0215	0.0162 ± 0.0013	0.0865 ± 0.0121
2015			
0	0.1215 ± 0.0086	0.0171 ± 0.0017	0.1407 ± 0.0101
1	0.3501 ± 0.0254	0.0392 ± 0.0035	0.1119 ± 0.0129
2	0.2469 ± 0.0224	0.0199 ± 0.0024	0.0806 ± 0.0122
2016			
0	0.1353 ± 0.0039	0.0202 ± 0.0007	0.1495 ± 0.0071
1	0.3889 ± 0.0112	0.0461 ± 0.0013	0.1187 ± 0.0048
2	0.2660 ± 0.0133	0.0229 ± 0.0009	0.0861 ± 0.0056

The efficiency for the $B_s^0 \to e^+ e^-$ mode is found to be significantly larger than for the normalisation mode. This difference originates mainly from the third track in the final state at three separate occasions. First, the additional kaon needs to be reconstructed and matched to the same vertex as the electrons, necessitating additional selection requirements and thus leading to a significantly lower reconstruction efficiency. Second, the selection applied to $B^+ \to K^+ J/\psi (\to e^+ e^-)$ differs from the one applied

to $B^0_{(s)} \rightarrow e^+ e^-$. The trigger selection is optimized for selecting a two-body decay, leading to a drop in efficiency for the three-body control mode, as no three-body triggers are considered. The additional preselection applied to $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$ before the sWeighting procedure leads to a further drop in efficiency compared to $B^0_{(s)} \rightarrow e^+ e^-$. Lastly, the additional track causes the electrons to be less significantly isolated, leading to worse separation power of signal and background from the BDT-based track isolation, which then causes a drop in BDT efficiency, as for the BDT $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$ looks more background-like.

Comparing the efficiencies between the years, they are very similar within each run period. However, the efficiencies tend to be significantly higher in Run 2 compared to Run 1. This is expected, since the triggers and PID variables were retuned for the Run 2 data taking, leading to better performance. The ratio of efficiencies is consistent between the years, as is expected, since the aforementioned improvements in Run 2 should affect $B^0_{(s)} \rightarrow e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$ in a similar manner and thus cancel in this ratio. However, the ratios show a clear hierarchy between the bremsstrahlung categories. This trend is consistent with the trends observed in the efficiency of the control channel specific preselection and BDT, so it also originates from the additional track present in $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$.

4.3.5 Normalisation constants

The measurement of the normalisation constant for $B^0_{(s)} \rightarrow e^+ e^-$ needs several inputs (as shown in Equation (4.2)):

- The ratio of selection efficiencies of $B^0_{(s)} \rightarrow e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$

- The signal yield in $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$ after the selection

- The branching fraction of $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$

- The ratio of hadronisation fractions of b quarks into B^0_s or B^+/B^0 mesons

While the ratio of efficiencies is given above in Table 4.4, the other inputs are described in the following.

The ratio of hadronisation fractions of b quarks into B^0_s or B^+/B^0 mesons has been measured by LHCb [123] as

$$\frac{f_s}{f_d} = 0.259 \pm 0.015 \tag{4.16}$$

in 2011 (where $f_u = f_d$ is assumed). Because of the difference in centre-of-mass energy compared to 2012 and Run 2, correction factors on this ratio have been

measured to be 1.026 ± 0.023 in 2012 and 1.068 ± 0.022 in Run 2 [124]. The correction factor for 2012 is neglected as it is negligible within the uncertainty on $\frac{f_s}{f_d}$.

The branching fraction of the control channel is taken from the PDG [107] as the product of the branching fractions for $B^+ \to K^+ J/\psi$ and $J/\psi \to e^+ e^-$ as

$$\mathcal{B}(B^+ \to K^+ J/\psi(\to e^+ e^-)) = (1.010 \pm 0.028) \cdot 10^{-3} \times (5.971 \pm 0.032) \cdot 10^{-2}$$
$$= (6.03 \pm 0.17) \cdot 10^{-5}. \tag{4.17}$$

This leaves only the yield of the control mode after the full selection as needed input for the normalisation constant. It is determined from a fit to the fully selected data using the same model as described in Section 4.2.2. The fit is visualised in Figure 4.11, the yields of each bremsstrahlung category are summarised in Table 4.5.
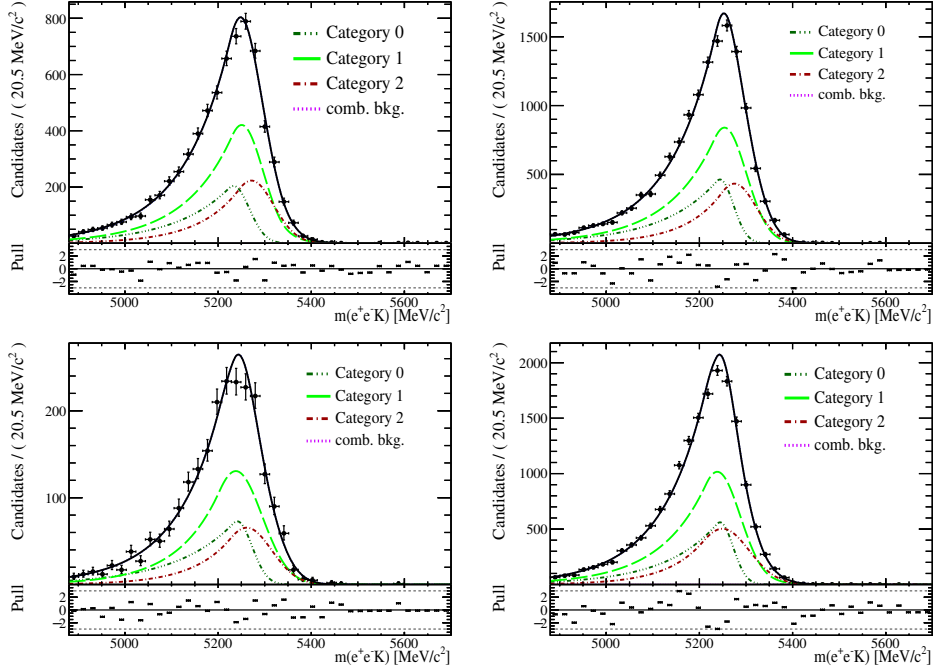
Figure 4.11: Fits to the control channel $K^+ e^+ e^-$ invariant mass distribution after the whole selection and split per year. 2011 (left) and 2012 (right) are shown on top, while 2015 (left) and 2016 (right) are shown on the bottom. The PDF consists of an exponential function to model the combinatorial background and one double-sided Crystal ball function per bremsstrahlung category, where either none ("0"), either ("1") or both ("2") of the electrons have bremsstrahlung corrections applied.

Table 4.5: Signal yields of the normalisation mode $B^+ \to K^+ J/\psi(\to e^+ e^-)$ computed from the fits shown in Figure 4.11.

Brem.	2011	2012	2015	2016
0	1594 ± 19	3451 ± 30	1006 ± 16	7035 ± 41
1	3503 ± 42	6970 ± 60	1962 ± 31	14829 ± 87
2	1744 ± 22	3219 ± 28	917 ± 15	7327 ± 43

With these values, the partial normalisation constants are determined per year. They are shown in Table 4.6.

Table 4.6: Partial normalisation constants $\alpha'_{B^0_{(s)}}$ split by bremsstrahlung category and year.

Brem.	$\alpha'(2011)$ $[10^{-4}]$	$\alpha'(2012)$ $[10^{-4}]$	$\alpha'(2015)$ $[10^{-4}]$	$\alpha'(2016)$ $[10^{-4}]$
0	1.003 ± 0.141	0.399 ± 0.046	1.398 ± 0.103	0.2125 ± 0.0102
1	0.405 ± 0.046	0.156 ± 0.013	0.570 ± 0.066	0.0800 ± 0.0033
2	0.483 ± 0.097	0.269 ± 0.038	0.879 ± 0.134	0.1175 ± 0.0077

These normalisation constants can then be combined for each run period as the harmonic sum over the years comprising that running period:

$$\frac{1}{\alpha_{\text{Run}}} = \sum_{years} \frac{1}{\alpha_{\text{year}}}, \tag{4.18}$$

leading to the partial normalisation constants shown in Table 4.7. They are combined with the hadronisation fraction and the branching fraction of the control mode into the full normalisation constants shown in Table 4.8

Table 4.7: Partial normalisation constants $\alpha'_{B^0_{(s)}}$ split by bremsstrahlung category and Run. These are used in the limit calculation.

Brem.	$\alpha'(\text{Run 1})$ $[10^{-5}]$	$\alpha'(\text{Run 2})$ $[10^{-5}]$
0	2.85 ± 0.24	1.84 ± 0.08
1	1.13 ± 0.08	0.702 ± 0.027
2	1.73 ± 0.20	1.04 ± 0.06

Table 4.8: Full normalisation constants $\alpha_{B^0_s}$ split by bremsstrahlung category and Run.

Brem.	α(Run 1) $[10^{-8}]$	α(Run 2) $[10^{-8}]$
0	0.66 ± 0.07	0.402 ± 0.031
1	0.262 ± 0.024	0.153 ± 0.012
2	0.40 ± 0.05	0.226 ± 0.020

4.3.6 Systematic uncertainties

In order to validate and check the selection steps outlined in Section 4.2, various cross-checks are performed on the different analysis steps. Where deviations are seen, systematic uncertainties are assigned. They are calculated separately for each year of data taking and each bremsstrahlung category where applicable.

Estimation of trigger efficiencies

While the binning done during estimation of the trigger efficiency, as described in Section 4.3.2, allows to reduce the effect of the correlation between triggered candidate and the rest of the event, it is not fine enough to fully describe this correlation. In order to quantify the remaining bias from not using infinitesimally small bins in the TISTOS method, the same method described in Section 4.3.2 is also applied to $B^+ \to K^+ J/\psi (\to e^+ e^-)$ simulation instead of data. While on simulation the trigger efficiency might be different from what it is on data due to mismodelling, the true efficiency on simulation is known and can be measured simply by applying the trigger requirements on simulation. Since applying the TISTOS method to simulation is subject to the same mismodelling as the cut-based efficiency determination, this allows to compare the true efficiency with the estimate from the TISTOS method and calculate the remaining bias. This bias is then assigned as a systematic uncertainty on the trigger efficiency. The bias is found to be up to $5.3\,\%$ and is one of the largest sources of systematic uncertainty of the analysis. The large bias is caused from the relatively low statistics in $B^+ \to K^+ J/\psi (\to e^+ e^-)$, especially of events being classified as TIS and TOS simultaneously, thus limiting the amount of bins used for the efficiency determination.

Differences between data and simulation

The kinematic corrections applied to $B^+ \to K^+ J\!/\!\psi(\to e^+e^-)$ and $B^0_{(s)} \to e^+e^-$ are tested to check how well the correction works by comparing the efficiencies of the selection estimated from data and simulation. Since the efficiency on data is obviously not accessible for $B^0_{(s)} \to e^+e^-$, this comparison is done on the control mode for the preselection and BDT selection. While data and simulation are found to be in agreement for the preselection, with only small deviations of up to $0.6\,\%$ observed, the BDT selection shows a larger discrepancy of up to $4\,\%$, showing that not all discrepancies could be corrected for during the reweighting process.

Comparison of $B^0_s \to e^+e^-$ and $B^0 \to e^+e^-$

While both, $B^0_s \to e^+e^-$ and $B^0 \to e^+e^-$ are analysed, the efficiencies are only determined for the B^0_s mode and assumed to be the same for the B^0 mode. To check this assumption, the efficiencies were compared on 2012 and 2016 simulation. Small differences of $2.5\,\%$ in 2012 and $2.1\,\%$ in 2016 were found and assigned as systematic uncertainty on the total $B^0 \to e^+e^-$ efficiency for their corresponding run periods.

Estimation of PID efficiencies

In the procedure described in Section 4.3.3 for estimating the PID efficiency from calibration data, there are two possible sources of systematic uncertainty. The first is the binning chosen for the efficiency calculation. To quantify how dependent the calculation is on the chosen binning, the efficiency is estimated again with the number of bins doubled in one variable at a time. The relative differences obtained from repeating this procedure for each binning variable are then added in quadrature, the result is assigned as systematic uncertainty. This leads to systematic uncertainties of up to $3\,\%$.

The second possible source of systematic uncertainty comes from the way background is subtracted from the calibration data. As the sPlot method is used, it relies on the variable used to discriminate signal from background and the variables used to parametrise the PID response to be uncorrelated. Since in the case of the electron calibration data, which come from $B^+ \to K^+ J\!/\!\psi(\to e^+e^-)$ decays, the discriminating variable is the invariant mass of the reconstructed B^+ meson, this is not the case here. The mass distribution shows long tails to either side from either under- or overcorrecting bremsstrahlung losses, which are directly correlated with the electron's kinematics. Together with the high background level present in the calibration

data, this can lead to incorrect projections of background-subtracted signal into the kinematic bins and thus bias the efficiency.

To check how much this affects the efficiency estimate, the efficiency is also estimated from fitting the $B^+ \to K^+ J\!/\!\psi (\to e^+e^-)$ data in each bin before and after applying the PID selection, thus circumventing the above problem. This study is carried out using a tag&probe approach, where one electron from the $J\!/\!\psi$ serves as a tag with strong PID requirements (PIDe > 5) applied, while the second electron is used as a probe on which the efficiency calculation is carried out. This procedure is done twice, with each electron serving as probe once, the results then being averaged to compare to the nominal efficiency. The relative difference is found to be up to $4.9\,\%$, which is assigned as a systematic uncertainty.

Corrections to the track reconstruction efficiency

The efficiencies of the reconstruction are taken from simulation. Since the signal and normalisation mode do not have the same final state, not all potential mismodelling of the track reconstruction efficiency, which is dependent on the detector condition, cancels in the ratio. To correct the track reconstruction efficiency of the kaon track, centrally produced correction maps [125] are used. These maps parametrise the track reconstruction efficiency in the particle's momentum and pseudorapidity and are convoluted with the kinematics of the investigated track to obtain a correction. The correction to the kaon track from $B^+ \to K^+ J\!/\!\psi (\to e^+e^-)$ was found to be unity within the uncertainty of the method, so no correction is applied. The only systematic uncertainty applied to the track reconstruction then comes from the uncertainty on hadronic interactions, where the kaon interacted with the detector material. From simulation of $B^0 \to K^{*0}(\to K^+\pi^-) J\!/\!\psi$, the fraction of kaons that can not be reconstructed due to this effect was found to be $11\,\%$ with a relative uncertainty of $10\,\%$. Thus, a systematic of $1.1\,\%$ is applied to the track reconstruction efficiency.

Fractions of bremsstrahlung categories

Since the selection efficiencies have been determined separately per bremsstrahlung category, precise knowledge of the relative fractions of these categories is important to determine the normalisation constant. Since these fractions differ between data and simulation on the control mode after the full selection, the impact of this difference on the overall normalisation needs to be checked. Therefore, an overall normalisation constant is computed from the normalisation constants reported in Section 4.3.5 with both sets of bremsstrahlung fractions. The relative difference

of 3.6 % in Run 1 and 4.0 % in Run 2 is assigned as a systematic uncertainty. The comparison is summarised in Table 4.9.

Table 4.9: Comparison of the relative fractions between bremsstrahlung categories as determined from control channel simulation and recorded data after the full selection. Since these fractions are required to sum up to unity, the uncertainties are fully correlated.

Brem.	Fraction			
	2011 [%]	2012 [%]	2015 [%]	2016 [%]
$B^+ \to K^+ J/\psi(\to e^+e^-)$ data (after full selection)				
0	23.04 ± 0.58	25.14 ± 0.42	24.35 ± 0.72	22.85 ± 0.26
1	50.73 ± 0.69	50.62 ± 0.48	49.79 ± 0.48	50.34 ± 0.31
2	26.23 ± 0.60	24.24 ± 0.41	25.86 ± 0.73	26.81 ± 0.27
$B^+ \to K^+ J/\psi(\to e^+e^-)$ simulation (after full selection)				
0	21.93 ± 0.67	23.46 ± 0.53	23.96 ± 0.29	25.58 ± 0.42
1	51.09 ± 0.76	50.06 ± 0.63	50.38 ± 0.35	50.56 ± 0.44
2	26.97 ± 0.67	26.48 ± 0.55	25.66 ± 0.30	23.85 ± 0.41
Combined normalisation constant				
Data	$(13.0 \pm 0.8) \times 10^{-10}$		$(8.3 \pm 0.4) \times 10^{-10}$	
Simulation	$(12.6 \pm 0.8) \times 10^{-9}$		$(8.7 \pm 0.4) \times 10^{-9}$	
Relative difference				
	$(3.60 \pm 0.31)\%$		$(4.0 \pm 0.4)\%$	

Signal parametrisation

The branching fraction of $B^0_{(s)} \to e^+e^-$ is determined from a fit to the dielectron invariant mass in data, therefore it is important that the expected signal shape is known beforehand. It is retrieved from a fit to $B^0_s \to e^+e^-$ simulation, where a double-sided Crystal Ball function per bremsstrahlung category is used. The same shape, shifted by the nominal mass difference between B^0_s and B^0 mesons, is also used for $B^0 \to e^+e^-$. Due to mismodelling of the detector response, the mass resolution differs between data and simulation. In order get the correct number of signal candidates, and by extension the correct branching fraction, this mismodelling needs to be corrected when fitting the data. While the tail parameters of the Crystal ball functions are fixed to their values obtained from simulation, the mean and

width are corrected, where the correction factor is determined from the dielectron mass distribution in $B^+ \to K^+ J/\psi(\to e^+ e^-)$. These correction factors are defined as

$$C_\mu = \frac{\mu_{\text{data}}}{\mu_{\text{MC}}} \tag{4.19}$$

$$C_\sigma = \frac{\left(\frac{\sigma}{\mu}\right)_{\text{data}}}{\left(\frac{\sigma}{\mu}\right)_{\text{MC}}}. \tag{4.20}$$

In order to quantify the uncertainty associated with this method, it is repeated 1000 times, where new data and simulation samples are obtained via the bootstrapping method [126]. For each of these samples, the fraction of the sample inside of the signal window defined in Table 4.1 is then measured. The width of the resulting efficiency distribution, centred around the nominal 68.3 %, is assigned as systematic uncertainty, which were found be be between 0.6 % and 1.1 %.

Summary of systematic uncertainties

The systematic uncertainties described above are summarised in Table 4.10. The combined systematic uncertainties per Run and bremsstrahlung category are shown in Table 4.11, where systematic uncertainties of the background model and fit range are not included, as they are not assigned to the normalisation directly. Instead, these uncertainties are assigned to the yields of the physical backgrounds in the fit to data instead. Uncertainties on the shape of the background model, described in Section 4.4 are listed here already. They are the dominant source of systematic uncertainty in the analysis. However, as no signal contribution is expected, the result is dominated by statistical uncertainty on the signal yield. The limits calculated in Section 4.5 have been checked to be robust with systematic uncertainties increased to even 20 %.

Table 4.10: Systematic uncertainties of the search for $B^0_{(s)} \to e^+e^-$. Systematic uncertainties are calculated separately per bremsstrahlung category where feasible. The systematic uncertainties on the trigger efficiencies have been determined with the nominal procedure as described in Section 4.3.2. The difference between B^0 and B^0_s only affects the B^0 normalisation.

		Syst. [%]			
Analysis step	Brem.	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 $B^0_{(s)} \to e^+e^-$	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 $B^+ \to K^+ J/\psi$	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 (sWeights)	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	
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 track reconstruction	all		1.1		
nSPDHits difference data - mc Figure 4.13	all		1		
B^0_s-B^0 difference	all	2.5		2.1	

Table 4.11: Combined relative systematic uncertainty on the normalisation constants. The uncertainty from the fit range and background shape are not included, since they affect the background estimation, not the normalisation constants.

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

4.3.7 Checks of absolute efficiency

In order to validate the estimation of efficiencies and uncertainties, the absolute efficiency on the control mode $B^+ \to K^+ J/\psi (\to e^+ e^-)$ is tested in two ways. First, the number of B^+ candidates in the analysed data is calculated from the observed yield in $B^+ \to K^+ J/\psi (\to e^+ e^-)$ after the selection and the determined total efficiency of the selection on $B^+ \to K^+ J/\psi (\to e^+ e^-)$. This value is then compared to the expectation from the $b\bar{b}$ production cross-section and integrated luminosity of the data. The second check is to compute the ratio of branching fractions between $B^+ \to K^+ J/\psi (\to e^+ e^-)$ and $B^0 \to K^+ \pi^-$, which does not rely on the measurements of cross-section and luminosity, thus avoiding the high uncertainties associated with both of these quantities.

Cross-check of $B^+ \to K^+ J/\psi (\to e^+ e^-)$ yields

To check the efficiency calculation, the B^+ yield in the data is compared to the expectation from the $b\bar{b}$ production cross-section and integrated luminosity. The measured $b\bar{b}$ cross sections at the centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ (2011) [127], $\sqrt{s} = 8\,\text{TeV}$ (2012) [128], and $\sqrt{s} = 13\,\text{TeV}$ (2016) [129] are

$$\sigma_{b\bar{b}}(2011) = (288 \pm 4 \pm 48)\,\mu\text{b}, \tag{4.21}$$
$$\sigma_{b\bar{b}}(2012) = (298 \pm 2 \pm 36)\,\mu\text{b}, \tag{4.22}$$
$$\sigma_{b\bar{b}}(2015/2016) = (495 \pm 2 \pm 52)\,\mu\text{b}. \tag{4.23}$$

Together with the hadronisation fraction of b quarks into B^+ mesons [130],

$$f_u = 0.402 \pm 0.007, \tag{4.24}$$

and the integrated luminosities (taken from [131])

$$2011 : 1.11 \, fb^{-1}, \tag{4.25}$$
$$2012 : 2.08 \, fb^{-1}, \tag{4.26}$$
$$2015 : 0.33 \, fb^{-1}, \tag{4.27}$$
$$2016 : 1.67 \, fb^{-1}, \tag{4.28}$$

the number of expected B^+ candidates can be calculated with the formula

$$N_{B^+, \text{ expected}} = \sum_{\text{year}} 2 \cdot \mathcal{L}(\text{year}) \cdot \sigma_{b\bar{b}}(\text{year}) \cdot f_u \tag{4.29}$$

to be

$$N_{11} = (2.28 \pm 0.05 \pm 0.37) \cdot 10^{11}, \tag{4.30}$$
$$N_{12} = (4.78 \pm 0.05 \pm 0.57) \cdot 10^{11}, \tag{4.31}$$
$$N_{15} = (1.13 \pm 0.02 \pm 0.14) \cdot 10^{11} \text{ and} \tag{4.32}$$
$$N_{16} = (6.55 \pm 0.12 \pm 0.71) \cdot 10^{11}. \tag{4.33}$$

The same numbers can be computed separately for each category of bremsstrahlung from the total selection efficiencies given in Table 4.4 and the yields extracted from the final fit to data shown in Figure 4.11 and summarised in Table 4.5 via the relation

$$N_{B^+} = \frac{N_{B^+ \to K^+ J/\psi(\to e^+ e^-), \text{category}}}{\varepsilon_{B^+ \to K^+ J/\psi(\to e^+ e^-), \text{category}} \cdot \mathcal{B}_{B^+ \to K^+ J/\psi(\to e^+ e^-)}}. \tag{4.34}$$

This cross-check showed a strong deviation between the expected and measured B^+ yields, which was found to originate from an incorrect reweighting of the number of hits in the Silicon Pad Detector (nSPDHits). In the preselection a tight requirement on this quantity is applied to ensure coverage with the PID calibration data, since most triggers on the hardware level have the same threshold. The nSPDHits is unfortunately not well modelled in simulation and also not correlated enough to the kinematic reweighting to be fully corrected by it, leading to a wrong efficiency being assigned. To estimate the effect, a sample of $B^+ \to K^+ J/\psi(\to e^+ e^-)$ decays is prepared, where at L0 level only the trigger L0DiMuon TIS is required, since the L0DiMuon trigger has a looser threshold of nSPDHits < 900 compared to other L0 triggers, where thresholds of 600(450) in Run 1(Run 2) are used. On this sample the requirement on nSPDHits in the preselection is also removed. After the application of the nominal reweighting from Section 4.2.3, a one-dimensional reweighting in nSPDHits is performed to calculate a correction factor to the efficiency.

This correction is calculated as the relative difference between the efficiency of the cut on nSPDHits when applied to simulation with and without the additional reweighting. The distributions for 2012 and 2016 entering the reweighting are shown in Figure 4.12.
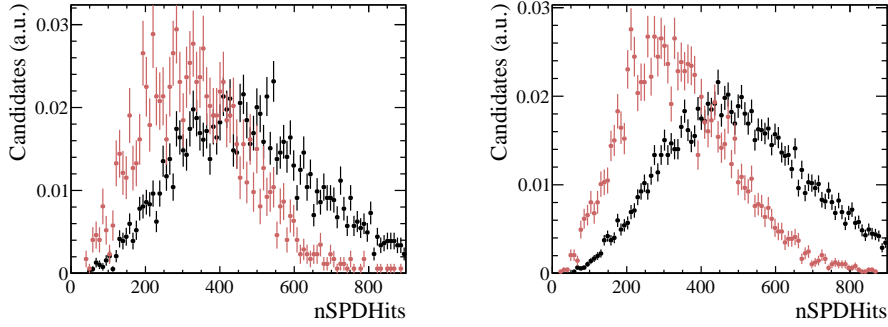
Figure 4.12: Comparison between sweighted data (black) and simulation with the nominal reweighting applied (blue). The difference is reweighted to obtain the correction factor $\Delta\varepsilon_{\mathrm{SPD}}$ on the efficiency of the cut on nSPDHits. 2012 is shown on the left, 2016 on the right.

The correction factors are found to be

$$\Delta\varepsilon_{\mathrm{SPD}}(\mathrm{Run\ 1}) = 0.86 \pm 0.01(\mathrm{stat.}) \pm 0.01(\mathrm{data\text{-}simulation\ overlap}),$$
$$\Delta\varepsilon_{\mathrm{SPD}}(2015) = 0.82 \pm 0.02(\mathrm{stat.}) \pm 0.10(\mathrm{data\text{-}simulation\ overlap}),$$
$$\Delta\varepsilon_{\mathrm{SPD}}(2016) = 0.76 \pm 0.01(\mathrm{stat.}) \pm 0.09(\mathrm{data\text{-}simulation\ overlap})$$

where the first uncertainty is obtained from the sample size and the second is obtained as the fraction of the data that falls in a region not covered by simulation.

Possible effects of the incorrect description of the efficiency of the nSPDHits requirement on the result of the analysis need to be checked as well. The simulated datasets after the full selection are therefore compared after application of the nominal reweighting. This comparison is shown in Figure 4.13.

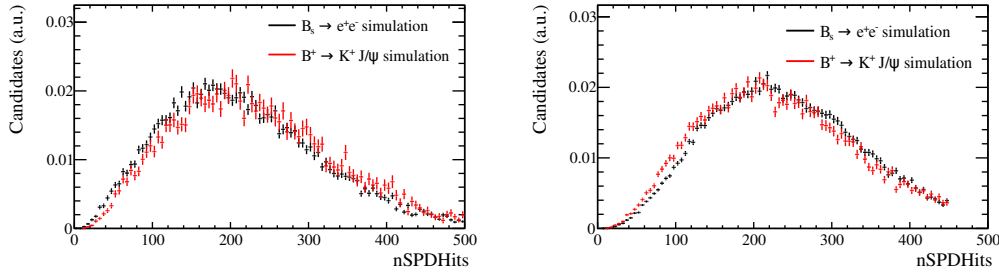**Figure 4.13:** Comparison of the nSPDHits distribution between signal and normalisation mode on simulation after the full selection. The left plot shows the distributions for Run 1 and the right one shows the distributions for Run 2. No large differences are observed.

The distributions of signal and normalisation mode are very similar in nSPDHits. The difference is quantified by applying the tighter requirement from Run 2 (nSPDHits< 450) also on the Run 1 sample, leading to a difference below $1\,\%$, confirming that the incorrect description of nSPDHits cancels in the efficiency ratio. A conservative systematic uncertainty of $1\,\%$ is assigned to the efficiency ratio to cover the small observed difference.

With the correction factors calculated above, the measured and expected B^+ yields can be compared, they are summarized in Table 4.12.

Table 4.12: B^+ candidates as determined from efficiencies and yields found in this analysis compared to the expected B^+ yields from the integrated luminosity and cross sections. The numbers for the bremsstrahlung categories are corrected by their fractions. The uncertainties include statistical and systematic uncertainties.

	B^+ yields (corrected for bremsstrahlung fractions)$[10^{11}]$			
Brem.	2011	2012	2015	2016
0	1.72 ± 0.33	4.3 ± 0.8	0.99 ± 0.17	7.6 ± 1.0
1	1.63 ± 0.30	4.3 ± 0.7	0.91 ± 0.14	7.2 ± 0.9
2	1.8 ± 0.4	4.3 ± 0.8	0.87 ± 0.17	7.1 ± 0.9
Expected	2.28 ± 0.37	4.78 ± 0.57	1.13 ± 0.14	6.55 ± 0.72

The measured and expected yields are in agreement within one standard deviation, they also match well between the bremsstrahlung categories. Since the cross section measurements are dominated by systematic uncertainty, which might well be underestimated, another check is performed against $B^0 \to K^+\pi^-$.

Ratio of $B^+ \to K^+ J\!/\!\psi (\to e^+ e^-)$ and $B^0 \to K^+ \pi^-$ branching fractions

The second cross check of the absolute efficiencies uses the ratio of branching fractions of $B^+ \to K^+ J\!/\!\psi (\to e^+ e^-)$ and $B^0 \to K^+ \pi^-$, computed as

$$\frac{\mathcal{B}(B^0 \to K^+ \pi^-)}{\mathcal{B}(B^+ \to K^+ J\!/\!\psi)} = \frac{N_{B^0 \to K^+ \pi^-}}{N_{B^+ \to K^+ J\!/\!\psi}} \times \frac{\epsilon_{B^+ \to K^+ J\!/\!\psi}}{\epsilon_{B^0 \to K^+ \pi^-}} \times \frac{f_u}{f_d} \qquad (4.35)$$

$$= 0.325 \pm 0.012 \text{ (PDG)}, \qquad (4.36)$$

with the expected value taken from the Particle Data Group [107]. The ratio of hadronisation fractions $\frac{f_u}{f_d}$ is assumed to be unity. The efficiencies and yields for the $B^0 \to K^+ \pi^-$ mode are taken from the $B^0_{(s)} \to \mu^+ \mu^-$ analysis published in 2017 [132], where a similar check was performed.

Since the $B^0_{(s)} \to \mu^+ \mu^-$ analysis used only part of the 2016 data ($\approx 1.1 \text{ fb}^{-1}$), the $B^0 \to K^+ \pi^-$ yield of that year is scaled by the luminosity ratio $1.66/1.1$, as there were no significant differences in data taking conditions between the parts of 2016 data.

The inputs needed for $B^0 \to K^+ \pi^-$ are listed in Table 4.13. Applying the same correction factor as in the comparison of the B^+ yields described before, the measured branching fraction ratios are shown in Table 4.14. They are in good agreement between the years and bremsstrahlung categories and also show good agreement with the expected value from the PDG.

Table 4.13: Yields and efficiencies for the mode $B^0 \to K^+ \pi^-$ [132]. 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 ± 0.07	1.80 ± 0.12	0.86 ± 0.08	4.3 ± 0.4
Efficiency [10^{-3}]	1.500 ± 0.033	1.79 ± 0.04	3.27 ± 0.11	2.97 ± 0.07

Table 4.14: Measured branching fraction ratios split by year and bremsstrahlung category compared to the nominal value taken from the PDG.

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

4.4 Study of physical backgrounds

In order to measure the branching fraction of $B^0_{(s)} \to e^+e^-$ in a fit to the data, the composition of the background remaining after the selection needs to be determined as precisely as possible. The backgrounds contributing to $B^0_{(s)} \to e^+e^-$ can be classified in three different categories:

1. Combinatorial background

2. Background from double misidentification

3. Partially reconstructed background

Combinatorial background means the random combination of tracks passing the reconstruction and selection by accident. This background is dominant after the preselection and is mainly handled by the BDT classifier, which is specifically trained against combinatorial background. Backgrounds from double misidentification, so of the type $B \to hh^{(\prime)}$ with hadrons h, are particularly dangerous as they peak in the signal region, where they can fake a signal peak. These backgrounds are strongly suppressed via the PID requirements applied to the data. Another distinguishing feature from real $B^0_{(s)} \to e^+e^-$ decays is that the final state particles in these backgrounds do not emit meaningful amounts of bremsstrahlung due to their higher mass compared to electrons. Thus, they can only contribute significantly to the case where no bremsstrahlung corrections were applied to the final state particles, distinguishing them from a potential $B^0_{(s)} \to e^+e^-$ signal.

Backgrounds from partial reconstruction come in two sub-groups. The first is of the type $B \to he^+e^-$, where the hadron is not reconstructed. The second follows the structure $B/\Lambda^0_b \to Xe^+\nu_e$, where one particle out of X is misidentified as an electron and the rest is not reconstructed. Both cases have in common that these decays do not show a clear peak in the data and are shifted towards lower masses because of the lost energy from not reconstructing every daughter particle. These

backgrounds can still contribute to the signal region and comprise most of the lower mass sideband after the selection. A table of the studied background processes is given in Table 4.15.

Table 4.15: Summary of the physical background processes studied in the search for $B^0_{(s)} \to e^+ e^-$.

MisID backgrounds	$B^0 \to \pi^+ \pi^-$	$B^0_s \to K^+ K^-$
	$B^0 \to K^+ \pi^-$	$B^0_s \to K^- \pi^+$
Semileptonic backgrounds	$B^0 \to \pi^- e^+ \nu_e$	$B^0_s \to K^- e^+ \nu_e$
	$\Lambda^0_b \to p e^- \overline{\nu}_e$	
Part. reco. backgrounds	$B^0 \to K^0_{\mathrm{S}} e^+ e^-$	$B^0 \to K^{*0} e^+ e^-$
	$B^+ \to K^+ e^+ e^-$	$B^+ \to K^{*+} e^+ e^-$
	$B^+ \to \pi^+ e^+ e^-$	$B^0 \to \pi^0 e^+ e^-$
	$B^+_c \to (J/\psi \to e^+ e^-) e^+ \nu_e$	$B^0_s \to \gamma e^+ e^-$

While combinatorial background can be described by a simple exponential function, as checked on same-sign data (reconstructed pairs of electrons of the same charge $e^\pm e^\pm$), the other background processes need to be studied on simulation, where the same selection and reconstruction as for $B^0_{(s)} \to e^+ e^-$ are applied. The expected contribution to the data can then be evaluated, similarly to how the signal branching fraction is calculated, as

$$N_{\mathrm{bkg}} = N_{B^+ \to K^+ J/\psi} \cdot \frac{\epsilon_{\mathrm{bkg}}}{\epsilon_{B^+ \to K^+ J/\psi}} \cdot \frac{\mathcal{B}(\mathrm{bkg})}{\mathcal{B}(B^+ \to K^+ J/\psi)} \cdot \frac{f_{\mathrm{bkg}}}{f_u}. \qquad (4.37)$$

Here, the total selection efficiencies (including the trigger efficiency) are determined directly from simulation. The only exception are the PID efficiencies, which are determined from calibration data as was done for the signal. The branching fractions of the studied decays are taken from the PDG where measurements exist. For the decays $B^0_s \to K^- e^+ \nu_e$, $B^0 \to \pi^0 e^+ e^-$, $B^+ \to \pi^+ e^+ e^-$ and $B^0_s \to \gamma e^+ e^-$ no measurements were available at the time of publication, so the Standard Model predictions [133–136] are used as input instead.

Since many of the studied backgrounds are expected to contribute significantly to the background after the full selection, the expected yield and shape of the respective backgrounds need to be estimated carefully, since they also need to be modelled in the final fit to the data. Several corrections are applied to the different decay channels, they are described below.

- $B^0 \to \pi^0 e^+ e^-$:
  No simulation was produced for $B^0 \to \pi^0 e^+ e^-$, since the decay, when considered as background for $B^0_{(s)} \to e^+ e^-$, is expected to have the same properties as the $B^+ \to \pi^+ e^+ e^-$ decay, where simulation was available. Since neither $B^0 \to \pi^0 e^+ e^-$ nor $B^+ \to \pi^+ e^+ e^-$ have been observed yet, the SM prediction for the branching fraction, taken from [135], is used for the estimation. The contribution of these decays to the background is found to be small.

- $B^+ \to \pi^+ e^+ e^-$ and $B^+_c \to J/\psi (\to e^+ e^-) e^+ \nu_e$:
  These decay modes were simulated with a phase-space model, instead of physical form-factors. In order to correctly model the shape and account in the selection efficiencies for potential kinematic differences, they are reweighted to physical models. $B^+ \to \pi^+ e^+ e^-$ simulation is reweighted to a model by Ball and Zwicky [137], while $B^+_c \to J/\psi (\to e^+ e^-) e^+ \nu_e$ is reweighted to a model by Kiselev [138].

- Partially reconstructed backgrounds:
  In all studied cases where at least one particle is not reconstructed, the simulations were produced requiring all produced final-state particles to be within the LHCb acceptance. Since this is not necessary to fake $B^0_{(s)} \to e^+ e^-$, where only two electrons are needed, the efficiency of this generator step needs to be corrected. This is studied from large samples of generator-level simulation generated in full 4π solid angle acceptance using RapidSim [139]. RapidSim is a fast-simulation tool, which allows the generation of large simulated samples without running the full detector geometry and reconstruction, thus saving significant time and disk-space over full simulation, while simultaneously allowing to place less constraints on the simulated decay. Here the relative difference between requiring all particles within the LHCb acceptance and only requiring the particles that fake $B^0_{(s)} \to e^+ e^-$ to be in the acceptance can then be determined and the efficiency of the nominal simulation be corrected. These correction factors were found to show a strong dependency on the decay topology, giving corrections between $16\,\%$ and $36\,\%$.

- $B^0 \to \pi^- e^+ \nu_e$ and $B^0_s \to K^- e^+ \nu_e$:
  While these decays are shifted towards lower masses from loosing energy to the neutrino, they can still contribute to the background by misidentification of either the π^- or K^- as an electron combined with relatively high branching fractions. These decays were simulated using the Isgur-Weiss form factor model [140], which is known to be inaccurate. Therefore, the simulation of $B^0 \to \pi^- e^+ \nu_e$ is corrected to recent calculation from Light Cone Sum rules [141], by reweighting the momentum transfer, defined as

$$q^2 = (p^\mu_{B^0} - p^\mu_{\pi^-})(p_{B^0,\mu} - p_{\pi^-,\mu}), \qquad (4.38)$$

69

where the p^μ_X denote the 4-momenta of the B^0 or π^- respectively.

This reweighting is found to have negligible impact on the estimation of $B^0 \to \pi^- e^+ \nu_e$ and is therefore dropped for the estimation of $B^0_s \to K^- e^+ \nu_e$, which is expected to have a significantly smaller contribution due to K-e misidentification being suppressed more strongly than π-e misidentification. Furthermore, no measurement exists for $B^0_s \to K^- e^+ \nu_e$, so the SM prediction from [135] is used.

- Semileptonic cascades containing D mesons:
  The class of backgrounds of the type $B \to D(\to Xe\nu)e\nu$, where X denotes at least one hadron, appear dominantly towards the low end of the studied mass range. While they are shifted towards these lower masses from the loss of energy from not reconstructing the neutrinos and hadrons, they also have large branching fractions, leading to large contributions in the studied mass range. They are studied using $B^+_c \to J/\psi (\to e^+ e^-) e^+ \nu_e$ as an efficiency proxy, because no simulated samples of sufficient size to measure any efficiencies exist. For the same reason, their shape is determined from a RapidSim sample of $B^+ \to D^0 (\to \pi e \nu) e \nu$, which is the most conservative choice as it looses the least amount of energy to missing particles. Since these decays are abundant at low masses but have almost no contribution in the signal region, they can affect the measurement of the $B^0_{(s)} \to e^+ e^-$ branching fraction only through influencing the description of the background shape in the sidebands.

The amount of expected background candidates for each studied decay in the $B^0_s \to e^+ e^-$ and $B^0 \to e^+ e^-$ signal windows are shown in Table A.7 for Run 1 and in Table A.8 for Run 2. The largest contributions to the signal window are found to come from the decays $B^0 \to \pi^- e^+ \nu_e$, $B^+_c \to J/\psi (\to e^+ e^-) e^+ \nu_e$ and $\Lambda^0_b \to p e^- \overline{\nu}_e$. The backgrounds in the signal regions sum up to

$$N_{\text{excl, } B^0_s \to e^+ e^-, \text{ Run 1}} = 24.91 \pm 3.97 (stat.) \pm 2.72 (\mathcal{B}), \tag{4.39}$$

$$N_{\text{excl, } B^0_s \to e^+ e^-, \text{ Run 2}} = 17.61 \pm 3.15 (stat.) \pm 2.96 (\mathcal{B}), \tag{4.40}$$

$$N_{\text{excl, } B^0 \to e^+ e^-, \text{ Run 1}} = 46.12 \pm 8.11 (stat.) \pm 6.49 (\mathcal{B}), \tag{4.41}$$

$$N_{\text{excl, } B^0 \to e^+ e^-, \text{ Run 2}} = 34.12 \pm 7.69 (stat.) \pm 7.07 (\mathcal{B}), \tag{4.42}$$

where the first uncertainty is statistical and the second comes from the uncertainty on the branching fractions of the backgrounds. The background contribution to $B^0 \to e^+ e^-$ is considerably higher than for $B^0_s \to e^+ e^-$ due to the lower B^0 mass, increasing the contribution from especially partially reconstructed backgrounds.

The description of the dielectron invariant mass distribution is done independently per category of bremsstrahlung. One advantage of this is that backgrounds might differ in relative fractions of the bremsstrahlung categories from $B^0_{(s)} \to e^+ e^-$. This

is especially useful for backgrounds of the type $B \to hh^{(\prime)}$, which only occur when no bremsstrahlung corrections were applied, as they only emit negligible amounts of bremsstrahlung. In case these backgrounds were to differ significantly from their expectation, this would allow to differentiate these backgrounds from a $B^0_{(s)} \to e^+e^-$ signal. The shapes of the backgrounds are described using gaussian kernels which are fitted to simulation and fixed in the fit to data. For combinatorial background as well as $B^+_c \to J/\psi(\to e^+e^-)e^+\nu_e$, the shape is taken from same-sign data. This data is reconstructed as $B^0_{(s)} \to e^\pm e^\pm$ and is expected to contain only random combinations of tracks passing the reconstruction criteria, which is the same as combinatorial background in opposite-sign data. The decay $B^+_c \to J/\psi(\to e^+e^-)e^+\nu_e$ is expected to have the same contribution to same-sign data as it has to opposite-sign data, since either the electron or the positron from the J/ψ can be combined with the positron from the B^+_c. It is thus included in the shape extracted from same-sign data instead of having an individual component added to the final fit.

In the fit to data, the yields of the various background components are constrained to their values from Tables A.7 and A.8 multiplied with the bremsstrahlung fractions taken from simulation. To improve the stability of the fit, backgrounds that are expected to have less than 0.1 candidates remaining in the whole mass range are neglected in the fit. For the semileptonic D cascades, no information about the bremsstrahlung fractions is available, as no reconstruction takes place in RapidSim samples, so the same shape is used for all categories and the relative bremsstrahlung fractions from $B^0_s \to e^+e^-$ simulation are used. With these inputs, the full background PDF is given as

$$
\begin{aligned}
\mathrm{PDF}_{\mathrm{Full}}(m, N_{SS}, N_i) = {} & \mathcal{P}(N_{\mathrm{meas}}|(N_{SS} + \sum_i N_i)) \times \frac{1}{N_{SS} + \sum_i N_i} \\
& \times \Big(N_{SS}\mathcal{G}(N_{SS,\mathrm{meas}}, \sigma_{N_{SS},\mathrm{meas}}|N_{SS})\mathrm{PDF}_{SS}(m) \\
& + \sum_i N_i \mathcal{G}(N_{i,\mathrm{meas}}, \sigma_{N_i,\mathrm{meas}}|N_i)\mathrm{PDF}_{i,\mathrm{bkg}}(m) \Big),
\end{aligned}
\tag{4.43}
$$

where the exponential function fitted to same-sign data, describing combinatorial background and $B^+_c \to J/\psi(\to e^+e^-)e^+\nu_e$, is denoted as PDF_{SS}, the various background PDFs are denoted as $\mathrm{PDF}_{i,\,\mathrm{bkg}}$ and the yields are denoted as N_i, with the summation being performed over all exclusive backgrounds. The yields are accounted for by a Poissonian term $\mathcal{P}$, while constraints of the yields are included as a Gaussian $\mathcal{G}$, where the width is given as the quadrature of the statistical and systematic uncertainties from Tables A.7 and A.8. The result of the fit to the mass sidebands is shown in Figure 4.14.

Overall, good agreement of the fit with the data is observed. The composition of the background varies with the category of bremsstrahlung, as the relative fractions
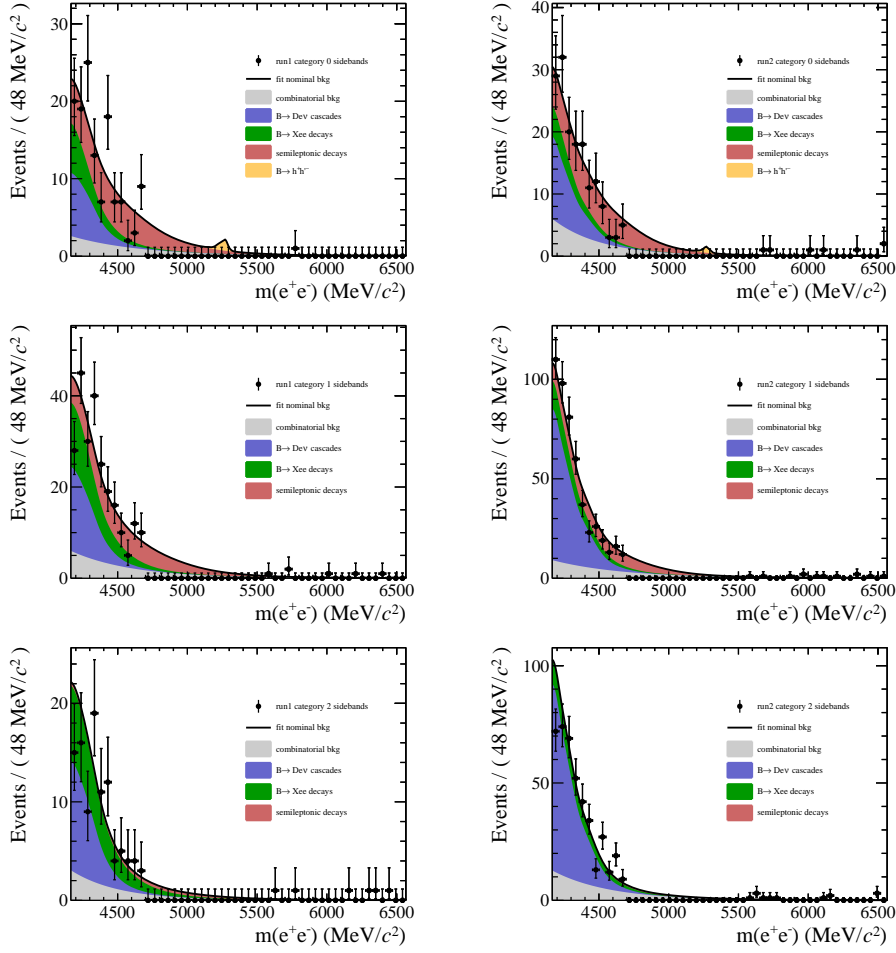
Figure 4.14: Fits of the full background model to the mass sidebands. From top to bottom the bremsstrahlung categories where none, either or both electrons corrections applied are shown, while the left column corresponds to Run 1 data and the right column to Run 2 data. The small peak in the signal region corresponds to the misidentified $B \to hh^{(\prime)}$ decays, which emit almost no bremsstrahlung and therefore only show up in the case where no bremsstrahlung corrections were applied. The relative fraction of backgrounds between bremsstrahlung categories follows the expectation from simulation.

of the backgrounds differ. The background composition also differs between Run 1 and Run 2 due to changes in the trigger system, the different BDT selection and the difference in tune of the PID variables.

It is also studied, whether differences between data and simulation in the shape of the backgrounds need to be accounted for. The fit is therefore repeated while the shape of each background is convoluted with a gaussian with a mean of 0 and variable width, which is allowed to fluctuate between 0 and the correction to the width of $B^0_{(s)} \to e^+e^-$ simulation determined in Section 4.3.6. Since the difference to the approach without including the resolution correction is found to be negligible, it is dropped from the fit.

Another cross-check performed concerns the uncertainty of the background estimates. The impact of these uncertainties on the background description is studied by fixing the yields of each background component to their nominal value reduced and increased by one standard deviation and repeating the fit to the mass sidebands. The maximum variation of the expected background candidates in the signal region is assigned as a systematic uncertainty to the background description. It is the major source of systematic uncertainty in the analysis with up to 12 %.

4.5 Results of the Search for $B^0_{(s)} \to e^+e^-$

The branching fractions for $B^0_{(s)} \to e^+e^-$ are obtained from an extended maximum likelihood fit to the invariant dielectron mass in the fully selected data. Six subsamples of all available data, defined by bremsstrahlung category and run period, are fitted simultaneously, where the only parameter shared between subsamples is the branching fraction. The normalisation constants are Gaussian constrained to their nominal values in the fit. The width of the gaussian is given as their statistic and systematic uncertainties added in quadrature.

Since the mass resolution is much larger than the mass difference between B^0_s and B^0 mesons (see Figure 4.15 for an illustration), possible contributions from $B^0_s \to e^+e^-$ and $B^0 \to e^+e^-$ can not be separated reliably. Thus the fit is performed assuming either only $B^0_s \to e^+e^-$ or $B^0 \to e^+e^-$ contributing, with the other being neglected in the fit.
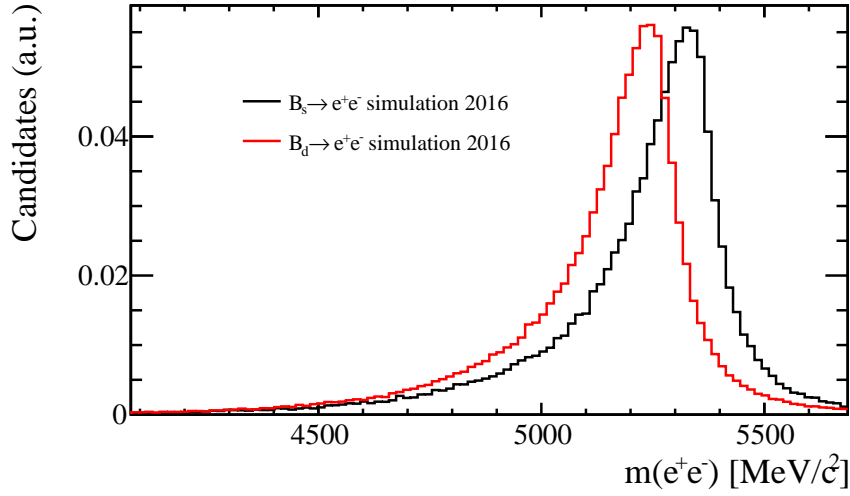
Figure 4.15: Comparison of the mass shapes of $B^0_s \to e^+ e^-$ and $B^0 \to e^+ e^-$ in 2016 simulation.

Before performing the simultaneous fit, a last check is performed. The fit is done separately per bremsstrahlung category and the ratio of the signal yields is compared to the ratio of normalisation constants, without inspecting the yields themselves. This serves as a check for the modelling of the backgrounds that differ strongly between bremsstrahlung categories (like double misidentification only occurring in category 0) and thus might influence these partial results differently. Consistency between the Run 1 and Run 2 measurements is also tested in the same way. Since no deviations more significant than 2.5σ were observed, where this threshold was defined before performing the checks, the final simultaneous fit to the unblinded data was performed. The resulting fit to the unblinded data for the $B^0_s \to e^+ e^-$ hypothesis is shown in Figure 4.16, the measured branching fractions are summarized in Table 4.16.
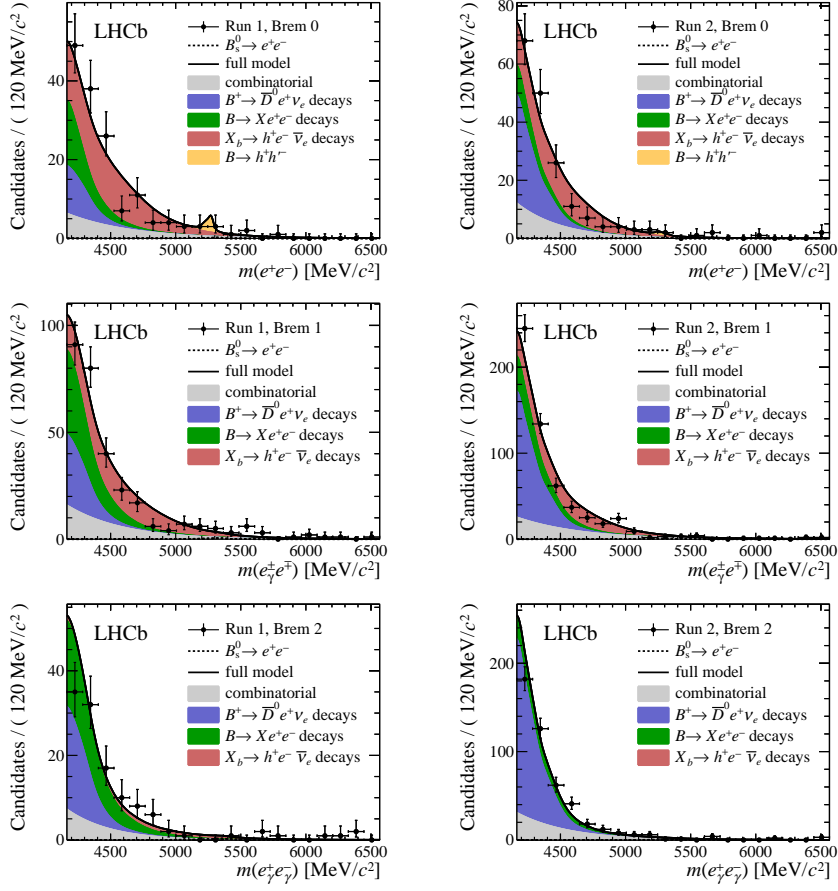
Figure 4.16: Fit to the unblinded dielectron invariant mass distribution with the $B^0_s \to e^+ e^-$ hypothesis. The data is shown in black, the complete fitted signal plus background PDF is shown as a black line. The background components are stacked in grey for combinatorial background, blue for semileptonic D cascades, green for partially reconstructed $B \to X e^+ e^-$ decays, red for semileptonic decays and yellow for $B \to h^+ h^-$ decays. Run 1 is shown on the top, Run 2 on the bottom. For both runs bremsstrahlung category 0 is shown on the left, category 1 in the middle and category 2 on the right.

Table 4.16: Measured branching fraction for $B^0_s \to e^+ e^-$ and $B^0 \to e^+ e^-$ from the fit to the unblinded data.

Decay	$\mathcal{B}$
$B^0_s \to e^+ e^-$	$(2.4 \pm 4.4) \cdot 10^{-9}$
$B^0 \to e^+ e^-$	$(0.30 \pm 1.29) \cdot 10^{-9}$

Since no significant signal is observed, upper limits on the branching fractions of $B^0_{(s)} \to e^+ e^-$ are set using the CL_s method [142], where all calculations are done in the GammaCombo framework [143]. For the limit, a one-sided test-statistic given as

$$
q = \begin{cases}
-2 \log \left(\frac{\mathcal{L}(\mu, \hat{\hat{\theta}} | x)}{\mathcal{L}(\hat{\mu}, \hat{\theta} | x)} \right) & \mu \geq \hat{\mu}, \\
0 & \mu < \hat{\mu}
\end{cases} \tag{4.44}
$$

is used. Here, $\mathcal{L}$ is the likelihood, $\hat{\mu}$ is the measured value (in this case the branching fraction $\mathcal{B}$) and x is the observable (dielectron invariant mass).

The notation $\hat{\theta}$ means that the nuisance parameters θ are fitted together with the parameter of interest μ, while $\hat{\hat{\theta}}$ describes that the nuisance parameters are profiled over. In the Wald approximation [144], the p-value can be derived from the test-statistics as [145]

$$
p = \int_{q_{obs}}^{\infty} \Phi \left(\frac{\mu' - \mu}{\sigma} \right) \delta(q) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q}} \exp \left[-\frac{1}{2} \left(\sqrt{q} - \frac{\mu - \mu'}{\sigma} \right)^2 \right], \tag{4.45}
$$

with the cumulative distribution of the standard Gaussian Φ, the tested hypothesis (in this case $\mathcal{B}(B^0_{(s)} \to e^+ e^-)$) μ and the potentially different hypothesis μ' the data distribution might follow. σ is the standard deviation of the measured value.

For the CL_s method, the p values of the signal+background hypothesis are normalised to the background only p value to take into account the compatibility of the test with the background only hypothesis, leading to

$$
CL_s = CL_{s+b}/CL_b. \tag{4.46}
$$

The CL_{s+b} value are obtained as the p-values assuming the data is distributed according to the tested hypothesis, so $\mu = \mu'$ in Equation (4.45). To determine the CL_b values, the data are assumed to follow the background-only hypothesis, i.e. $\mu' = 0$ (while μ is the hypothesis/branching fraction scanned over).

In order to avoid reliance on the Wald approximation, the test statistics for signal+background and background only hypotheses are sampled by generating pseudo-experiments, where nuisance parameters are plugged in at their best fit value at each given scan point (plugin method). The resulting distributions of CL_s for $B^0_s \to e^+ e^-$ and $B^0 \to e^+ e^-$ are shown in Figure 4.17, the resulting limits are given in Table 4.17.
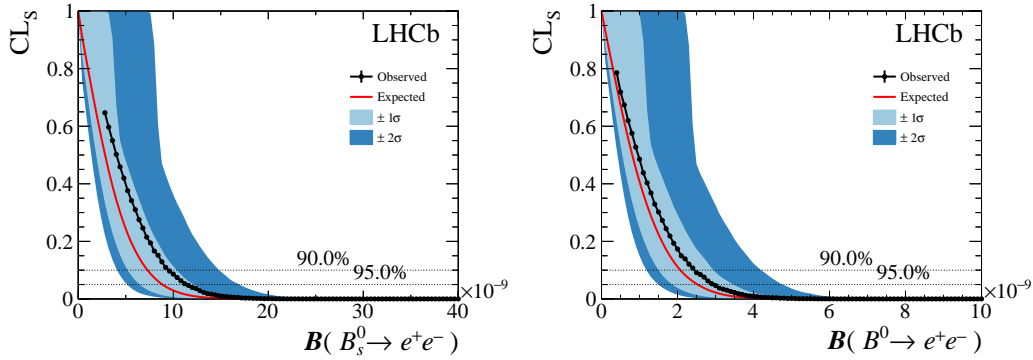
76

Figure 4.17: Distribution of CL_s for $B^0_s \to e^+ e^-$ (on the left) and $B^0 \to e^+ e^-$ (on the right) calculated using the GammaCombo framework. The observed values are plotted beginning at the measured branching fraction as the test statistic has non-zero values only for $\mathcal{B} > \mathcal{B}_{\text{meas}}$.

Table 4.17: Calculated observed upper limits on the branching fractions for $B^0_s \to e^+ e^-$ and $B^0_s \to e^+ e^-$ at a confidence level of 90 % (95 %).

	observed limit
$\mathcal{B}(B^0_s \to e^+ e^-)$	$9.4(11.2) \cdot 10^{-9}$
$\mathcal{B}(B^0 \to e^+ e^-)$	$2.5(3.0) \cdot 10^{-9}$

Since there are two mass eigenstates in the B^0_s system with a sizeable mass difference, the lifetime of the B^0_s meson depends on the CP parameter $\mathcal{A}_{CP}$. While in the SM only the heavy mass eigenstate contributes to $B^0_s \to e^+ e^-$, meaning $\mathcal{A}_{CP} = +1$, in theories beyond the SM this does not need to be true and all other values up to $\mathcal{A}_{CP} = -1$ can be allowed. Corrections to the nominal result are therefore calculated for values of $\mathcal{A}_{CP} = [-1; 0; +1]$, where 0 corresponds to the mean lifetime of the B^0_s meson the simulation was generated with. It is calculated to check that the approach to compute the corrections is not biased.

First, the time-dependent acceptance is extracted from simulation by dividing the decay time distribution of the B^0_s meson by the theoretical model, an exponential function where the parameter is given as the B^0_s lifetime used in the generation of the simulation. Here, a value of $\tau(B^0_s) = 1.512\,\text{ps}$ was used. The resulting shape is

then used to get the correction factor defined as

$$\delta = \frac{\epsilon^{\mathcal{A}_{CP}}}{\epsilon^{MC}} \tag{4.47}$$

$$= \frac{\int_0^\infty (R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t})\epsilon(t)dt}{\int_0^\infty (R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t})dt} \times \frac{\int_0^\infty e^{-\Gamma_{MC}t}dt}{\int_0^\infty e^{-\Gamma_{MC}t}\epsilon(t)dt}, \tag{4.48}$$

where $\Gamma_{H/L}$ are the widths of the heavy/light mass eigenstate and $R_{H/L}$ are the fractions of the contribution by the light and heavy mass eigenstate, defining $\mathcal{A}_{CP}$ as

$$\mathcal{A}_{CP} = \frac{R_H - R_L}{R_H + R_L}. \tag{4.49}$$

This procedure leads to corrections of $\mp 2.4\,\%$, corresponding to $\mathcal{A}_{CP} = \pm 1$ with respect to the nominal result in Table 4.17, which corresponds to an even mixture of the two eigenstates, i.e. $\mathcal{A}_{CP} = 0$. As the difference between the mass eigenstates in the B^0 system is much smaller than in the B^0_s system, this effect is neglected for $B^0 \to e^+ e^-$.

The limits found in this analysis are about 30 times smaller than the previous limits by the CDF collaboration, reaching the region predicted by universal New Physics scenarios in the scalar and pseudoscalar sector for electrons as illustrated in Figure 4.18. In particular new pseudoscalar operators in $b \to se^+ e^-$ transitions [146] are severely constrained by the result.
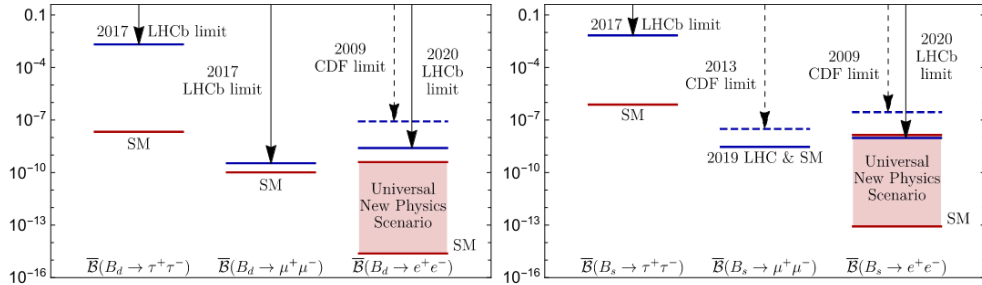

Figure 4.18: Comparison of the measured limits on the $B^0 \to \ell^+ \ell^-$ (left) and $B^0_s \to \ell^+ \ell^-$ (right) branching fractions with the SM prediction and (pseudo-)scalar New Physics scenarios [58]. The new measurement in the B^0_s mode reaches the allowed space by the New Physics scenario compared to.

5 Search for the lepton-flavour violating decays $B^+ \to K^+ e^\pm \mu^\mp$

This chapter contains the search for the lepton-flavour violating decays $B^+ \to K^+ e^\pm \mu^\mp$. The analysis follows a similar structure as the search for $B^0_{(s)} \to e^+ e^-$ presented in Chapter 4, with additional vetos against physical backgrounds and a two-step MVA strategy. At the time of writing, the analysis is not yet published. Thus the following chapter represents the current status of the search for $B^+ \to K^+ e^\pm \mu^\mp$ decays with the full LHCb dataset.

The analysis has been performed in cooperation with a postdoctoral researcher from TU Dortmund and a senior researcher from LPNHE, Paris. It makes use of two bachelor's theses, which focused on the expected sensitivity that can be achieved using the full Run 2 data [147] and improvements to the BDT strategy [148]. The focus of my work were the construction of the selection, efficiency calibration, studies of systematic uncertainties and studies on the strategy for the limit setting.

5.1 Analysis strategy

The aim of this search is to update the previous analysis performed by LHCb in this decay mode on 2011 and 2012 data [66]. Improvements of the sensitivity can be achieved two ways. The first and obvious one is the inclusion of all data collected by LHCb, meaning that in addition to 2011 and 2012, also data from the Run 2 period, meaning 2015 - 2018, is included. The second is a revised analysis strategy, for example including up-to-date isolation variables in the multivariate classifier and performing the multivariate analysis split by charge configuration. Similarly to the search for $B^0_{(s)} \to e^+ e^-$, the analysis is performed in a blinded fashion.

The region in the $Ke\mu$ invariant mass that is excluded from the data is taken directly from the previous analysis round. It is summarised in Table 5.1. Compared to the search for $B^0_{(s)} \to e^+ e^-$, the excluded region is much smaller, since there is only one electron in the final state, leading to less effect of the bremsstrahlung corrections.

Table 5.1: Regions of interest in the B^+ mass in the search for $B^+ \to K^+ e^\pm \mu^\mp$.

	mass region in MeV/c^2
Blinded region	[4985, 5385]
Upper mass sideband	[3780, 4985]
Lower mass sideband	[5385, 6780]

Since the backgrounds present in the data used for the search for $B^+ \to K^+ e^\pm \mu^\mp$ decays are expected to differ significantly depending on whether the K and μ have the same charge (SS or same-sign, $B^+ \to K^+ e^- \mu^+$) or opposite charge (OS or opposite-sign, $B^+ \to K^+ e^+ \mu^-$), the analysis is carried out separately for the two charge configurations. While the preselection and trigger (see Section 5.2.1) do not differ by charge configuration, the vetos applied to the data to remove physical backgrounds from the data differ between the charge configurations. The multivariate classifiers are trained and optimised separately for the two charge configurations as well. Limits on the branching fractions are in the end set on both charge configurations separately, as the underlying physics might be different in models beyond the SM. In addition to this split in charge configuration, the analysis is split by run period as well. Efficiencies are determined per year and then combined in their respective run period.

The branching fraction of $B^+ \to K^+ e^\pm \mu^\mp$ is measured relative to $B^+ \to K^+ J/\psi$. Instead of the J/ψ decaying further into a electron pair, as was the case for $B^0_{(s)} \to e^+ e^-$, here the case where the J/ψ decays into two muons is used, leading to the normalisation mode $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$. The decay mode $B^+ \to K^+ J/\psi(\to e^+ e^-)$ is also used however, mainly for checks concerning electron efficiencies and for corrections to the mass resolution of $B^+ \to K^+ e^\pm \mu^\mp$ in the fit to obtain a limit. The branching fraction is then given as

$$
\begin{aligned}
\mathcal{B}(B^+ \to K^+ e^\pm \mu^\mp) &= \frac{N(B^+ \to K^+ e^\pm \mu^\mp)}{\epsilon(B^+ \to K^+ e^\pm \mu^\mp)} \frac{\mathcal{B}(B^+ \to K^+ J/\psi) \cdot \epsilon(B^+ \to K^+ J/\psi)}{N(B^+ \to K^+ J/\psi)} \\
&= N(B^+ \to K^+ e^\pm \mu^\mp) \cdot \alpha' \cdot \mathcal{B}(B^+ \to K^+ J/\psi) \qquad (5.1) \\
&= N(B^+ \to K^+ e^\pm \mu^\mp) \cdot \alpha
\end{aligned}
$$

with the signal and normalisation yields after the selection $N(B^+ \to K^+ e^\pm \mu^\mp)$ and $N(B^+ \to K^+ J/\psi)$, respectively, as well as their associated selection efficiencies $\epsilon(B^+ \to K^+ e^\pm \mu^\mp)$ and $\epsilon(B^+ \to K^+ J/\psi)$.

Since both $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ are decays of a B^+ meson, the hadronisation fractions taken as input in the search for $B^0_{(s)} \to e^+ e^-$ cancel in the ratio. However, the branching fraction of the control mode $\mathcal{B}(B^+ \to K^+ J/\psi)$ is still

needed as an external input. In Equation (5.1) $B^+ \rightarrow K^+ e^\pm \mu^\mp$ is used a a stand-in for either $B^+ \rightarrow K^+ e^- \mu^+$ or $B^+ \rightarrow K^+ e^+ \mu^-$, as the measurement is performed split by charge.

The selection is described in detail in Section 5.2. The preselection described in Section 5.2.1 has the same structure as in $B^0_{(s)} \rightarrow e^+ e^-$, with the addition of dedicated vetos against semileptonic D meson cascades, which here can be vetoed specifically due to the presence of the K track, as well as vetos against decays involving charmonium resonances of the structure $B^+ \rightarrow K^+ c\bar{c} \, (\rightarrow \ell^+ \ell^-_{\text{ misid}})$. Background is then subtracted using sWeights as described in Section 5.2.2 and a kinematic reweighting is performed in Section 5.2.3. The MVA strategy consisting of two classifiers trained against different types of background is described in Section 5.2.4. Afterwards, stringent requirements on particle identification variables are applied in Section 5.2.5. The results of the search are presented in Section 5.4.

5.2 Selection

The following section describes the selection applied to $B^+ \rightarrow K^+ e^\pm \mu^\mp$ data and simulation to separate signal from background candidates. Unless specified otherwise, the same selections are applied to the $B^+ \rightarrow K^+ e^\pm \mu^\mp$, $B^+ \rightarrow K^+ J/\psi(\rightarrow \mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi(\rightarrow e^+ e^-)$ decay modes. The applied preselection (see Section 5.2.1) is split into four parts: the experiment-wide preselection, or stripping, requirements on the LHCb trigger system, fiducial cuts to align data and simulation with calibration data and vetos against specific physical backgrounds. Afterwards, remaining background in the control mode can be subtracted (see Section 5.2.2) in order to allow comparisons of $B^+ \rightarrow K^+ J/\psi(\rightarrow \mu^+ \mu^-)$ signal candidates in data and simulation and correct for observed differences (see Section 5.2.3). With these corrections applied, a multivariate analysis is performed in two steps (see Section 5.2.4) and requirements on particle identification are applied to remove remaining background from misidentification (see Section 5.2.5).

5.2.1 Preselection

In order to ensure coverage with the used calibration data while maintaining high signal efficiency, a loose preselection is applied to $B^+ \rightarrow K^+ e^\pm \mu^\mp$ data. It consists of a experiment-wide preselection, requirements on the LHCb trigger system, fiducial requirements on the daughter particles to ensure coverage with calibration data as well as vetos against physical backgrounds. These vetos help to reduce contributions

from semileptonic D meson cascades and decays of charmonia where one lepton is misidentified, creating an artificial $e\mu$ pair.

Stripping

The candidates for the search for $B^+ \to K^+ e^\pm \mu^\mp$ are formed during the experiment-wide preselection, where also a first set of selection requirements is performed, optimised for studies of $B^+ \to K^+ \ell^+ \ell^-$ decays. The selections applied to $B^+ \to K^+ e^\pm \mu^\mp$, $B^+ \to K^+ J/\psi(\to e^+ e^-)$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ are aligned on a per-particle basis, so differences between the decay modes only arise from the difference in final state. The applied selections are summarised in Table A.9.

Identical selections are applied to data and simulated samples, with the exception of the requirements on DLL variables. These requirements are removed from the selection on simulation and instead applied at a later stage to calculate this requirements efficiency from calibration data. For a detailed explanation of the used variables see Section 4.2.1.

Choice of triggers

The choice of triggers follows a similar logic in $B^+ \to K^+ e^\pm \mu^\mp$ as described in Section 4.2.1. Only TOS triggers are considered, leading to the trigger selection summarised in Table 5.2.

Table 5.2: Trigger selection applied to $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$. Triggers on the same trigger level are combined with a logical OR, the three trigger levels are combined with a logical AND. Only TOS (trigger on signal) is used. If a particle is present multiple times in the final state the stated triggers are applied to all occurrences.

Level	Run 1	Run 2
L0	μ L0Muon	μ L0Muon
HLT1	B TrackAllL0	B TrackMVA
	B TrackMuon	B TwoTrackMVA
HLT2	B Topo2BodyBBDT	B Topo2Body
	B Topo3BodyBBDT	B Topo3Body
	B TopoMu2BodyBBDT	B TopoMu2Body
	B TopoMu3BodyBBDT	B TopoMu3Body

On the hardware level L0, the three final state particles directly correlate to possible trigger strategies: trigger on the μ, e, K or a combination thereof. Since the efficiency of the muon trigger system is very high, it is chosen as baseline. Adding either electron or kaon triggers only increases the efficiency by a negligible amount. As estimating these triggers from data complicates the procedure, inflating uncertainties beyond the potential gain from adding them to the selection, they are not considered further.

The HLT selection is again based on muon triggers due to their high efficiency. Also here the inclusion of electron triggers was found to yield no significant increase in efficiency, similar as on the hardware level L0. Also, as $B^+ \to K^+ e^{\pm} \mu^{\mp}$ is a three body decay, the corresponding topological triggers are used in addition to the two-body triggers.

The triggers shown in Table 5.2 are applied to the signal mode $B^+ \to K^+ e^{\pm} \mu^{\mp}$ and the normalisation mode $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$, but are obviously not suitable to trigger $B^+ \to K^+ J/\psi (\to e^+ e^-)$. Thus, the following substitutions are made: On the hardware level, only L0Electron is required from both electrons, while on HLT2 the topological trigger lines with electron PID requirements are used instead of their muon counterparts, leading to the selection shown in Table 5.3. The L0Global TIS trigger, which was used in the selection applied to $B^+ \to K^+ J/\psi (\to e^+ e^-)$ during the search for $B^0_{(s)} \to e^+ e^-$, is not considered here. While an increase in the $B^+ \to K^+ J/\psi (\to e^+ e^-)$ efficiency could be achieved by including it, this would lead to a more significant difference compared to $B^+ \to K^+ e^{\pm} \mu^{\mp}$. However, $B^+ \to K^+ J/\psi (\to e^+ e^-)$ is not used in the normalisation and thus does not affect the sensitivity of the analysis. Thus similarity to $B^+ \to K^+ e^{\pm} \mu^{\mp}$ is more important than the efficiency of the selection applied to this decay, in order to perform cross-checks of electron efficiencies on the decay channel.

Table 5.3: Trigger selection applied to $B^+ \to K^+ J/\psi (\to e^+ e^-)$. Triggers on the same trigger level are combined with a logical OR, the three trigger levels are combined with a logical AND. Only TOS (trigger on signal) is used. If a particle is present multiple times in the final state the stated L0 triggers are applied to all occurrences.

Level	Run 1	Run 2
L0	*e* L0Electron	*e* L0Electron
HLT1	B TrackAllL0	B TrackMVA
	B TrackMuon	B TwoTrackMVA
HLT2	B Topo2BodyBBDT	B Topo2Body
	B Topo3BodyBBDT	B Topo3Body
	B TopoE2BodyBBDT	B TopoE2Body
	B TopoE3BodyBBDT	B TopoE3Body

Fiducial requirements

As a first step to reduce the amount of combinatorial background in the data, as well as ensure identical coverage with the samples used for PID calibration, a cut-based preselection is performed. Tightening the mass window around the B^+ and requiring the mass of the B^+ calculated with a constraint on the J/ψ mass above a certain threshold, in order to remove partially reconstructed decays was investigated. However, this biases the correct description of the mass tails and made it impossible to describe partially reconstructed background leaking into the tightened mass window after the selection. Thus, these cuts were discarded, as the dimuon final state is already very clean without these requirements. They are however listed in Table A.10 for completeness, marked with *. In order to remove candidates from clone tracks, a requirement on the angle between two final state tracks is applied for each pair of final state tracks. The specific version of the ProbNNk variable selected on depends on the run period. In Run 1 MC12TuneV2 is used, while in Run 2 MC15TuneV1 is used.

Vetos

In order to remove physical backgrounds from the data, vetos against certain classes of backgrounds are performed. The first considered class of background come from decays of the charmonium resonances J/ψ and $\psi(2S)$ into two leptons, where at least one final state particle is misidentified. Their contribution is removed by

applying mass vetos around the nominal J/ψ and $\psi(2S)$ masses, as summarised in Table A.11, where a shift from assigning the wrong mass hypothesis is taken into account for either swaps of the kaon with a leptons or the misidentification of one lepton. These vetos are only applied to $B^+ \to K^+ e^\pm \mu^\mp$ and not the control modes $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ and $B^+ \to K^+ J/\psi(\to e^+ e^-)$.

The other class of backgrounds consists of semileptonic decays of the type $B \to D\ell\nu$, where the D either decays semileptonically (with a different lepton flavour than the B meson it originated from) or decays to hadrons, where it might only get partially reconstructed and a hadron misidentified as a lepton. The three-body final state in $B^+ \to K^+ e^\pm \mu^\mp$ provides enough information to veto these decays by requiring the invariant mass of either kaon-lepton pair to be above the D^0 mass in the case of the pair having opposite charge from one another. The requirements are summarised in Table A.11. The contribution of the semileptonic cascades in the plane spanned by the Ke and $K\mu$ invariant masses are visualised in Figure 5.1. The contribution from D^0 decays is clearly visible right below the veto threshold.
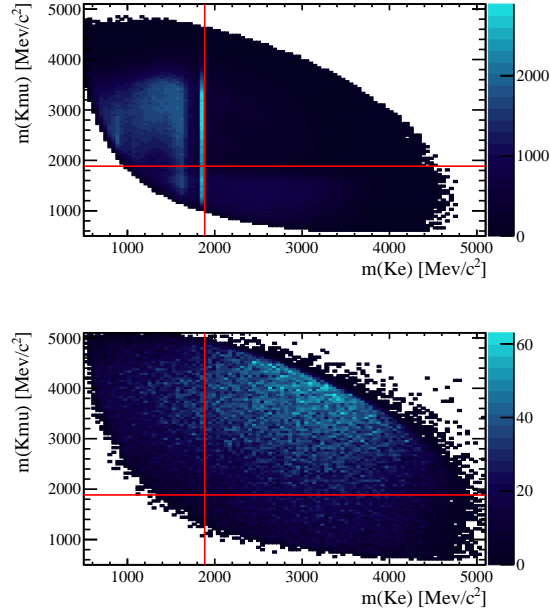

Figure 5.1: Plane spanned by the Ke and $K\mu$ invariant masses in $B^+ \to K^+ e^\pm \mu^\mp$. 2018 data is shown on the top and 2018 simulation on the bottom. The applied vetos against semileptonic cascades listed in Table A.11 are indicated by the red lines. The left, bottom and bottom left regions are discarded, the top right region is kept by vetos.

5.2.2 Background subtraction with the sPlot method

After application of the vetos against physical backgrounds, the remaining background in the control modes $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ and $B^+ \to K^+ J/\psi (\to e^+ e^-)$ can be subtracted using sWeights, similar to the procedure described in Section 4.2.2. While the same signal model is used for $B^+ \to K^+ J/\psi (\to e^+ e^-)$, it differs for $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$. Here, the peak is described using the sum of a Cruijff function [149, 150] and a Gaussian function. The Cruijff function itself is a distorted gaussian, where width and distortion are independent for the left and right side, defined as

$$C(x; \mu, \sigma_L, \sigma_R, \alpha_L, \alpha_R, \beta) = \begin{cases} \exp \frac{-(x-\mu)^2 \cdot (1+\alpha_L \cdot ((x-\mu)/\beta)^2)}{2 \cdot (\sigma_L^2 + \alpha_L \cdot (x-\mu)^2)}, & \text{for } (x-\mu) < 0 \\ \exp \frac{-(x-\mu)^2 \cdot (1+\alpha_R \cdot ((x-\mu)/\beta)^2)}{2 \cdot (\sigma_R^2 + \alpha_R \cdot (x-\mu)^2)}, & \text{for } (x-\mu) > 0 \end{cases}, \quad (5.2)$$

with the mean μ, the width of the left and right halves $\sigma_{L/R}$, the parameters governing the strength of the distortion of the gaussian $\alpha_{L/R}$ and a regularisation parameter β.

Another change compared to Section 4.2.2 is the description of partially reconstructed backgrounds. While most contributions from these backgrounds can be removed using requirements on the B^+ mass calculated with the J/ψ constrained to its nominal value, this comes with two issues. First, as this requirement does not remove the partially reconstructed background completely, a small contribution is left and can interfere with the modelling of combinatorial background and the lower tail of the signal peak. Second, the selection sculpts the lower tail of the signal peak. While these effects can in principle also affect $B^+ \to K^+ J/\psi (\to e^+ e^-)$, the selection to exclude partially reconstructed background is necessary here in order to be able to describe the data. As the dielectron mode does not enter the normalisation and is only used for cross-checks of the $B^+ \to K^+ e^{\pm} \mu^{\mp}$ signal shape, this slight mismodelling does not affect the result. However these effects do affect the much cleaner $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$, leading to instability of the description of the tails of the distribution, making careful treatment necessary.

The partially reconstructed background is described using a Cruijff function as well. As a proxy for the shape, simulation of the decay $B^0 \to K^{*0} (\to K^+ \pi^-) J/\psi (\to \mu^+ \mu^-)$, reconstructed as $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$, is used. Starting values for the fit to data are determined from a fit to simulation, while all parameters are left free to float in the fit to data. This shape is additionally convoluted with a Gaussian function to account for differences in resolution between data and simulation. In the fit to data, the complete background PDF consists of the Cruijff convoluted with a Gaussian for partially reconstructed background and an exponential function for combinatorial background. The resulting fits are shown in Figure 5.2.
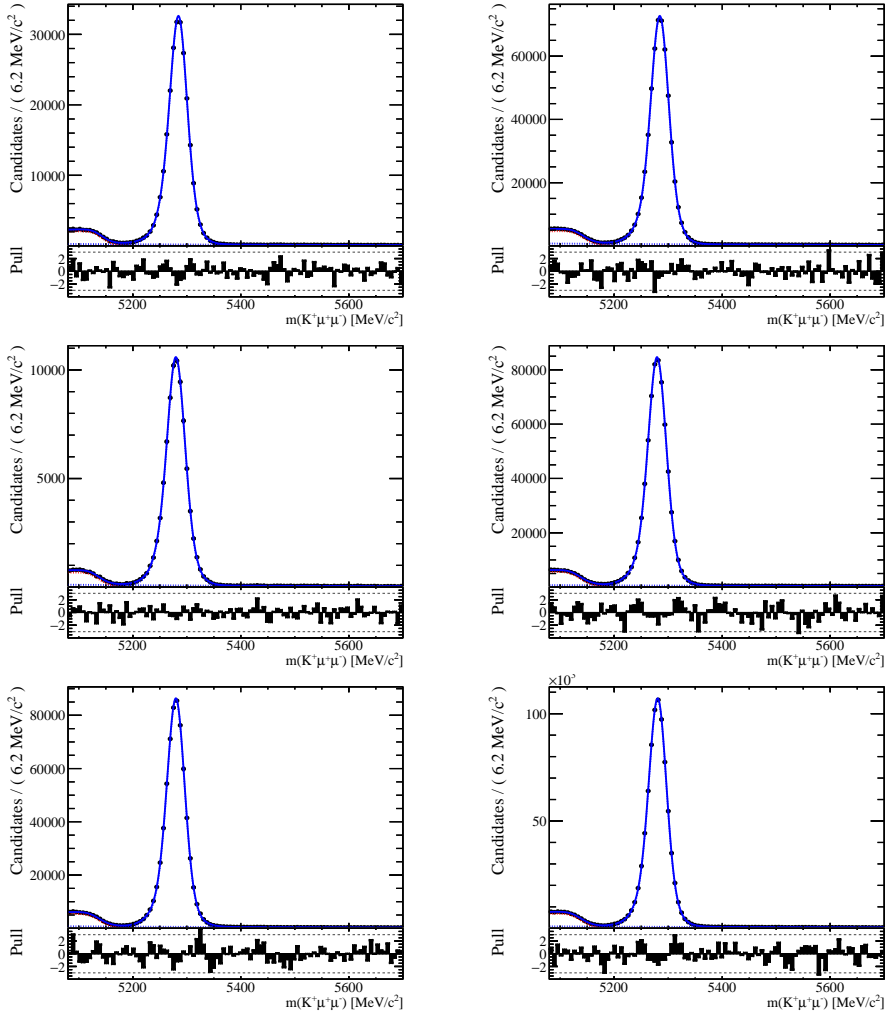
Figure 5.2: Fit to the control channel $K^+\mu^+\mu^-$ invariant mass distribution in data, split per year. 2011 (left) and 2012 (right) are shown on the top, 2015 (left) and 2016 (right) in the middle and 2017 (left) and 2018 (right) are shown on the bottom. The whole PDF is shown in blue, the signal component is shown as a blue dashed line, the background component as a red dashed line.

5.2.3 Kinematic reweighting

While in the search for $B^0_{(s)} \to e^+ e^-$ a resampling and a reweighting are performed, the former is omitted in $B^+ \to K^+ e^\pm \mu^\mp$ as the PID efficiencies are estimated from calibration data, making the resampling superfluous. The kinematic reweighting follows a similar strategy as explained in Section 4.2.3. The reweighting is performed using a GBReweighter, trained separately per year, where the inputs consist of the B^+ meson transverse momentum, pseudorapidity and vertex fit quality χ^2_{vtx} as well as the track multiplicity of the event.

A choice must be made which of the two control modes $B^+ \to K^+ J\!/\!\psi (\to e^+ e^-)$ and $B^+ \to K^+ J\!/\!\psi (\to \mu^+ \mu^-)$ is used for the reweighting. Both decays have been tested, and while the process works for both modes, a problem arises when applying the reweighting to $B^+ \to K^+ e^\pm \mu^\mp$. As the selection on the hardware trigger level is much tighter for electrons than for muons, many $B^+ \to K^+ e^\pm \mu^\mp$ candidates lie in a kinematic range not covered by $B^+ \to K^+ J\!/\!\psi (\to e^+ e^-)$, as the tighter L0 requirements are only applied to one of the final state particles. The reweighter trained on $B^+ \to K^+ J\!/\!\psi (\to e^+ e^-)$ is incapable of properly handling these candidates, assigning essentially random weights to them, leading to an incorrect reweighting. This can be avoided by using the reweighter trained on $B^+ \to K^+ J\!/\!\psi (\to \mu^+ \mu^-)$ also for $B^+ \to K^+ e^\pm \mu^\mp$.

The effect of the reweighting trained on $B^+ \to K^+ J\!/\!\psi (\to \mu^+ \mu^-)$ on the input variables is shown in Figure 5.3. The input variables to the multivariate classifiers show good agreement between data and simulation after the reweighting as well.

5.2.4 Multivariate Analysis

After the kinematic reweighting, a multivariate analysis is performed to significantly reduce the background level in $B^+ \to K^+ e^\pm \mu^\mp$ data. The strategy for the multivariate analysis performed in $B^+ \to K^+ e^\pm \mu^\mp$ differs from the one explained for $B^0_{(s)} \to e^+ e^-$ in Section 4.2.4. First, a BDT is trained with the upper mass sideband used as background proxy to reduce combinatorial background, as was done in $B^0_{(s)} \to e^+ e^-$. Afterwards however, a second BDT called BDTHOP is trained using the remaining lower mass sideband as background proxy. Since the lower mass sideband is dominated by partially reconstructed backgrounds at this point, this allows to specifically reduce these backgrounds. As these backgrounds might leak into the signal region similar to what is observed in the search for $B^0_{(s)} \to e^+ e^-$ (see Figure 4.16) this suppression leads to an increased sensitivity.
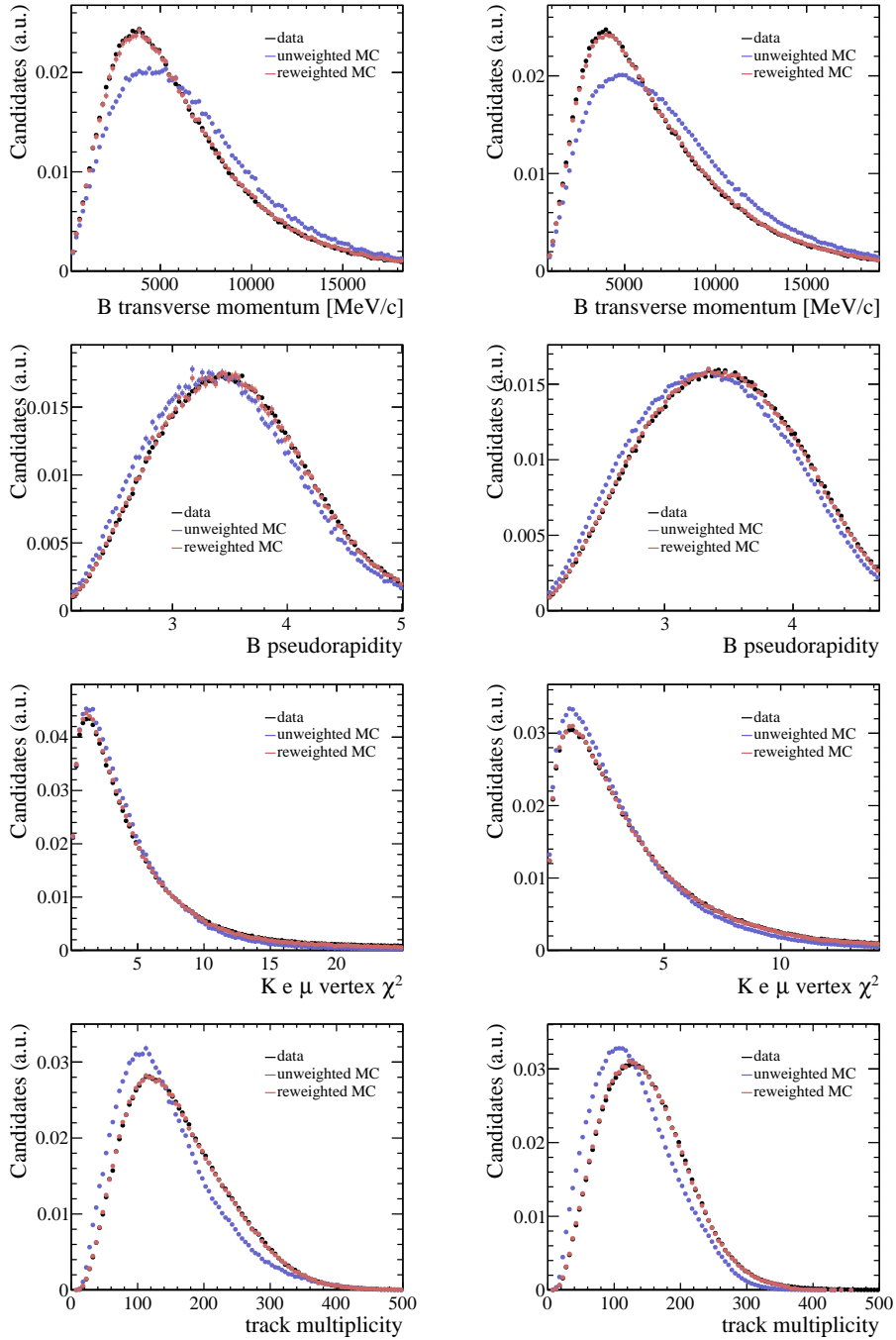
Figure 5.3: Input variables of the reweighting classifier trained on $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$, compared for sWeighted data (black), unweighted simulation (blue) and reweighted simulation (red). As an example for Run 1 the results for 2012 are shown on the left, as example for Run 2 2018 is shown on the right.

BDT against combinatorial background

The BDT against combinatorial background is trained using simulated $B^+ \to K^+ e^\pm \mu^\mp$ events as a signal proxy and the upper mass sideband in $B^+ \to K^+ e^\pm \mu^\mp$ data, which is dominated by combinatorial background, as background proxy. As classifier, the GradientBoostingClassifier from SCIKIT-LEARN [116] is used. The BDT is trained in 10 folds using the same strategy of k-folding explained in Section 4.2.4.

The BDT is trained separately for the $B^+ \to K^+ e^- \mu^+$ and $B^+ \to K^+ e^+ \mu^-$ charge configurations, since the signal and background distributions differ significantly depending on charge configuration, as illustrated in Figure 5.4. Splitting the BDT by charge thus allows to exploit these differences to increase the sensitivity of the search.

As inputs, the BDT uses:

- the B^+ meson momentum,

- the cosine of the B^+ direction angle,

- the B^+ χ^2_{IP},

- the B^+ χ^2_{vtx},

- the count-based track isolation from $B^0_s \to \mu^+ \mu^-$ for the leptons,

- the impact parameters of both leptons,

- the logarithm of the χ^2_{IP} of each daughter,

- the radial distance between two daughters for all combinations (Ke, $K\mu$, $e\mu$),

- the BDT-based track isolation from $B^0_s \to \mu^+ \mu^-$; Sum for e, μ and K tracks.

The BDT-based track isolations refer to the variables also used in the $B^0_{(s)} \to e^+ e^-$ BDT described in Section 4.2.4. However, an updated version compared to the search for $B^0_{(s)} \to e^+ e^-$ is used here. While the input variables stayed the same, the model for the BDT used to define the isolation was updated to increase performance. The radial distance between two particles is defined as

$$\Delta R(d1, d2) = \sqrt{(\phi(d1) - \phi(d2))^2 + (\eta(d1) - \eta(d2))^2}, \tag{5.3}$$

with the pseudorapidities η and azimuthal angles ϕ of the two final state particles $d1$ and $d2$. They are shown in the bottom row in Figure 5.4. The sharp cut-off at π originates from the cyclical nature of ϕ, where the maximum distance between two particles can be π radians.
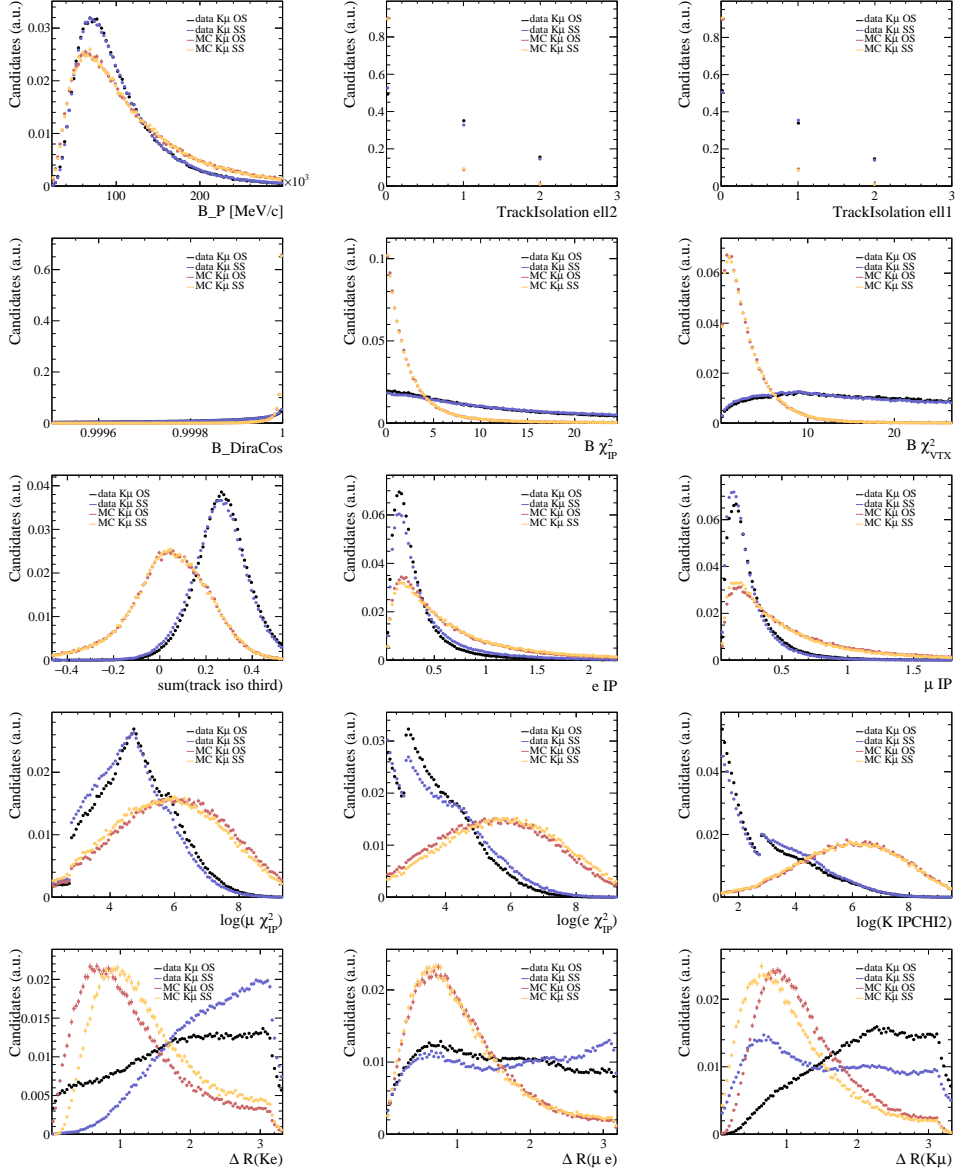
Figure 5.4: Input variables of the BDT classifier against combinatorial background for Run 2. The upper mass sideband in data is shown in black for the $K\mu$ OS sample and in blue for the $K\mu$ SS sample. The simulation is shown in red for the $K\mu$ OS sample and orange for the $K\mu$ SS sample.

In Figure 5.4, one of the main reasons for performing the analysis split in the charge configurations $B^+ \to K^+ e^- \mu^+$ and $B^+ \to K^+ e^+ \mu^-$ becomes immediately apparent, as the MVA inputs in data (shown in blue for the $B^+ \to K^+ e^- \mu^+$ case and black for the $B^+ \to K^+ e^+ \mu^-$ case) differ between the charge configurations. Similar disagreement between the charge configurations are observed in simulation (shown in yellow for the $B^+ \to K^+ e^- \mu^+$ case and red for the $B^+ \to K^+ e^+ \mu^-$ case). While this difference between charge configurations can be observed also before application of the offline preselection in data, this is not the case in simulation, where near perfect agreement was observed. The differences in simulation are instead introduced during the application of the vetos against partially reconstructed backgrounds, as the effect on the charge configurations is asymmetric due to the worse resolution in the invariant mass of the *Ke* system.

For the training of the BDT, separate years of data taking are combined according to their run period, as no large differences between the years are observed and the BDT can be trained on the largest sample available this way. The efficiencies are however calculated per year to check for unexpected differences. The classifier output for one exemplary fold in each run period and charge configuration is shown in Figure 5.5.

After the training, the selection requirement on the BDT is optimised by maximising the Punzi figure of merit as was also done in Section 4.2.4. The resulting distribution together with the signal efficiency and expected number of background candidates is shown in Figure 5.6.

As the figure of merit becomes unstable for high classifier values, where the background gets reduced to zero, the cut-point is not chosen as the maximum of the distribution, but rather the maximum in the region where at least one candidate is expected in the blinded region. This avoids an unnecessary reduction of the signal efficiency while keeping the sensitivity high. The cut point is marked in Figure 5.6 as a red dashed line.

BDT against partially reconstructed backgrounds

After application of the BDT selection, a second BDT (BDTHOP) is trained on the remaining sample. It uses $B^+ \to K^+ e^\pm \mu^\mp$ simulation as a signal proxy like the first BDT, however as background proxy the remainder of the lower mass sideband after the BDT selection is used. At this point most of the combinatorial background has been removed from the data, so the lower mass sideband is dominated by partially reconstructed backgrounds not removed by the vetos described in Section 5.2.1. The name of this second BDT originates from the addition of the so-called HOP variable
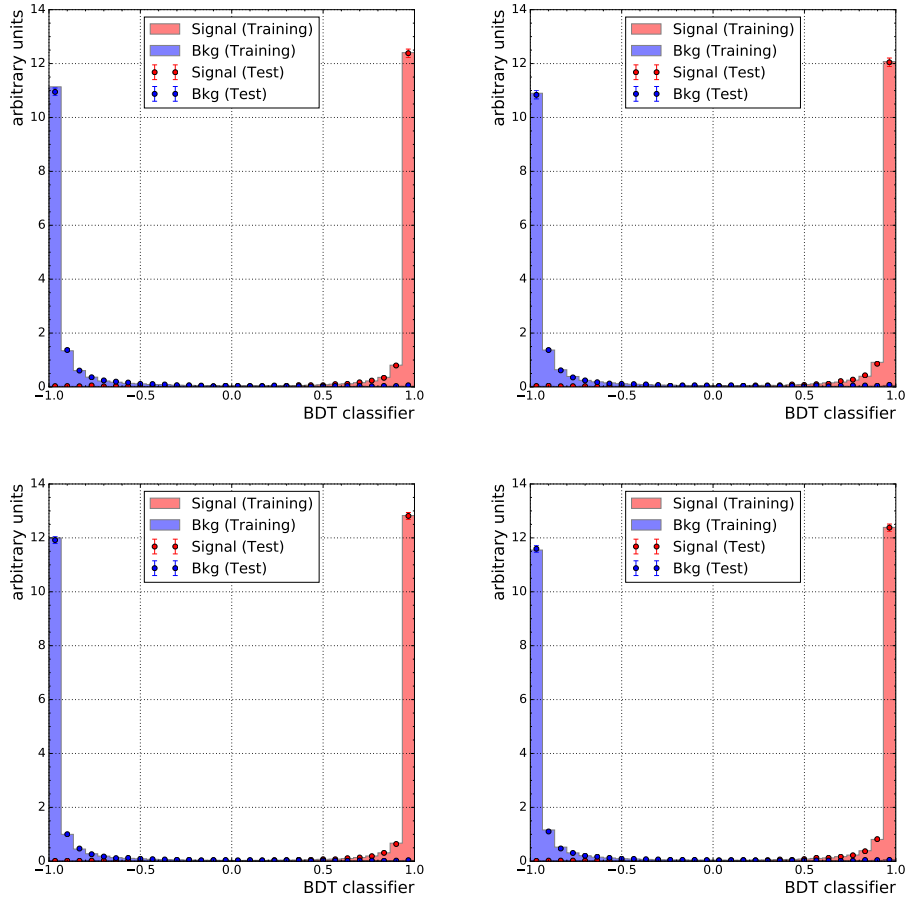
Figure 5.5: BDT classifier output for one exemplary fold. Run 1 is shown on the top, Run 2 on the bottom. The classifiers trained on the $B^+ \to K^+ e^+ \mu^-$ sample are shown on the left, the classifiers trained on the $B^+ \to K^+ e^- \mu^+$ sample on the right.

Figure 5.6: Result of the BDT classifier cut optimisation. The figure of merit is shown in blue, the signal efficiency in green and the number of backgro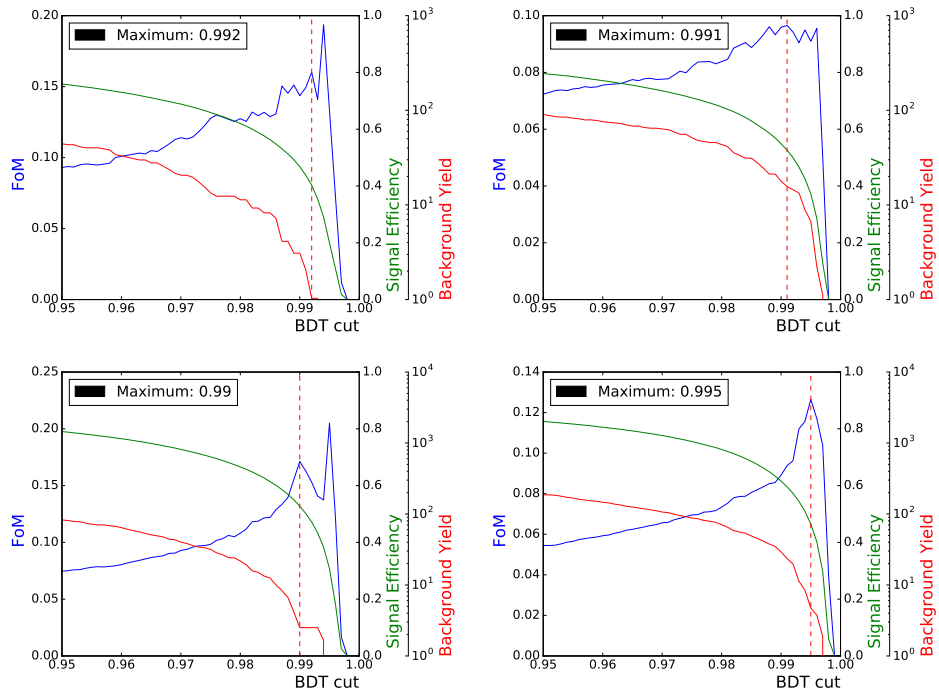und candidates in red. Run 1 is shown on the top, Run 2 on the bottom. The result for the classifiers trained on the $B^+ \to K^+ e^+ \mu^-$ sample are shown on the left, for the classifiers trained on the $B^+ \to K^+ e^- \mu^+$ sample on the right.

[151] to the inputs. It is defined as the ratio of momenta of the $K\mu$ system and the electron, balanced around the direction of the B^+ momentum. It is illustrated in Figure 5.7.
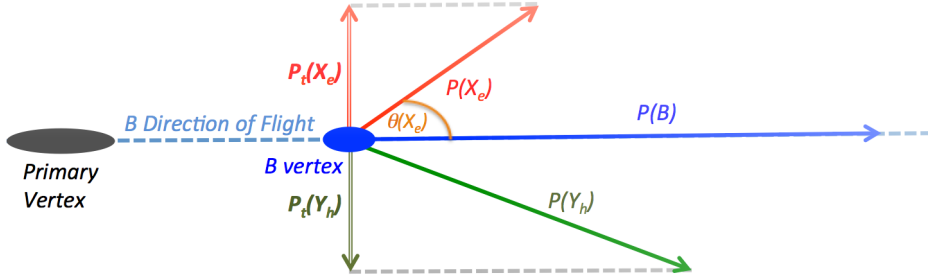
Figure 5.7: Schematic of the HOP variable.

As all momentum components transverse to the direction of the B^+ momentum should sum up to 0, this ratio is expected to be 1 for fully reconstructed $B^+ \to K^+ e^{\pm} \mu^{\mp}$ decays. In the case of partial reconstruction, where energy and thus momentum are lost from missing a particle, this balance does not hold and the ratio can deviate from unity. The same is true for missing bremsstrahlung, so the HOP is not perfectly 1 even for $B^+ \to K^+ e^{\pm} \mu^{\mp}$ signal. The effect is however smaller on average as for missing particles. The inclusion of this ratio into the BDT thus allows to increase the separation power of the BDTHOP against partially reconstructed backgrounds.

While the HOP ratio is added to the inputs of the BDTHOP, it was observed that some inputs of the BDT have little to no separation power left after the application of the BDT selection. These variables are removed from the classifier in order to reduce the model complexity, since the BDTHOP is trained on a much smaller data sample than the BDT, making it susceptible to overtraining. The BDTHOP input variables are:

- HOP

- the B^+ meson momentum,

- the B^+ χ^2_{IP},

- the B^+ χ^2_{vtx},

- the logarithm of the χ^2_{IP} of the kaon,

- the radial distance between the kaon and the same-sign lepton.

The choice on which variables to remove from the input samples was made by comparing the resulting area under the RoC with and without the feature in question included in the training. The feature with the lowest feature importance in the training is then removed and the process repeated until removing a feature leads to significant loss in separation power as measured by the area under the RoC.

The classifier for one exemplary fold is shown in Figure 5.8. The optimisation of the cut on the classifier is done as for the BDT, leading to the figure of merit shown in Figure 5.9. The separation power of the BDTHOP is considerably lower than that of the BDT, since the remaining physical backgrounds are more similar to $B^+ \to K^+ e^\pm \mu^\mp$ signal and the model complexity is reduced due to the smaller training sample. Still, by selecting on the BDTHOP, the background remaining in the lower mass sideband, which is also expected to leak into the signal region, can be reduced significantly while keeping reasonable signal efficiency, thus increasing the sensitivity of the search.

5.2.5 Requirements on particle identification

After the selection on BDT and BDTHOP, backgrounds from partial reconstruction as well as combinatorial background have been reduced significantly, leaving misidentification as a potential source of background. Especially backgrounds of the type $B^+ \to K^+ hh$ are dangerous, as they can have similar a topology as $B^+ \to K^+ e^\pm \mu^\mp$ and peak close to the B^+ mass, similar to the $B \to hh$ backgrounds in the search for $B^0_{(s)} \to e^+ e^-$. To suppress these backgrounds, requirements on the probability of each final state particle to be the particle it is assumed to be (ProbNN), are applied to the data. The selection is taken from the previous search for $B^+ \to K^+ e^\pm \mu^\mp$ in Run 1. As the tuning of the ProbNN changed between Run 1 and Run 2, the selection needs to be adjusted in Run 2 years. The selection is chosen to reproduce the Run 1 efficiency for each final state track in simulation. While MC12TuneV2 (V3 for electrons) is used in Run 1, MC15TuneV1 is used in Run 2 for all particles. The applied selection is summarised in Table A.12.
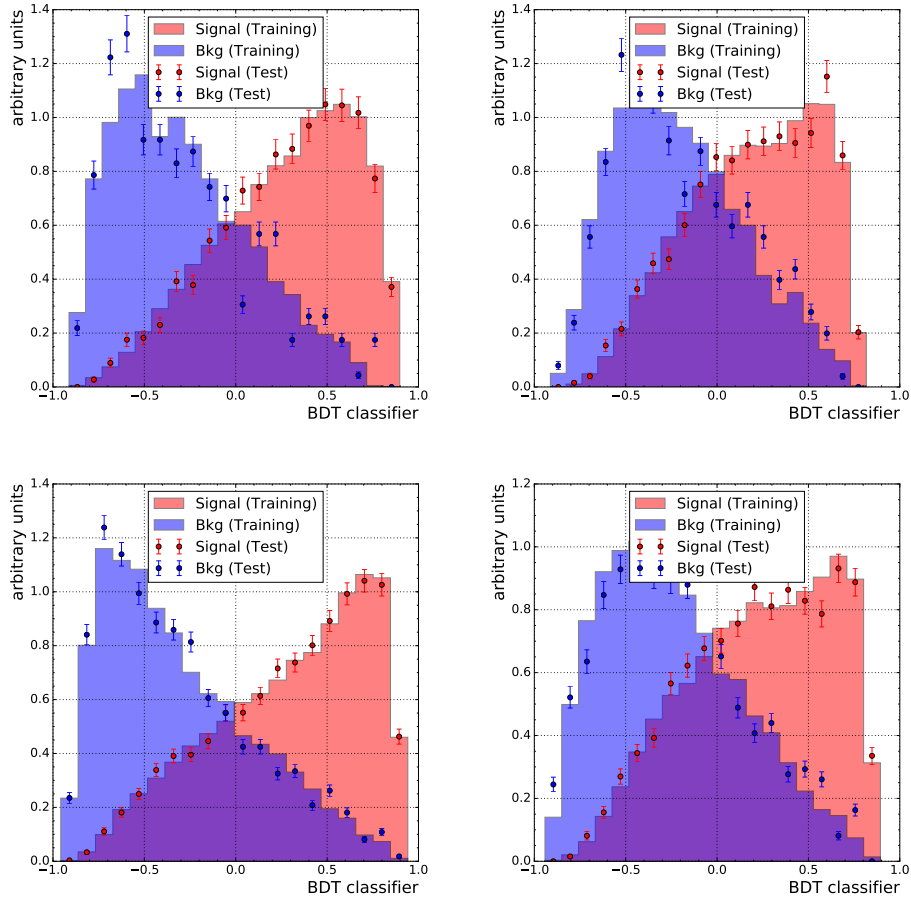
Figure 5.8: BDTHOP classifier output for one exemplary fold. Run 1 is shown on the top, Run 2 on the bottom. The classifiers trained on the $B^+ \to K^+ e^+ \mu^-$ sample are shown on the left, the classifiers trained on the $B^+ \to K^+ e^- \mu^+$ sample on the right.

Figure 5.9: Result of the BDTHOP classifier cut optimisation. The figure of merit is shown in blue, the signal efficiency in green and the number of backgro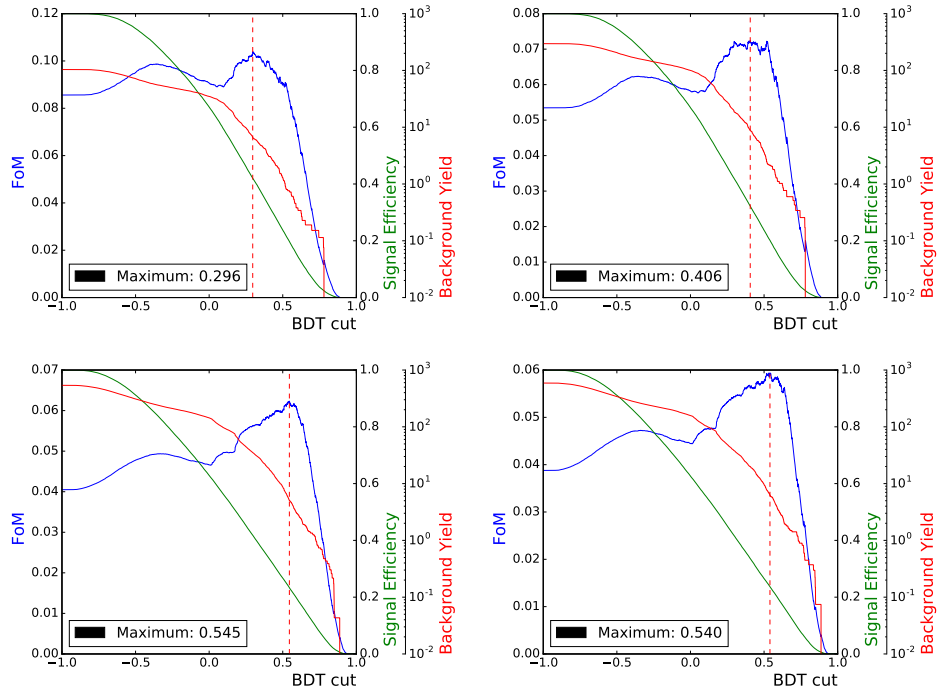und candidates in red. Run 1 is shown on the top, Run 2 on the bottom. The result for the classifiers trained on the $B^+ \to K^+ e^+ \mu^-$ sample are shown on the left, for the classifiers trained on the $B^+ \to K^+ e^- \mu^+$ sample on the right.

5.3 Normalisation constant determination

The following section describes the determination of inputs necessary for the calculation of an upper limit on the branching fraction of $B^+ \to K^+ e^\pm \mu^\mp$. First, efficiencies (see Section 5.3.1) and systematic uncertainties (see Section 5.3.5) are determined. Since the determination of the trigger efficiency (see Section 5.3.2) is done using a different approach from the other selection efficiencies, it is discussed separately. The determination of PID efficiencies follows the same procedure as described in Section 4.3.3, so it is not repeated here. The total selection efficiencies are discussed in Section 5.3.3.

5.3.1 Efficiencies

While the formulae given for the efficiencies and their uncertainties presented in Section 4.3.1 also hold here with the same reasoning, the overall efficiency is defined slightly differently. While in the search for $B^0_{(s)} \to e^+ e^-$, the trigger efficiency is defined after the rest of the selection, here the definition of efficiencies follows the order of selection steps, leading (together with the added selections steps compared to $B^0_{(s)} \to e^+ e^-$) to the total selection efficiency being defined as

$$
\begin{aligned}
\epsilon = \epsilon_{\text{geo}} &\times \epsilon_{\text{stripxreco | geo}} \times \epsilon_{\text{trig | stripxreco}} \times \epsilon_{\text{sel | trig}} \times \epsilon_{\text{veto | sel}} \\
&\times \epsilon_{\text{BDT | veto}} \times \epsilon_{\text{BDTHOP | BDT}} \times \epsilon_{\text{PID | BDTHOP}}.
\end{aligned}
\tag{5.4}
$$

While most efficiencies are determined from simulation with data driven corrections, the efficiencies of the trigger and PID selections are determined from data instead, since these efficiencies are not well described in simulation and do not get corrected by the kinematic reweighting. All uncertainties quoted below are purely statistical unless specified otherwise.

5.3.2 Trigger calibration

The trigger efficiency is calculated from data using the TISTOS method, following a similar procedure as described in Section 4.3.2. On $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ the binning is however not carried out in the B^+ transverse and longitudinal momenta, but rather the triggered muons transverse momentum and impact parameter, where the triggered muon is defined to be the muon that caused the hardware trigger to fire. In case both muons did, one is chosen at random. This is done to be more

similar to the procedure on $B^+ \to K^+ e^\pm \mu^\mp$, where the efficiency is transferred using a map in the muons transverse momentum and impact parameter.

Another difference compared to the search for $B^0_{(s)} \to e^+ e^-$ is the point in the analysis chain where the trigger efficiency is calculated. In the search for $B^0_{(s)} \to e^+ e^-$, the trigger efficiency is estimated after the whole selection chain has been applied. While this has the advantage of a clean sample of $B^+ \to K^+ J/\psi(\to e^+ e^-)$ data, it comes with the drawback of reduced statistics in the sample. While this tradeoff is necessary for $B^+ \to K^+ J/\psi(\to e^+ e^-)$, where the dielectron final state of the J/ψ causes difficulties to disentangle signal from background, $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ is much cleaner than its electron counterpart already after the stripping. This allows to perform the necessary fits to the data while keeping the largest possible data sample, leading to a more precise estimation of the trigger efficiency.

The results of the efficiency estimation are shown in Table 5.4.

Table 5.4: Efficiency of the trigger selection on $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$, estimated from data using the TISTOS method. The results from applying the TISTOS method to simulation, the true efficiency on simulation and the observed bias are reported as well.

Year	Data [%]	MC TISTOS [%]	MC TRUE [%]	Bias [%]
2011	71.33 ±1.06	73.30 ±0.94	72.10 ±0.09	1.63
2012	72.37 ±0.65	74.10 ±0.68	73.34 ±0.06	1.03
2015	58.07 ±1.12	58.32 ±0.50	57.57 ±0.07	1.28
2016	72.90 ±0.37	72.26 ±0.27	70.74 ±0.04	2.09
2017	75.08 ±0.40	75.41 ±0.18	74.23 ±0.03	1.57
2018	76.67 ±0.41	77.13 ±0.21	76.45 ±0.03	0.88

The estimation of the trigger efficiency for $B^+ \to K^+ e^\pm \mu^\mp$ is calculated by transferring the trigger efficiency from the control mode $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ by assigning to each candidate in $B^+ \to K^+ e^\pm \mu^\mp$ simulation an efficiency from a efficiency map calculated in the muons transverse momentum and impact parameter. Since in $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ there are two muons in the final state, the choice of variable to do the transfer in needs to be treated carefully to not bias the efficiency. Three choices were investigated:

- The maximum value of the two muons

- The triggered muons value

- A random muons value

The first two approaches were found to be biasing the efficiency towards higher efficiencies. This is expected, as having two muons in the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ final state the chance of one having a high p_T is increased. Since the L0 performance for muons is mainly driven by the transverse momentum, this increase in average transverse momentum biases the trigger efficiency upwards. The same is true when choosing the triggered muon for the transfer, as implicitly the same choice of the higher transverse momentum muon is made. Choosing the muon for the transfer randomly on the other hand allows to estimate the $B^+ \to K^+ e^\pm \mu^\mp$ trigger efficiency without large bias. As a cross-check, choosing either muon for the transfer and averaging the results for both muons was also tested and gave consistent results. The results are shown in Table 5.5. The bias is found to differ by year due to differences in the size of the data sample available to estimate the trigger efficiency and differences in the input variables between years.

Table 5.5: Efficiency of the trigger selection on $B^+ \to K^+ e^\pm \mu^\mp$, estimated from data using the TISTOS method. The results from applying the TISTOS method to simulation, the true efficiency on simulation and the observed bias are reported as well.

Year	Data [%]	MC TISTOS [%]	MC TRUE [%]	Bias [%]
2011	51.56 ±3.10	47.63 ±2.87	46.21 ±0.08	3.07
2012	49.32 ±1.83	46.06 ±1.99	45.41 ±0.08	1.43
2015	38.92 ±2.68	34.07 ±1.19	34.27 ±0.11	0.60
2016	47.84 ±1.24	45.62 ±0.81	45.56 ±0.12	0.14
2017	52.82 ±1.21	50.19 ±0.59	50.77 ±0.08	1.15
2018	51.53 ±1.09	49.96 ±0.60	49.85 ±0.09	0.22

5.3.3 Total selection efficiencies

The total selection efficiencies for $B^+ \to K^+ e^\pm \mu^\mp$ (split by charge configuration) and $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ are summarised in Table 5.6. The ratios entering the normalisation constants are summarised in Table 5.7.

Table 5.6: Total selection efficiencies split by decay channel, year and charge configuration.

Year	$\epsilon_{B^+ \to K^+ e^- \mu^+}$ [%]	$\epsilon_{B^+ \to K^+ e^+ \mu^-}$ [%]	$\epsilon_{B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)}$ [%]
2011	0.0629 ± 0.0039	0.0366 ± 0.0023	0.0934 ± 0.0017
2012	0.0624 ± 0.0025	0.0416 ± 0.0017	0.1434 ± 0.0017
2015	0.0575 ± 0.0042	0.0341 ± 0.0025	0.1570 ± 0.0032
2016	0.0781 ± 0.0025	0.0479 ± 0.0016	0.2435 ± 0.0017
2017	0.0860 ± 0.0023	0.0498 ± 0.0014	0.2519 ± 0.0017
2018	0.0780 ± 0.0020	0.0459 ± 0.0012	0.2408 ± 0.0016

Table 5.7: Efficiency ratios entering the normalisation constants, split by year and charge configuration.

Year	$\dfrac{\epsilon_{B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)}}{\epsilon_{B^+ \to K^+ e^- \mu^+}}$	$\dfrac{B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)}{B^+ \to K^+ e^+ \mu^-}$
2011	1.484 ± 0.096	2.551 ± 0.165
2012	2.296 ± 0.095	3.448 ± 0.144
2015	2.731 ± 0.206	4.608 ± 0.349
2016	3.118 ± 0.103	5.079 ± 0.171
2017	2.929 ± 0.081	5.059 ± 0.142
2018	3.086 ± 0.082	5.246 ± 0.143

It can be seen that the selection efficiency for the control mode $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ is considerably higher than for the signal modes $B^+ \to K^+ e^\pm \mu^\mp$, contrary to the search for $B^0_{(s)} \to e^+ e^-$, where the opposite trend was observed. This difference in efficiency between $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ originates from multiple points in the selection. The main difference comes from the application of vetos against semileptonic cascades. As $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ has two muons in the final state, the invariant K masses are more narrow than in $B^+ \to K^+ e^\pm \mu^\mp$, where the Ke mass is quite broad from losses of bremsstrahlung and generally worse momentum resolution of the electron track. This effect is directly visible in Figure 5.1, in the asymmetry of the distribution. Since only the narrow case exists for $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$, the efficiency of the vetos is higher. Another large difference in efficiency originates from the trigger, where the control mode shows considerably higher efficiency from the presence of the second muon, as visible from Tables 5.4 and 5.5. The last noteworthy difference comes in the BDTHOP. As the HOP variable itself is defined to be one for $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ on both data and simulation, no cut is applied to it in the BDTHOP, leading to an increase in

efficiency compared to $B^+ \to K^+ e^\pm \mu^\mp$.

In addition to the difference in efficiency observed between $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$, Table 5.7 also shows a difference in selection efficiencies between the $B^+ \to K^+ e^- \mu^+$ and $B^+ \to K^+ e^+ \mu^-$ charge configurations, originating from multiple points. While the thresholds of the vetos applied to reject physical backgrounds is identical, the distributions of the invariant K masses are asymmetric as seen in Figure 5.1. Since the vetos are only applied to the case where the kaon and lepton have opposite charge, the efficiency is different. Further differences come from the two BDTs, where the training and optimisation are done separately for the two charge configurations. Due to differences in the present backgrounds, the cut points needed to remove them are different, leading to the seen differences in efficiency.

5.3.4 Normalisation constants

Several Inputs, as shown in Equation (5.1) are needed to calculate the branching fraction of $B^+ \to K^+ e^\pm \mu^\mp$. This includes

- The ratio of selection efficiencies of $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$

- The signal yield in $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ after the selection

- The branching fraction of $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$

The efficiency ratios, split by charge are shown above in Table 5.7. The branching fraction of $B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)$ is the same as for $B^+ \to K^+ J/\psi(\to e^+ e^-)$, just replacing the branching fraction of $J/\psi \to e^+ e^-$ for the one of $J/\psi \to \mu^+ \mu^-$ taken from [107], leading to

$$\mathcal{B}(B^+ \to K^+ J/\psi(\to \mu^+ \mu^-)) = (1.010 \pm 0.028) \cdot 10^{-3} \times (5.961 \pm 0.033) \cdot 10^{-2}$$
$$= (6.02 \pm 0.17) \cdot 10^{-5}. \tag{5.5}$$

The yield of the control mode after the whole selection is obtained from repeating the fit described in Section 5.2.2 on the fully selected data. The fits are shown on Figure 5.10, the yields are summarised in Table 5.8.
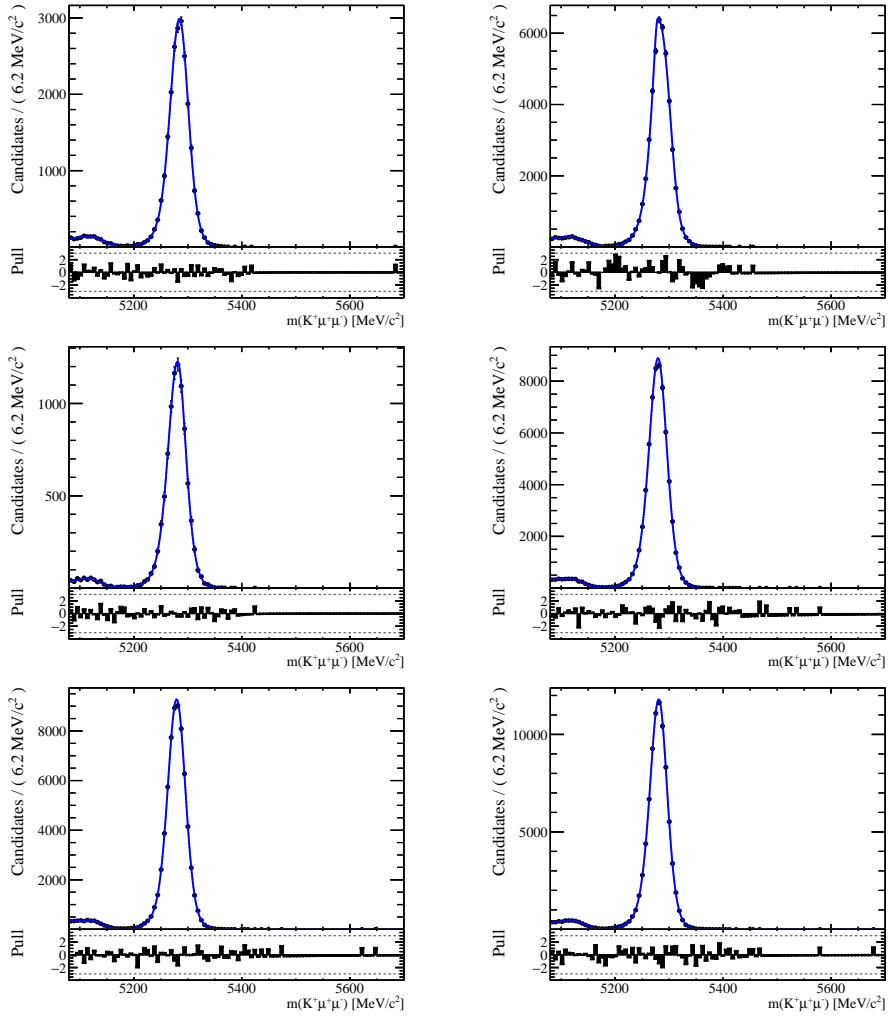
Figure 5.10: Fit to the control channel $K^+\mu^+\mu^-$ invariant mass distribution in data after the whole selection, split per year. 2011 (left) and 2012 (right) are shown on the top, 2015 (left) and 2016 (right) in the middle and 2017 (left) and 2018 (right) are shown on the bottom. The whole PDF is shown in blue, the signal component is shown as a blue dashed line, the background component as a red dashed line.

Table 5.8: Yields of the normalisation mode $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ computed from the fits shown in Figure 5.10.

Year	$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ yield
11	21772.0 ± 119.9
12	46668.0 ± 214.6
15	8809.2 ± 103.1
16	63610.0 ± 112.5
17	65372.0 ± 255.7
18	82268.0 ± 287.2

With these values, the (partial) normalisation constants $\alpha^{(\prime)}$ can be calculated per year. The partial normalisation constants separated per year and charge configuration are reported in Table 5.9, the full normalisation constants are shown in Table 5.10. As is expected from the low integrated luminosity recorded in 2015, the (partial) normalisation constant is significantly larger than in the other years. The normalisation constants are then also combined according to their run period as shown in Equation (4.18), the result of the combination is shown in Table 5.11.

Table 5.9: Partial normalisation constants α' for both charge configurations, split by year.

Year	$\alpha'(B^+ \to K^+ e^- \mu^+)[10^{-4}]$	$\alpha'(B^+ \to K^+ e^+ \mu^-)[10^{-4}]$
11	0.682 ± 0.044	1.172 ± 0.076
12	0.492 ± 0.021	0.739 ± 0.031
15	3.100 ± 0.237	5.231 ± 0.401
16	0.490 ± 0.016	0.799 ± 0.027
17	0.448 ± 0.013	0.774 ± 0.022
18	0.375 ± 0.010	0.638 ± 0.018

Table 5.10: Normalisation constants α for both charge configurations, split by year.

Year	$\alpha(B^+ \to K^+ e^- \mu^+)[10^{-8}]$	$\alpha(B^+ \to K^+ e^+ \mu^-)[10^{-8}]$
11	0.409 ± 0.029	0.703 ± 0.050
12	0.295 ± 0.015	0.443 ± 0.022
15	1.859 ± 0.151	3.137 ± 0.256
16	0.294 ± 0.013	0.479 ± 0.021
17	0.269 ± 0.011	0.464 ± 0.018
18	0.225 ± 0.009	0.382 ± 0.015

Table 5.11: Normalisation constants α for both charge configurations, split by run period.

	$\alpha(B^+ \to K^+ e^- \mu^+)[10^{-8}]$	$\alpha(B^+ \to K^+ e^+ \mu^-)[10^{-8}]$
Run 1	0.171 ± 0.007	0.272 ± 0.011
Run 2	0.083 ± 0.002	0.139 ± 0.003

5.3.5 Systematic uncertainties

Various possible sources of systematic uncertainty are considered in the search for $B^+ \to K^+ e^{\pm} \mu^{\mp}$. They are described in the following sections.

Estimation of trigger efficiencies

Two sources of systematic uncertainty are considered for the estimation of the trigger efficiency. The first is the relative bias observed when performing the estimation on $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ simulation, following the logic laid out in Section 4.3.6. This bias is found to be between 1% and 2%. For the signal mode, where the transfer is repeated using maps calculated from $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ simulation to calculate the bias, the difference is found to be between 1% and 3%. The observed bias is much smaller than what was observed in $B^0_{(s)} \to e^+ e^-$. This is the case mainly because of the considerably higher statistics in $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ compared to $B^+ \to K^+ J/\psi (\to e^+ e^-)$, which allows for a finer binning in the application of the TISTOS method, reducing the observed bias. The bias for $B^+ \to K^+ e^{\pm} \mu^{\mp}$ is listed in the last column of Table 5.5, for $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ in the last column of Table 5.4.

The second source of systematic uncertainty is related to the statistics of the data samples used for the efficiency estimation. For $B^+ \to K^+ e^\pm \mu^\mp$ it is quantified by varying the map used in the transfer from the control mode within its uncertainty and repeating the efficiency estimation. This process is repeated 1000 times and the relative spread of the obtained efficiency distribution is assigned as systematic uncertainty. This systematic was found to be up to 2%.

For the trigger efficiency on $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$, no intermediate map is used for the calculation. Instead, the yields of the TIS, TOS and TISTOS subsamples are varied within their uncertainty to test the robustness of the efficiency estimation. Similar effects as for the signal mode are observed, with slightly lower systematic uncertainties when compared to $B^+ \to K^+ e^\pm \mu^\mp$ of up to 0.8%.

Differences between data and simulation

The validity of taking efficiencies from simulation is tested by comparing selection efficiencies taken from $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ simulation with their counterparts from fitting the data before and after the selection step in question. For the preselection and vetos only small differences of up to $\approx 0.2\%$ are found. This comparison is also done for the efficiencies of the BDT and BDTHOP. Here, it serves also as a measure for the quality of the kinematic reweighting, as differences between data and simulation that were not corrected can become apparent in the BDT when applied to signal in $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ data and simulation. For the BDT, differences of $\approx 1-3\%$ in the efficiency are found. For the BDTHOP, differences of up to $\approx 3\%$ are found. Since the BDTHOP shares many inputs with the BDT, these differences in observed systematic uncertainties can be attributed to the limited statistics available for the training of the BDTHOP.

Estimation of PID efficiencies

In order to estimate a systematic uncertainty associated to the binning scheme used for the PID efficiency estimation, a similar approach as described in Section 4.3.6 is used. For each variable, the binning is halved compared to the nominal approach by removing every other bin border and the relative difference is calculated. This is repeated for each binning variable, the sum in quadrature is assigned as systematic uncertainty, leading to systematics of $\approx 1\%$. For 2015 data, the systematic is considerably larger with 4% because of the very limited amount of calibration data available due to the low integrated luminosity recorded in 2015.

In contrast to the search for $B^0_{(s)} \to e^+ e^-$, no systematic uncertainty is assigned related to the sWeighting of the electron calibration sample. Instead, the efficiency is directly estimated using a fit&count approach. In this approach, instead of relying on the sWeights calculated on the whole calibration data, the fit to subtract background is performed in each kinematic bin and the yield extracted. This allows to correctly account for correlations between mass and kinematics without relying on the sWeights to be valid throughout the whole phase-space.

Another source of systematic uncertainty similarly to the estimation of the trigger efficiency is the impact of the limited statistics in the calibration data on the efficiency estimation. For this systematic, the efficiencies in each kinematic bin are varied according to their statistical uncertainty and the efficiency estimation is repeated. This process is done 1000 times and the relative width of the efficiency distribution is assigned as systematic uncertainty. It was found to be $\approx 0.2 - 0.3\,\%$.

Corrections to the track reconstruction efficiency

The efficiency of the reconstruction are taken from simulation. As $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ do not have the same final state, possible mismodelling of the track reconstruction efficiency, which depends on the detector conditions, can not cancel for the electron and one of the control channel muons. To correct the track reconstruction efficiencies of the electron and muons, centrally produced correction maps [125] are used. These maps parametrise the track reconstruction efficiency in the particle's momentum and pseudorapidity for muons and the particle's transverse momentum, pseudorapidity and azimuthal angle for electrons. For each simulated candidate the correction from the corresponding kinematic bin of the map is assigned to calculate an overall correction factor.

The corrections are obtained after the reweighting is applied for muons and right after the stripping selection for electrons. They are summarised in Table 5.12. The corrections for the muons are found to be in good agreement between the muon in $B^+ \to K^+ e^\pm \mu^\mp$ and both muons in $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$. Systematic uncertainties on these corrections depend on the map used for the calculation of the correction. For the muon maps, these systematic uncertainties are provided with the tables as $0.4\,\%$ in Run 1 and $0.8\,\%$ in Run 2. In case of the electron maps, the same technique as for the trigger and PID systematics related to the limited statistics of available samples is used to assign a systematic uncertainty, which amounts to up to $0.2\,\%$.

Table 5.12: Corrections to the track reconstruction efficiency for the muons and electron in $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$.

| | $B^+ \to K^+ e^\pm \mu^\mp$ | | $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ | |
Year	μ	e	μ_1	μ_2
11	0.9978 ± 0.0018	0.9832 ± 0.0254	0.9982 ± 0.0019	0.9983 ± 0.0019
12	1.0028 ± 0.0019	0.9855 ± 0.0197	1.0047 ± 0.0020	1.0046 ± 0.0020
15	1.0007 ± 0.0015	0.9920 ± 0.0317	0.9989 ± 0.0013	0.9988 ± 0.0013
16	0.9950 ± 0.0119	0.9875 ± 0.0115	0.9929 ± 0.0112	0.9928 ± 0.0112
17	0.9967 ± 0.0007	0.9935 ± 0.0115	0.9959 ± 0.0008	0.9959 ± 0.0008
18	0.9961 ± 0.0006	0.9925 ± 0.0109	0.9948 ± 0.0007	0.9948 ± 0.0007

Summary of systematic uncertainties

The systematic uncertainties described above are summarised in Table 5.13.

Table 5.13: Systematic uncertainties in the search for $B^+ \to K^+ e^\pm \mu^\mp$. The systematic uncertainties have been calculated from the approach described in Section 5.3.2. The total systematic uncertainty is calculated assuming all systematic uncertainties to be uncorrelated.

| | Syst. [%] | | | | | |
Analysis step	2011	2012	2015	2016	2017	2018
Preselection	0.1	0.1	0.1	0.1	0.1	0.1
Vetos	0.1	0.1	0.1	0.2	0.2	0.1
Trigger $B^+ \to K^+ J/\psi$	1.6	1.0	1.3	2.1	1.6	0.9
Trigger $B^+ \to K^+ e^\pm \mu^\mp$	3.1	1.4	0.6	0.1	1.2	0.2
Trigger (stat.) $B^+ \to K^+ J/\psi$	0.8	0.5	0.8	0.3	0.3	0.3
Trigger (stat.) $B^+ \to K^+ e^\pm \mu^\mp$	1.4	1.0	2.0	0.6	0.6	0.5
BDT	0.2	2.3	1.2	1.8	0.9	2.1
BDTHOP	2.0	2.0	1.9	2.7	1.4	0.4
PID binning	1.1	0.5	4.4	0.3	0.2	0.8
PID statistics	0.2	0.2	0.3	0.2	0.2	0.2
Track reconstruction μ	0.4	0.4	0.8	0.8	0.8	0.6
Track reconstruction e	0.2	0.2	0.2	0.1	0.1	0.1
Total	4.5	3.7	5.6	4.0	2.8	2.6

Comparing these systematic uncertainties to their counterparts from the search

for $B^0_{(s)} \to e^+ e^-$ shown in Table 4.10, some differences become apparent. The systematic uncertainties on the trigger efficiencies are considerably lower here than in $B^0_{(s)} \to e^+ e^-$, even when taking into account the additional systematic assigned on the statistics of the sample used to estimate the efficiency. This reduction in systematics from the much larger statistics in $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ compared to $B^+ \to K^+ J/\psi (\to e^+ e^-)$, which allows for a finer binning to reduce the bias of the TISTOS method. Also, no single systematic ncertainty is found to be consistently dominant over the others.

The systematic uncertainties on the BDT and BDTHOP are smaller as well, hinting at a better performance of the reweighting procedure, holding also in reduced samples as tested with the BDTHOP. This is expected, as the $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ mode is much cleaner than the $B^+ \to K^+ J/\psi (\to e^+ e^-)$ decay, allowing for a relaxed selection to describe backgrounds that might leak into the signal. This in turn allows for better subtraction of these backgrounds and thus reduced pollution of the signal in data, which is input to the reweighting.

Lastly, the PID systematics should be mentioned. One source of systematic uncertainty considered in $B^0_{(s)} \to e^+ e^-$ is not listed above, which is the sWeighting of the electron calibration sample. Here, the efficiency is instead calculated using a fit&count approach directly, so the sWeighting has no effect on the result. The systematic on the binning used for the efficiency calibration is similar for $B^0_{(s)} \to e^+ e^-$ and $B^+ \to K^+ e^\pm \mu^\mp$, with 2015 being an outlier in the case of $B^+ \to K^+ e^\pm \mu^\mp$ due to very low statistics in all three used samples, leading to more pronounced changes to the efficiency from changes in the binning. In case the binning is kept constant however, the effect of the limited statistics is similar to other years.

5.4 Result of the Search for $B^+ \to K^+ e^\pm \mu^\mp$

While in the search for $B^0_{(s)} \to e^+ e^-$ the branching fraction could be measured from an extended maximum likelihood fit to the unblinded data, the analysis presented here for the search for $B^+ \to K^+ e^\pm \mu^\mp$ is not yet approved for unblinding at the time of writing. Therefore only an expected upper limit on the branching fraction is computed.

5.4.1 Background parametrisation

After the selection, only very little background (see Figure 5.13) remains in the $B^+ \to K^+ e^\pm \mu^\mp$ data sidebands, making the description of the shape difficult especially in the $B^+ \to K^+ e^+ \mu^-$ case. Thus, a suitable proxy to extract the shape from

is needed. One candidate as proxy for the remaining combinatorial background are the $e\mu$ same-sign data reconstructed as $B^+ \to K^+ e^+ \mu^+$. This data sample is expected to contain only combinatorial background. To validate the usage of the $B^+ \to K^+ e^+ \mu^+$ data as a proxy for the shape of combinatorial background, the same selection as for the $B^+ \to K^+ e^\pm \mu^\mp$ data is applied to the $B^+ \to K^+ e^+ \mu^+$ data. However, care must be taken when applying the vetos and following selection steps, as they are sensitive to the charge configuration of the kaon and muon. The charge configurations however are not defined in the $B^+ \to K^+ e^+ \mu^+$ data, as for every candidate by definition the kaon has the same sign as both leptons. Thus, two samples of $B^+ \to K^+ e^+ \mu^+$ data are prepared, having either the $B^+ \to K^+ e^- \mu^+$ or $B^+ \to K^+ e^+ \mu^-$ selection applied to it.

Additional care is necessary for the application of the vetos against partially reconstructed background, as they are only applied to the opposite-sign charge configuration in $B^+ \to K^+ e^\pm \mu^\mp$ data. Just applying the veto criteria without this distinction does not reproduce the shape observed in $B^+ \to K^+ e^\pm \mu^\mp$ data sidebands, rendering the $B^+ \to K^+ e^+ \mu^+$ data useless as a proxy for the shape. To emulate the effect of the charge distinction on the shape, each candidate rejected by the vetos has a chance to still be accepted as having passed the selection according to the fraction of the charge configuration observed in $B^+ \to K^+ e^\pm \mu^\mp$ data. This adjusted veto application reproduces the shape seen in $B^+ \to K^+ e^\pm \mu^\mp$ data quite well on $B^+ \to K^+ e^+ \mu^+$ data, as shown in Figure 5.11, although some deviations remain specifically in the $B^+ \to K^+ e^+ \mu^-$ sample in Run 2, which are alleviated later in the selection and affect mainly the lower mass sideband.

After application of the whole selection, the $B^+ \to K^+ e^+ \mu^+$ data also contains very low statistics. Therefore, the impact of the different analysis steps on the shape of the upper mass sideband, where pure combinatorial can be compared in both $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ e^+ \mu^+$ data, is studied. While both, the BDT and BDTHOP are found to significantly shape the $Ke\mu$ invariant mass, no significant effect is found for the ProbNN based selection. Therefore, the shape can be extracted without applying the ProbNN based selection and stays valid also after the full selection. This behaviour is consistent between $B^+ \to K^+ e^\pm \mu^\mp$ and $B^+ \to K^+ e^+ \mu^+$ data in all subsets, making the $B^+ \to K^+ e^+ \mu^+$ data a good proxy for pure combinatorial background, which can be fitted using an exponential function.

This fit to the $B^+ \to K^+ e^+ \mu^+$ data is performed separately for each run period and charge configuration. The resulting fits are shown in Figure 5.12. Afterwards, the fit can be repeated on the $B^+ \to K^+ e^\pm \mu^\mp$ data, where the slope of the exponential is fixed to the values obtained from the fit to the $B^+ \to K^+ e^+ \mu^+$ data, while the yield is left floating. These fits are shown in Figure 5.13.
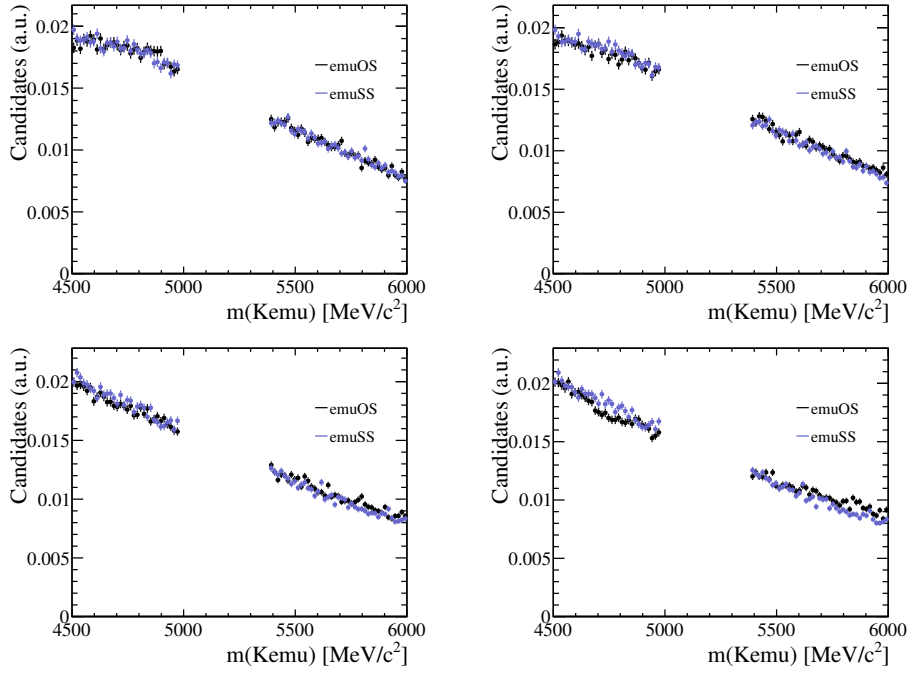
Figure 5.11: Comparison of the mass sidebands between $e\mu$ same-sign (blue) and opposite-sign (black) data after application of the vetos. The Run 1 data are shown on top, Run 2 data on the bottom. The $B^+ \to K^+ e^- \mu^+$ case is shown on the left, the $B^+ \to K^+ e^+ \mu^-$ case on the right.
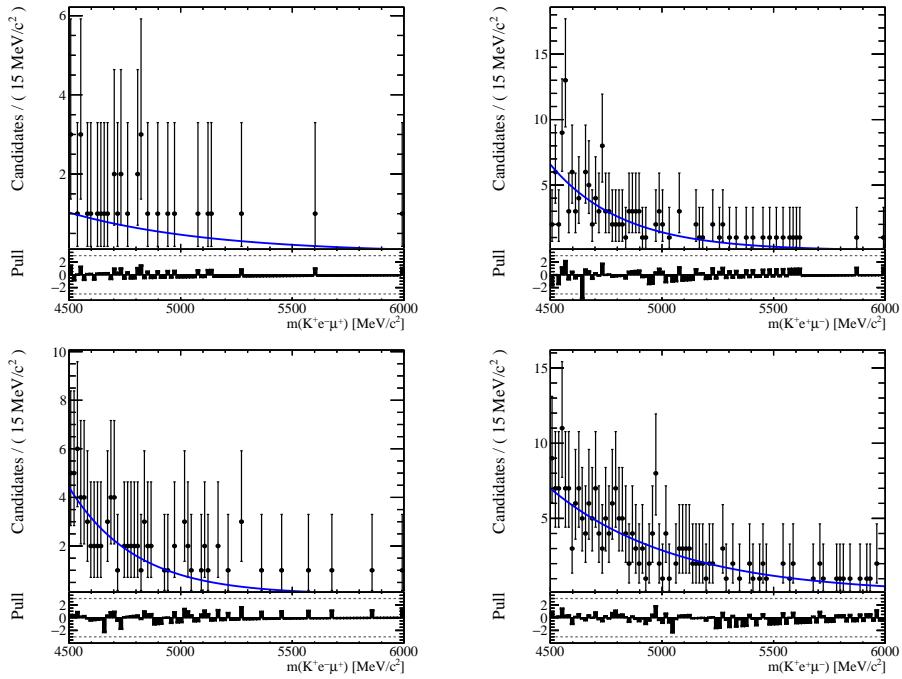
Figure 5.12: Fit to the B^+ mass after the full selection excluding the requirements on particle identification in $B^+ \to K^+ e^+ \mu^+$ data. The data is shown in black, while the background PDF is shown in blue. Run 1 is shown on the top, Run 2 at the bottom. While in all plots $B^+ \to K^+ e^+ \mu^+$ data is shown, the selection equivalent to $B^+ \to K^+ e^- \mu^+$ is shown on the left, the one equivalent to $B^+ \to K^+ e^+ \mu^-$ on the right.
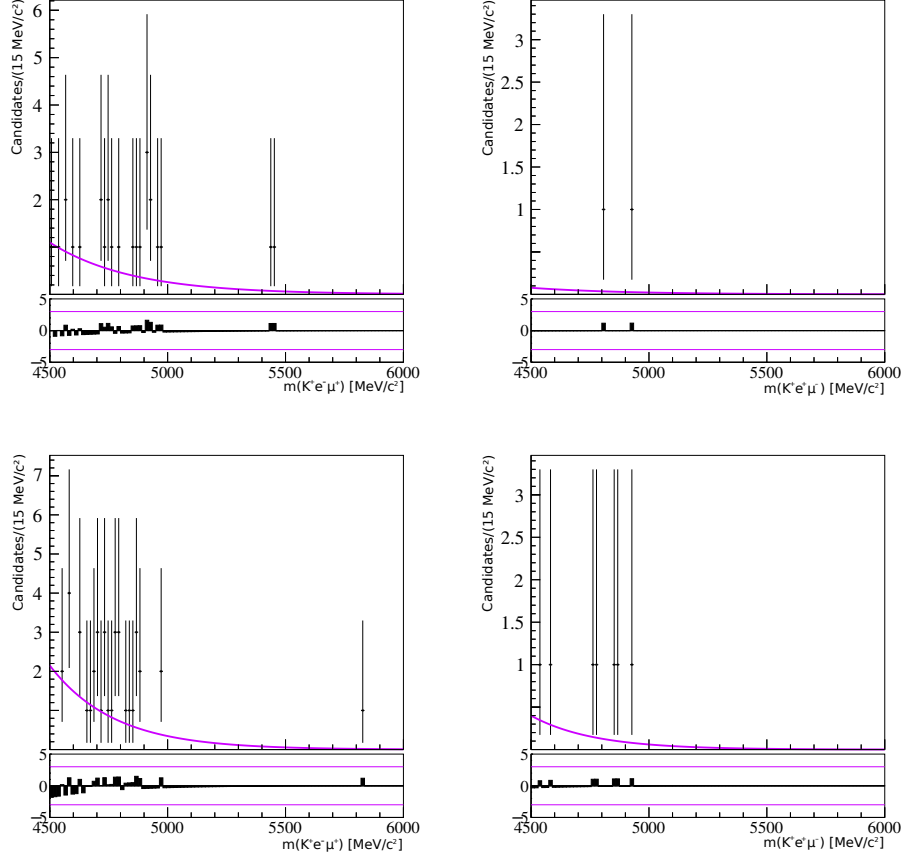
Figure 5.13: Fit to the blind B^+ mass after the full selection. The data is shown in black, while the background PDF is shown in purple. Run 1 is shown on the top, Run 2 at the bottom, with the $B^+ \to K^+ e^- \mu^+$ samples shown on the left and the $B^+ \to K^+ e^+ \mu^-$ samples on the right.

5.4.2 Signal parametrisation

For the calculation of the upper limit, it is necessary to describe the $B^+ \to K^+ e^\pm \mu^\mp$ signal and fix this shape in the fit to the generated toys. The shape is taken from simulation, with a model similar to the one used to fit the control mode $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ in Section 5.2.2. However, as in $B^+ \to K^+ e^\pm \mu^\mp$ there is an electron in the final state, the emission of bremsstrahlung needs to be taken into account. Therefore, the signal is parametrised by one Cruijff function and one Gaussian function per bremsstrahlung category. The resulting fits to $B^+ \to K^+ e^\pm \mu^\mp$

simulation are shown in Figure 5.14.
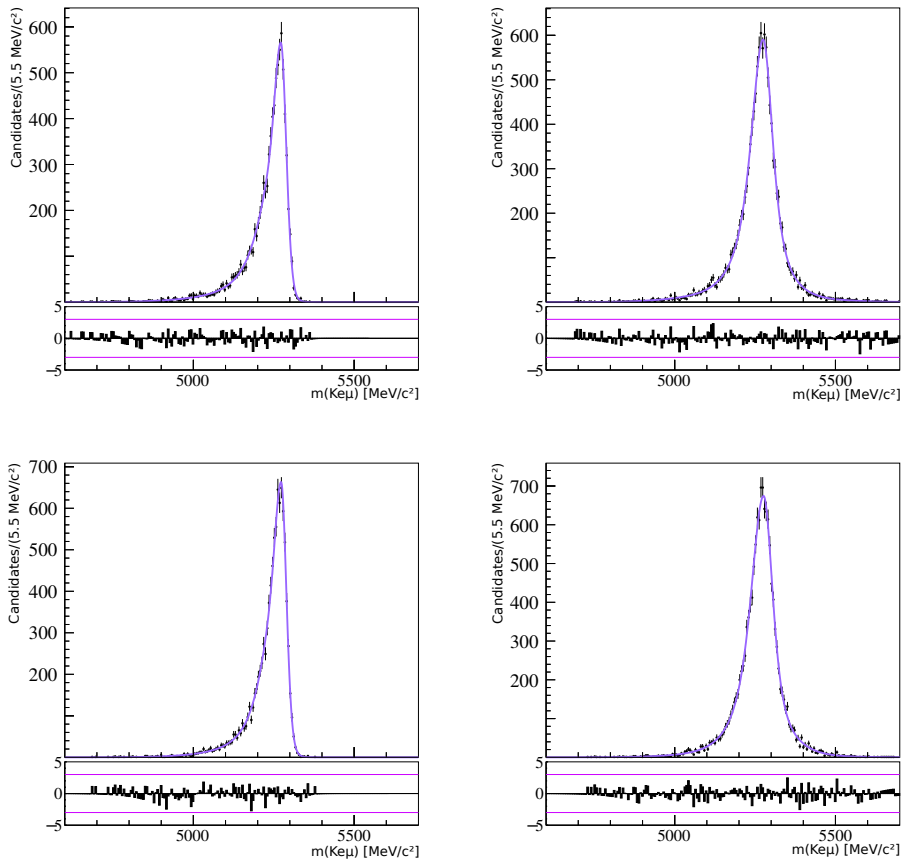
Figure 5.14: Fit to the $B^+ \to K^+ e^\pm \mu^\mp$ simulation after the full selection. The simulated data is shown in black, the fit model in purple. Run 1 is shown on the top, Run 2 at the bottom, with the case where the electron has no bremsstrahlung corrections applied on the left and the case where it has bremsstrahlung corrections applied on the right.

Similar to the search for $B^0_{(s)} \to e^+ e^-$, the resolution of the $B^+ \to K^+ e^\pm \mu^\mp$ invariant mass needs to be corrected before the limit is calculated to match a possible signal in data as close as possible. As the final state contains two leptons of different flavour, the correction can not simply be taken from one decay mode, but instead both $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ and $B^+ \to K^+ J/\psi (\to e^+ e^-)$ are needed. Thus the expression

for the correction factor for the width changes to

$$C = \frac{\sqrt{\left(\frac{\sigma}{\mu}\right)^2_{\text{ee, data}} + \left(\frac{\sigma}{\mu}\right)^2_{\text{mumu, data}}}}{\sqrt{\left(\frac{\sigma}{\mu}\right)^2_{\text{ee, MC}} + \left(\frac{\sigma}{\mu}\right)^2_{\text{mumu, MC}}}}, \tag{5.6}$$

where the parameters are obtained from a fit of the model described in Section 5.2.2 to the fully selected data. In case of $B^+ \to K^+ J/\psi(\to e^+ e^-)$, this fit is performed in two separate categories of bremsstrahlung correction, namely the cases where neither or both electrons have bremsstrahlung corrections applied to them. Correction factors are calculated separately for each bremsstrahlung category and are used to correct the two possible cases for the electron in $B^+ \to K^+ e^\pm \mu^\mp$. As the mixed case of only one electron in $B^+ \to K^+ J/\psi(\to e^+ e^-)$ having corrections applied to it has no counterpart in $B^+ \to K^+ e^\pm \mu^\mp$, it is not considered here.

Care needs to be taken in the definition of the width of the PDF used for the correction. In the fit model used there are three widths, the left and right width of the Cruijff function as well as the width of the gaussian. As these don't trivially sum up to one overall width, which also depends on the tails, the width of the PDF is instead obtained as the smallest region containing 68.3 % of the density. This approach also allows to take the full model uncertainty into account, instead of just the uncertainty of one fit parameter.

The corrections obtained are in line with what has been observed in the search for $B^0_{(s)} \to e^+ e^-$, with correction factors between 1.05 and 1.25. This is expected, as the corrections are dominated by the mismodelling of the electron mode, with a slight increase due to the effect from the muon mode. The correction factors are applied to each width of the signal model individually. The associated systematic uncertainties were found to be negligible.

5.4.3 Calculation of an expected upper limit

The expected upper limit is computed using the CL$_s$ method in the RooStats framework [152]. The p-value has been calculated in the Wald approximation using Equation (4.45). As a test statistic, the same one-sided test statistic as in Section 4.5 is used. The normalisation constants reported in Table 5.11 are gaussian constrained to their nominal values, with uncertainties consisting of their statistical uncertainties and the associated systematic uncertainties added in quadrature. The fit is performed simultaneously to the Run 1 and Run 2 data samples, where the branching fraction for $B^+ \to K^+ e^\pm \mu^\mp$ is shared between the samples. All parameters of the signal shape are fixed in the fit.

This approach leads to the CL$_s$ distribution shown in Figure 5.15 and the upper limits reported in Table 5.14.
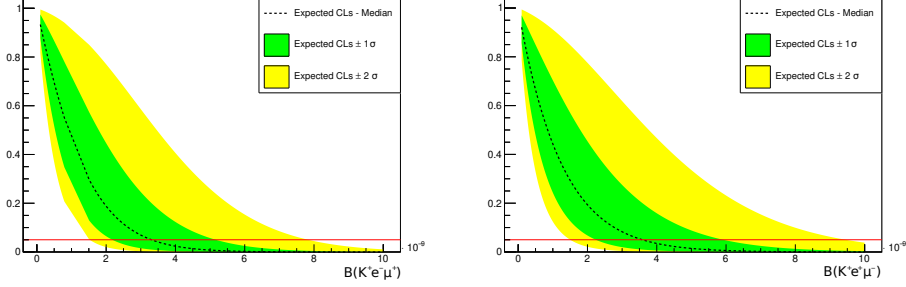
Figure 5.15: Distribution of CL_s as a function of the upper limit for the $B^+ \to K^+ e^- \mu^+$ case on the left and the $B^+ \to K^+ e^+ \mu^-$ case on the right.

Table 5.14: Expected upper limits on the branching fractions for $B^+ \to K^+ e^- \mu^+$ and $B^+ \to K^+ e^+ \mu^-$ at a confidence level of 95 %.

	expected limit
$\mathcal{B}(B^+ \to K^+ e^- \mu^+)$	$3.3 \cdot 10^{-9}$
$\mathcal{B}(B^+ \to K^+ e^+ \mu^-)$	$3.6 \cdot 10^{-9}$

As the search for $B^+ \to K^+ e^\pm \mu^\mp$ decays was already performed by the LHCb experiment on the Run 1 dataset, the limits shown in Table 5.14 can be compared to a projection to the full LHCb dataset based on the previous result [66]. From Equation (5.1) it can be seen that the size of the data sample indirectly enters at two points, namely the number of $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ candidates and the number of $B^+ \to K^+ e^\pm \mu^\mp$ candidates. However, the number of $B^+ \to K^+ e^\pm \mu^\mp$ candidates was found to be compatible with zero in the previous analysis (and is currently blinded in the one presented here). So instead upper limits are calculated, where the number of signal candidates (and thus limit) depend on the uncertainty of the background estimate in the signal region. Assuming the number of background candidates to follow a Poissonian distribution, this means the limit scales up with the square-root of the background in the signal region. Thus, assuming the amount of background to scale with the size of the dataset the same way as the signal and control modes do, with additional data the limit will scale proportional to the square-root of the size of the dataset and inversely proportional to the size of the dataset, meaning overall the limit can be expected to scale inversely proportional to the square-root of the size of the analysed dataset.

The full LHCb dataset corresponds to an integrated luminosity of $\approx 9\,\mathrm{fb}^{-1}$ compared to the $3\,\mathrm{fb}^{-1}$ of the Run 1 dataset. However, this does not reflect the increase in recorded B mesons, as the production cross-section of $b\bar{b}$ pairs increases with higher centre-of-mass energy. In the energy range the LHC operates in, this cross-section can be approximated to scale linearly with centre-of-mass energy. Because of this increase in $b\bar{b}$ production cross-section, the Run 2 dataset is not equivalent to twice the Run 1 dataset, as would be expected from just the increase in integrated luminosity. Instead, its size relative to the Run 1 dataset can be written as

$$\frac{6\,\mathrm{fb}^{-1} \cdot 13\,\mathrm{TeV}}{2\,\mathrm{fb}^{-1} \cdot 8\,\mathrm{TeV} + 1\,\mathrm{fb}^{-1} \cdot 7\,\mathrm{TeV}} \approx 3.4, \tag{5.7}$$

taking into account the change in centre-of-mass energy between 2011 and 2012, meaning the full LHCb dataset has a size about 4.4 times the size of the Run 1 dataset alone. The previous result, the projection to the full dataset and the expected upper limits from Table 5.14 are shown in Table 5.15.

Table 5.15: Upper limits on the branching fractions for $B^+ \to K^+ e^- \mu^+$ and $B^+ \to K^+ e^+ \mu^-$ at a confidence level of $95\,\%$ found in the previous search for $B^+ \to K^+ e^\pm \mu^\mp$. A projection to the full LHCb dataset as well as the expected upper limits found in the analysis presented here are also shown.

	previous upper limit
$\mathcal{B}(B^+ \to K^+ e^- \mu^+)$	$8.8 \cdot 10^{-9}$ @95 % CL
$\mathcal{B}(B^+ \to K^+ e^+ \mu^-)$	$9.5 \cdot 10^{-9}$ @95 % CL
	projected limit
$\mathcal{B}(B^+ \to K^+ e^- \mu^+)$	$4.2 \cdot 10^{-9}$ @95 % CL
$\mathcal{B}(B^+ \to K^+ e^+ \mu^-)$	$4.5 \cdot 10^{-9}$ @95 % CL
	expected limit updated analysis
$\mathcal{B}(B^+ \to K^+ e^- \mu^+)$	$3.3 \cdot 10^{-9}$ @95 % CL
$\mathcal{B}(B^+ \to K^+ e^+ \mu^-)$	$3.6 \cdot 10^{-9}$ @95 % CL

It can be seen that the expected upper limit found in the analysis presented here is lower than the projected expectation from an increase in the size of the analysed dataset. This suggests that the improvements that were made mainly in the MVA selection as well as performing the analysis split by charge configurations lead to an increase in sensitivity compared to the previous analysis.

6 Conclusion

In this thesis, two searches for rare decays of B mesons in the form of the searches for the rare $B^0_{(s)} \to e^+e^-$ and lepton-flavour violating $B^+ \to K^+ e^\pm \mu^\mp$ decays are presented.

The search for the rare decays $B^0_{(s)} \to e^+e^-$ has been performed on the data collected by the LHCb experiment during Run 1 in 2011 and 2012, and part of Run 2 in the years 2015 and 2016, corresponding to an integrated luminosity of $5\,\mathrm{fb}^{-1}$. This analysis has been published in [103] and set upper limits on the branching fractions of $B^0_{(s)} \to e^+e^-$ of

$$\mathcal{B}(B^0_s \to e^+e^-) < 9.4(11.2) \cdot 10^{-9} \text{ @ } 90(95)\,\% \text{ CL},$$
$$\mathcal{B}(B^0 \to e^+e^-) < 2.5(3.0) \cdot 10^{-9} \text{ @ } 90(95)\,\% \text{ CL},$$

improving the previous limits set by the CDF collaboration [55] by a factor of 30, reaching a region where the measurement becomes sensitive to potential New Physics contributions from universal scenarios including (pseudo-)scalar currents in $b \to s\ell^+\ell^-$ transitions.

The search for $B^+ \to K^+ e^\pm \mu^\mp$ decays has been performed on the full LHCb dataset corresponding to an integrated luminosity of $9\,\mathrm{fb}^{-1}$, effectively quadrupling the available B^+ mesons. While these decays are forbidden in the SM, and only occur with negligible rates when accounting for neutrino oscillations, their branching fractions can be enhanced up to values accessible with the current LHCb dataset in New Physics scenarios aimed at explaining anomalies seen in measurements of lepton-flavour universality.

At the time of writing, this search has not been published. However, expected upper limits on the branching fraction have been calculated as

$$\mathcal{B}(B^+ \to K^+ e^- \mu^+) < 3.3 \cdot 10^{-9} \text{ @ } 95\,\% \text{ CL},$$
$$\mathcal{B}(B^+ \to K^+ e^+ \mu^-) < 3.6 \cdot 10^{-9} \text{ @ } 95\,\% \text{ CL},$$

lowering them even further compared to the search performed on Run 1 data, which will allow to further constrain New Physics scenarios once published.

In the case of both analyses presented in this thesis, additional data will increase the possible precision, in turn further constraining potential effects from physics beyond the SM. With the LHC starting operations again in May 2022, this data is about to begin becoming available for physics analysis. In the case of the search for $B^0_{(s)} \to e^+ e^-$ however, a more direct route to increase the sensitivity is still available, as only part of the Run 2 data was used. A significant improvement in sensitivity can already be achieved from the inclusion of the remaining Run 2 data from 2017 and 2018, in total increasing the available data set by a factor of two. The inclusion of the 2017 and 2018 data has been studied in a master's thesis [153] together with an updated strategy for the BDT training, where the potential to lower the limit into the low $\mathcal{O}(10^{-9})$ range was found, which would allow significant constraints to potential New Physics in cases where the analysis presented here just started to impose constraints.

The arguably more exciting case than further constraints however would be the observation of $B^0_{(s)} \to e^+ e^-$ or $B^+ \to K^+ e^\pm \mu^\mp$, either of which would be an unambiguous sign for physics beyond the Standard Model, making new alternative theories not only an option, but rather a necessity.

A Appendix

A.1 Selections of the Search for the decays $B^0_{(s)} \to e^+ e^-$

Table A.1: Summary of stripping selection applied to the different particles.

Particle	Variable	$B^0_{(s)} \to e^+ e^-$	$B^+ \to K^+ J/\psi (\to e^+ e^-)$
$e^\pm$	islong	TRUE	TRUE
	χ^2_{IP}	> 25	> 25
	$\chi^2_{track}/\mathrm{ndf}$	< 3	< 3
	p_T	$> 250\,\mathrm{MeV}/c$	$> 250\,\mathrm{MeV}/c$
	$\mathrm{DLL}_{e,\pi}$	> -2	> 2
$B^0_{(s)}$ or J/ψ	$\chi^2_{vertex}/\mathrm{ndf}$	< 9	< 9
	$\|m_{e^+ e^-} - m_{pdg}\|$	$< 1200\,\mathrm{MeV}/c^2$	$< 1000\,\mathrm{MeV}/c^2$
	DOCA	$< 0.3\,\mathrm{mm}$	$< 0.3\,\mathrm{mm}$
	DIRA	> 0	> 0
	χ^2_{FD}	> 225	> 169
	χ^2_{IP}	< 25	-
B^+	χ^2_{vertex}	-	< 45
	$\|m_{K^+ e^+ e^-} - m_{pdg}\|$	-	$< 600\,\mathrm{MeV}/c^2$
	χ^2_{IP}	-	< 25
K^+	islong	-	TRUE
	$\chi^2_{track}/\mathrm{ndf}$	-	< 3
	χ^2_{IP}	-	> 25
	p_T	-	$> 250\,\mathrm{MeV}/c$
	p_{ghost}	-	< 0.3

Table A.2: Fiducial requirements applied to $B^0_{(s)} \to e^+e^-$ and $B^+ \to K^+ J/\psi (\to e^+e^-)$ for alignment with calibration data.

Particle	Variable	Run 1	Run 2
Event	nSPDHits	< 600	< 450
e	p_T	> 500 MeV/c	> 500 MeV/c
e	p	> 3000 MeV/c	> 3000 MeV/c
K	p	> 2000 MeV/c	> 2000 MeV/c
K	p_T	> 250 MeV/c	> 600 MeV/c
e, K	hasRich	TRUE	TRUE
e	hasCalo	TRUE	TRUE

Table A.3: Specific preselection applied to $B^+ \to K^+ J/\psi (\to e^+e^-)$.

Particle	Requirement
B^+	$4880\,\mathrm{MeV} < m_{B^+} < 5700\,\mathrm{MeV}$
B^+	$m_{B^+{}_\mathrm{constr}} > 5175\,\mathrm{MeV}$
J/ψ	$2450\,\mathrm{MeV} < m_{J/\psi} < 3176\,\mathrm{MeV}$
K^+	ProbNNk > 0.2

Table A.4: Summary of offline selection cuts applied prior to the training of the multivariate classifier.

Particle	cut
$e^\pm$	$\mathrm{prob}_\mathrm{ghost} < 0.3$
	$p_\mathrm{T} < 40$ GeV/c
	$p < 500$ GeV/c
B^0_s	$p_\mathrm{T} > 1$ GeV/c
	$\tau(B^0_s) < 9 \cdot \tau_\mathrm{true}(B^0_s)$

Table A.5: Binning schemes of the kinematic variables and multiplicities used for the resampling of the PID variables.

Variable	Bin borders
2011/2012	
$p\ (e)$	[0, 16000, 32000, 96000, 200000]
$\eta(e)$	[1.5, 2.66, 3.40, 3.89, 4.06, 5.5]
nTracks	[0, 88, 226, 500]
2015/2016	
$p_{\mathrm{T}}\ (e)$	[0, 800, 2400, 5600, 7200, 8800, 10400, 13600, 20000]
$\eta(e)$	[1.5, 2.19, 2.74, 3.23, 3.37, 3.55, 3.94, 4.14, 5.5]
nSPDHits	[0, 135, 292, 450]

Table A.6: Tightened ProbNNe cuts, used to suppress $B \to hh^{(\prime)}$ backgrounds. The cut-point differs as different tunings are used between Run 1 and Run 2 data.

Run	Selection
Run 1	MC12TuneV3 ProbNNe > 0.9
Run 2	MC15TuneV1 ProbNNe > 0.6

A.2 Results of the background studies in the Search for $B^0_{(s)} \to e^+ e^-$

Table A.7: Estimated number of exclusive background candidates in the $B^0_s \to e^+ e^-$ and $B^0 \to e^+ e^-$ signal regions for the Run 1 dataset. In the upper part of the table backgrounds from particle misID, that are peaking in the signal region, are reported. The decay $B^0 \to \pi^0 e^+ e^-$ is evaluated using the decay $B^+ \to \pi^+ e^+ e^-$ as a proxy. Branching fractions are taken from the Particle Data Group [107]. The first uncertainty comes from the limited amount of produced simulation, the second from the uncertainty on the branching fraction of the decay.

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

Table A.8: Estimated number of exclusive background candidates in the $B^0_s \to e^+e^-$ and $B^0 \to e^+e^-$ signal regions for the Run 1 dataset. In the upper part of the table backgrounds from particle misID, that are peaking in the signal region, are reported. The decay $B^0 \to \pi^0 e^+e^-$ is evaluated using the decay $B^+ \to \pi^+ e^+e^-$ as a proxy. Branching fractions are taken from the Particle Data Group [107]. The first uncertainty comes from the limited amount of produced simulation, the second from the uncertainty on the branching fraction of the decay. Since for some background contributions no Run 2 simulation was produced, for completeness the numbers from 2012 are included here, scaled by the mass region yield ratios of $B^+ \to K^+ e^+e^-$ (for $B^+/B^0 \to \pi^+/\pi^0 e^+e^-$) and $B^+ \to \pi^- e^+ \nu_e$ (for $\Lambda^0_b \to p e^- \bar{\nu}_e$).

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

A.3 Selections of the Search for the decays $B^+ \rightarrow K^+ e^\pm \mu^\mp$

Table A.9: Stripping selection applied to $B^+ \rightarrow K^+ e^\pm \mu^\mp$, $B^+ \rightarrow K^+ J/\psi (\rightarrow \mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$. The selections of the different modes are aligned per particle type. The requirement on nSPDHits is different between Run 1 (600) and Run 2 (450), to match the requirement in the L0 triggers.

Particle	Variable	Cut
B^+	$\|\|m_{K^+ e^\pm \mu^\mp} - m_{pdg}\|\|$	< 1500
	DIRA	> 0.995
	χ^2_{IP}	< 25
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf}$	< 9
	χ^2_{FD}	> 100
K	PT	> 400
	χ^2_{IP}	> 9
	$\mathrm{DLL}_{K,\pi}$	> -5
e	PT	> 350
	χ^2_{IP}	> 9
	$\mathrm{DLL}_{e,\pi}$	> 0
μ	PT	> 350
	χ^2_{IP}	> 9
	HasMuon	$== 1$
	IsMuon	$== 1$
$\ell^+ \ell^{-(\prime)}$ pair	PT	> 0
	M	< 5500
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf}$	< 9
	χ^2_{IP}	> 0
	χ^2_{FD}	> 16
Event	nSPDHits	$< 600(450)$

Table A.10: Cut-based preselection applied to all three channels. The specific version of the ProbNNk variable selected on depends on the run period, where in Run 1 MC12TuneV2 is used, while in Run 2 MC15TuneV1 is used.

Particle	Selection Run 1	Selection Run 2
K	InAccMuon == 1	InAccMuon == 1
	isMuonLoose == 0	isMuonLoose == 0
	ProbNNk > 0.2	ProbNNk > 0.2
	hasRich == 1	hasRich == 1
	$p > 2000$	$p > 2000$
	$p_{\mathrm{T}} > 250$	$p_{\mathrm{T}} > 600$
μ	PT > 800	PT > 800
e	P > 3000	P > 3000
	PT > 500	PT > 500
	isMuon == 0	isMuon == 0
	hasRich == 1	hasRich == 1
	hasCalo == 1	hasCalo == 1
$J/\psi\ (\to\mu^+\mu^-)$*	$3017 < \mathrm{m} < 3157$	$3017 < \mathrm{m} < 3157$
$J/\psi\ (\to e^+ e^-)$	$2600 < \mathrm{m} < 3300$	$2600 < \mathrm{m} < 3300$
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	$5080 < \mathrm{m} < 5700$	$5080 < \mathrm{m} < 5700$
$B^+ \to K^+ J/\psi(\to e^+ e^-)$	$4800 < \mathrm{m} < 5700$	$4800 < \mathrm{m} < 5700$
$B^+ \to K^+ J/\psi$*	> 5175	> 5175
Pair($p1$, $p2$)	$\theta(p1, p2) > 0.0005$	$\theta(p1, p2) > 0.0005$
Event	nSPDHits < 600	nSPDHits < 450

Table A.11: Vetos against physical backgrounds applied to $B^+ \to K^+ e^\pm \mu^\mp$. On the control modes, only the vetos against D-cascades are applied.

Background	Veto
D-cascades	$\mathrm{m}(K\mu) > 1885$
	$\mathrm{m}(Ke) > 1885$
Charmonia	$\mathrm{m}(K\mu)_{K\to\mu} < 3000 \parallel \mathrm{m}(K\mu)_{K\to\mu} > 3200$
	$\mathrm{m}(K\mu)_{K\to\mu} < 3630 \parallel \mathrm{m}(K\mu)_{K\to\mu} > 3740$
	$\mathrm{m}(Ke)_{K\to e} < 3000 \parallel \mathrm{m}(Ke)_{K\to e} > 3200$
	$\mathrm{m}(Ke)_{K\to e} < 3630 \parallel \mathrm{m}(Ke)_{K\to e} > 3740$
	$\mathrm{m}(\mu e)_{\mu\to e} < 2950 \parallel \mathrm{m}(\mu e)_{\mu\to e} > 3200$
	$\mathrm{m}(\mu e)_{\mu\to e} < 3630 \parallel \mathrm{m}(\mu e)_{\mu\to e} > 3740$
	$\mathrm{m}(\mu e)_{e\to\mu} < 3000 \parallel \mathrm{m}(\mu e)_{e\to\mu} > 3200$
	$\mathrm{m}(\mu e)_{e\to\mu} < 3630 \parallel \mathrm{m}(\mu e)_{e\to\mu} > 3740$

Table A.12: PID selection applied to $B^+ \to K^+ e^\pm \mu^\mp$ to suppress background from misidentification. MC12TuneV2 (V3 for electrons) is used in Run 1, MC15TuneV1 is used in Run 2 for all particles.

Particle	Selection	
	Run 1	Run 2
K	0.65	0.90
e	0.65	0.50
μ	0.70	0.95

Bibliography

[1] S. L. Glashow. "Partial-symmetries of weak interactions." In: *Nuclear Physics* 22.4 (1961), pp. 579–588.

[2] A. Salam and J. C. Ward. "Electromagnetic and weak interactions." In: *Physics Letters* 13.2 (1964), pp. 168–171.

[3] S. Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266.

[4] V. Trimble. "Existence and Nature of Dark Matter in the Universe." In: *Annual Review of Astronomy and Astrophysics* 25.1 (1987), pp. 425–472. DOI: `10.1146/annurev.aa.25.090187.002233`.

[5] Planck Collaboration. "Planck 2018 results - VIII. Gravitational lensing." In: *A&A* 641 (2020), A8. DOI: `10.1051/0004-6361/201833886`.

[6] A. J. Buras and J. Girrbach. "Towards the Identification of New Physics through Quark Flavour Violating Processes." In: *Rept. Prog. Phys.* 77 (2014), p. 086201. DOI: `10.1088/0034-4885/77/8/086201`. arXiv: `1306.3775`.

[7] R. Aaij et al. "Test of lepton universality in beauty-quark decays." In: *Nature Phys.* 18.3 (2022), pp. 277–282. DOI: `10.1038/s41567-021-01478-8`. arXiv: `2103.11769`.

[8] R. Aaij et al. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays." In: *JHEP* 08 (2017), p. 055. arXiv: `1705.05802`.

[9] R. Aaij et al. "Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays." In: *Journal of High Energy Physics* 2020.5 (May 2020), p. 40. ISSN: 1029-8479. DOI: `10.1007/JHEP05(2020)040`.

[10] R. Aaij et al. "Tests of Lepton Universality Using $B^0 \to K_S^0\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ Decays." In: *Phys. Rev. Lett.* 128 (19 May 2022), p. 191802. DOI: `10.1103/PhysRevLett.128.191802`.

[11] M. Algueró et al. "$b \to s\ell\ell$ Global Fits after R_{K_S} and $R_{K^{*+}}$." In: Apr. 2021. arXiv: `2104.08921`.

[12] S. L. Glashow, D. Guadagnoli, and K. Lane. "Lepton Flavor Violation in B Decays?" In: *Phys. Rev. Lett.* 114 (9 Mar. 2015), p. 091801. DOI: `10.1103/PhysRevLett.114.091801`.

[13] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos." In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567.

[14] Q. R. Ahmad et al. "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory." In: *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301.

[15] M. Aker et al. "Direct neutrino-mass measurement with sub-electronvolt sensitivity." In: *Nature Physics* 18.2 (Feb. 2022), pp. 160–166. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01463-1.

[16] J. Greensite. *An Introduction to the Confinement Problem*. Springer, 2011.

[17] S. K. Choi et al. "Observation of a resonance-like structure in the $\pi^{\pm}\psi'$ mass distribution in exclusive $B \to K\pi^{\pm}\psi'$ decays." In: *Phys. Rev. Lett.* 100 (2008), p. 142001.

[18] R. Aaij et al. "Observation of the Resonant Character of the $Z(4430)^-$ State." In: *Phys. Rev. Lett.* 112 (22 June 2014), p. 222002.

[19] R. Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays." In: *Phys. Rev. Lett.* 115 (7 Aug. 2015), p. 072001.

[20] B. Dey. "Overview on pentaquarks." In: *PoS* FPCP2017 (2017), p. 014. DOI: 10.22323/1.304.0014.

[21] N. Cabibbo. "Unitary Symmetry and Leptonic Decays." In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533.

[22] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction." In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.

[23] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge." In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247.

[24] B. Pontecorvo. "Mesonium and anti-mesonium." In: *Sov. Phys. JETP* 6 (1957), p. 429.

[25] Z. Maki, M. Nakagawa, and S. Sakata. "Remarks on the Unified Model of Elementary Particles." In: *Progress of Theoretical Physics* 28.5 (Nov. 1962), pp. 870–880. ISSN: 0033-068X. DOI: 10.1143/PTP.28.870.

[26] 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." In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[27] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Physics Letters B* 716.1 (2012), pp. 30–61.

[28]   P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509.

[29]   P. W. Higgs. "Broken symmetries, massless particles and gauge fields." In: *Physics Letters* 12.2 (1964), pp. 132–133.

[30]   F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323.

[31]   *Standard Model of Elementary Particles.* 2022. URL: https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg.

[32]   T. Blake, G. Lanfranchi, and D. M. Straub. "Rare B decays as tests of the Standard Model." In: *Progress in Particle and Nuclear Physics* 92 (2017), pp. 50–91. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2016.10.001.

[33]   G. Buchalla, A. J. Buras, and M. E. Lautenbacher. "Weak decays beyond leading logarithms." In: *Rev. Mod. Phys.* 68 (4 Oct. 1996), pp. 1125–1244. DOI: 10.1103/RevModPhys.68.1125.

[34]   R. Aaij et al. "Branching Fraction Measurements of the Rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2^{'}(1525)\mu^+\mu^-$ Decays." In: *Phys. Rev. Lett.* 127 (15 Oct. 2021), p. 151801. DOI: 10.1103/PhysRevLett.127.151801.

[35]   R. Aaij et al. "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^{*0}(892)\mu^+\mu^-$ differential branching fraction." In: *Journal of High Energy Physics* 2016.11 (Nov. 2016), p. 47. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2016)047. Erratum "Erratum to: Measurements of the S-wave fraction in B0 → K+π−μ+μ− decays and the B0 → K∗(892)0μ+μ− differential branching fraction." In: *Journal of High Energy Physics* 2017.4 (Apr. 2017), p. 142. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)142.

[36]   R. Aaij et al. "Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays." In: *Journal of High Energy Physics* 2015.6 (June 2015), p. 115. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2015)115. Erratum "Erratum to: Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays." In: *Journal of High Energy Physics* 2018.9 (Sept. 2018), p. 145. ISSN: 1029-8479. DOI: 10.1007/JHEP09(2018)145.

[37]   R. Aaij et al. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay." In: *Phys. Rev. Lett.* 125 (1 July 2020), p. 011802. DOI: 10.1103/PhysRevLett.125.011802.

[38]   R. Aaij et al. "Angular Analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay." In: *Phys. Rev. Lett.* 126 (16 Apr. 2021), p. 161802. DOI: 10.1103/PhysRevLett.126.161802.

[39] R. Aaij et al. "Angular analysis of the rare decay $B_s^0 \to \phi\mu^+\mu^-$." In: *Journal of High Energy Physics* 2021.11 (Nov. 2021), p. 43. ISSN: 1029-8479. DOI: `10.1007/JHEP11(2021)043`.

[40] M. Aaboud et al. "Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector." In: *Eur. Phys. J. C* 77.6 (2017), p. 367. DOI: `10.1140/epjc/s10052-017-4911-9`. arXiv: `1612.03016`.

[41] R. Aaij et al. "Measurement of $Z \to \tau^+\tau^-$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV." In: *JHEP* 09 (2018), p. 159. DOI: `10.1007/JHEP09(2018)159`. arXiv: `1806.05008`.

[42] P. Abreu et al. "Search for lepton flavor number violating Z0 decays." In: *Z. Phys. C* 73 (1997), pp. 243–251. DOI: `10.1007/s002880050313`.

[43] G. Aad et al. "Search for the lepton flavor violating decay Z $\to e\mu$ in pp collisions at $\sqrt{s}$ TeV with the ATLAS detector." In: *Phys. Rev. D* 90.7 (2014), p. 072010. DOI: `10.1103/PhysRevD.90.072010`. arXiv: `1408.5774`.

[44] M. Aaboud et al. "A search for lepton-flavor-violating decays of the Z boson into a τ-lepton and a light lepton with the ATLAS detector." In: *Phys. Rev. D* 98 (2018), p. 092010. DOI: `10.1103/PhysRevD.98.092010`. arXiv: `1804.09568`.

[45] V. M. Aulchenko et al. "Measurement of the ratio of the leptonic widths $\Gamma_{ee}/\Gamma_{\mu\mu}$ for the J/ψ meson." In: *Phys. Lett. B* 731 (2014), pp. 227–231. DOI: `10.1016/j.physletb.2014.02.046`. arXiv: `1311.5005`.

[46] M. Ablikim et al. "Search for the lepton flavor violation process $J/\psi \to e\mu$ at BESIII." In: *Phys. Rev. D* 87 (2013), p. 112007. DOI: `10.1103/PhysRevD.87.112007`. arXiv: `1304.3205`.

[47] M. Ablikim et al. "Search for the lepton flavor violation processes $J/\psi \to \mu\tau$ and $e\tau$." In: *Phys. Lett. B* 598 (2004), pp. 172–177. DOI: `10.1016/j.physletb.2004.08.005`. arXiv: `hep-ex/0406018`.

[48] W. Altmannshofer, P. Paradisi, and D. M. Straub. "Model-independent constraints on new physics in $b \to s$ transitions." In: *Journal of High Energy Physics* 2012.4 (Apr. 2012), p. 8. ISSN: 1029-8479. DOI: `10.1007/JHEP04(2012)008`.

[49] F. Beaujean et al. "Bayesian fit of exclusive $b \to s\bar{\ell}\ell$ decays: the standard model operator basis." In: *Journal of High Energy Physics* 2012.8 (Aug. 2012), p. 30. ISSN: 1029-8479. DOI: `10.1007/JHEP08(2012)030`.

[50] M. Beneke, C. Bobeth, and R. Szafron. "Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$." In: *Journal of High Energy Physics* 2019.10 (Oct. 2019), p. 232. ISSN: 1029-8479. DOI: `10.1007/JHEP10(2019)232`.

[51] J. Charles et al. "Current status of the standard model CKM fit and constraints on $\Delta F = 2$ new physics." In: *Phys. Rev. D* 91 (7 Apr. 2015), p. 073007. DOI: 10.1103/PhysRevD.91.073007.

[52] M. Bona et al. "Unitarity Triangle analysis in the Standard Model from the UTfit collaboration." In: *PoS* ICHEP2016 (2017), p. 554. DOI: 10.22323/1.282.0554.

[53] Y. Aoki et al. *FLAG Review 2021*. 2021. DOI: 10.48550/ARXIV.2111.09849.

[54] A. J. Buras. "Relations between $\Delta m_{s,d}$ and $B_{s,d} \to \mu^+\mu^-$ in models with minimal flavour violation." In: *Physics Letters B* 566.1 (2003), pp. 115–119. ISSN: 0370-2693. DOI: 10.1016/S0370-2693(03)00561-6.

[55] T. Aaltonen et al. "Search for the Decays $B^0_{(s)} \to e^+\mu^-$ and $B^0_{(s)} \to e^+e^-$ in CDF Run II." In: *Phys. Rev. Lett.* 102 (20 May 2009), p. 201801.

[56] L. J. Hall and M. B. Wise. "Flavor changing Higgs boson couplings." In: *Nuclear Physics B* 187.3 (1981), pp. 397–408.

[57] A. Datta et al. "New light mediators for the R_K and R_{K^*} puzzles." In: (2017). arXiv: 1705.08423.

[58] R. Fleischer, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi. "In Pursuit of New Physics with $B^0_{s,d} \to \ell^+\ell^-$." In: *JHEP* 05 (2017), p. 156. DOI: 10.1007/JHEP05(2017)156. arXiv: 1703.10160.

[59] M. C. Chang et al. "Search for $B^0 \to \ell^+\ell^-$ at BELLE." In: *Phys. Rev. D* 68 (2003), p. 111101. DOI: 10.1103/PhysRevD.68.111101. arXiv: hep-ex/0309069.

[60] S. Sandilya et al. "Search for the lepton-flavor-violating decay $B^0 \to K^{*0}\mu^\pm e^\mp$." In: *Phys. Rev. D* 98.7 (2018), p. 071101. DOI: 10.1103/PhysRevD.98.071101. arXiv: 1807.03267.

[61] B. Aubert et al. "Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$." In: *Phys. Rev. D* 73 (2006), p. 092001. DOI: 10.1103/PhysRevD.73.092001. arXiv: hep-ex/0604007.

[62] J. P. Lees et al. "A search for the decay modes $B^{+-} \to h^{+-}\tau^{+-}l$." In: *Phys. Rev. D* 86 (2012), p. 012004. DOI: 10.1103/PhysRevD.86.012004. arXiv: 1204.2852.

[63] B. Aubert et al. "Searches for the decays $B^0 \to \ell^\pm\tau^\mp$ and $B^+ \to \ell^+\nu$ (l=e, μ) using hadronic tag reconstruction." In: *Phys. Rev. D* 77 (2008), p. 091104. DOI: 10.1103/PhysRevD.77.091104. arXiv: 0801.0697.

[64] B. Aubert et al. "Search for decays of B^0 mesons into e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states." In: *Phys. Rev. D* 77 (2008), p. 032007. DOI: `10.1103/PhysRevD.77.032007`. arXiv: `0712.1516`.

[65] T. Aaltonen et al. "Search for the Decays $B^0_s \to e^+\mu^-$ and $B^0_s \to e^+e^-$ in CDF Run II." In: *Phys. Rev. Lett.* 102 (2009), p. 201801. DOI: `10.1103/PhysRevLett.102.201801`. arXiv: `0901.3803`.

[66] R. Aaij et al. "Search for Lepton-Flavor Violating Decays $B^+ \to K^+\mu^\pm e^\mp$." In: *Phys. Rev. Lett.* 123 (24 Dec. 2019), p. 241802. DOI: `10.1103/PhysRevLett.123.241802`.

[67] R. Aaij et al. "Search for the lepton-flavour violating decays $B^0_{(s)} \to e^\pm\mu^\mp$." In: *JHEP* 03 (2018), p. 078. DOI: `10.1007/JHEP03(2018)078`. arXiv: `1710.04111`.

[68] R. Aaij et al. "Search for the lepton-flavour-violating decays $B^0_s \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$." In: *Phys. Rev. Lett.* 123.21 (2019), p. 211801. DOI: `10.1103/PhysRevLett.123.211801`. arXiv: `1905.06614`.

[69] R. Aaij et al. "Search for the lepton flavour violating decay $B^+ \to K^+\mu^-\tau^+$ using B^{*0}_{s2} decays." In: *JHEP* 06 (2020), p. 129. DOI: `10.1007/JHEP06(2020)129`. arXiv: `2003.04352`.

[70] Roel Aaij et al. "Search for Majorana neutrinos in $B^- \to \pi^+\mu^-\mu^-$ decays." In: *Phys. Rev. Lett.* 112.13 (2014), p. 131802. DOI: `10.1103/PhysRevLett.112.131802`. arXiv: `1401.5361`.

[71] J. P. Lees et al. "Search for lepton-number violating processes in $B^+ \to h^-l^+l^+$ decays." In: *Phys. Rev. D* 85 (2012), p. 071103. DOI: `10.1103/PhysRevD.85.071103`. arXiv: `1202.3650`.

[72] R. Aaij et al. "Search for the lepton number violating decays $B^+ \to \pi^-\mu^+\mu^+$ and $B^+ \to K^-\mu^+\mu^+$." In: *Phys. Rev. Lett.* 108 (2012), p. 101601. DOI: `10.1103/PhysRevLett.108.101601`. arXiv: `1110.0730`.

[73] R. Aaij et al. "Searches for Majorana neutrinos in B^- decays." In: *Phys. Rev. D* 85 (2012), p. 112004. DOI: `10.1103/PhysRevD.85.112004`. arXiv: `1201.5600`.

[74] J. P. Lees et al. "Search for lepton-number violating $B^+ \to X^-l^+l^+$ decays." In: *Phys. Rev. D* 89.1 (2014), p. 011102. DOI: `10.1103/PhysRevD.89.011102`. arXiv: `1310.8238`.

[75] O. Seon et al. "Search for Lepton-number-violating $B^+ \to D^-l^+l'^+$ Decays." In: *Phys. Rev. D* 84 (2011), p. 071106. DOI: `10.1103/PhysRevD.84.071106`. arXiv: `1107.0642`.

[76] M. Bordone, G. Isidori, and A. Pattori. "On the standard model predictions for R_K and R_{K^*}." In: *The European Physical Journal C* 76.8 (Aug. 2016), p. 440. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4274-7.

[77] G. Isidori, S. Nabeebaccus, and R. Zwicky. "QED corrections in $\overline{B} \to \overline{K}\ell^+\ell^-$ at the double-differential level." In: *Journal of High Energy Physics* 2020.12 (Dec. 2020), p. 104. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2020)104.

[78] S. M. Boucenna, J. W. F. Valle, and A. Vicente. "Are the B decay anomalies related to neutrino oscillations?" In: *Physics Letters B* 750 (2015), pp. 367–371. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2015.09.040.

[79] G. Hiller, D. Loose, and K. Schönwald. "Leptoquark flavor patterns & B decay anomalies." In: *Journal of High Energy Physics* 2016.12 (Dec. 2016), p. 27. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2016)027.

[80] A. Crivellin et al. "Lepton-flavor violating B decays in generic $Z^{'}$ models." In: *Phys. Rev. D* 92 (5 Sept. 2015), p. 054013. DOI: 10.1103/PhysRevD.92.054013.

[81] A. Crivellin et al. "Correlating lepton flavor universality violation in B decays with $\mu \to e\gamma$ using leptoquarks." In: *Phys. Rev. D* 97 (1 Jan. 2018), p. 015019. DOI: 10.1103/PhysRevD.97.015019.

[82] L. Evans and P. Bryant. "LHC Machine." In: *JINST* 3 (2008), S08001.

[83] LHCb Speakers Bureau. Mar. 2022. URL: https://lhcb.web.cern.ch/lhcb/speakersbureau/html/PerformanceNumbers.html.

[84] LHCb Speakers Bureau. Mar. 2022. URL: http://cds.cern.ch/record/1087860/files/gene-2008-002_01.pdf?version=1.

[85] LHCb Speakers Bureau. Nov. 2017. URL: http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html.

[86] LHCb Speakers Bureau. Mar. 2022. URL: http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles/07_rad_acc_scheme_right.pdf.

[87] R. Aaij et al. "LHCb Detector Performance." In: *Int. J. Mod. Phys.* A30.07 (2015), p. 1530022.

[88] A. A. Alves et al. "The LHCb Detector at the LHC." In: *JINST* 3.LHCb-DP-2008-001. CERN-LHCb-DP-2008-001 (2008), S08005.

[89] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC." In: *The European Physical Journal C* 73.5 (May 2013), p. 2431. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2431-9.

[90] V. V. Gligorov and M. Williams. "Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree." In: *Journal of Instrumentation* 8.02 (2013), P02013.

[91] J. Albrecht, D. A. Berninghoff, and V. Gligorov. *Bremsstrahlung recovery of electrons using multivariate methods.* Tech. rep. LHCb-INT-2016-018. CERN-LHCb-INT-2016-018. CERN, Apr. 2016.

[92] The Gaudi Project. Nov. 2017. URL: http://gaudi.web.cern.ch/gaudi/.

[93] The Brunel Project. Nov. 2017. URL: http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/.

[94] R. Kalman. "A New Approach to Linear Filtering and Prediction Problems." In: *Transactions of the ASME - Journal of basic Engineering* 82 (Jan. 1960), pp. 35–45.

[95] The DaVinci Project. Nov. 2017. URL: http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/.

[96] W. D. Hulsbergen. "Decay chain fitting with a Kalman filter." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 552.3 (2005), pp. 566–575.

[97] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics.* 1868-4512. Springer Berlin, Heidelberg, 2010.

[98] M. Clemencic et al. "The LHCb Simulation Application, Gauss: Design, Evolution and Experience." In: *Journal of Physics: Conference Series* 331.3 (2011), p. 032023.

[99] T. Sjöstrand, S. Mrenna, and P. Skands. "A brief introduction to PYTHIA 8.1." In: *Computer Physics Communications* 178.11 (2008), pp. 852–867.

[100] D. J. Lange. "The EvtGen particle decay simulation package." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001), pp. 152–155.

[101] S. Agostinelli et al. "Geant4 - a simulation toolkit." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[102] The Boole Project. Nov. 2017. URL: http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/.

[103] R. Aaij et al. "Search for the Rare Decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$." In: *Phys. Rev. Lett.* 124 (21 May 2020), p. 211802. DOI: 10.1103/PhysRevLett.124.211802.

[104] N. Scharmberg. "Search for the rare decay $B_s^0 \to e^+e^-$ at the LHCb experiment." Master's thesis. Technische Universität Dortmund, September, 2016.

[105] A. Battig. "Search for the rare decay $B_s^0 \to e^+e^-$ with the LHCb experiment." Master's thesis. Technische Universität Dortmund, December, 2017.

[106] T. Mombaecher. "Beautiful leptons - setting limits to New Physics with the LHCb experiment." PhD thesis. Technische Universität Dortmund, September, 2020.

[107] J. Beringer et al. "Review of Particle Physics (RPP)." In: *Phys.Rev.* D86 (2012), p. 010001. DOI: `10.1103/PhysRevD.86.010001`.

[108] M. Pivk and F.R. Le Diberder. "Plots: A statistical tool to unfold data distributions." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1 (2005), pp. 356–369.

[109] B. P. Roe et al. "Boosted decision trees as an alternative to artificial neural networks for particle identification." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 543.2 (2005), pp. 577–584.

[110] T. Skwarnicki. "A study of the radiative CASCADE transitions between the Upsilon-prime and Upsilon resonances." DESY-F31-86-02. PhD thesis. Institute of Nuclear Physics, Krakow, 1986.

[111] A. Rogozhnikov. "Reweighting with Boosted Decision Trees." In: *Journal of Physics: Conference Series* 762 (Oct. 2016), p. 012036. ISSN: 1742-6596. DOI: `10.1088/1742-6596/762/1/012036`.

[112] J. H. Friedman. "Greedy function approximation: a gradient boosting machine." In: *Annals of statistics* (2001), pp. 1189–1232.

[113] A. Abulencia et al. "Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II." In: *Phys. Rev. Lett.* 95 (2005). [Erratum: Phys. Rev. Lett.95,249905(2005)], p. 221805. DOI: `10.1103/PhysRevLett.95.221805`. arXiv: `hep-ex/0508036`.

[114] L. Anderlini et al. *The PIDCalib package.* Tech. rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021. Geneva: CERN, July 2016. URL: `https://cds.cern.ch/record/2202412`.

[115] *The Urania Project.* `https://gitlab.cern.ch/lhcb/Urania`. May 2022.

[116] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python.* 2018. arXiv: `1201.0490`. URL: `http://scikit-learn.org/stable/`.

[117] Y. Freund and R. E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139.

[118] B. P. Roe et al. "Boosted decision trees as an alternative to artificial neural networks for particle identification." In: *Nucl.Instrum.Meth.* A543 (2005), pp. 577–584. DOI: 10.1016/j.nima.2004.12.018. arXiv: physics/0408124.

[119] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2009.

[120] L. Gavardi. "Search for lepton flavour violation in τ decays at the LHCb experiment." Presented 28 Nov 2013. Nov. 2013. URL: https://cds.cern.ch/record/1645251.

[121] G. Punzi. "Sensitivity of searches for new signals and its optimization." In: *Statistical Problems in Particle Physics, Astrophysics, and Cosmology.* Ed. by L. Lyons, R. Mount, and R. Reitmeyer. 2003, p. 79. arXiv: physics/0308063.

[122] S. Tolk et al. *Data-driven measurement of trigger efficiencies with the TisTos method.* Tech. rep. CERN-LHCb-INT-2013-038. Geneva: CERN, June 2013. URL: https://cds.cern.ch/record/1557354.

[123] R. Aaij et al. "Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics." In: *JHEP* 04 (2013). f_s/f_d value updated in LHCb-CONF-2013-011, p. 001. DOI: 10.1007/JHEP04(2013)001. arXiv: 1301.5286.

[124] R. Aaij et al. "Measurement of f_s/f_u variation with proton-proton collision energy and kinematics." In: (2019). submitted to Phys.Rev.Lett. arXiv: 1910.09934.

[125] The LHCb collaboration. "Measurement of the track reconstruction efficiency at LHCb." In: *Journal of Instrumentation* 10.02 (Feb. 2015), P02007–P02007. DOI: 10.1088/1748-0221/10/02/p02007.

[126] B. Efron. "Bootstrap Methods: Another Look at the Jackknife." In: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552.

[127] R. Aaij et al. "Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV." In: *Eur. Phys. J.* C71 (2011), p. 1645. DOI: 10.1140/epjc/s10052-011-1645-y. arXiv: 1103.0423.

[128] R. Aaij et al. "Production of J/ψ and Upsilon mesons in pp collisions at $\sqrt{(s)} = 8$ TeV." In: *JHEP* 06 (2013), p. 064. DOI: 10.1007/JHEP06(2013)064. arXiv: 1304.6977.

[129] R. Aaij et al. "Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 10 (2015). [Erratum: JHEP05,063(2017)], p. 172. DOI: `10.1007/JHEP05(2017)063,10.1007/JHEP10(2015)172`. arXiv: `1509.00771`.

[130] Y. Amhis et al. "Averages of *b*-hadron, *c*-hadron, and τ-lepton properties as of summer 2016." In: (2016). arXiv: `1612.07233`.

[131] A. Oyanguren. "B decay anomalies at LHCb." In: *EPJ Web of Conferences* 175 (Jan. 2018), p. 01004. DOI: `10.1051/epjconf/201817501004`.

[132] R. Aaij et al. "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays." In: *Phys. Rev. Lett.* 118.19 (2017), p. 191801. DOI: `10.1103/PhysRevLett.118.191801`. arXiv: `1703.05747`.

[133] J. M. Flynn et al. "$B \to \pi\ell\nu$ and $B_s \to K\ell\nu$ form factors and $|V_{ub}|$ from 2+1-flavor lattice QCD with domain-wall light quarks and relativistic heavy quarks." In: *Phys. Rev.* D91.7 (2015), p. 074510. DOI: `10.1103/PhysRevD.91.074510`. arXiv: `1501.05373`.

[134] C. M. Bouchard et al. "$B_s \to K\ell\nu$ form factors from lattice QCD." In: *Phys. Rev.* D90 (2014), p. 054506. DOI: `10.1103/PhysRevD.90.054506`. arXiv: `1406.2279`.

[135] W.-F. Wang and Z.-J. Xiao. "The semileptonic decays $B/B_s \to (\pi, K)(\ell^+\ell^-, \ell\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order." In: *Phys. Rev.* D86 (2012), p. 114025. DOI: `10.1103/PhysRevD.86.114025`. arXiv: `1207.0265`.

[136] A. Kozachuk, D. Melikhov, and N. Nikitin. "Rare FCNC radiative leptonic $B_{s,d} \to \gamma l^+l^-$ decays in the standard model." In: *Phys. Rev.* D97.5 (2018), p. 053007. DOI: `10.1103/PhysRevD.97.053007`. arXiv: `1712.07926`.

[137] P. Ball and R. Zwicky. "$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited." In: *Phys. Rev.* D71 (2005), p. 014029. DOI: `10.1103/PhysRevD.71.014029`. arXiv: `hep-ph/0412079`.

[138] V. V. Kiselev. "Exclusive decays and lifetime of B_c meson in QCD sum rules." In: (2002). arXiv: `hep-ph/0211021`.

[139] G. A. Cowan, D. C. Craik, and M. D. Needham. "RapidSim: an application for the fast simulation of heavy-quark hadron decays." In: *Comput. Phys. Commun.* 214 (2017), pp. 239–246. DOI: `10.1016/j.cpc.2017.01.029`. arXiv: `1612.07489`.

[140] D. Scora and N. Isgur. "Semileptonic meson decays in the quark model: An update." In: *Phys. Rev.* D52 (1995), pp. 2783–2812. DOI: `10.1103/PhysRevD.52.2783`. arXiv: `hep-ph/9503486`.

[141] A. Khodjamirian and A. V. Rusov. "$B_s \to K\ell\nu_\ell$ and $B_{(s)} \to \pi(K)\ell^+\ell^-$ decays at large recoil and CKM matrix elements." In: *JHEP* 08 (2017), p. 112. DOI: `10.1007/JHEP08(2017)112`. arXiv: `1703.04765`.

[142] A. L. Read. "Presentation of search results: The CL(s) technique." In: *J. Phys.* G28 (2002), pp. 2693–2704. DOI: `10.1088/0954-3899/28/10/313`.

[143] M. W. Kenzie et al. *GammaCombo - a statistical analysis framework.* 2022. URL: `https://gammacombo.github.io/`.

[144] A. Wald. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large." In: *Transactions of the American Mathematical Society* 54.3 (1943), pp. 426–482. ISSN: 00029947. DOI: `10.2307/1990256`.

[145] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics." In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: `10.1140/epjc/s10052-011-1554-0`. arXiv: `1007.1727`.

[146] A. K. Alok et al. "New physics in $b \to se^+e^-$: A model independent analysis." In: *Nuclear Physics B* 967 (2021), p. 115419. ISSN: 0550-3213. DOI: `10.1016/j.nuclphysb.2021.115419`.

[147] F. Koch. "Analyse des leptonzahlverletzenden Zerfalls $B^+ \to K^+e^\pm\mu^\mp$ mit Run 2 Daten des LHCb-Experiments." Bachelor thesis. Technische Universität Dortmund, December, 2019.

[148] J. Wendel. "Suche nach Leptonflavourzahlverletzung in $B^+ \to K^+e^\pm\mu^\mp$: Studien zur multivariaten Selektion." Bachelor thesis. Technische Universität Dortmund, January, 2020.

[149] P. del Amo Sanchez et al. "Study of $B \to X\gamma$ decays and determination of $|V_{td}/V_{ts}|$." In: *Phys. Rev. D* 82 (5 Sept. 2010), p. 051101. DOI: `10.1103/PhysRevD.82.051101`.

[150] LHCb Collaboration et al. *Constraints on the CKM angle γ from $B^\pm \to Dh^\pm$ decays using $D \to h^\pm h'^\mp \pi^0$ final states.* 2021. DOI: `10.48550/ARXIV.2112.10617`.

[151] M.-H. Schune, F. Polci, and M. Borsato. *HOP an additional tool for decays involving electrons.* Tech. rep. LHCb-INT-2015-037. CERN-LHCb-INT-2015-037. CERN, Nov. 2015.

[152] L. Moneta et al. *The RooStats Project.* 2010. DOI: `10.48550/ARXIV.1009.1003`.

[153] F. Koch. "Search for the very rare decay $B_s^0 \to e^+e^-$ with the LHCb experiment." Master's thesis. Technische Universität Dortmund, May, 2022.