The essence of rare beauty

Mulder, Mick

DOI:
10.33612/diss.149618058

The essence of rare beauty

Studying $B^0_{(s)} \to \mu^+\mu^-$ decays with the LHCb experiment

Mick Mulder

Cover: Puzzle pieces depicting a pp collision in the LHCb detector from June 2016, containing one of the most signal-like $B_s^0 \to \mu^+\mu^-$ candidates.

First edition.

This work is part of the research programme of the Netherlands Organisation for Scientific Research (NWO). The work is carried out at the National Institute of Subatomic Physics (Nikhef) in Amsterdam, The Netherlands.

rijksuniversiteit
groningen

The essence of rare beauty

Studying $B^0_{(s)} \to \mu^+\mu^-$ decays with the LHCb experiment

Proefschrift

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 15 januari 2021 om 14:30 uur

door

Mick Mulder

geboren op 17 januari 1993
te Haarlem

Promotores

Prof. dr. A. Pellegrino

Prof. dr. M.H.M. Merk

Copromotor

Dr. ir. C.J.G. Onderwater

Beoordelingscommissie

Prof. dr. P.J.G. Mulders

Prof. dr. ir. P.J. de Jong

Prof. dr. S. Hoekstra

Contents

Chapter 1

Introduction

From our childhood, we start asking questions about our world: why is the sky blue? what governs the cycle of the sun, the moon and the stars? why does it rain and what governs our weather? what are we and our world made of? For most of our history, these questions were answered by stories and myths, until we started to find answers to most of these questions using physics.

By measuring and experimenting, it was found that Nature behaves in predictable patterns and that laws can be defined that describe and explain those patterns, such as the laws of motion, gravity and electromagnetism. At the end of the 19th century, it seemed like these laws were sufficient to describe all of Nature, until certain aspects could not be explained by classical physics. The discovered structure of atoms, namely that of a dense, positively charged core and electrons orbiting around it, should be unstable in classical physics: electrons would lose electromagnetic radiation continously and fall into the core. Additionally, it was found that interactions of light with material, the so-called photo-electric effect, showed a threshold effect: a minimum photon energy was required to induce these interactions. Such an effect was unexplainable by classical physics, which predicted a continuous relation between energy and such interactions. Another issue was the UV problem: the prediction of classical physics of black body radiation included an arbitrarily high amount of energy in the UV spectrum, in which case all matter would quickly radiate all their energy. The resolution of these issues over the early 20th century resulted in a revolution in physics, culminating in the introduction of quantum physics. An essential component of this era was the discovery of fundamental constituents of the Universe, such as the electron in 1897 [1], the proton in 1917 [2], the anti-electron or positron in 1932 [3], and the neutron in the same year [4]. This started the field of elementary particle physics, which aims to describe the fundamental constituents of our world and their interactions, using the simplest possible set of physical principles.

The present knowledge on particle physics is described by a theory that is called the Standard Model (SM). The Standard Model describes the behaviour of all known fundamental particles and was completed with the discovery of a Higgs-like particle in July 2012: every particle that it predicts has been discovered and it has passed many experimental tests.

Still, there are many open questions about the Standard Model. It contains 26 fundamental constants, which are not determined by theory and have to be experimentally determined. Most of these parameters, 20 to be precise, are associated with the structure of matter, namely their masses (m_i, one per fermion, 12 in total) and mixing angles and phases (3 angles θ_{ij} and 1 complex phase δ, for quarks and for leptons, 8 in total). The other 6 are the coupling constants of the three interactions $\alpha_{\text{QED}}, \alpha_s, \alpha_{\text{weak}}$, the Higgs vacuum expectation value v and mass m_{H^0}, and the size of matter-antimatter differences in the strong interaction, θ_{QCD}. While there are only four types of matter particles (up- and down-type quarks, charged and neutral leptons), the Standard Model includes three "copies" of each matter particle, organised in three generations that interact similarly to each other. These particles show unexplained hierarchies and structures:

- Why are there three generations of matter particles? Could there be more? Curiously, at least three generations are needed to implement a difference between matter and anti-matter in the SM. Such a difference is required to generate our universe, which consists predominantly of matter.
- Why do the quark and charged lepton masses have a hierarchical structure, and do the masses of particles in separate generations differ so strongly? Why do neutrinos have much lower masses than the other matter particles and do those masses show less of a hierarchical structure?
- Why do the mass eigenstates of quarks coincide strongly with their weak interaction eigenstates, while this is not the case for leptons? Could this be related to the mass hierarchy?

Together, these questions are referred to as the 'flavour puzzle' of the Standard Model.

A promising approach to study the flavour puzzle is to investigate the interactions of matter from the third and heaviest generation. Their interactions have not been tested as much as those of the other two generations, and they might couple more strongly to New Physics contributions, for example if the flavour structure of New Physics is related to the size of the quark masses in the Standard Model.

Because of the uncertainty principle from quantum physics, particle decays are affected by virtual particles more massive than the decaying particles themselves. As a result, heavy new particles can be discovered by precision studies of the decays of lighter particles. Especially interesting are processes that are strongly suppressed in the Standard Model, as they may occur at tree level with new particles, while they can only occur through higher order quantum loop diagrams in the SM. Therefore, even small contributions from new massive quantum fields have relatively large effects, allowing highly suppressed processes to probe the effects of new particles up to mass scales far beyond the LHC collision energies.

The most famous example of a quantum loop process is the decay $B^0_{(s)} \to \mu^+\mu^-$, which is very strongly suppressed: in the Standard Model, it is predicted to occur around three times in every billion B meson decays. Its strong suppression, combined with the numerous production of B mesons at the LHC, and the precision of the theoretical predictions make measurements of $B^0_{(s)} \to \mu^+\mu^-$ decays stringent precision tests of the Standard Model. This thesis describes the search for $B^0_{(s)} \to \mu^+\mu^-$ decays at the LHCb detector, concluding with the first single-experiment observation of the $B^0_s \to \mu^+\mu^-$ decay and the world's strongest limit on the $B^0 \to \mu^+\mu^-$ decay.

The theoretical context for the study of $B^0_{(s)} \to \mu^+\mu^-$ decays is discussed in Chapter 2. The Standard Model is introduced, including the concept of Flavour Changing Neutral Currents, the strongly suppressed transitions that mediate $B^0_{(s)} \to \mu^+\mu^-$ decays in the Standard Model. Its amplitude is derived in detail and the Standard Model predictions for $B^0_{(s)} \to \mu^+\mu^-$ observables are given. Finally, the predictions are compared with the experimental measurements performed before the one described in this thesis.

To investigate $B^0_{(s)} \to \mu^+\mu^-$ decays, a source of $B^0_{(s)}$ mesons is needed. Due to their short lifetime, they have to be produced at a particle accelerator facility, such as the Large Hadron Collider (LHC), and studied with a particle detector, such as the LHCb detector, located at CERN. These are described in Chapter 3. Up to 2018, around 5.7×10^{11} $B^0_{(s)}$ mesons have been produced within the LHCb detector acceptance by the LHC, which makes it possible to study even very rare decays such as $B^0_{(s)} \to \mu^+\mu^-$. The subdetectors that can identify $B^0_{(s)}$ mesons and measure their lifetime are presented.

An analysis to search for $B^0_{(s)} \to \mu^+\mu^-$ decays, using data from the LHCb detector collected up to September 2016, is described in Chapter 4. It uses the unique signature of $B^0_{(s)} \to \mu^+\mu^-$ decays, namely a displaced vertex decaying to two muons with a dimuon mass around the $B^0_{(s)}$ masses, a machine learning algorithm to achieve

a strong separation of signal and background, and a normalisation of signal yield to a known mode, to determine the branching fraction of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$.

The results of the search for $B_{(s)}^0 \to \mu^+\mu^-$ decays are presented in Chapter 5. The world's first single experiment observation of the $B_s^0 \to \mu^+\mu^-$ decay is made and the world's strongest limit is set on the $B^0 \to \mu^+\mu^-$ decay. Additionally, the measurement of the effective lifetime of $B_s^0 \to \mu^+\mu^-$ decays using the same dataset is described, including its results.

In Chapter 6, the current knowledge on $B_{(s)}^0 \to \mu^+\mu^-$ decays is summarised, including recent measurements by ATLAS and CMS. The $B_{(s)}^0 \to \mu^+\mu^-$ decays are related to other measurements of the same quark process, and it is shown how these measurements hint at contributions of New Particles. Finally, an outlook to the future precision of $B_{(s)}^0 \to \mu^+\mu^-$ measurements at LHCb is presented, showing which milestones those measurements will achieve and how they will contribute to our knowledge of $B_{(s)}^0 \to \mu^+\mu^-$ decays.

Chapter 2

Theory

In this chapter, the fundamental concepts of the Standard Model (SM) and quantum loop processes such as $B^0_{(s)} \to \mu^+\mu^-$ decays are introduced. It is discussed why these decays are highly suppressed in the SM and how they probe New Physics at high mass scales. The $B^0_{(s)} \to \mu^+\mu^-$ amplitude is derived, its observables are introduced, and the SM predictions are discussed and compared with the experimental measurements performed before the analysis described in this thesis.

2.1 Standard Model

The Standard Model of particle physics is a quantum field theory that describes all known elementary particles and their interactions. A fundamental part of the SM is the concept of *flavour*, a quantum number which distinguishes the twelve particles that form matter. Interestingly, the twelve matter particles are divided into three so-called generations, which each contain an up- and down-type quark, a charged lepton and a neutral lepton, called a neutrino. While the electromagnetic and strong interactions couple matter to anti-matter of the same flavour, thus conserving flavour, the weak interaction does not, which allows it to probe the structure of matter.

Studies of the weak interaction have been essential to form the SM; for example, the beta decay of nuclear particles was used in the 1930s to propose the existence of the electron neutrino, and the indirect observation of the charm quark was proposed by Glashow, Iliopoulos and Maiani to explain the non-observation of the quantum loop decay of a kaon to two muons through a so-called Flavour Changing Neutral Current(FCNC) [5].

The interactions of particles in the SM are described through three imposed gauge symmetries, which describe the behaviour of the fundamental interactions:

$$\begin{aligned} \text{SM} = &\ \text{SU(3)}_{\text{Strong}} \otimes (\text{SU(2)} \times \text{U(1)})_{\text{Electroweak}}, \\ \xrightarrow{\text{SSB}} &\ \text{SU(3)}_{\text{Strong}} \otimes \text{U(1)}_{\text{EM}}. \end{aligned} \tag{2.1}$$

After strong symmetry breaking (SSB), specifically the Higgs mechanism, which is described in more detail in Ref. [6, 7], only the strong and electromagnetic interactions remain.

The strong interaction, often referrred to as Quantum Chromodynamics (QCD), is the strongest interaction at the energies that have been accessed in experiment up to now. It is the interaction that binds quarks together in hadrons, either as a quark-antiquark pair, which is called a meson, such as the neutral B mesons B_s^0 ($\bar{b}s$) and B^0 ($\bar{b}d$), or as a triplet of quarks, which is called a baryon, such as the proton (uud) and neutron (udd). In both cases, the colour charges can be combined to form a colour-neutral state. The strong interaction is mediated by massless vector bosons called gluons, which act on the so-called colour quantum number.

The weak interaction was discovered by studying the beta decay of nuclear particles, and is the weakest force out of the three. The weak interaction is the only interaction that affects all SM fermions. Because it is mediated by massive heavy bosons called $W^{\pm}$ and Z^0 bosons, it is the weakest interaction compared to others. It acts on the quantum number called weak isospin. Because only left-handed particles and right-handed anti-particles have a non-zero weak isospin, the weak interaction maximally violates parity, as discovered by Wu [8].

The electromagnetic interaction is mediated by massless particles called photons, which are denoted by γ, and which interact with electrically charged particles. The quantum field theory description of the electromagnetic interaction is commonly referred to as Quantum Electrodynamics(QED).

The Higgs boson is essential to the Standard Model. By interacting with all other particles (except possibly neutrinos), it confers a mass to them. The Higgs mechanism is described in more detail in Ref. [6, 7]. For gauge bosons, the interaction term originates directly from the covariant derivative in the kinematic term of the Higgs, which gives masses to the $W^{\pm}$ and Z^0 bosons, while leaving the γ massless. For the matter particles, an extra term is included in the Lagrangian in order to introduce this interaction; it is referred to as the Yukawa term. Importantly, in the SM the weak interaction does not conserve flavour because the Yukawa couplings

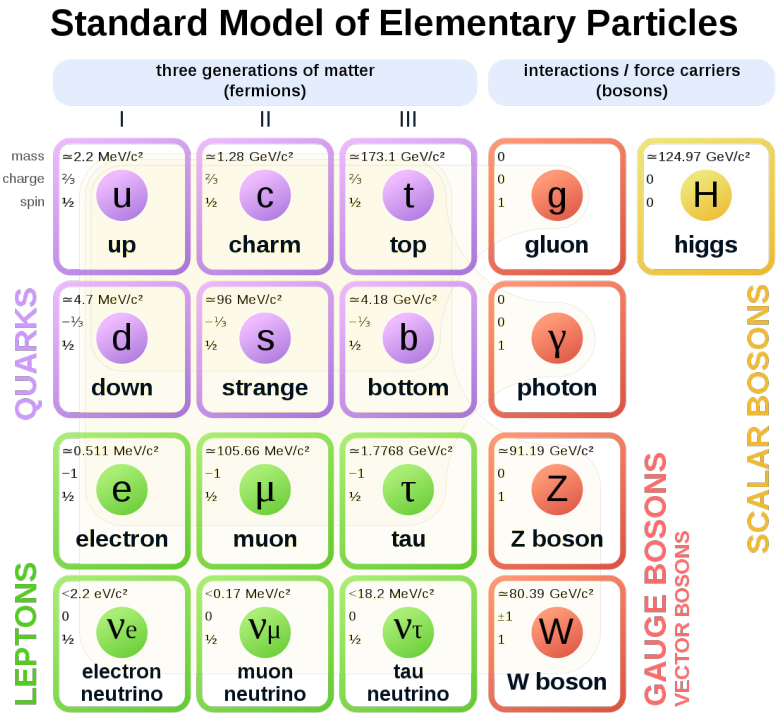

Figure 2.1: The particle content of the SM. On the left are the matter particles, consisting of quarks and leptons, divided over three "generations", with the mass increasing for each generation. On the right are the gauge bosons and the Higgs particle. An outline indicates which matter particles interact with each gauge boson. The Higgs boson interacts with all of the particles but the photon and gluon, giving a mass to those particles it interacts with.

and weak interaction coupling operate in different bases. As such, studying flavour changing transitions probes the Yukawa couplings.

The full particle content of the SM and their interactions are shown in Figure 2.1.

2.1.1 The weak interaction and flavour changing currents

In the Standard Model, the weak interaction acts on all left-handed fermion fields in the weak interaction basis. These are denoted as $u^I_{i,L}, d^I_{i,L}$ for up and down-type quarks respectively, indicating the generation by i and including the label I to indicate these fields are eigenstates of the weak interaction. The $W^\pm$ bosons moderate the flavour changing terms of the weak interaction by coupling down-type quarks to

up-type anti-quarks and vice versa, denoted as

$$\mathcal{L}_{\text{weak flavour changing}} = -\frac{g}{\sqrt{2}}\overline{u^I_{i,L}}\gamma_\mu W^{-\mu} d^I_{i,L} - \frac{g}{\sqrt{2}}\overline{d^I_{i,L}}\gamma_\mu W^{+\mu} u^I_{i,L}, \tag{2.2}$$

where g is the coupling strength of the weak interaction.

The masses of the quarks are caused by their interactions with the Higgs particle, which are described by the so-called Yukawa term. After spontaneous symmetry breaking, the Yukawa terms lead to mass terms for the quarks of the form

$$\mathcal{L}_{\text{Yukawa}} = -\overline{u^I_{i,L}} Y^u_{ij} \frac{v}{\sqrt{2}} u^I_{j,R} - \overline{d^I_{i,L}} Y^d_{ij} \frac{v}{\sqrt{2}} d^I_{j,R} \tag{2.3}$$

where Y^u_{ij}, Y^d_{ij} are the Yukawa couplings and v is the Higgs vacuum expectation value. Interestingly, the Yukawa couplings are not diagonal in the weak interaction basis. In order to associate a specific mass to each quark, the Lagrangian has to be expressed in a basis in which the mass matrices are diagonal, namely the mass basis. Defining the mass matrices $M^u_{ij}, M^d_{ij} = \frac{v}{\sqrt{2}} Y^u_{ij}, Y^d_{ij}$, they are diagonalised by using unitary matrices and the quarks are expressed in the mass basis:

$$\begin{aligned}
\mathcal{L}_{\text{Yukawa}} &= -(\overline{u^I_{i,L}} V^{u\dagger}_L)\ (V^u_L M^u_{ij} V^{u\dagger}_R)\ (V^u_R u^I_{j,R}) - (\overline{d^I_{i,L}} V^{d\dagger}_L)\ (V^d_L M^d_{ij} V^{d\dagger}_R)\ (V^d_R d^I_{j,R}) \\
&= -(\overline{u^I_{i,L}} V^{u\dagger}_L)\ M^u_{diag}\ (V^u_R u^I_{j,R}) - (\overline{d^I_{i,L}} V^{d\dagger}_L)\ M^d_{diag}\ (V^d_R d^I_{j,R}) \\
&= -\overline{u_{i,L}} M^u_{diag} u_{i,R} - \overline{d_{i,L}} M^d_{diag} d_{i,R}.
\end{aligned} \tag{2.4}$$

where V^u_L, V^u_R, V^d_L, V^d_R are the unitary matrices used to change bases, M^u_{diag},M^d_{diag} are the diagonalised mass matrices and $\overline{u_{i,L}}$,$\overline{d_{i,L}}$,$u_{i,R}$,$d_{i,R}$ the quark fields expressed in the mass basis.

Having determined the mass basis, the charged current in the weak interaction can be expressed in it as

$$\mathcal{L}_{\text{weak flavour changing}} = -\frac{g}{\sqrt{2}}\overline{u_{i,L}}\gamma_\mu W^{-\mu}(V^d_L V^{u\dagger}_L)_{ij} d_{j,L} - \frac{g}{\sqrt{2}}\overline{d_{i,L}}(V^u_L V^{d\dagger}_L)_{ij}\gamma_\mu W^{+\mu} u_{j,L}. \tag{2.5}$$

By convention, the interaction and mass eigenstates are chosen to be equal for up-type quarks, such that the change of bases affects only down-type quarks. The

associated matrix is referred to as the Cabbibo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d^I \\ s^I \\ b^I \end{pmatrix} = (V_L^d V_L^{u\dagger})\vec{d_L} = V^{\mathrm{CKM}}\vec{d_L} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.6}$$

which is a 3x3 unitary complex matrix.

The number of parameters needed to describe the CKM matrix is strongly reduced by the unitarity requirement to 3 real parameters and 1 complex phase. The complex phase changes sign under conjugation of the CKM matrix, which means that it acts differently on matter and anti-matter particles. In the SM, this is the only complex phase that has been measured to be non-zero. Curiously, at least three generations are needed to have a complex phase in the CKM matrix, which means that the SM contains the minimum number of generations to have a difference between matter and anti-matter in the weak interaction.

From measurements, the CKM matrix turns out to have a very strong hierarchy in its couplings. This is intriguing, especially given that the masses of the quarks also show a very strong hierarchy, and that both these coupling and the quark masses originate from the Yukawa couplings. One common way to describe the CKM matrix is with the so-called Wolfenstein parametrisation [9], which exploits this hierarchy. Including terms up to the order of λ^3, the Wolfenstein parametrisation of the CKM matrix is given as

$$V^{\mathrm{CKM}} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} \sim \begin{pmatrix} 0.97441 & 0.22475 & 0.00375 \\ 0.22461 & 0.97353 & 0.04240 \\ 0.00871 & 0.04169 & 0.99909 \end{pmatrix}. \tag{2.7}$$

The most recent values of these constants are A=$0.8403^{+0.0056}_{-0.0201}$, λ=$0.224747^{+0.000254}_{-0.000059}$, $\bar{\rho} = 0.1577^{+0.0096}_{-0.0074}$, $\bar{\eta} = 0.3493^{+0.0095}_{-0.0071}$, where $\bar{\rho}$ and $\bar{\eta}$ have been normalised following $\bar{\rho}, \bar{\eta} = \rho/(1-\frac{1}{2}\lambda^2), \eta/(1-\frac{1}{2}\lambda^2)$ [10]. The hierarchy in the CKM matrix is encoded in the order of λ associated with each element.

In the lepton sector, the same procedure can be applied in principle. The main difference between the lepton and the quark sector is that the masses of neutrinos have been measured to be less than 2 eV, far smaller than the masses of all other matter particles [11]. Additionally, it is not trivial to add mass terms for the

neutrino in the SM, as the right-handed neutrino is not charged under any of the SM interactions, not even under the weak interaction. Such a particle that only interacts via gravity is called sterile, and no sterile neutrino has been observed yet. Still, an equivalent of the CKM matrix can be defined for the lepton sector and is called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The off-diagonal terms in the PMNS matrix are on average far larger than those in the CKM matrix, showing another intriguing difference between the quark and lepton sector; their mass hierarchies and their mixing hierarchies are both very different.

The construction of the Standard Model leaves many open questions:

- Why are there three generations? Is this related to the fact that at least three generations are needed to generate matter-antimatter differences in the weak interaction?
- Why do the masses of the three generations differ significantly for quarks and charged leptons? Why do neutrinos have such small masses compared to all other fermions? Is it possible that the origin of their mass is different?
- Why does the mixing in the quark sector have a strong hierarchy, and why is such a hierarchy almost absent in the lepton sector?

Together, these questions form the flavour puzzle of the Standard Model: why do the parameters in the Standard Model related to matter have the structure that is found in experiments?

In this thesis, beauty quarks, which are part of the third generation, are investigated to search for deviations from the Standard Model. Such deviations which could lead to answers to the flavour puzzle. Specifically, decays of beauty quarks through Flavour Changing Neutral Currents are examined, as they are suppressed in the Standard Model and can be modified in models that address the flavour puzzle.

2.1.2 Flavour changing neutral currents

Flavour Changing Neutral Currents (FCNCs) are transitions in which a quark changes flavour, while the current that mediates the flavour change is neutral. Interestingly, FCNCs are very suppressed in the SM, which means that new heavy or weakly coupling particles would have a relatively large effect on FCNC decays. First, it will be detailed why FCNCs are suppressed and why it is interesting to look at suppressed

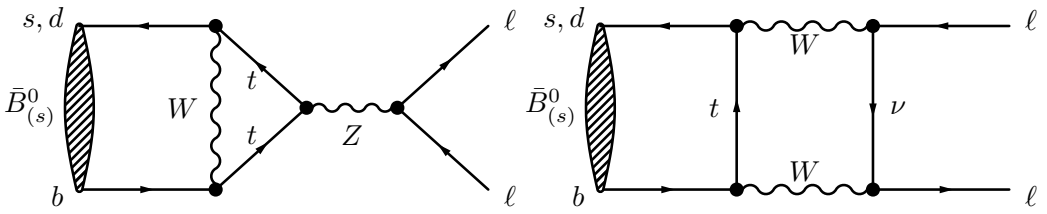

Figure 2.2: The two types of Feynman diagrams involved in $B^0_{(s)} \to \ell^+\ell^-$ decays. The left diagram is referred to as a "Z-penguin" diagram, the right diagram as a "box" diagram. The other penguin diagram is obtained by replacing internal t quarks by internal $W^\pm$ bosons and vice versa.

processes. Subsequently, the two types of FCNC that are important for $B^0_{(s)} \to \ell^+\ell^-$ decays are introduced.

The only transitions that change the flavour of particles in the SM proceed via the exchange of charged W bosons. Therefore, the only way to have a FCNC in the SM is via loop diagrams containing two charged current interactions. For example, the Feynman diagrams that contribute to $B^0_{(s)} \to \ell^+\ell^-$ decays are shown in Figure 2.2. In principle, all up-type quarks contribute to $B^0_{(s)} \to \ell^+\ell^-$ decays in the Z-penguin or box loop diagrams, with a strength determined by the CKM coupling constants and the quark masses:

$$\begin{aligned}\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-) &\propto V_{ub}V^*_{uq}Y_0(x_u) + V_{cb}V^*_{cq}Y_0(x_c) + V_{tb}V^*_{tq}Y_0(x_t) \\ &\propto +V_{cb}V^*_{cq}\left[Y_0(x_c) - Y_0(x_u)\right] + V_{tb}V^*_{tq}\left[Y_0(x_t) - Y_0(x_u)\right],\end{aligned} \tag{2.8}$$

where q is used to refer to either a d or s-type quark. $Y_0(x_i)$, with $x_i = m_i^2/m_W^2$, where m_i is the quark mass and m_W the $W^\pm$ boson mass, is a gauge-independent variable known as the Inami-Lim factor. It gives the first-order contribution from all diagrams with a certain up-type quark to the $B^0_{(s)} \to \ell^+\ell^-$ transition and was first calculated in Ref. [12]. As it is proportional to x_i at first order, it strongly favours contributions from heavier quarks. The second line uses the unitarity of the CKM matrix, $V_{ub}V^*_{uq} = -V_{cb}V^*_{cq} - V_{tb}V^*_{tq}$. If all the up-type quarks would have the same mass, the expression in Equation 2.8 would reduce to zero. Therefore, in the SM FCNCs occur only because of the inequality of the quark masses.

As the CKM elements associated with the charm and top quark both contribute at the same level ($\sim A\lambda^3$), and the mass of the top quark is far larger than the charm or up quark mass, the amplitude of $B^0_{(s)} \to \ell^+\ell^-$ is dominated by the top quark contribution.

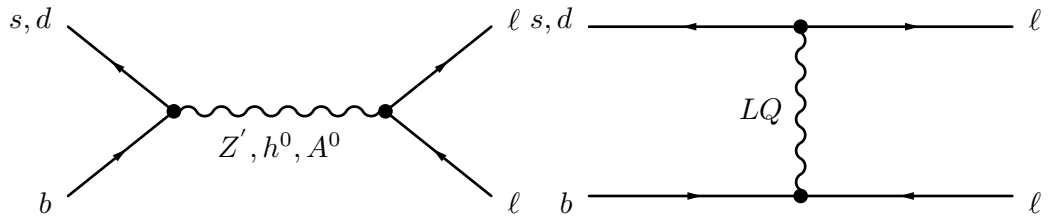

Figure 2.3: Two types of Feynman diagrams for $B^0_{(s)} \to \ell^+\ell^-$ decays with beyond the SM particles. The diagram on the left involves a Z' which directly couples quarks from different generations, the diagram on the right a leptoquark LQ which couples quarks and leptons.

The reach of FCNC decays to probe New Physics at tree-level is best illustrated with a toy model. For example, a new particle like a Z' or leptoquark could directly couple the quarks to leptons in $B^0_{(s)} \to \ell^+\ell^-$ and thus lead to a tree-level diagram like in Figure 2.3. As explained in Ref. [13], a constraint on the $B^0_s \to \mu^+\mu^-$ amplitude at the level of 20% translates to a constraint on $|g_{b\mu}g_{s\mu}|/M_U^2$, corresponding to a leptoquark with mass $M_U = 40\,\text{TeV}$ and couplings $g_{b\mu}g_{s\mu}$ of order 1 or a combined CKM-like coupling strength $|g_{b\mu}g_{s\mu}| = 7.5 \cdot 10^{-4}$ for a leptoquark with a mass $M_U = 1\,\text{TeV}$. These constraints complement direct searches at experiments like ATLAS and CMS for small couplings and surpass them for coupling of order 1.

For this thesis, two kinds of FCNC are important, namely the $b \to q\ell^+\ell^-$ transition and the $b\overline{q} \to \overline{b}q$ transition.

The $b \to q\ell^+\ell^-$ transition governs leptonic decays, $B^0_{(s)} \to \ell^+\ell^-$, in which neutral B mesons decay to two leptons. One of those decays is $B^0_{(s)} \to \mu^+\mu^-$, which is the topic of this thesis. Due to the strong suppression of the $b \to q\ell^+\ell^-$ transition in the SM, it is very interesting to measure decays that are mediated by this transition and to probe heavy new particles that lift some of these suppression factors. The $b \to q\ell^+\ell^-$ transition also governs semileptonic decays of the type $X_b \to X_q\ell^+\ell^-$, where X_b (X_q) is a hadron containing a b (s or d) quark. Semileptonic decays probe observables that are complementary to $B^0_{(s)} \to \ell^+\ell^-$ decays; in those observables, tensions have been observed with respect to SM predictions. The $b \to q\ell^+\ell^-$ transition is discussed in Section 2.2.

The $b\overline{q} \to \overline{b}q$ transition is essential to describe neutral B mesons, as it causes them to mix with their antipartners and propagate as CP eigenstates, a phenomenon that is referred to as neutral B meson mixing. As these two CP eigenstates have a different lifetime and mass, neutral B meson mixing modifies the time dependence decay rates of neutral B meson decays, such as $B^0_{(s)} \to \ell^+\ell^-$. The $b\overline{q} \to \overline{b}q$ transition and neutral B meson mixing are discussed in Section 2.4.

2.2 The $b \to q\ell^+\ell^-$ transition

The $b \to q\ell^+\ell^-$ transition is a specific example of a Flavour Changing Neutral Current. What makes the $b \to q\ell^+\ell^-$ transition interesting is that while it is strongly suppressed in the Standard Model, it is less suppressed than other semi-leptonic FCNC transitions such as $c \to u\ell^+\ell^-$, $t \to c\ell^+\ell^-$ or $s \to d\ell^+\ell^-$. This makes it experimentally possible to test the SM while searching for small deviations from its predictions with decays meditated by $b \to q\ell^+\ell^-$ transitions.

2.2.1 Effective field theory

For weak decays of B hadrons, with a momentum transfer of $\mathcal{O}(m_b)$, the scale of the weak interaction bosons of $\mathcal{O}(M_W)$ is much higher. Therefore, a common approach to calculations in decays of B hadrons is to use an effective low-energy theory. In such a theory, the heavy degrees of freedom (t quark, W boson, Z boson) are integrated out and the transitions discussed in this thesis reduce to an effective four-point interactions with associated Wilson coefficients $\mathcal{C}_i$ and operators $\mathcal{O}_i$ [14, 15]. The Wilson coefficients contain the short-distance effects, are independent of the initial and final state, and are calculated in pertubation theory, including higher order corrections. The Wilson operators contain the long-distance non-perturbative effects, and depend on the initial and final state. They can be computed in a number of different ways, for example with sum rules or lattice QCD.

The amplitude associated with a decay, given such an effective low-energy theory, is given by

$$\mathcal{A}(M \to f) = \langle f|\mathcal{H}_{\text{eff}}|M\rangle \propto \left[\sum_i \mathcal{C}_i^{(\prime)} \langle f|\mathcal{O}_i^{(\prime)}|M\rangle\right], \tag{2.9}$$

where $\mathcal{H}_{\text{eff}}$ is the effective Hamiltonian associated with a certain four-point interaction and the sum includes all possible effective Lorentz structures that can contribute.

For the $b \to q\ell^+\ell^-$ transition, the effective Hamiltonian is given by

$$\mathcal{H}_{\text{eff}} = -\frac{G_F\alpha}{\sqrt{2}\pi} V_{tq}^* V_{tb} \left[\sum_{i=7,9,10,S,P} \mathcal{C}_i\mathcal{O}_i + \mathcal{C}_i'\mathcal{O}_i'\right], \tag{2.10}$$

where G_F is the Fermi Constant, α is the electromagnetic strength constant, and V_{tq}^* and V_{tb} are the relevant CKM elements in the SM. The normalisation factor for

this sum can be related to the Feynman diagrams in Figure 2.2, as it shows the relevant CKM elements, and the weak coupling constant g squared is proportional to G_F. For New Physics contributions, like in Figure 2.3, the same normalisation factor is used for a direct comparison between SM and BSM contributions through their modifications of the Wilson coefficients $\mathcal{C}_i$.

The operators $\mathcal{O}_i^{(\prime)}$ are defined as [16]

$$\begin{aligned}
\mathcal{O}_7 &= \frac{m_b}{e}(\bar{q}\sigma_{\mu\nu}P_R b)F^{\mu\nu} & \mathcal{O}_7' &= \frac{m_b}{e}(\bar{q}\sigma_{\mu\nu}P_L b)F^{\mu\nu} \\
\mathcal{O}_9 &= (\bar{q}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell) & \mathcal{O}_9' &= (\bar{q}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu \ell) \\
\mathcal{O}_{10} &= (\bar{q}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\gamma_5 \ell) & \mathcal{O}_{10}' &= (\bar{q}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu\gamma_5 \ell) \\
\mathcal{O}_\mathcal{S} &= m_b(\bar{q}P_R b)(\bar{\ell}\ell) & \mathcal{O}_\mathcal{S}' &= m_b(\bar{q}P_L b)(\bar{\ell}\ell) \\
\mathcal{O}_\mathcal{P} &= m_b(\bar{q}P_R b)(\bar{\ell}\gamma_5 \ell) & \mathcal{O}_\mathcal{P}' &= m_b(\bar{q}P_L b)(\bar{\ell}\gamma_5 \ell)
\end{aligned} \tag{2.11}$$

with $P_{L,R} = 1 \pm \gamma_5$ the left- and right-handed projections of the quark current, and $\mathcal{C}_i^{(\prime)}$ are the Wilson coefficients, which are complex numbers associated with each of these operators. The operator labels indicate the type of leptonic current. Respectively, they are associated with the photon penguin contribution $\mathcal{O}_7$, the vector $\mathcal{O}_9$ and axial vector $\mathcal{O}_{10}$ lepton current, and the scalar $\mathcal{O}_\mathcal{S}$ and pseudo-scalar $\mathcal{O}_\mathcal{P}$ lepton current. The primed operators refer to contributions with a quark current with the opposite handedness from the SM contributions. These definitions are chosen as contributions from opposite-handed quark currents are negligible in the SM; the leptonic currents could have been defined as here or in terms of left- and right-handed currents.

In the Standard Model, the only operators with a relevant contribution are $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$, while $\mathcal{O}_\mathcal{P}$ and $\mathcal{O}_\mathcal{S}$ contribute at a negligible level as their Wilson coefficients are suppressed by a factor M_{B_q}/M_W[17], where M_{B_q} is the mass of the $B^0_{(s)}$ meson.

2.2.2 Types of $b \to q\ell^+\ell^-$ decays

The $b \to q\ell^+\ell^-$ transition is probed with three kinds of decays, namely $X_b \to X_q\gamma$ decays, $X_b \to X_q\ell^+\ell^-$ decays and $B^0_{(s)} \to \ell^+\ell^-$ decays.

$X_b \to X_q\gamma$ decays proceed through the $b \to q\gamma$ transition, producing a hadron and a photon. These so-called radiative decays directly probe the photon penguin contribution and are sensitive to $\mathcal{C}_7^{(\prime)}$ only. As they probe Wilson coefficients that are not found in $B^0_{(s)} \to \ell^+\ell^-$ decays, they are not described in more detail here.

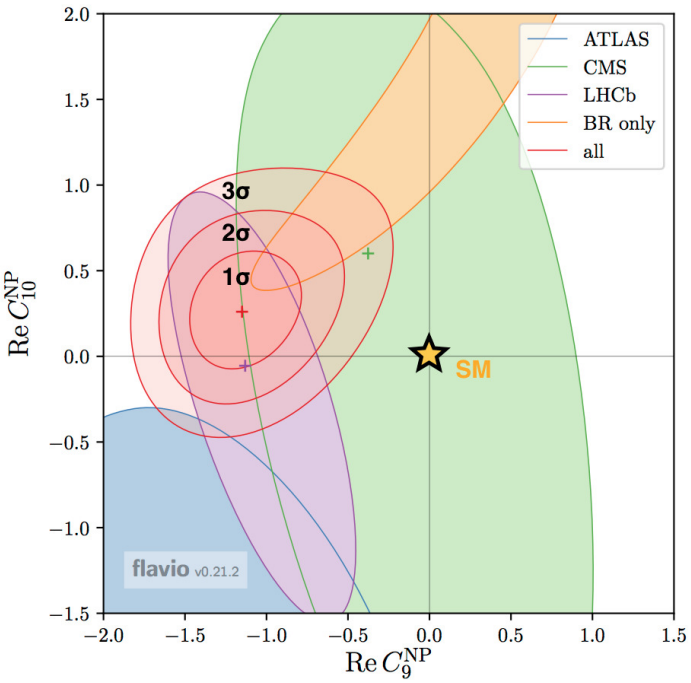

Figure 2.4: Constraints on New Physics contributions to $\mathcal{C}_9$ and $\mathcal{C}_{10}$ from a fit to $b \to q\ell^+\ell^-$ data, assuming both are real and that other Wilson coefficients are as in the SM. Individual constraints from branching fraction measurements only or from angular analyses by ATLAS, CMS, and LHCb, respectively, are indicated along with their contours at one standard deviation. Based on a figure from Ref. [18].

$X_b \to X_q\ell^+\ell^-$ decays, where the b quark is accompanied by one or more spectator quarks in the $b \to q\ell^+\ell^-$ transition, are referred to as semi-leptonic decays, as the final state contains one or more hadrons accompanied by two leptons. The three or more particles in the final state allow for enough degrees of freedom such that semi-leptonic decays are sensitive to all the operators that participate in the $b \to q\ell^+\ell^-$ transition. However, semi-leptonic decays also suffer from significant theoretical uncertainties, mainly from QCD effects and from interference with the resonant $X_b \to X_q J/\psi$ process. As there are many available modes and possible measurements with those modes, global fits to the data have been developed by multiple groups [18, 19]. For example, the global fit in [18] combines branching fraction measurements of $B^0 \to K^{*0}\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$, and angular measurements of $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$. Before the publication of the $B_s^0 \to \mu^+\mu^-$ analysis in this thesis, these fits were indicating a shift in $\mathcal{C}_9 = -\mathcal{C}_{10} = -0.68$, with a significance of around 4.5 standard deviations. One such fit is shown in Figure 2.4.

Finally, leptonic decays, where the q quark in the $b \to q\ell^+\ell^-$ transition is part of the initial state, namely the neutral $B_{(s)}^0$ meson, and the final state consists of only two leptons, with a general form of $B_{(s)}^0 \to \ell^+\ell^-$. The decays under discussion in this

thesis, $B^0_{(s)} \to \mu^+\mu^-$ decays, are the most famous example of a leptonic rare B decay, and are the topic of the rest of this Chapter.

2.3 $B^0_{(s)} \to \ell^+\ell^-$ decays

The Feynman diagrams that play a role in $B^0_{(s)} \to \ell^+\ell^-$ decays according to the SM are shown in Figure 2.2. What makes $B^0_{(s)} \to \ell^+\ell^-$ decays unique is that they only contain leptons and no quarks in the final state. As a result, the interactions between the initial and final state are limited to higher order QED corrections, which are small. The left- and right-hand side of the $B^0_{(s)} \to \ell^+\ell^-$ diagram can be factorised from each other for each operator $\mathcal{O}_i^{(\prime)}$:

$$\mathcal{C}_i^{(\prime)} \langle \ell^+\ell^- | \mathcal{O}_i^{(\prime)} | B^0_{(s)} \rangle = \mathcal{C}_i^{(\prime)} \langle \ell^+\ell^- | \mathcal{O}_{i,\ell\ell}^{(\prime)} | 0 \rangle \otimes \langle 0 | \mathcal{O}_{i,qq}^{(\prime)} | B^0_{(s)} \rangle \,, \tag{2.12}$$

where $\mathcal{O}_{i,\ell\ell}^{(\prime)}$ and $\mathcal{O}_{i,qq}^{(\prime)}$ represent the leptonic and hadronic part of the operator respectively. Because of the Ward identity [20], $\mathcal{O}_7^{(\prime)}$ does not contribute to $B^0_{(s)} \to \ell^+\ell^-$ decays.

As the $B^0_{(s)}$ meson is a pseudoscalar particle, its hadronic matrix elements either have to transform as axial vector or as pseudoscalar quark currents. The axial vector current is part of $\mathcal{O}_9^{(\prime)}$ and $\mathcal{O}_{10}^{(\prime)}$ and is expressed as

$$\langle 0 | \overline{q} \gamma^\mu \gamma_5 b | B^0_{(s)} \rangle = i f_{B^0_{(s)}} p^\mu, \tag{2.13}$$

where $f_{B^0_{(s)}}$ is the decay constant of the $B^0_{(s)}$ meson, which contains all the non-perturbative hadronic physics associated with the transition of a $B^0_{(s)}$ meson to a QCD vacuum.

The p_μ term that remains is contracted with the leptonic matrix elements. There are two leptonic matrix elements, one for the leptonic vector current($\mathcal{O}_9^{(\prime)}$) and one for the leptonic axial vector current($\mathcal{O}_{10}^{(\prime)}$). Respectively, they give

$$\begin{aligned} \langle \ell^+\ell^- | \overline{\ell} p_\mu \gamma^\mu \ell | 0 \rangle &= 0, \\ \langle \ell^+\ell^- | \overline{\ell} p_\mu \gamma^\mu \gamma_5 \ell | 0 \rangle &= 2 m_\ell \langle \ell^+\ell^- | \overline{\ell} \gamma_5 \ell | 0 \rangle = 2 \eta_\lambda m_\ell M_{B_q}, \end{aligned} \tag{2.14}$$

where the first contribution vanishes because of the Dirac equation of motion, $(\gamma^\mu p_\mu - m)\psi = 0$. For the second contribution, m_ℓ is the mass of the final state leptons, and η_λ is a term dependent on the helicity of the final state leptons, caused by γ_5 picking out the helicity of the state. η_λ is equal to -1 if both leptons are left-handed

and +1 if both are right-handed. As a result of the leptonic vector current operator vanishing, only $\mathcal{O}_{10}^{(\prime)}$ contributes to $B_{(s)}^0 \to \ell^+\ell^-$ decays in the Standard Model.

The pseudoscalar quark current is relevant for $\mathcal{O}_S^{(\prime)}$ and $\mathcal{O}_P^{(\prime)}$. It is obtained by contracting the definition in Equation 2.13 with p^μ on the left and right-hand side of the equation, which gives

$$\langle 0|\overline{q}\gamma^\mu p_\mu \gamma_5 b|B_{(s)}^0\rangle = i f_{B_{(s)}^0} p^\mu p_\mu, \tag{2.15}$$

$$-(m_b + m_q)\,\langle 0|\overline{q}\gamma_5 b|B_{(s)}^0\rangle = i f_{B_{(s)}^0} {M_{B_q}}^2, \tag{2.16}$$

$$\langle 0|\overline{q}\gamma_5 b|B_{(s)}^0\rangle = -i f_{B_{(s)}^0} \frac{{M_{B_q}}^2}{m_b + m_q}, \tag{2.17}$$

where the Dirac equation of motion is used to obtain the quark mass terms. It should be noted that the (pseudo-)scalar operators have the opposite chiral projectors, as shown in Equation 2.11, which is why $\mathcal{C}_{10}^{(\prime)}$ and $\mathcal{C}_P^{(\prime)}$ have the same sign in the $B_{(s)}^0 \to \ell^+\ell^-$ amplitude.

The contribution from the leptonic matrix elements for the pseudoscalar current are given by [21]:

$$\langle \ell^+\ell^-|\overline{\ell}\gamma_5\ell|0\rangle = \eta_\lambda M_{B_q}, \tag{2.18}$$

$$\langle \ell^+\ell^-|\overline{\ell}\ell|0\rangle = -M_{B_q}\sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}}. \tag{2.19}$$

The contributions from the (pseudo-)scalar currents differ from those found for the axial vector currents by scaling with a relative $M_{B_q}/2m_\ell$. Therefore, new contributions to (pseudo-)scalar operators in $B_{(s)}^0 \to \ell^+\ell^-$ decays are enhanced relative to the SM amplitude.

Combining the calculated vector and leptonic currents with Equations 2.9 and 2.10, the amplitude of $B_{(s)}^0 \to \ell^+\ell^-$ is found to be

$$\begin{aligned}\mathcal{A}(B_{(s)}^0 \to \ell^+\ell^-) = &-\frac{G_F\alpha}{2\sqrt{2}\pi} V_{tq}^* V_{tb} f_{B_{(s)}^0} M_{B_q} \\ &\times \left\{ \eta_\lambda \left[2m_\ell \left(\mathcal{C}_{10} - \mathcal{C}_{10}^{\prime}\right) + {M_{B_q}}^2 \left(\frac{m_b}{m_b + m_q}\right)\left(\mathcal{C}_P - \mathcal{C}_P^{\prime}\right)\right] \right. \\ &\left. + \sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}}\, {M_{B_q}}^2 \left(\frac{m_b}{m_b + m_q}\right)\left(\mathcal{C}_S - \mathcal{C}_S^{\prime}\right) \right\},\end{aligned} \tag{2.20}$$

From this amplitude, a few conclusions can be drawn. The only contribution to $B_{(s)}^0 \to \ell^+\ell^-$ decays in the SM comes from $\mathcal{C}_{10}$. This means that in the SM, the

amplitude of $B^0_{(s)} \to \ell^+\ell^-$ decays scales with the mass of the lepton. Compared to semileptonic $b \to s\ell\ell$ decays, the amplitude is suppressed by a factor $\frac{2m_\ell}{M_{B_q}}$; this phenomenon is called helicity suppression. Helicity suppression can be understood from the rest frame of the neutral B meson. As the neutral B meson has zero spin, the spin and momentum of the two leptons have to be opposite by the conservation of angular momentum and momentum respectively. Helicity is the inner product of the two, so a $B^0_{(s)} \to \ell^+\ell^-$ decay occurs with two right-handed or left-handed particles in terms of helicity. At the same time, the weak interaction selects chiral states, and only interacts with chirally left-handed particles or chirally right-handed antiparticles. For massless particles, the chiral and helicity states are the same, in which case the decay is forbidden. The larger the mass of the leptons the more the chiral and helicity states decouple, and the smaller the effect of helicity suppression.

However, any scalar or pseudoscalar contributions are not helicity-suppressed, which means that relative to the SM amplitude they are enhanced by a factor $\frac{{M_{B_q}}^2}{2m_\ell}$. This makes $B^0_{(s)} \to \ell^+\ell^-$ decays very sensitive to new (pseudo)-scalar contributions to $b \to q\ell^+\ell^-$ transitions.

Note that $B^0_{(s)} \to \ell^+\ell^-$ decays are only sensitive to combinations of the form $\mathcal{C}_i$-$\mathcal{C}'_i$, which is caused by the opposite-handedness of the quark current resulting in a relative minus sign. Therefore, additional measurements from semi-leptonic $b \to q\ell^+\ell^-$ decays or theoretical constraints are needed to fully constrain each Wilson coefficient entering into the $B^0_{(s)} \to \ell^+\ell^-$ amplitude.

Finally, the only non-perturbative QCD effects in this amplitude enter through the decay constant $f_{B^0_{(s)}}$. As it has recently been calculated in lattice QCD with a sub-percent level uncertainty [22], the theoretical predictions for $B^0_{(s)} \to \ell^+\ell^-$ decays are very clean and dominated by uncertainties on the input parameters.

In conclusion, $B^0_{(s)} \to \ell^+\ell^-$ decays are a very interesting and clean environment to measure the axial vector contributions to $b \to q\ell^+\ell^-$ transitions and compare them with the SM, while also probing new scalar or pseudo-scalar contributions.

For the following, it is useful to rewrite the $B^0_{(s)} \to \ell^+\ell^-$ amplitude from Equation 2.20 in terms of two separate amplitudes:

$$\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-) \propto V^*_{tq}V_{tb} f_{B^0_{(s)}} M_{B_q} m_\ell \mathcal{C}^{\mathrm{SM}}_{10}[\eta_\lambda P^q_{\ell\ell} + S^q_{\ell\ell}], \tag{2.21}$$

where $P^q_{\ell\ell}$ and $S^q_{\ell\ell}$ are the two complex amplitudes contributing to $B^0_{(s)} \to \ell^+\ell^-$ decays, defined as

$$P^q_{\ell\ell} = \frac{\mathcal{C}_{10} - \mathcal{C}'_{10}}{\mathcal{C}^{\rm SM}_{10}} + \frac{{M_{B_q}}^2}{2m_\ell}\left(\frac{m_b}{m_b + m_q}\right)\left(\frac{\mathcal{C}_P - \mathcal{C}'_P}{\mathcal{C}^{\rm SM}_{10}}\right), \tag{2.22}$$

$$S^q_{\ell\ell} = \sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}}\frac{{M_{B_q}}^2}{2m_\ell}\left(\frac{m_b}{m_b + m_q}\right)\left(\frac{\mathcal{C}_S - \mathcal{C}'_S}{\mathcal{C}^{\rm SM}_{10}}\right). \tag{2.23}$$

It should be noted that $P^q_{\ell\ell}$ and $S^q_{\ell\ell}$ have both a magnitude and a phase, and thus there are four separate parameters, two magnitudes and two phases, that play a role in $B^0_{(s)} \to \ell^+\ell^-$ decays. In the Standard Model, $\mathcal{C}_{10} = \mathcal{C}^{\rm SM}_{10}$, and all other Wilson coefficients are negligibly small, such that $P^q_{\ell\ell} = 1, S^q_{\ell\ell} = 0$.

Given the expression of the amplitude of $B^0_{(s)} \to \ell^+\ell^-$, it is interesting to discuss all possible observables in $B^0_{(s)} \to \ell^+\ell^-$ decays. Because they involve pseudoscalar mesons decaying to two fermions, there is no physical information that can be measured from angular observables. However, the effects of neutral meson mixing are required for a complete description of the time-dependent decay rate of $B^0_{(s)} \to \ell^+\ell^-$ decays.

2.4 Neutral B meson mixing

Hadronic particles are eigenstates of the electromagnetic and strong force Hamiltonian, and most also are eigenstates of the weak interaction. An exception to this rule are neutral mesons, such as the neutral B mesons (B^0 and B^0_s) discussed in this thesis. Such neutral mesons can transition into neutral anti-mesons via the $b\overline{q} \to \overline{b}q$ process, and thus they exist in a quantum superposition with their own antipartner. This superposition is commonly referred to as neutral meson mixing.

Neutral meson mixing affects how $B^0_{(s)}$ mesons propagate, and the time evolution of $B^0_{(s)}$ mesons under mixing is described by an effective Hamiltonian. Firstly, the states $B^0_{(s)}$ and $\overline{B}^0_{(s)}$ are defined, which are eigenstates of the strong and electromagnetic interaction, and have a common mass $m_{B^0_{(s)}}$, but opposite flavour content. These states are in superposition with each other because of the $b\overline{q} \to \overline{b}q$ transition, and thus together are described as

$$\Psi(t) = a(t)\,|B^0_{(s)}\rangle + b(t)\,|\overline{B}^0_{(s)}\rangle\,, \tag{2.24}$$

with $a(t)$ and $b(t)$ coefficients that as function of time describe the contributions from the $B^0_{(s)}$ and $\overline{B}^0_{(s)}$ states. Given this description, $\Psi(t)$ is denoted in the subspace of the flavour eigenstates as

$$\Psi(t) = \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \tag{2.25}$$

To describe the evolution of $\Psi(t)$ over time, the Schrödinger equation is used,

$$i\frac{\partial\Psi(t)}{\partial t} = \mathcal{H}\Psi(t), \tag{2.26}$$

where the effective Hamiltonian is given by

$$\mathcal{H} = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M^*_{12} - \frac{i}{2}\Gamma^*_{12} & M - \frac{i}{2}\Gamma \end{pmatrix}. \tag{2.27}$$

In this Hamiltonian, M and Γ are the mass and decay rate of the $B^0_{(s)}$ meson. The $B^0_{(s)}$-$\overline{B}^0_{(s)}$ transition implies that this Hamiltonian contains off-diagonal elements M_{12} and Γ_{12}, which describe different contributions to neutral meson mixing. In the rest of this thesis, it is assumed that M_{12} and Γ_{12} are real, which is the same as assuming that there is no CP violation in $B^0_{(s)}$-$\overline{B}^0_{(s)}$ mixing. In the SM, this holds to $\mathcal{O}(10^{-4})$, and LHCb has tested this assumption to $\mathcal{O}(10^{-3})$, not finding any deviation from the SM.

M_{12} describes all contributions to neutral meson mixing, including virtual transitions, and is dominated by box diagrams containing top quarks, similarly to the $b \to q\ell^+\ell^-$ transition. One such box diagram is shown in Figure 2.5. The other diagram is obtained by replacing the internal t quarks by internal $W^\pm$ bosons and vice versa. Γ_{12} describes contributions to neutral meson mixing from on-shell transitions through an intermediate final state, and is dominated by $B^0_{(s)} \to D^+_s D^-_{(s)}$ or $B^0_{(s)} \to D^- D^-_{(s)} K^0$ transitions, included excited versions of any of these final states [23].

From the effective Hamiltonian in Eq. 2.27, the eigenstates for the time evolution of the neutral B mesons in superposition are determined,

$$\begin{aligned} |B^0_{(s),H}\rangle &\propto |B^0_{(s)}\rangle - |\overline{B}^0_{(s)}\rangle, \\ |B^0_{(s),L}\rangle &\propto |B^0_{(s)}\rangle + |\overline{B}^0_{(s)}\rangle, \end{aligned} \tag{2.28}$$

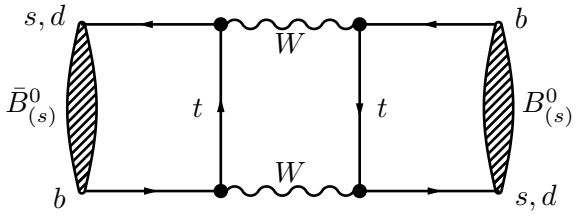

Figure 2.5: One of the two Feynman diagrams involved in the $b\overline{q} \to \overline{b}q$ transition. The other diagram is obtained by switching every internal t quark for an internal $W^{\pm}$ boson and vice versa.

where these two eigenstates have their own mass, lifetime and CP eigenvalue. These states are referred to as $B^0_{(s),H}$ and $B^0_{(s),L}$, as they are the heavy and the light eigenstate respectively. For the neutral B meson system, the $B^0_{(s),H}$ state is the CP-odd state and the state with the longer lifetime. This is because only the CP-even state $B^0_{(s),L}$ decays to $D^+_s D^-_{(s)}$, with a significant branching fraction, and thus decays faster. The differences in mass Δm_q and decay width $\Delta\Gamma_q$ between the two eigenstates are used as the parameters that describe the effects from neutral meson mixing. They are defined as

$$\Delta m_q = m_H - m_L = 2|M_{12}|, \tag{2.29}$$

$$\Delta\Gamma_q = \Gamma_L - \Gamma_H = 2|\Gamma_{12}|, \tag{2.30}$$

These parameters have been measured in experiment for B^0 and B^0_s mesons, and the current averages of each parameter normalised to the B^0 and B^0_s lifetimes, which are very similar, are shown in Table 2.1.

Table 2.1: The measured values of the neutral meson mixing parameters for B^0 and B^0_s mesons. Numbers taken from Ref. [24].

Parameter	B^0_s	B^0
$\Delta m_q/\Gamma_q$	26.79 ± 0.08	0.770 ± 0.004
$\Delta\Gamma_q/\Gamma_q$	0.132 ± 0.008	-0.002 ± 0.010

In Table 2.1, a few interesting patterns can be observed. First, in the SM, both $\Delta m_q/\Gamma_q$ and $\Delta\Gamma_q/\Gamma_q$ are enhanced by about a factor 35 for B^0_s compared to B^0 mesons, mainly due to the relative CKM elements involved in the transitions. This has allowed the measurement of both mixing observables in the B^0_s system, while in the B^0 system only Δm_d has yet been determined. Accordingly, throughout this thesis B^0 mesons are assumed to have $\Delta\Gamma_d = 0$. Additionally, the number of

oscillations per unit of decay time $\Delta m_q/\Gamma_q$ is a lot larger than the difference in decay width between the two eigenstates $\Delta\Gamma_q/\Gamma_q$. This is explained by the fact that Δm_q includes virtual contributions with top quarks, whereas $\Delta\Gamma_q$ does not.

Neutral B meson mixing is important to consider when investigating $B^0_{(s)} \to \ell^+\ell^-$ decays for two reasons. First, as the diagram for Δm_q is similar to that of $B^0_{(s)} \to \ell^+\ell^-$, it is instructive to examine the SM prediction for Δm_q. In Section 2.4.1, it is shown that $B^0_{(s)} \to \ell^+\ell^-$ branching fractions can be computed relative to the value of Δm_q, in which case some parametric uncertainties drop out. Second, the $B^0_{(s)}$ decay time dependence, which is caused by neutral meson mixing, is needed as input to determine $B^0_{(s)} \to \ell^+\ell^-$ observables from experimental measurements, as detailed in Section 2.4.2.

2.4.1 Calculating Δm_q in the SM

In the SM, the mass difference Δm_q is calculated to be

$$\Delta m_q = \frac{G_F^2 M_W^2}{6\pi^2} |V_{tq}^* V_{tb}|^2 f^2_{B^0_{(s)}} M_{B_q} S_0(x_t) \eta_{2B} \hat{B}^1_{B_q}, \tag{2.31}$$

where M_W is the mass of the W boson [25] and the last three factors are related to the calculation of the $b\overline{q} \to \overline{b}q$ transition. $S_0(x_t)$, with $x_t = (m_t/M_W)^2$, is a gauge-independent variable known as the Inami-Lim factor, which gives the size of the leading order diagram (save normalisation factors and CKM elements) and was first described in Ref. [12]. η_{2B} gives the perturbative QCD corrections to the diagram at next to leading order, and was determined to be $\eta_{2B} = 0.55210 \pm 0.00062$ in Ref. [22]. $\hat{B}^1_{B_q}$ is the so-called bag parameter, which quantifies the effect from non-perturbative QCD in neutral B meson mixing. The bag parameter is determined from theory and is currently the limiting factor on the precision of the SM prediction for Δm_q. Interestingly, as discussed in Section 2.6.3, the $B^0_{(s)} \to \ell^+\ell^-$ branching fraction can be related to the expression for Δm_q.

2.4.2 The time dependence of neutral B meson decays

The time dependence of neutral B meson decays due to the mixing process affects the experimental determination of the observables in $B^0_{(s)} \to \ell^+\ell^-$ decays, as is presented in Section 2.6.1. This time dependence, denoted as $\Gamma(\overline{B}^{(\overline{0})}_{(s)}(t) \to f)$, is determined in this Section.

The two eigenstates of neutral meson mixing, $B^0_{(s),H}$ and $B^0_{(s),L}$, have a definitive lifetime and mass. From the Schrödinger equation, their time dependence is

$$\begin{aligned} |B^0_{(s),H}\rangle\,(t) &= e^{-(im_H-\Gamma_H/2)t}\,|B^0_{(s),H}\rangle\,(0), \\ |B^0_{(s),L}\rangle\,(t) &= e^{-(im_L-\Gamma_L/2)t}\,|B^0_{(s),L}\rangle\,(0). \end{aligned} \tag{2.32}$$

Combining this with Equation 2.28, the time evolution of a $B^0_{(s)}$ or $\overline{B}^0_{(s)}$ state is given by

$$\begin{aligned} |B^0_{(s)}\rangle\,(t) &= g_+(t)\,|B^0_{(s)}\rangle + g_-(t)\,|\overline{B}^0_{(s)}\rangle\,, \\ |\overline{B}^0_{(s)}\rangle\,(t) &= g_-(t)\,|B^0_{(s)}\rangle + g_+(t)\,|\overline{B}^0_{(s)}\rangle\,, \end{aligned} \tag{2.33}$$

where $g_{+,-}(t)$ are the time dependent mixing amplitudes, which give the probability for a $B^0_{(s)}$ to be in the same or opposite state compared to its production state at a time t respectively. The amplitudes are found to be

$$g_\pm(t) = \frac{1}{2}e^{-iMt-\frac{1}{2}\Gamma t}(e^{-\frac{i}{2}\Delta m_q t+\frac{1}{4}\Delta\Gamma_q t} \pm e^{+\frac{i}{2}\Delta m_q t-\frac{1}{4}\Delta\Gamma_q t}), \tag{2.34}$$

where $M = (m_H + m_L)/2 = M_{B_q}, \Gamma = (\Gamma_\mathrm{H} + \Gamma_\mathrm{L})/2 = \Gamma_q$.

For $B^0_{(s)} \to \ell^+\ell^-$ decays, both the $B^0_{(s)}$ and $\overline{B}^0_{(s)}$ state decay to the same final state, which means that the relevant decay rates starting from a $B^0_{(s)}$ or $\overline{B}^0_{(s)}$ state are

$$\begin{aligned} \Gamma(B^0_{(s)}(t) \to f) &\propto |\,\langle f|B^0_{(s)}\rangle\,\langle B^0_{(s)}|B^0_{(s)}\rangle\,(t) + \langle f|\overline{B}^0_{(s)}\rangle\,\langle \overline{B}^0_{(s)}|B^0_{(s)}\rangle\,(t)|^2, \\ \Gamma(\overline{B}^0_{(s)}(t) \to f) &\propto |\,\langle f|\overline{B}^0_{(s)}\rangle\,\langle \overline{B}^0_{(s)}|\overline{B}^0_{(s)}\rangle\,(t) + \langle f|B^0_{(s)}\rangle\,\langle B^0_{(s)}|\overline{B}^0_{(s)}\rangle\,(t)|^2, \end{aligned} \tag{2.35}$$

where the amplitude $\mathcal{A}_f = \langle f|B^0_{(s)}\rangle$ and its CP conjugate $\bar{\mathcal{A}}_f$ have been determined for $B^0_{(s)} \to \ell^+\ell^-$ decays in Section 2.3. Using the mixing amplitudes $g_\pm(t)$ from Equation 2.34, and the so-called master equations that describe the decays, starting

from either state, are found to be

$$\begin{aligned}
\Gamma(B^0_{(s)}(t)\to f) &\propto |\mathcal{A}_f|^2|g_+(t)|^2 + |\bar{\mathcal{A}}_f|^2|g_-(t)|^2 + 2\mathrm{Re}[\mathcal{A}_f\bar{\mathcal{A}}_f g^*_+(t)g_-(t)] \\
&\propto \left(|\mathcal{A}_f|^2 + |\bar{\mathcal{A}}_f|^2\right)\times \\
&\frac{e^{-\Gamma t}}{2}[\cosh\frac{\Delta\Gamma_q}{2}t + A^f_{\Delta\Gamma_q}\sinh\frac{\Delta\Gamma_q}{2}t + C_f\cos\Delta m_q t - S_f\sin\Delta m_q t], \\
\Gamma(\overline{B}^0_{(s)}(t)\to f) &\propto |\bar{\mathcal{A}}_f|^2|g_+(t)|^2 + |\mathcal{A}_f|^2|g_-(t)|^2 + 2\mathrm{Re}[\mathcal{A}_f\bar{\mathcal{A}}_f g_+(t)g^*_-(t)]. \\
&\propto \left(|\mathcal{A}_f|^2 + |\bar{\mathcal{A}}_f|^2\right)\times \\
&\frac{e^{-\Gamma t}}{2}[\cosh\frac{\Delta\Gamma_q}{2}t + A^f_{\Delta\Gamma_q}\sinh\frac{\Delta\Gamma_q}{2}t - C_f\cos\Delta m_q t + S_f\sin\Delta m_q t],
\end{aligned} \tag{2.36}$$

where the observables $A^f_{\Delta\Gamma_q}$, C_f and S_f describe different components of the decay rate and are defined as

$$A^f_{\Delta\Gamma_q} = \frac{2\mathrm{Re}(\bar{\mathcal{A}}_f/\mathcal{A}_f)}{1+|\bar{\mathcal{A}}_f/\mathcal{A}_f|^2},\ S_f = \frac{2\mathrm{Im}(\bar{\mathcal{A}}_f/\mathcal{A}_f)}{1+|\bar{\mathcal{A}}_f/\mathcal{A}_f|^2},\ C_f = \frac{1-|\bar{\mathcal{A}}_f/\mathcal{A}_f|^2}{1+|\bar{\mathcal{A}}_f/\mathcal{A}_f|^2}. \tag{2.37}$$

From these definitions, it can be seen that the squared sum of $A^f_{\Delta\Gamma_q}$, C_f and S_f is equal to one, and thus that the allowed range for any of these observables is from -1 to +1.

In the above time dependent decay rate, which has been derived for $B^0_{(s)}$ or $\overline{B}^0_{(s)}$ states decaying to the same final state, it is useful to separately consider the total decay rate as well as the asymmetry between the two decay rates $\Gamma(B^0_{(s)}(t)\to f)$ and $\Gamma(\overline{B}^0_{(s)}(t)\to f)$. The total decay rate will be referred to as *untagged*, while the asymmetry requires *tagged* decay rates, *i.e.* rates with information on the starting flavour of the $B^0_{(s)}$ meson. As the only difference between the processes is whether the meson was produced as a $B^0_{(s)}$ or $\overline{B}^0_{(s)}$ state, it is not trivial for an experiment to measure the rate separately for $B^0_{(s)}\to\ell^+\ell^-$ and $\overline{B}^0_{(s)}\to\ell^+\ell^-$. The only method to measure the asymmetry is by using that the b-quark and the d or s-quark in the $B^0_{(s)}/\overline{B}^0_{(s)}$ meson were produced in coincidence with a partner anti-quark. By reconstructing the hadron containing the partner anti-quark, the flavour of the $B^0_{(s)}$ meson is determined.

The untagged decay rate is given as

$$\Gamma(B^0_{(s)}(t)\to f) + \Gamma(\overline{B}^0_{(s)}(t)\to f) \propto (|\mathcal{A}_f|^2 + |\bar{\mathcal{A}}_f|^2)e^{-\Gamma t}[\cosh\Delta\Gamma_q/2t + A^f_{\Delta\Gamma_q}\sinh\Delta\Gamma_q/2t], \tag{2.38}$$

which means its time dependence is fully determined by $A^f_{\Delta\Gamma_q}$, Γ_q and $\Delta\Gamma_q$, whereas there is no dependence on Δm_q. Γ_q and $\Delta\Gamma_q$ have been well measured with tree-level decays of $B^0_{(s)}$ mesons, which means that the time dependence of untagged $B^0_{(s)} \to \ell^+\ell^-$ decays can be used to determine $A^f_{\Delta\Gamma_q}$. $A^f_{\Delta\Gamma_q}$ quantifies how much each eigenstate contributes to the decay. For $A^f_{\Delta\Gamma_q}$ =+1, only the heavy eigenstate contributes; for $A^f_{\Delta\Gamma_q}$ =-1 only the light eigenstate does. For any value inbetween those extremes, both states are contributing at some level. Clearly, $A^f_{\Delta\Gamma_q}$ is only accessible from the untagged decay rate if there is a significant lifetime difference $\Delta\Gamma_q$. In that case, it can be determined by measuring the effective lifetime τ_{eff} of the decay rate, which is the average decay time of signal candidates in an experiment. $A^f_{\Delta\Gamma_q}$ and τ_{eff} are related through

$$\tau_{\text{eff}}/\tau_{B^0_{(s)}} = \frac{1}{1-{y_q}^2}\left[\frac{1+2A^f_{\Delta\Gamma_q}y_q+{y_q}^2}{1+A^f_{\Delta\Gamma_q}y_q}\right], \tag{2.39}$$

where y_q is defined as

$$y_q = \frac{\Delta\Gamma_q}{2\Gamma_q}, \tag{2.40}$$

as explained in Ref. [26]. As mentioned before, contrary to $\Delta\Gamma_d$, $\Delta\Gamma_s$ has been observed and found to be $y_s = 0.066 \pm 0.004$, see Table 2.1. As will be discussed in Section 2.6.1, as a consequence of the significant lifetime difference, the theoretical and experimental definition of a branching fraction are different. This difference has to be taken into account to compare theory with experiment.

Alternatively, to measure meson mixing observables is to examine the time-dependent CP asymmetry. In the case that the final state is the same as its anti-state, $f = \bar{f}$, as for $f = \ell^+\ell^-$, this asymmetry is given by

$$\frac{\Gamma(B^0_{(s)}(t)\to f)-\Gamma(\overline{B}^0_{(s)}(t)\to f)}{\Gamma(B^0_{(s)}(t)\to f)+\Gamma(\overline{B}^0_{(s)}(t)\to f)} = \frac{C_f\cos\Delta m_q t - S_f\sin\Delta m_q t}{\cosh\Delta\Gamma_q/2t + A^f_{\Delta\Gamma_q}\sinh\Delta\Gamma_q/2t}, \tag{2.41}$$

where C_f and S_f are included, as they are observables that quantify CP-violation in the interference between the B meson decay and the mixing process.

With the untagged decay rate in Equation 2.38 and the time-dependent CP asymmetry in Equation 2.41, all observables in neutral B meson decays have been introduced. They are now examined specifically for $B^0_{(s)} \to \ell^+\ell^-$ decays.

2.5 Observables in $B^0_{(s)} \to \ell^+\ell^-$ decays

In the previous Section, the untagged decay rate and time-dependent CP asymmetries of $B^0_{(s)}$ and $\overline{B}^0_{(s)}$ decays to the same final state have been determined in Equations 2.38 and 2.41 respectively. It is instructive to examine the four associated observables which are present in $B^0_{(s)} \to \ell^+\ell^-$ decays.

The first observable is the branching fraction, which is how often a decay of a particle to some final state happens relative to all the decays of a particle, and is the single observable probing the size of the $B^0_{(s)} \to \ell^+\ell^-$ amplitude. The branching fraction for a two-body decay of a particle to a particle-antiparticle pair, such as $B^0_{(s)} \to \ell^+\ell^-$, is given by [27]:

$$\mathcal{B}(B^0_{(s)} \to \ell^+\ell^-) = \tau_{B^0_{(s)}} \frac{\sqrt{1 - \left(\frac{2m_\ell}{M_{B_q}}\right)^2}}{8\pi M_{B_q}} \frac{|\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-)|^2 + |\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)|^2}{2}, \tag{2.42}$$

where $\tau_{B^0_{(s)}} = 1/\Gamma_{B^0_{(s)}}$ normalises the branching fraction to the total decay width. Predictions of the branching fraction for $B^0_{(s)} \to \ell^+\ell^-$ decays are theoretically clean and will be discussed in detail for $B^0_{(s)} \to \mu^+\mu^-$ decays in Section 2.6.

The other observables originate from neutral meson mixing, and they enter the $B^0_{(s)} \to \ell^+\ell^-$ decay rate through the meson mixing process. Their general expressions are given in Equation 2.37. The observable $A^{\ell\ell}_{\Delta\Gamma_q}$ is part of the time-dependent untagged decay rate. For $B^0_{(s)} \to \ell^+\ell^-$ decays it is related to the amplitudes as [28]:

$$\begin{aligned} A^{\ell\ell}_{\Delta\Gamma_q} &= \frac{2\mathrm{Re}(\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-))}{1 + |\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-)|^2} \\ &= \frac{|P^q_{\ell\ell}|^2 \cos 2\phi_S - |S^q_{\ell\ell}|^2 \cos 2\phi_P}{|P^q_{\ell\ell}|^2 + |S^q_{\ell\ell}|^2} \overset{\text{SM}}{\longrightarrow} +1, \end{aligned} \tag{2.43}$$

where the expressions for the $B^0_{(s)} \to \ell^+\ell^-$ amplitudes from Equations 2.21 and 2.23 are used. Note that ϕ_S and ϕ_P are the phases associated with $P^q_{\ell\ell}$ and $S^q_{\ell\ell}$ respectively, and that in the SM, $P^q_{\ell\ell} = 1, S^q_{\ell\ell} = 0$. As introduced in Equation 2.39, the effective lifetime is the experimental observable which probes $A^{\ell\ell}_{\Delta\Gamma_q}$ for $B^0_{(s)} \to \ell^+\ell^-$ decays. It is only sensitive to $A^{\ell\ell}_{\Delta\Gamma_q}$ in case of a significant lifetime difference between the two CP eigenstates from neutral meson mixing, which means this variable is only accessible in the foreseeable future for B^0_s mesons.

Finally, the two CP-violating observables, which are probed with time-dependent CP-asymmetries, are given by [29]:

$$C_{\ell\ell}^{\lambda} = \frac{1 - |\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-)|^2}{1 + |\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-)|^2} = -\eta_\lambda \frac{2|P^q_{\ell\ell} S^q_{\ell\ell}| \cos\phi_S - \phi_P}{|P^q_{\ell\ell}|^2 + |S^q_{\ell\ell}|^2} \xrightarrow{\text{SM}} 0, \tag{2.44}$$

$$S_{\ell\ell}^{\lambda} = \frac{2\text{Im}(\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-))}{1 + |\mathcal{A}(\overline{B}^0_{(s)} \to \ell^+\ell^-)/\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-)|^2} = \frac{|P^q_{\ell\ell}|^2 \sin 2\phi_S - |S^q_{\ell\ell}|^2 \sin 2\phi_P}{|P^q_{\ell\ell}|^2 + |S^q_{\ell\ell}|^2} \xrightarrow{\text{SM}} 0, \tag{2.45}$$

Only $C^{\lambda}_{\ell\ell}$ turns out to depend on the helicity η_λ of the leptons, while $S^{\lambda}_{\ell\ell}$ does not. This is crucial for the experimental sensitivity to these observables, as it is very difficult to measure the helicity of muons at the facilities where $B^0_{(s)} \to \ell^+\ell^-$ decays are investigated.

Out of the three $B^0_{(s)} \to \ell^+\ell^-$ decays, $B^0_{(s)} \to e^+e^-$, $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to \tau^+\tau^-$, only $B^0_{(s)} \to \mu^+\mu^-$ decays are within reach of the current experimental sensitivity. As they are also the subject of this thesis, the rest of this Chapter covers mainly $B^0_{(s)} \to \mu^+\mu^-$ decays. The branching fraction prediction for $B^0_{(s)} \to \mu^+\mu^-$ decays is discussed in Section 2.6, Subsequently, the effective lifetime τ_{eff} is discussed in Section 2.7. Following those CP-symmetric observables, CP-violating observables in $B^0_{(s)} \to \mu^+\mu^-$ decays are discussed in Section 2.8. Last but not least, the roles of $B^0_{(s)} \to e^+e^-$ and $B^0_{(s)} \to \tau^+\tau^-$ decays are briefly discussed in Section 2.3.

2.6 The branching fraction of $B^0_{(s)} \to \mu^+\mu^-$

In this Section, the branching fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays is discussed, which includes two aspects that are important to consider when comparing theory and experiment. In practice, the definition of a branching fraction differs between theory and experiment. Section 2.6.1 illustrates that, because of the lifetime difference between the B^0_s mass eigenstates, a correction factor has to be computed for decays of B^0_s mesons, in order to compare them. Additionally, one or more undetected photons can radiate from the $B^0_{(s)} \to \mu^+\mu^-$ diagram and affect the dimuon mass shape and branching fraction estimate, as discussed in Section 2.6.2. Accounting for both effects, different methods to compute the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are compared using the most up-to-date numerical inputs and theoretical calculations

in Section 2.6.3. The constraints on New Physics from previous measurements are discussed in Section 2.6.4.

2.6.1 The measured and theoretical branching fraction for B_s^0 mesons

The branching fraction of a decay is defined as the fraction of all decays of a particle that result in a given final state. Experimentally, the branching fraction is measured on a sample of particles decaying after some lifetime, while theoretically the branching fraction is defined as the decay rate at $t = 0$. Due to the sizable decay width difference $\Delta\Gamma_s$ between the light and heavy mass eigenstates of the B_s^0-$\bar{B}_s^0$ system, for B_s^0 mesons these definitions are not the same, as noted in Ref. [26]. In this Section, the effect from this difference is worked out to properly compare theory and experiment later.

For experiments, a branching fraction is determined by measuring the total event yield, without using the information on the lifetime of the decaying particle. This is equivalent to integrating over the lifetime of the particle, and thus the experimental branching fraction is defined as

$$\begin{aligned}
\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)_{\text{exp}} &= \frac{1}{2}\int_0^\infty \langle \Gamma(\overline{(B^0_{(s)})}(t) \to \mu^+\mu^-)\rangle \, \mathrm{d}t \\
&\propto \frac{1}{2}\int_0^\infty (|\mathcal{A}(B^0_{(s)} \to \mu^+\mu^-)|^2 + |\mathcal{A}(\overline{B}^0_{(s)} \to \mu^+\mu^-)|^2) \\
&\times \left[e^{-\Gamma t}(\cosh\frac{\Delta\Gamma_q}{2}t + A^{\mu^+\mu^-}_{\Delta\Gamma_s}\sinh\frac{\Delta\Gamma_q}{2}t)\right] \\
&\propto \frac{\tau_{B^0_{(s)}}}{2}(|\mathcal{A}(B^0_{(s)} \to \mu^+\mu^-)|^2 + |\mathcal{A}(\overline{B}^0_{(s)} \to \mu^+\mu^-)|^2)\left[\frac{1 + A^{\mu^+\mu^-}_{\Delta\Gamma_s} y_q}{1 - {y_q}^2}\right],
\end{aligned} \tag{2.46}$$

using the untagged decay rate from Equation 2.38. Note that the difference in lifetime of the heavy and light eigenstate is part of the last term through y_q, as defined in Equation 2.39.

The theoretical definition of a branching fraction is the CP-averaged decay rate in the flavour-eigenstate basis at $t = 0$ relative to the total decay rate. At $t = 0$, there is an equal mixture of heavy and light eigenstates and thus the theoretical

branching fraction is defined as

$$\begin{aligned}\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)_{\text{theo}} &= \frac{\tau_{B^0_{(s)}}}{2} \langle \Gamma(\overset{(-)}{B^0_{(s)}}(t) \to \mu^+\mu^-) \rangle \,|_{t=0} \\ &\propto \frac{\tau_{B^0_{(s)}}}{2} (|\mathcal{A}(B^0_{(s)} \to \mu^+\mu^-)|^2 + |\mathcal{A}(\overline{B}^0_{(s)} \to \mu^+\mu^-)|^2),\end{aligned} \tag{2.47}$$

which does not depend on the lifetime difference between the heavy and light eigenstates. Comparing the two definitions, it is found that

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)_{\text{theo}} \propto \left[\frac{1 - {y_q}^2}{1 + A^{\mu^+\mu^-}_{\Delta\Gamma_s} y_q}\right] \mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)_{\text{exp}}, \tag{2.48}$$

where in the extreme cases of $A^{\mu^+\mu^-}_{\Delta\Gamma_s} = \pm 1$ the branching fraction definitions differ by a relative $\pm y_s = \pm 6.6\%$, and even for $A^{\mu^+\mu^-}_{\Delta\Gamma_s} = 0$ the branching fraction definitions differ by $1 - {y_s}^2 = 0.5\%$. In Figure 2.6, the conversion factor between the two definitions given different values of $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$ is shown, for the value of y_s measured by LHCb at the time. In the rest of this thesis, the experimental definition of the branching fraction will be used.

2.6.2 Radiative photons in $B^0_{(s)} \to \mu^+\mu^-$ decays

The nominal prediction of the branching fraction of $B^0_{(s)} \to \ell^+\ell^-$ decays in the SM does not include the effect of photon emission from the Feynman diagram [30]. Two kinds of photon emission occur: emission from the final-state muons, called Final State Radiation(FSR), and emission of photons from any other particle in the diagram, called Initial State Radiation(ISR). The interference between these two decays has been investigated and found to be negligible [30].

A $B^0_{(s)} \to \mu^+\mu^-$ decay including ISR proceeds similarly to semi-leptonic $b \to s\ell\ell$ decays, such as $B^0 \to K^{*0}\mu^+\mu^-$. The emitted ISR photon carries a significant amount of the momentum released in the decay, which means that theoretically, it behaves like a semi-leptonic decay: all Wilson coefficients play a role in its decay and it is not affected by helicity suppression. Experimentally, its dimuon invariant mass shape is spread over a wide range, from the dimuon mass threshold up to the B^0_s mass, so the vast majority of this process is rejected by the selection for the $B^0_{(s)} \to \mu^+\mu^-$ analysis. In the discussion on exclusive backgrounds in Section 4.6, this background is investigated and found to be negligible.

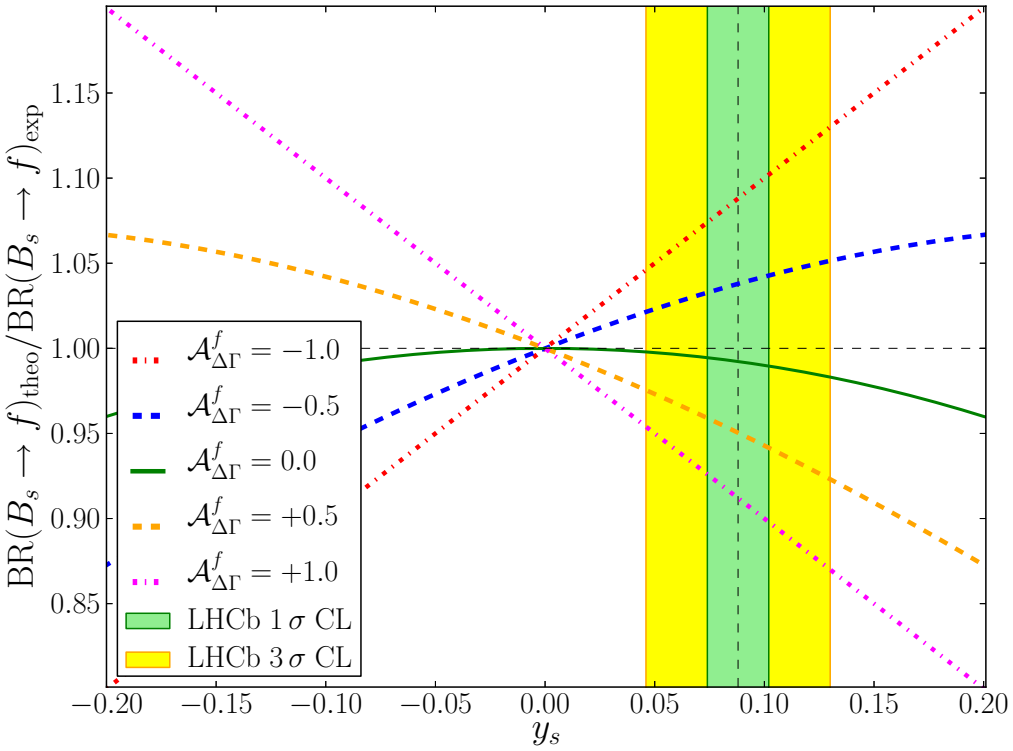

Figure 2.6: The conversion factor between the theoretical and experimental definition of a branching fraction, given a certain value of $A^f_{\Delta\Gamma_q}$ for a decay and y_s in the B^0_s-$\bar{B}^0_s$ system. The figure is taken from Ref. [26], and the green and yellow bands indicate the one and three sigma bands around the measured value for y_s at the time.

The FSR contribution behaves like a $B^0_{(s)} \to \ell^+\ell^-$ decay from a theoretical point of view, in that it has the same amplitude as $B^0_{(s)} \to \ell^+\ell^-$ without FSR. The emitted photon carries only a small fraction of the momentum, which means that in experiment it cannot be easily separated from $B^0_{(s)} \to \ell^+\ell^-$ decays without FSR, and behaves like an additional tail on the left-hand side of the $B^0_{(s)} \to \ell^+\ell^-$ mass shape. Therefore, it is an integral part of the $B^0_{(s)} \to \ell^+\ell^-$ branching fraction both theoretically and experimentally, and has to be included in both analyses.

Theoretically, the branching fraction is calculated without photon emission, which has been found to be equal to the branching fraction fully inclusive of FSR, up to a correction at the level of 1% [30], significantly below the current uncertainty on the branching fraction prediction.

In LHCb and other experiments, FSR is included in Monte Carlo simulations of $B^0_{(s)} \to \mu^+\mu^-$ decays with the photon emission generator PHOTOS [31]. The

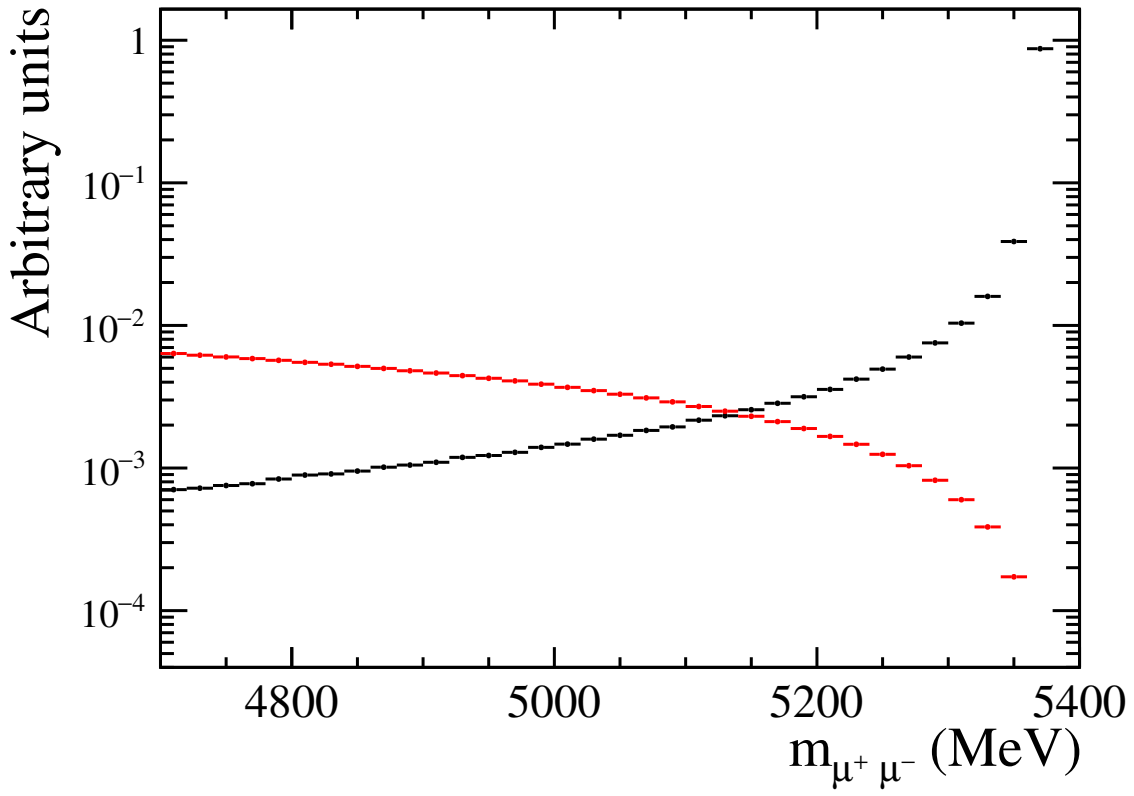

Figure 2.7: The dimuon invariant mass shape for the $B_s^0 \to \mu^+\mu^-$ decay including FSR (in black) and for the ISR process (in red), taken from LHCb simulation and scaled by their SM branching fraction.

branching fraction is then determined fully inclusive of FSR, which makes it equal to the theoretical branching fraction in this respect. The shapes for the simulated dimuon invariant mass in simulated samples used at LHCb, both for the signal including FSR and for the separately simulated ISR, are shown in Figure 2.7.

2.6.3 The branching fraction of $B^0_{(s)} \to \mu^+\mu^-$ in the SM

Correcting the theoretical branching fraction to be able to compare to the experimental results, the branching fraction of $B^0_{(s)} \to \ell^+\ell^-$ decays in the SM is usually determined by using the amplitude defined in Equation 2.21, squaring it and using

it in Equation 2.42:

$$\begin{aligned}\mathcal{B}(B^0_{(s)} \to \ell^+\ell^-) = \frac{G_F^2 M_W^4 \sin^4\theta_{\mathrm{W}}}{8\pi^5} \tau_{B^0_{(s)}} |V_{tq}^* V_{tb}|^2 f^2_{B^0_{(s)}} M_{B_q} m_\ell^2 \sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}} \\ \times |\mathcal{C}_{10}^{\mathrm{SM}}|^2 (|P^q_{\ell\ell}|^2 + |S^q_{\ell\ell}|^2) \left[\frac{1 + A^{\ell\ell}_{\Delta\Gamma_q} y_q}{1 - {y_q}^2}\right].\end{aligned} \tag{2.49}$$

In the Standard Model, $|P^q_{\ell\ell}| = 1, |S^q_{\ell\ell}| = 0, A^{\ell\ell}_{\Delta\Gamma_q} = 1$ and the expression in Equation 2.49 reduces to

$$\mathcal{B}(B^0_{(s)} \to \ell^+\ell^-)|_{\mathrm{SM}} = \frac{G_F^4 M_W^4 \sin^4\theta_{\mathrm{W}}}{8\pi^5} \tau_{B^0_{(s),H}} |V_{tq}^* V_{tb}|^2 f^2_{B^0_{(s)}} M_{B_q} m_\ell^2 \sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}} |\mathcal{C}_{10}^{\mathrm{SM}}|^2. \tag{2.50}$$

The main parametric uncertainties in this expression originate from the CKM elements and the decay constant $f_{B^0_{(s)}}$. Therefore, this method will be referred to as the CKM method.

A second method to calculate the branching fraction of $B^0_{(s)} \to \mu^+\mu^-$ is relative to the mass difference in neutral meson mixing Δm_q, which is described in Section 2.4, Equation 2.31, and which depends on the same CKM parameters. This method will be referred to as the Δm_q method. This relation holds in the Standard Model and more generally in MFV extensions of the SM [32], in which the flavour structure of nature is equivalent to the one found in the Standard Model [33]. The relation between Δm_q and $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ is given by

$$\begin{aligned}\mathcal{B}(B^0_{(s)} \to \ell^+\ell^-) = \frac{3 G_F^2 M_W^2 \sin^4\theta_{\mathrm{W}}}{8\pi^3} \tau_{B^0_{(s)}} m_\ell^2 \sqrt{1 - \frac{4m_\ell^2}{{M_{B_q}}^2}} \left[\frac{1 + A^{\ell\ell}_{\Delta\Gamma_q} y_q}{1 - {y_q}^2}\right] \Delta m_{q_{\mathrm{exp}}} \\ \times \frac{|\mathcal{C}_{10}^{\mathrm{SM}}|^2}{S_0(x_t)\eta_{2B}\hat{B}^1_{B_q}},\end{aligned} \tag{2.51}$$

where Equations 2.49 and 2.31 have been combined. For this method, the main parametric uncertainty is on the bag parameter $\hat{B}^1_{B_q}$, which is determined by theory.

In addition to choosing either method to determine $\mathcal{B}(B^0_{(s)} \to \ell^+\ell^-)|_{\mathrm{SM}}$, both methods require a choice for the input parameters. The branching fraction is calculated with both equations and different input parameters, in order to see how much the result varies and whether the results are consistent. The Wilson coefficient

$\mathcal{C}_{10}^{\mathrm{SM}}$ should be calculated for both methods, and therefore its calculation is first discussed.

Calculating the Wilson coefficient $\mathcal{C}_{10}^{\mathrm{SM}}$ for $B_{(s)}^0 \to \ell^+\ell^-$ In the SM, only the Wilson coefficient $\mathcal{C}_{10}$ contributes to $B_{(s)}^0 \to \ell^+\ell^-$ decays. Determining the value for this coefficient in the SM, $\mathcal{C}_{10}^{\mathrm{SM}}$, is essential for any $B_{(s)}^0 \to \ell^+\ell^-$ measurements, as it determines the scale of the Standard Model contributions. To leading order in the electroweak and QCD interactions, this coefficient is given as

$$\mathcal{C}_{10}^{\mathrm{SM}} = \frac{2\mathcal{C}_A^{\mathrm{SM}}}{\sin^2\theta_{\mathrm{W}}} = \frac{Y_0(x_t)}{\sin^2\theta_{\mathrm{W}}}, \tag{2.52}$$

where $Y_0(x_t)$ is a gauge-independent function of $x_t = (M_t/M_w)^2$, first determined by Inami and Lim [12], as introduced in this thesis in Section 2.1.2. It should be noted that the factor $\sin^2\theta_{\mathrm{W}}$ in this definition is canceled by its appearance in the branching fraction calculation in Equations 2.50, 2.51. Therefore, theoreticians directly determine $\mathcal{C}_A^{\mathrm{SM}}$ for the calculation of the $B_{(s)}^0 \to \ell^+\ell^-$ branching fraction. The most up-to-date calculations for $\mathcal{C}_A^{\mathrm{SM}}$, including NNLO QCD corrections and NLO QED corrections, are described in Ref. [34]. The difference between leading order and best estimate is a shift of -0.9%. More recently, power-enhanced QED corrections that are specific to $B_{(s)}^0 \to \ell^+\ell^-$ decays have been calculated as well [35], giving a further shift of -0.4%. Taking all of this into account, it is estimated that $\mathcal{C}_A^{\mathrm{SM}} = 0.4665 \pm 0.0020$.

Inputs for the CKM method For the CKM method, using Equation 2.50, the main uncertainty comes from the determination of $|V_{tq}^* V_{tb}|$. There are many possible inputs to determine this set of CKM elements from experiment; five inputs for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and two for $\mathcal{B}(B^0 \to \mu^+\mu^-)$.

For $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$, CKM elements are obtained with the Wolfenstein parametrisation from Equation 2.7. The Wolfenstein parameters are determined by performing a fit to all relevant experimental measurements. Here, the results from CKMFitter for ICHEP 2019 [10] are used. CKMFitter presents two determinations of the Wolfenstein parameters, one based on all available measurements and one based on measurements with tree-level decays only. Both sets of input parameters are used and referred to as "fitter all" and "fitter tree" respectively. The "fitter all" values are used for the nominal calculation of $\mathcal{B}(B^0 \to \mu^+\mu^-)$.

For $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, it is also possible to use a well measured ratio of CKM elements, namely

$$\frac{|V_{ts}V_{tb}|}{|V_{cb}|} = 0.982 \pm 0.004, \tag{2.53}$$

in which case the prediction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ requires an input for $|V_{cb}|$. There are two kinds of measurements of $|V_{cb}|$, which are referred to as inclusive and exclusive. Following CKMFitter [10] and the PDG [36], the exclusive $|V_{cb}|$ input is based on measurements using the so-called BGL scheme [37]. Note that a recent Belle result with the BGL scheme, not included in these $|V_{cb}|$ combinations, finds a smaller value for $|V_{cb}|$, around 9% less than the inclusive $|V_{cb}|$ value [38]. As $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ depends on ${V_{cb}}^2$, this highlights the need for a stable estimate of $|V_{cb}|$ for a proper estimate of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. Additionally, it is possible to average the results from exclusive and inclusive measurements, as they are consistent with each other. These three sets of inputs for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are referred to as $|V_{cb}|$ inclusive, exclusive and average. As the average measurement gives the highest precision, it is used for the nominal prediction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

In addition to the CKM elements, the CKM method requires an input for the decay constant $f_{B^0_{(s)}}$. This decay constant has been determined by multiple lattice QCD collaborations and averaged by the Flavour Lattice Averaging Group, or FLAG in short. The most recent averages by FLAG are determined in Ref. [39], where they find the decay constants to be

$$f_{B_s^0} = 230.3 \pm 1.3\,\text{MeV},\ f_{B^0} = (190.0 \pm 1.3)\ \text{MeV},\ f_{B_s^0}/f_{B^0} = 1.209 \pm 0.005. \tag{2.54}$$

Another average in the same paper does not use include loops of charm quarks in the lattice QCD calculations, and differs from the abovementioned results by about 1% for the individual decay constants and 0.7% for the ratio. This difference is small compared to the nominal uncertainty and thus neglected in the rest of this discussion.

Inputs for the Δm_q method For the Δm_q method, the bag parameter $\hat{B}^1_{B_q}$ is the relevant input and is determined by theory. The most precise and up-to-date results for this parameter have been obtained using either sum rules or lattice QCD [25, 40]. Both results are used and compared with each other.

Results and discussion All predictions for the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are shown in Table 2.2.

For $B^0_s \to \mu^+\mu^-$, the nominal BF prediction uses the CKM method with the average experimental value for $|V_{cb}|$ in order to determine $|V^*_{ts}V_{tb}|$:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-)|_{\text{SM}} = (3.58 \pm 0.13) \times 10^{-9}, \tag{2.55}$$

where the total uncertainty of 3.7% is dominated by $|V_{cb}|$ (3.4%), with small contributions from $f_{B^0_s}$ (1.1%) and m_t(0.7%) and the remainder at the level of 0.6%. The other predictions of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ are consistent with the nominal prediction within one standard deviation, although with larger uncertainties.

Comparing the CKM method predictions, it is interesting to note that varying the input for $|V_{cb}|$ affects the branching fraction at the level of 3%, close to the uncertainty of the nominal prediction. Therefore, a consistent picture will be required from the different $|V_{cb}|$ methods in the future to be fully confident of the predicted $B^0_s \to \mu^+\mu^-$ branching fraction. It is also interesting to note that the CKMFitter predictions have large and very asymmetrical uncertainties, which make them consistent with all other predictions, even though their central value is approximately equal to the $|V_{cb}|$ inclusive prediction.

The Δm_q method predictions are consistent with the nominal prediction, increasing the confidence in that prediction. Their uncertainties are fully dominated by the uncertainty on the bag parameter $\hat{B}^1_{B_q}$: with the more precise input for $\hat{B}^1_{B_q}$ from lattice QCD, which has an uncertainty of 4.3%, the BF prediction has an uncertainty of 4.4%.

The nominal prediction for the branching fraction of $B^0 \to \mu^+\mu^-$ uses the values for the Wolfenstein parameters determined by CKMFitter for ICHEP 2019 with all available measurements to determine $|V^*_{td}V_{tb}|$ and is

$$\mathcal{B}(B^0 \to \mu^+\mu^-)|_{\text{SM}} = (1.020 \pm^{+0.029}_{-0.055}) \times 10^{-10}, \tag{2.56}$$

where the total uncertainty of $+2.8, -5.4\%$ is dominated by the CKM elements $(+2.3, -5.2\%)$, with small contributions from $f_{B^0_s}$ (1.4%) and m_t(0.7%) and the remainder at the level of 0.5%. As for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$, the predictions based on the Wolfenstein parameters with the CKM method have significantly asymmetrical uncertainties. The "tree-level CKMFitter" prediction has a central value that differs by 9% from the nominal prediction, and is consistent with it at around 1.5 standard deviations, taking into account correlated inputs. The Δm_q method predictions

are consistent with the nominal prediction within one standard deviation. Their uncertainties are fully dominated by the uncertainty on the bag parameter $\hat{B}^1_{B_q}$: the uncertainty on the more precise input for $\hat{B}^1_{B_q}$ from lattice QCD has an uncertainty of 5.0%, compared to the BF prediction uncertainty of 5.0%.

Table 2.2: The branching fractions of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ as predicted in the SM. The different rows show results determined with different input parameter choices and with different methods. The conclusions are discussed in the text.

Method and parameter choice	$\mathcal{B}(B^0_s \to \mu^+\mu^-))$	$\mathcal{B}(B^0 \to \mu^+\mu^-))$
CKM, $\lvert V_{cb}\rvert$ inclusive	$(3.65 \pm 0.15) \times 10^{-9}$	-
CKM, $\lvert V_{cb}\rvert$ exclusive	$(3.48 \pm 0.19) \times 10^{-9}$	-
CKM, $\lvert V_{cb}\rvert$ average	$(3.58 \pm 0.13) \times 10^{-9}$	-
CKM, fitter all	$(3.69^{+0.07}_{-0.19}) \times 10^{-9}$	$(1.020^{+0.029}_{-0.055}) \times 10^{-10}$
CKM, fitter tree	$(3.67^{+0.09}_{-0.27}) \times 10^{-9}$	$(1.111^{+0.059}_{-0.096}) \times 10^{-10}$
Δm_q, $\hat{B}^1_{B_q}$ lattice	$(3.73 \pm 0.16) \times 10^{-9}$	$(1.016 \pm 0.051) \times 10^{-10}$
Δm_q, $\hat{B}^1_{B_q}$ sum rules	$(3.53 \pm 0.21) \times 10^{-9}$	$(0.942 \pm 0.057) \times 10^{-10}$

Particularly clean predictions are obtained for the ratio of branching fractions, as in this ratio uncertainties on the relative CKM elements, decay constants, bag parameters, and the Wilson coefficient partially cancel. Taking Equation 2.49, the ratio of branching fractions is given by

$$\frac{\mathcal{B}(B^0 \to \ell^+\ell^-)}{\mathcal{B}(B^0_s \to \ell^+\ell^-)} = \frac{\tau_{B^0} M_{B_d} f^2_{B^0} |V_{td}|^2}{\tau_{B^0_{s,H}} M_{B_s} f^2_{B^0_s} |V_{ts}|^2} \tag{2.57}$$

or alternatively, using Equation 2.51, by

$$\frac{\mathcal{B}(B^0 \to \ell^+\ell^-)}{\mathcal{B}(B^0_s \to \ell^+\ell^-)} = \frac{\tau_{B^0} \hat{B}^1_{B_d} {\Delta m_d}^2}{\tau_{B^0_{s,H}} \hat{B}^1_{B_s} {\Delta m_s}^2}, \tag{2.58}$$

where the kinematic term involving the mass of the leptons is neglected because it modifies this ratio by a negligible amount for $B^0_{(s)} \to \mu^+\mu^-$ decays, given the small muon mass.

These predictions on the ratio of branching fractions do not just hold for the Standard Model, but more generally for any theory that includes Minimal Flavour Violation (MFV). It is thus an important test of the flavour structure in Nature.

The ratio is calculated using the same inputs as were available for the calculation of $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The nominal prediction uses the Wolfenstein parameter inputs

from CKMFitter based on all available measurements and is

$$\frac{\mathcal{B}(B^0 \to \ell^+\ell^-)}{\mathcal{B}(B_s^0 \to \ell^+\ell^-)} = (2.775 \pm 0.069)\% \tag{2.59}$$

The predictions of the ratio are compared in Table 2.3. The other predictions of the ratio are consistent with the nominal prediction within two standard deviations. The prediction using the CKMFitter tree-only inputs differs by 10 to 14% from the others in central value, and is consistent at 1.8 standard deviations with the nominal prediction. It will be interesting to see whether this difference between the two CKMFitter estimates is resolved in the future. The results with the Δm_q method are close in central value to the nominal prediction, but the result using sum rules for the bag parameter $\hat{B}^1_{B_q}$ has such a high precision that it still disagrees with the nominal prediction by 1.7 standard deviations. It will be very interesting to measure this ratio to high precision in the future, both to understand the current tension between theory predictions and to serve as a test of Minimal Flavour Violation. The most precise result is the prediction using the Δm_q method and the relative $\hat{B}^1_{B_q}$ from sum rules, with an uncertainty of only 1.0%, dominated by the ratio of bag parameters $\hat{B}^1_{B_q}$ (0.8%), while the ratio of Δm_q (0.4%) and $\tau_{B^0_{(s)}}$ (0.4%) also contribute.

Table 2.3: The ratio of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fractions as predicted in the SM, determined with different input parameter choices and methods.

Method and parameter choice	$\mathcal{B}(B^0 \to \mu^+\mu^-)/\ \mathcal{B}(B_s^0 \to \mu^+\mu^-)$
CKM, fitter all	$(2.775 \pm 0.069)\%$
CKM, fitter tree	$(3.040 \pm 0.150)\%$
Δm_q, $\hat{B}^1_{B_q}$ lattice	$(2.725 \pm 0.070)\%$
Δm_q, $\hat{B}^1_{B_q}$ sum rules	$(2.668 \pm 0.027)\%$

2.6.4 Experimental status and constraints on New Physics

In this Section, the experimental status of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions before the measurement described in this thesis is discussed, and compared with the SM predictions given in the previous Section.

The two most up-to-date results at the time were obtained with the full Run 1 data, namely a joint publication by LHCb and CMS [41] and a publication by ATLAS [42].

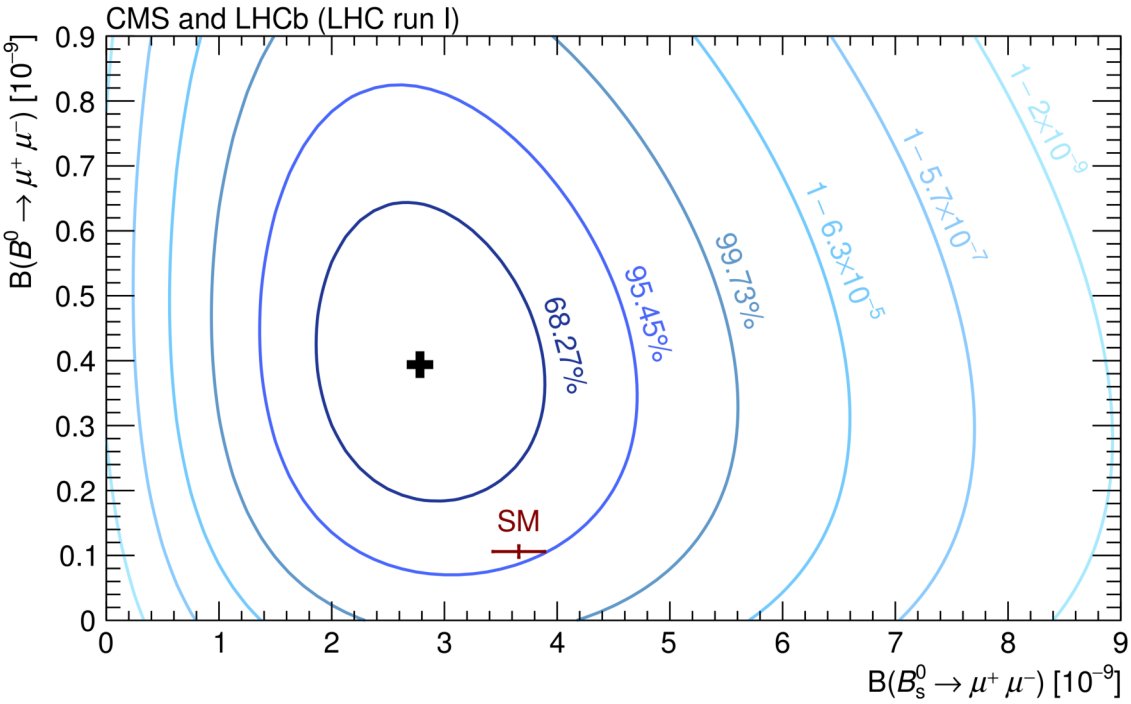

Figure 2.8: The result of the combined LHCb+CMS measurement in the two-dimensional $\mathcal{B}(B_s^0 \to \mu^+\mu^-) - \mathcal{B}(B^0 \to \mu^+\mu^-)$-plane. The Standard Model prediction, including uncertainties, is indicated by the red cross. The central value from the measurement is shown as a black plus sign, and the profile likelihood contours corresponding to a confidence level associated with a 1,2,3,...σ deviation are shown in blue.

The result from the combined LHCb+CMS measurement with the full Run 1 dataset was

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9}, \tag{2.60}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}, \tag{2.61}$$

$$\frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} = 0.14^{+0.08}_{-0.06}. \tag{2.62}$$

The accompanying two-dimensional likelihood of the measurement, compared to the SM prediction at the time, is shown in Figure 2.8, and shows that the two branching fraction measurements were largely uncorrelated. The $B_s^0 \to \mu^+\mu^-$ decay was observed with a significance of 6.2 standard deviations, constituting the worlds first observation of this decay, and evidence for the $B^0 \to \mu^+\mu^-$ decay was found with a significance of 3.2 standard deviations. Compared to the nominal SM predictions in this thesis in Equations 2.55, 2.56, the measured $B_s^0 \to \mu^+\mu^-$ branching fraction is compatible within 1.1 standard deviations, the $B^0 \to \mu^+\mu^-$ branching fraction at 2.3 standard deviations, and the relative branching fraction at 2.4 standard deviations.

The ATLAS collaboration also published a result with the full Run 1 dataset. For the $B^0 \to \mu^+\mu^-$ branching fraction, they found a negative central value. Restricting the $B^0 \to \mu^+\mu^-$ branching fraction to be positive or equal to zero, they found

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}, \tag{2.63}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.2 \times 10^{-10} \text{at } 95\% \text{ CL}. \tag{2.64}$$

The ratio of the two branching fractions was not reported because only a limit was set on the branching fraction of $B^0 \to \mu^+\mu^-$, and the central value of the $B^0 \to \mu^+\mu^-$ branching fraction was negative. Clearly, both the uncertainties were quite large compared to the combined LHCb+CMS measurement. The two-dimensional likelihood of the ATLAS measurement, compared to the combined LHCb+CMS measurement and the SM prediction at the time is shown in Figure 2.9. The two measurements as well as the SM agreed with each other at the level of around two standard deviations. Given the large uncertainties, large correlation between the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fraction, and negative central value for $\mathcal{B}(B^0 \to \mu^+\mu^-)$, the ATLAS result is not considered in the rest of this Section.

Using only the branching fraction information from the combined LHCb+CMS measurement, some assumption has to be made for $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ to convert the experimental measurement into a theoretical one. Here, $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ is assumed to be +1, as in the SM. In this case, a measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fraction constrains

$$R_{B_{(s)}^0 \to \mu^+\mu^-} = \frac{\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)_{\text{exp}}}{\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)_{\text{exp}}|_{\text{SM}}} = |P^q_{\mu^+\mu^-}|^2 + |S^q_{\mu^+\mu^-}|^2, \tag{2.65}$$

where the estimate of $R_{B_{(s)}^0 \to \mu^+\mu^-}$ is affected by the uncertainties on the SM prediction and on the experimental measurement. Using the LHCb+CMS measurement gives

$$R_{B_s^0 \to \mu^+\mu^-} = 0.78^{+0.20}_{-0.18}, \tag{2.66}$$

$$R_{B^0 \to \mu^+\mu^-} = 3.8^{+1.6}_{-1.4}, \tag{2.67}$$

$$\frac{R_{B^0 \to \mu^+\mu^-}}{R_{B_s^0 \to \mu^+\mu^-}} = 5.0^{+2.9}_{-2.2}, \tag{2.68}$$

which implies that the sum of the two amplitudes in $B_s^0 \to \mu^+\mu^-$ decay is consistent with the SM prediction, whereas the central value of the branching fraction ratio implies a strong breaking of Minimal Flavour Violation. As the uncertainties on $\mathcal{B}(B^0 \to \mu^+\mu^-)$-related observables were still large, more data was needed before drawing strong conclusions on them.

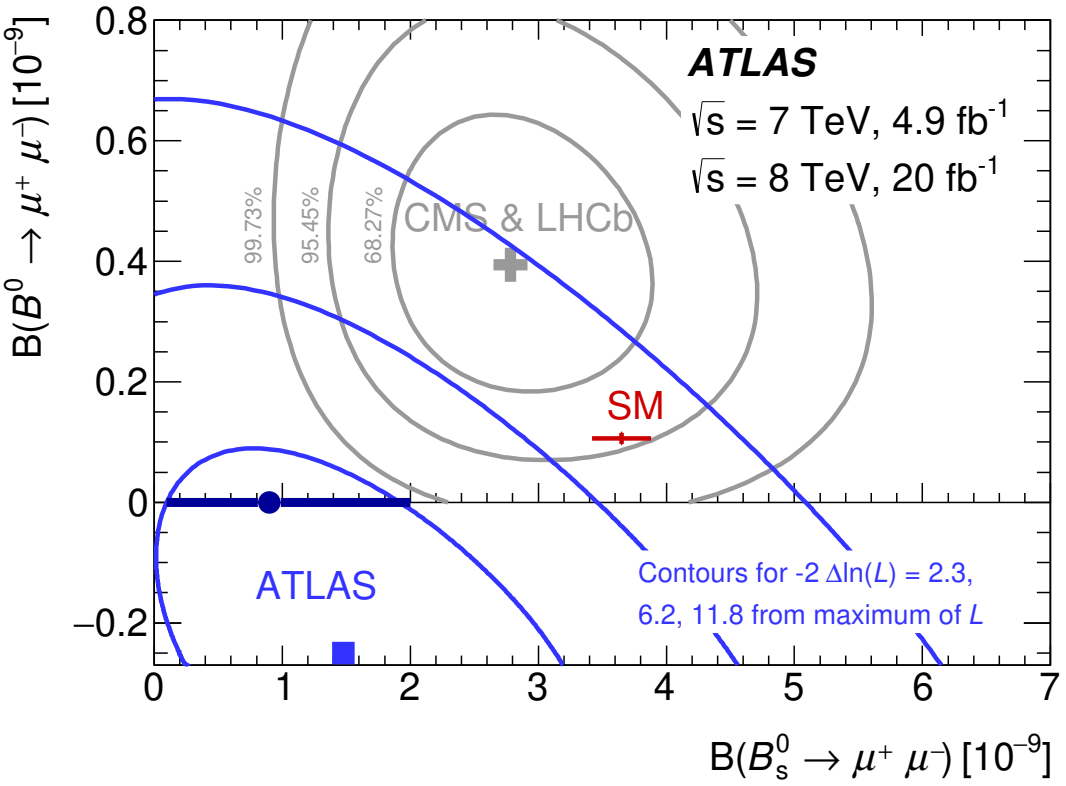

Figure 2.9: The result of the ATLAS measurement with Run 1 data in the two-dimensional $\mathcal{B}(B_s^0 \to \mu^+\mu^-) - \mathcal{B}(B^0 \to \mu^+\mu^-)$-plane. The Standard Model prediction, including uncertainties, is indicated by the red cross. The central value from the measurement is shown as the blue box, and the profile likelihood contours corresponding to a confidence level associated with a 1,2,3,...σ deviation are shown in blue. The CMS+LHCb measurement with Run 1 data is shown in grey.

There are two ways to interpret the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction. Firstly, by assuming that only $\mathcal{C}_{10}$ is affected by new physics, implying $\mathcal{C}_{10,\mathrm{NP}} = 0.5 \pm 0.5$, both consistent with the SM as well as with a central value similar to that obtained from a global fit to the other $b \to s\ell\ell$ measurements, like the one shown in Figure 2.4. An example of comparing the result from $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ with global fits in this scenario is given in Ref. [43] and shown in Figure 2.10. Secondly, assuming only new physics from scalar or pseudoscalar contributions, the difference in the total contribution from $\mathcal{C}_P - \mathcal{C}'_P$ and $\mathcal{C}_S - \mathcal{C}'_S$ is constrained at $< 0.029\,\mathrm{GeV}^{-1}$ at a 95% confidence level. The $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement was also combined with other $b \to s\ell\ell$ measurements, for example to fully constrain (pseudo-)scalar new physics as in Ref.[44].

Because of the tensions in other $b \to s\ell\ell$ decays, the different central values of the measurements and the SM prediction, and the theoretically clean predictions

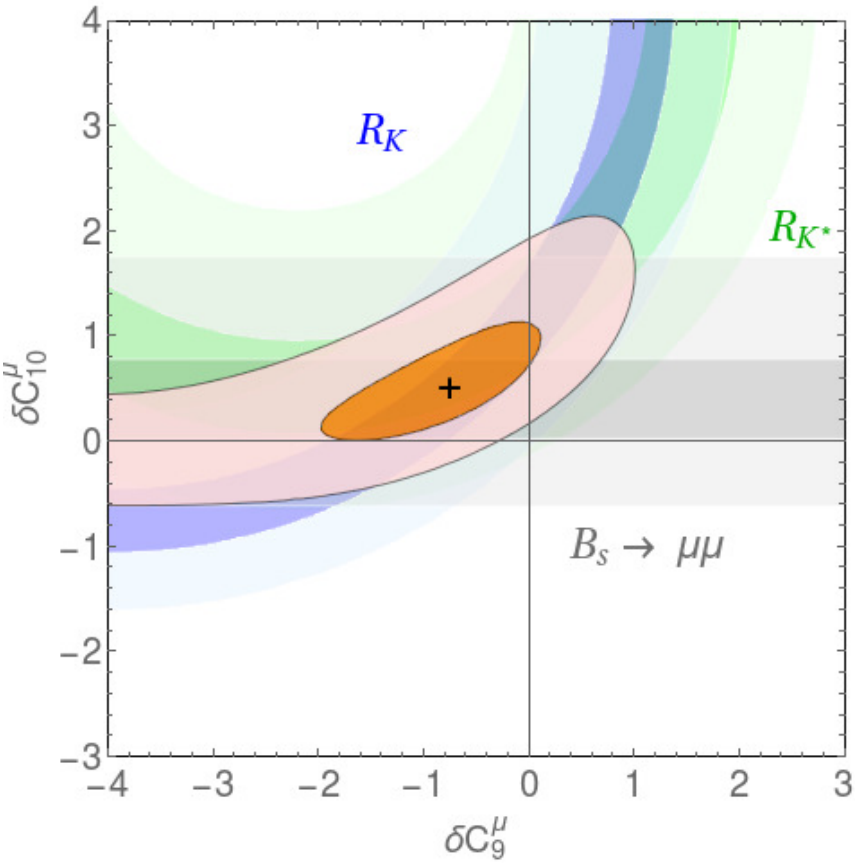

Figure 2.10: Constraints on New Physics contributions to $\mathcal{C}_9$ and $\mathcal{C}_{10}$ from a fit to $X_b \to X_q \mu^+\mu^-$ data, assuming both are real and that other Wilson coefficients are as in the SM, as indicated by orange contours at the level of one and three standard deviations. Individual constraints from R_K, R_K^* (which are tests of lepton flavour universality in semileptonic $b \to s\ell\ell$ decays, $R_{K^{(*)}} = \mathcal{B}(B \to K^{(*)}\mu^+\mu^-)/\mathcal{B}(B \to K^{(*)}e^+e^-)$), and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are indicated with labeled contours at the level of one and three standard deviations as well. Figure taken from Ref. [43].

for $B^0_{(s)} \to \mu^+\mu^-$ decays, it was very interesting to measure the branching fraction of both $B^0_{(s)} \to \mu^+\mu^-$ decays with more data. This directly motivates the analysis in this thesis, in which the worlds most precise measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions will be presented in Section 5.1.

2.7 The effective lifetime of $B_s^0 \to \mu^+\mu^-$

As introduced in Section 2.5, the effective lifetime is an observable which probes the $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ parameter in $B_s^0 \to \mu^+\mu^-$ decays. It is the second observable to examine for $B_s^0 \to \mu^+\mu^-$ decays, after having observed the decay and having measured its branching fraction.

In practice, measuring $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ directly or $\tau_{\text{eff}}^{\mu^+\mu^-}$ is just a convention, as they are related to each other through variables with negligible uncertainties. In this thesis the choice is made to analyse the implications of this measurement with $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$, which simplifies the equations on the theoretical side.

An instructive approach is to assume that there are no new *CP*-violating phases present and see how $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are connected, as is done for example in Ref. [45]. In this case, $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ from Equation 2.43 reduces to

$$A_{\Delta\Gamma_s}^{\mu^+\mu^-} = \frac{|P^s_{\mu^+\mu^-}|^2 - |S^s_{\mu^+\mu^-}|^2}{|P^s_{\mu^+\mu^-}|^2 + |S^s_{\mu^+\mu^-}|^2}, \tag{2.69}$$

and the following expressions are found for $|P^s_{\mu^+\mu^-}|^2$ and $|S^s_{\mu^+\mu^-}|^2$:

$$|P^s_{\mu^+\mu^-}|^2, |S^s_{\mu^+\mu^-}|^2 = \frac{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\text{exp}}}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\text{exp,SM}}} \frac{(1+y_s)(1 \pm A_{\Delta\Gamma_s}^{\mu^+\mu^-})}{2(1 + A_{\Delta\Gamma_s}^{\mu^+\mu^-} y_s)}, \tag{2.70}$$

where the + sign is valid for $|P^s_{\mu^+\mu^-}|$ and the − for $|S^s_{\mu^+\mu^-}|$; it should be noted that in the SM $P^s_{\mu^+\mu^-} = 1$ and $S^s_{\mu^+\mu^-} = 0$. This expression shows that measuring $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ allows to separate the pseudoscalar-type amplitude from the scalar-type amplitude, and that measuring any deviation from $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = +1$ directly implies new (pseudo-)scalar contributions or *CP*-violating effects in $B_{(s)}^0 \to \mu^+\mu^-$ decays, as noted in Ref. [28].

Another important reason to measure $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ is its effect on the theoretical as well as the experimental interpretation of the branching fraction. Theoretically, the conversion from the experimental to the theoretical definition of the branching fraction, as detailed in Section 2.6.1, can vary by up to 14% between the two extremes for $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$. Experimentally, the selection efficiency of events depends on the assumed lifetime of the $B_{(s)}^0$ meson candidate. Therefore, the estimate of the branching fraction depends on the lifetime of the decay under investigation. Within the theoretically allowed range of $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$, this can increase the experimental branching fraction estimate by around 11%. These two effects affect the interpretation of

$B_s^0 \to \mu^+\mu^-$ in the same relative direction. In the most extreme case, they could affect the ratio of measurement and prediction by up to 27%, at the level of the experimental uncertainty on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in the LHCb+CMS measurement.

Therefore, it is very interesting to measure the effective lifetime in $B_s^0 \to \mu^+\mu^-$ decays and probe new physics that would not be accessible with only the measurement of the branching fraction. In this thesis, the worlds first measurement of the effective lifetime will be presented in Section 5.2.5.

2.8 *CP* violation in $B^0_{(s)} \to \mu^+\mu^-$ decays

As discussed in Section 2.5, there are two observables in $B^0_{(s)} \to \ell^+\ell^-$ decays that probe CP violation: $C_{\ell\ell}$ and $S_{\ell\ell}$. Both are predicted to be zero with negligible uncertainties in the SM.

It should be noted that $C^\lambda_{\ell\ell}$ depends on the helicity of the final state leptons, which makes it experimentally very difficult to measure at the LHC for electrons and muons, while it might feasible for tau leptons [46]. Alternatively, $C_{\ell\ell}$ can estimated by measuring $S_{\ell\ell}$ and $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$, as the definitions of these observables in Equation 2.37 imply the constraint ${C_{\ell\ell}}^2 + {S_{\ell\ell}}^2 + {A^{\mu^+\mu^-}_{\Delta\Gamma_s}}^2 = 1$. Note that this would leave the sign of $C_{\ell\ell}$ ambiguous.

Without measuring the helicity, what is measured in experiment is the sum of both helicity states,

$$\Gamma(B^0_{(s)} \to \ell^+\ell^-) = \Gamma(B^0_{(s)} \to \mu^+_L\mu^-_L) + \Gamma(B^0_{(s)} \to \mu^+_R\mu^-_R), \tag{2.71}$$

and terms with $C^\lambda_{\ell\ell}$ cancel, as their sign is opposite for these two decay rates. The sensitivity for $S_{\ell\ell}$ remains, using the time-dependent CP-asymmetry defined in Equation 2.41. Even for this asymmetry, the required statistics are very large, such that neither $S_{\ell\ell}$ nor $C_{\ell\ell}$ are expected to be measured in the near future. In Section 6.2, the prospects for CP violation measurements with $B^0_{(s)} \to \mu^+\mu^-$ decays in the upgraded LHCb experiments are discussed.

2.9 Other $B^0_{(s)} \to \ell^+\ell^-$ decays

In addition to measurements with $B^0_{(s)} \to \mu^+\mu^-$ decays, it is interesting to measure the same observables in $B^0_{(s)} \to \tau^+\tau^-$ and $B^0_{(s)} \to e^+e^-$ decays. Either decay probes different aspects of leptonic rare B decays, as helicity suppression affects $B^0_{(s)} \to \tau^+\tau^-$ decays less and $B^0_{(s)} \to e^+e^-$ decays far more than $B^0_{(s)} \to \mu^+\mu^-$. Therefore, they each provide their own constraints.

The main role for $B^0_{(s)} \to \tau^+\tau^-$ decays is to probe lepton universality. Assuming lepton universality, the measurement of $B^0_s \to \mu^+\mu^-$ at the level of its SM branching fraction constrains $\mathcal{C}_{10}$ to be close to the SM value and new (pseudo-)scalar contributions to be small, strongly constraining the branching fraction of $B^0_{(s)} \to \tau^+\tau^-$. Without assuming lepton universality, $B^0_{(s)} \to \tau^+\tau^-$ decays are very interesting to investigate. As motivated by the lepton non-universality anomalies [47], new contributions to $\mathcal{C}^{\tau\tau}_{10}$ can possibly increase its branching fraction by two or three orders of magnitude.

On the other hand, the SM contributions to $B^0_{(s)} \to e^+e^-$ decays are so strongly suppressed that they are not affected by constraints from any $b \to s\ell\ell$ measurements. Any observation of $B^0_{(s)} \to e^+e^-$ in the foreseeable future would therefore be clear sign of (pseudo-)scalar new physics contributions, as detailed in Ref. [48].

Chapter 3

The Large Hadron Collider and the LHCb detector

The data used in the analysis described in this thesis was collected by the LHCb experiment. At the LHCb interaction point, proton beams were collided after being accelerated to 6.5 TeV/c^2 by the Large Hadron Collider (LHC).

In this Chapter, the LHC and the LHCb detector are described. The Large Hadron Collider and the LHCb interaction point are introduced in Section 3.1. The LHCb detector is described in Section 3.2 including the motivation for its construction, the layout of each of its subsystems, and its performance in the context of the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The conversion of raw data obtained with the LHCb detector to the information required for the $B^0_{(s)} \to \mu^+\mu^-$ analysis is discussed in Sections 3.3 to 3.6; the trigger system, the offline reconstruction and central selection, the simulation and the particle identification algorithms.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a 27 km long hadron collider, located close to Geneva and passing through France and Switzerland [49], at the Conseil Europeen pour la Recherche Nucleaire (CERN). The LHC was designed to collide beams of ions, mainly protons, at a centre-of-mass energy of 14 TeV with a peak luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}s^{-1}$.

The LHC has collided protons in two sets of runs, referred to as Run 1 and Run 2. These lasted from 2010 to 2012 and 2015 to 2018 respectively. In Run 1, the collisions took place at a centre-of-mass energy of 7 TeV in 2011 and 7 TeV in 2012, while in

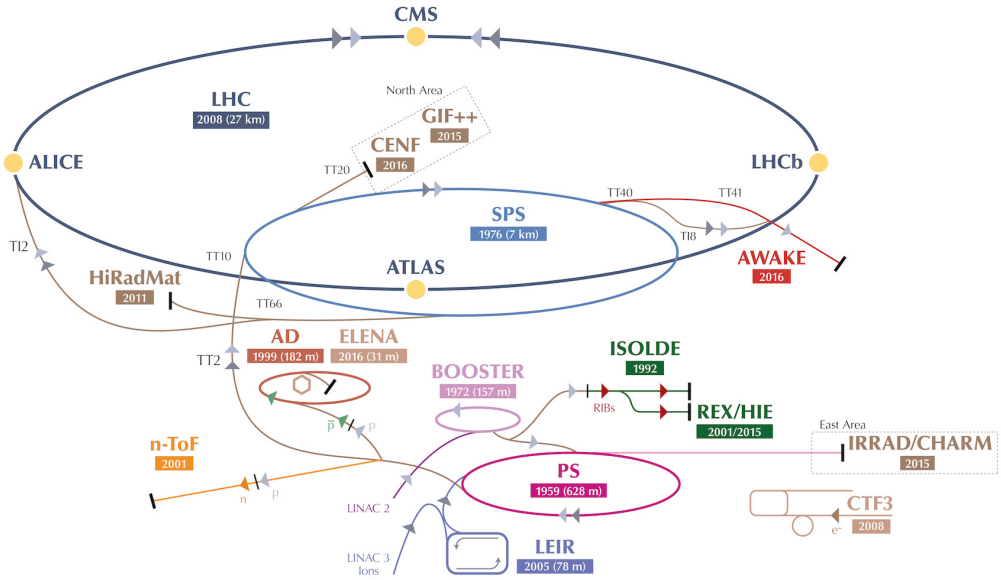

Figure 3.1: The CERN accelerator complex, showing the chain of accelerators used to inject protons into the LHC, as well as the other experiments using protons from those accelerators.

Run 2 collisions took place at a centre-of-mass energy of 13 TeV, almost reaching the design energy of 14 TeV. Before being injected into the LHC ring, bunches of about 1×10^{11} protons first pass through the CERN accelerator complex, during which they are accelerated up to an energy of 450 GeV each. The CERN accelerator complex is illustrated in Figure 3.1. The LHC operates with a spacing between bunches, which was 50 ns in Run 1 and 25 ns in Run 2. Accordingly, in Run 1 the LHC ring had a capacity of about 1400 bunches, while in Run 2 it could operate with 2732 bunches circulating. After injecting bunches, the LHC accelerates them to the collision energy, squeezes the beams to obtain the required luminosity, and collides them at the four interaction points of the main LHC experiments.

The ATLAS and CMS experiments were built to confirm or exclude the existence of the Higgs particle predicted in the Standard Model, and to search for the production of particles from theories Beyond the Standard Model (BSM). In July 2012, the experiments announced that they had observed a Higgs-like particle [50, 51], leading to the 2013 Nobel Prize in Physics being awarded to François Englert and Peter Higgs for the theoretical discovery of the Brout-Englert-Higgs mechanism. ATLAS and CMS have led the first investigations of Nature at the TeV scale and strongly increased the bounds on the direct production of BSM particles.

The ALICE experiment was designed to study the quark-gluon plasma; a state of matter of the strong interaction. This state of matter occurs when a system of strongly interacting matter has a high energy density, as happened during the Big Bang. By colliding heavy ions such as lead and gold, the LHC can recreate those conditions, which ALICE studies and compares with proton-proton collisions.

The LHCb experiment, which was built to investigated b-hadron decays, and for which the LHC conditions are discussed here. The LHCb detector was optimised to run with an average of one visible pp interaction per two bunch crossings, equivalent to $\mathcal{L} \sim 2 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$. As the LHC delivers significantly higher luminosities, it adjusts the separation and angle between the colliding beams at the LHCb interaction point during fills to keep the luminosity leveled around this average. In practice, LHCb has run with more visible pp interactions per bunch crossing than it was designed for, varying it over different years to make optimal use of the available computing resources and converging on a luminosity leveled at around $\mathcal{L} \sim 4 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$. The conditions of pp collisions at LHCb, for the data included in this thesis, are shown in Table 3.1. As the $b\bar{b}$ cross-section is expected to scale linearly with energy, the Run 2 dataset used here contains about 80% of the b hadron yield relative to the Run 1 dataset.

Table 3.1: The pp collision conditions at the LHCb experiment for each dataset used in this thesis, including data taken up to September 2016 for the $B^0_{(s)} \to \mu^+\mu^-$ analysis.

	Year			
	2011	2012	2015	2016
$\sqrt{s}$ (TeV)	7	8	13	13
$\int \mathcal{L}$ (fb^{-1})	1.11	2.08	0.33	1.10
Median $\mathcal{L}$ ($10^{32}\,\text{cm}^{-2}\text{s}^{-1}$)	3.4	3.9	3.4	3.7
μ (vis. pp int. / bunch crossing)	1.4	1.7	1.1	1.1
n_{bunches}	1320	1262	1866	2036
f_{rev}	11.2 kHz			
$f_{\text{collision}}$(MHz)	14.8	14.1	20.9	22.8

3.2 The LHCb detector

The motivation for building the LHCb detector was to perform precision measurements in beauty hadron decays, including differences between matter and antimatter,

known as CP violation, and to search for rare decays of beauty hadrons, for example the $B^0_{(s)} \to \mu^+\mu^-$ decays that are the topic of this thesis.

The LHCb experiment has produced a rich ensemble of physics measurements, including but not limited to

- CP violation measurements, discovering neutral meson mixing in the oscillations of B^0_s and D mesons and CP violation in the D meson system.
- Rare decays of beauty, charm and strange hadrons, with as highlight the observation of $B^0_{(s)} \to \mu^+\mu^-$ by a combined LHCb and CMS measurement with Run 1 data, and more recently hints of deviations in other measurements of the $b \to s\ell\ell$ transition, such as branching fractions of $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0_s \to \phi\mu^+\mu^-$, the angular observable P'_5 in $B^0 \to K^{*0}\mu^+\mu^-$ and lepton universality ratios in $B^+ \to K^+\ell^+\ell^-$ decays for muons relative to electrons.
- Semileptonic decays of beauty hadrons, which were expected to be too difficult for LHCb because of the neutrino in the final state, but have been used for measurements of CP violation, b hadron production fractions, $|V_{ub}|$ and lepton universality ratios with tau leptons versus muons, the last of which show hints of deviations from the SM.
- Spectroscopy of hadrons with beauty or charm quarks, including the observation of exotic QCD states in which four or five quarks are bound together, called tetra- and pentaquarks respectively.
- Searches for BSM physics from direct production of new particles such as dark photons or Majorana neutrinos, complementing measurements at ATLAS and CMS.
- Fixed target measurements performed by injecting gases such as helium into the detector, used by experiments that search for dark matter by measuring the amount of anti-matter in the universe.
- Heavy ion physics, measuring charm and beauty cross sections and complementing the measurements at ALICE.

In pp collisions at the LHC collision energies, $b\bar{b}$ pairs are mainly produced in the same forward or backward cone with respect to the direction of the beam pipe of the LHC, motivating the overall geometry of the LHCb detector [52]. The b hadrons in the forward direction have a significant boost of around 16 on average, which corresponds to an average decay length of 0.7 mm. This is sufficient to separate a b

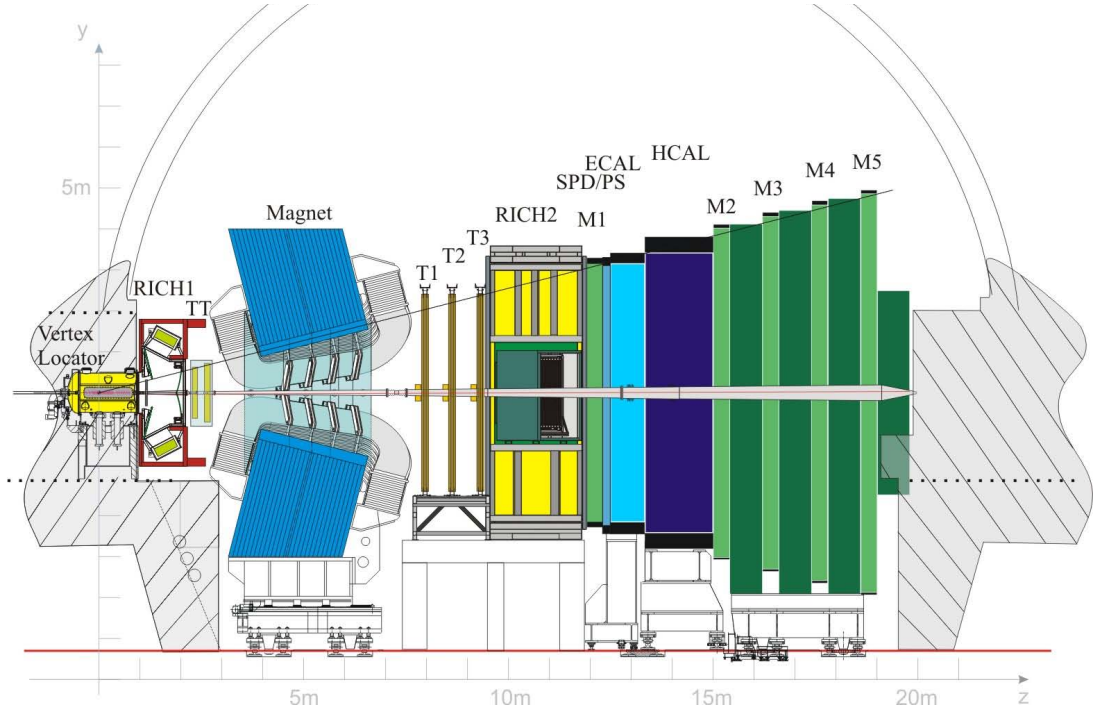

Figure 3.2: The LHCb detector in the yz-plane, showing each subdetector and the scale of the detector. Figure taken from Ref [52].

hadron decay vertex from the pp collision vertex, which is essential for the suppression of backgrounds.

All kinds of b hadrons are produced at the LHC, including B^+, B^0 and B_s^0 mesons and Λ_b^0 baryons. Due to isospin symmetry, B^+ and B^0 mesons are produced at similar rates (in this thesis, their production rates are assumed to be equal), while B_s^0 mesons are produced less because the strange quark is heavier. To measure the branching fraction of a given decay, the total number of produced b hadrons has to be determined. At LHCb, this is done by measuring branching fractions relative to decays with well-measured branching fractions, which are only those of B^+ and B^0 mesons (as those were measured at the B-factories, where B^+ and B^0 mesons were produced in equal rates). As a result, the production rate of of B_s^0 mesons relative to B^0 mesons, referred to as f_s/f_d, has to be measured for the normalisation of the $B_s^0 \to \mu^+\mu^-$ branching fraction. With 2011 data, at $\sqrt{s} = 7\,\text{TeV}$, it was found that $f_s/f_d = 0.259 \pm 0.015$ [53]. No measurement of f_s/f_d at 8 or 13 TeV was available for the $B_{(s)}^0 \to \mu^+\mu^-$ analysis, and its estimation for the 2012 and Run 2 datasets is discussed in Section 4.5.7.

The LHCb detector is shown in Figure 3.2. The coordinate system used is a right-handed Cartesian coordinate system x, y, z, where the z axis is pointing along

the beamline towards the muon stations, the y axis along the vertical and x axis away from the LHC ring. Alternatively, a cylindrical coordinate system is based on the same z axis. The two angles in this system are defined as the polar angle θ, relative to the z axis, and as the transverse azimuthal angle ϕ relative to the x axis, increasing in the direction of the y axis. The pseudorapidity η is defined as

$$\eta = -\ln\tan(\theta/2). \tag{3.1}$$

It is often used as differences in pseudorapidity are Lorentz invariant under boosts in the z direction in the ultra-relativistic limit. The LHCb detector acceptance covers the full range in ϕ and $\eta \in [2, 5]$, corresponding to $\theta \in [0.772°, 15.4°]$, which illustrates the geometry.

To reach the physics goals that motivated its construction, the LHCb detector had the following design requirements:

- Tracking system: reconstruct the tracks of charged particles originating from b hadron decays with high efficiency, to be able to measure decays with many final state particles. The relevant subdetectors are the Vertex Locator (VELO), Tracker Turicensis (TT), Inner Tracker (IT) and Outer Tracker (OT). The VELO reconstructs the primary and decay vertices from these tracks, to measure time-dependent properties of b and c hadrons and reject backgrounds. Using the other tracking detectors and the field of a dipole magnet, the momenta and invariant masses of particles are precisely measured, to suppress backgrounds and to separate, for example, B^0 from B^0_s decays. The tracking system and the magnet are described in Section 3.2.1.

- Particle identification (PID) system: determine the identity of charged particles with the Ring Imaging Cherenkov (RICH) and muon detectors. PID of muons is needed to find rare b hadron decays, such as $B^0_{(s)} \to \mu^+\mu^-$, and PID for hadrons is used to separate different final states, such as $B^0_{(s)} \to h^+h^{(\prime)-}$, where $h \in \pi^\pm, K^\pm$. The RICH detectors measure the velocity of charged particles, which combined with the momentum of the track gives an estimate of the mass, and thus the identity, of all charged particles. They are described in Section 3.2.2. As muons are minimally interacting particles at LHCb, most of them traverse the calorimeter system, while other charged particles are stopped. The muon stations reconstruct tracks after the calorimeter to obtain a clean signature of muons, both in the trigger and offline. They are described in Section 3.2.4.

- Calorimeter system: measure the energy of electrons, photons, and hadrons. The calorimeters serve as the source of all hardware triggers except the muon trigger. For example, the electronic trigger is used in lepton universality tests for b hadron decays containing electrons and the hadronic trigger on $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, which is a calibration channel for the $B^0_{(s)} \to \mu^+\mu^-$ analysis. For electromagnetically interacting particles, the Electronic Calorimeter (ECAL) system is used, consisting of the Scintillating Pad Detector (SPD), Pre-Shower Detector (PS), and the ECAL itself. The Hadronic Calorimeter (HCAL) is used to measure the energy of hadrons. The calorimeter system is described in Section 3.2.3.

- Trigger system: read out the detector and process the data from the collisions coming in with a frequency of $\sim 20\,\text{MHz}$ in order to bring down the data rate to around 10 kHz. The trigger system is described in Section 3.3.

In the subsequent sections, the subdetectors that compose the LHCb detector are described in more detail.

3.2.1 Tracking system and magnet

In LHCb, track types are defined based on which subdetectors are used to reconstruct them. The tracking stations downstream of the magnet (which contain the IT and the OT) are referred to as T stations in short. Figure 3.3 shows the different track types:

- Long tracks, which have hits in the VELO and the T stations, and possibly in the TT.
- VELO tracks, which are only reconstructed in the VELO.
- Downstream tracks, with hits in the TT and T stations.
- Upstream tracks, which have hits in the VELO and TT.
- T tracks, which only have a reconstructed segment in the T stations.

Most physics analyses use only long tracks to build their signal candidates, as b hadrons decay in the VELO and the momenta of the tracks are optimally measured if they traverse the full spectrometer. VELO tracks are used to reconstruct primary vertices.

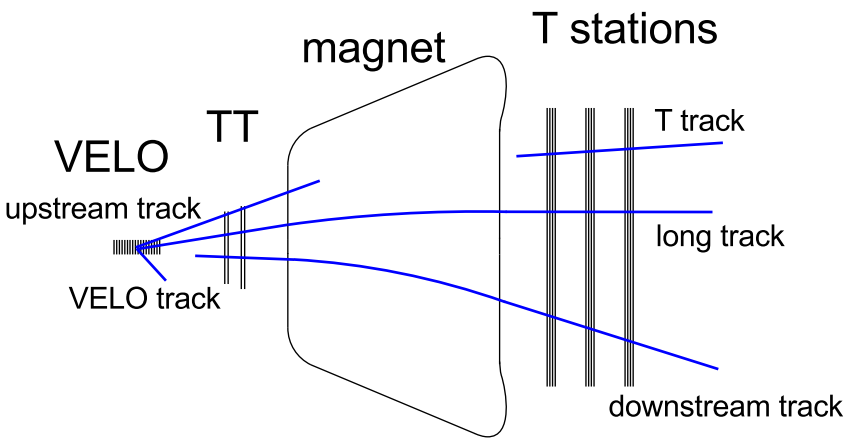

Figure 3.3: Track definitions used by LHCb. More details are given in the text. Figure taken from Ref. [54].

Vertex Locator The vertexing system is referred to as the Vertex Locator, or VELO in short. It is designed to have a high track reconstruction efficiency of around 98%, requiring a high hit reconstruction efficiency. To suppress prompt backgrounds and accurately measure decay times, it is a lightweight detector with a hit resolution of 4–20 µm, depending on angle and strip pitch. Its main task is to reconstruct primary vertices and decay vertices, specifically from b-hadrons, to separate signal from background and to make time-dependent measurements, such as the effective lifetime of $B_s^0 \to \mu^+\mu^-$, described in this thesis in Section 5.2.

The VELO uses silicon microstrip sensors in a semi-circular shape. Silicon microstrips are radiation hard, give lower occupancies in the detector compared to non-silicon detectors, and their readout can be placed outside the LHCb acceptance, which reduces the material budget compared to a pixel detector. The semi-circular shape, referred to as the $R-\Phi$ geometry, is motivated by the uniform occupancy per sensor, optimal resolution at small r, and ease of reconstruction.

For the optimal measurement of the impact parameter of tracks with respect to primary vertices, the innermost radius of the VELO is made as small as possible at 7 mm. The LHC vacuum has to be protected from outgasing of the VELO modules the VELO needs protection from their RF, and the solution with the minimal amount of material is to place the detector in a secondary vacuum shielded by the so-called RF-box, which is placed in between the beams and the VELO sensors, at 5 mm from the LHC beams during collisions.

Additionally, to protect the detector during injection of the LHC beams, the detector is constructed in two halves, which are retracted by 3 cm before the LHC

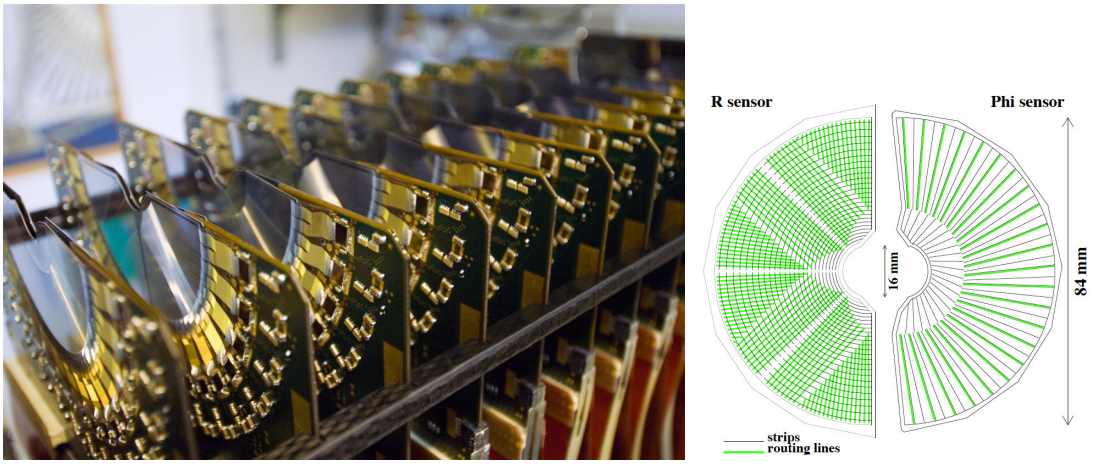

Figure 3.4: Left: A picture of half of the VELO detector, showing the sensors and readout hybrids. Figure taken from Ref. [55]. Right: The two types of VELO sensors, R and Φ sensors, are shown on the left and right respectively. The strips show the coverage of the sensors, and the routing lines are used to lead the signal from the strips to the read-out at the outside of the sensor. Figure taken from Ref. [56].

achieves stable beams. During stable beams, they are moved to a position at 7 mm from the beam in about 3.5 minutes. The VELO sensors and readout hybrids are shown in Figure 3.4.

Two types of VELO sensors are used: R and Φ sensors, which measured the r and ϕ coordinate respectively. The two types of sensors are shown in Figure 3.4. The z positions of the modules have been optimised to cover the angular acceptance of the LHCb detector with at least 3 VELO module hits per track, while minimising the amount of material.

The performance of the VELO in Run 1 is described in [55], and it achieved a similar performance in Run 2. Depending on the radial position and track angle, the VELO has a hit resolution of 5–25 µm with a tracking efficiency better than 98%, illustrated in Figure 3.5.It sustains a high radiation dose from operating close to the LHC pp collisions. The resolution of a primary vertex with 25 tracks is 13 µm in the transverse plane and 71 µm in the z-direction. The impact parameter resolution for a charged track is found to be $11.6 + 23.4/p_T$ μm with p_{T} in GeV/c, as shown in Figure 3.5. The performance of the VELO allowed to separate b hadron decays from the primary vertex and from backgrounds, which is essential for the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The decay time resolution measured with two-prong b-hadron decays is 50 fs, sufficient for the measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime.

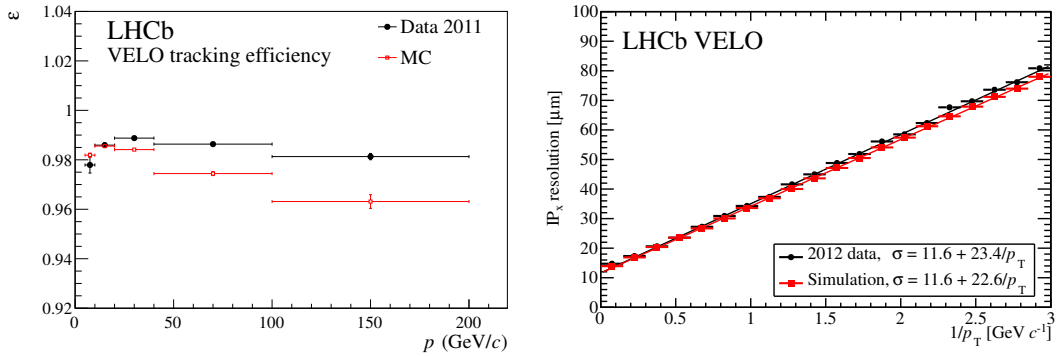

Figure 3.5: Left: VELO tracking efficiency in 2011 data and simulation(MC) as function of momentum. Right: VELO $\mathrm{IP_x}$ resolution in 2012 data and simulation as function of $1/p_\mathrm{T}$. Figures taken from Ref. [55].

Tracking detectors and magnet Downstream of the VELO, there are three more tracking subdetectors in LHCb to perform the tracking across the magnetic field. Together with the VELO, they form the LHCb tracking system. In addition to a high reconstruction efficiency, the tracking system served to determine track momenta with a precision of about 0.5%, achieving a mass resolution for $B^0_{(s)} \to \mu^+\mu^-$ decays of about 22 MeV/c^2, which is sufficient to separate B^0 and B^0_s candidates. This requirement led to a design criterion of a hit resolution of 50 µm in the TT and IT and 200 µm in the OT [57].

The TT is located in one station before the magnet, while the IT and OT are placed in three adjacent tracking stations after the magnet, referred to as T stations. Each tracking station contains four layers of detector material each. Of these four layers, the second and third are positioned at an angle of 5° and −5° respectively, in order to be sensitive to the y-coordinate of tracks passing through. This angle was optimised for the pattern recognition performance in the tracking algorithm [58].

The first tracking detector is the Tracker Turicensis (TT), which uses silicon microstrip sensors and is located just upstream of the magnet. It is pictured in Figure 3.6. It consists of a single detector box with four detection layers arranged in two pairs, with an active area of about 2 m^2 per layer. The main purpose of the TT is to detect the charged decay products of long-lived neutral particles, such as K^0_S and Λ, most of which decay outside of the acceptance of the VELO. It also improves the momentum resolution of long tracks, reduces ghost rates and speeds up the forward tracking in the HLT trigger. It operated with a hit resolution of 53.4 µm in Run 1 and an occupancy varying from 0.2% to 1.9%, increasing closer to the beamline [59], such that the TT provided noise-free extra hits which were sufficiently precise to improve the momentum resolution. Its importance in detecting long-lived neutral

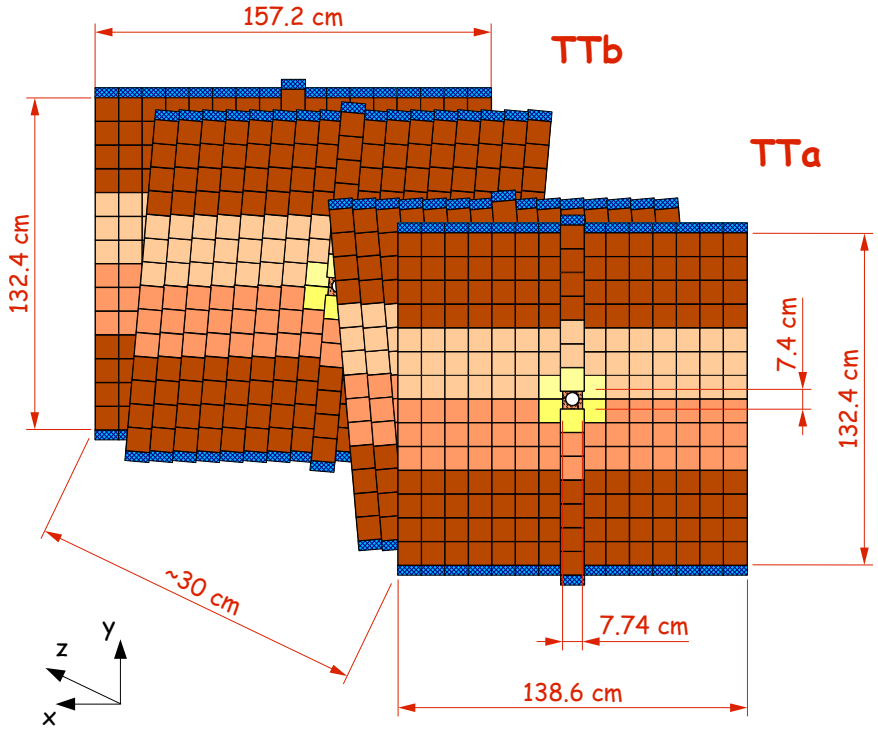

Figure 3.6: The layout of the TT, showing the two pairs of detection layers. Two of the layers are rotated by 5° and −5° with respect to the other two to be sensitive in the y-coordinate.

particles is clear from the fact that about two thirds of the reconstructed K^0_S decays in Run 1 or Run 2 are reconstructed using the TT.

Directly downstream from the magnet, the Inner Tracker (IT) and Outer Tracker (OT) are placed in three adjacent Tracking (T) stations. The OT layout is illustrated in Figure 3.7, including the cross-shaped gap in which the IT is located.

The IT covers a cross-shaped area close to the beam pipe of 120 cm wide and 40 cm high, where the number of particles per surface area is the highest, while the OT covers the rest of the acceptance. To operate in a high density tracking environment, the IT uses silicon microstrip sensors, similar to those used by the TT. In Run 1, the IT achieved a hit resolution of 54.9 µm, with an occupancy varying from 0.2% to 1.9%, depending on the distance from the beam-pipe, both very similar to the performance of the TT [59] and sufficient for the momentum resolution required.

The OT is a drift detector using 5 mm diameter straw tubes, which can be used to cover a large surface area while keeping an acceptable occupancy. The drift gas is a gas mixture of 70% Ar, 28.5% CO_2, and 1.5% O_2. By maintaining a potential

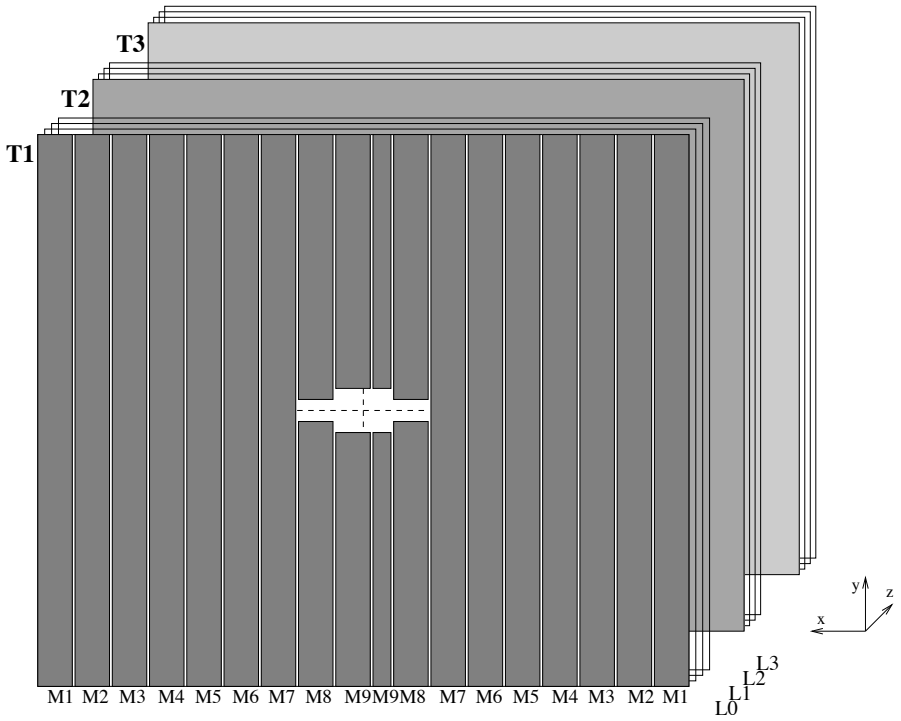

Figure 3.7: The layout of the OT, illustrating the three tracking stations, four layers per station and 18 modules per detector. Note that each second and third layer are rotated by 5° and −5°. The cross-shaped gap in the middle of the OT is where the IT is located. Figure taken from Ref. [60].

difference between the wire and walls of the tube, electrons produced by ionisation drift in the opposite direction of the ions. The current that this produces is measured as signal of a traversing charged particle, taking into account the delay times from time of flight and signal propagation. The drift-time varies depending on where the ionisation process occurred in the wire, up to about 35 ns, such that signals can spill over into the readout window of previous and next bunch crossings. Therefore, the OT is read out in time windows of 75 ns, corresponding to 1 to 3 bunch crossings, depending on the LHC bunch spacing. The long read-out time causes the OT to have a relatively high occupancy, from 5% for the least central to 25% for the most central modules. The OT covers an area of about $30\,\mathrm{m}^2$ with 14 long modules per layer and 8 shorter modules in the middle of the detector, as the gap left by those modules is covered by the IT. It performed up to expectations with a single hit efficiency of 99.2% and a hit resolution of 180 µm, slightly better than the design requirement of 200 µm [57].

Finally, the required magnetic field at LHCb is delivered by a warm dipole magnet, shown in Figure 3.8, designed to deliver an integrated magnetic field of 4 T m. The

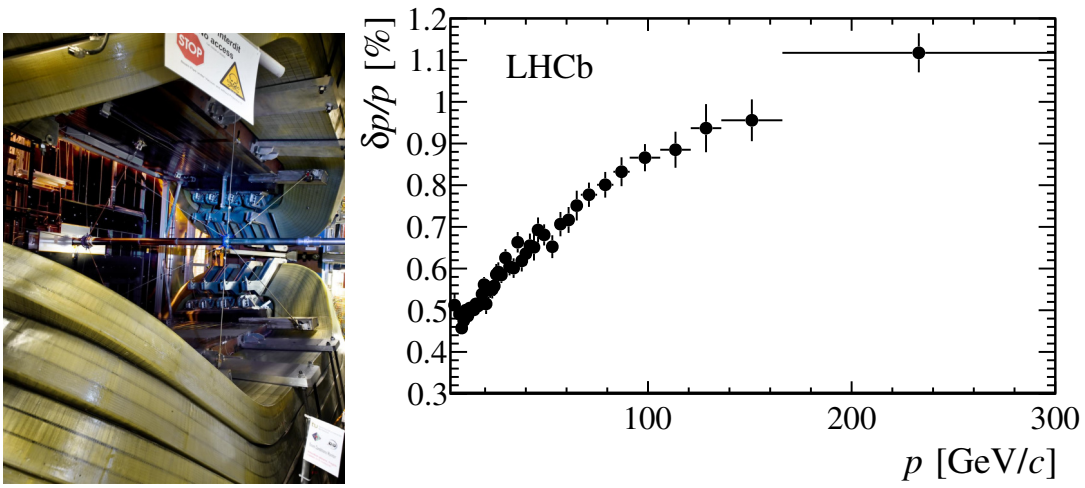

Figure 3.8: Left: A picture of the LHCb magnet, of which the coils are seen around the edges of the picture. The beam pipe and its support structure are shown in the middle, and on the left-hand side the T-stations are visible, with the IT close to the beam pipe in silver and the OT modules behind the IT in black. Right: Momentum resolution as a function of momentum, calibrated on $J/\psi \to \mu^+\mu^-$ decays in Run 1 data. Figure taken from Ref. [59].

advantages of a warm magnet are that it allows for a rapid ramping up of the field concurrently with the ramping-up of LHC magnets, and that its polarity can be inversed. The polarity determines which direction charged particles are bent, so running with both polarities enables to greatly reduce detection asymmetries between particles and anti-particles by cross-checking or canceling their effect between samples with opposite magnet polarity.

With the LHCb tracking system, tracks that reach the T stations are reconstructed with an efficiency of at least 96%, which is sufficient for measurements with 5 or 6-track final states [59]. The efficiency to reconstruct both muon tracks for $B_s^0 \to \mu^+\mu^-$ decays in simulation is around 95%, which is sufficient to not be a limiting factor. The momentum is measured with a precision from 0.5% at low momenta to 1.0% at 200 GeV/c, which resulted in a mass resolution on B^0 and B_s^0 mesons of around 22 MeV/c^2, essential to separate them from each other, such as with $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. The mass shape of $B_{(s)}^0 \to \mu^+\mu^-$ decays will be discussed further in Section 4.4.

3.2.2 Ring Imaging Cherenkov detectors

Two Rich Imaging Cherenkov detectors, referred to as RICH1 and RICH2, are included in the LHCb detector to provide PID for charged particles. RICH1 is

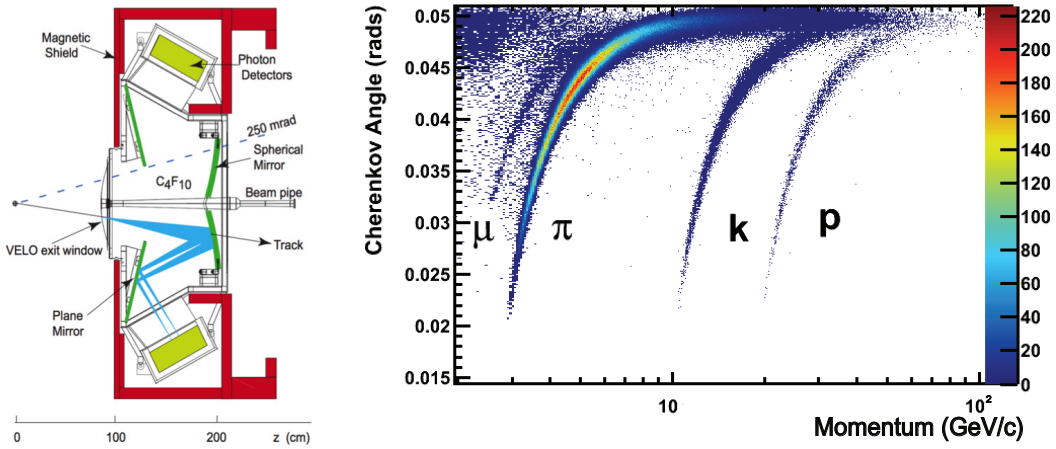

Figure 3.9: On the left, the layout of the RICH1 detector in Run 2. Figure taken from Ref. [61]. On the right, the Cherenkov angle measured by the RICH1 detector for different particle types. Figure taken from Ref [62].

located between the VELO and the TT to provide PID information for upstream tracks, while RICH2 is located after the T stations and before the calorimeter system, giving PID information for downstream and long tracks. Both provide PID for charged particles, especially for pions, protons and kaons, which otherwise are difficult to differentiate in physics analyses. A schematic drawing of the operating principle of RICH1 and the Cherenkov angle measured by RICH1 for different particle species are shown in Figure 3.9.

The RICH detectors identified charged particles by using Cherenkov radiation, which is emitted when they travel faster than the phase velocity of light in a medium. As the angle at which Cherenkov radiation is emitted depends on the velocity of the particle, and the momentum is measured by the tracking system, the RICH detectors provide a measurement of the particle mass and therefore identity. The medium used for each RICH detector was chosen such that the two detectors cover the momentum range of particles in LHCb from 2 to 100 GeV/c: RICH1 covers momenta from 2 to 40 GeV/c; RICH2 covers momenta from 15 to 100 GeV/c.

The RICH1 has an angular resolution of 1.618 mrad per photon and about 20 photons detected per track, while the RICH2 operates with an angular resolution of 0.68 mrad per photon and about 16 photons detected per track [62].In the range of momentum where it is sensitive, the RICH system by itself achieves a kaon identification efficiency of 85% with a pion misidentification rate of 3%, which enables physics measurements with minimal contaminations from misidentified backgrounds [62].

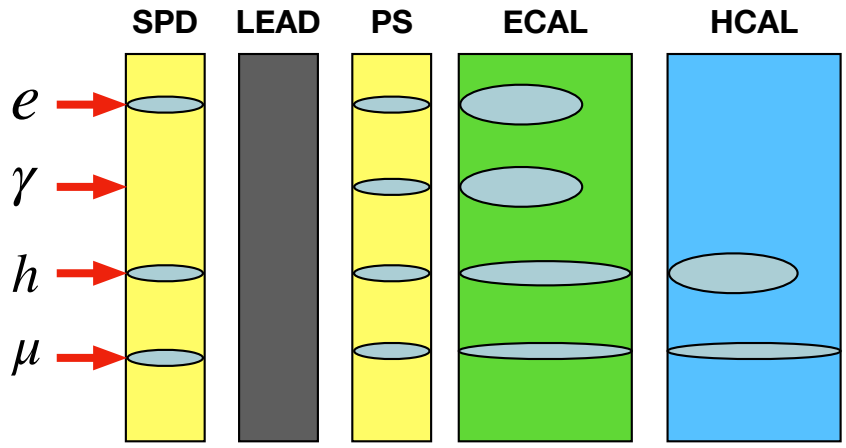

Figure 3.10: Illustration of the response of the LHCb calorimeter sytem to electrons, photons, hadrons and muons, enabling their separation. The deposited energies in the PS, ECAL and HCAL are used to determine the transverse energy of the different particles.

In the $B^0_{(s)} \to \mu^+\mu^-$ analysis, the RICH detectors reduce the misidentification of hadrons from $B^0_{(s)} \to h^+h'^-$ decays and separate different final states which is needed for the normalisation and calibration of two-body B decays. The PID performance is discussed in detail in Section 3.6.

3.2.3 Calorimeters

The calorimetry system at LHCb is located after the RICH2 detector. The main calorimeters at LHCb are sampling calorimeters. They contain a dense material such as lead (for the ECAL) or iron (for the HCAL) as an 'absorber', which starts the showering process, and scintillators as 'active' material, which are used to measure the energy deposited in the shower.

The goals of the calorimetry system at LHCb are to measure the energy of and trigger on hadrons, photons and electrons, and to separate these particles from each other.

The Scintillating Pad Detector (SPD) is a high granularity scintillator pad used to separate photons from electrons, as charged particles interact with scintillators, while neutral particles do not. The SPD only stores whether a signal is over threshold, as very little energy is deposited in the SPD.

The Pre-Shower (PS) detector is a scintillator pad, like the SPD. The main goal of the PS is to distinguish between electrons and charged pions. With that goal in mind, 15 mm of lead converter, equal to $2.5X_0$ thick, is placed inbetween the SPD and the PS. This causes showers that deposit significant amounts of energy for electrons and not for hadrons, thus distinguishing them from each other. The PS also records the energy deposited for each cluster to precisely estimate the energy contained in the electromagnetic shower.

The main goal of the Electronic Calorimeter (ECAL) is to measure the energy of electromagnetically interacting particles; electrons, photons and neutral pions. The amount of radiation lengths contained in the ECAL is $25X_0$, as required to fully contain the shower of highly energetic photons. The ECAL is read out using wavelength shifting fibers that lead the signal from the scintillators to photomultiplier tubes, located downstream in the ECAL and calibrated to properly convert a signal to an energy.

To take into account the large variation of particle density over the calorimeter surface, the ECAL is segmented into three different sections, each with the same sized modules but with a higher density of readout cells for sections closer to the beam. The corresponding geometry is also used in the SPD and PS detectors, scaled such that their transverse dimensions scaled relative to the distance from the pp collision point. In this configuration, the ECAL achieves an energy resolution on electromagnetic showers of

$$\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%, \tag{3.2}$$

where E is defined in GeV [63].

The Hadronic Calorimeter (HCAL) aims to trigger on the transverse energy deposited by hadrons and absorb their showers to limit the background from hadrons in the muon stations. The HCAL is a sampling calorimeter just like the ECAL, but with a different orientation where the layers are placed parallel to the beam axis, in the $yz-$plane. The layers contain alternating slabs of iron and scintillator pads, 3 of each, which are read out in the same way as the ECAL.

In total, the amount of radiation lengths contained in the HCAL is $5.6\lambda_{\text{int}}$, limited by space constraints in the LHCb cavern but enough to absorb most hadronic showers. The HCAL is segmented in two regions to operate well in the higher particle density environment close to the beamline. It achieves an energy resolution on hadronic

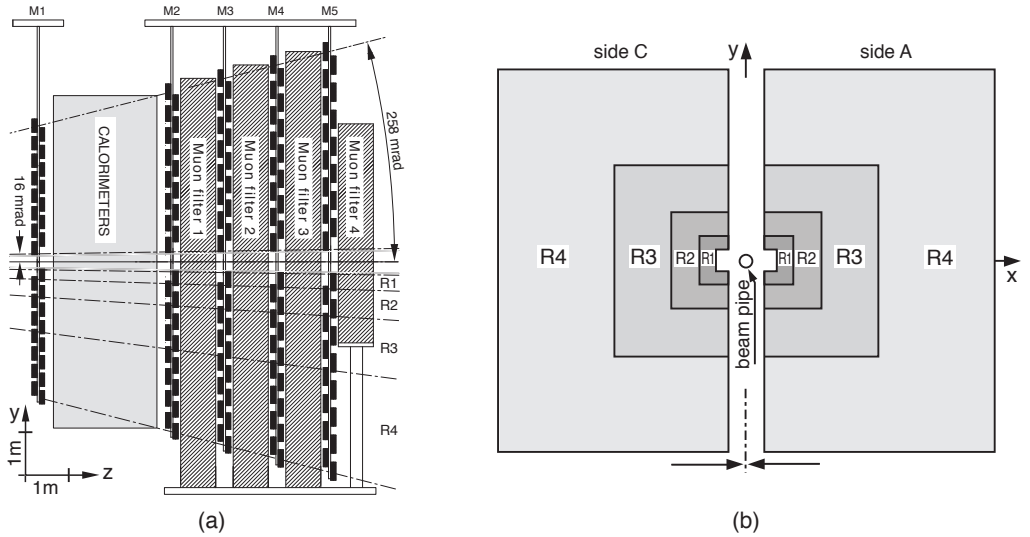

Figure 3.11: The muon stations, as seen from the side and from the front, on the left and right respectively. The layout of the stations is detailed in the text. Figure taken from Ref. [64].

showers of

$$\sigma_E/E = 80\%/\sqrt{E} \oplus 10\%, \tag{3.3}$$

where E is defined in GeV [63].

3.2.4 Muon stations

As muons are minimum ionising particles, they are the only particles that pass through the calorimeter system, if they have an energy of at least 3 GeV. Muons are identified by reconstructing their tracks in the muon stations, which strongly reduces background rates in the trigger and offline. The requirements for the muon system are that it should identify muon candidates with efficiencies around 95% and provide this information in time to be used in the trigger. The muon system uses five rectangular stations (M1-M5). The first is placed upstream of the calorimeters, while the other four are placed downstream of the calorimeters, as shown in Figure 3.11. The transverse size of each station is scaled to achieve a uniform acceptance in each station, given the distance to the pp interaction point. To account for the varying particle density in the detector depending on the distance to the beamline, the stations are divided in four regions, with different segmentations to achieve comparable occupancies in each region. These stations are interleaved with 80 cm thick iron absorbers to better separate muons from the few hadrons that pass through

the calorimeter system. The total absorber thickness is about $20\lambda_{\text{int}}$, such that muons must have a momentum of about 6 GeV/c to reach the most downstream station M5. The muon stations are equipped with multi-wire proportional chambers, except for the highest rate region in M1, which uses Gas Electron Multiplier detectors.

The hit efficiency per muon station is required to be above 99% to achieve the total muon detection efficiency required. This requirement is fulfilled in all but the most intense region in the M1 station, where the deadtime of around 1.5% from reading out the system is a limiting factor.

The muon identification algorithm based on muon station hits is called isMuon. It associates hits to reconstructed tracks, given their extrapolated trajectory in the muon system and requiring them to fall within a rectangular search window with an optimised range in x and y. IsMuon also takes into account that the number of hits in the muon stations depends on the momentum of the muon, because of the iron absorber layers in the muon systems. Those requirements are shown in Table 3.2.

The muon identification efficiency of isMuon in trigger and offline reconstruction is around 97%, with a pion to muon misidentification probability of only 1 to 3%, depending on the kinematics of the pion [64], For $B_s^0 \to \mu^+\mu^-$, this is sufficient to strongly suppress backgrounds at every stage until the final offline selection. More details on PID for muons, using the full detector information, are given in Section 3.6.

Table 3.2: The required number of associated hits in each muon station as function of the momentum of the muon candidate to pass isMuon.

Momentum	Consistent hits
$3 < p < 6$ GeV/c	M2 + M3
$6 < p < 10$ GeV/c	M2 + M3 + (M4 or M5)
$10 < p$ GeV/c	M2 + M3 + M4 + M5

3.3 The LHCb Trigger System

The LHCb trigger system uses all of the subdetectors to decide which events to save on disk [65]. It needed to bring the event rate down from the collision rate, around 20 MHz, by about three orders of magnitude, to 5 kHz in Run 1 and 12.5 kHz in Run 2. Events that are not accepted by a trigger are lost, so the trigger must be efficient on signal as well as robust to avoid losses that can not be recovered later. The LHCb trigger system in Run 1 and Run 2 consisted of two levels:

- The Level 0 (L0) trigger is implemented in hardware. It mainly uses input from the muon and calorimeter systems to bring the event rate down to 1 MHz, which is the rate at which the front-end electronics can be read out.
- The High Level Trigger (HLT) is a software application, called Moore. This application ran on a PC farm using information from the full detector. It reduces the rate of events down to 5 kHz (12.5 kHz) in Run 1 (Run 2).

3.3.1 L0 trigger

The L0 trigger is divided into three types of trigger: the calorimeter trigger, muon trigger and pile-up trigger. The last is not used to select physics signatures, but to determine the luminosity, which as will be discussed in Section 4.5 is not needed for the $B^0_{(s)} \to \mu^+\mu^-$ analysis, and thus not discussed any further in this thesis.

The calorimeter trigger searches for particles with a high transverse energy (E_{T}). It forms clusters based on 2×2 cells in the ECAL or HCAL and selects those with the largest E_{T} in each. Clusters are identified as electrons, photons, or hadrons, based on whether they are detected in the different sections of the calorimeter system, as illustrated in Figure 3.10. One trigger is defined for each particle type, each with its own tunable E_{T} threshold. L0Electron uses the cluster with the highest E_{T} in the ECAL. It requires 1 or 2 PS cells hit in front of the cluster and at least one hit in the SPD cells associated with those PS hits. In Run 1, the threshold was set varying from 2.5 to 3.0 GeV, and in Run 2 from 1.8 to 2.7 GeV. L0Photon is almost the same as L0Electron, but requires not having any hit in the associated SPD cells. Its threshold is set at the same level as L0Electron in Run 1 and 2015. In 2016 its threshold was tightened compared to L0Electron and kept around 2.8 GeV. L0Hadron uses the HCAL cluster with the largest E_{T} and adds the E_{T} measured in the associated ECAL cluster. Its threshold was set from 3.5 to 3.7 GeV in Run 1 and from 2.5 to 3.9 GeV in Run 2.

The muon trigger selects muons using information from the muon stations. Each quadrant of the muon detector is connected to a L0 muon processor, which searches for the two muon tracks with the largest and second largest p_{T} in each quadrant. They base their search on hits defining a straight line pointing towards the pp collision vertex in the $y - z$ plane, which is the non-bending plane of the magnet. The position of a track in the first two muon stations allows for a p_{T} esimate with an uncertainty of $\sim 25\%$, which is sufficient for the hardware trigger. Two triggers are implemented; L0Muon requires the largest p_{T} out of the p_{T} of the eight candidates

to pass a threshold, from 1.5 to 1.8 GeV/c in Run 1 and 1.4 to 2.9 GeV/c in Run 2. L0DiMuon requires the largest p_T times second largest p_T in any of the four quadrants to pass a threshold, from 1.7 to 2.6 GeV^2/c^2 in Run 1, while in Run 2 the threshold was set at 1.3 to 1.5 GeV^2/c^2.

In summary, LHCb used five different hardware triggers for B physics: L0Electron, L0Photon, L0Hadron, L0Muon and L0DiMuon. Together they reduced the event rate to 1 MHz, and passed events on to the HLT.

3.3.2 HLT

The software trigger used in Run 1 and Run 2 at LHCb, referred to as the HLT, is divided into two stages:

- HLT1, which took events passing the L0 trigger and ran a partial event reconstruction to bring down the rate to 40 kHz (100 kHz) in Run 1 (Run 2).
- HLT2, which processed events passing HLT1, running a full event reconstruction and reducing the rate to 5 kHz (12.5 kHz) in Run 1 (Run 2).

The HLT uses the concept of trigger lines. A trigger line is a set of reconstruction and selection steps which returns a positive or negative decision. Many lines can be run, but at least one positive decision by any trigger line is needed for an event to pass a trigger stage. There were of the order of 40(20) HLT1 and 130(500) HLT2 trigger lines in Run 1(Run 2). At LHCb, the configuration of the trigger is defined by the L0 thresholds in addition to the total set of HLT trigger lines and their configuration. It is referenced with a Trigger Configuration Key (TCK), a unique 32 bit identifier. For each event, the TCK used and trigger lines passed are recorded.

At LHCb, the HLT strategy changed substantially from Run 1 to Run 2 [66]. The main change is that all events from HLT1 are buffered to 10 PB of storage space in the online computing farm, sufficient to save events from two weeks of running HLT1 before having to run HLT2. The extra time gained from this buffer allowed running the alignment and calibration of all the subdetectors as input to the HLT2 reconstruction. As a result, the reconstruction and selection for high-yield modes could be run fully in the trigger.

HLT1 Due to the limited processing time available per event in HLT1, only a partial event reconstruction is run. The main goal of this reconstruction is to find displaced tracks with high p_T, as they are a signal of b hadron decays, or muons.

In Run 1, the HLT1 reconstruction is optimised for muons or displaced tracks. VELO tracks are reconstructed with a full 3D pattern recognition and primary vertices (PVs) are built from vertices with more than 5 VELO tracks within 300 μm of the nominal collision point per fill. For events triggered by L0Muon or L0DiMuon, a fast muon identification is run, which searches for associated muon hits assuming a momentum of at least 6 GeV/c. Any VELO-Muon track candidates has to pass a χ^2 cut using both the VELO and muon station hits. Independent of the L0 trigger, displaced VELO tracks are selected requiring a minimum χ^2_{IP} to any PV and a minimum number of hits on the track. Selected VELO tracks are extrapolated to the T stations with a search window defined by the maximum deflection of particles with a transverse momentum of 1.2 GeV/c. The track fit is run with a simplified material description and fewer iterations, such that for example the relative invariant mass resolution of dimuon candidates is $\sim 3\%$ worse than for offline reconstructed events. Therefore, cuts on IP significance, momentum, and mass could be applied with reasonable precision.Any remaining muon candidates are required to pass isMuon, which is the offline muon identification algorithm.

In Run 2, the partial event reconstruction started with VELO tracks and PVs reconstructed as in Run 1. However, in Run 2 all VELO tracks are used and extrapolated to the TT, where 3 hits in a small window around the track are required to create upstream tracks. Due to the small effect of the magnetic field in the TT, upstream tracks have a momentum resolution of $\sim 20\%$ and a defined charge, which allowed to loosen the p_{T} requirement for extrapolation to the T stations to 0.5 GeV/c. Good long track candidates are selected with a neural network, and the offline track fit is run, followed by another neural network to reject bad candidates based on the track fit quality. Muons with a p_{T} over 0.5 GeV/c are identified with the offline isMuon algorithm. For muons with lower p_{T}, a fast muon identification algorithm is run, similar to the one used in Run 1, which is described above. In Run 2 it is based on upstream tracks, and only considers hits that are not already associated with other candidates. It identifies muon tracks down to transverse momenta of only 80 MeV/c.

These are the main HLT1 lines used in this thesis:

- HLT1TrackAllL0, a one-track line for B physics analyses in Run 1. It selects displaced tracks (IP>0.1 mm and χ^2_{IP} >16) with a minimum p_T of 1.6 GeV/c, minimum p of 10(3) GeV/c in 2011(2012) and maximum track χ^2/ndf smaller than 2.

- HLT1TrackMuon, which selects muons similarly to Hlt1TrackAllL0, but with looser p, p_T thresholds and track χ^2/ndf requirements in Run 1. For Run 2, it requires a $p_{\mathrm{T}} > 1$ GeV/c, $p > 6$ GeV/c and $\chi^2_{\mathrm{IP}} > 10(35)$ in 2015(2016).

- HLT1TrackMVA, the improved version of HLT1TrackAllL0 for Run 2. In addition to requiring a track χ^2/ndf smaller than 2.5, it applies a 2D selection in IP significance and p_{T} (defined in GeV/c) using a hyperbolic function:

$$\log \chi^2_{\mathrm{IP}} = \frac{1}{(p_{\mathrm{T}} - 1)^2} + \frac{b}{25}(25 - p_{\mathrm{T}}) + \log(7.4). \tag{3.4}$$

  All tracks with $p_{\mathrm{T}} > 2.5\,\mathrm{GeV}/c$ and χ^2_{IP} >7.4 are accepted. A loose and tight configuration are defined with a different value for b, and the tight configuration is used when the buffer between HLT1 and HLT2 gets too full, as in 2016.

- HLT1TrackMuonMVA, which uses the same selection as HLT1TrackMVA. It was introduced in 2016 and used the loose configuration of HLT1TrackMVA throughout the year.

- HLT1SingleMuonHighPT, which selects muons without requiring any displacement, by requiring large tranverse momenta. In Run 1, a p_{T} of at least $4.8\,\mathrm{GeV}/c$ and a momentum of at least 8(3) GeV/c in 2011(2012) are required, and in Run 2 a p_{T} of $6(4.3)\,\mathrm{GeV}/c$ for 2015(2016) and a momentum of at least 6 GeV/c are needed to pass this line.

- HLT1DiMuonHighMass, a two-track line for dimuon candidates with an invariant mass above $2.7\,\mathrm{GeV}/c^2$, vertex $\chi^2 < 25$ and distance of closest approach (DOCA) for the two tracks of less than 0.2 mm. Both muons are required to have a minimum p_{T} of 0.5(0.3) GeV/c in Run 1 (Run 2) and a momentum of 6 GeV/c, except for 2012 where this requirement was loosened to $3\,\mathrm{GeV}/c$.

HLT2 In HLT2, a full event reconstruction is run. In Run 1, it is simplified compared to the offline reconstruction. It only uses a forward tracking algorithm with VELO tracks as a seed for long track reconstruction, instead of also matching independently found T track segments to the VELO track. This results in a 1-2% lower reconstruction efficiency per track. Additionally, only tracks with high enough momentum and transverse momentum are reconstructed to limit the number and size of search windows. Finally PID capabilities are limited, as the RICH reconstruction took too long to run in HLT2. Therefore, only muons and electrons could be well identified. Muons are identified with isMuon, while electrons are identified with a special algorithm that associated tracks with ECAL clusters.

In Run 2, the same algorithms are run offline and offline, with the exception of those to reconstruct neutral particles. Therefore, for this thesis, the offline and HLT2 reconstruction in Run 2 can be treated as being the same.

There are two types of trigger lines for B physics at the HLT2 level; topological and exclusive lines. The topological lines are designed to select b-hadrons with at least two charged tracks forming a good vertex based on their track fit quality, significant IP separation to the primary vertex and muon or electron identification. Additional tracks which pass similar requirements are included to make a 2 to 4-body candidate. The mass of the b-hadron candidate is reconstructed taking into account the possible omission of one or more daughters by using the method of the corrected mass. This method corrects for missing transverse momentum comparing the b hadron candidate and transverse momentum of all tracks to correct the b hadron mass estimate. In Run 1, a cut-based and an MVA-based topological line are implemented. In Run 2, only an MVA-based line is used, as it performs better, and is retrained for the conditions of Run 2. In this thesis, the MVA-based lines contribute significantly in the HLT2 selection for $B^0_{(s)} \to h^+h'^-$ and $B \to J/\psi X$ decays.

Exclusive lines are based on selecting a specific final state and define requirements specific to that final state. The $B^0_{(s)} \to \mu^+\mu^-$ analysis mostly relies on exclusive trigger lines. The main exclusive trigger lines in HLT2 relevant for this thesis are:

- HLT2DiMuonB, an exclusive line for $B^0_{(s)} \to \mu^+\mu^-$ and $\Upsilon(NS) \to \mu^+\mu^-$ decays. It requires a dimuon candidate with a vertex χ^2 smaller than 10 and a mass of at least 4.7 GeV/c^2. In Run 2, the vertex χ^2 cut was loosened to 25.

- HLT2DiMuonDetachedJpsi, an exclusive line for $B \to J/\psi X$ decays, which requires a dimuon candidate with a mass within 120 MeV/c^2 of the J/ψ mass, a χ^2_{vtx} smaller than 25 and a secondary vertex separated by 3σ from the primary vertex.

- HLT2DiMuonJPsiHighPT and HLT2DiMuonPsi2SHighPT, exclusive lines for $J/\psi \to \mu^+\mu^-$ and $\psi(2S) \to \mu^+\mu^-$ decays respectively. For Run 2, they require a dimuon candidate with χ^2_{vtx} smaller than 25, a p_{T} above 2(3.5) GeV/c for the J/ψ ($\psi(2S)$) line, and its mass in a window of 100 MeV/c^2 around the J/ψ or $\psi(2S)$ mass. The Psi2S line also requires a p_{T} above 2 GeV/c for both muon tracks. In 2015, the mass window range was increased to 120 MeV/c^2 and the muon p_{T} requirement for the $\psi(2S)$ line dropped. In 2016, the p_{T} cut on the dimuon candidate was also loosened for the $\psi(2S)$ line to be equal to the requirement for the J/ψ line.

- HLT2B2HH, an exclusive line for $B^0_{(s)} \to h^+h'^-$ decays. For Run 1, it requires a b-hadron candidate with a mass from 4.7 to 5.9 GeV/c^2 under the dipion hypothesis, p_{T} of at least 1.2 GeV/c, a maximum IP wrt the PV of 0.12 mm and tracks with a p_{T} >1 GeV/c and IP> 0.12 mm. In Run 2, the b-hadron candidate p_{T} requirement is tightened to be 4.5 GeV/c, and the candidate should have a secondary vertex distance significance of at least 10 and χ^2_{vtx} of at most 3. The IP distance cuts are replaced by χ^2_{IP} cuts, which respectively has to be at most 9 for signal and at least 16 for both tracks.

3.3.3 Global Event Cuts

During the running of LHCb, Global Event Cuts (GEC) are used to reject very high multiplicity events that take too much processing time. At the hardware trigger level, the number of hits in the SPD detector is required to be less than 600 in Run 1 and less than 450 in Run 2, except for the Dimuon line, which allows events having up to 900 SPD hits. In the HLT, events are rejected for having too many clusters in the VELO, OT or IT, each corresponding to an occupancy of about 20% in the respective detectors.

3.4 Offline reconstruction and selection

After an event passes the trigger, its raw data and trigger information are written out to storage. The Brunel application reconstructs objects such as tracks and clusters from the raw data. In Run 1, only the offline reconstruction uses the fully calibrated and aligned detector. For charged particle reconstruction, the resolution for tracking related parameters such as the mass, vertex quality, and displacement improve offline by about 1% relatively, compared to HLT2. The main differences between HLT2 and offline reconstruction in Run 1 are particle identification and fake track rejection, which are only fully exploited offline. In Run 2, the HLT2 and offline reconstruction are the same for charged particles.

The trigger information is saved to calibrate the trigger efficiency and to know the selection criteria applied to offline candidates. If an offline candidate is accepted by a certain trigger selection, it is defined as Trigger On Signal (TOS) for that selection. If the event still passes the criteria when the candidate hits are removed from the event, the candidate is defined as Trigger Independent of Signal (TIS). It is possible that an event is TIS and TOS at the same time.

Subsequently, a centrally organised selection procedure is applied to the data, to reduce the samples to a size that could be run over by analysts. This central selection is called the stripping, and divides the data into multiple streams, which share a common type of physics final state. For example, the main stripping lines for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are part of the dimuon stream.

Finally, analysts run the DaVinci application over data from a stream to obtain a set of candidates, and for those candidates extract or compute useful variables for their analysis, saving them in Root files. These jobs are run in parallel on the grid with the Ganga application. For the $B^0_s \to \mu^+\mu^-$ analysis, a final and optimised selection is applied on these root files, and referred to as offline selection. In this thesis, the stripping and offline selection are treated as one selection step in Section 4.2.2.

3.5 Simulation

To evaluate some of the efficiencies in the $B^0_s \to \mu^+\mu^-$ analysis and to obtain the kinematic distributions of signal, calibration and background channels, simulated samples are generated. The LHCb simulation framework is called Gauss. Pythia generates pp collisions until the requested b-hadron candidate is present. The b-hadron candidate is subsequently forced to decay with the EvtGen package according to a given model, for example according to two-body phase space for the $B^0_s \to \mu^+\mu^-$ decay. To ensure the signal candidate is reconstructible, the final state particles are required to be close to the LHCb detector acceptance, namely in a pseudorapidity range of 1.6 to 5.3. The generated particles are propagated through the detector with Geant4, which returns the truth information of the event as well as the hits left by the particles passing through the detector.

The digitisation application Boole converts detector hits into a format identical to real data; subsequently, simulated data is treated the same as real data, and the same trigger, reconstruction and stripping algorithms are run over each.

3.6 Particle Identification

For each particle hypothesis, the RICH detectors, calorimeters and muon stations return a likelihood. To obtain an improved PID performance, the likelihoods from these subdetectors are combined.

The first type of PID variables at LHCb are the Delta Log Likelihood (DLL) variables, which are a sum of the logarithms of the subdetector likelihoods, relative to the pion hypothesis. The second type is called ProbNN, as it is a Neural Network that returns a pseudo-probability from 0 to 1 for each hypothesis. ProbNN takes into account correlations between the subdetector likelihoods, and adds information from the tracking system to reject ghosts and decays in flight.

For the $B^0_{(s)} \to \mu^+\mu^-$ analysis, muon identification is essential for the signal and some of the calibration modes. Hadron identification, specifically separation of pions, kaons and protons, is needed for the $B^0_{(s)} \to h^+h'^-$ calibration sample.

3.6.1 Muon identification

The main backgrounds for muons are decays of pions or kaons to muons while propagating through the detector, random hits in the muon stations combined with tracks of hadrons before the calorimeter, and hadrons that punch through the calorimeter and reach the muon systems. The contribution from decays in flight dominates the misidentification rates.

There are three kinds of PID algorithms for muons at LHCb: isMuon, DLL and ProbNN. The first, isMuon, is a fast algorithm using only information from the muon stations, available offline and in the trigger, and is described in Section 3.2.4. It is used for $B^0_s \to \mu^+\mu^-$ to select muons in the trigger and stripping. To achieve the best separation of muons from hadrons, The DLL and ProbNN variables are considered in the offline selection, and the best performing variable combiation is used. The performance of DLL and ProbNN for muon identification in $\Sigma \to p\mu^+\mu^-$ is shown in Figure 3.12. Clearly, ProbNN rejects misidentified backgrounds more for the same identification efficiency, and is therefore used in the $B^0_s \to \mu^+\mu^-$ analysis. The PID selection for $B^0_s \to \mu^+\mu^-$ is described in Section 4.2.3.

3.6.2 Separation of charged hadrons

The separation of different charged hadrons at LHCb is based predominantly on the RICH systems, as described in Section 3.2.2. For the $B^0_{(s)} \to \mu^+\mu^-$ analysis, the DLL variables were used to separate kaons, pions and protons from each other. The performance of the DLL variables to separate kaons from pions, calibrated on $D^{*+} \to (D^0 \to K^+\pi^-)\pi^+$ decays, is illustrated in Figure 3.12. Clearly, $\mathrm{DLL}_{K\pi} > 5$ is sufficient to reject almost all pions while retaining most kaons.

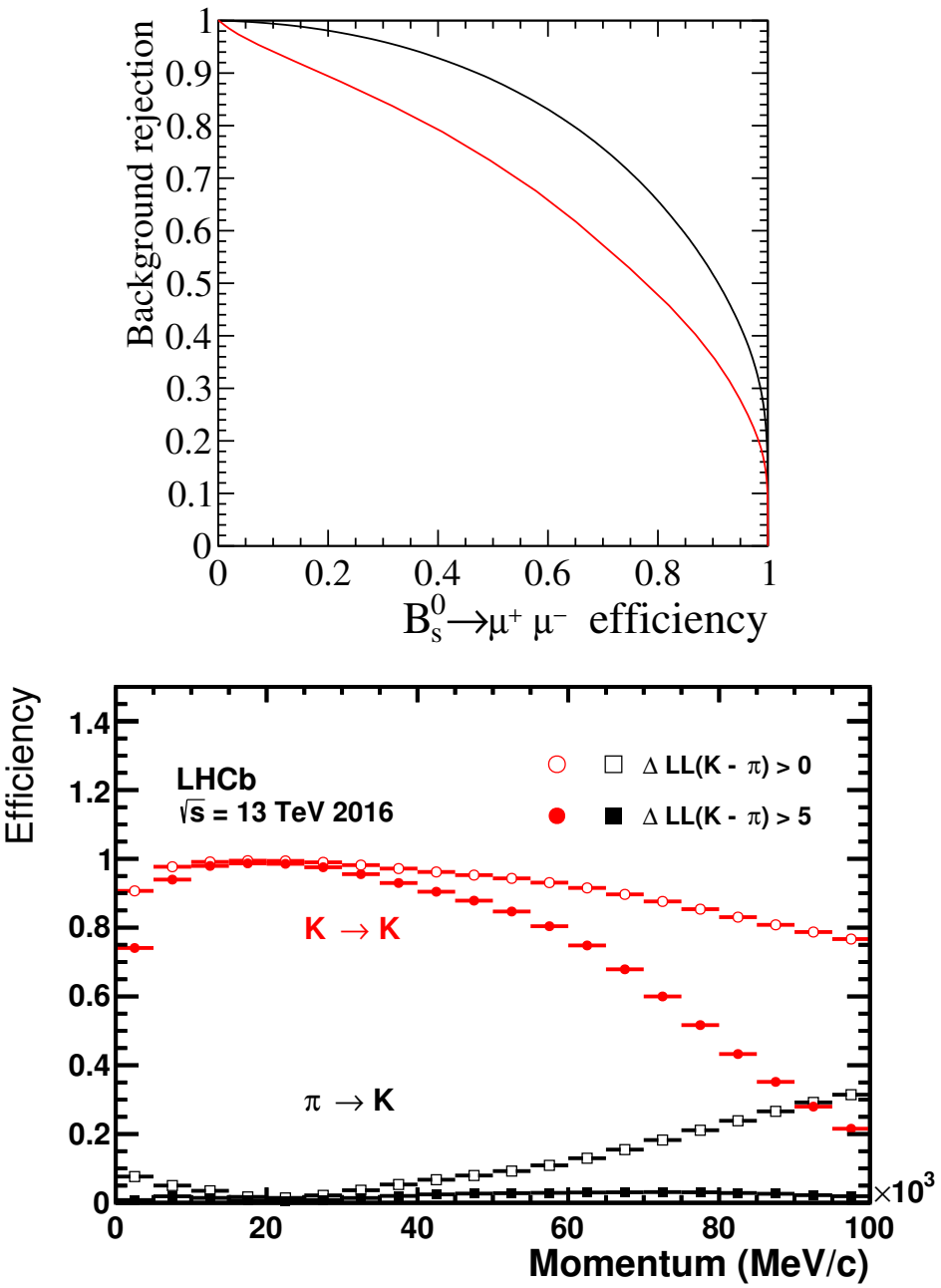

Figure 3.12: Top: the efficiency to select muons versus rejecting combinatorial background, using the DLL variable (in red) or ProbNN variable (in black) for muons on $B_s^0 \to \mu^+\mu^-$ simulated samples for signal and invariant mass sidebands in data. Bottom: the kaon identification and pion as kaon misidentification rate in 2016 data as function of momentum using the decay $D^{*+} \to D^0(\to K^+\pi^-)\pi^+$.

Chapter 4

$B^0_{(s)} \to \mu^+\mu^-$ analysis

4.1 Introduction

This chapter describes the main subject of this thesis, which is the search for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays and the measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction with the LHCb detector [67].

The measurement was made with data recorded from 2011 to 2016, corresponding to a total integrated luminosity of $4.4\,\text{fb}^{-1}$. The integrated luminosities and centre-of-mass energies per year of data-taking are listed in Table 3.1. The data taken in 2011 and 2012 (resp. 2015 and 2016) are referred to as the Run 1 (resp. Run 2) dataset. The data samples used for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are summarised in Table 4.1.

In this dataset, around 220 $B^0_s \to \mu^+\mu^-$ and 25 $B^0 \to \mu^+\mu^-$ decays were expected to occur within the acceptance of the LHCb detector. Luckily, the experimental signature of $B^0_{(s)} \to \mu^+\mu^-$ decays is very clean, namely two oppositely charged muons that form a well-reconstructed and clearly displaced vertex, with a dimuon invariant mass close to the B^0_s or B^0 mass of 5367 and 5280 MeV/c^2, respectively.

The $B^0_{(s)} \to \mu^+\mu^-$ analysis strategy is shown in Figure 4.1. First, a selection is made to obtain a clean sample containing most of the $B^0_{(s)} \to \mu^+\mu^-$ decay candidates, while suppressing backgrounds by many orders of magnitude. The most discriminating variables, in this analysis the dimuon invariant mass and the BDT, are used to separate signal and background in a maximum likelihood fit. To perform the fit, one needs to calibrate the efficiency of the selection and the fit shapes for signal, including the absolute normalisation, and to determine the expected background yields and fit shapes. Finally, the maximum likelihood fit is performed to search for $B^0_{(s)} \to \mu^+\mu^-$ decays and measure their branching fractions if they are found. A

detailed description of these steps is given in this chapter; an overview of the chapter is presented below, with references to the sections describing each specific subject.

The selection used to keep $B^0_{(s)} \to \mu^+\mu^-$ candidates and suppress backgrounds is described in Section 4.2. The largest background for $B^0_{(s)} \to \mu^+\mu^-$ is the "combinatorial" background, denoted by $b\bar{b} \to \mu^+\mu^- X$, which consists of two muons coming from two different b hadron decays. It is absolutely essential to suppress combinatorial, as about 3 billion $b\bar{b} \to \mu^+\mu^- X$ events are expected in the analysis dataset. Specific decays of b hadrons, referred to as exclusive backgrounds, also contribute. Both are suppressed by the selection to optimise the analysis sensitivity.

After the selection, a clean sample remains which is used to search for the $B^0_{(s)} \to \mu^+\mu^-$ signature. A multivariate classifier, called BDT, is used to divide the clean sample in more or less signal-like subsamples. It is trained to separate $B^0_{(s)} \to \mu^+\mu^-$ decays from combinatorial background using 7 variables, associated with the candidate topology and track isolation. The track isolation variables give the BDT most of its discriminating power, as $b\bar{b} \to \mu^+\mu^- X$ often have extra tracks close to the muon tracks, while real $B^0_{(s)} \to \mu^+\mu^-$ decays do not.

The branching fractions of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are determined from the $B^0_{(s)} \to \mu^+\mu^-$ sample with a two dimensional maximum likelihood fit to the dimuon invariant mass and the BDT, which are the two most discriminating variables. For this fit, the shapes of signal and background, the normalisation of the signal branching fraction and the expected background yields have to be determined.

The signal BDT shape is calibrated with fits to $B^0 \to K^+\pi^-$ data per BDT bin, as described in Section 4.3. These decays are topologically and kinematically similar to $B^0_{(s)} \to \mu^+\mu^-$ decays, and thus are a good candidate for a calibration channel. However, their trigger and particle identification requirements are very different than those for $B^0_{(s)} \to \mu^+\mu^-$, and the BDT shape from $B^0 \to K^+\pi^-$ data has to be corrected for these differences. To reduce uncertainties associated with this correction, only events Triggered Independent of Signal (TIS) are used for the BDT calibration.

The mass shape of $B^0_{(s)} \to \mu^+\mu^-$ decays is described by a Crystal Ball function [68], as described in Section 4.4. This shape contains four parameters: the mean, resolution, and two tail parameters. The mean is calibrated using $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ data, as it is not affected by different trigger and PID selections. The resolution is calibrated using a data sample with dimuon resonances at lower and higher masses than the $B^0_{(s)}$ mesons; the dimuon mass resolution at the B^0_s and B^0 masses is

determined through interpolation. Finally, the tail parameters are calibrated using $B^0_{(s)} \to \mu^+\mu^-$ simulated samples, including a smearing correction for the resolution difference between data and simulation.

To convert candidate yields into a branching fraction, the branching fraction is measured relative to other b hadron decays. This procedure is described in Section 4.5. The normalisation for the $B^0_{(s)} \to \mu^+\mu^-$ analysis is performed by measuring the yields of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ decays, which complement each other thanks to their different final states. As mentioned, the $B^0 \to K^+\pi^-$ sample has a different trigger and PID selection than the $B^0_{(s)} \to \mu^+\mu^-$ sample. Similarly to the BDT calibration, TIS candidates are required to reduce uncertainties coming from the efficiency correction. $B^+ \to J/\psi K^+$ decays have similar trigger and PID requirements to $B^0_{(s)} \to \mu^+\mu^-$. The effect of the extra kaon in the final state is easily and precisely corrected for using simulation samples and the tracking efficiency measured in data. Finally, to measure the $B^0_s \to \mu^+\mu^-$ branching fraction, the production fraction of B^0_s versus B^0 or B^+ mesons has to be known. At the time of this analysis, this ratio had been measured with sufficient precision in Run 1 using other B^0_s decays, but it had not yet been measured in Run 2. To determine if this ratio changed, the efficiency-corrected yields of $B^0_s \to J/\psi\,\phi$ and $B^+ \to J/\psi K^+$ decays were measured in Run 1 and Run 2 data.

Section 4.6 describes the procedure to constrain parameters used in the ML fit to describe the exclusive backgrounds. Simulation samples are used (except for $B^0_{(s)} \to h^+h^{(\prime)-}$) to estimate the number of events after selection, as well as the mass and BDT shape. For $B^0_{(s)} \to h^+h^{(\prime)-}$, the yields per BDT bin from $B^0 \to K^+\pi^-$ data are used. For the main backgrounds, namely $B^0_{(s)} \to h^-\mu^+\nu_\mu$ and $B^0_{(s)} \to h^+h^{(\prime)-}$, the estimates are checked using the $B^0_{(s)} \to h^+h^{(\prime)-}$ data sample with a PID selection requiring one hadron and one muon.

Using the expected signal yield, signal shape and background calibration, the maximum likelihood fit to the $B^0_{(s)} \to \mu^+\mu^-$ sample is performed. This fit is discussed in Section 4.7. The results of the $B^0_{(s)} \to \mu^+\mu^-$ analysis and the measurement of the effective lifetime of $B^0_s \to \mu^+\mu^-$ with the same dataset are discussed in Chapter 5.

Sample	Purpose
$B^0_{(s)} \to \mu^+\mu^-$	Search for $B^0_{(s)} \to \mu^+\mu^-$ decays Measurement of their branching fractions
$B^0 \to K^+\pi^-$ TIS	Normalisation, BDT calibration
$B^0 \to K^+\pi^-$	Momentum scale calibration for $B^0 \to \mu^+\mu^-$
$B^0_s \to K^+K^-$	Momentum scale calibration for $B^0_s \to \mu^+\mu^-$
$J/\psi \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$, $\Upsilon(NS) \to \mu^+\mu^-$	Momentum resolution calibration
$B^0_s \to K^-\mu^+\nu_\mu$	Calibration of $B^0_{(s)} \to K^\pm h^\mp$, $B^0_s \to K^-\mu^+\nu_\mu$ backgrounds
$B^0 \to \pi^-\mu^+\nu_\mu$	Calibration of $B^0_{(s)} \to \pi^\pm h^\mp$, $B^0 \to \pi^-\mu^+\nu_\mu$ backgrounds
$B^+ \to J/\psi K^+$	Normalisation
$B^0_s \to J/\psi\,\phi$	Hadronisation fractions for $B^0_s \to \mu^+\mu^-$ normalisation

Table 4.1: Data samples used for the study of $B^0_{(s)} \to \mu^+\mu^-$ decays.

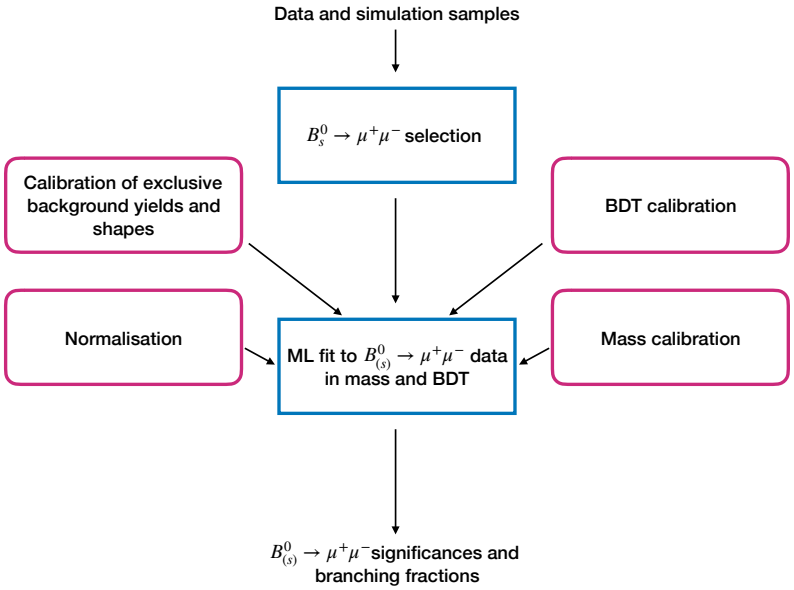

Figure 4.1: A flowchart of the main steps of the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The data is reconstructed and the $B^0_{(s)} \to \mu^+\mu^-$ data sample for the ML fit is obtained using the $B^0_{(s)} \to \mu^+\mu^-$ selection. The four calibration samples are selected in a way as similar as possible to the $B^0_{(s)} \to \mu^+\mu^-$ sample. Each calibration provides a set of calibration parameters, which are used in the ML fit. Finally, the ML fit returns the $B^0_{(s)} \to \mu^+\mu^-$ significances and branching fractions.

4.2 Selection

As discussed in Section 4.1, the $B^0_{(s)} \to \mu^+\mu^-$ decay signature in the LHCb detector consists of two oppositely charged muons with a dimuon invariant mass around the B^0 or B^0_s mass and a consistent, detached vertex because the $B^0_{(s)}$ meson decays after flying for about 0.7 cm. Although this is a straightforward and very clean signature, a strong background rejection is needed given the small branching fractions of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ ($\mathcal{O}(10^{-9})$), which results in the relative number of $B^0_s \to \mu^+\mu^-$ and background candidates before selection being of the order of 10^7 (see *e.g.* Figure 4.2).

In this Section, the backgrounds to $B^0_{(s)} \to \mu^+\mu^-$ decays are introduced and the selection for the $B^0_{(s)} \to \mu^+\mu^-$ sample and calibration samples are explained. For the calibration channels, the selection is made as similar as possible to the $B^0_{(s)} \to \mu^+\mu^-$ selection, to avoid systematic uncertainties from selection differences.

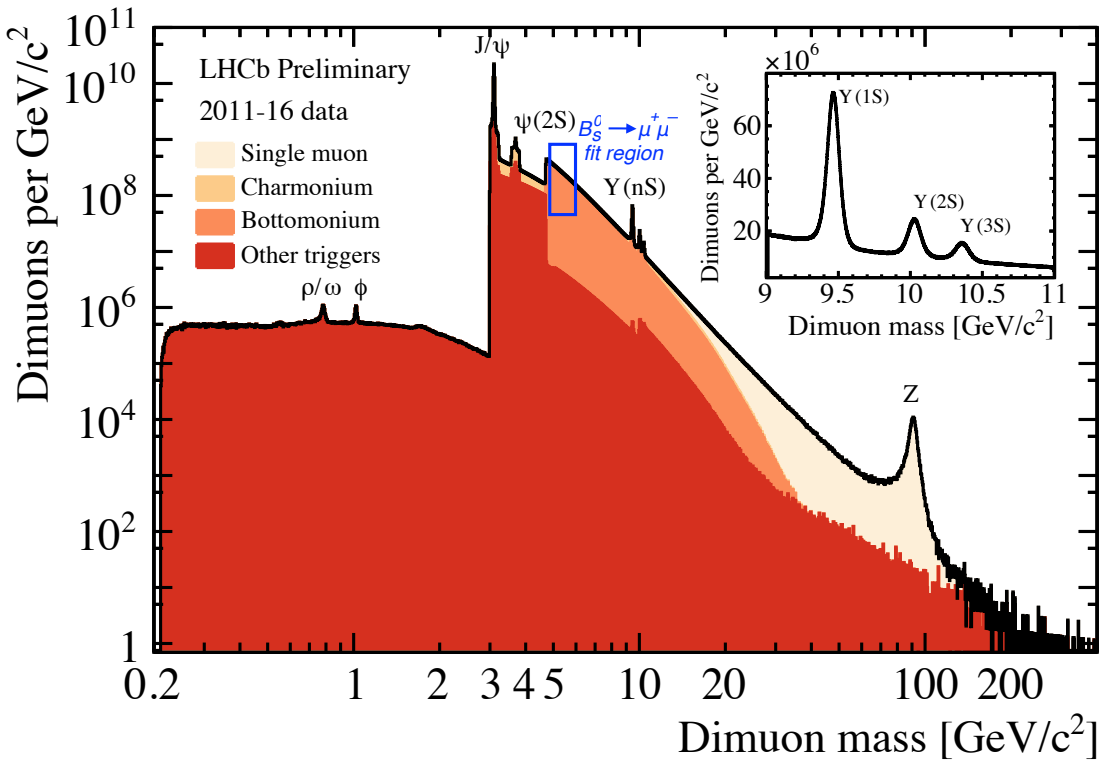

Figure 4.2: Dimuon mass distribution for candidates saved to disk at LHCb from 2011 to 2016 with a loose inclusive dimuon selection and requiring $\text{ProbNN}_\mu > 0.4$ and $\text{ProbNN(ghost)} < 0.1$. The $B^0_{(s)} \to \mu^+\mu^-$ fit region is indicated in blue, and the number of dimuon candidates in that region is around 3×10^8.

Background	Rejection method
$b\bar{b} \to \mu^+\mu^- X$	Vertex quality, muon displacement, muon isolation
$B^0_{(s)} \to h^+h'^-$	Muon identification
$H^0_b \to h^\pm\mu^\mp\nu_\mu$	Muon identification and pointing requirements
$B^+_c \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$	J/ψ veto, muon isolation
$B^+ \to \pi^+\mu^+\mu^-$	Pointing requirements, muon isolation
$B^0 \to \pi^0\mu^+\mu^-$	Pointing requirements

Table 4.2: The major backgrounds for the $B^0_{(s)} \to \mu^+\mu^-$ analysis and the main methods of rejecting them.

$B^0_{(s)} \to \mu^+\mu^-$ backgrounds The main sources of background for $B^0_{(s)} \to \mu^+\mu^-$ decays are shown in Table 4.2.

The dominant background is denoted as $b\bar{b} \to \mu^+\mu^- X$ and is due to two b hadrons that decay to two oppositely charged muons and other particles. As it is constructed from random combinations of muons, this background is referred to as "combinatorial" background. It is the main background for most of the $B^0_{(s)} \to \mu^+\mu^-$ selection, occuring about 10^7 times more often than $B^0_s \to \mu^+\mu^-$.

Other backgrounds originate from the decay of a single B hadron, in which some of the final state particles are missing or misidentified as muons. These backgrounds are referred to as "exclusive" backgrounds. This class of decays contains $B^0_{(s)} \to h^+h'^-$ decays, for which both hadrons are misidentified as muons with a dimuon invariant mass close to the B^0 mass. It occurs about 10^4 times more often than $B^0_s \to \mu^+\mu^-$, and is rejected by PID requirements on both hadrons. Another source of exclusive background are semileptonic decays, with the general form $H^0_b \to h^\pm\mu^\mp\nu_\mu$, where $(H^0_b, h) \in (B^0, \pi^-), (B^0_s, K^-), (\Lambda^0_b, p)$, for which the neutrino is not reconstructed and the hadron is misidentified as a muon. Although they appear about 5 times more often than $B^0_{(s)} \to h^+h'^-$ and only require one hadron to be misidentified, they are seldomly reconstructed with a dimuon invariant mass close to the B^0 or B^0_s mass. The last sources of exclusive background, $B^+_c \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$ and $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$, have missing final state particles, either a muon and a neutrino or a pion. Both are seldomly reconstructed with a dimuon mass in the mass window. The first occurs about 10^4 times more often than $B^0_s \to \mu^+\mu^-$ but is strongly suppressed with a J/ψ veto and muon isolation variables; the second occurs about 30 times more often than $B^0_s \to \mu^+\mu^-$, and the charged decay mode is suppressed with the muon isolation.

Selection procedure While taking data, a trigger is used to select candidates that resemble $B^0_{(s)} \to \mu^+\mu^-$ and calibration channel decays. It is as efficient on signal as possible, while still strongly reducing the event rate. The LHCb trigger is described in Section 3.3, and an overview of the relevant trigger lines for the $B^0_{(s)} \to \mu^+\mu^-$ analysis is given in Section 4.2.1. Following the offline reconstruction of the triggered data, a selection based on one-dimensional cuts is applied (see Section 4.2.2). This selection keeps most of the signal, reduces the main backgrounds and reduces the data to a manageable amount. Afterwards, particle identification (PID) information is used to reject misidentified b hadron decays. The PID variables in use at LHCb have been described in Section 3.6, and the PID selection is discussed in Section 4.2.3. Finally, two multivariate classifiers are used for an optimal separation of signal and combinatorial background. Both are Boosted Decision Trees (BDT). The principle of a BDT is explained in Section 4.2.4. The first BDT, referred to as BDTS, is discussed in Section 4.2.5 and is used for selection: cutting on it keeps 95% of the signal candidates and strongly reduces combinatorial background. A second BDT, discussed in Section 4.2.6, is used for classification of candidates passing the selection. in the maximum likelihood fit (see Section 4.1).

4.2.1 Trigger

Events that are used in this analysis have been triggered by specific lines at each level of the trigger: L0, HLT1 and HLT2. In this Section, an overview of the lines used to trigger the various samples used in this analysis is presented for convenience. The details of the trigger requirements can be found in Section 3.3. Unless otherwise specified, candidates are not required to have passed a specific trigger line in the $B^0_{(s)} \to \mu^+\mu^-$ selection. In Tables 4.3 and 4.4 the most relevant trigger lines are listed for Run 1 and Run 2, respectively.

$B^0_{(s)} \to \mu^+\mu^-$

Events containing $B^0_{(s)} \to \mu^+\mu^-$ candidates are triggered by muons or dimuons with high enough transverse momentum that are detached from the primary vertex. At the hardware (L0) level, $B^0_{(s)} \to \mu^+\mu^-$ candidates are triggered on by the Muon and DiMuon lines, which select a muon with a large p_T or dimuons with a large $p_{T,1} \times p_{T,2}$. In HLT1, $B^0_{(s)} \to \mu^+\mu^-$ is selected by looking for displaced muon tracks with sufficient p_{T}. The displacement is quantified with the significance of the impact parameter. In Run1 the specific lines are HLT1TrackMuon and HLT1TrackAllL0, while in Run2

Trigger level	L0	HLT1	HLT2
$B^0_{(s)} \to \mu^+\mu^-$	L0Muon or L0DiMuon	HLT1TrackMuon HLT1TrackAllL0	HLT2DiMuonB
$B \to J/\psi X$			HLT2TopoNBody HLT2DiMuonDetJpsi
$J/\psi \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$, $\Upsilon(NS) \to \mu^+\mu^-$		HLT1SingleMuonHighPT HLT1DiMuonHighMass	HLT2DiMuonJPsiHighPT HLT2DiMuonPsi2SHighPT HLT2DiMuonB
$B^0_{(s)} \to h^+h'^-$ $B^0_{(s)} \to h^+h'^-$	L0Hadron L0TIS	HLT1TrackAllL0	HLT2Topo2Body HLT2B2HH

Table 4.3: The main trigger lines used for this analysis in Run 1.

Trigger level	L0	HLT1	HLT2
$B^0_{(s)} \to \mu^+\mu^-$	L0Muon or L0DiMuon	HLT1TrackMuon(MVA) HLT1TrackMVA	HLT2DiMuonB
$B \to J/\psi X$			HLT2TopoNBody HLT2DiMuonDetJpsi
$J/\psi \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$, $\Upsilon(NS) \to \mu^+\mu^-$		HLT1SingleMuonHighPT HLT1DiMuonHighMass	HLT2DiMuonJPsiHighPT HLT2DiMuonPsi2SHighPT HLT2DiMuonB
$B^0_{(s)} \to h^+h'^-$ $B^0_{(s)} \to h^+h'^-$	L0Hadron L0TIS	HLT1TrackMVA	HLT2Topo2Body HLT2B2HH

Table 4.4: The main trigger lines used for this analysis in Run 2.

HLT1TrackMVA, HLT1TrackMuon and HLT1TrackMuonMVA are used. For HLT2, there are two trigger lines, each of which covering all $B^0_{(s)} \to \mu^+\mu^-$ signal candidates: HLT2DiMuonB and HLT2DiMuonDetachedHeavy. The first accepts dimuon pairs with a mass above 4.7 GeV/c^2 and a good dimuon vertex quality. The second does not require such a good dimuon vertex quality and has a lower dimuon mass bound, but requires a significant flight distance from the primary vertex.

$B \to J/\psi X$

For $B \to J/\psi X$ decays, the trigger selection is based on the muons from the J/ψ decay and therefore very similar to $B^0_{(s)} \to \mu^+\mu^-$. The overwhelming majority of $B \to J/\psi X$ candidates are triggered by the same L0 and HLT1 lines as $B^0_{(s)} \to \mu^+\mu^-$. At the HLT2 level, there are two lines that together triggered 99% of the HLT2 $B \to J/\psi X$ candidates: HLT2DiMuonDetachedJpsi, which triggers on dimuon pairs with a significant flight distance and an invariant mass around the J/ψ mass (similar

to the $B^0_{(s)} \to \mu^+\mu^-$ detached line but with different mass windows), and the inclusive trigger line for B decays, HLT2TopoNBody, with N $\in \{2, 3, 4\}$.

$J/\psi \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$, $\Upsilon(NS) \to \mu^+\mu^-$

These are prompt decays, for which no detachment is required in the trigger. Only the L0 lines are the same as for the $B^0_{(s)} \to \mu^+\mu^-$ sample. For charmonium decays, the offline selection requires HLT1DiMuonHighMass and HLT2DiMuonJpsiHighPT or HLT2DiMuonPsi2SHighPT. Bottomonium decays are triggered in HLT1 by HLT1DiMuonHighMass or HLT1SingleMuonHighPT for over 97% of the candidates, and in HLT2 by HLT2DiMuonB, the same line that triggers most $B^0_{(s)} \to \mu^+\mu^-$ decays, for over 95% of the candidates.

$B^0_{(s)} \to h^+h'^-$

For $B^0_{(s)} \to h^+h'^-$ decays, the trigger selection is very different from those described before, because it contains hadrons instead of muons in the final state. In L0, $B^0_{(s)} \to h^+h'^-$ candidates are saved because they passed the L0Hadron trigger or because anything else in the event fired the L0 trigger (TIS). In HLT1, 96% of the $B^0_{(s)} \to h^+h'^-$ candidates are triggered by the B decay lines of HLT1TrackAllL0 in Run 1 and HLT1TrackMVA in Run 2. In HLT2, there are two trigger lines that trigger 95% of the $B^0_{(s)} \to h^+h'^-$ candidates: HLT2Topo2Body and HLT2B2HH. HLT2Topo2Body is part of the inclusive topological trigger lines for B decays, while HLT2B2HH is a line specifically designed to select $B^0_{(s)} \to h^+h'^-$ decays.

Note that while no trigger requirements are included for the $B^0_{(s)} \to h^+h'^-$ samples used for the momentum scale calibration, the sample for the BDT calibration and normalisation uses only candidates that are TIS at L0 and HLT1 and triggered by HLT2B2HH in HLT2, in order to limit the systematic uncertainty on the trigger efficiency estimate.

4.2.2 Offline selection

The offline selection is used to reduce the data sample to a size that is manageable to produce ntuples and make an optimised selection. Just like the trigger selection, the offline selection should be as efficiency as possible on signal while reducing the background significantly. For the calibration channels, it should be as similar as possible to the signal selection. An important parameter for the offline selection is the mass window, which differs for each sample in order to tune the sample size.

The $B^0_{(s)} \to \mu^+\mu^-$ offline selection is made to reject combinatorial backgrounds and misidentified backgrounds. A number of cuts to reject candidates with a low reconstruction quality are applied, as shown in Table 4.5. Two good tracks with a transverse momentum range and maximum momentum in the detector acceptance are required. In addition, the B candidate needs to have a minimum transverse momentum, a positive lifetime by requiring a positive DIRA, and a maximum lifetime to reject unphysical candidates. Combinatorial background is rejected by requiring a B candidate with a good vertex, a significant flight distance χ^2, and an impact parameter consistent with the primary vertex, while requiring the tracks to have a significant impact parameter with respect to the primary vertex. Both muon tracks are required to pass isMuon to reduce background from misidentified paricles. To reduce the rate, the $B^0_{(s)} \to \mu^+\mu^-$ selection includes a loose mass requirement on the dimuon mass of the candidate, which should be within 1200 MeV/c^2 of the B^0_s mass.

A J/ψ veto is included to reject background from $B^+_c \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$ decays in the $B^0_{(s)} \to \mu^+\mu^-$ sample. If a dimuon candidate can be reconstructed from one of the muons from the $B^0_{(s)} \to \mu^+\mu^-$ candidate and one oppositely charged muon from the rest of the event, and it has a mass close to the J/ψ mass, it is likely that the $B^0_{(s)} \to \mu^+\mu^-$ candidate came from this background. Therefore, $B^0_{(s)} \to \mu^+\mu^-$ candidates where either of the muons forms a dimuon candidate with a mass within 30 MeV/c^2 of the J/ψ mass are rejected. Muons from the rest of the event are required to satisfy ProbNN$_\mu > 0.3$ in order to be well identified; if multiple background muons satisfy the PID requirement, the dimuon candidate with the mass closest to the J/ψ mass is kept. The J/ψ veto rejects 64% of this background while keeping 99.8% of $B^0_{(s)} \to \mu^+\mu^-$ signal candidates.

The $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to h^-\mu^+\nu_\mu$ samples have the same offline selection, similar to the $B^0_{(s)} \to \mu^+\mu^-$ selection and only differing by not having any PID requirements. As this increases the rate, the dimuon mass is required to be within 600 MeV/c^2 of the B^0_s mass.

The offline selection for $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\,\phi$ samples, hereafter referred to as the $B \to J/\psi X$ selection, is slightly different, as there are more final state particles. It applies the same requirements as the $B^0_{(s)} \to \mu^+\mu^-$ selection on each track, including isMuon for the tracks from the J/ψ, and includes the vertex quality, flight distance, DIRA and DOCA requirements for the J/ψ candidate, as those variables would behave differently for three-track vertices. An additional very loose cut on the B candidate vertex χ^2/ndf is included, and a mass window is required on both the dimuon and the total invariant mass.

Cut	Applied on	$B^0_{(s)} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h'^-$ $B^0_{(s)} \to h^-\mu^+\nu_\mu$	Applied on	$B^+ \to J/\psi K^+$
track χ^2/ndf	μ/h	$< 3(4)$	μ/K	$< 3(4)$
ghost probability	μ/h	$< 0.3(0.4)$	μ/K	$< 0.3(0.4)$
IP χ^2	μ/h	> 25	μ/K	> 25
p_{T}	μ/h	>0.25 GeV/c	μ/K	>0.25 GeV/c
p_{T}	μ/h	<40 GeV/c	μ/K	<40 GeV/c
p	μ/h	<500 GeV/c	μ/K	<500 GeV/c
isMuon	μ only	True	μ only	True
DIRA	B	>0	J/ψ	>0
DOCA	B	<0.3 mm	J/ψ	<0.3 mm
Vertex χ^2	B	< 9	J/ψ	< 9
Flight dist. χ^2	B	> 225	J/ψ	> 225
M($\mu^+\mu^-$)	B	$\lvert M(\mu^+\mu^-) - M_{B^0_s}\rvert$ < 1200 MeV/c^2	J/ψ	$\lvert M(\mu^+\mu^-) - M(J/\psi)\rvert$ < 60 MeV/c^2
τ	B	$< 9 \times \tau(B^0_s)$	B	$< 9 \times \tau(B^0_s)$
p_{T}	B	> 500 MeV/c	B	> 500 MeV/c
M($\mu^+\mu^-K^+$)			B	$\lvert M(\mu^+\mu^-K) - M(B^+)\rvert$ < 500 MeV/c^2
vertex χ^2			B	< 45

Table 4.5: The offline selection used for $B^0_{(s)} \to \mu^+\mu^-$,$B^0_{(s)} \to h^+h'^-$, $B^0_{(s)} \to h^-\mu^+\nu_\mu$, and $B^+ \to J/\psi K^+$ in Run 1. For Run 2, the thresholds for track χ^2/ndf and ghost probability changed and are indicated in brackets.

Quarkonium selection

A different kind of offline selection is used for the quarkonium samples, as most of the quarkonia candidates are not detached from the primary vertex. The goal of this selection is then to reject badly reconstructed candidates and prompt combinatorial background. Tracks are required to have a minimum momentum and transverse momentum; In the case of charmonium resonances, the dimuon candidate should also have a minimum transverse momentum. The track and vertex quality should be good. To reject misidentified combinatorial, a loose PID cut of $\mathrm{DLL}_{\mu\pi} > 0$ is included on both tracks. Finally, the mass of the dimuon candidate should either be close to a charmonium resonance mass or above 8.5 GeV/c^2 for the bottomonium resonances. All offline cuts for quarkonium samples are shown in Table 4.6.

Cut	Applied on	$J/\psi \to \mu^+\mu^-$	$\psi(2S) \to \mu^+\mu^-$	$\Upsilon(NS) \to \mu^+\mu^-$
p_{T}	μ	> 0.65 GeV/c	> 1 GeV/c	> 1 GeV/c
p	μ	> 10 GeV/c	> 10 GeV/c	> 8 GeV/c
track χ^2/ndf	μ	$< 3(4)$	$< 3(4)$	$< 3(4)$
isMuon	μ	True	True	True
DLL($\mu - \pi$)	μ	> 0	> 0	> 0
M($\mu^+\mu^-$)-M($q\overline{q}$)	$q\overline{q}$	< 80 MeV/c^2	< 100 MeV/c^2	
M($\mu^+\mu^-$)	$q\overline{q}$			8.5 GeV/c^2
p_{T}	$q\overline{q}$	> 3 GeV/c	> 3 GeV/c	-
Vertex χ^2	$q\overline{q}$	< 20	< 20	< 20

Table 4.6: The offline selection for each of the three $q\overline{q} \to \mu^+\mu^-$ samples in Run 1. For Run 2, the thresholds for track χ^2/ndf and ghost probability changed and are indicated in brackets.

4.2.3 Particle Identification

The goal of the Particle Identification (PID) selection is to further reduce the background from misidentified particles. For the $B^0_{(s)} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to h^-\mu^+\nu_\mu$ samples, this is essential. For $B^+ \to J/\psi K^+$, the only misidentified background is $B^+ \to J/\psi\pi^+$, which is small relative to $B^+ \to J/\psi K^+$ and properly modeled in the mass fit, so no additional PID selection is required. For $J/\psi \to \mu^+\mu^-$,$\psi(2S) \to \mu^+\mu^-$, and $\Upsilon(NS) \to \mu^+\mu^-$, no additional PID is required either, as the only background is the combinatorial one, which is sufficiently reduced by the offline selection.

There are three kinds of PID variables: isMuon, DLL and ProbNN. They are described in detail in Section 3.6.

For $B^0_{(s)} \to \mu^+\mu^-$, the goal of the PID selection is to reject backgrounds from $B^0_{(s)} \to h^+h'^-$ and $H^0_b \to h^\pm\mu^\mp\nu_\mu$, specifically $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$, and $\Lambda^0_b \to p\mu^-\nu$. The PID selection is reoptimised with respect to previous analyses to exploit ProbNN variables that were not previously available. A number of PID variable combinations are investigated:

- $\mathrm{DLL}_{K\pi} < 10$ and $\mathrm{DLL}_{\mu\pi} > -5$, which was already used in the previous analysis [41];

- ProbNN $\equiv$ ProbNN$_\mu \times (1 -$ ProbNN$_K) \times (1 -$ ProbNN$_p)$, tested at ProbNN $> 0.1, 0.2, 0.3, 0.4, 0.5$;
- ProbNN$_{\text{nop}} \equiv$ ProbNN$_\mu \times (1 -$ ProbNN$_K)$, tested at ProbNN$_{\text{nop}} > 0.3, 0.4, 0.5$, similar to ProbNN but with a slightly higher signal efficiency and a worse rejection of $\Lambda^0_b \to p\mu^-\nu$.

To assess the impact of the PID selection on the final $B^0_{(s)} \to \mu^+\mu^-$ measurement, toy datasets with the expected number of signal and background candidates are generated for different PID cuts and the expected sensitivity for the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions (at their SM values) is calculated. As $B^0_s \to \mu^+\mu^-$ is less affected by exclusive backgrounds than $B^0 \to \mu^+\mu^-$, the optimal working point is different for each and it is chosen to optimise for $B^0 \to \mu^+\mu^-$. The optimal PID cut is found to be ProbNN > 0.4.

Due to this optimisation, misidentified exclusive backgrounds are reduced by a factor 2 to 5 with respect to the previous analysis [41], and the sensitivity for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays improved by about 10% and 15% respectively [69]. A new tuning of the ProbNN variables, that is optimised on Run 2 data, is applied to the 2016 dataset. A different PID cut, ProbNN > 0.8, is chosen in order to have a similar signal efficiency for $B^0_s \to \mu^+\mu^-$ decays and an improved rejection of misidentified backgrounds.

For $B^0_{(s)} \to h^+h'^-$, the PID selection is needed to separate different $B^0_{(s)} \to h^+h'^-$ final states (KK, $K\pi$ and $\pi^+\pi^-$), as well as to reduce the background from $\Lambda^0_b \to p\pi^-$ and $\Lambda^0_b \to pK^-$ decays. The DLL variables are used in this case, specifically DLL$_{K\pi}$, DLL$_{p\pi}$ and the combination DLL$_{Kp}$ = DLL$_{Kp}$- DLL$_{p\pi}$. DLL$_{K\pi}$ is used to separate kaons and pions, while DLL$_{Kp}$ and DLL$_{p\pi}$ are used to separate either kaons or pions from protons. A cut κ is applied as DLL$_{K\pi}$, (resp. DLL$_{Kp}$) $> \kappa$, (resp. κ_p) for kaons and DLL$_{K\pi}$, (resp.DLL$_{p\pi}$) $< -\kappa$, (resp.κ_p) for pions. For the momentum scale calibration, both are used to avoid any bias on the mean of the $B^0_{(s)} \to h^+h'^-$ shapes and κ is varied from 5 to 20 while κ_p is varied from -2 to 2. For the normalisation and BDT calibration, as statistics are limited, only DLL$_{K\pi}$ is used, κ is varied from 5 to 10, and the contribution from $\Lambda^0_b \to p\pi^-$ and $\Lambda^0_b \to pK^-$ decays is separately modeled in the invariant mass fits.

For $B^0_{(s)} \to h^-\mu^+\nu_\mu$, the goal of the PID selection is to get $K\mu$ and $\pi\mu$ samples with a negligible contribution from misidentified backgrounds. As there is no PID in the offline selection, PID cuts are applied to the hadron and the muon candidates. Specifically, the requirement for muons is ProbNN$_\mu > 0.5$, for pions ProbNN$_\pi \times (1 -$

$\text{ProbNN}_K) \times (1 - \text{ProbNN}_p) > 0.5$, and for kaons $\text{ProbNN}_K \times (1 - \text{ProbNN}_\pi) \times (1 - \text{ProbNN}_p) > 0.5$. The resulting samples are well-identified and useable as a control region for misidentified backgrounds.

4.2.4 Boosted Decision Trees

In the $B^0_{(s)} \to \mu^+\mu^-$ analysis, two multivariate classifiers are used to classify data samples as being more like signal or more like combinatorial background. The main advantage of multivariate classifers is that they take into account correlations between variables to optimally separate two samples. Boosted Decision Trees (BDT) are used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis, as they are found to have the best signal-background separation of the investigated classifiers.

A decision tree is a classifier that separates signal and background by requiring a series of thresholds on individual variables. Using a measure of inequality, for example the Gini index, it takes the full sample and applies the selection on one of the variables that gives the biggest increase in inequality. For each subsample, the same procedure is followed recursively, until a condition is fulfilled for that sample which stops the procedure. The advantage of a decision tree is that it is simple to understand and examine, as it only applies a series of one-dimensional selections. Its main disadvantage is that it is prone to be unstable and will have a limited sensitivity.

The idea of boosting is to build one strong classifier from many weak ones by training many decision trees in series and emphasising previously misclassified entries in training new decision trees. Both BDTs used for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are trained using the AdaBoost algorithm [70]. AdaBoost gives more weight to previously misclassified entries when training new decision trees, based on how often they are misclassified. After the individual trees are trained, a single classifier, the BDT, is obtained. by combining the trees with weights assigned to each tree based on their performance.

4.2.5 BDT for selection (BDTS)

The BDTS is a BDT that is trained for selection. Its goal is to remove a significant part of the combinatorial background, while keeping most signal candidates. Addtionally, it should have a similar response for $B^0_{(s)} \to \mu^+\mu^-$,$B^0_{(s)} \to h^+h'^-$ and $B^+ \to J/\psi K^+$ candidates, so that it can be applied on all normalisation channels. It is trained with

simulated samples of $B^0_s \to \mu^+\mu^-$ decays as signal and of $b\bar{b} \to \mu^+\mu^- X$ candidates as background.

The input variables for the BDTS are chosen to have a similar response for $B^+ \to J/\psi K^+$ and $B^0_{(s)} \to \mu^+\mu^-$ decays. They are:

- The B meson candidate impact parameter with respect to the PV;
- The B meson candidate impact parameter χ^2 with respect to the PV;
- The dimuon vertex χ^2;
- The DIRA; *i.e.* the angle between the B momentum and the vector from the primary to the secondary vertex;
- The DOCA; *i.e.* the distance of closest approach for the two muon tracks;
- The minimum child muon impact parameter with respect to the PV.

By carefully selecting these input variables, the BDTS response is very similar for $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$, such that systematic uncertainties from mismodelling of these variables in simulation are negligible.

The BDTS distribution is "flattened" for signal, which means making it uniform from 0 to 1 for signal, with the most background-like candidate having a BDT value of 0. By flattening the distribution, it is intuitively simple to grasp the efficiency for signal in a given BDT range. The $B^0_{(s)} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h'^-$, $B^0_{(s)} \to h^-\mu^+\nu_\mu$ and $B^+ \to J/\psi K^+$ samples are all selected with the same requirement, namely BDTS > 0.05. It rejects 70% of the combinatorial background while keeping 93% of the $B^0_{(s)} \to \mu^+\mu^-$ candidates. For comparison, a cut-based selection would only have rejected 60% of the background for the same signal efficiency.

4.2.6 BDT for classification

The goal of the classification BDT (henceforth referred to simply as BDT) is to have the strongest possible separation of signal and combinatorial background, using the full event information. Therefore, the BDT only has a similar response for $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ decays and is not computed on any $B^+ \to J/\psi K^+$ samples.

In the selection procedure, the *isolation* of the signal candidate tracks is not yet used. An isolated track is one that is not associated with any other tracks in the event. The tracks from $B^0_{(s)} \to \mu^+\mu^-$ candidates are expected to be isolated with

respect to all tracks from the rest of the event. This is not the case for combinatorial background, which is dominated by $b\bar{b} \to \mu^+\mu^- X$ decays. For such decays, the muon tracks should be associated with tracks from the rest of the event; those tracks from the rest of the vent are non-isolated.

To determine the best possible isolation, two isolation BDTs are trained on simulated samples to score tracks from the rest of the event and return an isolation value with respect to either muon track. The two BDTs are trained on long and VELO tracks respectively, (defined in Section 3.2.1).

Background tracks are defined as non-isolated if they share an ancestor with the muon track being considered, and defined as isolated if they do not. Non-isolated tracks are taken from simulated samples of $b\bar{b} \to \mu^+\mu^- X$ candidates and isolated tracks are taken from simulated samples of $B^0_s \to \mu^+\mu^-$ decays.

The variables chosen for the long track isolation used all of the available topology and kinematical information in the event for the optimal identification of non-isolated tracks; they are

- The minimum impact parameter χ^2 of the track compared to any PV in the event;
- The (signed) distance between the track-muon vertex and the PV;
- The (signed) distance between the track-muon vertex and the $B^0_{(s)} \to \mu^+\mu^-$ vertex;
- The distance of closest approach (DOCA) between the track and the muon;
- The angle between the track and the muon;
- $f_c = \frac{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV})}{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV}) + P_{T_\mu} + P_{T_{trk}}}$, where $\alpha^{\mu+trk,PV}$ is the angle between the sum of the muon and track momenta $(\vec{P}_\mu, \vec{P}_{trk})$ and the direction defined by the PV and the (track,muon) vertex. If the track-muon system originates from the PV, $f_c \to 0$;
- The difference in pseudorapidity $\Delta\eta$ for the track and the muon;
- The difference in azimuthal angle $\Delta\phi$ for the track and the muon;
- Track p_T;

For the VELO track isolation, the first six variables are also used; adding the other variables does not improve its performance.

From the isolation BDTs, a set of isolation scores is obtained for each muon track; the larger the score, the less isolating a track from the rest of the event is for the muon track. The isolation of a $B^0_s \to \mu^+\mu^-$ candidate is defined as the sum of the largest isolation scores for each muon track, where those scores can be computed with respect to the same background track.

The classification BDT is trained on the same samples, including the isolation BDTs and adding information on the B candidate and muon tracks as follows:

1. Start with the most distriminating variable (long track isolation) as the first BDT input;
2. Add another variable to the BDT inputs temporarily;
3. Train the BDT and measure the background yield from $b\bar{b} \to \mu^+\mu^- X$ in the most sensitive bin (highest value of BDT). If this bin is zero, keep going to less sensitive bins until a bin has N > 0;
4. Perform steps 2. and 3. for every variable till all variables have been tested;
5. Add the best variable (with the smallest background yield) to the list of BDT input variables;
6. Run steps 2-5 until no improvement is found by adding more variables.

With this procedure, it is simple to quickly test a large number of variables and their combinations. However, due to statistical fluctuations in the most sensitive BDT bins, care is taken to interpret the results. The selected variables are:

- Long track isolation;
- VELO track isolation;
- The angular separation between the two muons ($\Delta_R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$);
- The minimum IP χ^2 of both muons with respect to the PV;
- The DIRA of the B with respect to the PV;
- The dimuon vertex χ^2;
- The B IP χ^2 with respect to the PV;

The last three variables are also used in the BDTS classifier.

After training, the BDT distribution was flattened for signal, analogously to what is done for the BDTS. To optimise the analysis sensitivity, the $B^0_{(s)} \to \mu^+\mu^-$

sample is divided into BDT bins. The binning of the BDT is chosen to be $[0.00, 0.25, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00]$, as this was found to be optimal in the previous $B^0_{(s)} \to \mu^+\mu^-$ analysis. The determination of the fraction of signal candidate in each bin (BDT calibration) will be discussed in Section 4.3.

To make sure the BDT and mass distribution can be used in the fit as independent variables, the correlation of the BDT with the invariant mass distribution was checked. A linear correlation of 3 to 5% was found for $b\bar{b} \to \mu^+\mu^- X$ candidates in simulated samples, which was deemed small enough to not affect the shape of this background. The linear correlation for simulated signal samples was found to be negligible.

4.3 BDT calibration

The BDT used to classify $B^0_{(s)} \to \mu^+\mu^-$ candidates as more signal- or background-like is described in Section 4.2.6. As the ML fit to determine the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions is performed in BDT bins, the fraction of $B^0_{(s)} \to \mu^+\mu^-$ signal candidates per BDT bin, $f_{B^0_{(s)}\to\mu^+\mu^-,\,\text{data}}$, has to be well determined. The procedure to determine these fractions is called the BDT calibration, and is preferably performed on data to take into account any differences between data and simulation.

As discussed in Section 4.2.6, the BDT is trained to separate two-body neutral B meson decays from background. Therefore, the BDT calibration is performed with data of $B^0_{(s)} \to h^+h'^-$ decays, where $h \in (\pi, K)$. From this data, the fraciton of $B^0 \to K^+\pi^-$ candidates per BDT bin, $f_{B^0\to K^+\pi^-,\,\text{data}}$ is determined, as described in Section 4.3.1. The $B^0 \to K^+\pi^-$ fractions are then converted into the $B^0_{(s)} \to \mu^+\mu^-$ fractions taking into account the differences between muons and hadrons, which originate from the trigger and the PID (see Sections 4.2.1 and 4.2.3, respectively):

$$f_{B^0_{(s)}\to\mu^+\mu^-,\,\text{data}} = f_{B^0\to K^+\pi^-,\,\text{data}} \frac{\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\text{PID|trig.}}}{\varepsilon_{B^0\to K^+\pi^-,\,\text{PID|trig.}}} \frac{\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\text{trig.}}}{\varepsilon_{B^0\to K^+\pi^-,\,\text{trig.}}} \, k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}, \qquad (4.1)$$

where $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\text{PID|trig.}}$ and $\varepsilon_{B^0\to K^+\pi^-,\,\text{PID|trig.}}$ are the PID efficiencies per BDT bin after requiring the trigger selection, and $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\text{trig.}}$ and $\varepsilon_{B^0\to K^+\pi^-,\,\text{trig.}}$ are the trigger efficiencies per BDT bin. One of these, $\varepsilon_{B^0\to K^+\pi^-,\,\text{PID|trig.}}$, is determined using fits to $B^0_{(s)} \to h^+h'^-$ data without a PID selection applied, as illustrated in Section 4.3.2. The procedure to determine the other efficiencies is described in Section 4.3.5. An additional correction stems from the lifetime of the $B^0_s \to \mu^+\mu^-$ decay, which had not been measured yet: as the BDT is correlated with the B candidate lifetime, a lifetime-hypothesis dependent correction, $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$, is included. This correction is discussed in Section 4.3.6. The BDT distribution is determined seperately for each year and combined for Run 1 and Run 2, weighted for the relative luminosity of each year.

4.3.1 The $B^0 \to K^+\pi^-$ BDT distribution in data

The BDT distribution of $B^0 \to K^+\pi^-$ candidates in the data sample, $f_{B^0\to K^+\pi^-,\,\text{data}}$, is determined by measuring the yields per BDT bin with ML fits using two approaches: the inclusive approach (see Section 4.3.3) uses the full $B^0_{(s)} \to h^+h'^-$ sample without PID requirements, whereas the exclusive approach, described here, uses

PID information to select well-identified $B^0 \to K^+\pi^-$ candidates. The inclusive and exclusive approach should give consistent results after correcting for the PID efficiency $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}}$ and for the rate of $B^0 \to K^+\pi^-$ decays relative to other $B^0_{(s)} \to h^+h'^-$ decays.

For the exclusive approach, $B^0 \to K^+\pi^-$ is used as it is the dominant $B^0_{(s)} \to h^+h'^-$ mode (around 60% of $B^0_{(s)} \to h^+h'^-$ decays at LHCb are $B^0 \to K^+\pi^-$). An advantage of the exclusive approach is that the fit model is greatly simplified, because the PID selection strongly reduces backgrounds from $B^0_{(s)} \to h^+h'^-$, $\Lambda^0_b \to ph^-$ decays and misidentified combinatorial background.

As the main mis-identified backgrounds to $B^0 \to K^+\pi^-$ are $B^0 \to \pi^+\pi^-$ and $B^0_s \to K^+K^-$, the PID selection is based on the $\mathrm{DLL}_{K\pi}$ variable, used to separate kaons from pions. Positive values of $\mathrm{DLL}_{K\pi}$ indicate hadrons that are more likely to be kaon than pion candidates. Kaons are required to satisfy $\mathrm{DLL}_{K\pi} > \kappa$ and pions $\mathrm{DLL}_{K\pi} < -\kappa$, where κ is the value of the PID cut. The nominal PID selection ($\kappa = 5$) makes the contribution from misidentified $B^0_{(s)} \to h^+h'^-$ decays negligible.

Exclusive fit model The mass window for the exclusive fit needs to be considered carefully because the preselection includes a requirement on the $\mu^+\mu^-$ invariant mass, that causes the mass acceptance not to be uniform under the $K\pi$ mass hypothesis. The mass acceptance is estimated by generating a uniform $\mu^+\mu^-$ mass distribution and calculating the $K\pi$ mass distribution with the momenta of both child hadrons taken from $B^0 \to K^+\pi^-$ simulated samples. It is found that the non-uniformity of the mass acceptance is negligible after requiring a minimum $K\pi$ mass of 5200 MeV/c^2. This requirement removes 0.5% to 1.1% of the signal candidates, estimated from the fraction of the $B^0 \to K^+\pi^-$ mass shapes in data fits outside the mass window. As a systematic check, the BDT calibration is performed after correcting for this effect and the difference is taken as a systematic uncertainty.

After the PID selection, the fit model consist only of shapes for $B^0 \to K^+\pi^-$, $\bar{B}^0_s \to K^+\pi^-$, and $\Lambda^0_b \to ph^-$ decays and for the combinatorial background. The $B^0 \to K^+\pi^-$ and $\bar{B}^0_s \to K^+\pi^-$ shapes are modeled with a double-sided Crystal Ball function, which is a Gaussian with enhanced tails, described in more detail in Appendix A.1. The mean and resolution of the $B^0 \to K^+\pi^-$ shape are left free in the fit, while its tail parameters are constrained using the results of fits to simulated samples and adding a Gaussian term to the likelihood of the data fit. Since it is relatively small, the $\bar{B}^0_s \to K^+\pi^-$ shape is fixed relative to the $B^0 \to K^+\pi^-$ shape: its mean is fixed using the B^0 and B^0_s meson mass difference from the PDG [36]; its

resolution is fixed using the ratio of B^0 and B^0_s resolutions determined in Section 4.4.2; its tail parameters are the same as those of the $B^0 \to K^+\pi^-$ shape. The $\Lambda^0_b \to ph^-$ shape is a sum of two Crystal Ball functions, one for $\Lambda^0_b \to pK^-$ and one for $\Lambda^0_b \to p\pi^-$, with their relative fraction fixed from their relative branching fractions [71]. The two shapes are determined with a fit to $\Lambda^0_b \to pK^-$ and $\Lambda^0_b \to p\pi^-$ simulated samples and the combined $\Lambda^0_b \to ph^-$ shape is fixed except for the mean, which is allowed to vary freely. Finally, the combinatorial background is modelled with an exponential whose yield and shape parameter are floating in the fit. As a systematic check of the fit model, the BDT calibration is repeated with the $B^0 \to K^+\pi^-$ and $\bar{B}^0_s \to K^+\pi^-$ shapes replaced by the sum of two Gaussians; the difference with respect to the nominal fraction per BDT bin is assigned as a systematic uncertainty.

Results The $B^0 \to K^+\pi^-$ yield is determined separately for each year. The exclusive fits to the 2016 data sample with the nominal PID cut, $\kappa = 5$, are shown in Figure 4.3. The fraction in the $i-$th BDT bin is determined as $f_{B^0\to K^+\pi^-,\,\mathrm{data}}$ (i) $= N_{B^0\to K^+\pi^-}(i)/N_{B^0\to K^+\pi^-,\mathrm{tot}}$.

4.3.2 $B^0 \to K^+\pi^-$ PID efficiency

As the PID requirements are not very well simulated, the $B^0 \to K^+\pi^-$ PID efficiency, $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}}$, is calibrated with data using the PIDCalib package. This package uses calibration samples of decay modes in which the final state particles are identified without any PID selection, namely $D^{*+} \to (D^0 \to K^+\pi^-)\pi^+$. To subtract background from these calibration samples, the *sPlot* method is used [72]: signal and background are separated using a maximum likelihood fit to the $D^0\pi^+$ invariant mass to obtain the PID variable distribution for signal.

To correct for kinematical differences between the PIDCalib control mode and $B^0 \to K^+\pi^-$, the PID efficiencies $\varepsilon_\pi, \varepsilon_K$ are determined in bins of ($p_{\mathrm{hadron}}, \eta_{\mathrm{hadron}}$, nTracks). The BDT calibration us repeated with a finer binning and the difference is included as a systematic uncertainty.

The $B^0 \to K^+\pi^-$ PID efficiency per candidate is then determined by

$$\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}} = \frac{\sum_i^{N_{\mathrm{sample}}} \varepsilon_{\pi,i}(p_-,\eta_-,nTracks)\ \varepsilon_{K,i}(p_+,\eta_+,nTracks)}{N_{\mathrm{sample}}}, \quad (4.2)$$

$$\varepsilon_{\bar{B}^0\to K^-\pi^+,\,\mathrm{PID|trig.}} = \frac{\sum_i^{N_{\mathrm{sample}}} \varepsilon_{\pi,i}(p_+,\eta_+,nTracks)\ \varepsilon_{K,i}(p_-,\eta_-,nTracks)}{N_{\mathrm{sample}}}. \quad (4.3)$$

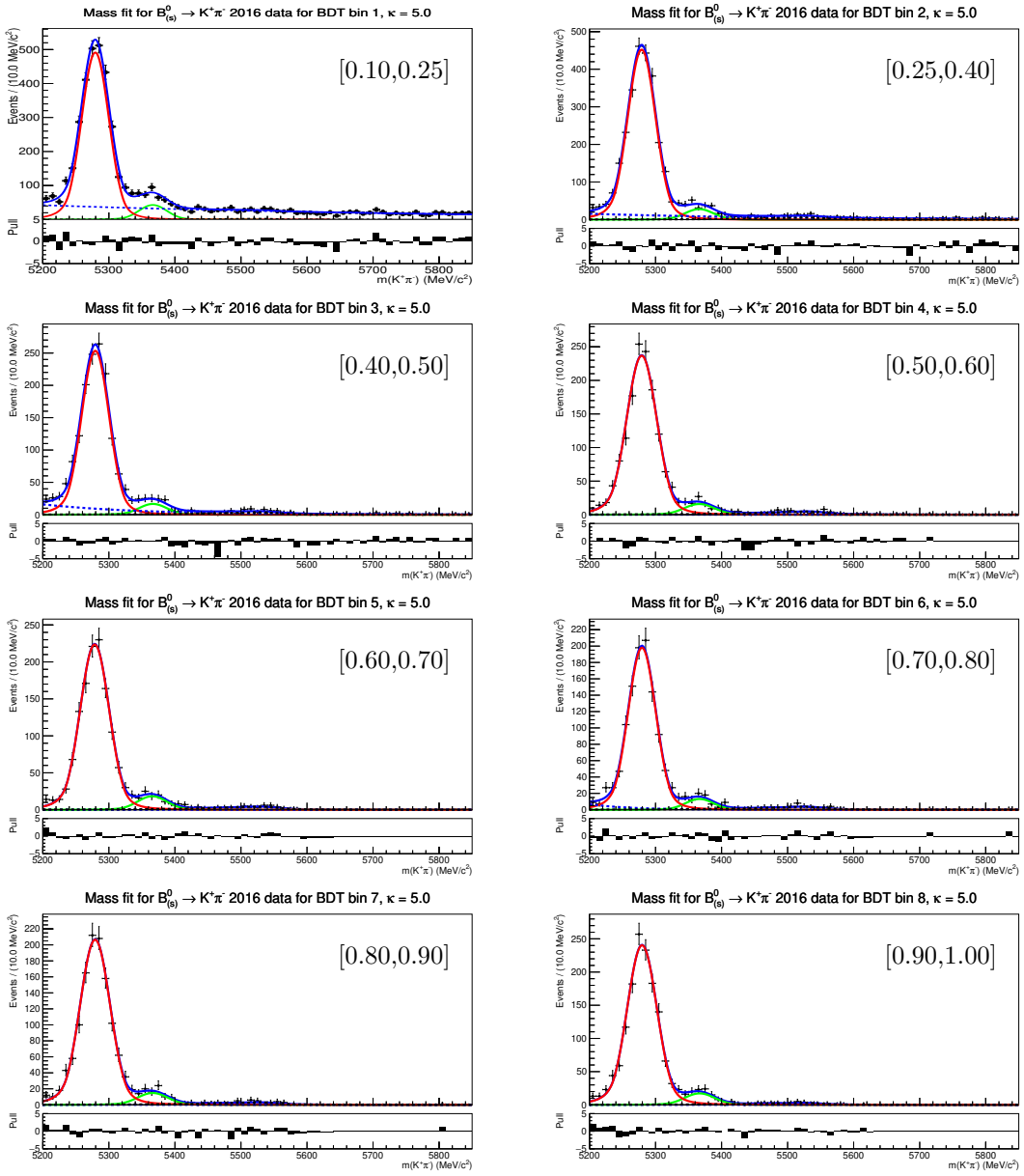

Figure 4.3: Invariant mass distributions of $B^0 \to K^+\pi^-$ candidates from 2016 data per BDT bin for $\kappa = 5$. The red and green solid lines shows the B^0 and B^0_s signal contribution, the red dashed line the $\Lambda^0_b \to ph^-$ background and the blue dashed line the combinatorial background.

The efficiency is normalised to the total size of the representative sample N_{sample}. The $B^0 \to K^+\pi^-$ PID efficiency, $\varepsilon_{B^0 \to K^+\pi^-,\,\text{PID|trig.}}$, is determined as the average of the $B^0 \to K^+\pi^-$ and $\overline{B}^0 \to K^-\pi^+$ efficiencies. The kinematic distributions of the positively and negatively charged hadron p_+, η_+, p_-, η_- and nTracks in Equation 4.3 must be taken from a sample that is representative of $B^0 \to K^+\pi^-$ data.

Simulation cannot be used as a representative $B^0 \to K^+\pi^-$ sample from which to extract the distributions of $(p, \eta, nTracks)$, since the trigger requirements are

not well simulated; instead, an inclusive $B^0_{(s)} \to h^+h'^-$ data sample without PID requirements is used.

4.3.3 $B^0 \to K^+\pi^-$ PID efficiency: inclusive $B^0_{(s)} \to h^+h'^-$ sample fits

An inclusive $B^0_{(s)} \to h^+h'^-$ sample without PID requirements is used to obtain the background-subtracted $(p, \eta, nTracks)$ distributions from $B^0_{(s)} \to h^+h'^-$ data, which are required to determine the $B^0 \to K^+\pi^-$ PID efficiency $\varepsilon_{B^0 \to K^+\pi^-,\,\mathrm{PID|trig.}}$ in Equation 4.3. To subtract background, the *sPlot* method [72] is used (exactly as it is done by the PIDCalib software): signal and background are separated using a maximum likelihood fit to the invariant mass.

$B^0_{(s)} \to h^+h'^-$ fit model Some complications arise from using a $B^0_{(s)} \to h^+h'^-$ sample without PID requirements: the fit model is more involved and the sample contains more combinatorial background. The $B^0_{(s)} \to h^+h'^-$ sample includes candidates from all $B^0_{(s)} \to h^+h'^-$ decays, both $\Lambda^0_b \to ph^-$ decays, partially reconstructed background (such as $B^{+/0} \to K^+\pi^-\pi^{+/0}$) and combinatorial background. The $\pi^+\pi^-$ mass hypothesis is used, as it is the closest to the $\mu^+\mu^-$ hypothesis used in the offline selection. Under this hypothesis, most candidates are reconstructed with the wrong mass hypothesis. The mass window is chosen to be $[4.95, 5.8]$ GeV/c^2 to properly model partially reconstructed background on the left hand side and combinatorial background on the right hand side.

The shapes of all $B^0_{(s)} \to h^+h'^-$ and $\Lambda^0_b \to ph^-$ decay modes are modeled by convolving their true mass shape (a RooKeys shape, see Appendix A.3), obtained from simulation, with a common Gaussian to model resolution effects. Partially reconstructed background is modeled with an Argus function (see Appendix A.2), and is convolved with the same resolution model. Finally, the combinatorial background is modeled with an exponential. All shape parameters are left free in the fit.

The normalisation of the total $B^0_{(s)} \to h^+h'^-$ shape, of the total $\Lambda^0_b \to ph^-$ shape, of the partially reconstructed background and of the combinatorial background are left free in the fit. The contributions of the different $B^0_{(s)} \to h^+h'^-$ components and $\Lambda^0_b \to ph^-$ are fixed relative to each other. Assuming similar efficiencies, their relative contributions are given by their production and branching fractions. The branching fractions of all $B^0_{(s)} \to h^+h'^-$ decays have been measured at LHCb relative to $B^0 \to K^+\pi^-$; for B^0_s mesons the production rate relative to B^0 mesons, f_s/f_d,

Decay	Prod. fraction	Branching fraction	Fraction of $B^0 \to K^+\pi^-$
$B^0 \to K^+\pi^-$	1	$(1.96 \pm 0.05) \times 10^{-5}$	1
$B^0 \to \pi^+\pi^-$	1	$(5.1 \pm 0.4) \times 10^{-6}$	0.262 ± 0.019
$B^0 \to K^+K^-$	1	$(1.1 \pm 0.7) \times 10^{-7}$	0.0057 ± 0.0038
$B^0_s \to \pi^+\pi^-$	0.259 ± 0.015	$(9.9 \pm 2.6) \times 10^{-7}$	0.0131 ± 0.0033
$\bar{B}^0_s \to K^+\pi^-$	0.259 ± 0.015	$(5.6 \pm 0.7) \times 10^{-6}$	0.074 ± 0.008
$B^0_s \to K^+K^-$	0.259 ± 0.015	$(2.39 \pm 0.21) \times 10^{-5}$	0.316 ± 0.021
$\Lambda^0_b \to p\pi^-$	0.230 ± 0.052	$(3.5 \pm 1.1) \times 10^{-6}$	0.041 ± 0.016
$\Lambda^0_b \to pK^-$	0.230 ± 0.052	$(5.6 \pm 1.7) \times 10^{-6}$	0.066 ± 0.025

Table 4.7: The relative production fractions and branching fractions for each $B^0_{(s)} \to h^+h'^-$ decay, and the resulting total fraction of the $B^0 \to K^+\pi^-$ rate.

also measured by LHCb [53], is taken into account. The relative fraction of the two $\Lambda^0_b \to ph^-$ decays is fixed with measurements from CDF [71]. The resulting fractions of $B^0_{(s)} \to h^+h'^-$ and $\Lambda^0_b \to ph^-$ decays are shown in Table 4.7.

Treatment of the first BDT bin The large amount of background in the $B^0_{(s)} \to h^+h'^-$ sample causes the first BDT bin to be too contaminated with background for the *sPlot* method to work well. Therefore, it is split into two bins, 1a and 1b, with a boundary at 0.10: bin 1b contains only 1% of the combinatorial background, so the *sPlot* method can be applied; bin 1a needs special treatment, as the *sPlot* method cannot be applied and the $B^0 \to K^+\pi^-$ PID efficiency cannot be determined for this bin as it is for the others. The $B^0 \to K^+\pi^-$ PID efficiency in bin 1a is computed using the distributions from reweighted $B^0 \to K^+\pi^-$ simulation. A systematic uncertainty is assigned to this PID efficiency by comparing to the PID efficiency computed with the distributions from the background-subtracted $B^0_{(s)} \to h^+h'^-$ sample. The yield in bin 1a is estimated subtracting the sum of the yields in bin 1b to 8 from the total $B^0_{(s)} \to h^+h'^-$ yield without BDT cut.

Results The fits per BDT bin and to the total sample for the 2016 dataset are shown in Figure 4.4. Using the fits performed to the $B^0_{(s)} \to h^+h'^-$ data sample, the background-subtracted $(p, \eta, nTracks)$ distributions are obtained. Together with the PIDCalib efficiencies (see Section 4.3.2), they are used as inputs to Equation 4.3 to measure the PID efficiency $\varepsilon_{B^0 \to K^+\pi^-,\,\mathrm{PID|trig.}}$.

The $B^0_{(s)} \to h^+h'^-$ yields from the same fits, shown in Table 4.8, are used to perform a consistency check with the PID-corrected $B^0 \to K^+\pi^-$ yields. From the relative rates in Table 4.7, it is determined that $\frac{\mathcal{B}(B^0 \to K^+\pi^-)}{\mathcal{B}(B^0_{(s)} \to h^+h'^-)} = 0.599 \pm 0.010$, which is used to compare inclusive $B^0_{(s)} \to h^+h'^-$ to exclusive $B^0 \to K^+\pi^-$ yields. The difference between the $B^0_{(s)} \to h^+h'^-$ yield in Table 4.8 and the PID-corrected $B^0 \to K^+\pi^-$ yield at $\kappa = 5$, corrected for the rate of $B^0 \to K^+\pi^-$ relative to $B^0_{(s)} \to h^+h'^-$, is assigned as a systematic uncertainty.

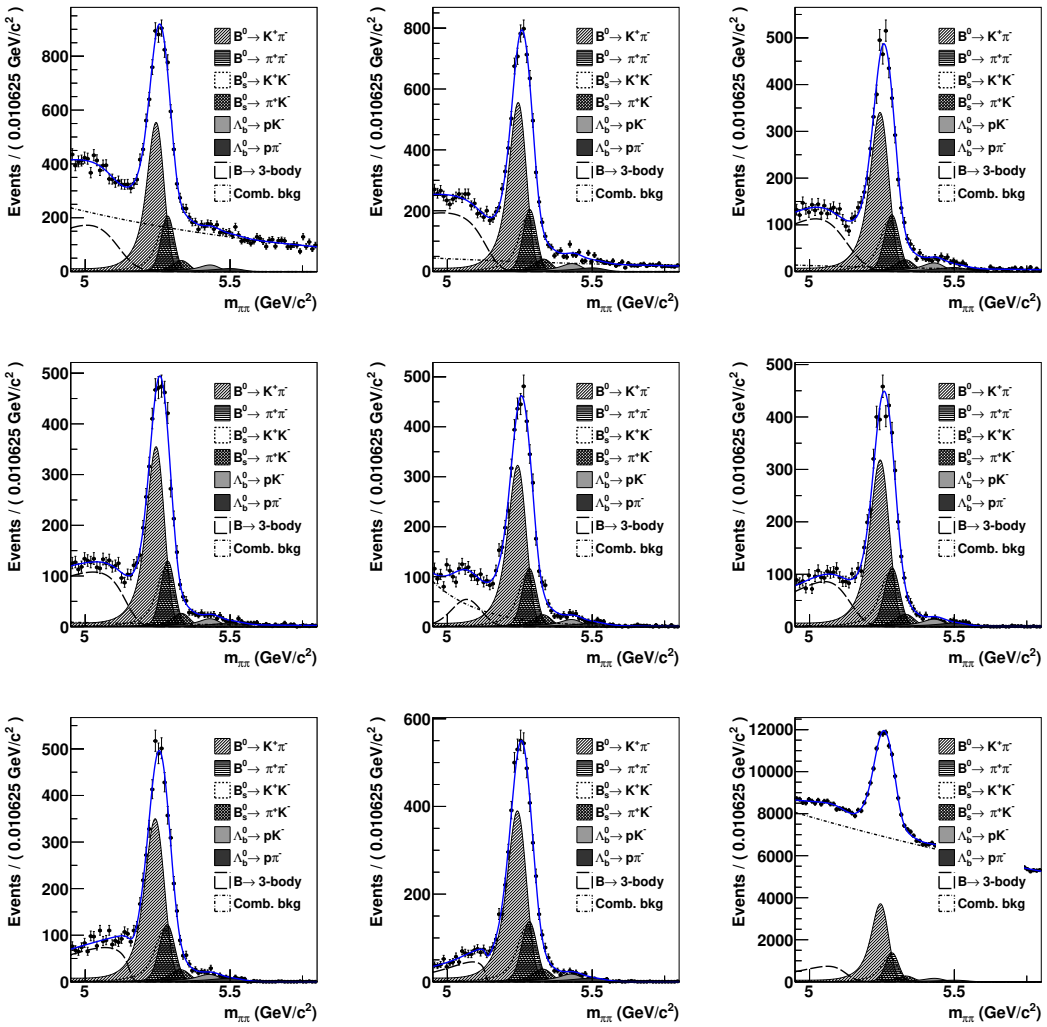

Figure 4.4: Invariant mass distributions for $B^0_{(s)} \to h^+h'^-$ candidates in the inclusive sample in 2016 data per BDT bin, increasing in bin number from left to right and top to bottom. Note that the first BDT bin here is taken to be from 0.10 to 0.25, in order to avoid the large amount of background at low BDT. The bottom right plot shows the result of the fit to the full sample. The fit result is superimposed on each plot.

BDT bin	2011	2012	2015	2016
[0.10, 0.25]	1718 ± 80	4824 ± 186	1921 ± 85	7384 ± 230
[0.25, 0.40]	1737 ± 68	4607 ± 189	2186 ± 81	7590 ± 155
[0.40, 0.50]	1144 ± 56	2997 ± 94	1383 ± 69	4906 ± 118
[0.50, 0.60]	1231 ± 47	3346 ± 81	1414 ± 64	4914 ± 97
[0.60, 0.70]	1261 ± 45	3222 ± 96	1497 ± 44	4547 ± 80
[0.70, 0.80]	1268 ± 53	3296 ± 83	1325 ± 47	4678 ± 89
[0.80, 0.90]	1531 ± 45	3874 ± 68	1604 ± 49	5227 ± 76
[0.90, 1.00]	1658 ± 41	4059 ± 69	1788 ± 44	5744 ± 76

Table 4.8: The $B^0_{(s)} \to h^+h'^-$ yield per BDT bin and per year from fit to the inclusive sample.

4.3.4 From $B^0 \to K^+\pi^-$ yields to PID-corrected fractions

The PID-corrected yields, $N_{B^0\to K^+\pi^-}/\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}}$, for $\kappa = 5$, per year and BDT bin, are shown in Table 4.9.

The fraction of $B^0 \to K^+\pi^-$ candidates per BDT bin i, r_i, for each κ is computed as the ratio of the PID-corrected yield in a bin and the total PID-corrected yield: $r_i(\kappa) = \frac{N_{B^0\to K^+\pi^-,\mathrm{i},\kappa}/\varepsilon_{B^0\to K^+\pi^-,\mathrm{PID|trig.,i},\kappa}}{N_{B^0\to K^+\pi^-,\mathrm{tot},\kappa}/\varepsilon_{B^0\to K^+\pi^-,\mathrm{PID|trig.,tot}\kappa}}$. To take into account the information of all r_i measurements at different values of κ, a weighted average $\overline{r_i}$ of all PID-corrected yields is calculated:

$$\overline{r_i} = \frac{\Sigma_{a,b}\rho^{-1}_{i(a,b)} r_{i(a)}}{\Sigma_{a,b}\rho^{-1}_{i(a,b)}} \tag{4.4}$$

The covariance matrix $\rho_i(a,b)$, containing the correlations between samples with different κ, is determined from the relative number of $B^0 \to K^+\pi^-$ candidates for different PID cuts a, b: $\rho_{i,a,b} = \sigma_{r_{i(a)}} \cdot \sigma_{r_{i(b)}} \cdot \sqrt{\frac{n_{i;b}}{n_{i;a}}}$.

To check this procedure, a linear function is fitted to the measurements of r_i as a function of κ. The PID-corrected $B^0 \to K^+\pi^-$ fractions r_i and the linear fits to them as function of κ are shown in Figure 4.5. From the fit, the fraction at $\kappa = 5$, r_{fit}, is determined. If it differs significantly from the weighted average, namely if $|r_{\mathrm{fit}} - \overline{r}| > \sqrt{\sigma^2_{\mathrm{fit}} + \sigma^2_{\overline{r}}}$, the difference is included as a systematic uncertainty. No significant variation of the fractions as function of κ was observed.

BDT bin	2011	2012	2015	2016
[0.00, 0.10]	390 ± 144	761 ± 233	966 ± 242	3040 ± 416
[0.10, 0.25]	860 ± 52	2498 ± 117	1072 ± 53	3885 ± 104
[0.25, 0.40]	940 ± 49	2523 ± 83	1169 ± 54	4136 ± 103
[0.40, 0.50]	639 ± 47	1613 ± 64	872 ± 50	2566 ± 87
[0.50, 0.60]	700 ± 38	2024 ± 75	769 ± 49	2795 ± 82
[0.60, 0.70]	688 ± 39	1852 ± 63	852 ± 45	2833 ± 81
[0.70, 0.80]	753 ± 42	2079 ± 73	845 ± 45	2609 ± 95
[0.80, 0.90]	848 ± 47	2240 ± 78	965 ± 50	3020 ± 89
[0.90, 1.00]	1015 ± 53	2422 ± 84	1042 ± 51	3527 ± 96

Table 4.9: The PID efficiency-corrected $B^0 \to K^+\pi^-$ candidate yield per BDT bin from the exclusive fit for $\kappa = 5$.

4.3.5 From $B^0 \to K^+\pi^-$ to $B^0_{(s)} \to \mu^+\mu^-$ fractions

Once $f_{B^0\to K^+\pi^-,\,\mathrm{data}}/\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}}$ is known, three more quantities (beside $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$) are needed in Equation 4.1: the $B^0 \to K^+\pi^-$ trigger efficiency, $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{trig.}}$, the $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiency, $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{trig.}}$, and the PID efficiency for $B^0_{(s)} \to \mu^+\mu^-$, $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{PID|trig.}}$.

The trigger efficiencies per BDT bin are determined from simulated samples. The $B^0_s \to \mu^+\mu^-$ trigger efficiencies are high and the muon trigger lines are well simulated, so the efficiencies per BDT bin $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{trig.}}$ are taken from simulation. However, the $B^0 \to K^+\pi^-$ sample has been selected as TIS in L0 and HLT1, which is sensitive to differences between data and simulation. Therefore, the $B^0 \to K^+\pi^-$ simulated sample is reweighted using the ratio of sWeighted data and simulated samples of $B^+ \to J/\psi K^+$ TIS candidates (in bins of $p_B, \eta_B, nTracks$). The $B^0 \to K^+\pi^-$ trigger efficiency is then determined directly on simulation as $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{trig.}} = \varepsilon_{B^0\to K^+\pi^-,\,\mathrm{L0\&HLT1\ TIS}} \cdot \varepsilon_{B^0\to K^+\pi^-,\,\mathrm{HLT2\ TOS|L0\&HLT1\ TIS}}$. The systematic uncertainties on the trigger efficiencies are estimated separately for the TIS and the HLT2 requirements with the TISTOS method on $B^0 \to K^+\pi^-$ data, which is defined in Section 4.5.4. The $B^0 \to K^+\pi^-$ L0 and HLT1 TIS efficiency are found to be consistent with the efficiencies from reweighted $B^0 \to K^+\pi^-$ simulation. A difference was found between the $B^0 \to K^+\pi^-$ HLT2 trigger efficiency, $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{HLT2\ TOS|L0\&HLT1\ TIS}}$, determined from data and simulation; it is assigned as a systematic uncertainty.

The trigger efficiencies as function of BDT, $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{trig.}}$ and $\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{trig.}}$, are shown in Figure 4.6.

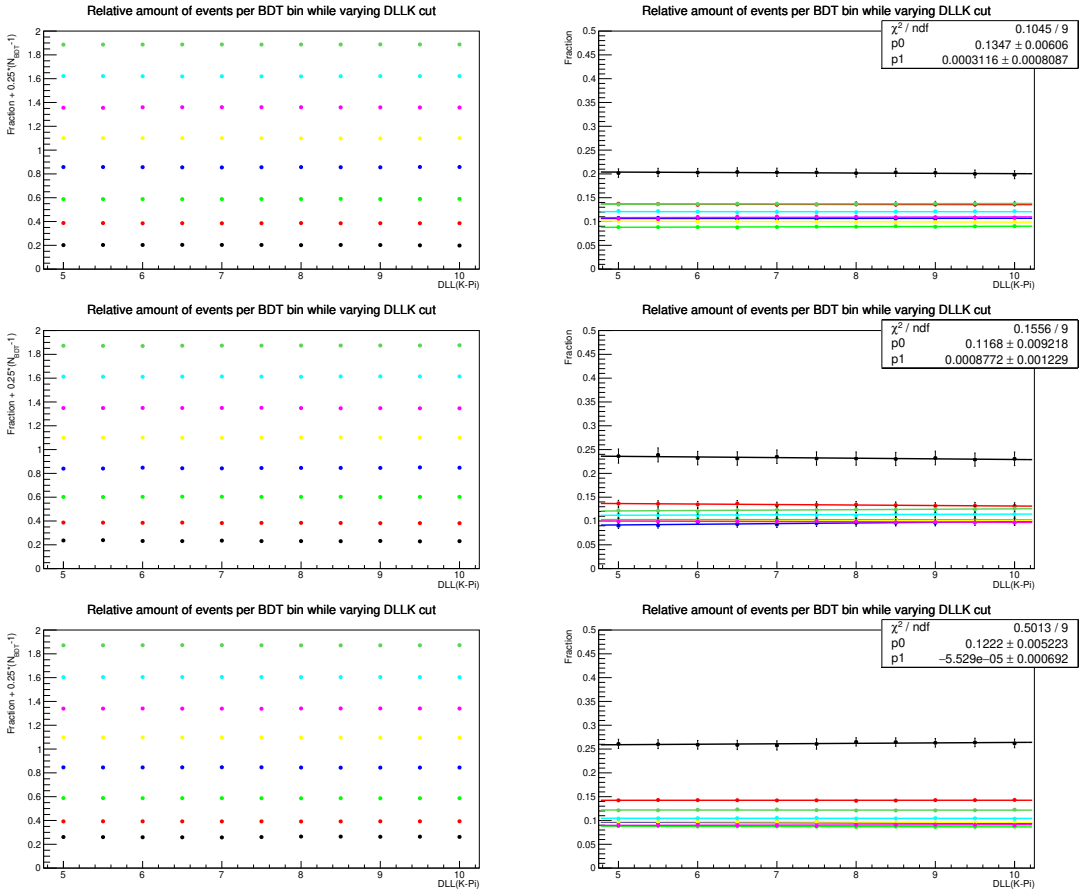

Figure 4.5: Fraction r_i of $B^0 \to K^+\pi^-$ signal events per BDT bin i corrected by the PID efficiency versus κ for Run 1 (top), 2015 (middle), and 2016 (bottom) data samples. Left: For better visualization the values are shifted by $0.25 \cdot (i-1)$; Right: a linear function is fitted to r_i to check if the fractions vary as function of PID.

As a check of the BDT calibration procedure, in Figure 4.7 the BDT distributions $f_{B^0\to K^+\pi^-,\,\mathrm{data}}/(\varepsilon_{B^0\to K^+\pi^-,\,\mathrm{PID|trig.}} \cdot \varepsilon_{B^0\to K^+\pi^-,\,\mathrm{trig.}})$ and $f_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{sim.}}$ are compared (without PID and trigger requirements). The distributions are consistent within their uncertainties.

The determination of the PID efficiency for $B^0_{(s)} \to \mu^+\mu^-$, $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{PID|trig.}}$, is discussed in the context of the normalisation in Section 4.5.5. The same method is applied per BDT bin to to determine the effect of the $B^0_{(s)} \to \mu^+\mu^-$ PID requirements on the BDT shape.

The final BDT fractions, $f_{B^0_{(s)}\to\mu^+\mu^-,\,\mathrm{data}}$, are calculated according to Equation 4.1 and are listed in Table 4.10.

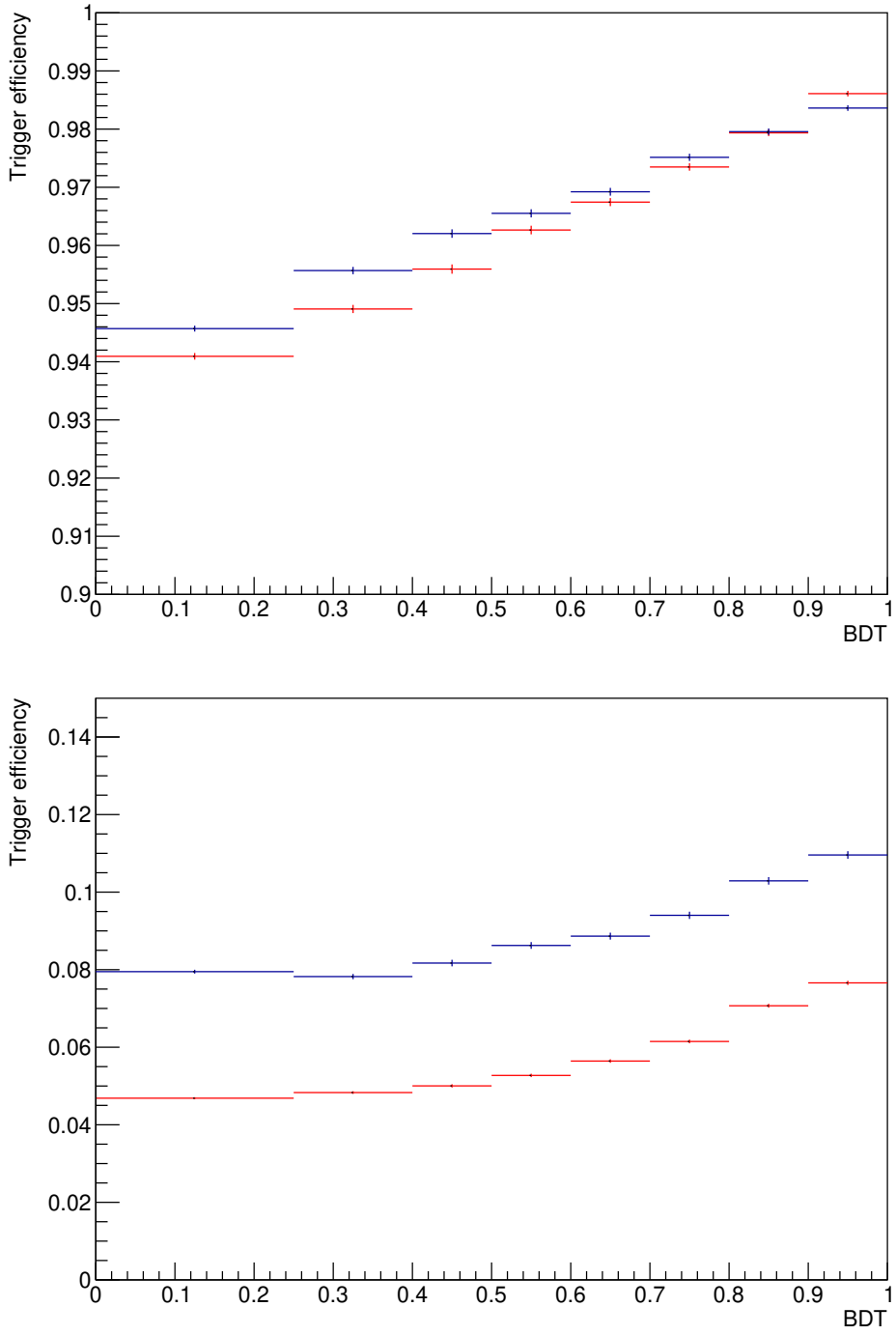

Figure 4.6: The total trigger efficiency as function of BDT extracted from 2012 (in red) and 2016 (in black) simulated samples of $B^0_s \to \mu^+\mu^-$ (top) and $B^0 \to K^+\pi^-$ (bottom) candidates.

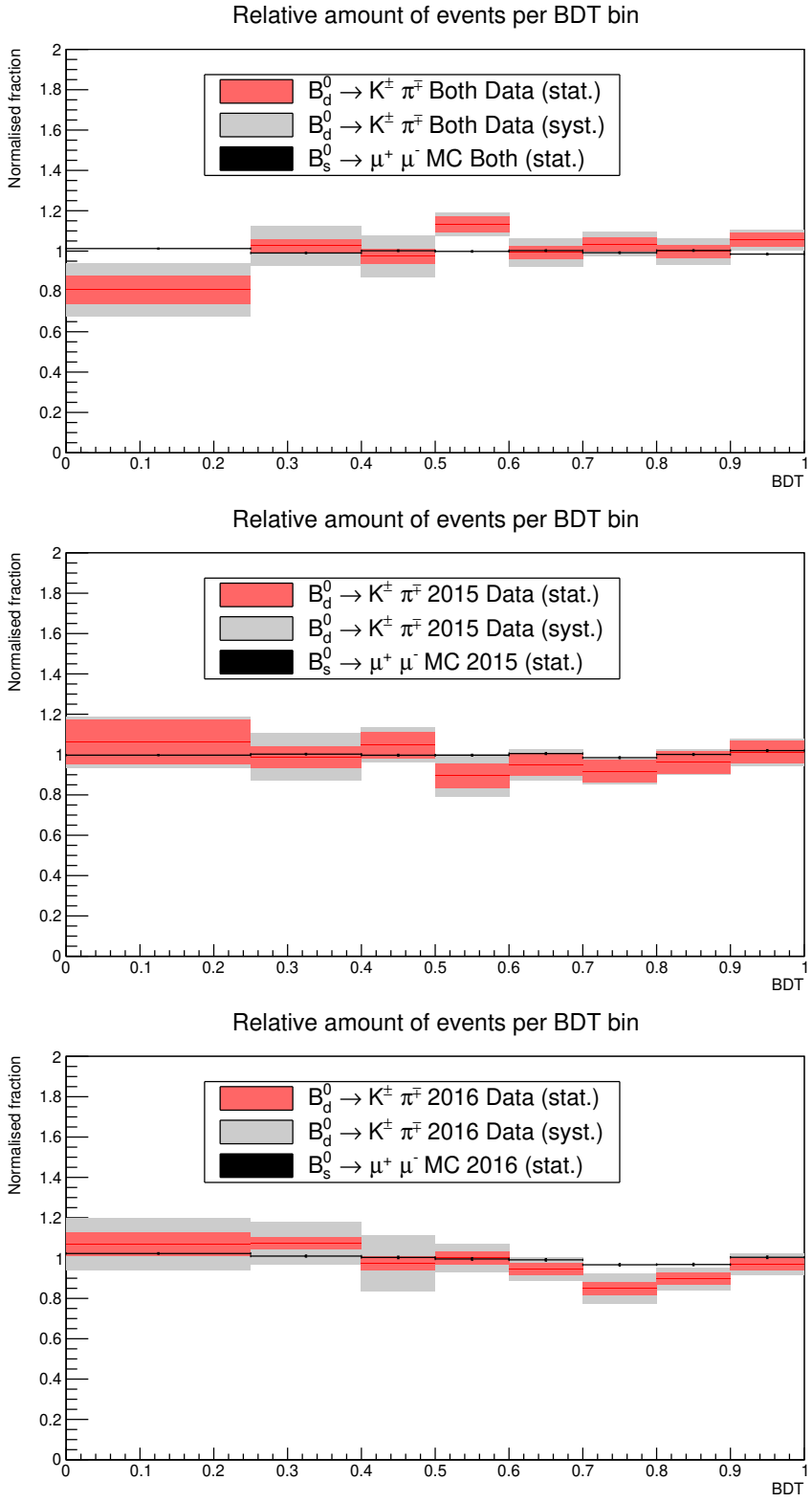

Figure 4.7: BDT signal fractions for Run 1 (top), 2015 (middle), and 2016 (bottom) data samples. The statistical uncertainty is included as a red band and systematic uncertainties, discussed in 4.3.7, as a grey band. For comparison, the $B^0 \to K^+\pi^-$ distribution (black points) from simulation is superimposed.

BDT bin	Run 1	2015	2016
$[0.00, 0.25]$	$0.1961 \pm 0.0171 \pm 0.0271$	$0.2573 \pm 0.0267 \pm 0.0157$	$0.2598 \pm 0.0142 \pm 0.0279$
$[0.25, 0.40]$	$0.1514 \pm 0.0048 \pm 0.0137$	$0.1457 \pm 0.0078 \pm 0.0155$	$0.1585 \pm 0.0045 \pm 0.0148$
$[0.40, 0.50]$	$0.0971 \pm 0.0037 \pm 0.0096$	$0.1042 \pm 0.0065 \pm 0.0056$	$0.0969 \pm 0.0035 \pm 0.0133$
$[0.50, 0.60]$	$0.1138 \pm 0.0039 \pm 0.0043$	$0.0898 \pm 0.0061 \pm 0.0087$	$0.1004 \pm 0.0033 \pm 0.0063$
$[0.60, 0.70]$	$0.1009 \pm 0.0033 \pm 0.0064$	$0.0961 \pm 0.0057 \pm 0.0053$	$0.0956 \pm 0.0030 \pm 0.0051$
$[0.70, 0.80]$	$0.1063 \pm 0.0036 \pm 0.0050$	$0.0935 \pm 0.0056 \pm 0.0029$	$0.0866 \pm 0.0033 \pm 0.0068$
$[0.80, 0.90]$	$0.1043 \pm 0.0034 \pm 0.0059$	$0.0996 \pm 0.0058 \pm 0.0030$	$0.0926 \pm 0.0030 \pm 0.0049$
$[0.90, 1.00]$	$0.1123 \pm 0.0036 \pm 0.0042$	$0.1057 \pm 0.0059 \pm 0.0040$	$0.1014 \pm 0.0031 \pm 0.0047$

Table 4.10: The BDT signal fractions, calculated according to Equation 4.1 for Run 1, 2015 and 2016 datasets, except for the correction applied for the $B^0_s \to \mu^+\mu^-$ BDT distribution. The first uncertainty is statistical, the second systematic.

4.3.6 Correcting the $B^0_s \to \mu^+\mu^-$ lifetime in simulation samples

To complete the discussion of Equation 4.1, the role of the lifetime correction $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$ is illustrated in this section. When generating $B^0_s \to \mu^+\mu^-$ simulation samples, the $B^0_s \to \mu^+\mu^-$ effective lifetime is needed, as all CP eigenstates must be averaged over to the compare to the measured branching fraction. However, in the standard simulated samples used for this analysis, the $B^0_s \to \mu^+\mu^-$ decay was simulated with the average B^0_s lifetime (as was known at that time, depending on the sample, ranging from 1.503 to 1.512 ps).

The relation between the average and the effective lifetime depends on the decay width asymmetry between the CP-odd and CP-even eigenstates y_s and the parameter $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$ (which is specific to the $B^0_s \to \mu^+\mu^-$ decay), because of the fact that due to neutral meson mixing, the lifetime of the B^0_s CP-odd and CP-even eigenstates is different ($y_s \neq 0$), as discussed in Section 2.4. The assumption made in the simulation corresponds to $y_s = 0$, which is in agreement with experimental data and theoretical predictions for the B^0 meson, but not quite the case for B^0_s mesons, $y_s = 0.066 \pm 0.004$. As a result, the yet unknown value of $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$ will play a role.

This is not only relevant to the comparison between theory and experiment through a measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime. As longer-lived $B^0_{(s)} \to \mu^+\mu^-$ candidates are more separated from the primary vertex and thus less background-like, the $B^0_s \to \mu^+\mu^-$ lifetime affects the selection efficiency and the BDT distribution. The effect on the efficiency per BDT bin was quantified by means of a correction factor, $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$. Three extreme assumptions were made: a CP-odd, CP-even, or average B^0_s lifetime, (1.411 ps, 1.511 ps and 1.599 ps respectively) such that the full range of B^0_s lifetimes is covered.

The correction factor is determined analytically by reweighting the simulation with weights given by

$$w_i = \frac{\tau_{\text{gen}}}{\tau_{\text{hypo}}} e^{-t_i(1/\tau_{\text{gen}} - 1/\tau_{\text{hypo}})}, \tag{4.5}$$

where t_i is the decay time of the B^0_s meson, $\tau_{\text{hypo}} = \{1.411, 1.511, 1.599\}$ ps and τ_{gen} is the lifetime used by the simulation software. These weights are calculated for $B^0_s \to \mu^+\mu^-$ simulated candidates that pass the full selection, and $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$ is

determined as

$$k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}} = \frac{1}{N}\sum_i^N w_i \tag{4.6}$$

where N is the number of events in BDT bin i.

The correction factor $k_{A^{\mu^+\mu^-}_{\Delta\Gamma_q}}$ is also determined without any requirement on the BDT to correct the $B^0_s \to \mu^+\mu^-$ selection efficiency for the normalisation. In the nominal ML fit to determine the branching fraction of $B^0_s \to \mu^+\mu^-$, the CP-odd hypothesis is used to determine the normalisation and BDT distribution. To test the effect of the variation in $B^0_s \to \mu^+\mu^-$ lifetime, the ML fit is repeated with the CP-averaged and CP-even hypotheses.

4.3.7 Systematic uncertainties

The BDT distribution for $B^0_{(s)} \to \mu^+\mu^-$, calibrated with $B^0_{(s)} \to h^+h'^-$ data, is subject to a number of systematic uncertainties:

1. **Consistency of yields from exclusive and inclusive approach**
   The PID-corrected $B^0 \to K^+\pi^-$ yield at $\kappa = 5$, scaled with the relative $B^0 \to K^+\pi^-$ branching fraction, is compared with the $B^0_{(s)} \to h^+h'^-$ yield. The difference is added as a systematic uncertainty per BDT bin (see Tables 4.8 and 4.9).

2. **Stability of the PID-corrected fractions as function of PID cut**
   It is investigated whether the PID-corrected fractions $\overline{r}_i$ (see Equation 4.4) are consistent as function of the PID cut κ (see Figure 4.5). Since they are, no systematic uncertainty is included.

3. **Binning used to determine PID efficiency**
   The PID efficiency of $B^0 \to K^+\pi^-$ is determined by binning the data and simulation samples, which means that the choice of binning introduces a systematic uncertainty. This uncertainty is estimated by repeating the BDT calibration with a different binning and assigning the difference for each BDT fraction as a systematic uncertainty (see Section 4.3.2).

4. **Fraction of signal events outside mass window**
   Depending on the fit, around 0.5% to 1.1% of the $B^0 \to K^+\pi^-$ signal candidates are rejected by the mass window for the $B^0 \to K^+\pi^-$ fit. This effect is not

corrected for in the nominal method. The BDT calibration is repeated while correcting for this effect and the difference with the nominal fractions is assigned as a systematic uncertainty (see Section 4.3.1).

5. **$B^0 \to K^+\pi^-$ fit model**
   To check the $B^0 \to K^+\pi^-$ fit model, the $B^0_{(s)} \to K^+\pi^-$ signal shapes are replaced by Double Gaussians (see Section 4.3.1). The difference with the nominal fraction per BDT bin is assigned as a systematic uncertainty.

6. **Statistical uncertainties on trigger and PID corrections**
   The statistical uncertainties from the sample size of the calibration samples and simulated sample are propagated to the trigger and PID efficiencies and considered as systematic uncertainties on the BDT calibration method and included as systematic uncertainties on the BDT distribution.

7. **HLT2 trigger efficiency from data and simulation**
   To estimate the systematic uncertainty from determining the trigger efficiencies from simulation, the BDT calibration is repeated with an estimate of the HLT2 trigger efficiency for $B^0 \to K^+\pi^-$ from data (see Section 4.3.5). The difference with the nominal fraction per BDT bin is assigned as a systematic uncertainty.

The dominating uncertainties for the BDT calibration originate from the the statistics of the $B^0 \to K^+\pi^-$ sample, the consistency of the yields from the exclusive and inclusive approach, and the effect of the binning on the PID efficiency of $B^0 \to K^+\pi^-$.

4.4 Mass calibration

The ML fit to determine the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions is performed in two dimensions, the dimuon invariant mass and the BDT. Therefore, the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass shapes have to be well determined. As the expected number of signal candidates in the $B^0_{(s)} \to \mu^+\mu^-$ analysis is small, it is not possible to determine the signal mass shapes in the $B^0_{(s)} \to \mu^+\mu^-$ ML fit. In addition, it is known that simulated samples do not reproduce well the momentum scale and resolution. Therefore, the $B^0_{(s)} \to \mu^+\mu^-$ mass shapes are calibrated with data from similar decays and the mass shape parameters are constrained in the $B^0_{(s)} \to \mu^+\mu^-$ ML fit to the values measured in this section.

Three effects determine the invariant mass shape of $B^0_{(s)} \to \mu^+\mu^-$ decays at LHCb: the true invariant mass shape and the detector bias and resolution. Concerning the first, the $B^0_{(s)}$ meson decay width is negligible (around $10^{-13}\,\text{GeV}/c^2$) and its shape is dominated by Final State Radiation (FSR), described in Section 2.6.2. The detector bias and resolution are dominated by the momentum scale and resolution, as the measured dimuon invariant mass is given by

$$\begin{aligned} m_{\mu^+\mu^-} &= \sqrt{m_\mu^2(2 + p_1/p_2 + p_2/p_1) + 2p_1p_2(1 - \cos\theta)} \\ &\sim \sqrt{2p_1p_2(1 - \cos\theta)}, \end{aligned} \tag{4.7}$$

where p_1, p_2 are the momenta of the two muons and θ is the angle between them in the detector frame, which has a negligible uncertainty. The relativistic approximation holds to a good level, as the momenta of muons at LHCb are far larger than their masses.

The invariant mass shapes of $B^0_{(s)} \to \mu^+\mu^-$ decays are described with a Crystal Ball function. This shape has a Gaussian core and an enhanced tail, given by the following formula [68]:

$$f(m|\mu,\sigma,\alpha,n) = N \begin{cases} \exp[-\frac{(m-\mu)^2}{2\sigma^2}], & \text{if } \frac{(m-\mu)}{\sigma} > -\alpha \\ (\frac{n}{|\alpha|})^n \exp[-\frac{\alpha^2}{2}](\frac{m-\mu}{\sigma} + \frac{n}{|\alpha|} - |\alpha|)^{-n}, & \text{otherwise} \end{cases} \tag{4.8}$$

Here, μ is the mean of the Gaussian, σ is the resolution of the Gaussian, α is the number of σ at which the exponential tail takes over, and n described how fast the exponential tail falls off. The role of the Gaussian core is to model the detector mass scale and resolution with μ and σ respectively, while the enhanced tail is included to describe the effect of FSR.

The Crystal Ball parameters calibrated separately. The detector bias is calibrated by measuring the mean of the $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ invariant mass distributions, as the bias at first order does not depend on the $B^0_{(s)}$ decay or selection. For the detector momentum resolution, since the trigger and PID selection for $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to \mu^+\mu^-$ decays are different, the resolution is calibrated with dimuon decays of quarkonium resonances at masses below and above the $B^0_{(s)}$ masses. To determine the $B^0_{(s)}$ resolutions, the measured resolutions are interpolated at the B^0 and B^0_s mass. As the tail parameters of the Crystal Ball function are dominated by FSR, which is a decay-dependent property, they are determined from $B^0_{(s)} \to \mu^+\mu^-$ simulation which is smeared to have a mass resolution similar to data.

4.4.1 Momentum scale calibration

The mean of the Crystal Ball function is determined by the momentum scale and is calibrated using the $B^0_{(s)} \to h^+h'^-$ sample, which contains both B^0 and B^0_s decays. Specifically, the B^0 and B^0_s mean are calibrated with $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ candidates, respectively. Both are the most abundant $B^0_{(s)} \to h^+h'^-$ decay for that $B^0_{(s)}$ meson.

The goal of the selection for the momentum scale calibration is to obtain as large a sample as possible with a negligible contribution from misidentified backgrounds, as they could bias the mean estimation. To keep a large sample, no trigger requirement is included in the selection. As PID selection, $\mathrm{DLL}_{K\pi} > \kappa$ and $\mathrm{DLL}_{K\pi} < -\kappa$ are used to select kaons and pions respectively at threshold κ. For the nominal fit, $\kappa = 10$, which is rejects any misidentified kaons or pions. In addition, $DLL_{ph} < 0$ is applied to each hadron to reject backgrounds from $\Lambda^0_b \to ph^-$ decays.

After the selection, the same fit model used for the BDT calibration is applied (see Section 4.3.1) with the difference that the $\Lambda^0_b \to ph^-$ shape is not included, because the proton PID selection rejects it. The fits to $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ data for Run 1 and Run 2 data are shown in Figures 4.8 and 4.9, respectively.

The selection that is expected to have the largest effect on the mean is the PID selection. To estimate the size of this effect, the mean is determined while varying κ from 0 to 20 in steps of 0.5, as shown in Figure 4.10. At small values of κ, a clear bias from mis-identified backgrounds is observed, which can be parametrised

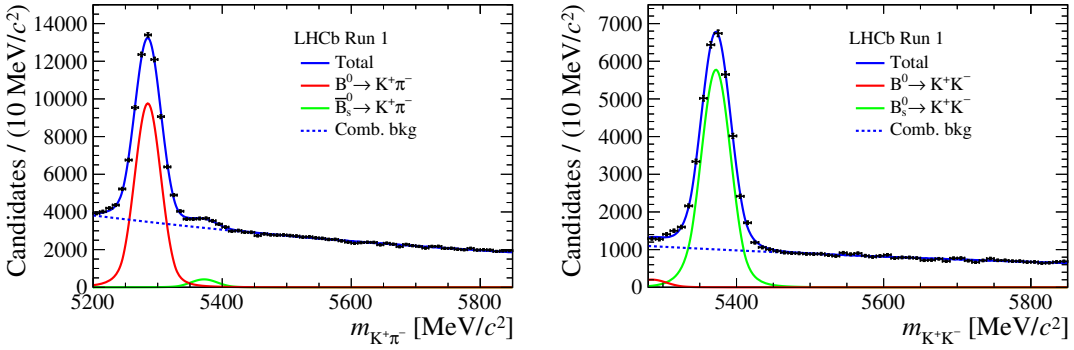

Figure 4.8: Invariant mass distributions for $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ candidates in Run 1 data. The fits are used to determine the mean of the invariant mass distribution for $B^0_{(s)} \to \mu^+\mu^-$. The red line indicates B^0 signal, the green B^0_s signal, and the blue dashed line the combinatorial background.

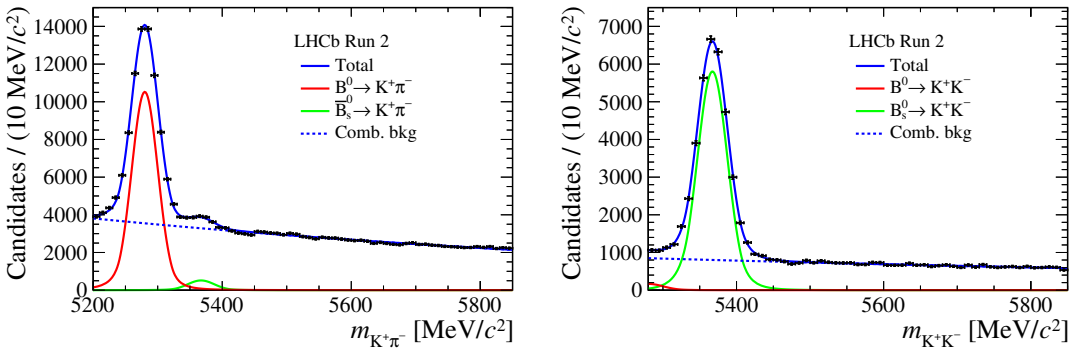

Figure 4.9: Invariant mass distributions for $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ candidates in Run 2 data. The fits are used to determine the mean of the invariant mass distribution for $B^0_{(s)} \to \mu^+\mu^-$. The red line indicates B^0 signal, the green B^0_s signal, and the blue dashed line the combinatorial background.

by fitting the following function:

$$\mu(\kappa) = \mu_0 + \frac{a_0}{1 + r \cdot \exp(\kappa/\kappa_0)}. \tag{4.9}$$

The systematic uncertainty from the PID selection is taken as the difference between the mean from the nominal fit, with $\kappa = 10$, and the mean at $\kappa = 20$. In addition, the fit is repeated with $DLL_{ph} > 2$ and -2, and the largest difference with the nominal result is taken as a systematic uncertainty. These two uncertainties are added in quadrature to obtain the total systematic uncertainty.

The means for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ in Run 1 and Run 2 are shown in Table 4.11. For comparison, the masses of the B^0 and B^0_s mesons have been measured

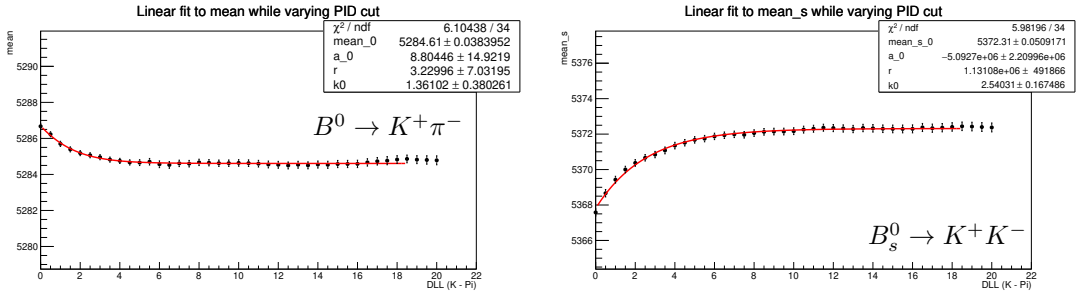

Figure 4.10: Dependence of the mean of the invariant mass distribution on the PID cut κ. The line denotes a parametrisatised fit based on Eq. 4.9 for illustrative purposes.

at 5279.63 ± 0.15 MeV/c^2 and 5366.89 ± 0.18 MeV/c^2 respectively [36]. In Run 1, the B^0 and B^0_s mean clearly differ from these measurements, by about 0.1%. In Run 2, the momentum scale calibration was improved and both means are consistent with the measured masses. The difference in momentum scale for Run 1 and Run 2 is one of the reasons that they are treated separately in the ML fit to determine the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions.

Dataset	B^0 mean (MeV/c^2)	B^0_s mean (MeV/c^2)
Run 1	$(5284.73 \pm 0.15_{\text{stat}} \pm 0.27_{\text{syst}})$	$(5372.05 \pm 0.16_{\text{stat}} \pm 0.36_{\text{syst}})$
Run 2	$(5279.95 \pm 0.13_{\text{stat}} \pm 0.08_{\text{syst}})$	$(5367.34 \pm 0.14_{\text{stat}} \pm 0.35_{\text{syst}})$

Table 4.11: Mean of the invariant dimuon mass distribution, measured with Run 1 and Run 2 data.

4.4.2 Momentum resolution calibration

The momentum resolution was calibrated with quarkonium resonances decaying to two muons, which have a similar selection to that of $B^0_{(s)} \to \mu^+\mu^-$ decays. , while $B^0_{(s)} \to h^+h'^-$ decays would have very different kinematics due to the different trigger and PID selection. The trigger and stripping selection of the quarkonium samples are described in Section 4.2. There are three distinct samples, each associated with specific resonances: J/ψ ,$\psi(2S)$ and $\Upsilon(NS)$(where $N \in \{1, 2, 3\}$).

The same fit model is used for all quarkonium samples. The J/ψ fit has two components: signal, described by a Double-Sided Crystal Ball function, and combinatorial background, described by an exponential. The tail parameters α and n are constrained to values obtained from simulation. For the $\psi(2S)$ fit, as there is more

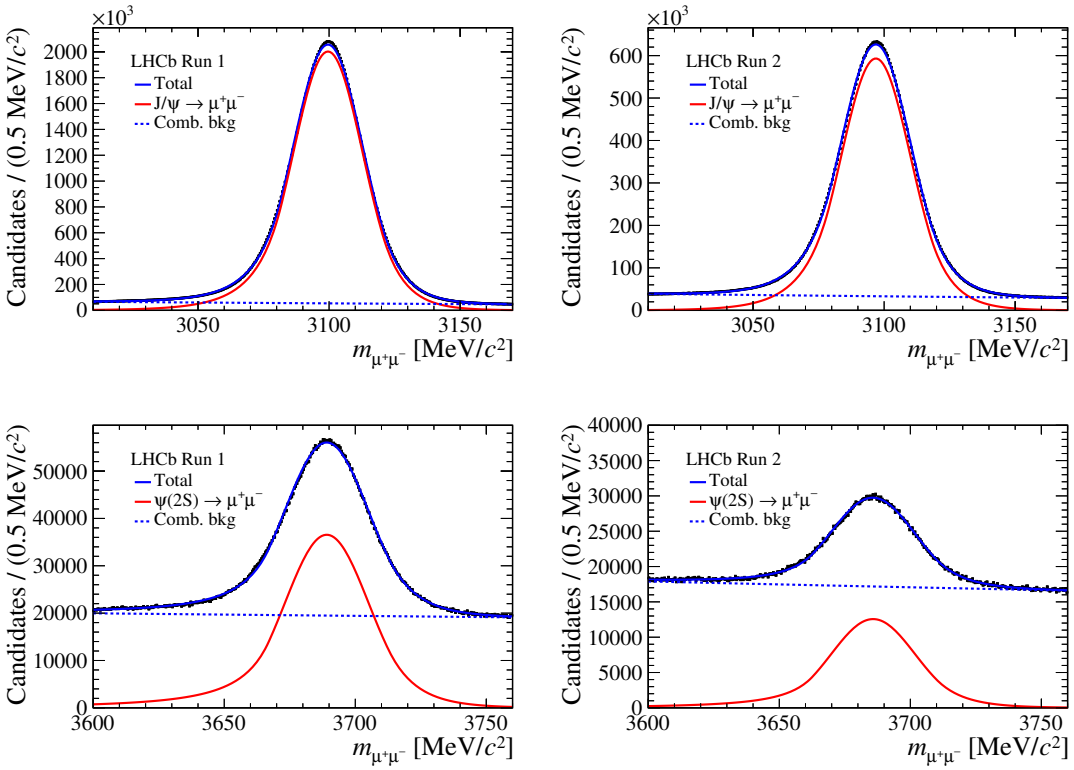

Figure 4.11: Fit to determine the invariant mass resolution for $J/\psi \to \mu^+\mu^-$ (top) and $\psi(2S) \to \mu^+\mu^-$ (bottom) on Run 1(left) and Run 2(right) data.

background in this sample, n has to be fixed to a value obtained from simulation. For the $\Upsilon(NS)$fit, a signal shape is included for each of the three resonances and n is fixed to the value from simulation. The fits are shown in Figures 4.11 and 4.12.

The resolutions determined at masses far above or below the $B^0_{(s)}$ masses have to be interpolated to the B^0 and B^0_s masses. In previous studies with Drell-Yan simulated samples, the dependence of the dimuon invariant mass resolution on the invariant mass was found to be a power law [73]:

$$\sigma_{\mu^+\mu^-} = a_0 + a_1 m^{\gamma}_{\mu^+\mu^-}. \tag{4.10}$$

Therefore, this power law is fitted to the five measured resolutions(leaving all parameters free), as shown in Figure 4.13.

A number of main systematic checks are performed for the momentum calibration. For the uncertainty from the selections, three cuts are varied; the minimum transverse momentum of both muons, from 800 to 1200 MeV/c in steps of 100 MeV/c; the maximum χ^2_{IP} of the dimuon resonance with respect to the primary vertex, from 8 to 12 in steps of 1; and the maximum track χ^2/ndf for each muon, from 3 to 7 in steps

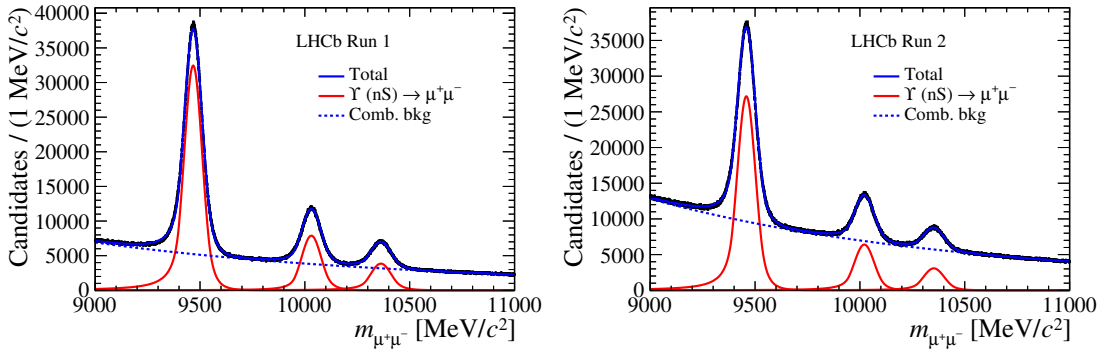

Figure 4.12: Fit to determine the invariant mass resolution for $\Upsilon(NS) \to \mu^+\mu^-$ on Run 1(left) and Run 2(right) data.

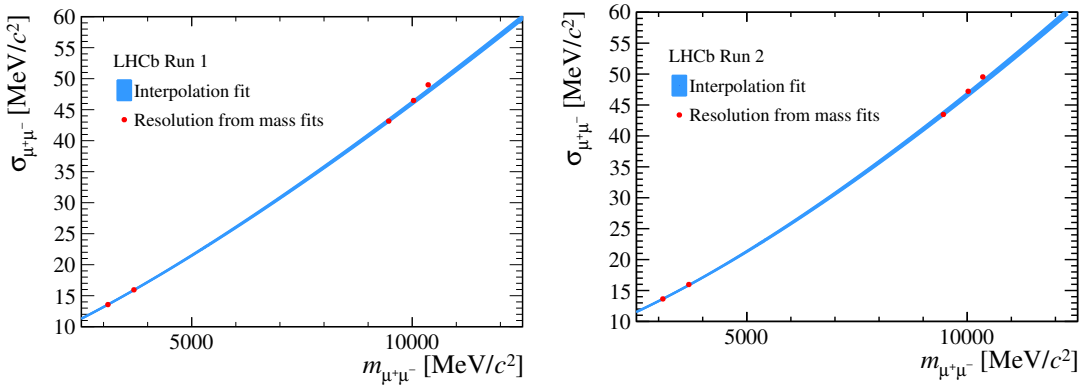

Figure 4.13: Power law fit to the invariant mass resolution of charmonium and bottomonium resonances for Run 1(left) and Run 2(right) data, used to determine the B^0 and B^0_s mass resolution.

of 1; The largest variation is taken as a systematic uncertainty. The fit model is varied in two ways to estimate a systematic uncertainty: first, fits where n was free, but constrained, fix it, and the other way around; second, the combinatorial shape is parametrised as a power law instead of an exponential. In both cases, the difference with the nominal result is included as a systematic uncertainty. The variation of the resolution in the region of ±60 MeV around the B mass is also taken as a systematic uncertainty.

The resolutions for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$, calibrated per Run, are shown in Table 4.12. The ratio of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ resolution is found to be compatible with the ratio of their means, and is used in the fit model for $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ decays.

Dataset	B^0 resolution (MeV/c^2)	B^0_s resolution (MeV/c^2)
Run 1	$(22.68 \pm 0.05_{\text{stat}} \pm 0.39_{\text{syst}})$	$(23.07 \pm 0.05_{\text{stat}} \pm 0.39_{\text{syst}})$
Run 2	$(22.46 \pm 0.08_{\text{stat}} \pm 0.41_{\text{syst}})$	$(22.85 \pm 0.08_{\text{stat}} \pm 0.42_{\text{syst}})$

Table 4.12: Invariant dimuon mass resolution measured with dimuon quarkonium decays per Run.

4.4.3 Tail parameter calibration

The tail parameters α and n are included in the $B^0_{(s)} \to \mu^+\mu^-$ mass shape (Eq. 4.8) to describe the FSR tail. As FSR is decay dependent, the tail parameters are obtained from $B^0_{(s)} \to \mu^+\mu^-$ simulated samples, without including the PID and trigger requirements.

Special care needs to be taken, since the tail parameters are correlated with the resolution of the Crystal Ball shape. The $B^0_{(s)} \to \mu^+\mu^-$ simulated samples are smeared to make them reproduce the resolution measured in data; first, the true mass distribution of $B^0_{(s)} \to \mu^+\mu^-$ decays is taken from the simulated samples; second, this distribution is smeared with a Gaussian, such that a fit with a Crystal Ball function reproduces the resolution measured in data. In practice a smearing with Gaussians with resolutions varying from 5 to 50 MeV/c^2 in steps of 5 is applied. The smeared distributions are fit with a Crystal Ball function with all parameters floating, and a linear fit is performed to parametrise the smearing width as function of the fitted resolution.

The tail parameters are not obtained from a single fit to the smeared simulation. To take into account the uncertainty on the resolution calibrated on data, toy studies are performed; the smearing is repeated 1000 times with a different smearing resolution, generated according to the measured resolution varied with its uncertainty. To each smeared sample, a Crystal Ball function is fit with all parameters left floating to obtain the tail parameters. Finally, the mean and uncertainty of each tail parameter are taken from the distribution of tail parameters from the toy experiments.

The tail parameters calibrated on simulated samples for Run 1 and Run 2 are shown in Table 4.13. All values are consistent with each other within the uncertainties.

Dataset	α_{B^0}	n_{B^0}	$\alpha_{B^0_s}$	$n_{B^0_s}$
Run 1	2.054 ± 0.013	1.141 ± 0.026	2.053 ± 0.007	1.156 ± 0.013
Run 2	2.063 ± 0.007	1.118 ± 0.014	2.062 ± 0.008	1.110 ± 0.017

Table 4.13: Tail parameters calibrated with Run 1 and Run 2 simulated samples.

4.5 Normalisation

In the previous sections, it was illustrated how the $B^0_{(s)} \to \mu^+\mu^-$ signal yield is distributed in the two fit variables, the invariant mass and the BDT. To determine the branching fractions, the fit needs an additional input, called a normalisation factor, denoted by $\alpha_{B^0_{(s)}\to\mu^+\mu^-}$; its determination is the subject of this Section.

In general, the branching fraction for a decay is defined as the fraction of decays to a specific final state:

$$\mathcal{B}(X \to YZ) = \frac{N_{X\to YZ}}{N_X}. \tag{4.11}$$

The experimental yield $N_{X\to YZ}$ is determined after reconstructing and selecting the signal, through a fit, and is corrected for the detection efficiency:

$$\mathcal{B}(X \to YZ) = \frac{N_{\text{fit},X\to YZ}}{N_X\ \varepsilon_{\text{det},X\to YZ}} \equiv \alpha_{X\to YZ}\ N_{\text{fit},X\to YZ}, \tag{4.12}$$

where $\alpha_{X\to YZ}$ is the normalisation factor, referred to as "the single-event sensitivity", as it gives the branching fraction corresponding to observing a single candidate.

There are two approaches to determine the single event sensitivity of $B^0_{(s)} \to \mu^+\mu^-$, $\alpha_{B^0_{(s)}\to\mu^+\mu^-}$. The direct approach is to determine the absolute number of $B^0_{(s)}$ mesons produced, $N_{B^0_{(s)}}$, from the measured cross section of $B^0_{(s)}$ mesons, $\sigma_{B^0_{(s)}}$, and the integrated luminosity $L = \int \mathcal{L}\ \mathrm{dt}$:

$$\alpha_{B^0_{(s)}\to\mu^+\mu^-} = \frac{1}{L\ \sigma_{B^0_{(s)}}\varepsilon_{\text{det},B^0_{(s)}\to\mu^+\mu^-}} \tag{4.13}$$

Alternatively, the branching fraction of $B^0_{(s)} \to \mu^+\mu^-$ can be measured relative to another $B^0_{(s)}$ decay, which is referred to as the normalisation channel:

$$\frac{\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)}{\mathcal{B}_{\text{norm}}} = \frac{N_{\text{fit},B^0_{(s)}\to\mu^+\mu^-}}{N_{\text{fit,norm}}} \frac{\varepsilon_{\text{det,norm}}}{\varepsilon_{\text{det},B^0_{(s)}\to\mu^+\mu^-}} \frac{N_{\text{norm}}}{N_{B^0_{(s)}}}. \tag{4.14}$$

If the experimental yields of $B^0_{(s)} \to \mu^+\mu^-$ and of the normalisation channel, $N_{\text{fit},B^0_{(s)}\to\mu^+\mu^-}$ and $N_{\text{fit,norm}}$, are extracted from the same data sample, then $N_{\text{norm}}/N_{B^0_{(s)}} = f_{\text{norm}}/f_{B^0_{(s)}}$, where f_{norm} and $f_{B^0_{(s)}}$ are the production fractions of

the normalisation and the signal B meson. As a consequence,

$$\alpha_{B^0_{(s)}\to\mu^+\mu^-} = \frac{\mathcal{B}_{\text{norm}}}{N_{\text{fit,norm}}} \frac{\varepsilon_{\text{det,norm}}}{\varepsilon_{\text{det},B^0_{(s)}\to\mu^+\mu^-}} \frac{f_{\text{norm}}}{f_{B^0_{(s)}}}. \tag{4.15}$$

The advantages of this method are that the uncertainty in the luminosity cancels out and that the uncertainties in the efficiencies and production fractions are greatly reduced by determining them for the normalisation and the signal mode relative to each other.

This Section describes the determination of the normalisation factor $\alpha_{B^0_{(s)}\to\mu^+\mu^-}$ through a normalisation channel. The normalisation channel should have a small uncertainty in the branching fraction and be as similar as possible to the signal decay, such that uncertainties cancel in the efficiency ratio. There is not a single decay that is exactly like $B^0_{(s)} \to \mu^+\mu^-$, so two complementary normalisation channels are used: $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$. Both have a well-determined branching fraction. $B^+ \to J/\psi K^+$ has a final state with two muons, just like $B^0_{(s)} \to \mu^+\mu^-$, so the trigger and PID selection are similar; however, because of the additional kaon and the intermediate resonance, the kinematics and reconstruction efficiencies are different. $B^0 \to K^+\pi^-$ is a two-body decay and has very similar kinematics to $B^0_{(s)} \to \mu^+\mu^-$. Because the final state particles are hadrons instead of muons, however, the trigger and PID requirements are different from $B^0_{(s)} \to \mu^+\mu^-$.

The yields $N_{\text{fit,norm}}$ are determined from maximum likelihood fits to the data. These fits and their results are discussed in Sections 4.5.1. The detection efficiency is separated into multiple terms, which are estimated independently:

$$\varepsilon_{\text{det}} = \varepsilon_{\text{acc.}} \times \varepsilon_{\text{rec.\&sel.|acc.}} \times \varepsilon_{\text{PID|rec.\&sel.,acc.}} \times \varepsilon_{\text{trig.|PID,rec.\&sel.,acc.}} \tag{4.16}$$

These terms in Equation 4.16 come from the acceptance of the detector (Section 4.5.2), the reconstruction and selection of the decay in the detector (Section 4.5.3), the trigger selection (Section 4.5.2), and finally the PID requirements that are imposed to $B^0 \to K^+\pi^-$ and $B^0_{(s)} \to \mu^+\mu^-$ candidates. The determination of the $B^0 \to K^+\pi^-$ PID efficiency has already been described in Section 4.3.2 and that of $B^0_s \to \mu^+\mu^-$ in Section 4.5.5. The trigger efficiencies, the track reconstruction efficiencies, and the PID efficiencies are corrected in a data-driven way, as they are not correctly reproduced in simulation. All other efficiencies are estimated from simulated samples. In Section 4.5.6, various cross-checks of the normalisation are discussed.

Assuming an equal production of B^+ and B^0 mesons, $f_{\rm norm}/f_{B^0} = 1$ and $f_{\rm norm}/f_{B^0_s} = f_d/f_s$. The value of f_s/f_d used in this analysis is taken from the LHCb measurement with 2011 data, $f_s/f_d = 0.259 \pm 0.015$ [53]. As 2012 and Run 2 data were taken at a different centre-of-mass energy, the variation of f_s/f_d with $\sqrt{s}$ is studied by measuring the efficiency-corrected yield ratio of $B^0_s \to J/\psi\,\phi$ and $B^+ \to J/\psi K^+$ in each year (Section 4.5.7).

Finally, the normalisation factors for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, $\alpha_{B^0_s \to \mu^+\mu^-}$ and $\alpha_{B^0 \to \mu^+\mu^-}$ are computed in Section 4.5.8.

4.5.1 Normalisation channel yields

The fit to determine the $B^+ \to J/\psi K^+$ candidate yield is performed constraining the dimuon mass to the J/ψ mass. This improves the mass resolution to about 10 MeV/c^2 and allows the fit to be performed in a small mass window from 5180 to 5380 MeV/c^2, such that the only backgrounds are combinatorial background and misidentified $B^+ \to J/\psi\pi^+$ decays.

The signal is modeled with a double-sided Hypatia function [74], which is a double-sided Crystal Ball function that takes into account the per-event mass error in the Gaussian core. The shape parameters α, n for each tail and λ for the core are constrained to values obtained from a fit to simulated samples including a Gaussian penalty term to the likelihood. As the $B^+ \to J/\psi K^+$ sample has no PID requirement on the kaon, the CKM-suppressed $B^+ \to J/\psi\pi^+$ decay contributes as a significant background. It is modeled with a non-parametric shape, called a RooKeys distribution, which uses one-dimensional kernel estimation to model arbitrary input datasets [75]. Its shape is thus fixed in the fit. Its relative contribution is constrained by using the measured $\mathcal{B}(B^+ \to J/\psi\pi^+)/\mathcal{B}(B^+ \to J/\psi K^+)$ branching fraction ratio [36] and their relative efficiencies. The combinatorial background is modeled with an exponential and is left free in the fit.

The fits to the Run 1 and Run 2 datasets are shown in Figure 4.14 and the yields obtained from the fit are listed in Table 4.14.

The $B^0 \to K^+\pi^-$ PID-corrected yields are obtained with the same procedure as described in Section 4.3.1. The fits to the $B^0 \to K^+\pi^-$ sample are shown in Figure 4.15 and the $B^0 \to K^+\pi^-$ candidate yields are given in Table 4.15. The systematic uncertainties in Table 4.15 originate mainly from the consistency check

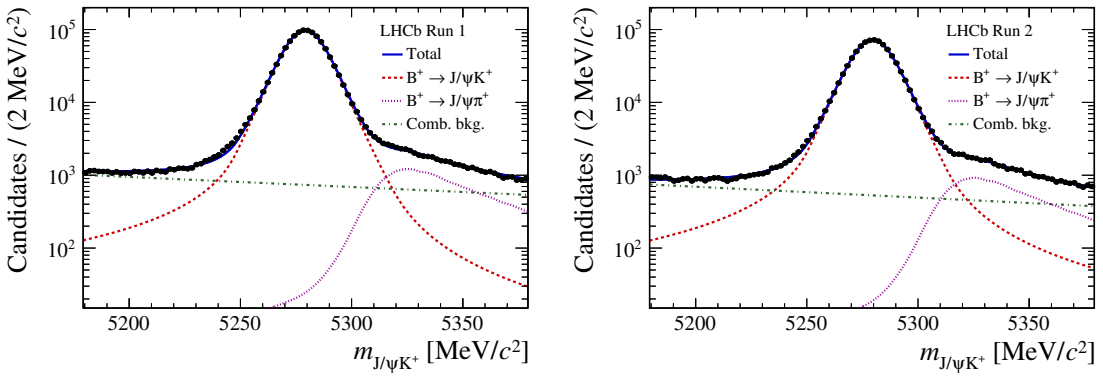

Figure 4.14: The fitted invariant mass distribution of the $B^+ \to J/\psi K^+$ candidates in Run 1 (left), Run 2 (right). The $B^+ \to J/\psi K^+$ component is denoted with a dashed red line, the mis-reconstructed $B^+ \to J/\psi \pi^+$ background component with a dashed magenta line, and the combinatorial background component with a dashed green line. The solid blue line denotes the sum of these components.

4

Table 4.14: The $B^+ \to J/\psi K^+$ yield per dataset as obtained from a maximum likelihood fit to the invariant mass distribution.

Dataset	Yield
2011	347312 ± 634
2012	774282 ± 930
2015	166689 ± 432
2016	684335 ± 878

with the inclusive $B^0_{(s)} \to h^+h'^-$ sample and from the variation of the PID efficiency with binning (see Section 4.3.1).

Table 4.15: The $B^0 \to K^+\pi^-$ yield per dataset, obtained from a maximum likelihood fit to the invariant mass distribution.

Dataset	Yield
2011	$6833 \pm 242\,(\text{stat}) \pm 616\,(\text{syst})$
2012	$18012 \pm 473\,(\text{stat}) \pm 1126\,(\text{syst})$
2015	$8552 \pm 305\,(\text{stat}) \pm 791\,(\text{syst})$
2016	$28411 \pm 549\,(\text{stat}) \pm 2488\,(\text{syst})$

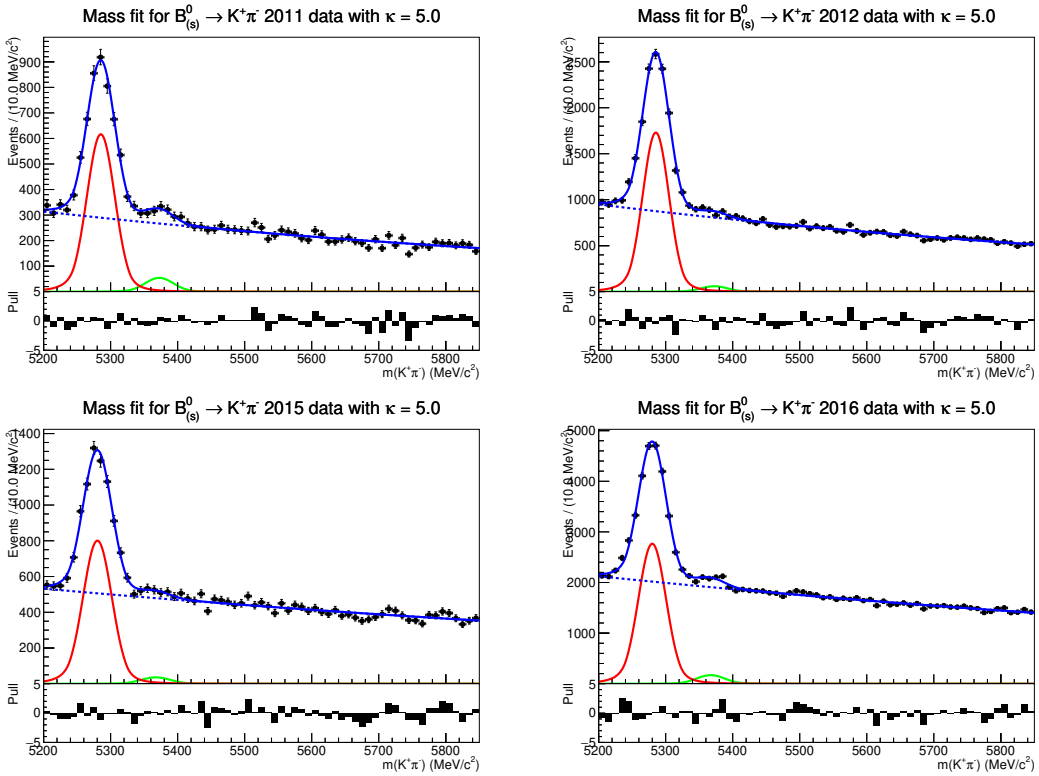

Figure 4.15: Invariant mass distribution of $B^0 \to K^+\pi^-$ candidates for 2011 (top left), 2012 (top right), 2015 (bottom left), and 2016 (bottom right) data samples. The yields are obtained from a binned maximum-likelihood fit. The red and green lines represent the $B^0 \to K^+\pi^-$ and $\bar{B}^0_s \to K^+\pi^-$ components, respectively, and the blue dashed line the combinatorial background. The sum of these components is denoted by the blue solid line.

4.5.2 Detector acceptance efficiency

In order to speed up the simulation, it is required that all final state particles from the simulated decay are contained in the acceptance of the detector, namely $\eta \in [1.62, 5.30]$, such that the detector response is only simulated for reconstructible decays. The efficiency of this requirement, $\varepsilon_{\text{acc.}}$, is taken from simulation, and shown per channel and year in Table 4.16.

The values of $\varepsilon_{\text{acc.}}$ are found to be similar for all modes with two final state particles. As $B^+ \to J/\psi K^+$ has an extra final state particle, its acceptance efficiency is slightly lower. As the centre-of-mass energy increased from 2011 to 2012 and from 2012 to Run 2, B mesons were produced more boosted and the acceptance increased.

Table 4.16: The geometrical detector acceptance efficiency, $\varepsilon_{\text{acc.}}$ determined from simulated samples.

Channel	2011	2012	2015	2016
$B^0_s \to \mu^+\mu^-$	$(17.45 \pm 0.04)\%$	$(18.67 \pm 0.02)\%$	$(19.29 \pm 0.04)\%$	$(19.33 \pm 0.05)\%$
$B^0 \to \mu^+\mu^-$	$(17.73 \pm 0.07)\%$	$(18.69 \pm 0.03)\%$	$(19.28 \pm 0.04)\%$	$(19.37 \pm 0.05)\%$
$B^+ \to J/\psi K^+$	$(15.46 \pm 0.03)\%$	$(16.66 \pm 0.05)\%$	$(17.28 \pm 0.04)\%$	$(17.34 \pm 0.05)\%$
$B^0 \to K^+\pi^-$	$(17.74 \pm 0.02)\%$	$(18.98 \pm 0.05)\%$	$(19.59 \pm 0.04)\%$	$(19.63 \pm 0.05)\%$

4.5.3 Reconstruction and selection efficiency

The efficiency of the reconstruction and selection, $\varepsilon_{\text{rec.\&sel.|acc.}}$, is taken from simulation, as the simulation reproduces the data well. Exceptions are the tracking efficiency and the isMuon efficiency (see Section 3.2.4), that need to be corrected for.

The ratio of tracking efficiencies in data and simulation is calibrated with $J/\psi \to \mu^+\mu^-$ candidates in bins of the p and η of the track [54]. The simulation is reweighted with a per candidate weight, calculated as the product of the track weights.

The efficiency of isMuon is calibrated on data with the PIDCalib package (similarly to what is done for $B^0 \to K^+\pi^-$ in Section 4.3.2). The ratio of the isMuon efficiency in data and simulation is used as a correction factor for the selection efficiency; at most, this correction factor is 1%.

The reconstruction and selection efficiencies are listed in Table 4.17. The relative efficiencies of different modes are found to be stable over the years.

Table 4.17: Reconstruction and selection efficiencies, $\varepsilon_{\text{rec.\&sel.|acc.}}$, for $B^0_{(s)} \to \mu^+\mu^-$ and the normalisation channels.

Channel	2011	2012	2015	2016
$B^0_s \to \mu^+\mu^-$	$(33.78 \pm 0.07)\%$	$(31.49 \pm 0.03)\%$	$(27.61 \pm 0.03)\%$	$(28.07 \pm 0.03)\%$
$B^0 \to \mu^+\mu^-$	$(33.62 \pm 0.07)\%$	$(30.95 \pm 0.07)\%$	$(27.09 \pm 0.03)\%$	$(27.54 \pm 0.03)\%$
$B^+ \to J/\psi K^+$	$(17.94 \pm 0.04)\%$	$(16.55 \pm 0.02)\%$	$(13.40 \pm 0.02)\%$	$(14.25 \pm 0.02)\%$
$B^0 \to K^+\pi^-$	$(23.29 \pm 0.05)\%$	$(21.81 \pm 0.01)\%$	$(18.65 \pm 0.02)\%$	$(19.15 \pm 0.01)\%$

4.5.4 Trigger efficiencies

No trigger line is explicitly required offline to select the $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ samples; therefore these samples contain events triggered by Trigger On Signal (TOS) and by Trigger Independent of Signal (TIS) requirements (see Section 3.4). As the events are actually predominantly triggered by muon and dimuon lines at every trigger level, the trigger efficiency for both modes, $\varepsilon_{\text{trig.}|\text{PID,rec.\&sel.,acc.}}$, can be calibrated with $B^+ \to J/\psi K^+$ data. In order to avoid biases from the signal hadron in the $B^+ \to J/\psi K^+$ decay when determining the $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiency, the $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiency is determined as from TOS efficiency for muon and dimuon triggers.

Because data only contains events passing the trigger, it is biased and an unbiased sample has to be constructed to measure the trigger efficiency from data. Assuming that events triggered by the signal hadron (TOS) do not influence the probability the event is triggered by something else in the event (TIS), a TIS requirement unbiases the sample for a TOS efficiency and vice versa. For the total trigger efficiency, which contains TOS and TIS candidates, the trigger is unbiased with respect to both (see Ref. [76]):

$$\varepsilon_{\text{trigger}} = \frac{N_{\text{trigger}}}{N_{\text{total}}} = \frac{N_{\text{trigger}}}{N_{\text{TIS}}}\frac{N_{\text{TIS}}}{N_{\text{total}}} = \frac{N_{\text{trigger}}}{N_{\text{TIS}}}\varepsilon_{\text{TIS}} = \frac{N_{\text{trigger}}}{N_{\text{TOS}}N_{\text{TIS}}}N_{\text{TISTOS}}, \tag{4.17}$$

where N_{trigger} is the number of events passing the trigger, N_{total} is the number of events in the sample without trigger selection, N_{TIS} is the number of events passing the TIS selection, N_{TOS} is the number of events passing the TOS selection, and N_{TISTOS} is the number of events passing the TIS and TOS selection.

As the kinematics of the B hadrons are correlated, the hypothesis of independence used in Equation 4.17 does not strictly hold. To mitigate this effect, the trigger efficiencies are determined in bins of kinematic variables. The total dimuon trigger efficiency is determined in bins of the IP and p_T of the muon with the maximum p_{T}.

Analogously to what is done for the PID efficiency (Section 4.3.2, the trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ are obtained by convoluting the binned efficiencies from $B^+ \to J/\psi K^+$ data with the kinematics of $B^0_{(s)} \to \mu^+\mu^-$ from the simulation, while the $B^+ \to J/\psi K^+$ efficiency is determined directly from the data.

The $B^+ \to J/\psi K^+$ efficiencies, determined with the TISTOS method, can also be used to determine the L0 and HLT1 trigger efficiency of $B^0 \to K^+\pi^-$. The events in the $B^0 \to K^+\pi^-$ sample are selected requiring the L0 and HLT1

trigger to be TIS; They are triggered mostly due to particles from the other b hadron; their efficiency $\varepsilon_{B^0\to K^+\pi^-,\,\text{L0\&HLT1 TIS}}$ can therefore be calibrated with any b hadron decay, and is calibrated using $B^+ \to J/\psi K^+$ data. The $B^0 \to K^+\pi^-$ HLT2 efficiency, $\varepsilon_{B^0\to K^+\pi^-,\,\text{HLT2 TOS|L0\&HLT1 TIS}}$ is determined analogously to as is described in Section 4.3.5. The trigger efficiency, $\varepsilon_{B^0\to K^+\pi^-,\,\text{trig.}}$, is determined as $\varepsilon_{B^0\to K^+\pi^-,\,\text{L0\&HLT1 TIS}} \times \varepsilon_{B^0\to K^+\pi^-,\,\text{HLT2 TOS|L0\&HLT1 TIS}}$.

The trigger efficiencies per year and mode are shown in Table 4.18. The trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ are stable within in Run 1 and Run 2. Their systematic uncertainties include the uncertainty of the TISTOS method and the fraction of $B^0_{(s)} \to \mu^+\mu^-$ events that are not TOS on muon or dimuon trigger lines. The $B^0 \to K^+\pi^-$ TIS efficiency $\varepsilon_{B^0\to K^+\pi^-,\,\text{L0\&HLT1 TIS}}$ varies a lot per year, as it depends on the exact configuration of the L0 and HLT1 trigger. The HLT2 efficiency improved in Run 2 as the trigger line was improved. Its systematic uncertainty is determined as the difference of the HLT2 efficiency in data and simulation.

Table 4.18: The trigger efficiencies, $\varepsilon_{\text{trig.|PID,rec.\&sel.,acc.}}$, per year and channel.

Channel	2011	2012	2015	2016
$B^0_s \to \mu^+\mu^-$	$(94.2 \pm 2.0)\%$	$(94.1 \pm 2.0)\%$	$(96.1 \pm 3.2)\%$	$(97.4 \pm 1.3)\%$
$B^0 \to \mu^+\mu^-$	$(94.1 \pm 2.2)\%$	$(93.9 \pm 2.1)\%$	$(95.9 \pm 3.3)\%$	$(97.1 \pm 1.5)\%$
$B^+ \to J/\psi K^+$	$(87.9 \pm 0.9)\%$	$(87.5 \pm 0.9)\%$	$(88.8 \pm 2.2)\%$	$(90.7 \pm 0.9)\%$
$B^0 \to K^+\pi^-$				
L0&HLT1 TIS	$(4.97 \pm 0.05)\%$	$(5.69 \pm 0.04)\%$	$(9.21 \pm 0.23)\%$	$(7.94 \pm 0.09)\%$
HLT2 TOS	$(72.31 \pm 0.38)\%$	$(75.49 \pm 0.10)\%$	$(88.85 \pm 0.11)\%$	$(91.20 \pm 0.07)\%$

4.5.5 $B^0_{(s)} \to \mu^+\mu^-$ PID efficiency

The efficiency of the $B^0_{(s)} \to \mu^+\mu^-$ PID requirements (Section 4.2.3), $\varepsilon_{\text{PID|rec.\&sel.,acc.}}$, is determined from data with the PIDCalib package, analogously to what is illustrated for the $B^0 \to K^+\pi^-$ PID efficiencies in Section 4.3.2. The muon PID efficiency is calibrated with detached $J/\psi \to \mu^+\mu^-$ decays produced in B hadron decays using a tag-and-probe method [77]. The tag-and-probe method relies on one well-identified or "tag" muon to identify a decay and reduce background, such that the "probe" muon can be used to determine the PID efficiency. To make sure that the muon trigger does not bias the PID estimates, events are required to be triggered by something else than the tag or probe muons, which is called "trigger unbiasing".

The efficiency is determined in 11 bins in p and 4 bins in p_{T} and convolved with the distributions of those variables from simulated $B^0_{(s)} \to \mu^+\mu^-$ samples. The $B^0_{(s)} \to \mu^+\mu^-$ PID efficiency is shown in Table 4.19.

The relative systematic uncertainty in the PID efficiency is estimated to be 2% and is dominated by the systematic uncertainty in the trigger unbiasing requirement. This effect is conservatively estimated by removing the requirement and taking the difference with the nominal result as a systematic uncertainty.

Table 4.19: The total PID efficiency, $\varepsilon_{\mathrm{PID|rec.\&sel.,acc.}}$, for $B^0_{(s)} \to \mu^+\mu^-$ decays in Run 1, 2015 and 2016 data.

Run 1	2015	2016
$(82.3 \pm 1.6)\%$	$(83.5 \pm 1.7)\%$	$(84.2 \pm 1.7)\%$

4.5.6 Normalisation cross-checks

As a check, the ratios of signal and normalisation channel efficiencies are computed separately for each term in Equation 4.16. Their stability over the datasets is checked. The biggest change is found to be the increase in TIS efficiency, $\varepsilon_{B^0 \to K^+\pi^-,\,\mathrm{L0\&HLT1\ TIS}}$, from Run 1 to Run 2, which is caused by the increase in trigger bandwidth. Except for this, no significant changes are observed.

Additionally, the efficiency corrected yields are computed for each decay and year. As expected, they are found to scale linearly with the centre-of-mass energy and the luminosity.

A stringent test of the normalisation is the comparison of the ratio of $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ branching fractions, with that measured at the B-factories with an uncertainty of 4.0% [36]. The ratios for each year are shown in Table 4.20, compared to the PDG value. They are found to be consistent for each year.

Table 4.20: The branching fraction ratio $\mathcal{B}(B^0 \to K^+\pi^-)/\mathcal{B}(B^+ \to J/\psi K^+)$ as found in the PDG and as determined from the efficiency corrected yields in this analysis.

PDG 2015	2011	2012	2015	2016
0.320 ± 0.013	0.333 ± 0.034	0.321 ± 0.024	0.359 ± 0.039	0.344 ± 0.033

4.5.7 Variation of f_s/f_d with centre-of-mass energy

At the time of this analysis, the ratio of production rates f_s/f_d had only been measured with 2011 data [53], at a different centre-of-mass energy than the 2012 and the Run 2 data. To investigate whether f_s/f_d varies with $\sqrt{s}$, the ratio of efficiency-corrected yields of $B^0_s \to J/\psi\,\phi$ and $B^+ \to J/\psi K^+$ is determined for 2011, 2012 and Run 2 data:

$$\frac{f_s/f_d(\text{X TeV})}{f_s/f_d(7\text{ TeV})} = \frac{\frac{N_{B^0_s\to J/\psi\,\phi,\text{X TeV}}}{\varepsilon_{B^0_s\to J/\psi\,\phi,\text{X TeV}}}}{\frac{N_{B^+\to J/\psi K^+,\text{X TeV}}}{\varepsilon_{B^+\to J/\psi K^+,\text{X TeV}}}} \Big/ \frac{\frac{N_{B^0_s\to J/\psi\,\phi,7\text{ TeV}}}{\varepsilon_{B^0_s\to J/\psi\,\phi,7\text{ TeV}}}}{\frac{N_{B^+\to J/\psi K^+,7\text{ TeV}}}{\varepsilon_{B^+\to J/\psi K^+,7\text{ TeV}}}} \tag{4.18}$$

The ratios of f_s/f_d at different $\sqrt{s}$, relative to 2011, are measured interally to be $\frac{f_s/f_d(8\text{ TeV})}{f_s/f_d(7\text{ TeV})} = 0.98 \pm 0.03$ and $\frac{f_s/f_d(13\text{ TeV})}{f_s/f_d(7\text{ TeV})} = 1.05 \pm 0.06$. The uncertainty in this estimate is dominated by the uncertainty in the tracking efficiency correction (see Section 4.5.3). A dedicated analysis has subsequently been performed and found evidence for the dependence of f_s/f_d on $\sqrt{s}$ [78].

4.5.8 Normalisation factors

The normalisation factors $\alpha_{B^0_{(s)}\to\mu^+\mu^-}$ are computed according to Eq. 4.15 and listed in Table 4.21. The $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ results have been averaged, taking into account the effect of the correlated uncertainties from the tracking efficiency correction (Section 4.5.3), f_s/f_d, and the $B^0_{(s)} \to \mu^+\mu^-$ signal efficiency.

The expected yields for $B^0_{(s)} \to \mu^+\mu^-$ in the $B^0_{(s)} \to \mu^+\mu^-$ data sample are thus estimated assuming either the Standard Model branching fractions or branching fractions measured by CMS and LHCb with Run 1 data [41]. Interestingly, the expected number of events differs significantly between the two scenarios (Standard Model or CMS+LHCb measurement), showing the additional sensitivity in this analysis.

Table 4.21: $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ normalisation factors and expected yields for each year. The expected signal yields are estimated assuming either the Standard Model branching fractions or the CMS+LHCb result with Run 1 data [41].

$B^0_s \to \mu^+\mu^-$	$\alpha_s^{\text{comb.}} \times 10^{10}$	$N^{\text{SM}}_{\text{expected}}$	$N^{\text{CMS+LHCb}}_{\text{expected}}$
2011	3.49 ± 0.25	10.5 ± 1.0	7.7 ± 2.0
2012	1.55 ± 0.11	23.6 ± 2.3	17.4 ± 4.6
2015	6.30 ± 0.53	5.8 ± 0.6	4.3 ± 1.2
2016	1.66 ± 0.13	22.0 ± 2.2	16.2 ± 4.4
Run 1	1.07 ± 0.07	34.2 ± 3.2	25.2 ± 6.8
Run 2	1.30 ± 0.10	28.0 ± 2.7	20.7 ± 5.6
Total:	0.59 ± 0.04	62.2 ± 5.6	45.9 ± 12.3
$B^0 \to \mu^+\mu^-$	$\alpha^{\text{comb.}} \times 10^{10}$	$N^{\text{SM}}_{\text{expected}}$	$N^{\text{CMS+LHCb}}_{\text{expected}}$
2011	9.21 ± 0.41	1.2 ± 0.1	4.2 ± 1.7
2012	4.20 ± 0.18	2.5 ± 0.2	9.3 ± 3.8
2015	17.00 ± 1.04	0.6 ± 0.1	2.3 ± 1.0
2016	4.48 ± 0.22	2.4 ± 0.2	8.7 ± 3.6
Run 1	2.88 ± 0.10	3.7 ± 0.3	13.6 ± 5.6
Run 2	3.52 ± 0.16	3.0 ± 0.3	11.1 ± 4.6
Total:	1.58 ± 0.04	6.7 ± 0.6	24.6 ± 10.1

4.6 Exclusive backgrounds

In the $B^0_{(s)} \to \mu^+\mu^-$ sample which was described in Section 4.2, some background candidates still remain. Most of the background in the sample arises from combinatorial background originating from the $b\bar{b} \to \mu^+\mu^- X$ process, which cannot be constrained from simulated samples. For this reason, the yield per BDT bin and the shape parameter of combinatorial background are included as unconstrained nuisance parameters in the maximum likelihood fit to obtain the branching fractions of $B^0_{(s)} \to \mu^+\mu^-$ decays.

The other backgrounds are exclusive decays, which are due to specific b hadron decays that mimic $B^0_{(s)} \to \mu^+\mu^-$. In this Section, the calibration of their yield and mass shape per BDT bin is discussed. In the maximum likelihood fit to obtain the results, the yields per BDT bin are constrained and the mass shapes fixed for each exclusive background. For this analysis, the following sources of exclusive background were taken into account:

1. $B^0_{(s)} \to h^+h'^-$ decays, where both hadrons are misidentified as muons. The dominant modes are $B^0 \to K^+\pi^-$, $B^0_s \to K^+K^-$, $B^0 \to \pi^+\pi^-$ and $\bar{B}^0_s \to K^+\pi^-$.

2. $H^0_b \to h^\pm\mu^\mp\nu_\mu$ decays, specifically $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$, $\Lambda^0_b \to p\mu^-\nu$. Here the neutrino is not reconstructed and the hadron is misidentified as a muon.

3. $B^+_c \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$ or $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ decays, where there are two real muons in the decay and there is a missing pion or muon in the final state.

Backgrounds that were considered, but not investigated further and assumed to be negligible are:

- $B^0_{(s)} \to p\overline{p}$, due to the small branching fraction of $\mathcal{O}(10^{-8})$ compared to $B^0_{(s)} \to h^+h'^-$.

- $\Lambda^0_b \to ph^-$, as it is produced less than B^0 or B^0_s mesons, the $p \to \mu$ misID rate is smaller, and it is not observed in the $B^0_{(s)} \to h^-\mu^+\nu_\mu$ calibration sample.

- $B^0_{(s)} \to h^+h'^-h''^0$, which is negligible compared to $B^0_{(s)} \to h^+h'^-$ as the missing particle implies that only a small fraction of candidates will be in the mass window.

- $B \to X_s\mu^+\mu^-$, where X_s is any strange hadron, as these candidates do not enter into the dimuon mass window after subtracting the mass of the lightest strange hadron, namely the kaon.
- $B^0_{(s)} \to \pi^+\pi^-\mu^+\mu^-$, which was checked to be negligible in the dimuon mass window due to the extra missing pion compared to $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ while having a similar branching fraction.
- $B^0_{(s)} \to \mu^+\mu^-\gamma$ from Initial State Radiation (ISR), which was discussed in Section 2.6.2. As described there, photons from Final State Radiation contribute to the $B^0_{(s)} \to \mu^+\mu^-$ signal as a radiative tail and are included in the signal simulation. The ISR component is separate as it has a distinctive mass shape, as illustrated by Figure 2.7. It is expected to contribute $\mathcal{O}(0.3)$ candidates in the most sensitive BDT bins (BDT>0.5), and is neglected.
- $B^+ \to \mu^+\mu^-\mu^+\overline{\nu}_\mu$ and $B^0_{(s)} \to \mu^+\mu^-\nu_\mu\overline{\nu}_\mu$, which have not yet been observed and are predicted to have a negligible branching fraction. Furthermore, they have two missing final state particles.

The mass shapes for each background are based on simulation, with specific corrections to the simulation for some background samples. For example, the $B^0_{(s)} \to h^+h'^-$ background contribution is dominated by decays of pions and kaons in flight, which are the most common cause of hadron misidentification as muon. In such a decay, the hadron track will have a kink from the momentum released in the decay, which means the momentum resolution is worse and the $B^0_{(s)} \to h^+h'^-$ mass shape is smeared accordingly. The mass shape calibration for each background is described in Section 4.6.1.

The yields of exclusive backgrounds are constrained in three different ways.

1. Each background apart from $B^0_{(s)} \to h^+h'^-$ is normalised relative to $B^+ \to J/\psi K^+$. This is the same approach as taken for the normalisation for the $B^0_{(s)} \to \mu^+\mu^-$ signal and is explained in more detail in Section 4.5. As the branching fraction is known for each background, the total yield for each background in the $B^0_{(s)} \to \mu^+\mu^-$ fit is constrained. The BDT shape of each background is taken from simulation. This approach is discussed in Section 4.6.2.
2. $B^0_{(s)} \to h^+h'^-$ is normalised using the most prominent decay, $B^0 \to K^+\pi^-$, as its yield in BDT bins is determined for the BDT calibration, described in Section 4.3. The $B^0_{(s)} \to h^+h'^-$ background estimation that is based on the BDT calibration yields is discussed in Section 4.6.3.

3. The $B^0_{(s)} \to h^-\mu^+\nu_\mu$ data sample serves as a validation region, which allows to check the yields per BDT bin of the $B^0_{(s)} \to h^-\mu^+\nu_\mu$ and $B^0_{(s)} \to h^+h'^-$ backgrounds from the first and second method respectively. The $B^0_{(s)} \to h^-\mu^+\nu_\mu$ data sample check is discussed in Section 4.6.4.

4.6.1 Background mass shape calibration

For each exclusive background, the mass shape is obtained from simulated samples after applying the full selection procedure. As the invariant mass shape for backgrounds depends on the BDT, it is determined in BDT bins. Almost every exclusive background misses one or more particles in the B reconstruction (apart from $B^0_{(s)} \to h^+h'^-$), which means that their mass shapes typically look like a shoulder on the left hand side of the mass window. Their mass distributions after the full selection are fitted with an Argus function (see Section A), which describes the true $h\mu$ mass shape, convolved with a Gaussian, which represents the experimental resolution.

$B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ The mass shapes for these decays are calibrated with the same procedure. For semileptonic decays, the effects of QCD on the hadronic $B^0_{(s)} \to h$ transition are described by so-called form factors, which depend on the momentum transfer in the decay $q = m(\mu\overline{\nu}_\mu)$. These form factors are calculated with non-perturbative QCD, and are checked with measurements at B-factories [79], [80]. Those measurements actually do not precisely agree with the form factor parametrisation used in the simulated samples, and the simulated samples are reweighted in bins of q^2 to correct for the difference. Because both mass shapes are very similar, $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ are included as one shape in the fit. The fits performed to obtain the $B^0 \to \pi^-\mu^+\nu_\mu$ mass distributions are shown in Figure 4.16.

$\Lambda^0_b \to p\mu^-\nu$ For $\Lambda^0_b \to p\mu^-\nu$, the relative production rate of Λ^0_b and B^+ mesons was found to depend on the transverse momentum of the Λ^0_b candidate [81],

$$f_{\Lambda^0_b}/f_u = (0.808 \pm 0.220) \times [1 - (0.031 \pm 0.005) \times p_{\mathrm{T}}(\,\mathrm{GeV}/c)], \tag{4.19}$$

and a per-candidate weight of $f_{\Lambda^0_b}/f_u$ is included to take this dependence into account. The fits performed to obtain the $\Lambda^0_b \to p\mu^-\nu$ mass distributions are shown in Figure 4.17.

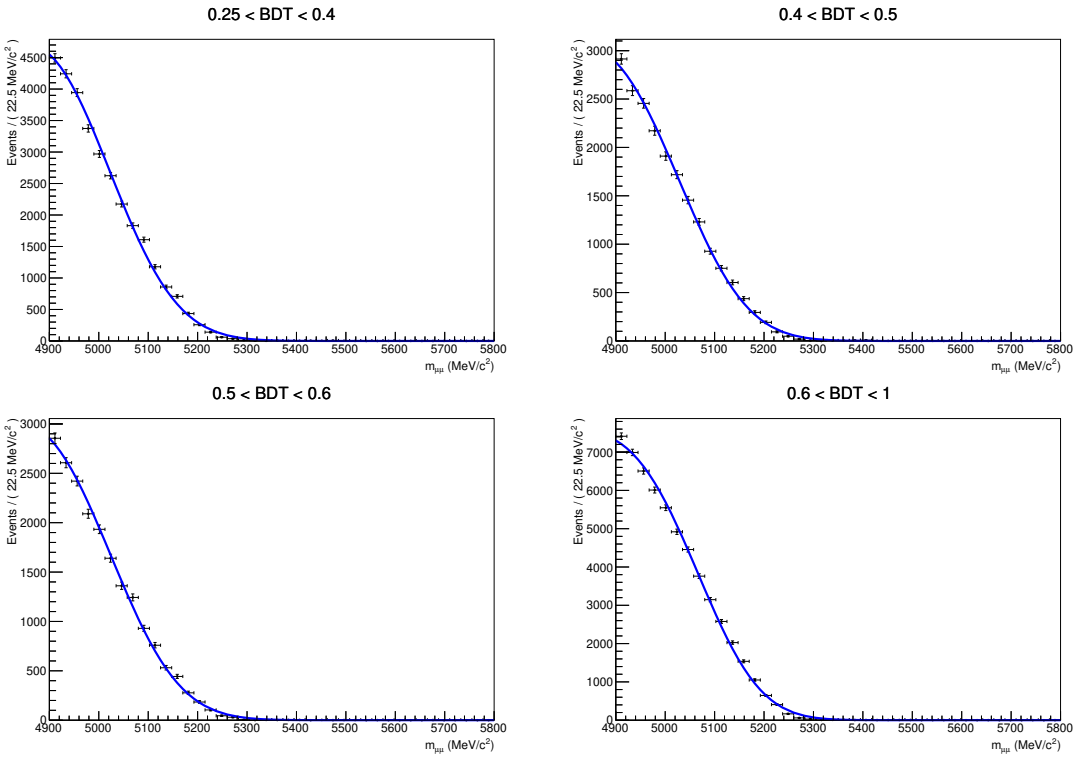

Figure 4.16: The mass shape for the $B^0_{(s)} \to h^-\mu^+\nu_\mu$ background in BDT bins, taken from simulation.

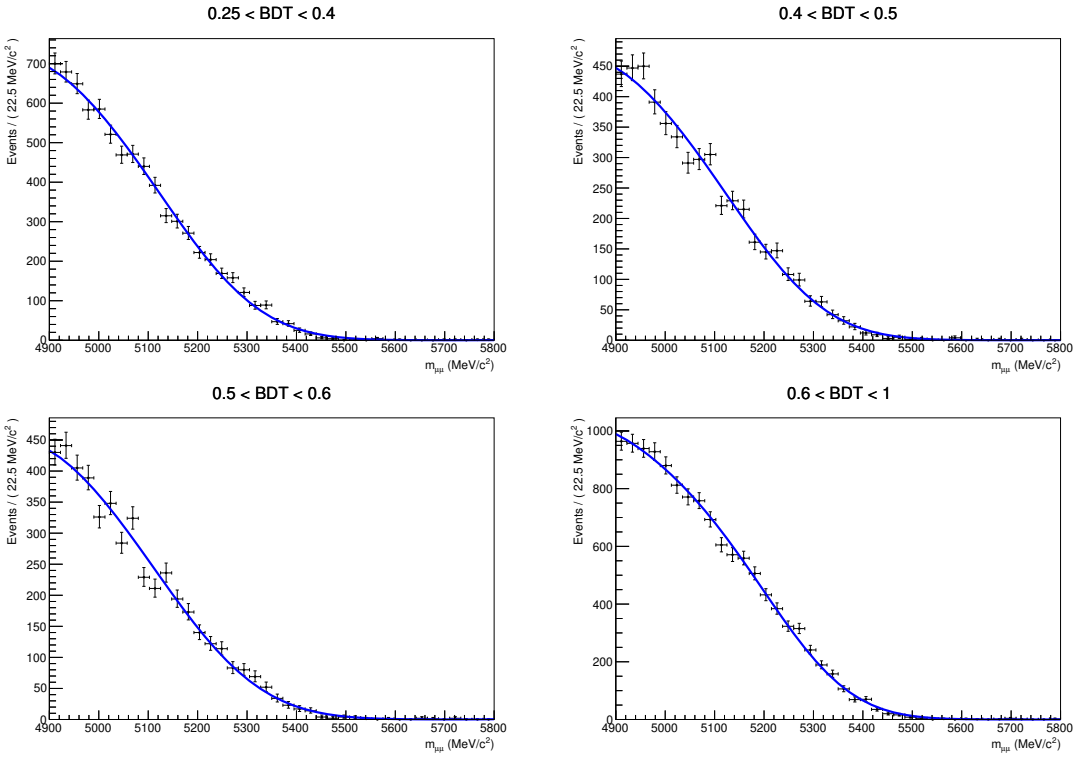

Figure 4.17: The mass shape for $\Lambda^0_b \to p\mu^-\nu$ background in BDT bins, taken from simulation.

$\boldsymbol{B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-}$ The $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ backgrounds both have a missing pion and two reconstructed muons. As a result, their mass shapes are very similar and a single mass shape is determined from their combined samples. The fits performed to obtain the $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ mass distributions are shown in Figure 4.18.

$\boldsymbol{B_c^+ \to J/\psi\mu^+\nu_\mu}$ For this background, one muon from the J/ψ and one from the B_c^+ decay are combined to form a dimuon candidate, and the other muon from this decay is part of the rest of the event. Due to the long track isolation in the BDT (see Section 4.2.6), which classifies candidates as more background-like if they are associated with a track from the rest of the event, most $B_c^+ \to J/\psi\mu^+\nu_\mu$ candidates are concentrated at low BDT values. At high BDT values, BDT>0.5, the statistics of the simulated sample are limited, so one fit is done to all candidates in that BDT region. The fits performed to obtain the $B_c^+ \to J/\psi\mu^+\nu_\mu$ mass distributions are shown in Figure 4.19.

$\boldsymbol{B^0_{(s)} \to h^+h'^-}$ For the $B^0_{(s)} \to h^+h'^-$ mass shape, it is important to realise that hadron to muon misidentification is dominated by decays in flight, which have a worse momentum resolution. This is taken into account by smearing the momentum of both hadrons in the simulated sample. As the double misidentification efficiency is small, applying the PID selection to the simulated sample would leave a tiny sample, and the simulation is weighted with the PID efficiency taken from data calibration samples. The $B^0_{(s)} \to h^+h'^-$ mass shape is found to be independent of the BDT bin and thus the same shape is used in each bin. It is non-parametric and is parametrised as a RooKeys distribution, which uses one-dimensional kernel estimation to model arbitrary input datasets [75]. The $B^0_{(s)} \to h^+h'^-$ mass shape used in the ML fit is shown in Figure 4.21.

4.6.2 Background normalisation with $B^+ \to J/\psi K^+$

For each exclusive background apart from $B^0_{(s)} \to h^+h'^-$, the expected yield is normalised with respect to $B^+ \to J/\psi K^+$, similar to the $B^0_{(s)} \to \mu^+\mu^-$ signal normalisation described in Section 4.5. The formula used to determine the number of background candidates per BDT bin i, $N_{\text{background},i}$ is

$$N_i^{\text{bkg}} = f_{\text{BDT},i}^{\text{bkg}}\ f_x/f_u\ \varepsilon_{\text{tot}}^{\text{bkg}}\ \mathcal{B}^{\text{bkg}}\ \beta_u \tag{4.20}$$

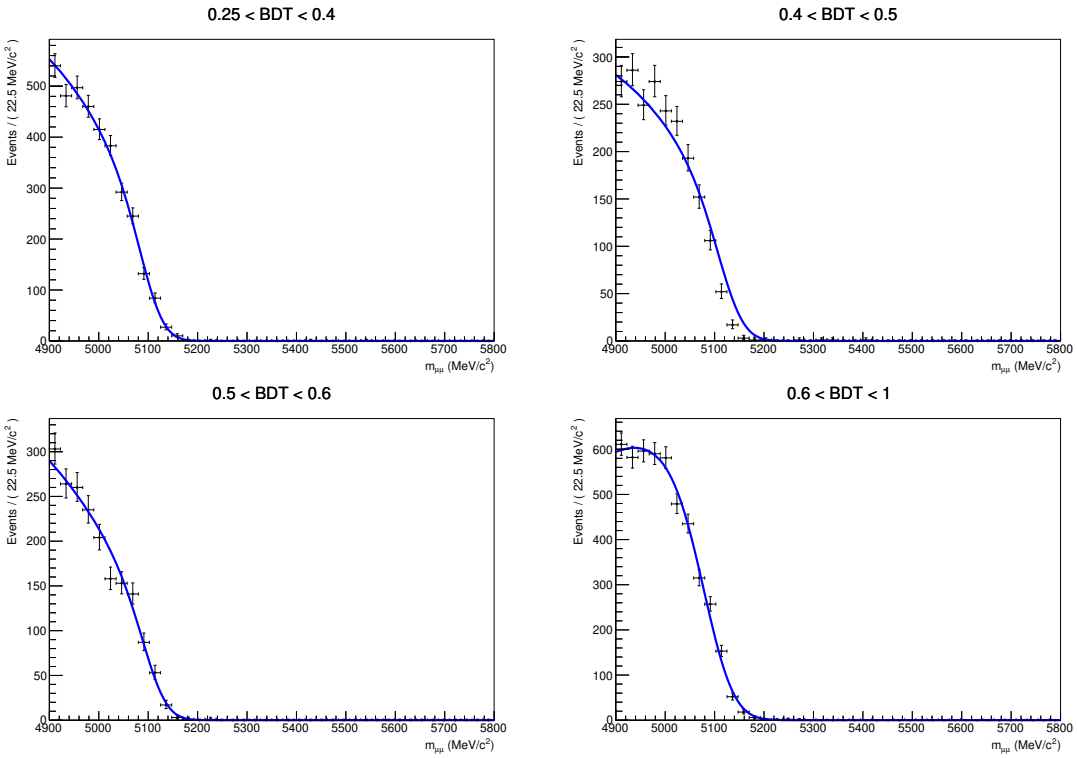

Figure 4.18: The mass shape for $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ background in BDT bins, taken from simulation.

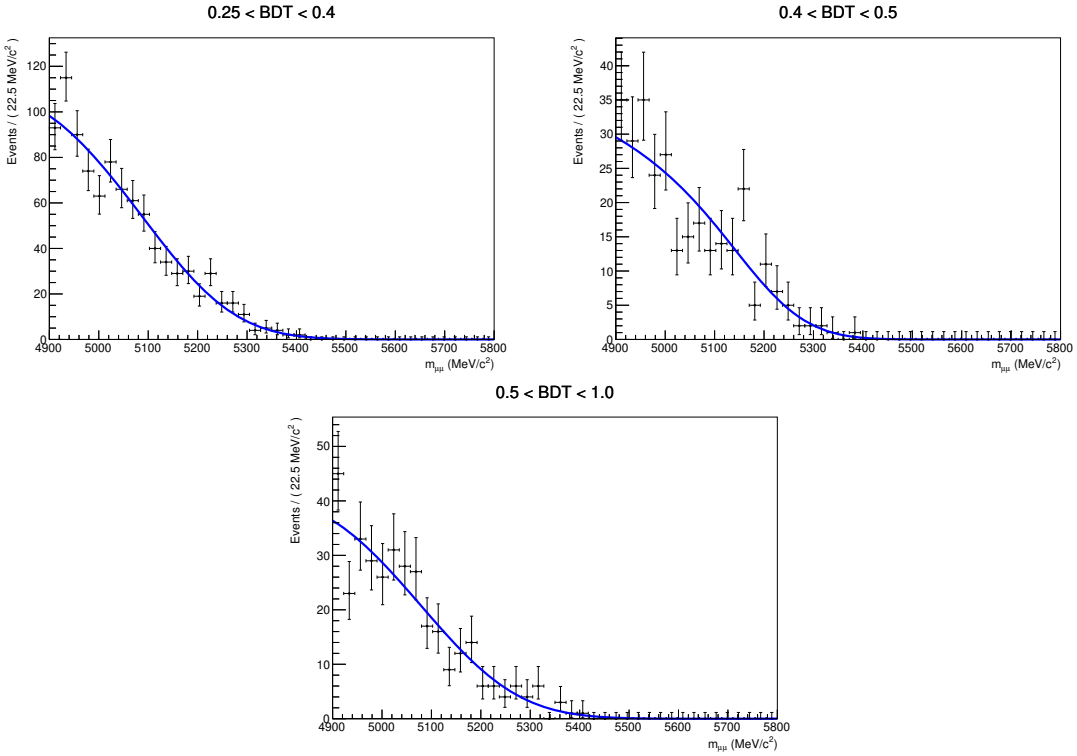

Figure 4.19: The mass shape for $B_c^+ \to J/\psi\mu^+\nu_\mu$ background in BDT bins, taken from simulation.

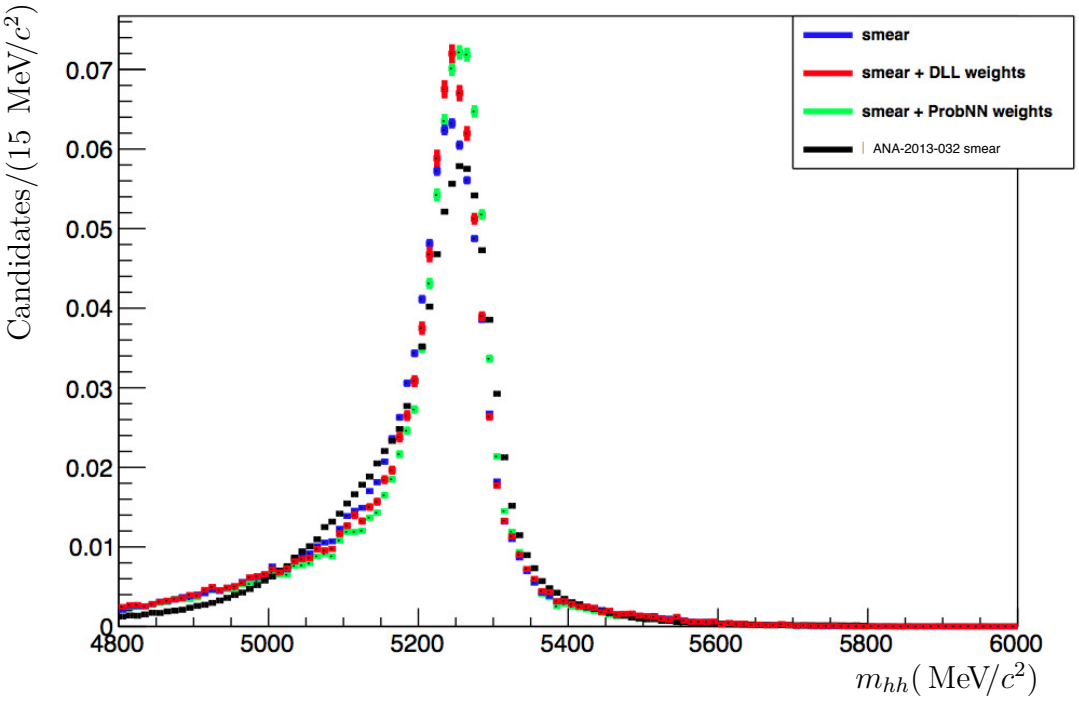

Figure 4.20: The mass shape for the $B^0_{(s)} \to h^+h'^-$ background with determined with smearing due to decays and flight and without, with DLL-based, and with ProbNN-based PID weights are shown in blue, red and green respectively. For comparison, the shape used in the Run 1 analysis, labeled 'ANA-2013-032', is shown in black. The green shape, denoted with 'smear + ProbNN weights', is used as the shape in the ML fit.

where $f^{\text{bkg}}_{\text{BDT},i}$ is the fraction of background candidates per BDT bin from simulation, f_x/f_u is the relative b hadron production fraction, $\varepsilon^{\text{bkg}}_{\text{tot}}$ is the efficiency to pass the total $B^0_{(s)} \to \mu^+\mu^-$ selection for a background, and $\mathcal{B}^{\text{bkg}}$ is the background branching fraction. Finally, β_u is the number of B^+ mesons produced at LHCb derived from a dataset, given by $\beta_u = \frac{N^{B^+\to J/\psi K^+}}{\mathcal{B}(B^+\to J/\psi K^+)\ \varepsilon^{B^+\to J/\psi K^+}_{\text{tot}}}$, with $N^{B^+\to J/\psi K^+}$ the number of observed $B^+ \to J/\psi K^+$ candidates per dataset determined in Section 4.5.1, $\mathcal{B}(B^+ \to J/\psi K^+)$ the branching fraction and $\varepsilon^{B^+\to J/\psi K^+}_{\text{tot}}$ the total efficiency for $B^+ \to J/\psi K^+$, which includes all data-driven corrections. For this analysis, β_u per dataset was given by $(7.18 \pm 0.24) \times 10^{11}$ for Run 1, $(1.27 \pm 0.06) \times 10^{11}$ for 2015, and $(4.62 \pm 0.21) \times 10^{11}$ for 2016.

The total efficiencies is split into three efficiencies, similar to Section 4.5; the acceptance efficiency ε_{acc}, the reconstruction and selection efficiency including PID ε_{sel}, and the trigger efficiency $\varepsilon_{\text{trig}}$. All of these efficiencies are taken from efficiency, with the exception of the PID efficiency, which is calibrated in a data-driven way with

the PIDCalib package. For most backgrounds, only simulated samples with 2012 or 2015 conditions were available. Therefore, samples with 2012 (2015) conditions were used to determine Run 1 (Run 2) efficiencies. As PID calibration samples were available for each year, the selection efficiencies including PID were determined separately per year.

$B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ The same reweighting is applied as used for the mass shapes in Section 4.6.1. For $B^0_s \to K^-\mu^+\nu_\mu$, the relative production rate f_s/f_d is taken into account. The inputs and resulting constrained yields are shown in Table 4.22.

$\Lambda^0_b \to p\mu^-\nu$ For $\Lambda^0_b \to p\mu^-\nu$, per-candidate weights are used for the production fraction ratio, $f_{\Lambda^0_b}/f_u$, as introduced in Section 4.6.1. By using β_u for normalisation and including these weights in the efficiency, $f_{\Lambda^0_b}/f_u$ is properly included in the normalisation of this background. The inputs and resulting constrained yields are shown in Table 4.22. The yield in 2016 is reduced relative to the size of the dataset due to the different ProbNN tuning used to suppress mis-identified backgrounds.

$B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ Both the neutral and charged mode need to be included for the total normalisation yield, but only the charged mode has been observed [82]. Therefore, a theoretical prediction of the relative branching fraction of $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ is used [83]:

$$\frac{\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-)}{\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)} = 0.47^{+0.22}_{-0.18} \quad , \tag{4.21}$$

Due to the long track isolation included in the BDT (see Section 4.2.6), which classifies candidates as more background-like if they are associated with a track from the rest of the event, the BDT shape for $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ is different. Therefore, they are evaluated separately, before constraining the total $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ yield to the sum of the two. The inputs and resulting constrained yields are shown in Table 4.23.

$B^+_c \to J/\psi\mu^+\nu_\mu$ For this decay, the relative production rate of B^+_c versus B^+ mesons needs to be taken into account. Fortunately, the relative production rate and branching fraction for $B^+_c \to J/\psi\mu^+\nu_\mu$ and $B^+ \to J/\psi K^+$ has been measured

at LHCb:

$$R_c = \frac{f_c}{f_u}\frac{\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)}{\mathcal{B}(B^+ \to J/\psi K^+)} = 0.152 \pm 0.032 \tag{4.22}$$

and the resulting normalisation formula is given by

$$N_{B_c^+\to J/\psi\mu^+\nu_\mu,i} = f^{\text{bkg}}_{\text{BDT},i}\ R_c\ \mathcal{B}(J/\psi \to \mu^+\mu^-)\ N^{B^+\to J/\psi K^+}\ \frac{\varepsilon^{B_c^+\to J/\psi\mu^+\nu_\mu}}{\varepsilon^{B^+\to J/\psi K^+}_{\text{tot}}} \tag{4.23}$$

In the normalisation table, $\mathcal{B}_{\text{eff}} = R_c \times \mathcal{B}(B^+ \to J/\psi K^+)\mathcal{B}(J/\psi \to \mu^+\mu^-)$ is included as an effective branching fraction. As the B_c^+ production rate had not been measured yet in Run 2, it is assumed to be the same as in Run 1. The inputs and resulting constrained yields are shown in Table 4.23.

4.6.3 Data-driven $B^0_{(s)} \to h^+h'^-$ normalisation from BDT calibration

The BDT calibration, which is used to determine the $B^0 \to K^+\pi^-$ yield per BDT bin, corrected for the hadron PID efficiency, is discussed in Section 4.3. It is used to get the expected $B^0_{(s)} \to h^+h'^-$ background yields after the $B^0_{(s)} \to \mu^+\mu^-$ selection, after correcting for the different efficiencies and rate of $B^0 \to K^+\pi^-$ relative to $B^0_{(s)} \to h^+h'^-$ decays:

$$N_i^{B^0_{(s)}\to h^+h'^-} = \frac{N_{K\pi,i}}{f_{K\pi}\ \varepsilon_{B^0\to K^+\pi^-,\ \text{trig.}}}\ \varepsilon_{B^0_{(s)}\to h^+h'^-\to\mu^+\mu^-}\ \varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\ \text{trig.}}, \tag{4.24}$$

where $f_{K\pi}$ is the relative fraction of $B^0 \to K^+\pi^-$ to $B^0_{(s)} \to h^+h'^-$, $\varepsilon_{B^0\to K^+\pi^-,\ \text{trig.}}$ is the trigger efficiency for $B^0_{(s)} \to h^+h'^-$, $\varepsilon_{B^0_{(s)}\to h^+h'^-\to\mu^+\mu^-}$ is the double misidentification rate for hadrons to muons and $\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\ \text{trig.}}$ is the $B^0_s \to \mu^+\mu^-$ trigger efficiency. $f_{K\pi}$ is determined by the relative production and branching fractions of each B meson and decay mode, as given in Section 4.3 in Table 4.7. The absolute value of $B^0 \to K^+\pi^-$ trigger efficiency was determined for the normalisation in Section 4.5.4, separated in the L0 and HLT1 TIS efficiency ($\varepsilon_{B^0\to K^+\pi^-,\ \text{L0\&HLT1 TIS}}$) and HLT2 TOS efficiency ($\varepsilon_{B^0\to K^+\pi^-,\ \text{HLT2 TOS|L0\&HLT1 TIS}}$). The variation of the trigger over BDT bins was taken from reweighted simulation, as described in Section 4.3.5. The double misidentification rate of $B^0_{(s)} \to h^+h'^-$ to $\mu^+\mu^-$ was determined for each final state with the PIDCalib package, in the same way as all exclusive backgrounds. The trigger efficiency for doubly misidentified $B^0_{(s)} \to h^+h'^-$ decays was assumed to be the same as for $B^0_s \to \mu^+\mu^-$ signal, as their kinematics are similar and the trigger

efficiency is determined after the PID selection. The estimated $B^0_{(s)} \to h^+h'^-$ yields per year are shown in Table 4.24.

4.6.4 Validation with $B^0_{(s)} \to h^-\mu^+\nu_\mu$ data sample

The estimated $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to h^-\mu^+\nu_\mu$ yields and mass shapes per BDT bin are checked by using the $B^0_{(s)} \to h^-\mu^+\nu_\mu$ data sample as a validation region. As discussed in Section 4.2, this sample differs from the $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to \mu^+\mu^-$ sample only in its PID selection. It consists of a $\pi^\pm\mu^\mp$ and a $K^\pm\mu^\mp$ data sample, which both contain three contributing processes. First, singly misidentified $B^0_{(s)} \to h^+h'^-$ decays, which appear as a peaking structure close to the B^0 mass. Second, semi-leptonic $B^0_{(s)} \to h^-\mu^+\nu_\mu$ decays, which appear as a shoulder at lower mass. Third, combinatorial background, which appears as an exponential contribution. A simultaneous fit to the invariant mass in all BDT bins is used to disentangle these components. The first BDT bin is not considered, as it is dominated by combinatorial background.

The $B^0_{(s)} \to h^+h'^-$ shape is a double-sided Crystal Ball function (see Section 4.4), with its mean and resolution floating and its tail parameters fixed from simulation. The $B^0_{(s)} \to h^-\mu^+\nu_\mu$ shapes are the same as used for the signal fit, described in Section 4.6.1, as they are found not to be affected by the PID selection. The combinatorial background shape is an exponential, and its shape parameter is shared over all BDT bins to improve the fit stability. One of the fits is shown in Figure 4.21. The $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to h^-\mu^+\nu_\mu$ background yields in each BDT bin are obtained from the fit, and are compared with the previous estimates after correcting for differences in PID and trigger efficiency. As for all backgrounds apart from $B^0_{(s)} \to h^+h'^-$, the PID efficiencies are calibrated in a data-driven way with the PIDCalib package, while the trigger efficiencies are taken from simulation.

The estimated $B^0_{(s)} \to h^+h'^-$ yields and $B^0 \to \pi^-\mu^+\nu_\mu$ yields per BDT bin, including their statistical and systematic uncertainty, are shown in Tables 4.25 and 4.26 respectively. The systematic uncertainty on the $B^0_{(s)} \to h^+h'^-$ estimate relative to the yield is 15 to 35% for the Run 1 and 2016 datasets, which together correspond to $\sim 90\%$ of the integrated luminosity, and higher for the 2015 samples, around 50 to 100%. On the $B^0 \to \pi^-\mu^+\nu_\mu$ estimate, the systematic uncertainty is below 15% for the higher BDT bins.

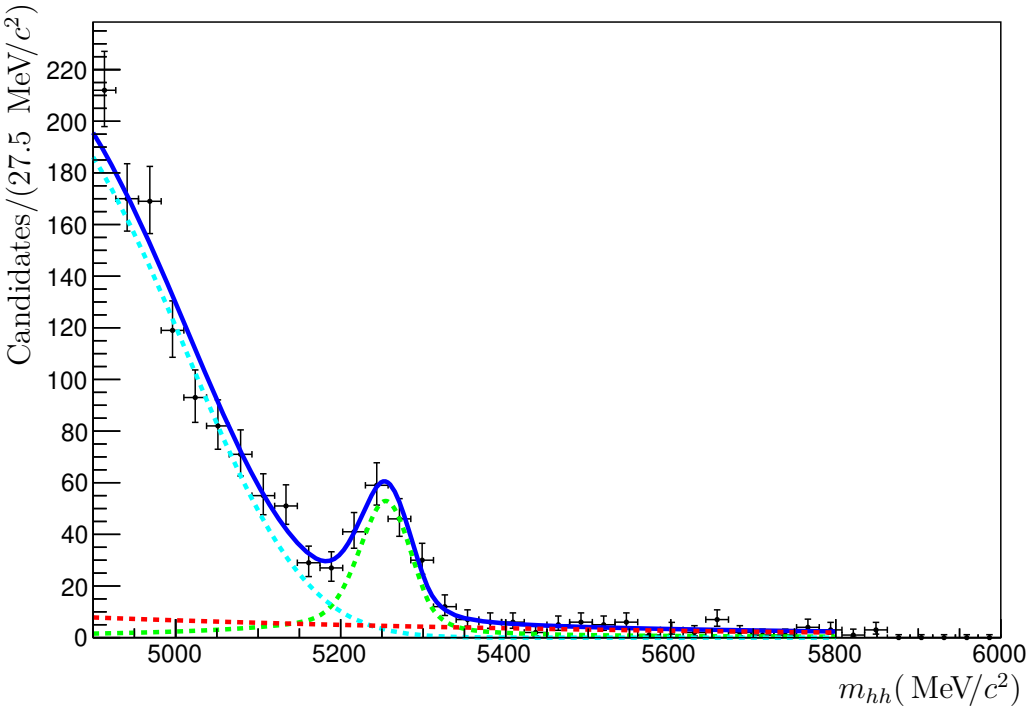

Figure 4.21: Bottom: a fit to $\pi^{\pm}\mu^{\mp}$ 2016 data in the BDT bin from 0.5 to 0.6. The cyan shoulder on the left describes $B^0 \to \pi^-\mu^+\nu_\mu$ decays, the green peaking structure $B^0_{(s)} \to h^+h'^-$ decays, and the red exponential combinatorial background.

Table 4.22: Inputs to and estimates of the $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$ and $\Lambda^0_b \to p\mu^-\nu$ peaking background yields for Run 1, 2015 and 2016 data.

	Run 1	2015	2016
$\mathcal{B}(B^0 \to \pi^-\mu^+\nu_\mu)$	$(1.44 \pm 0.05) \cdot 10^{-4}$		
ε_{acc}	$(6.745 \pm 0.004) \cdot 10^{-3}$	$(6.995 \pm 0.003) \cdot 10^{-3}$	$(6.995 \pm 0.003) \cdot 10^{-3}$
ε_{sel}	$(6.80 \pm 0.12) \cdot 10^{-5}$	$(5.16 \pm 0.17) \cdot 10^{-5}$	$(5.79 \pm 0.04) \cdot 10^{-5}$
$\varepsilon_{\text{trig}}$	0.926 ± 0.011	0.935 ± 0.011	0.935 ± 0.011
$\text{N}(B^0 \to \pi^-\mu^+\nu_\mu)$	44.0 ± 2.3	6.2 ± 0.4	25.2 ± 1.5
$\mathcal{B}(B^0_s \to K^-\mu^+\nu_\mu)$	$(1.42 \pm 0.35) \cdot 10^{-4}$		
f_s/f_d	0.259 ± 0.015	0.265 ± 0.017	0.265 ± 0.017
ε_{acc}	$(9.141 \pm 0.005) \cdot 10^{-3}$	$(9.465 \pm 0.004) \cdot 10^{-3}$	$(9.465 \pm 0.004) \cdot 10^{-3}$
ε_{sel}	$(2.34 \pm 0.07) \cdot 10^{-5}$	$(2.18 \pm 0.11) \cdot 10^{-5}$	$(1.82 \pm 0.02) \cdot 10^{-5}$
$\varepsilon_{\text{trig}}$	0.898 ± 0.020	0.961 ± 0.012	0.961 ± 0.012
$\text{N}(B^0_s \to K^-\mu^+\nu_\mu)$	5.04 ± 1.31	0.92 ± 0.24	2.81 ± 0.74
$\mathcal{B}(\Lambda^0_b \to p\mu^-\nu)$	$(4.1 \pm 1.0) \cdot 10^{-4}$		
ε_{acc}	$(1.34 \pm 0.01) \cdot 10^{-2}$	$(1.32 \pm 0.01) \cdot 10^{-2}$	$(1.32 \pm 0.01) \cdot 10^{-2}$
ε_{sel}	$(1.4 \pm 0.6) \cdot 10^{-6}$	$(4.0 \pm 1.5) \cdot 10^{-6}$	$(0.65 \pm 0.15) \cdot 10^{-6}$
$\varepsilon_{\text{trig}}$	0.801 ± 0.002	0.759 ± 0.002	0.759 ± 0.002
$\text{N}(\Lambda^0_b \to p\mu^-\nu)$	4.30 ± 2.11	2.09 ± 0.75	1.23 ± 0.48

Table 4.23: Inputs to and estimates of the $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ and $B^+_c \to J/\psi\mu^+\nu_\mu$ background yields for Run 1, 2015 and 2015 data.

	Run 1	2015	2016
$\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)$	$(1.83 \pm 0.25) \times 10^{-8}$		
ε_{acc}	0.2486 ± 0.0011	0.2503 ± 0.0010	0.2503 ± 0.0010
ε_{sel}	$(3.75 \pm 0.004) \cdot 10^{-3}$	$(3.51 \pm 0.004) \cdot 10^{-3}$	$(3.70 \pm 0.004) \cdot 10^{-3}$
$\varepsilon_{\text{trig}}$	0.958 ± 0.002	0.933 ± 0.003	0.933 ± 0.003
$\text{N}(B^+ \to \pi^+\mu^+\mu^-)$	11.8 ± 1.7	1.9 ± 0.3	7.3 ± 1.1
$\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-)$	$(0.86 \pm 0.36) \times 10^{-8}$		
ε_{acc}	0.2504 ± 0.0030	0.2501 ± 0.0012	0.2501 ± 0.0012
ε_{sel}	$(3.80 \pm 0.004) \cdot 10^{-3}$	$(3.54 \pm 0.004) \cdot 10^{-3}$	$(3.73 \pm 0.004) \cdot 10^{-3}$
$\varepsilon_{\text{trig}}$	0.954 ± 0.002	0.940 ± 0.003	0.940 ± 0.003
$\text{N}(B^0 \to \pi^0\mu^+\mu^-)$	5.6 ± 2.3	0.9 ± 0.4	3.5 ± 1.5
$\mathcal{B}_{B^+_c \to J/\psi\mu^+\nu_\mu,\text{eff}}$	$(9.5 \pm 2.0) \cdot 10^{-6}$		
ε_{acc}	$(4.35 \pm 0.006) \cdot 10^{-3}$	$(4.35 \pm 0.006) \cdot 10^{-3}$	$(4.35 \pm 0.006) \cdot 10^{-3}$
ε_{sel}	$(4.09 \pm 0.006) \cdot 10^{-3}$	$(2.64 \pm 0.006) \cdot 10^{-3}$	$(2.78 \pm 0.007) \cdot 10^{-3}$
$\varepsilon_{\text{trig}}$	0.964 ± 0.002	0.956 ± 0.003	0.956 ± 0.003
$\text{N}(B^+_c \to J/\psi\mu^+\nu_\mu)$	117.0 ± 25.1	13.3 ± 2.9	52.0 ± 11.1

Table 4.24: Inputs to and estimates of the $B^0_{(s)} \to h^+h'^-$ background yield for Run 1, 2015 and 2016 data.

	2011	2012	2015	2016
$f_{K\pi}$	0.562 ± 0.013			
$N_{K\pi,i}$	6833 ± 662	18012 ± 1221	8552 ± 848	28411 ± 2548
$\varepsilon_{B^0_{(s)}\to\mu^+\mu^-,\,\text{trig.}}$	$(94.2 \pm 2.0)\%$	$(94.1 \pm 2.0)\%$	$(96.1 \pm 3.2)\%$	$(97.3 \pm 1.3)\%$
$\varepsilon_{B^0\to K^+\pi^-,\,\text{trig.}}$	$(3.59 \pm 0.04)\%$	$(4.29 \pm 0.03)\%$	$(8.17 \pm 0.18)\%$	$(7.24 \pm 0.08)\%$
$\varepsilon_{B^0_{(s)}\to h^+h'^-\to\mu^+\mu^-}$ (10^{-6})	4.6 ± 0.2	4.6 ± 0.2	3.6 ± 0.2	3.4 ± 0.1
$N_{B^0_{(s)}\to h^+h'^-}$	1.47 ± 0.16	2.92 ± 0.32	0.59 ± 0.06	2.31 ± 0.21

Table 4.25: $N_{B^0_{(s)}\to h^+h'^-}$ constraint for the ML fit per BDT bin for Run 1, 2015 and 2016 data; the uncertainties are split into their statistical and systematic components.

BDT range	Run 1	2015	2016
[0-0.25]	0.640 ± 0.115 ± 0	0.112 ± 0.019 ± 0	0.441±0.068 ± 0
[0.25-0.4]	0.556 ± 0.067 ± 0.377	0.074 ± 0.012 ± 0.073	0.300±0.041 ± 0.204
[0.4-0.5]	0.398 ± 0.051 ± 0.210	0.060 ± 0.009 ± 0.033	0.202±0.034 ± 0.117
[0.5-0.6]	0.505 ± 0.044 ± 0.105	0.055 ± 0.009 ± 0.065	0.222±0.026 ± 0.129
[0.6-0.7]	0.482 ± 0.048 ± 0.120	0.064 ± 0.009 ± 0.041	0.221±0.025 ± 0.037
[0.7-0.8]	0.551 ± 0.050 ± 0.137	0.066 ± 0.009 ± 0.042	0.216±0.027 ± 0.093
[0.8-0.9]	0.585 ± 0.055 ± 0.097	0.075 ± 0.010 ± 0.041	0.242±0.027 ± 0.083
[0.9-1.0]	0.676 ± 0.057 ± 0.112	0.084 ± 0.012 ± 0.040	0.274±0.030 ± 0.040

Table 4.26: $N_{B^0\to\pi^-\mu^+\nu_\mu}$ constraint in the ML fit per BDT bin for Run 1, 2015 and 2016 data; the uncertainties are split into their statistical and systematic components.

BDT range	Run 1	2015	2016
[0-0.25]	$13.57 \pm 0.73 \pm 0$	$1.90 \pm 0.13 \pm 0$	$7.81 \pm 0.47 \pm 0$
[0.25-0.4]	$7.00 \pm 0.38 \pm 3.37$	$1.02 \pm 0.07 \pm 0.84$	$4.18 \pm 0.25 \pm 3.29$
[0.4-0.5]	$4.56 \pm 0.25 \pm 1.96$	$0.64 \pm 0.05 \pm 0.37$	$2.64 \pm 0.16 \pm 2.41$
[0.5-0.6]	$4.78 \pm 0.26 \pm 0.87$	$0.66 \pm 0.05 \pm 0.14$	$2.68 \pm 0.16 \pm 0.92$
[0.6-0.7]	$4.72 \pm 0.25 \pm 0.81$	$0.65 \pm 0.05 \pm 0.14$	$2.64 \pm 0.16 \pm 0.99$
[0.7-0.8]	$4.41 \pm 0.24 \pm 0.28$	$0.59 \pm 0.04 \pm 0.03$	$2.41 \pm 0.15 \pm 0.42$
[0.8-0.9]	$3.46 \pm 0.19 \pm 0.27$	$0.47 \pm 0.04 \pm 0.02$	$1.92 \pm 0.12 \pm 0.20$
[0.9-1.0]	$1.54 \pm 0.09 \pm 0.12$	$0.23 \pm 0.02 \pm 0.02$	$0.91 \pm 0.06 \pm 0.12$

4.7 $B^0_{(s)} \to \mu^+\mu^-$ maximum likelihood fit

The final part of the $B^0_{(s)} \to \mu^+\mu^-$ analysis is to search for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signal modes and measure their branching fractions with a maximum likelihood(ML) fit to the data. Before going into the specifics of the $B^0_{(s)} \to \mu^+\mu^-$ fit, the concepts behind a ML fit are explained.

The likelihood is the function that gives the probability to observe n measurements $\vec{x}_i$ given a model with unknown quantities $\vec{p}$, and is defined as

$$L(\vec{x}_i, \vec{p}) = \prod_i^n P(\vec{x}_i, \vec{p}), \tag{4.25}$$

where $P(\vec{x}_i, \vec{p})$ is a probability density function, which gives the probability to get a single measurement given the underlying model and parameters [84, 85]. The unknown quantities $\vec{p}$ include the parameters of interest, indicated as $\vec{\mu}$, as well as the nuisance parameters needed for the measurement, which are referred to as $\vec{\theta}$.

The best way to estimate the unknown quantities $\vec{p}$ is to find the set of parameter values that maximises the logarithm of the likelihood. The maximum likelihood estimator is given by the equation

$$\frac{d(\ln L(\vec{p}))}{d\vec{p}}|_{\vec{p}=\hat{\vec{p}}} = \frac{d(\sum_i^n \ln P(\vec{x}_i, \vec{p}))}{d\vec{p}}|_{\vec{p}=\hat{\vec{p}}} = 0. \tag{4.26}$$

In other words, the likelihood is maximised to estimate all the parameters $\hat{\vec{p}} = (\hat{\vec{\mu}}, \hat{\vec{\theta}}(\hat{\vec{\mu}}))$, where $\hat{\vec{p}}$ is the vector containing the set of parameter values that maximise the likelihood, $\hat{\vec{\mu}}$ is the vector of interesting parameters, and $\hat{\vec{\theta}}(\hat{\vec{\mu}})$ is the vector of nuisance parameters that maximises the likelihood given $\hat{\vec{\mu}}$. In practice, the maximum likelihood is computed by minimising the negative log likelihood, since taking the logarithm of the likelihood simplifies its calculation to a sum instead of a product and is numerically more stable.

The ML fit in this analysis is a simultaneous fit to the dimuon invariant mass in BDT bins in Run 1 and Run 2 data. The fit contains two parameters of interest, namely the branching fractions of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, and a large number of nuisance parameters, which are detailed in Table 4.27. Most of the nuisance parameters are measured, as discussed in Sections 4.3, 4.4, 4.5 and 4.6 for the BDT and mass distribution of signal, normalisation of signal over the full BDT range, and normalisation, mass and BDT distributions of exclusive backgrounds

Nuisance parameters	Calibration method	Status in ML fit
$f^{\text{BDT}}_{B^0_{(s)}}$	BDT calibration	Constrained
$M_{B^0_{(s)}}$	Mass calibration	Constrained
$\alpha_{B^0_s\to\mu^+\mu^-}$, $\alpha_{B^0\to\mu^+\mu^-}$	Normalisation	Constrained
N_{bkg}	Exclusive backgrounds	Constrained
M_{bkg}	Exclusive backgrounds	Fixed
N_{comb}, M_{comb}	-	Free

Table 4.27: Nuisance parameters in the $B^0_{(s)} \to \mu^+\mu^-$ ML fit.

respectively. In the fit, most nuisance parameters are constrained, which means that the uncertainties on these parameters are included in the fit as Gaussian constraints and the systematic uncertainties from these nuisance parameters are directly obtained from the fit. The only nuisance parameters that are fixed are the mass shapes for the exclusive backgrounds. Finally, the nuisance parameters associated with the combinatorial background, which is modeled with an exponential shape, are left free. Its shape parameter is shared over the BDT bins for each Run, while the yield is free to float in each bin.

Given the description of the fit model, the number of events in the data as function of the fit variables is given as

$$\begin{aligned} N_{\text{data}}(m, i_{\text{BDT}}, j_{\text{Run}}) = {} & \mathcal{B}(B^0_s \to \mu^+\mu^-) \times \alpha_{B^0_s\to\mu^+\mu^-}(j_{\text{Run}}) \times M_{B^0_s}(m, j_{\text{Run}}) \times f^{\text{BDT}}_{B^0_s}(i_{\text{BDT}}, j_{\text{Run}}) \\ & + \mathcal{B}(B^0 \to \mu^+\mu^-) \times \alpha_{B^0\to\mu^+\mu^-}(j_{\text{Run}}) \times M_{B^0}(m, j_{\text{Run}}) \times f^{\text{BDT}}_{B^0}(i_{\text{BDT}}, j_{\text{Run}}) \\ & + \sum_i N_{\text{bkg}}(i, i_{\text{BDT}}, j_{\text{Run}}) \times M_{\text{bkg}}(i, m, i_{\text{BDT}}, j_{\text{Run}}) \\ & + \sum_i N_{\text{comb}}(i_{\text{BDT}}, j_{\text{Run}}) \times M_{\text{comb}}(m, j_{\text{Run}}). \end{aligned} \tag{4.27}$$

To summarise, the role of the ML fit is to determine $\mathcal{B}$ $(B^0_s \to \mu^+\mu^-)$ and $\mathcal{B}$ $(B^0 \to \mu^+\mu^-)$, taking into account the uncertainties on most nuisance parameters and the statistical uncertainties in the $B^0_{(s)} \to \mu^+\mu^-$ sample. However, as both $B^0_{(s)} \to \mu^+\mu^-$ decays had not yet been observed with only data from LHCb, the significance of the signal should be determined, to decide whether an observation or evidence for either signal can be claimed.

4.7.1 Determining the significance of a signal

When searching for a new signal, two hypotheses should be compared; The background-only hypothesis and the hypothesis that includes both background and signal. For this analysis, to avoid cross-feed between the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ signals affecting the significance estimates, the null hypothesis for either significance estimate treats the other signal as an unconstrained background, which means leaving its branching fraction free.

The statistical significance of a signal is quantified with the p-value, which is the probability of finding data equally or less compatible with an hypothesis than the observed data. In particle physics, the p-value is converted into the equivalent number of standard deviations Z the measurement deviates from an hypothesis with

$$Z = \Phi^{-1}(1-p), \tag{4.28}$$

where Φ^{-1} is the inverse of the cumulative Gaussian distribution. In this thesis, significance is used to refer to Z. For a discovery, the p-value is computed with respect to the background-only hypothesis. To avoid claiming a discovery for random fluctuations of the data, particle physics has a stringent treshold before claiming the observation of a signal, namely a significance of at least $Z = 5$ or corresponding p-value of 2.87×10^{-7}.

The significance is determined with the so-called test statistic, which is a quantity that distinguishes two hypotheses under test and is denoted as t_μ, where μ is a given value for the parameter of interest that is tested. The test statistic used in this analysis is the profile likelihood ratio, as it has been shown to have the largest statistical power to separate different hypotheses [86]. The profile likelihood ratio is defined as

$$\lambda_\mu = \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\hat{\vec{\theta}}}(\hat{\mu}))}, . \tag{4.29}$$

The likelihood in the denominator is maximised unconditionally, and $\hat{\mu}$ and $\hat{\hat{\vec{\theta}}}(\hat{\mu})$ are the resulting values for the parameter of interest and all nuisance parameters. The numerator contains the likelihood for a hypothesised value of μ, and $\hat{\hat{\vec{\theta}}}(\mu)$ is the associated value for $\vec{\theta}$ that maximises the likelihood given μ. The associated test

statistic t_μ is defined as

$$t_\mu = -2 \ln \lambda_\mu, \tag{4.30}$$

Clearly, $t_{\hat{\mu}} = 0$, where the negative log likelihood is minimised, and is positive everywhere else. The larger the observed value of t_μ, the stronger the disagreement of the data with the model with the value of μ under test, and the smaller the p-value.

From the ML fit to the data, a certain value for the test statistic $t_{\mu,\text{obs}}$ is found. To convert it to a p-value, the distribution of t_μ given a value of μ for the data, $f(t_\mu|\mu)$, is required, after which the p-value is found by computing

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu) dt_\mu. \tag{4.31}$$

There are two ways to determine $f(t_\mu|\mu)$. The computationally intensive but conceptually easy way is to simulate toy data samples, which can be used to test the behaviour of the ML fit for an ensemble of data samples instead of just the single real one. This option is used to validate the fit and described in detail in Section 4.7.2.

The other approach is based on the large sample limit approximation, which turns out to hold already to good precision even for small samples and in which case the behaviour of the likelihood is known analytically [87]. Expanding around the maximum of the likelihood and ignoring higher order terms, it behaves as a parabola:

$$\log L(\mu, \hat{\vec{\theta}}(\mu)) = \log L(\hat{\mu}, \hat{\vec{\theta}}(\hat{\mu})) - \frac{(\mu - \hat{\mu})^2}{2\sigma^2_{\hat{\mu}}}, \tag{4.32}$$

or equivalently

$$t_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2_{\hat{\mu}}}, \tag{4.33}$$

where $\sigma_{\hat{\mu}}$ is the standard deviation of $\hat{\mu}$, which is obtained from the covariance matrix produced by the ML fit. As a result, the significance Z for a given μ is found to be

$$Z = \frac{|\mu - \hat{\mu}|}{\sigma_{\hat{\mu}}} = \sqrt{t_\mu}. \tag{4.34}$$

Applying these concepts to the significance for the discovery of a decay mode, the test statistic t_μ should be examined for the background-only model, for which $\mu = 0$. Note that a discovery should only be considered if $\hat{\mu} \geq 0$, because a negative

value for $\hat{\mu}$ would not be indicative of a negative and therefore unphysical branching fraction, but of a background fluctuation. Therefore, the formal definition of the test statistic for discovery q_0 is

$$q_0 = \begin{cases} t_0 = -2\ln\lambda_0 & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{4.35}$$

which takes both of these conditions into account. Converting this to a significance, we find the significance for a discovery

$$Z_0 = \sqrt{q_0} = \begin{cases} \sqrt{-2\ln\lambda_0} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0. \end{cases} \tag{4.36}$$

If Z_0 is found to be larger than five, an observation of a signal can be claimed, while if it is in between three and five, evidence for a signal can be claimed.

However, this approach to calculate the significance is dependent on the large sample limit approximation, which is validated with the other method, toy studies.

4.7.2 Fit validation and sensitivity estimate

The principle of fast Monte Carlo or "toy" studies is to generate many toy datasets with the expected signal and background yields and distributions. The fit performance was judged on this ensemble of datasets, to make sure it returned an unbiased estimate of the branching fractions with a proper estimate of their uncertainties.

For the toy studies, the signal and exclusive background distributions determined in the previous sections were used. The mass shape and yields for the combinatorial background and the yields of all exclusive backgrounds, using a fixed mass shape, were determined by fitting a background-only model to the data, which is blinded in the mass range from $m(\mu^+\mu^-) \in [5200, 5445]\,\text{MeV}/c^2$. The $B^0_{(s)} \to \mu^+\mu^-$ ML fit also includes both $B^0_{(s)} \to \mu^+\mu^-$ signals, generated with their Standard Model branching fraction.

The main questions studied to validate the $B^0_{(s)} \to \mu^+\mu^-$ ML fit are:

- Does the fit return an unbiased estimate of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions including a proper estimate of the uncertainty?

- What is the expected significance of the $B^0_{(s)} \to \mu^+\mu^-$ signals, assuming the SM values for their branching fractions?

These two questions are kept in mind while choosing the optimal set of BDT bins, as the BDT was reoptimised since the previous analysis. Instead of using the binning from the previous analysis, which was $[0.00, 0.25, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00]$ and was used in the previous chapters when calibrating nuisance parameters, it is found to be better to use a four-bin configuration, namely $[0.25, 0.40, 0.50, 0.60, 1.00]$. The first bin is removed, because it was dominated by combinatorial background, and the last four bins are merged, to make sure there is enough combinatorial in the high-mass sideband to reliably fit its yield. In this configuration, the fit returns an unbiased estimate with a proper uncertainty of the B^0_s and B^0 branching fractions, given the uncertainties from the finite number of toys.

The expected significance of the $B^0_s \to \mu^+\mu^-$ signal is found to be 9.7 σ, and all of the 1413 toys exceed the 5 σ significance threshold for claiming an observation, while the $B^0 \to \mu^+\mu^-$ signal has an expected significance of about 1.5 σ, with an 8% probability for a 3 σ signal significance, which would provide an evidence of $B^0 \to \mu^+\mu^-$ decays.

4.7.3 Setting a limit

In case the statistical significance of a signal is lower than the five standard devations required for claiming an observation, it is customary to determine an upper limit on the branching fraction. In this analysis, it is expected to put an upper limit on the $B^0 \to \mu^+\mu^-$ branching fraction, given the probability of finding evidence for this decay.

An upper limit is determined using the concept of a confidence interval. A confidence interval is an interval that, if the measurement would be repeated many times, contains the true value of the parameter of interest in a certain percentage, called the confidence level (CL), of the measurements.

Intimately related to a confidence interval is the concept of a p-value as defined in Section 4.7.1 If the p-value is found below a threshold α, given a hypothesised value μ for the parameter of interest, then that value μ is said to be excluded at a confidence level of $1 - \alpha$, and the set of not-excluded points form a confidence interval with the same confidence level.

An upper limit is then obtained by giving a one-sided confidence interval. First, a test statistic needs to be defined, in order to compute an associated p-value. In case of an upper limit, the null hypothesis is to have a signal with some signal strength μ and background, and the question is for which value of μ that hypothesis can be rejected in favour of the background-only hypothesis.

Therefore, the test statistic for upper limits is defined as

$$q_\mu = \begin{cases} -2\ln\lambda_\mu & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{4.37}$$

which analogously to the test statistic for discovery equals zero when testing an hypothesis where the signal strength μ does not make sense, in this case in case if μ is smaller than the signal strength observed in data $\hat{\mu}$, and otherwise uses $-2\ln\lambda_\mu$ as a measure of the discrepancy between the two hypotheses under test.

Using the same procedure as in Section 4.7.1 for the discovery test statistic, one finds the significance for the upper limit test statistic to be

$$Z_\mu = \sqrt{q_\mu} = \Phi^{-1}(1 - p_\mu) \tag{4.38}$$

where the definition of p_μ is taken from Equation (4.31) with q_μ as test statistic. Because an upper limit is a one-sided confidence interval at confidence level $1 - \alpha$, the upper limit on μ is simply found by setting as threshold $p_\mu = \alpha$ and solving for μ, using Equations (4.33) and (4.38) to find

$$\mu_{\text{up}} = \hat{\mu} + \sigma_{\hat{\mu}}\Phi^{-1}(1 - \alpha). \tag{4.39}$$

For example, setting $\alpha = 0.05$ gives $\Phi^{-1}(1 - \alpha) = 1.645$ and directly allows to determine the upper limit from the likelihood.

However, using this procedure, an experiment will have probability α to exclude hypotheses to which one is weakly sensitive or not sensitive at all. In this analysis, that corresponds to hypotheses with a small value for the $B^0 \to \mu^+\mu^-$ branching fraction.

To avoid such an over-exclusion, the CLs method is used to set an upper limit for this analysis [88]. This method is designed to properly take into account the sensitivity of a measurement to separate the background-only hypothesis from the signal plus background hypothesis. It does so by penalising the α term in Equation (4.39), based on the overlap between the two hypotheses under test.

The overlap is quantified by computing the p-value, assuming there is no signal in the data, to obtain a the same or a more background-like value for the test statistic compared to the one found in data. In case there is a large overlap, both p-values will be similar and the ratio of p-values will be too high for it to pass the exclusion threshold. This p-value is commonly referred to as $1 - p_b$:

$$1 - p_b = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|0) dq_\mu, \tag{4.40}$$

and the resulting upper limit is computed with

$$\frac{p_\mu}{1 - p_b} = \alpha, \tag{4.41}$$

such that Equation (4.39) is modified to

$$\mu_{\text{up}} = \hat{\mu} + \sigma_{\hat{\mu}} \Phi^{-1}[1 - \alpha + \alpha\Phi(-\hat{\mu}/\sigma_{\hat{\mu}})]. \tag{4.42}$$

Compared to that equation, a penalty term $\alpha\Phi(-\hat{\mu}/\sigma)$ is included, which means that the upper limit μ_{up} will be higher for cases where the two hypotheses are not fully separated and prevents an hypothesis from being excluded by a measurement that is not sensitive to it.

Chapter 5

Results

In this Chapter, the results of the $B^0_{(s)} \to \mu^+\mu^-$ analysis described in the previous Chapter are presented. In addition to the branching fraction measurement, the first measurement of the effective lifetime of the $B^0_s \to \mu^+\mu^-$ decay is presented. The results of the maximum likelihood fit leading to the measurement of the branching fraction of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are discussed in Section 5.1 and compared to theory. In Section 5.2, the measurement of the effective lifetime of the $B^0_s \to \mu^+\mu^-$ decay is outlined and the results (interpreted as an experimental constraint on the observable $A_{\Delta\Gamma}$, introduced in Equation 2.37) are compared to the theoretical predictions. The experimental correlation between the measurements of the branching fraction and the lifetime of $B^0_s \to \mu^+\mu^-$ is discussed in Section 5.3.

5.1 Branching fraction of $B^0_s \to \mu^+\mu^-$ and search for $B^0 \to \mu^+\mu^-$

The maximum likelihood fit to search for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays and measure their branching fractions has been described in Section 4.7. The mass distribution for $B^0_{(s)} \to \mu^+\mu^-$ candidates and the result of the unbinned maximum likelihood fit are shown in Figure 5.1 (for the most sensitive BDT bins combined) and in Figure 5.2 (separated per BDT bin).

As discussed in Section 2.6.1, the experimental definition for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is used throughout this thesis, assuming the Standard Model value for the effective lifetime and $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$. With the nominal fit, the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are found to be $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}$, where the first uncertainty is statistical and the second is

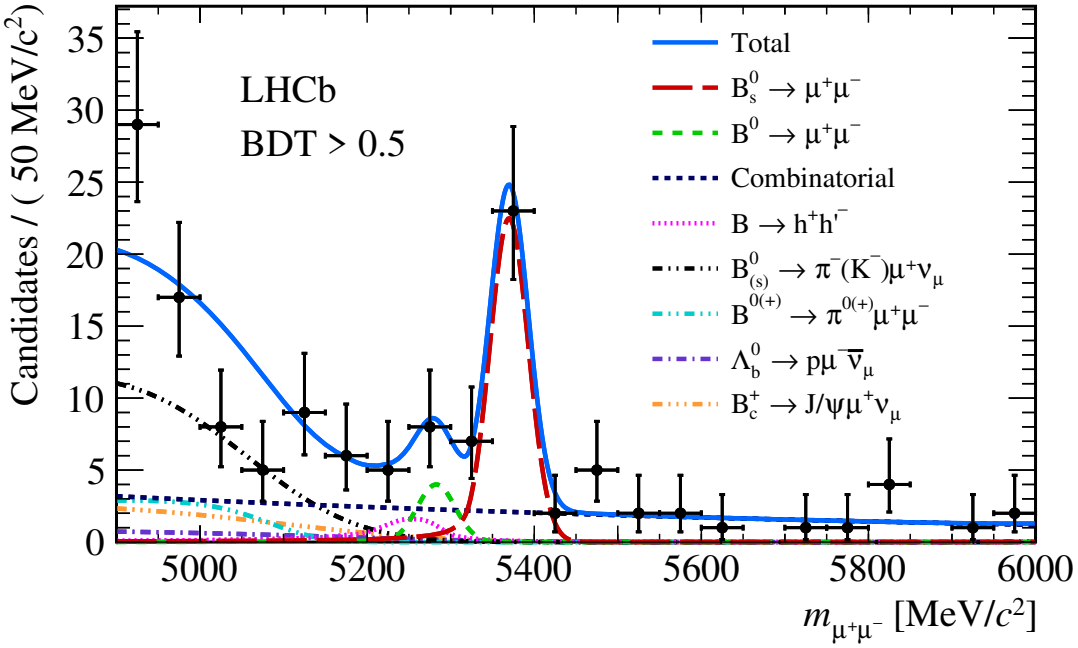

Figure 5.1: The dimuon invariant mass distribution of $B^0_{(s)} \to \mu^+\mu^-$ candidates which pass the requirement BDT > 0.5 for Run 1 and Run 2 data combined. The result of the unbinned maximum likelihood fit is overlaid.

systematic in nature. The $B^0_s \to \mu^+\mu^-$ decay is observed with a significance of 7.8 standard deviations compared to the background-only fit, which is the first single experiment observation of the $B^0_s \to \mu^+\mu^-$ decay, while the $B^0 \to \mu^+\mu^-$ signal is not significant enough to claim evidence for this decay. Therefore, a limit on its branching fraction is determined with the CL_s technique at the 95% confidence level, $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.4 \times 10^{-10}$.

The nominal fit returns the total uncertainty on each branching fraction, propagating systematic uncertainties on constrained nuisance parameters to the final result, as described in Section 4.7. An overview of all nuisance parameters in the fit is given in Table 4.27. The statistical uncertainty is isolated by performing the fit with all constrained nuisance parameters fixed, and the systematic uncertainty is given by quadratically subtracting the statistical from the total uncertainty. The main systematic uncertainty for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is the uncertainty on the ratio of B^0_s and B^0 production fractions f_s/f_d, while the main systematic uncertainty for $\mathcal{B}(B^0 \to \mu^+\mu^-)$ comes from the uncertainty on the exclusive background estimates for $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to h^-\mu^+\nu_\mu$.

Compared with the predictions of both branching fractions in the Standard Model as given in Section 2.6.3, $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.58 \pm 0.13) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) =$

$(1.02^{+0.029}_{-0.055}) \times 10^{-10}$, the measurements are found to be consistent with the Standard Model (see the prediction in Section 2.6.3), at the level of 1 σ. The two-dimensional confidence intervals for the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ vs. $\mathcal{B}(B^0 \to \mu^+\mu^-)$ measurement are shown in Figure 5.3, which shows that the correlation between the two measurements is small and confirms that the combined measurement is consistent with the Standard Model.

The central value of the $B^0 \to \mu^+\mu^-$ branching fraction measurement is different from the value found in a previous combined measurement by CMS and LHCb using Run 1 data, namely $\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}$ [41]. The samples used for the two measurements are close to uncorrelated, as the $B^0_{(s)} \to \mu^+\mu^-$ selection changed significantly. The reoptimised BDT results in the same level of combinatorial background at a BDT cut > 0.5 instead of > 0.7, equivalent to $\sim 40\%$ more signal candidates, while the optimised PID selection rejects five times more $B^0_{(s)} \to h^+h'^-$ background candidates at a cost of only 10% of the signal candidates. The two measurements are consistent at the level of 1.3 σ.

The main conclusion of this measurement is that the branching fractions of $B^0_{(s)} \to \mu^+\mu^-$ decays are at the level of their SM predictions, which strongly constrains contributions from scalar particles to $B^0_s \to \ell^+\ell^-$ decays specifically and the $b \to s\ell\ell$ transition in general. It is interesting to note that the central values of recent CMS and ATLAS measurements agree with this measurement, finding a $B^0_s \to \mu^+\mu^-$ branching fraction 20% smaller than the SM prediction. This is consistent with possible explanations for the tensions seen in other decays governed by the $b \to s\ell\ell$ transition, as discussed in more detail in Section 6.1.

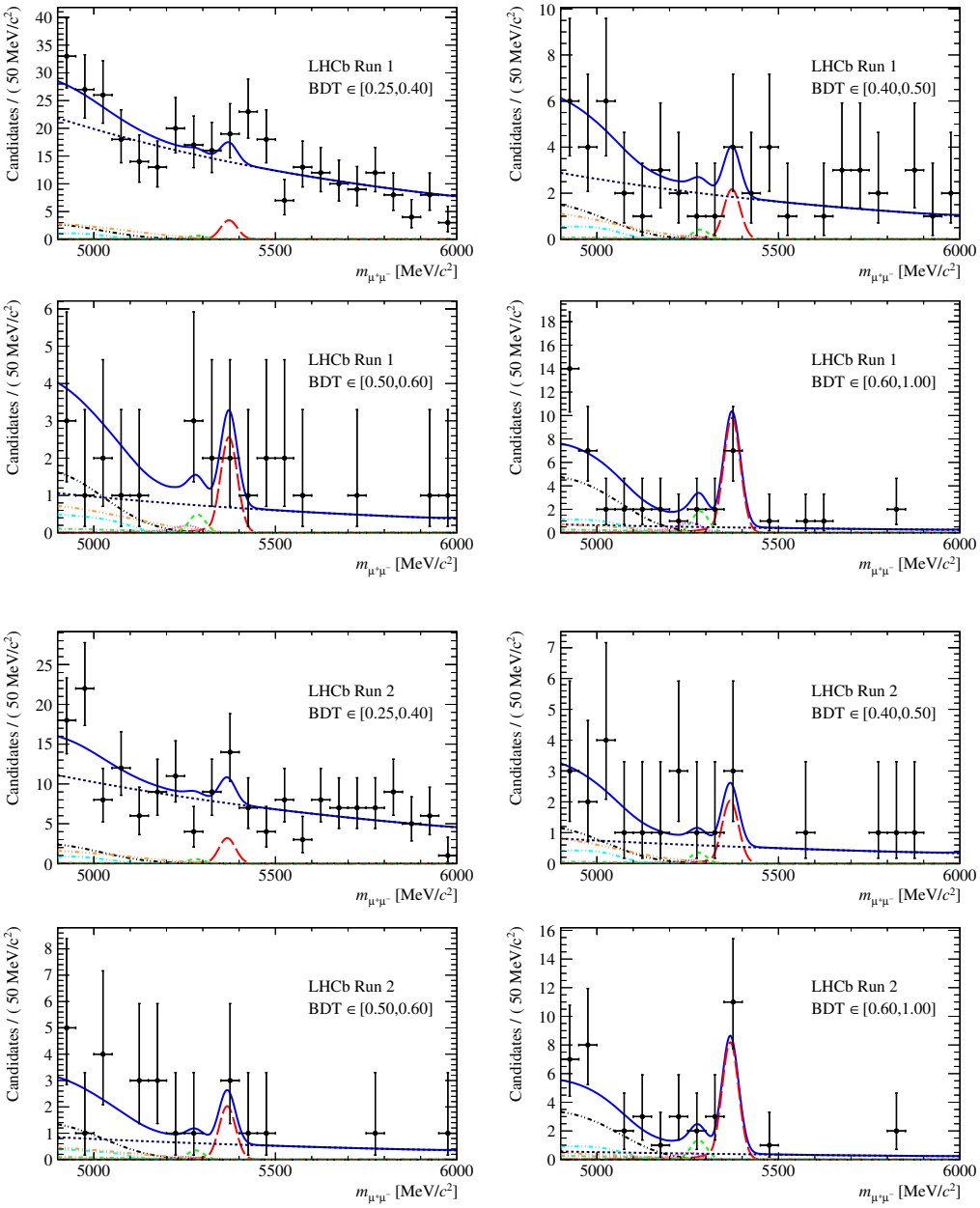

Figure 5.2: Mass distributions of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) in bins of BDT. The result of the fit is overlaid (blue solid line) and the different components are detailed: $B^0_s \to \mu^+\mu^-$ (red long dashed), $B^0 \to \mu^+\mu^-$ (green medium dashed), combinatorial background (dark blue short dashed), $B^0_{(s)} \to h^+h'^-$ (magenta dotted), $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ (black dot-dashed), $B^{(0/+)} \to \pi^{(0/+)}\mu^+\mu^-$ (cyan dot-dashed), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange dot-dashed) and $\Lambda^0_b \to p\mu^-\nu$ (violet dot-dashed).

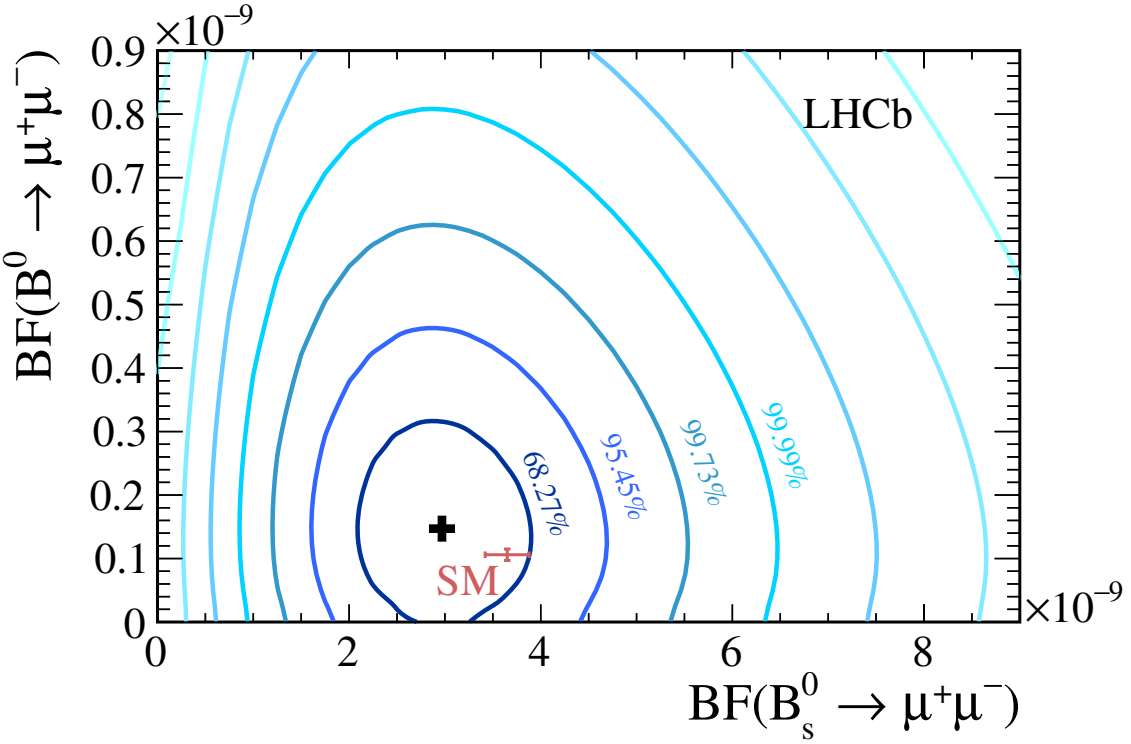

Figure 5.3: The result of the branching fraction measurement in the two-dimensional $\mathcal{B}(B_s^0 \to \mu^+\mu^-) - \mathcal{B}(B^0 \to \mu^+\mu^-)$-plane. The Standard Model prediction, including uncertainties, is indicated by the red cross. The central value from the measurement is shown as the black plus sign, and the profile likelihood contours corresponding to a confidence level associated with an number of standard deviations are shown as blue contours.

5.2 Effective lifetime of $B_s^0 \to \mu^+\mu^-$

5.2.1 Introduction

As part of the $B^0_{(s)} \to \mu^+\mu^-$ analysis, the effective lifetime of $B^0_s \to \mu^+\mu^-$ decays was measured as well. The effective lifetime of $B^0_s \to \mu^+\mu^-$ is defined as the time expectation value of the untagged decay rate $\Gamma(B^0_s(t) \to \mu^+\mu^-)$:

$$\tau_{\text{eff}}^{\mu^+\mu^-} = \frac{\int_0^\infty t \, \langle \Gamma(B^0_s(t) \to \mu^+\mu^-) \rangle \, dt}{\int_0^\infty \langle \Gamma(B^0_s(t) \to \mu^+\mu^-) \rangle \, dt}. \tag{5.1}$$

As discussed in Section 2.4.2, B^0_s mesons mix and decay as CP eigenstates which have two distinct lifetimes. Therefore, the effective lifetime can be used to determine the mixture of CP-odd and CP-even states contributing to $B^0_s \to \mu^+\mu^-$, known as $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$, using Equation 2.39 :

$$\tau_{\text{eff}}^{\mu^+\mu^-} / \tau_{B^0_s} = \frac{1}{1 - {y_s}^2} \left[\frac{1 + 2A^{\mu^+\mu^-}_{\Delta\Gamma_s} y_s + {y_s}^2}{1 + A^{\mu^+\mu^-}_{\Delta\Gamma_s} y_s} \right], \tag{5.2}$$

where y_s is the related to the difference in lifetimes between the two CP states, defined as

$$y_s = \frac{\Delta\Gamma_s}{2\Gamma_s}. \tag{5.3}$$

Although B^0 mesons mix as well, y_d is negligibly small and therefore the effective lifetime of $B^0 \to \mu^+\mu^-$ is not sensitive to $A^{\mu^+\mu^-}_{\Delta\Gamma_d}$. The constraint on $A^{\mu^+\mu^-}_{\Delta\Gamma_s}$, obtained by measuring the effective lifetime, is complementary to the branching fraction measurement, as it separates scalar from pseudoscalar contributions and acts a strong test of the Standard Model(see Section 2.7.

In the Standard Model, $A^{\mu^+\mu^-}_{\Delta\Gamma_s} = A^{\mu^+\mu^-}_{\Delta\Gamma_d} = +1$ and the decay time distribution of $B^0_s \to \mu^+\mu^-$ is a single exponential with the CP-odd B^0_s lifetime.

5.2.2 Strategy

Experimentally, the analysis strategy for the $B^0_s \to \mu^+\mu^-$ effective lifetime consists of three components. Firstly, the selection of the $B^0_{(s)} \to \mu^+\mu^-$ sample is re-optimised for $B^0_s \to \mu^+\mu^-$ decays only. Subsequently, the dimuon invariant mass is used to

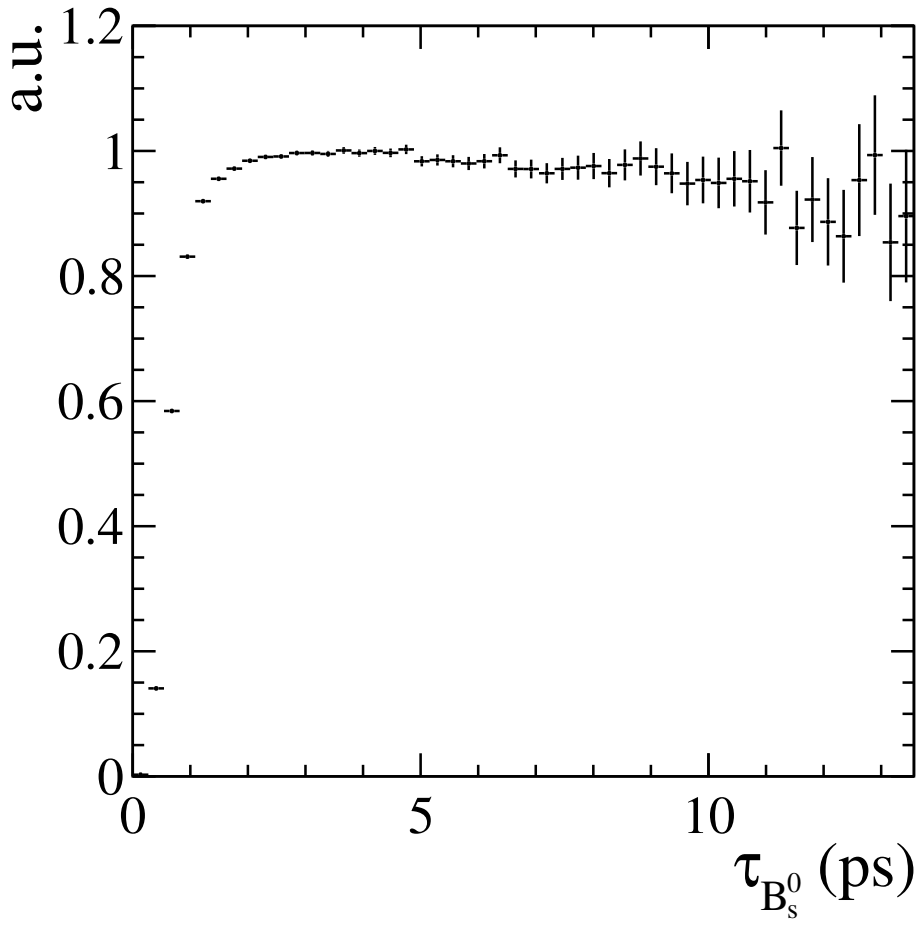

Figure 5.4: The lifetime acceptance distribution for simulated $B_s^0 \to \mu^+\mu^-$ candidates.

separate signal from background with a maximum likelihood fit, and the mass shapes are used by the *sPlot* method to statistically subtract the background and obtain a sample representative of $B_s^0 \to \mu^+\mu^-$ signal. Finally, the effective lifetime is measured from the background-subtracted decay time distribution using a maximum likelihood fit. As the efficiency to select and reconstruct $B_s^0 \to \mu^+\mu^-$ candidates is not uniform as function of the B_s^0 meson decay time, it is taken into account for the fit model. The non-uniformity of the efficiency is referred to as the decay time acceptance, and is shown in Figure 5.4 for simulated $B_s^0 \to \mu^+\mu^-$ decays. The fit model is a single exponential with the effective lifetime convolved with the decay time acceptance, as would be expected in the SM.

As the effective lifetime is measured for $B_s^0 \to \mu^+\mu^-$ candidates, *i.e.* ignoring $B^0 \to \mu^+\mu^-$, the selection and fit strategy are optimised accordingly, mostly using "toy" studies to compare different scenarios. This leads to a few essential differences compared to the selection for the branching fraction measurement:

- The PID selection is loosened, as the main misidentified backgrounds, $B^0_{(s)} \to h^+h'^-$ and $\Lambda^0_b \to p\mu^-\nu$, do not affect $B^0_s \to \mu^+\mu^-$ as much as they do $B^0 \to \mu^+\mu^-$. The optimal cut point is found at ProbNN = $\text{ProbNN}_\mu \times (1 - \text{ProbNN}_K) \times (1 - \text{ProbNN}_p) > 0.2(0.4)$ for Run 1 and 2015 (2016) data.

- Only $B^0_s \to \mu^+\mu^-$ candidates with a dimuon invariant mass of at least 5320 MeV/c^2 are considered, instead of the 4900 MeV/c^2 used for the branching fraction fit. As a result, $B^0 \to \mu^+\mu^-$ signal and exclusive backgrounds are excluded from the sample. This greatly simplifies the mass fit, as only combinatorial background remains, and was shown to have a negligible effect on the lifetime measurement with "toy" studies.

- The mass fit is a single fit to all candidates in Run 1 and Run 2 combined which pass BDT > 0.55. As shown in Figure 5.2, the part of the $B^0_{(s)} \to \mu^+\mu^-$ sample that does not pass this cut is predominantly combinatorial background. Fitting in BDT bins was investigated, but the fit results were found to be inaccurate due to the small number of $B^0_s \to \mu^+\mu^-$ candidates per bin.

5.2.3 Decay time acceptance

The efficiency to trigger, reconstruct and select $B^0_s \to \mu^+\mu^-$ candidates is not uniform in decay time. The relative efficiency as function of decay time is known as the decay time acceptance. It is modeled using the function

$$\varepsilon(t) = \frac{[a(t - t_0)]^n}{1 + [a(t - t_0)]^n}, \text{ for } t > t_0 \tag{5.4}$$

where t_0 is the starting point of the acceptance, and a and n are the coefficient and power associated with a time-dependent term. By definition, this function models the acceptance as starting at zero at $t = t_0$, smoothly increasing and approaching unity as $t \to \infty$.

The decay time acceptance is determined on simulated $B^0_s \to \mu^+\mu^-$ samples, and the effect of data-simulation differences on the acceptance was investigated in variables that can affect the effective lifetime selection, including the BDT. Although most variables were found to be well reproduced, the isolation variables included in the BDT were not. This disagreement is remedied by reweighting the nTracks variable, defined as the number of reconstructed tracks in an event, in $B^0 \to K^+\pi^-$ simulation to match $B^0 \to K^+\pi^-$ background-subtracted data. As the decay time acceptance varies for the datasets taken in different years, it is determined on a

cocktail of simulated samples, one per year. The relative weight of each sample is determined by the relative yields of $B_s^0 \to J/\psi\,\phi$ decays in each dataset before applying any PID or BDT requirements multiplied by the relative efficiency of the $B_s^0 \to \mu^+\mu^-$ PID and BDT selection, and divided by the relative size of each simulated sample.

5.2.4 Systematics

The main systematic uncertainties on the lifetime measurement are assigned for the accuracy of the lifetime fit, the decay time acceptance, the lack of modeling of exclusive backgrounds in the mass fit, and the decay time model. The systematic uncertainties and the total systematic uncertainty on the lifetime measurement are summarised in Table 5.1. Compared to the expected statistical uncertainty, the systematic uncertainty is close to negligible.

The accuracy of the lifetime fit is investigated with 10,000 toy experiments, and the measured lifetime is found to deviate from the true lifetime by 0.03 ps on average. This deviation is conservatively assigned as a systematic uncertainty.

The systematic uncertainy on the decay time acceptance is estimated by applying the analysis procedure to $B^0 \to K^+\pi^-$ data, including determining the acceptance parameters on reweighted $B^0 \to K^+\pi^-$ simulated samples. The $B^0 \to K^+\pi^-$ lifetime was measured to be

$$\tau(B^0 \to K^+\pi^-) = 1.52 \pm 0.03\text{ps}, \tag{5.5}$$

which is consistent with the PDG value of $\tau(B^0) = 1.520 \pm 0.004$ ps. To be conservative, the statistical uncertainty on this check is assigned as a systematic uncertainty.

In the mass fit to obtain the background-subtracted $B_s^0 \to \mu^+\mu^-$ sample, only shapes for the $B_s^0 \to \mu^+\mu^-$ signal and the combinatorial background are included. As a systematic check, toys are generated with all exclusive backgrounds included. The main exclusive backgrounds are $B_{(s)}^0 \to h^+h'^-$ and $\Lambda_b^0 \to p\mu^-\nu$, which contribute with 0.9 and 0.6 expected events respectively. The contribution from $B^0 \to \mu^+\mu^-$ is generated with 0.2 (0.7) events according to the Standard Model (world average) branching fraction for $B^0 \to \mu^+\mu^-$. Relative to the nominal fit, the effective lifetime measurement is found to be biased by 0.007 ps due to the presence of exclusive backgrounds. This bias is assigned as a systematic uncertainty.

The final systematic is assigned for the decay time model in the lifetime fit. The nominal decay time model is a single exponential, which is then combined with the

Systematic source	Systematic uncertainty on $\tau_{\text{eff}}^{\mu^+\mu^-}$
Limited size of data sample	0.03 ps
Decay time acceptance ($B^0 \to K^+\pi^-$ check)	0.03 ps
Exclusive backgrounds	0.007 ps
Mix of eigenstates	0.018 ps
Total	0.05 ps

Table 5.1: The different sources of systematic uncertainty and the size of each contribution.

lifetime acceptance. A bias may arise when $A_{\Delta\Gamma_s}^{\mu^+\mu^-} \neq \pm 1$, when both the heavy and light B_s^0 eigenstates contribute and the proper decay time model would be two exponentials. The decay time acceptance could favour one of the contributions over the other and thus bias the effective lifetime estimate. The bias is estimated by generating $B_s^0 \to \mu^+\mu^-$ candidates with $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = 0$, which is the scenario where the heavy and light B_s^0 eigenstates contribute equally. A total of two million candidate lifetimes are generated, one million with and one million without including the decay time acceptance. Both lifetime distributions are fitted to with the nominal model with or without the acceptance, as appropriate for the sample. The difference between the two lifetime measurements is 0.018 ps, which is assigned as a systematic uncertainty.

5.2.5 Results

The data used for the effective lifetime measurement and the results of the maximum likelihood fits to determine the *sWeights* and effective lifetime are shown in Figure 5.5

The effective lifetime of $B_s^0 \to \mu^+\mu^-$ is measured for the first time and found to be

$$\tau_{\text{eff}}^{\mu^+\mu^-} = 2.04 \pm 0.44 \pm 0.05\text{ps}, \tag{5.6}$$

where the uncertainties are statistical and systematic in nature, which shows the measurement uncertainty is dominated by statistics.

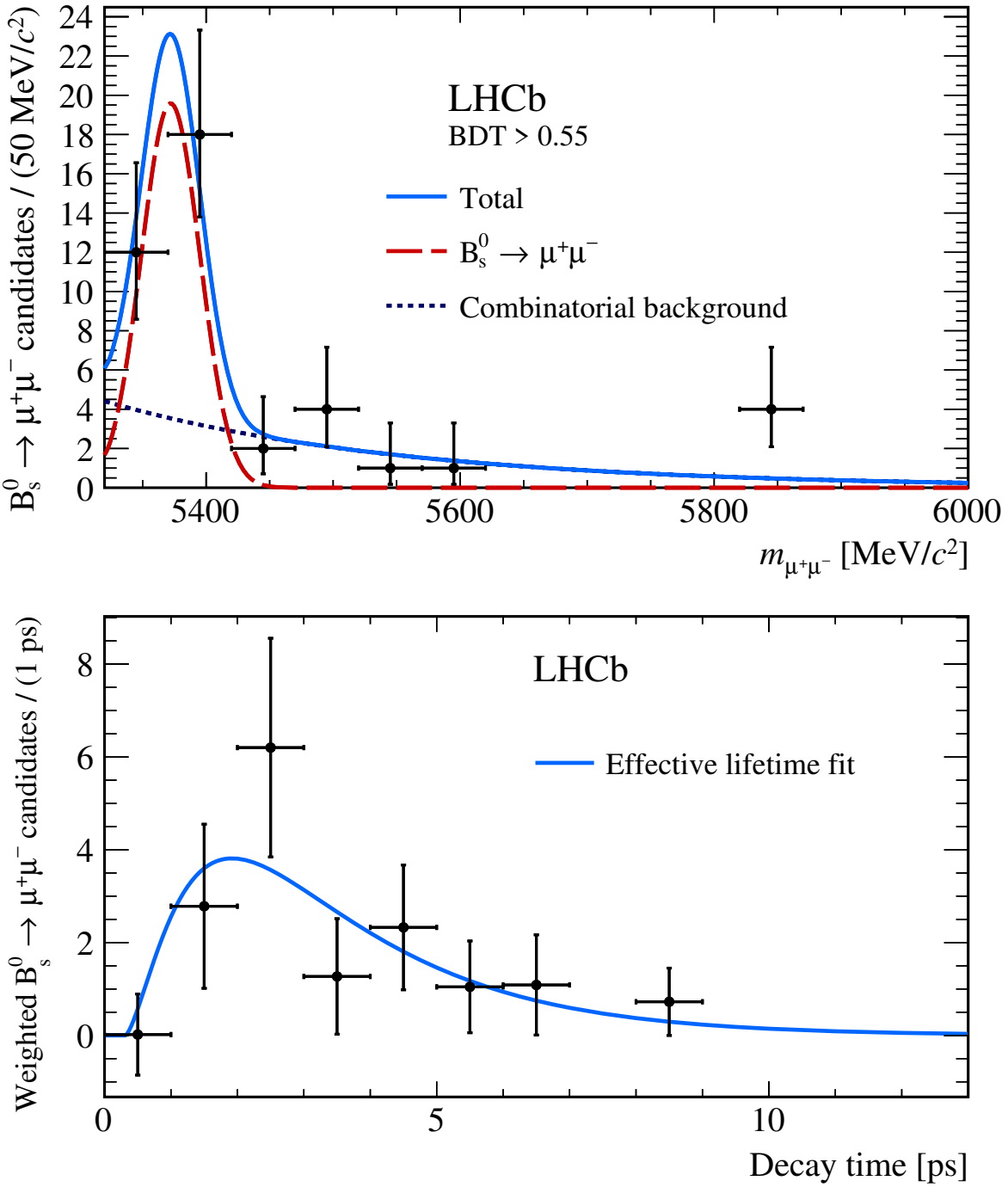

Figure 5.5: The two fits used to measure the effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay. The mass fit is used to subtract the background from the lifetime distribution. The fit to the background-subtracted lifetime distribution of $B_s^0 \to \mu^+\mu^-$ candidates is used to extract the effective lifetime.

To interpret the result of this effective lifetime measurement, the formula from Equation 2.39 is rewritten to

$$A_{\Delta\Gamma_s}^{\mu^+\mu^-} = \frac{1}{y_s}\frac{(1-{y_s}^2)\tau_{\text{eff}}^{\mu^+\mu^-} - (1+{y_s}^2)\tau_{B_s^0}}{2\tau_{B_s^0} - (1-y_s)\tau_{\text{eff}}^{\mu^+\mu^-}}, \tag{5.7}$$

where $y_s = \frac{\Delta\Gamma_s}{2\Gamma_s}$, half of the lifetime difference between the two lifetime eigenstates of the B_s^0 meson, and $\tau_{B_s^0} = \frac{1}{\Gamma_s}$ is the B_s^0 meson lifetime. Using the measured value of $\tau_{\text{eff}}^{\mu^+\mu^-}$, we find

$$A_{\Delta\Gamma_s}^{\mu^+\mu^-} = 8.2^{+19.3}_{-7.3}, \tag{5.8}$$

where uncertainty is very asymmetric because the formula for $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ is non-linear in $\tau_{\text{eff}}^{\mu^+\mu^-}$.

The measurement is consistent with the SM value for $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$, $+1$, at the level of 1 σ. The physically allowed range for $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$ is $[-1, 1]$, and the most extreme non-SM hypothesis, $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = -1$, is consistent with the data at the level of 1.4 σ. Therefore, this measurement does not yet put strong constraints on $A_{\Delta\Gamma_s}^{\mu^+\mu^-}$. More data is needed for a future analysis to obtain a more constraining measurement that could exclude parts of the physically allowed range.

5.3 Experimental correlation of branching fraction and effective lifetime

In this Section, the experimental correlation between the measurements of the branching fraction and effective lifetime of $B_s^0 \to \mu^+\mu^-$ is discussed.

In the current analysis, the branching fraction and effective lifetime are measured separately. However, the efficiency of detecting $B_s^0 \to \mu^+\mu^-$ decays depends on the lifetime of the $B_s^0 \to \mu^+\mu^-$ decay, which means that the $B_s^0 \to \mu^+\mu^-$ branching fraction measurement depends on the $B_s^0 \to \mu^+\mu^-$ effective lifetime, and in principle even has some sensitivity to it through the dependence of the BDT distribution for $B_s^0 \to \mu^+\mu^-$ on the effective lifetime.

In Section 5.1, the branching fraction of $B_s^0 \to \mu^+\mu^-$ is interpreted with $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = +1$, which is the value it takes in the Standard Model. In Section 5.2.5, the effective lifetime has been measured above this value by one standard deviation, outside of the physical range. Therefore, it is important to see how the branching fraction of $B_s^0 \to \mu^+\mu^-$ would be interpreted under the measured effective lifetime.

For the branching fraction analysis, the efficiency of $B_s^0 \to \mu^+\mu^-$ has been estimated for different hypotheses of $A_{\Delta\Gamma}$, namely -1, 0 and $+1$, by determining the BDT distribution and normalisation for each of these hypotheses, as described in Section 4.3.6, and fitting to the data with each hypothesis. Compared to the nominal

fit, the hypotheses of $A_{\Delta\Gamma} = -1$ and $A_{\Delta\Gamma} = 0$ increase the estimated branching fraction by 10.9% and 4.6% respectively.

For a future measurement with more precision, the correlation between the branching fraction and the lifetime should be taken into account more carefully. Ideally, the lifetime dependence of the BDT shape and normalisation can be parametrised as function of the lifetime. In that case, two approaches are possible to directly obtain a two-dimensional constraint on the branching fraction and effective lifetime of $B_s^0 \to \mu^+\mu^-$:

1. By performing a three-dimensional fit to the mass and lifetime in BDT bins, only considering a part of the sample for the lifetime fit in order to have to model only the signal and combinatorial background lifetime distributions. Such a fit would allow to determine a two-dimensional likelihood for the branching fraction and effective lifetime of $B_s^0 \to \mu^+\mu^-$.

2. As the effective lifetime fit uses information that is almost independent of the information in the branching fraction fit, it can be performed separately. The branching fraction fit can then be performed as described in this thesis, adding the effective lifetime as a parameter that is constrained to the value from the effective lifetime fit by adding a Gaussian term to likelihood. The branching fraction fit would then still allow to obtain the two-dimensional likelihood.

Chapter 6

Discussion and outlook

In this Chapter, the results of the $B^0_{(s)} \to \mu^+\mu^-$ analysis presented in this thesis are interpreted in the context of other measurements that probe the $b \to q\ell^+\ell^-$ transisition. The prospects for future measurements with $B^0_{(s)} \to \mu^+\mu^-$ decays are also discussed, including possible experimental improvements with the LHCb Upgrade and Upgrade-2 detectors. With the Upgrade-2 detector, the measurement of CP violation in $B^0_s \to \mu^+\mu^-$ decays will be possible, which will require information on the flavour of the B^0_s meson at production and a different analysis strategy.

6.1 Interpretation of current measurements

6.1.1 Status of $B^0_{(s)} \to \mu^+\mu^-$ measurements at the LHC

Since the publication of the measurement described in this thesis, ATLAS and CMS have also published measurements of $B^0_{(s)} \to \mu^+\mu^-$ decays, using Run 2 data, up to and including 2016.

The ATLAS Collaboration, using data up to 2016, reports [89]

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}, \tag{6.1}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (-1.9 \pm 1.6) \times 10^{-10}. \tag{6.2}$$

ATLAS measured a $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ value below the SM prediction, found a negative central value for the $B^0 \to \mu^+\mu^-$ branching fraction, and found both measurements to be strongly correlated because the B^0_s and B^0 separation in mass, around 87 MeV/c^2, is close to the dimuon mass resolution. The two-dimensional likelihood for the ATLAS measurement is shown in Figure 6.1.

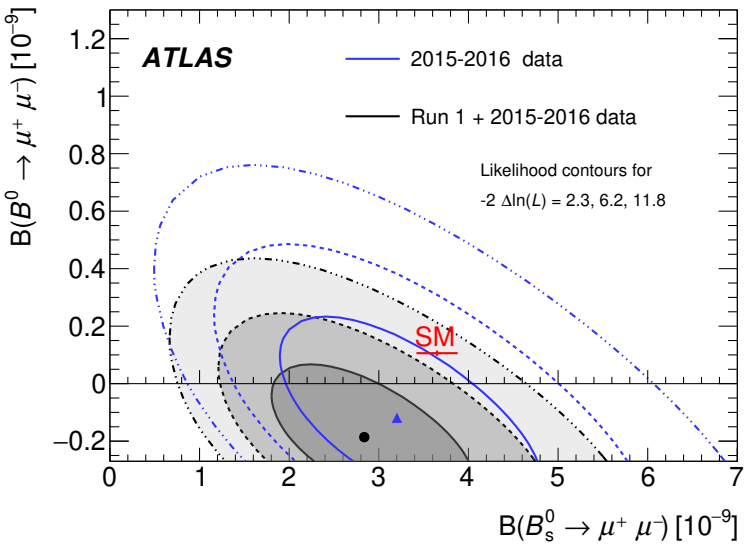

Figure 6.1: The result of the ATLAS measurement using data up to 2016 in the two-dimensional $\mathcal{B}(B_s^0 \to \mu^+\mu^-) - \mathcal{B}(B^0 \to \mu^+\mu^-)$-plane. The Standard Model prediction, including uncertainties, is indicated by the red cross. The central value from the measurement is shown as the black circle, and the profile likelihood contours corresponding to a confidence level associated with a 1,2,3,...σ deviation are shown in black. The measurement with Run 2 data only is shown in blue.

The CMS Collaboration, also using data up to 2016, reports [90]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.9^{+0.7}_{-0.6} \pm 0.2) \times 10^{-9}, \tag{6.3}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (0.8^{+1.4}_{-1.3}) \times 10^{-10}. \tag{6.4}$$

Both measurements by CMS are fully consistent with the SM and with the measurement of LHCb. The two-dimensional likelihood for the CMS measurement is shown in Figure 6.2 and shows a lower degree of correlation between the two measurements due to the fact that the mass resolution of the CMS experiment is better.

Recently, a combined measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions by LHCb, ATLAS and CMS has been published [91], by combining the two-dimensional likelihoods and taking into account the dominant and shared systematic of the value of the fragmentation fraction f_s/f_d, The current experimental averages for the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are found to be

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}, \tag{6.5}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (0.58^{+0.76}_{-0.72}) \times 10^{-10}, \tag{6.6}$$

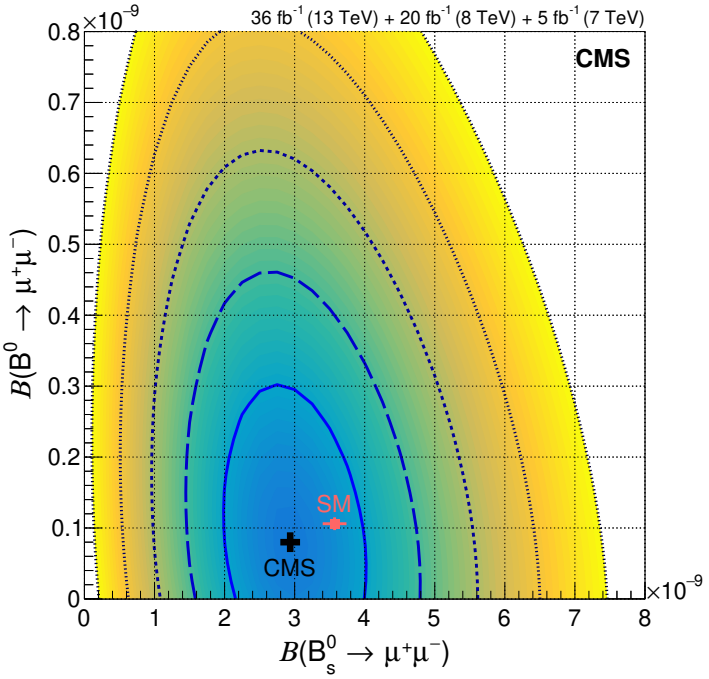

Figure 6.2: The result of the CMS measurement using data up to 2016 in the two-dimensional $\mathcal{B}(B_s^0 \to \mu^+\mu^-) - \mathcal{B}(B^0 \to \mu^+\mu^-)$-plane. The Standard Model prediction, including uncertainties, is indicated by the red cross. The central value from the measurement is shown as a black cross, and the profile likelihood contours corresponding to a confidence level associated with a 1,2,3,...σ deviation are shown by the blue lines.

The ratio of these measurements to the SM predictions in Equations 2.55, 2.56, is

$$R_{B_s^0 \to \mu^+\mu^-} = 0.75 \pm 0.10, \tag{6.7}$$

$$R_{B^0 \to \mu^+\mu^-} = 0.57^{+0.75}_{-0.71}, \tag{6.8}$$

where the uncertainties on both ratios are dominated by the experimental uncertainty. The result for $B_s^0 \to \mu^+\mu^-$ agrees with the SM prediction at 2.5 σ.

The CMS Collaboration also measured the effective lifetime of $B_s^0 \to \mu^+\mu^-$ and found

$$\tau_{\text{eff}}^{\mu^+\mu^-} = (1.70^{+0.61}_{-0.44})\ \text{ps}, \tag{6.9}$$

which is consistent with the SM and the LHCb measurement. Converting the asymmetric uncertainty on the CMS result into a symmetric uncertainty by averaging the two bounds and averaging with the LHCb result, the current experimental average

of the effective lifetime is

$$\tau_{\text{eff}} = (1.90 \pm 0.34)\ \text{ps}, \tag{6.10}$$

which is consistent with the case of a fully CP-odd decay(SM) at 0.8 σ and the case of a fully CP-even decay(non-SM) of 1.4 σ. Clearly, more data is needed in order to reduce the statistical uncertainty on the effective lifetime and significantly constrain New Physics.

Since the $B_s^0 \to \mu^+\mu^-$ decay has now been firmly established, it is interesting to examine the current status of measurements with other $b \to s\ell\ell$ decays and to compare with the measured value of the $B_s^0 \to \mu^+\mu^-$ branching fraction.

6.1.2 Status of $b \to s\ell\ell$ measurements and lepton-non universality

Rare leptonic B decays of the form $B_{(s)}^0 \to \ell^+\ell^-$ proceed through the quark-level $b \to q\ell^+\ell^-$ transition. The same transition mediates semileptonic decays of the form $X_b \to X_q\ell^+\ell^-$, where X_q is a hadron containing a q quark, as discussed in Section 2.2.2. Interestingly, the measurements of various observables in semileptonic decays currently show deviations from the SM predictions.

As there are many available modes and measurements, global fits to the data have been developed by multiple groups to interpret the results in terms of effective Wilson coefficients [92, 93]. The most recent global fit, from March 2020 [93], combines branching fraction measurements of $B^0 \to K^{*0}\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$, $B_s^0 \to \phi\mu^+\mu^-$, $B_s^0 \to \mu^+\mu^-$, $B \to X_s\mu^+\mu^-$, $B^0 \to K^{*0}\gamma$, $B_s^0 \to \phi\gamma$, angular measurements of $B^0 \to K^{*0}\mu^+\mu^-$, $B_s^0 \to \phi\mu^+\mu^-$, $B^0 \to K^{*0}e^+e^-$, and lepton universality ratios of branching fractions for $B^0 \to K^{*0}\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$. It assumes that pseudoscalar and scalar contributions to the $b \to s\ell\ell$ transition remain negligible, as in the SM. Under these assumptions, the $B_s^0 \to \mu^+\mu^-$ branching fraction is a strong probe of $\mathcal{C}_{10}^{\mu}$ and the angular measurements using $B^0 \to K^{*0}\mu^+\mu^-$ decays put the strongest constraints on $\mathcal{C}_9^{\mu}$.

First, a fit is performed to determine the Wilson coefficients $\mathcal{C}_7^{\mu(\prime)}, \mathcal{C}_9^{\mu(\prime)}, \mathcal{C}_{10}^{\mu(\prime)}$ in 6D, assuming that New Physics only contributes to the $b \to s\mu^+\mu^-$ transition. The p-value for the SM hypothesis given the data is 1.4%, and the pull of the fit result compared to the SM is 5.8 σ, mostly driven by a modification of $\mathcal{C}_9$.

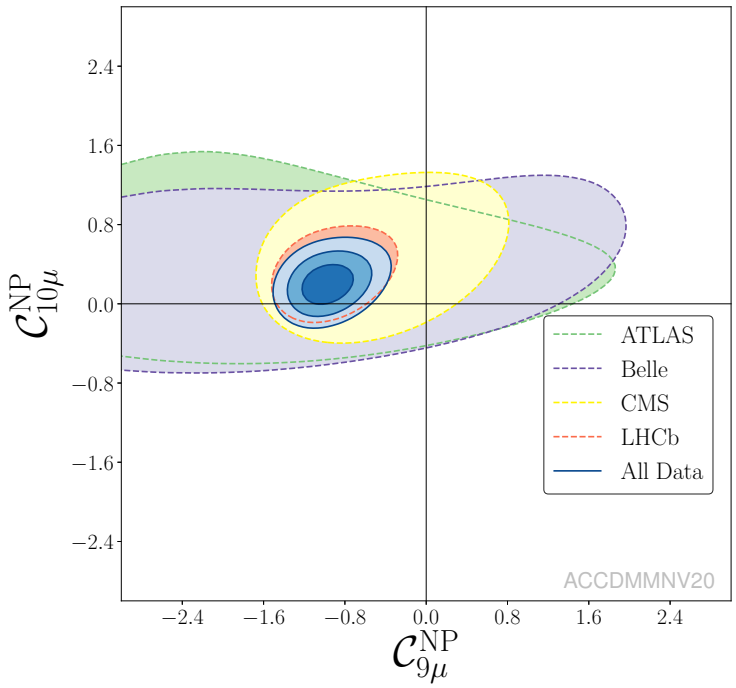

Figure 6.3: Constraints on New Physics contributions to $\mathcal{C}_9$ and $\mathcal{C}_{10}$ in the $b \to s\mu^+\mu^-$ transition from a fit to $b \to s\ell\ell$ data, assuming both are real and that other Wilson coefficients are as in the SM. Individual constraints from ATLAS, Belle, CMS and LHCb are shown with their contours at three standard deviations. Based on a figure from Ref. [18].

Given this indication of possible New Physics contributions, different combinations of coeffcients are tested. They are distinguished by which specific Wilson coefficients are fitted, specific relations between them (for example $\delta\mathcal{C}^{\mu}_{9} = -\delta\mathcal{C}^{\mu}_{10}$), and whether or not lepton flavour universal contributions ($\delta\mathcal{C}^{U}_{i}$), *i.e.* contributions to all three $b \to s\ell\ell$ transitions, appear as well.

Depending on the tested scenario, the magnitude of $\mathcal{C}^{\mu}_{10}$ in $b \to s\mu^+\mu^-$ transitions remains unmodified or is reduced by up to 11.6%, relative to its Standard Model value $\mathcal{C}^{\mu}_{10} \sim -4.31$. This corresponds to a reduction of the $B^0_s \to \mu^+\mu^-$ branching fraction by up to $\sim -22\%$. Indeed, the current experimental average (from Equation 6.8), namely $R_{B^0_s \to \mu^+\mu^-} = 0.75^{+0.13}_{-0.10}$, would favour some modification of $\mathcal{C}^{\mu}_{10}$. A more precise measurement of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ will be essential to distinguish different scenarios.

A simple scenario is a 2D fit to $\mathcal{C}^{\mu}_{9}$ and $\mathcal{C}^{\mu}_{10}$ in $b \to s\mu^+\mu^-$ transitions. The result of such a fit is shown in Figure 6.3. It shows that New Physics contributions to $\mathcal{C}^{\mu}_{9}$ are favoured by the current data, while new contributions to $\mathcal{C}^{\mu}_{10}$ are possible but not required.

On the other hand, the scenario that is the most favoured by current data contains a flavour universality breaking contribution, $\delta\mathcal{C}_9^\mu = -\delta\mathcal{C}_{10}^\mu = -0.34$, in $b \to s\mu^+\mu^-$ transitions, and a universal contribution $\delta\mathcal{C}_9^U = -0.80$, contributing to all $b \to s\ell\ell$ transitions. Such a scenario can be linked to lepton flavour anomalies in tree-level semileptonic decays, which also deviate from SM expectations.

The lepton flavour universality ratios in tree-level semileptonic decays compare $b \to c\tau\overline{\nu}_\tau$ to $b \to c\ell\overline{\nu}_\ell$ decays, with $\ell \in [e, \mu]$, and are found to deviate by around 3 standard deviations [24]. At first sight, these deviations indicate new physics at a far lower energy scale than those found in $b \to s\ell\ell$ decays. However, it is possible to connect the measurements in $b \to c\ell\overline{\nu}_\ell$ and $b \to s\ell\ell$ transitions, by matching the effective field theory at the b-hadron mass scale (as used in this thesis, associated with the Wilson coefficients $\mathcal{C}_i^{(\prime)}$ for $b \to s\ell\ell$ transitions), to an effective field theory at the weak scale (referred to as Standard Model Effective Theory or SMEFT). After doing so, it is found that a common effective interaction at the weak scale can explain the deviations observed in $b \to s\ell\ell$ and in $b \to c\ell\overline{\nu}_\ell$ decays in a consistent way, as shown in Figure 6.4 [92]. The particle mediating this effective interaction would be a U_1 vector leptoquark, which directly couples leptons and quarks, that would mainly couple to the third generation.

This result implies a suppression of the branching fraction of $B_s^0 \to \mu^+\mu^-$ by about 20%, similar to what is found in global fits to only $b \to s\ell\ell$ measurements. Notably, it also implies tauonic decays such as $B_s^0 \to \tau^+\tau^-$ to be enhanced by two to three orders of magnitude, up to $\sim 7 \times 10^{-4}$, compared to the experimental limit of $< 6.8 \times 10^{-3}$. It also predicts sizeable contributions to Lepton Flavour Violating decays such as $B^+ \to K^+\tau^\pm\mu^\mp$ and $B_s^0 \to \tau^\mp\mu^\pm$, enhancing them up to the 10^{-5} level, close to experimental limits, which are at the level of $3 - 5 \times 10^{-5}$. Specifically, $B_s^0 \to \tau^\mp\mu^\pm$ is very sensitive to new scalar contributions and is already constraining the allowed phase space for New Physics contributions.

Using the observed deviations to Wilson coefficients in the SMEFT framework, theorists can postulate new symmetries at higher energies and investigate the consequences. One such example is a new flavour symmetry, $\mathrm{U}(2)_q \times \mathrm{U}(2)_\ell$, which is explored in Ref. [94]. In this model, $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is modified both by changes in $\mathcal{C}_{10}^\mu$ and by new scalar contributions. As a result, the deviations observed in $R_{K^{(*)}}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are related to each other and moderated by a theory parameter sensitive to new scalar contributions, as shown in Figure 6.5. Similar to the SMEFT analysis, in this model decays such as $B_s^0 \to \tau^+\tau^-$ are expected to be strongly

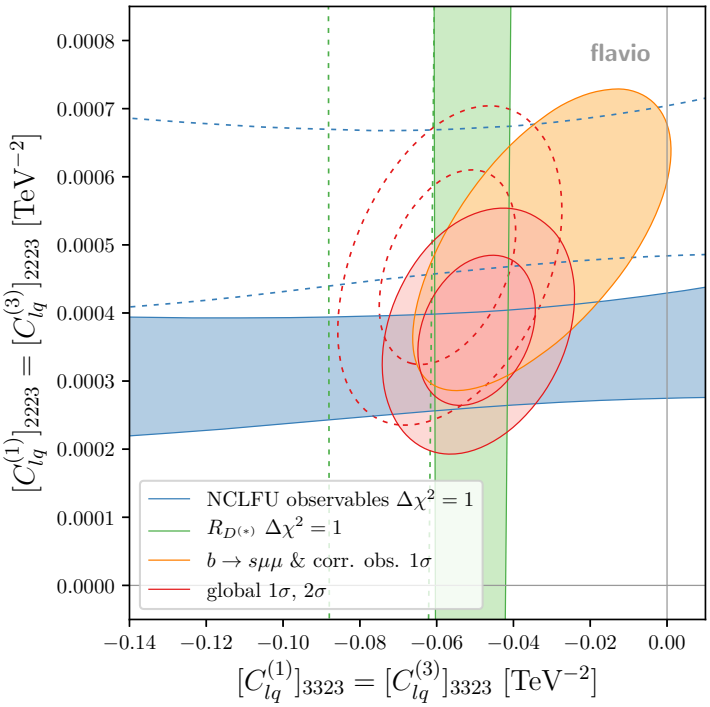

Figure 6.4: Constraints on effective Wilson coefficients in SMEFT with a fit performed to $b \to s\ell\ell$ and $b \to c\ell\overline{\nu}_\ell$ data. The Wilson coefficients $[C_lq^{(}1,3)]_{3323}$ introduce a lepton flavour non-universal contribution to $b \to c\tau\overline{\nu}_\tau$ and $b \to s\tau^+\tau^-$ transitions, the second of which induces a universal contribution to $b \to s\ell\ell$ transitions. The Wilson coefficients $[C_lq^{(}1,3)]_{2223}$ introduce a lepton flavour non-universal contribution to $b \to s\mu^+\mu^-$ transitions. Together, these two contributions provide a consistent explanation of the three types of measurements (LFU measurements in $b \to s\ell\ell$, $b \to c\ell\overline{\nu}_\ell$, and other measurements with $b \to s\mu^+\mu^-$ decays). Figure taken from Ref. [92].

enhanced and sizeable contributions to $B_s^0 \to \tau^\mp\mu^\pm$ and $B^+ \to K^+\tau^\pm\mu^\mp$ are expected.

It is interesting to consider the case of a UV-complete theory responsible for the observed effective couplings. An example of such a theory is the so-called PS^3 model, which is completely flavour non-universal at high energies and only appears flavour universal at lower energies [95]. Such a theory does not only explain the lepton universality anomalies but also the hierarchy of Standard Model Yukawa couplings: in this model they are both caused by the flavour symmetry breaking. For these models, $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ plays a critical role by probing the New Physics contribution to the Wilson coefficient $\mathcal{C}_{10}^\mu$ and simultaneously constraining new scalar contributions, caused by right-handed mixing in the lepton sector.

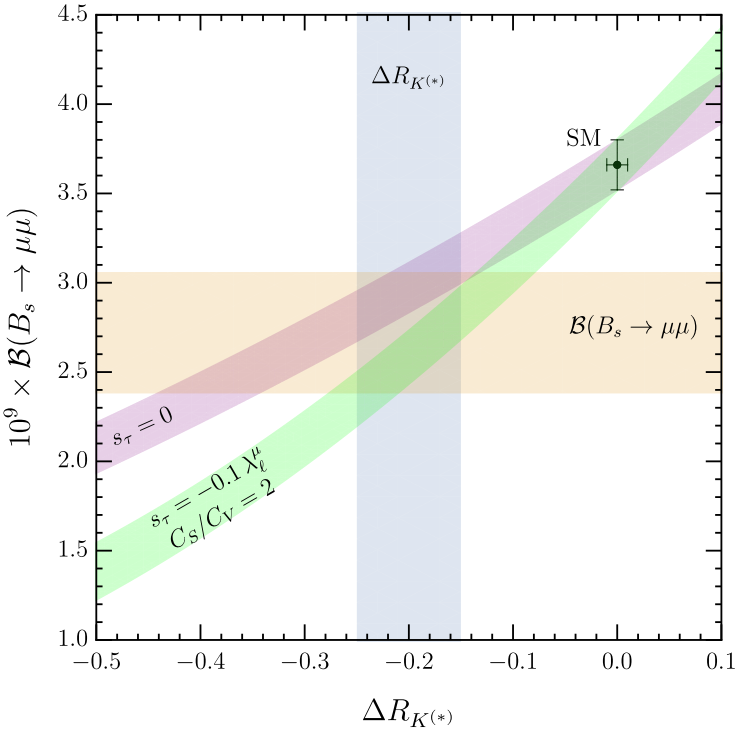

Figure 6.5: The relation between deviations in $R_{K^{(*)}}$ (which are tests of lepton flavour universality in semileptonic $b \to s\ell\ell$ decays, $R_{K^{(*)}} = \mathcal{B}(B \to K^{(*)}\mu^+\mu^-)/\mathcal{B}(B \to K^{(*)}e^+e^-)$) and the value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in the case of a new flavour symmetry. C_S/C_V is the relative contribution from scalar and vector new physics in the model; s_t and λ_ℓ^μ are parameters describing the flavour structure of New Physics contributions. Figure taken from Ref. [94].

Future measurements of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and other observables in $B_{(s)}^0 \to \ell^+\ell^-$ decays will be essential to further constrain the contributions from scalar New Physics in general and these kinds of models specifically. Experimentally, the prospects for stronger constraints look bright.

6.2 Prospects for future measurements

As the $B^0_{(s)} \to \mu^+\mu^-$ measurements presented in this thesis are statistically limited, it is interesting to discuss the prospects for future $B^0_{(s)} \to \mu^+\mu^-$ measurements with more data taken by the LHCb detector. Three datasets are considered to estimate future constraints on $B^0_{(s)} \to \mu^+\mu^-$: the current dataset, the dataset that will be obtained by the LHCb Upgrade detector, and the dataset that will be obtained by the LHCb Upgrade 2 detector.

The current dataset includes data from Run 1 and from Run 2 taken from 2015 to 2018, collecting a total of $6\,\text{fb}^{-1}$ at a centre-of-mass energy of 13 TeV. Currently, the LHCb detector is being upgraded, in order to take data without a hardware trigger and with an instantaneous luminosity around 5 times higher than before. It will take data during Run 3 and 4 of the LHC until around 2030, in order to collect $41\,\text{fb}^{-1}$ in total with this upgraded detector. Beyond that, a Upgrade 2 of the LHCb detector is being planned, increasing the instantaneous luminosity even further and aiming to collect $250\,\text{fb}^{-1}$ more, finishing data taking around 2038.

The LHCb Upgrade and Upgrade 2 aim at a performance for each triggered event similar to the LHCb detector in Run 2. Additionally, both detectors will benefit from operating with a pure software trigger and therefore save more interesting events. As the hardware trigger efficiency for $B^0_{(s)} \to \mu^+\mu^-$ decays was already good, a conservative prospect is that the gain in sensitivity for the $B^0_{(s)} \to \mu^+\mu^-$ analysis for the LHCb Upgrade and Upgrade 2 detectors will scale linearly with the increased integrated luminosity and possible increase in center of mass energy.

The relative size of each anticipated dataset, compared to the one used in this thesis, and the expected number of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates detected in each sample are shown in Table 6.1. By increasing the sample size two times by an order of magnitude, the measurements performed with future datasets can explore BSM physics far more deeply, effectively to very high mass scales of new particles.

$B^0_s \to \mu^+\mu^-$ branching fraction Since the $B^0_s \to \mu^+\mu^-$ decay has been observed, the aim of measuring $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is now to achieve a precision measurement to probe New Physics. Given the current flavour anomalies, a consistent modification of the $B^0_s \to \mu^+\mu^-$ branching fraction relative to the SM prediction is about 20%. A discovery of such a scenario by this measurement alone by five standard deviations thus requires a precision on the $B^0_s \to \mu^+\mu^-$ branching fraction of around 4%. The

Table 6.1: The current datasets and expectations for future datasets for the LHCb collaboration, along with the expected yields for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays.

	This thesis	Run 1+2	Run 3+4	Run 5+6
Total $\int \mathcal{L}$ (fb^{-1})	4.4	9	50	300
Average $\sqrt{s}$ (TeV)	9.4	11.2	13.5	13.9
Relative sample size	1.0	2.5	16.4	101
$N_{B^0_s \to \mu^+\mu^-,\mathrm{SM}}$	32	80	530	3300
$N_{B^0 \to \mu^+\mu^-,\mathrm{SM}}$	3.5	9	60	360

statistical uncertainty on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ with the LHCb Upgrade dataset is within this reach at around 4.3%, whereas it will be surpassed by Upgrade 2 and be improved to around 1.7%.

However, the current systematic uncertainty of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is larger than 4%, and will have to be reduced to fully exploit the LHCb Upgrade and Upgrade 2 datasets. The systematic uncertainty is dominated by the uncertainty on the relative production fraction of B^0_s and B^0 mesons, f_s/f_d, which is 5.8% for Run 1 and 4.3% for Run 2 data. The Run 2 estimate is dominated by the uncertainty on an external input, namely the D^+_s meson branching fraction, which is 3.3% and can be measured more precisely at charm factories. Sub-dominant uncertainties include the branching fractions of the normalisation channels at around 3%, which will be improved by the Belle 2 collaboration, and the particle identification efficiency at 2%, which can be improved with dedicated studies.

An important aspect of reducing the systematic uncertainty depends on the usage of the calibration channels $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$. Differences can occur between hadrons and muons in the trigger, particle identification and reconstruction and should be minimised. The selection procedure of the calibration channels should therefore be as similar as possible to the $B^0_s \to \mu^+\mu^-$ selection. The pure software trigger will be important in this regard, as it should enable to reduce discrepancies between the $B^0_{(s)} \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ selection and significantly increase the $B^0 \to K^+\pi^-$ sample size. Additionally, the smaller the differences between simulation and data, the systematics caused by differences in selection procedure. Large, high precision simulation samples will thus be essential.

Discovery of $B^0 \to \mu^+\mu^-$ The $B^0 \to \mu^+\mu^-$ decay has not been observed yet and its branching fraction has been shown to be not far above its SM expectation

value. The full Run 1+2 dataset can possibly provide evidence for this decay, after which the Upgrade dataset will allow to observe this decay, with around 60 expected candidates in the sample, and the Upgrade 2 dataset will put the minimal flavour violation hypothesis to the test, with an expected statistical uncertainty of around 5-10% on the ratio of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions.

Effective lifetime of $B_s^0 \to \mu^+\mu^-$ The first measurement of the effective lifetime of $B_s^0 \to \mu^+\mu^-$ decays is described in this thesis. It is clearly statistically limited and has a very small systematic uncertainty of around 3%. This makes it an interesting observable for the Upgrade and Upgrade 2 detectors. For the effective lifetime, two interesting scenarios correspond to $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = +1$ and $A_{\Delta\Gamma_s}^{\mu^+\mu^-} =$ -1, namely a fully CP-odd (SM) decay and a fully CP-even (NP) decay. These scenarios are favoured by effective field theory considerations and the fact that the measured value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is close to the SM prediction, as discussed in Ref. [45]: either new pseudo-scalar and scalar contributions do not exist, or they cancel each other in the branching fraction but modify the effective lifetime of $B_s^0 \to \mu^+\mu^-$. In the NP scenario, the lifetime is modified by 12% with respect to the SM.

The measurement with the full Run 1+2 data is expected to have an uncertainty of around 12% and will therefore not exclude either scenario yet. With the Upgrade dataset, the uncertainty should be reduced to around 4%, allowing to exclude $A_{\Delta\Gamma_s}^{\mu^+\mu^-} < -0.09$ at 95% CL if $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = +1$ is measured, or $A_{\Delta\Gamma_s}^{\mu^+\mu^-} > 0.09$ at 95% CL in case $A_{\Delta\Gamma_s}^{\mu^+\mu^-} = -1$ is measured. With the Upgrade 2 dataset, the uncertainty should be reduced to around 1.7%, and either it will be possible to claim an observation of New Physics from the effective lifetime alone (in the NP scenario) or to exclude $A_{\Delta\Gamma_s}^{\mu^+\mu^-} < 0.54$, so most non-SM scenarios, at 95% CL. The measurement of the effective lifetime will thus strongly constrain new (pseudo)-scalar contributions.

While the current systematic uncertainty is around 3%, larger than future statistical uncertainties, it is dominated by systematics cause by the limited size of the data and calibration samples as well as the simple lifetime model with a single exponential. Both will be reduced by the larger sample size of future datasets, when including the full lifetime model with two exponentials will be feasible, such that the future systematic uncertainty is expected to remain negligible compared to the statistical uncertainty.

6.2.1 CP violation in $B^0_{(s)} \to \mu^+\mu^-$ decays

With the Upgrade 2 dataset, it will become possible to search for CP violation in $B^0_s \to \mu^+\mu^-$ decays. CP violation measurements require measuring a CP-asymmetry, which means that the rate of a matter and anti-matter process have to be separated. As $B^0_{(s)} \to \mu^+\mu^-$ and $\overline{B}^0_{(s)} \to \mu^+\mu^-$ have the same final state, the only way to separate these rates is by determining the state of the $B^0_{(s)}$ candidate at production. At hadron colliders, this state is determined by reconstructing hadrons that participated in the $B^0_{(s)}$ hadronisation, a procedure called flavour tagging. Flavour tagging uses either the decay products of the other B hadron that was produced in the *pp* collision for an "opposite side" tag, or the flavour of the hadron that is produced in fragmentation together with the spectator quark in the $B^0_{(s)}$ meson, which means the "same side" hadron contains a d or s quark. Combining all of the tagging estimates, the effective efficiency of flavour tagging at LHCb for $B^0_s \to J/\psi\,\phi$, which should be similar to the efficiency for $B^0_s \to \mu^+\mu^-$, is around 4.7%. The moderate effective efficiency of flavour tagging in the LHCb environment implies that a large dataset is needed to be sensitive to CP violation in $B^0_s \to \mu^+\mu^-$ decays.

With the Upgrade dataset, an effective sample of around 26 $B^0_s \to \mu^+\mu^-$ tagged events might be collected, which would imply an uncertainty on $S_{\mu^+\mu^-}$ of around 40%. The Upgrade 2 dataset would allow a more stringent test with an effective sample of around 100 $B^0_s \to \mu^+\mu^-$ tagged events, and an implied uncertainty on $S_{\mu^+\mu^-}$ of around 16%. Such a precision would allow to strongly constrain contributions to $B^0_{(s)} \to \mu^+\mu^-$ with *CP*-violating phases and establish possible new (pseudo-)scalar contributions [96]. Which such a statistical uncertainty, systematic uncertainties are expected to be negligible.

The very rare quantum loop decay process $B^0_s \to \mu^+\mu^-$ has been discovered. It is now becoming a tool to subject the SM to a very stringent test through various precision measurements and will play a decisive role in distinguishing possible New Physics scenarios.

Appendix A

Commonly used PDFs for mass shapes of B decays

In this Appendix, the PDFs used to describe the invariant mass shapes of different fit components fits are described for reference. The three PDFs used for the results in this thesis (apart from the Crystal Ball shape used for $B^0_{(s)} \to \mu^+\mu^-$ signal, which is described in Section 4.4) are the Double-Sided Crystal Ball, Argus and RooKeys PDFs.

A.1 Double-Sided Crystal Ball

The Double-Sided Crystal Ball function is an extension of the Crystal Ball function described in Section 4.4, Equation 4.8. It includes an enhanced tail on both sides, according to

$$f(m|\mu,\sigma,\alpha_1,n_1,\alpha_2,n_2) = N \begin{cases} \left(\frac{n_1}{|\alpha_1|}\right)^{n_1} \exp[-\frac{\alpha_1^2}{2}](\frac{m-\mu}{\sigma} + \frac{n_1}{|\alpha_1|} - |\alpha_1|)^{-n_1}, & \text{if } \frac{(m-\mu)}{\sigma} < -\alpha_1 \\ \left(\frac{n_2}{|\alpha_2|}\right)^{n_1} \exp[-\frac{\alpha_2^2}{2}](\frac{m-\mu}{\sigma} + \frac{n_2}{|\alpha_2|} - |\alpha_2|)^{-n_1}, & \text{if } \frac{(m-\mu)}{\sigma} > +\alpha_2 \\ \exp[-\frac{(m-\mu)^2}{2\sigma^2}], & \text{otherwise} \end{cases} \tag{A.1}$$

It is used to describe mass shapes for signal samples with larger statistics, for which a smaller right-hand tail in the mass shape from material interactions needs to be included, or for samples where the mass of an intermediate particle is used as constraint in the calculation of the invariant mass, for which the radiative tail of the intermediate particle is overcorrected, causing a significant right-hand tail.

A.2 Argus

The Argus function is used to describe mass shapes with a fixed endpoint m_0 because of a kinematic limit, for example partially reconstructed backgrounds. It is given by

$$f(m|m_0, p, c) = Nm \left[1 - \left(\frac{m}{m_0}\right)^2\right]^p \times \exp\left[c(1 - \left(\frac{m}{m_0}\right)^2)\right]. \tag{A.2}$$

In this analysis, $p = 1/2$ is used and c and m_0 are left free to fit.

A.3 RooKeys

A RooKeys distribution is obtained by using one-dimensional kernel estimation to model arbitrary input datasets [75], resulting in a non-parametric, fixed shape that can be used in any fit.

Bibliography

[1] J. J. Thomson, *Xl. cathode rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44**, 293–316 (1897), eprint: https://doi.org/10.1080/14786449708621070.

[2] E. Rutherford, *Bakerian lecture: nuclear consitution of atoms*, Proc. R. Soc. Lond. A **97** (1920).

[3] C. D. Anderson, *The positive electron*, Phys. Rev. **43**, 491–494 (1933).

[4] J. Chadwick, *The existence of a neutron*, Proc. R. Soc. Lond. A **136**, 692–708 (1932), eprint: http://doi.org/10.1098/rspa.1932.0112.

[5] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2**, 1285–1292 (1970).

[6] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13**, 508–509 (1964).

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13**, 321–323 (1964).

[8] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental test of parity conservation in beta decay*, Phys. Rev. **105**, 1413–1415 (1957).

[9] L. Wolfenstein, *Parametrization of the kobayashi-maskawa matrix*, Phys. Rev. Lett. **51**, 1945–1947 (1983).

[10] CKMfitter group, J. Charles, A. Höcker, H. Lacker, S. Laplace, F. R. Le Diberder, J. Malclès, J. Ocariz, M. Pivk, and L. Roos, *CP Violation and the ckm matrix: assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41**, 1 (2005), arXiv:hep-ph/0406184, updated results and plots available at http://ckmfitter.in2p3.fr/.

[11] Troitsk Collaboration, V. N. Aseev *et al.*, *An upper limit on electron antineutrino mass from Troitsk experiment*, Phys. Rev. **D84**, 112003 (2011), arXiv:1108.5034.

[12] T. Inami and C. S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_L \to \mu\mu$, $K^+ \to \pi + \nu\nu$ and $K^0 \leftrightarrow \bar{K}^0$*, Progress of Theoretical Physics **65**, 297–314 (1981), eprint: `http://oup.prod.sis.lan/ptp/article-pdf/65/1/297/5252099/65-1-297.pdf`.

[13] S. Fajfer and N. Košnik, *Vector leptoquark resolution of R_K and $R_{D^{(*)}}$ puzzles*, Phys. Lett. **B755**, 270–274 (2016), arXiv:`1511.06024`.

[14] R. Fleischer, *CP violation in the B system and relations to $K \to \pi\nu\bar{\nu}$ decays*, Phys. Rept. **370**, 537–680 (2002), arXiv:`hep-ph/0207108`.

[15] L. Silvestrini, *Effective Theories for Quark Flavour Physics*, in Les Houches summer school: EFT in Particle Physics and Cosmology Les Houches, Chamonix Valley, France, July 3-28, 2017 (2019), arXiv:`1905.00798`.

[16] P. Böer, T. Feldmann, and D. van Dyk, *Angular Analysis of the Decay $\Lambda_b \to \Lambda(\to N\pi)\ell^+\ell^-$*, JHEP **01**, 155 (2015), arXiv:`1410.2115`.

[17] X.-Q. Li, J. Lu, and A. Pich, *$B^0_{s,d} \to \ell^+\ell^-$ Decays in the Aligned Two-Higgs-Doublet Model*, JHEP **06**, 022 (2014), arXiv:`1404.5865`.

[18] W. Altmannshofer, C. Niehoff, P. Stangl, and D. M. Straub, *Status of the $B \to K^*\mu^+\mu^-$ anomaly after Moriond 2017*, Eur. Phys. J. **C77**, 377 (2017), arXiv:`1703.09189`.

[19] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *The $b \to sl^+l^-$ Anomalies And Their Implications For New Physics*, in Proceedings, 51st Rencontres de Moriond on Electroweak Interactions and Unified Theories: La Thuile, Italy, March 12-19, 2016 (2016), pp. 31–36, arXiv:`1605.06059`.

[20] J. C. Ward, *An identity in quantum electrodynamics*, Phys. Rev. **78**, 182–182 (1950).

[21] R. J. Knegjens, *Strategies to Hunt for New Physics with Strange Beauty Mesons*, PhD thesis (Vrije U., Amsterdam, 2014).

[22] A. Bazavov *et al.*, *B- and D-meson leptonic decay constants from four-flavor lattice QCD*, Phys. Rev. **D98**, 074512 (2018), arXiv:`1712.09262`.

[23] C.-K. Chua, W.-S. Hou, and C.-H. Shen, *Long-Distance Contribution to $\Delta\Gamma s/\Gamma s$ of the $B_s - \bar{B}_s$ System*, Phys. Rev. **D84**, 074037 (2011), arXiv:`1107.4325`.

[24] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, Eur. Phys. J. **C77**, 895 (2017), arXiv:`1612 . 07233`, updated results and plots available at `http://www.slac.stanford.edu/xorg/hflav/`.

[25] D. King, A. Lenz, and T. Rauh, *B_s mixing observables and $|V_{td}/V_{ts}|$ from sum rules*, JHEP **05**, 034 (2019), arXiv:1904.00940.

[26] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, and N. Tuning, *Branching Ratio Measurements of B_s Decays*, Phys. Rev. **D86**, 014027 (2012), arXiv:1204.1735.

[27] F. Halzen and A. Martin, *Quarks & leptons: an introductory course in modern particle physics* (John Wiley & Sons, New York, USA, 1984).

[28] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109**, 041801 (2012), arXiv:1204.1737.

[29] A. J. Buras and R. Fleischer, *Quark mixing, CP violation and rare decays after the top quark discovery*, Adv. Ser. Direct. High Energy Phys. **15**, 65–238 (1998), arXiv:hep-ph/9704376, [,65(1997)].

[30] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for $BR(B_{s,d}^0 \to \mu^+\mu^-)$*, Eur. Phys. J. **C72**, 2172 (2012), arXiv:1208.0934.

[31] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45**, 97–107 (2006), arXiv:hep-ph/0506026.

[32] A. J. Buras, *Relations between $\Delta M(s, d)$ and $B(s, d) \to \mu\bar{\mu}$ in models with minimal flavor violation*, Phys. Lett. **B566**, 115–119 (2003), arXiv:hep-ph/0303060.

[33] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, Nucl. Phys. **B645**, 155–187 (2002), arXiv:hep-ph/0207036.

[34] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou, and M. Steinhauser, *$B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112**, 101801 (2014), arXiv:1311.0903.

[35] M. Beneke, C. Bobeth, and R. Szafron, *Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \to \mu^+\mu^-$*, Phys. Rev. Lett. **120**, 011801 (2018), arXiv:1708.09152.

[36] Particle Data Group, M. T. et al., *Review of particle physics*, Phys. Rev. D **98**, 030001 (2018).

[37] C. G. Boyd, B. Grinstein, and R. F. Lebed, *Constraints on form factors for exclusive semileptonic heavy to light meson decays*, Phys. Rev. Lett. **74**, 4603–4606 (1995).

[38] Belle Collaboration, E. Waheed *et al.*, *Measurement of the CKM matrix element $|V_{cb}|$ from $B^0 \to D^{*-}\ell^+\nu_\ell$ at Belle*, Phys. Rev. **D100**, 052007 (2019), arXiv:1809.03290.

[39] Flavour Lattice Averaging Group, S. Aoki *et al.*, *FLAG Review 2019*, (2019), arXiv:1902.08191.

[40] R. J. Dowdall, C. T. H. Davies, R. R. Horgan, G. P. Lepage, C. J. Monahan, J. Shigemitsu, and M. Wingate, *Neutral B-meson mixing from full lattice QCD at the physical point*, (2019), arXiv:1907.01025.

[41] CMS and LHCb collaborations, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522**, 68 (2015), arXiv:1411.4413.

[42] ATLAS Collaboration, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, Eur. Phys. J. **C76**, 513 (2016), arXiv:1604.04263.

[43] L.-S. Geng, B. Grinstein, S. Jäger, J. Martin Camalich, X.-L. Ren, and R.-X. Shi, *Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays*, Phys. Rev. **D96**, 093006 (2017), arXiv:1704.05446.

[44] F. Beaujean, C. Bobeth, and S. Jahn, *Constraints on tensor and scalar couplings from $B \to K\bar{\mu}\mu$ and $B_s \to \bar{\mu}\mu$*, Eur. Phys. J. **C75**, 456 (2015), arXiv:1508.01526.

[45] W. Altmannshofer, C. Niehoff, and D. M. Straub, *$B_s \to \mu^+\mu^-$ as current and future probe of new physics*, JHEP **05**, 076 (2017), arXiv:1702.05498.

[46] P. H. Chankowski, J. Kalinowski, Z. Was, and M. Worek, *CP violation in $B^0_{(s)} \to \tau^+\tau^-$ decays*, Nucl. Phys. **B713**, 555–574 (2005), arXiv:hep-ph/0412253.

[47] B. Capdevila, A. Crivellin, S. Descotes-Genon, L. Hofer, and J. Matias, *Searching for New Physics with $b \to s\tau^+\tau^-$ processes*, Phys. Rev. Lett. **120**, 181802 (2018), arXiv:1712.01919.

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

[49] L. Evans and P. Bryant, *LHC Machine*, JINST **3**, S08001 (2008).

[50] ATLAS Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716**, 1–29 (2012), arXiv:1207.7214.

[51] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716**, 30–61 (2012), arXiv:1207.7235.

[52] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3**, S08005 (2008).

[53] *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*, July 2013, LHCb-CONF-2013-011.

[54] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10**, P02007 (2015), arXiv:1408.1251.

[55] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9**, P09007 (2014), arXiv:1405.7808.

[56] A. Affolder *et al.*, *Radiation damage in the LHCb vertex locator*, JINST **8**, P08002 (2013), arXiv:1302.5259.

[57] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9**, P01002 (2014), arXiv:1311.3893.

[58] *LHCb outer tracker: Technical Design Report*, Geneva, 2001.

[59] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30**, 1530022 (2015), arXiv:1412.6352.

[60] D. van Eijk *et al.*, *Radiation hardness of the LHCb Outer Tracker*, Nucl. Instrum. Meth. **A685**, 62–69 (2012).

[61] LHCb RICH, J. He, *Real-time calibration and alignment of the LHCb RICH detectors*, Nucl. Instrum. Meth. **A876**, 13–16 (2017), arXiv:1611.00296.

[62] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73**, 2431 (2013), arXiv:1211.6759.

[63] *LHCb calorimeters: Technical Design Report*, Geneva, 2000.

[64] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8**, P02022 (2013), arXiv:1211.1346.

[65] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8**, P04022 (2013), arXiv:1211.3055.

[66] LHCb Collaboration, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14**, P04013 (2019), arXiv:1812.10790.

[67] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118**, 191801 (2017), arXiv:1703.05747.

[68] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis (Institute of Nuclear Physics, Krakow, 1986), DESY-F31-86-02.

[69] F. Archilli, M. Palutan, M. Rama, A. Sarti, B. Sciascia, and R. Vazquez Gomez, *Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization*, tech. rep. LHCb-INT-2014-047. CERN-LHCb-INT-2014-047 (CERN, Geneva, Nov. 2014).

[70] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences **55**, 119–139 (1997).

[71] CDF Collaboration, T. Aaltonen *et al.*, *Observation of New Charmless Decays of Bottom Hadrons*, Phys. Rev. Lett. **103**, 031801 (2009), arXiv:0812.4271.

[72] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555**, 356–369 (2005), arXiv:physics/0402083.

[73] C.Adrover et al., *Search for the $B^0_S \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with 3 fb^{-1} at LHCb*, LHCb-ANA-2013-032.

[74] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764**, 150–155 (2014), arXiv:1312.5000.

[75] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136**, 198–207 (2001), arXiv:hep-ex/0011057.

[76] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039 (CERN, Geneva, May 2014).

[77] L. Anderlini, A. Contu, C. R. Jones, S. S. Malde, D. Muller, S. Ogilvy, J. M. Otalora Goicochea, A. Pearce, I. Polyakov, W. Qian, B. Sciascia, R. Vazquez Gomez, and Y. Zhang, *The PIDCalib package*, LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021 (CERN, Geneva, July 2016).

[78] LHCb collaboration, R. Aaij *et al.*, *Measurement of f_s/f_u variation with proton-proton collision energy and B-meson kinematics*, Phys. Rev. Lett. **124**, 122002 (2020), arXiv:1910.09934.

[79] J. M. Flynn, T. Izubuchi, T. Kawanai, C. Lehner, A. Soni, R. S. Van de Water, and O. Witzel, *$B \to \pi\ell\nu$ 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*, Phys. Rev. **D91**, 074510 (2015), arXiv:1501.05373.

[80] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, (2016), arXiv:1607.00299.

[81] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in 7 TeV pp collisions*, Phys. Rev. **D85**, 032008 (2012), arXiv:1111.2357.

[82] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $B^+ \to \pi^+\mu^+\mu^-$*, JHEP **12**, 125 (2012), arXiv:1210.2645.

[83] 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*, Phys. Rev. **D86**, 114025 (2012), arXiv:1207.0265.

[84] A. Pellegrino, *Statistical data analysis (lecture notes)*, Sept. 2014.

[85] G. Cowan, *Statistical data analysis* (Oxford University Press, Oxford, United Kingdom, 1998).

[86] J. Neyman and E. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society, A **231**, 289–337 (1933).

[87] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71**, 1554 (2011), arXiv:1007.1727, [Erratum: Eur. Phys. J.C73,2501(2013)].

[88] A. L. Read, *Presentation of search results: the CLs technique*, Journal of Physics G: Nuclear and Particle Physics **28**, 2693–2704 (2002).

[89] ATLAS Collaboration, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04**, 098 (2019), arXiv:1812.03017.

[90] CMS Collaboration, *Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B_0 \to \mu^+\mu^-$ with the CMS experiment*, tech. rep. CMS-PAS-BPH-16-004 (CERN, Geneva, 2019).

[91] L. Collaboration, *Combination of the ATLAS, CMS and LHCb results on the $B_{(s)}^0 \to \mu^+\mu^-$ decays*, (2020).

[92] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, and D. M. Straub, *B-decay discrepancies after Moriond 2019*, (2019), arXiv:1903.10434.

[93] M. Algueró, B. Capdevila, A. Crivellin, S. Descotes-Genon, P. Masjuan, J. Matias, M. Novoa, and J. Virto, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79**, 714 (2019), arXiv:1903.09578.

[94] J. Fuentes-Martín, G. Isidori, J. Pagès, and K. Yamamoto, *With or without U(2)? Probing non-standard flavor and helicity structures in semileptonic B decays*, (2019), arXiv:1909.02519.

[95] M. Bordone, C. Cornella, J. Fuentes-Martín, and G. Isidori, *Low-energy signatures of the* PS^3 *model: from B-physics anomalies to LFV*, JHEP **10**, 148 (2018), arXiv:1805.09328.

[96] R. Fleischer, D. G. Espinosa, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi, *CP Violation in Leptonic Rare* B_s^0 *Decays as a Probe of New Physics*, Eur. Phys. J. **C78**, 1 (2018), arXiv:1709.04735.

Summary

Everything that we observe in our universe is made out of fundamental particles with different masses and charges and whose behaviour is predicted by the Standard Model of particle physics. Each of these particles is associated with an equivalent anti-particle: a particle that has the same mass but opposite charge for every interaction (interactions are commonly referred to as forces). Importantly, there are two types of matter: quarks and leptons. Quarks are particles that are affected by the strong interaction; leptons (such as electrons) are not. The strong interaction causes quarks to quickly form particles composed of configurations that are stable for the strong interaction; well-known examples are the protons and neutrons, the two particles found in atomic nuclei. The particles studied for this thesis, referred to as B_s^0 and B^0, are another type of composite particle, called mesons; particles consisting of a quark and an anti-quark. Specifically, they contain a beauty anti-quark ($\bar{b}$) and respectively a strange(s) or a down(d) quark. B_s^0 and B^0 particles are unstable and decay through the weak interaction after only around 1.5×10^{-12} seconds (*i.e.* around a millionth of a millionth of a second) or 1.5 picoseconds (ps) in short.

This thesis describes the measurement of the properties of decays of $B_{(s)}^0$ mesons to two leptons, specifically to a muon (μm) and an anti-muon (μ^+), referred to as $B_{(s)}^0 \to \mu^+\mu^-$ decays in short; those properties are the branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays and the effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay. These decays are extremely rare processes, mediated by the weak interaction; however, they are predicted very precisely in the Standard Model. They are expected to occur only around three times out of every billion B_s^0 decays, $(3.58 \pm 0.13) \times 10^{-9}$ and one time out of every ten billion B^0 decays, $(1.020 \pm^{+0.029}_{-0.055}) \times 10^{-10}$. We refer to these fractions as the branching fractions of $B_{(s)}^0 \to \mu^+\mu^-$ decays. The branching fractions of $B_{(s)}^0 \to \mu^+\mu^-$ decays are so small in the Standard Model because these decays can not proceed through first-order tree diagrams, only through higher-order quantum loop diagrams. Additionally, the $B_{(s)}^0 \to \mu^+\mu^-$ decay is affected by helicity suppression; the weak interaction prefers a specific spin configuration of the muons

that is not possible for the $B^0_{(s)} \to \mu^+\mu^-$ decay (because it is a pseudoscalar particle decaying to two fermions).

An essential property of $B^0_{(s)}$ mesons is that they can transform into their own anti-particle ($\overline{B}^0_{(s)}$) through another kind of quantum loop diagram. As a result, $B^0_{(s)}$ mesons are in a quantum superposition with $\overline{B}^0_{(s)}$ mesons: they propagate as a combination of the particle and anti-particle at the same time. It turns out there are two possible kinds of combinations, which are called eigenstates. For B^0_s mesons specifically, these eigenstates have lifetimes (the time before the meson decays) that differ by around 10%. Notably, only the long-lifetime eigenstate decays to a $\mu^+\mu^-$ pair according to the Standard Model, corresponding to an expected lifetime of 1.620 ps. By measuring the average decay time (referred to as effective lifetime) of the $B^0_s \to \mu^+\mu^-$ decay, this prediction can be tested.

Why is it interesting to measure the properties of $B^0_{(s)} \to \mu^+\mu^-$ decays? Precisely because these decays are so extremely rare in the Standard Model. Because of the uncertainty principle from quantum physics, $B^0_{(s)}$ decays are affected by particles heavier than the $B^0_{(s)}$ mesons themselves: they can indirectly probe the existence of heavy new particles. As $B^0_{(s)} \to \mu^+\mu^-$ decays are strongly suppressed, even particles much heavier than all known particles could significantly affect the branching fraction. Specifically, new scalar or pseudoscalar particles would not be affected by the helicity suppression affecting the Standard Model process, making $B^0_{(s)} \to \mu^+\mu^-$ decays very sensitive to such particles. The effective lifetime of $B^0_s \to \mu^+\mu^-$ decays provides an additional observable, which is specifically sensitive to detect scalar and pseudoscalar contributions that could interfere destructively and, therefore, remain hidden in the $B^0_s \to \mu^+\mu^-$ branching fraction. Unsurprisingly, $B^0_{(s)} \to \mu^+\mu^-$ decays have been of interest for a long time. The study described in this thesis follows over 30 years of searches for $B^0_{(s)} \to \mu^+\mu^-$ decays, as shown in Figure 1.

How can we study $B^0_{(s)} \to \mu^+\mu^-$ decays? As B^0_s and B^0 mesons decay very quickly, a dedicated facility is needed to produce them. In this case, the production facility is the Large Hadron Collider (LHC), a 27 km long ring located about 100 meters underground close to Geneva (a hadron is any kind of stable state under the strong interaction). The LHC uses the mass-energy equivalence first stated by Einstein, $E = mc^2$, to produce $B^0_{(s)}$ mesons. It accelerates protons up to incredible energies, giving them an energy almost 7000 times their mass. By colliding those protons, all this energy gets released and can be used to produce particles of masses much higher than the original protons, including $B^0_{(s)}$ mesons. Discoverying a very rare decay such as $B^0_{(s)} \to \mu^+\mu^-$ requires billions of $B^0_{(s)}$ mesons to be produced; at LHCb,

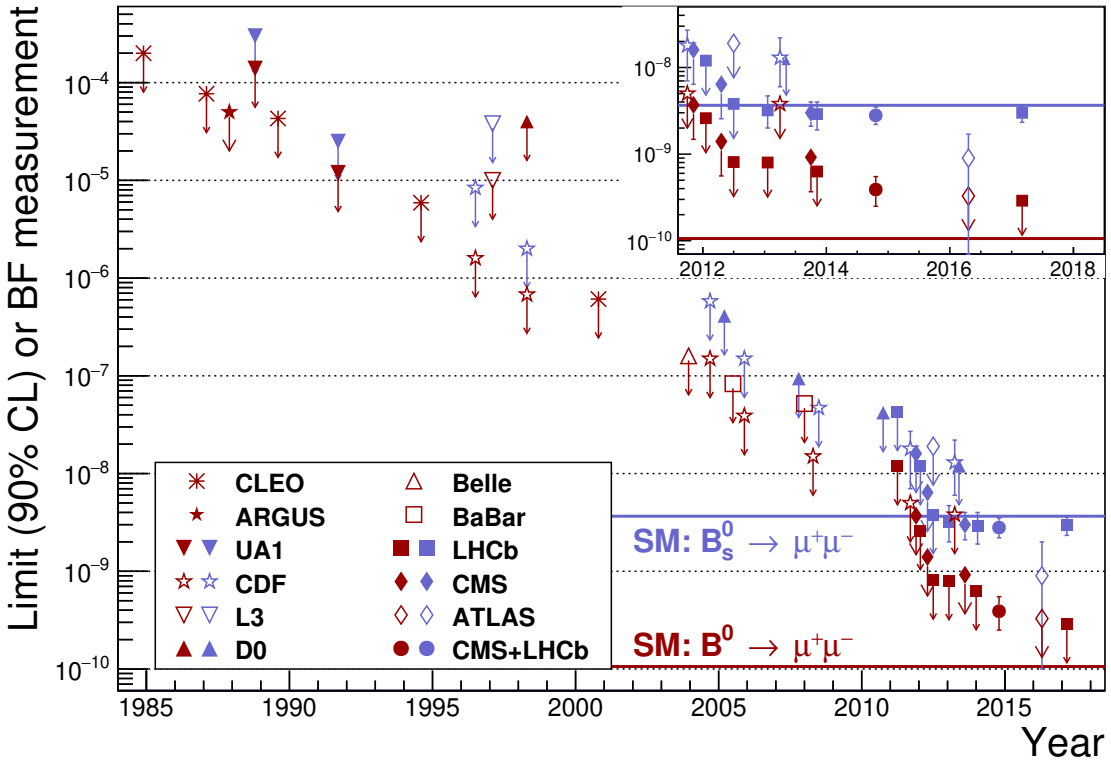

Figure 1: An overview of the results of searches for and measurements of the branching fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays, including the measurement described in this thesis.

around 60.000 $B^0_{(s)}$ mesons were produced per second during the operations running from 2015 to 2018.

LHCb is the detector that was used to detect $B^0_{(s)}$ mesons and collect the data used for this thesis. The LHCb detector is located in the forward direction, close to the axis of the proton-proton collisions, as many of the $B^0_{(s)}$ mesons are produced there. Because they travel at 98% of the speed of light and the time dilation the experience, $B^0_{(s)}$ mesons produced in the LHCb detector decay after travelling around 1 cm. Therefore, any $B^0_{(s)}$ decay is identified by using its decay products; in case of $B^0_{(s)} \to \mu^+\mu^-$ decays, these are the muon and anti-muon. An example event containing a signal-like $B^0_s \to \mu^+\mu^-$ candidate is shown in Figure 2.

The muons are electrically charged particles, which means that they can be detected by passing through a suitable material, for example silicon: they leave "tracks". The LHCb detector contains such materials so that it can detect muons from their origin in the $B^0_{(s)}$ meson decay and follow them through the detector. The detector around the proton-proton collision, the Vertex Locator, can determine the $B^0_{(s)}$ meson decay location by combining the information from the two muon

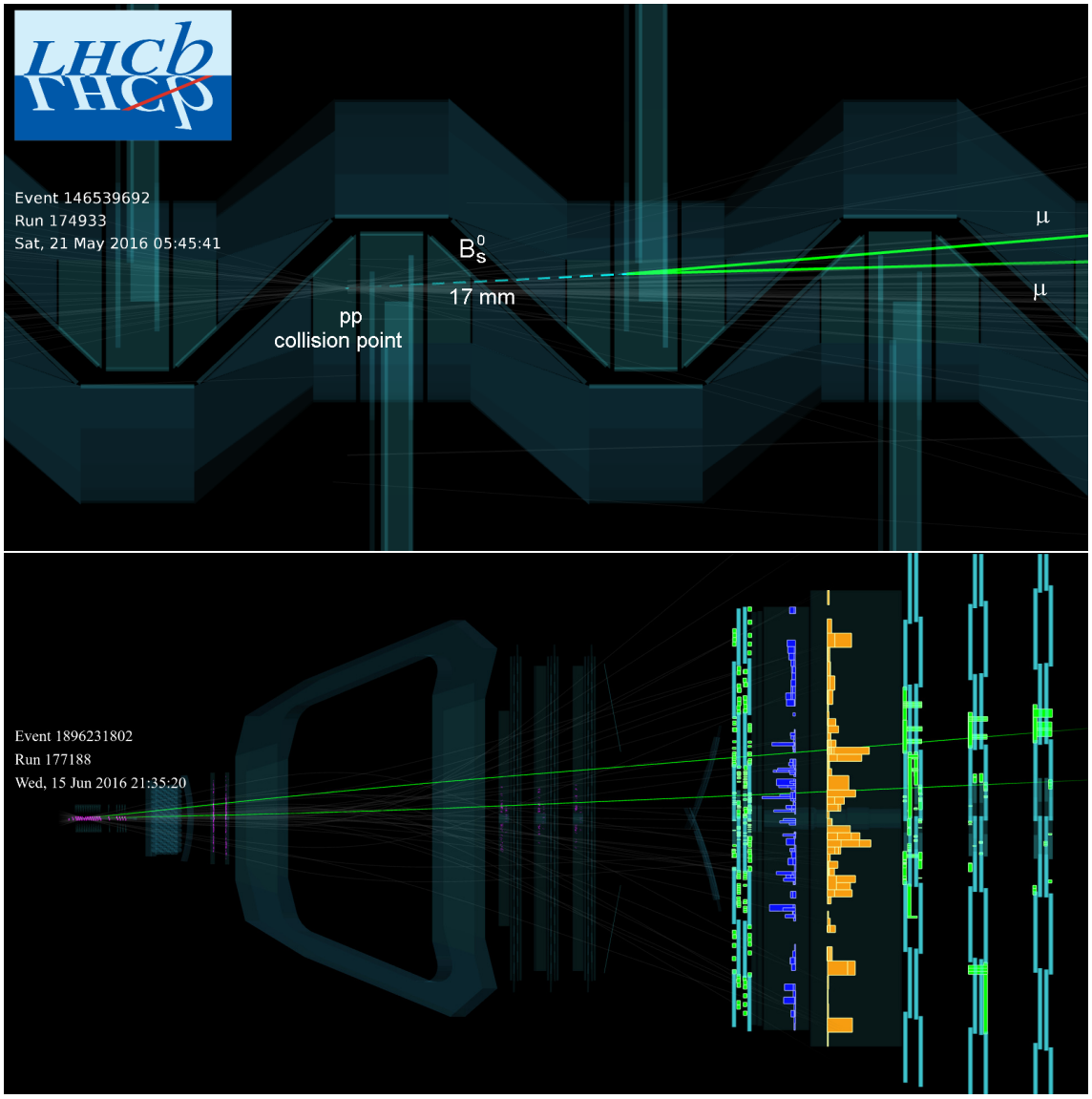

Figure 2: An example event containing a signal-like $B_s^0 \to \mu^+\mu^-$ candidate. Muon tracks are shown in green. Top: a close up of the event in the Vertex Locator, showing the decay of the B_s^0 meson after 17 mm. Bottom: the event in the full detector, showing the magnet bending the muons and the muons reaching the muon detectors at the end of the detector.

tracks, allowing to determine the $B^0_{(s)}$ meson decay time as required to measure the effective lifetime. To determine the momentum and charge of the two muons, their trajectory is affected by a strong magnet within LHCb; the more the magnet affects the trajectory, the smaller their momentum was. As a result, it can be determined whether the two muons were a muon and anti-muon (according to their charge) and whether the mass of the particle that decayed to those muons is consistent with the B_s^0 or B^0 mass. An important property of muons in the LHC environment is that they are least affected by interactions with materials, while all other charged

particles are stopped in other parts of the LHCb detector. Therefore, at the end of the LHCb detector a final set of tracking stations interlaced with iron slabs are located to detect the telltale sign of a muon: a track that persists through the full detector.

After reconstructing suitable $B^0_{(s)} \to \mu^+\mu^-$ candidates, it is essential to reject the many background candidates that also populate the data. The two main backgrounds are referred to as combinatorial background or exclusive backgrounds. Combinatorial background comes from combining a muon and anti-muon that actually came from two different decays of b-hadrons; it is very abundant, occuring about ten million times more often than $B^0_{(s)} \to \mu^+\mu^-$. It is rejected through a machine learning classifier, namely a Boosted Decision Tree (BDT). The main way the BDT identifies combinatorial background is by searching for other tracks that were produced in the b-hadron decay and can be associated with the muons, which are absent for true $B^0_{(s)} \to \mu^+\mu^-$ decays. Exclusive backgrounds originate mostly from $B^0_{(s)}$ meson decays that decay to two charged particles, which seem to be two muons (although one or both might actually be a hadron, such as a proton). Information from all of the components of the LHCb detector are combined by another machine learning classifier, specifically a neural network, such that only a few in a thousand hadrons are misidentified as muons.

The $B^0_{(s)} \to \mu^+\mu^-$ signal is then identified by using the BDT variable to divide the sample in separate ones that are more or less signal-like and using the reconstructed $B^0_{(s)}$ candidate mass which peaks around the B^0_s and B^0 mass for $B^0_{(s)} \to \mu^+\mu^-$ decays. The precise calibration of the signal and estimate of the remaining background are an essential part of the $B^0_{(s)} \to \mu^+\mu^-$ analysis. Similar decays are used to determine how many $B^0_{(s)}$ mesons were produced in the analysed dataset. Those are decays of B^+ mesons to a J/ψ particle and a charged kaon, K^+, where the J/ψ particle decays to two muons, just like the signal, and decays of $B^0_{(s)}$ mesons to two hadrons, $B^0_{(s)} \to h^+h'^-$, which have a similar spatial structure to $B^0_{(s)} \to \mu^+\mu^-$. Therefore, $B^0_{(s)} \to h^+h'^-$ decays look similar to $B^0_{(s)} \to \mu^+\mu^-$ decays with respect to the mass reconstruction and the BDT variable and are used to calibrate those two aspects for signal using the data. An important factor that is required to calibrate the normalisation for B^0_s mesons is their production rate relative to B^0 or B^+ mesons, referred to as f_s/f_d; it has been measured in other analyses and is used an input to the $B^0_s \to \mu^+\mu^-$ branching fraction measurement.

In the dataset recorded with the LHCb detector from 2011 to 2016, and using the strategy just described, the most signal-like $B^0_{(s)} \to \mu^+\mu^-$ candidates are shown

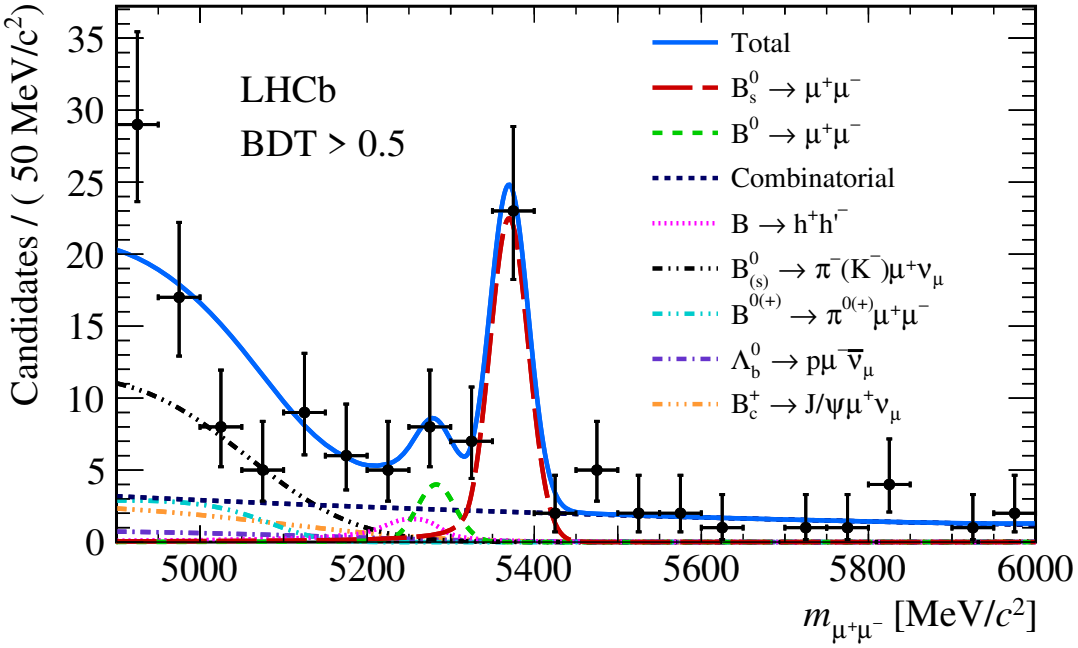

Figure 3: The mass distribution for the most signal-like $B^0_{(s)} \to \mu^+\mu^-$ candidates in the data sample used for this thesis. The result of the fit to the data is overlaid with their different components identified in the legend.

in Figure 3, including the result from a fit of the model describing the behaviour of signal and backgrounds in the dataset after the signal selection. The $B^0_s \to \mu^+\mu^-$ decay is observed for the first time by a single experiment; there is no conclusive sign of the $B^0 \to \mu^+\mu^-$ decay.

From the fit to the data candidates, the branching fractions for both decays are found to be $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}$; where the first uncertainty comes from the limited sample size of the data and the second from limitations of the analysis procedure. For the $B^0_s \to \mu^+\mu^-$ branching fraction, the main limitation is the uncertainty on the ratio of production fractions of B^0_s and B^0/B^+ mesons, f_s/f_d; for the $B^0 \to \mu^+\mu^-$ branching fraction, the main limitation is the uncertainty on the misidentification estimate, especially the resulting estimate of the contribution from $B^0_{(s)} \to h^+h'^-$ background. The uncertainties on both measurement are still dominated by the limited sample size. The measured value for the $B^0_s \to \mu^+\mu^-$ branching fraction is 16% smaller than the Standard Model prediction, similar to the uncertainty from the measurement and thus fully consistent. The measured value for the $B^0 \to \mu^+\mu^-$ branching fraction is fully consistent with the Standard Model, although it still has large uncertainties.

Using a dedicated selection for $B_s^0 \to \mu^+\mu^-$ candidates, the same dataset is used to measure the effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay. The decay time distribution for $B_s^0 \to \mu^+\mu^-$ candidates after background has been subtracted is shown in Figure 4. The lack of $B_s^0 \to \mu^+\mu^-$ candidates at small decay times is caused by the selection requirements required to reject the large amounts of combinatorial background, and is taken into account in the determination of the effective lifetime.

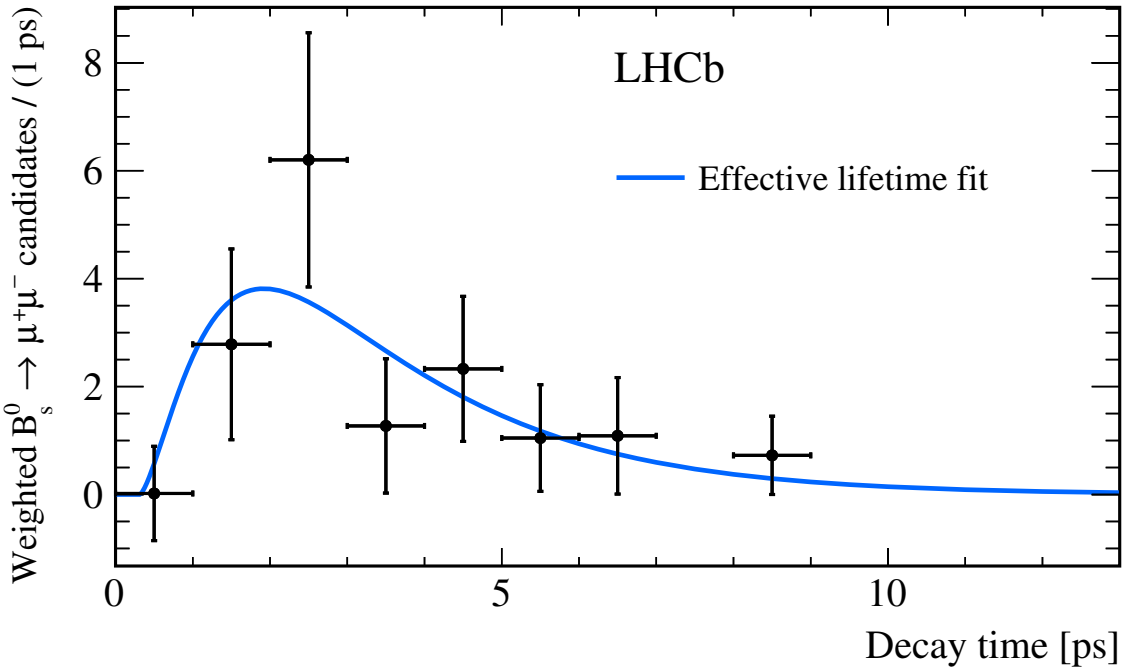

Figure 4: The fit to the background-subtracted lifetime distribution of $B_s^0 \to \mu^+\mu^-$ candidates used to extract the effective lifetime.

The measured value of the effective lifetime of the $B_s^0 \to \mu^+\mu^-$ decay is $\tau_{\rm eff}^{\mu^+\mu^-} = 2.04 \pm 0.44 \pm 0.05$ps, where the first uncertainty comes from the limited size of the dataset and the second from the limitations of the analysis procedure. The main limitation affecting the measurement is the validation of the effect of the selection requirements on the $B_s^0 \to \mu^+\mu^-$ decay time distribution. The measurement is consistent with the Standard Model, 1.62 ps, and very slightly favours it, as non Standard-Model values would be smaller than 1.62 ps.

How do these measurements of $B_s^0 \to \mu^+\mu^-$ with data from LHCb affect our understanding and what is next? Two other experiments at the LHC, ATLAS and CMS, have also measured the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions; CMS has also measured the $B_s^0 \to \mu^+\mu^-$ effective lifetime. Combining the results from all LHC experiments, the following averages are found: $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}, \mathcal{B}(B^0 \to \mu^+\mu^-) = (0.58^{+0.76}_{-0.72}) \times 10^{-10}$, and $\tau_{\rm eff}^{\mu^+\mu^-} = (1.70^{+0.61}_{-0.44})$ ps. The uncertainty on the $B_s^0 \to \mu^+\mu^-$

branching fraction takes into account that f_s/f_d is correlated between the measurements. Interestingly, the $B^0_s \to \mu^+\mu^-$ branching fraction is at some tension with, but still consistent with, the Standard Model. No significant contribution from the $B^0 \to \mu^+\mu^-$ decay is observed yet; its branching fraction estimate and the effective lifetime of the $B^0_s \to \mu^+\mu^-$ decay are consistent with the Standard Model predictions.

For all measurements of $B^0_{(s)} \to \mu^+\mu^-$ decays, the main limitation is still the sample size. All experiments are currently working on updating their results with the full dataset taken from 2011 to 2018, roughly tripling each dataset. The LHCb detector itself is currently being upgraded to a version that will be able to take five times as much data per year and enable to reduce the uncertainties on all $B^0_{(s)} \to \mu^+\mu^-$ measurements to 25% of their current values by 2030. This dataset will allow to strongly test new contributions to the branching fraction of $B^0_s \to \mu^+\mu^-$, to discover the $B^0 \to \mu^+\mu^-$ decay, and to start excluding some non Standard Model scenarios related to the $B^0_s \to \mu^+\mu^-$ effective lifetime. If we are lucky, it will be one of the key measurements showing us the existence of heavy new particles and open up a new portal into our understanding of the Universe.

Samenvatting

Alles dat wij observeren in ons universum bestaat uit fundamentele deeltjes met bepaalde massa's en ladingen, wiens gedrag voorspeld wordt door het Standaardmodel van de deeltjesfysica. Elk deeltje heeft een equivalent anti-deeltje: een deeltje met dezelfde massa maar tegenovergestelde lading voor elke interactie (in dagelijks taalgebruik worden interacties vaak krachten genoemd). Er zijn twee soorten materie: quarks en leptonen. Quarks zijn deeltjes die de sterke interactie voelen; leptonen (zoals electronen) niet. De sterke interactie maakt dat quarks snel samengestelde deeltjes vormen in configuraties die stabiel zijn voor de sterke interactie; beroemde voorbeelden zijn de protonen en neutronen, de twee deeltjes die voorkomen in atoomkernen. De deeltjes die centraal staan in dit proefschrift zijn B_s^0-en B^0-deeltjes; een ander soort samengesteld deeltje bestaand uit een quark en een anti-quark. Zij bestaan uit een "beauty" (schoonheid) anti-quark ($\overline{b}$) en respectivelijk een "strange" (vreemd,s) of een "beneden" (down,d) quark. B_s^0- en B^0-mesonen zijn instabiel en vervallen door de zwakke interactie na ongeveer 1.5×10^{-12} seconden (een miljoenste van een miljoenste van een seconde) of 1.5 picoseconden (ps) in het kort.

Dit proefschrift beschrijft de meting van een aantal eigenschappen van de vervallen van $B_{(s)}^0$-mesonen naar twee leptonen, specifiek naar een muon (μm) en een anti-muon (μ^+), waar in het kort naar verwezen wordt als $B_{(s)}^0 \to \mu^+\mu^-$-vervallen. Die eigenschappen zijn de vervalsfracties van $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-vervallen en de effectieve levensduur van het $B_s^0 \to \mu^+\mu^-$-verval. Deze vervallen zijn extreem zelfzaam en worden veroorzaakt door de zwakke interactie; toch kan zeer precies worden voorspeld hoe vaak ze voorkomen. De verwachting is dat ze ongeveer drie keer elke miljard B_s^0-vervallen voor komen, $(3.58 \pm 0.13) \times 10^{-9}$, en een enkele keer elke tien miljard B^0-vervallen, $(1.020 \pm^{+0.029}_{-0.055}) \times 10^{-10}$. Samen refereren we aan deze fracties als de vervalsfracties van $B_{(s)}^0 \to \mu^+\mu^-$-vervallen. De vervalsfracties zijn zo klein in het Standaardmodel omdat deze vervallen niet kunnen voorkomen als eerste orde "boom"-diagrammen, alleen als hogere orde quantum "lus"-diagrammen. Daarnaast wordt het $B_{(s)}^0 \to \mu^+\mu^-$-verval beïnvloed door heliciteitsonderdrukking; de zwakke interactie heeft een voorkeur voor een specifieke spinconfiguratie van de muonen die

niet mogelijk is in dit verval (omdat het een verval is van een pseudoscalair deeltje naar twee fermionen).

Een essentiële eigenschap van $B^0_{(s)}$-mesonen is dat zij kunnen transformeren naar hun eigen anti-deeltje ($\overline{B}^0_{(s)}$) via een ander quantum "lus"-diagram. Als gevolg daarvan bevinden $B^0_{(s)}$-mesonen zich in een quantumsuperpositie met $\overline{B}^0_{(s)}$-mesonen: zij bewegen zich voort als combinatie van deeltje en anti-deeltje. Het blijkt dat er twee soorten combinaties zijn, die eigentoestanden worden genoemd. Voor B^0_s-mesonen hebben deze eigentoestanden levensduren (de tijd voordat het meson uit elkaar valt) die ongeveer 10% verschillen van elkaar. Interessant genoeg vervalt volgens het Standaardmodel alleen de eigentoestand met een lange levensduur naar een $\mu^+\mu^-$-paar, wat correspondeert met een levensduur van 1.620 ps. Door de gemiddelde vervalstijd (ook effectieve levensduur) van $B^0_s \to \mu^+\mu^-$-vervallen te meten kan deze voorspelling worden getest.

Waarom is het interessant om de eigenschappen van $B^0_{(s)} \to \mu^+\mu^-$-vervallen te meten? Dat is precies omdat deze vervallen zo ongelofelijk zelfdzaam zijn in het Standaardmodel. Door het onzekerheidsprincipe uit quantumfysica kunnen $B^0_{(s)}$-vervallen veroorzaakt worden door deeltjes die zwaarder zijn dan de $B^0_{(s)}$-mesonen zelf: ze zijn indirect gevoelig voor het bestaan van nieuwe, zware deeltjes. Omdat $B^0_{(s)} \to \mu^+\mu^-$-vervallen zo zeldzaam zijn zouden zelfs deeltjes veel zwaarder dan alle bekende deeltjes de vervalsfracties sterk kunnen veranderen. In het bijzonder is het zo dat nieuwe scalaire of pseudoscalaire deeltjes geen heliciteitsonderdrukking zouden ondergaan, waardoor $B^0_{(s)} \to \mu^+\mu^-$-vervallen zeer gevoelig zijn voor zulke deeltjes. De effectieve levensduur van $B^0_{(s)} \to \mu^+\mu^-$-vervallen is een extra observabele die specifiek gevoelig is voor scalaire en pseudoscalaire contributies wanneer ze destructief zouden interfereren en op die manier verborgen zouden blijven in een meting van de $B^0_{(s)} \to \mu^+\mu^-$-vervalsfractie. Het is niet verrassend dat de zoektocht naar $B^0_{(s)} \to \mu^+\mu^-$-vervallen al lang duurt. De studie die wordt beschreven in dit proefschrift volgt meer dan 30 jaar van metingen op zoek naar $B^0_{(s)} \to \mu^+\mu^-$-vervallen, zoals getoond in Figuur 5.

Hoe bestuderen we $B^0_{(s)} \to \mu^+\mu^-$-vervallen? Omdat B^0_s- en B^0-mesonen zeer snel vervallen is het nodig om ze te produceren in een speciale faciliteit. In dit geval is de productiefaciliteit de Large Hadron Collider (LHC), een ring van 27 km lang, ongeveer 100 meter ondergrond in de buurt van Genève (een hadron is een samengesteld deeltje dat stabiel is volgens de sterke interactie). De LHC gebruikt de massa-energieverhouding die eerst door Einstein werd geponeerd, $E = mc^2$, om $B^0_{(s)}$-mesonen te produceren. Het versnelt protonen naar ongelofelijke energieën

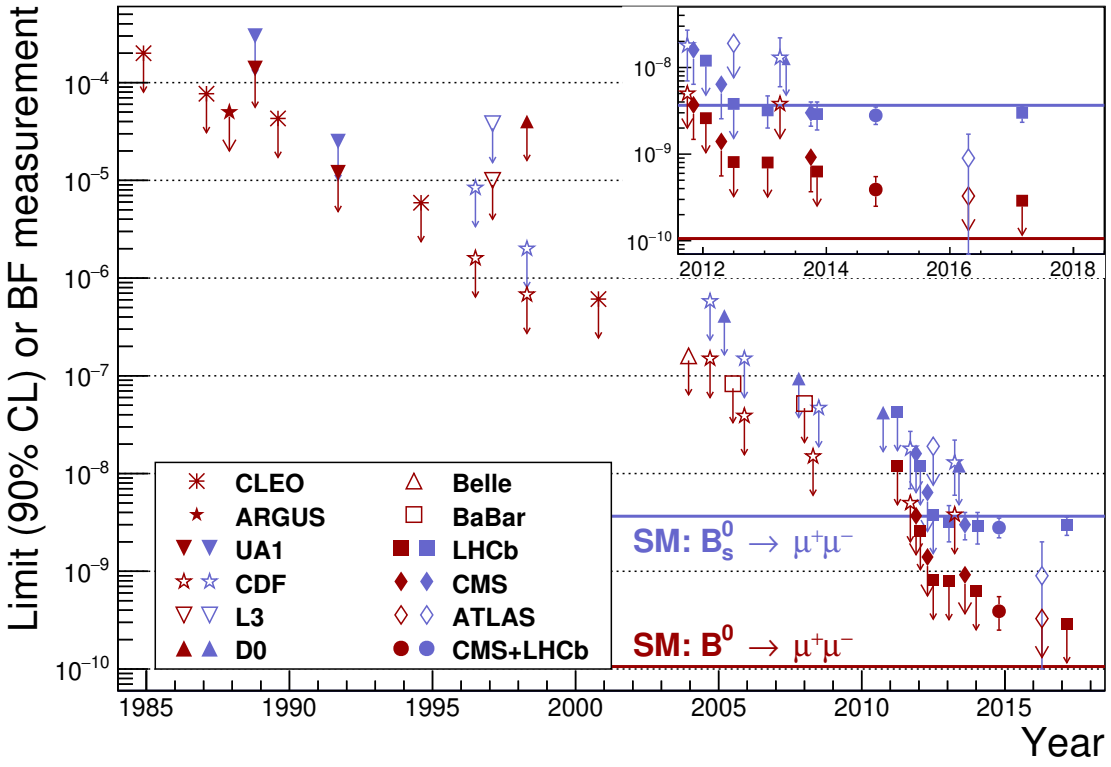

Figure 5: Een overzicht van de resultaten van vroegere metingen aan $B^0_{(s)} \to \mu^+\mu^-$-vervallen, zowel de limieten die eerder zijn gesteld en de recente metingen, tot en met de meting beschreven in dit proefschrift.

zodat ze een energie krijgen gelijk aan zevenduizend keer hun massa. Door zulke protonen met elkaar te laten botsen kan al deze energie vrijkomen en gebruikt worden om deeltjes te maken met massa's veel groter dan die van de protonen, inclusief $B^0_{(s)}$-mesonen. Het ontdekken van een zeer zeldzaam verval zoals $B^0_{(s)} \to \mu^+\mu^-$ maakt het nodig om miljarden $B^0_{(s)}$-mesonen te produceren; bij LHCb werden ongeveer 60.000 $B^0_{(s)}$-mesonen per seconde geproduceerd tijdens de operatie van de LHC van 2015 tot 2018.

LHCb is de detector die gebruikt is om de geproduceerde $B^0_{(s)}$-mesonen te detecteren. De LHCb-detector is voorwaarts geörienteerd rond de as van de proton-protonbotsingen omdat veel van de $B^0_{(s)}$-mesonen in die richting worden geproduceerd. Doordat de $B^0_{(s)}$-mesonen een snelheid hebben die rond 98% van de lichtsnelheid is en voor zulke snelle deeltjes tijd vertraagd is, vervallen ze na ongeveer 1 cm in de LHCb-detector te hebben gevlogen. Daarom wordt elk $B^0_{(s)}$-verval geïdentificeerd door de vervalsproducten te observeren; in het geval van $B^0_{(s)} \to \mu^+\mu^-$-vervallen zijn dat het muon en anti-muon. Een botsing met een waarschijnlijke $B^0_s \to \mu^+\mu^-$-kandidaat is te zien in Figuur 6.

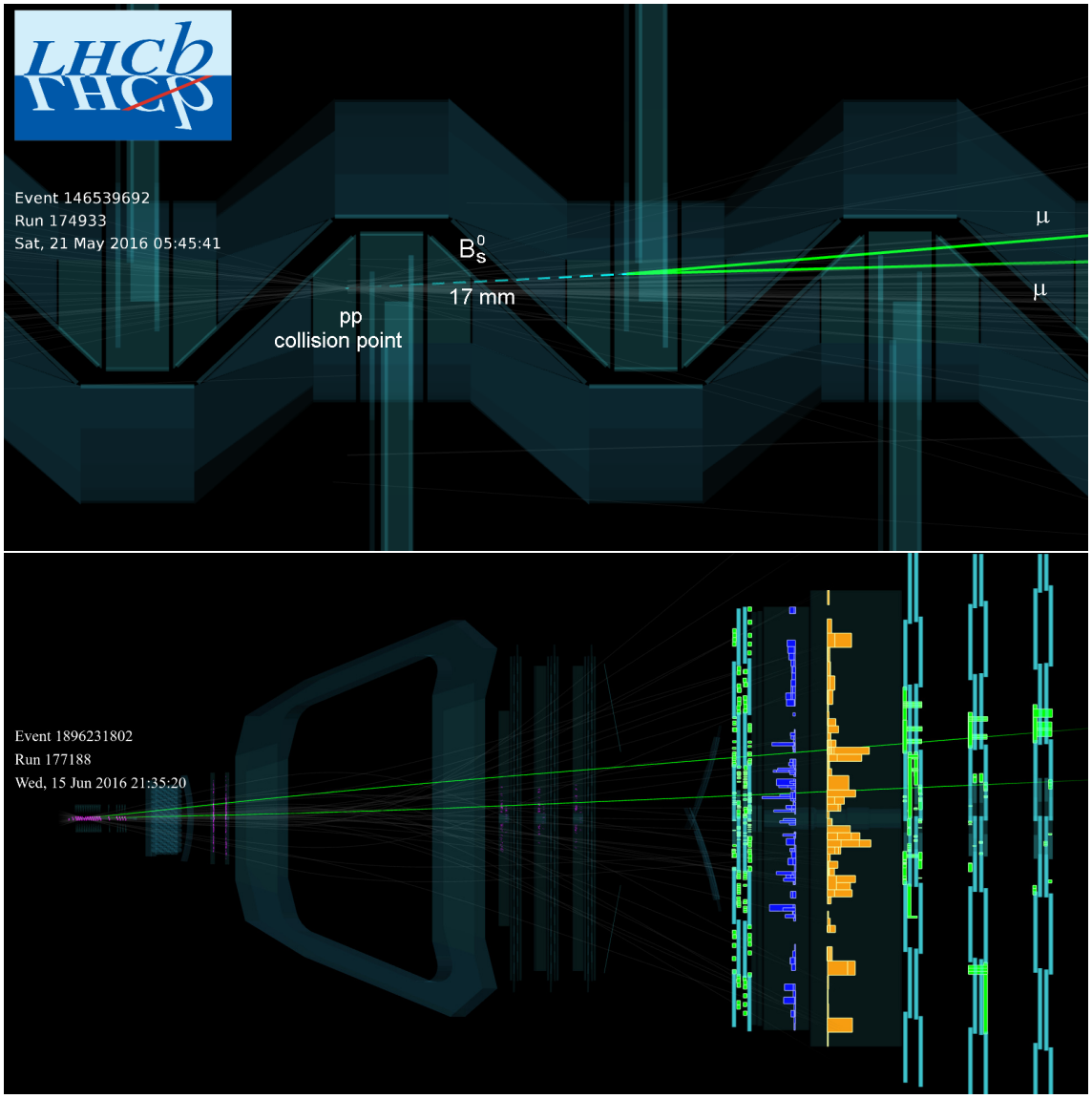

Figure 6: Een voorbeeldbotsing met een waarschijnlijke $B_s^0 \to \mu^+\mu^-$-kandidaat. De muonsporen zijn te zien als groene lijnen. Boven: een close-up van de botsing in de Vertex Locator, waarin het verval van het B_s^0 meson na 17 mm te zien is. Onder: de botsing in de volledige detector, waarin te zien is dat de magneet de muonen afbuigt en dat de muonen de muondetector bereiken aan het einde van de LHCb-detector.

Muonen hebben een elektrische lading, wat betekent dat ze gedetecteerd kunnen worden wanneer ze door een geschikt materiaal propagerent, bijvoorbeeld silicium: ze laten "sporen" achter. De LHCb-detector bevat zulk soort materialen zodat muonen gedetecteerd kunnen worden vanaf hun oorsprong in het $B_{(s)}^0$-meson verval tot aan het eind van de detector. De detector rond de proton-protonbotsing, de Vertex Locator, kan de $B_{(s)}^0$-meson vervalslocatie bepalen door de informatie van de twee muonsporen te combineren. Dit maakt het mogelijk om de vervalsduur van het $B_{(s)}^0$-meson te

bepalen, zoals nodig is om de effectieve levensduur van een verval te meten. Om de impuls en lading van de twee muonen te bepalen worden ze afgebogen door een sterke magneet binnen LHCb; hoe meer ze werden afgebogen door de magneet, hoe kleiner hun impuls was. Als gevolg van de afbuiging kan worden bepaald of de twee muonen een muon of anti-muon waren (volgens hun lading) en of de massa van het deeltje dat naar de twee muonen verviel consistent is met de B_s^0- of B^0-massa. Een belangrijke eigenschap van muonen bij de LHC is dat zij het minste beïnvloed worden door interacties met materiaal, terwijl alle andere geladen deeltjes gestopt worden in de LHCb-detector. Daarom zijn er een extra aantal sporendetectoren afgewisseld met ijzeren platen geplaatst aan het einde van de LHCb-detector, zodat muonen een zeer duidelijk teken achterlaten: een spoor dat door de hele detector persisteert.

Nadat geschikte $B_{(s)}^0 \to \mu^+\mu^-$-kandidaten zijn gereconstrueerd is het essentieel om de grote hoeveelheden achtergrondkandidaten te verwijderen die onderdeel zijn van de data. De twee belangrijkste achtergronden worden combinatorische en exclusive achtergrond genoemd. Combinatorische achtergrond komt van het combineren van een muon en anti-muon die eigenlijk kwamen van twee verschillende vervallen van b-hadronen; het is zeer overvloedig en komt ongeveer tien miljoen keer vaker voor dan $B_{(s)}^0 \to \mu^+\mu^-$. Het wordt verwijderd met behulp van een "machine-learning" algoritme, namelijk een Boosted Decision Tree (BDT). De belangrijkste manier waarop de BDT combinatorische achtergrond identificeert is door te zoeken naar de andere sporen geproduceerd in de b-hadronvervallen die geassocieerd kunnen worden met de muonen; zulke sporen zijn absent voor echte $B_{(s)}^0 \to \mu^+\mu^-$-vervallen. Exclusieve achtergronden worden vooral veroorzaakt door $B_{(s)}^0$-meson vervallen naar twee geladen deeltjes die er uitzien als twee muonen (hoewel één of beide mogelijk eigenlijk een hadron, bijvoorbeeld een proton, zouden kunnen zijn). Informatie van alle componenten van de LHCb-detector worden gecombineerd door een ander machine-learning algoritme, specifiek een neuraal netwerk, zodat slecht een paar van elke duizend hadronen als muonen worden geïdentificeerd.

Het $B_{(s)}^0 \to \mu^+\mu^-$-signaal wordt dan onderscheiden door de BDT-variabele te gebruiken om de data te verdelen in verschillende delen met verschillende signaal-achtergrondverhoudingen, en daarna de massa van de gereconstrueerde $B_{(s)}^0$-kandidaat te gebruiken; voor $B_{(s)}^0 \to \mu^+\mu^-$-vervallen piekt de massa rond de B_s^0- en B^0-massa's. De precieze kalibratie van het signaal en de inschatting van de resterende hoeveelheid achtergrond zijn een essentiëel onderdeel van de $B_{(s)}^0 \to \mu^+\mu^-$-analyse. Gelijkaardige vervallen worden gebruikt om te bepalen hoeveel $B_{(s)}^0$-mesonen werden geproduceerd in de data die wordt geanalyseerd. Die vervallen zijn van B^+-mesonen naar een J/ψ-deeltje en een geladen kaon, K^+, waarbij het J/ψ-deeltje vervalt naar twee muonen,

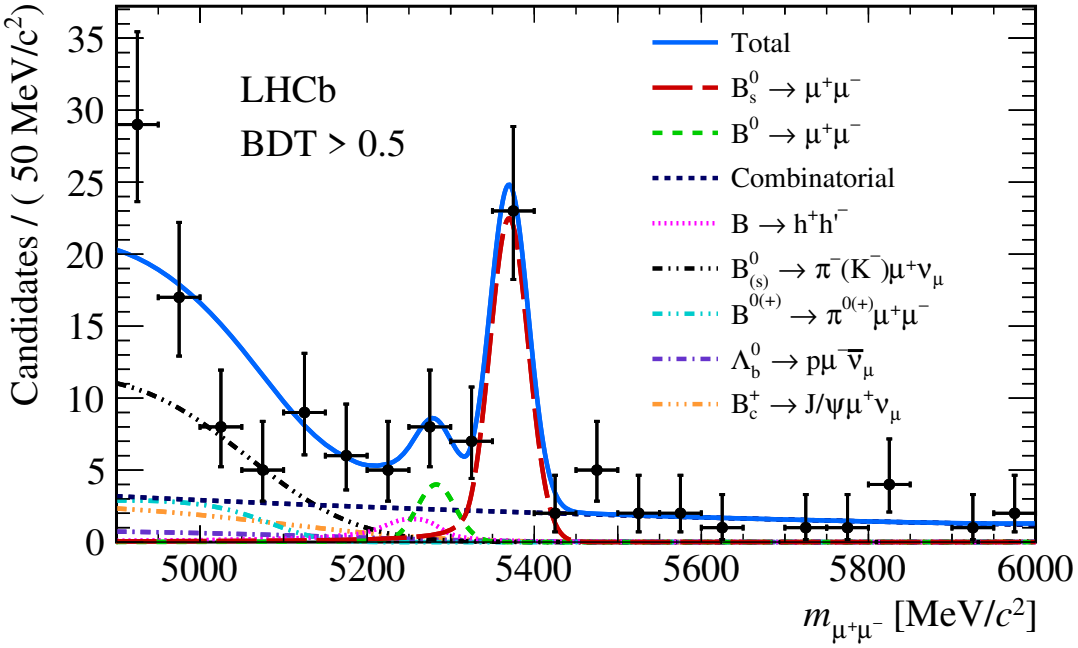

Figure 7: De massadistributie van de $B^0_{(s)} \to \mu^+\mu^-$-kandidaten die het meest op signaal lijken in de data gebruikt voor dit proefschrift. Het resultaat van de fit aan de data wordt getoond met de verschillende componenten uitgesplitst in de legenda.

zoals het signaal, en vervallen van $B^0_{(s)}$-mesonen naar twee hadronen, $B^0_{(s)} \to h^+h'^-$, met een gelijkaardige vervalsstructuur als $B^0_{(s)} \to \mu^+\mu^-$. Daarom lijken $B^0_{(s)} \to h^+h'^-$-vervallen ook op $B^0_{(s)} \to \mu^+\mu^-$-vervallen als het gaat om de massareconstructie en de BDT-variabele en worden ze gebruikt om die twee aspecten voor het signaal te kalibreren met de data. Een belangrijke factor die nodig is om de normalisatie voor B^0_s-mesonen te kalibreren is hun productiefractie ten opzichte van B^0- of B^+-mesonen, die f_s/f_d wordt genoemd; het is gemeten in andere analyses en is een input om de vervalsfractie van B^0_s-mesonen te bepalen.

Gebruikmakend van de data die door de LHCb-detector is verzameld van 2011 tot 2016, en met de strategie die zojuist is beschreven, worden de $B^0_{(s)} \to \mu^+\mu^-$-kandidaten die het meest op signaal lijken getoond in Figuur 7, inclusief het resultaat van een fit van het model dat het gedrag van signaal en achtergrond in de dataset na de signaalselectie beschrijft. Het resultaat van de meting is dat het $B^0_s \to \mu^+\mu^-$-verval voor het eerst wordt geobserveerd door een enkel experiment; er is geen significant signaal voor het $B^0 \to \mu^+\mu^-$-verval.

Uit de fit aan de data worden de vervalsfracties voor beide vervallen bepaald: $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$ en $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}$; daarbij komt de eerste onzekerheid van de gelimiteerde hoeveelheid data en de tweede

onzekerheid van de beperkingen van de analyseprocedure. Voor de $B_s^0 \to \mu^+\mu^-$-vervalsfractie is de hoofdbeperking de onzekerheid op de productiefractie van B_s^0 versus B^0 en B^+ mesons, f_s/f_d; voor de $B^0 \to \mu^+\mu^-$-vervalsfractie is de hoofdbeperking de onzekerheid op de inschatting van de misidentificatie, in het bijzonder de resulterende inschatting van de contributie van $B^0_{(s)} \to h^+h'^-$-achtergrond. De onzekerheden op allebei de metingen worden nog gedomineerd door de beperkte hoeveelheid data. De gemeten waarde voor de $B_s^0 \to \mu^+\mu^-$-vervalsfractie is 16% kleiner dan de Standaardmodelvoorspelling, ongeveer evenveel als de onzekerheid op de meting en dus consistent. De gemeten waarden voor de $B^0 \to \mu^+\mu^-$-vervalsfractie zijn volledig consistent met de voorspelling, hoewel de onzekerheden nog groot zijn.

Met een toegewijde selectie voor $B_s^0 \to \mu^+\mu^-$-kandidaten wordt met dezelfde dataset ook de effectieve levensduur van het $B_s^0 \to \mu^+\mu^-$-verval gemeten. De vervalsduurdistributie voor $B_s^0 \to \mu^+\mu^-$-kandidaten na het aftrekken van achtergrond wordt getoond in Figuur 8. Het gebrek aan $B_s^0 \to \mu^+\mu^-$-kandidaten met een korte vervalsduur wordt veroorzaakt door de selectie die nodig is om de grote hoeveelheden combinatorische achtergrond te verwijderen en wordt meegenomen bij het bepalen van de effectieve levensduur.

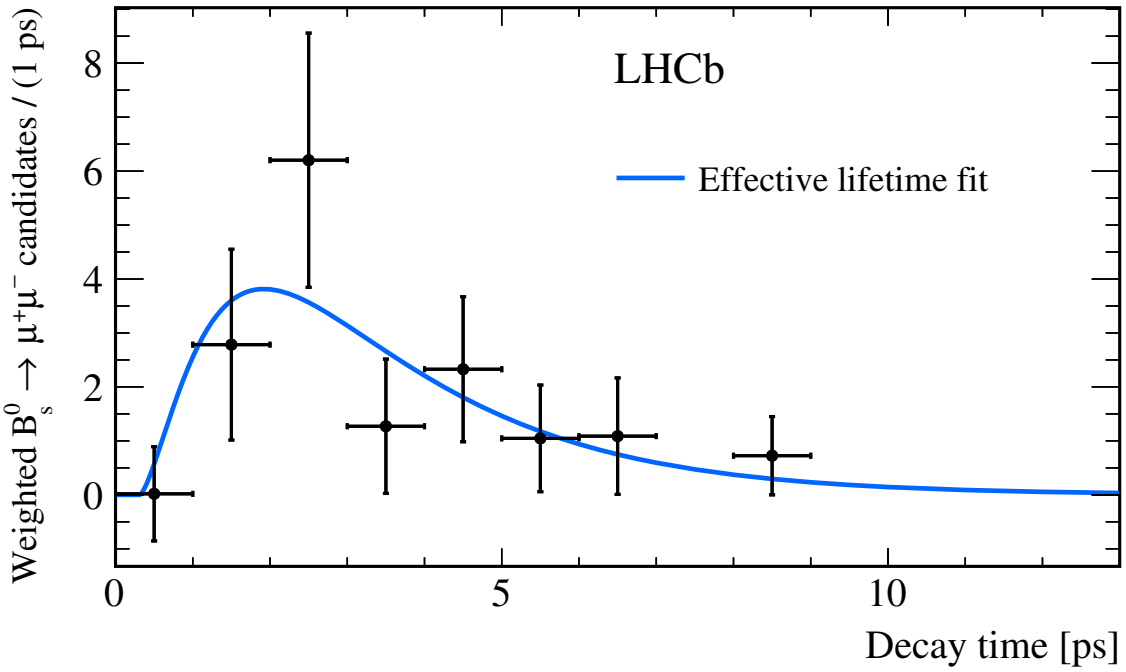

Figure 8: De fit aan de vervalsduurdistributie van $B_s^0 \to \mu^+\mu^-$-kandidaten na het aftrekken van achtergrond, die gebruikt wordt om de effectieve levensduur van $B_s^0 \to \mu^+\mu^-$-kandidaten te bepalen.

De gemeten waarde van de effectieve levensduur van het $B_s^0 \to \mu^+\mu^-$-verval is $\tau_{\text{eff}}^{\mu^+\mu^-} = 2.04 \pm 0.44 \pm 0.05$ps, waar de eerste onzekerheid komt van het beperkte aantal $B_s^0 \to \mu^+\mu^-$-kandidaten in de data en de tweede van de beperkingen van de analyseprocedure. De hoofdbeperking van de analyseprocedure is het valideren van het effect van de selectie op de $B_s^0 \to \mu^+\mu^-$-vervalsduurdistributie. De meting is consistent met de voorspelling uit het Standaardmodel, 1.62 ps, en een lichte aanwijzing ervoor, omdat anders de levensduur lager zou zijn geweest, niet groter.

Hoe vergroten de metingen van $B_{(s)}^0 \to \mu^+\mu^-$ met data van de LHCb detector ons begrip en wat komt er hierna? Twee andere experimenten bij de LHC, ATLAS en CMS, hebben ook de $B_{(s)}^0 \to \mu^+\mu^-$-vervalsfracties gemeten; CMS heeft ook de effectieve levensduur van $B_s^0 \to \mu^+\mu^-$ gemeten. Door de resultaten van alle LHC-experimenten te combineren worden de volgende gemiddelden gevonden: $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}, \mathcal{B}(B^0 \to \mu^+\mu^-) = (0.58^{+0.76}_{-0.72}) \times 10^{-10}$, en $\tau_{\text{eff}}^{\mu^+\mu^-} = (1.70^{+0.61}_{-0.44})$ ps. De onzekerheid op de $B_s^0 \to \mu^+\mu^-$-vervalsfractie neemt mee dat f_s/f_d gecorreleerd is tussen de verschillende metingen. Interessant genoeg is er een (nog niet significant) verschil tussen de voorspelling van het Standaardmodel en de $B_s^0 \to \mu^+\mu^-$-vervalsfractie. Er is nog geen significante contributie van het $B^0 \to \mu^+\mu^-$-verval geobserveerd; zowel de $B^0 \to \mu^+\mu^-$-vervalsfractie als de effectieve levensduur van $B_s^0 \to \mu^+\mu^-$ zijn consistent met de voorspellingen.

Alle metingen van $B_{(s)}^0 \to \mu^+\mu^-$-vervallen worden nog beperkt door de hoeveelheid $B_{(s)}^0 \to \mu^+\mu^-$-kandidaten. Alle LHC-experimenten werken op het moment aan een nieuwe meting met alle data die genomen is tot en met 2018, zodat de hoeveelheid kandidaten toeneemt met ongeveer een factor drie. De LHCb-detector zelf wordt op dit moment vervangen met een verbeterde versie die vijf keer zoveel data per jaar zal kunnen nemen en het mogelijk zal maken om alle $B_{(s)}^0 \to \mu^+\mu^-$-metingen te herhalen met een vier keer zo kleine onzekerheid op de metingen. Met zulke onzekerheden zal het mogelijk zijn om een sterke test te doen van nieuwe deeltjes die bijdragen aan de $B_s^0 \to \mu^+\mu^-$-vervalsfractie, om het $B^0 \to \mu^+\mu^-$-verval te ontdekken, en om sommige niet-Standaardmodel-scenarios voor de effectieve levensduur van $B_s^0 \to \mu^+\mu^-$ uit te sluiten. Als we geluk hebben zullen deze metingen de sleutel vormen om ons het bestaan van nieuwe zware deeltjes te laten zien en een nieuw portaal openen om ons begrip van het universum te verdiepen.

Acknowledgements

To become the physicist I am now and to write this thesis would not have been possible without the help of many people.

Antonio, thanks for giving me the opportunity to work on the $B^0_{(s)} \to \mu^+\mu^-$ analysis. To present the status of the analysis in front of the whole collaboration within a year of starting the PhD was quite the experience. And although at times it seemed like squeezing blood from a stone, thanks for polishing the thesis and helping me to make it to the finish line. Molte grazie per tutti!

Marcel, your on-point and rapid comments helped me when I was very stuck on the thesis and progress was slow; thanks for getting me through that time. Your professonial and personal advice has been very helpful to me; your enthusiasm and human interest are essential to the success of the bfys group at Nikhef.

Dear Flavio, molte molte grazie a ti. You joined me at Nikhef after half a year in my PhD as "my" postdoc and quickly helped me to find my way in the $B^0_{(s)} \to \mu^+\mu^-$ analysis. We almost made it to CKM, working days and nights (never again!), but the BDT calibration was difficult to crack. When we figured it out just after and published just one year after joining us. Thank you for all your advice and kind words, you are a great teacher and friend. Spero che tu sia felice e possa rilassarti un po 'negli anni a venire.

There are many people to thank for guiding me on the path to my the PhD and for passing on their love and enthusiasm for physics in general or particle physics specifically. Meneer Appelman, bedankt voor het studieadvies en uw droge humor. Studeren aan de VU heeft me goed voorbereid voor mijn latere onderzoek en was de perfecte stap op dat moment. Thanks to Rob Lambert for very enthusiastically introducing me to the world of particle physics, teaching me how to discover B meson mixing and how to present it in a concise and interesting presentation in only a few weeks. Thanks for Hella and Tim Head for supervising me in my first project at CERN as a Summer Student, the great group of Nikhef summer students we had at the time, and many thanks to the past and current contributors to the LHCb

StarterKit, which introduced me and many other PhD students and postdocs to the LHCb software.

Niels, thanks for your enthusiasm, drive, and for always keeping in mind the bigger picture. You made my masters project very enjoyable and made it possible to publish a paper on $B \to DD$ decays. It has been a pleasure to keep working with you on the OT and I look forward to many more such projects in the future.

Maarten, bedankt voor de gezelligheid, de eerste hulp bij uit huis wonen, de interessante discussies over fysica en voor alle tijd samen als kantoorgenoot, zelfs al waren we op een moment als een oud getrouwd stel. Ik hoop dat we nog veel samen kunnen blijven werken in de toekomst.

Katya, thank you for your great friendship, your perseverance and your very strong motivation. I am very interested to see what you will do in your coming CERN Fellow and beyond.

Elena, thanks for the many conversations, discussions, and mutual complaints we shared. I wonder what interesting projects you might lead on future detectors.

Jacco, nadat we elkaar al kenden van de Beta-PAL, was het een aangename verassing toen ik je in de LHCb-groep tegenkwam. Bedankt voor de slechte grappen, de fijne gesprekken en gefeliciteerd met je positie in Maastricht!

Georgios, as my first office mate, you have a special place in this story. $\Xi\varepsilon\rho\imath\sigma\ \tau\imath\ o\rho\alpha\ \imath\nu\eta$? $O\rho\alpha\ \gamma\imath\alpha\ \tau\sigma\alpha\imath/\phi\alpha\imath$! I hope to see you around at CERN again. Vasilis, it was great to be your office mate as well. You're always kind, happy and funny even if the last was sometimes unintentional. It is a pleasure every time to encounter you in Amsterdam.

Pieter, Veerle, Jeroen, Roel, Suvayu, Rose, Siim, Elise, Lennaert, Laurent, thanks for the warm welcome into the LHCb Nikhef group when I was just finding my way as a master student.

Gerhard, Patrick, Tjeerd, Jeroen en Wouter: bedankt voor jullie hulp, advies en kritiek waar nodig. Ook dankzij jullie was het een feest om onderdeel van de LHCb-groep te zijn.

Many people joined after me in the LHCb group. Thanks to Mauricio, Cristina, Igor, Michele, Hilbrand, Marjolein, Laís, Tenglin, Tim, Sean, Daniel, Carlos, Sevda and everyone else for keeping the Nikhef group welcoming, supportive and a little bit crazy.

On the analysis side, there has always been a significant Italian component. Molte grazie a il due Mattei, Barbara, Marco, Guido e Luca for the work together in our analyses, and showing the asymptotic freedom of the Italian interaction.

I am very lucky to have supervised and worked together with many students; specifically, thanks to Jennifer (een f_s/f_d expert tijdens de master), Jordy (eindelijk mijn masteranalyse af!), Giovanni (who now has a flavour of the difficulty of electron analyses at LHCb), Robin (le expert du per-event mass error), Silvia (has hecho el análisis como me hubiera gustado hacerlo la ultima vez), Lex (het is een feest om samen te werken), Vlad (good luck in Marseille!), Paul (let's see whether we find LFV) and Sietske (immediately part of our Λ_b^0 group). I am proud of what you have achieved and I hope you find the right place for yourselves, whether it's inside or outside particle physics.

Staying a year at the French/Swiss border was quite the adventure, as Maarten can attest to; it was especially nice to be welcomed by the Nikhef-Heidelberg lunch group at the time, with special thanks to Svende, Maria, Patrick and Bassem. Thanks a lot to Maarten van Dijk, Nathan, Greg and Marie for the Friday evening drinks. Special thanks to Greg for providing a listening ear and sage advice on the $B_s^0 \to \mu^+\mu^-$ analysis (or anything else really).

There many people to thank for various projects I worked on beside the analysis work. Andrea, Lucia, Marianna, Vladimir and Francesco Polci: thanks a lot for your help with the creation of the TrackCalibProduction package. Georgios, Marco Adinolfi, Dominik, Gloria, thanks for the collaboration during my time as simulation liaison. Thanks to Niels, Bassem and the great group of people at the pit for making it very enjoyable to do OT shifts and shift leader shifts at the pit. Additionally, the LHCb Masterclass live from the pit (literally next to the detector!) was an amazing experience and I would strongly recommend it to anyone else.

Aan mijn ouders en Suzanne, bedankt voor jullie hulp vanaf het begin met al mijn nieuwsgierigheid en vragen, jullie interesse tijdens mijn onderzoek (het bezoek aan LHCb hebben jullie nog te goed), en jullie voortdurende gezelligheid, steun en advies. Ik hou ontzettend veel van jullie.

A Naye, muchas gracias no es sufficiente para describir mi gratitud. Cuando siente desesperado con mi thesis o otras cosas, me ayudaste y motivaste cada vez. Espero que tenemos muchos años juntos por hacernos muy felices, por bailar, por abrazar y por rier. Soy muy suertudo estar contigo. Te amo tanto.