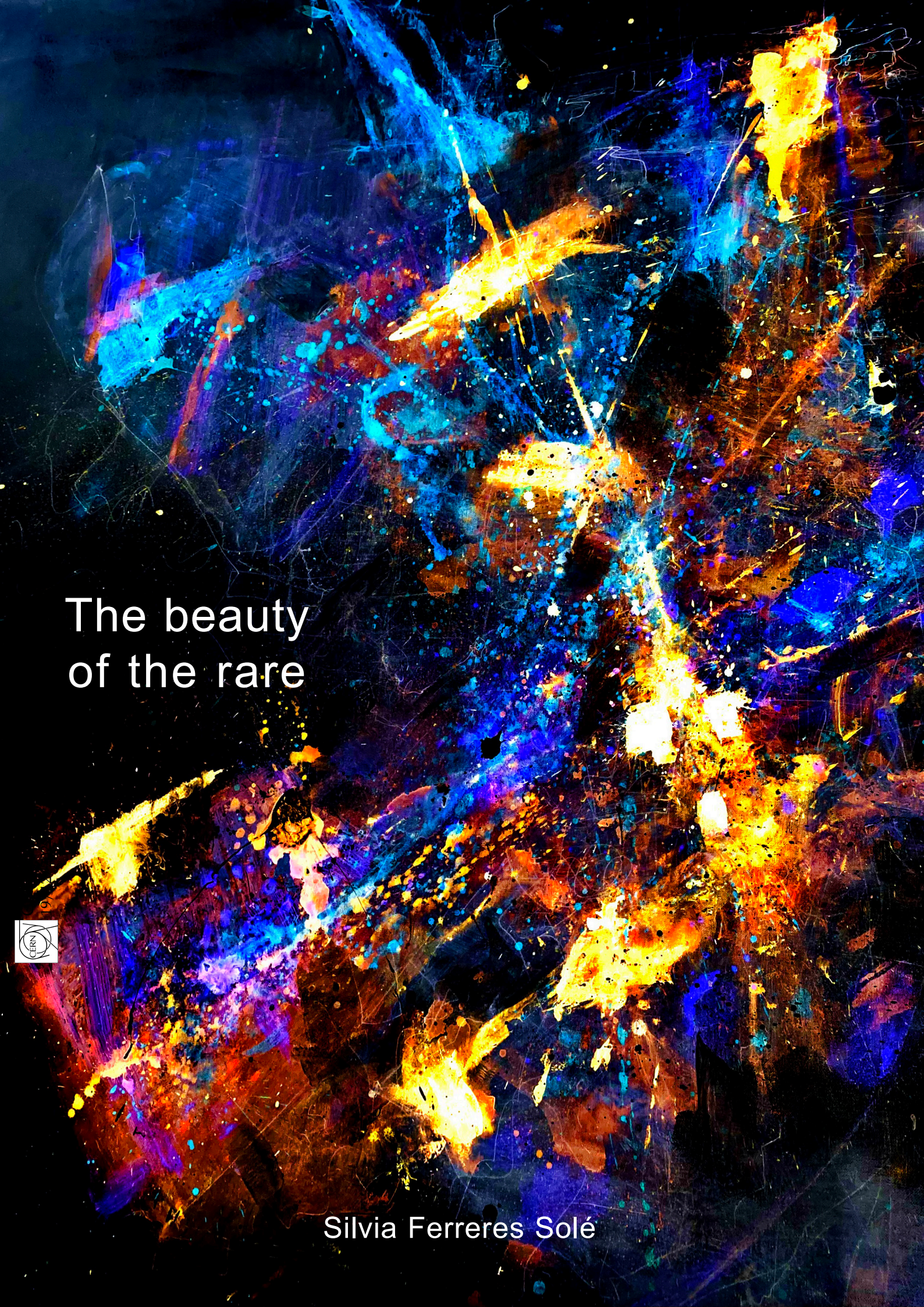

The beauty
of the rare
Silvia Ferreres Solé

Doctoral thesis

THE BEAUTY OF THE RARE

Silvia Ferreres Solé

2022

The beauty of the rare
Thesis, Maastricht University, Maastricht

Cover artwork by Ruska Balbuena, Silvia Ferreres Solé and Daniel Mangada Watt
Description: La obra representa de forma abstracta la colisión entre partículas y las trayectorias resultantes, pretendiendo reflejar la belleza de tal evento a través de la pintura y el color.
This artwork is an abstract interpretation of the collisions between particles and the resulting trajectories, reflecting the beauty of such an event through paint and colour.

Printed by Gildeprint
ISBN 978-94-6419-645-0

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 for Subatomic Physics (Nikhef) in Amsterdam, The Netherlands.

THE BEAUTY OF THE RARE

Dissertation

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. Pamela Habibović
in accordance with the decision of the Board of Deans,
to be defended in public
on Monday 19th December 2022, at 13:00 hours.

by

Silvia Ferreres Solé

Supervisor:
Prof. dr. M. Merk

Co-supervisor:
Dr. J. A. de Vries

Assessment Committee
Prof. dr. S. Hild (chair)
Prof. dr. R. Fleischer, Vrije Universiteit Amsterdam, The Netherlands
Prof. dr. R. Heeren
Dr. M. Palutan, Frascati National Laboratory, INFN, Italy
Dr. K. Vos

Contents

Introduction

Quarks and leptons are the elementary constituents of matter surrounding us. Their interaction is possible through the four fundamental forces ruling the universe: the gravitational, strong, electromagnetic and weak interactions. Besides gravity, interactions among elementary particles occur through the exchange of force-carrier particles mediating the process acting as messengers by transmitting four-momentum between the interacting particles. Offering an elegant mathematical framework, the Standard Model encloses the description of these interactions and the properties of the quarks and leptons.

Since its development in the 1970s, the Standard Model has been widely tested and established as a self-consistent theory [1]. It has been proven to describe and accurately predict a wide spectrum of physical phenomena observed in experiments. Despite being the best current description of nature, the Standard Model still leaves open questions as it fails to explain several aspects of fundamental nature. Although nature as we know it is mainly formed by matter, matter and antimatter should have been produced in equal amounts in the formation of the universe; where did the antimatter go? Why are there three generations of fundamental particles? What is the relation between the leptons and quarks, the fundamental components of matter, and why are they different?

Answering these questions will increase our understanding of the inner workings of nature and shed some light on how matter originated in the formation of the universe. Some of the answers might be found beyond the Standard Model through new models developed to give answers to these unexplained phenomena. These New Physics theories are built either as extensions to the Standard Model or as entirely new theories, such that the Standard Model would be an approximation at a low energy scale of a bigger picture, similar to the case of classical and relativistic mechanics.

Our mission as particle physicists is to try to find an explanation to these fundamental questions by constraining or excluding New Physics models or even finding new particles and new forces never predicted before. This is the primary purpose of the Large Hadron Collider (LHC) at CERN, built to collide particles, mainly protons, to produce thousands of different particles, recreating the conditions of the early universe. These collisions take place in one of the four detectors located on the LHC ring such that the various particle processes

occurring can be detected and analyzed further to search for New Physics.

Among all the particle processes occurring after the collisions, decays of unstable particles are very common. These processes consist in the transformation of an unstable particle into at least two other particles. Particle decays occur through the exchange of virtual particles that mediate the decay and can involve several interaction types. The mediating particles exist for such a short time that their properties are not well-defined due to the Heisenberg uncertainty principle. As a consequence, more massive particles than the decaying ones can act as mediators of their decays. New heavy particles not predicted by the Standard Model might enter these decays as mediators, altering, for instance, the rate at which they occur relative to what is predicted by the Standard Model.

One of such decays is the decay of neutral B mesons into two muons, which is the topic of this dissertation. While muons are elementary particles, in particular leptons, the B_s^0 (B^0)-hadrons are composite particles formed by the combination of two quark types: a b-quark and a s-quark (d-quark). In the Standard Model, the $B_s^0 \to \mu^+\mu^-$ decays are predicted to be significantly suppressed, and only around 3 out of every billion B_s^0 mesons are expected to decay into two muons. The $B^0 \to \mu^+\mu^-$ is even more suppressed, with only 1 every 10 billion B^0 mesons decaying through this decay mode. Although significantly suppressed in the Standard Model, new mediating virtual particles can affect the decay process, enhancing or reducing the frequency at which these decays occur. The idea is to measure the rate of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays produced in the LHC and compare it to the frequency predicted by the Standard Model. Finding any disagreement between theory and experiment would imply the existence of new particles mediating the process, giving a clearer picture of potential new models or theories.

This dissertation presents the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and the search for $B^0 \to \mu^+\mu^-$ and initial-state radiative $B_s^0 \to \mu^+\mu^-\gamma$ processes using the data collected in pp collision by the LHCb experiment. The theoretical framework describing these processes, jointly referred to as $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ decays, is introduced in Chapter 1 where the theoretical motivation for their study is given. Because $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ decays are very rare processes, a large amount of data must be studied in order to detect them. For this reason, the entire dataset gathered by the LHCb experiment from 2011 to 2018 is used in this analysis. The LHCb is one of the four detectors located in the LHC ring, whose layout and the different subdetectors forming it are shown in Chapter 2. An overview of the analysis is presented in Chapter 3 while a detailed explanation can be found in the subsequent chapters. A review of the statistical concepts and the final analysis results are presented in Chapter 7. In Chapter 8, the results obtained in the analysis are compared with the Standard Model and combined with other measurements to constrain New Physics models. Finally, an outlook with a summary of future improvements in the analysis is given in the same chapter.

The $B_s^0 \to \mu^+\mu^-\gamma$ analysis is a collaborative effort involving many people in the LHCb

experiment. Although a general overview is given, this manuscript mainly focuses on the research that was contributed by the author. In particular, the determination of the trigger efficiencies and the calibration of the BDT response. The trigger efficiencies are needed to correct for the fact that several $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ events do not pass the trigger requirements and are not stored for further analysis. The evaluation of these efficiencies is presented in Sec. 4.2. Since the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are very rare processes, the sensitivity of the data sample is maximized by fitting the dimuon invariant mass spectrum in bins of a BDT classifier (see Sec. 4.5). To ensure the fit convergence, the fraction of signal events ending up in each BDT bin must be constrained. The evaluation of these relative signal yields is determined in the so-called BDT calibration following a procedure detailed in Chapter 5.

1 | Theoretical framework

Decays of b-mesons into two leptons, such as $B^0_{(s)} \to \mu^+\mu^-$ decays, are interesting laboratories in the search for New Physics scenarios. In the Standard Model, these decays are predicted to be very rare processes mediated by the weak interaction, presented in Sec. 1.1. They are flavour-changing neutral current processes that can only take place through so-called quantum loop transitions, involving corresponding loop and weak interaction suppression factors. They are also affected by the so-called GIM mechanism and helicity suppression, further decreasing their decay rates.

Another quality that makes the $B^0_{(s)} \to \mu^+\mu^-$ processes of high interest is the presence of only leptons in the final state, while the hadronic interaction is restricted to the initial b-hadron. This factorization results in a clean theoretical computation of their branching fraction, which is performed using an effective field approach described in Sec. 1.2. Following the general expression obtained from the effective field approach, the Standard Model prediction is calculated to a high precision level. Consequently, any deviations from the measured branching fraction relative to the Standard Model value will serve as hints of physics beyond the Standard Model. In order to constrain New Physics models, the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction values are generally combined with other complementary measurements showing deviations from the Standard Model, such as the rare decay measurement involving $B \to K^{(*)}l^+l^-$ decays [2].

Photon emission is inevitable in $B^0_{(s)} \to \mu^+\mu^-$ decays. This photon radiation must be taken into account when calculating their decay branching fractions as well as in their experimental measurement. The two types of photon emission in the $B^0_{(s)} \to \mu^+\mu^-$ processes are presented in Sec. 1.3. Moreover, the effect of neutral B meson mixing is also considered in Sec. 1.4, where the difference between the theoretical and experimental approaches is discussed and their relation presented.

1.1 Standard Model: Weak interaction

The weak interaction is mediated by the W and Z bosons, which only couple to left-handed fermions and right-handed antifermions, resulting in the breaking of parity symmetry [3–5].

Moreover, processes mediated by the $W^{\pm}$ bosons cause the quarks to change their flavour to another type. This process is described in the interaction basis of the quarks with the charged current lagrangian term

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}\overline{u^I_{iL}}\gamma_\mu W^{-\mu} d^I_{iL} - \frac{g}{\sqrt{2}}\overline{d^I_{iL}}\gamma_\mu W^{+\mu} u^I_{iL} \tag{1.1}$$

where g is the coupling constant of weak interaction and u^I_{iL} (d^I_{iL}) stands for the Dirac spinors of the left-handed up-type (down-type) quarks in the weak interaction basis, denoted by the superscript I. The subscript i indicates the three generations. The presence of two terms in Eq. 1.1 follows from the existence of two different charged currents mediated by oppositely charged W bosons: the W^+ and W^- bosons.

The propagation states of the quarks are different from the interaction basis used in Eq. 1.1. Quark states acquire mass from their interaction with the Higgs field, described by the Yukawa coupling [6]. From this coupling, a relation between the mass eigenstates, denoted as d_i and u_i, and the interaction eigenstates, d^I_j and u^I_j, is determined. By convention, the interaction and mass eigenstates are equal for the up-type quarks and the change of basis is absorbed by the down-type quarks as [7]

$$\begin{aligned} u^I_i &= u_j \\ d^I_i &= V_{CKM,ij} d_j \end{aligned} \tag{1.2}$$

where $V_{CKM,ij}$ are the components of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [7] implementing flavour changing among different quark types in the charged current. To explicitly show the quark mixing relations, Eq. 1.2 can be written in matrix form as

$$\begin{pmatrix} d^I \\ s^I \\ b^I \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.3}$$

The magnitudes of the CKM matrix elements are determined from the combination of several experimental measurements resulting in [1]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix} \tag{1.4}$$

Since the quark mass eigenstates are the propagation states as observed in nature, the lagrangian from Eq. 1.1 can be expressed in the mass basis as

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}\overline{u_{iL}} V_{CKM,ij} \gamma_\mu W^{-\mu} d_{jL} - \frac{g}{\sqrt{2}}\overline{d_{iL}}\gamma_\mu V^*_{CKM,ij} W^{+\mu} u_{jL} \tag{1.5}$$

such that the lagrangian is expressed in the same basis as the Yukawa coupling, giving rise to the quark masses. By doing that, the quark mixing is explicitly included in the charged current Lagrangian.

1.1.1 Suppression of the $B^0_{(s)} \to \mu^+\mu^-$ decays in the Standard Model

The $B^0_{(s)} \to \mu^+\mu^-$ process requires the change of flavour between two down-type quarks, the $d(s)$ and the b quarks. Such transitions are known as *flavour changing neutral currents (FCNC)* because the charge of the two quarks is the same. These transitions are not allowed at tree level in the Standard Model since neutral currents only occur at tree level by the exchange of a Z boson between two same-type quarks. Instead, FCNC processes are only allowed in the Standard Model through loop transitions involving multiple charge currents; hence, they are highly suppressed.

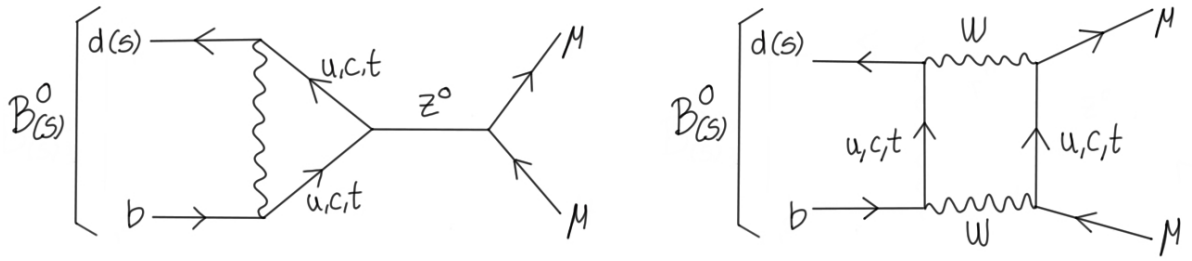


Figure 1.1: Standard Model Feynman diagrams of the lowest order contributions to $B^0_{(s)} \to \mu^+\mu^-$ decays: the penguin diagram (left) and the box diagram (right). A second box diagram contributes to the process obtained by replacing the W lines with u, c, t lines and vice versa.

The two main transitions contributing to the FCNC $B^0_{(s)} \to \mu^+\mu^-$ decays are displayed in the Feynman diagrams from Fig. 1.1. As with the rest of FCNC transitions, they are suppressed by the loop suppression factor $(g/4\pi)^2 \approx 10^{-3}$ arising from the two additional vertices needed in the process compared to tree-level transitions. At each vertex involving quark mixing, the corresponding CKM matrix element is added to the amplitude, further suppressing the $B^0_{(s)} \to \mu^+\mu^-$ processes, as can be seen from their magnitudes in Eq. 1.4. Moreover, virtual contributions from all up-type quarks in the loop contribute to the process such that the decay amplitude includes three terms with relative magnitudes according to:

$$\mathcal{A}(B^0_{(s)} \to \mu^+\mu^-) \propto V_{uq}V^*_{ub}\left(\frac{1}{q^2 - m_u^2}\right) + V_{cq}V^*_{cb}\left(\frac{1}{q^2 - m_c^2}\right) + V_{tq}V^*_{tb}\left(\frac{1}{q^2 - m_t^2}\right) \tag{1.6}$$

Due to the CKM unitarity condition, giving $V_{uq}V^*_{ub} + V_{cq}V^*_{cb} + V_{tq}V^*_{tb} = 0$, the $B^0_{(s)} \to \mu^+\mu^-$ processes would be completely forbidden if the masses of the quarks would be the same.

Fortunately, a complete cancellation is avoided thanks to the different masses of the quarks, a procedure known as the *GIM mechanism* [8].

Angular momentum conservation adds an additional source of suppression known as *helicity suppression*. The total angular momentum of the $B^0_{(s)}$ mesons is 0 since they are spin-0 particles with no angular momentum in the ground state. Due to angular momentum conservation, the total angular momentum of the final state must also be zero. This condition requires the final-state muons to have opposite spins and, hence, the same helicity in the centre of mass frame, as illustrated in Fig. 1.2. However, weak interaction only couples with left-handed (right-handed) particles (antiparticles), resulting in the muons from the decay having opposite chirality. Since chirality and helicity are identical in the ultra-relativistic limit, $B^0_{(s)} \to \mu^+\mu^-$ decays would be forbidden if muons were massless particles. Although intimately related, chirality and helicity are not identical for massive particles, lifting the cancellation of the $B^0_{(s)} \to \mu^+\mu^-$ process but adding a helicity suppression factor in the amplitude of the process of $(m_\mu/M_{B^0_{(s)}})^2 \approx 4 \times 10^{-4}$.

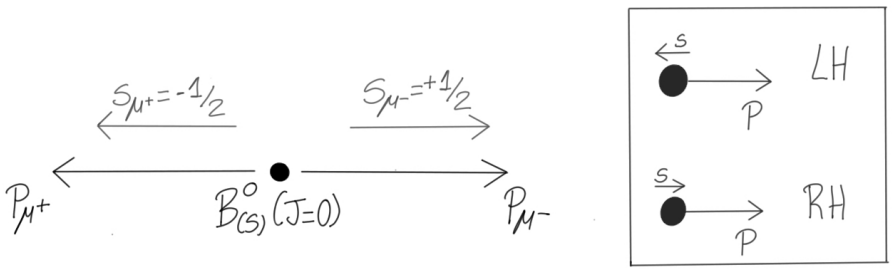


Figure 1.2: Illustration of the $B^0_{(s)} \to \mu^+\mu^-$ decay in the centre-of-mass frame resulting from the angular momentum conservation. The direction of the momenta of the muons is shown in black and their spin in grey. The helicity of the particles depending on the spin and the momentum are illustrated in the box.

Due to the loop suppression factors, the multiple weak interactions, the GIM mechanism and the helicity suppression, only 3 out of every billion B^0_s mesons are expected to decay to two muons in the Standard Model [9]. The number of B^0 decaying to two muons is expected to be around 1 out of 10 billion B^0 mesons [9].

1.2 Effective hamiltonian for weak interactions

The energy transfer in B meson decays is at the order of the b-quark mass, much smaller than the masses of the W and Z bosons mediating the decay. As a consequence, the effect of the propagators is negligible in the process such that they are "integrated out" of the theory and replaced by a four-point interaction containing all the perturbative information of the

process. In this effective field description [10], the weak mediators are then no longer degrees of freedom of the theory and the contributions to the $B^0_{(s)} \to \mu^+\mu^-$ processes are reduced to a four-point interaction.

Within the effective field description, the Hamiltonian describing the particle processes can be separated into two parts: the perturbative effects independent of the initial and final state particles are encoded in the Wilson coefficients C_i, while the long-distance part is parametrized with the Wilson operators, O_i. In the case of the $B^0_{(s)} \to \mu^+\mu^-$ decay, the effective field hamiltonian [11] is given by

$$\mathcal{H}_{eff} = -\frac{G_F \alpha}{\sqrt{2}\pi}\left(V^*_{tq}V_{tb}\sum_i \left(C_i^{(\prime)}O_i^{(\prime)}\right)\right) \tag{1.7}$$

where G_F is the Fermi coupling constant, α is the electromagnetic coupling constant and V^*_{tq} and V_{tb} the dominating CKM matrix elements present in the $B^0_{(s)} \to \mu^+\mu^-$ process. Using the effective hamiltonian, the amplitude for the $B^0_{(s)} \to \mu^+\mu^-$ decays is obtained as

$$\mathcal{M}(B^0_{(s)} \to \mu^+\mu^-) = \langle \mu^+\mu^-|\mathcal{H}_{eff}|B^0_{(s)}\rangle = -\frac{G_F\alpha}{\sqrt{2}\pi}V^*_{tq}V_{tb}\sum_i C_i^{(\prime)}\,\langle \mu^+\mu^-|O_i^{(\prime)}|B^0_{(s)}\rangle \tag{1.8}$$

The hadronic and leptonic parts of the $B^0_{(s)} \to \mu^+\mu^-$ decay are well separated and restricted to the initial and final state, respectively. Since there is no direct interaction between quarks and leptons, the decay amplitude is factorized into the leptonic and hadronic parts

$$\langle \mu^+\mu^-|O_i^{(\prime)}|B^0_{(s)}\rangle = \langle \mu^+\mu^-|O^{(\prime)}_{i,ll}|0\rangle\,\langle 0|O^{(\prime)}_{i,qq}|B^0_{(s)}\rangle \tag{1.9}$$

where $O^{(\prime)}_{i,ll}$ and $O^{(\prime)}_{i,qq}$ are the leptonic and hadronic components of the four-point operator. This factorization restricts the non-perturbative QCD effects to the initial state, allowing for a clean theoretical computation of the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction.

The specific form of the operators results from the combination of all the possible leptonic and hadronic currents defined by the theory [11]. Due to the pseudoscalar nature of the $B^0_{(s)}$ meson, the only operators contributing to the $B^0_{(s)} \to \mu^+\mu^-$ decay are [12, 13]:

$$\begin{aligned}
O_{10} &= (\bar{q}\gamma_\mu P_L b)(\bar{l}\gamma^\mu\gamma_5 l) & O'_{10} &= (\bar{q}\gamma_\mu P_R b)(\bar{l}\gamma^\mu\gamma_5 l)\\
O_S &= m_b(\bar{q}P_R b)(\bar{l}l) & O'_S &= m_b(\bar{q}P_L b)(\bar{l}l)\\
O_P &= m_b(\bar{q}P_R b)(\bar{l}\gamma_5 l) & O'_P &= m_b(\bar{q}P_L b)(\bar{l}\gamma_5 l)
\end{aligned} \tag{1.10}$$

where $q = s, d$ and $P_L = \frac{1}{2}(1-\gamma_5)$, $P_R = \frac{1}{2}(1+\gamma_5)$ are the chirality projectors such that chirality is flipped in the primed operators. These operators represent the Lorentz structure of the mediators contributing to the decay process that was integrated out. The O_{10} is related

to the axial vector lepton currents, such as the electroweak penguin shown in Fig. 1.1. The operators $O_{S,P}$ are (pseudo)-scalar lepton currents as penguin or box diagrams mediated by (pseudo) scalar particles such as the Higgs or possible contributions from hypothesized new particles.

1.2.1 Hadronic matrix elements evaluation

The hadronic part of the matrix element related to the four-point operator O_{10} is split into two parts:

$$\langle 0|O^{(\prime)}_{10,qq}|B^0_{(s)}\rangle = \frac{1}{2}\left[\langle 0|\overline{q}\gamma^\mu b|B^0_{(s)}\rangle \mp \langle 0|\overline{q}\gamma^\mu\gamma_5 b|B^0_{(s)}\rangle\right] \tag{1.11}$$

where the factor of 1/2 and the sign in the second term is given by the chiral projector, such that it is plus for $O'_{10,qq}$ and minus for $O_{10,qq}$. Because parity is conserved in the strong interactions, the hadronic matrix elements should result in pseudoscalar or axial vector currents due to the pseudoscalar nature of the $B^0_{(s)}$ meson. Hence, the first term in Eq. 1.11 vanishes as it is a vector current. The remaining term must be proportional to the $B^0_{(s)}$ four-momentum since it is the only variable available describing the $B^0_{(s)}$ meson. Non-perturbative effects are included in the decay constant $f_{B^0_{(s)}}$ such that the hadronic axial current is given by [13]

$$\langle 0|O^{(\prime)}_{10,qq}|B^0_{(s)}\rangle = \mp\frac{1}{2}\langle 0|\overline{q}\gamma^\mu\gamma_5 b|B^0_{(s)}\rangle = \pm i f_{B^0_{(s)}} p^\mu \tag{1.12}$$

The scalar and pseudoscalar hadronic matrix elements are also split into two parts:

$$\langle 0|O^{(\prime)}_{S,P,qq}|B^0_{(s)}\rangle = \frac{m_b}{2}\left[\langle 0|\overline{q}b|B^0_{(s)}\rangle \pm \langle 0|\overline{q}\gamma_5 b|B^0_{(s)}\rangle\right] \tag{1.13}$$

where the m_b factor comes from the definition of the scalar and pseudoscalar operators (see Eq. 1.10), and the factor of 1/2 and the sign in the second term is given by the chiral projector, such that it is plus for $O_{S,P,qq}$ and minus for $O'_{S,P,qq}$. The first term in Eq. 1.13 vanishes due to the pseudoscalar nature of the $B^0_{(s)}$ meson. To obtain the second term, Eq. 1.12 is contracted with the four-momentum of the $B^0_{(s)}$ meson, given by $p_\mu = p_{\overline{q},\mu} + p_{b,\mu}$, and using the relation $p^\mu p_\mu = M_{B^0_{(s)}}$:

$$\begin{aligned}
\langle 0|\overline{q}p_\mu\gamma^\mu\gamma_5 b|B^0_{(s)}\rangle &= i f_{B^0_{(s)}} M^2_{B^0_{(s)}} \\
\langle 0|\overline{q}p_{\overline{q},\mu}\gamma^\mu\gamma_5 b|B^0_{(s)}\rangle - \langle 0|\overline{q}\gamma_5 p_{b,\mu}\gamma^\mu b|B^0_{(s)}\rangle &= i f_{B^0_{(s)}} M^2_{B^0_{(s)}} \\
-(m_{\overline{q}} + m_b)\langle 0|\overline{q}\gamma_5 b|B^0_{(s)}\rangle &= i f_{B^0_{(s)}} M_{B^0_{(s)}} \\
\langle 0|\overline{q}\gamma_5 b|B^0_{(s)}\rangle &= -i f_{B^0_{(s)}} \frac{M^2_{B^0_{(s)}}}{m_{\overline{q}} + m_b}
\end{aligned} \tag{1.14}$$

where the the Dirac equations for particles, $(p_\mu\gamma^\mu - m)\psi = 0$, and antiparticles, $(p_\mu\gamma^\mu + m)\overline{\psi} = 0$, have been used. Notice the minus sign included in the second line of the equation, resulting from the anticommutation properties of the γ matrices. Combining Eq. 1.13 and Eq. 1.14, the hadronic scalar and pseudoscalar matrix element is given by

$$\langle 0|O^{(\prime)}_{S,P,qq}|B^0_{(s)}\rangle = \mp i f_{B^0_{(s)}} \frac{M^2_{B^0_{(s)}}}{2} \frac{m_b}{m_{\overline{q}} + m_b} \tag{1.15}$$

1.2.2 Leptonic matrix elements evaluation

The leptonic matrix element for the scalar and pseudoscalar contributions is obtained for the individual helicities using spinor contraction, resulting in [13]:

$$\begin{aligned} \langle \mu^+\mu^-|O^{(\prime)}_S|0\rangle &= \langle \mu^+\mu^-|\bar{l}l|0\rangle = -M_{B^0_{(s)}}\sqrt{1 - 4\frac{m^2_\mu}{M^2_{B^0_{(s)}}}} \\ \langle \mu^+\mu^-|O^{(\prime)}_P|0\rangle &= \langle \mu^+\mu^-|\bar{l}\gamma_5 l|0\rangle = \eta_\lambda M_{B^0_{(s)}} \end{aligned} \tag{1.16}$$

where η_λ indicates the helicity of the two muon spinors, such that it is $\eta_\lambda = +1$ for right-handed and $\eta_\lambda = -1$ for left-handed spinors.

The leptonic matrix element for $O^{(\prime)}_{10}$ is contracted with the hadronic part given in Eq. 1.12. Due to four-momentum conservation $p^\mu = p^\mu_{\bar{l}} + p^\mu_l$ and the matrix element for $O^{(\prime)}_{10}$ operator is given by:

$$\begin{aligned} \langle \mu^+\mu^-|O^{(\prime)}_{10}|B^0_{(s)}\rangle &= \mp i f_{B^0_{(s)}} p^\mu \langle \mu^+\mu^-|\bar{l}\gamma_\mu\gamma_5 l|0\rangle \\ &= \mp i f_{B^0_{(s)}} \left(\langle \mu^+\mu^-|\bar{l}p^\mu_{\bar{l}}\gamma_\mu\gamma_5 l|0\rangle - \langle \mu^+\mu^-|\bar{l}\gamma_5 p^\mu_l\gamma_\mu l|0\rangle\right) \\ &= \mp 2i f_{B^0_{(s)}} m_\mu \langle \mu^+\mu^-|\bar{l}\gamma_5 l|0\rangle = \mp 2i f_{B^0_{(s)}} m_\mu \eta_\lambda M_{B^0_{(s)}} \end{aligned} \tag{1.17}$$

where the anticommutation property of the Dirac matrices is used in the second line, the Dirac equation for particles and antiparticles is applied in the last line, and the expression for the pseudoscalar leptonic matrix element given in Eq. 1.16 is also used.

1.2.3 $B^0_{(s)} \to \mu^+\mu^-$ decay amplitude

The contributions from the operators derived above are combined using the Hamiltonian from Eq. 1.8 to obtain the matrix element for $B^0_{(s)} \to \mu^+\mu^-$ decays for independent helicities, which is given by

$$\mathcal{M}(B^0_{(s)} \to \mu^+_\lambda \mu^-_\lambda) = -i\frac{G_F \alpha}{\sqrt{2}\pi} f_{B^0_{(s)}} M_{B^0_{(s)}} m_\mu V^*_{tq} V_{tb} \times \left\{ \eta_\lambda \left(C_{10} - C'_{10} + \frac{M^2_{B^0_{(s)}}}{2m_\mu} \left(\frac{m_b}{m_b + m_q} \right) (C_P - C'_P) \right) + \frac{M^2_{B^0_{(s)}}}{2m_\mu} \sqrt{1 - \frac{4m^2_\mu}{M^2_{B^0_{(s)}}}} \left(\frac{m_b}{m_b + m_q} \right) (C_S - C'_S) \right\} \tag{1.18}$$

The value of the Wilson coefficients contributing to the $B^0_{(s)} \to \mu^+\mu^-$ decays are evaluated for the Standard Model by matching the effective amplitude with the full Standard Model one in the low energy region, $E << M_W$. From this matching, the only non-negligible contribution to the Standard Model comes from the C_{10} coefficient, given by $C_{10} = -4.134$ [12]. The C_{10} term in the decay amplitude is proportional to the muon mass, implying that the decay would be forbidden if muons were massless particles, reflecting helicity suppression already mentioned. The scalar and pseudoscalar contributions do not scale with the muon mass, illustrating that helicity suppression is lifted in these beyond the Standard Model cases. Although technically the Higgs penguin can contribute to the $B^0_{(s)} \to \mu^+\mu^-$ process in the Standard Model, its contribution is highly suppressed due to the Higgs coupling to the muons being proportional to $m_\mu/m_W \approx 10^{-3}$. Non-perturbative QCD effects are fully enclosed in the decay constant, $f_{B^0_{(s)}}$, and are calculated to high precision using lattice QCD [14], resulting in a clean computation of the $B^0_{(s)} \to \mu^+\mu^-$ amplitudes in the Standard Model.

Due to the sensitivity of the $B^0_{(s)} \to \mu^+\mu^-$ process to New Physics involving scalar, pseudoscalar and axial vector contributions, it is common to express the $B^0_{(s)} \to \mu^+\mu^-$ decay amplitude relative to the Standard Model C^{SM}_{10}contribution:

$$\mathcal{M}(B^0_{(s)} \to \mu^+\mu^-) = -i\frac{G_F \alpha}{\sqrt{2}\pi} C^{SM}_{10} f_{B^0_{(s)}} M_{B^0_{(s)}} m_\mu V^*_{tq} V_{tb} [\eta_\lambda P + S] \tag{1.19}$$

where

$$\begin{aligned} P &= \frac{C_{10} - C'_{10}}{C^{SM}_{10}} + \frac{M^2_{B^0_{(s)}}}{2m_\mu} \left(\frac{m_b}{m_b + m_q} \right) \frac{C_P - C'_P}{C^{SM}_{10}} \\ S &= \frac{M^2_{B^0_{(s)}}}{2m_\mu} \sqrt{1 - \frac{4m^2_\mu}{M^2_{B^0_{(s)}}}} \left(\frac{m_b}{m_b + m_q} \right) \frac{C_S - C'_S}{C^{SM}_{10}} \end{aligned} \tag{1.20}$$

with $C_{10} = C^{SM}_{10}$ in the Standard Model such that $P = 1$ and $S = 0$.

1.3 Radiative corrections

Photon emission takes place in the $B^0_{(s)} \to \mu^+\mu^-$ process. Two types of emission can be distinguished depending on the particles emitting the photons. *Final-state radiation* is emitted from the final-state muons due to bremsstrahlung, from interactions with the detector material or from the higher order QED corrections to the Feynman diagram. If photons are emitted by the initial-state quarks or the decay mediators, such as the W or Z bosons, the photon emission is known as *initial-state radiation*. Although interference between the two types is possible, it is found to be negligible for $B^0_{(s)} \to \mu^+\mu^-$ decays [15]. The magnitude of the interference contribution is visible in Fig. 1.3, which shows the breakup of the total $B^0_s \to \mu^+\mu^-\gamma$ spectrum in its components [16]. Two separate regions can be distinguished in Fig. 1.3 based on the dominant radiation contribution. Due to this clear separation, the two emission types are treated separately in the $B^0_{(s)} \to \mu^+\mu^-$ analysis.

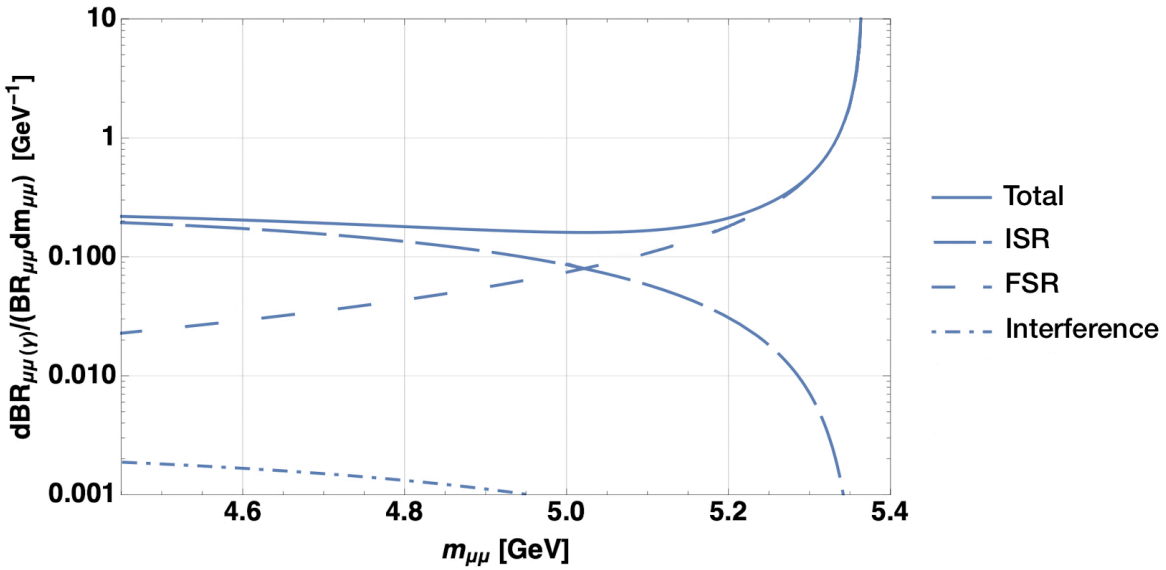


Figure 1.3: Breakup of the mass spectrum for $B^0_s \to \mu^+\mu^-\gamma$ decays with the different components: pure ISR (long-dashed line), pure FRS (medium-dashed line), interference between ISR and FSR (dot-dashed line). The y-axis corresponds to the differential branching fraction scaled by the total branching fraction. The full spectrum is given in solid blue. Adapted from Ref. [16].

1.3.1 Initial-State Radiation (ISR)

Initial-state radiation is dominated by hard photons with relatively large momenta, becoming the dominant photon emission at low dimuon invariant mass. When the photon is emitted by the $B^0_{(s)}$ meson, the helicity suppression is lifted, and the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decay is sensitive

to different Wilson coefficients than the $B^0_{(s)} \to \mu^+\mu^-$ decay, such as the C_7 from the QED penguin contributions. Therefore, the initial-state radiation should not be included in the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction as it de facto is a different decay process. In this analysis, the $B^0_s \to \mu^+\mu^-\gamma$ decay involving only initial-state radiation is treated as an independent signal component in the dimuon invariant mass spectrum, and its branching fraction assuming $m_{\mu^+\mu^-} > 4.9\,\text{GeV}/c^2$ is also measured, together with the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fraction.

1.3.2 Final-State Radiation (FSR)

Photons emitted by the final-state muons are generally soft photons. Consequently, final-state radiation is the primary source of photon emission in the high invariant mass region, closer to the $B^0_{(s)} \to \mu^+\mu^-$ peaks. The FSR transition is theoretically similar to the $B^0_{(s)} \to \mu^+\mu^-$ process in so far that they are sensitive to the same Wilson coefficients and with the same decay amplitude. Hence, FSR is an integral part of the $B^0_{(s)} \to \mu^+\mu^-$ processes and is included in the measurement of their branching fractions.

From the theoretical approach, the SM prediction of the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction is evaluated without including such final-state radiative contributions [15]. The radiative branching fraction, including final-state radiation with total energy ΔE, can be determined as [9, 15]:

$$\mathcal{B}^{phys}(\Delta E) = \mathcal{B}^0 \times \Omega(\Delta E) \tag{1.21}$$

where $\mathcal{B}^0$ is the non-radiative branching fraction and $\Omega(\Delta E)$ is the radiative factor [9] given by

$$\Omega(\Delta E) = \left(\frac{2\Delta E}{M_{B^0_{(s)}}}\right)^{-\frac{2\alpha_{em}}{\pi}\left(1+\ln\frac{m_\mu^2}{M^2_{B^0_{(s)}}}\right)} \tag{1.22}$$

with α_{em} as the electromagnetic coupling constant, and ΔE giving the experimental signal region around the $B^0_{(s)}$ mass peak in terms of the dimuon invariant mass. The fully inclusive FSR branching fraction has been found to be in agreement with the non-radiative value up to a 1% correction [9, 15].

From the experimental side, final-state radiation is included in simulation using a dedicated software known as Photos [17] (see Chapter 2). The branching fraction is measured by performing a fit over a specific invariant mass window. The FSR effect is included within the efficiency correction, in particular through the selection efficiency, which includes the requirement for the specific mass window. The selection efficiency is evaluated from the simulation samples. Hence, the measured branching fraction corresponds to the fully inclusive radiative branching fraction.

1.4 Theory versus experiment

The branching fraction is one of the observables used to probe the Standard Model and search for New Physics using $B^0_{(s)} \to \mu^+\mu^-$ decays. It gives the fraction of $B^0_{(s)}$ mesons decaying into two oppositely charged muons relative to the total number of decaying $B^0_{(s)}$ mesons. Although the definition holds for theory and experiment, a few differences in their evaluation must be considered when comparing the experimental and theoretical branching fraction evaluations, especially in the case of the $B^0_s \to \mu^+\mu^-$ decays.

In the experiment, the $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ mesons can be distinguished using the so-called *flavour-tagging* technique. Because the efficiencies associated with this technique are low and due to the rareness of the $B^0_{(s)} \to \mu^+\mu^-$ process, the $B^0_{(s)} \to \mu^+\mu^-$ analysis does not use flavour-tagging in order to include all the $B^0_s \to \mu^+\mu^-$ candidates in the analysis. Instead, no distinction is made between $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ initial states such that the *untagged decay rate* is defined as:

$$\langle \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-) \rangle = \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-) + \Gamma(\bar{B}^0_{(s)}(t) \to \mu^+\mu^-) \tag{1.23}$$

1.4.1 $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ mixing

The mixing between the $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ states is a particular characteristic of the $B^0_{(s)}$ meson system that must be considered when evaluating their time evolution. This mixing arises from the existence of two different eigenstate types for a hadronic particle: the flavour and the mass eigenstates. Flavour eigenstates have a well-defined quark content and are eigenstates of the strong and electromagnetic interactions. Nevertheless, the mesons propagate on the mass eigenstate basis, with well-defined lifetimes and masses. In the case of the $B^0_{(s)}$ meson, the $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ states are flavour eigenstates, while the light and heavy mass eigenstates, $B^0_{(s),H}$ and $B^0_{(s),L}$, are a linear combination of the flavour eigenstates. Therefore, the untagged decay rate can be written in terms of the heavy and light mass eigenstates as

$$\begin{aligned}
\langle \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-) \rangle =& \Gamma(B^0_{(s),H} \to \mu^+\mu^-) e^{-\Gamma_H t} + \Gamma(B^0_{(s),L} \to \mu^+\mu^-) e^{-\Gamma_L t} \\
=& (\Gamma(B^0_{(s),H} \to \mu^+\mu^-) + \Gamma(B^0_{(s),L} \to \mu^+\mu^-)) \\
& \times e^{-t/\tau_{B^0_{(s)}}} (\cosh(y_s t/\tau_{B^0_{(s)}}) + A^{\mu\mu}_{\Delta\Gamma_s} \sinh(y_s t/\tau_{B^0_{(s)}})) \\
=& (\Gamma(B^0_{(s)} \to \mu^+\mu^-) + \Gamma(\bar{B}^0_{(s)} \to \mu^+\mu^-)) \\
& \times e^{-t/\tau_{B^0_{(s)}}} (\cosh(y_s t/\tau_{B^0_{(s)}}) + A^{\mu\mu}_{\Delta\Gamma_s} \sinh(y_s t/\tau_{B^0_{(s)}}))
\end{aligned} \tag{1.24}$$

where $\tau_{B^0_{(s)}}$ is defined as the inverse of the average decay rate of the light and heavy mass eigenstates. Two new variables appear in the time-dependent factors from the untagged

decay rate. The first one, $y_{d(s)}$, is the decay time asymmetry between the mass eigenstates, a property of the B_s^0 meson system which is independent of the final state:

$$y_{d(s)} = \frac{\Gamma_{d(s),L} - \Gamma_{d(s),H}}{\Gamma_{d(s),L} + \Gamma_{d(s),H}} \tag{1.25}$$

where $\Gamma_{d(s),L}$ and $\Gamma_{d(s),H}$ are the decay rates of the light and heavy mass eigenstates. The second term, $A^{\mu\mu}_{\Delta\Gamma_s}$, is known as the decay rate asymmetry and is dependent on the specific decay process:

$$\mathcal{A}^{\mu^+\mu^-}_{\Delta\Gamma} = \frac{\Gamma(B^0_{(s),H} \to \mu^+\mu^-) - \Gamma(B^0_{(s),L} \to \mu^+\mu^-)}{\Gamma(B^0_{(s),H} \to \mu^+\mu^-) + \Gamma(B^0_{(s),L} \to \mu^+\mu^-)} \tag{1.26}$$

where $\Gamma(B^0_{(s),H} \to \mu^+\mu^-)$ and $\Gamma(B^0_{(s),L} \to \mu^+\mu^-)$ are the decay rates to two muons of the heavy and light mass eigenstates [18]. From its definition, the values of the decay rate asymmetry range from +1 to -1 depending on the contribution of each mass eigenstate to the two muons final-state.

The final ingredient in the evaluation of the untagged decay rate is the instantaneous decay rate for the $B^0_{(s)} \to \mu^+\mu^-$ and $\bar{B}^0_{(s)} \to \mu^+\mu^-$ processes. These are obtained using Fermi's Golden rule [19] summing over helicity states of the final-state muons as:

$$\Gamma(B^0_{(s)} \to \mu^+\mu^-)\Big|_{t=0} = \frac{1}{16\pi}\frac{1}{M_{B^0_{(s)}}}\sqrt{1 - \frac{4m_\mu^2}{M^2_{B^0_{(s)}}}} \sum_{\lambda=R,L} |\mathcal{M}(B^0_{(s)}(t) \to \mu^+_\lambda \mu^-_\lambda)|^2 \tag{1.27}$$

where $\mathcal{M}(B^0_{(s)}(t) \to \mu^+_\lambda \mu^-_\lambda)$ is the amplitude of the $B^0_{(s)} \to \mu^+\mu^-$ process for individual helicities given in Eq. 1.19. The decay amplitude for the $\overline{B}^0_{(s)}(t) \to \mu^+\mu^-$ is obtained analogously. Plugging Eq. 1.27 into Eq. 1.24, the untagged decay rates takes the form:

$$\begin{aligned}\langle \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-)\rangle =& \frac{G_F^2\alpha^2}{8\pi^3}|C_{10}^{SM} V_{t(d,s)} V_{tb}^*|^2 f^2_{B^0_{(s)}} M_{B^0_{(s)}} m_\mu^2 \sqrt{1 - \frac{4\,\mathrm{m}_\mu^2}{M^2_{B^0_{(s)}}}}(|P|^2 + |S|^2) \\ & \times e^{-t/\tau_{B^0_{(s)}}}(\cosh(y_s t/\tau_{B^0_{(s)}}) + A^{\mu\mu}_{\Delta\Gamma_s}\sinh(y_s t/\tau_{B^0_{(s)}}))\end{aligned} \tag{1.28}$$

where P and S are given by Eq. 1.20 and are defined such that in the Standard Model $P = 1$ and $S = 0$.

1.4.2 Relation between experimental and theoretical definitions

The theoretical branching fraction is calculated from the CP-average decay rate of the flavour eigenstates, $B^0_{(s)}$ and $\bar{B}^0_{(s)}$. It is evaluated assuming $t = 0$ in Eq. 1.28, such that it is not affected

by the $B^0_{(s)}$-$\bar{B}^0_{(s)}$ mixing, only by the $B^0_{(s)}$ mean lifetime [18]

$$\mathcal{B}_{th}(B^0_{(s)} \to \mu^+\mu^-) = \frac{\tau_{B^0_{(s)}}}{2} \langle \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-) \rangle \Big|_{t=0} \tag{1.29}$$

$$= \frac{\tau_{B^0_{(s)}} G_F^2 \alpha^2}{16\pi^3} |C_{10}^{SM} V_{t(d,s)} V_{tb}^*|^2 f^2_{B^0_{(s)}} M_{B^0_{(s)}} m_\mu^2 \sqrt{1 - \frac{4\,\mathrm{m}_\mu^2}{M^2_{B^0_{(s)}}}} (|P|^2 + |S|^2) \tag{1.30}$$

Experimentally, the branching fraction is measured from the total detected $B^0_{(s)} \to \mu^+\mu^-$ candidates generated in the proton-proton collision, regardless of their decay time. Since the information on the lifetime of the particle is ignored, the experimental branching fraction is defined as [18]

$$\mathcal{B}_{exp}(B^0_{(s)} \to \mu^+\mu^-) = \frac{1}{2} \int_0^\infty \langle \Gamma(B^0_{(s)}(t) \to \mu^+\mu^-) \rangle \, dt \tag{1.31}$$

Using the untagged decay rate computed in the previous subsection, given in Eq. 1.28, and performing the integral, the experimental branching fraction is found to be related to the theoretical one as [18]:

$$\mathcal{B}_{exp}(B^0_{(s)} \to \mu^+\mu^-) = \left[\frac{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_{d(s)}}{1 - y^2_{d(s)}} \right] \mathcal{B}_{th}(B^0_{(s)} \to \mu^+\mu^-) \tag{1.32}$$

1.4.3 Branching fraction predictions

The relation between the definitions of experimental and theoretical branching fractions depends on the decay time asymmetry and decay rate asymmetry parameters. In the case of the B^0 meson, the decay rates of the heavy and light mass eigenstates are approximately the same, resulting in a decay time asymmetry of zero: $y_d = 0$. Consequently, the experimental and theoretical branching fraction definitions are equal, as given from Eq. 1.32. The $B^0 \to \mu^+\mu^-$ branching fraction as predicted in the Standard Model is [9]

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10} \tag{1.33}$$

The case for the B^0_s meson system is more complex. Because the decay time asymmetry has been measured to be $y_s = 0.068 \pm 0.004$ [1], the role of the $A^{\mu\mu}_{\Delta\Gamma_s}$ becomes significant in Eq. 1.32. As shown in Eq. 1.26, the $A^{\mu\mu}_{\Delta\Gamma_s}$ parameter quantifies the decay rate asymmetry of the heavy and light mass eigenstates. Due to angular momentum conservation, the final state of the $B^0_{(s)} \to \mu^+\mu^-$ process is CP-odd since the two final-state muons have opposite spins and zero total angular momentum. Regarding the B^0_s meson system, the heavy (light) mass eigenstate is also identified as the CP-odd (CP-even) eigenstate. In the Standard Model, no CP violation is allowed in this decay, meaning that only the heavy mass eigenstate decays into two muons and $A^{\mu\mu}_{\Delta\Gamma_s} = +1$. Taking this into account in Eq. 1.32, the Standard Model

prediction for the $B_s^0 \to \mu^+\mu^-$ branching fraction is given by [9]

$$\begin{aligned} \mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.66 \pm 0.14) \times 10^{-9} \\ \mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.03 \pm 0.05) \times 10^{-10} \end{aligned} \tag{1.34}$$

Although outside of the scope of this thesis, it is important to highlight that the decay rate asymmetry $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ is essentially unknown. If CP violation occurs during the decay, the light mass eigenstate could also decay into two muons, and the decay rate asymmetry would then be different to its Standard Model value of +1. Therefore, $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ is also sensitive to New Physics scenarios. Its measurement is possible through the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime, defined as the average decay time of the $B_s^0 \to \mu^+\mu^-$ candidates in the experiment and given by [20]

$$\tau_{\mu^+\mu^-} = \frac{2\tau_{B_s^0} A_{\Delta\Gamma_s}^{\mu\mu} y_s + \left(1 + y_s^2\right) \tau_{B_s^0}}{\left(1 - y_s^2\right) + A_{\Delta\Gamma_s}^{\mu\mu} y_s \left(1 - y_s^2\right)} \tag{1.35}$$

The effective lifetime has recently been measured by the LHCb experiment [21, 22] and agrees with the Standard Model prediction.

2 | LHCb detector

Particle accelerators boost charged particles close to the speed of light to make them collide, either against a fixed target or against other particles. Thousands of different particles are created in these collisions whose behaviour can shed light on questions about the fundamental building blocks of matter and their interactions. The largest and most powerful accelerator is the Large Hadron Collider (LHC) at CERN, located near Geneva [23]. The LHC is a 27 km superconducting synchrotron designed to accelerate and collide proton beams, recreating the conditions of the early universe. The LHC started operating in 2011, and since then, two data-taking periods have taken place, known as Run 1 and Run 2. The Run 1 dataset consists of the collisions registered in 2011 and 2012, while Run 2 comprises the data gathered from 2015 to 2018. The centre-of-mass energies of the proton-proton (pp) collisions have increased with time, starting from 7 TeV in 2011 and 8 TeV in 2012 to 13 TeV in Run 2.

The proton bunches accelerated at the LHC are made to collide in four experiments located on the LHC ring: CMS, ATLAS, ALICE and LHCb. These experiments are designed to focus on diverse parts of fundamental particle physics research. The CMS and ATLAS experiments are general-purpose detectors studying a wide range of topics, from the properties of the Higgs boson to the existence of dark matter [24, 25]. The ALICE experiment is dedicated to heavy-ion physics and the study of quark-gluon plasma by analyzing the collisions involving heavy ions, mainly lead, accelerated in the LHC ring [26]. Finally, the LHCb experiment is designed to study heavy flavour physics involving beauty and charm hadrons, primarily focusing on CP violation and rare decays such as $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays, searching for phenomena beyond the Standard Model [27]. The entire data gathered by the LHCb detector in Run 1 and Run 2 is used in this analysis.

The LHCb detector is a forward single-arm spectrometer covering a pseudorapidity region of $2 < \eta < 5$, with the pseudorapidity defined using cylindrical coordinates as

$$\eta \equiv -\ln\tan(\theta/2) \tag{2.1}$$

where θ is the polar angle relative to the z axis. The forward geometry is motivated by the correlation of the b and $\overline{b}$-hadron production at high energies, resulting in a boosted $b\overline{b}$

pair propagating in the same forward (or backward) region of space along the beam axis. A sketch on the $y - z$ plane of the LHCb detector illustrating all subdetectors is displayed in Fig. 2.1, where the horizontal direction along the beam corresponds to the z-axis and the vertical to the y-axis. The x-axis completes a right-handed coordinate system.

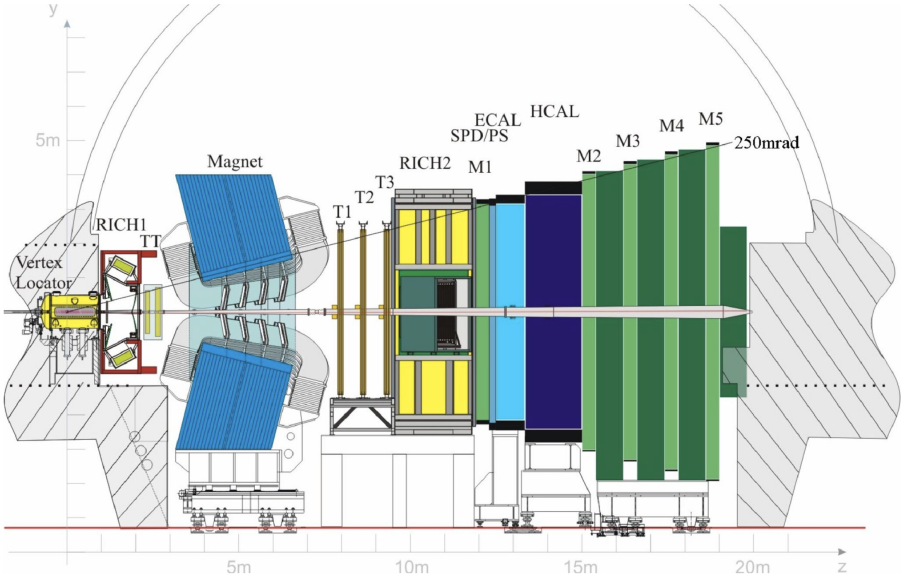


Figure 2.1: Layout of the LHCb detector with the different constituent subdetectors (taken from Ref. [28]). Visible are the Vertex Locator (VELO), two ring-imaging Cherenkov detectors (RICH1 and RICH2), the magnet, the Tracker Turicencis (TT) and the T1-T2-T3 tracking stations, the electromagnetic and hadronic calorimeters (ECAL and HCAL) and the five muon chambers (M1-M5).

The information from the different subdetectors is combined to reconstruct the particles produced in the pp collisions. The trajectories of the particles are reconstructed with the tracking system, presented in Sec. 2.1, and their momenta can be measured. The particle types are identified using a combination of various subdetectors, presented in Sec. 2.2, designed to distinguish different particle types as well as to measure their energies. Since protons are composite particles, a hadronic environment is generated after their collision, producing many different particles from which a significant fraction decay into other particles. Due to the huge data volume, only interesting events are selected through a fast and efficient online event selection, known as trigger, and stored on a disk for a detailed analysis. The trigger system consists of different levels combining software and hardware information, as detailed in Sec. 2.3. Simulation samples are also generated as explained in Sec. 2.4 and used for various estimations and calibrations in the analysis.

2.1 Tracking

The tracking system enables the reconstruction of the trajectories of the particles created in the pp collisions and propagating through the LHCb detector. It consists of the Vertex Locator (VELO), situated only 7 mm away from the LHC beams, and the tracking stations, one located upstream of the magnet and three downstream. A dipole magnet with an integrated field of 4 Tm deflects the trajectories of charged particles, allowing for the determination of their momenta. Combining the information in the tracking stations before and after the magnet allows for a precise reconstruction of the trajectories of charged particles and their momenta.

From the information in the tracking system, the tracks in the LHCb detector are classified depending on their lever arm in the spectrometer. The different track types are sketched in Fig. 2.2. Long tracks are mostly used in all the analyses, including the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, while VELO tracks are used to reconstruct primary vertices and in the real-time trigger selection (see Sec. 2.3). Upstream tracks are relevant for slow particles whose trajectories are largely deflected by the magnet, while downstream tracks are important for the detection of long-lived particles.

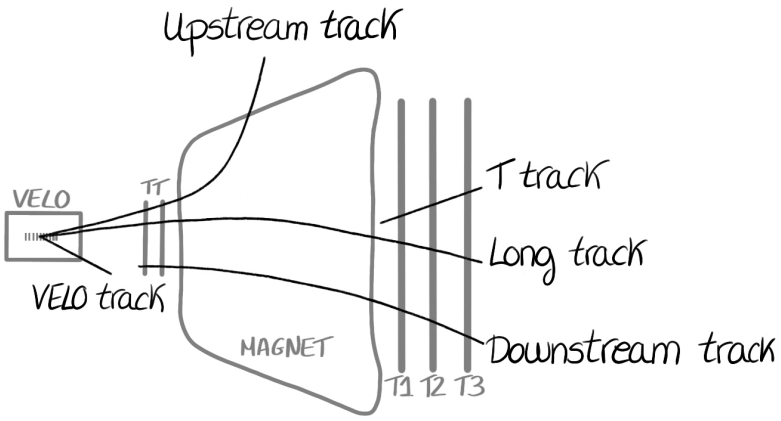


Figure 2.2: Illustration showing the track types in the LHCb. Figure adapted from [29].

2.1.1 Vertex Locator (VELO)

The VELO is a silicon multistrip tracking detector surrounding the pp interaction region. It measures the trajectories of the particles near the primary vertices (PV), or the interaction point where the particles are produced in the pp collision, providing the information needed for their reconstruction. It also allows making a clear distinction between the primary vertices and the vertices where unstable particles decay, known as secondary vertices.

Because a specific characteristic of the b and c-hadron decays is their displaced decay vertices from the primary vertex, a clear distinction between the two is essential to select the events containing such decays. The displacement between the vertices is quantified by the impact parameter, or the perpendicular distance between the trajectory of a particle and the interaction region, as exemplified for the $B_s^0 \to \mu^+\mu^-$ decay in Fig. 2.3.

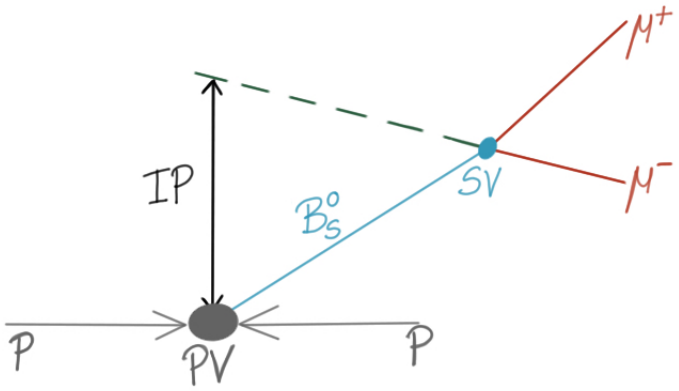


Figure 2.3: Definition of the impact parameter for the final-state μ^- in the case of the $B_s^0 \to \mu^+\mu^-$ decay.

The VELO consists of 42 semicircular modules placed around the beam axis. Each module contains two sensors, the R and Φ sensors, measuring the radial and azimuthal coordinates, respectively. The z coordinate is obtained from the known position of the sensors in the experiment. The only difference between the two sensors is the arrangement of the silicon strips, shown in Fig. 2.5. The strips are read out by the stripping routes, which take the information collected from the sensors to the electronics placed outside the detector.

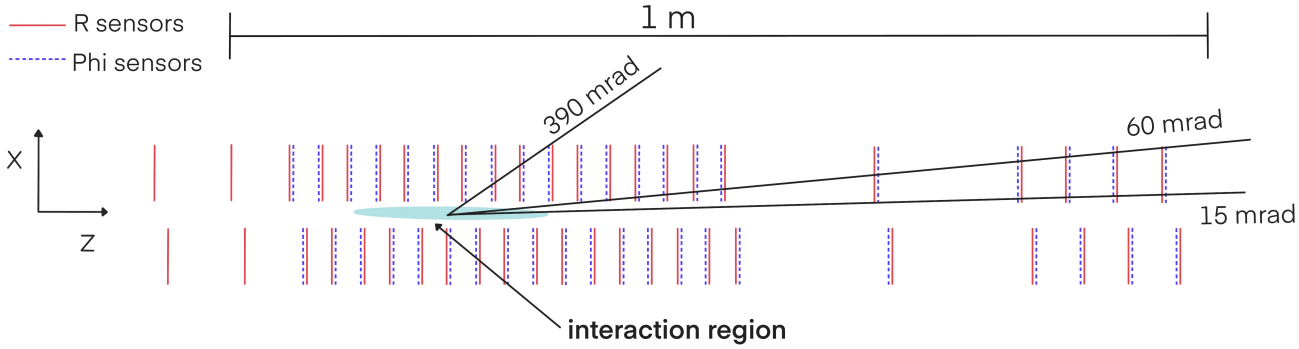


Figure 2.4: Cross section in the $x - z$ plane of the VELO [30]. The Φ and R sensors are displayed in dashed and solid lines, and the pp interaction region is highlighted in blue. Taken from Ref. [30].

The inner tip of the sensors is located only 7 mm away from the LHC beams. Because this distance is smaller than the beam aperture required by the LHC injection, the modules

are mounted in two halves on the left and right sides of the beam such that each half is retracted 3 cm during the beam injection and moved back to its position after the beam is stabilized. The open and closed positions are also sketched in Fig. 2.5. The so-called RF foils separate the LHC beam vacuum from the vacuum created around the VELO sensors. These corrugated foils serve as a two-sided protection: they protect the LHC vacuum from the outgassing of the VELO modules and the VELO sensors from the radiofrequency radiation induced by the moving protons in the beams. The left ($-x$) and right ($+x$) halves of the VELO overlap slightly for alignment purposes.

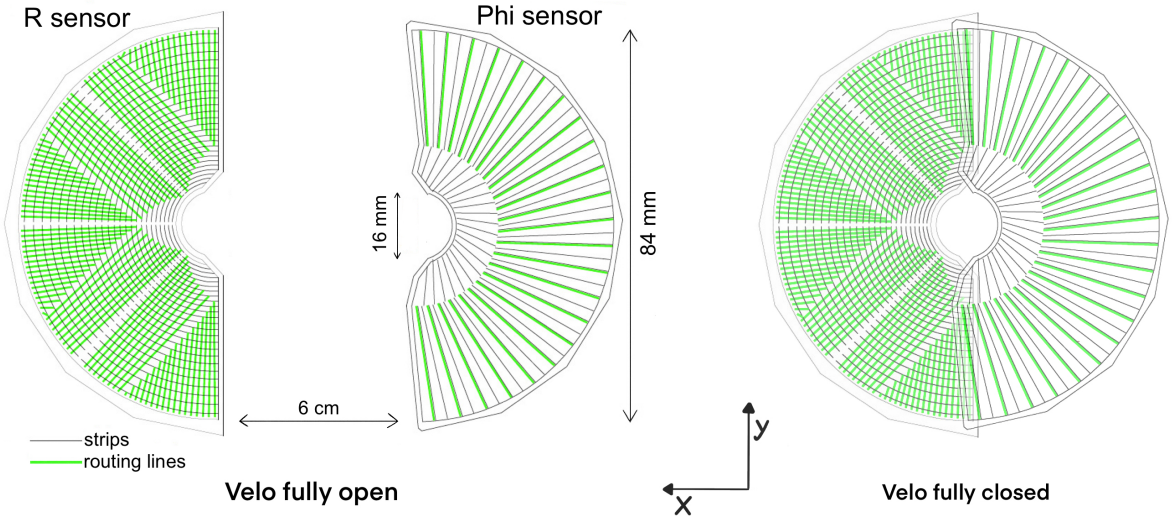


Figure 2.5: Representation of the front face of a VELO module in the open and close positions. The different layout of the R and Φ sensors is displayed: for the R sensor, the strips are placed into four circular segments, while for the Φ sensors, inner and outer zones are implemented. Figure adapted from Ref. [30]

The particles generated in the interaction point pass through several sensors while travelling through the VELO. A minimum of three hits in different sensors are required to consider a particle track reconstructible. With this requirement, the track-finding efficiency of the VELO is above 98% [30]. Due to its geometry and positioning, the VELO achieves a primary vertex resolution of 13 μm in the transverse plane and 71 μm along the beam axis for vertices with an average of 25 tracks [30]. An impact parameter resolution of less than 35 μm is achieved for tracks with transverse momentum larger than 1 GeV/c [30].

The information provided by the VELO is essential in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, firstly in the separation of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates from the background components, mainly the combinatorial background. This type of background, the most abundant one in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, results from the combination of two random muons generated in the pp interaction region. To reject this background, the impact parameter of the muon candidates relative to the primary vertex is used since the random combinatorics are expected to

originate mainly from the primary vertices. In addition, the lifetime of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidate, determined from the distance between the primary and secondary vertices, is used to select the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates. Moreover, the information obtained from the VELO is also used to determine the quality of the decay vertices, a quantity used in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis since only candidates with final-state muons forming a good-quality vertex are selected.

2.1.2 Tracking stations

Together with the VELO, four tracking stations form the LHCb tracking system. They are located on both sides of the magnet to measure the deflection of the trajectories of charged particles caused by the magnetic field and, from their curvature, infer the momenta of the charged particles. Out of the four tracking stations, one is located upstream of the magnet, known as Tracker Turicensis and the remaining three, the T1-T2-T3 stations, just after the magnet.

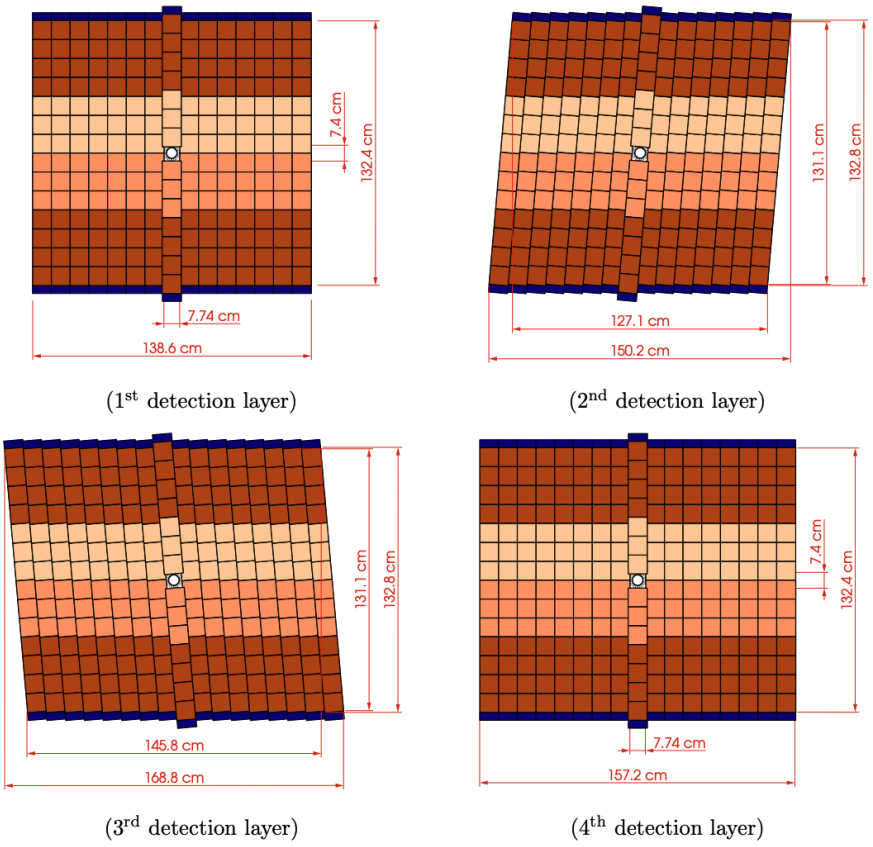


Figure 2.6: Layout of the x-u-v-x geometry of detection layers in the Track Turicencis and the TT stations. Taken from Ref. [31].

The Tracker Turicensis is a silicon microstrip detector located before the magnet, after

the VELO. It covers an active area of 8 m^2 and is formed by one station consisting of four detector layers [27]. The layers are arranged in an x-u-v-x geometry, shown in Fig. 2.6, with the strips in the first and last layers positioned vertically and the second and third layers rotated with a stereo angle of +5° and -5°, respectively. With this geometry, the sensitivity to the y coordinate is enough to reduce the combinatorics generated from the wrong track matching in the y plane while having a large momenta resolution.

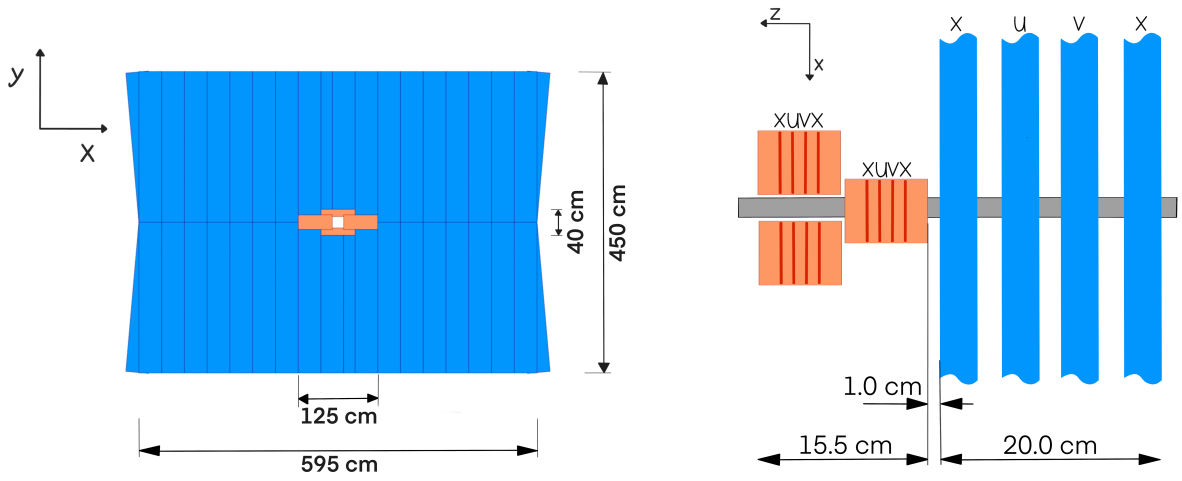
The three tracking stations positioned after the magnet are the T1, T2 and T3 stations. Each station is formed by the Inner Tracker (IT), located in the central region, surrounded by the Outer Tracker (OT). The IT uses silicon multistrip covering a cross-shaped area of 125 cm wide and 40 cm high centred around the beam axis. It consists of four detector boxes located above, below and on each side of the beam, each formed by four layers of material arranged in the x-u-v-x geometry, as explained for the TT. Although the IT only covers 1.3% of the whole tracking stations located downstream of the magnet, approximately 20% of the charged particles traverse it because they are predominantly produced collinearly to the beam direction [32].

A coarser detector is used in the rest of the tracking stations since the particle flux is lower than in the innermost part. The OT is built using drift tubes with a diameter of around 5 mm and covers an area of 595 cm wide and 480 cm high. When a charged particle passes through the drift tubes, it interacts with the gas inside them, generating electron-ion pairs. Due to the electric field inside the tube, the electrons drift to the central wire, connected to the front-end electronics. The position of the track is obtained from the drift time of the electrons to reach the central wire.

One of the discriminating variables used in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis is the dimuon invariant mass, evaluated from the four-momenta of the final-state muon candidates. The precise measurement of the momenta of the muon candidates obtained from the tracking stations allows for a good dimuon invariant mass resolution and a large background rejection in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis.

2.2 Particle Identification

Particle identification (PID) is essential to distinguish among decays and separate the signal and background components in physics analyses. At LHCb, the PID system is formed by three complementing types of detectors: the two RICH detectors, the electromagnetic and hadronic calorimeters and the muon chambers. The information from these detectors is combined with the momenta obtained from the tracking stations and used to effectively determine the mass of the particle.



(a) Front view.

(b) Top view.

Figure 2.7: Front (left) and top (right) view of one of the TT stations showing the IT and OT in orange and blue, respectively. The lateral dimensions in the top view are not scaled. Adapted from Ref. [32].

2.2.1 Ring-imaging Cherenkov detectors

There are two ring-imaging Cherenkov detectors at the LHCb experiment, RICH1 and RICH2, positioned upstream and downstream of the magnet covering particle identification for different momentum ranges. Cherenkov radiation is emitted by charged particles when travelling faster than the speed of light in a specific medium, such as the one inside the RICH detectors. The emitted radiation is transmitted in the shape of a cone whose angle depends on the velocity of the particle causing it. Using the known particle momentum measured by the tracking system, the mass of the particle can be estimated, leading to the particle type. Because the rest of the subdetectors cannot differentiate among hadron types, the RICH detectors are essential in the distinction among protons, pions and kaons.

The upstream RICH1 detector is located right after the VELO, before the TT stations. It is filled with C_4F_{10} and provides identification of charged particles with momenta ranging between 2 GeV/c and 60 GeV/c [33]. High momenta charged particles are identified by the RICH2 detector, placed downstream of the magnet after the T3 station. The radiator used in this case is CF_4, covering the identification of charged particles with momenta between 15 GeV/c and 100 GeV/c [33].

Identification of pions and kaons is important in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis to avoid their misidentification as muons and reduce misidentified backgrounds, mainly caused by $B^0_{(s)} \to h^+h'^-$ decays. Combining the information from RICH1 and RICH2, the kaon identification efficiency is found to be 85% with a pion misidentification rate of 3% [33]. The

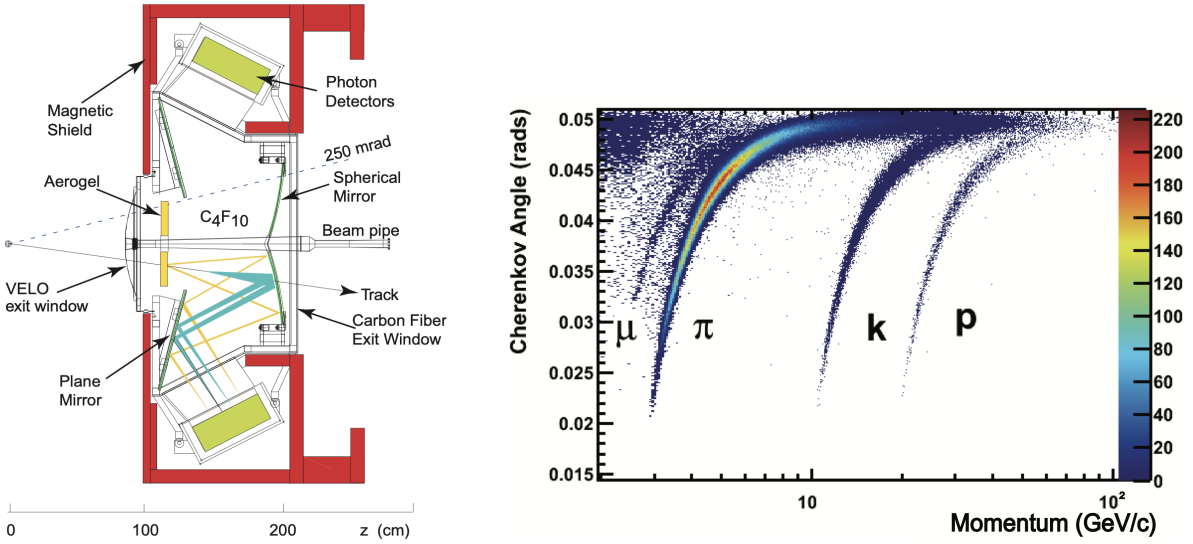


Figure 2.8: On the left, the schematic $x - y$ cross-section of RICH1. Taken from Ref. [28]. On the right, the reconstructed Cherenkov angle as a function of the track momentum for different particle types in the RICH1. Figure taken from Ref. [27].

probability of misidentifying a kaon (pion) as a muon is estimated to be 1.111 (1.025)%, with around 60% of these misidentifications caused by decays in flight due to the resulting muon being detected in the muon chambers [34].

2.2.2 Calorimeter system

The calorimeter system [35] is used to distinguish among electrons, photons and hadrons and measure the energies and position of these particles. It is formed by four components located downstream of the magnet: a scintillating pad detector (SPD), a preshower (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The response of each subdetector to the various particle types is different, enabling their distinction. An illustration of the energy deposits of electrons, photons, hadrons and muons in the various parts of the calorimeter system is shown in Fig. 2.9.

The SPD is a high-granularity scintillating pad detector used to separate photons and electrons as it only interacts with charged particles. It is separated from the PS by a wall of 15 mm of lead converter [37] to induce showers of traversing photons. These induced showers reach the PS, where their energy is deposited. The different radiation lengths and shower splittings left by electromagnetic and hadronic particles in the PS facilitate their distinction.

A sampling electromagnetic calorimeter is placed after the PS detector. The ECAL mea-

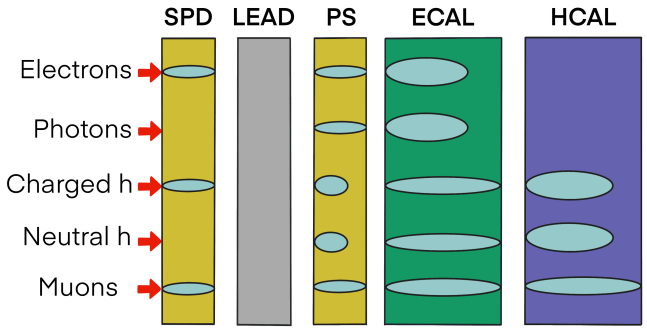
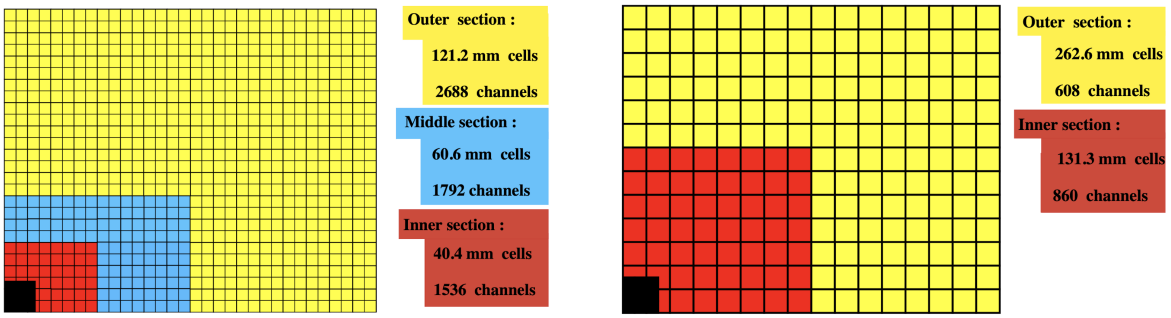


Figure 2.9: Sketch of the energy deposited by electrons, photons, hadrons and muons in different parts of the calorimeter system. Adapted from Ref. [36].

sures the energies from particles interacting electromagnetically, such as electrons, protons or neutral pions. It is composed of alternating layers of 4 mm of a lead converter and scintillating pads. When an electromagnetic interacting particle traverses the lead converter, it initiates a particle shower. The shower induces the emission of photons in the scintillating pad, which are read out to measure the induced light, used as a proxy for the total energy deposited by the incident particles [37]. The particle density hitting the calorimeter varies depending on the proximity to the beam axis, with the highest densities in the closest area to the LHC beam. To accommodate these various densities, the ECAL is segmented into three regions with a different number of readout cells, as shown in Fig. 2.10a.



(a) Electromagnetic calorimeter.

(b) Hadronic calorimeter.

Figure 2.10: Segmentation of the electromagnetic calorimeter (left) and the hadronic calorimeter (right) to accommodate the varying particle density. The dark region represents the beam axis. Taken from Ref. [28].

The hadronic calorimeter is placed after the ECAL to measure the energy of hadrons, both neutral and charged. It is a sampling calorimeter consisting of alternating layers of iron converter and scintillating pads. To accommodate the varying particle density, the HCAL is segmented into two regions, displayed in Fig. 2.10b.

The electrons, photons and hadrons leave energy deposits in various subdetectors of the calorimeter systems. The energy of the particle is measured by accounting for the energy deposits in the various parts of the calorimeter system, except from the SPD, since its readout only indicates whether a charged particle has passed through it.

The information from the calorimeter is also used in the L0 trigger to decide whether an event is discarded or stored for further analysis (see Sec. 2.3). Although the final-state muons from the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ are usually detected in the muon chambers, they can leave energy deposits in the calorimeter large enough to trigger the event.

2.2.3 Muon system

Only muons and neutrinos traverse the whole calorimeter system and reach the muon chambers. Although muons mainly interact electromagnetically with matter, their bremsstrahlung emission is negligible due to their larger mass compared to that of the electrons. Furthermore, their decay, mediated by the weak interaction, is slow such that they propagate large enough distances to reach the muon chambers before decaying. Neutrinos are weakly interacting particles propagating through the whole detector without leaving a trace and, hence, cannot be detected. Consequently, only muons leave hits in the muon chambers located after the calorimeter system, a clean signature that facilitates their identification.

The muon system consists of five stations, M1-M5, placed downstream of the magnet, surrounding the beam axis and covering a total active area of 435 m^2. The muon stations are rectangular-shaped, with their transverse dimension increased with the distance from the interaction point. Multiwire proportional chambers are used in the muon stations to detect traversing muons, which ionize the gas inside the chamber. Due to the presence of an electric field, the induced electrons drift towards the wires located inside the chamber where they are collected. The presence of hits in the muon stations is used in the hardware trigger and in the measurement of the momenta of the muon candidates.

The first muon station, M1, is located upstream of the calorimeter system, at 12 meters from the interaction point [38]. It improves the precision of the transverse momentum measurement of the muon candidates used in the hardware trigger before they scatter in the calorimeter material. The rest of the stations, M2-M5, are placed after the calorimeter system and are alternated with 80 cm thick layers of iron absorbers. The main goal of stations M1-M3 is to track the direction of the muon candidate and calculate its transverse momentum with a resolution of 20% [38]. Stations M4 and M5 are mainly used to detect the presence of highly penetrating muons since only muons with a minimum momentum of 6 GeV/c are expected to cross the five muon stations [28].

As is the case for the calorimeters, the particle density varies across the surface of the muon stations. This phenomenon is taken into account in the channel occupancy used in

different parts of the stations. Consequently, each muon station is divided into four regions, R1-R4, with respect to the beam axis such that the channel occupancy is approximately the same in all the regions. A side view of the muon system with the position of the muon stations and the four regions in each station is displayed in Fig. 2.11.

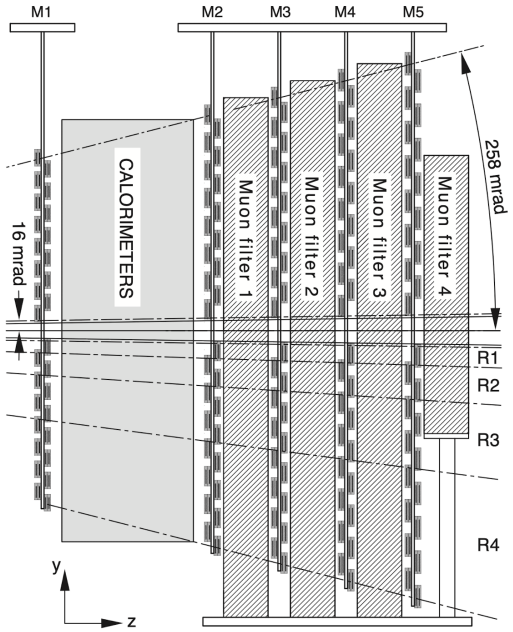


Figure 2.11: Side view of the muon stations with the four separated regions in each station. Taken from Ref. [28]

2.2.4 PID variables

Each of the PID subdetectors interacts differently with electrons, photons, hadrons and muons and returns a likelihood relative to a specific particle hypothesis for each reconstructed track. In the case of the RICH and the calorimeter system, the likelihood is given relative to the pion hypothesis. In contrast, for the muon stations, the likelihood ratio is determined between the muon and non-muon hypotheses [27].

The likelihoods of the various subdetectors are combined to create PID variables, which are applied in the selection procedure of physics analyses to separate the signal from the background. Two different PID variables are defined depending on the approach followed in the combination of the subdetector likelihoods: the *combined delta log-likelihoods (ΔLL)* and the *ProbNN* [27]. The ΔLL results from the linear combination of the logarithms of the subdetector likelihood ratios and is given relative to the pion hypothesis. For a reconstructed track, it is interpreted as how likely a specific mass hypothesis is relative to the pion hypothesis.

The ΔLL does not account for possible correlations among subdetectors or the information provided by the tracking systems. To take into account these correlations, the tracking information is combined with the PID likelihoods of each subdetector using neural networks, which result in the *ProbNN* variables for the specific particle hypothesis. These variables range from 0 to 1 and can be interpreted as a pseudo-probability of a particle track to have a specific particle hypothesis, i.e. the $ProbNN_K$ is the probability of the track to be a kaon.

A binary PID variable is built exclusively from the information of the muon systems: the *isMuon* variable [34]. It indicates whether a reconstructed track meets the minimum requirements to be considered a muon candidate. The *isMuon* is defined based on the number of muon stations with consistent hits, found by extrapolating the trajectories of the tracks to the muon stations inside an $x - y$ region. Due to the wall of lead interleaved with the muon stations, the required number of stations depends on the track momentum, as shown in Table. 2.1.

Table 2.1: Requirement of the *isMuon* on the number of hits in the muon station depending on the momentum of the muon candidate.

Momentum of the muon track	Muon stations required
$3\,\text{GeV}/c < p < 6\,\text{GeV}/c$	M2 & M3
$6\,\text{GeV}/c < p < 10\,\text{GeV}/c$	M2 & M3 & M4/M5
$10\,\text{GeV}/c < p$	M2 & M3 & M4 & M5

The final-state muons of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates are required to satisfy the *isMuon* requirement and a combination of *ProbNN* cuts, as explained in Sec. 2.2. This selection reduces the background caused by the misidentification of hadrons as muons.

2.3 Trigger

Although the total bunch crossing rate of the LHC is 40 MHz, the readout rate of the detector is limited by the bandwidth of the front-end electronics to around 1 MHz [29]. Moreover, the available space to store the data is limited to a few kHz. The trigger system decreases the collision rate to 12.5 kHz [29] by deciding whether an event should be analyzed further and stored on disk or discarded. To perform this task, the trigger system consists of ensembles of selection algorithms that classify which events are interesting. Each of these ensembles is known as a trigger line.

The trigger system is divided into two different and consecutive levels that implement various trigger lines. A first selection, called *L0 trigger level*, is based on the hardware information from the calorimeters and the muon stations. This selection allows for the readout of the rest of the detector as it brings the collision rate down to 1 MHz. After that, a more

complex selection procedure is carried out combining the information from several parts of the detector using software in the *HLT level*. An event must be accepted by at least one line in each trigger level to be stored for further analysis.

2.3.1 L0 trigger

Two different units can be distinguished in the L0 trigger based on the information used in the trigger decision: the L0-calorimeter and the L0-muon trigger. In the former, the calorimeter information, including the SPD, PD, ECAL and HCAL (see Sec. 2.2.2), is used to estimate the transverse energy deposited by the interacting particles. Events with large occupancies take a long time to be processed in the reconstruction, consuming too much of the available processing time for the HLT trigger. Hence, only events with a *minimum transverse energy* and a *maximum number of SPD hits* will fire the trigger and be selected. A trigger line is defined for each particle type depending on the interaction of the particle candidate with the different components of the calorimeter system. These lines are `L0Hadron`, `L0Electron` and `L0Photon`.

In the L0-muon trigger, the decision is based on the transverse momentum of the particles propagating through the muon stations. Each station is divided into quadrants in the x-y plane such that straight-line tracks are searched for in each quadrant by combining hits in the muon stations with an origin consistent with the interaction region. The p_{T} of the muon candidate is estimated from the direction of the track, assuming the candidate received a single bending kick from the magnet. The trigger decision can be based on the highest p_{T} muon in all the quadrants for the `L0Muon` trigger or on the product of the two largest p_{T} values for the `L0DiMuon` trigger. In the first case, the event is accepted if the p_{T} of the muon is above a threshold of 1.48-1.76 GeV/c in Run 1 and 1.4-2.8 GeV/c in Run 2, depending on the data-taking year. As for the `L0DiMuon`, the product of the two largest p_{T} values must be larger than 1.68-2.56 GeV$^2/c^2$ in Run 1 and 1.7-2.3 GeV$^2/c^2$ in Run 2 [29, 39].

The $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates are mainly triggered by the `L0Muon` and `L0DiMuon` lines. In some cases, events triggering the L0-muon lines also trigger the L0-calorimeter lines, mainly the `L0Hadron`, from the presence of other hadron-type tracks in the underlying events.

2.3.2 HLT trigger

Only events selected by the L0 pass to the HLT trigger level, which performs a more complex selection implemented in software. For timing reasons, the HLT trigger consists of two consecutive stages known as *HLT1* and *HLT2*.

HLT1

Events passing the L0 level are transferred to the HLT1 level, where the charged tracks and the primary vertices are reconstructed. This procedure starts with the reconstruction of VELO tracks using a 3D pattern recognition algorithm that combines the hits in the VELO, forming a straight line pointing to the interaction region. The primary vertices are subsequently reconstructed in this stage from the intersection of a minimum of 5 VELO tracks. Upstream tracks are formed by extrapolating the VELO tracks to the TT stations and requiring at least 3 hits in the straight-line extrapolated region. The upstream tracks with $p_{\mathrm{T}} > 500$ MeV/c are then extrapolated further to the T1-T3 stations, where clusters of hits are searched in the small region corresponding to the extrapolation. The classification of the tracks in the LHCb detector can be found in Fig. 2.2.

In events firing the `L0Muon` or `L0DiMuon`, the already reconstructed tracks are further extrapolated to search for hits in the muon stations. The isMuon algorithm is run at this stage to find hits in the stations consistent with the reconstructed tracks based on the momentum of the track candidate, as explained in Sec. 2.2.4. Only tracks with $p > 3$ GeV/c are considered in this procedure since they are the only candidates propagating through the detector and reaching the muon stations.

Different selection criteria are applied to the candidates' p, p_{T}, track quality and displacement from the interaction point to select whether a candidate passes the HLT1 trigger. Moreover, a trigger line is defined for high p_{T} muon candidates without requiring a minimum displacement from the interaction point. Events accepted by the `L0DiMuon` are selected in two different ways: either from their dimuon mass, required to be above 2.5 GeV/c^2, without requiring a displacement relative to the interaction point (`Hlt1DiMuonHighMass`) or by their displacement without a constraint on their dimuon mass (`Hlt1DiMuonLowMass`) [39].

In the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, all triggered events are considered, no matter which trigger lines they fired or if they were triggered by the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates or the underlying event. The selected events in this analysis mainly trigger the following lines:

- Dimuon lines `Hlt1DiMuonLowMass` and `Hlt1DiMuonHighMass`, described above;
- `Hlt1SingleMuonNoIP`, which selects single muon tracks with large momenta ($p > 6\,\mathrm{GeV}/c$ and $p_{\mathrm{T}} > 1.3\,\mathrm{GeV}/c$) without any requirement on the displacement with respect to the primary vertex [40];
- `Hlt1TrackAllL0` selecting good quality tracks displaced from the primary vertex ($IP > 0.1$ mm) and with a $p_{\mathrm{T}} > 1.6$ GeV/c [40];
- `Hlt1TrackMuon`, with a similar selection as `Hlt1TrackAllL0`. In this case, the tracks are required to have matching hits in the muon chambers, and the p_{T} threshold

is lowered to 1 GeV [40];

- `Hlt1TrackMVA`, an improvement of the `Hlt1TrackAllL0` in Run 2. Multivariate analysis techniques are used to combine the information on the impact parameter χ^2 and the p_T of the tracks [41];
- `Hlt1TrackMuonMVA`, similar to the `Hlt1TrackMVA`, with the additional requirement that the track should have matching hits in the muon chambers [41].

HLT2

At the HLT2 stage, events accepted by HLT1 are fully reconstructed, including the reconstruction of both charged and neutral candidates. The track reconstruction of charged candidates is similar to that of HLT1, with a second step added to reconstruct low-momentum tracks, ignored in the HLT1 due to timing constraints. In addition to the muon identification used in HLT1, particle types are identified using the information from the calorimeter system, the RICH detectors and the muon stations. Electron candidates result from the association of a charged track with an ECAL cluster, while photon candidates are reconstructed starting from the cluster built in the L0-Calorimeter.

Mainly three types of HLT2 trigger lines contribute to selecting the interesting b-hadron decays further used in physics analysis. Generic beauty lines are also known as topological lines since they target events with decays consistent with a b-hadron decaying to at least two charged particles. Exclusive lines are built to select specific decays, such as $B^+ \to J/\psi K^+$ or $B^0_{(s)} \to h^+h'^-$ decays. Muon trigger lines select events containing one or two muons based on their p_T or dimuon invariant mass.

The HLT2 lines that mostly trigger the events in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis are:

- `Hlt2SingleMuon`, a muon line selecting candidates with a muon candidate with high transverse momentum displaced from the interaction point [40];
- `Hlt2DiMuonDetachedHeavy`, a dimuon line triggering detached dimuon pairs with large masses (above 2.95 GeV/c^2) and a significant displacement from the primary vertex [40];
- `Hlt2DiMuonDetached`, a dimuon line analogue to the previous line for "light" muon pairs (masses above 1 GeV/c^2) [40];
- `Hlt2DiMuonB`, an exclusive line designed to trigger b-hadron decays to two muons. It selects dimuon pairs forming a good quality vertex and with a dimuon mass above 4.7 GeV/c^2 [40];
- `Hlt2B2HH`, the exclusive line to trigger $B^0_{(s)} \to h^+h'^-$ candidates. It requires the hadron pair mass to be between 4.7-5.9 GeV/c^2, together with requirements on the

p_T of the hadrons and their impact parameters [40];

- `Hlt2Topo2BodyBBDT`, a topological line selecting a combination of two good quality tracks based on the output of a Boosted Decision Tree [40];
- `Hlt2TopoMu2BodyBBDT`, a topological line analogous to `Hlt2Topo2BodyBBDT` requiring the two tracks to be identified as muons [40].

Even though all triggered events are included in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, the trigger efficiency must be carefully determined and included in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ branching fraction measurement. The trigger efficiency evaluation is described in Chapter 4.

2.4 Simulation

Simulated events are exploited in this analysis to determine the detection and selection efficiencies, calibrate the signal and background distributions and estimate the expected yield of signal and background contributions. Proton collisions are generated using Pythia [42, 43] tunned with a specific LHCb configuration [44]. This configuration induces data and simulation differences in the kinematic distributions of the b-hadrons and the event occupancy, which are corrected in this analysis. Subsequent decays of unstable hadrons are described by the EvtGen package [45] with the fina- state radiation modelled by Photos [17]. The response of the subdetectors while particles propagate through the LHCb and the LHCb detector geometry are simulated using the Geant4 toolkit [46, 47] as explained in Ref. [48].

Simulation samples for the background components, the $B^0_{(s)} \to h^+h'^-$ and $B^+ \to J/\psi K^+$ normalization channels and the three signal channels, $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$, are generated and used in this analysis. Different samples are produced for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ channels to account for the different average decay time and production mechanisms of the two decays in the experiment. All the samples generated are used to estimate and calibrate the nuisance parameters and the background shapes in the dimuon invariant mass fit, from which the branching fractions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are determined.

3 | Analysis overview

The analysis presented in this thesis includes the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and the search for $B^0 \to \mu^+\mu^-$ decays as well as the initial-state radiation $B_s^0 \to \mu^+\mu^-\gamma$ decays. The analysis is performed using the data collected in pp collisions by the LHCb detector at CERN. A previous analysis was performed by LHCb in 2016 using the data gathered from 2011 to 2016, with a total integrated luminosity of 4.4 fb^{-1} [49]. The dataset of the current analysis comprises the entire Run 1 and Run 2 datasets, corresponding to a total integrated luminosity of 9 fb^{-1}. The significant increase of the dataset size results in a more precise measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and a tightening of the limits on the $B^0 \to \mu^+\mu^-$ branching fraction. Additionally, the first limit on the initial-state radiative $B_s^0 \to \mu^+\mu^-\gamma$ decay using LHCb data is obtained.

The general analysis strategy is the same for $B_{(s)}^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$, jointly referred to as $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$, and is roughly in line with the previous analysis. The signal signature consists of a pair of oppositely charged muons forming a good-quality vertex displaced from the interaction point. The invariant mass of the two muons, $m_{\mu^+\mu^-}$, is required to be close to the mass of the B^0 or B_s^0 mesons, with lower and upper values of 4900 MeV/c^2 and 6000 MeV/c^2. Despite the very clean experimental signal signature, $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ processes are very rare decays and only a handful of signal events are expected. In contrast, the background events polluting the signal region are expected to be $10^7 - 10^8$ times more frequent than the signal. Therefore, one of the major challenges of this analysis is the separation of signal and background.

The backgrounds in this analysis can be classified into two main categories: combinatorial and physical backgrounds. The combinatorial background originates from the combination of two unrelated muons, generated either from two distinct decays or from the pp collision itself, forming a good-quality vertex with a similar signature as that of the signal. This type of background, distributed throughout the entire dimuon invariant mass window, is the most general and abundant one.

Physical backgrounds are specific b-hadron decays that might be wrongly identified as signal decays. They can be further classified into misidentified $B_{(s)}^0 \to h^+h'^-$ decays, where the h and h' stand for kaons and pions; and semileptonic backgrounds, including

$B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ with $J/\psi \to \mu^+\mu^-$, and $H_b \to p\mu^-\overline{\nu}_\mu$ where H_b is a b-hadron. In the case of $B^0_{(s)} \to h^+h'^-$ decays, the pions and kaons in the final state might be misidentified as muons if they decay into muons while travelling through the detector. Although these decays-in-flight are the most common cause of misidentification, they are not the only one. Misidentification can also occur from the combination of random hits in the muon chambers with the tracks left by the hadrons in the tracking stations or from the "punch-through" of the hadrons through the hadronic calorimeter reaching the muon chambers. Regarding the semileptonic backgrounds, some of the particles in the final state might not be reconstructed, or the final-state hadrons may be misidentified as a muon, leading to a similar signature as $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays. These semileptonic backgrounds pollute mainly the lower signal region in the dimuon invariant mass spectrum where they "peak".

The amount of combinatorial and physical background is reduced and separated from the signal decays in several ways. In the first place, the data sample size is reduced by requiring the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates to pass a cut-based selection. This selection involves requirements on different variables associated with the final-state muons or the reconstructed $B^0_{(s)}$ candidates, as well as trigger and particle identification (PID) requirements. A detailed description of the selection applied to the signal candidates can be found in Chapter 4.

After the selection, the background and signal samples are further separated using a multivariate classifier known as a Boosted Decision Tree (BDT). This machine learning technique is presented in Sec. 4.5. The BDT classifies all the signal candidates as more or less signal-like by assigning them a score such that the higher the score, the more likely the candidate is signal. This technique is applied twice in this analysis. A first BDT is used in the selection procedure to reject background events by requiring a certain threshold of the BDT output to the selected $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates. The second BDT classifier is used differently after the selection. Instead of rejecting candidates based on their score, the dataset is split into subsamples of the second BDT output to avoid losing signal candidates and maximize the sensitivity to the branching fraction measurement.

The $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ yields are determined from unbinned extended maximum likelihood fits to the dimuon invariant mass distribution of the selected candidates. The fit is performed simultaneously in the distinct bins of the BDT output. The fit parameters are calibrated using simulation and calibration samples in data and fixed in the fit to maximize the sensitivity to the branching fraction measurement. These calibrated parameters are the expected fraction of signal and background events in each BDT subsample and the parameters describing the signal and background components in the fit.

Because the fit is performed in bins of the BDT, an accurate normalization of the number of signal events ending up in each BDT subsample is needed to constrain the dimuon invariant mass fit. This procedure, known as the BDT calibration, is described in Chapter 5. The expected fraction of events per BDT bin is evaluated from signal simulation samples

corrected to compensate for the differences between real and simulated events. Further corrections are applied to account for the relative effect per BDT bin due to the muonic trigger and PID requirements.

The signal components are modelled by double-sided Crystal Ball functions. As explained in Appendix A.1, this function is built from a Gaussian core connected to an exponential tail on each side. Six shape parameters describe the double-sided Crystal Ball function: the mean and resolution of the Gaussian core and two parameters for each tail associated with its slope and the connection points with the Gaussian core. The shape parameters are calibrated to guarantee the fit stability as discussed in Sec. 6.1. The means of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ Gaussian shapes are obtained from fits to the invariant mass of the $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays in data. For the width, the high statistics charmonium (J/ψ, $\psi(2S)$ with $\sqrt{s} \sim 3\,\text{GeV}/c$) and bottonium ($\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ with $\sqrt{s} \sim 9.5\,\text{GeV}/c$) resonances decaying into two muons are exploited. The mass resolution of these resonances is obtained from fits to their dimuon invariant mass spectra in data samples. The signal widths are estimated from an interpolation to the mean of the B^0 and B_s^0 mesons ($\sqrt{s} \sim 5\,\text{GeV}/c$). Finally, the exponential tail parameters are evaluated from signal simulation samples previously smeared with the resolution evaluated from the $q\overline{q}$ resonances.

Although not of intrinsic interest, the background components must be modelled in the dimuon invariant mass fit to determine the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ branching fractions properly. While the combinatorial background is modelled using an exponential function, the shapes used to describe the physical backgrounds are obtained from simulation samples. The yields of each physical background per BDT subsample are also used to constrain the fit. Their evaluation is explained in Sec. 6.2.

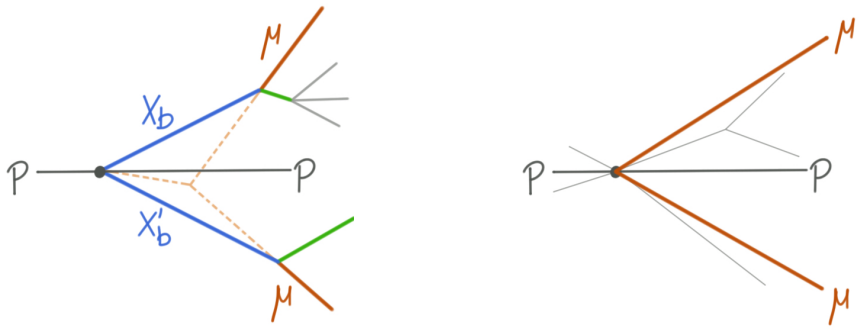
The $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ branching fractions are determined from the signal yields obtained in the fit, normalized to the branching fraction of other b-hadron decays. These decays are known as normalization channels and should have similar properties as the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays. This procedure, known as normalization, is explained in Sec. 6.3. Two complementary normalization channels are used: $B^0 \to K^+\pi^-$ decays and $B^+ \to J/\psi K^+$ decays with $J/\psi \to \mu^+\mu^-$. The former is a two-body decay with similar topology and kinematic behaviour as the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays. The main difference concerning the signal arises from the trigger and PID selection criteria. To verify the trigger and PID corrections, $B^+ \to J/\psi K^+$ signal events are used. Although these are three-body decays, their PID and trigger selection are similar to $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ due to the presence of the two muons from $J/\psi \to \mu^+\mu^-$ in the final state.

The mass calibration, the background estimation and the BDT calibration are included in the maximum-likelihood fit to constrain the fit input parameters. The dimuon invariant mass fit and the results obtained for the branching fractions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are presented in Sec. 7.

4 | Selection and efficiencies

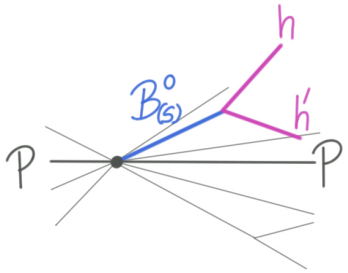
The $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ experimental signature consists of a pair of muons with opposite charges that form a good-quality vertex displaced from the pp interaction point. The trace left by the muons propagating in the detector is very distinctive since they are the only particles reaching the outermost part of the detector, where the muon chambers are located (see Chapter 2). The $B^0_{(s)}$ candidate is reconstructed from the trajectories and the four-momenta of the muon candidates.

Despite its distinctness, several backgrounds might have a signature similar to the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays. The trajectories of two oppositely charged muons generated directly in the pp collision or from two distinct (semileptonic) decays might be wrongly identified as coming from a $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidate. These two scenarios are jointly known as the combinatorial background and are sketched in Fig. 4.1. Alternatively, the final-state hadrons in $B^0_{(s)} \to h^+h'^-$ decays might be misidentified as muons coming from a $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidate if they decay-in-flight into muons while propagating through the detector. Finally, some semileptonic decays are also background components of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis when some of their final-state particles are not detected or reconstructed. The sketch of the physical backgrounds for the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays can be found in Fig. 4.2.

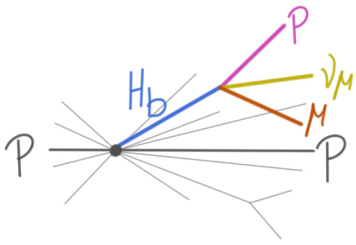


(a) Two decays (b) Prompt muons

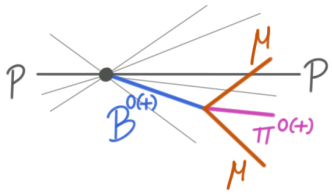
Figure 4.1: Illustration of the combinatorial background in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis.



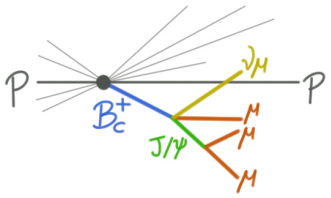
(a) Misidentified $B^0_{(s)} \to h^+h'^-$



(b) $H_b \to p\mu^-\overline{\nu}_\mu$



(c) $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$



(d) $B_c^+ \to J/\psi\,\mu^+\nu_\mu$

Figure 4.2: Illustration of the physical backgrounds in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis.

Because $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are very rare, they clearly occur much less frequently than the background events, requiring a stringent background rejection in their detection. The event candidates are required to pass several selection requirements to reduce the size of the dataset. An initial candidate selection based on a sequence of one-dimensional selection criteria is applied to reduce the amount of background in the data samples. This selection is presented in Sec. 4.1. The trigger strategies followed to select the signal and normalization candidates are described in Sec. 4.2. A specific selection in the particle identification (PID) variables, described in Sec. 4.3, is also included to reduce misidentified background events.

Part of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ events might be lost at the various stages of the selection procedure. The final signal yields obtained from the dimuon invariant mass fit should therefore be corrected by the efficiency of the criteria applied at each selection stage. The different efficiencies are presented in each corresponding section. The rest of the efficiencies resulting from the geometrical acceptance of the detector and the reconstruction are briefly discussed in Sec. 4.4

Following the one-dimensional conditions, the signal over background separation is further increased using Boosted Decision Trees (BDT). Two different multivariate classifiers are used in different stages of the analysis: the selection BDTS and the classification BDT. The first is used in the offline selection, where a cut is applied to the BDTS score to reduce the background significantly while keeping the majority of the signal. The classification BDT is

later used to classify the events as more or less signal-like and perform the dimuon invariant mass fit in subsamples of this BDT score. A detailed explanation of the BDT technique and the training variables used for the selection BDTS and classification BDT can be found in Sec. 4.5.

4.1 Candidate selection

An offline selection is applied to the signal candidates to reduce the amount of non-signal candidates and, hence, the sample size. It consists of an ensemble of one-dimensional cuts applied to different variables to remove non-physical or poorly reconstructed candidates, combinatorial and misidentified backgrounds from the data samples. The offline selection for the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ normalization channels is kept similar to the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates, with some differences in the $B^+ \to J/\psi K^+$ case since its decay topology is different from a two-body decay. The list of cuts applied in the offline selection for each channel is displayed in Table 4.1.

The non-physical candidates are eliminated by rejecting the reconstructed $B^0_{(s)}$ candidates with too large lifetimes as these $B^0_{(s)}$ candidates are highly unlikely in nature and probably come from wrongly reconstructed decays. The muon momenta are required to be below a certain threshold to reject tracks with low momenta resolution since their trajectories are not bent enough by the magnet. A particular case of non-physical candidates are ghost tracks resulting from random combinations of hits in the detector, not corresponding to original particle tracks. To reject them, a score is determined for each track reflecting the probability of it being a fake track. This score, known as the Ghost Probability (GP), is obtained from multivariate classifiers [50]. The muon tracks from the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates are requested to have low Ghost Probabilities.

Several cuts exclude poorly reconstructed candidates from the data sample. The reconstructed $B^0_{(s)}$ vertex is required to be a good quality vertex with a low χ^2 and large flight distance (FD) χ^2 to ensure that the B-decay vertex is well reconstructed and separated from the pp collision point. Since the two muon tracks should come from the same vertex, the candidates with a large distance of closest approach (DOCA) between the two daughter tracks are rejected. The χ^2 per degree of freedom (χ^2/DOF) of the track fit through the hits is required to be small to ensure the daughter tracks are well reconstructed. The transverse momenta of the daughter tracks are required to be within the LHCb acceptance.

After eliminating the non-physical and poorly reconstructed candidates, further cuts are included to reduce the combinatorial background. Random combinations of muon tracks originating from the pp collision are reduced by requiring a large impact parameter significance (IP χ^2) on the muon tracks relative to the primary vertex. The reconstructed $B^0_{(s)}$ candidates must have a minimum transverse momentum and a low IP significance relative

Table 4.1: Offline selection requirements for the signal and normalization channels. The values in parenthesis represent the thresholds for Run 2. A detail explanation of the selection variables and cuts can be found in the text.

Background reduced	Variable	On	$B^0_{(s)} \to \mu^+\mu^-$ $B^0_{(s)} \to h^+h'^-$	On	$B^+ \to J/\psi K^+$
Non-physical candidates	τ	$B^0_{(s)}$	$< 9 \times \tau_{B^0_s}$	B^+	$< 9 \times \tau_{B^0_s}$
	Track GP	μ /h	< 0.3 (0.4)	μ /K^+	< 0.3 (0.4)
	p	μ /h	< 500 GeV/c	μ /K^+	< 500 GeV/c
Poor reconstructed candidates	Vertex χ^2	$B^0_{(s)}$	< 9	J/ψ	< 9
	DOCA	$B^0_{(s)}$	< 0.3 mm	J/ψ	< 0.3 mm
	FD χ^2	$B^0_{(s)}$	> 225	J/ψ	> 225
	Track χ^2/DOF	μ /h	< 3 (4)	μ /K^+	< 3 (4)
	p_{T}	μ /h	[0.25, 40] GeV/c	μ /K^+	[0.25, 40] GeV/c
Combinatorial background	BDTS	$B^0_{(s)}$	> 0.05	B^+	> 0.05
	p_{T}	$B^0_{(s)}$	> 0.5 GeV/c	B^+	> 0.5 GeV/c
	Track IP χ^2	$B^0_{(s)}$	< 25	B^+	< 25
	Mass $m_{\mu^+\mu^-}$	$B^0_{(s)}$	[4.9, 6.0] GeV/c^2	J/ψ	[3.03, 3.16] GeV/c^2
	Track IP χ^2	μ /h	> 25	μ /K^+	> 25
Misidentified background	isMuon	μ	True	μ	True
$B^+_c \to J/\psi\,\mu^+\nu_\mu$ background	J/ψ veto	μ /h	> 30 MeV/c^2	-	-

to the primary vertex to guarantee they originated from the pp collision. The separation between signal and combinatorial background is increased using a multivariate classifier known as the BDTS (see Sec. 4.5). Only events with BDTS output above a certain threshold are kept to reduce the combinatorial background. This cut is optimized to keep most of the signal while reducing the background, mainly combinatorial, significantly.

To exclude misidentified background candidates, the muon tracks should pass the *isMuon* requirement (see Chapter 2). This background is further suppressed with the PID selection described in Sec. 4.3. A so-called J/ψ veto is applied to the muon candidates to

reduce $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ background events. This veto is built from the combination of each muon candidate coming from the $B^0_{(s)}$ candidate with each oppositely charged muon in the event. If the resulting dimuon invariant mass in one of these combinations is close to the J/ψ mass, the muon candidate is likely to come from a $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay, and the related $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidate is rejected.

Although the offline selection is optimized to keep the majority of the signal events, several real $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ candidates are rejected in this process. The efficiency associated with the offline selection is computed together with the reconstruction efficiencies using simulated events. Its evaluation and the associated values for the signal and normalization channels can be found in Sec. 4.4.2.

4.2 Trigger

The trigger system is used in real time to decide whether a candidate is stored for further analysis during data taking. Typically, the decay candidates are required to be accepted by a specific combination of trigger lines in order to reject background candidates. The trigger lines are defined at each trigger level as an ensemble of selection criteria based on reconstruction algorithms, as described in Sec. 2.3.

Two trigger strategies have been used in this analysis depending on the channel: an inclusive strategy is followed for $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ and $B^+ \to J/\psi\,K^+$ decays, and a combination of different line types in the L0&HLT1 and HLT2 levels is used for $B^0_{(s)} \to h^+h'^-$ decays. To avoid missing real $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ events, all triggered events are included in this analysis, regardless of the specific line they triggered. This inclusive strategy maximizes the trigger efficiency, an essential aspect in the search of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays. As shown in Sec. 4.2.1, the trigger efficiency associated with this strategy is around 95-97% for $B^0_{(s)} \to \mu^+\mu^-$ decays, depending on the decay and the data-taking period.

Ideally, the trigger selection should be the same for the signal, or $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays, and normalization channels, namely $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi\,K^+$ decays, to minimize systematic effects. This is not possible in the $B^0 \to K^+\pi^-$ case because the trigger performance depends strongly on the particle type of the final-state particles in the decay. To avoid a trigger selection based on the final-state hadrons, the selected $B^0 \to K^+\pi^-$ events are triggered by the underlying event, independently of the signal, at the L0&HLT1 levels. Although this trigger selection is considerably less efficient, a more unbiased procedure can be followed to determine the associated trigger efficiencies. At the HLT2 level, the $B^0 \to K^+\pi^-$ candidates are selected with topological HLT2 lines specifically built to select the inclusive $B^0_{(s)} \to h^+h'^-$ candidates. For $B^+ \to J/\psi\,K^+$, the same inclusive strategy as for $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ is used, retaining all events independently on the trigger lines fired.

The trigger efficiencies associated with the signal and normalization channels are computed on data samples using a technique known as the TISTOS method. This technique and the values of the efficiencies are detailed in the following section.

4.2.1 Efficiencies

Although in the experiment the first selection applied on data is the trigger, in simulation the trigger efficiency is determined for events already passing all the offline selection criteria. Therefore, the trigger efficiencies are defined from simulated samples as the fraction of selected events passing that trigger selection criterion:

$$\epsilon_{\text{Trg}} = \frac{N_{\text{Trg|Sel}}}{N_{\text{Sel}}} \tag{4.1}$$

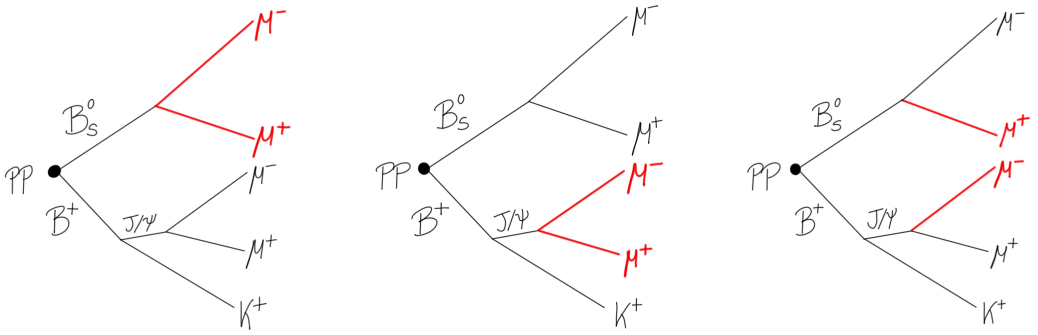
where |Sel indicates that the number of events is obtained from the already selected sample, including the acceptance, reconstruction, PID and offline selection effects. The expression in Eq. 4.1 is known as the *true trigger efficiency*.

The number of total selected events in Eq. 4.1, N_{Sel}, includes the events that were not triggered but passed all the previous stages of the selection process. These untriggered events are only present in simulation samples since only events firing the trigger are stored in real data. Although simulated events can be used to evaluate the trigger efficiencies, the mismodelling of some variables in these samples affects the reproduction of the trigger performance, possibly leading to biased efficiencies. Instead, the so-called TISTOS method is used as an alternative approach to evaluate the efficiencies from data samples.

TISTOS method

Once the event is triggered and stored, the particle tracks are reconstructed and the trigger response is analyzed relative to the signal candidate. The events are classified into three categories depending on the particle tracks triggering the event. For a given trigger line, the event is categorized as *triggered-on-signal (TOS)* when the particle tracks from the reconstructed signal candidate triggered that line. If something else in the event unrelated to the signal candidate triggers the line, the event is classified as *triggered-independent-of-signal (TIS)*. Finally, events triggered by a combination of particle tracks from the signal candidate and the rest of the event are considered *triggered-on-both (TOB)*. An example of the TIS, TOS and TOB categories is illustrated for the dimuon trigger line in Fig. 4.3, where the $B^0_{(s)} \to \mu^+\mu^-$ decay is treated as the signal. It is important to note that the TIS and TOS categories are not mutually exclusive since an event might simultaneously be TIS and TOS for the same trigger line. As shown in the Venn diagram from Fig. 4.4, these simultaneous TIS and TOS events are categorized as TISTOS events.

From its definition, the TIS decision is independent of the signal candidates. Therefore,



(a) Dimuon TOS (b) Dimuon TIS (c) Dimuon TOB

Figure 4.3: Example of the TOS, TIS and TOB categories for the dimuon trigger line with respect to the $B^0_{(s)} \to \mu^+\mu^-$ decay as the signal candidate. The trigger particles in each case are highlighted in red.

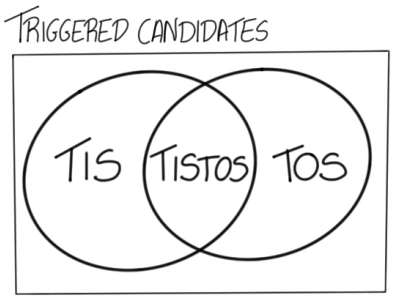


Figure 4.4: Venn diagram showing the overlap between the TIS and TOS categories for the triggered candidates for a particular trigger line. The TOB category is omitted since it is exclusive such that TOB events cannot also be classified as TIS or TOS.

the TIS selection is, to a good approximation, not affected by whether the signal candidates triggered the event or not. Taking this into account, the trigger efficiency expression in Eq. 4.1 can be rewritten as [51]:

$$\epsilon_{\text{Trg}} = \frac{N_{\text{Trg|Sel}}}{N_{\text{Sel}}} = \frac{N_{\text{Trg|Sel}}}{N_{\text{TIS|Sel}}} \times \frac{N_{\text{TIS|Sel}}}{N_{\text{Sel}}} = \frac{N_{\text{Trg}}}{N_{\text{TIS}}} \times \epsilon_{\text{TIS}} \tag{4.2}$$

where $N_{\text{TIS|Sel}}$ is the number of selected events passing the TIS criteria and $\epsilon_{\text{TIS|Sel}}$ is the efficiency of that TIS selection. As before, the |Sel subscript indicates that the efficiencies are evaluated on fully selected samples, including the acceptance, reconstruction, PID and offline selection. Since that is always the case, the |Sel subscript is already assumed in the last part of Eq. 4.2 and will henceforth be omitted.

The first factor in Eq. 4.2 can be evaluated from data samples since both terms are defined on triggered candidates. Unfortunately, that is not the case for the second term, correspond-

ing to the TIS efficiency, for which the total number of untriggered events is still required. Alternatively, the TIS efficiency can be determined within the TOS subsample using the TISTOS events as:

$$\epsilon_{\text{TIS|TOS}} = \frac{N_{\text{TISTOS}}}{N_{\text{TOS}}} \tag{4.3}$$

where N_{TISTOS} is the number of TISTOS events and N_{TOS} the number of TOS events.

Since the TIS selection is assumed to be independent of the signal candidates to a good extent, the TIS efficiency evaluated from the TOS subsample should be the same as the TIS efficiency obtained from the total sample, including the number of untriggered events; i.e. $\epsilon_{\text{TIS|TOS}} \approx \epsilon_{\text{TIS}}$. Taking this relation into account, the trigger efficiency in Eq. 4.2 can be evaluated from data samples as:

$$\epsilon_{\text{Trg}} = \frac{N_{\text{Trg}}}{N_{\text{TIS}}} \times \epsilon_{\text{TIS|TOS}} \tag{4.4}$$

The expression in Eq. 4.4 is known as the TISTOS method and is valid for any trigger selection where the TIS and TOS selections are uncorrelated. Three different trigger selection cases can be distinguished:

- the general case for which a mixed trigger strategy comprising a combination of TIS, TOS or TOB lines is used;
- a TOS-only trigger strategy, where only TOS lines are used to select the triggered events. In this case, N_{Trg} becomes N_{TOS} and Eq. 4.4 is simplified to $\epsilon_{\text{TOS}} = N_{\text{TISTOS}}/N_{\text{TIS}}$;
- a TIS-only trigger strategy, where the selection of events is only based on TIS lines triggering the underlying event. In such a case, N_{Trg} becomes N_{TIS} and Eq. 4.4 turns into Eq. 4.3.

The need to develop the TISTOS method is better illustrated by comparing the efficiencies obtained from data and simulation samples using $B^+ \to J/\psi K^+$ decays as the control channel. Besides being one of the main modes in terms of high statistics, the $B^+ \to J/\psi K^+$ channel is the optimal control mode for such a comparison since it fulfils the central assumption of the TISTOS method: the triggering of the muon TOS lines is significantly uncorrelated from the rest of the event, even from the muon TIS lines, allowing for the use of the TISTOS method to obtain unbiased trigger efficiencies. For completeness, the comparison between data and simulation is shown for the two distinct trigger strategies used in this analysis: the inclusive strategy characteristic of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays and the combination of L0&HLT1 TIS lines for $B^0 \to K^+\pi^-$ decays. The data samples are sWeighted to separate the background from the $B^+ \to J/\psi K^+$ candidates using the sPlot technique [52], allowing for the sWeighted data samples to be treated as signal-only. The efficiencies for the sWeighted data and simulation samples are evaluated using the TISTOS method and given in bins of variables that are correlated to the performance of the trigger lines used in each trigger strat-

egy. The true efficiencies (Eq. 4.1) evaluated from simulation samples are also included for a complete comparison.

The efficiencies for the inclusive trigger strategy of the $B^+ \to J/\psi K^+$ control channel are shown in Fig. 4.5. In this case, the efficiencies are given as a function of the maximum p_T and the product of the p_T of the two muon candidates separately. As can be seen, the agreement between the sWeighted data and simulation is excellent for the two binning variables considered, illustrating that the simulation accurately reproduces the trigger performance for the inclusive trigger strategy.

The same comparison for the combination of L0&HLT1 TIS lines gives different results. Fig. 4.6 displays the efficiencies for the L0&HLT1 TIS lines of $B^+ \to J/\psi K^+$ candidates as a function of the momentum and pseudorapidity of the B^+ candidate. The disagreement between sWeighted data and simulation is considerably large. The performance of the TIS lines depends on the underlying event and, for the simulation to be trusted, all the parameters describing it must be appropriately tuned. The discrepancy between the two distributions manifests that the tuning of these parameters is not reliable and evinces the need for the TISTOS method to evaluate the trigger efficiencies associated with a selection based mainly on TIS lines.

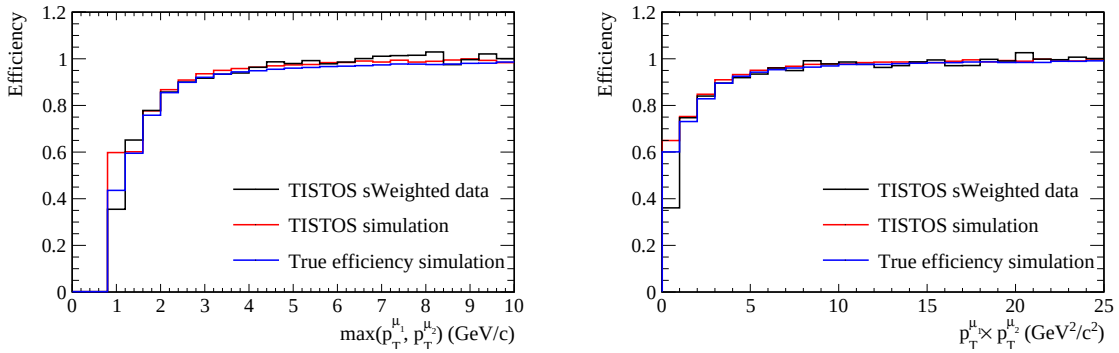


Figure 4.5: sWeighted data and simulation comparison between the trigger efficiencies of $B^+ \to J/\psi K^+$ events for the inclusive trigger strategy used for the signal and $B^+ \to J/\psi K^+$ channels. The efficiencies are obtained as a function of the maximum p_T of the muons (left) and the product of the p_T of the two muons in the final state (right). The black distribution corresponds to the efficiency from sWeighted data evaluated with the TISTOS method; the red and blue lines are the efficiencies for simulation computed with the TISTOS method and the true efficiency, respectively.

Maps application

The primary assumption in the development of the TISTOS method is that the TIS selection is independent of the signal candidates. Nevertheless, the correlation between the TIS decision and the signal candidates, although small, is not negligible. The b-hadrons in LHCb originate

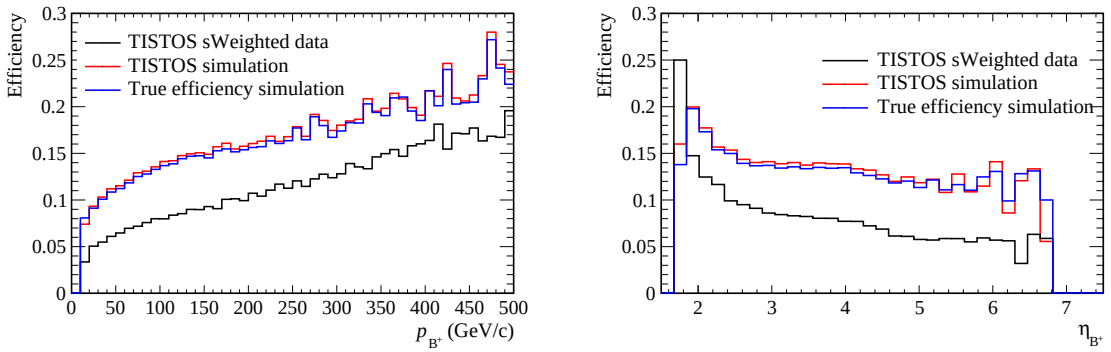


Figure 4.6: Comparison between sWeighted data and simulation samples of the L0&HLT1 trigger efficiencies of $B^+ \to J/\psi K^+$ events for the combination of TIS lines used to select $B^0_{(s)} \to h^+h'^-$ decays. The efficiencies are obtained as a function of the momentum (left) and pseudorapidity (right) of the B candidate. The black distribution corresponds to the efficiency from sWeighted data evaluated with the TISTOS method; the red and blue lines are the efficiencies for simulation computed with the TISTOS method and the true efficiency, respectively.

from $b\bar{b}$ pairs created in the pp collisions, mainly through gluon-gluon fusion. In some cases, the b-hadron inducing the TIS decision results from the same $b\bar{b}$ pair as the signal candidate, such that the kinematics of these two distinct b-hadrons are correlated because they generated from linked b quarks. Moreover, when dealing with hadronic trigger lines, the final-state hadron of the signal candidate might be detected in the same calorimeter cell as other hadrons from the underlying event, leading to a correlation between the underlying event and the signal candidates.

The modest correlation between the TIS and the signal candidates can be reduced if the TISTOS method is restricted to small regions of the phase space determined by some binning variables. These variables are chosen based on their correlation with the performance of the specific trigger selection and, thus, are dependent on the specific trigger strategy followed. Therefore, the trigger efficiencies evaluated from data samples using the TISTOS method are given in bins of these variables correlated to the trigger performance.

General procedure: $B^+ \to J/\psi K^+$ maps from data

In order to determine the trigger efficiencies from data samples, $B^+ \to J/\psi K^+$ decays are used as a proxy. As mentioned before, $B^+ \to J/\psi K^+$ decays are the optimal proxy mode due to the large statistics and the fact that the muon TOS is well isolated from the rest of the event. Thus, apart from being one of the normalization modes, the $B^+ \to J/\psi K^+$ decay is chosen as the proxy channel in the evaluation of the trigger efficiencies for all the signal and normalization modes with their respective trigger strategies.

The $B^+ \to J/\psi K^+$ signal and background components are unfolded in the $B^+ \to J/\psi K^+$ data samples by applying the sWeights, obtained from the sPlot technique [52]. The sWeights are determined from unbinned maximum likelihood fits to the mass distribution of the $B^+ \to J/\psi K^+$ data samples, where the J/ψ candidate mass is constrained to its measured value [1]. The invariant mass fits are performed for each data-taking year separately and are very similar among the different years in each corresponding run. As an example for Run 1 and Run 2, the $B^+ \to J/\psi K^+$ mass distributions with the corresponding fits are shown for the 2012 and 2016 data samples in Fig. 4.7. The signal shape is modelled with a Hypatia function (see Appendix A.2) whose parameters are constrained to the values obtained from simulation, except for the mean and the width, which are allowed to vary freely in the fit. An exponential function models the combinatorial background, while an analytical shape is used for the misidentified $B^+ \to J/\psi \pi^+$ background. This analytical shape consists of a Gaussian function obtained under the correct pion hypothesis modified to account for the mistaken kaon hypothesis by following an analytical description of the candidate kinematics [53].

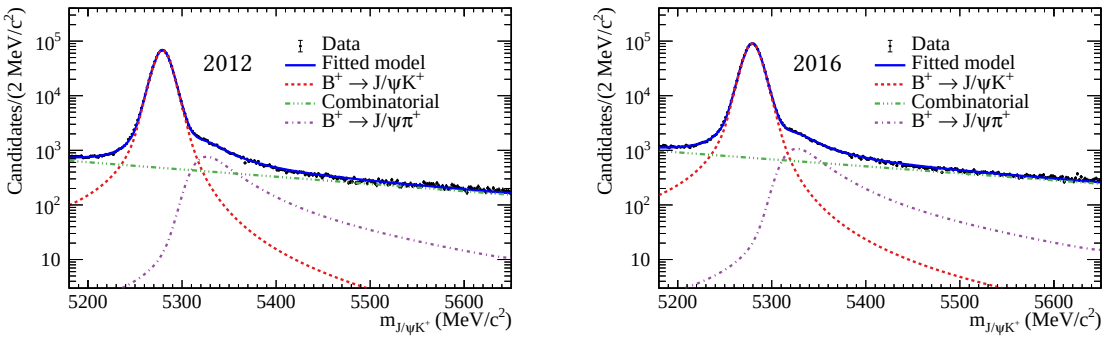


Figure 4.7: Mass distribution of the $B^+ \to J/\psi K^+$ candidates in the 2012 (left) and 2016 (right) data samples. The blue shape corresponds to the total fit model, in red the $B^+ \to J/\psi K^+$ component and in purple and green the $B^+ \to J/\psi \pi^+$ and combinatorial background components, respectively.

The trigger efficiencies are computed from $B^+ \to J/\psi K^+$ sWeighted data samples using the TISTOS method in bins of the specific binning variables. The resulting histograms from $B^+ \to J/\psi K^+$ data, referred to as *TISTOS maps*, are used as a look-up table to correct for the discrepancies observed in the trigger performance between data and simulation.

The distributions of the binning variables are different for the $B^+ \to J/\psi K^+$ control channel and the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ and $B^0_{(s)} \to h^+h'^-$ channels, for which the efficiencies must be calculated. This difference is important because the trigger performance is correlated with the binning variables, affecting the efficiency values. To take this effect into account, the trigger efficiencies for a specific decay mode result from the convolution between the corresponding $B^+ \to J/\psi K^+$ TISTOS map with the simulation samples of that decay mode. The specific details for each signal and normalization channel are given below.

Trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ and $B^+ \to J/\psi K^+$ decays

As previously mentioned, different binning variables are used depending on the trigger strategy. For the inclusive trigger strategy, used for the signal and $B^+ \to J/\psi K^+$ decays, the binning variables are the maximum transverse momentum of the two muons and the product of the transverse momenta of the two muons in the final state. These variables are chosen because they are used in the real-time selection of muon candidates in the hardware (L0) trigger level (see Sec. 2.3). The resulting TISTOS map for the inclusive trigger strategy is shown in Fig. 4.8. The trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ or $B^+ \to J/\psi K^+$ decays are evaluated from the convolution of this map with simulated $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ or $B^+ \to J/\psi K^+$ events. These simulation samples are further weighted to correct for the simulation and data differences observed in the momentum and pseudorapidity distributions of the B candidate, as described in Sec. 5.2.1. The trigger efficiencies for the signal and normalization modes are given by:

$$\epsilon^{B\to f}_{\text{Trg}} = \frac{\sum_{\text{iKin}} \epsilon^{\text{TISTOSMap}}_{\text{iKin}} \cdot N^{B\to f}_{\text{iKin}}}{N^{B\to f}} \tag{4.5}$$

where $\epsilon^{\text{TISTOSMap}}$ are the TISTOS maps for the inclusive trigger strategy obtained from $B^+ \to J/\psi K^+$ data and given in Fig. 4.8, i_{Kin} refers to the kinematic bin of the TISTOS map and $N^{B\to f}$ are the yields obtained from weighted simulation samples of the corresponding $B^+ \to J/\psi K^+$ or signal mode for which the trigger efficiencies are evaluated.

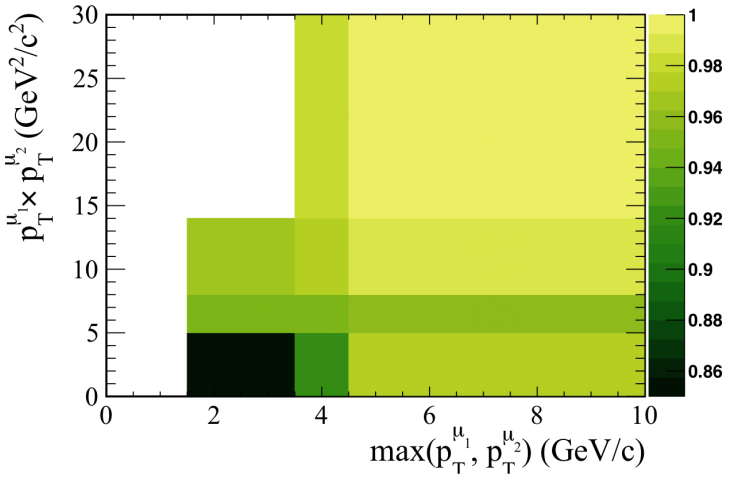


Figure 4.8: TISTOS map obtained from $B^+ \to J/\psi K^+$ sWeighted data with the trigger efficiencies for the inclusive trigger strategy as a function of the muon kinematics. The efficiencies correspond to the full trigger system, including L0, HLT1 and HLT2 levels. The sizes of the areas are chosen to have similar statistics available in each bin.

Trigger efficiencies for $B^0 \to K^+\pi^-$

A more complex procedure is followed for the $B^0 \to K^+\pi^-$ trigger efficiencies. As already mentioned, $B^0 \to K^+\pi^-$ candidates are selected from a combination of TIS lines at the L0&HLT1 levels, while specific TOS lines are used for the HLT2 level. Because the HLT2 lines used are topological lines optimized to select $B^0_{(s)} \to h^+h'^-$ candidates, $B^+ \to J/\psi K^+$ decays can only be used as a proxy to evaluate the L0&HLT1 TIS efficiencies from data samples. Moreover, the TISTOS method can be applied to evaluate the L0&HLT1 efficiencies since the muon TOS lines are uncorrelated from the lines triggered by the underlying event. That is not the case for the hadronic TOS lines, where there is a significant correlation with the underlying event from the overlap of several hadrons being detected in the same hadronic calorimeter cells. Therefore, the trigger efficiency for $B^0 \to K^+\pi^-$ is evaluated by determining the L0&HLT1 and HLT2 contributions separately as

$$\epsilon_{\text{Trg}} = \epsilon_{\text{L0\&HT1}} \times \epsilon_{\text{HLT2|L0\&HT1}} \tag{4.6}$$

The L0&HLT1 contribution is evaluated using a TISTOS map obtained from $B^+ \to J/\psi K^+$ sWeighted data. Because the TIS decision is correlated with the b-hadron kinematics, the TISTOS map for the L0&HLT1 TIS strategy is given in bins of the momentum and pseudorapidity of the B candidate. The resulting TISTOS map is illustrated in Fig. 4.9. The L0&HLT1 efficiency for the $B^0 \to K^+\pi^-$ decay results from the convolution of the TISTOS map with the $B^0 \to K^+\pi^-$ kinematic spectrum obtained from simulation samples as:

$$\epsilon^{B^0\to K^+\pi^-}_{\text{L0\&HLT1}} = \frac{\sum_{\text{i}_{\text{Kin}}} \epsilon^{\text{TISTOSMap}}_{\text{i}_{\text{Kin}}} \cdot N^{B^0\to K^+\pi^-}_{\text{i}_{\text{Kin}}}}{N^{B^0\to K^+\pi^-}} \tag{4.7}$$

where $\epsilon^{\text{TISTOSMap}}$ is the TISTOS map for the L0&HLT1 TIS strategy, i_{Kin} the kinematic bin of the TISTOS map and $N^{B^0\to K^+\pi^-}$ the $B^0_{(s)} \to h^+h'^-$ yields obtained from simulation samples.

Unlike the TIS effect from the L0&HLT1 levels, the HLT2 performance is expected to be appropriately modelled in simulation. Therefore, the HLT2 efficiency is evaluated from simulated $B^0 \to K^+\pi^-$ events as the fraction of events passing the HLT2 selection. Since the HLT2 selection is performed after the L0&HLT1 levels, the HLT2 efficiencies are evaluated on a sample of events already passing the TIS criteria.

Systematic checks

The stability of the trigger efficiency method is tested for several aspects through the systematics checks, which are performed separately for $B^0_{(s)} \to \mu^+\mu^-(\gamma)$, $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ decays. A summary of the systematic effects checked for the $B^0_s \to \mu^+\mu^-$ case can be found in Table 4.2, while their individual contributions for each channel are reported in Table 4.3.

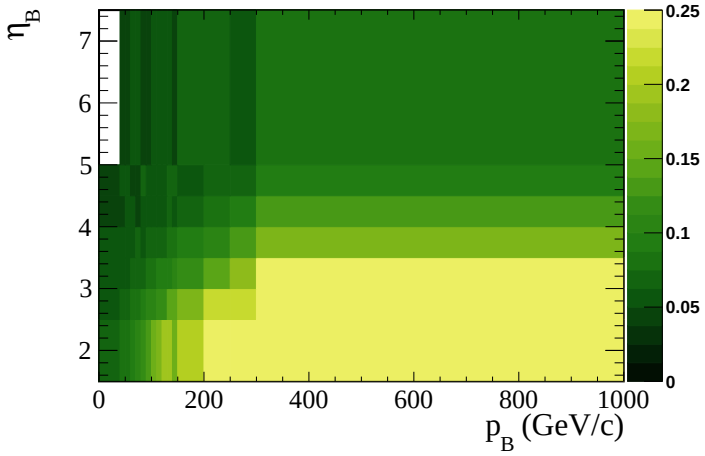


Figure 4.9: TISTOS map obtained from $B^+ \to J/\psi K^+$ data with the efficiencies for the combination of L0&HLT1 TIS lines.

One of the elements that could affect the trigger efficiencies is the binning scheme used in the TISTOS maps. To test this effect, the trigger efficiencies are estimated using a different binning scheme obtained by merging every two consecutive bins. The difference relative to the nominal values is taken as a systematic error. The choice of a binning scheme is found to be the primary source of systematic error of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ signal channels, as shown in Table 4.3.

The TISTOS maps are obtained from $B^+ \to J/\psi K^+$ sWeighted data to unfold the signal and background components. The influence of choosing a Hypatia function to model the $B^+ \to J/\psi K^+$ shape is analyzed using a double-sided Crystal Ball function instead and re-evaluating the efficiencies with the newly generated sWeights. The effect is negligible in all the cases and no systematic uncertainty is added.

Apart from the binning and the signal shape, three more systematics effects are tested in the evaluation of the trigger efficiencies: the TISTOS method, the maps application and the usage of the $B^+ \to J/\psi K^+$ channel as a proxy. Before presenting these checks, a new map type similar to the TISTOS map must be introduced: the *MC true map*. The main difference between the TISTOS and the MC true maps is the method used to evaluate the trigger efficiencies per kinematic bin. Instead of using the TISTOS method, the trigger efficiencies in each kinematic bin of the MC true map are evaluated as the true trigger efficiencies (Eq. 4.1). In order to obtain the absolute trigger efficiencies for a given decay, the MC true map is convolved with the kinematic spectrum from the simulation samples analogously to the TISTOS map. For clarity, the systematic checks are henceforth detailed for the $B^0_s \to \mu^+\mu^-$ trigger efficiencies, although the same procedure is followed for the rest of the channels.

The first systematic effect studied is the map usage to check whether it performs as expected and gives correct results. Two different values are compared in this check. For the

Table 4.2: Summary of the systematic effects checked for the procedure used to evaluate the trigger efficiencies for $B_s^0 \to \mu^+\mu^-$ decays. The same checks are applied to the $B_{(s)}^0 \to h^+h'^-$ and $B^+ \to J/\psi K^+$ channels. The results of procedure 1 and procedure 2 are compared to study the size of the systematic effects. When maps are used, the convolution with the $B_s^0 \to \mu^+\mu^-$ is assumed in the determination of the trigger efficiencies.

Systematic effect	Procedure 1	Procedure 2
Signal shape	Nominal TISTOS map from $B^+ \to J/\psi K^+$ data using DCB	TISTOS map from $B^+ \to J/\psi K^+$ data with Hypatia
Binning	Nominal TISTOS map from $B^+ \to J/\psi K^+$ data	TISTOS map from $B^+ \to J/\psi K^+$ data with merged bins
Use of maps	True trigger efficiencies from $B_s^0 \to \mu^+\mu^-$ simulation without maps	MC true map from $B_s^0 \to \mu^+\mu^-$ simulation
$B^+ \to J/\psi K^+$ as a proxy	MC true map from $B_s^0 \to \mu^+\mu^-$ simulation	MC true map from $B^+ \to J/\psi K^+$ simulation
TISTOS method	MC true map from $B^+ \to J/\psi K^+$ simulation	TISTOS map from $B^+ \to J/\psi K^+$ simulation

first one, the true trigger efficiencies are obtained from the $B_s^0 \to \mu^+\mu^-$ simulation samples without using the maps, directly from Eq. 4.1. For the second one, the efficiencies are evaluated from the convolution of the MC true map obtained from the $B_s^0 \to \mu^+\mu^-$ simulation samples with the very same sample. The discrepancy between the two sets of values is small and added as a systematic error.

A second check is the use of $B^+ \to J/\psi K^+$ decays as a proxy. This systematic check is performed for all the channels except for the $B^+ \to J/\psi K^+$ channel itself. As done to check the use of maps, the $B_s^0 \to \mu^+\mu^-$ trigger efficiencies resulting from two methods are compared. For the first one, the efficiencies result from the convolution of the MC true map obtained from the $B_s^0 \to \mu^+\mu^-$ simulation with the same $B_s^0 \to \mu^+\mu^-$ simulation sample. In the second method, the MC true map obtained from $B^+ \to J/\psi K^+$ simulation is convolved with the $B_s^0 \to \mu^+\mu^-$ simulation sample. The usage of $B^+ \to J/\psi K^+$ decays as a proxy is the primary

source of systematic uncertainty for the $B^0_{(s)} \to h^+h'^-$ trigger efficiencies, as expected from the different topologies of the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays and the correlation between the trigger performance with the B candidate kinematics.

A third and final check is the study of the possible bias added from the TISTOS method. Again, two sets of $B^0_s \to \mu^+\mu^-$ trigger efficiency values are compared. For the first set, the efficiencies are obtained from the convolution of the MC true map from $B^+ \to J/\psi K^+$ simulation with the $B^0_s \to \mu^+\mu^-$ simulation samples. For the second, the TISTOS map from $B^+ \to J/\psi K^+$ simulation is convolved with the $B^0_s \to \mu^+\mu^-$ simulation samples. The discrepancy between the two values is included as a systematic error.

Table 4.3: Individual listing of the uncertainties associated to the systematic effects for Run 1 (top) and Run 2 (bottom). These values are added in quadrature to evaluate the total systematic uncertainty.

Run 1					
	Signal shape	Binning	TISTOS method	Use of maps	Use of $B^+ \to J/\psi K^+$
$B^0_s \to \mu^+\mu^-$	0.002%	1.44%	0.25%	0.009%	0.73%
$B^0 \to \mu^+\mu^-$	0.002%	1.57%	0.24%	0.049%	0.72%
$B^0_s \to \mu^+\mu^-\gamma$	0.003%	1.86%	0.26%	0.054%	0.68%
$B^+ \to J/\psi K^+$	0.009%	0.49%	0.68%	0.010%	-
$B^0 \to K^+\pi^-$	0.001%	0.02%	0.09%	0.000%	0.121%

Run 2					
	Signal shape	Binning	TISTOS method	Use of maps	Use of $B^+ \to J/\psi K^+$
$B^0_s \to \mu^+\mu^-$	0.001%	0.77%	0.51%	0.009%	0.05%
$B^0 \to \mu^+\mu^-$	0.001%	0.81%	0.51%	0.050%	0.05%
$B^0_s \to \mu^+\mu^-\gamma$	0.001%	0.97%	0.54%	0.043%	0.06%
$B^+ \to J/\psi K^+$	0.003%	0.20%	0.86%	0.012%	-
$B^0 \to K^+\pi^-$	0.0006%	0.009%	0.13%	0.000%	0.14%

Trigger efficiency values

The trigger efficiencies for each signal and normalization channel are evaluated for each data-taking year separately. The systematic uncertainty assigned to each value results from the addition in quadrature of all the sources of systematic error. The Run 1 and Run 2 trigger efficiencies are obtained as the weighted luminosity averages of the data-taking years in each run. The resulting trigger efficiencies for each channel are listed in Table 4.4 for Run 1 and Run 2. The low efficiencies for the $B^0 \to K^+\pi^-$ decay result from the TIS strategy used in the L0&HLT1 levels, which is expected to be considerably less efficient than a TOS or inclusive

approach.

Table 4.4: Trigger efficiencies for the signal and normalization channels for the two data taking periods.

	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	$(95.79 \pm 0.33 \pm 1.64)\%$	$(97.12 \pm 0.14 \pm 0.93)\%$
$B^0 \to \mu^+\mu^-$	$(95.70 \pm 0.32 \pm 1.76)\%$	$(97.08 \pm 0.14 \pm 0.97)\%$
$B_s^0 \to \mu^+\mu^-\gamma$	$(95.38 \pm 0.32 \pm 1.95)\%$	$(96.94 \pm 0.13 \pm 1.11)\%$
$B^+ \to J/\psi K^+$	$(88.10 \pm 0.40 \pm 0.80)\%$	$(90.33 \pm 0.16 \pm 0.89)\%$
$B^0 \to K^+\pi^-$	$(4.33 \pm 0.02 \pm 0.16)\%$	$(7.27 \pm 0.02 \pm 0.20)\%$

4.3 Particle identification

The correct identification of particle types is crucial in rejecting physical backgrounds. Particle identification (PID) is also essential to isolate the $B^0 \to K^+\pi^-$ candidates used in the normalization procedure from inclusive $B^0_{(s)} \to h^+h'^-$ decays. In this analysis, three PID variables are used to identify the particle types from signal and normalization candidates: *isMuon*, *ProbNN* and ΔLL. A detailed description of the LHCb particle identification and these three variables can be found in Chapter 2.

The same muonic PID selection is applied to select $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ candidates. This PID selection is required on the two final-state muon candidates and consists of two steps. First, only muon candidates satisfying the *isMuon* algorithm are selected. Then, a tighter PID selection is required on these candidates based on a combination of *ProbNN* cuts. The performance of several *ProbNN* selections was studied on data samples by comparing the signal efficiency versus the background yields for each *ProbNN* criterion [54]. The *ProbNN* selection found to have the best performance is

$$\text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) > \begin{cases} 0.4 \text{ for Run 1 and 2015} \\ 0.8 \text{ for 2016, 2017, 2018} \end{cases} \tag{4.8}$$

where two distinct thresholds are applied over the years because the *ProbNN* algorithm was reoptimized in 2016, achieving a higher rejection power of non-muon candidates [54]. As explained in Sec. 2.2.4, the *ProbNN* variables are defined for each different particle type and give the probability of a reconstructed particle being of that type. The $(1-\text{ProbNN}_p)$ and the $(1-\text{ProbNN}_K)$ terms are included due to their strong rejection power of the $\Lambda_b^0 \to p\mu^-\overline{\nu}_\mu$ and $B^0_{(s)} \to h^+h'^-$ backgrounds. Since the *ProbNN* selection is optimized for the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays, the PID selection applied on the $B^+ \to J/\psi K^+$ channel requires only the *isMuon* criterion in the final-state muons.

A PID requirement is also needed to select the $B^0 \to K^+\pi^-$ decays used in the normalization procedure from inclusive $B^0_{(s)} \to h^+h'^-$ channels. The particle type of the final-state hadrons in $B^0_{(s)} \to h^+h'^-$ decays can be identified using the combined differential log-likelihood variable, $\Delta LL_{K-\pi}$. This variable is built from the likelihood of the PID subdetectors under the pion hypothesis. The selection is performed such that $\Delta LL_{K-\pi} > \kappa$ is required for the kaon and $\Delta LL_{K-\pi} < -\kappa$ for the pion candidates, where κ is a number that can be optimized for each particular decay. In the selection of $B^0 \to K^+\pi^-$ candidates, the default value $\kappa = 5$ is used.

Although the PID selection for the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ and the $B^0 \to K^+\pi^-$ channels is different, a similar procedure is used to evaluate their PID efficiencies. Their computation is performed on data samples from a control channel using the reconstructed candidates passing the offline selection and omitting the PID selection. The choice of the control channel depends on the PID selection applied and the particle type. For muon candidates, the $J/\psi \to \mu^+\mu^-$ decays with the J/ψ originating from b-hadron decays are exploited. The single muon PID efficiencies are evaluated using the tag-and-probe technique [55], which classifies the muons from the J/ψ decay as "tag" or "probe". While the tag muon is well-identified and used to ensure the presence of the J/ψ resonance, the probe muons have a loose selection. The single muon PID efficiency is evaluated as the fraction of probe muons passing the PID selection criteria.

The kinematics of the probe muon in the control channel are found to differ from that of the signal muon candidates. This difference is important because the *ProbNN* selection in Eq. 4.8 is strongly correlated to the muon kinematics and the track occupancy of the event. The correlation arises from the construction of the *ProbNN* selection as a combination of *ProbNN* variables associated with the pion, kaon and proton, which are mainly distinguished in the RICH detector, whose performance depends on the particle kinematics. To account for this correlation, the single muon PID efficiencies are evaluated as a function of three variables: the momentum and transverse momentum of the muon candidate and an occupancy variable, the *nTracks* variable. The resulting histograms are convolved with the binning variables distribution of the $B^0_{(s)} \to \mu^+\mu^-$ decays from simulation samples to obtain the PID efficiencies.

The *isMuon* algorithm is applied at the offline reconstruction level, meaning that only events satisfying this requirement are present in the data samples. Consequently, a correction to account for the *isMuon* efficiency must be applied to the muonic PID efficiencies. The *isMuon* efficiencies are calibrated using the tag-and-probe method from data and simulation $B^+ \to J/\psi K^+$ samples. The *isMuon* correction is evaluated as the ratio between the efficiencies from data and simulation and applied to the muonic PID efficiencies.

For the $B^0 \to K^+\pi^-$ mode, $D^{*\pm} \to (D^0 \to K^\pm\pi^\mp)\pi^\pm_s$ decays are used as a proxy to evaluate the PID efficiencies. In these decays, the slow pion from the $D^{*\pm}$ decays tags the flavour of

the D^0 candidate. The particle type of the decay products from the D^0 decay can then be distinguished from the charges of the tracks [55]. The efficiencies associated with a specific PID selection are evaluated for a reconstructed particle using this information.

As in the case of the muons, the final-state hadron kinematics of the $B^0 \to K^+\pi^-$ and $D^{*\pm} \to (D^0 \to K^\pm\pi^\mp)\pi_s^\pm$ decays are different. Since the hadron PID is mainly based on the information provided by the RICH, the single hadron PID efficiency for a specific PID selection is given in bins of the particle momentum, the pseudorapidity and the event occupancy; the variables correlated to the performance of the RICH. The single hadron efficiencies for a final-state hadron in the $B^0 \to K^+\pi^-$ decay are evaluated from the convolution of these histograms with the $B^0 \to K^+\pi^-$ kinematic spectra obtained from simulation samples. The PID efficiency of the $B^0 \to K^+\pi^-$ channel is the product of the single hadron efficiencies of the two hadrons in the final state.:

$$\epsilon_{\text{PID}}^{B^0\to K^+\pi^-} = \frac{\sum_i^{N_{\text{sample}}} \epsilon_i^\pi(p^-, \eta^-, \text{nTracks}) \cdot \epsilon_i^K(p^+, \eta^+, \text{nTracks})}{N_{\text{sample}}} \tag{4.9}$$

where N_{sample} is the total number of events in the simulation sample $\epsilon_i^\pi(p^-, \eta^-, \text{nTracks})$, $\epsilon_i^K(p^+, \eta^+, \text{nTracks})$ are the single pion and kaon PID efficiency given by the PID histograms determined from $D^{*\pm} \to (D^0 \to K^\pm\pi^\mp)\pi_s^\pm$ decays.

The PID efficiencies for the signal and $B^0 \to K^+\pi^-$ channels are listed in Table 4.5. The statistical errors include the statistics of the data and simulation samples. A variation of the choice of the binning scheme is studied and added as the systematic source of uncertainty in the PID efficiencies of the $B^0 \to K^+\pi^-$ channel. Apart from the binning, the effect of the correlation between the two muon tracks is also included as a systematic error in the signal channels. The *isMuon* correction is included in the PID efficiencies of the signal channels. For $B^+ \to J/\psi K^+$ candidates, only the *isMuon* correction is given as no *ProbNN* selection is applied.

Table 4.5: Particle identification (PID) efficiencies for the signal and normalization channels for the two data taking periods. The numbers reported for $B^+ \to J/\psi K^+$ correspond to the isMuon correction.

	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	$(85.80 \pm 0.06 \pm 0.53)\%$	$(88.22 \pm 0.03 \pm 0.39)\%$
$B^0 \to \mu^+\mu^-$	$(85.18 \pm 0.07 \pm 0.63)\%$	$(87.59 \pm 0.04 \pm 0.46)\%$
$B_s^0 \to \mu^+\mu^-\gamma$	$(84.87 \pm 0.06 \pm 0.88)\%$	$(87.85 \pm 0.03 \pm 0.64)\%$
$B^0 \to K^+\pi^-$	$(47.41 \pm 0.49 \pm 0.10)\%$	$(50.04 \pm 0.27 \pm 0.12)\%$
$B^+ \to J/\psi K^+$	1.0096 ± 0.0005	1.0026 ± 0.0002

4.4 Other efficiencies

4.4.1 Geometrical detector acceptance efficiency

Several decays are lost in the detection process because of the geometrical shape of the LHCb detector. This effect is quantified with the geometrical detector acceptance efficiency, evaluated as the fraction of decays whose final-state particles propagate inside the detector acceptance, defined by the polar region in the range [10, 400] mrad. The geometrical acceptance efficiencies are computed directly from simulation samples, and the values for the signal and normalization channels are given in Table 4.6. The slightly lower value for $B^+ \to J/\psi K^+$ decays compared to the $B^0_{(s)} \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ channels is due to the requirement of one more particle, namely the final-state kaon, to be in the detector acceptance.

Table 4.6: Geometrical acceptance efficiencies for the signal and normalization channels for the two data taking periods.

	Run 1	Run 2
$B^0_s \to \mu^+\mu^-$	$(18.58 \pm 0.020)\%$	$(19.35 \pm 0.026)\%$
$B^0 \to \mu^+\mu^-$	$(18.58 \pm 0.028)\%$	$(19.39 \pm 0.027)\%$
$B^+ \to J/\psi K^+$	$(16.58 \pm 0.034)\%$	$(17.35 \pm 0.026)\%$
$B^0 \to K^+\pi^-$	$(18.58 \pm 0.036)\%$	$(19.66 \pm 0.027)\%$
$B^0_s \to \mu^+\mu^-\gamma$	$(47.38 \pm 0.098)\%$	$(47.35 \pm 0.050)\%$

Since the final-state photon from $B^0_s \to \mu^+\mu^-\gamma$ candidates is not reconstructed, the geometrical detector acceptance efficiencies for the $B^0_s \to \mu^+\mu^-\gamma$ channel are generated using a different generator level requirement to avoid requiring the final state photon to be in acceptance. This results in large geometrical detector acceptance efficiencies associated with $B^0_s \to \mu^+\mu^-\gamma$ decays. These larger values are compensated by the reconstruction and selection efficiencies, which are significantly lower than for the rest of the channels.

4.4.2 Reconstruction and selection efficiencies

Particles are detected in the LHCb experiment from the hits they leave in the different subdetectors, which are combined to reconstruct particles and their trajectories. The reconstruction efficiency quantifies the fraction of candidates within the geometrical acceptance that are reconstructed successfully. These candidates are then required to pass the offline selection presented in Sec. 4.1. The selection efficiency gives the fraction of reconstructed candidates passing the offline selection.

The reconstruction and selection efficiencies are calculated jointly from simulation samples. A correction is applied to the resulting values to compensate for the simulation and

data differences observed in the track reconstruction efficiency. This correction is evaluated as the ratio between the data and simulation tracking efficiencies, evaluated on $J/\psi \to \mu^+\mu^-$ samples as a function of the kinematics of the final-state tracks. A track efficiency correction is evaluated for each channel and data-taking year, with values ranging from 0.9911 to 1.0210 for the data-over-simulation ratio.

Table 4.7: Reconstruction and selection efficiencies for the signal and normalization channels for the two data taking periods.

	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	$(32.43 \pm 0.15)\%$	$(33.07 \pm 0.21)\%$
$B^0 \to \mu^+\mu^-$	$(32.01 \pm 0.17)\%$	$(32.77 \pm 0.21)\%$
$B^+ \to J/\psi K^+$	$(17.54 \pm 0.18)\%$	$(17.59 \pm 0.18)\%$
$B^0 \to K^+\pi^-$	$(24.84 \pm 0.35)\%$	$(25.45 \pm 0.30)\%$
$B_s^0 \to \mu^+\mu^-\gamma$	$(10.72 \pm 0.07)\%$	$(11.52 \pm 0.08)\%$

The reconstructed and selection efficiencies for the signal and normalization channels are listed in Table 4.7 for the two data-taking periods. The values associated with $B^+ \to J/\psi K^+$ decays are considerably lower than for the rest of the channels considered. This difference is expected because the reconstruction efficiencies are intimately related to the single-track tracking efficiency, such that the more particles in the final state, the lower the reconstruction efficiency. Moreover, the three-body decay topology characteristic of $B^+ \to J/\psi K^+$ decays results in a lower total and transverse momentum of the final-state muons, leading to lower selection efficiencies.

4.5 Boosted Decision Trees

After the one-dimensional selection requirements, the signal and background components are further separated using Boosted Decision Trees (BDT) [56–58]. The BDT is a supervised machine learning technique that classifies the events depending on their signal-likeness. It assigns a score to each candidate such that the lower the score, the more likely the candidate is background.

As its name indicates, the BDT is a boosting algorithm. The boosting technique consists in building a powerful classifier from an ensemble of weak learners. For the BDT, the weak learners are decision trees, a supervised machine learning technique that classifies events as signal or background depending on a succession of one-dimensional cuts. A diagrammatic view of a decision tree is given in Fig. 4.10, showing the sequences of nodes. The goal at each node is to increase the signal over background separation by reducing the impurity of the sample, quantified with the Gini index [59]. This index offers a consistent way to measure

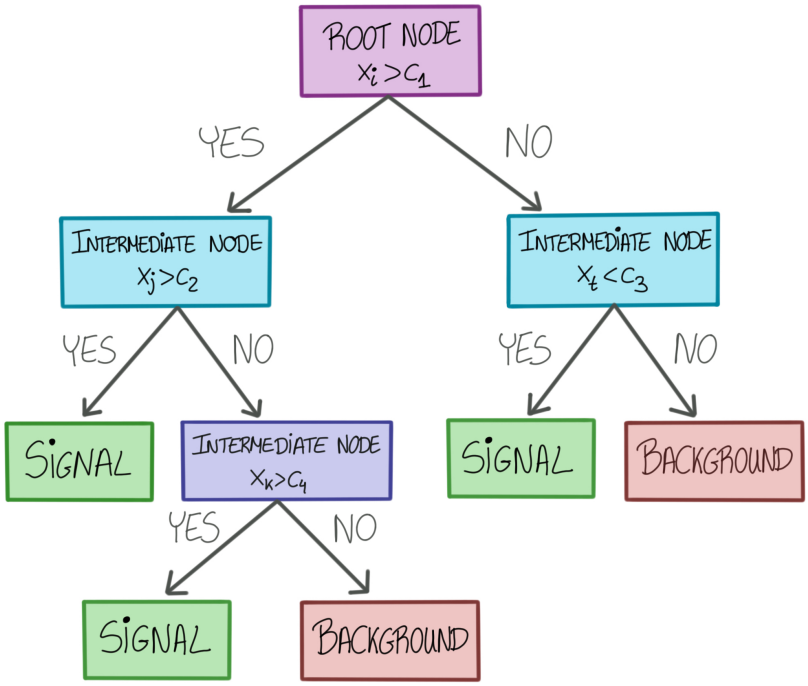


Figure 4.10: Diagrammatic illustration of a decision tree classifier. At each node, the data is separated into two subsets depending on a one-dimensional selection cut where x indicates the input variable to which the selection is applied and c_i the threshold applied.

how equitably a resource is distributed in a population. It is calculated from the area above and below the Lorenz curve, which represents the degree of equality in the distribution of such a resource graphically [59]. The values of the Gini index range from 0 to 1, with the largest values associated with the largest inequalities in the resource distribution and, hence, the highest impurities in the sample. Therefore, the variable (and its threshold) for the one-dimensional cut are chosen at each node to maximally reduce the Gini index value. The same procedure is performed sequence after sequence until a stop condition is satisfied.

The BDT results from the sequential combination of several decision trees. In each new sequence, the training sample is weighted to give more importance to the events wrongly classified in the previous step. The boosting technique is sketched in Fig. 4.11.

The BDT technique is used twice in the analysis. A first BDT is used at the offline selection, known as BDTS, and a one-dimensional cut is applied to its output to reduce the combinatorial background and the data size. After the selection procedure, a second BDT, referred to as the classification BDT, is used to further classify the selected events as more or less signal-like. The output of the classification BDT is used to maximize the sensitivity of the data sample by fitting the dimuon invariant mass spectrum in bins of this BDT classifier.

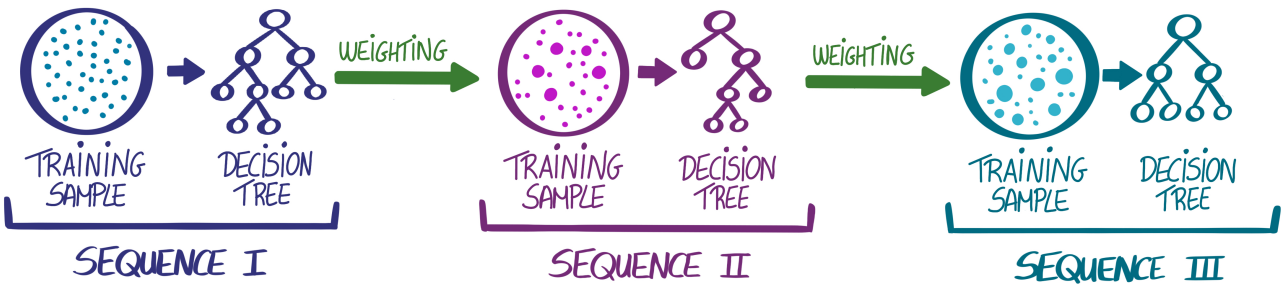


Figure 4.11: Schematic illustration of the boosting procedure with decision trees. In each new sequence, the data is weighted based on the results of the previous sequence.

4.5.1 Selection BDTS

The offline selection uses the BDTS output to reduce the sample size and reject the majority of the combinatorial background while keeping most of the signal candidates. The BDTS is trained on simulated samples using $B_s^0 \to \mu^+\mu^-$ decays as signal and inclusive $b\bar{b} \to \mu^+\mu^- X$ decays as backgrounds. The discriminating variables used in the training process are:

- the impact parameter of the B candidate relative to the primary vertex (PV);
- the impact parameter significance of the B candidate relative to the PV;
- the angle formed between the momentum of the B candidate and the vector connecting the primary and secondary vertices (DIRA);
- the distance of closest approach between the two muon candidates (DOCA);
- the minimum impact parameter of the two muon candidates with respect to the PV.

The BDTS performance strongly depends on the decay topology and the presence of a two-body decay vertex. Since it is used in the offline selection, its performance should be similar for all the signal and normalization channels. To satisfy this condition for the $B^+ \to J/\psi K^+$ mode, the impact parameter and its significance are evaluated on the J/ψ candidate and the DIRA on the J/ψ candidate relative to the PV.

To reduce the combinatorial background, a threshold on the BDTS output is required on the selected signal candidates. This cut was optimized in the previous analysis [49] to keep the most signal while rejecting most of the background. The most efficient threshold found is 0.05, which keeps 92% of the $B_s^0 \to \mu^+\mu^-$ candidates and rejects 70% of the combinatorial background. The same threshold is applied to the normalization channels to ensure maximal compatibility among samples.

4.5.2 Classification BDT

After the offline selection, another BDT classifies the selected events assigning each of them a probability of being signal. To improve the precision of the branching fraction measurement, the sensitivity of the data sample is maximized by classifying the events in subsamples of this new BDT classifier and performing the dimuon invariant mass fit in subsamples of the BDT output.

The BDT used for classification is trained on simulated samples using $B^0_{(s)} \to \mu^+\mu^-$ decays as signal and inclusive $b\bar{b} \to \mu^+\mu^- X$ decays as background. The offline selection is applied to the samples before the training with the exception of the particle identification in order to treat $B^0_s \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ equally. The variables found to optimize the BDT performance and chosen as the BDT discriminating variables are:

- the long track isolation;
- the VELO track isolation;
- the χ^2 of the dimuon vertex;
- the χ^2 of the impact parameter associated with the B candidate relative to the PV;
- the angle formed between the B candidate momentum and the vector connecting the primary and secondary vertices;
- the angular distance between the two muons, $\Delta_R(\mu^+, \mu^-) = \sqrt{\Delta\phi^2 + \Delta\eta^2}$;
- the minimum impact parameter χ^2 of the two muons relative to the PV.

The first two variables in the list quantify the isolation of the muon tracks in the event and deserve a special mention. The muon candidates originating from the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are expected to be well isolated, meaning no other particles are produced around the muon track, with the exception of the muon companion. The isolation variables are built using the output of different BDT classifiers and quantify how close to the muon track other particles are produced. The VELO track isolation is associated only with the VELO tracks in the event, while the long track isolation is related to the tracks reconstructed using the entire LHCb tracking system. For the definition of track types, see Sec. 2.1.

Once the BDT is trained, the BDT output is modified such that the signal yield is flat over the BDT range while the background peaks at zero. This modification is known as the flattening of the BDT and facilitates the separation of the data in BDT subsamples with a well-known efficiency.

The data samples are divided into BDT subsamples to increase the branching fraction measurement sensitivity. The BDT binning scheme used to create these subsamples is optimized to get the highest increase in sensitivity while keeping enough events in each BDT

bin to fit the background and signal components in the dimuon invariant mass fit. The performance of several BDT binning schemes was studied using fits to the dimuon invariant mass distributions from the sideband data samples. The optimal BDT binning scheme was found to consist of 6 bins with limits [0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 1.0]. The dimuon invariant mass fit is then performed simultaneously in these six BDT bins to determine the signal branching fractions, as described in Sec. 7. A finer binning is used in the calibration of the BDT response, presented in the next Chapter, to study in greater detail how various effects influence the BDT response. The BDT binning scheme used in the calibration of the BDT consisting of 8 bins with limits [0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0].

5 | BDT calibration

Due to the rareness of the $B^0_{(s)} \to \mu^+\mu^-$ processes, the signal selection efficiency must be maximized while keeping the background at a manageable level. This is done using the Boosted Decision Tree (BDT) technique presented in Sec. 4.5. The BDT classifies the signal candidates as more or less signal-like by assigning a score to each of them. The score ranks from 0 to 1 such that the higher (lower) the score of a candidate, the more likely it is signal (background).

The BDT score can be used to reduce the amount of background in different ways. A cut on the BDT output could be applied, removing candidates with low BDT scores following a similar approach as with the BDTS in the selection (see Chapter 4). With this approach, several real $B^0_{(s)} \to \mu^+\mu^-$ events with lower BDT values are lost, decreasing the sensitivity in the data samples. Instead, the data sample is divided into subsamples of the BDT classifier, also known as BDT bins, such that the dimuon invariant mass fit is performed simultaneously in these BDT bins. The same functions describing the different signal and background components are used for all the BDT bins, with their describing parameters shared across the BDT bins, except the yields of the signal and background components, which vary across the bins. An accurate prediction of the fraction of total candidates in each BDT bin for each signal and background component is needed to constrain the fit and ensure its stability and convergence. In the case of the signal channels, the estimation of these relative yields is known as the BDT calibration and is the main topic of this chapter. The evaluation of the expected yields for each background component is presented in Sec. 6.2.

The final goal of the BDT calibration is to estimate the expected fraction of signal events in each BDT bin. These values are needed to combine the fitted yields obtained for each BDT bin into a single branching fraction value for the $B^0_{(s)} \to \mu^+\mu^-$ decays. Consequently, the BDT calibration is one of the core parts of the analysis, with its results directly affecting the branching fraction measurement. Fig. 5.1 shows a flowchart summarizing the BDT calibration procedure. In order to minimize systematic effects, the expected signal yields per BDT bin must be determined from a sample whose characteristics are similar to the signal candidates. One possibility is to perform the BDT calibration on *real data* using a control channel with similar properties as the $B^0_{(s)} \to \mu^+\mu^-$ decay, such as the exclusive $B^0_{(s)} \to h^+h'^-$

decays, where $h, h' = \pi, K$. This procedure is explored in Sec. 5.1. The other possibility, discussed in Sec. 5.2, exploits the $B^0_{(s)} \to \mu^+\mu^-$ *simulated data* samples after correcting for the kinematic differences between simulation and real data. Since both approaches should be compatible, they should lead to the same results within uncertainties. Therefore, the comparison between the two different results, presented in Sec. 5.3, serves as a validation of the BDT calibration procedure.

Because the $B^0_{(s)} \to h^+h'^-$ events are rare decays, the $B^0_{(s)} \to \mu^+\mu^-$ simulation yields are higher than those obtained from the $B^0_{(s)} \to h^+h'^-$ real data samples. In addition, the selected $B^0_{(s)} \to h^+h'^-$ candidates are required to be triggered independently of the signal to reduce the difference in the trigger performance between hadrons and muons. The efficiencies associated with this trigger selection criterion are relatively low since the $B^0_{(s)} \to h^+h'^-$ event is triggered by other final-state objects apart from the $B^0_{(s)} \to h^+h'^-$ hadrons, resulting in a small expected number of $B^0_{(s)} \to h^+h'^-$ candidates. These two aspects directly affect the statistical uncertainties associated with the expected fraction of events per BDT bin obtained from $B^0_{(s)} \to h^+h'^-$ decays in data. Consequently, the $B^0_{(s)} \to \mu^+\mu^-$ simulation is the nominal sample used to perform the BDT calibration, whereas the $B^0_{(s)} \to h^+h'^-$ calibration serves as a validation test.

The $B^0_{(s)} \to \mu^+\mu^-$ candidates are selected in data by requiring them to pass a muonic PID and an inclusive trigger selection. Although the overall PID and trigger efficiencies are already included in the normalization procedure, their relative shifts per BDT bin should be accounted for to accurately predict the expected BDT fractions of signal events. To do this, muonic trigger and PID correction factors are determined in Sec. 5.4 and applied to the fractions obtained from $B^0_{(s)} \to \mu^+\mu^-$ simulation.

Another correction is yet needed for the calibration method using simulated $B^0_s \to \mu^+\mu^-$ events: the lifetime-dependence correction. The BDT response is correlated with the effective lifetime of the selected $B^0_s \to \mu^+\mu^-$ decays, which is the average decay time of the $B^0_s \to \mu^+\mu^-$ candidates in an experiment. As shown in Chapter 1, the effective lifetime is directly related to the mass-eigenstate rate asymmetry, $A^{\mu\mu}_{\Delta\Gamma_s}$, and to the decay width asymmetry, y_s. Since the effective lifetime of the $B^0_s \to \mu^+\mu^-$ has never been measured, its value in simulation is assumed to be Standard Model, meaning the decay width asymmetry is zero ($y_s = 0$). Although this assumption is in agreement with experimental observations for B^0, the value for B^0_s has been measured to be different from zero, such that the role of $A^{\mu\mu}_{\Delta\Gamma_s}$ becomes important (see Chapter 1). A correction for the BDT response is evaluated depending on different lifetime hypotheses. The determination of the lifetime-dependence correction factor is also discussed in Sec. 5.4.

As mentioned before, the optimal BDT binning scheme used in the dimuon invariant mass fit consists of 6 BDT bins with limits [0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 1]. In order to study the BDT response in more detail, especially at high BDT values, a finner binning scheme is used

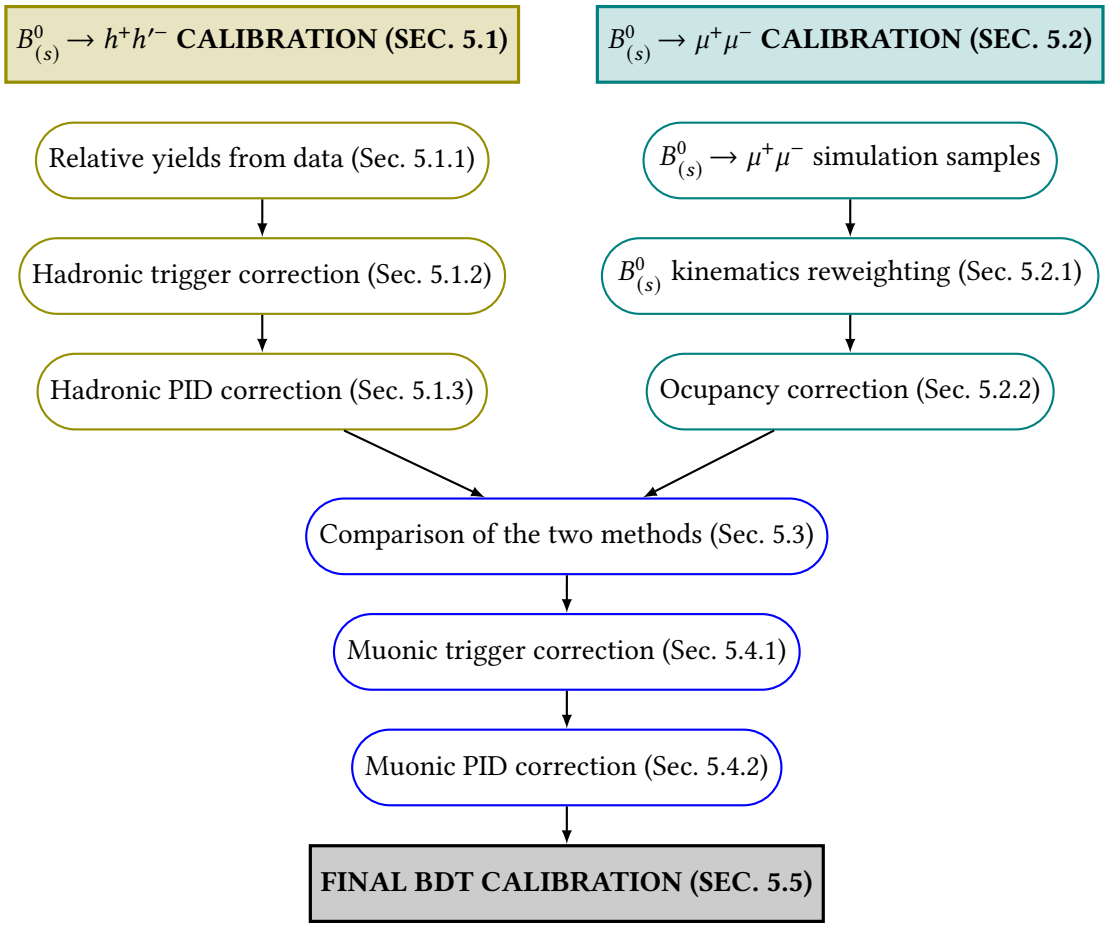


Figure 5.1: A flowchart of the BDT calibration procedure. The results can be derived from the $B^0_{(s)} \to h^+h'^-$ data sample corrected by the trigger and PID effects or the $B^0_{(s)} \to \mu^+\mu^-$ simulation samples, once corrected for the simulation and data differences. The results of both approaches are compared to test both methods. Two more corrections are applied to correct for the relative shifts per BDT bin of the muonic trigger and PID efficiencies.

for the BDT calibration which consists of 8 BDT bins with limits [0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]. After the studies, the expected relative signal yields are re-evaluated using the 6 BDT binning scheme characteristic of the dimuon invariant mass fit.

5.1 $B^0_{(s)} \to h^+h'^-$ calibration

The BDT calibration can be assessed from data samples using $B^0_{(s)} \to h^+h'^-$ decays as a proxy to $B^0_{(s)} \to \mu^+\mu^-$ decays. The selection procedure of the $B^0_{(s)} \to h^+h'^-$ candidates is kept similar to that of the $B^0_{(s)} \to \mu^+\mu^-$ candidates, with the only differences being the trigger and the

PID requirements. These two requirements depend on the final-state particles and how they are detected: while the tracking is similar for hadrons and muons, particle identification is obtained in different parts of the LHCb detector (see Chapter 2). Then, the fraction of events estimated in a BDT bin, $f_{\mathrm{i_{BDT}}}$, are determined by correcting the fraction of $B^0_{(s)} \to h^+h'^-$ events per BDT bin in *real data* for the hadronic trigger and PID effects as:

$$f_{\mathrm{i_{BDT}}} = \frac{N^{hh'}_{\mathrm{i_{BDT}}}/N^{hh'}}{\epsilon^{hh'}_{\mathrm{i_{BDT},trg}}/\epsilon^{hh'}_{\mathrm{trg}} \cdot \epsilon^{hh'}_{\mathrm{i_{BDT},PID}}/\epsilon^{hh'}_{\mathrm{PID}}} = \frac{f^{hh'}_{\mathrm{i_{BDT}}}}{\omega^{hh'}_{\mathrm{i_{BDT},trg}} \cdot \omega^{hh'}_{\mathrm{i_{BDT},PID}}} \tag{5.1}$$

where $f^{hh'}_{\mathrm{i_{BDT}}}$ is the fraction of $B^0_{(s)} \to h^+h'^-$ events for the $\mathrm{i_{BDT}}$ BDT bin, i.e. $f^{hh'}_{\mathrm{i_{BDT}}} = \frac{N^{hh'}_{\mathrm{i_{BDT}}}}{N^{hh'}}$, and $\omega^{hh'}_{\mathrm{i_{BDT},trg}}$ and $\omega^{hh'}_{\mathrm{i_{BDT},PID}}$ are the trigger and PID corrections, respectively, i.e. $\omega^{hh'}_{\mathrm{i_{BDT},trg(PID)}} = \frac{\epsilon^{hh'}_{\mathrm{i_{BDT},trg(PID)}}}{\epsilon^{hh'}_{\mathrm{trg(PID)}}}$. The hh' superscript indicates which exclusive $B^0_{(s)} \to h^+h'^-$ channel has been chosen to determine the relative yields. Only the two most dominant exclusive $B^0_{(s)} \to h^+h'^-$ decays, $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$, are used. To avoid any discrepancies arising from the different effective lifetimes and their correlation with the BDT, $B^0_s \to K^+K^-$ decays serve as calibration for the $B^0_s \to \mu^+\mu^-$ mode and the $B^0 \to K^+\pi^-$ channel for $B^0 \to \mu^+\mu^-$ decays.

Each component in Eq. 5.1 is evaluated independently. The $f^{hh'}_{\mathrm{i_{BDT}}}$ fractions are determined from $B^0_{(s)} \to h^+h'^-$ data samples as described in Sec. 5.1.1. The trigger correction, $\omega^{hh'}_{\mathrm{i_{BDT},trg}}$, is defined as the ratio between the trigger efficiency for each BDT bin and the overall trigger efficiency. Its computation is done from data-based techniques and described in Sec. 5.1.2. The PID correction, $\omega^{hh'}_{\mathrm{i_{BDT},PID}}$, is defined in a similar way and its calculation is detailed in Sec. 5.1.3.

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

One of the inputs needed for the BDT calibration obtained from data samples is the $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ relative event yields. They are determined per BDT bin as the ratio between the yields in each BDT bin and the total yield over the whole BDT range. The yields are obtained from binned maximum likelihood fits to the invariant mass of the hadrons from selected $B^0_{(s)} \to h^+h'^-$ data samples. The fits are performed separately on the total sample and the BDT bins. The selection criteria, given in Sec. 4.1, are kept similar to that of the $B^0_{(s)} \to \mu^+\mu^-$ candidates to ensure compatibility between the samples. The yields are determined for $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ decays by requiring different PID requirements on the final-state particles and performing the invariant mass fits for each channel separately.

Fig. 5.2 and 5.3 display the invariant mass distributions for the total sample and the BDT bins of the $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ data samples, with the corresponding fit model

overlaid. For both the $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ fits, the signal peak is modelled using a double-sided Crystal Ball function (see Appendix A.1) while an exponential function is used for the combinatorial background.

Apart from the combinatorial background, several exclusive background components have to be adequately modelled. In the case of the $B^0 \to K^+\pi^-$ fit, a structure can be found to the right of the signal peak, partly overlapping with the signal. It corresponds to the Cabibbo-suppressed $B^0_s \to \pi^+K^-$ decays and it is simulated using a double-sided Crystal Ball function. This component is removed in the $B^0_s \to K^+K^-$ data sample by the PID selection since it requires the presence of two kaons in the final state.

Two more background components are visible in both the $B^0_s \to K^+K^-$ and $B^0 \to K^+\pi^-$ fits. Although their overlap with the $B^0_s \to K^+K^-$ and $B^0 \to K^+\pi^-$ shapes is negligible, they might bias the combinatorial shape, affecting the signal peak. The small structure located at low mass, mostly visible in the $B^0 \to K^+\pi^-$ case, is caused by partly reconstructed three-body b-hadron decays and modelled by an ARGUS function [60]. A structure corresponding to the $\Lambda^0_b \to ph$ decays is observed at higher masses, particularly in the $B^0_s \to K^+K^-$ fits. A double-sided Crystal Ball function is used to model this background. All the parameters associated with the background components are fixed to the values obtained from simulation samples, except the total yields, which are left free in the fit.

Several consistency checks are performed to test the fit stability. In the nominal fit, the double-sided Crystal Ball tail parameters of the signal shapes are implemented as a Gaussian constraint, meaning they are treated as floating parameters whose values are allowed to vary following a Gaussian inside a specific range. The ranges of these values are determined from fits to the simulation samples. To study how much this biases the fit results, the fit is repeated by fixing the tail parameters to the central values obtained from simulation fits. Similarly, the effect of the invariant mass binning scheme is also tested by repeating the fit changing the binning scheme. In both cases, the differences concerning the nominal values are negligible and no systematic uncertainty is added.

The presence of the partly reconstructed background affects the combinatorial shape, which in turn might affect the signal peak. To study this effect, the partly reconstructed background component is removed by increasing the lower edge of the mass window to 5150 MeV/c^2. The difference with respect to the nominal fit is taken as a systematic uncertainty.

The fractions of events per BDT bin for $B^0 \to K^+\pi^-$ data are displayed for Run 1 and Run 2 in the second columns of Table 5.1 and Table 5.3, respectively. Tables 5.2 (Run 1) and 5.4 (Run 2) show the corresponding fractions for $B^0_s \to K^+K^-$ data. In both cases, the first uncertainty is statistical and the second systematic. The correlation between the BDT-binned and total yields is included in the calculation of the statistical and systematic errors associated with the relative yields. The statistical errors are relatively large, especially for the low BDT values. This is expected as the $B^0_{(s)} \to h^+h'^-$ decays are rare and selected using

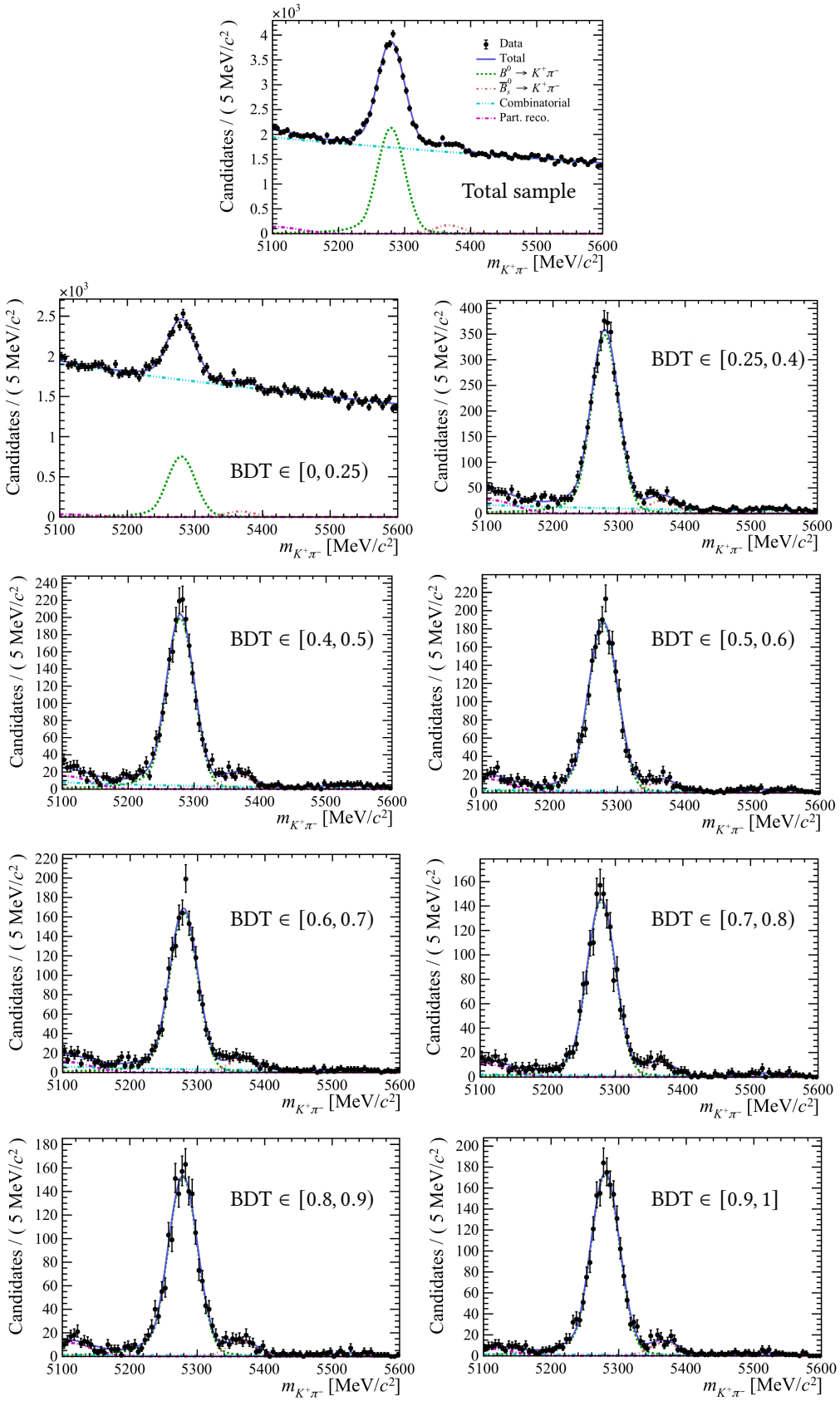


Figure 5.2: Invariant mass fits for $B^0 \to K^+\pi^-$ for the total sample and the BDT subsamples.

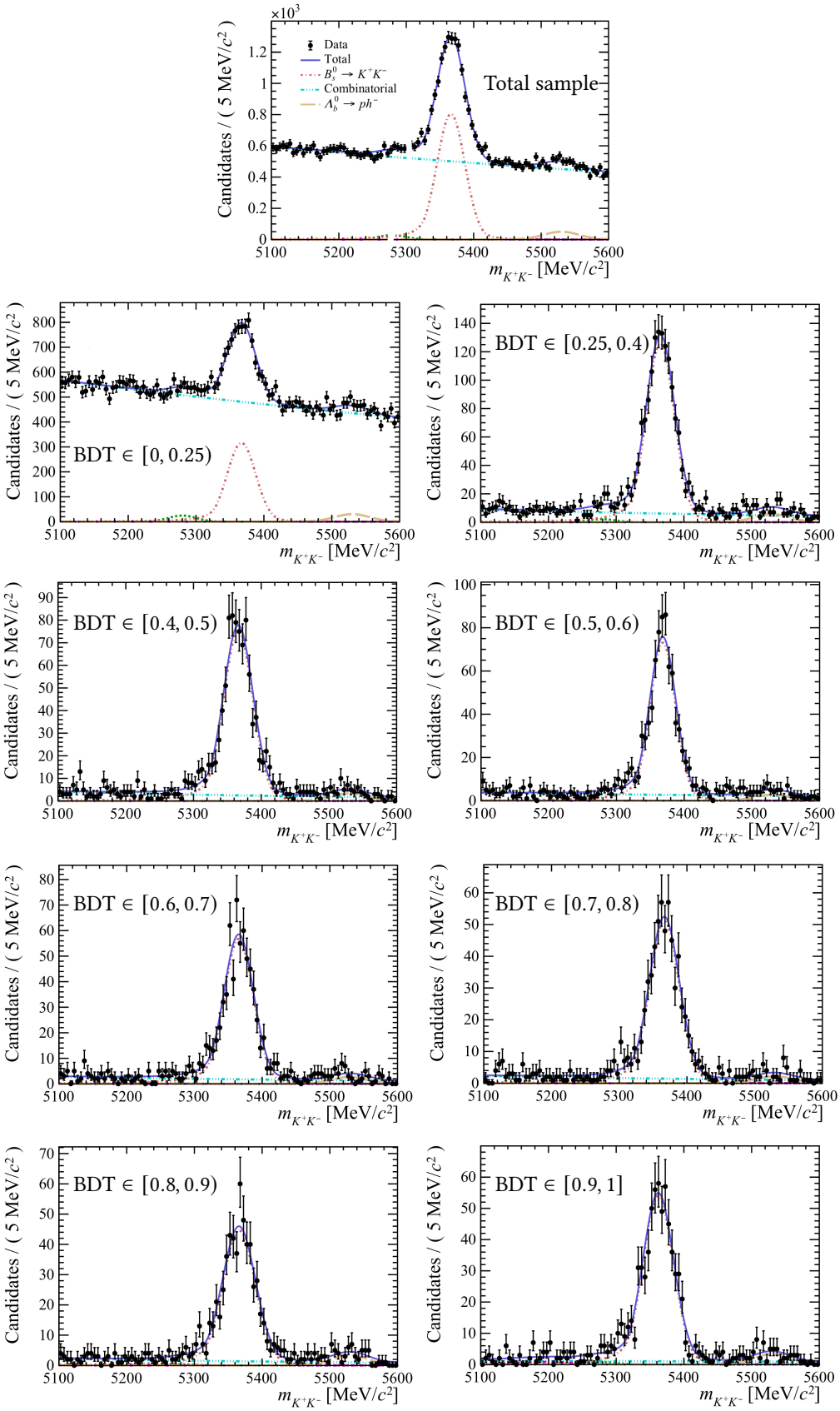


Figure 5.3: Invariant mass fits for $B^0_s \to K^+K^-$ for the total sample and the BDT subsamples.

TIS trigger lines, further limiting the available statistics, as explained in Sec. 4.2

5.1.2 Hadronic trigger correction

The trigger performance differs between muons and hadrons, and thus two trigger strategies are used to select the $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ candidates (see Sec. 4.2). As a consequence, the bin yields for $B^0_{(s)} \to h^+h'^-$ decays obtained from data will be different than for the $B^0_{(s)} \to \mu^+\mu^-$ channel, and a trigger correction factor per BDT bin has to be included to correct for this difference. For each BDT bin, the trigger correction is evaluated as the ratio between the BDT-binned and overall trigger efficiencies, i.e. $\omega^{hh'}_{\mathrm{i_{BDT}},\mathrm{trg}} = \frac{\epsilon^{hh'}_{\mathrm{i_{BDT}},\mathrm{trg}}}{\epsilon^{hh'}_{\mathrm{trg}}}$.

General strategy

The real-time selection applied to the $B^0_{(s)} \to h^+h'^-$ candidates consists of a combination of different types of trigger lines[1]. At the L0 and HLT1 levels, the $B^0_{(s)} \to h^+h'^-$ candidates are *triggered independently of the signal (TIS)*, or triggered by other final-state objects apart from the signal hadrons of the $B^0_{(s)} \to h^+h'^-$ candidate. This trigger strategy is chosen because it ignores the presence of the final-state hadrons in the $B^0_{(s)} \to h^+h'^-$, reducing the difference with respect to the triggering of $B^0_{(s)} \to \mu^+\mu^-$ candidates, although resulting in low event yields. At the HLT2 level, the selected events pass lines dedicated to the hadronic final state and are *triggered on the signal (TOS)* by the final-state hadrons from the $B^0_{(s)} \to h^+h'^-$ candidate. A complete explanation of the line categories and the trigger strategy used to select $B^0_{(s)} \to h^+h'^-$ candidates are given in Sec. 4.2.

The BDT-binned and overall trigger efficiencies are needed to evaluate the hadronic trigger correction factors. As shown in Sec. 4.2, the performance of the L0&HLT1 selection based on TIS lines is not accurately reproduced in simulation, affecting the evaluation of the L0&HLT1 trigger efficiencies. That is not the case for the HLT2 trigger performance, expected to be correctly modelled in simulation. Therefore, the trigger correction is determined by evaluating the L0&HLT1 and HLT2 contributions separately, following a similar procedure as for the absolute trigger efficiencies in Sec. 4.2. The relative L0&HLT1 efficiencies are determined from the TISTOS method by exploiting real data samples using $B^+ \to J/\psi K^+$ decays as a proxy. For the relative HLT2 efficiencies, simulation samples are used where the L0&HLT1 selection is already applied since the L0&HLT1 level precedes the HLT2 level in the real-time trigger system. The final trigger correction per BDT bin results from the combination of the L0&HLT1 and HLT2 contributions as

$$\omega^{hh'}_{\mathrm{i_{BDT}},\mathrm{trg}} = \omega^{hh'}_{\mathrm{i_{BDT}},\mathrm{L0\&HLT1}} \times \omega^{hh'}_{\mathrm{i_{BDT}},\mathrm{HLT2|L0\&HLT1}} = \frac{\epsilon^{hh'}_{\mathrm{i_{BDT}},\mathrm{L0\&HLT1}}}{\epsilon^{hh'}_{\mathrm{L0\&HLT1}}} \times \frac{\epsilon^{hh'}_{\mathrm{i_{BDT}},\mathrm{HLT2|L0\&HLT1}}}{\epsilon^{hh'}_{\mathrm{HLT2|L0\&HLT1}}} \tag{5.2}$$

[1] Collection of selection cuts based on the reconstruction algorithms at each trigger level.

where $\mathrm{i}_{\mathrm{BDT}}$ indicates the specific BDT bin and $\omega^{hh'}_{\mathrm{i}_{\mathrm{BDT}},\mathrm{L0\&HLT1}}$ and $\omega^{hh'}_{\mathrm{i}_{\mathrm{BDT}},\mathrm{HLT2|L0\&HLT1}}$ are the relative L0&HLT1 and HLT2 efficiencies, respectively.

Relative L0&HLT1 efficiencies

The relative L0&HLT1 efficiencies per BDT bin are evaluated as the ratio between the BDT-binned and overall L0&HLT1 efficiencies. Since the overall and BDT-binned efficiencies cannot be calculated directly from simulation, they are obtained from data samples using the TISTOS method. The L0&HLT1 efficiencies cannot be assessed from $B^0_{(s)} \to h^+h'^-$ data samples because the correlation between the signal final-state hadrons with the underlying event and the TIS lines is not negligible. In addition, the $B^0_{(s)} \to h^+h'^-$ decays are rare, and the low statistics in the data sample will affect the uncertainties associated with the L0&HLT1 efficiencies. Instead, the $B^+ \to J/\psi K^+$ mode is used as a proxy channel of the $B^0_{(s)} \to h^+h'^-$ decays.

As explained in Sec. 4.2, the overall L0&HLT1 efficiencies are calculated from the $B^+ \to J/\psi K^+$ data samples as a function of two kinematic variables: the momentum and the pseudorapidity of the B^+ candidate. The resulting 2-dimensional L0&HLT1 efficiency table, referred to as the *TISTOS map*, is shown in Sec. 4.2 where a detailed description of the procedure followed to obtain it is given. This map is used as a look-up table to obtain the overall L0&HLT1 efficiency for exclusive $B^0_{(s)} \to h^+h'^-$ decays by convolving it with the kinematic spectrum of the $B^0_{(s)} \to h^+h'^-$ decays obtained from simulation. The overall L0&HLT1 efficiency for $B^0_{(s)} \to h^+h'^-$ decays is

$$\epsilon^{hh'}_{\mathrm{L0\&HLT1}} = \frac{\sum_{\mathrm{i}_{\mathrm{Kin}}} \epsilon^{\mathrm{TISTOSMap}}_{\mathrm{i}_{\mathrm{Kin}}} \cdot N^{B^0_{(s)}\to h^+h'^-}_{\mathrm{i}_{\mathrm{Kin}}}}{N^{B^0_{(s)}\to h^+h'^-}} \tag{5.3}$$

where $\epsilon^{\mathrm{TISTOSMap}}$ is the TISTOS map, $\mathrm{i}_{\mathrm{Kin}}$ the kinematic bin of the TISTOS map and $N^{B^0_{(s)}\to h^+h'^-}$ the $B^0_{(s)} \to h^+h'^-$ yields obtained from simulation samples. In the case of the BDT-binned efficiencies, the convolution is done with the corresponding BDT bin, $\mathrm{i}_{\mathrm{BDT}}$, as

$$\epsilon^{hh'}_{\mathrm{i}_{\mathrm{BDT}},\mathrm{L0\&HLT1}} = \frac{\sum_{\mathrm{i}_{\mathrm{Kin}}} \epsilon^{\mathrm{TISTOSMap}}_{\mathrm{i}_{\mathrm{Kin}}} \cdot N^{B^0_{(s)}\to h^+h'^-}_{\mathrm{i}_{\mathrm{BDT}},\mathrm{i}_{\mathrm{Kin}}}}{N^{B^0_{(s)}\to h^+h'^-}_{\mathrm{i}_{\mathrm{BDT}}}} \tag{5.4}$$

Even though the application of the maps is found to yield the correct results for the overall trigger efficiencies (see Sec. 4.2), it does not properly reproduce the expected distributions of the L0&HLT1 efficiencies per BDT bin. The *MC true map*, introduced in Sec. 4.2, is used to illustrate and explain this discrepancy. The MC true map is identical to the TISTOS map, except that the L0&HLT1 efficiencies are evaluated as the true efficiencies[2]. Contrary to the

[2]Fraction of events passing the L0&HLT1 selection in the sample.

TISTOS map, the MC true map can only be obtained from simulation samples, as the total number of events, including untriggered events, must be known.

The application of the maps can be tested using simulated $B^0 \to K^+\pi^-$ signal samples. An MC true map is obtained from these samples and convolved with the very same simulated $B^0 \to K^+\pi^-$ events to determine the BDT-binned and overall L0&HLT1 efficiencies, as described in Eq. 5.3 and Eq. 5.4. The resulting relative L0&HLT1 efficiencies are compared with the relative L0&HLT1 efficiencies evaluated without the use of maps, directly from the simulated $B^0 \to K^+\pi^-$ samples as the ratio of the BDT-binned and overall true efficiencies:

$$\frac{\epsilon_{\mathrm{i_{BDT}},\mathrm{L0\&HLT1}}}{\epsilon_{\mathrm{L0\&HLT1}}} = \frac{N_{\mathrm{i_{BDT}},\mathrm{L0\&HLT1}}/N_{\mathrm{i_{BDT}}}}{N_{\mathrm{L0\&HLT1}}/N} \tag{5.5}$$

The comparison is illustrated in Fig. 5.4. Although both methods should mathematically give identical results, the disagreement between their values is large. This discrepancy implies that an existing correlation between the BDT and the L0&HLT1 TIS selection is not accounted for by either the MC true or TISTOS maps. This effect should be included to correctly evaluate the relative L0&HLT1 efficiencies for the $B^0_{(s)} \to h^+h'^-$ decays.

The correlation between the L0&HLT1 TIS selection and the BDT is not trivial as it is due to a third variable, concretely the occupancy of the event. Two of the variables used in the BDT training, the long and VELO track isolation variables, quantify the isolation of the final-state track candidates, which is closely related to the event occupancy. Furthermore, the L0&HLT1 TIS selection is also correlated with the event occupancy since TIS lines are triggered by the underlying event, which is characterized by the event occupancy among other variables.

The correlation between the BDT and L0&HLT1 TIS selection is too complex to generally correct for it in all the BDT bins. Instead, a BDT-dependent correction is applied to the TISTOS map when evaluating the BDT-binned efficiencies. Ideally, a different TISTOS map from $B^+ \to J/\psi K^+$ data samples could be determined for each BDT bin. Each map would be later convolved with the corresponding $B^0_{(s)} \to h^+h'^-$ kinematic spectrum for that BDT bin to calculate the BDT-binned efficiencies. Unfortunately, this is not a suitable solution because the BDT response strongly depends on the decay type it is applied to. The BDT response is different for $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays due to their distinct decay topology and the number of final-state particles in each channel.

Alternatively, a BDT-dependent correction is determined from simulated $B^0_{(s)} \to h^+h'^-$ events defined as the ratio between the MC true maps of the BDT binned and total samples, $\epsilon^{\mathrm{MCTrueMap}}_{(\mathrm{i_{BDT}})}$, such that for the $\mathrm{i_{Kin}}$ kinematic bin, the BDT-dependent correction, $h^{corr}_{\mathrm{i_{BDT}},\mathrm{i_{Kin}}}$, is

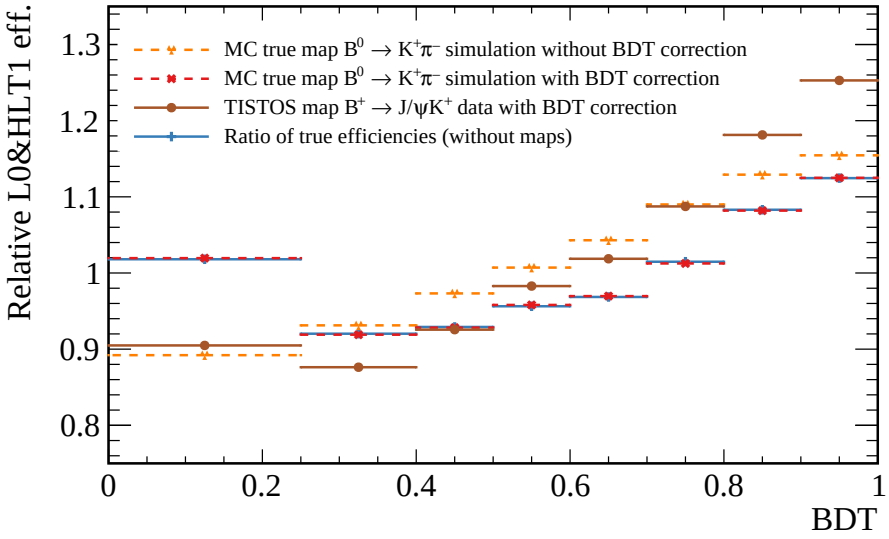


Figure 5.4: Relative L0&HLT1 efficiencies per BDT bin for $B^0 \to K^+\pi^-$ decays evaluated using four different approaches. The only difference between the orange and red distributions is the addition of the BDT-dependent correction in the latter case. The overlap between the values obtained from simulation without maps (blue) and the results from the MC true map from $B^0 \to K^+\pi^-$ simulation including the BDT-dependent correction (orange) is excellent. For completeness, the nominal procedure to evaluate the relative L0&HLT1 efficiencies, the use of the TISTOS map from $B^+ \to J/\psi K^+$ corrected with the BDT-dependence, is also added in brown. The difference between the nominal and the simulation values is due to the mismodelling of the L0&HLT1 TIS selection criterion in the simulation.

expressed as:

$$h^{corr}_{\mathrm{i_{BDT},i_{Kin}}} = \frac{\epsilon^{\mathrm{MCTrueMap}}_{\mathrm{i_{BDT},i_{Kin}}}}{\epsilon^{\mathrm{MCTrueMap}}_{\mathrm{i_{Kin}}}} \quad \text{where} \quad \epsilon^{\mathrm{MCTrueMap}}_{\mathrm{(i_{BDT}),i_{Kin}}} = \frac{N^{hh'}_{\mathrm{(i_{BDT}),i_{Kin},L0\&HLT1}}}{N^{hh'}_{\mathrm{(i_{BDT}),i_{Kin}}}}$$

For the calculation of the BDT-binned efficiencies, the TISTOS map is multiplied with the BDT-dependent correction such that a different factor corrects each kinematic bin. Adding the BDT-dependent correction, Eq. 5.4 becomes:

$$\epsilon^{hh'}_{\mathrm{i_{BDT},L0\&HLT1}} = \frac{\sum_{\mathrm{i_{Kin}}} \epsilon^{\mathrm{TISTOSMap}}_{\mathrm{i_{Kin}}} \cdot h^{corr}_{\mathrm{i_{BDT},i_{Kin}}} \cdot N^{B^0_{(s)}\to h^+h'^-}_{\mathrm{i_{BDT},i_{Kin}}}}{N^{B^0_{(s)}\to h^+h'^-}_{\mathrm{i_{BDT}}}} \tag{5.6}$$

which is the final expression used for the BDT-binned efficiencies.

The effect of the BDT-dependent correction is checked in simulated $B^0 \to K^+\pi^-$ events. The relative L0&HLT1 efficiencies are evaluated using the MC true map once corrected for

the BDT-dependent correction, with the BDT-binned and overall efficiencies determined as given in Eq. 5.6 and Eq. 5.3, respectively. The resulting distribution is plotted in Fig. 5.4 and compared with the distribution obtained from simulation without the use of maps. The perfect agreement between the two distributions shows that the BDT-dependent correction entirely captures the correlation effects between the BDT and the L0&HLT1 TIS selection criteria.

For completeness, the relative L0&HLT1 efficiencies for $B^0 \to K^+\pi^-$ decays obtained with the nominal method, namely the BDT-corrected TISTOS map from $B^+ \to J/\psi K^+$ sWeighted data, are also shown in Fig. 5.4. The difference relative to the values from simulation samples illustrates the mismodelling of the L0&HLT1 TIS performance in simulation.

Relative HLT2 efficiencies

The evaluation of the relative HLT2 efficiencies per BDT bin is simpler. They are defined as the ratio between the BDT-binned and overall HLT2 efficiencies. Since the HLT2 behaviour is well-modelled in simulation, both values are evaluated directly from simulation samples. The overall HLT2 efficiency is calculated as the fraction of events passing the HLT2 selection in the simulation samples. The BDT-binned values are computed in the same manner in the corresponding BDT bins. Prior to evaluating the HLT2 efficiencies, the simulation samples are required to pass the L0&HLT1 selection since the HLT2 selection is applied after the L0&HLT1 level.

Systematic checks

The robustness of the method to estimate the relative $B^0_{(s)} \to h^+h'^-$ trigger efficiencies is analyzed similarly to the overall trigger efficiencies, presented in Sec. 4.2. While no test is needed for the HLT2 procedure, several checks are performed to investigate possible systematic effects in the evaluation of the relative L0&HLT1 efficiencies. One of these effects is the binning scheme used for the TISTOS map, tested by changing the binning scheme merging two consecutive bins. The difference with respect to the nominal method is taken as systematic uncertainty.

Apart from the binning scheme, the usage of $B^+ \to J/\psi K^+$ decays as a proxy of the $B^0_{(s)} \to h^+h'^-$ channel is studied as a second systematic effect. This effect is examined using simulated $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ events. An MC true map is obtained from each simulation sample and convolved with the $B^0 \to K^+\pi^-$ simulation sample. Both MC true maps include the BDT-dependent correction. The resulting relative L0&HLT1 efficiency distributions are shown in Fig. 5.5. Although small, the discrepancy between the two shapes is not negligible. For this reason, the difference between both distributions is taken as systematic error.

A third source of systematic uncertainty is the precision of the TISTOS method. Its effect on the relative L0&HLT1 efficiencies is studied by obtaining an MC true and a TISTOS map from the $B^+ \to J/\psi K^+$ simulation samples. Each map is convolved with the $B^0 \to K^+\pi^-$ simulation samples to obtain the $B^0 \to K^+\pi^-$ trigger efficiencies. The BDT-dependent correction is applied in both cases to calculate the BDT-binned efficiencies. The distributions of the resulting relative L0&HLT1 efficiencies are illustrated in Fig. 5.5. The discrepancy between the two values is negligible and no further systematic is included.

Although not included in the systematic checks, the application of the maps to evaluate the relative L0&HLT1 efficiencies should also be tested. This was already checked in Fig. 5.4, together with the BDT-dependent correction. For completeness, the results obtained directly from simulation are also included in Fig. 5.5.

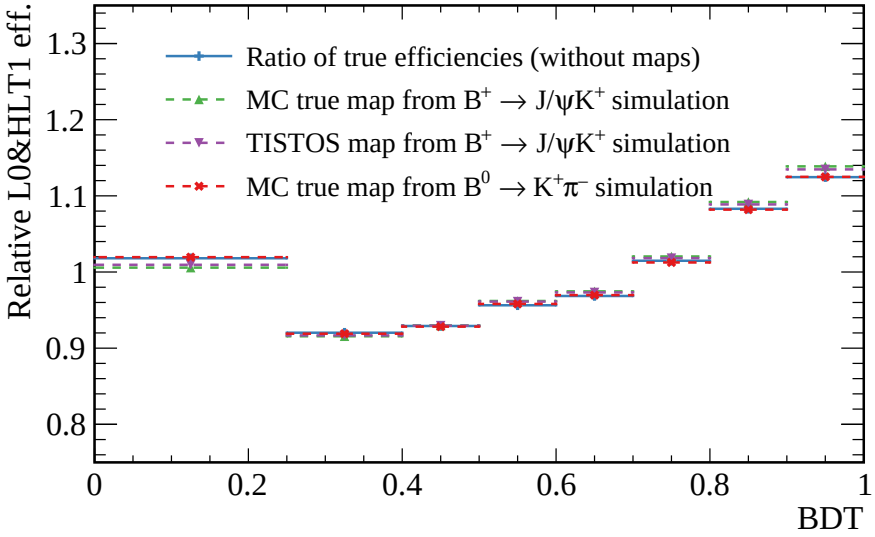


Figure 5.5: Relative TIS efficiencies per BDT bin for $B^0 \to K^+\pi^-$ decays obtained from the MC true map derived from $B^+ \to J/\psi K^+$ simulation (green), the TISTOS map from $B^+ \to J/\psi K^+$ simulation (purple), the MC true map from $B^0 \to K^+\pi^-$ simulation (red) and without the use of the maps (blue).

Final values

Even though the need for the BDT-dependent correction and the systematic checks have been presented for the $B^0 \to K^+\pi^-$ case, the same conclusions and procedures apply to the $B^0_s \to K^+K^-$ decay. The relative trigger efficiencies per BDT bin for the $B^0 \to K^+\pi^-$ mode can be found in the third column of Table 5.1 and Table 5.3, where the first uncertainty is statistical and the second systematic. The corresponding $B^0_s \to K^+K^-$ values are listed in Table 5.2 and Table 5.4.

5.1.3 Hadronic PID correction

The $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ candidates are selected in data by requiring the presence of either a pion and a kaon or two kaons as final-state particles. The hadronic PID criterion applied to the data samples is clearly different from that required for $B^0_{(s)} \to \mu^+\mu^-$ candidates and, hence, has to be corrected. Since this effect is expected to depend on the BDT values, a PID correction factor per BDT is determined. The correction for each BDT bin is evaluated as the ratio between the PID efficiency per BDT bin and the overall efficiency.

Both the BDT-binned and overall PID efficiencies for $B^0_{(s)} \to h^+h'^-$ exclusive decays are calculated from data samples using $D^{*\pm} \to (D^0 \to K^\pm\pi^\mp)\pi_s^\pm$ decays. From these decays, the single hadron PID efficiencies are obtained in bins of the occupancy of the event, the momentum and the transverse momentum of the hadron. The PID efficiency of a reconstructed particle results from the convolution of the corresponding single hadron PID efficiencies with the kinematic spectrum of that particle obtained from simulation. This convolution has to be done with each of the two hadrons in the final state, such that the PID efficiencies of the $B^0_{(s)} \to h^+h'^-$ decays are the product of the single hadron PID efficiencies of each hadron in the final state. A detailed explanation of the evaluation of the $B^0_{(s)} \to h^+h'^-$ PID efficiencies can be found in Sec. 4.3.

The kinematic spectrum for each hadron in the exclusive $B^0_{(s)} \to h^+h'^-$ decays is obtained from exclusive $B^0_{(s)} \to h^+h'^-$ simulation samples. While for the overall efficiencies, the total simulation sample is used, for the BDT-binned efficiencies, the kinematic spectrum is obtained for the BDT bins.

The PID corrections per BDT bin of the $B^0 \to K^+\pi^-$ decays with their statistical and systematic uncertainties can be found in the fourth column of Tables 5.1 and 5.3. The $B_s^0 \to K^+K^-$ values are listed in Tables 5.2 and 5.4. The PID efficiencies of each BDT subsample and the full sample are assumed to be correlated as the same PID histogram is used to evaluate them. Regarding the systematic uncertainties, only the choice of the binning in the PIDCalib histogram is considered a source of systematic error. This effect is studied by evaluating the PID efficiencies using a different binning scheme. The difference between these values and the nominal ones is taken as systematic uncertainty.

5.1.4 Results

The fraction of events estimated per BDT bin in each data-taking year using $B^0 \to K^+\pi^-$ data samples are listed in the fifth column of Tables 5.1 and 5.3. The same values estimated from $B_s^0 \to K^+K^-$ data samples are displayed in Tables 5.2 and 5.4. These values are determined per BDT bin from the correction of the relative $B^0_{(s)} \to h^+h'^-$ data yields by the PID and trigger correction factors using Eq. 5.1.

Two assumptions are made in the determination of the statistical and systematic un-

certainties associated with the corrected fraction of events per BDT bin. The trigger and PID corrections are considered to be fully correlated within each BDT bin since the same simulation samples are exploited as kinematic reference in their evaluation. Because the relative yields are obtained using data samples only, their correlation with the trigger and PID corrections is assumed to be negligible.

The systematic uncertainty includes the various sources of error affecting the relative data yields and the PID and trigger corrections. Only the presence of the partly reconstructed background is included as systematic error in the relative data yields. For the trigger correction, the effect of the TISTOS binning scheme and the use of the $B^+ \to J/\psi K^+$ decays as a proxy are the sources of systematic uncertainty. In the case of the PID correction, only the binning scheme used in the PID histograms is included as systematic error.

Table 5.1: Relative fitted yields, trigger and PID correction and final corrected relative yields for $B^0 \to K^+\pi^-$ for all Run1 data-taking years.

2011				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.279 \pm 0.037 \pm 0.014$	$0.786 \pm 0.013 \pm 0.017$	$1.275 \pm 0.036 \pm 0.002$	$0.278 \pm 0.039 \pm 0.016$
$[0.25, 0.4)$	$0.166 \pm 0.007 \pm 0.002$	$0.856 \pm 0.018 \pm 0.009$	$1.160 \pm 0.053 \pm 0.000$	$0.167 \pm 0.013 \pm 0.003$
$[0.4, 0.5)$	$0.105 \pm 0.006 \pm 0.002$	$0.917 \pm 0.024 \pm 0.002$	$1.065 \pm 0.066 \pm 0.006$	$0.108 \pm 0.011 \pm 0.002$
$[0.5, 0.6)$	$0.103 \pm 0.006 \pm 0.003$	$0.948 \pm 0.024 \pm 0.002$	$0.974 \pm 0.062 \pm 0.003$	$0.111 \pm 0.012 \pm 0.003$
$[0.6, 0.7)$	$0.081 \pm 0.005 \pm 0.001$	$1.044 \pm 0.026 \pm 0.007$	$0.935 \pm 0.058 \pm 0.002$	$0.083 \pm 0.009 \pm 0.001$
$[0.7, 0.8)$	$0.084 \pm 0.005 \pm 0.001$	$1.170 \pm 0.028 \pm 0.015$	$0.835 \pm 0.051 \pm 0.001$	$0.086 \pm 0.009 \pm 0.002$
$[0.8, 0.9)$	$0.087 \pm 0.005 \pm 0.002$	$1.306 \pm 0.030 \pm 0.016$	$0.780 \pm 0.046 \pm 0.004$	$0.086 \pm 0.009 \pm 0.002$
$[0.9, 1]$	$0.096 \pm 0.005 \pm 0.002$	$1.461 \pm 0.031 \pm 0.026$	$0.761 \pm 0.041 \pm 0.000$	$0.086 \pm 0.008 \pm 0.002$
2012				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.306 \pm 0.022 \pm 0.012$	$0.855 \pm 0.005 \pm 0.018$	$1.270 \pm 0.010 \pm 0.004$	$0.282 \pm 0.020 \pm 0.013$
$[0.25, 0.4)$	$0.170 \pm 0.004 \pm 0.002$	$0.874 \pm 0.006 \pm 0.008$	$1.156 \pm 0.016 \pm 0.002$	$0.169 \pm 0.006 \pm 0.003$
$[0.4, 0.5)$	$0.099 \pm 0.003 \pm 0.002$	$0.920 \pm 0.007 \pm 0.003$	$1.066 \pm 0.020 \pm 0.000$	$0.101 \pm 0.004 \pm 0.002$
$[0.5, 0.6)$	$0.099 \pm 0.003 \pm 0.002$	$0.976 \pm 0.008 \pm 0.003$	$0.973 \pm 0.019 \pm 0.000$	$0.104 \pm 0.005 \pm 0.002$
$[0.6, 0.7)$	$0.087 \pm 0.003 \pm 0.002$	$1.024 \pm 0.008 \pm 0.007$	$0.895 \pm 0.017 \pm 0.000$	$0.094 \pm 0.004 \pm 0.002$
$[0.7, 0.8)$	$0.080 \pm 0.003 \pm 0.001$	$1.119 \pm 0.009 \pm 0.013$	$0.811 \pm 0.016 \pm 0.004$	$0.089 \pm 0.004 \pm 0.002$
$[0.8, 0.9)$	$0.078 \pm 0.003 \pm 0.001$	$1.251 \pm 0.009 \pm 0.019$	$0.764 \pm 0.014 \pm 0.002$	$0.082 \pm 0.004 \pm 0.002$
$[0.9, 1]$	$0.081 \pm 0.003 \pm 0.001$	$1.362 \pm 0.011 \pm 0.027$	$0.758 \pm 0.014 \pm 0.003$	$0.078 \pm 0.004 \pm 0.002$

Table 5.2: Relative fitted yields, trigger and PID correction and final corrected relative yields for $B_s^0 \to K^+K^-$ for all Run1 data-taking years.

2011				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.326 \pm 0.054 \pm 0.006$	$0.789 \pm 0.010 \pm 0.018$	$1.291 \pm 0.024 \pm 0.002$	$0.320 \pm 0.054 \pm 0.010$
$[0.25, 0.4)$	$0.175 \pm 0.012 \pm 0.002$	$0.844 \pm 0.013 \pm 0.010$	$1.163 \pm 0.036 \pm 0.001$	$0.178 \pm 0.015 \pm 0.003$
$[0.4, 0.5)$	$0.086 \pm 0.008 \pm 0.000$	$0.938 \pm 0.018 \pm 0.002$	$1.072 \pm 0.044 \pm 0.006$	$0.086 \pm 0.010 \pm 0.001$
$[0.5, 0.6)$	$0.081 \pm 0.008 \pm 0.001$	$0.984 \pm 0.018 \pm 0.004$	$0.986 \pm 0.042 \pm 0.002$	$0.083 \pm 0.009 \pm 0.001$
$[0.6, 0.7)$	$0.091 \pm 0.008 \pm 0.000$	$1.052 \pm 0.019 \pm 0.010$	$0.927 \pm 0.039 \pm 0.002$	$0.093 \pm 0.010 \pm 0.001$
$[0.7, 0.8)$	$0.085 \pm 0.008 \pm 0.002$	$1.113 \pm 0.020 \pm 0.010$	$0.834 \pm 0.034 \pm 0.000$	$0.091 \pm 0.010 \pm 0.002$
$[0.8, 0.9)$	$0.076 \pm 0.008 \pm 0.000$	$1.326 \pm 0.022 \pm 0.021$	$0.771 \pm 0.030 \pm 0.005$	$0.075 \pm 0.008 \pm 0.002$
$[0.9, 1]$	$0.080 \pm 0.008 \pm 0.002$	$1.452 \pm 0.024 \pm 0.025$	$0.740 \pm 0.027 \pm 0.002$	$0.075 \pm 0.008 \pm 0.002$
2012				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.352 \pm 0.032 \pm 0.006$	$0.848 \pm 0.005 \pm 0.017$	$1.291 \pm 0.007 \pm 0.004$	$0.322 \pm 0.030 \pm 0.009$
$[0.25, 0.4)$	$0.173 \pm 0.007 \pm 0.001$	$0.870 \pm 0.005 \pm 0.008$	$1.161 \pm 0.011 \pm 0.002$	$0.171 \pm 0.008 \pm 0.002$
$[0.4, 0.5)$	$0.097 \pm 0.006 \pm 0.001$	$0.915 \pm 0.007 \pm 0.003$	$1.070 \pm 0.014 \pm 0.001$	$0.099 \pm 0.006 \pm 0.001$
$[0.5, 0.6)$	$0.080 \pm 0.005 \pm 0.001$	$0.975 \pm 0.007 \pm 0.004$	$0.974 \pm 0.013 \pm 0.000$	$0.084 \pm 0.006 \pm 0.001$
$[0.6, 0.7)$	$0.082 \pm 0.005 \pm 0.001$	$1.048 \pm 0.008 \pm 0.008$	$0.893 \pm 0.012 \pm 0.001$	$0.088 \pm 0.006 \pm 0.001$
$[0.7, 0.8)$	$0.067 \pm 0.004 \pm 0.001$	$1.122 \pm 0.008 \pm 0.013$	$0.804 \pm 0.010 \pm 0.003$	$0.074 \pm 0.005 \pm 0.001$
$[0.8, 0.9)$	$0.079 \pm 0.005 \pm 0.001$	$1.263 \pm 0.009 \pm 0.019$	$0.749 \pm 0.009 \pm 0.003$	$0.084 \pm 0.006 \pm 0.002$
$[0.9, 1]$	$0.069 \pm 0.005 \pm 0.001$	$1.354 \pm 0.010 \pm 0.024$	$0.734 \pm 0.009 \pm 0.003$	$0.070 \pm 0.005 \pm 0.002$

Table 5.3: Relative fitted yields, trigger and PID correction and final corrected relative yields for $B^0 \to K^+\pi^-$ for all Run2 data-taking years.

2015				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.335 \pm 0.033 \pm 0.006$	$0.809 \pm 0.010 \pm 0.017$	$1.314 \pm 0.018 \pm 0.003$	$0.315 \pm 0.032 \pm 0.009$
$[0.25, 0.4)$	$0.162 \pm 0.007 \pm 0.001$	$0.871 \pm 0.010 \pm 0.009$	$1.162 \pm 0.027 \pm 0.000$	$0.160 \pm 0.009 \pm 0.002$
$[0.4, 0.5)$	$0.099 \pm 0.005 \pm 0.001$	$0.932 \pm 0.012 \pm 0.003$	$1.069 \pm 0.035 \pm 0.002$	$0.099 \pm 0.007 \pm 0.001$
$[0.5, 0.6)$	$0.084 \pm 0.005 \pm 0.002$	$1.028 \pm 0.013 \pm 0.003$	$0.969 \pm 0.031 \pm 0.003$	$0.084 \pm 0.006 \pm 0.003$
$[0.6, 0.7)$	$0.074 \pm 0.004 \pm 0.002$	$1.069 \pm 0.014 \pm 0.008$	$0.866 \pm 0.028 \pm 0.003$	$0.080 \pm 0.006 \pm 0.002$
$[0.7, 0.8)$	$0.076 \pm 0.004 \pm 0.001$	$1.134 \pm 0.015 \pm 0.013$	$0.790 \pm 0.026 \pm 0.002$	$0.085 \pm 0.006 \pm 0.002$
$[0.8, 0.9)$	$0.078 \pm 0.005 \pm 0.002$	$1.221 \pm 0.015 \pm 0.017$	$0.771 \pm 0.025 \pm 0.002$	$0.083 \pm 0.006 \pm 0.002$
$[0.9, 1]$	$0.092 \pm 0.005 \pm 0.001$	$1.353 \pm 0.017 \pm 0.023$	$0.779 \pm 0.022 \pm 0.002$	$0.088 \pm 0.006 \pm 0.002$
2016				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.347 \pm 0.014 \pm 0.017$	$0.870 \pm 0.005 \pm 0.018$	$1.311 \pm 0.011 \pm 0.005$	$0.305 \pm 0.013 \pm 0.017$
$[0.25, 0.4)$	$0.163 \pm 0.003 \pm 0.004$	$0.873 \pm 0.006 \pm 0.007$	$1.162 \pm 0.018 \pm 0.003$	$0.161 \pm 0.005 \pm 0.004$
$[0.4, 0.5)$	$0.095 \pm 0.002 \pm 0.002$	$0.935 \pm 0.008 \pm 0.002$	$1.039 \pm 0.022 \pm 0.001$	$0.098 \pm 0.004 \pm 0.002$
$[0.5, 0.6)$	$0.090 \pm 0.002 \pm 0.002$	$0.997 \pm 0.008 \pm 0.005$	$0.935 \pm 0.020 \pm 0.001$	$0.097 \pm 0.004 \pm 0.003$
$[0.6, 0.7)$	$0.077 \pm 0.002 \pm 0.002$	$1.034 \pm 0.009 \pm 0.008$	$0.854 \pm 0.019 \pm 0.001$	$0.087 \pm 0.004 \pm 0.002$
$[0.7, 0.8)$	$0.069 \pm 0.002 \pm 0.002$	$1.110 \pm 0.009 \pm 0.011$	$0.794 \pm 0.018 \pm 0.005$	$0.078 \pm 0.003 \pm 0.002$
$[0.8, 0.9)$	$0.075 \pm 0.002 \pm 0.002$	$1.209 \pm 0.010 \pm 0.016$	$0.752 \pm 0.016 \pm 0.003$	$0.082 \pm 0.003 \pm 0.002$
$[0.9, 1]$	$0.084 \pm 0.002 \pm 0.002$	$1.286 \pm 0.010 \pm 0.023$	$0.761 \pm 0.015 \pm 0.004$	$0.085 \pm 0.003 \pm 0.003$
2017				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.344 \pm 0.015 \pm 0.015$	$0.858 \pm 0.005 \pm 0.014$	$1.305 \pm 0.011 \pm 0.004$	$0.307 \pm 0.014 \pm 0.015$
$[0.25, 0.4)$	$0.160 \pm 0.003 \pm 0.003$	$0.888 \pm 0.006 \pm 0.007$	$1.158 \pm 0.017 \pm 0.002$	$0.155 \pm 0.004 \pm 0.003$
$[0.4, 0.5)$	$0.096 \pm 0.002 \pm 0.003$	$0.944 \pm 0.008 \pm 0.003$	$1.029 \pm 0.021 \pm 0.000$	$0.099 \pm 0.004 \pm 0.003$
$[0.5, 0.6)$	$0.088 \pm 0.002 \pm 0.002$	$0.997 \pm 0.008 \pm 0.003$	$0.941 \pm 0.020 \pm 0.002$	$0.094 \pm 0.004 \pm 0.003$
$[0.6, 0.7)$	$0.078 \pm 0.002 \pm 0.002$	$1.036 \pm 0.008 \pm 0.006$	$0.861 \pm 0.019 \pm 0.001$	$0.087 \pm 0.003 \pm 0.002$
$[0.7, 0.8)$	$0.070 \pm 0.002 \pm 0.002$	$1.104 \pm 0.009 \pm 0.009$	$0.785 \pm 0.017 \pm 0.002$	$0.081 \pm 0.003 \pm 0.002$
$[0.8, 0.9)$	$0.074 \pm 0.002 \pm 0.002$	$1.208 \pm 0.009 \pm 0.014$	$0.761 \pm 0.016 \pm 0.003$	$0.081 \pm 0.003 \pm 0.002$
$[0.9, 1]$	$0.090 \pm 0.002 \pm 0.002$	$1.289 \pm 0.009 \pm 0.022$	$0.770 \pm 0.015 \pm 0.003$	$0.090 \pm 0.003 \pm 0.003$
2018				
BDT bin	$f^{K\pi}$	$w^{K\pi}_{\text{trg}}$	$w^{K\pi}_{\text{PID}}$	f
$[0, 0.25)$	$0.346 \pm 0.013 \pm 0.017$	$0.848 \pm 0.005 \pm 0.019$	$1.314 \pm 0.012 \pm 0.004$	$0.311 \pm 0.013 \pm 0.017$
$[0.25, 0.4)$	$0.159 \pm 0.003 \pm 0.004$	$0.880 \pm 0.006 \pm 0.007$	$1.166 \pm 0.018 \pm 0.003$	$0.155 \pm 0.004 \pm 0.004$
$[0.4, 0.5)$	$0.094 \pm 0.002 \pm 0.002$	$0.928 \pm 0.008 \pm 0.001$	$1.032 \pm 0.023 \pm 0.002$	$0.098 \pm 0.004 \pm 0.002$
$[0.5, 0.6)$	$0.085 \pm 0.002 \pm 0.002$	$0.998 \pm 0.009 \pm 0.004$	$0.947 \pm 0.021 \pm 0.002$	$0.090 \pm 0.003 \pm 0.002$
$[0.6, 0.7)$	$0.080 \pm 0.002 \pm 0.002$	$1.056 \pm 0.009 \pm 0.006$	$0.864 \pm 0.020 \pm 0.001$	$0.087 \pm 0.003 \pm 0.003$
$[0.7, 0.8)$	$0.072 \pm 0.002 \pm 0.002$	$1.110 \pm 0.010 \pm 0.013$	$0.782 \pm 0.019 \pm 0.002$	$0.083 \pm 0.003 \pm 0.002$
$[0.8, 0.9)$	$0.076 \pm 0.002 \pm 0.002$	$1.203 \pm 0.010 \pm 0.017$	$0.765 \pm 0.017 \pm 0.004$	$0.082 \pm 0.003 \pm 0.003$
$[0.9, 1]$	$0.088 \pm 0.002 \pm 0.002$	$1.321 \pm 0.010 \pm 0.025$	$0.758 \pm 0.015 \pm 0.003$	$0.088 \pm 0.003 \pm 0.003$

Table 5.4: Relative fitted yields, trigger and PID correction and final corrected relative yields for $B_s^0 \to K^+K^-$ for all Run2 data-taking years.

2015				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.293 \pm 0.056 \pm 0.007$	$0.812 \pm 0.010 \pm 0.018$	$1.309 \pm 0.012 \pm 0.002$	$0.276 \pm 0.053 \pm 0.009$
$[0.25, 0.4)$	$0.180 \pm 0.012 \pm 0.002$	$0.868 \pm 0.010 \pm 0.008$	$1.160 \pm 0.018 \pm 0.001$	$0.179 \pm 0.013 \pm 0.003$
$[0.4, 0.5)$	$0.101 \pm 0.009 \pm 0.001$	$0.963 \pm 0.013 \pm 0.002$	$1.071 \pm 0.023 \pm 0.001$	$0.098 \pm 0.009 \pm 0.001$
$[0.5, 0.6)$	$0.106 \pm 0.009 \pm 0.006$	$1.010 \pm 0.013 \pm 0.006$	$0.968 \pm 0.020 \pm 0.002$	$0.108 \pm 0.010 \pm 0.006$
$[0.6, 0.7)$	$0.073 \pm 0.008 \pm 0.000$	$1.073 \pm 0.014 \pm 0.007$	$0.866 \pm 0.019 \pm 0.002$	$0.079 \pm 0.009 \pm 0.001$
$[0.7, 0.8)$	$0.089 \pm 0.008 \pm 0.002$	$1.144 \pm 0.015 \pm 0.011$	$0.798 \pm 0.018 \pm 0.003$	$0.097 \pm 0.009 \pm 0.002$
$[0.8, 0.9)$	$0.073 \pm 0.008 \pm 0.000$	$1.216 \pm 0.016 \pm 0.015$	$0.771 \pm 0.016 \pm 0.001$	$0.077 \pm 0.009 \pm 0.001$
$[0.9, 1]$	$0.085 \pm 0.008 \pm 0.001$	$1.336 \pm 0.017 \pm 0.025$	$0.783 \pm 0.015 \pm 0.002$	$0.081 \pm 0.008 \pm 0.002$
2016				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.382 \pm 0.021 \pm 0.004$	$0.866 \pm 0.005 \pm 0.020$	$1.310 \pm 0.007 \pm 0.004$	$0.337 \pm 0.019 \pm 0.009$
$[0.25, 0.4)$	$0.157 \pm 0.005 \pm 0.001$	$0.877 \pm 0.006 \pm 0.007$	$1.160 \pm 0.012 \pm 0.003$	$0.155 \pm 0.005 \pm 0.002$
$[0.4, 0.5)$	$0.097 \pm 0.004 \pm 0.001$	$0.945 \pm 0.008 \pm 0.003$	$1.038 \pm 0.015 \pm 0.000$	$0.099 \pm 0.004 \pm 0.001$
$[0.5, 0.6)$	$0.084 \pm 0.003 \pm 0.000$	$0.984 \pm 0.008 \pm 0.006$	$0.938 \pm 0.014 \pm 0.001$	$0.091 \pm 0.004 \pm 0.001$
$[0.6, 0.7)$	$0.074 \pm 0.003 \pm 0.000$	$1.039 \pm 0.009 \pm 0.009$	$0.854 \pm 0.013 \pm 0.001$	$0.083 \pm 0.004 \pm 0.001$
$[0.7, 0.8)$	$0.068 \pm 0.003 \pm 0.000$	$1.115 \pm 0.009 \pm 0.013$	$0.794 \pm 0.012 \pm 0.005$	$0.077 \pm 0.004 \pm 0.001$
$[0.8, 0.9)$	$0.065 \pm 0.003 \pm 0.000$	$1.208 \pm 0.010 \pm 0.016$	$0.753 \pm 0.011 \pm 0.003$	$0.071 \pm 0.004 \pm 0.001$
$[0.9, 1]$	$0.073 \pm 0.003 \pm 0.001$	$1.289 \pm 0.010 \pm 0.022$	$0.765 \pm 0.010 \pm 0.003$	$0.074 \pm 0.004 \pm 0.002$
2017				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.374 \pm 0.022 \pm 0.005$	$0.867 \pm 0.005 \pm 0.016$	$1.296 \pm 0.007 \pm 0.004$	$0.333 \pm 0.020 \pm 0.009$
$[0.25, 0.4)$	$0.163 \pm 0.005 \pm 0.001$	$0.875 \pm 0.006 \pm 0.008$	$1.155 \pm 0.011 \pm 0.002$	$0.162 \pm 0.005 \pm 0.002$
$[0.4, 0.5)$	$0.100 \pm 0.004 \pm 0.001$	$0.940 \pm 0.008 \pm 0.002$	$1.030 \pm 0.014 \pm 0.000$	$0.103 \pm 0.004 \pm 0.001$
$[0.5, 0.6)$	$0.084 \pm 0.003 \pm 0.001$	$0.998 \pm 0.008 \pm 0.003$	$0.942 \pm 0.013 \pm 0.001$	$0.089 \pm 0.004 \pm 0.001$
$[0.6, 0.7)$	$0.073 \pm 0.003 \pm 0.001$	$1.051 \pm 0.008 \pm 0.007$	$0.864 \pm 0.012 \pm 0.000$	$0.080 \pm 0.004 \pm 0.001$
$[0.7, 0.8)$	$0.060 \pm 0.003 \pm 0.001$	$1.118 \pm 0.009 \pm 0.012$	$0.790 \pm 0.012 \pm 0.002$	$0.068 \pm 0.004 \pm 0.001$
$[0.8, 0.9)$	$0.070 \pm 0.003 \pm 0.001$	$1.186 \pm 0.009 \pm 0.016$	$0.765 \pm 0.010 \pm 0.003$	$0.077 \pm 0.004 \pm 0.001$
$[0.9, 1]$	$0.075 \pm 0.003 \pm 0.001$	$1.284 \pm 0.010 \pm 0.022$	$0.777 \pm 0.010 \pm 0.003$	$0.075 \pm 0.004 \pm 0.002$
2018				
BDT bin	f^{KK}	w_{trg}^{KK}	w_{PID}^{KK}	f
$[0, 0.25)$	$0.349 \pm 0.020 \pm 0.002$	$0.852 \pm 0.005 \pm 0.019$	$1.307 \pm 0.008 \pm 0.004$	$0.313 \pm 0.018 \pm 0.008$
$[0.25, 0.4)$	$0.166 \pm 0.004 \pm 0.000$	$0.863 \pm 0.006 \pm 0.008$	$1.160 \pm 0.012 \pm 0.003$	$0.165 \pm 0.005 \pm 0.002$
$[0.4, 0.5)$	$0.096 \pm 0.003 \pm 0.000$	$0.945 \pm 0.008 \pm 0.001$	$1.031 \pm 0.015 \pm 0.001$	$0.099 \pm 0.004 \pm 0.001$
$[0.5, 0.6)$	$0.086 \pm 0.003 \pm 0.000$	$0.994 \pm 0.009 \pm 0.003$	$0.948 \pm 0.014 \pm 0.002$	$0.091 \pm 0.004 \pm 0.001$
$[0.6, 0.7)$	$0.077 \pm 0.003 \pm 0.000$	$1.052 \pm 0.009 \pm 0.009$	$0.869 \pm 0.013 \pm 0.001$	$0.085 \pm 0.004 \pm 0.001$
$[0.7, 0.8)$	$0.072 \pm 0.003 \pm 0.000$	$1.106 \pm 0.010 \pm 0.014$	$0.786 \pm 0.012 \pm 0.002$	$0.083 \pm 0.004 \pm 0.001$
$[0.8, 0.9)$	$0.070 \pm 0.003 \pm 0.001$	$1.225 \pm 0.010 \pm 0.018$	$0.766 \pm 0.011 \pm 0.004$	$0.075 \pm 0.004 \pm 0.002$
$[0.9, 1]$	$0.084 \pm 0.003 \pm 0.001$	$1.310 \pm 0.010 \pm 0.023$	$0.764 \pm 0.010 \pm 0.003$	$0.083 \pm 0.004 \pm 0.002$

5.2 $B^0_{(s)} \to \mu^+\mu^-$ calibration

An alternative to the method presented in the previous section consists in performing the BDT calibration from simulated $B^0_{(s)} \to \mu^+\mu^-$ events. Unfortunately, the simulation does not reproduce the behaviour observed in data for various variables. The main discrepancies between data and simulation are observed in the transverse momentum, the pseudorapidity, the impact parameter significance of the vertex ($\chi^2_{\rm IP}$) associated with the $B^0_{(s)}$ candidate and the event occupancy. The mismodelling of these variables in the simulation affects the BDT distribution.

Thus, the $B^0_{(s)} \to \mu^+\mu^-$ simulation is weighted to compensate for the data and simulation differences. The weighting is done by first correcting the transverse momentum, pseudorapidity and $\chi^2_{\rm IP}$ of the $B^0_{(s)}$ candidate, jointly referred to as $B^0_{(s)}$ kinematics, using a multivariate classifier. This procedure is referred to as kinematic reweighting and is explained in Sec. 5.2.1. After that, the simulation is further corrected for the differences in the event occupancy using a two-dimensional correction obtained from $B^+ \to J/\psi K^+$ decays, as presented in Sec. 5.2.2.

5.2.1 $B^0_{(s)}$ kinematics reweighting

A machine learning technique known as a Gradient Boosted Reweighter (GBR) [61] is used to correct for the differences between data and simulation observed in the $B^0_{(s)}$ kinematics. The GBR distinguishes the discrepancies between data and simulation samples associated with the training variables and generates weights to align them.

The GBR is trained on simulation and data samples from a proxy channel for the $B^0_{(s)} \to \mu^+\mu^-$ candidates. The data samples are background-substracted using the sWeights from the sPlot technique [52]. Since the kinematic distributions of the B^0 and B^0_s hadrons might differ, two different control channels are exploited to treat $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ separately. These channels are the $B^+ \to J/\psi K^+$ decay for the $B^0 \to \mu^+\mu^-$ case and the $B^0_s \to J/\psi\,\phi$ mode for $B^0_s \to \mu^+\mu^-$ candidates. In both cases, the GBR training variables are the transverse momentum, $p_{\rm T}$, the pseudorapidity, η and the $\chi^2_{\rm IP}$ of the B candidates. These variables are taken as the input variables because of the significant disagreement between their sWeighted data and simulation distributions. The trained GBR is applied to the $B^0_{(s)} \to \mu^+\mu^-$ simulation samples to align them with the expected behaviour observed in data samples.

The distributions of the GBR training variables obtained for sWeighted data, simulation, and reweighted simulation are shown in Fig. 5.6 for $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\,\phi$ decays. Although the mismodelling of the kinematic variables in simulation is clearly visible, especially for the transverse momentum, the kinematic reweighting completely corrects for it, as can be seen from the perfect agreement between the data and reweighting simulation curves.

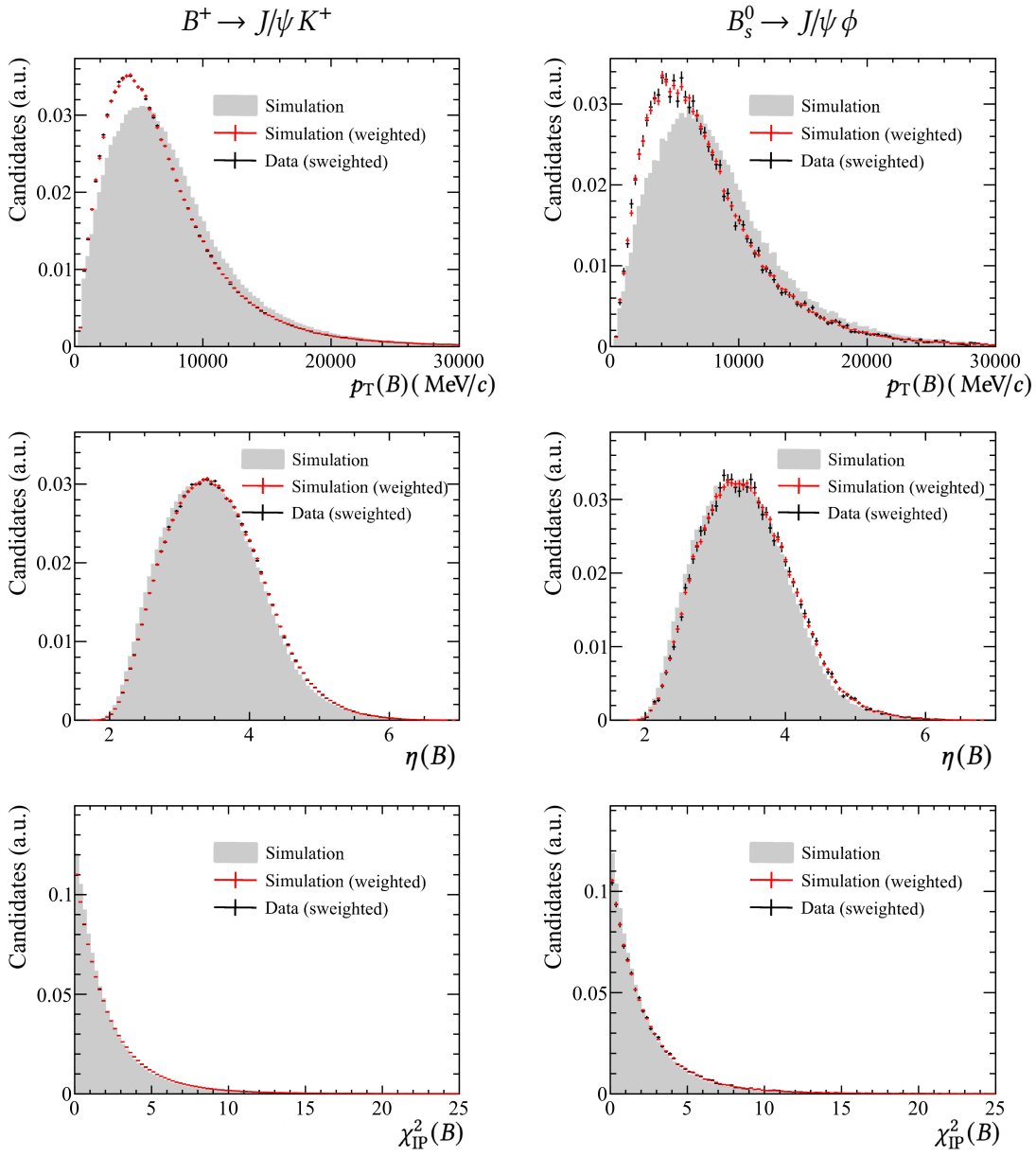


Figure 5.6: Distribution of the training variables used in the GBR for sWeighted data (black), simulation (grey), and kinematically reweighted simulation (red). The left column corresponds to $B^+ \to J/\psi K^+$ and the right to $B_s^0 \to J/\psi \phi$ decays.

By default, no error is associated with the weights determined from the GBR. To test the effect of the limited statistics of the training sample, a second GBR is trained on only half of the $B^+ \to J/\psi K^+$ ($B_s^0 \to J/\psi \phi$) samples. The simulated $B_s^0 \to \mu^+\mu^-$ events are reweighted using the new GBR and their BDT distribution is compared with the one originating from the entire training samples in Fig. 5.7. The differences between the two distributions are negligible and show the stability of the reweighting procedure. Therefore, no systematic

error is associated with the kinematic reweighting procedure.

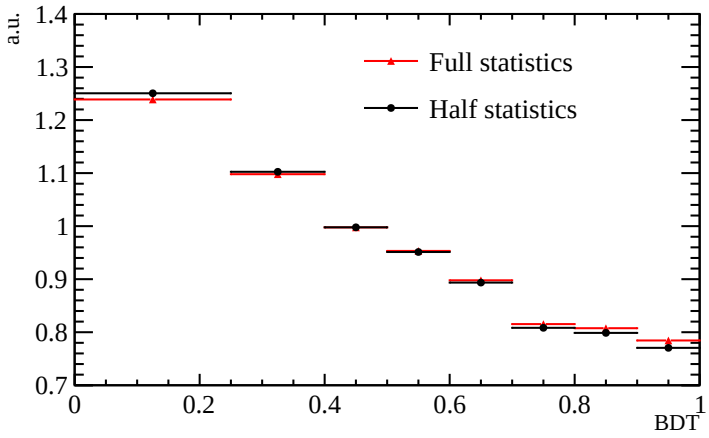


Figure 5.7: BDT distributions of the reweighted $B^0_s \to \mu^+\mu^-$ simulation using a GBR trained on the full (red) or half (black) of the training samples.

5.2.2 Occupancy correction

In addition to the $B^0_{(s)}$ kinematics, differences between simulation and data are also observed in the event occupancy. An appropriate variable to visualize this discrepancy is the *nTracks* variable, quantifying the number of tracks in the event. Fig. 5.8 shows its distribution for $B^+ \to J/\psi K^+$ sWeighted data, simulation and reweighted simulation. The difference between data and simulation is prominent and is not reduced by the kinematic reweighting, whose effect is found to be negligible for the event occupancy variables. This disagreement between data and simulation mainly results from a cutoff made for low momentum tracks in the simulation production to improve simulation times.

The mismodelling in simulation could be corrected by including the nTracks variable as a fourth training variable of the GBR from Sec. 5.2.1. Unfortunately, this is not a suitable solution because the GBR does not properly include the correlation between the nTracks and the BDT response. This correlation arises from the use of isolation variables as BDT training variables since they are directly linked to the event occupancy.

An alternative technique was developed to capture the correlation effect between the BDT and the event occupancy. A two-dimensional correction in bins of the nTracks and the BDT variables is used to correct the already kinematically reweighted simulation samples, further referred to as simply reweighted simulation samples. This correction is evaluated as the relative difference between the two-dimensional normalized event distributions from the sWeighted data and the kinematically reweighted simulation samples of a proxy channel. The data samples used are sWeighted to unfold the signal from the background contributions [52]. The occupancy correction is applied to the already kinematically reweighted

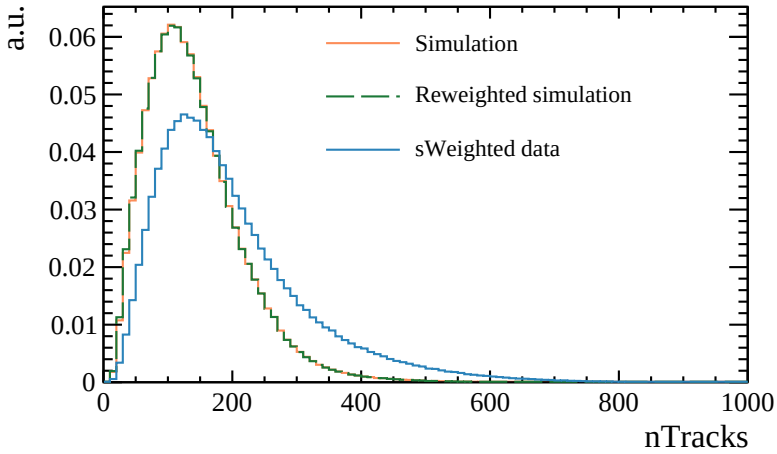


Figure 5.8: nTracks distribution of the $B_s^0 \to \mu^+\mu^-$ candidates from kinematically reweighted simulation (green) and $B^+ \to J/\psi K^+$ candidates from sWeighted data (blue) and kinematically reweighted simulation (red).

$B^0_{(s)} \to \mu^+\mu^-$ simulation by weighting each event with a factor given by the occupancy correction depending on the nTracks and BDT value of the event.

The evaluation of the occupancy correction is done using a proxy channel whose properties are similar to that of the $B^0_{(s)} \to \mu^+\mu^-$ candidates. Both normalization channels, $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$, fulfil this condition. Although the $B^0 \to K^+\pi^-$ decay is a two-body decay with a more similar topology to the $B^0_{(s)} \to \mu^+\mu^-$ process, the use of the $B^+ \to J/\psi K^+$ mode is preferred due to its significantly larger yield in the data sample as well as to avoid issues with different PID and trigger selections (see Chapter 4).

One of the main complications of using the $B^+ \to J/\psi K^+$ decays as a proxy is their distinct BDT response from that of the $B^0_{(s)} \to \mu^+\mu^-$ candidates. This difference is expected because the BDT distribution is highly dependent on the specific decay topology. While $B^+ \to J/\psi K^+$ are three-body decays, the $B^0_{(s)} \to \mu^+\mu^-$ processes have only two particles in the final state. In order to better mimic the $B^0_{(s)} \to \mu^+\mu^-$ BDT response, the BDT training variables for $B^+ \to J/\psi K^+$ decays are associated either with the B^+ candidate or with the final-state muons, always excluding the K^+ from the final state. An exception is applied to the χ^2_{IP} variable. Because the number of degrees of freedom of the J/ψ candidate is the same as for the B candidate in the $B^0_{(s)} \to \mu^+\mu^-$ decay, the χ^2_{IP} is evaluated on the J/ψ candidate. A detailed explanation of the BDT classifier and the training variables used for $B^0_{(s)} \to \mu^+\mu^-$ decays is given in Sec. 4.5.2 The normalized distributions for the BDT training variables are shown for $B^+ \to J/\psi K^+$ and $B^0_{(s)} \to \mu^+\mu^-$ reweighted simulation samples in Fig. 5.9. For completeness, the $B^+ \to J/\psi K^+$ sWeighted data and the $B^0_{(s)} \to \mu^+\mu^-$ unweighted simulation distributions are also displayed.

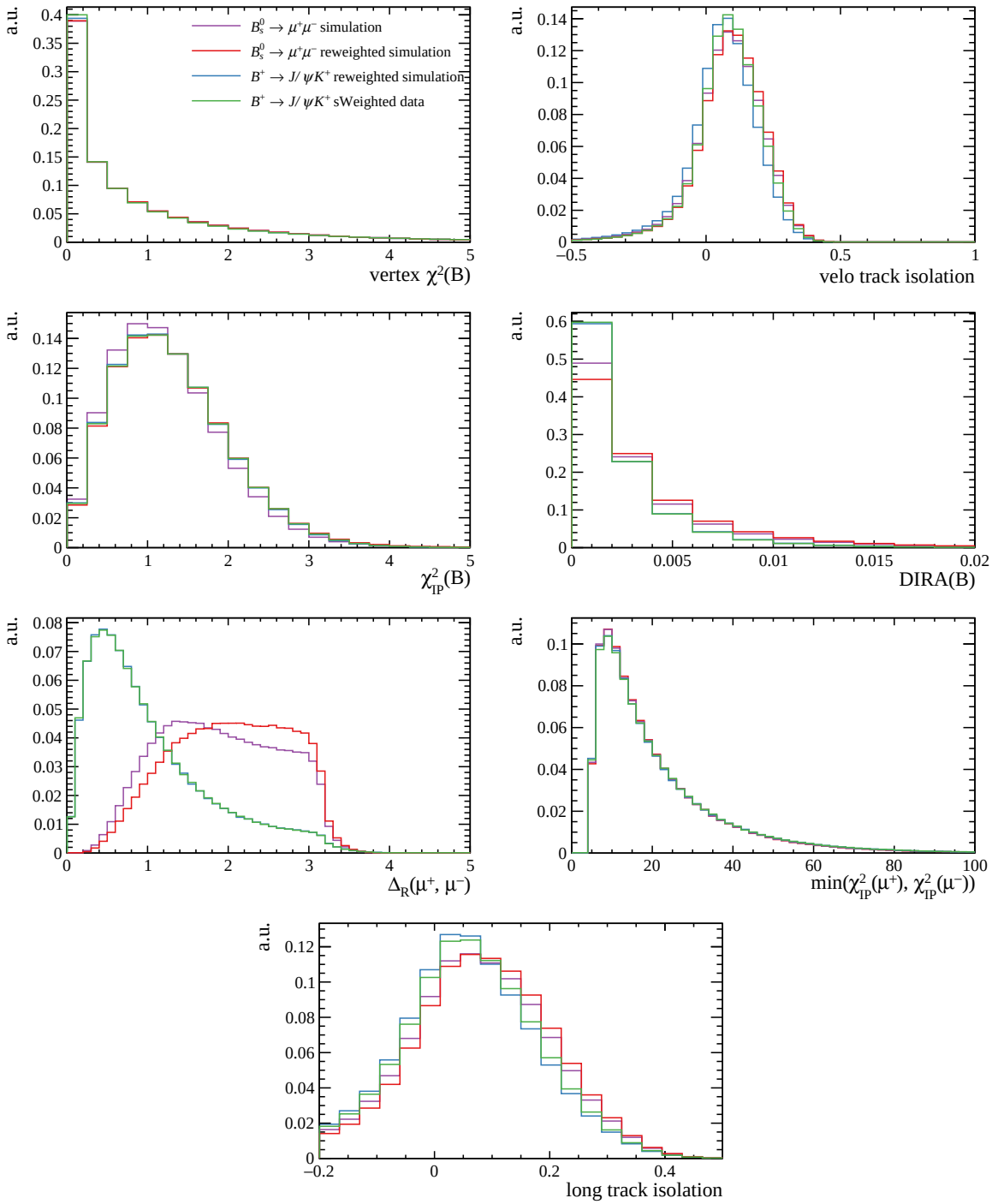


Figure 5.9: Distributions associated with the BDT training variables of the $B^0_s \to \mu^+\mu^-$ decays from kinematically reweighted simulation (green) and $B^+ \to J/\psi K^+$ decays from data (blue) and kinematically reweighted simulation (red).

As can be seen from Fig. 5.9, the distributions of the BDT training variables for reweighted $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ simulation samples are very similar, with the most significant discrepancies found in the isolation variables and mainly in the $\Delta_R(\mu^+, \mu^-)$ variable. The small discrepancy in the isolation variables originates from the two-body versus the three-body decay vertex for the $B^0_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ modes. Similarly, the large disagreement in $\Delta_R(\mu^+, \mu^-)$ comes from the different angular distribution of the muons coming from the J/ψ decay relative to the final-state muons from $B^0_s \to \mu^+\mu^-$ decays. As a consequence of these discrepancies in the distributions associated with $B^0_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays, the BDT distributions for $B^+ \to J/\psi K^+$ and $B^0_s \to \mu^+\mu^-$ decays are also expected to differ.

The resulting BDT response for $B^+ \to J/\psi K^+$ reweighted simulation is compared with the analogous from $B^0_s \to \mu^+\mu^-$ reweighted simulation in Fig. 5.10. Again, the BDT distributions for $B^+ \to J/\psi K^+$ data and $B^0_{(s)} \to \mu^+\mu^-$ simulation are also added for completeness. As anticipated, the differences between the BDT responses of the $B^+ \to J/\psi K^+$ and $B^0_s \to \mu^+\mu^-$ modes are large. Nevertheless, the relative difference between sWeighted data and simulation samples is expected to be the same for $B^+ \to J/\psi K^+$ and $B^0_{(s)} \to \mu^+\mu^-$ decays. Since the occupancy correction is a relative correction, the two-dimensional correction evaluated from $B^+ \to J/\psi K^+$ samples can be applied to the $B^0_{(s)} \to \mu^+\mu^-$ reweighted simulation samples.

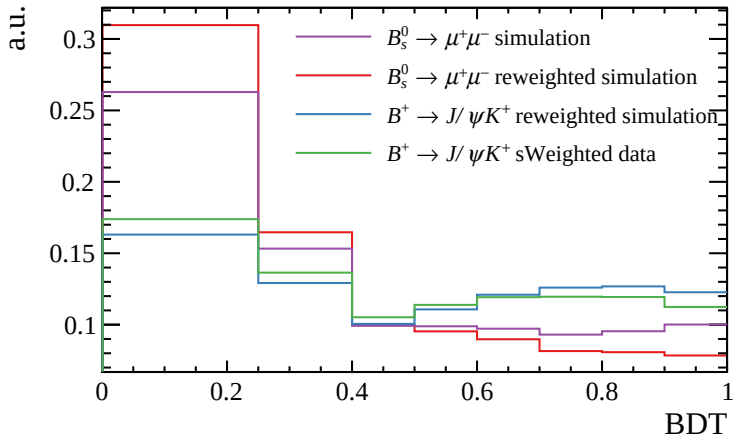


Figure 5.10: The BDT distributions of the $B^0_s \to \mu^+\mu^-$ decays from kinematically reweighted simulation (green) and $B^+ \to J/\psi K^+$ decays from data (blue) and kinematically reweighted simulation (red).

As previously mentioned, the occupancy correction is evaluated as the ratio between the normalized two-dimensional distributions of the $B^+ \to J/\psi K^+$ sWeighted data and reweighted simulation samples. Ideally, an occupancy correction should be determined for each data-taking year. However, the $B^+ \to J/\psi K^+$ event yields of the samples are significantly low at large BDT values, as can be seen in Fig. 5.10. In order to reduce the impact

of the limited statistics, only one occupancy correction is determined for Run 2 by merging all sWeighted data and reweighted simulation samples of the various Run 2 data-taking years. The resulting two-dimensional occupancy correction is shown in Fig. 5.11. This correction is later convolved with each data-taking $B^0_{(s)} \to \mu^+\mu^-$ reweighting simulation sample separately.

There are two primary sources of error in the occupancy correction evaluation. The effect of the binning scheme is studied by merging some bins and performing the BDT calibration again using $B^0_{(s)} \to \mu^+\mu^-$ simulation. The difference with respect to the use of the nominal occupancy correction is added as systematic uncertainty.

The $B^+ \to J/\psi K^+$ data samples used to determine the occupancy correction are sWeigthed to separate the signal from the background contributions [52]. The sWeights are determined from a maximum likelihood fit to the invariant mass of the final-state particles from the $B^+ \to J/\psi K^+$ candidates using the same fit strategy explained in Sec. 4.2. The $B^+ \to J/\psi K^+$ component is modelled using a Hypatia function. To test the effect of the specific shape, the BDT calibration is performed using a double-sided Crystal Ball function instead. The difference with respect to the nominal results is found negligible, and no further systematic uncertainty is added.

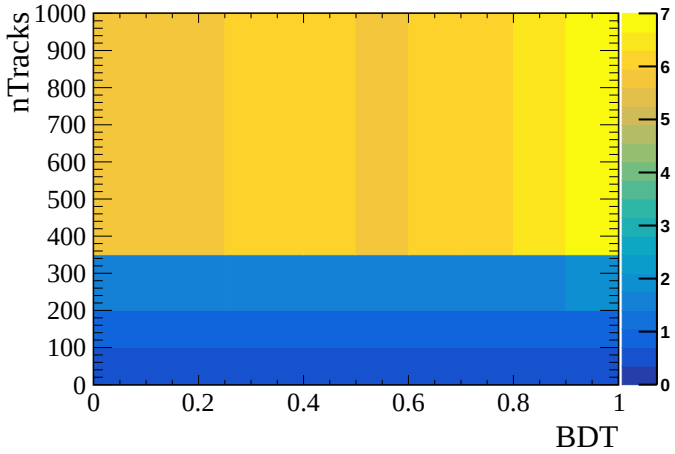


Figure 5.11: Occupancy correction determined from $B^+ \to J/\psi K^+$ sWeighted data and kinematically reweighted simulation samples.

5.2.3 Results

The fraction of events per BDT bin estimated using $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ simulation samples are presented in Table 5.5 for each data-taking year. As discussed in Sec. 5.2.1 and Sec. 5.2.2, the only non-negligible source of systematic uncertainty is the binning scheme used in the occupancy correction. The statistical uncertainty is propagated from the limited statistics of the $B^0_{(s)} \to \mu^+\mu^-$ simulation samples and the statistical error associated with the occupancy correction. The latter is found to be the main contribution, especially at large BDT values, due to the limited size of the $B^+ \to J/\psi K^+$ data sample used to evaluate the occupancy correction.

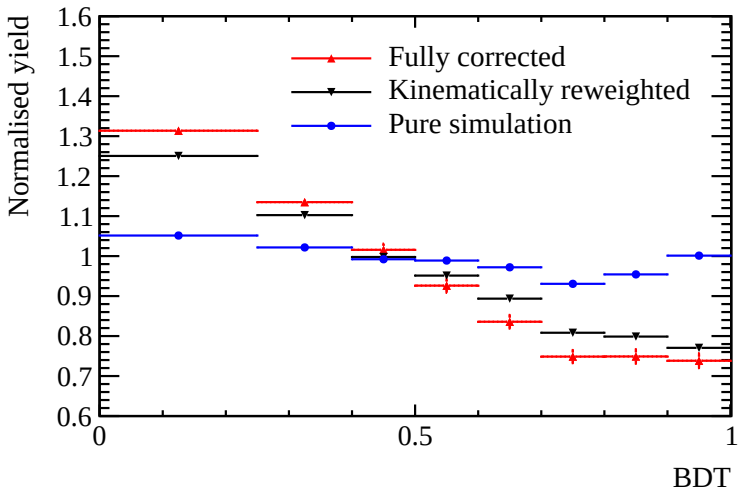


Figure 5.12: Check of the effect from the corrections added to the $B_s^0 \to \mu^+\mu^-$ simulation sample. The BDT distribution is shown for pure simulation (blue), kinematically reweighted simulation (black) and reweighted and occupancy-corrected simulation (red).

The effect of each correction in the BDT shape obtained from simulated $B^0_{(s)} \to \mu^+\mu^-$ events is illustrated in Fig. 5.12. The BDT distributions of the pure, kinematically reweighted and fully corrected $B^0_{(s)} \to \mu^+\mu^-$ simulation are shown. The flat BDT shape from pure simulation results from the flattening procedure applied to the BDT output explained in Sec. 4. The kinematic reweighting distorts the flatness significantly. The addition of the occupancy correction is found to have a significant impact and enlarges the difference with respect to the BDT response from pure simulated samples.

Table 5.5: Relative yields per BDT bin for $B^0_s \to \mu^+\mu^-$ (top) and $B^0 \to \mu^+\mu^-$ (bottom) as determined using $B^0_{(s)} \to \mu^+\mu^-$ simulation.

BDT bin	**2011**	**2012**	**2015**
$[0, 0.25)$	$0.2923 \pm 0.0018 \pm 0.0034$	$0.3082 \pm 0.0018 \pm 0.0068$	$0.3233 \pm 0.0018 \pm 0.0018$
$[0.25, 0.4)$	$0.1661 \pm 0.0015 \pm 0.0008$	$0.1643 \pm 0.0015 \pm 0.0005$	$0.1699 \pm 0.0016 \pm 0.0004$
$[0.4, 0.5)$	$0.1041 \pm 0.0013 \pm 0.0011$	$0.1043 \pm 0.0014 \pm 0.0012$	$0.1021 \pm 0.0014 \pm 0.0009$
$[0.5, 0.6)$	$0.0969 \pm 0.0015 \pm 0.0009$	$0.0960 \pm 0.0016 \pm 0.0014$	$0.0926 \pm 0.0016 \pm 0.0006$
$[0.6, 0.7)$	$0.0901 \pm 0.0015 \pm 0.0007$	$0.0895 \pm 0.0017 \pm 0.0012$	$0.0852 \pm 0.0017 \pm 0.0003$
$[0.7, 0.8)$	$0.0855 \pm 0.0016 \pm 0.0001$	$0.0820 \pm 0.0017 \pm 0.0009$	$0.0761 \pm 0.0017 \pm 0.0001$
$[0.8, 0.9)$	$0.0834 \pm 0.0016 \pm 0.0000$	$0.0813 \pm 0.0019 \pm 0.0006$	$0.0758 \pm 0.0018 \pm 0.0004$
$[0.9, 1]$	$0.0817 \pm 0.0016 \pm 0.0001$	$0.0746 \pm 0.0018 \pm 0.0009$	$0.0750 \pm 0.0019 \pm 0.0001$
BDT bin	**2016**	**2017**	**2018**
$[0, 0.25)$	$0.3284 \pm 0.0017 \pm 0.0015$	$0.3149 \pm 0.0016 \pm 0.0018$	$0.3136 \pm 0.0018 \pm 0.0011$
$[0.25, 0.4)$	$0.1703 \pm 0.0015 \pm 0.0004$	$0.1676 \pm 0.0015 \pm 0.0011$	$0.1684 \pm 0.0016 \pm 0.0010$
$[0.4, 0.5)$	$0.1014 \pm 0.0013 \pm 0.0008$	$0.1025 \pm 0.0013 \pm 0.0002$	$0.1034 \pm 0.0014 \pm 0.0005$
$[0.5, 0.6)$	$0.0926 \pm 0.0015 \pm 0.0010$	$0.0945 \pm 0.0016 \pm 0.0005$	$0.0928 \pm 0.0016 \pm 0.0005$
$[0.6, 0.7)$	$0.0836 \pm 0.0016 \pm 0.0006$	$0.0861 \pm 0.0016 \pm 0.0005$	$0.0865 \pm 0.0017 \pm 0.0002$
$[0.7, 0.8)$	$0.0749 \pm 0.0016 \pm 0.0004$	$0.0773 \pm 0.0017 \pm 0.0000$	$0.0778 \pm 0.0017 \pm 0.0001$
$[0.8, 0.9)$	$0.0749 \pm 0.0018 \pm 0.0003$	$0.0784 \pm 0.0019 \pm 0.0001$	$0.0789 \pm 0.0019 \pm 0.0007$
$[0.9, 1]$	$0.0738 \pm 0.0018 \pm 0.0005$	$0.0787 \pm 0.0019 \pm 0.0002$	$0.0786 \pm 0.0020 \pm 0.0003$

BDT bin	**2011**	**2012**	**2015**
$[0, 0.25)$	$0.2926 \pm 0.0019 \pm 0.0031$	$0.2986 \pm 0.0023 \pm 0.0069$	$0.3117 \pm 0.0019 \pm 0.0016$
$[0.25, 0.4)$	$0.1626 \pm 0.0017 \pm 0.0005$	$0.1621 \pm 0.0019 \pm 0.0012$	$0.1688 \pm 0.0018 \pm 0.0003$
$[0.4, 0.5)$	$0.1045 \pm 0.0014 \pm 0.0010$	$0.1037 \pm 0.0017 \pm 0.0013$	$0.1026 \pm 0.0016 \pm 0.0009$
$[0.5, 0.6)$	$0.0959 \pm 0.0016 \pm 0.0008$	$0.0973 \pm 0.0020 \pm 0.0019$	$0.0950 \pm 0.0018 \pm 0.0005$
$[0.6, 0.7)$	$0.0925 \pm 0.0017 \pm 0.0001$	$0.0917 \pm 0.0021 \pm 0.0007$	$0.0866 \pm 0.0019 \pm 0.0002$
$[0.7, 0.8)$	$0.0853 \pm 0.0018 \pm 0.0003$	$0.0826 \pm 0.0021 \pm 0.0010$	$0.0782 \pm 0.0019 \pm 0.0003$
$[0.8, 0.9)$	$0.0847 \pm 0.0019 \pm 0.0004$	$0.0839 \pm 0.0022 \pm 0.0001$	$0.0783 \pm 0.0021 \pm 0.0001$
$[0.9, 1]$	$0.0821 \pm 0.0018 \pm 0.0000$	$0.0801 \pm 0.0023 \pm 0.0006$	$0.0787 \pm 0.0022 \pm 0.0001$
BDT bin	**2016**	**2017**	**2018**
$[0, 0.25)$	$0.3191 \pm 0.0018 \pm 0.0025$	$0.3109 \pm 0.0017 \pm 0.0020$	$0.3114 \pm 0.0019 \pm 0.0023$
$[0.25, 0.4)$	$0.1678 \pm 0.0016 \pm 0.0009$	$0.1660 \pm 0.0016 \pm 0.0005$	$0.1666 \pm 0.0017 \pm 0.0011$
$[0.4, 0.5)$	$0.1022 \pm 0.0015 \pm 0.0003$	$0.1022 \pm 0.0015 \pm 0.0006$	$0.1019 \pm 0.0016 \pm 0.0011$
$[0.5, 0.6)$	$0.0937 \pm 0.0018 \pm 0.0008$	$0.0954 \pm 0.0018 \pm 0.0008$	$0.0951 \pm 0.0018 \pm 0.0007$
$[0.6, 0.7)$	$0.0860 \pm 0.0018 \pm 0.0002$	$0.0868 \pm 0.0019 \pm 0.0006$	$0.0872 \pm 0.0019 \pm 0.0000$
$[0.7, 0.8)$	$0.0766 \pm 0.0018 \pm 0.0003$	$0.0785 \pm 0.0019 \pm 0.0001$	$0.0788 \pm 0.0019 \pm 0.0002$
$[0.8, 0.9)$	$0.0781 \pm 0.0021 \pm 0.0000$	$0.0808 \pm 0.0021 \pm 0.0003$	$0.0794 \pm 0.0021 \pm 0.0001$
$[0.9, 1]$	$0.0766 \pm 0.0021 \pm 0.0001$	$0.0793 \pm 0.0021 \pm 0.0003$	$0.0797 \pm 0.0022 \pm 0.0004$

5.3 Comparison of the two methods

The results obtained from the two different methods used for the BDT calibration should lead to the same results within uncertainties. A direct comparison between the two methods is used to test if the various factors affecting the BDT response in each calibration method are included and properly evaluated. An intermediate alternative is included for this comparison to be fair and to avoid including biases from the use of two different decay modes, namely $B^0_{(s)} \to \mu^+\mu^-$ versus $B^0_{(s)} \to h^+h'^-$, in the two approaches presented. This new alternative consists in performing the BDT calibration using simulated $B^0_{(s)} \to h^+h'^-$ events and correcting them as done for the $B^0_{(s)} \to \mu^+\mu^-$ simulation samples. Therefore, three different methods to perform the BDT calibration are compared:

1. the $B^0_{(s)} \to h^+h'^-$ calibration method, described in Sec. 5.1, using $B^0_{(s)} \to h^+h'^-$ data samples corrected for the hadronic PID and trigger requirements;
2. the $B^0_{(s)} \to \mu^+\mu^-$ calibration method, presented in Sec. 5.2, using simulated $B^0_{(s)} \to \mu^+\mu^-$ events corrected for the data and simulation differences;
3. the new alternative, exploiting simulated $B^0_{(s)} \to h^+h'^-$ events corrected for the data and simulation difference following the same method as for $B^0_{(s)} \to \mu^+\mu^-$ simulation samples.

The comparison between the $B^0_{(s)} \to h^+h'^-$ calibration method and the new alternative serves as a validation of the two BDT calibration methods presented. The agreement between the two shapes has two separate implications for each method. In the $B^0_{(s)} \to \mu^+\mu^-$ calibration procedure, this agreement implies that all the effects from the simulation and data differences are correctly compensated. In the case of the $B^0_{(s)} \to h^+h'^-$ calibration, it indicates that the hadronic trigger and PID effects per BDT bin are under control and well-reproduced by the corresponding correction factors.

The comparison between the methods are shown in Fig. 5.13 for the $B^0 \to K^+\pi^-$ channel and Fig. 5.14 for the $B^0_s \to K^+K^-$ mode. The distributions using simulated $B^0_{(s)} \to \mu^+\mu^-$ events are also included for completeness. The results from $B^0 \to \mu^+\mu^-$ simulation samples are included in the $B^0 \to K^+\pi^-$ comparison, while $B^0_s \to \mu^+\mu^-$ simulation is added in the $B^0_s \to K^+K^-$ one. This separation is done because of the different decay width asymmetries for the B^0 and B^0_s mesons and their connection to the effective lifetimes of the $B^0_{(s)} \to \mu^+\mu^-$ decays, which are correlated to the BDT response. The statistical and systematic uncertainties for each shape are represented in solid and dashed lines, respectively. The systematic uncertainties are significantly small compared to the statistical errors, especially for the values obtained using simulation samples, meaning the systematic effects are under control.

The distributions from the exclusive $B^0_{(s)} \to h^+h'^-$ data and corrected simulation samples (methods 1 and 3 from the list) are in good agreement within uncertainties. This agreement

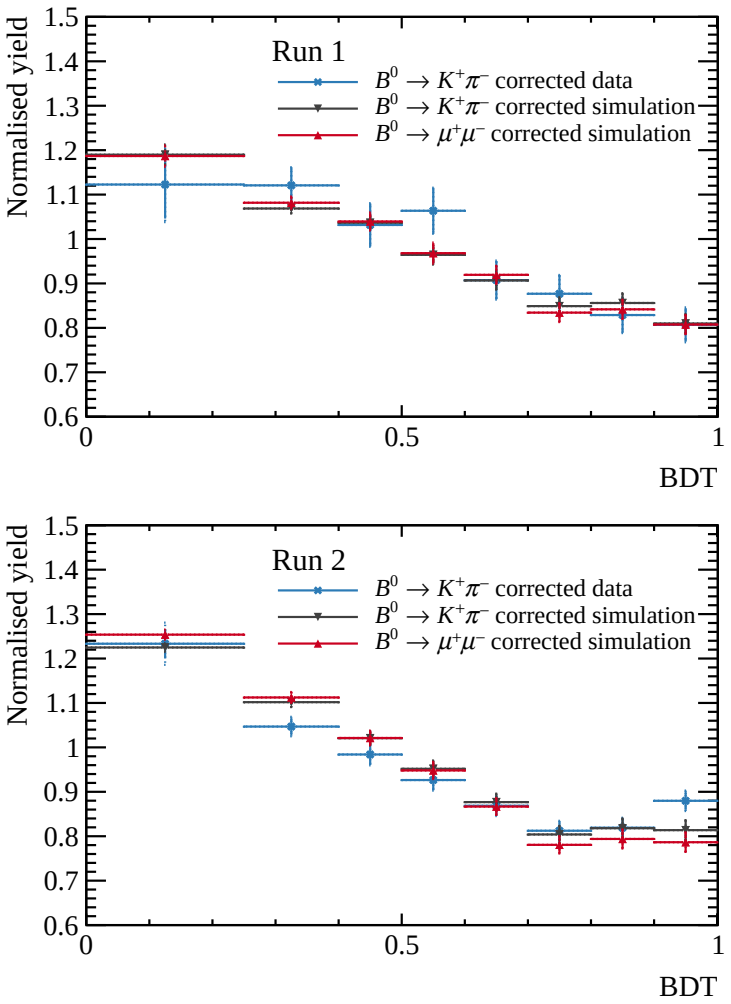


Figure 5.13: The BDT distributions for Run 1 (left) and Run 2 (right) obtained from the $B^0 \to K^+\pi^-$ corrected data (blue), as presented in Sec. 5.1; $B^0 \to \mu^+\mu^-$ corrected simulation (red), as explained in Sec. 5.2; and the alternative of using $B^0 \to K^+\pi^-$ simulation corrected as done for $B^0 \to \mu^+\mu^-$ simulation.

between two shapes derived following entirely different methods and samples shows the large control over the BDT response and the two different approaches presented. The overlap between $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ simulation distributions is expected because their selection procedure is identical since no trigger or PID criteria are applied in the simulation samples.

The results obtained from either $B^0_{(s)} \to h^+h'^-$ data or $B^0_{(s)} \to \mu^+\mu^-$ simulation samples can be used for the final BDT calibration. Nevertheless, the statistics of the $B^0_{(s)} \to h^+h'^-$ data sample are expected to be quite limited since $B^0_{(s)} \to h^+h'^-$ are rare decays. Furthermore, the TIS trigger requirement decreases further the expected yield for $B^0_{(s)} \to h^+h'^-$ candidates in the data samples. Since the total uncertainty is substantially lower, the method using

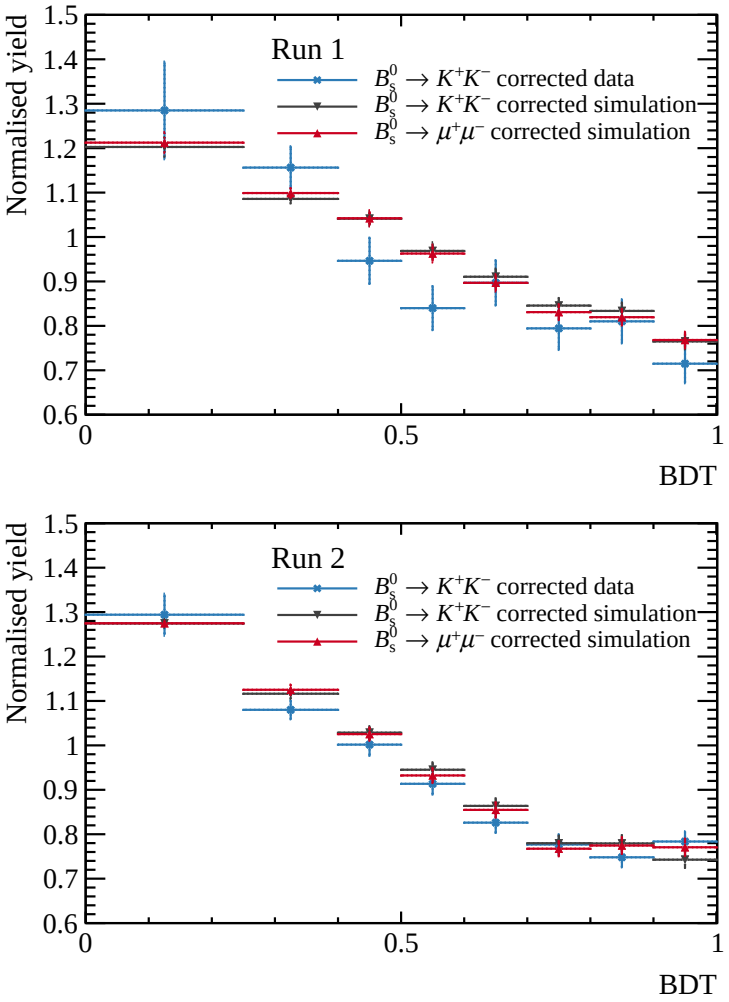


Figure 5.14: The BDT distributions for Run 1 (left) and Run 2(right) obtained from the $B_s^0 \to K^+K^-$ corrected data (blue), as presented in Sec. 5.1; $B_s^0 \to \mu^+\mu^-$ corrected simulation (red), as explained in Sec. 5.2; and the alternative of using $B_s^0 \to K^+K^-$ simulation corrected as done for $B_s^0 \to \mu^+\mu^-$.

simulated $B^0_{(s)} \to \mu^+\mu^-$ events is the approach chosen to perform the BDT calibration and determine the expected fraction of signal events in each BDT bin.

5.4 Muonic trigger and PID corrections per BDT bin

The $B^0_{(s)} \to \mu^+\mu^-$ candidates in data are required to pass a muonic trigger and PID selection criteria given in Chapter 4. Although the muonic trigger and PID efficiencies are already included in the normalization, their values might vary among the different BDT bins, altering the number of events expected in each BDT bin. Hence, the BDT calibration must take into account these relative shifts per BDT bin, such that the expected relative yield of the i_{BDT}

BDT bin for the signal candidates should be evaluated as:

$$f_{\mathrm{i_{BDT}},\mathrm{BDTCalib}} = f_{\mathrm{i_{BDT}}} \cdot \omega^{\mu\mu}_{\mathrm{i_{BDT}},\mathrm{trg}} \cdot \omega^{\mu\mu}_{\mathrm{i_{BDT}},\mathrm{PID}} \tag{5.7}$$

where $f_{\mathrm{i_{BDT}}}$ is the fraction of events for that BDT bin determined from $B^0_{(s)} \to \mu^+\mu^-$ corrected simulation, and $\omega^{\mu\mu}_{\mathrm{i_{BDT}},\mathrm{trg}}$ and $\omega^{\mu\mu}_{\mathrm{i_{BDT}},\mathrm{PID}}$ the correction factors for the relative shifts in the muonic trigger and PID selections, respectively. Both corrections are determined for each BDT bin as the ratio between the BDT-binned and overall efficiencies.

5.4.1 Muonic trigger correction

The muonic trigger efficiencies are determined from data using $B^+ \to J/\psi K^+$ decays as a proxy with the TISTOS method as a function of several muon kinematic variables. The resulting look-up table is referred to as the TISTOS map. The overall trigger efficiency of the $B^0_{(s)} \to \mu^+\mu^-$ decays is evaluated from the convolution of the TISTOS map with the kinematic spectra of the $B^0_{(s)} \to \mu^+\mu^-$ reweighted simulation, as detailed in Sec. 4.2. For the BDT-binned efficiencies, the TISTOS map is convolved with the corresponding BDT subsamples.

As shown in Sec 5.1.2, the L0&HLT1 trigger selection used to select $B^0_{(s)} \to h^+h'^-$ candidates and the BDT are correlated due to their relation with the event occupancy. The inclusive strategy followed to select $B^0_{(s)} \to \mu^+\mu^-$ events is expected to be less dependent on the underlying event as it is mainly based on the final-state muons (see Sec. 4.2). Therefore, the correlation effect is expected to be small.

Although the correlation between the trigger selection for $B^0_{(s)} \to \mu^+\mu^-$ decays and the BDT is expected to be minor, its effect on the relative $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiencies is analyzed using simulated $B^0_{(s)} \to \mu^+\mu^-$ signal samples. An MC true map[3] is obtained from simulated $B^0_{(s)} \to \mu^+\mu^-$ events and convolved with the same $B^0_{(s)} \to \mu^+\mu^-$ simulation samples to determine the relative trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ decays. As for the $B^0_{(s)} \to h^+h'^-$ case, the relative efficiencies are evaluated as the ratio of the BDT-binned efficiencies, $\epsilon^{\mu\mu}_{\mathrm{i_{BDT}}}$ and the overall efficiencies obtained, $\epsilon^{\mu\mu}$, which are given as:

$$\epsilon^{\mu\mu}_{(\mathrm{i_{BDT}})} = \frac{\sum_{\mathrm{i_{Kin}}} \epsilon^{\mathrm{MCTrueMap}}{}_{\mathrm{i_{Kin}}} \cdot N^{B^0_{(s)}\to\mu^+\mu^-}_{(\mathrm{i_{BDT}})\mathrm{i_{Kin}}}}{N^{B^0_{(s)}\to\mu^+\mu^-}_{(\mathrm{i_{BDT}})}} \tag{5.8}$$

The resulting relative trigger efficiencies are compared to the analogous values evaluated from the simulated $B^0_{(s)} \to \mu^+\mu^-$ signal samples without using the maps as the ratio of the BDT-binned and overall true trigger efficiencies. The relative trigger efficiency distribu-

[3] Identical to the TISTOS map with the only difference being the efficiencies in each kinematic bin are evaluated as the true efficiency.

tions for each of the two methods are compared in Fig. 5.15. Although small, the disagreement between distributions is due to the correlation between the trigger $B^0_{(s)} \to \mu^+\mu^-$ selection and the BDT response. Because the correlation effect is significantly smaller than the one observed for the L0&HLT1 TIS selection characteristic of $B^0_{(s)} \to h^+h'^-$ decays, the two-dimensional BDT-dependent correction is not applied here. Instead, the nominal relative trigger efficiencies at each BDT bin are multiplied by a factor taken as the ratio between the relative trigger efficiencies obtained directly from simulation and the values evaluated from the MC true maps from $B^0_{(s)} \to \mu^+\mu^-$ simulation.

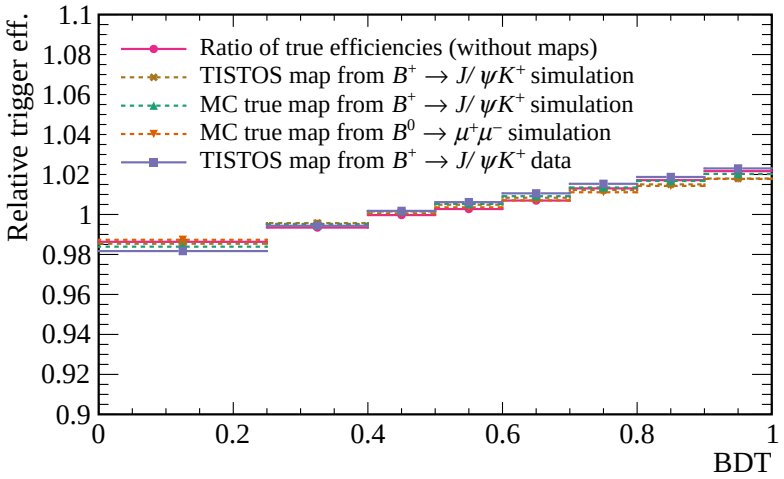


Figure 5.15: The different muonic trigger corrections per BDT bin for the $B^0_s \to \mu^+\mu^-$ decay obtained from the MC true map derived from $B^+ \to J/\psi K^+$ reweighted simulation (green), $B^0_s \to \mu^+\mu^-$ reweighted simulation (orange), the TISTOS map from $B^+ \to J/\psi K^+$ reweighted simulation (brown), $B^+ \to J/\psi K^+$ data (purple) and without the use of the maps (pink).

The statistical and systematic errors associated with the overall trigger efficiencies are already included in the normalization. Nevertheless, the systematic effects associated with the precision of the TISTOS method and the usage of $B^+ \to J/\psi K^+$ decays as a proxy of the $B^0_{(s)} \to \mu^+\mu^-$ are not uniformly distributed through the BDT bins. As a consequence, their systematic effect on the relative trigger efficiencies is analyzed.

The effect of using $B^+ \to J/\psi K^+$ decays as a proxy for the $B^0_{(s)} \to \mu^+\mu^-$ channel is studied on $B^+ \to J/\psi K^+$ and $B^0_{(s)} \to \mu^+\mu^-$ simulation samples. An MC true map is obtained from each sample and convolved with the simulated $B^0_{(s)} \to \mu^+\mu^-$ events to evaluate the relative trigger efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ decays. The results obtained from the $B^+ \to J/\psi K^+$ and $B^0_{(s)} \to \mu^+\mu^-$ MC true maps are compared in Fig. 5.15. The difference between the two values is included as systematic uncertainty in the relative $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiencies values.

The precision of the TISTOS method is examined using MC true and TISTOS maps obtained from $B^+ \to J/\psi K^+$ simulation samples. The relative $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiencies

are obtained using each map and result in the distributions shown in Fig. 5.15. The difference between the two curves is assigned to the relative $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiencies as systematic uncertainty.

The BDT distribution of the central values of the nominal trigger correction is also shown in Fig. 5.15 for completeness. Although not large, the discrepancy relative to the simulation values is not negligible. The trigger corrections for each BDT bin are given in Table 5.6 for each data-taking year. As previously mentioned, only systematic uncertainties are associated with it since the statistical errors are already included in the normalization procedure.

Table 5.6: Relative trigger correction factors per BDT bin for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays.

	$B^0_s \to \mu^+\mu^-$		$B^0 \to \mu^+\mu^-$	
BDT bin	Run 1	Run 2	Run 1	Run 2
$[0, 0.25)$	0.9820 ± 0.0015	0.9787 ± 0.0024	0.9816 ± 0.0016	0.9761 ± 0.0026
$[0.25, 0.4)$	0.9900 ± 0.0008	0.9928 ± 0.0005	0.9870 ± 0.0007	0.9925 ± 0.0005
$[0.4, 0.5)$	0.9972 ± 0.0005	1.0014 ± 0.0006	0.9978 ± 0.0006	1.0011 ± 0.0006
$[0.5, 0.6)$	1.0044 ± 0.0005	1.0059 ± 0.0010	1.0035 ± 0.0007	1.0077 ± 0.0011
$[0.6, 0.7)$	1.0091 ± 0.0007	1.0111 ± 0.0014	1.0087 ± 0.0008	1.0107 ± 0.0014
$[0.7, 0.8)$	1.0146 ± 0.0011	1.0161 ± 0.0017	1.0168 ± 0.0011	1.0182 ± 0.0018
$[0.8, 0.9)$	1.0207 ± 0.0016	1.0219 ± 0.0019	1.0199 ± 0.0016	1.0229 ± 0.0020
$[0.9, 1]$	1.0263 ± 0.0023	1.0277 ± 0.0021	1.0270 ± 0.0023	1.0293 ± 0.0022

5.4.2 Muonic PID correction

The PID corrections are also evaluated independently per BDT bin as the ratio between the BDT-binned and overall PID efficiencies. To evaluate them, $J/\psi \to \mu^+\mu^-$ data samples where the $J/\psi \to \mu^+\mu^-$ originates from b decays are used. The PID efficiencies are determined from this proxy channel as a function of the total and transverse momentum of the muon candidates and the nTracks variables. The resulting histogram is convolved with the $B^0_{(s)} \to \mu^+\mu^-$ kinematic spectrum obtained from reweighted simulation to evaluate the overall efficiencies (see Sec. 4.3). For the BDT-binned efficiencies, the PID histogram is convolved with the corresponding BDT subsample.

The statistical and systematic errors associated with the overall PID efficiency are already included in the normalization. No systematic uncertainty is associated with the PID correction since no correlation is observed between the PID and the BDT. The PID correction factors of each BDT bin for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are listed in Table 5.7 for Run 1 and Run 2.

Table 5.7: Relative PID correction factor per BDT bin for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$.

	$B_s^0 \to \mu^+\mu^-$		$B^0 \to \mu^+\mu^-$	
BDT bin	Run 1	Run 2	Run 1	Run 2
$[0, 0.25)$	0.958	0.964	0.954	0.962
$[0.25, 0.4)$	0.970	0.988	0.968	0.980
$[0.4, 0.5)$	0.986	1.002	0.987	0.996
$[0.5, 0.6)$	1.000	1.015	1.000	1.005
$[0.6, 0.7)$	1.014	1.022	1.016	1.014
$[0.7, 0.8)$	1.032	1.031	1.034	1.026
$[0.8, 0.9)$	1.055	1.043	1.055	1.043
$[0.9, 1]$	1.077	1.057	1.080	1.059

5.4.3 Effective lifetime correction

A further correction is needed in the $B_s^0 \to \mu^+\mu^-$ case to account for the effective lifetime dependence of the BDT shape. One of the inputs needed to generate simulation samples is the effective lifetime of the decay. In the Standard Model, the $B_{(s)}^0 \to \mu^+\mu^-$ effective lifetime can take any value between the lifetimes of the light and heavy mass eigenstates. The relation between the $B_{(s)}^0$ lifetime and the effective lifetime depends on the parameter $A_{\Delta\Gamma_s}^{\mu\mu}$ and the decay width asymmetry between heavy and light mass eigenstates, y_s, as shown in Chapter 1.

The simulation is generated by setting the effective lifetime to be the mean $B_{(s)}^0$ lifetime, meaning $y_s = 0$. Although this assumption is in agreement with the experimental observations for B^0, the corresponding B_s^0 value has been measured to be $y_s = 0.068 \pm 0.004$ [1]. As a result, the role of the $A_{\Delta\Gamma_s}^{\mu\mu}$ becomes significant in the effective lifetime evaluation. The value of $A_{\Delta\Gamma_s}^{\mu\mu}$ is still unknown, and it is predicted to be +1 in the Standard Moded.

The longer-lived $B_s^0 \to \mu^+\mu^-$ candidates will score higher in the BDT output than shorter-lived ones since their vertices are better reconstructed as they are more differentiated from the primary vertex. To assess the effect of this correlation on the BDT calibration, three different scenarios covering the entire effective lifetime range are considered: $A_{\Delta\Gamma_s}^{\mu\mu}$ = -1, 0 and 1. For each scenario, a correction factor per BDT is determined as

$$k_{\mathrm{iBDT}} = \frac{1}{N_{\mathrm{iBDT}}} \sum_{i=1}^{N} \omega_i \tag{5.9}$$

where N_{iBDT} is the number of events per BDT bin and ω_i is the weight associated with each event given by

$$\omega_i = \frac{\tau_{\mathrm{gen}}}{\tau_{\mu^+\mu^-}} e^{-t_i\left(1/\tau_{\mu^+\mu^-} - 1/\tau_{\mathrm{gen}}\right)} \tag{5.10}$$

where τ_{gen} is the lifetime used to generate the simulation, t_i the reconstructed decay time of the B_s^0 candidate and $\tau_{\mu^+\mu^-}$ the effective lifetime calculated using Eq. 1.35 and the specific value for $A_{\Delta\Gamma_s}^{\mu\mu}$.

The correction factors per BDT bin for the three different scenarios are listed for Run 1 and Run 2 in Table 5.8. The CP-odd hypothesis, $A_{\Delta\Gamma_s}^{\mu\mu} = +1$, in agreement with the SM, is assumed in the BDT calibration procedure. The dimuon invariant mass fit is repeated for the two remaining hypotheses as a check of the effective lifetime variation effect.

Table 5.8: Lifetime correction per BDT bin for Run 1 and Run 2 $B_s^0 \to \mu^+\mu^-$ simulation.

Run 1			
BDT bin	$k(A_{\Delta\Gamma_s}^{\mu\mu} = -1)$	$k(A_{\Delta\Gamma_s}^{\mu\mu} = 0)$	$k(A_{\Delta\Gamma_s}^{\mu\mu} = +1)$
$[0, 0.25)$	1.004	0.999	0.992
$[0.25, 0.4)$	0.989	1.002	1.012
$[0.4, 0.5)$	0.981	1.004	1.022
$[0.5, 0.6)$	0.977	1.005	1.027
$[0.6, 0.7)$	0.971	1.007	1.035
$[0.7, 0.8)$	0.964	1.008	1.044
$[0.8, 0.9)$	0.948	1.012	1.065
$[0.9, 1.0]$	0.909	1.021	1.119
Run 2			
BDT bin	$k(A_{\Delta\Gamma_s}^{\mu\mu} = -1)$	$k(A_{\Delta\Gamma_s}^{\mu\mu} = 0)$	$k(A_{\Delta\Gamma_s}^{\mu\mu} = +1)$
$[0, 0.25)$	1.00091	0.99963	0.99634
$[0.25, 0.4)$	0.98556	1.00226	1.01398
$[0.4, 0.5)$	0.98041	1.00318	1.02037
$[0.5, 0.6)$	0.97480	1.00417	1.02710
$[0.6, 0.7)$	0.96822	1.00531	1.03482
$[0.7, 0.8)$	0.96086	1.00658	1.04332
$[0.8, 0.9)$	0.94218	1.00989	1.06604
$[0.9, 1.0]$	0.89956	1.01773	1.12151

5.5 The final BDT calibration shape

The BDT calibration for $B_{(s)}^0 \to \mu^+\mu^-$ decays was performed following several consecutive steps. First, the $B_{(s)}^0 \to \mu^+\mu^-$ simulation is corrected for the simulation and data differences observed in the $B_{(s)}^0$ kinematics and event occupancy variables, as described in Sec. 5.2. Then, the relative muonic PID and trigger corrections per BDT bin are applied to the BDT response obtained from $B_{(s)}^0 \to \mu^+\mu^-$ corrected simulation samples. These corrections are evaluated from data samples as explained in Sec. 5.4. In the case of the $B_s^0 \to \mu^+\mu^-$, a further correction

is applied to account for the correlation between the effective lifetime and the BDT variable in the BDT response. This correction is determined for three different scenarios as detailed in Sec. 5.4. The CP-odd hypothesis, corresponding to $A_{\Delta\Gamma_s}^{\mu\mu} = +1$, is assumed for the BDT calibration.

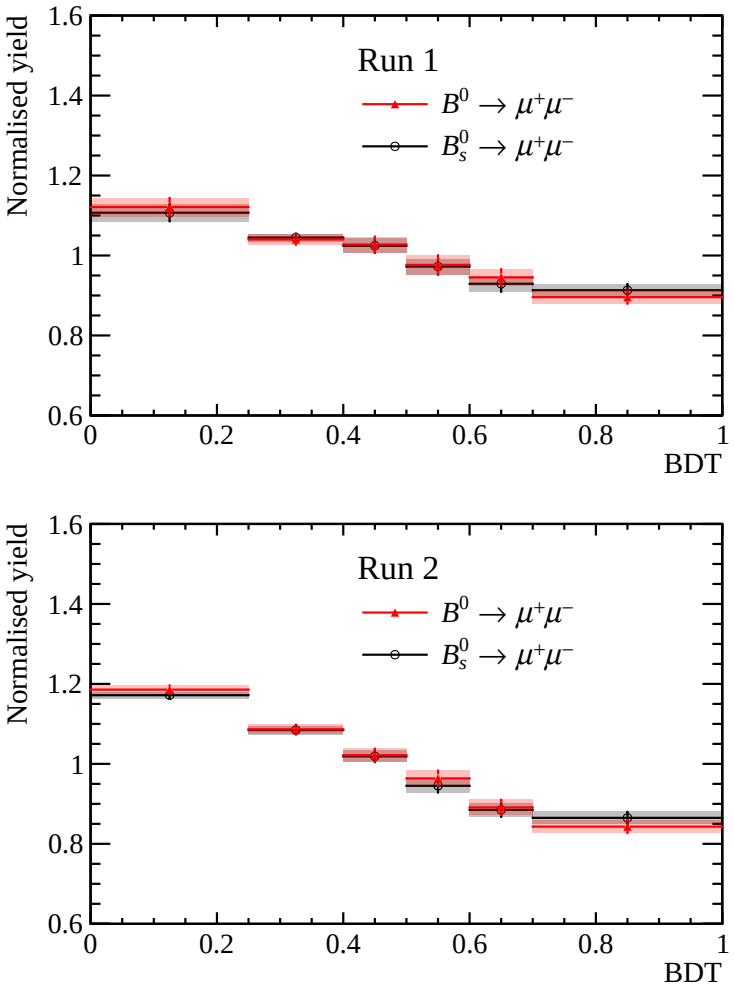


Figure 5.16: Calibrated BDT distribution of the $B_s^0 \to \mu^+\mu^-$ (black) and $B^0 \to \mu^+\mu^-$ (red) decays in Run 1 (up) and Run 2 (bottom). Both shapes are obtained using $B_{(s)}^0 \to \mu^+\mu^-$ corrected simulation samples, as detailed in Sec. 5.2, and include the muonic PID and trigger factors, explained in Sec. 5.4. The $B_s^0 \to \mu^+\mu^-$ shape is further corrected by the effective lifetime dependence correction, as described in Sec. 5.4.

The BDT calibration has been studied using a BDT binning scheme consisting of eight BDT bins to examine its performance in great detail, especially at high BDT values where the majority of the $B_{(s)}^0 \to \mu^+\mu^-$ events are expected. The procedure presented in this Chapter can be applied to any binning configuration and, in particular, to the six BDT bins binning configuration characteristic of the dimuon invariant mass fit. Therefore, the same approach

as described in this chapter is followed using a six BDT bins binning scheme to evaluate the expected fraction of signal events in each BDT bin needed to constrain the dimuon invariant mass fit, presented in Chapter 7.

The calibrated BDT distributions for $B_s^0 \to \mu^+\mu^-$ and $B_{(s)}^0 \to \mu^+\mu^-$ decays in Run 1 and Run 2 are shown in Fig. 5.16 for the six BDT bins binning configuration. The Run 1 and Run 2 distributions are similar as the BDT calibration gives the expected fraction of signal events per BDT bin, not the absolute yields. Due to the excellent agreement between $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, the BDT calibration for $B_s^0 \to \mu^+\mu^-$ mode is used for both channels. These values are used to combine the signal yields obtained from the dimuon invariant mass fit, described in Sec. 7, into a single branching fraction measurement.

The expected fraction of $B_s^0 \to \mu^+\mu^-\gamma$ events per BDT bin is also needed to constrain the dimuon invariant mass fit. The BDT calibration for $B_s^0 \to \mu^+\mu^-\gamma$ decays is obtained following the same procedure as the $B_{(s)}^0 \to \mu^+\mu^-$ cases, using the simulated $B_s^0 \to \mu^+\mu^-\gamma$ events and correcting for the simulation and data differences observed in the $B_{(s)}^0$ kinematics and the occupancy of the event. The muonic PID and trigger requirements and the lifetime correction are applied to the expected relative yields obtained from $B_s^0 \to \mu^+\mu^-\gamma$ simulation samples. The calibrated BDT response for $B_s^0 \to \mu^+\mu^-\gamma$ decays is shown in Sec. 5.17 for Run 1 and Run 2. The resulting distributions are different from the ones for the $B_{(s)}^0 \to \mu^+\mu^-$ decays due to the different decay topologies and final-state kinematics. The expected fractions of $B_s^0 \to \mu^+\mu^-\gamma$ events per BDT bin are used to constrain the dimuon invariant mass fit and determine the initial-state radiation $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction.

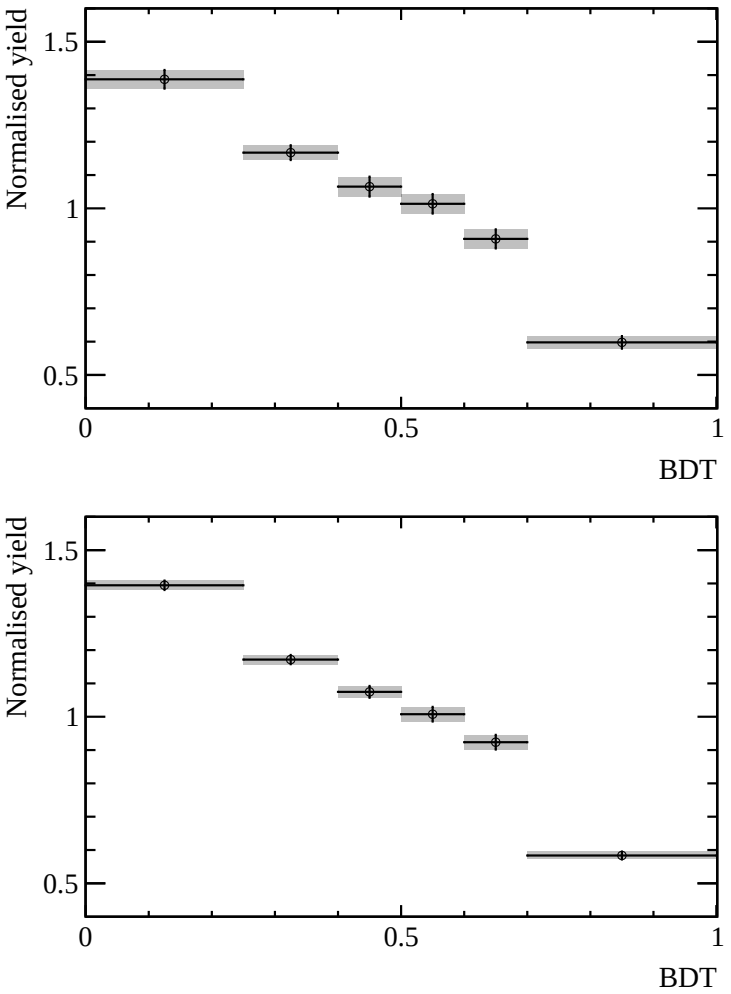


Figure 5.17: Calibrated BDT distribution of the $B_s^0 \to \mu^+\mu^-\gamma$ decay in Run 1 (up) and Run 2 (bottom). The shapes are obtained using $B_s^0 \to \mu^+\mu^-\gamma$ corrected simulation samples, as detailed in Sec. 5.2, and include the muonic PID and trigger factors, as well as the effective lifetime correction, as explained in Sec. 5.4.

6 | The mass spectrum

The dimuon invariant masses of the selected $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates are distributed in the range $m_{\mu^+\mu^-} \in [4900, 6000]\,\text{MeV}/c^2$, peaking at the masses of the B_s^0 and B^0 hadrons, respectively. This region is populated by several background components, as illustrated in Fig. 6.1, displaying the expected distribution of the different background components. Some of these backgrounds overlap with the region where the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ distributions are expected to peak, highlighted in green. In order to subtract them, the different background components must be adequately modelled in the fits to the dimuon invariant mass distributions, from which the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are determined. The fit stability and convergence are ensured by the application of some constraints to the fit input parameters, estimated from various calibration procedures.

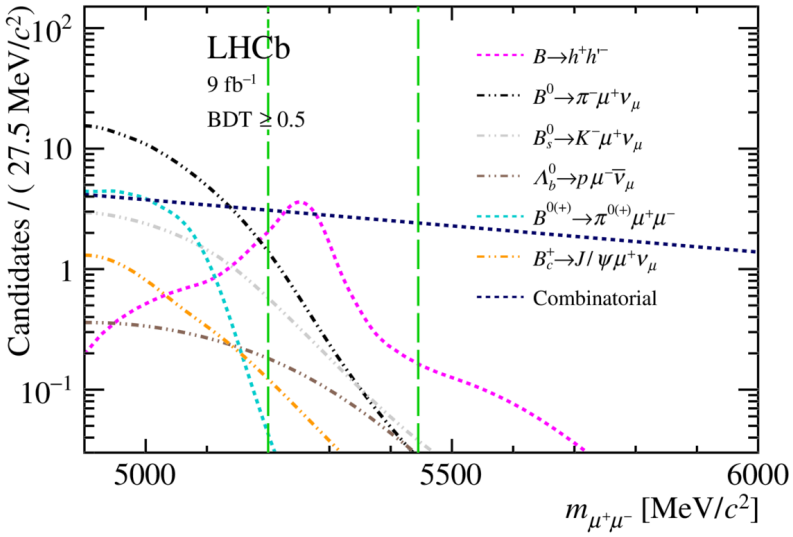


Figure 6.1: Expected dimuon invariant mass distributions of the background components for the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ distributions are expected to peak inside the region between green and dashed lines.

The $B^0_{(s)} \to \mu^+\mu^-$ signal components in the dimuon invariant mass fit are described using double-sided Crystal Ball functions. The parameters of these functions are calibrated for

the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ channels as presented in Sec. 6.1, while the shape used to describe the $B_s^0 \to \mu^+\mu^-\gamma$ signal component is presented in Chapter 7. Apart from the shape parameters, the expected fraction of $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ events ending up in each BDT bin is also estimated and constrained following the BDT calibration procedure described in Chapter 5.

As for the signal, the yields of the various background events in each BDT bin are also estimated and used to restrain the fit. The expected yields for each peaking background component are evaluated from data and simulation samples as described in Sec. 6.2, where the functions used to model their invariant mass distributions are also discussed.

The signal yields obtained from the dimuon invariant mass fits are normalized to the well-known branching fraction of other b-hadron decays to determine the signal branching fractions. This normalization procedure, presented in Sec. 6.3, is applied to reduce systematic effects in the ratio, as well as to avoid the dependence on the absolute luminosity in LHCb.

6.1 Mass calibration

The invariant mass distributions of $B_{(s)}^0 \to \mu^+\mu^-$ decays are modelled with double-sided Crystal Ball (DCB) functions. This function consists of a Gaussian shape connected to an enhanced exponential tail on each side. It is described by three sets of parameters: the mean and the width of the Gaussian core and two tail parameters for each exponential tail. For a detailed description of the double-sided Crystal Ball function, see Appendix A.1. The parameters describing the DCB for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ channels are estimated prior to the dimuon invariant mass fit to constrain their values and maximize the sensitivity to the signal yield. Different samples and decay channels are used to calibrate the mean, the width and the tail parameters. The $B_s^0 \to \mu^+\mu^-\gamma$ component is modelled in the fit using an empirical function presented in Chapter 7.

Because the dimuon invariant mass fit is performed in bins of the BDT classifier, the correlation between the mass shape parameters and the BDT is investigated. This is done using simulated $B_{(s)}^0 \to \mu^+\mu^-$ candidates by comparing the parameters describing the mass shapes in each BDT bin and the total mass shape. While the mean and the left tail parameters are invariant, the resolution and right tail parameters are found to depend on the BDT bin, and a BDT-dependent correction is applied to their total values.

6.1.1 Mean

The means of the DCB describing the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ invariant mass peaks are estimated from $B_s^0 \to K^+K^-$ and $B^0 \to K^+\pi^-$ data samples, respectively. The offline selection required on the $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ candidates, given in Sec.. 4.1, is the same as for the $B_{(s)}^0 \to \mu^+\mu^-$ candidates, with the only difference being the trigger and the PID re-

quirements, which depend on the final-state particles. In order to increase the event yield available in the data samples, the selected event candidates are required to be triggered by the particle tracks associated with the $B^0 \to K^+\pi^-$ ($B_s^0 \to K^+K^-$) candidate, following a TOS selection. Note that this trigger selection is different from the TIS trigger strategy used in the normalization and BDT calibration and described in Sec. 2.3.

Fig. 6.2 displays the fits to the mass distributions of the selected $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ candidates in Run 1 and Run 2 data. Several background components are present in both mass distributions. While the combinatorial background is described with an exponential shape, an ARGUS function (see Appendix A.3) is used for the partially reconstructed backgrounds. In the case of $B^0 \to K^+\pi^-$ decays, the background component at high mass corresponds to the $\overline{B}_s^0 \to K^+\pi^-$ decays and is modelled with the same shape as the $B^0 \to K^+\pi^-$ component. In the $B_s^0 \to K^+K^-$ fit, there is a small contribution from $\Lambda_b^0 \to ph$ decays, which are described with a double-sided Crystal Ball function. The shape parameters of all the backgrounds are constrained from fits to the simulation samples, except the slope of the exponential and the background yields, which are left free in the fit.

The resulting means for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ shapes are given in Table 6.1. The systematic uncertainty arises from an investigation of the correlation between the PID selection and the mean of the $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ peaks. To check the effect of this correlation, the mean is re-evaluated after varying the PID selection, both with tighter and looser criteria. The largest difference relative to the nominal value is assigned as systematic uncertainty. Because the mean estimations are found to be invariant under the BDT, the same values are used for all the BDT bins.

6.1.2 Resolution

The resolution of the $B_{(s)}^0 \to \mu^+\mu^-$ shapes is calibrated from data samples using Charmonium (J/ψ, $\psi(2S)$) and Bottonium ($\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$) resonances decaying to two muons. The mass distributions of these resonances are fitted to obtain the mean and the width describing the functions modelling each resonance. Fig. 6.3 shows the mass distributions of the $q\overline{q}$ resonances for Run 1 and Run 2 data with the fit model overlaid. The $q\overline{q}$ components are described by DCB functions, while an exponential is used for the combinatorial background.

The resolutions, $\sigma_{\mu\mu}$, for each $q\overline{q}$ resonance obtained from these fits are plotted versus the corresponding mean, $m_{\mu\mu}$. The resulting profile plots for Run 1 and Run 2 are shown in blue in Fig. 6.4. The two values are found to be correlated, following a relation modelled by the power law function

$$\sigma_{\mu\mu}(m_{\mu\mu}) = a_0 + a_1 \cdot m_{\mu\mu}^{\gamma} \tag{6.1}$$

where a_0, a_1 and γ are constants. Their values are extracted from a fit to the resolution data points versus the mean of the $q\overline{q}$ resonances, including the uncertainty on these values. The

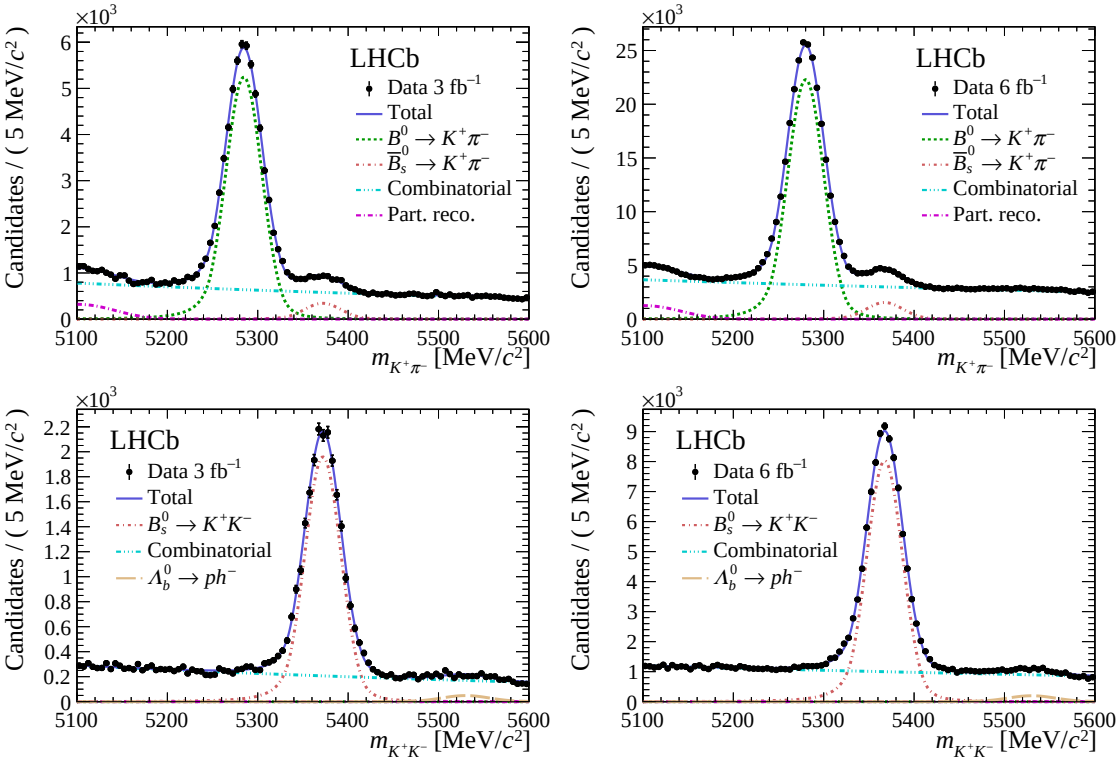


Figure 6.2: Mass distribution of the $B^0 \to K^+\pi^-$ (top) and $B^0_s \to K^+K^-$ (bottom) candidates in Run 1 (left) and Run 2 (right) samples. The blue shape corresponds to the total fit model and the constituent components are reported in the legend.

fits are superimposed in Fig. 6.4. The resolution of the $B^0_{(s)} \to \mu^+\mu^-$ shapes is extracted from the function in Eq. 6.1 by interpolation, requiring $m_{\mu^+\mu^-} = m_{B^0_{(s)}}$.

The resolutions estimated for the $B^0_{(s)} \to \mu^+\mu^-$ peaks are reported in Table 6.1. The total uncertainty results from the addition in quadrature of the statistical and systematic uncertainties. As can be seen from Fig. 6.4, the $\Upsilon(3S)$ data point is not entirely covered by the power-law function due to the large uncertainty associated with its mean value. Instead of a power law, a third-order polynomial is considered to better fit this data point. The difference in the $B^0_{(s)} \to \mu^+\mu^-$ resolution values is found to be 0.3 MeV/c^2 and is taken as systematic error.

As previously introduced, the resolution is found to vary depending on the BDT bin. This variation is related to the use of the angular distance between the two muons as one of the BDT training variables since the smaller the opening angle between the muon candidates, the worse the mass resolution of the decay. The variation per BDT bin is studied using simulated $B^0_{(s)} \to \mu^+\mu^-$ candidates smeared to match the resolution observed in data. The total and BDT-binned invariant mass shapes from the smeared simulation are compared. The ratios between their resolutions, ranging from 0.97 to 1.03, are used as a BDT-dependent

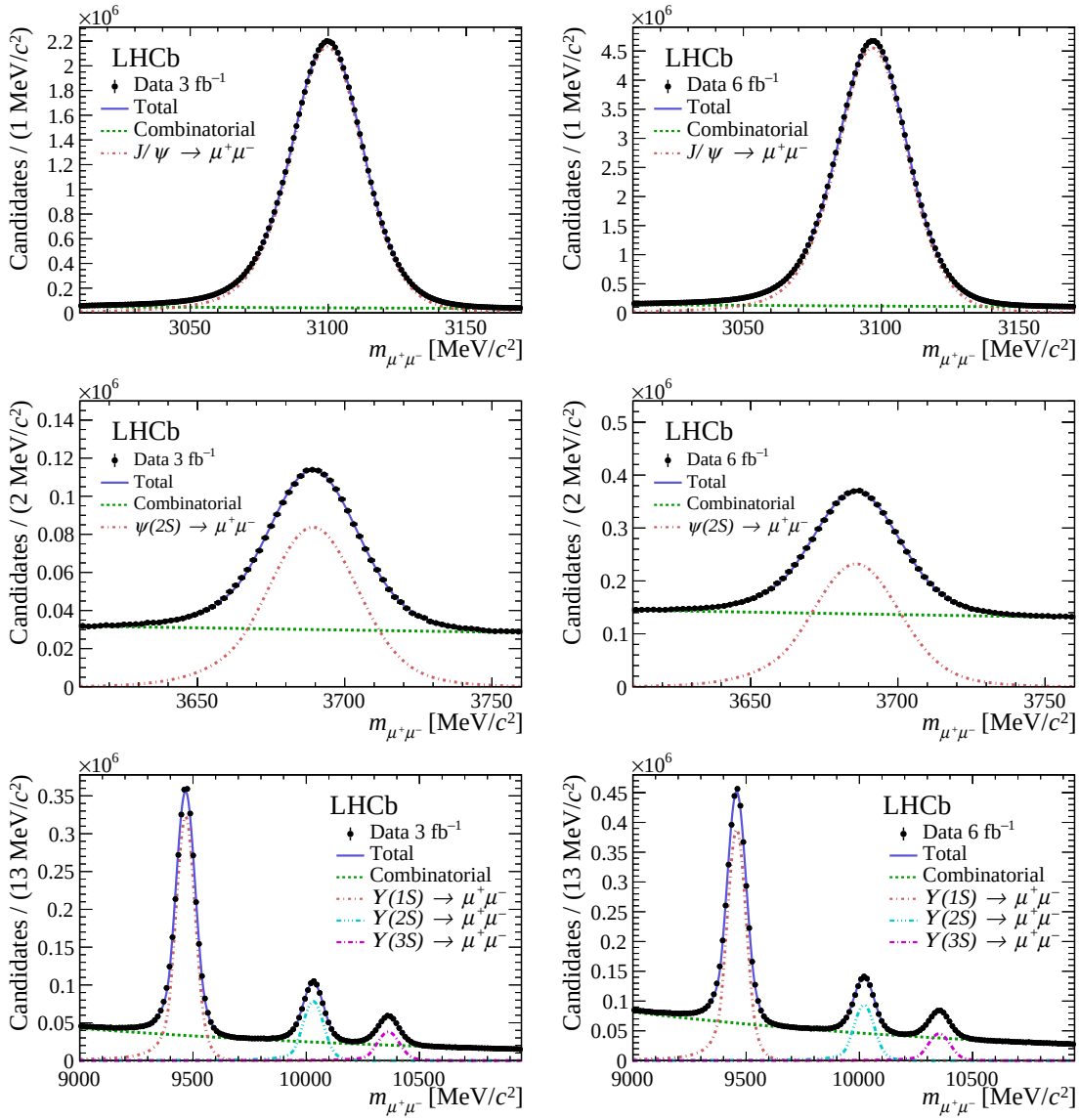


Figure 6.3: Mass distributions of the J/ψ (top), $\psi(2S)$ (center) and $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ (bottom) candidates decaying to two muons in Run 1 (left) and Run 2 (right) data. The blue line corresponds to the total fit model, with each component separated and reported in the respective legends.

correction factor to the values in Table 6.1.

6.1.3 Tail parameters

The tail parameters are common for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ shapes and are estimated from fits to simulation samples once smeared to result in a similar resolution as observed in data (see Sec. 6.1.2). The resulting tail parameter estimations for Run 1 and Run 2 are listed in

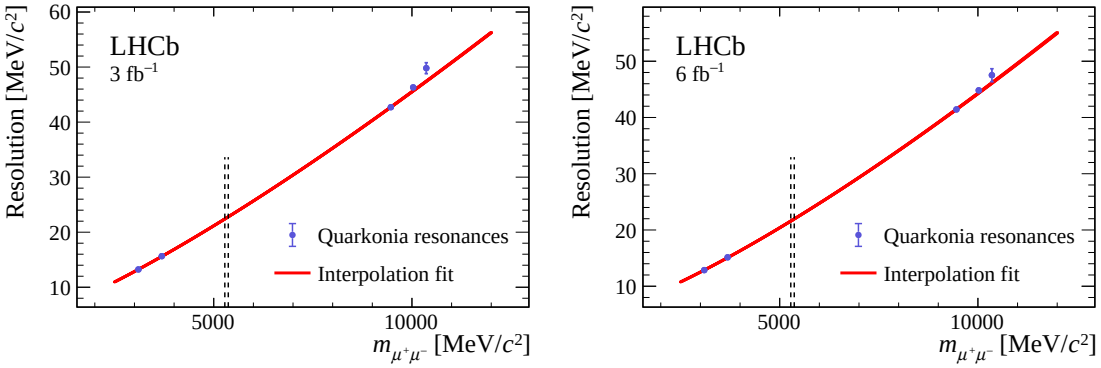


Figure 6.4: Fit to the width-versus-invariant mass distribution obtained from the quarkonia resonances in Run 1 (left) and Run 2 (right) data. The B^0 and B_s^0 mass values are marked in dashed lines.

Table 6.1. Since the Run 1 and Run 2 values are in good agreement, their values are combined into a common estimate used for both data-taking periods.

While the right tail parameters are found to be invariant, the left tail values show some correlation with the BDT. This correlation is taken into account by modifying the left tail parameters with a correction factor dependent on the BDT bin, whose values range from 0.9 to 1.2, as observed from simulation.

Table 6.1: Mass shape parameters for Run 1 and Run 2 with the statistical and systematic uncertainties added in quadrature. The difference in the mean values between Run 1 and Run 2 are due to the different momentum resolution of each data-taking period. The tail parameters for Run 1 and Run 2 are combined into a single estimate due to their excellent agreement.

	Run 1	Run 2	Combination
B_s^0 mean (MeV/c^2)	5372.27 ± 0.36	5367.54 ± 0.26	-
B^0 mean (MeV/c^2)	5284.61 ± 0.18	5280.13 ± 0.16	-
B_s^0 resolution (MeV/c^2)	22.8 ± 0.7	22.0 ± 0.6	-
B^0 resolution (MeV/c^2)	22.4 ± 0.6	21.6 ± 0.6	-
n left tail	1.55 ± 0.06	1.49 ± 0.02	1.50 ± 0.04
α left tail	1.79 ± 0.03	1.79 ± 0.01	1.79 ± 0.02
n right tail	5.86 ± 0.31	5.80 ± 0.24	5.81 ± 0.26
α right tail	2.12 ± 0.04	2.14 ± 0.03	2.14 ± 0.04

6.2 Peaking backgrounds

Because $B^0_{(s)} \to \mu^+\mu^-$ decays are very rare, a much larger number of background than signal events is expected to pollute the signal region. The background events are classified as combinatorial or physical backgrounds. The combinatorial background results from random combinations of two muon candidates originating from different decays in the event. This background is described with an exponential function whose slope is left free in the dimuon invariant mass fit. The physical backgrounds are specific b-hadron decays that might resemble the $B^0_{(s)} \to \mu^+\mu^-$ signature if they are partially reconstructed or if their final-state particles are misidentified as muons. The different sources of physical backgrounds, sketched in Fig. 4.2, are:

- $B^0_{(s)} \to h^+h'^-$ decays where $h, h' = K, \pi$, if the two hadrons in the final state are misidentified as muons. This misidentification occurs mainly from the decays-in-flight into muons of kaons and pions;
- $B_c^+ \to J/\psi\,\mu^+\nu$ with $J/\psi \to \mu^+\mu^-$, when the decay is partially reconstructed, missing one of the final-state muons;
- $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, when the final-state pion is missed and not reconstructed;
- $H_b \to h^\pm\mu^\mp\overline{\nu}_\mu$, if the hadron is misidentified as a muon, either from its punch-through in the hadronic calorimeter reaching the muon stations or due to its decay in flight to muons. The decays included in this type of background are $B^0 \to \pi^-\mu^+\nu_\mu$, $B_s^0 \to K^-\mu^-\nu_\mu$ and $\Lambda_b \to p\mu^-\nu$ decays.

In contrast to the combinatorial background, which is distributed through the whole invariant mass window, the physical background components peak at different regions in the dimuon mass spectrum. Although these backgrounds are strongly suppressed by the PID, isolation and BDT selections, they are not entirely discarded and their contributions must be modelled in the dimuon invariant mass fit. Each peaking background is modelled with a Gaussian kernel estimator obtained from simulation samples [62]. Apart from their shapes, the expected number of background events is needed to constrain the dimuon invariant mass fit and fix the mass distributions of each peaking background. Because the fit is done simultaneously in BDT bins, the background yields are determined for each BDT bin.

6.2.1 Misidentification rates

Despite the strong PID selection required for the final-state muon candidates, several background events are not rejected due to the misidentification of the kaons, pions and protons in their final state as muons. To estimate the expected number of background events in each BDT bin, the probability for these hadrons to pass the muon PID selection has to be computed. This probability, known as the *misidentification rate*, is evaluated as a function of

the momentum, p, and the transverse momentum, p_{T}, of the hadron tracks, using the same binning scheme as for the muonic PID efficiencies (see Sec. 4.3).

Two different strategies are followed to calculate the misidentification rates depending on how it occurs. The proton misidentification originates from random combinations of hits in the muon chambers with tracks left by the protons in the tracking stations or from the punch-through of protons in the calorimeters reaching the muon chambers. In the case of the kaons and pions, the leading cause of misidentification is their decays-in-flight to muons. To account for this difference, a more elaborate procedure is needed to evaluate the misidentification rate of pions and kaons.

Protons

The proton misidentification rate is computed from data samples using $\Lambda \to p\pi$ decays and Λ_c decays to protons as control channels [55]. The control channel is separated from the background components in the data samples using the sPlot technique [52]. The proton misidentification rate is evaluated in bins of the momentum and transverse momentum of the proton track as the fraction of proton tracks passing the muonic PID selection. The effect of the choice of a specific binning scheme is studied as a systematic by re-computing the misidentification rates with a different binning scheme. The difference concerning the nominal values is assigned as systematic uncertainty.

Pions and kaons

The single misidentification rates for kaons and pions are evaluated from a combination of data and simulation samples using $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ as the control channel. This channel is used due to its low background and from the charges of the different particles in the final state, which allows to determine the kaon (pion) track without requiring particle identification (see Sec. 4.3). Initially, a similar technique as for the protons was studied where only data samples were used. In these samples, the background-subtracted distribution for the misidentification performance was obtained from the m_{D^0} and $\Delta m = m_{D^{*+}} - m_{D^0}$ distributions using the sPlot technique. However, the sPlot technique is based on the assumption that the variables used to separate the different fit components, namely the m_{D^0} and Δm variables, are independent of the variable of interest, or the muonic PID criteria in this case [52]. This condition does not hold in the pion and kaon case since the primary cause of their misidentification are their decays-in-flight, which alter the m_{D^0} and Δm distributions. This phenomenon thus leads to a correlation between the muonic PID selection and the m_{D^0} and Δm distributions.

A new technique was developed to compute the misidentification rates of kaons and pions without applying the sPlot procedure. First, the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ data sample is split into subsamples of the momentum and transverse momentum (p, p_{T}) of the probe

track, either the final-state kaon or pion from the reconstructed $D^0 \to K^-\pi^+$ decay. Two different two-dimensional fits of the m_{D^0} and Δm distributions are then performed in each subsample: one where the muonic PID selection is applied to the probe tracks and the other where it is not required. The misidentification rate in each (p, p_T) subsample is evaluated as the ratio between the yields from these two fits.

The two-dimensional fit to m_{D^0} and Δm is performed inside a mass window, requiring the mass of the D^0 candidate to be centred around the D^0 mass, $m_{\pi K} \in [1825, 1910]$ MeV/c^2. With this mass-window selection, part of the hadrons decaying-in-flight is rejected because their associated D^0 mass distribution is broader. This effect is illustrated in Fig. 6.5, which shows the m_{D^0} distribution without requiring PID in the hadron (blue) and requiring the muonic PID (red), described in Sec. 4.3. The area in grey corresponds to the mass window used to perform the two-dimensional fits. The mass tail is longer when the kaon is misidentified as a muon due to the larger mass difference between kaons and muons. The number of events falling outside the mass window is evaluated from simulated $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ candidates both for the pion and kaon case. A correction factor is calculated from these samples as the inverse of the mass-window efficiencies, $N_{\text{total}}/N_{\text{in-mass}}$, and applied to the misidentification rates obtained above. The correction is computed for each p and p_T region, and the values range from 1 to 5.8 depending on the kinematic bin and the hadron decaying in flight.

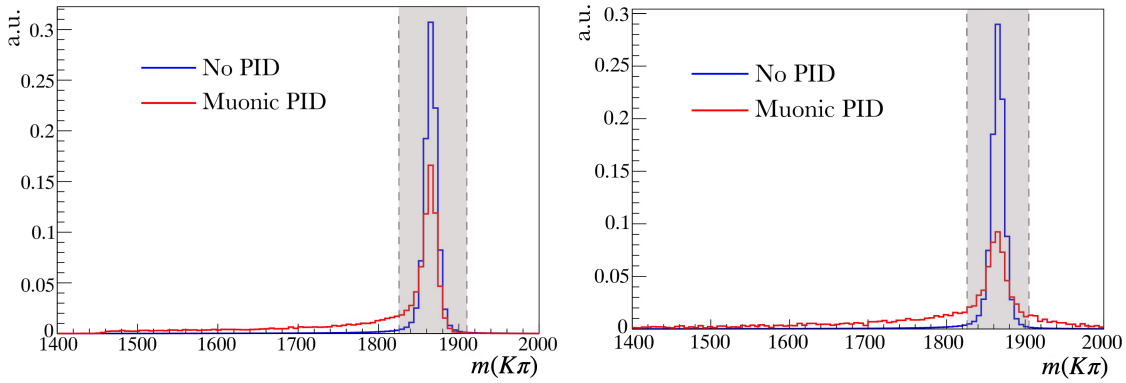


Figure 6.5: Mass distribution of the D^0 candidate when no PID selection is required (blue) and when the muonic PID selection (red) is required on the pion (left) and on the kaon (right).

To test the performance and the deviations arising from the modelling of the partially reconstructed backgrounds polluting the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ invariant mass distribution, the misidentification rates are evaluated from fits to the Δm variable only. Although more straightforward than the nominal case, the Δm fit is blinded to the presence of these partially reconstructed backgrounds. The fit is performed in a tighter mass window to reduce this background. The differences relative to the two-dimensional fit values are quoted as systematic uncertainty in the misidentification rates.

6.2.2 $B^0_{(s)} \to h^+h'^-$ background

The $B^0_{(s)} \to h^+h'^-$ decay is one of the peaking backgrounds in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis when both of their final-state hadrons are misidentified as muons. The yield of misidentified $B^0_{(s)} \to h^+h'^-$ decays after all the selection procedures described in Chapter 4 is evaluated for each BDT bin as

$$N_{B^0_{(s)}\to h^+h'^-\to\mu^+\mu^-} = \epsilon^{\text{trg}}_{B^0_s\to\mu^+\mu^-} \cdot \frac{N^{\text{trg}}_{hh}}{\epsilon^{\text{trg}}_{hh}} \cdot \epsilon_{hh\to\mu\mu} \tag{6.2}$$

where $\epsilon_{B^0_s\to\mu^+\mu^-}$ is the trigger efficiency of the $B^0_s \to \mu^+\mu^-$ decay computed in Sec. 4.2, N^{trg}_{hh} is the number of $B^0_{(s)} \to h^+h'^-$ triggered events, $\epsilon^{\text{trg}}_{hh}$ the trigger efficiency associated to the $B^0_{(s)} \to h^+h'^-$ decays evaluated per BDT bin in Sec. 5.1.2 and $\epsilon_{hh\to\mu\mu}$ is the double misidentification rate. The number of $B^0_{(s)} \to h^+h'^-$ triggered events is obtained by correcting the $B^0 \to K^+\pi^-$ yields determined in Sec. 5.1.1 by its expected fraction.

It is important to note that all the elements in Eq. 6.2 are evaluated per BDT bin. The double misidentification rate, $\epsilon_{hh\to\mu\mu}$, is obtained from the convolution of the corresponding kaon or pion single misidentification rates, evaluated as described in Sec. 6.2.1, with the spectrum of each hadron in the $B^0_{(s)} \to h^+h'^-$ final state. The BDT dependence is included by considering the spectra for each BDT bin separately.

A separate estimate of the $B^0_{(s)} \to h^+h'^-$ yields per BDT bin is performed to test the accuracy of the single misidentification rates and their possible correlation with the BDT response. This procedure exploits $B^0_{(s)} \to h^+h'^-$ data samples under the $K\mu$ ($\pi\mu$) selection, where one of the final-state hadrons is required to pass a strong kaon (pion) selection criteria, while the muonic PID selection is applied to the other. From the fits to the mass distributions of these candidates performed for each BDT bin separately, the $B^0_{(s)} \to h^+h'^-$ yields with one misidentified hadron are obtained per BDT bin. These values are corrected by the $K \to \mu$ ($\pi \to \mu$) misidentification rates, PID and trigger requirements to obtain the yield estimates. The results are compared with the BDT-binned yields obtained from Eq. 6.2 and the difference between the two values is applied as a correction factor.

The estimate of the expected number of $B^0_{(s)} \to h^+h'^-$ events per BDT bin are given for Run 1 and Run 2 in Tables 6.2 and 6.3, respectively. Although $B^0_{(s)} \to h^+h'^-$ is not the most abundant background, especially for Run 2, its proper estimate is essential due to its overlap with the $B^0 \to \mu^+\mu^-$ component in the dimuon invariant mass fit, for which even fewer events are expected (see the expected signal yields in Sec. 6.3.5).

6.2.3 Partially reconstructed background

Apart from the $B^0_{(s)} \to h^+h'^-$ decays, several semileptonic b-hadron decays with real muons in their final state are also backgrounds of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis. Some of these back-

grounds have two real muons among the final-state particle, such as the $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ or $B_c^+ \to J/\psi\,\mu^+\nu$ decays with the $J/\psi \to \mu^+\mu^-$. Others have one real muon and a hadron that might be misidentified as muon, such as the $B^0 \to \pi^-\mu^+\nu_\mu$, $B_s^0 \to K^-\mu^-\nu_\mu$ and $\Lambda_b \to p\mu^-\nu$ decays. Due to the presence of neutrinos or the missing of one of the final-state particles, these decays are partially reconstructed and their associated dimuon invariant mass distributions overlap with the $B^0_{(s)} \to \mu^+\mu^-$ signal peaks, though typically they have a broader spectrum around lower invariant mass values.

An accurate estimate of the expected background yields in each BDT bin is required to constrain the background mass shapes. For each background channel, the expected yields are evaluated by normalizing relative to the $B^+ \to J/\psi K^+$ yields as

$$N_{\text{bkg}} = \frac{f_{\text{bkg}}}{f_u} \cdot \frac{N_{B^+\to J/\psi K^+}}{\mathcal{B}_{B^+\to J/\psi K^+} \cdot \epsilon_{B^+\to J/\psi K^+}} \cdot \mathcal{B}_{\text{bkg}} \cdot \epsilon_{\text{bkg}} \tag{6.3}$$

where f_{bkg} and f_u are the hadronization fractions of the initial-state b-hadrons of the background and the $B^+ \to J/\psi K^+$ channels, the $\mathcal{B}_{\text{bkg}}$ and $\mathcal{B}_{B^+\to J/\psi K^+}$ their branching fractions, ϵ_{bkg} and $\epsilon_{B^+\to J/\psi K^+}$ the total efficiencies associated to these channels, and $N_{B^+\to J/\psi K^+}$ the number of $B^+ \to J/\psi K^+$ events determined from fits to the invariant mass distribution of selected $B^+ \to J/\psi K^+$ candidates from data samples.

The total efficiencies include the reconstruction, selection, PID and trigger efficiencies for both channels. The BDT dependence in Eq. 6.3 is embedded in these total efficiencies as they are determined for each BDT bin separately, except for the geometrical acceptance efficiencies, which account for the geometrical shape of the LHCb detector and the generation efficiency. All the efficiencies are evaluated from simulation samples except the PID efficiency, which is obtained using the misidentification rates, presented in Sec. 6.2.1, and the PID efficiencies given in Sec. 4.3. The rest of the factors in Eq. 6.3 are not BDT dependent.

The branching fractions for each background channel are obtained from a combination of theoretical and experimental references. The branching fractions and hadronization fractions evaluated from LHCb measurements are used for the $B_c^+ \to J/\psi\,\mu^+\nu$ [63], $\Lambda_b \to p\mu^-\nu$ [64, 65] and $B_s^0 \to K^-\mu^-\nu_\mu$ [66, 67] background modes. For the $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B^0 \to \pi^-\mu^+\nu_\mu$ channels, the estimates of the branching fractions on Refs. [1, 68] are used while for the hadronization fractions $f_u = f_d$ is assumed [67].

The expected yields per BDT bin for each semileptonic background are given for Run 1 and Run 2 in Tables 6.2 and 6.3, respectively. The most dominant backgrounds are partially reconstructed $B^0 \to \pi^-\mu^+\nu_\mu$ decays followed by $B^0_{(s)} \to h^+h'^-$, $B_s^0 \to K^-\mu^-\nu_\mu$ and $B_c^+ \to J/\psi\,\mu^+\nu$ while very low yields are expected for the $\Lambda_b \to p\mu^-\nu$ component both in Run 1 and Run 2. The scaling from Run 1 to Run 2 varies for different background components due to the different selections applied in each run, mainly the trigger and the muonic PID requirements, affecting the various background types differently.

Table 6.2: Estimated yields per BDT bin for the peaking background components in Run 1 data.

BDT bin	$B^0_{(s)} \to h^+h'^-$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda_b \to p\mu^-\nu$	$B^+_c \to J/\psi\,\mu^+\nu$
[0.25-0.4]	3.8 ± 0.9	8.8 ± 0.7	0.94 ± 0.20	2.72 ± 0.35	0.9 ± 0.4	0.29 ± 0.32	2.84 ± 0.21
[0.4-0.5]	1.57 ± 0.20	6.0 ± 0.4	0.76 ± 0.16	1.47 ± 0.19	0.65 ± 0.29	0.18 ± 0.20	1.38 ± 0.13
[0.5-0.6]	1.61 ± 0.21	5.8 ± 0.4	0.90 ± 0.19	1.22 ± 0.16	0.65 ± 0.28	0.20 ± 0.22	0.80 ± 0.09
[0.6-0.7]	1.65 ± 0.21	5.8 ± 0.4	1.03 ± 0.22	0.99 ± 0.13	0.65 ± 0.29	0.19 ± 0.21	0.58 ± 0.08
[0.7-1.0]	5.3 ± 0.7	11.6 ± 0.8	2.5 ± 0.5	1.46 ± 0.19	1.4 ± 0.6	0.32 ± 0.35	0.41 ± 0.06
[0.25-1.0]	13.9 ± 1.2	30.8 ± 1.8	9.0 ± 1.1	7.6 ± 0.5	4.2 ± 0.9	1.2 ± 0.6	6.01 ± 0.28

Table 6.3: Estimated yields per BDT bin for the peaking background components in Run 2 data.

BDT bin	$B^0_{(s)} \to h^+h'^-$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda_b \to p\mu^-\nu$	$B^+_c \to J/\psi\,\mu^+\nu$
[0.25-0.4]	6.2 ± 1.0	40.1 ± 2.6	3.3 ± 0.7	9.9 ± 1.3	3.5 ± 1.5	1.1 ± 1.2	11.7 ± 0.6
[0.4-0.5]	2.90 ± 0.26	24.6 ± 1.6	2.5 ± 0.5	5.2 ± 0.7	2.3 ± 1.0	0.8 ± 0.8	4.59 ± 0.27
[0.5-0.6]	2.82 ± 0.26	24.4 ± 1.6	2.9 ± 0.6	4.2 ± 0.5	2.3 ± 1.0	0.9 ± 0.9	2.85 ± 0.19
[0.6-0.7]	2.68 ± 0.24	23.5 ± 1.5	3.3 ± 0.7	3.3 ± 0.4	2.3 ± 1.0	0.9 ± 1.0	1.56 ± 0.12
[0.7-1.0]	8.1 ± 0.7	44.6 ± 2.9	7.7 ± 1.6	4.8 ± 0.6	4.7 ± 2.1	1.6 ± 1.8	0.85 ± 0.08
[0.25-1.0]	22.7 ± 1.3	130 ± 5	21.5 ± 2.5	26.5 ± 1.7	14.7 ± 3.0	5.3 ± 2.7	21.6 ± 0.7

6.3 Normalization

The branching fraction for a decay represents the probability of the mother particle to decay via the specific decay mode. For the $B^0_{(s)} \to \mu^+\mu^-$ case, the branching fraction is given by

$$\mathcal{B}_{B^0_{(s)}\to\mu^+\mu^-} = \frac{\Gamma(B^0_{(s)} \to \mu^+\mu^-)}{\Gamma(B^0_{(s)})} \tag{6.4}$$

Experimentally, the branching fraction is calculated as the fraction of the number of events decaying through the specific decay mode out of all the particles decaying, which can be evaluated from the subset of events taking place at the LHCb. The number of total $B^0_{(s)} \to \mu^+\mu^-$ decays in the LHCb is obtained from the well reconstructed, detected and selected $B^0_{(s)} \to \mu^+\mu^-$ candidates in LHCb, $N_{B^0_{(s)}\to\mu^+\mu^-}$, once corrected for the total detection efficiency associated with the $B^0_{(s)} \to \mu^+\mu^-$ candidates, $\epsilon_{B^0_{(s)}\to\mu^+\mu^-}$. Regarding the total number of $B^0_{(s)}$ hadrons produced in LHCb, it is calculated as

$$N_{B^0_{(s)}} = 2 \cdot \mathcal{L} \cdot \sigma_{pp\to b\bar{b}} \cdot f_{d(s)} \tag{6.5}$$

where $\mathcal{L}$ is the integrated LHCb luminosity giving the amount of pp collisions occurring at LHCb; $\sigma_{pp\to b\bar{b}}$ is the $b\bar{b}$ cross-section giving the probability for a $b\bar{b}$ pair to be generated in a pp collision; and $f_{d(s)}$ is the hadronization fraction accounting for the probability of a b quark to hadronize with a d (s) quark and form a $B^0_{(s)}$ hadron. The factor of 2 arises from the inclusion of the implicit charged conjugation as the $B^0_{(s)}$ and $\bar{B}^0_{(s)}$ are not distinguished experimentally, as explained in Chapter 1. Therefore, the expression for the branching fractions of $B^0_{(s)} \to \mu^+\mu^-$ decays is evaluated from the experiment as:

$$\mathcal{B}_{B^0_{(s)}\to\mu^+\mu^-} = \frac{N_{B^0_{(s)}\to\mu^+\mu^-}}{2 \cdot \mathcal{L} \cdot \sigma_{pp\to b\bar{b}} \cdot f_{d(s)} \cdot \epsilon_{B^0_{(s)}\to\mu^+\mu^-}} \tag{6.6}$$

The $\mathcal{L}$ and $\sigma_{pp\to b\bar{b}}$ in Eq. 6.6 have large uncertainties associated with them, reducing the precision of the measured $B^0_{(s)} \to \mu^+\mu^-$ branching fraction. An alternative to Eq. 6.6 is to normalize the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction relative to another b-hadron decay with a well-known branching fraction, known as the *normalization channel*. In this way, the two main sources of uncertainty in Eq. 6.6 cancel out as they are common to the signal and normalization channels. Using the normalization channel, the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction is expressed as:

$$\mathcal{B}_{B^0_{(s)}\to\mu^+\mu^-} = \frac{\mathcal{B}_{norm}}{N_{norm}} \frac{\epsilon_{norm}}{\epsilon_{B^0_{(s)}\to\mu^+\mu^-}} \frac{f_{norm}}{f_{d(s)}} \cdot N_{B^0_{(s)}\to\mu^+\mu^-} = \alpha_{d(s)} \cdot N_{B^0_{(s)}\to\mu^+\mu^-} \tag{6.7}$$

where N_{norm} are the yields, ϵ_{norm} the efficiencies, f_{norm} the hadronization fraction and $\mathcal{B}_{norm}$ the branching fraction of the normalization channel; and $\alpha_{d(s)}$ is the normalization factor gathering all the information concerning the normalization and representing, effectively, the single-event "power" of the selected $B^0_{(s)} \to \mu^+\mu^-$ events.

Apart from the luminosity and the $b\bar{b}$ cross-section, several other sources of uncertainty are avoided if the normalization channel is chosen wisely. A good normalization channel has a similar decay topology as the signal mode, is detected with similar efficiency as the signal decay and has a well-known and precise branching fraction value. These requirements are better fulfilled in this analysis using two complementing normalization channels: the $B^0 \to K^+\pi^-$ and the $B^+ \to J/\psi K^+$ decays.

$B^0 \to K^+\pi^-$ decays share the two-body decay topology also characteristic of the $B^0_{(s)} \to \mu^+\mu^-$ decays. The kinematic distributions of the B candidate are similar in both decays, as well as those associated with the final-state particles. Unfortunately, the particle identification and trigger are very different for hadrons and muons. Therefore, an additional and complementary channel is used, namely the $B^+ \to J/\psi K^+$ mode, where $J/\psi \to \mu^+\mu^-$. Although the decay topology is different, the trigger and particle identification criteria for the muons from the J/ψ in the $B^+ \to J/\psi K^+$ decay are very similar to the $B^0_{(s)} \to \mu^+\mu^-$ channel. Precise measurements exist for the branching fractions of both normalization channels. In this analysis their values are taken as $\mathcal{B}(B^+ \to J/\psi K^+) = (6.02 \pm 0.17) \times 10^{-5}$ [1] and $\mathcal{B}(B^0 \to K^+\pi^-) = (1.96 \pm 0.05) \times 10^{-5}$ [1,69–71].

Apart from the branching fractions of the normalization channels, other values such as the normalization channel yields, the ratio of detection efficiencies and the hadronization fractions are needed in the computation of the normalization factors, $\alpha_{d(s)}$ of Eq. 6.7. The evaluation of these factors is given in the following sections. A check, presented in Sec. 6.3.4, is performed to validate the ratio of the detection efficiencies and the yields of the normalization channels by evaluating the ratio of the $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ branching fractions and comparing it with its known value [1]. The resulting normalization factors are listed in Sec. 6.3.5.

6.3.1 Normalization channel yields

$B^+ \to J/\psi K^+$ yields

The $B^+ \to J/\psi K^+$ yields are obtained from data samples using fits to the reconstructed B^+ candidate mass with the J/ψ mass constrained to its measured value [1]. The $B^+ \to J/\psi K^+$ candidates are selected similarly to the signal. The selection criteria, including the offline selection and trigger requirements, are presented in Sec. 4.

The mass distributions of the reconstructed B candidates and their corresponding fits are displayed in Fig. 6.6 for all the data-taking years. The $B^+ \to J/\psi K^+$ shape is modelled with a

Hypatia function (see Appendix A.2). Except for the mean and the width, which are left free in the fit, the parameters describing the signal shape are allowed to vary following a Gaussian inside a specific range of values determined from simulation samples. Two background components pollute the $B^+ \to J/\psi K^+$ peak: the combinatorial and the misidentified $B^+ \to J/\psi \pi^+$ backgrounds. The combinatorial background is described by an exponential function with the slope treated as a free variable in the fit. To model the misidentified $B^+ \to J/\psi \pi^+$ background, a non-parametric function is used, which is obtained from simulated $B^+ \to J/\psi \pi^+$ events reconstructed using the wrong kaon hypothesis.

The systematic effect arising from the signal shape parameters is studied by repeating fit using two approaches and comparing the results: in the first approach, the parameters are kept fixed to the values obtained from simulation, while in the second, they are left free. Both methods lead to compatible results and no systematic uncertainty is added. As a second systematic effect, the shape modelling the misidentified $B^+ \to J/\psi \pi^+$ background is verified by repeating the fit using the analytical shape presented in Sec. 4.2.1. The resulting yields are found to be compatible with the nominal model and no systematic uncertainty is assigned.

The $B^+ \to J/\psi K^+$ yields for each data-taking year are combined to obtain the values for each data-taking period, listed in Table 6.4.

$B^0 \to K^+\pi^-$ yields

The $B^0 \to K^+\pi^-$ yields are evaluated from the mass distribution of the reconstructed B^0 candidates in data samples. The selection applied to $B^0 \to K^+\pi^-$ candidates is similar to the $B^0_{(s)} \to \mu^+\mu^-$ candidates, except for the PID and trigger criteria, which differ from hadrons and muons. To reduce this difference, the selected events are required to be triggered independently of the $B^0 \to K^+\pi^-$ candidate. The offline selection, PID and trigger criteria applied to the $B^0 \to K^+\pi^-$ candidates are described in Sec. 4.

Fig. 6.7 illustrates the mass distributions of $B^0 \to K^+\pi^-$ candidates for each data-taking year. The fit strategy is identical to the fit used for the estimation of the mean in the mass calibration in Sec. 6.2. The $B^0 \to K^+\pi^-$ yields for each data-taking year obtained from the fits are combined. The resulting values for each data-taking period are listed in Table 6.4.

6.3.2 Hadronization fractions

The hadronization fractions, f_q, represent the probability of a b quark to hadronize with a q quark to form a B_q meson. The hadronization fractions associated with the B^0, B^0_s and B^+ mesons, characteristic of the signal and normalization channels, are f_d, f_s and f_u, respectively. For the normalization procedure, only the ratio of hadronization fractions between the signal and normalization channels is needed, concretely f_d/f_s and f_u/f_s for the $B^0_s \to \mu^+\mu^-$ signal channel, and f_u/f_d for the $B^0 \to \mu^+\mu^-$ channel.

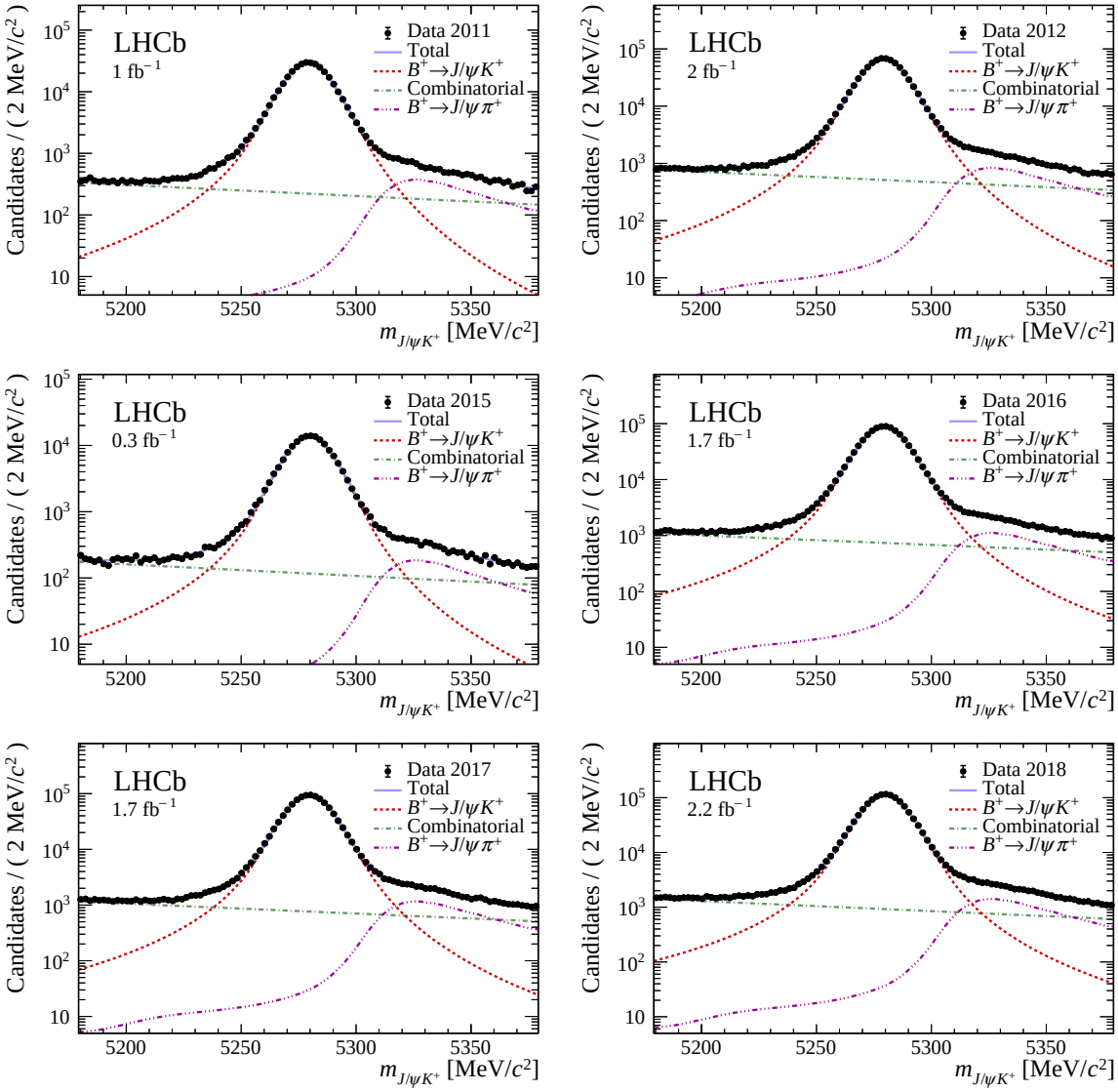


Figure 6.6: Mass distributions of the $B^+ \to J/\psi K^+$ selected candidates in data for each data-taking year used in the normalization. The blue line corresponds to the total fit model and the separated components are reported in the legend.

The f_s/f_d ratio is obtained at different pp centre-of-mass energies from measurements performed at the LHCb using different decays [65, 72–75]. These measurements are combined in Ref. [67]. Although the f_s/f_d ratio is found to be dependent on the pp centre of mass energy and the transverse momentum of the B candidate, only the p_{T}-integrated values for Run 1 and Run 2 are used in this thesis since the average is obtained from a sample of B mesons with the same spectrum as the signal channels.

The f_s/f_d ratios for Run 1 and Run 2 are strongly correlated because the same theoretical and experimental inputs are used in their computation at 7 TeV, 8 TeV and 13 TeV. To properly

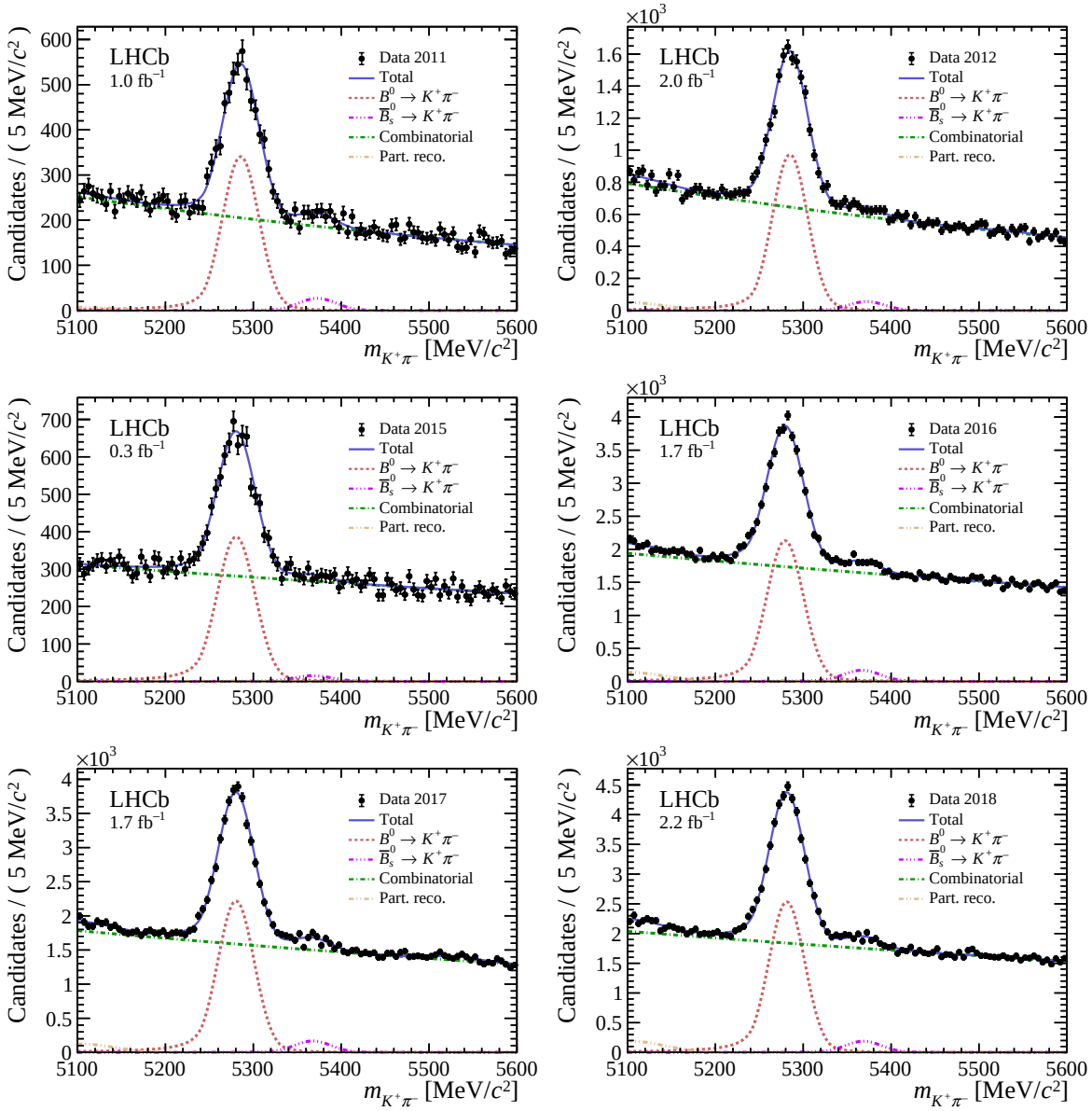


Figure 6.7: Mass distributions of the $B^0 \to K^+\pi^-$ selected candidates in data for each data-taking year used in the normalization. The blue line corresponds to the total fit model with the separated components reported in the legend.

deal with the uncertainty correlation, only the f_s/f_d ratio at 13 TeV is used, f_s/f_d (13 TeV) = 0.254 ± 0.008 [67], while for Run 1, the f_s/f_d ratio is normalized to the 13 TeV value as

$$\frac{f_s/f_d(13\text{ TeV})}{f_s/f_d(\text{Run 1})} = 1.064 \pm 0.007 \tag{6.8}$$

Isospin symmetry is assumed in the b hadronization processes taking place at the LHC [67], such that the f_d and f_u fractions are assumed to be identical. This implies that

Table 6.4: Yields for the $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ normalization channels for Run 1 and Run 2 obtained from the fits to data. The values are not corrected by the selection efficiencies. The uncertainties result from the addition in quadrature of the statistical and systematic uncertainties. The ratios between Run 1 and Run 2 are different for both channel since the trigger and PID efficiencies are larger for the $B^0 \to K^+\pi^-$ channel.

Channel	Run 1	Run 2
$B^0 \to K^+\pi^-$	$(14.05 \pm 0.26) \times 10^3$	$(7.99 \pm 0.10) \times 10^4$
$B^+ \to J/\psi K^+$	$(11.259 \pm 0.015) \times 10^5$	$(36.072 \pm 0.026) \times 10^5$

$f_u/f_d = 1$ and the values for f_s/f_d are also used for the $B^+ \to J/\psi K^+$ normalization channel. Although an updated value for the f_s/f_d hadronization ratio is used with respect to the previous analysis, its associated relative uncertainty of 3% is still the largest relative uncertainty associated with a factor in the normalization procedure and, therefore, it is still the dominant systematic uncertainty in the branching fraction measurements.

6.3.3 Efficiencies

Another ingredient in the normalization procedure is the detection efficiencies for the signal and normalization channels, particularly their ratio. The ratio of detection efficiencies can be split into distinct terms for each phase in the detection and selection processes as

$$\frac{\epsilon_{\text{norm}}}{\epsilon_{B^0_{(s)}\to\mu^+\mu^-}} = \frac{\epsilon^{\text{Acc}}_{\text{norm}}}{\epsilon^{\text{Acc}}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{\text{RecSel|Acc}}_{\text{norm}}}{\epsilon^{\text{RecSel|Acc}}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{\text{PID|RecSel}}_{\text{norm}}}{\epsilon^{\text{PID|RecSel}}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{\text{Trg|PID}}_{\text{norm}}}{\epsilon^{\text{Trg|PID}}_{B^0_{(s)}\to\mu^+\mu^-}} \tag{6.9}$$

where ϵ_{Acc} are the geometrical acceptance efficiencies, $\epsilon^{\text{RecSel|Acc}}$ the reconstruction and selection efficiencies, $\epsilon^{\text{PID|RecSel}}$ the PID efficiencies and $\epsilon^{\text{Trg|PID}}$ the trigger efficiencies. At each stage, the efficiency is computed for the subset of events passing the requirements from the previous step. The evaluation of these efficiencies for the signal and normalization channels is given in Chapter 4.

The first BDT bin, corresponding to the range [0, 0.25), is dominated by combinatorial background and excluded from the dimuon invariant mass fit. The $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ yields obtained from these fits should be corrected for the efficiencies associated with the exclusion of the first BDT bin. These efficiencies are obtained for each signal channel from the BDT calibration response in Chapter 5, and their values are listed in Table 6.5, with combined statistical and systematic uncertainties added in quadrature.

Table 6.5: Efficiency associated with the exclusion of the first BDT bin, i.e. BDT < 0.25, from the dimuon invariant mass fit. The values are reported for each signal channel in Run 1 and Run 2 data. The statistical and systematic uncertainties are added in quadrature.

	Run 1	Run 2
$B_s^0 \to \mu^+\mu^-$	0.723 ± 0.006	0.7071 ± 0.0026
$B^0 \to \mu^+\mu^-$	0.720 ± 0.006	0.7036 ± 0.0027
$B_s^0 \to \mu^+\mu^-\gamma$	0.656 ± 0.007	0.6531 ± 0.0035

6.3.4 Validation check

The ratio of the detection efficiencies and the normalization yields from Sec. 6.3.1 and 6.3.3 are validated using the branching fraction of the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ normalization channels. The ratio between the two branching fractions is evaluated from data as:

$$\frac{\mathcal{B}(B^0 \to K^+\pi^-)}{\mathcal{B}(B^+ \to J/\psi K^+)} = \frac{N_{B^0 \to K^+\pi^-}}{N_{B^+ \to J/\psi K^+}} \frac{\epsilon_{B^+ \to J/\psi K^+}}{\epsilon_{B^0 \to K^+\pi^-}} \frac{f_u}{f_d} \tag{6.10}$$

where the last factor cancels out since isospin asymmetry is assumed, i.e. f_u/f_d.

The ratio for Run 1, including the statistical uncertainty, is measured to be 0.340 ± 0.016 and for Run 2, 0.336 ± 0.012. Both values are in agreement with the ratio obtained using the known branching fractions for $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ decays, resulting in a ratio of 0.326 ± 0.012 [1]. This is a strong validation for the calibration of the various efficiencies, mainly trigger and PID efficiencies, involved in this analysis.

6.3.5 Normalization factors

The normalization factors obtained using the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays as normalization modes are given for BDT > 0.25 in Table 6.6. Both normalization factors are found to be compatible within uncertainties. A single normalization factor is determined by averaging the results from the two normalization channels and taking into account the full correlation from the f_s/f_d ratio, the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ efficiencies and the tracking efficiency corrections. Different normalization factors are determined for $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ to account for the different efficiencies values associated with each channel (see Chapter 4). The expected yields for each signal channel assuming the Standard Model branching fractions are also reported for BDT > 0.25 in Table 6.6. The low expected yields for $B^0 \to \mu^+\mu^-$ and especially for $B_s^0 \to \mu^+\mu^-\gamma$ decays anticipate that there is still not enough experimental sensitivity to claim a discovery if their branching fractions are Standard Model-like.

Table 6.6: Normalization factors for each signal channel and BDT > 0.25 resulting from the two normalization channels and the combined value. The values are given for Run 1, Run 2 and the full dataset. The efficiency derived from the exclusion of the first BDT bin is included. The expected yields obtained assuming the Standard Model branching fraction for the signal decays [1] are also reported. The uncertainties represent the statistical and systematic uncertainties.

Run 1			
	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-\gamma$
α_{B^+}	$(3.96 \pm 0.13 \pm 0.09) \times 10^{-11}$	$(1.57 \pm 0.07 \pm 0.03) \times 10^{-10}$	$(2.11 \pm 0.10 \pm 0.05) \times 10^{-10}$
α_{B^0}	$(3.79 \pm 0.14 \pm 0.16) \times 10^{-11}$	$(1.50 \pm 0.07 \pm 0.06) \times 10^{-10}$	$(2.01 \pm 0.10 \pm 0.09) \times 10^{-10}$
α_{comb}	$(3.93 \pm 0.10 \pm 0.08) \times 10^{-11}$	$(1.56 \pm 0.06 \pm 0.03) \times 10^{-10}$	$(2.09 \pm 0.09 \pm 0.05) \times 10^{-10}$
N_{exp}	$2.62 \pm 0.14 \pm 0.05$	$23.5 \pm 1.3 \pm 0.4$	$0.479 \pm 0.020 \pm 0.011$
Run 2			
	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-\gamma$
α_{B^+}	$(1.214 \pm 0.037 \pm 0.018) \times 10^{-11}$	$(4.54 \pm 0.20 \pm 0.07) \times 10^{-11}$	$(5.86 \pm 0.26 \pm 0.10) \times 10^{-11}$
α_{B^0}	$(1.176 \pm 0.035 \pm 0.037) \times 10^{-11}$	$(4.40 \pm 0.19 \pm 0.14) \times 10^{-11}$	$(5.67 \pm 0.25 \pm 0.18) \times 10^{-11}$
α_{comb}	$(1.204 \pm 0.023 \pm 0.014) \times 10^{-11}$	$(4.50 \pm 0.16 \pm 0.05) \times 10^{-11}$	$(5.81 \pm 0.21 \pm 0.08) \times 10^{-11}$
N_{exp}	$8.55 \pm 0.45 \pm 0.10$	$81.3 \pm 4.3 \pm 0.9$	$1.721 \pm 0.063 \pm 0.023$
All			
	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-\gamma$
α_{B^+}	$(9.27 \pm 0.28 \pm 0.12) \times 10^{-12}$	$(3.53 \pm 0.15 \pm 0.04) \times 10^{-11}$	$(4.61 \pm 0.20 \pm 0.07) \times 10^{-11}$
α_{B^0}	$(8.95 \pm 0.26 \pm 0.23) \times 10^{-12}$	$(3.41 \pm 0.15 \pm 0.09) \times 10^{-11}$	$(4.45 \pm 0.19 \pm 0.12) \times 10^{-11}$
α_{comb}	$(9.20 \pm 0.14 \pm 0.09) \times 10^{-12}$	$(3.51 \pm 0.12 \pm 0.03) \times 10^{-11}$	$(4.57 \pm 0.16 \pm 0.05) \times 10^{-11}$
N_{exp}	$11.20 \pm 0.57 \pm 0.11$	$104.4 \pm 5.4 \pm 1.0$	$2.186 \pm 0.077 \pm 0.026$

7 | Branching fraction measurement

This chapter presents the extended maximum likelihood fits to the dimuon invariant mass spectrum of the selected candidates resulting in three branching fraction measurements: $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and the initial-state radiation $B_s^0 \to \mu^+\mu^-\gamma$. The concepts of likelihood and maximum likelihood estimator are first introduced in Sec. 7.1, while the fit results are presented in Sec. 7.2. For each signal channel, the significance of the branching fraction measurement determined from the fit is obtained following the approach presented in Sec. 7.3. Since the significances associated with the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays are lower than the minimum required to claim a discovery, an upper limit is set to their branching fractions using the CL_s method, described in Sec. 7.4, resulting in the upper limits given in Sec. 7.5.

7.1 Maximum likelihood estimators

Consider an hypothesis $\mathcal{H}$ described by a set of model parameters $\boldsymbol{\theta} = \{\theta_j, ...\theta_m\}$ and a set of measurements $\mathbf{x} = \{x_i, ...x_n\}$ distributed according a probability density function (p.d.f.) $f(\mathbf{x}|\boldsymbol{\theta})$. Assuming $\mathcal{H}$, the *likelihood function* represents how well the parameters $\boldsymbol{\theta}$ describe the observed measurements $\mathbf{x}$. Supposing that the measurements are independent and described by the same probability density function $f(x_i|\boldsymbol{\theta})$, the likelihood function is defined as the product of the probability density functions of every single measurement [1, 76]:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{7.1}$$

Generally, one deals with *composite hypotheses*, where the shape of the p.d.f. $f(x_i|\boldsymbol{\theta})$ is known but not the value of all the $\boldsymbol{\theta}$ parameters describing it [76]. In these cases, the values of the unknown parameters can be estimated given the observed data using the *maximum likelihood estimator*[1] *(MLE)* method. From the definition of the likelihood, it is clear that the higher its value, the more probable the composite hypothesis is correct. Keeping this in

[1] An estimator is a function to evaluate a quantity or parameter of the p.d.f as a function of the observed data. [76].

mind, the $\boldsymbol{\theta}$ parameters that best describe the data are given by:

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_j} = 0, \text{ where } j = 1, ..., m \tag{7.2}$$

From the construction of the likelihood as a product of probability density functions, it is mathematically more convenient to maximize $\ln L(\boldsymbol{\theta})$ instead of the likelihood function itself. The use of the logarithm allows for each candidate to be a separate term in the $\ln L(\boldsymbol{\theta})$ as well as reducing the large numerical values associated with the $L(\boldsymbol{\theta})$. Using $\ln L(\boldsymbol{\theta})$, the maximum likelihood estimators give the parameter of the p.d.f. best describing the data as

$$\frac{\partial \ln(L(\boldsymbol{\theta}))}{\partial \theta_j} = 0, \text{ where } j = 1, ..., m \tag{7.3}$$

The number of observed measurements n, which in this analysis corresponds to the number of events, follows a Poisson distribution described by the average expected event rate, υ. To include this realization, the likelihood is smeared by adding the corresponding Poisson probability density function as an overall normalization of the likelihood function in Eq. 7.1. This addition results in the so-called *extended likelihood function* [76]

$$L(\upsilon, \boldsymbol{\theta}) = \frac{\upsilon^n}{n!} e^{-\upsilon} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{7.4}$$

where $f(x_i|\boldsymbol{\theta})$ are the p.d.f. of each independent measurement, n is the number of measurements and υ the estimated true yield, affecting the normalization of the likelihood function.

It is often the case that only a few of the $\boldsymbol{\theta}$ parameters describing the hypothesis $\mathcal{H}$ are of interest in an analysis, while the rest of the parameters, known as *nuisance parameters*, are needed to define the likelihood function properly [76]. Although not of intrinsic interest, the correct estimation of the nuisance parameters is essential as they might bias the predictions of the parameters of interest. Typically, the nuisance parameter estimation is done using different methods, from calibration samples to the use of control channels, or simply fitted in the final fit. A summary of the parameters describing the likelihood function used in the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis and the calibration methods followed in their estimation can be found in Table 7.1.

7.2 $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ branching fraction ML fit

The branching fractions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays are obtained from unbinned extended maximum likelihood fits to the dimuon invariant mass spectrum of the selected candidates in the data samples. The dataset comprises the data gathered by LHCb during the Run 1 and Run 2 data-taking periods. In order to maximize the sensitivity of the analysis, the en-

Table 7.1: Summary of parameters describing the likelihood function for the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis and the methods used in the estimation of the nuisance parameters.

Parameters θ_i	Type of parameter	Calibration
$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-(\gamma))$	Parameter of interest	-
Signal mass shape	Nuisance parameter	Mass calibration (sec. 6.1)
Fraction of signal events per BDT bin	Nuisance parameter	BDT calibration (sec. 5)
Normalization factors (including efficiencies)	Nuisance parameters	Normalization (sec. 6.3)
Peaking background yields	Nuisance parameters	Peaking backgrounds (sec. 6.2)
Peaking background mass shapes	Nuisance parameters	Fits to simulation (sec. 6.2)
Combinatorial background yields	Nuisance parameter	Free parameters in mass fit
Combinatorial background mass shape	Nuisance parameters	Free parameters in mass fit

tire dataset is divided into two subsamples for each data-taking period, Run 1 and Run 2, which are further split into six bins of the BDT classifier. As mentioned in Sec. 4.5, the BDT binning used is optimized to maximize the signal sensitivity while keeping the background manageable. The fits are performed simultaneously in all the data-taking periods and BDT bin subsets. The difference in luminosity for each data-taking period is accounted for by weighting each data-taking subset with the correct luminosity fraction. A similar procedure is followed for each BDT bin subset using the BDT calibration results in Chapter 5. The first BDT bin, corresponding to BDT $\in [0, 0.25)$, is not included in the fit because the combinatorial background dominates it. However, the fraction of events expected in this first BDT bin is included in the normalization factor as an efficiency correction, as described in Sec. 6.3.

The signal and background components are modelled independently in the fit using different probability density functions. The $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ components are mod-

elled with two independent double-sided Crystal Ball functions (see Appendix A.1) whose characterizing parameters are allowed to vary following a Gaussian with a specific mean and width defined by the mass calibration procedure in Sec. 6.1. The $B_s^0 \to \mu^+\mu^-\gamma$ component only includes initial-state radiation (see Chapter 1) and it is modelled using an empirical function given by

$$f(m_{\mu^+\mu^-}) \propto \left(1 - \frac{m_{\mu^+\mu^-}}{M_{B_s^0}}\right)^b - a\sqrt{1 - e^{\frac{m_{\mu^+\mu^-} - M_{B_s^0}}{s}}} \tag{7.5}$$

where the parameters b, a and s are determined from fits to dedicated simulation samples obtained based on the theoretical calculations from Refs. [77, 78]. The b parameter is found to be 0.5, while different values of the a and s parameters are obtained for different BDT bins. The detector resolution effects are included by convolving the empirical shape in Eq. 7.5 with a Gaussian function whose parameters are obtained from weighted simulation samples. All the parameters describing the $B_s^0 \to \mu^+\mu^-\gamma$ mass component are fixed in the fit to the data except the yield. The other two nuisance parameters associated with the signal channels, namely the normalization factors (see Sec. 6.3) and the expected fractions of signal events ending up in each BDT bin (see Chapter 5), are also Gaussian constrained in the fit.

Regarding the backgrounds, the combinatorial component is described with an exponential function whose slope is left free in the fit and shared among all the data-taking and BDT subsets. The shapes of the physical backgrounds are obtained from simulation samples using Gaussian kernel density estimators [62]. While the yields of the combinatorial background are left free in the fit, the yields of each physical background component are Gaussian constrained to the expected yields evaluated in Sec. 6.2.

The mass distributions of the signal candidates are shown in Fig. 7.1 and Fig. 7.2 for Run 1 and Run 2 and the corresponding BDT bins. The total fit and its separated components are also displayed. The branching fractions for the signal channels are obtained from these fits and found to be

$$\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9} \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.02^{+0.83}_{-0.74} \pm 0.14) \times 10^{-10} \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma; m_{\mu^+\mu^-} \geq 4.9\,\text{GeV}/c^2) &= (-2.5 \pm 1.4 \pm 0.8) \times 10^{-9}
\end{aligned}$$

where the statistical and systematic uncertainties are reported separately. It is important to note that the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction only includes initial-state radiation in the dimuon invariant mass window given by $m_{\mu^+\mu^-} \geq 4900\,\text{GeV}/c$. The negative result obtained is compatible with the very low sensitivity for that decay and suggests an under-fluctuation of the background on the low dimuon invariant mass region. As previously mentioned, both the statistical and systematic uncertainties of each nuisance parameter are taken into account by Gaussian constraining their values to the range given by their uncertainties.

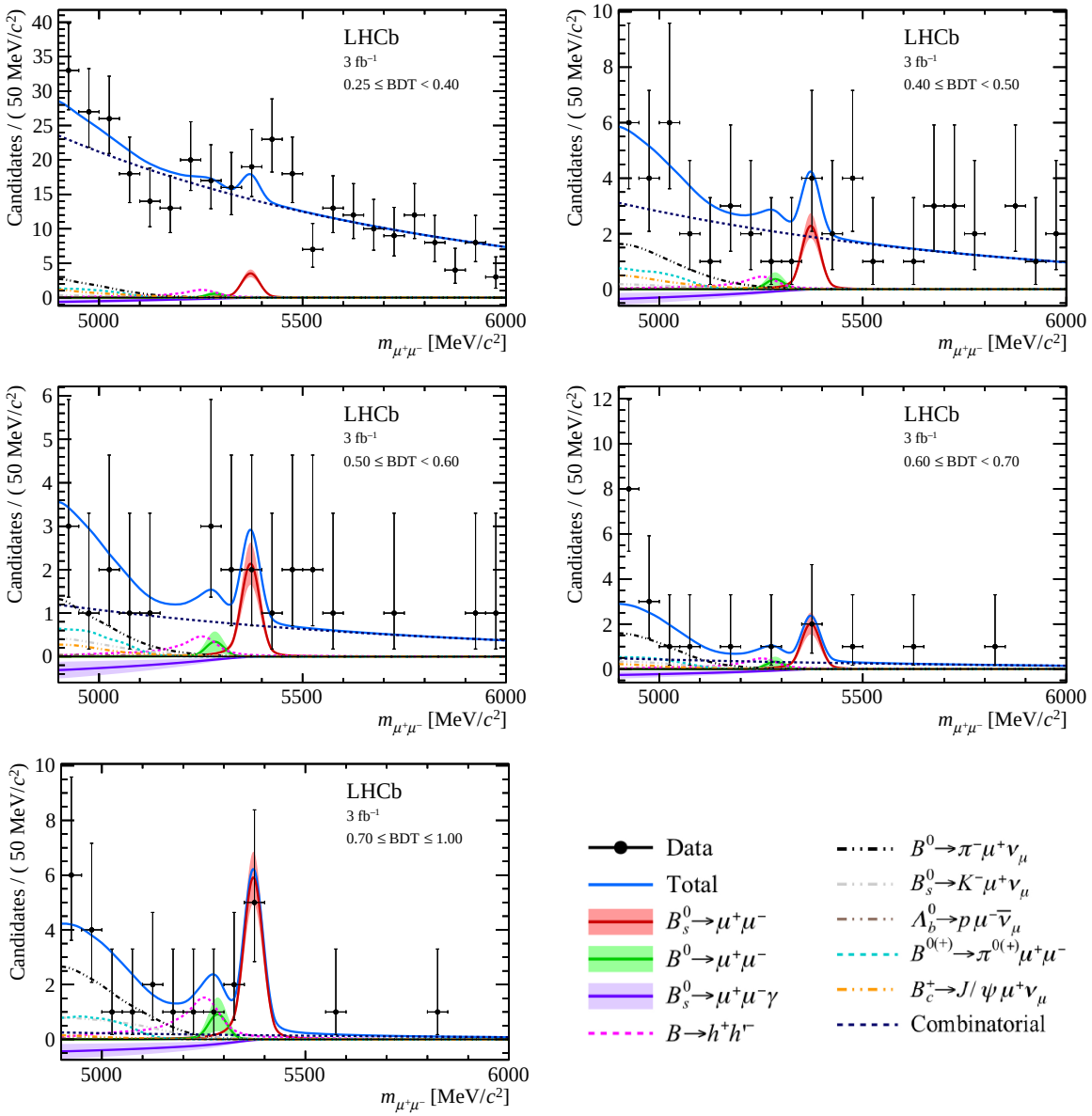


Figure 7.1: Mass distributions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ selected candidates for Run 1 for each different BDT bin. The blue line corresponds to the total fit model, and the separated components are reported in the legend. Solid bands are shown for the distributions of the signal channels indicating the total uncertainties.

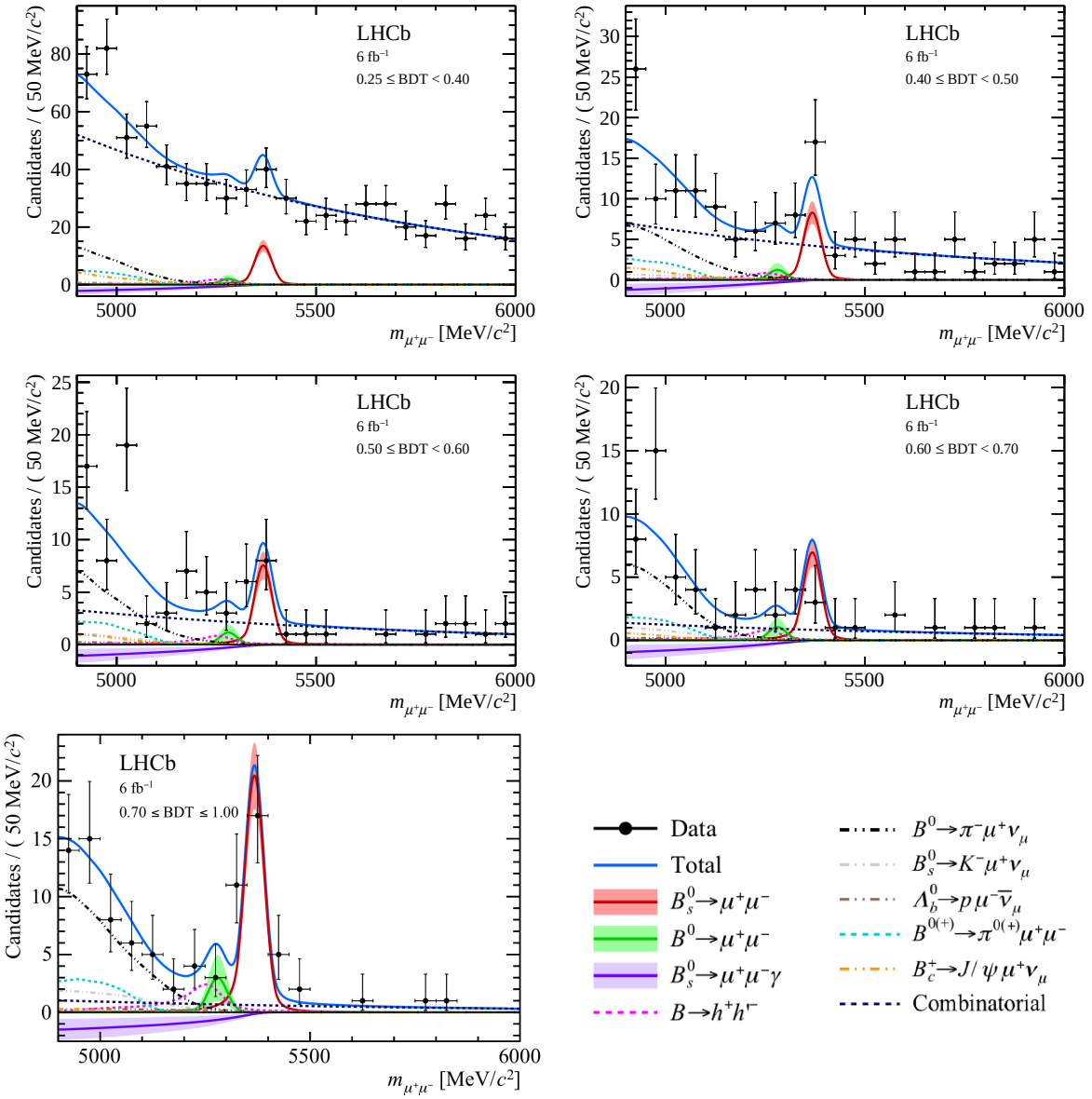


Figure 7.2: Mass distributions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ selected candidates for Run 2 for each different BDT bin. The blue line corresponds to the total fit model, and the separated components are reported in the legend. Solid bands are shown for the distributions of the signal channels indicating the total uncertainties.

As a consequence, the fit only returns the total uncertainty without isolating the statistical and systematic uncertainty associated with each branching fraction value. To separate the statistical component, the fit is repeated by fixing all the nuisance parameters to the values obtained from the nominal fit. The systematic uncertainties are then extracted by subtracting in quadrature the statistical uncertainties from the total ones in the nominal fit. The primary sources of systematic uncertainty in the branching fractions measurement arise from the f_s/f_d ratio and the estimation of the exclusive backgrounds yields, mainly the misidentified $B^0_{(s)} \to h^+h'^-$ decays due to their significant overlap with the $B^0 \to \mu^+\mu^-$ component.

7.3 Significance of a measurement

When particle physicists search for new phenomena, such as the $B^0_s \to \mu^+\mu^-\gamma$ or the $B^0 \to \mu^+\mu^-$ processes, two hypotheses are tested and compared: the hypothesis assuming that the data is consistent with the presence of only background, also known as the null hypothesis or H_0; and the hypothesis including both signal and background, referred to as H_1. The possible rejection of the null hypothesis in favour of the H_1 is done by measuring the discrepancy between the observed data and the null hypothesis with a *test statistic*[2]. The significance of such discrepancy is quantified by the *p-value* of H_0, giving the probability that the data obtained under the H_0 hypothesis is equally or less compatible than the observed data [76]. Considering a null hypothesis consisting of $\boldsymbol{\theta}$ parameters of interest and $\boldsymbol{\nu}$ nuisance parameters, the one-sided p-value for the parameters of interest $\boldsymbol{\theta}$ is [1]

$$p_{\boldsymbol{\theta}}(\boldsymbol{\nu}) = \int_{q_{\boldsymbol{\theta},obs}}^{\infty} f(q_{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\nu}) dq_{\boldsymbol{\theta}} \tag{7.6}$$

where $f(q_{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\nu})$ is the probability density distribution of a test statistic $q_{\boldsymbol{\theta}}$ and $q_{\boldsymbol{\theta},obs}$ is the value of the test statistic observed in the data.

The H_0 hypothesis is rejected if the p-value is less than α, which is the probability of rejecting the null hypothesis when it is true, also known as a *type-I error* [76]. Because the p-value in Eq. 7.6 depends on the nuisance parameters, the H_0 can only be rejected if the p-value is smaller than α for all the possible values of the nuisance parameters.

To remove this dependence of the p-value on the nuisance parameters, the test statistic $q_{\boldsymbol{\theta}}$ can be defined such that the p.d.f. $f(q_{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\nu})$ is independent of the nuisance parameters, i.e. $f(q_{\boldsymbol{\theta}}|\boldsymbol{\theta})$. One method to do so, used in this analysis, is to profile out the nuisance parameters and express them as a function of the parameters of interest to define the *profile likelihood* function [1] as

$$L_p(\boldsymbol{\theta}) = L(\boldsymbol{\theta}, \hat{\boldsymbol{\nu}}(\boldsymbol{\theta})) \tag{7.7}$$

[2] A test statistic is a quantity derived from data as a function of the measured variables that, together with its value for a given hypothesis, quantifies the compatibility of the observed data and the hypothesis tested.

where $\hat{\boldsymbol{\nu}}(\boldsymbol{\theta})$ are the nuisance parameters maximizing the likelihood for a given value of $\boldsymbol{\theta}$. A test statistic independent of the nuisance parameters is then the *profile likelihood ratio* defined as

$$\lambda_p(\boldsymbol{\theta}) = \frac{L_p(\boldsymbol{\theta})}{L_p(\hat{\boldsymbol{\theta}})} = \frac{L(\boldsymbol{\theta}, \hat{\boldsymbol{\nu}}(\boldsymbol{\theta}))}{L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\nu}}(\hat{\boldsymbol{\theta}}))} \tag{7.8}$$

with $\hat{\boldsymbol{\theta}}$ as the maximum likelihood estimator for the parameters of interest. While the likelihood in the denominator is maximized unconditionally, the likelihood in the numerator is given for the nuisance parameters that maximize the likelihood for a particular value of $\boldsymbol{\theta}$, assuming the null hypothesis H_0 [1, 79].

From its definition, the profile likelihood ratio ranges from 0 to 1, with higher values implying higher compatibility of the observed data with the null hypothesis H_0, i.e. $\lambda_p(\boldsymbol{\theta}) = 1$ if $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. To sculpt a profile of the likelihood around the maximum, the associated test statistic used is

$$t_{\boldsymbol{\theta}} = -2 \ln \lambda_p(\boldsymbol{\theta}) \tag{7.9}$$

where higher values of $t_{\boldsymbol{\theta}}$ correspond to higher incompatibility between the observed data and the hypothesized values of $\boldsymbol{\theta}$ [79]. Notice that $t_{\boldsymbol{\theta}}$ is not dependent on the nuisance parameters since they have been profiled out by expressing them as functions of the parameters of interest. Similarly, the probability density function of the test statistic can be written as $f(t_{\boldsymbol{\theta}}|\boldsymbol{\theta}) = f(t_{\boldsymbol{\theta}}|\boldsymbol{\theta}, \hat{\boldsymbol{\nu}}(\boldsymbol{\theta}))$. As a consequence, the evaluation of the p-value using Eq. 7.6 becomes:

$$p_{\boldsymbol{\theta}} = \int_{t_{\boldsymbol{\theta},obs}}^{\infty} f(t_{\boldsymbol{\theta}}|\boldsymbol{\theta}) dt_{\boldsymbol{\theta}} \tag{7.10}$$

Eq. 7.10 is defined for a set of parameter of interest $\boldsymbol{\theta} = (\theta_i, ...\theta_m)$. In the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis, there are three parameters of interest: the branching fractions of the $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ signal channels. They are treated independently, such that for any null hypothesis of one of the signal components, the others are treated as nuisance. Therefore, for each null hypothesis tested, there is only one parameter of interest $\boldsymbol{\theta} = \mu$, corresponding to the branching fractions of one signal decay. In these cases, the analytical behaviour of the test statistic in Eq. 7.9 in the large sample limit is known analytically using Wald's theorem [80] and can be approximated as [79]

$$t_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2} \tag{7.11}$$

where $\hat{\mu}$ and σ are the mean and the standard deviations of the estimated branching fraction obtained from the maximum likelihood fit. Moreover, Wilk's theorem states that the distribution of the test statistic t_μ in the large sample limit approaches a χ^2 distribution [81]. Therefore, the p-values can be adequately defined as the area under the χ^2 distribution.

Once the t_μ value is defined for a given hypothesis, the significance of the measurement

is determined with the p-value using Eq. 7.10. Although the p-value is already enough to decide whether to reject the null hypothesis, the significance in particle physics is typically given in terms of the *Z-score*. Interpreting the p-value in terms of a normal distribution, the Z-score is obtained as the number of standard deviations between the mean of the Gaussian of the parameter of interest, given by the hypothesis tested, and the measured value [1, 76]; as shown in Fig. 7.3. The Z-score is calculated from the p-value using the inverse of the cumulative Gaussian distribution Φ^{-1} as

$$Z = \Phi^{-1}(1 - p) \tag{7.12}$$

Both the p-value and the Z-score are evaluated under the assumption of the null hypothesis (background-only). The level of significance required to claim a discovery is set to a Z-score of five, corresponding to a p-value of 2.87×10^{-7}.

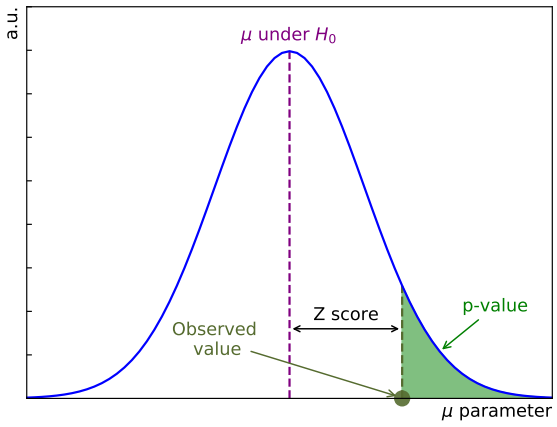


Figure 7.3: Example of a Gaussian distribution illustrating the relation between the p-value and the Z-score given an observed value of the parameter of interested μ relative to the value of μ under the null hypothesis.

The significance of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signals relative to the background-only hypothesis are obtained using the test statistic in Eq. 7.11. The significance for $B_s^0 \to \mu^+\mu^-$ decays is found to be 10σ, while for $B^0 \to \mu^+\mu^-$, it is lower than the minimum 5σ required to claim a discovery, with a value of 1.7σ.

7.4 Limit setting with the CL_s technique

The statistical significance of a particular signal process measurement is sometimes lower than the 5σ required to claim a discovery. In such cases, an upper limit for the signal branching fraction is reported using *confidence intervals* [76]. Considering the measurement is repeated multiple times, the confidence interval is the range of values the parameter of interest

will fall into in a fraction of the experiments given by the *confidence level*. In practice, the experiments are repeated using so-called "Montecarlo toys" or "pseudo"-experiments.

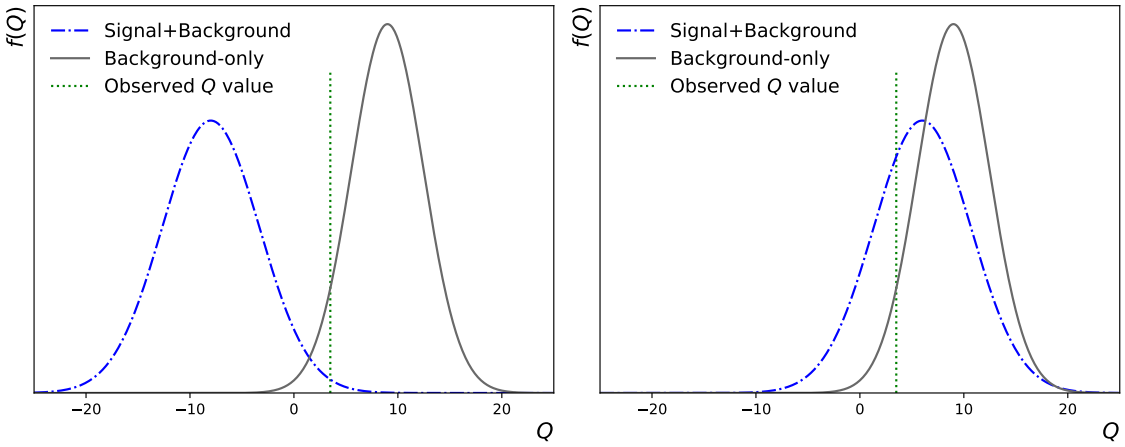


Figure 7.4: Examples of the Q distributions for the signal+background and the background-only hypothesis when they are well-separated (left) and when they are almost identical (right). The lower the branching fraction tested, the more similar the two distributions are. In the case of the overlap, excluding the hypothesized branching fraction value also implies an exclusion of the background-only hypothesis.

A one-sided confidence interval is used to set an upper limit of the branching fraction for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays. The goal is to find a value of the parameter of interest μ assuming the signal and background hypothesis such that it is rejected in favour of the background-only hypothesis with a certain confidence level. The confidence level is $1-\alpha$, where α is the type-I error rate or the probability of rejecting the null hypothesis when it is true. The values outside the confidence limit are assumed to be excluded with a $1-\alpha$ confidence level, and the signal+background hypothesis is rejected in those cases.

The test statistic adopted to set the confidence interval is given by

$$Q = -2\ln\left(\frac{L(s+b|\mathbf{x})}{L(b|\mathbf{x})}\right) \tag{7.13}$$

where $L(s+b|\mathbf{x})$ and $L(b|\mathbf{x})$ are the likelihood function for the signal+background and the background-only hypotheses, respectively. This test statistic is more convenient to use in the limit setting since it allows for a direct comparison of the strength of one hypothesis relative to the other. Using this test statistic, the idea is to find a branching fraction value such that the associated p-value, obtained using Eq. 7.10, is smaller or equal to α.

In order to find the upper limit, different branching fraction hypotheses are tested. The probability density function of the test statistic for the signal+background hypothesis must be known for each branching fraction value. This probability density function is constructed

from a set of simulated pseudo-experiments generated assuming the tested branching fraction value under the signal+background hypothesis. A similar procedure can be followed to construct the probability density function associated with the background-only hypothesis.

It is important to note that the distributions of the test statistic vary according to the branching fraction hypothesis, as can be deduced from Eq. 7.13, such that the lower the branching fraction tested, the more similar the signal+background and background-only distributions are. To illustrate the consequences of this effect, Fig. 7.4 shows two examples, one where the two distributions are well-separated (left) and the case where they are almost identical (right). In the latter case, excluding a value assuming the signal+background hypothesis will also exclude the background-only hypothesis, leading to an over-exclusion of low branching fraction values for which the experimental sensitivity is not enough.

An alternative to mitigate over-exclusion is the CL_s *method* [82,83]. With this approach, an *adjusted* confidence level is obtained, which takes into account the overlap between the two hypotheses

$$\mathrm{CL_s} = \frac{p_{s+b}}{1 - p_b} \tag{7.14}$$

where p_{s+b} and p_b are the p-values of the signal+background and background-only hypotheses, respectively. Assuming a specific value of the branching fraction for the signal+background hypothesis, these values are represented in Fig. 7.5 and are evaluated with the test statistic Q as

$$p_{s+b} = \int_{Q_{\mathrm{obs}}}^{\infty} f_{b+s}(Q)dQ \tag{7.15}$$

$$p_b = \int_{-\infty}^{Q_{\mathrm{obs}}} f_b(Q)dQ \tag{7.16}$$

where $f_{b+s}(Q)$ and $f_b(Q)$ are the probability density functions of the signal+background and background-only hypotheses and Q is the test statistic as defined in Eq. 7.13.

The confidence interval is determined using $\mathrm{CL_s} = \alpha$ by setting the confidence level to $1 - \mathrm{CL_s}$ [82, 83]. The probability density function of the Q test statistic is obtained from pseudo-experiments generated using the corresponding hypothesis.

7.5 Limit setting for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ branching fractions

Due to the low significance of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays, an upper limit is set to their branching fractions values using the $\mathrm{CL_s}$ method as implemented in the GammaCombo framework [84, 85]. Within this framework, a $\mathrm{CL_s}$ scan is performed for different values of

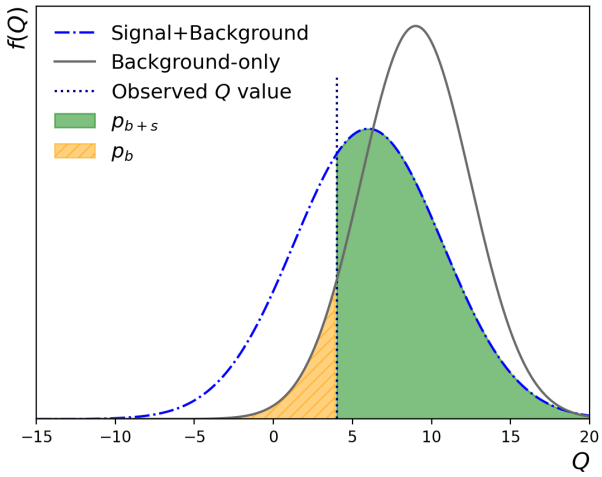


Figure 7.5: Illustration showing the p_b and p_{b+s} for the background-only and background+signal hypotheses given a observed value of Q.

the branching fractions to obtain two later compared curves: the expected and the observed curves. The expected or background-only curve is obtained by generating several samples using the background-only model and fitting them back using the signal+background model with the branching fraction fixed at a specific value. The CL_s is then calculated for this fixed value of the branching fractions as described in Sec. 7.4. Since several samples are fitted back with the same branching fraction, not all of them yield the same CL_s value. The 1σ (2σ) error bands associated to the expected curve are then defined as the vertical CL_s intervals where 16% (2.5%) of the fitted CL_s fall above and 16% (2.5%) below the boundaries. Regarding the observed curve, the CL_s values are determined from fits to the data sample by scanning the likelihood for various values of the branching fraction.

The resulting CL_s curves for $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays are shown in Fig. 7.6 from which the limits on their branching fractions are obtained as

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.3(2.6) \times 10^{-10} \text{ at } 90\%(95\%) \text{ CL}$$
$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)|_{m_{\mu^+\mu^-} \geq 4.9\,\text{GeV}/c^2} < 1.5(2.0) \times 10^{-9} \text{ at } 90\%(95\%) \text{ CL}$$

The results obtained for the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are in agreement with the Standard Model predictions of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.66 \pm 0.14) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10}$ [9]. This agreement is also visible in the two-dimensional confidence interval set for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$ in Fig. 7.7, where the Standard Model prediction is also represented together with the contours corresponding to the previous LHCb measurement [49]. The two-dimensional contours obtained for $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ versus $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are also presented in Fig. 7.7.

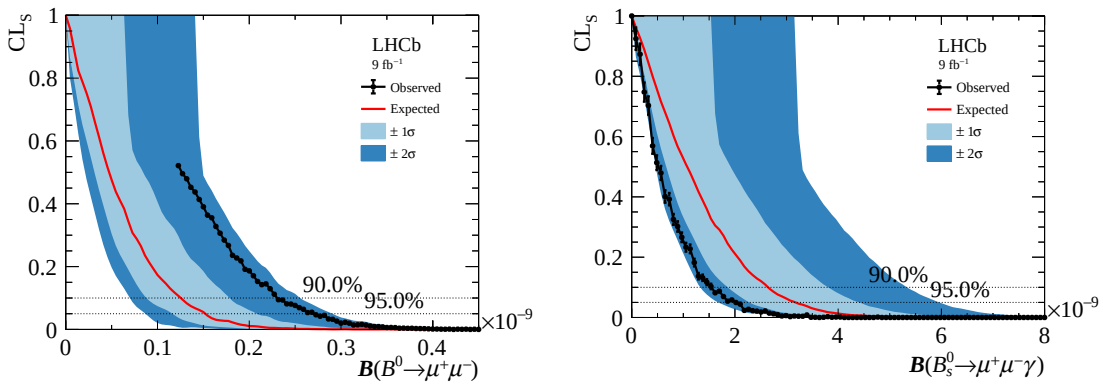


Figure 7.6: CL$_s$ curves for branching fractions of the $B^0 \to \mu^+\mu^-$ (left) and $B_s^0 \to \mu^+\mu^-\gamma$ (right) decays. The observed values in data are shown in black. The expected background-only values are shown in red with the 1σ and 2σ bands around the values in light and black blue, respectively.

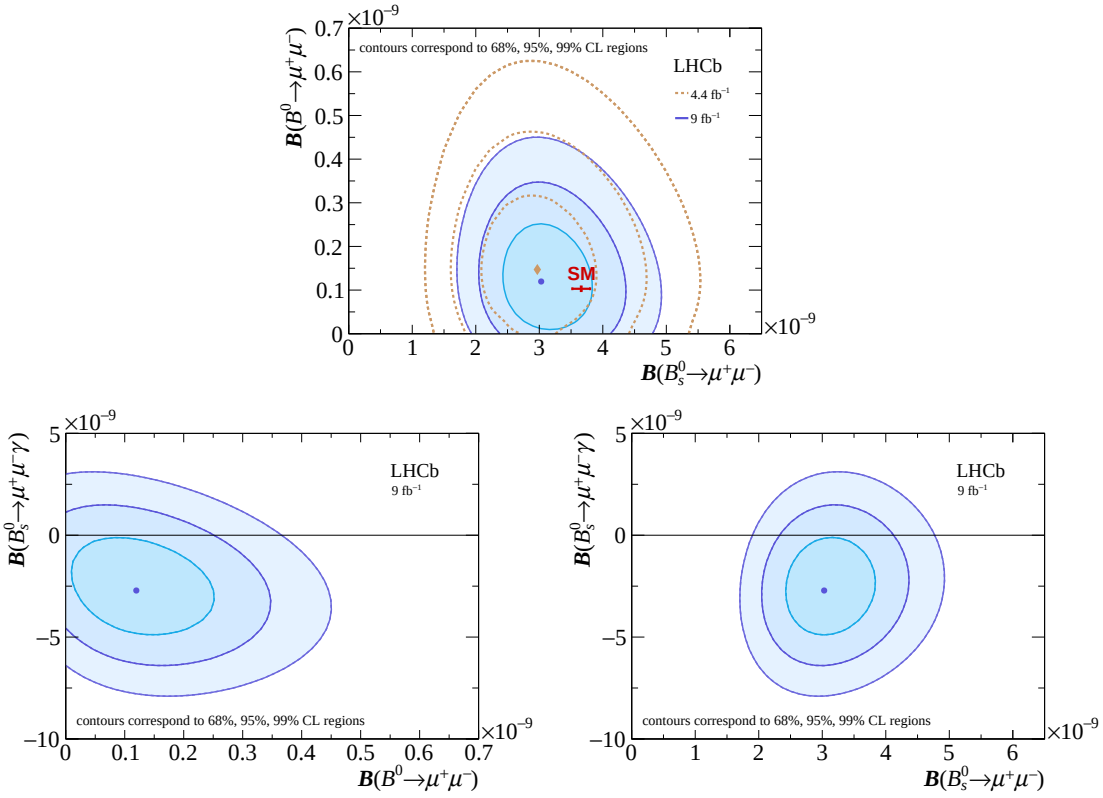


Figure 7.7: Two-dimensional confidence intervals of the branching fractions for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays (top) where the Standard Model predictions are shown in red, and the contours in brown indicate the results from the previous LHCb measurement [49]. The two-dimensional counters are also shown for the $B_s^0 \to \mu^+\mu^-\gamma$ and $B^0 \to \mu^+\mu^-$ decays (bottom left) and $B_s^0 \to \mu^+\mu^-\gamma$ and $B_s^0 \to \mu^+\mu^-$ decays (bottom right).

8 | Discussion and prospects

In this chapter, the results presented in this thesis are interpreted and compared to the Standard Model and to the latest and most recent CMS result in Sec. 8.1. The constraints to different New Physics scenarios obtained by combining the measured $B_s^0 \to \mu^+\mu^-$ branching fraction with other analyses are presented in Sec. 8.2. Although the $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ analysis is still statistically limited, several improvements, discussed in Sec. 8.3, can be performed in the future to increase the precision of the $B_{(s)}^0 \to \mu^+\mu^-(\gamma)$ branching fraction measurement and constrain New Physics models even further.

8.1 Interpretation of the results

This thesis presents the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and the search for $B^0 \to \mu^+\mu^-$ and initial-state radiation $B_s^0 \to \mu^+\mu^-\gamma$ decays using the entire dataset gathered by LHCb in Run 1 and Run 2. The dimuon mass distribution of the selected signal candidates is shown in Fig. 8.1 for BDT>0.5, together with each of the signal and background components overlaid.

The results obtained from the fit (see Chapter 7) for the $B_s^0 \to \mu^+\mu^-$ branching fraction and the upper limits for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ branching fractions are

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.3(2.6) \times 10^{-10} \text{ at } 90\%(95\%) \text{ CL} \tag{8.1}$$

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma; m_{\mu^+\mu^-} < 4.9\,\text{GeV}/c^2) \geq 1.5(2.0) \times 10^{-9} \text{ at } 90\%(95\%) \text{ CL} \tag{8.2}$$

Regarding the systematic uncertainties, the two most significant contributions arise from the f_s/f_d ratio and the $B_{(s)}^0 \to h^+h'^-$ misidentified backgrounds. The f_s/f_d ratio has a relative uncertainty of around 3%, the largest relative uncertainty in the fit. It only affects the $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ branching fractions since it is needed in their normalization procedure (see Chapter 6). The f_s/f_d ratio is not needed in the normalization for the $B^0 \to \mu^+\mu^-$ decays, leading to a smaller absolute error in its branching fraction. The latter is mainly affected by the $B_{(s)}^0 \to h^+h'^-$ misidentified background. Although the estimated yields of other

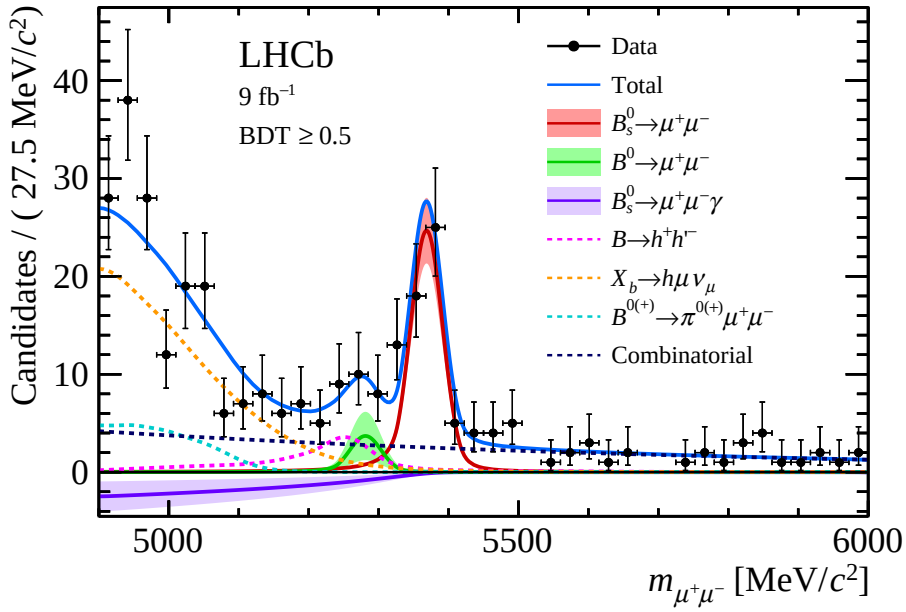


Figure 8.1: Mass distributions of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ selected candidates for BDT > 0.5. The blue line corresponds to the total fit model, and the separated components are reported in the legend. Solid bands are shown for the distributions of the signal channels indicating the total uncertainties.

backgrounds have larger uncertainties than the $B^0_{(s)} \to h^+h'^-$ background, its overlap with the $B^0 \to \mu^+\mu^-$ shape directly impacts the systematic uncertainty for the branching fraction value obtained from the fit.

The results from this analysis are compared with the Standard Model predictions of

$$\begin{aligned} \mathcal{B}(B^0_s \to \mu^+\mu^-) &= (3.66 \pm 0.14) \times 10^{-9} \\ \mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.03 \pm 0.05) \times 10^{-10} \end{aligned} \tag{8.3}$$

given in Chapter 1. Although a slightly lower central value of the $B^0_s \to \mu^+\mu^-$ branching fraction is measured, the measurement is found to agree with the Standard Model at the level of 1.2σ. The experimental sensitivity is not enough to reach the Standard Model prediction of the $B^0 \to \mu^+\mu^-$ branching fraction, and the upper limit obtained is also found to agree with the Standard Model prediction.

The latest measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions has recently been reported by the CMS collaboration just before the completion of this thesis [86]. In this analysis, the full Run 1 and Run 2 datasets gathered by the CMS collaboration are used. The

results reported are:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.95^{+0.39+0.27+0.21}_{-0.37-0.22-0.19}) \times 10^{-9} \tag{8.4}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.5(1.9) \times 10^{-10} \text{ at } 90\%(95\%) \text{ CL} \tag{8.5}$$

where for the $B_s^0 \to \mu^+\mu^-$ branching fraction, the first uncertainty is statistical, the second systematic and the third due to the external branching fractions used as inputs. Compared with the analysis presented in this thesis, the $B_s^0 \to \mu^+\mu^-$ branching fraction measured by CMS is 1.2σ higher [86]. Since both measurements are on opposite sides relative to the Standard Model value, the world average is shifted to a value closer to the Standard Model prediction. Although the evaluation of such average requires a proper combination of the CMS, LHCb and ATLAS results based on their likelihoods, a naive approximation is obtained combining the LHCb and latest CMS $B_s^0 \to \mu^+\mu^-$ branching fraction measurements for illustrative purposes. This naive combination is obtained as a weighted average using the inverse of the statistical error as the quantity related to the luminosity of each experiment. The average obtained is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 3.63^{+0.41+0.30}_{-0.39-0.24}$, where the f_s/f_d ratio has been subtracted from the systematic uncertainties of both measurements in order to combine them assuming uncorrelated uncertainties and added after the combination to the final systematic uncertainty. As can be seen, the naive average is in perfect agreement with the Standard Model prediction.

8.2 Constraint on New Physics models

As presented in Chapter 1, the $B_{(s)}^0 \to \mu^+\mu^-$ branching fraction defined in a model-independent way is only sensitive to contributions from the axial vector lepton currents ($C_{10}^{(\prime)}$), the scalar ($C_S^{(\prime)}$) and pseudoscalar ($C_P^{(\prime)}$) lepton currents. In the Standard Model, the only non-negligible contribution comes from the C_{10} Wilson coefficient. Since the theoretical computation of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fraction is very clean, the branching fraction values predicted in the Standard Model are very precise. Therefore, $B_{(s)}^0 \to \mu^+\mu^-$ decays are good laboratories to search for New Physics phenomena appearing as a contribution to the $C_{10}^{(\prime)}$, $C_S^{(\prime)}$ and $C_P^{(\prime)}$ Wilson coefficients.

In order to search for New Physics phenomena in a broader picture, the measured $B_s^0 \to \mu^+\mu^-$ branching fraction is combined with other measurements exhibiting deviations from the Standard Model predictions. These values are combined in the so-called global fits, which parametrize New Physics contributions in the form of Wilson coefficients using the effective field approach in Sec. 1.2. The results from the global fits presented in this section are obtained from Ref. [87]. It is important to note that the $B_s^0 \to \mu^+\mu^-$ branching fraction considered in these fits is the combination of the LHCb measurement presented in this thesis with the previous CMS result [88] and the latest ATLAS results [89]. Together with the $B_s^0 \to \mu^+\mu^-$ branching fraction, the measurement included in the global fits are the branch-

ing fractions of the $B \to K^{(*)}\mu^+\mu^-$, $B_s^0 \to \phi\mu^+\mu^-$ and $\Lambda_b \to \Lambda\mu^+\mu^-$, jointly referred to as $b \to s\mu\mu$; and the lepton-flavour universality (LFU) ratios R_K and R_{K^*}. These measurements complement the $B_s^0 \to \mu^+\mu^-$ decay in the search for New Physics since they are sensitive to vector lepton current contributions, given by the $C_9^{(\prime)}$ Wilson coefficients, in addition to the $C_{10}^{(\prime)}$, $C_S^{(\prime)}$ and $C_P^{(\prime)}$ coefficients. The New Physics contributions are tested as deviations to the values of the Wilson coefficients given by Standard Model as $C_i = C_i^{SM} + \delta C_i$, where δC_i is the New Physics contribution.

Different strategies are followed in the global fits allowing for New Physics phenomena in one, two or several Wilson coefficients. In the simplest approach, where only a single New Physics coefficient δC_i is allowed at a time, two different cases show an improvement of more than 4.5σ in the agreement between data and theory relative to the Standard Model: the $\delta C_{10}^{bs\mu\mu} \approx +0.60$ and $\delta C_9^{bs\mu\mu} = -\delta C_{10}^{bs\mu\mu} \approx -0.35$. Considering only possible new physics contributions in the right-handed quark projection operator $\delta C_{10}^{\prime bs\mu\mu}$, the global fits show a preference for results compatible with zero, i.e. no New Physics contributions. Possible scalar and pseudoscalar New Physics contributions are tested with the constraint $\delta C_S = -\delta C_P$ and $\delta C_S' = \delta C_P'$, in both cases showing a slight preference for small negative values of the scalar contributions, with pulls of 2.1σ.

Two-dimensional scenarios are also considered where two New Physics coefficients are allowed simultaneously. Fig. 8.2 shows the New Physics $\delta C_9^{bs\mu\mu}$, $\delta C_{10}^{bs\mu\mu}$ plane with the constraints imposed by the $B_s^0 \to \mu^+\mu^-$ branching fraction (yellow), the R_K and R_{K^*} ratios (blue) and the $b \to \mu\mu$ transitions (orange). The fit result combining all the measurements is shown in red and labelled as "rare B decays". The global fit result seems to show a preference for a slightly positive $\delta C_{10}^{bs\mu\mu}$ and a negative $\delta C_9^{bs\mu\mu}$. Concretely, the best-fit point is given by $(\delta C_9^{bs\mu\mu}, \delta C_{10}^{bs\mu\mu}) \approx (-0.51, 0.30)$ with an improvement in the agreement between theory and data of 5.3 σ compared to the Standard Model. Although the $B_s^0 \to \mu^+\mu^-$ branching fraction measurement is not sensitive to $\delta C_9^{bs\mu\mu}$, its sensitivity to the $\delta C_{10}^{bs\mu\mu}$ is significant, resulting in the $B_s^0 \to \mu^+\mu^-$ branching fraction playing a fundamental role in the constraint of such New Physics contribution.

Finally, more generic strategies are followed where four or six New Physics contributions are allowed. The results from these fits show a clear preference for New Physics models appearing in $\delta C_9^{bs\mu\mu}$, such as contributions from vector leptoquark or Z' models [90]. It is important to note that the new CMS result [86] recently published has not yet been included in the $B_s^0 \to \mu^+\mu^-$ branching fraction average used in these global fits. As previously mentioned, this measurement will bring the $B_s^0 \to \mu^+\mu^-$ branching fraction world average closer to the Standard Model prediction, limiting further the room available for the New Physics arising from axial vector lepton current contributions.

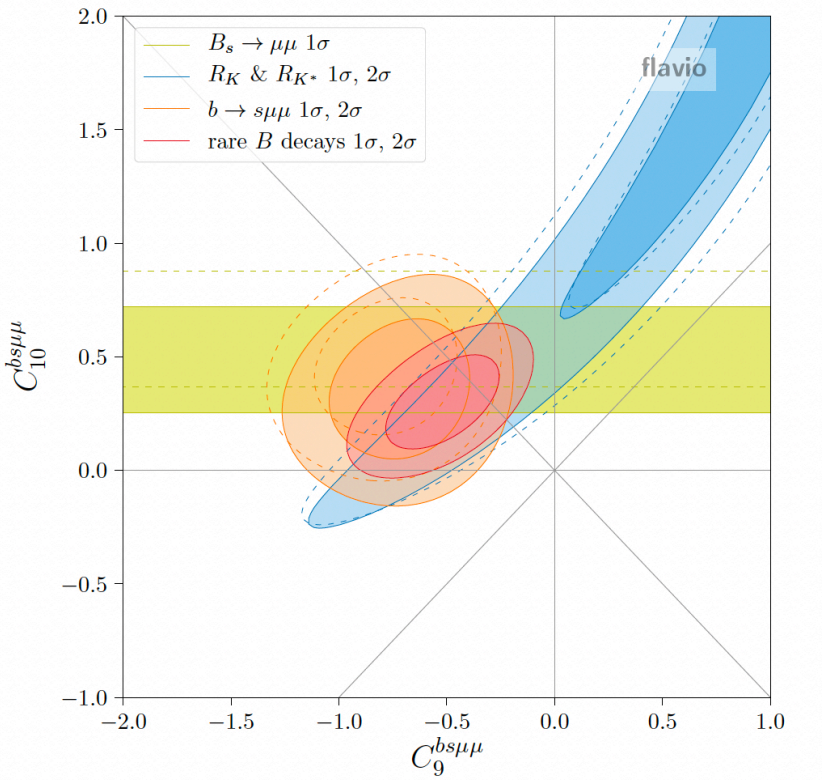


Figure 8.2: Constraints on New Physics models in the $\delta C_9^{bs\mu\mu} - \delta C_{10}^{bs\mu\mu}$ place. The combination of all results given by the global fits is shown in red. The 1σ and 2σ regions for each case are shown in dashed lines and the Standard Model value is located at (0,0). Figure taken from Ref. [87].

8.3 Future prospects

The anomalies between data and the Standard Model predictions observed in the $b \to sll$ transitions have been one of the hottest topics in particle physics in recent years. The $B_s^0 \to \mu^+\mu^-$ branching fraction measurement presented in this thesis is found to be consistent with the Standard Model, but it is still clearly statistically limited after Run 1 and Run 2. The Standard Model prediction is much more precise than the measured value, implying that there is still room for New Physics contributing to the $B_s^0 \to \mu^+\mu^-$ decay. With increasing data in the coming data-taking periods, the statistical uncertainty associated with the measured $B_s^0 \to \mu^+\mu^-$ branching fraction will be reduced. Combining these future results with other complementary measurements such as R_K or R_{K^*} will help to further constrain New Physics models.

With more data, the experimental sensitivity to the $B^0 \to \mu^+\mu^-$ will increase, allowing for tightening the upper limits on its branching fraction and, ultimately, claiming its discovery. The addition of the future $B^0 \to \mu^+\mu^-$ branching fraction measurement to the global fits

presented in the previous section will help to shed some light on the beyond the Standard Model physics, especially on the $\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ ratio. The overlap between the $B^0 \to \mu^+\mu^-$ invariant mass distribution with the $B^0_{(s)} \to h^+h'^-$ background will be significant, implying an extensive control of this background is required to identify the $B^0 \to \mu^+\mu^-$ candidates in the fit correctly. Moreover, the left-handed tail from the $B_s^0 \to \mu^+\mu^-$ decays, caused by the final-state radiation (see Sec. 1.3), also overlaps with the $B^0 \to \mu^+\mu^-$ shape. Currently, the number of possible $B_s^0 \to \mu^+\mu^-$ events mislabelled as $B^0 \to \mu^+\mu^-$ is negligible compared to the experimental sensitivity but, with the increased statistics of the Run 2 dataset, the experimental sensitivity will be large enough for this misidentification to become significant. At that moment, using a physical model of the final-state radiation tail, proposed in Sec. 8.3.1, will provide a more accurate description of the final-state radiation physics.

It is important to note that the analysis presented in this thesis has been optimized to measure the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for the $B^0 \to \mu^+\mu^-$ decay. Consequently, the search for initial-state radiation $B_s^0 \to \mu^+\mu^-\gamma$ decays has been constrained by the selection and reconstruction procedure used for $B^0_{(s)} \to \mu^+\mu^-$ decays. These constraints are mainly two: the narrow dimuon invariant mass window used to select the signal candidates ($m_{\mu^+\mu^-} \in [4900, 6000]\,\text{MeV}/c^2$); and the fact that the photon in the final state is not searched for in the reconstruction. A dedicated analysis for the $B_s^0 \to \mu^+\mu^-\gamma$ decay applying an optimized selection for such decay will increase the sensitivity in its search.

Together with the $B^0_{(s)} \to \mu^+\mu^-$ results, the search for the $B^0_{(s)} \to \tau^+\tau^-$ and $B^0_{(s)} \to e^+e^-$ decays will help to understand the nature of New Physics phenomena. The analysis of the $B^0_{(s)} \to \tau^+\tau^-$ decays helps in probing lepton flavour universality, an accidental symmetry in the Standard Model causing the electroweak coupling to be universal to all leptons. Although much less suppressed than the $B^0_{(s)} \to \mu^+\mu^-$ decays due to the larger mass of the τ lepton, the measurement of the $B^0_{(s)} \to \tau^+\tau^-$ decays is experimentally challenging due to the unstable τ particle, which must be reconstructed from different final states. The latest search for $B^0_{(s)} \to \tau^+\tau^-$ decays performed by the LHCb experiment using the Run 1 dataset set an upper limit to the branching fractions decays of $\mathcal{B}(B_s^0 \to \tau^+\tau^-) < 6.8 \times 10^{-3}$ and $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.1 \times 10^{-3}$ at 95% confidence level [91]. Contrary to $B^0_{(s)} \to \tau^+\tau^-$ decays, $B^0_{(s)} \to e^+e^-$ decays are strongly helicity suppressed in the Standard Model due to the lower mass of the electron compared to the muon. Consequently, an enhancement of their branching fraction will hint for New Physics contributions, affecting the lepton species differently. Although experimentally more accessible than $B^0_{(s)} \to \tau^+\tau^-$ decays, the final-state electrons of the $B^0_{(s)} \to e^+e^-$ decays are significantly affected by bremsstrahlung, complicating their reconstruction. The search for $B^0_{(s)} \to e^+e^-$ performed by LHCb using Run 1 and the 2015 and 2016 dataset resulted in an upper limit to the branching fractions of $\mathcal{B}(B_s^0 \to e^+e^-) < 11.2{\times}10^{-9}$ and $\mathcal{B}(B^0 \to e^+e^-) < 3.0{\times}10^{-9}$ at 95% confidence level [92], still much higher than the Standard Model predictions of $\mathcal{B}(B_s^0 \to e^+e^-) = (8.60 \pm 0.36) \times 10^{-14}$

and $\mathcal{B}(B^0 \to e^+e^-) = (2.41 \pm 0.13) \times 10^{-15}$ [9].

8.3.1 Radiative tail

As mentioned in Chapter 1, photon emission takes place in $B^0_{(s)} \to \mu^+\mu^-$ decays. Part of the final-state radiation emitted by the muons is due to the higher-order QED corrections in the Feynman diagrams of the $B^0_{(s)} \to \mu^+\mu^-$ processes. Experimentally, these FSR photons cause the dimuon invariant mass distributions of the $B^0_{(s)} \to \mu^+\mu^-$ decays to have a left exponential tail that must be included in the probability density function used in the fit to describe the $B^0_{(s)} \to \mu^+\mu^-$ decays.

Currently, the final-state radiation is modelled in simulation using PHOTOS while, in the invariant mass fit, the left tail generated from FSR photons is described by a double-sided Crystal Ball (see Appendix A.1). This shape is not giving a physical description of the processes happening in the final-state radiation but just serves as a parametrization for the FSR in the determination of the branching fraction.

As already shown in Sec. 1.3, the ultrasoft radiation generated by the FSR emission is modelled theoretically by the dimensionless radiative factor relating the radiative and non-radiative branching fractions and given by [9]

$$\Omega(\Delta E) = \left(\frac{2\Delta E}{M_{B^0_{(s)}}}\right)^{-\frac{2\alpha_{em}}{\pi}\left(1+\ln\frac{m_\mu^2}{M^2_{B^0_{(s)}}}\right)} \tag{8.6}$$

where $\Delta E = (M^2_{B^0_{(s)}} - m^2_{\mu^+\mu^-})/2M_{B^0_{(s)}}$, and α_{em} is the electromagnetic coupling constant. It is important to note that Eq. 8.6 is only assumed to be valid in a relatively narrow region around the $B^0_{(s)}$ mass of $\Delta E < 100\,\mathrm{MeV}/c^2$.

Conceptually, the radiative factor can be interpreted as the probability that a photon with energy $E_\gamma \leq \Delta E$ is emitted from the final-state muons in the $B^0_{(s)} \to \mu^+\mu^-$ decays. From this definition, the radiative factor represents the cumulative distribution of the probability density function describing the final-state radiation as a function of ΔE. Therefore, the probability density function describing the FSR as a function of the dimuon invariant mass can be obtained as the derivative of the radiative factor with respect to the dimuon invariant mass:

$$\frac{d\Omega(\Delta E)}{dm_{\mu^+\mu^-}} = \frac{a}{\Delta E}\left(\frac{2\Delta E}{M_{B^0_{(s)}}}\right)^a \frac{d\Delta E}{dm_{\mu^+\mu^-}} \qquad \text{where} \qquad a = -\frac{2\alpha_{em}}{\pi}\left(1+\ln\frac{m_\mu^2}{M^2_{B^0_{(s)}}}\right) \tag{8.7}$$

The resulting probability density function includes the theoretical description of the FSR radiative tail without including the detector effects. As mentioned earlier, the FSR arising

from the QED corrections is described in the simulation with PHOTOS. This simulation algorithm assumes the factorization of the decay amplitude into the non-bremsstrahlung and bremsstrahlung factors and computes the latter in terms of the four-momenta of the particles in the decay. The generated event is then modified to include the bremsstrahlung term [17] and the final-state radiation.

To illustrate the overlap between the $B_s^0 \to \mu^+\mu^-$ FSR radiative tail and the $B^0 \to \mu^+\mu^-$ peak, simulated $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ events are used where the number of events is scaled accordingly to the Standard Model branching ratio and the f_s/f_d ratio. The probability density distribution in Eq. 8.7 is convolved with a double Gaussian distribution to include the detector effects and used to model the $B_s^0 \to \mu^+\mu^-$ component. The $B^0 \to \mu^+\mu^-$ component is modelled analogously. The total model and the two distributions are shown in Fig. 8.3, where the overlap between the $B^0 \to \mu^+\mu^-$ peak and the radiative tail from $B_s^0 \to \mu^+\mu^-$ is clearly visible. At low invariant mass, the tail given by the radiative factor is enhanced with respect to the distribution obtained from simulation, implying a possible disagreement between PHOTOS and the expression obtained from the radiative factor in Eq. 8.7. Part of this disagreement might be caused by the region of validity of the radiative factor, which is expected to be around $\Delta E < 100\,\text{MeV}/c^2$ (or $m_{\mu^+\mu^-} > 5230\,\text{MeV}/c^2$). It is important to note the left tail in the simulation curve includes not only PHOTOS but also the bremsstrahlung radiation emitted by the muons while propagating through the detector. Although the bremsstrahlung effect is expected to be small in the case of muons, its impact in this comparison must be modelled correctly for a future study. The determination of the $\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ branching ratio is a very robust and sensitive probe to test for the concept of Minimal Flavour Violation, requiring the sources of CP violation in the New Physics scenarios to arise from the known Yukawa coupling structures of the Standard Model [93].

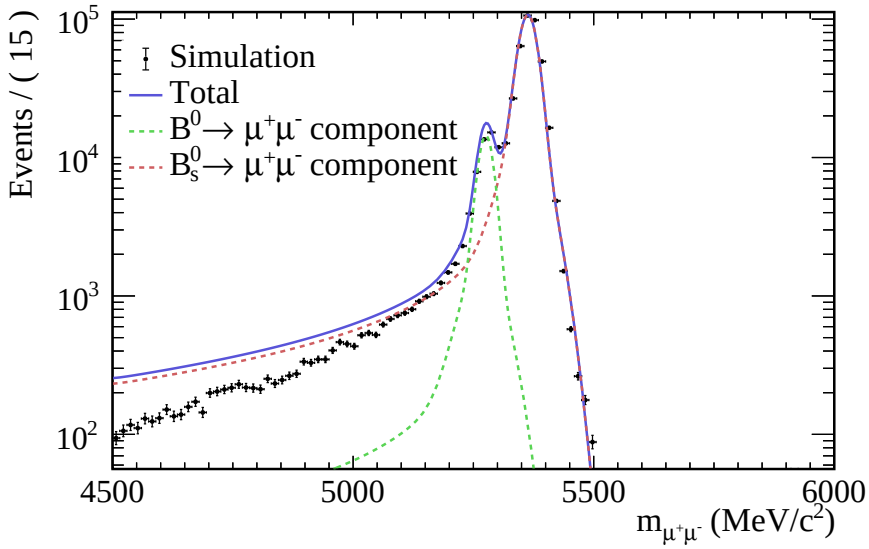


Figure 8.3: Overlap between the $B_s^0 \to \mu^+\mu^-$ radiative tail and the $B^0 \to \mu^+\mu^-$ peak. Both components are modelled with the radiative tail probability density function in Eq. 8.7 convolved with a double Gaussian. The total models and the simulation show a small disagreement at low dimuon invariant mass caused partly by the region of validity of the radiative factor.

A | Probability density functions for mass shapes

A.1 Double-sided Crystal Ball

The double-sided Crystal Ball function [94] consists of a Gaussian core connected to an exponential tail on each side. It is an extension of the Crystal Ball, which has only one exponential tail. The Double-Sided Crystal Ball function is given by

$$f(m|\mu,\sigma,\alpha_l,n_l,\alpha_r,n_r) = N\begin{cases}\left(\frac{n_l}{\alpha_l}\right)^{n_l}\exp\left[-\frac{\alpha_l^2}{2}\right]\left(-\frac{m-\mu}{\sigma}+\frac{n_l}{\alpha_l}-\alpha_l\right)^{-n_l}, & \text{if } \frac{m-\mu}{\sigma} < -\alpha_l \\ \left(\frac{n_r}{\alpha_r}\right)^{n_r}\exp\left[-\frac{\alpha_r^2}{2}\right]\left(\frac{m-\mu}{\sigma}+\frac{n_r}{\alpha_r}-\alpha_r\right)^{-n_r}, & \text{if } \frac{m-\mu}{\sigma} > \alpha_r \\ \exp\left[-\frac{(m-\mu)^2}{2\sigma^2}\right], & \text{otherwise,}\end{cases} \tag{A.1}$$

where μ and σ are the mean and resolution of the Gaussian core; α_r (α_l) define the starting points of the right (left) enhanced tail and n_r (n_l) determine the slope of the right (left) exponential tail.

A.2 Hypatia

The Hypatia function [95] is a generalization of the Double-Sided Crystal ball where the Gaussian core is modified to account for the spread in the per-event mass uncertainties. The function is given by

$$f(m;\mu,\sigma,\lambda,\zeta,\beta,\alpha_l,n_l,\alpha_r,n_r) \propto \begin{cases}\frac{G(\mu-\alpha_l\sigma;\mu,\sigma,\lambda,\zeta,\beta)}{\left(1-m/\left(n\frac{G(\mu-\alpha_l\sigma;\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-\alpha_l\sigma;\mu,\sigma,\lambda,\zeta,\beta)}-\alpha_l\sigma\right)\right)^{n_l}}, & \text{if } \frac{m-\mu}{\sigma} \leq -\alpha_l \\ \frac{G(\mu-\alpha_r\sigma;\mu,\sigma,\lambda,\zeta,\beta)}{\left(1-m/\left(n\frac{G(\mu-\alpha_r\sigma;\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-\alpha_r\sigma;\mu,\sigma,\lambda,\zeta,\beta)}-\alpha_r\sigma\right)\right)^{n_r}}, & \text{if } \frac{m-\mu}{\sigma} \geq \alpha_r \\ G(m;\mu,\sigma,\lambda,\zeta,\beta), & \text{otherwise,}\end{cases} \tag{A.2}$$

where α_r (α_l) and n_r (n_l) control the right (left) exponential tail, G' is the derivative of G with respect to m, which is the hyperbolic core, given by

$$G(m;\mu,\sigma,\lambda,\zeta,\beta) \propto \left[(m-\mu)^2 + A_\lambda^2(\zeta)\sigma^2\right]^{\frac{1}{2}\lambda-\frac{1}{4}} e^{\beta(m-\mu)} \times K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right) \quad \text{(A.3)}$$

where K_λ are special Bessel functions of the third kind and $A_\lambda^2 = \frac{\zeta K_\lambda(\zeta)}{K_{\lambda+1}(\zeta)}$.

A.3 ARGUS

The ARGUS function, introduced by the ARGUS collaboration, is mainly used to describe continuous backgrounds such as partially reconstructed background. It is given by

$$f(m|m_0,c,p) = Nm\left[1-\left(\frac{m}{m_0}\right)^2\right]^p \times \exp\left[c\left(1-\left(\frac{m}{m_0}\right)^2\right)\right] \quad \text{(A.4)}$$

Bibliography

[1] Particle Data Group, P. Zylae, *et al.*, "Review of particle physics," *Prog. Theor. Exp. Phys.*, vol. 83C01, 2020.

[2] R. Aaij *et al.*, "Test of lepton universality in beauty-quark decays," *Nature Phys.*, vol. 18, no. 3, pp. 277–282, 2022.

[3] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.

[4] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961.

[5] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967.

[6] H. Yukawa, "On the Interaction of Elementary Particles I," *Proc. Phys. Math. Soc. Jap.*, vol. 17, pp. 48–57, 1935.

[7] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973.

[8] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, 1970.

[9] M. Beneke, C. Bobeth, and R. Szafron, "Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$," *JHEP*, vol. 10, p. 232, 2019.

[10] T. Cohen, "As Scales Become Separated: Lectures on Effective Field Theory," *PoS*, vol. TASI2018, p. 011, 2019.

[11] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, "Weak decays beyond leading logarithms," *Rev. Mod. Phys.*, vol. 68, pp. 1125–1144, 1996.

[12] A. J. Buras, R. Fleischer, J. Girrbach, and R. Knegjens, "Probing New Physics with the $B_s^0 \to \mu^+\mu^-$ Time-Dependent Rate," *JHEP*, vol. 07, p. 077, 2013.

[13] R. Knegjens, *Strategies to Hunt for New Physics with Strange Beauty Mesons.* PhD thesis, Vrije Universiteit Amsterdam, 2014.

[14] A. Bazavov *et al.*, "B- and D-meson leptonic decay constants from four-flavor lattice QCD," *Phys. Rev. D*, vol. 98, no. 7, p. 074512, 2018.

[15] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, "On the Standard Model prediction for BR($B^0_{s,d} \to \mu^+\mu^-$)," *Eur. Phys. J. C*, vol. 72, p. 2172, 2012.

[16] F. Dettori, D. Guadagnoli, and M. Reboud, "$B^0_s \to \mu^+\mu^-\gamma$ from $B^0_s \to \mu^+\mu^-$," *Phys. Lett. B*, vol. 768, pp. 163–167, 2017.

[17] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays," *Eur. Phys. J. C*, vol. 45, pp. 97–107, 2006.

[18] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, and N. Tuning, "Branching Ratio Measurements of B_s Decays," *Phys. Rev. D*, vol. 86, p. 014027, 2012.

[19] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics*. 1984.

[20] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, "Probing New Physics via the $B^0_s \to \mu^+\mu^-$ Effective Lifetime," *Phys. Rev. Lett.*, vol. 109, p. 041801, 2012.

[21] R. Aaij *et al.*, "Analysis of Neutral B-Meson Decays into Two Muons," *Phys. Rev. Lett.*, vol. 128, no. 4, p. 041801, 2022.

[22] R. Aaij *et al.*, "Measurement of the $B^0_s \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ decays," *Phys. Rev. D*, vol. 105, no. 1, p. 012010, 2022.

[23] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, pp. S08001–S08001, aug 2008.

[24] CMS Collaboration, S. Chatrchyan, *et al.*, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[25] ATLAS Collaboration, G. Add, *et al.*, "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation*, vol. 3, pp. S08003–S08003, aug 2008.

[26] ALICE Collaboration, K. Aamodt, *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[27] LHCb Collaboration, R. Aaij, *et al.*, "LHCb Detector Performance," *Int. J. Mod. Phys. A*, vol. 30, no. 07, p. 1530022, 2015.

[28] LHCb Collaboration, A. A. Alves, *et al.*, "The LHCb detector at the LHC," *Journal of Instrumentation*, vol. 3, pp. S08005–S08005, aug 2008.

[29] R. Aaij *et al.*, "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC," *JINST*, vol. 14, no. 04, p. P04013, 2019.

[30] LHCb Collaboration, R. Aaij, *et al.*, "Performance of the LHCb Vertex Locator," *JINST*, vol. 9, p. P09007, 2014.

[31] J. Gassner, M. Needham, and O. Steinkamp, "Layout and Expected Performance of the LHCb TT Station," 4 2004.

[32] LHCb Collaboration, "LHCb: Inner tracker technical design report," *CERN-LHCC-2002-029*.

[33] M. Adinolfi *et al.*, "Performance of the LHCb RICH detector at the LHC," *Eur. Phys. J. C*, vol. 73, p. 2431, 2013.

[34] F. Archilli *et al.*, "Performance of the Muon Identification at LHCb," *JINST*, vol. 8, p. P10020, 2013.

[35] LHCb Collaboration, "LHCb calorimeters: Technical design report," *CERN-LHCC-2000-036*, 9 2000.

[36] E. Picatoste Olloqui, "LHCb preshower(PS) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring," *J. Phys. Conf. Ser.*, vol. 160, p. 012046, 2009.

[37] C. Abellán Beteta *et al.*, "Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC," 8 2020.

[38] LHCb Collaboration, "LHCb muon system technical design report," *CERN-LHCC-2001-010*, 5 2001.

[39] J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, "Performance of the LHCb High Level Trigger in 2012," *J. Phys. Conf. Ser.*, vol. 513, p. 012001, 2014.

[40] R. Aaij *et al.*, "The LHCb Trigger and its Performance in 2011," *JINST*, vol. 8, p. P04022, 2013.

[41] A. Alfonso Albero *et al.*, "Upgrade trigger selection studies," tech. rep., CERN, Geneva, Sep 2019.

[42] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[43] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[44] I. Belyaev *et al.*, "Handling of the generation of primary events in Gauss, the LHCb simulation framework," in *2010 IEEE Nuclear Science Symposium, Medical Imaging Conference, and 17th Room Temperature Semiconductor Detectors Workshop*, pp. 1155–1161, 2010.

[45] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth. A*, vol. 462, pp. 152–155, 2001.

[46] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003.

[47] J. Allison *et al.*, "Geant4 developments and applications," *IEEE Trans. Nucl. Sci.*, vol. 53, p. 270, 2006.

[48] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo, and P. Robbe, "The LHCb simulation application, Gauss: Design, evolution and experience," *J. Phys. Conf. Ser.*, vol. 331, p. 032023, 2011.

[49] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays," *Phys. Rev. Lett.*, vol. 118, no. 19, p. 191801, 2017.

[50] M. Needham, "Classification of ghost tracks," *LHCb-2007-128*.

[51] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, "Data driven trigger efficiency determination at LHCb," tech. rep., CERN, Geneva, May 2014.

[52] M. Pivk and F. R. Le Diberder, "SPlot: A Statistical tool to unfold data distributions," *Nucl. Instrum. Meth. A*, vol. 555, pp. 356–369, 2005.

[53] M. van Veghel, *Pursuing forbidden beauty: Search for the lepton-flavour violating decays* $B^0 \to e^\pm\mu^\mp$ *and* B_s^0ß$e^\pm\mu^\mp$ *and study of electron-reconstruction performance at LHCb*. PhD thesis, University of Groningen, 2020.

[54] 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., CERN, Geneva, Nov 2014.

[55] 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," tech. rep., CERN, Geneva, Jul 2016.

[56] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Chapman and Hall/CRC, 1984.

[57] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[58] A. Hocker *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," 3 2007.

[59] F. Farris, "The gini index and measures of inequality," *American Mathematical Monthly*, vol. 117, pp. 851–864, 12 2010.

[60] ARGUS Collaboration, H. Albrecht, *et al.*, "Measurement of the polarization in the decay $B \to J/\psi K^*$," *Physics Letters B*, vol. 340, no. 3, pp. 217–220, 1994.

[61] A. Rogozhnikov, "Reweighting with Boosted Decision Trees," *J. Phys. Conf. Ser.*, vol. 762, no. 1, p. 012036, 2016.

[62] K. S. Cranmer, "Kernel estimation in high-energy physics," *Comput. Phys. Commun.*, vol. 136, pp. 198–207, 2001.

[63] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of the B_c^- meson production fraction and asymmetry in 7 and 13 TeV pp collisions," *Phys. Rev. D*, vol. 100, no. 11, p. 112006, 2019.

[64] LHCb Collaboration, R. Aaij, *et al.*, "Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays," *Nature Phys.*, vol. 11, pp. 743–747, 2015.

[65] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of b hadron fractions in 13 TeV pp collisions," *Phys. Rev. D*, vol. 100, no. 3, p. 031102, 2019.

[66] LHCb Collaboration, R. Aaij, *et al.*, "First observation of the decay $B_s^0 \to K^-\mu^+\nu_\mu$ and Measurement of $|V_{ub}|/|V_{cb}|$," *Phys. Rev. Lett.*, vol. 126, no. 8, p. 081804, 2021.

[67] LHCb Collaboration, R. Aaij, *et al.*, "Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B_s^0 decay branching fractions," *Phys. Rev. D*, vol. 104, no. 3, p. 032005, 2021.

[68] 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. D*, vol. 86, p. 114025, 2012.

[69] BaBar Collaboration, B. Aubert, *et al.*, "Improved Measurements of the Branching Fractions for $B^0 \to \pi^+\pi^-$ and $B^0 \to K^+\pi^-$, and a Search for $B^0 \to K^+K^-$," *Phys. Rev. D*, vol. 75, p. 012008, 2007.

[70] Belle Colaboration, Y. T. Duh, *et al.*, "Measurements of branching fractions and direct CP asymmetries for B→Kπ, B→$\pi\pi$ and B→KK decays," *Phys. Rev. D*, vol. 87, no. 3, p. 031103, 2013.

[71] CLEO Collaboration, A. Bornheim, *et al.*, "Measurements of charmless hadronic two body B meson decays and the ratio BF($B \to DK$) / BF($B \to D\pi$)," *Phys. Rev. D*, vol. 68, p. 052002, 2003. [Erratum: Phys.Rev.D 75, 119907 (2007)].

[72] 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.*, vol. 124, no. 12, p. 122002, 2020.

[73] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of b-hadron production fractions in 7 TeV pp collisions," *Phys. Rev. D*, vol. 85, p. 032008, 2012.

[74] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics," *JHEP*, vol. 04, p. 001, 2013.

[75] LHCb Collaboration, R. Aaij, *et al.*, "Measurement of the branching fraction of the $B^0 \to D_s^+ \pi^-$ decay," *Eur. Phys. J. C*, vol. 81, no. 4, p. 314, 2021.

[76] G. Cowan, *Statistical data analysis.* Oxford University Press, USA, 1998.

[77] D. Melikhov and N. Nikitin, "Rare radiative leptonic decays $B_(d, s) \to l^+ l^- \gamma$," *Phys. Rev. D*, vol. 70, p. 114028, 2004.

[78] A. Kozachuk, D. Melikhov, and N. Nikitin, "Rare FCNC radiative leptonic $B_{s,d} \to \gamma l^+ l^-$ decays in the standard model," *Phys. Rev. D*, vol. 97, no. 5, p. 053007, 2018.

[79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[80] A. Wald, "Test of statistical hypotheses concerning several parameters when the number of observation is large," *Transactions of the American Mathematical Society*, vol. 54, pp. 462–482, Nov 1943.

[81] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Annals Math. Statist.*, vol. 9, no. 1, pp. 60–62, 1938.

[82] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys. G*, vol. 28, pp. 2693–2704, 2002.

[83] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method)," in *Workshop on Confidence Limits*, pp. 81–101, 8 2000.

[84] M. Kenzie *et al.*, "GammaCombo: A statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals [Computer software]," Sept. 2021.

[85] R. Aaij *et al.*, "Measurement of the CKM angle γ from a combination of LHCb results," *JHEP*, vol. 12, p. 087, 2016.

[86] CMS Collaboration, "Measurement of $\mathrm{B}_\mathrm{s}^0 \to \mu^+\mu^-$ decay properties and search for the $\mathrm{B}^0 \to \mu\mu$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-BPH-21-006, CERN, 2022.

[87] W. Altmannshofer and P. Stangl, "New physics in rare B decays after Moriond 2021," *Eur. Phys. J. C*, vol. 81, no. 10, p. 952, 2021.

[88] A. M. Sirunyan *et al.*, "Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment," *JHEP*, vol. 04, p. 188, 2020.

[89] 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*, vol. 04, p. 098, 2019.

[90] C. Cornella, J. Fuentes-Martin, and G. Isidori, "Revisiting the vector leptoquark explanation of the B-physics anomalies," *JHEP*, vol. 07, p. 168, 2019.

[91] R. Aaij *et al.*, "Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$," *Phys. Rev. Lett.*, vol. 118, no. 25, p. 251802, 2017.

[92] R. Aaij *et al.*, "Search for the Rare Decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$," *Phys. Rev. Lett.*, vol. 124, no. 21, p. 211802, 2020.

[93] G. Isidori and D. M. Straub, "Minimal Flavour Violation and Beyond," *Eur. Phys. J. C*, vol. 72, p. 2103, 2012.

[94] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances.* PhD thesis, Cracow, INP, 1986.

[95] D. Martínez Santos and F. Dupertuis, "Mass distributions marginalized over per-event errors," *Nucl. Instrum. Meth. A*, vol. 764, pp. 150–155, 2014.

Summary

Fundamental particle physics describes the fundamental constituents of the universe in terms of a few elementary particles, known as quarks and leptons, and their interactions. Both the properties of these particles and the workings of their interactions are encompassed in the Standard Model of particle physics, which offers a mathematical description of such elementary processes. Even though the Standard Model has been proven to be a self-consistent theory successfully predicting most particle physics phenomena, it leaves several open questions and is therefore not the complete picture. The answer to these questions might be found in the so-called "New Physics" or physics models beyond the Standard Model. The search of such New Physics models is performed by studying different fundamental particle processes to push the limits of our knowledge.

An interesting and important property of quarks must be taken into account in the search for New Physics: quarks cannot be detected alone, they are always "confined" with other quarks forming composite particles known as *hadrons*. Even though the most well-known hadrons are protons and neutrons, which make up most of the ordinary matter in the universe, there are many more different hadrons resulting from distinct combinations of quarks. All hadrons, except the protons and neutrons in atoms, are unstable and decay or "transform" into at least two other particles. Some of these decays are excellent laboratories in the search of physics beyond the Standard Model and their study gives a clearer picture of potential new theories.

Two of such less commonly known hadrons are the B_s^0 and B^0 hadrons. They are formed by the combination of two quarks: a *beauty* quark and a *strange* quark in the case of the B_s^0 and a *beauty* quark and a *down* quark in the case of the B^0. As all hadrons, the B^0 and B_s^0 hadrons decay into other particles and are known to have hundreds of different ways to decay. One possibility is their decay into two muons, a type of lepton. These decays, formally referred to as $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$, are predicted to be extremely rare in the Standard Model, with only 3 every billion B_s^0 hadrons decaying into two muons, while for B^0, the expected rate is 1 every 10 billion. Despite being very suppressed in the Standard Model, new quantum particles or forces might play a role in these decays, altering the rate at which they are expected to occur. Any disagreement between the measured decay rates and the

Standard Model predictions would imply the existence of new particles or forces altering the quantum process of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays.

Because the B_s^0 and B^0 mesons are very unstable, they decay very quickly, and the study of their decays is only possible in dedicated particle experiment producing them in a controlled environment. Currently, the best facility is the Large Hadron Collider located at CERN, a particle accelerator that accelerates particles, mainly protons, to high velocities and collides them in one of the four detectors positioned in its ring. Many different particles are produced in the proton-proton collisions, including B^0 and B_s^0 hadrons. These generated particles and their subsequent decays are detected in particle detectors, such as LHCb, which is specialized in the analysis of processes involving beauty or charm quarks. Since these decays are very rare processes, a huge amount of data is needed to detect them; thus the entire dataset collected by LHCb from 2011 to 2018 is exploited in this thesis.

Only a small fraction of the particles produced in the proton-proton collisions undergo processes that can be sensitive to New Physics and hence interesting for further analysis. In order to keep the data size manageable, the *trigger* system performs a real-time selection deciding whether the particle processes ocurring after the proton-proton collisions are interesting for further analysis, and stored in the disk, or discarded. In the case of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ analysis using LHCb data, the trigger is optimized to store the events containing $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. Nevertheless, several real $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays might not fire the trigger and get lost. The number of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates determined in the analysis must then be corrected by the efficiency of the trigger to account for these lost real candidates. The evaluation of the trigger efficiencies is performed using a combination of simulation and data samples and is one of the main topics of this thesis.

The detection of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, jointly referred to as signal, is relatively clean in the LHCb detector, as shown in the sketch displayed in Fig. 1. First, a B_s^0 or B^0 hadron must be produced in the proton-proton collision. These hadrons propagate a few millimetres in the vacuum before decaying in the detector volume into two oppositely charged muons. The muons leave a trace in different parts of the detector and are the only particles reaching its outermost part, allowing for their identification. Since muons are charged particles, their trajectories are bent by the dipole magnet, enabling the determination of their charges and trajectories. The point where the B_s^0 or B^0 hadron decayed can be determined by extrapolating the trajectories of the two final-state muons to the point where they meet.

Despite the signature left by the decays in the detector being very clean, some other different processes might be wrongly identified as $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. Because the signal decays are so rare, these so-called *background* events are expected to be $10^7 - 10^8$ times more likely than the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. In order to dis-

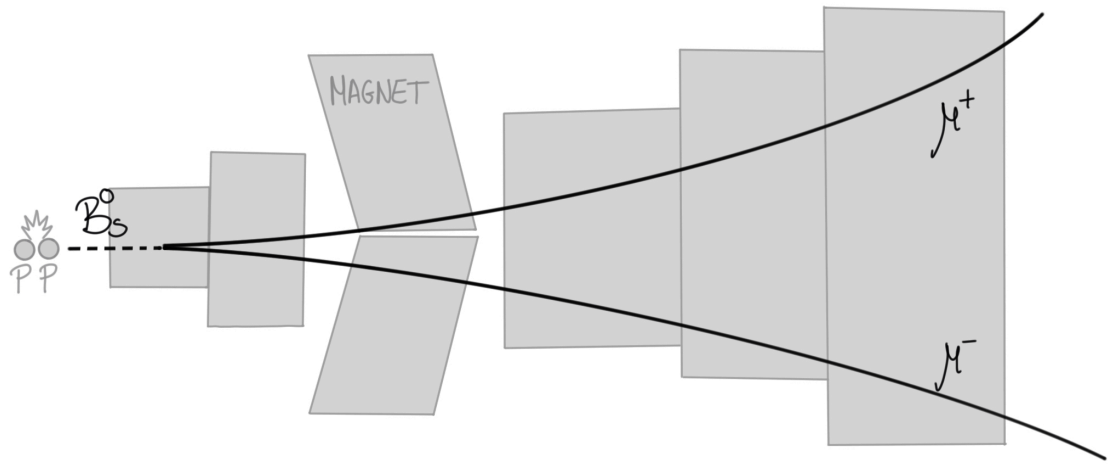


Figure 1: Sketch of the $B_s^0 \to \mu^+\mu^-$ decay signature in the LHCb detector. The B_s^0 originating in the proton-proton collision propagates a few milimeters inside the detector before decaying into two muons, which are the only particles reaching the outermost part of the detector. The trajectories of the muons are bent by the magnet depending on their charges.

criminate between the signal and background, the energy and momenta of the two final-state muons are combined into the so-called *invariant mass* variable. The majority of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ background candidates are distributed exponentially throughout the invariant mass variable, while the real $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates are observed as peaks or bumps in these distributions. The height of the peak corresponds to the number of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays obtained in this analysis, and is directly related to the quantity we are ultimately interested in: the rates of such decays, also known as *branching fractions.* A simple sketch illustrating the dimuon invariant mass distribution of the data candidates for a single signal component and a background component is given in Fig. 2.

The discriminating power offered by the invariant mass of the two muons is insufficient for clearly detecting the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. Since these decays are so rare, the sensitivity of the data sample must be maximized, meaning that the background reduction must be maximal without losing real signal candidates. To achieve this, a machine learning technique known as the Boosted Decision Tree or BDT is used. The BDT classifies the candidates as more or less signal-like by assigning a score to each of them. This score ranges from 0 to 1, such that the more likely the candidate is signal, the higher the score. The sensitivity of the data sample is maximized by separating the dataset into subsamples of the BDT score and analyzing the invariant mass distributions in each of these subsamples simultanously. The results obtained from each BDT subsample are combined into a single branching fraction measurement. To do this, an estimation of the expected fraction of signal events ending up in each BDT subsample is needed. This estimation is obtained through the *BDT calibration* procedure, one of the main topics of this thesis.

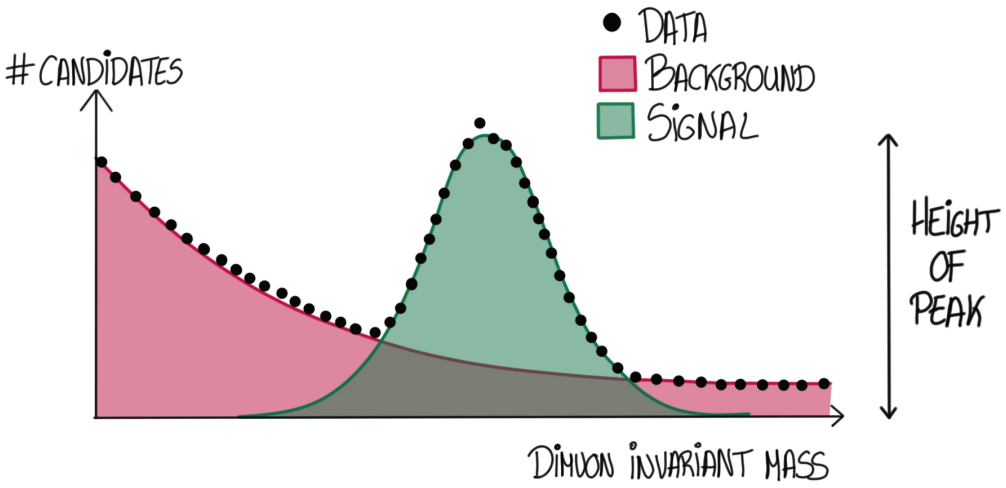


Figure 2: Illustration of signal and background components and the distribution of the candidates along the dimuon invariant mass. The height of the signal peak is related to its decay rate, also known as branching fraction.

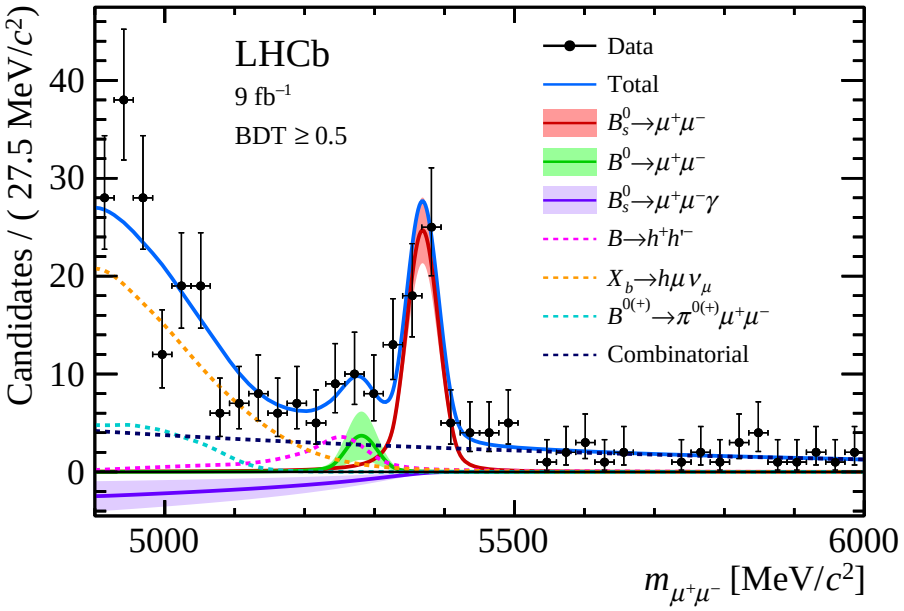


Figure 3: Invarint mass distributions of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ selected candidates for BDT > 0.5. The blue line corresponds to the total fit model, and the separated components are reported in the legend. The $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ components are shown in green and red, respectively.

The distribution of all $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ selected candidates are shown in Fig. 3 for candidates with BDT scores higher than 0.5. The $B_s^0 \to \mu^+\mu^-$ distribution, shown in red, significantly "peaks" over the background components, implying it is statistically significant. Although a small excess relative to the background is visible for $B^0 \to \mu^+\mu^-$ decays, its bump is not significantly differentiated from the rest of the background components, implying the experimental sensitivity is still not enough for detecting this decay process. From the $B_s^0 \to \mu^+\mu^-$ distribution and the height of the peak, the branching fraction for the $B_s^0 \to \mu^+\mu^-$ decays is measured to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$, where the first uncertainty is statistical and the second systematic. This measurement is found to be in agreement with the Standard Model prediction. In the case of the $B^0 \to \mu^+\mu^-$ decay, an upper limit of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.6 \times 10^{-10}$ at 95% CL is set to its branching fraction, still above the sensitivity needed to achieve the observation of the $B^0 \to \mu^+\mu^-$ decay as predicted by the Standard Model. These results are still statistically limited, meaning more data is needed to obtain a more accurate and precise measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and to ultimately discover the $B^0 \to \mu^+\mu^-$ decay. Although statistically limited, the results presented in this thesis limit the room available for New Physics models and constrain the future theoretical developments of extensions or new theories beyond the Standard Model.

Samenvatting

De fundamentele deeltjesfysica beschrijft het universum in termen van een paar elementaire deeltjes, bekend als quarks en leptonen en hun interacties. Zowel de eigenschappen van deze deeltjes als de werking van hun interacties zijn verwerkt in het Standaardmodel van de deeltjesfysica, dat een wiskundige beschrijving biedt van deze elementaire processen. Hoewel is aangetoond dat het Standaardmodel een zelfconsistente theorie is die met succes de meeste verschijnselen van de deeltjesfysica voorspelt, laat het verschillende vragen open en is het niet het volledige plaatje. Het antwoord op deze vragen zou gevonden kunnen worden in de zogenaamde "Nieuwe Fysica", natuurkundige modellen buiten het Standaardmodel. Het zoeken naar dergelijke Nieuwe Fysica modellen gebeurt door het bestuderen van verschillende fundamentele deeltjesprocessen om de grenzen van onze kennis te verleggen.

Bij het zoeken naar Nieuwe Fysica moet rekening worden gehouden met een interessante en belangrijke eigenschap van quarks: quarks kunnen niet alleen worden waargenomen, ze zijn altijd "opgesloten" met andere quarks en vormen samengestelde deeltjes die bekend staan als *hadronen*. Hoewel de bekendste hadronen protonen en neutronen zijn, die het grootste deel van de gewone materie in het heelal vormen, zijn er veel meer verschillende hadronen die voortkomen uit verschillende combinaties van quarks. Alle hadronen, behalve protonen en de neutronen in atomen, zijn onstabiel en vervallen of "transformeren" in ten minste twee andere deeltjes. Sommige van deze vervallen zijn uitstekend geschikt voor de zoektocht naar fysica buiten het Standaardmodel en hun studie geeft een duidelijker beeld van mogelijke nieuwe theorieën.

Twee van zulke minder bekende hadronen zijn de B_s^0 en B^0 hadronen. Zij worden gevormd door de combinatie van twee quarks: een *beauty* quark en een *strange* quark in het geval van de B_s^0 en een *beauty* quark en een *down* quark in het geval van de B^0. Zoals alle hadronen vervallen ook de B^0 en de B_s^0 in andere deeltjes en zijn er honderden verschillende manieren bekend hoe ze vervallen. Eén mogelijkheid is hun verval in twee muonen, een lepton soort. Deze vervalmogelijkheden, formeel aangeduid als $B^0 \to \mu^+\mu^-$ en $B_s^0 \to \mu^+\mu^-$, zijn volgens het Standaardmodel uiterst zeldzaam: slechts 3 per miljard B_s^0 hadronen vervallen in twee muonen, terwijl voor B^0 een verval per 10 miljard wordt verwacht. Hoewel het verval in het Standaardmodel zeer zeldzaam is, zouden nieuwe kwantumdeeltjes of krachten een

rol kunnen spelen in deze vervalprocessen, waardoor de kans waarmee ze naar verwachting zullen optreden, verandert. Ieder verschil tussen de gemeten vervalsnelheden en de voorspellingen van het Standaardmodel zou het bestaan impliceren van nieuwe deeltjes of krachten die het kwantumproces van het verval van $B^0 \to \mu^+\mu^-$ en $B_s^0 \to \mu^+\mu^-$ veranderen.

Omdat de mesonen B_s^0 en B^0 zeer onstabiel zijn, vervallen ze zeer snel en de studie van hun verval is daarom alleen mogelijk in speciale deeltjesexperimenten die deze mesonen in een gecontroleerde omgeving produceren. De beste faciliteit is momenteel de Large Hadron Collider in het CERN, een deeltjesversneller die deeltjes, voornamelijk protonen, versnelt tot hoge snelheden en ze laat botsen in een van de vier detectoren in een ring. Bij de proton-proton botsingen worden veel verschillende deeltjes geproduceerd, waaronder de hadronen B^0 en B_s^0. Deze deeltjes en hun latere verval worden gedetecteerd in deeltjesdetectoren, zoals de LHCb detector, die gespecialiseerd is in de analyse van processen waarbij beauty- of charm quarks betrokken zijn. Aangezien deze vervalprocessen zeer zeldzaam zijn, is een enorme hoeveelheid gegevens nodig om ze te detecteren. Daarom wordt de volledige dataset die LHCb van 2011 tot 2018 heeft verzameld in dit proefschrift benut.

Slechts een klein deel van de deeltjes die bij de proton-protonbotsingen worden geproduceerd ondergaan processen die gevoelig kunnen zijn voor nieuwe fysica en dus interessant zijn voor verdere analyse. Om de datagrootte beheersbaar te houden, voert het *trigger* systeem een real-time selectie uit om te beslissen of de deeltjesprocessen na de proton-proton botsingen interessant zijn voor verdere analyse en opgeslagen worden of niet. In het geval van de analyse van $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ is met behulp van LHCb-gegevens de trigger geoptimaliseerd om de gebeurtenissen met $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-verval op te slaan. Niettemin is het mogelijk dat verscheidene echte $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-decays de trigger niet activeren en verloren gaan. Het aantal in de analyse bepaalde kandidaten voor $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ moet dan worden gecorrigeerd met de efficiëntie van de trigger om rekening te houden met deze verloren gegane echte vervallen. De evaluatie van deze trigger-efficiënties wordt uitgevoerd met behulp van een combinatie van simulatie en detector gegevens en is een van de belangrijkste onderwerpen van dit proefschrift.

De detectie van de $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ verval, gezamenlijk aangeduid als signaal, is relatief schoon in de LHCb detector zoals te zien is in de schets in Fig. 4. Eerst moet een hadron B_s^0 of B^0 worden geproduceerd in de proton-proton botsing. Deze hadronen verplaatsen zich enkele millimeters in het vacuüm voordat ze in het detectorvolume vervallen in twee tegengesteld geladen muonen. De muonen laten een spoor achter in verschillende delen van de detector en zijn de enige deeltjes die het buitenste deel ervan bereiken, waardoor ze kunnen worden geïdentificeerd. Aangezien muonen geladen deeltjes zijn, worden hun banen afgebogen door de dipoolmagneet, waardoor hun ladingen en banen kunnen worden bepaald. Het punt waar het B_s^0 of B^0 hadron is vervallen kan worden bepaald door de banen van de twee muonen in de eindtoestand te extrapoleren naar het punt waar ze elkaar ontmoeten.

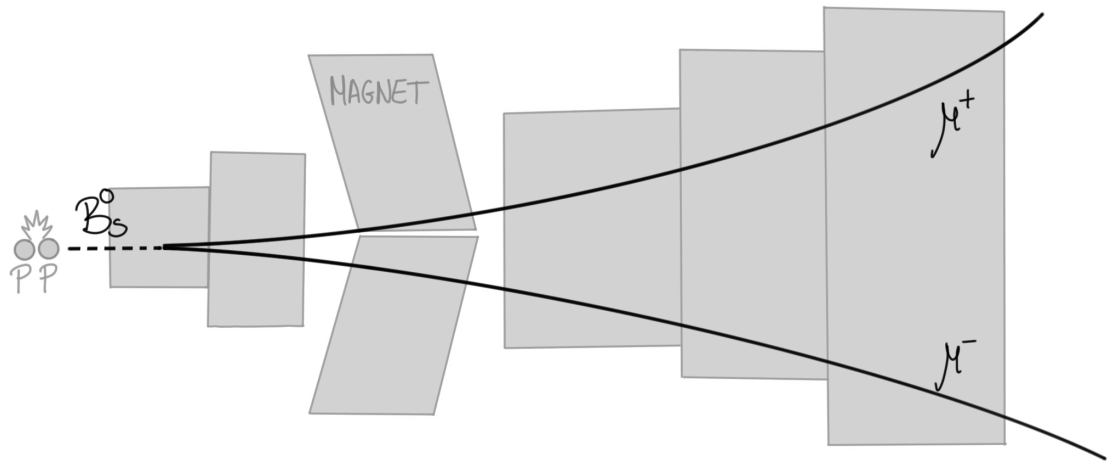


Figure 4: Schets van het $B_s^0 \to \mu^+\mu^-$-verval in de LHCb-detector. De uit de proton-protonbotsing afkomstige $B_s^0 \to \mu^+\mu^-$ verplaatst zich een paar milimeter in de detector alvorens te vervallen in twee muonen, de enige deeltjes die het buitenste deel van de detector bereiken. De banen van de muonen worden door de magneet afgebogen, afhankelijk van hun lading.

Ondanks het feit dat de sporen van het verval in de detector zeer schoon zijn, kunnen sommige andere processen ten onrechte als $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-verval worden geïdentificeerd. Omdat de signaal vervallen zo zeldzaam zijn, wordt verwacht dat deze zogenaamde achtergrondgebeurtenissen $10^7 - 10^8$ waarschijnlijker zijn dan de $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-decays. Om onderscheid te maken tussen het signaal en de achtergrond, worden de energie en de momenten van de twee muonen in de eindtoestand gecombineerd tot de zogenaamde *invariante massa* variabele. De meerderheid van de $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ achtergrondkandidaten zijn exponentieel verdeeld over de invariante massavariabele, terwijl de echte $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ kandidaten worden waargenomen als pieken of bulten in deze verdelingen. De hoogte van de piek komt overeen met het aantal in deze analyse verkregen $B_s^0 \to \mu^+\mu^-$- en $B^0 \to \mu^+\mu^-$-decays, en is direct gerelateerd aan de grootheid waarin we uiteindelijk geïnteresseerd zijn: de waarschijnlijkheid van zulke decays, ook bekend als *branching fractions*. Een eenvoudige schets van de dimuon invariante massaverdeling van de gegevenskandidaten voor een enkele signaalcomponent en een achtergrondcomponent wordt gegeven in Fig. 5.

Het onderscheidend vermogen van de invariante massa van de twee muonen is onvoldoende om het verval van de twee muonen duidelijk te detecteren. Omdat deze vervallen zo zeldzaam zijn, moet de gevoeligheid van het datamonster worden gemaximaliseerd, wat betekent dat de achtergrond maximaal moet worden gereduceerd zonder echte signaalkandidaten te verliezen. Om dit te bereiken wordt een machine-learningtechniek gebruikt die bekend staat als de Boosted Decision Tree of BDT. De BDT classificeert de kandidaten als

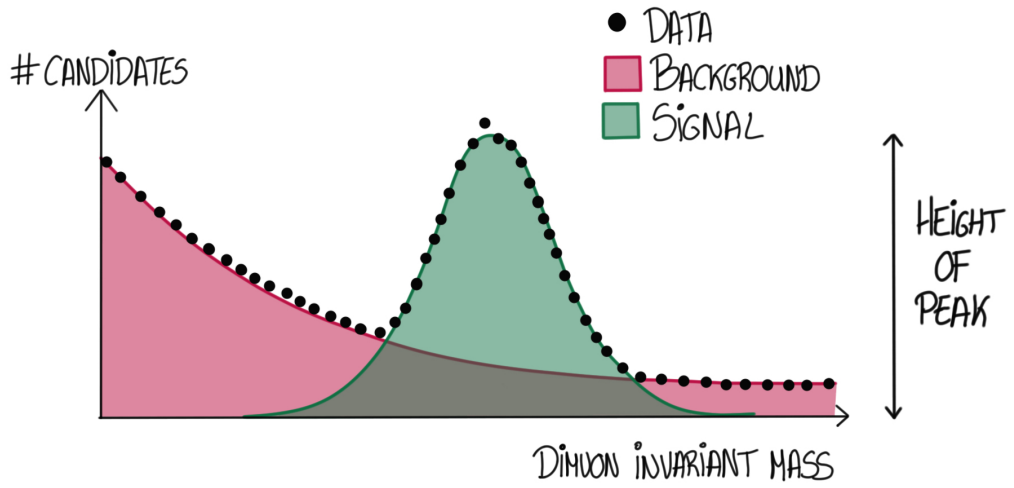


Figure 5: Illustratie van signaal- en achtergrondcomponenten en de verdeling van de kandidaten langs de dimuon invariante massa. De hoogte van de signaalpiek is gerelateerd aan de vervalsnelheid, ook wel vertakkingsfractie genoemd.

meer of minder signaalachtig door aan elk van hen een score toe te kennen. Deze score varieert van 0 tot 1, zodat de score hoger is naarmate het waarschijnlijker is dat de kandidaat een signaal is. De gevoeligheid van het datamonster wordt gemaximaliseerd door de dataset te verdelen in verschillende delen van BDT-score en de invariante massaverdeling in elk van deze delen gelijktijdig te analyseren. De resultaten van elk BDT-deel worden gecombineerd tot een enkele vertakkingsfractiemeting. Hiervoor is een schatting nodig van de verwachte signaal fractie die in elke BDT-deel terechtkomen. Deze schatting wordt verkregen via de ijking van de BDT calibratie, een van de hoofdonderwerpen van dit proefschrift.

De verdeling van alle geselecteerde $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ kandidaten wordt getoond in Fig. 6 voor kandidaten met BDT scores hoger dan 0.5. De $B_s^0 \to \mu^+\mu^-$ verdeling, weergegeven in rood, vertoont een duidelijke "piek" boven de achtergrondcomponenten, wat betekent dat deze statistisch significant is. Hoewel een kleine overmaat ten opzichte van de achtergrond zichtbaar is voor het verval van $B^0 \to \mu^+\mu^-$, is de piek ervan niet significant verschillend van de rest van de achtergrondcomponenten, wat betekent dat de experimentele gevoeligheid nog steeds niet voldoende is om dit vervalproces te detecteren. Uit de $B_s^0 \to \mu^+\mu^-$-verdeling en de hoogte van de piek is de vertakkingsfractie voor het $B_s^0 \to \mu^+\mu^-$-verval gemeten op $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$ waarbij de eerste onzekerheid statistisch is en de tweede systematisch. Deze meting komt overeen met de voorspelling van het Standaardmodel. In het geval van het $B^0 \to \mu^+\mu^-$-verval is een bovengrens van $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.6 \times 10^{-10}$ bij 95% CL ingesteld op zijn vertakkingsfractie, nog steeds boven de gevoeligheid die nodig is voor de waarneming van het $B^0 \to \mu^+\mu^-$-verval zoals voorspeld door het Standaardmodel. Deze resultaten zijn nog steeds statistisch beperkt,

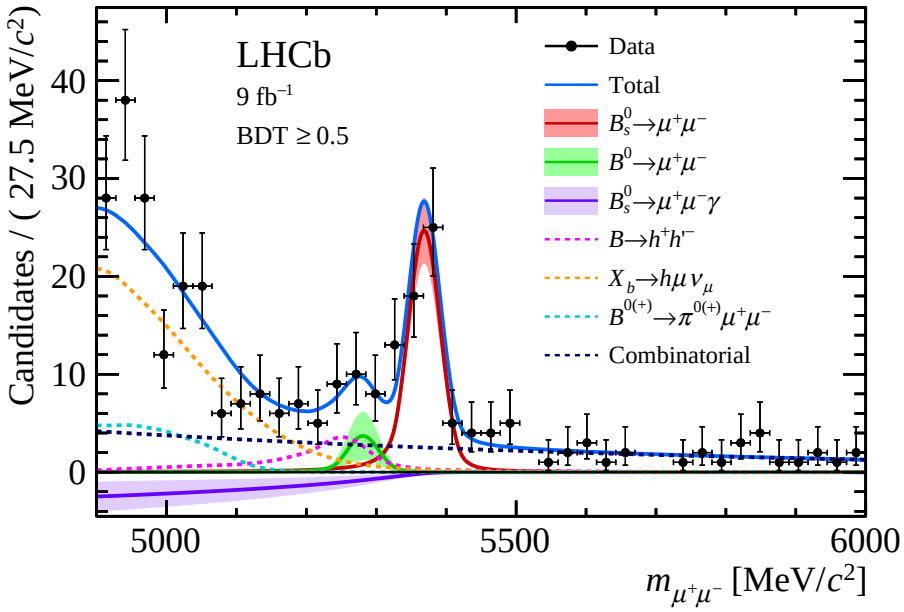


Figure 6: Invarintmassaverdeling van de $B_s^0 \to \mu^+\mu^-$ en $B^0 \to \mu^+\mu^-$ geselecteerde kandidaten voor BDT > 0.5. De blauwe lijn komt overeen met het totale fit-model, de afzonderlijke componenten staan in de legenda. De componenten $B^0 \to \mu^+\mu^-$ en $B_s^0 \to \mu^+\mu^-$ zijn respectievelijk groen en rood gekleurd.

wat betekent dat er meer gegevens nodig zijn om een nauwkeurigere en preciezere meting van de vertakkingsfractie van het $B^0 \to \mu^+\mu^-$-verval te verkrijgen en om uiteindelijk het $B^0 \to \mu^+\mu^-$-verval te ontdekken. Hoewel de resultaten statistisch beperkt zijn, beperken in dit proefschrift de beschikbare ruimte voor nieuwe natuurkundige modellen en beperken zij de toekomstige theoretische ontwikkelingen van uitbreidingen of nieuwe theorieën buiten het Standaard Model.

Resumen

La física de partículas fundamental es una rama de la física que ofrece una descripción del universo a partir de dos tipos de partículas elementales, llamadas quarks y leptones, y las interacciones entre ellas. Tanto las propiedades que definen a estas partículas como los mecanismos que rigen sus interacciones se engloban en el Modelo Estándar de la física de partículas, una teoría que proporciona una elegante descripción matemática de estos procesos elementales. El Modelo Estándar ha demostrado ser una teoría consistente que predice con éxito la mayoría de los fenómenos observados en física de partículas y, sin embargo, está lejos de ser la teoría absoluta, ya que deja varias incógnitas sin resolver. Algunas de ellas pueden ser ¿qué ha sido de toda la antimateria que necesariamente hubo de ser creada en igual proporción a la materia en la que consiste mayoritariamente nuestro universo? ¿Cuál es la relación entre los quarks y los leptones y por qué tiene diferentes propiedades?

Encontrar la respuesta a estas y otras preguntas es esencial para ampliar nuestro conocimiento sobre el funcionamiento interno de la naturaleza a un nivel fundamental y sobre el origen de la materia en la formación del universo. Algunas de estas respuestas podrían encontrarse en modelos que van más allá del Modelo Estándar, ya sean desarrollados como extensiones del mismo o como teorías completamente nuevas, de forma que dicho modelo sea una aproximación a baja energía de una teoría más amplia, como es el caso de la mecánica clásica y la relatividad. Nuestro deber como físicos de partículas es intentar encontrar una explicación a estos interrogantes, limitando modelos más allá del Modelo Estándar o incluso encontrando nuevas partículas cuya existencia nunca antes había sido predicha. Esta búsqueda se realiza a partir del estudio de diferentes procesos que involucran interacciones entre las partículas elementales anteriormente mencionadas.

Los quarks y los leptones consituyen el universo y su comportamiento es diferente. Una particularidad exclusiva de los quarks es el hecho de que no pueden detectarse individualmente, ya que siempre se combinan con otros quarks formando partículas compuestas llamadas hadrones. Los hadrones más conocidos son los protones y los neutrones, los cuáles forman los átomos y constituyen la mayor parte de la materia del universo. No obstante, hay muchos más tipos de hadrones que resultan de las diferentes combinaciones de varios tipos de quarks distintos. Excepto los protones y neutrones que se encuentran en los átomos, el

resto de hadrones son inestables y acaban desintegrándose o "transformándose" en múltiples partículas diferentes. Estas desintegraciones se producen a partir de un proceso cuántico que puede ser alterado por la presencia de partículas aún no conocidas. Por ello, algunas de estas desintegraciones son extremadamente sensibles a la posible existencia de nuevas partículas y su estudio permite obtener una visión más clara sobre posibles nuevos modelos más allá del Modelo Estándar.

Dos de estos hadrones menos conocidos son los hadrones B_s^0 y B^0, ambos formados por la combinación de dos quarks diferentes. El hadrón B_s^0 surge de la combinación de un quark tipo b (bello o *beauty* en inglés) y un quark tipo s (encanto o *charm* en inglés), mientras que el hadrón B^0 está formado por un quark tipo b y un quark tipo d (abajo o *down* en inglés). Como el resto de los hadrones, los hadrones B_s^0 y B^0 son inestables y se desintegran en otras partículas. De hecho, se sabe que estos hadrones tienen muchas formas diferentes de desintegrarse, transformándose en diferentes tipos de partículas. Una de estas formas es su desintegración en dos muones, un tipo específico de leptón, similar al electrón pero con mayor masa. La desintegración del hadrón B_s^0 en dos muones se expresa formalmente como $B_s^0 \to \mu^+\mu^-$ mientras que, en el caso de la desintegración del hadrón B^0, la expresión es $B^0 \to \mu^+\mu^-$. El Modelo Estándar predice la excepcional rareza de estas desintegraciones con tan solo 3 de cada mil millones (1,000,000,000) de hadrones B_s^0 desintegrándose en dos muones. El caso de $B^0 \to \mu^+\mu^-$ es aún más raro con solo 1 de cada 10 mil millones (10,000,000,000) de hadrones B^0 desintegrándose en dos muones. Estos valores, conocidos como la fracción de ramificación, indican la cantidad de hadrones B^0 o B_s^0 que se desintegran en dos muones respecto a todos los hadrones B^0 o B_s^0 que se desintegran en general. A pesar de que las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$ son muy raras en el Modelo Estándar, la existencia de nuevas y desconocidas partículas o fuerzas cuánticas podría desempeñar un papel decisivo en ellas, alterando la mencionada tasa de frecuencia con la que se espera que ocurran. Por tanto, el objetivo de esta tesis es medir la fracción de ramificación de los procesos $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$ y compararla con la que predice el Modelo Estándar. Cualquier discrepancia entre ambos valores implicaría la existencia de nuevas partículas o interacciones que alteran el proceso cuántico interno de las desintegraciones $B^0 \to \mu^+\mu^-$ y $B_s^0 \to \mu^+\mu^-$ y abriría una ventana a posibles nuevos modelos más allá del Modelo Estándar.

Para medir la fracción de ramificación de los procesos $B^0 \to \mu^+\mu^-$ y $B_s^0 \to \mu^+\mu^-$ es necesario poder detectar de alguna manera las partículas involucradas, es decir, los hadrones B^0 y B_s^0 y los dos muones. Puesto que son partículas muy inestables, los hadrones B_s^0 y B^0 se desintegran muy rápidamente, haciendo que el estudio de sus procesos de desintegración sea únicamente posible en instalaciones especializadas donde se provoca la generación de estas partículas en un entorno controlado. Actualmente, la mejor de dichas instalaciones es el Gran Colisionador de Hadrones (*Large Hadron Collider* en inglés o *LHC*), un acelerador de partículas situado en el CERN. El LHC acelera partículas, principalmente protones, hasta alcanzar altas velocidades para hacerlas colisionar entre sí en uno de los cuatro pun-

tos destinados para ello. En cada uno de estos puntos se localiza un detector que registra la gran cantidad de partículas que se genera en las colisiones de protones, entre ellas algunos hadrones B^0 y B_s^0, y sus posteriores desintegraciones. De los cuatro detectores que se encuentran en el LHC, esta tesis emplea los datos registrados por el detector conocido como LHCb, especializado en el análisis de procesos en los que intervienen quarks tipo b, como es el caso de los procesos que nos interesan, $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$. Debido a la rareza de estas desintegraciones, es necesaria una gran cantidad de datos para conseguir registrar un número relativamente significativo de los mencionados procesos y para determinar las correspondientes fracciones de ramificación. Por esta razón, en esta tesis se emplea el total de datos registrados por el detector LHCb desde el año 2011 hasta el año 2018.

De todas las partículas generadas en las colisiones de protones, sólo una pequeña parte sufre procesos que pueden ser sensibles a nuevas partículas o interacciones y que resultan interesantes para su posterior análisis. Para mantener la cantidad de datos a un nivel manejable, el sistema de *trigger* realiza una selección en tiempo real de los procesos que tienen lugar en las partículas generadas en las colisiones de protones. El trigger decide si estos procesos son de interés para su estudio posterior por ser sensibles a la existencia de nuevas partículas o si, por el contrario, no se consideran relevantes y se descartan. Únicamente los procesos clasificados como interesantes se guardan en el disco para ser analizados posteriormente. En el caso del análisis de las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$, utilizando el detector LHCb, el sistema de trigger está optimizado para almacenar los eventos que contienen desintegraciones de partículas compatibles con lo que cabría esperar de los procesos $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$. No obstante, la selección realizada por el trigger no es completamente perfecta y algunas de las desintegraciones de estos procesos podría no pasar esta selección y perderse. Por lo tanto, es necesario conocer la eficiencia del trigger para seleccionar las desintegraciones mencionadas y tener en cuenta este valor en la medición de sus fracciones de ramificación. La evaluación de la eficiencia del trigger se realiza utilizando una combinación de muestras de simulación y datos reales recogidos por el LHCb y es uno de los temas principales de esta tesis.

Para que el trigger pueda seleccionar correctamente las desintegraciones que estudiamos, es necesario comprender cómo se producen en el detector. Estos procesos de desintegración dejan una marca en el detector relativamente limpia, mostrada en el esquema de la Fig. 7, facilita su detección. Antes de desintegrarse en dos muones de carga opuesta, el hadrón B_s^0 o B^0 producido en la colisión entre protones se propaga dentro del detector unos milímetros. Los muones interaccionan con diferentes partes del detector, dejando un rastro de su trayectoria, y son las únicas partículas que llegan a la parte más externa del detector, permitiendo su clara identificación. Dado que los muones son partículas cargadas, sus trayectorias se desvián de una línea recta por el efecto del imán dipolar con el que cuenta el detector. A partir de las trazas que dejan en distintas partes del detector, sus trayectorias se reconstruyen con gran precisión y se determinan sus cargas. El vértice en el que el hadrón

B_s^0 o B^0 se desintegra se obtiene a partir de la extrapolación de las trayectorias de los dos muones hasta el punto en el que las mismas confluyen.

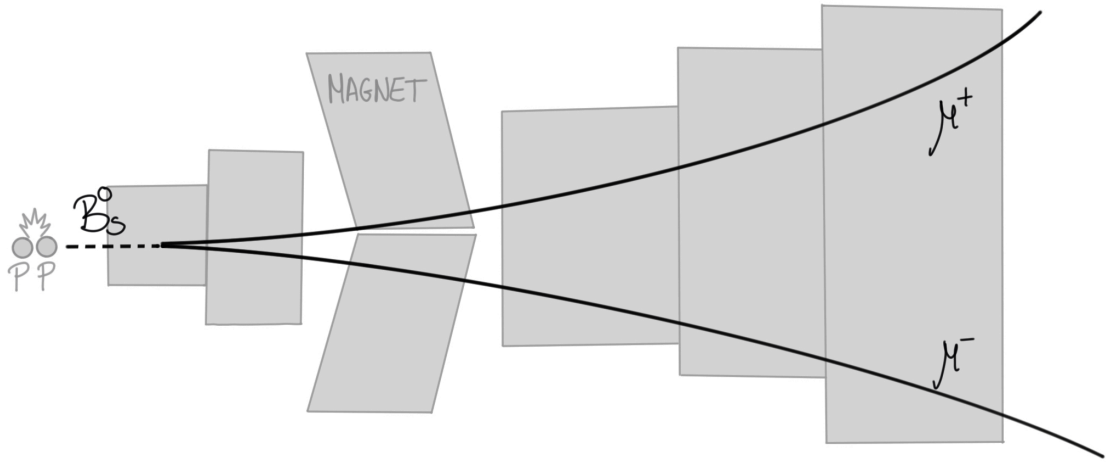


Figure 7: Ilustración del proceso de desintegración $B_s^0 \to \mu^+\mu^-$ en el detector LHCb. El hadron B_s^0 se origina en la colisión de protones y se propaga unos pocos milímetros dentro del detector antes de desintegrarse en dos muones, que son las únicas partículas que llegan a la parte más externa del detector. Las trayectorias de los muones son desviadas por el imán según su carga eléctica.

A pesar de que la traza y señal dejada por las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$ en el detector es muy clara, diferentes tipos de procesos, conocidos como procesos de fondo, podrían ser erróneamente identificados como $B_s^0 \to \mu^+\mu^-$ o $B^0 \to \mu^+\mu^-$. Debido a que estos procesos de fondo son alrededor de $10^7 - 10^8$ veces más probables que las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$, uno de los objetivos fundamentales del análisis es discriminar entre las señales, dadas por $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$, y los procesos de fondo. Esta distinción se consigue a partir del uso de diferentes variables cuyo comportamiento varía para la señal que nos intersa y los procesos de fondo. Una de esas variables es la masa invariante de los dos muones presentes en la desintegración, construida a partir de la combinación de la energía y el momento[1] de los dos muones. De forma explicativa, la ilustración en Fig. 8 muestra el comportamiento en la masa invariante para el caso simple en el que hay una única señal: mientras que la mayoría de procesos de fondo se distribuyen exponencialmente a lo largo de la masa invariante, la señal se manifiesta en forma de pico o protuberancia en estas distribuciones. La altura del pico corresponde al número de señales obtenidas en el análisis, en nuestro caso particular las desintegraciones $B_s^0 \to \mu^+\mu^-$ o $B^0 \to \mu^+\mu^-$, y está directamente relacionada con la cantidad que nos interesa en última instancia: la fracción de ramificación de dicha desintegración.

Debido a la rareza de las desintegraciones $B^0 \to \mu^+\mu^-$ y $B_s^0 \to \mu^+\mu^-$, el poder discrimi-

[1]Cantidad obtenida a partir de la masa de la partícula y su velocidad.

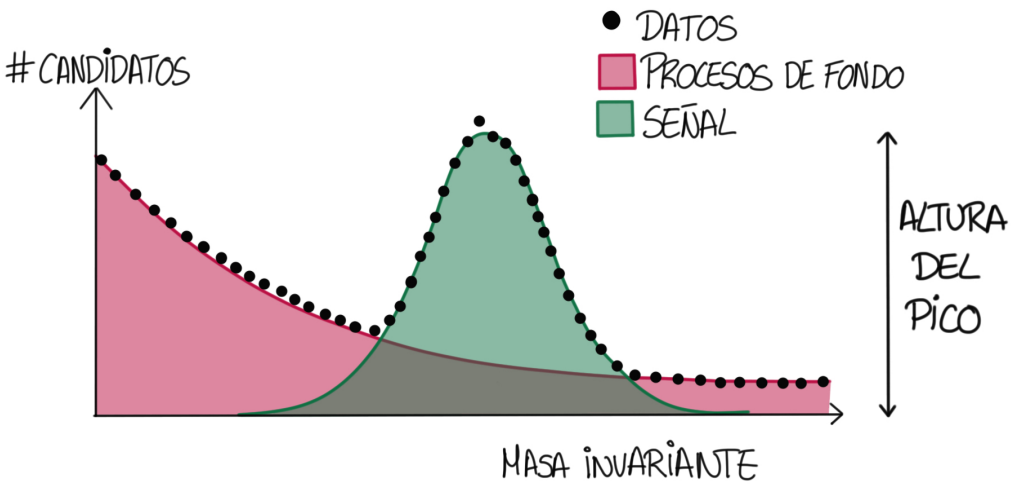


Figure 8: Ilustración del comportamiento de la señal y de los procesos de fondo a lo largo de la masa invariante de los dos muones. En el caso de las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$, la altura del pico de la señal está estrechamente relacionada con la fracción de ramificación de la correspondiente desintegración.

natorio que ofrece la masa invariante de los dos muones es insuficiente para diferenciarlas claramente de los procesos de fondo. Es por ello por lo que la sensibilidad de los datos debe ser maximizada, reduciendo la cantidad de procesos de fondo lo máximo posible sin perder las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$ reales. Para conseguirlo, se utiliza una técnica de aprendizaje automático conocida como Boosted Decision Tree o BDT. El BDT asigna una puntuación entre 0 y 1 a cada uno de los procesos seleccionados y usados en el análisis de forma que, cuanto mayor sea la puntación, mayor es la probabilidad de que el proceso en cuestión sea una desintegración $B^0 \to \mu^+\mu^-$ o $B_s^0 \to \mu^+\mu^-$. Para conseguir una mayor sensibilidad en el análisis, el conjunto total de datos recogidos por el detector LHCb se divide en submuestras según la puntuación dada por el BDT. El análisis de la masa invariante presentado anteriormente se realiza simultáneamente en cada una de estas submuestras. Los resultados obtenidos de cada submuestra se combinan en una única medición de la fracción de ramificación para cada una de las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$. Para poder realizar esta combinación es necesaria una estimación previa de la fracción de $B_s^0 \to \mu^+\mu^-$ o $B^0 \to \mu^+\mu^-$ que se espera en cada submuestra. Esta estimación se obtiene mediante la calibración del BDT, el segundo de los temas principales de esta tesis.

La distribución de todos los procesos seleccionados en este análisis, tanto procesos de fondo como las desintegraciones $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$, se muestra en Fig. 9. La distribución de las desintegraciones $B_s^0 \to \mu^+\mu^-$, mostrada en rojo, tiene un "pico" claramente diferenciado sobre los procesos de fondo, lo que implica que es estadísticamente significativo. A partir de la distribución de $B_s^0 \to \mu^+\mu^-$ y de la altura del pico, se puede obtener su frac-

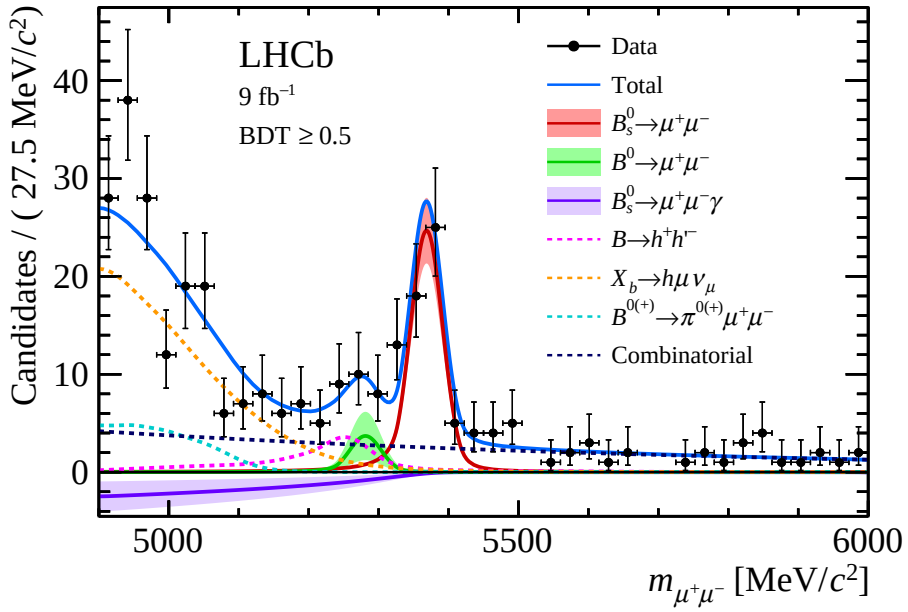


Figure 9: Distribuciones de la masa invariante de $B_s^0 \to \mu^+\mu^-$ y $B^0 \to \mu^+\mu^-$ de los candidatos seleccionados para BDT > 0.5. La línea azul corresponde al modelo de ajuste total, y los componentes separados se indican en la leyenda. Los componentes $B^0 \to \mu^+\mu^-$ y $B_s^0 \to \mu^+\mu^-$ se muestran en verde y rojo, respectivamente.

ción de ramificación, dada por $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$ donde el primer error es estadístico y el segundo sistemático. Esta medida indica que 3.09 de cada mil millones (1,000,000,000) de hadrones B_s^0 se desintegra en dos muones. Dado este valor y los errores asociades a su medida, la fracción de ramificación medida es compatible con la predicción dada por el Modelo Estándar. A pesar de que un pequeño exceso en relación con el los procesos de fondo es visible para las desintegración $B^0 \to \mu^+\mu^-$, mostrado en verde, su protuberancia no se diferencia significativamente del resto de los procesos de fondo, lo que implica que la sensibilidad experimental aún no es suficiente para detectar este proceso de desintegración. Por ello, se establece un límite superior de $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.6 \times 10^{-10}$ a 95% CL para su fracción de desintegración, lo que indica que se espera que menos de 2.6 de cada 10 mil millones (10,000,000,000) de hadrones B^0 se desintegre en dos muones, un valor aún por encima de la sensibilidad necesaria para lograr los valores predichos por el Modelo Estándar. Aún son necesarios más datos para obtener una medida más exacta y precisa de la fracción de desintegración del proceso $B_s^0 \to \mu^+\mu^-$ y para lograr descubrir finalmente la desintegración $B^0 \to \mu^+\mu^-$. Aunque estadísticamente limitados, los resultados presentados en esta tesis limitan el espacio disponible para los modelos desarrollados más allá del Modelo Estándar y restringen los futuros desarrollos teóricos de dichos modelos.

Acknowledgements

The first time I arrived in Amsterdam was the late afternoon of a Sunday. I only had a big suitcase full of clothes and basically zero knowledge of what I was going to do and the people I had to work with. It was my first time in the Netherlands, and I had no idea of the bumpy ride that the PhD was going to be.

First of all, I would like to express my immense gratitude to my supervisor and promotor, Jacco and Marcel. Jacco, I learned an incredible amount of things from you, both work-related and on a personal level. You have always been understanding, supportive and approachable to discuss any problem I had. Thanks for the enormous effort you made, especially at the end of the manuscript revision, to finish as soon as possible and for your always valuable feedback, which helped me improve in many different aspects. You helped me not only to become an independent physicist but also to have enough confidence to express myself freely, always being true to myself. I will never forget how you started singing "Des-pa-ci-to" as soon as you heard me talking in Spanish. Thank you for offering me the possibility to take part in very different experiences during my PhD, from the astronomy on tap at Leiden to the visit with the students to the accelerator. It was a great pleasure working with and learning from you.

Marcel, thank you for your always detailed and honest feedback, not only to this manuscript but to all the work I did. I remember getting the first corrections from you and thinking, "oh my God, this thesis is never going to be over", while you were extra nice, trying to calm me down and cheer me up. Countless are the times you told me "Silvia, it is going to be a great thesis" when I was not even sure if I was going to be able to finish it. I do not know if it ended up being great, but I am glad it gave me a chance to discuss more closely with you. You have been incredibly supportive in the writing phase and always encouraged and supported me to follow my slightly more creative side, even if it implied including some Wally's in the slides. I honestly admire your constant lively energy and your ability to transmit it to everybody in the group. I will miss hearing you shouting "Silviaaaaa!" from your office and your visits to my office to tease me a bit and make me laugh.

Next, I would like to thank my two paranymphs, Cristina and Miriam, for being my family abroad. Cristina, you were literally the third person I met when I arrived at Nikhef,

and I remember thinking, "oh wow, she is very nice. I am sure we will get along well". I was definitely not mistaken. You have been the best partner in crime I could have asked for, and my PhD would have never been the same without you. I cannot imagine my working day without going to your office a minimum of three times to complain about everything and always end up laughing. I will miss our walks around Oost (typically ending up buying some empanadas in the Argentinian place), the nights watching some trashy show (from classy real housewives to Supervivientes), and our immense happiness at parties when they put on that one reggeaton song, and we can finally dance!... I still need to know how you always find the perfect card for each occasion that always brings a big smile to the person receiving it. You are an incredible friend who truly cares about everybody who is around. I admire how strong and brave you are and how you stand up for yourself and the people you love. I am enormously proud of you for everything you have achieved and for overcoming every obstacle you got in your way. Thank you for filling these four years with unforgettable moments together, which bring a smile to my face every time I remember them. I am looking forward to collecting by your side many more memories everywhere we go.

Miriam, we started being office mates and immediately became friends. You were always ready to have a coffee with me after a long meeting and usually managed to cheer me up. I cannot count the times I went to your and Maxime's office seeking advice or just needing to vent, and how you always were there to listen to me. Your moving to Maastricht left a big space in the Amsterdam group that could not be filled with the few calls we had. Thank you for always helping me with everything, from reviewing some chapters of the thesis when I needed an extra opinion to listening to me when I needed to talk. I am grateful you gave me a chance to get to know the wonderful person you are.

During the PhD, I had the immense luck to be part of the bfys group at Nikhef, where I had the opportunity to meet fantastic people. Doing a PhD can sometimes be tricky, and sharing experiences with other PhD students makes it much easier. Lex, thank you for our countless walks+coffee combos at Sarphatipark, for the fun dinners at your place and my place (I genuinely love your kapsalon, we have to meet again) and for listening to me and my dramas. Everything is much more fun when you are around, and I hope we will have much more time to share. Alice, you are not only my shopping partner but a true and understanding friend. Thank you for all our Saturday/Sunday brunches accompanied by excellent chats, for trusting me and for all the time we shared. I wish that, although in the distance, we manage to keep this incredible friendship we built. Lera, you have been my office mate for the longest time and a great one! I will always be grateful to you for taking care of my baby plants as if they were yours, and for coming to rescue Cristina and me in Diemen and guiding us safely to Brían's house (never forget, we owe you a big one). Jordy, you will always be my SciFi body, even though we ended up being dead inside. Thank you for making the cracking comment at the right time to make everybody laugh. Aleksandra, after just arriving from a long trip from the UK, you did not let me down that night and came to have some beers with

Brían and Maarten. That's the spirit! Igor, I admire your determination to fight for your ideals. Carolina, you brought some Brazilian charm to the group. Thank you for always being available to have dinner with me after an intense day of teaching in Maastricht. Please take good care of our radiative tail baby once I leave. Robbert, we had very fruitful discussions where I could always openly talk with you. Jan, thanks for suggesting to have a PhD dinner. It was great and memorable! MD, you are quite a special person who always knows when to make that perfectly unexpected comment, always triggering a smile in everybody. Davide, we did not have much time to coincide in Maastricht or Amsterdam; however, I am grateful for how open and friendly you have been to me every time we have met. Efrén, thank you for coming to my office every day to check how I was doing and if I needed a coffee or a break. Michele, you have a great and unique sense of humour, which gives you that special charm you have. Thank you for giving me a chance to dig a bit under the surface, for your thesis template (which saved me hours of work) and for sharing the best advice I could have gotten: keep writing, no matter what. I still laugh when I remember that day when I broke some channels of a SciFi module and you went completely pale. I really miss having you around, sending memes and making jokes. Finally, although not directly from our corridor, I would never forget Brían, an essential part of the bfys group. I will miss your messages when I am on holiday asking "are you alive?". You continuously transmit that good vibe of yours in everything you do. Thank you for convincing me to join for beers countless times and always finding some time to meet and catch up, even though you are sometimes as busy as a minister. I promise that, from now on, I will answer the messages in due time!

The postdocs at the bfys group are not only friends but also very inspiring role models. Andrea Mauri, you happened to be a party animal and always fun to be around. I wish I had had the chance to get to know you earlier. Emmy, thank you for the times we had coffee together, especially the day when I had the chance to meet baby Leo, and for sharing your experiences with me. Suzanne, I would have loved to supervise a student with you. You were very inspiring and made me realise how cool giving outreach talks at schools can be. Andrii, you always had been friendly and helpful to me, and I admire your serenity and honesty. Maarten, you are not only usually in for a party or some beers, but also your door is always open if help is needed. Andrea García, I know you are not part of the bfys group, but I would also like to thank you for sharing your lively way of seeing life with me.

Mick, you deserve a special mention: this thesis would have never been possible without you. You were my daily supervisor at the beginning of the PhD and passed me all the knowledge I needed to perform the BDT calibration (and even more!). I really enjoyed our discussions about physics, which generally consisted in brainstorming and getting several new ideas to try. Thank you for giving me the confidence to ask all types of "silly" questions that came to my mind. You are very kind and extremely patient, always answering with a smile the same question I had already asked a hundred times. But more importantly, you genuinely care about the people you work with. I cannot count the times you came to my

office to check how I was doing and if I was stressed or stuck. Even during lockdown, you managed to keep checking on me virtually, especially at the very stressful moment of wrapping up the analysis. There are not enough pages to express my gratitude for all you have done. Además, estoy muy orgullosa de que estés aprendiendo español tan rápido. Muchas gracias por invitarnos a la fiesta de las enmoladas y por darme la oportunidad de compartir contigo tiempo fuera del trabajo. Espero que Naye y tú vengáis pronto a España a visitarme y así poner en práctica todo lo que estás aprendiendo.

The backbone of the bfys group at Nikhef is the senior staff who keep this group alive and full of amazing people. Roel, thank you for showing me that Dutch people can have good musical taste and even dance salsa! Gerhard, thank you for always finding the right question to ask in all presentations, triggering interesting discussions that lead to good ideas. Patrick, thank you for producing the nice anomaly plots, which help to see where we are in the search for New Physics with just a glance, and for your outreach contribution on Twitter. Kristoff, thank you for being the cameraman in the bfys outing and do that with a smile. Chris, I admire your ability to party one night and be completely fine the next day. Kazu, thank you for your awkward sense of humour. Tjeerd, thank you for always having a smile and for showing us what really enjoying your work means. Niels, thank you for your support as physics coordinator to finish the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis. Special mention to Wouter for having the nice detail to include two new plants to my collection after one died when I returned from the summer holidays. Mara, you were a real inspiration to me in starting this adventure as you were the first one to suggest I could apply to Nikhef for a PhD. You were right; it was a great time. Thank you for your always honest career advice and for having your door open for whatever we need. Keri, I am glad I had the opportunity to work closely with you. Thank you for always finding an answer to my most theoretical questions and for checking on me when things started to get stressful. I very much enjoyed our discussions with Carolina and Marcel, which always resulted in new excellent ideas to try. Daniel Cámpora, you brought some more Spanish energy to the group and, together with Vero, gave us the chance to discover swing. Antonio, do you remember that day you told me if I broke the vacuum machine, you would deduct the thousands of euros it costs from my monthly salary? In the end, I managed not to break it! I will miss having you around shouting, "Silvia, dónde estás!" to come to my office looking for either cashew nuts or a chat, always making me laugh. Thank you for making my time here much more fun and for caring about me. However, I will never forget I had to be your Secret Santa to get the Delizia di Limone I requested for half a year.

The worst part of being in such an incredible group is that people finishing their time here leave a huge gap that cannot be filled. Elena, although at the end of your thesis, I am glad for the time we could share. Even though you were very busy (now I can understand how you felt!), you always found some time to chat after lunch while eating the chocolate bars I was safely (?) keeping in my drawer. Thank you for making the beginning of the PhD

much easier. Sean and Katya, you were my first office mates when I started the PhD, and you received me with open arms. Daniel Hynds, I cannot believe I have never seen you wearing a T-shirt; that is a real commitment to elegance. Roman, working with you in SciFi was a pleasure, not only because you were extremely patient with us but also because we always got some chocolate after. Hilbrand, you were an essential one in all the parties and plans; we miss having you around. Carlos, thank you for constantly convincing people to grab some beers after work and for enlivening any party.

Maxime, I admire how calm and peaceful you are and how you transmit this to everyone around you. I literally went to your office every Thursday after the long analysis meetings to get a break and get some stress out, and you always had good advice to offer me. All the collaboration should be grateful to you for the big effort you have made to destroy the wrong LHCb logo. Thank you for giving us a chance to share with you and Marie-Cèline dinners and brunches and get to know both of you much better. You are amazing people, and I hope we can continue sharing such experiences.

Apart from the people at Nikhef, I had the chance to be part of the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ analysis team, which is formed by people I deeply admire. Flavio, thanks for trusting me from the beginning and for constantly acknowledging my work, even when I felt no real progress was achieved. You were always available for questions and willing to help when things were not working. Matteo Rama, you transmit calmness and serenity, even when wrapping up the analysis. Francesco, thank you for your help with the trigger efficiencies and for your quick and always friendly answers to all questions, also during the writing process. I really learned a lot from you. Marco, thank you for always sharing your work with me and for the feedback you gave me in all the meetings. Matteo Palutan, it has been a great honour that you accepted to be part of the committee to read my thesis. Harry we did not work very closely, but thanks for producing the lifetime corrections for the 8 BDT binning scheme to put it in my thesis! And, of course, I cannot forget the PhD crew. Our solo PhD meetings during the lockdown were very helpful, especially during the difficult time we were living. I wish we had met at CERN to share more experiences. Lauren, thank you for your lively energy in all the solo PhD meetings. Maik, we had to work closely, and it was always easy and pleasant to talk with you and discuss about our work. Ifan, thank you for your being the best stripping liaisonship partner and for your patience in explaining to me how things work, especially during the restripping campaigns. Titus, thank you for helping with the BDT calibration and answering my countless questions about the CLs method. Fabio, although you were a master's student, you did a fantastic job. I am sure all of you will have a promising career wherever you go.

The PhD is also an experience to be shared with people outside the working environment. Sara Pirroni, we managed to keep this unique friendship through messages and calls (and some trips to Sardinia and Amsterdam in between). Thank you for making me part of your life and for giving me the chance to share our PhD journeys closely, even at the dis-

tance. Ruska, gracias por compartir con nosotros tu creatividad y talento para hacer esta maravillosa portada. Gracias por regalarme un momento de inspiración y diversión que me ayudo a sacar la mente de este mundo tan técnico y tomar aire para poder continuar. Daniela, gracias por enseñarme a ser más paciente y amable conmigo misma, por ayudarme a bajar revoluciones con un "está bien" y por siempre poner facilidades y soluciones a los problemas. Admiro mucho la gran capacidad de trabajo que tienes, tu gran profesionalidad y lo ricos que te salen los bizcochos de tres leches.

A mis chicas de "Sex & the city", por compartir siempre todos los momentos de nuestras vidas, haciendo de las alegrías una auténtica fiesta y las penas más llevaderas. Aunque estemos desperdigadas por el mundo siempre os siento cerca. Os admiro a cada una de vosotras por las cosas que os hacen únicas y es un orgullo enorme teneros en mi vida y poder seguir creciendo juntas. Lara, Carla y Sofía, me gustaría agradeceros en especial a vosotras por ser un apoyo incondicional y por lo pendiente que habéis estado de mí, especialmente este último año. El viaje a París (espero que repitamos pronto) fue reparador y una recarga de energía brutal para ayudarme a centrarme y escribir. Me siento muy afortunada de teneros y de poder compartir con vosotras cada nueva aventura de nuestras vidas.

Blanca, Sara y Cris, hacéis que los cuatro o cinco meses que llevamos sin vernos parezcan solo unos cuantos días cuando estamos juntas. Gracias por esas quedadas por las calles de Madrid que siempre empiezan a las nueve de la noche y, entre charla y charla, acaban a las dos de la madrugada (y porque nos cierran el bar). Gracias por aceptarme tal cual soy, por hacerme sentir libre para explicar como me siento y por ayudarme a hacer autocrítica. Sara, estoy enormemente orgullosa de lo que has conseguido, tanto a nivel profesional (¡ser una doctora de verdad!) como a nivel personal. Eres un ejemplo de perserverancia, tenacidad y buen hacer. Blanca, gracias infinitas por hacer la mejor y más detallada correción que se podía hacer del resumen de esta tesis. Pero, sobre todo, gracias por apoyarme desde la distancia y por compartir conmigo reflexiones y todo el conocimiento brutal que tienes. Me encanta escucharte hablar y espero, en algún momento, conseguir expresarme sólo una cuarta parte de lo bien que te expresas tú. Es un placer veros crecer a las tres, y poder ser libres para compartir pensamientos, reflexiones y experiencias.

A mis padres, por enseñarme que el esfuerzo merece la pena y darme la confianza para ser quien quiero ser. Por estar cada día presentes, incluso desde la distancia, y ayudarme a ver las cosas desde otra perspectiva cuando me pongo cabezota. Gracias por enseñarnos a ser independientes y hacer de la familia un núcleo indestructible que nos impulsa hacia adelante pero que, a la vez, es el refugio donde siempre podemos volver cuando las cosas no van tan bien. Creo que nunca podré agradeceros lo suficiente el que ha sido vuesto mejor regalo: mi hermano. Keko, gracias por los abrazos de oso, por demostrarme tu apoyo incondicional y por enseñarme lo que realmente son el esfuerzo y la perseverancia. Admiro profundamente tu gran bondad y sensibilidad respecto hacia las personas a las que quieres y tu gran capacidad de continuar siempre hacia adelante, a pesar de todos los obstáculos que

te encuentras en el camino. Estoy muy orgullosa de la gran persona que eres y en la que te estás conviertiendo y deseando compartir contigo todos los grandes logros que te esperan a la vuelta de la esquina. Os pienso todos los días y os quiero infinito y más allá.

The (almost) last line goes to Daniel, who gives me the wings to fly, heals them when they get broken and gives me peace when I need it but also strength and courage to continue fighting for what I want. I guess I needed around a million power hugs and some dancing in the kitchen to finish this PhD. You know that this thesis is also a bit yours. Gracias simplemente por ser, por estar y por siempre conseguir hacerme reír.

Finally, I would like to thank myself, not only for finishing this thesis but also for how much I have personally grown during these four years. You must be doing something good to have such amazing people surrounding you. So remember: do not underestimate yourself and just enjoy the ride; many more incredible experiences are yet to come.

Curriculum vitae

Silvia Ferreres Solé was born in 1994 in Madrid, Spain. From 2012 to 2016, she attended the Universidad Complutense de Madrid, where she obtained her Bachelor's degree in Physics in 2016. After that, she enrolled in the MSc degree in Theoretical Physics at Lund University, where she graduated with distinction in 2018. During the summer of 2017, she took part in the Summer Student Programme at CERN, where she worked checking the performance of the muonic high-level trigger at the CMS experiment for three months.

In 2018, she started her PhD in experimental particle physics in the LHCb group at the Nikhef Institute, associated with Maastricht University. As part of her PhD, she presented her work at two high-profile conferences, Rencontres de Moriond in 2020 and Flavour Physics and CP violation in 2022, and took tutor roles in several courses at both Bachelor and Master levels.

The beauty of the rare

Silvia Ferreres Solé