Study of rare leptonic B meson decays with the LHCb experiment

Justine Serrano

AIX-MARSEILLE UNIVERSITÉ

ECOLE DOCTORALE 352

UMR 7346

Habilitation à Diriger les Recherches

Study of rare leptonic B meson decays with the LHCb experiment

Justine Serrano

Soutenue le 28 novembre 2019 devant le jury composé de:

Sébastien Descotes-Genon	Examinateur
Cristinel Diaconu	Tuteur
Francesco Forti	Rapporteur
Tim Gershon	Examinateur
Francois Le Diberder	Rapporteur
Lydia Roos	Rapporteuse

Contents

Preamble

This document presents my research activities in the last ten years, since I joined the LHCb collaboration. I obtained my PhD in 2008 at the Laboratoire de l'Accelerateur Linéaire in Orsay, studying charm semileptonic decays with the BaBar experiment. In october 2008, I moved to the Centre de Physique des Particules de Marseille in the LHCb team as a post-doctorate, and then as a chargé de recherche, to work on the very rare B decays. This activity is part of the global effort for the search of a physics beyond the Standard Model.

I first joined the search for the 'golden channel' $B_s^0 \to \mu^+\mu^-$. In a highly competitive environment, it has been a human, scientific and sociological adventure that last for five years, and materialized into five publications, including the first evidence of the $B_s^0 \to \mu^+\mu^-$ decay published in 2013. Combined with the CMS data, the Run1 analysis lead to the first observation of this decay, published in Nature.

In 2013, I started to tackle a new challenge : the search for rare B decays into τ leptons. These decays was thought to be impossible to study in a hadronic environment with a detector such as LHCb. After years of tries, failures, brainstorming, strategy development, hope and despair, the first experimental limit on the $B_s^0 \to \tau^+\tau^-$ channel was published in 2017.

This document describes the analyses of the $B^0_{d,s}\to \mu^+\mu^-$ and $B^0_{d,s}\to \tau^+\tau^-$ decays. While the former was performed by an international collaboration of about 15 physicists, the latter was done entirely at CPPM. The work presented here was performed in collaboration with several people, the first one being Giampiero Mancinelli, who started the rare B decay activity in Marseille. On the $B^0_{d,s}\to \mu^+\mu^-$ analysis, I worked in collaboration with three PhD students (Cosme Adrover, Mathieu Perrin-Terrin, and Alessandro Mordà). On the $B^0_{d,s}\to \tau^+\tau^-$ study, I supervised the post doctorate of Kristof de Bruyn, financed with my ANR grant 'Sorbet', and part of the PhD of Cédric Méaux, financed with a grant from the Labex OCEVU. During these ten years, I also supervised six undergraduate students.

Beside my activity on data analysis, I also participated to the data taking of LHCb, as a 'piquet' (on call expert) for the Level-0 trigger and shift leader. I was also appointed member of the LHCb editorial board for two years, and member of the *Heavy Flavour Averaging Group* (HFLAV) since 2015.

The rest of the document is organized as follows: Chapter 1 introduces the motivations and theoretical framework for the study of $B^0_{d,s}\to \ell^+\ell^-$ decays. Experimental analysis of $B^0_{d,s}\to \mu^+\mu^-$ and $B^0_{d,s}\to \tau^+\tau^-$ are described in Chapter 2 and 3, respectively. A conclusion presenting interpretations of the measurements and my personal projects is given in Chapter 4.

Chapter 1

Rare leptonic B decays : why and where

If the standard model (SM) of particle physics has been tremendously successful in explaining experimental data predicting the existence of several particles that were indeed observed, it is unable to answer some of the most important questions of fundamental physics as the nature of dark matter and dark energy, the origin of the strong mass hierarchy between the three families, or the matter/antimatter asymmetry in the Universe. A large consensus exists that the SM is not the most fundamental theory and that "new physics" (NP) should manifest itself at higher energies. Rare decays of B mesons into leptons constitute very powerful tool to look for NP. They can be classified into two categories: flavour changing neutral current (FCNC) decays, that correspond to the transition from a down-type quark (or up-type quark) to another one of same type but different flavour, and lepton flavour/number violating decays. The first ones are forbidden at tree level in the SM. They can only proceed via loop diagrams, involving non-diagonal CKM matrix elements, and are therefore suppressed with respect to tree-level mediated transitions, resulting in branching fractions ($\mathcal{B}$) of the order of 10^{-5} or lower. The FCNCs are sensitive probes of physics beyond the SM since the virtual SM particles present in the loops could be replaced by new physics particles, modifying properties of the decay such as branching fractions or angular observables. The decays of the second category are completely forbidden in the SM[1], providing a unambiguous manifestation of NP in case of observation.

The work presented in this document focuses on the FCNC $b \to d/s\ell^+\ell^-$ transitions, which dominantly proceed through the diagrams shown in Fig. 1.1. Section 1.1 details the theoretical aspects of these decays, including the SM predictions and sensitivity to NP, while Sec. 1.2 presents the experimental picture.

1.1 Theoretical aspects of $B^0_{d,s} \to \ell^+\ell^-$ decays

1.1.1 Phenomenology of rare B decays

Theoretical framework: the effective Hamiltonian approach

A very powerful tool to describe the FCNC decays of b-hadrons in a model-independent way is the effective Hamiltonian approach (see $e.g.$ [1]). The idea behind it is to separate the different scales intervening in the transitions: the electroweak scale, $M_W \sim 80$ GeV, that is responsible

[1] contribution from neutrino oscillations leads to branching fractions $< 10^{-40}$

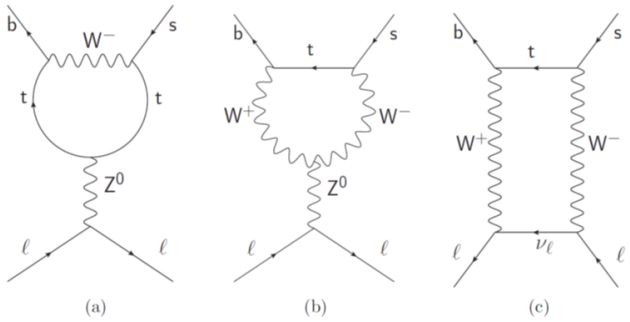

Figure 1.1: Dominant Feynman diagrams contributing to the $b \to s\ell^+\ell^-$ transitions in the Standard Model: Z penguin (a,b) and W box (c) diagrams.

for the change in flavour and the QCD scale, $\Lambda_{QCD} \sim 0.2$ GeV, responsible for the dynamic of the external states. The b quark scale, $m_b \sim 4$ GeV, is in between. From the point of view of the QCD and b quark scale, the weak interaction can be seen as point-like. An *operator product expansion* (OPE) allows to 'factorize' the long distance effects encoded in operators $\mathcal{O}_i$ and the short distance ones, proportional to the so-called Wilson coefficients C_i. While the operators are governed by non-perturbative theory and are usually computed by lattice QCD or light cone sum rules techniques, the Wilson coefficients can be computed perturbalively. Technically, these computations are made at a high scale $\mu \sim M_{W/t}$ and extrapolated to the relevant scale (here, $\sim m_b$) using the renormalization group equations. Based on the OPE, the amplitude for a $B \to f$ transition takes the form

$$A(B \to f) = \langle f | \mathcal{H}_{eff} | B \rangle = \frac{4G_F}{\sqrt{2}} \sum_i \lambda C_i(\mu_b) \langle f | \mathcal{O}_i(\mu_b) | B \rangle , \tag{1.1}$$

where $\mathcal{H}_{eff}$ is the effective Hamiltonian, λ is a CKM factor and $\langle f | \mathcal{O}_i(\mu_b) | B \rangle$ the matrix elements of the operators at the scale of the b quark. The general effective Hamiltonian governing rare B decays in the SM can be written:

$$\mathcal{H}_{eff} = \frac{4G_F}{\sqrt{2}} \left(\lambda_u^{(q)} \sum_{i=1}^{2} C_i \mathcal{O}_i^u + \lambda_c^{(q)} \sum_{i=1}^{2} C_i \mathcal{O}_i^c - \lambda_t^{(q)} \sum_{i=3}^{10} C_i \mathcal{O}_i - \lambda_t^{(q)} C_\nu \mathcal{O}_\nu + h.c \right), \tag{1.2}$$

where $q = d, s$ for processes based on the quark level $b \to d, s$ transitions and $\lambda_p^{(q)} = V_{pb} V_{pq}^*$. The operators $\mathcal{O}_{1,2}$ correspond to the current-current operators (Fig. 1.2(a)), $\mathcal{O}_{3-6}$ to the QCD penguin operators (Fig. 1.2(b)), $\mathcal{O}_{7,8}$ the electromagnetic and chromomagnetic operators (Fig. 1.2(c,d)) and $\mathcal{O}_{9,10,\nu}$ to the semileptonic operators (Fig. 1.2(e)). We can notice that due to the CKM matrix hierarchy, the $b \to d$ transitions are suppressed with respect to the $b \to s$ ones by $|V_{td}/V_{ts}|^2 \sim 0.22^2$. This is why the $b \to s$ transitions are of higher interest. We can also note that in $b \to s$ decays, $\lambda_{t,c}^{(s)} \gg \lambda_u^{(s)}$ and $\lambda_c^{(s)} \sim -\lambda_t^{(s)}$. The terms involving $\mathcal{O}_{1,2}^{(u)}$ can thus be neglected, which is not the case for $b \to d$ decays.

New physics contributions can affect the effective Hamiltonian in two ways: modifying the Wilson coefficients or generating new operators. These latters could be $\mathcal{O}'$ operators corresponding to chirality flipped operator (right-handed current), lepton-flavour-dependent operators $(\mathcal{O}_{9,10,\nu})^\ell$ in case of lepton flavour universality violation, scalar or pseudoscalar operators involving two quarks and two leptons, $\mathcal{O}_{\mathcal{S},\mathcal{P}}$, and lepton flavour violating operators $\mathcal{O}_9^{\ell_i \ell_j}$.

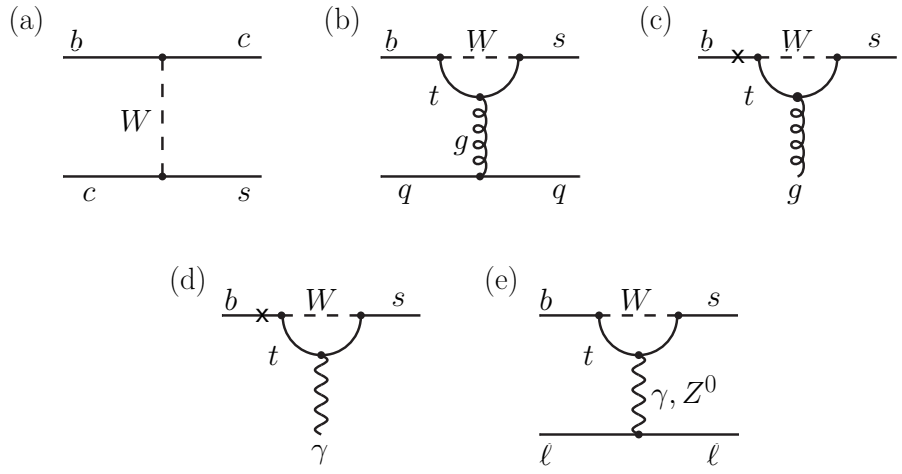

Figure 1.2: Example of tree and penguins diagrams contributing to $b \to s\ell\ell$ transitions.

1.1.2 The case of $B^0_{d,s} \to \ell^+\ell^-$ decays

Standard Model predictions

Within the framework of the effective Hamiltonian approach defined in Eq. (1.2), the operators contributing to the $B^0_q \to \ell^+\ell^-$ decays are the following :

$$\mathcal{O}_{10}^{\ell(\prime)} = \frac{e^2}{16\pi^2}(\bar{b}\gamma^\mu P_{L(R)} q)(\bar{\ell}\gamma_\mu\gamma^5\ell), \tag{1.3}$$

$$\mathcal{O}_{S}^{\ell(\prime)} = \frac{e^2}{16\pi^2} m_b(\bar{b} P_{L(R)} q)(\bar{\ell}\ell), \tag{1.4}$$

$$\mathcal{O}_{P}^{\ell(\prime)} = \frac{e^2}{16\pi^2} m_b(\bar{b} P_{L(R)} q)(\bar{\ell}\gamma^5\ell), \tag{1.5}$$

where the L and R indices refer to left- and righ-handed chiralities of the fermion. In the SM, only the C_{10} coefficient is non negligible. The branching fraction can then be expressed as:

$$\mathcal{BR}(B^0_q \to \ell^+\ell^-)_{SM} = \tau_{B_q} \frac{G_F^2 \alpha_{em}^2}{16\pi^2} f_{B_q}^2 m_\ell^2 m_{B_q} \sqrt{1 - \frac{4m_\ell^2}{m_{B_q}^2}} |V_{tb}V_{tq}^*|^2 |C_{10}^{\ell SM}|^2, \tag{1.6}$$

where

- τ_{B_q}, and m_{B_q} are the B_q meson lifetime and mass,
- α_{em} is the electromagnetic constant,
- m_ℓ is the mass of the final state lepton,
- f_{B_q} is the B_q meson decay contant.

The only non-perturbative quantity entering this expression is f_{B_q}. It is computed in numerical simulation of QCD on the lattice and is now known with a precision at the percent level [2]. It is an advantage of the purely leptonic decays on the semileptonic decays since the latter implies

calculation of form factors which are q^2 dependent and can hardly be computed on the lattice for low-q^2 values. We can see that due to the m_ℓ^2 factor, the branching fraction vanishes for massless leptons. This is the consequence of the helicity suppression, coming from the decay of a 0-spin particle to two spin-1/2 daughters.

The branching fraction expressed in Eq.(1.6) is a theoretical value that does not account for the fact that B mesons oscillate and is thus only valid at $t = 0$. Due to the sizeable lifetime difference between the light and heavy B_s mass eigentate ($\Gamma_{sL} - \Gamma_{sH} = 0.090 \pm 0.005$ ps^{-1} [3]), the time-integrated branching fraction of the B_s decay, $\bar{\mathcal{B}}$, should be expressed as [4]:

$$\bar{\mathcal{B}}(B_s^0 \to \ell^+\ell^-)_{SM} = \frac{1 + y_s \mathcal{A}_{\Delta\Gamma}}{1 - y_s^2} \mathcal{BR}(B_s^0 \to \ell^+\ell^-)_{SM}, \tag{1.7}$$

with

$$y_s = \frac{\Gamma_{sL} - \Gamma_{sH}}{\Gamma_{sL} + \Gamma_{sH}} = 0.065 \pm 0.005 \text{ [3]}, \tag{1.8}$$

and

$$\mathcal{A}_{\Delta\Gamma} = \frac{\Gamma_{B^0_{s,H} \to \ell^+\ell^-} - \Gamma_{B^0_{s,L} \to \ell^+\ell^-}}{\Gamma_{B^0_{s,H} \to \ell^+\ell^-} + \Gamma_{B^0_{s,L} \to \ell^+\ell^-}}, \tag{1.9}$$

where $\Gamma_{B^0_{s,L/H} \to \ell^+\ell^-}$ is the partial decay width of the light and heavy mass eigenstates. While it is predicted to be 1 in the SM, the value of $\mathcal{A}_{\Delta\Gamma}$ can be measured through the effective lifetime of $B_s^0 \to \ell^+\ell^-$ [4],

$$\tau_{\ell^+\ell^-} = \frac{\int_0^\infty t\Gamma(B_s^0(t) \to \ell^+\ell^-)dt}{\int_0^\infty \Gamma(B_s^0(t) \to \ell^+\ell^-)dt}, \tag{1.10}$$

using the relation,

$$y_s \mathcal{A}_{\Delta\Gamma} = \frac{(1 - y_s^2)\tau_{\ell^+\ell^-} - (1 + y_s^2)\tau_{B_s}}{2\tau_{B_s} - (1 - y_s^2)\tau_{\ell^+\ell^-}}. \tag{1.11}$$

The evaluation of branching fraction values in the SM requires the calculation of $C_{10}^{\ell SM}$ that corresponds to the Z and W contributions shown in Fig. 1.1. The SM predictions that will be used for reference in the rest of this document are taken from Ref. [5]. They include next-to-leading order electroweak [6] and next-to-next-to-leading order QCD [7] corrections. For the B_d^0 decays, they have a precision of 8.5%:

$$\mathcal{B}(B_d^0 \to e^+e^-) = (2.48 \pm 0.21) \times 10^{-15}, \tag{1.12a}$$
$$\mathcal{B}(B_d^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}, \tag{1.12b}$$
$$\mathcal{B}(B_d^0 \to \tau^+\tau^-) = (2.22 \pm 0.19) \times 10^{-8}. \tag{1.12c}$$

For the B_s^0, the precision is 6.3% :

$$\bar{\mathcal{B}}(B_s^0 \to e^+e^-) = (8.54 \pm 0.55) \times 10^{-14}, \tag{1.13a}$$
$$\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9}, \tag{1.13b}$$
$$\bar{\mathcal{B}}(B_s^0 \to \tau^+\tau^-) = (7.73 \pm 0.49) \times 10^{-7}. \tag{1.13c}$$

The largest sources of uncertainties entering these predictions are :

- The hadronic decay constants, which are computed on the lattice: $f_{B_s} = 227.7 \pm 4.5$ MeV and $f_{B_d} = 190.5 \pm 4.2$ MeV [8]. They account for 4.0 and 4.5% of the B_s^0 and B_d^0 uncertainty, respectively.

- The CKM matrix elements, $|V_{tb}^* V_{ts}/V_{cb}| = 0.980 \pm 0.001$, $|V_{cb}| = 0.0424 \pm 0.0009$ and $|V_{tb}^* V_{td}| = 0.0088 \pm 0.0003$ [9, 10]. They account for 6.9 and 4.3% of the B_d^0 and B_s^0 uncertainty, respectively.

- The top quark mass $M_t = 173.1 \pm 9$ GeV [11], which accounts for 1.6%.

Non-parametric uncertainties due to higher order corrections in α_s^3, α_{em}^2, $\alpha_s \alpha_{em}$ and higher order dimension operators m_b^2/m_W^2 in the effective langrangian are estimated to be about 1.5%. Interestingly, the value of $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ has been recently updated to include electromagnetic corrections below the b quark mass scale, which have been found larger than foreseen. The new SM prediction is $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = (3.57 \pm 0.17) \times 10^{-9}$ [12]. For consistency in the rest of the document, we will keep using the values from Ref. [5].

Sensitivity to NP

In case where contributions from particles beyond the SM are allowed, the $|C_{10}^{\ell SM}|^2$ factor of Eq.(1.6) is replaced by

$$|S|^2(1 - \frac{4m_\ell^2}{m_{B_q}^2}) + |P|^2, \tag{1.14}$$

where

$$S = \frac{m_{B_q}^2}{2m_\ell}(C_S^\ell - C_S^{\prime\ell}), \text{ and } P = (C_{10}^\ell - C_{10}^{\prime\ell}) + \frac{m_{B_q}^2}{2m_\ell}(C_P^\ell - C_P^{\prime\ell}). \tag{1.15}$$

From these equations, one can see that $C_{10}^{(\prime)\ell}$ is affected by the helicity suppression factor m_ℓ/m_{B_q}, so that the branching fraction vanish for massless leptons. This is not the case for the scalar and pseudoscalar contributions. The branching fraction is therefore very sensitive to these operators and can be largely enhanced by contributions of $C_S^{(\prime)\ell}$ and $C_P^{(\prime)\ell}$. It is actually a unique property of the $B_{d,s}^0 \to \ell^+\ell^-$ decay to be stronlgy helicity suppressed in the SM but not in the presence of NP scalar operators, making it a 'golden channel' to search for new physics. In particular, in supersymmetric (SUSY) extensions of the SM, the exchange of a heavy Higgs would give a scalar amplitude proportional to $tan^3\beta$, where $tan\beta$ is the ratio of vacuum expectation values [13]. Extended Higgs sector without SUSY could also gives a scalar amplitude enhanced by $tan^2\beta$. The $B_{d,s}^0 \to \ell^+\ell^-$ branching fractions can therefore be several oder of magnitude higher than in the SM. One may note that the pseudoscalar contribution can also suppress the branching fraction in case of negative interference with C_{10}^ℓ.

If the Wilson coefficients $C_{S,P}^{(\prime)}$ are in most scenarios independent, in the context of the SM effective field theory (SMEFT) [14], they are related by [15] $C_S = -C_P$ and $C_S' = -C_P'$. The SMEFT is justified by the fact that, given the absence of NP signs at the LHC, the NP scale should be much larger than the electroweak scale. In that case, the extended Lagrangian encoding NP terms is invariant under the full SM gauge symmetry, which allows to reduce the number of free parameters.

In case of NP, $\mathcal{A}_{\Delta\Gamma}$ is expressed as

$$\mathcal{A}_{\Delta\Gamma} = \frac{Re(P^2 - S^2)}{|P|^2 + |S|^2}. \tag{1.16}$$

The measurement of the branching fraction and $\mathcal{A}_{\Delta\Gamma}$ can therefore provide complementary informations, as illustrated in Fig. 1.3. Additional constraints could also be obtained from the

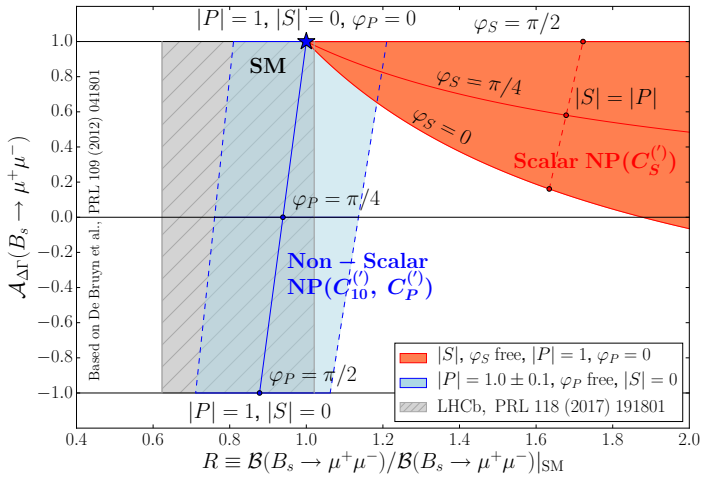

Figure 1.3: Illustration of new physics scalar and pseudoscalar contributions in the plane $\mathcal{A}_{\Delta\Gamma}$ vs $\mathcal{B}(B_s^0 \to \mu^+\mu^-)/\mathcal{B}_{SM}(B_s^0 \to \mu^+\mu^-)$. S and P are expressed here as $|P|e^{i\phi_P}$ and $|S|e^{i\phi_S}$, respectively. The hatched area corresponds to constraint from the latest LHCb measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [16].

mixing-induced CP asymmetry $S_{\ell^+\ell^-}$ [17],

$$S_{\ell^+\ell^-} = \frac{Im(P^2 - S^2)}{|P|^2 + |S|^2}, \tag{1.17}$$

but this requires the measurement of the time-dependent tagged rate, which would need much more statistics than currently available at the LHC experiments.

An additionnal observable, the ratio of the $B_s^0 \to \ell^+\ell^-$ over $B_d^0 \to \ell^+\ell^-$ can provide a powerful test of the Minimal Flavour Violation (MFV) hypothesis.

1.2 Experimental picture

1.2.1 The $B^0_{d,s} \to \ell^+\ell^-$ saga

The search for $B^0_{d,s} \to \ell^+\ell^-$ decays started in the eighties with the CLEO, ARGUS and UA1 experiments. It then continued at the B-factories, Belle and BaBar, for the B_d^0 modes and at hadronic colliders for both B_s^0 and B_d^0[2]. The former have the advantage of having very clean events and the possibility to reconstruct the missing energy, which makes them well suited to study the tauonic final state. On the other hand, the latter have an incredible statistics of b–hadrons produced and can reconstruct two-body final states with a high efficiency. The experimental picture at the start of the LHC is reported in Table 1.2.1. It was dominated by measurements from the CDF experiment for electronic and muonic final states, and by BaBar for the $B_d^0 \to \tau^+\tau^-$ mode.

BaBar and Belle stopped taking data in 2008 and 2010 respectively, while the Tevatron ran until september 2011. The LHC entered the game in 2010 and by end of 2011 had already reached a peak luminosity almost ten times higher than the Tevatron for a beam energy about 3.6 times larger. The LHC experiments focused on the search for the B_s^0 and $B_d^0 \to \mu^+\mu^-$ decays, rapidly improving the picture.

[2]Belle took data at the Υ(5S) allowing the study of B_s^0 decays, but the recorded statistics was way too small to compete with hadronic collider experiments for these rare decays.

Table 1.1: Best limits on rare dileptonic B decays as reported in PDG 2010[18]. Limits are given at 90% confidence level.

	e^+e^-	$\mu^+\mu^-$	$\tau^+\tau^-$
B_d^0	8.3×10^{-8} [19]	1.5×10^{-8} [20]	4.1×10^{-3} [21]
B_s^0	2.8×10^{-7} [19]	4.7×10^{-8}[20]	–

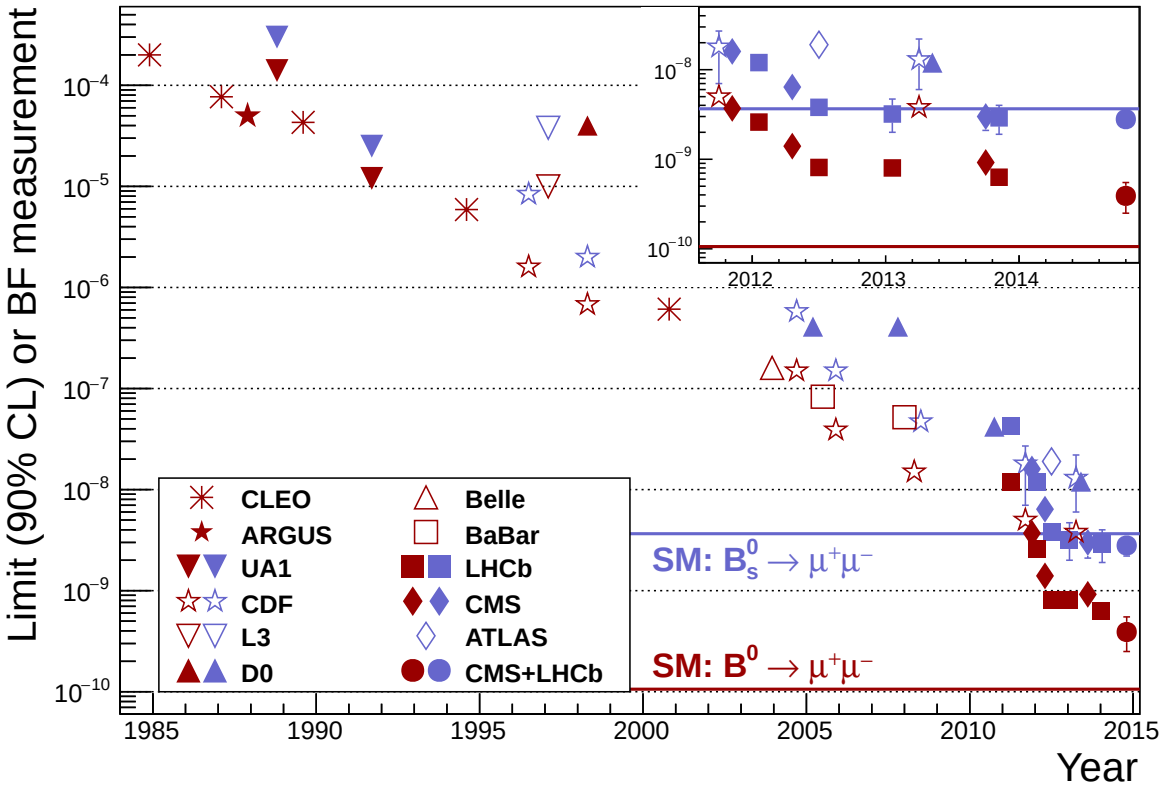

Figure 1.4: Search for the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays, reported by 11 experiments spanning more than three decades. Markers without error bars denote upper limits on the branching fractions at 90% confidence level, while measurements are denoted with errors bars delimiting 68% confidence intervals. The horizontal lines represent the SM predictions for the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ branching fractions [5]; the blue (red) lines and markers relate to the $B_s^0 \to \mu^+\mu^-$ ($B_d^0 \to \mu^+\mu^-$) decay.

Figure 1.4 summarizes the search for the $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays by the different experiments. The first evidence for the $B_s^0 \to \mu^+\mu^-$ decays was obtained by LHCb in 2012, and the first observation was published in 2015 in Nature [22] using a combined analysis of CMS and LHCb Run1 datasets. I contributed to the first five LHCb analyses published between 2011 and 2013 based successively on 37 pb^{-1}, 0.37, 1.0, 2.1 and 3.0 fb^{-1} [23, 24, 25, 26, 27]. The Chapter 2 of this document presents the analysis using the full Run1 dataset of 3.0 fb^{-1} [27]. Since then, LHCb published a new result adding 1.4 fb^{-1} of data recorded during the Run2, obtaining the first observation of the $B_s^0 \to \mu^+\mu^-$ channel by a single experiment [16]. The ATLAS collaboration published in 2019 an analysis based on 26.3 fb^{-1} of data collected in 2015 and 2016 [28]. Combined with the Run1 data, they obtain a measurement of the B_s^0 mode at 4.6σ and the most stringent

limit on the B_d^0 mode. The current world average of the $B_s^0 \to \mu^+\mu^-$ branching fraction decay, as obtained by HFLAV is $3.1 \pm 0.6 \times 10^{-9}$ [3].

The $B_s^0 \to \tau^+\tau^-$ decay was only searched for by LHCb, which published a first result in 2017, also based on the Run1 dataset. This analysis was entirely performed by the CPPM group and is detailed in Chapter 3. No results have been obtained on the electronic final states at the LHC so far. The current best measurements of rare leptonic B decays branching fractions are given in Table 1.2.1.

Table 1.2: Best measurements on rare dileptonic B decays (summer 2019). Limits are given at 90% CL, except for the $B_d^0 \to \mu^+\mu^-$ mode which is at 95% CL.

	e^+e^-	$\mu^+\mu^-$	$\tau^+\tau^-$
B_d^0	8.3×10^{-8} [19]	2.1×10^{-10} [28]	2.1×10^{-3} [29]
B_s^0	2.8×10^{-7} [19]	$3.0 \pm 0.6^{+0.3}_{-0.2} \times 10^{-9}$ [16]	5.2×10^{-3} [29]

1.2.2 B physics at the LHC

The LHC is a circular proton-proton collider working at a frequency of 40 MHz. It ran at an energy in the centre-of-mass of 7 TeV in 2011, 8 TeV in 2012, and 13 TeV in 2015-2018. Even though the first goal of the LHC is that of a discovery machine, the high $b\bar{b}$ production cross-section at its energy can be exploited to do high precision B-physics studies and searches for new physics with rare B decays. The total $b\bar{b}$ production cross-section has been measured by LHCb to be [30] $\sigma(b\bar{b}X) = (284 \pm 20 \pm 49)\mu b$[3]. While LHCb has been designed for the study of hadrons containing a b or c quark, ATLAS and CMS are general purpose detectors (GPD) for which flavour physics is a secondary objective. This result in different data taking conditions, detector geometry as well as different trigger strategy.

Regarding data taking conditions, the GPDs aim at recording as much luminosity as possible in order to maximize the chances of discovering a new particle. For example, the mean number of interactions per bunch crossing in 2012 in ATLAS and CMS was $\mu \sim$21 for a peak luminosity of 5×10^{33} cm^{-2} s^{-1}. In order to obtain flavour physics precision measurements, LHCb optimal conditions require few primary vertices (PV). This is obtained moving away the beams at the interaction point so that they do not collide head-on, as in the other experiments. A mechanism of 'luminosity levelling' allows to keep a constant peak luminosity during a fill of the LHC adjusting automatically the beam positions. The average μ value in 2012 was 1.7 for a luminosity of 4×10^{32} cm^{-2} s^{-1}. Consequently, LHCb and the GPDs record samples of very different statistics. The integrated luminosity per year for LHCb can be seen in Fig. 1.5 while the one of CMS is shown in Fig. 1.6. In particular, the Run 1 dataset corresponds to an integrated luminosity of $3\,\mathrm{fb}^{-1}$ for LHCb and $25\,\mathrm{fb}^{-1}$ for ATLAS/CMS.

The GPDs have a typical 4π detector geometry whereas LHCb exploits the fact that the b-quark pairs are mainly produced in the direction close to the beam axis. LHCb is a one-arm spectrometer with an acceptance in pseudo-rapidity between 1.8 and 4.9, which allows to have $\sim$25% of $b\bar{b}$ quark pairs in the acceptance, to be compared with 44% for the GPDs. The comparison of the acceptance can be seen in Fig. 1.7

[3] This value uses an extrapolation from the $2 < \eta < 6$ pseudorapidity range based on LEP fragmentation measurements.

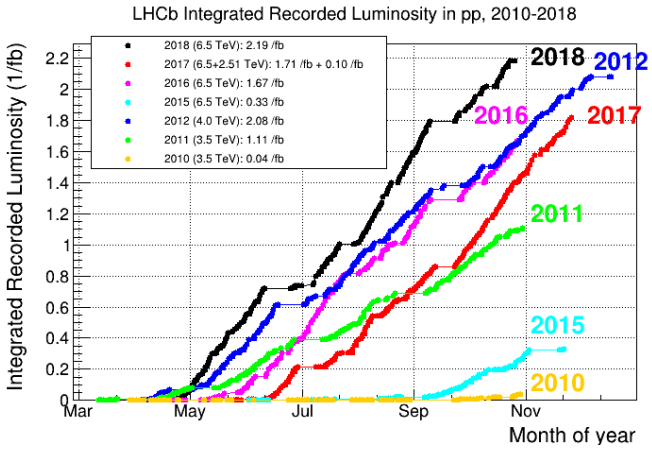

Figure 1.5: Integrated luminosity recorded per year by LHCb.

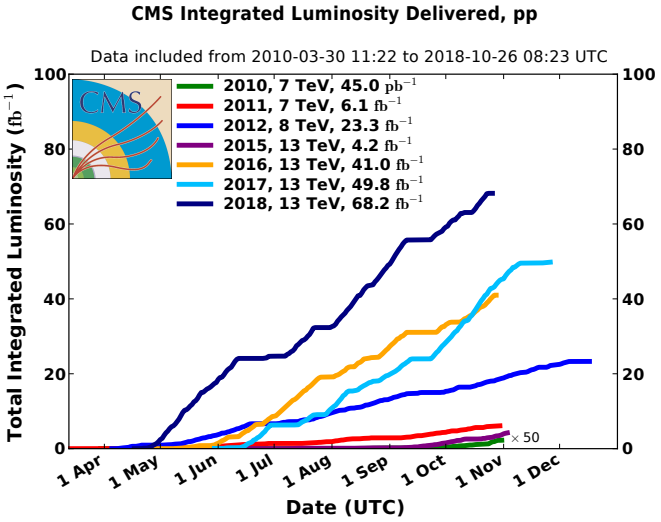

Figure 1.6: Integrated luminosity recorded per year by CMS.

Regarding the trigger strategy, the LHCb trigger focuses on the signature of b-hadron: high p_T tracks, displaced with respect to the PV. The muon trigger is the one having the highest efficiency for heavy flavour decays, but triggers based on hadron, electrons and photons are also used, which is not the case for the GPDs. The lower luminosity required by LHCb also allows to have lower trigger thresholds. For example, at the first trigger (hardware) level, during Run 1, the LHCb hardware muon trigger required a muon with a transverse momentum larger than 1.5 GeV/c while the one of CMS triggered on two muons with P_T>4(3) GeV/c for the leading (sub-leading) muons.

1.2.3 LHCb characteristics

The LHCb detector can be seen in Fig. 1.8. It consists of a tracking system with a dipole magnet that deflects charged particles in the horizontal plane, two Cherenkov detectors (RICH), a calorimetric system comprising a Scintillating Pad Detecor (SPD), a PreShower (PS), an Electromagnetic Calorimeter (ECAL) and an Hadronic Calorimeter (HCAL), and five muons stations

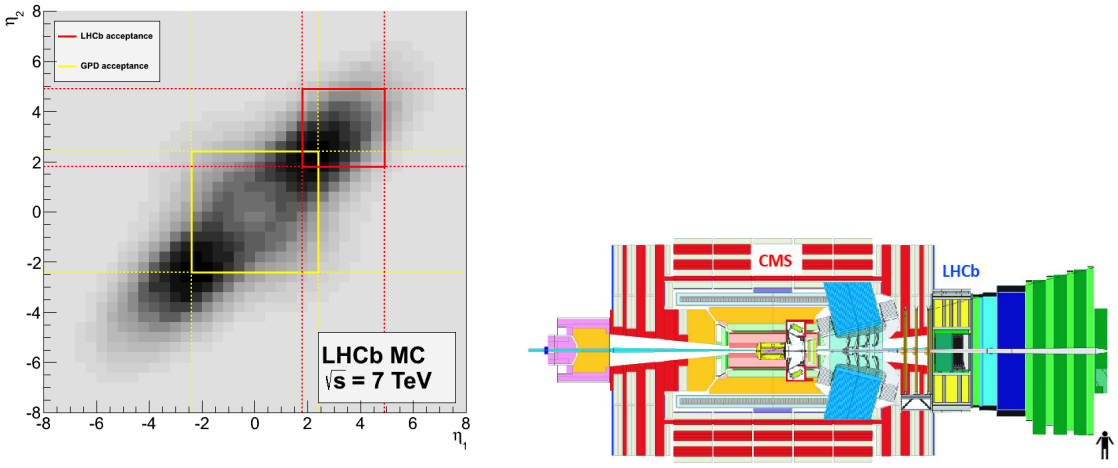

Figure 1.7: (left) 2D distribution of the b and $\bar{b}$ quark pseudorapidities obtained from simulation. The red (yellow) square represents the LHCb (GPDs) acceptance. (right) Schematic comparison of the CMS and LHCb geometry.

(M1-M5). Its angular coverage is between 15 and 300 (250) mrad in the bending (non-bending) plane.

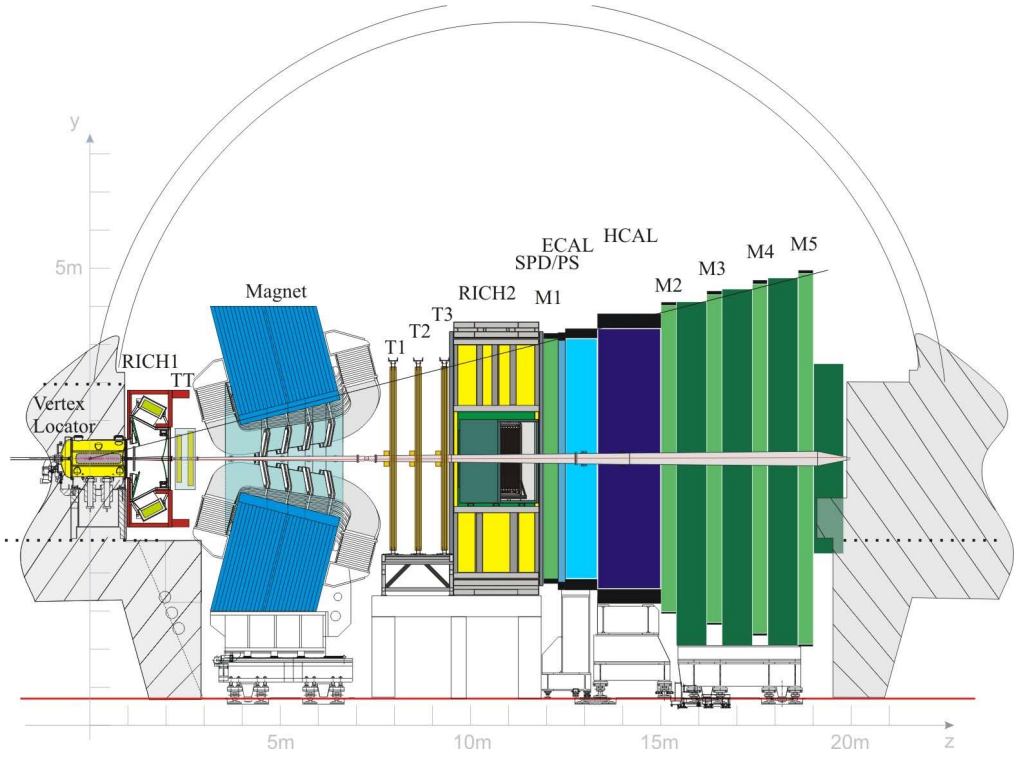

Figure 1.8: The LHCb detector.

Tracking system

The tracking system consists of the VErtex LOcator situated around the interaction point and 5 tracking stations, two before the magnet (TT) and 3 downstream of the magnet (T1-T3). The VELO contains 42 silicon semicircle modules perpendicular to the beam axis, providing measurements of the r and ϕ coordinates. The inner radius of the sensors are placed 8.2mm from the beam axis, which makes them vulnerable during the injection process. They are therefore retracted in a 'garage' position as long as the beams are not declared stable by the LHC. The TT and inner regions of the T1-T3 use silicon microstrip detector. The outer regions of T1-T3 are made of drift-tube gas detectors consisting of approximately 200 gas-tight straw-tube modules with drift-time read-out. The dipole is a 4Tm warm magnet whose polarity is regularly changed during the data taking in order to cancel systematic uncertainties for CP measurements. About half of the data are taken in each polarity.

The trajectories of the charged particles are reconstructed from their hits in the different tracking system detectors. Several types of tracks are defined, as illustrated in Fig. 1.9. The analyses reported here use only 'long tracks', which correspond to the tracks traversing the entire tracking system and having the most precise momentum measurement. The tracking algorithm contains several steps of reconstruction and extrapolation in the different sub-detectors, and a final fit based on a Kalman filter that takes into account multiple scattering and corrects for energy loss due to ionization. The χ^2 per degree of freedom of the fit is used to quantify the quality of the reconstructed track. The tracking efficiency, evaluated using a tag-and-probe technique with $J/\Psi \to \mu^+\mu^-$ from B decays, is found to be around 96% for tracks with momentum in the range $5<p<200\,\mathrm{GeV}/c$ [31]. The momentum resolution, also evaluated from $J/\Psi \to \mu^+\mu^-$ decays is measured to be about 5 per mille for long tracks with $p<20\,\mathrm{GeV}/c$, rising to 8 per mille for particles around $100\,\mathrm{GeV}/c$.

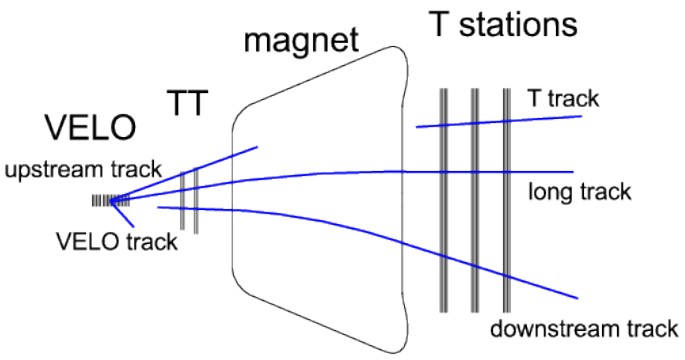

Figure 1.9: The different types of tracks reconstructed in LHCb.

A good resolution on impact parameters and vertex position is crucial for an efficient background rejection when searching for rare decays. The primary vertices (PV) are reconstructed using VELO tracks. The PV resolution is largely correlated with the number of tracks used, as seen in Fig. 1.10 (left), and is typically 13 μm in the transverse plane for a PV reconstructed with 25 tracks. The impact parameter (IP) of a track is defined as its distance from the primary vertex at its point of closest approach to the primary vertex. It allows to distinguish between tracks produced at the PV and tracks resulting from the decay of a long-lived particle such as a b- or a c-hadron. Figure 1.10 (right) shows the IP resolution in the transverse plane as a function of $1/p_T$. It is typically of a few tens of μm.

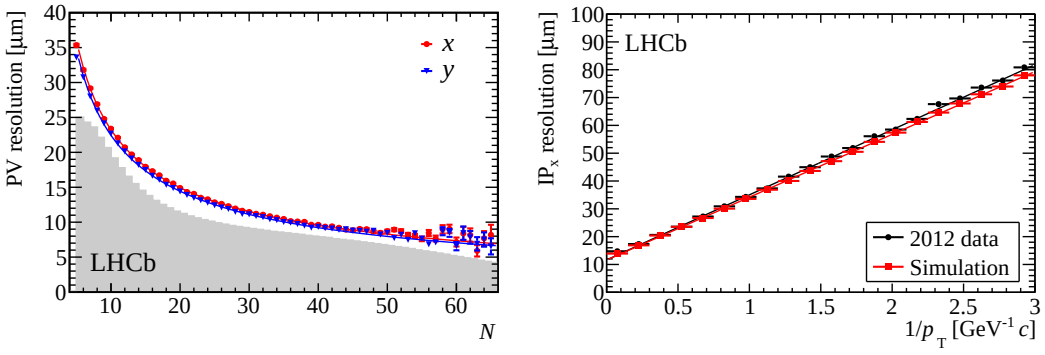

Figure 1.10: (left) The primary vertex resolution, for events with one reconstructed primary vertex, as a function of track multiplicity. The x (red) and y (blue) resolutions are separately shown and the superimposed histogram shows the distribution of number of tracks per reconstructed primary vertex for all events passing the high level trigger. (right) The impact parameter in x resolution as a function of $1/p_T$.

Particle identification

Particle identification is achieved using information from several subdetectors: the two Cherenkov detectors, the calorimeters and the muon stations. The distinction between charged and neutral particles is performed by studying the presence or absence of tracks in front of the energy deposit in the calorimeters. For energy deposit corresponding to neutral particles, the shape of the cluster is used to differentiate photon from π^0 candidates. For energy deposit corresponding to charged particles, an electron hypothesis is constructed to distinguish electrons from hadrons using information from the ECAL, PS and HCAL. The charged hadron identification is performed thanks to two RIng Imaging Cherenkov detectors (RICH1 and RICH2). The RICH system can contribute to the identification of charge leptons (e,μ) complementing information from the calorimeter and muon systems. RICH1 covers the low momentum range, from 2 to 60 GeV/c and uses Aerogel and C_4F_{10} as radiators. RICH2 covers the high momentum range, from 15 to 100 GeV/c using a CF_4 radiator. In both RICH, the photons emitted are collected by Hybrid Photon Detectors (HPDs), which are enclosed in iron shields. The identification of a track reconstructed in the tracking system as a muon (referred as "IsMuon" selection) is based on the association of hits around its extrapolated trajectory in the muon system. A search is performed for hits within rectangular windows around the extrapolation points where the x and y dimensions of the windows are parameterised as a function of momentum at each station and separately for each muon system region. The five muon chambers are multi-wire proportional chambers except the inner part of the M1 which is made of triple Gas Electron Multiplier detectors to cope with the high particle rate. The efficiency of the muon identification, evaluated from $J/\Psi \rightarrow \mu^+\mu^-$, can be seen in Fig. 1.11(left) for different p_T ranges. The probablity to misidentify a pion as a muon is evaluated from $D^0 \rightarrow K^-\pi^+$ decays from $D^{*+} \rightarrow D^0\pi^+$ and is shown in Fig. 1.11(right).

The informations from the RICH, calorimeters and muon systems are combined to build more powerful PID variables. In a first method, the likelihoods from each subdetector are linearly combined to provide a global likelihood named $DLL(X-\pi)$, which measures how likely is the X mass hypothesis compared to the pion hypothesis. A second method based on a multivariate approach was introduced during the Run1 data taking combining all the PID information into a single variable, named $ProbNNX$, corresponding to the probability value for each mass hypothesis.

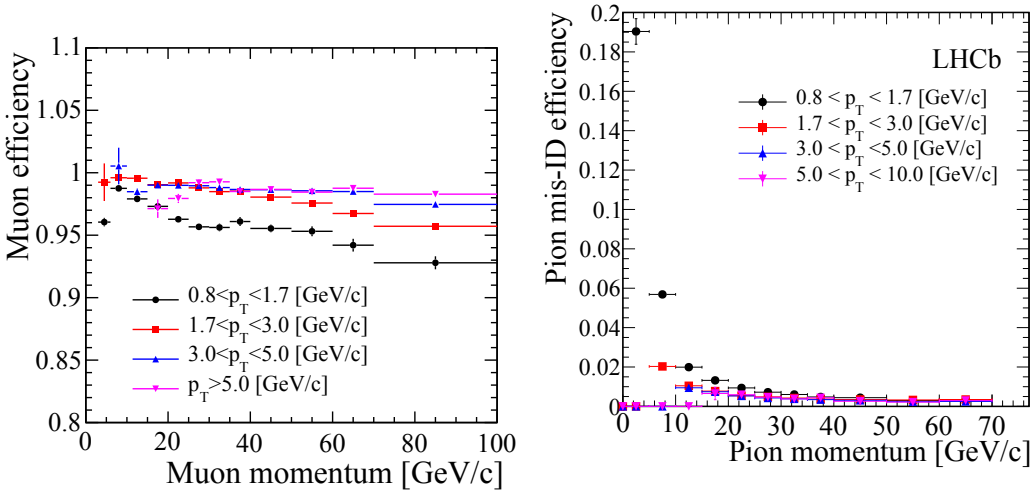

Figure 1.11: (left) Muon identification efficiency as function of momentum based on the matching of hits in the muon system to track extrapolation. (right) Misidentification probability of pion as muon candidates as a function of momentum, for different p_T ranges. Performance plots are from Ref. [32].

Trigger system

The LHCb trigger system consists of a hardware stage, the Level-0 (L0), which reduces the collision rate from 40 MHz to 1 MHz, followed by a software stage, the High Level Trigger (HLT), which brings it down to 5 kHz. The L0 exploits informations from the calorimeter and muon systems while at the HLT stage the whole detector is read-out. The events selected by the L0 and HLT are then stored for offline reconstruction.

The L0-calorimeter computes the transverse energy deposit in cluster of 2×2 cells and can build three types of candidates: *L0Hadron*, which corresponds to the highest E_T HCAL cluster, also containing the energy of the matching ECAL cluster, *L0Photon*, which is the highest E_T ECAL cluster with 1 or 2 PS hits in front of the ECAL cluster and no hit in the SPD cells corresponding to the PS, and *L0Electron* that is similar to *L0Photon* with the additional condition of at least one SPD cell hit in front of the PS cells. The number of SPD hits is also used to veto events with a too large multiplicity.

The L0-muon system comprises four processors each corresponding to one quadrant of the muon detector. They select the two muon candidates of highest p_T by searching for hits that define a straight line through the five muon stations and that point towards the interaction point. The position of a track in the first two stations allows the determination of its p_T with a measured resolution of 25% relative to offline reconstructed muon tracks. The trigger sets a single threshold on either the largest p_T of the eight candidates (*L0Muon*), or a threshold on the product of the two largest ones (*L0DiMuon*). The typical L0 thresholds used in Run1 are given in Table 1.2.3.

The HLT is divided into two steps : the HLT1 that performs a partial reconstruction of the event and reduces the rate to about 80 kHz, and the HLT2, where the complete event is reconstructed. The final output rate is 5 kHz. Most of physics analyses use an HLT1 algorithm based on the identification of one good quality track, with a p_T larger than 1.6-1.7 GeV/c and a displacement with respect to the PV larger than 0.1 mm. Additionaly, lines with tracks that match hits in the muon chambers are also implemented. A single muon trigger line selects good quality muon candidates displaced from the primary vertex and satisfy $p_T > 1\,\mathrm{GeV}/c$. Dimuon candidates

Table 1.3: Typical L0 thresholds used in Run1.

	p_T or E_T		SPDHits
	2011	2012	2011 and 2012
L0Muon	$1.48\,\mathrm{GeV}/c$	$1.76\,\mathrm{GeV}/c$	600
L0DiMuon	$(1.30\,\mathrm{GeV}/c)^2$	$(1.60\,\mathrm{GeV}/c)^2$	900
L0Hadron	$3.50\,\mathrm{GeV}/c$	$3.70\,\mathrm{GeV}/c$	600
L0Photon	$2.50\,\mathrm{GeV}/c$	$3.00\,\mathrm{GeV}/c$	600
L0Electron	$2.50\,\mathrm{GeV}/c$	$3.00\,\mathrm{GeV}/c$	600

are either selected based on their invariant mass ($m_{\mu\mu}$>2.5 GeV/c^2) without any displacement requirement, or based on their displacement without the mass restriction.

The HLT2 level comprises topological lines that can trigger on any b-hadron decay with at least two final state particles, as well as a large number of more specific lines. In the topological lines [33], tracks are selected based on their track fit $\chi^2/ndof$, IP and muon or electron identification. Two-, three- or four-body vertices are constructed from the selected tracks with a requirement on their distance of closest approach (DOCA). Candidates $n-$body combinations are then selected using a boosted decision tree that combines several kinematic and geometrical variables.

Conclusion

This chapter presented the state of the art of the dileptonic B meson decays. At the start of the LHC, the limits on the electronic final states were ∼7 orders of magnitude higher than their SM predictions. For the muonic final state, it was about 1-2 orders of magnitude. The best (and only) limit on the $B^0_d \to \tau^+\tau^-$ was ∼19000 times higher than the SM value and there were no experimental results on the $B^0_s \to \tau^+\tau^-$. From the characteristics presented previously, it is clear that LHCb is the perfect experiment to search for the $B^0_{d,s} \to \mu^+\mu^-$ decays thanks to the very high $b\bar{b}$ production cross section at the LHC, the low muon trigger thresholds and the excellent vertex and IP resolution. We can not claimed the same thing for the $B^0_{d,s} \to e^+e^-$ and $B^0_{d,s} \to \tau^+\tau^-$ channels. Indeed the LHCb electron reconstruction efficiency is quite poor, and, in the absence of a 4π detector coverage and a noisy hadronic environment, the study of decays with missing energy is particularly difficult. Nevertheless LHCb being the only running experiment where such a high rate of B^0_s meson could be produced, it would have been a pity not to tackle this challenge. Chapters 2 and 3 of this document present the search for the $B^0_d \to \mu^+\mu^-$ and $B^0_{d,s} \to \tau^+\tau^-$ channels at LHCb. For completeness, we inform the reader that the electronic final state is also currently under study at LHCb.

Chapter 2

Search for the $B^0_{d,s} \to \mu^+\mu^-$ decays

As discussed in Sec. 1.2.1, LHCb performed five successive searches for the $B^0_{d,s} \to \mu^+\mu^-$ decays based on Run1 data. Improvements have been made in the different versions, but the global analysis strategy did not change. This chapter presents the analysis based on the full Run1 statistics of $3\,\mathrm{fb}^{-1}$. Section 2.1 presents the analysis strategy, Sec. 2.2 and 2.3 detail the selection and signal yield extraction. The normalisation is presented in Sec. 2.4. The results are given in Sec. 2.5 and their combination with the CMS analysis is shown in Sec. 2.6. Future prospects on the $B^0_{d,s} \to \mu^+\mu^-$ branching fraction measurement are given in Sec. 2.7.

2.1 Analysis overview

2.1.1 Principle of a branching fraction measurement

By definition, the branching fraction of a $B_q \to f$ decay is

$$\mathcal{B}(B_q \to f) = \frac{N_{B_q \to f}}{N_{B_q^{tot}}}, \tag{2.1}$$

where $N(B_q \to f)$ is the number of B_q meson decaying into the f final state and $N(B_q^{tot})$ is the total number of B_q meson produced. This number can be expressed as function of the integrated luminosity $\mathcal{L}$, the $b\bar{b}$ cross section $\sigma_{b\bar{b}}$ and the hadronization fraction of a b quark into a B_q meson, f_q, as $N(B_q^{tot}) = \mathcal{L} \times \sigma_{b\bar{b}} \times f_q$. Given the uncertainties attached to the luminosity and cross section at a hadron collider, a more precise estimate of $N(B_q^{tot})$ is usually obtained through the use of a normalisation channel with a precisely known branching fraction, $\mathcal{B}_{norm}$, and a final state close to the signal, which allows the cancellation of sytematic uncertainties in the ratio.

The true number of signal $N(B_q \to f)$ and normalisation N_{norm} events are expressed as function of their corresponding observed events N^{obs} and efficiencies ϵ, so that the branching fraction takes the form :

$$\mathcal{B}_{B_q \to f} = \mathcal{B}_{norm} \times \frac{\epsilon_{\rm norm}}{\epsilon_{\rm sig}} \times \frac{f_{\rm norm}}{f_q} \times \frac{N^{obs}_{B_q \to f}}{N^{obs}_{\rm norm}} = \alpha_q \times N^{obs}_{B_q \to f}\,, \tag{2.2}$$

where f_q and $f_{\rm norm}$ are the probabilities that a b-quark hadronizes into the B_q meson and the b-hadron relevant for the chosen normalisation mode. $\mathcal{B}_{norm}$ and $N_{\rm norm}$ are the branching fration and the number of selected events of the normalisation channel, respectively. Finally, α_q is the normalisation factor (or single event sensitivity). The measurement of a branching fration then necessitates the determination of the observed number of events of signal and normalisation channel in data and their efficiencies. The other factors, $\mathcal{B}_{norm}$ and $\frac{f_{\rm norm}}{f_q}$, are usually external inputs measured in different analyses or experiments.

The analysis stategy should then be adapted to the final state considered in order to maximize the sensitivity. In case of searches for very rare decays such as $B^0_{d,s} \to \ell^+\ell^-$, the main challenge is the background rejection.

2.1.2 Main background sources polluting $B^0_{d,s} \to \mu^+\mu^-$ decays

The probability that a b-hadron decays semileptonically into a muon is ∼11% [34]. Before any selection, the ratio of expected $B^0_s \to \mu^+\mu^-$ events over the number of events where the two $b-$hadrons decay to a muon (labelled $bb \to \mu\mu X$ in the following) is therefore $\sim 4 \times 10^{-8}$. The topology of these two types of events is represented in Fig. 2.1, where one can see several caracteristics that can be exploited to distinguish between them:

- The fact that the two muon tracks form a good vertex. This is translated in the distance of closest approach (DOCA) of the two tracks and in the vertex quality ($\chi^2/ndof$).
- The presence of additional tracks coming from the $b-$hadrons for the background events, seen as light grey tracks. This will be exploited through isolation variables.
- The fact that the reconstructed B meson in background events will not necessary points toward the PV. This is translated in the impact parameter (IP), impact parameter χ^2 (IPχ^2) and the angle between the direction of the reconstructed B momentum and the direction defined by the primary and secondary vertices (DIRA).

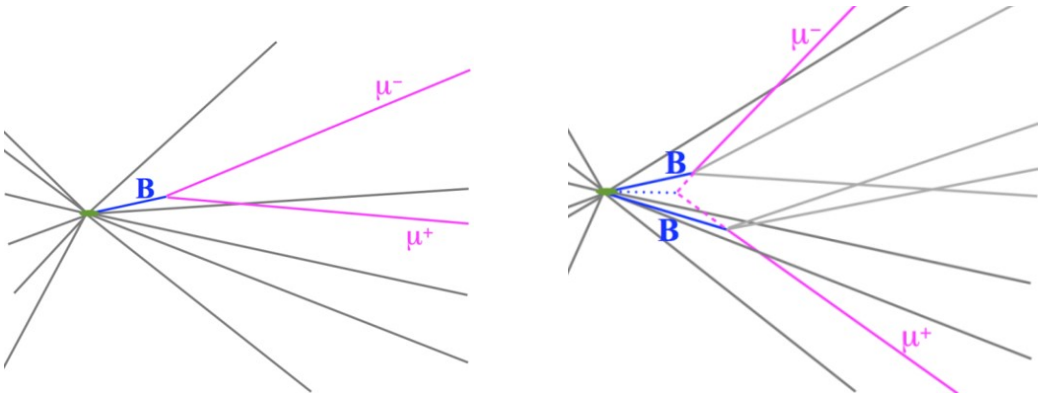

Figure 2.1: Schematic representations of (left) a signal event and (right) a $bb \to \mu\mu X$ background event.

In addition to this combinatorial background, *physics backgrounds*, less aboundant but eventually more signal-like, can also pollute the signal. These are:

- Exclusive dimuon events $pp \to p\mu\mu p$ coming from elastic processes that can mimic the signal if the two protons fall in the beam pipe and the muons point toward a PV.
- Cascade semileptonic decays of the type $B \to D(\to \mu Y)\mu X$.
- $B \to \mu^+\mu^-\gamma$ decays. These are not helicity suppressed and can easily mimic the signal in case of soft photon.
- $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, $h^{(\prime)}$ being a kaon or a pion, where the two hadrons are misidentified as muons because they decay in flight.
- Other hadronic or semileptonic decays with at least one final state particle misidentified as a muon ($B^0 \to \pi^-\mu^+\nu$, $\Lambda_b \to p\mu^+\nu$,...).

The main goal of the analysis is to find the best discrimination between signal events and the different background sources. This is achieved through a selection with high signal efficiency followed by the use of a boosted decision tree. Possible remaining physics backgrounds are then evaluated carefully and taken into account.

2.1.3 Analysis workflow

The main analysis steps, that will be detailed in the rest of this chapter, are the following:

- A selection designed to remove most of the obvious backgrounds while keeping a high signal efficiency.
- The classification of the candidates according to two variables: a boosted decision tree (BDT) and the dimuon invariant mass ($m_{\mu\mu}$).
- The computation of the normalisation factor using two normalisation channels, $B^+ \to J/\Psi(\to \mu^+\mu^-)K^+$, which contain two displaced muons in the final state as the signal but an additional kaon, and $B^0 \to K^-\pi^+$, which has the same topology as the signal but different final state particles.

- The determination of the signal yield using a fit to the invariant mass sidebands in bins of the BDT output. In case no significant amount of signal is obtained, a limit is set using the $\mathrm{CL_s}$ method [35].

The invariant mass fit is performed in the mass range $[4900, 6000]\,\mathrm{MeV}/c^2$ but in order to avoid unconscious bias, the signal region $m_{\mu\mu} = [m_{B^0_d} - 60\,\mathrm{MeV}/c^2, m_{B_s} + 60\,\mathrm{MeV}/c^2]$ is excluded ('blinded') until the completion of the analysis.

The selection and BDT optimization as well as the efficiency calculation are performed on simulated samples. Samples of 1M events are used for $B^0_s \to \mu^+\mu^-$ and $B^0_d \to \mu^+\mu^-$ signals, and a 8.6M events sample (equivalent to $7\,\mathrm{fb}^{-1}$) of high mass inclusive $bb \to \mu\mu X$ events is used as a proxy for background[1]. Additional samples for normalisation channels and exclusive backgrounds are also used. To save CPU time, requirements are applied at generator level so that all the final state particles of interest are in the LHCb acceptance. In addition, control samples consisting of $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, where $h^{(\prime)}$ is a pion or a kaon, and $B^+ \to J/\Psi(\to \mu^+\mu^-)K^+$ decays are largely used to calibrate signal distributions or correct the simulation.

2.2 Selection

The selection consists of a first cut-based step followed by a requirement on a boosted decision tree output. The selection should be designed to be as similar as possible for the signal and the normalisation channels to cancel as much as possible the systematic uncertainties appearing in Eq. 2.2.

2.2.1 Cut-based selection

The first step of the selection is done at the *stripping* level, an ensemble of algorithms that are run centrally and allow the reduction of the amount of data to be analysed. The requirements applied in the selection on the reconstructed signal and normalisation channels are listed in Table 2.1. To pass the stripping selection, candidates must be formed by two oppositely-charged tracks with a good χ^2/ndf, a small probability to be ghost tracks (i.e. tracks made of hits not deposed by a single particle), high momenta, having a DOCA less than 0.3 mm and a good vertex fit χ^2. To remove background candidates made of prompt tracks coming from the PV, the distance significance between the reconstructed B candidate and the PV (VDS) must be large enough and the B must point toward the PV.

Additional requirements are put on kinematic variables such as the momentum and transverse momentum. This allows to reduce exclusive dimuon $pp \to p\mu\mu p$ events as the dimuon system has a relatively low p_T [36]. Dimuon from J/Ψ in the $B^+ \to J/\Psi K^+$ channels are required to be within 60 MeV/c^2 of the nominal J/Ψ mass.

The different channels are then identified adding a PID selection: ISMUON selection [37] for the muons from $B^0_s \to \mu^+\mu^-$ and $B^+ \to J/\Psi K^+$ channel, and requirement on $DLL(K-\pi)$ and $DLL(\mu-\pi)$ for the hadrons from $B^0_{(s)} \to h^+h^{(\prime)-}$ channel. An additional requirement, $DLL(K-\pi) < 10$ and $DLL(\mu-\pi) > -5$, is added to the $B^0_s \to \mu^+\mu^-$ selection in order to reduce by a factor 5 the $B^0_{(s)} \to h^+h^{(\prime)-}$ events where the two hadrons are misidentified, at a price of a 3% signal loss.

[1] The specific cuts applied at generator level to this sample are $p_\mu > 3\,\mathrm{GeV}/c$, $4.7 < m_{\mu\mu} < 6.0\,\mathrm{GeV}/c^2$ and the distance of closest approach of the two muons less than 0.4mm.

Table 2.1: Selection for $B^0_{d,s} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h^{(\prime)-}$ and $B^+ \to J/\psi K^+$ channels.

Cut	applied on	value	applied on	value
		$B^0_{d,s} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h^-$		$B^+ \to J/\psi K^+$
track χ^2/ndf	μ / h	<3	μ / h	< 3
ghost prob		< 0.3		
DOCA		<0.3 mm		<0.3 mm
IPχ^2		>25		>25
p_T		> 0.25 and < 40 GeV/c		> 0.25 and < 40 GeV/c
p		<500 GeV/c		<500 GeV/c
ISMUON	μ only	true	μ only	true
vertex χ^2	$B_{(s)}$	<9	J/ψ	<9
VDS		> 15		> 15
ΔM		$\lvert M(hh,\mu\mu) - m_B\rvert <$ 60 MeV/c^2		$\lvert M(\mu\mu) - m_{J/\psi}\rvert <$ 60 MeV/c^2
IPχ^2	$B_{(s)}$	< 25	B^+	< 25
t		$< 9 \cdot \tau(B^0_s)$		$< 9 \cdot \tau(B^0_s)$
BDTS		> 0.05		> 0.05
DLL($K-\pi$)		< 10		
DLL($\mu-\pi$)		> -5		
ΔM				$\lvert M(J/\psi K) - m_B\rvert <$ 100 MeV/c^2
p_T (B^0_s)	$B^0_{d,s} \to \mu^+\mu^-$ $B^0_{(s)} \to h^+h^-$	> 0.5 GeV/c		

The analysis is also protected against non-physical candidates by removing all candidates which have either a momentum, a transverse momentum, or a lifetime too large. This fiducial selection is designed to be 99.9% efficient for the signal.

2.2.2 Trigger

Preliminary remark: even if all events containing a reconstructed candidates passed the trigger, a tighter selection is usually applied after the stripping requiring specific lines to be fired. The reason for that is to have a better control of the trigger efficiencies, which are usually evaluated from simulation. The requirement can be put on the event (e.g. the event passed the *L0Muon* trigger line), or to the candidate itself. It can be asked that the signal candidate is responsible for trigger (e.g. one of the two muons is the one used in the *L0muon* decision), which is called trigger-on-signal (TOS). It can be also required that the signal candidate has nothing to do with the fact that the event has been selected by the trigger, in that case it is trigger-independently-of-signal (TIS).

In order to maximize the signal efficiency, no choice of specific lines passed by signal candidates is made for this analysis. Most of selected events are triggered by a small number of line, listed in the following. The dominant L0 lines are the *L0Muon* and *L0Dimuon*, introduced in Sec. 1.2.3. At the HLT1 level, the most important lines are the single muon *HLT1TrackMuon* and the two dimuon lines *Hlt1DiMuonLowMass* and *Hlt1DiMuonHighMass*, where the latter selects dimuon combinations with an invariant mass at or above the J/Ψ mass without any requirement on the flight distance, whereas the former select lower masses requiring that the muon tracks have a significant IPχ^2 with respect to their best PV. For the HLT2, the signal is triggered by the

Hlt2DiMuonB line, which requires the dimuon invariant mass to be above $4.7\,\mathrm{GeV}/c^2$, and in smaller proportion by the topological lines. Additional informations about the HLT muon lines can be found in [38]. The control channels containing a J/Ψ are triggered by the *Hlt2DiMuonDetached* line.

For the $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, the situation is different as they are mainly triggered by hadronic lines, which have higher thresholds than the muon lines, as seen in Sec. 1.2.3. In order to keep the selection as similar as possible for the normalisation and signal channels, the $B_{(s)} \to h^+h^{\prime-}$ candidates are required to be TIS, meaning that the trigger was passed regardless the presence of the two final state hadrons. The TIS requirement can be applied at all trigger levels but has a quite low efficiency. To keep a sufficient statistics it is only required for the L0 and HLT1. The main HLT2 lines selecting $B^0_{(s)} \to h^+h^{(\prime)-}$ candidates are the 2-body topological line and the exclusive *Hlt2B2hh* line.

2.2.3 BDT based selection

After this set of requirements, the signal is still polluted by a huge amount of backgrounds and a more refined selection, based on a BDT, is applied. This BDT, named 'BDTS', is trained using the TMVA package [39] on signal MC and background candidates from the dihadron mass spectra sidebands $[4800-5000]$ and $[5500-6000]\,\mathrm{MeV}/c^2$. It uses the following input variables:

- the IP and IPχ^2 of the B candidate,
- the vertex χ^2 of the B meson or of the J/ψ for the $B^+ \to J/\psi K^+$ channel,
- the angle between the direction of the momentum of the B candidate and the direction given by the B candidate production and decay vertices (DIRA),
- the DOCA of the two muon tracks,
- the smallest impact parameter of the two muon tracks with respect to any reconstructed primary vertex in the event.

The selection criteria based on the output of this BDT, allows a background rejection of 70% in the whole dimuon invariant mass fit region while keeping an efficiency on the signal of 95%. The efficiencies are found to be consistent between signal, normalisation, and control channels within 0.4%.

2.3 Signal yield extraction

After the selection, the signal is classified in a 2D plane defined by a multivariate discriminant, based again on a boosted decision tree, and the reconstructed B invariant mass. The number of $B^0_s \to \mu^+\mu^-$ and $B^0_d \to \mu^+\mu^-$ events are obtained from a simultaneous likelihood fit to the invariant mass in eight bins of BDT. For each BDT bin, the fit function can be written as

$$F_{tot}(m_{\mu\mu}) = N_{bkg}F_{bkg} + N_{B^0_d}F_{B^0_d}(m_{\mu\mu}) + N_{B^0_s}F_{B^0_s}(m_{\mu\mu}), \tag{2.3}$$

where N_i are the fitted number of events and $F_i(m_{\mu\mu})$ are the invariant mass PDFs. This section describes the procedure followed to define the BDT, obtain the fraction of signal events in each BDT bins and the invariant mass PDFs calibration.

2.3.1 BDT definition

The BDT is trained on simulated signal and $bb \to \mu\mu X$ background, which allows to keep an independent sample for the background determination in data. The choice of the input variables and the BDT parameters (maximum number of trees, minimum number of events in a leaf, maximum allowed depth of the tree, number of steps in the optimization of each cut and parameter related to the boosting algorithm) are optimized using a figure-of-merit based on a test statistics. The BDT is found to be optimal when including the following 12 topological and kinematical variables :

- the B meson proper time, transverse momentum and impact parameter,
- the minimum impact parameter significance of the muons,
- distance of closest approach between the two muons,
- the isolation of the two muons with respect to any other track in the event. This variable corresponds to the number of long tracks that can make a good vertex with one of the two muons.
- the cosine of the angle between the muon momentum in the dimuon rest frame and the vector perpendicular to the B momentum and the beam axis,
- the B isolation based on the CDF definition [40],

$$I(B) = \frac{p_T(B)}{p_T(B) + \sum_{tracks} p_T(tracks)}, \tag{2.4}$$

- the angle between the B candidate's momentum and the thrust momentum of the B, defined as the sum of momenta of all the long tracks coming from the B PV and excluding those coming from long lived particles. If no such tracks are available, the variable is set to 0 (other B angle),
- the angle between the direction of the positive muon candidate in the rest frame of the B and the thrust momentum in the B rest frame,
- the absolute values of the differences between the pseudorapidity and spherical ϕ coordinate of the two muon candidates.

As a final step, the BDT is transformed to be uniformly distributed for the signal between 0 and 1. The background is then peaking toward 0.

The agreement between data and simulation for the input variables has been checked on the $B^0_{d,s} \to \mu^+\mu^-$ sidebands for the background sample and the sideband-subtracted $B^0_{(s)} \to h^+h^{(\prime)-}$ for the signal. Only the isolation variables are found to differ slightly due to different data taking conditions in the data and simulation. This can result is suboptimal BDT performances but does not bias the analysis since, as it will be explained in Sec. 2.3.2, the BDT is calibrated on data. Given that the number of events in the signal mass region is extrapolated from the sidebands, the correlations between the BDT and the mass should be carefully evaluated and taken into account in the optimization process. If a linear correlation is not problematic, a peaking distribution would clearly bias the background determination.

A correlation is observed for signal in the low mass region due to the radiative tail. The correlation for $bb \to \mu\mu X$ and signal events in the signal region is negligible.

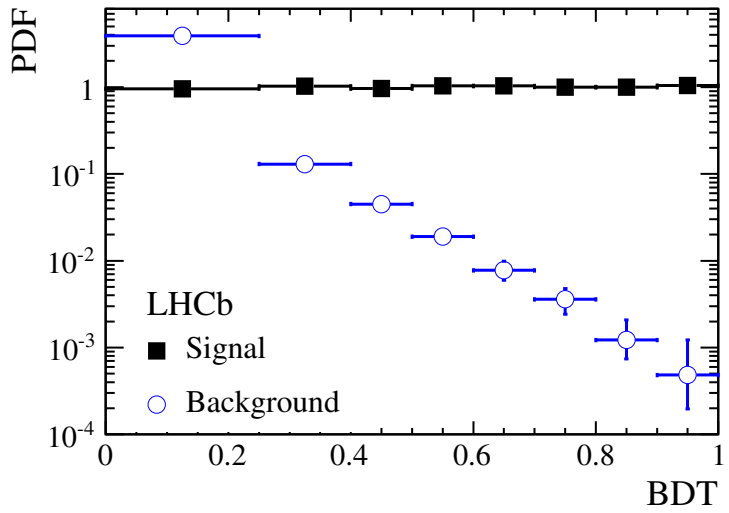

Figure 2.2: Expected distribution of the BDT output for the $B_s^0 \to \mu^+\mu^-$ signal (black square) and the combinatorial background (blue circles).

The BDT discriminating power can be seen in Fig. 2.2, which shows the expected distribution for signal and background events. The method to obtain these distributions is described in the following section. The BDT binning has been optimized in order to reach the highest sensitivity [41]. The bin boundaries are [0,0.25,0.4,0.5,0.6,0.7,0.8,0,9,1.0].

The distribution of events in the 2D plane consisting of the BDT output and the invariant mass is shown in Fig. 2.3.

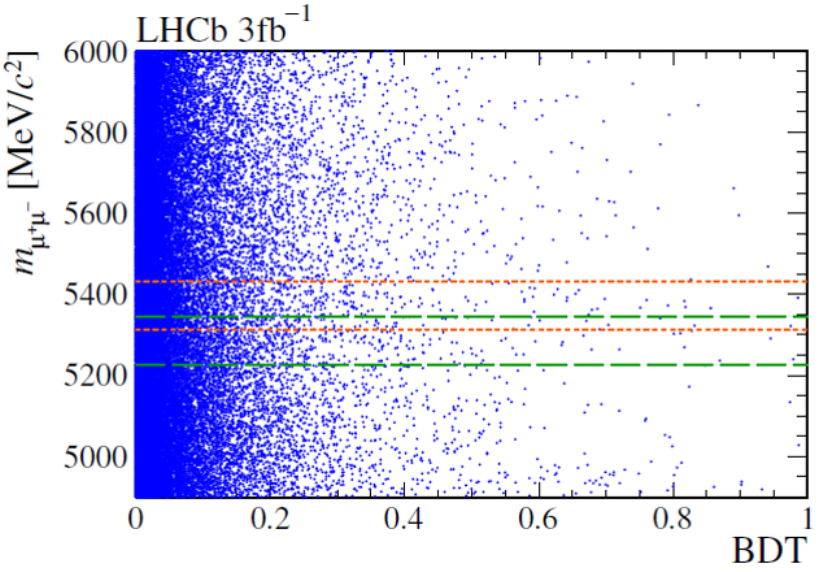

Figure 2.3: Data distribution in the 2-dimensional plane of the BDT output versus the dimuon invariant mass. The signal regions (unblinded here) are shown by orange (green) lines for the B_s^0 (B_d^0) decays.

Table 2.2: Main decays contributing to the $B^0_{(s)} \to h^+h^{(\prime)-}$ sample and their branching fractions.

Decay	$\mathcal{B}$ (10^{-6})
$B^0_d \to K^+\pi^-$	19.6 ± 0.5
$B^0_s \to \pi^+K^-$	5.7 ± 0.6
$B^0_s \to K^+K^-$	25.9 ± 1.7
$B^0_d \to \pi^+\pi^-$	5.12 ± 0.19

2.3.2 BDT calibration

Since the BDT is based on geometrical or B kinematic variables, the signal shape can be obtained from the $B^0_{(s)} \to h^+h^{(\prime)-}$ events, which have a similar topology. To further reduce the differences between $B^0_{(s)} \to h^+h^{(\prime)-}$ and $B^0_{(s)} \to \mu^+\mu^-$ events due to the reconstruction and selection, only the $B^0_{(s)} \to h^+h^{(\prime)-}$ candidates inside the muon detector acceptance are used. In addition, the candidates are required to be TIS at the L0 and HLT1 level to avoid trigger biases. Remaining discrepancies due to the trigger and PID selection are corrected for. Another bias due to the different lifetime acceptance in BDT bins is also taken into account.

The $B^0_{(s)} \to h^+h^{(\prime)-}$ cocktail is actually composed of four main decays listed in Table 2.2 and polluted by $\Lambda^0_b \to pK^-$ and $\Lambda^0_b \to p\pi^-$ decays where the proton is identified as a kaon or a pion. Figure 2.4 illustrates the contributions from these different channels.

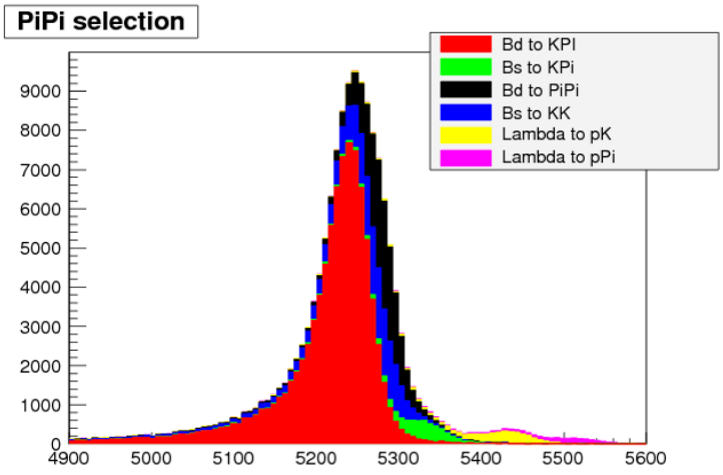

Figure 2.4: Invariant mass under the $\pi\pi$ hypothesis for the different channels contributing to the $B^0_{(s)} \to h^+h^{(\prime)-}$ sample as obtained from simulation.

The BDT distribution is obtained for the $K\pi$ final state using requirements on the $DLL(K-\pi)$ variable. An example of a fit for the highest BDT bin is shown in Fig. 2.5. The B^0_s and B^0_d signals as well as the $\Lambda_b \to ph$ decays are each described by the sum of two Crystal Ball functions, the combinatorial background is described by an exponential function and the partially reconstructed backgrounds by a function defined for an independent analysis of $B^0_{(s)} \to h^+h^{(\prime)-}$ decays [42].

Figure 2.6 shows the signal BDT PDF with the statistical uncertainty as red band and the systematic uncertainty as grey band for the different BDT bins. The systematic uncertainty takes

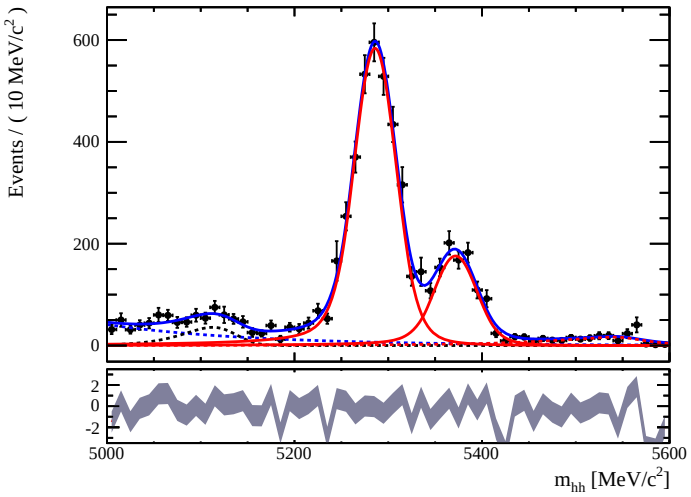

Figure 2.5: Example of a fit to the $m_{K\pi}$ invariant mass distribution when requiring $DLL(K-\pi) > 5$ for the kaon and $-DLL(K-\pi) > 5$ for the pion, in the BDT bin [0.9,1.0]. The red solid line shows the B^0_d and B^0_s signals, the red dashed one the one from $\Lambda^0_b \to ph$ decays. The combinatorial background is shown by the blue dashed line, while the partially reconstructed background is shown by the black solid line.

into account uncertainties related to the $DLL(K-\pi)$ requirements and due to the fit model.

Three corrections need to be made to this PDF. The first one concerns the PID requirement added to the $B^0_{d,s} \to \mu^+\mu^-$ signal selection, $DLL(K-\pi)$<10 and $DLL(\mu-\pi)$>-5. The corresponding efficiency is obtained from inclusive J/Ψ events.The second correction concerns the trigger efficiency. The BDT response calibrated with $B^0_{(s)} \to h^+h^-$ (L0×HLT1) TIS events has to be divided by the following correction factor:

$$\frac{\epsilon_{\rm TIS}(L0\times HLT1)_{hh} \times \epsilon_{\rm trigger}(HLT2)}{\epsilon_{\rm trigger}(L0\times HLT1\times HLT2)_{\mu\mu}} \tag{2.5}$$

where the numerator is the efficiency correction to be applied to $B^0_{(s)} \to h^+h^-$ TIS events to get a trigger unbiased sample, while the denominator account for the trigger efficiency of $B^0_{(s)} \to \mu^+\mu^-$ signal events. The evaluation of the above factors has been performed using $B^0_d \to K^+\pi^-$ and $B^0_{(s)} \to \mu^+\mu^-$ Monte Carlo events. The last correction is related to the lifetime of the B mesons, which is an input variable to the BDT. The control channels[2] and the B^0_s signal actually have different lifetime distributions. Each bin i of the BDT PDF should be corrected by

$$\delta^i_{PDF} = \frac{\epsilon^i_{\mu\mu}(\mathcal{A}_{\Delta\Gamma}, y_s)}{\epsilon^i_{\rm hh}} \times \frac{\epsilon_{\rm hh}}{\epsilon_{\mu\mu}(\mathcal{A}_{\Delta\Gamma}, y_s)}, \tag{2.6}$$

where ϵ_i is the efficiency in the bin i and ϵ, the total efficiency. This correction depends on $\mathcal{A}_{\Delta\Gamma}$and y_s and is thus model dependent, as shown in Fig. 2.7. Assuming the SM values ($\mathcal{A}_{\Delta\Gamma} = 1$ and $y_s = 0.0615 \pm 0.085$), the correction ranges from -3.1% to 4.7% depending on the BDT bin. For $B^0_d \to \mu^+\mu^-$ no correction has to be applied since the control channel is also a B^0_d decay and has therefore the same decay rate as the signal one.

[2]Given that the B^0_s modes are about three times less abundant than the B^0_d ones, they are neglected here.

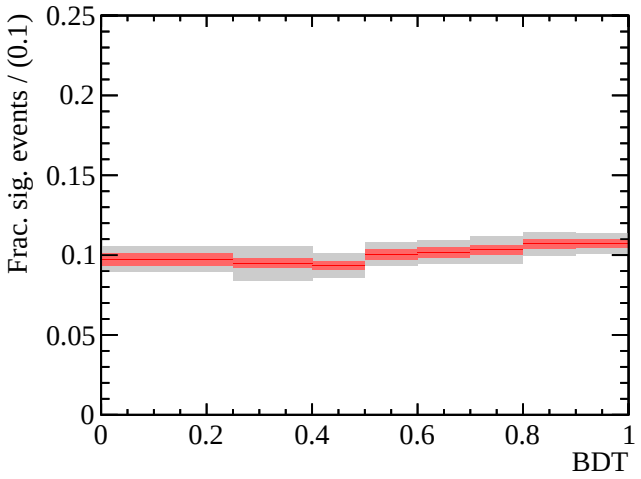

Figure 2.6: Signal PDF for the BDT output. The red band shows the statistical uncertainty, the grey one the systematic one. Values are normalized to a bin size of 0.1.

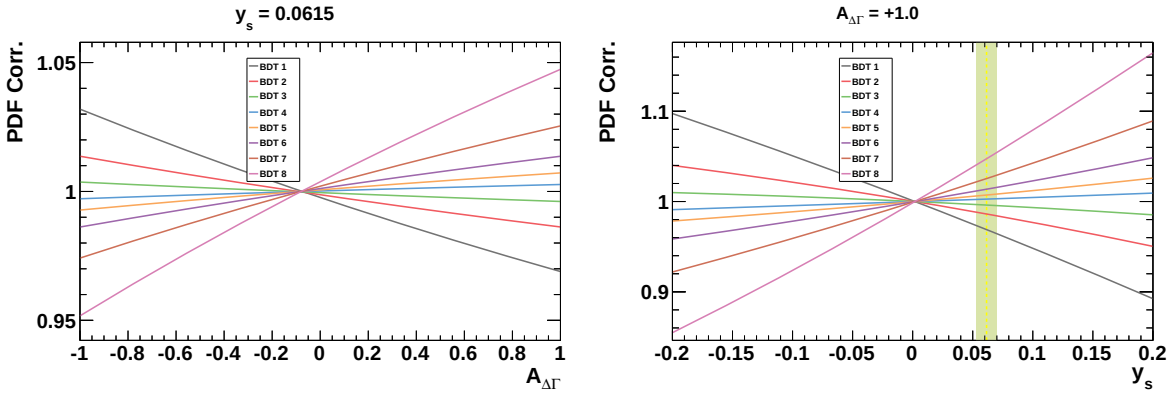

Figure 2.7: (a) Correction to be applied to BDT PDF as a function of $\mathcal{A}_{\Delta\Gamma}$ for the experimental average y_s value and (b) as a function of y_s for the SM $\mathcal{A}_{\Delta\Gamma}$ value (+1). The vertical green band is the experimental average y_s value and its uncertainty.

2.3.3 Invariant Mass Calibration

Signal PDF

The $m_{\mu\mu}$ invariant mass PDF for the B_s^0 and B_d^0 signal is modelled by a Crystal Ball function, which has four parameters: the central value μ, the width σ, the transition point α between the Gaussian function and the exponent in unit of σ and the tail exponent n.

The central value is obtained from $B^0_{(s)} \to h^+h^{(\prime)-}$ events after having applied a cut on the BDTS>0.1 and PID requirements to separate the different final states. Figure 2.8 shows the invariant mass distribution for the four mass hypotheses ($m_{\pi\pi}$, $m_{K\pi}$, $m_{\pi K}$ and m_{KK}) from the selected events in the 2012 $B^0_{(s)} \to h^+h^{(\prime)-}$ sample. The central value for the B_s^0 is taken from the $B_s^0 \to K^+K^-$ channel whereas the one for the B_d^0 is obtained averaging the three final states $\pi\pi$, $K\pi$ and πK. Systematic uncertainties are computed varying the BDTS and PID cuts.

The invariant mass resolution is estimated from two methods. One uses the interpolation of

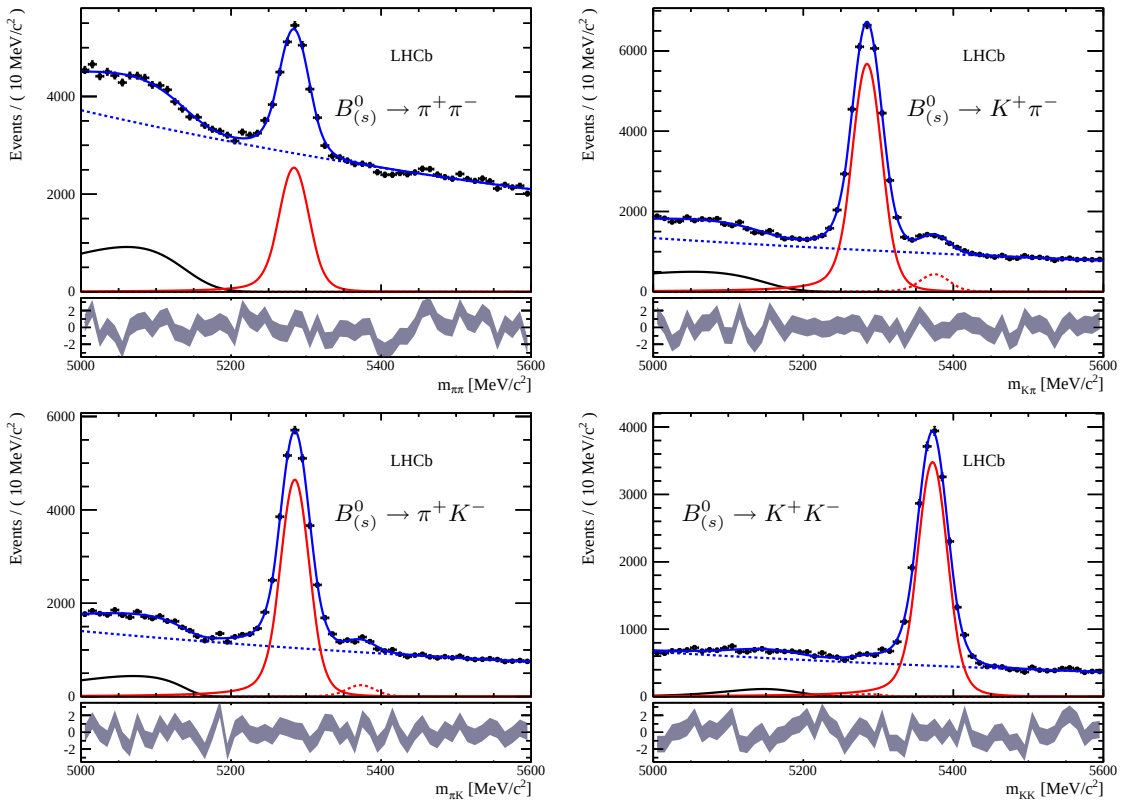

Figure 2.8: Invariant mass distribution of $B^0_{(s)} \to h^+h^{(\prime)-}$ for 2012 data separated into the different decay channels. These distributions are used for the determination of the central value of the invariant mass distribution $m_{\mu\mu}$ The full red lines indicates the dominant signal model, the dashed red the sub-dominant (e.g. upper right: full is $B^0_d \to K^+\pi^-$ and dashed is $B^0_s \to K^+\pi^-$). The black curve to the left indicates partially reconstructed background.

Table 2.3: Central values and resolutions for 2011 and 2012 data.

	2011	2012
m_{B^0}	$(5284.89 \pm 0.12 \pm 0.22)$ MeV/c^2	$(5284.91 \pm 0.17 \pm 0.19)$ MeV/c^2
$m_{B_s^0}$	$(5371.96 \pm 0.22 \pm 0.22)$ MeV/c^2	$(5371.96 \pm 0.22 \pm 0.22)$ MeV/c^2
σ_{B^0}	$(22.84 \pm 0.14 \pm 0.41)$ MeV/c^2	$(22.83 \pm 0.08 \pm 0.43)$ MeV/c^2
σ_{Bs}	$(23.22 \pm 0.18 \pm 0.44)$ MeV/c^2	$(23.24 \pm 0.09 \pm 0.43)$ MeV/c^2

the invariant mass resolution of charmonium and bottomonium decays, as shown in Fig. 2.9. The second one uses the same technique as for the BDT calibration with $B^0_{(s)} \to h^+h^{(\prime)-}$ decays. The final value is taken as the weighted average of the two methods. A summary of the central values and resolutions is given in Table 2.3 for 2011 and 2012 data. As they are compatible, they are combined using their average weighted by the statistical error. The central values are found to be about 0.1% higher than the nominal values of the B^0_s and B^0_d mass.

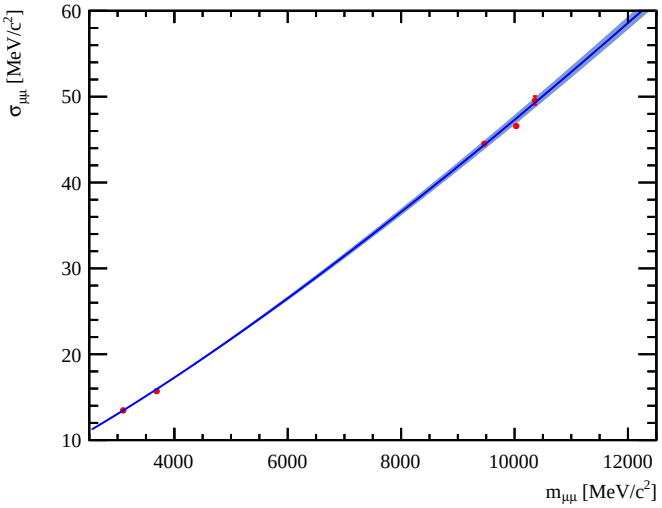

Figure 2.9: Interpolation of the invariant mass resolution between charmonium and bottomonium resonances (red points) for 2012 data. The blue line represents a fit of a power-law function.

The parameters α and n are obtained from simulation, smearing the dimuon invariant mass so that it has the same resolution as the one obtained from data. They are found to be similar for the B^0_s and B^0_d and 2011 and 2012 dataset. The combined values are:

$$\alpha = 2.065 \pm 0.005 \pm 0.010, \quad n = 1.118 \pm 0.013 \pm 0.038. \tag{2.7}$$

Background PDFs

The invariant mass fit is performed in the range $[4900, 6000]\,\mathrm{MeV}/c^2$. The lower boundary was chosen to avoid contamination from cascade $b \to c\mu X \to \mu\mu X$ decays. If the majority of the background entering the fit region is combinatorial and can be modelled by an exponential function, physics backgrounds can pollute the left sidebands region and/or the B^0_d windows and have thus to be taken into account in the fit. They can be split into three categories:

- $B^0_{(s)} \to h^+h^{(\prime)-}$ decays where the two final states hadrons are misidentified as muons. This mainly happen when the two hadrons decay in flight. To estimate the number of expected background from these decays, the pion and kaon misidentification probabilities are determined from a data sample of $D^{*\pm} \to D^0\pi^\pm$ with $D^0 \to K\pi$ decays and convolved with the momentum spectrum of simulated $B^0_{(s)} \to h^+h^{(\prime)-}$ events. The double misid probability $hh \to \mu\mu$ is of the order of 10^{-5}. The dimuon mass lineshape is obtained smearing the track momenta of the two hadrons to emulate the missing momentum due to the neutrino and rebuilding the B invariant mass. The BDT shape is evaluated from the double misidentification probability in each BDT bin.

- Semileptonic B decays with one misidentified hadron: $B^0_d \to \pi^-\mu^+\nu$, $B^0_s \to K^-\mu^+\nu$, $\Lambda^0_b \to p\mu^-\nu$. The branching fraction of these channels are of the order of 10^{-4} but their expected contributions differ due to different misidentification rates, mass shifts, and the different production fraction of B^0_d, B^0_s and Λ^0_b hadrons. The expected number of events is obtained by normalising to the $B^+ \to J\Psi K^+$ channel, taking the selection efficiency from simulation and the misididentification rate from data control samples. The invariant mass PDFs and the fraction of events in each BDT bin are taken from simulation.

- Semileptonic B decays with two muons coming from the same vertex: $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, which are FCNC and have branching fraction of the order of 10^{-8}. Their contributions are evaluated in the same way as the previous category. An additional component could come from $B^+_c \to J\Psi(\to \mu\mu)\mu^+\nu$ decays but it has a negligible contribution in the most sensitive BDT bins.

Background from $B \to \mu^+\mu^-\gamma$ decays is expected to be negligible since the predicted branching fraction for photons with an energy smaller than $60\,\mathrm{MeV}/c^2$ is $\mathcal{B}(B \to \mu^+\mu^-\gamma) = 1.6\times10^{-12}$ [43].

The number of events for all the exclusive decays described in this section estimated in the mass range $[4900, 6000]\,\mathrm{MeV}/c^2$ and normalised to a luminosity of $3\,\mathrm{fb}^{-1}$ are listed in Table 2.4. In the final strategy adopted for the invariant mass fit, the $B^0_{(s)} \to h^+h^{(\prime)-}$ and $B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$ backgrounds are added as separate PDFs, the $B^0_d \to \pi^-\mu^+\nu$ and $B^0_s \to K^-\mu^+\nu$ contributions are added into a single PDF as they have very similar shapes and the $\Lambda^0_b \to p\mu^-\nu$ component is only added as a systematic uncertainty.

Table 2.4: Number of events expected in the 2011 and 2012 data samples for all the dominant exclusive background sources estimated in the mass range [4900-6000] MeV/c^2 and fraction of events with BDT>0.7.

	$2011+2012,\ 3\,\mathrm{fb}^{-1}$	BDT>0.7(%)
$B^0_d \to \pi^-\mu^+\nu$	115 ± 6	15
$B^0_s \to K^-\mu^+\nu$	10 ± 4	21
$\Lambda^0_b \to p\mu^-\nu$	70 ± 30	11
$B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$	28 ± 8	15
$B^0_{(s)} \to h^+h^{(\prime)-}$	15 ± 1	28

2.4 Normalisation

The number of observed signal events is translated into a branching fraction according to Eq. 2.2. The normalisation factor $\alpha_{(s)}$ takes the form:

$$\alpha_{(q)} = \frac{\mathcal{B}_{norm}}{N^{obs}_{\rm norm}} \times \frac{\epsilon_{\rm norm}}{\epsilon_{\rm sig}} \times \frac{f_{\rm norm}}{f_q}, \tag{2.8}$$

where 'norm' refers to the two normalisation channels $B^+ \to J/\psi K^+$ and $B^0 \to K^-\pi^+$ with corresponding branching fractions given in Table 2.5.

Table 2.5: Branching fraction of the two normalisation channels as taken from the PDG 2012.

$\mathcal{B}\,(B^+ \to J/\psi K^+)$	$\mathcal{B}\,(B^0 \to K^-\pi^+)$
$(6.025 \pm 0.205) \times 10^{-5}$	$(1.94 \pm 0.06) \times 10^{-5}$

The efficiencies $\epsilon_{\rm sig/norm}$ are separated into three contributions refering to three different steps of the selection process, $\epsilon = \epsilon^{gen} \times \epsilon^{rec,sel|gen} \times \epsilon^{trig|sel}$:

- ϵ^{gen} is the efficiency to have all the final state particles inside the LHCb acceptance. It is evaluated from the simulation and is of the order of 16%.
- $\epsilon^{rec,sel|gen}$ is the efficiency of reconstruction and selection for the candidates that are inside the acceptance. It is evaluated from simulation taking into account possible data/MC differences in the systematic uncertainties. Since these efficiencies appear as ratio in the normalisation factor, most of systematic effects cancel. The residual effects are coming from the extra track in the case of the $B^+ \to J/\psi K^+$ channel and the differences between muon and hadron tracking system and PID requirement for the $B^0 \to K^-\pi^+$ channel.
- $\epsilon^{trig|sel}$ is the efficiency of triggering events that have been reconstructed and selected. The trigger efficiencies are estimated using the TIS and TOS candidates, according to the method explained in [44]. For the $B^+ \to J/\psi K^+$ control channel, this measurement can be performed directly on the data and the efficiency is found to be $\sim$ 87%. For the $B^0_s \to \mu^+\mu^-$ signal, a map from detached $J/\psi \to \mu^+\mu^-$ is created in the maximum p_T and IP of the muons which is then applied to the $B^0_s \to \mu^+\mu^-$ spectrum taken from the simulation, resulting in an efficiency of $\sim$ 92%. Only muon trigger lines are used in this process, the fraction of events which are selected by non-muon triggers (about 2%) is taken as additional uncertainty.

  As for the BDT calibration, $B^0 \to K^-\pi^+$ events are required to be TIS at the L0 and HLT1 level to minimize the biases from the hadronic triggers The TIS efficiency is obtained from the $B^+ \to J/\psi K^+$ data and is at the level of $\sim$ 5%. The HLT2 efficiency on top of the (L0xHLT1) TIS events is taken from simulation and is of the order of 92%.

The number of candidates for the two normalisation channels are obtained from fits to the invariant mass distributions. Figure 2.10 shows the invariant mass distribution of the events passing the $B^+ \to J/\psi K^+$ selection and the fit result. The signal distribution is modelled with a double Crystal Ball function, the combinatorial background is modelled by an exponential function and physical background coming from misidentified $B^+ \to J/\psi \pi^+$ by a Gaussian function. The fit to

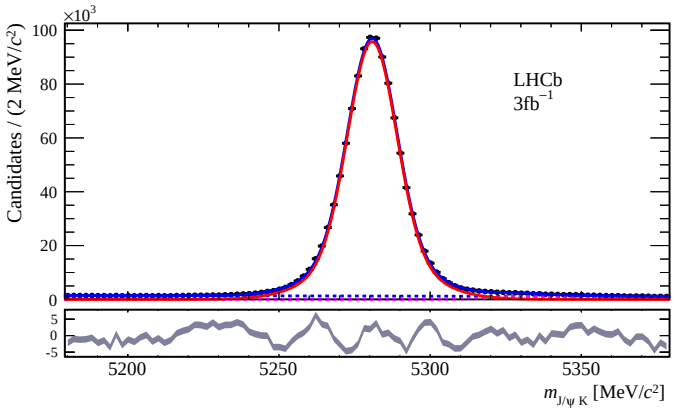

Figure 2.10: Invariant mass distribution of the $B^+ \to J/\psi K^+$ candidates, after constraining the J/Ψ mass to the PDG value. A fit to the distribution is superimposed (blue continuous line) with signal (red continuous line), combinatorial background (blue dashed line) and misidentification background from $B^+ \to J/\psi\pi^+$ decays (magenta dashed line).

Table 2.6: Number of observed events for the two normalisation channels.

year	$N_{B^+\to J/\psi K^+}$	$N_{B^0\to K^-\pi^+}$
2011	355233 ± 608	10809 ± 439
2012	761122 ± 891	26749 ± 447

the $B^0 \to K^-\pi^+$ candidates follows the same procedure as for the BDT calibration.The number of events for the two normalisation channels are summarized in Table 2.6.

The last missing ingredient to the normalisation factor is the ratio of hadronization fractions. They have been measured by LHCb in two ways: using the relative abundance of $B^0_s \to D^-_s\pi^+$, $B^0 \to D^-K^+$ and $B^0 \to D^-\pi^+$ [45] and using semileptonic $B \to DX$ decays [46]. The combination of these two measurements results in the value [47, 48]

$$f_s/f_d = 0.259 \pm 0.015\,. \tag{2.9}$$

The two channels give compatible values for the normalisation factor. They are averaged taking the uncertainty on f_s/f_d to be correlated, giving:

$$\alpha_{B^0_s\to\mu^+\mu^-} = (9.01 \pm 0.62) \times 10^{-11}\,, \tag{2.10}$$

$$\alpha_{B^0_d\to\mu^+\mu^-} = (2.40 \pm 0.09) \times 10^{-11}\,. \tag{2.11}$$

Assuming the SM $B^0_{(s)} \to \mu^+\mu^-$ branching fractions, the full Run1 data sample is expected to contain 40 ± 4 $B^0_s \to \mu^+\mu^-$ and 4.5 ± 0.4 $B^0_d \to \mu^+\mu^-$ events in the mass range $[4900, 6000]\,\mathrm{MeV}/c^2$.

2.5 Results

The result of the fit to the invariant mass distribution of unblinded data is shown in Fig. 2.11 for each BDT bin. The corresponding branching fraction values are:

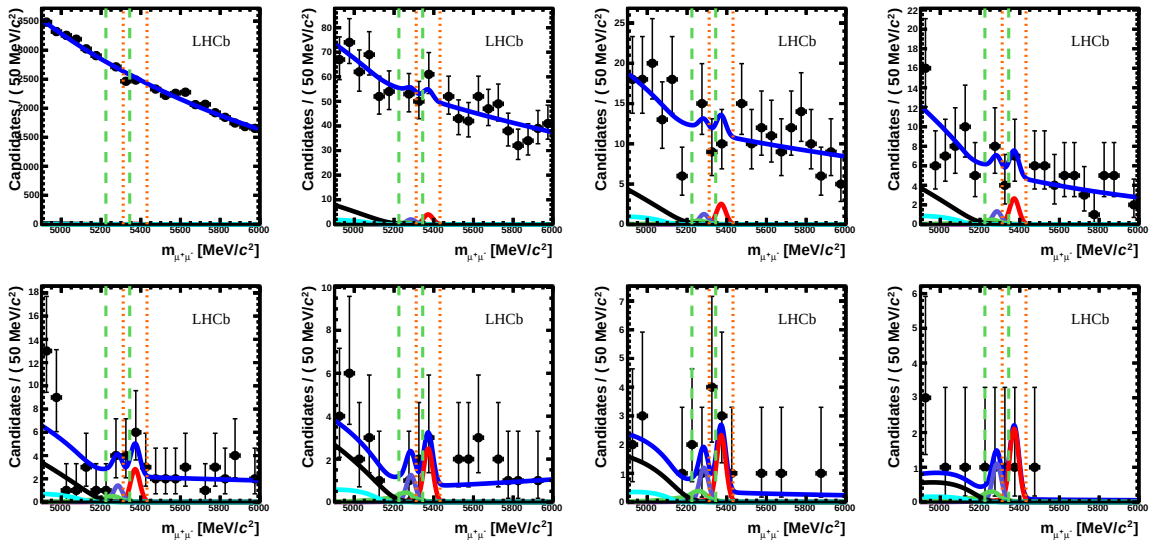

Figure 2.11: Invariant mass distribution of the selected signal candidates (black dots) for the 8 BDT bins. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \to \mu^+\mu^-$ (red), $B_d^0 \to \mu^+\mu^-$ (purple), $B_{(s)}^0 \to h^+h^{(\prime)-}$ (green), $B^0 \to \pi^-\mu^+$ and $B_s^0 \to K^-\mu^+$ (black), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan).

$$\begin{aligned}\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (2.9\,^{+1.1}_{-1.0}(\text{stat})\,^{+0.3}_{-0.1}(\text{syst})) \times 10^{-9},\\ \mathcal{B}(B_d^0 \to \mu^+\mu^-) &= (3.7\,^{+2.4}_{-2.1}(\text{stat})\,^{+0.6}_{-0.4}(\text{syst})) \times 10^{-10}.\end{aligned}$$

The statistical uncertainty reflects the interval corresponding to a change of 0.5 in the log-likelihood after fixing all the nuisance parameters to their expected values, and the systematic uncertainty takes into account the uncertainties due to the signal and background mass PDFs as well as the addition of the $\Lambda_b^0 \to p\mu^-\nu$ component. The signal significance obtained from the difference in log-likelihood between signal and null hypotheses corresponds to a $4.0\ \sigma$ evidence of $B_s^0 \to \mu^+\mu^-$ and $2.0\ \sigma$ for $B_d^0 \to \mu^+\mu^-$. The values of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions obtained from the fit are in agreement with the SM expectations. To better visualize the signal, the invariant mass distribution of the $B_{(s)}^0 \to \mu^+\mu^-$ candidates with $\text{BDT} > 0.7$ is shown in Fig. 2.12.

As no significant excess of $B_d^0 \to \mu^+\mu^-$ events is found, a modified frequentist approach, the CL_s method [35] is used, to set an upper limit on the branching fraction. The method provides $\text{CL}_\text{s+b}$, a measure of the compatibility of the observed distribution with the signal plus background hypothesis, CL_b, a measure of the compatibility with the background-only hypothesis, and $\text{CL}_\text{s} = \text{CL}_\text{s+b}/\text{CL}_\text{b}$. A search region is defined around the B_d^0 invariant mass as $m_{B_d^0} \pm 60\,\text{MeV}/c^2$. For each BDT bin the invariant mass signal region is divided into nine bins with boundaries $m_{B^0} \pm 18, 30, 36, 48, 60\,\text{MeV}/c^2$, leading to a total of 72 search bins.

An exponential function is fitted, in each BDT bin, to the invariant mass sidebands. The uncertainty on the expected number of combinatorial background events per bin is determined by applying a Poissonian fluctuation to the number of events observed in the sidebands and by varying the exponential slopes according to their uncertainties. In each bin, the expectations for $B_s^0 \to \mu^+\mu^-$ decays assuming the SM branching fraction and for $B_{(s)}^0 \to h^+h^{(\prime)-}$ background are accounted for. For each branching fraction hypothesis, the expected number of signal events is estimated from the normalisation factor. In each bin, the expected numbers of signal and

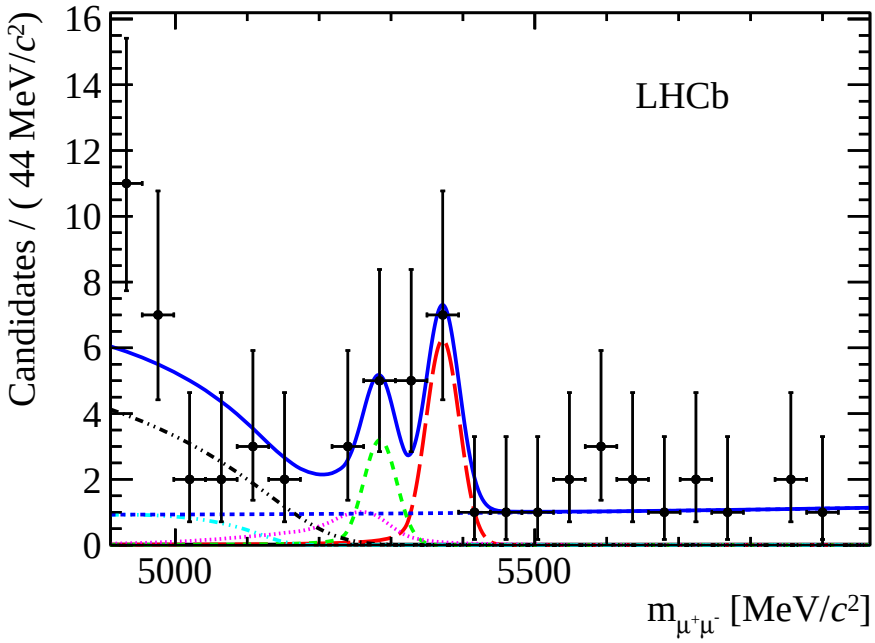

Figure 2.12: Invariant mass distribution of the selected signal candidates (black dots) with $\mathrm{BDT} > 0.7$. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \to \mu^+\mu^-$ (red long dashed), $B_d^0 \to \mu^+\mu^-$ (green medium dashed), $B_{(s)}^0 \to h^+h^{(\prime)-}$ (magenta dotted), $B^0 \to \pi^-\mu^+$ and $B_s^0 \to K^-\mu^+$ (black dot-dashed), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (light blue dot dashed), and the combinatorial background (blue medium dashed).

background events are computed and compared to the number of observed candidates using $\mathrm{CL_s}$. The expected and observed upper limits for the $B_d^0 \to \mu^+\mu^-$ channel are summarised in Table 2.7 and the expected and observed $\mathrm{CL_s}$ values as functions of the branching fraction are shown in Fig. 2.13.

2.6 Combination with CMS

The CMS experiment also searched for the $B_{(s)}^0 \to \mu^+\mu^-$ decays using a similar strategy on the Run1 data corresponding to a luminosity of $25\,\mathrm{fb}^{-1}$ [49]. The combination of LHCb and CMS

Table 2.7: Expected limits for the background only (bkg) and background plus SM signal (bkg+SM) hypotheses, and observed limits on the $B_d^0 \to \mu^+\mu^-$ branching fraction.

	90 % C.L.	95 % C.L.
Exp. bkg	3.5×10^{-10}	4.4×10^{-10}
Exp. bkg+SM	4.5×10^{-10}	5.4×10^{-10}
Observed	6.3×10^{-10}	7.4×10^{-10}

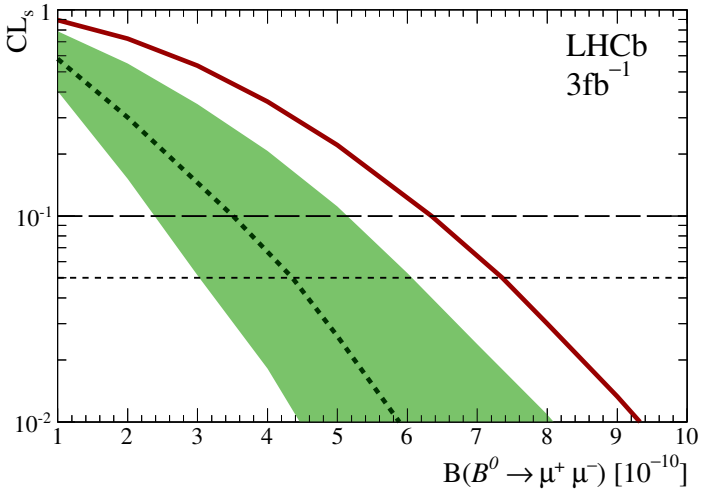

Figure 2.13: $\mathrm{CL_s}$ as a function of the assumed $B^0_d \to \mu^+\mu^-$ branching fraction. The dashed curve is the median of the expected $\mathrm{CL_s}$ distribution for background-only hypothesis. The green area covers, for each branching fraction value, $34.1\,\%$ of the expected $\mathrm{CL_s}$ distribution on each side of its median. The solid red curve is the observed $\mathrm{CL_s}$.

results has been performed fitting simultaneously the two datasets [22], split according to 20 categories : experiment, BDT value, and for CMS, the data taking period and central or forward region. Figure 2.14 shows the fit result overlaid to the dimuon invariant mass weighted by the factor $s/\sqrt{s+b}$, where s is the expected number of B^0_s signal and b the background yield under the B^0_s peak in that category. An excess of event at 6.2σ is seen for the B^0_s decay, giving the first observation. The corresponding measured branching fraction is

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = 2.8^{+0.7}_{-0.6} \times 10^{-9}. \tag{2.12}$$

A first evidence for the B^0_d mode is seen at 3.0σ, with a measured branching fraction

$$\mathcal{B}(B^0_d \to \mu^+\mu^-) = 3.9^{+1.6}_{-1.4} \times 10^{-8}. \tag{2.13}$$

The likelihood contours of $\mathcal{B}\,(B^0_s \to \mu^+\mu^-)$ vs $\mathcal{B}\,(B^0_d \to \mu^+\mu^-)$, as well as as the one-dimensional likelihood scans are shown in Fig. 2.15. The measurements are compatible with the SM branching fractions of the $B^0_s \to \mu^+\mu^-$ and $B^0_d \to \mu^+\mu^-$ at 1.2 and 2.2 σ respectively. The ratio of the two branching fractions, $\mathcal{R} = \frac{\mathcal{B}(B^0_d \to \mu^+\mu^-)}{\mathcal{B}(B^0_s \to \mu^+\mu^-)}$, provides an observable sensitive to the minimal flavour violation hypothesis. It is found to be $\mathcal{R} = 0.14^{+0.08}_{-0.06}$, which is compatible at 2.3 σ with the SM value.

2.7 Prospects

As said previously, the LHCb measurement has been updated adding $1.4\,\mathrm{fb}^{-1}$ recorded at $\sqrt(s) = 13\ \mathrm{TeV}$ [16]. Thanks to improved isolation variables and the optimisation of the muon identification requirements, a gain equivalent to the reduction of the statistical uncertainty by 30% on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ has been obtained. There have been a total of $5.9\,\mathrm{fb}^{-1}$ recorded at $\sqrt(s) = 13\,\mathrm{TeV}$; these additional data are being analysed.

On a longer term timescale, the improvement of the $B^0_{d,s} \to \mu^+\mu^-$ results will depend on the integrated luminosity. LHCb is currently performing an upgrade (called Upgrade I) [50], which

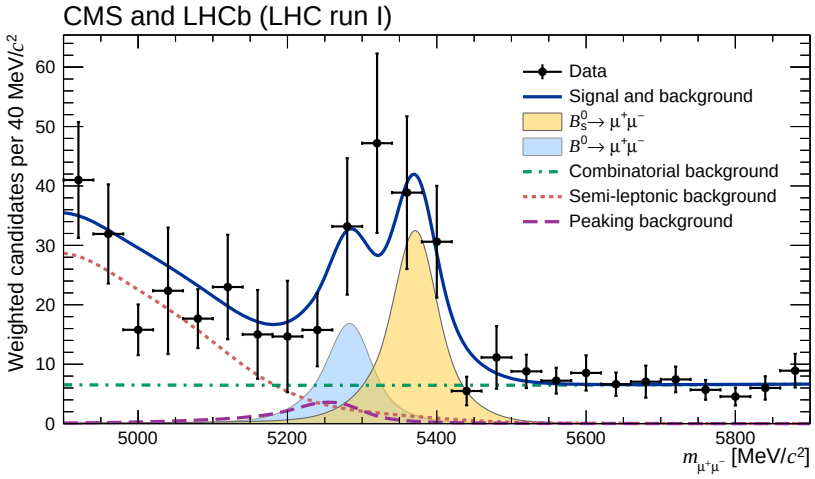

Figure 2.14: Weighted distribution of the dimuon invariant mass. Superimposed on the data points in black are the combined fit (solid blue line) and its components: the B_s^0 (yellow shaded area) and B_d^0 (light-blue shaded area) signal components; the combinatorial background (dash-dotted green line); the sum of the semileptonic backgrounds (dotted salmon line); and the peaking backgrounds (dashed violet line).

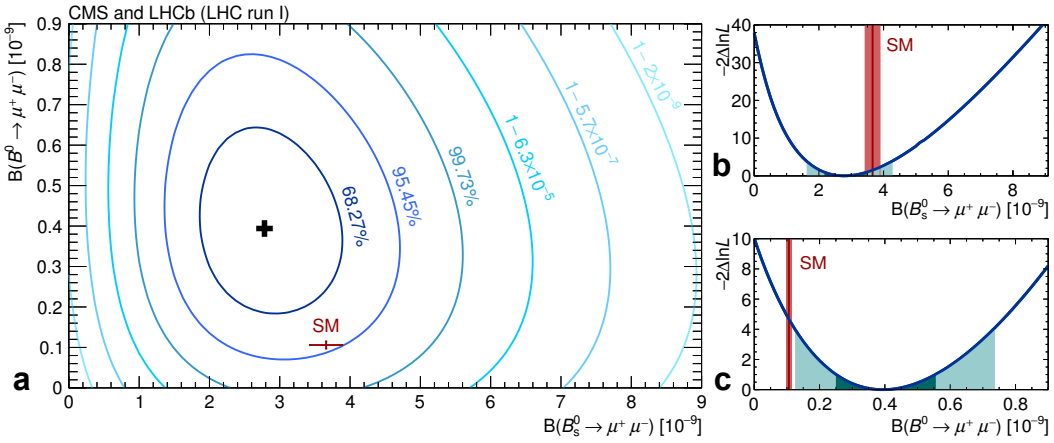

Figure 2.15: (a) Likelihood contours in the $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ versus $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ plane. The (black) cross in a marks the best-fit central value. The SM expectation and its uncertainty is shown as the (red) marker. Each contour encloses a region approximately corresponding to the reported confidence level. Variations of the test statistic $-2\Delta Ln$ for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ (b) and $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ (c). The dark and light (cyan) areas define the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals for the branching fraction, respectively. The SM prediction and its uncertainty for each branching fraction is denoted with the vertical red band.

consists in removing the L0 trigger and reading out the entire detector at the LHC bunch crossing rate. The main detector evolutions concern the electronic and the tracking system, that require higher granularity. By implementing all trigger decisions in software, a luminosity of 2×10^{33} cm^{-2} s^{-1} can be achieved without suffering from efficiency loss. An integrated luminosity of $23\,\mathrm{fb}^{-1}$ is expected at the end of Run3 in and $50\,\mathrm{fb}^{-1}$ at the end of Run4. A second upgrade of the LHCb detector (Upgrade II) that could take place during the long shutdown 4, in 2030, is under

discussion, targetting an instantaneous luminosity of 2×10^{-34} cm^{-2}s^{-1} [51]. This Upgrade II would exploit the so-called High-Luminosity LHC (HL-LHC) operational period and record a final data set of 300 fb^{-1}. This will necessitate major improvements of the different sub-detectors to increase the granularity and make use of precision timing. Additional tracking stations could be installed in the magnet, enhancing the reconstruction efficiency of low momentum particles[3]. The luminosity projection of LHCb can be seen on Fig. 2.16.

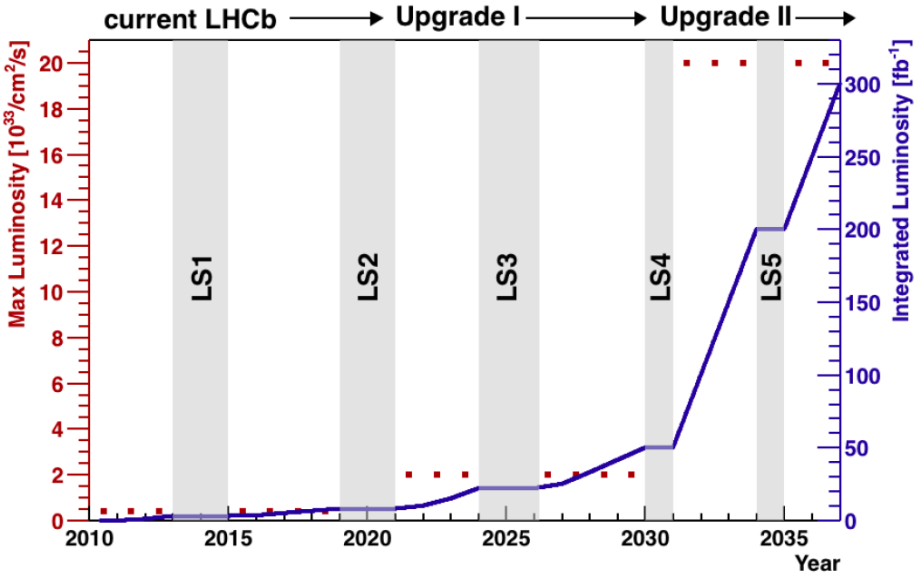

Figure 2.16: Luminosity projections for the original LHCb, Upgrade I and Upgrade II experiments as function of time.

While the $\mathcal{B}(B^0_d \to \mu^+\mu^-)$ measurement will remain statistic-dominated, the projection of the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ one depends on the assumption made for the systematic uncertainties. The dominant one is currently due to the uncertainty on f_s/f_d (5.8%), followed by the branching fraction of the normalization modes (3%). A conservative estimate leads to an uncertainty on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ of 0.30×10^{-9} for 23 fb^{-1} and 0.16×10^{-9} for 300 fb^{-1} [51], reaching the current precision of the SM prediction. The uncertainty on the ratio of the two decay modes is expected to be 34% for 23 fb^{-1} and 10% for 300 fb^{-1}.

As LHCb, the ATLAS and CMS experiments have not yet made public the analysis of the full Run2 statistics. Results are expected in the coming months. They both provide projections for the end of the HL-LHC period, assuming 3 ab^{-1} of data recorded. Both experiments will benefit from an upgraded tracking system with more granularity, improving the dimuon mass resolution. In the ATLAS projection, three values are envisaged for the dimuon trigger thresholds, leading to three scenari for the extrapolated statistics: conservative (15×Run1 statistics), intermediate (60×Run1 statistics), and high-yield (75×Run1 statistics). The total uncertainty obtained on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ varies between 0.46×10^{-9} to 0.55×10^{-9} depending on the scenario [52]. CMS performed an equivalent study [53] based on an improved analysis strategy developped for Run2 data. They reach a relative uncertainty of 7% on the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$, which corresponds to 0.25×10^{-9} for 3 ab^{-1}.

[3]This tracking stations could already be installed during LS3.

Personnal contributions
My main contributions were in the BDT calibration and the evaluation of trigger efficiencies and biases for signal and normalisation channels. I also participated to other steps of the analysis, in particular: optimisation of the selection, comparison of different reprocessing and data taking periods, crosscheck of different NTuples productions, setup of a blinding procedure.

Chapter 3

Search for the $B^0_{d,s} \to \tau^+\tau^-$ decays

If the measurement principle is the same as for the $B^0_{d,s} \to \mu^+\mu^-$ decays (see Sec. 2.1.1), the $B^0_{d,s} \to \tau^+\tau^-$ case is much trickier. Indeed, the di-tau invariant mass can not be used as a discriminating variable due to the (at least) two final-state neutrinos that can not be detected. Another complication comes from the fact that the amount of background is too large to be simulated so the analysis must be entirely data-driven. A different strategy than that for the $B^0_{d,s} \to \mu^+\mu^-$ decay needs then to be set up to optimize the background rejection, measure the signal yield and have an estimate of the expected background. In addition, in the absence of sufficient invariant mass resolution it is not possible to distinguish the B^0_s from the B^0_d mesons, as illustrated in Fig. 3.1. An hypothesis must be made on the branching fraction of one of the two to set a limit on the other. In the rest of this chapter, the $B^0_d \to \tau^+\tau^-$ mode is omitted for practical purpose and only appears when presenting the results. This analysis is performed on the Run 1 data sample. Section 3.1 presents the analysis strategy. The selection, signal yield extraction and normalisation are detailed in Sec. 3.2, Sec. 3.4 and Sec. 3.3. Section 3.5 presents the results. Finally, some prospects are given in Sec. 3.6.

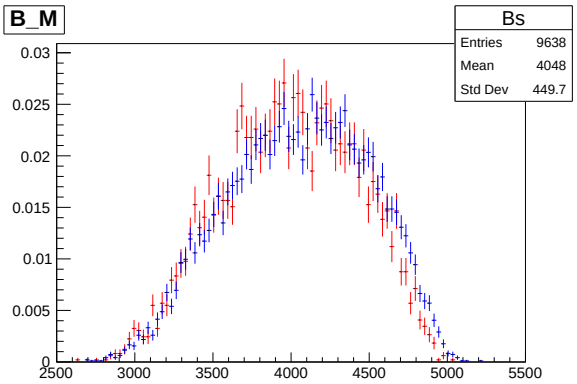

Figure 3.1: Reconstructed (visible) invariant mass of the B_s^0 (blue) and B_d^0 (red) meson when both τ leptons decay into three charged pions, as obtained from the simulation.

Table 3.1: Branching fraction of the main τ decays [34].

$\tau^+ \to \mu^+ \bar{\nu}_\tau \nu_\mu$	17.39 ± 0.04
$\tau^+ \to e^+ \bar{\nu}_\tau \nu_e$	17.82 ± 0.04
$\tau^+ \to \pi^+ \bar{\nu}_\tau$	10.82 ± 0.05
$\tau^+ \to \pi^+ \pi^0 \bar{\nu}_\tau$	25.49 ± 0.09
$\tau^+ \to \pi^+ \pi^0 \pi^0 \bar{\nu}_\tau$	9.26 ± 0.10
$\tau^+ \to \pi^+ \pi^- \pi^+ \bar{\nu}_\tau$	9.31 ± 0.05
$\tau^+ \to \pi^+ \pi^- \pi^+ \pi^0 \bar{\nu}_\tau$	4.62 ± 0.05

3.1 Analysis strategy

3.1.1 Choice of τ decay

The main τ decays and their corresponding branching fractions are reported in Table 3.1. The relatively low π^0 and electron reconstruction efficiency of the LHCb detector makes the $\tau^+ \to \mu^+ \bar{\nu}_\tau \nu_\mu$, $\tau^+ \to \pi^+ \bar{\nu}_\tau$ and $\tau^+ \to \pi^+ \pi^- \pi^+ \bar{\nu}_\tau$ decays the best options. The muonic final state has the highest trigger and reconstruction efficiency. The one-pion final state has a better trigger and reconstruction efficiency than the three-pion one but the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ decay has two important advantages. Firstly, it proceeds predominantly via intermediate $a_1^-(1260)$ and $\rho^0(770)$ resonances as [34]

$$\tau^- \to a_1^-(1260)\nu_\tau \to (\rho^0(770) \to \pi^+\pi^-)\pi^-\nu_\tau \,. \tag{3.1}$$

This feature is extensively exploited in the analysis as it offers interesting handles to discriminate signal from background and provides a control region, as it will be explained in Sec 3.1.4. Secondly, the τ decay vertex can be reconstructed. In the case where both τ leptons are reconstructed in the 3π final state, as illustrated in Fig. 3.2, an (approximate) reconstruction of the complete decay chain can be performed [54]. This in turn introduces additional variables that are well suited to discriminate between signal and background.

The $B \to \tau^+(\to \pi^+\pi^-\pi^+\bar{\nu}_\tau)\tau^-(\to \mu^-\bar{\nu}_\mu\nu_\tau)$ could appear as a good compromise, benefiting both from the high efficency of the muon trigger and the intermediate ρ resonance but deeper investigations, performed in the PhD of Cédric Méaux, have demonstrated that it is not the case [55]. In the following, the analysis performed with both tau leptons reconstructed into the three

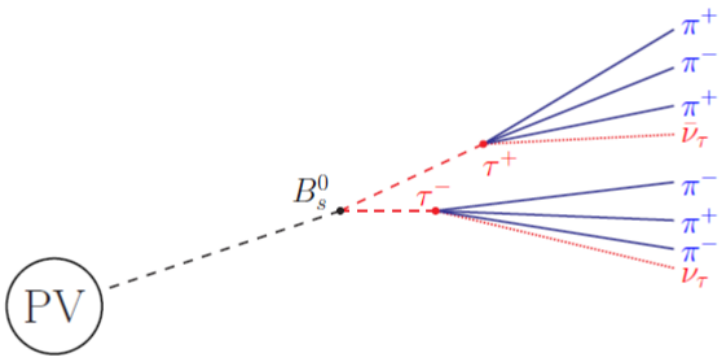

Figure 3.2: Schematic representation of the $B_s^0 \to \tau^+\tau^-$ decay, where both τ leptons are reconstructed in the 3π final state.

pions final state is presented. This analysis has been the main subject of the post doctorate Kristof de Bruyn and has been published in 2017. The study of the $(3\pi, \mu)$ final state is briefly reviewed in the last Section of this Chapter.

3.1.2 Analysis overview

The analysis workflow is pretty similar to the muonic final state, the major analysis steps being the selection of signal and normalisation channel, a fit to a discriminating variable to measure the signal yield and the limit computation (in the absence of signal). The major difference stands in the fitted variable. For the τ case, the dilepton invariant mass is poorly discriminating and can not be used alone. Instead a multivariate discriminant combining kinematic and geometrical variable is built, and fitted to the data. The detailed analysis workflow is the following:

- A selection applied at the stripping level, followed by trigger requirements. It is then refined by a cut-based selection followed by a selection on the output of a neural network (NN), labelled NN1 in the following.
- The building of a second neural network, NN2, maximizing the signal to background separation.
- The computation of the normalisation using the channel $B_d^0 \to D^-(\to \pi^- K^+ \pi^-) D_s^+(\to K^+ K^- \pi^+)$. This decay has the same number of final state particles as the signal and a similar two-vertex topology.
- The fit of NN2 to the data and the computation of the limit using the $\mathrm{CL_s}$ method.

The selection and NN optimization are performed using simulated samples for the signal and normalisation channels, and background from a control region as detailed in the following. In the $B_s^0 \to \tau^+\tau^-$ simulated sample, the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ decays are generated using the Resonance Chiral Lagrangian model [56] with a tuning based on the BaBar results for the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ decays [57], implemented in the TAUOLA generator [58][1]. Generator-level cuts are applied on final state daughters in order to save disk space and CPU time. They require all pions to be in the LHCb acceptance, and that the pion (transverse) momenta are larger than 2 GeV/c ($250\,\mathrm{MeV}/c^2$). These cuts have an efficiency of about 3%.

[1]The analysis has first been performed using a simulated sample based on the TauHadNu model of EVTGEN [59].

3.1.3 Main background sources

The average number of tracks produced in LHCb at each collision being around 200, the combinatorial background is enormous when searching for a signal with six tracks in the final state,. It is however easy to reduce thanks to requirement on the τ vertex quality and invariant mass. The remaining backgrounds are mainly "physical", meaning that the six tracks are coming from the same original b-hadron. When this b-hadron has more than 6 daughters, isolation variables constitute a powerful tool to eliminate them. Decays with daughters that are not all real pions can be reduced thanks to hard requirements on PID variables. The most dangerous backgrounds are therefore signal-like $b-$hadron decays giving 6 final-state pions and additionnal undetected particle(s) mimicking missing energy. They can be classified into two categories, having different topology : semileptonic B decays, such as $B^0 \to D^{(\star)-}\tau^+\nu$ with $D^{(\star)-} \to (D^- \to (3\pi)^-\pi^0)\pi^0$ and hadronic B decays, such as $B^0 \to D^-(\to (3\pi)^- K^0)(3\pi)^+$, where the three pions from the B mostly go through the a_1 resonance.

In the early days, we naively thought that the number of these exclusive backgrounds was small enough so that we could explicitely simulate them. Thirteen channels with the highest branching frations have thus been generated. Nevertheless, this approach suffers from two problems. On one hand, this list may not be exhaustive as some decays are not well known: their branching fraction or decay model are not or badly measured. This is for example the case of the $D^- \to (3\pi)^-\pi^0$, for which the contribution of $a_1\pi^0$ or $\rho^+\rho^0$ has to be guessed. On the other hand, it requires the generation of very high statistics samples. Another approach was then developed to have a more 'inclusive' background caracterisation: a custom generic $b\bar{b}$ MC sample has been created from many available simulated samples of specific B decays. The idea is not to look at this specific B decay but at the other b-hadron produced in the events, since they are always produced in pairs. The simulated samples are chosen so that the specific decay is as different from our signal decay as possible (for example, $B_s^0 \to \mu^+\mu^-$). In this way, the reconstructed $B_s^0 \to \tau^+\tau^-$ candidate will, in the large majority of cases, only be made of the other b-hadron daughters. A total of 80 specific samples were used, corresponding to a total number of 744 millions generated events. Of course, even if this sample is more exhaustive, it is also affected by the problem of poorly known decays. In addition, due to the low signal efficiency, only 25 candidates remain in the signal region after having applied the full selection. This sample is therefore only used as a cross-check in the analysis and a data driven strategy need to be defined for the background evaluation.

3.1.4 Control region

Numerous options have been studied in order to find a control region allowing to caracterise and estimate the amount of background in data. Same sign data, where the two τ leptons are reconstructed with the same charge, turned out to be useless as the discriminating variables were found to have different distributions than in opposite sign data. This is particularly true for the isolation variables, and can be understood from the fact that same-sign candidates are necessarily formed from a $b-$hadron decay having at least seven final-state tracks (if it is a B^+) or eight (if it is a B_d^0).

Other regions like τ mass sidebands, or from an anti-selection based on isolation variables, were found to be too different from the signal region as well. The best solution was obtained using the two-dimensional distribution of the invariant masses $m_{\pi^+\pi^-}$ of the oppositely-charged final state pions. Labelling them π_1, π_2, π_3 as $\tau^+ \to a_1^+(1260)\bar{\nu}_\tau \to (\rho^0(770) \to \pi_1^+\pi_2^-)\pi_3^+\bar{\nu}_\tau$, true τ events form a plus shape in a "Dalitz"-like plane[2] spanned by the two invariant mass variables

[2]For ease of use, we refer to this plot as the Dalitz plane, even though a Dalitz plot is classically defined in

$m_{\pi_1^+\pi_2^-}$ and $m_{\pi_3^+\pi_2^-}$, as illustrated in Fig. 3.3. Using two cuts on each of the two invariant mass variables, the Dalitz plot can be divided into nine rectangles. Based on this division of both the

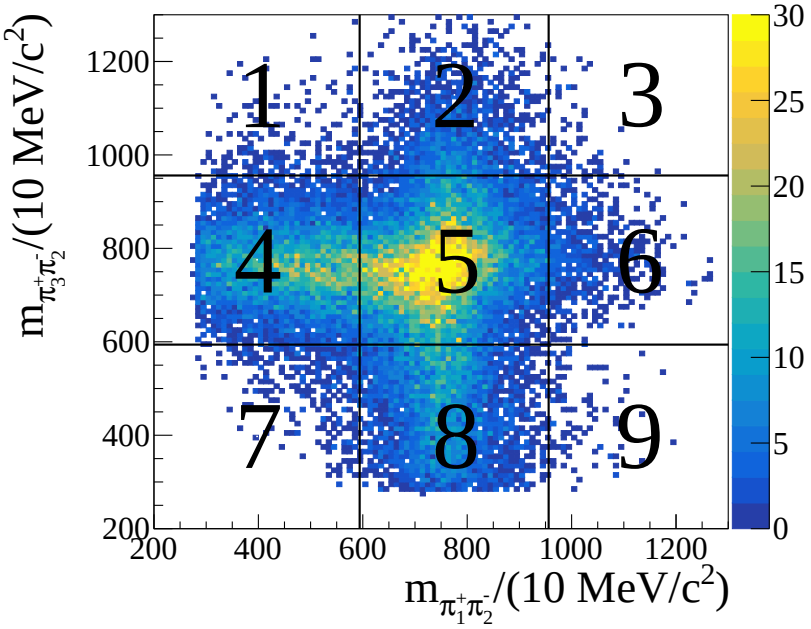

Figure 3.3: Illustration of the Dalitz-like plot for simulated $B_s^0 \to \tau^+\tau^-$ events, used in this analysis to divide the data into three mutually exclusive subsamples. The definitions of these subregions are given in terms of the nine different rectangles highlighted in this cartoon. This notation is used throughout this document.

τ^+ and τ^- Dalitz planes, the data can be split into mutually exclusive subsamples. Specifically, the analysis uses the following three subregions:

Signal region: (τ^+ in box 5) and (τ^- in box 5)
Both τ candidates are required to lie inside the middle rectangle, highlighted as "5" in Fig. 3.3. These events are considered to be the most signal-like, and are the ones remaining in the final selection.

Background region: (τ^+ in boxes 1, 3, 7 or 9) or (τ^- in boxes 1, 3, 7 or 9)
At least one of the two τ candidates must end up in one of the four corners of its Dalitz plot. These events are considered to be the least signal-like, and will represent the background when optimising the selection.

Control region: (τ^+ in boxes 4, 5 or 8) and (τ^- in boxes 4 or 8), or vice versa
These events will provide a control region to estimate the background shape. As it is not completely free of signal, the signal component will be subtracted.

The 'left-over' region defined as (τ^+ in boxes 2 or 6) and (τ^- in boxes 2, 4, 5, 6 or 8) (or vice versa) was initially used as the control region but turn out to be less 'signal-like' than the one previously defined. It is therefore ignored in the following.

terms of the squares of these masses.

3.2 Selection

The selection consists of the event reconstruction and stripping, detailed in Sec. 3.2.1, the trigger requirements, listed in Sec. 3.2.2, and a further offline selection, detailed in Sec. 3.2.4.

3.2.1 Stripping

An overview of the different selection cuts for the signal and normalisation channel is given in Table 3.2. The selection uses standard variables except for the corrected mass, M_{corr}. The corrected mass variable was used for the first time in the SLD experiment [60] to correct for the missing energy from neutral and/or undetected particles in the decay chain. It includes the transverse momentum relative to the flight direction as the minimum missing momentum to the invariant mass of the charged tracks, such that the momentum vector and flight direction align. This variable is calculated as[3]

$$M_{corr} = \sqrt{M_B^2 + \left(P_B \times \sqrt{1-\theta_{DIRA}^2}\right)^2} + P_B \times \sqrt{1-\theta_{DIRA}^2}\,, \tag{3.2}$$

where M_B is the invariant mass calculated from the charged tracks, and θ_{DIRA} is the cosine of the angle between the momentum of the B and its flight direction.

The stripping is designed in order to keep the highest signal efficiency while maintaining a retention rate below an imposed limit of 0.05%. The stripping requirements select a total of ∼18 million events with an average multiplicity of 7 in the $3\,\mathrm{fb}^{-1}$ of Run1 data. The reconstruction and stripping efficiency for simulated signal events is between 2.7 and 3.0% depending on the data taking year.

3.2.2 Trigger

The signal $B_s^0 \to \tau^+\tau^-$ is mainly selected by the hadronic *L0Hadron* line, the *HLT1TrackAllL0* line, which requires one high p_T track, not attached to the PV, and the HLT2 topological lines. All these trigger lines are required to be fired by the reconstructed signal candidate (TOS). This allows to minimize the uncertainties associated to the trigger efficiencies and helps the background rejection. The HLT2 TOS requirement is already applied at the stripping level for this reason. Additional events are selected using the L0Global TIS requirement, meaning that they can have been selected by any L0 line if the reconstructed signal candidate is not involved in this decision. In that case, the trigger is generally fired by the other b-hadron of the event.

The efficiency of L0 (HLT) requirements on stripped signal candidates is around 50 (75)%.

3.2.3 Discriminating variables

Beside standard kinematic and geometric variables, this analysis makes extensive use of custom-made isolation variables and variables designed specifically for the reconstruction of the $B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$ decay. In more details they are:

- track isolations: Initially built for the $B_s^0 \to \mu^+\mu^-$ analysis, they have been reoptimized for the $B_s^0 \to \tau^+\tau^-$ case. The basic isolation variable corresponds to the number of tracks that can make a good vertex with a considered final state signal particles. In addition, we

[3]The expression $M = \sqrt{m^2+p^2}+p$ represents the invariant mass of a particle decaying into a particle with mass m and a massless particle.

Table 3.2: Overview of the different selection cuts for the signal and normalisation channel. The vertex distance are calculated between the PV and the end-vertex of the considered particle.

		$B_s^0 \to \tau^+\tau^-$		$B_d^0 \to D^-(\to \pi^-K^+\pi^-)D_s^+(\to K^+K^-\pi^+)$
Cut	on	value	on	value
Track χ^2/ndf	π	<3	π or K	<3
Ghost probability		< 0.3		< 0.3
IPχ^2		>16		>16
p_T		> 0.25 GeV/c		$> 0.25\,\mathrm{GeV}/c$
p		> 2 GeV/c		> 2 GeV/c
ProbNNpi		>0.55	π	>0.55
DLL(K-π)		N/A	K	>-5
at least 1 daughter with p_T	τ	>0.8 GeV/c	D	>0.8 GeV/c
vertex chi^2		<16		<16
maxDOCA		<0.2 mm		0.2 mm
M		[500-2000] MeV/c		[1800-2030] MeV/c
vertex distance in χ^2		>16		>16
vertex distance in ρ		$[0.1 - 7]$ mm		$[0.1 - 7]$ mm
vertex distance in z		>5.0 mm		>5.0 mm
θ_{DIRA}		>0.99		>0.99
p_T		>1 GeV/c		>1 GeV/c
p_T	B	> 2 GeV/c	B	> 2 GeV/c
M		[2-7] GeV/c		[5-7] GeV/c
M_{corr}		<10 GeV/c		<10 GeV/c
vertex χ^2		<90		<90
θ_{DIRA}		>0.99		>0.99
vertex distance in chi^2		>225		>225
vertex distance		<90 mm		<90 mm
max p_T of D/τ		>4 GeV/c		>4 GeV/c
max(min IPχ^2 D/τ_1,min IPχ^2 D/τ_2)		>150		>150
min(min IPχ^2 D/τ_1,min IPχ^2 D/τ_2)		>16		>16
max(min (IPχ^2 $K/\pi_{1,2,3}$),min (IPχ^2 $K/\pi_{4,5,6}$))		>20		>20
max p_T of K/π		>2 GeV/c		>2 GeV/c
sum p_T of K/π		>7 GeV/c		>7 GeV/c
TOS on Hlt2Topo[2,3,4]Body		yes		N/A

also define more sophisticated variables related to a BDT output combining kinematic and geometrical quantities of the 'non-isolating' tracks.

- Vertex isolations: For their construction, the tracks making up the τ or B candidate's decay vertex $\mathcal{V}$ are combined with a single other track from the event, and fitted together into a new vertex $\mathcal{V}^*$. Several characteristics of this new vertex or of the additional track provides discriminating power.

- Neutral isolations: They corresponds to the neutral activity in a cone of size 0.5 in pseudo-rapidity η and polar angle ϕ around the B candidate.

- $B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$ reconstruction : A method for the full reconstruction of the $B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$ decay has been developed by Alessandro Mordà during his PhD [54]. It is based on geometrical information from measurable quantities of the decay, combined with mass constraints for the particles in the decay chain. With this method, the reconstruction of the two τ momenta is equivalent to finding the roots of a 4^{th} degree polynomial:

$$\mathcal{P}^{(4)}(\xi) = \sum_{i=0}^{4} a^{(i)}(\theta)\xi^i = 0 \,, \tag{3.3}$$

  where the explicit expressions for the coefficients $a^{(i)}(\theta)$ as well as the definition of the unknown parameters ξ are shown in Appendix D of Ref. [54], and θ is a rotation angle physically related to the asymmetry in the triangle formed by the PV and the two τ decay vertices. The solutions of the fourth degree equation are in the form

$$\xi_1 = a - ib \,, \quad \xi_2 = a + ib \,, \quad \xi_3 = c - id \,, \quad \xi_4 = c + id \,, \tag{3.4}$$

  with $a, b, c, d \in \mathcal{C}$.

  The angle θ is not known and has to be approximated. The approximated value of the θ parameter is used to compute the coefficients of the polynomial in Eq. (3.3) and to solve for the unknown ξ. Once the correct solution is chosen among the ξ_i, the two (approximate) τ momenta can be reconstructed, as well as the topology of the decay. The knowledge of the τ momenta allows to reconstruct their common origin vertex and the B candidate momentum. From a strictly conceptual point of view, the only new variables that can be computed to discriminate signal against background are the decay times of the two τ leptons and of the B candidate. All the kinematic variables such as the masses of the τ leptons or the B meson, as well the ones of the two ν have already been imposed as external constraint in the derivation of Eq. (3.3). Nevertheless, in the various steps of the reconstruction algorithm, several variables which are functions of observable quantities appear, and turn out to be useful in discriminating signal against background. These variables do not have a physical meaning, being much more related to the mathematics needed to derive the relevant equations.

The complete definition of the discriminating variables used further in this analysis can be found in Appendix A.1.

3.2.4 Offline selection

The first phase of the offline selection aims at removing the most obvious background, while still keeping a relatively high signal efficiency. Eleven variables, including eight isolation variables and three kinematical variables, are identified from a long list of discriminating variables. They are

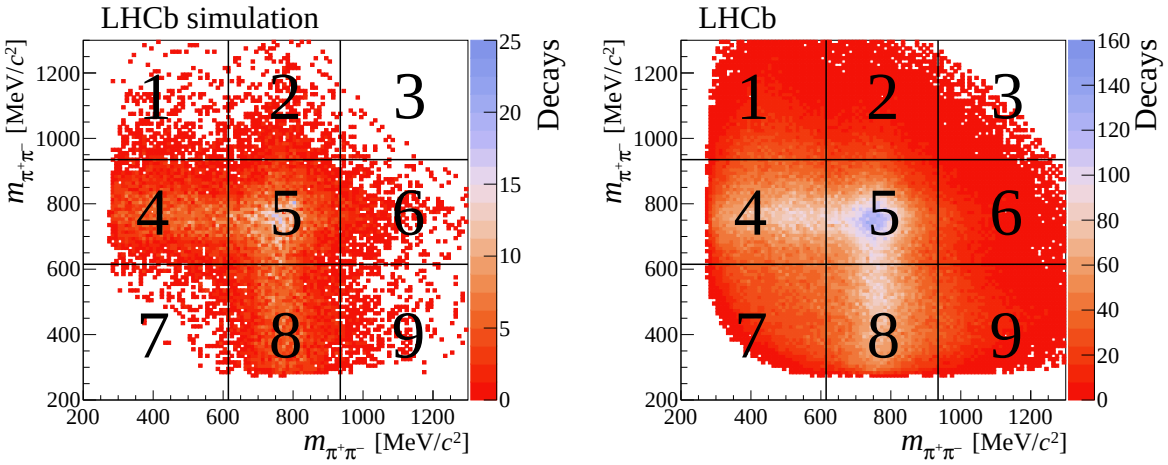

Figure 3.4: Dalitz plane of the τ^+ candidates in the signal MC [Left] and data [Right].

chosen for their capability to individually remove more than 50% of the candidates in data while still maintaining a signal efficiency of more than 95%. In total, the cut based selection reject 98.6% of the candidates in data while keeping 83% of stripped and triggered signal candidates. The distributions of the selected variables for simulated signal and data can be found in Appendix A.2.

The second phase of the offline selection consists of a multivariate analysis based on an artificial neural network (NN1). The chosen implementation is that of NeuroBayes [61]. The NN1 is trained on the signal MC sample and the background sample obtained from the data background region described in Section 3.1.4. The cuts on the invariant masses $m_{\pi_1^+\pi_2^-}$ and $m_{\pi_3^+\pi_2^-}$, used to define the different regions, are optimised on the signal MC so that the signal region includes 80% of the signal events. The boundaries are set to 615 and 935 MeV/c^2. The Dalitz distributions of candidates passing the stripping and trigger selection are shown in Fig. 3.4.
The choice of variables entering the NN is made from the ranking performed by the NeuroBayes algorithm. Seven variables as retained: the τ masses and lifetimes, two isolation variables and a variable from the full decay chain reconstruction (the detailed list is given Table A.2 of the Appendix). The selection cut is optimised to retain 80% of the signal MC, which corresponds to a background rejection of 91% for candidates in the background region of the data. The NN1 distribution for simulated signal and data from the background region is given in Fig 3.5 (left).

The total selection efficiency, including geometrical acceptance of the detector, trigger, stripping, offline selection and selection of the signal Dalitz region is $(24.09 \pm 0.70)^{-6}$. After the full selection, 14690 events remain in the data signal region.

3.3 Signal yield extraction

3.3.1 Neural network definition

The number of signal events is obtained from a binned maximum likelihood fit to a second neural network, named NN2. It uses 29 variables, listed in Table A.3 of the Appendix, including general event properties, τ properties, B properties, isolation variables and τ reconstruction variables. The signal events on which the NN2 is trained are used later on to model the signal template in the likelihood fit. To avoid overtraining, the k-folding technique (with $k = 10$) is therefore applied,

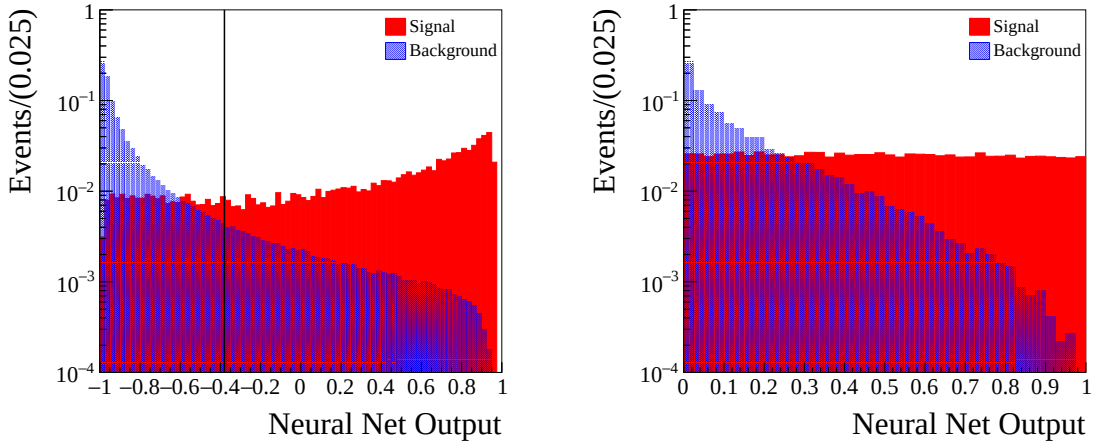

Figure 3.5: Flattened output distribution of the NN1 (Left) and NN2 (Right) for signal (Red) and background (Blue).

meaning that the NN2 is trained on 9/10 of the events and then applied to the remaining 1/10. The output of the neural network being a non-physical variable in the range $x \in [-1, 1]$, before combining the information of the 10 trained neural networks the individual NNs are transformed to uniformise their distributions. The final NN2 distribution is given in Fig 3.5 (right).

3.3.2 Fit strategy

The signal yield is obtained from a maximum likelihood fit to the NN2 output using ten equally sized bins in the range [0.0, 1.0]. It is not possible to perform a fully blind analysis but in order to minimize possible biases, the high NN2 region [0.7,1.0] was not looked at until the fit strategy was fixed. After the selection, the signal, background and control regions contain 16%, 13% and 58% of the simulated signal decays, respectively. The corresponding fractions in data are 6.9%, 37% and 47%. Even if most signal decays fall into the control region, the signal region is the most sensitive one thanks to its lower background contamination. In the fit, the signal model is taken from the $B_s^0 \to \tau^+\tau^-$ simulation, while the background model is taken from the data control region, subtracting the signal contribution.

The fit model is given by

$$\mathrm{NN}^{\mathrm{SR}}_{\mathrm{data}} = s \times \widehat{\mathrm{NN}}^{\mathrm{SR}}_{\mathrm{sim}} + f_b \times \left(\mathrm{NN}^{\mathrm{CR}}_{\mathrm{data}} - s \times \frac{\epsilon^{\mathrm{CR}}}{\epsilon^{\mathrm{SR}}} \times \widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{sim}} \right) , \tag{3.5}$$

where $\mathrm{NN}^{\mathrm{SR}}_{\mathrm{sim/data}}$ ($\mathrm{NN}^{\mathrm{CR}}_{\mathrm{sim/data}}$) is the NN2 distribution in the signal (control) region from simulation/data, s is the signal yield in the signal region, f_b is the scale factor for the background template, and ϵ^{SR} (ϵ^{CR}) is the signal efficiency in the signal (control) region. The corresponding normalised distributions, $\widehat{\mathrm{NN}}^{\mathrm{SR}}_{\mathrm{sim}}$, $\widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{sim}}$, $\widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{data}}$ are shown in Fig. 3.6. Distributions of the NN2 output in the signal and control regions are compared in Fig. 3.7 and found to be statistically compatible. Larger discrepancies are seen between the background and signal or control regions. As the background region is only used to optimize the selection and train the neural networks, this could lead to suboptimal performances of the analysis but can not bias the result. The composition of the background in the control and signal region have been examined in the exclusive and

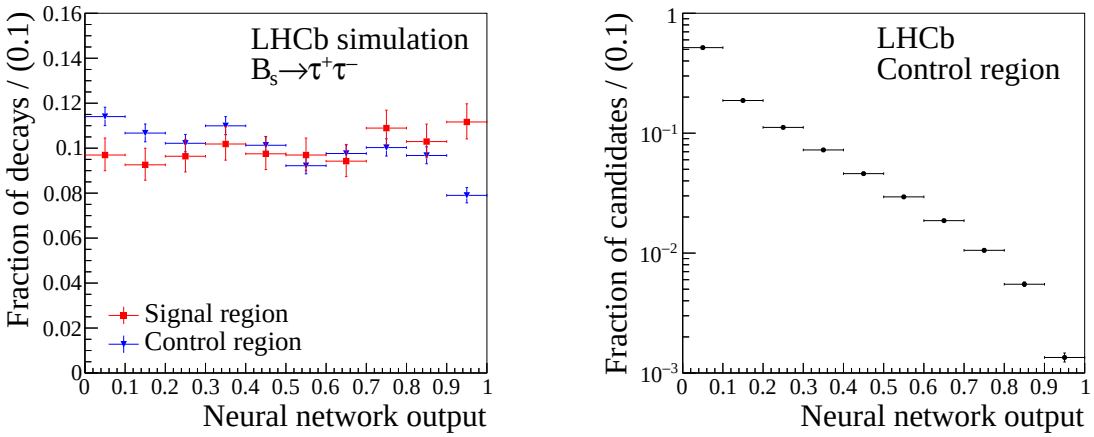

Figure 3.6: (Left) NN2 output distribution in the signal ($\widehat{\mathrm{NN}}^{\mathrm{SR}}_{\mathrm{sim}}$) and control ($\widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{sim}}$) region for $B^0_s \to \tau^+\tau^-$ simulated events. (Right) NN2 output distribution in the data control region $\widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{data}}$.

inclusive simulated samples mentionned in Sec. 3.1.3 and found to be in agreement. This cross check is however limited by the low statistics of these samples.

The chosen fit framework uses histogram templates to describe the signal and background component, and is set up using the `HistFactory` package [62] in `RooFit` [63]. The minimisation of the likelihood and the limit calculation, using the $\mathrm{CL_s}$ method [35, 64], is done by `RooStats` [65].

Differences between the shape of the background distribution in the signal and control regions of the data form the main systematic uncertainty on the background model. This uncertainty is taken into account allowing each bin of the $\widehat{\mathrm{NN}}^{\mathrm{CR}}_{\mathrm{data}}$ distribution to vary according to a Gaussian constraint. The width of this Gaussian function is determined splitting the control region into two approximately equally-populated samples (one closer to the a_1 resonance and one more distant) and taking, for each bin, the maximum difference between the NN2 output of the two subregions and the unsplit sample.

The signal can be mismodelled in the simulation. The $B^0_d \to D^- D^+_s$ decay is used to compare data and simulation for the variables entering the NN2. Ten variables are found to be slightly mismodelled and are weighted following an iterative procedure. The difference in the shape of the NN2 output distribution from the original, unweighted, sample is assigned as a systematic uncertainty. Dependence of the NN2 signal output distribution on the τ decay model has been studied and found to be negligible. The fit procedure is validated with pseudoexperiments and is found to be unbiased. Assuming no signal contribution, the expected statistical (systematic) uncertainty on the B^0_s signal yield is $^{+62}_{-40}$ ($^{+40}_{-42}$).

3.4 Normalisation

Following Eq. 2.2, the observed number of $B^0_s \to \tau^+\tau^-$ events, s, is related to the $B^0_s \to \tau^+\tau^-$ branching ratio through the normalisation factor

$$\mathcal{B}(B^0_s \to \tau^+\tau^-) = \alpha_s \times s\,, \tag{3.6}$$

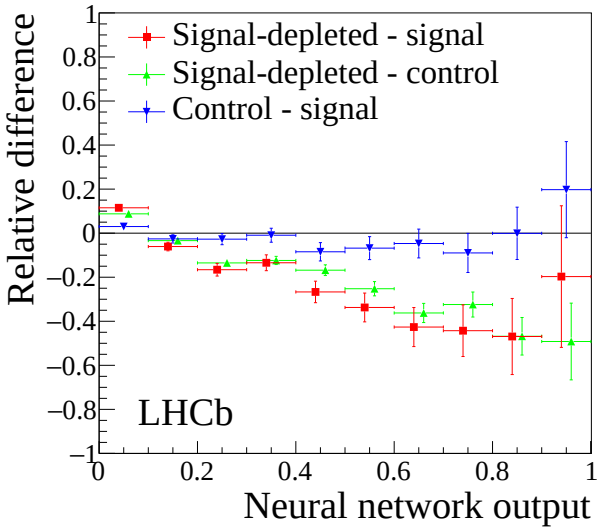

Figure 3.7: Relative difference of the NN2 output distributions in the three regions used in this analysis. The 'signal-depleted' label refers to the background region.

with

$$\alpha_s \equiv \frac{\epsilon^{D^-D_s^+} \cdot \mathcal{B}(B_d^0 \to D^- D_s^+) \cdot \mathcal{B}(D^+ \to \pi^+ K^- \pi^+) \cdot \mathcal{B}(D_s^+ \to K^+ K^- \pi^+)}{N^{\text{obs}}_{D^-D_s^+} \cdot \epsilon^{\tau^+\tau^-} \cdot [\mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\nu_\tau)]^2} \cdot \frac{f_d}{f_s}\,, \qquad (3.7)$$

where the total efficiencies $\epsilon_{D^-D_s^+}$ and $\epsilon_{\tau^+\tau^-}$ are determined from simulation, applying correction factors derived from data. The branching ratios are taken from the PDG [34], and $f_s/f_d = 0.259 \pm 0.015$ [47, 48] is the ratio of B_s^0 to B_d^0 production fractions.

The $B_d^0 \to D^- D_s^+$ yield, $N^{\text{obs}}_{D^-D_s^+}$, is obtained by fitting the mass distribution with the sum of four contributions: the $B_d^0 \to D^- D_s^+$ component modelled by an Hypatia function [66], a combinatorial background component described by an exponential function, and two partially reconstructed backgrounds, $B_d^0 \to D^{*-} D_s^+$ and $B_d^0 \to D^- D_s^{*+}$, modelled as in Ref. [67]. The resulting fit is shown in Fig. 3.8 and gives a yield of $N^{\text{obs}}_{D^-D_s^+} = 10629 \pm 114$.

Uncertainties on α_s arise from the $B_d^0 \to D^- D_s^+$ fit model, the finite size of the simulated samples, the uncertainty from the corrections to the simulation, and external inputs, which include the branching fractions and hadronisation fractions of Eq. (3.7). The latter contribution is the dominant with a relative uncertainty of 13% on α_s. The $B_d^0 \to D^- D_s^+$ fit model is varied using a double Gaussian function with power-law tails instead of the Hypatia function for the signal, a second-order Chebychev polynomial instead of an exponential function for the combinatorial background, and adding two other background components from $B_s^0 \to D^- D_s^{*+}$ and $B_d^0 \to a_1(1260)^- D_s^{*+}$ decays. The difference in signal yield compared to the nominal fit is taken as a systematic uncertainty, adding the contributions from the four variations in quadrature. The overall relative uncertainty on α_s due to $N^{\text{obs}}_{D^-D_s^+}$ (including the fit uncertainty) is 1.6%. Corrections determined from $J/\psi \to \mu^+\mu^-$ and $D^0 \to K^-\pi^+$ data control samples are applied for the tracking, particle identification and the hadronic hardware trigger efficiencies. The associated uncertainties correspond to the precision with which they are determined. The relative uncertainty on α_s due to selection efficiencies is 2.9%, taking into account both the limited size of the simulated samples and the systematic uncertainties.

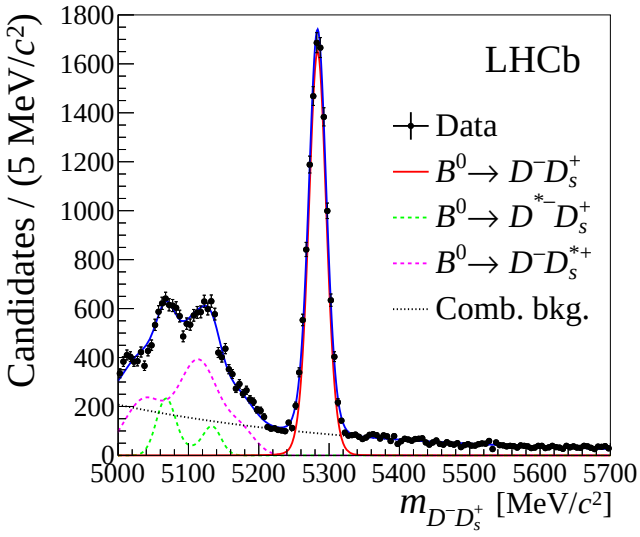

Figure 3.8: Invariant mass distribution of the reconstructed $B_d^0 \to D^- D_s^+$ events in data (black points), together with the total fit result (blue) used to extract the $B_d^0 \to D^- D_s^+$ yield. The individual components are described in the text.

The normalisation factor is found to be $\alpha_s = (4.07 \pm 0.70) \times 10^{-5}$.

3.5 Results

The fit to the data, shown in Fig. 3.9, gives a signal yield of $s = -23^{+63}_{-53}(stat)\ ^{+41}_{-40}(syst)$. The split between the statistical and systematic uncertainties is based on the ratio expected from pseudoexperiments. Assuming no contribution from the $B_d^0 \to \tau^+\tau^-$ decays, it is translated into an upper limit on the $B_s^0 \to \tau^+\tau^-$ decay branching fraction :

$$\mathcal{B}(B_s^0 \to \tau^+\tau^-) < 5.2(6.8) \times 10^{-3} \text{ at } 90(95\%)\text{CL}. \tag{3.8}$$

This is the first experimental limits on this decay.
Figure 3.10 (left) shows the fit result using only the background model. A likelihood ratio test is performed comparing the nominal fit with the background-only alternative. The p-value of the likelihood ratio test is 0.06, showing that the result is consistent with the background-only hypothesis. Figure 3.10 (right) shows the profile likelihood of the nominal fit. Figure 3.11 shows the expected and observed CL_s value as function of the branching fraction. The expected limit for the B_s^0 mode is $\mathcal{B}(B_s^0 \to \tau^+\tau^-) < 5.7\,(7.4) \times 10^{-3}$ at 90 (95)% CL.

The fit is performed replacing the signal model by that derived from simulated $B_d^0 \to \tau^+\tau^-$ decays, giving $s = -15^{+67}_{-56}(stat)\ ^{+44}_{-42}(syst)$. The fit result is shown in Fig. 3.12 (right). The corresponding normalisation factor is computed to be $\alpha_d = (1.16 \pm 0.19) \times 10^{-5}$. Figure 3.12 (left) gives the expected and observed CL_s value as function of the branching fraction. The obtained limit is $\mathcal{B}(B_d^0 \to \tau^+\tau^-) < 1.6\,(2.1) \times 10^{-3}$ at 90 (95)% CL, which constitutes a factor 2.6 improvement with respect to the BaBar result [21] and is the current best experimental limit.

Additional remarks

It is important to mention here two contributions that can have an effect on the analysis result. The first one concerns the contribution from $\tau^+ \to \pi^+\pi^-\pi^+\pi^0\bar{\nu}_\tau$ decays. Even if the π^0 is not

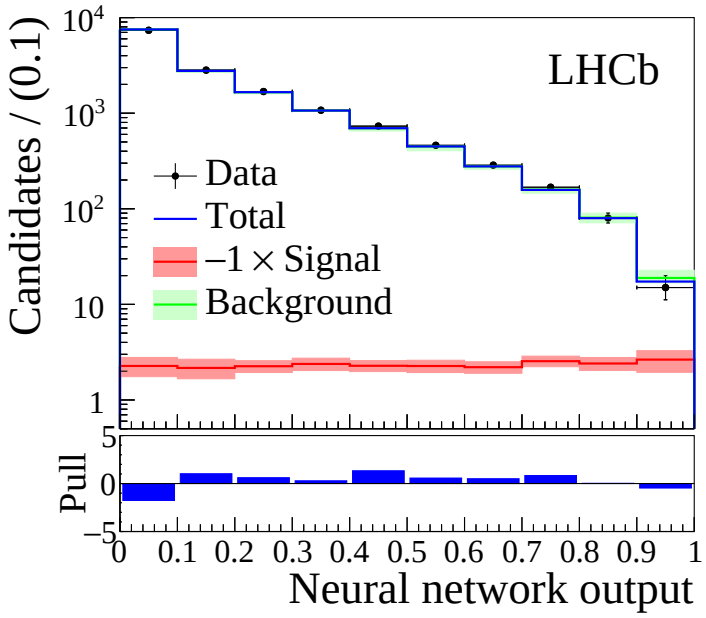

Figure 3.9: Distribution of the NN2 output in the signal region $\mathrm{NN}^{\mathrm{SR}}_{\mathrm{data}}$ (black points), overlaid with the total fit result assuming the presence of $B^0_s \to \tau^+\tau^-$ only (blue), and the background component (green). The signal component is also shown (red), multiplied by -1.

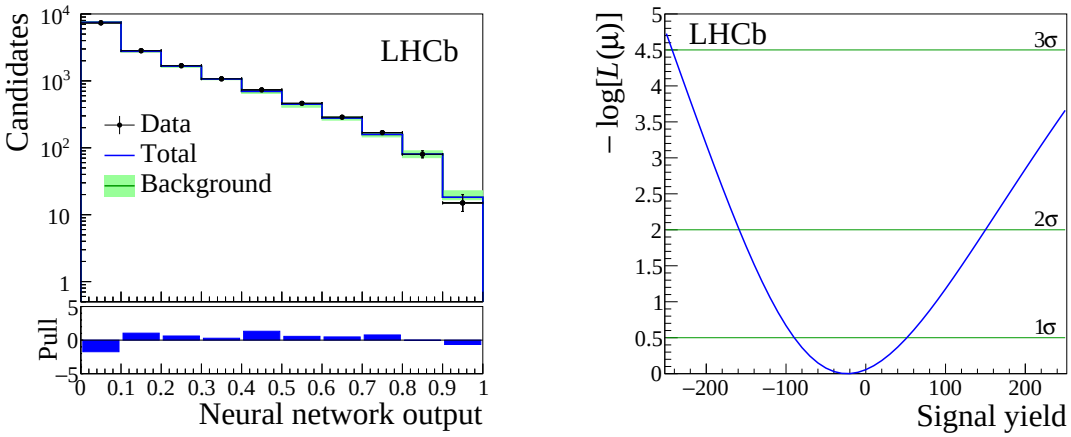

Figure 3.10: (Left) Distribution of the NN2 output in the signal region $\mathrm{NN}^{\mathrm{SR}}_{\mathrm{data}}$ (black points), overlaid with the total fit result (blue), and the background component (green). Shown is the fit using the "background only" model. (Right) Likelihood scan for the fit to the signal region for the $B^0_s \to \tau^+\tau^-$ decay.

explicitely reconstructed, this final state should be considered as signal. The efficiency of the $B^0_s \to \tau^+(\pi^+\pi^-\pi^+\bar{\nu}_\tau)\tau^-(\pi^-\pi^+\pi^-\pi^0\nu_\tau)$ decays is evaluated to be only 11.5% of the main signal one. In addition, its distribution in the NN2 output is more background like. At the end, the contribution from $\tau^+ \to \pi^+\pi^-\pi^+\pi^0\bar{\nu}_\tau$ has a negligible impact on the limit.

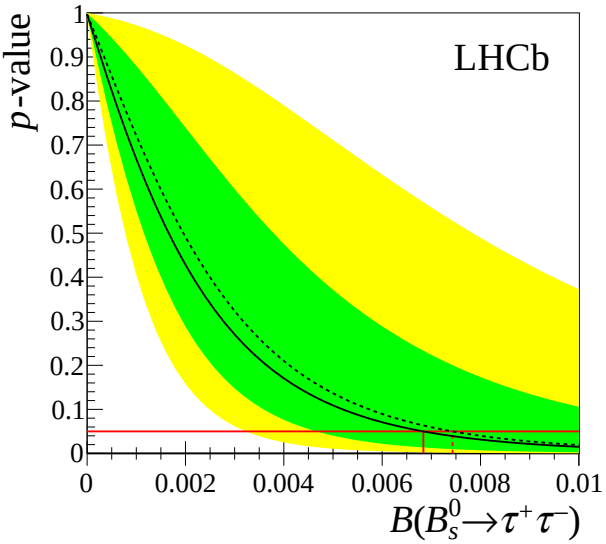

Figure 3.11: The p-value derived with the $\mathrm{CL_s}$ method as a function of $\mathcal{B}(B_s^0 \to \tau^+\tau^-)$. Expected (observed) values are shown by a dashed (plain) black line. The green (yellow) band covers the regions of 68% and 95% confidence for the expected limit. The red horizontal line corresponds to the limit at 95% CL.

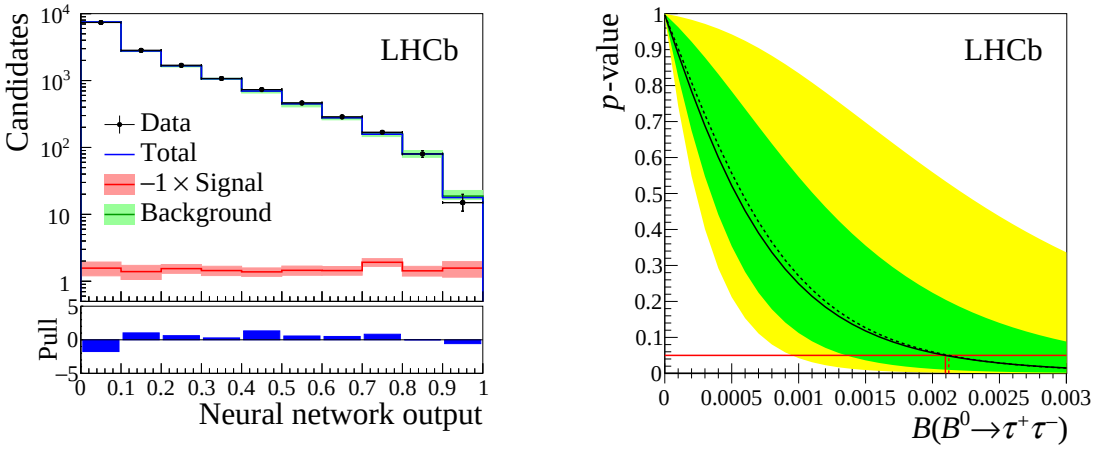

Figure 3.12: (Right) Distribution of the NN2 output in the signal region $\mathrm{NN}^{\mathrm{SR}}_{\mathrm{data}}$ (black points), overlaid with the total fit result assuming the presence of $B_d^0 \to \tau^+\tau^-$ only (blue), and the background component (green). The signal component is also shown (red), multiplied by -1. (Left)The p-value derived with the $\mathrm{CL_s}$ method as a function of $\mathcal{B}(B_d^0 \to \tau^+\tau^-)$. Expected (observed) values are shown by a dashed (plain) black line. The green (yellow) band covers the regions of 68% and 95% confidence for the expected limit. The red horizontal line corresponds to the limit at 95% CL.

The second contribution regards the modeling of the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ decay used in the simulation, which can affect both the efficency and the NN2 shape. The result obtained with the TAUOLA BaBar-tune is therefore compared to alternatives available within TAUOLA [68], which

Table 3.3: Efficiency (in %) at different stage of the selection. The numbers given here are not corrected from data/MC differences.

selection step	$(3\pi,3\pi)$	$(3\pi,\mu)$
Generator level cuts	3.12 ± 0.01	5.83 ± 0.01
Reco + Stripping	1.5 ± 0.01	0.71 ± 0.01
L0 trigger	49.95 ± 0.29	92.41 ± 0.19
HLT	77.83 ± 0.034	91.57 ± 0.21
cut-based + MVA	63.63 ± 0.43	11.78 ± 0.26
Signal region	15.82 ± 0.42	34.41 ± 1.10
TOTAL	$(1.79 \pm 0.05)10^{-3}$	$(1.43 \pm 0.06)10^{-3}$

are based on CLEO data for the $\tau^- \to \pi^-\pi^0\pi^0\nu_\tau$ decay [69]. The selection efficiency for this alternative models can be up to 20% higher, due to different structures in the two-pion invariant mass, resulting in lower limits. Dependence of the NN2 signal output distribution on the τ-decay model is found to be negligible. Given that the alternative models are based on a different τ decay, the BaBar-tune model is chosen as default and no systematic uncertainty is assigned.

3.6 Prospects

3.6.1 Use of $(3\pi, \mu)$ final state

As seen in Sec. 3.1, the muonic decay of the τ lepton can be of interest at LHCb. The study of the $B_s^0 \to \tau^+\tau^-$ decay reconstructed in the $(3\pi, \mu)$ final state was started by several internship students under my supervision and has been pursued by Cédric Méaux during his PhD. The strategy developed is similar to the one of the $(3\pi,3\pi)$ final state. The definition of the different regions is adapted as only one Dalitz plane is available in that case: the signal region corresponds to box 5 of Fig. 3.3, the background region to 1, 3, 7 or 9 and the control region to 4 or 8.

The stripping selection is harder than for the $(3\pi,3\pi)$ final state in order to cope with the retention rate limit, having an efficiency of 0.7%. Detail of the stripping selection can be found in Appendix A.3. As expected, the efficiency of L0 (HLT) requirements on stripped signal candidates is high, around 92 (92)%. The offline selection comprises a cut-based selection followed by a requirement on a boosted decision tree output, based on 20 variables. In order to improve the background rejection, two other BDTs are trained on simulation to remove specifically backgrounds from hadronic B decay (e.g $B_d^0 \to D^-(\to K^0\mu^-\nu)\pi^+\pi^-\pi^+$) and semileptonic B decays (e.g. $B_d^0 \to D^-(\to \pi^+\pi^-\pi^-\pi^0)\mu^+\nu$). This latter is particularly problematic because of its very high branching fraction, at the % level, and the fact that the τ lepton and D meson have similar lifetime (0.29 versus 0.41 ps). Even if rather good performances are obtained on simulation, the performances of these BDTs turn out to be disappointing when applied to data and only a mild cut is kept on the BDT trained against hadronic B decays. After the full selection, 13715 candidates remain in the signal region. The total signal efficiency is 1.4×10^{-5}. A comparison of the efficiencies at different stages of the selection process for the $(3\pi,3\pi)$ and $(3\pi,\mu)$ final state can be seen in Table 3.3.

A final BDT is optimized and used as a fit variable. Its distribution for the simulated signal, data control region and data signal region can be see in Fig. 3.13. The expected signal yield uncertainty is computed from pseudoexperiments and found to be 444, without including the systematic uncertainties. The corresponding expected limit, computed using the $B_d^0 \to D^-(\to \pi^+\pi^-\pi^+)\pi^+$

as a normalisation channel, is found to be $\mathcal{B}(B_s^0 \to \tau^+\tau^-) < 3.1 \times 10^{-2}$ at 95%CL This is four times higher than the expected limit of the $(3\pi, 3\pi)$ final state.

Several other analysis strategies were tried but did not lead to better results. This bad outcome comes mainly from the too poor discriminating power of the final BDT. The muon provides very few discriminating variables (only its p_T is retained in the analysis optimization process), and the semileptonic B decay background has a topology too similar to the signal one. We therefore decided to abandon this analysis to focus on the update of the $(3\pi, 3\pi)$ final state one with the Run2 data. This study has been documented in an internal note [55].

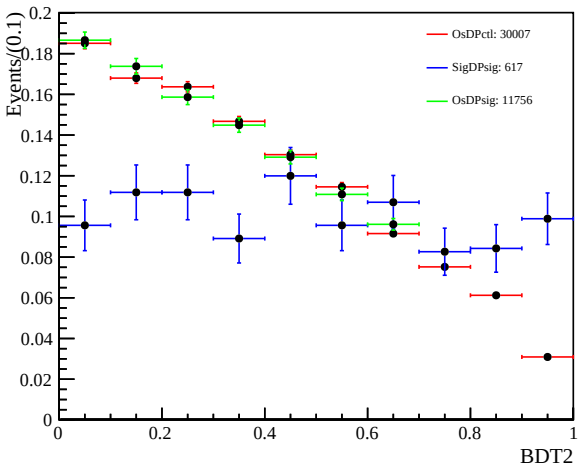

Figure 3.13: Distribution of the BDT2 output for data in control region (red), in the signal region (green) and MC signal in the signal region (blue). The region corresponding to BDT values higher than 0.7 is blinded.

3.6.2 Prospects for the $(3\pi, 3\pi)$ final state

The Run 2 analysis will benefit from the almost twice higher $b\bar{b}$ cross section coming from the centre-of-mass energy increase. As discussed in Sect. 2.7, detector upgrades are planned during LS2 and LS4. They could help improving the $B_s^0 \to \tau^+\tau^-$ search thanks to higher trigger and reconstruction efficiency. On the other hand, especially for the HL-LHC phase, the huge multiplicity (2500 charged tracks at the maximum luminosity) can necessitate a tighter selection. The red curve in Fig. 3.14 shows the expected limit as function of time only taking into account the luminosity increase. A value of 4.5×10^{-4} could be reached by 2036. The blue curve in Fig. 3.14 assumes a (certainly optimistic) factor 4 improvement to the reconstruction, trigger and analysis efficiency.

The Belle II experiment could also perform the search for the $B_d^0 \to \tau^+\tau^-$ decays. Extrapolating the BaBar analysis, the Belle II physics book states an expected limit of 9.6×10^{-5} for the full dataset of $50\,\mathrm{fb}^{-1}$ collected by 2027 [70]. To perform this analysis in BaBar, the opposite B meson is reconstructed in a hadronic mode. The efficiency could certainly be improved using the 'full event interpretation' developed by the Belle collaboration [71]. The statistics collected at the $\Upsilon(5s)$ resonance might be too low to have a competitive result on the $B_s^0 \to \tau^+\tau^-$ branching fraction.

In a longer time scale, the future circular collider (FCC) would be an ideal machine to search for this kind of decays. Indeed, in its electron-positron configuration, it would produce 5×10^{12} Z boson decay resulting in 250 billions of B_s^0 mesons produced in a clean environment [72]. There is no

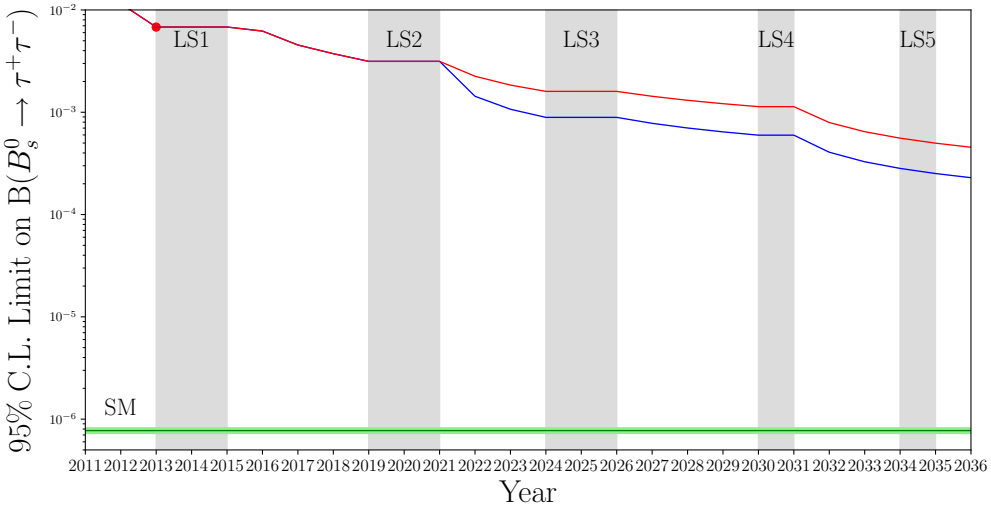

Figure 3.14: 95% C.L expected limit on the $B_s^0 \to \tau^+\tau^-$ branching fraction as function time. The red curve takes only into account the increase in recorded luminosity, while the blue curve assume an additional factor 4 improvement in efficiency. The green band represents the SM prediction.

estimate of the possible limit that could be reached for the $B_s^0 \to \tau^+\tau^-$ decays. However the $B \to K^*\tau\tau$ decay have been studied more deeply and about one thousand events could be observed, assuming the SM prediction of the branching fraction, $\mathcal{B}(B \to K^*\tau\tau) = (0.98 \pm 0.10) \times 10^{-7}$ [73].

3.6.3 Study of other decays at LHCb

In parallel to the $B_s^0 \to \tau^+\tau^-$ analysis, other rare B decays into τ leptons have been and are still under study at LHCb. Even if I have been much less involved in these ones, they are mentionned here for completness.

$B_{d,s} \to \tau\mu$

This is a lepton flavour violating decay, which is highly suppressed in the SM. Indeed it can only proceed through a one loop-diagram involving neutrino oscillation, making its branching fraction prediction to be of the order of $\mathcal{O}(10^{-50})$. In presence of new physics this value could enhanced up to $\sim 10^{-5}$, in particular in models with Z' or leptoquarks [74, 75]. When reconstructing the τ into three pions, the $B_{d,s} \to \tau\mu$ decay has the same 'visible' final state as the $B_{d,s} \to \tau^+(\pi^+\pi^-\pi^+\nu_\tau)\tau^-(\mu^-\nu_\mu\bar{\nu}_\tau)$ and can thus use from the same stripping line. Knowing that the B_s^0 decay vertex must lies on the trajectory of the muon and imposing a mass contraint on the τ lepton, the invariant B_s^0 mass can be reconstructed analytically with a 2-fold ambiguity, and used as the main discriminating variable. The limit obtained on the branching fractions are $\mathcal{B}(B_s \to \tau\mu) < 4.2 \times 10^{-5}$ and $\mathcal{B}(B_d \to \tau\mu) < 1.4 \times 10^{-5}$ at 95%CL (assuming no contribution from the B_d^0 or B_s^0 decay) [76]. This represents the first limit on the B_s^0 decay and the world best limit on the B_d^0. The analysis, performed by the CPPM group, was the subject of Joan Arnau's PhD defended in september 2018.

$B_d^0 \to K^*\tau\tau$

The $B_d^0 \to K^*\tau\tau$ decay is a FCNC whose branching fraction in the SM is expected to be $(0.98 \pm 0.10) \times 10^{-7}$ [73]. New physics models predict an enhancement of up to 3 orders of magnitude [73, 77, 78]. The efficiency for this mode is expected to be quite low due to the high number of tracks in the final state. However it has the advantage that the K^* decay allows to measure the position of the B vertex. I participated to the definition of two stripping lines, for the $(3\pi, 3\pi)$ and $(3\pi, \mu)$ final states[4], and worked with Andrey Tayduganov during its post doctorate at CPPM on the definition of the analysis strategy. This study has now been taken over by a new PhD student, Jacopo Cerasoli.

$B_{d,s} \to K^*\tau\mu$

This decays has a similar physics case as the $B_s^0 \to \tau\mu$ one and has the same visible final state as the $B \to K^*\tau(3\pi)\tau(\mu)$ decay. The analysis thus uses the same stripping line as $B \to K^*\tau(3\pi)\tau(\mu)$. It is lead by the LPNHE group in Paris. The results are not public yet but a limit of the order of 10^{-5} is expected.

Personal contributions
I worked on almost all analysis steps, except for the fit and limit computation that were done entirely by Kristof De Bruyn. In particular, I participated for both $(3\pi, 3\pi)$ and $(3\pi, mu)$ final states, to the stripping definition, writting of the decay files for the production of simulated signal and background samples, definition of the analysis strategy and selection, and background caracterisation.

[4] I also performed a feasibility study of the (μ, μ) final state, to have some fun.

Chapter 4

Conclusion

An interpretation of the measurements presented in this document is given both in a model-independent and model-dependent scheme in Sec 4.1. Section 4.2 gives a brief description of my personal projects for future.

4.1 Interpretation of the measurements

4.1.1 Muonic decay

The measurement of the $B^0_{d,s} \to \mu^+\mu^-$ branching fraction can first be interpreted in a model independent way using Eqs. 1.14 and 1.15. Neglecting possible NP effects in $C_{10}^{(\prime)}$ and in the context of SMEFT, the measurement defines a circle in the plane $(C_S - C'_S,\ C_S + C'_S)$, as shown in Fig. 4.1. While the experimental values are compatible with an absence of NP contributions, future measurements could reveal a deviation from the SM. It should be stressed that an increased experimental precision will reduce the width of the ring but only the measurements of additional observables such as $\mathcal{A}_{\Delta\Gamma}$ will allow to break the degeneracy.

As seen in Sec. 1.1.2, in the context of SUSY, the $B^0_{d,s} \to \mu^+\mu^-$ branching fraction scales as $tan^6\beta$. The agreement of the experimental values with the SM predictions therefore drastically reduces the parameter space of SUSY models. Given the amount of free parameters of these models, it is interesting to make some assumptions to demonstrate the impact of flavour observables. This is what is done in Ref. [79], where the authors assume a minimal supersymmetric extension of the

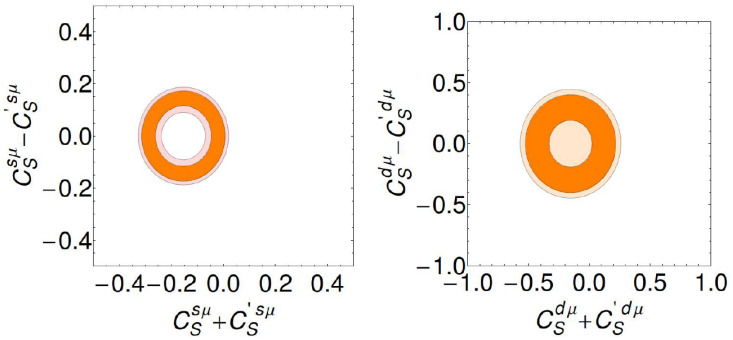

Figure 4.1: Constraints at 1 and 3σ obtained from the measurement of the $B^0_s \to \mu^+\mu^-$ (left) and $B^0_d \to \mu^+\mu^-$ (right) branching fractions. Figure from [15].

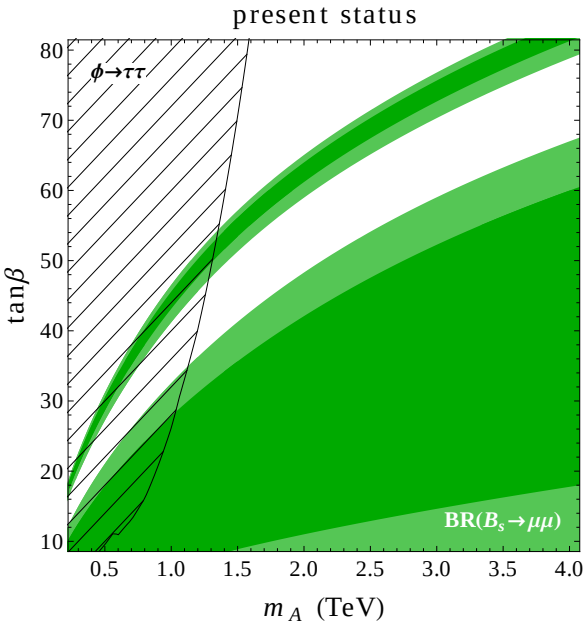

Figure 4.2: Constraints from $B_s^0 \to \mu^+\mu^-$ decays in the plane $tan\beta$ vs m_A. The green bands correspond to the regions allowed by $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the 1 and 2σ level. The hatched area is excluded by direct searches for $\tau^+\tau^-$ resonances. Figure from [79].

SM with minimal flavour violation and no new sources of CP violation. In that case, NP contribution are induced by the exchange of a heavy scalar and pseudoscalar Higgs boson, H and A, whose masses are approximately degenerated in the decoupling limits. Assuming all SUSY masses to be degenerated and set to 5 TeV, the sensitivity of the current measurements in the parameter space $tan\beta$ vs m_A is shown in Fig. 4.2[1]. For masses below $\sim$1.1 TeV the constraints from the direct searches are stronger than the indirect ones while for higher masses the $B_s^0 \to \mu^+\mu^-$ branching fraction measurement covers unconstrained parameter space, showing the complementarity of the two approaches.

The $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement can also be interpreted in the context of leptoquark models. Leptoquarks (LQ) are scalar or vectors states that can couple to a quark-lepton current. Among the ten possible representations respecting the SM gauge invariance [80], two, named U_1 and V_2, can generate a scalar operator contributing at tree level to the $B_s^0 \to \mu^+\mu^-$ decay. The contraints on the coupling versus mass for these two scenarios are shown in Fig. 4.3 [79]. One can see that the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement can access much higher masses than the direct searches.

4.1.2 Tauonic decay

The result presented in this document on the $B_s^0 \to \tau^+\tau^-$ decay is four orders of magnitude higher than the SM prediction and provides weak constraints on NP. However it can be seen as a first point on a picture similar to Fig. 1.4 and demonstrates that the study of rare B decays into τ leptons are possible in a hadronic environment.

It is interesting to see what could be the expected branching fraction in the context of the anomalies observed in tests of lepton flavour universality (LFU), both in loop-level and tree-level

[1] the experimental measurements used here is the combination of the CMS Run1 analysis and the latest LHCb analysis [16].

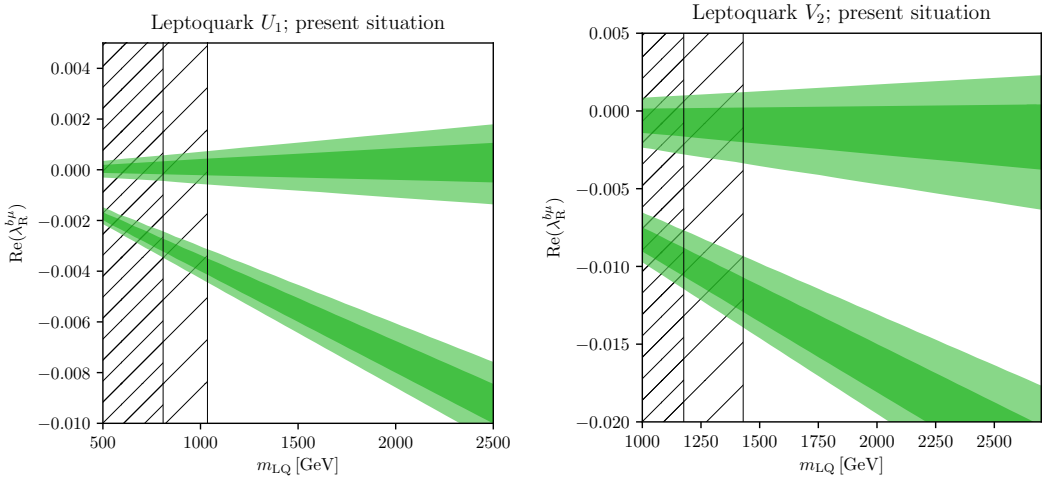

Figure 4.3: Current constraints from $B_s^0 \to \mu^+\mu^-$ decays in the plane defined by the mass and coupling for the LQs U_1 (left) and V_2 (right). The green bands correspond to the regions allowed by $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the 1 and 2σ level. The black hatched regions show the exclusion from direct searches. Figure from [79].

transitions. A detailed description of these anomalies is beyond the scope of this document, but we recall the reader that in $b \to s\ell^+\ell^-$ decays, deviations are seen in measurements of differential branching ratios and angular observables of muonic decays [81, 82, 83, 84] and in tests of lepton flavour universality when comparing electronic and muonic final state of $B \to K^{(*)}\ell^+\ell^-$ [85, 86, 87, 88]. Hints of LFU violation are also seen by several experiments in the tree-level processes $b \to c\ell\nu$ when comparing tau and muon or electron final states [89, 90, 91, 92, 93, 94, 95]. The latest HFLAV combination of the $R(D^{(*)})$ measurements corresponds to a 3.1σ deviation from the SM prediction.

A model independent discussion of $b \to s\tau\tau$ processes can be found in [73]. NP is assumed to modify only the $C_9^{(\prime)}$ and $C_{10}^{(\prime)}$ Wilson coefficient as favoured by the anomalies, modifying the $B_s^0 \to \tau^+\tau^-$ SM branching fraction as:

$$\mathcal{B}(B_s^0 \to \tau^+\tau^-) = (\Delta/C_{10}^{SM})^2 \times \mathcal{B}^{SM}(B_s^0 \to \tau^+\tau^-), \tag{4.1}$$

with $\Delta = \frac{2\pi}{\alpha}\frac{V_{cb}}{V_{tb}V_{ts}^*}\sqrt{(R_X/R_X^{SM}) - 1}$. The quantity R_X corresponds to the R_D, R_{D^*} and $R_{J/\psi}$ observables, defined as

$$R_X = \frac{B \to X\tau\nu}{B \to Xe/\mu\nu}. \tag{4.2}$$

The predicted branching fraction values as function of R_X/R_X^{SM} are shown in Fig. 4.4. The vertical green bands corresponds to the current experimental range obtained from the weighted average of R_D, R_{D^*} and $R_{J/\psi}$ measurements. In particular the $B_s^0 \to \tau^+\tau^-$ branching fraction can be enhanced up to $\sim 10^{-3}$ staying compatible at 2σ with the experimental results.

In [96], the flavour anomalies are analysed in the context of a simplified model with a vector leptoquark U_1 that can couple to both left and right-handed SM fields. A fit of five parameters of the model (four couplings and the scale) is performed to low-energy observables. The best fit region is consistent with the explanation of charged and neutral current anomalies, as seen in Fig. 4.5 (left). Interestingly, it implies large NP effects in LFV decays and $b \to s\tau\tau$ transitions. Figure 4.5 (right) shows the fit prefered region in the plane $\mathcal{B}(B_s^0 \to \tau^+\tau^-)$ versus ΔR_D, where

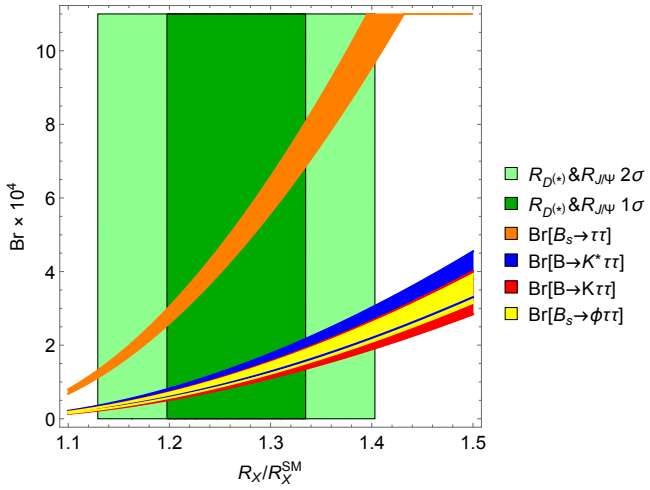

Figure 4.4: Prediction of the branching fractions of the $b \to s\tau\tau$ processes as function of R_X [73].

$\Delta R_D \equiv R_D/R_D^{SM} - 1$. In that case also, the branching fraction of $B_s^0 \to \tau^+\tau^-$ could be enhanced up to few $\times 10^{-3}$. This values could be reached with Run3 or Run4 data.

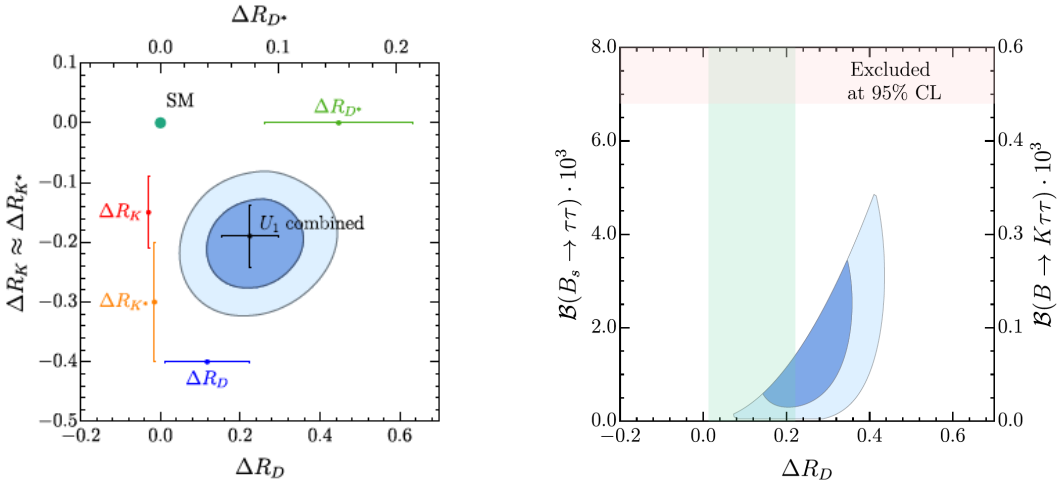

Figure 4.5: Prefered fit region at 1 (light blue) and 2 (dark blue) σ. The red band shows the excluded region at 95%C.L. The green band represents the experimental measurement at 1σ.

4.2 Personal project

Since I started the study of $B_{s/d} \to \ell^+\ell^-$ decays ten years ago, the landscape of heavy flavour physics became even more interesting thanks to the apparition of the anomalies previously discussed. Moreover, given that no unexpected signal emerged from the direct searches, the hope to find NP in a near future relies on the intensity frontier experiments. My personal interest is therefore to stay in this field. More precisely, my project is driven by the fact that all models proposed to explain the flavour anomalies imply

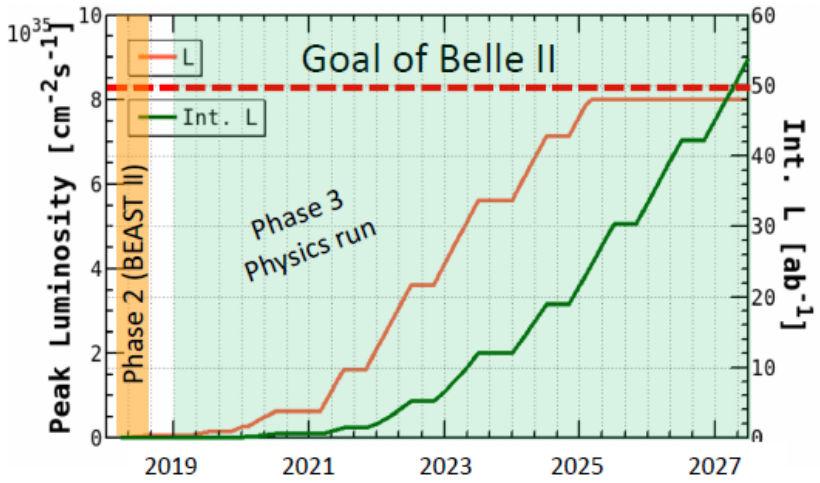

Figure 4.6: Projected instantaneous (green curve) and integrated (yellow curve) luminosities at SuperKEKB through 2027 assuming nine months operation per year.

- violation of the charged lepton flavour involving the third generation (e.g $\tau \to \mu\mu\mu$ or $B \to \tau\mu$) [97, 98, 77, 99, 100, 101, 96]

  and/or

- large enhancement of $b \to s\tau^+\tau^-$ processes [73, 77, 102, 78, 103, 96].

It thus looks critical to improve the experimental constraints on decays with τ leptons as soon as possible. I wrote a proposal for an ERC grant in that sense, that has been accepted in 2018. While LHCb is the only experiment able to study rare B_s^0 decays, Belle II, which started data taking in march 2019, is an ideal place to look for τ LFV decay and B_d^0 decays into τ leptons. Its expected peak and integrated luminosity as function of time is shown in Fig. 4.6 [70]. A first ab^{-1} is expected to be recorded by mid-2021, and the final statistics of $50\ ab^{-1}$ by 2027, multiplying by a factor 50 the Belle dataset. The grant will start in october 2019 and last for five years, allowing to build a team of analysts that will improve the current limits on these decays.

My activities in the longer term will depend on the outcome of the LHC and Belle II experiments. Discussions just started about an upgrade of the Belle II detector, with the goal of collecting $250\,\mathrm{fb}^{-1}$. Interesting results could also appear in low energy experiments such as NA62 and KOTO for the kaon sector, or in lepton flavour violating decays of the muon from the MEG2, Mu2e or COMET experiments.

Appendix A

$B^0_{d,s} \to \tau^+\tau^-$ selection

A.1 Variable definitions

The complete definition of the isolation and full reconstruction variables are the following:

- BDT based track isolation variables

  For the training of this BDT, the non-signal tracks in the event are divided into two categories: *non-isolating* tracks, i.e. they make the signal track non-isolated, versus *isolating* tracks, i.e. they leave the signal track isolated. The non-isolating tracks are all tracks coming from displaced B and D decay vertices that are part of the same true decay chain as the signal track. The isolating tracks are all remaining tracks, primarily coming directly from the primary vertex (PV), which are essentially unrelated to the signal track.

  For each non-signal track the common vertex $\mathcal{V}$ with the signal track is defined as the midpoint between the two tracks at their point of closest approach.[1]. The BDT combines the following seven discriminating variables:

  - The minimum IP χ^2 of the track with respect to any PV.
  - The transverse momentum of the track
  - The angle between the track and the signal track
  - The parameter

    $$f_c = \frac{|P_S + P_{tr}|\alpha}{|P_S + P_{tr}|\alpha + P_{TS} + P_{Ttr}}\,, \tag{A.1}$$

    where P and P_T are the momentum and transverse momentum of the tracks, S identifies the signal track, tr identifies the considered track, α is the angle between the momentum of the combined track + signal track system and the vector between the PV and vertex $\mathcal{V}$.
  - The distance of closest approach of the track and signal track.
  - The distance between the vertex $\mathcal{V}$ and the B decay vertex.
  - The distance between the vertex $\mathcal{V}$ and the PV.

  The BDT is now trained using the non-isolating tracks as signal target, and the isolating tracks as background target. For the analysis, isolating tracks (low BDT values) are preferred, whereas non-isolating tracks (high BDT values) are harmful.

[1] An actual refit of this vertex for every track in the event is too CPU intense, and not practical.

For each long track in the event which is not part of the signal candidate, the BDT response is calculated. Let a, b and c be the number of long tracks with a BDT value smaller than -0.09, -0.05 and 0, respectively, the the track isolation variables are defined as:

BDT Iso 1: The sum $a + 100 \times b + 1000 \times c$.
The values of a, b and c are recovered for the selection as the 1st digit (a), 3rd digit (b) and 5th or 10k digit (c) of BDT Iso 1.

BDT Iso 2: The sum $\sum_{\mathrm{BDT}(x)<-0.05} \mathrm{BDT}(x)$ of BDT values for all tracks with a BDT output smaller than -0.05.

BDT Iso 3: The sum of BDT Iso 2 and the minimal BDT value of all events with a BDT output in the range $[-0.05, 0]$.

- Vertex isolation variables:

NumVtxWithinChi2WindowOneTrack: The number of other tracks in the event for which the χ^2 of the new vertex fit is less than nine, $\chi^2_{\mathcal{V}^*} < 9$.

SmallestDeltaChi2MassOneTrack: The invariant mass of the tracks making up the τ or B candidate and the track leading to the smallest difference in χ^2

$$\Delta\chi^2 \equiv \chi^2_{\mathcal{V}} - \chi^2_{\mathcal{V}^*} \tag{A.2}$$

between the $\mathcal{V}$ and $\mathcal{V}^*$ vertex fits.

SmallestDeltaChi2MassTwoTracks: The invariant mass of the tracks making up the τ or B candidate and the two tracks leading to the smallest difference in χ^2 between the $\mathcal{V}$ and $\mathcal{V}^{**}$ vertex fits, where $\mathcal{V}^{**}$ is constructed iteratively from the $\mathcal{V}^*$ vertex that itself has the smallest $\Delta\chi^2$.

- Neutral isolation variables:

Mult: The number of neutral objects found inside the cone.

vPT: The vector-summed transverse momentum of the neutral objects inside the cone.

PZ: The z component of the total momentum of the neutral objects inside the cone.

- $B^0_s \to \tau^+(3\pi)\tau^-(3\pi)$ reconstruction:

Among the reconstruction variables, the most discriminating ones, which are further consider in this analysis, are [54]:

  - $\Re[x_3]$: $Re(|\vec{\tilde{p}}_+(\theta = \bar{\theta}, \xi_3)|)$
  - $\Im[\tilde{s}^{\pm}_1]$: $Im(|\vec{\tilde{p}}_+(\theta = \theta^\star, \xi_1)|)$
  - $\Im[x_1]$: $Im(\tilde{p}_+(\theta = \theta^\star, \xi_1)\tilde{p}_-(\theta = \theta^\star, \xi_1))$
  - $\Im[\tilde{s}^{\pm}_3]$: $Im(\tilde{p}_+(\theta = \theta^\star, \xi_3)\tilde{p}_-(\theta = \theta^\star, \xi_3))$
  - $\Re[\tilde{s}^{\pm}_1]$: $Re(\tilde{p}_+(\theta = \pi/4, \xi_1)\tilde{p}_-(\theta = \pi/4, \xi_1))$
  - $\Re[\xi_1]$
  - $\bar{\theta}$
  - $p_+ \cdot p_-$

– B_s^0 corrected mass: combination of the 6-pion invariant mass and the transverse momentum estimate, using the decay plane of the $B_s^0 \to \tau^+\tau^-$ system, for the two missing neutrinos.

where $\bar{\theta}$ and $\theta^\star$ are different approximations of the angle θ, $s \equiv p_+ \cdot p_-$ (for signal events $s = \frac{M_B^2 - 2m_\tau^2}{2}$) and $\tilde{p}_\pm^2 \equiv M_\pm^2 = m_\tau^2 \pm \sin(2\theta)s$, $p_\pm$ being the four momenta of the $\tau^\pm$.

A.2 Selection of the $(3\pi, 3\pi)$ final state

The variables used in the cut-based selection are given in Table A.1. Distributions of simulated

Table A.1: Overview of the cuts used in the loose selection. The cuts are optimised to retains as close to, but more than 98% of the signal candidates in simulation. Other highly discriminating variables exist, but are fully correlated with those used here.

ID	Description
0	$802 \leq m_\tau \leq 1598\,\mathrm{MeV}/c^2$
1	$m_B - m_{\tau^+} - m_{\tau^-} \geq 756\,\mathrm{MeV}/c^2$
2	$\tau_\tau \geq -0.12$ps
3	B SmallestDeltaChi2MassOneTrack ≥ 3499
4	τ SmallestDeltaChi2MassTwoTracks ≥ 1791
5	Number of Candidates ≤ 6
6	$\sum$ Iso$(\pi) \leq 3$
7	$\sum$ BDT Iso$_1(\pi)_{\text{10k digit}} \leq 16$
8	$\sum$ BDT Iso$_1(\pi)_{\text{1 digit}} \leq 5$
9	$\sum$ BDT Iso$_3(\pi) \geq -0.96$
10	B Neutral Isolation Cone $vPT \leq 3850$

signal and data for the variables entering the cut-based selection are given in Fig. A.1 and A.2.

Tables A.2 and A.3 list the variables entering the NN1 nd NN2.

Table A.2: List of input variables for the multivariate selection.

Rank		Variable
1		τ reconstruction variable $\Re[x_3]$
2	τ^-	Mass
3	τ^+	Mass
4	B_s^0	Neutral isolation cone variable vPT
5		$\sum_\pi$ third isolation BDT output
6	τ^-	Lifetime
7	τ^+	Lifetime

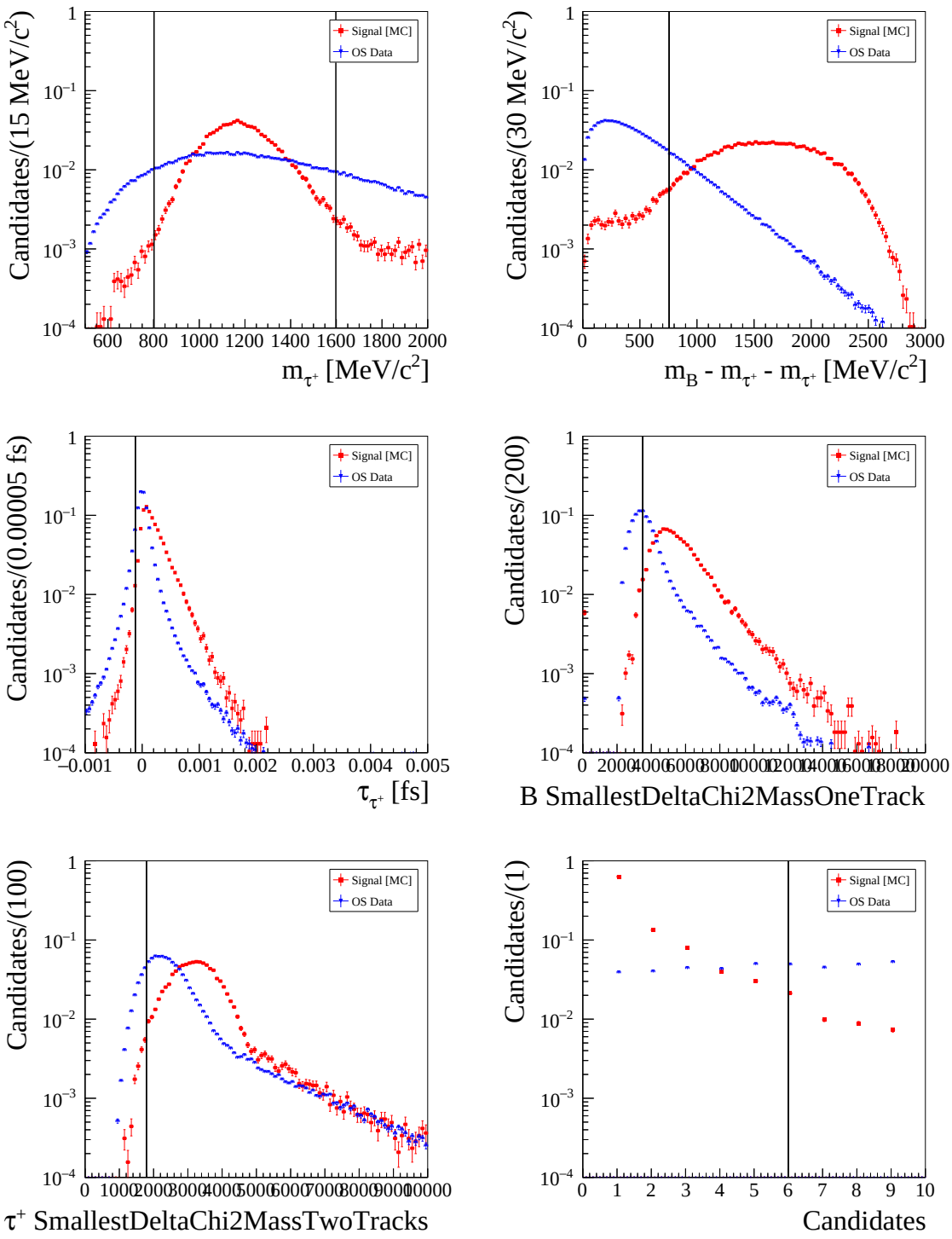

Figure A.1: Comparison between the observable distributions of the $B^0_{d/s} \to \tau^+\tau^-$ signal MC (Red) and the OS data (Blue), for the observables used in the cut-based loose selection. The vertical black lines indicate the position of the selection cuts.

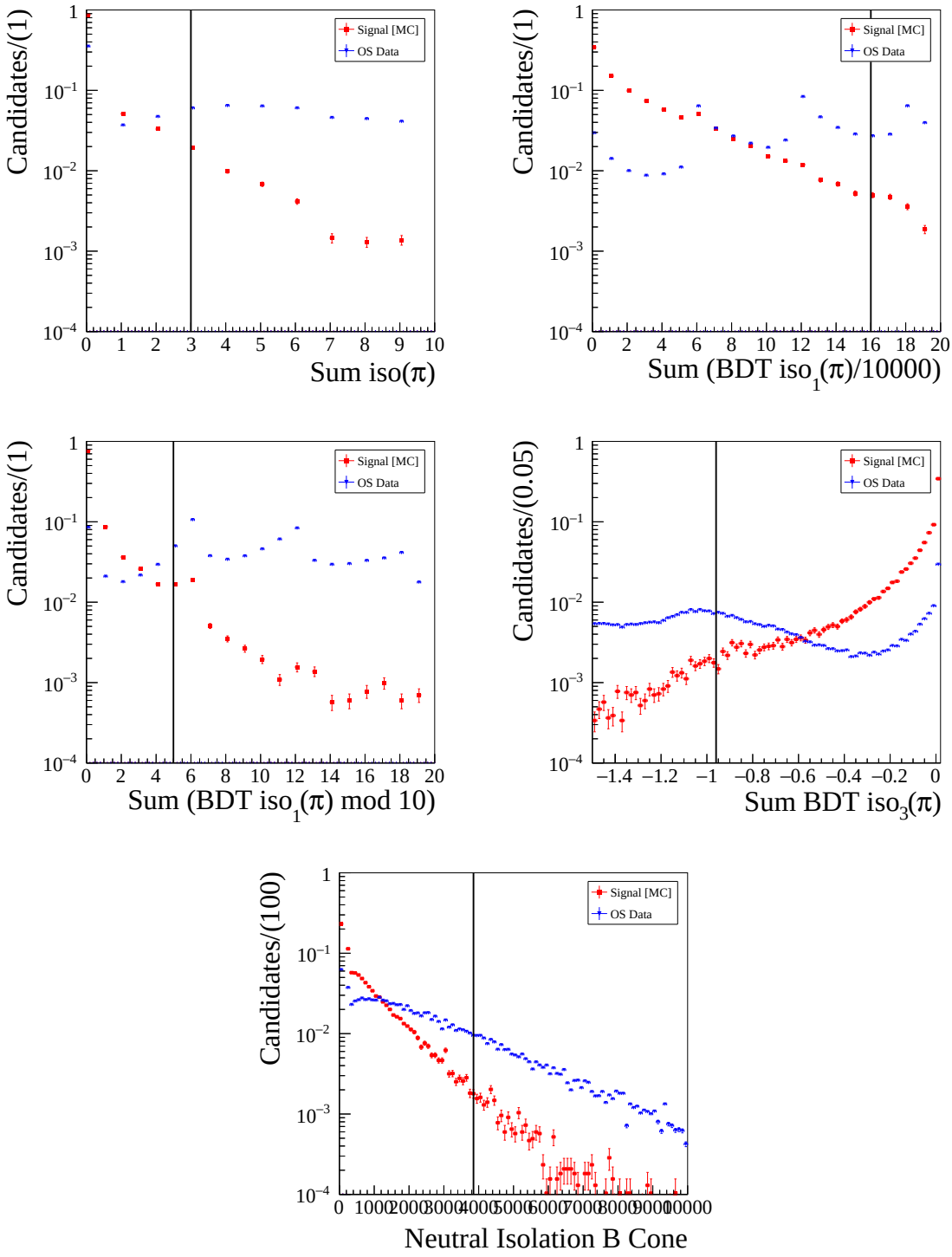

Figure A.2: Comparison between the observable distributions of the $B^0_{d/s} \to \tau^+\tau^-$ signal MC (Red) and the OS data (Blue), for the observables used in the cut-based loose selection. The vertical black lines indicate the position of the selection cuts.

Table A.3: List of input variables for the optimization of NN2. The variables below the horizontal line are present at the preprocessing stage of the NN, but not included in the actual training stage.

Rank		Variable
1	τ^+	Lifetime
2	τ^-	Lifetime
3	B_s^0	Neutral isolation cone variable vPT
4	τ^-	Mass
5	τ^+	Mass
6		τ reconstruction variable $\Re[x_3]$
7		$\sum_\pi$ third isolation BDT output
8	B_s^0	Transverse momentum
9	B_s^0	DIRA
10	τ^+	Isolation smallest $\Delta\chi^2$
11	τ^+	Decay vertex χ^2
12	τ^-	Decay vertex χ^2
13		τ reconstruction variable $\Im[\tilde{s}_3^{\pm}]$
14		Missing mass
15	B_s^0	Decay vertex χ^2
16	B_s^0	Corrected mass
17		τ reconstruction variable $\bar{\theta}_W$
18	τ^-	Distance of closest approach
19	τ^+	Distance of closest approach
20		τ reconstruction variable $p^+ \cdot p^-$
21		Number of real solutions for $p^+ \cdot p^-$
22	B_s^0	Mass
23		$\sum_\pi$ first isolation BDT output, 5th digit
24		Total number of candidates
25	τ^+	z displacement
26		τ reconstruction variable $\Im[x_1]$
27	τ^-	z displacement
28		$\sum_\pi$ first isolation BDT output, 1st digit
29		τ reconstruction variable $\Re[\xi_1]$
30	B_s^0	Isolation smallest $\Delta\chi^2$
31	τ^-	Transverse momentum
32	τ^+	Transverse momentum
33		τ reconstruction variable $\Re[\tilde{s}_1^{\pm}]$
34	B_s^0	Neutral isolation cone variable PZ
35		$\sum_\pi$ first isolation BDT output, 3rd digit
36	B_s^0	Neutral isolation cone variable mult

A.3 Selection of the $(3\pi, \mu)$ final state

The stripping selection is given in Table A.4.

Table A.4: Stripping requirements for $B^0_{(s)} \to \tau\tau(3\pi, \mu)$ and the normalisation channel $B^0_d \to D^-(\to \pi^+\pi^-\pi^-)\pi^+$.

		$B^0_{(s)} \to \tau\tau(3\pi, \mu)$		$B^0_d \to D^-(\to \pi^+\pi^-\pi^-)\pi^+$
cut	on	value	on	value
p	π	$> 2000\,\mathrm{MeV}/c$	π / K	$> 2000\,\mathrm{MeV}/c$
p_T		$> 250\,\mathrm{MeV}/c$		$> 250\,\mathrm{MeV}/c$
IP χ^2		> 16		> 16
track χ^2/ndf		< 3		< 3
ghost probability		< 0.3		< 0.3
PROBNNpi		> 0.55	π	> 0.55
PIDK		-	K	> -5
p	μ	$> 6000\,\mathrm{MeV}/c$	π	$> 2000\,\mathrm{MeV}/c$
p_T		$> 1000\,\mathrm{MeV}/c$		$> 250\,\mathrm{MeV}/c$
IP χ^2		> 16		> 16
track χ^2/ndf		< 3		< 3
ghost probability		< 0.3		< 0.3
PIDmu		> 0		-
p_T	B	$> 5000\,\mathrm{MeV}/c$	B	$> 5000\,\mathrm{MeV}/c$
M		$[2000, 7000]\,\mathrm{MeV}/c^2$		$[2000, 7000]\,\mathrm{MeV}/c^2$
M_{corr}		$< 10000\,\mathrm{MeV}/c^2$		$< 10000\,\mathrm{MeV}/c^2$
IPχ^2		< 200		< 200
vertex distance		< 35mm		< 35mm
sum p_T of π, μ		$> 2500\,\mathrm{MeV}/c$		$> 2500\,\mathrm{MeV}/c$
M	τ	$[400, 2100]\,\mathrm{MeV}/c^2$	D	$[1750, 2080]\,\mathrm{MeV}/c^2$
At least 1 daughter with p_T		$> 800\,\mathrm{MeV}/c$		$> 800\,\mathrm{MeV}/c$
IPχ^2		> 50		> 50
MaxDOCA		< 0.2 mm		< 0.2 mm
p_T		$> 1000\,\mathrm{MeV}/c$		$> 1000\,\mathrm{MeV}/c$
θ_{DIRA}		> 0.99		> 0.99
vertex χ^2		< 12		< 12
vertex distance in χ^2		> 16		> 16
vertex distance in χ^2		< 4000		< 4000
vertex distance in ρ		$[0.1, 7]$mm		$[0.1, 7]$mm
vertex distance in z		5mm		5mm
TOS on HLT2(Topo*BodyBBDT or TopoMu or SingleMuon)		yes		no

Bibliography

[1] A. J. Buras, "Weak Hamiltonian, CP violation and rare decays," in *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*, pp. 281–539, 1998.

[2] S. Aoki *et al.*, "Review of lattice results concerning low-energy particle physics," *Eur. Phys. J.*, vol. C77, no. 2, p. 112, 2017.

[3] Y. Amhis *et al.*, "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016 and updates at https://hflav.web.cern.ch/," *Eur. Phys. J.*, vol. C77, no. 12, p. 895, 2017.

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

[5] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou, and M. Steinhauser, "$B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty," *Phys. Rev. Lett.*, vol. 112, p. 101801, 2014.

[6] C. Bobeth, M. Gorbahn, and E. Stamou, "Electroweak Corrections to $B_{s,d} \to \ell^+\ell^-$," *Phys. Rev.*, vol. D89, no. 3, p. 034023, 2014.

[7] T. Hermann, M. Misiak, and M. Steinhauser, "Three-loop QCD corrections to $B_s \to \mu^+\mu^-$," *JHEP*, vol. 12, p. 097, 2013.

[8] S. Aoki *et al.*, "Review of lattice results concerning low-energy particle physics," *Eur. Phys. J.*, vol. C74, p. 2890, 2014.

[9] U. collaboration, "http://www.utfit.org/utfit/."

[10] C. collaboration, "http://ckmfitter.in2p3.fr/."

[11] J. Beringer *et al.*, "Review of Particle Physics (RPP)," *Phys. Rev.*, vol. D86, p. 010001, 2012.

[12] M. Beneke, C. Bobeth, and R. Szafron, "Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \to \mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 120, no. 1, p. 011801, 2018.

[13] S. R. Choudhury and N. Gaur, "Dileptonic decay of B(s) meson in SUSY models with large tan Beta," *Phys. Lett.*, vol. B451, pp. 86–92, 1999.

[14] W. Buchmuller and D. Wyler, "Effective Lagrangian Analysis of New Interactions and Flavor Conservation," *Nucl. Phys.*, vol. B268, pp. 621–653, 1986.

[15] R. Alonso, B. Grinstein, and J. Martin Camalich, "$SU(2) \times U(1)$ gauge invariance and the shape of new physics in rare B decays," *Phys. Rev. Lett.*, vol. 113, p. 241802, 2014.

[16] 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.

[17] A. J. Buras, R. Fleischer, J. Girrbach, and R. Knegjens, "Probing New Physics with the $B_s \to \mu^+\mu^-$ Time-Dependent Rate," *JHEP*, vol. 07, p. 77, 2013.

[18] K. Nakamura *et al.*, "Review of particle physics," *J. Phys.*, vol. G37, p. 075021, 2010.

[19] T. Aaltonen *et al.*, "Search for the Decays $B_s^0 \to e^+\mu^-$ and $B_s^0 \to e^+e^-$ in CDF Run II," *Phys. Rev. Lett.*, vol. 102, p. 201801, 2009.

[20] T. Aaltonen *et al.*, "Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays with $2fb^{-1}$ of $p\bar{p}$ collisions," *Phys. Rev. Lett.*, vol. 100, p. 101802, 2008.

[21] B. Aubert *et al.*, "A search for the rare decay $B^0 \to \tau^+\tau^-$ at BABAR," *Phys. Rev. Lett.*, vol. 96, p. 241802, 2006.

[22] V. Khachatryan *et al.*, "Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data," *Nature*, vol. 522, pp. 68–72, 2015.

[23] R. Aaij *et al.*, "Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$," *Phys. Lett.*, vol. B699, pp. 330–340, 2011.

[24] R. Aaij *et al.*, "Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$," *Phys. Lett.*, vol. B708, pp. 55–67, 2012.

[25] R. Aaij *et al.*, "Strong constraints on the rare decays $B_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 108, p. 231801, 2012.

[26] R. Aaij *et al.*, "First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 110, no. 2, p. 021801, 2013.

[27] R. Aaij *et al.*, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment," *Phys. Rev. Lett.*, vol. 111, p. 101805, 2013.

[28] 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.

[29] 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.

[30] R. Aaij *et al.*, "Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region," *Phys. Lett.*, vol. B694, pp. 209–216, 2010.

[31] R. Aaij *et al.*, "Measurement of the track reconstruction efficiency at LHCb," *JINST*, vol. 10, no. 02, p. P02007, 2015.

[32] R. Aaij *et al.*, "LHCb Detector Performance," *Int. J. Mod. Phys.*, vol. A30, no. 07, p. 1530022, 2015.

[33] V. V. Gligorov, C. Thomas, and M. Williams, "The HLT inclusive B triggers," Tech. Rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016. LHCb-INT-2011-030, CERN, Geneva, Sep 2011. LHCb-INT-2011-030.

[34] C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[35] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys.*, vol. G28, pp. 2693–2704, 2002. [,11(2002)].

[36] "Central Exclusive Dimuon Production at $\sqrt{s} = 7$ TeV," May 2011. LHCb-ANA-2011-033.

[37] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez, E. Polycarpo, and A. Sarti, "The Muon Identification Procedure of the LHCb Experiment for the First Data," Tech. Rep. LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013, CERN, Geneva, Aug 2009.

[38] R. Aaij and J. Albrecht, "Muon triggers in the High Level Trigger of LHCb," Tech. Rep. LHCb-PUB-2011-017. CERN-LHCb-PUB-2011-017, CERN, Geneva, Sep 2011.

[39] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," *PoS*, vol. ACAT, p. 040, 2007.

[40] A. Abulencia *et al.*, "Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II," *Phys. Rev. Lett.*, vol. 95, p. 221805, 2005. [Erratum: Phys. Rev. Lett.95,249905(2005)].

[41] M. Perrin-Terrin and G. Mancinelli, "Optimisation of the binning of the discriminating variables used in the computation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ upper limits with the modified frequentist approach," Tech. Rep. LHCb-INT-2012-003. CERN-LHCb-INT-2012-003, CERN, Geneva, Feb 2012.

[42] A. Carbone, D. Galli, U. Marconi, S. Perazzini, A. Sarti, V. Vagnoni, and G. Valenti, "Invariant mass line shape of $B \to PP$ decays at LHCb," Tech. Rep. LHCb-PUB-2009-031. CERN-LHCb-PUB-2009-031, CERN, Geneva, Jan 2010.

[43] Y. G. Aditya, K. J. Healey, and A. A. Petrov, "Faking $B_s \to \mu^+\mu^-$," *Phys. Rev.*, vol. D87, p. 074028, 2013.

[44] J. A. Hernando Morata, E. Lopez Asamar, D. Martinez Santos, H. Ruiz-Pérez, and F. Teubert, "Measurement of trigger efficiencies and biases," Tech. Rep. LHCb-2008-073. CERN-LHCb-2008-073, CERN, Geneva, Mar 2010.

[45] 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.

[46] R. Aaij *et al.*, "Measurement of b hadron production fractions in 7 TeV pp collisions," *Phys.Rev.*, vol. D85, p. 032008, 2012.

[47] 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.

[48] "Updated average f_s/f_d b-hadron production fraction ratio for $7TeV$ pp collisions," Jul 2013.

[49] S. Chatrchyan *et al.*, "Measurement of the $(B_s^0 \to \mu^+\mu^-)$ branching fraction and search for $B_d^0 \to \mu^+\mu^-)$ with the CMS Experiment," *Phys. Rev. Lett.*, vol. 111, p. 101804, 2013.

[50] T. L. collaboration, "Framework TDR for the LHCb Upgrade: Technical Design Report," Tech. Rep. CERN-LHCC-2012-007. LHCb-TDR-12, Apr 2012.

[51] L. E. The LHCb Collaboration, "Physics case for an LHCb Upgrade II," Tech. Rep. LHCb-PUB-2018-009. CERN-LHCb-PUB-2018-009, CERN, Geneva, May 2018.

[52] "Prospects for the $\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)$ measurements with the ATLAS detector in the Run 2 and HL-LHC data campaigns," Tech. Rep. ATL-PHYS-PUB-2018-005, CERN, Geneva, May 2018.

[53] C. Collaboration, "Measurement of rare $B \to \mu^+\mu^-$ decays with the Phase-2 upgraded CMS detector at the HL-LHC," 2018.

[54] A. Mordá, "Rare dileptonic $B_{(s)}^0$ meson decays at LHCb, PhD Thesis," *Aix-Marseille Université*, 2015. CERN-THESIS-2015-264.

[55] C. Meaux, J. Cogan, K. A. M. De Bruyn, G. Mancinelli, and J. Serrano, "Search for $B_{(s)}^0 \to \tau^+\tau^-$ in the $(3\pi, \mu)$ final state," Tech. Rep. LHCb-INT-2018-021. CERN-LHCb-INT-2018-021, CERN, Geneva, Jun 2018.

[56] I. M. Nugent, T. Przedzinski, P. Roig, O. Shekhovtsova, and Z. Was, "Resonance chiral Lagrangian currents and experimental data for $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$," *Phys. Rev.*, vol. D88, p. 093012, 2013.

[57] I. M. Nugent, "Invariant mass spectra of $\tau^- \to h^-h^-h^+\nu_\tau$ decays," *Nucl. Phys. Proc. Suppl.*, vol. 253-255, pp. 38–41, 2014.

[58] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was, and Z. Was, "Universal interface of TAUOLA technical and physics documentation," *Comput. Phys. Commun.*, vol. 183, pp. 821–843, 2012.

[59] K. De Bruyn, "Search for the rare decays $B_{(s)}^0 \to \tau^+\tau^-$," Tech. Rep. LHCb-CONF-2016-011. CERN-LHCb-CONF-2016-011, CERN, Geneva, Sep 2016.

[60] E. Etzion, "Measurement of R_b at SLD," pp. 139–145, 1996. Proceedings, 31st Rencontres de Moriond, Leptonic Session, Les Arcs, France, March 16-23, 1996.

[61] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," *Nucl. Instrum. Meth.*, vol. A559, pp. 190–194, 2006.

[62] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, "HistFactory: A tool for creating statistical models for use with RooFit and RooStats," 2012.

[63] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling," *eConf*, vol. C0303241, p. MOLT007, 2003. [,186(2003)].

[64] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J.*, vol. C71, p. 1554, 2011. [Erratum: Eur. Phys. J. **C73** (2013) 2501].

[65] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, "The RooStats project," *PoS*, vol. ACAT2010, p. 057, 2010.

[66] D. Martínez Santos and F. Dupertuis, "Mass distributions marginalized over per-event errors," *Nucl.Instrum.Meth.*, vol. A764, p. 150, 2014.

[67] R. Aaij *et al.*, "First observations of $\bar{B}^0_s \to D^+D^-$, $D^+_sD^-$ and $D^0\bar{D}^0$ decays," *Phys. Rev.*, vol. D87, p. 092007, 2013.

[68] Z. Was and J. Zaremba, "Study of variants for Monte Carlo generators of $\tau \to 3\pi\nu$ decays," *Eur. Phys. J.*, vol. C75, p. 566, 2015.

[69] D. M. Asner *et al.*, "Hadronic structure in the decay $\tau^- \to \nu_\tau\pi^-\pi^0\pi^0$ and the sign of the tau-neutrino helicity," *Phys. Rev.*, vol. D61, p. 012002, 2000.

[70] W. Altmannshofer *et al.*, "The Belle II Physics Book," 2018.

[71] T. Keck *et al.*, "The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment," 2018.

[72] M. Mangano, P. Azzi, M. Benedikt, A. Blondel, D. A. Britzger, A. Dainese, M. Dam, J. de Blas, D. Enterria, O. Fischer, C. Grojean, J. Gutleber, C. Gwenlan, C. Helsens, P. Janot, M. Klein, U. Klein, M. P. Mccullough, S. Monteil, J. Poole, M. Ramsey-Musolf, C. Schwanenberger, M. Selvaggi, F. Zimmermann, and T. You, "Future Circular Collider," Tech. Rep. CERN-ACC-2018-0056, CERN, Geneva, Dec 2018. Submitted for publication to Eur. Phys. J. C.

[73] B. Capdevila, A. Crivellin, S. Descotes-Genon, L. Hofer, and J. Matias, "Searching for New Physics with $b \to s\tau^+\tau^-$ processes," *Phys. Rev. Lett.*, vol. 120, no. 18, p. 181802, 2018.

[74] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski, and J. Rosiek, "Lepton-flavour violating B decays in generic Z' models," *Phys. Rev.*, vol. D92, no. 5, p. 054013, 2015.

[75] D. Bečirević, N. Košnik, O. Sumensari, and R. Zukanovich Funchal, "Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to s\ell_1\ell_2$ modes," *JHEP*, vol. 11, p. 035, 2016.

[76] R. Aaij *et al.*, "Search for the lepton-flavour-violating decays $B^0_s \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$," 2019.

[77] A. Crivellin, D. Müller, and T. Ota, "Simultaneous explanation of R(D$^{(*)}$) and $b \to s\mu\mu$: the last scalar leptoquarks standing," *JHEP*, vol. 09, p. 040, 2017.

[78] D. Buttazzo, A. Greljo, G. Isidori, and D. Marzocca, "B-physics anomalies: a guide to combined explanations," *JHEP*, vol. 11, p. 044, 2017.

[79] W. Altmannshofer, C. Niehoff, and D. M. Straub, "$B_s \to \mu^+\mu^-$ as current and future probe of new physics," *JHEP*, vol. 05, p. 076, 2017.

[80] W. Buchmuller, R. Ruckl, and D. Wyler, "Leptoquarks in Lepton - Quark Collisions," *Phys. Lett.*, vol. B191, pp. 442–448, 1987. [Erratum: Phys. Lett.B448,320(1999)].

[81] R. Aaij *et al.*, "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction," *JHEP*, vol. 11, p. 047, 2016. [Erratum: JHEP04,142(2017)].

[82] R. Aaij *et al.*, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity," *JHEP*, vol. 02, p. 104, 2016.

[83] R. Aaij *et al.*, "Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$," *JHEP*, vol. 09, p. 179, 2015.

[84] R. Aaij *et al.*, "Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays," *JHEP*, vol. 06, p. 115, 2015. [Erratum: JHEP09,145(2018)].

[85] R. Aaij *et al.*, "Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays," *Phys. Rev. Lett.*, vol. 122, no. 19, p. 191801, 2019.

[86] R. Aaij *et al.*, "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays," *JHEP*, vol. 08, p. 055, 2017.

[87] A. Abdesselam *et al.*, "Test of lepton flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle," 2019.

[88] J. T. Wei *et al.*, "Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for $B \to K^{(*)}\ell^+\ell^-$," *Phys. Rev. Lett.*, vol. 103, p. 171801, 2009.

[89] A. Abdesselam *et al.*, "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method," 2019.

[90] S. Hirose *et al.*, "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle," *Phys. Rev.*, vol. D97, no. 1, p. 012004, 2018.

[91] R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$," *Phys. Rev. Lett.*, vol. 115, no. 11, p. 111803, 2015. [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)].

[92] M. Huschle *et al.*, "Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle," *Phys. Rev.*, vol. D92, no. 7, p. 072014, 2015.

[93] J. P. Lees *et al.*, "Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons," *Phys. Rev.*, vol. D88, no. 7, p. 072012, 2013.

[94] R. Aaij *et al.*, "Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays," *Phys. Rev.*, vol. D97, no. 7, p. 072013, 2018.

[95] R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$," *Phys. Rev. Lett.*, vol. 120, no. 12, p. 121801, 2018.

[96] C. Cornella, J. Fuentes-Martin, and G. Isidori, "Revisiting the vector leptoquark explanation of the B-physics anomalies," 2019.

[97] D. Choudhury, A. Kundu, R. Mandal, and R. Sinha, "$R_{K^{(*)}}$ and $R(D^{(*)})$ anomalies resolved with lepton mixing," *Nucl. Phys.*, vol. B933, pp. 433–453, 2018.

[98] B. Bhattacharya, A. Datta, D. London, and S. Shivashankara, "Simultaneous Explanation of the R_K and $R(D^{(*)})$ Puzzles," *Phys. Lett.*, vol. B742, pp. 370–374, 2015.

[99] D. Guadagnoli and K. Lane, "Charged-Lepton Mixing and Lepton Flavor Violation," *Phys. Lett.*, vol. B751, pp. 54–58, 2015.

[100] F. Feruglio, P. Paradisi, and A. Pattori, "Revisiting Lepton Flavor Universality in B Decays," *Phys. Rev. Lett.*, vol. 118, no. 1, p. 011801, 2017.

[101] S. L. Glashow, D. Guadagnoli, and K. Lane, "Lepton Flavor Violation in B Decays?," *Phys. Rev. Lett.*, vol. 114, p. 091801, 2015.

[102] R. Alonso, B. Grinstein, and J. Martin Camalich, "Lepton universality violation and lepton flavor conservation in B-meson decays," *JHEP*, vol. 10, p. 184, 2015.

[103] L. Calibbi, A. Crivellin, and T. Li, "Model of vector leptoquarks in view of the B-physics anomalies," *Phys. Rev.*, vol. D98, no. 11, p. 115002, 2018.