



Search for the $B_d^0 \rightarrow \mu^+ \mu^-$ decay
and measurement of the
 $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and
effective lifetime

Ph.D. thesis

XXX ciclo di dottorato

Università degli studi di Roma Tre - Dipartimento di Matematica e Fisica

Candidate:

Marco Santimaria

Supervisor:

Matteo Palutan

Referees:

Patrick Koppenburg

Aleandro Nisati

Contents

Introduction	2
1 Theory of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays	3
1.1 Introduction to the Standard Model	3
1.1.1 Particle masses: the Higgs mechanism	5
1.1.2 Quark masses and mixing	6
1.1.3 The CKM matrix	7
1.2 $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ in the Standard Model	10
1.2.1 An Effective Field Theory for B decays	10
1.2.2 B^0 mixing and $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ branching fractions	13
1.2.3 $B_s^0 \rightarrow \mu^+ \mu^-$ Effective Lifetime	15
1.2.4 The Standard Model branching fraction	16
1.3 $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ and new physics models	19
1.3.1 Non CP-violating new physics ($\varphi_P, \varphi_S \in \{0, \pi\}$)	20
1.3.2 Pseudo-scalar dominated new physics ($S = 0$)	22
1.3.3 Scalar dominated new physics ($P = 1$)	24
1.3.4 Mixed scalar/pseudo-scalar new physics ($P \pm S = 1$)	24
1.4 Current experimental status	26
2 The LHCb experiment	31
2.1 The Large Hadron Collider	31
2.1.1 b physics at the LHC	35
2.2 Overview of the LHCb detector	40
2.3 The Tracking System	40
2.3.1 The Magnet	42
2.3.2 The Vertex Locator	43
2.3.3 The Silicon Tracker	44
2.3.4 The Outer Tracker	47
2.4 The Particle Identification System	48
2.4.1 The Ring Imaging Cherenkov Detector	48
2.4.2 The Calorimeters	49

2.4.3	The Muon System	52
2.5	The Trigger System	58
2.5.1	The Level-0 Trigger	59
2.5.2	The High Level Trigger	62
2.5.3	Trigger efficiency measurement from data	64
2.6	Offline Analysis	65
Overview of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis		71
3	Muon Identification	73
3.1	The muonID algorithm	73
3.1.1	IsMuon selection	73
3.1.2	Muon Likelihood	76
3.1.3	Combined PID	77
3.1.4	The PID selection	79
3.2	Measurement of the PID efficiency from data	80
3.2.1	Calibration samples	80
3.2.2	PID efficiency computation	82
3.3	PID efficiencies for $B_{d,s}^0 \rightarrow \mu^+ \mu^-$	83
3.3.1	Muon identification	85
3.3.2	Hadron misidentification	87
4	Signal selection and characterisation	91
4.1	Selection	91
4.2	The BDT for combinatorial background rejection	94
4.2.1	Boosted decision trees	95
4.2.2	Muon isolation	96
4.2.3	The global BDT	98
4.2.4	BDT calibration	99
4.2.5	Time-dependent effects	102
4.3	Mass calibration	105
4.3.1	Mean	105
4.3.2	Resolution	106
4.3.3	Tail parameters	107
4.4	Normalisation	108
4.4.1	Normalisation factors	110
5	Exclusive backgrounds	113
5.1	$B_{d,s}^0 \rightarrow h^+ h'^-$	114
5.2	Semileptonic backgrounds	117
5.2.1	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	119

5.2.2	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	121
5.2.3	$B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B^0 \rightarrow \pi^0 \mu^+ \mu^-$	123
5.2.4	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	124
5.2.5	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	126
5.3	Data-driven background estimate	128
5.3.1	The $h - \mu$ selection	130
5.3.2	The $h - \mu$ fit model	131
5.3.3	Yield computation	132
5.4	Results and systematic uncertainties	133
5.4.1	$B_{d,s}^0 \rightarrow h^+ h'^- \rightarrow \mu^+ \mu^-$	133
5.4.2	Semileptonic backgrounds	134
5.4.3	Background yields in the signal region	135
6	Results	155
6.1	The signal fit	155
6.1.1	Fit model	156
6.1.2	Toy Monte Carlo studies	157
6.1.3	Statistical significance of a peak	160
6.2	Branching fraction results	160
6.2.1	2D likelihood scan	163
6.2.2	$B_d^0 \rightarrow \mu^+ \mu^-$ limit computation with CL_s	164
6.3	Effective lifetime measurement	166
6.3.1	Analysis strategy	167
6.3.2	Results	169
6.4	Implications and future prospects	171
6.5	Conclusions	175
	List of figures	177
	List of tables	184
	References	187

Introduction

Our current knowledge on the nature of matter and its interactions is described by the Standard Model (SM) of particle physics, a quantum field theory which has been able to describe the wide variety of experimental results achieved in the last decades.

Besides the outstanding predictions that have established the SM as the paradigm of particle physics, there are a few relevant observations which are not accounted for in the theory:

1. the nature of 95% of the matter and energy that make up the universe: dark matter and dark energy,
2. the huge disproportion between matter and anti-matter we observe in the universe,
3. gravity force,
4. the nature of neutrino masses.

Many New Physics (NP) models have been proposed to explain these and other SM shortcomings.

The search for NP is the current aim of particle physics and might be pursued *directly*, by producing new possible particles in high energy collisions, or *indirectly*, by measurements of processes in which loops of new virtual particles might affect, for example, the decay rate. Being not limited by the collision energy, indirect searches are sensitive to particle masses which are larger than those accessible in direct searches. For this reason, indirect searches are a powerful tool to probe heavy particles that cannot be produced at colliders.

The $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays are among the most sensitive probes to physics beyond the SM. Such decays are extremely rare, occurring few times in billions of B decays, due to loop and helicity suppressions. The decay probability is however precisely predicted in the SM, as the purely leptonic final state allows to condensate hadronic interactions into a single constant. The quest for $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays started more than 30 years ago at the CLEO experiment [1], received a significant

boost at Fermilab with D0 [2] and CDF [3], and finally came into success at LHC with the LHCb and CMS joint observation of the $B_s^0 \rightarrow \mu^+\mu^-$ process, at a rate in agreement with the SM expectation [4]. The precision achieved in the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction measurement enormously constrained NP models. However, the search continues: there is still room for new physics effects in the $B_s^0 \rightarrow \mu^+\mu^-$ mode but new observables and precision measurements are required. In addition, a possible observation of the $B_d^0 \rightarrow \mu^+\mu^-$ decay at a rate above its SM prediction, still allowed by the present experimental constraints, would be an unambiguous sign of NP. In particular, the result from [4] suggests a slight enhancement of the $B_d^0 \rightarrow \mu^+\mu^-$ rate, which strongly pushed for a step further in the analysis, concerning both statistics and background rejection power.

In this thesis, the LHCb measurement using Run 1 and a fraction of Run 2 data is presented. This work represents a substantial step towards the aforementioned goal, since rejection and estimation of the backgrounds have been greatly improved. Thanks to the enlarged statistics, together with a new and optimised analysis, the first single experiment observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay was made possible.

In **Chapter 1**, after a brief introduction to the SM, the theory of $B_{d,s}^0 \rightarrow \mu^+\mu^-$ decays is discussed in both SM and NP scenarios, and the present experimental picture is given. In **Chapter 2**, a short description of the LHC accelerator complex and the LHCb detector is provided, focussing on the parts which are directly related to this work. **Chapter 3** is devoted to the particle identification, giving details on all the aspects which are relevant for the analysis. The signal selection and normalisation are described in **Chapter 4**, while **Chapter 5** is devoted to the background estimate, with each background source investigated in detail. Results, implications and future prospects are finally given in **Chapter 6**.

The main contributions I brought to the current work pertain to Chapters 3, 5 and 6.

Chapter 1

Theory of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays

In this chapter, a brief description of the theory elements needed to understand the processes which determine the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays is given.

After a historical introduction to the Standard Model, the origin of modern flavour physics is presented through the Higgs mechanism and the CKM matrix.

The argument then evolves towards the theory of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays: an introduction to effective field theories allows to understand how the branching fraction and effective lifetime measurements are connected to the expectations from the SM as well as from many new physics models.

1.1 Introduction to the Standard Model

The pursuit of a theory unifying the principles of Quantum Mechanics and the principle of Special Relativity culminated in the Quantum Field Theory (QFT) [5–7]. The established QFT of particle physics, the Standard Model (SM), is based on quarks, leptons and bosonic force carriers, as depicted in Fig. 1.1. A quantized field is associated to each one of these particles, which at the present energy scale are considered to be elementary. Four separated classes of interactions have been observed in nature: electromagnetic, weak, strong and gravitational. A description of the gravity force is not included within the SM framework, as its QFT has not been yet developed.

The Yang and Mills theory [8], which extended the concept of covariant derivative to local and non-abelian symmetry groups, set the basis for the development of a gauge theory of weak interactions. In analogy with the isospin symmetry, leptons and neutrinos were grouped to form doublets:

$$\psi = \begin{pmatrix} \nu_l \\ l \end{pmatrix}, \quad (1.1)$$

Standard Model of Elementary Particles

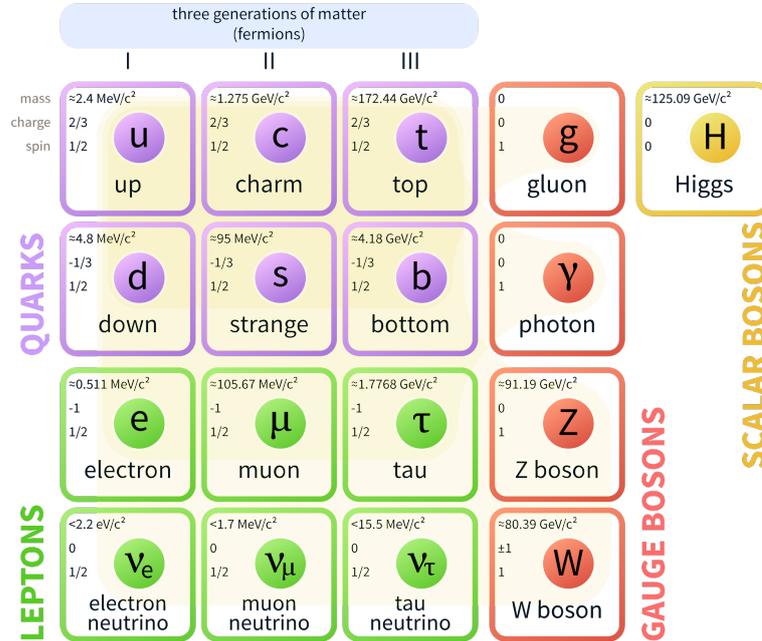


Figure 1.1: The elementary particles of the Standard Model.

with the aim of finding the symmetries which can lead, through the Noether theorem [9], to the weak current. Glashow firstly identified in his theory [10] that the proper symmetry group to describe both electromagnetic and weak interactions was $SU(2)_L \otimes U(1)_Y$,¹ with the ensuing need of a neutral intermediate boson, the Z^0 , alongside the charged W^\pm bosons and the photon. The key problem of his formulation was that gauge bosons were massless: including “ad hoc” mass terms, which explicitly violate the gauge symmetry, led to a non-renormalisable theory. The issue was only solved with the development of the Higgs mechanism (Sec. 1.1.1), so that a unified, renormalisable theory of electromagnetic and weak interactions could be finally completed [11, 12].

Starting from the pioneering experiment of R. W. McCallister and R. Hofstadter [13], and the theory from R. Feynman [14], J.D. Bjorken [15] and M. Gell-Mann [16], we discovered that hadrons, i.e. all the strongly-interacting particles, are made of elementary constituents called quarks. Hadrons can be mesons or baryons, depending on whether they are made of 2 or 3 valence quarks. Along with the valence quarks, building hadrons and their quantum numbers, a sea quark component is

¹Weak hypercharge is defined from electromagnetic charge Q and the third component of the weak isospin I_3^W according to the Gell-Mann - Nishijima formula: $Y = 2Q - 2I_3^W$.

also present: quark-antiquark couples are continuously generated inside the strong field. The mediators of the strong interactions are the gluons, massless bosons which in their turn carry the colour charge. Quarks appear in three different colours and arrange themselves into the colourless hadrons we observe, following the hadronisation process.² The theory of the strong interactions, the Quantum Chromodynamics (QCD), is represented by the $SU(3)$ symmetry group. The complete symmetry group of the SM therefore is

$$SU(2)_L \otimes U(1)_Y \otimes SU(3)_c. \quad (1.2)$$

Each interaction term in the SM Lagrangian is shaped by its underlying symmetry, all of them being non-abelian gauge transformations. This common denominator for such phenomenologically diverse interactions leads to think of the SM as the manifestation of an even more symmetric and unified structure.

1.1.1 Particle masses: the Higgs mechanism

The missing piece of the Electroweak theory was how its gauge bosons could acquire a mass. A mechanism was needed to explain how the EW symmetry breaks, preserving the gauge invariance of the electromagnetism, according to the scheme:

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{em}. \quad (1.3)$$

The idea was to extend the Goldstone theorem to local gauge transformations, so that a *spontaneous symmetry breaking* could happen.

In 1964, P. Higgs, R. Brout and F. Englert proposed [17] the existence of a complex scalar field ϕ with a potential:

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \quad (1.4)$$

which is symmetric under rotation in the $\phi_1 - \phi_2$ plane, ϕ_1 and ϕ_2 being the components of the field ϕ . The theory behaves differently for positive and negative values of μ^2 :

$\mu^2 > 0$

The potential has a parabolic shape and the fields ϕ_1 and ϕ_2 represent two degenerate particles with mass μ , the minimum energy configuration being at the bottom of the parabola.

²The top quark is an exception: its lifetime is so small that it decays before hadronising.

$$\mu^2 < 0$$

The potential has a *sombrero* shape (Fig. 1.2) : the previous energy configuration is unstable (local maximum) and the minimum resides in all the points in the circular region around the origin. ϕ acquires a vacuum expectation value (VEV) of $\eta \equiv \mu/\sqrt{2\lambda}$ which breaks the initial rotation symmetry. The fluctuations of ϕ

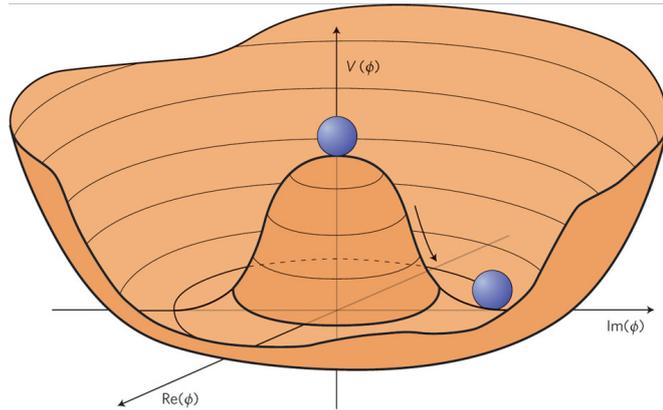


Figure 1.2: Higgs potential for $\mu^2 < 0$ [18].

around the VEV give rise to the so-called Higgs particle, which in the EW theory is incorporated into the gauge bosons providing them their masses.

In July 2012, almost 50 years after this theory was formulated, the ATLAS and CMS collaborations at CERN observed [19,20] a new spin 0 [21,22] particle with a mass of about $125 \text{ GeV}/c^2$, whose properties are compatible with the Higgs boson.

1.1.2 Quark masses and mixing

The idea of quark mixing as a consequence of symmetry breaking was introduced by N. Cabibbo in 1963 [23], in the framework of the recently developed theory by M. Gell-Mann, for which quarks came in 3 flavours: *up*, *down* and *strange*. Cabibbo observed that symmetry breaking could lead to a mixing between d_L and s_L , the left-chirality components of the *down* and *strange* quark fields, so that the weak charged current could be expressed as a function of the parameter θ_C , the Cabibbo angle:

$$J_\mu^1 + iJ_\mu^2 = \bar{u}_L \gamma_\mu (\cos \theta_C d_L + \sin \theta_C s_L). \quad (1.5)$$

However, such theory could not be merged, as it allowed *strangeness changing neutral currents*, for which $K \rightarrow \mu^+ \mu^-$ decays should be as probable as $K^+ \rightarrow \mu^+ \nu$ ones, in open contrast with the experimental observations.

This issue was solved in 1970 by S. L. Glashow, J. Iliopoulos e L. Maiani [24], who predicted the existence of a fourth quark, named *charm*, such that the weak neutral current became diagonal. The GIM mechanism was therefore able to explain the suppression in $K \rightarrow \mu^+ \mu^-$ decays by the negative interference of the charm contribution, and the first prediction of its mass yielded 2-3 GeV/c². The charm quark was discovered in 1974 under the form of a bound $c\bar{c}$ state: the J/ψ meson [25, 26].

Another open issue waiting for an explanation was the CP symmetry violation, observed in 1964 in neutral K meson decays [27]. In 1973, M. Kobayashi and T. Maskawa [28] found that at least 3 quark doublets, i.e. 6 quark fields, were mandatory to produce the complex phase in the mixing matrix that was able to explain the observed CP violation. The discovery of *beauty* [29] and *top* [30] quarks and a third family of leptons, made up by the τ [31] and its neutrino ν_τ [32], plainly confirmed that quarks and leptons can be depicted as:

$$\begin{aligned} Q_L &\equiv \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L; \\ Q_R &\equiv u_R, d_R, c_R, s_R, t_R, b_R; \end{aligned} \tag{1.6}$$

$$\begin{aligned} L_L &\equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L; \\ L_R &\equiv e_R, \mu_R, \tau_R. \end{aligned} \tag{1.7}$$

Measurements carried out by the four LEP experiments [33–36] have confirmed that the number of light neutrino families is 3, although this number is not constrained in the SM.

The assignments (1.6) and (1.7) specify quark and lepton interactions with the EW gauge fields: left-handed fields have *weak isospin* 1/2, while right-handed fields are singlets ($I^W = 0$). However, since the Higgs field provokes the symmetry breaking, a distinction has to be made between isospin-defined fields and physical fields, which actually create and destroy physical particles.

1.1.3 The CKM matrix

The interaction between the Higgs field and the quarks can be described by a Yukawa term, which is $SU(2)_L \otimes U(1)_Y$ invariant:

$$\mathcal{L}_Y = \sum_{ij} g_{ij}^D \bar{Q}_i \phi D_j + \sum_{ij} g_{ij}^U \epsilon_{ab} \bar{U}_i Q_j^a \phi^b + h.c., \tag{1.8}$$

where Q_i denotes the generic left-handed doublet of Eq. (1.6), while U_i and D_i represent the right-handed *up* and *down* type singlets, i.e. $(u, c, t)_R$ and $(d, s, b)_R$, respectively. $i, j = 1, 2, 3$ are the indices running on the 3 quark generations, and ϵ is the 2×2 antisymmetric tensor.

Upon substitution of the Higgs VEV into Eq. (1.8), the quark mass Lagrangian is obtained:

$$\mathcal{L}_m = \bar{D}_L M^d D_R + \bar{U}_R M^u U_L + h.c., \quad (1.9)$$

where the $M^{d,u} = g_{ij}^{D,U} \eta$ are the mass matrices in the space of quarks generations, so that a misalignment manifests between the *interaction base*, or EW base, and the *physical base*, or mass base, identified by the fields that diagonalize the mass matrices.

A switch to the mass base (denoted with the apex) is performed by means of the unitary transformations:

$$D_L \rightarrow V_L^d D'_L, \quad U_L \rightarrow V_L^u U'_L; \quad D_R \rightarrow V_R^d D'_R, \quad U_R \rightarrow V_R^u U'_R; \quad (1.10)$$

so that Eq. (1.9) becomes diagonal:

$$\mathcal{L}_m = \bar{D}'_L m^d D'_R + \bar{U}'_R m^u U'_L + h.c., \quad (1.11)$$

where $m^{d,u} = (V^{d,u})_L^\dagger M^{d,u} (V^{d,u})_R$ are diagonal.

When the weak isospin lowering operator is applied to a physical field, it produces a superposition of fields with different masses, which is not an eigenstate of the Hamiltonian. The charged weak current mediated by the W^\pm transforms in fact according to:

$$\bar{U}_L \gamma_\mu D_L \quad \rightarrow \quad \bar{U}'_L \gamma_\mu V_{CKM} \bar{D}'_L \quad (1.12)$$

where $V_{CKM} = V_L^u V_L^{d\dagger}$ is the Cabibbo-Kobayashi-Maskawa matrix [23, 28] which describes the quark mixing:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (1.13)$$

Besides including all quark transition magnitudes, an irreducible complex phase in the CKM matrix accounts for CP violation.

Starting from the experimental observation that all the diagonal elements of V_{CKM} are of $\mathcal{O}(1)$, Wolfenstein proposed a parametrisation [37] that directly exhibits the hierarchy of the quark couplings:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \quad (1.14)$$

Unlike charged currents, weak neutral currents are not affected by the base change (1.10), so that no flavour mixing terms are present. Therefore, Flavour Changing Neutral Current (FCNC) processes are only possible at higher orders, meaning that direct transitions between down or up type quarks are highly suppressed within the SM, as shown in Sec. 1.2.

1.2 $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ in the Standard Model

$B_d^0(\bar{b}d)$ and $B_s^0(\bar{b}s)$ decays into a pair of oppositely charged muons, $B_{d,s}^0 \rightarrow \mu^+ \mu^-$, are especially interesting and extremely rare in the SM.

Given the quark compositions of the B_d^0 and B_s^0 mesons, their dimuon decay implies a weak transition between two down-type quarks, $b \rightarrow d$ or $b \rightarrow s$, which is forbidden at the tree level in the SM (Fig. 1.4a), as deduced in 1.1.2.

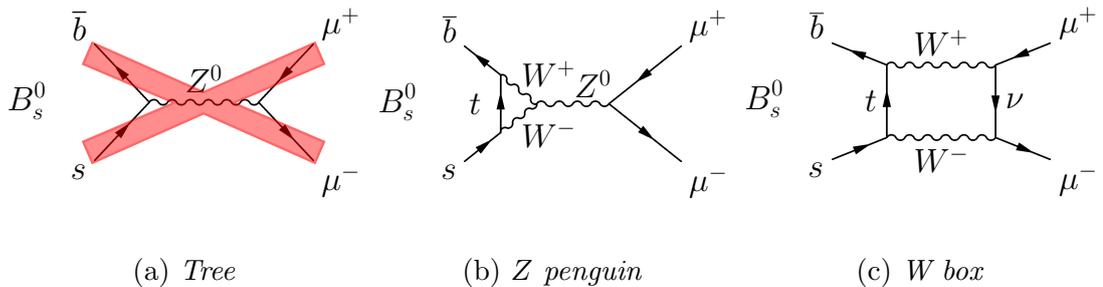


Figure 1.4: Dominant Feynman diagrams for $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays

Nevertheless, $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ can occur in the SM in higher order processes, the dominant ones being Z penguin with top loop (75%) and W box (24%) [40], as depicted in Fig. 1.4. In addition to being loop and CKM suppressed, $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays suffer significant helicity suppression. The neutral B mesons are pseudoscalars ($J^P = 0^-$), so that the two muons in the final state are forced to have the same helicity. The helicity state of one of the two muons is therefore always disfavoured by a factor $(m_\mu/M_B)^2 \sim 4 \times 10^{-4}$ with respect to the other.

1.2.1 An Effective Field Theory for B decays

The main obstacle in evaluating amplitudes for hadronic weak decays such as $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ is strong interaction. Conversely to QED, where higher order processes are suppressed by powers of $\alpha_{EM} \simeq 1/137$, the strong coupling of QCD largely depends on the transferred momentum scale of the process. At sufficiently

high energy, the strong coupling is small enough to allow for a perturbative approach, but at the low energy scale of the meson binding processes, where quarks are confined, non-perturbative methods such as lattice calculations are needed. Fortunately, the high and low energy scales can be disentangled in many cases [41]. A stratagem to separate *short-distance* (perturbative) and *long-distance* (generally non-perturbative) effects is used in the Effective Field Theory (EFT). Let us consider the simplest case of the neutron β -decay. The Feynman diagram in

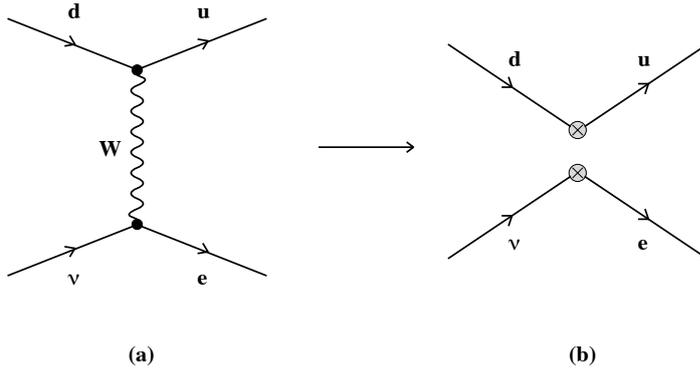


Figure 1.5: Neutron β -decay at the quark level in the full (a) and effective (b) theory.

Fig. 1.5a with full W-propagator represents the situation at very short distance scales of $\mathcal{O}(M_W)$, whereas the true picture of a decaying neutron, whose mass is $M_n \ll M_W$, is more properly described by effective point-like vertices which are represented by the *local operator* of Fig. 1.5b. An effective Hamiltonian can therefore be written as [42]

$$\mathcal{H}_{eff}^{(\beta)} = \frac{G_F}{\sqrt{2}} \cos \theta_c [\bar{u} \gamma_\mu (1 - \gamma_5) d \otimes \bar{e} \gamma^\mu (1 - \gamma_5) \nu_e], \quad (1.18)$$

which is the familiar Fermi theory for β -decays.

Analogously to Fermi theory, a generic effective weak Hamiltonian can be written as

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\lambda) \mathcal{O}_i(\lambda), \quad (1.19)$$

where \mathcal{O}_i are the local operators relevant for the decay and C_i are called Wilson coefficients, which, together with the CKM matrix elements, describe the strength with which a given operator enters the Hamiltonian. \mathcal{H}_{eff} is thus represented as a series, known as Operator Product Expansion (OPE), of effective vertices multiplied by effective coupling constants C_i .

A factorization scale λ decouples the high energy and low energy contributions in the process. The non-perturbative, i.e. *long-distance* part is described by the operators $\mathcal{O}_i(\lambda)$, while $C_i(\lambda)$ encode the physics contributions at scales higher than λ , i.e. *short-distance*, and can be calculated in perturbation theory as long as λ is not too small.

With the Hamiltonian (1.19), the decay amplitude of a meson M into a final state F is given by

$$\mathcal{A}(M \rightarrow F) = \langle F | \mathcal{H}_{eff} | M \rangle = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\lambda) \langle F | \mathcal{O}_i(\lambda) | M \rangle, \quad (1.20)$$

where the hadronic matrix elements $\langle F | \mathcal{O}_i(\lambda) | M \rangle$ are usually computed with lattice QCD and generally represent the largest source of uncertainty in the amplitude evaluation.

The scale λ is customarily set to the order of the decaying hadron mass, $\lambda = \mathcal{O}(m_b)$ for B -decays. As the choice of the scale must not affect the amplitude, the λ -dependence of the Wilson coefficients $C_i(\lambda)$ has to cancel the λ -dependence of the hadronic matrix elements. When the energy scale is lowered from $\lambda = \mathcal{O}(M_W)$, some high energy contributions are transferred from the hadronic matrix elements into the $C_i(\lambda)$: in other words, it is a matter of choice what exactly belongs to $C_i(\lambda)$ and what to $\mathcal{O}_i(\lambda)$. In addition, renormalisation of the local operators is necessary when QCD or QED corrections, i.e. higher order processes, are taken into account. Therefore, as for λ , the hadronic matrix element dependence on the renormalisation scheme must be cancelled out by the one of the $C_i(\lambda)$, so that the physical amplitude is independent from the chosen scheme.

The values of Wilson coefficients are calculated by *matching* the decay amplitude in the effective theory onto the one evaluated in the full theory. This matching is performed at $\lambda = \mathcal{O}(M_W)$, to get rid of the large logarithms $\ln M_W/\lambda$, and the $C_i(\lambda)$ are afterwards evolved down to the $\mathcal{O}(m_b)$ scale by means of renormalisation group equations. Such procedure is known under the name of *renormalisation group improved* perturbative expansion [43]. It is important to notice that the construction of \mathcal{H}_{eff} is fully done in the perturbative framework, irrespective of the complicated momentum configuration of the quarks bound in a meson state. The C_i coefficients are in fact independent on the external states.

Concerning $B_q^0 \rightarrow \mu^+ \mu^-$ processes ($q = d, s$), the only operators giving non-vanishing contributions to the decay amplitudes are [44]:

$$\begin{aligned} \mathcal{O}_{10} &= (\bar{q} \gamma_\mu P_L b) (\bar{\mu} \gamma^\mu \gamma_5 \mu), \\ \mathcal{O}_S &= m_b (\bar{q} P_R b) (\bar{\mu} \mu), \\ \mathcal{O}_P &= m_b (\bar{q} P_R b) (\bar{\mu} \gamma_5 \mu). \end{aligned} \quad (1.21)$$

The operator $\mathcal{O}_9 = (\bar{q}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu\mu)$, corresponding to the photon penguin, has a vanishing contribution in the matrix element calculation due to the Ward identity [45]. A model-independent effective Hamiltonian can therefore be written as

$$\mathcal{H}_{eff} = \frac{G_F\alpha_{EM}}{\sqrt{2}\pi} \left\{ V_{ts}^* V_{tq} \sum_{i \in [10, S, P]} (C_i \mathcal{O}_i + C'_i \mathcal{O}'_i) + \text{h.c.} \right\}, \quad (1.22)$$

where the unitarity of V_{CKM} has been exploited, and terms proportional to $V_{ub}V_{us}^*$ have been neglected. The \mathcal{O}'_i operators are obtained from the \mathcal{O}_i by replacing $P_L \leftrightarrow P_R$, where $P_{L,R} \equiv (1 \pm \gamma_5)/2$ denotes the left and right handed chiral projectors.

1.2.2 B^0 mixing and $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fractions

The simplest observable for $B_q^0 \rightarrow \mu^+\mu^-$ processes is the branching fraction (\mathcal{B}), namely the probability of a specific decay to occur or, in other terms, the relative frequency of a specific decay channel.

Experimentally, the branching fraction is extracted from the total event yield ignoring the meson lifetime:

$$\mathcal{B}(B_q^0 \rightarrow \mu^+\mu^-)_{\text{exp}} \equiv \frac{1}{2} \int_0^\infty \langle \Gamma(B_q^0(t) \rightarrow \mu^+\mu^-) \rangle dt, \quad (1.23)$$

where the integrand is the *untagged* (irrespective of the flavour) and time-dependent decay rate. Conversely, in the theory the rate is usually CP-averaged and calculated in the flavour eigenstate basis:

$$\langle \Gamma(B_q^0(t) \rightarrow \mu^+\mu^-) \rangle |_{t=0} = \Gamma(B_q^0(t) \rightarrow \mu^+\mu^-) + \Gamma(\bar{B}_q^0(t) \rightarrow \mu^+\mu^-), \quad (1.24)$$

which leads to the branching fraction definition:

$$\mathcal{B}(B_q^0 \rightarrow \mu^+\mu^-)_{\text{theo}} \equiv \frac{\tau_{B_q}}{2} \langle \Gamma(B_q^0(t) \rightarrow \mu^+\mu^-) \rangle |_{t=0}, \quad (1.25)$$

where τ_{B_q} is the mean lifetime of the B_q^0 meson.

A key feature of the B^0 meson system is *mixing* [46]: quantum-mechanical and time-dependent oscillations between the B^0 and \bar{B}^0 states. Neutral B mesons evolve in time as mass eigenstates (or *physical* eigenstates) which do not correspond to flavour eigenstates (or *interaction* eigenstates), in which they are produced. Flavour eigenstates can be expressed as a superposition of mass eigenstates, which have well-defined mass and lifetime. The time-dependent untagged decay rate R can therefore be written as a sum of two exponentials:

$$R \equiv \langle \Gamma(B_q^0(t) \rightarrow \mu^+\mu^-) \rangle = R_H^{\mu^+\mu^-} e^{-\Gamma_H^q t} + R_L^{\mu^+\mu^-} e^{-\Gamma_L^q t}, \quad (1.26)$$

where the Heavy (H) and Light (L) mass eigenstates have different decay widths:

$$y_q \equiv \frac{\Gamma_L^q - \Gamma_H^q}{2\Gamma_q} \equiv \frac{\Delta\Gamma_q}{2\Gamma_q} \quad (1.27)$$

and $\Gamma_q \equiv \tau_{B_q}^{-1} \equiv (\Gamma_L^q + \Gamma_H^q)/2$. The parameter y measures the difference of the lifetimes between the two mass eigenstates. This difference is measured to be less than 1% in the $B_d^0 - \bar{B}_d^0$ system but sizeable in the $B_s^0 - \bar{B}_s^0$ system,³ for which [39, 47]:

$$y_s = 0.062 \pm 0.006. \quad (1.28)$$

In view of this decay width difference, an expression relating the definitions (1.23) and (1.25) is needed for the B_s^0 [48].

The untagged rate (1.26) can be written as

$$\begin{aligned} \langle \Gamma(B_s^0(t) \rightarrow \mu^+\mu^-) \rangle &= (R_H^{\mu^+\mu^-} + R_L^{\mu^+\mu^-}) \\ &\times e^{-\Gamma_s t} \left[\cosh\left(\frac{y_s t}{\tau_{B_s}}\right) + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} \sinh\left(\frac{y_s t}{\tau_{B_s}}\right) \right], \end{aligned} \quad (1.29)$$

where

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} \equiv \frac{R_H^{\mu^+\mu^-} - R_L^{\mu^+\mu^-}}{R_H^{\mu^+\mu^-} + R_L^{\mu^+\mu^-}}, \quad (1.30)$$

so that the experimentally measurable branching fraction can be converted into the theoretical one through

$$\mathcal{B}(B_s \rightarrow \mu^+\mu^-)_{\text{theo}} = \left[\frac{1 - y_s^2}{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_s} \right] \mathcal{B}(B_s \rightarrow \mu^+\mu^-)_{\text{exp}}. \quad (1.31)$$

The two branching fraction definitions coincide only in the case of a vanishing decay width difference.

The $\mu^+\mu^-$ state is CP odd, so that the SM predicts $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$, i.e. only the heavy mass eigenstate contributes to the $B_s^0 \rightarrow \mu^+\mu^-$ decay. $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ may be moved away from its SM prediction by new physics effects even in the event that the branching fraction agrees with the SM, as explained in Sec. 1.3. Figure 1.6 illustrates Eq. (1.31) for different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$, where differences as large as $\mathcal{O}(10\%)$ may arise [48].

³ The width differences are caused by the existence of final states to which both the B_q^0 and \bar{B}_q^0 mesons can decay to. Such decays involve $b \rightarrow c\bar{c}q$ quark-level transitions, which are Cabibbo suppressed if $q = d$ and Cabibbo-allowed if $q = s$.

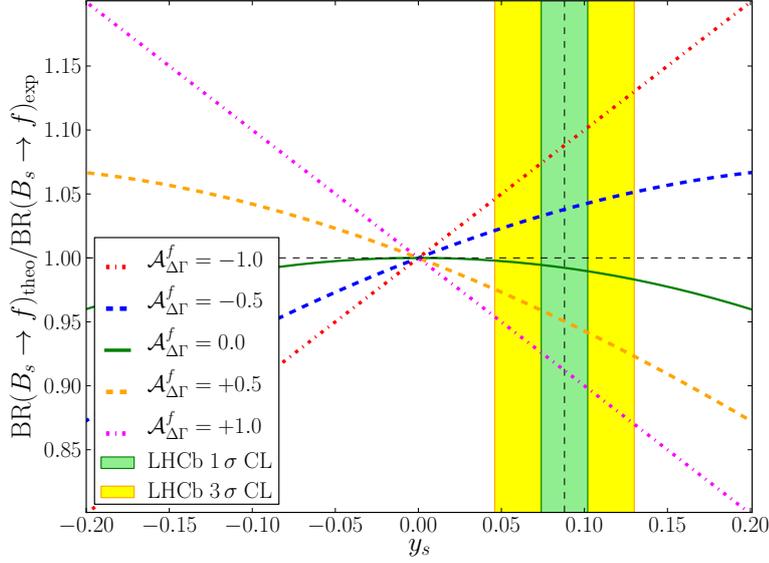


Figure 1.6: Ratio between the theoretical and experimental branching fraction as a function of y_s for different $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ scenarios [48]. The value of y_s is taken from the LHCb measurement published in [49].

1.2.3 $B_s^0 \rightarrow \mu^+\mu^-$ Effective Lifetime

$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ is an observable that can be extracted by measuring the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, defined as the mean decay time of an unbiased sample of $B_s^0 \rightarrow \mu^+\mu^-$ decays

$$\tau_{\mu^+\mu^-} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt} = \frac{\tau_{B_s}}{1 - y_s^2} \left[\frac{1 + 2\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_s + y_s^2}{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_s} \right], \quad (1.32)$$

where t is the proper decay time of the B_s^0 meson. Eq. (1.32) shows that $\tau_{\mu^+\mu^-}$ allows an efficient extraction of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$.

Note that by using the definition (1.32), the relation (1.31) can be rewritten as

$$\mathcal{B}(B_s \rightarrow \mu^+\mu^-)_{\text{theo}} = \left[2 - (1 - y_s^2) \frac{\tau_{\mu^+\mu^-}}{\tau_{B_s}} \right] \mathcal{B}(B_s \rightarrow \mu^+\mu^-)_{\text{exp}}. \quad (1.33)$$

The $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, which has never been measured before, therefore represents a new and interesting observable, “orthogonal” to the branching fraction, from which new physics effects might be spotted.

1.2.4 The Standard Model branching fraction

From the effective Hamiltonian (1.22), the *time-integrated*, *untagged* and *helicity-summed* branching fraction (1.23) can be worked out by evaluating the amplitude (1.20). Within the SM, the only non-negligible contribution to $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays comes from the operator \mathcal{O}_{10} , whose magnitude in the effective Hamiltonian is represented by the real Wilson coefficient C_{10}^{SM} . Scalar (\mathcal{O}_S) and pseudo-scalar (\mathcal{O}_P) contributions are in fact absent in the SM, with the only exception of the Higgs penguin process, which is however negligible due to the smallness of the muon mass. The left-handedness of the charged current also implies that the Wilson coefficients C'_i corresponding to the \mathcal{O}'_i operators are suppressed by $\mathcal{O}(m_q/m_b)$, where $q = d, s$. The SM branching fraction can therefore be expressed as [44]:

$$\begin{aligned} \mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}^{\text{SM}} &= \frac{\tau_{B_q} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} |C_{10}^{\text{SM}} V_{tb} V_{tq}^*|^2 \\ &\times f_{B_q}^2 m_{B_q} m_\mu^2 \sqrt{1 - \frac{4m_\mu^2}{m_{B_q}^2} \frac{1 + y_q}{1 - y_q^2}}, \end{aligned} \quad (1.34)$$

where, as stated in Sec. 1.2.2, the mixing effect correction $(1 + y_q)/(1 - y_q^2)$ is sizeable only in the $B_s^0 \rightarrow \mu^+ \mu^-$ case ($q = s$).

C_{10}^{SM} comprises the contributions from Z penguin and W box diagrams of Fig. 1.4, and has a value of ~ -4.1 [44]. Since Higgs boson couplings are proportional to the fermion masses (Eq. (1.8)), its only substantial contributions are those in which H^0 is coupled at both end of its propagator to the top quark. The main processes for such contributions appear at two-loop level in EW interactions and can be safely neglected [42].

The Hadronic Matrix Element

As the final state of $B_q^0 \rightarrow \mu^+ \mu^-$ is purely leptonic, the hadronic sector of the decay can be expressed in terms of a single non-perturbative *decay constant* f_{B_q} , defined by the matrix element [50]

$$\langle 0 | \bar{q} \gamma_\mu \gamma_5 b | \bar{B}_q(p) \rangle = i p_\mu f_{B_q}, \quad (1.35)$$

which contracted with p^μ on both sides gives

$$\langle 0 | \bar{q} \gamma_5 b | \bar{B}_q(p) \rangle = -i f_{B_q} \frac{M_{B_q}^2}{m_b + m_s}. \quad (1.36)$$

The decay constant used to be the largest source of uncertainty in the amplitude calculation, but recent advances in lattice QCD calculations brought this error

down to the 2% level [51]:

$$\begin{aligned} f_{B_d} &= 190.5 \pm 4.2 \text{ MeV} \\ f_{B_s} &= 227.7 \pm 4.5 \text{ MeV}. \end{aligned} \tag{1.37}$$

Radiative corrections

$B_{d,s}^0 \rightarrow \mu^+ \mu^-$ transitions are inevitably accompanied by real photon emissions, which can generally be distinguished in two types:

1. Initial State Radiation (ISR), or *direct-emission*, represents real photon emission by quarks in the initial state. This component (blue line in Fig. 1.7) vanishes in the limit of small photon energies and represents a background for both the theory and the experiment. Experimentally, it is neglected within the signal window (vertical green lines in Fig. 1.7), while it is just excluded in the theoretical branching fraction by definition.
2. Final State Radiation (FSR), or *bremstrahlung*, denotes photon emission by muons in the final state. FSR (red line in Fig. 1.7) is largely dominant for small photon energies and, in the soft-photon approximation ($E_{max} \ll m_{B_q}/2$), leads to a multiplicative correction factor to the non-radiative rate $\mathcal{B}^{(0)}$ (Eq. (1.38)) [52].

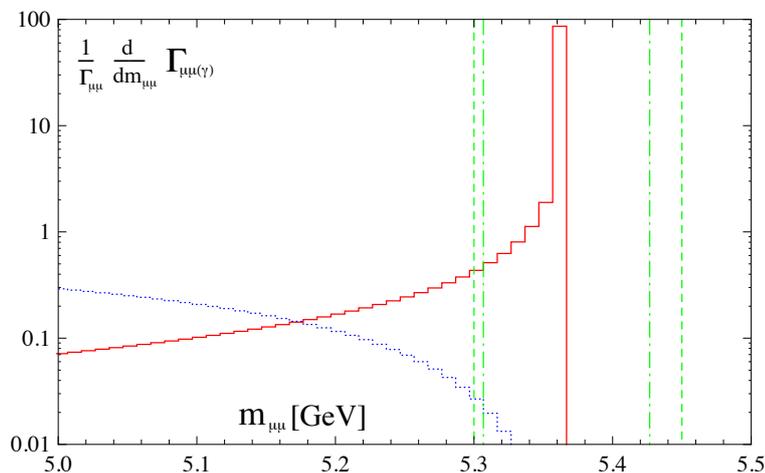


Figure 1.7: ISR (dotted blue line) and FSR (solid red line) contributions to the dimuon invariant mass spectrum in $B_{d,s}^0 \rightarrow \mu^+ \mu^- + n\gamma$. The vertical green lines indicate the CMS (dashed) and LHCb (dot-dashed) signal windows [53].

To treat the soft-photon radiation, a branching fraction can be defined to account for an arbitrary number n of undetected photons with a total energy in the B_q rest frame not larger than E_{max} :

$$\begin{aligned}\mathcal{B}(B_{d,s}^0 \rightarrow \mu^+ \mu^-) &= \mathcal{B}(B_{d,s}^0 \rightarrow \mu^+ \mu^- + n\gamma) |_{\sum E_\gamma < E_{max}} \\ &= \mathcal{B}^{(0)}(B_{d,s}^0 \rightarrow \mu^+ \mu^-) \times \omega(E_{max}).\end{aligned}\quad (1.38)$$

For $E_{max} = 60$ MeV, the correction factor yields $\omega(60 \text{ MeV}) \approx 0.89$, which corresponds to a $\approx 11\%$ suppression of the non-radiative rate up to corrections of order 1%. On the experimental side, both CMS and LHCb use PHOTOS [54] to simulate photon emission, and the corresponding correction is included in the signal detection efficiency.

Numerical result

The most precise branching fraction prediction in the SM yields [53]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{theo}} = (3.65 \pm 0.23) \times 10^{-9}, \quad (1.39)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}. \quad (1.40)$$

This result profits from recent two-loop EW corrections [55] and three-loop QCD corrections [56] to C_{10}^{SM} , which significantly brought down its uncertainty as well as the normalisation scheme dependence of the matching calculation. Considering the increased precision on the decay constant computation (Eq. (1.37)), the largest source of uncertainty now arises from the CKM parameters, accounting for 4.3% and 6.9% relative uncertainty in the (1.39) and (1.40), respectively, with $|V_{cb}|$ being one of the main limiting factors for the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ precision.

A theoretically cleaner quantity that can be built is the ratio between the B_d and B_s decay modes [4]:

$$\frac{\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{theo}}} = \frac{\tau_{B_d} m_{B_d} f_{B_d}^2}{\tau_{B_s} m_{B_s} f_{B_s}^2} \left| \frac{V_{td}}{V_{ts}} \right|^2 = 0.0295^{+0.0028}_{-0.0025}, \quad (1.41)$$

being free from the uncertainties stemming from the Wilson coefficient calculations. The (1.41) also holds in NP models where the flavour interaction retains the SM structure, and therefore constitutes a powerful observable to test such mechanism, as discussed in Sec. 1.3.

1.3 $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ and new physics models

The model-independent expression for the $B_q^0 \rightarrow \mu^+ \mu^-$ branching fraction can be written as [44]:

$$\begin{aligned} \mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} &= \frac{\tau_{B_q} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} |C_{10}^{\text{SM}} V_{tb} V_{tq}^*|^2 f_{B_q}^2 m_{B_q} m_\mu^2 \\ &\times \sqrt{1 - \frac{4m_\mu^2}{m_{B_q}^2}} \times (|P|^2 + |S|^2) \times \frac{1 + y_q \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}}{1 - y_q^2}, \end{aligned} \quad (1.42)$$

where the SM contribution C_{10}^{SM} has been factorised out thanks to the Wilson coefficient combinations:

$$\begin{aligned} P &= \frac{C_{10} - C'_{10}}{C_{10}^{\text{SM}}} + \frac{m_{B_q}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_s} \right) \left(\frac{C_P - C'_P}{C_{10}^{\text{SM}}} \right) \equiv |P| e^{i\varphi_P}, \\ S &= \sqrt{1 - \frac{4m_\mu^2}{m_{B_q}^2}} \frac{m_{B_q}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_s} \right) \left(\frac{C_S - C'_S}{C_{10}^{\text{SM}}} \right) \equiv |S| e^{i\varphi_S}. \end{aligned} \quad (1.43)$$

With these definitions, the coefficients C_{10} and C'_{10} are dimensionless, while C'_S and C'_P have dimensions of GeV^{-1} . The corresponding expression for the $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ observable is

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = \frac{|P|^2 \cos(2\varphi_P - \phi_s^{\text{NP}}) - |S|^2 \cos(2\varphi_S - \phi_s^{\text{NP}})}{|P|^2 + |S|^2}, \quad (1.44)$$

where the phase ϕ_s^{NP} represents the CP-violating contribution due to new physics in the $B_s^0 - \bar{B}_s^0$ mixing. Unlike the branching fraction, the f_{B_q} dependence cancels in $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$, which is also not affected by CKM uncertainties. As a result, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ is theoretically clean and independent from the ratio of fragmentation fractions f_s/f_d , which is the major limitation on the precision of $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction measurement at hadronic colliders (see Chapter 6).

Through the effective theory formulation, it is clear from Eq. (1.42) how sensitive these decays are to any new scalar (S) and pseudo-scalar (P) contributions. Effective theory description of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ processes is also particularly suitable to describe new physics affecting the branching fraction or $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. To this end, it is useful to introduce the ratio between the experimental and theoretical $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction [57]:

$$\begin{aligned} R &\equiv \frac{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{theo}}^{\text{SM}}} = \left[\frac{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} y_s (|P|^2 + |S|^2)}{1 - y_s^2} \right] = \\ &= \left[\frac{1 + y_s \cos(2\varphi_P - \phi_s^{\text{NP}})}{1 + y_s} \right] |P|^2 + \left[\frac{1 - y_s \cos(2\varphi_S - \phi_s^{\text{NP}})}{1 + y_s} \right] |S|^2, \end{aligned} \quad (1.45)$$

which is derived from the (1.42) and (1.34) with the definitions (1.31) and (1.43). A variety of NP models allow R to stray from $R_{\text{SM}} = 1$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ from $+1$ by affecting the Wilson coefficients P, S as well as their phases φ_P and φ_S . Measuring the branching fraction and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ therefore allows to put stringent limits to such models, given their strong dependence on pseudo-scalar (P) and scalar (S) operators highlighted in Eqs. (1.42) and (1.44).

Following [44], NP models can be categorised according to their structure in terms of P and S . This allows to reduce the number of free parameters and highlights the generic features of the models, which directly translates in distinct phenomenologies for the observables R and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$, as shown in Fig. 1.8.

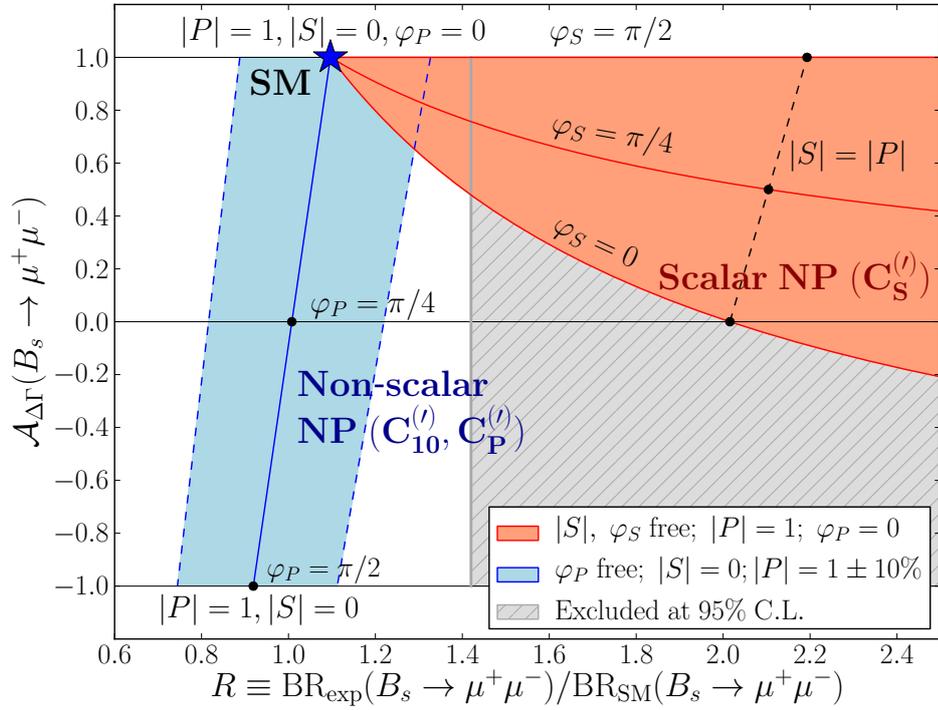


Figure 1.8: Illustration of allowed regions in the $R - \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ plane for scenarios with scalar or non-scalar NP contributions [57].

1.3.1 Non CP-violating new physics ($\varphi_P, \varphi_S \in \{0, \pi\}$)

When there are no CP violating phases in the $B_s^0 \rightarrow \mu^+\mu^-$ decay (φ_P and φ_S are either 0 or π), S and P are allowed to take arbitrary but *real* values. The

observables R and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ read in this case:

$$R = |P|^2 \left[\frac{1 + y_s \cos \phi_s^{\text{NP}}}{1 + y_s} \right] + |S|^2 \left[\frac{1 - y_s \cos \phi_s^{\text{NP}}}{1 + y_s} \right] \quad (1.46)$$

and

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = \cos \phi_s^{\text{NP}} \left[\frac{|P|^2 - |S|^2}{|P|^2 + |S|^2} \right]. \quad (1.47)$$

Fig. 1.9 shows how their measurement can be used to constrain the parameters S and P in the case $\phi_s^{\text{NP}} = 0$.⁴

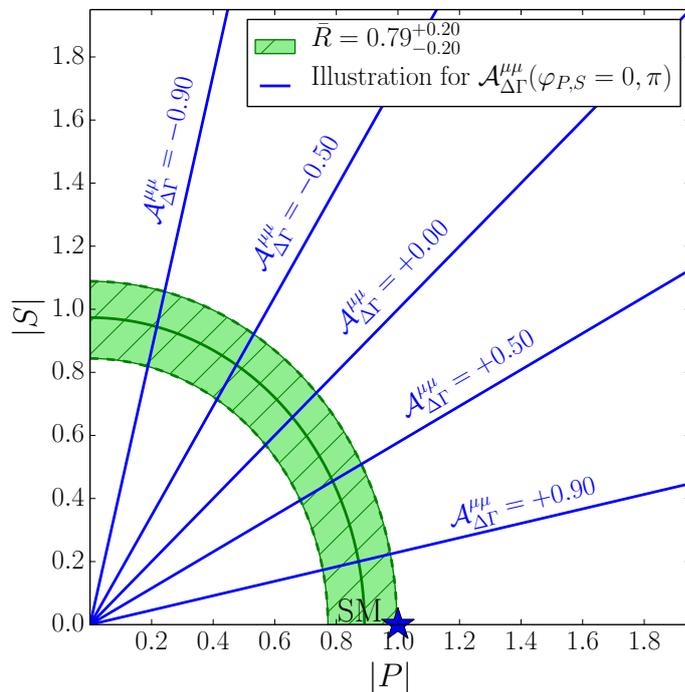


Figure 1.9: The relation between P and S for different $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ cases [59]. The measured R is taken from the combination of CMS [60] and LHCb [61] results. The observable \bar{R} follows the definition (1.45).

Minimal Flavour Violation (MFV)

In MFV models, flavour interactions and CP violation are described by the CKM matrix through the known Yukawa terms (1.8), i.e. there are not any additional

⁴Such assumption is justified by the well measured $\sin \phi_s$, which is known to be small [58].

structures besides the SM ones [62]. These prescriptions still allow for additional higher-dimensional operators [63], which, conversely, are not included in the so-called *constrained* (cMFV) models [62, 64]. For cMFV models, all Wilson coefficients are zero with the exception of C_{10} , which is real. New physics therefore enters only in the ratio R :

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1, \quad R = |P|^2. \quad (1.48)$$

The ratio (1.41) also provides a powerful and clean observable that can be used to test the MFV assumption, as $|V_{td}/V_{ts}|$ can be directly accessed without knowledge on the Wilson coefficients.

1.3.2 Pseudo-scalar dominated new physics ($S = 0$)

In this scenario $C_S - C'_S = 0$, while C'_{10} and C'_P are free to depart from zero. As any model with new gauge bosons or pseudo-scalars is contemplated, a number of popular NP models fall into this category. Note that also models with scalars can qualify if their quark coupling is left-right symmetric, so that $C_S = C'_S$.

R and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ simplify to

$$R = |P|^2 \left[\frac{1 + y_s \cos(2\varphi_P - \phi_s^{\text{NP}})}{1 + y_s} \right] \quad (1.49)$$

and

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = \cos(2\varphi_P - \phi_s^{\text{NP}}). \quad (1.50)$$

Z' Model

The simplest extension to the Standard Model that introduces new flavour interactions as well as CP violation sources is the addition of a $U(1)$ gauge symmetry to the SM gauge group. If the resulting heavy gauge boson, named Z' , can mediate FCNC processes at tree-level (like in the Feynman diagram of Fig. 1.4a), a rich pattern of deviations from the SM expectations emerges, which only depend on Z' coupling to fermions and on its mass [65]. For $M_{Z'} < 3 \text{ TeV}/c^2$, sizeable deviations from the SM occur, while if the scale is higher, $M_{Z'} > 5 \text{ TeV}/c^2$, Z' effects in rare B_d and B_s decays are typically below 10%. Large effects on the $B_s^0 \rightarrow \mu^+\mu^-$ decay are already ruled out by the latest branching fraction measurement [4].

Randall-Sundrum Model (RS)

With the aim of solving the gauge hierarchy problem, i.e. the large imbalance between EW and gravitational forces, L. Randall and R. Sundrum proposed a

model [66] in which the addition of a warped extra dimension reduces the *effective* 4-dimensional Planck scale, $M_{Planck} = 2 \times 10^{18}$ GeV, to the scale of EW interactions. In the RS model with custodial protection, NP contributions to $B_s^0 \rightarrow \mu^+ \mu^-$ are dominated by right-handed flavour violating couplings of the Z boson to quarks, $Z d_R^i \bar{d}_R^j$, resulting in departures of the branching fraction from the SM values at most of order 15% [67]. The custodial P_{LR} symmetry suppresses the tree level Z boson couplings to left-handed quarks ($Z d_L^i \bar{d}_L^j$) which would otherwise have been dominant.

Four Generation Model (SM4)

Even though LEP measurements confirmed that only three neutrino families with masses not larger than $M_Z/2$ are present, the existence of a fourth family of fermions is not yet completely ruled out. The only sources of flavour symmetry breaking are the Yukawa couplings, which are represented by 4×4 matrices [68]. The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction being already strongly constrained [4], an enhancement of the $B_d^0 \rightarrow \mu^+ \mu^-$ mode could point to SM4 (Fig. 1.10).

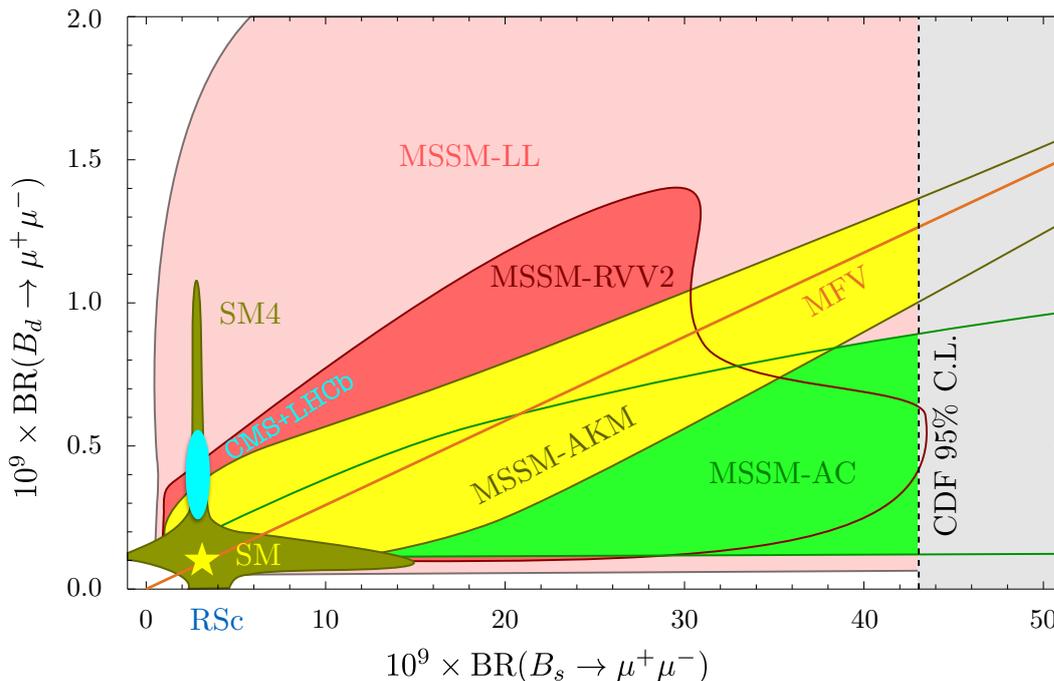


Figure 1.10: $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ branching fraction predictions in MFV, RS, SM4 and several MSSM models [69]. The grey area on the right shows the CDF measurement [70] before the LHC results, while the CMS+LHCb measurement [4] is superimposed in cyan.

1.3.3 Scalar dominated new physics ($P = 1$)

Complementary to the $S = 0$ class is the $P = 1$ case, where only scalar operators drive new physics effects. Its simplest realisation is $C_{10} = C_{10}^{\text{SM}}$ and $C'_{10} = C'_P = 0$, but also pseudoscalars that couple left-right symmetrically to quarks can participate, realising $C_P = C'_P$ or even $C_{10} - C'_{10} = C_{10}^{\text{SM}}$.

In this scenario the observables read

$$R = \left[\frac{1 + y_s \cos(2\varphi_P - \phi_s^{\text{NP}})}{1 + y_s} \right] + |S|^2 \left[\frac{1 - y_s \cos(2\varphi_S - \phi_s^{\text{NP}})}{1 + y_s} \right], \quad (1.51)$$

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = \frac{\cos \phi_s^{\text{NP}} - |S|^2 \cos(2\varphi_S - \phi_s^{\text{NP}})}{1 + |S|^2}. \quad (1.52)$$

Note that from Eq. (1.43), while pseudoscalar contributions might both enhance or reduce the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fractions, scalar contributions can only lead to an increase of the observed decay rate.

Scalar Dominance Model

This class includes every model in which new physics is dominated by tree-level FCNC contribution of a scalar. As explored in [71], heavy neutral scalars give raise to new sources of flavour and CP violations, as well as left- and right-handed currents, where the deviations from the SM for FCNC processes depend only on the couplings of the scalars to fermions and on their masses. While Z' masses already have a lower bound of 1-2 TeV/ c^2 , new neutral scalars with masses of few hundred GeV/ c^2 are not yet excluded [71]. Note also that in Z' models, operators with modified Wilson coefficients can participate, while for these models all operators are new. Concerning flavour violating couplings of the SM Higgs, given the smallness of its coupling to muons, the effects on $B_s^0 \rightarrow \mu^+\mu^-$ are small but can still reach the level of 8% [71].

1.3.4 Mixed scalar/pseudo-scalar new physics ($P \pm S = 1$)

In the last scenario considered here, new physics affecting S and P are on the same footing. By defying $P = 1 + \tilde{P}$, all NP contributions are contained into \tilde{P} , and the relation $P \pm S = 1$ reads in fact $\tilde{P} = \mp S$. Such condition is fulfilled if $C'_S = \pm C'_P$, neglecting C'_{10} contributions and m_μ/m_{B_s} in the (1.43).

The observables now read:

$$R = \frac{1 \mp 2|S| \cos \varphi_S + 2|S|^2 + y_s [\cos \phi_s^{\text{NP}} \mp 2|S| \cos(\varphi_S - \phi_s^{\text{NP}})]}{1 + y_s} \quad (1.53)$$

and

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = \frac{\cos\phi_s^{\text{NP}} \mp 2|S| \cos(\varphi_S - \phi_s^{\text{NP}})}{1 \mp 2|S| \cos\varphi_S + 2|S|^2}. \quad (1.54)$$

Two Higgs-Doublet Model (2HDM)

The SM Higgs mechanism is based on one Higgs doublet and provides the simplest description of the electroweak symmetry breaking (EWSB). However, the resulting neutral Higgs boson with SM couplings might be the lightest scalar of a two-Higgs doublet, whose vacuum expectation values, v_1 and v_2 , define the well-known free parameter β of the 2HDM, through the relation $\tan\beta = v_2/v_1$. In the decoupling limit, i.e. when the mass of the lightest Higgs boson is significantly lower than the masses of the other Higgs bosons ($m_h \ll \Lambda_{\text{2HDM}}$), the conditions $C_S = -C_P$ and $C'_S = C'_P$ are realized, which correspond to the $P \pm S = 1$ class of models when the couplings of the heavy Higgs bosons are not left-right symmetric, as the MFV hypothesis wants. There are different types of 2HDMs, categorised according to the Higgs boson couplings to fermions. As an example, in type-II models the neutral member of one Higgs doublet couples only to up-type quarks and the neutral member of the other doublet couples only to down-type quarks and leptons. In this case, tree-level Higgs mediated FCNC are absent and [50]

$$C_S = -C_P \propto \tan^2\beta \frac{\ln m_{H^\pm}^2/m_t^2}{1 - m_{H^\pm}^2/m_t^2}, \quad (1.55)$$

hence the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fractions can be substantially enhanced as they depend on $\tan^4\beta$ and do not suffer helicity suppression [72]. In type-III models, the most general Higgs-fermion Yukawa couplings are contemplated, allowing for tree-level FCNC which are however suppressed in the decoupling limit [73]. An example of neutral and charged Higgs mediated $B_s^0 \rightarrow \mu^+\mu^-$ decay is shown in Fig. 1.11a.

Minimal Supersymmetric Standard Model (MSSM)

A *supersymmetry* (SUSY) is a transformation that turns a bosonic state into a fermionic state and vice-versa. One of main issue that SUSY model addresses is the Higgs mass computation, which in the SM implies large cancellations between corrections stemming from any particle that couples to the Higgs field [74]. In fact, if a supersymmetry is introduced, a systematic cancellation of such corrections occurs because of the relative minus sign between fermion and boson loop contributions to m_H , thus stabilising its value. In the minimal supersymmetric extension of the Standard Model (MSSM), each one of the known fundamental particles is a member of a *supermultiplet*, and must therefore have a *superpartner*

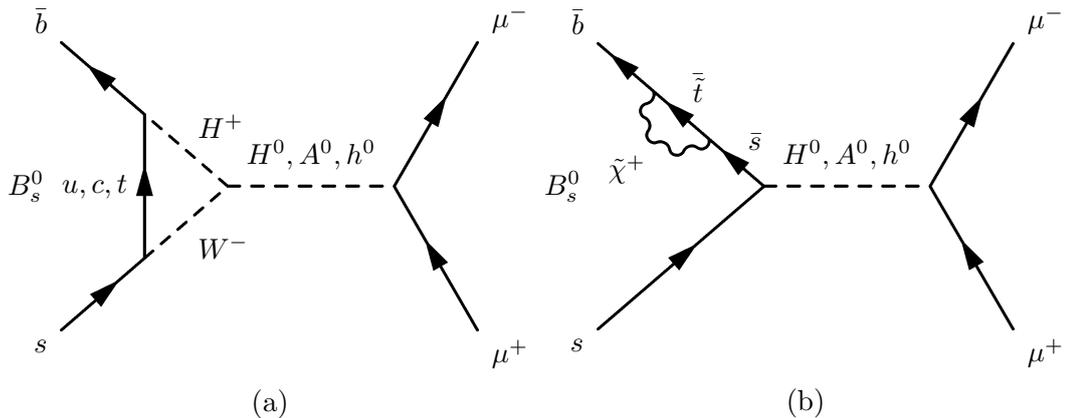


Figure 1.11: (a) Higgs mediated $B_s^0 \rightarrow \mu^+\mu^-$ process in the 2HDM. (b) Higgs mediated $B_s^0 \rightarrow \mu^+\mu^-$ decay in SUSY.

with a spin differing by $1/2$ unit. As all members in a supermultiplet must have the same mass [74], the supersymmetry is clearly broken. The MSSM comprises two Higgs doublets, where the heavy Higgs bosons have mass m_A , thus allowing flavour violating couplings analogously to the aforementioned 2HDMs [75] (Fig. 1.10). A typical $B_s^0 \rightarrow \mu^+\mu^-$ process in SUSY is shown in Fig. 1.11b: in this class of models the branching fraction is proportional to $\tan^6 \beta$, hence a strong enhancement is foreseeable. A plot showing the constraint due to the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction measurement in the plane $m_A - \tan \beta$ is shown in Fig. 1.12. As it is visible, even the first $B_s^0 \rightarrow \mu^+\mu^-$ observation by LHCb [76] wiped out an extensive region of this parameter space.

1.4 Current experimental status

The history of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ measurements is summarised in Fig. 1.13. The LHCb experiment was the first to report an evidence⁵ for the $B_s^0 \rightarrow \mu^+\mu^-$ decay in 2012 [76], by analysing 1.0 fb^{-1} of LHC data collected at a centre-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$ in 2011 plus 1.1 fb^{-1} collected at $\sqrt{s} = 8 \text{ TeV}$ in 2012. In 2013, both LHCb and CMS published their results obtained with the full Run 1 data of LHC, 3.0 fb^{-1} and 25 fb^{-1} , respectively, reaching the same level of precision. The combined analysis of CMS and LHCb data, performed in 2014, led to the first observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay and the resulting branching fraction

⁵As a convention, a statistical significance of 3σ denotes an *evidence*, while more than 5σ are needed to claim for an *observation*.

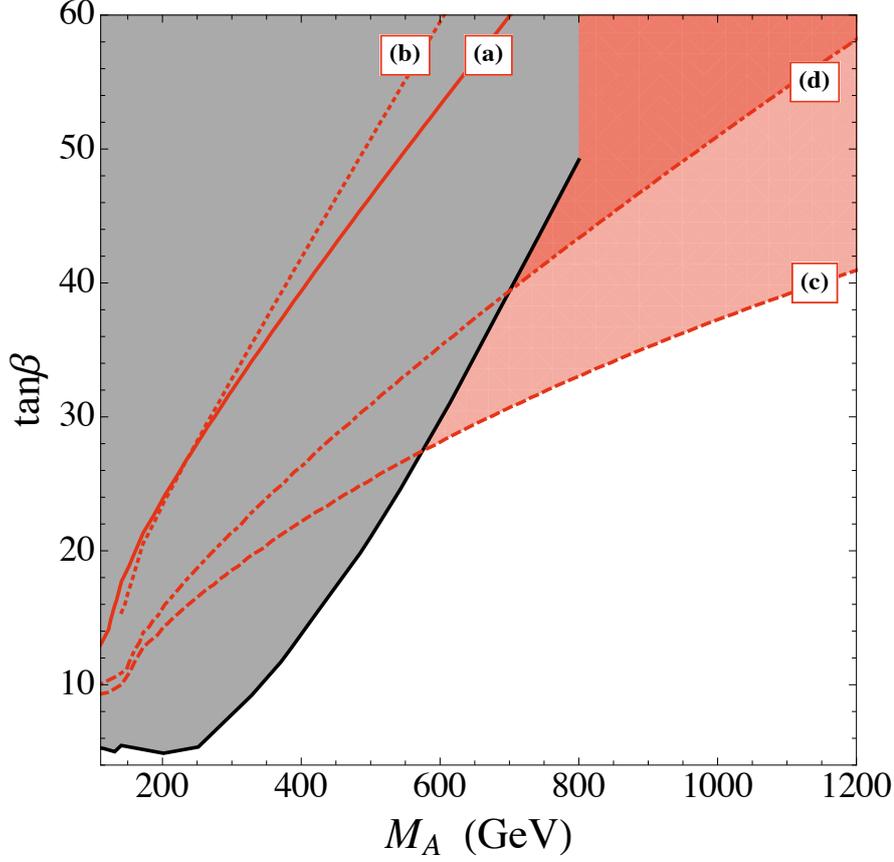


Figure 1.12: Constraints in the $m_A - \tan\beta$ plane from the first observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay at LHCb [76]. The grey area is excluded by direct searches of $\tau^+\tau^-$ resonances, while the red shaded areas marked with the letters $a - d$ correspond to exclusions of the different MSSM scenarios considered in [77].

measurement was very close to the SM prediction [4]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.8_{-0.6}^{+0.7}) \times 10^{-9} \quad (1.56)$$

with a statistical significance of 6.2σ . An evidence for the $B_d^0 \rightarrow \mu^+\mu^-$ decay was also reported with 3.2σ of statistical significance:

$$\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) = (3.9_{-1.4}^{+1.6}) \times 10^{-10}. \quad (1.57)$$

In 2016, the ATLAS collaboration also reported a measurement using 25 fb^{-1} of data collected during the Run 1 of LHC [79]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (0.9_{-0.8}^{+1.1}) \times 10^{-9} \quad (1.58)$$

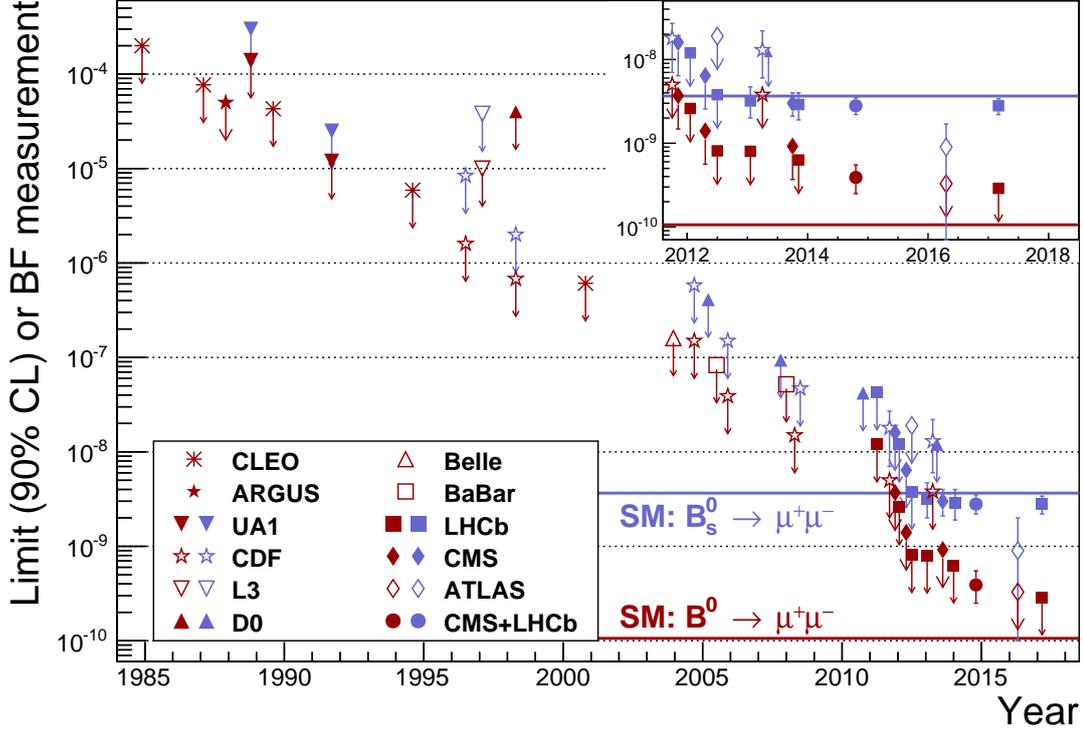


Figure 1.13: History of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ limits and measurements across the years [78].

with a statistical significance of 1.4σ , while the B_d mode of the decay was limited to

$$\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10} \quad (95\% \text{ CL}). \quad (1.59)$$

These experimental measurements as well as the theoretical prediction [53] are shown together in Fig. 1.14, which represents the “state of the art” before the LHCb analysis reported in this thesis. While the general picture is the one of an overall agreement between SM predictions and measurements, one might claim a mild tension between the expected and measured $B_d^0 \rightarrow \mu^+\mu^-$ branching fraction and another one between the ATLAS $B_s^0 \rightarrow \mu^+\mu^-$ measurement against the SM one. To address this debate, a more precise measurement is needed. With the data available from Run 2 of LHC, and in the context of an optimised analysis aimed to background reduction, a new LHCb analysis is presented.

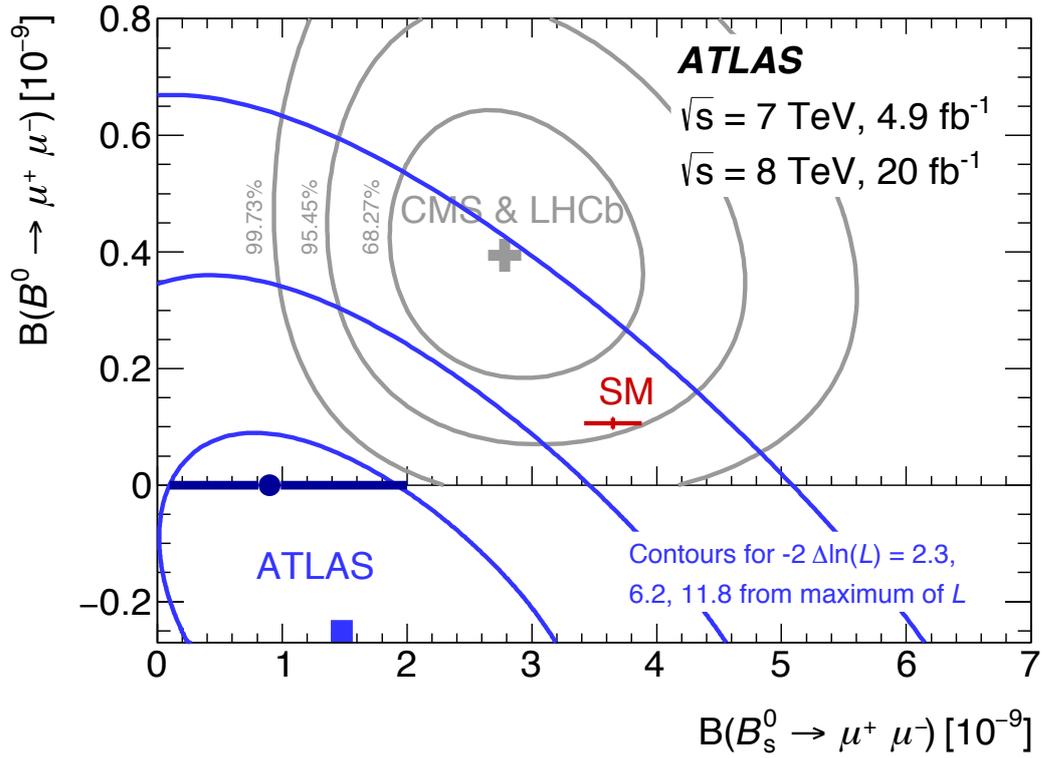


Figure 1.14: The experimental and theoretical picture of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements as of 2016 [79]. In red the SM prediction [53], in blue the ATLAS measurement [79] and in grey the combined CMS+LHCb one [4], together with their respective confidence ranges.

Chapter 2

The LHCb experiment

The LHCb detector [80] is designed to study the physics of b and c quarks arising from proton-proton collisions produced with the Large Hadron Collider (LHC) at CERN, now operating at a centre-of-mass energy of 13 TeV. The LHCb collaboration [81] counts more than 1200 members across 72 institutes from 16 countries. Among the key elements of the LHCb physics program there are the study of neutral B meson oscillations, CP violation and rare b -hadron decays, with the common goal of finding any hint of physics beyond the Standard Model. The study of the rare $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays is one of the flagship measurements of LHCb, actually being one of the reasons for which the detector was designed.

A description of the LHC and the LHCb detector is given in the following, focussing on the aspects which are of particular relevance for the presented analysis.

2.1 The Large Hadron Collider

The LHC [82] is a 26.7 km-long superconductive synchrotron located near Geneva, running across the French and Swiss border and installed inside a 100 m underground tunnel, which was originally built to host the Large Electron-Positron collider (LEP). The machine comprises two accelerating rings where hadron beams circulate in opposite directions and collide into four *Intersection Points*, where LHCb and the three other main CERN experiments are installed:

1. ATLAS [83] (A Toroidal Lhc ApparatuS) is a multi-purpose detector for Higgs physics and NP direct searches;
2. CMS [84] (Compact Muon Solenoid) is a multi-purpose detector that complements and competes with ATLAS;
3. ALICE [85] (A Large Ion Collider Experiment) is specifically designed to study heavy nuclei collisions.

The LHC design energy in the centre-of-mass is 14 TeV for proton collisions, while 2.76 TeV per nucleon is achieved in the case of lead ion collisions, yielding a total energy in the centre-of-mass of 1.15 PeV. During Run 1, LHC operated at energies of 7 and 8 TeV, while it is currently running at 13 TeV for Run 2.

The CERN accelerator complex

The LHC represents the last element in the chain of the CERN accelerator complex, sketched in Fig. 2.1. The source of protons is provided by gaseous hydrogen, from which electrons are stripped off by means of an electric field. These protons are accelerated to an energy of 50 MeV by the linear accelerator Linac 2, which also focuses the beam by using quadrupole magnets. The beam is then injected into the Proton Synchrotron Booster (PSB), made up of four synchrotron rings, that accelerates it to 1.4 GeV, the energy at which the beam is transferred to the Proton Synchrotron (PS). The PS pushes the beam to 25 GeV and also forms proton bunches with a time spacing of 25 ns, where a single bunch typically contains $\sim 10^{11}$ protons. Proton bunches are then sent to the Super Proton Synchrotron (SPS), a 7 km-long accelerator ring that ramps up their energy to 450 GeV before they can finally be fed into the LHC.

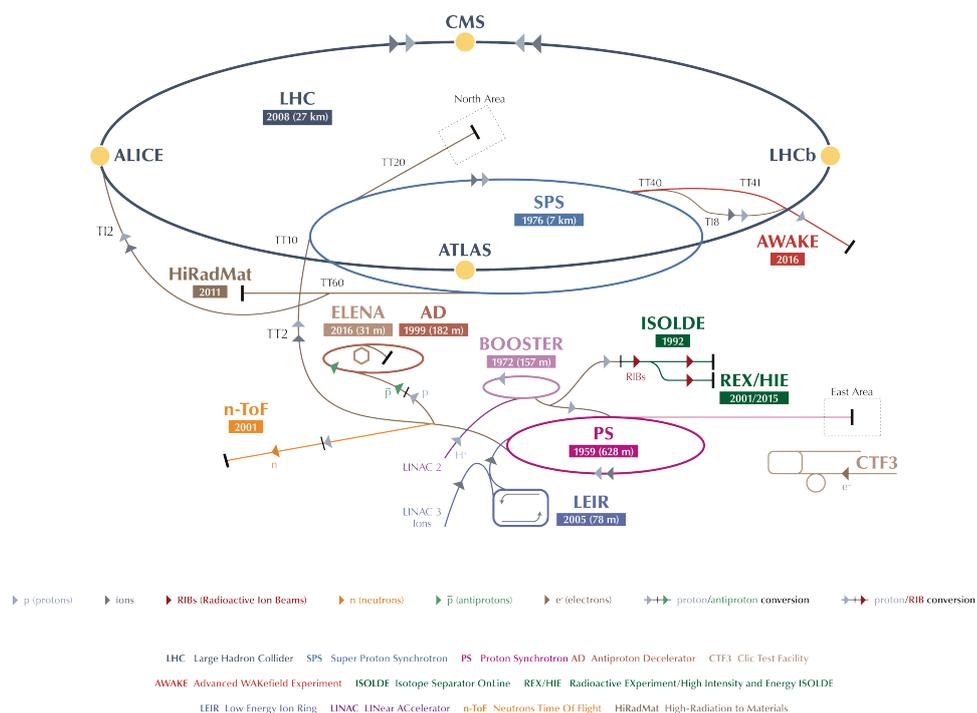


Figure 2.1: The CERN accelerator complex [86].

The luminosity

From the physics viewpoint, the interesting quantity is the rate at which a certain process occurs during collisions, which is given by the cross section¹ of the process times the *instantaneous luminosity* of the accelerator:

$$\dot{N} = \sigma \times \mathcal{L}. \quad (2.1)$$

By using machine parameters, the instantaneous luminosity for head-on collisions occurring at a frequency f can be written as [87]

$$\mathcal{L} = \frac{N_p^2 n_b f}{4\pi\sigma_x\sigma_y}, \quad (2.2)$$

where N_p stands for the number of protons contained in a single bunch and n_b is the number of colliding bunches. The quantity at the denominator represents the effective transverse area where the collisions take place, which can be evaluated from the overlap integral of the two transverse beam spatial distributions. For equal gaussian beams the result reduces to $4\pi\sigma_x\sigma_y$. The design luminosity of LHC, $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, has been recently surpassed by reaching $1.58 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [88].

By integrating Eq. (2.1) over time, the number of events occurring for a specific process can be expressed by its cross section times the *integrated luminosity*, i.e. the integral of the instantaneous luminosity over the period in which collisions took place:

$$N = \sigma \times \int_t \mathcal{L} = \sigma \times \mathcal{L}_{int}. \quad (2.3)$$

During regular operations, the beam intensity degrades over time due to collisions but also, for example, because of proton scattering on residual gas in the beam pipe. As a consequence, the rate of collisions decays with a lifetime of $\mathcal{O}(10 \text{ h})$, and, when it becomes too low, the beam is dumped on an absorber and a new *fill* is prepared.

Due to the particular geometry of the LHCb detector (Sec. 2.2), the luminosity at point 8, where the LHCb cavern is located, is kept below the maximum deliverable from LHC by tuning the transverse separation between the beams, also to prevent radiation damage of the detector elements closer to the beam pipe. To this end, a luminosity levelling technique [89] is used, for which the beams are progressively brought closer to each other in the transverse plane, so that the rate of collisions is almost constant over the beam lifetime and the luminosity is set to a level around $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, as shown in Fig. 2.2. For Run 2, the detector is now operating

¹The cross section (σ) has the dimensions of an area and is typically measured in *barns*, where $1 \text{ b} = 10^{-24} \text{ cm}^2$.

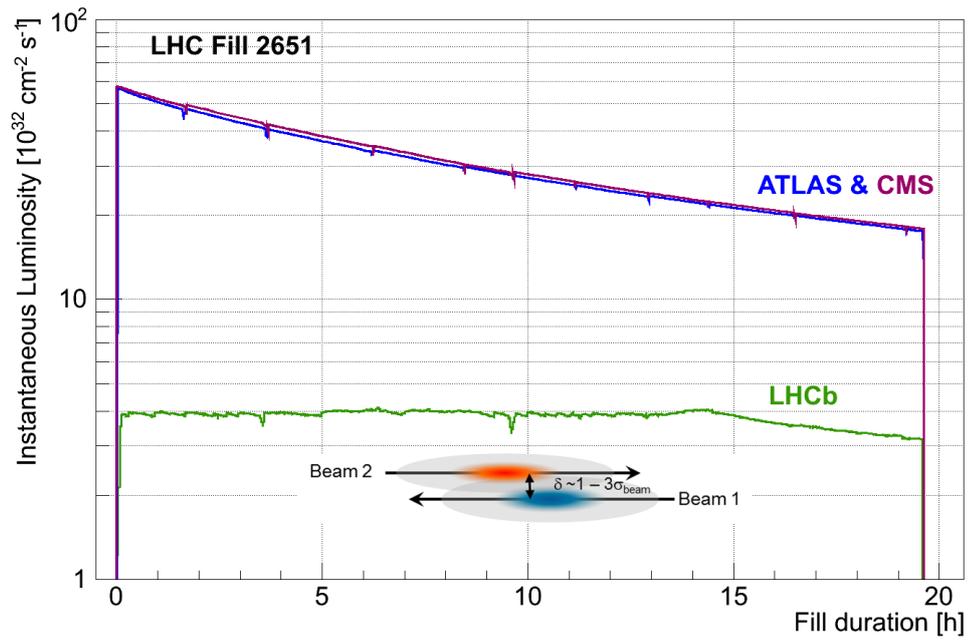


Figure 2.2: Development of the instantaneous luminosity for ATLAS, CMS and LHCb during a typical LHC fill in Run 1 [90]. On the lower side of the figure a cartoon shows how the LHCb luminosity is levelled by adjusting the transverse beam overlap for about 15 hours, after which the beams are colliding head-on. After almost 20 h, the beam is eventually dumped.

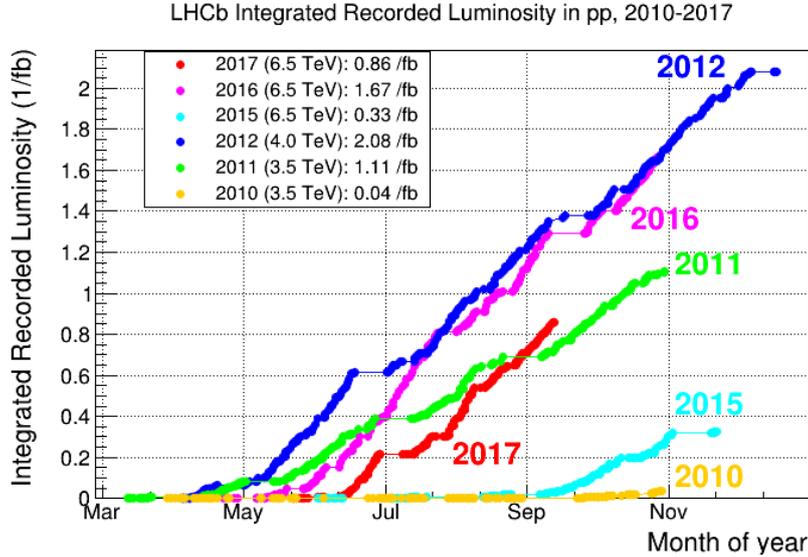


Figure 2.3: Integrated luminosity collected at LHCb per year [92].

at a luminosity of $4.4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, almost the double of the design value. The LHCb experiment measured the LHC luminosity with a precision of $\sim 1\%$ [91], which is the most precise luminosity measurement achieved so far at a bunched-beam hadron collider.

The total integrated luminosity collected with the LHCb detector is summarised in Fig. 2.3.

2.1.1 b physics at the LHC

When two protons collide, they can undergo an *elastic* or an *inelastic* scattering. While in the former case the kinetic energy is conserved, in the latter the internal degrees of freedom of the proton are excited, eventually leading to its fragmentation. The elementary quarks and gluons that build up protons follow the so-called *parton density functions* (*pdfs*), which depend on the fraction of the proton momentum they carry, x , and on the scale of the interaction process, Q^2 , i.e. the transferred momentum. Following the S.D. Drell and T.-M. Yan factorization theorem [93], the cross section of a process occurring in a proton-proton *hard* scattering, i.e. which involves its constituents, can be written as the product of

the partonic cross section and their *pdfs*, integrated over x :

$$\sigma_{AB} = \int dx_a dx_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma}_{ab \rightarrow X}. \quad (2.4)$$

A representation of the process is sketched in Fig. 2.4a.

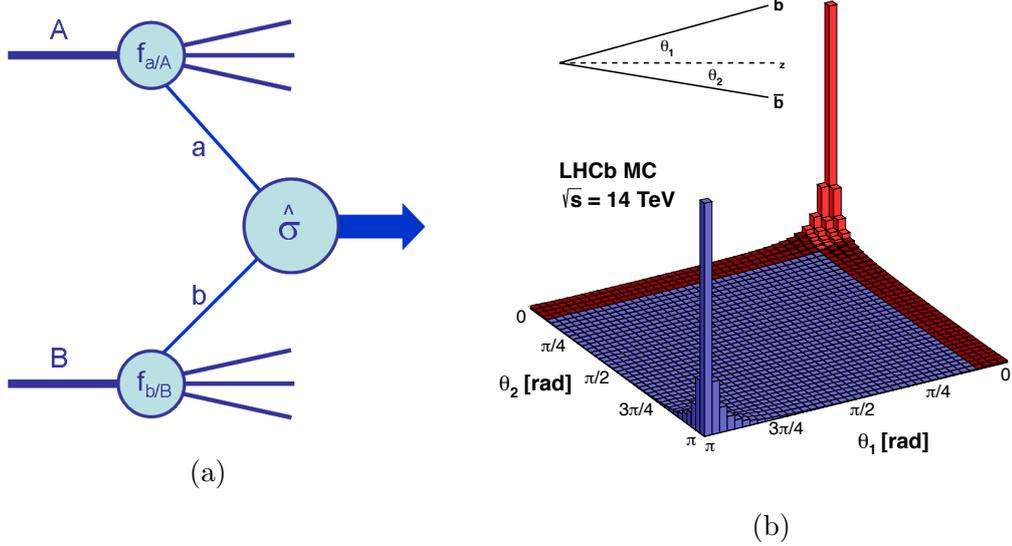


Figure 2.4: (a) Diagram of two interacting partons a and b inside the hard-scattered protons A and B [94]. (b) Polar angular distribution of the $b\bar{b}$ pairs produced at LHC for $\sqrt{s} = 14$ TeV as simulated with the PYTHIA8 event generator [95, 96]. The acceptance of the LHCb detector is highlighted in red.

The partonic cross section $\hat{\sigma}$ represents the elementary process and can be calculated perturbatively in QCD theory, while the partonic *pdfs* as a function of x need to be experimentally probed, usually in deep inelastic electron-proton scattering. The Q^2 evolution of the *pdfs* is determined by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [97], so that starting from an experimentally measured value $f(x, Q_0^2)$, the value at the desired scale can be calculated. The total and inelastic cross sections at LHC have been measured by the TOTEM experiment to be 98.0 ± 2.5 mb and 72.9 ± 1.5 mb, respectively [98]. A very mild energy dependence of the inelastic cross section is expected, as measured by ATLAS at 13 TeV [99].

b -hadron production

Three processes dominate the hadronic beauty production [100]:

- **pair creation** occurs in hard QCD scattering at leading order [101] as in Fig. 2.5a. The corresponding $q\bar{q}$ annihilation is less significant as the gluon $pdfs$ are dominant at the LHC energies [102].
- **flavour excitation** occurs when a b quark from one proton is excited on mass shell by scattering against a parton of the other proton, as shown in Fig. 2.5b.
- **gluon splitting** is when a $g \rightarrow b\bar{b}$ branching occurs in the initial or final state shower, as depicted in Fig. 2.5c.

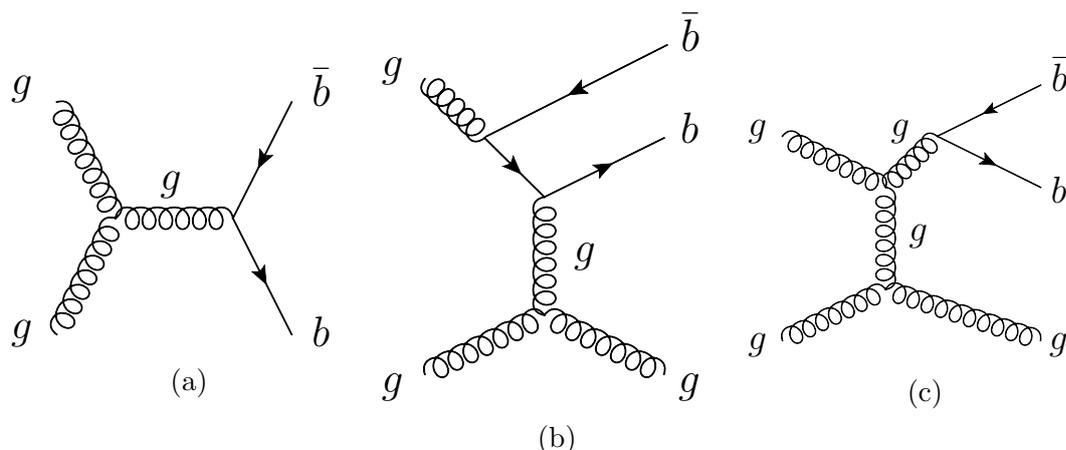


Figure 2.5: Pair creation (a), flavour excitation (b) and gluon splitting (c) processes.

Fig. 2.6 shows the relative importance of these production mechanisms for pp collisions over a wide range of energies, indicating the flavour excitation as the most favorable at LHC.

The $b\bar{b}$ pair production peaks at small angles with respect to the beam direction, as shown in Fig. 2.4b. In a recent paper, the LHCb collaboration reported two $b\bar{b}$ production cross section measurements which, extrapolated to the full solid angle, give [103]:

$$\begin{aligned}\sigma_{pp \rightarrow b\bar{b}} &\sim 295 \mu\text{b} \quad (\sqrt{s} = 7 \text{ TeV}), \\ \sigma_{pp \rightarrow b\bar{b}} &\sim 600 \mu\text{b} \quad (\sqrt{s} = 13 \text{ TeV}).\end{aligned}\tag{2.5}$$

The above results display a linear behaviour of the $b\bar{b}$ cross section as a function of the energy, with the consequent strong advantage of increasing the collision energy. Following Eq. (2.5), about 6×10^{11} $b\bar{b}$ pairs are produced at LHC per fb^{-1} .

Once a b quark is produced, it will interact with another quark in the strong field to

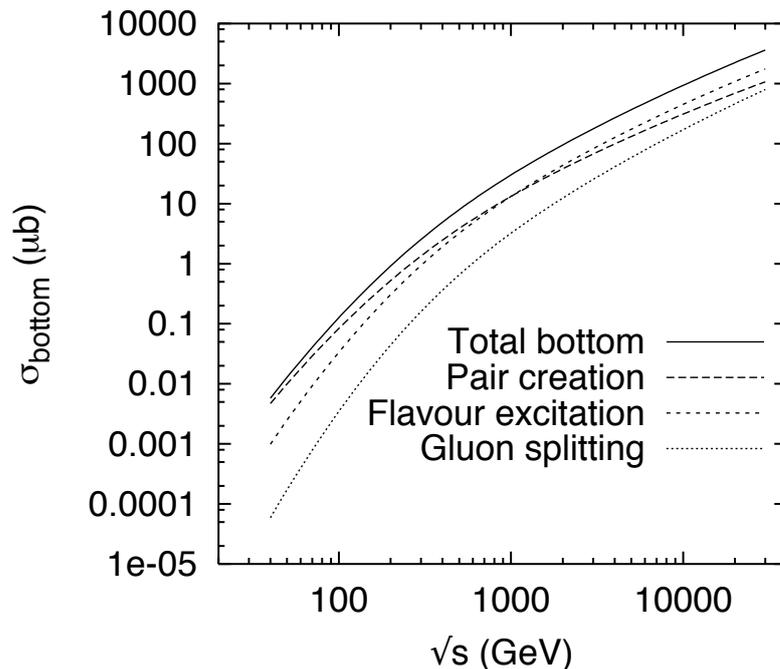


Figure 2.6: Total bottom cross sections for pp collisions as a function of the centre-of-mass energy, showing the contributions from pair creation, flavour excitation and gluon splitting processes [100].

hadronise into a colourless hadron. The probabilities for a quark to hadronise into specific b -hadrons are called *fragmentation fractions* or *hadronisation fractions*, and their theoretical prediction is burdened by the non-perturbative regime of the strong dynamics. However, combinations of the hadronisation fractions have been measured at LHCb at $\sqrt{s} = 7$ TeV [104, 105]:

$$\frac{f_s}{f_d} = 0.259 \pm 0.015, \quad (2.6)$$

which assumes isospin symmetry, i.e. $f_d = f_u$, and

$$\left[\frac{f_{\Lambda_b}}{f_u + f_d} \right] (p_T) = (0.404 \pm 0.017 \pm 0.027 \pm 0.105) \times [1 - (0.031 \pm 0.004 \pm 0.003) \times p_T(\text{GeV})], \quad (2.7)$$

where the first uncertainty is statistical, the second one is systematic, and the third one represents the error on the $\Lambda_c^+ \rightarrow pK^-\pi^+$ branching fraction, being the main source of uncertainty. For the Λ_b production, the transverse momentum dependence is also given, whereas is absent in f_s/f_d . The relevant quantity for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis (Sec. 4.4) is the ratio (2.6), which quantifies the B_s^0 over B_d^0 production.

B-factories versus LHC

Before the advent of LHC and the LHCb experiment, the most copious sources of B mesons were the so-called B -factories KEKB and PEP-II, where the Babar [106] and Belle [107] detectors took data until 2008 and 2010, respectively. The two machines were asymmetric e^+e^- colliders, with an energy in the centre-of-mass of 10.58 GeV, corresponding to the mass of the $\Upsilon(4s)$ resonance. The $\Upsilon(4s)$ decays predominantly into B meson pairs, and the boost provided by the asymmetric beams ensured the needed spatial resolution for their identification. The main advantages of a lepton collider is the cleanness of the event, being free from QCD backgrounds arising from hadronic collisions, as well as the possibility to fully reconstruct the kinematics of the decay products. The b -tagging efficiency at B -factories is also higher with respect to the one achieved in hadronic collisions, which produce significantly more crowded events. In addition, collision events at LHC are fully dominated by QCD background, and the kinematic can only be closed in the plane transverse to the beam direction, as the boost of the interacting partons is unknown. Nonetheless, there are several advantages in producing b -hadrons at the LHC: a very high $b\bar{b}$ cross section and the high energy, which opens the possibility of producing b -hadrons heavier than the B_s , and also provides a huge boost along the longitudinal direction, which is a crucial feature to distinguish the b -hadrons from the overwhelming background. B mesons at LHCb in fact travel for distances of $\mathcal{O}(1\text{ cm})$ before decaying.

A comparison of the invariant mass spectrum of $B \rightarrow D\pi$ events reconstructed at BaBar and LHCb (2011 data only) is made in Fig. 2.7, showing that LHCb has very competitive performances in terms of yield and background to signal ratio, with only a small fraction of the collected data. The Belle experiment is now starting

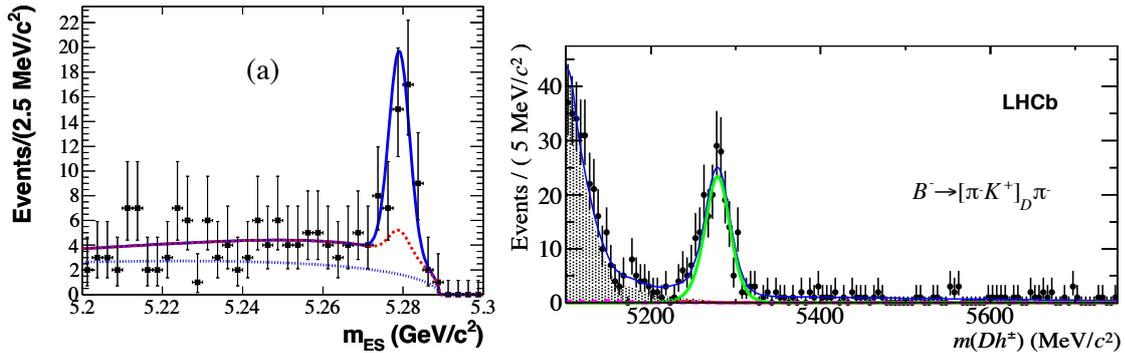


Figure 2.7: (Left) Energy-substituted mass distribution of $B^\pm \rightarrow D(\pi^- K^+) \pi^\pm$ decays at Babar [108], $m_{ES} = \sqrt{(s/2 + \vec{p}_\Upsilon \cdot \vec{p}_B)^2 / E_\Upsilon^2 - p_B^2}$. (Right) Invariant mass distribution of $B^- \rightarrow D(\pi^- K^+) \pi^-$ decays at LHCb [109].

its second phase: the Belle II detector [110] is collecting data at the SuperKEKB

collider, with the aim of competing with LHCb in the coming years.

2.2 Overview of the LHCb detector

The LHCb detector [80,111] is a single-arm forward spectrometer covering 10 mrad to 300(250) mrad in the bending (non-bending) plane. The forward coverage of the acceptance follows the angular distribution of the $b\bar{b}$ pairs produced at LHCb, shown in Fig. 2.4b. The pseudorapidity acceptance is $2 < \eta < 5$, where the pseudorapidity is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \quad (2.8)$$

and θ is the polar angle, i.e. the angle with respect to the beam axis z . The detector layout is shown in Fig. 2.8: a right-handed coordinate system is adopted, with y identifying the vertical and z is along the beam.

LHCb is comprised of multiple sub-detectors adopting diverse technologies in order to track and identify different particles across multiple energy regimes. Sub-detectors can be divided into two classes based on their purpose:

- The **tracking system** comprises the magnet, the Vertex Locator (VELO), the Tracker Turicensis (TT) and three tracking stations T1-T3 which are divided into Inner Tracker (IT) and Outer Tracker (OT). The TT and the IT are also collectively called Silicon Tracker (ST), as they share the same silicon technology.
- The **particle identification system** is made up of two Ring Imaging Cherenkov detectors (RICH1 and RICH2), an electromagnetic calorimeter (SPD+PS+ECAL), a hadronic calorimeter (HCAL) and five muon stations (M1-M5), the first of which (M1) is placed before the calorimeters and the other four (M2-M5) are interspersed with iron absorbers.

A brief description of each subdetector is given in the following sections, with a special attention on the muon detector.

2.3 The Tracking System

The reconstruction of particle trajectories starts with the VELO in the very proximity of the interaction point. A series of silicon trackers measure the coordinates of particles flying close to the beam axis, while an outer tracker made of straw-tubes covers the external acceptance. By means of a dipole magnet, particle momentum is measured with very high precision.

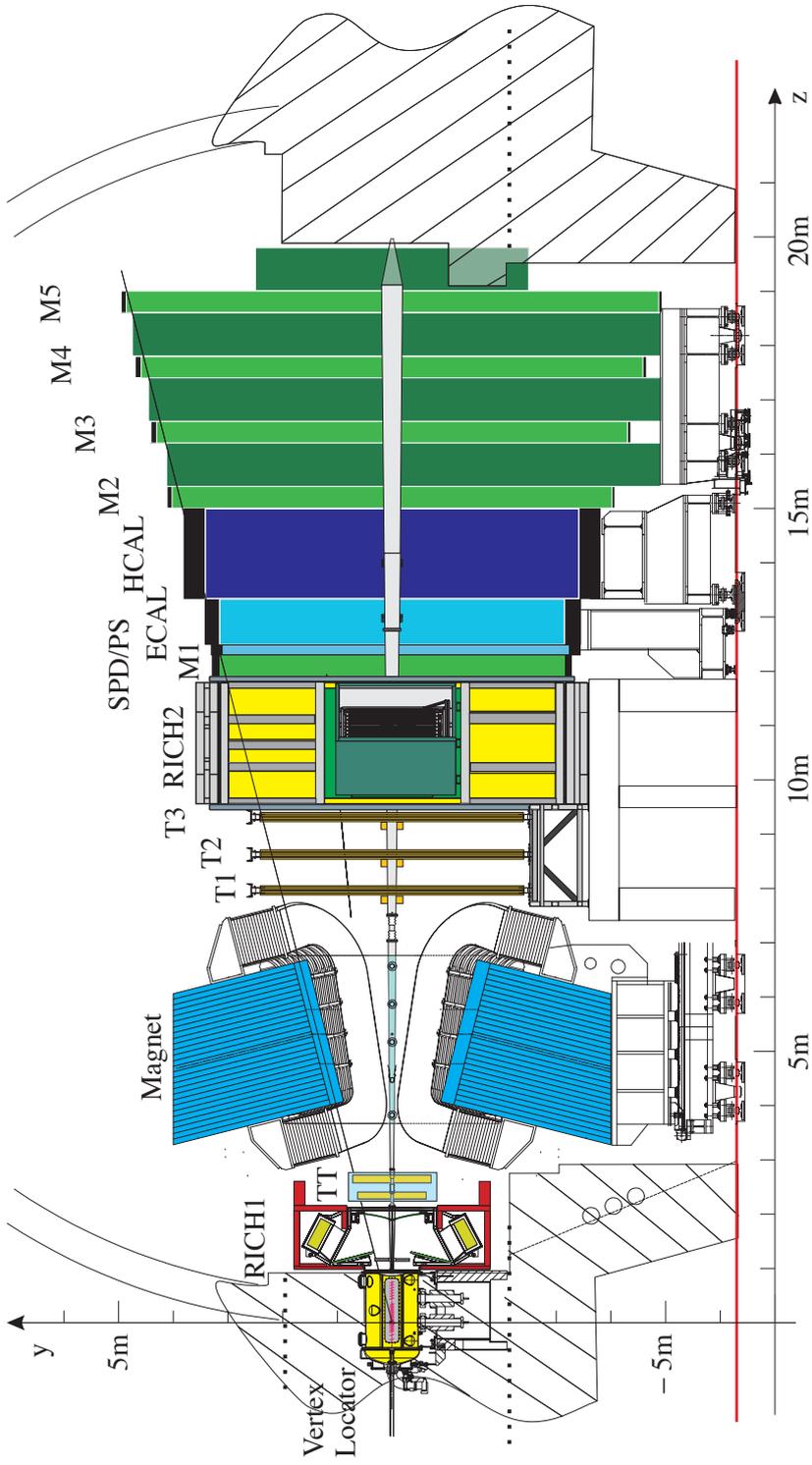


Figure 2.8: Side view of the LHCb detector.

2.3.1 The Magnet

The LHCb dipole magnet [112] consists of two saddle-shaped warm coils (2×25 ton) inside a window-frame iron yoke (1450 ton), as shown in Fig. 2.9. The coils are designed to minimise the magnetic field inhomogeneities, which are measured to be below 5% across the acceptance. The electric current flowing in the conductor is about 5.8 kA, with a total power dissipation of about 4.2 MW.

The generated magnetic field is directed along the vertical axis y and bends

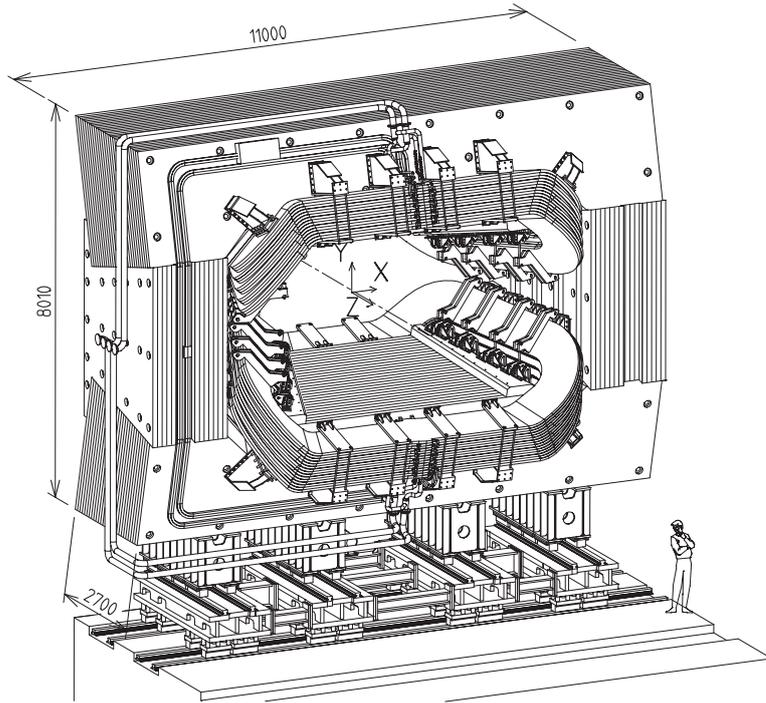


Figure 2.9: Perspective view of the LHCb magnet [112].

charged particle so that their momentum p can be measured from their angular deflection α . A particle with unitary charge inside a uniform magnetic field B moves, in the plane perpendicular to B , in circular motion with radius of curvature $R = p/0.3B$, and, using $dl = R d\alpha$, the momentum resolution can be written as

$$\frac{\delta p}{p} = \frac{\delta \alpha}{\alpha} = \frac{p}{0.3 \int B dl} \delta \alpha. \quad (2.9)$$

The resolution is thus proportional to the angular resolution, which receives contributions from both the multiple scattering and the spatial resolution of the tracking detectors [113], described in the following sections. The LHCb magnet has a field

integral of $\int Bdl = 4 \text{ Tm}$ for particles originating close to the interaction point and travelling through the entire tracking system ($\sim 10 \text{ m}$). To ensure high momentum resolution, a strong magnetic field is needed but also the field integral must be measured with a precision of 10^{-4} , which is achieved by means of an array of Hall probes. The resulting momentum resolution is about $\delta p/p = 0.5\%$ for momenta up to $20 \text{ GeV}/c$ and $\delta p/p = 0.8\%$ for momenta of about $100 \text{ GeV}/c$ [90]. The polarity of the magnet is swapped every two weeks of data taking to better control the detection asymmetries.

2.3.2 The Vertex Locator

Displaced secondary vertices are a distinctive feature of b -hadron decays: their reconstruction is therefore a fundamental ingredient at LHCb to identify secondary vertices associated with long-lived particle, against the primary vertices of short-lived ones. Time-dependent analyses also rely on a precise measurement of particle lifetimes, hence a very high spatial resolution is mandatory. The VELO detector [114] fulfills these requirements by providing an accurate measurement of the particle coordinates in the proximity of the interaction region, which are used in the trigger (Sec. 2.5) to reconstruct the primary vertex²(PV) and the track impact parameter³ (IP).

The VELO system is comprised of a series of silicon modules displaced perpendicularly along the beam axis, as shown in Fig. 2.10. The silicon technology guarantees the needed spatial resolution and a strong resistance to radiations. Each module is separated into two halves, and each half is equipped with $300 \mu\text{m}$ thick R and Φ sensors. The two halves are staggered by 1.5 cm in z so they can overlap to ensure no acceptance holes and for alignment purposes. There are 23 modules in total, two of which are equipped with R sensors and placed upstream of the interaction point to measure the number of interactions per collision, thus producing the *pile-up* information for the trigger. The R sensors are segmented into concentric silicon strips, providing a measurement of the r coordinate, i.e. the distance from the beam axis z , while the Φ sensors measure the azimuthal coordinate thanks to their radial segmentation (Fig. 2.11a). The 3D track reconstruction is completed by knowing the position of each sensor plane within the experiment. The cylindrical ($R - \Phi$) geometry was chosen to speed-up the determination of the IP in the trigger, but also naturally allows to use the smallest strip pitch ($\sim 40 \mu\text{m}$) close to the beam axis, where the highest resolution is needed, thus reducing the number of readout channels and balancing the sensor occupancy. A better IP resolution is achieved when the track extrapolation is short: for this reason the sensitive area

²The PV is the point where pp collisions take place, as described in Sec. 2.5.2.

³The IP is defined as the distance between a track and a vertex at the track point of closest approach to that vertex.

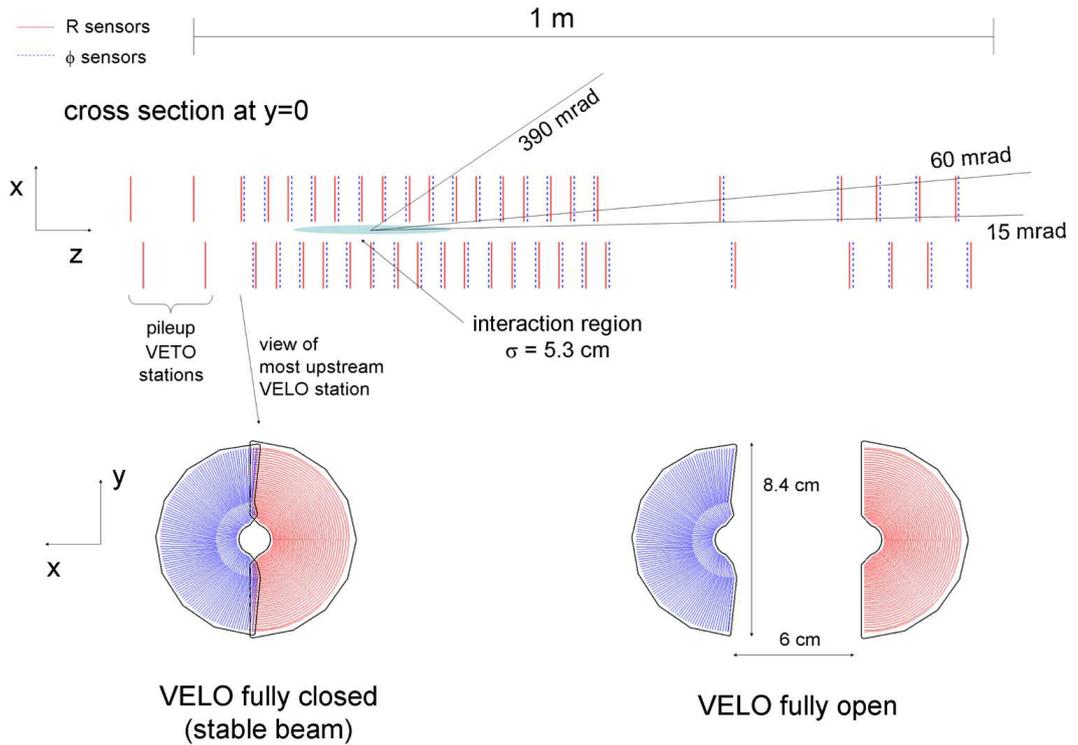


Figure 2.10: The VELO detector [80].

of the VELO sensors begins at only 7 mm from the beam. Due to the larger LHC aperture during injection, the VELO system retracts to a safety distance of 3 cm and only closes when the beams are stable. These requirements, along with the need of reducing as much as possible the amount of material within the detector acceptance, led to the choice of operating the silicon sensors inside a vacuum vessel with 0.5 mm thick aluminum walls (Fig. 2.11b).

In the smallest strip pitch region, the spatial resolution is as high as $4 \mu\text{m}$, which translates into high IP and PV resolution, as reported in Fig. 2.12a and Fig. 2.12b.

2.3.3 The Silicon Tracker

The ST [116] comprises the Tracker Turicensis (TT), placed upstream of the magnet, and the inner region of the three tracking stations T1-T3, named Inner Tracker (IT), downstream of the magnet (Fig. 2.8).

The Tracker Turicensis The TT system measures the transverse momentum (p_T) of the traversing particle for the trigger and is employed for track reconstruc-

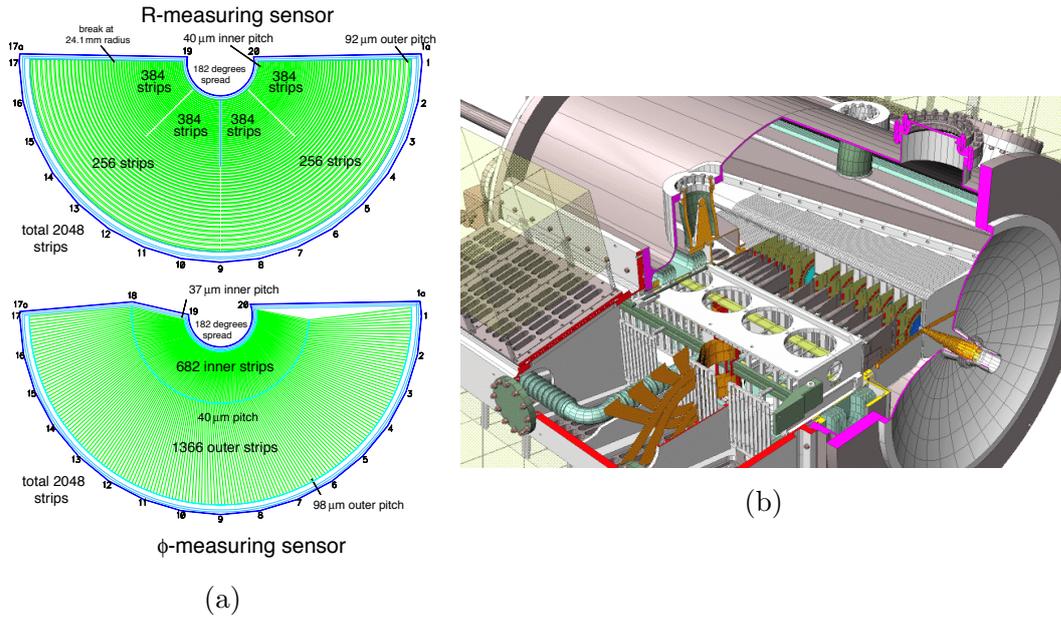


Figure 2.11: (a) Schematics of the R and Φ VELO sensors. (b) The VELO vacuum vessel.

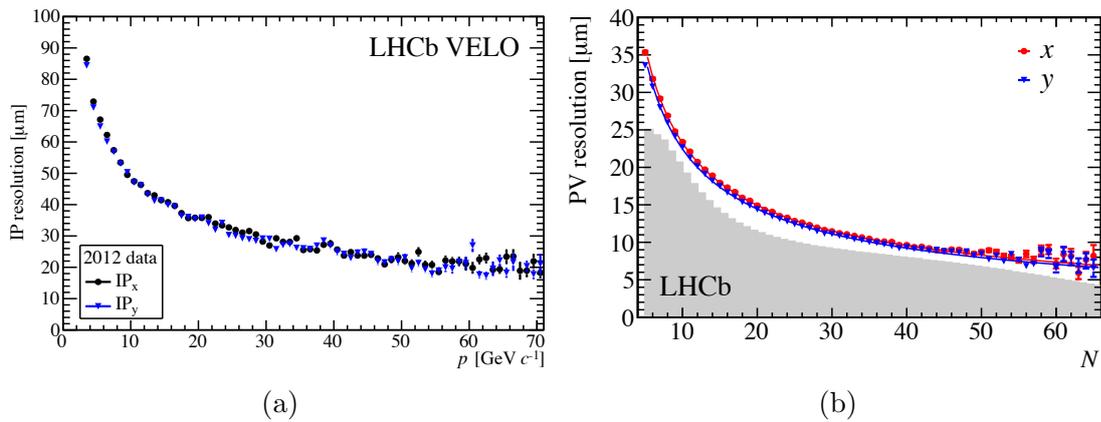


Figure 2.12: (a) Resolution of the x and y components of the IP as a function of momentum [115]. (b) PV resolution as a function of track multiplicity. The superimposed histogram shows the distribution of the number of tracks per reconstructed PV [90]. Both plots are made using data collected in 2012.

tion of long-lived neutral particles that decay outside of the VELO, but also for charged low-momentum tracks that are bent out of the acceptance by the magnetic field.

The TT is a 150 cm wide and 130 cm high planar detector placed at the entrance

of the magnet and is comprised of two stations, TTa and TTb, each of which is made up of two silicon layers (Fig. 2.13a). The first and the fourth layers are segmented into vertical strips (x layers), while in the second and in the third ones the strips are rotated by a $\pm 5^\circ$ stereo angle respectively (u/v layers), as shown in Fig. 2.13a. The single sensors are $\sim 9.64 \times 9.44 \text{ cm}^2$ wide and $500 \mu\text{m}$ thick silicon tiles carrying 512 readout strips with a pitch of $183 \mu\text{m}$. The sensors are arranged in groups of 7 into extractable adjacent modules overlapped by few mm along x to avoid acceptance gaps and to facilitate the relative alignment. The single hit resolution of the TT reaches $\sim 60 \mu\text{m}$.

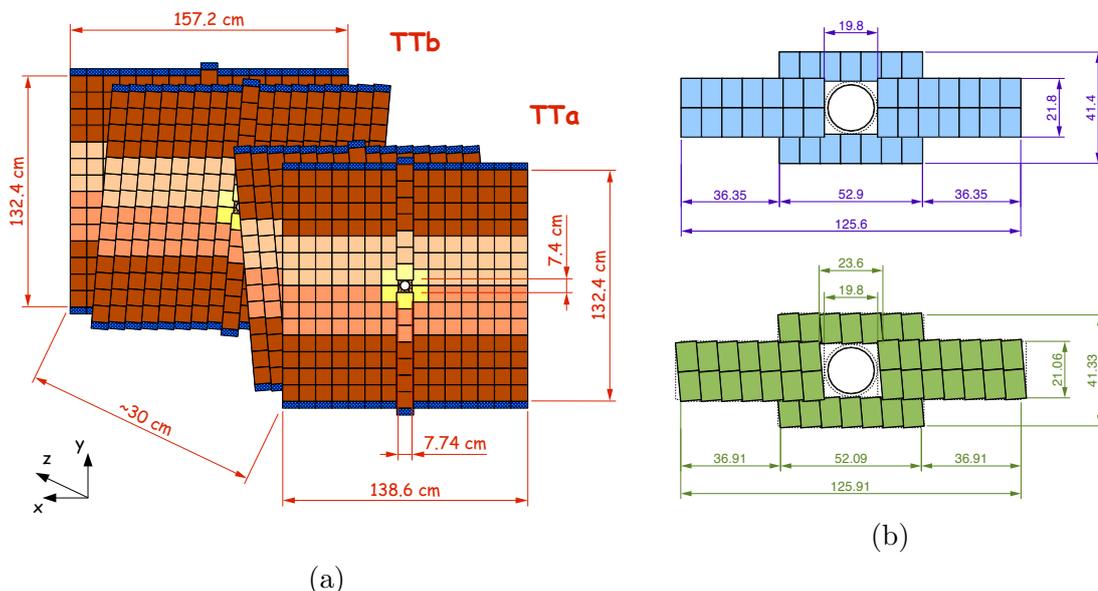


Figure 2.13: (a) Layout of the four TT stations. (b) Layout of the x layer (top) and v layer (bottom) of the IT in the T2 station.

The Inner Tracker The IT constitutes the region closer to the beam pipe of the T1-T3 tracking stations and completes the silicon tracking system, providing precise momentum and coordinate measurements.

Each one of the three IT stations is comprised of four detector boxes settled around the beam pipe, where each box contains four detection layers with the usual $xuvx$ topology. Each layer consists of seven modules with one or two silicon sensors depending on their position. The IT silicon sensors share the same technology of the TT: they have an area of $7.6 \times 11 \text{ cm}^2$ with 384 readout strips and $198 \mu\text{m}$ pitch. Their thickness is $320 \mu\text{m}$ for modules with one sensor and $410 \mu\text{m}$ for modules with two sensors, to optimise the signal-to-noise ratio while minimising

the material budget. The arrangement of the modules is schematised in Fig. 2.13b for a x and a v layer. The single hit resolution of the IT amounts to $\sim 50 \mu\text{m}$.

2.3.4 The Outer Tracker

The OT [117] constitutes the outer region of the three tracking stations T1-T3, as shown in Fig. 2.14a. The OT is still responsible for reconstructing track segments

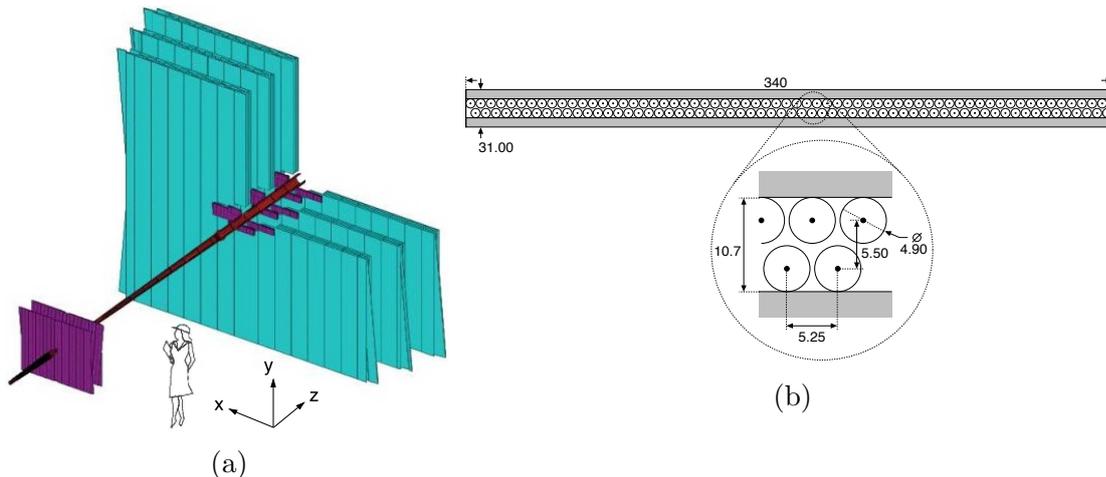


Figure 2.14: (a) Overview of the tracking stations: the TT, IT (purple) and the OT (cyan). (b) Cross section of an OT module with a zoom on the straw-tubes. The dimensions are in mm.

with high resolution and for providing precise momentum measurements, but its larger distance from the beam pipe allows the use of a cheaper technology with respect to the silicon used in the IT, as the granularity requirements are less stringent where the particle flux is lower. For this reason, the detector consists of 55000 straw-tubes that cover a total active area of $\sim 87 \text{ m}^2$.

Each OT station comprises four layers of straw-tube modules arranged in the usual $xvwx$ geometry, where each module is made up of two staggered layers of 64 drift tubes each, as schematised in Fig. 2.14b. The single straw-tube is a gas-tight cylinder with a diameter of 4.9 mm enclosing a single conductive wire. The internal surface of the tube acts as a cathode and is made up of carbon-doped Kapton externally clad with an aluminum foil, for a total thickness of $\sim 75 \mu\text{m}$. The internal gold-plated tungsten anode wire has a diameter of $25.4 \mu\text{m}$ and is strung with 0.7 N tension, as it has to be centered with respect to the straw-tube within $50 \mu\text{m}$ over the entire straw length. The total OT material budget sums up to 9.6% of X_0 , achieved with the important aim of reducing the multiple scattering and the material in front of the calorimeters. The gas that fills the straws is a

mixture of Argon (70%) and CO₂ (30%) that guarantees a drift time below 50 ns and a drift-coordinate resolution of 200 μm .

Invariant mass resolution

The precision on the invariant mass of two oppositely charged muons is a key element for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis, and determined by the performances of tracking system described above. The resolution at the B^0 mass has been estimated by interpolating the measurements of many dimuon resonances, as described in Sec. 4.3. The result yields about 23 MeV/c², i.e. below 5 per mille.

2.4 The Particle Identification System

A redundant and unambiguous particle identification (PID) is a fundamental requirement for the LHCb detector. The discrimination between pions, kaons and protons is realized by exploiting their different Cherenkov radiation in two Cherenkov detectors, RICH1 and RICH2. Electron and photons are detected by means of the calorimeter system, while muons are identified in the muon stations in the terminal part of the detector.

2.4.1 The Ring Imaging Cherenkov Detector

When a charged particle traverses a medium with a velocity higher than the local phase velocity of the light it emits a Cherenkov light cone. The Cherenkov photons are in fact emitted at an angle θ_c with respect to the particle direction, which depends on the particle velocity βc and the refractive index n of the medium, according to

$$\cos \theta_c = \frac{1}{n\beta}. \quad (2.10)$$

A measurement of the Cherenkov angle therefore allows to infer the mass of a particle once its momentum and charge are measured.

To this end, LHCb employs two ring-imaging Cherenkov detectors: RICH1 and RICH2 [118]. At small polar angles, the particle momentum spectrum is harder while it is softer at large angles, hence two detectors and different radiator materials are used to cover the full momentum range. The RICH1 is placed upstream of the magnet (Fig. 2.8) and covers the lower momentum range, $\sim 1\text{-}60$ GeV/c, with a gaseous fluorobutane (C₄F₁₀) radiator. The RICH2, downstream of the magnet, employs gaseous CF₄ to cover the highest momentum range: $\sim 15\text{-}100$ GeV/c. In both RICH detectors, the Cherenkov photons are deflected by means of a combination of spherical and flat mirrors, and then collected by Hybrid Photon

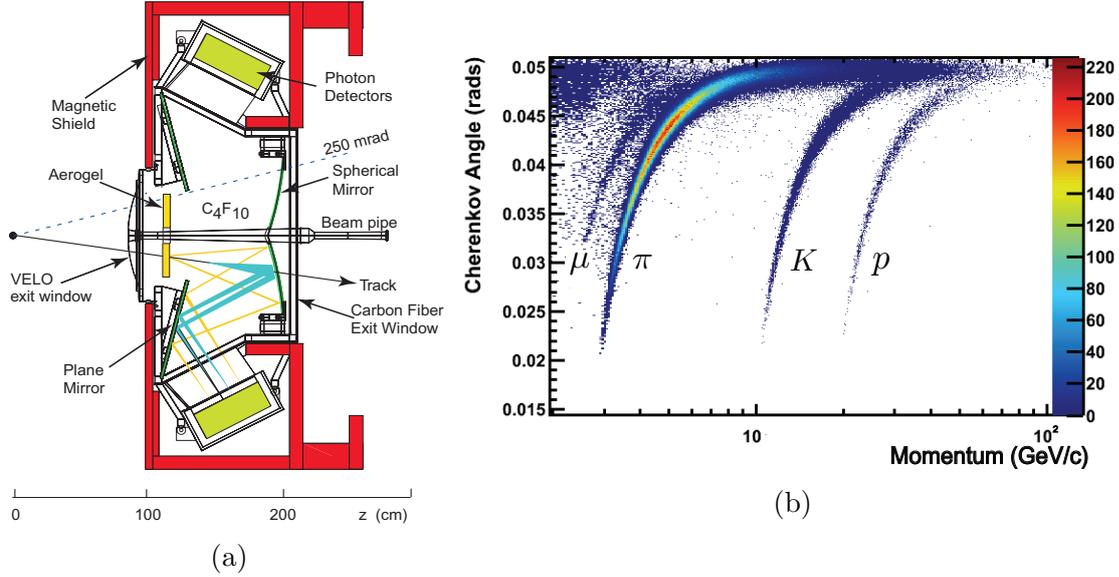


Figure 2.15: (a) Side view of the RICH1 detector. (b) Reconstructed Cherenkov angle as a function of track momentum in RICH1 [119].

Detectors (HPDs) in the wavelength range 200-600 nm, as shown in Fig. 2.15a. An iron shield provides a strong reduction of the residual magnetic field to ensure the correct operation of the HPDs, without affecting the field integral in the region between the VELO and the TT.

Fig. 2.15b shows how particles populate distinct bands in the $\theta_c - p$ plane according to their masses. Even though RICH detectors are primarily used for hadron identification, a muon band can also be distinguished. The kaon identification efficiency and pion to kaon misidentification efficiency are shown in Fig. 2.16 as a function of the particle momentum.

2.4.2 The Calorimeters

The calorimeters [121] complement the RICH PID by identifying and measuring the position of photons, electrons and hadrons thanks to their different energy deposits and shower shapes. In addition, the CALO selection based on energy deposit is used in the Level-0 trigger and is performed within $4 \mu\text{s}$ from the interaction. Starting from the interaction point, the calorimeter system is composed of a Scintillator Pad Detector (SPD), a lead converter, a Preshower (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL), all positioned after the RICH2 and the first muon station (M1), as shown in Fig. 2.8. All the sub-detectors share the same principle of operation: the scintillation light produced

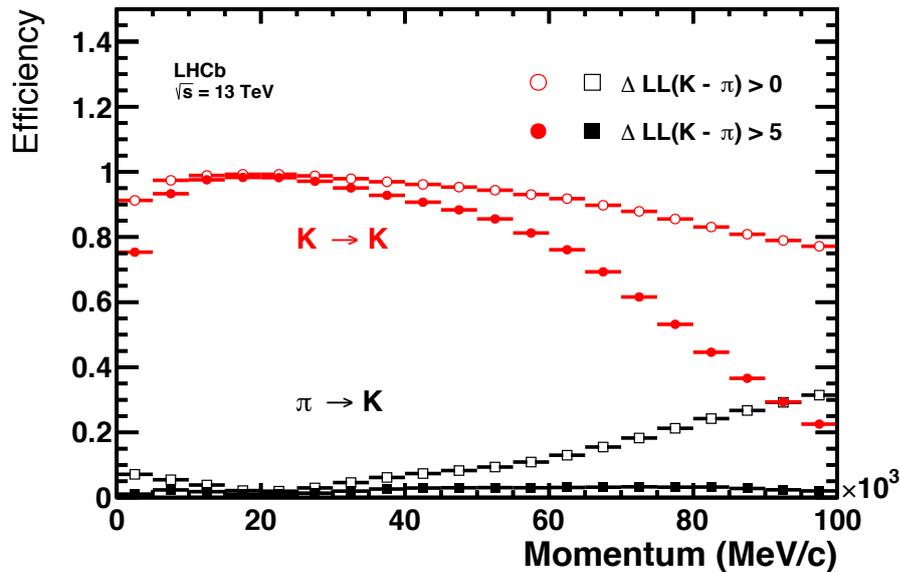


Figure 2.16: Efficiency of the kaon identification (red) and pion to kaon misidentification (black) as a function of the momentum, as measured from 2015 data. Two cuts on the delta log-likelihood (ΔLL) between kaon and pion hypotheses are shown [120].

by the traversing particle is guided with wavelength-shifting (WLS) fibers to a Photo-Multiplier (PMT).

SPD/PS

The SPD discriminates charged and neutral particles, mainly neutral pions with high E_T , while the PS precedes ECAL to provide the longitudinal segmentation needed to reject the large background of charged pions.

The SPD/PS detector is comprised of two almost identical planes⁴ of rectangular scintillator pads which enclose a 15 mm ($2.5 X_0$) lead converter. Each plane is divided into 3 sections of increasing cell size, according to the scheme shown in Fig. 2.17. The cells are packed into $\sim 48 \times 48 \text{ cm}^2$ boxes that are grouped into supermodules arranged into 2 rows and 13 columns. This variable lateral segmentation, that matches the ECAL one, is naturally adopted as the hit density varies by two orders of magnitude over the calorimeter surface. Each cell is made up of a 15 mm thick polystyrene-based scintillator pad, which also incorporate a coiled WLS fiber. The WLS fiber is coupled via clear fibers to multianode photomultiplier tubes (MAPMT), which are housed at the detector periphery.

⁴Owing to projectivity requirements, all the SPD plane dimensions are 0.45% smaller than those of the PS plane.

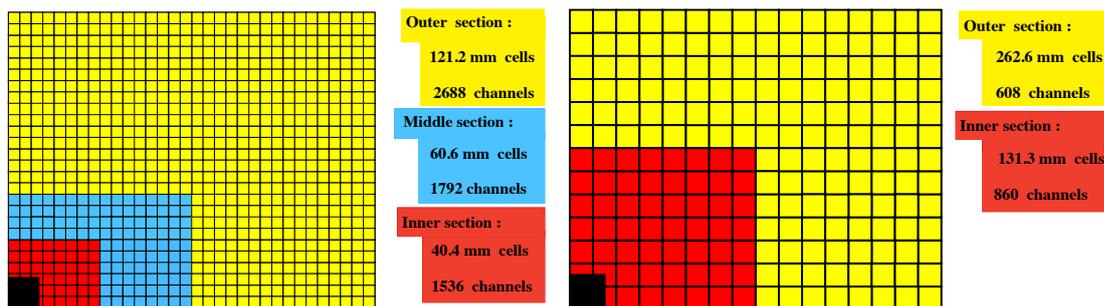


Figure 2.17: Lateral segmentation of one quadrant of SPD/PS and ECAL (left) and HCAL (right) as seen by front. Cell sizes and number of channels refer to ECAL (left) and HCAL (right).

ECAL

In order for the energy resolution to be optimal, full containment of the showers produced by high energy photons is required. For this reason, ECAL has a thickness of 25 radiation lengths. The detector makes use of the well established shashlik calorimeter technology, consisting in a sampling lead/scintillator frame readout by plastic WLS fibers, to achieve a modest energy resolution and a fast response time. ECAL adopts a sampling structure of 2 mm lead sheets interspersed with 4 mm thick polystyrene-based scintillator plates forming a 42 cm stack with a Moliere radius of 3.5 cm. Each layer incorporates a pattern of holes to house the traversing WLS fibers responsible for collecting the scintillation light, as shown for the module in Fig. 2.18a. As far as the segmentation is concerned, ECAL follows the same scheme of Fig. 2.17.

The energy of the electromagnetic showers can be measured by ECAL with a resolution of

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \quad (2.11)$$

where E is measured in GeV.

HCAL

Whereas the SPD/PS and ECAL are divided into 3 regions, HCAL only has 2 regions with larger cell sizes (Fig. 2.17), since hadronic showers are typically larger than electromagnetic ones. Moreover, the HCAL detector has a thickness of only 1.2 m, corresponding to 5.6 interaction lengths (λ_I),⁵ as the trigger requirements do not impose a high energy resolution and therefore there is no need for a full

⁵The ECAL accounts for an additional $1.2 \lambda_I$.

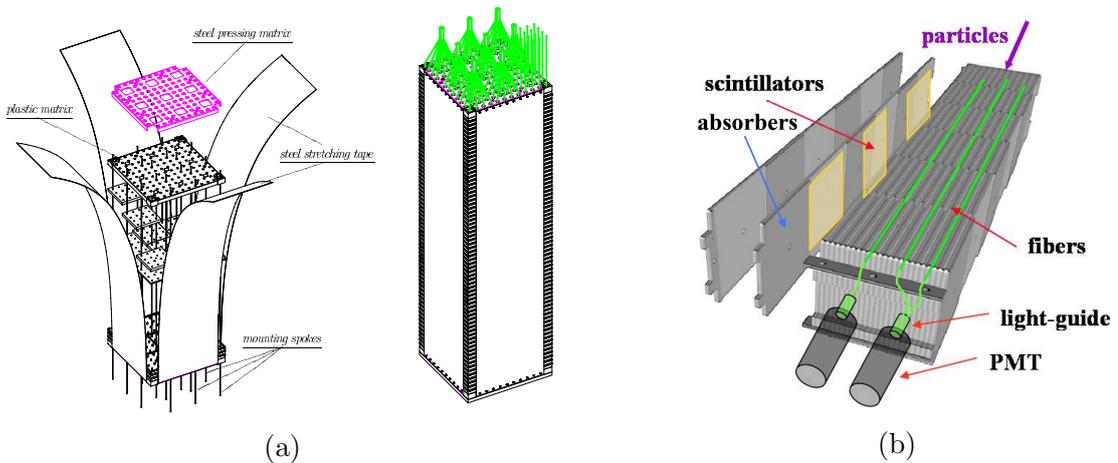


Figure 2.18: (a) Exploded view of one ECAL module. On the right, the assembled stack with inserted fibers. (b) A HCAL module where two ejected layers highlight its elementary periodic structure.

containment of the hadronic shower. The detector still features a sampling structure but alternates 4 mm scintillator planes with 16 mm iron plates, as shown for the module in Fig. 2.18b, with the peculiar feature that the scintillating tiles run parallel to the beam. The scintillation light is once more collected by means of WLS fibers running along the modules towards the PMTs. It is worth noting that the light yield of the HCAL module is a factor ~ 30 less with respect to the ECAL one: for this reason, the PMTs need to operate at a higher gain. HCAL is capable of detecting hadrons with an energy resolution of

$$\frac{\sigma_E}{E} = \frac{69 \pm 5\%}{\sqrt{E}} \oplus (9 \pm 2)\%, \quad (2.12)$$

as determined from a fit to test-beam data (energy dimensions are in GeV).

2.4.3 The Muon System

Muons are fundamental for the LHCb physics program. They are present in the final state of many CP-sensitive B decays and are also used in oscillation measurements, since the flavour of the decaying B hadron can be inferred from the charge of the muon originating from its semileptonic decay (*flavour tag*). Needless to say, muon triggering and identification play a fundamental role in rare decays searches, and in particular the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays.

The muon system provides both the Level-0 trigger for muons above a given p_T threshold and muon identification information for the High Level Trigger.

Layout

The muon system [122–124] is placed in the terminal part of the LHCb detector, as shown in Fig. 2.8, and is based on gas chambers. There are five muon stations, labelled M1–M5, the first of which is located upstream of the calorimeters, while M2–M5 are interspersed with 80 cm thick iron absorbers to filter penetrating muons,⁶ according to the layout shown in Fig. 2.19. The stations have a rect-

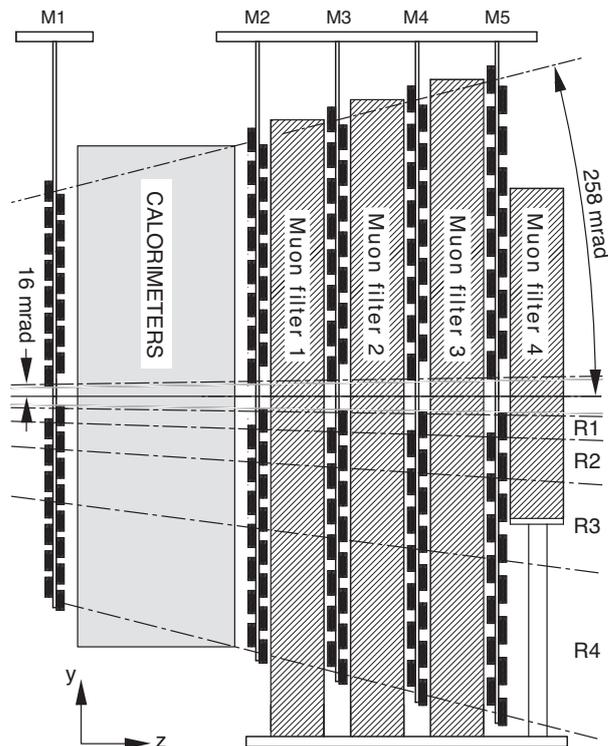


Figure 2.19: Lateral view of the five muon station arrangement.

angular shape and follow a projective geometry, i.e. their transverse dimensions scale with increasing z , and are divided into four regions, R1–R4, with increasing distance from the beam pipe, as can be seen in Fig. 2.20. The largest station, M5, has a dimension of about 10×9 m². The full system comprises 1380 chambers for a total area of 435 m². The granularity of the detector, which determines the x and y coordinate resolution, is defined by rectangular *logical pads* (Fig. 2.20), which are obtained by grouping physical channels together, as explained in the following. The M1–M3 stations have a fine granularity along x , which is used to reconstruct the muon track direction as well as to provide a stand-alone measurement of its transverse momentum with a resolution of 20% for the Level-0 trigger.

⁶The minimum momentum that allows a muon to cross the entire detector is around 6 GeV/c.

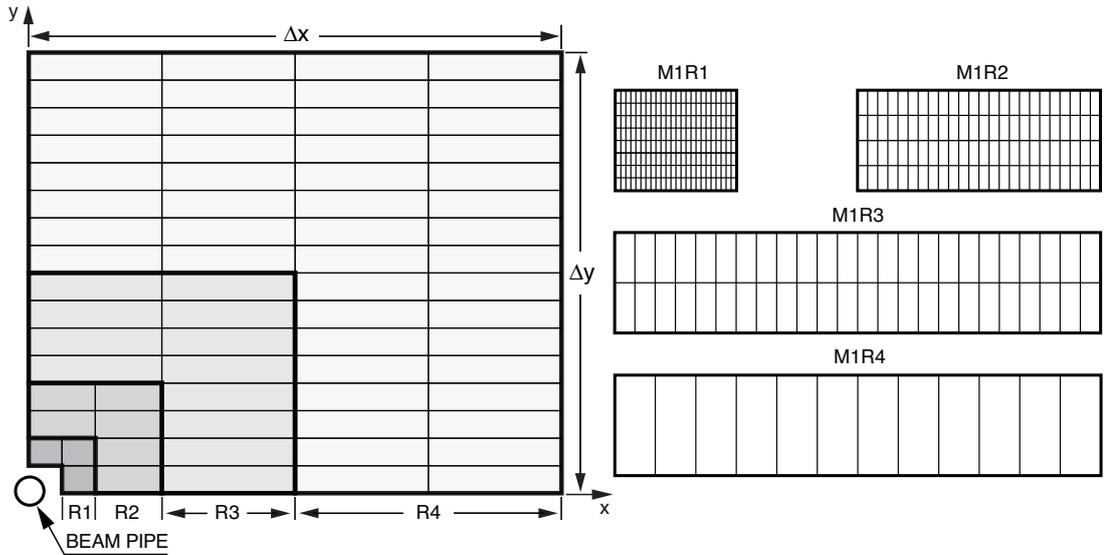


Figure 2.20: On the left, front view of one quadrant of a muon station. The regions R1-R4 are defined by groups of chambers, which are represented by a rectangle. On the right, the division of the M1 chambers into logical pads. The number of pad rows per chamber is the same for each station, while the number of pad columns, which define the x granularity, is doubled in M2-M3 and halved in M4-M5 with respect to the M1 station shown in the picture.

The granularity is coarser for the M4 and M5 stations, as their main purpose is the identification of penetrating particles. Logical pad size is also larger for the outer regions of the detector since the spatial resolution is anyway dominated by the multiple scattering.

The Chambers

The muon chambers must provide efficient and fast muon detection while sustaining an intense particle flux. To cope with these requirements, 1368 Multi-Wire Proportional Chambers (MWPCs) are employed across the whole detector, with the exception of the innermost region (R1) of the M1 station, where 12 Gas Electron Multiplier (GEM) chambers are installed.

Since a fivefold coincidence between all muon stations is required by the Level-0 trigger, high trigger efficiency demands a high chamber efficiency within a time window of 25 ns. For this reason, the chambers in M2-M5 are comprised of two layers with independent readout, where each layer is made up of two gas gaps: the resulting four gaps per station are set in OR configuration. In order to minimize the amount of material before the calorimeters, the M1 station has chambers with

only one layer, i.e. two OR-ed gas gaps. This setup provides both flexibility and redundancy to the muon system.

The chambers are segmented into *physical pads*, each read out by a single Front-End (FE) electronics channel. Physical pads are defined by the readout technology used in the MWPCs: since the spatial resolution requirements and particle rates are very different across the detector, different solutions are adopted. Wire pads are used in the R4 region, mixed wire-cathode pads are used in R1-R2 of M2 and M3, while cathode pads are used everywhere else. GEM chambers use anode pads. Fig. 2.21a shows how logical pads are obtained for the inner regions of M2 and M3, while everywhere else up to four adjacent physical pads are OR-ed to make a logical pad. In this way, the size and therefore the electrical capacitance of a given physical pad can be limited, thus reducing noise and dead-time on FE channels. With the exception of M1, where the occupancy is high, several contiguous logical pads are further OR-ed to build x and y strips named *logical channels*. Logical pads are then reconstructed from logical channels by the coincidence of two crossing strips. 55296 logical pads are used for the muon tracking, whose dimensions are summarized in Fig. 2.21b.

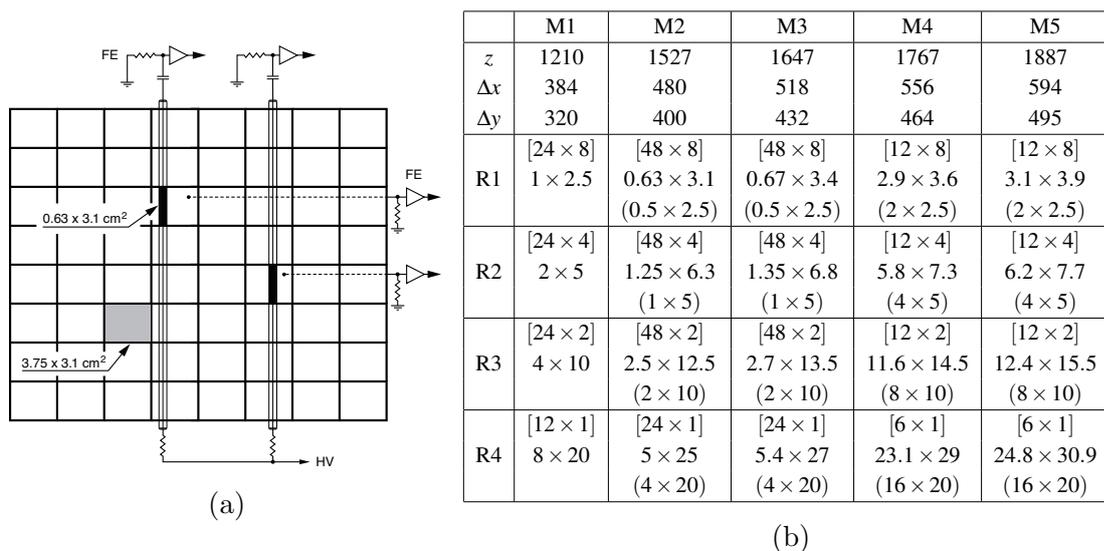


Figure 2.21: (a) Mixed wire-cathode pads readout for a M2R1 chamber. The logical pads, shown in black, are obtained by the coincidence of crossing vertical wire pads and cathode pads. (b) Summary table of the muon system layout. z is the distance from the interaction point, while Δx and Δy represent the dimensions of a quadrant. The R1-R4 rows contain the number of logical pads per chamber (square brackets) and their size along x and y , while the sizes projected onto M1 are reported in parenthesis. All dimensions are given in cm.

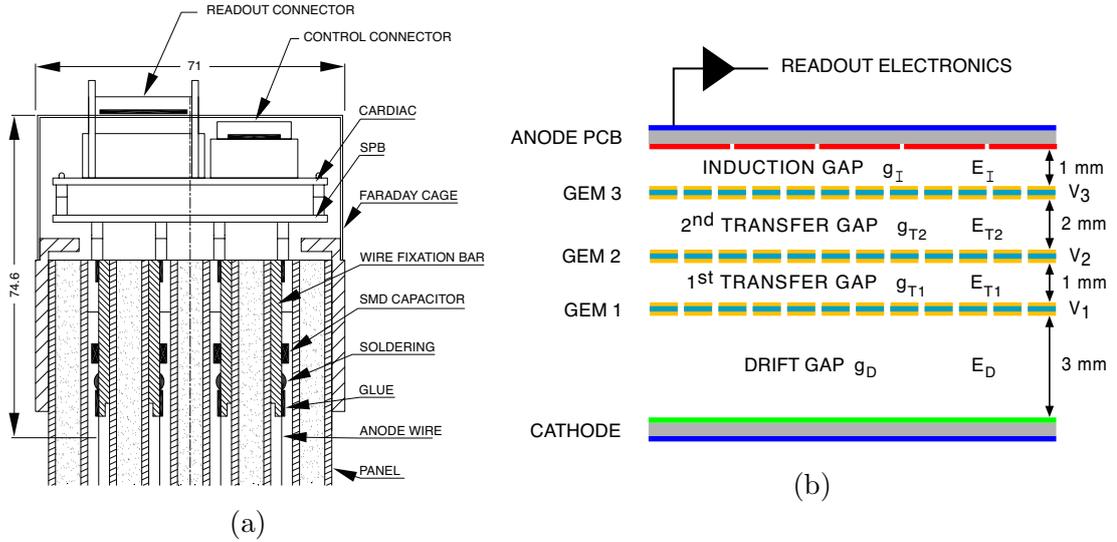


Figure 2.22: (a) Schematic cross section of a four gap MWPC. SPB stands for Spark Protection Board, while CARDIAC is the front-end electronic board. (b) Layout of the triple-GEM detector showing the three GEM foils and the corresponding gaps, together with their dimensions.

The Multi-Wire Proportional Chambers (MWPC) The muon system comprises a total of 1368 MWPCs⁷ made up of one or two layers. Each layer is made up of two adjacent 5 mm gas gaps with their corresponding readout electrodes hard-wired together in OR configuration, as depicted in Fig. 2.22a. The filling gas mixture was chosen to be Ar/CO₂/CF₄ (40:55:5), which allows fast drift time and the possibility to reach high gain with a low discharge probability. Inside the gaps, gold-plated tungsten wires with 30 μ m diameter and 2 mm spacing are tensed at 0.7 N and kept at an operating voltage of 2.6-2.7 kV. The resulting double gap layer reaches an efficiency higher than 95% in a 20 ns time window at a gain of $\simeq 10^5$. As already mentioned, M2-M5 stations employ chambers with two OR-ed layers (four OR-ed gaps), whereas M1 has one-layer chambers (two OR-ed gaps). Each layer has an independent readout and, to improve flexibility, each gap has its individual high voltage line. The efficiency of the chambers is summarised in Fig. 2.23 for each station and region.

The Triple Gas Electron Multiplier (GEM) chambers Twelve chambers made up of two OR-ed triple-GEM detectors [126] replace the MWPCs in M1R1, where the charged particle flux is as high as 500 kHz/cm². The triple-GEM de-

⁷ For a thorough review of the principles of operation of gaseous detectors, the reader is referred to [125].

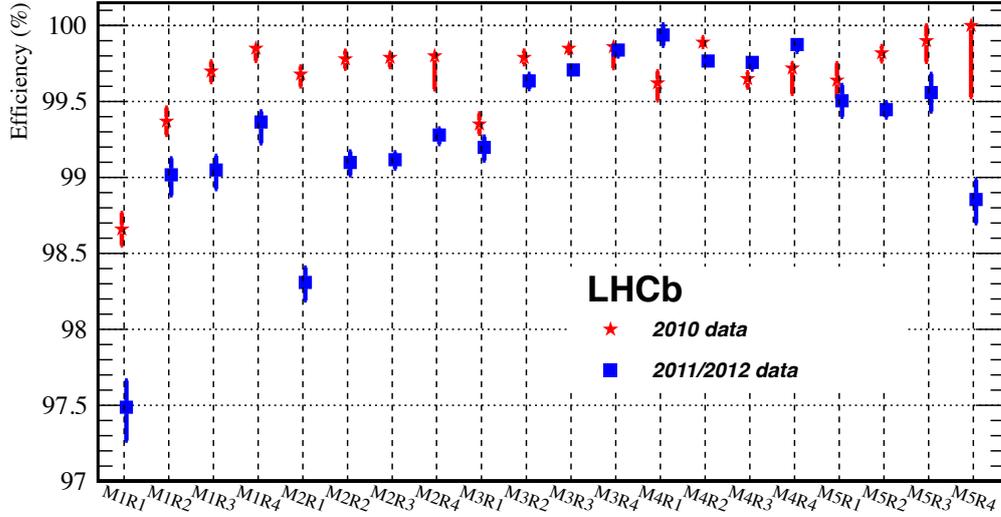


Figure 2.23: Average hit efficiency, measured in 2011 and 2012 data, for all the chamber types in the muon detector.

tectors are constituted by an anode and a cathode plane enclosing three GEM foils [127], according to the scheme in Fig. 2.22b. The GEM is a $50 \mu\text{m}$ Kapton foil clad on each side by $5 \mu\text{m}$ copper, on which bi-conical holes with external (internal) diameter of $70 \mu\text{m}$ ($50 \mu\text{m}$) and a pitch of $140 \mu\text{m}$ are applied. A charged particle traversing the gas inside the detector produces ionisation electrons, which drift towards the GEM foils, where a high voltage is applied. For a voltage of 500 V , the electric field inside the holes reaches $\sim 100 \text{ kV/cm}$, thus inducing the avalanche multiplication of the charges. The electron cloud moves through the other GEM foils where it is further amplified until its arrival at the induction gap, where it induces an electric signal on the anode. By virtue of the smallness of the holes, ions collection is fast and space charge effect, which affects MWPC under high particle rates, is averted. Moreover, the use of three GEM foils in cascade allows to split the total gain of the detector, hence reducing the discharge probability. For more details concerning GEM chambers and their performances at LHCb, the reader is referred to [128, 129].

Electronics

The readout of the electric signals generated inside the muon chambers starts with the FE boards, named CARDIAC. Each CARDIAC board comprises 16 inputs and 8 digital outputs and is directly plugged onto the chambers. The boards include two CARIOCA chips and one DIALOG chip, while a diode circuit, installed in a separate Spark Protection Board (SPB), protects the CARDIAC from sparks,

as shown in Fig. 2.22a. The CARIOCA⁸ is an Amplifier-Shaper-Discriminator (ASD) chip with 8 channels. The two CARIOCAs send their 16 digital signals to the DIALOG chip that performs the logical OR between the corresponding pads in the two layers of a chamber. In the regions R2-R4 of M2-M5 stations, an additional logical layer is provided by 168 Intermediate Boards (IBs), as logical channels might span more than one FE board. By means of a 4-bit TDC, the DIALOG allows to adjust the delays of its inputs so that a time alignment within 1.6 ns is possible, as demanded by the muon trigger to be fully-synchronous with the 25 ns machine clock. Moreover, the individual CARIOCA channel thresholds can be set by the DIALOG, whose control is managed by the Service Boards (SBs) via I²C communication. Each SB houses four Embedded Local Monitoring Boards (ELMBs) based on 8-bit microcontrollers. In each of the ten crates housing the total 156 SBs, a Pulse Distribution Module (PDM), again based on an ELMB, generates the low-jitter pulses phased with the LHC clock by means of a TTCrx chip.

The logical channels then reach the 152 Off Detector Electronics (ODE) boards, housed in the same crates of the IBs, which contain the Level-0 pipelines and DAQ interfaces embedded in the SYNC chips. They also include an ELMB for board control. The ODE board synchronises the signals via a TTCRx chip before routing them to the Level-0 trigger, while the data are parallelly stored into 4 μ s pipelines, awaiting for the positive trigger decision. A label containing the bunch crossing identifier is also applied at this stage. Once a positive decision is made, the data are written into a FIFO, capable of storing up to 16 data words: such derandomiser allows a regular readout rate of 1.1 MHz, providing a safe margin against the average trigger rate of 1 MHz. The data are formatted and sent to the TELL1 boards, in which they are processed by FPGAs and finally dispatched to the DAQ by means of Gb Ethernet ports.

In Fig. 2.24 the architecture of the muon detector readout electronics is shown. Since FE boards operate under a high particle flux, all the employed ASICs were produced using radiation-hard technology and employ triple-voting and auto-corrected registers to improve the Single Event Upset (SEU) immunity. Besides a strongly reduced radiation dose, near detector electronics also exhibit radiation resistant FPGAs.

2.5 The Trigger System

As described in Sec. 2.1.1, B meson production occurs in a small fraction of the pp collisions: about 0.6% at a centre-of-mass energy of 13 TeV. Moreover, the

⁸GEM chambers employ a dedicated CARDIAC-GEM board which integrates the spark protection circuit as well as a special CARIOCAGEM chip with a lower charge threshold.

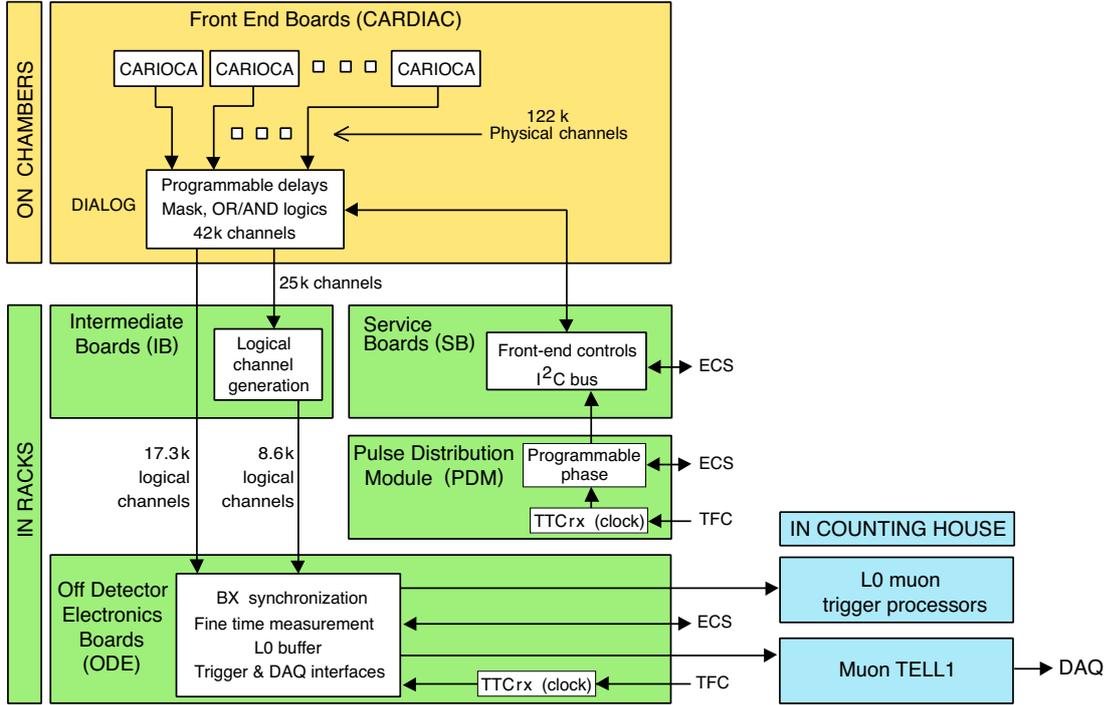


Figure 2.24: Architecture of the muon system electronics.

interesting physics channels may represent a small subset of the B meson decays, which is especially true in the study of rare decays. To isolate the interesting events among the vastly dominating background, LHCb employs a trigger system [130] structured in Level-0 (L0) and High Level Trigger (HLT).

Only a fraction of the total interactions is actually *visible* by the LHCb detector, where an interaction is said to be visible if at least two charged tracks are reconstructed in the spectrometer. As a first step, the rate of visible interactions, of the order of 20 MHz [131], is brought down to 1 MHz by the L0 trigger. At this rate, the entire detector can be readout, allowing for a more refined selection by the HLT, which sends data to the storage at a rate of ~ 12.5 kHz.

2.5.1 The Level-0 Trigger

The Level-0 trigger is hardware implemented in custom made electronics and designed to operate synchronously with the LHC bunch crossing rate of 40 MHz. Due to the large mass of the B meson, its decay products are often characterised by large transverse momentum (p_T) and transverse energy (E_T). In addition, the VELO pile-up system is able to provide the number of primary interactions, while the event multiplicity is measured by the number of hits in the SPD. The L0 ex-

exploit these global features to reduce the input rate to the HLT. The L0 architecture is highly parallelised and divided into three separate components:

1. Level-0 pile-up,
2. Level-0 calorimeter,
3. Level-0 muon,

connected to their respective detectors. A Decision Unit (DU) gathers all the informations and evaluates the final decision. The L0 decision must arrive within $4 \mu\text{s}$ from the bunch crossing, i.e. its latency is fixed to $4 \mu\text{s}$. Once the time-of-flight of the particles, the cable and the FE electronics delays are taken into account, $2 \mu\text{s}$ are left for processing the data and derive a decision. The Level-0 pile-up system is not used for flavour physics and will therefore be neglected in the following.

The Level-0 calorimeter

The calorimeter component of the L0 trigger uses the informations of the SPD, PS, ECAL and HCAL detectors. Since the size of an ECAL cell is approximately one Moliere radius, the E_T is computed in 2×2 cell blocks, as they contain most of the energy and avoid overlaps with neighbour showers. The transverse energy is defined as

$$E_T = \sum_{c=1}^4 E_c \sin \theta_c, \quad (2.13)$$

where E_c is the energy deposited in a cell and θ_c defines the cell orientation with respect to the z axis. Only the candidate with the highest transverse energy is kept. The E_T information is merged with the PS and SPD informations to infer the particle type: hadron, photon or electron according to:

- **L0Hadron** is the candidate with the highest E_T cluster in HCAL. The highest E_T of ECAL is added to the candidate if the corresponding cluster is located in front of the HCAL cluster.
- **L0Photon** candidate is the highest E_T cluster in ECAL with 1 or 2 (up to 4 in the inner zone) PS hits in front of the cluster and no hits in the SPD cells aligned with the PS ones. The E_T of the photon is the E_T as measured by the ECAL alone.
- **L0Electron** candidate has the same requirements of a **L0Photon** with the extra request of at least one SPD cell hit in front of the PS cells.

The SPD also records the total number of hits, providing information on the event multiplicity.

Typical thresholds for the L0 calorimeter trigger are $E_T > 4$ GeV for hadrons and $E_T > 2.7$ GeV for electrons and photons, while the number of hits in the SPD is required to be smaller than 450 to reject crowded events which are difficult to reconstruct.

The Level-0 muon

The L0 muon trigger searches for high p_T muon tracks traversing the five muon stations. The search starts in the central muon station: each fired logical pad in M3 defines a *seed* from which a track is searched. Extrapolation points are defined in the M2, M4 and M5 stations along a straight line connecting the track seed to the interaction point. In these stations, hits are searched for in a Field Of Interest (FOI) defined around the extrapolation points. By making a straight-line extrapolation from M2 and M3, a further hit is searched for in a FOI on the M1 station. The direction indicated by the selected hits in M1 and M2 is then used to measure the track p_T for the L0 trigger, with a precision of about 20%. A cartoon of the muon track finding is shown in Fig. 2.25. Each Processing Unit

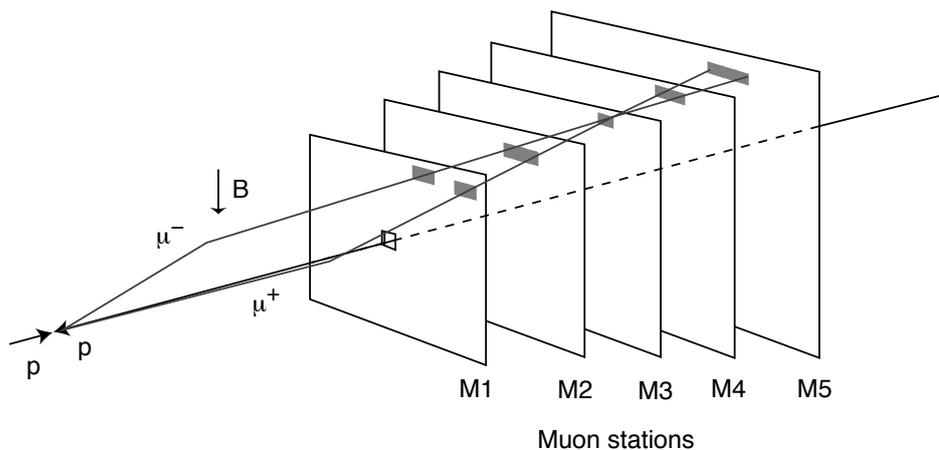


Figure 2.25: Scheme of the L0-muon trigger [130].

(PU) composing the Level-0 muon trigger performs the track search in a *tower* pointing to the interaction point, which is made up of 288 logical pads across the five muon stations. The track search is performed in parallel in the 192 towers in which the muon system is segmented. Moreover, by exploiting the projectivity of the logical layout, i.e. pad mapping between stations, the track search is performed using logical operations only.

The L0 muon trigger sets a threshold on either the largest p_T candidate (L0Muon)

or on the product of the largest and second largest p_T candidates (L0DiMuon), with typical values of $p_T > 2.8$ GeV and $\sqrt{p_{T_1} \times p_{T_2}} > 1.5$ GeV, respectively. The L0 trigger efficiency on $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays is at the level of 80%.

2.5.2 The High Level Trigger

The L0 output data are sent at a rate of 1 MHz to the Event Filter Farm (EFF), where the HLT software [130, 132] performs the event reconstruction in two steps of increasing complexity: HLT1 and HLT2. The EFF was improved during the first long shutdown of LHC (2013-2014) and now consists of approximately 1800 nodes and a total of 5.2 PB of hard disk space. Each server node contains 12-16 physical processor cores and 24-32 logical cores. HLT has access to the full event information and uses the same algorithms that are used in the offline reconstruction. For some data, the quality of the HLT2 reconstruction is high enough that no offline reconstruction is needed, as discussed in Sec. 2.6.

A sequence of reconstruction algorithms and selections defines a *trigger line*, while combinations of trigger lines together with a L0 configuration form a unique Trigger Configuration Key (TCK). A TCK is a 32 bit word pointing to the database that stores all the parameters that configure the trigger lines.

HLT1

The first sequence of HLT, named HLT1, performs a partial event reconstruction and selects displaced and/or high p_T charged particles and high E_T photons. A full 3D pattern recognition of all events entering the HLT1 from L0 is carried out by the VELO reconstruction software. During the LHC fill, the mean position of the pp interaction region is determined by using VELO tracks and is found to be stable within few μm for the fill duration. VELO tracks are also used to reconstruct vertices, and those which are less than 300 μm away from PV are equally marked as primary vertices.

VELO tracks are matched to hits in the TT stations, which lie in the fringe of the magnetic field and allow for a first estimate of track charge and momentum, with a relative uncertainty of around 20% on the latter. The charge estimate allows to reduce the size of the hit search window in the downstream tracking stations, which in turn allows to use lower p_T thresholds.

Different types of tracks are defined according to particle trajectories, as depicted in Fig. 2.26. Track candidates are then fitted using a Kalman filter [134] which accounts for multiple scattering and corrects for ionisation energy losses. The resolution on the track parameters achieved at the HLT1 stage is sufficiently high to allow for selective cuts on IP, momentum and invariant mass. Moreover, an inclusive selection named *TrackMVA* searches for two-track combinations that form a

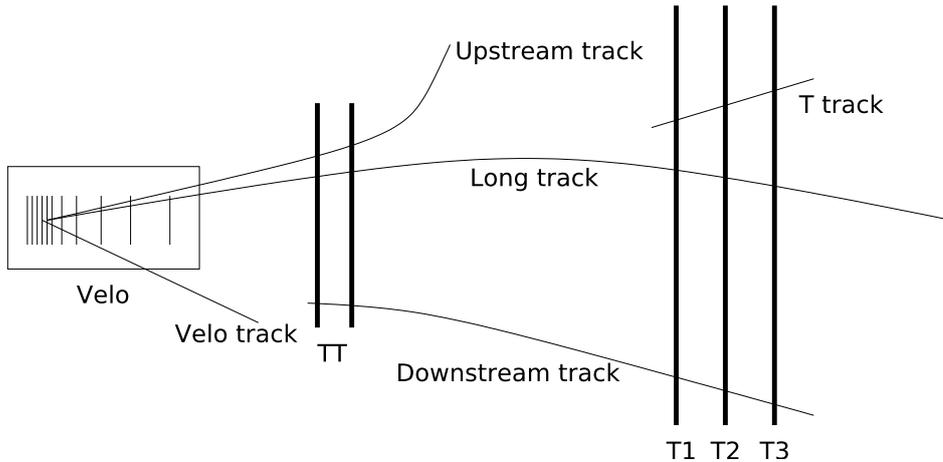


Figure 2.26: Track types at LHCb [133].

good quality vertex, and both the one- and two-track triggers make use of multivariate classifiers with improved performance over simple cuts [135].

A track can also be reconstructed from random combination of hits in the tracking stations (*ghost track*). A tool has been developed at LHCb to identify ghost tracks: a multivariate operator called *ghost probability* combines track and event parameters and returns a variable which is used as a discriminant.

To define a muon candidate at the HLT1 level, a positive match between the track extrapolation points in the muon stations and the hits in the detector is required, as will be explained in Sec. 3.1.1. Different muonic HLT1 lines contribute in selecting $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays. `HLT1TrackMuon` line selects candidates having $p > 6$ GeV/c and $p_T > 600$ MeV/c, with an $\text{IP}\chi^2$ larger than 7.4: this quantity measures how much a track is detached from the PV by repeating the PV fit with an without the track in question. Only good quality tracks, fulfilling $\chi^2/\text{ndof} < 3$, are retained. `HLT1TrackMVA` employs similar cuts and profits from several other event informations to further improve the selection. Both these single-muon lines have typical efficiencies of 94% on $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays. In addition, the `HLT1DiMuonHighMass` line selects muon couples having invariant masses greater than 2.9 GeV/ c^2 without the need of an IP cut. The two tracks are required to have a distance of closest approach (DOCA) of less than 0.2 mm. The resulting efficiency on $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ events amounts to $\sim 80\%$.

HLT2

The output rate of the HLT1 stage (~ 150 kHz) allows the HLT2 to perform the full reconstruction of RICH detectors and calorimeters [136].

The HLT2 is comprised of many trigger lines that can be grouped into *inclusive*

trigger lines and *exclusive trigger lines*.

- Inclusive trigger lines are defined by topological cuts aimed to trigger on partially reconstructed b -hadron decays, and represent the main resource for b -physics analyses. These lines select all b -hadrons with at least 2 charged particles in the final state and a displaced decay vertex. The *topological tracks* are selected with additional requirements on track fit quality (e.g. χ^2/ndf), IP and particle identification. N-body objects are built by requiring a small DOCA between the daughter tracks. The same multivariate classifiers techniques adopted in HLT1 are used here to improve the HLT2 trigger performances [135].
- Exclusive lines target specific final states and require all particles to be reconstructed, like it happens for prompt charm decays.

To exploit the inter-fill time of LHC, a buffer stores the HLT1 processed data which are produced synchronously with the collisions (*deferred trigger* [137]). The HLT2 later processes these data asynchronously so that the idle time of the EFF is reduced, i.e. HLT2 tasks run at lower priority in the farm independently from the data acquisition process. This allows a precise detector alignment and calibration procedure to run between HLT1 and HLT2 [138], ensuring offline-quality to the HLT2 reconstruction. Output data from HLT2 are written in the storage at a rate of ~ 12.5 kHz, where they can be further analysed offline.

Concerning the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ processes, the main HLT2 line is HLT2DiMuonB, which selects dimuon tracks with an invariant mass greater than 4.7 GeV/ c^2 and requires good quality dimuon vertices with the cut $\chi_{vtx}^2/ndf < 10$. The resulting signal efficiency is at the level of 100%.

2.5.3 Trigger efficiency measurement from data

As described throughout Chapter 2, in order for a particle to be detected it must first lie within the detector acceptance, then be triggered, reconstructed and eventually pass the offline selection. The trigger efficiency can be expressed as the fraction of triggered events among the number of signal events contained in the acceptance, but the latter is not observable as only triggered events are recorded. For this reason, the trigger efficiency is usually estimated from simulated samples after the selection. Nonetheless, to validate the simulation of the trigger process, a fully data-driven technique, named TISTOS method [139], has been developed at LHCb to measure the trigger efficiency.

Principles of the TISTOS method

Events accepted by the trigger can be split into three categories:

1. Triggered On Signal (TOS): events for which the signal presence is sufficient to induce a positive trigger decision.
2. Triggered Independent of Signal (TIS): the rest of the event is sufficient to induce a positive trigger decision, where the rest of the event is obtained by removing the signal and all detector hits belonging to it from the event.

Events which are simultaneously TIS and TOS (TISTOS) are those for which both the signal and the rest of the event are able to generate a positive trigger decision. By making use of these categories, the trigger efficiency can be written as

$$\epsilon_{Trig} = \frac{N_{Trig}}{N_{Sel}} = \frac{N_{Trig}}{N_{TIS}} \times \frac{N_{TIS}}{N_{Sel}} = \frac{N_{Trig}}{N_{TIS}} \times \epsilon_{TIS}, \quad (2.14)$$

where it is implied that all efficiencies are evaluated on a sample of selected events N_{Sel} , i.e. the conditional $|Sel$ has been omitted in Eq. 2.14. ϵ_{TIS} is not measurable from data but can be determined within the TOS subsample:

$$\epsilon_{TIS|TOS} = \frac{N_{TISTOS}}{N_{TOS}} = \epsilon_{TIS}, \quad (2.15)$$

where the last equality holds if ϵ_{TIS} is independent from the criterion used to select the signal. To this end, what is typically done is to subdivide the phase space into bins of the quantity used by the trigger to select events (e.g. p and p_T), so that in each bin the signal and the rest of the event can be assumed to be uncorrelated. The trigger efficiency can therefore be computed as

$$\epsilon_{Trig} = \frac{N_{Trig}}{N_{TIS}} \times \frac{N_{TISTOS}}{N_{TOS}}. \quad (2.16)$$

2.6 Offline Analysis

Triggered events from HLT2 follow three data streams towards the storage [140].

- **FullStream**: this is typically used for signals that are not fully reconstructed, such as b -hadron decays selected by the inclusive trigger or for analyses that need to (re-)run dedicated event reconstructions, as in the case of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis.⁹
- **TurboStream**: this is used for event selections which do not need to re-run an offline dedicated reconstruction, so that the raw event information can be discarded.

⁹ The raw event information was needed to compute the isolation variable described in Sec. 4.2.2.

- **TurboCalib:** this is a collection of calibration selections for online monitoring and measurement of particle identification (Sec. 3.2.1) and tracking efficiencies. Differently from the TurboStream, the raw event information here is retained and the offline reconstruction is also performed.

Events from each stream are permanently saved in the LHC grid computing infrastructure [141] in the Data Summary Tape (DST) format.

Before being analysed, LHCb data are splitted into *stripping lines*. Each stripping line is characterised by a series of cuts used to build specific candidates and to perform a first selection of interesting events.

Monte Carlo data

Together with real data acquisition, a large amount of simulated data is produced at LHCb for control purposes. Such data are produced with Monte Carlo (MC) simulations of physical processes and detector interactions. At LHCb, the MC simulation is handled by the GAUSS project [142], which makes use of the PYTHIA [95] generator and the EvtGen package [143] for particle generations and decays, where the bremsstrahlung is simulated by the PHOTOS [54] package. The simulation of particle interactions in the detector is carried out with the GEANT4 toolkit [144].

In conclusion of the chapter devoted to the detector description, a typical LHCb event display is shown in Fig. 2.27.

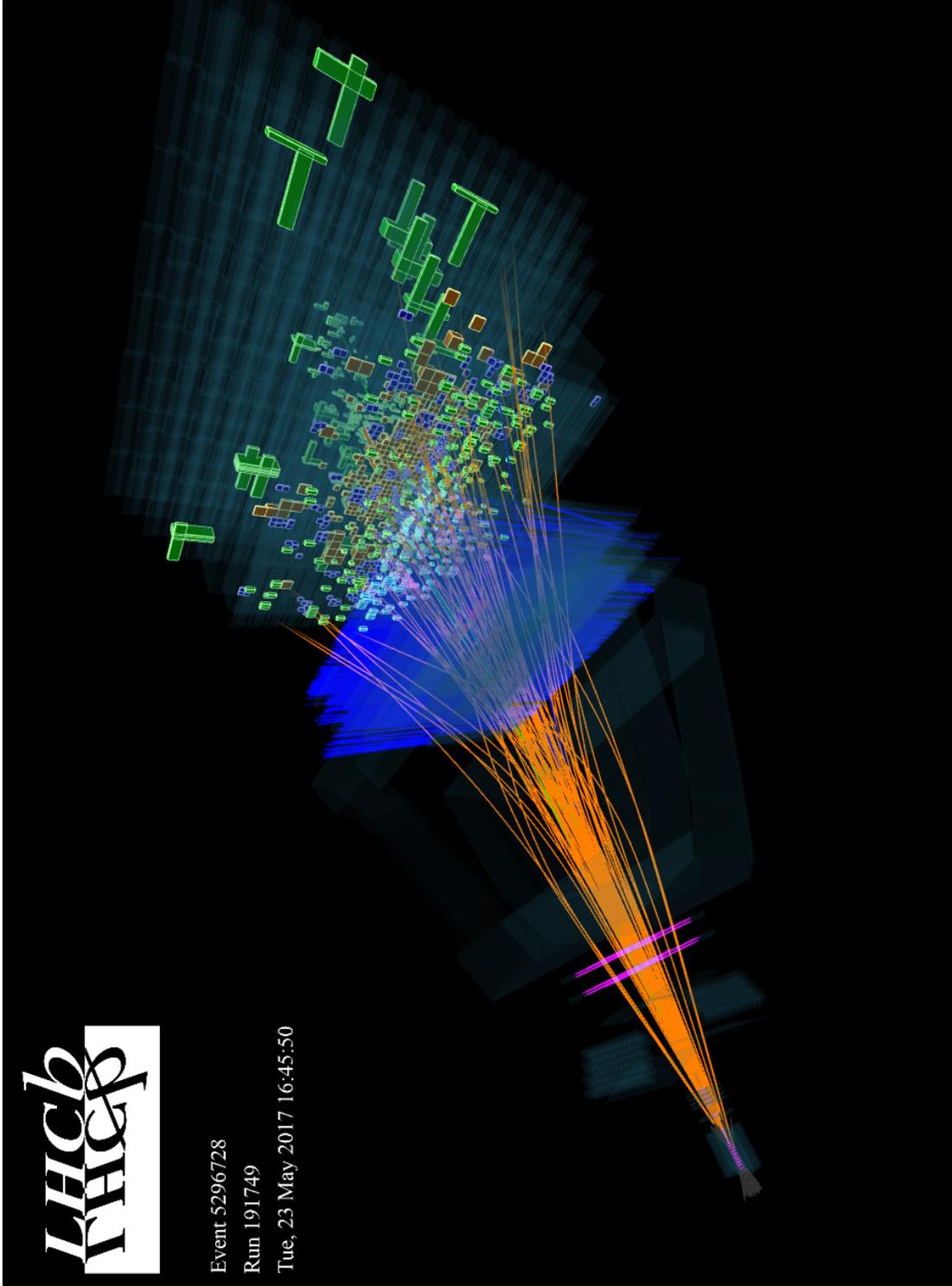


Figure 2.27: Display of a typical LHCb event, recorded during the first day of data taking in 2017.

Overview of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis

In this paragraph, a brief overview of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis is given. The whole analysis strategy is developed by keeping the signal region ($m(\mu^+ \mu^-) \in [5200, 5445]$ MeV/c²) blinded, to avoid the non-measurable experimenter's bias [145].

The analysis procedure starts with a loose event selection, aimed to reduce the data sample size to a manageable level while keeping a very high signal efficiency. The selection is mainly based on topological cuts that exploit the characteristics of a two-body B meson decay, and a loose cut on a multivariate classifier is already applied at this stage. Besides this, a strong requirement on the muon identification is imposed, in addition to the loose muon hit matching criterium applied at the HLT1 level. This is obtained by combining the informations from all the PID detectors into a neural network. As a result, all background sources from exclusive decays with misidentified hadrons in the final state are strongly reduced, at the price of a reasonable signal loss.

The selected events are then classified according to their dimuon invariant mass and to the output of a Boosted Decision Tree (BDT), which makes use of the topological and kinematic information of the event. The BDT is very powerful in separating the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signals from the overwhelming combinatorial background, where the two oppositely-charged muons arise from two different B decays. The BDT output is shown in Fig. I, with signal events having a flat distribution (by construction) and background events strongly peaking at zero. The number of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signal events is evaluated via an invariant mass fit performed simultaneously in different BDT regions of increasing signal sensitivity, in order to fully exploit the statistical power of the data samples. The signal mass shape used in the fit is calibrated on data. While the combinatorial background is left free in the signal fit, all the other exclusive background sources must be accurately estimated and are constrained in the fit.

Several decays are found to be relevant, some of which also reaching the signal mass region. To the above category belong the $B_{d,s}^0 \rightarrow h^+ h'^-$ decays ($h = \pi, K$) which, in the event that both hadrons are misidentified as muons, give origin to a broad

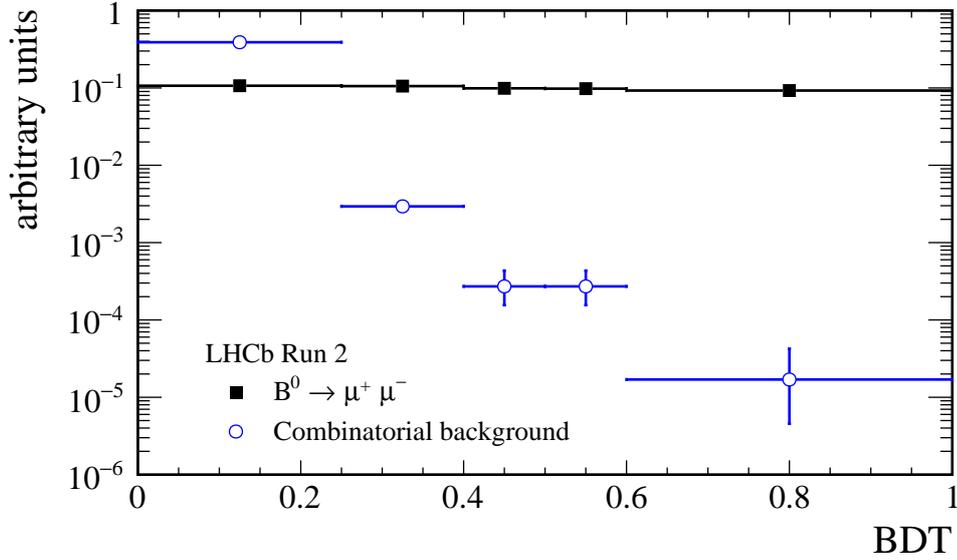


Figure I: BDT distribution of combinatorial and $B_d^0 \rightarrow \mu^+ \mu^-$ signal events.

peak almost centered below the $B_d^0 \rightarrow \mu^+ \mu^-$ signal peak. A careful evaluation of this contribution is made by using $B_{d,s}^0 \rightarrow h^+ h'^-$ data and double misidentification rates, as measured on data calibration samples. In addition, semileptonic decays, with one or two real muons in the final state, mainly populate the left invariant mass sideband, thus influencing the combinatorial background determination in the fit. Their yields are evaluated using Monte Carlo simulation, and normalising to the $B^+ \rightarrow J/\psi K^+$ channel. For the $B_{d,s}^0 \rightarrow h^+ h'^-$ and the most abundant semileptonic decays, an independent data-driven estimate has also been carried out, so that a robust systematic uncertainty can be assigned. To compute the branching fractions, the number of signal events are normalised to well-known processes. Two normalisation channels, $B^+ \rightarrow J/\psi K^+$ and $B_d^0 \rightarrow K^+ \pi^-$, are employed for this purpose, their selection being as similar as possible to that of the signal, to allow for cancellation of systematic uncertainties. Once the full analysis procedure has been thoroughly developed, the data are unblinded and the signal fit is performed.

For the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime measurement, a BDT cut has been firstly performed to select the most sensitive region. Then, with the *sPlot* [146] technique, an invariant mass fit is carried out to determine the signal decay time distribution, which is then fitted by taking into account the lifetime dependent acceptance.

The $B_{d,s}^0 \rightarrow \mu^+\mu^-$ data sample

The data used for the analysis correspond to a total of 4.4 fb^{-1} , as reported in Table I. Note that Run 2 data collected during the years 2015 and 2016 are enriched in B meson candidates due to the increase of the $b\bar{b}$ production cross section, as discussed in Sec. 2.1.1.

Year	\sqrt{s} (TeV)	\mathcal{L} (fb^{-1})
2011	7	1.0
2012	8	2.0
2015	13	0.3
2016	13	1.1

Table I: Data collected during the LHC Run 1 and Run 2 used for the present analysis.

Assuming the SM branching fraction for both $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ decays, the following number of signal events are expected with the total available statistics:

$$N_{\text{expected}}^{SM}(B_s^0 \rightarrow \mu^+\mu^-) = 62.2 \pm 5.6,$$
$$N_{\text{expected}}^{SM}(B_d^0 \rightarrow \mu^+\mu^-) = 6.7 \pm 0.6.$$

Chapter 3

Muon Identification

The muon identification (muonID) is one of the LHCb strengths as well as a key ingredient for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis, and will be discussed in this chapter. With respect to the previous analysis [4], a significant improvement has been made in the muon selection, which led to an important reduction of the exclusive backgrounds. Besides this, particular attention has been also made in reducing possible systematic effects in the background evaluation, which mostly relies on the correctness of the hadron to muon misidentification probability estimate. The first part of this thesis work was concentrated on both the above aspects, which will be described with some detail in the following.

3.1 The muonID algorithm

The muonID algorithm [147] is essentially realised in two steps: a binary selection called IsMuon and the computation of a likelihood for the muon hypothesis named muDLL.¹ The latter is then combined with the log-likelihoods of the RICH and CALO, either as a plain sum, called DLL, or by using a neural network, whose output is called ProbNN, where in this last case other informations from the tracking are also used. The definition of the muonID variables is given in this chapter, highlighting their performances.

3.1.1 IsMuon selection

With the exception of the elusive neutrinos, muons are the only particles able to traverse the whole LHCb detector. Muons mainly loose energy due to ionisation, as radiative losses start to be significant for momenta above hundreds of

¹The muDLL is defined as the logarithm of the ratio between the likelihoods of the muon and non-muon hypotheses.

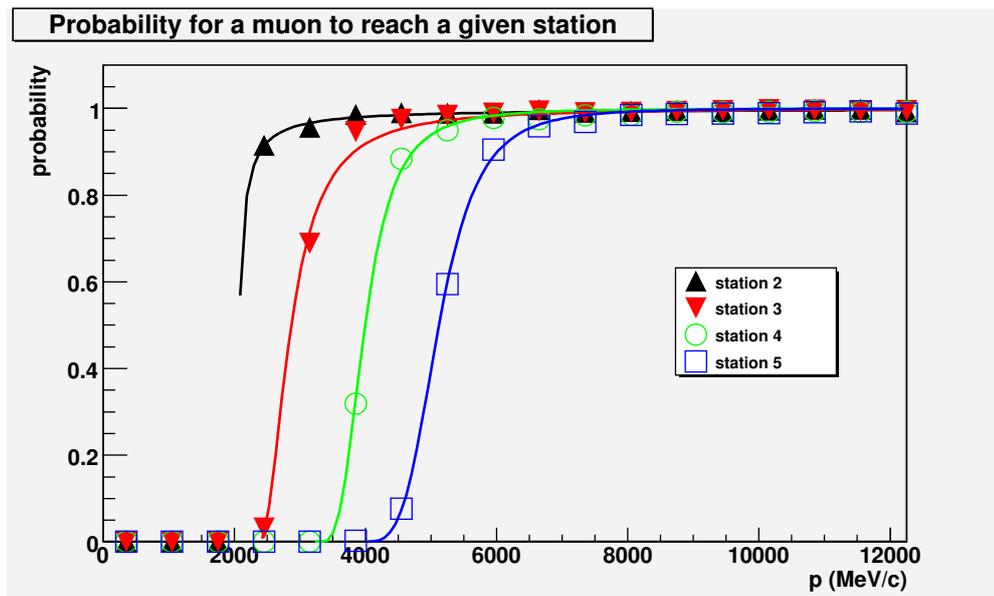


Figure 3.1: Probability for a muon to reach a given station as a function of its momentum, from a simulated $J/\psi \rightarrow \mu^+\mu^-$ sample [148]. Note that the calorimeters, located upstream of M2, accounts for almost 7 interaction lengths (Sec. 2.4.2), while the calorimeters plus iron absorbers thickness amount to 20 interaction lengths.

GeV/c.² Being Minimum Ionising Particles (MIPs) for momenta greater than ~ 100 MeV/c [39], muons with $p \sim$ few GeV/c can cross the LHCb CALO system. Depending on their momentum, they can also cross the iron absorbers of the muon stations, as shown in Fig. 3.1: for $p > 6$ GeV/c, a muon has a probability larger than 95% to reach the M5 station. For muons coming from $B_s^0 \rightarrow \mu^+\mu^-$ decays, this occurs about 98% of the times, as can be deduced from the simulated spectrum shown in Fig. 3.2.

The IsMuon selection aims at identifying muon candidates by exploiting their penetration power: the algorithm is a binary selection which depends on the number of muon stations traversed by a muon candidate as a function of its momentum, according to Table 3.1. The muon hit matching is performed on each station within a field of interest (FOI) around the track extrapolation. The x and y sizes of the FOI depend on the particle momentum and on the expected multiple scattering which muons undergo while traversing the iron absorbers. FOI parameters are

² The critical energy of a muon, at which radiative and ionisation losses are equal, is about 350 GeV/c for iron ($Z = 26$) [39].

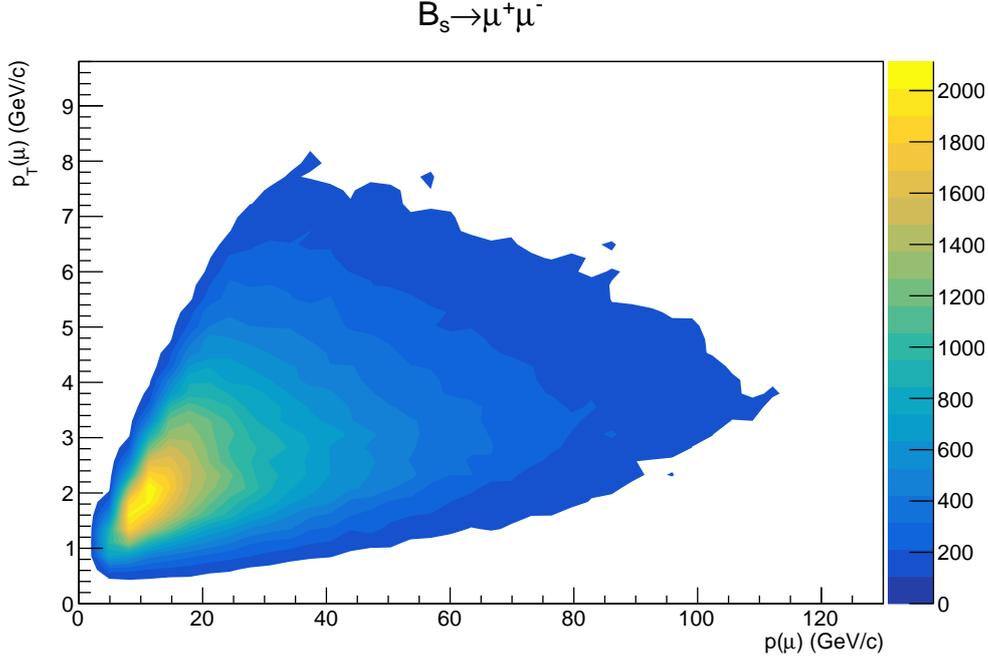


Figure 3.2: Simulated momentum and transverse momentum of muons originating from $B_s^0 \rightarrow \mu^+ \mu^-$ decays at $\sqrt{s} = 13$ TeV. Both tracks are required to lie within the muon detector acceptance.

separately tuned in the four muon detector regions as

$$\text{FOI} = a + be^{-cp}, \quad (3.1)$$

where p is the momentum and the numerical values of a , b and c are determined from a full Monte Carlo simulation of the LHCb detector.

The IsMuon selection efficiency and hadron to muon misidentification probabilities, evaluated from calibration samples (as explained in Sec. 3.2.2) are shown in Fig. 3.3. The average efficiency is about 98%, while the average misidentification probability is at the level of 1%. The efficiency values drop slightly below unity only in the low p_T regimes, where tracks might easily be scattered outside the acceptance. For protons, the misidentification probability is largely due to combinatorial hits, and its increase at low p and p_T is provoked by the larger FOI size. For pions and kaons, there is an additional contribution due to decays in flight, which is dominant at low momentum.

Momentum range (GeV/c)	Muon stations
$3 < p < 6$	M2 and M3
$6 < p < 10$	M2 and M3 and (M4 or M5)
$p > 10$	M2 and M3 and M4 and M5

Table 3.1: Muon stations that a muon candidate must traverse to satisfy the IsMuon criterion, as a function of the momentum. 3 GeV/c is the threshold to traverse the first absorber, while 6 GeV/c are needed to reach the last station, as shown in Fig. 3.1.

3.1.2 Muon Likelihood

For tracks accomplishing the IsMuon request, a more refined selection is made by exploiting the pattern of hits in the muon stations by means of the D^2 variable. The D^2 is the average squared distance significance of the hits in the muon detector with respect to the linear extrapolation coming from the tracking system:

$$D^2 = \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{x_{\text{closest}}^i - x_{\text{track}}^i}{\text{pad}_x^i} \right)^2 + \left(\frac{y_{\text{closest}}^i - y_{\text{track}}^i}{\text{pad}_y^i} \right)^2 \right]. \quad (3.2)$$

The i index runs over the N stations containing hits inside the FOI, while the *closest* coordinates are the coordinates of the closest hit to the track extrapolation points. The hit residuals are normalised to the chamber pad size in the x and y directions.

The D^2 distribution for true muons exhibits a narrow peak at 0, while hadrons selected by the IsMuon criterion tend to have a broader distribution, as shown in Fig. 3.4a. A likelihood is then defined as the cumulative distribution of the D^2 , i.e. the integral of the D^2 distribution from 0 to the observed D_{obs}^2 . The likelihood for the muon hypothesis is evaluated with muons from $J/\psi \rightarrow \mu^+ \mu^-$ decays. The likelihood for the non-muon hypothesis is computed using the D^2 distribution of protons from a simulated sample of $\Lambda \rightarrow p\pi^-$ decays. The choice of calibrating the non-muon hypothesis with a sample of protons is dictated by the fact that pions and kaons have a true muon component stemming from their decays in flight. Finally, the muDLL variable is defined as the logarithm of the ratio between the likelihoods for the muon and non-muon hypotheses: its discriminating power is evident from Fig. 3.4b. For pions and kaons, a very clear component due to decays in flight, being true muons in the detector, is also visible.

Note that, as the D^2 depends on multiple scattering, hence on the momentum and polar angle distributions of the calibration tracks, different tunings of the muon and non-muon hypotheses are made separately in momentum bins for each

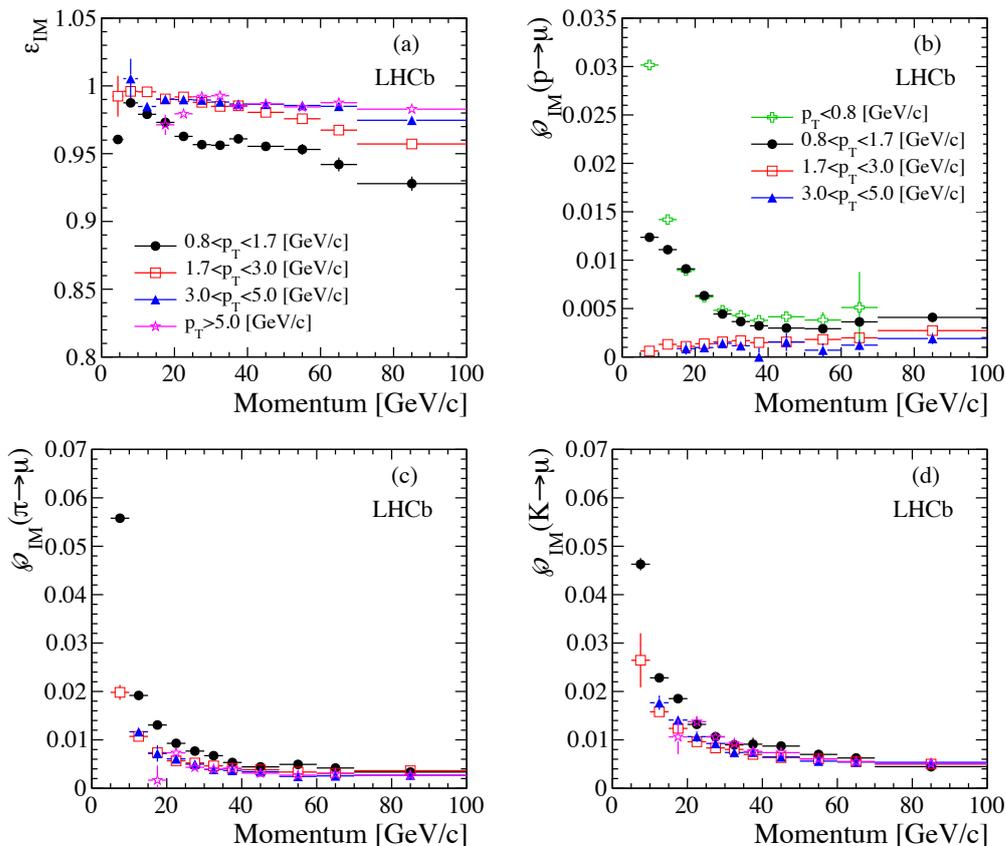


Figure 3.3: IsMuon efficiency (a) and proton (b), pion (c), kaon (d) misidentification probabilities as a function of momentum for different transverse momentum regions. Non-physical efficiency values may arise from the background subtraction procedure, as discussed in Sec. 3.2.2.

detector region.

3.1.3 Combined PID

The discriminating power of the muon and non-muon likelihoods are improved by their combination with the likelihoods computed in the RICH and CALO detectors. The Cherenkov angle measured in the two RICH detectors, combined with the particle momentum, is used to build a likelihood for each mass hypothesis (proton, pion, kaon, muon and electron), whose discriminating power is particularly strong in the low momentum regime, as shown in Fig. 2.16. On the other hand, the energy deposit in the calorimeters allows to separate muons, being MIPs, from electrons and hadrons.

The combined value of the log-likelihood is computed for each track by adding

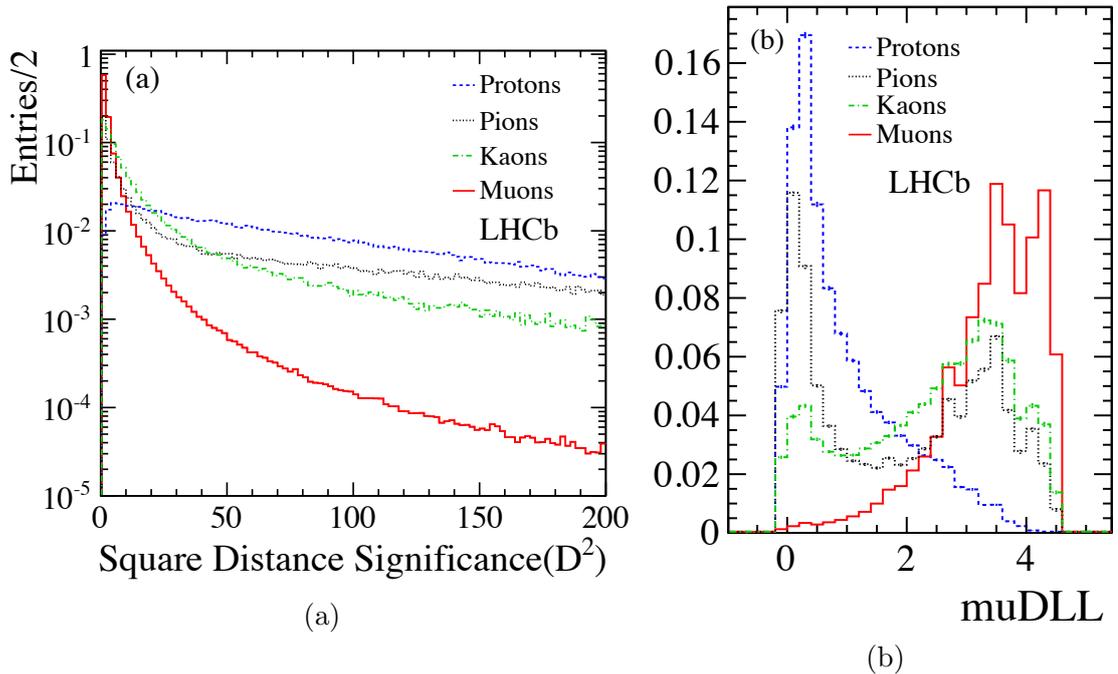


Figure 3.4: (a) D^2 distribution for protons, pions, kaons and muons. (b) The corresponding muDLL distribution. The values are computed on data from 2011 calibration samples after the IsMuon selection [147].

the log-likelihoods for each particle hypothesis. The difference of the combined log-likelihoods between muon and pion hypotheses is named $DLL(\mu - \pi)$.

Two main drawbacks affect the DLL variable. Firstly, it does not include the full detector information, and other bits might add discriminating power. In addition, the DLL neglects the correlation between the input variables. For these reasons, a neural network has been used to define new variables, named ProbNN [90], which use as inputs the DLLs from the individual PID detectors and also add more informations from the tracking system, such as track fit χ^2 and ghost probability.

The training of the neural network is performed on simulated samples of inclusive B decays, where the samples representing the signal and background depend on the particle type to be selected. In addition, different tunings of the same variables are defined, following the specific features of the different data taking periods. The tunings employed for the current analysis are named MC12TuneV2 and MC15TuneV1, which are trained with MC samples generated with Run 1 and Run 2 LHCb conditions, respectively. The NN has been developed with the TMVA package [149] within the ROOT framework [150].

	PID $_{\mu} > 0.2$	PID $_{\mu} > 0.3$	PID $_{\mu} > 0.4$
$B_{d,s}^0 \rightarrow h^+ h'^-$	53%	34%	22%
$B_d^0 \rightarrow \pi^- \mu^+ \nu_{\mu}$	71%	61%	50%
$B_s^0 \rightarrow K^- \mu^+ \nu_{\mu}$	77%	52%	35%
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_{\mu}$	30%	21%	15%
Signal efficiency	99%	95%	89%

Table 3.2: Background levels with respect to the DLL cut used in the previous analysis [4]. The signal efficiency is also indicated in the last line, again relatively to the DLL one.

3.1.4 The PID selection

The $B_d^0 \rightarrow \mu^+ \mu^-$ analysis performance critically depends on the ability to reject $B_{d,s}^0 \rightarrow h^+ h'^-$ events under the signal peak, as will be discussed in detail in Sec. 5.1. The signal sensitivity can also benefit from a better rejection of the semileptonic modes where a hadron is misidentified as a muon: as described in Sec. 5.2, these events pollute the left mass sideband, in which a strong correlation with the combinatorial background is present. In addition, the $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_{\mu}$ decays, albeit suppressed with respect to the $B_{d,s}^0 \rightarrow h^+ h'^-$, leak into the signal region.

In the last analysis [4], the PID selection after IsMuon consisted in requiring both tracks to satisfy $\text{DLL}(K - \pi) < 10$ and $\text{DLL}(\mu - \pi) > -5$, which ensured enough rejection on $B_{d,s}^0 \rightarrow h^+ h'^-$ events (about a factor 5 with respect to the IsMuon criterion only), while keeping a high $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signal efficiency, about 95%. With the aim of enhancing the $B_d^0 \rightarrow \mu^+ \mu^-$ sensitivity, tighter selection strategies have been investigated. To this purpose, different combinations of the ProbNN cuts were studied, and the operator

$$\text{PID}_{\mu} \equiv \text{ProbNN}_{\mu} \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) \quad (3.3)$$

was chosen due to its strong rejection of the $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_{\mu}$ background component. Table 3.2 reports the results of this study for different PID cuts, showing how the background levels get reduced with respect to the DLL benchmark values. Among the shown values, the chosen working point for the operator (3.3) was $\text{PID}_{\mu} > 0.4$, as toy MC studies conducted with Run 1 data yielded the highest sensitivity on the $B_d^0 \rightarrow \mu^+ \mu^-$ signal. When compared to the DLL, the $\text{PID}_{\mu} > 0.4$ selection rejects the $B_{d,s}^0 \rightarrow h^+ h'^-$ background by a factor of almost 5, at the price of a moderate signal loss of about 11%. The $B_d^0 \rightarrow \pi^- \mu^+ \nu_{\mu}$ and $B_s^0 \rightarrow K^- \mu^+ \nu_{\mu}$ get reduced by a factor ~ 2 and ~ 3 , respectively, due to the hadron in the final state. Moreover, the $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_{\mu}$ is reduced by a factor ~ 6 , reflecting the fact that the DLL cut used no proton information.

3.2 Measurement of the PID efficiency from data

The particle identification efficiency could be evaluated from MC sample, but the full simulation of the detectors devoted to particle identification, as well as their response to a traversing particle, is highly non-trivial. In the simulation, many effects have to be taken into account such as detector occupancy, alignments, temperatures and gas pressures, from which the detector response is dependent. The above consideration motivated the use of a data-driven technique to provide a precise and reliable measurement of the PID efficiency. To this end, many *calibration samples* act as proxies for the “*signal*” samples (*reference samples* in the following), so that an arbitrary PID selection can be studied.

3.2.1 Calibration samples

As anticipated in Sec. 2.6, a dedicated HLT2 stream called TurboCalib is devoted to the selection of calibration samples for PID efficiency estimation and for reconstruction studies. These HLT2 selections aim to collect pure samples of the most common charged, long-lived particle species produced in LHCb: protons, pions, kaons, muons and electrons, the latter being excluded from the following discussion as they are not relevant for the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis. Their purpose being the computation of PID efficiencies, calibration samples are exclusively selected by means of kinematic cuts and, where higher purity is required and the sample size is large enough, the *tag and probe* method is further applied. Moreover, special care is put to decorrelate the PID selection from the trigger, so that PID and trigger contributions to the total efficiency can be factorised.

For more details about the calibration samples, the reader is referred to [147] for Run 1 and [151] for Run 2 data.

Protons

Proton calibration samples are obtained from $\Lambda \rightarrow p\pi^-$ decays.³ In these decays, powerful background discrimination is provided by the long Λ lifetime, which allows a selection based on displaced decay vertices. As the Λ production cross section in LHCb is very high, a *prescale* of the corresponding HLT2 lines is mandatory. However, lines with different p_T requirements and prescale factors have been implemented, resulting in an optimised kinematic coverage. To increase the purity, the pion is *tagged* by means of a PID selection, while the other track constitutes the *probe*.

³ Unless explicitly stated, charge conjugation is implied hereinafter.

Pions and Kaons

An abundant source of K^\pm and π^\pm is provided by $D^{*+} \rightarrow \pi^+ D^0 (\rightarrow K^- \pi^+)$ decays, which are selected without the use of PID information for any tracks. High impact parameter is required for the D^0 daughter tracks, while the D^0 flight direction must point to the primary vertex. The decay vertices of the D^0 and D^{*+} must fulfill quality criteria, and a $25 \text{ MeV}/c^2$ invariant mass window around the nominal D^0 mass is used to exclude the doubly Cabibbo suppressed mode as well as the $K^+ K^-$ and $\pi^+ \pi^-$ channels.

Muons

A large source of muons arises from $J/\psi \rightarrow \mu^+ \mu^-$ decays. To reject most of the combinatorial background originating from the primary vertex, the J/ψ must have a large flight distance significance and good decay vertex quality, while the two muons are required to have a high impact parameter with respect to the primary vertex. In addition, the DOCA between the two tracks has to be small, and their invariant mass must lie within $210 \text{ MeV}/c^2$ from the nominal J/ψ mass. To further enhance the purity of the sample, one of the two muons is *tagged* with the IsMuon selection, while the other one, being the *probe*, is only required to have $p_T > 800 \text{ MeV}/c$.

Examples of the invariant mass distribution for the aforementioned calibration samples are shown in Fig. 3.5 for 2015 data.

Trigger decorrelation

Because the information arising from the CALO and MUON subdetectors is exploited in the L0 and HLT1 stages of the trigger, special attention is required to ensure that calibration samples are not biased by the trigger selections. To this end, TIS requirements (Sec. 2.5.3) are used when selecting muons and hadrons.

- **Muons** The charged track used to measure the PID efficiency, i.e. the *probe*, is required to be TIS with respect to L0 and HLT1. In a typical $J/\psi \rightarrow \mu^+ \mu^-$ event, one of the muons triggers L0 and HLT1 and the remaining one is therefore used for calibration.
- **Hadrons** When computing pion and kaon to muon misidentification probabilities, the trigger unbiasing is guaranteed by requiring the TIS condition on the probe track at L0 and HLT1 stages. Since a PID-trigger correlation is observed in the proton to muon misID, the HLT2 TIS condition also needs to be requested in this case.

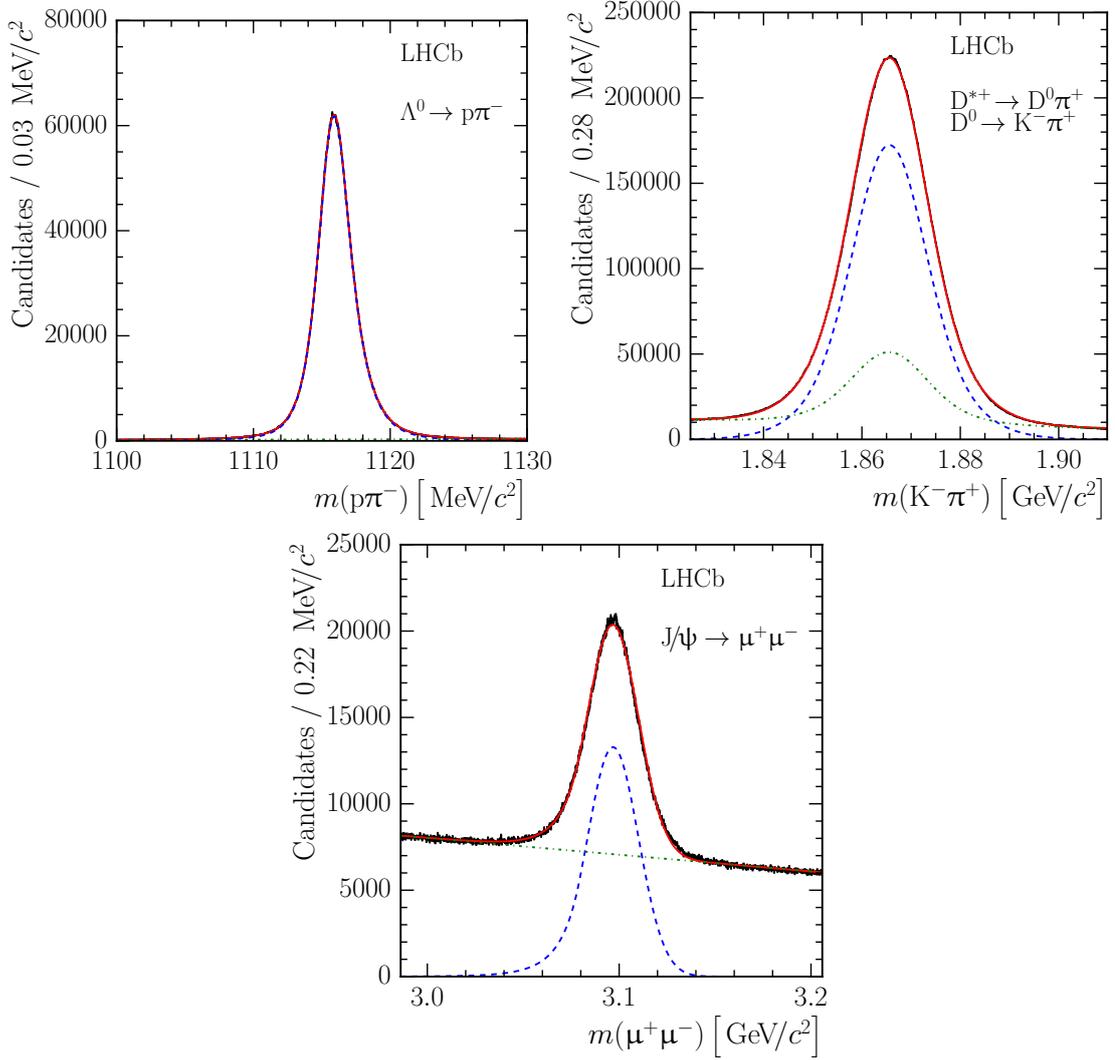


Figure 3.5: Invariant mass distribution for the $\Lambda \rightarrow p\pi^-$ (top left), $D^{*+} \rightarrow \pi^+D^0(\rightarrow K^-\pi^+)$ (top right) and $J/\psi \rightarrow \mu^+\mu^-$ (bottom) samples. The superimposed fit (red line) shows the signal (dashed blue) and background (dotted dash green) components. All distributions are populated with 2015 data [151].

3.2.2 PID efficiency computation

The PID response for a particle of a given species is not flat across the phase space: its efficiency in fact varies as a function of kinematic variables and also depends on event multiplicity. If the distribution of these variables is different between the calibration sample and the reference sample under scrutiny, the average PID efficiency will be different.

To overcome this misalignment, it can be assumed that the response of a PID variable is fully parametrised by a known set of variables such as p , p_T , η and event multiplicity, the latter being quantified for example by the number of hits in the SPD. If the calibration sample is then binned with sufficient granularity in the parametrising variables, the efficiency of a PID selection will be constant within each bin, and given by

$$\epsilon_i^{\text{PID}} = \frac{N'_i}{N_i}, \quad (3.4)$$

where N'_i is the number of probe tracks surviving the cut and N_i is the total number of probe tracks. To evaluate the PID efficiency on the signal sample, an average over the parameter space is defined according to:

$$\epsilon^{\text{PID}} = \frac{\sum_i R_i \epsilon_i^{\text{PID}}}{\sum_i R_i} = \frac{1}{R} \sum_i R_i \epsilon_i^{\text{PID}}, \quad (3.5)$$

where R_i denotes the number of signal tracks inside the i -th bin and R is the total number of signal tracks.

Background subtraction

Eq. (3.4) relies on counting the number of events before and after the PID selection, which is only possible if the sample of calibration tracks is pure. Each calibration sample inevitably retains a certain amount of background, for example due to random track combinations. The adopted strategy is to compute ${}_s\mathcal{Weights}$ [146] from a fit to the invariant mass of the full sample, so that the number of signal candidates in the i -th bin can be extracted as

$$C_i = \sum_{\text{cands} \in i} {}_s\mathcal{W}_i, \quad (3.6)$$

where ${}_s\mathcal{W}_i$ is the signal ${}_s\mathcal{Weight}$ for candidate i . Provided the PID selection is not correlated with the mass, the above Eq. (3.6) can be used to compute the signal events before and after the PID selection, thus giving the wanted efficiency.

The choice of the parameter space binning should guarantee enough statistics of calibration events and a granularity such that the efficiency within each bin can be treated as uniform. A dedicated package, named PIDCalib [152, 153], implements the above procedure and is available within the LHCb software.

3.3 PID efficiencies for $B_{d,s}^0 \rightarrow \mu^+ \mu^-$

Several PID selection efficiencies are needed in the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis and will be used throughout the next chapters:

- The PID selection efficiency on the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signals, i.e. the single muon efficiency of the PID_μ cut (Eq. 3.3) convoluted with the $B_s^0 \rightarrow \mu^+\mu^-$ kinematic spectra⁴ (Sec. 3.3.1);
- the PID selection efficiency of the IsMuon request on the two muons of the $B^+ \rightarrow J/\psi K^+$ normalisation channel (Sec. 4.4);
- the PID_μ cut selection efficiency for double-hadron to muon misidentification in case of $B_{d,s}^0 \rightarrow h^+h^-$ decays (Sec. 5.1);
- the PID_μ cut selection efficiency for single-hadron to muon misidentification for various semileptonic backgrounds (Sec. 5.2);
- the PID selection efficiency for the normalisation/control channel $B_d^0 \rightarrow K^+\pi^-$ (Sec. 4.2.4);

In this section, the method to evaluate the PID efficiency on a specific channel is presented, which essentially consists in convoluting the data-driven efficiencies with the simulated kinematic spectrum of the process of interest.

The calibration data samples used to evaluate the PID efficiencies correspond to the full statistics of Run 1 and 2015, while 600 pb⁻¹ of calibration data were available for 2016. They are split into $p - p_T$ subsets according to the following *binning scheme*:

$$\begin{aligned}
 p &= [2, 5] [5, 10] [10, 15] [15, 20] [20, 25] [25, 30] \\
 &\quad [30, 35] [35, 40] [40, 50] [50, 60] [60, 500] \text{ GeV}/c; \\
 p_T &= [0.8, 1.7] [1.7, 3] [3, 5] [5, 40] \text{ GeV}/c.
 \end{aligned} \tag{3.7}$$

For each one of these 44 bins, the PID efficiency is evaluated as described in Sec. 3.2.2, and a convolution of the resulting values with the $p - p_T$ spectra of the signal daughters is then performed.

To take into account the event-by-event correlations in propagating the calibration errors to the average signal PID efficiency, a toy MC technique is used:

- the signal daughter efficiencies in each kinematic bin are gaussianly sampled according to their error, and are then multiplied to give the per event PID efficiency, which is in turn averaged over the signal sample to give the signal PID efficiency corresponding to that particular toy;
- the first step is repeated for several toys, in which the sampling of the daughter efficiencies over the bin is changed;

⁴ To the purpose of evaluating the PID efficiency, the kinematic spectra of $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ processes are equivalent.

Cut name	Value	ProbNN Tuning	Applied on	Used for
$\text{PID}_{\mu,4}^{\text{MC12}}$	$\text{PID}_{\mu} > 0.4$	MC12TuneV2	Run 1, 2015	\mathcal{B}
$\text{PID}_{\mu,8}^{\text{MC15}}$	$\text{PID}_{\mu} > 0.8$	MC15TuneV1	2016	\mathcal{B}
$\text{PID}_{\mu,2}^{\text{MC12}}$	$\text{PID}_{\mu} > 0.2$	MC12TuneV2	Run 1, 2015	τ_{eff}
$\text{PID}_{\mu,4}^{\text{MC15}}$	$\text{PID}_{\mu} > 0.4$	MC15TuneV1	2016	τ_{eff}

Table 3.3: muonID cuts employed for the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ branching fraction (\mathcal{B}) and $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime (τ_{eff}) measurements on the analysed data samples.

- as a result, a distribution of signal PID efficiencies is obtained, whose average value and RMS are quoted as the wanted signal PID efficiency central value and error, respectively.

In the following sections, the efficiencies for the muons and the hadrons (pions, kaons and protons) will be shown under the muon PID selection defined in Sec. 3.1.4.

3.3.1 Muon identification

As discussed in Sec. 3.1.4, the muon PID selection working point has been optimized on Run 1 data, resulting in the cut $\text{PID}_{\mu} > 0.4$ ($\text{PID}_{\mu,4}^{\text{MC12}}$ in the following). The MC12 label refers to the MC sample used to train the ProbNN variables. A looser cut has been applied instead in the lifetime analysis, as discussed in Sec. 6.3: $\text{PID}_{\mu} > 0.2$ ($\text{PID}_{\mu,2}^{\text{MC12}}$ in the following). This is motivated by the fact that such analysis selects the B_s^0 mass region only, which is less affected by the exclusive backgrounds, whose rejection strongly benefits from a tighter PID selection. Following the different detector conditions in Run 2, a new tuning of the ProbNN variables has been trained, which makes use of a different MC sample, called MC15. For this tuning, a new working point has been identified, corresponding to the same signal efficiency obtained in Run 1: $\text{PID}_{\mu} > 0.8$ ($\text{PID}_{\mu,8}^{\text{MC15}}$) for the branching fraction analysis and $\text{PID}_{\mu} > 0.4$ ($\text{PID}_{\mu,4}^{\text{MC15}}$) for the lifetime analysis, respectively. This selection has been then used for 2016 data. For 2015 data, the Run 1 selection was used instead, since the new ProbNN tuning was not available. This choice may lead to a suboptimal PID performance on the 2015 sample, but since it represents only the 7% of the total statistics, this is accepted.

The outcome of the above discussion is summarized in Table 3.3, where the values of the various PID cuts are given, together with the data samples and ProbNN tunings.

In Fig. 3.6, the single muon efficiency is shown as a function of momentum in bins of p_T , for the PID selection of the branching fraction analysis, and for the

kinematic bins defined by Eq. (3.7). Note that the IsMuon condition is implied in

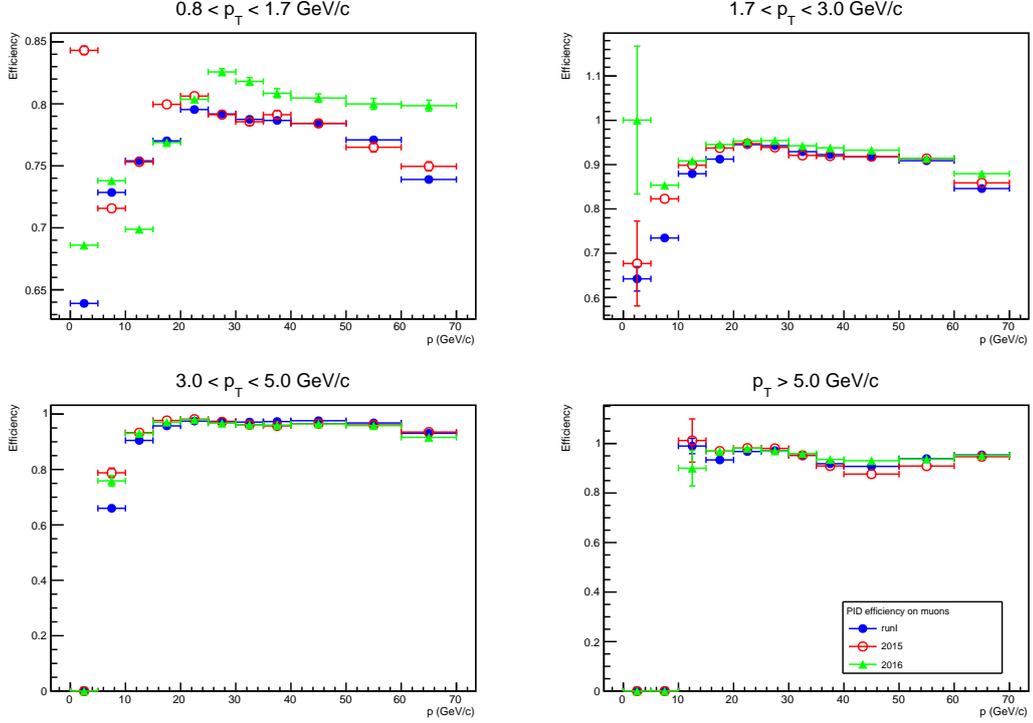


Figure 3.6: Efficiency of the muon PID selection of the branching fraction analysis on muon calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T .

the PID_μ cut, i.e. the reported values show the combined efficiency of IsMuon & PID_μ .⁵ The efficiency worsens in the low momentum regime following the degradation of the muDLL discriminating power for low p_T tracks, where the multiple scattering is larger [147].

When convoluted with the $B_s^0 \rightarrow \mu^+ \mu^-$ kinematics (Fig. 3.2), the following PID selection efficiencies are obtained:

$$\begin{aligned}
 \epsilon_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}^{\text{PID, Run 1}} &= 82.3 \pm 1.6\%, \\
 \epsilon_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}^{\text{PID, 2015}} &= 83.5 \pm 1.7\%, \\
 \epsilon_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}^{\text{PID, 2016}} &= 84.3 \pm 1.7\%,
 \end{aligned} \tag{3.8}$$

⁵ The track acceptance is factorised out of the computation by requiring the calibration tracks to lie within the geometrical acceptance of the detector.

where quoted uncertainty includes the calibration sample statistics, as described in the previous section, and a systematic uncertainty due to the binning scheme choice and the trigger unbiasing conditions. Despite the appreciable differences in the single muon efficiencies for 2016 calibration data (green points in Fig. 3.6), the average results of Eq. 3.8, which are integrated over the full phase space, are in good agreement.

A PID efficiency degradation is expected in high multiplicity events [119, 147]. Since the PID selection dependence on the event multiplicity was not accounted for in the extraction of the single particle efficiencies from the calibration samples, the corresponding effect has been investigated by further splitting the calibration sample according to three bins in the number of tracks:

$$N_{\text{tracks}} = [2, 150] [150, 250] [250, \infty]. \quad (3.9)$$

By comparing the resulting PID efficiencies, no difference has been found with the N_{tracks} -integrated value, hence no systematic error is needed.

3.3.2 Hadron misidentification

Pions and kaons

The single-pion and -kaon misID under the muon PID selection of the branching fraction analysis (Table 3.3) are shown in Fig. 3.7 and Fig. 3.8, respectively. Given the very small value of the misID, some points fluctuate to negative values due to the background subtraction procedure, but they are still compatible with zero within the quoted error. For both pions and kaons, the 2016 tuning of the ProbNN gives an equal or better background rejection with an equal or higher muon efficiency (Fig. 3.6), thus demonstrating a real performance improvement. By integrating the above misID on the $B_{d,s}^0 \rightarrow h^+h^-$ spectra (more details are given in Sec. 5.1), the following average double-hadron misidentification rates are found:

$$\begin{aligned} \epsilon_{hh \rightarrow \mu\mu}^{\text{PID, Run 1}} &= (4.6 \pm 0.2) \times 10^{-6}, \\ \epsilon_{hh \rightarrow \mu\mu}^{\text{PID, 2015}} &= (3.6 \pm 0.2) \times 10^{-6}, \\ \epsilon_{hh \rightarrow \mu\mu}^{\text{PID, 2016}} &= (3.39 \pm 0.07) \times 10^{-6}. \end{aligned} \quad (3.10)$$

The reduced uncertainty in the Run 2 result is due to the much larger statistics available for the respective calibration samples, which is also visible in Fig. 3.7 and Fig. 3.8.

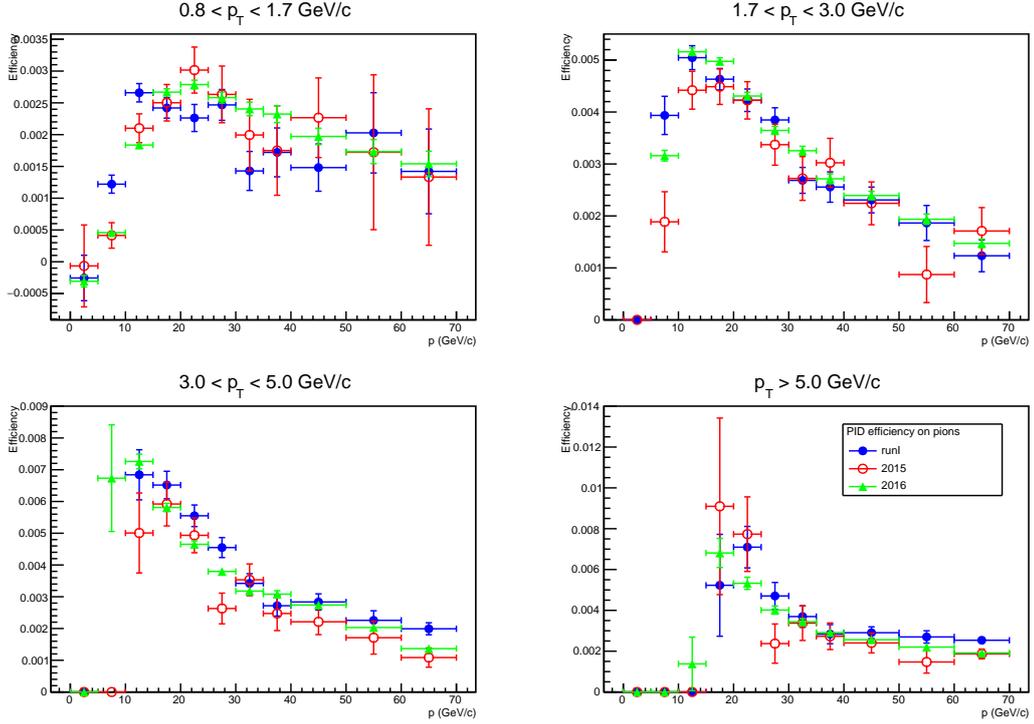


Figure 3.7: Efficiency of the muon PID selection of the branching fraction analysis on pion calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T .

Protons

A separated treatment is reserved for proton to muon misID, since the strong PID request heavily cuts down the statistics of the relative calibration samples. For this reason, a different binning scheme is adopted:

$$\begin{aligned}
 p &= [2, 40] [40, 500] \text{ GeV}/c, \\
 p_T &= [0.8, 1.7] [1.7, 3] [3, 40] \text{ GeV}/c,
 \end{aligned}
 \tag{3.11}$$

i.e. only two momentum regions are used and the last two p_T bins of Eq. (3.7) are merged. This coarser division is meant to make the most out of the available calibration sample, so that reasonable values can be computed, as shown in Fig. 3.9. The strenght of the PID_μ cut in rejecting protons is evident: the misID rate is at the per mille level, and bigger statistical uncertainties arise where the $p - p_T$ coverage of the calibration sample is weaker.

The proton to muon misID only pertains the computation of the $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ background yield (Sec. 5.2.4) and also constitutes its biggest source of uncertainty.

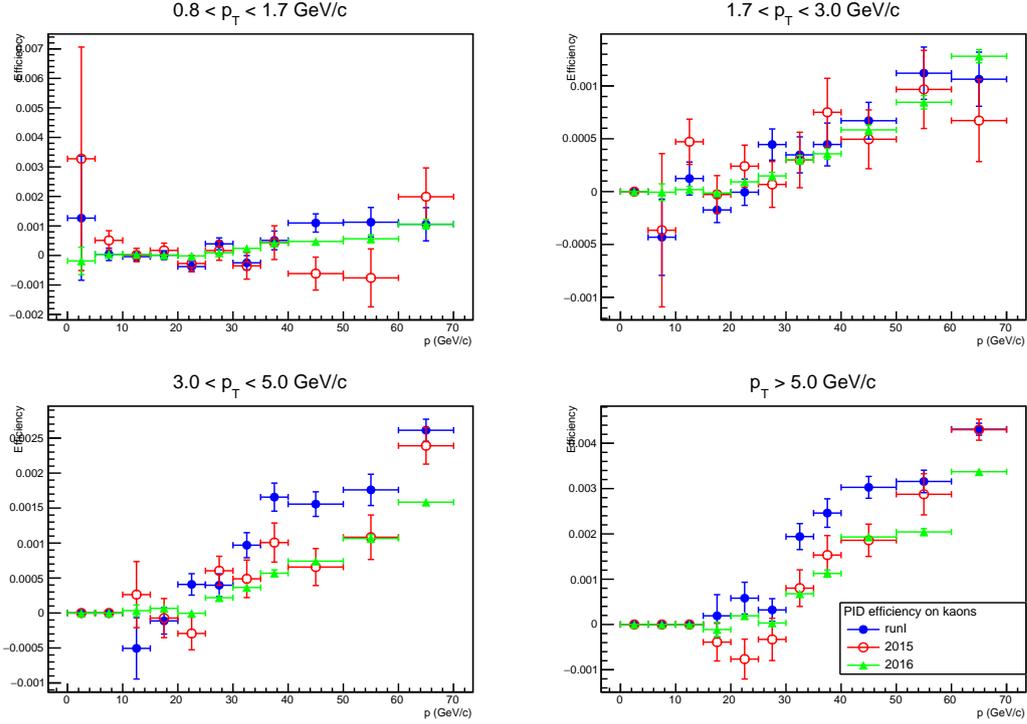


Figure 3.8: Efficiency of the muon PID selection of the branching fraction analysis on kaon calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T .

When integrated over the $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ kinematic spectrum, the following PID selection efficiencies are found:

$$\begin{aligned}
 \epsilon_{\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu}^{\text{PID, Run 1}} &= (0.9 \pm 0.4) \times 10^{-4}, \\
 \epsilon_{\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu}^{\text{PID, 2015}} &= (2.6 \pm 0.5) \times 10^{-4}, \\
 \epsilon_{\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu}^{\text{PID, 2016}} &= (3.4 \pm 0.8) \times 10^{-4}.
 \end{aligned} \tag{3.12}$$

To the quoted statistical uncertainty, a systematic error for the binning scheme is added in the computation of Sec. 5.2.4 by taking the difference with the values obtained with the usual binning scheme, and amounts to 44%, 10% and 28% for Run 1, 2015 and 2016, respectively. Since the proton calibration samples are populated with Λ decays, and the background source involves a Λ_b baryon, which has a shorter lifetime, the dependence of the proton misID on the Λ production point was investigated, but no effect was found.

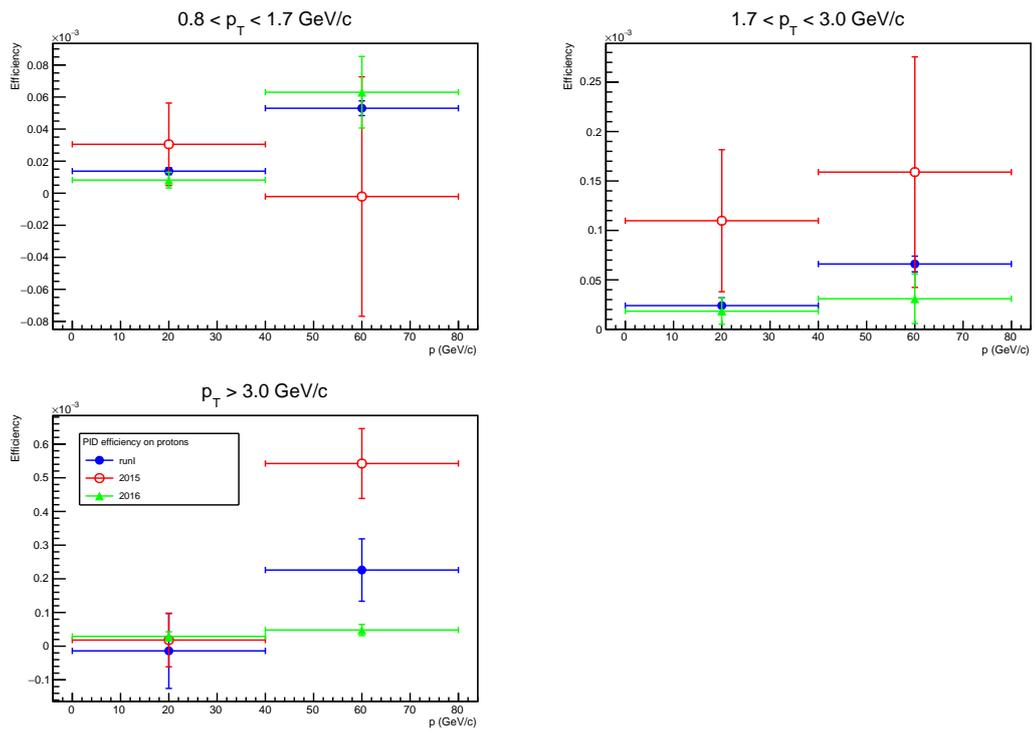


Figure 3.9: Efficiency of the muon PID selection of the branching fraction analysis on proton calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T .

Chapter 4

Signal selection and characterisation

This chapter describes the selection criteria used to isolate the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signals, as well as the definition and calibration of their mass and BDT *pdfs*. After a first loose selection aimed to reduce the sample size, the data are classified into regions according to the response of a multivariate classifier named BDT, with increasing signal over background ratio. The discussion continues with the calibration of the signal mass *pdf* and the normalisation procedure used to extract the branching fraction.

Besides the signals, two main control channels are used in this analysis: $B^+ \rightarrow J/\psi K^+$ for signal and exclusive background normalisation, and $B_d^0 \rightarrow K^+ \pi^-$ for signal normalisation and BDT calibration. These channels are selected using criteria which are as similar as possible to the signal ones, as will be explained in the following sections.

4.1 Selection

After the data are filtered by the trigger, they undergo an offline selection, called *stripping*, in which soft cuts are applied to reduce the data size to a manageable level while keeping the signal efficiency as high as possible. The signal and normalisation channels are selected in a similar way to minimise the systematic uncertainties. The following spurious events are easily rejected by the stripping:

1. **Non-physical background**, which originates from incorrect event reconstruction, is reduced by imposing good quality of the reconstructed tracks, for example limiting the track χ^2/ndf , the DOCA between the two tracks and the χ^2 of their vertex. Artefacts of the reconstruction (*ghosts*, Sec. 2.5.2) are rejected by means of a ghost probability cut, while upper limits on p and

p_T are applied to discard reconstructed events which fall out of the LHCb acceptance.

2. **Quasi-elastic pp collisions**, which can produce couples of oppositely charged muons with good quality vertex, following the process $pp \rightarrow p\mu^+\mu^-p$. As in this case the protons undergo a soft interaction, they travel close to the beam axis almost undisturbed, so that this background source is easily reduced by imposing a minimum p_T on the B meson candidate.
3. **Prompt muons**, which are produced in inelastic collisions. These can be rejected by exploiting their topology in comparison to the one that characterises a B decay: since B mesons travel for distances of $\mathcal{O}(1 \text{ cm})$, selections based on the significance of the impact parameter ($\text{IP}\chi^2$) for the muon candidates, as well as flight distance significance of the secondary vertex (VDS) are employed to reject this class of backgrounds.

Besides the above categories, the two main sources of background are combinatorial events, where two muons arise from different B decays, and exclusive background decays, treated in detail in Chapter 5. The combinatorial events are partly reduced in this first selection by cutting on the DOCA between the tracks, and using a dedicated MVA variable called BDTS. After this, a much better separation is achieved by using the main BDT of the analysis, which implements the full topological information of the event. The exclusive background decays are mostly reduced by the tight PID selection, as discussed in Chapter 3.

Three stripping lines are relevant for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis: $B_s^0 \rightarrow \mu^+\mu^-$, $B_{d,s}^0 \rightarrow h^+h^-$ and $B^+ \rightarrow J/\psi K^+$, whose selections are kept as similar as possible. The stripping selections are summarised in Table 4.1. For Run 2 data, the track χ^2/ndf and the ghost probability requirements have been loosened to take advantage of the reconstruction improvements.

The BDTS¹ variable is the outcome of a Boosted Decision Tree² classifier used to further reduce the size of the background sample before the application of the final BDT (Sec. 4.2). The variables used to train the BDTS are:

1. IP of the B candidate,
2. $\text{IP}\chi^2$ of the B candidate,
3. DOCA of the two daughter tracks (the two muons in the $B^+ \rightarrow J/\psi K^+$ case),

¹The BDTS has been developed during the first $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis in LHCb and never changed [154].

²More details about boosted decision trees are given in Sec. 4.2.1.

Table 4.1: Selection for $B_{d,s}^0 \rightarrow \mu^+\mu^-$, $B_{d,s}^0 \rightarrow h^+h'^-$ and $B^+ \rightarrow J/\psi K^+$ channels. The values in parenthesis for track χ^2/ndf and ghost probability show the softer cuts used for Run 2 data.

Cut	applied on	value	applied on	value
		$B_s^0 \rightarrow \mu^+\mu^-$ and $B_{d,s}^0 \rightarrow h^+h'^-$		$B^+ \rightarrow J/\psi K^+$
track χ^2/ndf	μ / h	< 3 (< 4)	μ / h	< 3
ghost prob		< 0.3 (< 0.4)		
DOCA		< 0.3 mm		< 0.3 mm
IP χ^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	$B_{(s)}^0$	< 9	J/ψ	< 9
VDS		> 15		> 15
ΔM		$ M(hh, \mu\mu) - m_B < 60$ MeV/c ²		$ M(\mu\mu) - m_{J/\psi} < 60$ MeV/c ²
IP χ^2	$B_{(s)}^0$	< 25	B^+	< 25
t		$< 9 \cdot \tau(B_s)$		$< 9 \cdot \tau(B_s)$
BDTS		> 0.05		> 0.05
ΔM				$ M(J/\psi K) - m_B < 100$ MeV/c ²
$p_T(B_s)$	$B_{(s)}^0$	> 0.5 GeV/c		

4. B vertex χ^2 (the J/ψ vertex χ^2 in the $B^+ \rightarrow J/\psi K^+$ case),
5. the angle between the direction of the B candidate momentum and the direction defined by the secondary and primary vertices (DIRA),
6. the minimum impact parameter (minIP) of each daughter track with respect to any primary vertex (muon tracks in the $B^+ \rightarrow J/\psi K^+$ case).

The training has been performed using simulated samples of $B_s^0 \rightarrow \mu^+\mu^-$ for the signal and $b\bar{b} \rightarrow \mu^+\mu^- X$ for the background, on which the selection in Table 4.1 is applied. The cut on the BDTS output retains $\sim 93\%$ of the signal, while rejecting $\sim 70\%$ of the background. The BDTS cut efficiencies on signal and normalisation samples agree within 0.4%.

As a final step, the PID requirement of Sec. 3.3.1 is applied to the muon candidates of the signal selection, in order to reduce the contribution from the exclusive

backgrounds. The remaining combinatorial background cannot be reduced by means of rectangular cuts without losing a significant amount of sensitivity. For this reason, the events surviving the full selections are analysed in bins of a BDT output, which is described in the following section.

4.2 The BDT for combinatorial background rejection

The most abundant source of fake $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signals is represented by the combinatorial background. Since b and \bar{b} are always produced in pairs, their semileptonic decay can originate two oppositely charged muons. If the muon track extrapolations cross to form a detached vertex, the event can be recognised as a $B_{d,s}^0 \rightarrow \mu^+\mu^-$ one, as sketched in Fig. 4.1. Given the arbitrariness of the momentum combination,

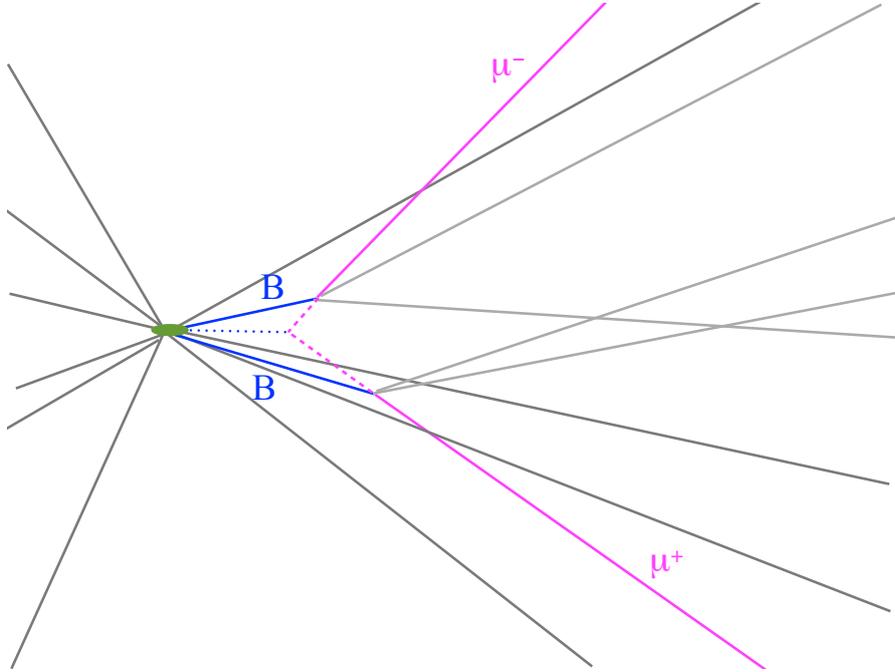


Figure 4.1: Cartoon of a combinatorial $B_{d,s}^0 \rightarrow \mu^+\mu^-$ event. Two B mesons produced at the PV (green ellipse) decay and produce two muons, whose track extrapolations (dashed pink) form a B -like vertex (dashed blue).

the invariant mass of the two muons has an exponentially decreasing distribution, i.e. the mass spectrum of the combinatorial background sharply decreases within the signal mass region.

To fight this background, a BDT has been defined that exploits the full event

topology. In particular, a huge improvement has been made with respect to the previously published analysis by developing a new muon isolation variable. Before discussing the details about the BDT implementation and its performances, a brief introduction on the boosted decision tree technique is given in the next section.

4.2.1 Boosted decision trees

Multivariate techniques [155] are gaining more and more interest in particle physics analyses, where big data samples characterised by many variables constitute a fertile ground for their application. As a matter of fact, more information from the detector can be exploited by accounting for variable correlations.

The building block of the BDT developed for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis is the *decision tree*. It is natural and intuitive to classify an event through a sequence of questions, in which the next question asked depends on the answer to the current one. A sequence of *yes/no* questions is depicted as a decision tree: starting from a *root node*, repeated decisions are taken on a single variable at each *leaf node*, as schematised in Fig. 4.2. In this way, the parameter space is splitted into many regions that are labelled as *signal* or *background*, depending on the majority of *training* events that fall into the final leaf node. A *training sample* is a collection of events for which the classification is known, for example the one obtained via a MC simulation.

To enhance the classification power of a decision tree, an ensemble of decision trees, a *forest*, can be built so that their joint decision rule is more accurate. The trees in a forest are derived from the same training sample by weighting events, and the final classifier is made up by the weighted average of the individual decision trees. Besides performance improvement, the boosting technique also stabilises against fluctuation of the training samples, making the classifier less prone to *overtraining*. As a side effect, the intuitive decision algorithm which characterises the single tree is lost.

The boosting method adopted for the BDT is the popular adaptive boost, AdaBoost [156]. Starting with the event weights obtained after the training of the first tree, the next tree is trained using an event sample in which the weights of the previously misclassified events are multiplied by a common *boost factor* α , given by

$$\alpha = \frac{1 - err}{err}, \quad (4.1)$$

where *err* is the misclassification rate. Hence AdaBoost focuses on “informative” events, i.e. the ones that are harder to classify.

This boosting method can turn a set of weak decision trees into a strong learner, but it is sensitive to noisy data and outliers. To mitigate this drawback, *slow learning* can be imposed by tuning the parameter β which is assigned to the boost

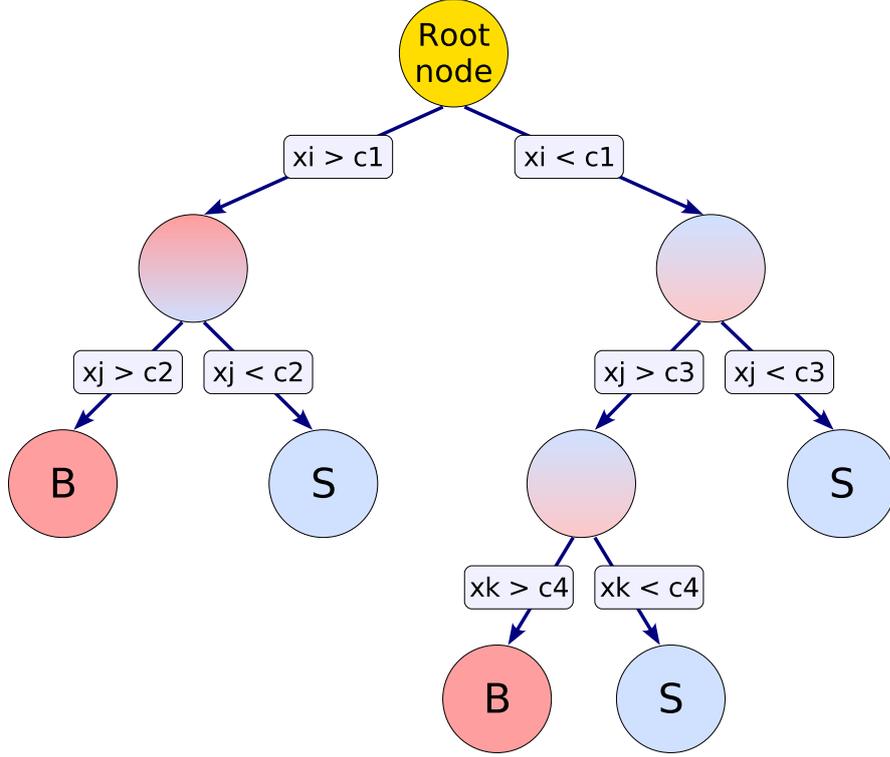


Figure 4.2: Scheme of a decision tree [149]. At each node, the split is performed by cutting on the variable which gives the best separation between the signal (S) and background (B) classes. The same variable may thus be used in several nodes, while others might not be used at all.

factor as $\alpha \rightarrow \alpha^\beta$. A small learning rate and a large number of boost steps are therefore advisable to improve the classifier performance [149].

4.2.2 Muon isolation

A powerful discriminant for the combinatorial background is the *isolation*, as muons from $B_{d,s}^0 \rightarrow \mu^+\mu^-$ events tend to be far from other tracks in the event, whereas in the topology of Fig. 4.1, muon tracks are often closely accompanied by track arising from the same B decay. In the previously published analysis, the muon isolation was defined by a series of rectangular cuts [4]. For the present analysis, two new isolation variables, based on boosted decision trees, have been developed, targeting long tracks and VELO tracks separately. The isolation score is computed for each track in the event against the μ^+ (iso^+) and the μ^- (iso^-), with isolated tracks being characterised by smaller scores. The event isolation is then defined as $\max(iso_i^+ + iso_j^-)$, i.e. the maximum value of the two isolations

computed for each pair of tracks i, j in the event.

For the long track isolation, the following variables are used as an input for the classifier:

1. the minimum of the $\sqrt{\text{IP}\chi^2}$ of the track with respect to any PV,
2. the signed distance between the track vertex and the PV,
3. the signed distance between the track vertex and the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ vertex,
4. the DOCA between the track and the muon,
5. the angle between the track and the muon,
6. $f_c = \frac{|\vec{p}_\mu + \vec{p}_{\text{trk}}| \sin(\alpha^{\mu+\text{trk},\text{pV}})}{|\vec{p}_\mu + \vec{p}_{\text{trk}}| \sin(\alpha^{\mu+\text{trk},\text{pV}}) + p_{T\mu} + p_{T\text{trk}}}$, where $\alpha^{\mu+\text{trk},\text{pV}}$ is the angle between the sum of the muon and track momenta and the direction defined by the PV and the track-muon vertex. f_c is close to zero if the track and the muon originate from the PV,
7. the absolute value of the difference between the track and the muon azimuthal angles,
8. the absolute value of the difference between the track and the muon pseudo-rapidities,
9. the p_T of the track,

whereas the VELO track classifier only uses the first 6 variables in the list.³ The classifiers have been trained using simulated $b\bar{b} \rightarrow \mu^+\mu^- X$ events as background and $B_s^0 \rightarrow \mu^+\mu^-$ events as signal.

The Receiver Operating Characteristic (ROC) curves in Fig. 4.3a and Fig. 4.3b show the long track and VELO track isolation classifier performances as background rejection versus signal efficiency. The performances have been evaluated on simulated $B_s^0 \rightarrow \mu^+\mu^-$ events as signal and on right-sideband data for the background, where the right sideband is defined as the invariant mass region $m(\mu^+\mu^-) \in [5447, 6000] \text{ MeV}/c^2$, which is populated by combinatorial background only. The selection described in Sec. 4.1 is applied to the data along with the Is-Muon request. The isolation performance has significantly increased from the previous analysis: for a signal efficiency of 80%, about 40% more background events are rejected, as can be deduced from Fig. 4.3a. No performance degradation is observed among the data samples.

³ Since the p_T cannot be measured for VELO tracks, it is set to 400 MeV/c in the computation of the f_c variable, a value which is not far from the mean of the long track p_T distribution.

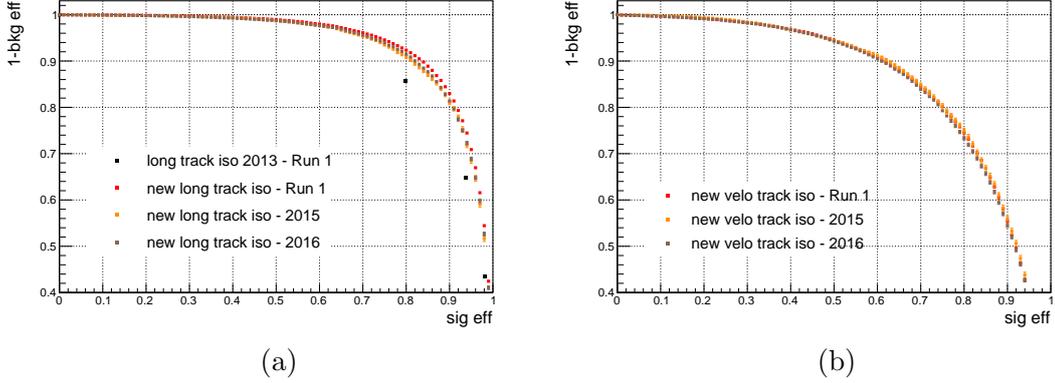


Figure 4.3: ROC curves for the long track (a) and VELO track (b) isolations evaluated on Run 1 (red), 2015 (orange) and 2016 (brown) data. The performance of the cut-based isolation (black) used in the past analysis is superimposed in (a).

4.2.3 The global BDT

The global BDT (simply “BDT” in the following) constitutes the final tool for signal and background classification, and is meant to distinguish between two-body B decays from multi-body decays of the class represented in Fig. 4.1. To this end, only geometrical variables are used for its training.

The long track isolation variable described in Sec. 4.2.2 gives the best separation power between $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signals and combinatorial background. Through an iterative procedure, more variables are added to the BDT training set to find the one that gives the best background rejection. Such procedure quickly allows to find many discriminating variables but, since it is prone to statistical fluctuation of the background yield, the final choice among the best configurations is made by hand.

A total of 7 variables are selected for the BDT definition:

1. Long track isolation (Sec. 4.2.2);
2. $\Delta R = \sqrt{\Delta\Phi^2 + \Delta\eta^2}$, $\Delta\Phi$ and $\Delta\eta$ being the azimuthal angle and pseudorapidity differences between the two muons (`mu_deltaR`);
3. the minimum IP significance of the two muons with respect to the PV associated to the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidate (`mu_MINIPS`);
4. the angle between the B direction and the vector joining the primary and secondary vertices (`B_ACOSDIRA_OWNPV`);
5. the vertex χ^2 of the B candidate (`B_ENDVERTEX_CHI2`);

- 6. IP significance of the B candidate with respect to the PV (B_IPS_OWNPV);
- 7. VELO track isolation (Sec. 4.2.2).

The training is performed on simulated $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu^+\mu^-X$ events for signal and background, respectively, to which the selection of Sec. 4.1 is applied. The input variable distributions are shown in Fig. 4.4 for signal MC and right data sidebands. The BDT output variable is subsequently *flattened* in the range $[0, 1]$ such that the signal is uniformly distributed, while the background peaks at zero. In this way, the signal/background separation is preserved and the BDT output can be splitted into regions (bins) of known signal efficiency once the calibration is applied, as explained in Sec. 4.2.4. The data are in fact classified according to the following binning scheme in the BDT output variable:

$$[0, 0.25] [0.25, 0.4] [0.4, 0.5] [0.5, 0.6] [0.6, 0.7] [0.7, 0.8] [0.8, 0.9] [0.9, 1.0], \quad (4.2)$$

which will be adopted throughout the rest of the analysis.

The final invariant mass fit for the signal yield extraction will be performed in several BDT bins simultaneously. For this reason, the correlation between BDT and invariant mass must be carefully controlled. If, on one hand, linear BDT-mass correlations can be accounted for in the fit, on the other hand non-linear correlations must be avoided to prevent biases in signal and background yield extraction. To avoid false signal peaks, the BDT must not be capable of reconstructing the invariant mass, i.e. the input variables concerning kinematics must not be exhaustive enough to allow the BDT to calculate the invariant mass. A linear correlation of $\sim -3\%$ is found, while no peaking structures are present within the signal mass region, as checked from simulated data. In addition, no biases on data are added since the BDT training and optimisation is fully done on simulated events.

The ROC curve of the BDT is displayed in Fig. 4.5 for Run 1 data sidebands, together with the BDT used in the previous analysis: the background rejection has increased by about 50%. Fig. 4.6 shows the performance on the right data sidebands of all the data subsamples, where no significant performance disparity is observed.

4.2.4 BDT calibration

While the BDT is fully trained on simulated events, its distribution is evaluated on data by using $B_d^0 \rightarrow K^+\pi^-$ events (the most abundant among the four $B_{d,s}^0 \rightarrow h^+h'^-$ channels), which act as a proxy for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signals. This data-driven BDT *calibration* provides a reliable estimate of the signal fractions in each BDT bin. The $B_d^0 \rightarrow K^+\pi^-$ events are selected by using the $B_{d,s}^0 \rightarrow h^+h'^-$ cuts reported in Table 4.1, with the additional requirements of L0 and HLT1 TIS, HLT2

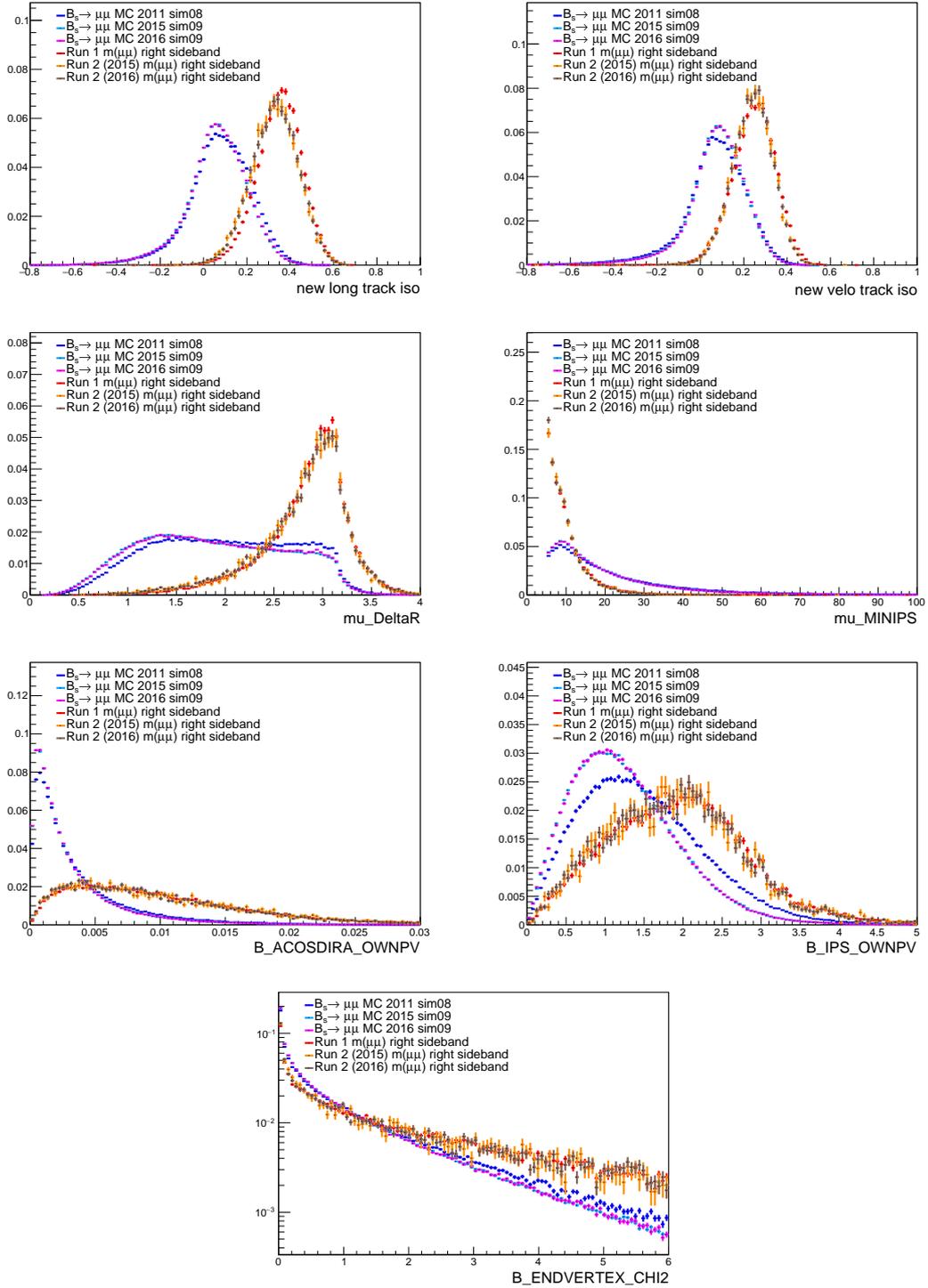


Figure 4.4: Distribution of the BDT input variables in $B_s^0 \rightarrow \mu^+\mu^-$ MC events (2011, 2015 and 2016 conditions) and in the right mass data sideband (Run 1, 2015 and 2016).

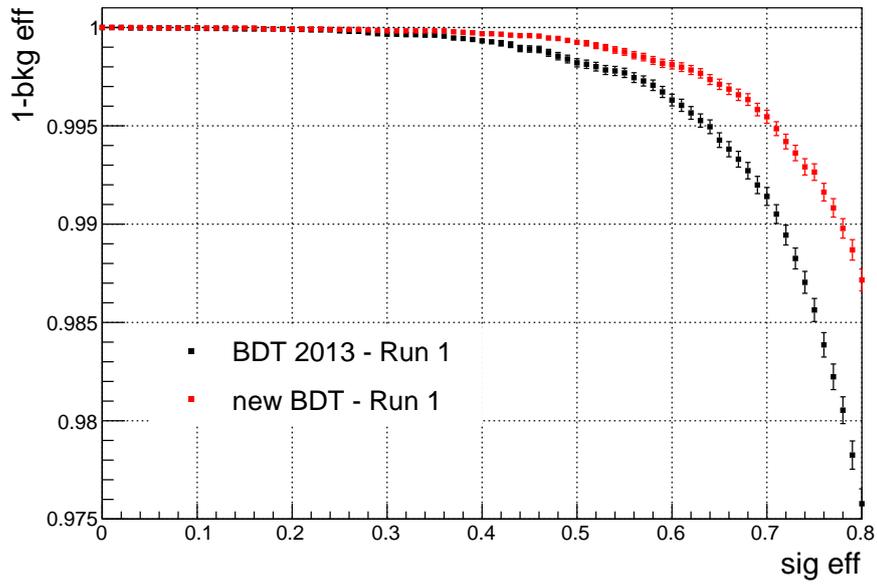


Figure 4.5: ROC of the new BDT in Run 1 data and comparison with the ROC of the BDT used in the past analysis.

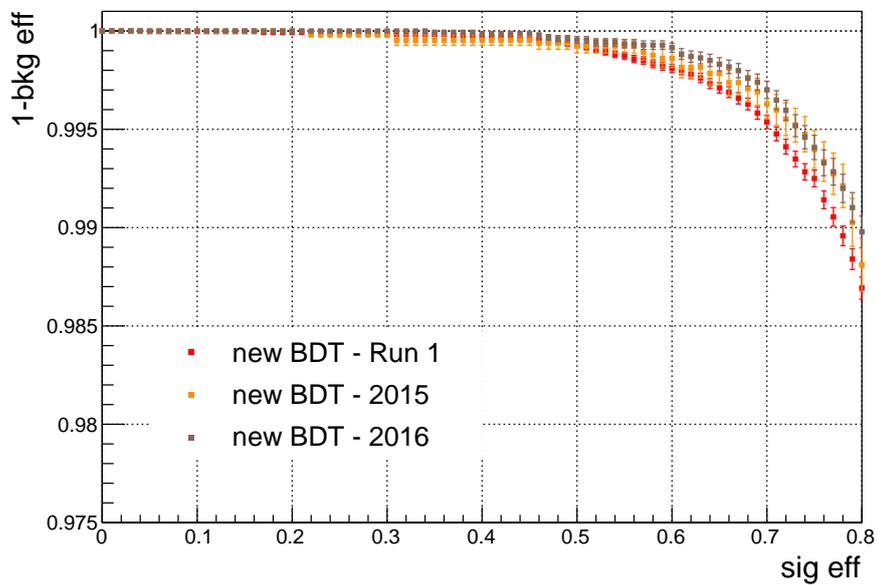


Figure 4.6: ROC of the new BDT in Run 1, 2015 and 2016 data.

TOS and both hadrons within the muon detector acceptance. A PID cut is also applied to identify the kaon and the pion in the final state. The goal of the BDT calibration is then to estimate the $B_d^0 \rightarrow K^+\pi^-$ yield in each BDT bin by means of an invariant mass fit. The fit mass window has been chosen to run from 5200 to 5850 MeV/c², in order to rule out the partially reconstructed backgrounds on the left sideband. The first BDT bin is further splitted as $[0, 0.25] \rightarrow [0, 0.1] + [0.1, 0.25]$, and the $B_d^0 \rightarrow K^+\pi^-$ yield in $\text{BDT} \in [0, 0.1]$, which is fully dominated by combinatorial background, is evaluated by subtracting the yields in the other bins to the yield evaluated in the full BDT range.

The fit model comprises the following components:

- Double-sided Crystal Ball functions [157] for the $B_{d,s}^0$ peaks, whose tail parameters are constrained from $B_d^0 \rightarrow K^+\pi^-$ MC events. The masses are fixed to the PDG values [158], while the ratio of the resolutions is fixed to the one obtained in mass calibration procedure, as reported in Sec. 4.3.
- A double-sided Crystal Ball for the $\Lambda_b \rightarrow ph$ misidentified background, i.e. when the proton is wrongly selected as a pion or a kaon. The *pdf* parameters are evaluated from a fit to the corresponding simulated sample and kept fixed.
- An exponential function for the combinatorial background, whose slope is left free in the fit.

The $B_d^0 \rightarrow \pi^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ components are found to be negligible, given their small branching fraction and the low hadron misidentification rate of the chosen PID cut. In Fig. 4.7, a typical fit result on 2016 data is shown.

Results

The fit yields in each bin are corrected for the relative PID efficiency and divided by the total yield to compute the BDT fractions. A further correction is applied to take into account the different trigger requirements on $B_{d,s}^0 \rightarrow h^+h'^-$ and $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signals, as evaluated from simulated data. The results are displayed in Fig. 4.8 and show no significant differences with respect to the fractions obtained from the $B_d^0 \rightarrow K^+\pi^-$ MC. Systematic uncertainties for the PID correction, trigger and fit model are included.

4.2.5 Time-dependent effects

The BDT output is flattened using simulated events, so that the expected signal yield in each BDT bin is simply proportional to the bin width. As explained in

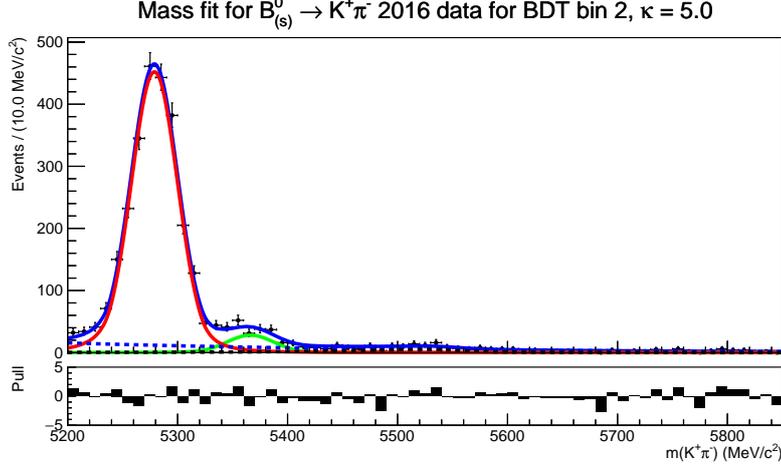


Figure 4.7: Invariant mass distribution of $B_d^0 \rightarrow K^+\pi^-$ events in 2016 data for $\text{BDT} \in [0.25, 0.4]$. The fit result is overlaid, showing the B_d^0 (red), B_s^0 (green), Λ_b (dashed red) and combinatorial (dashed blue) components. $\kappa = 5$ indicates the PID cut value: $\text{DLL}(K - \pi) > \kappa$ for the kaon and $\text{DLL}(K - \pi) < -\kappa$ for the pion.

Sec. 1.2.2, the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime ranges from the heavy to the light mass eigenstate lifetimes. However, the MC sample is generated using the B_s^0 lifetime. Since the BDT output is correlated with the decay time of the candidate, with long-lived candidates tending to have larger BDT values, a bias can be introduced. To account for this *time-dependent* effect, numerical factors have been calculated to correct the BDT distribution under the different hypotheses $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1, 0, +1$. Simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays are required to pass the trigger and offline selections, then, a per-event weight is evaluated as

$$\omega_i = \frac{\tau_{gen}}{\tau_{\mu^+\mu^-}} e^{-t_i(1/\tau_{\mu^+\mu^-} - 1/\tau_{gen})}, \quad (4.3)$$

where t_i is the reconstructed decay time, τ_{gen} the decay time used for the event generation and $\tau_{\mu^+\mu^-}$ is the effective lifetime defined by Eq. (1.32). Each BDT bin is then corrected by a factor

$$k = \sum_i^N \omega_i / N = \frac{\epsilon_{\tau_{\mu^+\mu^-}}}{\epsilon_{\tau_{gen}}}, \quad (4.4)$$

N being the number of candidates within a BDT bin. The correction factors are therefore represented by the ratios of the total efficiencies (trigger, reconstruction and selection) of a sample whose mean lifetime corresponds to one of the three

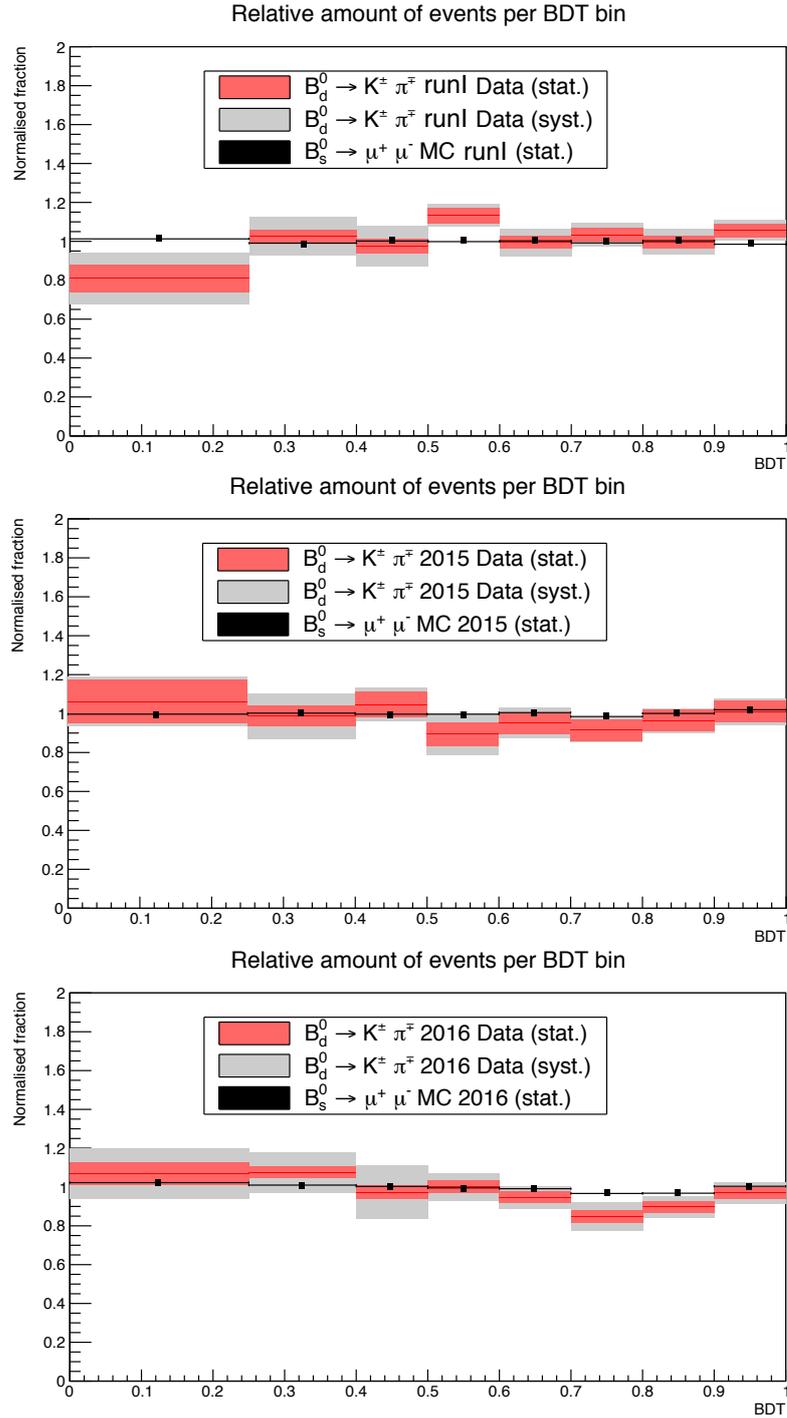


Figure 4.8: BDT distributions for Run 1 (top), 2015 (middle) and 2016 (bottom) data samples. The statistical uncertainty is shown in red, while the systematic one is in grey. The black points mark the resulting distribution for the $B_d^0 \rightarrow K^+ \pi^-$ simulated sample.

$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ values and a sample where the mean lifetime is the one used in the MC generation. Different corrections are thus evaluated for each data subsample for all the three $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ hypotheses.

4.3 Mass calibration

The invariant mass distribution of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signal is parametrised by a Crystal Ball [157] function, which is well-suited to describe the effect of the detector resolution as well as the radiative losses due to FSR (Sec. 1.2.4). The Crystal Ball function is in fact made up of a Gaussian core and a left power-law tail which is activated above a certain energy threshold. In this section, the determination of the four parameters (μ, σ, α, n) defining the signal Crystal Ball function is discussed.

Two different sets of parameters are used for Run 1 and Run 2, respectively. In both cases, the result is obtained from a weighted average of the parameters from the single years of data taking, as they yield consistent numbers. The systematic error is treated as fully correlated in the average.

4.3.1 Mean

The mean value of the B mass is evaluated from a fit to the $B_d^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ invariant masses for the $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ *pdfs*, respectively. The selection of the control sample is similar to the BDT calibration one, the only difference being that no trigger unbias requirement is applied in this case. For this reason, a much larger statistics is available, and a harder PID cut can be applied to select pions and kaons in the final state. The same fit model of Sec. 4.2.4 is employed. The fit results for 2016 data are shown in Fig. 4.9, for both the B_d^0 and the B_s^0 . Table 4.2 summarises the mass calibration results, in which the systematic error is dominated by the PID selection.

Table 4.2: Mean values of the B_d^0 and B_s^0 masses for Run 1 and Run 2 data.

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

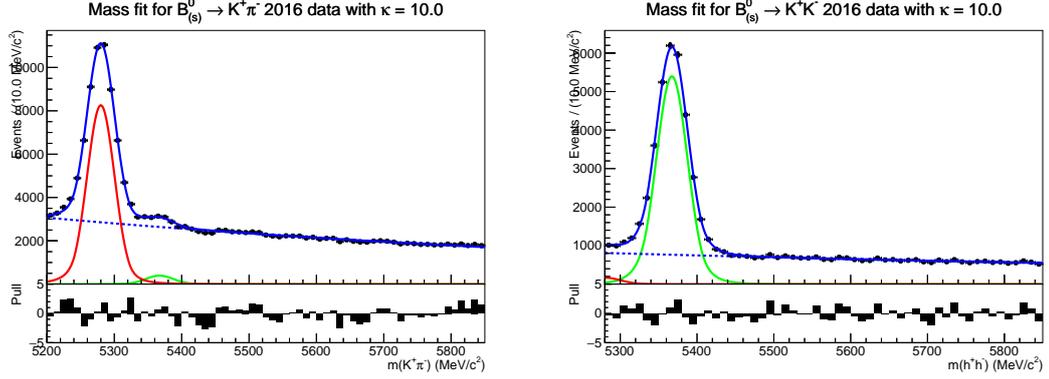


Figure 4.9: Invariant mass fits to $B_d^0 \rightarrow K^+\pi^-$ (left) and $B_s^0 \rightarrow K^+K^-$ (right) 2016 data. The red line identifies the B_d^0 pdf, the green line is the B_s^0 pdf and the blue one represents the combinatorial background. $\kappa = 10$ indicates the PID cut value: $\text{DLL}(K - \pi) > \kappa$ for the kaon and $\text{DLL}(K - \pi) < -\kappa$ for the pion.

4.3.2 Resolution

The invariant mass resolution of the $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ peaks is interpolated from charmonium and bottomonium resonances. These include $J/\psi(1S) \rightarrow \mu^+\mu^-$, $\psi(2S) \rightarrow \mu^+\mu^-$, $\Upsilon(1S) \rightarrow \mu^+\mu^-$, $\Upsilon(2S) \rightarrow \mu^+\mu^-$ and $\Upsilon(3S) \rightarrow \mu^+\mu^-$ decays. A fit model with a double-sided Crystal Ball function on top of a power-law distribution describing the combinatorial background is used in all cases. Fig. 4.10 shows the Υ resonances as obtained from 2016 data. By analysing a Drell-Yan MC

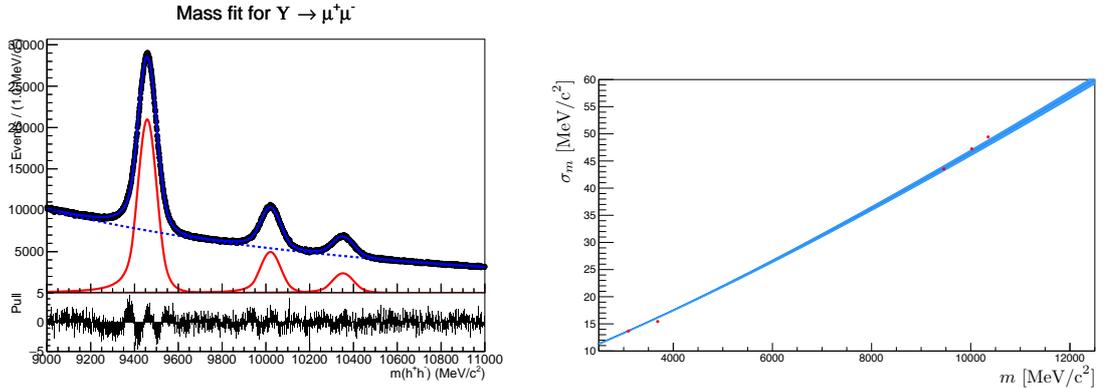


Figure 4.10: (Left) Invariant dimuon mass fits on $\Upsilon(1, 2, 3S) \rightarrow \mu^+\mu^-$ 2016 data. (Right) Interpolation of the invariant mass resolution of the dimuon resonances for 2016 data.

sample, the invariant mass resolution as a function of the dimuon mass is found to be well described by the empirical law [159]

$$\sigma_{\mu\mu}(m_{\mu\mu}) = a_0 + a_1 \cdot m_{\mu\mu}^\gamma. \quad (4.5)$$

The a_0 and a_1 parameters are obtained from a fit to the resolution values obtained from the dimuon resonances, as shown in Fig. 4.10. The resolution values interpolated at the B_d^0 and B_s^0 masses from Eq. (4.5) are shown in Table 4.3, where the systematic uncertainty accounts for the fit model and the selection cuts.

Table 4.3: Dimuon invariant mass resolutions for Run 1 and Run 2 data, as obtained from the interpolation of charmonium and bottomonium resonances.

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

4.3.3 Tail parameters

Two parameters determine the behaviour of the Crystal Ball tail: α represents the *transition point* at which the gaussian tail is replaced by the power-law, with n being its *exponent*. To evaluate their values, the true⁴ invariant mass distribution of a signal MC sample is smeared according to the measured resolutions of Table 4.3, and then fitted with a Crystal Ball function to extract the parameters α and n . The results are listed in Table 4.4, where the quoted errors include a systematic contribution obtained by repeating the above procedure with the mass resolution value fluctuating within its uncertainty.

Table 4.4: Crystal Ball tail parameters in Run 1 and Run 2 data.

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

⁴The *true* Monte Carlo values are the ones obtained from the generator, thus independently from the detector effects.

4.4 Normalisation

One can count the number of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ events via an invariant mass fit of the candidates and calculate the branching fraction of the process via the basic equation (2.3), as:

$$\mathcal{B}(B_{d,s}^0 \rightarrow \mu^+ \mu^-) = \frac{N_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}}{\mathcal{L}_{int} \times \sigma_{pp \rightarrow b\bar{b}} \times 2 \times f_{d,s} \times \epsilon_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}}, \quad (4.6)$$

i.e. the number of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ events divided by the total number of produced $B_{d,s}^0$ (or $\bar{B}_{d,s}^0$), given by the product of the number of produced $b\bar{b}$ pairs and the hadronisation fractions $f_{d,s}$, times the total efficiency and geometrical acceptance of the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ channel. Although feasible, Eq. (4.6) is affected by large uncertainties stemming from the measurements of the cross section and the integrated luminosity.

To improve the precision, a *normalisation channel* is used: the number of events of a well-known process is measured so that the branching fraction can be expressed as the ratio between the observed $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidates and the normalisation candidates, as

$$\mathcal{B}(B_{d,s}^0 \rightarrow \mu^+ \mu^-) = \underbrace{\frac{\mathcal{B}_{norm}}{N_{norm}}}_{\alpha_d} \times \overbrace{\frac{\epsilon_{norm}}{\epsilon_{sig}} \times \frac{f_{norm}}{f_{d,s}}}_{\alpha_s}} \times N_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}, \quad (4.7)$$

where α_d and α_s are called *normalisation factors* for $B_d^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$, respectively.

The approach of Eq. (4.7) requires to calculate the normalisation channel yield and efficiency, but avoids to use the absolute number of produced B mesons. To minimise the systematic error, the normalisation channel has to be similar to the signal as far as trigger, reconstruction and selection are concerned. To this purpose, two normalisation channels are employed in the present analysis:

1. $B^+ \rightarrow J/\psi K^+$, with $J/\psi \rightarrow \mu^+ \mu^-$, which has a very similar muon trigger selection with respect to $B_{d,s}^0 \rightarrow \mu^+ \mu^-$,
2. $B_d^0 \rightarrow K^+ \pi^-$, which is a two-body B decay and therefore exhibits a similar reconstruction and topology with respect to $B_{d,s}^0 \rightarrow \mu^+ \mu^-$.

Both channels have large yields and precisely measured branching fractions.⁵ The resulting normalisation factors are then combined, as described in Sec. 4.4.1.

⁵ $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.026 \pm 0.031) \times 10^{-3}$, $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033)\%$ and $\mathcal{B}(B_d^0 \rightarrow K^+ \pi^-) = (1.96 \pm 0.05) \times 10^{-5}$ [158].

Selection efficiencies

The total efficiencies of Eq. (4.7) factorise as the product of detector acceptance, reconstruction and selection, and trigger, so that

$$\frac{\epsilon_{norm}}{\epsilon_{sig}} = \frac{\epsilon_{norm}^{Acc}}{\epsilon_{sig}^{Acc}} \times \frac{\epsilon_{norm}^{RecSel|Acc}}{\epsilon_{sig}^{RecSel|Acc}} \times \frac{\epsilon_{norm}^{Trig|RecSel}}{\epsilon_{sig}^{Trig|RecSel}}, \quad (4.8)$$

where the efficiency of each step is conditional with respect to the previous one. The detector acceptance is evaluated with simulated samples, while the combined effect of reconstruction and selection is firstly evaluated on simulated samples and then corrected for data-MC differences where needed. The PID selection efficiency is evaluated from data control samples and the trigger efficiencies are evaluated from data with the TISTOS method described in Sec. 2.5.3.

Yields of the normalisation channels

Candidates from the two normalisation channels are required to pass a selection similar to the signal one, as reported in Table 4.1, with the exception of the tight muon PID selection (3.3), which is only applied to the signal. A fit to the invariant mass of the candidates surviving the selection is then performed to extract the yields.

The fit model used to estimate the number of $B^+ \rightarrow J/\psi K^+$ candidates consists of the following components:

- An *Hypatia* [160] function for the $B^+ \rightarrow J/\psi K^+$, whose parameters are constrained from MC sample within their errors,
- A *RooKeysPdf*⁶ for the misidentified $B^+ \rightarrow J/\psi \pi^+$ component, evaluated by reconstructing a sample of $B^+ \rightarrow J/\psi \pi^+$ decays under the $B^+ \rightarrow J/\psi K^+$ hypothesis,
- An exponential for the combinatorial background, whose slope is free to vary.

The $B^+ \rightarrow J/\psi \pi^+$ yield is constrained to the $B^+ \rightarrow J/\psi K^+$ one by means of the ratio measured in [163], corrected for the ratio of the selection efficiencies of the two channels. The well-known mass of the J/ψ is also constrained. The fit results are shown in Fig. 4.11, which directly give the normalisation yield for Eq. (4.7). The yield of the $B_d^0 \rightarrow K^+ \pi^-$ channel is evaluated with the same selection and fit model described in Sec. 4.2.4 for the BDT calibration.

⁶The *RooKeys* [161] is a gaussian kernel density estimate implemented in the *Roofit* package [162] for ROOT.

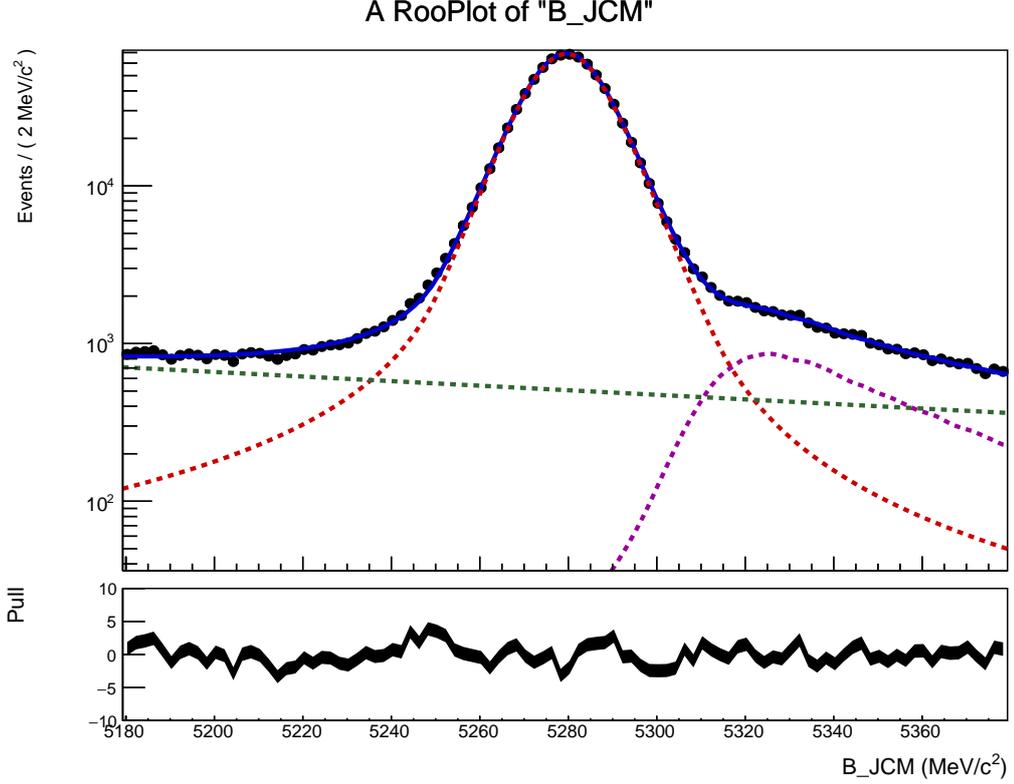


Figure 4.11: Invariant mass spectrum of the $B^+ \rightarrow J/\psi K^+$ candidates in 2016 data. The fit (solid blue) is superimposed together with its components: $B^+ \rightarrow J/\psi K^+$ (dashed red), $B^+ \rightarrow J/\psi \pi^+$ (dashed purple) and combinatorial (dashed green).

4.4.1 Normalisation factors

In the α_s computation, the f_s/f_d ratio (2.6) has been used, and a check on its energy dependence has been performed by comparing the efficiency corrected ratio of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates: their relative production is found to be stable when passing from 7 to 13 TeV.

The normalisation factors are evaluated by combining the results from $B^+ \rightarrow J/\psi K^+$ and $B_d^0 \rightarrow K^+ \pi^-$, where the correlations between the systematic uncertainty in the tracking efficiency and f_s/f_d are accounted for. The results are shown in Table 4.5, where the combined numbers for Run 1 and Run 2 account for the correlation due to the branching fraction of the normalisation channel. In the same table, the number of expected $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ signal events are listed, for the SM branching fraction and for the one measured in [4]. The main uncertainty in the normalisation procedure stems from the hadronisation

Table 4.5: $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ normalisation factors and expected yields for Run 1 and Run 2 data. The expected signal yields are estimated assuming standard model branching fractions (SM) and the latest most precise experimental branching fraction measurements [4] (LHC).

$B_s^0 \rightarrow \mu^+\mu^-$	$\alpha_s \times 10^{10}$	$N_{expected}^{SM}$	$N_{expected}^{LHC}$
Run 1	1.071(72)	34.2 ± 3.2	25.2 ± 6.8
Run 2	1.306(95)	28.0 ± 2.7	20.7 ± 5.6
Total	0.588(38)	62.2 ± 5.6	45.9 ± 12.3
$B_d^0 \rightarrow \mu^+\mu^-$	$\alpha_d \times 10^{11}$	$N_{expected}^{SM}$	$N_{expected}^{LHC}$
Run 1	2.877(101)	3.7 ± 0.3	13.6 ± 5.6
Run 2	3.521(155)	3.0 ± 0.3	11.1 ± 4.6
Total	1.583(44)	6.7 ± 0.6	24.6 ± 10.1

fraction ratio f_s/f_d . This quantity can be avoided if another B_s^0 decay is used as a normalisation channel: the use of $B_s^0 \rightarrow J/\psi\phi$ as a third normalisation channel was studied, and the improvement was found to be negligible due to its branching fraction uncertainty, even after including the new measurement from BELLE [164].

Chapter 5

Exclusive backgrounds

Besides combinatorial background, several exclusive decays pollute the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signal mass region, as shown in Fig. 5.1. The most relevant are:

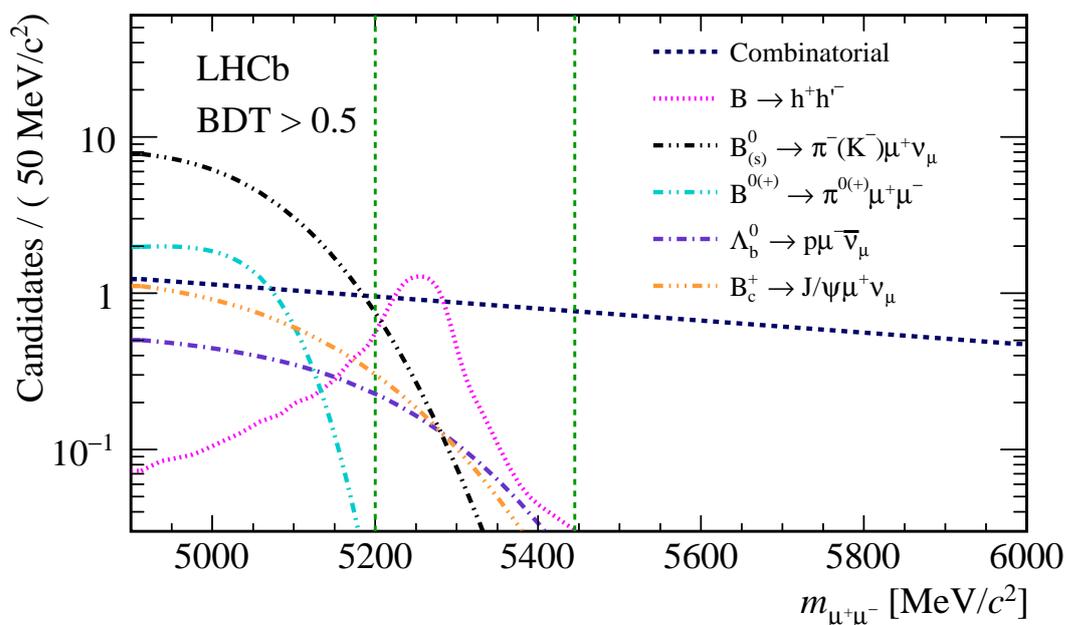


Figure 5.1: Mass *pdfs* of all the backgrounds sources in the region $\text{BDT} > 0.5$. The signal region is indicated by the green vertical lines.

- $B_{d,s}^0 \rightarrow h^+h'^-$ ($h = \pi, K$) decays, when both the pion and the kaon are misidentified as muons, represent a peaking background which mainly affects the $B_d^0 \rightarrow \mu^+\mu^-$ signal region;

semileptonic decays:

- $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays, when the pion or the kaon is misidentified as a muon, pollute the left mass sideband;
- $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ decays, when the proton is misidentified as a muon, pollute both the left mass sideband and the $B_d^0 - B_s^0$ mass regions;
- $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ and $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays, having two real muons in the final state, pollute the left mass sideband and the full mass region, respectively.

Since the signal fit (Sec. 6.1) is performed in bins of BDT, the yields and the mass shapes for all the above background sources have to be estimated in each BDT bin.

In this work, significant improvements have been shown in rejecting the backgrounds since the last published analysis [4]. The tight muon PID selection based on the ProbNN variables strongly reduces the dangerous $B_{d,s}^0 \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ decays, which enter into the signal mass region. In addition, not only the background rejection is relevant to the determination of the signal sensitivity, but also the systematic uncertainties on their yield for each BDT bin, and especially those affecting the peaking $B_{d,s}^0 \rightarrow h^+ h'^-$. For this reason, a data-driven method has been developed to estimate the most abundant background sources: $B_{d,s}^0 \rightarrow h^+ h'^-$, $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, providing a reliable yield estimate and ultimately a more accurate signal fit.

The sections 5.1 and 5.2 are devoted to the estimate of all the background sources using the same strategy of the past analysis editions. In Sec. 5.3, a discussion on the data-driven background estimate is given, while in Sec. 5.4 the systematic uncertainties are discussed and a detailed summary of all of the results is given.

5.1 $B_{d,s}^0 \rightarrow h^+ h'^-$

$B_{d,s}^0 \rightarrow h^+ h'^-$ events represent the most dangerous background source, since they are characterised by a peaking shape almost centered at the B_d^0 mass. The $B_{d,s}^0 \rightarrow h^+ h'^- \rightarrow \mu^+ \mu^-$ yield is evaluated according to:

$$N_{B_{d,s}^0 \rightarrow h^+ h'^- \rightarrow \mu^+ \mu^-} \equiv N_{hh \rightarrow \mu\mu} = \frac{N_{B_{d,s}^0 \rightarrow h^+ h'^-}^{\text{TIS}}}{\epsilon^{\text{TIS}} \times \epsilon^{\text{HLT2}}} \times \epsilon_{B_{d,s}^0 \rightarrow \mu^+ \mu^-}^{\text{Trig|Sel}} \times \epsilon_{hh \rightarrow \mu\mu}, \quad (5.1)$$

where the number of $B_{d,s}^0 \rightarrow h^+ h'^-$ TIS events is evaluated by correcting the number of $B_d^0 \rightarrow K^+ \pi^-$ TIS events, measured in the normalisation (Sec. 4.4), for the expected fraction of this mode. The first factor in Eq. 5.1 thus represents

the number of $B_{d,s}^0 \rightarrow h^+h'^-$ events corrected by their trigger efficiency, i.e. the L0 and HLT1 TIS efficiency, computed from data control sample, and the HLT2 efficiency, evaluated from MC. Since the estimate is made for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ selection, including trigger, the $B_{d,s}^0 \rightarrow h^+h'^-$ corrected yield has to be multiplied by the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ trigger efficiency, which is obtained from simulation after applying the selection. The last ingredient for the computation is the probability to misidentify both hadrons as muons, which is evaluated from data with the method described in Sec. 3.3 for each one of the four $B_{d,s}^0 \rightarrow h^+h'^-$ modes: $B_d^0 \rightarrow \pi^+\pi^-$, $B_d^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$. The double misidentification probability, $\epsilon_{hh \rightarrow \mu\mu}$, is then evaluated by weighting each mode according to its expectation, as

$$\begin{aligned} \epsilon_{hh \rightarrow \mu\mu} = & \left[\epsilon_{B_d^0 \rightarrow \pi^+\pi^- \rightarrow \mu^+\mu^-} \times \mathcal{B}(B_d^0 \rightarrow \pi^+\pi^-) + \epsilon_{B_d^0 \rightarrow K^+\pi^- \rightarrow \mu^+\mu^-} \times \mathcal{B}(B_d^0 \rightarrow K^+\pi^-) \right. \\ & \left. + \epsilon_{B_s^0 \rightarrow K^+\pi^- \rightarrow \mu^+\mu^-} \times \mathcal{B}(B_s^0 \rightarrow K^+\pi^-) \frac{f_s}{f_d} + \epsilon_{B_s^0 \rightarrow K^+K^- \rightarrow \mu^+\mu^-} \times \mathcal{B}(B_s^0 \rightarrow K^+K^-) \frac{f_s}{f_d} \right] \\ & \times \frac{1}{\left[\mathcal{B}(B_d^0 \rightarrow \pi^+\pi^-) + \mathcal{B}(B_d^0 \rightarrow K^+\pi^-) + \mathcal{B}(B_s^0 \rightarrow K^+\pi^-) \frac{f_s}{f_d} + \mathcal{B}(B_s^0 \rightarrow K^+K^-) \frac{f_s}{f_d} \right]}. \end{aligned} \quad (5.2)$$

The results are given in Table 5.1, where the uncertainties are dominated by the error on the PID determination. Table 5.2 summarises all the inputs needed

Table 5.1: Double misID probability in units of 10^{-6} for Run 1, 2015 and 2016 data. The $\text{PID}_{\mu,4}^{\text{MC12}}$ selection is used for Run 1 and 2015 data, whereas $\text{PID}_{\mu,8}^{\text{MC15}}$ is used for 2016 data, as discussed in Sec. 3.3.1. The quoted errors include MC statistics and PID efficiency statistical uncertainties for the single modes, and the branching fractions and f_s/f_d for the average.

	$B_d^0 \rightarrow \pi^+\pi^-$	$B_s^0 \rightarrow K^+K^-$	$B_d^0 \rightarrow K^+\pi^-$	$B_s^0 \rightarrow K^+\pi^-$	$\epsilon_{hh \rightarrow \mu\mu}$
Run 1	10.8 ± 0.4	1.9 ± 0.1	3.9 ± 0.2	4.0 ± 0.2	4.6 ± 0.2
2015	7.1 ± 0.5	1.9 ± 0.2	3.2 ± 0.3	3.3 ± 0.3	3.6 ± 0.2
2016	8.6 ± 0.1	1.22 ± 0.03	2.79 ± 0.04	2.83 ± 0.05	3.39 ± 0.07

for the $B_{d,s}^0 \rightarrow h^+h'^-$ background estimate as for Eq. (5.1), which yields the results reported in the last row.

Table 5.2: Inputs for the computation of the $B_{d,s}^0 \rightarrow h^+h'^-$ peaking background (Eq. (5.1)) for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
$N_{B_d^0 \rightarrow K^+\pi^-}^{\text{TIS}}$	24845 ± 1389	8552 ± 848	28411 ± 2548
$\epsilon_{B_{d,s}^0 \rightarrow \mu^+\mu^-}^{\text{Trig Sel}}$	$(94.1 \pm 2.0)\%$	$(96.1 \pm 3.2)\%$	$(97.3 \pm 1.3)\%$
ϵ^{TIS}	$(5.48 \pm 0.04)\%$	$(9.19 \pm 0.20)\%$	$(7.94 \pm 0.09)\%$
ϵ^{HLT2}	$(74.61 \pm 0.17)\%$	$(88.85 \pm 0.11)\%$	$(91.20 \pm 0.07)\%$
$\epsilon_{hh \rightarrow \mu\mu}$	$(4.6 \pm 0.2) \times 10^{-6}$	$(3.6 \pm 0.2) \times 10^{-6}$	$(3.39 \pm 0.07) \times 10^{-6}$
$N_{hh \rightarrow \mu^+\mu^-}$	4.39 ± 0.36	0.59 ± 0.06	2.12 ± 0.10

Estimate as a function of the BDT

In order to take into account the dependence of the double misID from the BDT, the former has been evaluated as a function of the BDT bin, and the resulting values are listed in Table 5.3. The observed dependency is large, and it is given by the PID-BDT correlation through the momentum of the selected candidates. The number of $B_{d,s}^0 \rightarrow h^+h'^-$ events in each BDT bin is then computed using the same Eq. (5.1), where the double misID in bins of BDT is used. However, the BDT dependence of Eq. (5.1) is more complex, and appears in two more factors:

- The ratio of trigger efficiencies $\epsilon_{B_{d,s}^0 \rightarrow \mu^+\mu^-}^{\text{Trig|Sel}} / \epsilon^{\text{TIS}} \times \epsilon^{\text{HLT2}}$: while the BDT-integrated values are partially extracted from data control samples, their BDT dependence is taken from $B_{d,s}^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow K^+\pi^-$ MC samples, since not enough statistics is available for a reliable data-driven evaluation.
- The actual value of the BDT *pdf* for $B_{d,s}^0 \rightarrow h^+h'^-$ double misID events before the PID selection: this is assumed to be the same as the signal and is given by the fractions obtained in the BDT calibration, shown in Fig. 4.8.

The detailed estimate in bins of BDT, as well as the relative discussion, are given in Sec. 5.4.1, where the comparison with the data-driven estimate (Sec. 5.3) is also

Table 5.3: Double misID probability in units of 10^{-6} as a function of the BDT bin for Run 1, 2015 and 2016 data. The quoted errors include MC statistics, PID efficiency statistical uncertainties, and the errors from the branching fractions and f_s/f_d .

BDT range	Run 1	2015	2016
[0-0.25]	3.30 ± 0.15	2.52 ± 0.16	2.59 ± 0.08
[0.25-0.4]	3.77 ± 0.15	2.97 ± 0.17	2.92 ± 0.08
[0.4-0.5]	4.26 ± 0.15	3.39 ± 0.18	3.26 ± 0.08
[0.5-0.6]	4.65 ± 0.16	3.67 ± 0.19	3.49 ± 0.08
[0.6-0.7]	5.06 ± 0.16	4.00 ± 0.20	3.67 ± 0.07
[0.7-0.8]	5.57 ± 0.17	4.31 ± 0.22	3.99 ± 0.07
[0.8-0.9]	6.12 ± 0.18	4.66 ± 0.25	4.23 ± 0.07
[0.9-1.0]	6.69 ± 0.19	4.94 ± 0.28	4.44 ± 0.07

performed, so that a robust systematic uncertainty is assigned. The conclusive results are shown in Tables 5.18, 5.19 and 5.20.

Mass shape

The invariant mass *pdf* for doubly-misidentified $B_{d,s}^0 \rightarrow h^+h'^-$ decays is determined from simulated events by applying a momentum smearing that accounts for the hadron decays in flight. To overcome the lack of statistics which would occur after the PID selection, the events are not required to satisfy the IsMuon criterion: a per-event PID weight is instead applied. The mass spectra of each $B_{d,s}^0 \rightarrow h^+h'^-$ mode is evaluated separately, and the resulting distributions are combined according to the weights (5.2) to give the shape shown in Fig. 5.1.

5.2 Semileptonic backgrounds

Although semileptonic background decays mostly populate the left side of the invariant mass range, they directly affect the estimate of the combinatorial component, which is instead present inside the signal region. A careful evaluation of these components is mandatory so that the combinatorial yield extracted from the fit is not biased.

For each semileptonic channel x , its yield N_x is evaluated by normalising to the

$B^+ \rightarrow J/\psi K^+$ channel, according to:

$$\begin{aligned} N_x &= N_{B^+ \rightarrow J/\psi K^+} \frac{f_x}{f_d} \frac{\mathcal{B}_x}{\mathcal{B}_{B^+ \rightarrow J/\psi K^+}} \frac{\epsilon_x^{Tot}}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Tot}} \\ &\equiv \beta_x \times \epsilon_x^{Tot} \times \mathcal{B}_x, \end{aligned} \quad (5.3)$$

where the β normalisation factor is given by Eq. (5.4), and $\beta_s = \beta_u \times f_s/f_d$ is used for B_s^0 decays. The values of the normalisation factors needed for the background estimates are:

$$\begin{aligned} \beta_u^{bkg} &= \frac{N_{B^+ \rightarrow J/\psi K^+}}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \times \frac{1}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Gen} \times \epsilon_{B^+ \rightarrow J/\psi K^+}^{RecSel} \times \epsilon_{B^+ \rightarrow J/\psi K^+}^{Trig}}, \\ \beta_u^{bkg}(\text{Run 1}) &= (7.18 \pm 0.24) \times 10^{11}, \\ \beta_u^{bkg}(2015) &= (1.27 \pm 0.06) \times 10^{11}, \\ \beta_u^{bkg}(2016) &= (4.62 \pm 0.21) \times 10^{11}. \end{aligned} \quad (5.4)$$

For Λ_b^0 (Sec. 5.2.4) and B_c^+ (Sec. 5.2.5) decays, β_u will be used and the specific hadronisation factors will be absorbed in the selection efficiency, as explained in the relative sections.

The total efficiency entering Eq. (5.3) includes geometrical acceptance (generation), reconstruction and selection, PID and trigger:

$$\epsilon_x^{Tot} = \epsilon_x^{Gen} \times \epsilon_x^{RecSel|Gen} \times \epsilon_x^{PID|RecSel\&Gen} \times \epsilon_x^{Trig|PID\&RecSel\&Gen}. \quad (5.5)$$

All the above efficiencies are evaluated from simulated events but the PID, which is determined from data with the usual method of Sec. 3.3. In the following, the estimate (5.3) will be therefore referred to as *MC-driven*, given its strong dependence on the simulation. Such estimate will quote a total error which includes the uncertainties from:

- The branching fraction of the channel,
- The $B^+ \rightarrow J/\psi K^+$ normalisation, i.e. the β factors (5.4),
- The pertinent hadronisation factor,
- The total efficiency evaluated from the simulation,
- The data-driven PID efficiency.

Estimate as a function of the BDT

Since the yield estimate has to be performed in bins of BDT, so has to be to total efficiency. Moreover, the mass shapes of each channel are obtained via a fit to the corresponding MC sample for each BDT bin, as they are represented by separate components in the signal fit. As for the $B_{d,s}^0 \rightarrow h^+h'^-$ background, the detailed results of the semileptonic yields as a function of the BDT are given in Sec. 5.4.2, where, for the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ channels, the data-driven estimate (Sec. 5.3) is also discussed. The conclusive results are shown in Tables 5.18, 5.19 and 5.20.

5.2.1 $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$

$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays occur with a relative rate of $\mathcal{B} = (1.44 \pm 0.05) \times 10^{-4}$ [165], and constitute a background in the event that the pion is misidentified as a muon. The invariant mass of the two muon candidates is shifted to the left sideband due to the missing neutrino energy. Moreover, the BDT values of these candidates is low, since less “pointing” decays, i.e. when the two candidates are less likely to point towards the B vertex, are rejected more by the topologic discrimination of the BDT. Since the two effects are correlated, the invariant mass distribution has to be determined as a function of the BDT.

MC samples of about 6 million events each for Run 1 and Run 2 conditions of the detector are available for studying this channel. The samples were produced with a cut at the generation level of $m(\pi\mu) > 4500 \text{ MeV}/c^2$, which corresponds to about 190 million events produced within the detector acceptance. The form factors¹ used in the simulation follow the ISGW2 model, which is found not to be in agreement with recent data [166], as shown in Fig. 5.2. To fix this issue, per-event weights in two bins of q^2 have been calculated according to the ratio between the ISGW2 model and the fit curve [167]. The events are therefore scaled by

$$\begin{aligned} w_1 &= 0.747 \pm 0.019 & (q^2 \in [0, 2.5] \text{ GeV}^2/c^2), \\ w_2 &= 0.791 \pm 0.021 & (q^2 \in [2.5, 5] \text{ GeV}^2/c^2), \end{aligned} \quad (5.6)$$

where the errors are due to the MC statistics. Only the $q^2 < 5 \text{ GeV}^2/c^2$ region is interested by the reweight (5.6), as no larger q^2 values survive the invariant mass cut of the signal fit: $m(\mu\mu) > 4.9 \text{ GeV}/c^2$, as shown in Fig. 5.3. Clearly, this weighting procedure has to be validated, and this is done in Sec. 5.4.2 by

¹The form factor parametrises the hadronic contribution to the decay and needs to be computed non-perturbatively, for example with lattice QCD. It can be expressed as a function of the momentum transferred to the outgoing lepton-neutrino pair: $q = p_B - p_\pi$.

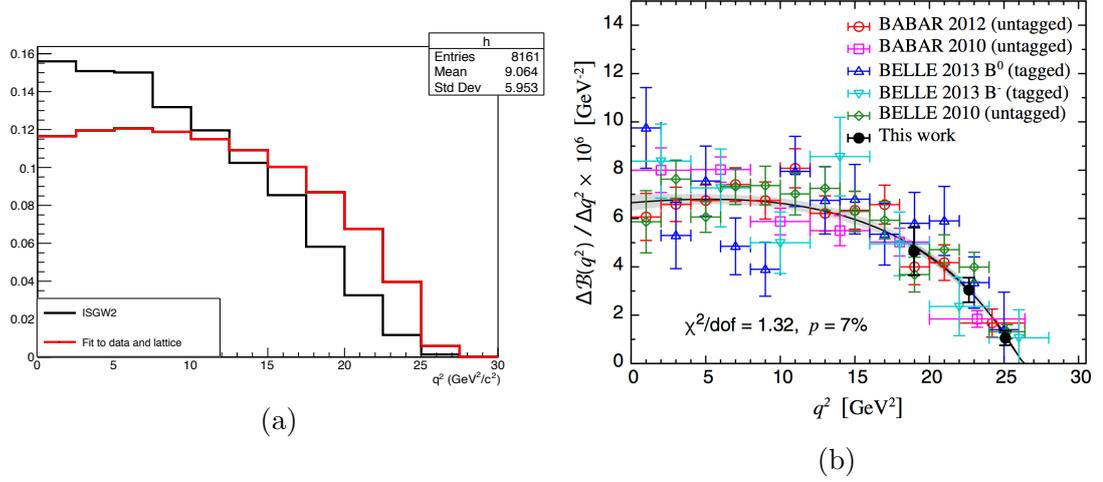


Figure 5.2: (a) q^2 spectra of $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays for the ISGW2 model (black) and for a fit to lattice data and recent results (red); both histograms are normalised to the same area. (b) Fit to lattice data and recent results [166] used for the left plot.

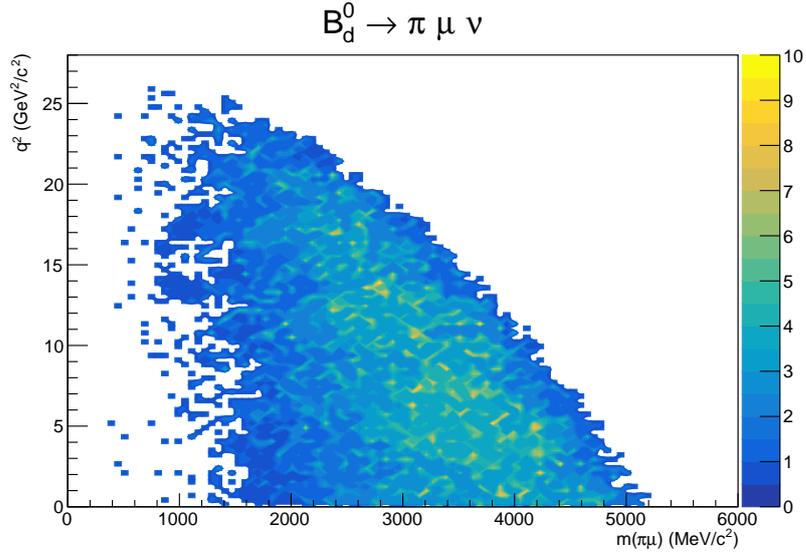


Figure 5.3: $\pi - \mu$ invariant mass versus q^2 in simulated $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays.

comparison with the independent data-driven estimate of Sec. 5.3. The inputs for the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ estimate are listed in Table 5.4, together with the resulting number of events, while the estimate as a function of the BDT is given in Tables 5.18, 5.19 and 5.20.

Table 5.4: Inputs to Eq. 5.3 for the computation of the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_u	$(7.18 \pm 0.24) \times 10^{11}$	$(1.27 \pm 0.06) \times 10^{11}$	$(4.62 \pm 0.21) \times 10^{11}$
\mathcal{B}	$(1.44 \pm 0.05) \times 10^{-4}$	$(1.44 \pm 0.05) \times 10^{-4}$	$(1.44 \pm 0.05) \times 10^{-4}$
ϵ^{Gen}	$(6.745 \pm 0.004) \times 10^{-3}$	$(6.995 \pm 0.003) \times 10^{-3}$	$(6.995 \pm 0.003) \times 10^{-3}$
ϵ^{Sel}	$(6.80 \pm 0.12) \times 10^{-5}$	$(5.16 \pm 0.17) \times 10^{-5}$	$(5.79 \pm 0.04) \times 10^{-5}$
ϵ^{Trig}	0.926 ± 0.011	0.935 ± 0.011	0.935 ± 0.011
$N_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}$	44.0 ± 2.3	6.2 ± 0.4	25.2 ± 1.5

The invariant mass *pdfs* of this channel are determined from a fit to MC samples after the full signal selection is applied. Given the small q^2 range accessible by $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays, no bias in the invariant mass distribution is introduced by the event reweight. A dependence on the PID cut was also investigated and the effect was found to be negligible, so no correction is needed for this effect either. The results are shown in Fig. 5.4, whereas the fit model, a convolution between an Argus [168] function and a gaussian, is discussed in Sec. 5.3.

5.2.2 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$

Similarly to the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays constitute a background when the kaon is misidentified as a muon. The branching fraction of this channel, although never measured, is expected to be of the same order of the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ one. Nevertheless, the expected background yield for this channel is lower, given the smaller fragmentation fraction of the B_s^0 meson and the slightly larger mass shift due to the kaon-muon mass difference.

Using an average of recent determinations of the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ form factors from lattice QCD [166, 169], and $V_{ub} = 4.09 \pm 0.39$ [158], the following estimate is made: $\mathcal{B} = (1.42 \pm 0.35) \times 10^{-4}$, which will be used for the background evaluation purpose. MC samples of about 6 million events for both Run 1 and Run 2 detector conditions are used, which are produced with the invariant mass cut $m(K\mu) > 4500 \text{ MeV}/c^2$. Analogously to the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$, a per-event weight is applied to correct the superseded ISGW2 form factor model, as shown in Fig. 5.5. However, in this case only lattice data are available [170], which are the same used in the branching fraction computation. The weights for the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ obtained from the above

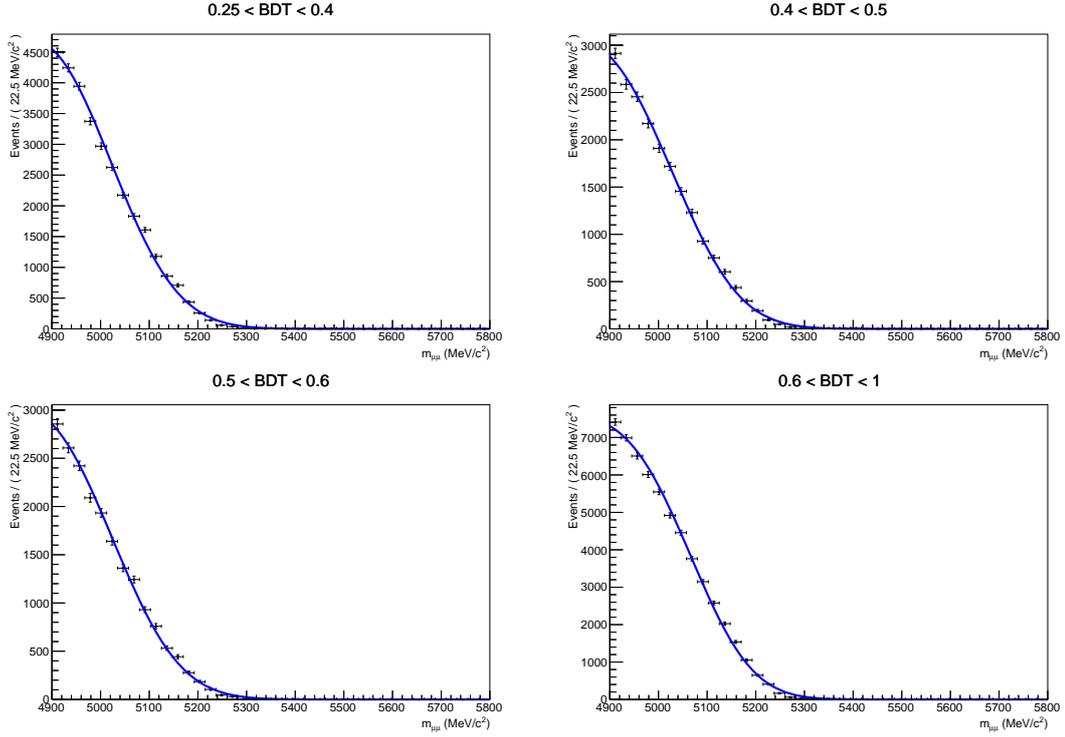


Figure 5.4: Argus*Gaussian fit to the invariant mass distribution of simulated $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays in the BDT bins used for the signal fit.

comparison are

$$\begin{aligned}
 w_1 &= 0.637 \pm 0.020 & (q^2 \in [0, 2.5] \text{ GeV}^2/c^2), \\
 w_2 &= 0.750 \pm 0.025 & (q^2 \in [2.5, 5] \text{ GeV}^2/c^2),
 \end{aligned} \tag{5.7}$$

where the errors are due to the statistics of the simulated sample. Even though form factors are affected by a large uncertainty, it is not included in Eq. (5.7), given its correlation with the branching fraction estimate. This choice is also dictated by the fact that most of the uncertainty affecting the yield is given by the comparison with the data-driven method.

The inputs for the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ estimate are listed in Table 5.5, as well as the total expected number of events. The estimate as a function of the BDT is given in Tables 5.18, 5.19 and 5.20.

The fact that the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ yield is about a factor four smaller with respect to the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ one, and that the mass distributions for these two decays are very similar, led to the choice of including this background component together with the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ one, i.e. the same mass shape is used and the resulting

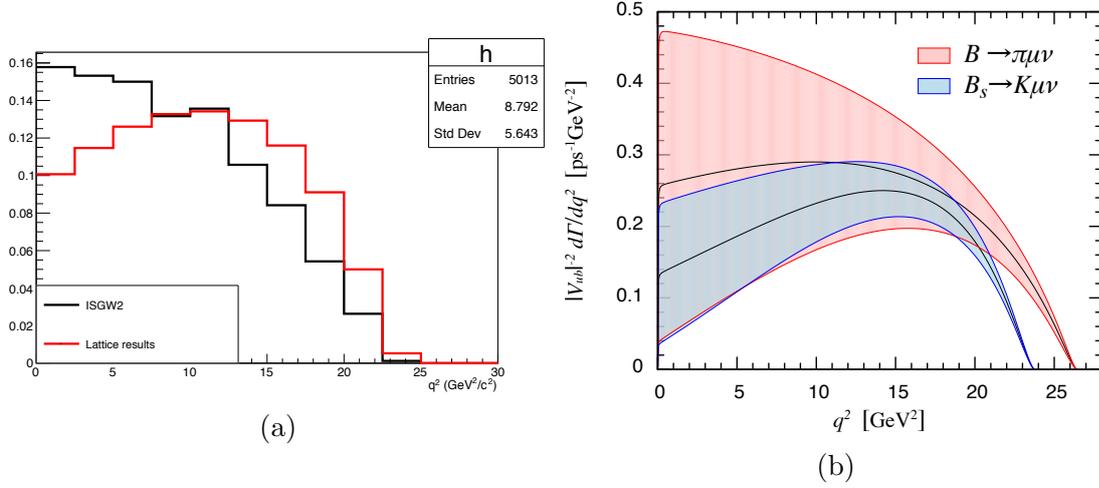


Figure 5.5: (a) q^2 spectra of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays for the ISGW2 model (black) and for a fit to recent lattice results (red); both histograms are normalised to the same area. (b) $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ q^2 spectra taken from [166].

Table 5.5: Inputs to Eq. 5.3 for the computation of the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_s	$(1.86 \pm 0.12) \times 10^{11}$	$(3.28 \pm 0.25) \times 10^{10}$	$(1.20 \pm 0.09) \times 10^{11}$
\mathcal{B}	$(1.42 \pm 0.35) \times 10^{-4}$	$(1.42 \pm 0.35) \times 10^{-4}$	$(1.42 \pm 0.35) \times 10^{-4}$
ϵ^{Gen}	$(9.141 \pm 0.005) \times 10^{-3}$	$(9.465 \pm 0.004) \times 10^{-3}$	$(9.465 \pm 0.004) \times 10^{-3}$
ϵ^{Sel}	$(2.34 \pm 0.07) \times 10^{-5}$	$(2.18 \pm 0.11) \times 10^{-5}$	$(1.82 \pm 0.02) \times 10^{-5}$
ϵ^{Trig}	0.898 ± 0.020	0.961 ± 0.012	0.961 ± 0.012
$N_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}$	5.03 ± 1.31	0.92 ± 0.24	2.81 ± 0.74

yields are summed up in the signal fit, as was already done in the previously published analysis [4].

5.2.3 $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B^0 \rightarrow \pi^0 \mu^+ \mu^-$

$B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ decays can mimic the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ signals owing to two real muons in the final state, which are characterised by a good vertex. The dimuon invariant mass cannot however reach the signal region, and it affects the left sideband only.

The branching fraction of the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ decay has been measured at LHCb

[171]: $\mathcal{B} = (1.83 \pm 0.25) \times 10^{-8}$, but the neutral mode $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ has not been yet observed. Nevertheless, a theoretical prediction on the branching fraction ratio of the charged over neutral mode is available [172]:

$$\frac{\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-)}{\mathcal{B}(B^0 \rightarrow \pi^0 \mu^+ \mu^-)} = 0.47^{+0.22}_{-0.18}, \quad (5.8)$$

from which $\mathcal{B} = (0.86 \pm 0.36) \times 10^{-8}$ is computed.

Since the two modes exhibit similar mass spectra and total selection efficiencies, in the previous analysis [4] the neutral mode yield was simply obtained by scaling the charged mode yield according to Eq. (5.8). However, the improved performance of new isolation algorithm (Sec. 4.2.2) led to a better rejection of the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ decays over $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ ones, given the extra charged track. Therefore, the strategy adopted in the previous analysis would lead to an underestimation of this background component. Although the total effect would have been small compared to the dominant $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ background source, the two modes are evaluated independently in the present analysis, and are summed up into a single component in the signal fit, given their very similar mass spectra.

Simulated samples of about 2 million events for each mode have been used for studying this channel, for both Run 1 and Run 2 detector conditions, in which the two muons are required to fly within the detector acceptance. The inputs for the $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ estimates are listed in Table 5.6 and Table 5.7, together with the resulting number of events. The estimate as a function of the BDT is given in Tables 5.18, 5.19 and 5.20.

Table 5.6: Inputs to Eq. 5.3 for the computation of the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_u	$(7.18 \pm 0.24) \times 10^{11}$	$(1.27 \pm 0.06) \times 10^{11}$	$(4.62 \pm 0.21) \times 10^{11}$
\mathcal{B}	$(1.83 \pm 0.25) \times 10^{-8}$	$(1.83 \pm 0.25) \times 10^{-8}$	$(1.83 \pm 0.25) \times 10^{-8}$
ϵ^{Gen}	0.2486 ± 0.0011	0.2503 ± 0.0010	0.2503 ± 0.0010
ϵ^{Sel}	$(3.75 \pm 0.004) \times 10^{-3}$	$(3.51 \pm 0.004) \times 10^{-3}$	$(3.70 \pm 0.004) \times 10^{-3}$
ϵ^{Trig}	0.958 ± 0.002	0.933 ± 0.003	0.933 ± 0.003
$N_{B^+ \rightarrow \pi^+ \mu^+ \mu^-}$	11.8 ± 1.7	1.9 ± 0.3	7.3 ± 1.1

The fits performed to extract the mass shapes for the sum of the two channels are shown in Fig. 5.6.

Table 5.7: Inputs to Eq. 5.3 for the computation of the $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_u	$(7.18 \pm 0.24) \times 10^{11}$	$(1.27 \pm 0.06) \times 10^{11}$	$(4.62 \pm 0.21) \times 10^{11}$
\mathcal{B}	$(0.86 \pm 0.36) \times 10^{-8}$	$(0.86 \pm 0.36) \times 10^{-8}$	$(0.86 \pm 0.36) \times 10^{-8}$
ϵ^{Gen}	0.251 ± 0.003	0.251 ± 0.003	0.251 ± 0.003
ϵ^{Sel}	$(3.80 \pm 0.004) \times 10^{-3}$	$(3.54 \pm 0.004) \times 10^{-3}$	$(3.73 \pm 0.004) \times 10^{-3}$
ϵ^{Trig}	0.954 ± 0.002	0.940 ± 0.003	0.940 ± 0.003
$N_{B^0 \rightarrow \pi^0 \mu^+ \mu^-}$	5.6 ± 2.3	0.9 ± 0.4	3.5 ± 1.5

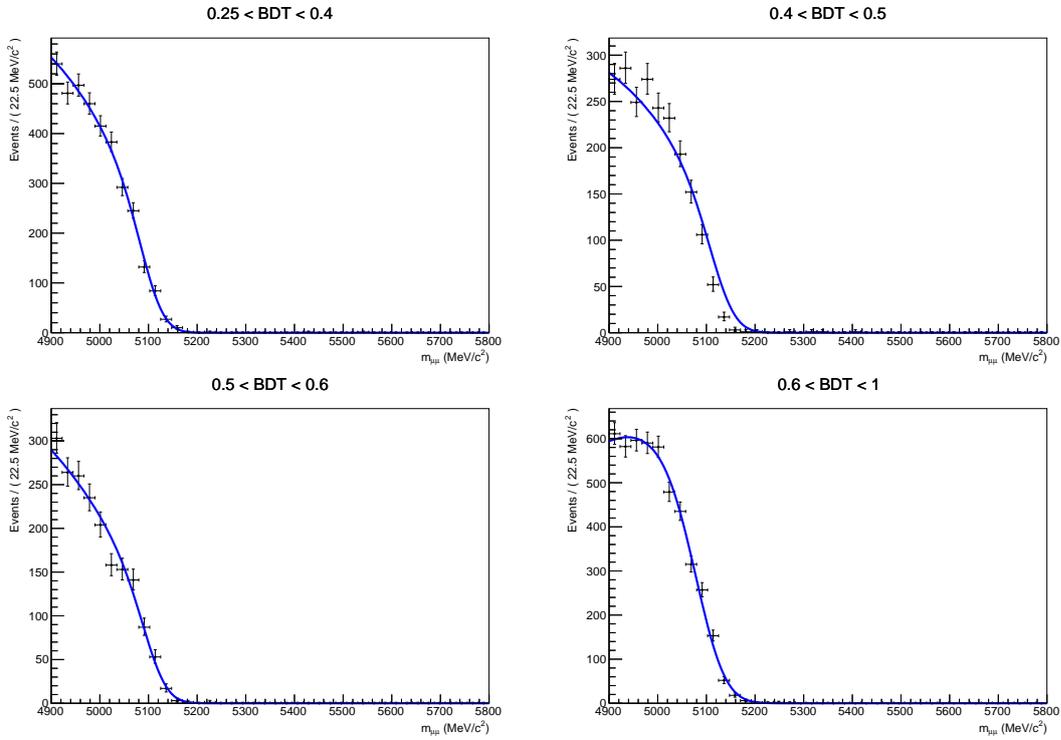


Figure 5.6: Argus*Gaussian fit to the invariant mass distribution of simulated $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ decays in the BDT bins used for the signal fit.

5.2.4 $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$

$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ decays enter into the signal selection in the event that the proton is misidentified as a muon. With respect to the old analysis, the proton to muon

misidentification rate has been decreased by a factor ~ 6 by the new PID cut (3.3), due to the $(1 - \text{ProbNN}_p)$ factor.

The branching fraction of this channel has been recently measured by the LHCb collaboration [173] to be $\mathcal{B} = (4.1 \pm 1.0) \times 10^{-4}$, where the uncertainty is dominated by the theoretical error on the extrapolation to the full phase space. As far as the hadronisation fraction is concerned, Eq. (2.7) is used, comprehensive of its p_T dependence, within the normalisation constant, as: $\beta_{\Lambda_b} = \beta_u \times 2 \times f_{\Lambda_b}/(f_u + f_d)$. However, since the $2 \times f_{\Lambda_b}/(f_u + f_d)$ factor is p_T dependent, it has been included as a per-event weight during the computation of the PID efficiency, and is therefore incorporated in the total selection efficiency. For this reason, the plain β_u values are used for the normalisation.

The study of the $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ has been conducted on samples of about 2 million events simulated using LQCD form factors, for both Run 1 and Run 2 detector conditions. An invariant mass cut $m(p\mu) > 4500 \text{ MeV}/c^2$ has been imposed at the generation level.

The total number of expected events is given in Table 5.8, as well as all the inputs needed for the computation. The estimate as a function of the BDT is given in Tables 5.18, 5.19 and 5.20. The yield uncertainty is in this case dominated by the

Table 5.8: Inputs to Eq. 5.3 for the computation of the $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_u	$(7.18 \pm 0.24) \times 10^{11}$	$(1.27 \pm 0.06) \times 10^{11}$	$(4.62 \pm 0.21) \times 10^{11}$
\mathcal{B}	$(4.1 \pm 1.0) \times 10^{-4}$	$(4.1 \pm 1.0) \times 10^{-4}$	$(4.1 \pm 1.0) \times 10^{-4}$
ϵ^{Gen}	$(1.34 \pm 0.01) \times 10^{-2}$	$1.32 \pm 0.01 \times 10^{-2}$	$1.32 \pm 0.01 \times 10^{-2}$
ϵ^{Sel}	$(1.4 \pm 0.6) \times 10^{-6}$	$(1.3 \pm 0.5) \times 10^{-6}$	$(0.65 \pm 0.15) \times 10^{-6}$
ϵ^{Trig}	0.801 ± 0.002	0.759 ± 0.002	0.759 ± 0.002
$N_{\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu}$	4.30 ± 2.11	2.09 ± 0.75	1.23 ± 0.48

proton to muon misID systematic, as explained in Sec. 3.3.2.

The invariant mass fits to the simulated samples are shown in Fig. 5.7, where the long tail entering the signal mass region is visible.

5.2.5 $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$

The last background component included in the signal fit is represented by $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ decays, which contain two oppositely-charged muons in the final state. In fact, a fake $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signal can be produced when the muon from the

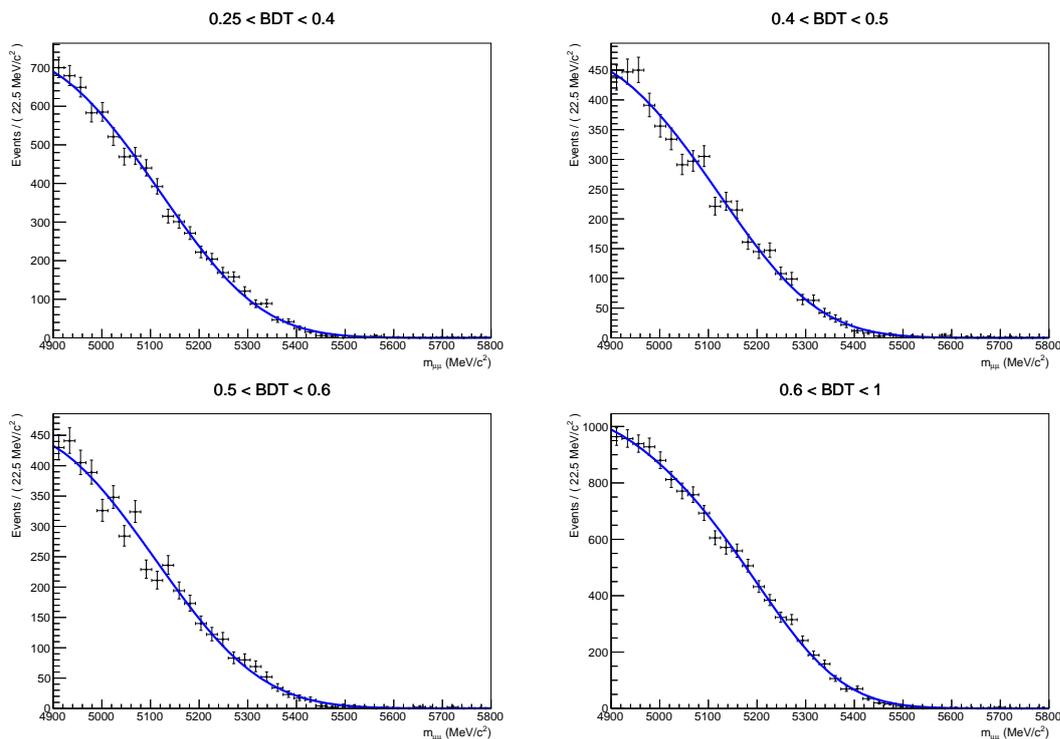


Figure 5.7: Argus*Gaussian fit to the invariant mass distribution of simulated $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ decays in the BDT bins used for the signal fit.

$J/\psi \rightarrow \mu^+\mu^-$ decay forms a good vertex with the oppositely charged one arising from the semileptonic decay.

Given the above topology, the isolation discriminant (Sec. 4.2.2) is expected to be rather effective in rejecting this component. Nonetheless, a simple J/ψ veto has been designed to further reduce this background. The veto consists of cutting away events in which a candidate muon, coupled to any other oppositely-charged muon within the event, forms an invariant mass which is $30 \text{ MeV}/c^2$ close to the nominal J/ψ mass, i.e. the event is rejected if the condition $(m_{\mu^+\mu^-} - m_{J/\psi}) < 30 \text{ MeV}/c^2$ is fulfilled. The veto rejects about 64% of the $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ events with a negligible $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signal loss of about 0.2%: for this reason, this veto is added to the selection.²

The absolute branching fraction of this channel has never been measured. However, since the $B^+ \rightarrow J/\psi K^+$ is used as a normalisation channel, two LHCb

²It should be noted that the majority of the vetoed events are however characterised by small BDT values, given the effectiveness of the isolation discriminant.

measurements can be exploited: the ratio [174]

$$\mathcal{R}_1 = \frac{\sigma(B_c^+) \mathcal{B}(B_c^+ \rightarrow J/\psi\pi)}{\sigma(B^+) \mathcal{B}(B^+ \rightarrow J/\psi K^+)} = (0.68 \pm 0.12) \times 10^{-2}, \quad (5.9)$$

which has been measured for $p_T > 4$ MeV/c and $2.5 < \eta < 4.5$, and the ratio [175]

$$\mathcal{R}_2 = \frac{\mathcal{B}(B_c^+ \rightarrow J/\psi\pi)}{\mathcal{B}(B_c^+ \rightarrow J/\psi\mu^+\nu_\mu)} = 0.0469 \pm 0.0054. \quad (5.10)$$

These two measurements can be combined into an *effective* $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ branching fraction as

$$\begin{aligned} \mathcal{B} &= \mathcal{R}_1/\mathcal{R}_2 \times \alpha \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) \\ &= (9.5 \pm 2.0) \times 10^{-6}, \end{aligned} \quad (5.11)$$

where α is a correction factor accounting for the ratio of $B^+ \rightarrow J/\psi K^+$ and $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ acceptances under the kinematic region of the measurement (5.9). This factor has been evaluated from an ad-hoc simulated sample without acceptance cuts.

The inputs for the $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ estimate are listed in Table 5.9, together with the resulting number of events, while the estimate as a function of the BDT is given in Tables 5.18, 5.19 and 5.20.

Table 5.9: Inputs to Eq. 5.3 for the computation of the $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.

	Run 1	2015	2016
β_u	$(7.18 \pm 0.24) \times 10^{11}$	$(1.27 \pm 0.06) \times 10^{11}$	$(4.62 \pm 0.21) \times 10^{11}$
Effective \mathcal{B}	$(9.5 \pm 2.0) \times 10^{-6}$	$(9.5 \pm 2.0) \times 10^{-6}$	$(9.5 \pm 2.0) \times 10^{-6}$
ϵ^{Gen}	$(4.350 \pm 0.006) \times 10^{-3}$	$(4.350 \pm 0.006) \times 10^{-3}$	$(4.350 \pm 0.006) \times 10^{-3}$
ϵ^{Sel}	$(4.09 \pm 0.006) \times 10^{-3}$	$(2.64 \pm 0.006) \times 10^{-3}$	$(2.78 \pm 0.007) \times 10^{-3}$
ϵ^{Trig}	0.964 ± 0.002	0.956 ± 0.003	0.956 ± 0.003
$N_{B_c^+ \rightarrow J/\psi\mu^+\nu_\mu}$	117.0 ± 25.1	13.3 ± 2.9	50.8 ± 11.1

The mass shapes are shown in Fig. 5.8, where the last two BDT bins have been merged because of lack of MC statistics.

5.3 Data-driven background estimate

The peaking background yields and mass *pdfs* have to be carefully estimated to improve the accuracy of the signal fit. In the current analysis, a new method

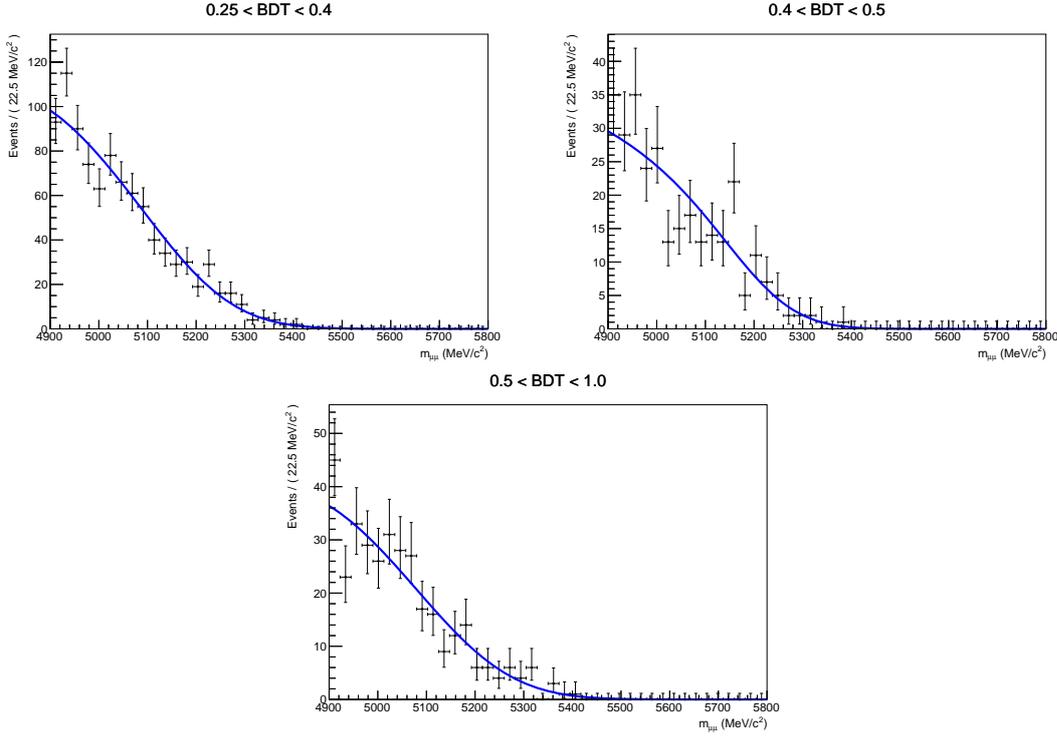


Figure 5.8: Argus*Gaussian fit to the invariant mass distribution of simulated $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ decays in the BDT bins used for the signal fit.

to evaluate the most crucial and abundant background sources ($B_{d,s}^0 \rightarrow h^+h'^-$, $B_d^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$) has been developed. By comparison with the MC-driven estimates (5.1) and (5.3), a robust systematic can be assigned.

The data driven method exploits $B_{d,s}^0 \rightarrow h^+h'^-$ data, whose selection is reported in Table 4.1, and a strong PID cut to build a $h - \mu$ ($h = \pi, K$), i.e. single misID, selection. In fact, when $B_{d,s}^0 \rightarrow h^+h'^-$ data are selected with one muon in the final state, the following components can be disentangled via an invariant mass fit:

- a *shoulder*, populated by the single misidentified backgrounds, i.e. $B_d^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ when the data are selected under the $\pi - \mu$ and $K - \mu$ PID selection, respectively,
- a *peak*, made up by $B_{d,s}^0 \rightarrow h^+h'^-$ events where one of the two hadrons is misidentified as a muon,
- *combinatorial background*.

An example is reported in Fig. 5.9, and the fit model is described in Sec. 5.3.2. The background yields evaluated in $h - \mu$ data, *data-driven* estimate in the following,

are then corrected by the efficiency ratios needed to obtain the yields entering the signal, i.e. $\mu - \mu$, selection, as explained in Sec. 5.3.3.

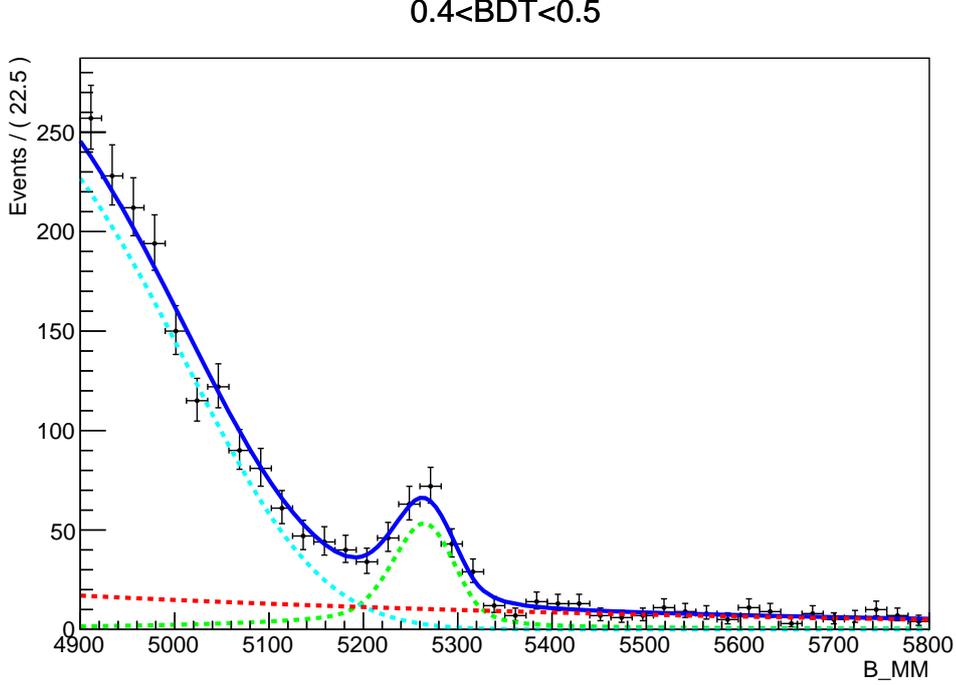


Figure 5.9: Example of a $\pi - \mu$ fit showing the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ shoulder (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ peak with single muon misidentification (green) and combinatorial (red) components. The fit (blue) is performed on Run 1 data in the BDT bin $[0.4, 0.5]$. The invariant mass is expressed in MeV/c^2 .

5.3.1 The $h - \mu$ selection

The data-driven background estimate makes use of $B_{d,s}^0 \rightarrow h^+ h'^-$ data, which are selected by the cuts reported in Table 4.1. For these events, the following PID operators have been built to realise the $h - \mu$ selections:

$$\begin{aligned}
 \overline{\text{PID}}_\pi &\equiv \text{ProbNN}_\pi \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K), \\
 \overline{\text{PID}}_K &\equiv \text{ProbNN}_K \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_\pi), \\
 \overline{\text{PID}}_\mu &\equiv \text{ProbNN}_\mu,
 \end{aligned} \tag{5.12}$$

where the overline is used to distinguish the $h - \mu$ PID selection from the $\mu - \mu$ one (Eq. (3.3)) used in the signal fit. The operators are aimed to remove the contamination coming from the wrong hadron type, and their cut values are reported

Cut name	Value	ProbNN Tuning	Applied on
$\overline{\text{PID}}_{\pi,5}^{\text{MC12}}$	$\overline{\text{PID}}_{\pi} > 0.5$	MC12TuneV2	Run 1, 2015 (π)
$\overline{\text{PID}}_{\mu,5}^{\text{MC12}}$	$\overline{\text{PID}}_{\mu} > 0.5$	MC12TuneV2	Run 1, 2015 (μ)
$\overline{\text{PID}}_{K,7}^{\text{MC12}}$	$\overline{\text{PID}}_K > 0.7$	MC15TuneV1	2016 (K)
$\overline{\text{PID}}_{\mu,7}^{\text{MC15}}$	$\overline{\text{PID}}_{\mu} > 0.7$	MC15TuneV1	2016 (μ)

Table 5.10: PID cuts defying the $h - \mu$ selection. Since two hadrons are present in the final state, the cut is applied in the form $(\overline{\text{PID}}_h \& \overline{\text{PID}}_{\mu}) \text{OR} (\overline{\text{PID}}_{\mu} \& \overline{\text{PID}}_h)$. As for the $\mu - \mu$ PID selection, the MC15TuneV1 of the ProbNN variables (Sec. 3.1.3) is adopted for 2016 data.

in Table 5.10. The efficiencies of all the reported cuts are evaluated, in each BDT bin, from calibration data with the usual technique described in Sec. 3.3, i.e. the data-driven muon identification and hadron to muon misidentification efficiencies are convoluted with the MC spectra of the specific background channel.

5.3.2 The $h - \mu$ fit model

The fit model for $h - \mu$ selected data has to deal with the interplay between the combinatorial and the shoulder component and with a $B_{d,s}^0 \rightarrow h^+ h'^-$ peak which is smeared by the PID selection. To this end, the three components described in the following are included in the total *pdf*.

The shoulder

The shoulder component constitutes the semileptonic signal in the $h - \mu$ fit: $B_d^0 \rightarrow \pi^- \mu^+ \nu_{\mu}$ ($h = \pi$) or $B_s^0 \rightarrow K^- \mu^+ \nu_{\mu}$ ($h = K$). Since partially reconstructed B decays are included as separate components in the signal fit, their shape is evaluated as a function of the BDT for both the $h - \mu$ and the signal fit using the same functional form. In the previous version of the analysis, the RooPhysBkg [176] *pdf* was used to model these decays in the signal fit. In the present analysis, this function is replaced by a convolution between an Argus function [168] and a gaussian, which exhibits more stability in the fit. The Argus function

$$A = x \sqrt{1 - \frac{x^2}{m^2}} \times e^{c(1 - \frac{x^2}{m^2})} \quad (5.13)$$

describes partially reconstructed B decays up to the kinematical endpoint m , with the parameter c driving its slope. The convolution with a gaussian centered at zero

smears the Argus endpoint to account for the single-misID selection. The parameters of the shoulder *pdfs* are determined from simulated samples of $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays for the $\pi - \mu$ and $K - \mu$ selections, respectively. The invariant mass shape is fitted in each BDT bin by splitting the sample according to the usual binning scheme (4.2), after applying the usual selection of Table 4.1. Since the PID cut (5.12) does not affect the shoulder shape, it is not applied so that enough simulated events are retained for the fit.

The peak

The peak exhibited by the $h - \mu$ data is composed by the peaks of the four $B_{d,s}^0 \rightarrow h^+ h'^-$ modes, smeared by the single misidentification selection (5.12). Each $B_{d,s}^0 \rightarrow h^+ h'^-$ component contributes to the total peak according to its branching fraction, multiplied by f_s/f_d for the B_s^0 modes, and PID efficiency, i.e. according to the weights (5.2) but under the single misID selection. The single misidentification efficiency, $\epsilon_{PID}(h - \mu)$, is evaluated as usual by convoluting the efficiencies obtained from calibration data with the MC samples of each one of the four $B_{d,s}^0 \rightarrow h^+ h'^-$ modes. The peak mass *pdf* is modelled by a double-sided Crystal Ball function, whose parameters are determined from a fit to a mixture of $B_{d,s}^0 \rightarrow h^+ h'^-$ simulated events where each mode contributes according to its relative abundancy. While the tail parameters are kept fixed, the mean and the width of the *pdf* are left free to vary in the $h - \mu$ data fit. Since the peak shape is not distorted by the BDT binning, the same *pdf* can be safely used across all the BDT bins.

Combinatorial background

Combinatorial background essentially affects only the second and third BDT bin³ of the $h - \mu$ fit. For these bins, care is required to handle the interplay between the combinatorial and shoulder shape. Since the exponential slopes are found from MC to be compatible across all the BDT bins, a single slope is determined from a simultaneous fit to all the bins. In this way, the fit stability is improved and a more reliable result is obtained.

The $h - \mu$ fit results are shown in Figs. 5.13, 5.14, 5.15, 5.16, 5.17 and 5.18 at the end of the chapter.

³The first BDT bin is combinatorial-dominated and not included in the $h - \mu$ and signal fits.

5.3.3 Yield computation

Once the yields of the peak ($N_{B \rightarrow hh}^{h-\mu}$) and shoulder ($N_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{\pi-\mu}$ and $N_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{K-\mu}$) components are obtained from the $h - \mu$ data fits, two corrections are applied to evaluate the corresponding yields in the signal selection, according to

$$\begin{aligned} N_{B \rightarrow hh}^{(1)} &= N_{B \rightarrow hh}^{\pi-\mu} \times \frac{\epsilon_{B \rightarrow hh}^{PID^{\mu\mu}}}{\epsilon_{B \rightarrow hh}^{PID^{\pi\mu}}} \times \frac{\epsilon_{B \rightarrow \mu\mu}^{Trig|\mu\mu Sel}}{\epsilon_{B \rightarrow hh}^{Trig|h\mu Sel}}, \\ N_{B \rightarrow hh}^{(2)} &= N_{B \rightarrow hh}^{K-\mu} \times \frac{\epsilon_{B \rightarrow hh}^{PID^{\mu\mu}}}{\epsilon_{B \rightarrow hh}^{PID^{K\mu}}} \times \frac{\epsilon_{B \rightarrow \mu\mu}^{Trig|\mu\mu Sel}}{\epsilon_{B \rightarrow hh}^{Trig|h\mu Sel}}, \end{aligned} \quad (5.14)$$

for the $B_{d,s}^0 \rightarrow h^+ h'^-$ ((1) and (2) label the estimates obtained from the $\pi - \mu$ and $K - \mu$ fits, respectively) and

$$\begin{aligned} N_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu} &= N_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{\pi-\mu} \times \frac{\epsilon_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{PID^{\mu\mu}}}{\epsilon_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{PID^{\pi\mu}}} \times \frac{\epsilon_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{Trig|\mu\mu Sel}}{\epsilon_{B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu}^{Trig|\pi\mu Sel}}, \\ N_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu} &= N_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{K-\mu} \times \frac{\epsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{PID^{\mu\mu}}}{\epsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{PID^{K\mu}}} \times \frac{\epsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{Trig|\mu\mu Sel}}{\epsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}^{Trig|K\mu Sel}}, \end{aligned} \quad (5.15)$$

for $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$. The first correction factor on both estimates concerns the PID: since the backgrounds are evaluated under the $h - \mu$ PID selection, their yields have to be corrected by the ratio between the $\mu - \mu$ (double misidentification) and $h - \mu$ (single misidentification) PID efficiencies. The second correction is represented by the ratio of the trigger efficiencies: since signal and $B_{d,s}^0 \rightarrow h^+ h'^-$ data are triggered by different lines and under two different PID selections, the ratio between the trigger efficiencies is evaluated on MC after applying the two PID selections. For the $B_{d,s}^0 \rightarrow h^+ h'^-$ estimate, the signal and $B_d^0 \rightarrow K^+ \pi^-$ trigger efficiencies are used in the ratio.

5.4 Results and systematic uncertainties

In this section, detailed results on the background estimates as a function of the BDT are given, following the methods explained in Sec. 5.1 and Sec. 5.2. Upon comparison with the data-driven estimate described in Sec. 5.3, systematic uncertainties are evaluated and added to the $B_{d,s}^0 \rightarrow h^+ h'^-$, $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ expected yields. The estimates of each background source as a function of the BDT are summarised in Tables 5.18, 5.19 and 5.20, which are used as inputs for the signal fit, described in Sec. 6.1.

5.4.1 $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$

The $B_{d,s}^0 \rightarrow h^+h'^-$ background yields are evaluated in each one of the 8 BDT bins defined by the scheme (4.2) from $B_d^0 \rightarrow K^+\pi^-$ events via Eq. (5.1), and are reported in Table 5.11. To the total uncertainty of all the factors entering this computation (Table 5.2), a systematic error is computed by comparison⁴ with the data-driven estimates 5.14, whose results are given in Table 5.12 and Table 5.13 for the $\pi - \mu$ and $K - \mu$ fit, respectively. The comparison is shown in Fig. 5.10. No systematic uncertainty can be computed for the first BDT bin, since the $h - \mu$ fits are dominated by combinatorial background. This will not anyway affect the final result, as the first BDT bin will not be used in the signal fit (Sec. 6.1). Since the $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$ yield can be extracted from both $\pi - \mu$ and $K - \mu$ data, a fit systematic (s_{syst}) is firstly defined as the error needed to recover a 1σ difference between the two data-driven estimates:

$$\frac{\left| N_{B \rightarrow hh}^{(\pi\mu)} - N_{B \rightarrow hh}^{(K\mu)} \right|}{\sqrt{\left(\Delta N_{B \rightarrow hh}^{(\pi\mu)} \right)^2 + \left(\Delta N_{B \rightarrow hh}^{(K\mu)} \right)^2 + s_{syst}^2}} = 1. \quad (5.16)$$

s_{syst} is then included as a systematic to the $K - \mu$ results, which are used as reference values due to their higher peak over shoulder ratio in the fit. The $B_{d,s}^0 \rightarrow h^+h'^-$ estimates from TIS events (Eq. (5.1)) are finally compared to the ones obtained with the $K - \mu$ data fit, comprehensive of the fit systematic. The same criterion used in Eq. (5.16) is adopted to assign the final systematic to the yield, to which s_{syst} is added in quadrature. The statistical error amounts to 6%, 11% and 5% for Run 1, 2015 and 2016 data respectively. Conversely, the systematic error in the last four, i.e. the most significant, BDT bins lies in the range 15 – 25% for Run 1 and 15 – 35% for 2016 data, respectively, whereas the situation is worse in the 2015 sample, which however has a negligible impact in the result as it corresponds to $\sim 7\%$ of the total integrated luminosity.

The procedure for calculating the systematic error gives zero for a couple of bins, since no discrepancy is observed in the comparison between the two estimates, nor in the semileptonic fit itself ($s_{syst} = 0$). This effect can be of course due to fluctuations in the uncertainties of the different estimates, and since it is not reasonable to have large discontinuities in the systematic error as a function of the BDT, the largest systematic error among the two adjacent bins is assigned.

The final $B_{d,s}^0 \rightarrow h^+h'^-$ estimates, which include both statistical and systematic uncertainties, are reported in Tables 5.18, 5.19 and 5.20.

⁴ Note that to evaluate this systematic uncertainty, the errors on the PID and trigger efficiencies are removed, as they are used in both estimates and hence are 100% correlated.

5.4.2 Semileptonic backgrounds

The semileptonic background yields per BDT bin are evaluated with the MC-driven formula (5.3), with an uncertainty that includes all the components listed in Sec. 5.2. The detailed results for the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ channels are given in Table 5.14 and Table 5.16: for these two channels, a systematic uncertainty is added by taking the difference between the MC-driven and data-driven estimates, Table 5.15 and Table 5.17, with the same criterion defined by Eq. 5.16. The comparisons between the two estimates are shown in Fig. 5.11 and Fig. 5.12 for $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, respectively. For the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ channel, the resulting systematic errors are below 15% in the last four BDT bins for all the datasets, confirming the reliability of the MC-driven estimate. Concerning the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ channel, the systematic errors are significantly larger than the statistical ones, but the final impact is however expected to be small given the fact that this background source amounts to only $\sim 20 - 25\%$ of the dominant $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ in the most significant BDT bins.

The final $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ estimates, comprehensive of both statistical and systematic uncertainties, are reported in Tables 5.18, 5.19 and 5.20, together with the other semileptonic decays.

5.4.3 Background yields in the signal region

It is worth to investigate the background contamination level inside the signal region, defined by the mass range [5200, 5445] MeV/c². To this purpose, the MC-driven estimate for the semileptonic backgrounds has been repeated by tightening the mass range to the signal region, while for the $B_{d,s}^0 \rightarrow h^+ h'^-$ background, a factor of $\sim 67\%$ takes into account the portion of the $B_{d,s}^0 \rightarrow h^+ h'^-$ mass spectrum lying within the signal mass region with respect to the total. The results, given in Table 5.21, Table 5.22 and Table 5.23, show how all the background sources directly impact the $B_d^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ signals. The $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ channels are not listed as their contribution is negligible. The $B_{d,s}^0 \rightarrow h^+ h'^-$ background is by far the dominant, also being loosely rejected by the BDT. In the signal region, the copious $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ background is surpassed by the $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ one, given its long right tail in the mass spectrum (Fig. 5.7), while the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ component accounts for less than 50% of the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ in the most significant BDT bins. The $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ component is also relevant in the 5th and 6th bins, given its long mass tail (Fig. 5.8). However, the combined effect of the J/ψ veto and the isolation variable reduces this component to negligible levels in the last two bins.

Table 5.11: Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$ events as extracted from $B_d^0 \rightarrow K^+\pi^-$ TIS events as a function of the BDT bin for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0-0.25]	0.64 ± 0.12	0.11 ± 0.02	0.44 ± 0.07
[0.25-0.4]	0.56 ± 0.07	0.07 ± 0.01	0.30 ± 0.04
[0.4-0.5]	0.40 ± 0.05	0.06 ± 0.01	0.20 ± 0.03
[0.5-0.6]	0.51 ± 0.04	0.06 ± 0.01	0.22 ± 0.03
[0.6-0.7]	0.48 ± 0.05	0.06 ± 0.01	0.22 ± 0.03
[0.7-0.8]	0.55 ± 0.05	0.07 ± 0.01	0.22 ± 0.03
[0.8-0.9]	0.59 ± 0.06	0.08 ± 0.01	0.24 ± 0.03
[0.9-1.0]	0.68 ± 0.06	0.08 ± 0.01	0.27 ± 0.03

Table 5.12: Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$ events as extracted from the $\pi - \mu$ fit as a function of the BDT bin for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0.25-0.4]	0.80 ± 0.11	0.20 ± 0.03	0.44 ± 0.05
[0.4-0.5]	0.64 ± 0.07	0.09 ± 0.02	0.33 ± 0.03
[0.5-0.6]	0.67 ± 0.07	0.10 ± 0.02	0.31 ± 0.03
[0.6-0.7]	0.64 ± 0.06	0.11 ± 0.02	0.27 ± 0.03
[0.7-0.8]	0.68 ± 0.07	0.08 ± 0.02	0.27 ± 0.04
[0.8-0.9]	0.82 ± 0.07	0.07 ± 0.02	0.39 ± 0.03
[0.9-1.0]	0.73 ± 0.06	0.09 ± 0.02	0.33 ± 0.03

Table 5.13: Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$ events as extracted from the $K - \mu$ fit as a function of the BDT bin for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0.25-0.4]	0.95 ± 0.10	0.13 ± 0.03	0.51 ± 0.04
[0.4-0.5]	0.62 ± 0.06	0.10 ± 0.02	0.33 ± 0.03
[0.5-0.6]	0.53 ± 0.06	0.13 ± 0.03	0.36 ± 0.03
[0.6-0.7]	0.63 ± 0.06	0.06 ± 0.02	0.28 ± 0.03
[0.7-0.8]	0.52 ± 0.07	0.06 ± 0.02	0.32 ± 0.03
[0.8-0.9]	0.70 ± 0.07	0.11 ± 0.03	0.31 ± 0.03
[0.9-1.0]	0.71 ± 0.07	0.13 ± 0.03	0.27 ± 0.03

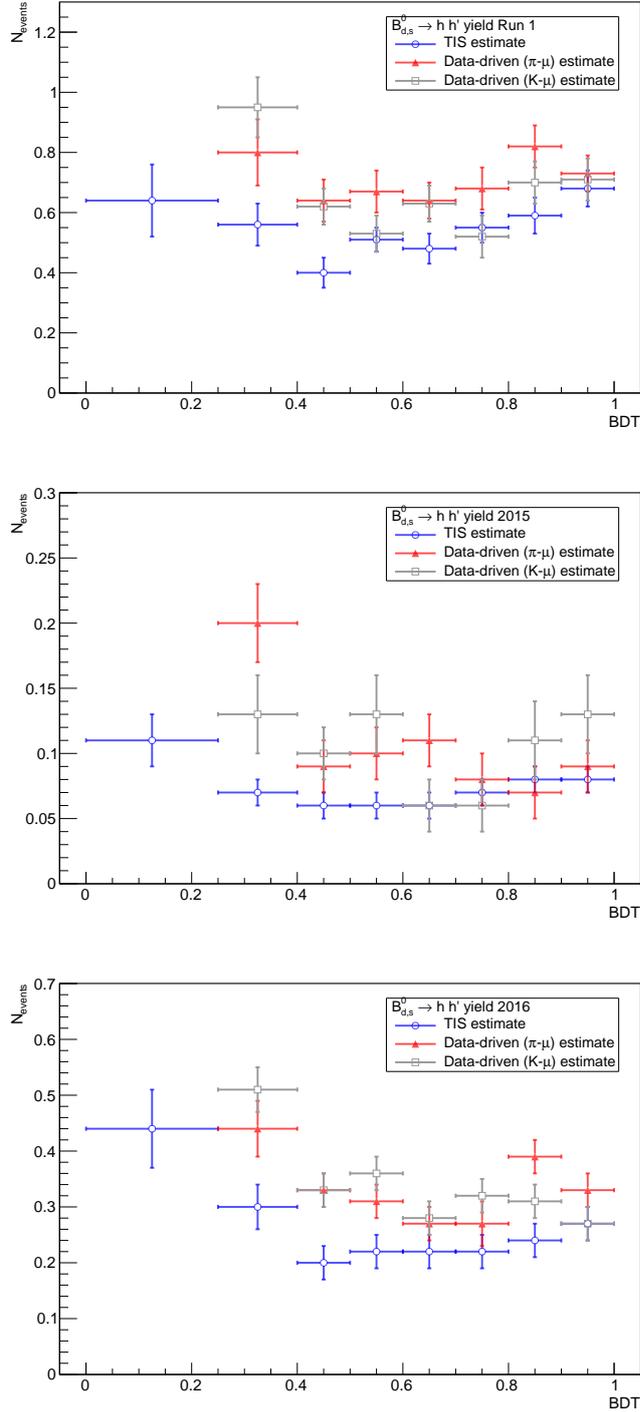


Figure 5.10: Comparison between the number of $B_{d,s}^0 \rightarrow h^+ h'^- \rightarrow \mu^+ \mu^-$ events from the $B_d^0 \rightarrow K^+ \pi^-$ TIS estimate (blue) and from the data-driven estimates (red and grey) as a function of the BDT bin for Run 1, 2015 and 2016 data.

Table 5.14: $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ MC-driven yields per BDT bin for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0-0.25]	13.57 ± 0.73	1.90 ± 0.13	7.81 ± 0.47
[0.25-0.4]	7.00 ± 0.38	1.02 ± 0.07	4.18 ± 0.25
[0.4-0.5]	4.56 ± 0.25	0.64 ± 0.05	2.64 ± 0.16
[0.5-0.6]	4.78 ± 0.26	0.66 ± 0.05	2.68 ± 0.16
[0.6-0.7]	4.72 ± 0.25	0.65 ± 0.05	2.64 ± 0.16
[0.7-0.8]	4.41 ± 0.24	0.59 ± 0.04	2.41 ± 0.15
[0.8-0.9]	3.46 ± 0.19	0.47 ± 0.04	1.92 ± 0.12
[0.9-1.0]	1.54 ± 0.09	0.23 ± 0.02	0.91 ± 0.06

Table 5.15: $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ yields per BDT bin as extracted from the $\pi - \mu$ fit for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0.25-0.4]	10.40 ± 0.60	1.87 ± 0.12	7.59 ± 0.34
[0.4-0.5]	6.56 ± 0.27	1.03 ± 0.08	5.12 ± 0.19
[0.5-0.6]	5.71 ± 0.22	0.81 ± 0.07	3.68 ± 0.14
[0.6-0.7]	5.60 ± 0.21	0.69 ± 0.06	3.71 ± 0.14
[0.7-0.8]	4.83 ± 0.20	0.60 ± 0.06	2.93 ± 0.14
[0.8-0.9]	3.66 ± 0.17	0.54 ± 0.05	2.21 ± 0.10
[0.9-1.0]	1.72 ± 0.11	0.18 ± 0.03	0.93 ± 0.07

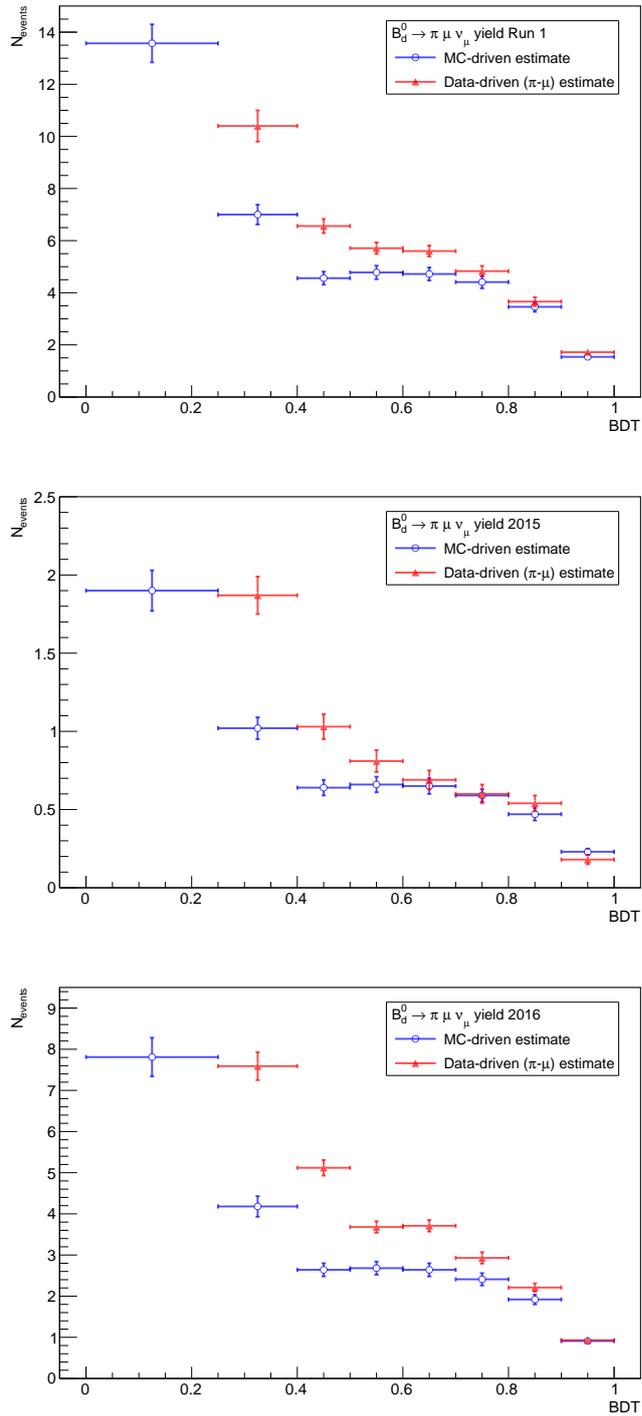


Figure 5.11: Comparison between the number of $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ events from the MC-driven estimate (blue) and from the data-driven estimate (red) as a function of the BDT bin for Run 1, 2015 and 2016 data.

Table 5.16: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ MC-driven yields per BDT bin for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0-0.25]	0.93 ± 0.24	0.16 ± 0.04	0.50 ± 0.13
[0.25-0.4]	0.62 ± 0.16	0.12 ± 0.03	0.36 ± 0.09
[0.4-0.5]	0.47 ± 0.12	0.10 ± 0.03	0.29 ± 0.08
[0.5-0.6]	0.63 ± 0.16	0.11 ± 0.03	0.34 ± 0.09
[0.6-0.7]	0.71 ± 0.18	0.13 ± 0.03	0.40 ± 0.10
[0.7-0.8]	0.76 ± 0.20	0.14 ± 0.04	0.41 ± 0.11
[0.8-0.9]	0.65 ± 0.17	0.12 ± 0.03	0.35 ± 0.09
[0.9-1.0]	0.27 ± 0.07	0.05 ± 0.01	0.16 ± 0.04

Table 5.17: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ yields per BDT bin as extracted from the $K - \mu$ fit for Run 1, 2015 and 2016 data.

BDT range	Run 1	2015	2016
[0.25-0.4]	1.07 ± 0.13	0.23 ± 0.06	0.92 ± 0.07
[0.4-0.5]	0.82 ± 0.07	0.19 ± 0.03	0.72 ± 0.05
[0.5-0.6]	1.03 ± 0.08	0.10 ± 0.03	0.60 ± 0.05
[0.6-0.7]	0.97 ± 0.08	0.17 ± 0.04	0.64 ± 0.05
[0.7-0.8]	1.13 ± 0.10	0.24 ± 0.04	0.72 ± 0.06
[0.8-0.9]	0.91 ± 0.09	0.10 ± 0.03	0.62 ± 0.06
[0.9-1.0]	0.32 ± 0.06	0.07 ± 0.03	0.31 ± 0.04

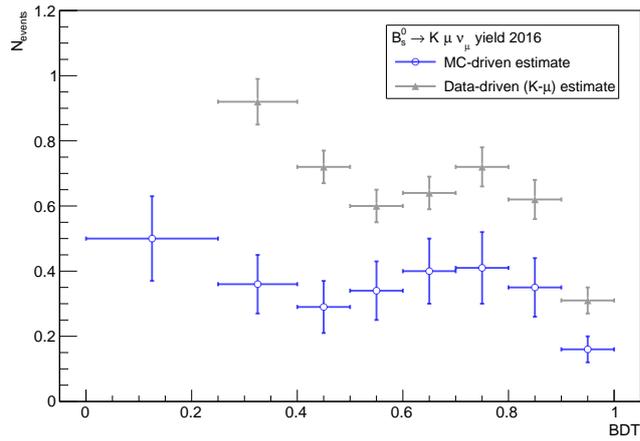
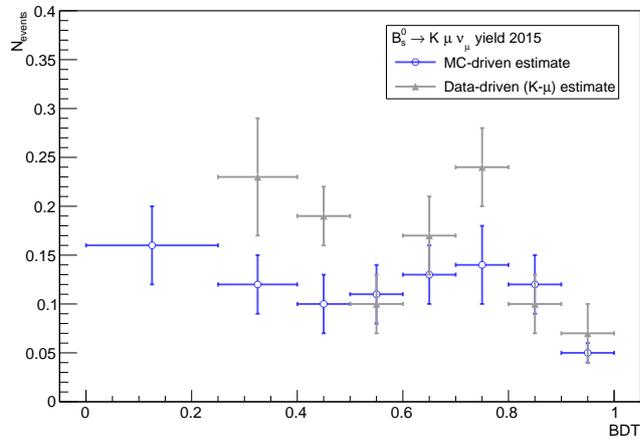
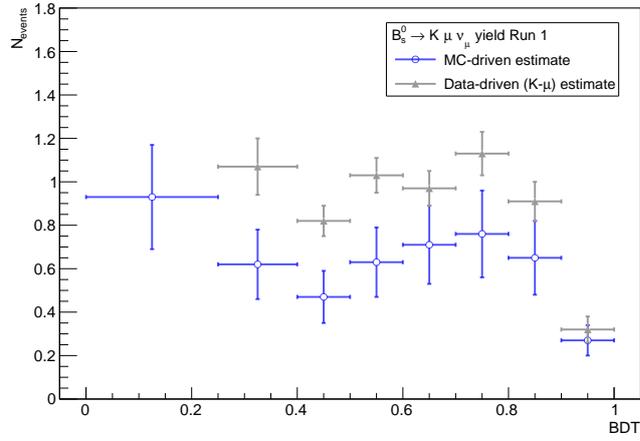


Figure 5.12: Comparison between the number of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ events from the MC-driven estimate (blue) and from the data-driven estimate (grey) as a function of the BDT bin for Run 1, 2015 and 2016 data.

Table 5.18: Background yields with total uncertainty for each BDT bin in Run 1 data.

BDT range	$B_{d,s}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$
[0-0.25]	0.64 ± 0.12	13.57 ± 0.74	0.93 ± 0.24	5.06 ± 0.72	1.66 ± 0.70	92.32 ± 19.9	1.11 ± 0.55
[0.25-0.4]	0.56 ± 0.38	7.00 ± 3.39	0.62 ± 0.43	2.33 ± 0.33	0.86 ± 0.36	13.17 ± 2.88	0.62 ± 0.31
[0.4-0.5]	0.40 ± 0.22	4.56 ± 1.98	0.47 ± 0.34	1.24 ± 0.18	0.60 ± 0.25	4.68 ± 1.06	0.43 ± 0.22
[0.5-0.6]	0.51 ± 0.11	4.78 ± 0.91	0.63 ± 0.40	1.06 ± 0.15	0.60 ± 0.25	3.45 ± 0.80	0.56 ± 0.27
[0.6-0.7]	0.48 ± 0.13	4.72 ± 0.85	0.71 ± 0.25	0.84 ± 0.12	0.61 ± 0.26	2.06 ± 0.50	0.56 ± 0.27
[0.7-0.8]	0.55 ± 0.15	4.41 ± 0.37	0.76 ± 0.35	0.61 ± 0.09	0.57 ± 0.24	1.02 ± 0.27	0.50 ± 0.24
[0.8-0.9]	0.59 ± 0.11	3.46 ± 0.33	0.65 ± 0.25	0.44 ± 0.07	0.47 ± 0.20	0.30 ± 0.11	0.36 ± 0.18
[0.9-1.0]	0.68 ± 0.13	1.54 ± 0.15	0.27 ± 0.11	0.19 ± 0.03	0.19 ± 0.08	0	0.12 ± 0.06

Table 5.19: Background yields with total uncertainty for each BDT bin in 2015 data.

BDT range	$B_{d,s}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$
[0-0.25]	0.11 ± 0.02	1.90 ± 0.13	0.16 ± 0.04	0.82 ± 0.12	0.26 ± 0.11	10.10 ± 2.22	0.16 ± 0.08
[0.25-0.4]	0.07 ± 0.07	1.02 ± 0.84	0.12 ± 0.10	0.38 ± 0.06	0.15 ± 0.06	1.56 ± 0.36	0.09 ± 0.05
[0.4-0.5]	0.06 ± 0.04	0.64 ± 0.38	0.10 ± 0.09	0.19 ± 0.03	0.09 ± 0.04	0.71 ± 0.17	0.07 ± 0.03
[0.5-0.6]	0.06 ± 0.07	0.66 ± 0.14	0.11 ± 0.10	0.17 ± 0.03	0.10 ± 0.04	0.44 ± 0.11	0.09 ± 0.04
[0.6-0.7]	0.06 ± 0.04	0.65 ± 0.14	0.13 ± 0.09	0.13 ± 0.02	0.10 ± 0.04	0.22 ± 0.07	0.09 ± 0.04
[0.7-0.8]	0.07 ± 0.04	0.59 ± 0.05	0.14 ± 0.10	0.11 ± 0.02	0.10 ± 0.04	0.19 ± 0.06	0.08 ± 0.04
[0.8-0.9]	0.08 ± 0.04	0.47 ± 0.04	0.12 ± 0.08	0.08 ± 0.01	0.08 ± 0.03	0.03 ± 0.02	0.06 ± 0.03
[0.9-1.0]	0.08 ± 0.04	0.23 ± 0.03	0.05 ± 0.04	0.03 ± 0.01	0.04 ± 0.02	0	0.02 ± 0.01

Table 5.20: Background yields with total uncertainties for each BDT bin in 2016 data.

BDT range	$B_{d,s}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$
[0-0.25]	0.44 ± 0.07	7.81 ± 0.47	0.50 ± 0.13	3.14 ± 0.46	0.98 ± 0.41	38.74 ± 8.48	0.25 ± 0.10
[0.25-0.4]	0.30 ± 0.21	4.18 ± 3.40	0.36 ± 0.55	1.47 ± 0.21	0.56 ± 0.24	5.97 ± 1.36	0.15 ± 0.06
[0.4-0.5]	0.20 ± 0.12	2.64 ± 2.47	0.29 ± 0.42	0.74 ± 0.11	0.36 ± 0.15	2.70 ± 0.65	0.11 ± 0.05
[0.5-0.6]	0.22 ± 0.13	2.68 ± 0.99	0.34 ± 0.25	0.64 ± 0.10	0.38 ± 0.16	1.69 ± 0.44	0.18 ± 0.07
[0.6-0.7]	0.22 ± 0.04	2.64 ± 1.06	0.40 ± 0.23	0.50 ± 0.07	0.38 ± 0.16	0.85 ± 0.25	0.19 ± 0.08
[0.7-0.8]	0.22 ± 0.10	2.41 ± 0.50	0.41 ± 0.31	0.40 ± 0.06	0.37 ± 0.16	0.75 ± 0.23	0.17 ± 0.07
[0.8-0.9]	0.24 ± 0.09	1.92 ± 0.28	0.35 ± 0.26	0.29 ± 0.05	0.31 ± 0.13	0.10 ± 0.06	0.12 ± 0.05
[0.9-1.0]	0.27 ± 0.05	0.91 ± 0.13	0.16 ± 0.14	0.13 ± 0.02	0.15 ± 0.07	0	0.05 ± 0.05

Table 5.21: Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/c², for Run 1 data.

BDT range	$B_{d,s}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$
[0-0.25]	0.43 ± 0.08	0.201 ± 0.013	0.030 ± 0.008	11.10 ± 2.43	0.142 ± 0.071
[0.25-0.4]	0.37 ± 0.26	0.091 ± 0.007	0.020 ± 0.005	1.37 ± 0.35	0.090 ± 0.045
[0.4-0.5]	0.27 ± 0.15	0.068 ± 0.006	0.016 ± 0.004	0.36 ± 0.12	0.062 ± 0.031
[0.5-0.6]	0.34 ± 0.08	0.076 ± 0.006	0.023 ± 0.006	0.41 ± 0.13	0.082 ± 0.040
[0.6-0.7]	0.33 ± 0.09	0.078 ± 0.007	0.024 ± 0.006	0.10 ± 0.05	0.085 ± 0.041
[0.7-0.8]	0.37 ± 0.10	0.077 ± 0.007	0.029 ± 0.008	0.13 ± 0.06	0.079 ± 0.038
[0.8-0.9]	0.39 ± 0.08	0.062 ± 0.006	0.033 ± 0.009	0	0.079 ± 0.038
[0.9-1.0]	0.46 ± 0.08	0.051 ± 0.005	0.022 ± 0.006	0	0.035 ± 0.018

Table 5.22: Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/ c^2 , for 2015 data.

BDT range	$B_{d,s}^0 \rightarrow h^+h'^-$	$B_d^0 \rightarrow \pi^-\mu^+\nu_\mu$	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	$B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$	$\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$
[0-0.25]	0.076 \pm 0.013	0.027 \pm 0.002	0.005 \pm 0.001	1.25 \pm 0.29	0.046 \pm 0.017
[0.25-0.4]	0.050 \pm 0.050	0.014 \pm 0.001	0.004 \pm 0.001	0.17 \pm 0.05	0.035 \pm 0.013
[0.4-0.5]	0.040 \pm 0.023	0.008 \pm 0.001	0.003 \pm 0.001	0.065 \pm 0.027	0.031 \pm 0.011
[0.5-0.6]	0.037 \pm 0.045	0.010 \pm 0.001	0.004 \pm 0.001	0.051 \pm 0.024	0.037 \pm 0.014
[0.6-0.7]	0.043 \pm 0.028	0.010 \pm 0.001	0.004 \pm 0.001	0.034 \pm 0.019	0.043 \pm 0.016
[0.7-0.8]	0.045 \pm 0.029	0.010 \pm 0.001	0.005 \pm 0.001	0.025 \pm 0.016	0.045 \pm 0.017
[0.8-0.9]	0.051 \pm 0.028	0.009 \pm 0.001	0.005 \pm 0.001	0	0.045 \pm 0.017
[0.9-1.0]	0.057 \pm 0.028	0.008 \pm 0.001	0.004 \pm 0.001	0	0.025 \pm 0.009

Table 5.23: Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/c², for 2016 data.

BDT range	$B_{d,s}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow \mu^+ \nu_\mu$	$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$
[0-0.25]	0.30 ± 0.05	0.112 ± 0.008	0.016 ± 0.004	4.80 ± 1.10	0.030 ± 0.012
[0.25-0.4]	0.20 ± 0.14	0.057 ± 0.005	0.011 ± 0.003	0.64 ± 0.20	0.020 ± 0.008
[0.4-0.5]	0.14 ± 0.08	0.033 ± 0.003	0.010 ± 0.003	0.25 ± 0.10	0.015 ± 0.006
[0.5-0.6]	0.15 ± 0.09	0.042 ± 0.004	0.011 ± 0.003	0.20 ± 0.09	0.023 ± 0.009
[0.6-0.7]	0.15 ± 0.03	0.039 ± 0.004	0.013 ± 0.004	0.13 ± 0.07	0.026 ± 0.010
[0.7-0.8]	0.15 ± 0.07	0.040 ± 0.004	0.015 ± 0.004	0.10 ± 0.06	0.027 ± 0.011
[0.8-0.9]	0.16 ± 0.06	0.036 ± 0.003	0.015 ± 0.004	0	0.025 ± 0.010
[0.9-1.0]	0.18 ± 0.03	0.030 ± 0.003	0.013 ± 0.003	0	0.013 ± 0.006

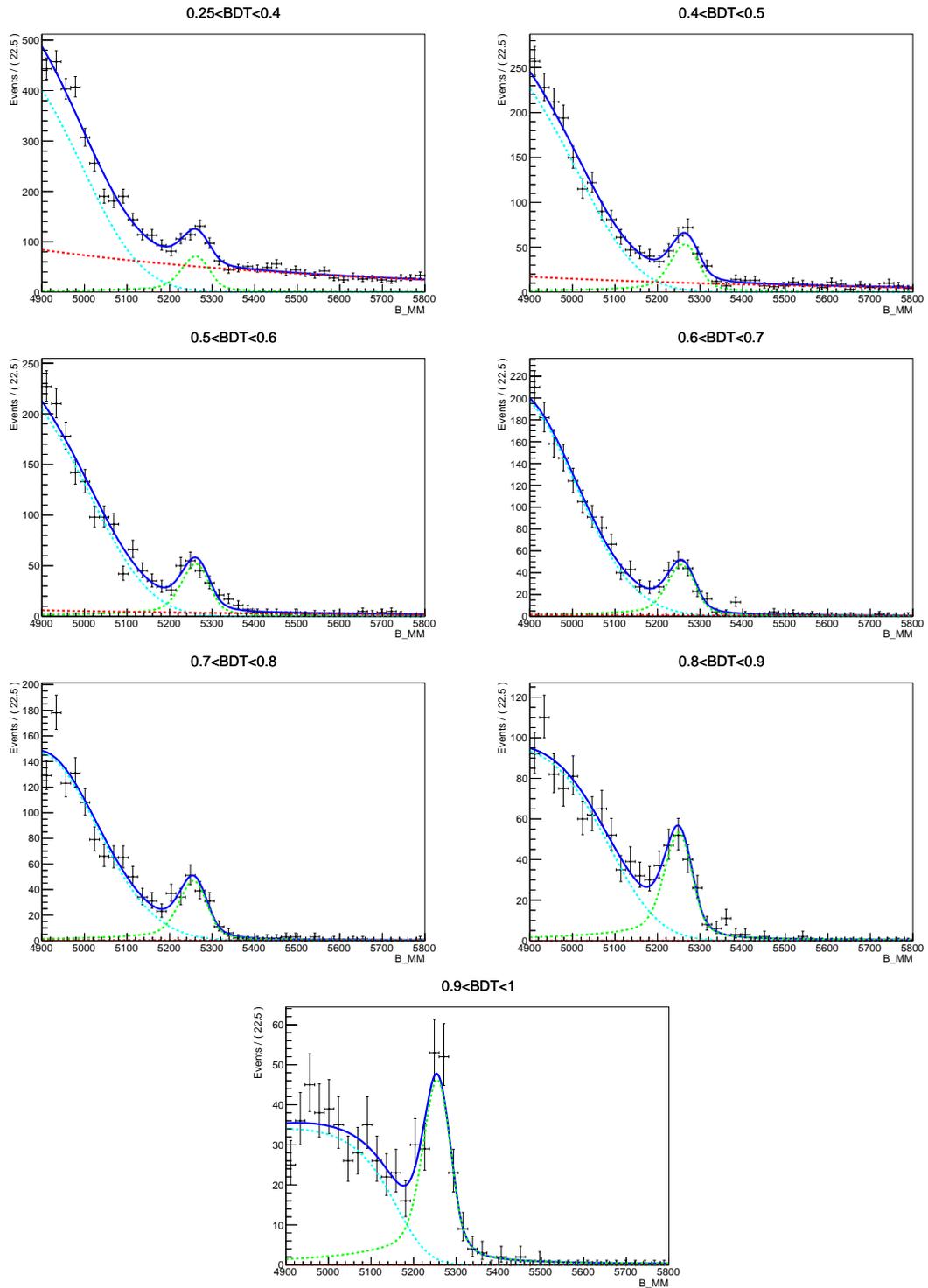


Figure 5.13: Invariant mass fit to Run 1 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

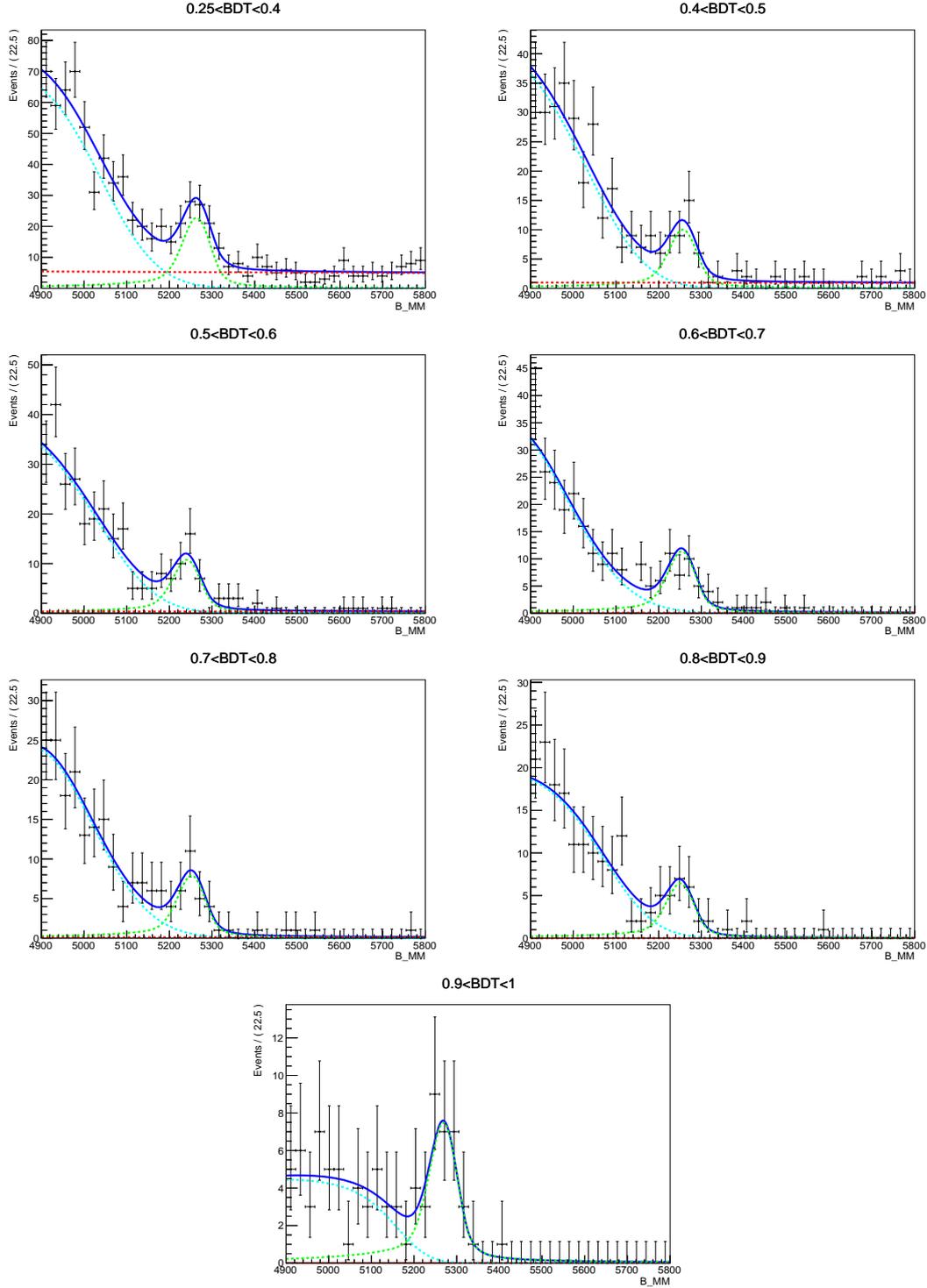


Figure 5.14: Invariant mass fit to 2015 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

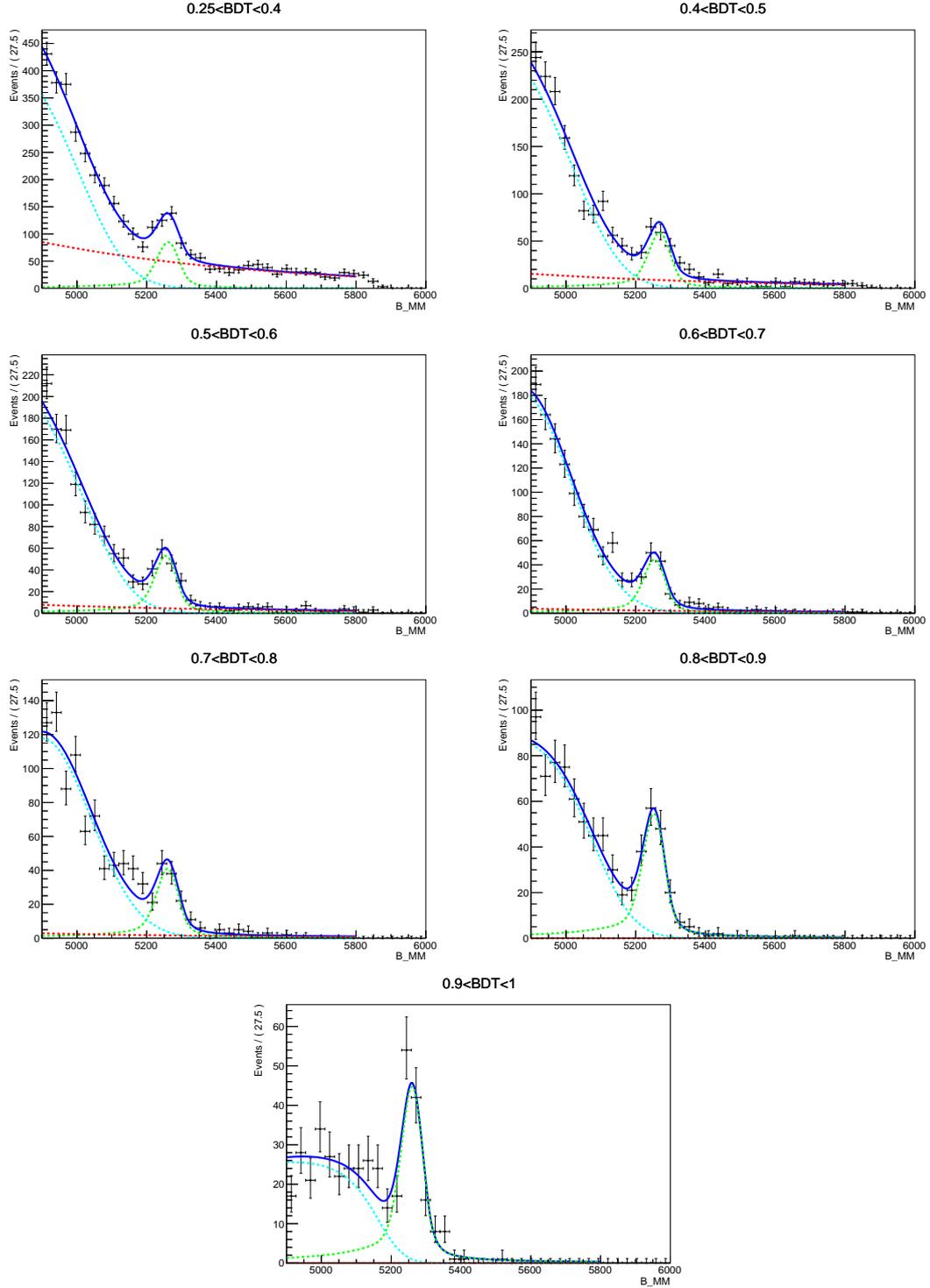


Figure 5.15: Invariant mass fit to 2016 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

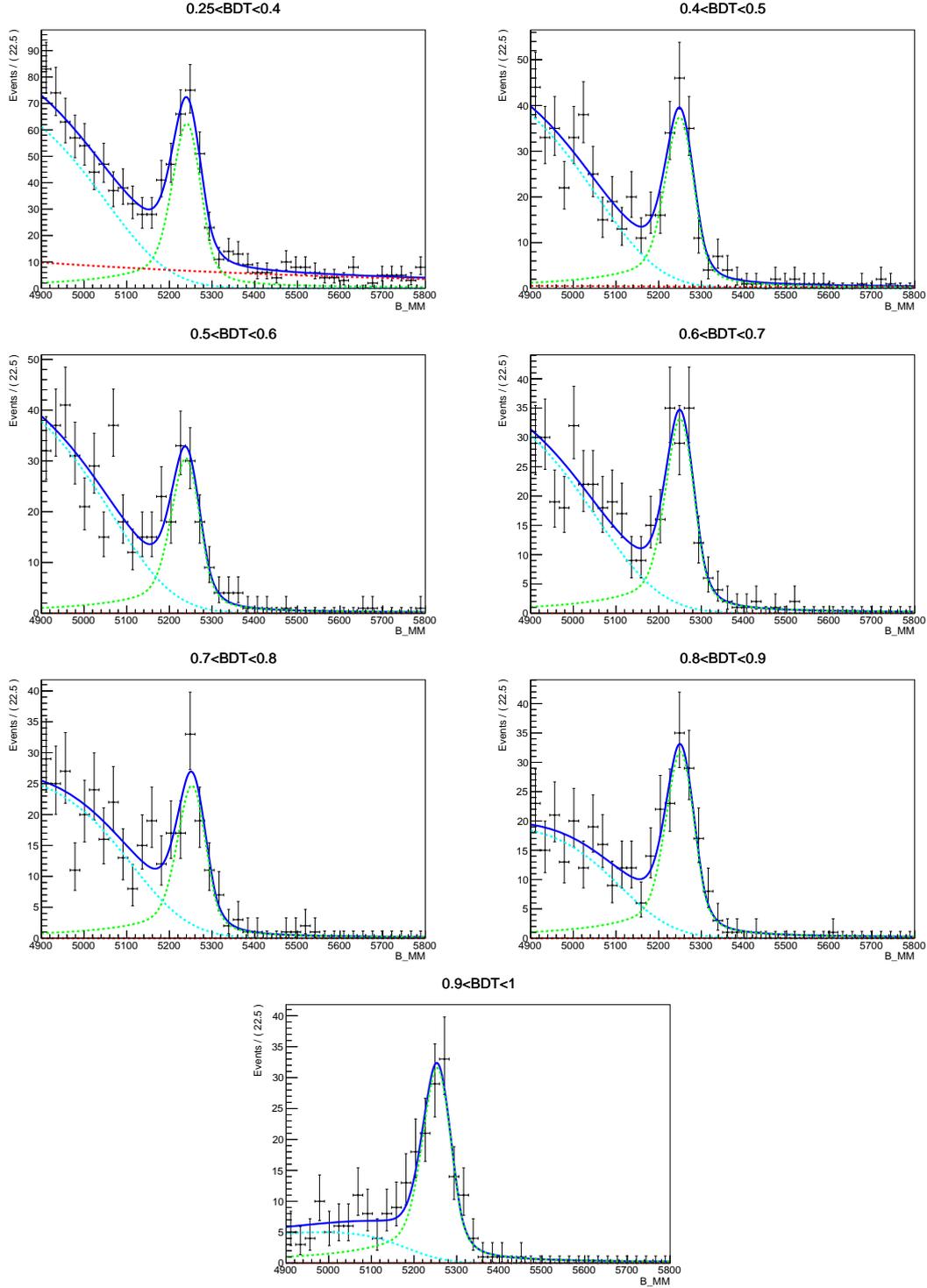


Figure 5.16: Invariant mass fit to Run 1 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

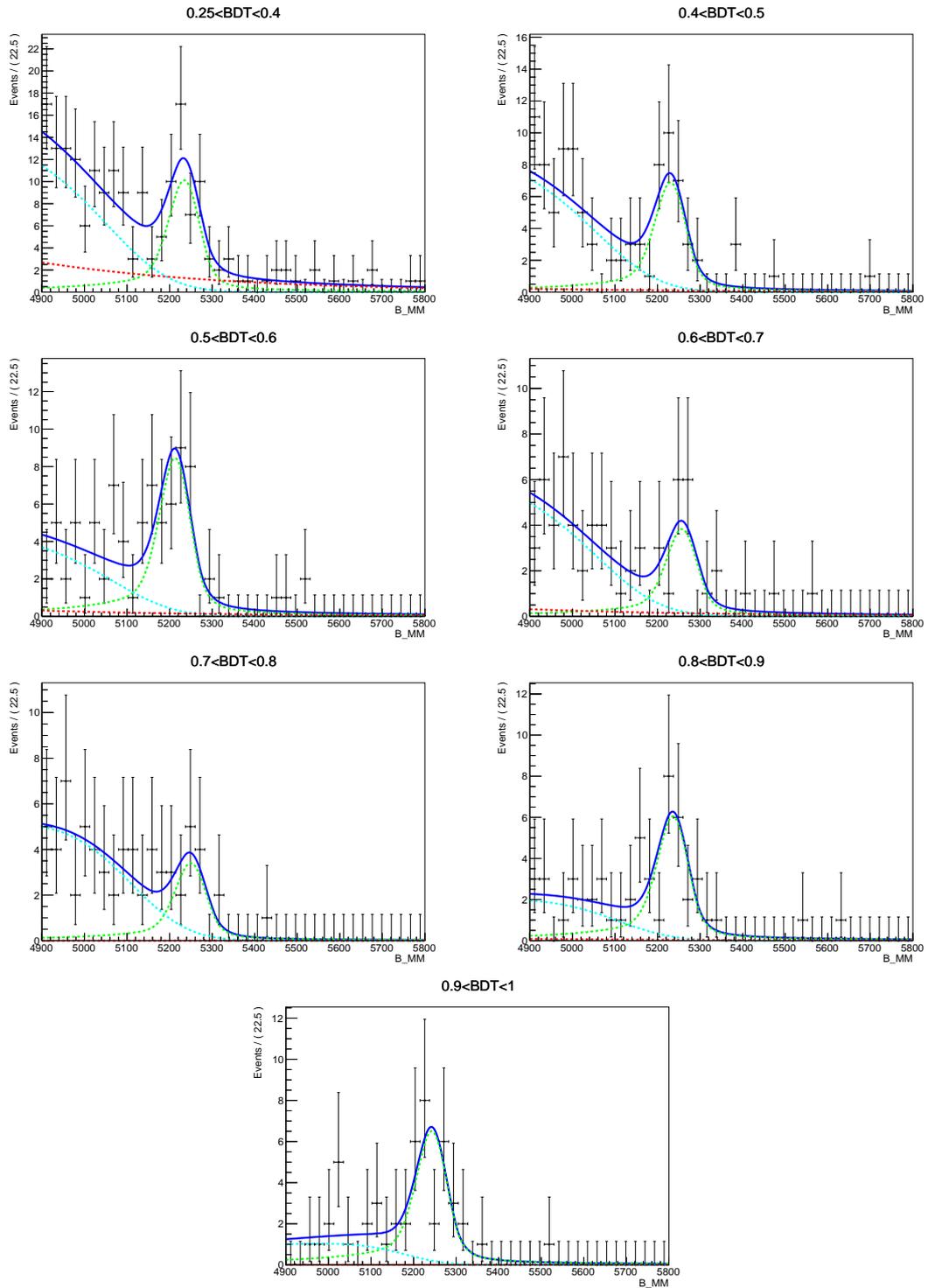


Figure 5.17: Invariant mass fit to 2015 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

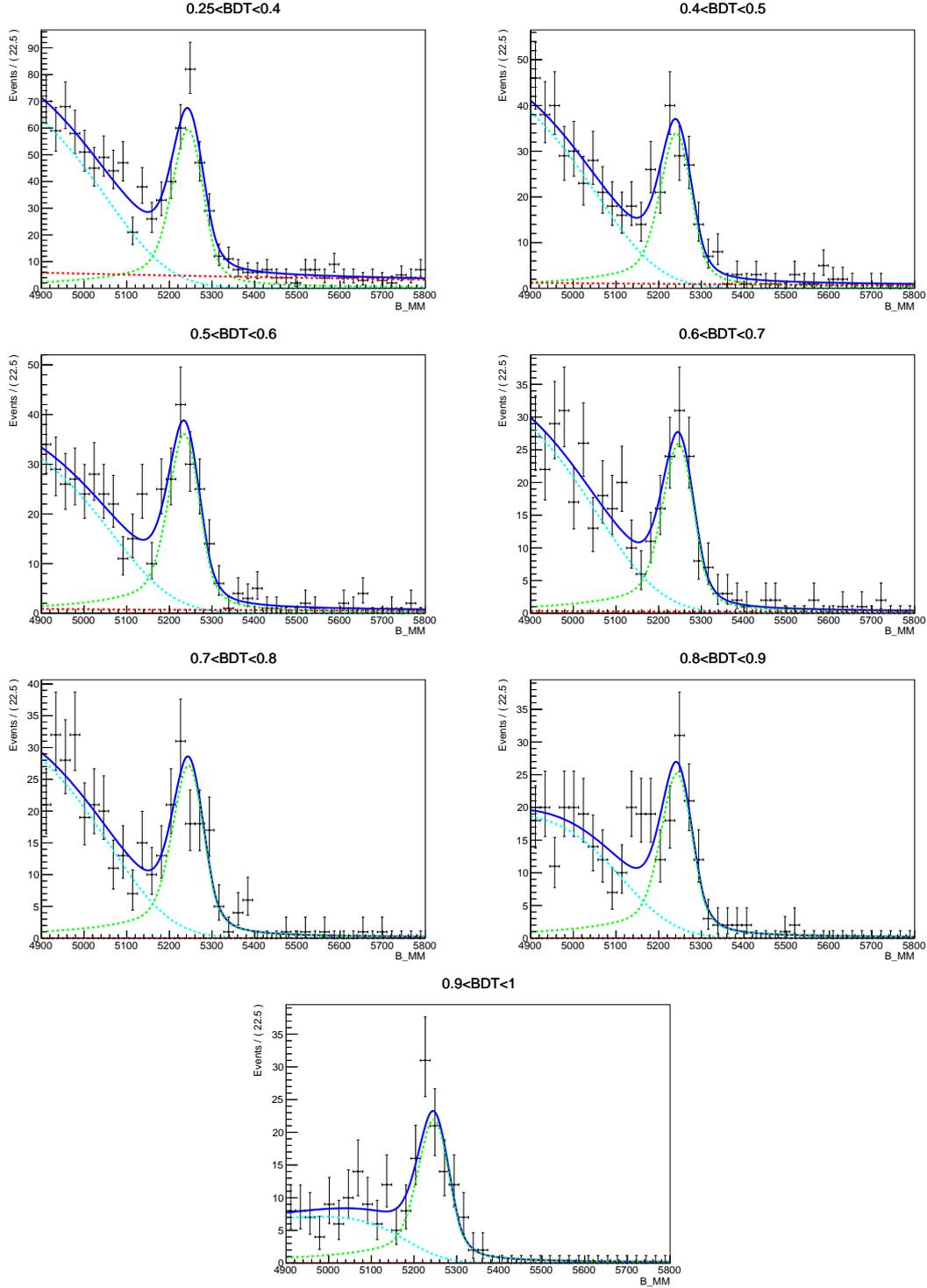


Figure 5.18: Invariant mass fit to 2016 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 .

Chapter 6

Results

In this final chapter, the results of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ analysis are given, together with a discussion on the uncertainties. Sec. 6.1 is devoted to the description of the signal fit and the consequent extraction of the branching fractions, while the procedure to set the limit on the $B_d^0 \rightarrow \mu^+\mu^-$ branching fraction is described in Sec. 6.2.2. A summary of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime analysis and its result is given in Sec. 6.3. The implications of the measurements and the future prospects are finally discussed in Sec. 6.4.

6.1 The signal fit

The branching fractions of both $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ decays are measured via an unbinned extended maximum likelihood fit to the invariant mass distribution of the dimuon candidates, selected with the cuts described in Sec. 4.1 and categorised into 8 BDT bins: 4 for Run 1 and 4 for Run 2 data, which are simultaneously fitted. The first BDT bin, $[0, 0.25]$, is excluded from both datasets, since it gives no contribution to the total sensitivity, while it significantly slows down the fit due to the large number of combinatorial events added. In addition, the last 4 BDT bins of the original scheme (Eq. (4.2)) were merged in order to have enough combinatorial events that allow for a reliable background estimate. This is shown in Fig. 6.1 and Fig. 6.2, where the dimuon sidebands for Run 1 and 2016 data are displayed, with the background components of the fit superimposed. Thanks to the excellent performances of the BDT, the last two bins of the data samples have no combinatorial events on the right mass sidebands, which clearly represents a problem in the fit. For this reason, and following the conclusion of toy MC studies discussed in Sec. 6.1.2, the BDT bins which define the fit regions have been rearranged for both Run 1 and Run 2 data samples as:

$$[0.25, 0.4] [0.4, 0.5] [0.5, 0.6] [0.6, 1.0]. \quad (6.1)$$

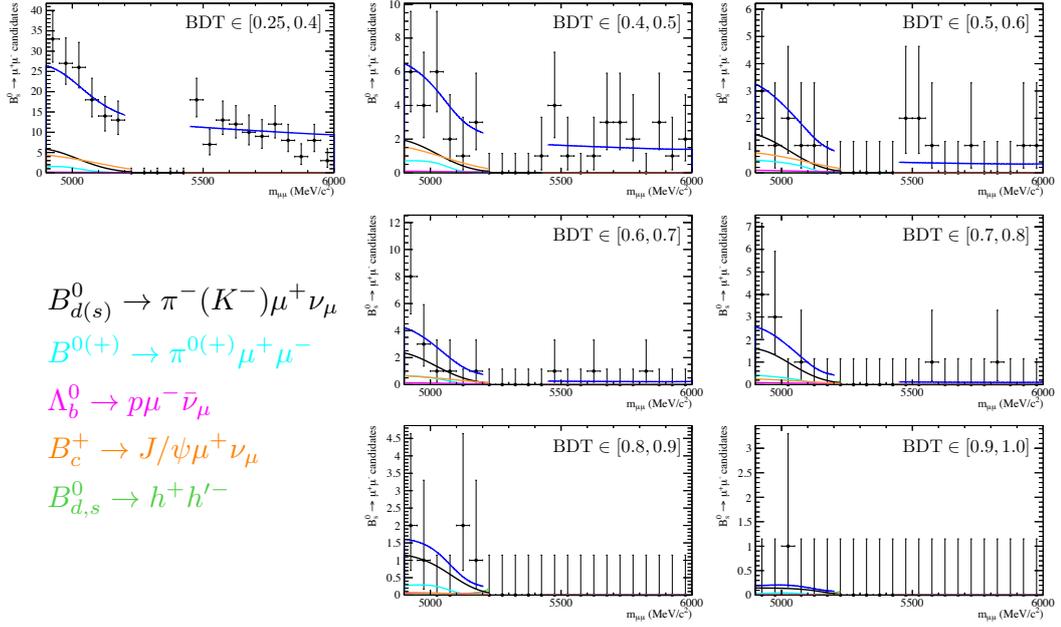


Figure 6.1: Invariant mass fits of the dimuon sidebands for the 8 BDT bins of Run 1 data. The fit components are indicated with their respective colours in the legend.

6.1.1 Fit model

The parameters of interest in the fit are the branching fractions of the $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays, which are determined from the number of signal candidates via the normalisation factors, as from Eq. (4.7). The slope and yield of the combinatorial background in each BDT region are also free parameters. In addition, the following nuisance parameters are needed to build the fit model:

- the mass shape parameters of the $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ peaks (Sec. 4.3);
- the BDT calibration parameters, i.e. the signal fractions in each BDT bin (Sec. 4.2.4);
- the mass shape parameters and yields of the $B_{d,s}^0 \rightarrow h^+h'^-$ (Sec. 5.1) and semileptonic (Sec. 5.2) backgrounds;
- the normalisation factors α , used to convert the signal yields into branching fractions (Sec. 4.4);
- the time-dependent corrections (Sec. 4.2.5).

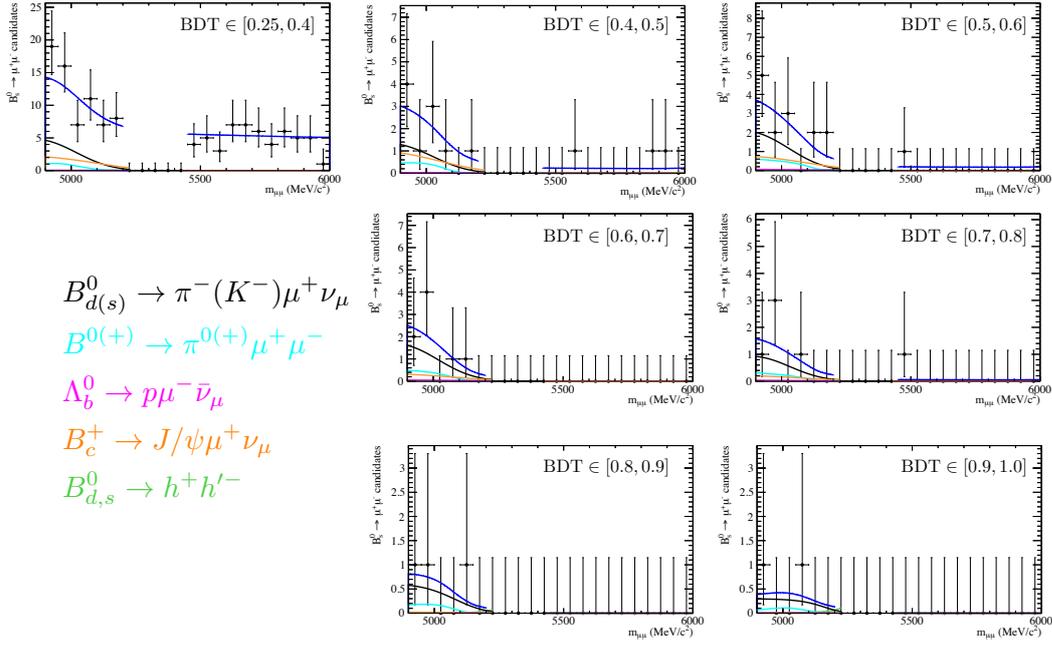


Figure 6.2: Invariant mass fits of the dimuon sidebands for the 8 BDT bins of 2016 data. The fit components are indicated with their respective colours in the legend.

All the above parameters are constrained to their expected values according to their gaussian uncertainties. In addition, common terms which are present in the definition of the various nuisance parameters, such as the normalisation channel branching fraction, have been treated as correlated.

6.1.2 Toy Monte Carlo studies

In order to assess the expected sensitivity as well as the optimal BDT binning configuration for the signal fit, toy MC studies based on the Run 1 dataset were carried out. A toy MC consists of generating several pseudoexperiments in which all background components are fluctuated around their expectations, while the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ signals are generated according to their SM predictions. The resulting toy data are then fitted according to the model of Sec. 6.1.1. The following BDT binning configurations were investigated:

1. **7-bins:** the starting configuration of Eq. (4.2), with the exclusion of the bin $[0.0, 25]$;
2. **5-bins:** last three bins ($\text{BDT} > 0.7$) merged and excluding the bin $[0.0, 25]$: with this configuration 2 events are present in the right sideband of the last bin;

3. **4-bins**: last four bins ($\text{BDT} > 0.6$) merged and excluding the bin $[0.0, 25]$: with this configuration 5 events are present in the right sideband of the last bin.

A small sensitivity loss is expected by merging the last bins, where the combinatorial background is negligible, while an improvement of the fit stability is foreseen. Since the right sideband is populated by combinatorial background only, its yield is assumed to be always positive definite. For each configuration, ~ 1000 toys are generated. The pulls for the $B_s^0 \rightarrow \mu^+\mu^-$ BF are shown in Fig. 6.3 for the three binning configurations, and show no significant biases. For each toy, the statistical

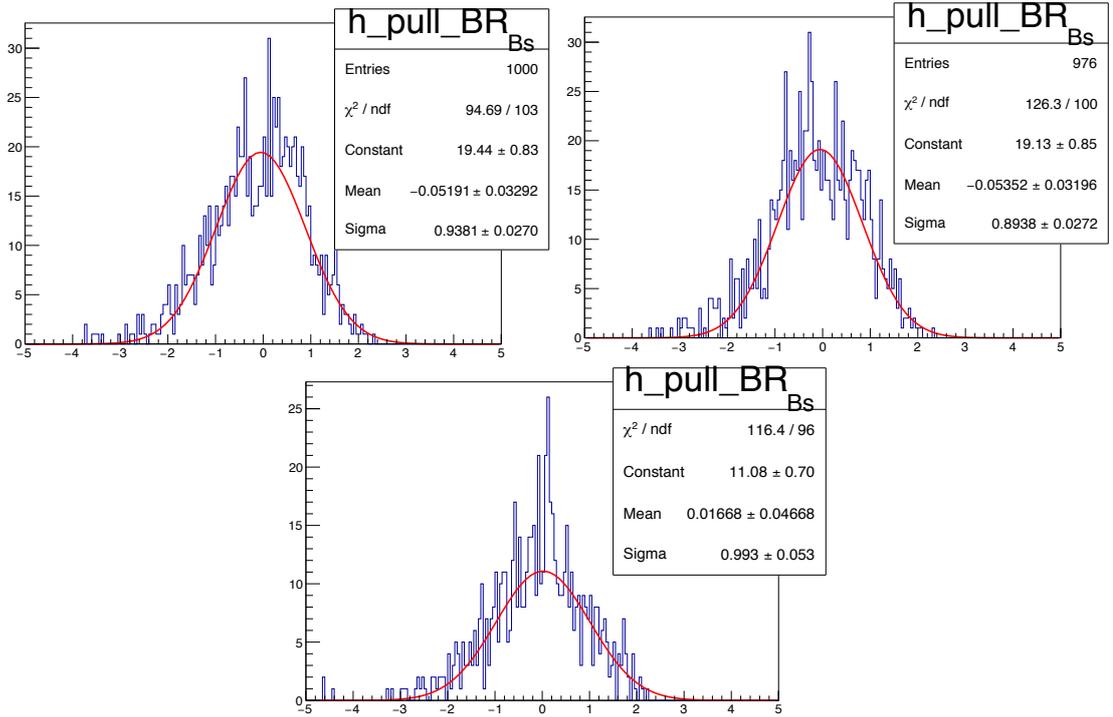


Figure 6.3: Pulls of the $B_s^0 \rightarrow \mu^+\mu^-$ BF for Run 1 toys in the 7-bins (left), 5-bins (right) and 4-bins (bottom) configurations.

significance of the $B_s^0 \rightarrow \mu^+\mu^-$ signal is also evaluated using Wilks theorem, as described in Sec. 6.1.3. In Table 6.1, the expected significance and the probability of having more than 5 standard deviations are listed. The expected significance is evaluated as the median value of the statistical significance distribution obtained with each binning configuration. In the 4-bin configuration, the merging of the last BDT bins seems to produce a positive effect on the overall fit stability with only a small sensitivity decrease. This is confirmed by the pulls of the combinatorial background for the last bins of the 5- and 4-bin configurations, shown in Fig. 6.4.

BDT binning configuration	Expected significance	$> 5\sigma$ probability
7-bins	6.5	89.2%
5-bins	6.3	88.3%
4-bins	6.1	83.8%

Table 6.1: Expected significance and observation probability of the $B_s^0 \rightarrow \mu^+\mu^-$ decay mode for Run 1 toys.

In the 5-bin configuration, the pull distribution in the most significant bin is much

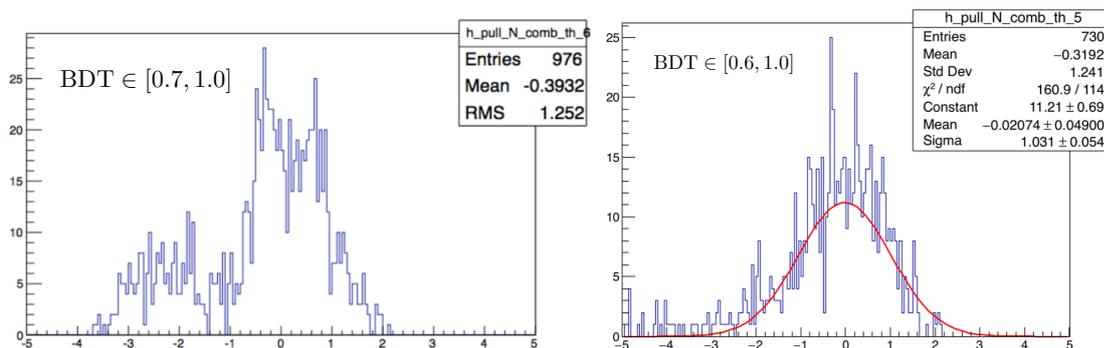


Figure 6.4: Pulls of the combinatorial in the most significant BDT bin for Run 1 toys in the 5-bin (left) and 4-bin (right) configurations.

more distorted with respect to the 4-bin case.

Finally, toys have been run on the full dataset, using the 4-bin configuration for all samples. The expected sensitivities are shown in Fig. 6.5, for both $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$. While it is very unlikely not to have a 5σ observation of the

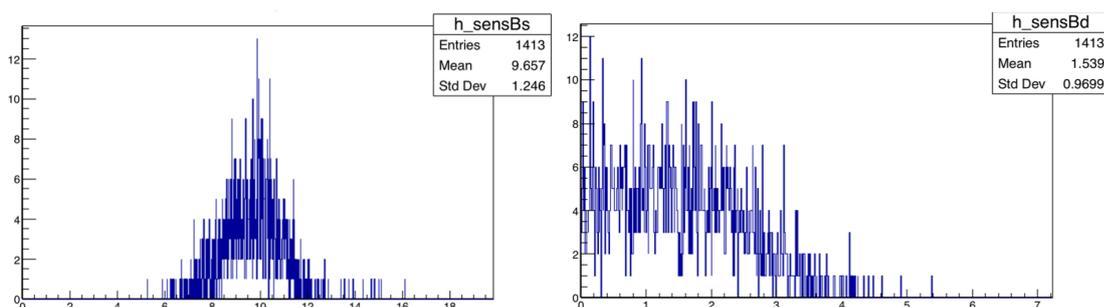


Figure 6.5: Expected sensitivities on $B_s^0 \rightarrow \mu^+\mu^-$ (left) and $B_d^0 \rightarrow \mu^+\mu^-$ (right) decays from the toys on the full dataset with the 4-bin configuration.

$B_s^0 \rightarrow \mu^+\mu^-$ decay assuming the SM, the toys indicate a $\sim 8\%$ probability of

having a 3σ evidence for the $B_d^0 \rightarrow \mu^+\mu^-$ decay assuming SM.

6.1.3 Statistical significance of a peak

While the principle of maximum likelihood provides a method to estimate parameters [177], it does not directly suggest a method of testing goodness-of-fit. A possible solution is to pick the value of the likelihood at its maximum, \mathcal{L}_{max} , as a goodness-of-fit statistic. However, the \mathcal{L}_{max} distribution is a priori unknown. If the data are modelled with a likelihood \mathcal{L} that depends on a set of N parameters $\mu = (\mu_1, \dots, \mu_N)$, the *likelihood ratio* can be used as a test statistic:

$$t_\mu = -2 \ln \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}, \quad (6.2)$$

where the hat denotes the maximum likelihood (ML) estimator. When the fit agreement to data is poor, $\hat{\mu}$ will be far from μ , and t_μ will be large. Larger values of t_μ thus indicate increasing incompatibility between the data and the hypothesized μ .

Wilks' theorem [178] states that if the parameter values μ are true, then, in the asymptotic limit of a large data sample, t_μ follows a χ^2 distribution with N degrees of freedom:

$$f(t_\mu|\mu) \sim \chi_N^2. \quad (6.3)$$

Hence, if the observed value of the statistic yields $t_{\mu,obs}$, the level of compatibility between the parameters and the observed data can be quantified by computing the *p-value* [179]:

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f_{\chi_N^2}(t_\mu|\mu) dt_\mu. \quad (6.4)$$

In the practical case of the $B_s^0 \rightarrow \mu^+\mu^-$ signal, the χ^2 difference observed between the full fit and the fit where the signal is forced to zero, i.e. $-2 \ln \mathcal{L}(\text{sig} = 0)/\mathcal{L}(\hat{\mu})$ in Eq. 6.2, is computed to test the background only hypothesis, thus yielding the signal significance.

6.2 Branching fraction results

From the fit to the unblinded data the following results are obtained:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.0 \pm 0.6_{-0.2}^{+0.3}) \times 10^{-9}, \quad (6.5)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) = (1.5_{-1.0}^{+1.2} \pm 0.2_{-0.1}^{+0.2}) \times 10^{-10}, \quad (6.6)$$

where the first uncertainty is statistical and the second one is systematic. The nominal fit assumes $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$, as predicted by the SM, for the evaluation of

the signal acceptance. In order to investigate the dependence of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction on the $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ assumption, the fit is repeated under the hypotheses $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$ using the corrections discussed in Sec. 4.2.5. The central value of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction increases by 4.6% and 10.9% for $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$, respectively. The statistical significances of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ signals are 7.8 and 1.6 standard deviations, respectively, computed according to Sec. 6.1.3.

Fig. 6.6 shows the fit result in the most sensitive region, $\text{BDT} > 0.5$, while the results for all BDT bins are shown in Fig. 6.7.

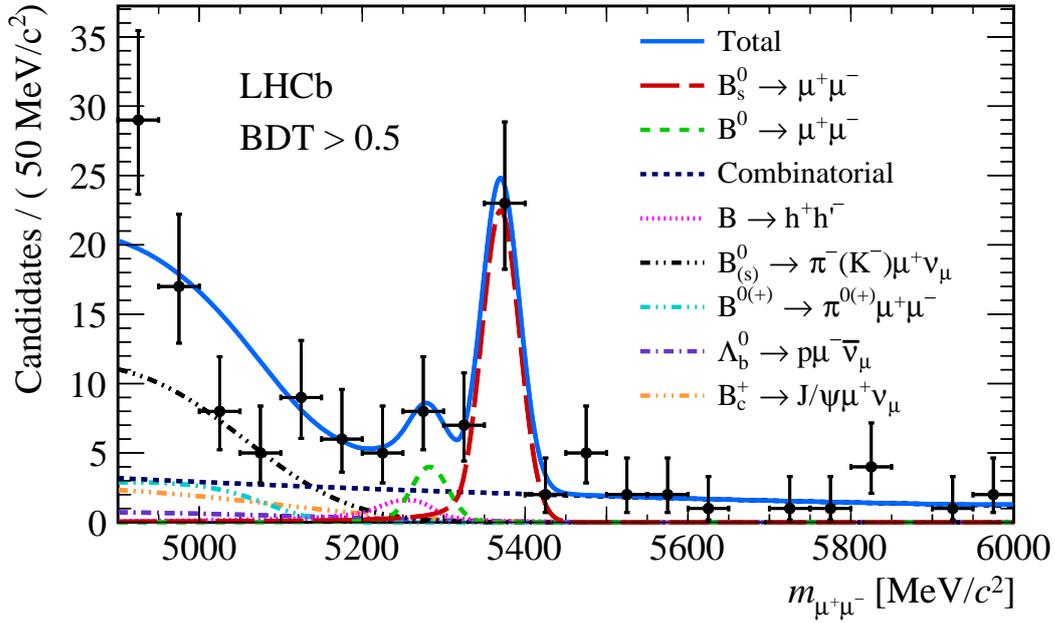


Figure 6.6: Mass distribution of the selected $B_{d,s}^0 \rightarrow \mu^+\mu^-$ candidates for $\text{BDT} > 0.5$. The fit curve is overlaid, together with its components.

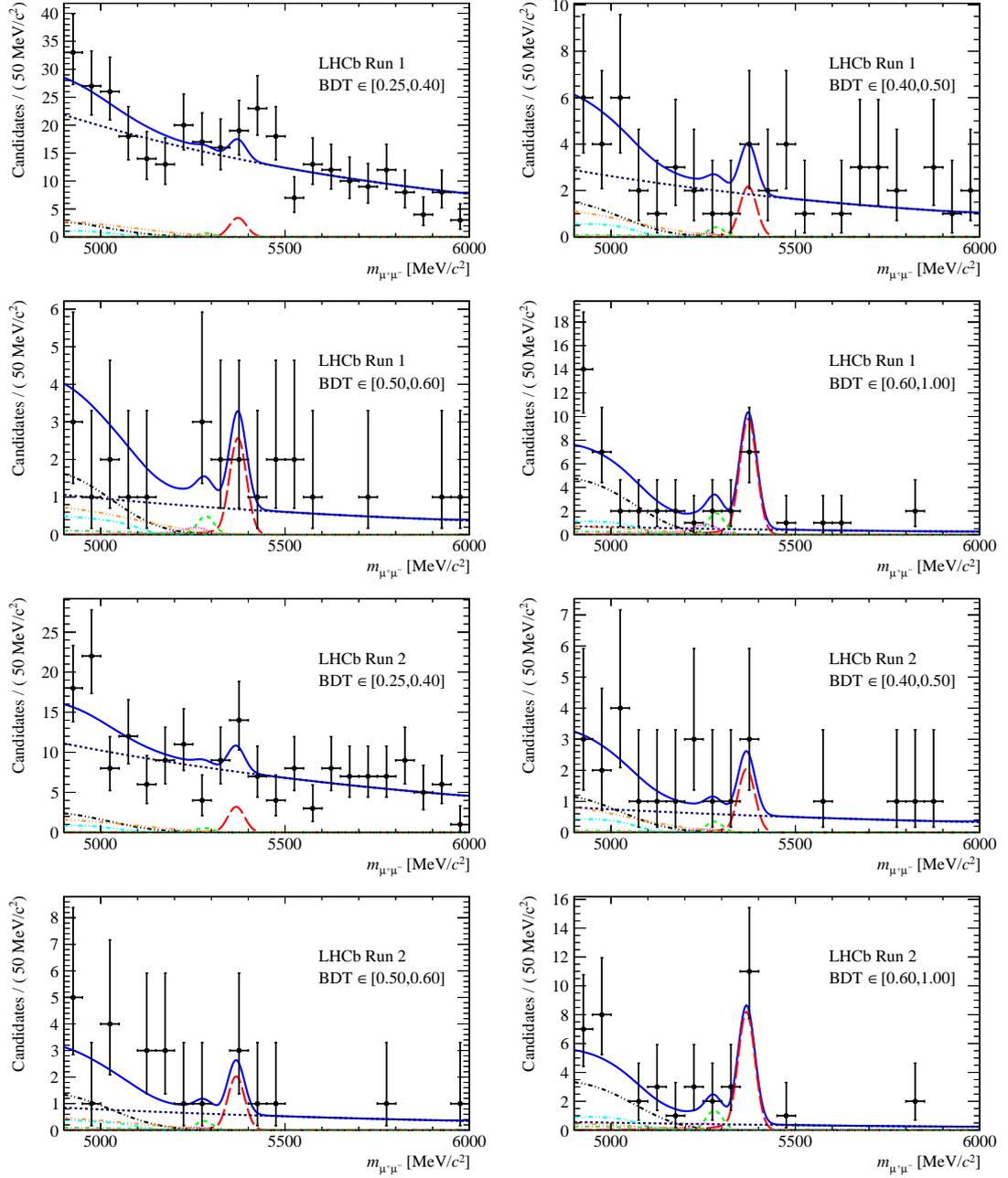


Figure 6.7: Mass distribution of the selected $B_{d,s}^0 \rightarrow \mu^+\mu^-$ candidates in each BDT bin used in the fit. The fit curve is overlaid together with its components, following the same legend of Fig. 6.6.

The prominent peak marks the first single-experiment observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay, and the most precise measurement of its branching fraction to date. The results are in good agreement with the previous LHCb+CMS combination [4] and the latest Standard Model predictions [53]. The statistical uncertainty is almost halved with respect to the previous LHCb measurement [61], as a consequence of the increased integrated luminosity, B meson production in Run 2, and background rejection.

Systematic uncertainty

The systematic uncertainty is obtained by subtracting in quadrature the statistical uncertainty to the total one, the former being computed with MINOS [180] after fixing all the nuisance parameters in the fit. As shown in the results (6.5), it amounts to 13.6% and 41.7% of the total branching fraction error for $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$, respectively. For the $B_s^0 \rightarrow \mu^+\mu^-$ measurement, the systematic error is dominated by the knowledge of the hadronisation fraction parameter f_s/f_d , while the $B_d^0 \rightarrow \mu^+\mu^-$ systematic error is mainly due to the exclusive backgrounds. A subdominant systematic uncertainty indicates large room for precision improvement as more statistics is gathered, hence the previous considerations point out the quantities that next analyses should primarily address.

6.2.1 2D likelihood scan

The information inequality [177] sets a lower bound to the variance of any estimator for the parameter θ :

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 / E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right], \quad (6.7)$$

where b is the bias and the operator E denotes the expectation value. Since the ML method ensures minimum variance in the large sample limit, the information inequality simply reads

$$\hat{\sigma}_{\hat{\theta}}^2 = \left(-1 / \frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right) \Big|_{\theta=\hat{\theta}}, \quad (6.8)$$

under the further assumption that the estimator is unbiased, as legitimated by the toy studies of Sec. 6.1.2 for the branching fractions.

The RooFit framework employed for the signal fit uses the MINUIT [181] program to find the maximum¹ value of the likelihood on the data. MINUIT calls the MIGRAD

¹In practice, $-\log \mathcal{L}$ is minimised so that large numerical differences are mitigated and products are converted into sums.

and HESSE routines [180] to determine numerically the matrix of second derivatives of $\log \mathcal{L}$ at the maximum likelihood estimates, so that the standard deviation of the estimator, which quantifies its statistical uncertainty, is given by the square root of the (6.8). A gaussian behaviour of the likelihood *pdf*, corresponding to a parabolic shape of $-\log \mathcal{L}$ around its minimum, is therefore assumed. In the case of multiple parameters, HESSE hence takes into account their correlations, but neglects the non-linearities, i.e. when the likelihood *pdf* is not gaussian². To this end, the MINOS routine [180] is run after the minimum is found to compute the proper statistical error: it searches for the change in the parameter that causes the log-likelihood value to decrease by 1/2. It is in fact true [177] that even if the likelihood is not a gaussian, the confidence interval can be approximated by

$$\log \mathcal{L}(\hat{\theta}_{-\Delta\theta^-}^{+\Delta\theta^+}) = \log \mathcal{L}_{max} - \frac{1}{2}. \quad (6.9)$$

Note that the resulting asymmetric interval, $[\hat{\theta} - \Delta\theta^-, \hat{\theta} + \Delta\theta^+]$, has the meaning of a coverage probability only in the limit of large sample.

In the signal fit, there are two parameters of interest: the two branching fractions of the $B_d^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ channels. In order to obtain a meaningful statistical coverage on their simultaneous measurement, a 2D likelihood contour has also been computed in Fig. 6.8. The plot displays the contours of constant likelihood ratio, Eq. 6.2, in the estimated parameter space, i.e. the $\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-) - \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ plane. The SM expectation value lies within the first contour, delimiting the 68% confidence region. Note that in this two-dimensional case the confidence region is given by

$$\log \mathcal{L}(\hat{\theta}) = \log \mathcal{L}_{max} - \frac{Q_\gamma}{2}, \quad (6.10)$$

where Q_γ is the quantile of order $1-\gamma$ of the χ^2 distribution³, as follows from Wilks' prescription (6.3). To compute the contours, an extensive profile likelihood scan has been performed by analysing many possible values of the branching fractions around the minimum, and regaining the maximum likelihood from a fit where the two BFs are kept fixed.

6.2.2 $B_d^0 \rightarrow \mu^+\mu^-$ limit computation with \mathbf{CL}_s

Since the statistical significance of the observed $B_d^0 \rightarrow \mu^+\mu^-$ peak is below the conventional threshold of 3 standard deviations needed to claim for an evidence,

²Asymmetries around the minimum of $-\log \mathcal{L}$ arise when a large number of parameters is estimated within a limited statistics.

³In the single-parameter case (Eq. (6.9)), $Q_\gamma = 1$ corresponds to $1 - \gamma = 0.683$, whereas for two parameters the confidence level decreases to 0.393. To obtain the 68.3% confidence level one needs $Q_\gamma = 2.30$.

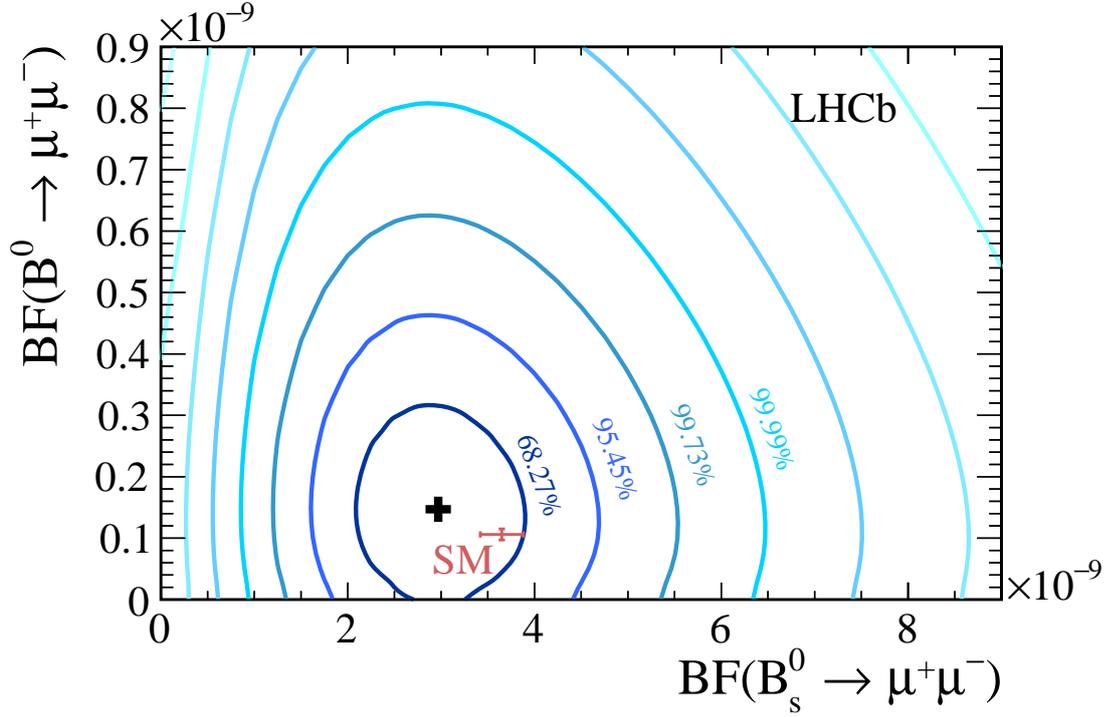


Figure 6.8: Contours of constant likelihood ratio in the space of the two parameters of interest. The values of the confidence regions are indicated at each contour, according to the formula (6.10). The SM expectation is superimposed in red.

an upper limit is set to its branching fraction according to the CL_s prescription, as described below.

The p-value (6.4) can be used to quantify the confidence in the signal plus background hypothesis according to

$$CL_{s+b} = P_{s+b}(t_\mu \leq t_{\mu,obs}) = \int_{-\infty}^{t_{\mu,obs}} f_{\chi_N^2}(t_\mu|\mu) dt_\mu = 1 - p_\mu, \quad (6.11)$$

i.e. the probability that the test statistic (6.2) is less than or equal to the value observed in data. Small values of CL_{s+b} indicate poor compatibility with the signal plus background hypothesis, thus favoring the background-only hypothesis. Analogously, CL_b quantifies the confidence in the background-only hypothesis, and is computed by performing the fit with the signal forced to zero.

When an experimental result appears consistent with little or zero signal, and is accompanied by a downward fluctuation of the background, the exclusion limit may be so strong to even exclude zero signal with high confidence levels. Although a perfectly valid result from a statistical point of view, it tends to say more about

the probability of observing a similar or stronger exclusion in future experiments with the same expected signal and background than about the non-existence of the signal itself, and it is the latter which is of more interest to the physicist [182]. The CL_s method [182, 183] is a modified frequentist approach that proposes to normalise the confidence level observed for the signal plus background hypothesis to the one observed for the background-only hypothesis:

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b}. \quad (6.12)$$

Being a ratio of confidence levels, CL_s is not a confidence level itself, but the signal hypothesis can be considered excluded at the level CL when

$$1 - \text{CL}_s \leq \text{CL}. \quad (6.13)$$

The normalisation in (6.12) provides a conservative limit and avoids the undesirable property of CL_{s+b} that of two experiments with the same (small) expected signal rate but different backgrounds, the experiment with the larger background might set a more stringent limit.

The CL_s method is especially suited for the $B_d^0 \rightarrow \mu^+ \mu^-$ analysis scenario, where upward statistical fluctuations of the background levels could likely cover the signal. By choosing a threshold of 95% confidence level, the following limit is obtained:

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}, \quad (6.14)$$

as shown in Fig. 6.9.

6.3 Effective lifetime measurement

In this section, the analysis procedure and results of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime are presented. The strategy adopted for the analysis is informed to a large extent by the expected precision of the measurement. In fact, as explained in Sec. 1.2.2, the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is predicted by the Standard Model to be equal to the one of the heavy mass eigenstate of the B_s meson: $\tau_{\mu^+ \mu^-} = \tau_H = (1.610 \pm 0.012)$ ps [184]. New physics effects may move $\tau_{\mu^+ \mu^-}$ towards the lifetime of the light B_s mass eigenstate, which amounts to $\tau_{\mu^+ \mu^-} = \tau_L = (1.422 \pm 0.008)$ ps [184]. The difference between these two extreme cases, 0.188 ps, corresponds to the change in $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ (Eq. (1.30)) from +1 ($\tau_{\mu^+ \mu^-} = \tau_H$) to -1 ($\tau_{\mu^+ \mu^-} = \tau_L$). Therefore, a precision of the order of 0.038 ps is required on the effective lifetime to discriminate between $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = +1$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = -1$ at five standard deviations. On the other hand, preliminary studies on the sensitivity, conducted on Run 1 data, have indicated a likely precision of around 0.4 ps, i.e. about ten times worse

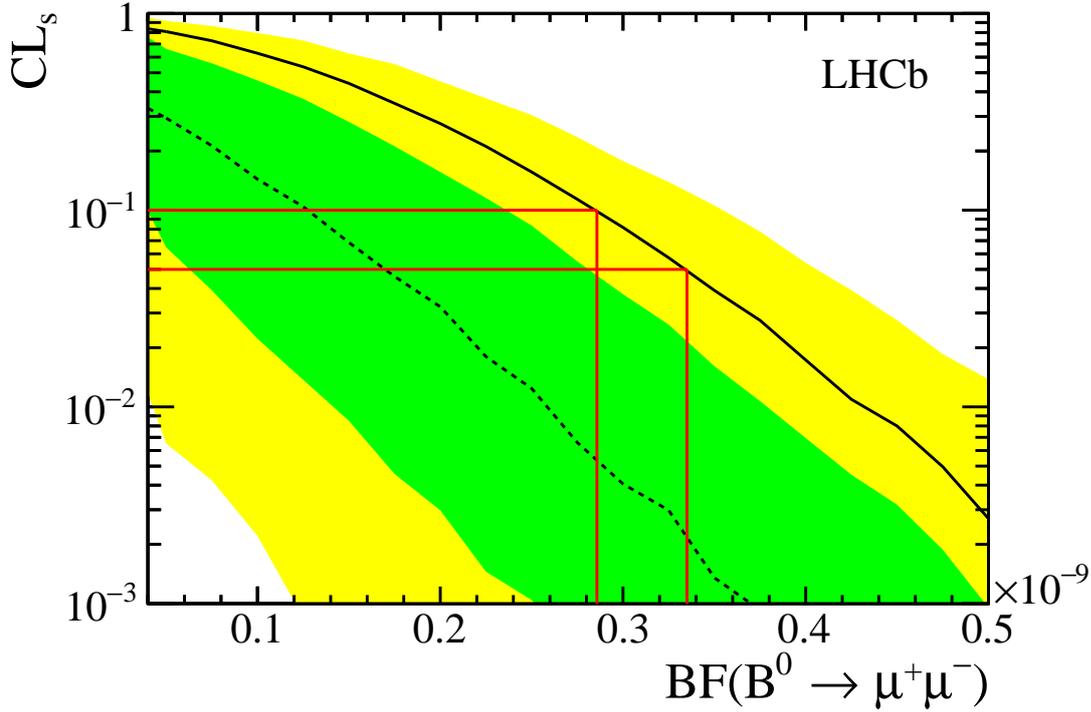


Figure 6.9: The CL_s confidence belt for the $B_d^0 \rightarrow \mu^+\mu^-$ branching fraction. The two thresholds at 90% and 95% are indicated in red.

than the one required from the previous considerations.

The approach to the measurement is therefore driven by simplicity. A cut is made on the BDT to reject the combinatorial background, and a maximum likelihood fit is performed on the dimuon invariant mass so that $sWeights$ are extracted with the $sPlot$ method. The considered mass region begins at $5320 \text{ MeV}/c^2$, where all background sources but the combinatorial one are negligible. Finally, a maximum likelihood fit is performed on the $sWeighted$ decay time distribution to extract $\tau_{\mu^+\mu^-}$.

6.3.1 Analysis strategy

The effective lifetime measurement employs the same dataset used for the branching fraction measurement, which is reported in Table I. The data selection is similar to the one described in Sec. 4.1, with the exception of a softer PID cut (Table 3.3) and a cut on the BDT output has been imposed, rather than performing a simultaneous fit in multiple BDT bins. The softening of the PID cut is justified by the fact that $B_{d,s}^0 \rightarrow h^+h^-$ and semileptonic backgrounds are less relevant in

the proximity of the $B_s^0 \rightarrow \mu^+\mu^-$ signal then they are in the $B_d^0 \rightarrow \mu^+\mu^-$ region, to which the branching fraction analysis is aiming, as can be seen from the toy data of Fig. 6.10. In addition, the fit has been chosen to be performed on data

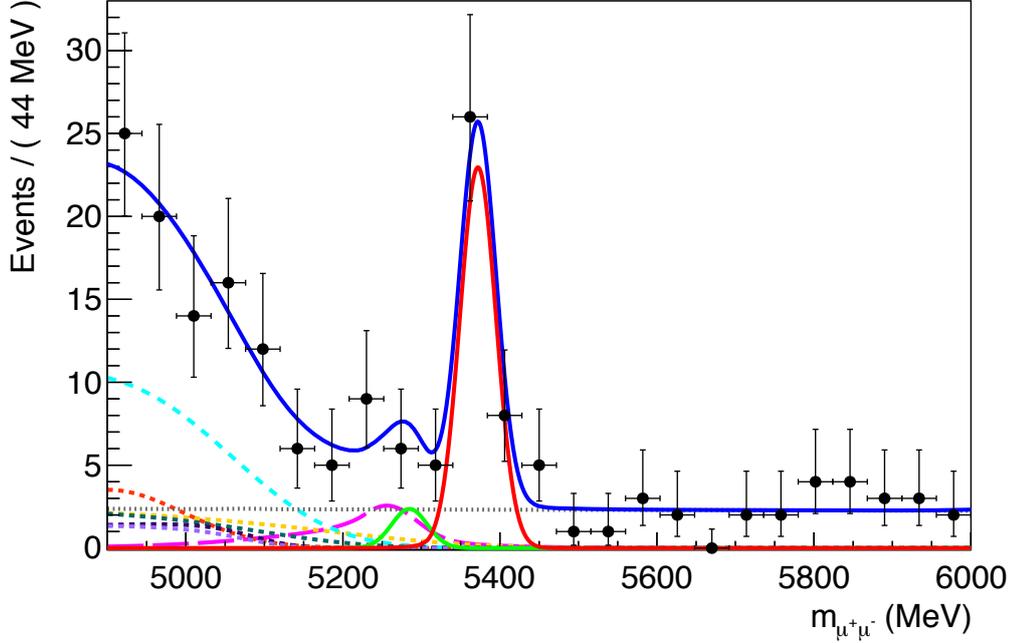


Figure 6.10: Dimuon invariant mass spectrum generated by a typical toy MC. The *pdfs* used for the generation are superimposed: the most relevant being $B_d^0 \rightarrow \mu^+\mu^-$ (solid green), $B_s^0 \rightarrow \mu^+\mu^-$ (solid red), combinatorial (narrow-dashed grey) and $B_{d,s}^0 \rightarrow h^+h'^-$ (dashed magenta). The other components describe the semileptonic backgrounds: $B_d^0 \rightarrow \pi^-\mu^+\nu_\mu$ (dashed cyan), $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (dashed red), $B^+ \rightarrow \pi^+\mu^+\mu^-$ (dashed violet), $B^0 \rightarrow \pi^0\mu^+\mu^-$ (dashed lilac), $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ (dashed orange) and $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ (dashed green). The full selection including the PID is applied, together with the cut $\text{BDT} > 0.55$.

having an invariant mass greater than $5320 \text{ MeV}/c^2$, a cut which removes almost all semileptonic backgrounds, whose components are therefore not included in the fit model.

The value of the BDT cut has been optimised by means of toy MC experiments, with the aim of minimising the statistical uncertainty on the effective lifetime. About 10000 simulated datasets were generated for each BDT cut ranging from 0.4 to 0.6 in steps of 0.05, and each dataset has been fitted to extract $\tau_{\mu^+\mu^-}$. The results of these investigations showed that the cut $\text{BDT} > 0.55$ yields the highest signal sensitivity as well as the lowest statistical error on the effective lifetime.

Decay time acceptance

Since the signal selection depends on the lifetime of the $B_s^0 \rightarrow \mu^+\mu^-$ candidates, a bias is introduced into the decay time distribution, and must be corrected for in the fit. To this end, an *acceptance function* models the dependency of the efficiency on the decay time t as follows:

$$\epsilon(t) = \frac{[a(t - t_0)]^n}{1 + [a(t - t_0)]^n}, \quad (6.15)$$

where the parameters have been determined on a simulated $B_s^0 \rightarrow \mu^+\mu^-$ sample. To validate the above acceptance function, the same analysis strategy has been used to measure the well known lifetime of the $B_d^0 \rightarrow K^+\pi^-$ decay, which is equal to the mean B_d^0 lifetime. A fit is thus performed on the invariant mass of the $B_d^0 \rightarrow K^+\pi^-$ candidates, selected with the same criteria of the calibration procedure (Sec. 4.2.4), and shown in Fig. 6.11. The ${}_s\mathcal{W}eights$ extracted from the fit are then used to obtain the background subtracted decay time distribution, also shown in Fig. 6.11. The latter is then fitted using the acceptance function (6.15), whose parameters are determined from simulated $B_d^0 \rightarrow K^+\pi^-$ events. The lifetime fit yields:

$$\tau(B_d^0 \rightarrow K^+\pi^-) = 1.52 \pm 0.03 \text{ ps}, \quad (6.16)$$

with a very good agreement with the PDG value of $\tau(B_d^0) = 1.520 \pm 0.004$ [158]. The quoted uncertainty represents the difference observed between data and simulation in the acceptance function, and is assigned as a systematic error. A further cross-check is made by measuring the $B_s^0 \rightarrow K^+K^-$ lifetime, once more yielding a value consistent with the PDG.

The $B_s^0 \rightarrow \mu^+\mu^-$ signal decay time acceptance function, being modelled from MC events, assumes the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime to be the one of the heavy B_s^0 mass eigenstate, i.e. $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$. Since both trigger and offline selection tend to reject short-lived particles, the efficiency on the eventual light mass component would be overestimated. To estimate the size of this effect, which can potentially disguise new physics effects, a fit is performed to one million $B_s^0 \rightarrow \mu^+\mu^-$ events generated under $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0$, i.e. with equal mixture of light and heavy mass eigenstates. The fit employs the same acceptance function used in the analysis, and the observed shift in the fitted effective lifetime, 0.018 ps, is taken as a systematic uncertainty.

6.3.2 Results

A fit is performed to the invariant mass of the $B_s^0 \rightarrow \mu^+\mu^-$ candidates to extract the ${}_s\mathcal{W}eights$, which are then used to weight the decay time distribution. The

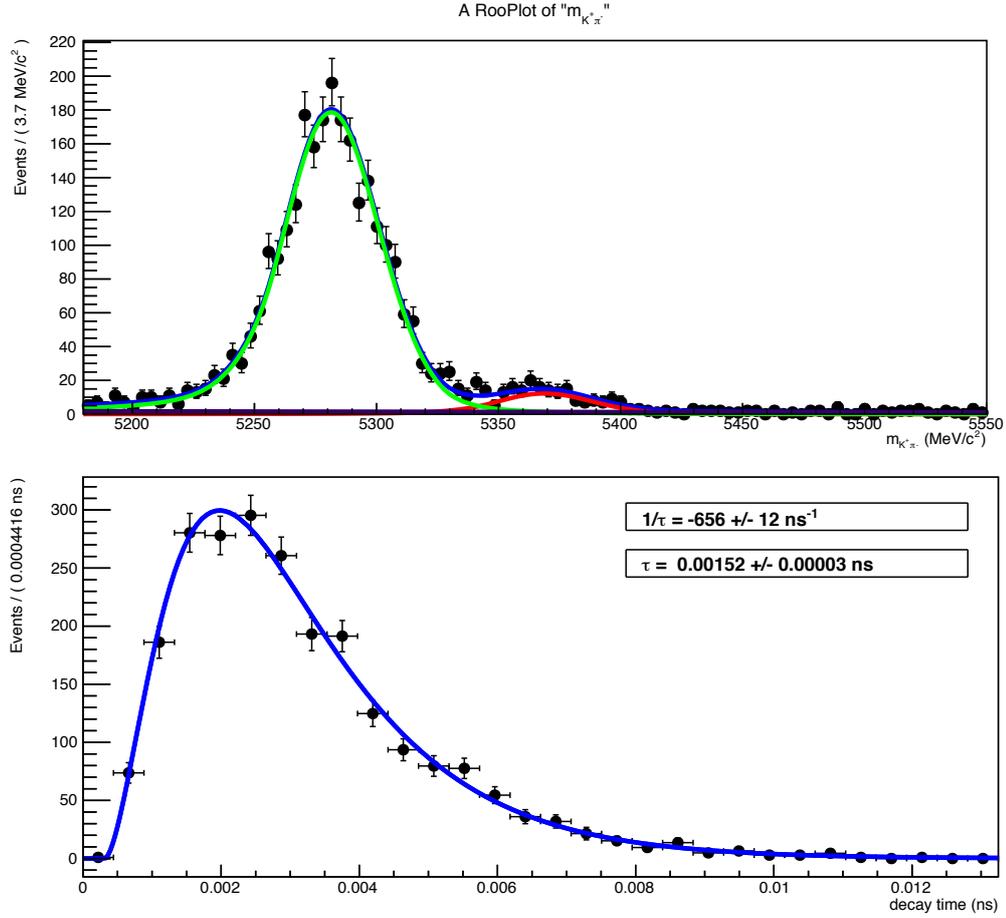


Figure 6.11: (Top) Fit to the invariant mass distribution of $B_d^0 \rightarrow K^+\pi^-$ candidates used to extract the $sWeights$. The fit model includes the $B_d^0 \rightarrow K^+\pi^-$ (green), $B_s^0 \rightarrow K^+\pi^-$ (red) and combinatorial (dark blue) components. (Bottom) Decay time distribution of the $B_d^0 \rightarrow K^+\pi^-$ candidates, as determined by the $sWeights$.

correlation between mass and lifetime has been measured to be very small on both signal MC and data sidebands, hence confirming the needed requirement for the $sPlot$ application [146]. The fit model includes the signal Crystal Ball of Sec. 4.3 and an exponential function for the combinatorial background, all the other components being neglected, given their small contribution beyond the 5320 MeV/c² cut. The systematic uncertainty for neglecting the remaining background components is evaluated by producing toy experiments where such components are generated according to a Poisson distribution with a mean value given by the integral of their mass pdf above 5320 MeV/c², i.e. their mean yield is computed

as from Sec. 5.1 and Sec. 5.2 with a selection efficiency that includes the harder mass cut. The resulting systematic uncertainty amounts to only 0.007 ps.

The results of the invariant mass and s Weighted decay time fits are shown in Fig. 6.12. The first measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime yields:

$$\tau(B_s^0 \rightarrow \mu^+\mu^-) = 2.04 \pm 0.44 \pm 0.05 \text{ ps}, \quad (6.17)$$

the first uncertainty being statistical and the second one systematic. By using Eq. (1.32), the measurement can be converted into [185]:

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 8.24 \pm 10.72, \quad (6.18)$$

which is consistent with the $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$ hypothesis at 1.0σ and with $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$ at 1.4σ . The current statistical uncertainty therefore prevents to assess a constraint on $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$, but this result enestablishes the potential of the effective lifetime measurement in probing new physics scenarios with the data that LHCb is expected to collect in the coming years, as discussed in Sec. 6.4.

6.4 Implications and future prospects

The study of the rare $B_{d,s}^0 \rightarrow \mu^+\mu^-$ processes is of paramount importance to test the flavour interactions: both the branching fraction and effective lifetime observables are highly sensitive to new physics effects, which can variously alter the Standard Model predictions. To this end, the discussion of Sec. 1.3 continues here in the light of the results presented in the previous section, and of the work of [185] and [186]. The most precise experimental evaluation of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction can be obtained by combining the CMS measurement [60] from 2013 with the result (6.5) presented in this work, to give [185]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)_{\text{LHCb17+CMS13}} = (3.0 \pm 0.5) \times 10^{-9}. \quad (6.19)$$

The ratio (1.45) between the experimental and SM branching fraction then yields⁴:

$$R_{\text{LHCb17+CMS13}} = 0.84 \pm 0.16. \quad (6.20)$$

Therefore, under the assumption that no new sources of CP violation are present, i.e. Wilson coefficients are real, the constraint from the measured value of R can be represented in the $S - P$ plane as a circular band, shown in Fig. 6.13. A future precise measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime will thus pin down values

⁴Note that the updated SM prediction $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = 3.57 \pm 0.16$ has been used in [185].

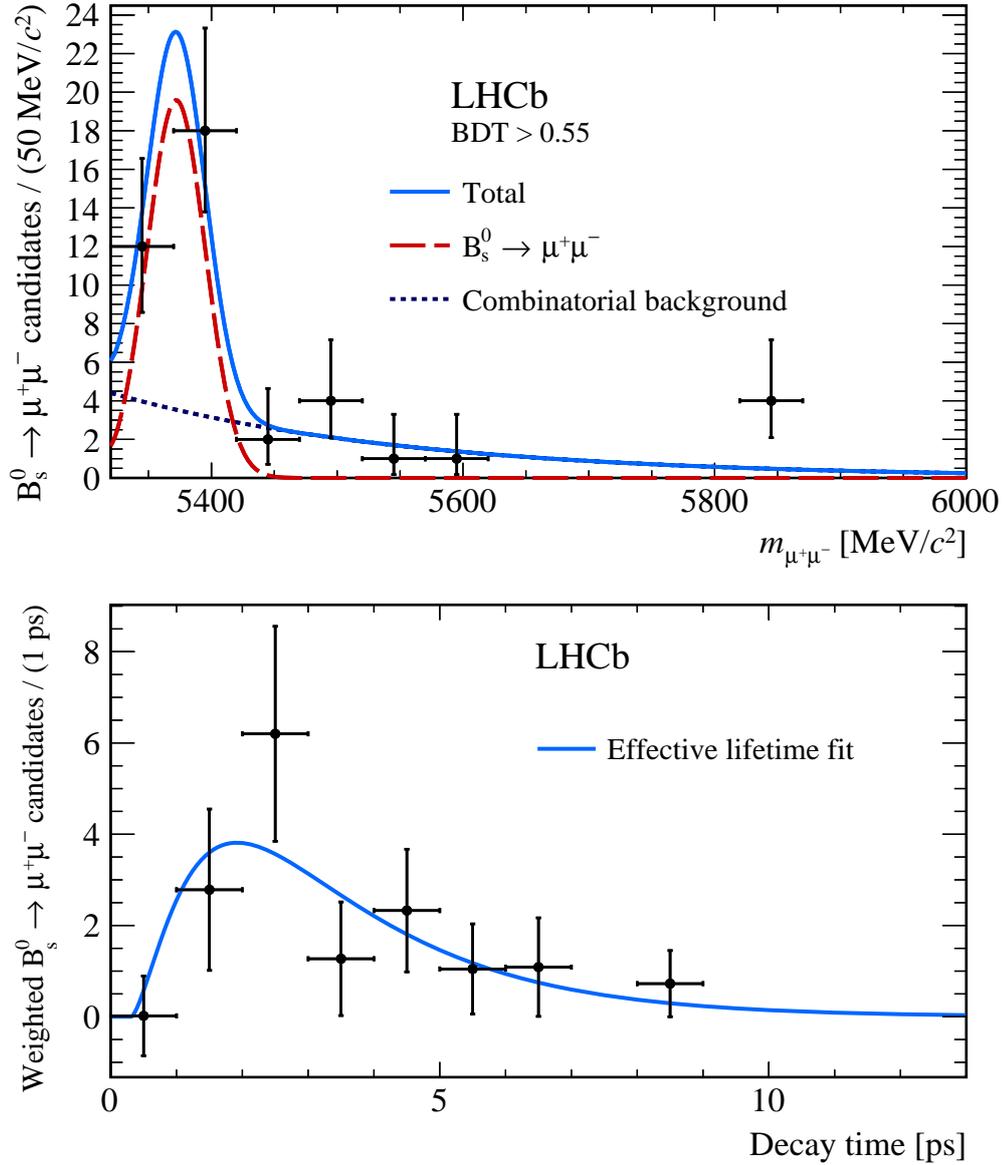


Figure 6.12: (Top) Mass distribution of the selected $B_s^0 \rightarrow \mu^+\mu^-$ candidates with $\text{BDT} > 0.55$. The fit curve is overlaid together with its $B_s^0 \rightarrow \mu^+\mu^-$ (red dashed) and combinatorial (blue dashed) components. (Bottom) Background subtracted decay-time distribution of the selected $B_s^0 \rightarrow \mu^+\mu^-$ candidates with the fit result superimposed.

for the P and S coefficients up to a two-fold ambiguity⁵.

⁵ This remaining ambiguity can be solved by measuring the CP-violating asymmetry, as described in [44, 186], which however requires flavour tagging and is hence prohibitive at LHC.

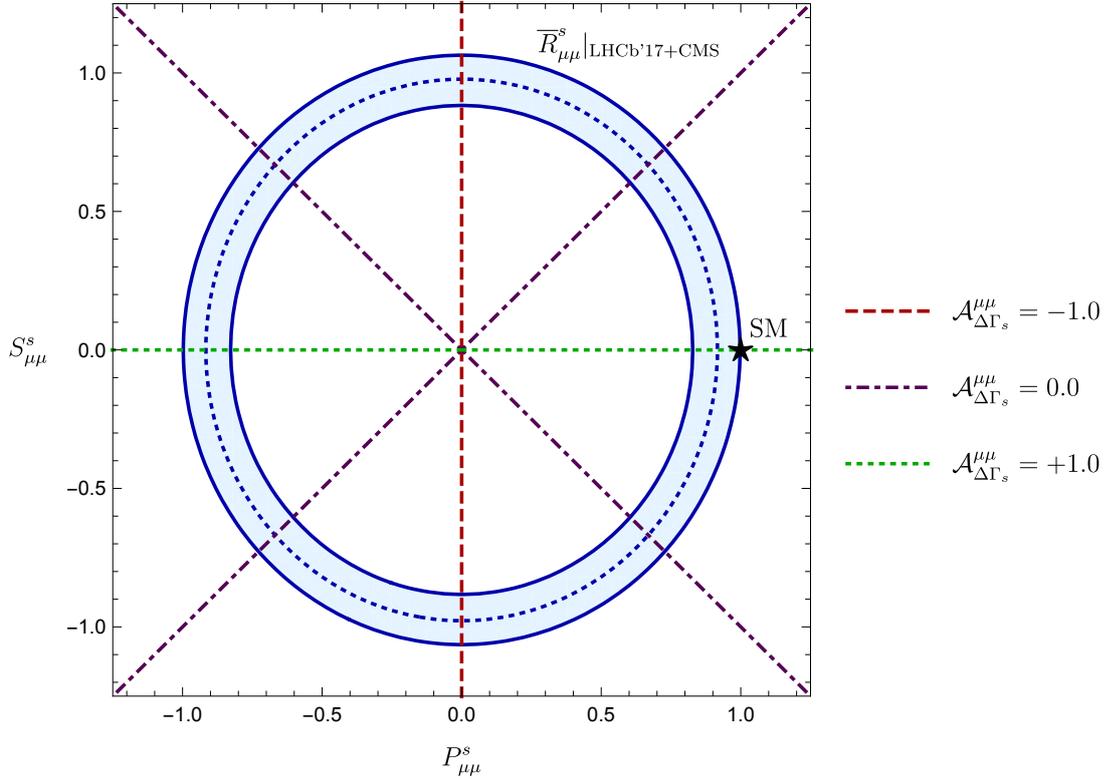


Figure 6.13: Constraints in the $P - S$ plane, where the blue circular band is given by the current experimental information on R.

Fig. 6.14 reports the current experimental constraints in the $m_A - \tan\beta$ plane, under the MFV assumption and without new CP violating phases [186]. Being proportional to $\tan^6\beta$, the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction is in fact a powerful tool to investigate the parameter space of SUSY models. The plot shows a broad allowed region at large m_A and small $\tan\beta$, corresponding to the SM scenario, $S \sim 0$ and $P \sim 1$, and a thin region at larger $\tan\beta$ originating in the case of sizeable NP amplitudes, $S \sim 1$ and $P \sim 1$. The latter cannot be excluded by the branching fraction measurement alone, as discussed in the following.

The branching fraction as well as the effective lifetime measurements are dominated by statistical uncertainty, hence there is large room for improvement in the coming years of LHCb data-taking. On the theory side, the branching fraction precision is limited by the uncertainties on the CKM element V_{cb} and the decay constant f_B : significant improvements in lattice QCD calculations, on which both these quantities strongly depend, are expected in the coming years [187]. The authors of [186] have evaluated an interesting projection at 50 fb^{-1} (“Run 4”) and 300

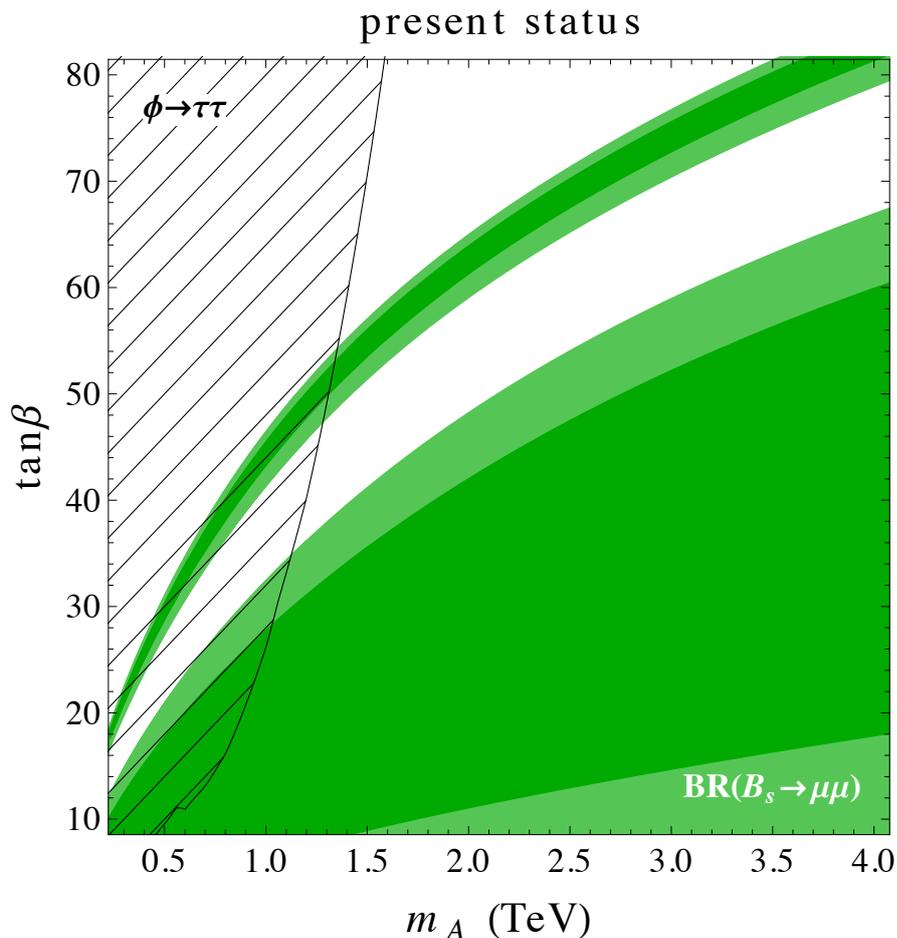


Figure 6.14: Constraints in the $m_A - \tan\beta$ plane for the MSSM scenario of [186], with the light Higgs mass fixed to $m_H = 125 \text{ GeV}/c^2$. The dark and light green regions are allowed by the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ measurement at the 1 and 2 standard deviations level. The black hatched region is excluded by direct searches for $\tau^+\tau^-$ resonances.

fb^{-1} (“Run 5”) of data collected at LHCb, at which the following precisions are realistically foreseen:

$$\begin{aligned} \Delta\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) &= 0.19 \times 10^{-9}, \quad \Delta\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0.8 \quad (\text{“Run 4”}), \\ \Delta\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) &= 0.08 \times 10^{-9}, \quad \Delta\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0.3 \quad (\text{“Run 5”}), \end{aligned} \quad (6.21)$$

provided the systematic uncertainties can be sufficiently reduced. Within these two scenarios, the precision on the Wilson coefficients under the MFV assumption will be reduced according to Fig. 6.15, for the scenario of a single real Wilson coefficient $C_S = -C_P$. Two regions are allowed by a SM-like branching fraction,

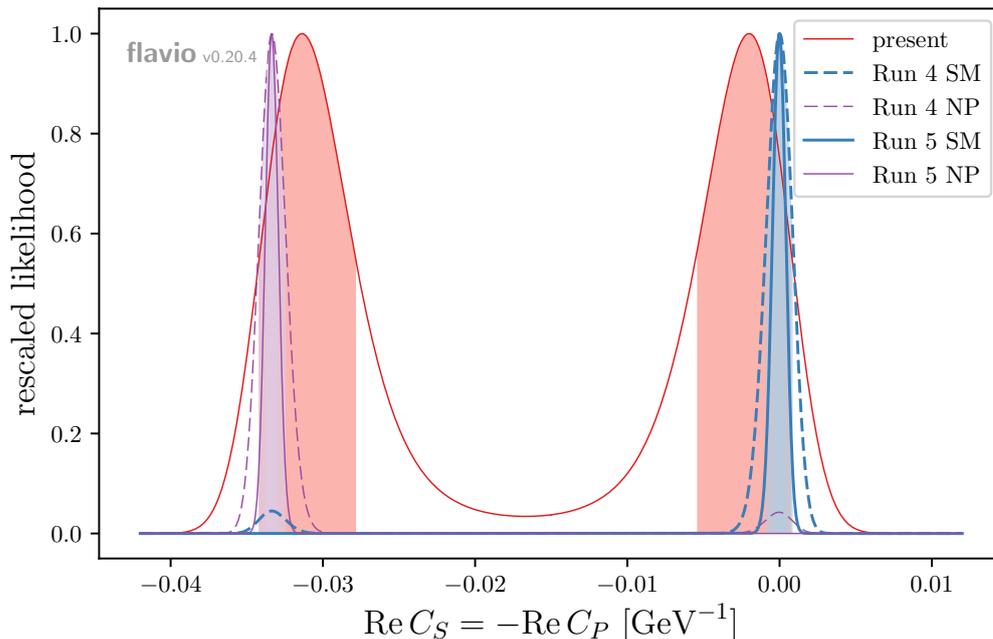


Figure 6.15: Present and future constraints on the real part of the Wilson coefficient C_S , given as $\mathcal{L}/\mathcal{L}_{max}$, where the SM scenario assumes $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$ and the NP one considers $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$. The shaded areas correspond to 1 standard deviation, i.e. the regions where the likelihoods contain 68.3% of their total integral.

for which the shaded areas indicate the expected precision in the higher statistics scenarios. Only a measurement of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ will be able to exclude one of the two solutions.

The same projections are made for the considered SUSY model in Fig. 6.16. While both the allowed regions at low and high $\tan\beta$ are reduced by increasing the statistics, the branching fraction measurement alone is not able to exclude the latter: to do so, a precise measurement of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ is needed.

6.5 Conclusions

The $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fractions and $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurements have been published in [188]. With Run 2 data and an optimised analysis, the first single-experiment observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay as well as the most precise measurement of its branching fraction were made possible. As such measurements entered the precision era, more and more NP models are falsified, or at least strongly constrained. At the same time, the pioneering analysis of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime demonstrates the feasibility of a measurement of

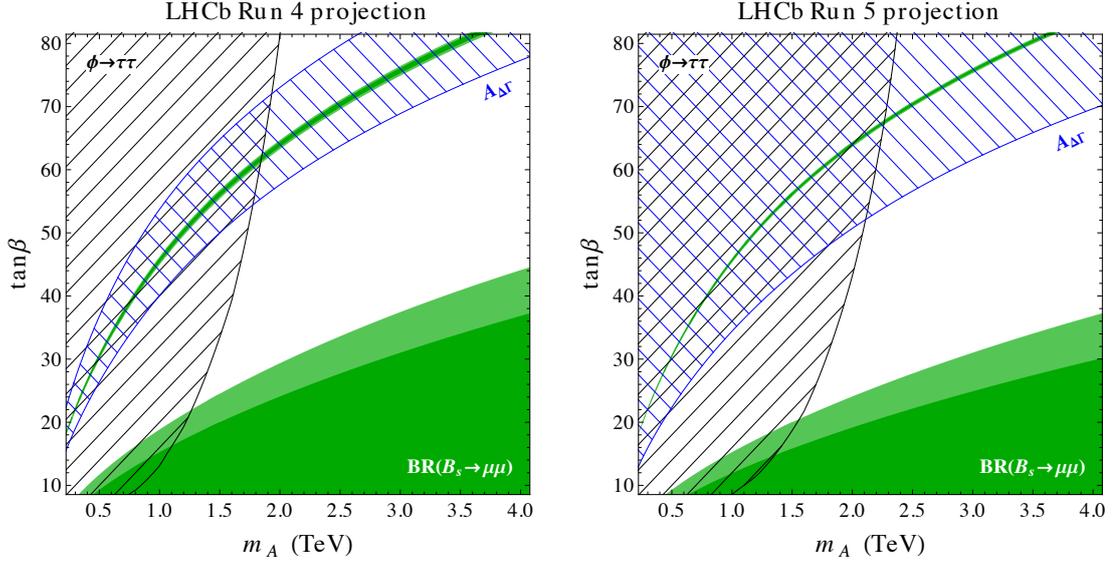


Figure 6.16: Constraints in the $m_A - \tan\beta$ plane for the MSSM scenario of [186] in the “Run 4” and “Run 5” projections. The blue hatched regions are probed by the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime: $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} < -0.6$ (left) and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} < 0.4$ (right).

the mass-eigenstate rate asymmetry $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$. This observable probes regions in the parameter space of many NP models which cannot be accessed by the branching fraction measurement alone, i.e. even if an agreement is found with the SM expectation of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction, NP may still be revealed if the $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ measurement deviates from +1.

Concerning $B_d^0 \rightarrow \mu^+\mu^-$ decays, with an improved analysis and enhanced statistics, the excess observed in the previous measurement [4] is not confirmed. However, a strong limit on the $B_d^0 \rightarrow \mu^+\mu^-$ branching fraction has been set, which starts to approach the SM value. Although affected by the exclusive backgrounds estimates, the measurement precision is dominated by the available statistics, hence large room for improvement is foreseen in the coming years.

List of Figures

1.1	The elementary particles of the Standard Model.	4
1.2	Higgs potential for $\mu^2 < 0$ [18].	6
1.3	State of the art of the $\bar{\rho} - \bar{\eta}$ constraints as from the CKMFitter group [38].	9
1.4	Dominant Feynman diagrams for $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ decays	10
1.5	Neutron β -decay at the quark level in the full (a) and effective (b) theory.	11
1.6	Ratio between the theoretical and experimental branching fraction as a function of y_s for different $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ scenarios [48]. The value of y_s is taken from the LHCb measurement published in [49].	15
1.7	ISR (dotted blue line) and FSR (solid red line) contributions to the dimuon invariant mass spectrum in $B_{d,s}^0 \rightarrow \mu^+ \mu^- + n\gamma$. The vertical green lines indicate the CMS (dashed) and LHCb (dot-dashed) signal windows [53].	17
1.8	Illustration of allowed regions in the $R - \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ plane for scenarios with scalar or non-scalar NP contributions [57].	20
1.9	The relation between P and S for different $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ cases [59]. The measured R is taken from the combination of CMS [60] and LHCb [61] results. The observable \bar{R} follows the definition (1.45).	21
1.10	$B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ branching fraction predictions in MFV, RS, SM4 and several MSSM models [69]. The grey area on the right shows the CDF measurement [70] before the LHC results, while the CMS+LHCb measurement [4] is superimposed in cyan.	23
1.11	(a) Higgs mediated $B_s^0 \rightarrow \mu^+ \mu^-$ process in the 2HDM. (b) Higgs mediated $B_s^0 \rightarrow \mu^+ \mu^-$ decay in SUSY.	26
1.12	Constraints in the $m_A - \tan\beta$ plane from the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay at LHCb [76]. The grey area is excluded by direct searches of $\tau^+ \tau^-$ resonances, while the red shaded areas marked with the letters $a-d$ correspond to exclusions of the different MSSM scenarios considered in [77].	27

1.13	History of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ limits and measurements across the years [78].	28
1.14	The experimental and theoretical picture of the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fraction measurements as of 2016 [79]. In red the SM prediction [53], in blue the ATLAS measurement [79] and in grey the combined CMS+LHCb one [4], together with their respective confidence ranges.	29
2.1	The CERN accelerator complex [86].	32
2.2	Development of the instantaneous luminosity for ATLAS, CMS and LHCb during a typical LHC fill in Run 1 [90]. On the lower side of the figure a cartoon shows how the LHCb luminosity is levelled by adjusting the transverse beam overlap for about 15 hours, after which the beams are colliding head-on. After almost 20 h, the beam is eventually dumped.	34
2.3	Integrated luminosity collected at LHCb per year [92].	35
2.4	(a) Diagram of two interacting partons a and b inside the hard-scattered protons A and B [94]. (b) Polar angular distribution of the $b\bar{b}$ pairs produced at LHC for $\sqrt{s} = 14$ TeV as simulated with the PYTHIA8 event generator [95,96]. The acceptance of the LHCb detector is highlighted in red.	36
2.5	Pair creation (a), flavour excitation (b) and gluon splitting (c) processes.	37
2.6	Total bottom cross sections for pp collisions as a function of the centre-of-mass energy, showing the contributions from pair creation, flavour excitation and gluon splitting processes [100].	38
2.7	(Left) Energy-substituted mass distribution of $B^\pm \rightarrow D(\pi^- K^+)\pi^\pm$ decays at Babar [108], $m_{ES} = \sqrt{(s/2 + \vec{p}_\Upsilon \cdot \vec{p}_B)^2/E_\Upsilon^2 - p_B^2}$. (Right) Invariant mass distribution of $B^- \rightarrow D(\pi^- K^+)\pi^-$ decays at LHCb [109].	39
2.8	Side view of the LHCb detector.	41
2.9	Perspective view of the LHCb magnet [112].	42
2.10	The VELO detector [80].	44
2.11	(a) Schematics of the R and Φ VELO sensors. (b) The VELO vacuum vessel.	45
2.12	(a) Resolution of the x and y components of the IP as a function of momentum [115]. (b) PV resolution as a function of track multiplicity. The superimposed histogram shows the distribution of the number of tracks per reconstructed PV [90]. Both plots are made using data collected in 2012.	45

2.13	(a) Layout of the four TT stations. (b) Layout of the x layer (top) and v layer (bottom) of the IT in the T2 station.	46
2.14	(a) Overview of the tracking stations: the TT, IT (purple) and the OT (cyan). (b) Cross section of an OT module with a zoom on the straw-tubes. The dimensions are in mm.	47
2.15	(a) Side view of the RICH1 detector. (b) Reconstructed Cherenkov angle as a function of track momentum in RICH1 [119].	49
2.16	Efficiency of the kaon identification (red) and pion to kaon misidentification (black) as a function of the momentum, as measured from 2015 data. Two cuts on the delta log-likelihood (ΔLL) between kaon and pion hypotheses are shown [120].	50
2.17	Lateral segmentation of one quadrant of SPD/PS and ECAL (left) and HCAL (right) as seen by front. Cell sizes and number of channels refer to ECAL (left) and HCAL (right).	51
2.18	(a) Exploded view of one ECAL module. On the right, the assembled stack with inserted fibers. (b) A HCAL module where two ejected layers highlight its elementary periodic structure.	52
2.19	Lateral view of the five muon station arrangement.	53
2.20	On the left, front view of one quadrant of a muon station. The regions R1-R4 are defined by groups of chambers, which are represented by a rectangle. On the right, the division of the M1 chambers into logical pads. The number of pad rows per chamber is the same for each station, while the number of pad columns, which define the x granularity, is doubled in M2-M3 and halved in M4-M5 with respect to the M1 station shown in the picture.	54
2.21	(a) Mixed wire-cathode pads readout for a M2R1 chamber. The logical pads, shown in black, are obtained by the coincidence of crossing vertical wire pads and cathode pads. (b) Summary table of the muon system layout. z is the distance from the interaction point, while Δx and Δy represent the dimensions of a quadrant. The R1-R4 rows contain the number of logical pads per chamber (square brackets) and their size along x and y , while the sizes projected onto M1 are reported in parenthesis. All dimensions are given in cm.	55
2.22	(a) Schematic cross section of a four gap MWPC. SPB stands for Spark Protection Board, while CARDIAC is the front-end electronic board. (b) Layout of the triple-GEM detector showing the three GEM foils and the corresponding gaps, together with their dimensions.	56
2.23	Average hit efficiency, measured in 2011 and 2012 data, for all the chamber types in the muon detector.	57

2.24	Architecture of the muon system electronics.	59
2.25	Scheme of the L0-muon trigger [130].	61
2.26	Track types at LHCb [133].	63
2.27	Display of a typical LHCb event, recorded during the first day of data taking in 2017.	67
I	BDT distribution of combinatorial and $B_d^0 \rightarrow \mu^+\mu^-$ signal events.	70
3.1	Probability for a muon to reach a given station as a function of its momentum, from a simulated $J/\psi \rightarrow \mu^+\mu^-$ sample [148]. Note that the calorimeters, located upstream of M2, accounts for almost 7 interaction lengths (Sec. 2.4.2), while the calorimeters plus iron absorbers thickness amount to 20 interaction lengths.	74
3.2	Simulated momentum and transverse momentum of muons originating from $B_s^0 \rightarrow \mu^+\mu^-$ decays at $\sqrt{s} = 13$ TeV. Both tracks are required to lie within the muon detector acceptance.	75
3.3	IsMuon efficiency (a) and proton (b), pion (c), kaon (d) misidentification probabilities as a function of momentum for different transverse momentum regions. Non-physical efficiency values may arise from the background subtraction procedure, as discussed in Sec. 3.2.2.	77
3.4	(a) D^2 distribution for protons, pions, kaons and muons. (b) The corresponding muDLL distribution. The values are computed on data from 2011 calibration samples after the IsMuon selection [147].	78
3.5	Invariant mass distribution for the $\Lambda \rightarrow p\pi^-$ (top left), $D^{*+} \rightarrow \pi^+D^0(\rightarrow K^-\pi^+)$ (top right) and $J/\psi \rightarrow \mu^+\mu^-$ (bottom) samples. The superimposed fit (red line) shows the signal (dashed blue) and background (dotted dash green) components. All distributions are populated with 2015 data [151].	82
3.6	Efficiency of the muon PID selection of the branching fraction analysis on muon calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T	86
3.7	Efficiency of the muon PID selection of the branching fraction analysis on pion calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T	88
3.8	Efficiency of the muon PID selection of the branching fraction analysis on kaon calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T	89
3.9	Efficiency of the muon PID selection of the branching fraction analysis on proton calibration data for Run 1, 2015 and 2016, as a function of momentum in bins of p_T	90

4.1	Cartoon of a combinatorial $B_{d,s}^0 \rightarrow \mu^+\mu^-$ event. Two B mesons produced at the PV (green ellipse) decay and produce two muons, whose track extrapolations (dashed pink) form a B -like vertex (dashed blue).	94
4.2	Scheme of a decision tree [149]. At each node, the split is performed by cutting on the variable which gives the best separation between the signal (S) and background (B) classes. The same variable may thus be used in several nodes, while others might not be used at all.	96
4.3	ROC curves for the long track (a) and VELO track (b) isolations evaluated on Run 1 (red), 2015 (orange) and 2016 (brown) data. The performance of the cut-based isolation (black) used in the past analysis is superimposed in (a).	98
4.4	Distribution of the BDT input variables in $B_s^0 \rightarrow \mu^+\mu^-$ MC events (2011, 2015 and 2016 conditions) and in the right mass data side-band (Run 1, 2015 and 2016).	100
4.5	ROC of the new BDT in Run 1 data and comparison with the ROC of the BDT used in the past analysis.	101
4.6	ROC of the new BDT in Run 1, 2015 and 2016 data.	101
4.7	Invariant mass distribution of $B_d^0 \rightarrow K^+\pi^-$ events in 2016 data for $\text{BDT} \in [0.25, 0.4]$. The fit result is overlaid, showing the B_d^0 (red), B_s^0 (green), Λ_b (dashed red) and combinatorial (dashed blue) components. $\kappa = 5$ indicates the PID cut value: $\text{DLL}(K - \pi) > \kappa$ for the kaon and $\text{DLL}(K - \pi) < -\kappa$ for the pion.	103
4.8	BDT distributions for Run 1 (top), 2015 (middle) and 2016 (bottom) data samples. The statistical uncertainty is shown in red, while the systematic one is in grey. The black points mark the resulting distribution for the $B_d^0 \rightarrow K^+\pi^-$ simulated sample.	104
4.9	Invariant mass fits to $B_d^0 \rightarrow K^+\pi^-$ (left) and $B_s^0 \rightarrow K^+K^-$ (right) 2016 data. The red line identifies the B_d^0 pdf, the green line is the B_s^0 pdf and the blue one represents the combinatorial background. $\kappa = 10$ indicates the PID cut value: $\text{DLL}(K - \pi) > \kappa$ for the kaon and $\text{DLL}(K - \pi) < -\kappa$ for the pion.	106
4.10	(Left) Invariant dimuon mass fits on $\Upsilon(1, 2, 3S) \rightarrow \mu^+\mu^-$ 2016 data. (Right) Interpolation of the invariant mass resolution of the dimuon resonances for 2016 data.	106
4.11	Invariant mass spectrum of the $B^+ \rightarrow J/\psi K^+$ candidates in 2016 data. The fit (solid blue) is superimposed together with its components: $B^+ \rightarrow J/\psi K^+$ (dashed red), $B^+ \rightarrow J/\psi \pi^+$ (dashed purple) and combinatorial (dashed green).	110

5.1	Mass <i>pdfs</i> of all the backgrounds sources in the region $\text{BDT} > 0.5$. The signal region is indicated by the green vertical lines.	113
5.2	(a) q^2 spectra of $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays for the ISGW2 model (black) and for a fit to lattice data and recent results (red); both histograms are normalised to the same area. (b) Fit to lattice data and recent results [166] used for the left plot.	120
5.3	$\pi - \mu$ invariant mass versus q^2 in simulated $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays. .	120
5.4	Argus*Gaussian fit to the invariant mass distribution of simulated $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ decays in the BDT bins used for the signal fit. . . .	122
5.5	(a) q^2 spectra of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays for the ISGW2 model (black) and for a fit to recent lattice results (red); both histograms are normalised to the same area. (b) $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ q^2 spectra taken from [166].	123
5.6	Argus*Gaussian fit to the invariant mass distribution of simulated $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ decays in the BDT bins used for the signal fit. .	125
5.7	Argus*Gaussian fit to the invariant mass distribution of simulated $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ decays in the BDT bins used for the signal fit.	127
5.8	Argus*Gaussian fit to the invariant mass distribution of simulated $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays in the BDT bins used for the signal fit. . . .	129
5.9	Example of a $\pi - \mu$ fit showing the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ shoulder (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ peak with single muon misidentification (green) and combinatorial (red) components. The fit (blue) is performed on Run 1 data in the BDT bin $[0.4, 0.5]$. The invariant mass is expressed in MeV/c^2	130
5.10	Comparison between the number of $B_{d,s}^0 \rightarrow h^+ h'^- \rightarrow \mu^+ \mu^-$ events from the $B_d^0 \rightarrow K^+ \pi^-$ TIS estimate (blue) and from the data-driven estimates (red and grey) as a function of the BDT bin for Run 1, 2015 and 2016 data.	137
5.11	Comparison between the number of $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ events from the MC-driven estimate (blue) and from the data-driven estimate (red) as a function of the BDT bin for Run 1, 2015 and 2016 data. . . .	139
5.12	Comparison between the number of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ events from the MC-driven estimate (blue) and from the data-driven estimate (grey) as a function of the BDT bin for Run 1, 2015 and 2016 data. .	141
5.13	Invariant mass fit to Run 1 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 . .	148
5.14	Invariant mass fit to 2015 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c^2 . .	149

5.15	Invariant mass fit to 2016 $\pi - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c ² .	150
5.16	Invariant mass fit to Run 1 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c ² .	151
5.17	Invariant mass fit to 2015 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c ² .	152
5.18	Invariant mass fit to 2016 $K - \mu$ data in bins of BDT. The three components of the fit model (blue) are $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (cyan), $B_{d,s}^0 \rightarrow h^+ h'^-$ (green) and combinatorial (red). The dimensions are MeV/c ² .	153
6.1	Invariant mass fits of the dimuon sidebands for the 8 BDT bins of Run 1 data. The fit components are indicated with their respective colours in the legend.	156
6.2	Invariant mass fits of the dimuon sidebands for the 8 BDT bins of 2016 data. The fit components are indicated with their respective colours in the legend.	157
6.3	Pulls of the $B_s^0 \rightarrow \mu^+ \mu^-$ BF for Run 1 toys in the 7-bins (left), 5-bins (right) and 4-bins (bottom) configurations.	158
6.4	Pulls of the combinatorial in the most significant BDT bin for Run 1 toys in the 5-bin (left) and 4-bin (right) configurations.	159
6.5	Expected sensitivities on $B_s^0 \rightarrow \mu^+ \mu^-$ (left) and $B_d^0 \rightarrow \mu^+ \mu^-$ (right) decays from the toys on the full dataset with the 4-bin configuration.	159
6.6	Mass distribution of the selected $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidates for BDT>0.5. The fit curve is overlaid, together with its components.	161
6.7	Mass distribution of the selected $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidates in each BDT bin used in the fit. The fit curve is overlaid together with its components, following the same legend of Fig. 6.6.	162
6.8	Contours of constant likelihood ratio in the space of the two parameters of interest. The values of the confidence regions are indicated at each contour, according to the formula (6.10). The SM expectation is superimposed in red.	165
6.9	The CL _s confidence belt for the $B_d^0 \rightarrow \mu^+ \mu^-$ branching fraction. The two thresholds at 90% and 95% are indicated in red.	167

6.10	Dimuon invariant mass spectrum generated by a typical toy MC. The <i>pdfs</i> used for the generation are superimposed: the most relevant being $B_d^0 \rightarrow \mu^+\mu^-$ (solid green), $B_s^0 \rightarrow \mu^+\mu^-$ (solid red), combinatorial (narrow-dashed grey) and $B_{d,s}^0 \rightarrow h^+h'^-$ (dashed magenta). The other components describe the semileptonic backgrounds: $B_d^0 \rightarrow \pi^-\mu^+\nu_\mu$ (dashed cyan), $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (dashed red), $B^+ \rightarrow \pi^+\mu^+\mu^-$ (dashed violet), $B^0 \rightarrow \pi^0\mu^+\mu^-$ (dashed lilac), $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ (dashed orange) and $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ (dashed green). The full selection including the PID is applied, together with the cut $\text{BDT} > 0.55$	168
6.11	(Top) Fit to the invariant mass distribution of $B_d^0 \rightarrow K^+\pi^-$ candidates used to extract the ${}_s\mathcal{W}eights$. The fit model includes the $B_d^0 \rightarrow K^+\pi^-$ (green), $B_s^0 \rightarrow K^+\pi^-$ (red) and combinatorial (dark blue) components. (Bottom) Decay time distribution of the $B_d^0 \rightarrow K^+\pi^-$ candidates, as determined by the ${}_s\mathcal{W}eights$	170
6.12	(Top) Mass distribution of the selected $B_s^0 \rightarrow \mu^+\mu^-$ candidates with $\text{BDT} > 0.55$. The fit curve is overlaid together with its $B_s^0 \rightarrow \mu^+\mu^-$ (red dashed) and combinatorial (blue dashed) components. (Bottom) Background subtracted decay-time distribution of the selected $B_s^0 \rightarrow \mu^+\mu^-$ candidates with the fit result superimposed.	172
6.13	Constraints in the $P-S$ plane, where the blue circular band is given by the current experimental information on R.	173
6.14	Constraints in the $m_A - \tan\beta$ plane for the MSSM scenario of [186], with the light Higgs mass fixed to $m_H = 125 \text{ GeV}/c^2$. The dark and light green regions are allowed by the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ measurement at the 1 and 2 standard deviations level. The black hatched region is excluded by direct searches for $\tau^+\tau^-$ resonances.	174
6.15	Present and future constraints on the real part of the Wilson coefficient C_S , given as $\mathcal{L}/\mathcal{L}_{max}$, where the SM scenario assumes $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$ and the NP one considers $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$. The shaded areas correspond to 1 standard deviation, i.e. the regions where the likelihoods contain 68.3% of their total integral.	175
6.16	Constraints in the $m_A - \tan\beta$ plane for the MSSM scenario of [186] in the ‘‘Run 4’’ and ‘‘Run 5’’ projections. The blue hatched regions are probed by the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime: $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} < -0.6$ (left) and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} < 0.4$ (right).	176

List of Tables

I	Data collected during the LHC Run 1 and Run 2 used for the present analysis.	71
3.1	Muon stations that a muon candidate must traverse to satisfy the IsMuon criterion, as a function of the momentum. 3 GeV/c is the threshold to traverse the first absorber, while 6 GeV/c are needed to reach the last station, as shown in Fig. 3.1.	76
3.2	Background levels with respect to the DLL cut used in the previous analysis [4]. The signal efficiency is also indicated in the last line, again relatively to the DLL one.	79
3.3	muonID cuts employed for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$ branching fraction (\mathcal{B}) and $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime (τ_{eff}) measurements on the analysed data samples.	85
4.1	Selection for $B_{d,s}^0 \rightarrow \mu^+\mu^-$, $B_{d,s}^0 \rightarrow h^+h'^-$ and $B^+ \rightarrow J/\psi K^+$ channels. The values in parenthesis for track χ^2/ndf and ghost probability show the softer cuts used for Run 2 data.	93
4.2	Mean values of the B_d^0 and B_s^0 masses for Run 1 and Run 2 data.	105
4.3	Dimuon invariant mass resolutions for Run 1 and Run 2 data, as obtained from the interpolation of charmonium and bottomonium resonances.	107
4.4	Crystal Ball tail parameters in Run 1 and Run 2 data.	107
4.5	$B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ normalisation factors and expected yields for Run 1 and Run 2 data. The expected signal yields are estimated assuming standard model branching fractions (SM) and the latest most precise experimental branching fraction measurements [4] (LHC).	111

5.1	Double misID probability in units of 10^{-6} for Run 1, 2015 and 2016 data. The $\text{PID}_{\mu,4}^{\text{MC12}}$ selection is used for Run 1 and 2015 data, whereas $\text{PID}_{\mu,8}^{\text{MC15}}$ is used for 2016 data, as discussed in Sec. 3.3.1. The quoted errors include MC statistics and PID efficiency statistical uncertainties for the single modes, and the branching fractions and f_s/f_d for the average.	115
5.2	Inputs for the computation of the $B_{d,s}^0 \rightarrow h^+h'^-$ peaking background (Eq. (5.1)) for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	116
5.3	Double misID probability in units of 10^{-6} as a function of the BDT bin for Run 1, 2015 and 2016 data. The quoted errors include MC statistics, PID efficiency statistical uncertainties, and the errors from the branching fractions and f_s/f_d	117
5.4	Inputs to Eq. 5.3 for the computation of the $B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	121
5.5	Inputs to Eq. 5.3 for the computation of the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	123
5.6	Inputs to Eq. 5.3 for the computation of the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	124
5.7	Inputs to Eq. 5.3 for the computation of the $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	125
5.8	Inputs to Eq. 5.3 for the computation of the $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	126
5.9	Inputs to Eq. 5.3 for the computation of the $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ background for Run 1, 2015 and 2016 data. The total estimated events are given in the last row.	128
5.10	PID cuts defying the $h-\mu$ selection. Since two hadrons are present in the final state, the cut is applied in the form $(\overline{\text{PID}}_h \& \text{PID}_\mu) \text{OR} (\text{PID}_\mu \& \overline{\text{PID}}_h)$. As for the $\mu-\mu$ PID selection, the <code>MC15TuneV1</code> of the ProbNN variables (Sec. 3.1.3) is adopted for 2016 data.	131
5.11	Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+ \mu^-$ events as extracted from $B_d^0 \rightarrow K^+ \pi^-$ TIS events as a function of the BDT bin for Run 1, 2015 and 2016 data.	136
5.12	Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+ \mu^-$ events as extracted from the $\pi-\mu$ fit as a function of the BDT bin for Run 1, 2015 and 2016 data.	136

5.13	Number of $B_{d,s}^0 \rightarrow h^+h'^- \rightarrow \mu^+\mu^-$ events as extracted from the $K - \mu$ fit as a function of the BDT bin for Run 1, 2015 and 2016 data.	136
5.14	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ MC-driven yields per BDT bin for Run 1, 2015 and 2016 data.	138
5.15	$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ yields per BDT bin as extracted from the $\pi - \mu$ fit for Run 1, 2015 and 2016 data.	138
5.16	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ MC-driven yields per BDT bin for Run 1, 2015 and 2016 data.	140
5.17	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ yields per BDT bin as extracted from the $K - \mu$ fit for Run 1, 2015 and 2016 data.	140
5.18	Background yields with total uncertainty for each BDT bin in Run 1 data.	142
5.19	Background yields with total uncertainty for each BDT bin in 2015 data.	143
5.20	Background yields with total uncertainties for each BDT bin in 2016 data.	144
5.21	Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/c ² , for Run 1 data.	145
5.22	Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/c ² , for 2015 data.	146
5.23	Background yields and statistical uncertainties per BDT bin in the mass range [5200, 5450] MeV/c ² , for 2016 data.	147
6.1	Expected significance and observation probability of the $B_s^0 \rightarrow \mu^+\mu^-$ decay mode for Run 1 toys.	159

Bibliography

- [1] **CLEO** Collaboration, R. Giles *et al.*, *Two-Body Decays of B Mesons*, *Phys. Rev.* **D30** (1984) 2279.
- [2] **D0** Collaboration, V. M. Abazov *et al.*, *Search for the rare decay $B_s \rightarrow \mu\mu$* , *Phys. Rev.* **D87** no. 7, (2013) 072006, [arXiv:1301.4507 \[hep-ex\]](#).
- [3] **CDF** Collaboration, T. Aaltonen *et al.*, *Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ decays with the full CDF Run II data set*, *Phys. Rev.* **D87** no. 7, (2013) 072003, [arXiv:1301.7048 \[hep-ex\]](#).
- [4] **LHCb**, **CMS** Collaboration, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \rightarrow \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, *Nature* **522** (2015) 68–72, [arXiv:1411.4413 \[hep-ex\]](#).
- [5] M. Testa and O. Benhar, *Interazioni Elettrodeboli*. Università di Roma “La Sapienza”, 2010.
- [6] F. Mandl and G. Shaw, *Quantum Field Theory*. Wiley, 1984.
- [7] L. Maiani, *Interazioni Elettrodeboli*. Università di Roma “La Sapienza”, 2011.
- [8] C.-N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, *Phys. Rev.* **96** (1954) 191–195.
- [9] E. Noether, *Invariant Variation Problems*, *Gott. Nachr.* **1918** (1918) 235–257, [arXiv:physics/0503066 \[physics\]](#). [Transp. Theory Statist. Phys.1,186(1971)].
- [10] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579–588.
- [11] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264–1266.

- [12] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc.* **C680519** (1968) 367–377.
- [13] R. W. Mcallister and R. Hofstadter, *Elastic Scattering of 188-MeV Electrons From the Proton and the α Particle*, *Phys. Rev.* **102** (1956) 851–856.
- [14] R. P. Feynman, *Very high-energy collisions of hadrons*, *Phys. Rev. Lett.* **23** (1969) 1415–1417.
- [15] J. D. Bjorken and E. A. Paschos, *Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon*, *Phys. Rev.* **185** (1969) 1975–1982.
- [16] M. Gell-Mann, *Symmetries of Baryons and Mesons*, *Phys. Rev.* **125** (Feb, 1962) 1067–1084.
<https://link.aps.org/doi/10.1103/PhysRev.125.1067>.
- [17] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509.
- [18] CMS Collaboration, W. de Boer, *The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology*, in *Time and Matter 2013 (TAM2013) Venice, Italy*. 2013.
 arXiv:1309.0721 [hep-ph]. <https://inspirehep.net/record/1252561/files/arXiv:1309.0721.pdf>.
- [19] ATLAS Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].
- [20] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].
- [21] ATLAS Collaboration, G. Aad *et al.*, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, *Phys. Lett.* **B726** (2013) 120–144, arXiv:1307.1432 [hep-ex].
- [22] CMS Collaboration, S. Chatrchyan *et al.*, *Measurement of the properties of a Higgs boson in the four-lepton final state*, *Phys. Rev.* **D89** no. 9, (2014) 092007, arXiv:1312.5353 [hep-ex].

- [23] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533.
<https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [24] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, *Phys. Rev.* **D2** (1970) 1285–1292.
- [25] **E598** Collaboration, J. J. Aubert *et al.*, *Experimental Observation of a Heavy Particle J* , *Phys. Rev. Lett.* **33** (1974) 1404–1406.
- [26] **SLAC-SP-017** Collaboration, J. E. Augustin *et al.*, *Discovery of a Narrow Resonance in e^+e^- Annihilation*, *Phys. Rev. Lett.* **33** (1974) 1406–1408. [Adv. Exp. Phys.5,141(1976)].
- [27] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, *Phys. Rev. Lett.* **13** (1964) 138–140.
- [28] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.
- [29] S. W. Herb *et al.*, *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) 252–255.
- [30] **CDF** Collaboration, F. Abe *et al.*, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626–2631, [arXiv:hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002) [hep-ex].
- [31] M. L. Perl *et al.*, *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1489–1492.
- [32] **DONUT** Collaboration, K. Kodama *et al.*, *Observation of tau neutrino interactions*, *Phys. Lett.* **B504** (2001) 218–224, [arXiv:hep-ex/0012035](https://arxiv.org/abs/hep-ex/0012035) [hep-ex].
- [33] **ALEPH** Collaboration, D. Decamp *et al.*, *Determination of the Number of Light Neutrino Species*, *Phys. Lett.* **B231** (1989) 519–529.
- [34] **DELPHI** Collaboration, P. Aarnio *et al.*, *Measurement of the Mass and Width of the Z^0 Particle from Multi - Hadronic Final States Produced in e^+e^- Annihilations*, *Phys. Lett.* **B231** (1989) 539–547.
- [35] **OPAL** Collaboration, R. Akers *et al.*, *Measurement of single photon production in $e^+ e^-$ collisions near the Z^0 resonance*, *Z. Phys.* **C65** (1995) 47–66.

- [36] **L3** Collaboration, M. Acciarri *et al.*, *Determination of the number of light neutrino species from single photon production at LEP*, *Phys. Lett.* **B431** (1998) 199–208.
- [37] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, *Phys. Rev. Lett.* **51** (1983) 1945.
- [38] **CKMfitter Group** Collaboration, J. Charles, A. Hocker, H. Lacker, S. Laplace, F. R. Le Diberder, J. Malcles, J. Ocariz, M. Pivk, and L. Roos, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, *Eur. Phys. J.* **C41** no. 1, (2005) 1–131, arXiv:hep-ph/0406184 [hep-ph].
- [39] **Particle Data Group** Collaboration, C. Patrignani *et al.*, *Review of Particle Physics*, *Chin. Phys.* **C40** no. 10, (2016) 100001.
- [40] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martinez Santos, *Supersymmetry confronts $B_s \rightarrow \mu^+ \mu^-$: Present and future status*, *Phys. Rev.* **D87** no. 3, (2013) 035026, arXiv:1212.4887 [hep-ph].
- [41] U. Nierste, *Three Lectures on Meson Mixing and CKM phenomenology*, in *Heavy quark physics. Proceedings, Helmholtz International School, HQP08, Dubna, Russia, August 11-21, 2008*, pp. 1–38. 2009. arXiv:0904.1869 [hep-ph].
<https://inspirehep.net/record/817820/files/arXiv:0904.1869.pdf>.
- [42] 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. arXiv:hep-ph/9806471 [hep-ph].
<http://alice.cern.ch/format/showfull?sysnb=0282793>.
- [43] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, *Rev. Mod. Phys.* **68** (1996) 1125–1144, arXiv:hep-ph/9512380 [hep-ph].
- [44] A. J. Buras, R. Fleischer, J. Girrbach, and R. Knegjens, *Probing New Physics with the $B_s \rightarrow \mu^+ \mu^-$ Time-Dependent Rate*, *JHEP* **07** (2013) 77, arXiv:1303.3820 [hep-ph].

- [45] A. Dedes, J. Rosiek, and P. Tanedo, *Complete One-Loop MSSM Predictions for $B \rightarrow \text{lepton lepton}'$ at the Tevatron and LHC*, *Phys. Rev.* **D79** (2009) 055006, arXiv:0812.4320 [hep-ph].
- [46] G. Branco, L. Lavoura, and J. Silva, *CP violation*. Clarendon Press - Oxford, 1999.
- [47] Y. Amhis *et al.*, *Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016*, arXiv:1612.07233 [hep-ex].
- [48] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, and N. Tuning, *Branching Ratio Measurements of B_s Decays*, *Phys. Rev.* **D86** (2012) 014027, arXiv:1204.1735 [hep-ph].
- [49] **LHCb** Collaboration, *Tagged time-dependent angular analysis of $B_s^0 \rightarrow J/\psi\phi$ decays at LHCb*,.
- [50] C. Bobeth, T. Ewerth, F. Kruger, and J. Urban, *Analysis of neutral Higgs boson contributions to the decays $\bar{B}_s \rightarrow \ell^+\ell^-$ and $\bar{B} \rightarrow K\ell^+\ell^-$* , *Phys. Rev.* **D64** (2001) 074014, arXiv:hep-ph/0104284 [hep-ph].
- [51] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, *Eur. Phys. J.* **C74** (2014) 2890, arXiv:1310.8555 [hep-lat].
- [52] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for $BR(B_{s,d} \rightarrow \mu^+\mu^-)$* , *Eur. Phys. J.* **C72** (2012) 2172, arXiv:1208.0934 [hep-ph].
- [53] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou, and M. Steinhauser, *$B_{s,d} \rightarrow l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty*, *Phys. Rev. Lett.* **112** (2014) 101801, arXiv:1311.0903 [hep-ph].
- [54] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, *Eur. Phys. J.* **C45** (2006) 97–107, arXiv:hep-ph/0506026 [hep-ph].
- [55] C. Bobeth, M. Gorbahn, and E. Stamou, *Electroweak Corrections to $B_{s,d} \rightarrow \ell^+\ell^-$* , *Phys. Rev.* **D89** no. 3, (2014) 034023, arXiv:1311.1348 [hep-ph].
- [56] T. Hermann, M. Misiak, and M. Steinhauser, *Three-loop QCD corrections to $B_s \rightarrow \mu^+\mu^-$* , *JHEP* **12** (2013) 097, arXiv:1311.1347 [hep-ph].

- [57] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, and N. Tuning, *Probing New Physics via the $B_s^0 \rightarrow \mu^+ \mu^-$ Effective Lifetime*, *Phys. Rev. Lett.* **109** (2012) 041801, [arXiv:1204.1737 \[hep-ph\]](#).
- [58] **LHCb** Collaboration, R. Aaij *et al.*, *Resonances and CP violation in B_s^0 and $\bar{B}_s^0 \rightarrow J/\psi K^+ K^-$ decays in the mass region above the $\phi(1020)$* , [arXiv:1704.08217 \[hep-ex\]](#).
- [59] R. J. Knegjens, *Strategies to Hunt for New Physics with Strange Beauty Mesons*. PhD thesis, Vrije U., Amsterdam, 2014. <http://inspirehep.net/record/1287675/files/dissertation.pdf>.
- [60] **CMS** Collaboration, S. Chatrchyan *et al.*, *Measurement of the $B(s)$ to $\mu^+ \mu^-$ branching fraction and search for B^0 to $\mu^+ \mu^-$ with the CMS Experiment*, *Phys. Rev. Lett.* **111** (2013) 101804, [arXiv:1307.5025 \[hep-ex\]](#).
- [61] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment*, *Phys. Rev. Lett.* **111** (2013) 101805, [arXiv:1307.5024 \[hep-ex\]](#).
- [62] A. J. Buras, *Minimal flavor violation*, *Acta Phys. Polon.* **B34** (2003) 5615–5668, [arXiv:hep-ph/0310208 \[hep-ph\]](#).
- [63] G. D’Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, *Nucl. Phys.* **B645** (2002) 155–187, [arXiv:hep-ph/0207036 \[hep-ph\]](#).
- [64] A. J. Buras, P. Gambino, M. Gorbahn, S. Jager, and L. Silvestrini, *Universal unitarity triangle and physics beyond the standard model*, *Phys. Lett.* **B500** (2001) 161–167, [arXiv:hep-ph/0007085 \[hep-ph\]](#).
- [65] A. J. Buras, F. De Fazio, and J. Girrbach, *The Anatomy of Z' and Z with Flavour Changing Neutral Currents in the Flavour Precision Era*, *JHEP* **02** (2013) 116, [arXiv:1211.1896 \[hep-ph\]](#).
- [66] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370–3373, [arXiv:hep-ph/9905221 \[hep-ph\]](#).
- [67] M. Blanke, A. J. Buras, B. Duling, K. Gemmler, and S. Gori, *Rare K and B Decays in a Warped Extra Dimension with Custodial Protection*, *JHEP* **03** (2009) 108, [arXiv:0812.3803 \[hep-ph\]](#).

- [68] A. J. Buras, B. Duling, T. Feldmann, T. Heidsieck, C. Promberger, and S. Recksiegel, *Patterns of Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons*, *JHEP* **09** (2010) 106, [arXiv:1002.2126](https://arxiv.org/abs/1002.2126) [hep-ph].
- [69] D. M. Straub, *New physics correlations in rare decays*, in *CKM unitarity triangle. Proceedings, 6th International Workshop, CKM 2010, Warwick, UK, September 6-10, 2010*. 2010. [arXiv:1012.3893](https://arxiv.org/abs/1012.3893) [hep-ph]. <https://inspirehep.net/record/881964/files/arXiv:1012.3893.pdf>.
- [70] CDF Collaboration, *CDF public note 9892*.
- [71] A. J. Buras, F. De Fazio, J. Girrbach, R. Kneijens, and M. Nagai, *The Anatomy of Neutral Scalars with FCNCs in the Flavour Precision Era*, *JHEP* **06** (2013) 111, [arXiv:1303.3723](https://arxiv.org/abs/1303.3723) [hep-ph].
- [72] H. E. Logan and U. Nierste, *$B_{s,d} \rightarrow \ell^+ \ell^-$ in a two Higgs doublet model*, *Nucl. Phys.* **B586** (2000) 39–55, [arXiv:hep-ph/0004139](https://arxiv.org/abs/hep-ph/0004139) [hep-ph].
- [73] J. F. Gunion and H. E. Haber, *The CP conserving two Higgs doublet model: The Approach to the decoupling limit*, *Phys. Rev.* **D67** (2003) 075019, [arXiv:hep-ph/0207010](https://arxiv.org/abs/hep-ph/0207010) [hep-ph].
- [74] S. P. Martin, *A Supersymmetry primer*, [arXiv:hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356) [hep-ph]. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [75] K. S. Babu and C. F. Kolda, *Higgs mediated $B^0 \rightarrow \mu^+ \mu^-$ in minimal supersymmetry*, *Phys. Rev. Lett.* **84** (2000) 228–231, [arXiv:hep-ph/9909476](https://arxiv.org/abs/hep-ph/9909476) [hep-ph].
- [76] LHCb Collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \rightarrow \mu^+ \mu^-$* , *Phys. Rev. Lett.* **110** no. 2, (2013) 021801, [arXiv:1211.2674](https://arxiv.org/abs/1211.2674) [hep-ex].
- [77] W. Altmannshofer, M. Carena, N. R. Shah, and F. Yu, *Indirect Probes of the MSSM after the Higgs Discovery*, *JHEP* **01** (2013) 160, [arXiv:1211.1976](https://arxiv.org/abs/1211.1976) [hep-ph].
- [78] F. Archilli, *Observation of B_s^0 to $\mu^+ \mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data. Observation of B_s^0 to $\mu^+ \mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data*. Feb, 2017. <https://cds.cern.ch/record/2253469>.

- [79] **ATLAS** Collaboration, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, *Eur. Phys. J.* **C76** no. 9, (2016) 513, [arXiv:1604.04263 \[hep-ex\]](#).
- [80] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.
- [81] **LHCb** Collaboration, *LHCb public webpage*.
<http://lhcb-public.web.cern.ch/lhcb-public/>.
- [82] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.
- [83] **ATLAS** Collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [84] **CMS** Collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [85] **ALICE** Collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.
- [86] C. De Melis, *The CERN accelerator complex*. Jul, 2016.
<https://cds.cern.ch/record/2197559>. General Photo.
- [87] H. Burkhardt and P. Grafstrm, *Absolute Luminosity from Machine Parameters*,. <https://cds.cern.ch/record/1056691>.
- [88] C. Pralavorio, *The LHC racks up records*. Jun, 2017.
<http://cds.cern.ch/record/2272474>.
- [89] F. Follin and D. Jacquet, *Implementation and experience with luminosity levelling with offset beam*, in *Proceedings, ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders (BB2013): CERN, Geneva, Switzerland, March 18-22 2013*, pp. 183–187. 2014. [arXiv:1410.3667 \[physics.acc-ph\]](#). <https://inspirehep.net/record/1322139/files/arXiv:1410.3667.pdf>. [183(2014)].
- [90] **LHCb** Collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, *Int. J. Mod. Phys.* **A30** no. 07, (2015) 1530022, [arXiv:1412.6352 \[hep-ex\]](#).
- [91] **LHCb** Collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, *JINST* **9** no. 12, (2014) P12005, [arXiv:1410.0149 \[hep-ex\]](#).

- [92] **LHCb** Collaboration, *LHCb Operations Plots Webpage*.
<http://lhcb-operationsplots.web.cern.ch/lhcb-operationsplots/index.htm>.
- [93] S. D. Drell and T.-M. Yan, *Partons and their Applications at High-Energies*, *Annals Phys.* **66** (1971) 578. [Annals Phys.281,450(2000)].
- [94] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, *Rept. Prog. Phys.* **70** (2007) 89, arXiv:hep-ph/0611148 [hep-ph].
- [95] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].
- [96] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].
- [97] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298–318.
- [98] **TOTEM** Collaboration, G. Antchev *et al.*, *Luminosity-independent measurements of total, elastic and inelastic cross-sections at $\sqrt{s} = 7$ TeV*, *Europhys. Lett.* **101** (2013) 21004.
- [99] **ATLAS** Collaboration, M. Aaboud *et al.*, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** no. 18, (2016) 182002, arXiv:1606.02625 [hep-ex].
- [100] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*, *Eur. Phys. J.* **C17** (2000) 137–161, arXiv:hep-ph/0005110 [hep-ph].
- [101] B. L. Combridge, *Associated Production of Heavy Flavor States in $p p$ and $anti-p p$ Interactions: Some QCD Estimates*, *Nucl. Phys.* **B151** (1979) 429–456.
- [102] V. P. Andreev, *Heavy quark production at CMS and ATLAS*, *PoS BEAUTY2009* (2009) 023.
- [103] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of the b -quark production cross-section in 7 and 13 TeV pp collisions*, *Phys. Rev. Lett.* **118** no. 5, (2017) 052002, arXiv:1612.05140 [hep-ex].

- [104] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of b-hadron production fractions in 7 TeV pp collisions*, *Phys. Rev.* **D85** (2012) 032008, arXiv:1111.2357 [hep-ex].
- [105] **LHCb** Collaboration, B. Storaci, *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*,.
- [106] **BaBar** Collaboration, B. Aubert *et al.*, *The BaBar detector*, *Nucl. Instrum. Meth.* **A479** (2002) 1–116, arXiv:hep-ex/0105044 [hep-ex].
- [107] A. Abashian *et al.*, *The Belle Detector*, *Nucl. Instrum. Meth.* **A479** (2002) 117–232.
- [108] **BaBar** Collaboration, P. del Amo Sanchez *et al.*, *Search for $b \rightarrow u$ transitions in $B^- \rightarrow DK^-$ and $D^* K^-$ Decays*, *Phys. Rev.* **D82** (2010) 072006, arXiv:1006.4241 [hep-ex].
- [109] **LHCb** Collaboration, R. Aaij *et al.*, *Observation of CP violation in $B^\pm \rightarrow DK^\pm$ decays*, *Phys. Lett.* **B712** (2012) 203–212, arXiv:1203.3662 [hep-ex]. [Erratum: *Phys. Lett.* B713,351(2012)].
- [110] **Belle-II** Collaboration, T. Abe *et al.*, *Belle II Technical Design Report*, arXiv:1011.0352 [physics.ins-det].
- [111] **LHCb** Collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*,. <http://cds.cern.ch/record/630827>.
- [112] **LHCb** Collaboration, *LHCb magnet: Technical Design Report*,. <https://cds.cern.ch/record/424338>.
- [113] R. L. Gluckstern, *Uncertainties in track momentum and direction, due to multiple scattering and measurement errors*, *Nucl. Instrum. Meth.* **24** (1963) 381–389.
- [114] **LHCb** Collaboration, *LHCb VELO (VERtex LOcator): Technical Design Report*,. <http://cds.cern.ch/record/504321>.
- [115] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, *JINST* **9** (2014) 09007, arXiv:1405.7808 [physics.ins-det].
- [116] **LHCb** Collaboration, *LHCb inner tracker: Technical Design Report*,. <http://cds.cern.ch/record/582793>. revised version number 1 submitted on 2002-11-13 14:14:34.

- [117] **LHCb** Collaboration, *LHCb outer tracker: Technical Design Report*,
<https://cds.cern.ch/record/519146>.
- [118] **LHCb** Collaboration, *LHCb RICH: Technical Design Report*,
<http://cds.cern.ch/record/494263>.
- [119] **LHCb RICH Group** Collaboration, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, *Eur. Phys. J.* **C73** (2013) 2431, arXiv:1211.6759 [physics.ins-det].
- [120] **LHCb RICH** Collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, arXiv:1703.08152 [physics.ins-det].
- [121] **LHCb** Collaboration, *LHCb calorimeters: Technical Design Report*,
<http://cds.cern.ch/record/494264>.
- [122] **LHCb** Collaboration, *LHCb muon system: Technical Design Report*,
<http://cds.cern.ch/record/504326>.
- [123] **LHCb** Collaboration, *LHCb muon system: addendum to the Technical Design Report*,
<http://cds.cern.ch/record/600536>.
- [124] **LHCb** Collaboration, *LHCb muon system: second addendum to the Technical Design Report*,
<http://cds.cern.ch/record/831955>.
 Submitted on 9 Apr 2005.
- [125] F. Sauli, *Principles of operation of multiwire proportional and drift chambers*, p. 92 p, CERN. CERN, Geneva, 1977.
<http://cds.cern.ch/record/117989>. CERN, Geneva, 1975 - 1976.
- [126] W. Bonivento, G. Bencivenni, G. Felici, F. Murtas, P. Valente, A. Cardini, A. Lai, D. Pinci, B. Saitta, and C. Bosio, *A triple GEM detector with pad readout for high rate charged particle triggering*, *Nucl. Instrum. Methods Phys. Res., A* **488** no. 1-2, (2002) 493–502.
<http://cds.cern.ch/record/592972>.
- [127] F. Sauli, *GEM: A new concept for electron amplification in gas detectors*, *Nucl. Instrum. Meth.* **A386** (1997) 531–534.
- [128] D. Pinci and M. Santimaria, *Analysis of the performance of GEM chambers for the upgrade of the LHCb muon system*, Tech. Rep. LHCb-PUB-2015-014. CERN-LHCb-PUB-2015-014, CERN, Geneva, May, 2015. <https://cds.cern.ch/record/2018404>.

- [129] M. Santimaria and D. Pinci, *Study of the GEM detector for the LHCb experiment upgrade*, Nov, 2013. <https://cds.cern.ch/record/1690550>. Presented 19 Dec 2013.
- [130] **LHCb** Collaboration, *LHCb trigger system: Technical Design Report*,. <http://cds.cern.ch/record/630828>. revised version number 1 submitted on 2003-09-24 12:12:22.
- [131] **LHCb** Collaboration, *Nu, Mu and Pile-Up. The LHCb definitions of what we see and what we don't see*. <https://twiki.cern.ch/twiki/bin/view/LHCb/NuMuPileUp>.
- [132] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, *JINST* **8** (2013) P04022, arXiv:1211.3055 [hep-ex].
- [133] **LHCb** Collaboration, *LHCb Tracking Strategies*. <https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingStrategies>.
- [134] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth.* **A262** (1987) 444–450.
- [135] T. Likhomanenko, P. Ilten, E. Khairullin, A. Rogozhnikov, A. Ustyuzhanin, and M. Williams, *LHCb Topological Trigger Reoptimization*, *J. Phys. Conf. Ser.* **664** no. 8, (2015) 082025, arXiv:1510.00572 [physics.ins-det].
- [136] O. Lupton, *Studies of $D^0 \rightarrow K_S^0 h^+ h'^-$ decays at the LHCb experiment*. PhD thesis, Oxford U., 2016-07-11. <http://inspirehep.net/record/1503518/files/CERN-THESIS-2016-156.pdf>.
- [137] M. Frank, C. Gaspar, E. v Herwijnen, B. Jost, and N. Neufeld, *Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization*, *Journal of Physics: Conference Series* **513** no. 1, (2014) 012006. <http://stacks.iop.org/1742-6596/513/i=1/a=012006>.
- [138] G. Dujany and B. Storaci, *Real-time alignment and calibration of the LHCb Detector in Run II*, *J. Phys. Conf. Ser.* **664** no. 8, (2015) 082010.
- [139] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014. <https://cds.cern.ch/record/1701134>.

- [140] R. Aaij *et al.*, *Tesla : an application for real-time data analysis in High Energy Physics*, *Comput. Phys. Commun.* **208** (2016) 35–42, arXiv:1604.05596 [physics.ins-det].
- [141] *The Worldwide LHC Computing Grid*.
<http://wlcg-public.web.cern.ch/>.
- [142] LHCb Collaboration, *The GAUSS project*,. <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/>.
- [143] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152–155.
- [144] GEANT4 Collaboration, S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250–303.
- [145] A. Roodman, *Blind analysis in particle physics*, *eConf C030908* (2003) TUIT001, arXiv:physics/0312102 [physics.data-an]. [,166(2003)].
- [146] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, *Nucl. Instrum. Meth.* **A555** (2005) 356–369, arXiv:physics/0402083 [physics.data-an].
- [147] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, *JINST* **8** (2013) P10020, arXiv:1306.0249 [physics.ins-det].
- [148] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez, E. Polcarpo, and A. Sarti, *The Muon Identification Procedure of the LHCb Experiment for the First Data*,.
- [149] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, *PoS ACAT* (2007) 040, arXiv:physics/0703039 [PHYSICS].
- [150] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, *Nucl. Instrum. Meth.* **A389** (1997) 81–86.
- [151] O. Lupton, L. Anderlini, B. Sciascia, and V. Gligorov, *Calibration samples for particle identification at LHCb in Run 2*, Tech. Rep. LHCb-PUB-2016-005. CERN-LHCb-PUB-2016-005, CERN, Geneva, Mar, 2016. <http://cds.cern.ch/record/2134057>.
- [152] L. Anderlini, A. Contu, C. R. Jones, S. S. Malde, D. Muller, S. Ogilvy, J. M. Otalora Goicochea, A. Pearce, I. Polyakov, W. Qian, B. Sciascia, R. Vazquez Gomez, and Y. Zhang, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016. <https://cds.cern.ch/record/2202412>.

- [153] **LHCb** Collaboration, *PIDCalib Package*,
<https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackages>.
- [154] **LHCb** Collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$* , *Phys. Lett.* **B708** (2012) 55–67, [arXiv:1112.1600](https://arxiv.org/abs/1112.1600) [hep-ex].
- [155] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.
- [156] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, *Journal of Computer and System Sciences* **55** no. 1, (1997) 119 – 139. <http://www.sciencedirect.com/science/article/pii/S002200009791504X>.
- [157] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*. PhD thesis, Cracow, INP, 1986. http://lss.fnal.gov/cgi-bin/find_paper.pl?other/thesis/skwarnicki.pdf.
- [158] **Particle Data Group** Collaboration, K. A. Olive *et al.*, *Review of Particle Physics*, *Chin. Phys.* **C38** (2014) 090001.
- [159] C. Elsasser, *The Rare Decays $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and Z Boson Production at LHCb and Radiation Damage in its Silicon Tracker*. PhD thesis, Zurich U., 2015-02-03. <http://inspirehep.net/record/1381389/files/CERN-THESIS-2015-020.pdf>.
- [160] D. Martinez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, *Nucl. Instrum. Meth.* **A764** (2014) 150–155, [arXiv:1312.5000](https://arxiv.org/abs/1312.5000) [hep-ex].
- [161] W. Verkerke and D. P. Kirkby, *RooFit Users Manual v2.91*, https://root.cern.ch/download/doc/RooFit_Users_Manual_2.91-33.pdf.
- [162] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, *eConf C0303241* (2003) MOLT007, [arXiv:physics/0306116](https://arxiv.org/abs/physics/0306116) [physics]. [,186(2003)].
- [163] **LHCb** Collaboration, R. Aaij *et al.*, *Measurements of the branching fractions and CP asymmetries of $B^\pm \rightarrow J/\psi \pi^\pm$ and $B^\pm \rightarrow \psi(2S)\pi^\pm$ decays*, *Phys. Rev.* **D85** (2012) 091105, [arXiv:1203.3592](https://arxiv.org/abs/1203.3592) [hep-ex].

- [164] **Belle** Collaboration, F. Thorne *et al.*, *Measurement of the decays $B_s^0 \rightarrow J/\psi\phi(1020)$, $B_s^0 \rightarrow J/\psi f_2'(1525)$ and $B_s^0 \rightarrow J/\psi K^+ K^-$ at Belle*, *Phys. Rev.* **D88** no. 11, (2013) 114006, [arXiv:1309.0704](#) [hep-ex].
- [165] **Particle Data Group** Collaboration, J. Beringer *et al.*, *Review of Particle Physics (RPP)*, *Phys. Rev.* **D86** (2012) 010001.
- [166] J. M. Flynn, T. Izubuchi, T. Kawanai, C. Lehner, A. Soni, R. S. Van de Water, and O. Witzel, *$B \rightarrow \pi\ell\nu$ and $B_s \rightarrow K\ell\nu$ form factors and $|V_{ub}|$ from 2+1-flavor lattice QCD with domain-wall light quarks and relativistic heavy quarks*, *Phys. Rev.* **D91** no. 7, (2015) 074510, [arXiv:1501.05373](#) [hep-lat].
- [167] Marcello Rotondo, *Private communication*.
- [168] **ARGUS** Collaboration, H. Albrecht *et al.*, *Search for Hadronic $b \rightarrow u$ Decays*, *Phys. Lett.* **B241** (1990) 278–282.
- [169] C. M. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, *$B_s \rightarrow K\ell\nu$ form factors from lattice QCD*, *Phys. Rev.* **D90** (2014) 054506, [arXiv:1406.2279](#) [hep-lat].
- [170] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, *Eur. Phys. J.* **C77** no. 2, (2017) 112, [arXiv:1607.00299](#) [hep-lat].
- [171] **LHCb** Collaboration, R. Aaij *et al.*, *First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \rightarrow \pi^\pm \mu^+ \mu^-$ decay*, *JHEP* **10** (2015) 034, [arXiv:1509.00414](#) [hep-ex].
- [172] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \rightarrow (\pi, K)(\ell^+ \ell^-, \ell\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, *Phys. Rev.* **D86** (2012) 114025, [arXiv:1207.0265](#) [hep-ph].
- [173] **LHCb** Collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, *Nature Phys.* **11** (2015) 743–747, [arXiv:1504.01568](#) [hep-ex].
- [174] **LHCb** Collaboration, R. Aaij *et al.*, *Measurements of B_c^+ production and mass with the $B_c^+ \rightarrow J/\psi\pi^+$ decay*, *Phys. Rev. Lett.* **109** (2012) 232001, [arXiv:1209.5634](#) [hep-ex].

- [175] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of the ratio of B_c^+ branching fractions to $J/\psi\pi^+$ and $J/\psi\mu^+\nu_\mu$* , *Phys. Rev.* **D90** no. 3, (2014) 032009, [arXiv:1407.2126](https://arxiv.org/abs/1407.2126) [[hep-ex](#)].
- [176] S. Tolk, *Discovery of Rare B Decays*. PhD thesis, Groningen U., 2016.
<https://inspirehep.net/record/1491544/files/CERN-THESIS-2015-331.pdf>.
- [177] G. Cowan, *Statistical Data Analysis*. Clarendon Press, 1998.
- [178] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Ann. Math. Statist.* **9** no. 1, (03, 1938) 60–62. <http://dx.doi.org/10.1214/aoms/1177732360>.
- [179] G. Cowan, *Goodness of fit and Wilks' theorem*,. https://wwwusers.ts.infn.it/~milotti/Didattica/StatisticaAvanzata/Cowan_2013.pdf.
- [180] F. James, *The interpretation of errors in MINUIT*,.
<http://seal.web.cern.ch/seal/documents/minuit/mnerror.pdf>.
- [181] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343–367.
- [182] A. L. Read, *Modified frequentist analysis of search results (The CL(s) method)*, in *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*, pp. 81–101. 2000.
<http://weplib.cern.ch/abstract?CERN-OPEN-2000-205>.
- [183] G. Zech, *Upper Limits in Experiments with Background Or Measurement Errors*, *Nucl. Instrum. Meth.* **A277** (1989) 608.
- [184] **Heavy Flavor Averaging Group** Collaboration, Y. Amhis *et al.*, *Averages of B-Hadron, C-Hadron, and tau-lepton properties as of early 2012*, [arXiv:1207.1158](https://arxiv.org/abs/1207.1158) [[hep-ex](#)].
- [185] R. Fleischer, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi, *In Pursuit of New Physics with $B_{s,d}^0 \rightarrow \ell^+\ell^-$* , *JHEP* **05** (2017) 156, [arXiv:1703.10160](https://arxiv.org/abs/1703.10160) [[hep-ph](#)].
- [186] W. Altmannshofer, C. Niehoff, and D. M. Straub, *$B_s \rightarrow \mu^+\mu^-$ as current and future probe of new physics*, *JHEP* **05** (2017) 076, [arXiv:1702.05498](https://arxiv.org/abs/1702.05498) [[hep-ph](#)].

- [187] **USQCD** Collaboration, T. Bloom *et al.*, *Lattice QCD at the intensity frontier*,. <https://www.usqcd.org/documents/13flavor.pdf>.
- [188] **LHCb** Collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays*, *Phys. Rev. Lett.* **118** no. 19, (2017) 191801, arXiv:1703.05747 [hep-ex].

