

Discovery of rare B decays

First observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay,
first evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay



This work is part of the research program of the Stichting voor Fundamenteel onderzoek der Materie (FOM), which is part of the Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO). It was carried out at the Nationaal Instituut voor Subatomaire Fysica (Nikhef) in Amsterdam, the Netherlands.

Copyright © 2016 by Siim Tolk
Printed in the Netherlands by Gildeprint Printing

ISBN: 978-90-367-8722-2 Printed version
ISBN: 978-90-367-8721-5 Electronic version



/ university of
groningen

Discovery of rare B decays

First observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay,
first evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 8 April 2016 at 14.30 hours

by

Siim Tolk

born on 16 March 1987
in Tallinn, Estonia

Supervisor

Prof. dr. A. Pellegrino

Co-supervisor

Dr. F. Dettori

Assessment Committee

Prof. dr. K.H.K.J. Jungmann

Prof. dr. P.J. de Jong

Prof. dr. A.N.J.J. Schellekens

Propositions

1. The Standard Model is not a complete theory of particle physics. There are fundamental physical phenomena in nature that the Standard Model can not explain and a complete theory must involve physics explanations beyond it. The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays can be used to study the possible Standard Model extensions.
2. $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are suppressed in the Standard Model. The decays are theoretically clean and have precise branching fractions predictions in the Standard Model.
3. The LHCb detector at the Large Hadron Collider in CERN is well suited for detecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays.
4. The measured $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions were normalised to the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$ channels. Normalising significantly improves the precision of the measured branching fractions.
5. In LHCb, the trigger efficiency can be estimated from the data with the TISTOS method. The bias of the method can be reduced by dividing the candidates into sub-samples according to the momentum of the B meson, applying the method in each sub-sample, and combining the results.
6. $B_s^0 \rightarrow \mu^+ \mu^-$ decays are experimentally confirmed. Albeit subject to large uncertainties, the measured $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction in the combined CMS and LHCb analysis is in agreement with the Standard Model prediction.
7. The combined CMS and LHCb data show evidence of $B^0 \rightarrow \mu^+ \mu^-$ decays. The detected number of $B^0 \rightarrow \mu^+ \mu^-$ candidates is higher than expected in the Standard Model. The uncertainties of the measurement are large and the measured $B^0 \rightarrow \mu^+ \mu^-$ branching fraction is compatible with the Standard Model prediction.

8. Those New Physics scenarios where the branching fraction predictions for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are larger than predicted in the Standard Model are disfavoured by the data. More precise measurements are necessary to study the viability of New Physics scenarios with the branching fraction predictions close to the Standard Model.
9. Studying what the universe is made of is as expensive as playing football. We should do both.

Contents

I Overview of theory and LHCb experiment	v
1 Physics of the elementary particles, rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$	1
1.1 Physics of elementary particles	1
1.1.1 The Standard Model	3
1.1.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in the Standard Model	7
1.2 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction definition	9
1.3 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction computation	11
1.3.1 The effective approach	13
1.3.2 The non-perturbative matrix element	14
1.3.3 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction	16
1.3.4 The Standard Model $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction .	17
1.3.5 Radiative corrections	18
1.4 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in New Physics models	19
1.4.1 Pseudo-scalar dominated New Physics ($S = 0$)	20
1.4.2 Scalar dominated New Physics ($P = 1$)	22
1.4.3 Mixed (pseudo-)scalar New Physics ($P \pm S = 1$)	22
2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ production and detection: The Large Hadron Collider and the LHCb detector	25
2.1 The Large Hadron Collider (LHC)	26
2.1.1 B meson production	28
2.2 The LHCb detector and tracking	29
2.2.1 The Outer Tracker (OT)	34
2.2.2 Cherenkov detectors (RICH1 and RICH2)	34
2.2.3 The Muon system (M1-M5)	35
2.2.4 The Calorimeter system (CALO)	37
2.3 Overview of the LHCb trigger	38
2.3.1 Level-0 trigger	39

2.3.2	The High Level Trigger	40
2.4	Offline reconstruction	43
2.5	Particle Identification	44
2.5.1	Muon identification	44
2.5.2	Muon identification performance	47
2.5.3	Kaon and pion identification	48
II	LHCb $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis	51
3	Trigger in $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis	57
3.1	Level 0	57
3.1.1	$B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$	58
3.1.2	$B_{(s)}^0 \rightarrow h^+h^-$	59
3.2	HLT1	59
3.2.1	$B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$	62
3.2.2	$B_{(s)}^0 \rightarrow h^+h^-$	62
3.3	HLT2	62
3.3.1	$B_{(s)}^0 \rightarrow \mu^+\mu^-$	64
3.3.2	$B^+ \rightarrow J/\psi K^+$	64
3.3.3	$B_{(s)}^0 \rightarrow h^+h^-$	66
3.4	Global event cuts	67
4	Signal selection	69
4.1	Selection based on cuts	71
4.2	Multivariate operators	73
4.2.1	Decision Trees and Boosting	74
4.3	Multivariate selection: BDTS	75
4.4	Multivariate classification: BDT	76
4.4.1	BDT binning, training, and performance	83
5	Data-driven trigger efficiency estimation (TISTOS method)	85
5.1	TISTOS trigger efficiency	85
5.1.1	Estimating the trigger efficiency uncertainty	89
5.1.2	Binning the phase space	90
5.2	Performance on simulation	91
5.2.1	Signal separation	91
5.2.2	Results with different binning schemes	92
5.3	Performance on data	96

6 Signal yield extraction	99
6.1 The di-muon mass model	99
6.2 Signal BDT distribution	101
6.2.1 From the $B_{(s)}^0 \rightarrow h^+h^-$ BDT to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT	104
6.3 Signal invariant mass distribution	105
6.3.1 B_s^0 and B^0 central mass value	106
6.3.2 B_s^0 and B^0 mass resolution	108
6.4 Doubly mis-identified background model	109
6.4.1 Mass distribution	113
6.5 Other background sources	113
6.5.1 BDT and mass distribution	117
6.6 Combinatorial background BDT and mass distributions	117
7 Signal branching fraction normalisation	119
7.1 Normalisation channels	120
7.2 Normalisation channel yields	121
7.2.1 $B^+ \rightarrow J/\psi K^+$ mass model	122
7.2.2 $B^0 \rightarrow K^+\pi^-$ mass model	122
7.3 Detection efficiencies	125
7.3.1 Geometrical detector acceptance	126
7.3.2 Reconstruction and selection efficiencies	126
7.3.3 Trigger efficiencies	131
7.4 Hadronisation factors	133
7.5 Normalisation factors	134
8 Results of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis	137
8.1 The un-binned maximum likelihood fit	137
8.1.1 Nuisance parameters	139
8.2 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions	140
8.3 The significances of the measured branching fractions	141
8.4 The exclusion limit from CL_s	145
III Combined CMS and LHCb analysis	149
9 Overview of the CMS $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis	153
9.1 CMS detector	153
9.2 Trigger	155
9.3 Selection	156
9.4 Likelihood model	157
9.5 Normalisation and results	159

10 Combined CMS and LHCb analysis	161
10.1 Changes to the published analysis models	162
10.2 Shared parameters and combined model.	163
10.3 Results of the combined analysis	164
10.3.1 Branching fractions	166
10.4 Comparison to theory	169
11 Summary and outlook	177
12 Samenvatting en Vooruitzicht	185
13 Kokkuvõte	193

Introduction

The decays of B_s^0 and B^0 mesons to two muons are the topic of this dissertation. The rates of these decays are precisely predicted and extremely small in the Standard Model. As these rates may hold the key to yet unseen physics effects, they have been searched for with great enthusiasm by many experiments. This dissertation describes the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay and the first evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay, bringing to an end the experimental searches lasting for more than three decades.

Often, our understanding of the world improves when an unexpected experimental result contradicts the established theory and prefers another explanation [1]. The “established” theory of elementary particles, the “Standard Model” (SM), is the culmination of our understanding of the nature of matter. It combines the insights from measurements and theory alike. The latest addition to the SM was the Brout-Englert-Higgs [2,3] mechanism, which describes how the elementary particles acquire masses and predicts another elementary particle called Higgs boson; the Higgs boson was recently experimentally discovered by the ATLAS and CMS collaborations [4,5].

The aim of modern particle physics is to understand those experimental observations that are not explained by the SM. Arguably, the most striking of these are: the nature of matter and energy, most notably those that contribute the most to the mass of the universe: the Dark Matter and Dark Energy; the large imbalance between matter and anti-matter in the universe; and the nature of the non-zero neutrino masses implied by the mass differences required to explain neutrino oscillations. Several theoretical models, collectively referred to as “New Physics” (NP), have been proposed to explain the SM shortcomings, and particle physicists try to validate them either by a *top-down* (i.e. deductive) or by a *bottom-up* (i.e. inductive) approach.

In the top-down approach, specific New Physics models, such as Supersymmetry (SUSY), are studied to find processes with distinct and measurable deviations from the SM. In the bottom-up approach, no specific model is chosen, but instead, the possible deviations are classified in a general way,

and, if observed, are used to look for patterns that could match the predictions of a new physics model, or lead to the construction of a new model.

The experimental particle physicist studies how subatomic processes proceed in nature. The known “particle zoo” is often extended in new physics with new unobserved particles; if the new particles exist, they would be produced in high-energy collisions, decay, and leave traces in particle detectors. This is how the Higgs boson was discovered. Because the mass of the heaviest particle that can be produced is determined by the collision energy, these so-called “direct” searches form the *high-energy frontier* of experimental particle physics. Alternatively, new physics could be looked for indirectly, that is, by studying the new particle contributions to “quantum loop” processes. Since these “indirect” searches are performed on well understood processes with precise SM predictions, they form the *precision frontier* of experimental particle physics. Indirect searches are sensitive to particle masses that are orders of magnitude higher than those in direct searches.

The search for the rare $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays is a good example of an indirect search: the well-understood and modest SM decay rates are very precisely predicted, in the order of one $B_s^0 \rightarrow \mu^+ \mu^-$ decay every 250 million B_s^0 mesons, and much less for $B^0 \rightarrow \mu^+ \mu^-$. With small expected SM rates and loop processes dominating the transition, it is likely that New Physics, especially models with extended Higgs sectors or additional bosons, can significantly alter the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay probability. This is the reason why these decays have been actively searched for by many experiments over the last three decades. The experimentally clear di-muon signature and the abundant B meson production at the Large Hadron Collider, make it finally possible to search for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in a wide branching-fraction range, down to the minuscule SM predictions.

Outline

This dissertation describes the search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays in three parts: the first part describes the theoretical framework and the LHCb experiment; the second part presents the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis of the data samples collected in 2011 and 2012; the third part presents the $B_s^0 \rightarrow \mu^+ \mu^-$ discovery, and the first $B^0 \rightarrow \mu^+ \mu^-$ evidence in the combined CMS and LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.

Part I: Overview of theory and the LHCb experiment

Chapter 1 introduces the theory of elementary particle physics. The mechanisms affecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are reviewed in more detail, and the

theoretical $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction computation is presented in a general, model-independent way. **Chapter 2** focusses on the experimental side of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ searches; the chapter gives an overview of the B_s^0 and B^0 meson production in the Large Hadron Collider (LHC) at CERN, and of the design of the LHCb detector. The LHCb trigger system and particle identification strategies are discussed in detail.

Part II: LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

Chapter 3 reviews the LHCb trigger strategies used to select the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, and the $B_{(s)}^0 \rightarrow h^+ h^-$, $B^+ \rightarrow J/\psi K^+$, and $B^0 \rightarrow K^+ \pi^-$ decays, that are used as control channels in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. **Chapter 4** describes the signal selection. A correct estimate of the trigger efficiency is crucial to normalise the final result, and in **Chapter 5**, a method is presented to estimate trigger efficiencies from data. **Chapter 6** discusses the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yield extraction from the di-muon invariant mass distribution. Possible mis-identified background sources are considered, and, if significant, included in the model. The signal yields are normalised with respect to the normalisation channel ($B^0 \rightarrow K^+ \pi^-$ and $B^+ \rightarrow J/\psi K^+$) yields, to calculate the signal branching fraction. **Chapter 7** discusses the normalisation procedure. Part II of this thesis ends with **Chapter 8**, which presents the results of the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. The chapter also describes the statistical aspects of the determination of the significance of the measured $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction.

Part III: Combined CMS and LHCb analysis

Chapter 9 is devoted to the CMS detector and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ searches. The differences with respect to the LHCb analysis are emphasised. **Chapter 10** discusses the combination of the CMS and LHCb analyses. The combined $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements are presented, together with a measurement of the ratio between the two. Finally, the results are discussed, and compared with SM and NP predictions.

Part I

Overview of theory and LHCb experiment

Chapter 1

Physics of the elementary particles, rare decays

$$B_s^0 \rightarrow \mu^+ \mu^- \text{ and } B^0 \rightarrow \mu^+ \mu^-$$

This chapter sets the context for the rest of the thesis. The general discussion relies on references [1, 6–11]. It starts with introducing the domain of elementary particles as it is known from the experimental evidence in Sec. 1.1. The physical model built to explain the observed elementary particles and their qualities, the Standard Model, is presented in the following section, Sec. 1.1.1.

Section 1.1.2 discusses the main topic of this thesis: neutral B meson decays to muons, $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. As will be shown, many of the key mechanisms of the Standard Model are expected to play an important role in these decays and their decay probabilities are expected to be tiny. The formal definition of the decay probability, the branching fraction, is given in Sec. 1.2. Their computation in a general and model independent manner is presented in Sec. 1.3.

Because of the small Standard Model rates, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are sensitive probes for models beyond the Standard Model. The sensitive elements in the branching fraction and the relevant SM extensions are discussed in Sec. 1.4.

1.1 Physics of elementary particles

The interactions of matter define the dynamics of our universe. According to their strength, physical range, and the forms of matter involved, four types of interactions can be distinguished: gravitational, electromagnetic, weak,

and strong. The last three have been observed to affect solely what we call “visible” matter. Gravity on the other hand affects all the matter forms; according to astronomical observations, visible matter only amounts to 5% of the mass of our universe, while the remaining 95% is distributed between the yet unknown Dark Energy (69%) and Dark Matter (26%) [12].

Present experimental observations allow to identify 24 elementary units of visible matter: 12 fermions, and for each fermion, a corresponding anti-fermion with equal mass but opposite electric charge. The twelve fermions are identified by their “flavour” and grouped according to which interactions they partake into six quarks, that interact strongly:

$$\text{up } (u), \quad \text{charmed } (c), \quad \text{top } (t), \quad (1.1)$$

$$\text{down } (d), \quad \text{strange } (s), \quad \text{bottom } (b), \quad (1.2)$$

and six leptons, that do not:

$$e\text{-neutrino } (\nu_e), \quad \mu\text{-neutrino } (\nu_\mu), \quad \tau\text{-neutrino } (\nu_\tau). \quad (1.3)$$

$$\text{electron } (e), \quad \text{muon } (\mu), \quad \text{tau } (\tau), \quad (1.4)$$

The electromagnetic force affects particles with an electric charge, and the quarks can be further divided into “up” types carrying $+2/3$ of the electron charge (e), Eq. (1.1), and “down” types carrying $-(1/3)e$, Eq. (1.2). Electron, muon, and tau leptons are negatively charged ($-1e$), whereas neutrinos are neutral and do not interact electromagnetically. Contrary to leptons, quarks have not been observed as free particles. The simplest observed quark combinations are particles consisting of quark anti-quark pairs, called “mesons”.

Weak force is a general term for the interactions, which, among other processes, are responsible for transforming a d quark to an u quark in nuclear β -decays. Experiments with neutrino beams have shown that the neutrinos can also scatter from the electrons and nuclei [13]. The weak interactions, if responsible for the scattering, must thus also have a neutral component in addition to the charged interactions mediating nuclear β decays. Weak force is unique in many respects:

- it affects all the twelve fermions, including neutrinos, e.g. in neutrino scattering $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$;
- it can transform a down-type quark into an up-type quark (and the other way around), e.g. in a nuclear β -decay $n(udd)\nu_e \rightarrow p(udu) + e^-$;
- it can transform a charged lepton into a neutrino (and the other way around), e.g. in a nuclear β -decay;

- it has been observed to be different for a specular version of the same process [14], e.g. for decays of aligned cobalt-60 atoms $\text{Co}_{27}^{60} \rightarrow \text{Ni}_{28}^{60} + e^- + \bar{\nu}_e + 2\gamma$.

Studies on the relative strength of the weak interactions between different quark or lepton pairs has revealed an underlying structure among the 12 elementary particles, according to which the elementary particles can be grouped in three pairs of quarks: (^u_d) , (^c_s) , (^t_b) ; and three pairs of leptons: (ν_e) , (ν_μ) , (ν_τ) . The pairs are called “generations” or “families” [15].

1.1.1 The Standard Model

At the fundamental level, the observations on the visible matter interactions and its elementary constituents, can be described within the framework of Quantum Field Theory (QFT) by a model called the “Standard Model” (SM) [16–19].

In QFT, the observed elementary particles are seen as excitations of the corresponding fields. The 12 fermion fields form the cornerstones of the Standard Model. The Standard Model does not attempt to explain the total number of fermion flavours, their underlying structure, or explain their properties. The 12 fermions are taken as elementary and are required to satisfy the Dirac equation, which offers a compact way of describing the behaviour of massive spin- $\frac{1}{2}$ particles together with their anti-particles¹.

In QFT, a force between particles is seen as an exchange of a (virtual) mediator. The mediating particles are called bosons, and unlike fermions, they carry an integer spin. Boson are excitations of the relevant gauge fields that must be included in the model in order to preserve its invariance under local phase transformations. The gauge fields and their interactions with fermion fields can thus be said to arise naturally, on the ground of symmetry requirements imposed on the model.

The Standard Model is required to be invariant under specific local phase transformations of type:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \quad (1.5)$$

This is guaranteed by introducing gauge fields, associated with the force mediators: the photon for the electromagnetic, the W^\pm and Z^0 for the weak, and the gluons for the strong force. A description of gravity is not included in the Standard Model. The gluon gauge fields leave the model invariant

¹In the Dirac equation the anti-particles are seen as “reflections” of the matter fermions with equal mass, but opposite internal quantum numbers (charge, lepton number, baryon number, etc.).

under the $SU(3)_C$ local phase transformations, where C stands for the “colour” charge of the strong force. Quarks carry either “red”, “green”, or “blue” colour charge, and anti-quarks the opposite colour charge. The theory of quarks and gluons is known as Quantum Chromodynamics (QCD) [19], and it postulates that only colourless states called hadrons are allowed to exist in nature. In QCD, a single quark or anti-quark cannot exist by itself, but forms a bound colour neutral state in a process called hadronisation. Invariance under local $SU(2)_L \times U(1)_Y$ phase transformations requires introduction of four new gauge fields; these fields mix and give rise to two bosons of charged weak interaction (W^\pm), one boson of neutral weak interaction (Z^0), and one boson of electromagnetic force (γ).

The unique parity violating feature of the weak interaction needs special treatment in the model. The fermion fields are decomposed into right- and left-handed chiral components, e.g. for an up type quark:

$$u_L \equiv \frac{1}{2}(1 - \gamma^5)u = \mathcal{P}_L u, \quad u_R \equiv \frac{1}{2}(1 + \gamma^5)u = \mathcal{P}_R u, \quad (1.6)$$

where γ^5 is a product of the four Dirac matrices ($\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$), and \mathcal{P}_L and \mathcal{P}_R are the left- and right-handed chiral projection operators. The observed parity violation of the weak force is modelled by only allowing the weak bosons to interact with the left-handed chiral components of the fermion fields, that is, requiring only the left handed chiral components of the fermion fields to be invariant under local phase transformations $SU(2)_L$. According to the strength of the weak interaction, the left-handed chiral components can be grouped in three generations of quark and lepton doublets:

$$\mathcal{Q}_{L,i} \equiv \begin{pmatrix} u_L \\ d_L \end{pmatrix}_{i=1}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}_{i=2}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}_{i=3}, \quad (1.7)$$

$$L_{L,i} \equiv \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}_{i=1}, \begin{pmatrix} \nu_\mu \\ \mu_L \end{pmatrix}_{i=2}, \begin{pmatrix} \nu_\tau \\ \tau_L \end{pmatrix}_{i=3}, \quad (1.8)$$

where i is the generation index. The corresponding right handed fields do not couple to the charged weak bosons (W^\pm) and are described as singlets $u_{R,i}$ and $d_{R,i}$, denoting right-handed up- and down-type quarks, and $\nu_{R,i}$ and $l_{R,i}$, denoting right-handed neutral and charged leptons². Left-handed doublets and right-handed singlets are both required to remain invariant under $U(1)_Y$

²The right-handed neutrinos do not participate in any of the Standard Model interactions. Although not observed, they can be used to explain the non-zero neutrino masses and are included here for completeness.

phase transformations, where Y denotes the weak hypercharge³.

Particle masses in the Standard Model

Invariance under the local phase transformation Eq. (1.5) is an elegant way to describe the observed forces, yet it only works for the unphysical case of massless gauge bosons and fermions. Boson mass terms are not invariant under the local phase transformations, and fermion mass terms include both chiral components and the different behaviour of the left/right chiral components prevents them from being symmetric under the $SU(2)_L$ local phase transformations.

A solution to this is to extend the model by adding a $SU(2)_L$ doublet, consisting of a charged and neutral complex scalar fields [2, 3]:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.9)$$

The doublet ϕ is given a specific field potential profile and required to satisfy the Klein-Gordon field equation for a massive scalar field. The potential profile is constructed such that the model remains invariant under the SM local phase transformations (Eq. (1.5)), but when the scalar fields are at the ground state of the potential, only invariance under the subgroup $U(1)_Q$ is preserved:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \quad (1.10)$$

where v denotes the expectation value of the field in the ground state and h a real and massive scalar field. The shift in the ground state is called “spontaneous symmetry breaking” and it generates the necessary mass terms for the weak gauge bosons. The excitation of the massive scalar field h is the recently observed Higgs boson [4, 5]. Preserving the ground state invariance under the global $U(1)_Q$ phase transformation ensures that the photon is massless.

Fermion mass terms are included in the model through specific interaction terms. These are built from the left-handed fermion doublets in Eq. (1.8), the scalar doublet in Eq. (1.9), the right-handed fermion singlets, and a corresponding coupling strength for each combination (Y^{ij}). The mass terms for the up- and down-type quark are:

$$\mathcal{L}_{Yukawa} \equiv \sum_{i,j} Y_u^{ij} (\bar{Q}_{L,i} i\sigma_2 \phi^*) u_{R,j} + \sum_{i,j} Y_d^{ij} (\bar{Q}_{L,i} \phi) d_{R,j} + h.c., \quad (1.11)$$

³The weak hypercharge is defined as a combination of the electromagnetic charge Q and the third component of the weak isospin I_W^3 : $Y = 2Q + 2I_W^3$. It is introduced in a unified description of electromagnetic and weak forces to model the observed couplings of weak and electromagnetic forces to different fermions.

where σ_2 is one of the Pauli matrices, and *h.c.* stands for the hermitian conjugate terms. At the ground state of ϕ (Eq. (1.10)) the quark mass terms become:

$$\mathcal{L}_{Yukawa} = \sum_{i,j} \frac{v}{\sqrt{2}} Y_u^{ij} \bar{u}_{L,i} u_{R,j} + \sum_{i,j} \frac{v}{\sqrt{2}} Y_d^{ij} \bar{d}_{L,i} d_{R,j} + h.c. \quad (1.12)$$

Eq. (1.12) contains terms which involve different quark flavours, i.e where $i \neq j$. It is known that quarks are produced by strong interactions and these mass (or strong) eigenstates do not necessarily need to overlap with the interaction (or weak) eigenstates of Eq. (1.12). The mass eigenstates (u^m and d^m) can be expressed as combinations of the interaction eigenstates:

$$u_{L,i}^m = \sum_j V_{L,ij}^u u_{L,j} \quad , \quad u_{R,i}^m = \sum_j V_{L,ij}^u u_{R,j}, \quad (1.13)$$

$$d_{L,i}^m = \sum_j V_{L,ij}^d d_{L,j} \quad , \quad d_{R,i}^m = \sum_j V_{L,ij}^d d_{R,j}. \quad (1.14)$$

In the terms of mass eigenstates, Eq. (1.12) can be written as:

$$\mathcal{L}_{Yukawa} = \sum_{i,j} m_u^{ij} \bar{u}_{L,i}^m u_{R,j}^m + \sum_{i,j} m_d^{ij} \bar{d}_{L,i}^m d_{R,j}^m + h.c., \quad (1.15)$$

where $m_{u,d}^{ij} = \frac{v}{\sqrt{2}} (V_{L,ij}^{u,d})^\dagger Y_{u,d}^{ij} V_{R,ij}^{u,d}$, and the unitary transition matrices $V_{L,R}^{u,d}$ have been chosen such that they diagonalise the mass matrices $m_{u,d}^{ij}$ and lead to quark mass terms. Leptons acquire mass through an analogous mechanism.

In the mass basis, the charged weak currents mediated by W^\pm will include terms describing transitions between quarks from different generations (quark mixing) in mass basis:

$$\underbrace{i \frac{g_W}{\sqrt{2}} W_\mu^+ \bar{\mathbf{u}}_{\mathbf{L}} \gamma^\mu \mathbf{d}_{\mathbf{L}} + h.c.}_{\text{Weak basis}} \rightarrow \underbrace{i \frac{g_W}{\sqrt{2}} W_\mu^+ \bar{\mathbf{u}}_{\mathbf{L}}^m (\overbrace{V_L^u V_L^{d\dagger}}^{V_{CKM}}) \gamma^\mu \mathbf{d}_{\mathbf{L}}^m + h.c.}_{\text{Mass basis}}. \quad (1.16)$$

The V_{CKM} is called the Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix [20, 21]. It is a 3-by-3 unitary matrix and determines the relative strength of the weak charged-current transitions between the left-handed quarks:

$$V_{CKM} \equiv V_L^u V_L^{d\dagger} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \quad (1.17)$$

where the non-zero off-diagonal elements give rise to transitions between the generations. The analogous (left-handed) lepton mixing matrix is known as Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [22, 23].

The unitarity of the transition matrices $V_{L,R}^{u,d}$ has a direct consequence on the neutral currents mediated by Z^0 and γ bosons. Unlike the charge currents, these currents are left unchanged when expressed in the mass basis:

$$\underbrace{\bar{u}_{\mathbf{L},\mathbf{R}} \gamma^\mu u_{\mathbf{L},\mathbf{R}}}_{\text{Int. basis}} \rightarrow \underbrace{\bar{u}_{\mathbf{L},\mathbf{R}}^m \overbrace{(V_{L,R}^{u,d} V_{L,R}^{u,d\dagger})}^{\mathbb{1}}}^{\text{Mass basis}} \gamma^\mu u_{\mathbf{L},\mathbf{R}}^m \quad (1.18)$$

The absence of flavour mixing terms in the neutral current correctly describes the unobserved direct quark transitions within the up and down type quarks [24]. These Flavour Changing Neutral Currents (FCNC) are possible in higher order processes involving two up-down transitions, but remain suppressed because of the CKM unitarity. The suppression of FCNC is one of the main reasons why $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are so rare in the Standard Model.

1.1.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in the Standard Model

Mesons are the simplest particles made of quarks. They consist of a quark and an anti-quark, and differ from each other by the quark flavours. The focus of this thesis is on the decays of neutral mesons containing a b or a \bar{b} quark: B_s^0 ($\bar{b}s$), \bar{B}_s^0 ($b\bar{s}$), B^0 ($\bar{b}d$), and \bar{B}^0 ($b\bar{d}$). Unless mentioned otherwise, in the following B_s^0 and B^0 refer to both mesons and their anti-mesons, and because the treatment is analogous, $B_{(s)}^0$ is used to refer to either of the neutral B mesons. This chapter introduces meson decays and the mechanisms behind the rare nature of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in the Standard Model.

$B_{(s)}^0$ meson decays to two muons and nothing else require an interaction between down type quarks of different flavour. The weak interactions are the only interactions capable of changing the quark flavours, but as could be seen in Eq. (1.18), not directly from one down (or up) type quark to another down (or up) type quark (see Fig. 1.1). Transitions between down type quarks are possible in quantum loops involving an up type quark, as shown in Fig. 1.2. These transitions are rare because they involve multiple weak interactions and a loop suppression factor $(g_W/4\pi)^2 \sim 10^{-3}$. They are further suppressed by the unitary condition of the CKM matrix:

$$V_{ub}^* V_{us} + V_{cb}^* V_{cs} + V_{tb}^* V_{ts} = 0, \quad (1.19)$$

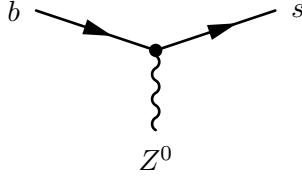


Figure 1.1: Forbidden direct (tree level) flavour changing neutral current (FCNC) in the Standard Model.

although a complete FCNC cancellation by Eq. (1.19) is avoided due to the different masses of the quarks in internal lines, as illustrated in Fig. 1.2 (GIM suppression, named after the proponents Glashow-Iliopoulos-Maiani [24]).

The decay processes contributing to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in the Standard Model are shown in Fig. 1.3. The “Z-penguin” diagrams contribute the most, whereas the analogous “Higgs-penguin” diagrams are suppressed due to small lepton and b-quark mass. Because of the large top-quark mass, contributions from diagrams involving other up-type quarks are negligible.

An additional suppression of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays arises from the angular momentum conservation. The neutral B mesons are spin-0 particles with no angular momentum in their ground state ($J^P = 0^-$). Only those decay paths that produce a final state with zero total angular momentum are allowed. In the helicity basis it means the muons must be both either in positive or negative helicity states. The chiral nature of the weak interactions, however, produces muons with opposite chirality, and for the allowed helicity configurations, the helicity state of one of the muons is always unfavoured. The helicity suppression factor is $(m_\mu/M_B)^2 \sim 4 \times 10^{-4}$.

Altogether, the absence of the FCNC interactions at the tree level, the GIM suppression of the FCNC at the loop level, and the helicity suppression

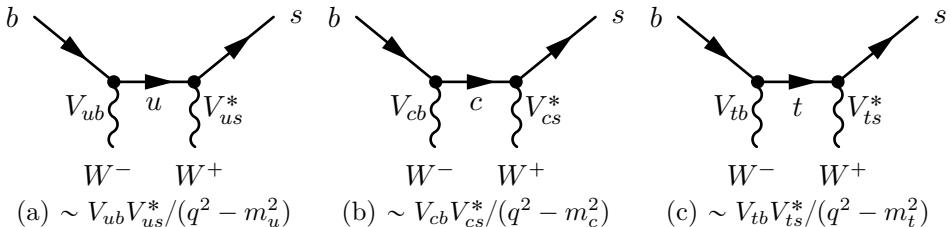


Figure 1.2: Processes contributing to the 1-loop flavour changing neutral transitions.

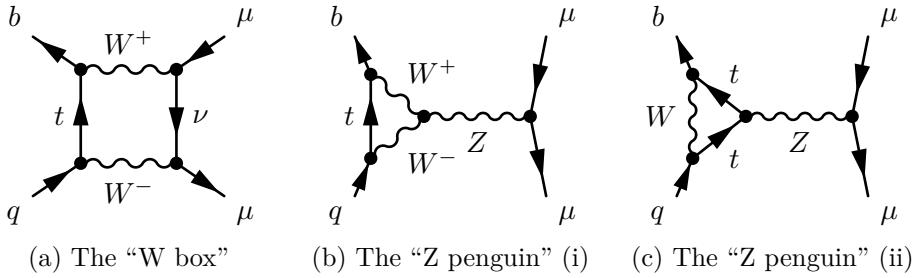


Figure 1.3: The main Standard Model diagrams contributing to the $B_s^0 \rightarrow \mu^+ \mu^-$ decays, with $q = s$, and $B^0 \rightarrow \mu^+ \mu^-$ decays, with $q = d$. Largest contributions come from Z penguin top loop (75%), and W box diagram (24%); the contributions from u and c loops, and Higgs penguins are negligible [25].

reduce the fraction of $B_{(s)}^0$ decays to two muons to the level of 10^{-9} for the $B_s^0 \rightarrow \mu^+ \mu^-$, and of 10^{-10} for the $B^0 \rightarrow \mu^+ \mu^-$.

1.2 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction definition

The branching fraction of a decay is the relative frequency of a particular decay mode. The branching fraction of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is experimentally determined without distinguishing between $B_{(s)}^0$ and anti- $B_{(s)}^0$ meson decays:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \equiv \frac{N(B_{(s)}^0 \rightarrow \mu^+ \mu^-) + N(\bar{B}_{(s)}^0 \rightarrow \mu^+ \mu^-)}{N(B_{(s)}^0) + N(\bar{B}_{(s)}^0)}. \quad (1.20)$$

Moreover, all decays are counted irrespectively of their decay time. The branching fraction in Eq. (1.20) can be calculated in terms of a universal decay probability at a given time

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \equiv \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle dt, \quad (1.21)$$

where $\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle$ is the “untagged”⁴ time-dependent decay rate, which can be calculated from theory.

The theoretical and experimental approaches to the branching fraction definition have subtle but important differences that need to be accounted for [26–29]. The theoretical branching fraction calculation is performed using the flavour states with a well defined quark structure.

Quarks are produced in their mass eigenstates and bound into a meson by strong processes. Because the flavour and anti-flavour states of the meson can

⁴Tagging is the experimental term for identifying a $B_{(s)}^0$ state from an anti- $B_{(s)}^0$ state.

transform into one-another ($B_{(s)}^0 \leftrightarrow \bar{B}_{(s)}^0$), the “physical” meson propagates as a mixture of these states. These mixtures have a well defined mass and lifetime, and are called meson mass eigenstates $B_{q,H}^0$ and $B_{q,L}^0$, where the labels stand for heavy and light and refer to the mass hierarchy of the mixtures. The exact composition of the mixtures depends on the mixing strength. The mean mass, decay rate, and lifetime of a $B_{(s)}^0$ meson are all defined in terms of the mass eigenstate properties:

$$M_{B_{(s)}^0} = \frac{M_{q,H} + M_{q,L}}{2}, \quad \Gamma_q = \frac{\Gamma_{q,H} + \Gamma_{q,L}}{2}, \quad \tau_{B_{(s)}^0} = \frac{1}{\Gamma_q}. \quad (1.22)$$

The theoretical branching fraction is defined in terms of $B_{(s)}^0$ flavour eigenstate decay rates. A time-dependent decay rate, i.e. the probability for a $B_{(s)}^0$ -meson flavour eigenstate to decay into a given final state f at any time t , can be defined as

$$\Gamma(B_{(s)}^0(t) \rightarrow f) \equiv \frac{dN(B_{(s)}^0(t) \rightarrow f)/dt}{N_{B_{(s)}^0}}, \quad (1.23)$$

where the rate is given with respect to the total number of $B_{(s)}^0$ mesons ($N_{B_{(s)}^0}$) and $dN(B_{(s)}^0 \rightarrow f)$ is the number of decays to the final state f in a time window $[t, t+dt]$. The definition of $\Gamma(\bar{B}_{(s)}^0 \rightarrow f)$ is analogous⁵. Lifetime and decay width cannot be defined for a flavour eigenstate, and instead, the total meson decay width $\Gamma_q = (\Gamma_{q,H} + \Gamma_{q,L})/2$ is used to calculate the theoretical branching fraction:

$$\begin{aligned} \mathcal{B}_{th.}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) &\equiv \frac{\tau_{B_{(s)}^0}}{2} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle|_{t=0} \\ &= \frac{1}{2} \left[\frac{\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0}}{(\Gamma_{q,H} + \Gamma_{q,L})/2} + \frac{\Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0}}{(\Gamma_{q,H} + \Gamma_{q,L})/2} \right]. \end{aligned} \quad (1.24)$$

The experimentally measured branching fraction in Eq. (1.20) is instead expressed in terms of (heavy and light) mass eigenstates

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{1}{2} \left[\frac{\Gamma(B_{(s),H}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0}}{\Gamma_{q,H}} + \frac{\Gamma(B_{(s),L}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0}}{\Gamma_{q,L}} \right]. \quad (1.25)$$

⁵For a state with a well defined lifetime (and thus well defined decay width) the time-dependent decay rate in Eq. (1.23) drops exponentially with time as $\Gamma(t) = \Gamma(0)e^{-\Gamma_q t}$, where Γ_q is the total decay rate of a meson to any of the possible final states.

The theoretical and measured branching fractions in Eq. (1.24) and Eq. (1.25) agree in case the heavy and light mass eigenstates have the same lifetime. This is approximately true for the B^0 meson system [19]. For the B_s^0 meson system, however, the mass eigenstates lifetime difference has been measured [30, 31] to be

$$y_s \equiv \frac{\Gamma_{s,L} - \Gamma_{s,H}}{\Gamma_{s,L} + \Gamma_{s,H}} = 0.075 \pm 0.012. \quad (1.26)$$

Using the lifetime asymmetry in Eq. (1.26), the physical and theoretical branching fraction can be related [26] by

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \left[\frac{1 + y_{(d,s)} \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}}{1 - y_{(d,s)}^2} \right] \times \mathcal{B}_{th.}(B_{(s)}^0 \rightarrow \mu^+ \mu^-), \quad (1.27)$$

where $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ is the mass-eigenstate rate asymmetry, which depends on both the final state and the physics model considered, and for $B_s^0 \rightarrow \mu^+ \mu^-$ is defined as

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} \equiv \frac{\Gamma(B_{s,H}^0 \rightarrow \mu^+ \mu^-) - \Gamma(B_{s,L}^0 \rightarrow \mu^+ \mu^-)}{\Gamma(B_{s,H}^0 \rightarrow \mu^+ \mu^-) + \Gamma(B_{s,L}^0 \rightarrow \mu^+ \mu^-)}. \quad (1.28)$$

In the SM, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ in Eq. (1.28) takes the maximal value $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = +1$, which means that only the heavy mass-eigenstate $B_{s,H}^0$ contributes to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. As described in Refs. [27, 28], $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ can also be constrained experimentally through the measurement of the effective lifetime. The conversion factors between the experimental and the theoretical branching fractions are shown in Fig. 1.4 for different values of the mass-eigenstate and lifetime asymmetry.

1.3 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction computation

This section discusses how the theoretical branching fraction is computed in a model-independent framework. A detailed discussion can be found in Refs. [32–39].

The instantaneous decay rate of a B meson flavour eigenstate into two muons is given by Fermi’s golden rule [6]:

$$\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0} = \frac{1}{16\pi} \cdot \frac{1}{M_{B_{(s)}^0}} \cdot \sqrt{1 - 4 \left(\frac{m_\mu}{M_{B_{(s)}^0}} \right)^2 \cdot |\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|^2}, \quad (1.29)$$

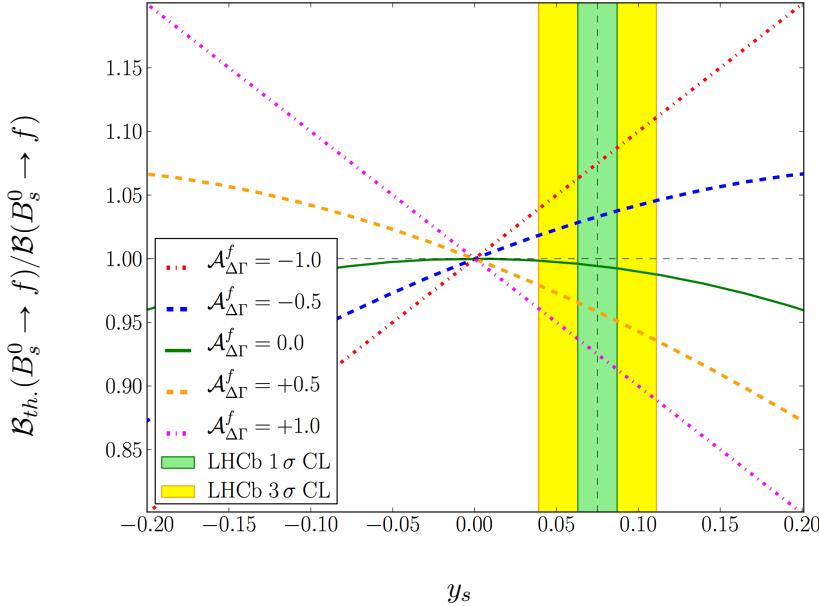


Figure 1.4: The conversion factor from the experimentally determined $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction to the theoretical prediction, shown for different values of $\mathcal{A}_{\Delta\Gamma}^f$ and y_s . The figure is taken from Ref. [29], and includes the latest LHCb y_s measurements [30, 31].

where m_μ is the mass of a muon, $M_{B_{(s)}^0}$ the mass of the relevant B meson, and $\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ the invariant transition matrix element. $\Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0}$ is expressed in an analogous way.

The matrix element $\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ describes the transition processes responsible for the decays. The processes contributing to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ matrix element can be expressed as:

$$\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \underbrace{B_{(s)}^0 \text{ (loop diagram)}}_{\text{Strong\&EW}} + \underbrace{B_{(s)}^0 \text{ (loop diagram)}}_{\text{Strong\&EW}} + \dots \quad (1.30)$$

The computation of $\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is complicated by the presence of the strong interaction. Unlike the electroweak interactions, where the small and almost constant coupling ($1/137 \ll 1$) means that more complicated higher-order diagrams contribute little and can be neglected, the strong coupling depends on the momentum transfer scale in the processes; at the low energy scales of the meson-binding processes ($\Lambda_{QCD} \simeq 0.3$ GeV), it becomes large

enough ($\mathcal{O}(1)$ at 1 GeV) to undermine such a perturbative approach. In the case of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, the treatment of the strong interaction is simplified because the final state has no strongly interacting particles, and is thus (from the point of view of the strong interactions) decoupled from the initial state. This results in factorisation, as will be discussed in Sec. 1.3.2. The role of the strong interaction in binding the quarks into a meson still needs to be included.

1.3.1 The effective approach

A way of dealing with non-perturbative effects is to separate high- and low-energy contributions to the matrix element, as is done in the “effective approach”: in high momentum transfer processes, the strong coupling is small enough to allow for a perturbative approach; in low momentum transfer processes, the effects are calculated using non-perturbative techniques like Lattice QCD.

In Quantum Field Theory, interactions between particles are mediated by the exchange of other particles, and the transition from an initial state to a final state, passes through an intermediate state. In an effective approach, the effects of these mediators are omitted, and, instead, the interaction matrix is expressed in terms of effective “four-point” vertices, that can be understood as initial state particles scattering off a static potential and emerging as final state particles:

$$\begin{aligned} \mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) &\equiv A_1 \text{ } \begin{array}{c} \diagup \\ \diagdown \end{array} \text{ } + A_2 \text{ } \begin{array}{c} \diagdown \\ \diagup \end{array} \text{ } + \dots \\ &\simeq \sum_i^{OPE} \left\{ \underbrace{A_i}_{\text{Perturbative}} \otimes \underbrace{\text{ } \begin{array}{c} \text{ } \\ \text{ } \end{array} \text{ } \circ_i \text{ } \begin{array}{c} \text{ } \\ \text{ } \end{array}}_{\text{Non-perturbative}} \right\} \end{aligned} \quad (1.31)$$

The so-called Operator Product Expansion (OPE) in Eq. (1.31), represents the effective theory in which the computations are performed. While the mediators are physical (e.g. W^\pm), and the effective four-point view is known to be a simplified picture, it does, when given proper weights A_i , reproduce the total effects of the QFT. A factorisation scale λ separates the high-energy and low-energy regimes: the perturbative physics effects (momentum transfer above λ) are contained in the weight coefficients, A_i , while the corresponding low momentum transfer non-perturbative contributions are contained in the four-point matrix elements.

The ellipsis in Eq. (1.31) represents higher order operators. These scale with the mediator mass as $\mathcal{O}(p^2/M_{\text{mediator}}^2)$, and for heavy mediators such as W^\pm , become negligible at the typical meson momentum scales of $B_{(s)}^0$ decays ($p = \Lambda_{QCD}, m_s, m_b$).

In a more formal notation:

$$\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \langle \mu^+ \mu^- | \mathcal{H}_{\text{eff},q} | B_{(s)}^0 \rangle + \mathcal{O}(p^2/M_W^2) \quad (1.32)$$

$$\simeq \frac{G_F}{\sqrt{2}} \sum_i \underbrace{C_i(\lambda)}_{\substack{\text{Wilson coef.} \\ (\text{Perturbative})}} \times \underbrace{\langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle}_{\substack{\text{4-point matrix element} \\ (\text{Non-perturbative})}}, \quad (1.33)$$

where \mathcal{O}_i denotes the four-point operator, $G_F \equiv (\sqrt{2}g_W^2)/(8M_W^2)$ is the Fermi constant for the weak processes common to all operators, C_i are the perturbatively calculable “Wilson coefficients” describing short distance effects, and λ is the scale at which short and long distance effect are factorised.

The Wilson coefficients describe the effects of the heavier mediators (e.g. W^\pm, Z, t, H). Their values are obtained by matching the effective theory expression to the decay-amplitude calculated in the full theory (see Ref. [32]). The matching is performed at high factorisation scales⁶, after which the coefficient values are evolved down to the energy scales of the decay ($\lambda = \mathcal{O}(m_b)$) at which the (hadronic) non-perturbative part of the amplitude can be estimated.

1.3.2 The non-perturbative matrix element

The final state of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ is purely leptonic and does not couple directly to the initial state quarks. The four-point matrix element can therefore be factorised into a hadronic and a leptonic part:

$$\langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle = \underbrace{\langle \mu^+ \mu^- | \mathcal{O}_i^{ll} | 0 \rangle}_{\text{Leptonic matrix element}} \otimes \underbrace{\langle 0 | \mathcal{O}_i^{qq} | B_{(s)}^0 \rangle}_{\text{Hadronic matrix element}}, \quad (1.34)$$

where now all non-perturbative strong interaction effects are contained in the hadronic part.

Lorentz invariance limits the possible $\mathcal{O} = \mathcal{O}^{ll} \otimes \mathcal{O}^{qq}$ interaction forms to

⁶Because the Wilson coefficient calculation involves logarithmic $\alpha_s(\lambda) \ln(M_W^2/\lambda^2)$ terms, the matching is performed at factorisation scales where the large logarithmic terms disappear ($\lambda = \mathcal{O}(M_W)$).

the following set:

$$S^{ll} \otimes S^{qq} , \quad S^{ll} \otimes P^{qq} , \quad P^{ll} \otimes S^{qq} , \quad P^{ll} \otimes P^{qq}, \quad (1.35)$$

$$V^{ll} \otimes V^{qq} , \quad V^{ll} \otimes A^{qq} , \quad A^{ll} \otimes V^{qq} , \quad A^{ll} \otimes A^{qq}, \quad (1.36)$$

$$T^{ll} \otimes T^{qq} , \quad (1.37)$$

with S, P, V, A , and T denoting scalar ($\mathbb{1}$), pseudo-scalar (γ_5), vector (γ_μ), axial-vector ($\gamma_\mu\gamma_5$), and tensor ($\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu$) forms of the hadronic and leptonic operators.

The hadronic part $\langle 0 | \mathcal{O}_i^{qq} | B_{(s)}^0 \rangle$ must be expressed in terms of the B meson four-momentum (p_B^μ), the only quantity available to describe the pseudo-scalar mesons $B_{(s)}^0$. The four-momentum is a vector (parity odd) quantity, and, when contracted with itself, a scalar (parity even) quantity can be built. Therefore, the hadronic matrix element can only be a vector or a scalar quantity, and the $T^{ll} \otimes T^{qq}$ form in Eq. (1.37) cannot contribute to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Also the terms with hadronic scalar and vector operators in Eq. (1.35) and Eq. (1.36) ($S^{ll} \otimes S^{qq}, P^{ll} \otimes S^{qq}, V^{ll} \otimes V^{qq}, A^{ll} \otimes V^{qq}$) must vanish, since, given the pseudo-scalar nature of the $B_{(s)}^0$, the corresponding hadronic matrix elements transform as pseudo-vector and pseudo-scalar.

The hadronic axial-vector (A^{qq}) contribution can be expressed as

$$\langle 0 | \bar{b} \gamma_\mu (\pm \gamma_5) s | B_{(s)}^0 \rangle \equiv \pm i p_\mu f_{B_{(s)}^0}, \quad (1.38)$$

where all non-perturbative hadronic effects are contained in one constant, $f_{B_{(s)}^0}$, called the “decay constant”. The decay constant can be calculated using Lattice QCD methods [40]. From the simple form of the axial-vector matrix element (Eq. (1.38)), it follows that the $V^{ll} \otimes A^{qq}$ form in Eq. (1.36) cannot contribute either:

$$\langle \mu^+ \mu^- | V^{ll}(\mu) | 0 \rangle \langle 0 | A^{qq}(\mu) | B_{(s)}^0 \rangle = \langle \mu^+ \mu^- | V^{ll}(\mu) | 0 \rangle i f_{B_{(s)}^0} p_\mu, \quad (1.39)$$

which vanishes due to the Ward identity [38]. As a consequence, $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays cannot proceed through photon-penguin diagrams.

The expression of hadronic pseudo-scalar (P^{qq}) contribution can be obtained contracting the decay constant definition in Eq. (1.38) with p_μ on both sides:

$$\langle 0 | \bar{b} (\pm \gamma^5) q | B_{(s)}^0 \rangle = \mp \frac{i M_{B_{(s)}^0}^2}{m_b + m_q} f_{B_{(s)}^0}. \quad (1.40)$$

In summary, the operators that contribute to the non-perturbative $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ matrix element $\langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle$, on the left side of Eq. (1.34), are

$$\mathcal{O}_S^L = (\bar{b}P_L q) \otimes (\bar{\mu}\mu) , \quad \mathcal{O}_S^R = (\bar{b}P_R q) \otimes (\bar{\mu}\mu), \quad (1.41)$$

$$\mathcal{O}_P^L = (\bar{b}P_L q) \otimes (\bar{\mu}\gamma_5\mu) , \quad \mathcal{O}_P^R = (\bar{b}P_R q) \otimes (\bar{\mu}\gamma_5\mu), \quad (1.42)$$

$$\mathcal{O}_{10}^L = (\bar{b}\gamma_\mu P_R q) \otimes (\bar{\mu}\gamma_\mu\gamma_5\mu) , \quad \mathcal{O}_{10}^R = (\bar{b}\gamma_\mu P_L q) \otimes (\bar{\mu}\gamma_\mu\gamma_5\mu) \quad (1.43)$$

where $q = d, s$. The operators have been written separately for the right- and left-handed \bar{b} quark chirality states. Using the operators in Eqs. (1.41)-(1.43), one can construct the effective Hamiltonian for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in Eq. (1.32):

$$\mathcal{H}_{eff,q} = \frac{G_F}{\sqrt{2}} V_{t(d,s)}^* V_{tb} \sum_i^{10,S,P} (\mathcal{C}_i^L(\lambda) \mathcal{O}_i^L + \mathcal{C}_i^R(\lambda) \mathcal{O}_i^R). \quad (1.44)$$

Since the other up-type quark contributions in the internal loop diagrams are negligible, only the top quark contribution has been retained, and the relevant CKM matrix elements are explicitly indicated. Equation (1.44) is used in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ transition matrix calculation in Eq. (1.32).

1.3.3 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction is expressed in the centre-of-mass frame of the meson, where $p^\mu = M_{B_{(s)}^0} \hat{v}$. It can be calculated from Eq. (1.29), using the effective hamiltonian in Eq. (1.44), and the hadronic matrix elements (in terms of the decay constant) in Eqs. (1.38) and (1.40). The time-dependent untagged decay rate summed over all helicity states is obtained contracting the lepton Dirac spinor terms and taking their sum over all helicity states [29]:

$$\begin{aligned} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle &\equiv \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) + \Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \\ &= \frac{G_F^4 M_W^4 \sin^4 \theta_W}{4\pi^5} |C_{10}^{SM} V_{t(d,s)} V_{tb}^*|^2 \\ &\times f_{B_{(s)}^0}^2 M_{B_{(s)}^0} m_\mu^2 \cdot \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}^0}^2}} \\ &\times (|P|^2 + |S|^2) \\ &\times e^{-t/\tau_{B_{(s)}^0}} [\cosh(y_{(d,s)} t/\tau_{B_{(s)}^0}) + \\ &\quad \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} \sinh(y_{(d,s)} t/\tau_{B_{(s)}^0})], \end{aligned} \quad (1.45)$$

where the pseudo-scalar and scalar operator effects are described by the coefficients

$$P = \frac{C_{10}^R - C_{10}^L}{C_{10}^{SM}} + \frac{M_{B_{(s)}^0}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{C_P^R - C_P^L}{C_{10}^{SM}} \right), \quad (1.46)$$

$$S = \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}^0}^2}} \cdot \frac{M_{B_{(s)}^0}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{C_S^R - C_S^L}{C_{10}^{SM}} \right). \quad (1.47)$$

1.3.4 The Standard Model $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction

In the SM, the only non-negligible contribution to the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decay rates comes from the $A^{ll} \otimes A^{qq}$ operator \mathcal{O}_{10}^R in Eq. (1.36). The SM Higgs boson contribution to \mathcal{O}_{10}^R is negligible when compared to the Z-penguin or W-box contributions in Fig. 1.3. In the SM, the coefficients P and S in Eqs. (1.46) and (1.47) are $P^{SM} = 1$ and $S^{SM} = 0$. Therefore, the SM $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction expression obtained from Eqs. (1.24) and (1.48) reads simply

$$\begin{aligned} \mathcal{B}_{th.}^{SM}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) &= \frac{\tau_{B_{(s)}^0}}{2} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle|_{t=0, P=1, S=0}, \\ &= \frac{\tau_{B_{(s)}^0} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} |C_{10}^{SM} V_{t(d,s)} V_{tb}^*|^2 \\ &\quad \times f_{B_{(s)}^0}^2 M_{B_{(s)}^0} m_\mu^2 \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}^0}^2}}. \end{aligned} \quad (1.48)$$

Please note that the SM prediction in Eq. (1.48) needs to be corrected for the B_s^0 meson mixing effects ($\sim 8\%$) before comparing to the experimental value (see Eq. (1.27)).

The most recent branching fractions calculations account for higher-order electromagnetic and strong interaction effects [41], use lattice QCD to compute the meson decay constants [40], and include the combined top quark mass measurements [42]. Including the B_s^0 meson mixing effects, the most recent SM predictions are

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.66 \pm 0.23) \times 10^{-9}, \quad (1.49)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}. \quad (1.50)$$

The dominant uncertainty in the theory predictions stems from the CKM matrix elements (see Fig. 1.5).

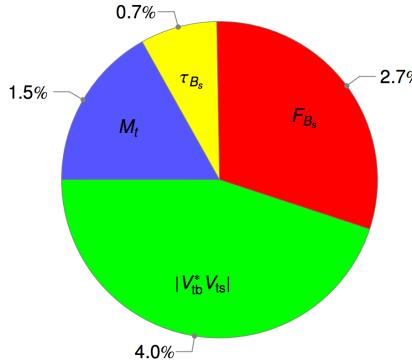


Figure 1.5: The error budget of the $B_s^0 \rightarrow \mu^+\mu^-$ SM branching fraction estimate [28].

The accuracy of the theoretical prediction is improved when taking a ratio of the $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ branching fractions, where several sources of uncertainty (e.g. top mass, V_{tb} , Wilson coefficients) cancel out [43]:

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)} = \frac{\tau_{B^0}}{1/\Gamma_H^s} \left(\frac{f_{B^0}}{f_{B_s^0}} \right)^2 \left| \frac{V_{td}}{V_{ts}} \right|^2 \frac{M_{B^0}}{M_{B_s^0}} \sqrt{1 - \frac{4m_\mu^2}{M_{B^0}^2}} = 0.0295^{+0.0028}_{-0.0025}. \quad (1.51)$$

1.3.5 Radiative corrections

Photon emission inevitably occurs in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. The dominant mechanism affecting the experimental $B_{(s)}^0 \rightarrow \mu^+\mu^-$ search is the final state (Bremsstrahlung) radiation from the muons [41, 44, 45]. Additionally to the final state emission, the muons may radiate photons due to the interactions with the detector material and the magnetic field. If not reconstructed, the photons emitted at higher energies considerably reduce the invariant mass of the reconstructed di-muon pair, and could reduce the measured signal yield in the search widow. Experimentally, the photon emission is simulated with the PHOTOS package [46]. It is included in the signal detection efficiency correction, and the loss of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ events due to Bremsstrahlung radiation is accounted for in the experimental branching fraction through the efficiency correction.

From the theoretical point of view, allowing a photon in the final state in

the measurement of the branching fraction means including $B_{(s)}^0 \rightarrow \mu^+ \mu^- \gamma$ decays. The calculation of $B_{(s)}^0 \rightarrow \mu^+ \mu^- \gamma$ branching fraction is similar to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction calculation, but involves contributions from additional operators and intermediate resonances; in some cases, the helicity suppression affecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay can be overcome in $B_{(s)}^0 \rightarrow \mu^+ \mu^- \gamma$ decays [45]. In addition to the Bremsstrahlung radiation from final state muons, the photon in $B_{(s)}^0 \rightarrow \mu^+ \mu^- \gamma$ can be radiated directly from the initial state quarks, or from an intermediate operator. Moreover, constructive or destructive interference between the initial-, intermediate-, and final state emission affects the decay rate.

The total branching fraction of $B_{(s)}^0 \rightarrow \mu^+ \mu^- \gamma$ decays is expected to be one order of magnitude higher than that of the fully leptonic mode [47]. This, however, has little (sub-percent) impact on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurement because radiative effects other than the final state Bremsstrahlung from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ have been shown to produce photons with higher energy, and therefore, have no significant impact on the experimental di-muon search region (Fig. 1.6).

1.4 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in New Physics models

New Physics could either suppress or enhance the small $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions predicted by the SM (see Fig. 1.7). Possible deviations from the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions predicted by the SM (Eqs. (1.49), (1.50)) can be described in terms of the signal strength:

$$\begin{aligned} \mathcal{S}_{\text{SM}}^{B_{(s)}^0} &= \frac{\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}^{\text{SM}}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)} \\ &= \left[\frac{1 + y_{(d,s)} \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}}{1 + y_{(d,s)}} \right] \times (|P|^2 + |S|^2), \end{aligned} \quad (1.52)$$

where the coefficients P and S are given in Eqs. (1.46) and (1.47). In the SM, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = +1$, $P = +1$ and $S = 0$. The New Physics effects can alter the signal strength through $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$, $|P|$, and $|S|^2$ ⁷. A large variety of different models can affect the signal strength. A comprehensive review can be found in Refs. [28, 29]. Here, three (limiting) scenarios from Ref. [29] will be discussed: (i) $S = 0$, (ii) $P = 1$, and (iii) $P \pm S = 1$, each dominated by different New Physics contributions.

⁷Or, equivalently, through P , S , and their respective phases, φ_P and φ_S .

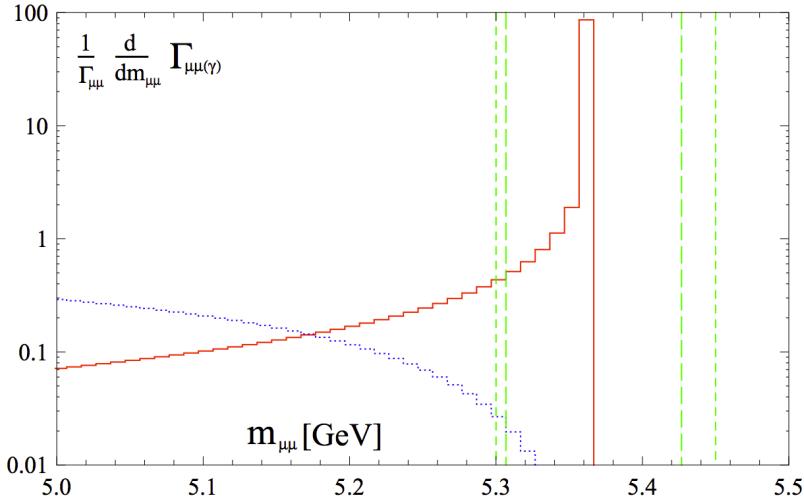


Figure 1.6: The di-muon invariant mass spectrum for $B_{(s)}^0 \rightarrow \mu^+ \mu^- + n\gamma$, with $n = 0, 1, 2, \dots$. The dotted (blue) curve corresponds to the photon emission from the initial state quarks, the solid (red) curve is dominated by the low-energy photon radiation from the muons (Bremsstrahlung), and corresponds to the theoretical branching fraction estimate. The vertical dashed and dash-dotted (green) lines indicate the CMS and LHCb signal windows, respectively [41]

1.4.1 Pseudo-scalar dominated New Physics ($S = 0$)

$$S \propto \left(\frac{C_S^R - C_S^L}{C_{10}^{SM}} \right) = 0$$

The $S = 0$ scenario includes New Physics models in which no scalar particles contribute to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, and leaves $C_{10}^{L,R}$ and $C_P^{L,R}$ free to take non-SM values. Models including scalar contributions can qualify, provided the coupling to quarks is left-right symmetric ($C_S^R = C_S^L$).

Constrained Minimal Flavor Violation (CMFV) is the simplest example: in CMFV, the flavor structure as well as the contributing operators are assumed to be SM-like. The contributions to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction can only arise from C_{10}^R . Models with SM operator structure but altered CP violating phases can contribute (albeit little) to the signal

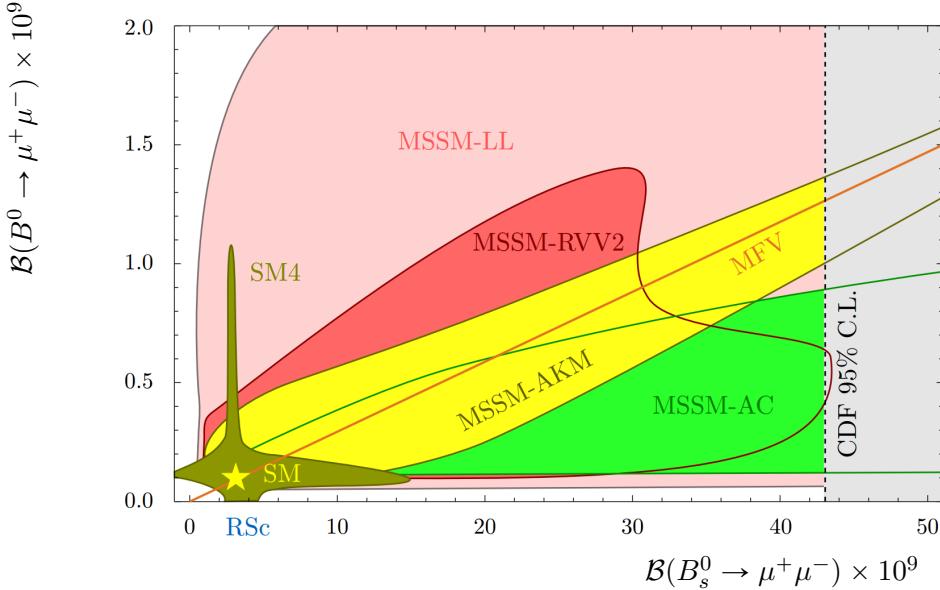


Figure 1.7: Possible values of the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions in the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ plane. The Standard Model prediction is shown as a star. For models with CKM-like couplings, called Minimal Flavor Violation, the ratio of the branching fractions is fixed along the orange diagonal. In SM4, the rates are modified through a fourth generation [48]. The four MSSM-X domains represent the branching fraction in different supersymmetric models [49]. The grey area denotes the experimentally excluded region before the results from Large Hadron Collider (LHC). The figure is taken from [50].

strength through $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$. Examples are the Littlest Higgs Model with T-Parity (LHT) [51] and non-minimal “331” models [52].

Other models alter the operator structure and introduce contributions beyond the operator associated to coefficient C_{10}^R . For example, in the Randall-Sundrum model, New Physics contributions to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ could arise through right-handed flavour violating Z boson couplings to quarks [53].

New operators can also arise if FCNC are made possible at tree level, e.g. in Z' models [54], or models with new pseudo-scalar particles [55] (see Fig. (1.8a) and (1.8c)). The latter includes models with additional Higgs doublets, which arise in many SM extensions, including the Minimal supersymmetric extension of the SM (MSSM), twin Higgs models, and certain composite Higgs models (see Ref. [56] and references therein).

More generally, two Higgs doublet models (2HDMs) provide a simple way to parametrize extensions of the Higgs sector. Two Higgs doublet models

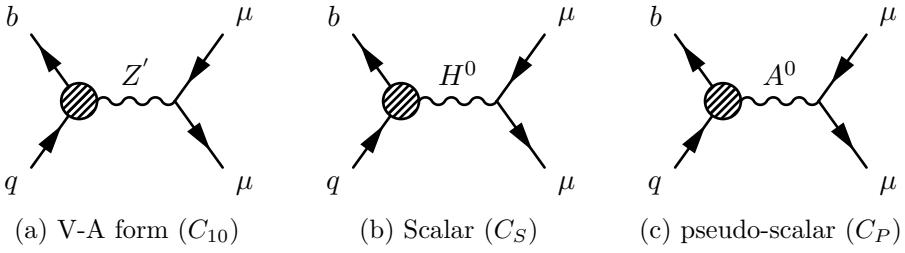


Figure 1.8: Possible tree-level FCNC contributing to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in Z' models (a), two-Higgs doublet models with heavy scalar and pseudo-scalar bosons, H^0 (b) and A^0 (c).

(2HDM) involve five physical scalars: two neutral scalars, h, H^0 ; one neutral pseudo-scalar, A^0 ; and two charged scalars, H^+, H^- . The 2HDM qualify for the $S = 0$ scenario in case the pseudo-scalar A^0 is considerably lighter than the scalar particle H^0 .

The correlation between the $B_s^0 \rightarrow \mu^+ \mu^-$ signal strength and mass-eigenstate rate asymmetry observables for the $S = 0$ scenario is shown in Fig. 1.9, on the left plot.

1.4.2 Scalar dominated New Physics ($P = 1$)

$$P = \frac{C_{10}^R - C_{10}^L}{C_{10}^{SM}} + \frac{M_{B_{(s)}^0}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_{(d.s)}} \right) \left(\frac{C_P^R - C_P^L}{C_{10}^{SM}} \right) = 1$$

In the simplest case, the $P = 1$ scenario is realised if $C_{10}^R = C_{10}^{SM}$ and $C_{10}^L = C_P^{R,L} = 0$, but also left-right symmetric pseudo-scalars are included ($C_P^R = C_P^L$). The $P = 1$ scenario is complementary to the pseudo-scalar dominated $S = 0$ scenario, as here New Physics effects are driven by scalar operators. 2HDMs with a heavy pseudo-scalar A^0 and with a considerably lighter scalar H^0 particle are a good example (see Fig. (1.8b)). Also, various models with scalar induced tree-level FCNC could contribute to $P = 1$, as analysed in Ref. [55].

The correlation between the $B_s^0 \rightarrow \mu^+ \mu^-$ signal strength and mass-eigenstate rate asymmetry observables for the $P = 0$ scenario is shown in Fig. 1.9, on the right.

1.4.3 Mixed (pseudo-)scalar New Physics ($P \pm S = 1$)

$$\begin{aligned} (C_P^R + C_S^R) - (C_P^L + C_S^L) &= 0 \\ (C_P^R - C_S^R) - (C_P^L - C_S^L) &= 0 \end{aligned}$$

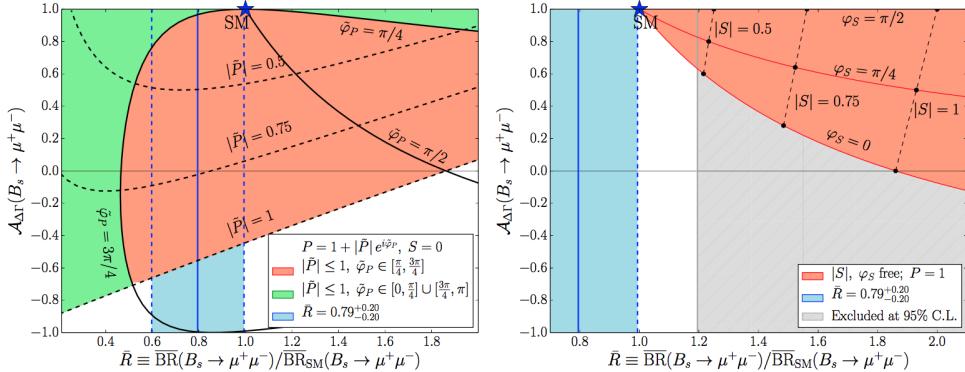


Figure 1.9: The correlation between the $B_s^0 \rightarrow \mu^+ \mu^-$ signal strength (\bar{R}) and the mass-eigenstate rate asymmetry ($\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$) in the $S = 0$ scenario (left), and in the $P = 1$ scenario (right). The figure is taken from Ref. [27, 29], and includes the latest CMS and LHCb $B_s^0 \rightarrow \mu^+ \mu^-$ measurements [57, 58].

After taking $(4m_\mu^2/M_{B_{(s)}^0}^2) \simeq 0$ (in the definition of S , Eq. (1.47)), and assuming $C_{10}^R - C_{10}^L \simeq C_{10}^{SM}$ (in the definition of P , Eq. (1.46)), the $P \pm S = 1$ scenario is realised if New Physics effects in S and P are of similar size. For example, the coefficients of 2HDMs in the decoupling regime ($M_{H^0} \simeq M_{A^0} \simeq M_{H^\pm} \gg M_h$) (see App. A of Ref. [29]), are related by

$$C_S^R = -C_P^R \quad , \quad C_S^L = C_P^L. \quad (1.53)$$

These 2HDMs will qualify for the $P \pm S = 1$ scenario in case the couplings of the H^0 and A^0 Higgs bosons are left-right asymmetric, and either $C_{S,P}^L$ or $C_{S,P}^R$ are dominant. This condition is satisfied in Minimal Flavor Violation (MFV) models, where the couplings are related by $C_{S,P}^L/C_{S,P}^R \sim m_{s,b}/m_b$.

MSSM models can contribute to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays similarly, if contributions to C_{10}^L are negligible. In the MSSM, the 2HDM decoupling regime is realised separating the vacuum expectation values of the Higgs doublets. Substantially different expectation values at the electroweak-symmetry breaking scale, quantified by their ratio $\tan\beta \equiv v_2/v_1$, can, also enlarge scalar operator contributions [59, 60], which are proportional to $(\tan\beta)^3$.

Chapter 2

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ production and detection: The Large Hadron Collider and the LHCb detector

The largest laboratory for particles physics research is the European Organisation for Nuclear Research (CERN), established in 1954. CERN hosts the Large Hadron Collider (LHC), an accelerator built in a 27 km long circular underground tunnel, partly in France, partly in Switzerland (see Fig. 2.1). Its primary goal is to study processes taking place in proton-proton (pp) collisions at record high energies, including the rare B meson decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. The LHC accelerator and B meson production in LHC will be discussed in Sec. 2.1.

The four main experiments at the LHC (ALICE, ATLAS, CMS, and LHCb) are located at points where the two proton beams are brought into collision. The collisions at the LHC Point 8, as shown in Fig. 2.1, are observed by the LHCb detector, optimised to detect rare B meson decays such as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$.

The LHCb detector will be the main topic of this chapter: the sub-detectors are described in Sec. 2.2 and the trigger system in Sec. 2.3; an overview of the offline reconstruction of the events for the final physics analysis is presented in Sec. 2.4, while Sec. 2.5 focusses on the particle identification, crucial to separate true $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays from other decays.

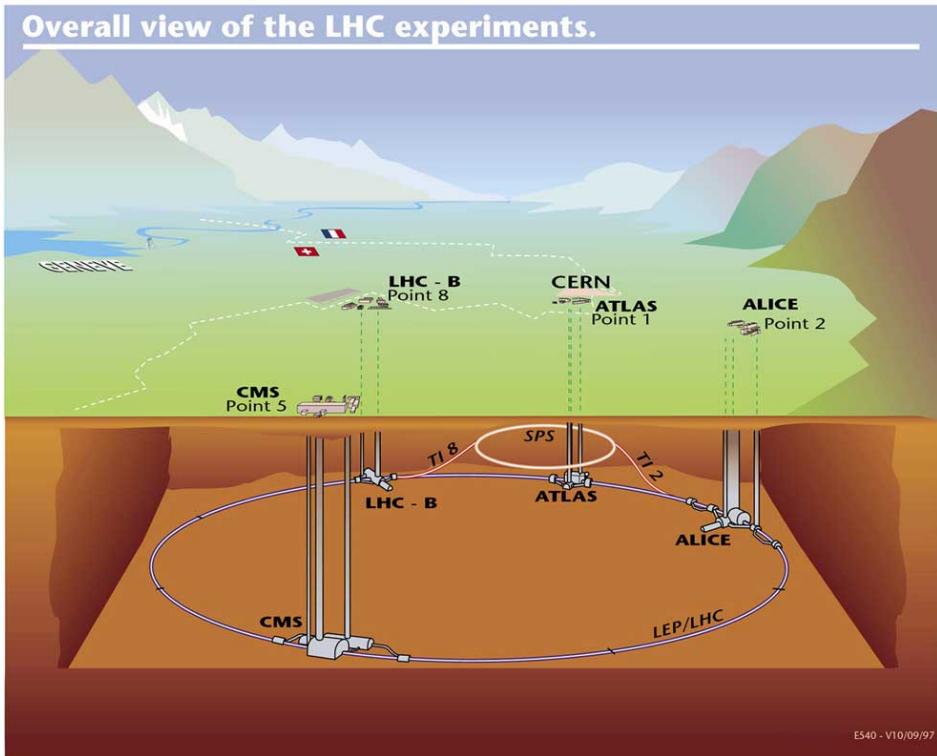


Figure 2.1: A schematic view of the Large Hadron Collider (LHC), the largest particle accelerator in the world. Two proton beams are accelerated in a 27 km circular underground tunnel. The two proton beams cross each other at four points, where four detectors are located: ALICE, ATLAS, CMS, and LHCb.

2.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is a 27 km long circular accelerator built in an underground tunnel on the border between France and Switzerland (see Fig. 2.1). In LHC, two proton beams are accelerated in opposite directions and brought into collision at four points where the beams cross each other. The LHC became operational in 2010. In 2011, the four experiments recorded proton collisions at a centre-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$. In 2012, the collision energy was increased to $\sqrt{s} = 8 \text{ TeV}$. At present, the LHC is recording pp collision data at $\sqrt{s} = 13 \text{ TeV}$.

Each of the two proton beams in LHC consists of 3564 slots. A filled slot contains about 10^{11} protons, collectively called a proton “bunch”. Slots of the

two beams (filled or not) cross each other every 25 ns. Not all available slots are filled. For reasons related to machine safety, about half the slots were filled during 2011 and 2012. The total number of bunches per beam was first gradually increased to 1380 during the first part of 2011, and then kept such. Thus, the effective bunch-bunch crossing rate, ν_{filled} , was $\frac{1380}{3564} \times 40 \text{ MHz} \approx 15 \text{ MHz}$.

The number of pp collisions is determined by the (instantaneous) luminosity (\mathcal{L}) and the pp cross-section (σ_{pp}):

$$N_{pp} = \sigma_{pp} \int_{\Delta t} \mathcal{L} dt. \quad (2.1)$$

The luminosity is a key parameter describing a collider: it describes the compactness of the beams, the ability of the magnets to focus the beams at the interaction points, the number of protons per bunch, and the number of bunch crossings per second. The design luminosity of the LHC is $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The luminosity at the LHCb interaction point, however, is kept lower, such that events are dominated by a single pp interaction per bunch crossing and thereby reducing the occupancy as well as the radiation damage to the detector and allowing the proper resolution of vertex structure and affordable reconstruction times in the trigger.

Lower luminosity ($\mathcal{L} = 4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, in 2012) is achieved by beam defocussing at the LHCb interaction point. A fill can last for several hours, during which the luminosity is kept stable by compensating the reduction in the number of protons with a smaller transverse beam separation (see Fig. 2.2).

The number of pp collisions in the bunch-bunch crossings also depends on the pp cross-section (σ_{pp}). The total pp cross section σ_{pp}^{tot} has been measured in LHC at both collision energies, $\sqrt{s} = 7 \text{ TeV}$ and $\sqrt{s} = 8 \text{ TeV}$, to be close to 100 mb [61–63].

This also includes elastic pp collisions where the two protons stay intact, deflect only little after the collision, and miss the detector. The visible (inelastic) pp cross-section, σ_{pp}^{inel} , leading among other processes also to B meson production, is about 75 mb [64–68]. The number of inelastic collisions per bunch crossing follows a Poisson distribution with a mean of:

$$\mu = \frac{N_{pp}}{N_{bx}} = \frac{\hat{\mathcal{L}} \cdot \sigma_{pp}^{inel}}{\nu_{filled}}, \quad (2.2)$$

where N_{bx} is the number of filled bunch crossings, and N_{pp} the number of inelastic pp collisions.

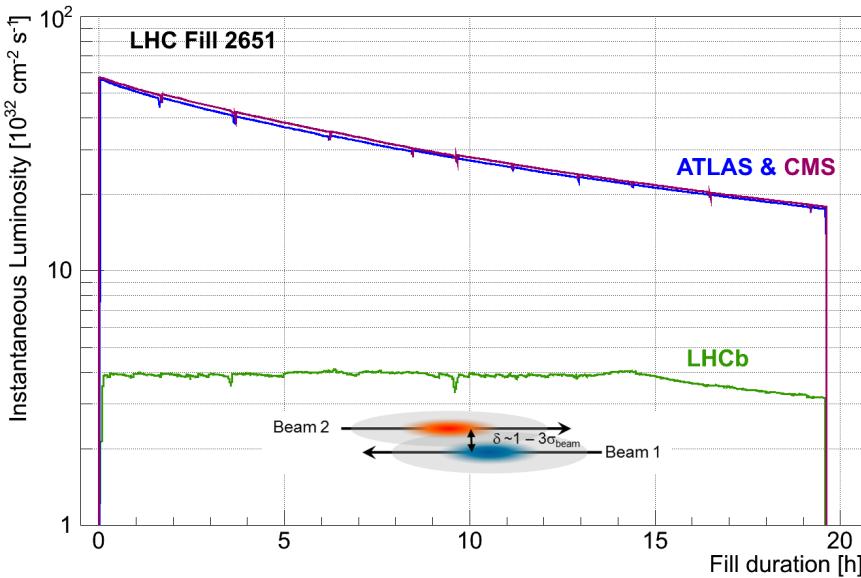


Figure 2.2: Instantaneous luminosity for ATLAS, CMS, and LHCb during an LHC fill of 20 hours. After reaching the desired luminosity value of $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, the LHCb luminosity is kept stable within 5% for about 15 hours by adjusting the transverse beam-overlap. During the last 5 hours, the beam overlap adjustments are not enough to compensate for the reduced number of protons.

2.1.1 B meson production

In the course of 2011 and 2012, more than 800 billion $b\bar{b}$ quark pairs were produced in the LHCb interaction point. A typical B_s^0 meson detected in LHCb had a momentum of about $70 \text{ GeV}/c$, with a transverse component of about $4 \text{ GeV}/c$. Because of the high boost factor, γ , of about 15, a B_s^0 meson travels around 6 mm in the detector before decaying. The dominant processes leading to $b\bar{b}$ pair production in pp collisions are shown in Fig. 2.3. As shown in Fig. 2.4, the $b\bar{b}$ pairs are produced forward or backward, and one in every four $b\bar{b}$ pairs lies in the LHCb detector acceptance, which is coloured red. A review of b -hadron production in LHCb is given in Ref. [69].

The production rate of B mesons (and b hadrons in general) is given by the $pp \rightarrow b\bar{b}$ cross-section times the probability that a b quark forms a particular B meson. These probabilities describe the process in which a primary b (or

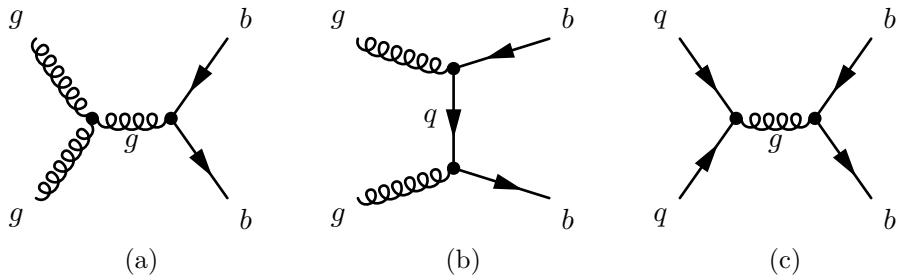


Figure 2.3: Main $b\bar{b}$ pair production processes in proton-proton collisions. Dominant contributions arise from the gluon-fusion diagrams (a) and (b); the quark-quark annihilation contribution (c) is small.

\bar{b}) quark pairs with other quarks into a color neutral object, a hadron, and are called “hadronisation” fractions: $f_{s,d,u}$, respectively for the B_s^0 , B^0 , and B^\pm mesons.

Hadronisation is a strong process occurring at energy scales of $\mathcal{O}(1 \text{ GeV})$. Large strong coupling constant values at such low scales complicate the theoretical calculations; the hadronisation fractions for B mesons and b baryons are determined experimentally. The relative sizes of the measured hadronisation fractions are shown in Fig. 2.4b. The most precise experimental values are achieved by combining [70] the measurements from LEP [71–78], all using $e^+e^- \rightarrow Z \rightarrow b\bar{b}$ decays, with the measurements from CDF [79–81], which use $b\bar{b}$ pairs from $p\bar{p}$ collisions, and the measurement from LHCb [82], which use $b\bar{b}$ pairs from pp collisions:

$$f_{u,d} = (40.1 \pm 0.7)\%, \quad (2.3)$$

$$f_s = (10.7 \pm 0.5)\%, \quad (2.4)$$

$$f_{baryon} = (9.1 \pm 1.5)\%. \quad (2.5)$$

The baryon hadronisation fraction in Eq. (2.5) can be subdivided as $f_{baryon} = f_{\Lambda_b^0} + f_{\Sigma_b^0} + f_{\Sigma_b^-} + f_{\Omega_b^+}$.

2.2 The LHCb detector and tracking

The LHCb detector is a spectrometer in the forward region. It covers a range from 10 mrad to 300 (resp. 250) mrad in the bending (resp. non-bending) plane (see Fig. 2.5). The LHCb detector design and performance is described in detail in Refs. [83, 84]. This section gives an overview of the common variables used to describe the reconstructed candidate tracks and the components of the LHCb detector systems.

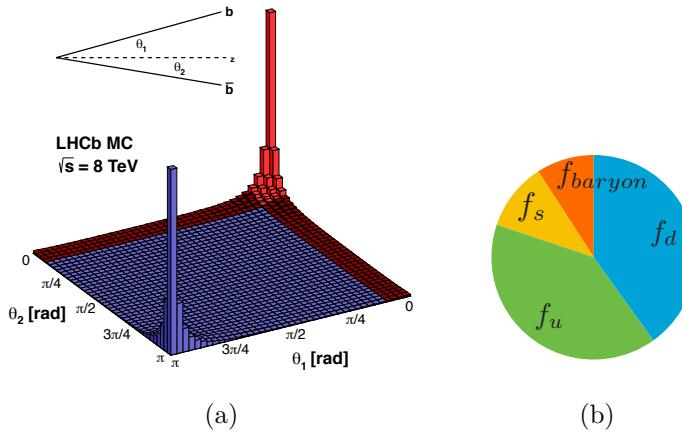


Figure 2.4: (a) Angular distribution of the produced $b\bar{b}$ pairs, with respect to the proton beam direction. The LHCb detector acceptance is shown in red. (b) Relative sizes of the B meson and b baryon hadronisation fractions [70].

The particle tracks reconstructed in LHCb are described in a right-handed coordinate system, with z axis pointing along the beam. The direction of the emerging particles is described by the “pseudo-rapidity”, defined as:

$$\eta \equiv -\ln(\tan(\theta/2)). \quad (2.6)$$

This is preferred to the polar angle θ as the differences in pseudo-rapidity are independent of the reference frame and the measured angular distribution in η will preserve its shape in any frame (in proton collisions, the proton constituents collide with very different momenta, and thus, in different centre-of-mass frames). The pseudo-rapidity is zero for particle tracks perpendicular to the beam, and approaches infinity for tracks parallel to the beam. The region close to the beam and with high η values is referred to as “forward” region. The LHCb detector covers the pseudorapidity region $1.9 < \eta < 4.9$.

The region where the two protons collide is called the “interaction region”. The actual pp collision point, called the “primary vertex” (PV), can vary from event to event and is reconstructed from the particle tracks originating from the interaction region. An unstable particle produced in the primary vertex decays in a “secondary vertex”. The closest distance between the particle track and the associated primary vertex, called the “impact parameter” (IP), is often used to identify particles originating from the secondary vertices. The transverse component, p_T , of the reconstructed particle momentum, p , (the

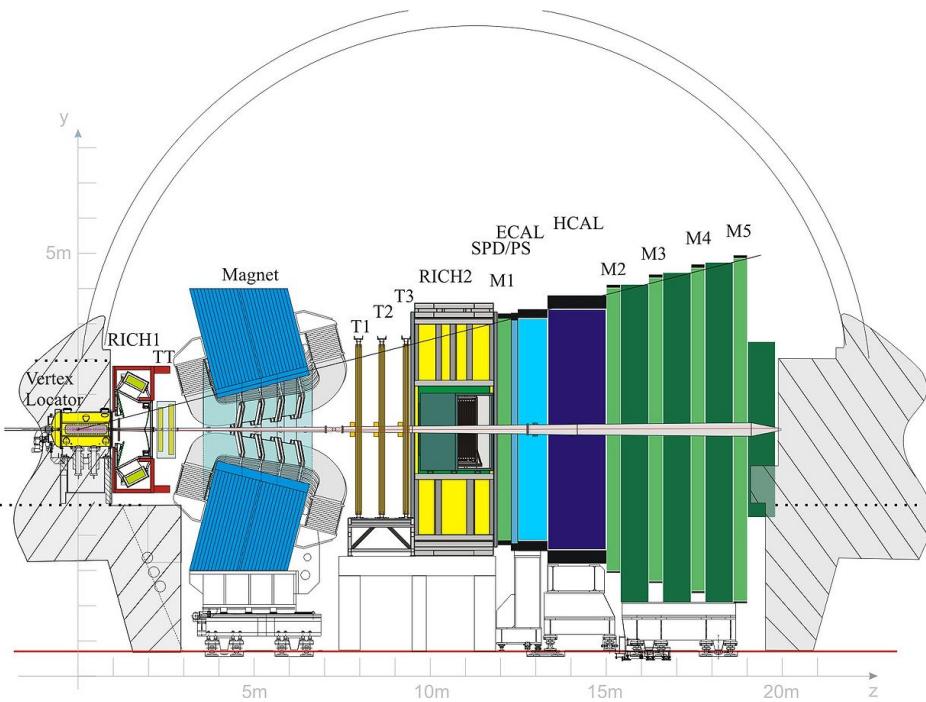


Figure 2.5: A side-view of the LHCb detector, showing the different sub-detectors. The right-handed coordinate system adopted has the z axis along the beam and the y axis along the vertical.

transverse direction is defined perpendicular to the beam axis) is also used to select hard scattering processes.

According to the main purpose, the detector elements can be grouped in two:

- Tracking system: Vertex Locator (VELO), Tracker Turicensis (TT), the dipole magnet, and three tracking stations (T1-T3).
- Particle identification (PID) system: Ring Imaging Cherenkov detectors (RICH1 and RICH 2), electromagnetic and hadronic calorimeters (ECAL and HCAL), and the muon detector (M1-M5).

The goal of the tracking system is to reconstruct the tracks of the charged particles traversing the magnetic field region and the tracking stations. The principal component of the magnetic field is vertical, created by two symmetric custom-made saddle-shaped coils, and the polarity of the magnet is changed after every two weeks of taking data to be able to control detection

asymmetries in CP violation measurements [85]. The magnetic field integral for tracks passing through the entire tracking system is $\int B dl = 4 \text{ Tm}$.

The sub-detectors close to the beam-pipe, i.e. VELO, TT, and the innermost part of T1-T3 (Inner Tracker, IT), are silicon micro-strip detectors, whereas the Outer Tracker (OT) in the region where the occupancy is lower, employs gas straw tubes. The dimensions of the silicon strips are chosen such that occupancies are sufficiently low, while the signal-to-noise ratio is greater than 10. The VELO, TT, IT, and OT resolutions and the magnetic field strength lead to an overall charged particle momentum resolution $\delta p/p \sim 0.5\%$ at $p = 20 \text{ GeV}/c$, and $\delta p/p \sim 0.8\%$ at $p = 100 \text{ GeV}/c$. This in turn provides a B mass resolution of typically $25 \text{ MeV}/c^2$ for the di-muon modes, and a B meson decay time (τ_B , defined in the meson frame) resolution of about 50 fs.

The strategies behind particle identification and its performance estimation will be discussed in detail in Sec. 2.5. The following will describe the LHCb sub-detectors.

Vertex Locator (VELO)

The VELO comprises a series of silicon modules positioned perpendicularly to the beam line close to the interaction point, as shown in Fig. 2.6. It is used to identify the primary pp interaction vertices and the secondary vertices of the b -hadron decays with high proper-time resolution. The modules are positioned to detect the majority of the tracks originating from the interaction region. In order to allow fast reconstruction of tracks and vertices in the LHCb trigger, the modules have a cylindrical geometry with silicon strips measuring $r\phi$ coordinates. Each VELO module consists of two $300 \mu\text{m}$ thick half-disc silicon sensors, called r and ϕ -sensors (Fig. 2.6). The r -sensor has circular strips segmented into four sectors around the beam axis measuring the r coordinate of the hit; the ϕ -sensor has (almost) radial strips measuring the azimuthal ϕ -coordinate of the hit. The sensor is divided into inner and outer parts in order to reduce the strip occupancy and overlap between the inactive regions. The strips in both regions are at an angle of 5° with respect to the radius. The set of 21 modules on each side of the beam is housed in a box made out of thin aluminium separating the beam vacuum from the VELO vessel vacuum and also shielding the detector from the radiation near the beam.

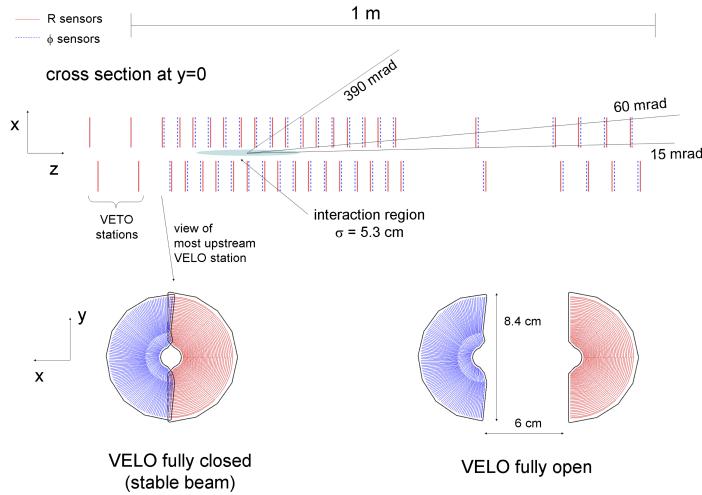


Figure 2.6: The set-up of the VELO silicon modules along the beam line (up) together with the layout of the r (bottom right, red) and ϕ (bottom left, blue) sensors in closed and open positions. Indicated are the average crossing angle for minimum bias events (60 mrad), and the minimal (15 mrad) and maximal (390 mrad) angle for which at least 3 VELO stations are crossed.

The Tracker Turicensis (TT) and the Inner Tracker (IT)

The TT stands between the VELO and the dipole magnet. It is used to improve the momentum estimate for charged particles and to perform track measurements of long lived neutral particles (in particular longer living particles as K_S and Λ mesons). The TT consists of four layers of silicon sensor ladders separated by a 27 cm gap into two pairs: (x,u) and (v,x). For better track reconstruction, the strips are vertical in the first and last layer (x), whereas the other two layers (u,v) are rotated by stereo angles of $\pm 5^\circ$, as shown in Fig. 2.7. The dimensions of the TT are chosen to cover the full detector acceptance (150 cm wide and 130 cm high).

The IT is located behind the magnet, at the centre of the tracking stations T1-T3. It provides accurate momentum estimates, important for the invariant mass and decay-time resolutions. The three 120 cm wide and 40 cm high IT stations (Fig. 2.7) consist of four boxes arranged in the cross shape around the beam pipe. Each box contains four layers in the $x-u-v-x$ stereo angle configuration.

The silicon sensors of IT and TT are single sided p⁺-on-n sensors. The TT sensors are 500 μm thick, 9.64 cm \times 9.44 cm long and carry 512 readout strips

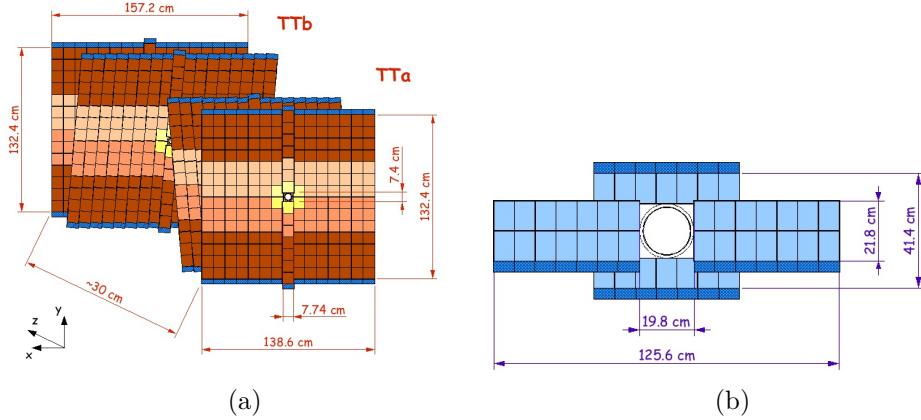


Figure 2.7: (a) The x - u - v - x geometry of the TT. (b) A cross-section of the TT x -layer.

with a strip pitch of $183\text{ }\mu\text{m}$. In case of IT, the sensors are $7.6\text{ cm} \times 11\text{ cm}$ long and $310\text{ }\mu\text{m}$ (for one-sensor modules) or $410\text{ }\mu\text{m}$ (for two-sensor modules) thick. Each sensor has 384 readout strips with strip pitch of $198\text{ }\mu\text{m}$. The spatial resolution is about $50\text{ }\mu\text{m}$ for TT and $60\text{ }\mu\text{m}$ for IT.

2.2.1 The Outer Tracker (OT)

The OT consists of gaseous straw-tube detectors and covers the outer region of the tracking stations T1-T3. The layers are arranged with the same geometry as the TT and IT: x - u - v - x , and designed as arrays of individual, gas-tight straw-tube modules, as shown in Fig. 2.8. Each module contains two layers of drift-tubes with an inner diameter of 4.9 mm. Given the bunch crossing rate of 25 ns and the diameter of the tube, and in order to guarantee a fast drift time (below 50 ns and a sufficient drift-coordinate resolution ($200\text{ }\mu\text{m}$), a mixture of Argon (70%), CO₂ (28.5%), and O₂ (1.5%) is used as counting gas.

The front-end (FE) electronics measures the drift time of the ionisation clusters produced by charged particles traversing the straw tubes. The drift times allow to determine the hit positions in the OT.

2.2.2 Cherenkov detectors (RICH1 and RICH2)

The Ring Imaging Cherenkov (RICH) system distinguishes between kaons, pions, protons, and muons. The system is composed of two detectors which, when combined, cover the full momentum range. The upstream detector

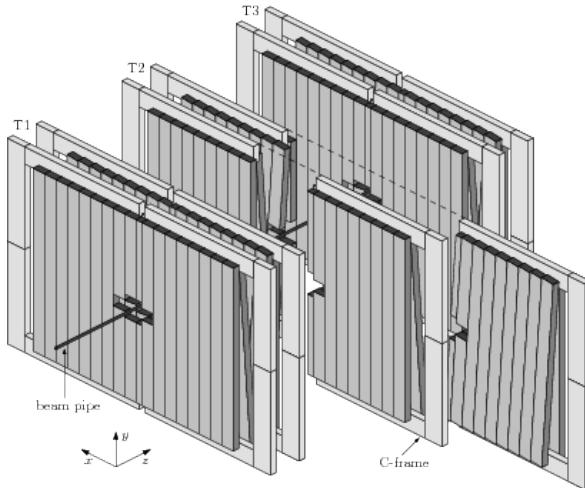


Figure 2.8: The three OT stations, the C-side of the middle station is shown in its open position.

(RICH 1) covers the low momentum particle range ($1 - 60 \text{ GeV}/c$), using aerogel and C_4F_{10} radiators. The downstream detector (RICH 2) covers the high momentum range from $15 \text{ GeV}/c$ up to and beyond $100 \text{ GeV}/c$, using a CF_4 radiator.

Both detectors use the same principle to determine the particle type: when particles traverse a medium with a speed larger than the macroscopic speed of light in the same medium, Cherenkov radiation in the shape of a cone is emitted. To analyse the emitted Cherenkov radiation, the cones of emitted light are reflected by spherical and flat mirrors, as shown in Fig. 2.9, for RICH 1. Hybrid Photon Detectors, detecting the circular projection of the reflected cones, are placed outside the detector acceptance.

Measuring the angle of the Cherenkov cone (θ_C) and knowing the refractive index of the medium (n) allows to determine velocity of the particle (v) relative to the speed of light ($\beta = v/c$): $\cos(\theta_C) = 1/(n\beta)$. Combining this with the momentum information from the tracking system one can calculate the mass and determine the particle type.

2.2.3 The Muon system (M1-M5)

The muon identification is a fundamental requisite to trigger events with muons in the final state and perform the offline selection of rare $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. The muon system, shown in Fig. 2.10, comprises five stations (M1-M5) of rectangular shape. The stations are placed along the beam axis at the

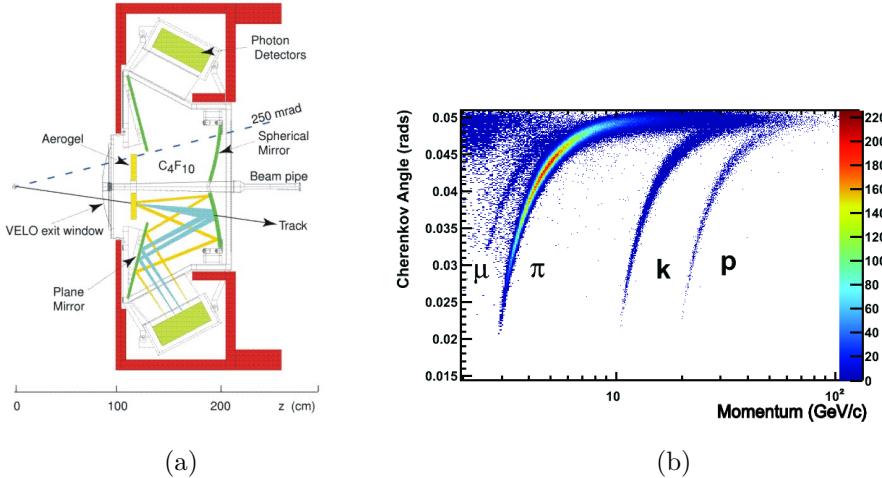


Figure 2.9: (a) A schematic layout of the RICH 1 detector. (b) Graph of the Cherenkov angle (θ_C) depending on the momentum of different particles. Combining the information from the RICH system with the momentum information provided by the tracking system, provides the particle mass.

rear end of the LHCb spectrometer. Their transverse dimensions scale with the distance from the interaction point, such that the full detector acceptance is covered. Stations M2-M5 are placed downstream of the calorimeters and are interleaved with 80 cm thick iron absorbers to select penetrating muons; station M1 is placed in front of the calorimeters.

The first level muon trigger uses the muon station hits for quick muon candidate p_T estimate. Stations M1-M3 have a high spatial resolution along the x coordinate; they are used to separate between positive and negative candidates and to calculate the p_T of the candidate muon in the first level muon trigger. Stations M4 and M5 have a limited spatial resolution, and their main purpose is to tag false muon candidates.

The layout of a muon station is shown in Fig. 2.10. Multi-wire proportional chambers (MWPC) are used in all regions, except in the innermost region of station M1, where the occupancy is highest and where triple-GEM detectors are used. Each muon station is divided into four regions (R1-R4) with increasing distance from the beam axis. The linear dimensions of the regions and their segmentation scale in the ratio 1:2:4:8 are such that the particle flux and channel occupancy are roughly the same over the four regions of a given station.

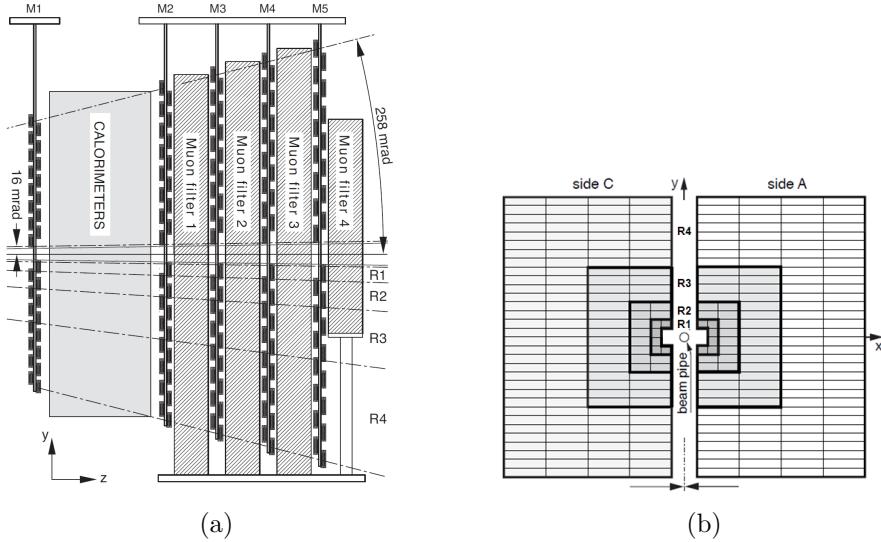


Figure 2.10: (a) The set-up of the muon stations (side view). (b) Layout of one muon station.

2.2.4 The Calorimeter system (CALO)

The calorimeter system identifies electrons, photons, and hadrons. The energy deposits it measures are also used in the first trigger level to trigger on particles with large transverse energy deposits. The transverse energy is computed for a 2-by-2 cell cluster as

$$E_T = \sum_{i=1}^4 E_i \sin \theta_i, \quad (2.7)$$

where E_i is the energy deposited in the i -th calorimeter cell, and θ_i the angle of the cell with respect to the beam.

In increasing distance from the interaction point, the calorimeter system consists of two scintillator layers, the Scintillator Pad Detector (SPD) and the Pre-Shower (PS), and two calorimeters, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). Because the hit density varies by two orders of magnitude over the calorimeter surface, the SPD, PS, ECAL, and HCAL adopt a variable lateral segmentation (Fig. 2.11). The ECAL is segmented into three different sections. The SPD and PS detector segments match the ECAL segments, when looked from the interaction point. The HCAL is segmented into two zones and has larger cells to accommodate for the wider hadronic showers.

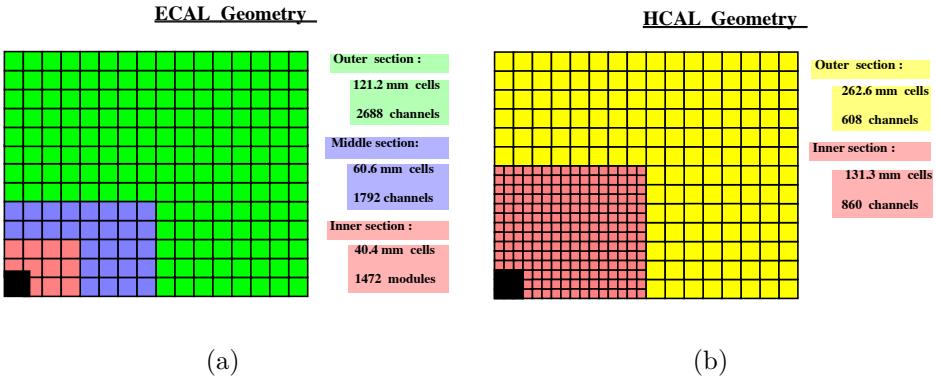


Figure 2.11: (a) Segmentation of the LHCb calorimeter system. SPS, PS, and ECAL, cells are shown for a single quadrant in (a), whereas the cell sizes for SPD and PS are 1.5% smaller. (b) The HCAL segmentation is coarser, to account for the wider hadronic showers.

The SPD and PS consist of rectangular scintillating pads. A thin layer of lead, with a thickness of 2.5 radiation lengths¹ (X_0), is inserted between the two detectors to initiate a photon shower and improve the separation between electrons and photons. Hadrons are not likely to deposit energy in the SPD/PS. The ECAL and HCAL have both a sampling scintillator/lead (active-layer/absorber) structure. The lead layers used in ECAL amount to a total thickness of $25 X_0$. The HCAL absorber layers are made of iron, and the thickness of HCAL is limited to 5.6 interaction lengths² (λ_i) because of space constraints. In all cases, the light produced in the scintillators is transmitted to photomultiplier tubes (PMT) by optical fibres.

2.3 Overview of the LHCb trigger

Every analysis in LHCb begins with a decision whether to store the event or not. The decision must be taken right after the pp collision has taken place. The events that are kept for storage are called “triggered” events, and the trigger is the system designed to make the decision. The particles produced in the detector acceptance can deposit energy in the detection cells and produce so-called “hits”. A collection of all the recorded hits forms the event information, typically between 50 and 100 kbytes per event.

¹A radiation length is both the mean distance over which a high-energy electron loses all but $1/e$ (0.368) of its energy by bremsstrahlung, and $7/9$ of the mean free path for pair production by a high-energy photon.

²Nuclear interaction length is the mean path length required to reduce the numbers of relativistic charged particles by the factor $1/e$ (0.368) as they pass through matter.

In 2011 (resp. 2012) the average number of visible pp interactions³ per bunch crossing in LHCb, ν , was 1.5 (resp. 1.6). The rate of visible collisions processed by the trigger was on average 13 MHz (resp. 13.5 MHz) in 2011 (resp. 2012). The LHCb trigger has to reduce this rate to 2-5 kHz, at which the events can be stored (about 500 Mbytes/s).

In the 2011 and 2012 running conditions approximately 600 000 $c\bar{c}$ and 30 000 $b\bar{b}$ pairs are produced in the LHCb acceptance every second. Nine times out of ten, a \bar{b} quark hadronises into a B meson, which decays to various final states of which only a fraction are interesting; e.g. the typical branching fractions of B meson decays used to study CP violation are less than a per mille (and orders of magnitude less for rare decays). The trigger is optimised to keep the interesting B meson decays and reject uninteresting background events in early stages.

The LHCb trigger system is designed in two stages: the Level-0 (L0) and the High Level Trigger (HLT). Because there is not enough time to read out and process all the detector information for every single event, the L0 makes a quick preliminary selection based on the information only from the fastest sub-detectors: the calorimeters (Sec. 2.2.4) and muon stations (Sec. 2.2.3). At the L0 output rate of about 1 MHz, the HLT is then able to process the complete event information.

The stored events are later processed with a more elaborate reconstruction software, that uses accurate sub-detectors alignment and calibration parameters. This later stage of reconstruction and the following event selection will be referred to as the *offline* reconstruction and selection.

2.3.1 Level-0 trigger

The first level of the LHCb trigger (L0) is implemented in custom-designed hardware, designed to take exactly 4 μ s (latency) to analyse the event information. This latency includes the time of flight of the decay products, the cable length, and all delays in the front-end (FE) electronics, which leaves 2 μ s for the actual processing. The events that take place during the processing time are kept in pipeline memories.

In 4 μ s, L0 produces a decision for each of the three independent L0 sub-triggers: L0-Muon, L0-Calorimeter, and L0-PileUp trigger. If any of the three sub-trigger decisions is positive, the event is passed to the next trigger level; otherwise the event is not processed further and is cleared from memory. The hits in the five muon stations are processed by the L0-Muon sub-trigger. Four

³A visible interaction is defined as one in which at least two reconstructed tracks in LHCb point to the interaction region.

processors, one for each quadrant, look for two muon tracks in their quadrants with the largest and the second largest transverse momentum. Each processor searches for hits that define a straight line through the five muon stations while pointing towards the interaction point (correcting for magnetic field effects). Once a suitable track (candidate) is found, its hits in the first two stations where the lateral resolution is higher are used to estimate the track p_T . The resolution of the p_T estimate is around 20%, which is enough for the L0 purpose (see Fig. 2.12). If any of the muon track candidates passes a p_T threshold, the event is triggered with a single muon candidate LOMuon decision; in case the candidates with the largest and the second largest p_T pass a combined threshold on $p_T^{largest} \times p_T^{2ndlargest}$, the event is triggered by an LODiMuon decision (see Sec. 3).

The L0-Calorimeter sub-trigger uses information from SPD, PS, ECAL, and HCAL. These four detector layers are positioned in increasing distance from the collision point. The different penetration power of different particles allows to separate between photon, electron, and hadron showers (Sec. 2.5). The calorimeter system computes the transverse energy (see Sec. 2.2.4) deposited in ECAL and HCAL clusters of 2×2 cells. From these clusters, three types of candidates are built:

- *hadron* candidate (**L0Hadron**) is the highest- E_T HCAL cluster;
- *photon* candidate (**LOPhoton**) is the highest- E_T ECAL cluster, with 1–2 PS cell hits in front of the ECAL cluster, but with no hits in the corresponding SPD cells;
- *electron* candidate (**LOElectron**) has the same requirements as **LOPhoton**, but with the additional requirement of at least one SPD cell hit in front of the PS cells.

The E_T of each candidate is compared to a given threshold and events containing at least one candidate above threshold are accepted by the L0.

Besides the L0-Calorimeter and L0-Muon requirements, an additional veto is applied on the number of SPD hits to exclude events with numerous tracks that are much more likely to contain random combinations passing our signal selection criteria and would take disproportionately large fraction of the processing time available in the HLT. The exact criteria are described in Sec. 3.

2.3.2 The High Level Trigger

The events accepted by the L0 trigger are further processed by the HLT. At the L0 output rate of about 1 MHz, the HLT is able to read out the full

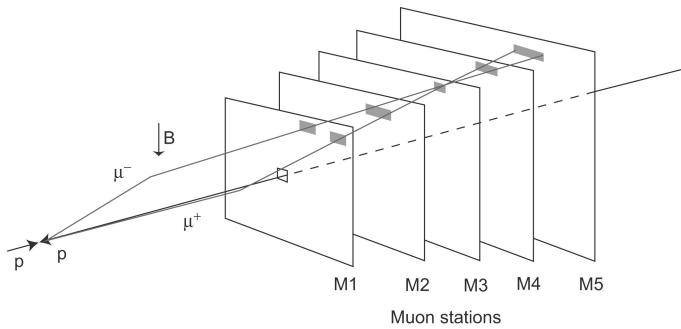


Figure 2.12: A schematic view of muons changing direction in the magnetic field and passing through the five muon stations (M1-M5), shown for magnet-down polarity. The lateral resolution of M1 and M2 is higher and is used for a muon- p_T estimate in the first trigger level. The hits in the last muon stations help to select true muon candidates.

event information. The HLT is a software (mainly in C++) running the same algorithms that are used later in offline stage for reconstructing the triggered events. The HLT program runs in 26 110 copies, simultaneously on a farm of multiprocessor PCs called Event Filter Farm (EFF). The offline reconstruction takes about 2 seconds per event but the parallel computing resources of the EFF allow about 30 ms and therefore the HLT application has to be simplified to meet the time requirements. This is achieved in two steps, referred to as HLT1 and HLT2. In the following, the term *trigger line* refers to a specified sequence of reconstruction algorithms and selections in a particular trigger level.

HLT1

The HLT1 is based on the VELO reconstruction software. It is fast enough to run the full 3D pattern recognition for all the L0-accepted events. At the start of each LHC fill, the mean position of the pp interactions in the $x - y$ plane (PV^{mean}) is determined using the VELO tracks. This position remains stable within a few micrometers during the fill. The VELO tracks are used to construct vertices with at least 5 tracks originating from them. Vertices within a radius of 300 μm around the PV^{mean} are considered as primary vertices (PV).

The pattern recognition in HLT1 limits the execution time by selecting only VELO tracks with high probability to originate from a signal decay. For events triggered by the L0 muon lines, the fast muon identification algorithm

defines a search window in the central muon station, by extrapolating each VELO track in a straight line to the station M3. The magnet does not bend the particle trajectories in the *vertical plane* and therefore it is multiple scattering from the detector material that dominates the vertical size of the search window. The horizontal search window size is defined by the deflection of a muon with a $6 \text{ GeV}/c$ momentum⁴. Hits found in the search window are then combined with the VELO tracks to search for hits in the other muon stations, M2, M4, and M5. A candidate track is provisionally accepted if it contains at least one hit in addition to the M3 hit. In the *horizontal plane*, the VELO track together with all the muon hits is used in a simplified track fit. As soon as a muon candidate track with a χ^2 per degree of freedom smaller than 25 is found, the algorithm selects the VELO track and stops. HLT1 trigger lines that do not require muons (e.g. `Hlt1TrackAllL0`), select tracks with large enough impact parameter with respect to any PV. The quality of the VELO tracks is further assessed using the measured VELO hits; only good quality tracks are selected.

For the selected VELO tracks, the hits in the tracking stations downstream the magnet (OT and IT) are used to determine the track momentum. This process is called *forward tracking* [86]. The online algorithm used by the HLT1 is similar to the offline one, but, to run faster, considers tracks with enough momentum and transverse momentum.

Tracks are fitted using a Kalman filter [87, 88] track fit. In this process, the HLT1 uses a simpler detector material geometry and fewer iterations than in the offline track fit. Studies on the $J/\psi \rightarrow \mu^+\mu^-$ decays show that the invariant mass resolution in the HLT1 is only 3% larger than that obtained offline ($14 \text{ MeV}/c^2$), and is sufficient to allow selection cuts in IP, momentum, and mass.

Around 38 different HLT1 trigger lines select events based on their track variables (IP, number of VELO hits, number of OT and IT hits, track IP χ^2 , track p , track χ^2 per degree of freedom, etc.) For muon candidate tracks, HLT1 also applies the offline muon identification algorithm to improve the purity of the muon sample. The rate of accepted events after HLT1 was 43 kHz (resp. 80 kHz) in 2011 (resp. 2012).

HLT2

HLT2 can perform the forward tracking for the majority of the VELO tracks. To reduce the size of the IT and OT search windows, most HLT2 lines consider only VELO tracks with high enough p or p_T . There are only a few

⁴The layers of iron between the muon stations absorb most of the muons with momentum smaller than $6 \text{ GeV}/c^2$.

simplifications with respect to the offline reconstruction algorithms, and they lead to 1–2% lower reconstruction efficiency per track. The 130 HLT2 trigger lines can be classified into two groups: *exclusive trigger lines* target specific final states and require all final state particles to be reconstructed; *inclusive trigger lines* trigger on b-hadron decays with at least two charged particles in the final state and a displaced decay vertex (*topological* inclusive lines can also trigger on partially reconstructed b decays). The total HLT2 output rate of 5 kHz in 2012 (3 kHz in 2011) can be split among three contributions:

- *Generic beauty trigger lines* take the majority of the bandwidth. These are mostly inclusive trigger lines designed to trigger on partially reconstructed b-hadron decays with at least two reconstructed charged tracks (topological lines). Examples of decays triggered by generic beauty lines include $B^0 \rightarrow D^+ \pi^-$, $B^\pm \rightarrow D^0 \pi^\pm$, $B_{(s)}^0 \rightarrow h^+ h^-$.
- *Muonic trigger lines* select events with one or two identified muons in the final state, e.g. $B^+ \rightarrow J/\psi K^+$. HLT2 applies the offline muon identification algorithm (see Sec. 2.5.1) to all tracks from the forward tracking. Single muon candidates are selected if the muon passes a p_T threshold or if its origin is sufficiently separated from the PV. The di-muon candidates are accepted if their invariant mass is compatible with a particular mother particles mass (e.g. J/ψ). There are separate lines for prompt and detached J/ψ (the prompt line keeps only a pre-determined fraction of all selected events because of the high rate).
- *Charm trigger lines* require all the final state particles to be reconstructed, and because of high charm production rate, include only one inclusive trigger line: a line selecting the decay chain $D^{*+} \rightarrow D^0 \pi^+$. The inclusive rate is reduced by requiring that the two charged tracks from the D^0 decay match to a slow pion from the D^{*+} . The main exclusive trigger lines for prompt charm selection (around 30 trigger lines) are hadronic two- and three-body lines designed to trigger on $D^0 \rightarrow K^- \pi^+$ and $D^+ \rightarrow K^- \pi^+ \pi^-$ decays. These lines accept candidates based on the track and vertex quality, and the mass of the parent D mesons.

2.4 Offline reconstruction

The information from the triggered events is stored on tape at the rate of 500 MB/s. This includes the complete hit information from the detector cells and the trigger decision reports, important for the trigger efficiency determination (Ch. 5).

The offline reconstruction algorithms use the recorded information to reconstruct the particle tracks and the decay vertices and to determine their energy and momentum. Compared to the online (trigger) reconstruction algorithms, the offline reconstruction uses better calibration of the sub-detectors; the extra time available allows to use multiple strategies, which all together improves the offline reconstruction efficiency by 1 – 2% per track, when compared to the online reconstruction. The hit information is processed by the reconstruction algorithm BRUNEL [89] and the High Level Trigger (HLT) application MOORE [90]. BRUNEL reconstructs the tracks, identifies the particles, and passes the track and particle objects on to the event selection application, DAVINCI [91], where they are combined into particles and vertices.

2.5 Particle Identification

A typical pp collision produces many particles (typically around 100) in the LHCb acceptance. Pions are the most abundant: in $b\bar{b}$ inclusive events they represent about three out of four charged particles reconstructed as tracks traversing the full tracking system (from VELO to IT and OT). Other tracks belong to kaons (15%), protons (4%), electrons (6%), and muons (less than 1%). Different sub-detectors contribute to identifying these particles: the RICH distinguishes between charged hadrons (K , π , and p , see Sec. 2.2.2); the calorimeter system identifies electrons, photons, and neutral pions (see Sec. 2.2.4); the muon detector identifies muons (see Sec. 2.2.3). To a lesser extent, the RICH also adds to the discrimination of leptons from hadrons.

2.5.1 Muon identification

This section focusses on the strategy and performance of muon identification, a crucial ingredient of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis. The LHCb muon identification relies on three approaches [92]:

- **Muon penetration power.** Muons lose their energy predominantly due to ionisation and are able to transverse the calorimeters as well as the iron shields between the muon stations. Electrons are lighter than muons and lose their energy by radiating photons (through bremsstrahlung in the electric field of nuclei) and are stopped by the ECAL. Hadrons, just as muons, are too heavy for considerable bremsstrahlung radiation but do interact strongly with nuclei and thus rarely reach beyond the HCAL (Fig. 2.13).

- **Muon hit-pattern in the muon stations.** Muon station hits from true muons that originate from the interaction region tend to be closer to the track prediction (based on the hits in the tracking station) than those due to kaons and pions decaying into muons (decays in flight). Also, in the region of high occupancy, hits caused by muons, other particles, or detector effects can be wrongly associated to hadron tracks. These, however, exhibit in general a spread of hits larger than the true muon tracks.
- **Muon mass.** The Cherenkov radiation angles measured in the RICH can be combined with the track momentum measurement to give an estimate on the charged particle masses. This works especially well in separating slower muons (muon momentum below $5 \text{ GeV}/c$, see Fig. 2.9) from kaons and protons.
- **Calorimeter energy deposits.** Muons deposit only little energy in the calorimeter and in a proximity to the extrapolated track. Electrons, on the other hand, deposit all their energy in SPD, PS, and ECAL, and affect a much wider area. Hadrons deposit their energy predominantly in HCAL and over a larger calorimeter area.

The LHCb muon identification algorithms use all the aforementioned identification strategies and are applied to all tracks. The high penetration power of muons is exploited by a loose binary muon selection algorithm, called `isMUON`. `isMUON` requires muon station hits in the field of interest (FOI) that is defined around the extrapolated muon candidate track. The size of a FOI is defined for different momentum ranges depending on the expected multiple scattering suffered by the muons before the muon stations. The track momentum also determines which muon stations are required to be hit for a positive `isMUON` decision, as shown in Tab. 2.1.

Muon identification and mis-identification are further improved by including all the muon identification strategies (Sec. 2.5). Individual muon

Table 2.1: The muon station hits required by the `isMUON` criteria for different track momentum ranges.

Particle momentum	Required muon station hits
$3 < p < 6 \text{ (GeV}/c)$	M2 and M3
$6 < p < 10 \text{ (GeV}/c)$	M2 and M3 and (M4 or M5)
$p > 10 \text{ (GeV}/c)$	M2 and M3 and M4 and M5

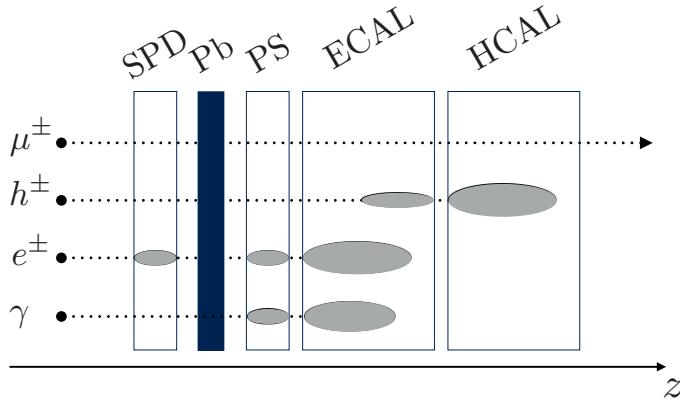


Figure 2.13: A schematic view of the energy deposits in LHCb calorimeter layers, shown for a muon, hadron, electron, and photon. Muons are minimum ionising particles and traverse the calorimeters.

and non-muon likelihoods are computed for each muon candidate, using the information from muon stations, RICH detectors, and different calorimeter layers.

Muon detector likelihoods are computed using the muon station hit patterns, specifically the distribution of the average squared hit distance significance (D^2) with respect to the extrapolated track. The likelihood for the muon hypothesis of a track (the muon likelihood) is calculated by integrating over the muon D^2 probability density function from 0 to the measured value, D_{cand}^2 ; the non-muon likelihood is calculated by integrating over the non-muon D^2 probability density function instead. The D^2 distribution for muon tracks is measured on $J/\psi \rightarrow \mu^+\mu^-$ data using the tag-and-probe method (see Sec. 2.5.2). The D^2 distribution for the other tracks is measured using protons from the simulated $\Lambda^0 \rightarrow p\pi^-$ decays. Because the muon D^2 distribution depends on multiple scattering, it is determined separately in the four muon detector regions (θ intervals) and in bins of muon momentum. For protons, the hits in the muon stations are caused by (i) true muons pointing in the same direction, (ii) random detector hits, or (iii) punch-through protons, and the momentum dependence can be neglected. The muon station likelihoods are combined with the likelihoods provided by RICH and calorimeters. The

RICH likelihoods are calculated for each track, using different particle mass hypotheses. The calorimeters use the size and the spread of energy depositions to build the likelihoods for muon, electron, and hadron tracks.

Muons are identified on the basis of the difference between the combined muon (log-)likelihood, and the combined pion, kaon, or proton (negative log-)likelihoods. The exact choice of the non-muon hypothesis depends on the application. Unless stated otherwise, $\Delta LL(\mu - h)$, will refer to the combined (negative log-)likelihood difference.

2.5.2 Muon identification performance

Knowing the efficiency of the particle identification is crucial for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. The performance of the muon identification algorithms is measured on two-body decay data: the muon identification efficiency is determined on a $J/\psi \rightarrow \mu^+ \mu^-$ sample, the kaon and pion identification efficiencies on a $D^0 \rightarrow K^- \pi^+$ sample, and the proton identification efficiencies on a $\Lambda^0 \rightarrow p \pi^-$ sample.

The tag-and-probe method is applied in all cases. The two daughter tracks in the decays are called “tag” and “probe”, where probe is the muon, pion, kaon, or proton track used for the efficiency estimation. The tag track, together with overall kinematic selection requirements and a maximum likelihood fit on the mass distribution, is used to be sure that one selects the true decay candidate. Only events triggered independently of the probe track are used to remove any bias from the trigger.

In addition to muon, kaon, pion, and proton identification efficiencies, the tag-and-probe method is also used to determine how often non-muon tracks pass the muon identification criteria. This, as will be discussed later in Sec. 6.4, is a dangerous source of background.

The muon identification efficiency of the `isMUON` requirement and the `isMUON` mis-identification probabilities for hadrons depend on the track momentum. The values determined for different momentum ranges are given in Tab. 2.2. For muons with p_T higher than $1.7 \text{ GeV}/c$, the `isMUON` efficiency is above 97% in the whole momentum range ($3 \text{ GeV}/c - 100 \text{ GeV}/c$); hadron tracks are mis-identified as muon in less than a percent of the cases. More than half of the kaon and pion mis-identifications are caused by decays in flight before the muon stations, and are therefore very difficult to suppress. The other causes of hadron mis-identification are punch-through and random muon station hits in the fields of interest. The chance to mis-identify a hadron as a muon is therefore higher in the larger occupancy regions, and also depends on the number of tracks in the event.

The `isMUON` performance for kaon and pion decays in flight is improved by

Table 2.2: The average isMUON (mis-)identification probabilities for 2011 analysis, shown in percentages and together with statistical uncertainties.

Particle p_T interval	$\epsilon(\mu \rightarrow \mu)$	$\epsilon(p \rightarrow \mu)$	$\epsilon(\pi \rightarrow \mu)$	$\epsilon(K \rightarrow \mu)$
$p_T < 0.8$ (GeV/c)		1.393 ± 0.005	6.2 ± 0.1	4.3 ± 0.1
$0.8 < p_T < 1.7$ (GeV/c)	96.94 ± 0.07	0.737 ± 0.003	2.19 ± 0.01	1.93 ± 0.1
$1.7 < p_T < 3.0$ (GeV/c)	98.53 ± 0.05	0.149 ± 0.004	0.61 ± 0.01	0.93 ± 0.01
$3.0 < p_T < 5.0$ (GeV/c)	98.51 ± 0.06	0.12 ± 0.02	0.40 ± 0.01	0.72 ± 0.01
$5.0 < p_T$ (GeV/c)	98.51 ± 0.07		0.33 ± 0.02	0.69 ± 0.01

a requirement on muon and non-muon likelihood difference: $\Delta\text{LL}(\text{K} - \pi) < 10$ and $\Delta\text{LL}(\mu - \pi) > -5$. The kaon, pion, and proton mis-identification probabilities after isMUON and the ΔLL requirements are measured in bins of p and p_T (Fig. 2.14), and the results were found to be independent of the event multiplicity.

2.5.3 Kaon and pion identification

The difference between kaon and pion log-likelihoods, $\Delta\text{LL}(\text{K} - \pi)$, is used in the analysis to separate the $B_{(s)}^0 \rightarrow h^+h^-$ decay modes (used in Ch. 6). The efficiency of the $\Delta\text{LL}(\text{K} - \pi)$ cut for kaons and pions is determined for different thresholds in bins on the track momentum, pseudo-rapidity, and the number of tracks in the event. The RICH detector contributes the most to the kaon-pion separation performance (see Fig. 2.15).

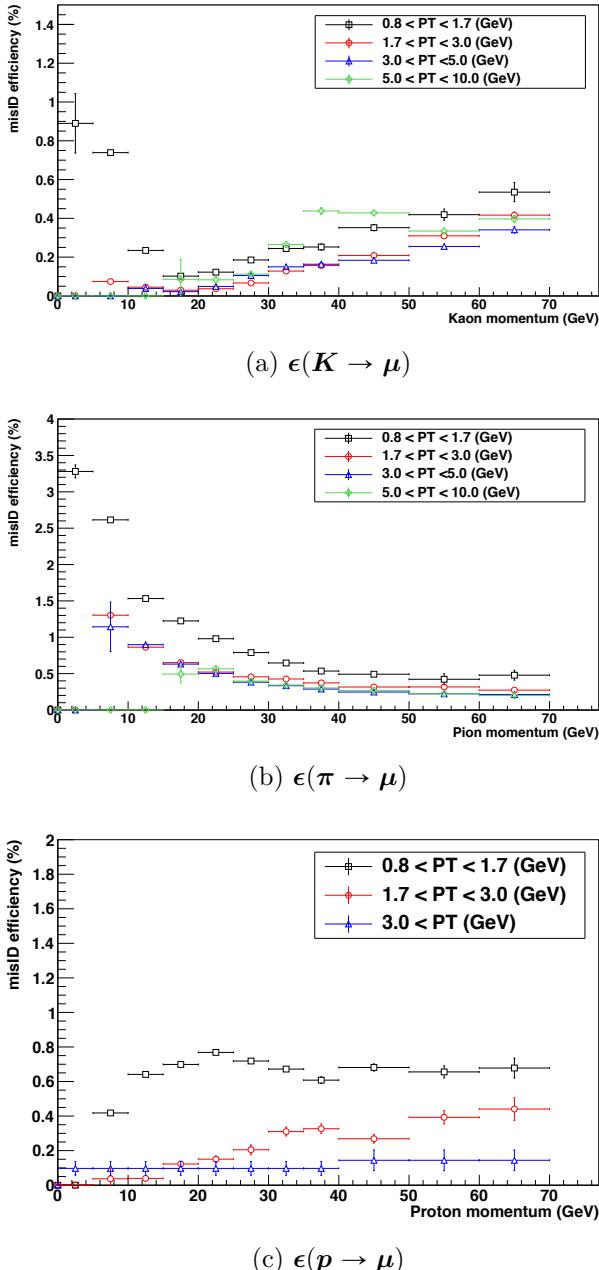


Figure 2.14: The probability to mis-identify a hadron as a muon, depending on the track momentum (2012 data), shown for kaons (a), pions (b), and protons (c). The probabilities were determined from $D^0 \rightarrow K^- \pi^+$ and $\Lambda^0 \rightarrow p \pi^-$ decays, using the isMUON requirement together with $\Delta LL(K - \pi) < 10$ and $\Delta LL(\mu - \pi) > -5$.

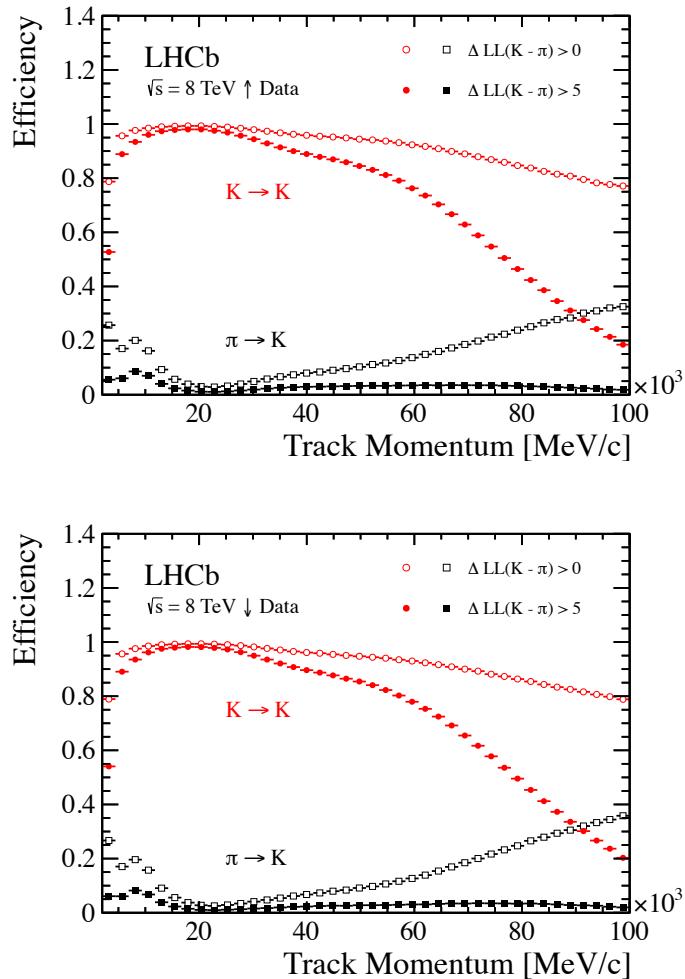


Figure 2.15: Kaon identification efficiency and pion mis-identification rate for two different RICH $\Delta LL(K - \pi)$ cuts, measured from $D^{*+} \rightarrow \pi^+ D^0 (\rightarrow K^- \pi^+)$ decays (2012 data), for magnet-up polarity (top), and magnet-down polarity (bottom).

Part II

LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

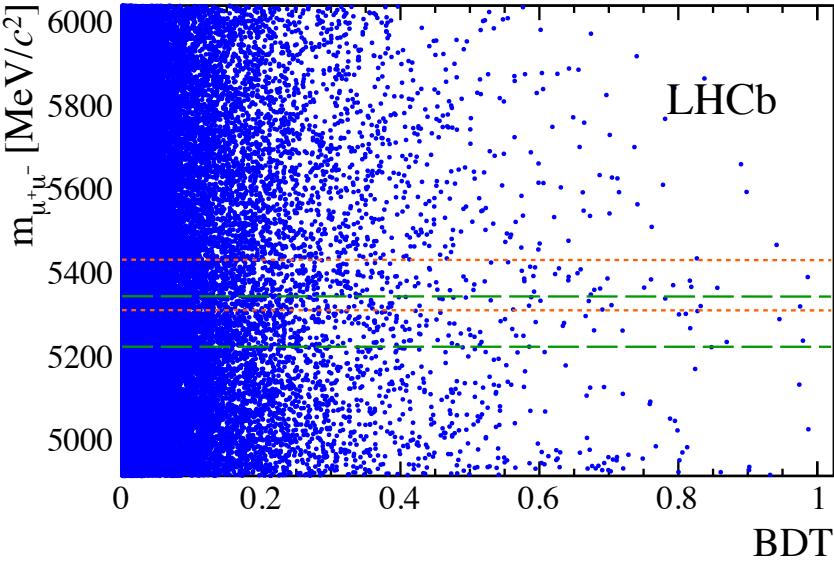
Introduction to Part II

This part of the dissertation describes the measurement of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields, $N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$, and the calculation of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. The data was collected by the LHCb detector at two different proton-proton collision energies: 1018 pb^{-1} at $\sqrt{s} = 7 \text{ TeV}$ (2011), and 2028 pb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ (2012). In addition to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, the analysis relies on the recorded $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^+ \rightarrow J/\psi K^+$ decays for calibration and normalisation purposes.

The data recording started with a few proton bunches in April 2011, after which the LHC gradually increased the number of bunches until the bunch separation in the colliding bunch trains was 50 ns. The instantaneous luminosity was kept between 3 and $3.5 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, and on average 1.5 pp interactions per bunch crossing were visible in the detector (see Sec. 2.1). After a technical stop at the beginning of 2012, LHC raised the collision energy to $\sqrt{s} = 8 \text{ TeV}$, and protons were collided in LHCb again in April 2012. The bunches remained separated by 50 ns, but the instantaneous luminosity was slowly raised to $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ during the first 100 pb^{-1} . The remaining 2 fb^{-1} were recorded at very stable conditions, with about 1.6 visible pp interactions per bunch crossing.

The trigger strategies for the decay channels relevant in this analysis are reviewed in Ch. 3. The triggered events are stored, reconstructed (as described in Sec. 2.4), and split into $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B_{(s)}^0 \rightarrow h^+ h^-$, and $B^+ \rightarrow J/\psi K^+$ samples. The samples collected at different collision energies are treated separately to account for possible variations in the detection efficiencies.

The analysis of the reconstructed samples is described in Ch. 4. The goal of the first candidate selection stage is to retain maximum $B_s^0 \rightarrow \mu^+ \mu^-$ selection efficiency while suppressing the main sources of background and reducing the data sample to a manageable size. In subsequent selection steps, no more candidates are discarded. Rather, candidates are assigned a probability of being background or signals according to the value of a multivariate operator called ‘‘BDT’’. The BDT combines the discrimination power of twelve



The selected $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ events in the 2011 and 2012 data, spread on the two dimensional analysis plane formed by the Multivariate operator output and di-muon (invariant) mass. $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ mass region is between the green lines, the $B^0 \rightarrow \mu^+ \mu^-$ mass region between the orange lines.

geometrical and kinematic variables, and is trained on the simulated samples to recognise false signal candidates arising from combinations of two random muon tracks.

The total efficiency with which the candidate yields are measured is given by the detector acceptance, the trigger efficiency, the reconstruction efficiency, and the selection efficiency. These efficiencies are mainly determined from Monte Carlo simulations, with the notable exception of the trigger efficiency, that is determined from the data itself with the “TISTOS” method described in Ch. 5.

The best signal sensitivity is achieved by distributing the selected $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates on a two-dimensional plane, defined by the BDT and the di-muon (invariant) mass. The candidates are grouped, according to their BDT value, into eight categories, and a di-muon mass distribution model is used to extract the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields. Chapter 6 describes how the model is built. In order to avoid the experimenter bias, the analysis has been developed by “blinding” the signal regions, which were only un-blinded once the analysis strategies had been established.

The total number of $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ mesons in LHCb needs to be known for the branching fraction calculation. Since its estimate from the luminosity and cross section carries a sizable uncertainty, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are normalised relative to other decay channels, namely $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$, with well known branching fractions, as discussed in Ch. 7.

Finally, the results of the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis are presented in Ch. 8.

Simulated samples

In many cases, the methods discussed in this part of the dissertation rely on Monte Carlo simulation of decay samples for validation and calibration: the trigger selection is demonstrated on a simulated $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ sample; the trigger efficiency estimation with the TISTOS method is validated and calibrated on simulated $B^+ \rightarrow J/\psi K^+$ samples; the BDT multivariate classifier is trained using a simulated $B_s^0 \rightarrow \mu^+ \mu^-$ and an inclusive $b\bar{b} \rightarrow \mu\mu X$ sample; the mass distributions of mis-identified backgrounds are calibrated on the corresponding simulated decay samples.

The simulation of the Monte Carlo samples includes all the stages of the experiment: particle production in the pp collisions, particle decays, interactions with the detector material and the magnetic field, trigger, and event reconstruction [93]. The LHCb event simulation is performed by a software package called GAUSS [94]. In the first step, GAUSS simulates the pp collisions, the particle production and decays. It relies on PYTHIA [95] to simulate the (hard) processes in the pp collisions and the subsequent hadronisation. The output of PYTHIA is particles with a given four-momentum. The decays of the B mesons are simulated with a dedicated software package, EVTGEN [96], that also describes mixing and CP violation effects. The simulation is sped up by excluding particles outside of the LHCb detector acceptance. The polar angle of the tracks of the two muons from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ decays, the hadrons from $B_{(s)}^0 \rightarrow h^+ h^-$ decays, and the two (oppositely charged) muons in the $b\bar{b} \rightarrow \mu\mu X$ background sample is required to be between 10 and 40 mrad.

Next, the particles and their decay products are propagated through the detector material and magnetic field, both simulated by the software package GEANT4 [97]. GEANT4 includes a detailed description of the detector geometry, the materials of which the detector cells are made, and how different particles interact with the materials. GEANT4 also simulates decays of the longer-lived particles, such as K_S . The output of GEANT4 is a collection of

hits in the detector cells.

The charged decay products can radiate photons and lose energy as they propagate through the detector. The simulation of the final state radiation is necessary for a reliable comparison between the experimental results and the theoretical predictions, and is performed by a dedicated software package, PHOTOS [46]⁵.

Finally, the software package BOOLE [98] simulates the detector response to the cell hits produced by GEANT4. BOOLE includes spill-over effects⁶, LHC backgrounds, and a simulation of the Level 0 trigger. For each sub-detector, dedicated algorithms describe the physical processes of the analogue signal formation and collection and the electronics response. The output of these algorithms has been calibrated on test-beam data.

Further processing steps are identical for Monte Carlo simulation and real pp collision data: all data is processed by the same offline reconstruction algorithms, described in Sec. 2.4. The simulation conditions are matched to the conditions during the data acquisition: the samples are simulated separately for the different collision energies and accounting for the changes in trigger configuration, as well as the changes in reconstruction and selection criteria.

⁵The $B_s^0 \rightarrow \mu^+ \mu^-$ signal simulation uses PHOTOS version 215.4.

⁶Spill-over effects are detector effects that originate from physical processes that last longer than the time span between two bunch-bunch crossings. The spill-over effects include hits originating from slow particles and particle drift in the OT.

Chapter 3

Trigger in $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

The LHCb trigger system decides which proton-proton collisions will be stored for further analysis. Its main purpose is to reduce the incoming high data rate keeping only the events interesting for LHCb’s physics analysis. The LHCb trigger philosophy and its technical implementation was described in Sec. 2.3. An overview of the trigger performance can be found in Refs. [99–101]. This chapter focuses on the trigger lines relevant for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, namely those used to trigger the signal channels, $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, and those used to trigger the normalisation and the calibration channels, $B^+ \rightarrow J/\psi K^+$ and $B_{(s)}^0 \rightarrow h^+ h^-$ ($h = \pi, K$).

3.1 Level 0

The first trigger level, Level 0 (L0), is based on hardware (Sec. 2.3.1). The L0 consists of dedicated trigger lines¹ designed to select muons, hadrons, photons, and electrons

The L0 muon trigger lines, `L0Muon` and `L0DiMuon`, use the fast muon detector information to trigger on one high p_T muon candidate or two lower p_T muon candidates. The muon search begins with a hit in the middle muon station (M3), which is used to define fields of interest in the other four muon stations. The algorithm looks for at least one hit in each of the five muon stations, and when found, uses the hits in the first two stations, where the lateral resolution is higher, for a quick muon p_T determination. Two muons,

¹A “trigger line” is a unique set of selection criteria and/or reconstruction criteria within a trigger level.

one with the largest and the other with the second largest p_T , are selected in each quadrant. Eight selected candidates are sent to the L0 muon trigger decision, where the **LOMuon** trigger line compares the p_T of the largest- p_T muon candidate to a single p_T threshold, and the **LODiMuon** trigger line compares the p_T product² of the top candidates, $p_T^{largest} \times p_T^{2nd\ largest}$, to a fixed threshold.

The L0 calorimeter trigger lines trigger on electrons, photons, or hadrons, using the transverse energy depositions in ECAL and HCAL cells. The electromagnetic candidate type is determined including the hit information from PS and SPD (see Fig. 2.13). Only the highest E_T candidate per type is selected for the trigger decision, whereas for the hadron candidates, the ECAL and HCAL E_T deposits are summed.

The L0 thresholds for the relevant trigger lines are listed in Tab. 3.1. Events with large occupancy in the OT and IT consume a disproportionately large fraction of the available processing time in the subsequent trigger levels. The number of SPD hits is a good measure of the occupancy, permitting an early rejection of events that would otherwise take a long time to be processed. The optimal number of SPD hits is below 900 for events triggered by **LODiMuon**, and below 600 for all the other L0 lines. In 2012, the **LOMuon** and **LODiMuon** thresholds were raised in order to reduce the higher muon yield at larger pp collision energies.

Table 3.1: L0 trigger line definitions in 2011 and 2012.

Trigger line	Cut	2011 [GeV/c]	2012 [GeV/c]	SPD [hits]
LOMuon	p_T	> 1.48	> 1.76	< 600
LODiMuon	$\sqrt{p_T(\mu_1) * p_T(\mu_2)}$	> 1.3	> 1.3	< 900
LOHadron	E_T	> 3.5	> 3.7	< 600
LOElectron	E_T	> 2.5	> 3.0	< 600
LOPhoton	E_T	> 2.5	> 3.0	< 600

3.1.1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$

More than 99% of the L0 accepted $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates are triggered by **LOMuon** or **LODiMuon** lines (Fig. 3.1). Around $\sim 20\%$ of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates also satisfy the **LOHadron** trigger criteria because of other (had-

²Compared to a threshold set on the muon p_T sum, a threshold on the product was found to reduce the accepted event rate by 30%, while retaining the same signal efficiency.

ron) tracks in the event; these events would have been triggered even if no signal were present. The $B^+ \rightarrow J/\psi K^+$ candidates, necessary for normalisation (see Ch. 7), are mostly triggered by the muon trigger lines (`L0Muon` or `L0DiMuon`). `L0Hadron` and other L0 lines trigger exclusively 1.9% of the candidates (Fig. 3.1).

3.1.2 $B_{(s)}^0 \rightarrow h^+ h^-$

The `L0Hadron` line triggered on 70% of the L0 accepted $B_{(s)}^0 \rightarrow h^+ h^-$ decay candidates. The daughter hadron decays in flight produce muons, electrons, photons, and other hadrons, which is why the $B_{(s)}^0 \rightarrow h^+ h^-$ decays can also be triggered by muon lines (`L0Muon`, 19%) and electron lines (`L0Electron`, 5%). The unrelated muon, electron, photon, and hadron candidates in the event cause the triggers independent of the $B_{(s)}^0 \rightarrow h^+ h^-$ decays.

3.2 HLT1

The first software trigger level, HLT1 (see Sec. 2.3.2), processes the L0 accepted events. The events used in the $B_s^0 \rightarrow \mu^+ \mu^-$ analysis are triggered by the inclusive one track trigger line, `Hlt1TrackAllL0`, and various muon lines. The muon lines are executed in case the event was triggered by `L0Muon` or `L0DiMuon` line; the `Hlt1TrackAllL0` line will be executed for all L0 accepted events. The complete selection criteria for these trigger lines are listed in Tab. 3.2.

The track segments in the VELO that can be built in a short time, are used in HLT1 for further event analysis. `Hlt1TrackAllL0` selects VELO tracks with good quality: based on the smallest impact parameter (IP) to any primary vertex (PV) ($IP > 0.1$ mm). The HLT1 muon lines first validate the L0 muon hypothesis with a quick algorithm that tries to “match” each VELO track to the muon station hits (see Sec. 2.3.2). The `Hlt1TrackMuon` line accepts the events containing at least one matched muon candidate with a significant IP to any primary vertex ($IP > 0.1$ mm); the `Hlt1SingleMuonHighPT` line specialises on muons from heavy particles, and if very high p_T muons ($p_T > 4.8$ GeV/ c) are present, triggers the event regardless of the displacement.

The VELO tracks selected either by `Hlt1TrackAllL0` or by the various muon lines, become seeds for the forward tracking algorithm (see Sec. 2.3.2). Tracks accepted by the muon lines have looser momentum requirements in the *forward tracking*, but are additionally asked to pass the offline `isMuon` muon identification (see Sec. 2.5).

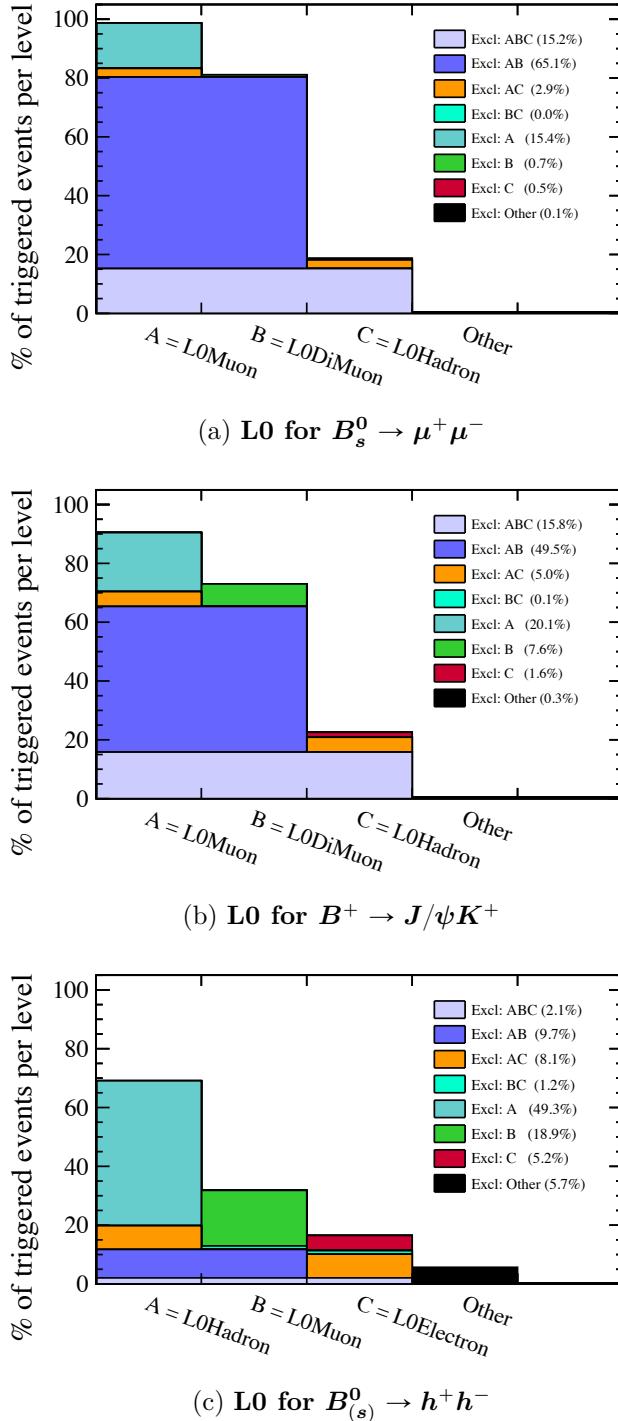


Figure 3.1: The trigger lines in L0, shown for simulated $B_s^0 \rightarrow \mu^+ \mu^-$ (top), $B^+ \rightarrow J/\psi K^+$ (middle) and $B_{(s)}^0 \rightarrow h^+ h^-$ candidates (bottom) in the 2012 (2 fb^{-1}) data. In $B_{(s)}^0 \rightarrow h^+ h^-$ sample, only events with hadron pair mass matching the B meson mass are considered. The colours denote candidates exclusively triggered by a certain set of lines, e.g. “Excl. ABC” denotes the candidates with a positive trigger decision from all three lines: A, B, and C.

Table 3.2: The main HLT1 trigger line definitions in 2011 and 2012.

HLT1 line	TrackAll	TrackMuon	SingleMuonHighPT	DiMuonHighMass	DiMuonLowMass
Configuration in 2011					
isMUON	-	true	true	true	true
Nr. of VELO hits/track	> 9	-	-	-	-
Nr. of missed VELO hits/track	< 3	-	-	-	-
Nr. of OT+ITx2 hits/track	> 16	> 0	-	-	-
Track $IP[mm]$	> 0.1	> 0.1	-	-	-
Track $IP\chi^2$	> 16	> 16	-	-	-
Track $p_T[\text{GeV}/c]$	> 1.7	> 1	> 4.8	> 0.5	> 0.5
Track $p[\text{GeV}/c]$	> 10	> 8	> 8	> 6	> 6
Track χ^2/ndf	< 2	< 2	< 4	< 4	< 4
DOCA [mm]	-	-	-	< 0.2	< 0.2
χ^2_{vertex} [mm]	-	-	-	< 25	< 25
Mass [GeV/c^2]	-	-	-	> 2.7	> 1
Configuration in 2012					
isMUON	-	true	true	true	true
Nr. of VELO hits/track	> 9	-	-	-	-
Nr. of missed VELO hits/track	< 3	-	-	-	-
Nr. of OT+ITx2 hits/track	> 16	> 0	-	-	-
Track $IP[mm]$	> 0.1	> 0.1	-	-	-
Track $IP\chi^2$	> 16	> 16	-	-	-
Track $p_T[\text{GeV}/c]$	> 1.6	> 1	> 4.8	> 0.5	> 6
Track $p[\text{GeV}/c]$	> 3	> 3	> 3	> 3	> 0
Track χ^2/ndf	< 2	< 2.5	< 3	< 3	< 0
DOCA [mm]	-	-	-	< 0.2	< 0.2
χ^2_{vertex} [mm]	-	-	-	< 25	< 25
Mass [GeV/c^2]	-	-	-	> 2.7	> 0

The tracks found are then fitted with a Kalman filter [87,88] that accounts for the multiple scattering and corrects for the energy loss in the detector material. The resulting χ^2 of each track is a good estimate of the track reconstruction quality, and, depending on the trigger line, low quality tracks with too high χ^2 or $\text{IP}\chi^2$, or both, will be discarded.

A subset of the muon lines is dedicated to events with two good muon candidates. These lines try to reconstruct di-muon vertexes, and, if a good vertex can be built, select the di-muon events by either of the two criteria:

- `Hlt1DiMuonLowMass` is based on the track displacement from the primary vertex, and has a low³ di-muon mass cut, $m_{\mu\mu} > 1 \text{ GeV}/c$);
- `Hlt1DiMuonHighMass` requires $m_{\mu\mu} > 2.7 \text{ GeV}/c$, and no additional vertex requirements.

3.2.1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates in HLT1 (Fig. 3.2) are triggered predominantly by `Hlt1TrackMuon` and `Hlt1SingleMuonHighPT` lines, which together accept about 98% of the signal candidates. The `Hlt1TrackAllL0` line recovers an additional 0.7% of the candidates that do not pass the muon identification criteria. The $B^+ \rightarrow J/\psi K^+$ candidates in the 2012 data (Fig. 3.2) were mainly triggered by the inclusive single-track trigger lines `Hlt1TrackAllL0` and `Hlt1TrackMuon` (more than 98%). The remaining events were accepted by dedicated HLT1 di-muon lines (1%).

3.2.2 $B_{(s)}^0 \rightarrow h^+ h^-$

More than 95% of the $B_{(s)}^0 \rightarrow h^+ h^-$ candidates accepted by HLT1 were triggered by the `Hlt1TrackAllL0` line; 2.5% were accepted exclusively by the `Hlt1TrackMuon` line, and 0.6% by a photon line (Fig. 3.2). The muon and photon trigger decisions were based on the tracks associated to the signal decays, which can be explained by hadron decays in flight.

3.3 HLT2

The HLT1 output rate allows HLT2, the final trigger level, to apply forward tracking to the majority of the VELO tracks, and offline `isMUON` muon identification to all the tracks from the forward tracking. The relevant HLT2 lines for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis are single muon lines (Tab. 3.3), di-muon lines

³In 2012, the mass cut for `Hlt1DiMuonLowMass` was omitted completely.

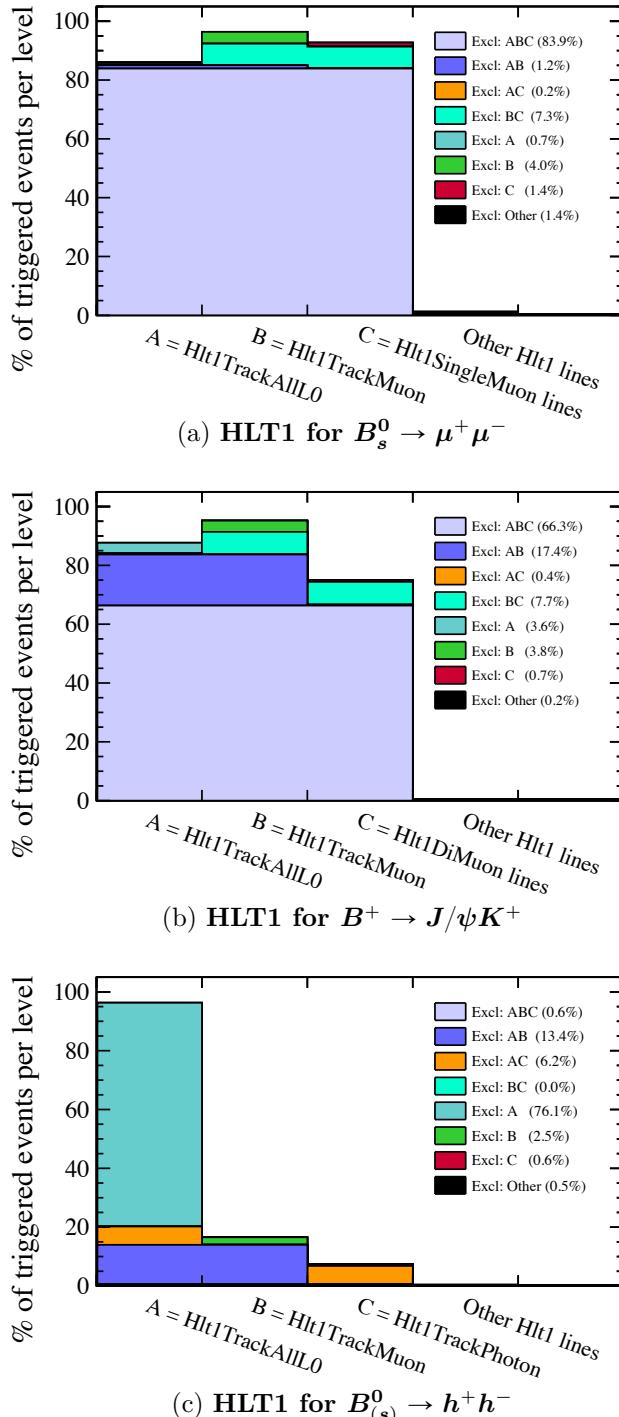


Figure 3.2: The trigger lines in HLT1, shown for simulated $B_s^0 \rightarrow \mu^+ \mu^-$ (top), $B^+ \rightarrow J/\psi K^+$ (middle), and $B_{(s)}^0 \rightarrow h^+ h^-$ candidates (bottom) in the 2012 (2 fb^{-1}) data. In $B_{(s)}^0 \rightarrow h^+ h^-$ sample, only events with hadron pair mass matching the B meson mass are considered. The colours denote candidates exclusively triggered by a certain set of lines, e.g. “Excl. ABC” denotes the candidates with a positive trigger decision from all three lines: A, B, and C.

Table 3.3: HLT2 single-muon trigger line definitions. TOM stands for a positive trigger decision on the same muon.

HLT2Single:	Muon	MuonHighPT	MuonLowPT
Configuration in 2011 and 2012			
Pre-scale	0.5	1.	0.002
isMUON	true	true	true
Hlt1TrackMuon	TOM	-	-
Track IP [mm]	> 0.5	-	-
Track $IP\chi^2$	> 200	-	-
Track p_T [GeV/c]	> 1.3	> 10	> 4.8
Track χ^2/ndf	< 2	-	< 10

(Tab. 3.4), topological lines, and an inclusive trigger line dedicated for the $B_{(s)}^0 \rightarrow h^+h^-$ decays - Hlt2B2HH. The HLT2 lines are designed having specific decay signatures in mind, and are therefore best discussed in the context of each decay channel separately.

3.3.1 $B_{(s)}^0 \rightarrow \mu^+\mu^-$

The Hlt2DiMuonB line is specifically designed to trigger on B decays to two muons. It requires two identified muon candidates, with a combined mass above $4.7\text{ GeV}/c^2$, and a well reconstructed di-muon vertex. It triggers 97% of the HLT2-accepted $B_s^0 \rightarrow \mu^+\mu^-$ candidates (Fig. 3.3).

These candidates are in large part also selected by the Hlt2DiMuonDetachedHeavy line, which triggers on “heavy” muon pairs from a particle with long enough flight-distance significance ($\text{FD}\chi^2$). The difference with Hlt2DiMuonDetached is that the latter applies a lower threshold on the muon pair mass, but includes a requirement on the muon transverse momentum. The selection criteria are given in Tab. 3.4.

The majority of the remaining 3% of the candidates are triggered by the Hlt2SingleMuon lines, which are able to recover signal candidates in cases in which the di-muon candidate does not pass the requirements, but there is a high transverse-momentum or detached muon in an event. Different single muon lines have been designed to target muon candidates with low and high transverse momentum (see Tab. 3.3).

3.3.2 $B^+ \rightarrow J/\psi K^+$

Muon pairs from the J/ψ in the $B^+ \rightarrow J/\psi K^+$ sample are too “light” to be triggered by the Hlt2DiMuonB line. Instead, the candidates are selected by

Table 3.4: HLT2 di-muon trigger lines definitions.

HLT2DiMuon:	B	JPsi	JPsiHighPT	Detached	DetachedHeavy	DetachedJPsi
Configuration in 2011						
Pre-scale	1.	0.2	1.	1.	1.	1.
Track χ^2/ndf	< 5	< 5	< 5	< 5	< 5	< 5
Track $IP\chi^2$	-	-	-	> 9	-	-
Mass [GeV/c ²]	> 4.7	$M_{J/\psi} \pm 0.12$	$M_{J/\psi} \pm 0.10$	> 1	> 2.95	$M_{J/\psi} \pm 0.12$
χ^2_{vertex} [mm]	< 10	< 25	< 25	< 25	< 25	-
Track p_T [GeV/c]	-	-	-	> 0.5	> 0.5	-
Track $p_T^{\mu\mu}$ [GeV/c]	-	-	> 2	> 1.5	-	-
$F D\chi^2$	-	-	-	< 49	< 25	< 9
Configuration in 2012						
Pre-scale	1.	0.2	1.	1.	1.	1.
Track χ^2/ndf	< 4	< 4	< 4	< 4	< 4	< 4
Track $IP\chi^2$	-	-	-	> 9	-	-
Mass [GeV/c ²]	> 4.7	$M_{J/\psi} \pm 0.12$	$M_{J/\psi} \pm 0.10$	> 1	> 2.95	$M_{J/\psi} \pm 0.12$
χ^2_{vertex} [mm]	< 10	< 25	< 25	< 8	< 8	-
Track p_T [GeV/c]	-	-	-	> 0.3	> 0.3	-
Track $p_T^{\mu\mu}$ [GeV/c]	-	-	> 2	> 0.6	-	-
$F D\chi^2$	-	-	-	< 49	< 25	< 9

a very similar line, `Hlt2DiMuonDetachedJPsi`. This line fires if the di-muon mass is close to the J/ψ mass, and if the di-muon vertex is well separated from the primary vertex. Compared to the `Hlt2DiMuonDetachedHeavy` line (see Tab. 3.4), the efficiency for $J/\psi \rightarrow \mu^+ \mu^-$ decays is enhanced by including lines with reduced flight distance significance ($FD\chi^2$) requirement (see Tab. 3.4).

Besides the muon lines, the so called inclusive topological lines play an important role in triggering the $B^+ \rightarrow J/\psi K^+$ candidates. The topological lines are designed to trigger on partially reconstructed b -hadron decays with a displaced decay vertex and at least two charged particles in the final state. First, tracks are selected based on their track fit quality (χ^2/ndf), impact parameter (IP), and muon or electron identification. Two, three, or four-body vertices are constructed, starting with a two-body vertex, and including additional tracks. Good vertex candidates need to pass the requirements on $\sum |p_T|$, $p_{T\min}$, n-body invariant mass, distance of closest approach of the added track (*DOCA*), impact parameter significance ($IP\chi^2$), and flight distance significance ($FD\chi^2$). Additionally, a corrected mass (m_{cor}) of an n-body vertex is evaluated by imposing momentum conservation at the n-body vertex (using the closest PV) and estimating the missing transverse momentum. Inclusion of m_{cor} enables the topological trigger lines to trigger also on partially reconstructed decays.

A multivariate operator combines the separation power of the variables. The operator used in the topological lines is trained with Monte Carlo simulated signal data that contains B^+ , B^0 , B_s^0 , and Λ_b^0 decays, and on the measured `NoBias` data sample as background⁴. The topological lines and the `Hlt2DiMuonDetachedJPsi` line together triggered 99% of the $B^+ \rightarrow J/\psi K^+$ candidates in HLT2 (Fig. 3.3). The single-muon trigger lines recovered an extra 0.5% of the $B^+ \rightarrow J/\psi K^+$ decays.

3.3.3 $B_{(s)}^0 \rightarrow h^+ h^-$

The topological lines trigger 85% of the $B_{(s)}^0 \rightarrow h^+ h^-$ candidates. A dedicated `Hlt2B2HH` line triggers exclusively about 10% of the candidates (Fig. 3.3); it does not identify the tracks, but suppresses the abundant hadron background originating from the primary vertex imposing stringent requirements on the mass of the hadron pair, the hadron and the B meson impact parameters, and the B meson lifetime. For a fast estimate of the hadron pair mass, the algorithm assumes all hadrons are pions. The `Hlt2B2HH` line definition is given by

⁴`NoBias` events are randomly triggered pp collisions. The random decision is made by L0, and the events are passed through the HLT.

in Tab. 3.5.

Table 3.5: H1t2B2HH trigger line definitions.

Configuration in 2011 and 2012	H1t2B2HH
Pre-scale	1.
Track $IP[\text{mm}]$	> 0.12
Mass $M(h^+h^-)[\text{GeV}/c^2]$	$4.7 - 5.9$
$\chi^2_{vertex} [\text{mm}]$	< 10
B meson $p_T[\text{GeV}/c]$	> 1.2
B meson $\tau[\text{ps}]$	> 0.0006
B meson $IP[\text{mm}]$	> 0.12

3.4 Global event cuts

In 2011, some “global event cuts” (GEC) were applied at the L0 level (on the number of SPD hits) to discard events that would take very long processing time in HLT. In 2012, these were introduced at all trigger levels, as shown in Tab. 3.6.

Table 3.6: Global events cuts per trigger level in 2012.

Trigger level	Cut
L0 (LOMuon/LODiMuon)	SPD hits $< 900/600$
HLT1	VELO hits < 10000
HLT1	IT hits < 3000
HLT1	OT hits < 15000
HLT2	VELO tracks < 350

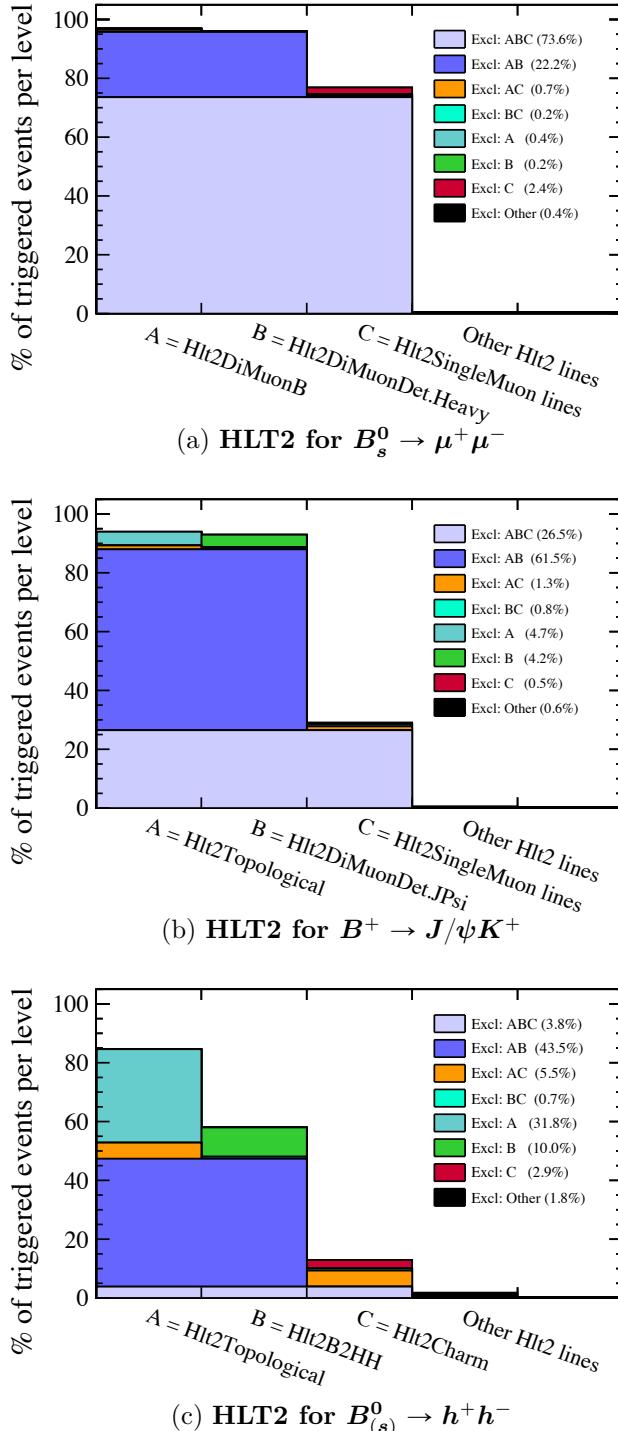


Figure 3.3: The trigger lines in HLT2, shown for simulated $B_s^0 \rightarrow \mu^+ \mu^-$ (top), $B^+ \rightarrow J/\psi K^+$ (middle), and $B_{(s)}^0 \rightarrow h^+ h^-$ candidates (bottom) in the 2012 (2 fb^{-1}) data. In $B_{(s)}^0 \rightarrow h^+ h^-$ sample, only events with hadron pair mass matching the B meson mass are considered. The colours denote candidates exclusively triggered by a certain set of lines, e.g. ‘‘Excl. ABC’’ denotes the candidates with a positive trigger decision from all three lines: A, B, and C.

Chapter 4

Signal selection

Although the signature of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in the detector is straightforward (Fig. 4.1), the low branching fractions in the order of 10^{-9} make it a challenging task to distinguish the signal decays from similar and more abundant processes (Fig. 4.2). The selection of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates begins with a cut based selection, described in Sec. 4.1. It is optimised for the signal efficiency, reduces the main backgrounds, and provides a sample with a manageable size for the following multivariate-operator selection.

Two multivariate operators are used: “BDTS” and “BDT”, both described in Sec. 4.2. The BDTS output is used to further reduce the sample by keeping most of the signal while considerably reducing the background. After the BDTS selection, no more candidates are discarded, and the BDT output will be used in Ch. 6 to classify the candidates according to the background likelihood. As the calibration of the BDT output (Sec. 4.4) and the branching fraction normalisation (Sec. 7) rely on $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^+ \rightarrow J/\psi K^+$ data samples, the selection of these two decay channels is kept as similar as possible to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection.

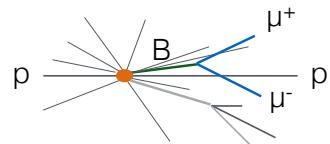


Figure 4.1: $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal decay signature.

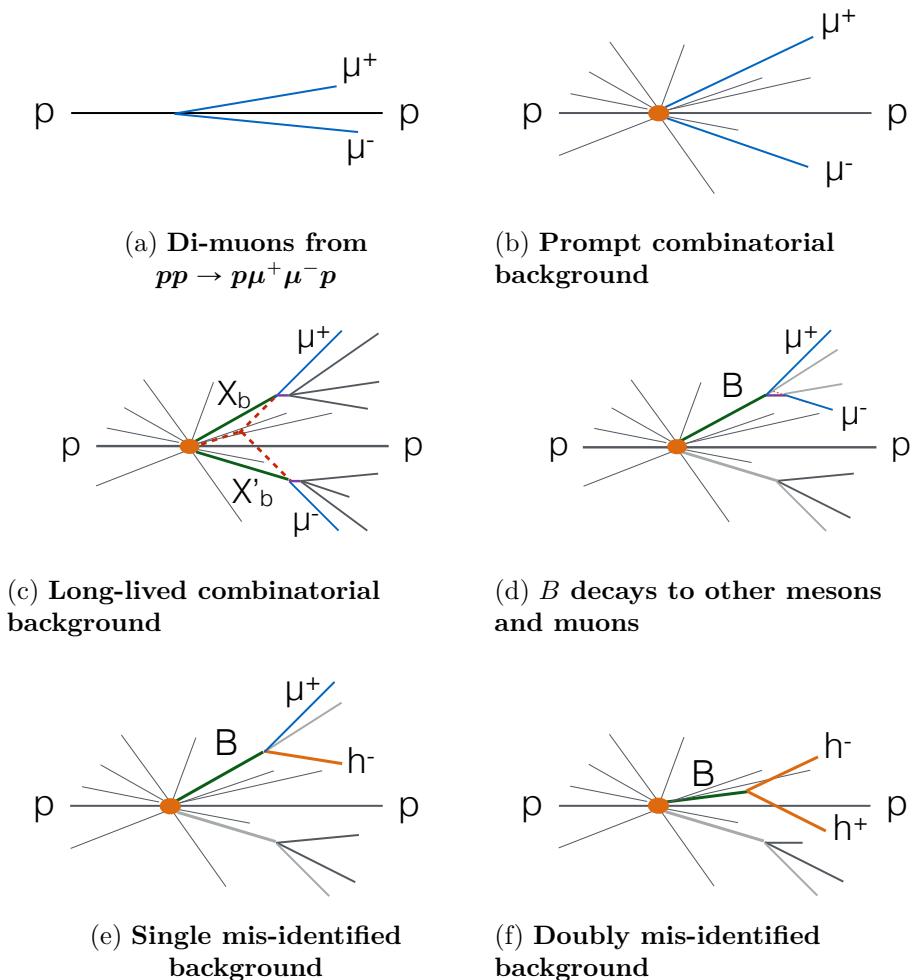


Figure 4.2: Different $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ backgrounds.

4.1 Selection based on cuts

The information in the reconstructed event, track and vertex quality, particle identity, particle masses, and particle momenta, is used to select pure samples of signal candidates as well as candidates of the calibration and normalisation channels. The selection criteria have been optimised to retain high $B_s^0 \rightarrow \mu^+ \mu^-$ selection efficiency while reducing the background sources shown in Fig. 4.2. The criteria for the normalisation and calibration channels are kept as similar as possible to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection, and, in the context of the selection, we refer to any of these channels as *signal*. The selection criteria and the exact cut values for the different signal samples are listed in Tab. 4.1.

The first selection goal is to eliminate candidates with low reconstruction quality.

Only the tracks with small track χ^2/DOF are kept; the two daughter tracks are then required to have a small distance of closest approach (DOCA) and to form a vertex with a good χ^2 . Next, the selection excludes any *non-physical candidates* by removing tracks with momentum (p) or transverse momentum (p_T) values that are too large for the LHCb acceptance, and discarding the B meson candidates with too long proper time. Non-physical tracks can also be created as artefacts of the event reconstruction (“ghosts”). Ghost tracks do not represent physical particle paths. Multiple track and event parameters¹ are combined in a single multivariate operator, called Ghost Probability (GP), which, for each track, evaluates the probability for it to be a ghost track. The tracks likely to be ghosts are excluded.

A complementary approach to reducing the reconstruction artefacts is to consider the information each track adds when compared to other tracks. The difference in information content is described by a variable called the “Kullback-Leibner” distance (KL) and it is estimated for all the possible pairs of tracks in the event. Two tracks are called “clones” if they provide the same information (i.e have a small KL) in which case only the one with the better quality is kept.

The second selection goal is to reduce the physical background sources.

Muons from *elastic pp* collisions, in which the protons remain intact and fly close to the beam, can lead to a di-muon pair with a well reconstructed vertex and an invariant mass in the B meson mass range (Fig. 4.2a). However, the resulting di-muon object will have a very low transverse momentum, and the $pp \rightarrow p\mu^+\mu^-p$ background is rejected by a B mesons transverse momentum

¹The GP operator combines 22 parameters, including different track-fit quality quantifiers, the number of sub-detector hits, comparison between the measured and observed number of sub-detector hits, and several kinematic variables.

Table 4.1: Selection criteria applied to the signal, normalisation, and calibration candidate samples.

Cut	Applied on	Threshold for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$	Applied on	Threshold for $B^+ \rightarrow J/\psi K^+$
Reduce: Candidates with poor reconstruction quality				
Track χ^2/DOF	μ/h	< 3	μ/h	< 3
Track DOCA[mm]		< 0.3		< 0.3
Vertex χ^2	$B_{(s)}^0$	< 9	J/ψ	< 9
Reduce: Non-physical candidates				
Track p [GeV/ c]	μ/h	< 500	μ/h	< 500
Track p_T [GeV/ c]		< 40		< 40
Proper time t	$B_{(s)}^0$	$< 9 \cdot \tau(B_s^0)$	B^+	$< 9 \cdot \tau(B_s^0)$
Track GP		< 0.4373		< 0.4373
Track KL		< 5000		< 5000
Reduce: $pp \rightarrow p\mu^+\mu^-p$				
Track p_T	$B_{(s)}^0$	$> 0.5 \text{ GeV}/c^2$	B^+	$> 0.5 \text{ GeV}/c$
Reduce: Prompt combinatorial background				
Track $IP\chi^2$	μ/h	> 25	μ/h	< 25
Track $VD\chi^2$	$B_{(s)}^0$	> 15	J/ψ	> 15
Track $IP\chi^2$	$B_{(s)}^0$	< 25	B^+	< 25
$ M^{PDG} - M $ [GeV/ c^2]		< 0.5	$B^+/J/\psi$	< 0.5
Reduce: Mis-identified background				
isMUON	μ only	true	μ only	true
Track $\Delta LL(\mu, \pi)$		> -5		—
Track $\Delta LL(K, \pi)$		< 10		—

requirement.

The *inelastic* pp collisions produce a great number of muons at the primary vertex (Fig. 4.2b). These so-called “prompt” muons, are produced in a wide momentum range, and, in the right combination, the two oppositely charged muons can look as muons from an instantaneously decaying B meson. One clear difference between the prompt and the interesting muons is that the B mesons produced in LHCb are strongly boosted and cover distances in the order of a centimeter before decaying. The prompt combinatorial background is thus strongly reduced by selecting only tracks with large impact parameter significance (IP χ^2), large distance significance of the reconstructed B (or J/ψ) decay vertex ($VD\chi^2$), and small B candidate impact parameter significance². Then, depending on the sample, the knowledge of the B and J/ψ meson mass is used to keep only the candidates with matching di-muon invariant mass.

The number of mis-identified muon candidates in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ samples is significantly reduced by an ISMUON requirement (see Sec. 2.5.1). Furthermore, it was found that requiring a PID cut, $\Delta LL(K - \pi) < 10$ and $\Delta LL(\mu - \pi) > -5$ cut (see Sec. 2.5.1), reduces the yields of $B_{(s)}^0 \rightarrow h^+ h^-$ events in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ sample by a factor of five, while retaining 98% of the $B_s^0 \rightarrow \mu^+ \mu^-$ events.

4.2 Multivariate operators

The selection based on cuts, described in Sec. 4.1, is optimised for the $B_s^0 \rightarrow \mu^+ \mu^-$ decays and has nearly 100% signal efficiency. While it considerably reduces non-physical, prompt, and mis-identified backgrounds, the long-lived combinatorial background (Fig. 4.3) is still dominating the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ sample.

This background could be further reduced by even tighter cuts, which are able to reject up to 50% of the background while still retaining a relatively high signal efficiency at 95%. A Boosted Decision Tree (BDT) multivariate classifier, however, is able to reject 60% of the background with the same signal efficiency. Therefore, a BDT multivariate approach was chosen for the further background rejection.

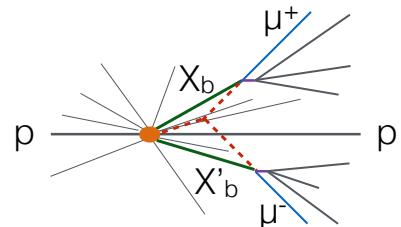


Figure 4.3: Cartoon of a long-lived combinatorial background.

²All the parameters are calculated relative to the primary vertex.

The multivariate selection is performed in two steps: first, the output of a BDT, called BDTS, is used to *exclude* candidates with very low signal likelihood (Sec. 4.3); in the second step, another BDT (simply called BDT) is used to *classify* the candidates in eight categories with increasing signal-to-background ratio (Sec. 4.4). A classification rather than rejection is motivated, since after the selection and the BDTS requirement, any additional sample reduction would exclude a significant amount of signal candidates, and thereby, lower the analysis sensitivity.

4.2.1 Decision Trees and Boosting

The multivariate operators used in the analysis have been built and optimised in The Toolkit for Multivariate Analysis (TMVA) [102] environment. Using TMVA, a wide range of classification techniques were compared and the Boosted Decision Tree operators were found to provide the best background rejection and signal efficiency.

Decision trees are classifiers that can be visualised by a two-dimensional tree structure (Fig. 4.4). They are based on multiple yes/no decisions, just as a cut-based analysis, but are capable of selecting more than one region in the space defined by the discriminating variables. Starting from the *root* node, the sample is split between two sub-nodes using a discriminating variable x_i . The node-splitting is repeated for the sub-nodes, until a stop criterion is fulfilled³.

Decision trees are “trained” on pure samples of signal and background candidates, typically obtained from Monte Carlo simulations. Each split uses a variable and a corresponding cut value, which, at that node, give the best separation between signal and background. Some variables might thus be used several times, while others not at all. The nodes at the bottom of the tree, called *leaf* nodes, are classified as signal or background depending on the majority of the candidates they contain.

Decision trees are prone to statistical fluctuations in the training sample. These fluctuations could alter the variable choice in the process of node splitting, and therefore, the entire tree structure below that particular node. In this case, the classifier response on actual data will (on average) be sub-optimal, and the tree is called “over-trained”.

The instability of a single decision tree is overcome by constructing a large number of decision trees and defining the classifier response based on the majority vote. The trees are trained, one after another, on the same

³The stop criterion can be based on the number of candidates in each sub-node, or on the fraction of signal candidates (purity).

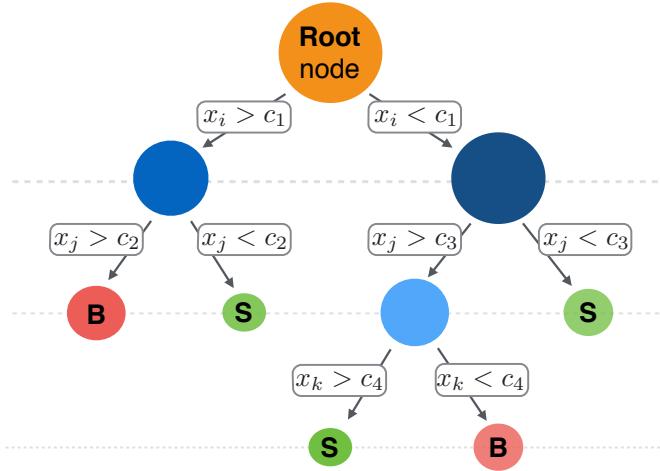


Figure 4.4: Schematic view of a decision tree. In each split, the variable and threshold which give the best separation between signal and background are used.

training sample. The candidates misclassified by the first tree are given a higher weight in training the second tree, and so forth. This process is called *boosting*. After the trees have been trained, the weights given to the training samples are taken into account when combining all tree decisions into a single classifier - a Boosted Decision Tree (BDT).

4.3 Multivariate selection: BDTS

The BDTS uses six variables to discriminate the background:

- B candidate impact parameter;
- B candidate impact parameter χ^2 ;
- di-muon vertex χ^2 (either B or J/ψ vertex, depending on the decay);
- the angle between the B candidate momentum, and the line connecting the secondary and primary vertex (DIRA);
- the shortest distance between the two daughter tracks (DOCA), considering only muon tracks in case of $B^+ \rightarrow J/\psi K^+$;
- the smallest daughter track impact parameter (*minIP*) with respect to any primary vertex (muon tracks for $B^+ \rightarrow J/\psi K^+$).

The BDTS is trained using the Monte Carlo $B_s^0 \rightarrow \mu^+ \mu^-$ sample as signal and an inclusive $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo sample as a background. The candidates used for the training are required to pass the $B_{(s)}^0 \rightarrow h^+ h^-$ initial selection (Tab. 4.1).

The trained BDTS is applied to all candidates, including normalisation and calibration decays, and only the candidates passing a cut on the BDTS output are kept. The cut value ($\text{BDTS} > 0.05$) was optimised to have the best performance of the subsequent multivariate operator BDT, described in Sec. 4.4. It was verified, that the BDTS cut efficiency is the same for the different channels.⁴

4.4 Multivariate classification: BDT

The BDT is trained on a simulated $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo sample, as signal, and an inclusive Monte Carlo $b\bar{b} \rightarrow \mu\mu X$ sample, as (long-lived combinatorial) background.

Both training samples consist of about 7×10^4 candidates after the $B_{(s)}^0 \rightarrow h^+ h^-$ selection and the $\text{BDTS} > 0.05$ cut. Independent signal and background Monte Carlo samples are used for the post-training BDT testing and optimisation.

The BDT deliberately does not include particle identification information: it is trained to separate the true 2-body B decay candidates from multi-body decays (e.g Fig. 4.3), and treat $B_{(s)}^0 \rightarrow h^+ h^-$ and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ equally as signals.

Many possible variables have been considered for the BDT definition. Twelve variables were chosen, based on their good discrimination power between the two-body B decays and the combinatorial background, and their low correlation to the di-muon mass⁵.

The twelve BDT variables can be split in four groups according to their signal and background separation strategies. The simulated signal distributions for the variables are validated on the measured $B_{(s)}^0 \rightarrow h^+ h^-$ sample;

⁴The small expected variations in *minIP* distribution between the channels were further studied and found to be negligible. The BDTS cut efficiency ratios between the normalisation channels and signal channels are equal to 1 within 0.4%. The study included the IP “smearing” effects.

⁵The di-muon mass will later be used in the analysis to estimate the number of combinatorial background events in each BDT category, and extract the signal yield (Ch. 6). The non-linear BDT dependency on the di-muon mass could alter the exponential shape and lead to over- or underestimation of the signal yield. A small linear dependence, however, can be accounted for by the exponential fit function.

the simulated background distributions are validated on the di-muon data side-bands, as shown in Figs. 4.5, 4.6, 4.7, and 4.8.

The four variables in the first group identify the signal decays based on the longer flight distance of the B mesons, as well as on the quality of the reconstructed vertex:

- $\text{IPS}(\mu/h)$: minimum daughter impact-parameter significance;
- $\tau(B)$: B meson decay time;
- $\text{IP}(B)$: B meson impact parameter;
- **DOCA**: distance of closest approach between the two daughters.

The distributions of this first group of variables are shown in Fig. 4.5.

The second group of variables uses the differences between true B candidates and fake B candidates built with tracks from multiple decay vertices. Because the b quarks are produced together with \bar{b} quarks, a B meson decay is accompanied by another particle, containing the \bar{b} quark. The combinatorial background arises when daughter tracks from the two separate decays are combined to form a “fake” B meson candidate. These fake B ’s, however, tend to have a low transverse momentum, and to fly close to the direction of the $b\bar{b}$ pair. The $b\bar{b}$ pair direction can be measured with a variable called “thrust” momentum. The thrust momentum is defined as the momentum sum over good tracks coming from the B candidate primary vertex; the B candidate daughters, prompt tracks, and tracks from the long-lived particles are excluded. For a true signal decay, the thrust momentum approximates the direction of the other b , and therefore, forms an angle with the signal candidate larger than that of a fake B . The thrust momentum is also used to measure the angles at which the daughter particles are emitted in the B candidate rest-frame: for background, angles with respect to the thrust momentum show a larger spread.

The variables in the second group are

- $p_T(B)$: B meson transverse momentum;
- **B thrust**: the angle between the B candidate momentum and the thrust momentum, calculated in the laboratory frame;
- **B boost**: the angle between the direction of the positive daughter candidate and the thrust-momentum direction, calculated in the B candidate rest-frame.

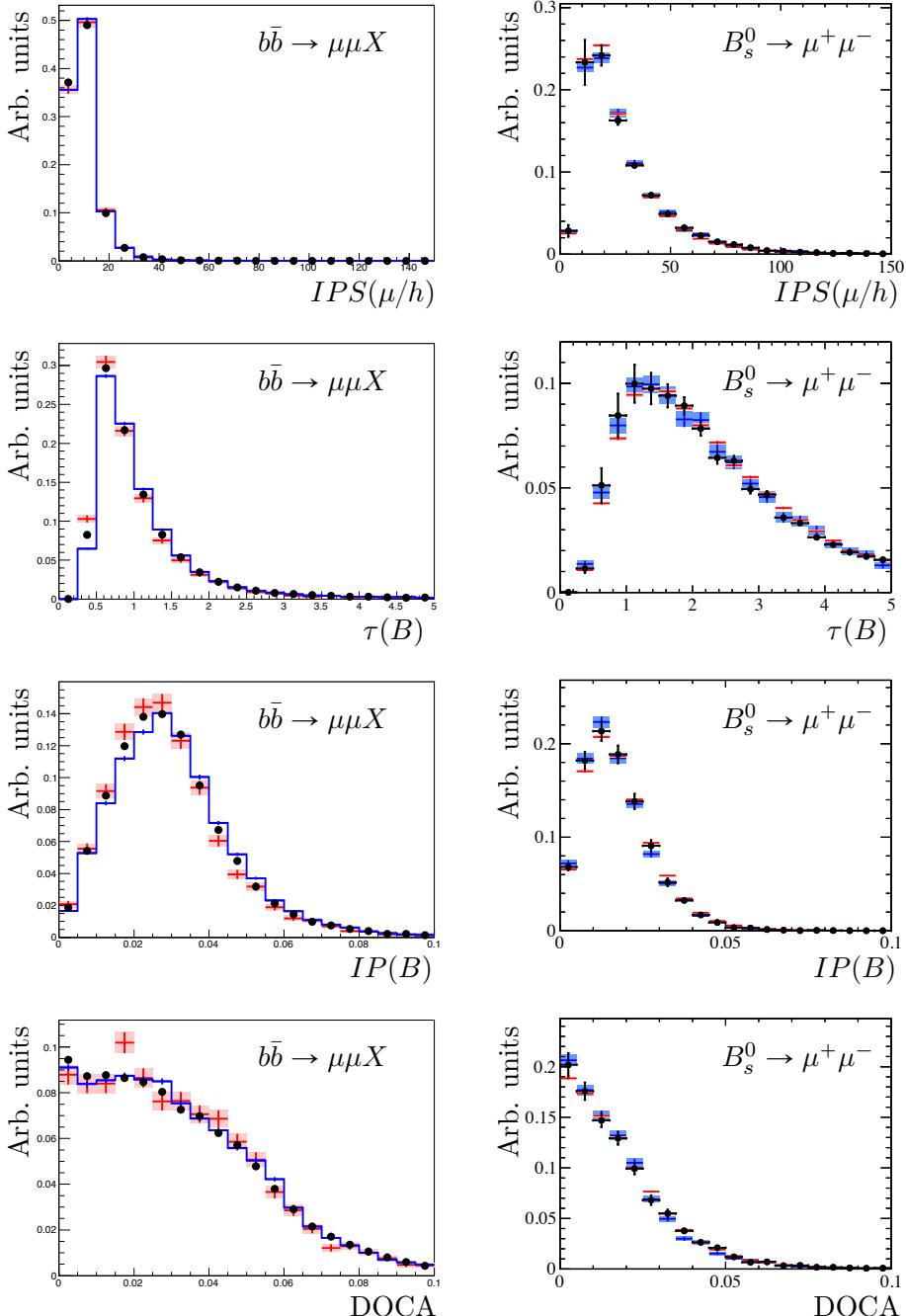


Figure 4.5: The distributions for the BDT “flight distance” variables, shown for background (left) and signal samples (right). The background samples include $b\bar{b} \rightarrow \mu\mu X$ MC samples for 2011 and 2012 (left, red and blue dots resp.) and 2012 di-muon mass sideband data (left, black dots). The signal samples include $B_s^0 \rightarrow \mu^+ \mu^-$ MC samples for 2011 and 2012 (right, blue and red dots resp.) and sideband subtracted 2012 $B_{(s)}^0 \rightarrow h^+ h^-$ data (right, black dots).

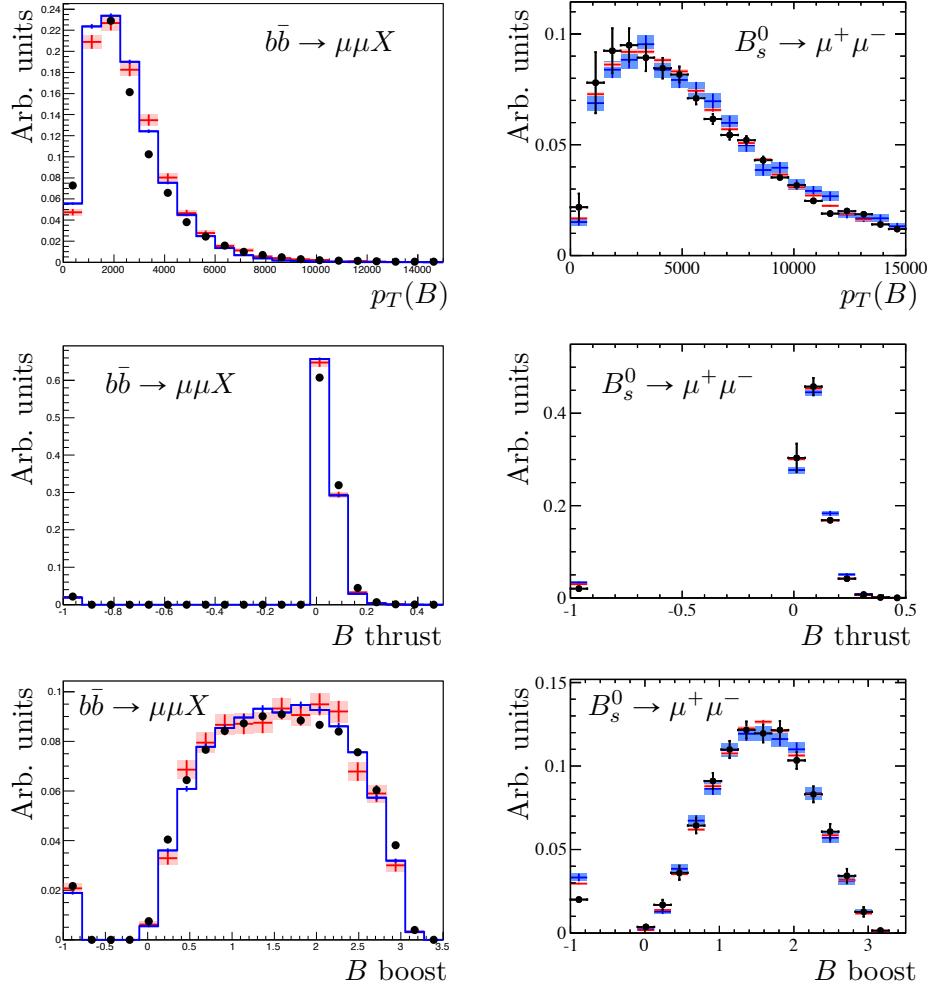


Figure 4.6: The distributions of the BDT “ $b\bar{b}$ pair” variables, shown for background (left) and signal samples (right). The background samples include $b\bar{b} \rightarrow \mu\mu X$ MC samples for 2011 and 2012 (left, red and blue dots resp.) and 2012 di-muon mass sideband data (left, black dots). The signal samples include $B_s^0 \rightarrow \mu^+\mu^-$ MC samples for 2011 and 2012 (right, blue and red dots resp.) and sideband subtracted 2012 $B_{(s)}^0 \rightarrow h^+h^-$ data (right, black dots).

If no good tracks are found originating from the B candidate vertex, the value is set to zero for B thrust, and $\pi/2$ for B boost. The distributions of this second group of variables are shown in Fig. 4.6.

In case the other tracks from the two decay vertices (X_b and X'_b in Fig. 4.3) are well reconstructed, the long lived combinatorial background can be identified by variables belonging to a third group.

The “daughter isolation” counts all possible good-quality vertices a daughter track can form in addition to the B candidate vertex; this tends to be higher for background. The “ B isolation”, on the other hand, estimates the fraction of transverse momentum carried by the B candidate with respect to the total transverse momentum of the tracks in a cone around the B candidate direction. For signal, the B itself contributes most of the transverse momentum in the cone.

The third group of variables includes:

- **Daughter isolation:** the sum of the daughter-track isolations. Track isolation is defined as the number of good two-track vertices a daughter track can form. Only vertices where the daughter track and the other track have a total momentum pointing back to the primary vertex are counted:

$$\frac{|\vec{p}_{\mu/h} + \vec{p}_{\text{tr}}| \cdot \sin(\alpha^{(\mu/h,\text{tr})})}{|\vec{p}_{\mu/h} + \vec{p}_{\text{tr}}| \cdot \sin(\alpha^{(\mu/h,\text{tr})}) + \vec{p}_{T,\mu/h} + \vec{p}_{T,\text{tr}}} < 0.6, \quad (4.1)$$

where $\alpha^{(\mu/h,\text{tr.})}$ is the angle between the sum of the momentum of the daughter and the other track, and the line connecting their vertex to the (B candidate) primary vertex;

- **B isolation:** defined as:

$$I(B) = \frac{p_T(B)}{p_T(B) + \sum_{\text{tracks}} p_T}, \quad (4.2)$$

where the sum runs over all the tracks in a cone of $\sqrt{(\delta\eta)^2 + (\delta\phi)^2} < 1.0$ around the B candidate direction, excluding the daughter tracks.

The distributions of the “isolation” variables are shown in Fig. 4.7.

The fourth group of variables uses the geometrical properties of the two-body B meson decays. Because of the LHCb limited angular acceptance, the two daughters will miss the detector when emitted at certain angles with respect the beam and the B meson momentum direction. The background decays, built from two random daughter tracks, need not obey that constraint.

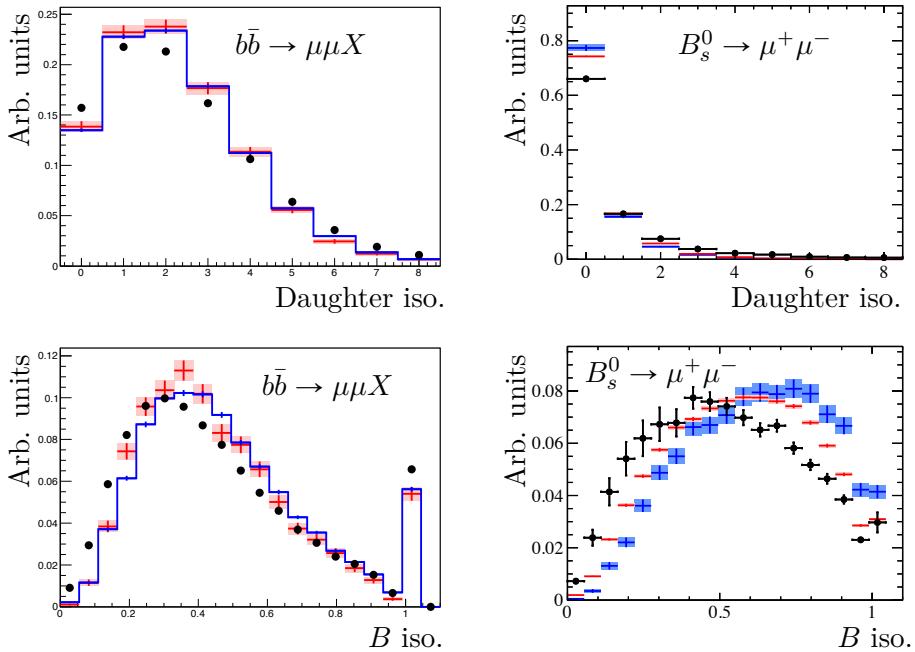


Figure 4.7: The distributions of the BDT ‘isolation’ variables, shown for background (left) and signal samples (right). The background samples include $b\bar{b} \rightarrow \mu\mu X$ MC samples for 2011 and 2012 (left, red and blue dots resp.) and 2012 di-muon mass sideband data (left, black dots). The signal samples include $B_s^0 \rightarrow \mu^+\mu^-$ MC samples for 2011 and 2012 (right, blue and red dots resp.) and sideband subtracted 2012 $B_{(s)}^0 \rightarrow h^+h^-$ data (right, black dots).

Additionally, owing to the relatively high longitudinal and transverse momentum of the B meson, the daughters from a true signal decay tend to be closer to each other in the pseudo-rapidity (η) and spherical angle (ϕ). The three geometrical variables included in the BDT are:

- **Polarisation angle:** cosine of the angle between the momentum of the lowest- p_T daughter (calculated in the B candidate rest-frame) and the vector perpendicular to the B momentum and the beam axis;
- **$\Delta\phi$:** the azimuthal angle difference between the daughter tracks.
- **$\Delta\eta$:** the pseudo-rapidity difference between the daughter tracks.

The distributions of the ‘geometrical’ variables are shown in Fig. 4.8.

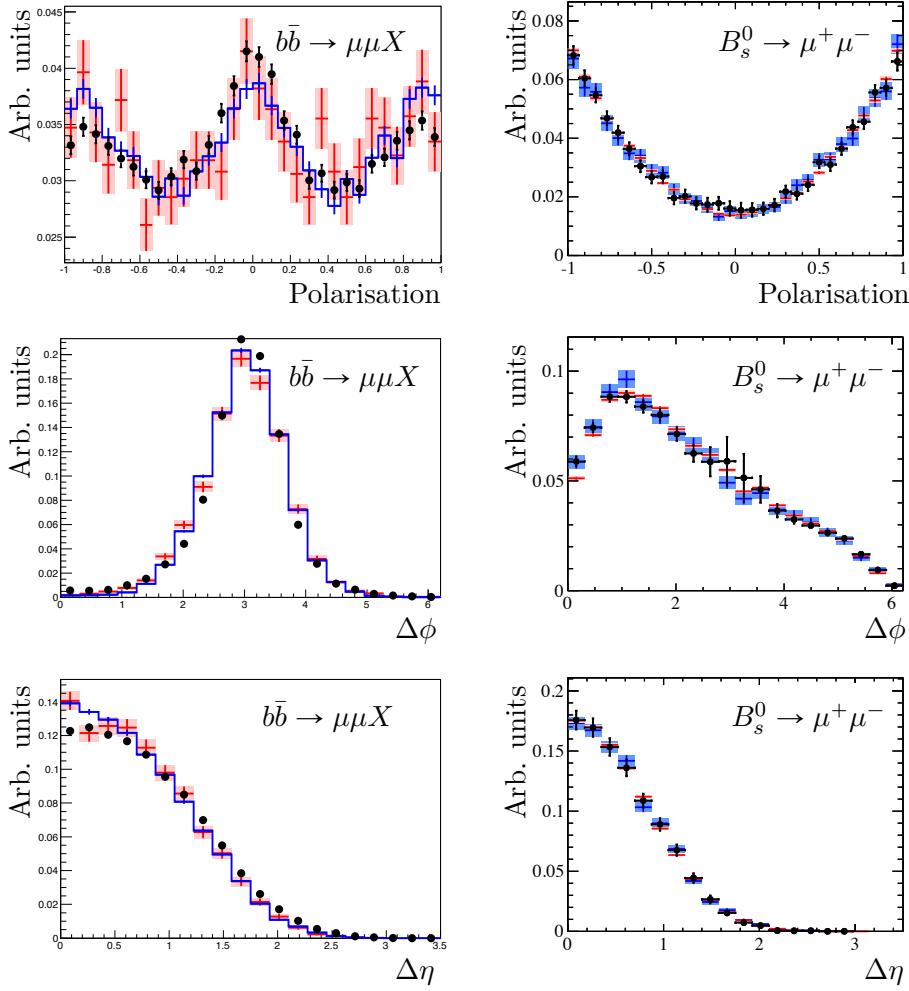


Figure 4.8: The distributions of the “geometrical” BDT variables, shown for background (left) and signal samples (right). The background samples include $b\bar{b} \rightarrow \mu\mu X$ MC samples for 2011 and 2012 (left, red and blue dots resp.) and 2012 di-muon mass sideband data (left, black dots). The signal samples include $B_s^0 \rightarrow \mu^+\mu^-$ MC samples for 2011 and 2012 (right, blue and red dots resp.) and sideband subtracted 2012 $B_{(s)}^0 \rightarrow h^+h^-$ data (right, black dots).

4.4.1 BDT binning, training, and performance

The BDT output lies in the range between zero and one. It is designed to be uniformly distributed for signal while for background it peaks around zero and drops fast towards higher BDT values. All candidates are classified according to the BDT output into 8 BDT categories, with bin boundaries [0.00, 0.25, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00].

The BDT performance depends on various training and boosting parameters in the TMVA package⁶. A number of separate BDT's were trained with different TMVA configurations, and the best performing BDT was chosen by optimising simultaneously for the signal efficiency, the background rejection, and the signal sensitivity over the BDT categories. To avoid over-training, the performance was evaluated on independent signal and background Monte Carlo test samples.

The chosen BDT should not be correlated to the di-muon invariant mass, in order to allow for a simple factorisation of the two-dimensional likelihood model (see Ch. 6). In the case of background, a correlation between the BDT and the di-muon mass could distort the background estimation in the signal region, and thus lead to an over- or underestimated signal yield. While small *linear* correlations (up to 10%) can be accounted for by the fit model, the *non-linear* correlations must be avoided.

The BDT dependence on the di-muon mass is minimised by adjusting the TMVA training parameters. For signal decays, the BDT output has a small linear correlation to the di-muon mass because of the radiative effects (see Sec. 1.3.5), which, in case the photon is not reconstructed, result in worse B -meson reconstruction quality and lower candidate mass (Fig. 4.9).

⁶Including the number of decision trees, the minimum number of events required in a leaf node, the maximum allowed depth, the number of steps during the node cut optimisation, and the parameters for the boosting algorithms, e.g AdaBoost.

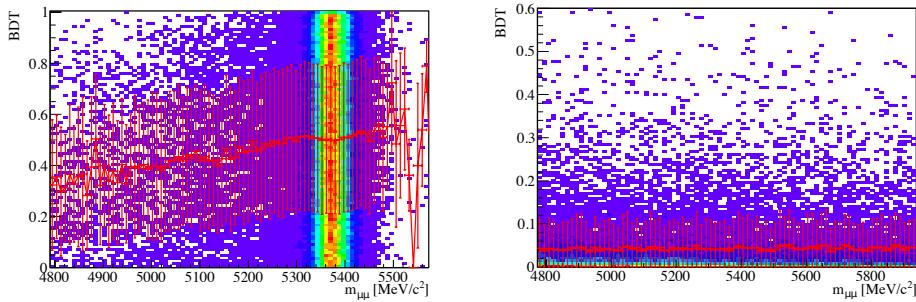


Figure 4.9: Correlation between the BDT and invariant mass of the two daughters, shown for Monte Carlo $B_s^0 \rightarrow \mu^+ \mu^-$ (left) and Monte Carlo $b\bar{b} \rightarrow \mu\mu X$ background (right). The points show the average BDT values for the candidates within the mass bins.

Chapter 5

Data-driven trigger efficiency estimation (TISTOS method)

In LHCb, every analysis begins with the trigger accepting an event. Not all interesting events are accepted though, and a correct normalisation of the event yields requires the determination of the trigger efficiency. This is a rather difficult task as it would require knowledge of those events that are missed. At LHCb, this delicate task can be performed directly from data with the TISTOS method. This chapter describes the principles of the TISTOS method and the underlying assumptions.

The main assumption of uncorrelated TIS and TOS decisions is better satisfied in small regions of phase space. Therefore, the total phase space is split in small regions and the TISTOS efficiency formula is applied accordingly. Section 5.1 explains how the TISTOS efficiency and its uncertainty can be calculated. In Sec. 5.2, the performance of the TISTOS method is demonstrated on simulated $B^+ \rightarrow J/\psi K^+$ samples, where the true efficiency is known. In Sec. 5.3, the performance of the TISTOS method is evaluated with $B^+ \rightarrow J/\psi K^+$ data.

5.1 TISTOS trigger efficiency

Various effects contribute to the detection efficiency of a given decay channel: the particles in the candidate events must be within the detector acceptance, then be triggered, reconstructed, and pass the offline selection requirements. Each consecutive step further reduces the sample leaving us with a subset of all the events. The overall efficiency can be written as a product:

$$\epsilon_{Tot} = \epsilon_{Acc} \cdot \epsilon_{Trig|Acc} \cdot \epsilon_{Rec|Trig} \cdot \epsilon_{Sel|Rec}. \quad (5.1)$$

The (conditional) trigger efficiency $\epsilon_{Trig|Acc}$ for a given decay channel is defined as the fraction of candidates within the acceptance that pass the trigger:

$$\epsilon_{Trig|Acc} \equiv \frac{N_{Trig|Acc}}{N_{Acc}}. \quad (5.2)$$

The detector records only triggered events, and the total number of signal events that the trigger processes (N_{Acc}) is not known. A standard solution to the problem of determining $\epsilon_{Trig|Acc}$ is that of a complete simulation of the trigger decision process. Here, an alternative procedure that makes use of measured data sets will be explained.

The idea is to use a data sample that would be triggered even in the absence of the signal. One can find out whether given trigger line would have triggered even in the absence of signal by comparing the “trigger report” for that line to the complete offline reconstruction. The event that pass the selection are used to evaluate the trigger efficiency for a particular decay channel. The total detection efficiency given in Eq. (5.1), is expressed as

$$\epsilon_{Tot} = \epsilon_{Trig|Sel} \cdot \epsilon_{Sel|Rec} \cdot \epsilon_{Rec|Acc} \cdot \epsilon_{Acc}, \quad (5.3)$$

where the trigger efficiency is now defined on the sample of selected (thus also reconstructed and accepted) events:

$$\epsilon_{Trig} \equiv \epsilon_{Trig|Sel} \equiv \frac{N_{Trig|Sel}}{N_{Sel}}. \quad (5.4)$$

To estimate the trigger efficiency in Eq. (5.4) from data, the triggered events are split into categories.

Trigger-line decision categories

The categorisation is performed in the offline reconstruction stage (Sec. 2.4), using the trigger decision reports stored for every triggered event. The reports contain a pass or no-pass decision for each trigger line, as well as the (online) reconstructed candidate on which the decision was based. Depending on the trigger line, a trigger line candidate can be either a single online track or a collection of online tracks (e.g. for di-muon or topological trigger lines).

After specifying a signal decay, all the trigger line candidates from the event trigger report are retrieved and one after the other compared to the signal tracks reconstructed offline. Each trigger-line decision is categorised as triggered-on-signal (TOS), if a substantial amount of hits are shared between

the trigger-line candidate tracks and the offline signal candidate tracks¹, and as triggered-independent-of-signal (TIS) if there is no overlap. Therefore, for a TIS decision, the presence of the signal candidate was not necessary for the trigger line to trigger the event. An example of the categorisation procedure is given in Fig. 5.1.

The categories are not mutually exclusive, e.g. a trigger line decision can simultaneously be categorised as TIS and TOS (TISTOS). The trigger line decisions that are neither TIS nor TOS, that is, neither the presence of the signal alone nor the rest of the event alone are sufficient to generate a positive trigger decision, are categorised as triggered-on-both (TOB)².

The candidates with at least one trigger-line decision categorised as TIS per trigger level are called TIS candidates. Their selection can be, to a good approximation, considered not biased by the trigger. Using these categories, the following partial efficiencies can be defined:

$$\epsilon_{TOS} \equiv \frac{N_{TOS|Sel}}{N_{Sel}}, \quad \epsilon_{TIS} \equiv \frac{N_{TIS|Sel}}{N_{Sel}}, \quad \epsilon_{TISTOS} \equiv \frac{N_{TISTOS|Sel}}{N_{Sel}}. \quad (5.5)$$

In terms of the TIS, TOS, and TISTOS categories, the trigger efficiency in Eq. (5.4) becomes:

$$\epsilon_{Trig} = \frac{N_{Trig|Sel}}{N_{Sel}} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \frac{N_{TIS|Sel}}{N_{Sel}} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \epsilon_{TIS}. \quad (5.6)$$

Henceforth, we will omit the “ $|Sel$ ” subscript with the understanding that all efficiencies are defined on a sample of selected events. Note that Eq. (5.6) is formally correct, but includes the unknown TIS efficiency, ϵ_{TIS} . While unknown for the overall sample, the TIS efficiency within the TOS sub-sample, defined as

$$\epsilon_{TIS|TOS} \equiv \frac{N_{TISTOS}}{N_{TOS}}, \quad (5.7)$$

can be measured. Provided that the TIS decision is truly independent of the signal candidates, its efficiency will be the same in any sub-sample of triggered signal events and the TIS efficiency for the total signal sample can be determined from the sub-sample of TOS events:

$$\epsilon_{TIS} \simeq \epsilon_{TIS|TOS}. \quad (5.8)$$

¹Typically at least 70% of the subdetector hits, the exact requirement varies between the sub-detectors. For a composite candidate, the combination of all individual trigger candidates is compared to the set of offline candidates.

²This applies to only a small fraction (0.05%) of the triggered events, and is irrelevant for the TISTOS method.

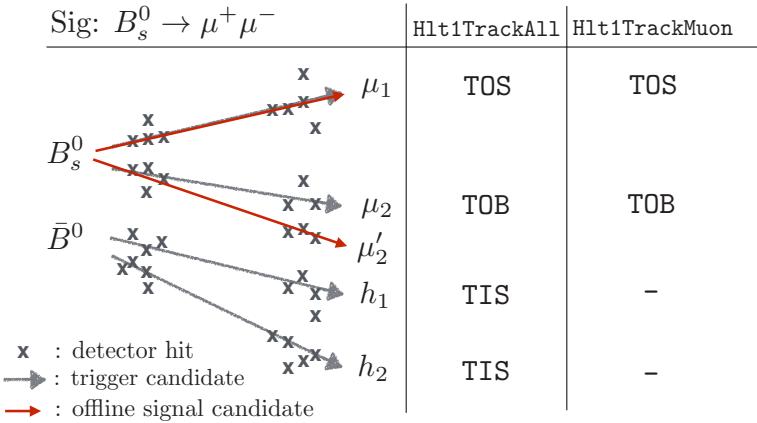


Figure 5.1: An illustration of the TIS-TOS categorisation for two trigger lines in HLT1, for an event containing a $B_s^0 \rightarrow \mu^+ \mu^-$ candidate (chosen as signal) and a $B_{(s)}^0 \rightarrow h^+ h^-$ candidate (from the other b quark). The trigger line candidates (grey arrows) are retrieved from the trigger reports and compared to the offline reconstructed signal candidate (red arrows). In case the offline signal candidates overlap with the trigger candidates, the trigger line decision on that candidate is classified as triggered-on-signal (TOS), if there is no overlap, the decision is classified as trigger-independent-of-signal (TIS); for rare cases of partial overlap, the decision is categorized as trigger-on-both (TOB).

The trigger efficiency can then be determined as

$$\epsilon_{Trig} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \frac{N_{TISTOS}}{N_{TOS}}, \quad (5.9)$$

where now all four quantities can be measured from data!

Equation 5.8 contains the main assumption of the method: it is valid if the TIS decisions for the lines containing a signal candidate can be considered random. The TIS decision is often caused by the other b -hadron decay in the event and this b -hadron originates from the same $b\bar{b}$ pair as the signal; therefore, the TIS decision could be related to the TOS decision through the “linked” kinematics of the signal B meson and the other b -hadron (the trigger selection often relies on transverse momentum and impact parameter cuts). This correlation is modest, but not negligible. It can be reduced by restricting

the method to small regions of the signal B meson phase space:

$$\epsilon_{Trig} = \frac{N_{Trig|Sel}}{\sum_i N_{Sel}^i} = \frac{N_{Trig|Sel}}{\sum_i \frac{N_{TIS|Sel}^i}{\epsilon_{TIS}^i}} = \frac{N_{Trig|Sel}}{\sum_i \frac{N_{TIS|Sel}^i N_{TOS|Sel}^i}{N_{TISTOS|Sel}^i}}, \quad (5.10)$$

where the sum is performed on all the bins in the phase space of the signal B meson.

5.1.1 Estimating the trigger efficiency uncertainty

Given all the necessary trigger yields, the trigger efficiency can be computed using Eq. (5.10). The computation of its uncertainty needs care since the TRIG, TIS, TOS, and TISTOS samples overlap. The trigger efficiency is re-written indicating the contribution of disjointed sets, $N_{Trig|sel} = n$ and $N_{sel} = n + m$:

$$\epsilon_{Trig} = \frac{n}{n + m}, \quad (5.11)$$

which leads to the following uncertainty expression:

$$\begin{aligned} \sigma_{\epsilon_{Trig}}^2 &= \left(\frac{m}{(n + m)^2} \right)^2 \cdot \sigma_n^2 + \left(\frac{-n}{(n + m)^2} \right)^2 \cdot \sigma_m^2, \\ \sigma_{\epsilon_{Trig}}^2 &= \left(\frac{m}{(n + m)^2} \right)^2 \cdot N_{Trig|Sel} + \left(\frac{-n}{(n + m)^2} \right)^2 \cdot (\sigma_{N_{Sel}}^2 - N_{Trig|Sel}). \end{aligned} \quad (5.12)$$

In the last step, $\sigma_{N_{Sel}}^2 = \sigma_n^2 + \sigma_m^2$ was used to express σ_m in terms of measurable yields.

The uncertainty on N_{SEL} is also calculated by expressing it first in terms of disjoint sets:

$$N_{Sel} = \sum_i N_{Sel}^i = \sum_i \frac{N_{TIS|Sel}^i N_{TOS|Sel}^i}{N_{TISTOS|Sel}^i} = \sum_i \frac{(b_i + d_i)(c_i + d_i)}{d_i}, \quad (5.13)$$

where we have denoted $N_{TISTOS|Sel}^i$ by d_i and the non-overlapping part of $N_{TIS|Sel}^i$ and $N_{TOS|Sel}^i$ by b_i and c_i (see Fig. 5.2). The uncertainty can be

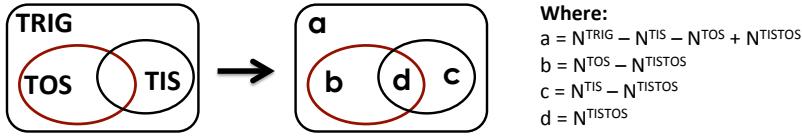


Figure 5.2: Redefining yields with independent terms in uncertainty calculation for selected number of events.

calculated as follows³:

$$\begin{aligned}
 \sigma_{N_{Sel}}^2 &= \sum_i \sigma_{N_{Sel}^i}^2 \\
 &= \sum_i \left(\frac{\partial N_{Sel}^i}{\partial b_i} \right)^2 \sigma_{b,i}^2 + \left(\frac{\partial N_{Sel}^i}{\partial c} \right)^2 \sigma_{c,i}^2 + \left(\frac{\partial N_{Sel}^i}{\partial d} \right)^2 \sigma_{d,i}^2, \quad (5.14) \\
 &= \sum_i \left(\frac{c_i + d_i}{d_i} \right)^2 b_i + \left(\frac{b_i + d_i}{d_i} \right)^2 c_i + \left(1 - \frac{b_i c_i}{d_i^2} \right)^2 d_i.
 \end{aligned}$$

5.1.2 Binning the phase space

The phase space of the B meson can be parametrised by the transverse and longitudinal momentum. The binning boundaries are defined independently for the two variables and optimised such that about the same number of TISTOS events fall into each bin⁴. After optimisation, the number of events over the bins agrees within few percents. When using these independently optimised bin boundaries to bin the two-dimensional phase space, the bin yields show a greater variation because transverse and longitudinal momentum are not independent variables. This correlation does not jeopardise the performance of the TISTOS method for small bin areas.

³Here we assume the yield uncertainties follow a Gaussian distribution and use the first-order Taylor expansion for the uncertainty propagation. For the Gaussian assumption to be valid, the yields in each bin must be large enough ($\mathcal{O}(10)$).

⁴TISTOS is the category with the smallest statistics and therefore influences the method performance the most. Dividing events equally among the bins minimises the chance of encountering a bin with too low or no statistics.

5.2 Performance on simulation

In this section, the **TISTOS** method is demonstrated on simulated $B^+ \rightarrow J/\psi K^+$ events. In a simulated sample, the results can be compared to the true efficiency. Interactions are accepted if one or more⁵ $B^+ \rightarrow J/\psi K^+$ candidates are within the LHCb detector acceptance. The accepted MC events are processed by the same trigger, reconstruction, and selection algorithms as the events from real pp collisions (see Sec. 2.4).

Two $B^+ \rightarrow J/\psi K^+$ MC samples are produced with different sample size and using different detector configurations. The smaller sample (MC127k) contains 127×10^3 events generated in the detector acceptance and uses the detector configuration during the data taking in May and June 2012. The larger MC sample of 10^6 events (MC1000k) uses the detector configuration from July, August, and September 2012. For both samples, the pp interactions have been simulated assuming a beam energy of 4 TeV and an average number of visible interactions per crossing $\nu = 1.75$.

In MC, one knows also how many selected signal events do not pass the trigger requirements and the trigger efficiency defined in Eq. (5.4) can be evaluated directly from N^{TRIG}/N^{SEL} . In the following, this is referred to as the *true trigger efficiency*. The true trigger efficiency is an important benchmark for the MC study and allows us to test how the **TISTOS** method performs. The difference in the results (bias) is treated as the systematic uncertainty of the **TISTOS** method.

5.2.1 Signal separation

The trigger efficiency is defined for a specific signal decay and is calculated on a sample of true signal events. The signal candidates are distinguished from the other processes in the pp collisions using the selection criteria and the B mesons invariant mass techniques: the side-band Subtraction (SB) or the Maximum Likelihood (ML) fit. In this study, the SB serves as the main method. This is mainly because of its robustness when dealing with bins containing few events, but also because of its generality and its simple application on binned phase space. The ML is applied as a cross check.

For each selected $B^+ \rightarrow J/\psi K^+$ candidate, the B^+ invariant mass is calculated. Only candidates with B masses in a mass window of $\pm 100 \text{ MeV}/c^2$ around the B^+ meson mass value of $5279.17 \text{ MeV}/c^2$ [103] are considered. In the SB method, candidates with reconstructed B^+ mass off by more than $55 \text{ MeV}/c^2$ from the mean mass value⁶, belong to the *side-bands*. The yield

⁵The fraction of events with more than one signal decay is at 0.1% level.

⁶The B^+ invariant mass resolution is LHCb is about $20 \text{ MeV}/c^2$

in the side-bands is considered as a background proxy, extrapolated over the whole mass window, and subtracted from the total yield to derive the signal candidate yield.

In the ML fit, the B^+ meson invariant mass distribution is modelled with a probability density function (*pdf*) that is fitted to the mass distribution in the sample. The *pdf* for $B^+ \rightarrow J/\psi K^+$ (signal) consists of two Gaussians with a common mean, but different (independent) widths. The background model has two parts: an exponential function to describe the combinatorial background, and a Crystal Ball function⁷ to describe cases in which a pion is mis-identified as a kaon. This latter component has a mean that is fixed with respect to the signal mean mass value. The `RooFit` package [104] evaluates the total likelihood of the sample, and the `MINUIT` minimizer [105] finds the set of *pdf* parameter values that maximise the total likelihood. The $B^+ \rightarrow J/\psi K^+$ yield is one of the model parameters.

The SB and ML are tested on the MC sample, where the true $B^+ \rightarrow J/\psi K^+$ candidates can be identified by a MC matching procedure. The matching procedure relies on the MC truth information, available for every particle in an simulated event. For $B^+ \rightarrow J/\psi K^+$ to be matched as a signal candidate, the particle identities must match and the particles must be properly linked in the decay chain (e.g. the J/ψ needs to originate from the B^+ decay, etc.). The SB and ML are applied on (i) matched, (ii) not-matched, and (iii) the total MC $B^+ \rightarrow J/\psi K^+$ sample. The results from the two methods, as shown in Tab. 5.1, are almost indistinguishable. As expected, the efficiency calculated on the not-matched sample differs due to the larger background contribution.

The compatible results from the total and matched sample show that the SB and ML methods are suitable to determine the trigger efficiency from the total sample without further MC matching. The ϵ_{true} calculated on the total sample is therefore used as benchmark in the following MC studies.

5.2.2 Results with different binning schemes

In the LHCb collaboration, the `TISTOS` method is widely used but mostly without binning in phase space. This results in a considerable increase of the systematic uncertainty assigned to the trigger efficiency estimation. For the MC127k sample, the `TISTOS` method without binning gives efficiencies more than 5% higher than the true efficiency. The relative bias of the `TISTOS` method (and thus its systematic uncertainty) can be significantly reduced by

⁷A function with a Gaussian core and an exponential tail. The exponential tail is usually used at the low end below a certain threshold to model the radiative energy loss processes, see Sec. 6.3.

Table 5.1: True un-binned trigger efficiency on matched (ϵ_{match}), on the remaining non-matched ($\epsilon_{notmatch}$), and on the whole MC127k sample (ϵ_{true}).

Signal separation method	ϵ_{match}	$\epsilon_{notmatch}$	ϵ_{true}
SB	$(87.39 \pm 0.22)\%$	$(84.44 \pm 1.61)\%$	$(87.32 \pm 0.22)\%$
ML	$(87.37 \pm 0.22)\%$	$(85.56 \pm 1.54)\%$	$(87.32 \pm 0.22)\%$

binning the B meson phase space, applying the method in every bin, and combining the results into an overall efficiency of the sample as described in Sec. 5.1.

The TISTOS method was applied to both MC samples. The number of the B meson transverse and longitudinal momentum bins was increased in steps, as shown in Figs. 5.3 and 5.4. The TISTOS results converge to the true efficiency value if the number of bins in the B meson phase space is increased. As a cross check, an identical study was performed using the ML method instead of the SB. The comparison between the SB and the ML on the MC127k sample is shown in Fig. 5.5. For the lower numbers of bins, the results are in good agreement, but, as the number of bins increases, the individual bins contain less and fewer statistics and eventually leave the ML fit too little statistics to reliably separate the signal contribution.

The optimal number of transverse and longitudinal momentum bins is a compromise between the smallest relative bias and the statistical uncertainty of the efficiency. The optimal binning scheme depends on the sample size and could be sample specific. The dependence on the sample is studied on MC samples with similar configurations but different sizes (MC127k and MC1000k). The best binning scheme for the smaller sample (MC127k) is 4 bins in p_z and 5 in p_T . Comparing to the un-binned results, the relative bias of the method was reduced from $(5.7 \pm 1.4)\%$ to $(0.5 \pm 2.5)\%$. The best binning scheme for the larger sample (MC1000k) was found to be 4 bins in p_z and 9 in p_T , and it reduced the relative bias from $(3.9 \pm 0.6)\%$ to $(0.3 \pm 0.8)\%$ (Tab. 5.2). The best binning schemes on the smaller and larger MC sample are very similar. Moreover, a change in the number of bins has a small impact on the bias when the number of bins exceeds 3 in both dimensions (see Figs. 5.3 and 5.4). The optimal binning scheme can therefore be considered independent of the sample, and can be pre-determined on a MC sample.

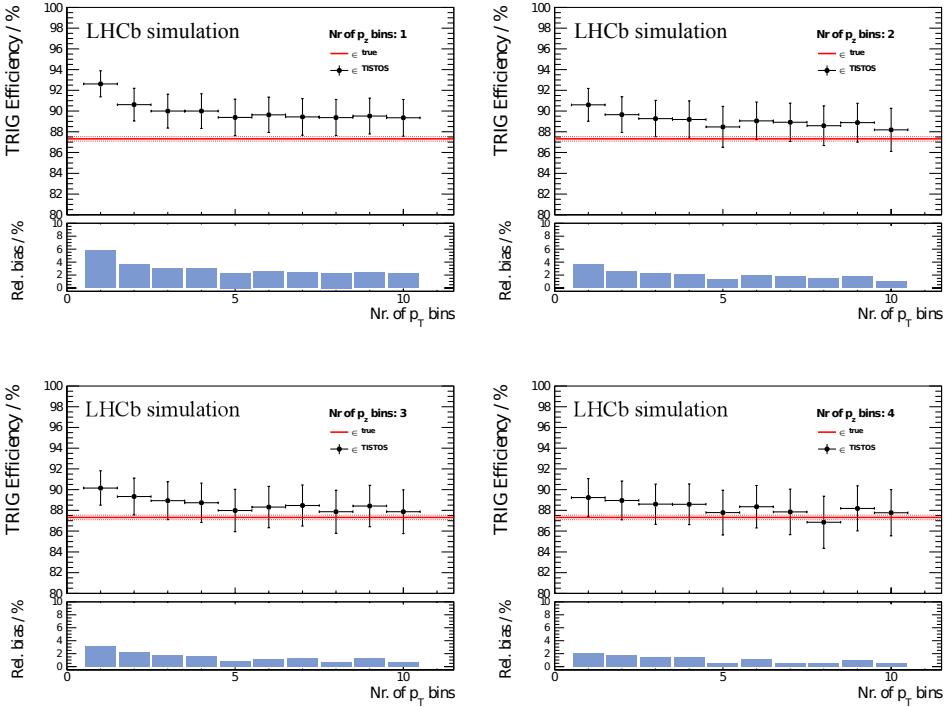


Figure 5.3: Trigger efficiency calculated with the TISTOS method as a function of the number of bins in the B meson p_T , for different binning in the p_Z (MC127k sample). The true efficiency is denoted by the red line.

Table 5.2: Efficiency evaluated with the TISTOS method (ϵ_{TisTos}) and its relative bias with respect to the ϵ_{true} ($Bias(\epsilon)$) without binning and with optimal binning schemes.

Sample / Binning	ϵ_{true}	ϵ_{TisTos}	$Bias(\epsilon)$
MC127k/ No binning	$(87.3 \pm 0.2)\%$	$(92.6 \pm 1.3)\%$	$(5.7 \pm 1.4)\%$
MC127k/ 4x5 binning in (p_Z, p_T)	$(87.3 \pm 0.2)\%$	$(87.8 \pm 2.2)\%$	$(0.5 \pm 2.5)\%$
MC1000k/ No binning	$(87.6 \pm 0.1)\%$	$(91.2 \pm 0.5)\%$	$(3.9 \pm 0.6)\%$
MC1000k/ 4x9 binning in (p_Z, p_T)	$(87.6 \pm 0.1)\%$	$(87.9 \pm 0.7)\%$	$(0.3 \pm 0.8)\%$

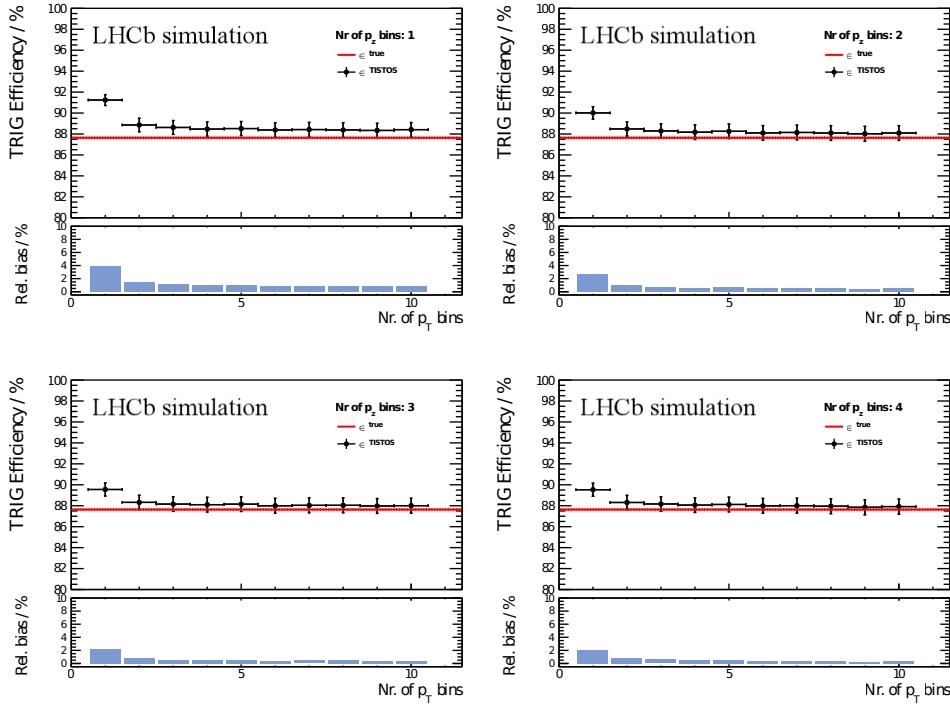


Figure 5.4: Trigger efficiency calculated with the TISTOS method as a function of the number of bins in the B meson p_T , for different binning in the p_Z (MC1000k sample). The true efficiency is denoted by the red line.

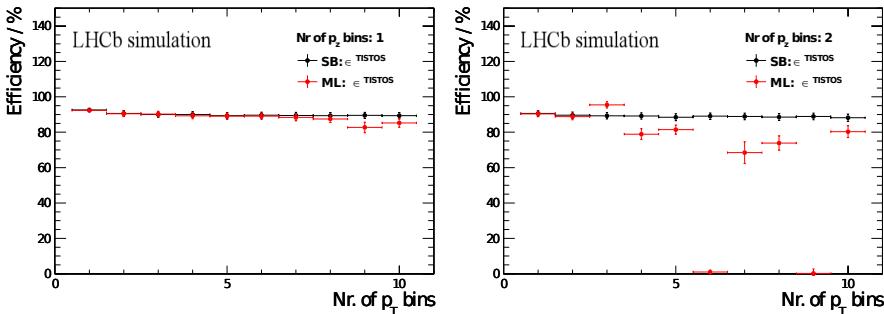


Figure 5.5: The TISTOS trigger efficiency (MC127k sample) for different number of B meson p_Z and p_T bins, calculated with the SB and the ML methods. The black points represent the results of the SB method, and the red points the results of the ML method.

5.3 Performance on data

In this section, the **TISTOS** method is applied to the full LHCb $B^+ \rightarrow J/\psi K^+$ data sample collected in 2011 and 2012. The systematic uncertainty of the method and the optimal phase space binning scheme are determined from the simulated MC1000k sample, as described in Sec. 5.2. The systematic uncertainty is assumed not to vary with the small trigger configuration changes over the data taking period. The optimal binning *scheme* is determined on MC, bin *boundaries* are optimised for the 2011 and 2012 samples individually, as described in Sec. 5.1.2.

The trigger efficiencies for the $B^+ \rightarrow J/\psi K^+$ sample in 2011 and 2012 are given in Tab. 5.3. The estimates from the binned **TISTOS** method are (4–5)% lower, similarly to what was observed in the MC samples (Tab. 5.2). The systematic uncertainty of the method, which mainly arises due to correlations between TIS and TOS trigger decisions, can be reduced by applying the method in bins of phase space. The total systematic uncertainty of the **TISTOS** method in 2011 and 2012 $B^+ \rightarrow J/\psi K^+$ data can be reduced from 3.6% to 0.2%, which is to the same level as the statistical uncertainty.

Table 5.3: Trigger efficiencies from the $B^+ \rightarrow J/\psi K^+$ data, estimated using the **TISTOS** method, with and without binning the B meson phase space.

Binning	ϵ_{TisTos}	<i>Abs.Stat.Unc.</i>	<i>Abs.Syst.Unc.</i>	<i>Rel.Syst.Unc.</i>
Data: 2011 S20r1				
No binning	92.9%	0.5%	3.6%	3.9%
4x9 binning in (p_L, p_T)	87.8%	0.6%	0.2%	0.3%
Data: 2012 S20				
No binning	92.0%	0.3%	3.6%	3.9%
4x9 binning in (p_L, p_T)	87.8%	0.4%	0.2%	0.3%

Chapter 6

Signal yield extraction

The invariant di-muon mass is a powerful way of selecting muon pairs from the B_s^0 and B^0 meson decays. Its discrimination power is further enhanced by dividing the candidates in categories according to the output of the multivariate classifier BDT, discussed in Ch. 4. The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yield is then extracted with a Maximum Likelihood Fit (ML) from the di-muon mass distribution. This chapter describes the likelihood model, while the parameter estimation with the ML is described in Ch. 8.

6.1 The di-muon mass model

The invariant mass distribution of the candidates in the selected $B_s^0 \rightarrow \mu^+ \mu^-$ sample (Tab. 4.1) is modelled by the likelihood function

$$\mathcal{L}^{LHCb}(\vec{\lambda}) = \overbrace{\frac{(\mu_{Sig} + \mu_{Bkg})^{N_{obs}} \cdot e^{-(\mu_{Sig} + \mu_{Bkg})}}{N_{obs}!}}^{\text{Poisson distr.}} \times \prod_{i=1}^{N_{obs}} pdf \left(\text{BDT}^i, m_{\mu^+ \mu^-}^i; \vec{\lambda} \right). \quad (6.1)$$

The *pdf* in Eq. (6.1) is a function of two variables: BDT^i and $m_{\mu^+ \mu^-}^i$, the BDT output and the di-muon mass of the i -th $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidate; $\vec{\lambda}$ is a set of parameters defining the *pdf*; N_{obs} is the total number of candidates in the 3 fb^{-1} data sample.

The BDT output is designed to be independent of the di-muon invariant mass, and the correlation has been cross-checked to be very small (Sec. 4.4.1). This allows to factorise the two dimensional probability density function in

Eq. (6.1) into two independent terms:

$$pdf(BDT^i, m_{\mu^+\mu^-}^i; \vec{\lambda}) = pdf(BDT^i; \vec{\lambda}_1) \times pdf(m_{\mu^+\mu^-}^i; \vec{\lambda}_2). \quad (6.2)$$

The first term in Eq. (6.2) models the BDT distribution and the second one models the di-muon mass distribution. Separating the signal and background components, Eq. (6.2) reads

$$\begin{aligned} pdf(BDT^i, m_{\mu^+\mu^-}^i; \vec{\lambda}) = & f_{Sig} \left\{ pdf_{Sig}(BDT^i; \vec{\lambda}_{1,S}) \times pdf_{Sig}(m_{\mu^+\mu^-}^i; \vec{\lambda}_{2,S}) \right\} + \\ & f_{Bkg} \left\{ pdf_{Bkg}(BDT^i; \vec{\lambda}_{1,B}) \times pdf_{Bkg}(m_{\mu^+\mu^-}^i; \vec{\lambda}_{2,B}) \right\}. \end{aligned} \quad (6.3)$$

where

$$f_{Sig} = \frac{\mu_{Sig}}{N_{obs}}, \quad f_{Bkg} = \frac{\mu_{Bkg}}{N_{obs}}. \quad (6.4)$$

Using the eight categories (Sec. 4.4.1) to describe the BDT dependence, Eq. (6.3) becomes

$$\begin{aligned} pdf(BDT^i, m_{\mu^+\mu^-}^i; \vec{\lambda}) = & \sum_{cat=1}^8 \left\{ f_{Sig} \left(f_{Sig}^{cat} \times pdf_{Sig}^{cat}(m_{\mu^+\mu^-}^i; \vec{\lambda}_{2,S}) \right) + \right. \\ & \left. f_{Bkg} \left(f_{Bkg}^{cat} \times pdf_{Bkg}^{cat}(m_{\mu^+\mu^-}^i; \vec{\lambda}_{2,B}) \right) \right\}. \end{aligned} \quad (6.5)$$

Note, that in order to account for possible remaining BDT and mass correlations, the mass pdf parameters are allowed to differ from category to category, as denoted by the “cat” superscript.

The $B_s^0 \rightarrow \mu^+\mu^-$ decays and $B^0 \rightarrow \mu^+\mu^-$ decays are simultaneously searched for and therefore the signal mass pdf has two independent components, one for each decay. The BDT fractions, however, are the same for the two components. The signal component in Eq. (6.5) can thus be re-written as

$$\begin{aligned} pdf_{Sig}(BDT^i, m_{\mu^+\mu^-}^i) = & \frac{1}{N_{obs}} \sum_{cat}^8 \left\{ f_{Sig}^{cat} \cdot N_{Bd} \times pdf_{Bd}^{cat}(m_{\mu^+\mu^-}^i) + \right. \\ & \left. f_{Sig}^{cat} \cdot N_{Bs} \times pdf_{Bs}^{cat}(m_{\mu^+\mu^-}^i) \right\}. \end{aligned} \quad (6.6)$$

The background pdf is expressed in terms of three background sources: doubly mis-identified component from $B_{(s)}^0 \rightarrow h^+h^-$ decays (Sec. 6.4), singly

mis-identified component from semi-leptonic $B \rightarrow h\mu X$ decays¹ (Sec. 6.5), and the combinatorial background from arbitrary muon combinations (Sec. 6.6):

$$\begin{aligned} pdf_{Bkg}(BDT^i, m_{\mu^+\mu^-}^i) = & \frac{1}{N_{obs}} \sum_{cat}^8 \left\{ f_{hh}^{cat} \cdot N_{hh} \times pdf_{hh}^{cat}(m_{\mu^+\mu^-}) + \right. \\ & f_{h\mu}^{cat} \cdot N_{h\mu} \times pdf_{h\mu}^{cat}(m_{\mu^+\mu^-}) + f_{Comb.}^{cat} \cdot N_{Comb.} \times pdf_{Comb.}^{cat}(m_{\mu^+\mu^-}) \left. \right\}. \end{aligned} \quad (6.7)$$

The signal yields N_{Bs} and N_{Bd} can be extracted from the fit if all the other model components are well defined. The signal fraction in each BDT category, f_{Sig}^{cat} , is determined from the fraction of $B_{(s)}^0 \rightarrow h^+h^-$ decays in the same BDT category (Sec. 6.2); the shape of the signal mass distribution is deduced from simulations, with central values and resolutions calibrated on data (Sec. 6.3). The semi-leptonic background contributes little to the $B^0 \rightarrow \mu^+\mu^-$ mass region and is negligible in the $B_s^0 \rightarrow \mu^+\mu^-$ region, but, because of the peaking shape of its di-muon mass pdf , it needs to be modelled separately from the combinatorial; only then can the di-muon mass distribution of the combinatorial background be modelled with an exponential function. The shape of the semi-leptonic background invariant mass distribution is determined from simulated samples (Sec. 6.5). The doubly mis-identified $B_{(s)}^0 \rightarrow h^+h^-$ background has also a peaking shape, and thus needs a separate model from the combinatorial component. The yield of doubly mis-identified events is determined from $B_{(s)}^0 \rightarrow h^+h^-$ data (Sec. 6.4); the di-muon mass distribution of these candidates is studied and calibrated with a simulated $B_{(s)}^0 \rightarrow h^+h^-$ sample.

6.2 Signal BDT distribution

The BDT deliberately does not include particle identification information: it is trained to separate true 2-body B decays from the multi-body decays, and to treat $B_{(s)}^0 \rightarrow h^+h^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ equally (see Sec. 4.4). Instead of relying on the simulation to determine the signal BDT distribution, the $B_{(s)}^0 \rightarrow h^+h^-$ candidates can be used as proxies for the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays, and the signal BDT distribution can be measured on data with a procedure referred to as “BDT calibration”.

The BDT was trained, optimised, and tested on simulated samples (Sec. 4.4.1). The operator was then applied to a data sample containing the candidates that pass the $B_{(s)}^0 \rightarrow h^+h^-$ selection (Tab. 4.1), including the

¹Including the $B^{0/+} \rightarrow \pi^{0/+} \mu^+\mu^-$ component.

BDTS cut. Although the fractions of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in the 8 BDT categories could be determined on simulated samples, this data-driven approach provides a more reliable BDT calibration, and is thus preferred.

The number of $B_{(s)}^0 \rightarrow h^+h^-$ candidates in each BDT category is extracted from the $B_{(s)}^0 \rightarrow h^+h^-$ invariant mass distribution, in the range [5000, 5600] MeV/ c^2 . The main difference between $B_{(s)}^0 \rightarrow h^+h^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays is that they are triggered by very different trigger lines (see Ch. 3). The hadron-trigger bias could be avoided if TIS $B_{(s)}^0 \rightarrow h^+h^-$ candidates were used in the $B_{(s)}^0 \rightarrow h^+h^-$ BDT calibration. In practice, however, the low statistics of the TIS sample obliges us to use a $B_{(s)}^0 \rightarrow h^+h^-$ sample which is TIS for L0 and HLT1, but not HLT2 (this compromise raises the trigger efficiency from less than 2% to more than 5%).

The di-hadron mass is calculated using one of four possible final state hypotheses: $\pi^+\pi^-$, π^+K^- , $K^+\pi^-$, or K^+K^- , where the hadrons are identified as kaons if $\Delta LL(K - \pi) > \kappa$, and as pions if $\Delta LL(K - \pi) < -\kappa$, and rejected otherwise; the value of κ is varied from 0 to 10. These four final state configurations get contributions from the different B and Λ_b^0 decays listed in Tab. 6.1. A di-hadron mass model is built for each final state hypothesis, and it consists of four components:

- the B_s^0 and B^0 decays are described by two independent double-sided “Crystal Ball” functions (Sec. 6.3). All parameters are fixed (as determined from a Monte Carlo sample) except a common resolution² parameter;
- the $\Lambda_b^0 \rightarrow ph^-$ contamination, with a proton mis-identified as a kaon or a pion, is described by a double-sided Crystal Ball function, with all parameters fixed (from a Monte Carlo sample);
- the combinatorial background is described by an exponential function;
- the partially reconstructed B decays (e.g. $B^0 \rightarrow K^+\pi^-\pi^0$) are described by a specific function, called ROOPHYSBKG, first used in Ref. [107]:

$$\mathcal{F}_P(m_{hh}) = \frac{1}{A} \cdot m' \left(1 - \frac{m'^2}{m_0^2} \right) \cdot \Theta(m_0 - m') \cdot e^{-C_P \cdot m'} \otimes G(m_{hh} - m'; \sigma_P), \quad (6.8)$$

where A is a normalisation constant, Θ the Heaviside step function, m_0 a free fitted parameter, the symbol \otimes represents a convolution product

²The expected difference in the B_s^0 and B^0 resolutions is less than 2%.

over m' , G is a Gaussian function with a standard deviation σ_P , and C_P a shape parameter. The parameter m_0 acts as the higher limit of the partially reconstructed mass distribution in the absence of resolution effects, and in practice is close to the value of the B mass minus the masses of the missing particles [108].

The kaon and pion identification requirements are necessary to identify the four $B_{(s)}^0 \rightarrow h^+h^-$ final states. The efficiency of these requirements, however, depends on the candidate kinematics and on the number of tracks, and therefore will affect the distribution of the $B_{(s)}^0 \rightarrow h^+h^-$ events in the BDT categories and bias the BDT shape. The effect of the kaon and pion requirements is corrected for by giving each candidate an “efficiency weight”:

$$w = \frac{1}{\epsilon_{\Delta LL}(p, \eta, nTr|h^+, \kappa) \cdot \epsilon_{\Delta LL}(p, \eta, nTr|h^-, \kappa)}, \quad (6.9)$$

where the $\Delta LL(K - \pi)$ efficiency, $\epsilon_{\Delta LL}$, is obtained from $D^0 \rightarrow K^\mp \pi^\pm$ decays (see Sec. 2.5) in bins of the hadron momentum, the hadron pseudo-rapidity, and the number of tracks in the event.

After the $\Delta LL(K - \pi)$ efficiency correction, the BDT shape should not depend on the strength of the $\Delta LL(K - \pi)$ threshold, κ . The only remaining dependence is of statistical nature, and, to reduce the statistical fluctuations, the fraction of $B_{(s)}^0 \rightarrow h^+h^-$ candidates in a given BDT category is calculated as a weighted average over different ΔLL cut values (see Fig. 6.1). The BDT fraction uncertainty includes the uncertainties from the PID correction in Eq. (6.9) and from the fitting procedure; the averaging procedure takes into account the correlation between the samples at different κ values. The BDT shape determined from $B_{(s)}^0 \rightarrow h^+h^-$ decays and corrected for the particle identification effects is shown in Fig. 6.1.

Table 6.1: Decays contributing to the h^+h^- final state, in the invariant mass range $[5000, 5600]$ MeV/ c^2 . As Λ_b^0 decays contribute only in case the proton is mis-identified as a kaon or a pion, their contribution is small. The branching fractions are taken from Ref. [106]

Final state	B^0 branching fraction	B_s^0 branching fraction	Λ_b^0 branching fraction
$\pi^+\pi^-$	$(5.12 \pm 0.19) \times 10^{-6}$	$(7.6 \pm 1.9) \times 10^{-7}$	—
$K^+\pi^-$	$(1.96 \pm 0.05) \times 10^{-5}$	—	—
π^+K^-	—	$(5.5 \pm 0.6) \times 10^{-6}$	—
K^+K^-	$(1.3 \pm 0.5) \times 10^{-7}$	$(2.49 \pm 0.17) \times 10^{-5}$	—
$p^+\pi^-$	—	—	$(4.1 \pm 0.8) \times 10^{-6}$
p^+K^-	—	—	$(4.9 \pm 0.9) \times 10^{-6}$

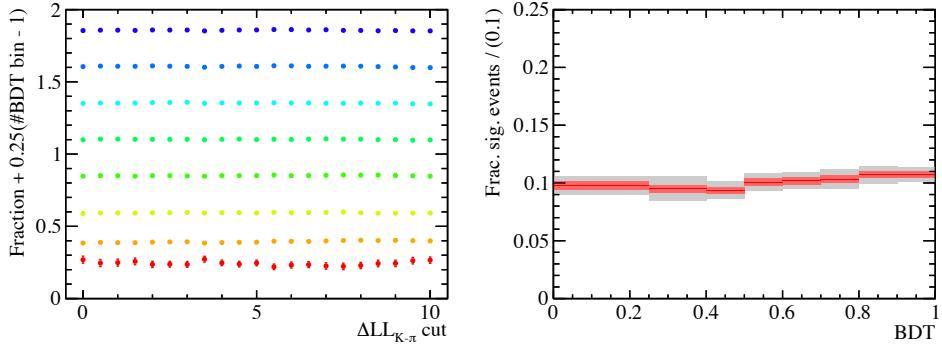


Figure 6.1: Left: the fraction of $B_{(s)}^0 \rightarrow h^+h^-$ candidates in the BDT categories for different ΔLL cut values. Note that the candidate fractions in the BDT categories are shifted vertically for illustrative purpose. Right: The probability distribution of the signal events in the BDT bins, as calculated on the $B_{(s)}^0 \rightarrow h^+h^-$ sample. The red (grey) band demonstrates the statistical (systematic) uncertainty, and the values are normalised to bin size of 0.1.

6.2.1 From the $B_{(s)}^0 \rightarrow h^+h^-$ BDT to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT

The signal BDT distribution is determined by correcting the BDT distribution obtained from $B_{(s)}^0 \rightarrow h^+h^-$ decays (Fig. 6.1, right) for the differences between the hadron and muon channels. These differences arise from the different phase space distribution, trigger selection, event reconstruction, and candidate selection.

The signal and the $B_{(s)}^0 \rightarrow h^+h^-$ decays have almost identical geometrical properties, with minor differences arising from the different masses of the decay products. The selection of both samples is identical, up until the muon identification cuts (Tab. 4.1), and the different material interactions and reconstruction efficiencies have been estimated to be the same at the few per-cent level from simulation studies.

The hadron trigger effects on the $B_{(s)}^0 \rightarrow h^+h^-$ BDT distribution have been reduced by the TIS requirement in L0 and HLT1; the remaining trigger bias (mostly from HLT2) and the effects introduced by the trigger lines selecting the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are corrected for. The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT distribution is obtained multiplying the $B_{(s)}^0 \rightarrow h^+h^-$ BDT distribution in Fig. 6.1 by a correction factor:

$$\epsilon^{Trig.Cor.} = \frac{\epsilon^{Trig.}(B_{(s)}^0 \rightarrow \mu^+\mu^-)}{\epsilon_{L0HLT1}^{TIS} \cdot \epsilon_{HLT2}^{Trig.}(B_{(s)}^0 \rightarrow h^+h^-)} \quad (6.10)$$

where the numerator (resp. denominator) is estimated from Monte Carlo $B_s^0 \rightarrow \mu^+ \mu^-$ (resp. $B^0 \rightarrow K^+ \pi^-$) sample. The trigger correction fraction is evaluated and applied separately in every BDT category; the variation is smaller than 1% over the full BDT range.

An additional difference between signal and $B_{(s)}^0 \rightarrow h^+ h^-$ decays is introduced by the muon identification requirement, applied to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ sample to reduce the doubly mis-identified $B_{(s)}^0 \rightarrow h^+ h^-$ background (Tab. 4.1). The muon identification requirement affects the BDT distribution of the signal because its efficiency depends on the phase space. The correction factors per BDT category are measured from $J/\psi \rightarrow \mu^+ \mu^-$ data (see Sec. 2.5.2) and corrected for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ phase space by using a simulated $B_s^0 \rightarrow \mu^+ \mu^-$ sample. The corrections were found to be comparable for 2011 and 2012, and the variation of the average correction over the BDT range is below 1.5% (Fig. 6.2).

The BDT uses the decay time information to identify (short-lived) combinatorial background candidates, and thus the lifetime of the candidates in higher BDT categories is on average longer. In Sec. 1.2, it was shown that the neutral B -meson decay time distribution depends on (i) the lifetime difference between the heavy and light mass eigenstate, (ii) the final state, and (iii) the theoretical model considered. The differences in the decay time distribution between $B^0 \rightarrow h^+ h^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays were studied on simulated samples³. To obtain the $B_s^0 \rightarrow \mu^+ \mu^-$ BDT distribution, the $B_{(s)}^0 \rightarrow h^+ h^-$ BDT distribution is multiplied by the correction factors listed in Tab. 6.2. No corrections are needed for the $B^0 \rightarrow \mu^+ \mu^-$ BDT distribution.

6.3 Signal invariant mass distribution

The invariant mass distribution of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is affected by radiative energy loss and detector resolution effects. It is described by a ‘‘Crystal Ball’’ function, that, to account for both effects, consists of a Gaussian core and a power-law tail. The power-law tail is mostly used at the low side below a certain threshold, but can also be defined at the high-side above a certain threshold, or at both sides (double-sided Crystal Ball).

The Crystal Ball function is defined by four parameters: central value, resolution, and the threshold and power of the power-law tail. The tail para-

³The $B_s^0 \rightarrow h^+ h^-$ modes were not included in the simulated sample, as their contribution to $B_{(s)}^0 \rightarrow h^+ h^-$ is three times smaller than that of B^0 modes, and the effective lifetimes of $B_s^0 \rightarrow K^+ K^-$ and $B_s^0 \rightarrow \pi^+ K^-$ are almost (within 5%) equal to those of the $B^0 \rightarrow h^+ h^-$ modes [109].

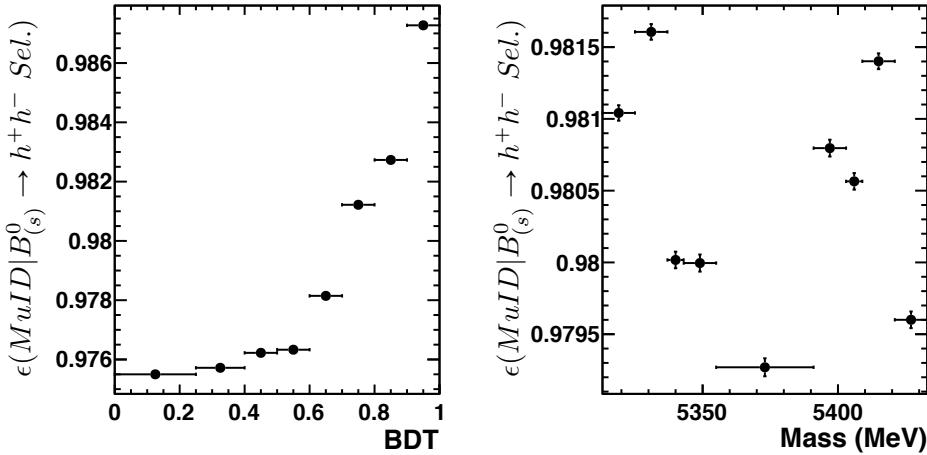


Figure 6.2: The muon identification ($isMUON$ and $\Delta LL(K - \pi) < 10$ and $\Delta LL(\mu - \pi) > -5$) efficiency in the BDT categories (left), and in the di-muon mass bins (right). Determined on a simulated $B_s^0 \rightarrow \mu^+\mu^-$ sample.

meters of the signal di-muon distribution are determined using the Monte Carlo $B_s^0 \rightarrow \mu^+\mu^-$ candidate mass distribution⁴.

6.3.1 B_s^0 and B^0 central mass value

The central values of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ mass distributions are determined by a fit to the $B_{(s)}^0 \rightarrow h^+h^-$ data sample, selected with a higher BDTS cut than that in Tab. 4.1 ($BDTS > 0.1$), in order to reduce the background.

The hadrons are identified as kaons (resp. pions) if $\Delta LL(K - \pi) > 10$ $\&\&$ $\Delta LL(K - p) > 2$ (resp. $\Delta LL(K - \pi) < -10$ $\&\&$ $\Delta LL(\pi - p) > 2$). The sample is split into four groups, according to the final state hypothesis ($\pi^+\pi^-$, π^+K^- , $K^+\pi^-$, or K^+K^-). The particle-identification efficiency is accounted for by per-event weights (see Eq. (6.9)). The weighted sample is then fitted with a likelihood model consisting of two double-sided Crystal Ball functions, describing the B_s^0 and the B^0 decays, an exponential function, describing the combinatorial background, and a ROOPHYSBKG function (see Eq. (6.8)), describing the background from partially reconstructed B decays. The central mass values extracted from the fit (Fig. 6.3) were found stable

⁴The tail parameter values for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ are compatible. Therefore, the same values are used.

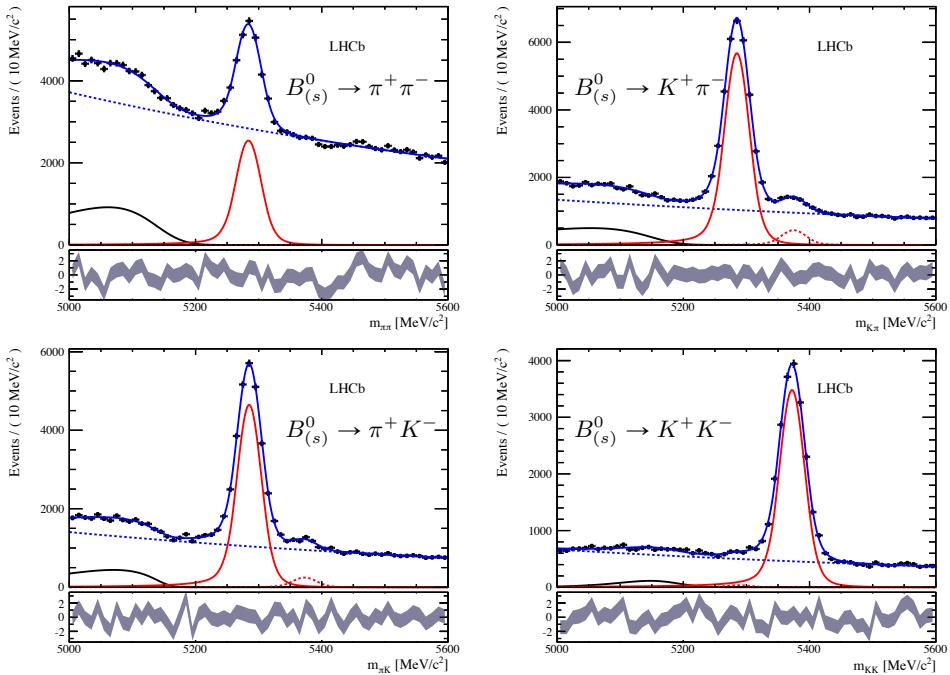


Figure 6.3: The fitted invariant mass distribution of the four $B_{(s)}^0 \rightarrow h^+ h^-$ mass hypotheses (2012 data). These mass distributions are used to determine the central value of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ invariant mass. The solid blue line denotes the full fit result, the solid red line the dominant signal component for the mass hypothesis, the dashed red line the sub-dominant signal component, and the solid black line the partially reconstructed background. The corresponding pull distributions are also shown.

Table 6.2: Corrections due to lifetime effects, applied to the $B_{(s)}^0 \rightarrow h^+h^-$ BDT distribution in order to obtain the $B_s^0 \rightarrow \mu^+\mu^-$ BDT distribution. They are calculated assuming the Standard Model value of $\mathcal{A}_{\Delta\Gamma} = 1$ for $B_s^0 \rightarrow \mu^+\mu^-$, the relative decay-width difference in the B_s^0 system, $(\Gamma_L - \Gamma_H)/(\Gamma_L + \Gamma_H) = 0.0615 \pm 0.0085$, the average B^0 lifetime of $\tau(B^0) = 1.519 \pm 0.007$, and the average B_s^0 lifetime of $\tau(B_s^0) = 1.516 \pm 0.011$ [70].

BDT category	Correction factor $\delta_{PDF}^i - 1 (\%)$
1	-3.106 ± 0.020
2	-1.378 ± 0.029
3	-0.389 ± 0.039
4	$+0.270 \pm 0.042$
5	$+0.719 \pm 0.045$
6	$+1.365 \pm 0.046$
7	$+2.542 \pm 0.046$
8	$+4.737 \pm 0.043$

across 2011 and 2012. The average values for 2011 and 2012 data:

$$M(B^0) = (5284.90 \pm 0.10_{stat} \pm 0.20_{syst}) \text{ MeV}/c^2, \quad (6.11)$$

$$M(B_s^0) = (5371.85 \pm 0.17_{stat} \pm 0.19_{syst}) \text{ MeV}/c^2. \quad (6.12)$$

The systematic uncertainty was evaluated by varying the BDTS and the particle-identification requirements.

6.3.2 B_s^0 and B^0 mass resolution

The invariant mass resolution is determined from data by combining the results of two different methods.

The first method deducts B^0 and B_s^0 mass resolutions from the mass resolutions of lighter and heavier di-muon resonances: $J/\psi(1S) \rightarrow \mu^+\mu^-$, $\psi(2S) \rightarrow \mu^+\mu^-$, $\Upsilon(1S) \rightarrow \mu^+\mu^-$, $\Upsilon(2S) \rightarrow \mu^+\mu^-$, and $\Upsilon(3S) \rightarrow \mu^+\mu^-$. The resolutions are determined by a fit in which each di-muon resonance is described by a double-sided Crystal Ball function, and the combinatorial background is modelled by an exponential function. The dependence of the resolution on the di-muon mass has been studied with a Drell-Yan Monte Carlo sample, and a power-law,

$$\sigma_{\mu^+\mu^-} = a_0 + a_1 \cdot m_{\mu^+\mu^-}^\gamma, \quad (6.13)$$

was found to perform well over a large mass range. The parameters a_0 , a_1 , and γ , are determined from a fit to the resolution values obtained from the fit to the resonances, and then used to estimate the resolutions at the B_s^0 and the B^0 masses (see Fig. 6.4).

The second method uses $B_{(s)}^0 \rightarrow h^+h^-$ decays to extract the B_s^0 and B^0 mass resolutions. The procedure is identical to that described in Sec. 6.2, except that, in the likelihood model, the B_s^0 mass resolution is parametrized as the B^0 mass resolution times a correction factor (obtained from the interpolation method), and that the full $B_{(s)}^0 \rightarrow h^+h^-$ sample is used (not only the L0 and HLT1 TIS candidates). The hadrons are identified as kaons if $\Delta LL(K - \pi) > \kappa$, and pions if $\Delta LL(K - \pi) < -\kappa$. The κ value is increased from 0 to 20 in steps of 0.5, and at every step, the resolution is extracted from a fit to the mass distribution. The threshold dependency is introduced by the $K - \pi$ mis-identification, and is expected to decrease with better identification (higher κ threshold). The true B^0 mass resolution is extracted by fitting the threshold dependency (see Fig. 6.5).

The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass resolutions are calculated by combining the results from the interpolation method and the $B_{(s)}^0 \rightarrow h^+h^-$ fit; as the 2011 and 2012 results are comparable, they are averaged. We obtain:

$$\sigma(B^0) = (22.83 \pm 0.07_{stat} \pm 0.42_{syst}) \text{ MeV}/c^2, \quad (6.14)$$

$$\sigma(B_s^0) = (23.24 \pm 0.08_{stat} \pm 0.44_{syst}) \text{ MeV}/c^2. \quad (6.15)$$

The systematic uncertainty of the interpolation method stems from the size of the mass window and from the mass fit model; the systematic uncertainty of the second method stems from the $\Delta LL(K - \pi)$ efficiency corrections and the $B_{(s)}^0 \rightarrow h^+h^-$ mass fit model.

6.4 Doubly mis-identified background model

The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ sample consists of $B_{(s)}^0 \rightarrow h^+h^-$ candidates that also pass the muon identification criteria: ISMUON and $\Delta LL(K - \pi) < 10$ and $\Delta LL(\mu - \pi) > -5$. This, however, also includes a small fraction of doubly mis-identified K^+K^- , $K^+\pi^-$, π^+K^- , and $\pi^+\pi^-$ final states from B^0 or B_s^0 decays. This doubly mis-identified $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-$ background has a peaking shape in the di-muon invariant mass distribution overlapping with the $B^0 \rightarrow \mu^+\mu^-$ invariant mass distribution, and therefore, its impact needs to be well understood and modelled.

The total yield of the doubly mis-identified background is determined by applying the hadron mis-identification efficiencies to the $B_{(s)}^0 \rightarrow h^+h^-$ data

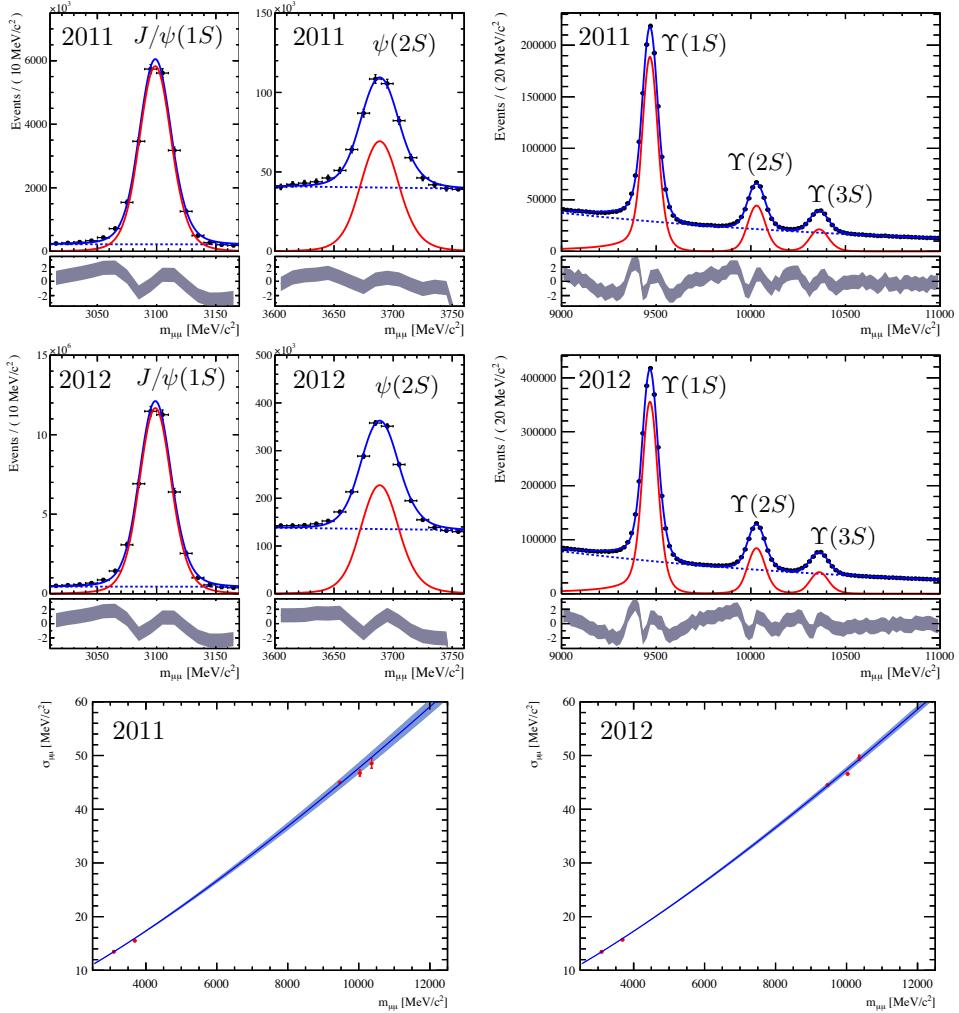


Figure 6.4: The invariant mass distribution of $J/\psi(1S) \rightarrow \mu^+ \mu^-$, $\psi(2S) \rightarrow \mu^+ \mu^-$, $\Upsilon(1S) \rightarrow \mu^+ \mu^-$, $\Upsilon(2S) \rightarrow \mu^+ \mu^-$, and $\Upsilon(3S) \rightarrow \mu^+ \mu^-$, shown for 2011 (top row) and 2012 (middle row) data. The lower plots illustrate the fit of Eq. (6.13) to the mass resolutions determined from the di-muon resonances.

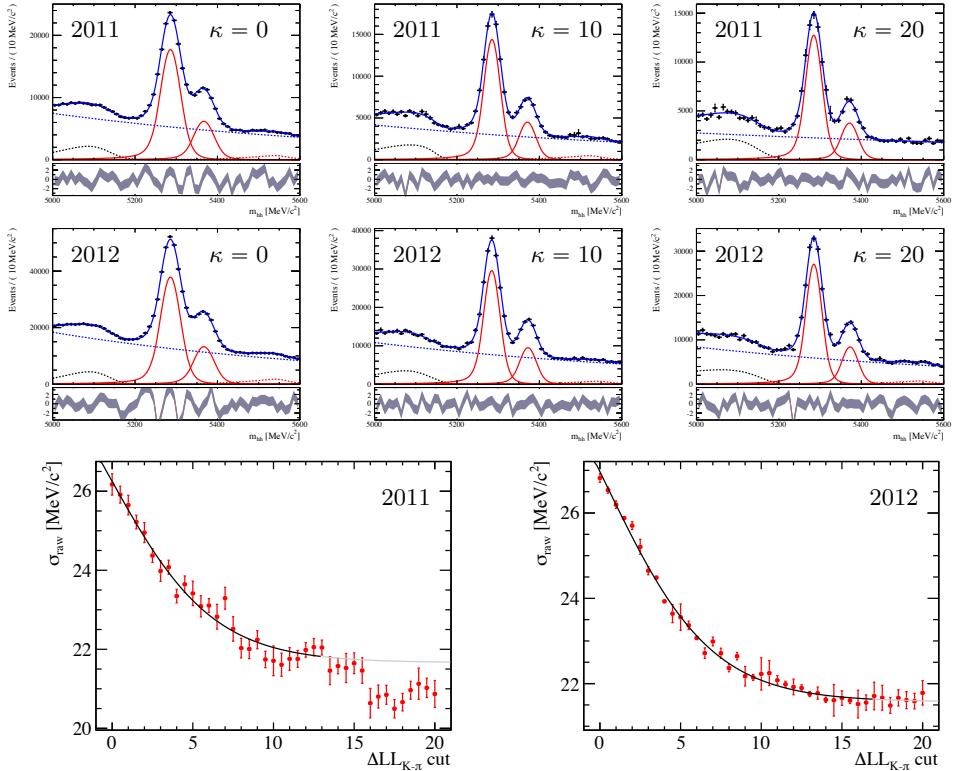


Figure 6.5: The invariant mass distributions of $B_{(s)}^0 \rightarrow h^+h^-$ candidates, shown for three $\Delta LL(K - \pi)$ cut thresholds (κ). The hadrons are identified as kaons if $\Delta LL(K - \pi) > \kappa$, and pions if $\Delta LL(K - \pi) < -\kappa$, and rejected otherwise; the cut efficiency is corrected for by applying per-event weights (see Eq. (6.9)). The two plots at the *bottom* show the width of the B^0 mass distribution as a function of $\Delta LL(K - \pi)$ for 2011 (bottom left) and 2012 (bottom right) data. The solid black line on the bottom plots denotes the full fit result, the dashed grey line denotes its extrapolation into the region excluded from the fit (too few hadrons in the higher momentum range).

sample. To minimise the hadronic trigger bias, only the candidates triggered independently of the $B_{(s)}^0 \rightarrow h^+h^-$ candidate are considered. As already mentioned in Sec. 6.2, in practice the low statistics of the TIS sample obliges us to use a $B_{(s)}^0 \rightarrow h^+h^-$ sample which is TIS for L0 and HLT1, but not HLT2.

The total yield of the doubly mis-identified background is obtained multiplying the $B_{(s)}^0 \rightarrow h^+h^-$ yield (corrected for the trigger efficiency, as in Eq. (6.10)) by the double mis-identification efficiency, $\epsilon_{hh \rightarrow \mu\mu}$:

$$N_{hh \rightarrow \mu\mu} = \epsilon_{hh \rightarrow \mu\mu} \cdot \frac{\epsilon^{Trig.}(B_{(s)}^0 \rightarrow \mu^+\mu^-)}{\epsilon_{L0HLT1}^{TIS} \cdot \epsilon_{HLT2}^{Trig.}(B_{(s)}^0 \rightarrow h^+h^-)} \cdot N_{hh}^{TIS} \quad (6.16)$$

Notice that the mis-identified hadrons have mostly decayed to muons, and are therefore triggered by the same trigger lines as $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays (see Ch. 3).

The quantities used in the yield calculation are given in Tab. 6.3. The double mis-identification efficiency, $\epsilon_{hh \rightarrow \mu\mu}$, has been calculated from the kaon and pion mis-identification efficiencies, as discussed in Sec. 2.5.2. As these mis-identification efficiencies are given in bins of p and p_T , the hadron momentum and transverse momentum distributions for each of the four final state hypotheses are taken from Monte Carlo samples respective mis-identification efficiencies for each final state. These efficiencies are then weighted according to the relative K^+K^- , $K^+\pi^-$, π^+K^- , and $\pi^+\pi^-$ production rates (Tab. 6.1).

The fraction of mis-identified $B_{(s)}^0 \rightarrow h^+h^-$ decays in every BDT category is found by evaluating the inclusive $\epsilon_{h+h-\rightarrow\mu^+\mu^-}$ for each category separately.

Table 6.3: Quantities entering the doubly mis-identified $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ peaking background yield calculation.

Parameter	2012 data	2011 data
N_{hh}^{TIS}	49653 ± 507	20143 ± 572
$\epsilon_{MC}^{Trig.}(B_s^0 \rightarrow \mu^+\mu^-)$	$(92.4 \pm 0.3_{\text{stat}} \pm 1.9_{\text{syst}})\%$	$(92.1 \pm 0.5_{\text{stat}} \pm 1.6_{\text{syst}})\%$
$\epsilon_{L0HLT1}^{TIS}(B^+ \rightarrow J/\psi K^+)$	$(5.92 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}}) \%$	$(5.05 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}}) \%$
$\epsilon_{HLT2,MC}^{Trig.}(B^0 \rightarrow K^+\pi^-)$	$(91.6 \pm 0.2)\%$	$(91.5 \pm 0.3)\%$
$\epsilon_{hh \rightarrow \mu\mu}$	$(0.12 \pm 0.01) \times 10^{-4}$	$(0.11 \pm 0.01) \times 10^{-4}$
$N_{hh \rightarrow \mu\mu}$	Combined: 14.6 ± 1.3	

6.4.1 Mass distribution

The doubly mis-identified hadrons from a $B_{(s)}^0 \rightarrow h^+h^-$ decay leads to wrong daughter-mass assumptions, and because muons are lighter than hadrons, to lower invariant di-muon masses. The di-muon mass distribution for doubly mis-identified $B_{(s)}^0 \rightarrow h^+h^-$ decays is determined from the $B^0 \rightarrow K^+\pi^-$ Monte Carlo sample. The shape of the distribution does not significantly depend on the BDT, and the same shape is thus used for all BDT categories (Fig. 6.6).

The procedure to determine the invariant mass distribution of doubly mis-identified $B_{(s)}^0 \rightarrow h^+h^-$ decays is complicated, because the simulated $B^0 \rightarrow K^+\pi^-$ sample does not contain enough doubly mis-identified candidates. Instead, a work-around is used to emulate the effects of the double-misidentification on the $B^0 \rightarrow K^+\pi^-$ sample.

The wrong mass hypothesis is emulated by reconstructing the B mass of the $B^0 \rightarrow K^+\pi^-$ candidates with di-muon hypothesis. Moreover, the hadron decays into muons and neutrinos before the calorimeters affect the measurement of the hadron momentum. This effect is emulated by using a momentum smearing determined separately for kaon and pion decays⁵. The effect of each step on the $B^0 \rightarrow K^+\pi^-$ mass distribution can be seen in Fig. 6.6.

6.5 Other background sources

Single muon mis-identification could also fake the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal. The singly mis-identified backgrounds have been studied with the inclusive $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo sample determining their true identity from the Monte Carlo truth information. The results listed in Tab. 6.4 show that the contribution from decays where the muons originate from the same b quark ($b \rightarrow \mu\mu X$), is 0.4% for events with di-muon mass above 4700 MeV/ c^2 , and amounts to 0.1% in the fitted mass range [4900, 6000] MeV/ c^2 .

The $b \rightarrow \mu\mu X$ type contribution stems from $B_c^+ \rightarrow J/\psi(\mu^-\mu^+)X$ decays. This component has been separately studied using a dedicated $B_c^+ \rightarrow J/\psi(\mu^+\mu^-)\mu^+\nu_\mu$ Monte Carlo sample, and found to be identified by the BDT into background dominated categories: the total expected yield in BDT > 0.8 region is 0.5 events. Furthermore, the $B_c^+ \rightarrow J/\psi(\mu^+\mu^-)\mu^+\nu_\mu$ mass distribution is well described by an exponential function. The same considerations

⁵The relative fraction of the momentum smearing, $(p_{rec} - p_{true})/p_{true}$, is determined for the $B^0 \rightarrow K^+\pi^-$ candidates, in which either a kaon or a pion has decayed. The distribution of the smearing fraction is modelled with two Gaussian functions.

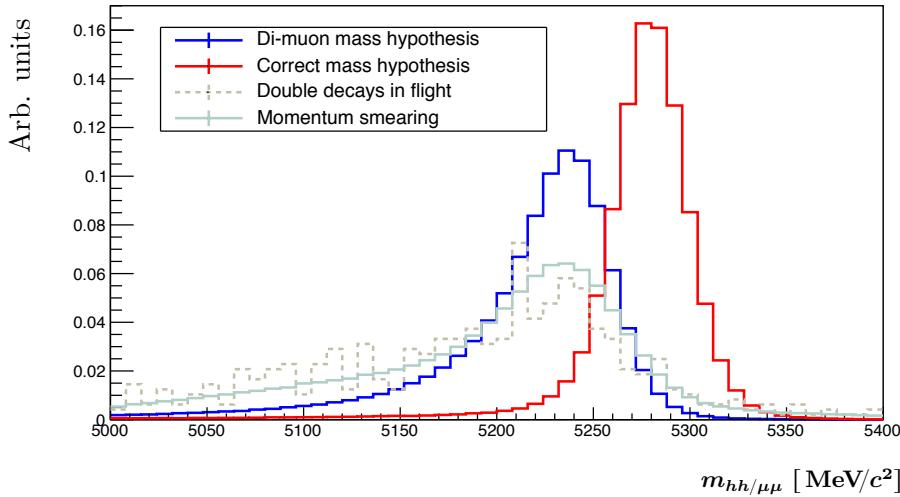


Figure 6.6: Mass distribution of Monte Carlo $B^0 \rightarrow K^+ \pi^-$ candidates, in the case where the hadrons have been reconstructed with the correct daughter hypothesis (red line), with the di-muon hypothesis (blue line), with the di-muon hypothesis and momentum smearing (light blue line), and with the di-muon mass hypothesis only for the candidates with double decays in flight (grey dashed line).

Table 6.4: Breakdown of the singly mis-identified backgrounds in the $b\bar{b} \rightarrow \mu\bar{\mu}X$ Monte Carlo sample for two ranges of the di-muon invariant mass.

Background source	$m_{\mu\mu} > 4700 \text{ MeV}/c^2$	$m_{\mu\mu} = [4900, 6000] \text{ MeV}/c^2$
$b\bar{b} \rightarrow \mu\bar{\mu}X$	103277	82414
\downarrow		
$\bar{b} \rightarrow \mu^+ \mu^- X$	427	74
\downarrow		
$B_c^+ \rightarrow J/\Psi(\mu^- X)\mu^+ X$	149	71
$B_s^0 \rightarrow D_s^-(\mu^- X)\mu^+ X$	139	3
$B^+ \rightarrow \mu^+ \mu^- X$	52	0
$B^0 \rightarrow \mu^+ \mu^- X$	43	0
$B_s^0 \rightarrow D_s^{-*}(D_s^-(\mu^- Y)\gamma/\pi^0)\mu^+ X$	28	0
$B^0 \rightarrow D^-(\mu^- X)\mu^+ X$	13	0
$B^+ \rightarrow \bar{D}^0(\mu^- Y)\mu^+ X$	1	0
$B^0 \rightarrow \pi^- \mu^+ \mu^- X$	1	0
$\Upsilon(1S) \rightarrow \tau^+(\mu^+ Y)\tau^-(\mu^- X)$	1	0

apply to the $B_s^0 \rightarrow D_s^- (\mu^- Y) \mu^+ X$ background.

The semi-leptonic B and Λ_b^0 decays ($B^0 \rightarrow \pi^- \mu^+ \nu$, $B_s^0 \rightarrow K^- \mu^+ \nu$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$, and $\Lambda_b^0 \rightarrow p \mu^- \nu$) do not contribute significantly to the background as compared to the main background sources listed in Tab. 6.4. The mis-identified hadron, however, disguises the true nature of the background candidates, and ‘‘confuses’’ the BDT; if mis-identified, these decays also contribute to the higher BDT categories, and distort the exponential behaviour of the combinatorial background distribution. Their direct impact on the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ invariant mass distributions is instead small, as their di-muon mass is shifted towards lower values.

For each semi-leptonic background channel, the selection efficiency, ϵ_{bkg} , is determined from the corresponding Monte Carlo sample (Tab. 6.5). The candidates are asked to pass the $B_{(s)}^0 \rightarrow h^+ h^-$ selection (Tab. 4.1); the muon identification (isMUON and $\Delta LL(K - \pi) < 10$ and $\Delta LL(\mu - \pi) > -5$) is required from all the final-state muons, the hadron mis-identification efficiencies are applied in bins of momentum and transverse momentum of the final-state hadron (see Sec. 2.5.2). The semi-leptonic background events that pass the selection are separated from the small mis-reconstructed component according to their true Monte Carlo particle identity.

The total number of mis-identified semi-leptonic background candidates that contaminate the 2011 and 2012 $B_s^0 \rightarrow \mu^+ \mu^-$ sample is estimated by normalising to the measured $B^+ \rightarrow J/\psi K^+$ candidate yield:

$$N_{bkg} = \underbrace{\frac{N_{B^+ \rightarrow J/\psi K^+}}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \times \frac{f_{bkg}}{f_u} \times \frac{1}{\epsilon_{B^+ \rightarrow J/\psi K^+}}}_{\beta_{bkg}} \times \mathcal{B}(bkg) \times \epsilon_{bkg}, \quad (6.17)$$

where the subscript ‘‘ bkg ’’ refers to a specific background decay, and the normalisation factors are:

$$\beta_d = (8.19 \pm 0.33) \times 10^{11}, \quad (6.18)$$

$$\beta_s = (2.12 \pm 0.16) \times 10^{11}, \quad (6.19)$$

$$\beta_\Lambda = \beta_d \cdot 2r_\Lambda, \quad (6.20)$$

$$r_\Lambda \equiv \frac{f_\Lambda}{f_u + f_d} = (0.404 \pm 0.110) \times [1 - (0.031 \pm 0.004) \times p_T(\text{GeV}/c)]. \quad (6.21)$$

The hadronisation ratio, r_Λ , is measured by LHCb in Ref. [82]. The $B^+ \rightarrow J/\psi K^+$ yield and efficiencies are given in Sec. 7.2.1 and Sec. 7.3, respectively.

Table 6.5: Expected mis-identified semi-leptonic background yields in the selected $B_s^0 \rightarrow \mu^+ \mu^-$ sample, and in the di-muon mass range of $m_{\mu\mu} = [4900, 6000]\text{MeV}/c^2$. The efficiency for $\Lambda_b^0 \rightarrow p\mu^-\nu$ includes the $2r_\Lambda$ factor (see Eq. (6.20).)

Parameter	$B^0 \rightarrow \pi^- \mu^+ \nu$	$B_s^0 \rightarrow K^- \mu^+ \nu$	$B^{0/+} \rightarrow \pi^{0/+} \mu^+ \mu^-$	$\Lambda_b^0 \rightarrow p\mu^-\nu$
β_{bkg}	$(8.19 \pm 0.33) \times 10^{11}$	$(2.12 \pm 0.16) \times 10^{11}$	$(8.19 \pm 0.33) \times 10^{11}$	$(8.19 \pm 0.33) \times 10^{11}$
\mathcal{B}	$(1.44 \pm 0.05) \times 10^{-4}$	$(1.27 \pm 0.49) \times 10^{-4}$	$(3.4 \pm 0.8) \times 10^{-8}$	$(4.75 \pm 2.11) \times 10^{-4}$
ϵ_{bkg}	$(9.69 \pm 0.14) \times 10^{-7}$	$(3.80 \pm 0.09) \times 10^{-7}$	$(1.011 \pm 0.016) \times 10^{-3}$	$(1.74 \pm 0.07) \times 10^{-7}$
<hr/>				
N_{bkg}	114.4 ± 6.3	10.2 ± 4.0	$28.01^{+8.51}_{-8.16}$	67.6 ± 30.3
\hookleftarrow				
N^{BDT}_1	42.9 ± 2.4	3.2 ± 1.3	$10.7^{+3.2}_{-3.1}$	31.0 ± 13.9
N^{BDT}_2	21.6 ± 1.2	1.7 ± 0.7	$5.2^{+1.6}_{-1.5}$	12.7 ± 5.7
N^{BDT}_3	12.4 ± 0.7	1.1 ± 0.4	$3.0^{+0.9}_{-0.9}$	6.8 ± 3.0
N^{BDT}_4	10.7 ± 0.6	1.0 ± 0.4	$3.0^{+0.9}_{-0.9}$	5.4 ± 2.4
N^{BDT}_5	9.5 ± 0.5	1.0 ± 0.4	$2.1^{+0.7}_{-0.6}$	4.5 ± 2.0
N^{BDT}_6	8.2 ± 0.5	1.0 ± 0.4	$2.1^{+0.6}_{-0.6}$	3.7 ± 1.6
N^{BDT}_7	6.0 ± 0.3	0.7 ± 0.3	$1.4^{+0.4}_{-0.4}$	2.4 ± 1.1
N^{BDT}_8	3.1 ± 0.2	0.4 ± 0.2	$0.8^{+0.2}_{-0.2}$	1.2 ± 0.5

6.5.1 BDT and mass distribution

The invariant di-muon mass distribution of mis-identified semi-leptonic backgrounds is described by the RoOPHYSBKG bkg function, originally developed to describe partially reconstructed backgrounds candidates (Eq. (6.8)). The parameters of the RoOPHYSBKG functions describing the $B^0 \rightarrow \pi^- \mu^+ \nu$, $B_s^0 \rightarrow K^- \mu^+ \nu$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$, and $\Lambda_b^0 \rightarrow p \mu^- \nu$ di-muon mass distributions, are extracted in every BDT category from the corresponding Monte Carlo samples [110].

The candidates in the Monte Carlo samples are asked to pass the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection (Tab. 4.1). However, asking both muon candidates to pass the muon identification would leave too few candidates for the mass distribution fit. Therefore, the muon identification is only applied on the true final-state muons. The selected candidates are then weighted with the mis-identification efficiency (see Sec. 2.5.2) to emulate the effect of the hadron mis-identification on the mass distribution without biasing the BDT distribution.

The fraction of candidates in each BDT category is obtained from the fit in every category (requiring the total to add up to one). The mis-identified semi-leptonic background yields are listed in Tab. 6.5.

6.6 Combinatorial background BDT and mass distributions

The combinatorial background in Fig. 4.2c dominates the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ sample. In most cases, the muons originate from two different B mesons, but, as shown in Tab. 6.6, they can also originate from Λ_b^0 baryons, or heavier hadrons. Once the background components with peaking mass distributions are accounted for, the di-muon mass *pdf* for the remaining background is well described by an exponential function:

$$pdf_{Comb}(m) = Ae^{-k \cdot m}, \quad (6.22)$$

where A is the normalisation constant and k the slope parameter. The slope parameter and the combinatorial yields in the BDT categories, $f_{Comb}^{cat} \cdot N_{Comb}$ in Eq. (6.7), are determined by a Maximum Likelihood fit to the di-muon distribution. The fraction of candidates in each BDT category and the slope parameter are determined independently in every BDT category, except in the last two, where the same slope parameter was used because of the low statistics.

Table 6.6: The main decay channels contributing to the combinatorial background in the selected $B_s^0 \rightarrow \mu^+ \mu^-$ sample, shown for the mass range $m_{\mu\mu} = [4900, 6000] \text{ MeV}/c^2$. The fractions are determined from the $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo sample by using the truth information.

Channels	Contribution to combinatorial background
$B^0 \rightarrow \mu^+ X + B^- \rightarrow \mu^- Y$	$(26.0 \pm 0.2)\%$
$B^+ \rightarrow \mu^+ X + B^- \rightarrow \mu^- Y$	$(16.2 \pm 0.1)\%$
$B^0 \rightarrow \mu^+ X + \bar{B}^0 \rightarrow \mu^- Y$	$(11.3 \pm 0.1)\%$
$B^+ \rightarrow \mu^+ X + \Lambda_b^0 \rightarrow \mu^- Y$	$(6.2 \pm 0.1)\%$
$B^0 \rightarrow \mu^+ X + \Lambda_b^0 \rightarrow \mu^- Y$	$(5.2 \pm 0.1)\%$
$B_s^0 \rightarrow \mu^+ X + B^- \rightarrow \mu^- Y$	$(4.0 \pm 0.1)\%$
$B_s^0 \rightarrow \mu^+ X + \bar{B}^0 \rightarrow \mu^- Y$	$(3.8 \pm 0.1)\%$
other	27.3%

Chapter 7

Signal branching fraction normalisation

A reaction rate of a sub-atomic process can be calculated from the instantaneous luminosity (\mathcal{L}), the physical quantity that relates the reaction cross-section (σ) to the reaction rate (R): $R = \mathcal{L} \cdot \sigma$ (see Sec. 2.1). The number of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays measured in LHCb, can be expressed as

$$N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-} = \underbrace{L^{LHCb} \times \sigma_{pp \rightarrow b\bar{b}} \times 2 \times f_{d(s)} \times \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \times \epsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}}_{\text{Produced } N(b) + N(\bar{b})} \underbrace{\text{Prob}(\bar{b} \rightarrow B_{(s)}^0 \rightarrow \mu^+ \mu^- \rightarrow \text{Detected})}_{(7.1)}$$

where we have used:

- the total integrated luminosity collected by the LHCb (L^{LHCb});
- the $b\bar{b}$ production cross-section in pp collisions ($\sigma_{pp \rightarrow b\bar{b}}$);
- the probability of a b quark to form a B meson, i.e. the hadronisation fraction f_q (where $q = u, d, s, c, \Lambda_q$);
- the branching fraction of the decay channel (\mathcal{B});
- the total detection efficiency of the signal decay ($\epsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$).

The measurement of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidate yields, $N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$, has been described in Ch. 6. This chapter describes the calculation of the branching fractions $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ from the number of measured signal candidates.

The branching fraction can be calculated using Eq. (7.1); however, the large uncertainties in L^{LHCb} and $\sigma_{pp \rightarrow b\bar{b}}$ limit the precision. A better alternative is to consider another B meson decay channel with a well known

branching fraction, measure its yield over the same period, and normalise the signal branching fraction with respect to this channel:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{\mathcal{B}_{norm}}{N_{norm}} \times \frac{\epsilon_{norm}}{\epsilon_{sig}} \times \frac{f_{norm}}{f_{s.d.}} \times N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}. \quad (7.2)$$

Compared to the direct branching fraction extraction from Eq. (7.1), the relative uncertainty in the signal branching fraction from Eq. (7.2) is around three times smaller because many sources of uncertainty, such as $\sigma_{pp \rightarrow b\bar{b}}$, L^{LHCb} , and the detection efficiencies common to signal and normalisation channels, are avoided. This motivates the choice of a normalisation channel “as similar as possible” to the signal decays, with common trigger, reconstruction, and selection procedures.

Section 7.1 describes the choice of the normalisation channels and introduces the normalisation factors used to calculate the signal branching fractions. The various components of the normalisation factors are then described: the normalisation-channel candidate yields in Sec. 7.2; detection-efficiency differences between the signal and the normalisation channels in Sec. 7.3; the differences arising from the different hadronisation probabilities in Sec. 7.4. The normalisation factors are calculated in Sec. 7.5.

7.1 Normalisation channels

An ideal normalisation channel is a channel identical to the signal decay in terms of production and detection probabilities, and with a relatively large and well known branching fraction. The best compromise between the requirements is found by using two normalisation channels instead of one: $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$.

The $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$ channel was chosen because the trigger criteria on the two muons from the J/ψ are almost identical to those on the muons from B_s^0 or B^0 mesons, and the branching fraction of the decay is well known: $\mathcal{B}(B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+) = (6.025 \pm 0.205) \times 10^{-5}$, using $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.93 \pm 0.06) \times 10^{-2}$ [106]. Minor differences arise in the detection and selection efficiencies, because the additional track in the $K^+ \mu^+ \mu^-$ final state lowers the overall tracking efficiency and alters the kinematic distribution of the muons.

The $B^0 \rightarrow K^+ \pi^-$ channel was chosen because, unlike $B^+ \rightarrow J/\psi K^+$, it is a two-body B -meson decay and the kinematic distribution of the two hadrons in the final state resemble that of the two muons from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. On the other hand, the trigger criteria for kaons and pions are very different from those for muons, and these differences need to be accounted for in the

normalisation. To minimise the difference arising from the hadronic trigger, only candidates triggered independently of the $B^0 \rightarrow K^+\pi^-$ candidate are considered. The branching fraction of $B^0 \rightarrow K^+\pi^-$ is well measured: $\mathcal{B}(B^0 \rightarrow K^+\pi^-) = (1.94 \pm 0.06) \times 10^{-5}$ [106].

A common shortcoming of both $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ comes from the different B meson flavour: the b (or \bar{b}) quarks hadronise into B^+ and B^0 mesons about four times more often than into B_s^0 mesons. This difference has to be taken into account in the normalisation, and is at present the largest source of systematic uncertainty in the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ measurement. Notice that henceforth the hadronisation fractions to B^+ and B^0 (f_u and f_d) are taken to be equal. This will be discussed in Sec. 7.4. Normalising with respect to B_s^0 decays is not an option because B_s^0 decay branching fractions are not known with sufficient precision.

The normalisation factors used to normalise $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$, α_s and α_d , are defined through the following formulas:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = \underbrace{\frac{\mathcal{B}(\text{norm})}{N_{\text{norm}}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \times \frac{f_d}{f_s}}_{\alpha_d} \times N_{B_s^0 \rightarrow \mu^+\mu^-}, \quad (7.3)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = \alpha_d \times N_{B^0 \rightarrow \mu^+\mu^-}.$$

These factors are calculated separately for each normalisation channel and collision energy, and finally combined (Sec. 7.5).

7.2 Normalisation channel yields

$$\boxed{\alpha_s = \underbrace{\frac{\mathcal{B}(\text{norm})}{N_{\text{norm}}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \times \frac{f_d}{f_s}}_{\alpha_d}}$$

The normalisation-channel yields (N_{norm}) are measured on the data sample that is collected in the same period as the signal sample. For each channel, the distribution of the reconstructed B meson mass is modelled with a probability density function (pdf) that takes into account detector resolution effects as well as different background sources. The B -meson mass is reconstructed, using the momenta and masses of the final state particles, for every triggered candidate that passes the selection criteria. The likelihood of measuring an invariant mass distribution with certain pdf parameters is calculated with the `RooFit` [104] software package; the candidate yield is one of the model parameters, and is extracted by maximising the total likelihood with `MINUIT` [105].

7.2.1 $B^+ \rightarrow J/\psi K^+$ mass model

The $B^+ \rightarrow J/\psi K^+$ invariant mass is calculated from the masses and momenta of the final state muons and kaon. A constraint is imposed on the J/ψ mass ($|m_{\mu^+\mu^-} - m_{J/\psi}| < 60 \text{ MeV}/c^2$, with $m_{J/\psi} = 3096.916 \text{ MeV}/c^2$) to be certain that the muons come from a J/ψ , and to reduce background from random two-muon combinations. The $B^+ \rightarrow J/\psi K^+$ mass *pdf* is taken to be a double-sided Crystal Ball function (see Sec. 6.3). The background model consists of two parts: first, an exponential function to describe the background arising from random combinations of two muons and a kaon; second, a Crystal Ball function to describe $B^+ \rightarrow J/\psi \pi^+$ decays with a pion mis-identified as a kaon (see Sec. 2.5).

The $B^+ \rightarrow J/\psi K^+$ yield is one of the fit model parameters, and is obtained from a Maximum Likelihood fit in a mass window of $\pm 100 \text{ MeV}/c^2$ around the B^+ meson mass $5279.17 \text{ MeV}/c^2$ [103]. The $B^+ \rightarrow J/\psi K^+$ mass distributions and the corresponding fits are shown in Fig. 7.1. The $B^+ \rightarrow J/\psi K^+$ candidate yields extracted from the fit are:

$$\begin{aligned} N_{B^+ \rightarrow J/\psi K^+}^{2011} &= 355232 \pm 608_{\text{stat}} \pm 1066_{\text{syst}}, \\ N_{B^+ \rightarrow J/\psi K^+}^{2012} &= 761122 \pm 891_{\text{stat}} \pm 2283_{\text{syst}}, \end{aligned} \quad (7.4)$$

where the relative systematic uncertainty of 0.3% arises from the treatment of the background in the likelihood model.

7.2.2 $B^0 \rightarrow K^+ \pi^-$ mass model

Candidates of the second normalisation channel, $B^0 \rightarrow K^+ \pi^-$, are selected from the trigger independent (TIS) $B_{(s)}^0 \rightarrow h^+ h^-$ calibration sample, where h denotes a kaon or a pion. The same $B_{(s)}^0 \rightarrow h^+ h^-$ sample was used to calibrate the signal BDT in Sec. 6.2. The $B^0 \rightarrow K^+ \pi^-$ sub-sample is separated by using the particle identification information (ΔLL), and the ΔLL selection efficiency is corrected for with efficiency weights, defined in Eq. (6.9).

The Maximum Likelihood mass fit to the $B_{(s)}^0 \rightarrow h^+ h^-$ invariant mass is carried out in a mass window of $\pm 300 \text{ MeV}/c^2$ around the B_s^0 meson mass $5366.3 \text{ MeV}/c^2$. Besides the $B^0 \rightarrow K^+ \pi^-$ candidates, the invariant mass distribution in Fig. 7.2 also includes a contamination from $B_s^0 \rightarrow \pi^+ K^-$ decays; both decays are modelled with a double-sided Crystal Ball function (see Sec. 6.3). The background consists of three sources: combinatorial background (modelled with an exponential distribution), a physical background arising from partially reconstructed B meson decays (modelled with a

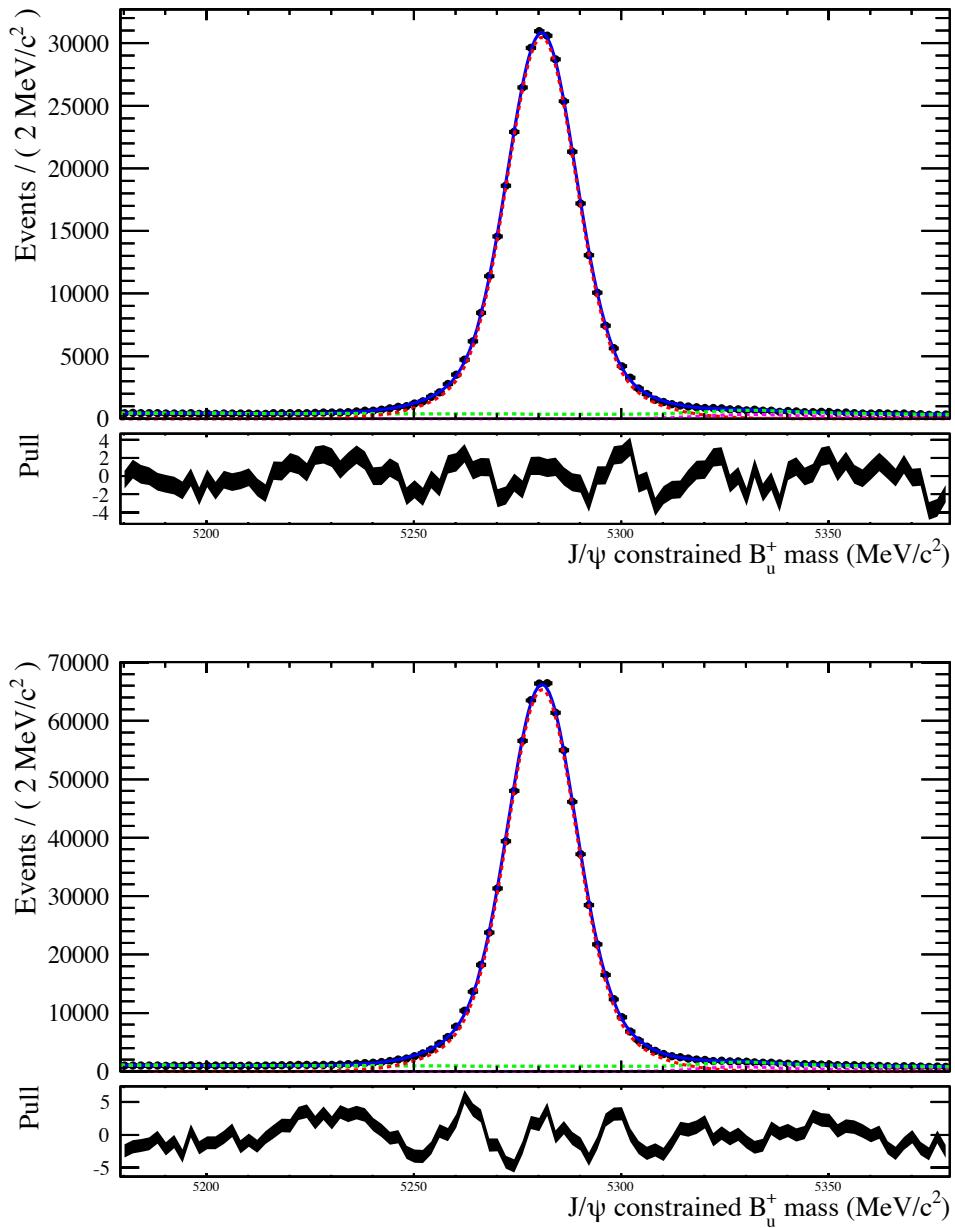


Figure 7.1: $B^+ \rightarrow J/\psi K^+$ invariant mass fit on 1 fb^{-1} of data collected in 2011 (top), and 2 fb^{-1} collected in 2012 (bottom). The continuous black curve denotes the fitted likelihood model, the dashed red curve the signal contribution, and the dashed green curve the background contribution. The 2011 sample (top) contains 355232 ± 1227 signal candidates and 2012 sample (bottom) 761122 ± 2451 .

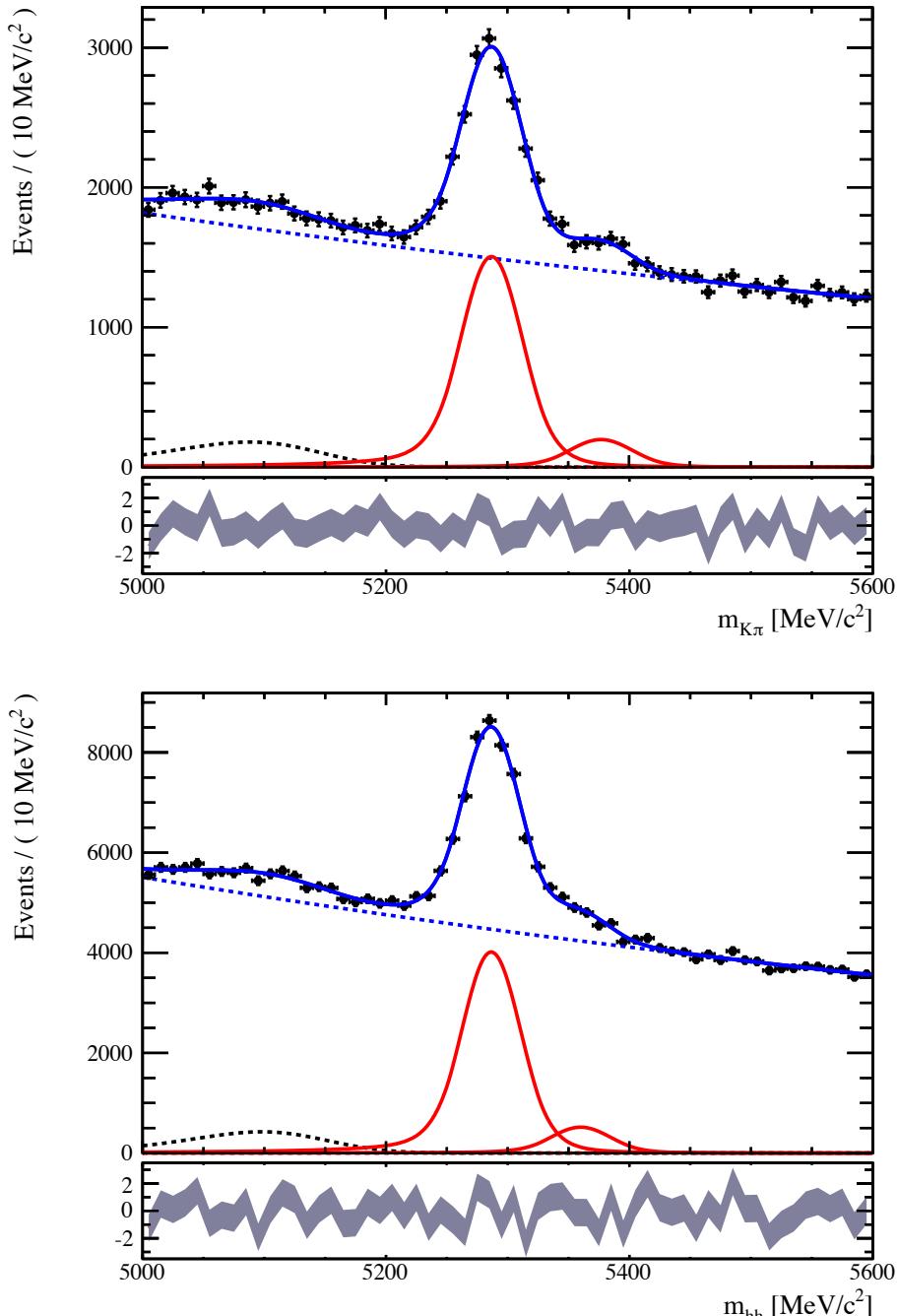


Figure 7.2: $B^0 \rightarrow K^+\pi^-$ invariant mass fit on 1 fb^{-1} of data collected in 2011 (top), and 2 fb^{-1} collected in 2012 (bottom). The continuous blue curve denotes the fitted likelihood model, the solid red curve the contribution from $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \pi^+K^-$, the dashed black curve from the partially-reconstructed background, and the dashed blue curve from the combinatorial background. The 2011 sample (top) contains 10809 ± 439 signal candidates and 2012 sample (bottom) 26749 ± 447 signal candidates.

RooPhysBkg function, given in Eq. (6.8)), and a minor contribution from mis-identified $\Lambda_b^0 \rightarrow ph^-$ decays, with $h = K, \pi$, and a proton mis-identified as a kaon or pion (modelled with a double sided Crystal Ball function, parameters of which are determined from a simulated sample).

The $B^0 \rightarrow K^+\pi^-$ mass distribution with the corresponding fits are shown in Fig. 7.2. The $B^0 \rightarrow K^+\pi^-$ candidate yields extracted from the fits are:

$$\begin{aligned} N_{B^0 \rightarrow K^+\pi^-}^{2011} &= 10809 \pm 439_{\text{stat+syst}}, \\ N_{B^0 \rightarrow K^+\pi^-}^{2012} &= 26749 \pm 447_{\text{stat+syst}}, \end{aligned} \quad (7.5)$$

where the systematic uncertainty stems mainly from the ΔLL cut efficiency correction¹.

7.3 Detection efficiencies

$$\alpha_s = \underbrace{\frac{\mathcal{B}(\text{norm})}{N_{\text{norm}}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \times \frac{f_d}{f_s}}_{\alpha_d}$$

The detection efficiencies enter the normalisation factors only as efficiency ratios ($\epsilon_{\text{norm}}/\epsilon_{\text{sig}}$), and thus common factors cancel out. A crucial step in calculating the normalisation factors is to evaluate the detection efficiencies for the signal and normalisation channels, and to assess possible differences. The detection efficiencies are split into three contributions, corresponding to the three stages of the detection: acceptance, reconstruction and selection, and trigger, such that

$$\frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} = \frac{\epsilon_{\text{norm}}^{\text{Acc}}}{\epsilon_{\text{sig}}^{\text{Acc}}} \times \frac{\epsilon_{\text{norm}}^{\text{RecSel|Acc}}}{\epsilon_{\text{sig}}^{\text{RecSel|Acc}}} \times \frac{\epsilon_{\text{norm}}^{\text{Trig|RecSel}}}{\epsilon_{\text{sig}}^{\text{Trig|RecSel}}}, \quad (7.6)$$

where the efficiency for each subsequent stage is estimated for events that pass the previous stages.

The $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ detection efficiencies were studied on simulated samples and found to be equal within 1% for all the three terms in Eq. (7.6). Therefore, the $B_s^0 \rightarrow \mu^+\mu^-$ detection efficiencies, determined from the simulated sample will also be used for $B^0 \rightarrow \mu^+\mu^-$, and referred to as the *signal* efficiency in the subsequent sections.

¹The relative uncertainty in 2012 is lower because of a more precise determination of the ΔLL cut efficiency on a larger control sample.

7.3.1 Geometrical detector acceptance

$$\frac{\epsilon_{norm}}{\epsilon_{sig}} = \frac{\epsilon_{norm}^{Acc}}{\epsilon_{sig}^{Acc}} \times \frac{\epsilon_{norm}^{RecSel|Acc}}{\epsilon_{sig}^{RecSel|Acc}} \times \frac{\epsilon_{norm}^{Trig|RecSel}}{\epsilon_{sig}^{Trig|RecSel}}$$

In first approximation, the detector acceptance is defined as the fraction of the decays having all the decay products in the geometrical detector acceptance. The total detector acceptance, however, will also be affected by the magnetic field and by the interactions with the detector material; these effects will be evaluated as part of the reconstruction efficiency in Sec. 7.3.2.

The geometrical detector acceptances for the signal and normalisation channels have been estimated with simulated samples, and are listed in Tab. 7.1. The decay products are required to fly in the LHCb detector acceptance, defined by the polar angle in the range of [10, 400] mrad, which is chosen to be larger than the physical LHCb detector acceptance in order to allow for the recovery of particles by the magnetic field.

As expected from the kinematic distribution of the final decay products, the geometrical acceptance efficiencies in Tab. 7.1 are similar for the signal and the $B^0 \rightarrow K^+ \pi^-$ decay, but lower for $B^+ \rightarrow J/\psi K^+$. The efficiencies were found to be stable across 2011 and 2012.

Table 7.1: Geometrical detector acceptance, estimated as the fraction of decays contained in the polar angle region of [10, 400] mrad.

	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	$B^+ \rightarrow J/\psi K^+$	$B^0 \rightarrow K^+ \pi^-$
$\epsilon^{Acc.}$	$(17.75 \pm 0.09)\%$	$(15.78 \pm 0.08)\%$	$(18.08 \pm 0.07)\%$

7.3.2 Reconstruction and selection efficiencies

$$\frac{\epsilon_{norm}}{\epsilon_{sig}} = \frac{\epsilon_{norm}^{Acc}}{\epsilon_{sig}^{Acc}} \times \frac{\epsilon_{norm}^{RecSel|Acc}}{\epsilon_{sig}^{RecSel|Acc}} \times \frac{\epsilon_{norm}^{Trig|RecSel}}{\epsilon_{sig}^{Trig|RecSel}}$$

The reconstruction efficiency is the fraction of decay candidates in the detector acceptance that are successfully reconstructed (Sec. 2.4); the selection efficiency is the fraction of reconstructed decay candidates that pass the selection (Ch. 4). Both efficiencies depend on the characteristics of the

decay channel: the number of particles in the final state, their kinematic distributions, the track finding algorithm efficiency, the particle identification efficiency, etc.

The determination of the selection and reconstruction efficiency is based both on simulations and real data.

Reconstruction and selection efficiencies from simulation

The combined effect of reconstruction and selection is first evaluated on simulated samples, considering only candidates in the detector acceptance (see Tab. 7.2). For the normalisation samples, a mass fit is performed before and after reconstruction and selection, to separate the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ candidates from a (small) background contribution; the mass models have been described in Sec. 7.2. The small background in the simulated normalisation samples arises from the other particles in the event (e.g. from the other B meson decay) because the selection has been optimised for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays, and not for the normalisation channels. The background in the simulated $B_s^0 \rightarrow \mu^+\mu^-$ sample is negligible.

The reconstruction efficiency of $B^+ \rightarrow J/\psi K^+$ is expected to be lower than the reconstruction efficiency of $B_s^0 \rightarrow \mu^+\mu^-$, because the efficiency to reconstruct a decay candidate depends on the number of tracks in the final state, as well as on their kinematic distribution. The efficiencies were found to be the same for 2011 and 2012 data.

Reconstruction efficiency from data

The efficiency measured on the data is expressed as a correction to that from the simulation, separately for muons and hadrons. The individual muon- and hadron-track reconstruction efficiency ratios are combined according to the final state composition, and multiplied by the simulated reconstruction efficiencies to obtain the reconstruction efficiency for a given decay channel.

The muon-track reconstruction efficiency is determined on a clean sample

Table 7.2: Reconstruction and selection efficiencies for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and the normalisation channels, evaluated on the Monte Carlo simulated samples, together with the statistical uncertainties.

$\epsilon^{RecSel Acc.}$	$B_{(s)}^0 \rightarrow \mu^+\mu^-$	$B^+ \rightarrow J/\psi K^+$	$B^0 \rightarrow K^+\pi^-$
	$(31.42 \pm 0.05)\%$	$(16.51 \pm 0.06)\%$	$(25.80 \pm 0.04)\%$

of $J/\psi \rightarrow \mu^+ \mu^-$ decays with the tag-and-probe method², described in Ref. [111]; the hadron-track reconstruction efficiency is determined by modifying the measured muon-track reconstruction efficiency to account for the hadron interactions with the detector material. This is done by means of simulations. Because of possible inaccuracies in the simulated detector material budget, the simulation of hadron interactions with the detector material introduces the largest systematic uncertainty to the selection and reconstruction efficiency.

The reconstruction efficiency depends on the kinematics, which differs from final state to final state. The muon- and hadron-track efficiencies are determined in bins of track pseudo-rapidity and momentum. The resulting muon-, kaon-, and pion-tracking efficiency maps, when applied to the simulated signal and normalisation channel samples, take into account the momentum and pseudo-rapidity distributions of the final state particles.

The reconstruction efficiencies measured from data yield the following (multiplicative) correction factors:

$$\begin{aligned} C_{Track} \left(\frac{B^+ \rightarrow J/\psi K^+}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 0.9914 \pm 0.0070_{stat} \pm 0.004_{syst} \pm 0.013_{material}, \\ C_{Track} \left(\frac{B^0 \rightarrow K^+ \pi^-}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 0.9996 \pm 0.0059_{stat} \pm 0.004_{syst} \pm 0.027_{material}. \end{aligned} \quad (7.7)$$

The first uncertainty in Eq. (7.7) is statistical, the second is the systematic uncertainty from the tag-and-probe procedure. The third uncertainty is the dominant contribution, and arises from the simulation of the hadronic interactions with the detector material. The increase in the collision energy may have an effect on the reconstruction and selection efficiencies and the factors in Eq. (7.7) were calculated separately for 2011 and 2012 samples. The results were found compatible, and were finally combined.

Corrections to the simulated selection efficiency

The selection criteria for signal and normalisation channels are kept as similar as possible (Tab. 4.1). The selection efficiency is determined partly on simulated samples, and partly on data. All differences that could lead to

²The tag-and-probe method uses two-prong decays, where one of the decay products, the “tag”, is fully reconstructed as a track, while the particle, the “probe”, is only partially reconstructed, i.e. not using the tracking information with which the tracking efficiency will be evaluated.

wrong efficiency estimates in case the simulation fails to reproduce the correct selection variable distributions were studied, and, if necessary, corrected for.

The choice of the *mass range* ($[4900, 6000]$ MeV/ c^2) might have an effect on the signal efficiency depending on the spread of the signal events in the di-muon spectrum. The effect of the mass range cut has been studied on the simulated signal samples, accounting for possible inaccuracies in simulating mass resolution effects. No efficiency correction was found necessary for the signal channels. The same procedure was repeated for the normalisation channels by investigating the effect of the di-muon mass cut on $B^+ \rightarrow J/\psi K^+$. Again, the correction was found negligible because of the narrow J/ψ mass peak.

The IP *distribution* of the simulated samples differs from the distribution observed in data. If the differences are large, they could lead to incorrect reconstruction and selection- efficiency. We reweighted the simulated IP distribution to match the measured IP distribution and studied the effect of reweighting on the reconstruction and selection efficiencies. These were found to be dependent on the IP distribution reweighting, but the effect was shown to cancel in the ratio, and thus no additional correction was deemed necessary. Also, the simulated distribution of the Kaon IP χ^2 in $B^+ \rightarrow J/\psi K^+$ decay was studied. A cut on the kaon IP χ^2 is used only for $B^+ \rightarrow J/\psi K^+$ selection (Tab. 4.1), and therefore is not “balanced” in the efficiency ratio. Comparison between the distributions in data and in simulation showed no need for an additional correction.

The *decay-time distribution* of the B_s^0 mesons in the Monte Carlo $B_s^0 \rightarrow \mu^+ \mu^-$ sample is simulated using a single exponential with a mean lifetime 1.469 ps. As explained in Sec. 1.2, the heavy and light B_s^0 meson mass eigenstates have a different lifetime, and the relative decay-width asymmetry (see Eq. (1.26)) is measured to be non-zero in the B_s^0 system [30, 31]. Because the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection (Tab. 4.1) depends on the B meson lifetime, the signal selection efficiency determined from the simulated $B_s^0 \rightarrow \mu^+ \mu^-$ sample (with a zero decay-width asymmetry) must be corrected for [109]. The correction factor for the $B_s^0 \rightarrow \mu^+ \mu^-$ selection efficiency is physics-model dependent, and is obtained assuming the Standard Model value of $\mathcal{A}_{\Delta\Gamma} = 1$ for $B_s^0 \rightarrow \mu^+ \mu^-$, the measured values for the relative decay-width asymmetry in the B_s^0 system $(\Gamma_L - \Gamma_H)/(\Gamma_L + \Gamma_H) = 0.0615 \pm 0.0085$, the average B^0 lifetime of $\tau(B^0) = 1.519 \pm 0.007$, and the average B_s^0 lifetime of

$\tau(B_s^0) = 1.516 \pm 0.011$ [70]. The signal selection efficiency corrections are:

$$\begin{aligned} C_{B_s^0 \rightarrow \mu^+ \mu^-}^{Sel} &= 1.0150 \pm 0.00006_{stat}, \\ C_{B_s^0 \rightarrow \mu^+ \mu^-}^{Sel} &= 1.0457 \pm 0.00020_{stat}. \end{aligned} \quad (7.8)$$

The correction for $B^0 \rightarrow \mu^+ \mu^-$ is not model dependent; it is only necessary because the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ detection efficiencies are taken to be equal, and determined from the $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo sample.

Selection efficiency from data

The isMUON requirement is applied on signal and normalisation channels alike³ (Tab. 4.1). The simulated isMUON identification efficiency (including the precise muon-detector acceptance) is compared to the efficiencies measured on data (see Sec. 2.5). The differences with simulation are accounted for with correction maps, built in bins of the muon track momentum (p) and transverse momentum (p_T). The (multiplicative) correction terms that need to be applied to the isMUON efficiencies determined from the simulated samples are computed by folding the correction maps with the p and p_T spectra of the simulated decays:

$$\begin{aligned} C_{MuID}^{2011} \left(\frac{B^+ \rightarrow J/\psi K^+}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 1.0027 \pm 0.0013_{stat} \pm 0.0072_{syst}, \\ C_{MuID}^{2012} \left(\frac{B^+ \rightarrow J/\psi K^+}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 0.9959 \pm 0.0011_{stat} \pm 0.0140_{syst}, \\ C_{MuID}^{2011} \left(\frac{B^0 \rightarrow K^+ \pi^-}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 0.9935 \pm 0.0010_{stat} \pm 0.0034_{syst}, \\ C_{MuID}^{2012} \left(\frac{B^0 \rightarrow K^+ \pi^-}{B_{(s)}^0 \rightarrow \mu^+ \mu^-} \right) &= 0.9889 \pm 0.0008_{stat} \pm 0.0130_{syst}, \end{aligned} \quad (7.9)$$

where the first uncertainty is statistical and the second systematic, determined varying the trigger requirements of the tag-and-probe tracks.

Furthermore, to reduce the hadron mis-identification, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection includes a ΔLL requirement (see Tab. 4.1). The simulated efficiency

³In case of $B^0 \rightarrow K^+ \pi^-$, only the acceptance efficiency map is used, and no muon identification is required.

$B_s^0 \rightarrow \mu^+ \mu^-$ efficiency in Tab. 7.2 has to be multiplied by the ΔLL cut efficiencies, determined by folding the muon efficiencies measured from the $J/\psi \rightarrow \mu^+ \mu^-$ data (Sec. 2.5.2) with the $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo phase space distribution:

$$\begin{aligned}\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\Delta\text{LL}}(2011) &= (97.49 \pm 0.01_{\text{stat}} \pm 0.27_{\text{syst}})\%, \\ \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\Delta\text{LL}}(2012) &= (97.98 \pm 0.01_{\text{stat}} \pm 0.28_{\text{syst}})\%.\end{aligned}\quad (7.10)$$

The systematic uncertainty in Eq. (7.10) includes trigger effects and is estimated by the difference between requiring a TIS probe or a TOS tag muon in $J/\psi \rightarrow \mu^+ \mu^-$.

A *ghost probability* cut was included in the selection to reduce the rate of fake tracks in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ reconstruction (Sec. 4.1). The ghost-probability cut efficiency is defined as the efficiency of the cut when all other selection criteria are fulfilled. Just as the ΔLL efficiency, it is not included in the simulated efficiencies (Tab. 7.2) and is determined from data instead. For signal decays, it is measured using the tag-and-probe method and $J/\psi \rightarrow \mu^+ \mu^-$ decays, and the efficiency map is folded with the simulated $B_s^0 \rightarrow \mu^+ \mu^-$ di-muon spectrum in bins of the muon p and p_T . The ghost-probability cut efficiency for signal channels was $(99.70 \pm 0.01_{\text{stat}} \pm 0.04_{\text{syst}})\%$ in 2011, and $(99.59 \pm 0.01_{\text{stat}} \pm 0.09_{\text{syst}})\%$ in 2012. The systematic uncertainty is estimated by comparing the efficiencies from a $B^+ \rightarrow J/\psi K^+$ sample, with those from an inclusive J/ψ sample from B decays. The ghost probability for the $B^0 \rightarrow K^+ \pi^-$ is estimated using as a probe the kaon from $B^+ \rightarrow J/\psi K^+$. The efficiency is assumed to be the same for pions and kaons, $(99.057 \pm 0.015)\%$ for both years. The ghost probability cut is not applied to the $B^+ \rightarrow J/\psi K^+$ normalisation channel.

Total reconstruction and selection efficiencies

The total reconstruction and selection efficiencies relative to $B_s^0 \rightarrow \mu^+ \mu^-$ are given in Tab. 7.3. They are obtained multiplying the efficiencies in Tab. 7.2 by the “correction” terms given in Eqs. (7.7)-(7.10), as well as the ghost-probability cut efficiency.

7.3.3 Trigger efficiencies

$$\frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} = \frac{\epsilon_{\text{norm}}^{\text{Acc}}}{\epsilon_{\text{sig}}^{\text{Acc}}} \times \frac{\epsilon_{\text{norm}}^{\text{RecSel}|\text{Acc}}}{\epsilon_{\text{sig}}^{\text{RecSel}|\text{Acc}}} \times \frac{\epsilon_{\text{norm}}^{\text{Trig}|\text{RecSel}}}{\epsilon_{\text{sig}}^{\text{Trig}|\text{RecSel}}}.$$

The LHCb trigger system records all information necessary to measure the trigger efficiencies directly from data with the **TISTOS** method described in Ch. 5. The trigger efficiencies are calculated for events that have been selected and reconstructed.

The $B_s^0 \rightarrow \mu^+ \mu^-$ trigger efficiency is measured from data. The majority of the recorded $B^+ \rightarrow J/\psi K^+$ and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates are triggered by the same muon trigger lines (Ch. 3). The muon trigger line efficiency is measured directly from the $B^+ \rightarrow J/\psi K^+$ sample in bins of muon p and p_T , using the **TISTOS** method.

The trigger efficiency for the signal channels $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, is found by folding the result with the simulated muon spectrum of $B_s^0 \rightarrow \mu^+ \mu^-$ decays:

$$\begin{aligned}\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2011) &= (92.1 \pm 0.5_{stat} \pm 1.6_{syst})\%, \\ \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2012) &= (92.4 \pm 0.3_{stat} \pm 1.9_{syst})\%,\end{aligned}\quad (7.11)$$

where the systematic uncertainty is a combination of the **TISTOS** method uncertainty, and the uncertainty stemming from the small fraction of $B_s^0 \rightarrow \mu^+ \mu^-$ events that are not triggered by the muon trigger lines. The total trigger efficiency for $B^+ \rightarrow J/\psi K^+$ is the efficiency determined for muon lines, with an additional systematic uncertainty from the small fraction of events triggered by other than muon lines:

$$\begin{aligned}\epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2011) &= (88.0 \pm 0.5_{stat} \pm 2.5_{syst})\%, \\ \epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2012) &= (86.6 \pm 0.3_{stat} \pm 2.3_{syst})\%,\end{aligned}\quad (7.12)$$

where the systematic uncertainty also includes the uncertainty of the **TISTOS** method.

$B^0 \rightarrow K^+ \pi^-$ candidates are required to be triggered independently of the signal candidate in the first trigger levels, L0 and HLT1. This reduces the differences between the signal and the $B^0 \rightarrow K^+ \pi^-$ trigger lines, and, at the same time, provides enough statistics for the normalisation (see Sec. 6.2). The efficiency to trigger independent of the signal candidate is by definition the same for all B decays⁴. The L0 and HLT1 TIS efficiency can thus be measured from $B^+ \rightarrow J/\psi K^+$ decays:

$$\begin{aligned}\epsilon_{L0HLT1}^{TIS}(2011) &= (5.046 \pm 0.04_{stat} \pm 0.4_{syst})\%, \\ \epsilon_{L0HLT1}^{TIS}(2012) &= (5.920 \pm 0.03_{stat} \pm 0.4_{syst})\%,\end{aligned}\quad (7.13)$$

⁴Once the kinematic distribution of the decaying meson is unfolded, as discussed in Ch. 5.

where the systematic uncertainty is due to the **TISTOS** method. The HLT2 trigger efficiency is estimated from a simulated $B^0 \rightarrow K^+\pi^-$ sample, for events passing the L0 and HLT1 TIS requirement:

$$\begin{aligned}\epsilon_{B^0 \rightarrow K^+\pi^-}^{HLT2|L0HLT1TIS}(2011) &= (91.5 \pm 0.3_{stat})\%, \\ \epsilon_{B^0 \rightarrow K^+\pi^-}^{HLT2|L0HLT1TIS}(2012) &= (91.6 \pm 0.2_{stat})\%.\end{aligned}\quad (7.14)$$

The trigger efficiency ratios used to calculate the normalisation factors are in Tab. 7.3.

7.4 Hadronisation factors

$$\alpha_s = \underbrace{\frac{\mathcal{B}(norm)}{N_{norm}} \times \frac{\epsilon_{norm}}{\epsilon_{sig}}}_{\alpha_d} \times \frac{\mathbf{f}_d}{\mathbf{f}_s}$$

The b (or \bar{b}) hadronisation probabilities are determined experimentally. The hadronisation factors for B^0 and B^+ mesons, f_d and f_u , are measured to be very similar (see Eq. (2.5)), and are taken to be equal⁵: $f_d = f_u \equiv f_d$. Therefore, the hadronisation-fraction ratio in the $B^0 \rightarrow \mu^+\mu^-$ normalisation factor is equal to one. The B_s^0 meson hadronisation fraction f_s , however, differs from f_d and needs to be accounted for in the $B_s^0 \rightarrow \mu^+\mu^-$ normalisation factor α_s (see Eq. (7.3)).

The production ratio f_s/f_d has been measured by LHCb from the relative abundance $B_s^0 \rightarrow D_s^-\pi^+$ and $B^0 \rightarrow D^-K^+$ decays [112], and also from inclusive semi-leptonic $B_s^0 \rightarrow D\mu X$ and $B^{0,-} \rightarrow D\mu X$ decays [82]. The combination of these results [113],

$$f_s/f_d = 0.259 \pm 0.015, \quad (7.15)$$

is used in this analysis.

The ratio in Eq. (7.15) was determined at $\sqrt{s} = 7$ TeV. The effect of the increase in the collision energy to $\sqrt{s} = 8$ TeV in 2012 was studied by measuring the ratio of $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ candidates at both energies. The yield ratio is directly proportional to the hadronisation ratio f_s/f_d and remained stable after the increase in the collision energy. The f_s/f_d value in Eq. (7.15) was therefore used for the whole data sample.

Furthermore, LHCb has measured an evidence for a possible f_s/f_d dependence on the B meson transverse momentum [112]. In the simulated

⁵The equality between the light B meson production fractions is additionally supported by the isospin symmetry of the strong force.

$B_s^0 \rightarrow \mu^+ \mu^-$ sample, the transverse momentum of the selected B mesons ranges from 3 to 9 GeV/ c . The average B meson transverse momentum in Ref. [112] was 10.4 GeV/ c . Using the transverse momentum dependence from Ref. [112], it was estimated that the value of f_s/f_d would change by about 0.02 (about one standard deviation) for a B -meson transverse momentum of 3 GeV/ c . Therefore, the f_s/f_d has been assumed to be constant in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.

7.5 Normalisation factors

The normalisation factors, α_s for $B_s^0 \rightarrow \mu^+ \mu^-$ and α_d for $B^0 \rightarrow \mu^+ \mu^-$, are calculated separately for each normalisation channel, and for the data at $\sqrt{s} = 7$ TeV and at $\sqrt{s} = 8$ TeV. The results as well as the values of the parameters entering the calculations are shown in Tab. 7.3⁶. The values from the different samples and normalisation channels are finally combined:

$$\begin{aligned} N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-} &= N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}^{2011} + N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}^{2012}, \\ &= \left(\frac{1}{\alpha_{d(s)}^{2011}} + \frac{1}{\alpha_{d(s)}^{2012}} \right) \times \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-), \\ &\equiv \frac{1}{\alpha_{d(s)}} \times \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-). \end{aligned} \quad (7.16)$$

The f_s/f_d is taken to be fully correlated in the combination, and the combined normalisation parameter values are

$$\begin{aligned} \alpha_d &= (2.38 \pm 0.09) \times 10^{-11}, \\ \alpha_s &= (8.93 \pm 0.62) \times 10^{-11}. \end{aligned} \quad (7.17)$$

⁶The numbers in Tab. 7.3 exclude the decay-time distribution corrections. These are taken into account in the MLL fit.

Table 7.3: Normalisation factors for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, calculated for each normalisation channel and collision energy.

$\mathcal{B}(norm)$ $\times 10^{-5}$	N_{norm}	$\frac{\epsilon_{norm}^{AccRecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{AccRecSel}}$	$\frac{\epsilon_{norm}^{TRIG AccRecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG AccRecSel}}$	α_d $\times 10^{-11}$	α_s $\times 10^{-10}$
2011 data					
$B^+ \rightarrow J/\psi K^+$ 6.025 ± 0.205	355232 ± 1227	$(47.8 \pm 1.2)\%$	$(95.5 \pm 2.0)\%$	7.63 ± 0.35	2.86 ± 0.22
$B^0 \rightarrow K^+ \pi^-$ 1.94 ± 0.06	10809 ± 439	$(84.7 \pm 2.4)\%$	$(5.01 \pm 0.41)\%$	7.51 ± 0.76	2.81 ± 0.33
2012 data					
$B^+ \rightarrow J/\psi K^+$ 6.025 ± 0.205	761122 ± 2451	$(47.3 \pm 1.3)\%$	$(93.7 \pm 2.6)\%$	3.46 ± 0.18	1.29 ± 0.11
$B^0 \rightarrow K^+ \pi^-$ 1.94 ± 0.06	26749 ± 447	$(84.0 \pm 2.6)\%$	$(5.87 \pm 0.42)\%$	3.52 ± 0.30	1.32 ± 0.14

Chapter 8

Results of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

The results of the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis are presented in this chapter. The chapter begins introducing the Maximum Likelihood fit in Sec. 8.1. The measured $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are presented in Sec. 8.2; the significance of the results with respect to the null (no signal) hypothesis is calculated in Sec. 8.3. Since the measured $B^0 \rightarrow \mu^+ \mu^-$ was found compatible with the null hypothesis, the $B^0 \rightarrow \mu^+ \mu^-$ measurement is used to exclude more extreme $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ values with the CL_s technique (Sec. 8.4).

8.1 The un-binned maximum likelihood fit

The procedure of finding the model parameter values that best describe the data is called parameter estimation [19,114]. In a Maximum Likelihood (ML) fit, the model parameters are estimated by maximising the agreement of the model with the observed data. The parameters selected are those that maximise the so-called “likelihood” function (\mathcal{L}). The model likelihood is thus a function of the parameters given the observed data, and it is maximised¹ by solving:

$$\frac{d \ln \mathcal{L}(\vec{\lambda})}{d \vec{\lambda}} \Bigg|_{\vec{\lambda}=\hat{\vec{\lambda}}} = 0, \quad (8.1)$$

¹Statistical software packages typically minimise a function rather than maximise, and in practice the likelihood maximum is found by *minimising* the *negative logarithm* of the likelihood. Using logarithmic functions is numerically more convenient because it reduces large likelihood values and converts the product of candidate likelihoods to a sum.

where $\vec{\lambda}$ is the set of parameters on which the likelihood depends on, and $\hat{\vec{\lambda}}$ the set of parameter values satisfying Eq. (8.1), called the ML “estimates” (MLE).

The MLE parameters for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ likelihood model are found by the MIGRAD function of MINUIT [105]. The uncertainty in the parameter estimates is calculated using the Rao-Cramer-Frechet (RCF) inequality, which gives a lower bound on the estimator variance. For a single parameter λ_i , it can be expressed as:

$$V[\hat{\lambda}_i] \geq \frac{(1 + \partial b / \partial \lambda_i)^2}{E[-\partial^2 \ln \mathcal{L}(\vec{\lambda}) / \partial \lambda_i^2]}, \quad (8.2)$$

where E denotes the expectation value, and b is the bias of the estimator, defined as $b \equiv E[\hat{\lambda}_i] - \lambda_i^{true}$. In practice, ML estimators are often assumed to be unbiased; this assumption was verified for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ likelihood model with Monte Carlo toy experiments, and the estimator biases were indeed found negligible. The RFC bound becomes an equality for the ML estimators when the sample size tends to infinity, in which case the MLE $\hat{\lambda}_i$ is Gaussian distributed around the true value λ_i^{true} with a variance

$$(\hat{\sigma}_{\hat{\lambda}_i})^{-2} = \left. \frac{\partial^2 \ln \mathcal{L}(\vec{\lambda})}{\partial \lambda_i^2} \right|_{\lambda_i = \hat{\lambda}_i}. \quad (8.3)$$

In the ML fit, the MLE uncertainty is calculated with the MINOS function of MINUIT [105]. MINOS employs the “graphical technique”, in which Eq. (8.3) represents the second term in the Taylor expansion of $\ln \mathcal{L}(\lambda_i)$ around its maximum ($\lambda_i = \hat{\lambda}_i$):

$$\begin{aligned} \ln \mathcal{L}(\lambda_i) &\simeq \ln \mathcal{L}(\lambda_i)|_{\lambda_i = \hat{\lambda}_i} + \left[\frac{\partial \ln \mathcal{L}(\lambda_i)}{\partial \lambda_i} \right]_{\lambda_i = \hat{\lambda}_i} (\lambda_i - \hat{\lambda}_i) \\ &\quad + \frac{1}{2!} \left[\frac{\partial^2 \ln \mathcal{L}(\lambda_i)}{\partial \lambda_i^2} \right]_{\lambda_i = \hat{\lambda}_i} (\lambda_i - \hat{\lambda}_i)^2, \end{aligned} \quad (8.4)$$

i.e.

$$\ln \mathcal{L}(\lambda_i) = \ln \mathcal{L}_{max} - \frac{(\lambda_i - \hat{\lambda}_i)^2}{2(\hat{\sigma}_{\hat{\lambda}_i})^2}, \quad (8.5)$$

and thus

$$\ln \mathcal{L}(\hat{\lambda}_i \pm \hat{\sigma}_{\hat{\lambda}_i}) = \ln \mathcal{L}_{max} - \frac{1}{2}. \quad (8.6)$$

Eq. (8.6) shows that the largest variation of $\hat{\lambda}_i$ inside a 68.3% (central) confidence interval leads to a change of $\frac{1}{2}$ of the log-likelihood around its maximum.

It is this relation that is exploited by MINOS to determine the MLE uncertainty. In the large sample limit, the likelihood function becomes a Gaussian centred around the ML estimate, and the variation corresponds to plus or minus one standard deviation.

8.1.1 Nuisance parameters

The likelihood model used to extract signal branching fractions was described in Ch. 6. It includes components for the signal decays (Sec. 6.2 and Sec. 6.3), for the combinatorial background (Sec. 6.6), for the doubly mis-identified background from $B_{(s)}^0 \rightarrow h^+ h^-$ decays (Sec. 6.4), and for the single mis-identified semi-leptonic decays, $B^0 \rightarrow \pi^- \mu^+ \nu$, $B_s^0 \rightarrow K^- \mu^+ \nu$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$, and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ (Sec. 6.5). In addition to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, the model contains many other parameters: the signal and background normalisation factors, various invariant mass distribution shape parameters, and the BDT category fractions. These parameters are necessary to define the likelihood model, but are not of direct interest, and are called “nuisance” parameters ($\vec{\nu}$). The values of the nuisance parameters are constrained by auxiliary measurements, theoretical predictions, or both, and therefore carry an uncertainty. These uncertainties introduce a systematic uncertainty in the measurement of the signal branching fraction.

The nuisance parameter uncertainties are included into the likelihood function through a “constraint” probability density function:

$$pdf(\vec{\nu}; \vec{\mu}_\nu, \vec{\sigma}_\nu) = G(\nu_1; \mu_{\nu_1}, \sigma_{\nu_1}) \cdot G(\nu_2; \mu_{\nu_2}, \sigma_{\nu_2}) \cdot \dots , \quad (8.7)$$

where it is assumed that the nuisance parameter values, ν_i , are Gaussian distributed, with standard deviations, σ_{ν_i} , around² the central nuisance parameter values μ_{ν_i} . The constraint term in Eq. (8.7) is added to the model log-likelihood³:

$$\ln \mathcal{L}^{Tot}(\vec{\lambda}) = \ln \mathcal{L}(\vec{\lambda}) + \ln \underbrace{\left(\prod_{i=1}^{N_{obs}} pdf(\vec{\nu}; \vec{\mu}_\nu, \vec{\sigma}_\nu) \right)}_{\text{Constraint}}, \quad (8.8)$$

The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fraction uncertainties that include the nuisance parameter uncertainties are obtained by applying the

²In case the nuisance parameter uncertainties are asymmetric, an asymmetric Gaussian function is used.

³In RooFIT, the presence of the `ExternalConstraints()` argument in the fitting function ensures that the constraint term Eq. (8.7) is included to the likelihood before starting the likelihood maximising with MIGRAD.

graphical method once the likelihood is written as a function of only the parameters of interest. This is achieved by writing the nuisance parameters as functions of the parameter of interest, θ :

$$\ln \mathcal{L}^{Tot}(\theta) = \ln \mathcal{L}^{Tot}(\theta, \hat{\nu}(\theta)). \quad (8.9)$$

Equation 8.9 defines a “profile likelihood” function; $\hat{\nu}(\theta)$ denotes the nuisance parameter estimates that maximise the total likelihood for a given θ value, and θ can be either of the signal branching fractions.

8.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions

The signal branching fraction ML estimates are obtained from an un-binned ML fit to the di-muon invariant mass distribution, performed simultaneously in the eight BDT categories. The signal branching fractions are free fit parameters defined as the signal candidate yields times the normalisation factors, as in Eq. (7.3). The branching fraction values from the ML fit are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.87_{-0.95}^{+1.11} (\text{stat + syst}_{\text{base}})) \times 10^{-9}, \quad (8.10)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.74_{-2.05}^{+2.42} (\text{stat + syst}_{\text{base}})) \times 10^{-10}, \quad (8.11)$$

where the uncertainty is evaluated with the graphical technique from the profile likelihood (Sec. 8.1.1) scan. The errors in Eqs. (8.10) and (8.11) include the statistical uncertainty and the uncertainties from the nuisance parameters. The contribution of the statistical uncertainty alone is measured by performing the fit with all the nuisance parameters fixed to their mean values and excluding the constraint term (Eq. (8.8)) from the likelihood (see Eqs. (8.12) and (8.12)). The systematic uncertainty from the nuisance parameters was found by subtracting (in quadrature) the statistical uncertainty from the total branching fraction uncertainty in Eqs. (8.10) and (8.11).

Because of the imprecise branching fraction knowledge and the complicated normalisation procedure, the $\Lambda_b^0 \rightarrow p\mu^-\nu$ component (see Sec. 6.5) is not included in the baseline fit model. Moreover, the parameters describing the shape of the invariant mass distribution of the mis-identified background components, were, for technical reasons, kept constant at the mean values in the baseline fit. The additional systematic uncertainty arising from these simplifications is estimated from the difference in the central values, calculated between the branching fraction values obtained with the baseline model, and the branching fraction values obtained with a model that includes the $\Lambda_b^0 \rightarrow p\mu^-\nu$ component and where the mis-identified semi-leptonic

background mass distribution parameters are increased or decreased by one standard deviation.

The final results together with the statistical and systematic uncertainties are [58]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.9_{-1.0}^{+1.1}(\text{stat})_{-0.1}^{+0.3}(\text{syst})) \times 10^{-9}, \quad (8.12)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.7_{-2.1}^{+2.4}(\text{stat})_{-0.4}^{+0.6}(\text{syst})) \times 10^{-10}. \quad (8.13)$$

The systematic uncertainty contains the uncertainty from the nuisance parameters, the $\Lambda_b^0 \rightarrow p\mu^-\nu$ background component, and the modelling of the mis-identified semi-leptonic background mass distribution. The projection of the baseline model on the di-muon invariant mass distribution is shown in Fig. 8.1. The di-muon mass distribution in the last three BDT categories, where the fit is the most sensitive to the signals, is presented in Fig. 8.2.

8.3 The significances of the measured branching fractions

In searches of new phenomena, the aim is to reject the hypothesis that the data are consistent with the null (no signal) hypothesis. Therefore, the discrepancy between the data and the null hypothesis is a measure of the signal “significance”, and it is expressed by the probability (“p-value”) that under the null hypothesis, one would obtain a result equal to or more extreme than that observed:

$$p = \int_{Q_0^{obs}}^{\infty} pdf(Q_0|H_0)dQ_0, \quad (8.14)$$

where $pdf(Q_0|H_0)$ is the probability density distribution of the test-statistic, Q_0 , under the null hypothesis, H_0 , and Q_0^{obs} , is the value of the same test-statistic as observed in the data (see Fig. 8.3).

Typically, the p-value is expressed in terms of Gaussian standard deviations, as the number of standard deviations counted from the mean of the Gaussian to the point where the area under the remaining tail⁴ is equal to the p-value (see Fig. 8.3).

The null hypotheses for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are defined through the likelihood model without the signal decay under test⁵ (the other signal decay component is included to exclude the possible cross-feed between the signal components).

⁴This is the definition of a one-sided significance.

⁵The signal component branching fraction is fixed to zero.

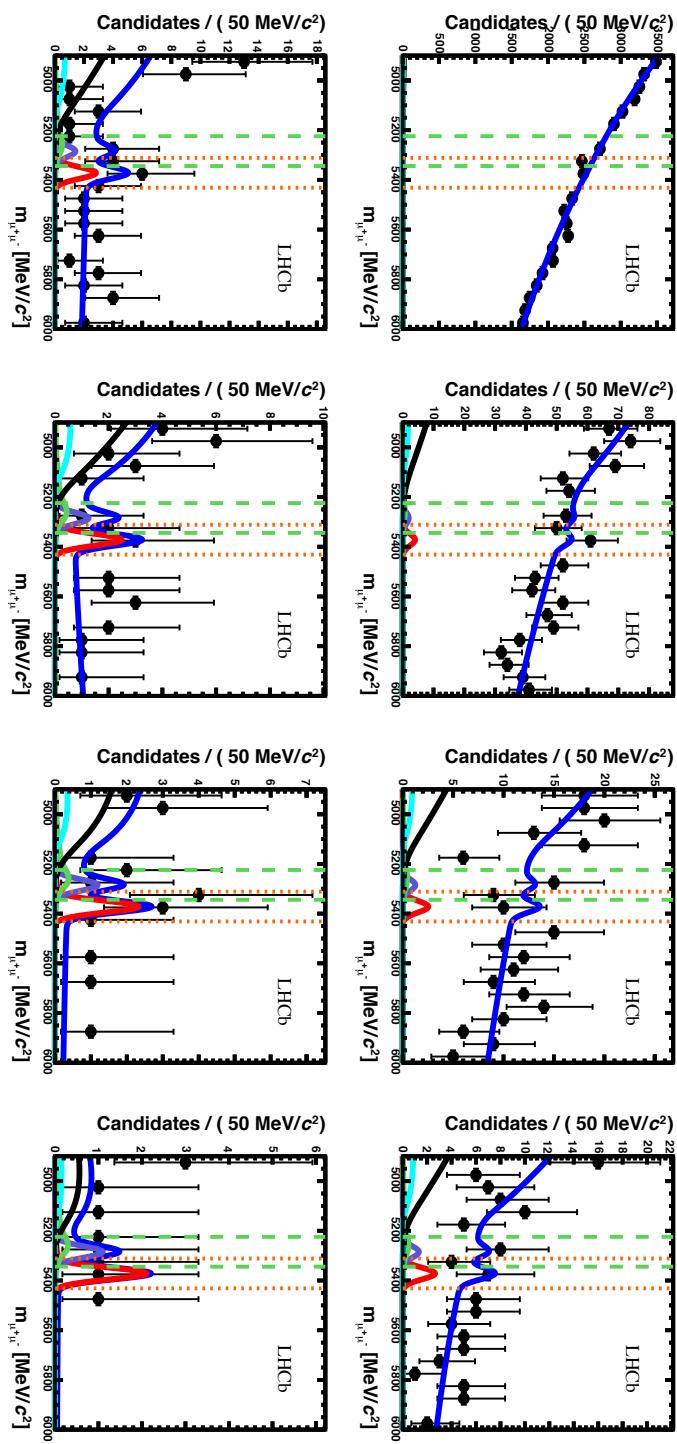


Figure 8.1: Di-muon invariant mass distribution of the signal candidates selected in the 3 fb^{-1} LHCb data-sample, shown for the 8 BDT categories (black dots). The (baseline) likelihood model projection is denoted by the solid blue lines; the different model components are denoted by the other solid lines: $B_s^0 \rightarrow \mu^+\mu^-$ (red), $B^0 \rightarrow \mu^+\mu^-$ (purple), $B_{(s)}^0 \rightarrow h^+h^-$ (green), $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (black), and $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ (cyan).

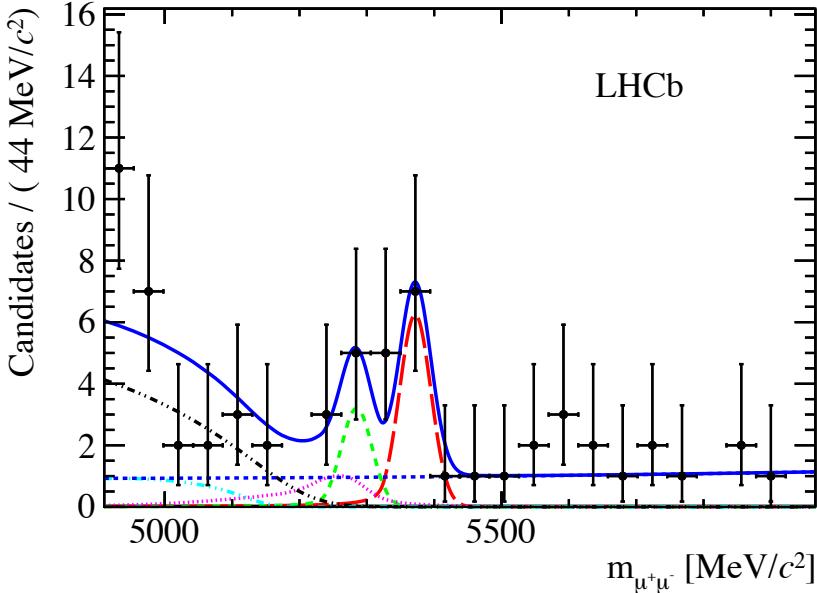


Figure 8.2: Di-muon invariant mass distribution of the signal candidates selected in the 3 fb^{-1} LHCb data-sample, shown for $\text{BDT} > 0.7$ (black dots). The (baseline) likelihood model projection is denoted by the solid blue line, and the different model components by the other lines: $B_s^0 \rightarrow \mu^+ \mu^-$ (long dashed red), $B^0 \rightarrow \mu^+ \mu^-$ (medium dashed green), $B_{(s)}^0 \rightarrow h^+ h^-$ (dotted magenta), $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (dot-dashed black), $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ (dot-dashed light blue), combinatorial background (medium dashed blue).

The compatibility between the observed data and a signal branching fraction hypothesis, \mathcal{B} , is described by the test statistic $Q_{\mathcal{B}}$, defined using the profile likelihood (defined in Eq. (8.9)) to include the effect of the nuisance parameter uncertainties into the calculation of the significance of the measurement. The test statistic $Q_{\mathcal{B}}$ is defined as a profile-likelihood ratio:

$$Q_{\mathcal{B}} \equiv Q(\mathcal{B}) \equiv \frac{\mathcal{L}(\mathcal{B}, \hat{\bar{\nu}}(\mathcal{B}))}{\mathcal{L}(\hat{\mathcal{B}}, \hat{\bar{\nu}}(\hat{\mathcal{B}}))}, \quad (8.15)$$

The statistic is zero at the ML branching-fraction estimate, and decreases as the compatibility between the branching fraction hypotheses and the data decreases (see Fig. 8.4, and note that $\Delta \text{LL} = -Q_{\mathcal{B}}$). The inclusion of the nuisance parameter uncertainties “smears” the profile-likelihood ratio and thus raises the compatibility between the data and the hypothesis.

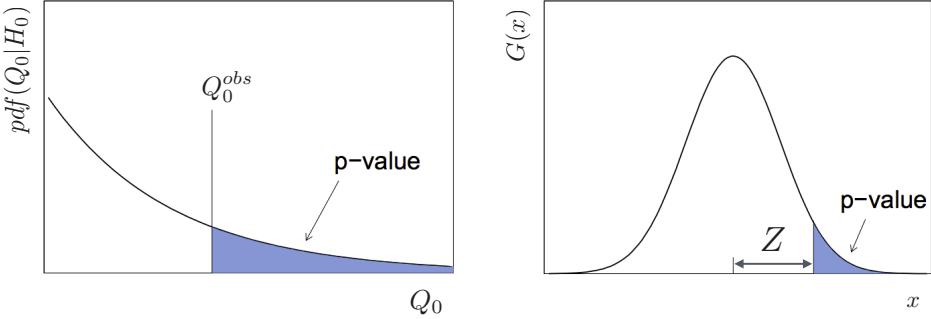


Figure 8.3: The *pdf* of the test-statistic, Q_0 , under the null hypothesis, together with the observed test-statistic value, Q_0^{obs} , and the p-value (left). A Gaussian distribution, $G(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$, illustrating the relation between the significance, Z , and the p-value (right) [115].

If the data sample is large enough, and consists of independent and identically distributed events⁶, then according to Wald [116], the distribution of $-2\ln Q_0$ under null hypothesis approaches a non-central χ^2 distribution with one degree of freedom. As shown by Wilks in Ref. [117], the non-centrality of the χ^2 vanishes if the signal branching fraction estimates are unbiased and the test statistic distribution is given by

$$pdf(-2 \ln Q_0 | H_0) = \frac{1}{2} \delta(-2 \ln Q_0) + \frac{1}{2\sqrt{2\pi}} \frac{1}{\sqrt{-2 \ln Q_0}} e^{\ln Q_0}, \quad (8.16)$$

where the first term models the fact that only positive branching-fraction estimates are considered⁷. As shown in Ref. [115], the observed significance with respect to null hypothesis can be computed from $\sqrt{-2 \ln Q_0^{obs}}$.

The measured p-values and the corresponding significances of the measured $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are:

$$p_{B_s^0 \rightarrow \mu^+ \mu^-} = 3.17 \times 10^{-5} \xrightarrow{\text{1sided}} 4.0\sigma, \quad (8.17)$$

$$p_{B^0 \rightarrow \mu^+ \mu^-} = 2.28 \times 10^{-2} \xrightarrow{\text{1sided}} 2.0\sigma. \quad (8.18)$$

⁶A large number of independent and identically distributed random variables will be, according to the Central Limit Theorem, approximately Gaussian distributed regardless of the underlying individual distributions.

⁷In case the fluctuations in the background give rise to a negative (and therefore non-physical) branching fraction estimates, the test statistic is set to zero.

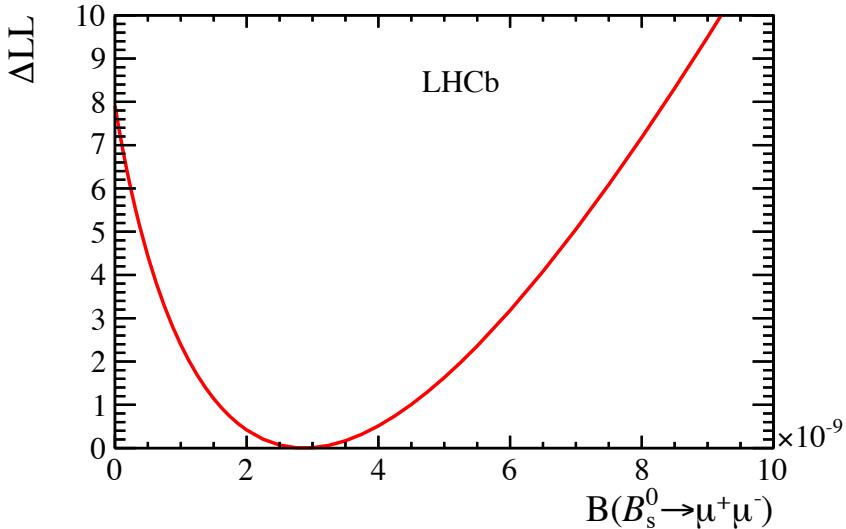


Figure 8.4: The observed profile likelihood ratio, expressed as the difference between the log-likelihoods of the hypotheses under the test: $\Delta LL = \ln\mathcal{L}(\hat{\mathcal{B}}, \hat{\nu}) - \ln\mathcal{L}(\mathcal{B}, \hat{\nu}(\mathcal{B}))$. The value depends on the branching fraction chosen for the numerator and for the null-hypothesis can be read from the intercept at $\mathcal{B} = 0$.

In the traditional language of high-energy physics, one would express the results in Eqs. (8.17), (8.18) by stating that the data show a clear evidence for the $B_s^0 \rightarrow \mu^+\mu^-$ decay ($Z > 3.0\sigma$), but no evidence for $B^0 \rightarrow \mu^+\mu^-$.

8.4 The exclusion limit from CL_s

The measured $B^0 \rightarrow \mu^+\mu^-$ branching fraction is not statistically significant (Eq. (8.18)). The observed candidates in the B^0 search window, however, can be used to set a limit on the highest possible $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ value, called the “exclusion limit”. This section reviews the main steps of this procedure and presents the results. A detailed account can be found in Sec. 3.3 of Ref. [109].

The $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ exclusion limits are set with the CL_s method. The CL_s uses a test statistic which depends on the number of expected background (b) and signal (s) candidates, and on the number of observed candidates (d):

$$Q_{\mathcal{B}}^{CL} \equiv Q(\mathcal{B})^{CL} \equiv \prod_i^{72} \frac{\mathcal{P}(d_i, b_i + s_i(\mathcal{B}))}{\mathcal{P}(d_i, b_i)}, \quad (8.19)$$

where $\mathcal{P}(x, y)$ is the Poissonian probability to observe x candidates in a bin,

when y candidates are expected. The product runs over the 72 di-muon mass bins⁸.

The distribution of the test statistic needs to be known for every $B^0 \rightarrow \mu^+ \mu^-$ branching fraction hypothesis. The distributions are found by generating, for each $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ hypothesis, two different sets of simulated experiments (pseudo-experiments): one generated according to the expected background yields in the 72 bins, and another set generated according to the expected signal-plus-background yields in the 72 bins. The test-statistic distributions in the first set is used as a proxy for the test-statistic distribution under the background only hypothesis, and the test-statistic distribution in the second set, for the test-statistic distribution assuming the signal-plus-background hypothesis.

The expected number of background candidates in each bin, b_i , is calculated from the expected combinatorial background, doubly mis-identified $B_{(s)}^0 \rightarrow h^+ h^-$, and cross-feed of the $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in the B^0 search region. The expected number of signal candidates in a mass bin, s_i , depends on the chosen branching fraction hypothesis, and is calculated by fixing the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction to the desired value in the likelihood model (Ch. 6). The expected background and signal yields and the observed candidates yields are shown in Tab. 8.1.

Once the test statistic distribution for the background hypothesis, $pdf_b(Q_B^{CL})$, and for the signal-plus-background hypotheses, $pdf_{s+b}(Q_B^{CL})$, have been obtained from the pseudo-experiments, one can determine the compatibility of these hypotheses with the observed value of the test statistics, $Q_B^{CL,obs.}$:

$$1 - CL_b \equiv \int_{Q_B^{CL,obs.}}^{+\infty} pdf_b(Q_B^{CL}) dQ_B, \quad (8.20)$$

$$CL_{s+b} \equiv \int_{-\infty}^{Q_B^{CL,obs.}} pdf_{b+s}(Q_B^{CL}) dQ_B. \quad (8.21)$$

According to the CL_s method, the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ hypothesis that are less than 5% compatible with the measurement, i.e.

$$CL_s \equiv \frac{CL_{s+b}}{CL_b} < 0.05, \quad (8.22)$$

are excluded at 95% confidence level (Fig. 8.5). The $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ exclusion limits are given in Tab. 8.2.

⁸The 72 di-muon mass bins are defined by splitting in 8 BDT categories the $60\text{ MeV}/c^2$ search region around the B^0 mass, itself split into nine bins with boundaries $m_{B^0} \pm 18, 30, 36, 48, 60\text{ MeV}/c^2$.

Table 8.1: The total number of expected combinatorial background, $B_{(s)}^0 \rightarrow h^+ h^-$ background, and $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ candidates under the SM hypothesis, shown for the $B^0 \rightarrow \mu^+ \mu^-$ search regions (above), and $B_s^0 \rightarrow \mu^+ \mu^-$ search regions (below). Also, the observed number of candidates in the 3 fb^{-1} sample is shown for both regions.

Invariant mass [MeV/c ²]		BDT							
		0.0 – 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
$B_s^0 : m_{\mu^+\mu^-} \in [5224, 5344]$	Exp. comb. bkg	6520^{+119}_{-117}	$127.0^{+5.2}_{-5.0}$	$29.4^{+2.4}_{-2.3}$	$12.8^{+1.7}_{-1.5}$	$4.9^{+1.2}_{-1.1}$	$2.14^{+0.88}_{-0.77}$	$0.82^{+0.53}_{-0.37}$	$0.29^{+0.35}_{-0.18}$
	Exp. $B_{(s)}^0 \rightarrow h^+ h^-$ bkg	$1.97^{+0.64}_{-0.47}$	$1.25^{+0.41}_{-0.31}$	$0.83^{+0.27}_{-0.20}$	$0.88^{+0.29}_{-0.21}$	$0.83^{+0.28}_{-0.20}$	$0.77^{+0.25}_{-0.18}$	$0.66^{+0.21}_{-0.16}$	
	Exp. $B_s^0 \rightarrow \mu^+ \mu^-$ c-f.	$1.38^{+0.18}_{-0.16}$	$0.85^{+0.12}_{-0.11}$	$0.554^{+0.075}_{-0.067}$	$0.590^{+0.078}_{-0.070}$	$0.591^{+0.076}_{-0.070}$	$0.567^{+0.077}_{-0.069}$	$0.579^{+0.076}_{-0.069}$	$0.595^{+0.077}_{-0.069}$
	Exp. $B^0 \rightarrow \mu^+ \mu^-$ cand.	$0.99^{+0.12}_{-0.11}$	$0.610^{+0.081}_{-0.075}$	$0.398^{+0.049}_{-0.046}$	$0.424^{+0.050}_{-0.047}$	$0.425^{+0.050}_{-0.047}$	$0.408^{+0.050}_{-0.047}$	$0.416^{+0.049}_{-0.046}$	$0.428^{+0.050}_{-0.046}$
	Observed	6280	127	30	16	7	3	6	3
$B_s^0 : m_{\mu^+\mu^-} \in [5311, 5431]$	Exp. comb. bkg	6138^{+114}_{-112}	$121.6^{+4.8}_{-4.7}$	$28.2^{+2.2}_{-2.1}$	$11.9^{+1.5}_{-1.4}$	$4.77^{+1.11}_{-0.95}$	$2.17^{+0.79}_{-0.65}$	$0.79^{+0.48}_{-0.34}$	$0.29^{+0.32}_{-0.18}$
	Exp. $B_{(s)}^0 \rightarrow h^+ h^-$ bkg	$0.330^{+0.105}_{-0.089}$	$0.210^{+0.068}_{-0.058}$	$0.140^{+0.045}_{-0.038}$	$0.148^{+0.048}_{-0.040}$	$0.147^{+0.047}_{-0.040}$	$0.140^{+0.045}_{-0.038}$	$0.130^{+0.042}_{-0.035}$	$0.111^{+0.035}_{-0.030}$
	Exp. $B_s^0 \rightarrow \mu^+ \mu^-$ cand.	$8.77^{+0.9}_{-0.99}$	$5.40^{+0.75}_{-0.67}$	$3.52^{+0.46}_{-0.41}$	$3.75^{+0.47}_{-0.43}$	$3.76^{+0.47}_{-0.43}$	$3.61^{+0.46}_{-0.42}$	$3.68^{+0.46}_{-0.42}$	$3.79^{+0.46}_{-0.42}$
	Observed	5885	135	18	16	13	5	4	2

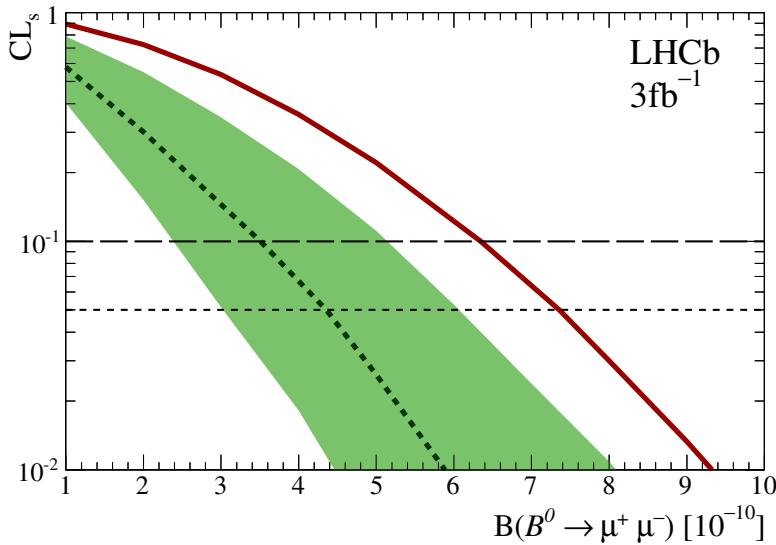


Figure 8.5: The CL_s values for different $B^0 \rightarrow \mu^+\mu^-$ branching fraction hypotheses. The solid red curve represents the compatibility between the different $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ hypotheses and the observed data; the dashed black curve illustrates the compatibility between the different $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ hypotheses and the simulated background-only samples, and the green band shows the central 68% confidence region of the CL_s values from the simulated background-only samples.

Normalising to the CL_b in Eq. (8.22) reduces the risk of excluding those signal branching fraction hypotheses for which the measurement has low sensitivity, i.e. for which the signal hypothesis compatibility (CL_{s+b}) is small enough to be excluded, but compatibility between the data and the background only hypothesis (CL_b) is large.

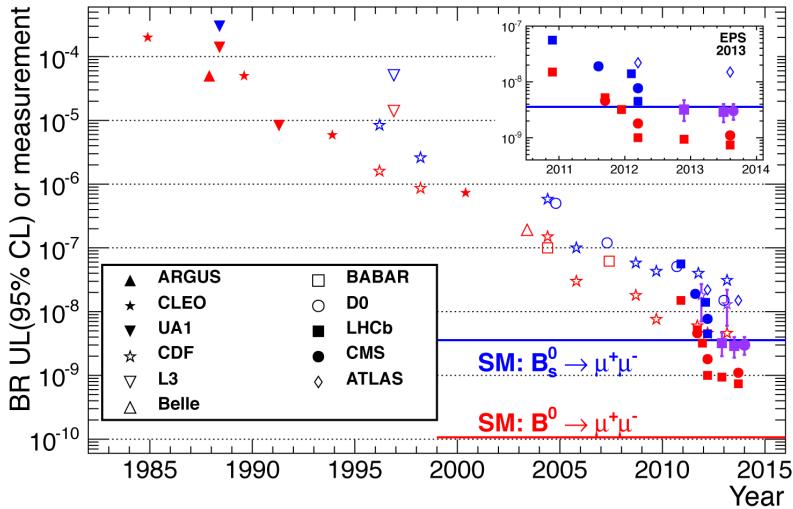
Table 8.2: Observed and expected $B^0 \rightarrow \mu^+\mu^-$ branching fraction exclusion limits at two different confidence levels.

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

Part III

Combined CMS and LHCb analysis

Introduction to Part III



The long history of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ searches. The blue and red marks denote the results of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ searches, respectively.

Over the course of the last 30 years, $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays have been searched for by more than ten experiments. The searches have been closing in on the branching fraction values expected in the Standard Model with an impressive average rate of an order of magnitude every five years. The first evidence of the $B_s^0 \rightarrow \mu^+\mu^-$ decay was reported by the LHCb experiment in 2013 [118], and confirmed by the latest LHCb results (this dissertation, and Ref. [58]) and by CMS [57]. The $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ values measured by CMS and LHCb are in good agreement. In the $B^0 \rightarrow \mu^+\mu^-$ channel, neither experiment has gathered enough statistics to refute the null hypothesis, as both report measurements at the significance level of 2σ .

This part of the dissertation describes the combination of LHCb and CMS $B_{(s)}^0 \rightarrow \mu^+\mu^-$ searches, starting with an overview of the CMS detector and the CMS analysis in Ch. 9; Chapter 10 describes the combination procedure that led to the world most precise $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching-fraction measurement up to date [43].

Chapter 9

Overview of the CMS $B^0_{(s)} \rightarrow \mu^+ \mu^-$ analysis

9.1 CMS detector

The CMS detector, described in detail in Ref. [119] and shown in Fig. 9.1, is designed to look for phenomena beyond the Standard Model (SM) searching for new heavy particles with masses ranging from $100\,\text{GeV}/c^2$ to a few TeV. If observed, these particles would be a direct manifestation of new physics. Since many new heavy particles could decay into b -quarks or muons in the final state, CMS has been designed with the capabilities that enable it also to detect the muonic decays of B^0 and B_s^0 mesons.

The CMS detector covers a very large range of angles and momenta to reconstruct high-mass states efficiently. To that extent, it employs a 13 m long superconducting solenoidal magnet with a 6 m diameter that is operated at a field of 3.8 T. The magnet is centred around the collision point with its field axis along the beam direction, and covers both hemispheres.

The silicon tracker is the first sub-detector that the particles from the pp collision traverse. A series of silicon tracking layers, consisting of silicon pixel detectors near the beam and silicon strips farther from the collision point, measures the tracks of charged particles emerging from the pp collisions. The tracking layers are organised in concentric cylinders around the beam, extending to a radius of 1.1 m, completed at each end by planar detectors (disks) perpendicular to the beam. Tracking coverage starts from the direction perpendicular to the beam and extends to within 220 mrad from it on both sides of the collision point. Coverage is the main difference between the two experiments: CMS has an overall larger acceptance, covering the pseudo-

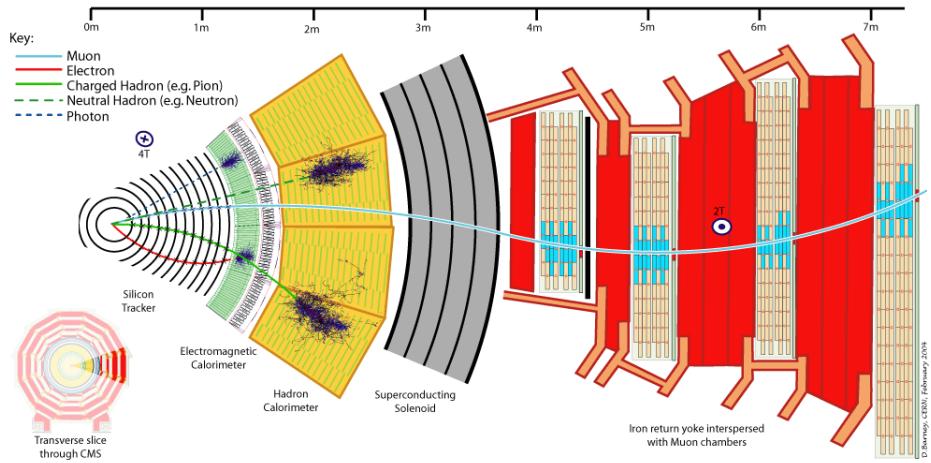


Figure 9.1: A transverse slice of the CMS detector, showing the tracks and energy depositions for a muon, electron, charged hadron, neutral hadron, and a photon.

rapidity region $|\eta| < 2.4$, whereas LHCb covers the extreme forward region, $1.9 < |\eta| < 4.9$ (see Fig. 9.2).

The impact parameter is measured with a precision of $10 \mu\text{m}$ for transverse momenta of $100 \text{ GeV}/c$, and $20 \mu\text{m}$ for $10 \text{ GeV}/c$. The typical muon transverse momenta from B meson decays in the CMS detector are in the range of $4\text{-}20 \text{ GeV}/c^2$. The momentum resolution, provided mainly by the silicon strips, changes with the angle with respect to the beam direction, resulting in an invariant mass resolution for $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ decays that varies from $32 \text{ MeV}/c^2$ for $B_{(s)}^0$ mesons produced perpendicularly to the proton beams to $75 \text{ MeV}/c^2$ for those produced in the direction along the beams.

After the tracking system, at a larger distance from the collision point, there is the electromagnetic calorimeter, followed by the hadron calorimeter. Muons are identified by their ability to penetrate the steel return yoke of the solenoid magnet and to produce signals in gaseous particle detectors interspersed within the steel - the muon detectors. Unlike LHCb, the CMS detector has practically no capability to discriminate between charged hadron species (pions, kaons, or protons), effective at the typical particle momenta in this analysis.

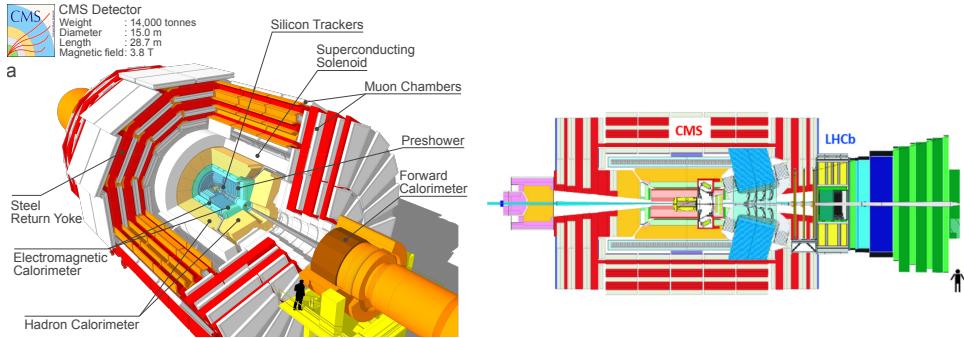


Figure 9.2: Exploded view of the CMS detector (left) and a cartoon sketch superimposing CMS and LHCb to show the CMS acceptance compared to that of the LHCb detector (right).

9.2 Trigger

Similarly to the LHCb trigger strategy (see Sec. 2.3), CMS implements a two-level trigger dedicated to selecting specifically events containing two muons. The hardware stage, based on information from the calorimeter and muon systems, is followed by a software stage, HLT, consisting of a large computing cluster that uses all the information from the detector. The software stage includes tracking in the final selection.

Since CMS is designed to look for much heavier objects than $B_{(s)}^0$ mesons, its trigger typically aims at events that contain muons with higher transverse momenta than those selected by the LHCb trigger. The CMS trigger records an event if (i) the di-muon invariant mass is between 4.8 and $6.0 \text{ GeV}/c^2$, (ii) both muons pass the momentum requirements, and (iii) the muon candidates originate from a common well-reconstructed vertex. The exact muon momentum thresholds depend on the collision energy and the direction of the muon track in the detector, but range between 3 and $4 \text{ GeV}/c$ for single-muon transverse momentum, and between 4 and $7 \text{ GeV}/c$ in the di-muon case¹. The higher transverse momentum requirement in CMS trigger lowers the trigger efficiency for $B_{(s)}^0$ decays with respect to that of LHCb, but allows CMS to record data at much higher proton-proton collision rate.

¹The LHCb trigger can recover muons with much lower transverse momenta, larger than $> 1.5 - 1.8 \text{ GeV}/c$, or $> 1.3 - 1.6 \text{ GeV}/c$ for the di-muon case; the exact p_T cut values depend on the collision energy (see Ch. 3).

9.3 Selection

The CMS analysis is performed on 5 fb^{-1} of data recorded at $\sqrt{s} = 7\text{ TeV}$ (2011), and 20 fb^{-1} of data recorded at $\sqrt{s} = 8\text{ TeV}$ (2012). The peak luminosity varied from 3.5×10^{30} to $7.7 \times 10^{33}\text{ cm}^{-2}\text{s}^{-1}$, on average leading to nine pp interactions per bunch-bunch crossing in the 2011 data sample, and twenty-one in the 2012 data sample. The large number of pp interaction per bunch-crossing can lead to a wrong estimation of the B meson decay-time and kinematic variables. Therefore, the candidates associated to badly reconstructed primary vertices are removed from the data sample.

The CMS analysis strategy is similar to that of the LHCb analysis (see Part II): the signal is searched for in the di-muon invariant mass spectrum, focussing on the region around the B -meson mass, and in categories of a multivariate classifier, “BDT”. Because of their relatively long lifetime, B -mesons on average travel a few cm before decaying. Therefore, the experimental signature for a signal consists of two isolated muons originating from a common displaced vertex, with the total di-muon momentum pointing away from the primary vertex. The offline reconstruction builds $B_{s,d}^0 \rightarrow \mu^+\mu^-$ candidates from two oppositely charged muons. Both muon candidates are required to have $p_T > 4\text{ GeV}/c$.

A BDT is constructed to further separate genuine muons from those arising from mis-identified charged hadrons. A low muon mis-identification probability is crucial for reducing the backgrounds. The muon identification BDT is based on four complementary classes of variables: basic kinematic quantities, silicon-tracker fit information, combined silicon and muon track-fit information, and muon detector information. The BDT classifier is trained on MC simulated samples of B -meson decays to kaons and muons. Compared to the offline reconstructed muon candidates, the muon selection based on the BDT reduces the mis-identification probability by 50% while retaining 90% of the true muons.

The probability to mis-identify a charged hadron as a muon, e.g. because of a decay in flight or detector punch-through, is measured with data from samples of well-identified pions, kaons, and protons. This probability is less than 0.13% for pions, 0.22% for kaons, and 0.15% for protons [57]: the exact mis-identification probabilities depend on the detector region, the running period, and the hadron momentum².

Since the background level and the mass resolution depend on the pseudo-

²For comparison, these hadron mis-identification probabilities in LHCb analysis are measured to be less than 0.5% for pions, 0.25% for kaons, and 0.5% for protons; the exact values depend on the detector region, hadron momentum, and running period (see Sec. 2.5).

rapidity of the B meson candidate, candidates are separated into two groups: the “barrel channel”, containing candidates where both muons have $|\eta| < 1.4$, and the “endcap channel”, containing those candidates for which at least one of the muons lie in the forward region, $|\eta| > 1.4$. Furthermore, these samples are separated into two sub-groups based on the collision energy.

The di-muon mass resolution in the different regions, as determined from simulated events, ranges from $32 \text{ MeV}/c^2$ for $\eta_{\mu\mu} \approx 0$ to $75 \text{ MeV}/c^2$ for $|\eta_{\mu\mu}| > 1.8$ ³. The analysis considered candidates with di-muon masses between 4.9 and $5.9 \text{ GeV}/c$.

The vertex and isolation properties are used to discriminate against combinatorial backgrounds, arising from combinations of muons from unrelated sources, most notably, from two separate $b \rightarrow c\mu\nu$ decays (e.g. $B \rightarrow D\mu\nu X$) or from one such decay in combination with a mis-identified hadron. Instead of applying thresholds to the selection variables, these are used to build a second BDT multivariate discriminator. Separate BDTs are constructed for the barrel and endcap channels in the 2011 and 2012 data. Each of the four BDTs is trained using signal candidates from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Monte Carlo simulation, and background candidates from the di-muon mass sideband data. The BDT discriminant output, a single number in the range $(0.0 - 1.0)$, is used to define together 12 BDT categories with different signal-to-background ratios. The candidates in the barrel and endcap channels are separated into 2 (resp. 4) categories for 2011 (resp. 2012) data, while the binning has been chosen such that it gives the same signal yield in each category.

9.4 Likelihood model

The di-muon invariant mass distribution in every BDT category is described by a model containing the signal components and the possible background components in the interesting mass range. The signal branching fractions are obtained from a maximum likelihood fit (see Ch. 8) to the di-muon invariant mass spectrum, simultaneously in the 12 BDT categories. In each category, the signal decays appear as peaks around the B_s^0 and B^0 masses, with their expected mass-distribution shape defined from Monte Carlo simulations. The calibration of the di-muon mass resolution is performed using the di-muon resonances analogously to what is done in the LHCb analysis, which also uses $B_{(s)}^0 \rightarrow h^+ h^-$ decays (Sec. 7.1). The central values for the signal mass distributions in the categories are taken from simulations, while they are measured on $B_{(s)}^0 \rightarrow h^+ h^-$ sample in LHCb.

³This is to be compared with the $23 \text{ MeV}/c^2$ invariant mass resolution of LHCb (see Sec. 6.3)

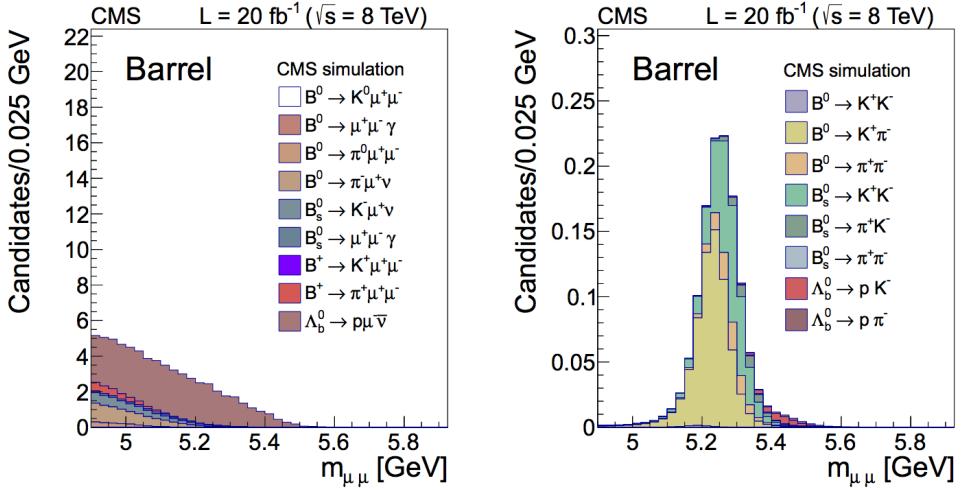


Figure 9.3: Processes contributing to the non-peaking (left) and peaking (right) backgrounds, caused by hadron mis-identification. The estimation has been made with simulated samples.

As discussed in Sec. 6.5, certain b -hadron decays can mimic the signal. In particular, the semi-leptonic decays $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, and $\Lambda_b^0 \rightarrow p \mu^- \nu$ can have reconstructed masses that are near the signal if one of the hadrons is mis-identified as a muon and combined with a genuine muon. Also, the two muons coming from the rare $B^{0/+} \rightarrow \pi^{0/+} \mu^+ \mu^-$ decays could fake the signal. When reconstructed as a di-muon final state, the invariant masses of these channels are mostly lower than those of the B^0 and B_s^0 , except for $\Lambda_b^0 \rightarrow p \mu^- \nu$, that can also populate higher-mass regions. Additionally, the doubly-mis-identified hadronic two-body decays $B_{(s)}^0 \rightarrow h^+ h'^-$, where $h^{(\prime)} = \pi^\pm$ or K^\pm , contribute to a di-muon invariant-mass region close to the B^0 mass value (see Sec. 6.4).

The signal distributions in di-muon mass and BDT discriminant are derived from the Monte Carlo simulated $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ samples with a detailed description of the detector response. The distributions for the mis-identified backgrounds are also obtained from Monte Carlo simulated samples (see Fig. 9.3)⁴. The combinatorial background has a monotonic dependence on the di-muon invariant mass, and can be obtained from the di-muon mass sidebands once the other background components have been accounted for.

⁴For comparison, the LHCb analysis extracts the contamination of doubly mis-identified $B_{(s)}^0 \rightarrow h^+ h^-$ decays from data (Sec. 6.4).

9.5 Normalisation and results

The CMS $B^0 \rightarrow \mu^+ \mu^-$ branching fraction is obtained normalising with respect to the number of $B^+ \rightarrow J/\psi K^+$ candidates⁵, using the branching fraction value $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (6.10 \pm 0.19) \times 10^{-5}$ [106]. The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction can be obtained analogously, accounting for the hadronisation ratio, f_s/f_d (see Sec. 2.1.1). CMS uses $f_s/f_d = 0.256 \pm 0.020$ from the LHCb measurement in Ref. [112]. The value of f_s/f_d could depend on the pseudo-rapidity and the kinematic range of the B mesons, and the CMS acceptance differs notably from that of LHCb. A study was performed using the ratio of $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ candidates, and an additional systematic uncertainty of 5% was assigned to the f_s/f_d to account for the extrapolation of the LHCb result to the CMS acceptance. The branching fraction results from the CMS analysis [57],

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9} \quad (4.3\sigma), \quad (9.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.5^{+2.1}_{-1.8}) \times 10^{-10} \quad (2.0\sigma), \quad (9.2)$$

are in good agreement with the LHCb results [58], described in Ch. 8.

⁵LHCb also uses the $B^0 \rightarrow K^+ \pi^-$ channel (Sec. 7.1)

Chapter 10

Combined CMS and LHCb analysis

This chapter describes how the results of the CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis (Ch. 9) and the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis (Part II) are combined [43]. The discussion begins with the motivation for a combined likelihood model. Section 10.1 explains the few modifications to the CMS and LHCb analysis models [57, 58] necessary to prepare them for the combination. The correlations between the two models are identified in Sec. 10.2. The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, their ratio, and their relative strengths with respect to the Standard Model expectations, are all obtained from a fit to the combined LHCb and CMS data sample; the results are presented in Sec. 10.3.

Two independent measurements of a Gaussian-distributed quantity may be averaged by weighting the results according to their uncertainties. The weighted average of the published CMS and LHCb results,

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.95 \pm 0.70) \times 10^{-9} , \quad (10.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.59 \pm 1.47) \times 10^{-10} , \quad (10.2)$$

treat the two measurements as independent, and assume symmetric uncertainties. Under the same assumptions, the combined significances may be estimated with [120]

$$\chi^2(\text{DOF} = 2k) = -2 \sum_{i=1}^k \ln(p_i), \quad (10.3)$$

to be 5.7σ and 2.6σ , for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, respectively.

The CMS and LHCb results, however, are not independent, and as may be inferred from the asymmetric uncertainties, nor are the measured branching

fraction *pdf*'s strictly Gaussian. The correct branching fractions and significances can be obtained by combining the two likelihood functions, and fitting the combined likelihood to the merged CMS and LHCb data sample. If the correlations between the two models are identified and accounted for, the combined likelihood can lead to better separation between alternative hypotheses, and thus to higher significances.

The CMS and LHCb likelihood models are both built using the `ROOFIT` framework [104], where the concept of workspace permits to share the models with ease, and simplifies the technical aspects of the combination procedure. Minor changes were necessary in order to make the original CMS and LHCb analysis models compatible with each other.

10.1 Changes to the published analysis models

The LHCb analysis model remains in a large part unchanged with respect to the most recent analysis [58]. The semi-leptonic background arising from mis-identified $\Lambda_b^0 \rightarrow p\mu^-\nu$ decay (previously only accounted for by the systematic uncertainty) was included in the combined LHCb model.

The $\Lambda_b^0 \rightarrow p\mu^-\nu$ background is normalised with respect to the $B^+ \rightarrow J/\psi K^+$ channel (Sec. 6.5), and requires a correction to the hadronisation fractions ratio that depends on the p_T of the Λ_b^0 baryon. Furthermore, the probability to reconstruct a proton and a muon from Λ_b^0 as a $B_s^0 \rightarrow \mu^+\mu^-$ candidate, depends on the p_T of the proton-muon pair from the Λ_b^0 . The p_T distribution of the proton-muon pair was simulated but turned out to differ from the theory predictions and had to be corrected for.

After this change, the LHCb results become:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)^{LHCb} = (2.7^{+1.1}_{-0.9}) \times 10^{-9}, \quad (10.4)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)^{LHCb} = (3.3^{+2.4}_{-2.1}) \times 10^{-10}. \quad (10.5)$$

The relative change in the central values with respect to the published results amounts to -6% (resp. -11%) for $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ (resp. $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$).

The CMS analysis model also underwent a few changes. The treatment of the $\Lambda_b^0 \rightarrow p\mu^-\nu$ background was carried out similarly to the LHCb analysis. Moreover, it was realised that as described in Sec. 1.2, the measured time-integrated and CP averaged branching fractions depend uniquely on the decay-time distribution of the channel, and therefore introduce corrections to the experimental decay-time acceptance. In the original CMS analysis, the signal selection is performed using variables correlated to the decay-time (e.g. IP of the candidate tracks), but the time-acceptance effects were not corrected

for. For the combination, the time-dependent signal efficiency was obtained from Monte Carlo, and the corrections were included into the model.

After the changes to the published analysis, CMS results become:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{CMS} = (2.8^{+1.0}_{-0.9}) \times 10^{-9}, \quad (10.6)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{CMS} = (4.4^{+2.2}_{-1.9}) \times 10^{-10}. \quad (10.7)$$

The relative change in the central values with respect to the published results amounts to -6% (resp. $+26\%$) for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (resp. $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$).

The weighted averages of the LHCb and CMS results in Eqs. (10.4)-(10.7) are

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.95 \pm 0.70) \times 10^{-9} \quad (5.9\sigma), \quad (10.8)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.59 \pm 1.47) \times 10^{-10} \quad (2.9\sigma). \quad (10.9)$$

10.2 Shared parameters and combined model.

By default, all the CMS and LHCb model parameters are kept independent in the combined likelihood. However, some parameters are common to both models; this arises from the description of the same physical background sources and from the use of the same normalisation channel, $B^+ \rightarrow J/\psi K^+$.

In the combined model, the branching fractions of the signal channels, the branching fraction of the normalisation channel, $\mathcal{B}(B^+ \rightarrow J/\psi K^+)$, and the hadronisation fraction ratio, f_s/f_d , are shared (see Tab. 10.1). The additional systematic uncertainty (5%) assigned to f_d/f_s in the CMS likelihood (see Sec. 9.5) remains.

The estimated background yields in the final data sample are calculated by normalising with respect to the $B^+ \rightarrow J/\psi K^+$ channel either measured or predicted branching fraction values¹. In principle, all the background-channel branching fractions used by the models are common and should be shared. However, this could not have been done without modifying the analysis significantly (in the CMS semi-leptonic backgrounds description, these are treated as a whole and not expressed as separate model parameters). Instead of sharing, these branching fraction values were synchronised; the values used by CMS were updated to the most recent measurements or predictions (see Tab. 10.2).

¹Except the doubly mis-identified component yield in LHCb analysis, which is evaluated from the measured $B_{(s)}^0 \rightarrow h^+ h^-$ sample (see Sec. 6.4)

Table 10.1: Common shared parameters in the combined CMS and LHCb likelihood model, together with the values and the uncertainties used as constraints in the fit.

Parameter	Value	Reference
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	floating	-
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	floating	-
$\mathcal{B}(B^+ \rightarrow J/\psi K^+)$	$(5.93 \pm 0.06) \cdot 10^{-2}$	Ref. [106]
$\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$	$(1.028 \pm 0.031) \cdot 10^{-3}$	Ref. [106]
$\mathcal{B}(B^+ \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) K^+)$	$(6.10 \pm 0.19) \cdot 10^{-5}$	
f_d/f_s	3.86 ± 0.22	Refs. [82, 112, 113]

Table 10.2: Parameters that are not shared, but have common values in the combined model, together with the latest values and uncertainties used as constraints in the fit.

Parameter	Value	Reference
$\mathcal{B}(B^0 \rightarrow K^+ \pi^-)$	$(1.96 \pm 0.05) \cdot 10^{-5}$	Ref. [106]
$\mathcal{B}(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)$	$(1.27 \pm 0.59) \cdot 10^{-4}$	Ref. [121]
$\mathcal{B}(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)$	$(1.44 \pm 0.05) \cdot 10^{-4}$	Ref. [106]
$\mathcal{B}(\Lambda_b^0 \rightarrow p \mu^- \nu)$	$(4.94 \pm 2.19) \cdot 10^{-4}$	Ref. [122] and updates
$\mathcal{B}(B^{0/+} \rightarrow \pi^{0/+} \mu^+ \mu^-)$	$(1.47 \pm 0.20) \times (2.3 \pm 0.6) \cdot 10^{-8} = (3.38 \pm 0.99) \cdot 10^{-8}$	Refs. [121, 123]

10.3 Results of the combined analysis

The combined results are obtained from a Maximum Likelihood fit with the combined likelihood model. The fit is performed on the di-muon mass plane, simultaneously in eight LHCb and twelve CMS BDT categories, as shown in Fig. 10.1.

The combined fit is also shown on a single plot (see Fig. 10.3), after weighting the contributions from each category according to the expected $S/(S+B)$ ratio for $B_s^0 \rightarrow \mu^+ \mu^-$ in the SM. The weights are estimated by the ratio of the signal *pdf* value (S) and the complete *pdf* ($S+B$), both evaluated at the B_s^0 mass. This was shown to be equivalent to using the ratio of the corresponding *pdf* integrals in the range $[m_{B_s^0} - \sigma_{B_s^0}, m_{B_s^0} + \sigma_{B_s^0}]$ (the weights also take into account the signal mass resolution). The weights for the twenty CMS and LHCb categories are given in Fig. 10.2.

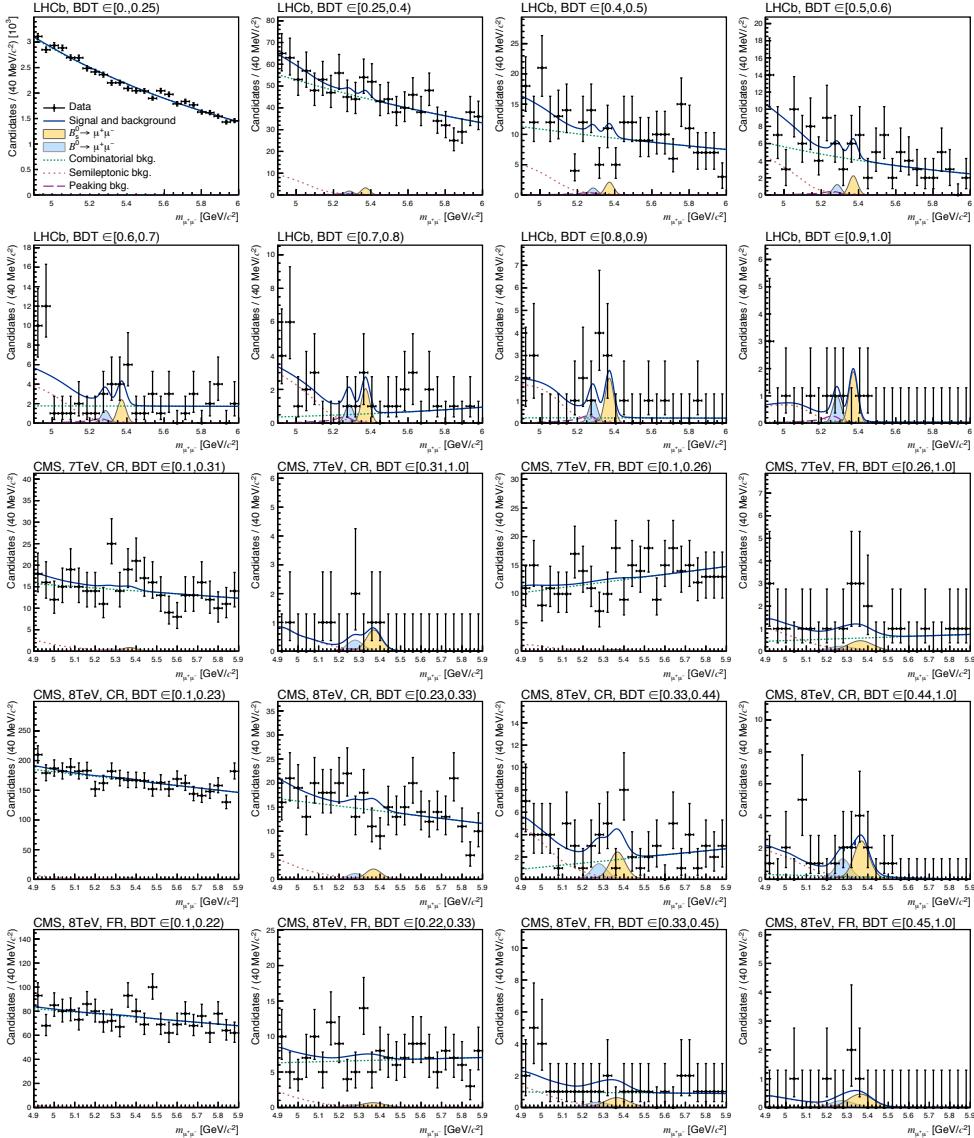


Figure 10.1: Di-muon invariant mass in each of the 20 categories. Superimposed on the data points (black dots) are lines denoting the combined fit (solid blue) and its components: the B_s^0 (yellow shaded) and B^0 (light-blue shaded) signal components; the combinatorial background (dash-dotted green); the sum of the semi-leptonic backgrounds (dotted pink); and the peaking backgrounds (dashed violet). Categories for which both muons are detected in the central region of the CMS detector (barrel channel) are denoted with “CR”, those for which at least one muon was detected into the forward region (end-cap channel) with “FR”.

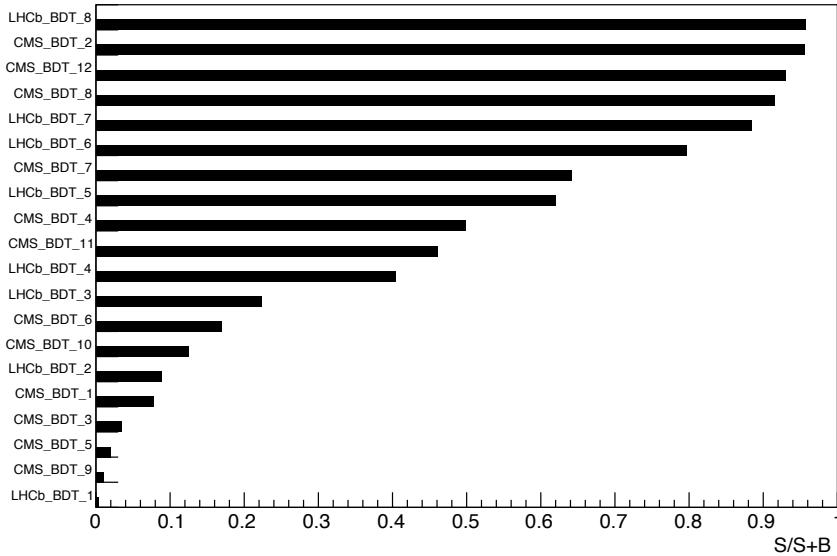


Figure 10.2: The $S/(S+B)$ weights, assigned to the eight LHCb and twelve CMS BDT categories. The weights are calculated using the $B_s^0 \rightarrow \mu^+ \mu^-$ signal *pdf* with Standard Model $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ prediction (S), and the complete model *pdf* (S+B), both evaluated at the B_s^0 meson mass. These weights are then used to represent the combined analysis model in a single di-muon mass plot (Fig. 10.3).

Three main results are extracted from the combined fit: the combined $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions and their significances, the branching fractions relative to the SM predictions, and the ratio between the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction.

10.3.1 Branching fractions

The combined fit leads to the following branching fraction measurements

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{LHC} = (2.78^{+0.66}_{-0.60}(\text{stat})^{+0.27}_{-0.18}(\text{syst})) \times 10^{-9} \quad (6.2\sigma), \quad (10.10)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{LHC} = (3.94^{+1.58}_{-1.41}(\text{stat})^{+0.31}_{-0.24}(\text{syst})) \times 10^{-10} \quad (3.2\sigma). \quad (10.11)$$

The number of signal and background candidates listed in Table 10.3.

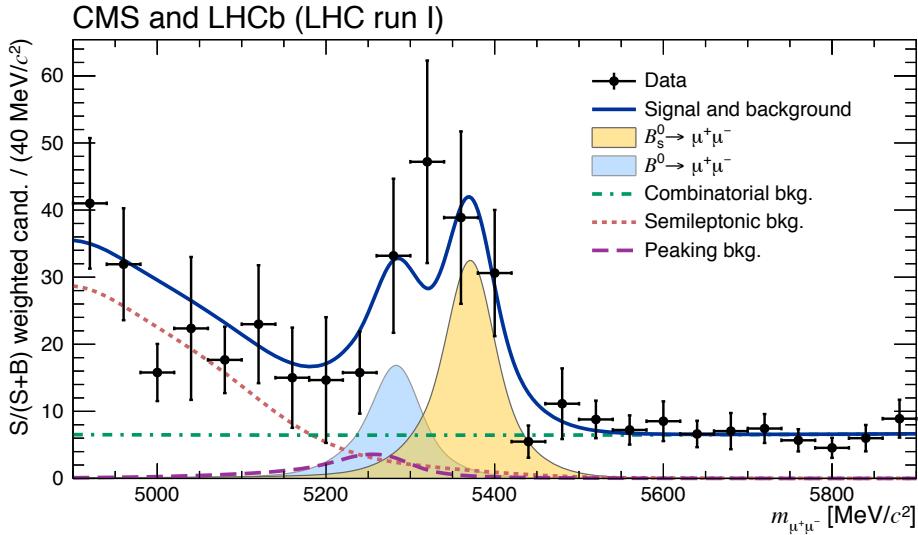


Figure 10.3: Di-muon invariant mass for the eight LHCb and twelve CMS BDT categories. The distribution has been obtained using the weights in Fig. 10.2. Superimposed on the data points (black dots) are the combined fit (solid blue line) and its components: the B_s^0 (yellow shaded peak) and B^0 (light-blue shaded peak) signal components; the combinatorial background (dash-dotted green line); the sum of the semi-leptonic backgrounds (dotted pink line); and the peaking backgrounds (dashed violet line). The data point uncertainties are calculated as $\sigma = \sqrt{\sum_{cat}^{20} w_{cat}^2 N_{cat}}$.

Table 10.3: The number of observed candidates in the combined CMS and LHCb analysis.

Model component	Observed candidates
$N(B_s^0 \rightarrow \mu^+ \mu^-)$	73 ± 4
$N(B^0 \rightarrow \mu^+ \mu^-)$	39 ± 4
$N(hh \rightarrow \mu\mu)$	19 ± 1
$N(h\mu\nu \rightarrow \mu\mu)$	446 ± 23
$N(Comb.\mu\mu)$	62668 ± 252

The uncertainties in the branching fractions include both statistical and systematic sources, the latter contributing 35% and 18% to the total uncertainty of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions, respectively. The significances in Eqs. (10.10) and (10.11) have been estimated using the

Wilks' theorem [117], as explained in Sec. 8.3. The median expected significances, assuming the SM branching fractions, are 7.4σ and 0.8σ for the B_s^0 and B^0 mode, respectively. Likelihood contours for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ versus $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ are shown in Fig. 10.4. One-dimensional likelihood scans for both decay modes are also displayed in the same figure.

For the first time, the $B_s^0 \rightarrow \mu^+ \mu^-$ decay has been observed, with a strong significance of 6.2σ . The observed significance for $B^0 \rightarrow \mu^+ \mu^-$ (3.2σ) is above the accepted evidence “threshold” of 3.0σ . The Wilks' theorem used to estimate this significance involves making asymptotic assumptions about the distribution of the test statistic (i.e. profile likelihood ratio) under the null-hypothesis. In order to verify the validity of these assumptions, the statistical significance and confidence intervals for the B^0 branching fractions are also determined using toy experiments. The toy experiments, simulated with $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = 0$, provide us with a realistic (independent of asymptotic assumptions) profile-likelihood ratio distribution under the null hypothesis. Using the ordering principle proposed by Feldman and Cousins (FC) [124], the p-value was calculated from the profile-likelihood ratio, and it was found to correspond to the significance of 3.0σ (see Fig. 10.5).

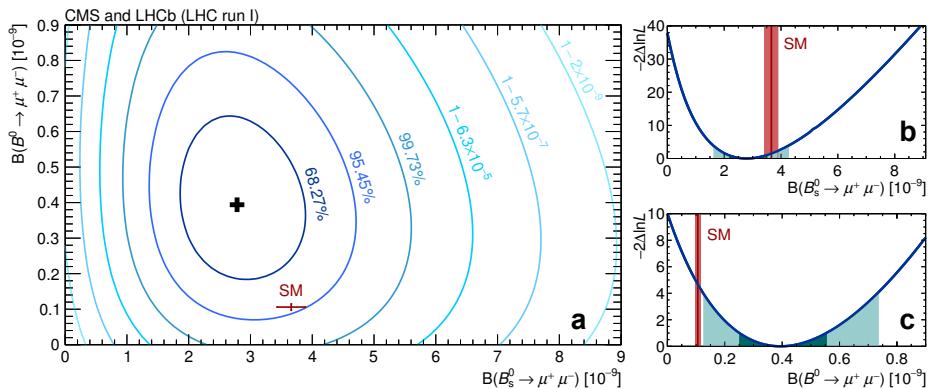


Figure 10.4: Likelihood contours in the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ versus $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ plane. The (black) cross in panel (a) denotes the best-fit value. The Standard Model expectation is denoted by the marker (red, SM). Each contour encloses a region approximately corresponding to the indicated confidence level. Variations of $-2\Delta \ln L$ for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ are shown in panels (b) and (c), respectively: the dark and light (cyan) areas define the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals for the branching fraction, respectively; the SM prediction for each branching fraction is denoted by the vertical (red) band.

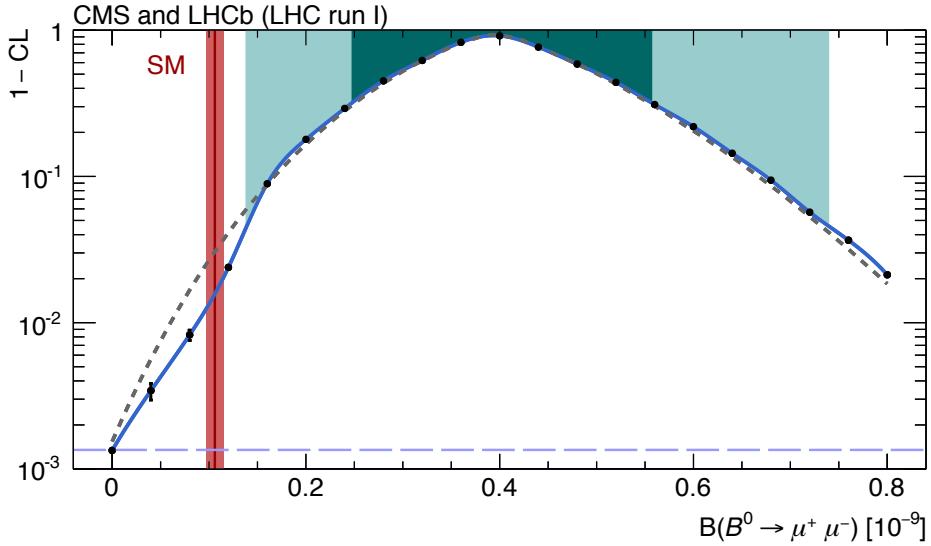


Figure 10.5: Confidence level as a function of the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ hypothesis. CL is the confidence level obtained with the Feldman–Cousins procedure [124]. The points denote the $1 - CL$ values found with the toy experiments, and the curve is their (spline) interpolation. The dark and light (cyan) areas denote the two-sided $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals, while the dashed horizontal line defines the confidence level for the 3σ one-sided interval. The dashed (grey) curve shows the $1 - CL$ values computed using Wilks’ theorem [117]. Differences between these confidence-level values and those from the FC procedure illustrate the degree of approximation implied by the asymptotic assumptions inherent to Wilks’ theorem [117].

10.4 Comparison to theory

The signal strength, defined as the ratio between the observed branching fraction and the one predicted by the SM, illustrates the compatibility between the observed results and the SM predictions in Eqs. (1.49) and (1.50). The SM predictions are introduced into the model as nuisance parameters, and used to define the signal strengths. The fitted signal strengths are

$$\mathcal{S}_{\text{SM}}^{B_s^0} = 0.76^{+0.20}_{-0.18} \quad , \quad \mathcal{S}_{\text{SM}}^{B^0} = 3.7^{+1.6}_{-1.4}. \quad (10.12)$$

The signal strengths in Eq. (10.12) show that the measurements are compatible with the SM predictions, at the 1.2σ and 2.2σ level for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, respectively.

As the last step, the fit model is adapted to fit directly to the ratio of branching fractions. In the case of the ratio common normalisation para-

meters will cancel out: e.g. the detection efficiencies, which are very similar (within 1%) and were taken equal in the normalisation. The hadronisation fraction ratio, f_s/f_d , will still be needed. In summary:

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{f_s}{f_d} \times \frac{N_{B_d}}{N_{B_s}}. \quad (10.13)$$

The ratio measured in the combined analysis,

$$\mathcal{R} = 0.14^{+0.08}_{-0.06}, \quad (10.14)$$

is compatible with the SM prediction (see Eq. (1.51)) at the 2.3σ level. The one-dimensional likelihood scan for the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction ratio is shown in Fig. 10.6.

The latest combined CMS and LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction results are shown together with various New Physics models in Fig. 10.7. The impact of the latest $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ results is clearly seen when comparing Fig. 10.7 with the analogous plot displaying the regions experimentally excluded by previous data (see Fig. 1.7). The measurements are compatible with the SM predictions, and large enhancements to the decay rates are excluded.

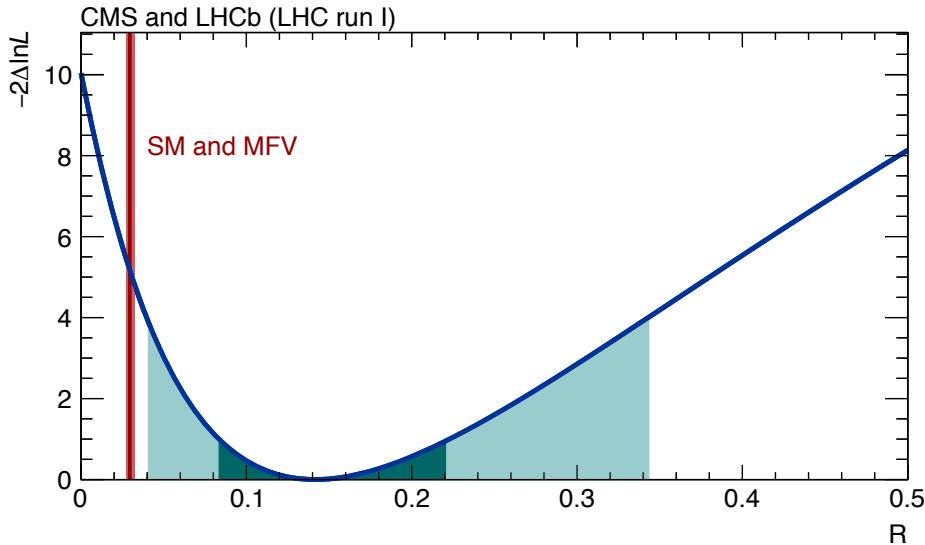


Figure 10.6: Variation of $-2\Delta\ln L$ as a function of the ratio of branching fractions $\mathcal{R} \equiv \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)/\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The dark and light (cyan) areas denote the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals for \mathcal{R} , respectively. The value of \mathcal{R} predicted in the SM is denoted with the vertical (red) band.

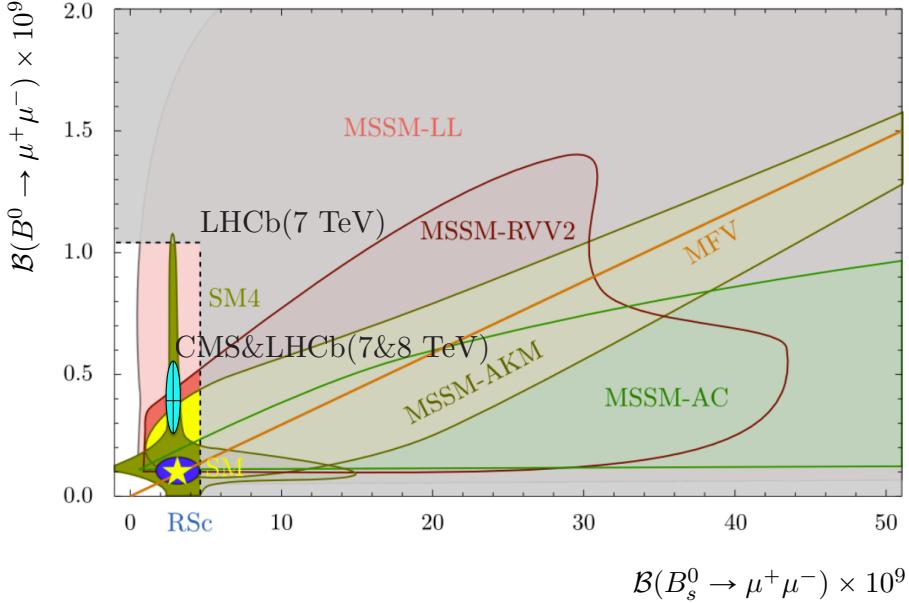


Figure 10.7: Possible values of the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions on the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ plane in different New Physics models. The Standard Model prediction is shown as a star. The shaded grey area denotes the region that has been excluded by the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis with 2011 data [125]; the light blue ellipse denotes the latest combined CMS and LHCb measurement [43].

In Sec. 1.4, several (limiting) New Physics scenarios were discussed. They may lead to deviations from the SM which can be expressed through the signal strength

$$\mathcal{S}_{\text{SM}}^{B_s^0} = \left[\frac{1 + y_{(d,s)} \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}}{1 + y_{(d,s)}} \right] \times (|P|^2 + |S|^2), \quad (10.15)$$

namely the $S = 0$, $P = 1$, and $S \pm P = 1$ scenarios [28, 29, 126].

The $P = 1$ scenario is realised in 2HDM models with a dominating scalar Higgs (H^0). Possible tree-level scalar exchange contributions (Fig. 1.8b) to the $B_s^0 \rightarrow \mu^+ \mu^-$ have been studied considering a left-handed scheme, $M(H^0) = 1$ TeV, quark-scalar couplings allowed by the constraints from B_s^0 mixing, and muon-scalar couplings allowed by the recent $B_s^0 \rightarrow \mu^+ \mu^-$ measurements (see

Sec. 6.4.3 in Ref. [29]). Contributions of this type look the least favourable in the light of the latest measurements (see Fig. 10.8).

The models where both scalar (H^0) and pseudo-scalar (A^0) Higgses contribute can belong to the $S \pm P = 1$ scenario if either left- or right-handed couplings dominate. This is realised in 2HDMs in a decoupling regime, where $M(H^0) \simeq M(A^0) \gg M(h)$, and the Minimal Flavor Violation (MFV) is retained to suppress the $C_{S,P}^L$ coefficients (analogously to the SM): $C_{S,P}^L/C_{S,P}^R \sim m_{s,b}/m_b$. Possible tree-level contributions have been studied with H^0 and A^0 masses equal to 1 TeV, with the quark coupling values allowed by the B_s^0 mixing constraints in the left-handed scheme, and with various muon coupling values (Fig. 10.8).

The models without scalar contributions, but with additional gauge bosons or pseudo-scalar exchanges, contribute to the $S = 0$ scenario. Models with an additional heavy neutral gauge boson, Z' , are a good example, and could lead to tree-level FCNC.

A tree-level Z' exchange (Fig. 1.8) with $M(Z') = 1$ TeV, quark couplings constrained by the B_s^0 mixing in the left-handed scheme, and muon couplings in the range $[0.3 - 0.7]$, leads to predictions that in majority of the cases are compatible with observed $B_s^0 \rightarrow \mu^+ \mu^-$ signal strength region, as illustrated in Fig. 10.8. A detailed analysis of Z' models can be found in Ref. [54]. Note that also 2HDMs with pseudo-scalar (A^0) dominance may contribute to the $S = 0$ scenario. However, the predicted signal strength values are mostly higher or lower than the observed signal strength (see Fig. 10.8).

One should notice that the signal strength parameter alone does not allow to separate P and S contributions (see e.g. Eq. (10.15) and Fig. 1.9). Moreover, sizeable NP contributions may still be present even if the signal strength is measured close to the SM value $\mathcal{S}_{\text{SM}}^{B_s^0} = 1$. The situation can be considerably improved by measuring the mass-eigenstate asymmetry, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. Depending on the result, a measurement of $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ could distinguish between scalar, pseudo-scalar, and gauge boson exchange in $B_s^0 \rightarrow \mu^+ \mu^-$ [27, 28]. In the case of $B^0 \rightarrow \mu^+ \mu^-$, the negligible lifetime asymmetry does not allow NP contributions through $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$.

Concerning supersymmetric models, MSSM includes two Higgs doublets and can contribute to the three discussed $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ NP scenarios. The measured $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions can thus be used to constrain the allowed SUSY parameters space. In testing the validity of the various MSSM models, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions provide information complementary to the direct ATLAS and CMS SUSY-searches, the Higgs mass measurements, and $H^0/A^0 \rightarrow \tau^+ \tau^-$ searches; especially in the high

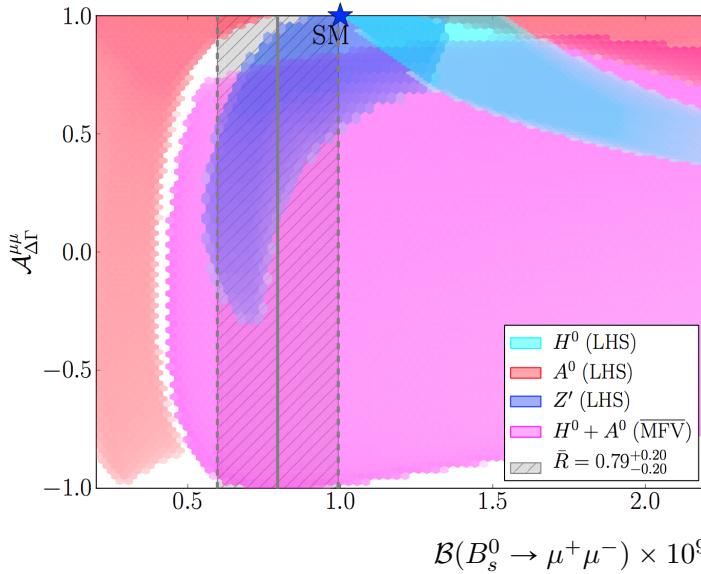


Figure 10.8: The mass-eigenstate asymmetry $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ versus the $B_s^0 \rightarrow \mu^+\mu^-$ signal strength in different tree-level NP models: a gauge boson exchange (Z'), a scalar exchange (H^0), a pseudo-scalar (A^0) exchange, and a mixed scalar-pseudo-scalar exchange. The models here include left-handed quark couplings, with values compatible with the B_s^0 mixing constraints, and various muon couplings. All particle masses are set to 1 TeV. The striped grey band denotes the $B_s^0 \rightarrow \mu^+\mu^-$ signal strength from CMS and LHCb, and the star the Standard Model values. The figure is taken from Ref. [29].

$\tan\beta$ regions (Sec. 1.4) with $\tan\beta > 50$ [25,127]). The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions are among the most relevant observables in combined MSSM fits, e.g. in [128].

Summary

Chapter 11

Summary and outlook

Modern physics successfully describes the structure and the processes of the microscopic world. According to the established theory of the microscopic world, The Standard Model (SM), the ordinary matter in the universe consists of six quarks and six leptons, and its dynamics is dictated by the forces between these constituents. In the SM, these forces are described by an exchange of force carriers, called bosons.

Even though successful in general, the SM does fall short in answering some important questions: little is known about the matter that constitutes the majority of the mass of the universe or why we live in a matter rather than anti-matter dominated universe. In search of these answers, physicists have come up with many clever solutions. The successful SM could be extended by postulating yet unseen New Physics (NP). For instance, these NP models can be built by postulating supersymmetric partners of the observed particles (SUSY), or additional Higgs-boson-like particles.

The NP models must be compatible with the verified SM predictions, while they may predict new measurable but yet unverified effects. Particle physics experiments look further than the already verified SM domain in various ways: reaching for higher collision energies, devising new more precise measurements of already observed processes, or searching for yet unobserved processes; in all cases, the effects in which the NP predictions can differ from the SM ones are of interest. The rare B_s^0 and B^0 meson decays into two muons (i.e. $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in short) are a good example of a search for yet unobserved processes.

The results from the Large Hadron Collider (LHC) experiments at CERN have taken us closer to solving the big open problems. Perhaps the most memorable results of 2011 and 2012, the so called “Run 1”, were the discovery of the Higgs boson [4, 5] and the first evidence of B_s^0 meson decays to two

muons ($B_s^0 \rightarrow \mu^+ \mu^-$) [118], which were both long searched for. At the same time, several hints of discrepancy with the SM hold promises of new exciting results in the years to come.

The history of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ searches over 30 years is shown in Fig. 11.1. This dissertation describes the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ searches at the LHC. It explains why $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are interesting from the theoretical perspective, how these decays were looked for in the LHCb experiment and what was found [58], and how the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis results from the LHCb experiment were combined with the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis results from another experiment, CMS. The parts to which the author has contributed personally are discussed in greater detail. In particular, these include improving the trigger efficiency estimation (Ch. 5), constructing the LHCb likelihood model (Ch. 6), normalising the LHCb signal yields (Ch. 7), fitting it to the data to obtain the results (Ch. 8), combining the LHCb and CMS likelihood models and obtaining the combined results (Ch. 10). As is shown in Sec. 10.3, the combination of the results leads to the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay and the first evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay [43].

The rate at which the decays proceed is expressed through the branching fraction (\mathcal{B}). The $B_s^0 \rightarrow \mu^+ \mu^-$ (or $B^0 \rightarrow \mu^+ \mu^-$) branching fraction into a dimuon final state is the fraction of $B_s^0 \rightarrow \mu^+ \mu^-$ (or $B^0 \rightarrow \mu^+ \mu^-$) decays with respect to all the possible B_s^0 (or B^0) decays. The SM processes contributing to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are well understood. In the SM, $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are suppressed by the GIM mechanism and proceed through higher order loop diagrams (see Ch. 1 and Fig. 1.3). The SM B_s^0 and B^0 meson branching fractions to two muons are precisely predicted [41, 43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.66 \pm 0.23) \times 10^{-9}, \quad (11.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}. \quad (11.2)$$

With these very low expected SM rates and loop processes dominating the transition, it is likely that NP, especially in models with extended Higgs sectors or additional bosons, significantly alters the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay probability.

The main results of the LHC $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ searches are the measurements of the B_s^0 and B^0 meson branching fractions [43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{LHC} = (2.78^{+0.66}_{-0.60}(\text{stat})^{+0.27}_{-0.18}(\text{syst})) \times 10^{-9} \quad (6.2\sigma), \quad (11.3)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{LHC} = (3.94^{+1.58}_{-1.41}(\text{stat})^{+0.31}_{-0.24}(\text{syst})) \times 10^{-10} \quad (3.2\sigma). \quad (11.4)$$

The branching fractions and the significances are extracted from the combined CMS and LHCb likelihood model, shown together with the data in

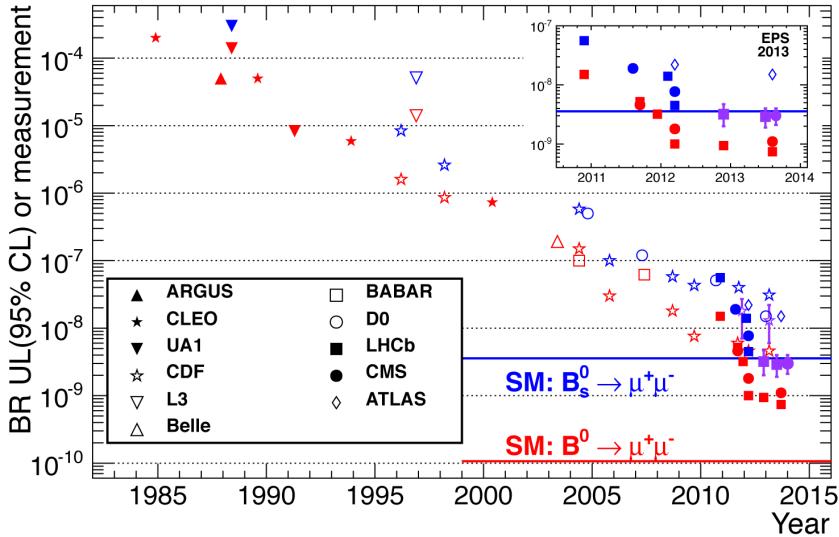


Figure 11.1: The long history of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ searches. The blue and red marks denote the results of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ searches, respectively. Upper limits, set by various searches throughout the years, are shown together with the latest measurements by CMS and LHCb.

Fig. 11.2. The main contribution to the uncertainties is of statistical nature and can be reduced in the coming years. Compared to the total uncertainty, the systematic uncertainty amounts to 35% and 18% in the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ branching fractions, respectively. It arises from the signal normalisation, the mis-identified background yield estimation, and the di-muon mass model. The uncertainty in the measured hadronisation fraction ratio (f_s/f_d) from Ref. [113] is the dominant systematic uncertainty.

The impact of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ measurements can be seen by comparing the excluded parameter space of various NP models before and after the LHC Run 1 results (see Fig. 11.3): with the results from LHC, a large part of the parameter space is excluded. Given the still large uncertainty, the current measurement is compatible with the SM predictions. However, the NP models tend to have complex parameter spectra and they can seldom be excluded by the measurement of a single observable. Sizeable NP contributions can still be present in other observables even if the measured $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction is close to the SM value. Therefore, future efforts have to be directed on one hand to improve the precision of the branching fraction measurements and on the other to measure additional observables.

The observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay and the evidence of the $B^0 \rightarrow \mu^+\mu^-$ decay are the culmination of 30 years of experimental $B_{(s)}^0 \rightarrow \mu^+\mu^-$

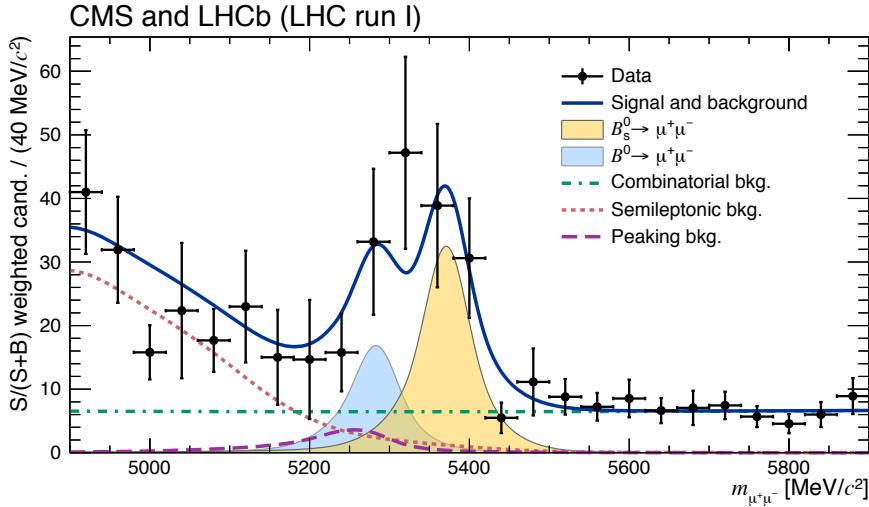


Figure 11.2: Di-muon invariant mass distribution of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates in the combined CMS and LHCb analysis. Superimposed on the data points (black dots) are the combined fit (solid blue line) and its components: the B_s^0 (yellow shaded peak) and B^0 (light-blue shaded peak) signal components; the combinatorial background (dash-dotted green line); the sum of the semi-leptonic backgrounds (dotted pink line); and the peaking backgrounds (dashed violet line).

searches, and open the possibility of precise measurements of these channels. As pointed out in Ref. [129], the $B_s^0 \rightarrow \mu^+\mu^-$ decay is a theoretically clean probe for the Wilson coefficient C_{10}^R . A more precise determination of this coefficient is mandatory in the light of several tensions in other decays involving $b \rightarrow s$ transitions, such as $B^0 \rightarrow K^{*0}\mu^+\mu^-$ [130], and $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$ [131]; more stringent constraints on C_{10}^R are necessary in order to identify NP contributing to other Wilson coefficients, such as C_9^R . Moreover, the current best $B^0 \rightarrow \mu^+\mu^-$ branching fraction measurement is more than three times higher than what the SM predicts. This excess needs to be investigated and measured with a better precision.

More precise measurements will already be possible in the LHC Run 2. At the time of writing, the LHCb has recorded 0.32 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ (equivalent to 0.5 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$). The expected uncertainties for the future LHCb analysis are shown in Fig. 11.4. Two possible near-future scenarios for the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ studies in LHCb are shown in Fig. 11.5. Assuming the current analysis sensitivity, detector performance, and the measured $B^0 \rightarrow \mu^+\mu^-$ branching fraction, the LHCb experiment has

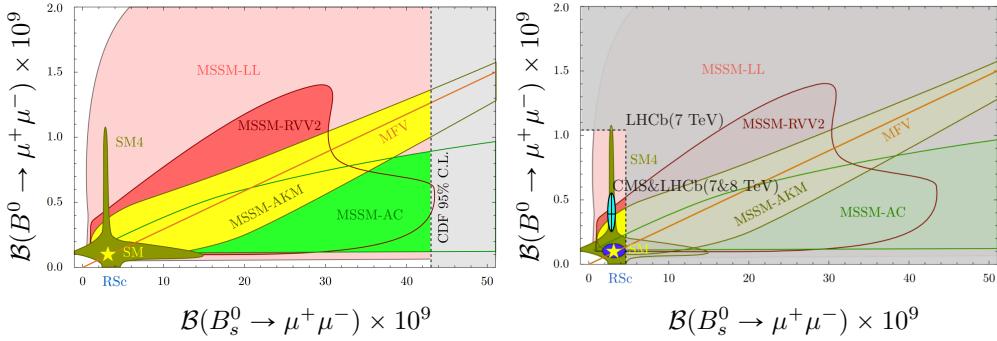


Figure 11.3: Allowed $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions values on the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ plane in different New Physics models. The Standard Model prediction is shown as a star. The shaded grey area on the left figure denotes the region experimentally excluded before the LHC results, the shaded grey area on the right figure denotes the region experimentally excluded by the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis with 2011 data [125]; the light blue ellipse on the right figure denotes the latest combined CMS and LHCb measurement [43].

a good chance to measure an evidence for a non-SM $B^0 \rightarrow \mu^+ \mu^-$ branching fraction value! The possibility of an NP enhanced $B^0 \rightarrow \mu^+ \mu^-$ branching fraction makes the $B^0 \rightarrow \mu^+ \mu^-$ measurement the most awaited result of the next LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.

The impact of the $B_s^0 \rightarrow \mu^+ \mu^-$ measurements has also been studied in models with tree-level contributions from different types of new particles [28, 29]. Such contributions could arise from new heavy vector bosons, as predicted in various Z' and Little Higgs models, or from new scalar or pseudo-scalar particles, as predicted in different types of two-Higgs doublet (2HDM) models (see Sec. 1.4, Fig. 1.8).

The current situation can be considerably improved by measuring an additional observable, the mass-eigenstate asymmetry in $B_s^0 \rightarrow \mu^+ \mu^-$, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. At this moment, the size of the still allowed parameter space suggests that the pseudo-scalar Higgs (A^0) and scalar Higgs (H^0) dominated 2HDMs are less favoured by the measurement, and that models with new heavy gauge bosons, such as Z' , are less constrained by the measurement. Together with a more precise $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction measurement, the $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ measurement will be an important future step in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses. As suggested in Ref. [27], $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ can be extracted from the lifetime distribution

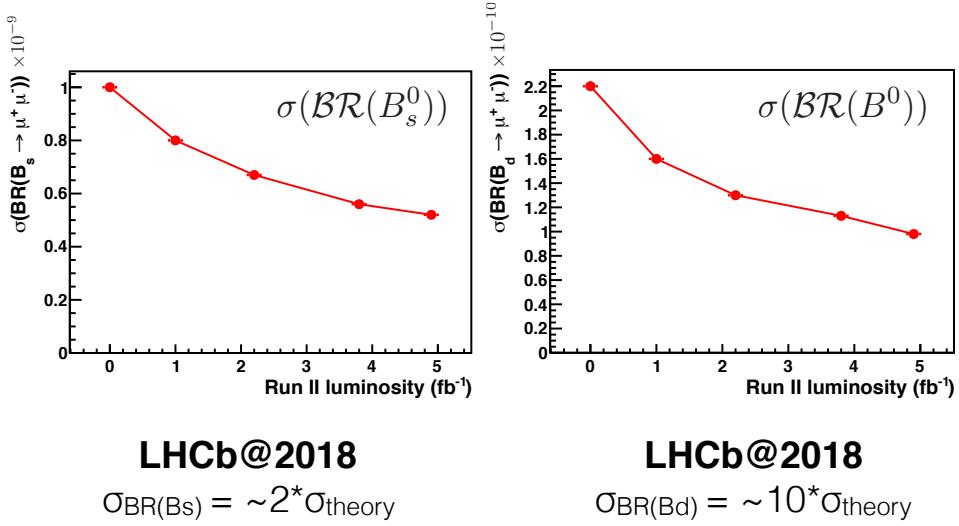


Figure 11.4: The expected uncertainties in the $B_s^0 \rightarrow \mu^+ \mu^-$ (left) and the $B^0 \rightarrow \mu^+ \mu^-$ (right) branching fractions, shown as a function of the pp collision data expected in the LHCb Run 2. The studies assume SM branching fraction values, and Run 1 analysis sensitivity and detector performance.

of the $B_s^0 \rightarrow \mu^+ \mu^-$ decays. Depending on the result, a measurement of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ could distinguish between scalar, pseudo-scalar, and gauge boson exchange in $B_s^0 \rightarrow \mu^+ \mu^-$ [27, 28].

In summary, the results of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements presented in this dissertation, have had a significant impact in the search of NP, reducing a large part of the parameter space of NP models. The improvement in the measurements expected in the coming years, as well as the possibility of measuring new observables such as $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$, hold promises of new exciting results and will be a strategical goal of the LHCb experiment.

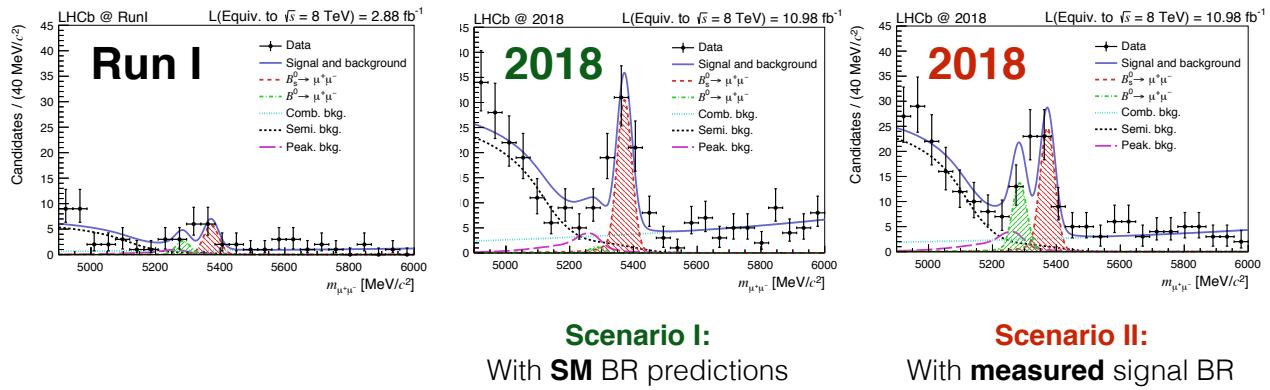


Figure 11.5: The invariant di-muon mass distribution in the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis on Run 1 data sample (left), and on the simulated data samples expected to be collected by the LHCb at the end of 2018, shown for the SM $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction hypothesis (middle), and the measured branching fraction hypothesis (right). The simulated data samples include the Run 1 sample, and assume the Run 1 detector performance and analysis sensitivity.

Hoofdstuk 12

Samenvatting en Vooruitzicht

De huidige theorieën van de fundamentele natuurkunde schieten te kort als het aankomt op sommige van de meest interessante vraagstukken binnen de elementaire deeltjesfysica. Er is nog maar weinig bekend over de materie die verantwoordelijk is voor de meeste massa in ons universum of waarom we leven in een universum dat voornamelijk bestaat uit materie maar nauwelijks anti-materie. In de zoektocht naar de antwoorden hebben natuurkundigen een tal van slimme oplossingen bedacht, waarmee ze naast het gevestigde Standaard Model (SM) ook nieuwe, nog onbevestigde theorieën, ook wel Nieuwe Fysica (NP) genoemd, postuleren. Bijvoorbeeld, door het postuleren van supersymmetrische partners die horen bij de geobserveerde deeltjes (SUSY) of het bestaan van een extra Higgs doublet (2HDM).

De NP-modellen moeten compatibel zijn met de bevestigde SM-voorspellingen. De experimenten in de deeltjesfysica kijken al verder dan het gevestigde domein van de SM door middel van verscheidene manieren: door het bereiken van nog hogere energetische botsingen tussen deeltjes; door het opzetten van ingenieuze, nieuwe precisie-experimenten; of door het zoeken naar nog niet geobserveerde processen waar NP-voorspellingen kunnen verschillen van de SM-voorspelling. Dat laatste is het geval voor het zeldzame verval van B_s^0 en B^0 -mesonen naar twee muonen. Omdat deze $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ vervallen zo zeldzaam zijn en de frequentie nauwkeurig wordt voorspeld door het SM, dragen ze in belangrijke mate bij aan de informatie over NP-modellen.

De resultaten van CERNs Large Hadron Collider (LHC) brachten ons dichter bij het oplossen van de grote onopgeloste vraagstukken. Misschien één van de meest memorabele resultaten van 2011 en 2012, de zogenaamde “Run 1” periode, was de ontdekking van het Higgs boson [4,5] en het eerste bewijs voor het verval van B_s^0 -mesonen naar twee muonen ($B_s^0 \rightarrow \mu^+ \mu^-$) [118]. Er werd al lang gezocht naar deze twee verschijningen. Tegelijkertijd, ver-

scheidene resultaten laten afwijkingen zien met het SM en beloven daarmee opwindende uitkomsten voor de komende jaren!

De resultaten van de zoektocht naar $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ van de afgelopen 30 jaar is te zien in Fig. 12.1. Dit proefschrift beschrijft de zoektocht naar $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in de LHC. Het behandelt de theoretische motivatie, de analyse in het LHCb-experiment [58] en hoe de combinatie met de Run 1 resultaten van het CMS-experiment hebben geleid tot de eerste observatie van $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ verval en het eerste bewijs voor $B^0 \rightarrow \mu^+ \mu^-$ verval [43].

De SM-processen die bijdragen aan het $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ verval zijn goed begrepen. In het SM worden de $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ vervallen onderdrukt door het GIM-mechanisme en kunnen alleen verlopen via hogere orde lusdiagrammen (zie Ch. 1 Fig. 1.3). De SM B_s^0 en B^0 -meson vertakkingsverhoudingen naar twee muonen heeft een zeer precieze voorspelling [41, 43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.66 \pm 0.23) \times 10^{-9}, \quad (12.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}. \quad (12.2)$$

Met deze zeer lage verwachting voor het verval via het SM en doordat vooral lusdiagrammen de transitie mogelijk maken, is het waarschijnlijk dat NP, vooral de modellen met extra Higgs sectors of extra bosonen, de kans op $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ verval aanzienlijk veranderen.

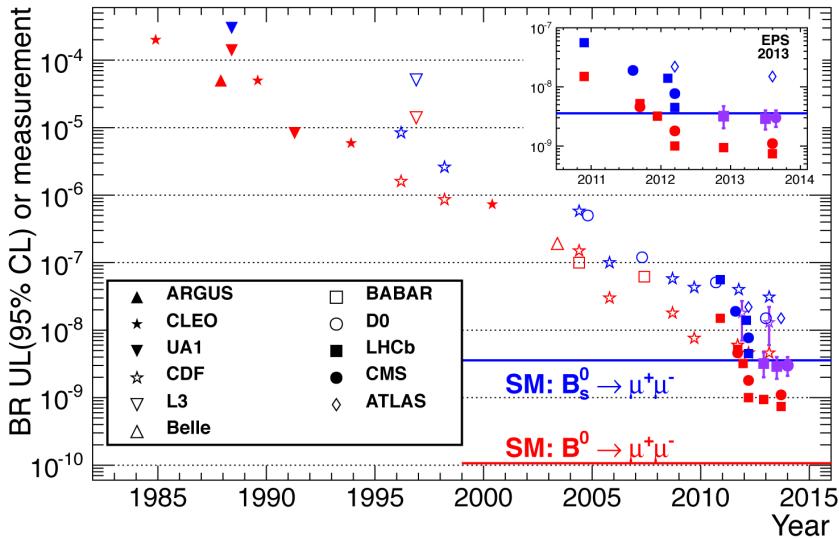
Het hoofdresultaat van de LHC $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ zoektocht zijn de metingen aan de vertakkingsverhoudingen van de B_s^0 en B^0 -mesonen [43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{LHC} = (2.78^{+0.66}_{-0.60}(\text{stat})^{+0.27}_{-0.18}(\text{syst})) \times 10^{-9} \quad (6.2\sigma), \quad (12.3)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{LHC} = (3.94^{+1.58}_{-1.41}(\text{stat})^{+0.31}_{-0.24}(\text{syst})) \times 10^{-10} \quad (3.2\sigma). \quad (12.4)$$

De vertakkingsverhouding en de significanties zijn geëxtraheerd uit het gecombineerde CMS en LHCb likelihood-model, welke samen te zien zijn met de data in Fig. 12.2. De grootste bijdrage aan de onzekerheid op het resultaat is statistisch van aard en kan worden gereduceerd met meer nieuwe data de komende jaren. De systematische fouten dragen respectievelijk 35% en 18% bij aan de totale onzekerheid van de $B_s^0 \rightarrow \mu^+ \mu^-$ en $B^0 \rightarrow \mu^+ \mu^-$ vertakkingsverhoudingen. Ze zijn afkomstig van de signaalnormalisatie, de geschatte hoeveelheid fout-geïdentificeerde achtergrond en het di-muon massamodel. De onzekerheid gemeten in de hadronisatie vertakkingsverhouding (f_s/f_d) is de grootste onzekerheid voor het bepalen van $B_s^0 \rightarrow \mu^+ \mu^-$ vertakkingsverhouding.

De impact van de $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ meting is te zien door het vergelijken van de uitgesloten parameter-ruimte van verscheidene NP-modellen voor en

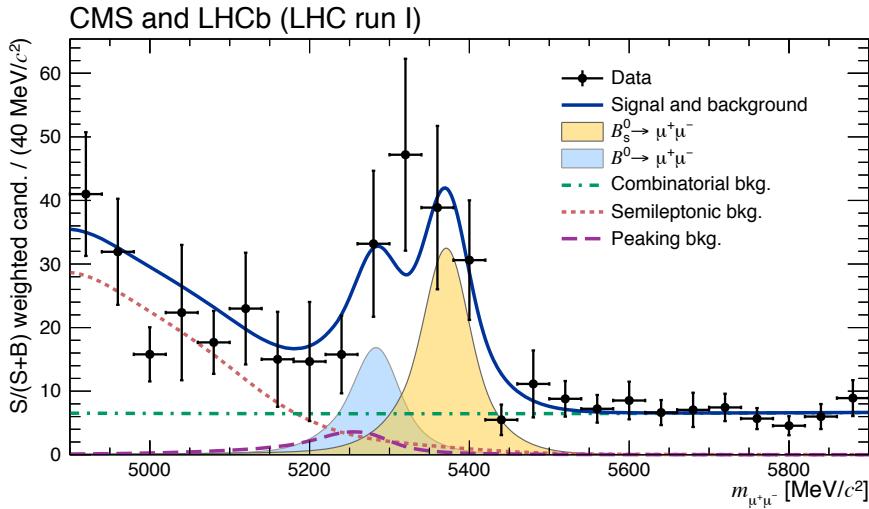


Figuur 12.1: De lange geschiedenis van de zoektocht naar $B_{(s)}^0 \rightarrow \mu^+ \mu^-$.

na de Run 1 resultaten van de LHC (zie Fig. 12.3); Er is al een groot gedeelte van de parameter-ruimte uitgesloten met de Run 1 data. Echter, gegeven de nog grote onzekerheid is de huidige meting compatibel met het SM en nog verscheidene NP-modellen, welke geneigd zijn complexe spectra te vormen in de parameter-ruimte en zelden uitgesloten worden door de meting van één enkele observabele. Daardoor kunnen op moment de NP-bijdrages nog steeds van belang zijn, zelfs al is de gemeten vertakingsverhouding dichtbij de SM-waarde.

Om de impact van de $B_s^0 \rightarrow \mu^+ \mu^-$ meting te bestuderen, kan men kijken naar simpele eerste-orde bijdrages van verschillende nieuw-type deeltjes. Deze bijdrages kunnen afkomstig zijn van zware vectorbosonen, zoals wordt voorspeld in verscheidene Z' en Little Higgs modellen, of van nieuwe scalar of pseudo-scalar deeltjes, welke zijn voorspeld in verschillende typen two-Higgs doublets modellen (2HDM) (zie Sec. 1.4, Fig. 1.8). De toegestane waarden voor de twee prominentste obersvabelen in het $B_s^0 \rightarrow \mu^+ \mu^-$ kanaal, de vertakingsverhouding $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ en de massa-eigenstate-asymmetrie $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$, zijn samen met de laatste $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ meting afgebeeld in Fig. 10.8 (nota bene, de waarde van $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ is experimenteel nog steeds onbekend).

De observatie van het $B_s^0 \rightarrow \mu^+ \mu^-$ verval en het bewijs voor het $B^0 \rightarrow \mu^+ \mu^-$ verval zijn de opeenstapeling van de 30-jarige, experimentele zoektocht naar $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ en maken het mogelijk om zeer precieze metingen te doen aan deze vervalkanalen.

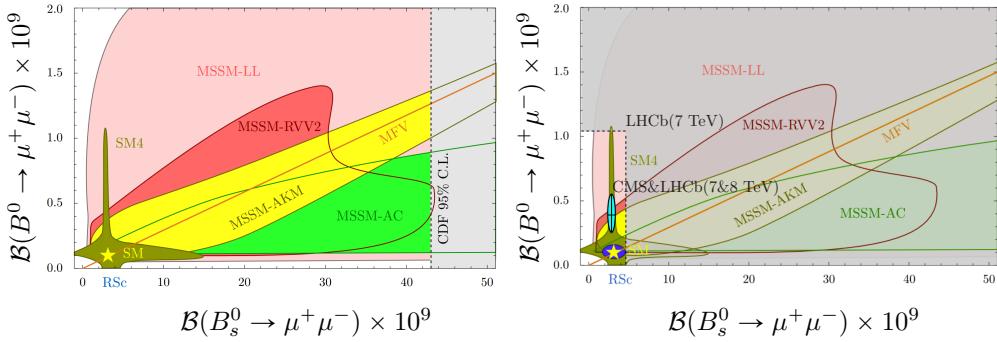


Figuur 12.2: Di-muon invariante massaspectrum voor de acht LHCb en de twaalf CMS BDT-categorieën. De data, hier aangegeven met de zwarte cirkels, is samen getekend met de gecombineerde fit (blauwe lijn), welke is opgebouwd uit de volgende elementen: de B_s^0 (gele gebied) en B^0 -signaal (lichtblauwe gebied) componenten; de combinatorische achtergrond (doorbroken groene lijn); de som van de semi-leptonische achtergronden (roze stippellijn); en de piekende achtergronden (doorbroken paarse lijn).

Zoals aangegeven in Ref. [129], het $B_s^0 \rightarrow \mu^+\mu^-$ verval is een zuivere, theoretische indicator voor de C_{10}^R Wilson-coëfficiënt. Een meer nauwkeurige bepaling van deze coëfficiënt is noodzakelijk, gezien de verschillende spanningen met het SM in andere, interessante vervalprocessen, waarbij $b \rightarrow s$ overgangen, zoals $B^0 \rightarrow K^{*0}\mu^+\mu^-$ [130], $B^+ \rightarrow K^+\mu^+\mu^-$ en $B^+ \rightarrow K^+e^+e^-$ zijn betrokken [131]. Strengere restricties op C_{10}^R zijn nodig voor het identificeren van NP die bijdraagt aan andere Wilson-coëfficiënten, zoals C_9^R .

Sterker nog, de huidige, beste meting aan de $B^0 \rightarrow \mu^+\mu^-$ vertakkingssverhouding komt drie keer hoger uit dan wat het SM voorspelt en moet daarom zeker met een nog hogere precisie worden gemeten.

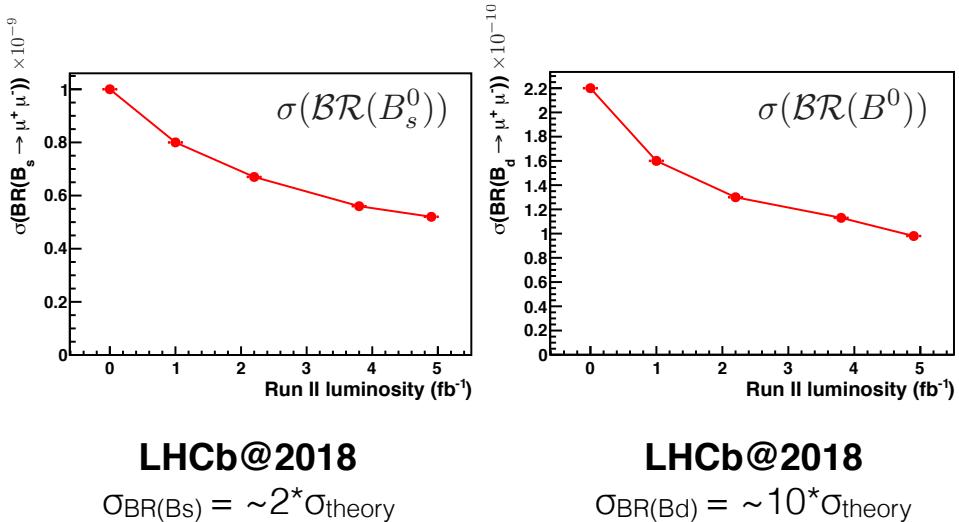
Nog preciezere metingen zullen al mogelijk zijn met de LHC Run 2 data. Op moment dat dit geschreven wordt, heeft de LHCb-detector al 0.32 fb^{-1} aan proton-protonbotsingen geregistreerd voor $\sqrt{s} = 13 \text{ TeV}$ (wat overeenkomt met 0.5 fb^{-1} voor $\sqrt{s} = 8 \text{ TeV}$). De verwachte onzekerheden voor de toekomstige LHCb-analyse zijn te zien in Fig. 12.4. Twee mogelijke toekomstscenario's zijn geschat in Fig. 12.5. Ervanuitgaande dat de huidige analyse-gevoeligheid en detector-prestatie hetzelfde blijven, heeft het LHCb-



Figuur 12.3: De mogelijke waarden voor de $B^0 \rightarrow \mu^+ \mu^-$ en $B_s^0 \rightarrow \mu^+ \mu^-$ vertakingsverhoudingen in het $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ en $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ vlak voor verschillende NP-modellen. De SM-voorspelling is aangegeven met de gele ster. Het grijze gebied op de linkerfiguur geeft aan welke regio's zijn uitgesloten met experimenten voor het LHC-tijdperk; het vergrijsde gebied op de figuur rechts zijn de regio's die door de LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyse over de 2011 data worden uitgesloten [125]; en de lichtblauw-gekleurde ellips in de rechterfiguur represeneert de gecombineerde CMS en LHCb-metingen [43].

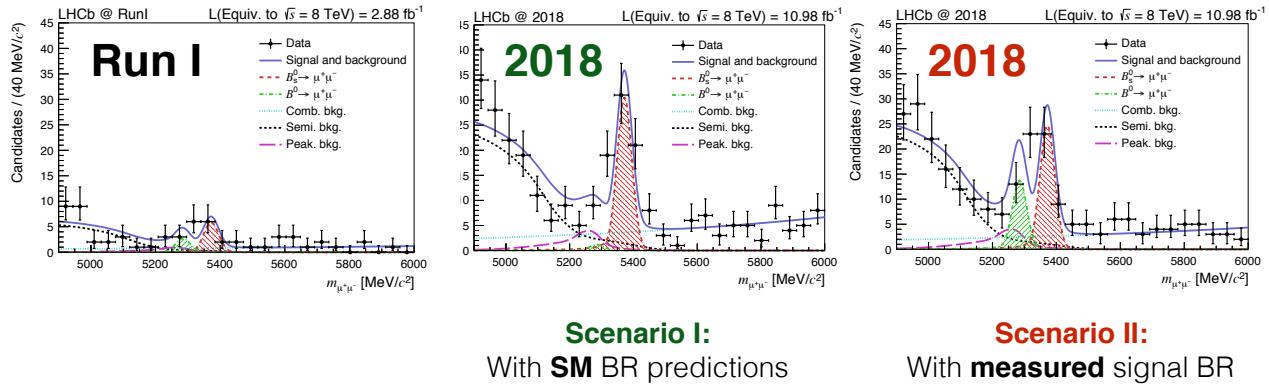
experiment zelf een kans van rond de 85% om een 5σ $B_s^0 \rightarrow \mu^+ \mu^-$ observatie te bereiken (met als aanname dat de $B_s^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding gelijk is aan de SM-verwachting) en een kans van rond de 80% om een 3σ $B^0 \rightarrow \mu^+ \mu^-$ bewijs te vinden (ervan uitgaande dat de $B^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding gelijk is aan de huidig gemeten vertakingsverhouding). Verder nog, als de gemeten $B^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding stand houdt, dan is er een één op vijf kans om het bewijs te vinden voor een niet-SM $B^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding-waarde! De mogelijkheid voor een versterkte waarde van de $B^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding via NP maakt de $B^0 \rightarrow \mu^+ \mu^-$ meting het resultaat waar het meest naar wordt uitgekeken bij de volgende $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyse.

De huidige situatie zoals aangegeven in Fig. 10.8, kan aanzienlijk worden verbeterd door het meten van de massa-eigenstate-asymmetrie in $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. Op dit moment, de grootte van de toegestane gebieden in Fig. 10.8 suggereren dat de dominerende pseudo-scalar Higgs (A^0) en de scalar Higgs (H^0) van de 2HDMs niet geprefereerd worden door de meting, maar dat modellen met een nieuwe, zware ijkboson, zoals Z' , minder restricties krijgen opgelegd. Samen met de nog preciezere meting van de $B_s^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding zal de $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ -meting een belangrijke stap



Figuur 12.4: De verwachte onzekerheden voor de $B_s^0 \rightarrow \mu^+ \mu^-$ (links) en $B^0 \rightarrow \mu^+ \mu^-$ (midden) vertakkingverhoudingen en hoe die zich tot elkaar verhouden (rechts) als functie uitgezet tegen de pp -botsingsdata verwacht voor LHCb Run 2. De studies gaan uit van de SM-waarden voor de vertakkingsverhoudingen en de analysegevoeligheid en detector-prestaties van Run 1.

in de toekomst zijn van de $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyse. Zoals wordt gesuggereerd in Ref. [27], kan $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ worden geëxtraheerd uit de levensduur-distributie van de $B_s^0 \rightarrow \mu^+ \mu^-$ vervallen. Afhankelijk van het resultaat kan de meting van $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ het onderscheid maken tussen scalar, pseudo-scalar en ijkboson-uitwisseling in $B_s^0 \rightarrow \mu^+ \mu^-$ [27, 28].



Figuur 12.5: De gefitte invariante di-muon massadistributie voor de LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyse met de Run 1 data-sample in de linkerfiguur en in de overige twee figuren voor de gesimuleerde data die verwacht verzameld te worden door de LHCb-detector aan het einde van 2018. De middenfiguur gaat uit van een SM-verwachting voor de $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ vertakingsverhouding en de rechterfiguur gaat uit van de huidig gemeten vertakingsverhouding. De gesimuleerde sets zijn inclusief de Run 1 sample en de aanname is dat de detector-prestatie en analyse-gevoeligheid van Run 1 hetzelfde blijven voor Run 2.

Peatükk 13

Kokkuvõte

Kaasaegne füüsika kirjeldab edukalt mikromaailma struktuuri ja mikromaailmas toimuvaid protsesse. Mikromaailma nüüdismudel, Standardmudel (SM), ei suuda aga vastata paljudele olulistele küsimustele universumi koostise ja selle arengu kohta: vähe on teada tumedast ainest ja tumedast energiast, mis koos moodustavad enamuse universumi massist või sellest, miks universumi koostises on nii vähe anti-ainet. Füüsikud on nendele küsimustele välja pakkinud palju kavalaid selgitusi ja laiendanud SM'it võttes appi uue ning veel tundmatu füüsika, nn. Uue Füüsika (UF). KAESOLEVA töö eesmark on otsida erinevate UF mudelite ilminguid ja seelabi jõuda lähemale universumi mõistmisele.

Mendelejevi tabeli elemendid koosnevad prootonitest, neutronitest ja elektronidest. Protonid ja neutronid omakorda koosnevad u ja d tüüpi kvarkidest. Seega keemiliste elementide koostises on kahte tüüpi kvarke ja ühte tüüpi leptoneid - elekrone. SM'i järgi koosneb nähtav aine universumis kuute tüüpi kvargist ja kuute tüüpi leptonist. Nähtava aine dünaamika määravad kvarkide ning leptonite vahelised jõud ja jõudude mõju ainele kirjeldab SM'i jõu osakeste ehk bosonite vahetuse kaudu. Üheks SM'is kirjeldatud jõududest on nt. nõrk vastastikmõju, mis kirjeldab aatomituumade lagunemist.

Postuleeritud SM edasiarendusi UF mudelite abil on mitmeid ja väga erinevaid. Näiteks Super-sümmeetrilised UF mudelid (SUSÜ) oletavad, et igal nähtud SM'i osakesel (nt. elektronil) on oma super-sümmeetriline partnerosaake (nt. selektron). Kõige kergem neist partnerosakestest on stabilne ja elektriliselt neutraalne, ja sobiks oma omaduste poolest tumeda aine ehituskiviks.

Kuna UF mudelid peavad olema kooskõlas SM'i ennustustega, vaatavad osakestefüüsika eksperimendid juba kinnitust leidnud SM'i ennustustest kauemale. Uurimata alale võib jõuda mitmel viisil: kas tõstes põrgatavate osakeste energiat, kavandades uusi täpsusmõõtmisi või otsides seni märkamatuks

jääenud protsesse, mis UF mudelites käituvald teisiti kui SM'is. Hea näide loetelus viimasesest viisist on antud töös otsitud harvad B mesoni lagunemised kaheks müüöniks. Koosnedes kvargist ja anti-kvargist, on meson väikseim kvarkidest moodustatud osake. B meson on meson mille koostises üks kvarkidest on b -kvark. Tänu üli-väikestele lagunemistõenäosustele, on $B_s^0 \rightarrow \mu^+ \mu^-$ ja $B^0 \rightarrow \mu^+ \mu^-$ lagunemised (ühiselt $B_{(s)}^0 \rightarrow \mu^+ \mu^-$) eriti tundlikud UF mõjutustele ja annavad väärthuslikku teavet universumi koostise mõistmiseks.

Suures Hadronite Põrgutis (LHC) ajavahemikul 2011 - 2012 kogutud andmed on viinud meid sammu võrra lähemale universumi mõistmissele. Kaks kõige meeldejäävamat tulemust LHC esimesest katse perioodist ("Run 1") on Higgsi bosoni avastamine [4, 5] ja esimesed tõendid B_s^0 mesoni lagunemisest kaheks müüöniks ($B_s^0 \rightarrow \mu^+ \mu^-$) [118]. Nii Higgsi bosonit kui ka $B_s^0 \rightarrow \mu^+ \mu^-$ lagunemist on eelnevate eksperimentidega otsitud aastakümneid: Joonisel 13.1 on välja toodud $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ otsingute tulemused alates esimestest otsingu-test 30 aastat tagasi.

Käesolev doktoritöö kirjeldab $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ otsinguid LHCs. Doktoritöös selgitatakse miks just need harvad lagunemised on universumi mõistmisel olulised, kuidas LHCb eksperiment $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemisi otsib, milliste tulemusteni jõuti [58] ning kuidas kahe erineva eksperimendi, CMSi ja LHCb, ühisanalüüsiga avastati $B_s^0 \rightarrow \mu^+ \mu^-$ lagunemine. Lisaks kirjaldatakse, kuidas samas analüüsits leiti ootamatult esimesed tõendid veelgi haruldasemast $B^0 \rightarrow \mu^+ \mu^-$ lagunemisest [43].

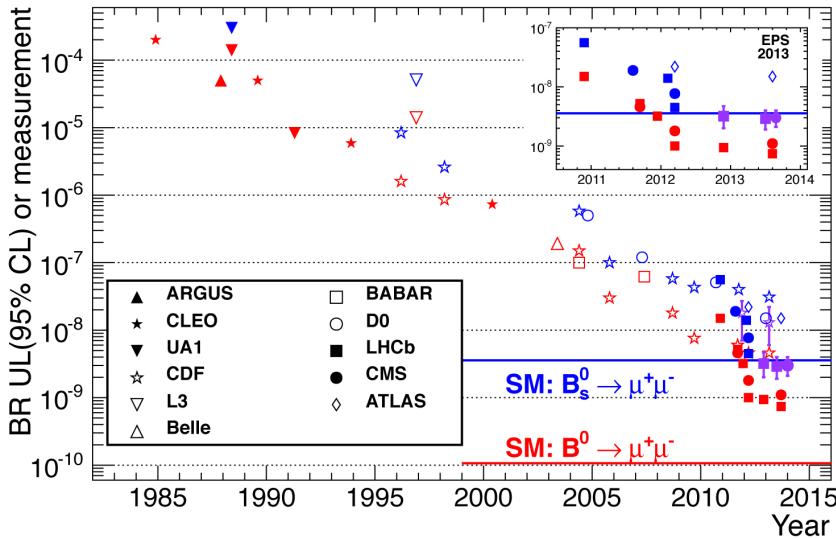
SM'i protsesse mis on võimalised lagundama B_s^0 või B^0 mesoni kaheks müüöniks on pikalt uuritud. Nende esinemissagedused on teoreetiliselt täpselt ennustatud ja põhjus miks $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemised SM'is nii harvad on peitub nn. GIM mehhanismis (vt. Peatükk 1 ja Joonis 1.3), mis lubab ainult kõrgema järgu (nn. silmus) protsessidel toimuda.

Lagunemiste sagedust kirjeldavad lagunemismurruud. B_s^0 (või B^0) mesoni lagunemismurruud kaheks müüöniks on antud lagunemiste arvu osakaal kõikidest võimalikest B_s^0 (või B^0) mesoni lagunemistest. SM'iga on võimalik antud lagunemismurde täpselt ennustada [41, 43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.66 \pm 0.23) \times 10^{-9}, \quad (13.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}. \quad (13.2)$$

Paljud UF mudelid aga ennustavad märgatavalalt suuremat lagunemissagedust. Eriti suur on erinevus laiendatud Higgsi sektoriga UF (2HDM) mudelites, kus $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurruud võivad olla kuni kümneid kordi võimendatud. Teoreetiliselt täpselt ennustatud ja pea olematute SM'i lagunemismurdude taustal on erinevaid UF mudeliteid $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemistes väga tõhus uurida.



Joonis 13.1: Pikk $B_s^0 \rightarrow \mu^+ \mu^-$ ja $B^0 \rightarrow \mu^+ \mu^-$ otsingute ajalugu: $B_s^0 \rightarrow \mu^+ \mu^-$ ja $B^0 \rightarrow \mu^+ \mu^-$ lagunemismurdude ülempiirid (vastavalt, sinised ja punased märgid) läbi aastate koos hiljutiste mõõtmistulemustega.

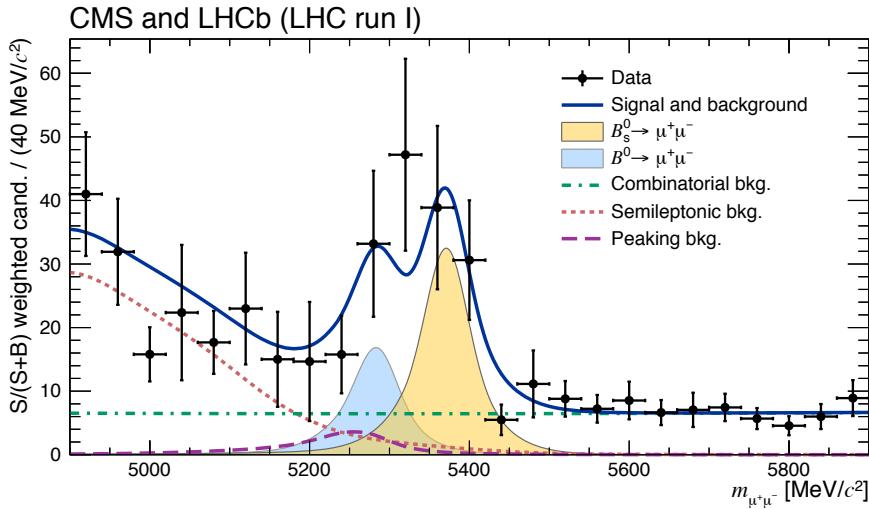
$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ otsingute kõige oluliseim mõõtmistulemus on käesolevas doktoritöös (vt. Peatükk 10) mõõdetud B_s^0 ja B^0 lagunemismurrud [43]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{LHC} = (2.78^{+0.66}_{-0.60}(\text{stat})^{+0.27}_{-0.18}(\text{syst})) \times 10^{-9} \quad (6.2\sigma), \quad (13.3)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{LHC} = (3.94^{+1.58}_{-1.41}(\text{stat})^{+0.31}_{-0.24}(\text{syst})) \times 10^{-10} \quad (3.2\sigma). \quad (13.4)$$

Mõõdetud lagunemismurdude statistilised mõjukused (σ) on arvutatud ühtset CMSi ja LHCb võimalikkusmudelite kasutades. Antud mudeli parim lähendus ühendatud katseandmetele on näidatud Joonisel 13.2. Katseandmed kinnitavad esmakordselt $B_s^0 \rightarrow \mu^+ \mu^-$ lagunemist. Lisaks sellele, leiti andmetest tõendeid veelgi harvemini esinevast $B^0 \rightarrow \mu^+ \mu^-$ lagunemisest.

Antud mõõtmistulemuste olulisuse toob hästi esile Joonis 13.3, mis võrdleb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurdude mõõtmistulemusi võimalike UF mudelite ennustustega nii enne kui ka pärast antud doktoritöös leitud tulemusi. Suur hulk UF mudelite lahendid on nüüdseks katseandmete poolt välisstatud. Mõõdetud $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurdude määramatusi arvestades on mõlemad mõõdetud lagunemismurrud kooskõlas nii SM'i ennustustega kui ka UF mudelitega. UF mudelite parameetrite spekter on tihti keeruline ja enamasti ei ole UF mudeliteid võimalik täielikult välisstatada pugalt ühe vaadeldava suuruse abil. Seetõttu võivad UF mudelid eksisteerida ka näiteks siis, kui mõõdetud $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurrud on SM'i ennustuste läheosal.



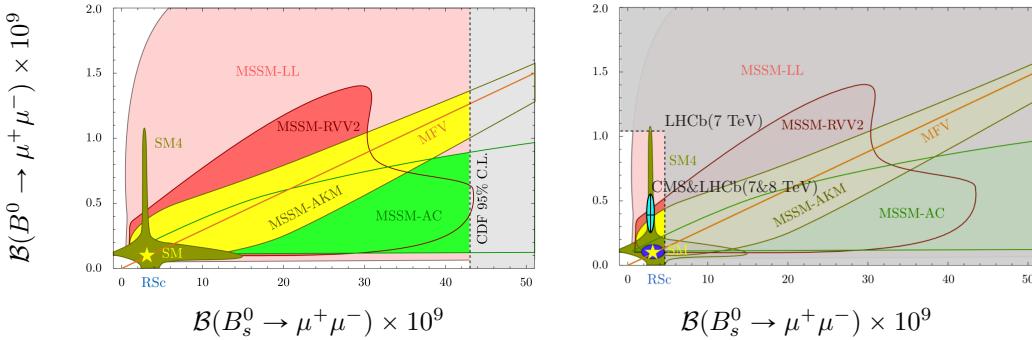
Joonis 13.2: Müünionipaari invariantse massi jaotus LHCb ja CMSi poolt 2011 ja 2012 aastal kogutud andmetes. Katseandmete taustal (mustad punktid) on näidatud ühendatud võimalikkusemudeli parim lähend katseandmetele (sinine pidev joon) ja selle mudeli koostisosad: B_s^0 (kollane varjutatud kühm) ja B^0 (helesinine varjutatud kühm) lagunemised müünioniteks; kombinatsioon-taust (rohelise kriips-puntiirjoon); pool-leptoniline valesti tuvastatud taust (roosa puntiirjoon) ja kuhjuv valesti tuvastatud taust (violetne puntiirjoon).

Mõlema tulemuse puhul on enamus määramatusest statistilist laadi ja väheneb uute andmete lisandudes. Süsteematalise määramatuse osakaal mõõdetud $B_s^0 \rightarrow \mu^+ \mu^-$ ja $B^0 \rightarrow \mu^+ \mu^-$ lagunemismurdudes on vastavalt 35% ja 18%, ning tuleneb mõõdetud lagunemiste arvu normeerimisest, valesti tuvastatud müünionite tausta osakaalu hindamisest ja müünion-paari massijaotuse modelleerimisest. Suurim süsteemataline määramatus pärineb mõõdetud B_s^0 ja B^0 hadronisatsiooniprotsesside suhtarvust (f_s/f_d) [113].

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analüüsi tulevik

$B_s^0 \rightarrow \mu^+ \mu^-$ lagunemise esmakordne mõõtmine ja esimesed tõendid $B^0 \rightarrow \mu^+ \mu^-$ lagunemisest tähistavad 30 aasta pikkuste otsingute lõppu ja annavad võimaluse täpismõõtmisteks antud lagunemiskanalites.

$B_s^0 \rightarrow \mu^+ \mu^-$ teoreetiliselt väga puhas kanal. UF effektide puudumisel sobib $B_s^0 \rightarrow \mu^+ \mu^-$ hästi Wilsoni kordaja C_{10}^R mõõtmiseks (vt. Osa 1.3 ja [129]). Antud kordaja on oluline kuna see mängib olulist rolli ka mitmes teises $b \rightarrow s$

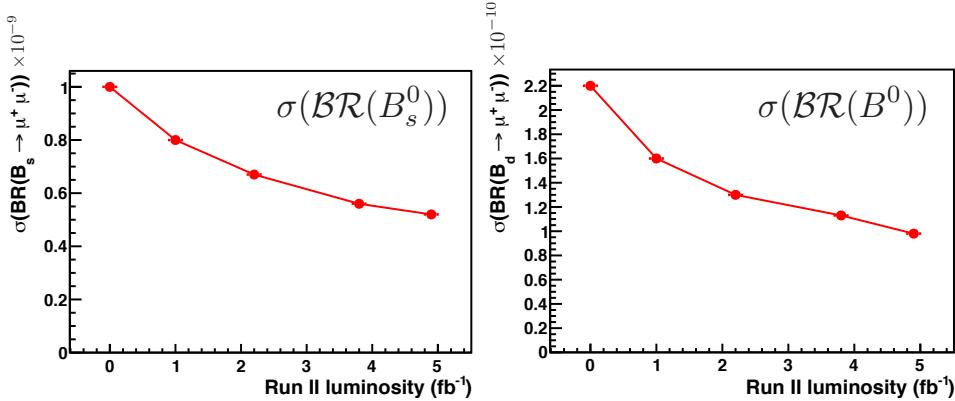


Joonis 13.3: Võimalikud $B^0 \rightarrow \mu^+ \mu^-$ ja $B^0 \rightarrow \mu^+ \mu^-$ lagunemismurdude väärtsused erinevates Uue Füüsika mudelites. Standardmudeli ennustus on märgitud lagunemismurdude tasandile tähega. Varjutatud hallala vasakul joonisel tähistab eksperimentaalselt välistatud ala enne LHC möötmistulemusi, varjutatud hallala paremal joonisel tähistab 2011 aasta ($\sqrt{s} = 7$ TeV) LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analüüsitlemuste poolt välistatud ala [125]. Antud doktoriöös leitud ja seni täpseimad $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurrud on määrdgitud paremal joonisel helesinise ellipsiga [43].

tüüpi lagunemistes, nagu näiteks $B_s^0 \rightarrow K^{*0} \mu^+ \mu^-$ [130] ning $B^+ \rightarrow K^+ \mu^+ \mu^-$ ja $B^+ \rightarrow K^+ e^+ e^-$ [131], kus viimased möötmised näitavad huvitavaid eba-kõlasid SM'i ennustuste ja katsetulemuste vahel. Möötmised viitavad ühiselt pingetele Wilsoni kordajas C_9^R , ja seega võimalikule UF'le. Täpsem C_{10}^R kordaja väärthus aitaks paremini uurida võimalikke UF effekte Wilsoni kordajas C_9^R .

Lisaks $B_s^0 \rightarrow \mu^+ \mu^-$ lagunemisele on hetkel möödetud $B^0 \rightarrow \mu^+ \mu^-$ lagunemismurd pea neli korda kõrgem SM'i ennustusest. Seda erinevust on vaja kindlasti edasi uurida. Täpsemaid möötmistulemusi võib oodata juba lähitulevikus. Hetkel on LHCs käimas nn. teine katseperiood ("Run 2") ja LHCb detektoriga on kogutud 0.32 fb^{-1} prooton-prooton põrgete andmeid rekordilisel energial $\sqrt{s} = 13$ TeV (võrdväärne umbes 0.5 fb^{-1} suuruse esimesel katseperioodil kogutud andmehulgaga).

Joonisel 13.4 on näidatus lagunemismurdude möõtemääramatuste vähenemine andmehulga suurenedes. Kaks võimalikku LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analüüsi tulevikustsenaariumi on kujutatud Joonisel 13.5. Eeldades, et analüüs tundlikkus jäab samale tasemele ja detektor toimib sama hästi kui siiani, on LHCb-l hea võimalus mööta kindlaid UF märke (3σ või enam) $B^0 \rightarrow \mu^+ \mu^-$ lagunemiskanalis. $B^0 \rightarrow \mu^+ \mu^-$ lagunemismurd on seetõttu üks enimoodatud tulemus järgmistes $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analüüsides.

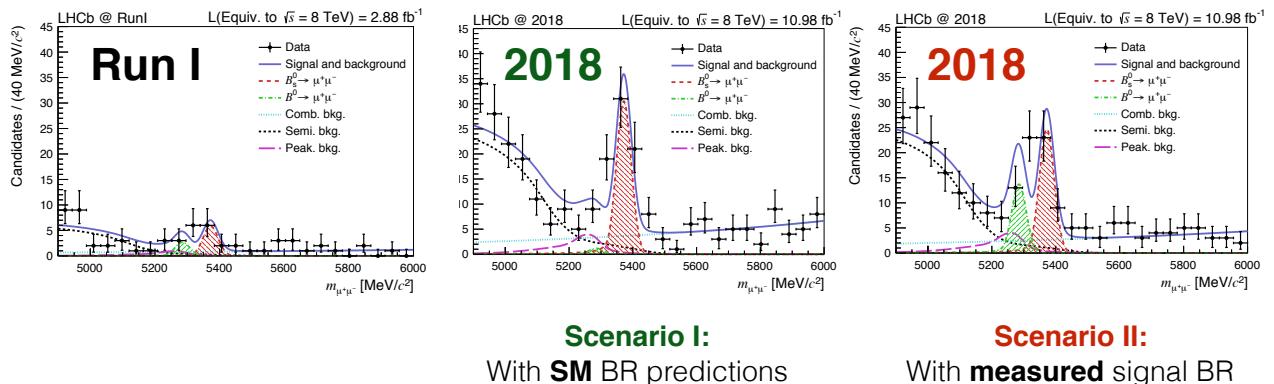
**LHCb@2018**

$$\sigma_{\text{BR}(B_s)} = \sim 2^* \sigma_{\text{theory}}$$

LHCb@2018

$$\sigma_{\text{BR}(B_d)} = \sim 10^* \sigma_{\text{theory}}$$

Joonis 13.4: Oodatud $B_s^0 \rightarrow \mu^+ \mu^-$ (vasakul) ja $B^0 \rightarrow \mu^+ \mu^-$ (keskel) lagunemismurdude mõõtemääramatused andmehulgaga kasvades. Simulatsioon eeldab, et analüüsitudlikkus jäab samaks ja detektor toimib sarnaselt praegusele. Paremal on näidatud lagunemismurdude suhte mõõtemääramatuse oodatud langus uute katseandmete lisandudes.

**Scenario I:**With **SM** BR predictions**Scenario II:**With **measured** signal BR

Joonis 13.5: Müüionipaari invariantse massi jaotus LHCb's 2011 ja 2012 aastal kogutud andmetes ("Run 1", vasakul), ja aastaks 2018 ("Run 2") eeldatavasti kogutavates katseandmetes eeldades Standardmudeli ennustatud $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ lagunemismurde (keskel) või mõõdetud lagunemismurde (paremal). Simuleeritud andmehulgad sisaldavad aastal 2011 ja 2012 kogutud andmeid ja eeldavad sarnast detektori toimimist ja analüüsitudlikkust.

Biography

In 2009, Siim defended his B.Sc. degree in physics at the *University of Tartu* and continued his physics studies in the *Master of Particle Physics* programme at the *University of Amsterdam*.

Siim joined the LHCb collaboration in 2010. Under the supervision of prof. dr. Antonio Pellegrino, he focussed on determining the difference between matter and anti-matter using the $B^0 \rightarrow J/\psi K_S$ decays in the first LHCb data. Siim defended his master thesis at the *University of Amsterdam* and attained a M.Sc. degree in physics in 2011 (*cum-laude*).

In 2011, Siim continued his research under the supervision of prof. dr. Antonio Pellegrino as a doctoral student. The research focus changed from matter-anti-matter differences to searches of New Physics effects in very rare B meson decays to two muons. This thesis summarises Siim's research and findings during his four doctoral years.

Acknowledgements

This dissertation was made possible thanks to the guidance, support, and help I received from the people around me. It would take another volume to properly acknowledge everyone who has helped me during these four years. Instead, I thank all of you who are reading this and ask you to pass on my gratitude (and a copy of my thesis) to those who are not, but should be thanked.

I would like to personally express the gratitude to the two people who directly contributed to the making of my thesis: my supervisor prof. dr. Antonio Pellegrino and my co-supervisor dr. Francesco Dettori.

Antonio - thank you. You have been the most helpful when help was needed the most. I appreciate your ability to see the bigger picture and focus on things that matter. Also, I admire the elegance and ease with which you manoeuvre through the administrative procedures. I am especially grateful to your assistance and patience when working with the final versions of the manuscript and for never failing to keep up the good spirit. Francesco, thank you for always having time for discussions, never failing to ask the right question, and most importantly, for your good-natured attitude and passion for physics.

I would also like to thank all the people involved in the LHCb $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, we make a good team. And finally, warm greetings and many thanks to all the former, present, and future Nikhef people for keeping your office doors open and for generating a welcoming and inquisitive atmosphere. I hope to see you soon in the Nikhef coffee corner, at the well known meeting places in Amsterdam, or why not both.

Many thanks to all of you for your support over the years. It has been a pleasure working with you and I hope we will continue collaborating (and discovering) in the future.

Bibliography

- [1] J. Iliopoulos, “Introduction to the Standard Model of the Electro-Weak Interactions,” in *Proceedings, 2012 European School of High-Energy Physics (ESHEP 2012)*, pp. 1–30. 2014.
[arXiv:1305.6779 \[hep-ph\]](https://arxiv.org/abs/1305.6779). <http://inspirehep.net/record/1236083/files/arXiv:1305.6779.pdf>. Cited on pages: i and 1.
- [2] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.
<http://link.aps.org/doi/10.1103/PhysRevLett.13.508>. Cited on pages: i and 5.
- [3] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.
<http://link.aps.org/doi/10.1103/PhysRevLett.13.321>. Cited on pages: i and 5.
- [4] **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** (2013) 1–29,
[arXiv:1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214). Cited on pages: i, 5, 177, 185, and 194.
- [5] **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 \[hep-ex\]](https://arxiv.org/abs/1207.7235). Cited on pages: i, 5, 177, 185, and 194.
- [6] M. Thomson, *Modern particle physics*. Cambridge University Press, New York, 2013. <http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013>. Cited on pages: 1 and 11.

- [7] N. G. Cooper and G. B. West, eds., *Particle Physics. A Los Alamos Primer*. Cambridge, USA: Univ. Pr. (1988) 199p, 1988. <http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=0521345421>. Cited on page: 1.
- [8] D. Griffiths, *Introduction to elementary particles*. Weinheim, Germany: Wiley-VCH (2008) 454 p, 2008. Cited on page: 1.
- [9] D. Green, “Lectures in particle physics,” *World Sci. Lect. Notes Phys.* **55** (1994) 1–475. Cited on page: 1.
- [10] J. F. Donoghue, E. Golowich, and B. R. Holstein, “Dynamics of the standard model,” *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **2** (1992) 1–540. Cited on page: 1.
- [11] L. O’Raifeartaigh, *Group Structure Of Gauge Theories*. Cambridge University Press, 1988. <http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=0521252938>. Cited on page: 1.
- [12] **Planck** Collaboration, P. A. R. Ade *et al.*, “Planck 2015 results. XIII. Cosmological parameters,” [arXiv:1502.01589 \[astro-ph.CO\]](https://arxiv.org/abs/1502.01589). Cited on page: 2.
- [13] **Gargamelle** Collaboration, F. J. Hasert *et al.*, “Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment,” *Phys. Lett.* **B46** (1973) 138–140. Cited on page: 2.
- [14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, “Experimental test of parity conservation in beta decay,” *Phys. Rev.* **105** (Feb, 1957) 1413–1415. <http://link.aps.org/doi/10.1103/PhysRev.105.1413>. Cited on page: 3.
- [15] H. Harari, “Three Generations of Quarks and Leptons,” in *Boston Meson Conf.1977:0170*, p. 0170. 1977. <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=SLAC-PUB-1974>. Cited on page: 3.
- [16] S. L. Glashow, “Partial Symmetries of Weak Interactions,” *Nucl. Phys.* **22** (1961) 579–588. Cited on page: 3.

- [17] S. Weinberg, “A model of leptons,” *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.
<http://link.aps.org/doi/10.1103/PhysRevLett.19.1264>. Cited on page: 3.
- [18] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C***680519** (1968) 367–377. Cited on page: 3.
- [19] **Particle Data Group** Collaboration, K. A. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys. C***38** (2014) 090001. Cited on pages: 3, 4, 11, and 137.
- [20] N. Cabibbo, “Unitary symmetry and leptonic decays,” *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533.
<http://link.aps.org/doi/10.1103/PhysRevLett.10.531>. Cited on page: 6.
- [21] M. Kobayashi and T. Maskawa, “CP-violation in the renormalizable theory of weak interaction,” *Prog. Theor. Phys.* **49** no. 2, (1973) 652–657. Cited on page: 6.
- [22] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the unified model of elementary particles,” *Prog. Theor. Phys.* **28** (1962) 870–880. Cited on page: 7.
- [23] B. Pontecorvo, “Inverse beta processes and nonconservation of lepton charge,” *Sov. Phys. JETP* **7** (1958) 172–173. [Zh. Eksp. Teor. Fiz.34,247(1957)]. Cited on page: 7.
- [24] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak interactions with lepton-hadron symmetry,” *Phys. Rev. D* **2** (Oct, 1970) 1285–1292.
<http://link.aps.org/doi/10.1103/PhysRevD.2.1285>. Cited on pages: 7 and 8.
- [25] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martinez Santos, “Supersymmetry confronts $B_s \rightarrow \mu^+ \mu^-$: Present and future status,” *Phys. Rev. D* **87** no. 3, (2013) 035026, [arXiv:1212.4887 \[hep-ph\]](https://arxiv.org/abs/1212.4887). Cited on pages: 9 and 173.
- [26] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, and N. Tuning, “Branching Ratio Measurements of B_s Decays,” *Phys. Rev. D* **86** (2012) 014027, [arXiv:1204.1735 \[hep-ph\]](https://arxiv.org/abs/1204.1735). Cited on pages: 9 and 11.

- [27] 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 \[hep-ph\]](https://arxiv.org/abs/1204.1737). Cited on pages: 9, 11, 23, 172, 181, 182, 190, and 190.
- [28] 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 \[hep-ph\]](https://arxiv.org/abs/1303.3820). Cited on pages: 9, 11, 18, 19, 171, 172, 181, 182, and 190.
- [29] 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>. Cited on pages: 9, 12, 16, 19, 19, 23, 23, 171, 172, 173, and 181.
- [30] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \rightarrow J/\psi K^+ K^-$ and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays,” *Phys. Rev.* **D87** no. 11, (2013) 112010, [arXiv:1304.2600 \[hep-ex\]](https://arxiv.org/abs/1304.2600). Cited on pages: 11, 12, and 129.
- [31] **LHCb** Collaboration, R. Aaij *et al.*, “Determination of the sign of the decay width difference in the B_s system,” *Phys. Rev. Lett.* **108** (2012) 241801, [arXiv:1202.4717 \[hep-ex\]](https://arxiv.org/abs/1202.4717). Cited on pages: 11, 12, and 129.
- [32] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, “Weak decays beyond leading logarithms,” *Rev. Mod. Phys.* **68** (Oct, 1996) 1125–1244. <http://link.aps.org/doi/10.1103/RevModPhys.68.1125>. Cited on pages: 11 and 14.
- [33] M. Neubert, “B decays and the heavy quark expansion,” *Adv. Ser. Direct. High Energy Phys.* **15** (1998) 239–293, [arXiv:hep-ph/9702375 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9702375). Cited on page: 11.
- [34] A. J. Buras and R. Fleischer, “Quark mixing, CP violation and rare decays after the top quark discovery,” *Adv. Ser. Direct. High Energy Phys.* **15** (1998) 65–238, [arXiv:hep-ph/9704376 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9704376). Cited on page: 11.
- [35] A. J. Buras, “Operator product expansion, renormalization group and weak decays,” *Lect. Notes Phys.* **558** (2000) 65–85, [arXiv:hep-ph/9901409 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9901409). 65(1999). Cited on page: 11.

- [36] G. Buchalla, “Heavy quark theory,” in *Heavy flavor physics: Theory and experimental results in heavy quark physics and CP violation. Proceedings, 55th Scottish Universities Summer School in Physics, SUSSP 2001, St. Andrews, UK, August 7-23, 2001*, pp. 57–104. 2002. [arXiv:hep-ph/0202092 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0202092).
<http://weblib.cern.ch/abstract?CERN-TH-2002-018>. Cited on page: 11.
- [37] R. Fleischer, “Flavour Physics and CP Violation: Expecting the LHC,” in *High-energy physics. Proceedings, 4th Latin American CERN-CLAF School, Vina del Mar, Chile, February 18-March 3, 2007*, pp. 105–157. 2008. [arXiv:0802.2882 \[hep-ph\]](https://arxiv.org/abs/0802.2882).
<http://weblib.cern.ch/abstract?CERN-PH-TH-2008-034>. Cited on page: 11.
- [38] A. Dedes, J. Rosiek, and P. Tanedo, “Complete One-Loop MSSM Predictions for $B \rightarrow$ lepton lepton’ at the Tevatron and LHC,” *Phys. Rev.* **D79** (2009) 055006, [arXiv:0812.4320 \[hep-ph\]](https://arxiv.org/abs/0812.4320). Cited on pages: 11 and 15.
- [39] T. Blake, T. Gershon, and G. Hiller, “Rare b hadron decays at the LHC,” *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 8007, [arXiv:1501.03309 \[hep-ex\]](https://arxiv.org/abs/1501.03309). Cited on page: 11.
- [40] S. Aoki *et al.*, “Review of lattice results concerning low-energy particle physics,” *Eur. Phys. J.* **C74** (2014) 2890, [arXiv:1310.8555 \[hep-lat\]](https://arxiv.org/abs/1310.8555). Cited on pages: 15 and 17.
- [41] 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 \[hep-ph\]](https://arxiv.org/abs/1311.0903). Cited on pages: 17, 18, 20, 178, 186, and 194.
- [42] **ATLAS, CDF, CMS, D0** Collaboration, “First combination of Tevatron and LHC measurements of the top-quark mass,” [arXiv:1403.4427 \[hep-ex\]](https://arxiv.org/abs/1403.4427). Cited on page: 17.
- [43] **LHCb and 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 \[hep-ex\]](https://arxiv.org/abs/1411.4413). Cited on pages: 18, 151, 161, 171, 178, 178, 178, 181, 186, 186, 186, 189, 194, 194, 195, and 197.

- [44] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, “On the Standard Model prediction for BR(Bs,d to mu+ mu-),” *Eur. Phys. J.* **C72** (2012) 2172, [arXiv:1208.0934 \[hep-ph\]](#). Cited on page: 18.
- [45] Y. G. Aditya, K. J. Healey, and A. A. Petrov, “Faking $B_s \rightarrow \mu^+ \mu^-$,” *Phys. Rev.* **D87** (2013) 074028, [arXiv:1212.4166 \[hep-ph\]](#). Cited on pages: 18 and 19.
- [46] 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 \[hep-ph\]](#). Cited on pages: 18 and 56.
- [47] D. Melikhov and N. Nikitin, “Rare radiative leptonic decays $B_s^0 \rightarrow l^+ l^- \gamma$,” *Phys. Rev.* **D70** (2004) 114028, [arXiv:hep-ph/0410146 \[hep-ph\]](#). Cited on page: 19.
- [48] W.-S. Hou, “Source of CP Violation for the Baryon Asymmetry of the Universe,” *Chin. J. Phys.* **47** (2009) 134, [arXiv:0803.1234 \[hep-ph\]](#). Cited on page: 21.
- [49] S. P. Martin, “A Supersymmetry primer,” [arXiv:hep-ph/9709356 \[hep-ph\]](#). [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. Cited on page: 21.
- [50] D. M. Straub, “New Physics Searches in Flavour Physics,” *Nuovo Cim.* **C035N1** (2012) 249–256, [arXiv:1107.0266 \[hep-ph\]](#). Cited on page: 21.
- [51] M. Blanke, A. J. Buras, and S. Recksiegel, “Quark flavour observables in the Littlest Higgs model with T-parity after LHC Run 1,” [arXiv:1507.06316 \[hep-ph\]](#). Cited on page: 21.
- [52] A. J. Buras, F. De Fazio, J. Girrbach, and M. V. Carlucci, “The Anatomy of Quark Flavour Observables in 331 Models in the Flavour Precision Era,” *JHEP* **02** (2013) 023, [arXiv:1211.1237 \[hep-ph\]](#). Cited on page: 21.
- [53] M. Blanke, A. J. Buras, B. Duling, K. Gemmeler, and S. Gori, “Rare K and B Decays in a Warped Extra Dimension with Custodial Protection,” *JHEP* **03** (2009) 108, [arXiv:0812.3803 \[hep-ph\]](#). Cited on page: 21.

- [54] A. J. Buras, F. De Fazio, and J. Gиррбах, “The Anatomy of Z' and Z with Flavour Changing Neutral Currents in the Flavour Precision Era,” *JHEP* **02** (2013) 116, [arXiv:1211.1896 \[hep-ph\]](https://arxiv.org/abs/1211.1896). Cited on pages: 21 and 172.
- [55] A. J. Buras, F. De Fazio, J. Gиррбах, R. Knegjens, and M. Nagai, “The Anatomy of Neutral Scalars with FCNCs in the Flavour Precision Era,” *JHEP* **06** (2013) 111, [arXiv:1303.3723 \[hep-ph\]](https://arxiv.org/abs/1303.3723). Cited on pages: 21 and 22.
- [56] N. Craig, J. Galloway, and S. Thomas, “Searching for Signs of the Second Higgs Doublet,” [arXiv:1305.2424 \[hep-ph\]](https://arxiv.org/abs/1305.2424). Cited on page: 21.
- [57] **CMS** Collaboration, S. Chatrchyan *et al.*, “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ Branching Fraction and Search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS Experiment,” *Phys. Rev. Lett.* **111** (Sep, 2013) 101804. <http://link.aps.org/doi/10.1103/PhysRevLett.111.101804>. Cited on pages: 23, 151, 156, 159, and 161.
- [58] **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 \[hep-ex\]](https://arxiv.org/abs/1307.5024). Cited on pages: 23, 141, 151, 159, 161, 162, 178, 186, and 194.
- [59] G. Isidori and A. Retico, “Scalar flavor changing neutral currents in the large tan beta limit,” *JHEP* **11** (2001) 001, [arXiv:hep-ph/0110121 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0110121). Cited on page: 23.
- [60] A. J. Buras, P. H. Chankowski, J. Rosiek, and L. Slawianowska, “ $\Delta M_{d,s}, B^0 d, s \rightarrow \mu^+ \mu^-$ and $B \rightarrow X_s \gamma$ in supersymmetry at large $\tan \beta$,” *Nucl. Phys.* **B659** (2003) 3, [arXiv:hep-ph/0210145 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0210145). Cited on page: 23.
- [61] G. Antchev, P. Aspell, I. Atanassov, V. Avati, J. Baechler, *et al.*, “First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s} = 7$ TeV,” *Europhys.Lett.* **96** (2011) 21002, [arXiv:1110.1395 \[hep-ex\]](https://arxiv.org/abs/1110.1395). Cited on page: 27.
- [62] **TOTEM** Collaboration, G. Antchev *et al.*, “Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8\text{--}13\text{ TeV}$,” *Phys. Rev. Lett.* **111** no. 1, (2013) 012001. Cited on page: 27.

- [63] **ATLAS** Collaboration, G. Aad *et al.*, “Measurement of the total cross section from elastic scattering in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,” *Nucl. Phys.* **B889** (2014) 486–548, [arXiv:1408.5778 \[hep-ex\]](https://arxiv.org/abs/1408.5778). Cited on page: 27.
- [64] **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. Cited on page: 27.
- [65] **ALICE** Collaboration, B. Abelev *et al.*, “Measurement of inelastic, single- and double-diffraction cross sections in proton–proton collisions at the LHC with ALICE,” *Eur. Phys. J.* **C73** no. 6, (2013) 2456, [arXiv:1208.4968 \[hep-ex\]](https://arxiv.org/abs/1208.4968). Cited on page: 27.
- [66] **ATLAS** Collaboration, G. Aad *et al.*, “Measurement of the Inelastic Proton-Proton Cross-Section at $\sqrt{s} = 7$ TeV with the ATLAS Detector,” *Nature Commun.* **2** (2011) 463, [arXiv:1104.0326 \[hep-ex\]](https://arxiv.org/abs/1104.0326). Cited on page: 27.
- [67] **CMS** Collaboration, S. Chatrchyan *et al.*, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV,” *Phys. Lett.* **B722** (2013) 5–27, [arXiv:1210.6718 \[hep-ex\]](https://arxiv.org/abs/1210.6718). Cited on page: 27.
- [68] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of the inelastic pp cross-section at a centre-of-mass energy of $\sqrt{s} = 7$ TeV,” *JHEP* **02** (2015) 129, [arXiv:1412.2500 \[hep-ex\]](https://arxiv.org/abs/1412.2500). Cited on page: 27.
- [69] R. Koopman, *Ageing of the LHCb Outer Tracker and b-hadron production and decay at $\sqrt{s} = 7$ TeV*. PhD thesis, Vrije U., Amsterdam, Oct, 2015. <http://cds.cern.ch/record/>. Presented 03 Nov 2015. Cited on page: 28.
- [70] **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 \[hep-ex\]](https://arxiv.org/abs/1207.1158). Cited on pages: 29, 30, 108, and 130.
- [71] **DELPHI** Collaboration, P. Abreu *et al.*, “Lifetime and production rate of beauty baryons from Z decays,” *Z. Phys.* **C68** (1995) 375–390. Cited on page: 29.
- [72] **ALEPH** Collaboration, R. Barate *et al.*, “Measurement of the B baryon lifetime and branching fractions in Z decays,” *Eur. Phys. J.* **C2** (1998) 197–211. Cited on page: 29.

- [73] **ALEPH** Collaboration, D. Buskulic *et al.*, “Strange b baryon production and lifetime in Z decays,” *Phys. Lett.* **B384** (1996) 449–460. Cited on page: 29.
- [74] **DELPHI** Collaboration, J. Abdallah *et al.*, “Production of $\Xi_0(c)$ and $\Xi(b)$ in Z decays and lifetime measurement of X(b),” *Eur. Phys. J.* **C44** (2005) 299–309, [arXiv:hep-ex/0510023 \[hep-ex\]](#). Cited on page: 29.
- [75] **DELPHI** Collaboration, P. Abreu *et al.*, “Evidence for B_s^0 meson production in Z^0 decays,” *Phys. Lett.* **B289** (1992) 199–210. Cited on page: 29.
- [76] **OPAL** Collaboration, P. D. Acton *et al.*, “Evidence for the existence of the strange b flavored meson B_s^0 in Z^0 decays,” *Phys. Lett.* **B295** (1992) 357–370. Cited on page: 29.
- [77] **ALEPH** Collaboration, D. Buskulic *et al.*, “Measurement of the B_s 0 lifetime and production rate with $D_s - \ell^+$ combinations in Z decays,” *Phys. Lett.* **B361** (1995) 221–233. Cited on page: 29.
- [78] **DELPHI** Collaboration, J. Abdallah *et al.*, “A Measurement of the branching fractions of the b quark into charged and neutral b hadrons,” *Phys. Lett.* **B576** (2003) 29–42, [arXiv:hep-ex/0311005 \[hep-ex\]](#). Cited on page: 29.
- [79] **CDF** Collaboration, T. Affolder *et al.*, “Measurement of b quark fragmentation fractions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV,” *Phys. Rev. Lett.* **84** (2000) 1663–1668, [arXiv:hep-ex/9909011 \[hep-ex\]](#). Cited on page: 29.
- [80] **CDF** Collaboration, T. Aaltonen *et al.*, “Measurement of Ratios of Fragmentation Fractions for Bottom Hadrons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ -TeV,” *Phys. Rev.* **D77** (2008) 072003, [arXiv:0801.4375 \[hep-ex\]](#). Cited on page: 29.
- [81] **CDF** Collaboration, T. Aaltonen *et al.*, “First Measurement of the Ratio of Branching Fractions $B(\Lambda_b^0 \rightarrow \Lambda_c^+ \mu^- \bar{\nu}_\mu) / B(\Lambda_b^0 \rightarrow \Lambda_c^+ \pi^-)$,” *Phys. Rev.* **D79** (2009) 032001, [arXiv:0810.3213 \[hep-ex\]](#). Cited on page: 29.

- [82] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of b hadron production fractions in 7 TeV pp collisions,” *Phys. Rev. D* **85** no. arXiv:1111.2357. CERN-PH-EP-2011-172. LHCb-PAPER-2011-018, (Nov, 2011) 032008. 28 p. Comments: 28 pages, 12 figures. Cited on pages: 29, 115, 133, and 164.
- [83] **LHCb** Collaboration, J. Alves, A. Augusto *et al.*, “The LHCb Detector at the LHC,” *JINST* **3** (2008) S08005. Cited on page: 29.
- [84] **LHCb** Collaboration, R. Aaij *et al.*, “LHCb Detector Performance,” *Int. J. Mod. Phys. A* **30** no. 07, (2015) 1530022, arXiv:1412.6352 [hep-ex]. Cited on page: 29.
- [85] M. Vesterinen, “Considerations on the LHCb dipole magnet polarity reversal,” Tech. Rep. LHCb-PUB-2014-006. CERN-LHCb-PUB-2014-006, CERN, Geneva, Apr, 2014. <http://cds.cern.ch/record/1642153>. On behalf of the LHCb Collaboration. Cited on page: 32.
- [86] O. Callot and S. Hansmann-Menzemer, “The Forward Tracking: Algorithm and Performance Studies,” Tech. Rep. LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May, 2007. <http://cds.cern.ch/record/1033584>. Cited on page: 42.
- [87] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” *Trans. ASME, Ser. D, J. Basic Eng* (1961) 109. Cited on pages: 42 and 62.
- [88] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl.Instrum.Meth.* **A262** (1987) 444–450. Cited on pages: 42 and 62.
- [89] **LHCb** Collaboration. <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/>. Cited on page: 44.
- [90] **LHCb** Collaboration. <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/>. Cited on page: 44.
- [91] **LHCb** Collaboration. <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/>. Cited on page: 44.
- [92] F. Archilli, W. Baldini, G. Bencivenni, N. Bondar, W. Bonivento, *et al.*, “Performance of the Muon Identification at LHCb,” *JINST* **8** (2013) P10020, arXiv:1306.0249 [physics.ins-det]. Cited on page: 44.

- [93] J. Van Tilburg, *Track simulation and reconstruction in LHCb*. PhD thesis, Vrije U. Amsterdam, Amsterdam, 2005.
<https://cds.cern.ch/record/885750>. Presented on 01 Sep 2005.
 Cited on page: 55.
- [94] **LHCb** Collaboration, S. Eisenhardt, H. Luo, A. Morris, M. Needham, and J. Neill, “Characterisation and magnetic field properties of multianode photomultiplier tubes,” *Nucl.Instrum.Meth.* **A766** (2014) 167–170. Cited on page: 55.
- [95] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP* **0605** (2006) 026, [arXiv:hep-ph/0603175 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0603175).
 Cited on page: 55.
- [96] A. Ryd, D. Lange, *et al.*, “EvtGen, A Monte Carlo Generator for B-Physics,” [http://robbep.web.cern.ch/robbep/EvtGen/ GuideEvtGen.pdf](http://robbep.web.cern.ch/robbep/EvtGen/GuideEvtGen.pdf). Cited on page: 55.
- [97] J. Allison *et al.*, “Geant4 developments and applications,” *Nuclear Science, IEEE Transactions on* **53** no. 1, (Feb, 2006) 270–278. Cited on page: 55.
- [98] **LHCb** Collaboration. <http://lhcb-release-area.web.cern.ch/lhcb-release-area/DOC/bole/>. Cited on page: 56.
- [99] R. Aaij, J. Albrecht, F. Alessio, S. Amato, E. Aslanides, *et al.*, “The LHCb Trigger and its Performance in 2011,” *JINST* **8** (2013) P04022, [arXiv:1211.3055 \[hep-ex\]](https://arxiv.org/abs/1211.3055). Cited on page: 57.
- [100] J. Albrecht, V. Gligorov, G. Raven, and S. Tolk, “Performance of the LHCb High Level Trigger in 2012,” [arXiv:1310.8544 \[hep-ex\]](https://arxiv.org/abs/1310.8544).
 Cited on page: 57.
- [101] A. Puig, “The LHCb trigger in 2011 and 2012,” Tech. Rep. LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov, 2014. <https://cds.cern.ch/record/1970930>. Cited on page: 57.
- [102] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS ACAT* (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039). Cited on page: 74.
- [103] **Particle Data Group** Collaboration, K. Nakamura *et al.*, “The review of particle physics,” *Journal of Physics G* **37** (2010) 075021. <http://pdg.lbl.gov>. Cited on pages: 91 and 122.

- [104] W. Verkerke and D. P. Kirkby, “The RooFit toolkit for data modeling,” *eConf* **C0303241** (2003) MOLT007, [arXiv:physics/0306116 \[physics\]](https://arxiv.org/abs/physics/0306116). Cited on pages: 92, 121, and 162.
- [105] F. James, “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1,” Tech. Rep. CERN-D-506, CERN-D506, CERN, 1994. Cited on pages: 92, 121, 138, and 138.
- [106] **Particle Data Group** Collaboration, J. Beringer *et al.*, “Review of particle physics,” *Phys. Rev. D* **86** (2012) 010001. Cited on pages: 103, 120, 121, 159, 164, 164, 164, and 164.
- [107] **ARGUS** Collaboration, H. R. Albrecht *et al.*, “Search for hadronic $b \rightarrow u$ decays,” *Physics Letters B* **241** no. 2, (1990) 278 – 282. <http://www.sciencedirect.com/science/article/pii/037026939091293K>. Cited on page: 102.
- [108] **LHCb** Collaboration, B. Adeva *et al.*, “Roadmap for selected key measurements of LHCb,” [arXiv:0912.4179 \[hep-ex\]](https://arxiv.org/abs/0912.4179). Cited on page: 103.
- [109] M. Perrin-Terrin, *Searches for B meson decays to purely leptonic final states*. PhD thesis, Aix-Marseille U., Jul, 2013. <http://cds.cern.ch/record/1642541>. Presented 23 Sep 2013. Cited on pages: 105, 129, and 145.
- [110] S. Oggero, *Beauty in the Crowd: Commissioning of the LHCb Pile-Up detector and First evidence of $B_s^0 \rightarrow \mu^+ \mu^-$* . PhD thesis, Vrije U., Amsterdam, Apr, 2013. <http://cds.cern.ch/record/1635658>. Presented 03 Oct 2013. Cited on page: 117.
- [111] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of the track reconstruction efficiency at LHCb,” *JINST* **10** no. 02, (2015) P02007, [arXiv:1408.1251 \[hep-ex\]](https://arxiv.org/abs/1408.1251). Cited on page: 128.
- [112] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of the fragmentation fraction ratio, f_s/f_d and its dependence on B meson kinematics,” *J. High Energy Phys.* **04** no. arXiv:1301.5286. CERN-PH-EP-2013-006. LHCB-PAPER-2012-037, (Jan, 2013) 001. 18 p. Cited on pages: 133, 133, 134, 134, 159, and 164.

- [113] **LHCb** Collaboration, R. Aaij *et al.*, “Updated average f_s/f_d b -hadron production fraction ratio for 7 TeV pp collisions,” Tech. Rep. LHCb-CONF-2013-011, CERN, Jul, 2013.
<https://cds.cern.ch/record/1559262>. Cited on pages: 133, 164, 179, and 196.
- [114] G. Cowan, *Statistical Data Analysis*. Oxford University Press, Oxford, 1998. Cited on page: 137.
- [115] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J.* **C71** no. 2, (2011) .
<http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>. Cited on page: 144 and 144.
- [116] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large,” *Transactions of the American Mathematical Society* **54** no. 3, (1943) pp. 426–482.
<http://www.jstor.org/stable/1990256>. Cited on page: 144.
- [117] 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>. Cited on pages: 144, 168, 169, and 169.
- [118] **LHCb** Collaboration, R. Aaij *et al.*, “First Evidence for the Decay $B_s^0 \rightarrow \mu^+ \mu^-$,” *Phys. Rev. Lett.* **110** (2013) 021801, [arXiv:1211.2674 \[hep-ex\]](https://arxiv.org/abs/1211.2674). Cited on pages: 151, 178, 185, and 194.
- [119] **CMS** Collaboration, S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *JINST* **3** (2008) S08004. Cited on page: 153.
- [120] R. Fisher, *Statistical methods for research workers*. Edinburgh Oliver & Boyd, 1925. Cited on page: 161.
- [121] W.-F. Wang and Z.-J. Xiao, “The semileptonic decays $B/B_s \rightarrow (\pi, K)(l^+ l^-, l\nu, \nu\bar{\nu})$,” *Phys. Rev. D* **86** (Dec, 2012) 114025.
<http://link.aps.org/doi/10.1103/PhysRevD.86.114025>. Cited on page: 164 and 164.
- [122] A. Khodjamirian, C. Klein, T. Mannel, and Y.-M. Wang, “Form factors and strong couplings of heavy baryons from QCD light-cone sum rules,” *JHEP* **09** (2011) 106, [arXiv:1108.2971 \[hep-ph\]](https://arxiv.org/abs/1108.2971). Cited on page: 164.

- [123] **LHCb** Collaboration, R. Aaij *et al.*, “First observation of the decay $B^+ \rightarrow \pi^+ \mu^+ \mu^-$,” *JHEP* **12** (2012) 125, [arXiv:1210.2645 \[hep-ex\]](https://arxiv.org/abs/1210.2645). Cited on page: 164.
- [124] G. Feldman and R. Cousins, “Unified approach to the classical statistical analysis of small signals,” *Phys. Rev. D* **57** (Apr, 1998) 3873–3889. <http://link.aps.org/doi/10.1103/PhysRevD.57.3873>. Cited on pages: 168 and 169.
- [125] **LHCb** Collaboration, R. Aaij *et al.*, “Strong constraints on the rare decays $B_s \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$,” *Phys. Rev. Lett.* **108** (2012) 231801, [arXiv:1203.4493 \[hep-ex\]](https://arxiv.org/abs/1203.4493). Cited on pages: 171, 181, 189, and 197.
- [126] R. Fleischer, “Probing new physics with $B_s^0 \rightarrow \mu^+ \mu^-$: Status and perspectives,” *Int. J. Mod. Phys. A* **29** (2014) 1444004, [arXiv:1407.0916 \[hep-ph\]](https://arxiv.org/abs/1407.0916). Cited on page: 171.
- [127] D. Ghosh, M. Guchait, S. Raychaudhuri, and D. Sengupta, “How Constrained is the cMSSM?,” *Phys. Rev. D* **86** (2012) 055007, [arXiv:1205.2283 \[hep-ph\]](https://arxiv.org/abs/1205.2283). Cited on page: 173.
- [128] O. Buchmueller *et al.*, “The CMSSM and NUHM1 after LHC Run 1,” *Eur. Phys. J. C* **74** no. 6, (2014) 2922, [arXiv:1312.5250 \[hep-ph\]](https://arxiv.org/abs/1312.5250). Cited on page: 173.
- [129] W. Altmannshofer and D. M. Straub, “Implications of $b \rightarrow s$ measurements,” in *Proceedings, 50th Rencontres de Moriond Electroweak interactions and unified theories*, pp. 333–338. 2015. [arXiv:1503.06199 \[hep-ph\]](https://arxiv.org/abs/1503.06199). <http://inspirehep.net/record/1353682/files/arXiv:1503.06199.pdf>. Cited on pages: 180, 188, and 196.
- [130] **LHCb** Collaboration, “Angular analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay,” Tech. Rep. LHCb-CONF-2015-002, CERN, Mar, 2015. <http://cds.cern.ch/record/2002772>. Cited on pages: 180, 188, and 197.
- [131] **LHCb** Collaboration, R. Aaij *et al.*, “Test of lepton universality using $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays,” *Phys. Rev. Lett.* **113** (2014) 151601, [arXiv:1406.6482 \[hep-ex\]](https://arxiv.org/abs/1406.6482). Cited on pages: 180, 188, and 197.

