

UNIVERSITY OF CAMBRIDGE

**Measurements of $B \rightarrow \mu^+ \mu^-$ decays using
the LHCb Experiment**

Hannah Mary Evans

Selwyn College

Supervisor
Prof. Valerie Gibson

This dissertation is submitted for the degree of Doctor of Philosophy
June 2017

Abstract

This dissertation documents a study of very rare B -meson decays at the LHCb experiment, using data taken during the first experiment run of the Large Hadron Collider (LHC) and during the second experiment run until September 2016.

The LHCb experiment was designed to test the Standard Model of particle physics and to search for New Physics effects that go beyond the scope of the Standard Model through the decay of b hadrons produced in high energy proton-proton collisions at the LHC. The measurements described in this dissertation are made using data samples of proton-proton collisions with integrated luminosities of 1.0, 2.0 and 1.4 fb^{-1} , collected at centre-of-mass energies of 7, 8 and 13 TeV, respectively.

The branching fractions of the very rare $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the effective lifetime of $B_s^0 \rightarrow \mu^+ \mu^-$ decays are precisely predicted by the Standard Model and are sensitive to effects from New Physics. New Physics processes could influence the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime independently, and therefore the two observables are complementary.

The $B_s^0 \rightarrow \mu^+ \mu^-$ decay is observed with a statistical significance of 7.8σ and the branching fraction is measured to be $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$. The $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is measured for the first time as $2.04 \pm 0.44 \pm 0.05$ ps. The $B^0 \rightarrow \mu^+ \mu^-$ branching fraction is measured as $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}$ with a statistical significance of 1.6σ . An upper limit is set for the branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$ at the 95% confidence level. All results are consistent with the predictions of the Standard Model.

Declaration

This dissertation is the result of my own work, except where work done in collaboration with others is specified in the text. No part of it has been submitted for another qualification at this or any other University. Finally, this dissertation does not exceed the word limit set by the respective Degree Committee.

Hannah Evans

May 2017

Acknowledgements

I have never been one for effusive speeches or declarations but after almost 4 years working for my PhD there are many acknowledgements I would like to make. I would like to thank Valerie Gibson for giving me the opportunity to complete a PhD at the University of Cambridge, for her guidance throughout my research and for the repeated reading of this dissertation. To Marc-Olivier Bettler and Harry Cliff, without whose support and encouragement this PhD would never have been completed and without whom I would not have had the opportunity to work on one of the flagship analyses of the LHCb experiment. Thanks also go to the $B \rightarrow \mu\mu$ LHCb analysis group for the tireless effort that went into the branching fraction measurements and for the guidance I received for my work. I have benefited greatly from the experience of those within this group. To the Science and Technologies Faculties Council for funding my research, enabling me to spend a year living in Geneva and to attend international conferences. To Selwyn College for support throughout my PhD, particularly with respect to the financial support and encouragement I received from the college during the time I spent training with Cambridge University Women's Boat Club. Past and present members of the Cambridge HEP Group, thank you to the good friends I've made and for making coffee and lunch breaks entertaining. Particular thanks go to John Hill for the computing infrastructure I greatly benefited from during my PhD and for putting up with me filling up a lot of disk space. Thanks also go to Susan Haines for answering my questions and giving me hope that everything would be OK. To my friends from Oxford and Cambridge, rowing friends, church friends and physics friends. To list them all would take some time, but special thanks go to Anouska Bartlett for voluntarily reading through parts of my dissertation. Finally, I would like to thank my family. Thank you for coping with me being very absent during my PhD, although you did move to the opposite side of the country!

Preface

In 2013 I started my PhD at the University of Cambridge and now, almost four years later, it is finally coming to an end. This dissertation describes the research I have undertaken during those four years; studying $B \rightarrow \mu\mu$ decays with the LHCb experiment. Two different measurements of these decays are presented; the measurement of $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ branching fractions and the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. These results have been published in reference [1] and are the first measurements of these decays using data collected during the second experiment run of the Large Hadron Collider (LHC). My contributions have predominately been to the effective lifetime measurement, however there is a significant overlap of the two analysis strategies therefore both measurements are described in this dissertation.

A short introduction to particle physics is presented in Chapter 1 along with a summary of the past experimental searches and recent measurements of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays¹. The theoretical motivation for measuring the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions and the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is discussed in Chapter 2. Chapter 3 describes the LHC and the LHCb experiment, which provided the data used in the measurements for this dissertation. The criteria used to identify $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in LHCb data are described in Chapter 4. This work was carried out over several years by many members, including myself, of the $B \rightarrow \mu\mu$ LHCb analysis group. My contributions are the study of the ‘stripping’ selection described in Sections 4.3.2.1 and 4.3.2.2 and the criteria used to identify $B_s^0 \rightarrow \mu^+\mu^-$ decays for the effective lifetime measurement in Section 4.4.

The measurements of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions are described in Chapter 5; this work was performed by members of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ LHCb analysis group and the description focuses in more detail on the parts of the analysis strategy that are also used for the effective lifetime measurement. My contributions include the technical aspects of this measurement; producing the ROOT ntuples containing data and simulated decays and maintaining the stripping selection applied to data.

¹Throughout this dissertation $B_{(s)}^0$ refers to both the particle and anti-particle states of B^0 and B_s^0 mesons and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ refers to both the particle and anti-particle decays of B^0 and B_s^0 mesons into two oppositely charged muons

The measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime and the systematic uncertainties associated with this measurement are described in Chapters 6 and 7. The work documented in these chapters is the result of my own efforts, although it uses inputs from the branching fraction measurements; the mass shapes for signal and background decays, the yields of $B_s^0 \rightarrow J/\psi \phi$ decays in data and the expected yields of signal and background decays in data.

Finally, a summary is given in Chapter 8 of the main results documented in this dissertation and the future prospects for the branching fraction and effective lifetime measurements are also discussed.

Table of contents

Abstract	iii
Acknowledgements	vii
Preface	ix
1 Introduction	1
2 Theory of $B \rightarrow \mu^+ \mu^-$ decays; the Standard Model and beyond	7
2.1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in the Standard Model	7
2.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Branching Fraction	10
2.3 Quark mixing	13
2.3.1 Time evolution of the $B_{(s)}^0$	15
2.3.2 Impact on the Branching Fraction	18
2.4 $A_{\Delta\Gamma}$ and the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime	19
2.5 The Standard Model predictions	20
2.6 New Physics models and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays	21
3 The LHC and the LHCb experiment	25
3.1 The Large Hadron Collider	26
3.2 The LHCb experiment	29
3.2.1 Tracking	32
3.2.2 Particle identification	39
3.2.3 Trigger	47
3.2.4 Software and simulation	50
3.3 Summary	52
4 Event selection	55
4.1 Background sources	56
4.2 Simulated particle decays	57
4.3 Event selection for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements	59
4.3.1 Trigger requirements	60

4.3.2	Cut-based selection	61
4.3.3	Particle identification	75
4.3.4	Multivariate Classifiers	76
4.3.5	Summary	84
4.4	Selection for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement	88
4.4.1	Trigger requirements	89
4.4.2	Mass range	89
4.4.3	Particle identification	90
4.4.4	Multivariate classifier	91
4.4.5	Summary	102
5	Measurement of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions	105
5.1	Analysis strategy	105
5.2	$B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass and BDT PDFs	108
5.2.1	Mass PDFs	108
5.2.2	BDT PDFs	111
5.3	Background mass PDFs and expected yields	113
5.3.1	Mis-identified $B \rightarrow h^+h'^-$ decays	114
5.3.2	Exclusive backgrounds	115
5.3.3	Combinatorial background	116
5.4	Normalisation	116
5.4.1	$B^0 \rightarrow K^+\pi^-$ and $B^+ \rightarrow J/\psi K^+$ yields	117
5.4.2	Efficiency ratio	118
5.4.3	Hadronisation factors	118
5.4.4	Normalisation parameters	119
5.5	Results	119
6	Measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime	123
6.1	Analysis strategy	123
6.2	Mass PDFs	125
6.3	Decay time distributions	126
6.3.1	$B_s^0 \rightarrow \mu^+\mu^-$ decay time PDF	127
6.3.2	Background decay time PDFs	137
6.4	Measurement strategy	139
6.4.1	To fit for τ or τ^{-1} ?	140
6.4.2	Optimisation of fit configuration	143
6.5	Results	149

7 Systematic uncertainties and cross checks on the effective lifetime	151
7.1 Accuracy of the fit	151
7.1.1 Fit stability with $\tau_{\mu\mu}$ values	152
7.1.2 $B_s^0 \rightarrow \mu^+ \mu^-$ yield estimation	152
7.1.3 Overall bias on $\tau_{\mu\mu}$	154
7.2 Background contamination	154
7.3 Mass PDF parameters	157
7.4 Acceptance function accuracy	158
7.5 Incorrectly assigned primary vertices and the detector resolution	162
7.6 Combinatorial background decay time model	164
7.7 Mix of B_s^0 mass eigenstates	167
7.8 Production asymmetry of B_s^0 and \bar{B}_s^0 mesons	168
7.9 Summary	169
8 Summary and Outlook	171
8.1 Summary	171
8.2 Outlook	173
Bibliography	177
Appendix A Distributions of input variables for the global BDT	189
Appendix B Multivariate classifier development	193
B.1 Input variables	193
B.2 Training parameters	196
B.3 Overtraining test	196

Chapter 1

Introduction

The Standard Model (SM) [2–4] of particle physics is a Quantum Field Theory that describes the building blocks of matter and their interactions. It has been developed over several decades from a combination of theoretical progress and experimental discoveries. The SM predicts that all matter is constructed from combinations of particles called quarks or leptons and their anti-particles. The interactions between these particles are governed by the strong, weak and electromagnetic forces. There are a total of 6 quarks (u, d, s, c, t, b) and 6 anti-quarks ($\bar{u}, \bar{d}, \bar{s}, \bar{c}, \bar{b}, \bar{t}$) in the SM that can interact via all three forces. There are also 6 leptons ($e^-, \mu^-, \tau^-, \nu_e, \nu_\mu, \nu_\tau$) and 6 anti-leptons ($e^+, \mu^+, \tau^+, \bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$), which, unlike quarks, do not interact via the strong force. The interactions of each force are described by the exchange of gauge bosons. The electromagnetic force is mediated by the photon (γ), the weak force by the W^\pm and Z^0 bosons and the strong force by eight gluons (g). The final particle in the SM is the Higgs boson (H^0). It is interactions with the field associated with this boson that are responsible for the intrinsic masses of the particles. The properties of the particles in the SM are summarised in Tables 1.1 and 1.2.

The SM can be used to predict how particles will interact and decay. These predictions have been tested over the past decades and so far the SM has proved to be extremely successful. However, despite its success, there are a number of experimental observations that the SM does not explain. In its current form, the SM cannot explain the observed oscillation of neutrinos between different types [6–9] and it does not provide a particle or mechanism that could account for the observed presence of dark matter and dark energy in the universe [10–13]. The SM includes three fundamental forces but the final force, the gravitational force, is not included in its framework. Furthermore, at the start of the universe matter and anti-matter should have been produced in equal amounts but that is not what is observed in the universe today and the SM does not include a mechanism large enough to account for this asymmetry [14, 15]. As well as experimental observations there are more fundamental questions about the SM that are unanswered. For example the SM

Name	Symbol	Mass/MeV/ c^2	Charge / e
Up	u	$2.2^{+0.6}_{-0.4}$	$\frac{2}{3}$
Down	d	$4.7^{+0.5}_{-0.4}$	$-\frac{1}{3}$
Charm	c	1270 ± 30	$\frac{2}{3}$
Strange	s	96^{+8}_{-4}	$-\frac{1}{3}$
Top	t	$173210 \pm 510 \pm 710$	$\frac{2}{3}$
Bottom	b	4180^{40}_{30}	$-\frac{1}{3}$
Electron	e^-	$0.5109989461 \pm 0.0000000031$	-1
Electron neutrino	ν_e	< 0.000002	0
Muon	μ^-	$105.6583745 \pm 0.0000024$	-1
Muon neutrino	ν_μ	< 0.000002	0
Tau	τ^-	1776.86 ± 0.12	-1
Tau neutrino	ν_τ	< 0.000002	0

Table 1.1 Mass and electric charge of quarks and leptons in the SM taken from reference [5]. Only particles are listed; anti-particles have the same mass and opposite electric charge as their corresponding particles. All masses listed are measured values or limits. The quark masses use the $\overline{\text{MS}}$ renormalisation scheme, apart from the t quark mass that comes from direct measurements.

Name	Symbol	Mass/GeV/ c^2	Charge / e
Photon	γ	0	0
W boson	W^\pm	80.385 ± 0.015	± 1
Z boson	Z^0	91.1876 ± 0.0021	0
Gluons	g	0	0
Higgs boson	H^0	$125.09 \pm 0.21 \pm 0.11$	0

Table 1.2 Mass and electric charge of bosons in the SM taken from reference [5]. All masses listed are measured values except for gluons and the photon. In the SM the gluons and the photon do not couple to the Higgs field and are therefore massless. An upper limit has been placed on the photon mass at 1×10^{-18} eV and gluon masses up to a few MeV/ c^2 are not excluded.

does not explain why there are very large differences in the coupling strengths of the electromagnetic, weak and strong forces or why there is a large range in the masses of quarks and leptons. The examples given here illustrate that despite the success of the SM it is not sufficient to describe the universe and indicate that it could be a low energy approximation of a more fundamental theory [16]. A more complete discussion of the shortcomings of the SM can be found in references [17, 18].

There exist many theories that go beyond the scope of the SM and seek to explain what the SM cannot. These theories predict the presence of new particles and phenomena that are collectively called New Physics (NP). At the moment there is no clear indication of which Beyond the SM (BSM) theory may give the correct description of the universe and the search for NP is ongoing. The Large Hadron Collider (LHC) is the latest machine built to study the predictions of the SM and to search for NP in high energy particle collisions. There are two different approaches used to search for NP effects at the LHC and other experiments; direct searches and indirect searches.

Direct searches involve looking for the production of on-shell NP particles and phenomena in high energy collisions. This type of search is limited by the centre-of-mass energy of the collisions that dictates the energy available for the creation of new particles. The Higgs boson was discovered in 2012 by the ATLAS and CMS collaborations using this type of search [19, 20] but no NP has been observed from direct searches to date. The lack of observations enables constraints to be placed on the parameter space of BSM theories.

Indirect searches precisely measure SM processes and look for deviations in the measured values from the predicted values. Deviations can be caused by the presence of NP that modifies the SM process. Indirect searches are not as limited by the centre-of-mass energy as direct searches because NP or SM particles influencing these processes are off-shell. In a similar way to direct searches, indirect searches that do not reveal NP place constraints on the parameter space of the theoretical models. Although indirect searches are yet to reveal any significant deviations from SM predictions, some interesting anomalies have been seen in measured results in heavy flavour physics. In $b \rightarrow sll$ transitions, where $\ell = e$ or μ , deviations from the SM predictions have been seen in measurements of the angular distribution of $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays [21–25], the branching fraction of $B_s^0 \rightarrow \phi \mu^+ \mu^-$ decays [26] and the ratios $R(K) = \frac{B^+ \rightarrow K^+ \mu^+ \mu^-}{B^+ \rightarrow K^+ e^+ e^-}$ [27] and $R(K^*) = \frac{B^0 \rightarrow K^{*0} \mu^+ \mu^-}{B^0 \rightarrow K^{*0} e^+ e^-}$ [28]. Also measurements of the ratios $R(D)$ and $R(D^*)$ for the branching fractions of $B^0 \rightarrow D^{(*)} \tau^- \nu_\tau$ and $B^0 \rightarrow D^{(*)} \mu^- \nu_\mu$ [29–33] differ from the SM predictions. The individual measurements of these processes are all $\sim 2\text{--}3$ standard deviations from the SM predictions however, combining the results from different decays and experiments increases the difference. The combined results of different $b \rightarrow sll$ transitions are 4–5 standard deviations

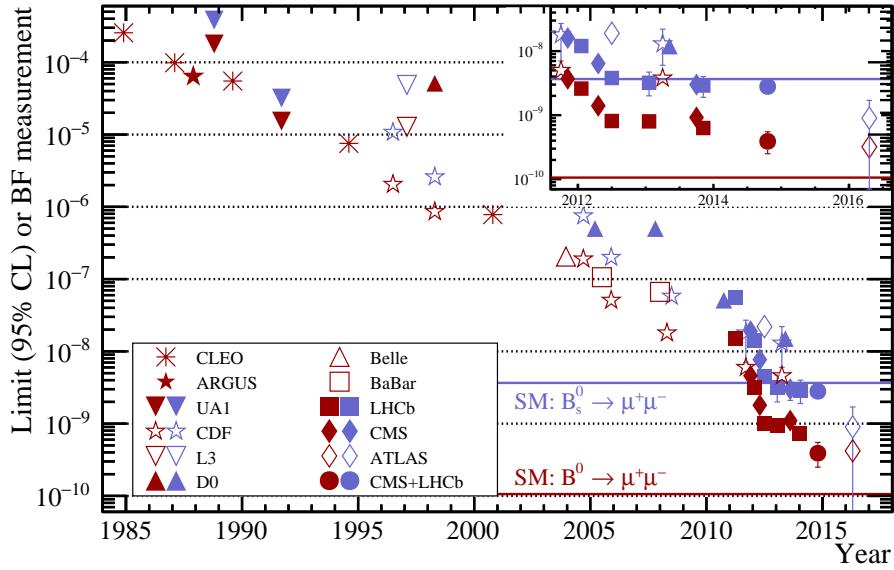


Fig. 1.1 Results from searches for and measurements of $B^0 \rightarrow \mu^+\mu^-$ (red) and $B_s^0 \rightarrow \mu^+\mu^-$ (purple) decays. Upper limits are shown without error bars at the 95% confidence level. The figure is taken from reference [47] and has been updated to include the latest result from the ATLAS experiment [38].

from the SM predictions [34, 35] and the combined values of $R(D)$ and $R(D^*)$ measurements from different experiments are ~ 3.9 standard deviations from the SM predictions [36].

Particle decays and interactions that are suppressed in the SM offer excellent areas for indirect searches for NP because in these processes the contributions from BSM theories can be of a similar order of magnitude to contributions from the SM. The rare decays of B^0 and B_s^0 mesons into two oppositely charged muons are examples of such processes. The purely leptonic final states of these decays lead to precise theoretical predictions for their branching fractions and the two muons have an identifiable signature in particle detectors. The search for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays began over 30 years ago and the experimental sensitivity to these decays has dramatically increased since then as shown in Figure 1.1. The latest experiments to join the search are the ATLAS, CMS and the LHCb experiments [37–47]. The high energy pp collisions of the LHC has enabled these experiments to reach unprecedented sensitivities to $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays.

The first evidence for $B_s^0 \rightarrow \mu^+\mu^-$ decays was found in 2012 by the LHCb experiment [45]. Since then, the LHCb experiment has measured the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction to be $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.9^{+1.1}_{-1.0}) \times 10^{-9}$ at a statistical significance of 4.0σ and placed an upper limit on the $B^0 \rightarrow \mu^+\mu^-$ branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 7.4 \times 10^{-10}$ at the 95% confidence level [46]. These measurements were performed using data collected during 2011 and 2012 at the centre-of-mass energies of 7 and 8 TeV, respectively. Searches for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays performed by the

CMS experiment using data recorded in the same time period corroborated the results from the LHCb experiment. The CMS experiment measured the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction to be $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$ with a statistical significance of 4.3σ and placed a limit on the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.1 \times 10^{-9}$ at the 95% confidence level [41]. The combined analysis of the CMS and LHCb data sets, collected at centre-of-mass energies of 7 and 8 TeV, resulted in the first observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the first evidence for $B^0 \rightarrow \mu^+ \mu^-$ decays [47]. The measured branching fractions were

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{CMS+LHCb}} = 2.8^{+0.7}_{-0.6} \times 10^{-9} \quad (1.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{\text{CMS+LHCb}} = 3.9^{+1.6}_{-1.4} \times 10^{-10} \quad (1.2)$$

with a statistical significance of 6.2σ for the B_s^0 and 3.0σ for the B^0 . The ATLAS experiment also searched for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay using data collected during the same period [38], measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction as

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{ATLAS}} = 0.9^{+1.1}_{-0.8} \times 10^{-9} \quad (1.3)$$

with a statistical significance of 2.0σ . An upper limit was placed on the $B^0 \rightarrow \mu^+ \mu^-$ decay of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) > 4.2 \times 10^{-10}$ at the 95 % confidence level.

Although it was hoped that large deviations from the SM predictions would be found in these decays this has not yet been observed. All the measured values are consistent with the expectations of the SM and have enabled constraints to be placed on the parameter space available for New Physics models. Nevertheless, the precision of the measurements allows plenty of room for NP effects to be revealed. Furthermore, there is some tension between the separate measurements themselves and the SM prediction as shown in Figure 5.9. Therefore, the study of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays continues to be a very interesting topic in the search for NP effects.

The observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays opens the way for other properties of this decay to be studied. In particular the effective lifetime of $B_s^0 \rightarrow \mu^+ \mu^-$ decays provides a search for NP complementary to the branching fraction measurement, as the presence of NP could be revealed in either both or only one of these measurements. The search for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is complete and the precise study of this decay has begun. This dissertation documents the latest study of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment. Measurements of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions and the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime are presented using data collected during pp collisions with centre-of-mass energies of 7, 8 and 13 TeV.

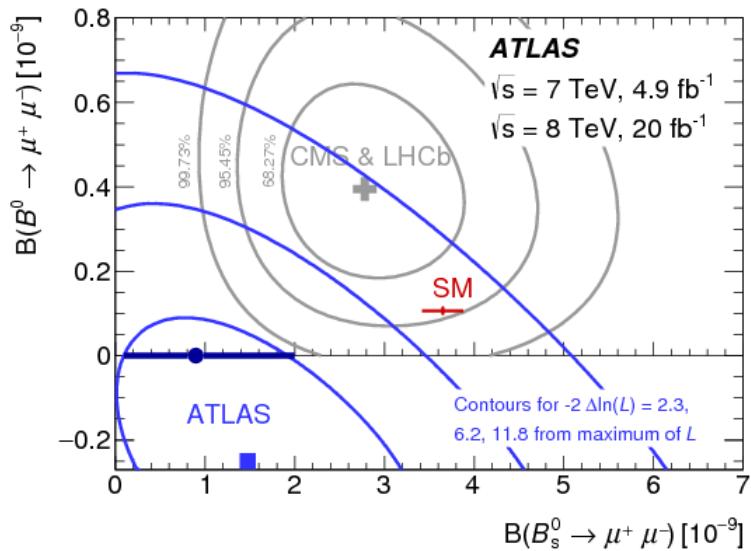


Fig. 1.2 Measurements of the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction from the ATLAS experiment and the combined analysis of CMS and LHCb data alongside the predictions of the SM [38]. The measurements were performed using data collected during 2011 and 2012 and centre-of-mass energies of 7 and 8 TeV, respectively.

Chapter 2

Theory of $B \rightarrow \mu^+\mu^-$ decays; the Standard Model and beyond

This chapter describes the theoretical motivation for the study of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. In the SM these decays are predicted to occur very rarely, Section 2.1 describes why $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are suppressed compared to other decay modes of $B_{(s)}^0$ mesons in the SM. The branching fraction of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays provides a good observable to compare experimental measurements with predictions of the SM. The determination of the theoretical predictions of branching fractions is outlined in Section 2.2 and the discussion is based on references [48, 49]. Quark mixing leads to oscillations between B_s^0 and \bar{B}_s^0 states over time and therefore a difference between the values of the predicted and measured $B_s^0 \rightarrow \mu^+\mu^-$ branching fractions. These oscillations and the influence on the branching fractions values are described in Section 2.3 and follows the material in references [49–51]. A parameter, $A_{\Delta\Gamma}$, arises from the difference between B_s^0 and \bar{B}_s^0 mesons, this observable is complementary to the branching fraction of $B_s^0 \rightarrow \mu^+\mu^-$ decays in the study of the SM and it can be measured through the effective lifetime of $B_s^0 \rightarrow \mu^+\mu^-$ decays as described in Section 2.4. The SM predictions for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions and the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime are given in Section 2.5 and the ways in which NP can influence these observables is briefly discussed in Section 2.6.

2.1 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in the Standard Model

In the SM, quarks and anti-quarks can be combined in pairs to form mesons that are held together by the strong force. The neutral B mesons, B^0 and B_s^0 , consist of a \bar{b} quark combined with a d quark for the B^0 and an s quark for the B_s^0 . Their anti-particles, \bar{B}^0 and \bar{B}_s^0 , are formed by swapping over which quark in the pair is the anti-quark. These particles are unstable and exist for $\sim 10^{-12}$ s before decaying into leptons, lighter mesons or a combination of both. One decay mode is when

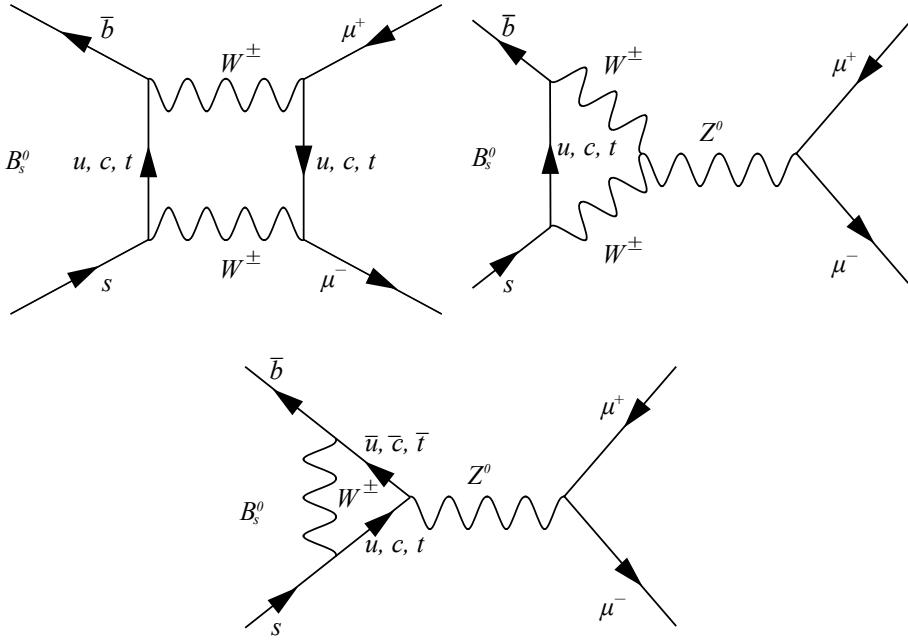


Fig. 2.1 Feynman diagrams for $B_s^0 \rightarrow \mu^+ \mu^-$ decays in the SM via W -box and Z^0 -penguin processes. The same diagrams apply to $B^0 \rightarrow \mu^+ \mu^-$ decays where the s quark is exchanged for a d quark.

the $B_{(s)}^0$ decays into two oppositely charged muons as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$. These decay modes occur very rarely in the SM compared to other decay modes of the $B_{(s)}^0$. The fraction of B^0 and B_s^0 mesons that decay into two muons is $\sim 10^{-10}$ and $\sim 10^{-9}$, respectively. The suppression of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays comes from three sources that are described in the following.

The first source of suppression is due to the quarks that form the B^0 and B_s^0 mesons. The composite quarks of a $B_{(s)}^0$ both have the same charge magnitude, therefore in the decay $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ only quark flavour and not quark charge changes. This type of decay is called a flavour changing neutral current (FCNC). These decays must proceed via the weak force because it is the only interaction in which quark flavour is not conserved via the exchange of a W boson. However, FCNCs are forbidden in the SM to occur at the tree level by the GIM mechanism [4]. Therefore $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays proceed via more complex diagrams such as W -box and Z^0 -penguin diagrams as shown in Figure 2.1. The decays can also proceed via Higgs-penguin diagrams however the contributions from these diagrams are negligible [52]. The lack of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at the tree level causes them to be suppressed compared to other $B_{(s)}^0$ decay modes that can occur at the tree level.

Although the weak force allows quark flavour to change, the coupling strengths between different quark flavours are not all the same magnitude. This leads to the second source of suppression for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. The coupling strengths of different quarks are described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [53,

54]. Quarks can be separated into two types depending on their charge: up-type quarks include u , c and t ; and down-type quarks include d , s and b . The weak force couples all up-type quarks to the weak eigenstate of the down-type quark in the same family with the same strength, where the quark families are; u and d , c and s , t , and b . The weak quark eigenstates are not the same as the mass eigenstates and the two types of states are related via the CKM matrix as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \mathbf{V}_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (2.1)$$

where d' , s' and b' are weak eigenstates and d , s and b are mass eigenstates. The CKM matrix is a unitary matrix with complex elements that ensure no tree level FCNCs occur. Each element of the matrix gives the coupling strengths of transitions between the mass eigenstates of quarks, for example the amplitude of a u quark changing into a d quark is proportional to $|V_{ud}|$.

The difference in the coupling strength sizes can be illustrated through the Wolfenstein parametrisation of the CKM matrix [55], which parametrises the matrix elements in powers of the small parameter of $\lambda = 0.22 \approx |V_{us}|$. The CKM matrix then becomes

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (2.2)$$

This parametrisation shows that the CKM matrix is almost diagonal. For a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay to occur, a single off-diagonal element is needed to describe the quark transitions in Figure 2.1, thus introducing an additional source of suppression to the decay. The internal quark lines in Figure 2.1 can have contributions from u , c and t quarks. However, in the SM the contributions from u and c quarks are negligible when compared to the t quark. This is due to the large t quark mass compared to the other quarks, as given in Table 1.1, and because the coupling strength of the b quark to any quark except the t is extremely small due to additional powers of λ in Equation 2.2.

The final source of suppression of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays comes from the helicities of the muons in the final state. Both B^0 and B_s^0 are spin zero particles and for angular momentum to be conserved in the decay the spins of the two muons must be oppositely aligned. This leads to the muons having opposite helicities. The weak force only couples to left-handed particle states and right-handed anti-particle states.

In the high energy limit where particles are massless, negative helicity states are equal to left-handed states and positive helicity states are equal to right-handed states. Therefore if the muons were massless the weak interaction could only produce a μ^- and a μ^+ with opposite helicities which cannot conserve angular momentum in $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. Muons are not massless, therefore $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays can occur but are suppressed because $m_\mu \ll M_{B_{(s)}}$ [5], leading to one of the helicity states of the muons always being disfavoured.

Overall $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are highly suppressed within the framework of the SM compared to other decay modes of $B_{(s)}^0$ mesons. Therefore these decays offer excellent processes in which to search for NP because the contribution of BSM theories to these decay rates can be of a similar order of magnitude to those from the SM.

2.2 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ Branching Fraction

The branching fraction of a particle decay offers an excellent observable through which predictions of the SM can be compared to measured values. The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction is defined as the fraction of the total number of $B_{(s)}^0$ particles that decay into two muons. It can be calculated from the time-dependant decay rate, $\Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-)$, which is the probability per unit of time that a $B_{(s)}^0$ decays into two muons, as

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-) \equiv \frac{\Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-) + \Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+\mu^-)}{\Gamma(B_{(s)}^0) + \Gamma(\bar{B}_{(s)}^0)}. \quad (2.3)$$

where $\Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+\mu^-)$ is defined in an analogous way to $\Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-)$, $\Gamma(B_{(s)}^0)$ is the total decay rate for $B_{(s)}^0$ mesons and $\Gamma(\bar{B}_{(s)}^0)$ is the total decay rate for $\bar{B}_{(s)}^0$ mesons. The SM predictions are calculated from the ‘prompt’ decay rate which ignores any evolution with time of $B_{(s)}^0$ mesons. The branching fractions are calculated using [56]

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)_{\text{th}} = \frac{\tau_{B_{(s)}}}{2} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-) \rangle \Big|_{t=0}, \quad (2.4)$$

where $\tau_{B_{(s)}}$ is the mean lifetime of the $B_{(s)}^0$ and $\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-) \rangle$ is the untagged decay rate, defined as

$$\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-) \rangle = \Gamma(B_{(s)}^0(t) \rightarrow \mu^+\mu^-) + \Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+\mu^-). \quad (2.5)$$

The untagged decay rate makes no distinction between the particle and anti-particle decays, and is accessible by experiments provided $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ are produced in equal

numbers. The branching fractions are calculated this way to enable easy comparison of different B meson branching fractions including B^0 , B_s^0 and B^+ [56].

The prompt decay rate is evaluated from Fermi's golden rule, relating the transition amplitude, $|\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|$, and the kinematics of the decay to the decay rate as [57]

$$\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \Big|_{t=0} = \frac{1}{16\pi} \frac{1}{M_{B_{(s)}^0}} \sqrt{1 - 4 \left(\frac{m_\mu}{M_{B_{(s)}^0}} \right)^2} |\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|^2, \quad (2.6)$$

where m_μ and $M_{B_{(s)}^0}$ are the masses of the muon and the $B_{(s)}^0$, respectively. The factor $m_\mu/M_{B_{(s)}^0}$ is from the helicity suppression discussed in Section 2.1.

Weak decays like $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ include interactions that occur at different energy scales, from the weak propagators at $M_W \approx 80$ GeV/ c^2 to the strong coupling in the $B_{(s)}^0$ meson at $\Lambda_{QCD} \sim 0.2$ GeV [5]. The Operator Product Expansion [58, 59] is used to create the Effective Hamiltonian, \mathcal{H}_{eff} , which splits the interaction into two very different energy levels. The transition amplitude then becomes

$$|\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)| \equiv \langle \mu\mu | \mathcal{H}_{\text{eff}} | B_{(s)}^0 \rangle = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i \langle \mu\mu | \mathcal{C}(\lambda)_i \mathcal{O}(\lambda)_i | B_{(s)}^0 \rangle \quad (2.7)$$

where G_F is the Fermi coupling constant, V_{CKM}^i are CKM matrix elements, C_i are Wilson coefficients, \mathcal{O}_i are local operators and i is the sum over all possible Wilson coefficient and operator pairs. The energy scale λ separates the two energy levels in the interaction. The Wilson coefficients describe short scale processes with energies above λ . This incorporates the internal structure and loops of Feynman diagrams leading to the dependence of Wilson coefficients on the W^\pm , Z^0 , H^0 and t quark masses. The long distance processes are described by the local operators \mathcal{O}_i for energies less than λ . The local operators link the initial and final states of the decay. Wilson coefficients can be calculated using perturbation theory. However, this cannot be used for the local operators which can lead to large theoretical uncertainties on their values. The choice of λ is arbitrary; however the final transition amplitude must be independent of λ . Often the mass of the decaying particle is used.

In the Effective Hamiltonian in Equation 2.7 the CKM matrix elements are factored out of the Wilson coefficients and operators, so the same coefficients and operators can be used to describe both the B^0 and B_s^0 decays. The Effective Hamiltonian for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is [60]

$$\mathcal{H}_{\text{eff}} = -\frac{G_F \alpha}{\sqrt{2\pi}} V_{tq}^* V_{tb} \sum_i^{10,S,P} (\mathcal{C}_i \mathcal{O}_i + C'_i \mathcal{O}'_i), \quad (2.8)$$

where α is the fine structure constant and q corresponds to the d quark in the B^0 or the s quark in the B_s^0 . The terms proportional to $V_{cq}^*V_{cb}$ and $V_{cq}^*V_{ub}$ are negligible and can be neglected in Equation 2.8. The operators that can contribute to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ Effective Hamiltonian due to the initial and final decay states are

$$\mathcal{O}_{10} = (\bar{q}\gamma^\mu P_L b)(\bar{l}\gamma_\mu\gamma_5 l), \quad \mathcal{O}'_{10} = (\bar{q}\gamma^\mu P_R b)(\bar{l}\gamma_\mu\gamma_5 l), \quad (2.9)$$

$$\mathcal{O}_S = m_b(\bar{q}P_R b)(\bar{l}l), \quad \mathcal{O}'_S = m_b(\bar{q}P_L b)(\bar{l}l), \quad (2.10)$$

$$\mathcal{O}_P = m_b(\bar{q}P_R b)(\bar{l}\gamma_5 l), \quad \mathcal{O}'_P = m_b(\bar{q}P_L b)(\bar{l}\gamma_5 l). \quad (2.11)$$

The operator \mathcal{O}_{10} encompasses the only significant contributions in the SM that come from W -box and Z^0 penguin diagrams. The operator \mathcal{O}'_{10} describes the equivalent interactions as \mathcal{O}_{10} but for right handed currents that are forbidden in the SM. Finally, the operators \mathcal{O}'_S and \mathcal{O}'_P correspond to the exchange of scalar and pseudo-scalar particles which are negligible in the SM.

The purely leptonic final state $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays means that the computation of the transition amplitude can be split in two so that all uncertainties arising from the bound $B_{(s)}^0$ states are encompassed into one parameter, $F_{B_{(s)}}$, the hadronic decay factor. This leads to a theoretically clean prediction for the branching fraction.

The branching fractions for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays can therefore be written as [61]

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-) = \frac{\tau_{B_{(s)}} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} \left| \mathcal{C}_{10}^{\text{SM}} V_{tq}^* V_{tb} \right|^2 F_{B_{(s)}} M_{B_{(s)}^0} m_\mu^2 \times \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}}^2} (|P|^2 + |S|^2)}, \quad (2.12)$$

where θ_W is the weak mixing angle and M_W the mass of the W boson. The branching fraction has been parametrised in terms of $\mathcal{C}_{10}^{\text{SM}}$, P and S , where $\mathcal{C}_{10}^{\text{SM}}$ is the SM value of the operator \mathcal{C}_{10} and

$$P \equiv |P| e^{i\varphi_P} \equiv \frac{\mathcal{C}_{10} - \mathcal{C}'_{10}}{\mathcal{C}_{10}^{\text{SM}}} + \frac{M_{B_{(s)}^0}^2}{2m_\mu} \frac{m_b}{m_b + m_q} \frac{\mathcal{C}_P - \mathcal{C}'_P}{\mathcal{C}_{10}^{\text{SM}}}, \quad (2.13)$$

and

$$S \equiv |S| e^{i\varphi_S} \equiv \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}^0}^2}} \frac{M_{B_{(s)}^0}^2}{2m_\mu} \frac{m_b}{m_b + m_q} \frac{\mathcal{C}_S - \mathcal{C}'_S}{\mathcal{C}_{10}^{\text{SM}}}. \quad (2.14)$$

In the SM P and S are real, with the magnitudes $P = 1$ and $S = 0$, although the branching fractions are parametrised in terms of P and S because BSM theories can significantly alter their values. The presence of scalar particles could increase the branching fractions above the SM expectation through $\mathcal{C}_S^{(')}$ leading to $S > 0$. Furthermore, the contributions from scalar particles are not subject to helicity

constraints. Additionally, pseudoscalar particles can either enhance or suppress the branching fractions compared to the SM prediction depending on how the values of $\mathcal{C}_P^{(')}$ interfere with $\mathcal{C}_{10}^{(')}$ in NP models.

As well as the individual branching fractions of $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays, the ratio of the two branching fractions is also an interesting observable to test the flavour structure of the SM. In the ratio the dependence of the branching fractions on Wilson coefficients cancels, leaving it dependant on the elements $|V_{td}|$ and $|V_{ts}|$ of the CKM matrix. The ratio of branching fractions is given by

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{\tau_B}{\tau_{B_s}} \left| \frac{V_{td}}{V_{ts}} \right|^2 \frac{M_B^2}{M_{B_s}^2} \sqrt{\frac{1 - \frac{4m_\mu^2}{M_B^2}}{1 - \frac{4m_\mu^2}{M_{B_s}^2}}}. \quad (2.15)$$

Therefore the ratio provides a test of the coupling strengths of different quarks types as described by the CKM matrix in the SM and could reveal whether or not BSM theories follow this structure as well. An additional advantage of using the ratio of branching fractions is that the theoretical uncertainty on the ratio is less than the individual branching fractions because sources of uncertainties including those from Wilson coefficients and $|V_{tb}|$ cancel.

2.3 Quark mixing

The theoretical prediction for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions described in Section 2.2 does not take into account the evolution of the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ mesons with time. Once a $B_{(s)}^0$ is created it will oscillate between the particle and anti-particle states as it propagates through time, the same is true for the $\bar{B}_{(s)}^0$. Therefore the states that travel through time are a superposition of the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$. Oscillations occur as the constituents quarks transition between different flavours through the exchange of W bosons as illustrated in Figure 2.2. The branching fractions are measured from data where $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ decays are not separated, which is called an untagged sample of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Since a $B_{(s)}^0$ or $\bar{B}_{(s)}^0$ lives for $\sim 10^{-12}$ s before decaying, the state that decays will not necessarily be the same as the one that was produced. The measured branching fraction is not the same as the ‘prompt’ branching fraction used for the theoretical prediction. The measured value corresponds to the time integrated branching fraction given by [56]

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} \equiv \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle dt. \quad (2.16)$$

Therefore for a meaningful comparison between the measured and predicted branching fraction values, the difference in the two definitions must be evaluated [56, 60, 61].

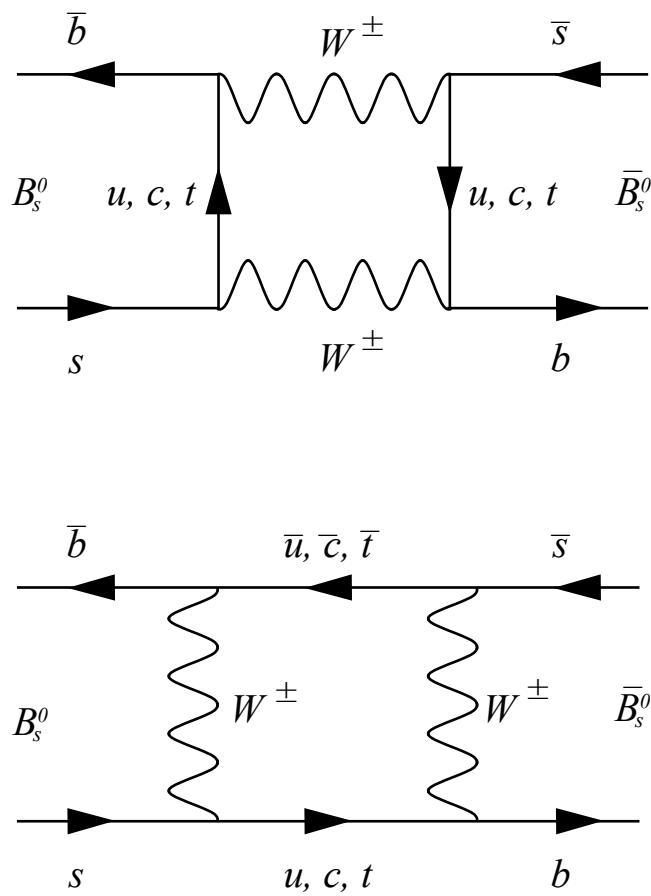


Fig. 2.2 Oscillation of B_s^0 and \bar{B}_s^0 quarks through the exchange of W bosons. The same diagrams apply to B^0 and \bar{B}^0 oscillations but with the s quark exchanged for a d quark.

2.3.1 Time evolution of the $B_{(s)}^0$

In pp collisions, initially each b and \bar{b} quark hadronises to form a $B_{(s)}^0$ or $\bar{B}_{(s)}^0$ described at $t = 0$ by the states $|B_{(s)}^0\rangle$ and $|\bar{B}_{(s)}^0\rangle$. The time evolution of these states must be evaluated in order to determine the time integrated branching fractions. The Time-Dependent Schrödinger Equation (TDSE) describes the time evolution of the particle and anti-particle states as

$$i\frac{d}{dt} \begin{pmatrix} |B_{(s)}^0(t)\rangle \\ |\bar{B}_{(s)}^0(t)\rangle \end{pmatrix} = \left(\mathbf{M} - \frac{i\Gamma}{2} \right) \begin{pmatrix} |B_{(s)}^0(t)\rangle \\ |\bar{B}_{(s)}^0(t)\rangle \end{pmatrix}, \quad (2.17)$$

where \mathbf{M} and Γ are 2×2 Hermitian matrices describing mass and decay time with the properties $M_{12}^* = M_{21}$ and $\Gamma_{12}^* = \Gamma_{21}$. Invariance under charge, parity and time inversion introduces additional constraints of $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$.

The $B_{(s)}^0$ - $\bar{B}_{(s)}^0$ oscillations ensure that for any $t > 0$ the particles are a superposition of $|B_{(s)}^0\rangle$ and $|\bar{B}_{(s)}^0\rangle$ states. The off-diagonal elements in the mass and decay time matrices mean that the eigenstates of the TDSE have different masses and lifetime to the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$. The eigenstates can be defined at $t = 0$ as

$$|B_H\rangle = p|B_{(s)}^0\rangle - q|\bar{B}_{(s)}^0\rangle, \quad |B_L\rangle = p|B_{(s)}^0\rangle + q|\bar{B}_{(s)}^0\rangle \quad (2.18)$$

with eigenvalues of $(m_{H,L} - i\Gamma_{H,L}/2)$ and the coefficients p and q are constrained by $|p|^2 + |q|^2 = 1$. The eigenstates are known as heavy, $|B_H\rangle$, and light, $|B_L\rangle$, mass eigenstates. The eigenvalues are different for the B^0 and B_s^0 systems, however the treatment of the two systems is identical. To simplify the notation only the B_s^0 system will be described in the following discussion. The time evolution of the heavy and light mass eigenstates is given by

$$|B_H(t)\rangle = |B_H\rangle e^{-i(m_H - i\Gamma_H/2)t}, \quad |B_L(t)\rangle = |B_L\rangle e^{-i(m_L - i\Gamma_L/2)t} \quad (2.19)$$

from the TDSE. Therefore, the time evolution of the flavour states can now be determined from Equations 2.18 and 2.19 as

$$|B_s^0(t)\rangle = \frac{1}{2p} (|B_L(t)\rangle + |B_H(t)\rangle) = f_+(t)|B_s^0\rangle + \frac{q}{p}f_-(t)|\bar{B}_s^0\rangle \quad (2.20)$$

$$|\bar{B}_s^0(t)\rangle = \frac{1}{2q} (|B_L(t)\rangle - |B_H(t)\rangle) = \frac{p}{q}f_-(t)|B_s^0\rangle + f_+(t)|\bar{B}_s^0\rangle \quad (2.21)$$

where

$$f_{\pm} = \frac{1}{2} e^{-i(m_s - i\Gamma_s)t} \left\{ e^{i(\Delta m_s + i\Gamma_s)t/2} \pm e^{-i(\Delta m_s + i\Gamma_s)t/2} \right\}. \quad (2.22)$$

The relationships

$$m_s \equiv \frac{m_H + m_L}{2}, \quad \Delta m_s \equiv m_H - m_L, \quad (2.23)$$

$$\Gamma_s \equiv \frac{(\Gamma_H + \Gamma_L)}{2}, \quad \Delta\Gamma_s \equiv \Gamma_L - \Gamma_H, \quad (2.24)$$

have been used in the expressions of $|B_s^0(t)\rangle$ and $|\bar{B}_s^0(t)\rangle$. The difference Δm_s is defined so that it is always positive whereas $\Delta\Gamma_s$ can take either sign. The time evolution is written in terms of these variables because Δm_s and $\Delta\Gamma_s$ are measurable quantities.

Theoretical predictions can be calculated for M_{12} and Γ_{12} and it is therefore useful to express the measurable quantities in terms of these parameters. This is done by solving the characteristic equation of the TDSE, $|\mathbf{M} - i\mathbf{\Gamma}/2 - \lambda\mathbf{I}| = 0$, which has the solutions

$$\Delta m^2 - \frac{\Delta\Gamma^2}{4} = 4(|M_{12}|^2 - \frac{1}{4}|\Gamma_{12}|^2) \quad (2.25)$$

$$\Delta m\Delta\Gamma = 4|\Gamma_{12}||M_{12}|\cos\phi \quad (2.26)$$

where $\phi \equiv \arg(-M_{12}/\Gamma_{12})$.

The observed relationships $\Delta\Gamma \ll \Delta m$ and $\Gamma_{12} \ll M_{12}$ [51] are used to separate the expressions for Δm and $\Delta\Gamma$ to give

$$\Delta m = 2|M_{12}|\left(1 + \mathcal{O}\left(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\right)\right) \quad (2.27)$$

and

$$\Delta\Gamma = 2|\Gamma_{12}|\cos\phi\left(1 + \mathcal{O}\left(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\right)\right). \quad (2.28)$$

The values of p and q can also be related to the measurable quantities and Γ_{12} and M_{12} by diagonalising $(\mathbf{M} - \frac{i}{2}\mathbf{\Gamma})$ to produce

$$\frac{q}{p} = -\frac{\Delta m_s^2 + i\Delta\Gamma_s/2}{2M_{12} - i\Gamma_{12}} \approx -e^{-i\phi_M}\left(1 - \frac{a}{2}\right) + \mathcal{O}\left(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\right) \quad (2.29)$$

where $\phi_M \equiv \arg(M_{12}/|M_{12}|)$ and $a \equiv |\Gamma_{12}/M_{12}|\sin\phi$ and the relationships $\Delta\Gamma \ll \Delta m$ and $\Gamma_{12} \ll M_{12}$ have been used. The value of ϕ_M is related to the elements of the CKM matrix and $\phi_M = \arg(V_{tb}^*V_{td})$ for the B^0 and $\phi_M = \arg(V_{tb}^*V_{ts})$ for the B_s^0 . The ratio of p and q is given in terms of the small parameter a which is needed to evaluate some SM processes but can be neglected in this process.

The necessary parameters used to describe the time evolution of $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ states have now been expressed in terms of measurable or predictable quantities, and therefore the time-dependant decay rates can now be evaluated. The decay rates

can be expressed as

$$\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) = \mathcal{N} |\langle \mu\mu | B_{(s)}^0 \rangle|^2 \quad (2.30)$$

and

$$\Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+ \mu^-) = \mathcal{N} |\langle \mu\mu | \bar{B}_{(s)}^0 \rangle|^2, \quad (2.31)$$

where \mathcal{N} encompasses the additional terms in Equation 2.6 from kinematic parameters. For the evaluation of the time-dependant decay rates, the exact form of the transition amplitude is not needed. A new parameter is defined

$$\lambda_{\mu\mu} = \frac{q}{p} \left| \frac{\bar{A}_{\mu\mu}}{A_{\mu\mu}} \right| \quad (2.32)$$

where $A_{\mu\mu} = \langle \mu^+ \mu^- | B_s^0 \rangle$ and $\bar{A}_{\mu\mu} = \langle \mu^+ \mu^- | \bar{B}_s^0 \rangle$ to simplify the decay rate expression. Combining the information in Equations 2.20, 2.21 and 2.31, and using $\lambda_{\mu\mu}$ the time-dependant decay rates are

$$\begin{aligned} \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) = & \frac{1}{2} \mathcal{N} |A_{\mu\mu}|^2 e^{-\Gamma_s t} \left\{ (1 + |\lambda_{\mu\mu}|^2) \cosh \left(\frac{\Delta\Gamma_s t}{2} \right) \right. \\ & + (1 - |\lambda_{\mu\mu}|^2) \cos(\Delta m_s t) - 2\Re(\lambda_{\mu\mu}) \sinh \left(\frac{\Delta\Gamma_s t}{2} \right) \\ & \left. - 2\Im(\lambda_{\mu\mu}) \sin(\Delta m_s t) \right\} \end{aligned} \quad (2.33)$$

and

$$\begin{aligned} \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-) = & \frac{1}{2} \mathcal{N} (1 + a) |A_{\mu\mu}|^2 e^{-\Gamma_s t} \left\{ (1 + |\lambda_{\mu\mu}|^2) \cosh \left(\frac{\Delta\Gamma_s t}{2} \right) \right. \\ & - (1 - |\lambda_{\mu\mu}|^2) \cos(\Delta m_s t) - 2\Re(\lambda_{\mu\mu}) \sinh \left(\frac{\Delta\Gamma_s t}{2} \right) \\ & \left. + 2\Im(\lambda_{\mu\mu}) \sin(\Delta m_s t) \right\}. \end{aligned} \quad (2.34)$$

The time integrated branching fraction depends on the sum of the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ time-dependant decay rates. Using Equation 2.34 and ignoring terms $\mathcal{O}(a)$, the total decay rate is

$$\langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle = \mathcal{N} |A_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) e^{-\Gamma_s t} \left(\cosh \left(\frac{\Delta\Gamma_s t}{2} \right) + A_{\Delta\Gamma} \sinh \left(\frac{\Delta\Gamma_s t}{2} \right) \right). \quad (2.35)$$

The parameter, $A_{\Delta\Gamma}$, has been introduced into the total decay rate and it is defined as

$$A_{\Delta\Gamma} = \frac{2\Re(\lambda_{\mu\mu})}{1 + |\lambda_{\mu\mu}|^2}. \quad (2.36)$$

The meaning of $A_{\Delta\Gamma}$ can be seen when the total decay rate is written in terms of the heavy and light $B_{(s)}^0$ mass eigenstates as

$$\begin{aligned} \langle \Gamma(B_s^0(t) \rightarrow \mu^+\mu^-) \rangle &= \mathcal{N}|A_{\mu\mu}|^2(1 + |\lambda_{\mu\mu}|^2) \left((1 - A_{\Delta\Gamma})e^{-\Gamma_L t} + (1 + A_{\Delta\Gamma})e^{-\Gamma_H t} \right) \\ &= R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t}. \end{aligned} \quad (2.37)$$

This final expression for the decay rates shows how $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays can be described in terms of the sum of the decays of the heavy and light mass eigenstates. The parameter $A_{\Delta\Gamma}$ is therefore related to the number of heavy and light mass eigenstates that decay as

$$A_{\Delta\Gamma} = \frac{R_H - R_L}{R_H + R_L}. \quad (2.38)$$

The values $A_{\Delta\Gamma}$ can take range from +1 when only heavy mass eigenstates decay as $B_s^0 \rightarrow \mu^+\mu^-$ and -1 when only light mass eigenstates decay as $B_s^0 \rightarrow \mu^+\mu^-$. The same is true for the $B^0 \rightarrow \mu^+\mu^-$ decay.

2.3.2 Impact on the Branching Fraction

The time-dependent decay rates are used to understand the difference between the two branching fraction definitions. The final form of the decay rates in Equation 2.37 is used in the evaluation of the branching fractions. The ‘prompt’ branching fraction used in the theoretical predictions is

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)_{\text{th}} = \frac{\tau_{B_s}}{2} \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle \quad (2.39)$$

$$= \frac{\tau_{B_s}}{2} (R_H + R_L). \quad (2.40)$$

The time integrated branching fraction that is measured is

$$\begin{aligned} \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)_{\text{exp}} &= \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0 \rightarrow \mu^+\mu^-) \rangle dt \\ &= \frac{1}{2} \left(\frac{R_H}{\Gamma_H} + \frac{R_L}{\Gamma_L} \right) \\ &= \frac{\tau_{B_{(s)}}}{2} (R_H + R_L) \left[\frac{1 + A_{\Delta\Gamma} y_{(s)}}{1 - y_{(s)}^2} \right], \end{aligned} \quad (2.41)$$

where $y_{(s)}$ relates the heavy and light mass eigenstate decay times as $y_{(s)} = \Delta\Gamma_{(s)}/2\Gamma_{(s)}$. Therefore the measured and prompt branching fraction values are related as

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)_{\text{exp}} = \left[\frac{1 + A_{\Delta\Gamma}y_{(s)}}{1 - y_{(s)}^2} \right] \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)_{\text{th.}} \quad (2.42)$$

For B^0 - \bar{B}^0 oscillations the difference in the lifetimes of the heavy and light mass eigenstates is extremely small. Therefore y is negligible and the prompt branching fraction is equivalent to the experimental branching fraction. However, for B_s^0 - \bar{B}_s^0 oscillations there is a large difference in the lifetimes of the mass eigenstates and $y_s = 0.062 \pm 0.006$ [36]. The prompt branching fraction must therefore be corrected to account for the oscillations before it is compared to the experimental value.

2.4 $A_{\Delta\Gamma}$ and the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime

The definition of $A_{\Delta\Gamma}$ in Equation 2.36 shows that it depends upon the transition amplitude of $B_s^0 \rightarrow \mu^+\mu^-$ decays. In Section 2.2 the Effective Hamiltonian for this decay was discussed and the branching fraction given in terms of the complex variables P and S . Therefore $A_{\Delta\Gamma}$ can also be expressed in terms of these parameters as [60]

$$A_{\Delta\Gamma} = \frac{|P| \cos \varphi_P + |S| \cos \varphi_S}{|P|^2 + |S|^2}, \quad (2.43)$$

where P and S are defined in Equations 2.13 and 2.14, respectively. In the SM, $P = 1$ and $S = 0$, therefore $A_{\Delta\Gamma}$ takes the maximal value of $+1$ and only the heavy mass eigenstate decays as $B_s^0 \rightarrow \mu^+\mu^-$. This can be understood because the final state of a $B_s^0 \rightarrow \mu^+\mu^-$ decay is a \mathcal{CP} odd state and the heavy B_s^0 mass eigenstate is a \mathcal{CP} odd state.

As discussed in Section 2.2, NP models can alter the values of P and S , moving them away from the SM expectations. A change in these values could alter both the measured branching fraction and $A_{\Delta\Gamma}$ or just one of these observables. Since the comparison of the measured branching fraction to the SM prediction relies on $A_{\Delta\Gamma}$, in order to understand possible NP contributions to the branching fraction, $A_{\Delta\Gamma}$ must be measured as well.

The value of $A_{\Delta\Gamma}$ can be measured directly from the time-dependant decay rate of $B_s^0 \rightarrow \mu^+\mu^-$ decays. This method involves separating the $B_s^0 \rightarrow \mu^+\mu^-$ decays into those with $|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$ initial states, which requires a large number of $B_s^0 \rightarrow \mu^+\mu^-$ decays. Since $B_s^0 \rightarrow \mu^+\mu^-$ are very rare decays this approach is currently not viable. Alternatively $A_{\Delta\Gamma}$ can be measured through the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime [56]. The effective lifetime is the mean decay time of an untagged sample of $B_s^0 \rightarrow \mu^+\mu^-$

decays, defined as [62]

$$\tau_{\mu\mu} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt}. \quad (2.44)$$

It can be measured by fitting a single exponential to the same set of decays used to measure the branching fraction [56]. The effective lifetime can be expressed in terms of $A_{\Delta\Gamma}$ using the decay rates in Equation 2.37 as

$$\tau_{\mu\mu} = \frac{\tau_{B_s}}{1 - y_s^2} \left(\frac{1 + 2A_{\Delta\Gamma}y_s + y_s^2}{1 + A_{\Delta\Gamma}y_s} \right). \quad (2.45)$$

In the SM only the heavy B_s^0 mass eigenstate decays as $B_s^0 \rightarrow \mu^+\mu^-$ and the effective lifetime equals the lifetime of the heavy mass eigenstate, $\tau_{\mu\mu} = \tau_H = \frac{1}{\Gamma_H}$. The effective lifetime offers a measurement complementary to the branching fractions to study the SM and NP models in $B_s^0 \rightarrow \mu^+\mu^-$ decays due to the dependence of the effective lifetime on $A_{\Delta\Gamma}$.

2.5 The Standard Model predictions

The SM provides precise predictions of the time-integrated $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions [63]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9} \quad (2.46)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10} \quad (2.47)$$

where the latest progress in lattice QCD calculations have been taken into account [64–66]. The largest contributions to the theoretical uncertainties come from the CKM matrix elements and the decay constants of the B_s^0 and the B^0 . The ratio of the branching fraction values, defined in Equation 2.15 is [47]

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)} = 0.0295^{+0.0028}_{-0.0025}. \quad (2.48)$$

The $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is determined from Equation 2.45. Using the SM value of $A_{\Delta\Gamma} = +1$, the B_s^0 mean lifetime of $\tau_{B_s} = 1.510 \pm 0.005$ ps and $y_s = 0.062 \pm 0.006$ [5, 36] the SM effective lifetime is

$$\tau_{\mu\mu} = 1.609 \pm 0.010 \text{ ps}, \quad (2.49)$$

which is equal to the lifetime of the heavy B_s^0 mass eigenstate as, in the SM, only the heavy mass eigenstate of the B_s^0 decays into two muons. The same calculation

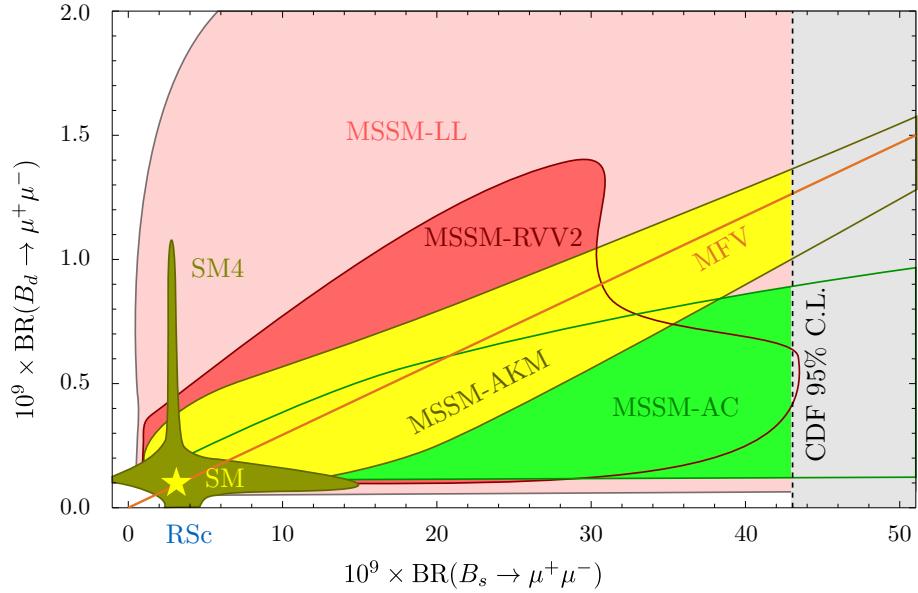


Fig. 2.3 Correlations between the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions including the SM prediction, the Minimal Flavour Violation hypothesis (MFV), four Minimal Supersymmetric Standard Models [70] and the SM extended to constrain four generations of fermions (SM4) [71]. Figure is taken from reference [72].

can be performed for $A_{\Delta\Gamma} = -1$, giving the lifetime of the light mass B_s^0 eigenstate as $\tau_L = 1.414 \pm 0.006$ ps. The difference in these lifetimes is 0.195 ps and therefore a precision of 0.38 ps would be needed to distinguish between $A_{\Delta\Gamma} = +1$ and $A_{\Delta\Gamma} = -1$ a 5σ with the effective lifetime.

2.6 New Physics models and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays

There exist a large number of BSM theories that can influence $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in different ways. Measurements of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions and the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime can constrain the parameter space available for NP and in doing so could reveal which theories are viable extensions of the SM. This section will briefly introduce some NP models that could influence $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays given the current precision of the branching fraction measurements. For a more detailed discussions of NP models relevant to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays and constraints on these models from measurements see references [61, 67, 68].

As discussed in Section 2.2, the ratio of the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ provides an excellent test of the flavour structure of the SM and BSM theories. Figure 2.3 shows possible values accessible by BSM theories alongside the SM prediction. The prediction of the Minimal Flavour Violation (MFV) [69] hypothesis is included. The MFV hypothesis predicts that the coupling of quark flavour and \mathcal{CP} violation follow the same Yukawa structure as the SM in NP models. It is a popular theory to describe the flavour structure in NP models due to the current agreement of

measurements with the SM predictions. A significant deviation of the branching fraction ratio from the SM or MFV hypothesis predictions would indicate the need for a new flavour structure in theoretical models.

Additionally, NP models could move the branching fractions and $A_{\Delta\Gamma}$ away from the SM predictions by providing new particles that could contribute to the decays either through loop diagrams, similar to those in Figure 2.1, or allowing the decays to occur at the tree level. These new particles would change the Wilson coefficients included in the parameters P and S . The dependence of the branching fractions and $A_{\Delta\Gamma}$ on P and S are different, as shown in Equations 2.12 and 2.43, and therefore NP models can influence the observables independently. The allowed values of $A_{\Delta\Gamma}$ and the ratio of the measured branching fraction to prompt SM prediction for $B_s^0 \rightarrow \mu^+\mu^-$ decays are shown in Figure 2.4 for possible situations where $S = 0$, $P = 1$, $P \pm S = 1$ and $\varphi_P, \varphi_S \in \{0, \pi\}$. These figures illustrate that if NP effects are not revealed in the branching fraction measurements, they could still appear in $A_{\Delta\Gamma}$.

Amongst the BSM theories that can influence the values of P and S are the Two Higgs Doublet Model (2HDM) [74], supersymmetric models [75] and models including leptoquarks.

The 2HDM extends the Higgs sector of the SM by introducing two complex scalar field doublets both with non-zero vacuum expectation values. Spontaneous symmetry breaking then produces two charged, one neutral pseudoscalar and two neutral scalar Higgs bosons. The new particles can enter the loops of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays and allow FCNCs to occur at the tree level. Different scenarios of this model depend on the Higgs-quark interactions and can incorporate the MFV hypothesis. This model can produce scenarios where $S = 0$, $P = 1$ or $P \pm S = 1$ [61, 67] and the corresponding branching fraction and $A_{\Delta\Gamma}$ values for $B_s^0 \rightarrow \mu^+\mu^-$ decays as shown in Figures 2.4a, 2.4b and 2.4c.

Supersymmetric (SUSY) models extend the SM by giving each SM particle a supersymmetric partner. The resulting theory is symmetric in terms of the transformation of fermions to bosons and bosons to fermions. So far no evidence for SUSY particles has been found, and therefore the symmetry must be broken and the mass of SUSY particles are greater than their SM partners. The Minimal Supersymmetric Standard Model (MSSM) includes a Higgs sector similar to the 2HDM and there are scenarios where it obeys the MFV hypothesis. $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are sensitive to this model provided the ratio of the vacuum expectation values of the Higgs doublet is large [76–78]. The MSSM can produce values for $A_{\Delta\Gamma}$ and the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction shown in Figure 2.4c and 2.4d for situations where $P \pm S = 1$ and $\varphi_P, \varphi_S \in \{0, \pi\}$ [61, 67].

Models including leptoquarks are currently popular to explain the anomalies observed in heavy flavour measurements [79–83]. A leptoquark is a boson that carries

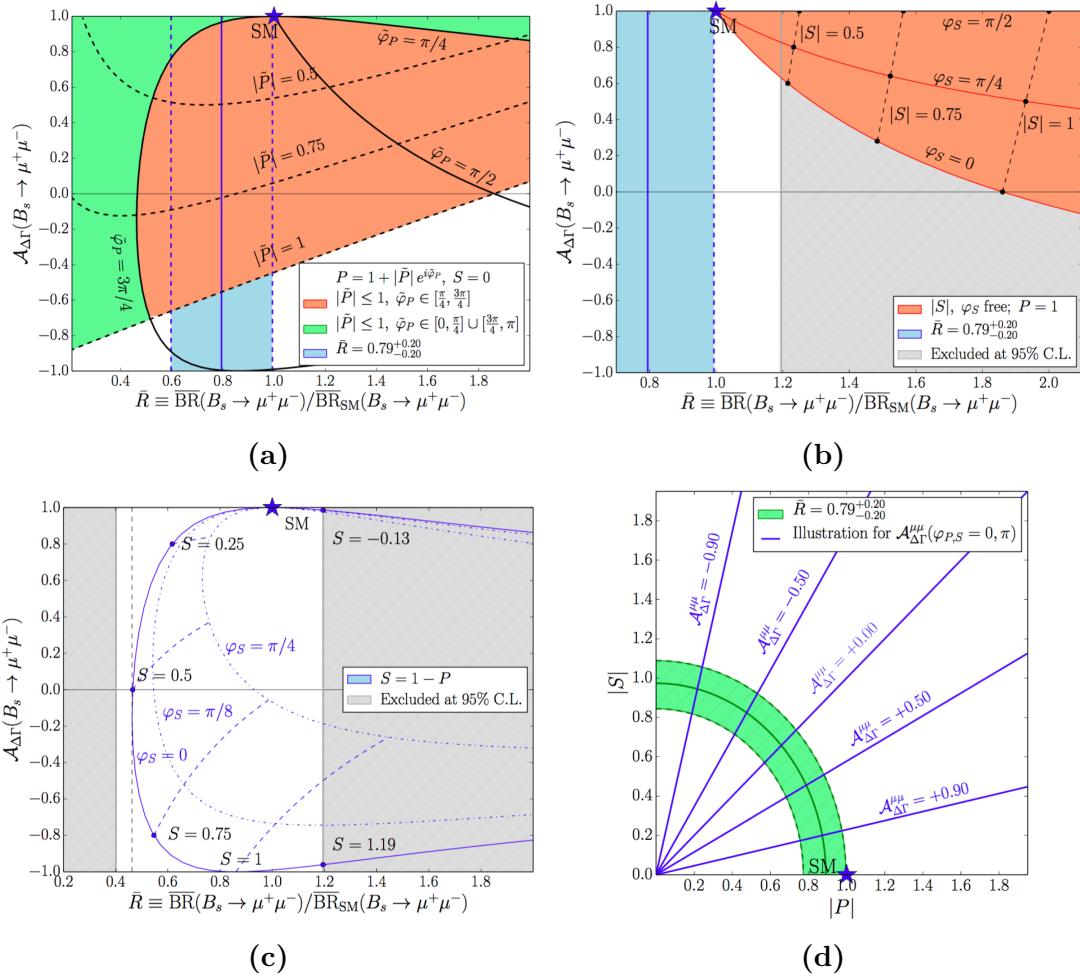


Fig. 2.4 Allowed values for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $A_{\Delta\Gamma}$ for situations where a) $S = 0$, b) $P = 1$, c) $P \pm S = 1$ and d) $\varphi_P, \varphi_S \in \{0, \pi\}$ (d) [61, 67]. The ratio \bar{R} plotted is from an average of the individual results from the CMS and LHCb collaborations from [73], and the results from the combined analysis of the CMS and LHCb data gives $\bar{R} = 0.76^{+0.20}_{-0.18}$ [47].

both lepton and baryons numbers, therefore leptoquarks can allow FCNCs to occur at the tree level. The exact quantum numbers of these particles depend on their interactions with SM fermions. Therefore leptoquarks could enhance $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays but information from $A_{\Delta\Gamma}$ is necessary for the study of leptoquarks with $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays because it resolves degeneracies that are present when only the branching fraction measurements are considered [84].

Although $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are yet to reveal NP, the current experimental precision still leaves plenty of room for NP to be revealed. The observation of $B_s^0 \rightarrow \mu^+\mu^-$ decays makes it possible to start investigating $A_{\Delta\Gamma}$ through the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. A measurement of $A_{\Delta\Gamma}$ will provide important information, complementary to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction measurements in the search for NP in $B_s^0 \rightarrow \mu^+\mu^-$ decays.

Chapter 3

The LHC and the LHCb experiment

The European Organisation for Nuclear Research (CERN) was founded in 1954 and began with 12 member states as an organisation to encourage European collaboration and to study nuclear physics. The collaborative nature of CERN has enabled large-scale expensive experiments to be built that individual member states would not have been able to afford. In 1959 the Proton Synchrotron (PS) was CERN's flagship synchrotron accelerator. It had a circumference of 628 m and accelerated protons up to a centre-of-mass energy of $\sqrt{s} = 25$ GeV, making the PS the highest energy particle accelerator at that time. The PS was succeeded by the Super Proton Synchrotron (SPS) in 1976. The SPS was 7 km in circumference and designed to accelerate protons up to a centre-of-mass energy of $\sqrt{s} = 400$ GeV. The most notable achievements of the SPS were the discoveries of the W and Z^0 bosons in 1983 after the accelerator had been converted into a proton - anti-proton collider. After the SPS, precise measurements of the W and Z^0 bosons were performed using the Large Electron Positron collider (LEP) from 1989. The largest e^+e^- collider built to date, LEP was 27 km in circumference and designed to operate around a centre-of-mass energy of $\sqrt{s} = M_Z$ and M_W .

Now 63 years since its foundation, CERN has grown to include 22 member states¹ and is still at the forefront of high energy physics research. CERN's latest accelerator, the Large Hadron Collider (LHC), is the most energetic particle accelerator yet to be built. It was built in the tunnel that housed LEP and is designed to collide protons at a centre-of-mass energy of $\sqrt{s} = 14$ TeV.

This chapter introduces the LHC and the LHCb experiment, one of the experiments that studies the products of particle collisions produced at the LHC. The LHC is described in Section 3.1 along with the experimental runs that have occurred so

¹Countries and organisations that are unable to become member states can still participate in scientific research as observer states [85].

far. The LHCb experiment and the sub-detectors it is composed of are described in Section 3.2. The study of particles produced in proton-proton (pp) collisions requires information about the passage of charged particles and identity of particles travelling through the detector. The sub-detectors in LHCb experiment that are designed to track particles and identify the particle types are described in Sections 3.2.1 and 3.2.2, respectively. Not all events that occur when protons collide contain particles that are interesting for the study of the SM and the search for NP effects, therefore the LHCb experiment uses a trigger system, presented in Section 3.2.3, to identify interesting events that are saved to be later analysed. The analysis of data recorded by the experiment requires the use of custom software packages which are described in Section 3.2.4. Finally, the data recorded by the LHCb experiment and used for the analyses described in this dissertation are given in Section 3.3.

3.1 The Large Hadron Collider

The LHC is a proton synchrotron designed to accelerate and collide two beams of protons with a centre-of-mass energy of 14 TeV. Although operation of the LHC began in 2010 it is yet to reach the design energy. The purpose of the LHC is to provide high energy pp collisions, the products of which are used for precision tests of the SM and to search for NP effects that cannot be explained within the context of the SM. There are four interaction points on the LHC ring where the beams are collided, at these points various experiments detect and study the products of particle collisions. In addition to protons, the LHC can also accelerate lead-nuclei up to 2.76 TeV per nucleon. It is only the products from pp collisions that are studied in this dissertation.

The protons accelerated by the LHC originate from hydrogen gas, the hydrogen atoms are ionised to strip away the electrons. The protons are then accelerated through a chain of particle accelerators of increasing energy before being injected into the LHC. The chain of accelerators, shown in Figure 3.1, consists of machines that were used in experiments throughout the second half of the last century and have been modified to meet the requirements needed to provide protons to the LHC. The protons leave the chain of accelerators with an energy of 450 GeV per proton and in bunches of $\sim 10^{11}$ protons. As the bunches are injected into the LHC they are split into two oppositely circulating beams. The LHC accelerates the protons to the desired centre-of-mass energy using sixteen radio frequency cavities and guides them around the ring with 1232 superconducting dipole magnets. Once the required energy has been reached, the bunches are focused using 392 quadrupole magnets before being collided at 4 interaction points around the LHC ring.

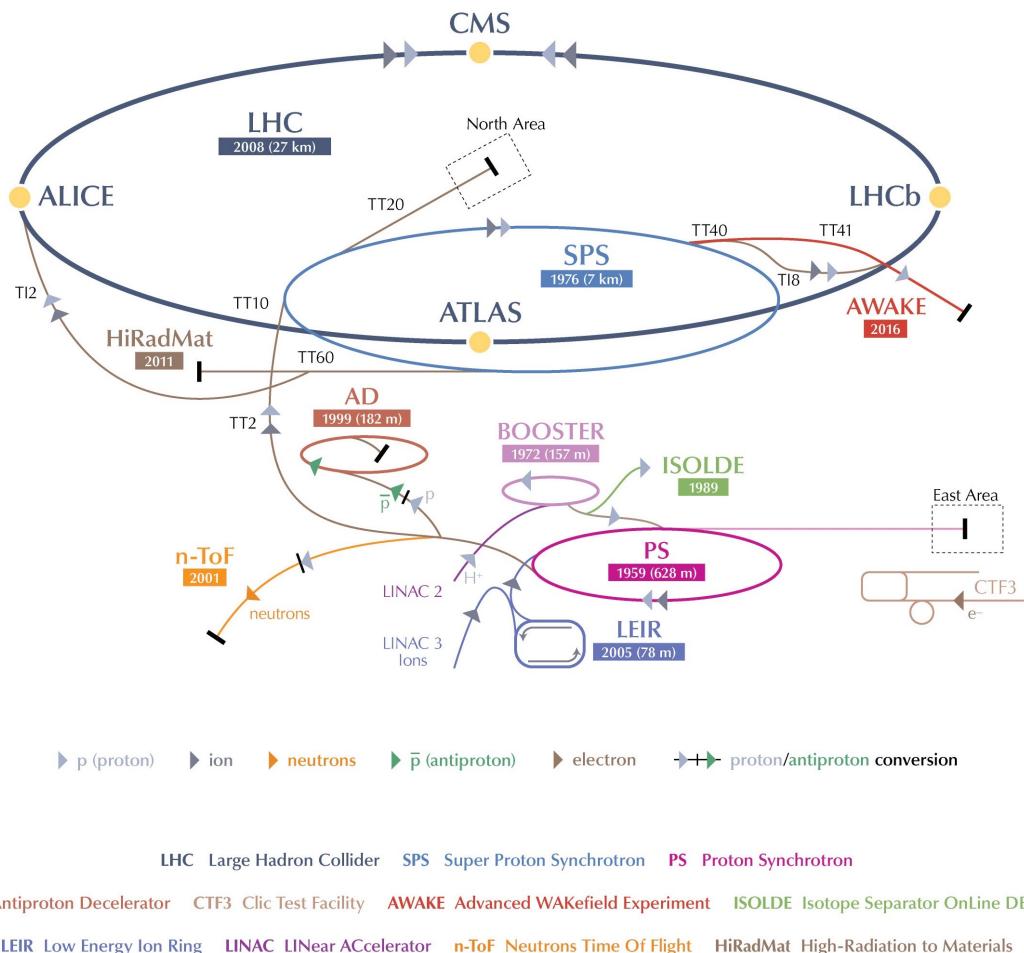


Fig. 3.1 The accelerator complex at CERN. The chain of accelerators used to provide protons to the LHC begins with the Linac 2 which accelerates protons to 50 MeV, the protons are passed to the Proton Synchrotron Booster that accelerates them to 1.4 GeV. The Proton Synchrotron is next in the chain, accelerating protons to 25 GeV and creating the desired spacing between proton bunches. Then finally the Super Proton Synchrotron accelerates protons to 450 GeV ready for injection into the LHC. Source: CERN.

The centre-of-mass energy of a collider is an important measure of its performance as it describes the energy available to create particles during pp collisions. Another important measure of collider performance is the instantaneous luminosity a collider can provide. The instantaneous luminosity, \mathcal{L} , is a measure of how many collisions occur per second, and is given by

$$\mathcal{L} = \frac{N^2 f n_b}{\mathcal{F}}, \quad (3.1)$$

where N is the number of protons per bunch, n_b the number of bunches per beam, f the bunch revolution frequency and \mathcal{F} contains information about the beam geometry. The LHC is designed to operate at a maximum instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. To reach this luminosity the LHC can have up to 2808 proton bunches per beam with a revolution frequency of 11.245 kHz and a separation of 25 ns between each proton bunch. The higher the luminosity, the more collisions happen in a second and the more particles can be produced, this is either advantageous or disadvantageous depending on the particle interactions that are being studied. Therefore the luminosity delivered at each interaction point can be tuned using the quadrupole magnets by altering the shape and the orbit of each bunch to suit the experiments at each interaction point.

There are seven experiments around the LHC that study particles produced in proton and heavy ion collisions at the interaction points. These detectors are:

- ATLAS [86] and CMS [87], are general purpose detectors designed to search for the Higgs boson and NP effects that are beyond the scope of the SM. These experiments operate at the full instantaneous luminosity of the LHC;
- ALICE [88] studies quark-gluon plasma produced in heavy ion collisions to understand conditions similar to those present in the early universe, products from pp and lead-proton collisions are used to understand collisions of lead ions;
- TOTEM [89] studies properties of protons as they collide head on at the LHC, to measure the pp interaction cross section and understand the internal structure of protons;
- MOEDAL [90] is designed to detect magnetic monopoles and massive stable charged particles predicted by theories that go beyond the scope of the SM;
- the LHCf experiment [91] studies particles that are created at very small angles to the incident proton beams to understand similar processes that occur in cosmic rays; and

Run	Year	\sqrt{s} / TeV	Integrated luminosity / fb^{-1}		
			ATLAS	CMS	LHCb
Run 1	2010	7	0.04	0.04	0.04
	2011	7	5.08	5.55	1.11
	2012	8	21.3	21.79	2.08
Run 2	2015	13	3.9	3.81	0.32
	2016	13	35.6	37.76	1.67

Table 3.1 Centre-of-mass energy for each year of data taking and the integrated luminosity recorded by ATLAS, CMS and LHCb during pp collisions [93–95].

- the final experiment is the LHCb experiment [92] is designed to study rare b -hadron decays and \mathcal{CP} violating processes. It operates at a lower luminosity and has a smaller angular acceptance than the general purpose detectors.

Proton beams first circulated the LHC in 2008 and since then the experiments have recorded data during two experiment runs separated by a long shutdown period. The amount of data recorded by an experiment is given in terms of the time integrated luminosity. Run 1 began in 2010 and continued until 2013, during this time protons were collided with a centre-of-mass energies of $\sqrt{s} = 7$ and $\sqrt{s} = 8$ TeV. After Run 1 there was a period of long shut down during which work was done to prepare the LHC to operate at higher energies, renovation work was performed on the accelerators that provide the LHC with protons and the experiments were prepared for the second experiment run. Run 2 began in 2015 with proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, and will continue until 2018 when a second period of upgrades and maintenance will begin. The centre-of-mass energies for pp collisions for each year of operation so far are summarised in Table 3.1, alongside the integrated luminosity recorded by the ATLAS, CMS and LHCb experiments.

3.2 The LHCb experiment

The LHCb experiment was built to study the SM and search for NP phenomena through the study of \mathcal{CP} -violating decays and rare b -hadron decays. At the LHC the dominant production mechanisms of $b\bar{b}$ pairs are gluon-gluon fusion, quark anti-quark annihilation and gluon-gluon splitting, as shown in Figure 3.2. The $b\bar{b}$ pairs produced travel at small angles relative to the beam pipe and the angular distribution is shown in Figure 3.3. The $b\bar{b}$ pair is created at the primary vertex of an event, where the protons collide. The quarks then hadronise to form a range of b -hadrons, including B^+ , B_s^0 and Λ_b^0 . A typical b hadron will exist for approximately 1.5 ps [5] before

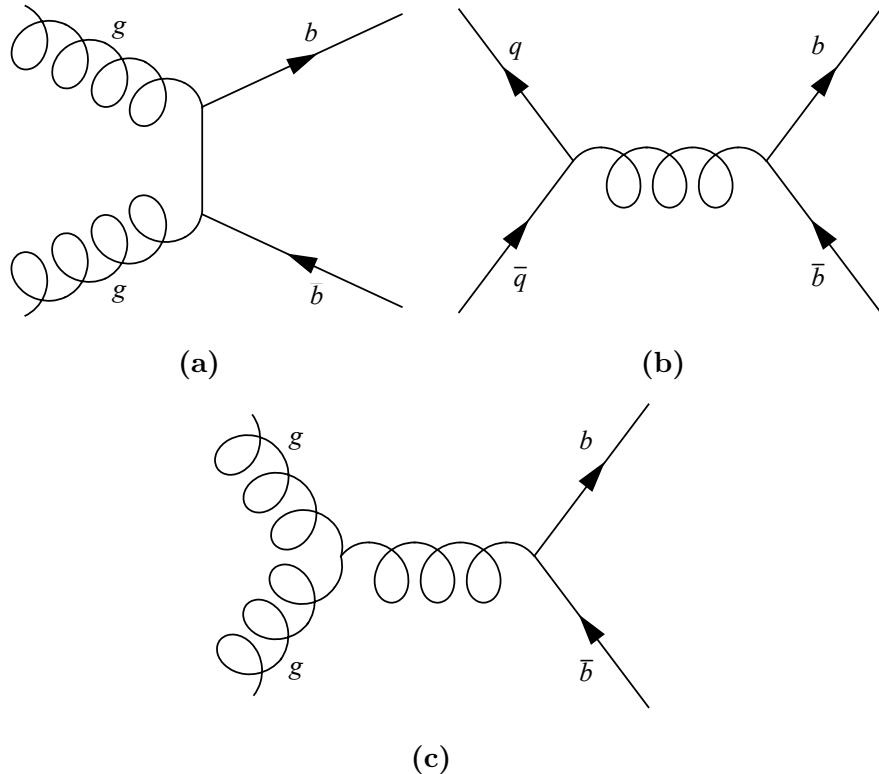


Fig. 3.2 Dominant production mechanisms for $b\bar{b}$ pairs created in pp collisions are: gluon-gluon fusion (a); quark anti-quark annihilation (b); and gluon-gluon splitting (c).

decaying into leptons, light hadrons or a mixture of both. The vertex where the b -hadron decays is known as the secondary vertex and is displaced by a few centimetres from the primary vertex. The LHCb experiment was designed considering both the cost of the detector and the physics processes that could be studied with it.

The LHCb experiment is a single arm forward spectrometer, with an angular coverage of 10 to 300 mrad in the vertical direction and 10 to 250 mrad in the horizontal direction relative to the beam pipe. A cross-section of the LHCb detector is shown in Figure 3.4, where a right-handed coordinate system is used. The z and y directions are shown in the figure and the x -axis points into the page. Protons collide at the interaction point on the left-hand side of the diagram, a subset of particles produced during the collisions travel through the detector leaving information in the sub-detectors. The coordinate system is chosen so that the interaction point is approximately at $x = y = z = 0$. The information deposited in the sub-detectors is reconstructed to determine what happened during the pp collisions. Although the angular coverage of the LHCb experiment is only 4% of the total solid angle, it covers the smaller angles relative to the beam pipe at which $b\bar{b}$ pairs are produced, as illustrated by the red shaded area in Figure 3.3.

The different sub-detectors have been chosen to exploit the characteristics of b -hadron decays and fall into two distinct categories; tracking detectors and particle

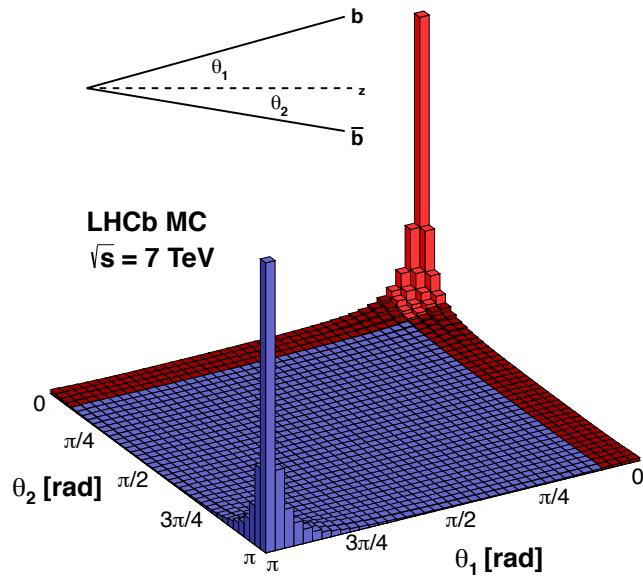


Fig. 3.3 Simulated angular distribution for $b\bar{b}$ production at the LHC at center-of-mass energies of 7 TeV, angles are relative to the beam pipe with $\theta = 0$ in the forward direction and $\theta = \pi$ in the backward direction. The red area shows $b\bar{b}$ pairs that are produced within the angular coverage of the LHCb detector and the purple shaded area shows $b\bar{b}$ pairs outside the detector coverage. Source: LHCb.

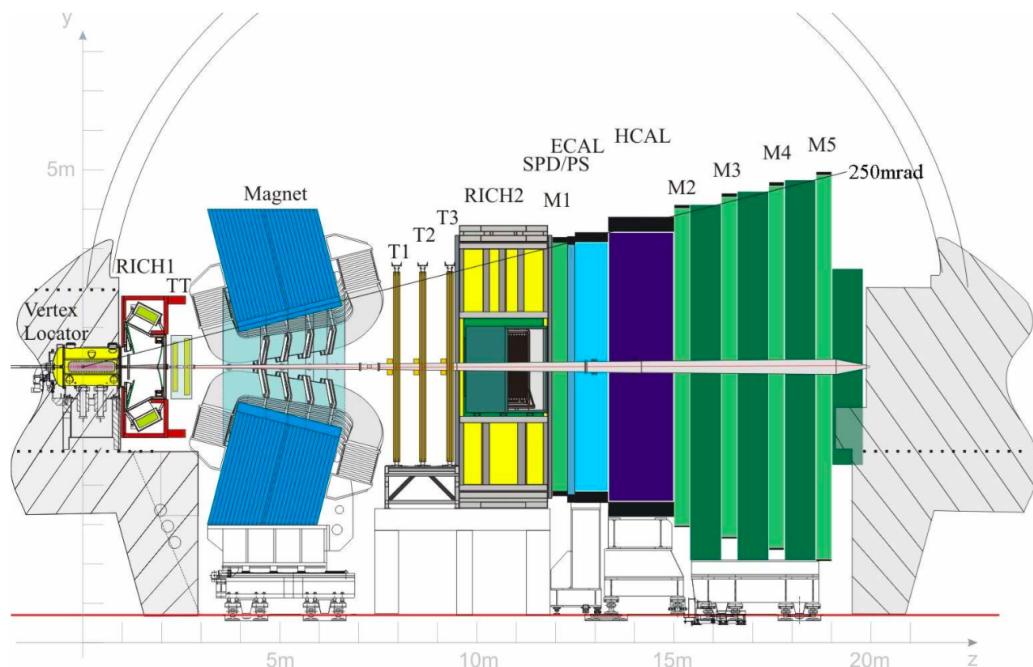


Fig. 3.4 Schematic of the LHCb detector [92]. A right-handed coordinate system is chosen where the x and z directions are shown and the x direct points into the page.

identification detectors. The sub-detectors and their performances are described briefly in the following sections along with the trigger system and software needed to analyse the data collected by the experiment. For a more detailed description of the detector and its performance during Run 1, see references [96, 92, 97, 98] and those given in the text.

3.2.1 Tracking

The tracking system within the LHCb experiment consists of the vertex locator, a dipole magnet and three tracking stations after the magnet. The vertex locator and tracking stations are described in Sections 3.2.1.1 and 3.2.1.2 and the magnet is described in Section 3.2.1.3. Together, the sub-detectors provide precise information on the passage of charged particles through the detector, a measurement of the particle momentum and electric charge. The tracking detectors work on the principle that the passage of high energy charged particles through silicon or ionised gas causes the excitation or ionisation of atoms in the material. The release of this energy is recorded and translated into an electrical signal that reveals the path of a particle. Hits left in the different detectors are combined to form tracks that are used to identify particle decays as described in Section 3.2.1.4.

3.2.1.1 The vertex locator

The VErtex LOcator (VELO) [99, 100] is a silicon detector surrounding the interaction point. Its main purpose is to provide precise information about the pp interaction vertices and secondary decay vertices of the particles produced. Information from the VELO enables precise measurements of particle lifetimes and the impact parameters of particle tracks necessary for the study of particle decays.

The VELO is made of two identical halves, each half consists of 21 stations containing two silicon sensors arranged along the beam pipe. The two halves of the VELO slot together and there is a small gap in the centre for the beams to pass through. The arrangement of sensors along the z -axis, shown in Figure 3.5, is designed so that the sensors cover the full LHCb acceptance and a charged particle within the detector acceptance will pass through at least three stations.

Each station is composed of two sensors: the R-sensor that measures the radial distance of charged particles from the beam axis; and the ϕ -sensor that measures the azimuthal angle of the particle with respect to the z -axis of the detector as shown in Figure 3.4. The information from the sensors is combined with the distance of sensors along the z -axis is used to reconstruct charged particle trajectories. The cylindrical geometry used for the VELO sensors, shown in Figure 3.6, is chosen to allow fast reconstruction of particle trajectories in the VELO.

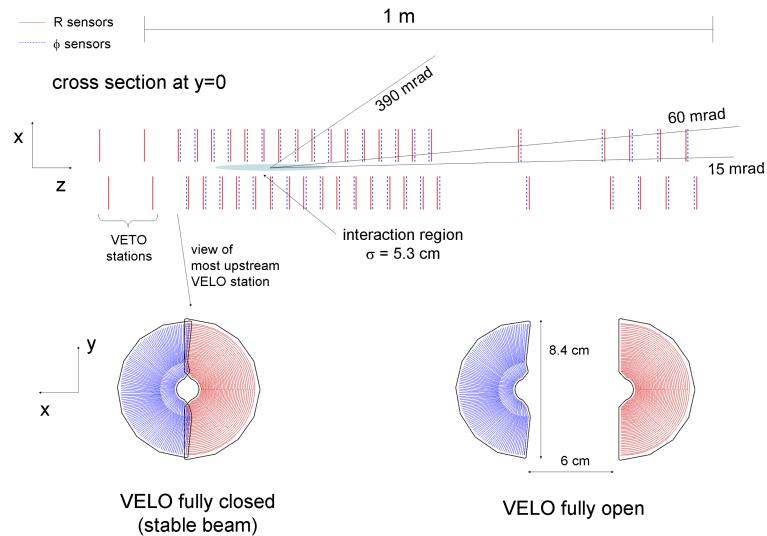


Fig. 3.5 The VELO layout and position of sensors along the beam axis [92].

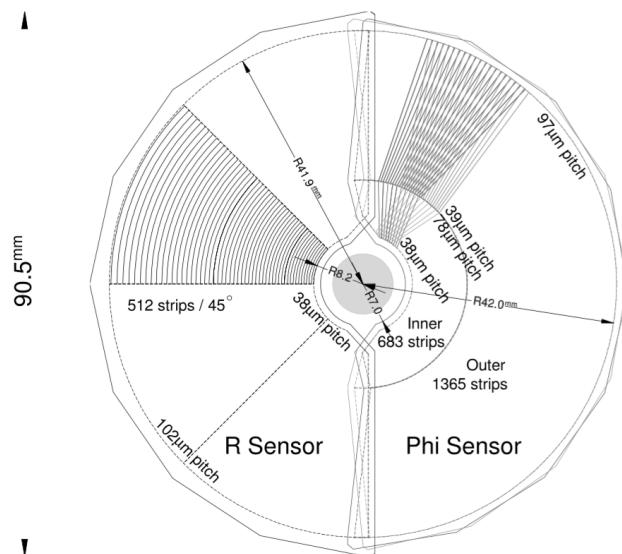


Fig. 3.6 Diagram of r and ϕ sensor layouts [92].

The momentum resolution achievable for charged tracks by the LHCb experiment is limited by the multiple scattering of particles as they travel through material the detector is made from. Therefore, to ensure good momentum resolution throughout the detector, the VELO is kept in a vacuum to reduce the material each particle must travel through. Each half of the VELO is enclosed inside an aluminium box, which keeps it in a vacuum and shields the electronic readouts from radio frequencies generated by the beam.

Excellent resolution of the position of each vertex is required in the VELO to identify the particle decays that occur. To achieve this the sensors need to be as close as possible to the interaction point. This is achieved by making the VELO out of two retractable halves and including the pp interaction point within the coverage of the VELO. During data taking, when the VELO is recording particle tracks, the inner most part of the sensors are 8 mm from the beam axis. However, during the injection phase the width of the beam is much larger than 8 mm, therefore the halves of the VELO are retracted to 3 cm from the nominal beam axis. This keeps the VELO safe from unnecessary radiation damage. The two halves of the VELO are displaced by 150 mm in the z -direction, as shown in Figure 3.5, so that when the VELO is closed, the sensors in each half overlap to help with detector alignment and reduced edge effects.

An additional purpose of the VELO is to identify high pile-up events. There are two VELO sensors upstream of the interaction point that provide information to the trigger about how many pp interactions there were in a bunch crossing. This information can be used to identify events with high numbers of primary vertices.

The VELO achieves a resolution on the position of each vertex of 10 - 20 μm transverse to the z direction and 50 - 100 μm along the z direction, the resolution of each track depends on the number of tracks in each event as shown in Figure 3.7. The VELO also gives measurements on the impact parameters of particles tracks; the impact parameter (IP) is the distance of closest approach between a particle track and the primary vertex. Figure 3.8 shows the IP resolution for 2012 data; a track with transverse momentum of 1 GeV/c has an impact parameter resolution of 35 μm .

3.2.1.2 Tracking stations

The LHCb experiment has 4 tracking stations in addition to the VELO. The Tracker Turicensis (TT) is located upstream of the magnet and the T stations, T1-T3, located downstream of the magnet [101, 102]. These tracking stations provide complementary information to the VELO, and the presence of the magnetic field allows the momentum and the sign of electric charge of particles to be determined.

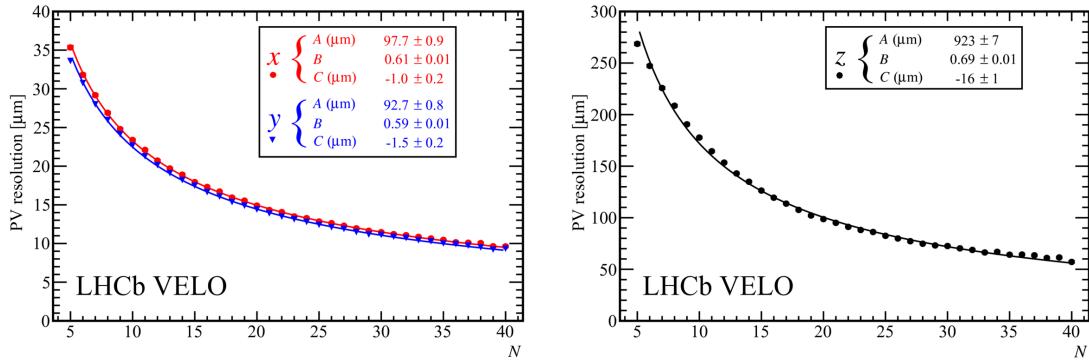


Fig. 3.7 Resolution of the position of vertices in an event achieved by the VELO in 2012 data. The resolution is shown for primary vertices perpendicular (left) and parallel (right) to the beam axis as a function of the number of tracks in an event [100].

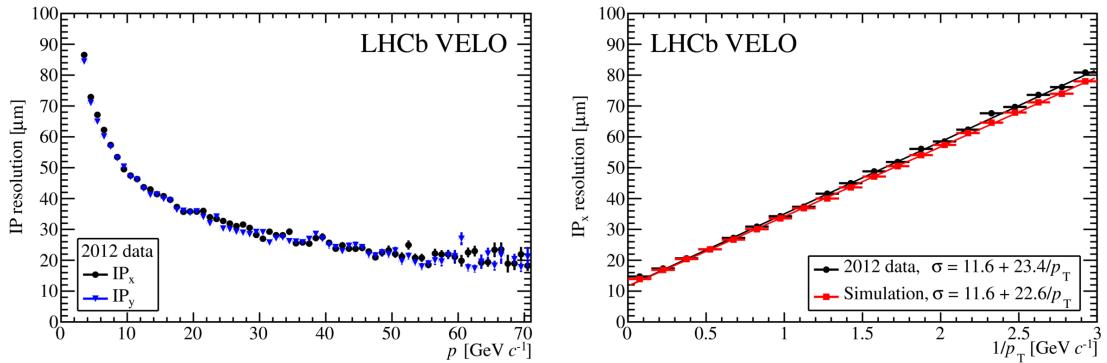


Fig. 3.8 Impact parameter resolution as a function of momentum (left) and inverse transverse momentum (right) achieved by the VELO for 2012 data [100].

The TT is made up of 4 layers of silicon trackers that cover the full LHCb angular acceptance. The TT is located just within the influence of the magnetic field of the dipole magnet, which provides the detector with two main purposes. First, the TT tracks the passage of charged particles with high momentum to enable good momentum resolution of tracks when the information is combined with that from other tracking stations. The TT has a resolution of 50 μm for a single hit. This resolution was chosen so that multiple scattering in the detector material rather than detector resolution is the limiting factor for the momentum resolution. The second purpose of the TT is to record tracks of low momentum particles that are then swept out of the detector acceptance as they continue through the magnetic field. These tracks will have a lower momentum resolution and help with pattern recognition within the RICH detectors.

The T stations, T1-T3, use two different technologies. The central part of each station is covered by the Inner Tracker (IT), made of silicon, while the external part is covered by the Outer Tracker and composed of straw drift tubes. There is a large increase in the size of the tracking stations between the TT and the T3 to ensure

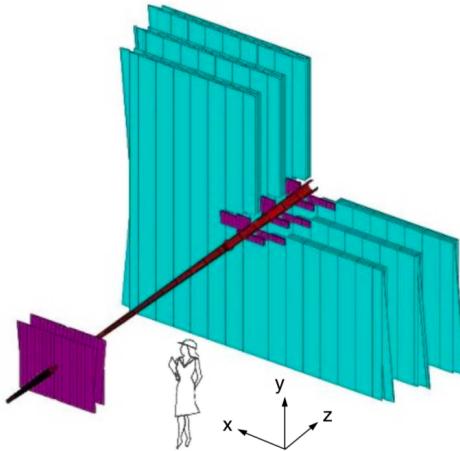


Fig. 3.9 Illustration of the relative sizes of the tracking stations. The TT is in magenta on the left hand side of the diagram, the IT is in purple surrounded in turquoise by the OT and the beam pipe is shown in red [92].

all stations cover the full angular acceptance of the detector. The size of the TT is $150 \text{ cm} \times 130 \text{ cm}$ and the T3 station is $600 \text{ cm} \times 490 \text{ cm}$. The relative sizes of the tracking stations are illustrated in Figure 3.9 as well as the IT and OT sections of the T1-T3 stations. The IT has very similar in design to the TT, each station is made of 4 layers of silicon trackers with an overall track resolution of $50 \mu\text{m}$. The silicon trackers are arranged in a cross shape around the beam pipe, as shown in Figure 3.9. Although the IT covers less than 2% of the T stations, 20% of tracks pass through it. This allows the occupancy of the OT to be less than 10% enabling a good overall track resolution from the OT despite it not being made of silicon. The OT of each tracking station is made of 2 staggered layers of straw tubes, covering the remaining area required for full coverage of the LHCb angular acceptance, including tracks bent by the magnetic field. The straw tubes have a fast drift time of 50 ns giving a better than $200 \mu\text{m}$ track resolution.

3.2.1.3 Dipole magnet

A warm dipole magnet [103] is used to measure the momentum and electric charge of particles travelling through the LHCb detector. The magnet was designed to have an integrated field strength of 4 Tm for track that travels 10 m through the detector.

The magnet is located between the TT and the T stations and magnetic field influences all charged particles in the LHCb angular coverage. The field is in the vertical direction therefore bending tracks in the horizontal direction. The magnet was designed so that the field strength in the RICH detectors is negligible (less than 2 mT) and to have the largest strength possible between the TT and T stations. Figure 3.10 shows a plot of the magnet strength along the length of the detector.

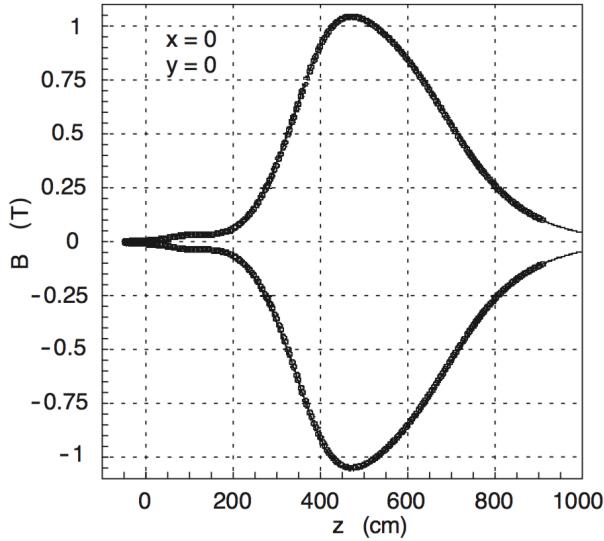


Fig. 3.10 Magnet field of the dipole magnet along the length of the LHCb detector. The peak strength of the field occurs between the TT and T1-T3 station.

The polarity of the magnetic field is periodically switched so that charged tracks are bent in the opposite direction. The polarity is switched to enable approximately equal amounts of data to be recorded for each magnet polarity during every year of data taking. This is done to measure detection asymmetries in each half of the detector and to help understand systematic uncertainties of \mathcal{CP} violation measurements.

3.2.1.4 Track reconstruction and performance

The information left by the passage of charged particles through the VELO, TT and T stations is combined using track reconstruction algorithms to find trajectories of charged particles through the length of the LHCb detector and infer the particle momentum. The tracking algorithms start with either segments of tracks in the VELO or the T stations and extrapolate from these segments into the other tracking detectors using specific search windows. Once all the segments of a track have been found, the trajectory is fitted with a Kalman Filter [104, 105] which takes into account multiple scattering and energy loss within the detector. For each track the Filter returns the χ^2 per degree of freedom, which is a measure of quality for the track. This parameter is used to ensure that only good quality tracks are used in physics analyses. The reconstructed tracks are classified into five types depending on which detectors they travelled through, as shown in Figure 3.11.

The different track classifications are:

- VELO tracks are formed by particles produced at large angles to the beam axis or travelling in the negative z direction from the interaction point, therefore

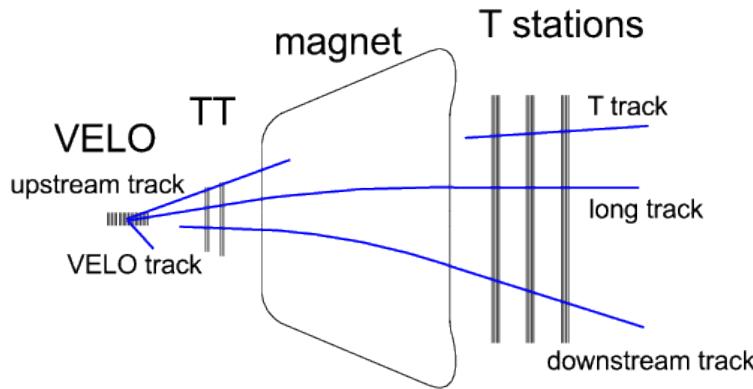


Fig. 3.11 Different types of tracks that are reconstructed at the LHCb experiment [106].

these particles only leave tracks in the VELO. VELO tracks are useful for reconstructing primary vertices;

- Upstream tracks are made by low momentum particles that leave hits in the VELO and TT stations. The absence of hits further down the detector is because the magnetic field sweeps the particles out of the detector acceptance. Upstream tracks have poor momentum resolution but are useful for understanding backgrounds and pattern recognition in the RICH-1 detector located between the VELO and the TT;
- Downstream tracks are produced by the decays of long-lived neutral particles, that travel out of the VELO before decaying. These particles only leave tracks in the TT and T stations;
- T tracks are tracks that cross only the T1-T3 stations and are formed from particles created in interactions with the detector material. Similar to upstream tracks, T tracks can help to understand backgrounds and pattern recognition in the RICH-2 detector located just before the T stations; and
- Long tracks are the most useful for understanding particles decays recorded by the detector because they are formed by particles that travel through the VELO, TT and T1-T3 stations. Information from all the tracking stations is combined in these tracks therefore they have the best momentum resolution.

The efficiency to correctly reconstruct tracks varies with the particle momentum and the number of tracks present in an event, as shown in Figure 3.12 for 2012 data. In Run 1 long tracks were correctly reconstructed an average of 96% of the time.

Inevitably not all tracks that are reconstructed are correct, there are two main types of incorrectly reconstructed tracks. The first are clone tracks that occur when two tracks have many hits in common. When this happens the track with the highest

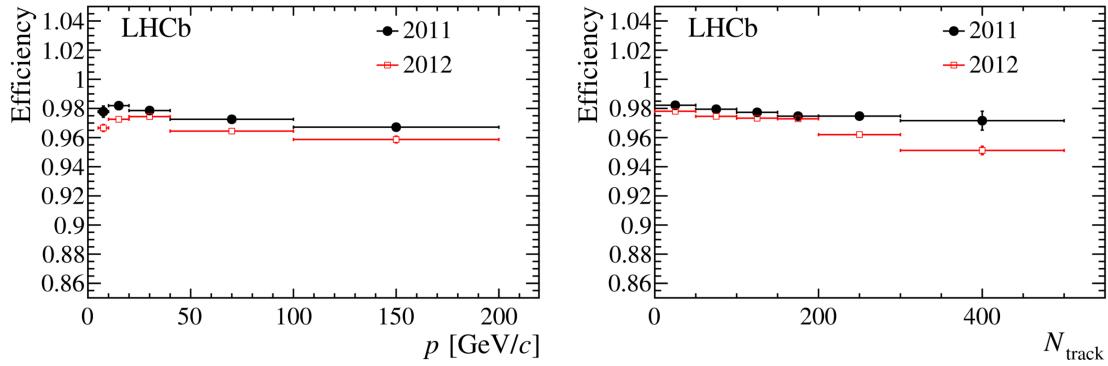


Fig. 3.12 Long track reconstruction efficiency as a function of momentum (left) and number of tracks in the event (right) for 2012 data [106].

number of total hits is kept and the other is discarded. The second type are ghost tracks that are formed when track segments in different detectors are incorrectly joined together. This most often occurs with segments in the VELO and T1-T3 stations, the number of ghost tracks in an event depends on the event multiplicity. These tracks are removed by cutting on the output of a neural network that returns a probability of how likely a track is to be a ghost.

The combined tracking systems achieve a momentum resolution of $\delta p/p = 0.5\%$ for particles with $p = 20 \text{ GeV}/c$ and a resolution of $\delta p/p = 0.8\%$ for particles with $p = 100 \text{ GeV}/c$. This momentum resolution, when combined with vertex information from the VELO, gives a decay time resolution of around 50 fs [98].

3.2.2 Particle identification

The particle identification (PID) detectors consist of two Ring Imaging Cherenkov (RICH) detectors, electromagnetic and hadronic calorimeters, and muon stations. Together these detectors distinguish between different charged leptons and hadrons and between neutral particles such as photons and neutral pions. Good particle identification is necessary to determine which b -hadron decayed and to distinguish between topologically similar decays, such as $B^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+K^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. The separate particle identification detectors are described in Sections 3.2.2.1, 3.2.2.2 and 3.2.2.3 and the methods used to combine information from each detector to identify different particle types is described in Section 3.2.2.4.

3.2.2.1 Ring Imaging Cherenkov detectors

The LHCb experiment uses two RICH detectors [107, 108] that are shown in Figure 3.13. The RICH-1 detector is located between the VELO and the TT station and RICH-2 detector is located between the last tracking station and the first muon station. Together the detectors provide particle identification information for particles

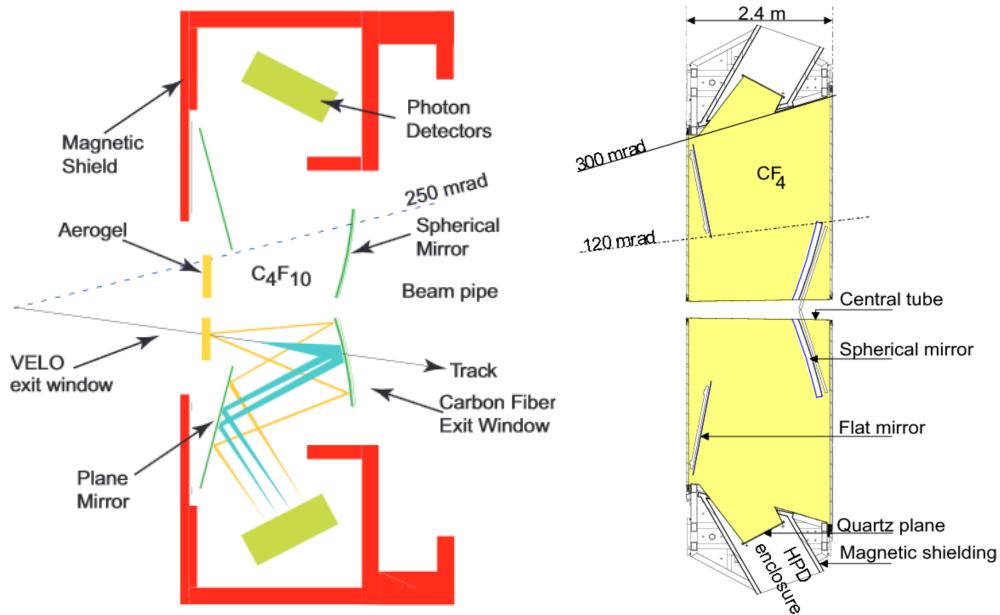


Fig. 3.13 RICH-1 detector (left) and the RICH-2 detector (right) [92]. For Run 2 the aerogel radiator in the RICH-1 detector was removed.

that fall in the momentum range of 2 to 100 GeV/c . The RICH detectors are vital to distinguish between pions, kaons and protons frequently produced in b -hadron decays. The energy range of the RICH detectors was chosen to cover the momentum range of decay products coming from multi-body and 2-body b -hadron decays.

These detectors are based on the following principle; when a charged particle travels through a dielectric medium the atoms excited by its passage are polarised. If the particle is travelling faster than the speed of light in the medium, the excitation energy is released as a coherent wavefront. The angle the wavefront travels at relative to the particle trajectory, θ_c , depends particle velocity, v , and the refractive index of the medium, n , as $\cos(\theta_c) = c/nv$. The light is produced in a ring and is called Cherenkov radiation. The angle at which Cherenkov radiation is produced gives a measurement of a particle's speed, which when combined with the particle's momentum, as measured by the tracking detectors, the particle mass and consequently its identity can be determined. However, many particles travel through the RICH detectors and create overlapping rings of light making particle identification complex. As charged particles travel through RICH detectors, the rings of light produced are focused by spherical and planar mirrors onto Hybrid Photon Detectors (HPDs) [109], as shown in Figure 3.13. The radii of the detected rings provides information about how fast the particle was travelling. Particle trajectories through the RICH detectors are inferred from information in the tracking stations and the expected pattern of Cherenkov radiation is calculated for each possible particle type. The expected patterns of light are compared to the observed pattern to find the likelihood for each

particle type, all possible particle types are compared to maximise the likelihood. An in-depth description of the reconstruction algorithm used in the RICH detectors can be found in [110].

The two RICH detectors provide identification information for different particle momentum ranges. The RICH-1 is sensitive to particles in the momentum range of 2 to 40 GeV/c and the RICH-2 is sensitive to particles in the range 15 - 100 GeV/c, to achieve this the detectors are made of different materials. The RICH-1 detector is composed of an array of 16 aerogel tiles sensitive to particles with a momentum between 2 and 10 GeV/c, behind the aerogel is a C₄F₁₀ gas radiator sensitive to particles in the momentum range 10 to 40 GeV/c. The aerogel radiator was removed after Run 1, therefore the RICH-1 is only sensitive to particles in the momentum range 10 to 40 GeV/c in Run 2. Unlike the RICH-1, the RICH-2 detector is composed of only one radiator, a CF₄ gas radiator.

As well as covering different momentum ranges the RICH detectors also have different angular acceptances. The RICH-1 covers the full LHCb angular acceptance whereas the RICH-2 covers ± 120 mrad in the horizontal and ± 100 mrad in the vertical direction. Although the angular acceptance for the RICH-2 detector is smaller, the area it covers contains the higher momentum particles it is sensitive to because the low momentum particles have been bent out of the acceptance by the magnetic field.

The performance of HPDs used in the RICH detectors is sensitive to the magnetic field of the dipole magnet. Therefore the HPDs are shielded from the using iron sheets ensuring the field is less than 2mT across them.

The expected Cherenkov angles for particle types with different momenta is shown in Figure 3.14 for the radiator materials in the RICH detectors. The reconstructed Cherenkov angle as a function of momentum is shown in Figure 3.15 for particles transversing the RICH-1 detector in 2011 data, the data shows distinct bands for each particle type as expected for the detector.

3.2.2.2 Calorimeters

The calorimeter system [111] consists of four detectors: a Scintillating Pad Detector (SPD); a Pre-Shower (PS); an electromagnetic calorimeter (ECAL); and a hadronic calorimeter (HCAL). Information from the calorimeters is used to identify electrons, photons and hadrons with high transverse momentum to be used in the first level of the trigger and to help with the reconstruction and identification of these particles. The ECAL is particularly important in the reconstruction of photons and neutral pions because it is the only part of the detector that measures the energy and momentum of these particles.

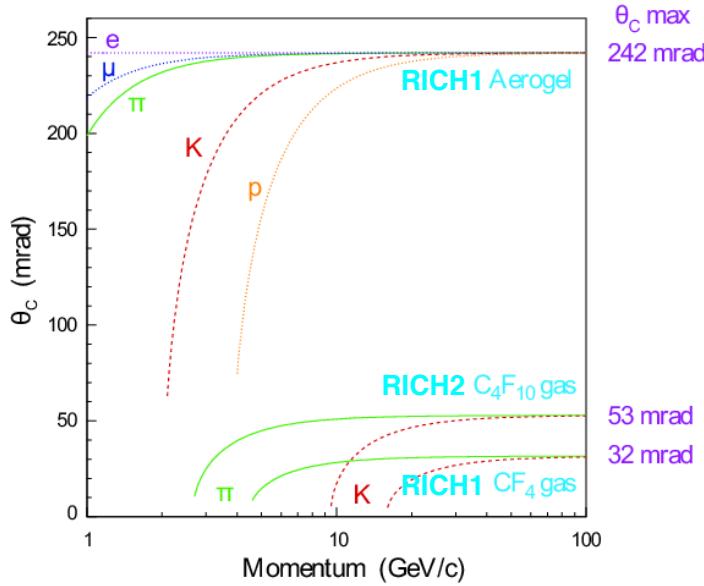


Fig. 3.14 Expected Cherenkov angles produced by different particles travelling through the radiators in the RICH detectors [92].

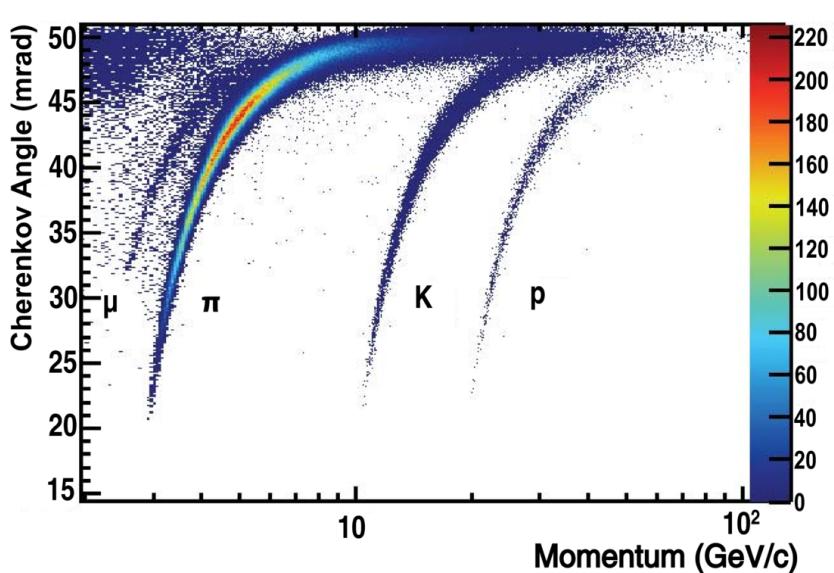


Fig. 3.15 Cherenkov angles for isolated tracks as a function of momentum in the RICH-1 detector for 2011 data [108].

The calorimeters in LHCb are sampling calorimeters consisting of layers of lead absorbers and scintillating material. In lead, incident particles create showers of secondary particles, the charged particles produced in the absorbers create light as they pass through the scintillators. The light travels through wavelength shifters where it is collected by photo-multiplier tubes and turned into an electrical signal. Electromagnetic showers caused by photons, electrons or positrons are started by ionisation, bremsstrahlung radiation or pair production depending on the energy and the type of the incident particle. Hadronic showers are caused by the interaction of hadrons with the detector via the strong force which produces showers of secondary particles. The showers produced in the calorimeters are along the direction of flight of the incident particle.

The SPD, PS and ECAL identify electrons, positrons and photons. The SPD is a layer of scintillating material at the start of the calorimeter system. It separates electron and photon showers created later in the calorimeter because only charged particles will produce light in the SPD. Next in the calorimeter system is the PS, and consists of a lead absorber followed by another scintillator similar to the SPD. The length of the lead absorber is chosen so that electrons will start showers in the absorber but charged pions will not. There is only a 1% chance of a pion creating a shower in the PS. Information collected by the PS enables showers created by pions in the ECAL to be separated from those created by electrons and positrons. The ECAL is designed to contain the entire shower of high energy photons so that it can provide a good energy resolution for photons passing through the detector. The ECAL has an energy resolution of $\delta E/E = 9\%/\sqrt{E} \oplus 0.8\%$ provided information from the PS and SPD are used.

The HCAL is predominately designed for use in the trigger and there is no requirement that the HCAL contains the full hadronic showers, therefore it was designed with a lower energy resolution of $\delta E/E = 69\%/\sqrt{E} \oplus 9\%$.

3.2.2.3 Muon stations

Muons are produced in many b -hadron decays; good muon identification is necessary to trigger events containing muons and to distinguish topologically similar decays such as $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$ in the analysis of particle decays. Compared with other particles, muons have a high penetrating power due to their relatively large mass and because they do not interact via the strong force, these properties are exploited in the muon detectors. There are five muon stations [112] located at the far end of the detector from the interaction point.

The layout of the muons stations is shown in Figure 3.16. The first muon station, M1, is located before the calorimeters. The inner section, where the fluence is greatest, is made of Gas-Electron-Multiplier detectors [113] and the outer section

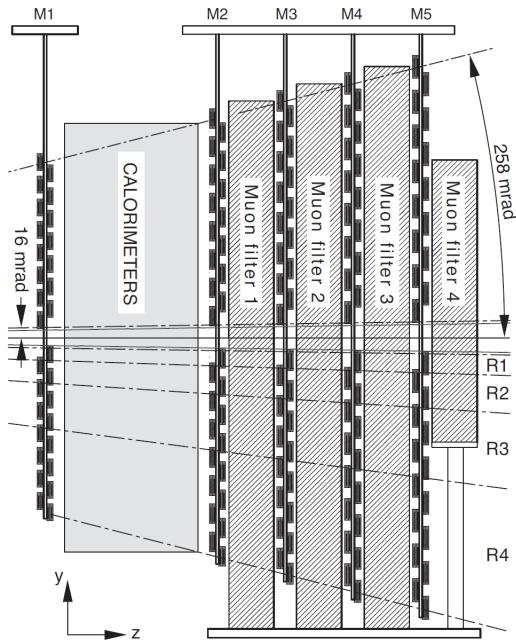


Fig. 3.16 Layout of the muon stations [92].

is made from Multi-Wire Proportional Chambers (MWPCs) [114]. The remaining stations, M2-5, are located after the HCAL, by which point most other particles have been absorbed by the calorimeters. These stations are made from MWPCs interleaved with 80 cm of lead absorber to filter out any remaining hadrons.

A muon must have a momentum of at least 3 GeV/c to pass through the calorimeters and the M2 and M3 stations. To travel through all the muons stations a muon must have a momentum of 6 GeV/c. Momentum information collected in the muon stations can be used in the trigger because the stations lie outside the magnetic field, which allows for fast reconstruction of the tracks and a muons. M1 is located before the calorimeters to improve the transverse momentum measurement of the muons that is needed in the trigger.

After the muon stations there is an iron wall to stop any particles from travelling downstream of the detector. The active area of the muon stations increases with distance from the interaction point to ensure the full angular acceptance of the detector is covered.

3.2.2.4 Particle identification and performance

The information collected in the PID detectors is combined to provide several discriminating variables that can be used to identify muons, protons, kaons, pions and electrons. There are three types of variables: the isMuon criteria, DLL variables; and ProbNN variables, which are described in the following. The performance of

these variables at distinguishing between different particles types is described in detail in references [98, 115]

The muon stations are used, along with information from the tracking system, to produce a binary selection to identify muons called the `isMuon` criteria. The tracking system is used to extrapolate a field of interest within the muon stations, a muon is identified if hits in the muon stations can be combined with those from the tracking system within the field of interest. The number of hits required in the muon stations depends on the momentum of the muon. Muons with momentum in the range $3 < p < 6 \text{ GeV}/c$ are required to leave hits in M2-M3, those in the momentum range $6 < p < 10 \text{ GeV}/c$ leave hits in M2-M3 and either M4 or M5, and finally muons with momentum above $10 \text{ GeV}/c$ must be observed in all the muon stations. Figure 3.17 shows the efficiency of the `isMuon` selection at identifying muons and probabilities of mis-identifying different hadrons as muons. The efficiencies and mis-identification probabilities are computed using the *tag and probe technique*. This technique uses two tracks from a decay. Particle identification requirements are applied to one track, the tag track, and the other track, the probe track, is used to evaluate the efficiency or mis-identification probability. The computation of the efficiency of the `isMuon` criteria to correctly identify muons uses $J/\psi \rightarrow \mu^+ \mu^-$ decays. Proton mis-identification probabilities are computed using $\Lambda^0 \rightarrow p \pi^-$ decays, and pion and kaon mis-identification probability are computed from $D^{*+} \rightarrow \pi^+ D^0 (\rightarrow K^- \pi^+)$ decays. The mis-identification rate is higher for lower momentum particles, which is expected given there are less hits in the muon detectors. The main contribution to misidentifying hadrons as muons comes from the kaons and pions that decay as they travel through the detector, the muons from these decays are then detected in the muon stations.

The information from all the PID detectors is combined using two different methods to provide global particle identification variables. One method is based on likelihood fits, producing the DLL variables, and the other is based on Neural Networks [116], producing ProbNN variables. In the first method, likelihood fits are performed in each sub-detector comparing charged particle tracks to different particle hypotheses. The resulting variable is the difference in the log-likelihoods between the track corresponding to a pion and a kaon, proton, muon or electron. The likelihood information from each sub detector is added linearly to form a combined likelihood. These variables are called $\text{DLL}_{X\pi}$ variables and give a measure of how likely a particle hypothesis of X is compared to that of a pion, where X can be a muon, kaon, proton or electron.

The second approaches uses Neural Networks to combine information from different sub-detectors and to provide a global probability of a track having a particular particle hypothesis. This method takes into account correlations between detector

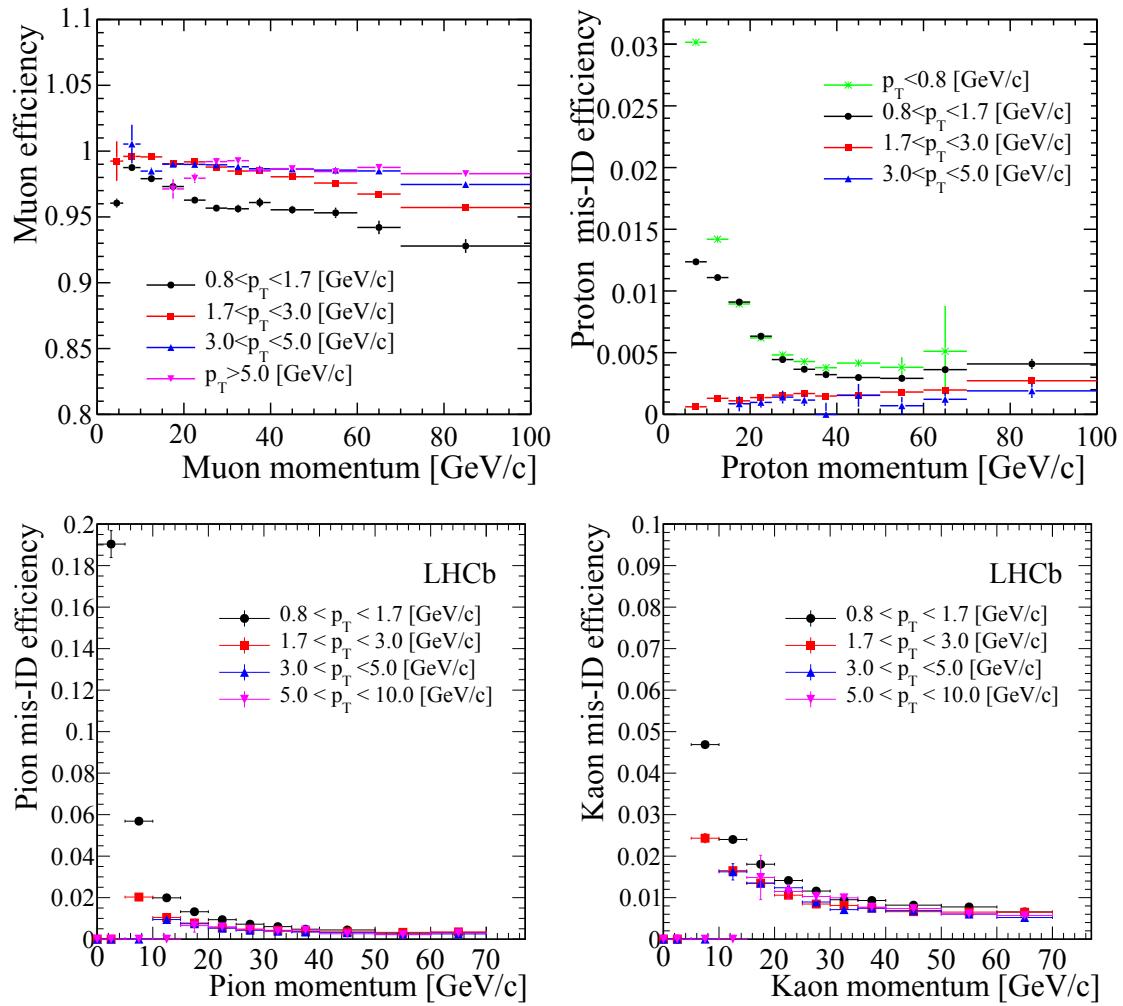


Fig. 3.17 Muon efficiency (top left) and mis-identification probabilities for protons (top right), pions (bottom left) and kaons (bottom right) for isMuon criteria [115].

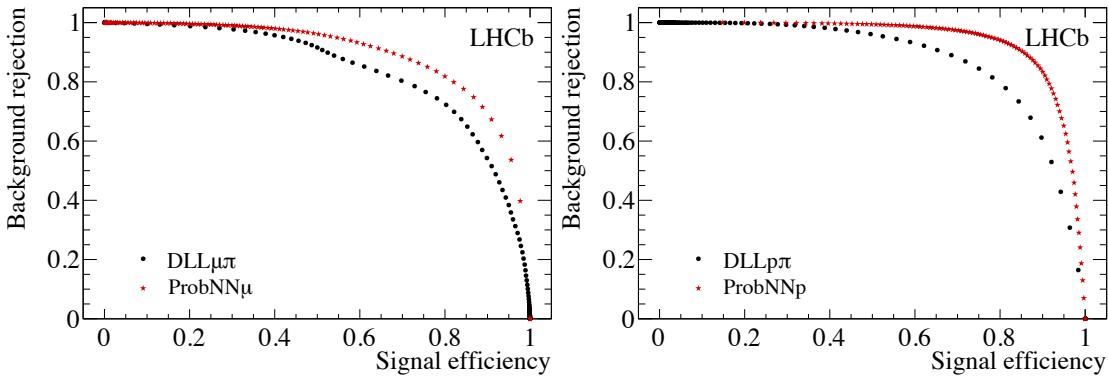


Fig. 3.18 Muon (left) and proton (right) signal efficiency vs background rejection for DLL and ProbNN PID variables [92].

systems and extra detector information that is not considered in the likelihood method. The Neural Networks are trained on simulated inclusive b -hadron decays and can be tuned to suit different situations, such as the data taking year. The variables produced by the Neural Networks are called ProbNN variables and ProbNN X corresponds to the probability of a track belonging to a particle hypothesis of X where X is a pion, kaon, proton, muon or electron.

Figure 3.18 shows a comparison of the performance of the DLL and ProbNN variables in selecting protons and muons. Although the performance to the two types of variables are quite different, the efficiencies of each variable varies with different kinematic properties of the decay. The most appropriate PID variable type to use depends on the physics analysis it is being used in.

3.2.3 Trigger

The LHC was designed to collide protons at a rate of 40 MHz, which is too high for information to be read out by the original design of the LHCb detector². However, most pp collisions do not produce particles within the detector acceptance that are interesting for the physics processes studied at LHCb. A trigger system [92, 117, 118] is used to identify pp collisions that contain potentially interesting events, the information from these events is saved and later analysed. The trigger has been designed to select interesting physics events with a high efficiency whilst reducing the event rate to one where information from the full detector can be read out. There are two levels to the LHCb trigger; the hardware trigger and the software trigger. The hardware trigger is known as the Level-zero (L0) trigger and reduces the 40 MHz collision rate to 1 MHz at which the full detector can be read out. The software trigger is known as the High-Level-Trigger (HLT) and has two stages that run on the

²After the upgrade to the LHCb detector during the second long shut down after 2018, the detector read out will be at 40 MHz

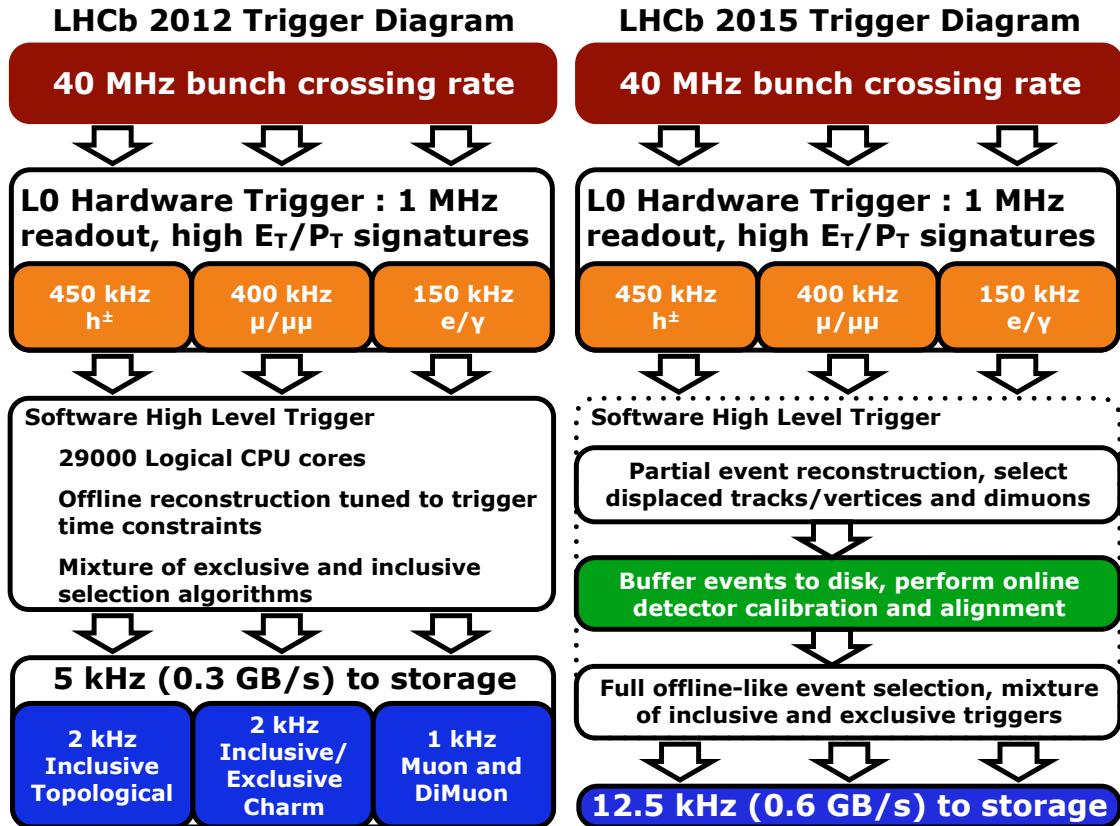


Fig. 3.19 Diagrams of the trigger systems used in 2012 (left) and 2015 (right). Source: LHCb.

output of the L0 further reducing the event rate by utilising information for all the detector sub-systems. Each level of the trigger is composed of trigger ‘lines’; these lines are made up of reconstruction and selection algorithms and either accept or reject each event. Only events that are accepted by a trigger line at both the L0 and HLT are available for use in physics analyses. During the long shut down between Run 1 and Run 2 of the LHC significant changes were made to the reconstruction of particle decays used to make decisions within the HLT. Diagrams of the trigger system used in 2012 and 2015 are shown in Figure 3.19 and are good illustrations of trigger systems used throughout Run 1 and Run 2, respectively.

3.2.3.1 L0 trigger

The L0 trigger runs synchronously to the LHC bunch crossing. Its purpose is to reduce the event rate to 1 MHz, at which point information from the full detector can be read out. The L0 is limited to use information from the detector that can be read at the same rate as the LHC collision rate. Therefore the L0 uses information from only the VELO, calorimeters and the muon stations to make decisions about the relevance of each event.

The heavy masses of b -hadrons means that their decays are characterised by the production of daughter particles with large transverse momentum (p_T) and transverse energy (E_T). The L0 trigger is designed to identify high p_T and E_T hadrons, electrons, photons and muons produced in events. Information from the PS, SPD, ECAL and HCAL is used to identify electrons, photons and hadrons in each event. Events are then accepted by the trigger lines if there is an electron, photon or hadron with E_T above a threshold value, where the E_T thresholds are different for each particle type. In a similar way to the calorimeters, the muon stations are used to identify muons with high p_T for trigger lines. There are two L0 trigger lines for muons that accept events based on muon p_T if either a single muon has a p_T above a threshold value or if the two muon combination $\sqrt{p_{T1} \times p_{T2}}$ is above a threshold value.

An additional requirement is placed on events before they can be accepted by any L0 trigger line, this requirement is based on the number of tracks present in an event. Events containing a large number of tracks, high multiplicity events, take a long time to reconstruct and process in the HLT, therefore it is not efficient to save these events. The multiplicity is measured by the number of hits in the SPD detector (nSPD) and only events with nSPD lower than a specified value can pass an L0 trigger line. The E_T and p_T thresholds and the event multiplicity limit for the L0 trigger lines vary for each year of data taking depending on the bandwidth available for the trigger.

Finally, there is one other type L0 trigger line that is not used to identify events containing interesting physics processes. The pileup veto stations in the VELO are used in L0 pileup trigger line, these lines identify the number of collisions in an event and are predominately used for luminosity measurements [119].

3.2.3.2 HLT trigger

Events that are accepted by the L0 trigger lines are moved to a farm of multiprocessor computing nodes, called the Event Filter Farm, where the HLT is run. The HLT is a software trigger that is split into two levels that are run successively.

The HLT1 is the first level of the HLT. It runs on the output of the L0 checking the decisions and reducing the event rate. The HLT1 trigger lines are composed of generic selection criteria, making decisions that confirm those made in the L0 about particular particle types and also identify generic types of particle decays such as inclusive b -hadron decays. The second level of the HLT, the HLT2, runs on the output of the HLT1 trigger and consists of trigger lines designed to select decays relevant to specific physics analyses or particle decay topologies.

During Run 1 time constraints in the HLT1 trigger to process the output of the L0 did not allow for full event reconstruction using all LHCb sub-detectors. Instead the HLT1 ran reconstruction and selection algorithms on information only from the

VELO and tracking stations. The reduced output of the HLT1 provided an event rate that was low enough to allow event reconstruction that includes all detector subsystems to be used in the HLT2. However, the reconstruction used in the HLT2 was different to the offline reconstruction that is applied to data before it is analysed. Significant changes were made in the reconstruction used in the HLT between Run 1 and Run 2, the details of the changes made can be found in [120]. The majority of the changes to the HLT for Run 2 are not relevant for the measurements presented in this dissertation, but the overall change is that the same reconstruction is used in the HLT and the offline reconstruction.

Just like the L0 trigger, trigger lines in the HLT vary for each year of data taking; both the selection criteria used in the lines and also new trigger lines are introduced. The number of HLT2 lines increases with each year of data taking as understanding of the capabilities of the experiment increases; there were about 100 HLT2 lines in 2011, 200 in 2012, and 450 in 2015.

3.2.4 Software and simulation

The data that is read out of the LHCb experiment needs further processing before it can be analysed to study the SM and search for NP effects. The GAUDI framework [121] is a C++ framework that is the basis for the software applications needed to process data recorded by the LHCb experiment [122]. This framework ensures the necessary software is available to all users and changes to the software are implemented across all applications, it is suited to the distributed computing system used in LHCb [123].

Once events have been accepted by the trigger, the first step in processing the output of the detector is reconstructing events, this is done by the BRUNEL [124] application. The reconstruction is applied to events that are accepted by the trigger for both Run 1 and Run 2 data, however the reconstruction used for Run 2 data is the same as that used in the trigger. The BRUNEL application takes the digitised detector read out, reconstructs hits in the tracking stations to find particle trajectories and momenta, and combines information from the RICH detectors, calorimeters and muon stations to compute PID variables. The output of processing by the BRUNEL application are stored in ‘Data Summary Type’ (DST) files.

Next the DAVINCI application [125] is used to fit the tracks reconstructed in BRUNEL with primary and secondary vertices. This application assigns particle hypotheses to each track and reconstructs the decay trees of particles in the detector, computing the kinematic properties of decays. The reconstructed output of the trigger is too large to be stored in one place and to be used by all analysts, therefore a ‘stripping’ procedure is used to break up the data into a manageable size for each the different analyses performed on the data. Each analyst designs a set of loose selection

requirements, called stripping lines, specific to their decays of interest. The selections are applied centrally to the reconstructed events and are designed to keep as much of the signal relevant to the analysis as possible but reduce the number background events. Only events that pass the selection criteria of a stripping line are available to be used in analyses. The output of this process are smaller DST files, events passing the stripping selections can either be saved with the full event information or with just the tracks related to the signal candidate. The choice depends on the physics process the stripping line is relevant for. The stripping selection is run a limited number of times and is applied separately to data collected in different years. Requirements are imposed on the amount of data each stripping line can retain, typically the output of a line must be less than 0.05% of the original data set size if the full event information is saved. Each analyst then uses the DA VINCI application one last time to produce ROOT [126] files from the output of their stripping lines, these files display the data and properties of particle decays in histograms and are used to analyse the data.

As well as data collected by the experiment, simulated data that mirrors what is expected in the experiment is needed to understand the detector performance and for the analysis of the data. There is a set of software applications that are dedicated to the production of Monte Carlo simulated events within the GAUDI framework. Events are generated using the GAUSS application [127, 128], which uses PYTHIA [129, 130] to model pp collisions and the production of particles, and then the EVTGEN application [131] is used to calculate the decay trees and kinematics of these particles. Final state radiation is modelled using PHOTOS [132]. Both PYTHIA and EVTGEN have been tuned for the production and decay of particles within the LHCb detector. The GEANT4 [133, 134] toolkit is used to model the interaction of particles as they travel through the LHCb sub-detectors and the hits made by particles in the detector. In the simulation the type of particles generated and how they decay can be specified so that the simulated events of a particular decay can be generated. The BOOLE application [135] then produces the digitised detector read out based on the information from GEANT4 that mimics the detector read out when data is recorded. The output of BOOLE encompasses the detector response to the different hits, the electronic read out and the L0 hardware trigger, as well as including additional hits from event spillover and LHC backgrounds. The digitised response of the detector is then processed by BRUNEL and DA VINCI in the same way as the real data to produce the ROOT files.

The LHCb software framework is set up so that it can be used on the Worldwide LHC Computing Grid [136, 137], the Grid is made up of computers across the world that each store part for the LHCb data set and simulation data. Despite the stripping process the data produced at LHCb is too large to be stored in one place. The

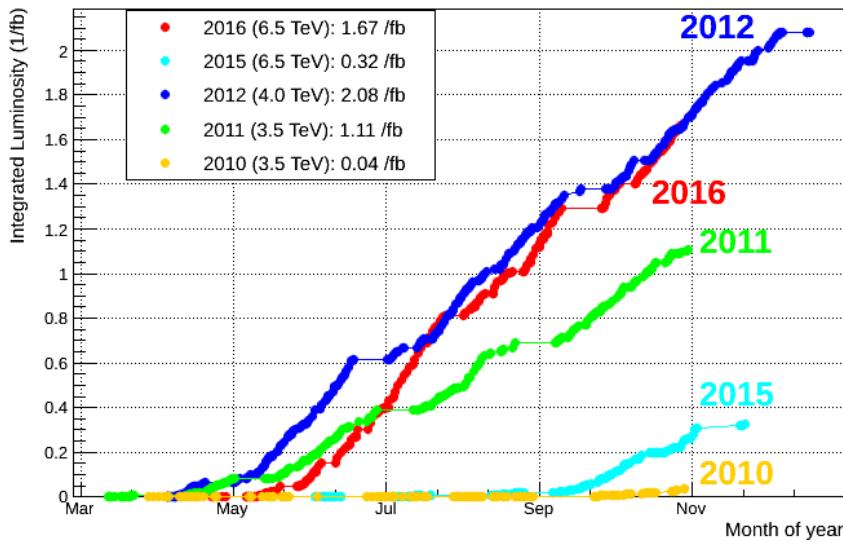


Fig. 3.20 Integrated luminosity collected by the LHCb experiment in each year of data taking. Source: LHCb.

DIRAC [138] system manages grid sites and the GANGA [139, 140] project allows the submission analysis code to different grid sites. The grid enables analysts to process and study the large amounts of data produced by LHCb without having to store the data where the analyst is.

3.3 Summary

The data collected so far by the LHCb experiment during pp collisions is summarised in Figure 3.20. The physics analyses described in this dissertation use an integrated luminosity of 4.4 fb^{-1} that consists of data recorded during 2011, 2012 and 2015 and up until September of 2016. The break down of the integrated luminosities for each year of data used in this dissertation are given in Table 3.2. The total integrated luminosity of Run 2 is currently less than the total from Run 1. However the production cross section for b -hadrons approximately doubled with the increase in centre-of-mass energy between Run 1 and Run 2, therefore Run 2 data will already contain a comparable number of b -hadron decays as Run 1 data.

Run	Year	\sqrt{s} / TeV	Integrated luminosity / fb^{-1}
Run 1	2011	7	1.11
	2012	8	2.08
Run 2	2015	13	0.32
	2016	13	1.10

Table 3.2 Integrated luminosity of data collected by the LHCb experiment during pp collisions used in the analyses documented in Chapters 5 and 6.

Chapter 4

Event selection

This chapter describes the criteria used to select and identify B -meson decays needed for two analyses: the measurement of the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions; and the measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime. In order to measure properties of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays from data collected by the LHCb experiment, $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays must be separated from backgrounds in the data. The main sources of backgrounds are described in Section 4.1. The development of the selection criteria and analysis strategies rely on information from simulated particle decays, these are documented in Section 4.2.

The selection criteria used to identify $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays and decays used as normalisation channels for the branching fraction measurement are described in Section 4.3. There are four main steps in the selection process. First, requirements are applied to the output of the trigger and then the data is refined by cut-based selection criteria. The last two steps in the selection process uses particle identification variables and multivariate classifiers to separation signal and background decays.

The selection of decays for the effective lifetime measurement differs from that used for the branching fraction measurement due to the different analysis strategies described in Chapters 5 and 6. The criteria used to identify decays needed for the branching fraction measurements are adapted for the effective lifetime measurement and the changes made to each step of the branching fraction selection process are documented in Section 4.4.

During the development of the selection criteria, $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates in data that have an invariant mass of the two muons within specified window around the B_s^0 or B^0 meson masses are not revealed. This is done to avoid introducing biases into the selection procedure based on statistical fluctuations in the data. An analysis performed in this way is known as a ‘blinded analysis’. The mass windows are defined as ± 80 MeV/ c^2 around the B_s^0 and B^0 masses of 5367 MeV/ c^2 and 5280 MeV/ c^2 , respectively [5].

4.1 Background sources

The reconstruction of the data collected by the LHCb experiment, described in Section 3.2.4, produces numerous $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates from pairs of muons in the detector. Some candidates will have come from real $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays but there are other processes that occur during pp collisions that leave a signature in the detector which can be reconstructed incorrectly as a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay. The selection aims to separate real $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays from these backgrounds to produce a set of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates with a high signal purity. The main background sources are:

- Elastic collisions of protons that produce a pair of muons via the exchange of a photon, $pp \rightarrow p\mu^+\mu^-p$. The protons travel down the beam pipe and are undetected leaving the muons to be reconstructed as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$. Typically the muons produced in this way have low transverse momentum;
- Inelastic proton collisions that create two muons at the primary vertex. The muons form a good vertex and can be combined to form a $B_{(s)}^0$ that decays instantaneously. This type of background is prompt combinatorial background;
- $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ decays where the photon is not reconstructed. The presence of the photon in the decay means that $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ decays are not helicity suppressed and could therefore be a sizable background. However, the photon gains a large transverse momentum resulting in the reconstructed $B_{(s)}^0$ mass being much lower than the expected B_s^0 mass. The branching fraction of $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ varies with the energy of the photon energy and is approximately an order of magnitude higher than the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction [63, 141, 142];
- Random combinations of muons produced by separate semi-leptonic decays. The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates formed in this way are called long-lived combinatorial background because the reconstructed $B_{(s)}^0$ will have a significantly longer lifetime than the $B_{(s)}^0$ candidate of prompt combinatorial background;
- Semi-leptonic decays where one of the decay products is mis-identified as a muon and/or is not detected. The resulting mass of the $B_{(s)}^0$ candidate is lower than expected due to the missing particle information. The semi-leptonic decays that contribute to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ backgrounds in this way are $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ where $J/\psi \rightarrow \mu^+ \mu^-$; and
- $B \rightarrow h^+ h'^-$ decays, where $h^{(\prime)} = K, \pi$, when both hadrons are mis-identified as muons. This usually occurs when the hadrons decay whilst travelling through

Decay	Branching fraction
$B_s^0 \rightarrow \mu^+ \mu^- \gamma$	$\sim 10^{-8}$
$B_s^0 \rightarrow K^+ K^-$	$(2.52 \pm 0.17) \times 10^{-5}$
$B_s^0 \rightarrow K^+ \pi^-$	$(5.6 \pm 0.6) \times 10^{-6}$
$B^0 \rightarrow K^+ \pi^-$	$(1.96 \pm 0.05) \times 10^{-5}$
$B^0 \rightarrow \pi^+ \pi^-$	$(5.12 \pm 0.19) \times 10^{-6}$
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$(1.45 \pm 0.05) \times 10^{-4}$
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$(1.42 \pm 0.35) \times 10^{-4}$
$\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$	$(4.1 \pm 1.0) \times 10^{-4}$
$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	$(1.83 \pm 0.25) \times 10^{-8}$
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	$(8.6 \pm 3.6) \times 10^{-9}$
$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	$(9.5 \pm 0.2) \times 10^{-6}$

Table 4.1 Branching fractions for background decays. The estimate of the $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ branching fraction comes from reference [141]. The measured values of the $B \rightarrow h^+ h^-$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ branching fractions are taken from references [5, 143, 144]. The theoretical prediction for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ branching fraction combines information from references [145, 146], the $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ branching fraction is estimated from references [147, 148] and the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ branching fraction is evaluated from references [144, 149].

the detector. Similar to mis-identified semi-leptonic decays, the reconstructed $B_{(s)}^0$ candidate mass is lower than expected.

The branching fractions of the backgrounds from mis-identified decays are shown in Table 4.1. The separation of $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays from the backgrounds is challenging because these decays are much less abundant than the backgrounds. Therefore reconstructed candidates are predominately made from background decays.

4.2 Simulated particle decays

Simulated particle decays, as described in Section 3.2.4, are used to develop the selection and analysis of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Large samples of simulated decays are needed to separate signal from background decays and to evaluate the efficiency of the selection criteria to identify different particle decays. The simulated decays used for studies performed for this dissertation are listed in Table 4.2 alongside the data taking conditions and simulation versions used to generated the decays.

There exist multiple versions of the simulation because it is updated as understanding of the detector improves and to incorporate differences in data taking

Decay	Generator level cuts	Data taking conditions	Simulation version	Events ($\times 10^6$)
<i>Cut-based selection studies</i>				
$B_s^0 \rightarrow \mu^+ \mu^-$		2012	sim06b	2.0
$B^0 \rightarrow \mu^+ \mu^-$		2012	sim06b	2.0
$B^0 \rightarrow K^+ \pi^-$		2012	sim06b	1.0
$B^+ \rightarrow J/\psi K^+$		2012	sim06b	1.0
<i>Multivariate classifier training</i>				
$b\bar{b} \rightarrow \mu^+ \mu^- X$	$p > 3 \text{ GeV}/c$ $4.7 < m_{\mu^+ \mu^-} < 6.0 \text{ GeV}/c^2$ DOCA < 0.4mm $1 < \text{PtProd} < 16 \text{ GeV}/c$	2012	sim06b	8.0
$b\bar{b} \rightarrow \mu^+ \mu^- X$	$p > 3 \text{ GeV}/c$ $4.7 < m_{\mu^+ \mu^-} < 6.0 \text{ GeV}/c^2$ DOCA < 0.4mm $\text{PtProd} > 16 \text{ GeV}/c$	2012	sim06b	6.6
$B_s^0 \rightarrow \mu^+ \mu^-$		2012	sim06b	2.0
<i>Analysis method development</i>				
$B_s^0 \rightarrow \mu^+ \mu^-$		2011	sim08a	0.6
		2012	sim08i	2.1
		2015	sim09a	2.1
		2016	sim09a	1.1
$B^0 \rightarrow K^+ \pi^-$		2011	sim08b	0.8
		2012	sim08g	8.6
		2015	sim09a	4.6
		2016	sim09a	8.1
$B_s^0 \rightarrow K^+ K^-$		2012	sim08g	7.1
		2015	sim09a	4.0

Table 4.2 Simulated samples used for developing the selection and analysis of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays listed according to the study the decays are used in. Cuts are applied to $b\bar{b} \rightarrow \mu^+ \mu^- X$ to the magnitude muon momenta (p), invariant mass of the two muons ($m_{\mu^+ \mu^-}$), the distance of closest approach of the tracks of the muons (DOCA) and the product of the transverse momenta of the muons (PtProd).

conditions, such as new trigger lines or changes in the centre-of-mass energy. Similar versions are chosen for decay samples used within each study listed in Table 4.2, so that differences between different decays are not masked by variations in the simulation.

Simulated $b\bar{b} \rightarrow \mu^+\mu^-X$ decays are used to understand the long-lived combinatorial background. However, producing a large enough sample of these decays to be useful is computational expensive and produces large output files. Therefore cuts are applied as the decays are generated to reduce the size of the samples and to speed up the simulation process. The cuts, listed in Table 4.2, are applied to the magnitude of the muon momenta, the reconstructed mass of the muon pair, the product of the transverse momenta of the muons and the distance of closest approach of the tracks of the two muons. In addition, these samples are ‘stripping filtered’ which means that only candidates that pass the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ stripping selection criteria, discussed in Section 4.2, are saved to further reduce the size of the output files. The cuts applied in the stripping selection are given in Table 4.4.

Overall simulated decays accurately model what occurs in data. However, the distributions of particle identification variables and properties of the underlying pp collision, such as the number of tracks in an event, are not well modelled in the simulation. The mis-modelling of particle identification variables is corrected for using the PIDCalib package [150] and simulated decays are re-weighted using information from data to accurately model the underlying event, as described in Section 6.3.1.

4.3 Event selection for the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction measurements

As well as identifying $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in data, the branching fraction measurements, described in Chapter 5, require $B^+ \rightarrow J/\psi K^+$ and $B \rightarrow h^+h'^-$ decays as normalisation modes to determine the branching fractions from the observed number of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in data. Furthermore $B_s^0 \rightarrow J/\psi\phi$ decays are used to verify steps of the measurement process.

This section describes the selection criteria used to identify $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi\phi$ decays in data. The trigger requirements used to identify these decays are given in Section 4.3.1. Section 4.3.2 describes a cut based selection, tailored for each decay mode, that is used to refine to candidates that pass the trigger requirements. Included in this section is an investigation into the selection efficiency of cuts used in this step of the selection process. Up until the cut-based selection the process for selection $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi\phi$ decays is similar but the selection of $B_{(s)}^0 \rightarrow \mu^+\mu^-$

decays diverges from the other decays with the requirements placed on particle identification variables described in Section 4.3.3. The last step in the selection process uses two multivariate classifiers that are described in Section 4.3.4 to separate signal and background decays. One classifier is applied to all decays needed for the branching fraction analysis whereas the other classifier is used only to separate $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays from backgrounds. Finally, the selection criteria used for the branching fraction measurements are summarised in Section 4.3.5.

4.3.1 Trigger requirements

The trigger, described in Section 3.2.3, selects events that could contain interesting particle decays. Candidates consistent with different particle decay hypothesis are reconstructed from events that were accepted by the trigger. For each candidate it is useful to know whether it was a component in that candidate that caused the event to be selected by a trigger line or if it was another particle in the event. Each trigger line produces different decisions that identify this. The possible trigger decisions are:

- TOS, triggered on signal - a candidate is identified as TOS if only information from the candidate was enough to cause a trigger line to save the event;
- TIS, triggered independent of signal - a candidate is identified as TIS if part of the event independent of the candidate was sufficient to cause a trigger line to save the event; and
- DEC - a candidate is identified as DEC if anything in the event caused a trigger line to save the event. This includes TIS and TOS decisions and also when a combination of information from the candidate and something else in the event caused a trigger line to save the event.

Since $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are very rare decays and the trigger requirements are chosen to keep a high efficiency for selection $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at this step of the selection. Individual trigger lines are not used for the selection, instead global trigger lines that combine the information from many separate lines are used. Furthermore candidates are required to be identified as DEC for each level of the trigger to ensure a high efficiency is achieved. The combined trigger lines used at each level of the trigger are the L0Global, Hlt1Phys and Hlt2Phys lines. The L0Global combines all trigger lines present in the L0 trigger. It selects an event provided at least one L0 trigger line selects it and rejects an event if no L0 trigger selects it. The Hlt1Phys and Hlt2Phys triggers are very similar to the L0Global trigger except that decisions are based only trigger lines related to physics processes and HLT trigger lines used for calibration are excluded.

Trigger Line	Trigger decision
$B_{(s)}^0 \rightarrow \mu^+\mu^-, B^+ \rightarrow J/\psi K^+, B_s^0 \rightarrow J/\psi \phi$	
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC
$B \rightarrow h^+h'^-$	
L0Global	TIS
Hlt1Phys	TIS
Hlt2B2HHDecision	TOS

Table 4.3 Trigger decisions used to select $B_s^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays.

The trigger requirements to identify $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are also used to select $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays and slightly different trigger requirements are used for $B \rightarrow h^+h'^-$ decays. $B \rightarrow h^+h'^-$ decays are required to be TIS by the L0Global and Hlt1Phys trigger lines and TOS by at the HLT2 level by specific trigger lines designed to select $B \rightarrow h^+h'^-$ decays. The TIS decision is used for $B \rightarrow h^+h'^-$ decays to reduce the difference in selection efficiencies between the dominant lines that trigger $B \rightarrow h^+h'^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. However, the efficiency of TIS decisions is quite low at the HLT2 level, therefore TOS decisions are used so that there is a large enough samples of decays.

In summary, the requirements imposed on the trigger to select $B_s^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays are shown in Table 4.3.

4.3.2 Cut-based selection

The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates that pass the required trigger decisions are refined by a cut-based selection. The selection criteria for $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays are kept as similar as possible to that used to identify $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in order to reduce systematic uncertainties from selection efficiencies in the normalisation procedure described in Section 5.4. The cut-based selection is composed of two parts; the stripping selection and the offline selection.

The stripping selection described in Section 3.2.4, is applied to all events that pass the trigger. It consists of individual stripping lines that select reconstructed candidates for specific decays. The stripping selection used to select $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays used the branching fraction measurements published in references [46, 47] are described in Sections 4.3.2.1.

These selections requirements were designed at the start of Run 1 by studying the efficiencies of different selection cuts from simulated events [151]. However since then improvements have been made to the simulation of particle decays at LHCb, therefore it is prudent to check the accuracy of the selection efficiencies with updated simulated events and also to investigate where improvements can be made to the efficiency of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ stripping selection. An investigation into the choice of cuts used in the stripping selection is described in Section 4.3.2.2.

The offline selection cuts are applied to the output of the stripping selection. Overall the stripping selection imposes loose selection requirements onto $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates so that as much information as possible is still available to develop the analysis and understand background events after the stripping selection. Therefore the offline selection further refines the data, removing background candidates. The offline selection cuts are presented in Section 4.3.2.3 and difference in selection criteria applied to Run 1 and Run 2 data are detailed.

4.3.2.1 Development of the stripping selection

There are four separate stripping lines that are designed to select $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ and $B \rightarrow h^+h'^-$ candidates, respectively. Although the selection of all decays is kept as similar as possible to the signal selection, the selection of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays diverges from the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ selection due to the additional particles in the final state of the decay.

The stripping selection cuts applied for the Run 1 branching fraction analysis [47, 46] to select $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ and $B \rightarrow h^+h'^-$ candidates are listed in Table 4.4 and 4.5.

The variables used in the stripping selection are:

- the reconstructed mass (M) - the mass and momenta of the decay products of the B meson (or J/ψ) are combined to provide its reconstructed mass. Cuts on the mass remove candidates with a reconstructed mass far from expected, which are consistent with background. Loose mass requirements are made in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ selection to allow for the study of semi-leptonic backgrounds in data that have a mass lower than the $B_{(s)}^0$ mass when mis-identified as a $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decay;
- the “direction cosine” (DIRA) - this is the cosine of the angle between the momentum vector of the particle and the vector connecting the production and decay vertices¹ of the particle. For correctly reconstructed candidates the

¹The production vertex of the B or the primary vertex is identified by extrapolating the B meson momentum vector towards the beam axis. The closest vertex to the intersection of the B momentum and the beam axis is assigned as the primary vertex.

Particle	$B_{(s)}^0 \rightarrow \mu^+\mu^-$	$B \rightarrow h^+h'^-$
$B_{(s)}^0$	$ M_{B_{(s)}^0} - M_{B_{(s)}^0}^{\text{PDG}} < 1200 \text{ MeV}/c^2$	$ M_B - M_B^{\text{PDG}} < 500 \text{ MeV}/c^2$
	DIRA > 0	DIRA > 0
	$\chi_{\text{FD}}^2 > 225$	$\chi_{\text{FD}}^2 > 225$
	$\chi_{\text{IP}}^2 < 25$	$\chi_{\text{IP}}^2 < 25$
	$\chi_{\text{VTX}}^2 < 9$	$\chi_{\text{VTX}}^2 < 9$
	DOCA < 0.3 mm	DOCA < 0.3 mm
		$\tau < 13.248 \text{ ps}$
		$p_T > 500 \text{ MeV}/c$
μ or h	$\chi_{\text{trk}}^2 < 3$	$\chi_{\text{trk}}^2 < 3$
	isMuon = True	Ghost probability < 0.3
	Minimum $\chi_{\text{IP}}^2 > 25$	Minimum $\chi_{\text{IP}}^2 > 25$
	$p_T > 0.25 \text{ GeV}/c$	$0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$

Table 4.4 Selection requirements applied during the stripping selection for Run 1 data used in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction analysis [46, 47] to select $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays. M_{PDG} corresponds to the Particle Data Group [5] mass of each particle.

direction cosine should be very close to one. Requiring candidates to have positive value ensures candidates are travelling in the wrong direction are removed;

- the flight distance χ^2 (χ_{FD}^2) - this is computed by performing the fit for the production vertex of a particle with and without the tracks originating from the decay vertex of the particle. For a B meson the χ_{FD}^2 is likely to be large because B mesons have long lifetimes therefore the tracks from its decays vertex will not point towards the production vertex;
- track fit χ^2 per degree of freedom (χ_{trk}^2) - provides a measure of the quality of a fitted track, placing an upper limit on this variable removes poor quality tracks and backgrounds composed of poorly reconstructed decays;
- vertex fit χ^2 per degree of freedom (χ_{VTX}^2) - provides a measure of how well tracks can be combined to form a vertex, placing an upper limit on this variable removes poorly constrained vertices and backgrounds composed of poorly reconstructed decays;
- distance of closest approach (DOCA) - this is the distance of closest approach of the tracks of the two daughter particles that make up a parent particle. It

Particle	$B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$	Particle	$B_s^0 \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$
B^+	$ M_{B^+} - M_{B^+}^{\text{PDG}} < 500 \text{ MeV}/c^2$ $\chi_{VTX}^2 < 45$ $\chi_{\text{IP}}^2 < 25$	B_s^0	$ M_{B_s^0} - M_{B_s^0}^{\text{PDG}} < 500 \text{ MeV}/c^2$ $\chi_{VTX}^2 < 75$ $\chi_{\text{IP}}^2 < 25$
J/ψ	$ M_{J/\psi} - M_{J/\psi}^{\text{PDG}} < 100 \text{ MeV}/c^2$ DIRA > 0 $\chi_{\text{FD}}^2 > 225$ $\chi_{\text{VTX}}^2 < 9$ DOCA < 0.3 mm	J/ψ	$ M_{J/\psi} - M_{J/\psi}^{\text{PDG}} < 100 \text{ MeV}/c^2$ DIRA > 0 $\chi_{\text{FD}}^2 > 225$ $\chi_{\text{VTX}}^2 < 9$ DOCA < 0.3 mm
μ^\pm	$\chi_{\text{trk}}^2 < 3$ isMuon = True Minimum $\chi_{\text{IP}}^2 > 25$ $p_T > 0.25 \text{ GeV}/c$	μ^\pm	$\chi_{\text{trk}}^2 < 3$ isMuon = True Minimum $\chi_{\text{IP}}^2 > 25$ $p_T > 0.25 \text{ GeV}/c$
K^+	$\chi_{\text{trk}}^2 < 3$ $p_T > 0.25 \text{ GeV}/c$ Minimum $\chi_{\text{IP}}^2 > 25$	ϕ	$ M - M_\phi^{\text{PDG}} < 20 \text{ MeV}/c^2$ Minimum $\chi_{\text{IP}}^2 > 4$
		K^\pm	$\chi_{\text{trk}}^2 < 3$ $p_T > 0.25 \text{ GeV}/c$

Table 4.5 Selection requirements applied during the stripping selection for Run 1 data used in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction analysis [46, 47] to select $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays. M_{PDG} corresponds to the Particle Data Group [5] mass of each particle.

is computed from straight tracks in the VELO. For the decay products of a particle, for example the muons from $B_{(s)}^0 \rightarrow \mu^+\mu^-$, this distance would ideally be zero because the muons originate from the same vertex;

- decay time (τ) - is the length of time a particle lives as it travels from its production vertex to its decay vertex. Applying an upper decay time cut removes unphysical background decays;
- isMuon - particle identification variable, defined in Section 3.2.2, that returns True for muons and False for other particles;
- transverse momentum (p_T) - the component of a particle's momentum perpendicular to the beam axis. Decay products of B mesons are expected to have relatively high p_T due to the heavy masses of B mesons, however, an upper limit removes unphysical backgrounds;
- momentum (p) - an upper limit is placed on the momentum of a particle removes unphysical backgrounds;
- ghost probability - defined in Section 3.2.1.4, provides the probability of a track being composed of random hits in the detector, tracks from the passage of real particles will have a low ghost probability;
- impact parameter χ^2 (χ_{IP}^2) - this is the change in the fit χ^2 for a primary vertex (PV) caused by removing one track in the fit. In a $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decay, the $B_{(s)}^0$ is produced at the PV therefore it should have a small χ_{IP}^2 value whereas the muons will be displaced from the PV and will have a large χ_{IP}^2 because of the relatively long lifetime of the $B_{(s)}^0$;
- minimum impact parameter (χ_{IP}^2) - this is the χ_{IP}^2 of the muons with respect to all PVs in the event, this parameter is used to remove prompt muons created at any PV in the event and therefore reduce the prompt combinatorial background.

The stripping selection imposes a greater number of cuts to select $B \rightarrow h^+h^-$ decays compared to $B_s^0 \rightarrow \mu^+\mu^-$ because $B \rightarrow h^+h^-$ decays are much more abundant. Therefore extra cuts are needed to reduce the number of events passing the stripping to an acceptable level. The cuts applied to only $B \rightarrow h^+h^-$ decays in the stripping are applied to $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates in the offline selection.

4.3.2.2 Investigation of the stripping selection

An investigation into the efficiency of selection cuts used in the stripping lines to select $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h^-$ and $B^+ \rightarrow J/\psi K^+$ decays is presented in this section. The efficiency of the $B_s^0 \rightarrow J/\psi\phi$ stripping line is not investigated because

it is not used as a normalisation channel for the branching fraction measurements. First, the efficiencies of selection cuts used in each stripping line to select its signal decay are evaluated using simulated decays. This is done to identify which cuts can be changed to improve the selection efficiency of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Then, the efficiencies of cuts used in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line are compared to the efficiencies of the $B \rightarrow h^+ h'$ and $B^+ \rightarrow J/\psi K^+$ stripping lines for different cut values. This is done because the selection efficiencies of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and the normalisation channels must be similar to keep systematic uncertainties of the normalisation procedure under control and any change made to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line must be propagated through to the other stripping lines. These studies show that improvements to the selection efficiency of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line can be made by changing the cuts on the $B_{(s)}^0 \chi_{\text{FD}}^2$ and the muon minimum χ_{IP}^2 of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates. Therefore the efficiencies of a set of different cut values applied to these variables are investigated. One of the main purposes of the stripping selection is to reduce the size of the data collected by LHCb to a manageable level that can be analysed. Therefore the change in the amount of data retained by the stripping lines is evaluated and new cut values are chosen. Finally the cuts used to select candidates in the stripping for the branching fraction measurements described in Chapter 5 are summarised at the end of this section.

Stripping line efficiency

The efficiencies of the selection cuts in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'$ and $B^+ \rightarrow J/\psi K^+$ stripping lines at selecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'$ and $B^+ \rightarrow J/\psi K^+$, respectively, are evaluated using simulated decays. The efficiencies are evaluated as the fraction of reconstructed decays within the detector angular acceptance that pass a stripping selection cut. Several selection requirements are applied to the simulated decays before the efficiencies are evaluated. These are: the p_T requirement on the daughter particles in the decays; the χ_{trk}^2 requirement; and all muons are required to have `isMuon = True`. These cuts are applied when ROOT files are created, as described in Section 3.2.4, and cannot be changed. No trigger requirements have been applied so that only the effect of the stripping selection on the efficiencies can be assessed. During the simulation of particle decays the trigger is run in *pass through* mode so that all reconstructed decays are saved, not just those that have passed a trigger line.

The efficiencies have been evaluated for each cut that is used in all three stripping lines and also for the total selection efficiency of each stripping line, the results are shown in Table 4.6.

The efficiencies for most of the stripping cuts are $\sim 97\%$ or higher. However, the efficiencies of the cuts on the χ_{FD}^2 of the $B_{(s)}^0$ or J/ψ and the daughter χ_{IP}^2 of the muon

Requirement	Efficiency			
	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$	$B \rightarrow h^+ h'^-$	$B^+ \rightarrow J/\psi K^+$
$ M_B - M_B^{PDC} $	(100.00 \pm 0.00)%	(100.00 \pm 0.00)%	(98.25 \pm 0.02)%	(99.73 \pm 0.02)%
$B_{(s)}^0$ or J/ψ DIRA	(99.41 \pm 0.01)%	(99.47 \pm 0.01)%	(99.47 \pm 0.01)%	(95.83 \pm 0.08)%
$B_{(s)}^0$ or J/ψ χ_{FD}^2	(83.74 \pm 0.06)%	(83.96 \pm 0.06)%	(83.83 \pm 0.06)%	(82.90 \pm 0.15)%
$B_{(s)}^0$ or J/ψ χ_{IP}^2	(96.78 \pm 0.03)%	(96.93 \pm 0.03)%	(97.44 \pm 0.03)%	(97.52 \pm 0.06)%
$B_{(s)}^0$ or J/ψ χ_{vtx}^2	(97.21 \pm 0.03)%	(97.18 \pm 0.03)%	(97.68 \pm 0.02)%	(96.78 \pm 0.07)%
$B_{(s)}^0$ or J/ψ DOCA	(99.82 \pm 0.01)%	(99.80 \pm 0.01)%	(99.83 \pm 0.01)%	(99.58 \pm 0.03)%
μ, h, K^+ minimum χ_{IP}^2	(80.16 \pm 0.06)%	(80.62 \pm 0.06)%	(79.66 \pm 0.07)%	(86.98 \pm 0.14)%
Total after above cuts	(71.29 \pm 0.07)%	(71.82 \pm 0.07)%	(70.97 \pm 0.07)%	(71.30 \pm 0.18)%
Total after all cuts	(71.29 \pm 0.07)%	(71.82 \pm 0.07)%	(70.70 \pm 0.07)%	(62.25 \pm 0.20)%

Table 4.6 Signal efficiencies of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$ and $B^+ \rightarrow J/\psi K^+$ stripping lines using simulated $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow K^+ \pi^-$ and $B^+ \rightarrow J/\psi K^+$ decays, respectively. The stripping selection cuts are listed in Tables 4.4 and 4.5 and efficiencies are evaluated for cuts are shared between all stripping lines and the total efficiency for each line.

or hadron pair are lower at 83% and 80%, respectively. Therefore improvements to the stripping selection efficiencies could be achieved by altering these two selection requirements.

Comparison of different stripping lines

The selection efficiencies are very similar for stripping cuts across the different decays, fitting the requirement that the selection of signal and normalisation decays used in the branching fraction measurement are as similar as possible.

The similarity of the selection efficiencies for the signal and normalisation decays is further illustrated in Figures 4.1 and 4.2, which show the ratio of selection efficiencies of $B_s^0 \rightarrow \mu^+\mu^-$ decays to $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ decays for a range of selection cuts. With the exception of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ χ_{IP}^2 cuts on the daughter particles, the ratio of efficiencies is well within 3% of unity for the range of cuts values shown. The ratio of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ efficiencies for the daughter particle χ_{IP}^2 , Figure 4.1d, markedly deviates from unity, showing that the χ_{IP}^2 distribution of the muons and kaon are very different as seen previously in reference [151]. If the χ_{FD}^2 , B_s^0 or J/ψ χ_{IP}^2 and χ_{VTX}^2 selection cuts are applied to the simulated events before the daughter χ_{IP}^2 requirement the ratio of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ efficiencies is much closer to 1. The stability of the ratios of selection efficiencies across a large range of cuts values shows that changing a cut value in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ selection will have a similar impact on the efficiencies of the normalisation decays.

Efficiencies of different cuts values

The set of events removed by each cut in the stripping selection is not independent. Therefore the effect of changing one cut on the total efficiency of a stripping selection must be considered. Figure 4.3 shows the total efficiency of the $B_s^0 \rightarrow \mu^+\mu^-$ stripping line on simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays for a range of χ_{FD}^2 and daughter χ_{IP}^2 cut values. As expected the lower the cut values the more efficient the stripping line becomes. It is important that any increase in $B_s^0 \rightarrow \mu^+\mu^-$ selection efficiency from the stripping is not removed when the trigger requirements are applied. Figure 4.4 shows that the trigger efficiencies are relatively flat across a large range of χ_{FD}^2 and daughter χ_{IP}^2 cut values. Therefore the efficiency gained by a change in the stripping selection is not lost when trigger requirements are imposed. The selection efficiency for $B^0 \rightarrow \mu^+\mu^-$ is very similar to $B_s^0 \rightarrow \mu^+\mu^-$ as seen in Table 4.6, therefore only $B_s^0 \rightarrow \mu^+\mu^-$ have been studied for different stripping selection cut values.

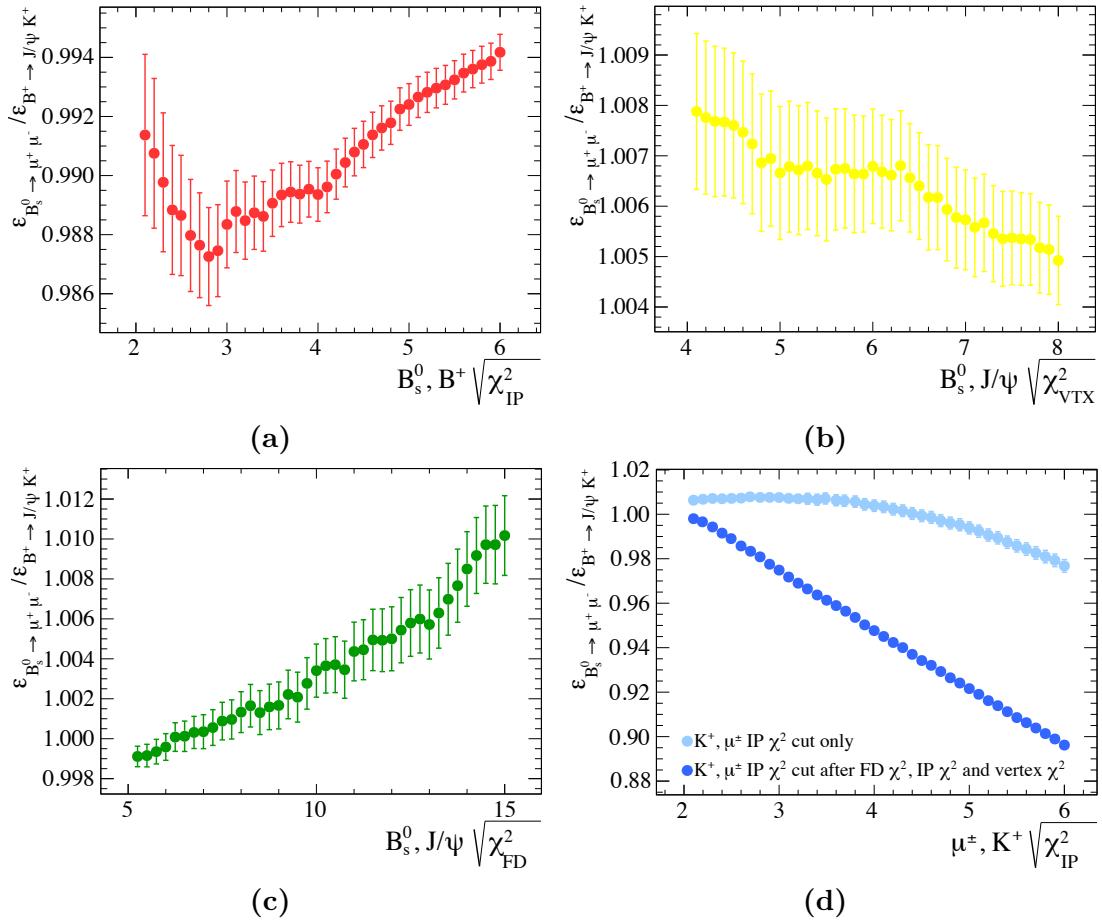


Fig. 4.1 Ratio of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ to $B^+ \rightarrow J/\psi K^+$ stripping efficiencies when each cut has been applied independently of all other cuts. The ratios are shown for cuts on the B meson χ_{IP}^2 (a), B meson and J/ψ χ_{VTX}^2 (b) and χ_{FD}^2 (c) and the muon and kaon χ_{IP}^2 (d). The square root of each χ^2 is used to condense the x -axis of the plots and follows the previous work in reference [151].

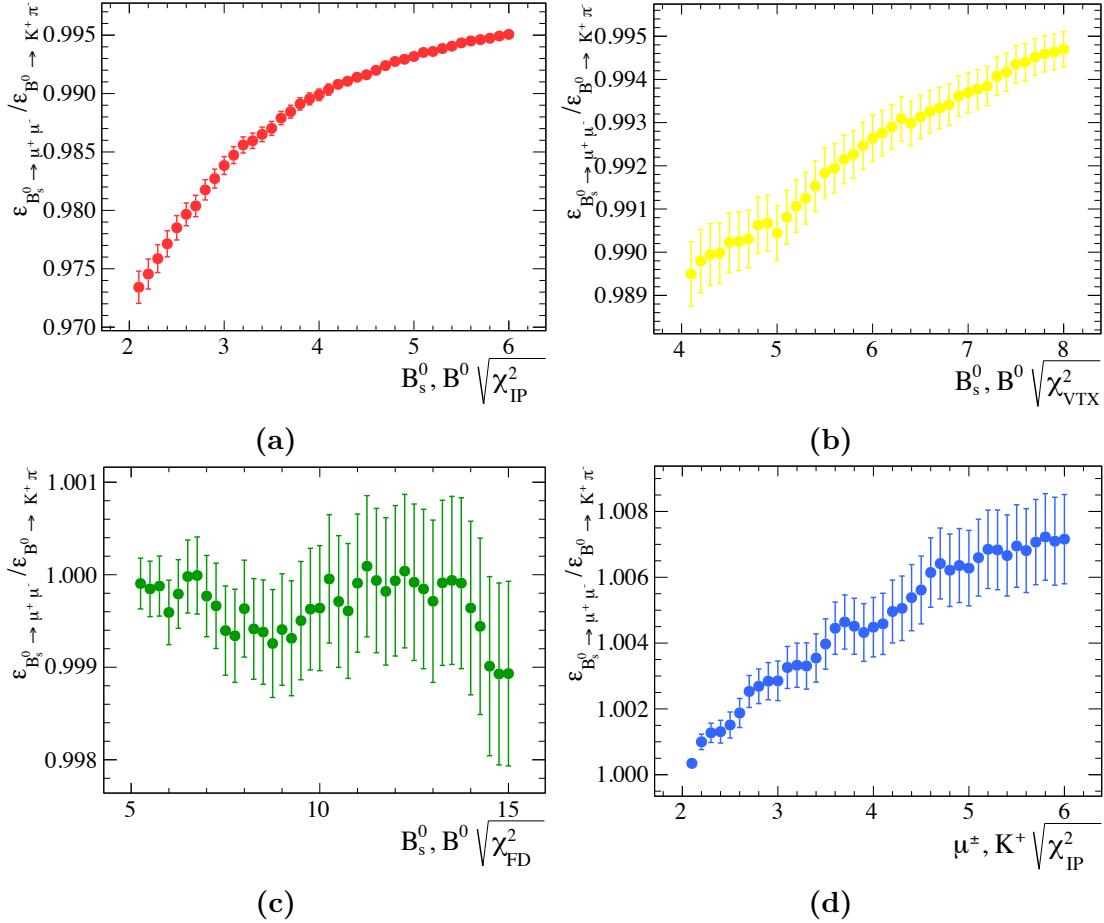


Fig. 4.2 Ratio of $B_s^0 \rightarrow \mu^+ \mu^-$ to $B^0 \rightarrow K^+ \pi^-$ stripping efficiencies when each cut has been applied independently of all other cuts. The ratios are shown for cuts on the B meson χ^2_{IP} (a), χ^2_{VTX} (b) and χ^2_{FD} (c) and the muon and kaon χ^2_{IP} (d). The square root of each χ^2 is used to condense the x -axis of the plots and follows the previous work in reference [151].

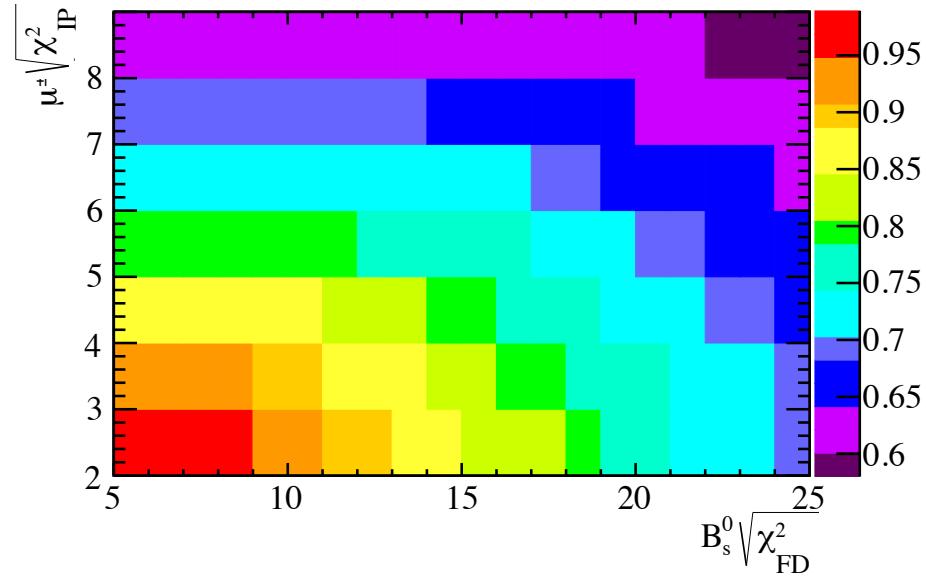


Fig. 4.3 Efficiency of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line to select $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays for a range of cuts on the $B_s^0 \sqrt{\chi^2_{\text{FD}}}$ and the minimum muon $\sqrt{\chi^2_{\text{IP}}}$.

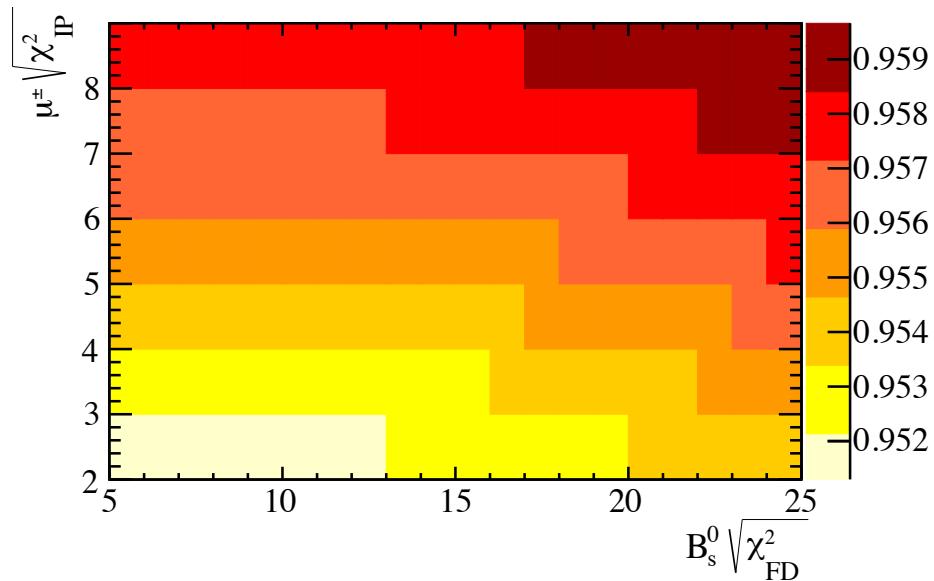


Fig. 4.4 Trigger efficiencies of $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays across a range of $B_s^0 \sqrt{\chi^2_{\text{FD}}}$ and the minimum muon $\sqrt{\chi^2_{\text{IP}}}$ cut values for the trigger requirements listed in Table 4.3 for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays.

Data retention

One of the main purposes of the stripping selection is to reduce the size of the data set, therefore the cuts cannot be set as loose as possible and the amount of data passing the selection must be considered. Any change applied to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line must be propagated through into the stripping lines for $B \rightarrow h^+ h^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays. Therefore the retention of all stripping lines must be evaluated. The efficiencies of the $B_s^0 \rightarrow J/\psi \phi$ stripping line have not been presented because this decay is not used as a normalisation mode for the branching fraction measurements but only to cross check the results, therefore the efficiency of this stripping line compared to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line is less important.

Table 4.7 shows the total selection efficiency of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line along side the amount of data retained for the set of cuts on the χ_{FD}^2 and daughter χ_{IP}^2 for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ stripping lines. The set of chosen cuts used in Table 4.7 aims to keep both cuts as tight as possible for a certain $B_s^0 \rightarrow \mu^+ \mu^-$ efficiency.

The data retention is computed by applying the stripping selection to a sub-set of 2012 data to find the number of events that pass the stripping lines for each pair of χ_{FD}^2 and daughter χ_{IP}^2 cuts. No trigger requirements are imposed on trigger lines because the stripping selection run on the full output of the trigger. The number of events for each set of cuts is normalised to the number of events passing the original Run 1 stripping line requirements to show the fractional increase caused by loosening the cut values.

An increase of 15% can be gained in the stripping selection efficiencies by using the loosest cuts in Table 4.7. However, the loosest cuts increases the amount of data passing the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping selection by a factor of 7 and the $B \rightarrow h^+ h^-$ stripping selection by a factor of 4. Table 4.8 shows the number of Run 1 candidates passing the original stripping selection listed in Tables 4.4 and 4.5. The $B \rightarrow h^+ h^-$ stripping line lets through the most candidates where as the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line saves far fewer candidates, therefore a change in the retention of the $B \rightarrow h^+ h^-$ line is more significant than the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ line.

The final set of cuts used in the stripping selection must be a compromise between the selection efficiency and the amount of data that passes the selection. The studies detailed here show that using selection cuts of $B_s^0 \chi_{\text{FD}}^2 > 121$ ($\sqrt{\chi_{\text{FD}}^2} > 11$) and minimum muon $\chi_{\text{IP}}^2 > 9$ ($\sqrt{\chi_{\text{IP}}^2} > 3$) in the stripping lines would increase the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection efficiency by from 71% to 82% and the amount of data retained would be doubled. The increase of the data retained by the $B \rightarrow h^+ h^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ lines for equivalent cut values is smaller and the efficiencies are similar to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection efficiencies. Therefore, the cuts of $B_s^0 \chi_{\text{FD}}^2 > 121$ and minimum muon $\chi_{\text{IP}}^2 > 9$ offer a good compromise between

		Stripping cut		Stripping line efficiency		Stripping line retention			
$B_{(s)}^0$	$\sqrt{\chi_{\text{FD}}^2}$	μ^\pm	$\sqrt{\chi_{\text{IP}}^2}$	$B_s^0 \rightarrow \mu^+ \mu^-$	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow K^+ \pi^-$	$B^+ \rightarrow J/\psi K^-$	$B_s^0 \rightarrow J/\psi \phi$	
15	5.00	(71.29 \pm 0.07)	%	1.0	1.0	1.0	1.0	1.0	
14	4.25	(74.91 \pm 0.07)	%	1.5	1.3	1.1	1.1	1.3	
13	4.00	(76.84 \pm 0.07)	%	1.8	1.5	1.2	1.2	1.4	
12	3.50	(79.76 \pm 0.07)	%	2.6	1.8	1.3	1.3	1.7	
11	3.00	(82.72 \pm 0.06)	%	3.7	2.4	1.6	1.6	1.9	
10	2.75	(84.86 \pm 0.06)	%	4.7	3.0	1.7	1.7	2.1	
9	2.50	(86.96 \pm 0.06)	%	6.8	3.9	2.0	2.0	2.2	

Table 4.7 Efficiency of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line to select $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the change in the data retention for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ stripping lines for a range of χ_{FD}^2 and daughter χ_{IP}^2 cut values. The amount of data passing each selection has been normalised to the original set of stripping select cuts of $\sqrt{\chi_{\text{FD}}^2} > 15$ and $\sqrt{\chi_{\text{IP}}^2} > 5$. The uncertainty on the normalised retention is less than 1%.

Stripping Lines	Selected events	Retention / %
$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	898880	0.0022
$B \rightarrow h^+ h^-$	14502295	0.0831
$B^+ \rightarrow J/\psi K^+$	3344568	0.0087
$B_s^0 \rightarrow J/\psi \phi$	456787	0.0011
Total	18745743	-

Table 4.8 Number of events passing stripping lines used for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurement in reference [46, 47] that are listed in Tables 4.4 and 4.5 and the percentage of the total LHCb data set that they correspond to. The total does not include correlation between lines and the requirements of $\chi_{\text{FD}}^2 > 225$ and daughter $\chi_{\text{IP}}^2 > 25$ are used.

signal efficiency and the amount of data retained. The stripping lines have been updated to include the new, looser cut values that will be used in future studies of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays.

Final stripping selection

Although the new looser cut values would improve the efficiency of identifying $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at this step in the selection process, the new looser cut values are not used in the measurements presented in this dissertation. The multivariate classifier used to separate signal and combinatorial background decays, described in Section 4.3.4.3, is trained on simulated $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays. As discussed in Section 4.2, cuts are applied to $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays when the decays are simulated. Only decays that pass the original χ_{FD}^2 and daughter χ_{IP}^2 requirements are available in the simulated sample. Therefore to ensure the best performance of the classifier on data, the same cuts are applied to data that are applied to the simulated samples. Therefore the original cuts on χ_{FD}^2 and daughter χ_{IP}^2 listed in Table 4.4 are used to select $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates.

4.3.2.3 Additional offline cuts

Additional selection requirements are applied after the stripping to remove specific backgrounds. A lower bound is placed on the B meson transverse momentum to remove pairs of muons originating from $pp \rightarrow p\mu^+\mu^-p$ decays and a J/ψ veto is used to remove backgrounds from $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays. Semi-leptonic $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays, where $J/\psi \rightarrow \mu^+ \mu^-$, are backgrounds for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays when a muon from the J/ψ forms a good vertex with the muon from the B_c^+ decay. Due to the high mass of the B_c^+ this could place mis-reconstructed candidates within the B_s^0 mass

window. A ‘ J/ψ veto’ is used to remove background events from $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays. The veto removes events where one muon from the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidate combined with any other oppositely charged muon in the event has $|m_{\mu\mu} - m_{J/\psi}| < 30 \text{ MeV}/c^2$.

The offline selection of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays includes the momentum, ghost track probability and decay time cuts made in the $B \rightarrow h^+ h'^-$ stripping line, but were absent in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ stripping line. Also a narrower mass range of 4900 - 6000 MeV/c^2 is imposed to remove $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ backgrounds. The stripping selection for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is kept loose to allow for the study of background decays in data.

The selection applied to Run 1 and Run 2 data is the same for all variables except the track ghost probability and χ^2_{trk} . Slightly looser cuts of track ghost probability < 0.4 and $\chi^2_{\text{trk}} < 4$, are used in Run 2 to take advantage of changes in the reconstruction that were introduced for Run 2.

Table 4.13 summarises all selection cuts used to identify $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at the end of this section.

4.3.3 Particle identification

In the selection of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, particle identification variables are particularly useful to reduce the backgrounds coming from mis-identified semi-leptonic decays and $B \rightarrow h^+ h'^-$ decays and also help to reduce the number of combinatorial background decays. On top of the isMuon requirement used in the stripping selection, ProbNN variables, defined in Section 3.2.2.4, are used. A linear combination of these variables

$$\text{PID}_\mu = \text{ProbNN}\mu \times (1 - \text{ProbNN}K) \times (1 - \text{ProbNN}p) \quad (4.1)$$

is used to refine the selection of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates. The ProbNN K variable is effective at removing mis-identified $B \rightarrow h^+ h'^-$ backgrounds and the ProbNN p variable is effective at removing $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ backgrounds.

Different tunings of the algorithms used in the ProbNN variables are used to select candidates in Run 1 and 2015 data compared to 2016 data. The tunings have different efficiencies to select particles therefore the cut values placed on PID_μ are different for each tuning. The cuts applied to data are $\text{PID}_\mu > 0.4$ for Run 1 and 2015 data and $\text{PID}_\mu > 0.8$ for 2016 data. The cut value on PID_μ for the Run 1 and 2015 is optimised using pseudoexperiments to sufficiently reduce the background decays and give the highest sensitivity to the $B^0 \rightarrow \mu^+ \mu^-$ decays. Accurate particle identification is most important for $B^0 \rightarrow \mu^+ \mu^-$ decays because the backgrounds from $B \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ pollute the B^0 mass window. The cut value for 2016 was chosen to have the same or lower background rejection as the Run 1 and

2015 cut, however the 2016 tuning has a better performance therefore the final cut choice has a higher efficiency for selecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays.

4.3.4 Multivariate Classifiers

The selection described so far removes a large number of background candidates. However, because $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays occur very rarely, the data is still dominated by long-lived combinatorial background. To improve the separation of signal and background decays multivariate classifiers are used.

A multivariate classifier is an algorithm that learns differences between signal and background decays. The classifier is given two input samples, one containing only signal decays and the other containing only background decays and a set of input variables. The input variables have different distributions for signal and background decays. The classifier uses the distributions of the input variables along with its knowledge of which decays are signal and background to learn the difference between the two types of decays. The algorithm is then applied to a data set containing an unknown mixture of signal and background decays to separate them. For each decay the algorithm produces a number, typically between -1 and $+1$, where high numbers indicate signal-like decays and low numbers indicate background-like decays.

Two multivariate classifiers are used to identify $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Both classifiers are a type called a Boosted Decision Tree (BDT), described in Section 4.3.4.1. The first classifier, described in Section 4.3.4.2, is called the BDTS. It is used to remove candidates that are very unlikely to be signal by placing a cut on the BDTS output. The second classifier, described in Section 4.3.4.3, it called the global BDT. The output of the global BDT is used to classify candidates into bins containing increasing proportions of signal candidates. The branching fractions are measured from the invariant mass distribution of the two muons in bins of BDT output as described in Chapter 5. The BDTS is necessary to reduce the background to a more manageable level for the global BDT.

4.3.4.1 Boosted Decision Trees

A BDT is made up of the combined outputs of separate decision trees. A decision tree begins with a data sample, where each decay is known to be either signal or background and a set of variables describing them. The decision tree applies a cut on a variable that will be the most effective at separating the signal and background in the sample and creates two sub-samples. Another cut is then applied to each of the sub-samples to further separate signal from background. This process is repeated until either a certain number of cuts, defined as the depth of the tree, or the number of candidates in each sub-sample has reached a minimum value. Each sub-sample

produced at the end of the tree is called a leaf. The tree uses the knowledge of whether decays are signal or background to assign a value of +1 or -1 to every decay. A decay is given a value +1 if it is in a leaf where the majority of decays are signal and the value -1 if it is in a leaf that has a majority of background decays. The final decisions made by the tree are not perfect, some signal (background) decays will be mis-classified as background and given the value of -1 (+1).

Often a single decision tree is not particularly good at classifying decays; there is no way to correct mis-classified decays in the leaves, and it is particularly sensitive to statistical fluctuations in the training samples. A BDT combines the output of numerous decision trees to improve the classification of decays and reduce the dependence of the final decisions on statistical fluctuations. A BDT starts with a decision tree and assigns weights to decays in the signal and background samples depending on whether the output of the first decision tree classified them correctly. The weighted sample is then used as the input for the training of the next decision tree. The weights are designed so that the next tree is more likely to correctly classify previously mis-classified decays. This process is repeated until a certain number of trees have been trained. The re-weighting process is known as “boosting” and the weights applied to the samples are taken into account when combining the output value of each decision tree into the overall output of the BDT. The output of a BDT will be a number between -1 and +1 where high numbers indicate signal and low numbers indicate background.

The TMVA package [152] is used to develop and train the BDTs. The package provides several different methods of boosting that can be used. The adaptive boosting method was found to produce the most effective BDT at separating $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ from combinatorial background. This method of boosting assigns decays incorrectly classified by one tree the weight, w , before being used as the input to the next decision tree. The weights assigned are given by

$$w = \frac{1-f}{f}, \text{ where } f = \frac{\text{misclassified events}}{\text{total events}}. \quad (4.2)$$

Therefore, incorrectly assigned candidates are given a higher weight than correctly classified candidates. The ‘speed’ at which the boosting occurs is controlled by the parameter β where $w \rightarrow w^\beta$. The parameter β is specified in the training of the decision tree and a large number of boosting steps can improve the performance of the BDT.

The ability of a BDT to correctly identify signal and background candidates depends on three main factors:

- the size of the training samples - a large training sample is useful to prevent the BDT from being sensitive to statistical fluctuations and contains more

information the classifier can use to learn the difference between signal and background;

- the input variables - different distributions in the input variables for signal and background candidates enable the classifier to easily separate the types of candidates. The overall performance is insensitive to poorly discriminating variables that are included; and
- parameters that dictate the BDT training - the training of a BDT is specified by several parameters; the number of trees (NTrees), the tree depth (MaxDepth), the minimum number of events a leaf can contain (nEventsMin or MinNodeSize²); the ‘speed’ at which the boosting occurs (β) and the number of cut values that a tree tries for a variable before making a decision (nCuts).

These three factors affect the performance of the BDT. However, the importance of each varies. Together they are used to prevent the BDT being very sensitive to the statistical fluctuations in the training sample. This is called overtraining; an overtrained BDT is extremely accurate at classifying the candidates in the training sample but performs poorly at classifying candidates in a statistically independent sample. Although this is less common in BDTs than single decision trees, it can be avoided by having a sufficiently large training sample or by limiting the depth of trees or the number of trees in the BDT.

4.3.4.2 The BDTS

The BDTS uses input variables similar to those in the stripping selection to classify events:

- χ^2_{IP} of the $B_{(s)}^0$;
- χ^2_{VTX} of the $B_{(s)}^0$;
- direction cosine of $B_{(s)}^0$;
- distance of closest approach of the tracks of the muons;
- minimum χ^2_{IP} of the muons with respect to all primary vertices in the event; and
- impact parameter (IP) of the $B_{(s)}^0$, this is the distance of closest approach of the B to the primary vertex.

²nEventsMin is the minimum number of decays in a leaf and MinNodeSize is the number of decays in a leaf given as a percentage of the training sample size. The parameter specified in the training depends on the version of the TMVA package used.

Selection applied to BDTS training samples.	
B_s^0	μ^\pm
$\chi_{\text{FD}}^2 > 225$	$p_T > 500 \text{ MeV}/c$
$\chi_{\text{IP}}^2 < 25$	$\chi_{\text{trk}}^2 < 3$
$\chi_{\text{VTX}}^2 < 9$	Minimum $\chi_{\text{IP}}^2 > 25$
$\text{DOCA} < 0.3 \text{ mm}$	$0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$
$\tau < 13.248 \text{ ps}$	$p < 500 \text{ GeV}/c$
$p_T > 500 \text{ MeV}/c$	
DIRA > 0	
Trigger line	Decision
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC

Table 4.9 Selection cuts applied to select the signal and background samples used to train the BDTS. The isMuon requirement is not applied to the muons so that the BDTS can be used on $B \rightarrow h^+h'^-$ decays.

The signal and background samples used to train the BDTS are simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays and background candidates in a sample of Run 1 data from the mass ranges 4800 - 5000 MeV/ c^2 and 5500 - 6000 MeV/ c^2 . The selection cuts listed in Table 4.9 are applied to the training samples and the training parameters used are listed in Table 4.10. The output of the BDTS is flattened between 0 and 1 so that signal is uniformly distributed across the range and background is peaked at zero as illustrated in Figure 4.5. The BDTS is applied to all candidates passing the $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$ and $B^+ \rightarrow J/\psi K^+$ stripping lines, and candidates are required to have a BDTS value above 0.05. When the BDTS is applied to $B^+ \rightarrow J/\psi K^+$ decays the distance of closest approach of the muons refers to the muons in the J/ψ and the χ_{VTX}^2 is of the J/ψ . The performance of the BDTS at removing backgrounds is illustrated in Figure 4.6.

4.3.4.3 Global BDT

The global BDT is the final step in identifying $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays and it is very effective at separating them from long-lived combinatorial background decays. The discriminating power achieved by the global BDT is mostly dependent on isolation criteria. Isolation criteria provide a measure of how far away each muon from a $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidate is from other tracks in the event. The tracks of the muons from a real $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays will be, in general, far from other tracks in the event

Parameter	Value
nTrees	250
nEventsMin	400
MaxDepth	3
β	1.0
nCuts	20

Table 4.10 Training parameters used to specify the training of the BDTS.

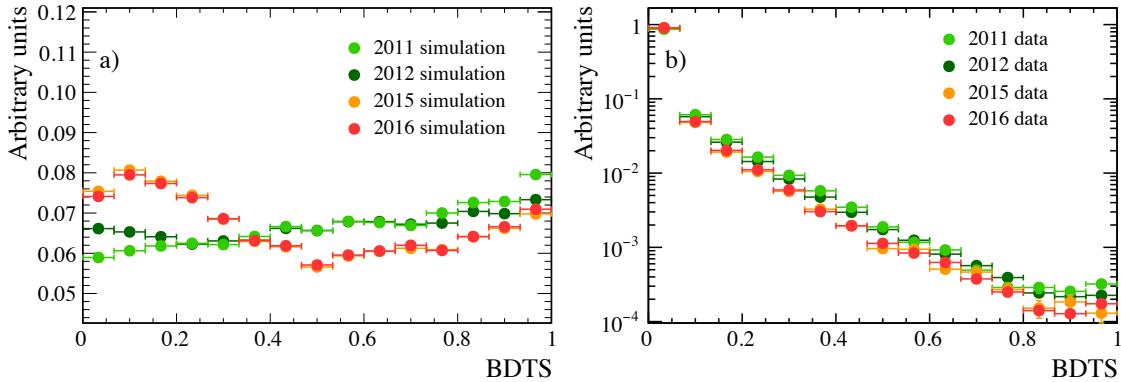


Fig. 4.5 Normalised BDTS response for a) simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays and b) $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates in data with a mass above 5447 MeV/ c^2 consisting of background decays.

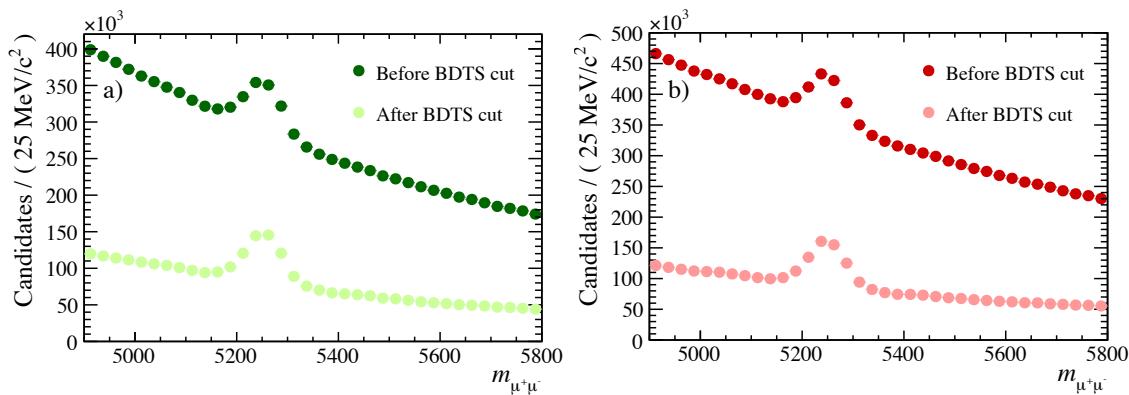


Fig. 4.6 Invariant mass spectrum for $B \rightarrow h^+ h^-$ decays in a) 2012 and b) 2016 data passing the selection requirements in Table 4.9 before and after the BDTS cut is applied.

because the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decay tree contains no other tracks apart from the muons. However, long-lived combinatorial background arises from semi-leptonic decays where the muon tracks are likely to be close to other tracks that originate from the same decay tree. Isolation criteria are very useful in the selection of very rare decays like $B_s^0 \rightarrow \mu^+\mu^-$ because they enable background to be removed whilst keeping a high efficiency for signal decays.

Two isolation criteria are used in the global BDT, one compares long tracks in the event to the muons in $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates and the other compares VELO tracks in the event to the muons. The definition of the track types are given in Section 3.2.1.4. The isolation variables are built from the output of BDTs. For each type of track a BDT is trained on simulated $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu^+\mu^- X$ decays using a set of input variables that describe track and vertex properties and the separation between muons in a $B_s^0 \rightarrow \mu^+\mu^-$ candidate and other tracks in the event. The BDT for the long track isolation criteria compares the μ^+ from a $B_s^0 \rightarrow \mu^+\mu^-$ candidate with all other long tracks in the event, excluding the track of the μ^- , and gives an output for each possible μ^+ and track pairing. The process is repeated for the μ^- . The BDT is designed to produce high output values for muons from $b\bar{b} \rightarrow \mu^+\mu^- X$ decays and a low value for muons from $B_s^0 \rightarrow \mu^+\mu^-$ decays. The long track isolation criteria of a $B_s^0 \rightarrow \mu^+\mu^-$ candidate is then composed of the sum of the highest BDT output values produced for the μ^+ and the μ^- . The same setup is used for the VELO track isolation criteria expect muons are compared to VELO tracks rather than long tracks. The separation power of these isolation criteria are shown in Figure 4.7. Full details of the isolation variables can be found in reference [153].

The isolation criteria are used along with five other variables in the global BDT. The full list of input variables used are:

- long track isolation criteria;
- VELO track isolation criteria;
- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ is the difference in azimuthal angles of the muons and $\Delta\eta$ the difference in the pseudo-rapidity of the muons;
- the smallest χ_{IP}^2 with respect to the primary vertex of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ of the muons;
- χ_{VTX}^2 of the $B_{(s)}^0$;
- χ_{IP}^2 of the $B_{(s)}^0$ with respect to the primary vertex; and
- the angle, θ , between the momentum vector of the $B_{(s)}^0$ and the vector connecting the production and decay vertices of the $B_{(s)}^0$.

Parameter	Value
nTrees	1000
MinNodeSize	1%
MaxDepth	3
β	0.75
nCuts	30

Table 4.11 Training parameters used to specify the training of the global BDT.

A comparison of the signal and background distributions of the input variables in the training samples are shown in Figures 4.7. These variables were chosen by training a BDT beginning with the most discriminating variable, the long track isolation criteria, and adding variables to determine which improved the performance to the classifier. Only variables that significantly improved the performance were included in the global BDT. The training parameters used in the BDT are listed in Table 4.11. These parameters were chosen by scanning across a range of variables and choosing those that gave the best performance. The performance of each BDT with different input variables and training parameters was evaluated by comparing the number of background decays from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates in data with masses above 5447 MeV/ c^2 remaining after different cuts on the BDT output values. The cut values compared have the same efficiency for each BDT to select simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the best performing BDT removed the lowest number of background decays for the highest signal efficiency.

Simulated $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays are used to provide large signal and background training samples for the global BDT. The complete list of selection requirements applied to the training samples used to develop global BDT are listed in Table 4.12, the same selection is applied to $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays.

The global BDT is applied to data taken in all years and in the same way as the BDTS. The final output of the global BDT is flattened to have a response between 0 and 1 that is uniform for signal and the background peaks at zero. The global BDT output for signal and background is shown in Figure 4.8 for each year of data taking. The flattening is useful for the branching fraction measurements because a simultaneous fit is applied to the invariant mass of the two muons in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidate in bins of BDT. Flattening the BDT output enables bins containing equal proportions of signal decays to be created. The signal efficiency versus the background rejection of the global BDT is shown in Figure 4.9 for all years of data taking, the performance is similar across all the years but Run 2 data has a slightly better background rejection for a given signal efficiency than Run 1.

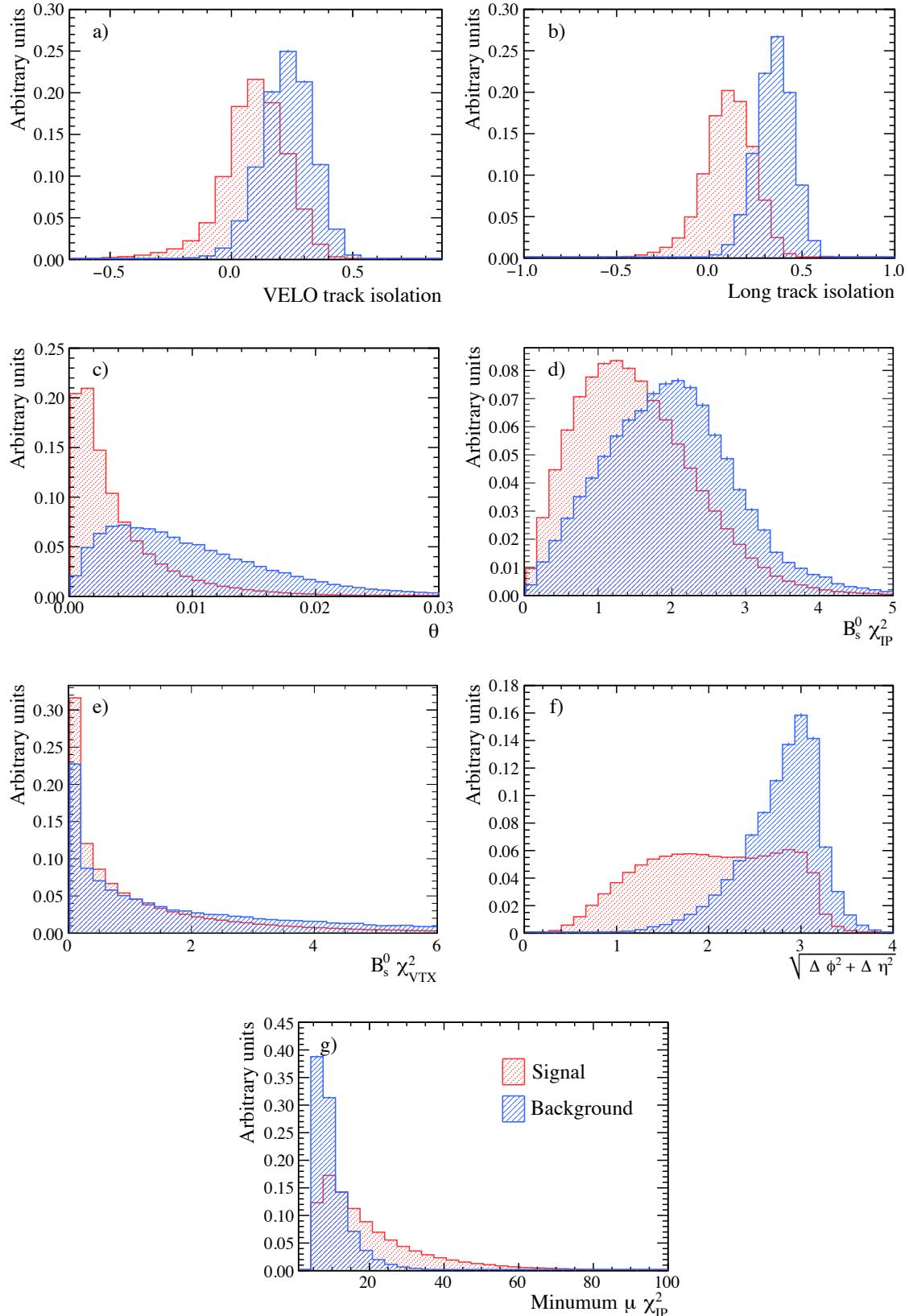


Fig. 4.7 Distributions of input variables of the global BDT from simulated $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays used to train the global BDT passing cuts in Table 4.12. Input variables are VELO track isolation criteria (a), long track isolation criteria (b) θ (c), $B_s^0 \chi^2_{\text{IP}}$ (d), $B_s^0 \chi^2_{\text{VTX}}$ (e), $\sqrt{\Delta \phi^2 + \Delta \eta^2}$ (f) and minimum muon χ^2_{IP} (g).

Selection applied to global BDT training samples.	
B_s^0	μ^\pm
$\chi_{\text{FD}}^2 > 225$	$p_T > 500 \text{ MeV}/c$
$\chi_{\text{IP}}^2 < 25$	$\chi_{\text{trk}}^2 < 3$
$\chi_{\text{VTX}}^2 < 9$	Minimum $\chi_{\text{IP}}^2 > 25$
$\text{DOCA} < 0.3 \text{ mm}$	$0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$
$\tau < 13.248 \text{ ps}$	$p < 500 \text{ GeV}/c$
$p_T > 500 \text{ MeV}/c$	isMuon = True
$\text{DIRA} > 0$	$\text{BDTS} > 0.05$
$4900 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$	
Trigger line	Decision
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC

Table 4.12 Selection cuts applied to select candidates for signal and background samples used to train the global BDT. $m_{\mu\mu}$ is the invariant mass to the two muons in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidate.

A comparison of the input variables used in the global BDT for each year of data taking is given in Appendix A.

4.3.5 Summary

The complete set of selection criteria used for identify $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in Run 1 and Run 2 data for the branching fraction measurements is listed in Tables 4.13. The selection requirements do not remove all backgrounds decays from the data set but reduce them to a level at which the branching fractions can be measured. Figure 4.10 shows a scatter plot of the mass and global BDT values for all candidates that pass the selection criteria in Run 1 and Run 2 data. The criteria to select $B \rightarrow h^+ h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ decays are composed of the trigger requirements listed Table 4.3, the stripping selection in Tables 4.4 and 4.5 and the cut on the BDTS output.

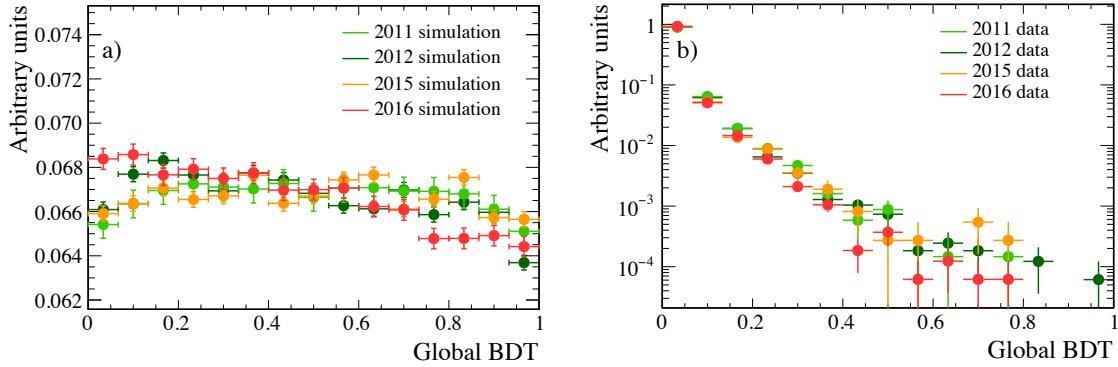


Fig. 4.8 Normalised output distributions for the global BDT for a) $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays and b) $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays in simulation and data.

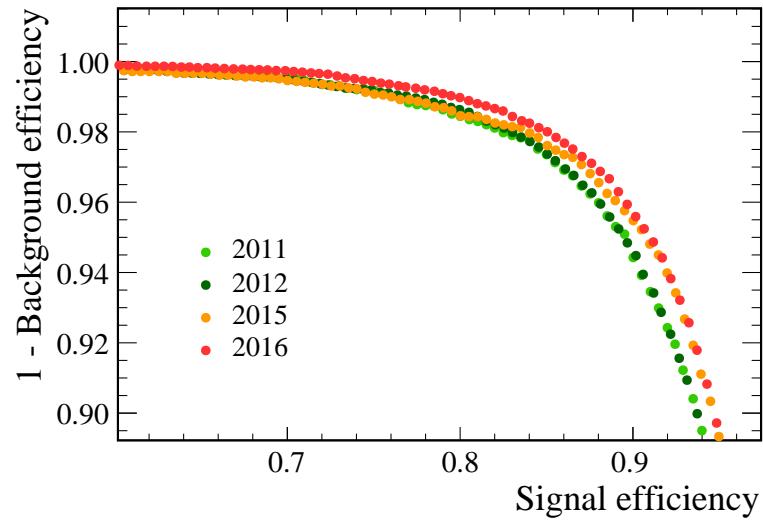


Fig. 4.9 Global BDT performance for 2011, 2012, 2015 and 2016 data taking conditions. Signal efficiency is calculated from $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays and background rejection from data passing the $B_s^0 \rightarrow \mu^+ \mu^-$ selection with $m_{\mu^+ \mu^-} > 5447 \text{ MeV}/c^2$. The performance is very similar for the different data taking years therefore only the most sensitive region is shown. The full range of BDT output values is from 0 to 1.

Particle	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$
$B_{(s)}^0$	$4900 \text{ MeV}/c^2 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$ $\text{DIRA} > 0$ $\chi_{\text{FD}}^2 > 225$ $\chi_{\text{IP}}^2 < 25$ $\chi_{\text{VTX}}^2 < 9$ $\text{DOCA} < 0.3 \text{ mm}$ $\tau < 13.248 \text{ ps}$ $p_T > 500 \text{ MeV}/c$ $\text{BDTS} > 0.05$ $ m_{\mu\mu} - m_{J/\psi} < 30 \text{ MeV}/c^2$
μ	$\chi_{\text{trk}}^2 < 3 \text{ (4)}$ $\text{Minimum } \chi_{\text{IP}}^2 > 25$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ $p < 500 \text{ GeV}/c$ $\text{ghost probability} < 0.3 \text{ (0.4)}$ $\text{isMuon} = \text{True}$ $\text{PID}_{\mu}^{\text{Run1+2015}} > 0.4 \text{ or } \text{PID}_{\mu}^{2016} > 0.8$
Trigger requirements	L0Global = DEC Hlt1Phys = DEC Hlt2Phys = DEC

Table 4.13 Selection requirements applied to select $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ for the branching fraction measurements, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

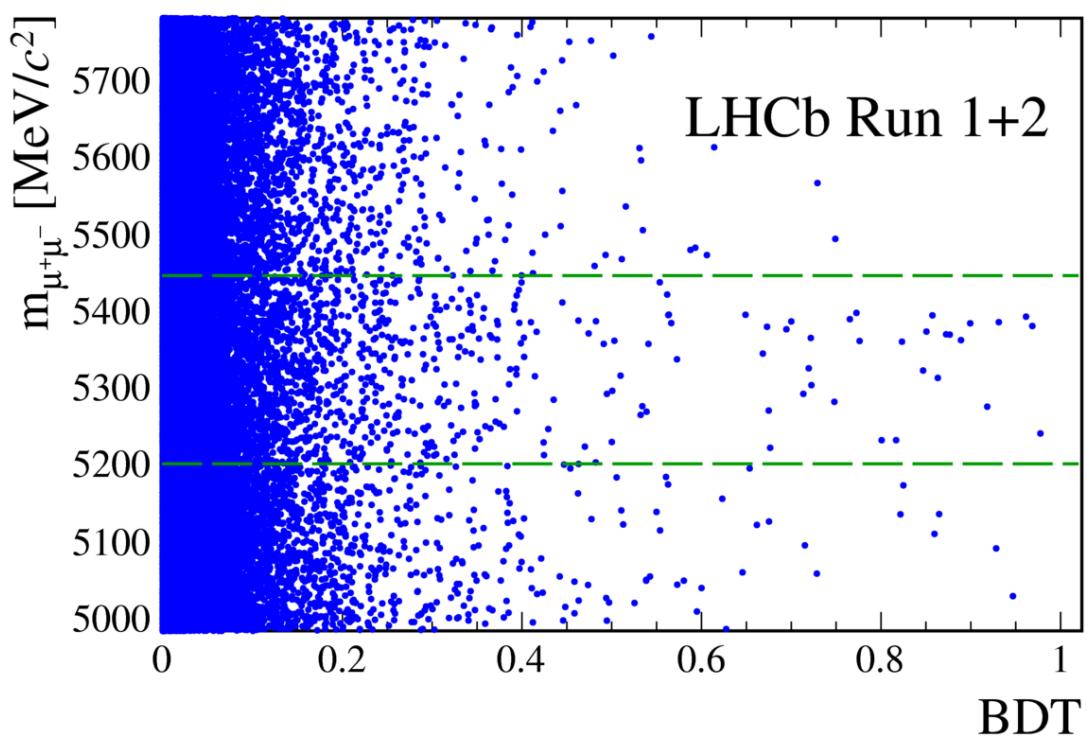


Fig. 4.10 Mass and global BDT values for candidates in Run 1 and Run 2 data that pass the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ selection criteria. The green dashed lines show the B_s^0 and B^0 mass windows.

4.4 Selection for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement

The selection criteria used to identify particles decays for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement is based on the selection used to identify candidates for the branching fraction measurements.

As well as $B_s^0 \rightarrow \mu^+\mu^-$ decays, $B^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+K^-$ and $B_s^0 \rightarrow J/\psi\phi$ decays are used to develop and validate the effective lifetime analysis strategy. There are some differences in the selection of $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays for the effective lifetime measurement compared to the branching fraction measurement to account for the different measurement strategies and because only the B_s^0 decay mode is required for the effective lifetime measurement. The selection of $B_s^0 \rightarrow J/\psi\phi$ decays is kept the same as that used for the branching fraction measurement.

The selection criteria used for $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays uses the cut based selection in Section 4.3.2 and the BDTS requirement in Section 4.3.4.2. Changes are made to the trigger requirements, the mass range of candidates and the particle identification requirements. The differences in these selection requirements compared to the selection for the branching fraction measurements and the motivation for these changes are described in Section 4.4.1, 4.4.2 and 4.4.3, respectively. Similar to the branching fraction measurement, the selection for the effective lifetime measurement uses a multivariate classifier as the final step in the selection process to separate signal from combinatorial background. A study into the best classifier for the effective lifetime measurement is described in Section 4.4.4. The selection criteria used to identify decays in data for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement are summarised in Section 4.4.5.

One important consideration for the selection of candidates for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement is the efficiency of the selection as a function of the B_s^0 decay time. This efficiency is not uniform across different decay times because the selection relies on parameters such as the χ_{IP}^2 and χ_{FD}^2 that are correlated with the decay time of $B_s^0 \rightarrow \mu^+\mu^-$ candidates. Since the effective lifetime is measured from the decay time distribution of the candidates, the selection efficiency as a function of decay time must be accurately evaluated in order to perform the measurement. The efficiency is evaluated using simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays in Chapter 6 and the procedure to evaluate the efficiency and the analysis strategy is validated using $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays. The impact of selection requirements on the decay time efficiency of $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays is important in the choice of the trigger requirements and the multivariate classifier as described in the following sections.

4.4.1 Trigger requirements

The same global triggers used to select decays for the branching fraction measurements are used to select $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ candidates for the effective lifetime measurement but different trigger decisions are used.

Candidates from $B_s^0 \rightarrow \mu^+\mu^-$ decays are required to be identified as TOS or TIS at each level of the trigger. The change in trigger decisions is motivated by the use of simulated decays in evaluating the selection efficiency as a function of decay time. The trigger efficiencies for candidates that are triggered as DEC, but not as TIS or TOS, are not well modelled in simulated decays because they depend on the underlying pp event. Therefore only candidates triggered by TOS or TIS decisions are used so that the selection as a function of decay time can be accurately modelled. Candidates triggered by DEC decisions, but not TIS or TOS do not pose the same problem for the branching fraction analysis because the selection and trigger efficiencies are evaluated using different methods as discussed in Section 5.4.

Candidates from $B \rightarrow h^+h'^-$ decays are required to be identified as TIS at each level of the trigger. In general trigger lines designed to select particle decays containing muons have a uniform efficiency for candidates with different decay times, however this is not the case for trigger lines designed to select $B \rightarrow h^+h'^-$ decays. These lines rely on information about candidate IP and χ_{IP}^2 to make decisions at the HLT level. For $B \rightarrow h^+h'^-$ decays to be useful as a validation channel the efficiency of the trigger requirements as a function of the decay time should be similar to the $B_s^0 \rightarrow \mu^+\mu^-$ triggers. This is achieved by requiring decays to be TIS at each level of the trigger.

In summary, the requirements imposed on the trigger to select $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays are shown in Table 4.14.

4.4.2 Mass range

The mass of $B_s^0 \rightarrow \mu^+\mu^-$ candidates is restricted to the range 5320 - 6000 MeV/ c^2 , the motivation for the narrower mass window compared to the selection of candidates for the branching fraction measurements comes from the optimisation of the measurement strategy detailed in Section 6.4. The lower mass bound now lies on the low edge of the B_s^0 mass window, therefore $B^0 \rightarrow \mu^+\mu^-$ candidates and backgrounds from mis-identified decays are almost completely removed. The dominant background left in the data set is from combinatorial background.

Similarly, $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays have a reduced mass range compared to the selection of $B \rightarrow h^+h'^-$ decays for the branching fraction measurements; $B \rightarrow h^+h'^-$ decays must be in the mass range 5100 - 5500 MeV/ c^2 in order to remove contributions from exclusive backgrounds.

Trigger Line	Trigger decision
$B_s^0 \rightarrow \mu^+ \mu^-$	
L0Global	TIS or TOS
Hlt1Phys	TIS or TOS
Hlt2Phys	TIS or TOS
$B \rightarrow h^+ h'^-$	
L0Global	TIS
Hlt1Phys	TIS
Hlt2Phys	TIS

Table 4.14 Trigger lines used to select $B_s^0 \rightarrow \mu^+ \mu^-$ and $B \rightarrow h^+ h'^-$ decays for the effective lifetime measurement.

4.4.3 Particle identification

The particle identification requirements used for selecting candidates for the branching fraction measurements were optimised to give the greatest sensitivity to $B^0 \rightarrow \mu^+ \mu^-$ decays. Backgrounds from $B \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ decays pollute the B^0 mass window and must be reduced as much as possible to enable good sensitivity of $B^0 \rightarrow \mu^+ \mu^-$ decays. The requirement placed on the linear combination of ProbNN variables in Section 4.3.3 is a compromise between background rejection and signal efficiency. However, for the effective lifetime measurement, the B^0 mode is not relevant and the mass region of selected candidates removes the majority of $B^0 \rightarrow \mu^+ \mu^-$ decays as well as backgrounds from $B \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ decays. Therefore, looser particle identification requirements can be used leading to a higher signal efficiency.

The same linear combination of ProbNN variables, PID_μ , is used because there is still a small contribution from mis-identified $B \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ decays within the mass range. Also the particle identification requirements help to reduce the number of combinatorial background decays. Different ProbNN tunes, and consequently cut values, are used for Run 1 and 2015 data compared to 2016 data. The cuts are chosen to give similar efficiencies for each data set at selecting signal and removing background and are listed in Table 4.15. The cut values have not been optimised because there are too few candidates in data after the selection and simulated decays are not used because particle identification variables are not well modelled in simulation. However, the chosen particle identification requirements are tight enough to make the expected number of mis-identified decays in the data set after the full selection negligible, as shown in Section 7.2.

Decay	Year	Particle	PID requirements
$B_s^0 \rightarrow \mu^+\mu^-$	2011, 2012, 2015	μ^\pm	$\text{PID}_\mu > 0.2$
$B_s^0 \rightarrow \mu^+\mu^-$	2016	μ^\pm	$\text{PID}_\mu > 0.4$
$B^0 \rightarrow K^+\pi^-$	2011, 2012, 2015, 2016	K^+	$\text{DLL}_{K\pi} > 10$
		π^-	$\text{DLL}_{K\pi} < -10$
$B_s^0 \rightarrow K^+K^-$	2011, 2012, 2015, 2016	K^\pm	$\text{DLL}_{K\pi} > 10$

Table 4.15 Particle identification requirements to select $B_s^0 \rightarrow \mu^+\mu^-$, $B_s^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement.

The separation of $B \rightarrow h^+h'^-$ decays into $B_s^0 \rightarrow K^+K^-$ and $B_s^0 \rightarrow K^+\pi^-$ decays is done using the $\text{DLL}_{K\pi}$ variable, defined in Section 3.2.2.4. The DLL variables are useful to separate $B \rightarrow h^+h'^-$ decays where h is either a pion or kaon because the variables compare different particle hypotheses with the pion hypotheses. The selection requirements used are given in Table 4.15 and are the same for each year of data taking.

4.4.4 Multivariate classifier

Two multivariate classifiers are used in the selection for the branching fraction measurements to separate signal and combinatorial background decays. The BDTS is used first to remove candidates that are very unlikely to be signal and to reduce the size of the data set. The global BDT is then used to classify candidates into separate bins and a simultaneous fit is then applied across the BDT bins to measure the branching fractions.

A different, simpler strategy is used to identify candidates for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement. Combinatorial background is reduced by placing a cut on the output of a multivariate classifier. Only candidates passing the selection cut are used to measure the effective lifetime. The measurement strategy is given in more detail in Chapter 6.

As a consequence of the different selection methods, two classifiers may not be necessary for the measurement of the effective lifetime. Alternative classifiers were developed for the effective lifetime measurement in parallel to the development of the global BDT, with a particular focus on how the cuts placed on the output of the classifiers effect the selection efficiency as a function of decay time.

The development of classifiers for the effective lifetime measurement is described in Section 4.4.4.1 and a study into whether using data or simulated decays as

the background training sample produces a more effective classifier is presented in Section 4.4.4.2. The impact of cuts placed on the output of the classifiers on the selection efficiency as a function of decay time is investigated in Section 4.4.4.3. A comparison between the performances of the classifiers developed for the effective lifetime measurement and the global BDT developed for the branching fraction measurement is made in Section 4.4.4.4 and the classifier with the best performance at separating signal and combinatorial background decays is chosen. Finally, the optimal cut value placed on the chosen classifier is determined in Section 4.4.4.5.

4.4.4.1 Development of effective lifetime multivariate classifiers

Several types of multivariate classifiers were investigated for the effective lifetime selection and BDTs gave the best performance at separating signal from background. A range of boosting methods for the decision trees were studied and the adaptive boosting method once again yielded the best results. However, a boosting method of particular interest for the effective lifetime measurement was the uBoost technique [154]. The uBoost method produces a classifier output that has a uniform efficiency for a specified variable. The most effective input variables for achieving good signal and background separation with a BDT are also correlated with the decay time. These include the B_s^0 IP, χ_{IP}^2 , χ_{FD}^2 and isolation criteria. If the output of a BDT is correlated with the B_s^0 decay time, the efficiency as a function of decay time may not have a smooth or easily understandable distribution. The uBoost method could provide a way to make modelling the efficiency as a function of decay time easier by requiring the algorithm output to have a uniform efficiency across the decay time distribution.

The input variables used in the adaptive boosting and uBoost BDTs were chosen separately, starting from a large set of variables including kinematic and geometric variables and isolation criteria. Initially the BDTs were trained using all input variables within the set and variables that had no impact on the BDT performance were removed until removing any of the remaining variables had a negative impact on the BDT performance. The performance of each BDT was evaluated from the integrated Receiver Operating Characteristic curve, which is the signal efficiency versus background rejection. The final variable sets were different for the two boosting methods; the adaptive boosting BDT uses 11 input variables and the uBoost BDT uses 21 input variables. The full list of input variables used and the definition of those variables are given in Appendix B.

Simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays were used as the signal training sample and simulated $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays were used as the background training sample to determine the input variables used in the two types of BDTs. The selection requirements listed in Table 4.12, except the BDTS requirement, were applied to training samples of

Sample	Number of decays
Simulated $B_s^0 \rightarrow \mu^+\mu^-$	668292
Simulated $b\bar{b} \rightarrow \mu^+\mu^- X$	586586
Data	189077

Table 4.16 Number of candidates present in each training sample after the selection cuts have been applied. Simulated decays and decays in data were identified as candidates that pass the selection requirements listed in Table 4.12, except the BDTS cut was not applied and the decays in data must be in the mass range 5447 - 6000 MeV/ c^2 .

simulated decays. The simulated decays were split into two samples for both signal and background, so that the BDTs could be trained on one sample and tested on the other.

The training parameters of the adaptive BDT have been optimised by iterating over a large range of different values, whereas the training parameters of the uBoost were not optimised because changing the parameters has a small impact of the overall BDT performance [154]. The values used are given in Appendix B.

4.4.4.2 Investigation of background training samples

The performance of the BDTs with different boosting methods trained on two different background training samples is compared. One sample consists of simulated $b\bar{b} \rightarrow \mu^+\mu^- X$ decays and the other combinatorial background decays in Run 1 data. At the time of the BDT development, only Run 1 data was available. The selection requirements listed in Table 4.12, except the BDTS requirement, are applied to training samples of simulated decays. Combinatorial background decays in data are identified as $B_s^0 \rightarrow \mu^+\mu^-$ candidates that pass the selection requirements listed in Table 4.12, except the BDTS requirement, and have an invariant mass of the two muon in the range 5447 - 6500 MeV/ c^2 , outside to the B_s^0 mass window. The number of events in each training sample is given in Table 4.16.

The different background samples are investigated to determine which would produce the BDT with the best performance. The final BDT is used to separate signal from combinatorial background in data therefore using data as the background training sample could lead to better performance of the BDT. However there are fewer events in the data background sample than the sample from simulation which could limit the performance. Also, a BDT trained on a combination of data and simulated decays could be sensitive to differences between data and simulation which could lead to a worse performance of the BDT.

The performance of the BDTs using either simulated decays or data as the background sample was evaluated using $B \rightarrow h^+h^-$ decays in Run 1 data. No

particle identification variables are used as input variables of the BDTs due to the mis-modelling of particle identification variables in simulated decays, therefore the performance on the BDTs on $B \rightarrow h^+h^-$ decays should be very similar to $B_s^0 \rightarrow \mu^+\mu^-$ decays. $B \rightarrow h^+h^-$ decays in data were identified by the same selection requirements used applied to the BDT training samples of simulated decays except the isMuon requirement was not applied. Also, no particle identification requirements were used to separate different $B \rightarrow h^+h^-$ decays.

The performance of the BDTs is evaluated by determining the signal significance for a range of cuts placed on the output of each BDT. The signal significance is given by

$$\mathcal{S} = \frac{S}{\sqrt{S + B}} \quad (4.3)$$

where S (B) are the number of signal (background) decays. An unbinned maximum likelihood fit is performed on the $B \rightarrow h^+h^-$ mass distribution, where all $B \rightarrow h^+h^-$ decays are reconstructed as $B_s^0 \rightarrow \mu^+\mu^-$, to find the signal and background yields for each cut value. In the mass fit, the $B \rightarrow h^+h^-$ mass distribution is modelled with a Gaussian function and the combinatorial background decays with an exponential function, an example of the mass fit is given in Figure 4.11. The number of signal and background decays used to calculate the signal significance are found as the signal and background yields within 3σ of the centre of the $B \rightarrow h^+h^-$ mass peak, where σ is the width of the Gaussian function.

The signal significances as a function of the cut value placed on the BDT output for the BDTs trained on the different background samples are shown in Figure 4.12. The outputs of the BDTs are not flattened, adaptive boosting BDT gives output values between -1 and $+1$ and uBoost BDT gives output values between 0 and $+1$. The signal significance of BDTs trained on simulated decays is higher than that of the BDTs trained on data. Therefore, from now on only BDTs trained using simulated decays as the background training sample will be considered.

Both the adaptive and uBoost BDTs shown in Figure 4.12 were trained without applying the BDTS cut to the training samples. However, the signal significance on $B \rightarrow h^+h^-$ decays has also been evaluated with the BDTS cut applied both after the BDT training and before the BDT training. The improvement in the overall performances of the BDTs is small but applying the BDTS cut to $B \rightarrow h^+h^-$ after the BDT training produces the highest signal significances.

4.4.4.3 Selection efficiency with decay time

The selection efficiency as a function of decay time has been evaluated in simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays after all selection requirements and a range of different cut values on the outputs of the adaptive and uBoost BDTs trained on simulated decays. The

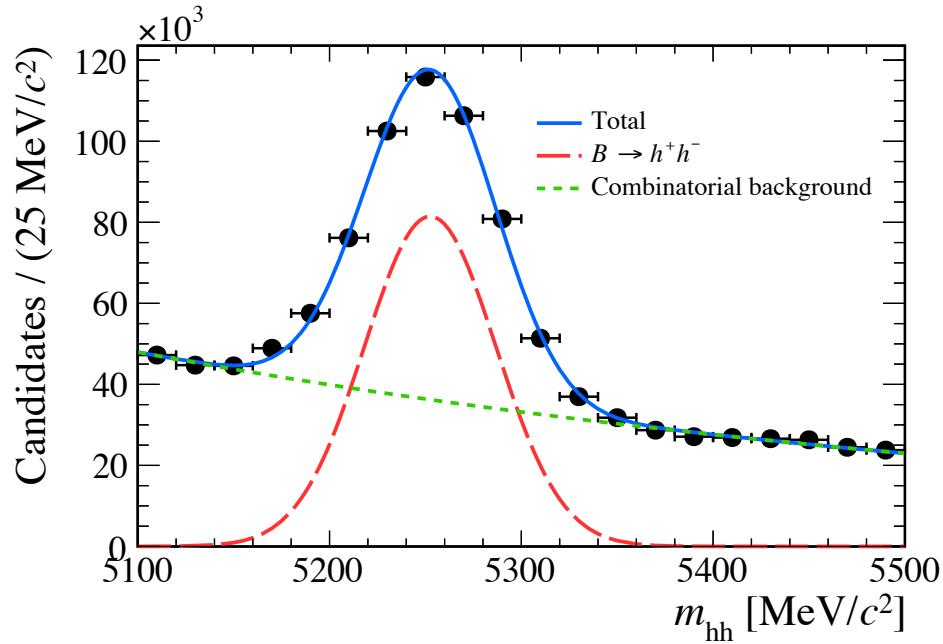


Fig. 4.11 Example of the mass fit to $B \rightarrow h^+ h^-$ Run 1 data to find the signal significance for the adaptive boosting BDT with a cut value at 0.0 on the BDT output. The BDT produces a response between -1 and $+1$.

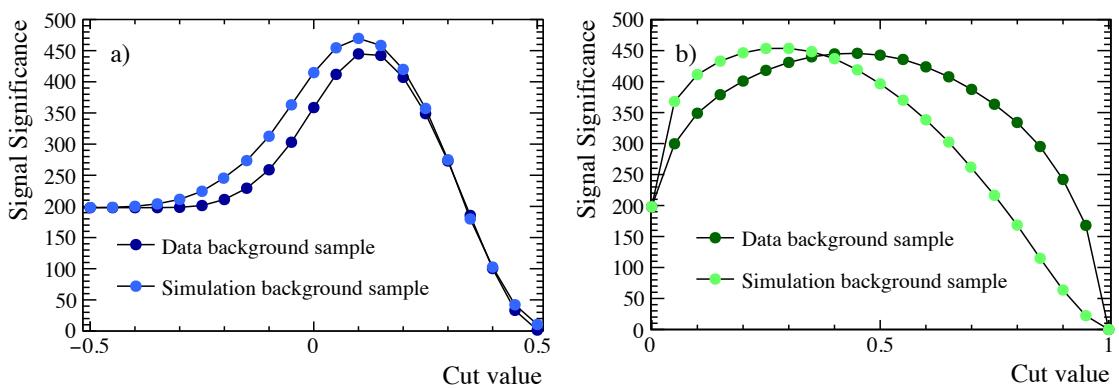


Fig. 4.12 Signal significance from $B \rightarrow h^+ h^-$ decays in Run 1 data of the a) adaptive boosting and b) uBoost BDTs trained using simulated decays and data as the background training samples.

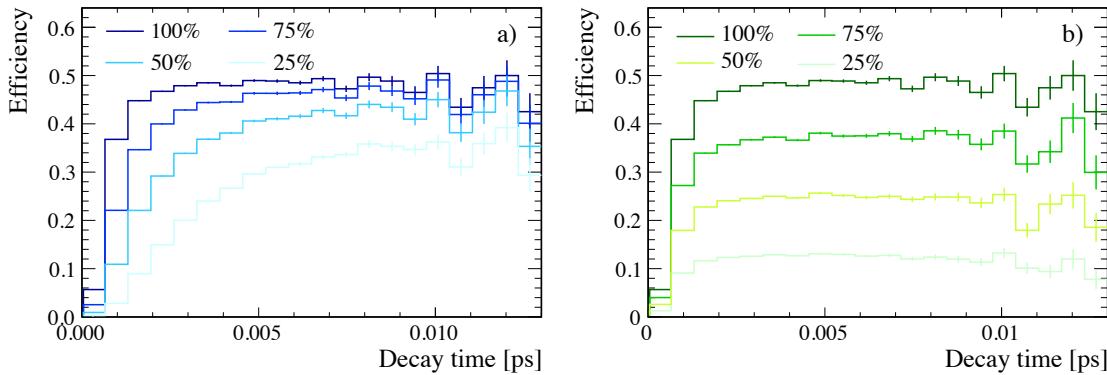


Fig. 4.13 Selection efficiency as a function of decay time of simulated 2012 $B_s^0 \rightarrow \mu^+\mu^-$ decays for the a) adaptive and b) uBoost BDTs. The selection requirements applied to the training sample are applied to the simulated decays and cuts are placed on the BDT output so that the efficiency of the cut on decays passing the other selection requirements is 100, 75, 50 and 25%.

cut values are chosen to have the same selection efficiencies for each algorithm. The efficiencies are shown in Figure 4.13. The selection efficiency as a function of decay time is uniform for the uBoost BDT whereas the adaptive boosting BDT removes a greater proportion of decays with short decay times than the uBoost method.

Ideally, to reduce systematic uncertainties on the measurement of the effective lifetime described in Chapter 7, the selection would not bias the decay time distribution. However, with the expected statistics of the data set an unbiased selection would not be appropriate. Both algorithms have a smooth efficiency as a function of decay time therefore, with either algorithm the efficiency as function of decay time can be well modelled.

4.4.4.4 Classifier performance comparison

The final classifier used to select $B_s^0 \rightarrow \mu^+\mu^-$ candidates is the BDT that has the greatest separation power between signal and combinatorial background decays and consequently removing the most combinatorial background decays for a given signal efficiency, provided the selection efficiency as a function of decay time can be accurately modelled. The performances of the BDTs developed for the effective lifetime measurement are compared to that of the global BDT used to classify candidates for the branching fraction measurements. Two different approaches are used to evaluate the performances: the signal significance of $B \rightarrow h^+h^-$ decays as a function of BDT cut values; and the rejection of $B_s^0 \rightarrow \mu^+\mu^-$ backgrounds in data. In order to enable easy comparison of the different BDTs, the outputs of the BDTs developed for the effective lifetime measurement have been flattened in the same way as the global BDT, described in Section 4.3.4.

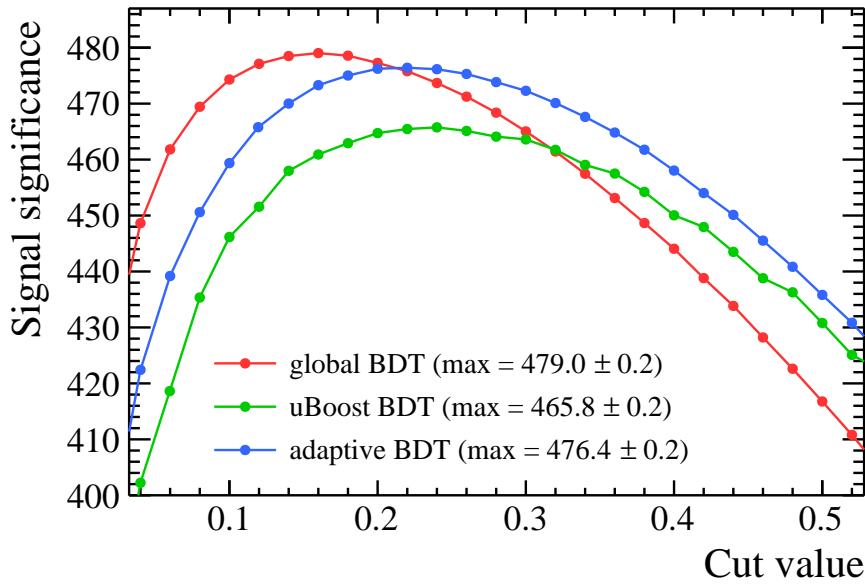


Fig. 4.14 Signal significance from $B \rightarrow h^+h'^-$ decays in Run 1 data of the adaptive and uBoost BDTs trained using simulated decays as the background sample and the signal significance of the global BDT developed for the branching fraction measurement. The selection requirements listed in Table 4.12 are used, apart from the isMuon requirement.

The signal significance for each BDT is evaluated on $B \rightarrow h^+h'^-$ decays in Run 1 data and the maximum signal significance is found in the same way as described earlier for comparing BDTs trained with data or simulated decays as the background sample. The BDTS cut is applied in the selection process because the global BDT was designed to be used with the BDTS and the performance of the BDTs developed for the effective lifetime is best when the BDTS requirement is used. The results are shown in Figure 4.14; the global BDT produces the highest signal significance, but is closely followed by the adaptive boosting BDT developed for the effective lifetime measurement.

The purpose of the BDT is to remove combinatorial background decays passing the $B_s^0 \rightarrow \mu^+\mu^-$ selection, therefore an additional comparison of the different algorithms is made. The number of combinatorial background decays present in Run 1 data passing the effective lifetime selection criteria but in the mass range 5447 - 6550 MeV/ c^2 are found for a range of cuts on the output of the BDTs. The same cut values are applied to each BDT and, since all the BDTs are flattened to have a uniform distribution of signal decays between 0 and 1, the cut values will have very similar signal efficiencies for each BDT. The results are given in Table 4.17. The global BDT is most effective at removing background decays for a given signal efficiency. The same comparisons were made with the BDT used in the previous analysis [46, 47] and all BDTs described in this dissertation have a better performance at removing background decays.

BDT	Number events above BDT output value								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Global BDT	2261	597	229	89	34	13	4	1	0
Adaptive BDT	4623	1395	513	215	77	32	15	4	2
uBoost BDT	7904	3344	1535	630	268	92	27	7	0

Table 4.17 Number of candidates in Run 1 data passing the effective lifetime selection and the BDTS cut in the mass range $5447 - 6000 \text{ MeV}/c^2$. The output of each BDT is flattened to have a uniform response between 0 and 1, therefore the cuts applied to each BDT will have approximately the same efficiency.

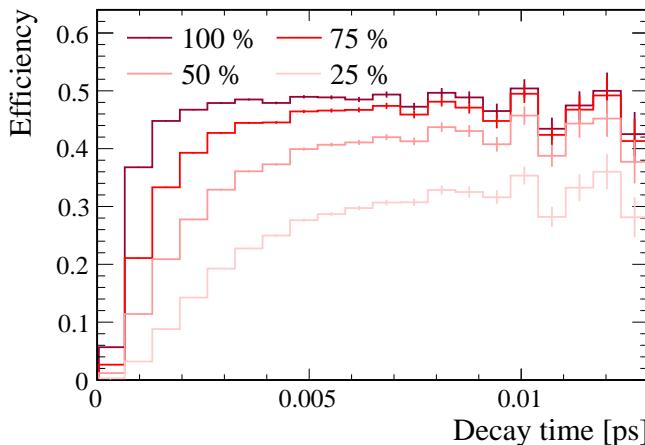


Fig. 4.15 Selection efficiency as a function of decay time of simulated 2012 $B_s^0 \rightarrow \mu^+ \mu^-$ decays for the global BDT. The selection requirements applied to the training sample are applied to the simulated decays and cuts are places on the BDT output so that the efficiency of the cut on already selected event is 100, 75, 50 and 25%.

Although the global BDT, combined with the BDTS, performs best at separating signal from background decays, the efficiency as a function of decay time must also be evaluated for this algorithm to ensure it does not exhibit any strange behaviour which would make modelling the decay time efficiency challenging. The decay time efficiency is shown in Figure 4.15 for several cut values on the BDT output and gives a smooth distribution as a function of decay time. For the data set used to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime, the expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ decays is very low. Therefore the benefits of using the uBoost method are outweighed by its poor performance. Since the global BDT developed for the branching fraction measurements has the best performance in both tests and a smooth decay time efficiency it is the best BDT to use for the selection of events for effective lifetime measurement.

4.4.4.5 Optimisation of BDT cut choice

A cut is placed on the output of the global BDT to select $B_s^0 \rightarrow \mu^+\mu^-$ decays. The cut value is optimised to give the smallest expected uncertainty on the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$. The optimisation is done using pseudoexperiments generating the expected number of $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays in Run 1 and Run 2 data for different cuts on the global BDT output.

The fit procedure to extract $\tau_{\mu\mu}$ from the data is described in Chapter 6. The pseudoexperiments used to optimise the global BDT cut value are performed following the steps:

- the mass and decay time distributions for number of expected $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background events are generated using the expected mass and decay time probability density functions (PDFs);
- an unbinned maximum likelihood fit is performed to the invariant mass distribution of the two muons, where the $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background yields are free to float in the fit along with the slope, λ , of the combinatorial background PDF; and
- the mass fit is used to compute sWeights using the sPlot method [155] and a maximum likelihood fit is performed to the sWeighted decay time distribution to extract $\tau_{\mu\mu}$.

The number of expected $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays for different BDT cut values is derived from the expected number of candidates that pass the effective lifetime selection cuts and a cut on the global BDT of $\text{BDT} > 0.55$ ³. These predictions are given in Table 4.18 and assume the SM branching fraction for $B_s^0 \rightarrow \mu^+\mu^-$ decays. The methods used to evaluate the expected number of each decay are detailed in Chapter 5.

The expected number of $B_s^0 \rightarrow \mu^+\mu^-$ decays after different BDT cut values is straight forward to compute from the information in Table 4.18. This is because the flattening procedure applied to the global BDT output means that the number of $B_s^0 \rightarrow \mu^+\mu^-$ decays is evenly distributed across the BDT range. The number of combinatorial background decays expected after each BDT cut is determined from the number of decays in Table 4.18 and using information from simulated $b\bar{b} \rightarrow \mu^+\mu^- X$ decays that have had all effective lifetime selection requirements applied up until the

³Initially the observed yields from published the Run 1 branching fraction measurements were used to determine the expected number of decays present in 4.4 fb^{-1} of Run 1 and Run 2 data and a global BDT cut of 0.55 was found to be optimal. However, the expected number of decays was then re-evaluated using the more sophisticated techniques described in Chapter 5 and using global BDT cut of 0.55 and the pseudoexperiments were repeated to check the optimal BDT cut was the same.

Decay	Expected number of candidates
$B_s^0 \rightarrow \mu^+ \mu^-$	30.5
Combinatorial background	40.6
Total	71.1

Table 4.18 Expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background candidates after the $B_s^0 \rightarrow \mu^+ \mu^-$ selection requirement and with a global BDT value greater than 0.55 in the mass range $5320 < m_{\mu^+ \mu^-} < 6000 \text{ MeV}/c^2$.

cut on the global BDT. A ratio is evaluated from $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays, given by

$$R(X) = \frac{\epsilon(BDT > X)}{\epsilon(BDT > 0.55)}, \quad (4.4)$$

where $\epsilon(BDT > X)$ is the fraction of simulated $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays that have a global BDT value greater than X and $\epsilon(BDT > 0.55)$ is the fraction of simulated $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays that have a global BDT value greater than 0.55. The expected number of combinatorial background decays after a BDT cut of X is then evaluated from multiplying $R(X)$ by the number of decays in Table 4.18. The ratios for the different cuts values are shown in Table 4.19 along with the expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays for each BDT cut. Simulated decays had to be used to compute the efficiencies rather than data because there are too few candidates left after tight BDT cuts in data to enable meaningful studies.

The mass distribution of the combinatorial background is described by an exponential function, it was observed from the simulated $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays that the slope of the mass distribution changed with the BDT cut value as illustrated in Figure 4.16. The change in the slope value is accounted for in the mass distribution used in the pseudoexperiment by changing the slope parameter (λ) for each BDT cut. Table 4.19 shows the slope of the mass distribution for different BDT cuts values evaluated from $b\bar{b} \rightarrow \mu^+ \mu^- X$ simulated decays.

The results from 10,000 pseudoexperiments for BDT cut values every 0.05 in the range 0.4 - 0.65 are shown in Table 4.20 with the median uncertainty of the fits for $\tau_{\mu\mu}$ and the signal significance for each BDT cut. The median uncertainties are used rather than the mean because the distribution of uncertainties is asymmetric. The highest signal significance and lowest expected uncertainties occur for a BDT cut of 0.55. Therefore this cut value is used to select $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the same cut is applied to the global BDT to select $B \rightarrow h^+ h'^-$ decays.

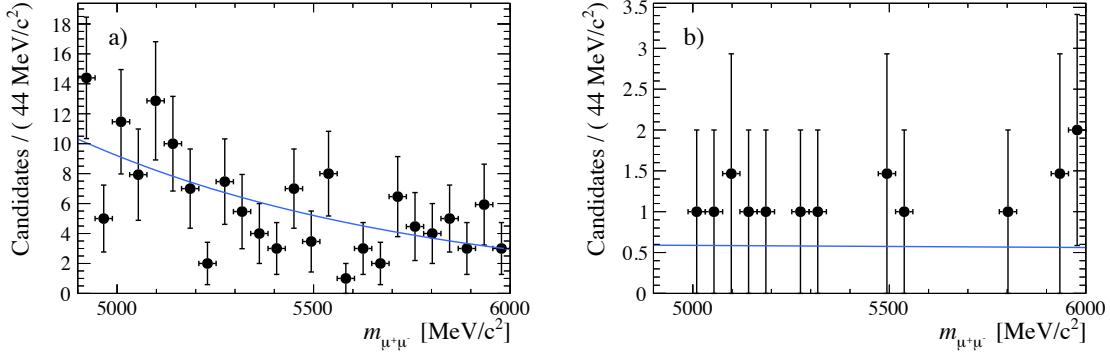


Fig. 4.16 Mass distribution of simulated decays that have passes the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime selection requirements and after global BDT cuts of a) 0.4 and b) 0.55.

BDT cut	$\mathcal{N}(B_s^0 \rightarrow \mu^+\mu^-)$	$R(X)$	$\mathcal{N}(\text{Comb.})$	$\lambda / c^2 \text{MeV}^{-1}$
0.40	40.5	8.69	269.1	-0.00114 \pm 0.00028
0.45	37.2	3.91	116.2	-0.00129 \pm 0.00041
0.50	33.8	1.91	56.3	-0.00132 \pm 0.00060
0.55	30.5	1.00	40.6	-0.00004 \pm 0.00089
0.60	27.1	0.55	22.5	-0.00000 \pm 0.00114
0.65	23.8	0.32	12.4	-0.00024 \pm 0.00122

Table 4.19 Inputs to the pseudoexperiments used to determine the optimum global BDT cut value for the effective lifetime analysis. For each BDT cut value the number of expected $B_s^0 \rightarrow \mu^+\mu^-$ decays ($\mathcal{N}(B_s^0 \rightarrow \mu^+\mu^-)$), the ratio ($R(X)$) used to find the number of combinatorial background decays, the expected number of combinatorial background decays ($\mathcal{N}(\text{Combinatorial})$) and the slope of the combinatorial background mass distribution (λ) are given.

Global BDT cut	$\frac{S}{\sqrt{S+B}}$	$\sigma(\tau_{\mu\mu}) / \text{ps}$
0.40	3.87	0.345
0.45	4.51	0.309
0.50	4.85	0.291
0.55	4.94	0.285
0.60	4.86	0.297
0.65	4.65	0.309

Table 4.20 The signal significance for each cut value in the global BDT and median of the expected uncertainties for $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$ from 10,000 pseudoexperiments for the expected number of events. The expected number of $B_s^0 \rightarrow \mu^+\mu^-$ ($\mathcal{N}(B_s^0 \rightarrow \mu^+\mu^-)$) and combinatorial background ($\mathcal{N}(\text{Comb.})$) decays that are generated in the pseudoexperiments are also listed for each BDT cut.

4.4.5 Summary

The complete set of selection criteria used for identify $B_s^0 \rightarrow \mu^+\mu^-$ decays in Run 1 and Run 2 data for the effective lifetime measurement are listed in Table 4.21. The selection requirements do not remove all backgrounds decays from the data set but reduce them to a level at which the effective lifetime can be measured. The selection criteria for $B \rightarrow h^+h'^-$ decays used to verify the measurement strategy are very similar to the selection used to identify $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays the differences are in the mass range used and the trigger and particle identification requirements as discussed in Sections 4.4.2, 4.4.1 and 4.4.3. The mass and decay time distributions for $B_s^0 \rightarrow \mu^+\mu^-$ candidates passing the selection criteria in 4.4 fb^{-1} of Run 1 and Run 2 data are shown in Figure 4.17.

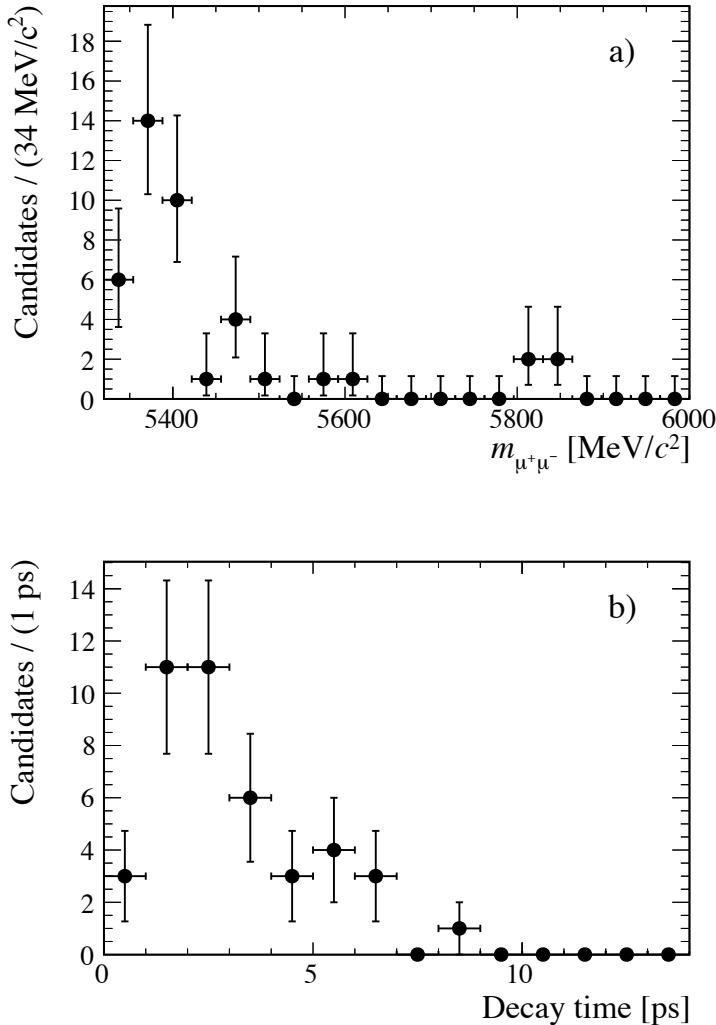


Fig. 4.17 Dimuon invariant mass (a) and decay time (b) distributions for $B_s^0 \rightarrow \mu^+\mu^-$ candidates in 4.4 fb^{-1} of Run 1 and Run 2 data passing the selection requirements in Table 4.21.

Particle	$B_s^0 \rightarrow \mu^+\mu^-$
B_s^0	$5320 \text{ MeV}/c^2 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$ $\text{DIRA} > 0$ $\chi_{\text{FD}}^2 > 225$ $\chi_{\text{IP}}^2 < 25$ $\chi_{\text{VTX}}^2 < 9$ $\text{DOCA} < 0.3 \text{ mm}$ $\tau < 13.248 \text{ ps}$ $p_T > 500 \text{ MeV}/c$ $\text{BDTS} > 0.05$ $\text{PID}_{\mu}^{\text{Run1+2015}} > 0.2 \text{ or } \text{PID}_{\mu}^{\text{2016}} > 0.4$ $ m_{\mu\mu} - m_{J/\psi} < 30 \text{ MeV}/c^2$ Global BDT > 0.55
μ	$\chi_{\text{trk}}^2 < 3 \text{ (4)}$ Minimum $\chi_{\text{IP}}^2 > 25$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ $p < 500 \text{ GeV}/c$ ghost probability < 0.3 (0.4) isMuon = True
Trigger requirements	L0Global = TIS or TOS Hlt1Phys = TIS or TOS Hlt2Phys = TIS or TOS

Table 4.21 Selection cuts applied to select $B_s^0 \rightarrow \mu^+\mu^-$ for the effective lifetime measurement, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

Chapter 5

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

This chapter presents the measurements of the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions, focusing in more detail on the parts of the analysis that are also used for the measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime in Chapter 6. Section 5.1 gives an overview of the analysis strategy and a description of how the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yield is extracted from the data is given in Section 5.2. The estimation of the background decays that must be understood for the branching fraction measurements are detailed in Section 5.3 and the normalisation procedure to convert the number of observed $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays into the branching fractions is explained in Section 5.4. Finally, the results are presented in Section 5.5. The $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime measurement uses the mass distributions of signal and background decays and the expected yields described in Sections 5.2 and 5.3.

5.1 Analysis strategy

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, are defined as the fraction of $B_{(s)}^0$ mesons which decay into two muons. In reality, not every $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay produced in pp collisions will be within the LHCb detector acceptance or be reconstructed and pass the selection criteria of Chapter 4. Therefore, the number of observed $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays at LHCb is reduced by the efficiency, ϵ , of the detector, trigger, reconstruction and selection criteria. The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are measured as

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{\mathcal{N}_{B_{(s)}^0 \rightarrow \mu\mu}}{\mathcal{N}_{B_{(s)}^0}} = \frac{\mathcal{N}_{B_{(s)}^0 \rightarrow \mu\mu}^{obs}}{\epsilon \mathcal{N}_{B_{(s)}^0}} \quad (5.1)$$

where $\mathcal{N}_{B_{(s)}^0 \rightarrow \mu\mu}$ is the total number of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays that occur, $\mathcal{N}_{B_{(s)}^0}$ is the total number of $B_{(s)}^0$ mesons and $\mathcal{N}_{B_{(s)}^0 \rightarrow \mu\mu}^{obs}$ is the number of observed $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays.

The number of $B_{(s)}^0$ mesons produced can be calculated from the integrated luminosity, \mathcal{L}_{int} , and the $b\bar{b}$ production cross-section, $\sigma_{b\bar{b}}$, via

$$\mathcal{N}_{B_{(s)}^0} = 2 \times \mathcal{L}_{int} \times \sigma_{b\bar{b}} \times f_{d(s)}, \quad (5.2)$$

where $f_{d(s)}$ is the hadronisation factor, giving the probability for a b or \bar{b} quark to form a B^0 (B_s^0) or a \bar{B}^0 (\bar{B}_s^0) meson. The factor of 2 arises because no distinction is made between the $B_{(s)}^0$ and the $\bar{B}_{(s)}^0$. Although the number of $B_{(s)}^0$ mesons can be computed in this way the measured cross-section is not precisely known and neither are the hadronisation factors. Therefore, in order to achieve more precise branching fraction measurements, an alternative approach is used. Another decay with a well known branching fraction is used to normalise the observed number of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays and obtain the branching fractions. The normalisation channel can be chosen in such a way to allow several uncertainties associated with the measurement to cancel in the normalisation process. The extraction of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$ from the number of observed decays is therefore

$$\begin{aligned} \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-) &= \mathcal{B}_{norm} \cdot \frac{f_{norm}}{f_{d(s)}} \cdot \frac{\epsilon_{norm}}{\epsilon_{B_{(s)}^0 \rightarrow \mu\mu}} \cdot \frac{\mathcal{N}_{B_{(s)}^0 \rightarrow \mu\mu}^{obs}}{\mathcal{N}_{norm}^{obs}} \\ &= \alpha_{d(s)} \cdot \mathcal{N}_{obs B_{(s)}^0 \rightarrow \mu\mu}, \end{aligned} \quad (5.3)$$

where *norm* indicates the normalisation channel. The normalisation factors can be combined into one normalisation parameter $\alpha_{d(s)}$ for each of the B_s^0 and B^0 decays. The normalisation procedure removes the uncertainty from $\sigma_{b\bar{b}}$, the systematic uncertainties in the efficiencies cancel out in the ratio, as well as uncertainties on $f_{d(s)}$ depending on the choice of the normalisation channel. Therefore, to measure the branching fractions in this way the number of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in data and the normalisation parameters $\alpha_{d(s)}$ must be measured.

The number of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in data is found by performing a simultaneous unbinned extended maximum likelihood fit [126, 156] to the dimuon invariant mass distribution of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates in data in four BDT bins. For this measurement data from Run 1 and Run 2 are kept separate in the fit, therefore the simultaneous fit is performed over 8 categories; four BDT bins for each run. The BDT bin boundaries used in the fit are

$$[0.25, 0.4, 0.5, 0.6, 1.0]. \quad (5.4)$$

Pseudoexperiments based on the expected number of signal and background decays were performed to determine the BDT bin configuration that gave the best fit stability and sensitivity to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions. The bins were optimised for the $B^0 \rightarrow \mu^+ \mu^-$ sensitivity because this decay is yet to be observed. Candidates with BDT values between 0 and 0.25 are not included in the fit because this bin is dominated by backgrounds from random combinations of muons in the event. The inclusion of this bin does not improve the branching fraction sensitivity and reduces the stability of the fit.

The candidates in data that pass the selection criteria described in Chapter 4 consist of background as well as signal decays. Therefore in order to find the number of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays from the simultaneous fit information is needed about both the signal and background. For the fit, the mass PDFs of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays must be known as well as the BDT PDF, which describes the fraction of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays expected in each BDT bin. The evaluation of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ mass and BDT PDFs in Run 1 and Run 2 data are described in Section 5.2. The backgrounds present in the data must be modelled in the mass fit, therefore the mass PDFs and expected yields of the backgrounds in each BDT bin must be evaluated for Run 1 and Run 2 data. The backgrounds included in the fit are described in Section 5.3 along with the mass PDFs and expected yields for each background.

The normalisation parameters are evaluated for two different normalisation decays and are then combined to measure the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions. A normalisation decay is chosen to be as similar as possible to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, in order to reduce systematic uncertainties introduced by different detection and selection efficiencies between the signal and normalisation channels. Furthermore, the chosen decay needs to be abundant and have a precisely measured branching fraction so that the precision of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements are not limited by the uncertainties of the normalisation channel. Two decays are chosen as normalisation channels: $B^+ \rightarrow J/\psi K^+$, where $J/\psi \rightarrow \mu^+ \mu^-$; and $B^0 \rightarrow K^+ \pi^-$. Both decays have large, precisely measured branching fractions and are similar to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in complementary ways. The $B^+ \rightarrow J/\psi K^+$ decay has a very similar trigger efficiency due to the two muons from the J/ψ , although the extra particle in the final state leads to different selection and reconstruction efficiencies. The $B^0 \rightarrow K^+ \pi^-$ decay has a very similar topology to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, therefore the selection and reconstruction efficiencies will be similar, but the trigger efficiency for hadrons is quite different to muons.

The normalisation factors $\alpha_{d(s)}$ for $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays are evaluated independently for each normalisation channel and year of data taking, the factors are combined to produce an overall normalisation factor for Run 1 and Run 2. The evaluation of the normalisation factors is described in Section 5.4. As mentioned

in Chapter 4, the analysis is performed as a ‘blind’ analysis and the mass regions ± 80 MeV/c 2 either side of the B_s^0 and B^0 mass peaks are not revealed until each step in the analysis procedure has been finalised.

5.2 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass and BDT PDFs

5.2.1 Mass PDFs

The mass PDFs for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays are modelled by a Crystal Ball function [157]. A Crystal Ball function is a Gaussian function that has a power-law tail on the low mass side to model radiative energy loss in the final state. The parameters defining the function are: the mean, μ ; and resolution, σ , of the Gaussian; the slope of the exponential, n ; and a parameter α , defined in terms of σ , that determines the transition point between the Gaussian and the exponential function.

The signal shape parameters are evaluated in the following ways:

- μ - the means of B^0 and B_s^0 decays are evaluated separately from fits to $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays;
- σ - the resolution is interpolated from the resolutions of quarkonia resonances. The resolutions for the J/ψ , $\Psi(2S)$ and $\Upsilon(1, 2, 3S)$ decaying into two muons are measured from fits to data. The B^0 and B_s^0 resolutions are interpolated from the power-law relationship between quarkonia mass and resolution and using the mean B^0 and B_s^0 values from $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays, respectively; and
- n and α - these parameters are evaluated from the mass spectrum of $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ simulated decays where the mass distributions are smeared to have the same resolution as that measured from the quarkonia decays in data.

All parameters are evaluated separately for the B^0 and B_s^0 for the Run 1 and Run 2 data sets. The resulting parameter values are given in Tables 5.1 and 5.2 and the mass fits are shown in Figures 5.1, 5.2 and 5.3. The systematic uncertainties on the means come from varying the particle identification cuts used to separate the different $B \rightarrow h^+h^-$ decays and the systematic uncertainties on the σ come from the mass windows chosen for the quarkonia mass fits and the chosen mass fit model. The same mass PDFs are used for all BDT bins because $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays are distributed uniformly across the BDT output range.

Parameter	$B^0 \rightarrow \mu^+\mu^-$	$B_s^0 \rightarrow \mu^+\mu^-$
μ (MeV/c ²)	$5284.73 \pm 0.15_{stat} \pm 0.27_{syst}$	$5372.05 \pm 0.16_{stat} \pm 0.36_{syst}$
σ (MeV/c ²)	$22.68 \pm 0.05_{stat} \pm 0.39_{syst}$	$23.07 \pm 0.05_{stat} \pm 0.39_{syst}$
n	1.141 ± 0.026	1.156 ± 0.013
α	2.054 ± 0.013	2.053 ± 0.007

Table 5.1 Parameter values for Crystal Ball functions used to describe the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass PDFs in Run 1 data.

Parameter	$B^0 \rightarrow \mu^+\mu^-$	$B_s^0 \rightarrow \mu^+\mu^-$
μ (MeV/c ²)	$5279.95 \pm 0.13_{stat} \pm 0.08_{syst}$	$5367.34 \pm 0.14_{stat} \pm 0.35_{syst}$
σ (MeV/c ²)	$22.46 \pm 0.08_{stat} \pm 0.41_{syst}$	$22.85 \pm 0.08_{stat} \pm 0.42_{syst}$
n	1.118 ± 0.014	1.110 ± 0.017
α	2.063 ± 0.007	2.062 ± 0.008

Table 5.2 Parameter values for Crystal Ball functions used to describe the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass PDFs in Run 2 data.

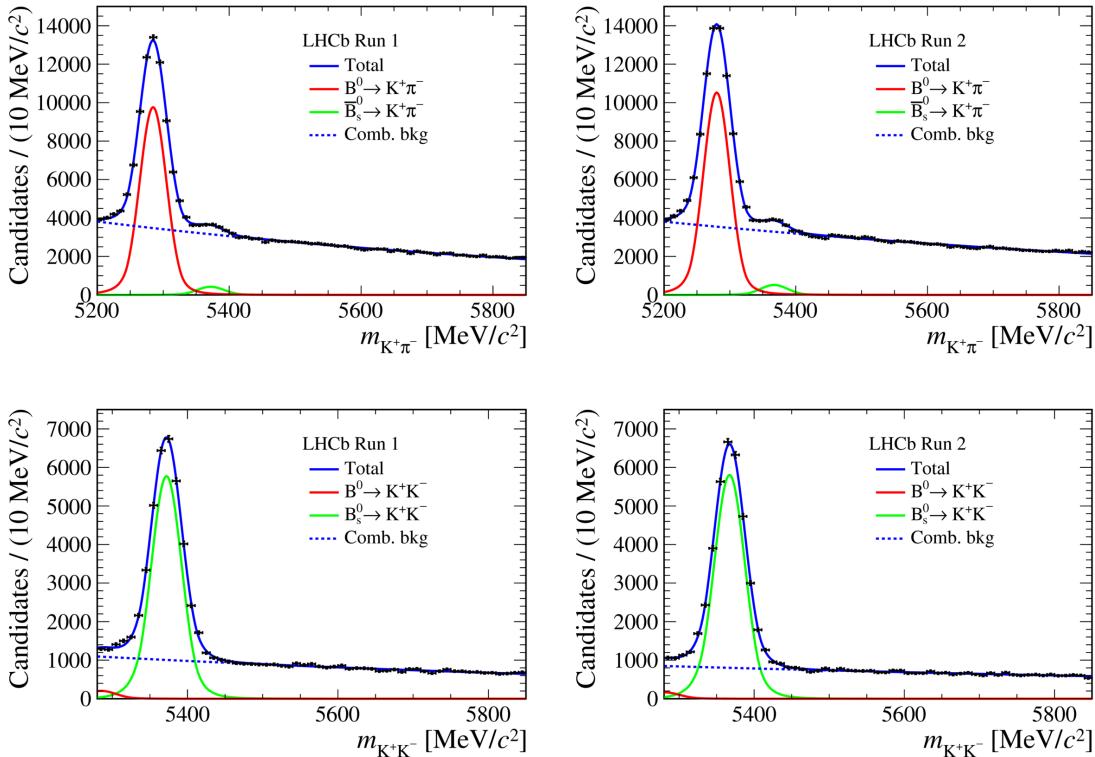


Fig. 5.1 Maximum likelihood fits to $B^0 \rightarrow K^+\pi^-$ (top) and $B_s^0 \rightarrow K^+K^-$ (bottom) for Run 1 and Run 2 data to measure B^0 and B_s^0 masses.

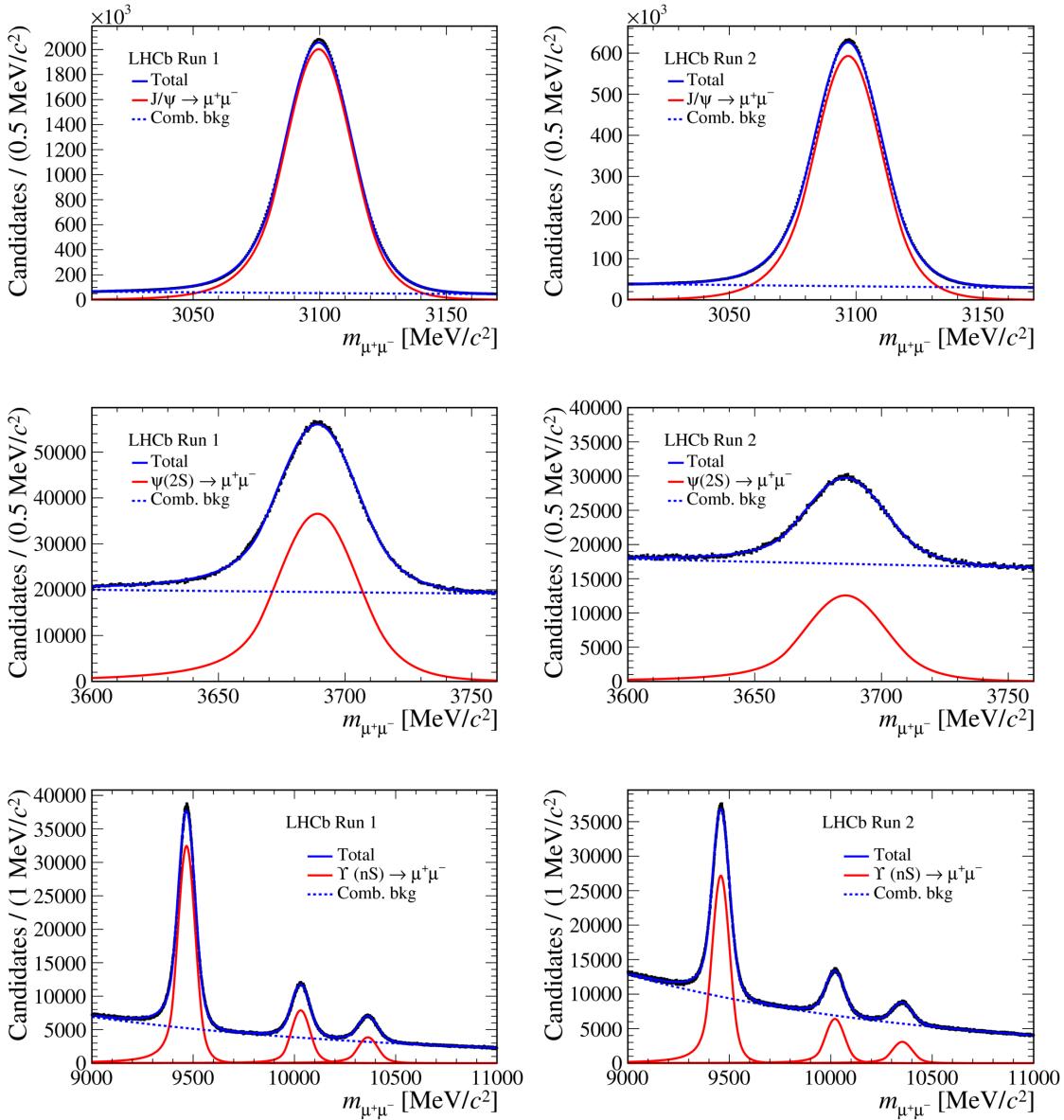


Fig. 5.2 Maximum likelihood fit to the mass spectrum of J/ψ (top), $\Psi(2S)$ (centre) and $\Upsilon(1, 2, 3S)$ (bottom) decaying into two muons in Run 1 and Run 2 data.

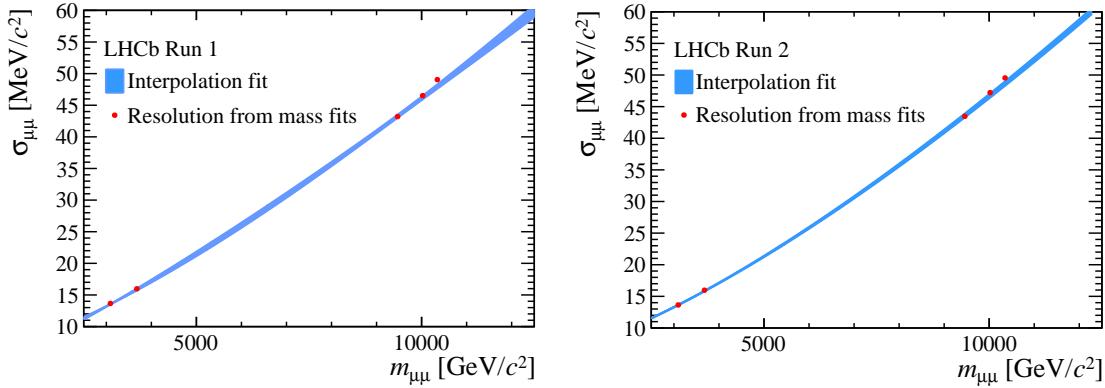


Fig. 5.3 Power law fit to the resolution of quarkonia resonances to determine mass resolution for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays on Run 1 and Run 2 data.

5.2.2 BDT PDFs

The global BDT distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays is expected to be uniform between 0 and 1 as designed by the flattening procedure described in Section 4.3.4.2. The fraction of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in a BDT bin should be proportional to the bin width. However, the global BDT was trained and flattened using simulated decays, therefore to avoid differences between simulated decays and data affecting the expected fraction of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in each BDT bin, the BDT PDF is evaluated from $B^0 \rightarrow K^+\pi^-$ decays in data. This process is known as the BDT calibration. The global BDT is designed to use only kinematic and geometric information to classify candidates and includes no PID information. Therefore the BDT distributions of $B^0 \rightarrow K^+\pi^-$ decays will be the same to a good approximation as $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays since they are kinematically identical.

The number of $B^0 \rightarrow K^+\pi^-$ decays is extracted from data by fitting the mass distribution of $B^0 \rightarrow K^+\pi^-$ candidates in each BDT bin for Run 1 and Run 2 data. The $B^0 \rightarrow K^+\pi^-$ candidates must pass the standard $B \rightarrow h^+h^-$ selection outlined in Section 4.3.1 and are separated from other $B \rightarrow h^+h^-$ modes using the $\text{DLL}_{K\pi}$ variable.

The particle identification and trigger efficiencies are different for $B^0 \rightarrow K^+\pi^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays. Therefore, the $B^0 \rightarrow K^+\pi^-$ yields in each BDT bin are corrected for the different efficiencies. The same calibration is used for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays. The calibration is performed for each year separately then combined to give the Run 1 and Run 2 fractions per BDT bin. Figure 5.4 shows the BDT distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays calibrated with $B^0 \rightarrow K^+\pi^-$ data for Run 1 and Run 2. The systematic uncertainties on the BDT calibration arise from the mass range used in the fit, the choice of the fit model and particle identification requirements and the trigger and particle identification efficiency corrections.

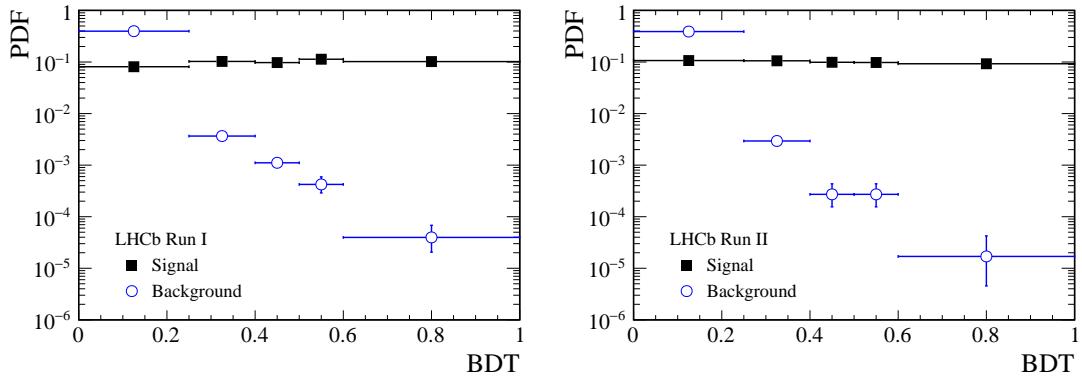


Fig. 5.4 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT PDFs (black squares) for Run 1 and Run 2 data calibrated using $B^0 \rightarrow K^+\pi^-$ decays and the combinatorial background decays (blue circles) for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates in data with a dimuon mass above 5477 MeV/ c^2 . The uncertainties on the signal fractions are included on the plots but are too small to be visible.

Although the BDT is calibrated, the dependence of the BDT response on the $B_{(s)}^0$ candidate decay time must also be considered. The response of the global BDT for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays is correlated with their decay time due to the use of the B_s^0 IP and χ_{IP}^2 and isolation criteria as inputs to the BDT. This correlation will lead to slightly incorrect estimations of the $B_s^0 \rightarrow \mu^+\mu^-$ BDT PDF. In the SM the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$, is equal to the lifetime of the heavy B_s^0 mass eigenstate, τ_H . However in reality $\tau_{\mu\mu}$ could be somewhere in between the lifetimes of the heavy and light mass eigenstates. As described in Chapter 2 the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is related to the parameter $A_{\Delta\Gamma}$, where $A_{\Delta\Gamma} = +1$ for $\tau_{\mu\mu} = \tau_H$ and $A_{\Delta\Gamma} = -1$ for $\tau_{\mu\mu} = \tau_L$, where τ_L is the lifetime of the light $B_s^0 \rightarrow \mu^+\mu^-$ mass eigenstate.

The simulated decays used to train and flatten the global BDT use as the $B_s^0 \rightarrow \mu^+\mu^-$ lifetime the mean of the measured τ_H and τ_L values at the time of the simulation production [5]. Therefore, the lifetime used is different between simulation versions. Since the BDT output is correlated with the lifetime, the fraction of $B_s^0 \rightarrow \mu^+\mu^-$ decays in each BDT bin will depend on the lifetime used in the simulation. Numerical correction factors are computed for each year to scale the fraction of $B_s^0 \rightarrow \mu^+\mu^-$ decays in each BDT bin for the situations where $A_{\Delta\Gamma} = -1$, 0 or +1, so that the dependence on $A_{\Delta\Gamma}$ of the measured branching fractions can be evaluated.

No corrections are needed for $B^0 \rightarrow \mu^+\mu^-$ because the difference in lifetime of the heavy and light B^0 mass eigenstates is negligible and the need for correction cancels out with the BDT calibration that uses the B^0 decay $B^0 \rightarrow K^+\pi^-$.

5.3 Background mass PDFs and expected yields

The selection described in Chapter 4 is effective at reducing the backgrounds in the data set to a suitable level so that number of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays can be measured. However, some background decays remain in the final data set that cannot be completely removed without drastically reducing the signal efficiency. The backgrounds present in the final data set originate from:

- $B \rightarrow h^+h'^-$ decays when both hadrons are mis-identified as muons, commonly caused by hadrons decaying semi-leptonically during their flight through the detector after leaving the VELO. This background falls within the B^0 mass window but not the B_s^0 mass window due to the missing energy from the undetected neutrino;
- semi-leptonic decays where one hadron is mis-identified as a muon that include
 - $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ decays where the final state hadrons are mis-identified as muons. The mass of these backgrounds falls below the B^0 mass window in the left mass sideband; and
 - $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$ decays when the proton is mis-identified as a muon. The large mass of the Λ_b^0 means that this background pollutes the B_s^0 and B^0 mass windows and below;
- decays which contain two muons that form a good vertex that include
 - $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ decays where the pion is not detected. The missing hadron means that these backgrounds fall well below the B^0 mass window; and
 - $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ decays where $J/\psi \rightarrow \mu^+\mu^-$. The large mass of the B_c^+ causes this background to cover the full mass range 4900 - 6000 MeV/ c^2 ; and
- combinatorial background formed by the random combination of any two muons in the event, this background is distributed across the full mass range.

The backgrounds present in the data set must be included in the fit to the dimuon invariant mass in order to accurately measure the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions. Therefore, the mass PDFs and expected yields of each background must be evaluated. The following sections summarise the information used for each background source in the branching fraction fit and Figure 5.5 shows the mass distributions of the background sources for the expected number of candidates with BDT values of $BDT > 0.5$.

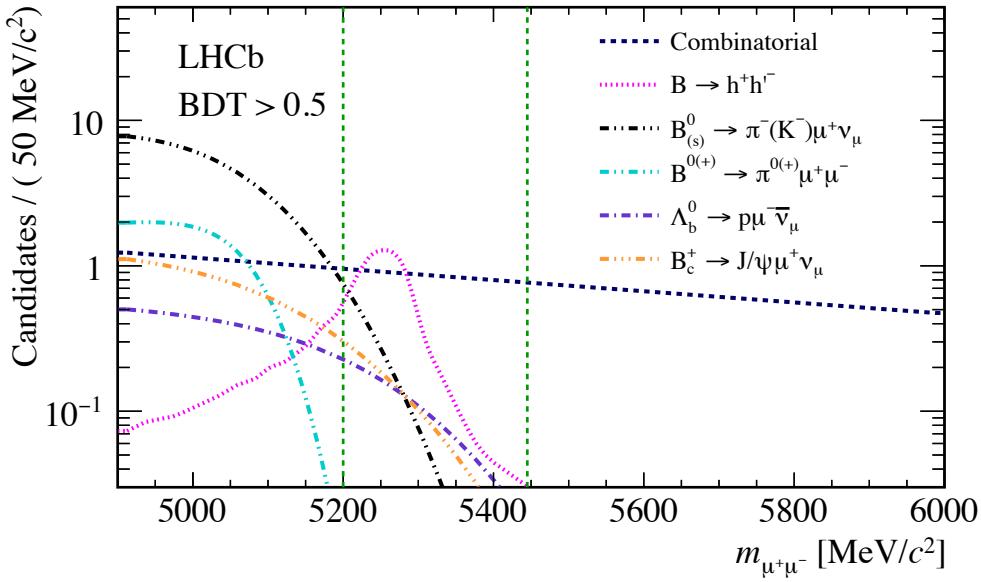


Fig. 5.5 Mass distributions for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ backgrounds with global BDT values of $\text{BDT} > 0.5$. The backgrounds shown are from $B \rightarrow h^+h'^-$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ and combinatorial background. The green dashed lines show the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass window.

5.3.1 Mis-identified $B \rightarrow h^+h'^-$ decays

The mass PDF describing mis-identified $B \rightarrow h^+h'^-$ decays is formed of two Crystal Ball functions. The two functions have the same mean but all other parameters can be different and the power-law tails for each function are on opposite sides of the mean value. This combination of functions is called a double Crystal Ball function. The parameter values are evaluated from simulated $B^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+K^-$, $B^0 \rightarrow \pi^+\pi^-$ and $B_s^0 \rightarrow K^+\pi^-$ decays in which the momenta of tracks are smeared to model the hadrons decaying in flight. The parameters are evaluated separately for each $B \rightarrow h^+h'^-$ decay and combined using the branching fractions and the particle identification efficiencies for each decay. The mass PDFs are evaluated separately for Run 1 and Run 2 data and the same PDF is used for all BDT bins.

The total number of mis-identified $B \rightarrow h^+h'^-$ decays expected in Run 1 and Run 2 data, $\mathcal{N}_{B \rightarrow hh \rightarrow \mu\mu}$, is found using the relationship

$$\mathcal{N}_{B \rightarrow hh \rightarrow \mu\mu} = \epsilon_{B_{(s)}^0 \rightarrow \mu\mu, B \rightarrow hh}^{TRIG} \cdot \frac{\mathcal{N}_{B \rightarrow hh}}{\epsilon_{B \rightarrow hh}^{TRIG}} \cdot \epsilon_{B \rightarrow hh \rightarrow \mu\mu} \quad (5.5)$$

where $\mathcal{N}_{B \rightarrow hh}$ is the number of TIS $B \rightarrow h^+h'^-$ decays in data, $\epsilon_{B_{(s)}^0 \rightarrow \mu\mu, B \rightarrow hh}^{TRIG}$ are the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ trigger efficiencies and $\epsilon_{B \rightarrow hh \rightarrow \mu\mu}$ is the probability that a $B \rightarrow h^+h'^-$ decay is mis-identified as $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The number of $B \rightarrow h^+h'^-$ decays is calculated from a fit to the mass distribution of $B^0 \rightarrow K^+\pi^-$ candidates

identified in data, the number of $B^0 \rightarrow K^+\pi^-$ candidates is translated into the total number of $B \rightarrow h^+h^-$ by scaling the $B^0 \rightarrow K^+\pi^-$ yields by the relative production rates of $B^0 \rightarrow K^+\pi^-$ decays compared to the other $B \rightarrow h^+h^-$ decays. The trigger efficiencies are calculated from simulated decays and the mis-identification probabilities are evaluated using the PIDCalib package [150]. Once the total number of mis-identified $B \rightarrow h^+h^-$ decays in Run 1 and Run 2 data are found, the probability that a $B \rightarrow h^+h^-$ decay is mis-identified as a $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decay is then evaluated for each BDT bin in the data sets. The number of mis-identified $B \rightarrow h^+h^-$ decays expected in each BDT bin is calculated by multiplying the total number of mis-identified $B \rightarrow h^+h^-$ decays with the mis-identification probability for each BDT bin. The output of this procedure gives the expected number of mis-identified $B \rightarrow h^+h^-$ decays in each BDT bin for Run 1 and Run 2 data separately.

5.3.2 Exclusive backgrounds

The mass PDFs and expected yields for exclusive backgrounds from $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$, $B^0 \rightarrow \pi^0\mu^+\mu^-$ and $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ decays are all evaluated using the same techniques.

The mass PDFs of exclusive backgrounds vary across the BDT range. Therefore these PDFs are evaluated using simulated decays for each BDT bin separately. An Argus function [158] convoluted with a Gaussian function is used to describe the mass distributions. The Gaussian function accounts for the mass resolution of the detector. The shapes of $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ are extremely similar and therefore these backgrounds are modelled with one common PDF. Similarly one mass PDF is used to model $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$ decays. The mass PDFs are shown in Figure 5.5 for high BDT output values.

The expected yields of the exclusive backgrounds in each BDT bin are estimated by using to the number of $B^+ \rightarrow J/\psi K^+$ decays observed in data via

$$\mathcal{N}_x^{exp} = \mathcal{N}_{B^+ \rightarrow J/\psi K^+} \cdot \frac{f_x}{f_u} \cdot \frac{\mathcal{B}_x}{\mathcal{B}_{B^+ \rightarrow J/\psi K^+}} \cdot \frac{\epsilon_x}{\epsilon_{B^+ \rightarrow J/\psi K^+}} \quad (5.6)$$

where x represents each background decay. The background estimation can be factorised as

$$\mathcal{N}_x^{exp} = \beta \cdot f_x \cdot \epsilon_x \cdot \mathcal{B}_x \quad (5.7)$$

where β combines the yield, selection efficiency and hadronisation factor of $B^+ \rightarrow J/\psi K^+$ decays and is the same for all backgrounds. The β term is evaluated using the same method as the normalisation of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions described in Section 5.4. The $B^+ \rightarrow J/\psi K^+$ efficiencies and yields are evaluated

across the full BDT range whereas the detection and selection efficiency of each background, ϵ_x , are evaluated separately for each BDT bin using information from both data and simulated decays. The hadronisation factors and branching fractions are specific to each background and where possible measured, rather than predicted, branching fractions are used. The branching fraction values used for each background are given in Table 4.1. The output of the procedure gives the expected yields of each background in each BDT bin for both Run 1 and Run 2 data.

5.3.3 Combinatorial background

The combinatorial background is the most straight forward background to model and it is described by an exponential function. The slope of the function is required to have the same value for all BDT bins but the slope can be different for Run 1 and Run 2 data. These slopes are determined from a simultaneous fit to candidates in data across the BDT bins in the mass ranges $[4900, (m_{B_{(s)}^0} - 80)]$ MeV/ c^2 and $[(m_{B_{(s)}^0} + 80), 6000]$ MeV/ c^2 , where the mass shapes and yields of the remaining backgrounds are constrained to their expected values. The expected number of combinatorial background decays does not need to be evaluated because the yield is not constrained in the fit to measured the branching fractions.

5.4 Normalisation

The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions are measured by normalising the number of observed $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays to the number of observed $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ decays. The normalisation parameters $\alpha_{d(s)}$, in Equation 5.3 for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays depend on the yields of the normalisation decays, the ratio of the detection and selection efficiencies and the hadronisation factors. The yields of $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ decays are evaluated separately for Run 1 and Run 2 data and this process is described in Section 5.4.1. The evaluation of the selection efficiencies is described in Section 5.4.2 and this is done separately for each year of data taking. The hadronisation factors used for the normalisation discussed in Section 5.4.3 and depend on the centre-of-mass energy of collisions. The different inputs are combined to obtain the final normalisation parameters for Run 1 and Run 2 data that are presented in Section 5.4.4. In addition to the normalisation channels, $B_s^0 \rightarrow J/\psi\phi$ decays are used to check the normalisation parameters. Consequently the yields of decays and the detection and selection efficiencies must also be evaluated. This is done in the same way as the normalisation channels.

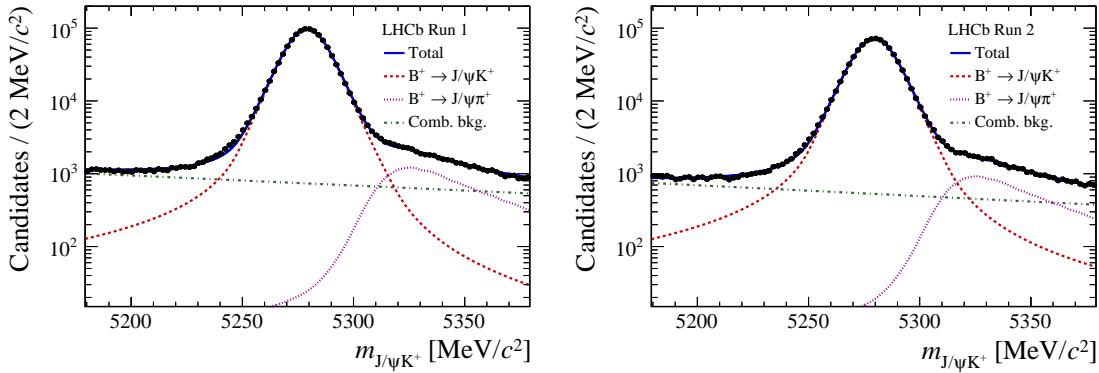


Fig. 5.6 Mass fit to measure the $B^+ \rightarrow J/\psi K^+$ yield for the normalisation for Run 1 and Run 2 data. The total PDF is made up of components of $B^+ \rightarrow J/\psi K^+$ and $B^+ \rightarrow J/\psi \pi^+$ decays and combinatorial background.

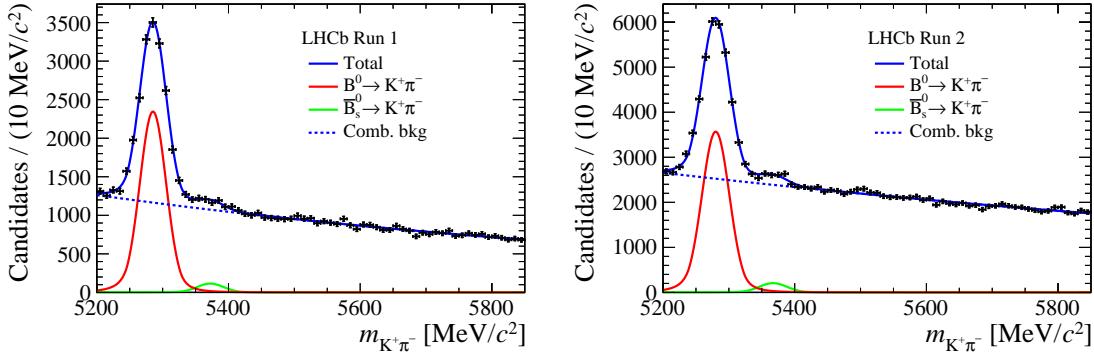


Fig. 5.7 Mass fit to measure the $B^0 \rightarrow K^+ \pi^-$ yield for the normalisation for Run 1 and Run 2 data. The total PDF is made up of components for $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ \pi^-$ decays and combinatorial background.

5.4.1 $B^0 \rightarrow K^+ \pi^-$ and $B^+ \rightarrow J/\psi K^+$ yields

The yields of $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$ decays are measured from data from fitting the mass distributions of these decays in Run 1 and Run 2 data. The $B^+ \rightarrow J/\psi K^+$ mass PDF is modelled by an Ipathia function [159] and the fit includes components for combinatorial background and $B^+ \rightarrow J/\psi \pi^+$ decays that are mis-identified as $B^+ \rightarrow J/\psi K^+$ decays. The mass PDF parameters are determined from a mixture of information from data and simulated decays. The $B^0 \rightarrow K^+ \pi^-$ mass fit includes components for $B_s^0 \rightarrow K^+ \pi^-$ decays and combinatorial background as well as $B^0 \rightarrow K^+ \pi^-$ decays. Both $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ \pi^-$ decays are modelled by double Crystal Ball functions and the parameters are determined from a mixture of information from data and simulated decays. Figure 5.6 and 5.7 show the mass fits used to calculate the Run 1 and Run 2 $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$ yields.

5.4.2 Efficiency ratio

The efficiency ratio in Equation 5.3 is divided into several separate efficiency terms

$$\frac{\epsilon_{norm}}{\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{Acc}} = \frac{\epsilon_{norm}^{Acc}}{\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{Acc}} \cdot \frac{\epsilon_{norm}^{RecSel|Acc}}{\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{RecSel|Acc}} \cdot \frac{\epsilon_{norm}^{Trig|RecSel}}{\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{Trig|RecSel}}, \quad (5.8)$$

where ϵ^{Acc} is the detector acceptance efficiency, $\epsilon^{RecSel|Acc}$ the reconstruction and selection efficiency given the detector efficiency, and $\epsilon^{Trig|RecSel}$ the trigger efficiency given the reconstruction and selection efficiency.

The detector acceptance efficiency gives the efficiency for the decay products to be within the LHCb detector angular acceptance. This efficiency is evaluated on simulated decays for decay products that fall within the range [10,400] mrad in both x and y directions. The range is chosen to be slightly larger than the detector acceptance so that particles recovered by the magnetic field are included. To keep this efficiency similar for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$ decays, the hadrons from $B^0 \rightarrow K^+\pi^-$ decays are required to be within the muon detector acceptance.

The reconstruction and selection efficiency of decays within the detector acceptance is evaluated from a combination of information from data and simulated decays. Similar to the fraction of $B_s^0 \rightarrow \mu^+\mu^-$ in each BDT bin, a correction is applied for the lifetime used in simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays assuming $A_{\Delta\Gamma} = +1$.

The trigger efficiencies for decays passing the reconstruction and selection are evaluated for each decay by data driven methods as described in [57, 160].

The efficiencies are calculated for $B_s^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow K^+\pi^-$ and $B^+ \rightarrow J/\psi K^+$ separately to account for differences between the decay kinematics. The ratio of efficiencies between signal and normalisation channels used in the normalisation parameters ensures that systematic uncertainties arising from the use of simulated decays cancel out and will not affect the measurements of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions.

5.4.3 Hadronisation factors

The normalisation factors depend on the hadronisation factors, f_u , f_s , f_d , that give the fraction of $b\bar{b}$ pairs that produce B^\pm , B^0 or \overline{B}^0 and B_s^0 or \overline{B}_s^0 mesons. The factors f_d and f_u are equal to good approximation, therefore the $B^0 \rightarrow \mu^+\mu^-$ branching fraction does not depend on any hadronisation factors. For the $B_s^0 \rightarrow \mu^+\mu^-$ decay the ratio f_s/f_d is used in the normalisation, since $f_d = f_u$. This ratio has been measured at LHCb for pp collisions at $\sqrt{s} = 7$ TeV [161]. The stability of this ratio at different centre-of-mass energies was tested using the ratio of $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ decays at 8 and 13 TeV relative to their ratio at 7 TeV. The ratios are stable across the different collision energies and f_s/f_d is assumed to be identical for

Normalisation Parameters	Run 1	Run 2
$\alpha_d \times 10^{11}$	2.877 ± 0.101	3.521 ± 0.155
$\alpha_s \times 10^{10}$	1.071 ± 0.072	1.306 ± 0.095

Table 5.3 Normalisation parameters for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ for Run 1 and Run 2.

$\sqrt{s} = 8$ TeV. For Run 2, the f_s/f_d ratio is modified due to a small relative production difference in $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ decays observed for Run 2 compared to Run 1. The uncertainty on the hadronisation factor ratio contributes the largest uncertainty to the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction.

Alternatively, the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction could be normalised using a different B_s^0 decay, such as $B_s^0 \rightarrow J/\psi\phi$. However the precision of the measured branching fractions and abundance of such decays is not high enough at present to provide a lower overall uncertainty on the measured branching fraction.

5.4.4 Normalisation parameters

The yields, efficiencies and hadronisation factors are combined to produce separate normalisation factors for each year of data taking and each normalisation channel. The consistency of the efficiencies and yields for each normalisation channel are checked for each year by comparing the ratios $\mathcal{B}(B^0 \rightarrow K^+\pi^-)/\mathcal{B}(B^+ \rightarrow J/\psi K^+)$ and $\mathcal{B}(B^+ \rightarrow J/\psi K^+)/\mathcal{B}(B_s^0 \rightarrow J/\psi\phi)$ with the average of previously measured values of these quantities in reference [5]. The efficiencies and yields are consistent with the measured values for these decays.

The yearly normalisation factors are combined for each channel to produce the overall normalisation factors for Run 1 and Run 2, taking into account correlations between the parameters. A weighted average of the normalisation factors for $B^0 \rightarrow K^+\pi^-$ and $B^+ \rightarrow J/\psi K^+$ decays are used to produce the overall normalisation factors for Run 1 and Run 2 as shown in Table 5.3.

5.5 Results

The $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ branching fractions are measured by a simultaneous fit to the dimuon invariant mass distribution across eight categories: Run 1, Run 2 and four BDT bins, as described in Section 5.1. The fit results are shown in Figure 5.8.

In the mass fit all PDFs, except the combinatorial background, are constrained within Gaussian limits of their expected values based on the uncertainties PDF parameters. The fraction of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays in each BDT bin is constrained

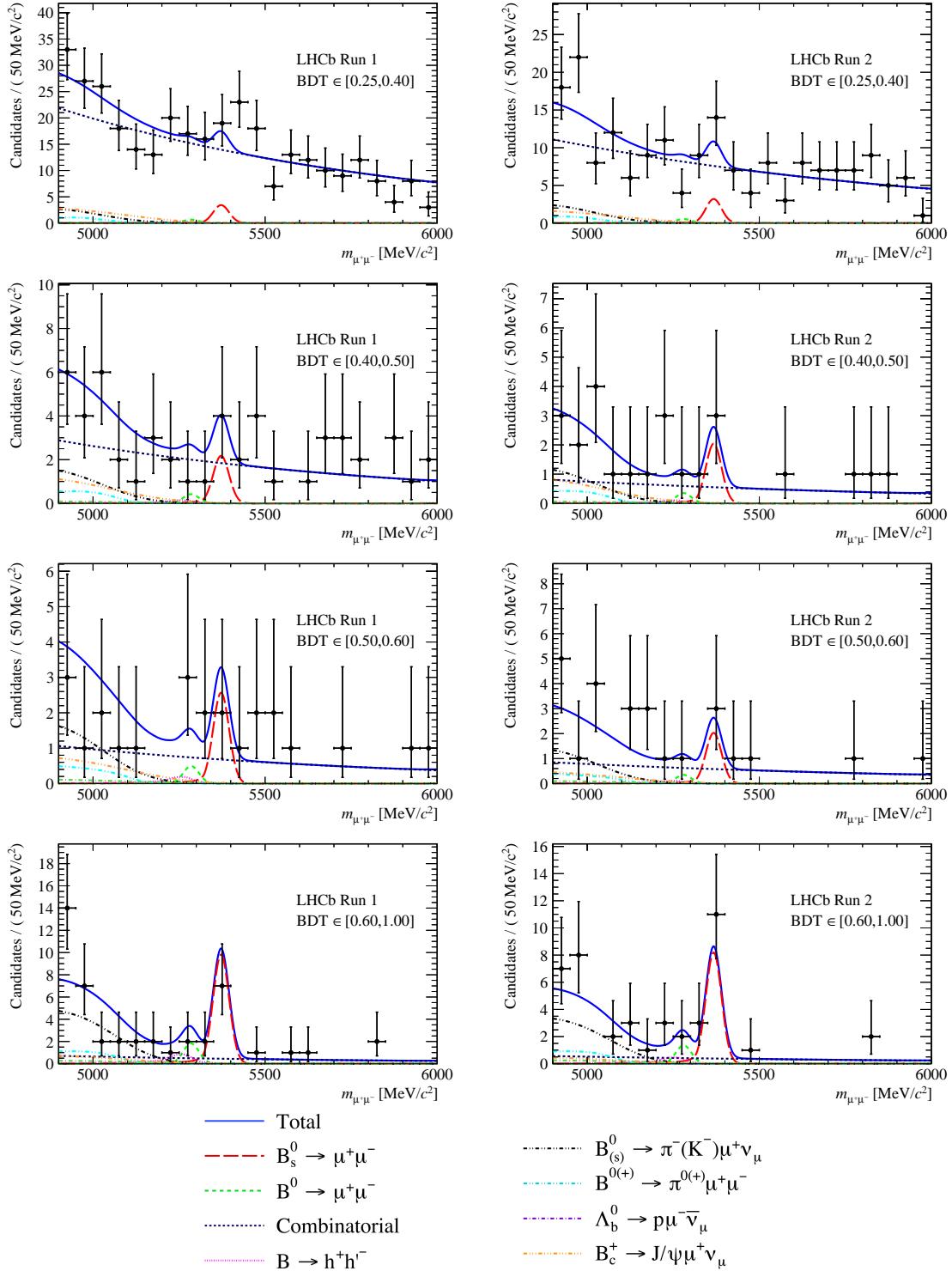


Fig. 5.8 Mass distribution in BDT bins for selected $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ candidates with the fit overlaid for Run 1 and Run 2 data. The fit includes components for $B^0 \rightarrow \mu^+\mu^-$, $B_s^0 \rightarrow \mu^+\mu^-$, combinatorial backgrounds and backgrounds from mis-identified decays.

using the BDT PDF and the yields of the mis-identified background are constrained around their expected values. The combinatorial background yields are left free in the fit and the slope of the mass distribution is required to have the same value across all bins for each data set. The measured branching fractions are

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.5^{+1.2}_{-1.0}{}^{+0.2}_{-0.1}) \times 10^{-10}\end{aligned}\quad (5.9)$$

where the first quoted uncertainty is the statistical uncertainty and the second is the systematic uncertainty. The dominant contributions to the systematic uncertainties are from the ratio f_s/f_d and the uncertainty on the background yields for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions, respectively. The statistical significance of the $B_s^0 \rightarrow \mu^+ \mu^-$ signal is 7.8σ making this measurement the first single experiment observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The significance of the $B^0 \rightarrow \mu^+ \mu^-$ signal is 1.6σ , therefore the CL_s method [162] is used to place an upper limit on the branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$ at the 95% confidence level.

The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction in Equation 5.9 assumes the Standard Model value for $A_{\Delta\Gamma}$ of $A_{\Delta\Gamma} = +1$, applying the corrections detailed in Section 5.2.2; $A_{\Delta\Gamma}$ values of 0 and -1 shift the central value of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ by 4.6% and 10.9%, respectively. All results are consistent with the predictions of the SM as illustrated in Figure 5.9.

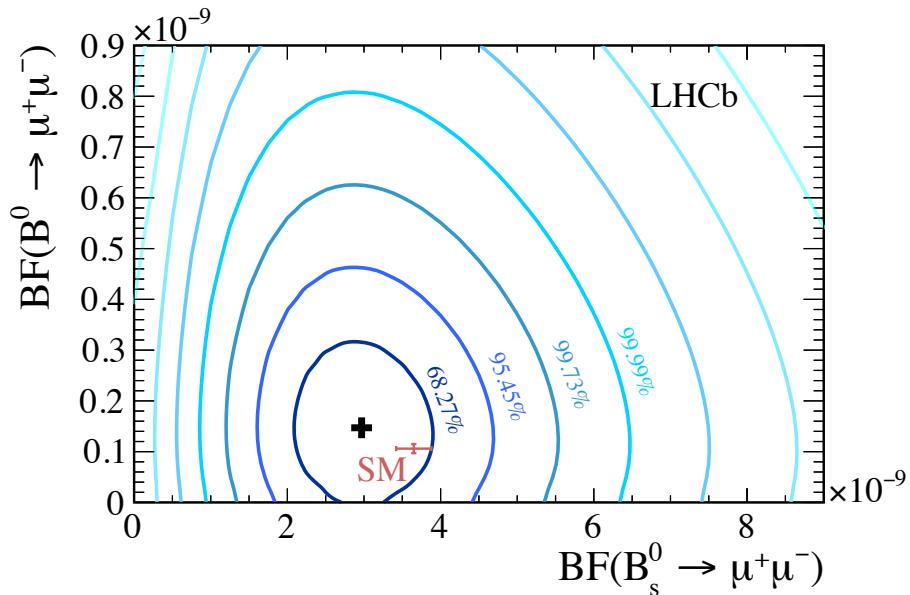


Fig. 5.9 $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ 2-dimensional likelihood plot for the simultaneous branching fraction fit to Run 1 and Run 2 data.

Chapter 6

Measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime

This chapter describes the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. Section 6.1 presents an overview of the analysis strategy. The PDFs of the mass and decay time distributions for signal and background candidates are described in Sections 6.2 and 6.3. Due to the very rare nature of $B_s^0 \rightarrow \mu^+\mu^-$ decays, the fit configuration used to measured the effective lifetime is optimised to produce the smallest expected statistical uncertainty on the measured value, the optimisation studies are detailed in Section 6.4. Finally, the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is presented in Section 6.5.

6.1 Analysis strategy

The $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is measured from the decay time distribution of $B_s^0 \rightarrow \mu^+\mu^-$ candidates passing the selection criteria described in Section 4.4. Since the selection requirements do not completely separate real $B_s^0 \rightarrow \mu^+\mu^-$ decays from the backgrounds, in order to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, either the PDFs describing the decay time distributions of the signal and all backgrounds must be known, or the background candidates must be removed from the data leaving only the signal distribution. Several approaches were investigated to determine which would produce stable results for the measured $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime for the data set and yield the smallest expected statistical uncertainty on the result. The most successful approach was found to be the sPlot statistical weighting method, described in reference [155], which allows the signal and background components of a dataset to be disentangled in a statistically rigorous way.

The sPlot method produces a two step strategy to measure the effective lifetime. The first step is an unbinned extended maximum likelihood fit to the dimuon invariant mass spectrum, where components are included in the PDF for $B_s^0 \rightarrow \mu^+\mu^-$ decays

Year	$B_s^0 \rightarrow \mu^+\mu^-$ correlation	$b\bar{b} \rightarrow \mu^+\mu^- X$ correlation
2011	-0.008	0.003
2012	-0.006	0.008
2015	-0.006	0.010
2016	0.008	0.002

Table 6.1 Correlation between mass and decay time for candidates from $B_s^0 \rightarrow \mu^+\mu^-$ simulated decays and combinatorial background decays from data for 2011, 2012, 2015 and 2016 data taking conditions. The selection given in Table 4.21 is applied to simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays and the same selection is applied to decays in data apart from the global BDT cut and have a dimuon invariant mass of $> 5447 \text{ MeV}/c^2$.

and each background decay in the data. The mass fit determines the yields of the signal and background decays and from the fit sWeights are calculated for each component. In the second step the sWeights are applied to the data, effectively removing all background decays. An unbinned maximum likelihood fit to the decay time distribution of the sWeighted data is performed to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. In the final fit only the $B_s^0 \rightarrow \mu^+\mu^-$ decay time PDF is needed to measure the effective lifetime as any background will have been removed by the sPlot technique. The PDF describing the decay time will need to take into account the selection efficiency as a function of decay time. Due to the low statistics expected for the data set, Run 1 and Run 2 data are combined and the fit to the mass and weighted decay time distributions are performed to the combined data.

The approach outlined here is suited to the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime because the mass PDFs are accurately known for the signal and background decays from the branching fraction analysis. Furthermore, no knowledge is required for the decay time PDFs of the backgrounds in the final fit. This is advantageous because decay time distributions of combinatorial background decays is challenging to accurately model.

A requirement of the sPlot procedure is that the variable used to calculate the sWeights and the variable from which the observable is measured must be independent. The correlation of the mass and decay time for $B_s^0 \rightarrow \mu^+\mu^-$ decays and combinatorial background decays has been evaluated using simulated decays and data. The correlation is negligible as shown in Table 6.1. Therefore the invariant mass of the $B_s^0 \rightarrow \mu^+\mu^-$ candidate can be used to accurately determine sWeights applied to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime.

The sWeights used in the fit are calculated using the RooFit package [163]. However, the raw sWeights from the mass fit cannot be used directly in the maximum likelihood fit to measure the effective lifetime. The normalisation of the sWeights will

not produce the correct statistical uncertainty on the effective lifetime measurement. The uncertainty will be related to the sum of the weights rather than the number of candidates in the data set [164, 165]. Therefore, the fit uses re-normalised sWeights given by

$$\omega'_i = \omega_i \cdot \frac{\sum_j \omega_j}{\sum_j \omega_j^2}, \quad (6.1)$$

where ω_i are the sWeight values for each decay and ω_j are sWeights summed over all decays. The re-normalised sWeights will produce the correct statistical uncertainty in a maximum likelihood fit to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime.

The measurement strategy of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime requires the mass and decay time PDFs of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the mass PDFs of background that pass the selection criteria in Chapter 4. However, the overall performance of this strategy depends on the fit to the invariant mass distribution and how accurately the sWeights are calculated. Pseudoexperiments are performed to determine the best mass fit configuration to measure the effective lifetime and the fit is studied for different $B_s^0 \rightarrow \mu^+ \mu^-$ mass ranges, the largest of which is 4900 - 6000 MeV/ c^2 . Therefore the pseudoexperiments require the decay time PDFs of the backgrounds as well as the signal and the signal and background mass PDFs in the range 4900 - 6000 MeV/ c^2 . The mass PDFs used for the effective lifetime measurement are described in Section 6.2 and the decay time PDFs are described in Section 6.3. The pseudoexperiments used to determine the fit configuration are described in Section 6.4 along with the final fit configuration and the expected sensitivity of the measurement. Finally, the results are presented in Section 6.5.

6.2 Mass PDFs

The selection criteria used to identify $B_s^0 \rightarrow \mu^+ \mu^-$ candidates for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction and $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime measurements are very similar. Therefore, the background decays passing for the effective lifetime measurement are the same as those passing the branching fraction, although the yields are different. Consequently the mass fit used to extract the sWeights is very similar to the fit used for the branching fraction measurements in Sections 5.2 and 5.3. The PDF used in the mass fit has the form

$$\mathcal{P}_{tot}(m_{\mu\mu}) = N_{sig} \mathcal{P}_{sig}(m_{\mu\mu}) + \sum_i N_{bkg}^i \mathcal{P}_{bkg}^i(m_{\mu\mu}), \quad (6.2)$$

where i represents a particular background, $N_{sig(bkg)}$ are the signal (background) yields and $P_{sig(bkg)}$ are the signal (background) PDFs. The background decays

include; $B^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$, $B^0 \rightarrow \pi^0\mu^+\mu^-$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ and combinatorial background decays. For the effective lifetime measurement the $B^0 \rightarrow \mu^+\mu^-$ decay is included as a background.

The $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ mass PDFs are described by the same Crystal Ball functions used in the branching fraction measurements, with the Run 1 parameters given in Table 5.1. The choice of Run 1 or Run 2 parameters in the PDF has a negligible affect on the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime as shown in Section 7.3.

Mis-identified semi-leptonic decays, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$, $B^0 \rightarrow \pi^0\mu^+\mu^-$ and $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$ are each described by an Argus function convoluted with a Gaussian function evaluated from simulated decays using the same method as described in Section 5.3. The particle identification requirements and the cut on the global BDT used in the selection of candidates for the effective lifetime measurement are taken into account in the evaluation of the PDF shapes.

Backgrounds from mis-identified $B \rightarrow h^+h'^-$ decays are described by a double Crystal Ball function evaluated using the method described in Section 5.3 with the effective lifetime particle identification requirements applied. Finally, the combinatorial background is modelled with a decaying exponential where the slope is not constrained in the final fit.

6.3 Decay time distributions

The efficiency of the selection criteria for both signal and background decays varies as a function of decay time which biases the decay time distribution, as described in Chapter 4. Therefore, the PDF describing the decay time is not just an exponential but is described by

$$\mathcal{P}(t) = \epsilon(t) \times e^{-t/\tau}, \quad (6.3)$$

where $\epsilon(t)$ is the selection efficiency as a function of decay time. The decay time distribution and selection efficiency as a function of decay time are shown in Figure 6.1 for simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays at different stages through the selection. The cut on the global BDT causes the biggest decay time bias which is as expected since it is the hardest selection cut applied.

To measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, the efficiency of the selection of $B_s^0 \rightarrow \mu^+\mu^-$ decays as a function of decay time must be accurately modelled. The determination of $\epsilon(t)$ for $B_s^0 \rightarrow \mu^+\mu^-$ decays is described in Section 6.3.1. Although the sPlot method used to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime means that the decay time PDFs of the backgrounds present in the data set are not needed, realistic descriptions of the background decay time PDFs are necessary for optimising the

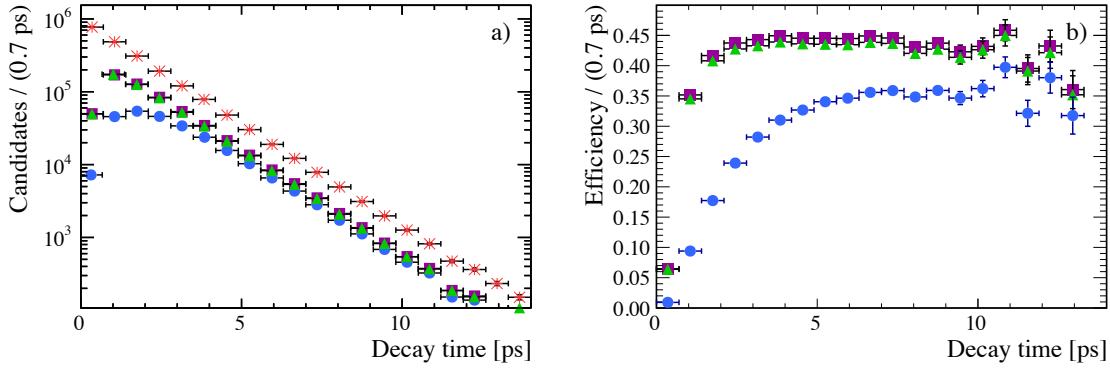


Fig. 6.1 Decay time distribution (a) and selection efficiency as a function of decay time (b) for 2012 $B_s^0 \rightarrow \mu^+\mu^-$ simulated decays at different stages of the selection process. The decay time distributions and efficiencies are shown for reconstructed decays that pass the trigger, stripping and pre-selection cuts (magenta squares), the decays that go on to pass PID requirements (green triangles) and decays that pass all selection requirement including the global BDT cut (blue circles). Also the decay time distribution is shown for all generated simulated decays (red stars).

mass fit configuration. The background PDFs, both the efficiency as a function of decay time and the lifetime used for each background, are described in Section 6.3.2.

6.3.1 $B_s^0 \rightarrow \mu^+\mu^-$ decay time PDF

The selection efficiency of $B_s^0 \rightarrow \mu^+\mu^-$ decays as a function of decay time is modelled by an acceptance function and a range of different models were investigated. The model that described the $B_s^0 \rightarrow \mu^+\mu^-$ decay time efficiency best was the parametrised acceptance used in reference [166]

$$\epsilon(t) = \frac{[a(t - t_0)]^n}{1 + a(t - t_0)]^n}, \quad (6.4)$$

where t is the decay time, a and n describe the curvature of the efficiency at low decay times and t_0 is the decay time below which the efficiency is zero.

The acceptance function parameters are determined from a fit to simulated decay and these parameters are then fixed in the fit to data to measure the effective lifetime. A systematic uncertainty describing how well the acceptance function is understood is detailed in Section 7.4.

There are several steps performed in the evaluation of the acceptance function parameters to ensure the final function describes the selection efficiency of $B_s^0 \rightarrow \mu^+\mu^-$ decays accurately. First, simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays are weighted using information from $B^0 \rightarrow K^+\pi^-$ decays in data and simulation to correct for mis-modelling of the underlying event in simulation. Second, simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays from each year of data taking are combined into one sample to account for different selection

efficiencies for each year. Third, a unbinned maximum likelihood fit is performed to the decay time distribution of the weighted simulated decays to find the acceptance function parameters. Each of the steps is described in the following.

$B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays weighted using $B^0 \rightarrow K^+ \pi^-$ decays

In general simulated decays model distributions in data reasonably well, however the number of tracks present in an event are not well modelled in the simulation. Although the $B_s^0 \rightarrow \mu^+ \mu^-$ decay time distribution does not depend on the number of tracks present in the event, the isolation criteria used in the global BDT do. Therefore, the selection efficiency as a function of decay time depends on the number of tracks in the event and cannot be accurately described by simulated decays alone. To overcome this the number of tracks for simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays are weighted using information from the number of tracks per event for $B^0 \rightarrow K^+ \pi^-$ decays in both data and simulation. The number of tracks for $B^0 \rightarrow K^+ \pi^-$ decays is found for decays in data and simulated decays for each year of data taking where no requirement is made on the output of the BDT. The difference between the distributions from simulation and data are used to evaluated weights which are applied to the simulated decays so that the distribution of number of tracks in an event matches the distribution in data. These weights are then applied to simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays before the BDT cut is applied. The affect of the weights can be seen in the different decay time distribution of weighted and unweighted decays that have a BDT value of $\text{BDT} > 0.55$.

The selection requirements described in Chapter 4 are used to identify $B^0 \rightarrow K^+ \pi^-$ decays in data and simulation but, importantly the global BDT cut is not applied. The distribution of the number of tracks $B^0 \rightarrow K^+ \pi^-$ decays in data is obtained by performing a fit to the $K\pi$ mass distribution and extracting sWeights. The sWeights are applied to data to find the distribution of the number of tracks in an event for $B^0 \rightarrow K^+ \pi^-$ decays. The distribution of the weighted number of tracks per event in data is compared with the distribution in simulated $B^0 \rightarrow K^+ \pi^-$ decays. The mass fits to $B^0 \rightarrow K^+ \pi^-$ decays in data are shown in Figure 6.2 and the normalised distributions of the number of tracks per event in weighted data and simulated decays are shown in Figure 6.3. Each year of data taking is kept separate and the same simulation version is used for $B^0 \rightarrow K^+ \pi^-$ simulated decays as available for $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

The distributions of the number of tracks per event for $B^0 \rightarrow K^+ \pi^-$ decays in data and simulated decays are used to weight $B^0 \rightarrow K^+ \pi^-$ decays so that the distribution in simulation matches that in data. The weights are evaluated by taking the ratio of the normalised histograms in Figure 6.3 for each year. The affect on the decay time distribution of using these weights and then applying the global BDT cut

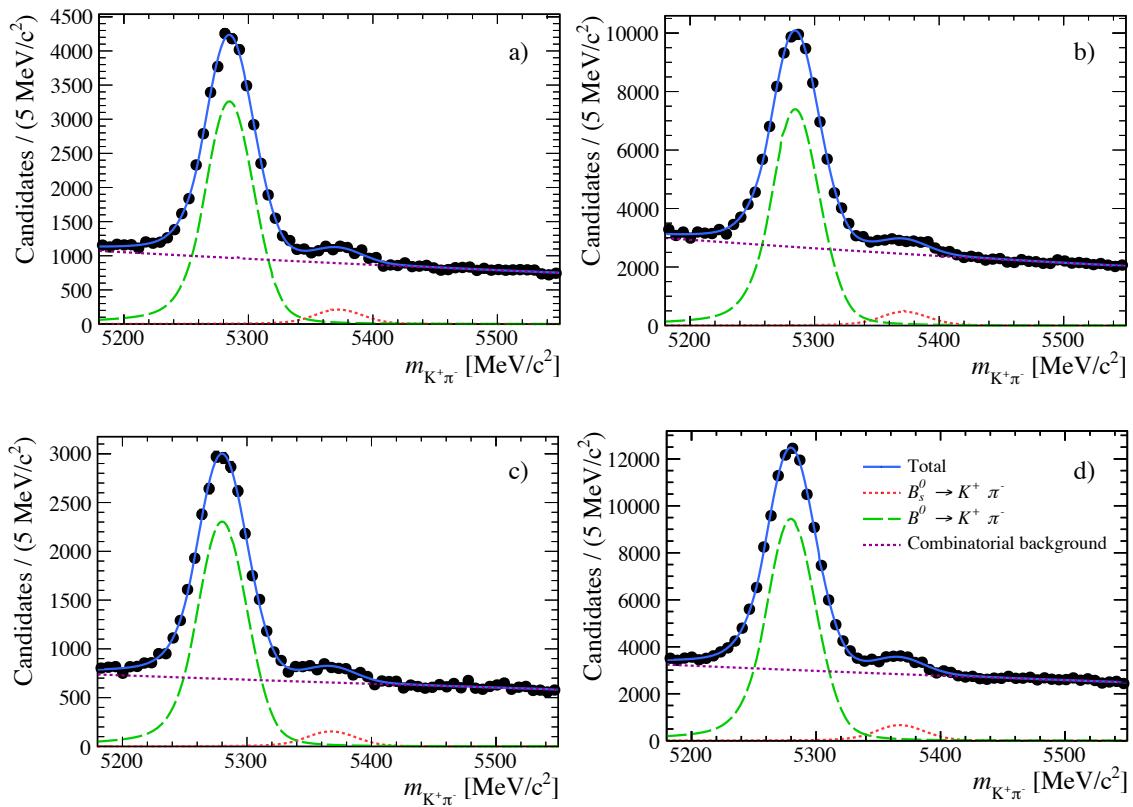


Fig. 6.2 Maximum likelihood fits to the mass distribution of $B^0 \rightarrow K^+\pi^-$ candidates in a) 2011, b) 2012, c) 2015 and d) 2016 data. The mass PDF includes components for $B^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+\pi^-$ and combinatorial background.

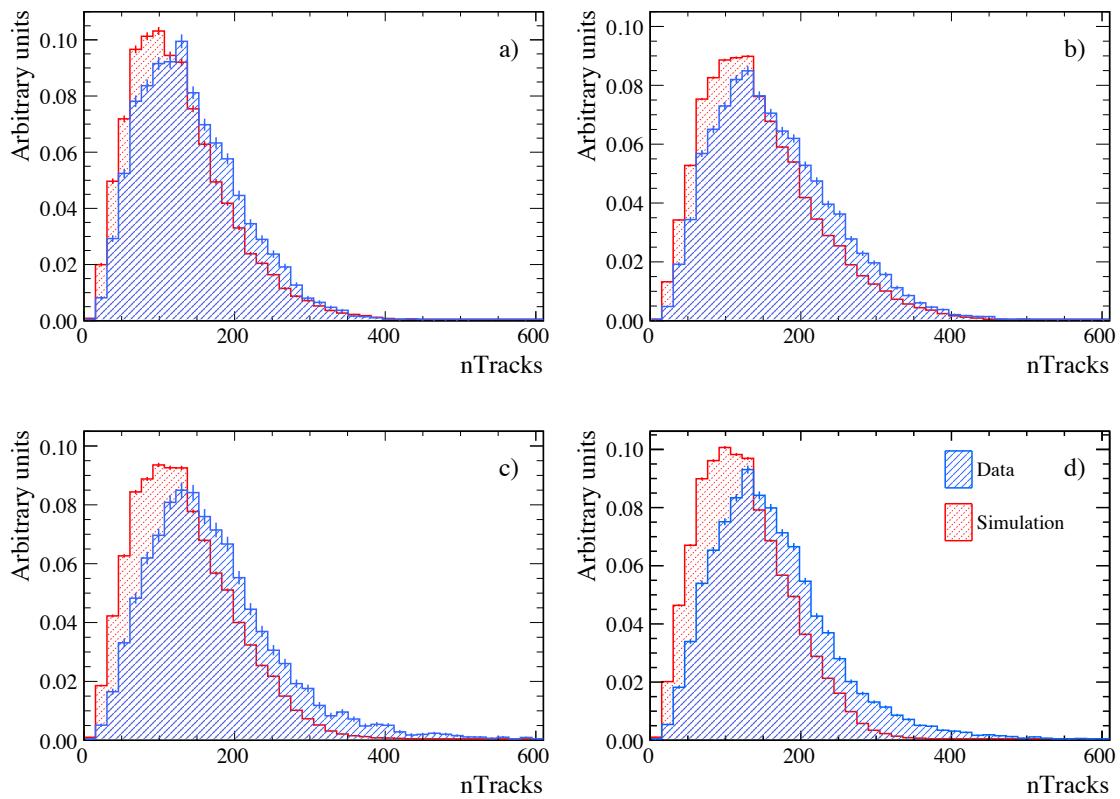


Fig. 6.3 Normalised histograms of the number of tracks per event (nTracks) in simulated $B^0 \rightarrow K^+ \pi^-$ decays and weighted $B^0 \rightarrow K^+ \pi^-$ decays in data for a) 2011, b) 2012, c) 2015 and d) 2016 data taking conditions.

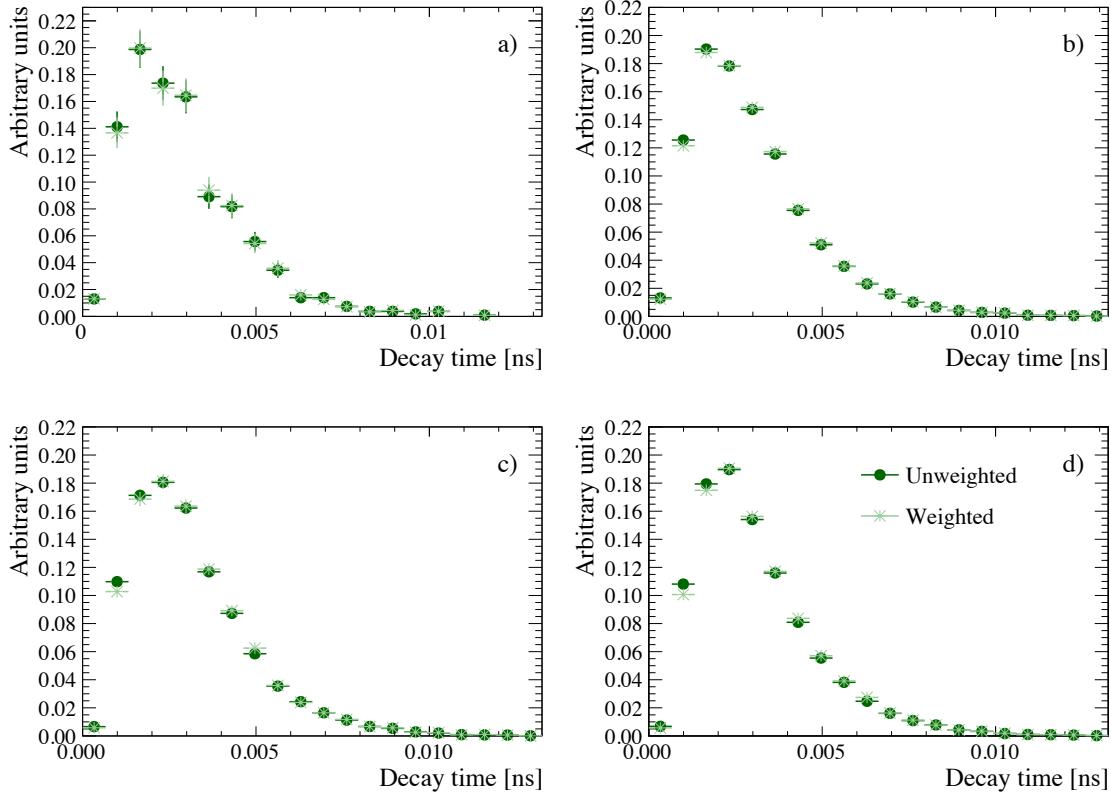


Fig. 6.4 Decay time distributions for weighted and unweighted $B^0 \rightarrow K^+\pi^-$ simulated decays for a) 2011, b) 2012, c) 2015 and d) 2016 data taking conditions.

is shown in Figure 6.4 for the simulated $B^0 \rightarrow K^+\pi^-$ decays. The difference between the decay time distributions with and without the weights is not large but clearly noticeable at low decay times where the change in selection efficiency is greatest.

The same weights are applied to simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays by binning the number of tracks per event for $B_s^0 \rightarrow \mu^+\mu^-$ decays in the same way as used for $B^0 \rightarrow K^+\pi^-$ decays. The weights are applied to decays that pass the selection but before the global BDT cut is applied. The decay time distributions for weighted and unweighted $B_s^0 \rightarrow \mu^+\mu^-$ simulated decays after the global BDT cut has been applied is shown in Figure 6.5. Similar to $B^0 \rightarrow K^+\pi^-$ decays, the largest effect is at low decay times where the change in selection efficiency, caused by the BDT cut, is greatest as seen in Figure 6.1.

The re-weighting relies on the number of tracks per event being similar for $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ decays. This cannot be evaluated in data due to the small number of $B_s^0 \rightarrow \mu^+\mu^-$ decays in data. However, Figure 6.6 shows a comparison of the number of tracks per event for simulated $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$ decays in each year and resulting distributions are similar.

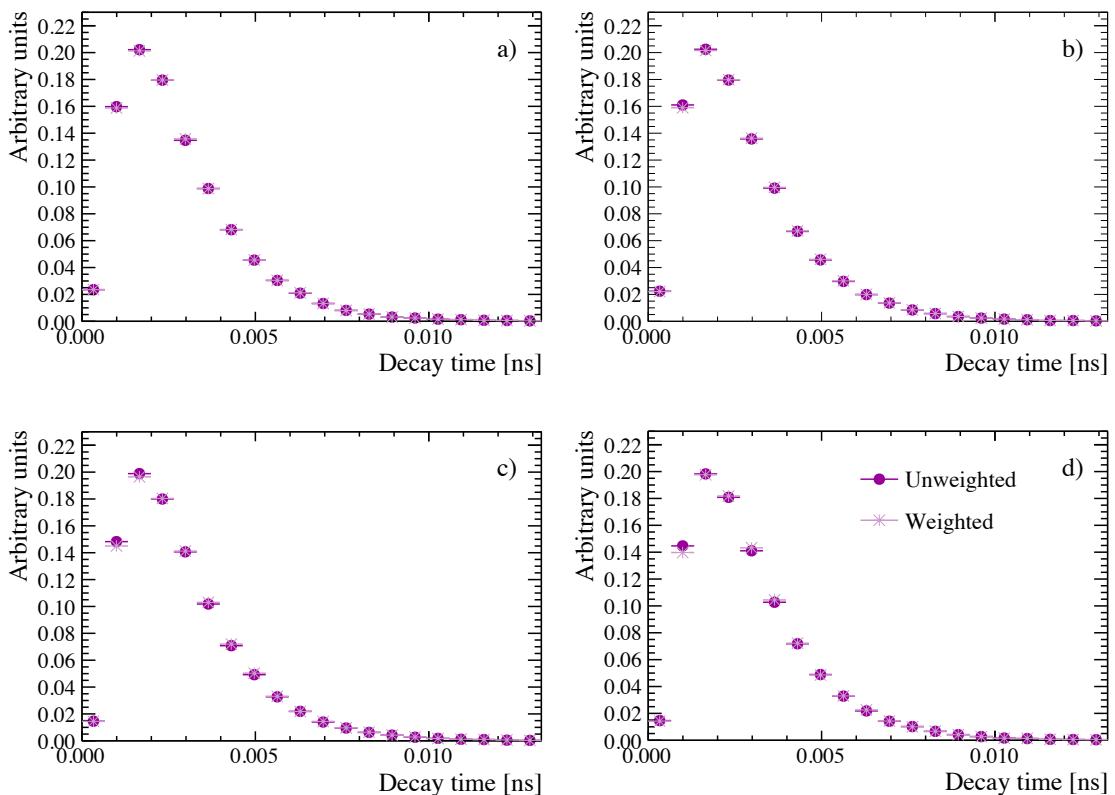


Fig. 6.5 Decay time distributions for weighted and unweighted $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays for a) 2011, b) 2012, c) 2015 and d) 2016 data taking conditions. Distributions have been normalised to have unit area.

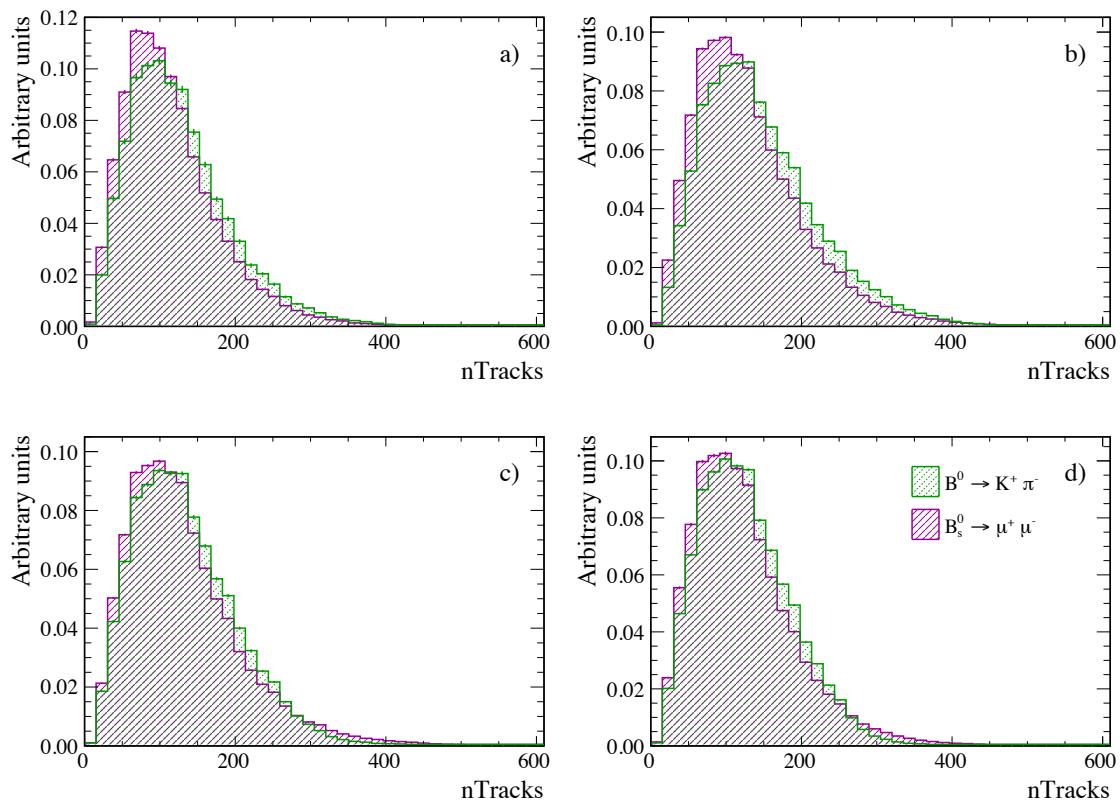


Fig. 6.6 Normalised histograms of the number of tracks per event in simulated $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays in data for a) 2011, b) 2012, c) 2015 and d) 2016 data.

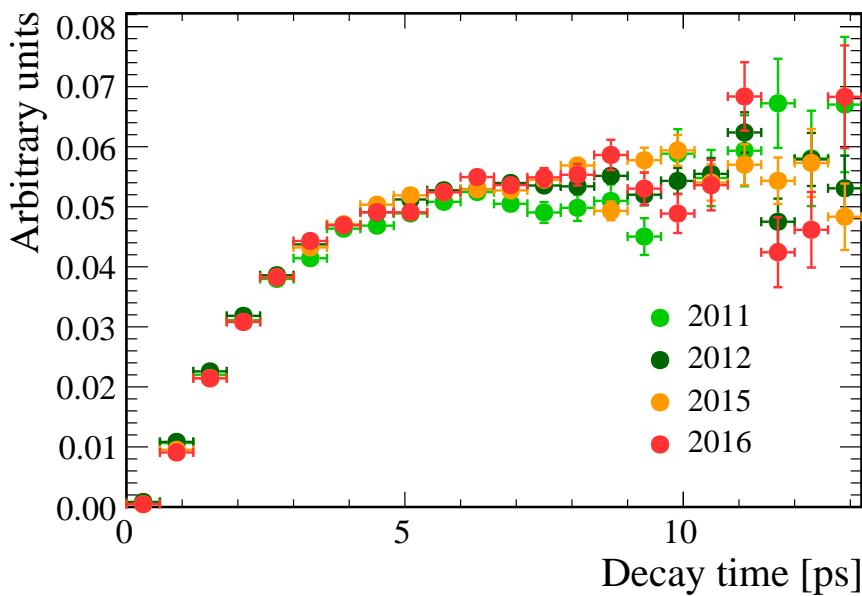


Fig. 6.7 Selection efficiency histograms for each year of data taking for weighted simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

Combining $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays from each year of data taking

So far, each year of data taking is treated separately. This is due to the different B_s^0 lifetimes used in the generation of simulated decays as well as different selection and trigger efficiencies. The different selection efficiencies for each data taking year are shown in Figure 6.7. Therefore to accurately model the $B_s^0 \rightarrow \mu^+ \mu^-$ acceptance function, simulated decays from each year of data taking must be used. The number of simulated decays available for each year does not correspond to the proportions of decays present in each year of the data. Therefore, additional weights are used to combine the simulated decays so that the combined set of decays has the same proportions of decays for each year as the complete data set.

The proportion of events expected in each year of data is taken from the number of $B_s^0 \rightarrow J/\psi \phi$ decays in data for each year corrected for the selection differences for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow J/\psi \phi$ decays. The $B_s^0 \rightarrow J/\psi \phi$ yields are extracted from fits to the mass spectrum of candidates in each year of data taking. The selection applied to identify $B_s^0 \rightarrow J/\psi \phi$ candidates is very similar to that applied to $B_s^0 \rightarrow \mu^+ \mu^-$ decays apart from the particle identification and global BDT requirements. This decay is chosen because the ratio of the efficiencies for the stripping, trigger and pre-selection requirements of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow J/\psi \phi$ decays is uniform across the different years making $B_s^0 \rightarrow J/\psi \phi$ decays a good proxy for $B_s^0 \rightarrow \mu^+ \mu^-$. The

Year (i)	Y_i	ϵ_i	N_i	ω_i	$\mathcal{N}_i \equiv N_i \omega_i$
2011	19190	0.412	70448	1.72	131364
2012	42103	0.406	254822	1.03	262461
2015	8571	0.410	222820	0.24	53917
2016	37765	0.406	124870	1.88	235218

Table 6.2 Weights used to combine simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays for each year to determine the acceptance function. Weights ensure the proportion of simulated events for each year matches what is expected in data.

weights applied to simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays are

$$\omega_i = \frac{Y_i^{J/\psi\phi} \epsilon_i}{\sum_j Y_j^{J/\psi\phi} \epsilon_j} \cdot \frac{\sum_k N_k^{\mu^+ \mu^-}}{N_i^{\mu^+ \mu^-}}, \quad (6.5)$$

where i represents the year, $Y^{J/\psi\phi}$ the $B_s^0 \rightarrow J/\psi\phi$ yields, $N^{\mu^+ \mu^-}$ the number of simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays available for the year passing the full $B_s^0 \rightarrow \mu^+ \mu^-$ selection and ϵ_i the efficiency of the particle identification and global BDT requirements for $B_s^0 \rightarrow \mu^+ \mu^-$ decays that have passed all other selection requirement evaluated from simulated decays. The sums, over k and j , are performed over all years of data taking. The weights applied to simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays and values of the different components of the weights are given in Table 6.2.

Acceptance parameter fit

The weights applied to $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays from $B^0 \rightarrow K^+ \pi^-$ decays ensure that the decay time efficiency is described accurately and the simulated decays have been combined to represent the proportion of decays present from each year of data. Therefore the acceptance parameters can now be determined. An unbinned maximum likelihood fit is performed to the decay time distribution of the weighted simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays to determine the acceptance parameters in Equation 6.4. In the fit the acceptance parameters are free and the $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime is constrained to the weighted average of lifetimes used to generate each year of simulated decays. The fit result is shown in Figure 6.8 and the acceptance parameters are given in Table 6.3. Figure 6.9 shows the shape of the fitted acceptance function.

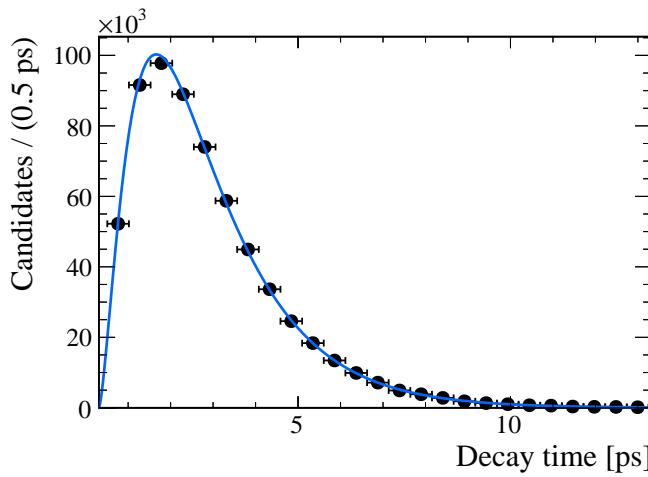


Fig. 6.8 Maximum likelihood fit to the combined decay time distribution of weighted 2011, 2012, 2015 and 2016 simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

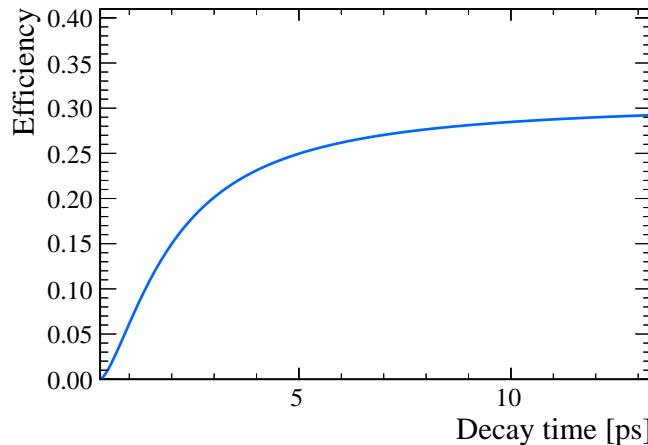


Fig. 6.9 Acceptance function determined from the fit to the combined decay time distribution of weighted 2011, 2012, 2015 and 2016 simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

Parameter	Value
a / ps^{-1}	0.574 ± 0.011
n	1.49 ± 0.03
t_0 / ps	0.313 ± 0.007

Table 6.3 Parameters for the $B_s^0 \rightarrow \mu^+ \mu^-$ acceptance function determined from weighted simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

6.3.2 Background decay time PDFs

The selection biases the decay time distributions of the backgrounds in a similar way as the $B_s^0 \rightarrow \mu^+ \mu^-$ decay time. Therefore, they are described by the same PDF as in Equation 6.3 and the acceptance function in Equation 6.4 is used to model the selection efficiency as a function of decay time. However, different acceptance function parameters are needed to describe the different backgrounds.

The backgrounds from mis-identified and $B^0 \rightarrow \mu^+ \mu^-$ decays are assigned the same acceptance parameters as $B_s^0 \rightarrow \mu^+ \mu^-$ decays because the decay time efficiency of these backgrounds is approximately the same as the signal. The acceptance function of these backgrounds do not need to be as accurately known as the acceptance function of the signal because very few background decays from these sources will be present in the dataset after the selection and the final result does not depend on the acceptance function of the backgrounds. The lifetimes of these background decays are taken from a fit to simulated decays. For $B \rightarrow h^+ h^-$ the fit is performed to a combined set of $B \rightarrow h^+ h^-$ decays representing what is expected in data.

The decay time PDF of the combinatorial background is more challenging to determine. This background arises from random combinations of muons in the event and not from one source, therefore there is no single lifetime that describes the background. Furthermore, the global BDT which is designed to separate $B_s^0 \rightarrow \mu^+ \mu^-$ decays from combinatorial background decays will have a different efficiency as a function of decay time for the combinatorial background compared to the $B_s^0 \rightarrow \mu^+ \mu^-$ decays. The decay time PDF of the combinatorial background cannot be evaluated from simulated decays or decays in data that pass the $B_s^0 \rightarrow \mu^+ \mu^-$ selection because there are too few candidates left after the selection has been applied. Therefore, it is evaluated from the combinatorial background of $B \rightarrow h^+ h^-$ decays using candidates in data that pass the $B \rightarrow h^+ h^-$ selection and have a dimuon invariant mass greater than 5447 MeV/ c^2 .

The decay time PDF for combinatorial background decays is modelled by

$$P_{\text{cbg}}(t) = \epsilon(t) \times \left(f \cdot e^{-t/\tau_1} + (1 - f) \cdot e^{-t/\tau_2} \right) \quad (6.6)$$

where τ_1 and τ_2 are two independent lifetimes used to describe the background, f describes the fraction of candidates with lifetime τ_1 and the same acceptance parametrisation as in Equation 6.4 is used for describe the decay time efficiency $\epsilon(t)$. The lifetimes in the PDF are different, one describes a long-lived component and the other a short-lived component that are evident in the data.

Two fits are applied to $B \rightarrow h^+ h^-$ combinatorial background to find the decay time PDF parameters. The first fit determines f , τ_1 and τ_2 and the second is used to find a and n . The value of t_0 is fixed in the fit to improve fit stability. The biases

Parameter	Value
a / ps^{-1}	1.45 ± 0.12
n	1.92 ± 0.17
t_0 / ps	0.290
τ_1 / ps	17 ± 16
τ_2 / ps	1.3 ± 0.3
f	0.032 ± 0.027

Table 6.4 Parameters used to describe the background decay time distribution from combinatorial background decays in data passing the $B \rightarrow h^+ h^-$ selection.

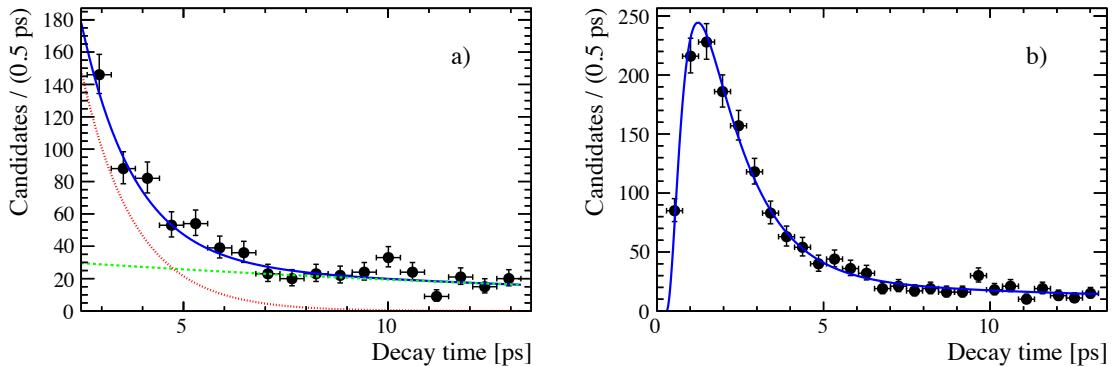


Fig. 6.10 Maximum likelihood fit to determine the lifetimes to describe the combinatorial background (a), with the long live component (green) and the short-lived component (red), and the fit to determine the acceptance function parameters (b) from the decay time distribution of combinatorial background decays in data passing the $B \rightarrow h^+ h^-$ selection requirements.

to the decay time distribution from the selection criteria have the greatest affect at low decay times, therefore the decay time acceptance is flat at large decay times. The lifetimes of the combinatorial background decays and the fraction of decays with each lifetime are determined from a fit using Equation 6.6, setting $\epsilon(t) = 1$, to candidates with a decay time above 2.5 ps. The acceptance function parameters are then determined from a fit to the full decay time range using Equation 6.6, where the lifetimes and the fraction of candidates with each lifetime are fixed. The results are shown in Figure 6.10 and the PDF parameters in Table 6.4.

This model for the background assumes that the decay time distribution of $B \rightarrow h^+ h^-$ candidates formed by random combinations of kaons and pions is the same as that of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates formed by randomly combining muons. There are too few candidates passing the $B_s^0 \rightarrow \mu^+ \mu^-$ selection to verify this assumption, although the validity of this model and the impact on the measured $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is investigated in Section 7.6.

6.4 Measurement strategy

The strategy to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is described in Section 6.1. The extremely rare nature of $B_s^0 \rightarrow \mu^+ \mu^-$ decays means that the stability and performance of the final fit will be highly dependent on the fit to the invariant mass distribution used to find the sWeights. Pseudoexperiments are performed to determine which mass range would produce an accurate fit with the smallest expected uncertainty on the measured effective lifetime for the dataset. The choice of the mass range determines which background sources need to be included into the fit.

The expected number of signal and background decays in data passing the $B_s^0 \rightarrow \mu^+ \mu^-$ selection in the mass range 4900 - 6000 MeV/ c^2 are used as the basis for the pseudoexperiments. The expected background yields are calculated using the same methods described in Section 5.3 but taking into account the looser particle identification requirement and the cut placed on the global BDT. The number of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are calculated using the normalisation factors in Section 5.4 and assuming the branching fraction values predicted by the SM. The expected yields are shown in Table 6.5.

The pseudoexperiments are performed by generating the mass and decay time distributions for the expected number of signal and background decays in the data. The number of decays generated in the pseudoexperiments are fluctuated using Poisson distributions for each decay type where the mean values are taken from the expected yields in Table 6.5. The PDFs described in Section 6.2 and 6.3 are used and the slope of the combinatorial background mass PDF is taken from simulated

Decay	Expected yield
$B_s^0 \rightarrow \mu^+ \mu^-$	30.94
$B^0 \rightarrow \mu^+ \mu^-$	3.27
$B \rightarrow h^+ h'^-$	9.68
$\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$	13.34
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	40.50
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	9.13
$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	6.01
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	4.86
$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	9.79
Combinatorial background	66.23
Total	193.75

Table 6.5 Number of expected decays in data passing the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime selection.

$b\bar{b} \rightarrow \mu^+ \mu^- X$ decays. The SM prediction for $\tau_{\mu\mu}$ is used to describe the lifetime of $B_s^0 \rightarrow \mu^+ \mu^-$ decays. The effective lifetime is then measured using the strategy described in Section 6.1, where the sWeights are computed from a fit to the invariant mass distribution and the lifetime is measured by a fit to the sWeighted decay time distribution. A series of different mass ranges and background components are tested. For each possible configuration a study containing 10,000 pseudoexperiments is performed and the performance of each configuration is evaluated using two of different metrics. The first, is the median expected uncertainty of the $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime. The median rather than the mean uncertainty is used due to the asymmetric spread of uncertainties observed for the expected statistics. The second measure, is the pull distributions of any free parameters in the fit, where the pull is defined as $(x - \mu)/\sigma$ with x the measured parameter value, μ the value used in the generation and σ the uncertainty on the measured parameter value. Ideally the pull distributions will be Gaussian in shape with a mean at 0 and a width of 1.

The details of the pseudoexperiments and mass fits tested are given in Section 6.4.2. However, first is a discussion of whether the lifetime or inverse lifetime should be measured given the expected number of decays present in the data set.

6.4.1 To fit for τ or τ^{-1} ?

During the development of the fit strategy, the pseudoexperiments produced biased pull distributions for the measured $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime no matter what mass fit configuration or acceptance function was used. The pull distribution for the effective lifetime, $\tau_{\mu\mu}$, is shown in Figure 6.11 for a simplified configuration where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/ c^2 . The distribution is not Gaussian in shape, and the bias was more pronounced in early stages of the analysis development which was performed assuming the expected signal and background yields of only the Run 1 data set.

The log-likelihood profile of the fit as a function of $\tau_{\mu\mu}$ reveals the cause of the biased pull distribution. For the simplified studies illustrated in Figure 6.11, the decay time is modelled by

$$N(t, \tau_{\mu\mu}) = N_0 e^{-t/\tau_{\mu\mu}}. \quad (6.7)$$

The likelihood profile as a function of decay time for this model is shown in Figure 6.11 and there is a clear discontinuity at the zero. The discontinuity arises because the value of $N(t, \tau)$ approaches zero as τ reduces in value until at the origin where $\tau = 0$ and $N(t, \tau)$ is undefined.

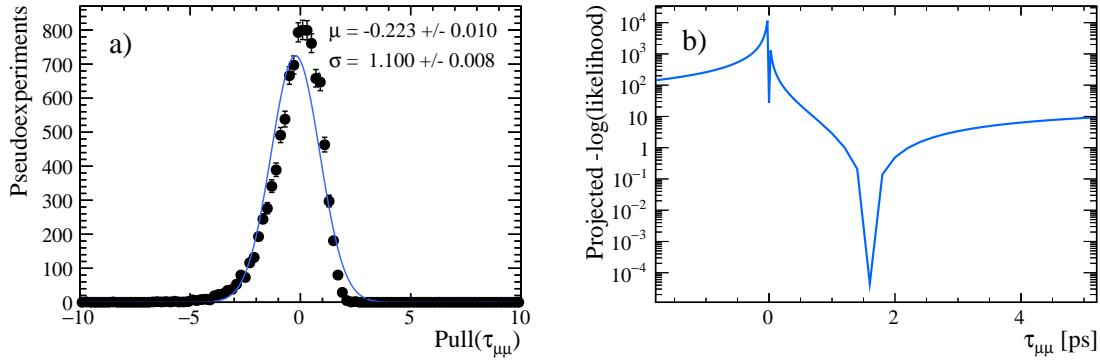


Fig. 6.11 Pull distribution (a) for $\tau_{\mu\mu}$ using a simplified configuration where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/ c^2 with the expected statistics for 4.4 fb $^{-1}$ of data used to measured the effective lifetime and the likelihood profile for $\tau_{\mu\mu}$ (b).

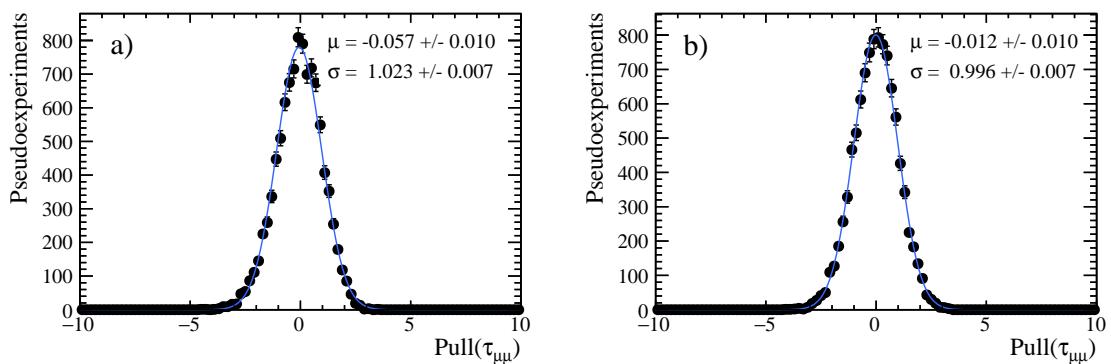


Fig. 6.12 Pull distribution for $\tau_{\mu\mu}$ using simplified mass and decay time model for 50 fb $^{-1}$ (a) and 300 fb $^{-1}$ (b) where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/ c^2 .

	$\tau_{\mu\mu}$	$\Gamma_{\mu\mu} \equiv \frac{1}{\tau_{\mu\mu}}$	Gaussian
1σ	$68.50 \pm 0.08\%$	$67.92 \pm 0.08\%$	68.27%
2σ	$93.44 \pm 0.10\%$	$95.91 \pm 0.10\%$	95.45%
3σ	$98.06 \pm 0.10\%$	$99.55 \pm 0.10\%$	99.73%

Table 6.6 Coverage of the statistical uncertainties evaluated as the number of pseudoexperiments for the expected number of decays with measured $\tau_{\mu\mu}$ ($\Gamma_{\mu\mu}$) values that are with 1, 2 and 3 times $\sigma_{\tau_{\mu\mu}}$ ($\sigma_{\Gamma_{\mu\mu}}$) of the generated lifetime value.

For the low number of $B_s^0 \rightarrow \mu^+ \mu^-$ decays expected in data the fitted value for $\tau_{\mu\mu}$ is only a few standard deviations away from this discontinuity, thereby biasing the estimation of the statistical uncertainty and changing the pull distribution from the expected Gaussian shape. However, as the number of expected signal and background decays increases, the $\tau_{\mu\mu}$ pull distributions become Gaussian in shape as shown in Figure 6.12. This is as expected because when the statistical uncertainty decreases, the discontinuity of Figure 6.11 is no longer within a few standard deviations of the measured $\tau_{\mu\mu}$.

The bias in the $\tau_{\mu\mu}$ pull distribution shows that the distribution cannot be interpreted in the usual way and also that the statistical uncertainties from the maximum likelihood to the weighted decay time distribution may not be correct.

Another way to assess the accuracy of the statistical uncertainties returned by the fit is the coverage of the uncertainties; the percentage of fitted $\tau_{\mu\mu}$ values from the pseudoexperiments that fall within 1, 2 and 3 standard deviations of the lifetime used to generate the pseudoexperiments. Table 6.6 shows the coverage of the statistical uncertainties for $\tau_{\mu\mu}$ from a set of 10,000 pseudoexperiments for the expected $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background yields with the data set alongside the intervals expected for a Gaussian distribution. The simple toy configuration used to produce the log-likelihood function is used. A comparison between the coverage of $\tau_{\mu\mu}$ and the Gaussian intervals shows that the coverage of the statistical uncertainties is very close to the expected values. Therefore the uncertainty on $\tau_{\mu\mu}$ is still reasonably estimated from the fit despite the poor pull distribution.

Alternatively, a biased pull distribution can be avoided by fitting for the inverse of the effective lifetime, $\tau_{\mu\mu}^{-1} \equiv \Gamma_{\mu\mu}$. The pull distributions for $\Gamma_{\mu\mu}$ are shown in Figure 6.13 and produce unbiased pull values regardless of the amount of data. This is unsurprising given the smooth log-likelihood profile as a function of $\Gamma_{\mu\mu}$ also shown in Figure 6.13. Furthermore, the statistical coverage of $\Gamma_{\mu\mu}$ is closer to the expected Gaussian coverage than the coverage of $\tau_{\mu\mu}$.

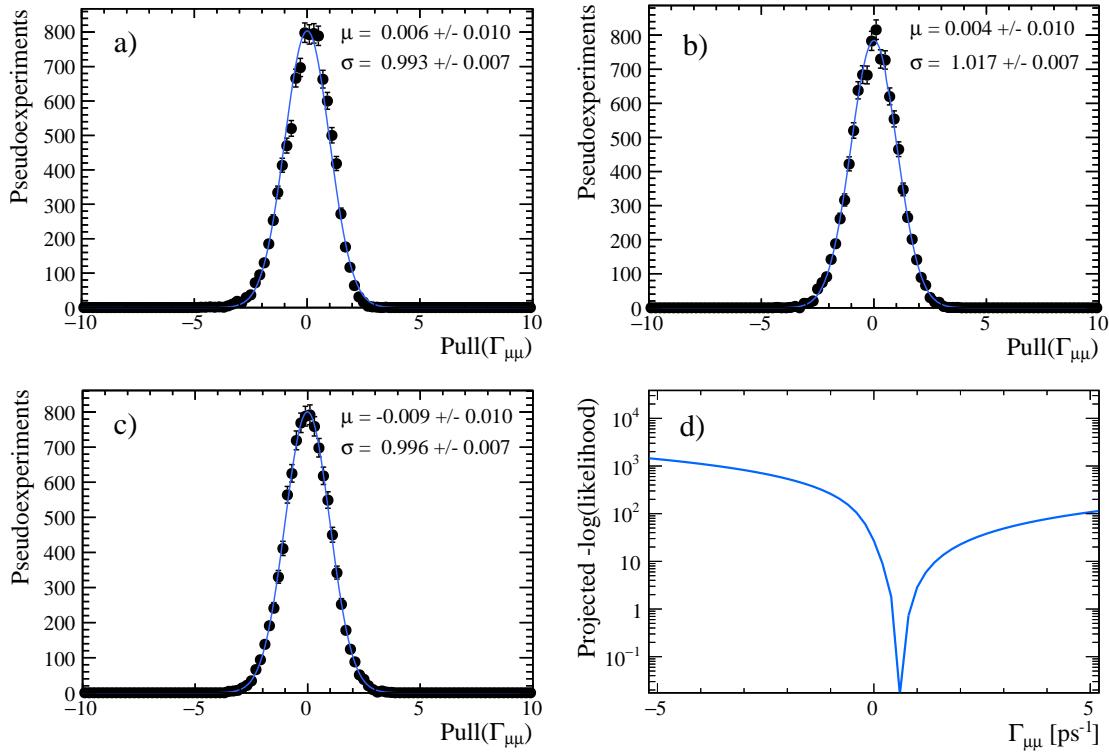


Fig. 6.13 Pull distribution for $\Gamma_{\mu\mu}$ using simplified pseudoexperiments for the Run 1 and Run 2 data set of 4.4 fb^{-1} (a) as well as 50 (b) and 300 (c) fb^{-1} and the likelihood profile as a function of $\Gamma_{\mu\mu}$ (d).

Ideally, the fit strategy would be performed to extract the lifetime not the inverse lifetime. However, for the moment the fit for both $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ is used in the pseudoexperiments. The statistical coverage for both parameters is good and using either will give a reasonable estimate of the statistical uncertainty on the measured value. The final decision is made based on the statistical coverage for the observed number of decays in the data set.

6.4.2 Optimisation of fit configuration

The mass distribution of the expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates passing the effective lifetime selection is shown in Figure 6.14 alongside the corresponding decay time distribution for one pseudoexperiment for the combined Run 1 and Run 2 data set. The contributions from the different signal and background sources are shown and the backgrounds beneath the B_s^0 mass peak are the combinatorial background and the tails of the $B \rightarrow h^+ h'^-$, $B^0 \rightarrow \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ backgrounds. The expected mass distribution is used to determine a range of mass fit configurations to be tested using toy studies to find the configuration that produces the smallest expected uncertainty on the measurement of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ provided the pull distributions of free parameters in the fit are accurate.

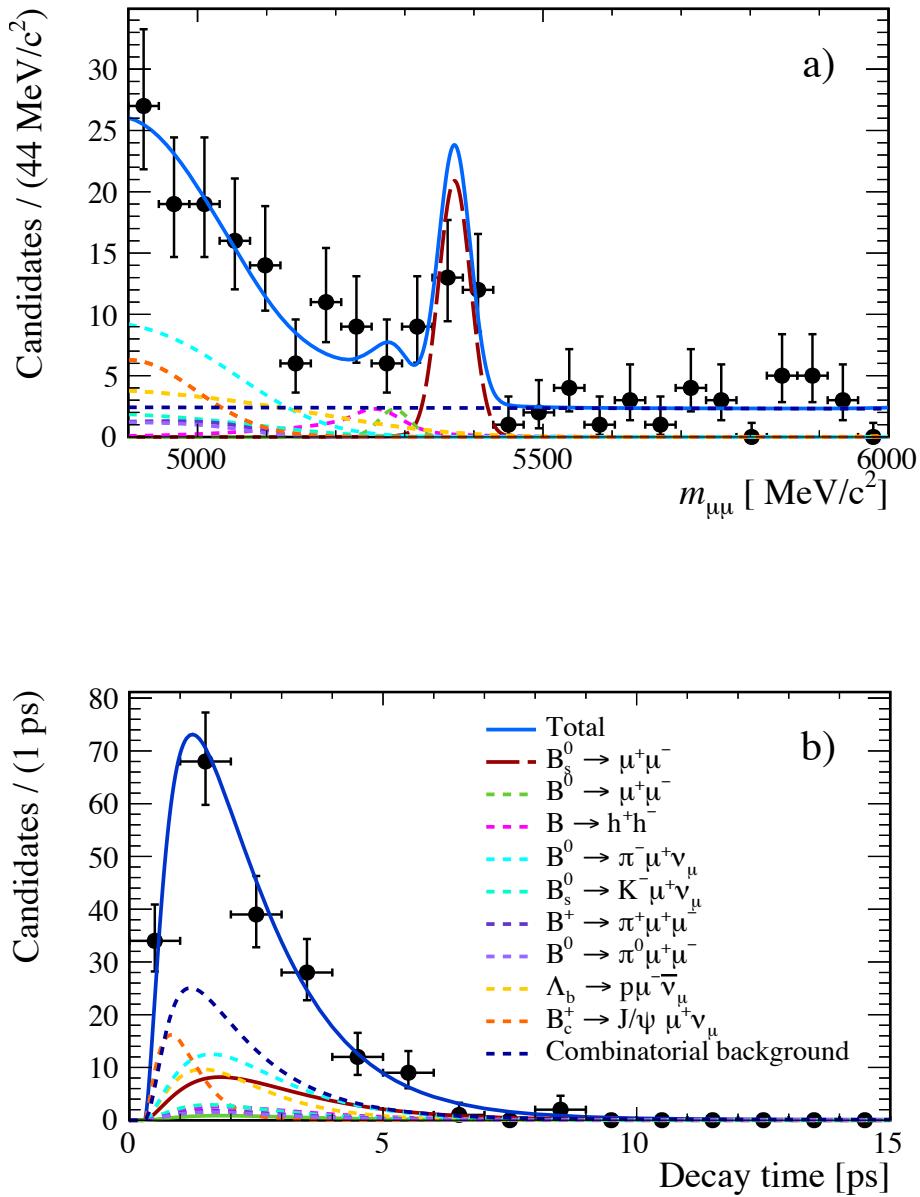


Fig. 6.14 Distributions of a) mass and b) decay time for the generated decays in the mass range 4900 - 6000 for one pseudoexperiment based on the expected number of signal and background decays in the data set.

Thirteen mass configurations are studied. In each mass fit configuration the mass PDF in Equation 6.2 is used and the mass ranges and backgrounds included in the PDF for the different configurations are given in Table 6.7.

For each possible mass fit configuration the $B_s^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$ and combinatorial background yields are left free in the fit whereas the yields of any other backgrounds are constrained to their expected values. The mass shapes of all components are fixed in the maximum likelihood fit except the slope of the combinatorial background because this is not accurately known in data. The SM predicts that $\tau_{\mu\mu}$ is equal to the lifetime of the heavy B_s^0 mass eigenstate, τ_H , the average value calculated by the Particle Data Group of $\tau_{\mu\mu} = 1.661$ ps, is used to generate events for the pseudoexperiments [5]. Also, regardless of which background components are included in the mass fit all backgrounds are generated for each mass range.

A total of 10,000 pseudoexperiments are performed for each mass configuration and the results are given in Table 6.8. The mean and widths of the pull distributions of $\Gamma_{\mu\mu}$, the $B_s^0 \rightarrow \mu^+\mu^-$ yield, the combinatorial background yield and the slope of the background mass PDF as well as the median expected uncertainties on $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are used to measure the performance of each mass fit configuration. The pull distribution of the fit for $\tau_{\mu\mu}$ is not used to assess the performance of each mass fit configuration given the discussion in Section 6.4.1.

The expected statistical uncertainties for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are smallest for fit configuration 11, where the mass range is restricted to 5320 - 6000 MeV/ c^2 and only the $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background components are used in the mass fit PDF. The mean and widths for the different pull distributions are consistent with the expected mean of 0 and width of 1 for this fit configuration. The larger mass ranges with more background components included in the mass PDF have larger uncertainties for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ as well as biased pull distributions.

Therefore, the fit configuration 11 is chosen to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. Figure 6.15 gives an example of the mass and decay time fits for the chosen configuration. Figure 6.14 shows that the number of background decays from $B^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h'^-$ and $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$ is extremely low above 5320 MeV/ c^2 , therefore these backgrounds do not need to be modelled in the mass PDF. The effect on the final result of not modelling these backgrounds is estimated in Section 7.2.

The expected uncertainties for the chosen fit configuration for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are $\sigma(\tau_{\mu\mu}) = 0.28$ ps and $\sigma(\Gamma_{\mu\mu}) = 0.11$ ps $^{-1}$, respectively. However, due to the low expected number of decays there is a large spread in the expected uncertainties as shown in Figure 6.16. Therefore, the uncertainties on the measurements would range between 0.1 - 0.8 ps for $\tau_{\mu\mu}$ and 0.07 - 0.2 ps $^{-1}$ for $\Gamma_{\mu\mu}$.

Fit no.	Mass Range /MeV/ c^2	Components included in the mass PDF	Yields free	Yields fixed
		$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
1.	4900 - 6000	$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$, $B^0 \rightarrow \pi^0\mu^+\mu^-$		✓
2.	4900 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
		$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$, $B_{(s)}^0 \rightarrow \pi^-(K^-)\mu^+\nu_\mu$, $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$		✓
3.	5150 - 6000	$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$, $B^0 \rightarrow \pi^0\mu^+\mu^-$		✓
4.	5150 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
		$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$, $B_{(s)}^0 \rightarrow \pi^-(K^-)\mu^+\nu_\mu$, $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$		✓
5.	5200 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
		$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$, $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$		✓
6.	5200 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
		$B \rightarrow h^{+-}$, $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$		✓
7.	5200 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
8.	5200 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
9.	5250 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
10.	5300 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
11.	5320 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
12.	5340 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	
13.	5350 - 6000	$B_s^0 \rightarrow \mu^+ \mu^-$, comb. bkg.	✓	

Table 6.7 Mass ranges and components included in the different mass fit configurations tested using pseudoexperiments. The final two columns indicate which components in the mass fits have fixed yields at the expected value and which are left free in the fit. In fit configurations 2 and 4 a single mass PDF is used to describe both $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and a single PDF is used to describe $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$ decays, all other configurations have one mass PDF per component. The shapes for all mass PDFs are fixed in the mass fit except the slope of the combinatorial background mass PDF.

Fit	$\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)$			$\mathcal{N}(\text{Comb. bkg.})$			Comb. bkg. slope	$\tau_{\mu\mu}$	$\Gamma_{\mu\mu}$	
	Mean	Width	Mean	Width	Mean	Width				
1.	-0.043	1.002	-0.048	1.050	-0.071	1.005	0.40	0.021	0.946	0.15
2.	-0.064	1.015	-0.018	1.048	-0.092	1.006	0.35	0.013	0.970	0.13
3.	-0.034	0.973	-0.067	1.023	-0.019	0.999	0.42	0.031	0.935	0.12
4.	-0.042	0.981	-0.066	1.024	-0.018	0.999	0.41	0.028	0.942	0.15
5.	-0.094	0.997	0.100	1.007	-0.228	1.018	0.40	0.017	0.933	0.40
6.	-0.124	1.024	0.110	1.009	-0.242	1.021	0.32	-0.008	0.973	0.12
7.	-0.367	1.045	1.248	0.923	-1.823	1.104	0.33	-0.091	0.975	0.12
8.	-0.521	1.049	1.770	0.983	-2.425	1.075	0.34	-0.114	0.969	0.12
9.	-0.473	1.044	1.296	0.918	-1.883	1.126	0.34	-0.126	-0.993	0.12
10.	-0.101	1.013	0.396	0.989	-0.571	1.068	0.31	-0.043	0.985	0.11
11.	0.050	1.006	0.060	1.013	-0.123	1.013	0.29	0.024	0.982	0.11
12.	0.020	1.009	0.015	1.007	-0.066	0.991	0.30	0.021	0.995	0.11
13.	0.007	1.001	-0.039	1.033	-0.029	0.995	0.34	0.023	0.983	0.13

Table 6.8 Results for the pseudoexperiments testing the mass fit configurations. The mean and width of the pull distributions for the $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background yields and the slope of the combinatorial background mass PDF are shown along with the expected statistical uncertainty on $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$. The uncertainties on the means are 0.010 and the widths are 0.007 for all configurations.

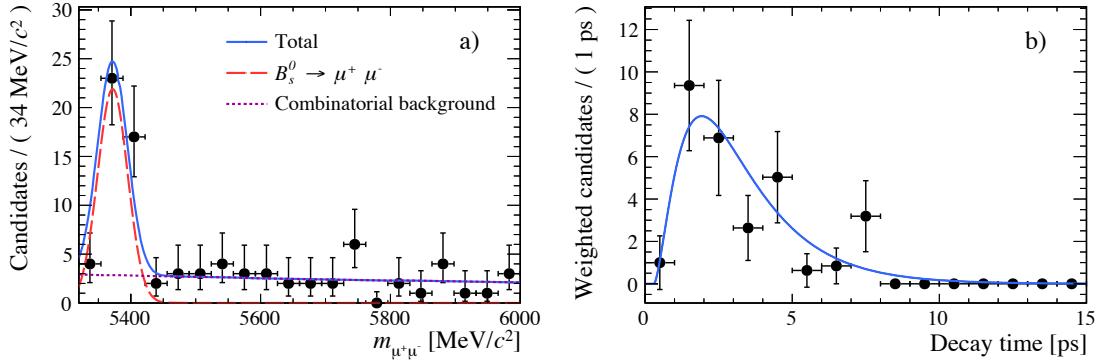


Fig. 6.15 Example of a) the mass and b) the decay time maximum likelihood fits for one pseudoexperiment using the chosen fit configuration where only components for $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background are modelled in the mass PDF.

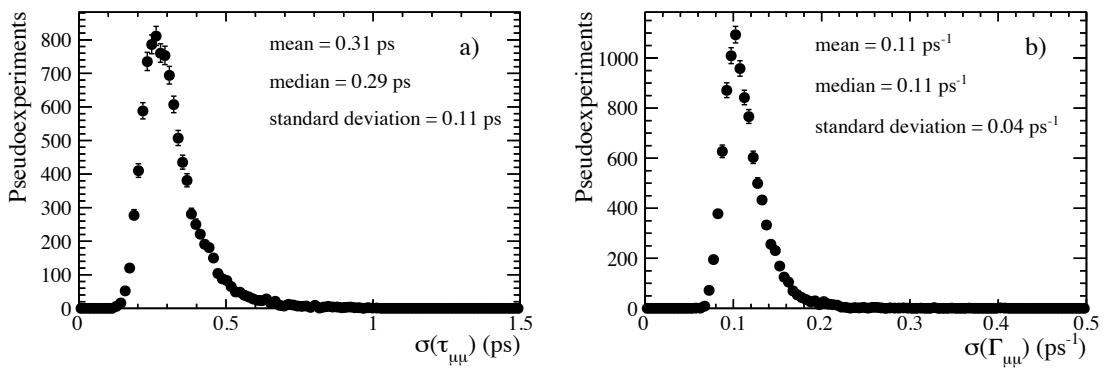


Fig. 6.16 Expected statistical uncertainties for a) $\tau_{\mu\mu}$ and b) $\Gamma_{\mu\mu}$ using fit configuration 11 in Table 6.7.

6.5 Results

The results of the unbinned maximum likelihood fit to the dimuon mass distribution and the sWeighted decay time of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates for Run 1 and Run 2 data are shown in Figure 6.17. The number of observed decays is 22 ± 6 $B_s^0 \rightarrow \mu^+ \mu^-$ decays and 20 ± 6 combinatorial background decays. The measured values of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are

$$\tau_{\mu\mu} = 2.04 \pm 0.44 \text{ ps} \quad (6.8)$$

$$\Gamma_{\mu\mu} = 0.489 \pm 0.117 \text{ ps}^{-1} \quad (6.9)$$

where the uncertainties are statistical only. The results are consistent with the SM prediction given in Chapter 2 and the measured value is 1σ from $A_{\Delta\Gamma} = +1$ and 1.4σ from $A_{\Delta\Gamma} = -1$.

The observed number of signal and background decays are lower than the expected yields given in Table 6.5 for the data. Therefore, it is important to check whether the statistical coverage for $\tau_{\mu\mu}$ uncertainties is still good. The pseudoexperiments in Section 6.4.1 are repeated, this time using the observed number of $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays and the results are shown in Table 6.9. The statistical coverage of the $\tau_{\mu\mu}$ uncertainty is good and therefore $\tau_{\mu\mu}$ and its statistical uncertainty can be trusted as accurate and the measurement can be quoted in terms of $\tau_{\mu\mu}$.

	$\tau_{\mu\mu}$	$\Gamma_{\mu\mu}$	Gaussian
1σ	$68.83 \pm 0.08\%$	$67.76 \pm 0.08\%$	68.27%
2σ	$93.11 \pm 0.10\%$	$95.55 \pm 0.10\%$	95.45%
3σ	$97.92 \pm 0.10\%$	$99.67 \pm 0.10\%$	99.73 %

Table 6.9 Coverage of the statistical uncertainties evaluated as the number of pseudoexperiments using the observed number of decays with measured $\tau_{\mu\mu}$ ($\Gamma_{\mu\mu}$) values that are with 1, 2 and 3 times $\sigma_{\tau_{\mu\mu}}$ ($\sigma_{\Gamma_{\mu\mu}}$) of the generated lifetime value.

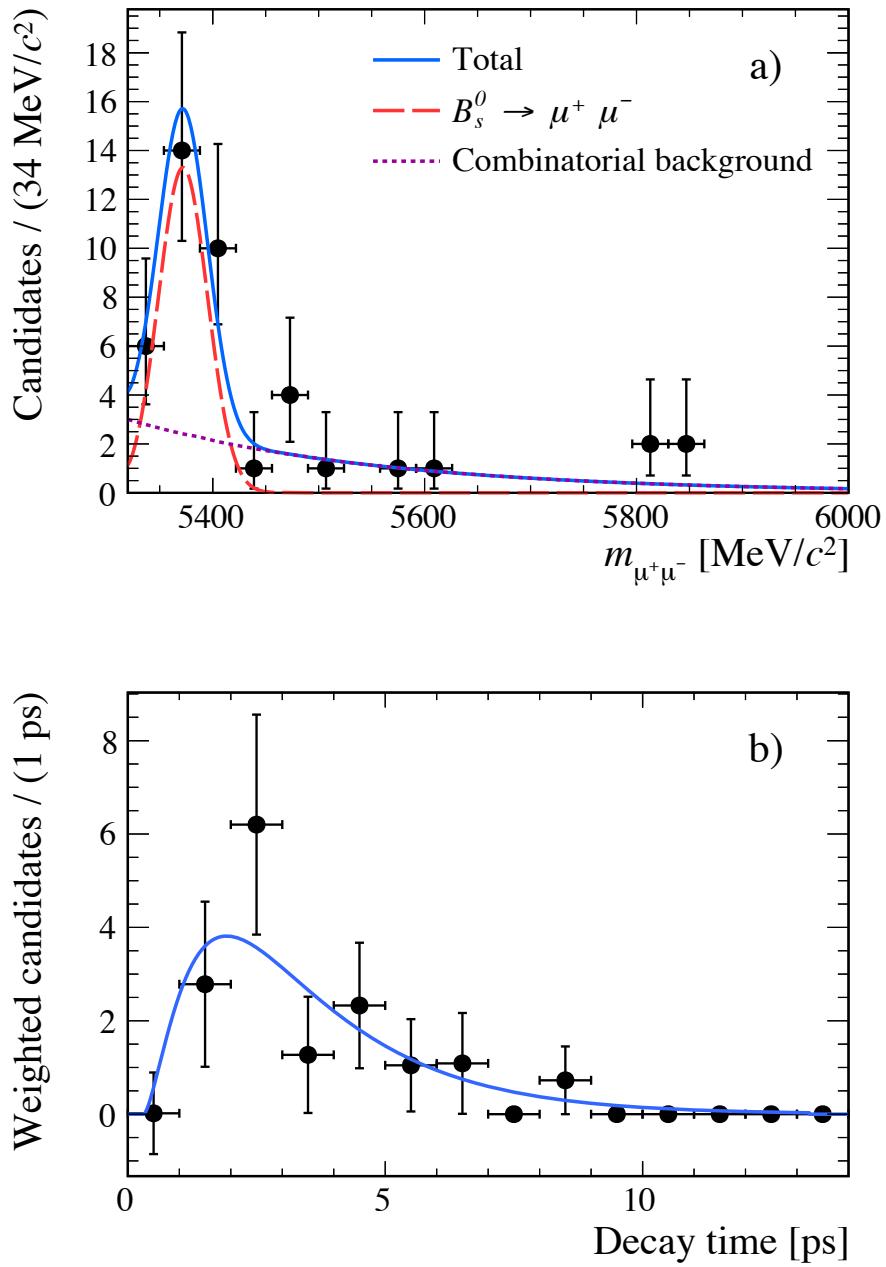


Fig. 6.17 Maximum likelihood fit to a) the invariant mass distribution and b) weighted decay time distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates using an integrated luminosity of 4.4 fb^{-1} of data collected by the LHCb experiment.

Chapter 7

Systematic uncertainties and cross checks on the effective lifetime

The measured $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime presented in Chapter 6 is influenced by various systematic biases arising from different areas of the analysis procedure. In this chapter the sizes of the different systematic uncertainties are estimated and several cross checks are made on the measurement strategy to ensure the uncertainty quoted on the final result is correct. The total systematic uncertainty on the measurement of $\tau_{\mu\mu}$ is given at the end of the chapter.

As discussed in Section 6.4.1, pseudoexperiments showed that the pull distributions of for fitting for $\tau_{\mu\mu}$ are biased. Therefore the pull distributions of $\Gamma_{\mu\mu}$ are used, when needed, as a measure of the fit performance instead.

7.1 Accuracy of the fit

The final configuration used to measure the effective lifetime was chosen using pseudoexperiments for the expected number of decays and by optimising two different figures of merit: the mean and width of the pull distributions of free parameters in the fit; and the expected uncertainties on $\tau_{\mu\mu}$. To test the fit accuracy, the values for the figures of merit for a set of 10,000 pseudoexperiments using the final fit configuration and assuming the SM $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions are given in Table 7.1. The same set up as described in Section 6.4 is used but in these pseudoexperiments only $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays are generated, all other backgrounds are ignored. Based on these results several aspects of the fit deserve further investigation. This includes the stability of the fit with different $\tau_{\mu\mu}$ values, the slightly biased pull distribution for $B_s^0 \rightarrow \mu^+\mu^-$ yields and the overall bias in the measured value of $\tau_{\mu\mu}$.

τ	$\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)$		$\mathcal{N}(\text{Comb.})$		Comb. slope		$\Gamma_{\mu\mu}$	
	Mean	Width	Mean	Width	Mean	Width	Mean	Width
τ_H	-0.097	1.020	-0.062	1.030	-0.013	0.992	-0.000	0.993
τ_L	-0.098	1.019	-0.062	1.018	-0.003	0.987	-0.010	1.018
$\tau_{B_s^0}$	-0.102	1.017	-0.059	1.032	-0.027	0.996	0.001	0.989

Table 7.1 Results from 10,000 pseudoexperiments using the final fit configuration for the expected number of decays and using as the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime; the SM predication (τ_H); the lifetime of the light B_s^0 mass eigenstate (τ_L); and the mean lifetime of the B_s^0 ($\tau_{B_s^0}$). The mean and width of the pull distributions for $\Gamma_{\mu\mu}$, the $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background yields and the slope of the combinatorial background mass PDF are shown. The uncertainties on the means are 0.010 and widths are 0.007 for all configurations.

7.1.1 Fit stability with $\tau_{\mu\mu}$ values

The pull distribution for $\Gamma_{\mu\mu}$ and the coverage of the statistical uncertainties given in Section 6.4 show that the fit gives a good estimation of $\tau_{\mu\mu}$ for the expected number of decays. However, it is necessary to understand if this is due to accurate background subtraction by the sPlot method or if it could be resulting from similarities between the decay time distributions of $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays. As a test, pseudoexperiments are performed for a range of generated $B_s^0 \rightarrow \mu^+ \mu^-$ lifetimes different to the SM prediction. Only $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays are generated so that the small contamination from mis-identified backgrounds does not mask the effects of using different lifetimes. The bias arising from the small contribution of mis-identified backgrounds in the mass range of the fit is evaluated in Section 7.2.

The results of 10,000 pseudoexperiments are shown in Table 7.1 and the different lifetimes return accurate pull distributions for the fitted $\Gamma_{\mu\mu}$ values with means and widths consistent with 0 and 1, respectively. Therefore, the fit returns an accurate measured value for the $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime independent of the lifetime chosen for the B_s^0 .

7.1.2 $B_s^0 \rightarrow \mu^+ \mu^-$ yield estimation

The pull distributions for $B_s^0 \rightarrow \mu^+ \mu^-$ yields in Table 7.1 have slightly biased mean values of ~ 0.010 ps. implying that the mass fit does not accurately estimate the $B_s^0 \rightarrow \mu^+ \mu^-$ yield. However, the pull distribution for $\Gamma_{\mu\mu}$ is accurate therefore this bias in the $B_s^0 \rightarrow \mu^+ \mu^-$ yield could originate from a different source.

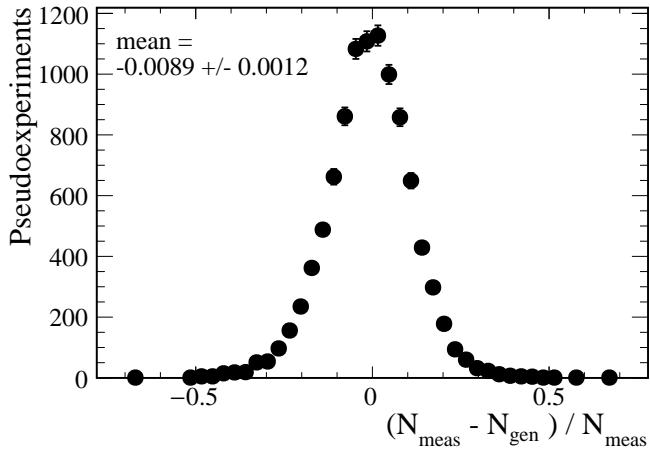


Fig. 7.1 Fractional bias of the measured $B_s^0 \rightarrow \mu^+\mu^-$ yield from pseudoexperiments for the expected number of decays with 4.4 fb^{-1} . Only $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays are included in the pseudoexperiments.

In the pseudoexperiments the number of expected decays in the mass range 4900 - 6000 MeV/ c^2 are given in Table 6.5. These numbers are used as the basis for the pseudoexperiments. However, the number of decays generated is fluctuated for each study about the expected value using a Poisson distribution. Therefore the number of decays generated, N_{gen} , is different to the expected number. This enables an extended maximum likelihood fit to be used to fit the mass distribution where the total number of events is a free parameter. To achieve a Gaussian shaped pull distribution of the measured $B_s^0 \rightarrow \mu^+\mu^-$ yields, the uncertainties on the measured yields must be distributed according to a Gaussian function. This will be true when there are a large number of $B_s^0 \rightarrow \mu^+\mu^-$ decays, in the high statistics limit, where a Poisson distribution is a good approximation of a Gaussian distribution. However, the current data contains a small number of $B_s^0 \rightarrow \mu^+\mu^-$ decays. The uncertainty on the measured yields is proportional to $\sqrt{N_{gen}}$, where N_{gen} is the number of generated decays, and does not have a Gaussian distribution. This effect would shift the mean value of the pull distribution but not lead to an incorrect estimation of the $B_s^0 \rightarrow \mu^+\mu^-$ yield. The fractional bias, $(N_{meas} - N_{gen})/N_{gen}$, where N_{meas} is the measured $B_s^0 \rightarrow \mu^+\mu^-$ yield is shown in Figure 7.1 and supports this explanation by producing a mean consistent at zero, with a negligible bias of 0.8 %.

Furthermore, the pull distributions for $B_s^0 \rightarrow \mu^+\mu^-$ yields for pseudoexperiments with higher statistics produce means that tend towards zero as the number of decays increases as shown in Figure 7.3. Therefore the mass fit returns accurate yields for $B_s^0 \rightarrow \mu^+\mu^-$ and the biased pull distribution arises from the low statistics of the data set. The same reasoning can be applied to the pull distribution of the yields of combinatorial background decays that have a slightly less biased mean value of ~ 0.006 compared with the $B_s^0 \rightarrow \mu^+\mu^-$ yields.

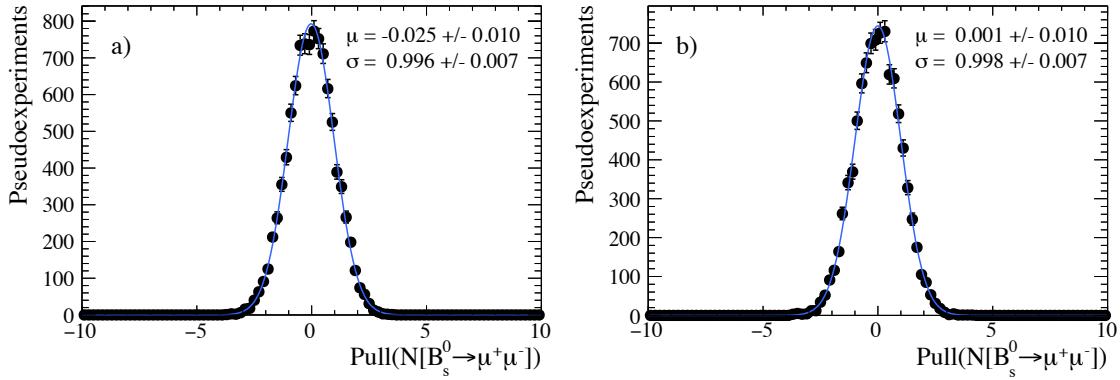


Fig. 7.2 Pull distribution for $B_s^0 \rightarrow \mu^+\mu^-$ measured yields from 10,000 pseudoexperiments for the expected number of decays in a) 50 fb^{-1} and b) 300 fb^{-1} . Only $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays are included in the pseudoexperiments.

7.1.3 Overall bias on $\tau_{\mu\mu}$

The remaining area of the fit to investigate is any underlying bias from the fit on the measured value of $\tau_{\mu\mu}$.

The overall bias in the fit for measuring $\tau_{\mu\mu}$ is evaluated from the difference between the measured and generated values of $\tau_{\mu\mu}$ from pseudoexperiments. A total of 10,000 pseudoexperiments are performed generating only $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays. The difference between the measured and generated $\tau_{\mu\mu}$ values is evaluated for pseudoexperiments generated with the expected and also the observed number of $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays. The fit bias is evaluated for the observed number of decays as well as the expected number of decays because there are fewer than expected therefore the bias may be larger. The resulting distributions are shown in Figure 7.3. The mean of the difference in $\tau_{\mu\mu}$ values is 0.02 ps for the expected number of decays and 0.03 ps for the observed number of decays. Therefore the larger uncertainty of 0.03 ps is used as the measured of the systematic uncertainty caused by the fit on the final result for $\tau_{\mu\mu}$.

7.2 Background contamination

The mass fit configuration used to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime includes components for $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial background decays to fit candidates with a B_s^0 mass between $5320 - 6000 \text{ MeV}/c^2$. Although the majority of background decays from mis-identified and $B^0 \rightarrow \mu^+\mu^-$ decays fall outside this mass window, as shown in Figure 6.14, the tails of some backgrounds are still within this region. The backgrounds of particular importance for the chosen mass range

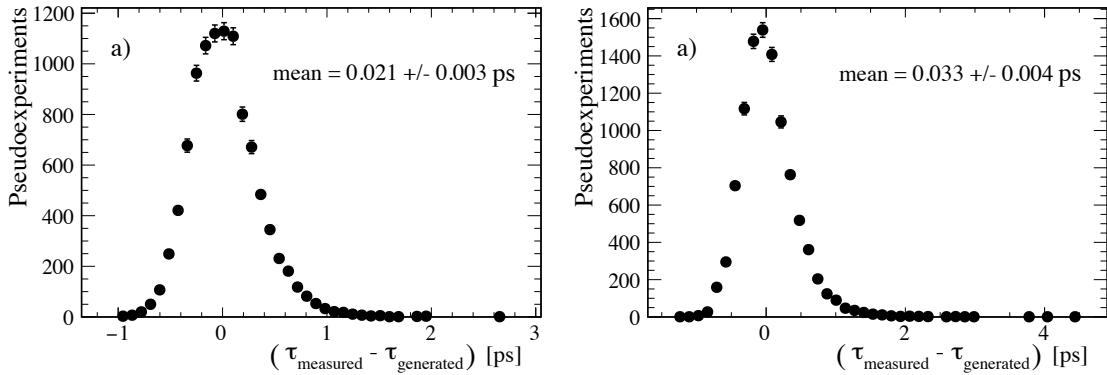


Fig. 7.3 Overall bias in $\tau_{\mu\mu}$, evaluated as the difference between the measured, τ_{measured} , and the generated, $\tau_{\text{generated}}$, lifetimes for pseudoexperiments using a) the expected and b) the observed $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays.

include $B^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$, $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$.

The number of expected background decays and their mass PDFs in the range 4900 - 6000 MeV/ c^2 are computed using the methods described in Chapter 5 and the number of decays expected in the smaller mass range 5320 - 6000 MeV/ c^2 are determined by integrating the mass PDFs over this mass range. The expected yields for each background are given in Table 7.2 and are less than 1 candidate, except the combinatorial background, in the range 5320 - 6000 MeV/ c^2 .

The impact of backgrounds not modelled in the mass fit on the measured $\tau_{\mu\mu}$ value is evaluated using two sets of pseudoexperiments. The pseudoexperiments have the same general set up as described in Section 6.4. One set of pseudoexperiments assumes there are no backgrounds other than the combinatorial background and therefore only $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background candidates are generated. The second set of pseudoexperiments uses a more realistic model and generates all possible background decays. The expected yields are fluctuated using a Poisson distribution around their expected values to two decimal places. For each configuration 10,000 pseudoexperiments are performed and the pull distributions for $\Gamma_{\mu\mu}$ of each set up are compared.

The inclusion of all the background decays causes a shift in the mean of the $\Gamma_{\mu\mu}$ pull distribution of 0.024 ps^{-1} as shown in Figure 7.4. Therefore, assuming the expected uncertainty in Section 6.4.2 for $\tau_{\mu\mu}$, the systematic shift from not including all backgrounds in the fit configuration is 0.007 ps.

The expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays assumes the SM branching fractions. However, the measured values for the branching fractions are slightly different from the SM predictions. Using the measured values from the combined analysis of the Run 1 CMS and LHCb data [47], the expected number of

Decay	Expected yields in mass ranges	
	4900 - 6000 MeV/ c^2	5320 - 6000 MeV/ c^2
$B_s^0 \rightarrow \mu^+ \mu^-$	30.94	30.47
$B^0 \rightarrow \mu^+ \mu^-$	3.27	0.20
$B \rightarrow h^+ h'^-$	9.68	0.92
$\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$	13.34	0.64
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	40.50	0.06
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	9.13	0.03
$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	6.01	0.00
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	4.86	0.60
$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	9.79	0.00
Combinatorial background	66.2	40.6

Table 7.2 Number of expected signal and background decays in data passing the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime selection in the mass ranges 4900 - 6000 MeV/ c^2 and 5320 - 6000 MeV/ c^2 .

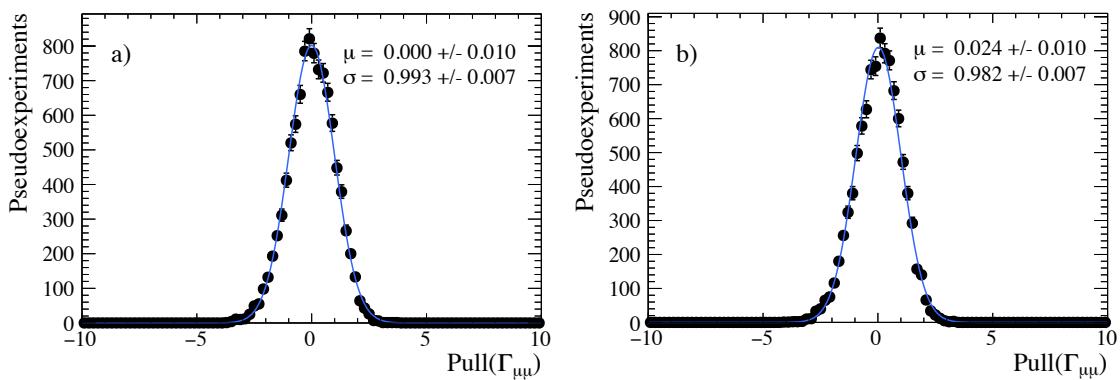


Fig. 7.4 Pull distribution for $\Gamma_{\mu\mu}$ from 10,000 pseudoexperiments, using background sources of a) combinatorial background only and b) combinatorial background, mis-identified decays and $B^0 \rightarrow \mu^+ \mu^-$ decays.

Decay	Expected yield in 5320 - 6000 MeV/ c^2	
	SM	World average
$B_s^0 \rightarrow \mu^+ \mu^-$	30.47	22.50
$B^0 \rightarrow \mu^+ \mu^-$	0.20	0.73

Table 7.3 Number of expected decays in data passing the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime selection in the mass range 5320 - 6000 MeV/ c^2 using the SM predictions and branching fraction measurements from the combined analysis of the Run 1 CMS and LHCb data [47].

$B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays decreases and increases, respectively, compared to the SM expectations. The changes in the yields are given in Table 7.3. The pseudoexperiments were repeated with the measured branching fractions but the shift in the mean of the pull distribution was smaller. Therefore the larger value from the SM predictions is used as the systematic uncertainty for the background contamination.

7.3 Mass PDF parameters

The data collected in Run 1 and Run 2 are combined for the measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime and the mass and decay time fits are applied to the combined data. However, the parameters used in the mass PDF in Table 5.1 were evaluated specifically for Run 1 data and different parameters are available for Run 2 data. Therefore, the influence of the choice of mass PDF parameters on the measured $\tau_{\mu\mu}$ value must be evaluated.

Once again pseudoexperiments are performed to understand the size of the impact of the mass model choice on the effective lifetime measurement. Only $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays are generated to separate mass PDF effects from the contamination of mis-identified backgrounds and $B^0 \rightarrow \mu^+ \mu^-$ decays. $B_s^0 \rightarrow \mu^+ \mu^-$ candidates are generated using the Run 1 parameters in Table 5.1 but the mass fit is performed using the Run 2 parameters in Table 5.2. The pull distribution for 10,000 pseudoexperiments for $\Gamma_{\mu\mu}$ from this configuration is compared with those from pseudoexperiments where Run 1 parameters are used to generate and fit the mass distribution. The change in the measured lifetime with the mass PDF parameters is negligible as shown in Figure 7.5. Therefore, no systematic uncertainty is assigned.

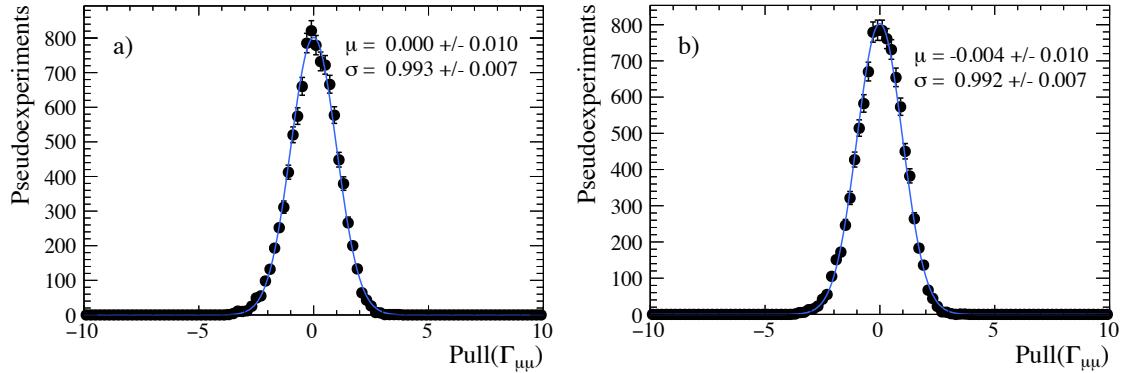


Fig. 7.5 Pull distribution for $\Gamma_{\mu\mu}$ from 10,000 pseudoexperiments where the $B_s^0 \rightarrow \mu^+\mu^-$ mass distribution is generated using the Run 1 parameters and the mass fit is performed a) Run 1 parameters and b) Run 2 parameters.

7.4 Acceptance function accuracy

The $B_s^0 \rightarrow \mu^+\mu^-$ decay time acceptance function is determined from weighted simulated decays, as described in Section 6.3.1. It relies on the assumption that weighted simulated decays model the data reasonably well and weights taken from the number of tracks in $B^0 \rightarrow K^+\pi^-$ decays can be used for $B_s^0 \rightarrow \mu^+\mu^-$ decays. To test this assumption, as well as the strategy used to measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime, the lifetimes of the more abundant $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays are measured using the same approach as the $B_s^0 \rightarrow \mu^+\mu^-$ lifetime. The measurement of the $B^0 \rightarrow K^+\pi^-$ lifetime is used to define a systematic uncertainty associated with the accuracy of the acceptance function and $B_s^0 \rightarrow K^+K^-$ decays are used as a cross-check to ensure the weights computed from $B^0 \rightarrow K^+\pi^-$ decays can be used to measure the lifetime of other decays as well.

The selection requirements used to identify $B^0 \rightarrow K^+\pi^-$ decays in 2011, 2012, 2015 and 2016 data are detailed in Chapter 4 and are kept very similar to the $B_s^0 \rightarrow \mu^+\mu^-$ selection. Although considerably reducing the statistics, all candidates are required to be triggered as TIS, in order to keep the $B \rightarrow h^+h^-$ trigger efficiency similar to that of $B_s^0 \rightarrow \mu^+\mu^-$ decays. The $DLL_{K\pi}$ variable is used to separate $B^0 \rightarrow K^+\pi^-$ decays from other $B \rightarrow h^+h^-$ decays and candidates are reconstructed with the daughter with the highest $DLL_{K\pi}$ value assigned the kaon mass hypothesis and the daughter particle with the lowest $DLL_{K\pi}$ value the pion mass hypothesis.

The measurement of the lifetime is performed in the same way as the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement and all years of data are combined into one data set. An extended unbinned maximum likelihood fit is performed to the $B^0 \rightarrow K^+\pi^-$ mass distribution. Components for $B^0 \rightarrow K^+\pi^-$, $B_s^0 \rightarrow K^+\pi^-$ and combinatorial background decays are included in the mass fit. Both B -meson decays are modelled by Crystal Ball functions and combinatorial background decays are modelled by an

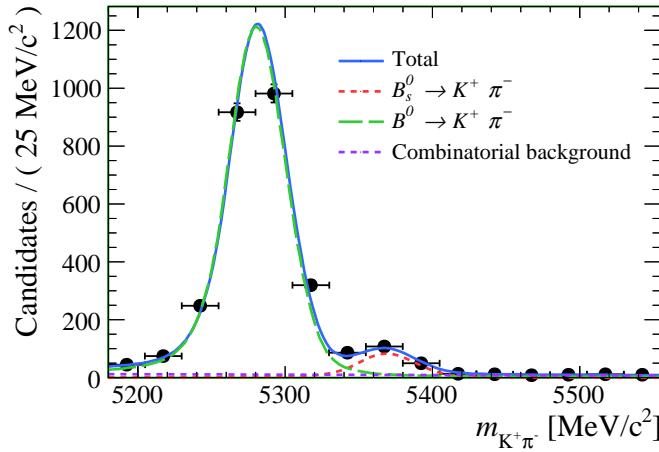


Fig. 7.6 Maximum likelihood fit to the mass distribution of $B^0 \rightarrow K^+\pi^-$ decays for data taken in 2011, 2012, 2015 and 2016.

exponential function. The mass fit, shown in Figure 7.6, is used to calculate sWeights that are re-normalised using Equation 6.1. The lifetime, $\tau_{K\pi}$, is measured from the sWeighted decay time distribution.

The decay time PDF has the same form as the one used for the $B_s^0 \rightarrow \mu^+\mu^-$ decays in Equation 6.3. The acceptance parameters are found from a fit to weighted $B^0 \rightarrow K^+\pi^-$ simulated decays using the same method described in Section 6.3.1 with the number of tracks in the event weighted using the same weights as $B_s^0 \rightarrow \mu^+\mu^-$ decays. However, the weights applied to combine simulated $B^0 \rightarrow K^+\pi^-$ decays from each year are not dependent on $B_s^0 \rightarrow J/\psi\phi$ decays. Since $B^0 \rightarrow K^+\pi^-$ decays have a high yield in data, mass fits to each year are used to find the yields and combine the simulated decays from each year. The same mass PDF used to fit the combined mass distribution is applied to each year. The weights used to combine simulated decays in different years are

$$\omega_i = \frac{Y_i^{data}}{\sum_j Y_j^{data}} \cdot \frac{\sum_k N_k^{K\pi}}{N_i^{K\pi}}, \quad (7.1)$$

where Y_i^{data} are the $B^0 \rightarrow K^+\pi^-$ yields in each year of data and $N_i^{K\pi}$ the number of simulated decays available for each year. The acceptance function fit is shown in Figure 7.7 and for consistency the same simulation versions are used for $B^0 \rightarrow K^+\pi^-$ decays as $B_s^0 \rightarrow \mu^+\mu^-$ decays.

The results from the decay time fit are shown in Figure 7.8 and the measured $B^0 \rightarrow K^+\pi^-$ lifetime is

$$\tau_{K\pi} = 1.52 \pm 0.03 \text{ ps}, \quad (7.2)$$

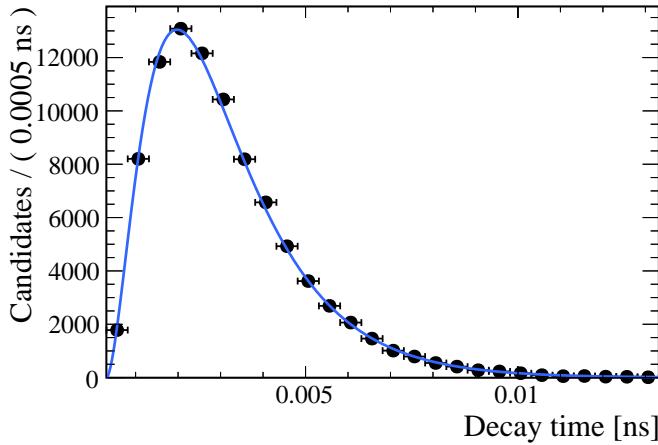


Fig. 7.7 Decay time distribution in weighted 2011, 2012, 2015 and 2016 simulated decays and the maximum likelihood fit results to determine the acceptance function parameters.

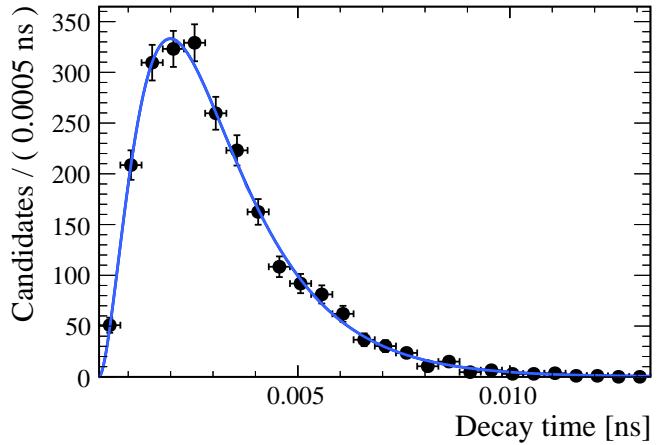


Fig. 7.8 Maximum likelihood fit to the signal weighted decay time distribution of $B^0 \rightarrow K^+\pi^-$ decays for data taken in 2011, 2012, 2015 and 2016.

where only the statistical uncertainty is given. The measured results are consistent with the expected value of 1.520 ± 0.004 ps [5] and the measurement strategy used to find the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime has been shown to work. The statistical uncertainty of the measured $B^0 \rightarrow K^+\pi^-$ decay time is assigned as a systematic uncertainty to provide a measure of how well the acceptance function can be determined from weighted simulated decays for measuring $\tau_{\mu\mu}$.

The $B_s^0 \rightarrow K^+K^-$ lifetime is measured using 2012 and 2015 data with $B_s^0 \rightarrow K^+K^-$ candidates identified using the selection requirements in Chapter 4. The same measurement strategy is used to measure the $B_s^0 \rightarrow K^+K^-$ lifetime as used for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. Only 2012 and 2015 data are used due to the available simulation versions of simulated $B_s^0 \rightarrow K^+K^-$ decays. Once again TIS triggers are used to keep a relatively lifetime unbiased trigger efficiency and candidates are reconstructed assuming both daughters are kaons. The mass PDF

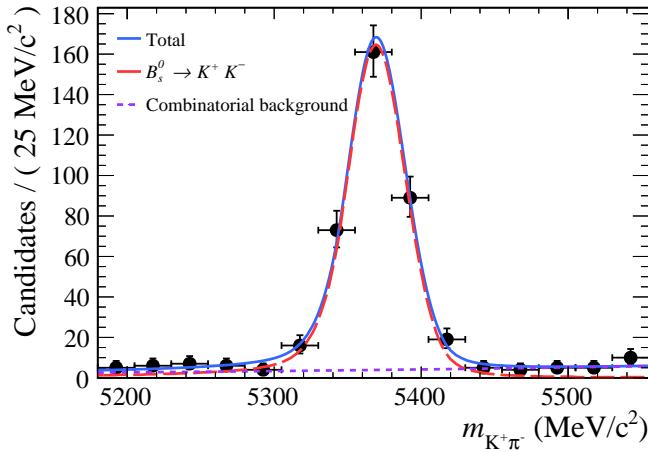


Fig. 7.9 Maximum likelihood fit to the mass distribution of $B_s^0 \rightarrow K^+K^-$ decays for data taken in 2012 and 2015.

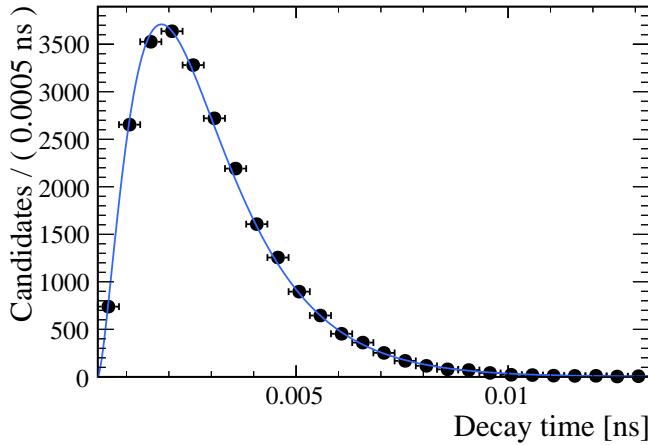


Fig. 7.10 Decay time distribution in weighted 2012 and 2015 simulated decays and the fit results to determine the acceptance function parameters.

includes $B_s^0 \rightarrow K^+K^-$ and combinatorial background decays and the same PDF is used for $B_s^0 \rightarrow K^+K^-$ as for $B_s^0 \rightarrow K^+\pi^-$. The unbinned extended maximum likelihood fit used to extract the sWeights is shown in Figure 7.9.

The $B_s^0 \rightarrow K^+K^-$ acceptance is found using the same method as $B_s^0 \rightarrow \mu^+\mu^-$ with $B_s^0 \rightarrow J/\psi\phi$ decays used to determine the relative proportions of decays in each year of data. Figure 7.10 shows the acceptance fit and the results of the decay time fit is shown in Figure 7.11. The measured values for the lifetime, τ_{KK} , is

$$\tau_{KK} = 1.39 \pm 0.06 \text{ ps}, \quad (7.3)$$

where only the statistical uncertainty is given. The measured value is consistent with the predicted value of 1.395 ± 0.020 ps [167] and shows that $B^0 \rightarrow K^+\pi^-$ weights can be used for other decays as well as $B^0 \rightarrow K^+\pi^-$.

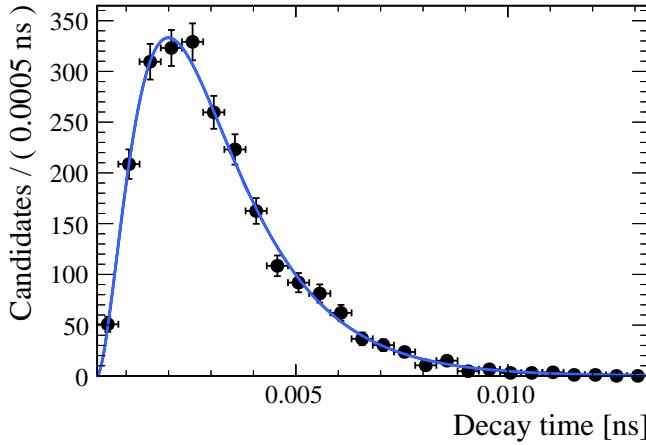


Fig. 7.11 Maximum likelihood fit to the signal weighted decay time distribution of $B_s^0 \rightarrow K^+ K^-$ decays for 2012 and 2015 data.

7.5 Incorrectly assigned primary vertices and the detector resolution

Measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime accurately relies on the B_s^0 candidate being assigned to the correct primary vertex in the event; an incorrect assignment would lead to the wrong value for the B_s^0 decay time. In references [168, 169] that study the lifetimes of $B \rightarrow J/\psi X$ decays at LHCb, a component is included into the decay time PDF to model the number of incorrectly assigned primary vertices (PVs) as well as the resolution of the detector. The decay time PDF is convoluted by the sum of three Gaussian functions; two narrow Gaussian functions model the detector resolution and a third wider Gaussian function corresponds to < 1% of decays assigned incorrect PVs. The decay time fit to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime does not explicitly model incorrectly assigned PVs or the detector resolution, although these effects will, to some degree, be included into the acceptance function.

A similar model to references [168, 169] is used to check the affect on the measured lifetime of decays with incorrectly assigned PVs and detector resolution effects that are not included in the acceptance function. A set of 1 million decays are generated using the decay time model

$$\epsilon(t)[\mathcal{R}(t) \otimes e^{-t/\tau}], \quad (7.4)$$

where $\epsilon(t)$ is the acceptance function with parameters given in Table 6.3 and $\mathcal{R}(t)$ is the resolution function composed of 3 Gaussian functions. Decays are generated assuming the SM prediction of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime of $\tau_{\mu\mu} = 1.610$ ps. A fit is then performed to the generated decay time distribution but the resolution term is no longer included in the PDF. The measured $\tau_{\mu\mu}$ value is compared to the value used to generate the decays.

Parameter	Fit value
μ (ps)	0.00063 ± 0.00005
$\sigma_1(ps)$	5.62 ± 0.07
f_1	0.006
$\sigma_2(ps)$	0.0573 ± 0.0003
f_2	0.313
$\sigma_3(ps)$	0.0294 ± 0.0001
f_3	0.681

Table 7.4 Parameters from the fit to the difference between the reconstructed decay time and the true decay time for simulated decays that pass the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime selection. The mean used for all Gaussian is μ and σ_i are the widths of each Gaussian that make up a fraction f_i of the total sum.

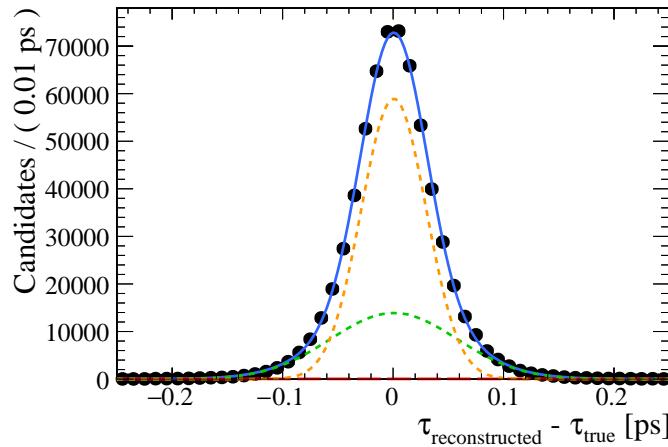


Fig. 7.12 Fit result to the difference between the reconstructed decay time and the true decay time for simulated decays that pass the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime selection.

The resolution function is determined from weighted simulated $B_s^0 \rightarrow \mu^+\mu^-$ decays that were used to compute the acceptance function in Section 6.3.1. The difference between the reconstructed decay time and the ‘true’ decay time that each decay is generated with is evaluated for decays that passes the full selection. The resulting distribution is fitted with the resolution function. Each Gaussian function has the same mean value, which is left free in the fit. The widths of the Gaussian functions are different and these are also free in the fit. The fit parameters are shown in Table 7.4 and the results in Figure 7.12. The resulting distribution has a similar form to those used in references [168, 169].

The result from the fit to the generated decays without the resolution function included is $\tau_{\mu\mu} = 1.6098 \pm 0.0014$ ps, which is consistent with the lifetime of generate

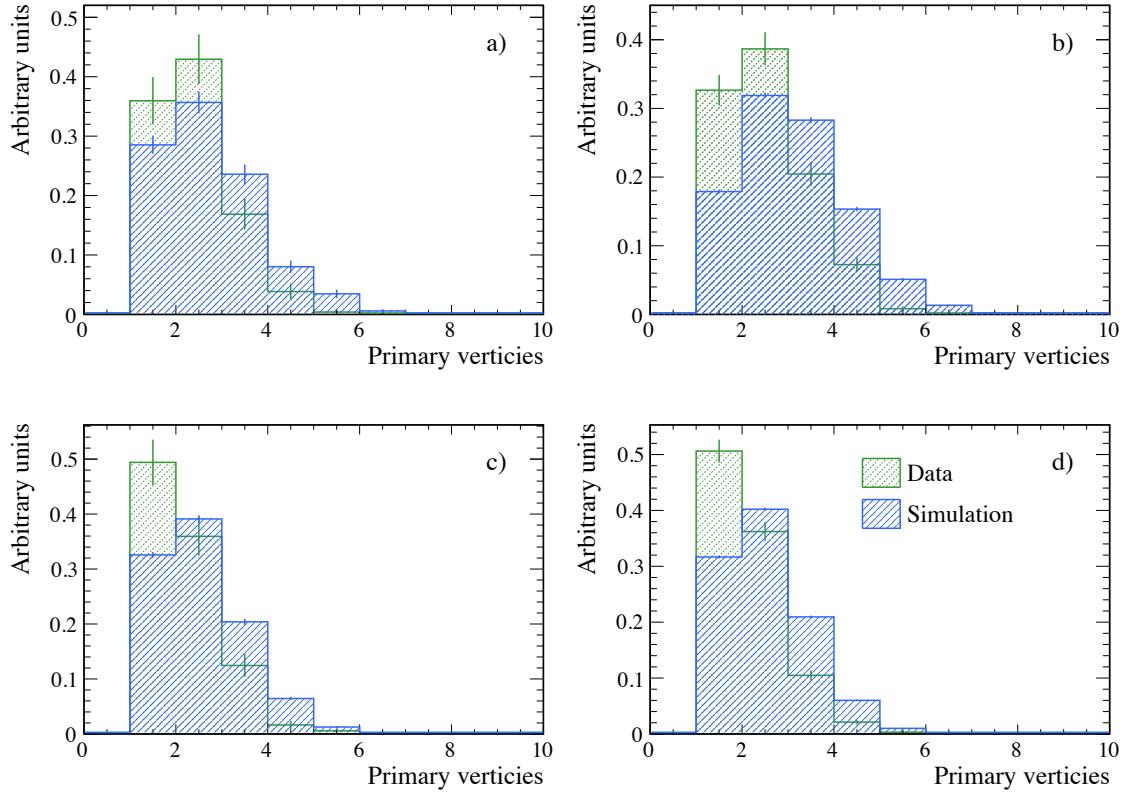


Fig. 7.13 Distributions for the number of primary vertices in an event $B^0 \rightarrow K^+\pi^-$ data and simulated decays for a) 2011, b) 2012, c) 2015 and d) 2016.

events. The difference between the lifetime used to generate events and the fitted value is 0.0002 ps, a factor of 10 smaller than the smallest systematic uncertainty. This cross check shows that the presence of incorrectly assigned PVs or detector resolution effects that are not included in the acceptance function have a negligible effect on the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime.

However, this check assumes that simulated decays provide a good estimate of the number of incorrectly assigned PVs. Figure 7.13 shows the number of $B^0 \rightarrow K^+\pi^-$ decays passing the selection for simulated $B^0 \rightarrow K^+\pi^-$ decays and sWeighted decays data for each year. On average there are more PVs per event in simulated decays compared to data, therefore using simulation would give an overestimation of the number of incorrectly assigned PVs expected in data.

7.6 Combinatorial background decay time model

The decay time distribution of combinatorial background decays is largely unknown due to the nature of the background. The model used for this distribution in the pseudoexperiments is described in Section 6.3.2. The decay time distribution of

combinatorial background decays consists of mostly a short-lived component with a lifetime of 1.3 ps and a long-lived component with a lifetime of 17 ps.

The sWeighting method is sensitive to backgrounds that are significantly longer lived than the signal and this can lead to a biased estimate of $\tau_{\mu\mu}$. During the selection an upper decay time cut is applied to remove long-lived backgrounds. This cut is very effective at removing the bias entering the final results, although the remaining effect must be evaluated.

As discussed in Section 6.3.2 determining the decay time distribution of combinatorial background decays is challenging because there are too few decays left in either data or $b\bar{b} \rightarrow \mu^+\mu^-X$ simulated decays after the selection requirements to determine the decay time PDF. Therefore, the combinatorial background of $B \rightarrow h^+h'^-$ decays in the mass range 5600 - 6000 MeV/ c^2 is used. The validity of using $B \rightarrow h^+h'^-$ combinatorial background to model $B_s^0 \rightarrow \mu^+\mu^-$ combinatorial background is studied by comparing the average lifetime of decays in bins of global BDT. The average lifetimes are shown in Table 7.5 for decays in data passing the selection requirements for $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays and in the mass ranges 5447 - 6000 MeV/ c^2 and 5600 - 6000 MeV/ c^2 , respectively. At low values of the global BDT the average lifetimes are similar and the lifetime of backgrounds of both decays increases with the output of the global BDT. Overall $B \rightarrow h^+h'^-$ combinatorial background decays are longer lived than $B_s^0 \rightarrow \mu^+\mu^-$ combinatorial background decays therefore the $B \rightarrow h^+h'^-$ decay time model is a conservative estimate for $B_s^0 \rightarrow \mu^+\mu^-$ combinatorial background as far as the affect of long-lived components in concerned.

The model used for the combinatorial background decay time currently introduces no significant bias into the pull distribution of $\Gamma_{\mu\mu}$ for pseudoexperiments as shown in Table 6.8. However, the size of a systematic bias from the choice of the lifetimes used in the combinatorial background decay time distribution is estimated by two sets of pseudoexperiments. The first uses the background decay time distribution in Table 6.4 and the second set uses a modified version this distribution: τ_1 and τ_2 are both increased by 1σ ; and the fraction of decays with lifetime τ_1 is increased by 1σ . For both sets of pseudoexperiments only combinatorial background and $B_s^0 \rightarrow \mu^+\mu^-$ decays are generated and 10,000 studies are performed for each configuration.

The resulting pull distributions for $\Gamma_{\mu\mu}$ are shown in Figure 7.14, the difference in the mean value of the distributions for the two studies is negligible and the width changes by 0.008 ps⁻¹ between the two studies. The change in the width is the largest and therefore the systematic uncertainty is estimated from this. Assuming the expected uncertainty for $\tau_{\mu\mu}$, a systematic uncertainty of 0.002 ps is assigned due to the combinatorial background decay time model.

$B_s^0 \rightarrow \mu^+\mu^-$			$B \rightarrow h^+h'^-$	
BDT bin	mean decay time / ps	Number of candidates	mean decay time / ps	Number of candidates
1	1.178 ± 0.005	50,695	1.124 ± 0.001	964,502
2	1.94 ± 0.10	244	2.394 ± 0.022	8,838
3	2.6 ± 0.3	46	2.78 ± 0.05	2,373
4	2.2 ± 0.4	17	3.02 ± 0.08	1,125
5	2.6 ± 1.1	4	3.42 ± 0.11	655
6	2.5 ± 0.4	3	3.98 ± 0.19	313
7	2.9 ± 1.1	2	4.6 ± 0.4	109
8	-	0	5.7 ± 0.7	35

Table 7.5 The mean decay time of $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ candidates in 2011, 2012, 2015 and 2016 data in bins of the global BDT output. The mass ranges 5447 - 6000 MeV/ c^2 and 5600 - 6000 MeV/ c^2 are used for $B_s^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h'^-$ decays, respectively.

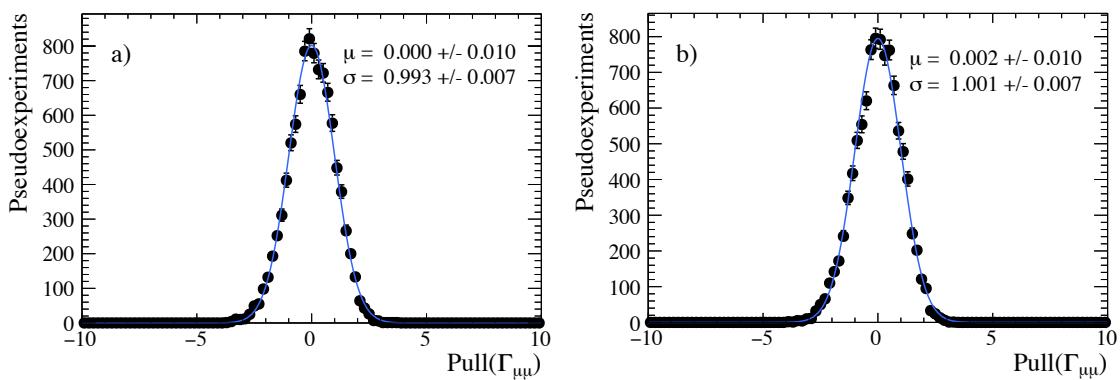


Fig. 7.14 Pull distributions for $\Gamma_{\mu\mu}$ from 10,000 pseudoexperiments using a) the nominal combinatorial background decay time model and b) the nominal model with the lifetimes and fraction of longer lived decays increased by one standard deviation.

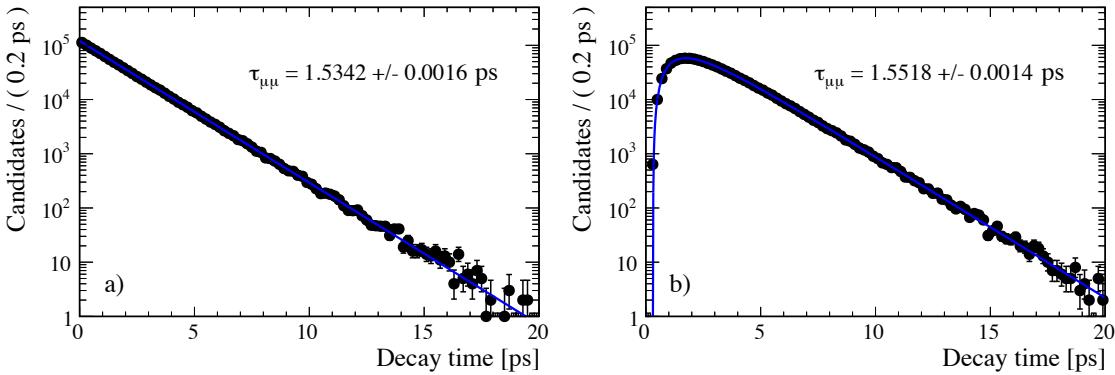


Fig. 7.15 Maximum likelihood fits to the decay time distribution to measure $\tau_{\mu\mu}$ for $B_s^0 \rightarrow \mu^+\mu^-$ decays that are composed of an equal mix from the heavy and light mass eigenstates. Decay time distributions are generated assuming a) a flat acceptance function and b) the acceptance function used to describe $B_s^0 \rightarrow \mu^+\mu^-$ decays in data.

7.7 Mix of B_s^0 mass eigenstates

In the SM the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is equal to the lifetime of the heavy B_s^0 mass eigenstate. However, the real $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime could be different due to the mixture of the light and heavy mass eigenstates.

As shown in Section 6.3 the selection efficiency used to identify $B_s^0 \rightarrow \mu^+\mu^-$ candidates is not uniform across the decay time range. The selection rejects a greater proportion of candidates with short lifetimes compared to candidates with longer lifetimes. Therefore, the presence of the light B_s^0 mass eigenstate decaying as $B_s^0 \rightarrow \mu^+\mu^-$ could be masked by the bias in the decay time distribution, since the efficiency to select the light B_s^0 mass eigenstate is lower than the efficiency to select the heavy B_s^0 mass eigenstate.

The size of this effect has been estimated using two simple pseudoexperiments. The first assumes that the selection has no bias on the decay time distribution and 1 million candidates are generated with equal contributions from the heavy and light B_s^0 mass eigenstates. A second set of 1 million candidates are generated with the same mix of eigenstates but with a more realistic decay time model including the $B_s^0 \rightarrow \mu^+\mu^-$ acceptance function. A fit is performed to the first set of candidates with a single exponential function and the second set with the acceptance function and exponential function in order to find $\tau_{\mu\mu}$ for each distribution. The acceptance parameters are fixed in the second fit.

The values of $\tau_{\mu\mu}$ are compared for the two studies and a systematic uncertainty is assigned for the change in $\tau_{\mu\mu}$ caused by the inclusion of the acceptance function. The fit results are shown in Figure 7.15 and the difference between the measured lifetimes for the two studies is 0.018 ps and is assigned as a systematic uncertainty.

7.8 Production asymmetry of B_s^0 and \bar{B}_s^0 mesons

The $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is the mean lifetime of an unbiased sample of $B_s^0 \rightarrow \mu^+ \mu^-$ decays, as discussed in Chapter 2, and is given by

$$\tau_{\mu\mu} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle dt}, \quad (7.5)$$

where the untagged decay rate is

$$\langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle = \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) + \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-), \quad (7.6)$$

and assumes that B_s^0 and \bar{B}_s^0 mesons are produced at equal rates. This assumption is made for the measured value of $\tau_{\mu\mu}$. However, since the LHC is a pp collider, B_s^0 and \bar{B}_s^0 mesons are not produced at equal rates. The effect of such a production asymmetry of B_s^0 and \bar{B}_s^0 mesons on the measured results must therefore be evaluated.

The production asymmetry is given by

$$A_p \equiv \frac{\sigma(B_s^0) - \sigma(\bar{B}_s^0)}{\sigma(B_s^0) + \sigma(\bar{B}_s^0)}, \quad (7.7)$$

where $\sigma(B_s^0)$ and $\sigma(\bar{B}_s^0)$ are the production cross-sections for B_s^0 and \bar{B}_s^0 mesons, respectively. The production asymmetry was measured by LHCb in 2011 at a centre-of-mass energy of 7 TeV as $A_p = (1.09 \pm 2.61 \pm 0.66)\%$ [170]. The presence of the production asymmetry modifies the untagged $B_s^0 \rightarrow \mu^+ \mu^-$ decay rate to

$$\begin{aligned} \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle &= (1 + A_p) \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \\ &\quad + (1 - A_p) \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-). \end{aligned} \quad (7.8)$$

The effect of the production asymmetry on the measured lifetime is determined from Equations 7.5 and 7.8 using the decay rates of $B_s^0 \rightarrow \mu^+ \mu^-$ and $\bar{B}_s^0 \rightarrow \mu^+ \mu^-$ given in Equations 2.33 and 2.34.

The total decay rate with the production asymmetry is

$$\begin{aligned} \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle &= \mathcal{N} |A_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) e^{-\Gamma_s t} \left\{ \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + A_{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\quad \left. + A_p [C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)] \right\} + \mathcal{O}(a), \end{aligned}$$

where $C_\lambda = (1 - |\lambda_{\mu\mu}|^2)/(1 + |\lambda_{\mu\mu}|^2)$ and $S_\lambda = 2\Im/(1 + |\lambda_{\mu\mu}|^2)$ are \mathcal{CP} asymmetries and are related to $A_{\Delta\Gamma}$ by $|A_{\Delta\Gamma}|^2 + |C_\lambda|^2 + |S_\lambda|^2 = 1$ [49].

The production asymmetry introduces an oscillatory term into the decay rate which disappears when $A_p = 0$. Using the relationships $\Delta\Gamma_s = \Gamma_L - \Gamma_H$ and $\Gamma_s = (\Gamma_L + \Gamma_H)/2$ and ignoring terms $\mathcal{O}(a)$, the decay rate becomes

$$\begin{aligned} \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \simeq \mathcal{N}' \left\{ (1 - A_{\Delta\Gamma}) e^{-\Gamma_L t} + (1 + A_{\Delta\Gamma}) e^{-\Gamma_H t} \right. \\ \left. + 2A_p e^{-\Gamma_s t} [C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)] \right\}, \end{aligned} \quad (7.10)$$

where $\mathcal{N}' \equiv \frac{1}{2}\mathcal{N}|A_{\mu\mu}|^2(1+|\lambda_{\mu\mu}|^2)$. This decay rate is used to calculate the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime in the presence of a production asymmetry. Using integration by parts the contributing terms to the effective lifetime become

$$\begin{aligned} \int_0^\infty t \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle dt = \mathcal{N}' \left\{ \frac{1 - A_{\Delta\Gamma}}{\Gamma_L^2} + \frac{1 + A_{\Delta\Gamma}}{\Gamma_H^2} \right. \\ \left. + \frac{2A_p}{(\Delta m_s^2 + \Gamma_s^2)^2} [C_\lambda(\Gamma_s^2 - \Delta m_s^2) + 2S_\lambda\Gamma_s\Delta m_s] \right\} \end{aligned} \quad (7.11)$$

and

$$\begin{aligned} \int_0^\infty \langle \Gamma(B_s(t) \rightarrow \mu^+ \mu^-) \rangle dt = \mathcal{N}' \left\{ \frac{1 - A_{\Delta\Gamma}}{\Gamma_L} + \frac{1 + A_{\Delta\Gamma}}{\Gamma_H} \right. \\ \left. + 2A_p \left[C_\lambda \frac{\Gamma_s}{\Delta m_s^2 + \Gamma_s^2} + S_\lambda \frac{\Delta m_s}{\Delta m_s^2 + \Gamma_s^2} \right] \right\}. \end{aligned} \quad (7.12)$$

The effect of the production asymmetry on the effective lifetime can now be calculated using the values of $\Delta m_s = 17.717 \text{ ps}^{-1}$, $\Gamma_s = 0.662 \text{ ps}^{-1}$, $\Gamma_L = 0.703 \text{ ps}^{-1}$ and $\Gamma_H = 0.621 \text{ ps}^{-1}$ [5]. A value of $A_p = 0.40$ is used, which is 1 standard deviation greater than the value measured by LHCb, and $A_{\Delta\Gamma} = 0.0$ is chosen. Since $(A_{\Delta\Gamma})^2 + (C_\lambda)^2 + (S_\lambda)^2 = 1$, it is assumed that $C_\lambda = S_\lambda = \sqrt{0.5}$.

In the presence of the production asymmetry, the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is found to be 1.520 ps, whereas when there is no production asymmetry and $A_p = 0$, the effective lifetime is 1.522 ps. Therefore the production asymmetry introduces a bias on 0.002 ps into the measurement of $\tau_{\mu\mu}$, this value is assigned as a systematic uncertainty.

7.9 Summary

The complete list of systematic uncertainties for $\tau_{\mu\mu}$ are summarised in Table 7.6. Adding the uncertainties in quadrature gives a total uncertainty of 0.05 ps for $\tau_{\mu\mu}$, which corresponds to 11 % of the observed statistical uncertainty. The small size of

Uncertainty source	Uncertainty/ps
Fit accuracy	0.033
Background contamination	0.007
Acceptance function	0.028
Combinatorial background decay time model	0.008
Mix of B_s^0 eigenstates	0.018
Production asymmetry	0.002
Total	0.048

Table 7.6 Summary of the systematic uncertainties on $\tau_{\mu\mu}$, the total uncertainty is achieved by adding the separate uncertainties in quadrature.

the total of the systematic uncertainties compared to the statistical uncertainty is expected given the small number of observed number of decays.

Chapter 8

Summary and Outlook

8.1 Summary

The LHCb experiment was built to test the predictions of the SM and search for NP effects through the study of \mathcal{CP} violating and rare decays of b -hadrons. So far measurements performed using the LHCb experiment, and other LHC experiments, have not revealed conclusive evidence for NP effects, although some interesting anomalies have been seen in measured results in heavy flavour physics [21–36]. The search for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays was identified as one of the key measurements to be made with the LHCb experiment [171] as an indirect search for NP. In 2011 LHCb joined the search for these decays, that began over 30 years ago, using the unprecedented energies available at the LHC. The first evidence for $B_s^0 \rightarrow \mu^+ \mu^-$ decays was found by the LHCb experiment with 2.1 fb^{-1} of Run 1 data from pp collisions [45]. A combined analysis of the Run 1 data from the CMS and LHCb experiments produced the first observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and the first evidence for $B^0 \rightarrow \mu^+ \mu^-$ decays [47]. The measured branching fractions of these decays are consistent with the SM predictions and place constraints on BSM theories. However, the precision of the measurements still leaves room for NP effects to be revealed, therefore it is important to improve the precision of the branching fraction measurements. With the observation of the B_s^0 mode the search for $B_s^0 \rightarrow \mu^+ \mu^-$ decays is complete and properties of this decays, including the effective lifetime, can now be studied. The effective lifetime offers a new observable to test the SM in $B_s^0 \rightarrow \mu^+ \mu^-$ decays that is complementary to the branching fraction.

The measurements of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions and the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime with 4.4 fb^{-1} of Run 1 and Run 2 data collected by the LHCb experiment are presented in this dissertation. Figure 8.1 shows a summary of fit used to measure the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions for candidates in Run 1 and Run 2

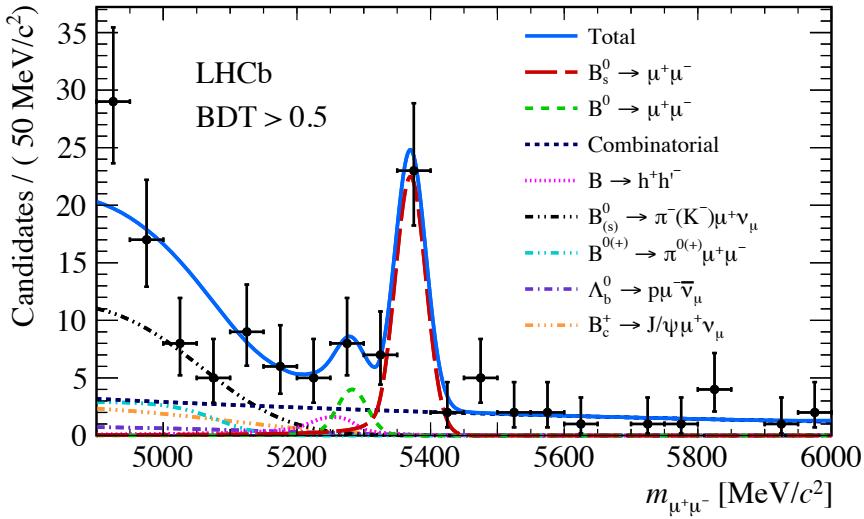


Fig. 8.1 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates with a global BDT value of $\text{BDT} > 0.5$ in 4.4 fb^{-1} of Run 1 and Run 2 data collected by the LHCb experiment and the mass fit used to measure the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions overlaid.

data with a global BDT value of $\text{BDT} > 0.5$. The measured branching fractions are

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) &= (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) &= (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}.\end{aligned}\quad (8.1)$$

The B_s^0 mode is observed with a statistical significance of 7.8σ , making this result the first single experiment observation of this decay and the most precise measurement to date. The B^0 mode is observed with a significance of 1.6σ , therefore a limit is placed on the branching fraction of $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 3.4 \times 10^{-10}$ at the 95% confidence level. The measured values are consistent with the SM predictions.

The effective lifetime of $B_s^0 \rightarrow \mu^+\mu^-$ decays is measured for the first time to be

$$\tau_{\mu\mu} = 2.04 \pm 0.44 \pm 0.05 \text{ ps}, \quad (8.2)$$

which is within 1.0σ of the SM prediction. The result is consistent with $A_{\Delta\Gamma} = +1$ hypothesis at 1.0σ and with $A_{\Delta\Gamma} = -1$ hypothesis at 1.4σ . Although the current precision of the measurement does not enable constraints to be placed on BSM theories it is important to illustrate the ability of the LHCb experiment to make this measurement.

The measured values of the branching fraction presented in this dissertation have already been used to help constrain parameters in BSM theories [84, 172–174]. The possible values available for the parameters P and S , defined in Section 2.2, are shown in Figure 8.2, where the values are constrained from the branching fraction results in Equation 8.1 and the results from the CMS collaboration in reference [41].

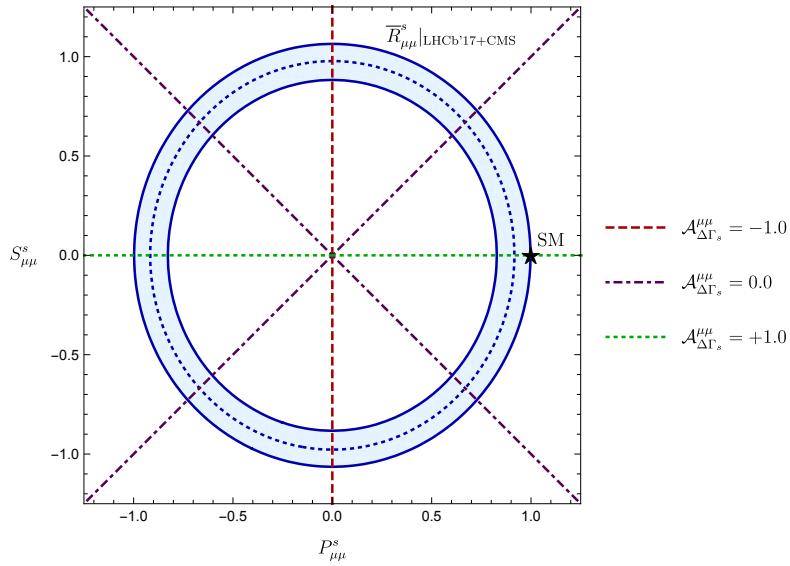


Fig. 8.2 Constraints in the P - S plane where P and S are defined in Section 2.2. The blue band corresponds to constraints from the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements presented in this dissertation and the CMS experiment Run 1 results in reference [41]. The dashed lines show different values for $A_{\Delta\Gamma}$ and the constraints assume $\varphi_P, \varphi_S \in 0, \pi$. The figure is taken from [172].

The current measurements produce a circular band of possible P and S values, and a precise measurement of $A_{\Delta\Gamma}$ through the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime will help resolve the ambiguities.

8.2 Outlook

The measured values of the branching fractions and the effective lifetime still leave plenty of room for NP effects to be observed with these decays. At the end of Run 2 of the LHC, the LHCb dataset will have almost doubled to 8 fb^{-1} , enabling the precision of these measurements to be improved. Looking further ahead, LHCb is expected to collect 50 fb^{-1} of data by the end of Run 4 and with the high luminosity LHC up to 300 fb^{-1} could be recorded.

With more data, the expected precision of the branching fraction measurements is expected to be reduced to $\sim 0.19 \times 10^{-9}$ with 50 fb^{-1} [175]. Not only are the branching fraction measurements in themselves interesting to test the SM but the ratio of the branching fractions of the two modes is also useful to test the SM, in particular the MFV hypothesis. The current precision of the ratio of branching fractions is $\sim 60\%$ [47], the future runs of the LHC will enable the precision of the ratio of branching fractions to be reduced to 40% with 50 fb^{-1} of pp data and 20% with 300 fb^{-1} of data [176].

The expected uncertainty achievable by the LHCb experiment for the effective lifetime at the end of Run 2 and after future runs of the LHC is estimated using pseudoexperiments based on the observed numbers of decays with 4.4 fb^{-1} and the current measurement strategy. At the end of Run 2, with 8 fb^{-1} , the median uncertainty of the effective lifetime will be $\sim 0.2 \text{ ps}$ which is reduced to $\sim 0.08 \text{ ps}$ with 50 fb^{-1} and $\sim 0.03 \text{ ps}$ with 300 fb^{-1} . Therefore, with 300 fb^{-1} the precision on the effective lifetime will enable NP effects to be distinguished with it. The expected mass and decay time distributions for 8 , 50 and 300 fb^{-1} are shown in Figure 8.3. The expected uncertainties for the effective lifetime measurement are conservative estimates because they are based on the current measurement strategy which was designed for low expected statistics. Therefore the precision of the measurement could be much better as different analysis methods can be taken advantage of with more statistics.

The current systematic uncertainty on the effective lifetime is 0.05 ps which is too large to allow possible NP effects to be observed. However, several components contributing to the total, such as the fit accuracy and the acceptance function systematic, will be reduced with the availability of more data enabling greater precision on the measurement. With more data, an alternative analysis approach could be used which would reduce the systematic uncertainties on the effective lifetime, the selection criteria can be designed so that it does not bias $B_s^0 \rightarrow \mu^+ \mu^-$ decay time distribution, therefore removing the need for an acceptance function and the associated systematic.

The study of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays has been in progress for over 30 years and with the energy and luminosity available at the LHC, the study of these decays is just as interesting as it ever was. As more data is collected by the LHCb experiment NP effects will have less and less space to hide, it will either be seen in $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays or these decays will place ever tighter constraints on BSM theories.

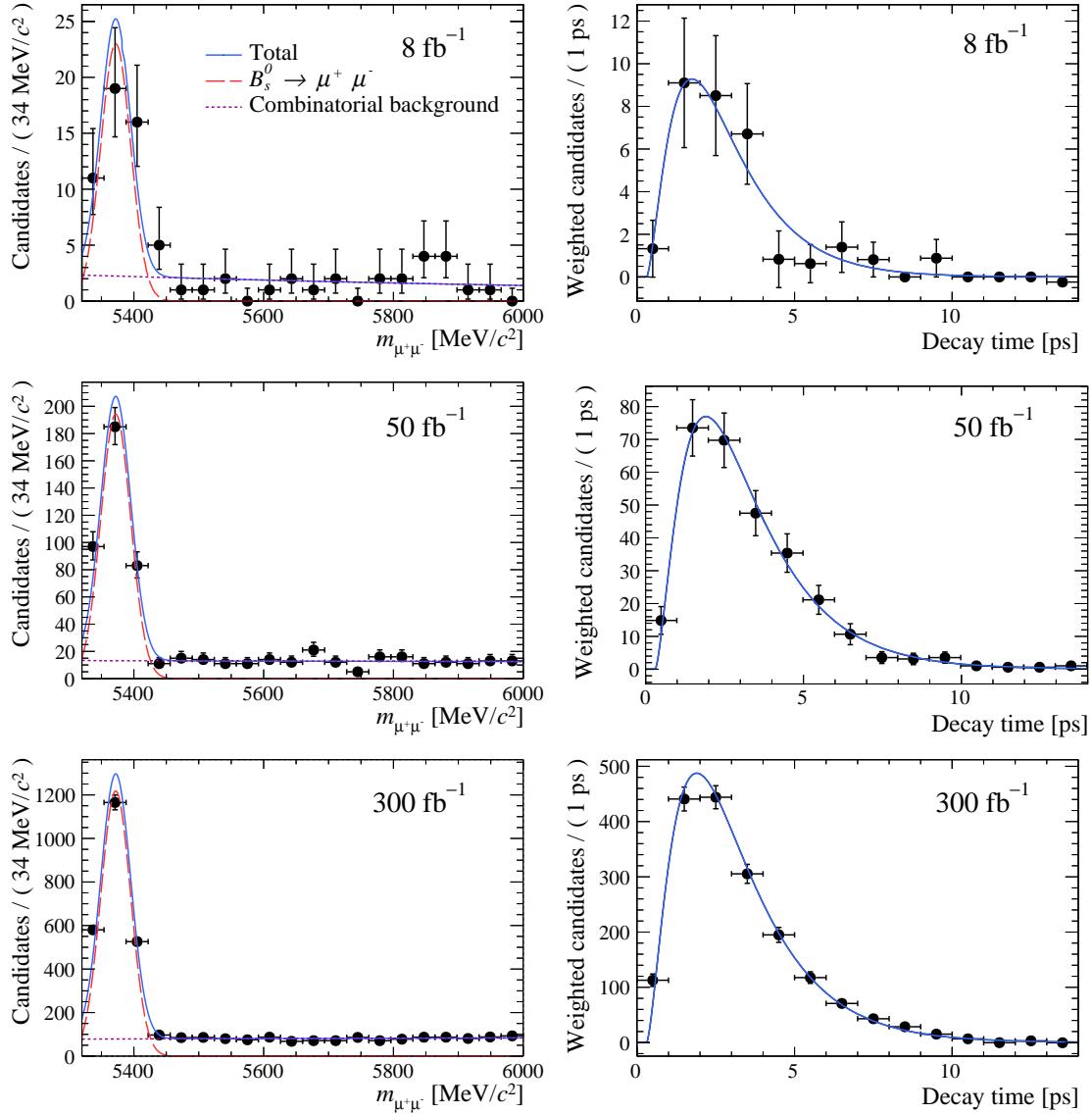


Fig. 8.3 The expected mass and decay time distributions for $B_s^0 \rightarrow \mu^+ \mu^-$ candidates to measure the effective lifetime with 8, 50 and 300 fb^{-1} . The signal and background yields are determined from the observed number of $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background decays and the expected number of mis-identified backgrounds and $B^0 \rightarrow \mu^+ \mu^-$ decays in 4.4 fb^{-1} of Run 1 and Run 2 data.

Bibliography

- [1] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801.
- [2] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961), no. 4 579 .
- [3] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [4] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2** (1970) 1285.
- [5] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.
- [6] R. Davis, D. S. Harmer, and K. C. Hoffman, *Search for neutrinos from the sun*, Phys. Rev. Lett. **20** (1968) 1205.
- [7] Super-Kamiokande collaboration, Y. Fukuda *et al.*, *Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days*, Phys. Rev. Lett. **81** (1998) 1158, arXiv:hep-ex/9805021, [Erratum: Phys. Rev. Lett. 81,4279(1998)].
- [8] Super-Kamiokande Collaboration, S. Fukuda *et al.*, *Constraints on neutrino oscillations using 1258 days of super-kamiokande solar neutrino data*, Phys. Rev. Lett. **86** (2001) 5656.
- [9] SNO Collaboration, Q. R. Ahmad *et al.*, *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by 8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301.
- [10] F. Zwicky, *Spectral displacement of extra galactic nebulae*, Helv. Phys. Acta **86** (1933).
- [11] M. S. Roberts and A. H. Rots, *Comparison of rotation curves of different galaxy types*, Astronomy and Astrophysics **26** (1973) 483.
- [12] J. Dunkley *et al.*, *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Likelihoods and Parameters from the WMAP data*, .
- [13] P. A. R. Ade *et al.*, *Planck 2015 results. XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13, arXiv:1502.01589.
- [14] A. D. Sakharov, *Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32, [Usp. Fiz. Nauk 161,61(1991)].

- [15] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry*, Mod. Phys. Lett. **A9** (1994) 795, [arXiv:hep-ph/9312215](#).
- [16] M. Dine, *Supersymmetry and String Theory: Beyond the Standard Model*, Cambridge University Press, 2007.
- [17] J. R. Ellis, *Limits of the standard model*, in *PSI Zuoz Summer School on Exploring the Limits of the Standard Model Zuoz, Engadin, Switzerland, August 18-24, 2002*, 2002. [arXiv:hep-ph/0211168](#).
- [18] A. Pomarol, *Beyond the Standard Model*, in *Proceedings, High-energy Physics. Proceedings, 18th European School (ESHEP 2010): Raseborg, Finland, June 20 - July 3, 2010*, pp. 115–151, 2012. [arXiv:1202.1391](#).
- [19] 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, [arXiv:1207.7235](#).
- [20] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, [arXiv:1207.7214](#).
- [21] Belle Collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of $b \rightarrow K^* \ell^+ \ell^-$* , Phys. Rev. Lett. **118** (2017) 111801.
- [22] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \rightarrow K^{(*)} \mu^+ \mu^-$ decays*, JHEP **06** (2014) 133, [arXiv:1403.8044](#).
- [23] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, [arXiv:1512.04442](#).
- [24] ATLAS Collaboration, *Angular analysis of $B_d^0 \rightarrow K^* \mu^+ \mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2017-023, CERN, Geneva, Apr, 2017.
- [25] CMS Collaboration, *Measurement of the P_1 and P'_5 angular parameters of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Tech. Rep. CMS-PAS-BPH-15-008, CERN, Geneva, 2017.
- [26] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **09** (2015) 179, [arXiv:1506.08777](#).
- [27] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601.
- [28] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays*, [arXiv:1705.05802](#).
- [29] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)$* , Phys. Rev. Lett. **115** (2015), no. 11 111803, [arXiv:1506.08614](#), [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)].

- [30] Belle Collaboration, M. Huschle *et al.*, *Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, Phys. Rev. **D92** (2015), no. 7 072014, [arXiv:1507.03233](#).
- [31] BaBar Collaboration, J. P. Lees *et al.*, *Evidence for an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays*, Phys. Rev. Lett. **109** (2012) 101802, [arXiv:1205.5442](#).
- [32] BaBar Collaboration, J. P. Lees *et al.*, *Measurement of an Excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons*, Phys. Rev. **D88** (2013), no. 7 072012, [arXiv:1303.0571](#).
- [33] Belle Collaboration, Y. Sato *et al.*, *Measurement of the branching ratio of $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method*, Phys. Rev. **D94** (2016), no. 7 072007, [arXiv:1607.07923](#).
- [34] W. Altmannshofer, P. Stangl, and D. M. Straub, *Interpreting Hints for Lepton Flavor Universality Violation*, [arXiv:1704.05435](#).
- [35] B. Capdevila *et al.*, *Patterns of New Physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data*, [arXiv:1704.05340](#).
- [36] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ -lepton properties as of summer 2016*, [arXiv:1612.07233](#).
- [37] ATLAS collaboration, G. Aad *et al.*, *Search for the decay $B_s^0 \rightarrow \mu\mu$ with the ATLAS detector*, Phys. Lett. **B713** (2012) 387, [arXiv:1204.0735](#).
- [38] ATLAS collaboration, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, Eur. Phys. J. **C76** (2016), no. 9 513, [arXiv:1604.04263](#).
- [39] CMS collaboration, S. Chatrchyan *et al.*, *Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **107** (2011) 191802, [arXiv:1107.5834](#).
- [40] CMS collaboration, S. Chatrchyan *et al.*, *Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ decays*, JHEP **04** (2012) 033, [arXiv:1203.3976](#).
- [41] 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** (2013) 101804, [arXiv:1307.5025](#).
- [42] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$* , Phys. Lett. **B699** (2011) 330, [arXiv:1103.2465](#).
- [43] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$* , Phys. Lett. **B708** (2012) 55, [arXiv:1112.1600](#).
- [44] 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](#).
- [45] LHCb collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \rightarrow \mu^+\mu^-$* , Phys. Rev. Lett. **110** (2013), no. 2 021801, [arXiv:1211.2674](#).
- [46] 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](#).

- [47] LHCb collaboration, 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, arXiv:1411.4413.
- [48] T. Blake, G. Lanfranchi, and D. M. Straub, *Rare B Decays as Tests of the Standard Model*, Prog. Part. Nucl. Phys. **92** (2017) 50, arXiv:1606.00916.
- [49] K. Anikeev *et al.*, *B physics at the Tevatron: Run II and beyond*, in *Workshop on B Physics at the Tevatron: Run II and Beyond Batavia, Illinois, September 23-25, 1999*, 2001. arXiv:hep-ph/0201071.
- [50] I. Dunietz, R. Fleischer, and U. Nierste, *In pursuit of new physics with B_s decays*, Phys. Rev. **D63** (2001) 114015, arXiv:hep-ph/0012219.
- [51] U. Nierste, *Three Lectures on Meson Mixing and CKM phenomenology*, in *Heavy quark physics. Proceedings, Helmholtz International School, HQP08, Dubna, Russia, August 11-21, 2008*, pp. 1–38, 2009. arXiv:0904.1869.
- [52] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martínez Santos, *Supersymmetry confronts $B_s \rightarrow \mu^+ \mu^-$: Present and future status*, Phys. Rev. **D87** (2013), no. 3 035026, arXiv:1212.4887.
- [53] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.
- [54] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Progress of Theoretical Physics **49** (1973), no. 2 652.
- [55] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Phys. Rev. Lett. **51** (1983) 1945.
- [56] K. De Bruyn *et al.*, *Branching Ratio Measurements of B_s Decays*, Phys. Rev. **D86** (2012) 014027, arXiv:1204.1735.
- [57] S. Tolk, *Discovery Of Rare B decays. First observation of the $B_s \rightarrow \mu^+ \mu^-$ decays, first evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decays.*, PhD thesis, Groningen U., Dec, 2015, Presented 08 Apr 2016.
- [58] K. G. Wilson, *Non-lagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.
- [59] K. G. Wilson, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Communications in Mathematical Physics **24** (1972), no. 2 87.
- [60] K. De Bruyn *et al.*, *Probing New Physics via the $B_s^0 \rightarrow \mu^+ \mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.
- [61] A. J. Buras, R. Fleischer, J. Girrbach, and R. Knegjens, *Probing New Physics with the $B_s^0 \rightarrow \mu^+ \mu^-$ Time-Dependent Rate*, JHEP **07** (2013) 77, arXiv:1303.3820.
- [62] R. Fleischer and R. Knegjens, *Effective Lifetimes of B_s Decays and their Constraints on the B_s^0 - \bar{B}_s^0 Mixing Parameters*, Eur. Phys. J. **C71** (2011) 1789, arXiv:1109.5115.

- [63] C. Bobeth *et al.*, $B_{s,d} \rightarrow l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty, Phys. Rev. Lett. **112** (2014) 101801, [arXiv:1311.0903](#).
- [64] Fermilab Lattice, MILC, A. Bazavov *et al.*, B - and D -meson decay constants from three-flavor lattice QCD, Phys. Rev. **D85** (2012) 114506, [arXiv:1112.3051](#).
- [65] H. Na *et al.*, The B and B_s Meson Decay Constants from Lattice QCD, Phys. Rev. **D86** (2012) 034506, [arXiv:1202.4914](#).
- [66] O. Witzel, B -meson decay constants with domain-wall light quarks and non-perturbatively tuned relativistic b -quarks, PoS **LATTICE2013** (2014) 377, [arXiv:1311.0276](#).
- [67] R. J. Knegjens, Strategies to Hunt for New Physics with Strange Beauty Mesons, PhD thesis, Vrije U., Amsterdam, 2014.
- [68] W. Altmannshofer and D. M. Straub, New physics in $b \rightarrow s$ transitions after LHC run 1, Eur. Phys. J. **C75** (2015), no. 8 382, [arXiv:1411.3161](#).
- [69] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, Minimal flavor violation: An Effective field theory approach, Nucl. Phys. **B645** (2002) 155, [arXiv:hep-ph/0207036](#).
- [70] S. P. Martin, A Supersymmetry primer, [arXiv:hep-ph/9709356](#), [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [71] W.-S. Hou, Source of CP Violation for the Baryon Asymmetry of the Universe, Chin. J. Phys. **47** (2009) 134, [arXiv:0803.1234](#).
- [72] D. M. Straub, New physics correlations in rare decays, in CKM unitarity triangle. Proceedings, 6th International Workshop, CKM 2010, Warwick, UK, September 6-10, 2010. [arXiv:1012.3893](#).
- [73] LHCb collaboration, CMS collaboration, Combination of results on the rare decays $B_{(s)}^0 \rightarrow \mu^+\mu^-$ from the CMS and LHCb experiments, .
- [74] L. J. and M. B. Wise, Flavor changing higgs boson couplings, Nuclear Physics B **187** (1981), no. 3 397 .
- [75] E. Witten, Dynamical Breaking of Supersymmetry, Nucl. Phys. **B188** (1981) 513.
- [76] A. J. Buras, P. H. Chankowski, J. Rosiek, and L. Slawianowska, Correlation between ΔM_s and $B_{s,d}^0 \rightarrow \mu^+\mu^-$ in supersymmetry at large $\tan\beta$, Phys. Lett. **B546** (2002) 96, [arXiv:hep-ph/0207241](#).
- [77] K. S. Babu and C. F. Kolda, Higgs mediated $B^0 \rightarrow \mu^+\mu^-$ in minimal supersymmetry, Phys. Rev. Lett. **84** (2000) 228, [arXiv:hep-ph/9909476](#).
- [78] G. Isidori and A. Retico, Scalar flavor changing neutral currents in the large $\tan\beta$ limit, JHEP **11** (2001) 001, [arXiv:hep-ph/0110121](#).
- [79] R. Barbieri, C. W. Murphy, and F. Senia, B -decay Anomalies in a Composite Leptoquark Model, Eur. Phys. J. **C77** (2017), no. 1 8, [arXiv:1611.04930](#).

- [80] D. Bećirević, S. Fajfer, N. Košnik, and O. Sumensari, *Leptoquark model to explain the B -physics anomalies, R_K and R_D* , Phys. Rev. **D94** (2016), no. 11 115021, [arXiv:1608.08501](https://arxiv.org/abs/1608.08501).
- [81] G. Hiller and M. Schmaltz, *R_K and future $b \rightarrow s\ell\ell$ physics beyond the Standard Model opportunities*, Phys. Rev. **D90** (2014) 054014, [arXiv:1408.1627](https://arxiv.org/abs/1408.1627).
- [82] M. Bauer and M. Neubert, *Minimal Leptoquark Explanation for the $R_{D^{(*)}}$, R_K , and $(g-2)_g$ Anomalies*, Phys. Rev. Lett. **116** (2016), no. 14 141802, [arXiv:1511.01900](https://arxiv.org/abs/1511.01900).
- [83] S. Fajfer and N. Košnik, *Vector leptoquark resolution of R_K and $R_{D^{(*)}}$ puzzles*, Phys. Lett. **B755** (2016) 270, [arXiv:1511.06024](https://arxiv.org/abs/1511.06024).
- [84] W. Altmannshofer, C. Niehoff, and D. M. Straub, *$B_s \rightarrow \mu^+ \mu^-$ as current and future probe of new physics*, [arXiv:1702.05498](https://arxiv.org/abs/1702.05498).
- [85] CERN. <http://home.cern/about/member-states>.
- [86] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [87] CMS, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [88] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [89] TOTEM, G. Anelli *et al.*, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007.
- [90] MoEDAL, J. Pinfold *et al.*, *Technical Design Report of the MoEDAL Experiment*, , CERN-LHCC-2009-006, MoEDAL-TDR-001.
- [91] LHCf, O. Adriani *et al.*, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) S08006.
- [92] LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.
- [93] LHCb collaboration, <http://lhcb-public.web.cern.ch/lhcb-public/>.
- [94] CMS collaboration, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResultsOverview>.
- [95] ATLAS collaboration, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults> and <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [96] LHCb Collaboration, *LHCb : Technical Proposal*, Tech. Proposal, CERN, Geneva, 1998. CERN-LHCC-98-04, CERN-LHCC-P-4.
- [97] LHCb Collaboration, *LHCb technical design report: Reoptimized detector design and performance*, , CERN-LHCC-2003-030.
- [98] LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015), no. 07 1530022, [arXiv:1412.6352](https://arxiv.org/abs/1412.6352).

- [99] LHCb Collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2001.
- [100] LHCb collaboration, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, [arXiv:1405.7808](https://arxiv.org/abs/1405.7808).
- [101] LHCb Collaboration, *LHCb inner tracker: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2002. CERN-LHCC-2002-029.
- [102] LHCb Collaboration, *LHCb outer tracker: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2001. CERN-LHCC-2001-024.
- [103] LHCb Collaboration, *LHCb magnet: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2000. CERN-LHCC-2000-007.
- [104] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering **82** (1960) 35.
- [105] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.
- [106] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015), no. 02 P02007, [arXiv:1408.1251](https://arxiv.org/abs/1408.1251).
- [107] LHCb Collaboration, *LHCb RICH: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2000. CERN-LHCC-2000-037.
- [108] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, [arXiv:1211.6759](https://arxiv.org/abs/1211.6759).
- [109] TA2, M. Alemi *et al.*, *First operation of a hybrid photon detector prototype with electrostatic cross-focussing and integrated silicon pixel readout*, Nucl. Instrum. Meth. **A449** (2000) 48.
- [110] R. W. Forty and O. Schneider, *RICH pattern recognition*, Tech. Rep. LHCb-98-040, CERN, Geneva, Apr, 1998.
- [111] LHCb Collaboration, *LHCb calorimeters: Technical design report*, , CERN-LHCC-2000-036.
- [112] LHCb Collaboration, *LHCb muon system: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2001. CERN-LHCC-2001-010.
- [113] G. Bencivenni *et al.*, *A triple-GEM detector with pad readout for the inner region of the first LHCb muon station*, Tech. Rep. LHCb-2001-051, CERN, Geneva, Jul, 2001.
- [114] B. Botchine *et al.*, *Wire pad chambers and cathode pad chambers for the LHCb muon system*, , LHCb-2000-114, CERN-LHCb-2000-114.
- [115] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249).
- [116] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190.

- [117] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, [arXiv:1211.3055](https://arxiv.org/abs/1211.3055).
- [118] LHCb HLT project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, [arXiv:1310.8544](https://arxiv.org/abs/1310.8544).
- [119] LHCb collaboration, R. Aaij *et al.*, *Absolute luminosity measurements with the LHCb detector at the LHC*, JINST **7** (2012) P01010, [arXiv:1110.2866](https://arxiv.org/abs/1110.2866).
- [120] O. Lupton and G. Wilkinson, *Studies of $D^0 \rightarrow K_S^0 h^+ h^-$ decays at the LHCb experiment*, PhD thesis, Oxford U., Jul, 2016, Presented 14 Sep 2016.
- [121] P. Mato, *GAUDI-Architecture design document*, .
- [122] LHCb collaboration, R. Antunes-Nobrega *et al.*, *LHCb computing: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2005. CERN-LHCC-2005-019.
- [123] F. Stagni *et al.*, *LHCbDirac: Distributed computing in LHCb*, J. Phys. Conf. Ser. **396** (2012) 032104.
- [124] <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/>.
- [125] <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/>.
- [126] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.
- [127] LHCb collaboration, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, Journal of Physics: Conference Series **331** (2011) 032047.
- [128] LHCb collaboration, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
- [129] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
- [130] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
- [131] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.
- [132] 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, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- [133] GEANT4, S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [134] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.
- [135] <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/>.

- [136] I. Bird, *Computing for the Large Hadron Collider*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 99.
- [137] Worldwide LHC Computing Grid, <http://www.cern.ch/LHCgrid>.
- [138] S. Paterson and A. Tsaregorodtsev, *DIRAC Infrastructure for Distributed Analysis*, , <https://cds.cern.ch/record/1397926/files/LHCb-TALK-2006-013.pdf>.
- [139] <http://ganga.web.cern.ch/ganga/>.
- [140] K. Harrison *et al.*, *GANGA: a user-grid interface for ALTAS and LHCb*, CoRR **cs.SE/0306085** (2003).
- [141] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays $B(d,s) \rightarrow l+l-gamma$* , Phys. Rev. **D70** (2004) 114028, [arXiv:hep-ph/0410146](https://arxiv.org/abs/hep-ph/0410146).
- [142] Y. G. Aditya, K. J. Healey, and A. A. Petrov, *Faking $B_s \rightarrow \mu^+\mu^-$* , Phys. Rev. **D87** (2013) 074028, [arXiv:1212.4166](https://arxiv.org/abs/1212.4166).
- [143] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11** (2015) 743, [arXiv:1504.01568](https://arxiv.org/abs/1504.01568).
- [144] LHCb, R. Aaij *et al.*, *First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \rightarrow \pi^\pm \mu^+ \mu^-$ decay*, JHEP **10** (2015) 034, [arXiv:1509.00414](https://arxiv.org/abs/1509.00414).
- [145] C. M. Bouchard *et al.*, *$B_s \rightarrow K\ell\nu$ form factors from lattice QCD*, Phys. Rev. **D90** (2014) 054506, [arXiv:1406.2279](https://arxiv.org/abs/1406.2279).
- [146] RBC and UKQCD Collaborations, J. M. Flynn *et al.*, *$b \rightarrow \pi\ell\nu$ and $b_s \rightarrow k\ell\nu$ form factors and $|v_{ub}|$ from 2 + 1-flavour lattice qcd with domainwall light quarks and relativistic heavy quarks*, Phys. Rev. D **91** (2015) 074510.
- [147] LHCb, R. Aaij *et al.*, *Measurements of B_c^+ production and mass with the $B_c^+ \rightarrow J/\psi\pi^+$ decay*, Phys. Rev. Lett. **109** (2012) 232001, [arXiv:1209.5634](https://arxiv.org/abs/1209.5634).
- [148] LHCb, R. Aaij *et al.*, *Measurement of the ratio of B_c^+ branching fractions to $J/\psi\pi^+$ and $J/\psi\mu^+\nu_\mu$* , Phys. Rev. **D90** (2014), no. 3 032009, [arXiv:1407.2126](https://arxiv.org/abs/1407.2126).
- [149] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \rightarrow (\pi, K)(\ell^+\ell^-, \ell\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, Phys. Rev. **D86** (2012) 114025, [arXiv:1207.0265](https://arxiv.org/abs/1207.0265).
- [150] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.
- [151] D. Martinez Santos, *Study of the very rare decay $B_s \rightarrow \mu^+\mu^-$ in LHCb*, PhD thesis, Santiago de Compostela, Universidade de Santiago de Compostela, Santiago de Compostela, 2010, Presented on 05 May 2010.
- [152] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [153] F. Archilli *et al.*, *Background studies for $B^0 \rightarrow \mu^+\mu^-$ analysis optimization*, Tech. Rep. LHCb-INT-2014-047. CERN-LHCb-INT-2014-047, CERN, Geneva, Nov, 2014.

- [154] J. Stevens and M. Williams, *uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers*, JINST **8** (2013) P12013, [arXiv:1305.7248](#).
- [155] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](#).
- [156] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) 343.
- [157] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [158] ARGUS, H. Albrecht *et al.*, *Search for hadronic $b \rightarrow u$ decays*, Phys. Lett. **B241** (1990) 278.
- [159] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, [arXiv:1312.5000](#).
- [160] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data-driven measurement of trigger efficiencies with the TisTos method*, Tech. Rep. LHCb-INT-2013-038, CERN-LHCb-INT-2013-038, CERN, Geneva, Jun, 2013.
- [161] LHCb collaboration, *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*, .
- [162] A. L. Read, *Presentation of search results: the cl s technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002), no. 10 2693.
- [163] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [arXiv:physics/0306116](#).
- [164] M. Baak and W. Verkerke. https://twiki.cern.ch/twiki/pub/Main/RooFit/baak_eventweights.pdf.
- [165] Y. Xie. https://indico.cern.ch/event/201535/contributions/384769/attachment_s/301972/421937/sFit_analsysis_week_feb_2013_MAC_version.pdf.
- [166] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP violating phase ϕ_s in $\bar{B}_s^0 \rightarrow J/\psi f_0(980)$* , Phys. Lett. **B707** (2012) 497, [arXiv:1112.3056](#).
- [167] LHCb collaboration, R. Aaij *et al.*, *Effective lifetime measurements in the $B_s^0 \rightarrow K^+K^-$, $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \pi^+K^-$ decays*, Phys. Lett. **B736** (2014) 446, [arXiv:1406.7204](#).
- [168] LHCb collaboration, R. Aaij *et al.*, *First study of the CP -violating phase and decay-width difference in $B_s^0 \rightarrow \psi(2S)\phi$ decays*, Phys. Lett. **B762** (2016) 253, [arXiv:1608.04855](#).
- [169] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in $B^0 \rightarrow J/\psi K_S^0$ decays*, Phys. Rev. Lett. **115** (2015), no. 3 031601, [arXiv:1503.07089](#).
- [170] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $\bar{B}^0 - B^0$ and $\bar{B}_s^0 - B_s^0$ production asymmetries in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. **B739** (2014) 218, [arXiv:1408.0275](#).

- [171] LHCb collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, arXiv:0912.4179.
- [172] R. Fleischer, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi, *In Pursuit of New Physics with $B_{s,d}^0 \rightarrow \ell^+ \ell^-$* , arXiv:1703.10160.
- [173] C. Bobeth, A. J. Buras, A. Celis, and M. Jung, *Yukawa enhancement of Z -mediated New Physics in $\Delta S = 2$ and $\Delta B = 2$ Processes*, arXiv:1703.04753.
- [174] C.-W. Chiang, X.-G. He, F. Ye, and X.-B. Yuan, *Constraints and Implications on Higgs FCNC Couplings from Precision Measurement of $B_s \rightarrow \mu^+ \mu^-$ Decay*, arXiv:1703.06289.
- [175] *Impact of the LHCb upgrade detector design choices on physics and trigger performance*, Tech. Rep. LHCb-PUB-2014-040. CERN-LHCb-PUB-2014-040. LHCb-INT-2013-024, CERN, Geneva, Aug, 2014.
- [176] LHCb collaboration, R. Aaij *et al.*, *Expression of Interest for a Phase-II LHCb Upgrade: Opportunities in flavour physics, and beyond, in the HL-LHC era*, Tech. Rep. CERN-LHCC-2017-003, CERN, Geneva, Feb, 2017.
- [177] O. Augusto, *LHCb; LHCb Jet Reconstruction*, Dec, 2012.
- [178] W. Barter and D. Ward, *Z boson and associated jet production at the LHCb experiment*, PhD thesis, Cambridge U. (main), Mar, 2014, Presented 27 May 2014.
- [179] X. Cid Vidal, *Using jets to identify semi-leptonic B decays*, Tech. Rep. LHCb-INT-2015-022. CERN-LHCb-INT-2015-022, CERN, Geneva, Jun, 2015.
- [180] D. J. Jackson, *A Topological vertex reconstruction algorithm for hadronic jets*, Nucl. Instrum. Meth. **A388** (1997) 247.
- [181] A. Mordà and G. Mancinelli, *Rare dileptonic $B0(s)$ meson decays at LHCb*, PhD thesis, Aix-Marseille U., 2015, presented 28 Sep 2015.
- [182] CDF, A. Abulencia *et al.*, *Search for $B_s \rightarrow \mu^+ \mu^-$ and $B_d \rightarrow \mu^+ \mu^-$ decays in $p\bar{p}$ collisions with CDF II*, Phys. Rev. Lett. **95** (2005) 221805, arXiv:hep-ex/0508036, [Erratum: Phys. Rev. Lett.95,249905(2005)].

Appendix A

Distributions of input variables for the global BDT

Comparison of the signal and background distributions of input variables used in the global BDT for 2011, 2012, 2015 and 2016 data taking conditions. Signal distributions are from simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays for each year that have passed the selection cuts in Table 4.12. The background distributions are from $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays in 2011, 2012, 2015 and 2016 data with $m_{\mu\mu} > 5447$ MeV/ c^2 and passing the selection cuts in Table 4.12. The input variables used for the global BDT are:

- long track isolation criteria;
- VELO track isolation criteria;
- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ is the difference in azimuthal angles of the muons and $\Delta\eta$ the difference in the pseudo-rapidity of the muons;
- the smallest χ^2_{IP} with respect to the primary vertex of the $B_s^0 \rightarrow \mu^+ \mu^-$ of the muons;
- χ^2_{VTX} of the B_s^0 ;
- χ^2_{IP} of the B_s^0 with respect to the primary vertex; and
- the angle, θ , between the momentum vector of the B_s^0 and the vector connecting the production and decay vertices of the B_s^0 .

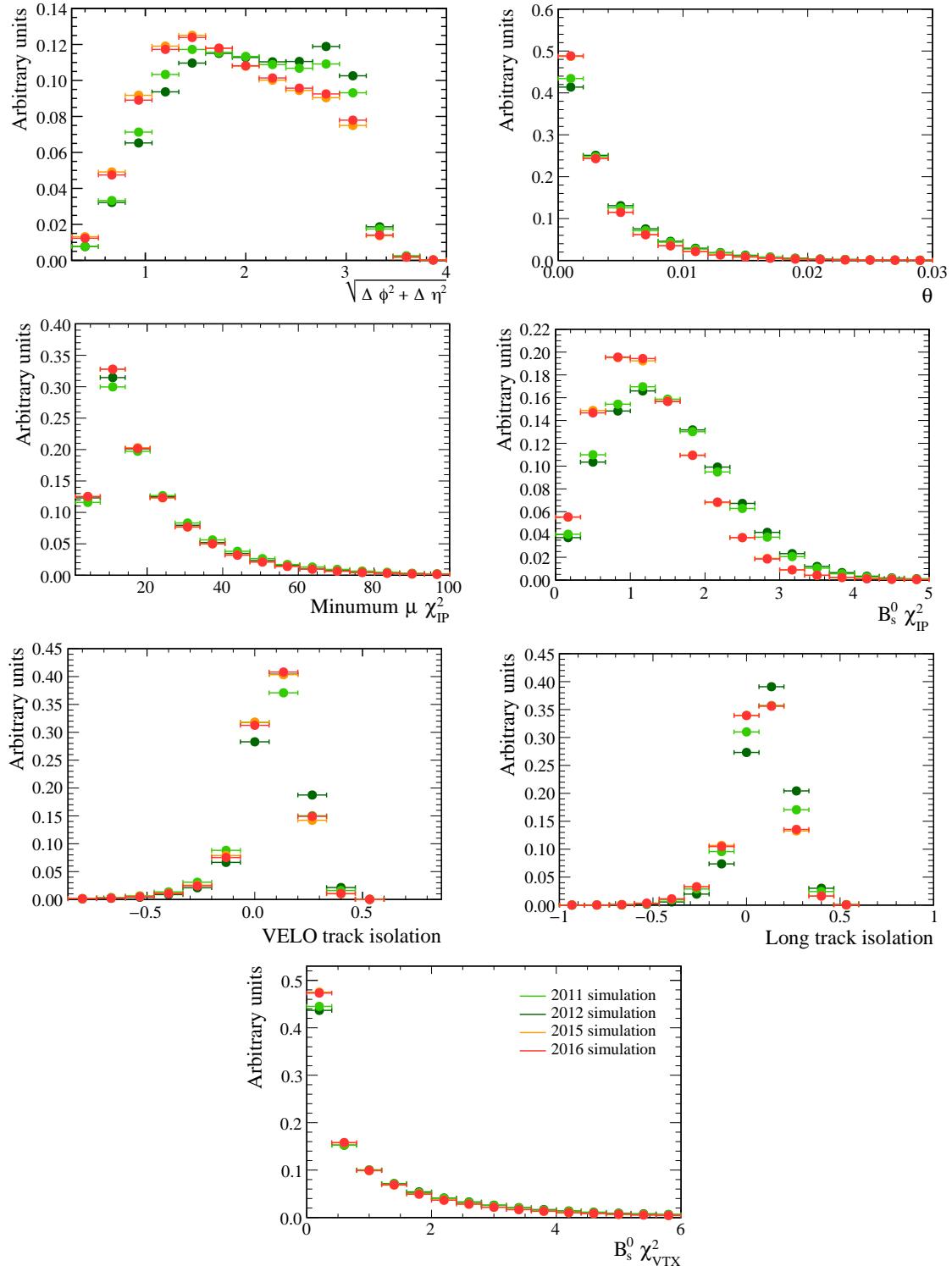


Fig. A.1 Signal distributions for input variables for the global BDT for $B_s^0 \rightarrow \mu^+ \mu^-$ simulated decays in 2011, 2012, 2015 and 2016.

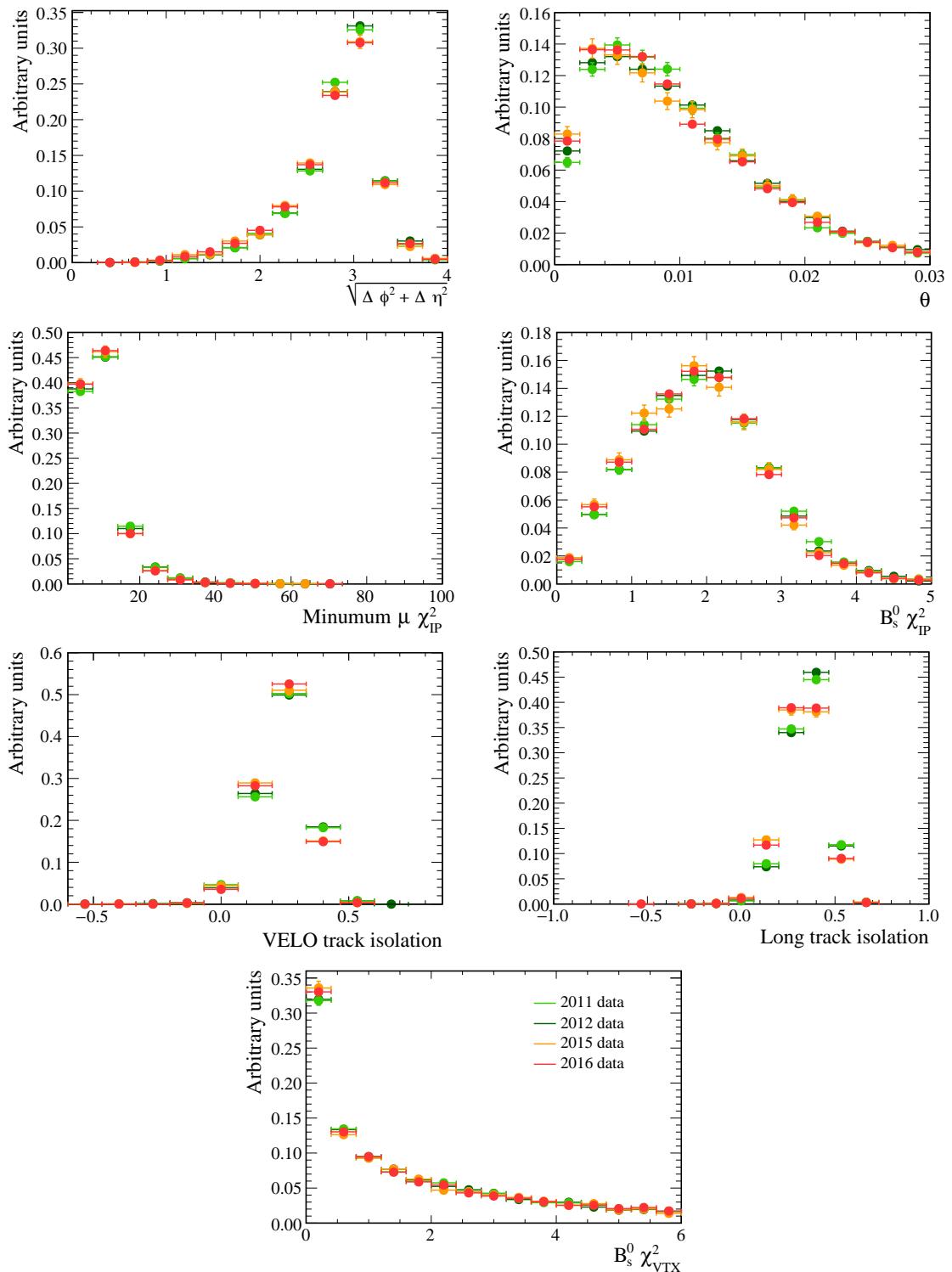


Fig. A.2 Background distributions for input variables from $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays in 2011, 2012, 2015 and 2016 data with $m_{\mu\mu} > 5447 \text{ MeV}/c^2$.

Appendix B

Multivariate classifier development for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement

B.1 Input variables

The input variables used in the adaptive boosting and uBoost BDTs were chosen separately, starting from a large set of variables. Initially the BDTs were trained using all input variables within the set and variables that had no impact on the BDT performance were removed until removing any of the remaining variables had a negative impact on the BDT performance. The performance of each BDT was evaluated from the integrated Receiver Operating Characteristic curve, which is the signal efficiency versus (1 - background rejection).

The adaptive boosting BDT uses 11 input variables and the uBoost BDT uses 21 variables which includes all variables used by adaptive boosting BDT. The input variables used in both algorithms are related to the B_s^0 , the muons, isolation variables and properties of jets reconstructed in the event. Isolation variables, discussed in Section 4.3.4.3, give a measure of how busy an event is and the separation of the tracks in a $B_s^0 \rightarrow \mu^+\mu^-$ candidate from other tracks in the event.

The reconstruction of jets in an event provides the most inclusive way to reconstruct semi-leptonic decays, where both the neutral particles and hadrons produced in the decays can be included into one jet. The BDTs are designed to remove combinatorial background decays formed from combining muons produced by semi-leptonic $b\bar{b} \rightarrow \mu^+\mu^- X_1 X_2$ processes, therefore information about semi-leptonic decays from reconstructed jets can help to separate signal and background decays. The reconstruction of jets at LHCb is detailed in reference [177, 178] and, during the reconstruction, constraints can be placed on the jets as to whether one or both muons in the $B_s^0 \rightarrow \mu^+\mu^-$ candidate or the B_s^0 is within a jet. Once the jets in an event

have been created variables can be constructed based on the properties of the jets and the comparisons between the jets and muons in the event [179]. These variables can be used to exploit differences in the jets created for semi-leptonic decays and $B_s^0 \rightarrow \mu^+\mu^-$ decays. For example a jet containing the B_s^0 of a $B_s^0 \rightarrow \mu^+\mu^-$ candidate created from two semi-leptonic decays by $b\bar{b} \rightarrow \mu^+\mu^- X$ will include both the b and \bar{b} created in the pp interaction, whereas a jet containing a real $B_s^0 \rightarrow \mu^+\mu^-$ decay is less likely to include both the original b and \bar{b} quarks. The information included in jet variables is complementary to that contained in the isolation variables.

The adaptive boosting and uBoost BDTs use input variables that are also used in the cut based selection. These variables are:

- IP and χ_{IP}^2 of the B_s^0 ;
- χ_{VTX}^2 of the B_s^0 ;
- the flight distance of the B_s^0 ;
- the p_T of the B_s^0 and the minimum p_T of the two muons; and
- the minimum χ_{IP}^2 of the two muons.

The definitions of these variables are given in Section 4.3.2.1. The additional input variables used in both BDTs are:

- the ‘polarisation angle’ which is the cosine of the angle between a vector perpendicular to a plane containing the B_s^0 momentum and the beam axis and the muon momentum in the B_s^0 rest frame;
- $(\Delta\phi)^2$, where $\Delta\phi$ is the difference in azimuthal angles of the muons;
- a BDT isolation variable designed in the same way to those described in Section 4.3.4.3, using information from long tracks. This isolation version was produced during the development of the final isolations used in the global BDT, the details of this variable can be found in reference [153]¹; and
- ZVtop isolation variable which uses a topological vertex algorithm [180] and is defined in reference [181].

The remaining input variables used in the uBoost BDT are:

- the direction cosine, DIRA, as defined in Section 4.3.2.1;
- $(\Delta\eta)^2$, where $\Delta\eta$ is the difference in the pseudorapidity of the muons;

¹Replacing this isolation variable with the Long track and VELO track isolations does not significantly improve the overall performance of either BDT.

- an isolation variable of the B_s^0 candidate based on the definition use by the CDF collaboration in the search for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays [182]. The isolation is computed from the transverse momentum of the B_s^0 , $p_T(B)$, and transverse momenta of tracks, $p_T(\text{tracks})$, in an event that fall within a cone around the B_s^0 . The cone is defined as $\sqrt{\delta\eta^2 + \delta\phi^2} > 1.0$ where $\delta\eta$ and $\delta\phi$ are the differences in pseudorapidity and azimuthal angle of a track in the event and the B_s^0 candidate. The isolation variable is defined as

$$I_{CDF} = \frac{p_T(B_s^0)}{p_T(B_s^0) + \sum_{\text{track} \in \text{cone}} p_T(\text{track})}; \quad (\text{B.1})$$

- a cut based muon isolation, I_μ , this isolation variable was the precursor of the BDT based isolation variables and is based on placing cuts on variables relating long tracks in the event to the muons in $B_s^0 \rightarrow \mu^+ \mu^-$ candidates. The definition of this variable can be found in reference [182?]Morda:2120795;
- the angle between the B_s^0 momentum and the sum of the momenta of all tracks in the event, excluding tracks from long lived particles and tracks associated with a different primary vertex than the primary vertex of the B_s^0 . Since b and \bar{b} quarks are produced in pairs in pp collisions, the angle is effectively the angle between the B_s^0 and the other b quark from the pair produced. Therefore this variable is called the ‘other B angle’, if there are too few candidates in the event to compute this variable the value is set to 1;
- the angle between the μ^+ candidate in the B_s^0 rest frame and the sum of the momenta in the B_s^0 rest frame of all tracks in the event, excluding tracks from long lived particles and tracks associated with a different primary vertex than the primary vertex of the B_s^0 . If there are too few tracks in the event to compute this variable the angle is set to $\pi/2$;
- the jet width of jets that are forced to contain both muons in the $B_s^0 \rightarrow \mu^+ \mu^-$ candidate and are constructed around the B_s^0 . The width is defined as the average value of $\sqrt{\eta^2 + \phi^2}$ where the difference is computed between each component in the jet and the jet total, in the total width each component in the jet is a weighted by its p_T ;
- distance, $\sqrt{\eta^2 + \phi^2}$, between a jet forced to contain one lepton and the other lepton in the $B_s^0 \rightarrow \mu^+ \mu^-$ candidate
- the transverse momentum of jets that are forced to contain both muons in the $B_s^0 \rightarrow \mu^+ \mu^-$ candidate and are constructed around the B_s^0 ; and

- the ratio of the transverse momenta of the muons and the jet that contains the muons and is constructed around the B_s^0 .

The distributions of the input variables for $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu^+\mu^- X$ 2012 simulated decays are shown in Figure B.3.

B.2 Training parameters

The training parameters discussed in Section 4.3.4.1 put constraints on how a BDT separates signal and background decays. The training parameters used in the adaptive boost BDT were optimised by iterating over different training parameter values and choosing the BDT that gave the best signal significance for identifying $B \rightarrow h^+h^-$ decays in Run 1 data. The computation of the signal significance is described in Section 4.4.4.1. The final set of training parameters are given in Table B.1. The training parameters used in the uBoost BDT have not been optimised and are given in Table B.1. The parameter values suggested in reference [154] have been used where it was shown that different training parameters had a small impact of the overall BDT performance.

Adaptive Boost BDT		uBoost BDT	
Parameter	Value	Parameter	Value
nTrees	1000	nTrees	100
MinNodeSize	5%	nEventsMin	100
MaxDepth	3	MaxDepth	4
β	0.1	β	1.0
nCuts	30	nCuts	200

Table B.1 Training parameters used to specify the training of the adaptive boost and uBoost BDT.

B.3 Overtraining test

As discussed in Section 4.3.4.1, it is important that BDTs are not overtrained. To test this assumption the signal and background samples are both split in two, to create a training set and a testing set. A BDT is trained using the training set, and the BDT is then applied to both the training and testing sets. The distribution of BDT output values for signal and background decays in the training and testing sets are compared. If the BDT is overtrained the response of the BDT will be quite

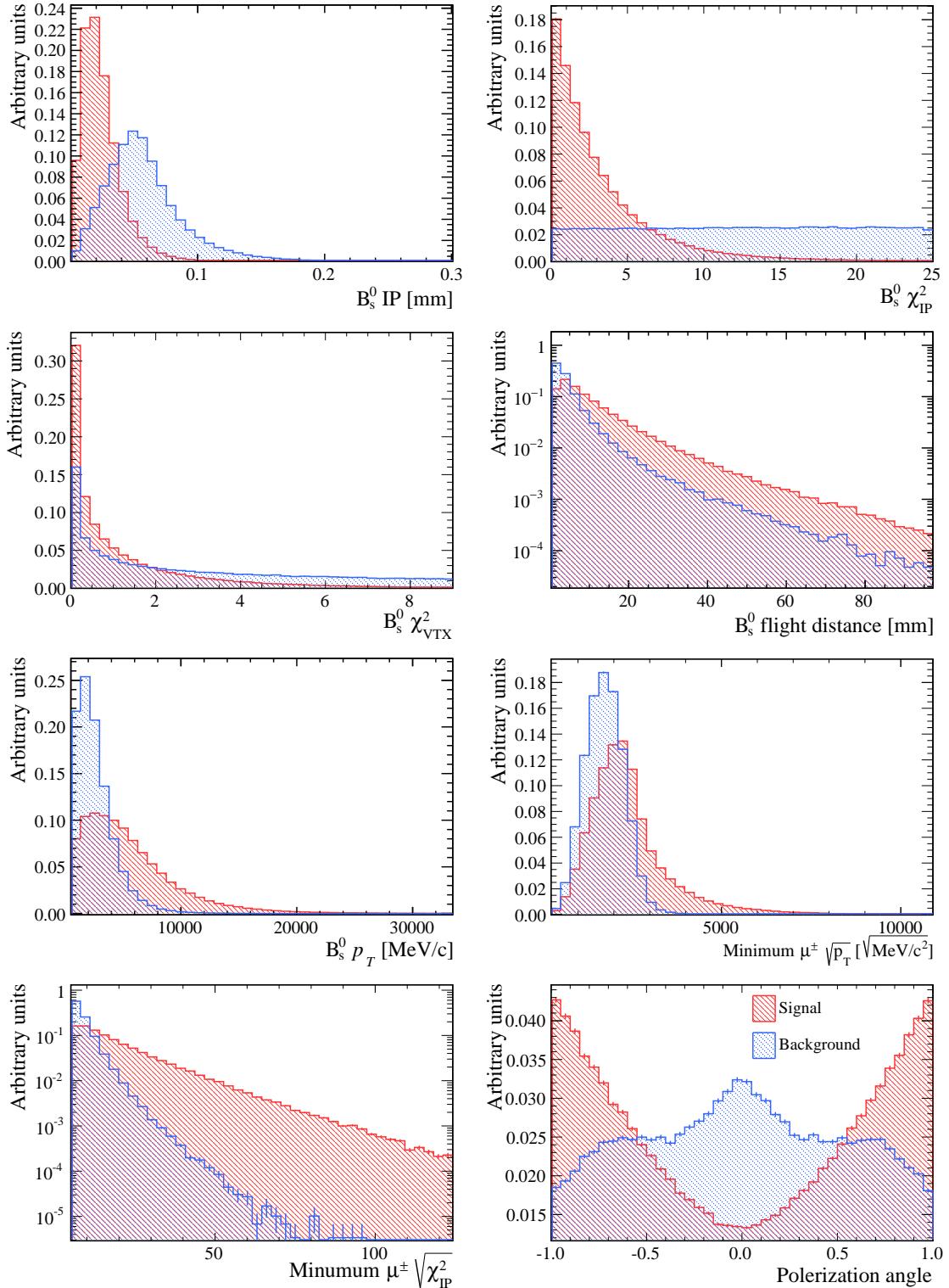


Fig. B.1 The distributions of the input variables used in the adaptive boost and uBoost BDTs for $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu^+ \mu^- X$ 2012 simulated decays.

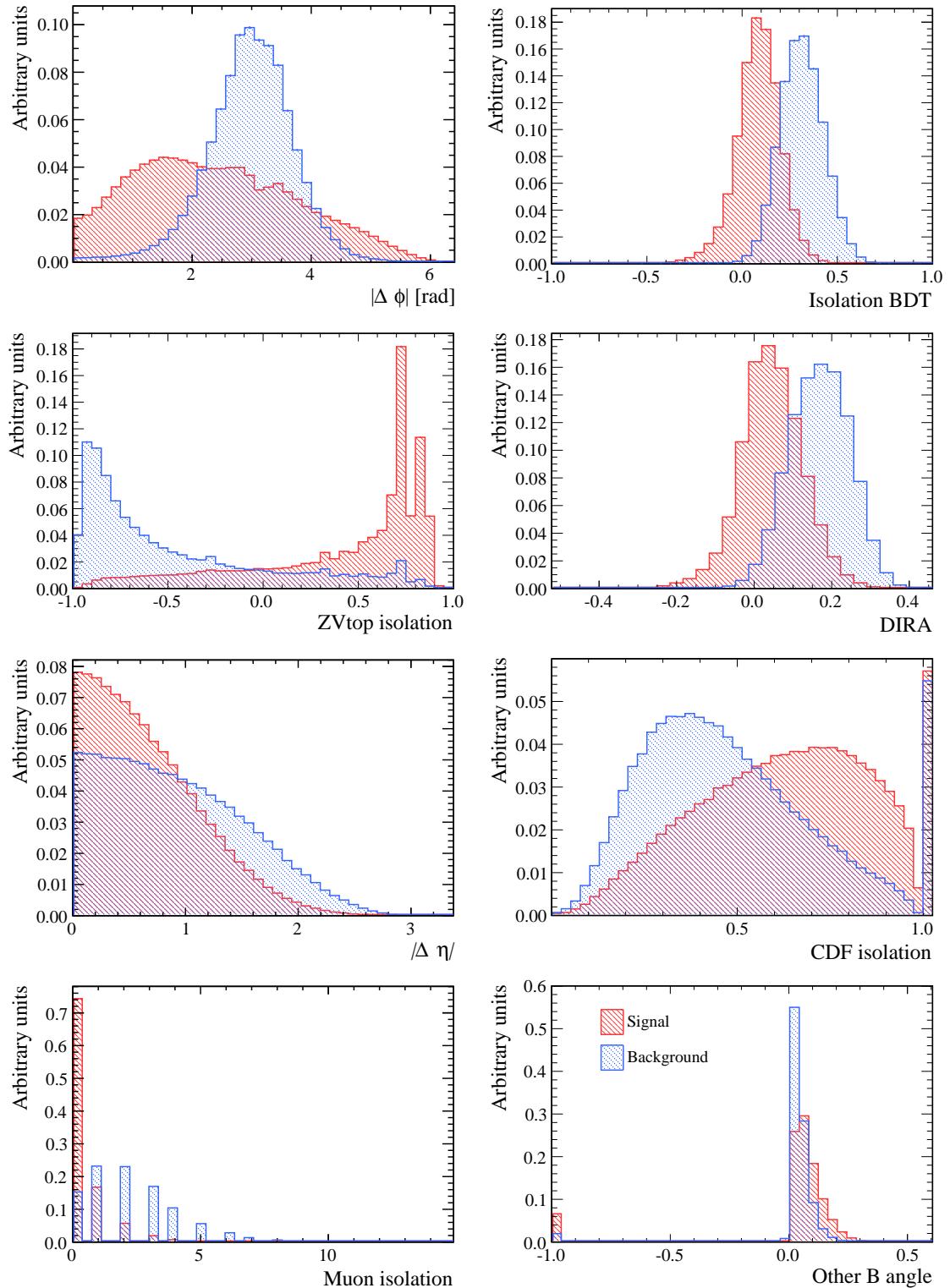


Fig. B.2 The distributions of the input variables used in the adaptive boost and uBoost BDTs for $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu^+ \mu^- X$ 2012 simulated decays.

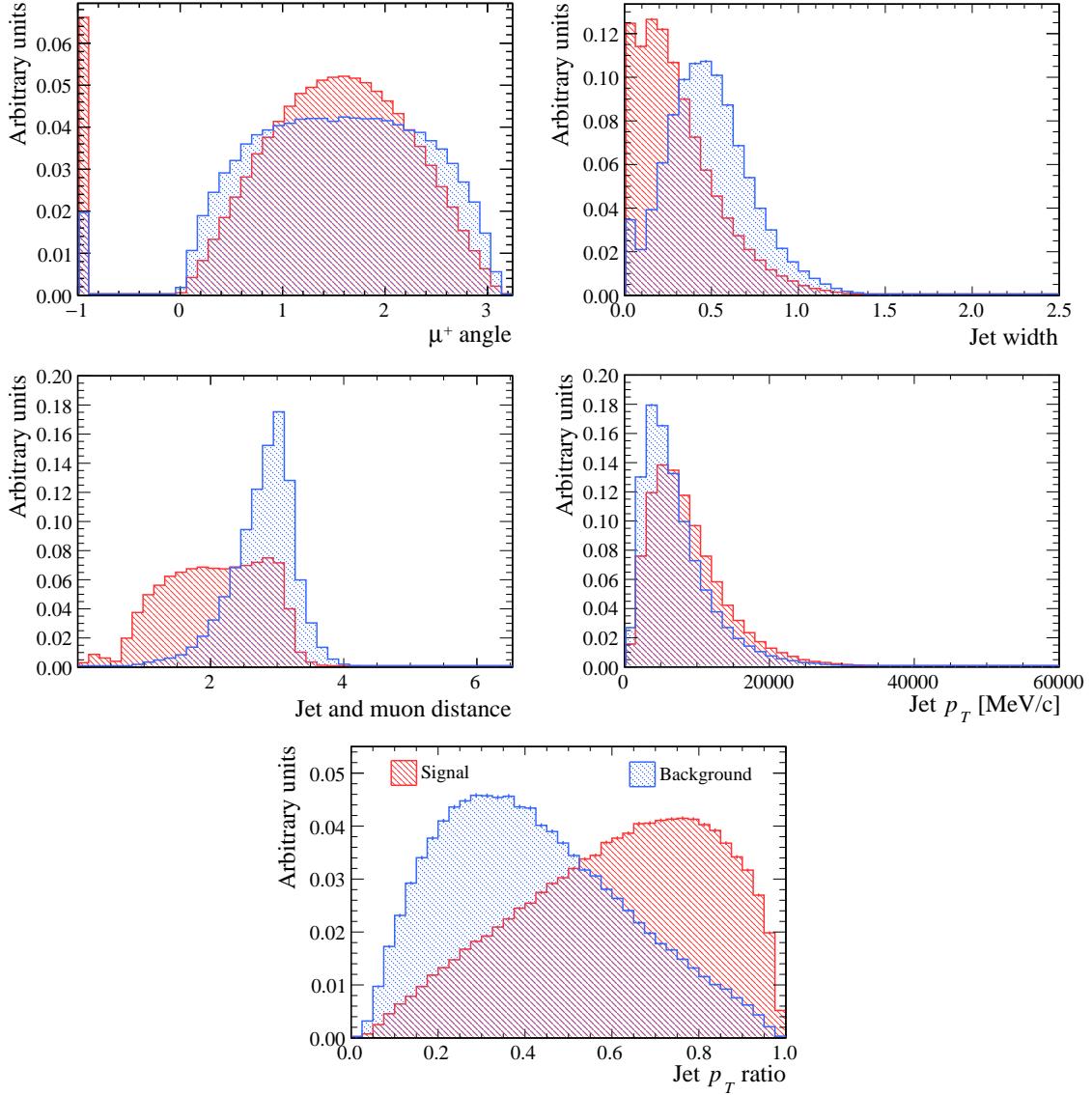


Fig. B.3 The distributions of the input variables used in only the uBoost BDT for $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu^+\mu^-X$ 2012 simulated decays.

different for the training and testing sets for signal and background decays. However, if the BDT is not overtrained the distributions will be similar for the training and testing sets.

Figure B.4 shows the results of this test where the responses for the training and testing samples lie on top of each other. Therefore, neither the uBoost BDT or the adaptive boosting BDT developed for the effective lifetime measurement are overtrained. The same test was performed for the global BDT developed of the branching fraction measurements and the global BDT is not overtrained.

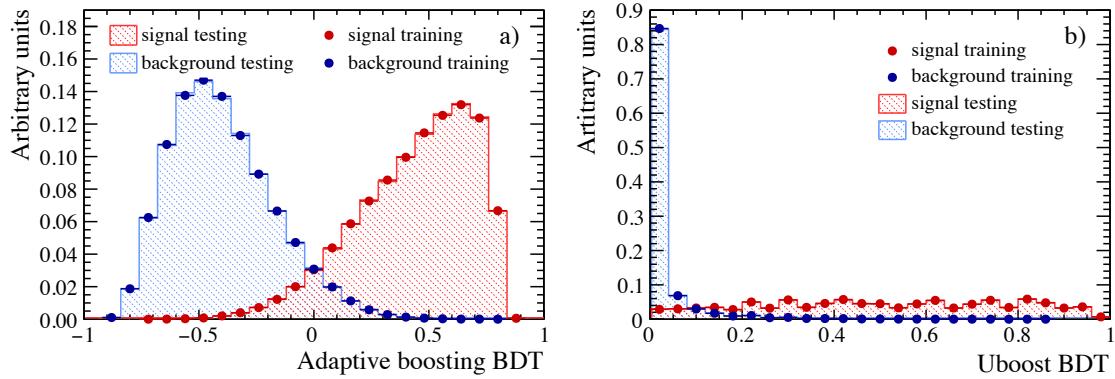


Fig. B.4 BDT response for training and testing samples of signal and background decays for a) the adaptive boost BDT and b) the uBoost BDT.