

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 two experiment runs of the Large Hadron Collider (LHC).

The LHCb experiment was designed to test the Standard Model of particle physics and 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 were made using data samples of proton-proton collisions with integrated luminosities of 1.0, 2.0 and  $1.4 \text{ fb}^{-1}$ , collected at centre-of-mass energies of 7, 8 and 13 TeV, respectively.

The branching fractions of the very rare  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays and the effective lifetime of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays are sensitive to effects from new physics. New physics processes could influence the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime and branching fraction independently, and therefore the two observables are complementary in the search for new physics.

The  $B_s^0 \rightarrow \mu^+ \mu^-$  decay is observed with a statistical significance of  $7.8\sigma$  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  $2.04 \pm 0.44 \text{ (stat)} \pm 0.05 \text{ (syst)} \text{ ps}$ . An upper limit is placed on the branching fraction  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$  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, it makes me feel awkward, but after almost 4 years working for my PhD there are many acknowledgements I would like to make. However I will keep this section quite brief. I would like to thank Valerie Gibson for giving me the opportunity to complete a PhD at the University of Cambridge, and for her guidance throughout my research and the repeated reading of this dissertation. Marc-Olivier Bettler and Harry Cliff, without their support and encouragement my PhD would never have ended and I would not have had the opportunity to work on one of the LHCb experiments flagship analyses. The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  LHCb analysis group, for the tireless effort that went into the branching fraction analysis and the guidance and support I received for my work. I have benefited a huge from the experience of those within this group. The Science and Technologies Faculties Council for funding my research, enabling me to spend a year living in Geneva and travel to international conferences. 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 rowing with Cambridge University Women's Boat Club. Whilst on the topic of rowing, I would like to thank again Valerie Gibson and Marc-Olivier Bettler for their support whilst I spent much of the first two years at Cambridge training for the Boat Races. Past and present members of the Cambridge HEP group, for the good friends I've made and making coffee and lunch breaks entertaining. Particular thanks go to John Hill and Steve Wotton for the computing infrastructure I greatly benefited from during my PhD and for putting up with me filling up a lot of disk space. Also thanks go to Susan Haines for answering my dumb questions and giving me hope that everything would be OK. My friends at Crossroad Church Saint Genis-Pouilly, for giving me a home and support whilst I lived at CERN. 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 my dissertation despite having no physics background. Finally, I would like to thank my parents and my

sister. 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 by 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  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays with the LHCb experiment.

The first three chapters set the scene for measurements described in the later chapters. Firstly, there is a short introduction to particle physics and current measurements of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays, then the theoretical motivation for studying these decays is described in Chapter 2. The research documented in this dissertation uses the Large Hadron Collider and the LHCb experiment which are described in the Chapter 3. The work presented in these chapters is not my own but summaries the work performed by others and the citations show where credit is due.

The study of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays has been undertaken using data from the LHCb experiment since it began operation, the latest measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions and the first measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime are described in the remaining chapters. 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 of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  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 and 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 part of the analysis strategy that are also used for the effective lifetime measurement. Although I did not directly perform the work documented in this chapter, my contributions were to the technical aspects of this measurement; producing the ROOT ntuples containing data and simulated decays; and maintaining the ‘stripping’ selection applied to data.

The measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime and the systematic uncertainties associated with the measurement are described in Chapters 6 and 7. The work documented in these chapters is result of my own effort but it uses inputs from the branching fraction

analysis which includes; 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. The results for the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fraction and  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime lifetime results are published in the paper [1].

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

I am grateful to have been a member of the LHCb collaboration over the last 4 years and to have been fortunate to work on the study of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays, one of the LHCb experiment's flagship analyses. Completing my PhD has been an experience and I am extremely appreciative of the opportunities being at the University of Cambridge has offered outside of time spent doing research. My time spent rowing with Cambridge University Women's Boat Club included some of the most challenging and enjoyable experiences of my life and Cambridge University Cycling Club showed me that sport can continue after rowing and perhaps be even more fun. Apart from sport, I have found great enjoyment in physics outreach events and undergraduate teaching, particularly seeing other people become excited about particle physics or understand new concepts. My year spent at CERN also showed me what is it really like to work at an international center of research. I am glad that my PhD is finally finished, although this dissertation documents the research outcomes of my PhD it does not do justice to all that has gone into my time at Cambridge.

# Table of contents

|  |            |
|--|------------|
| <b>Abstract</b>  | <b>iii</b> |
| <b>Acknowledgements</b>  | <b>vii</b> |
| <b>Preface</b>   | <b>ix</b>  |
| <b>1 Introduction</b>  | <b>1</b>   |
| <b>2 Theory of <math>B \rightarrow \mu^+\mu^-</math> decays; the Standard Model and beyond</b> | <b>7</b>   |
| 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 $B_{(s)}^0$ . . . . .  | 13         |
| 2.3.2 Impact on the Branching Fraction . . . . .   | 18         |
| 2.4 $A_{\Delta\Gamma}$ and the effective lifetime . . . . .                                    | 19         |
| 2.5 The Standard Model predictions . . . . .   | 20         |
| 2.6 New Physics models and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays . . . . .                 | 21         |
| <b>3 The LHC and the LHCb experiment</b>   | <b>25</b>  |
| 3.1 The LHC . . . . .  | 26         |
| 3.2 The LHCb experiment . . . . .  | 28         |
| 3.2.1 Tracking . . . . .   | 30         |
| 3.2.2 Particle identification . . . . .  | 38         |
| 3.2.3 Trigger . . . . .  | 44         |
| 3.2.4 Software and simulation . . . . .  | 49         |
| 3.3 Summary . . . . .  | 50         |
| <b>4 Event selection</b>   | <b>53</b>  |
| 4.1 Backgrounds . . . . .  | 54         |

---

|          |  |            |
|----------|--|------------|
| 4.2      | Simulated particle decays . . . . .  | 55         |
| 4.3      | Event selection for branching fraction measurements . . . . .                          | 57         |
| 4.3.1    | Trigger requirements . . . . .   | 58         |
| 4.3.2    | Cut-based selection . . . . .  | 59         |
| 4.3.3    | Particle identification . . . . .  | 72         |
| 4.3.4    | Multivariate Classifiers . . . . .   | 73         |
| 4.3.5    | Summary . . . . .  | 82         |
| 4.4      | Selection for the effective lifetime measurement . . . . .                             | 85         |
| 4.4.1    | Trigger requirements . . . . .   | 85         |
| 4.4.2    | Mass range . . . . .   | 86         |
| 4.4.3    | Particle identification . . . . .  | 87         |
| 4.4.4    | Multivariate classifier . . . . .  | 87         |
| 4.4.5    | Summary . . . . .  | 99         |
| <b>5</b> | <b>Measurement of <math>B \rightarrow \mu\mu</math> branching fractions</b>            | <b>103</b> |
| 5.1      | Analysis strategy . . . . .  | 103        |
| 5.2      | $B_{(s)}^0 \rightarrow \mu^+\mu^-$ mass and BDT PDFs . . . . .                         | 106        |
| 5.2.1    | Mass PDFs . . . . .  | 106        |
| 5.2.2    | BDT PDFs . . . . .   | 106        |
| 5.2.3    | Decay time dependence . . . . .  | 111        |
| 5.3      | Background mass PDFs and expected yields . . . . .                                     | 111        |
| 5.3.1    | $B \rightarrow h^+h'^-$ . . . . .  | 112        |
| 5.3.2    | Semi-leptonic decays . . . . .   | 113        |
| 5.3.3    | Combinatorial background . . . . .   | 114        |
| 5.4      | Normalisation . . . . .  | 114        |
| 5.4.1    | $B^0 \rightarrow K^+\pi^-$ and $B^+ \rightarrow J/\psi K^+$ yields . . . . .           | 115        |
| 5.4.2    | Efficiency ratio . . . . .   | 116        |
| 5.4.3    | Hadronisation factors . . . . .  | 117        |
| 5.4.4    | Normalisation parameters . . . . .   | 117        |
| 5.5      | Results . . . . .  | 118        |
| <b>6</b> | <b>Measurement of the <math>B_s^0 \rightarrow \mu^+\mu^-</math> effective lifetime</b> | <b>121</b> |
| 6.1      | Analysis strategy . . . . .  | 121        |
| 6.2      | Mass PDFs . . . . .  | 123        |
| 6.3      | Decay time PDFs . . . . .  | 124        |
| 6.3.1    | $B_s^0 \rightarrow \mu^+\mu^-$ . . . . .   | 125        |
| 6.3.2    | Background decay time PDFs . . . . .   | 133        |

|          |   |            |
|----------|---|------------|
| 6.4      | Fit optimisation . . . . .  | 137        |
| 6.4.1    | To fit for $\tau$ or $\tau^{-1}$ ? . . . . .                                | 138        |
| 6.4.2    | Optimisation results . . . . .  | 141        |
| 6.5      | Results . . . . .   | 147        |
| <b>7</b> | <b>Systematic uncertainties and cross checks</b>                            | <b>149</b> |
| 7.1      | Fit Accuracy . . . . .  | 149        |
| 7.1.1    | Fit stability with $\tau_{\mu\mu}$ values . . . . .                         | 150        |
| 7.1.2    | $B_s^0 \rightarrow \mu^+ \mu^-$ yields . . . . .                            | 151        |
| 7.1.3    | Overall bias on $\tau_{\mu\mu}$ . . . . .                                   | 151        |
| 7.2      | Background contamination . . . . .  | 153        |
| 7.3      | Mass PDF parameters . . . . .   | 156        |
| 7.4      | Acceptance function accuracy . . . . .                                      | 156        |
| 7.5      | Incorrectly assigned primary vertices and the detector resolution . . . . . | 162        |
| 7.6      | Combinatorial background decay time model . . . . .                         | 163        |
| 7.7      | Mix of $B_s^0$ mass eigenstates . . . . .                                   | 166        |
| 7.8      | Production asymmetry of $B_s^0$ and $\bar{B}_s^0$ . . . . .                 | 168        |
| 7.9      | Summary . . . . .   | 170        |
| <b>8</b> | <b>Summary and Outlook</b>  | <b>171</b> |
| 8.1      | Summary . . . . .   | 171        |
| 8.2      | Outlook . . . . .   | 172        |
|          | <b>Bibliography</b>   | <b>175</b> |
|          | <b>A Distributions of input variables for the global BDT</b>                | <b>185</b> |
|          | <b>B Multivariate classifier development</b>                                | <b>189</b> |
| B.1      | Input variables . . . . .   | 189        |
| B.2      | Training parameters . . . . .   | 192        |
| B.3      | Overtraining test . . . . .   | 192        |
|          | <b>C Details of pseudoexperiment configuration</b>                          | <b>197</b> |



# 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 made up of 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 ( $\bar{u}$ ,  $\bar{d}$ ,  $\bar{s}$ ,  $\bar{c}$ ,  $\bar{t}$ ,  $\bar{b}$ ) 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 ( $e^+$ ,  $\mu^+$ ,  $\tau^+$ ,  $\bar{\nu}_e$ ,  $\bar{\nu}_\mu$ ,  $\bar{\nu}_\tau$ ) however unlike quarks, no leptons can 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 ( $\gamma$ ), the weak force by the  $W^\pm$  and  $Z^0$  bosons and the strong force by eight gluons ( $g$ ). This 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 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 from one type into another [5–8] and it does not provide a particle or mechanism that could account of the observed presence of dark matter and dark energy in the universe [9–12]. The SM includes three fundamental forces but the final force, the gravitational force, is not included in its formalism. 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 to this asymmetry [13, 14]. As well as experimental observations

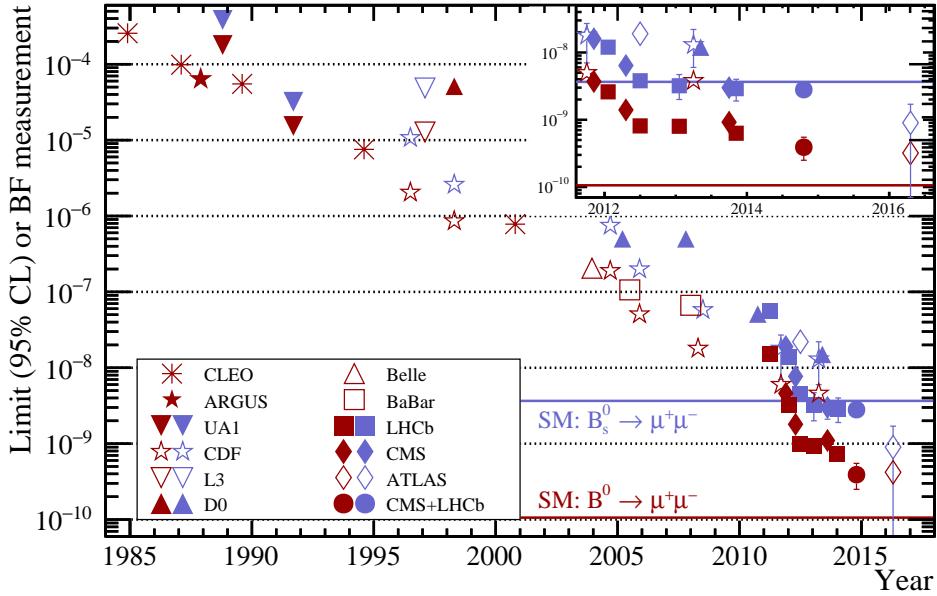
there are more fundamental questions about the SM that are unanswered. There is no reason to explain why there are very large differences in the coupling strengths of the electromagnetic, weak and strong forces or why there is a large range in the masses of quarks and leptons. The examples given here 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 [15]. A more complete discussion of the shortcomings of the SM can be found in reference [16, 17].

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 can collectively be called New Physics (NP). At the moment there is no clear indication of which beyond the SM (BSM) theory gives 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 of the predictions 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 found in 2012 by the ATLAS and CMS collaborations using this type of search [18, 19] but no NP has been observed from direct searches yet. 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 constraints the parameters 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 the measured results in heavy flavour physics [20–29]. Although these deviations are far from conclusive evidence of NP, it will be very interesting to see if and how these measurements change in the future.

Particle decays and interactions that are suppressed in the SM offer excellent places for indirect searches for NP. The possible contributions from BSM theories can be at a similar order of magnitude to the SM contributions in these processes. The rare decays of  $B^0$  and  $B_s^0$  mesons into two oppositely charged muons are two examples of such processes and have long been interesting studied as a test of the SM. The purely leptonic final states



**Fig. 1.1** Results from searches for  $B^0 \rightarrow \mu^+\mu^-$  (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 from reference [40] and has been updated to include the latest result from the ATLAS experiment [31].

leads to precise theoretical predictions for the branching fractions of these decays and the 2 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 in Figure 1.1. The latest experiments to join the search were the ATLAS, CMS and the LHCb experiments [30–40]. 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 [38]. 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  $4.0\sigma$  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 [39]. The 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. Producing a measurement of the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction of  $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$  at a statistical significance of  $4.3\sigma$  and placing 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 [34]. The combined analysis of the CMS and LHCb data sets resulted in the first observation of  $B_s^0 \rightarrow \mu^+\mu^-$

decays and the first evidence of  $B^0 \rightarrow \mu^+ \mu^-$  [40]. The measured branching fractions were

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

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

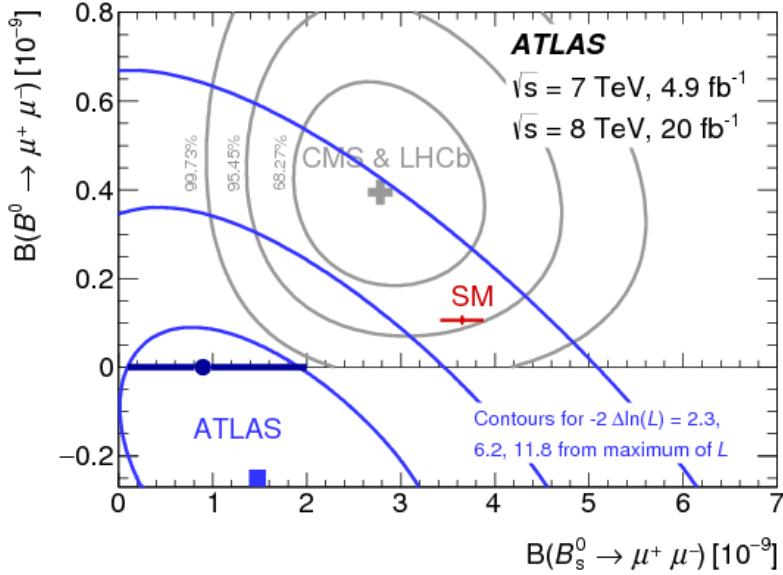
with a statistical significance of  $6.2\sigma$  for the  $B_s^0$  and  $3.0\sigma$  for the  $B^0$ . The ATLAS experiment also searched for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decay using data collected during the same period [31], measuring the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction 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  $2\sigma$ . 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 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 both the separate measurements and each measurement and the SM prediction as shown in Figure 1.2. Therefore the study of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays continues to be an very interesting topic in the search for NP effects.

The data collected during from  $pp$  collisions with centre-of-mass energies of  $pp$  13 TeV at the LHC, will enable more precise measurements of the branching fractions of these decays to be made. Furthermore 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, the presence of NP could be revealed in either both or only one of these measurements. The search for  $B_s^0 \rightarrow \mu^+ \mu^-$  decays is over and the 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 fraction 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. The theoretical motivation for studying these decays is given in Chapter 2 and the LHC and LHCb experiment are described in Chapter 3. The criteria used to identify these decays in the data are detailed in Chapter 4 and the measurement of the branching fraction is briefly described in Chapter 5. The measurement of the



**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 [31]. The measurements were performed using data collected during 2011 and 2012 and centre-of-mass energies of 7 and 8 TeV, respectively.

$B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime is discussed in Chapter 6 and the systematic uncertainties of this measurement are given in Chapter 7. Finally a summary of the results and prospects for future measurements of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays are given in Chapter 8.



# Chapter 2

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

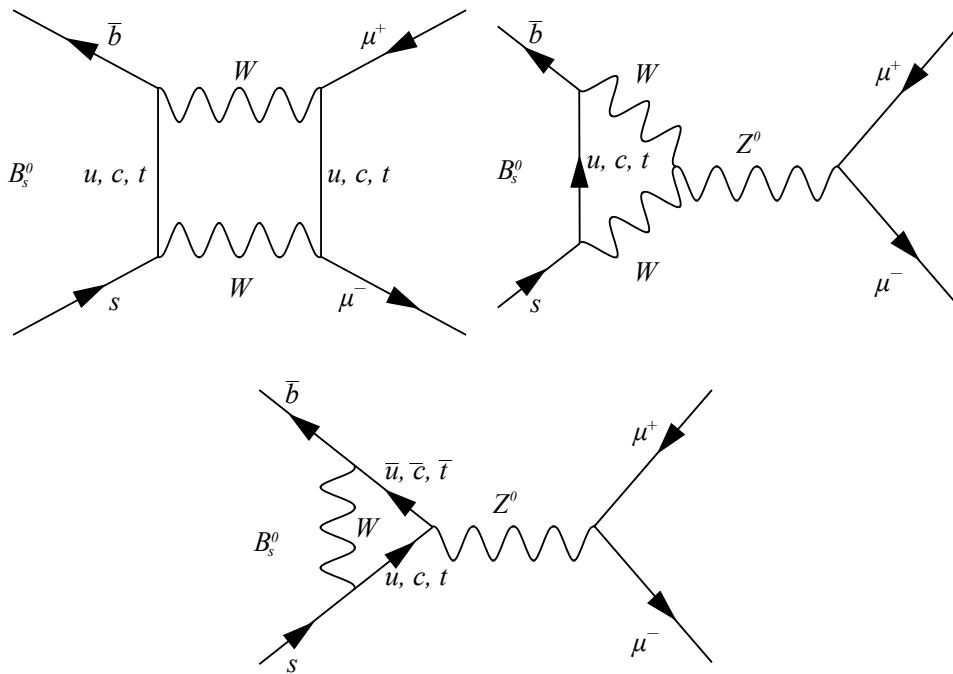
This Chapter describes the theoretical motivation for the study of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays. The description of these decays within the SM framework is presented in Section 2.1 and the determination of the theoretical predictions of branching fractions is outlined in Section 2.2. The discussion of the theoretical branching fractions is based on references [41, 42]. Quark mixing leads to oscillations of a  $B_s^0$  to a  $\bar{B}_s^0$  over time and 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 [43, 42, 44]. A new parameter,  $A_{\Delta\Gamma}$ , arises from the  $B_s^0$ - $\bar{B}_s^0$  oscillations 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 branching fractions and the 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  $B_s^0$ , are made up 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 flavour 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 the  $B_{(s)}^0$  decays into two

oppositely charged muons as  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ <sup>1</sup>. This decay mode occurs very rarely in the SM compared to other decay modes of the  $B_{(s)}^0$ , the suppression of this mode arises from several different sources.

The composite quarks of a  $B_{(s)}^0$  both have the same charge, 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  $W$ -box and  $Z^0$ -penguin diagrams as shown in Figure 2.1.



**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 but the  $s$  quark is exchanged for a  $d$  quark.

The decays can also proceed via Higgs-penguin diagrams however the contributions from these diagrams are negligible. 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.

---

<sup>1</sup> $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  refers to both the particle and anti-particle decays of the  $B^0$  and  $B_s^0$  unless otherwise specified.

Although the weak force allows quark flavour to change the coupling strengths between different quark flavours are not all the same magnitude. The coupling strengths are described by the CKM matrix. Quarks can be separated into two types depending on their charge; up-type quarks include  $u$ ,  $c$  and  $t$ , 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 [45, 46] 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 which ensures 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 [47], 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 one off diagonal element is needed to describe the quark transitions in Figure 2.1. Therefore 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 and because the coupling strength of the  $b$  quark to any quark except the  $t$  is extremely small.

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  $\mu^-$  and a  $\mu^+$  with opposite helicities which cannot conserve angular momentum in  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays. Muons are not massless therefore  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decay can occur but are suppressed because  $m_\mu \ll M_{B_{(s)}}$  [48] 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 at 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 decay rate, which is the probability per unit time that a  $B_{(s)}^0$  decays into two muons, as

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

The SM predictions are calculated from the ‘prompt’ decay rate that ignore any evolution with time of the  $B_{(s)}^0$  particles. The branching fractions are calculated using [49]

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

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 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 branching fractions are calculated this way to enable easy comparison of different  $B$  meson branching fractions including  $B^0$ ,  $B_s^0$  and  $B^+$  [49].

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

give [50]

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

where  $m_\mu$  and  $M_{B_{(s)}}$  are the masses of the muon and the  $B_{(s)}^0$ , respectively. The factor of  $\frac{m_\mu}{M_{B_{(s)}}}$  comes 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 [48]. The Operator Product Expansion [51, 52] is used to create the effective Hamiltonian,  $\mathcal{H}_{eff}$ , which splits the interaction into two energy levels. The transition amplitude then becomes

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

where  $G_F$  is the Fermi coupling constant,  $V_{CKM}^i$  are CKM matrix elements,  $C_i$  are Wilson coefficients and  $\mathcal{O}_i$  are local operators. The energy scale  $\mu$  separates the two energy levels in the interaction. The Wilson coefficients describe short scale processes with energies above  $\mu$ . 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  $\mu$ . 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  $\mu$  is arbitrary however the final transition amplitude must be independent of  $\mu$ , 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, therefore the same coefficients and operators can be used to describe the  $B^0$  and  $B_s^0$  decays. The effective Hamiltonian for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays is [53]

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

where  $\alpha$  is the fine structure constant and  $q$  corresponds to the  $d$  quark in the  $B^0$  or the  $s$  quark in the  $B_s^0$ . Terms proportional to  $V_{cq}^* V_{cb}$  and  $V_{cq}^* V_{ub}$  neglected because they are negligible. The operators that can contribute to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  effective Hamiltonian

due to the initial and final decay states are

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

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

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

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 is negligible in the SM.

The purely leptonic final state of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  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 [54]

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

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

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

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

In the SM  $P = 1$  and  $S = 0$ , however 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  $S > 0$ , also the contributions from scalar particles are not subject to helicity constraints. 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. The ratio of branching fractions is given by

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

and the uncertainty on the ratio is less than the individual branching fractions because sources of uncertainties including those from Wilson coefficients and  $|V_{tb}|$  cancel out. The ratio does not depend on Wilson coefficients and provides an excellent observable to test the flavour structure of the SM and BSM theories.

## 2.3 Quark mixing

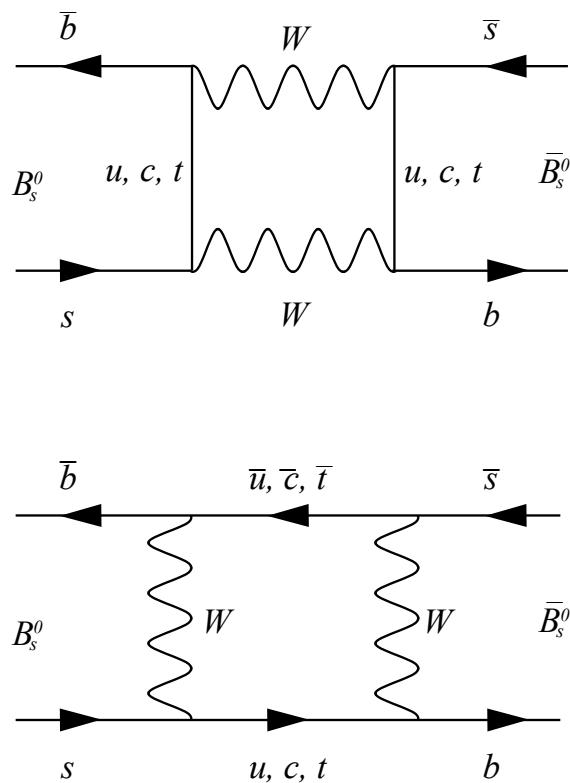
The theoretical prediction for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions does not take into account the evolution of the  $B_{(s)}^0$  and  $\bar{B}_{(s)}^0$  mesons with time. Once a  $B_{(s)}^0$  is created it will oscillate between the particle and anti-particle states as it propagates 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 the  $\bar{B}_{(s)}^0$ . These 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 as 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 [49]

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

Therefore for a meaningful comparison between the measured and predicted branching fraction values, the difference in the two definitions must be evaluated [53, 54, 49].

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

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  $|\bar{B}_{(s)}^0\rangle$ . In order to evaluate the time integrated branching fractions the



**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.

evolution of these states with time must be evaluated. The time dependant Schrödinger equation (TDSE) describes the time evolution of the particle and anti-particle states as

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

$\mathbf{M}$  and  $\Gamma$  are  $2 \times 2$  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 given by heavy,  $H$ , and light,  $L$ , mass states 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 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  $|\bar{B}_s^0\rangle$ . The difference  $\Delta m_s$  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  $\Delta m_s$  and  $\Delta \Gamma_s$  are measurable quantities.

Theoretical predictions can be calculated for  $M_{12}$  and  $\Gamma_{12}$  therefore it is useful to express the measurable quantities in terms of them. 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 relationship  $\Delta \Gamma \ll \Delta m$  as well as  $\Gamma_{12} \ll M_{12}$  [44] are used to separate the expressions for  $\Delta m$  and  $\Delta \Gamma$  to give

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

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

The values of  $p$  and  $q$  can also be related to the measurable quantities and  $\Gamma_{12}$  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  $\phi_M$  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.

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 therefore the time dependant decay rates can now be evaluated. The decay rates can be expressed as

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

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 parameters is defined

$$\lambda_{\mu\mu} = \frac{q}{p} \left| \frac{\bar{A}_{\mu\mu}}{A_{\mu\mu}} \right| \quad (2.31)$$

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 ?? and using  $\lambda_{\mu\mu}$  the time dependant decay rates are

$$\begin{aligned} \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) &= \frac{1}{2} \mathcal{N} |A_{\mu\mu}|^2 e^{-\Gamma_s t} \left\{ (1 + |\lambda_{\mu\mu}|^2) \cosh \left( \frac{\Delta\Gamma_s t}{2} \right) + (1 - |\lambda_{\mu\mu}|^2) \cos(\Delta m_s t) \right. \\ &\quad \left. - 2\text{Re}(\lambda_{\mu\mu}) \sinh \left( \frac{\Delta\Gamma_s t}{2} \right) - 2\text{Im}(\lambda_{\mu\mu}) \sin(\Delta m_s t) \right\} \end{aligned} \quad (2.32)$$

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

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 2.33 and ignoring terms  $\mathcal{O}(a)$  the total decay rate is

$$\langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle = \mathcal{N} |A_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) e^{-\Gamma_s t} \left( \cosh \left( \frac{\Delta\Gamma_s t}{2} \right) + A_{\Delta\Gamma} \sinh \left( \frac{\Delta\Gamma_s t}{2} \right) \right). \quad (2.34)$$

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

$$A_{\Delta\Gamma} = \frac{2\text{Re}(\lambda_{\mu\mu})}{1 + |\lambda_{\mu\mu}|^2}. \quad (2.35)$$

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

$$\begin{aligned}\langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle &= \mathcal{N} |A_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) \left( (1 - A_{\Delta\Gamma}) e^{-\Gamma_L t} + (1 + A_{\Delta\Gamma}) e^{-\Gamma_H t} \right) \\ &= R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t}\end{aligned}\quad (2.36)$$

The 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

$$A_{\Delta\Gamma} = \frac{R_H - R_L}{R_H + R_L}. \quad (2.37)$$

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  $B_s^0 \rightarrow \mu^+ \mu^-$ .

### 2.3.2 Impact on the Branching Fraction

The time dependant decay rates are used to understand the difference between the two branching fraction definitions. The final form of the decay rates in Equation 2.36 is used in the evaluations of the branching fractions. The ‘prompt’ branching fraction used in the theoretical predictions is

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{th}} = \frac{\tau_{B_s}}{2} \langle \Gamma(B_s^0 \rightarrow \mu^+ \mu^-) \rangle \quad (2.38)$$

$$= \frac{\tau_{B_s}}{2} (R_H + R_L). \quad (2.39)$$

The time integrated branching fraction that is measured is

$$\begin{aligned}\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} &= \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \rangle dt \\ &= \frac{1}{2} \left( \frac{R_H}{\Gamma_H} + \frac{R_L}{\Gamma_L} \right) \\ &= \frac{\tau_{B_{(s)}}}{2} (R_H + R_L) \left[ \frac{1 + A_{\Delta\Gamma} y_{(s)}}{1 - y_{(s)}^2} \right]\end{aligned}\quad (2.40)$$

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

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} = \left[ \frac{1 + A_{\Delta\Gamma} y_{(s)}}{1 - y_{(s)}^2} \right] \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{th}} \quad (2.41)$$

For the  $B^0$ - $\bar{B}^0$  oscillations the difference in the lifetimes of the heavy and light mass eigenstates is extremely small therefore  $y$  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  $y_s = 0.062 \pm 0.006$  [55]. 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 effective lifetime

The definition of  $A_{\Delta\Gamma}$  in Equation 2.35 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 [53]

$$A_{\Delta\Gamma} = \frac{|P| \cos \varphi_P + |S| \cos \varphi_S}{|P|^2 + |S|^2} \quad (2.42)$$

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 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 effects in 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 needs 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 [49]. The

effective lifetime is the mean decay time of an untagged sample of  $B_s^0 \rightarrow \mu^+\mu^-$  decays, defined as [56]

$$\tau_{\mu\mu} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0 \rightarrow \mu^+\mu^-) \rangle dt}. \quad (2.43)$$

It can be measured by fitting a single exponential to the same set of decays used to measure the branching fraction [49]. The effective lifetime can be expressed in terms of  $A_{\Delta\Gamma}$  using the decay rates in Equation 2.36 as

$$\tau_{\mu\mu} = \frac{\tau_{B_s}}{(1 - y_s^2)} \frac{(1 + 2A_{\Delta\Gamma}y_s + y_s^2)}{(1 + A_{\Delta\Gamma}y_s)}. \quad (2.44)$$

In the SM only the heavy  $B_s^0$  mass eigenstates 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 dependant of the effective lifetime on  $A_{\Delta\Gamma}$ .

## 2.5 The Standard Model predictions

The SM provides precise predictions of the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fractions of [57–59]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.57 \pm 0.16) \times 10^{-9} \quad (2.45)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (1.02 \pm 0.06) \times 10^{-10} \quad (2.46)$$

where quark oscillations have been accounted for in the quoted value of the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction. 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 [40]

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)} = 0.0295^{+0.0028}_{-0.0025}. \quad (2.47)$$

In the SM the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime is predicted to be [59]

$$\tau_{\mu\mu} = 1.61 \pm 0.01 \text{ ps}. \quad (2.48)$$

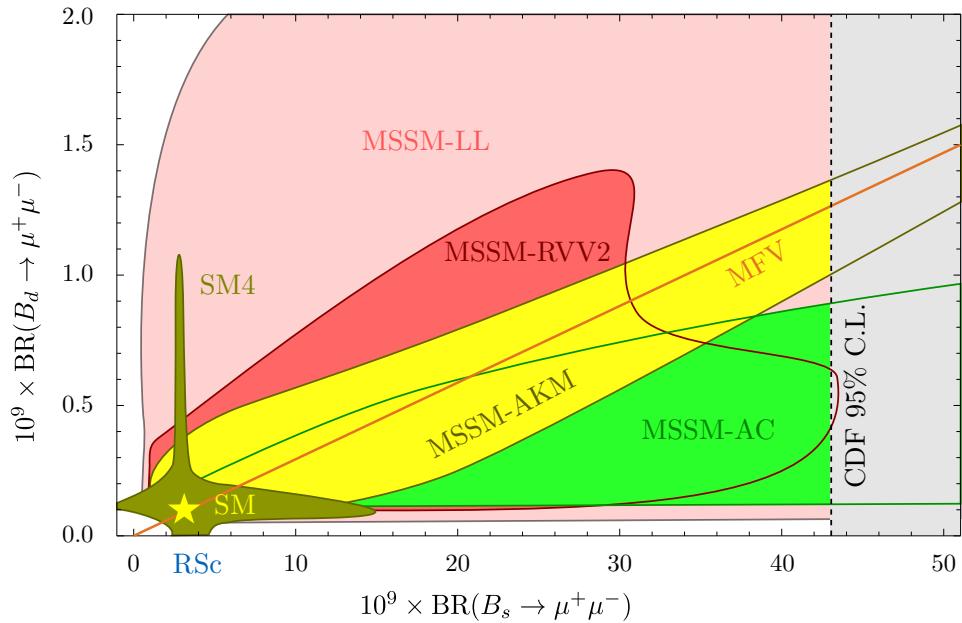
The Heavy Flavour Averaging group provides the world averages for the measured values of the lifetime of the light and heavy  $B_s^0$  mass eigenstates to be  $\tau_H = 1.609 \pm 0.010$  ps

and  $\tau_L = 1.414 \pm 0.006$  ps [55]. The difference in these lifetimes is 0.195 ps. Therefore a precision of 0.38 ps would be needed to distinguish between  $A_{\Delta\Gamma} = +1$  and  $A_{\Delta\Gamma} = -1$  a  $5\sigma$  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  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime can constrain the parameter space available for NP and could reveal which theories are the correct extension of the SM. This section will briefly introduce how some NP models that could still be seen with  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  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 [54, 60, 61].

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) [62] hypothesis is included in Figure 2.3. The



**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 (MSSM) [63] and the SM extended to constrain four generations of fermions (SM4) [64]. Figure is taken from [65].

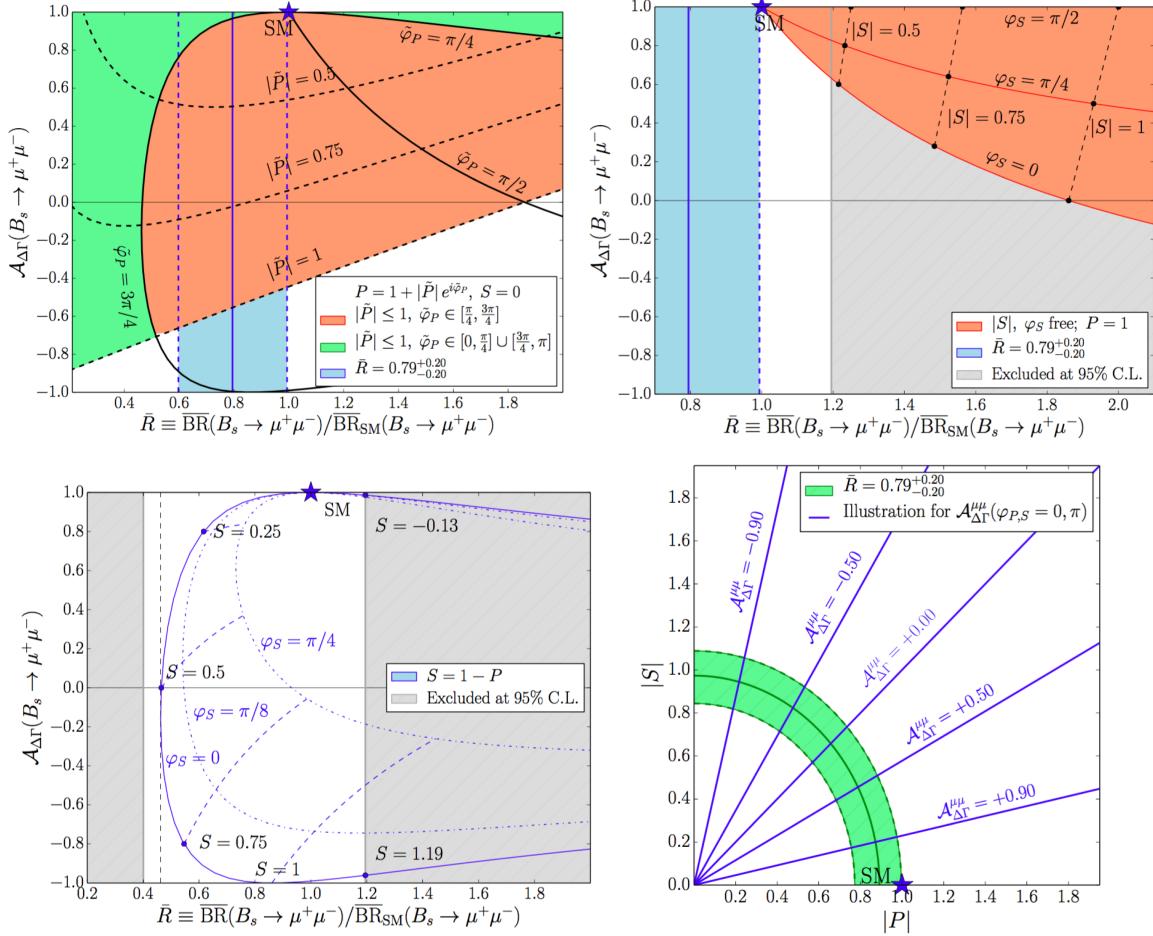
MFV hypothesis predicts that the coupling of quark flavour and  $\mathcal{CP}$  violation follow the same Yukawa structure as the SM in NP models. It is a popular theory to describe the flavour structure in NP models due to the current agreement of measurements with the SM predictions. A significant deviation of the branching fraction ratio from the SM or MFV hypothesis predictions would indicate the need to 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 can contribute to the decays. 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 2.42, 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 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  $A_{\Delta\Gamma}$ . Furthermore in some scenarios  $A_{\Delta\Gamma}$  is needed to resolve degeneracies that arise from information from the branching fraction alone.

Amongst the BSM theories that can influence the values of  $P$  and  $S$  are the two Higgs doublet model (2HDM) [67], supersymmetric models [68], models including leptoquarks and models that obey the MFV hypothesis as mentioned earlier.

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 2 charged, one neutral pseudoscalar and 2 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  $P \pm S = 1$  [54, 60] and the corresponding branching fraction and  $A_{\Delta\Gamma}$  values for  $B_s^0 \rightarrow \mu^+ \mu^-$  decays as shown in Figure 2.4.

Supersymmetric (SUSY) models extend the SM by giving each SM particle a supersymmetric partner. The resulting theory is symmetric under the transformation of fermions to boson and bosons to fermions. So far no evidence for SUSY particles has been found therefore the symmetry must be broken and the mass of SUSY particles is 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 expectations values of the Higgs doublet is large [69–71]. The MSSM can



**Fig. 2.4** Allowed values for  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  and  $A_{\Delta\Gamma}$  for situations where  $S = 0$  (top left),  $P = 1$  (top right),  $P \pm S = 1$  (bottom left) and  $\varphi_P, \varphi_S \in 0, \pi$  (bottom right) [54, 60]. The ratio  $\bar{R}$  plotted is from an average of the individual results from the CMS and LHCb collaborations from [66], the results from the combined analysis of the CMS and LHCb data gives  $\bar{R} = 0.76^{+0.2}_{-0.18}$  [40].

produce values for  $A_{\Delta\Gamma}$  and the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction shown in Figure 2.4 for situations where  $P \pm S = 1$  and  $\varphi_P, \varphi_S \in [0, \pi]$  [54, 60].

Models including leptoquarks are currently popular to explain the anomalies observed in heavy flavour measurements [72–76]. A leptoquark is a boson that carries both lepton and baryons numbers. The exact quantum numbers depend on the interactions with SM fermions and leptoquarks can enable FCNCs to occur at the tree level. 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 with just the branching fraction measurements [77].

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 prove important information, complementary to the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fraction measurements to 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 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 400 GeV. The most notable achievements of the SPS were the discoveries of the  $W$  and  $Z^0$  bosons in 1983 when the accelerator had been converted into a protons - anti-proton collider. Precise measurement of the  $W$  and  $Z^0$  bosons and the electroweak force 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 at 100 GeV. Now 62 years since its foundation, CERN has grown to include 22 member states<sup>1</sup> 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 ever built. It was built in the tunnel that housed LEP and the LHC was designed to collide protons at a centre-of-mass energy of 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.

---

<sup>1</sup>Countries and organisations that are unable to become member states can still participate in scientific research as observer states [78].

### 3.1 The LHC

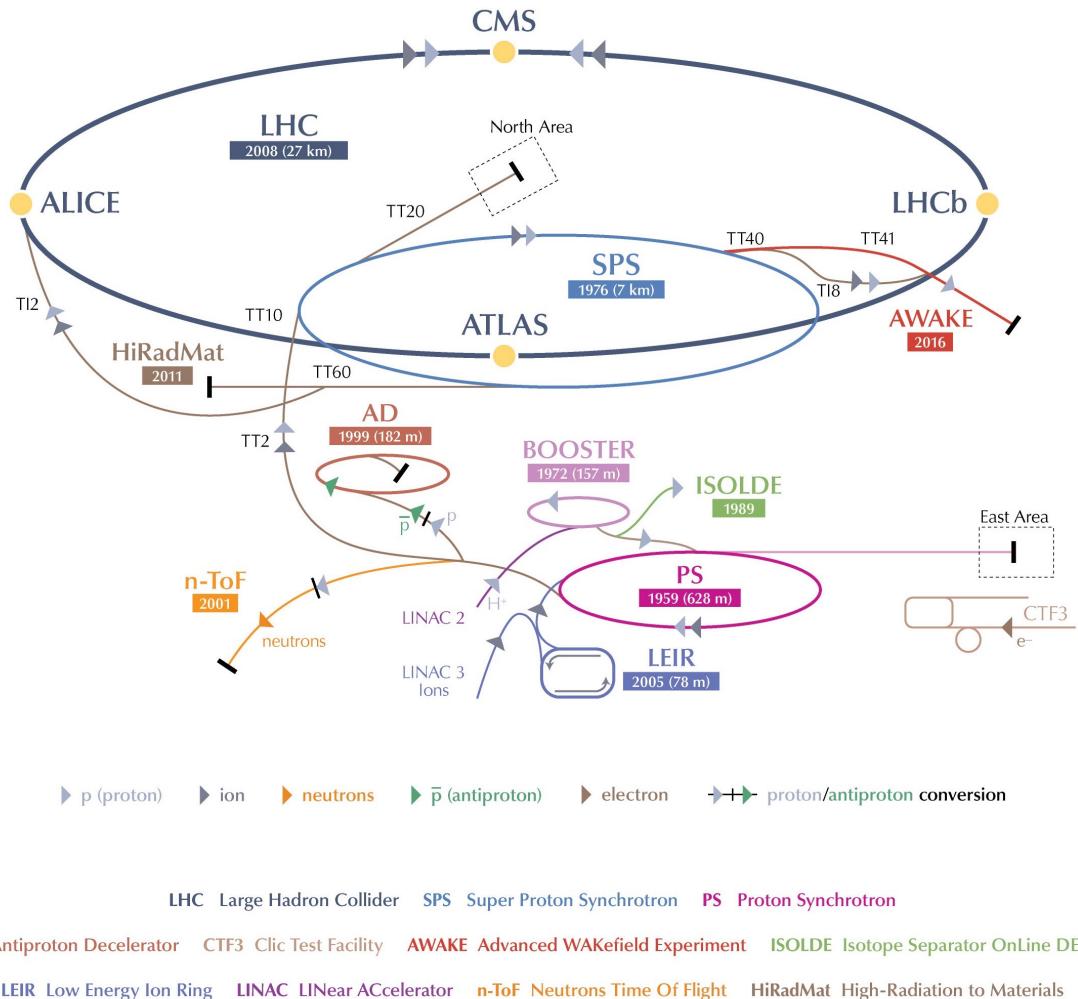
The LHC is a proton synchrotron designed to accelerate and collide two beams of protons with a centre-of-mass energy of 14 TeV. Although operation of the LHC began in 2010 it is yet to reach the design energy. The purpose of the LHC is to provide high energy  $pp$  collisions, the products of which are used for precision tests of the Standard Model (SM) and to search for new physics effects that cannot be explained within the context of the SM. There are four interaction points on the LHC ring where the beams are brought to collide, 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 proton 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 and then the protons are 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 of 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 radio frequency cavities and guides them around the ring with superconducting dipole magnets. Once the required energy has been reached, the bunches are focused using quadrupole magnets before being collided at 4 interaction points around the LHC ring.

The centre-of-mass energy of a collider is an important measure of its performance as it describes the energy available to create new 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, it 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



**Fig. 3.1** The accelerator complex at CERN. The chain of accelerators used to inject protons into 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 accelerating protons to 450 GeV ready for injection into the LHC. Source: CERN.

bunch. The higher the luminosity, the more collisions happen in a second and the more particles will be produced, this can either be advantageous or disadvantageous depending on the physics process that is being studied. Therefore the luminosity delivered at each interaction point can be tuned using the quadrupole magnets by altering the shape of each bunch to suit the experiments at each point.

There are 7 experiments on the LHC that detect particles produced in proton and heavy ion collisions. There are two general purpose detectors, ATLAS and CMS, that were designed to search for the Higgs boson and new effects that are beyond the scope of the SM. These two experiments operate at the full instantaneous luminosity of the LHC. ALICE studies quark-gluon plasma produced in heavy ion collisions to understand conditions similar to those present in the early universe. The TOTEM experiment studies properties of protons as they collide head on at the LHC and the MOEDAL experiment is aimed to detect magnetic monopoles. The LHCf experiment studies particles that are created at very small angles to the incident proton beams to understand similar processes that occur in cosmic rays. The final experiment of the LHCb experiment that was designed to study rare  $b$ -hadron decays and  $\mathcal{CP}$  violating processes and operates at a lower luminosity and has a smaller angular acceptance than the general purpose detectors.

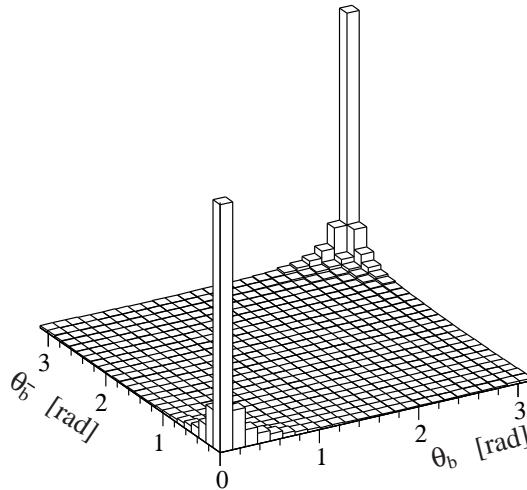
Proton beams first circulated the LHC in 2008 and since then the experiments have recorded data during two physics runs separated by a long shutdown period. Run 1 began in 2010 and continued until 2013, during this time protons were collided with a centre-of-mass energies of  $\sqrt{s} = 7$  and  $\sqrt{s} = 8$  TeV. After Run 1 there was a period of long shut down (LS1) during which work was done to prepare the LHC to operate at higher energies and renovation work was preformed on the accelerators that provide the LHC with protons. Run 2 began in 2015 with proton collisions at a centre-of-mass energy of  $\sqrt{s} = 13$  TeV, this run will continue until 2018 when a second period of upgrades and maintenance, the Long Shutdown 2, 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 for the experiments.

## 3.2 The LHCb experiment

The LHCb experiment was built to study the SM and search for new physics 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. The  $b\bar{b}$  pairs produced travel at small angles

| Run   | Year | $\sqrt{s}$ / TeV | Integrated luminosity / $\text{fb}^{-1}$ |       |      |
|-------|------|------------------|--|-------|------|
|       |      |                  | 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 |

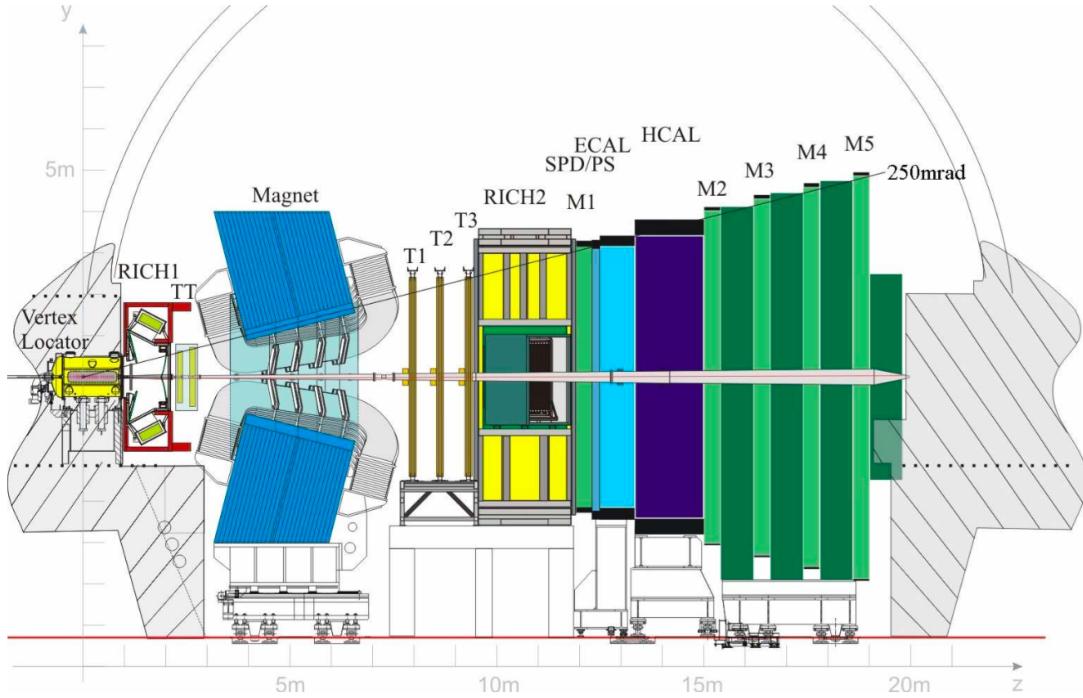
**Table 3.1** Centre-of-mass energy for each year of data taking and the integrated luminosity recorded by ATLAS, CMS and LHCb during  $pp$  collisions [79–81].



**Fig. 3.2** Simulated angular distribution for  $b\bar{b}$  production at the LHC, angles are relative to the beam pipe with  $\theta = 0$  in the forward direction and  $\theta = \pi$  in the backward direction [82].

relative to the beam pipe as shown in Figure 3.2 and hadronize to form a range of  $b$ -hadrons, including  $B^+$ ,  $B_s^0$  and  $\Lambda_b^0$ , that are studied by LHCb.

The LHCb experiment was built as 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. The angular coverage was chosen to exploit the small angles at which  $b\bar{b}$  pairs are produced. A cross-section of the LHCb detector is shown in Figure 3.3, where a right handed coordinate system is used. Protons collide at the interaction point on the left hand side of the diagram, the products of the collisions travel through the detector leaving information in the sub-detectors along the



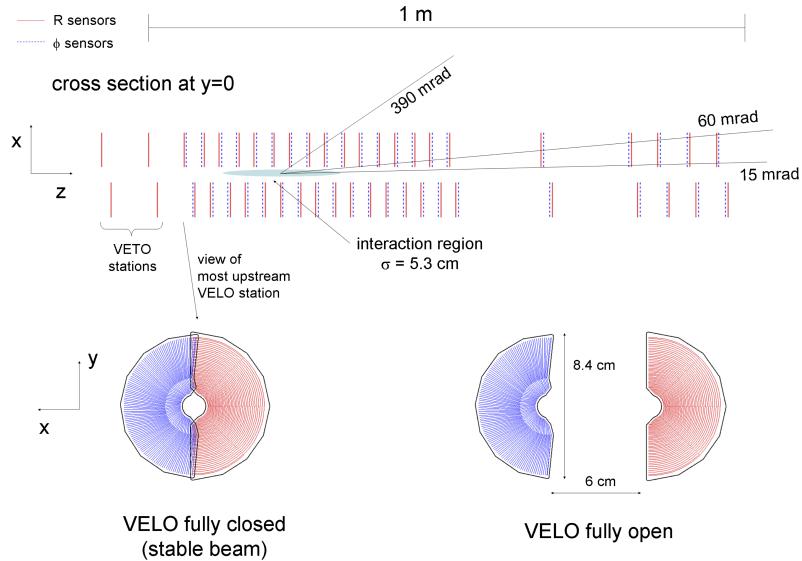
**Fig. 3.3** Cross section of the LHCb detector [83].

length of the detector. The information deposited in the sub-detectors is reconstructed to determine what happened during the  $pp$  collisions.

The different sub-detectors have been chosen to exploit the characteristics of  $b$ -hadron decays and fall into 2 distinct categories; tracking detectors and particle identification detectors. Each sub-detector and its performance are described 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 [84, 85].

### 3.2.1 Tracking

The tracking system within the LHCb experiment consists of the vertex locator (VELO), a dipole magnet and the tracking stations. Together, the sub-detectors provide precise information on the passage of charged particles through the detector and a measurement of the particle momentum. 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 atoms in the material. The release of this energy is recorded and translated into an electrical signal that reveals the path of a particle.



**Fig. 3.4** The VELO layout and position of sensors along the beam axis [83].

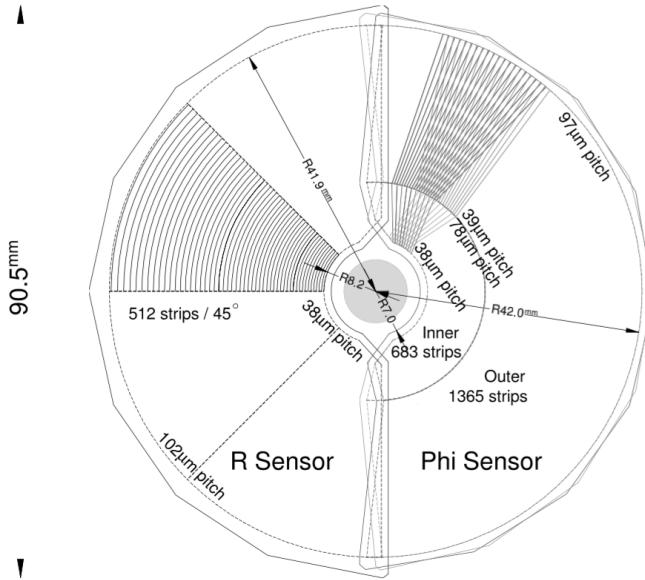
### 3.2.1.1 The vertex locator

The VErtex LOcator (VELO) is a silicon detector surrounding the interaction point. Its main goal is to provide precise information about the  $pp$  interaction vertices and secondary decay vertices of particles produced. Information from the VELO enables precise measurements of particle lifetimes and impact parameters of particle tracks necessary for physics analyses.

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 3.4, 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 the radial distance of charged particles from the beam axis; and the  $\phi$  sensor that measures the azimuthal angle of the particle. The information from the sensors is combined with the  $z$  placement of the sensors is used to reconstruct charged particle trajectories. The cylindrical geometry used for the VELO sensors, shown in Figure 3.5, is chosen to allow fast reconstruction of particle trajectories in the VELO.

The momentum resolution achievable for charged tracks by the LHCb experiment is limited by multiple scattering of particles as they travel through material in the detector.

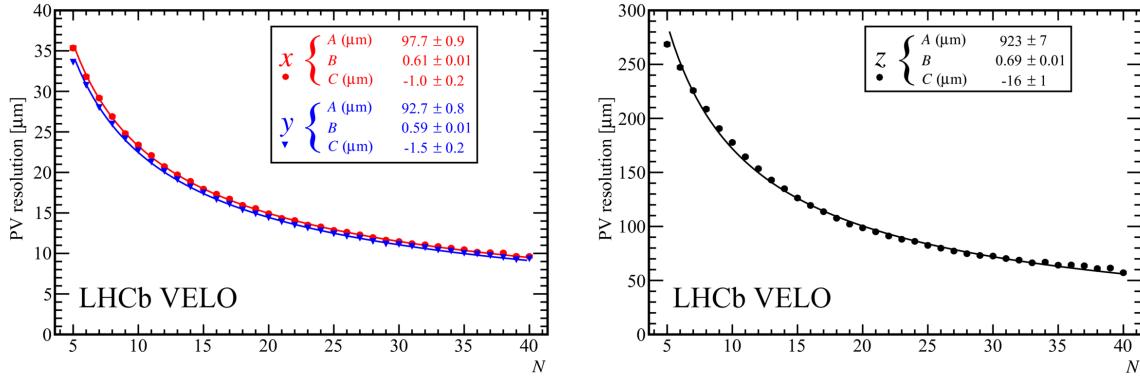


**Fig. 3.5** Diagram of  $r$  and  $\phi$  sensor layouts [83].

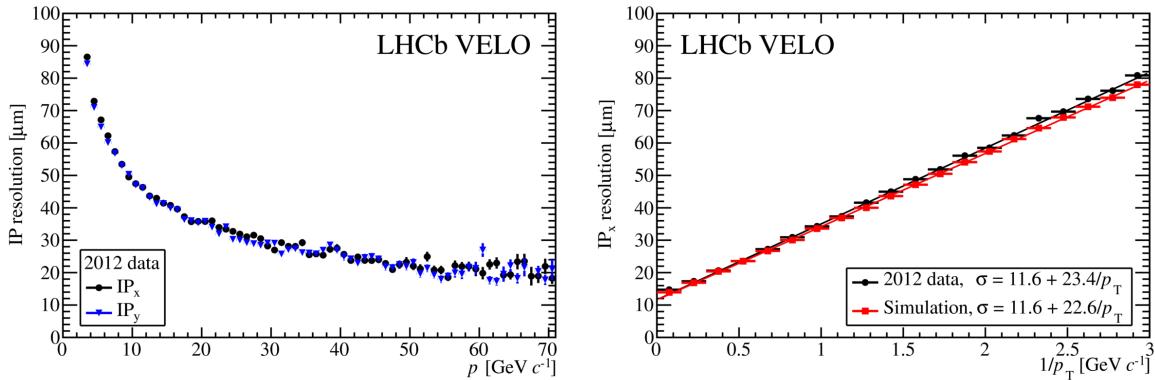
Therefore, to ensure good momentum resolution throughout the detector, the VELO is kept in a vacuum to reduce its material budget. Each half of the VELO is enclosed inside an aluminium box, which keeps it in a vacuum and shields the electronic readouts from radio frequencies generated by the beam. The overall material budget of the VELO comes to 17.5 % of a radiation length.

Excellent vertex resolution is required in the VELO. To achieve this the sensor needs 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 8mm from the beam axis. However during the injection phase the width of the beam is much larger, therefore the halves of the VELO can retract to be 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 3.4 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 2 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.



**Fig. 3.6** VELO performance for primary vertex resolution perpendicular (left) and parallel (right) to the beam axis as a function of the number of tracks in an event for 2012 data [86].

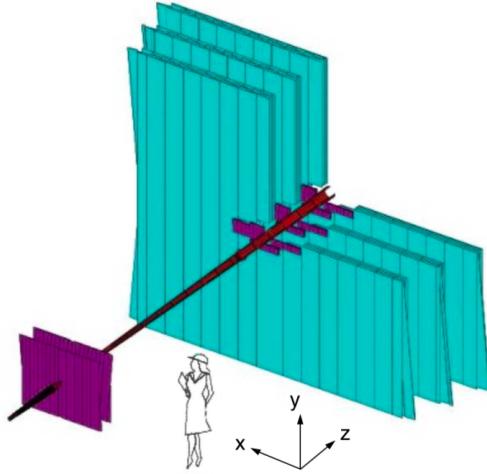


**Fig. 3.7** VELO performance for impact parameter resolution as a function of momentum (left) and inverse transverse momentum (right) for 2012 data [86].

The VELO achieves a vertex resolution 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 3.6. 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 3.7 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 down stream of the magnet. These tracking stations provide complementary information

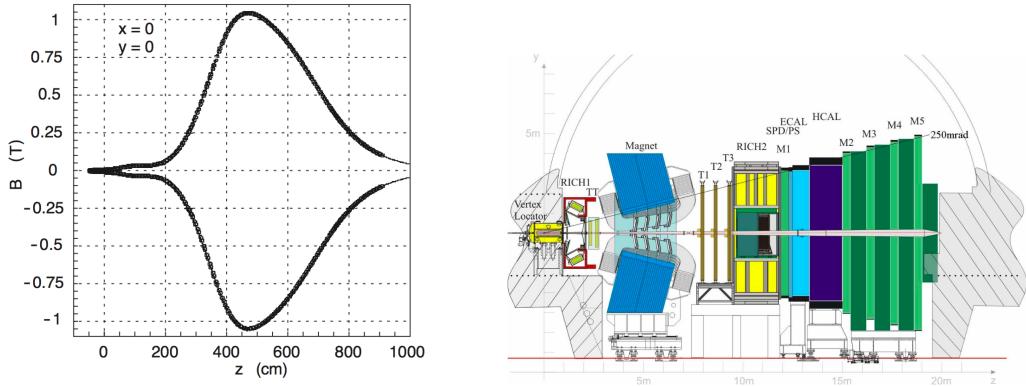


**Fig. 3.8** Illustration of the relative sizes of the tracking stations. The TT is in purple on the left hand side of the diagram, the IT is in purple surrounded in turquoise by the OT and the beam pipe is shown in red [83].

to the VELO, and the presence of the magnetic field allows the momentum of charged particles to be determined.

The TT is made up of 4 layers of silicon trackers spaced 27 cm apart 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  $50 \mu\text{m}$  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 but help with pattern recognition within the RICH detectors.

The T stations, T1-3, are split into two sections. Each station is composed of an Inner Tracker (IT) made of silicon and an Outer Tracker (OT) 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 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  $600 \text{ cm} \times 490 \text{ cm}$ , as shown in Figure 3.8. Therefore the T stations are not composed of only silicon due to its high cost. 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



**Fig. 3.9** Magnet field of the dipole magnet along the length of the LHCb detector (left) and the layout to the LHCb detector [83]. The peak strength of the field occurs between the TT and T1-3 station.

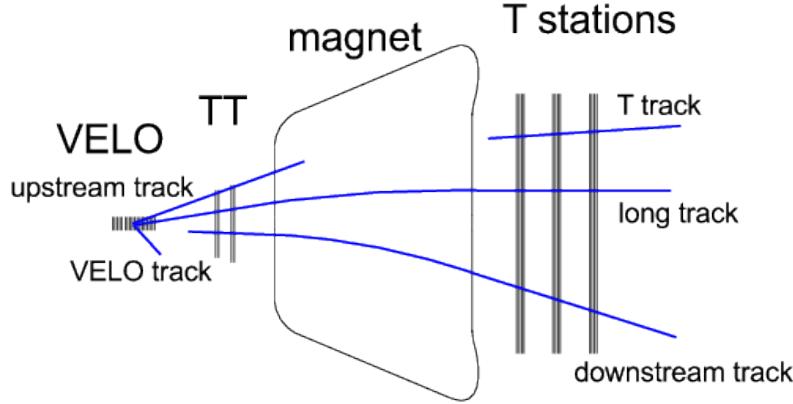
cross shape around the beam pipe, as shown in Figure 3.8, although the IT covers less than 2% of the T stations, 20% of tracks pass through it. This allows the occupancy of the OT to be less than 10% enabling a good overall track resolution from the OT despite it not being made of silicon. The OT of 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 is used to measure the momentum of charged particles travelling through the LHCb detector. The magnet was designed to have an integrated field strength is 4 Tm for track that travels 10 m through the detector.

The magnet is located between the TT and the T stations and its field covers the full LHCb acceptance. 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 3.9 shows a plot of the magnet strength alongside the detector layout. A small magnetic field is achieved in the RICH detectors by iron shielding. The magnet was designed to have an integrated field strength is 4 Tm for track that travels 10 m through the detector.

The polarity of the magnetic field is periodically switched so that approximately equal amounts of data is recorded for each magnet polarity during every year of data



**Fig. 3.10** Different types of tracks that are reconstructed at LHCb [89].

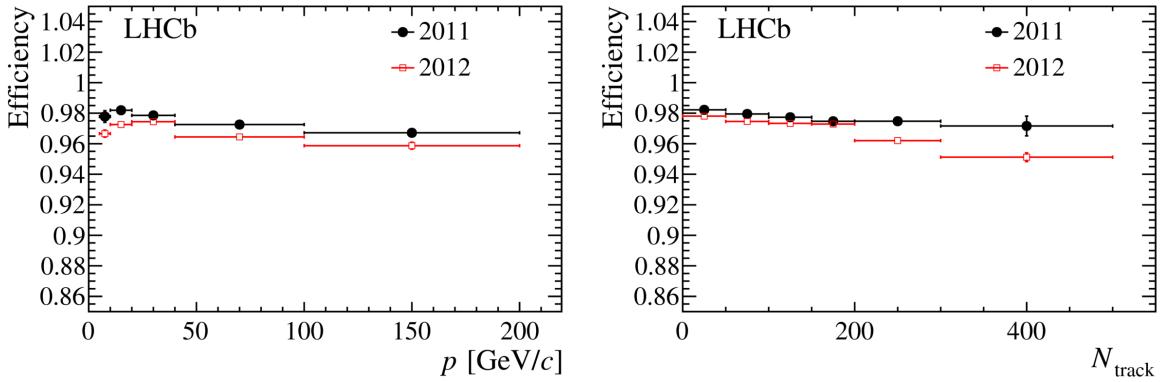
taking. This is done to measure left-right detection asymmetries and to help understand systematic uncertainties of  $\mathcal{CP}$  violation measurements.

#### 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 though the length of the LHCb detector and the particle momentum. The 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 the segments of the track have been found, the trajectory is fitted with a Kalman Filter [87, 88] with takes into account multiple scattering and energy loss within the detector. For each track the Filter returns the  $\chi^2$  per degree of freedom, a measure of quality for the track. In LHCb 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, 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 only leave hits in the VELO and TT stations upstream of the magnet. The absence of tracks further down the detector is because the magnetic field sweeps the particles out



**Fig. 3.11** Long track reconstruction efficiency as a function of momentum (left) and number of tracks in the event (right) for 2012 data [89].

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 only cross the T1-3 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.
- **Long tracks** are the most useful for physics analyses because they are formed by particles that travel through the VELO, TT and T1-3 stations. Information from all the tracking stations is combined so these tracks 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 3.11 for 2012 data. In Run 1 long tracks were correctly reconstructed on 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 number of total hits it used and the other is discarded. The second type of incorrect tracks 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 T1-3 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.

Once the tracks have been reconstructed, parameters that are necessary for identifying and measuring different particle decays in an event can be computed from the tracks. 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 ns.

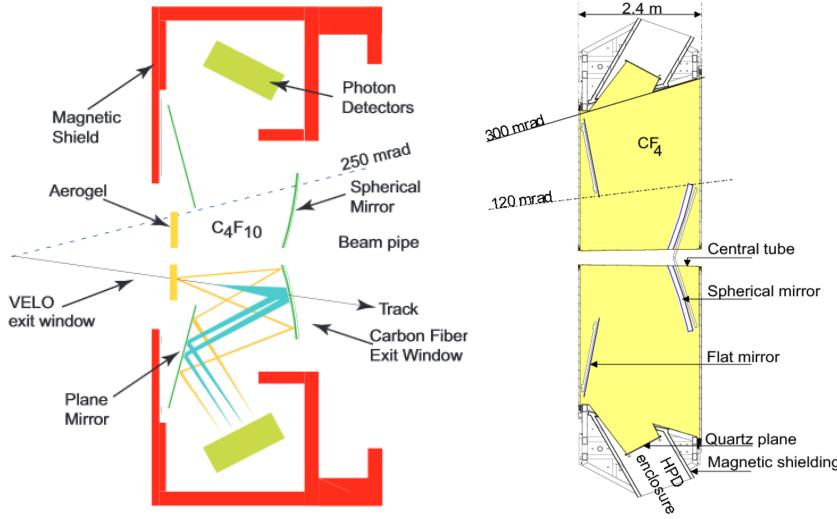
### 3.2.2 Particle identification

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

#### 3.2.2.1 Ring Imaging Cherenkov detectors

RICH detectors are used at LHCb to distinguish charged hadrons and leptons that have a momentum between 2 and 100  $\text{GeV}/c$ . 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 because the typical decay products of 2-body and multi-body  $b$ -hadron decays is around 50  $\text{GeV}$  [](TDR).

The RICH detectors are based on the following principle; when a charged particle travels with velocity  $v$  through a dielectric medium with a refractive index  $n$ , 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,  $\theta_c$  depends on the speed at which the particle was travelling 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, 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

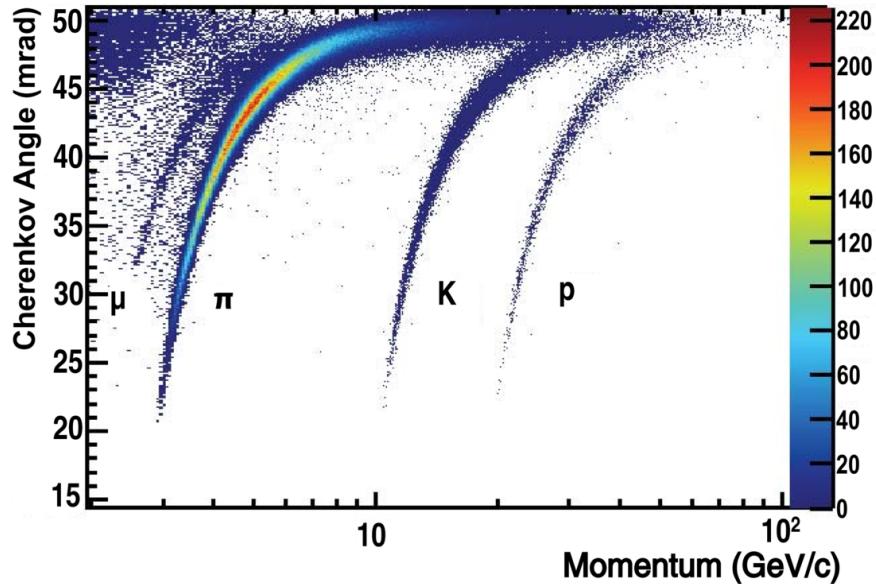


**Fig. 3.12** RICH-1 detector (left) and the RICH-2 detector (right) [83]. For Run 2 the aerogel radiator in the RICH-1 detector was removed.

particle identification complex. 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 [90].

The two RICH detectors cover complementary momentum regions. The RICH-1 detector is located between the VELO and the TT station, it covers the full LHCb angular acceptance and provides PID information on particles in the momentum range 2 to 40 GeV/c. The RICH-1 detector illustrated in Figure 3.12, contains two different radiator materials; at the front of the detector is an array of 16 aerogel tiles sensitive to particles with a momentum between 2 and 10 GeV/c, behind the aerogel is a C<sub>4</sub>F<sub>10</sub> 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. As charged particles travel through RICH-1, the rings of light produced are focused by spherical and planar mirrors onto Hybrid Photon Detectors (HPDs) [6]. The radii of the detected rings provides information about how fast the particle was travelling.

The RICH-2 detector is located upstream of RICH-1, between the last tracking station and the first muon station. RICH-2 consists of a CF<sub>4</sub> gas radiator sensitive to particles with a momentum range 15 - 100 GeV/c and the detection of the light produced is the



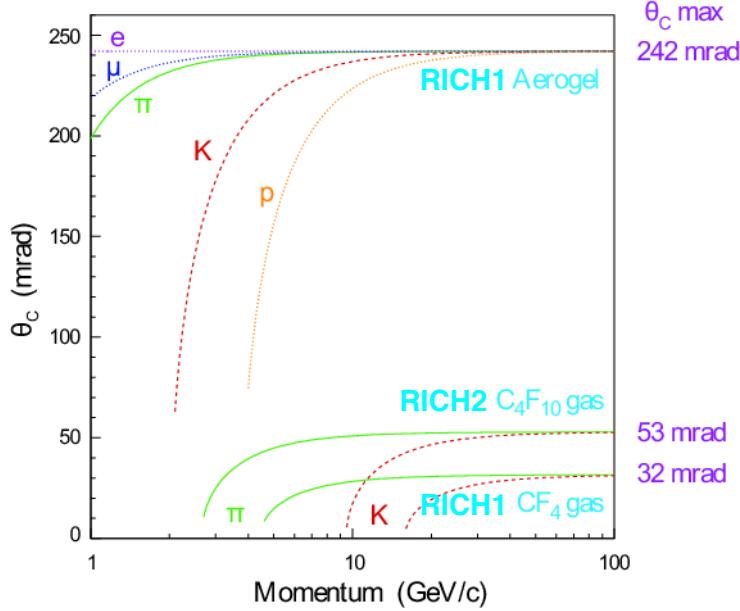
**Fig. 3.13** Cherenkov angles for isolated tracks as a function of momentum in the RICH-1 detector for 2011 data [91].

similar to RICH-1, as illustrated in Figure 3.12. Unlike RICH-1, the RICH-2 detector does not cover the full LHCb angular acceptance but only  $\pm 120$  mrad in the horizontal and  $\pm 100$  mrad in the vertical direction. This area contains the higher momentum particles the RICH-2 is sensitive to, the low momentum particles have been bent out of the acceptance by the magnetic field.

The rings of light collected by the RICH detectors when combined with information about particle momentum and tracks from the tracking stations enables the particle type to be identified. Figure 3.13 shows how the Cherenkov angle and momentum can be combined to identify different types of particles in the RICH-1 detector, there are distinct bands for each particle mass. Figure 3.14 shows what is expected for the different radiators.

### 3.2.2.2 Calorimeters

The calorimeter system consists of 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



**Fig. 3.14** Expected Cherenkov angles produced by different particles travelling through the radiators in the RICH detectors [83].

reconstruction and identification of these particles. The ECAL is the only part of the LHCb detector that measures the position and energy of photons and neutral pions.

The calorimeters in LHCb are sampling calorimeters that consist 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 photon multiplier tubes and turned into an electrical signal. In the ECAL showers are started by ionisation, bremsstrahlung radiation or pair production depending on the energy of the incident particle and whether it is a  $e^\pm$  or a photon. In the HCAL it is interaction via the strong force that leads to 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 shower in the PS. Information collected by the PS enables showers created by pions in the ECAL to

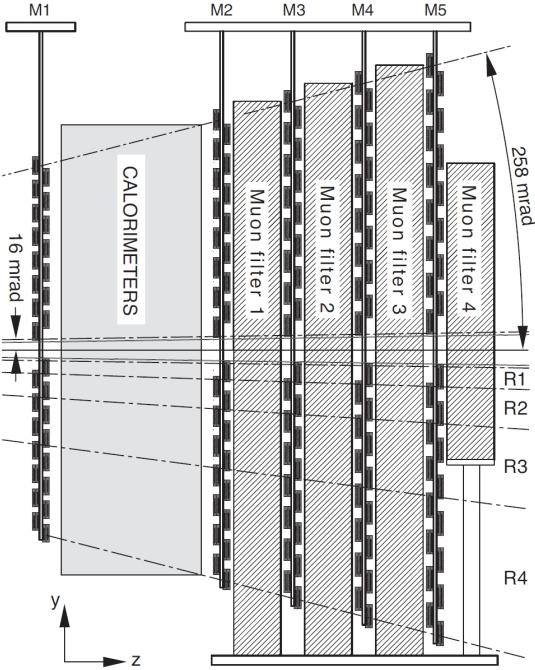
be separated from those created by electrons and positrons. The ECAL is designed to contain the entire shower of high energy photons so that it can provide good energy resolutions of photons passing through the detector. The ECAL has an energy resolution of  $\delta E/E = 9\%/\sqrt{E} \oplus 0.8\%$  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

The muon stations are designed to identify highly penetrating muons for use in the trigger and offline analyses. 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 physics analyses. Compared to other particles muons have a high penetrating power due to their relatively large mass and because muons do not interact via the strong force, these properties are exploited in the muon detectors.

There are 5 muon stations, M1-5, shown in Figure 3.15 that track and identify muons. The first muon station is located before the calorimeters, the inner section, where the fluence is greatest, is made of gas electron multiplier foils and the outer section is made from multi-wire proportional chambers (MWPCs). 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 and between each station is 80cm of lead absorber ensuring only high energy muons pass through the muon detector. 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. The first 3 stations have a high spatial resolution and provide track and transverse momentum information to be used in the trigger. M1 is located before the calorimeters to improve the transverse momentum measurement of the muons. The last two stations have a lower spatial resolution and are designed to identify muons with the greatest transverse momentum. 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. Tracking 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.



**Fig. 3.15** Layout of the muon stations [83].

### 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.

The muon stations are used, along with information from the tracking system, to produce a binary selection (`isMuon`) to identify muons. 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 the hits required in the muon stations depends on the momentum of the muon. Muons with momentum in the range  $3 < p < 6$  GeV/ $c$  must leave hits in M2-3, those in the momentum range  $6 < p < 10$  GeV/ $c$  leave hits in M2-3 and either M4 or M5, and finally muons with momentum above 10 GeV/ $c$  must be observed in all the muon stations. Figure 3.16 shows the efficiency for the `isMuon` selection at selecting muons and probabilities of mis-identifying hadrons as muons. The efficiencies and mis-identification probabilities are computed using the *tag and probe technique* [6]. This technique uses two tracks from a decay and 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 muon efficiency uses  $J/\psi \rightarrow \mu^+ \mu^-$  decays, proton mis-identification probabilities are computed using  $\Lambda^0 \rightarrow p\pi^-$  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 muons detectors. The main contribution to misidentifying hadrons as muons comes from the kaons and pions that decay in flight, 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 and the other is based on Neural Networks. In the first method, likelihood fits are performed in each sub-detector comparing each charged particle track 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 approach uses Neural Networks [93] 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 systems and extra detector information that are not considered in the likelihood method. The Neural Networks are trained on simulated inclusive  $b$  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.

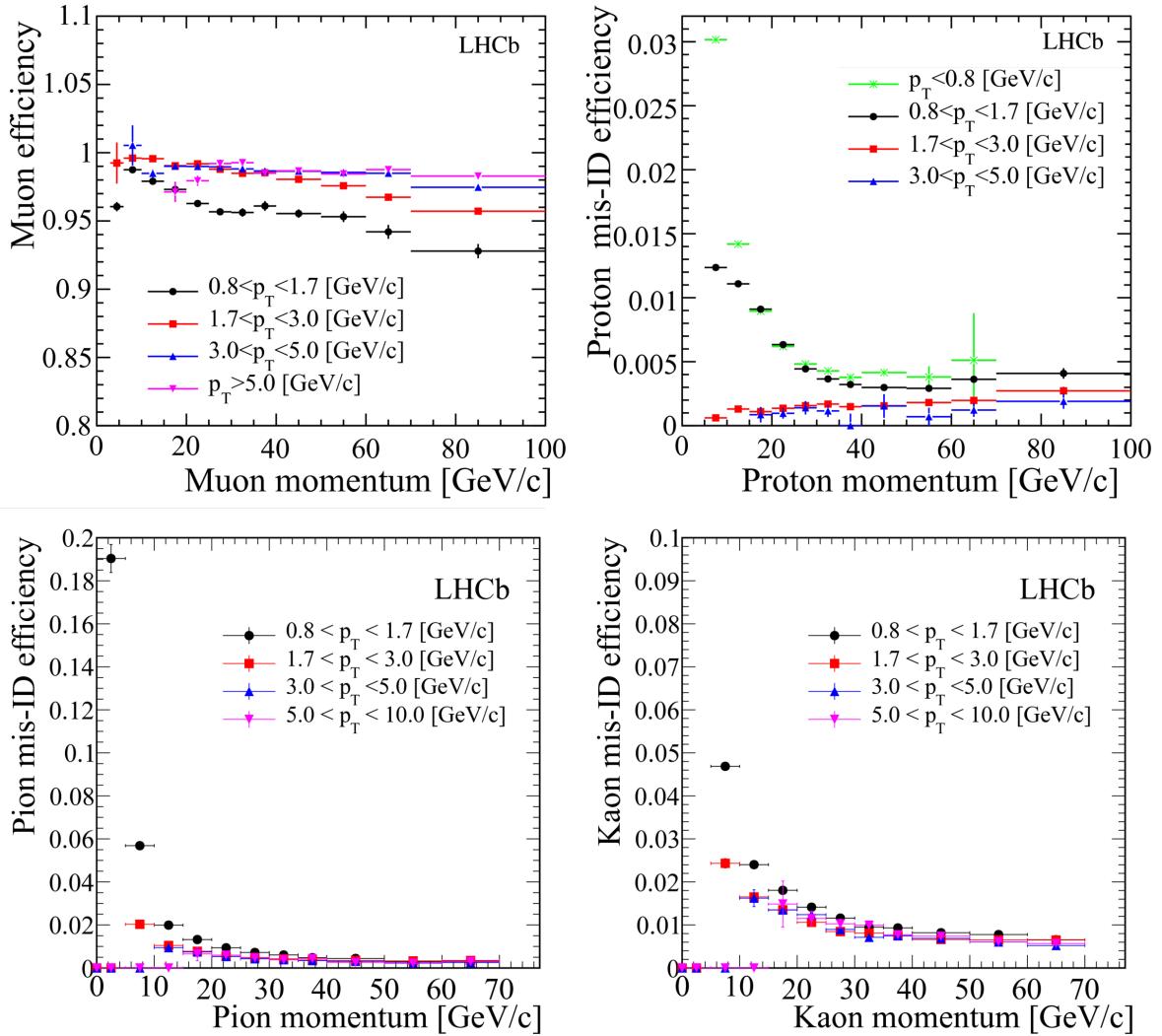
Figure 3.17 shows a comparison of the performance of the DLL and ProbNN variables in selecting protons and muons. Although the performance of the two types of variables are quite different, the efficiencies of each variable varies with different kinematic properties of the decay. The most appropriate PID variable type to use depends on the physics analysis it is being used in.

### 3.2.3 Trigger

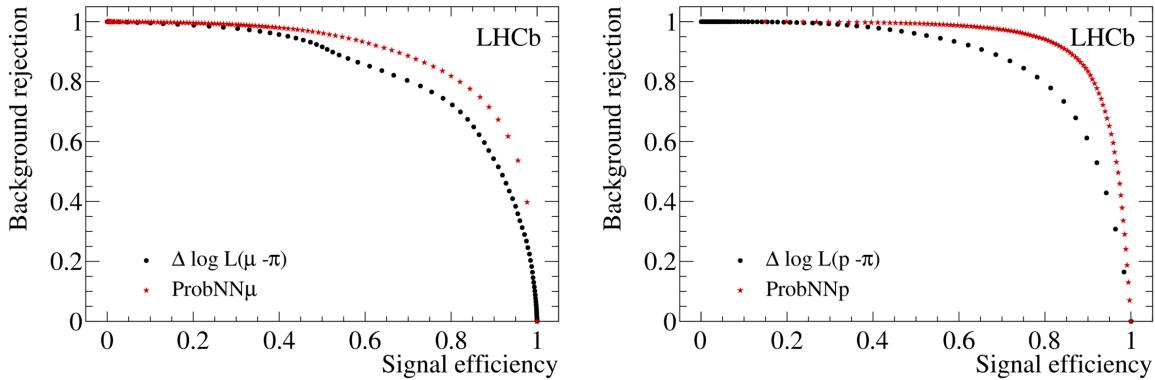
The LHC was designed to collide protons at a rate of 40 MHz, this rate is too high for information to be read out by the original design of the LHCb detector<sup>2</sup>. However, most  $pp$  collisions do not produce particles within the detector acceptance that are

---

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



**Fig. 3.16** Muon efficiency (top left) and mis-identification probabilities for protons (top right), pions (bottom left) and kaons (bottom right) for isMuon criteria [92].

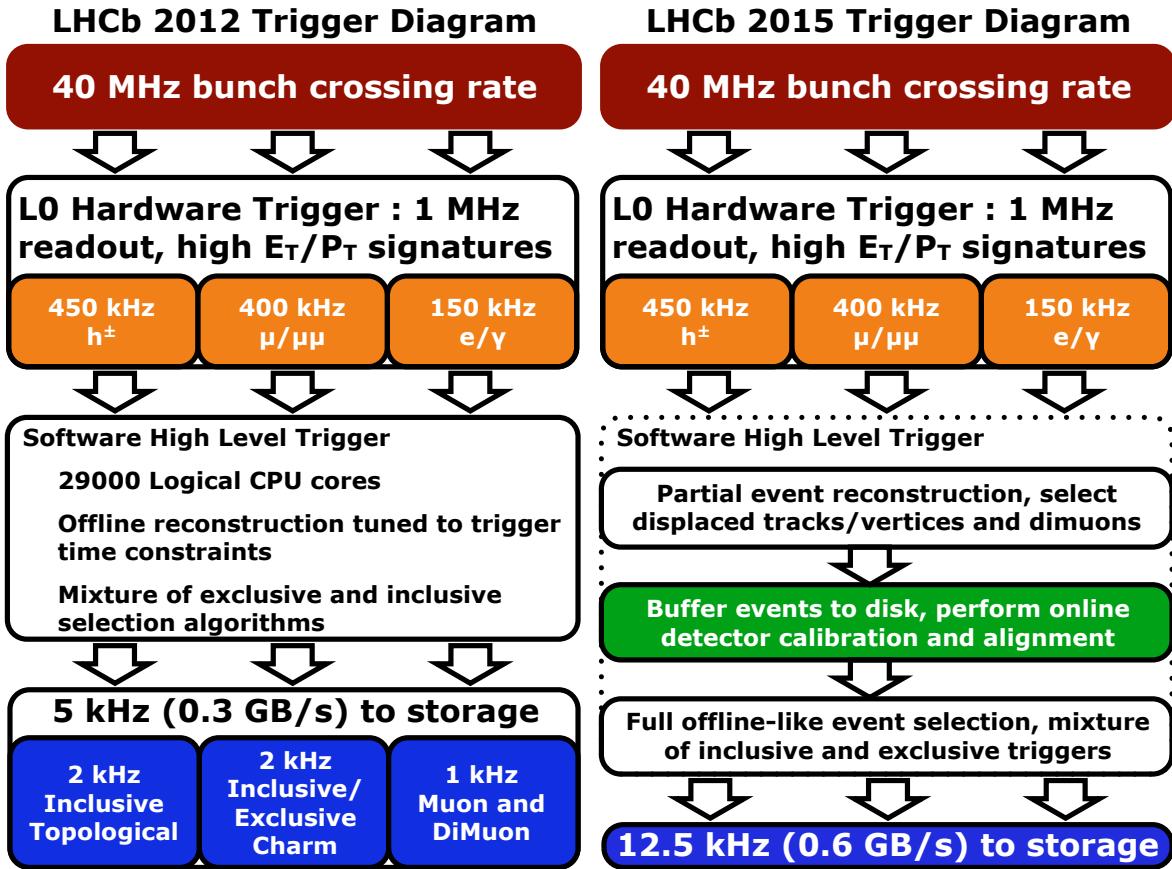


**Fig. 3.17** Muon (left) and proton (right) signal efficiency vs background rejection for DLL and ProbNN PID variables [83].

interesting for physics analyses at LHCb. A trigger system [83, 94, 95] is used to identify  $pp$  collisions that contain potentially interesting physics processes, the information from these events are saved for later use in physics analyses. 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 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 3.18 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 events rate to 1 MHz, where information from the full detector can be read out. Therefore the L0 is limited to use information from the detector that can be read at the same rate as the LHC collision rate. The L0 uses information from three parts of the



**Fig. 3.18** Diagrams of the trigger systems used in 2012 (left) and 2015 (right).

detector, the VELO, calorimeters and the muon stations, to make decisions about the relevance of each event.

The pileup veto stations in the VELO are used in L0 pileup trigger lines, these lines identify the number of collisions in an event and are predominately used for luminosity measurements [96].

The other L0 trigger lines are based on the kinematic properties of  $b$ -hadron decays. 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 calorimeters are used in trigger lines that select events containing high  $E_T$  electrons, photons or hadrons. 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 provided the event multiplicity is not too high. The  $E_T$  thresholds are different for each particle type. Events with high multiplicity take a long time to reconstruct and process in the HLT, therefore it is not efficient to keep these events. The multiplicity is measured by

the number of hits in the SPD detector (nSPD), only events with nSPD lower than a specified value can pass an L0 trigger line.

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 muons  $\sqrt{p_{T1} \times p_{T2}}$  above a threshold value, provided the event multiplicity is not too high.

The  $E_T$  and  $p_T$  thresholds and the multiplicity limit for the L0 trigger lines vary for each year of data taking depending on the bandwidth available for the trigger.

### 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 made by the L0 trigger lines 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 event information only from the VELO and tracking stations. The reduced output of the HLT1 then 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 used in physics analyses. 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 [97]. The majority of the changes to the HLT for Run 2 are not relevant for the analysis discussed 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 HTL2 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 used in physics analyses. The GAUDI framework [98] is a C++ framework that is the basis for the software applications needed to process the data at LHCb [99]. This framework ensures that 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 [100].

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 application. It 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 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 that are needed for physics analyses. 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 physics analysis. Each physics analysis 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 a stripping line selection are available to be used in physics 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 DaVinci application one last time to produce ROOT [101] files from the output of

their stripping lines, these files display the data in histograms and are used for physics analyses.

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 physics analyses. 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 [102, 103], which uses PYTHIA [104, 105] to model  $pp$  collisions and the production of particles, and then the EVTGEN [106] application to calculate the decay trees and kinematics of these particles. Final state radiation is modelled using PHOTOS [107]. Both PYTHIA and EVTGEN have been tuned for the production and decay of particles within the LHCb detector. The GEANT4 [108, 109] 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 are relevant to particular physics decays. The BOOLE application 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 DAVINCI in the same way as the real data to produce the ROOT files used in physics analyses.

The LHCb software framework is set up so that it can be used on the Worldwide LHC Computing Grid [110, 111], 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 DIRAC [112] system manages grid sites and the GANGA [113, 114] 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.

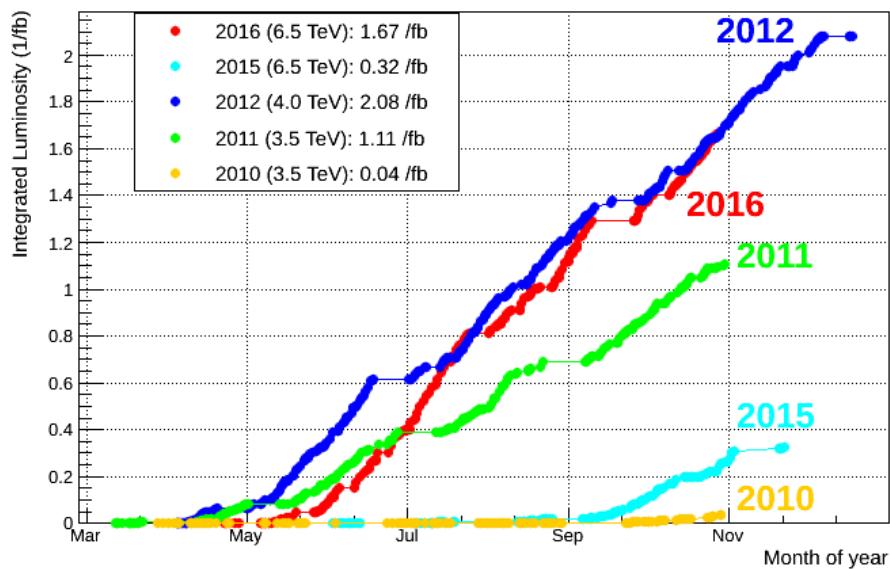
### 3.3 Summary

The data collected by the LHCb experiment during  $pp$  collisions is summarised in Figure 3.19. The physics analyses described in this dissertation use an integrated luminosity of  $4.4 \text{ fb}^{-1}$  which consists of data recorded during 2011, 2012 and 2015 and up until September of 2016. The integrated luminosities used for each year are given in Table 3.2. The integrated luminosity of Run 2 is currently less than the total from

| Run   | Year | $\sqrt{s}$ TeV | Integrated luminosity / $\text{fb}^{-1}$ |
|-------|------|----------------|--|
| 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.

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 the Run 2 data set will already contain more  $b$ -hadrons useful for physics analyses than the Run 1 data set.



**Fig. 3.19** Integrated luminosity collected by the LHCb experiment in each year of data taking.  
Source: LHCb.



# Chapter 4

## Event selection

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

The criteria used to identify  $B$ -meson decays for the branching fraction measurements are described in Section 4.3. As well as  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays this analysis also uses ,  $B_s^0 \rightarrow J/\psi\phi$  and  $B \rightarrow h^+h'^-$  decays, where  $h = K, \pi$ . Firstly the trigger requirements to select these decays are given in Section 4.3.1. Then in Section 4.3.2 cut-based criteria used to identify each  $B$ -meson decay are described alongside a study into the optimisation of signal efficiency for several cuts used in this stage of the selection process. The particle identification requirements used to separate  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays from backgrounds are given in Section 4.3.3. The last step in the selection relies on multivariate classifiers that are described in Section 4.3.4. A summary of the criteria used to select  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays for the branching fraction measurements is given in Section 4.3.5.

The identification and selection criteria of decays used for the branching fraction measurements are adapted for the effective lifetime measurement as described in Section 4.4. Similar to the branching fraction analysis, the effective lifetime measurement requires  $B_s^0 \rightarrow J/\psi\phi$ ,  $B^0 \rightarrow K^+\pi^-$  and  $B_s^0 \rightarrow K^+K^-$  decays as well as  $B_s^0 \rightarrow \mu^+\mu^-$  decays. The changes made to the trigger requirements, mass range of  $B_s^0 \rightarrow \mu^+\mu^-$  candidates and the particle identification requirements are presented in Sections 4.4.1, 4.4.2 and 4.4.3, respectively. Section 4.4.4 describes an investigation into multivariate classifiers for the effective lifetime measurement. Finally the complete set of selection criteria are summarised in Section 4.4.5.

## 4.1 Backgrounds

The reconstruction of the data, 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 expected.
- 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 not decay instantaneously.
- 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  $J/\psi \rightarrow \mu^+ \mu^-$ .
- $B \rightarrow h^+ h'^-$  decays, where  $h^{(\prime)} = K, \pi$ , when both hadrons are mis-identified as muons. This usually occurs when the hadrons decay whilst travelling through 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  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays from the backgrounds is challenging because  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays are much less abundant than the backgrounds therefore reconstructed candidates are predominately made from background decays.

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

**Table 4.1** Branching fractions for background decays. The 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 [48, 115, 116]. The theoretical prediction for  $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$  branching fraction combines information from references [117, 118], the  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  branching fraction is estimated from references [119, 120] and the  $B^+ \rightarrow \pi^+ \mu^+ \mu^-$  branching fractions is evaluated from [116, 121]

## 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. 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 conditions, such as the trigger lines or centre-of-mass energy. Similar versions are chosen for decay samples

| Decay  | Generator level cuts   | Data taking conditions       | Simulation version                   | Events ( $\times 10^6$ ) |
|--|--|------------------------------|--------------------------------------|--------------------------|
| <i>Stripping selection studies selection</i> |  |                              |                                      |                          |
| $B_s^0 \rightarrow \mu^+ \mu^-$              |  | 2012                         | sim06b                               | 2                        |
| $B^0 \rightarrow \mu^+ \mu^-$                |  | 2012                         | sim06b                               | 2                        |
| $B^0 \rightarrow K^+ \pi^-$                  |  | 2012                         | sim06b                               | 1                        |
| $B^+ \rightarrow J/\psi K^+$                 |  | 2012                         | sim06b                               |                          |
| <i>Multivariate classifier training</i>      |  |                              |                                      |                          |
| $b\bar{b} \rightarrow \mu^+ \mu^- X$         | $p > 3 \text{ GeV}/c$<br>$4.7 < M_{\mu^+ \mu^-} < 6.0 \text{ GeV}/c^2$<br>DOCA < 0.4mm<br>$1 < \text{PtProd} < 16 \text{ GeV}/c$ | 2012                         | sim06b                               | 8                        |
| $b\bar{b} \rightarrow \mu^+ \mu^- X$         | $p > 3 \text{ GeV}/c$<br>$4.7 < M_{\mu^+ \mu^-} < 6.0 \text{ GeV}/c^2$<br>DOCA < 0.4mm<br>$\text{PtProd} > 16 \text{ GeV}/c$     | 2012                         | sim06b                               | 7                        |
| $B_s^0 \rightarrow \mu^+ \mu^-$              |  | 2012                         | sim06b                               | 2                        |
| <i>Analysis method development</i>           |  |                              |                                      |                          |
| $B_s^0 \rightarrow \mu^+ \mu^-$              |  | 2011<br>2012<br>2015<br>2016 | sim08a<br>sim08i<br>sim09a<br>sim09a | 0.6<br>2<br>2<br>2       |
| $B^0 \rightarrow K^+ \pi^-$                  |  | 2011<br>2012<br>2015<br>2016 | sim08b<br>sim08g<br>sim09a<br>sim09a | 8<br>9<br>4<br>8         |
| $B_s^0 \rightarrow K^+ K^-$                  |  | 2012<br>2015                 | sim08g<br>sim09a                     | 7<br>4                   |

**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 two muons ( $M_{\mu^+ \mu^-}$ ), the distance of closest approach of the muons (DOCA) and the product of the transverse momenta of the muons (PtProd).

used within each study listed in Table 4.2, so that differences are not masked by variations in the simulation of the decays.

Simulated  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays are used to understand 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 at the generator level to reduce the size of the samples and to speed up the simulation process. The cuts, listed in Table 4.2, are applied on 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 to 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 [122] 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 branching fraction measurements

The selection of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays occurs in several steps. The first step is choosing the trigger requirements (Sect. 4.3.1), which is followed by a cut-based selection to remove some background events (Sect. 4.3.2). Particle identification variables (Sect. 4.3.3) are then used to reduce backgrounds from mis-identified semi-leptonic and  $B \rightarrow h^+h'^-$  decays. Finally multivariate classifiers (Sect. 4.3.4) are used to reduce the backgrounds to a low enough level so that the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  branching fractions can be measured.

The branching fraction measurements are described in Chapter 5 and 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. The selection criteria for these decays are kept as similar as possible to the selection of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays and will be described alongside the signal selection. Furthermore  $B_s^0 \rightarrow J/\psi\phi$  decays are used to verify steps of the measurement process.

### 4.3.1 Trigger requirements

The trigger is the first step in the selection, which selects events that could contain an interesting particle decays. Candidates from different particle decays are reconstructed from events that have passed 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 part of the event. There are several different decisions that identify this:

- TOS, triggered on signal - a candidate is identified as TOS if only information from the candidate was enough to cause a trigger line to select 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 select the event; and
- DEC - a candidate is identified as DEC if anything in the event caused a trigger line to select an 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 select the event.

$B_s^0 \rightarrow \mu^+ \mu^-$  decays are very rare decays and therefore trigger requirements are chosen to keep a high efficiency at this step of the selection.

Candidates are required to be identified as DEC for each level of the trigger. The trigger lines L0Global, Hlt1Phys and Hlt2Phys are used in the analysis. These trigger lines combine the decisions of many individual lines, which allows a high efficiency to be achieved for selecting  $B_s^0 \rightarrow \mu^+ \mu^-$  decays. The L0Global trigger 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 on trigger lines related to physics processes and HLT trigger lines used for calibration are excluded.

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 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 there to high a high enough number 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.

| 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              |
| <hr/>  |                  |
| $B \rightarrow h^+ h'^-$   |                  |
| 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.

### 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 development of the stripping selection is described in Sections 4.3.2.1 and 4.3.2.2. The primary purpose of the stripping selection is to reduce to size of the data set produced from  $pp$  collisions to a manageable size from which properties of particle decays can be measured.

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.

#### 4.3.2.1 Development of the stripping selection

The stripping selections applied to all decays needed for the branching fraction measurements were designed at the start of Run 1 by studying the efficiencies of different selection cuts from simulated events [123]. 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.

There are four separate stripping lines that 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. 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. Any changes made to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping selection to improve the selection efficiency must be included in the selection of the other decays to keep the systematic uncertainties under control, this is particularly important for  $B \rightarrow h^+ h'$  and  $B^+ \rightarrow J/\psi K^+$  decays. The stripping selection cuts applied for the Run 1 branching fraction analysis [40, 39] 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  $J/\psi$ ) 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 that have a mass less 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 vertices<sup>1</sup> of the particle. For correctly reconstructed candidates the direction cosine should be very close to one, requiring candidates to have positive value ensuring events are travelling in the wrong direction are removed;

<sup>1</sup>The 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 - M_{PDG}  < 1200 \text{ MeV}/c^2$<br>DIRA > 0<br>$\chi_{\text{FD}}^2 > 225$<br>$\chi_{\text{IP}}^2 < 25$<br>$\chi_{\text{VTX}}^2 < 9$<br>DOCA < 0.3 mm | $ M - M_{PDG}  < 500 \text{ MeV}/c^2$<br>DIRA > 0<br>$\chi_{\text{FD}}^2 > 225$<br>$\chi_{\text{IP}}^2 < 25$<br>$\chi_{\text{VTX}}^2 < 9$<br>DOCA < 0.3 mm<br>$\tau < 13.248 \text{ ps}$<br>$p_T > 500 \text{ MeV}/c$ |
| $\mu$ or $h$ | $\chi_{\text{trk}}^2 < 3$<br>isMuon = True<br>Minimum $\chi_{\text{IP}}^2 > 25$<br>$p_T > 0.25 \text{ GeV}/c$   | $\chi_{\text{trk}}^2 < 3$<br>Minimum $\chi_{\text{IP}}^2 > 25$<br>$0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$<br>Ghost probability < 0.3  |

**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 [40, 39] to select  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h'^-$  decays.  $M_{PDG}$  corresponds to the Particle Data Group [48] mass of each particle.

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

**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 [40, 39] to select  $B^+ \rightarrow J/\psi K^+$  and  $B_s^0 \rightarrow J/\psi \phi$  decays.  $M_{PDG}$  corresponds to the Particle Data Group [48] mass of each particle.

- the flight distance ( $\chi^2_{\text{FD}}$ ) - this is computed by performing the fit for the production vertex of a particle with and without including the tracks from its decay products that originate from the decay vertex in the fit as well. For a  $B$  meson the  $\chi^2_{\text{FD}}$  is likely to be large because  $B$  mesons have long lifetimes therefore the tracks of its decays products will not point towards the production vertex;
- track fit  $\chi^2/\text{ndof}$  ( $\chi^2_{\text{trk}}$ ) - provides a measure of the quality of a fitted track, placing an upper limit removes poor quality tracks and backgrounds composed of poorly reconstructed decays;
- vertex fit  $\chi^2/\text{ndof}$  ( $\chi^2_{\text{VTX}}$ ) - provides a measure of how well tracks can be combined to form a vertex, placing an upper limit removes poorly constrained vertices and backgrounds composed of poorly reconstructed decays;
- distance of closest approach (DOCA) - this is the distance of closest approach of two daughter particles that make up the parent particle, 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 ( $\tau$ ) - 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  $B$  meson masses however an upper limit removes unphysical backgrounds;
- momentum ( $p$ ) - an upper limit 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 on random hits in the detector, tracks from the passage of real particles will have a low ghost probability;
- impact parameter  $\chi^2$  ( $\chi^2_{\text{IP}}$ ) - this is the change in the fit  $\chi^2$  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  $\chi^2_{\text{IP}}$  value whereas the muons will be displaced from the PV because of the relatively long lifetime of the  $B_{(s)}^0$  and therefore will have a large  $\chi^2_{\text{IP}}$ ;

- minimum impact parameter ( $\chi^2_{\text{IP}}$ ) - this is the  $\chi^2_{\text{IP}}$  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 Optimisation of $B_s^0 \rightarrow \mu^+\mu^-$ stripping selection

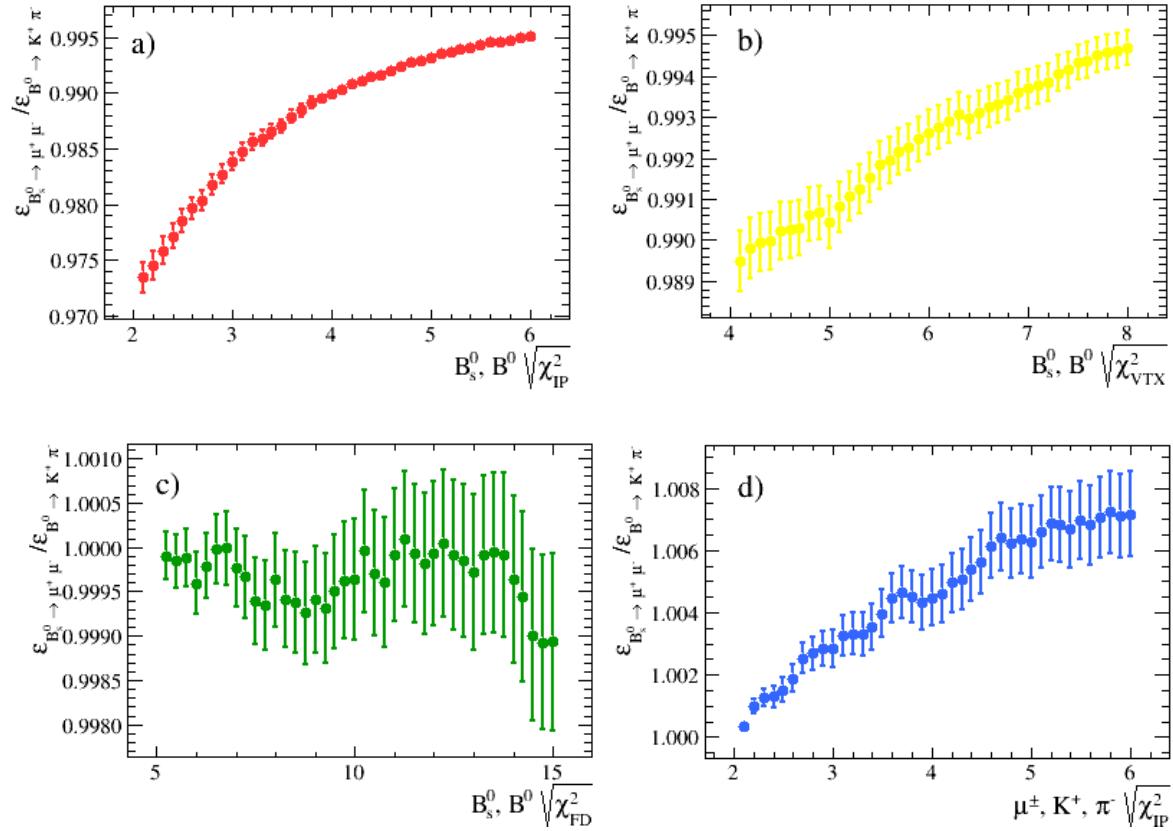
The  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  stripping line is designed to have a high efficiency for selecting  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays whilst removing backgrounds. The  $B \rightarrow h^+h^-$  and  $B^+ \rightarrow J/\psi K^+$  stripping lines are designed to have similar efficiencies for selecting  $B \rightarrow h^+h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays similar to the signal efficiency of the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  stripping line in order to reduce systematic uncertainties in the normalisation procedure. The signal 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 have been evaluated using simulated  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ ,  $B \rightarrow h^+h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays, respectively, listed in Table 4.2. The daughter particle of the simulated decays have  $p_t > 0.25$  GeV/c,  $\chi^2_{\text{trk}} < 3$  and isMuon true for muons in  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  and  $B^+ \rightarrow J/\psi K^+$  decays. The signal efficiencies for individual cuts are evaluated as well as the total signal efficiency of the stripping line to understand whether the lines can be made more efficient and the results are shown in Table 4.6. 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 are saved, not just those that have passed a trigger line.

The selection efficiencies are very similar for each stripping cut 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 4.2 and 4.1 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

| Requirement   | Efficiency                      |                               |                          |                              |
|---|---------------------------------|-------------------------------|--------------------------|------------------------------|
|   | $B_s^0 \rightarrow \mu^+ \mu^-$ | $B^0 \rightarrow \mu^+ \mu^-$ | $B \rightarrow h^+ h'^-$ | $B^+ \rightarrow J/\psi K^+$ |
| $ B  M - M_{PDG} $                                  | (100.00 $\pm$ 0.00)%            | (100.00 $\pm$ 0.00)%          | (98.25 $\pm$ 0.02)%      | (99.73 $\pm$ 0.02)%          |
| $B_{(s)}^0$ or $J/\psi$ 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 $J/\psi$ $\chi_{\text{FD}}^2$        | (83.74 $\pm$ 0.06)%             | (83.96 $\pm$ 0.06)%           | (83.83 $\pm$ 0.06)%      | (82.90 $\pm$ 0.15)%          |
| $B_{(s)}^0$ or $J/\psi$ $\chi_{\text{IP}}^2$        | (96.78 $\pm$ 0.03)%             | (96.93 $\pm$ 0.03)%           | (97.44 $\pm$ 0.03)%      | (97.52 $\pm$ 0.06)%          |
| $B_{(s)}^0$ or $J/\psi$ vertex $\chi^2/\text{ndof}$ | (97.21 $\pm$ 0.03)%             | (97.18 $\pm$ 0.03)%           | (97.68 $\pm$ 0.02)%      | (96.78 $\pm$ 0.07)%          |
| $B_{(s)}^0$ or $J/\psi$ DOCA                        | (99.82 $\pm$ 0.01)%             | (99.80 $\pm$ 0.01)%           | (99.83 $\pm$ 0.01)%      | (99.58 $\pm$ 0.03)%          |
| $\mu, h, K^+$ minimum $\chi_{\text{IP}}^2$          | (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)%          |

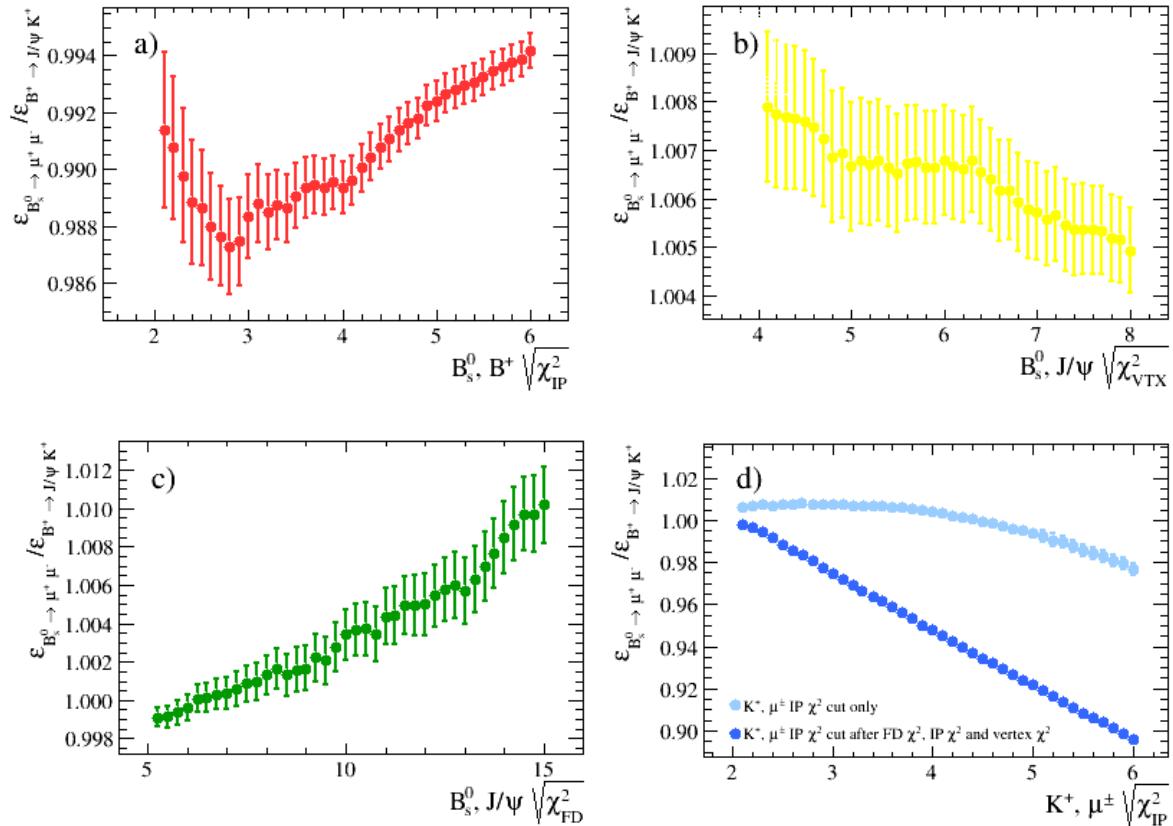
**Table 4.6** Signal efficiencies of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow h^+ h'^-$  and  $B^+ \rightarrow J/\psi K^+$  stripping lines using simulated  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B^0 \rightarrow K^+ \pi^-$  and  $B^+ \rightarrow J/\psi K^+$  decays respectively. The stripping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

exception of the  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^+ \rightarrow J/\psi K^+$   $\chi_{\text{IP}}^2$  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  $\chi_{\text{IP}}^2$ , Figure ??, markedly deviates from unity, showing that the  $\chi_{\text{IP}}^2$  distribution of the muons and kaon are very different as seen previously in reference [123]. If the  $\chi_{\text{FD}}^2$ ,  $B_s^0$  or  $J/\psi$   $\chi_{\text{IP}}^2$  and  $\chi_{\text{VTX}}^2$  selection cuts are applied to the simulated events before the daughter  $\chi_{\text{IP}}^2$  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.

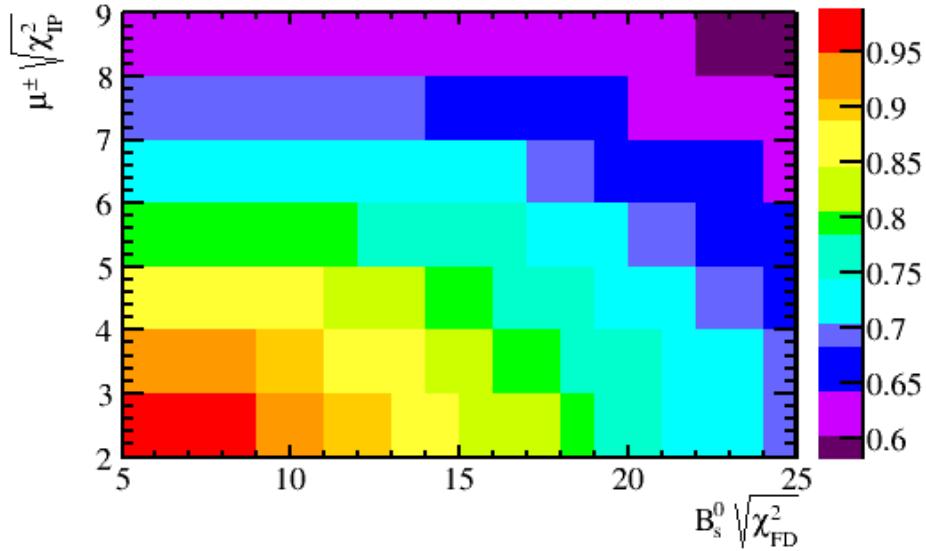


**Fig. 4.1** The ratio of  $B_s^0 \rightarrow \mu^+ \mu^-$  to  $B^0 \rightarrow K^+ \pi^-$  stripping efficiencies when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines. The square root of each  $\chi^2$  is used to condense the  $x$ -axis of the plots.

The efficiencies for most of the stripping cuts are  $\sim 97\%$  or higher. However, the efficiencies of the cuts on the  $\chi_{\text{FD}}^2$  of the  $B_{(s)}^0$  or  $J/\psi$  and the daughter  $\chi_{\text{IP}}^2$  of the muon or hadron pair are lower at 83% and 80%, respectively. Therefore improvements to



**Fig. 4.2** The 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 current cut values are marked by the blue lines. The square root of each  $\chi^2$  is used to condense the  $x$ -axis of the plots.

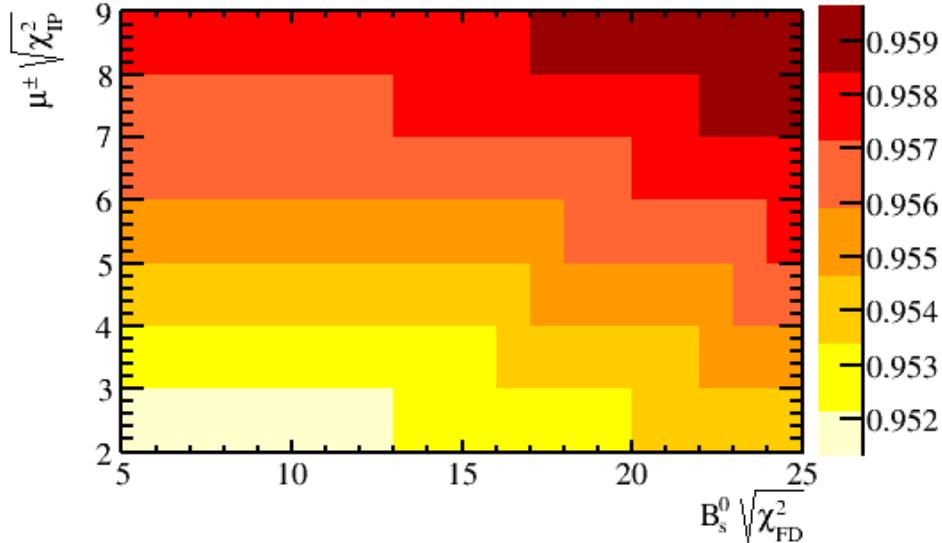


**Fig. 4.3** Efficiency of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping selection for  $B_s^0 \rightarrow \mu^+ \mu^-$  simulated decays for a range of cuts on the  $B_s^0 \chi_{FD}^2$  and the minimum muon  $\chi_{IP}^2$ .

the stripping selection efficiencies could be achieved by altering these two selection requirements.

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  $\chi_{FD}^2$  and daughter  $\chi_{IP}^2$  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  $\chi_{FD}^2$  and daughter  $\chi_{IP}^2$  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.

One of the main purposes of the stripping selection, as described in Section 3.2.4, 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. Also, 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.



**Fig. 4.4** The trigger efficiencies of  $B_s^0 \rightarrow \mu^+ \mu^-$  simulated decays across a range of  $B_s^0 \chi^2_{FD}$  and the minimum muon  $\chi^2_{IP}$  cut values for the trigger requirements used to select  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays for the branching fraction measurement.

Table 4.7 shows the total 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  $\chi^2_{FD}$  and daughter  $\chi^2_{IP}$  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 for a set of  $\chi^2_{IP}$  and  $\chi^2_{FD}$  cuts. 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  $\chi^2_{FD}$  and daughter  $\chi^2_{IP}$  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 for the last published analysis [40]. 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.

| $\sqrt{\chi_{FD}^2}$ | Stripping cut<br>Daughter $\sqrt{\chi_{IP}^2}$ | Stripping line efficiency       |                                     |                                   | Stripping line retention     |                                 |     |
|----------------------|--|---------------------------------|-------------------------------------|-----------------------------------|------------------------------|---------------------------------|-----|
|                      |  | $B_s^0 \rightarrow \mu^+ \mu^-$ | $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ | $B_{(s)}^0 \rightarrow K^+ \pi^-$ | $B^+ \rightarrow J/\psi K^+$ | $B_s^0 \rightarrow J/\psi \phi$ |     |
| 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** The 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  $\chi_{FD}^2$  and daughter  $\chi_{IP}^2$  cut values. The amount of data passing each selection has been normalised to the original set of stripping select cuts. The fractional uncertainty on the retention is less than 1 %.

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  $B_s^0 \chi_{\text{FD}}^2 > 121$  and minimum muon  $\chi_{\text{IP}}^2 > 9$  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 is smaller and the efficiencies are similar to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  selection efficiencies. Therefore the cuts on  $B_s^0 \chi_{\text{FD}}^2 > 121$  and minimum muon  $\chi_{\text{IP}}^2 > 9$  offer a good compromise between signal efficiency and the amount of data retained. The stripping lines have been updated to include these new looser cut values which will be used in future studies of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. The new looser cut values are not used in the analyses presented in this dissertation because the multivariate classifier use to separate signal and background decays, described in Section 4.3.4.3, is trained on simulated  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays and only decays that pass the original  $\chi_{\text{FD}}^2$  and daughter  $\chi_{\text{IP}}^2$  requirements are available in the simulated sample. In order to gain the best performance of the multivariate classifier on data the same cuts are applied to data that are applied to the samples used to train the classifier. Therefore the original cuts on  $\chi_{\text{FD}}^2$  and daughter  $\chi_{\text{IP}}^2$  listed in Table 4.4 must be used to select  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates.

| Stripping Lines                     | 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** The number of events passing stripping lines used for the  $B_s^0 \rightarrow \mu^+ \mu^-$  analysis in reference [39] from the selection 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  $\chi_{\text{FD}}^2 > 225$  and daughter  $\chi_{\text{IP}}^2 > 25$  are used.

### 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  $J/\psi$  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  $J/\psi \rightarrow \mu^+ \mu^-$ , contribute to the background of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays when a muon from the  $J/\psi$  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 ‘ $J/\psi$  veto’ can be used to remove background events from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays. The veto works by removing events where one muon from the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidate combined with any other oppositely charged muon in the event has  $|m_{\mu^+ \mu^-} - m_{J/\psi}| < 30$  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  $\chi^2_{\text{trk}}$ . 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 summaries 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  $\text{PID}_\mu$  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  $\text{PID}_\mu$  for the Run 1 and 2015 tuning was optimised 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  $B^0 \rightarrow \mu^+ \mu^-$  because the background decays 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 backgrounds. To improve the separation of signal and background decays two 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. These 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. The algorithm can then be 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 indicating 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 (Sect. 4.3.4.2), called the BDTS, is used to remove candidates that are very unlikely to be signal by placing a cut on the BDTS output. The second classifier (Sect. 4.3.4.3), called the global BDT, is used to classify candidates into bins containing increasing proportions of signal candidates. 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 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 number. 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 is 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).

One decision tree on its own is often 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 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 [124] 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. 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  $\beta$  where  $w \rightarrow w^\beta$ . The parameter  $\beta$  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 MinNodeSize<sup>2</sup>); the ‘speed’ at which the boosting occurs ( $\beta$ ) 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 can be 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:

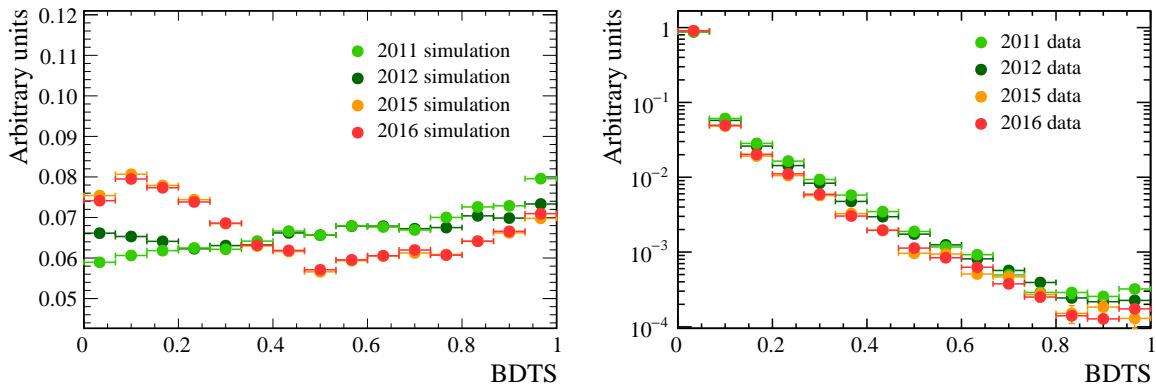
- impact parameter  $\chi^2$  of the  $B_{(s)}^0$ ;
- $\chi_{\text{VTX}}^2$  of the  $B_{(s)}^0$ ;
- direction cosine of  $B_{(s)}^0$ ;

---

<sup>2</sup>nEventsMin is the minimum number of decays in a lead where as 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.

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

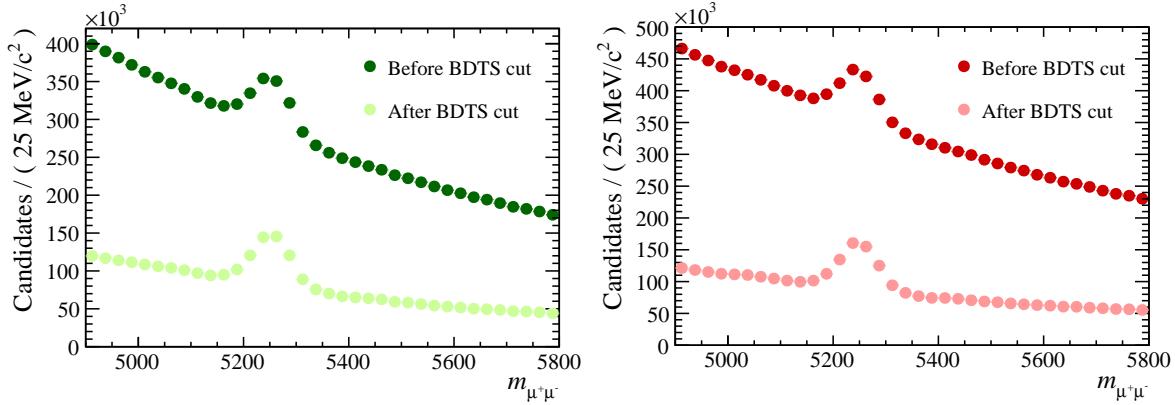
The signal and background samples used to train the BDTS are simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and background candidates in a sample of Run 1 data from the mass ranges 4800 - 5000 MeV/ $c^2$  and 5500 - 6000 MeV/ $c^2$ . The selection cuts listed in Table 4.9 are applied to the training 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  $J/\psi$  and the  $\chi^2_{\text{VTX}}$  is of the  $J/\psi$ . The performance of the BDTS at removing backgrounds is illustrated in Figure 4.6.



**Fig. 4.5** BDTS response for simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  decays (left) and data with a mass above 5447 MeV/ $c^2$  consisting of background decays.

#### 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 dependant on isolation variables. Isolation criteria provide a measure of how far away each muon from a  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidate is



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

| Selection applied to BDTS training samples. |   |
|---|---|
| $B_s^0$                                     | $\mu^\pm$                                     |
| $\chi_{\text{FD}}^2 > 225$                  | $p_T > 500 \text{ MeV}/c$                     |
| $\chi_{\text{IP}}^2 < 25$                   | $\chi_{\text{trk}}^2 < 3$                     |
| Vertex $\chi^2/n\text{dof} < 9$             | minimum $\chi_{\text{IP}}^2 > 25$             |
| 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.

| Parameter  | Value |
|------------|-------|
| nTrees     | 250   |
| nEventsMin | 400   |
| MaxDepth   | 3     |
| $\beta$    | 1.0   |
| nCuts      | 20    |

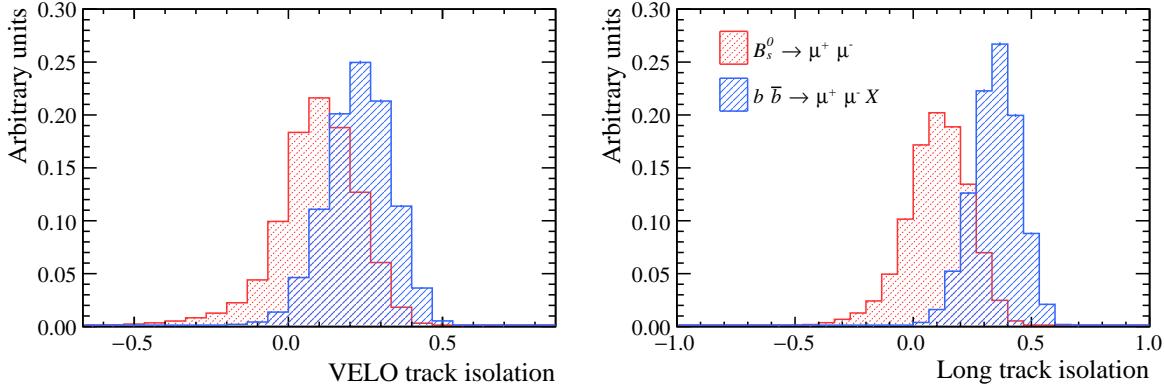
**Table 4.10** Training parameters used to specify the training of the BDTS.

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 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 can be found 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 tracks in the event. The BDT for the long track isolation criteria compares the  $\mu^+$  from a  $B_s^0 \rightarrow \mu^+\mu^-$  candidate with all other long tracks in the event, excluding the track of the  $\mu^-$ , and gives an output for each possible  $\mu^+$  and track pairing. The process is repeated for the  $\mu^-$ . 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 isolation criteria of a  $B_s^0 \rightarrow \mu^+\mu^-$  candidate is then composed of the sum of the highest BDT output values for the  $\mu^+$  and the  $\mu^-$ . 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 isolations are shown in Figure 4.7. Full details of the isolation variables can be found in reference [125].

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;

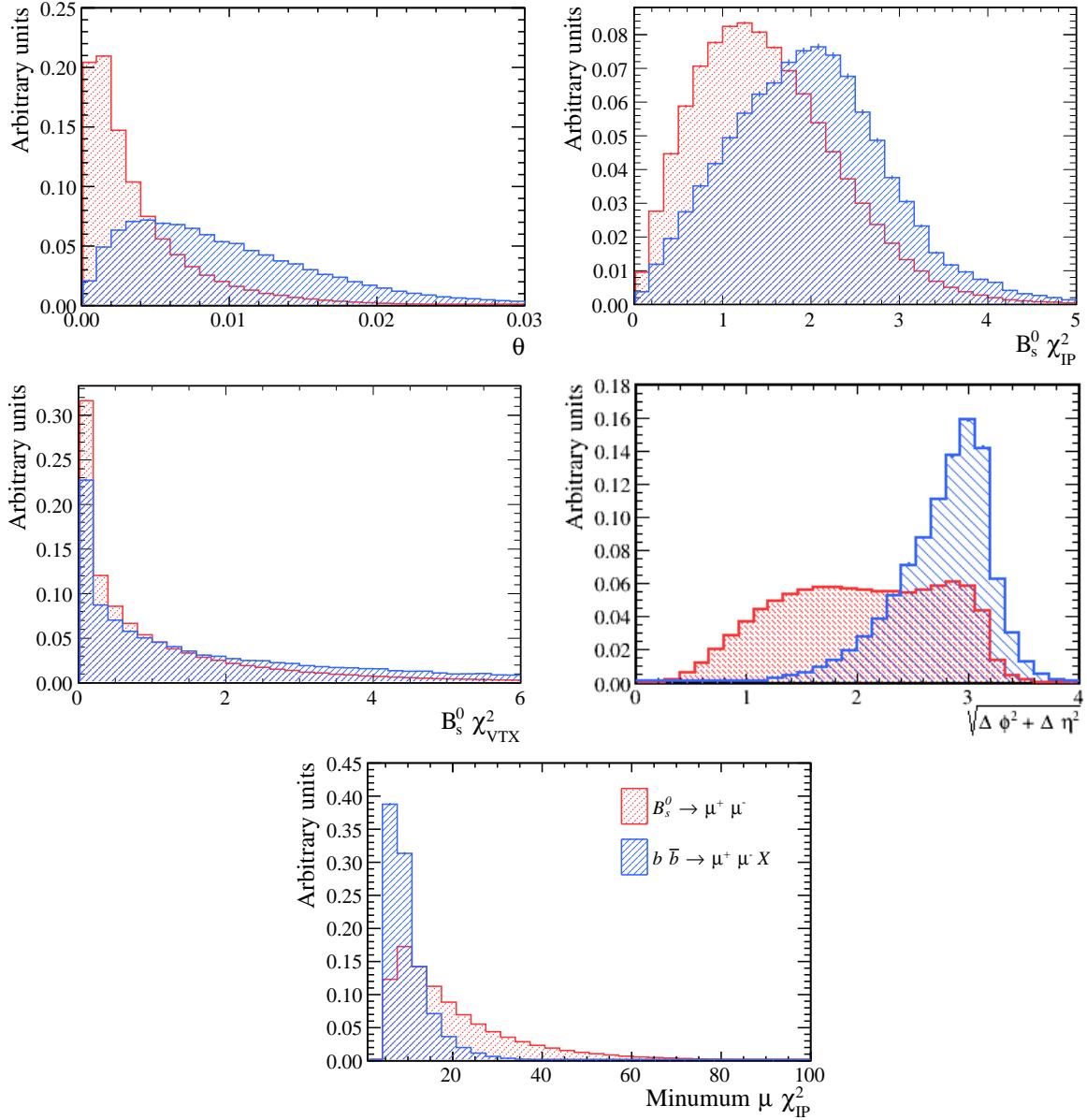


**Fig. 4.7** VELO track (left) and Long track (right) isolation distributions of simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays used to train the global BDT passing cuts in Table 4.12.

- VELO track isolation;
- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$ , 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  $\chi_{\text{IP}}^2$  with respect to the primary vertex of the  $B_s^0 \rightarrow \mu^+ \mu^-$  of the muons;
- $\chi_{\text{VTX}}^2$  of the  $B_s^0$ ;
- $\chi_{\text{IP}}^2$  of the  $B_s^0$  with respect to the primary vertex; and
- the angle between the momentum vector of the  $B_s^0$  and the vector connecting the production and decay vertices of the  $B_s^0$ .

A comparison of the signal and background distributions of the input variables in the training samples are shown in Figures 4.7 and 4.8. These variables were chosen by training a BDT beginning with the most discriminating variable, the Long track isolation, 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.

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.



**Fig. 4.8** 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.

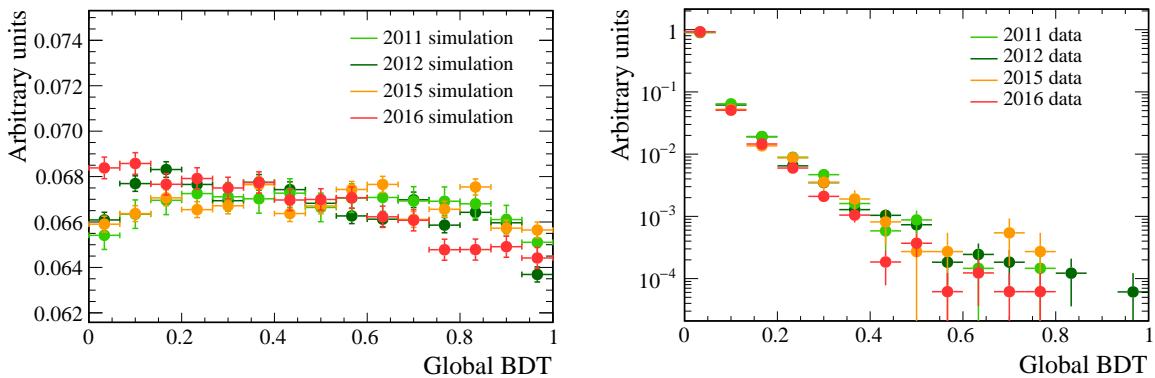
| Parameter   | Value |
|-------------|-------|
| nTrees      | 1000  |
| MinNodeSize | 1%    |
| MaxDepth    | 3     |
| $\beta$     | 0.75  |
| nCuts       | 30    |

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

| Selection applied to global BDT training samples. |   |
|---|---|
| $B_s^0$   | $\mu^\pm$                                     |
| $\chi_{\text{FD}}^2 > 225$                        | $p_T > 500 \text{ MeV}/c$                     |
| $\chi_{\text{IP}}^2 < 25$                         | $\chi_{\text{trk}}^2 < 3$                     |
| Vertex $\chi^2/ndof < 9$                          | minimum $\chi_{\text{IP}}^2 > 25$             |
| DOCA < 0.3 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                                 |
| DIRA > 0  | 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 BDT.

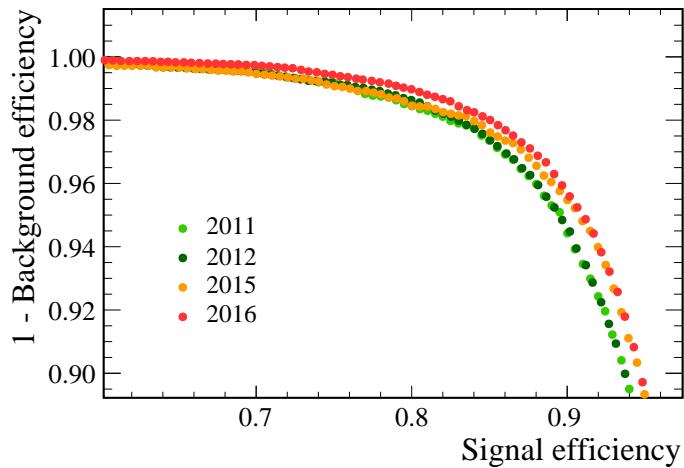
The global BDT is applied to data taken in all years and in the same way as the BDTS the final output of the global BDT is flattened to have a response between 0 and 1 that is uniform for signal and the background peaks at zero. The global BDT output for signal and background is shown in Figure 4.9 for each year of data taking. The flattening is important for the branching fraction measurements because a simultaneous fit is applied to the dimuon invariant mass in bins of BDT, flattening the BDT output enables bins containing equal proportions of signal decays to be easily created. The signal efficiency versus the background rejection of the global BDT is shown in Figure 4.10 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 then Run 1. A comparison of the input variables used in the global BDT for each year of data taking is given in Appendix A.



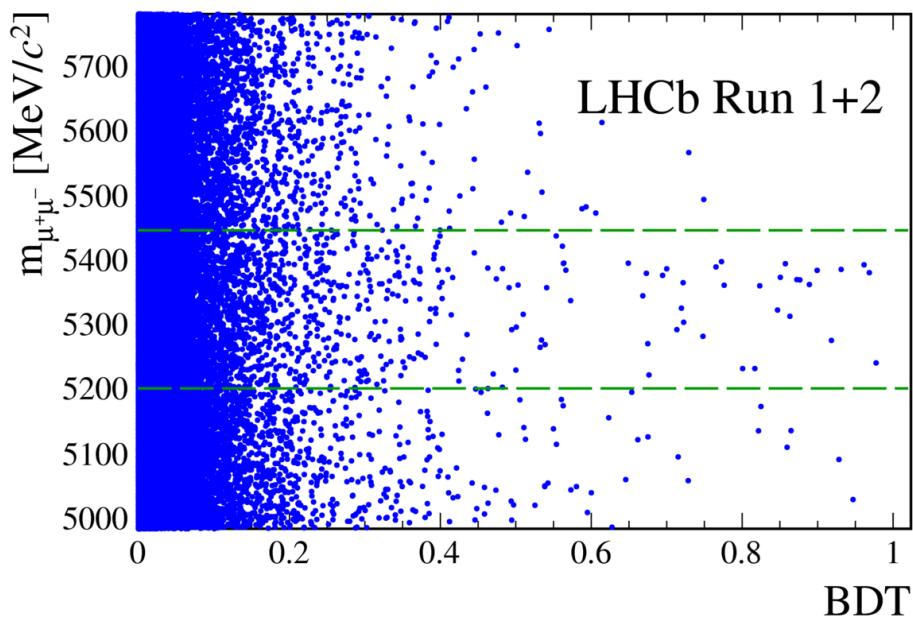
**Fig. 4.9** Global BDT output distributions for  $B_s^0 \rightarrow \mu^+\mu^-$  simulated decays (left) and  $b\bar{b} \rightarrow \mu^+\mu^- X$  decays in simulation and data.

### 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 4.11 shows a scatter plot of the mass and BDT values for all candidates that pass the selection criteria in Run 1 and Run 2 data.



**Fig. 4.10** 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.



**Fig. 4.11** 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.

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

**Table 4.13** Selection cuts applied to select  $B_s^0 \rightarrow \mu^+ \mu^-$ , where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

## 4.4 Selection for the effective lifetime measurement

The selection used to identify candidates for the effective lifetime measurement is based on the selection used to identify candidates for the branching fraction measurements. However, several changes are made to account for the different measurement strategies that are described in Chapters 5 and 6 and also only the  $B_s^0$  decay mode is required for the effective lifetime measurement. Similar to the branching fraction measurements,  $B \rightarrow h^+h'^-$  and  $B_s^0 \rightarrow J/\psi\phi$  decays are used to develop and validate the effective lifetime analysis strategy. The selection of  $B_s^0 \rightarrow J/\psi\phi$  decays is kept the same as that used for the branching fraction measurement but there are a few differences in the selection of  $B \rightarrow h^+h'^-$  decays for the effective lifetime measurement.

The majority of the selection of  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h'^-$  decays is kept the same as the selection for the branching fraction measurements; the same cut based selection in Section 4.3.2 is used and the BDTS requirement is applied. The differences in the selection are: the trigger requirements; the mass ranges; the particle identification requirements; and the use of multivariate classifiers. The differences are outlined in the following sections and the full selection criteria are summarised in Section 4.4.5.

### 4.4.1 Trigger requirements

The same trigger lines used in the branching fraction measurements selection are used to select candidates for the effective lifetime measurement but different trigger decisions are used. Candidates are required to be identified as TOS or TIS at each level of the trigger. The change in trigger decisions is motivated by the dependence of the effective lifetime measurement on simulated decays. The efficiency of the selection criteria varies with the decay time of each candidate, and therefore the selection efficiency as a function of decay time must be well understood in order to measure the effective lifetime. Simulated decays are used in the determination of this efficiency, as described in Section 6.3. However, the trigger efficiencies for candidates that are triggered as DEC, but not as TIS or TOS, are not well modelled in simulated decays. Therefore only candidates triggered by TOS or TIS decisions are used in order to accurately model the selection as a function of decay time. 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.

The analysis strategy used to measure the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime is verified by measuring the lifetimes of the more abundant  $B^0 \rightarrow K^+\pi^-$  and  $B_s^0 \rightarrow K^+K^-$  decays. Although the same trigger lines are used, slightly different trigger decisions are used

to select  $B \rightarrow h^+h'^-$  decays. To be useful as a validation channel the efficiency of the trigger requirements as a function of the decay time for  $B \rightarrow h^+h'^-$  decays will ideally 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.

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.

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

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

#### 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  $B \rightarrow h^+h'^-$  and semi-leptonic decays are almost completely removed. The dominant background left in the data set is from combinatorial background.

Similarly,  $B \rightarrow h^+h'^-$  decays used to verify the measurement strategy for the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime 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.

### 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,  $\text{PID}_\mu$ , 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.

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 bins of BDT and a simultaneous fit is then applied across the BDT bin to measure the branching fractions.

| Decay   | Particle  | PID requirements          |
|---|-----------|---------------------------|
| $B_s^0 \rightarrow \mu^+ \mu^-$ (Run 1 and 2015)              | $\mu^\pm$ | $\text{PID}_\mu > 0.2$    |
| $B_s^0 \rightarrow \mu^+ \mu^-$ (2016)                        | $\mu^\pm$ | $\text{PID}_\mu > 0.4$    |
| $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ \pi^-$ | $K^+$     | $\text{DLL}_{K\pi} > 10$  |
|   | $\pi^-$   | $\text{DLL}_{K\pi} < -10$ |
| $B_s^0 \rightarrow K^+ K^-$                                   | $K^\pm$   | $\text{DLL}_{K\pi} > 10$  |

**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.

A different, simpler strategy is used to identify candidates for the measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime. Combinatorial backgrounds are reduced by placing a cut on the output of a multivariate classifier. Then all 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. Therefore alternative classifiers were developed in parallel to the development of the global BDT and the performance was compared to determine the most effective way to select candidates for the effective lifetime measurement.

#### 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 [126]. The uBoost method produces a classifier output that has a uniform efficiency for a specified variable. The selection used to identify  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates uses variables that are correlated with the decay time of the  $B_s^0$  in the stripping selection and the most effective input variables for achieving good signal and background separation with a BDT are also correlated with the decay time. Therefore, the overall selection efficiency varies as a function of decay time. The final measurement of the effective lifetime relies on the efficiency being well

understood. 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 has a uniform efficiency across the decay time distribution.

Simulated 2012  $B_s^0 \rightarrow \mu^+ \mu^-$  decays were used as the signal training sample and two different samples were tested as the background training sample. One sample consisted of simulated 2012  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays and the other was 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, were applied to training samples of simulated decays. Combinatorial background decays in data were identified as candidates that pass the selection requirements listed in Table 4.12, except the BDTS requirement and candidates are required to have a dimuon invariant mass 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.

| 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/ $c^2$ .

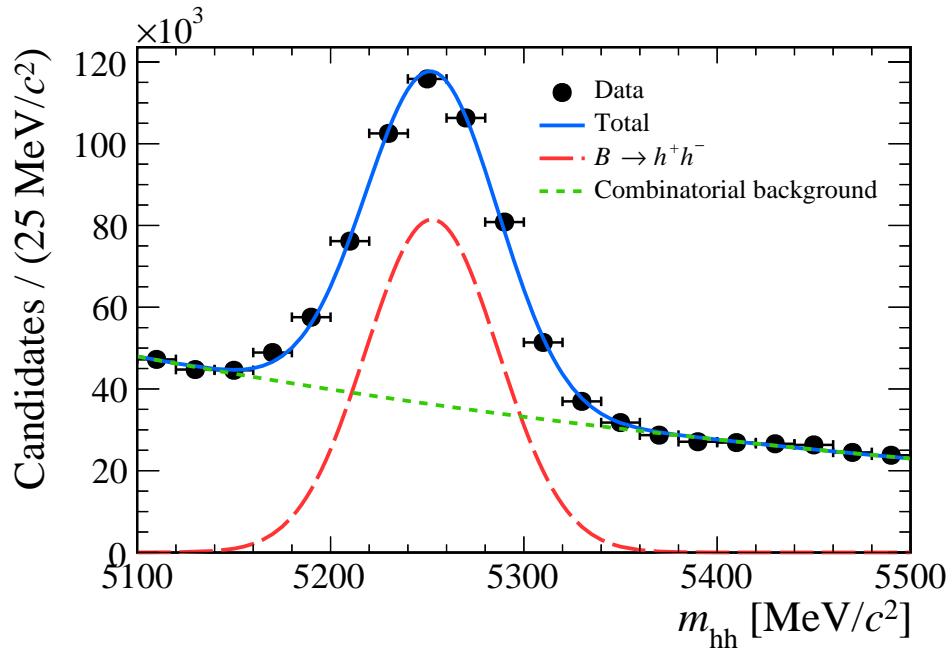
The input variables used in the adaptive boosting and uBoost BDTs were chosen separately, starting from a large set of variables including kinematic, geometric and isolation variables. Initially the BDTs were trained using all input variables within the set and variables that had no impact on the BDT 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 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.

The performance of the BDTs with different boosting methods and training using either simulated decays or data as the background sample was evaluated using  $B \rightarrow h^+h^-$  decays in data. No particle identification variables were used in the input variables of the BDTs due to the poor 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 and no particle identification requirements were used to separate different  $B \rightarrow h^+h^-$  decays. An unbinned maximum likelihood fit was applied to the  $B \rightarrow h^+h^-$  mass distribution, where all  $B \rightarrow h^+h^-$  decays are reconstructed as  $B_s^0 \rightarrow \mu^+\mu^-$ , for a range cuts on the BDT outputs and the signal significance  $\mathcal{S}$  was evaluated for each cut value. In the mass fit, the  $B \rightarrow h^+h^-$  mass distribution was modelled with a Gaussian function and the combinatorial background decays with an exponential function, an example of the mass fit is given in Figure 4.12. The signal significance is given by

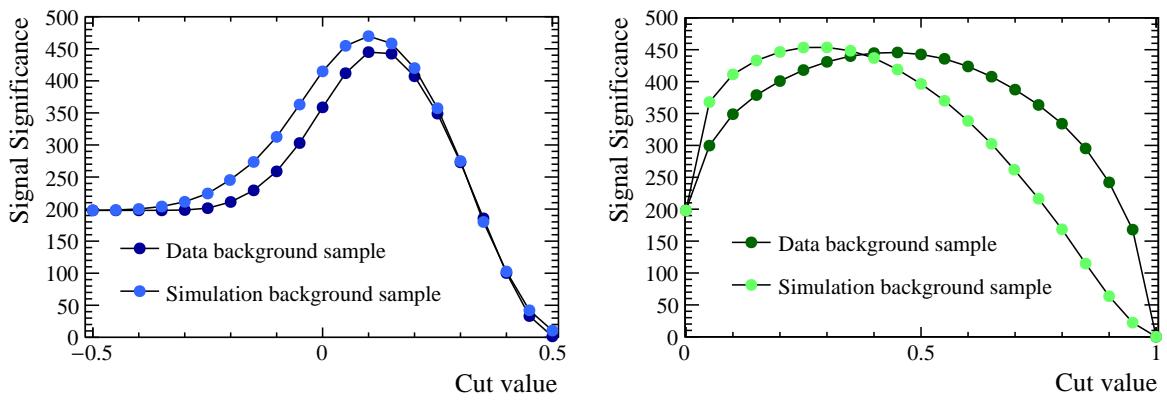
$$\mathcal{S} = \frac{S}{\sqrt{S+B}} \quad (4.3)$$

where  $S$  ( $B$ ) are the number of signal (background) decays within  $3\sigma$  of the centre of the  $B \rightarrow h^+h^-$  mass peak. The signal significance 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 4.13. 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. It is clear from Figure 4.13 that the performance of BDTs trained on simulated decays is better than that of the BDTs trained on data, this is due to the higher statistics available for simulated decays, as shown in Table 4.16. Furthermore the adaptive boosting BDT performs better than the uBoost BDT. This is expected because the adaptive BDT is not constrained to have a uniform efficiency across the decay time range. 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 4.13 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 gives the highest signal significances.



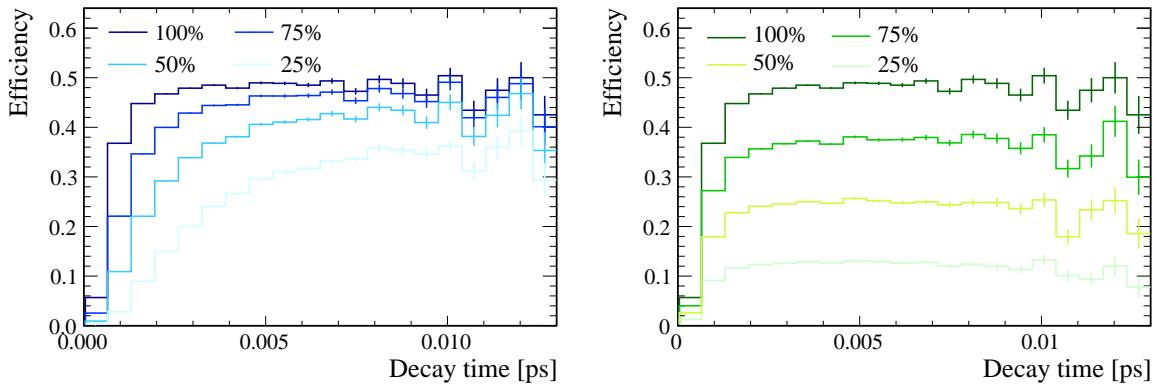
**Fig. 4.12** Example of the mass fit to  $B \rightarrow h^+ h^-$  Run 1 data to find the signal significance for the Adaptive BDT with a cut value at 0.0 on the BDT output.



**Fig. 4.13** Signal significance from  $B \rightarrow h^+ h^-$  decays in Run 1 data of the adaptive (left) and uBoost (right) BDTs trained using simulated decays and data as the background training samples.

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 [126]. The values used are given in Appendix B.

The understanding of the selection efficiency as a function of decay time was the main motivation for investigating the uBoost boosting method. 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 cut values are chosen to have the same selection efficiencies for each algorithm. The efficiencies are shown in Figure 4.14, although the shapes are quite different, the uniform efficiency of the uBoost BDT is evident. 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, 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.



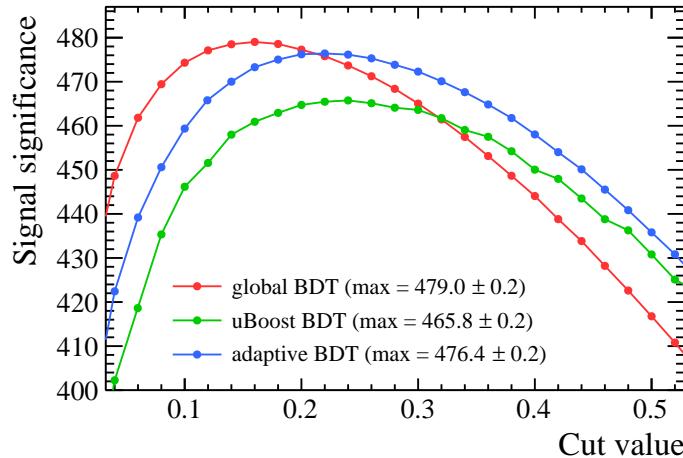
**Fig. 4.14** Selection efficiency as a function of decay time of simulated 2012  $B_s^0 \rightarrow \mu^+ \mu^-$  decays for the adaptive (left) and uBoost (right) BDTs. The selection requirements applied to the training sample are applied to the simulated decays and cuts are placed on the BDT output so that the efficiency of the cut on decays passing the other selection requirements is 100, 75, 50 and 25%.

#### 4.4.4.2 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. The performance of the two BDTs developed for the effective lifetime measurement is 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.

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. 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 4.15; the global BDT produces the highest signal significance, but is closely followed by the adaptive BDT developed for the effective lifetime measurement.



**Fig. 4.15** 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 and the mass range.

However, since the purpose of the BDT is to remove combinatorial background decays passing the  $B_s^0 \rightarrow \mu^+\mu^-$  selection, 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 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 [40] and all BDTs described in this dissertation have a better performance at removing background decays.

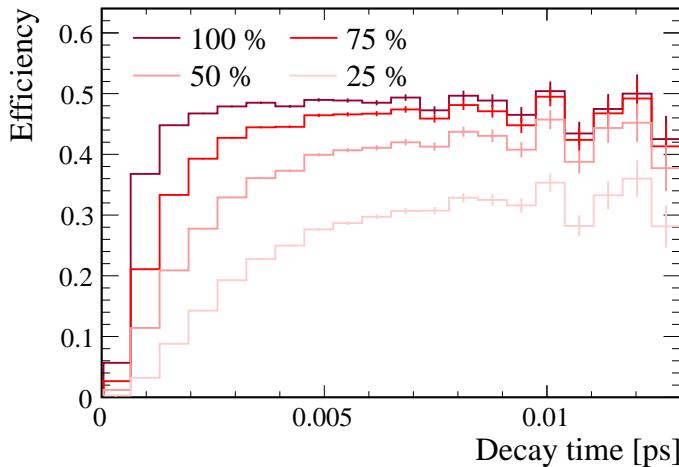
| BDT          | Number events above BDT output value |      |      |     |     |     |     |     |     |  |
|--------------|--------------------------------------|------|------|-----|-----|-----|-----|-----|-----|--|
|              | 0.1                                  | 0.2  | 0.3  | 0.4 | 0.5 | 0.6 | 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 MeV/c<sup>2</sup>. 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.

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 4.16 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 and a smooth decay time efficiency it is the best BDT to use for the selection of events for effective lifetime measurement.

#### 4.4.4.3 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 has been optimised to give the smallest expected uncertainty on the measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime,  $\tau_{\mu\mu}$ , and its inverse,  $\tau_{\mu\mu}^{-1}$ . This is done by using pseudoexperiments for the expected number of  $B_s^0 \rightarrow \mu^+ \mu^-$  combinatorial background decays for different cuts on the global BDT output.



**Fig. 4.16** 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 placed on the BDT output so that the efficiency of the cut on already selected events is 100, 75, 50 and 25%.

The fit procedure to extract  $\tau_{\mu\mu}$  from the data is described in Chapter 6 along with a discussion of whether it is best to fit for  $\tau_{\mu\mu}$  or  $\tau_{\mu\mu}^{-1}$ . The pseudoexperiment 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;
- an unbinned maximum likelihood fit is performed to the dimuon invariant mass spectrum, where the  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background yields are free to float in the fit along with the slope,  $\lambda$ , of the combinatorial background mass distribution; and
- the mass fit is used to compute sWeights using the sPlot method [127] and a maximum likelihood fit is performed to the sWeighted decay time distribution to extract  $\tau_{\mu\mu}$  and  $\tau_{\mu\mu}^{-1}$ .

Full details of the toy experiment set up and the probability density functions used are given in Appendix C.

The number of expected  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background events for different BDT cut values is derived from the expected number of decays passing in all the selection cuts and  $\text{BDT} > 0.55$ <sup>3</sup> for Run 1 and Run 2 data in the mass range 4900

<sup>3</sup>Initially the observed yields from published branching fraction measurements were used to determine the expected number of decays present in  $4.4 \text{ fb}^{-1}$  of Run 1 and Run 2 data and a global

$< m_{\mu^+\mu^-} < 6000 \text{ MeV}/c^2$ . These predictions assume the SM branching fraction for  $B_s^0 \rightarrow \mu^+\mu^-$  and are given in Table 4.18. The methods used to evaluate the expected number of each decay in the data are detailed in Chapter 5. Since the output of the global BDT is flattened, the number of  $B_s^0 \rightarrow \mu^+\mu^-$  decays is evenly distributed across the BDT range. Therefore the expected number of  $B_s^0 \rightarrow \mu^+\mu^-$  decays is straight forward to calculate for each BDT cut value. The number of combinatorial background decays expected after each BDT is computed from simulated  $b\bar{b} \rightarrow \mu^+\mu^- X$  decays, that have pass all the other selection criteria, using the ratio

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

where  $\epsilon(BDT > X)$  is the efficiency of the cut  $BDT > X$ . The ratios for the different cuts values are shown in Table 4.19. Simulated decays had to be used to compute the efficiencies rather than data because there were too few candidates left after the higher BDT cuts were applied to data to enable meaningful studies.

| Decay                          | Expected number of candidates |
|--------------------------------|-------------------------------|
| $B_s^0 \rightarrow \mu^+\mu^-$ | 30.94                         |
| Combinatorial background       | 66.23                         |
| Total                          | 97.17                         |

**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  $4900 < m_{\mu^+\mu^-} < 6000 \text{ MeV}/c^2$ .

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 4.17. The change in the slope value is accounted for in the mass distribution used in the pseudoexperiment by changing the slope parameter ( $\lambda$ ) for each BDT cut. Table 4.20 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 4.21 along with the expected number of  $B_s^0 \rightarrow \mu^+\mu^-$

---

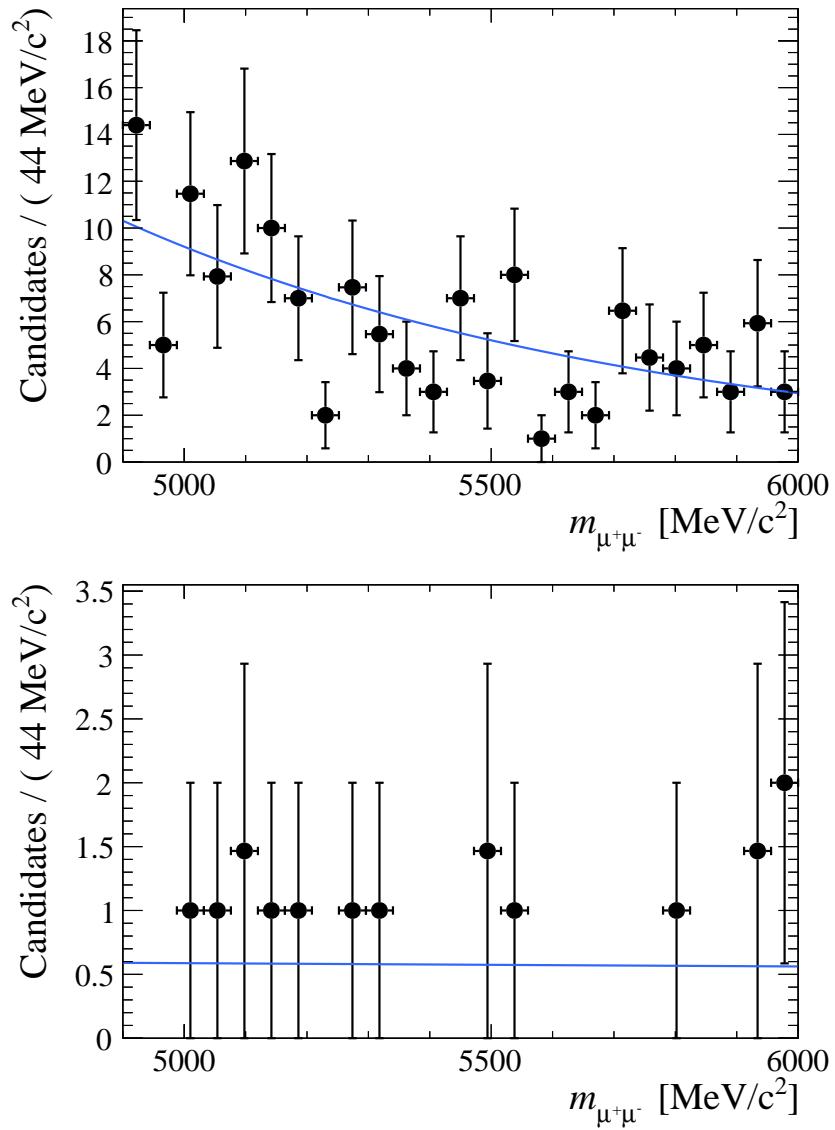
BDT cut of 0.55 was found to be optimal. However the expected number of decays was then re-evaluated using the more sophisticated techniques described in Chapter 5 and using global BDT cut of 0.55 and the pseudoexperiments were repeated to check the optimal BDT cut was the same.

| Global BDT cut | $R_\epsilon$ |
|----------------|--------------|
| 0.40           | 8.69         |
| 0.45           | 3.91         |
| 0.50           | 1.91         |
| 0.55           | 1.00         |
| 0.60           | 0.55         |
| 0.65           | 0.32         |

**Table 4.19** The ratio of efficiencies of cuts on the global BDT to select  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays relative to a cut of 0.55 on the global BDT.

| Global BDT cut | $\lambda / c^2 \text{MeV}^{-1}$ |
|----------------|---------------------------------|
| 0.40           | -0.00114 $\pm$ 0.00028          |
| 0.45           | -0.00129 $\pm$ 0.00041          |
| 0.50           | -0.00132 $\pm$ 0.00060          |
| 0.55           | -0.00004 $\pm$ 0.00089          |
| 0.60           | -0.00000 $\pm$ 0.00114          |
| 0.65           | -0.00024 $\pm$ 0.00122          |

**Table 4.20** The slope of the combinatorial background mass distribution for different cut value on the global BDT evaluated from  $b\bar{b} \rightarrow \mu^+\mu^-X$  simulated decays.



**Fig. 4.17** Mass distribution of simulated decays after global BDT cuts of 0.4 and 0.55 and the  $B_s^0 \rightarrow \mu^+\mu^-$  selection.

and combinatorial background decays for each BDT cut value. The median uncertainty of the fit for  $\tau_{\mu\mu}$  and  $\tau_{\mu\mu}^{-1}$  are given along with the signal significance ( $\mathcal{S} = S/\sqrt{S+B}$ ) 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.

| Global BDT cut | $\frac{S}{\sqrt{S+B}}$ | $\sigma(\tau_{\mu\mu}) / \text{ps}$ | $\sigma(\tau_{\mu\mu}^{-1}) / \text{ps}^{-1}$ |
|----------------|------------------------|-------------------------------------|---|
| 0.40           | 3.87                   | 0.345                               | 0.128   |
| 0.45           | 4.51                   | 0.309                               | 0.114   |
| 0.50           | 4.85                   | 0.291                               | 0.108   |
| 0.55           | 4.94                   | 0.285                               | 0.106   |
| 0.60           | 4.86                   | 0.297                               | 0.109   |
| 0.65           | 4.65                   | 0.309                               | 0.115   |

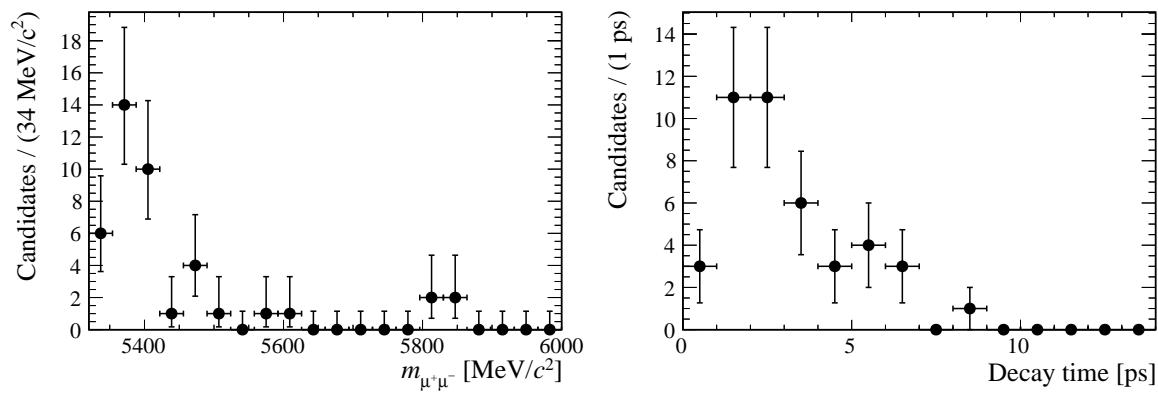
**Table 4.21** The signal significance for each cut value in the global BDT and median of the expected uncertainties for  $\tau_{\mu\mu}$  and  $\tau_{\mu\mu}^{-1}$  from 10,000 pseudoexperiments for the expected number of events.

#### 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 4.22. 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. The mass and decay time distributions for  $B_s^0 \rightarrow \mu^+\mu^-$  candidates passing the selection criteria in 4.4  $\text{fb}^{-1}$  of Run 1 and Run 2 data are shown in Figure 4.18.

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

**Table 4.22** Selection cuts applied to select  $B_s^0 \rightarrow \mu^+ \mu^-$ , where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.



**Fig. 4.18** Dimuon invariant mass (left) and decay time (right) distributions for  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates in  $4.4 \text{ fb}^{-1}$  of Run 1 and Run 2 data passing the selection requirements in Table 4.22.



# Chapter 5

## Measurement of $B \rightarrow \mu\mu$ 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 effective lifetime. 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 is 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 for these decays 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,  $\epsilon$ , of the detector, trigger, reconstruction and selection. 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  $\mathcal{N}_{B^0(s) \rightarrow \mu^+ \mu^-}$  is the total number of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays,  $(\mathcal{N}_{B_{(s)}^0})$  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  $B_{(s)}^0$  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 many uncertainties to cancel out. 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 ratios cancel out in the ratio, as well as uncertainties on  $f_{d(s)}$  depending on the choice of the normalisation channel. Therefore the number of observed  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays and the normalisation parameters,  $\alpha_{d(s)}$ , need to be evaluated to measure the branching fractions. The selection described in Chapter 4 allows  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidates to be classified by their dimuon invariant mass and global BDT output.

The branching fractions are measured by performing a simultaneous unbinned extended maximum likelihood fit [101, 128] to the dimuon invariant mass distribution in Run 1 and Run 2 data in four BDT bins. The simultaneous fit requires the probability density functions (PDFs) describing the mass distributions of signal and background decays, the fraction of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays in each BDT bin and the expected background

yields. The evaluation of the PDFs and expected yields are detailed in Sections 5.2 and 5.3.

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. 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 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  $J/\psi$ , 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.

The analysis is performed as a ‘blind’ analysis and the mass regions  $\pm 60$  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 [129]. 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,  $\mu$ ; and resolution,  $\sigma$ , of the Gaussian; the slope of the exponential,  $n$ ; and a parameter  $\alpha$ , defined in terms of  $\sigma$ , that determines the transition point between the Gaussian and the exponential function.

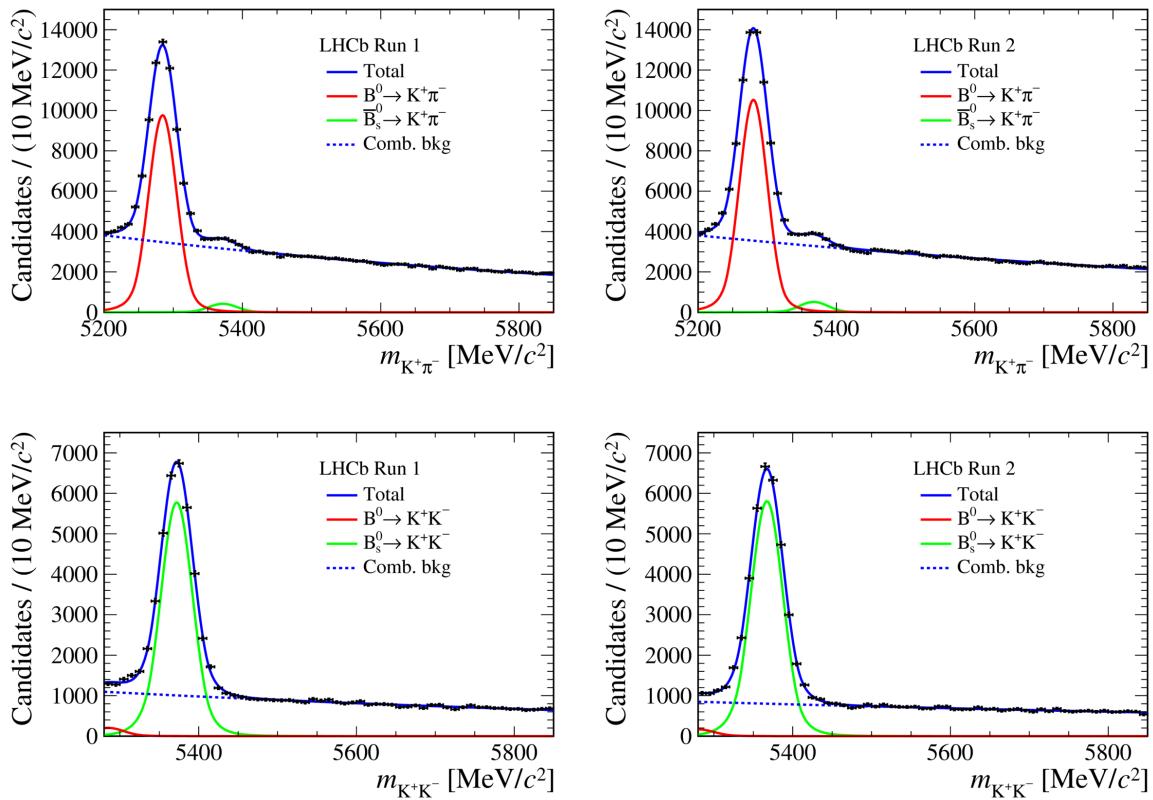
The signal shape parameters are evaluated in the following ways:

- $\mu$  - the mean's 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 in data;
- $\sigma$  - the resolution is interpolated from the resolutions of quarkonia resonances. The resolutions for the  $J/\psi$ ,  $\Psi(2S)$  and  $\Upsilon(1, 2, 3S)$  decaying into two muons are measured from fits to data. The  $B^0$  and  $B_s^0$  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  $\alpha$  - 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.

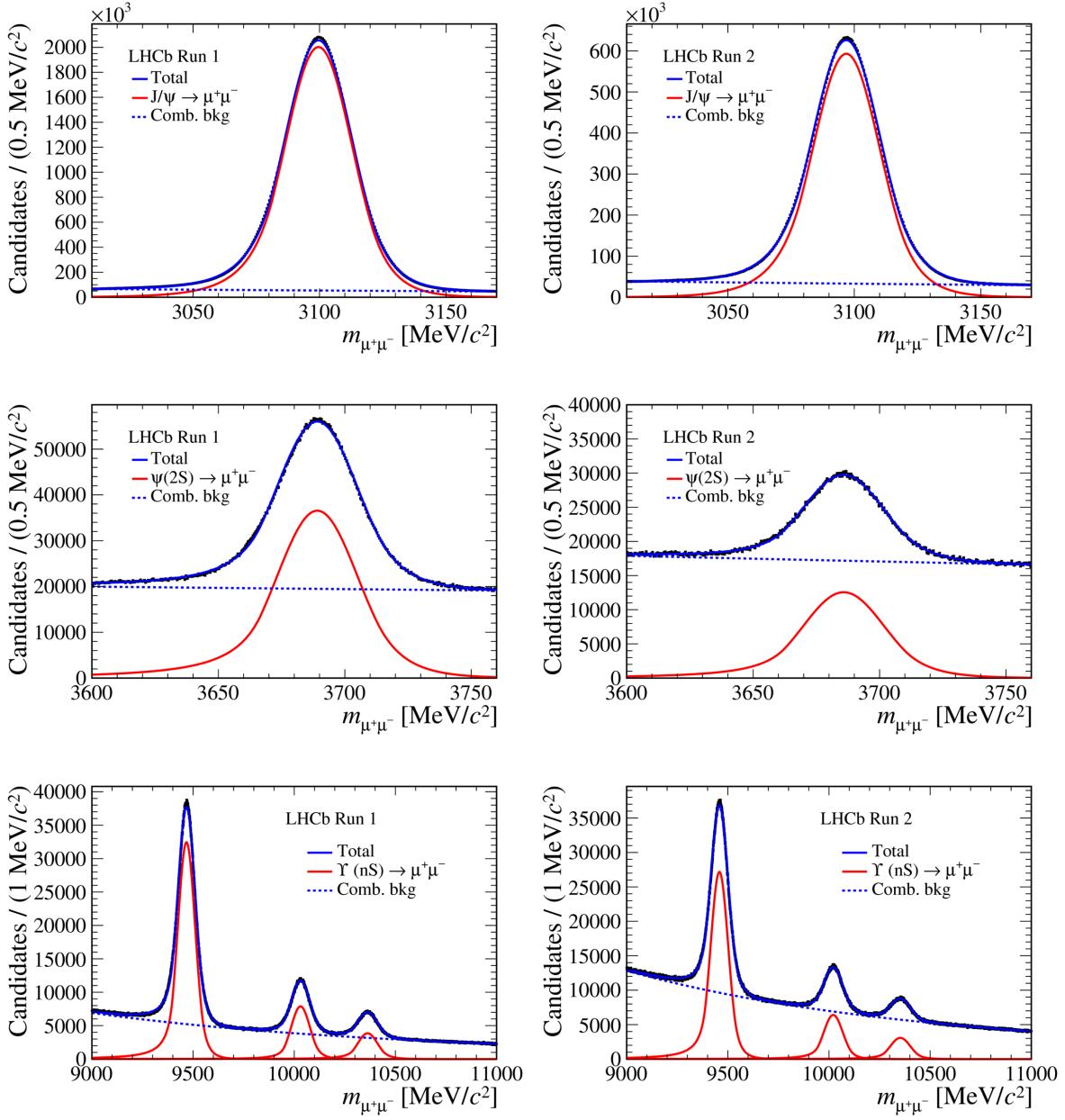
All parameters are evaluated separately for the  $B^0$  and  $B_s^0$  for the Run 1 and Run 2 data sets. The resulting parameter values are given in Tables 5.1 and 5.2 and the mass fits are shown in Figures 5.1, 5.2 and 5.3. The systematic uncertainties on the  $\mu$  values come from varying the particle identification cuts used to separate the different  $B \rightarrow h^+h^-$  decays and the systematic uncertainties on the  $\sigma$  come from the mass windows chosen for the quarkonia mass fits and the chosen mass fit model.

### 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 simply 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



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



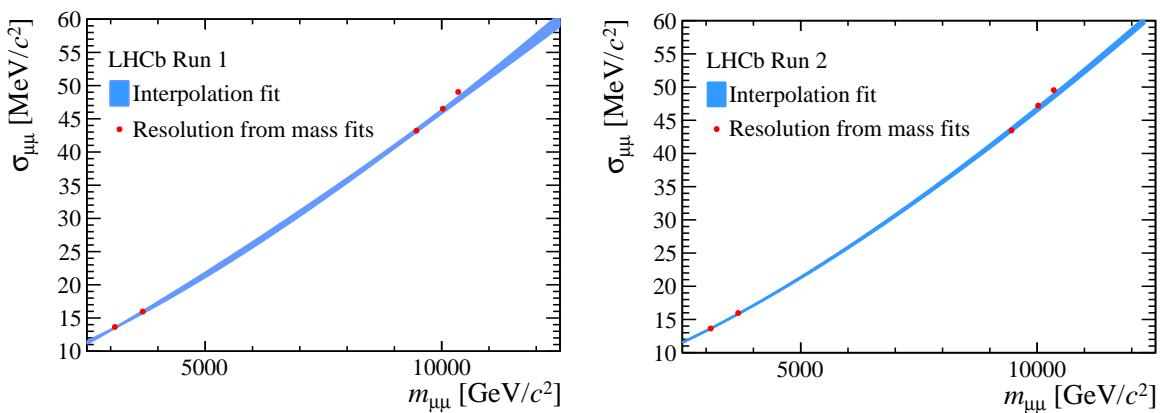
**Fig. 5.2** Maximum likelihood fit to the mass spectrum of  $J/\psi$  (top),  $\Psi(2S)$  (centre) and  $\Upsilon(1, 2, 3S)$  (bottom) decaying into two muons in Run 1 (left) and Run 2 (right) data.

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

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

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

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

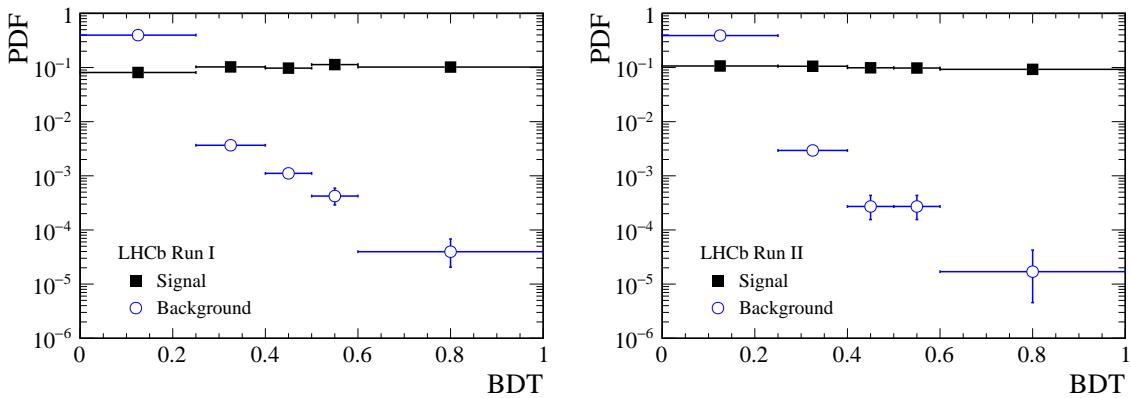


**Fig. 5.3** Power law fit to the resolution of quarkonia resonances to determine mass resolution for  $B^0 \rightarrow \mu^+\mu^-$  and  $B_s^0 \rightarrow \mu^+\mu^-$  decays on Run 1 (left) and Run 2 (right) data.

$B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays in each BDT bin, the BDT PDF is evaluated from 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 will be the same to a good approximation as  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays as they are kinematically identical.  $B^0 \rightarrow K^+\pi^-$  decays are used to calibrate the BDT response because it is the most abundant  $B \rightarrow h^+h^-$  decay.

The number of  $B^0 \rightarrow K^+\pi^-$  decays is extracted from data using maximum likelihood fits in each BDT bin for each year of data taking. 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. To reduce the difference in the trigger efficiency between and  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays,  $B^0 \rightarrow K^+\pi^-$  candidates are required to be TIS at L0 and Hlt1, but TOS at Hlt2, to ensure enough statistics.

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  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow \mu^+\mu^-$  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 window used, 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  $B^0 \rightarrow K^+\pi^-$  decay 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 but are too small to be visible on the plots.

### 5.2.3 Decay time dependence

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 IP  $\chi^2$  and isolation variables 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,  $\tau_H$ , 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  $\tau_L$  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  $\tau_H$  and  $\tau_L$  values from the PDG [48] at the time of the simulation production. 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  $= -1, 0$  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 which 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 (where  $h = K, \pi$ ) where 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<sup>1</sup> due to the missing energy from the undetected neutrino;
- semi-leptonic decays where one hadron is mis-identified as a muon that include

---

<sup>1</sup> $B^0$  and  $B_s^0$  mass windows are defined as  $\pm 60$  MeV/ $c^2$  of the  $B^0$  and  $B_s^0$  masses.

- $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$  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
- $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$  decays when the proton is mis-identified as a muon. The large mass of the  $\Lambda_b^0$  means that this background pollutes the  $B_s^0$  and  $B^0$  mass windows and below;
- decays where muons in the decay 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
  - $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  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/ $c^2$ ;
- combinatorial background formed by the random combination of any two muons in the event, this background is distributed across the full mass range.

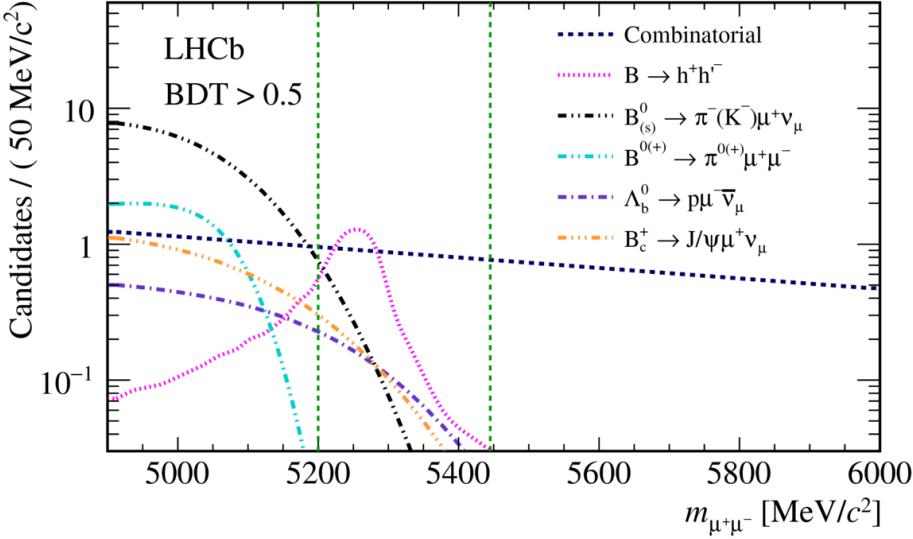
The backgrounds present in the data set must be included in the fit to the dimuon invariant mass in order to accurately measure the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions. Therefore the mass PDFs and expected yields of each background must be evaluated. The following sections summarise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

### 5.3.1 $B \rightarrow h^+ h'^-$

The mass PDF describing mis-identified  $B \rightarrow h^+ h'^-$  decays is formed of two Crystal Ball functions. The parameter values are evaluated from simulated decays for  $B^0 \rightarrow K^+ \pi^-$ ,  $B_s^0 \rightarrow K^+ K^-$ ,  $B^0 \rightarrow \pi^+ \pi^-$  and  $B_s^0 \rightarrow K^+ \pi^-$  in which the momenta of tracks is smeared to model the hadrons decaying in flight. The parameters are evaluated separately for each decay and combined using the branching fractions and the particle identification efficiencies for each decay.

The number of mis-identified  $B \rightarrow h^+ h'^-$  decays in each BDT bin,  $\mathcal{N}_{B \rightarrow hh \rightarrow \mu\mu}$ , is found using the relationship

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



**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  $B \rightarrow h^+h'^-$ ,  $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ ,  $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ ,  $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ ,  $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ ,  $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$  and combinatorial background. The green dashed lines show the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  mass window.

where  $\mathcal{N}_{B \rightarrow hh}$  is the number of TIS  $B \rightarrow h^+h'^-$  decays in data,  $\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-, B \rightarrow hh}^{\text{TRIG}}$  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  $B \rightarrow h^+h'^-$  decays is mis-identified as  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ . The mis-identification probabilities are evaluated using the PIDCalib package[122]. The number of TIS  $B \rightarrow h^+h'^-$  decays is calculated from the number of  $B^0 \rightarrow K^+\pi^-$  decays in the full BDT range. The number of TIS  $B^0 \rightarrow K^+\pi^-$  decays in the full BDT range is used to determine the number of inclusive  $B \rightarrow h^+h'^-$  decays 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. Apart from the trigger and particle identification requirements the same selection is used for  $B^0 \rightarrow K^+\pi^-$  decays as  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ , therefore only the trigger and particle identification efficiencies are accounted for. The efficiencies are calculated using a combination of data and simulated decays for each BDT bin and the same BDT PDF as  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays is assumed for  $B \rightarrow h^+h'^-$  decays.

### 5.3.2 Semi-leptonic decays

The mass PDFs of semi-leptonic backgrounds vary across the BDT range. Therefore these PDFs are evaluated using simulated decays separated into each BDT bin. An Argus function [130] convoluted with a Gaussian function is used to describe the mass

distributions. 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 expected yields of the semi-leptonic backgrounds in each BDT bin is estimated by normalising to the number of  $B^+ \rightarrow J/\psi K^+$  decays observed 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  $\beta$  combines the background yield, detection and selection efficiency and hadronisation factors of  $B^+ \rightarrow J/\psi K^+$  decays and it is the same for all backgrounds. The  $\beta$  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,  $\epsilon_x$ , are evaluated separately for each BDT bin from information from both data and simulated decays. The hadronisation factors and branching fractions are specific to each background and were possible measured, rather than predicted, branching fractions are used.

### 5.3.3 Combinatorial background

In the fit to the dimuon invariant mass the combinatorial background is modelled by an exponential function. The combinatorial background yield is not constrained in the fit and the slope is required to have the same value across all BDT bins for each data set. These parameters are determined from a simultaneous fit to candidates in data in BDT bins for the mass ranges  $[4900, (m_{B^0} - 60)]$  MeV/ $c^2$  and  $[(m_{B_s^0} + 60, 6000]$  MeV/ $c^2$ , where the mass shapes and yields of the remaining backgrounds are constrained as described in the following sections.

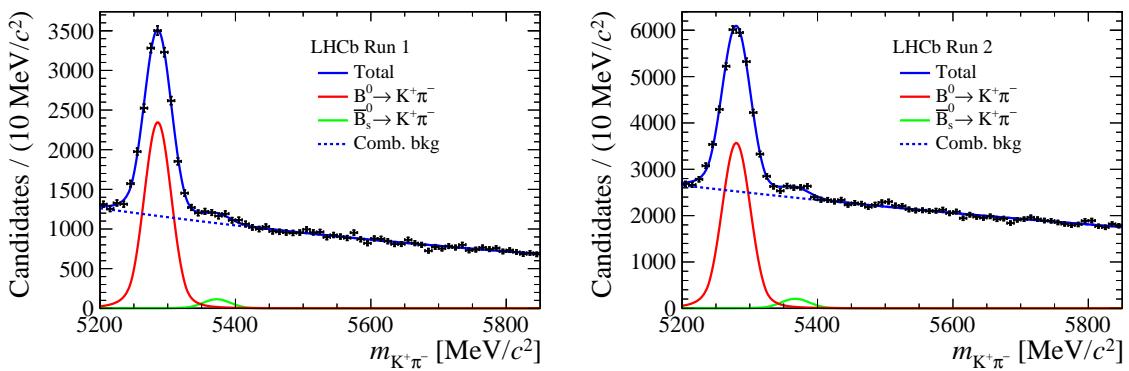
## 5.4 Normalisation

As introduced earlier 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  $B_{(s)}^0 \rightarrow$

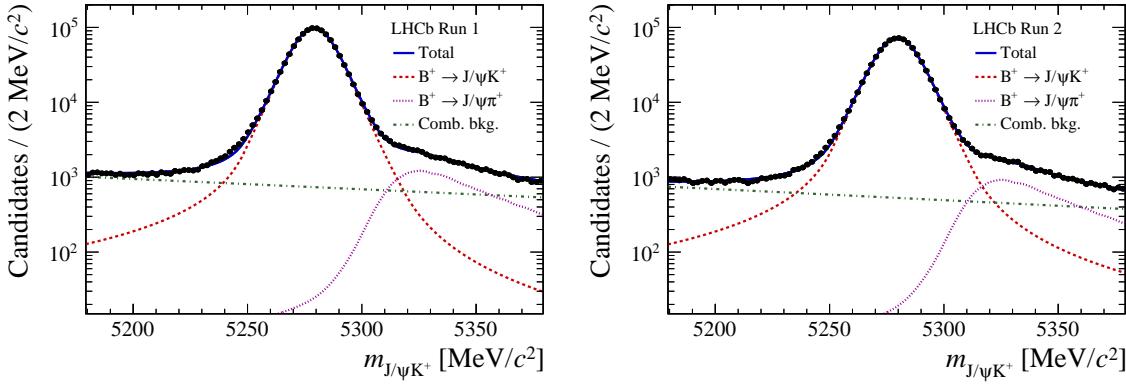
$\mu^+\mu^-$  decays depend on the yields of the normalisation decays, the ratio of the detection and selection efficiencies and the hadronisation factors. The evaluation of each of these terms is described in the following sections. In addition to the normalisation channels,  $B_s^0 \rightarrow J/\psi\phi$  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.

#### 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,  $\mathcal{N}_{norm}^{obs}$ , are calculated from data using an extended maximum likelihood fits to Run 1 and Run 2 data separately. The  $B^+ \rightarrow J/\psi K^+$  mass PDF is modelled by an Ipathia function [131] and the fit includes components for combinatorial background and  $B^+ \rightarrow J/\psi\pi^+$  decays that are mis-reconstructed as  $B^+ \rightarrow J/\psi K^+$ . The mass PDF parameters are determined from using a mixture of information from data and simulated decays. The  $B^0 \rightarrow K^+\pi^-$  yields are calculated in the same way as the BDT calibration with the same trigger requirements. However, to get the normalisation, the total number of  $B^0 \rightarrow K^+\pi^-$  decays across the full BDT range is required rather than the bin-by-bin yields computed previously. Figure 5.6 and 5.7 show the mass fits used to calculate the Run 1 and Run 2  $B^0 \rightarrow K^+\pi^-$  and  $B^+ \rightarrow J/\psi K^+$  yields.



**Fig. 5.6** Mass fit to measure  $B^0 \rightarrow K^+\pi^-$  yield for the normalisation for Run 1 (left) and Run 2 (right) 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.



**Fig. 5.7** Mass fit to measure  $B^+ \rightarrow J/\psi K^+$  yield for the normalisation for Run 1 (left) and Run 2 (right) 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.

#### 5.4.2 Efficiency ratio

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

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

where  $\epsilon^{Acc}$  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 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  $B^0 \rightarrow K^+\pi^-$  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 [132, 50].

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 in the normalisation parameters ensures that systematic uncertainties arising from the use of simulated decays cancel 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 which produce  $B^\pm$ ,  $B^0$  or  $\bar{B}^0$  and  $B_s^0$  or  $\bar{B}_s^0$  mesons. Given the hadronisation factors  $f_d$  and  $f_u$  are equal to good approximation, 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 [133]. 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 decays at 8 and 13 TeV relative to their ratio at 7 TeV. The reconstruction, selection and trigger efficiencies were corrected for in the ratios. The ratios are stable across the different collision energies and  $f_s/f_d$  is assumed to be identical for  $\sqrt{s} = 8$  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 decays 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  $B_s^0 \rightarrow \mu^+\mu^-$  decay 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 [48]. 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  $B^+ \rightarrow J/\psi K^+$  are used to produce the overall normalisation factors for Run 1 and Run 2 as shown in Table 5.3.

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

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

## 5.5 Results

The  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow \mu^+\mu^-$  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. 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  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  in each BDT bin is constrained using the BDT PDF and the yields of the mis-identified background are constrained around their expectations. 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

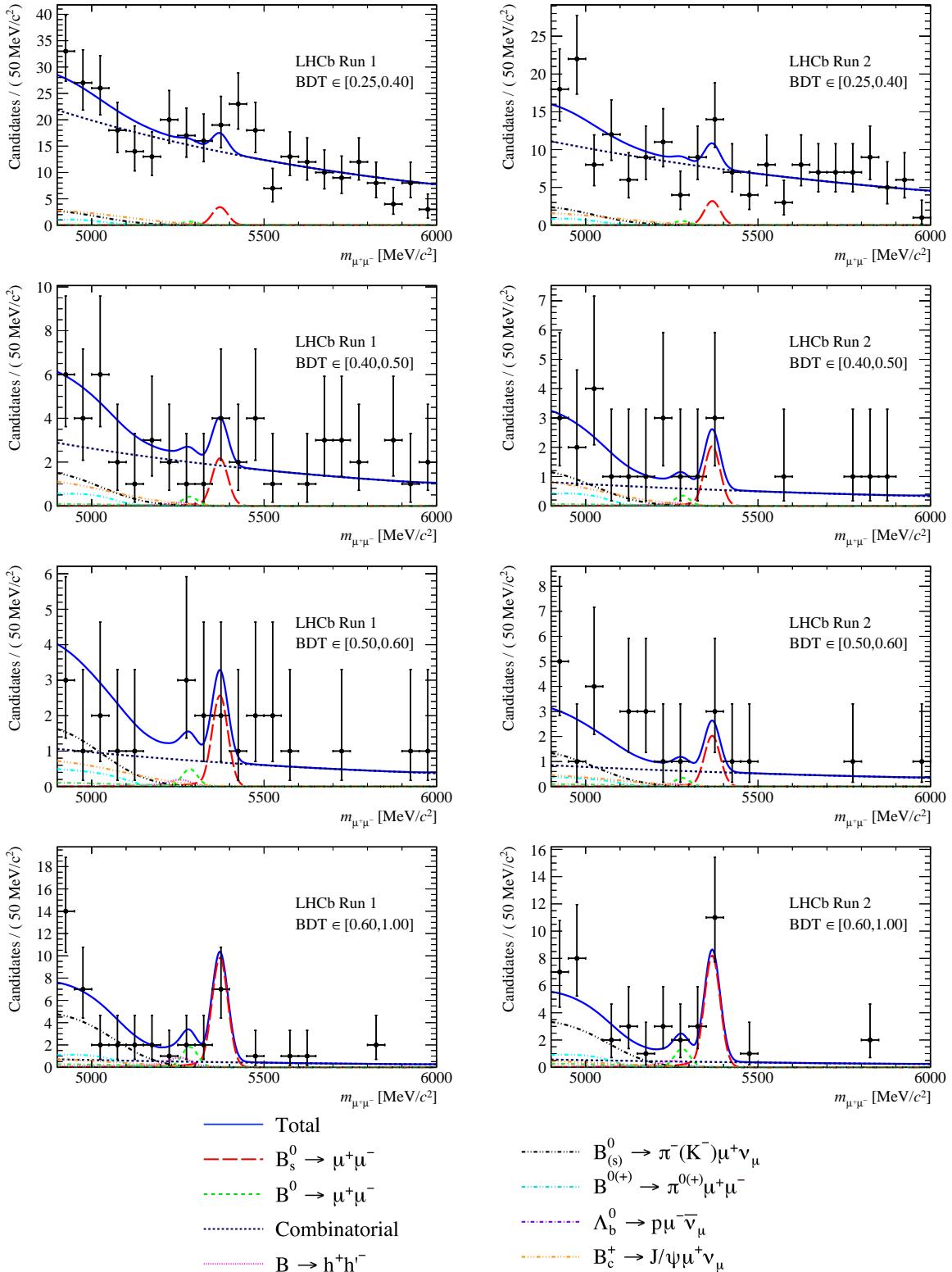
$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) &= (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) &= (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}. \end{aligned} \quad (5.9)$$

where the statistical and systematic uncertainties are given. The dominant contributions to the systematic uncertainties are from the ratio  $f_s/f_d$  and the uncertainty on the backgrounds for the  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow \mu^+\mu^-$  branching fractions, respectively. Figure 5.8 shows the fit results for  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  candidates in the four BDT bins for both Run 1 and Run 2 data. The statistical significance of the  $B_s^0 \rightarrow \mu^+\mu^-$  signal is  $7.8\sigma$  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  $1.6\sigma$ , therefore the CL<sub>s</sub> method [134] 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 quoted  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction assumes the Standard Model value for  $A_{\Delta\Gamma}$ , applying the corrections detailed in Section 5.2.3;  $A_{\Delta\Gamma}$  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.



**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, mis-identified  $B \rightarrow h^+h^-$  decays and backgrounds from semi-leptonic decays.

# 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 are described in Sections 6.2 and 6.3 for signal and background candidates. Due to the very rare nature of  $B_s^0 \rightarrow \mu^+\mu^-$  decays, the measurement strategy has been optimised to produce the smallest expected statistical uncertainty on the measured  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime, 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 dataset 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 using the integrated luminosity of  $4.4 \text{ fb}^{-1}$  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 [127], 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 and each background decay. 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 dataset, 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 should have been removed by the sPlot technique. Due to the low statistics expected for the data set, Run 1 and Run 2 data are combined and the maximum likelihood fit to the mass and weighted decay time distributions are performed to the combined data.

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 dimuon invariant mass can be used to accurately determine sWeights applied to measure the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime.

| 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 full effective lifetime selection is applied to simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays and decays in data must pass the effective lifetime selection requirements apart from the global BDT cut and have a dimuon invariant mass of  $> 5447$  MeV/ $c^2$ .

The sWeights are calculated using the RooFit package [135]. 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. Therefore the

sWeights are re-normalised using

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

where  $\omega_i$  are the sWeights values for each decay and  $\omega_j$  are weights 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 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 in the dataset 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. However, the overall performance of this strategy depends on the maximum likelihood fit to the invariant mass distribution; how many background components are included in the fit and the mass range the fit covers. The determination of the final fit configuration was performed using toy studies, described in Section 6.4, that study a variety of different mass ranges, the largest being 4900 - 6000 MeV/ $c^2$ . Therefore, the development of the fit configuration requires the mass and decay time PDFs of all backgrounds within the largest mass range to be known as well as the signal PDFs.

## 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 various components of the background decays passing the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime selection and in the mass range 4900 - 6000 MeV/ $c^2$  are the same as those passing the branching fraction selection, although the yields will be different. Consequently the invariant mass fit used to extract the sWeights is very similar to the fit to that used for the branching fraction measurements in Sections 5.2 and 5.3. The PDF used in the mass fit has the form

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

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 Run 1 and Run 2 data sets are combined for the measurement of the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime and only one mass PDF is needed to describe  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays in data. 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.

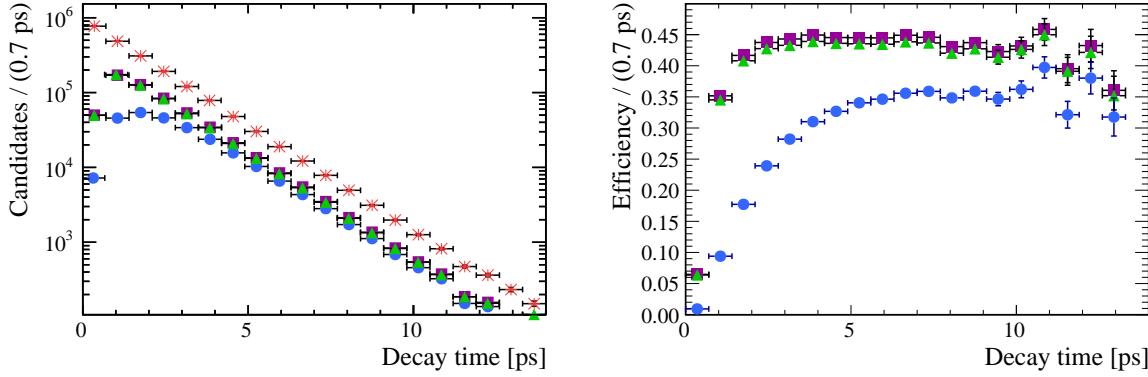
Backgrounds from mis-identified  $B \rightarrow h^+h'^-$  decays are described by the double Crystal Ball function evaluated using the method described in Section 5.3 with the effective lifetime particle identification requirements applied. Finally, the combinatorial background is modelled with a decaying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

The mass PDFs used in the toy studies, described in Section 6.4, for the signal and backgrounds are evaluated for the mass range 4900 to 6000 MeV/ $c^2$ .

### 6.3 Decay time PDFs

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. The bias arises because variables used in the selection and the global BDT, such as the isolation criteria, the  $B$  meson impact parameter and  $B$  meson flight distance significance, are correlated with the decay time. Consequently, cuts placed on these variables have a non-uniform efficiency across the decay time range. Therefore, the PDF describing the decay time changes from a decaying exponential to

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



**Fig. 6.1** Decay time distribution (left) and selection efficiency as a function of decay time (right) 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).

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 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 on  $B_s^0 \rightarrow \mu^+\mu^-$  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 mass fit configuration. The background PDFs used are described in Section 6.3.2.

### 6.3.1 $B_s^0 \rightarrow \mu^+\mu^-$

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 [136]

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

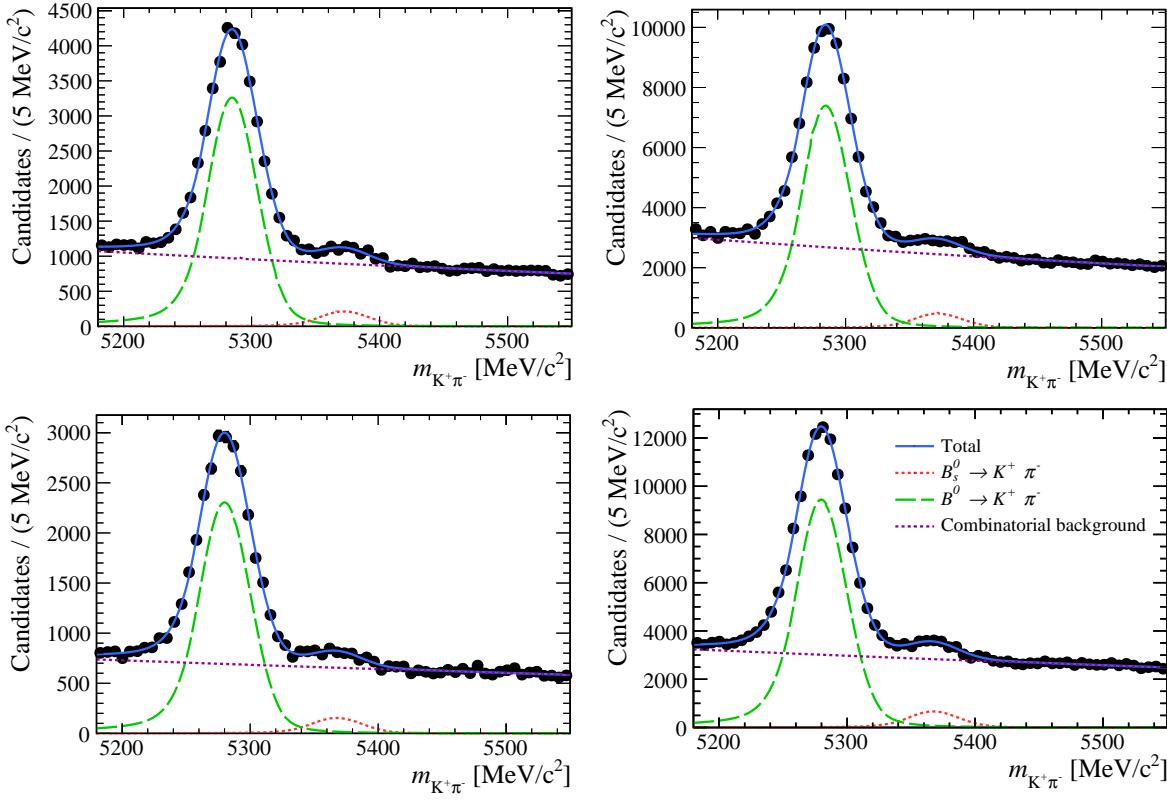
where  $t$  is the decay time,  $t_0$  the shortest decay time allowed for the function and  $a$  and  $n$  are parameters to be determined. was found to best describe the  $B_s^0 \rightarrow \mu^+\mu^-$  decay time efficiency. The acceptance function parameters are taken from a fit to simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays and are fixed in the fit to data. The parameters could not be determined from data because there are too few  $B_s^0 \rightarrow \mu^+\mu^-$  decays in data and the efficiency distribution of the more abundant  $B \rightarrow h^+h^-$  decays after the selection is quite different to that of  $B_s^0 \rightarrow \mu^+\mu^-$ . A systematic uncertainty describing how well the acceptance function is understood is detailed in Section 7.4. The decay time efficiency for each year of data taking is slightly different therefore simulated decays from each year of data taking must be used to determine the acceptance parameters.

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 in an event 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 selection requirements listed in Table 4.22 are used to identify  $B^0 \rightarrow K^+\pi^-$  decays in data and simulation but importantly the global BDT cut is not applied. The  $DLL_{K\pi}$  variable is used to separate  $B^0 \rightarrow K^+\pi^-$  decays from other  $B \rightarrow h^+h^-$  decays in data and the loose trigger requirements used for the branching fraction analysis are applied to data and simulation to keep a high trigger efficiency<sup>1</sup>. The same requirements are applied to simulated decays. The distribution of the number of tracks present in events containing  $B^0 \rightarrow K^+\pi^-$  decays is obtained from data by performing a maximum likelihood fit to the  $B^0$  mass distribution and extracting sWeights. 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.

---

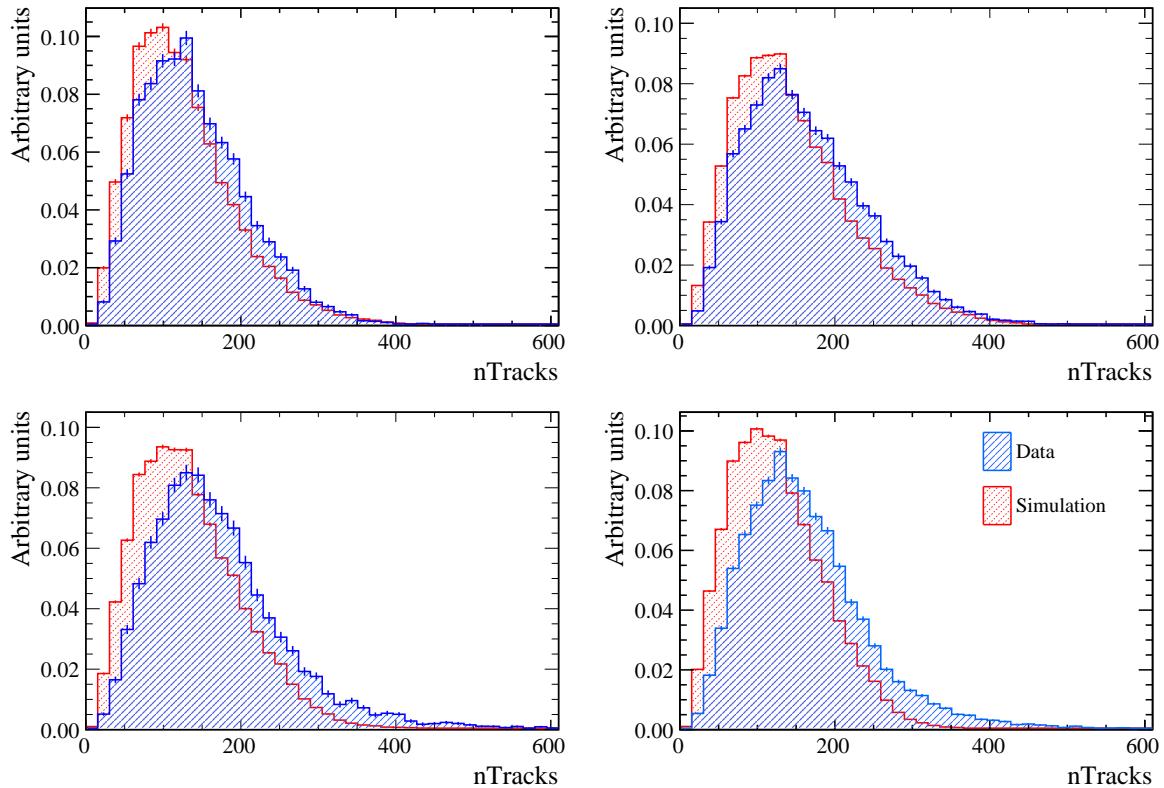
<sup>1</sup>The Hlt2Phys Dec trigger decision was not correctly implemented in 2016 simulated decays, therefore the DEC decisions of a combination of trigger lines designed to select  $B \rightarrow h^+h^-$  are used to emulate the Hlt2Phys DEC trigger decision. The trigger lines are Hlt2Topo2BodyDecision, Hlt2B2HH\_Lb2PPiDecision, Hlt2B2HH\_Lb2PKDecision Dec, Hlt2B2HH\_B2PiPiDecision, Hlt2B2HH\_B2PiKDecision, Hlt2B2HH\_B2KKDecision and Hlt2B2HH\_B2HHDecision. These trigger lines are applied to both data and simulated decays.



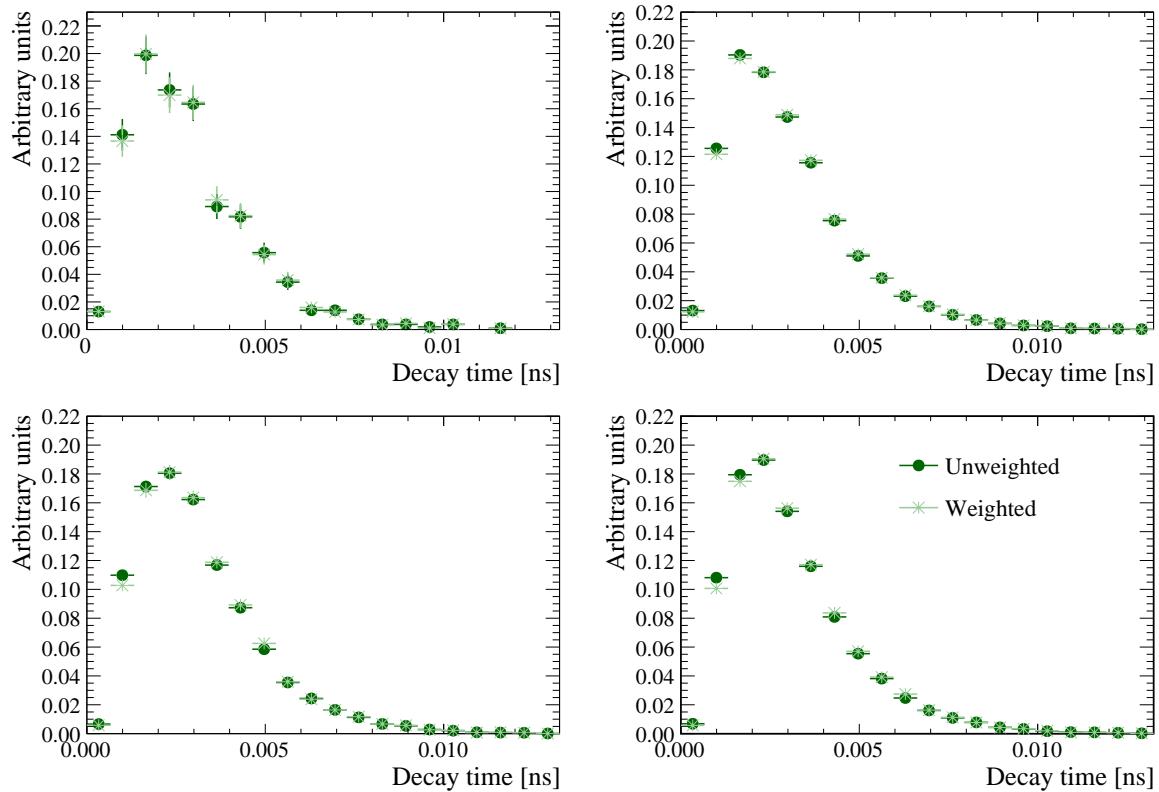
**Fig. 6.2** Maximum likelihood fits to the mass distribution of  $B^0 \rightarrow K^+\pi^-$  candidates in 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data. The mass PDF includes components for  $B^0 \rightarrow K^+\pi^-$  (green),  $B_s^0 \rightarrow K^+\pi^-$  (red) and combinatorial background (purple).

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 the number of tracks per event in data and simulation for each year. The affect on the decay time distribution of using these weights and then applying the global BDT cut 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.

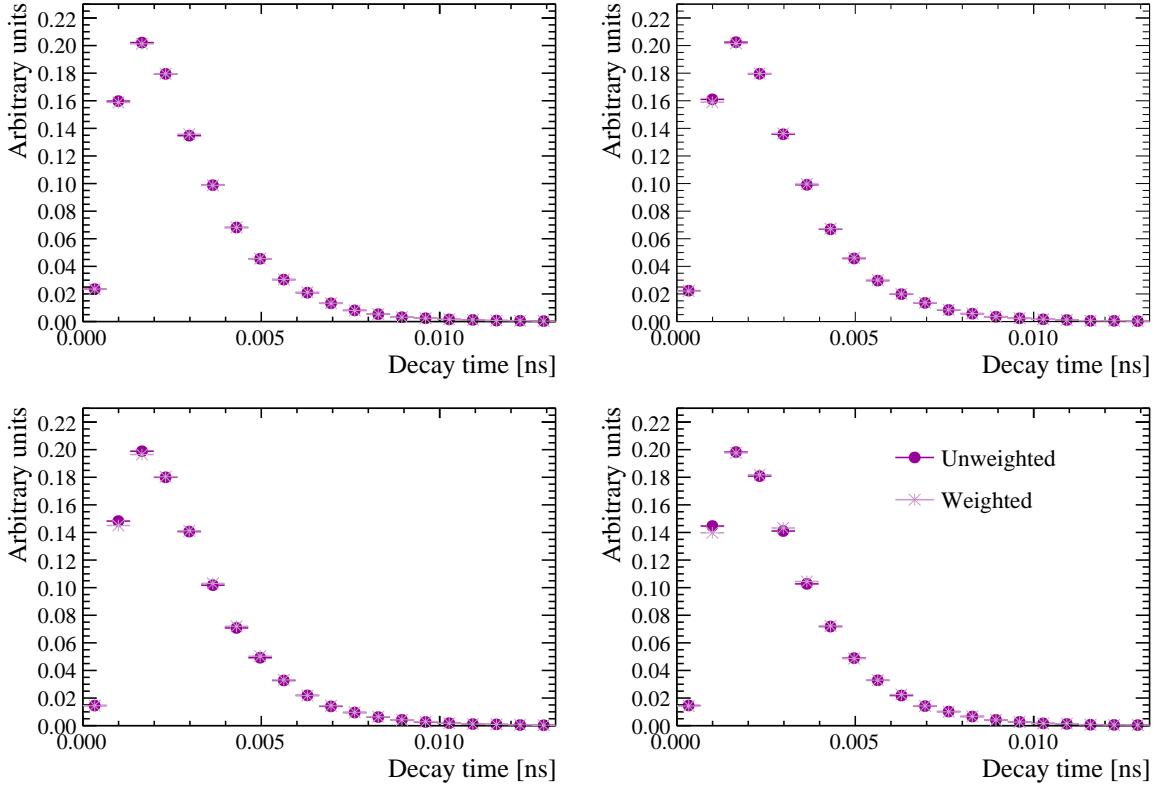
The same weights are applied to simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays by binning the number of tracks per event for  $B_s^0 \rightarrow \mu^+\mu^-$  decays in the same way as used for  $B^0 \rightarrow K^+\pi^-$  decays. The weights are applied to decays that pass the selection but before the global BDT cut is applied. The change in the decay time distribution before and after reweighting for simulated decays after the global BDT cut is shown in Figure 6.5. Similar to  $B^0 \rightarrow K^+\pi^-$



**Fig. 6.3** Normalised histograms of the number of tracks per event in simulated  $B^0 \rightarrow K^+\pi^-$  decays and weighted  $B^0 \rightarrow K^+\pi^-$  decays in data for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data.



**Fig. 6.4** Decay time distributions for weighted and un-weighted  $B^0 \rightarrow K^+ \pi^-$  simulated decays for for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data taking conditions.

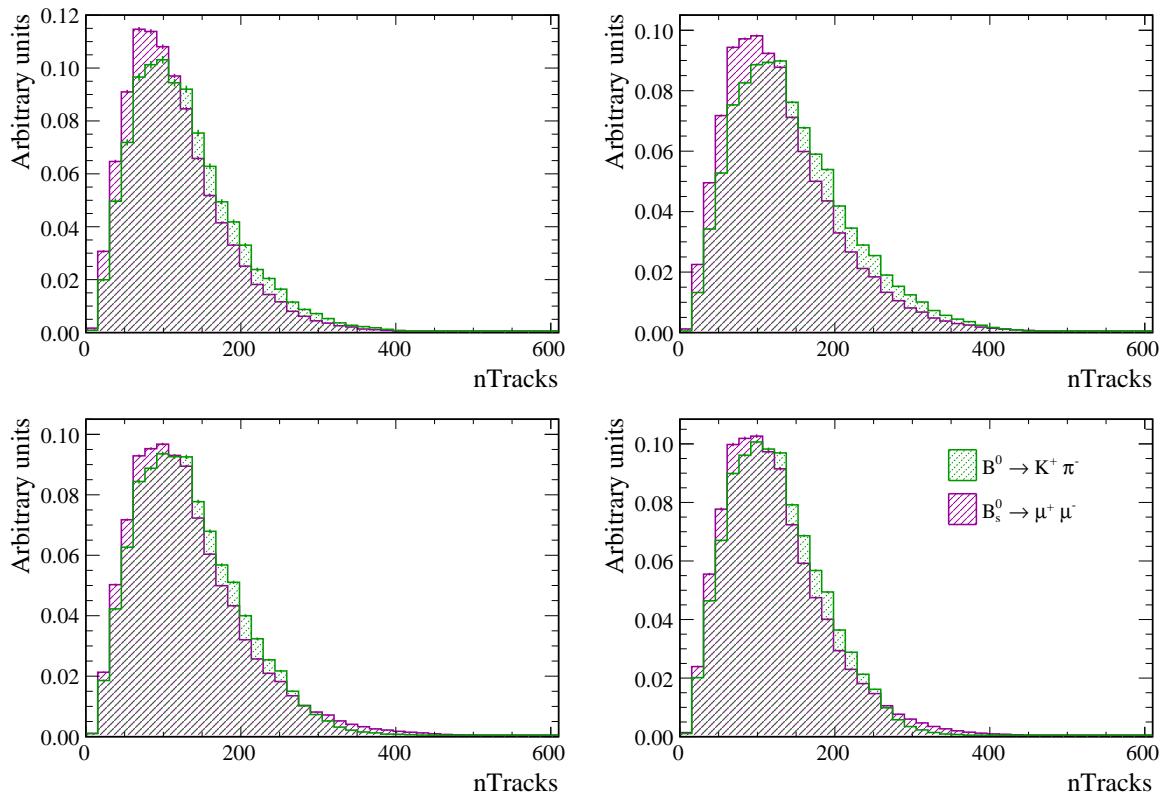


**Fig. 6.5** Decay time distributions for weighted (blue) and un-weighted (red)  $B_s^0 \rightarrow \mu^+\mu^-$  simulated decays for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data taking conditions. Distributions have been normalised to have unit area.

decays the largest effect is at low decay times where the change in selection efficiency is greatest as seen in Figure 6.1.

The reweighting relies on the number of tracks per event being very 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 rather similar.

The decay time efficiency will now be accurately modelled in the weighted simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays and the parameters in the acceptance function can be evaluated. Each year of data taking is treated separately due to the different  $B_s^0$  lifetimes used in the simulation generation, selection and trigger efficiencies. 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, 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 in each year is taken from the number



**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 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data.

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,  $Y^{J/\psi\phi}$ , are extracted from maximum likelihood fits to the mass spectrum of candidates in each year of data taking. The selection applied to identify 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 weights applied to simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays are

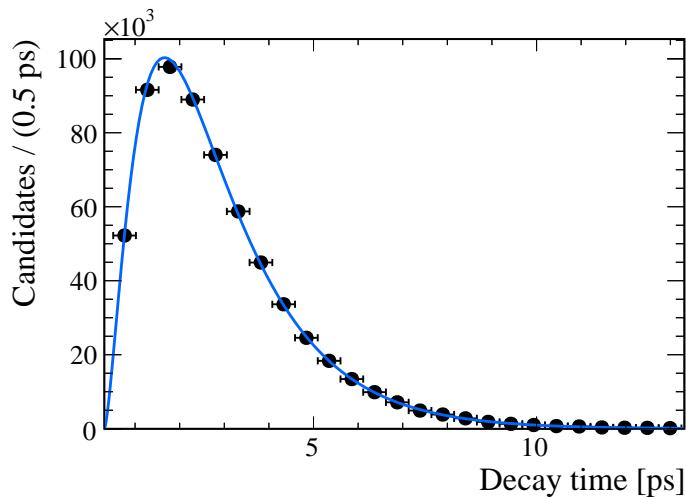
$$\omega_i = \frac{Y_i^{J/\psi\phi} \epsilon_i}{\sum_j Y_j^{J/\psi\phi} \epsilon_j} \cdot \frac{\sum_k N_k^{\mu^+\mu^-}}{N_i^{\mu^+\mu^-}}, \quad (6.5)$$

where  $i$  represents the year and  $N_i^{\mu^+\mu^-}$  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  $\epsilon_i$  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.

| Year ( $i$ ) | $Y_i$ | $\epsilon_i$ | $N_i$  | $\omega_i$ | $\mathcal{N}_i \equiv N_i \omega_i$ |
|--------------|-------|--------------|--------|------------|-------------------------------------|
| 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.

An unbinned maximum likelihood fit is performed to the combined 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 results are shown in Figure 6.7 and the acceptance parameters are given in Table 6.3. Figure 6.8



**Fig. 6.7** Maximum likelihood fit to the combined decay time distribution (left) of 2011, 2012, 2015 and 2016 simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays and the pull distribution of the fit.

| Parameter            | Value             |
|----------------------|-------------------|
| $a / \text{ps}^{-1}$ | $0.574 \pm 0.011$ |
| $n$                  | $1.49 \pm 0.03$   |
| $t_0 / \text{ps}$    | $0.313 \pm 0.007$ |

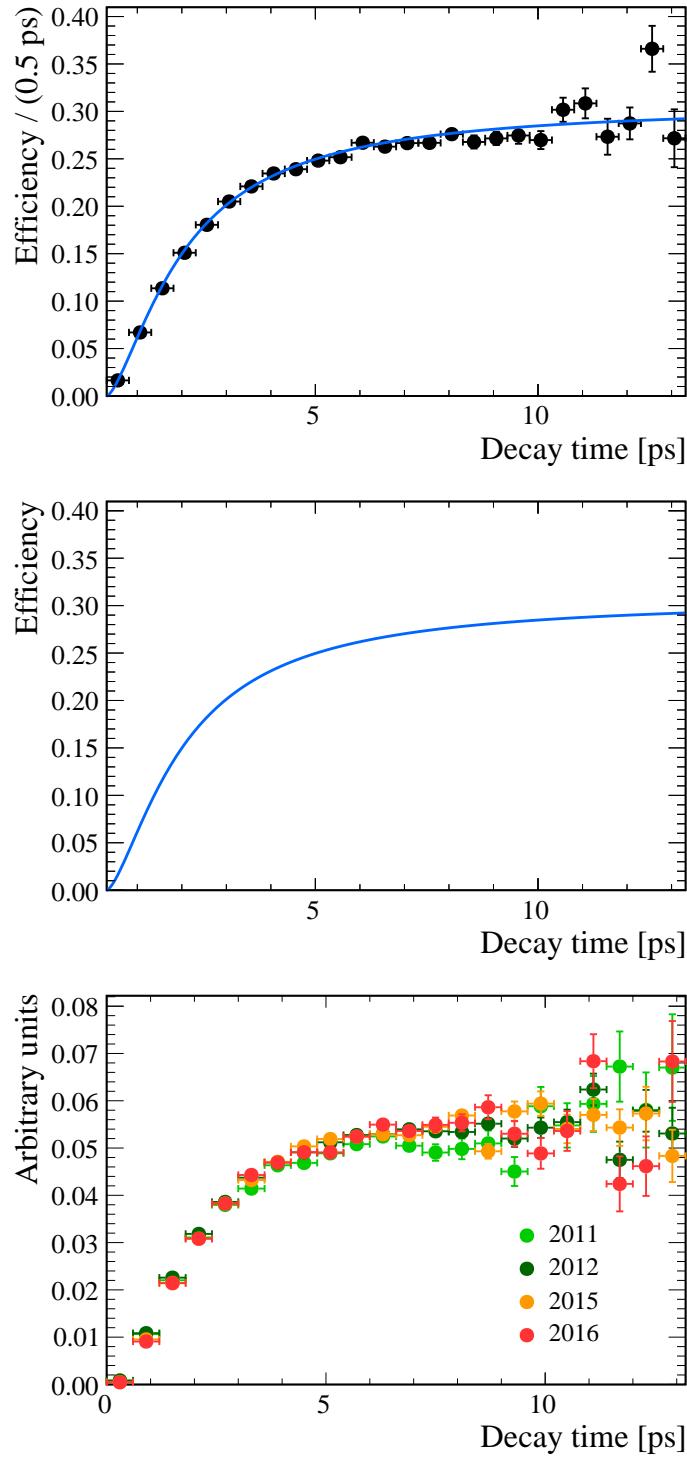
**Table 6.3** Parameters for the  $B_s^0 \rightarrow \mu^+\mu^-$  acceptance function determined from weighted simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays.

shows the selection efficiency histogram as a function of decay time with the acceptance function overlaid.

### 6.3.2 Background decay time PDFs

The final fit to measure the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime does not require knowledge of the decay time PDFs of the backgrounds. However, the fit configuration is developed using toy studies that use the mass and decay time distributions of both signal and background decays. Therefore realistic models of the decay time PDFs are needed to determine the optimal fit configuration.

The selection biases the decay time distributions of the backgrounds in the same way as the  $B_s^0 \rightarrow \mu^+\mu^-$  decay time. Therefore, they are described by the same PDFs as in Equation 6.3.



**Fig. 6.8** Selection efficiency histogram as a function of decay time with the acceptance PDF overlaid for weighted 2011, 2012, 2015 and 2016 simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays (left) and the efficiency histograms for each year separately for weighted simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays.

The backgrounds from mis-identified and  $B^0 \rightarrow \mu^+ \mu^-$  decays are assigned the same acceptance function as  $B_s^0 \rightarrow \mu^+ \mu^-$  decays, Equation 6.4, because the decay time efficiency of these backgrounds is approximately the same as the signal. The acceptances of these backgrounds does 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.

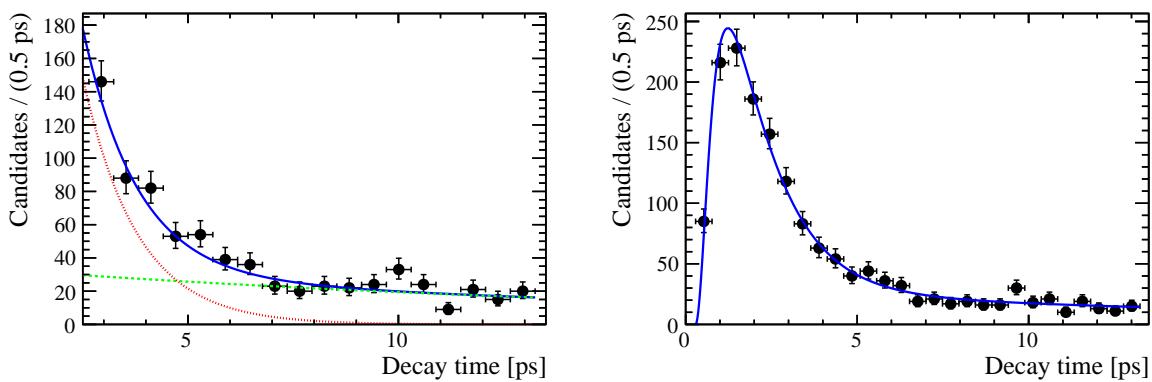
The decay time PDF of the combinatorial background is more challenging to determine. This background arises from random combinations of muons in the event and not from one source, therefore there is no single lifetime that describes the background. Furthermore, the global BDT which is designed to separate  $B_s^0 \rightarrow \mu^+ \mu^-$  decays from combinatorial background decays will have a different efficiency as a function of decay time for the combinatorial background compared to the signal. 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. Therefore, the decay time PDF of the combinatorial background for  $B_s^0 \rightarrow \mu^+ \mu^-$  decays 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 reconstructed mass greater than 5447 MeV/ $c^2$ , above the  $B_s^0$  signal region. The decay time PDF for combinatorial background decays is modelled by

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

where  $\tau_1$  and  $\tau_2$  are two independent lifetimes used to describe the background,  $f$  describes the fraction of candidates with lifetime  $\tau_1$  and the same acceptance shape as in Equation 6.4 is used for describe the decay time efficiency  $\epsilon(t)$ . The lifetimes are different, one describes a long-lived component and the other a short-lived component that are evident in the data. The decay time acceptance is flat at large decay times, therefore the lifetimes of the combinatorial background decays are determined from a maximum likelihood 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 maximum likelihood 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 6.9 and the PDF parameters in Table 6.4, the  $t_0$  parameter is fixed in the fit to improve fit stability.

| Parameter            | Value             |
|----------------------|-------------------|
| $a / \text{ps}^{-1}$ | $1.45 \pm 0.12$   |
| $n$                  | $1.92 \pm 0.17$   |
| $t_0 / \text{ps}$    | 0.290             |
| $\tau_1 / \text{ps}$ | $17 \pm 16$       |
| $\tau_2 / \text{ps}$ | $1.3 \pm 0.3$     |
| $f$                  | $0.032 \pm 0.027$ |

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



**Fig. 6.9** Maximum likelihood fit to determine the lifetimes of the background (top), with the long live component (bottom) and the short-lived component (red). The decay time distribution acceptance (right) of the combinatorial background decays in data passing the  $B \rightarrow h^+h'^-$  selection requirements.

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 in the event. 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 of the toy studies is investigated in Section 7.6.

## 6.4 Fit optimisation

The strategy to measure the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime was described in Section 6.1. However, given the extremely rare nature of  $B_s^0 \rightarrow \mu^+\mu^-$  decays, the stability and performance of the final fit will be highly dependant on the fit to the invariant mass distribution. Pseudoexperiments were 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 the data set passing the  $B_s^0 \rightarrow \mu^+\mu^-$  selection in the mass range 4600 - 6000 MeV/ $c^2$  were used as the basis for the toy studies. The expected background yields were 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 toy studies are performed by generating the mass and decay time distributions for the expected number of signal and background decays using the PDFs described in Section 6.2 and 6.3, assuming the Standard Model prediction for  $\tau_{\mu\mu}$ , and taking the slope of the combinatorial background mass PDF from simulated decays. sWeights are computed from an unbinned maximum likelihood fit to the invariant mass distribution. The lifetime and its inverse are measured by a unbinned maximum likelihood fit to the sWeighted decay time distribution. A series of different mass ranges and background components were tested. For each possible configuration a study containing 10,000 pseudoexperiments were performed and the performance of each configuration was evaluated using a couple of different metrics. The first, is the median expected uncertainty of the  $B_s^0 \rightarrow \mu^+\mu^-$  lifetime and inverse 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

| 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          |

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

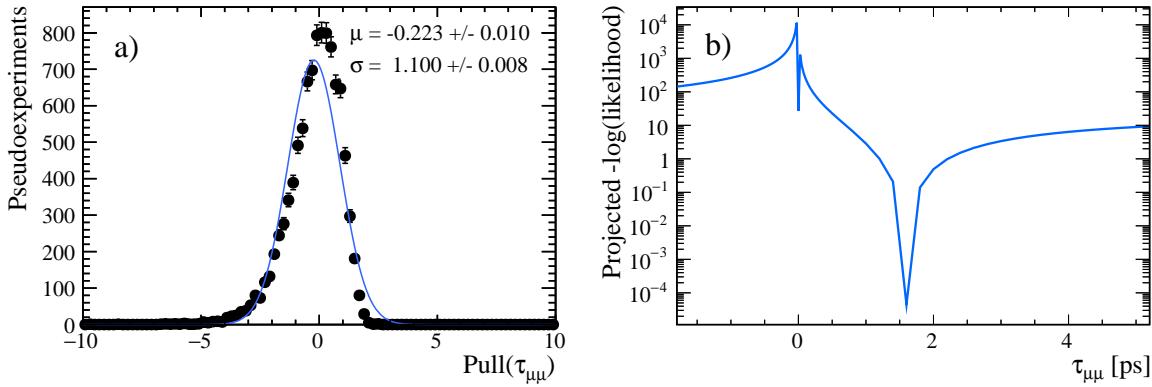
in the fit, where the pull is defined as  $(x - \mu)/\sigma$  with  $x$  the measured parameter value,  $\mu$  the value used in the generation and  $\sigma$  the uncertainty on the measured parameter value. Ideally the pull distributions will be Gaussian in shape with a mean at 0 and a width of 1.

The details of the toy studies performed 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 $\tau$ or $\tau^{-1}$ ?

During the development of the fit strategy, the toy studies 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 6.10 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 clearly not Gaussian in shape. 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 at a function of  $\tau_{\mu\mu}$  reveals the cause of the biased pull distribution. For the simplified studies illustrated in Figure 6.10 the decay time is



**Fig. 6.10** Pull distribution (left) 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  $4900 - 6000 \text{ MeV}/c^2$  with the expected statistics for  $4.4 \text{ fb}^{-1}$  of data. Likelihood profile for  $\tau_{\mu\mu}$  (right).

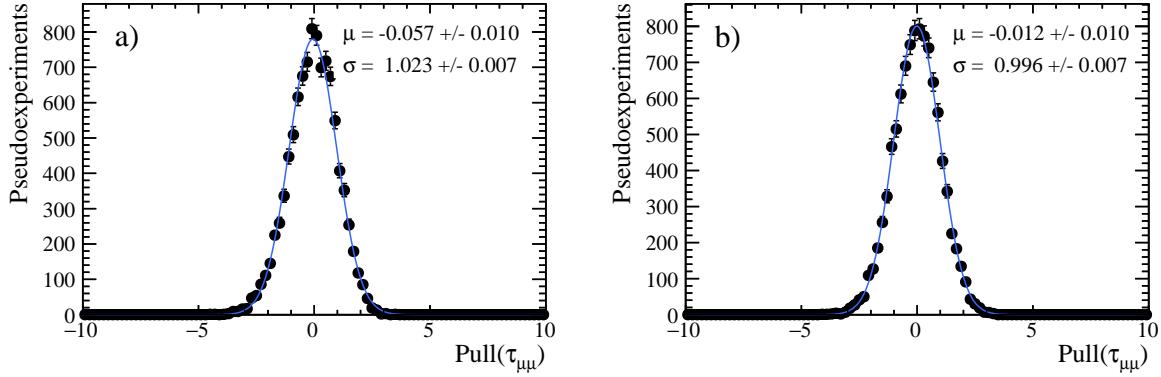
modelled by

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

The likelihood profile as a function of decay time for this model is shown in Figure ?? and there is a clear discontinuity at the zero. The discontinuity arises because the value of  $N(t, \tau)$  approaches zero as  $\tau$  reduces in value until at the origin when  $\tau = 0$  and  $N(t, \tau)$  is undefined. At the low statistics expected for the dataset, particularly when only Run 1 data was considered, the fitted value for  $\tau_{\mu\mu}$  is only a few standard deviations away from this discontinuity, therefore 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 are increased the  $\tau_{\mu\mu}$  pull distributions become Gaussian in shape as shown in Figure 6.11. This is as expected because when the statistical uncertainty decreases, the discontinuity of Figure 6.10 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 maximum likelihood 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 as the input value. Table 6.6 shows the coverage of the statistical uncertainties for  $\tau_{\mu\mu}$  for a set of 10,000 pseudoexperiments for the expected  $B_s^0 \rightarrow \mu^+ \mu^-$



**Fig. 6.11** Pull distribution for  $\tau_{\mu\mu}$  using simplified mass and decay time model for  $50 \text{ fb}^{-1}$  (left) and  $300 \text{ fb}^{-1}$  (right) where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range  $4900 - 6000 \text{ MeV}/c^2$ .

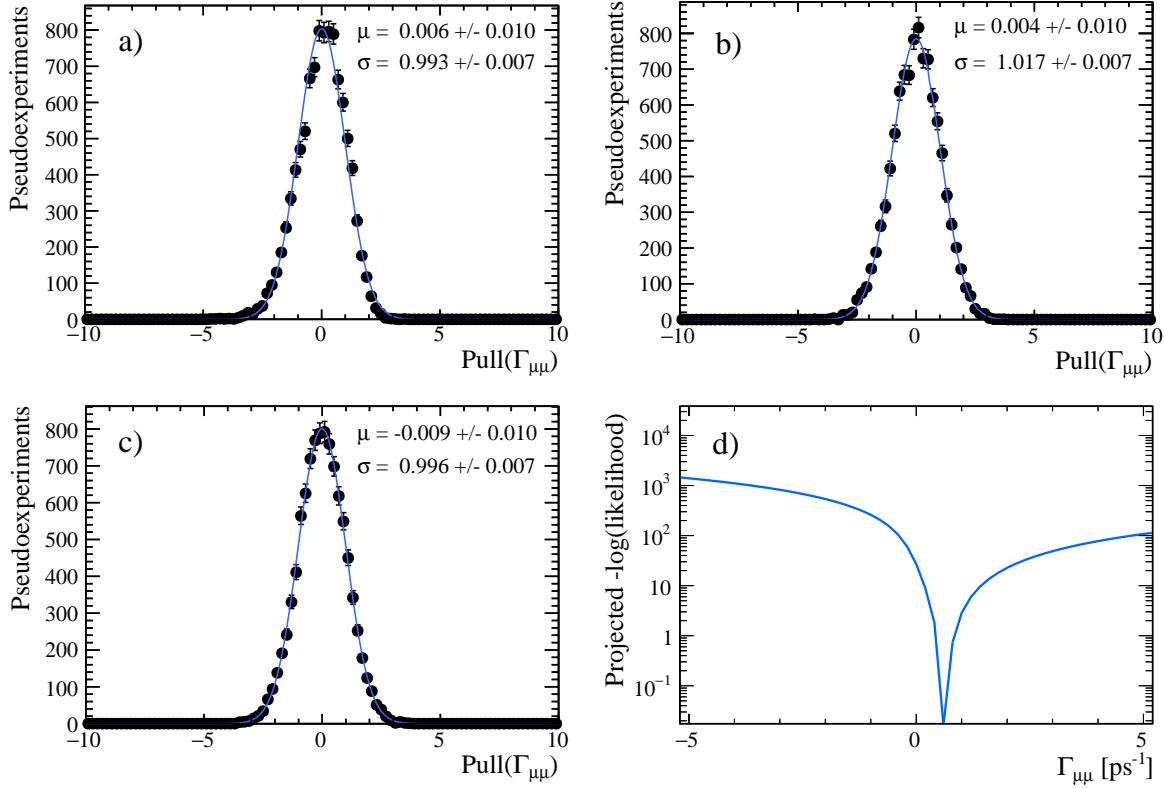
and combinatorial background yields with  $4.4 \text{ fb}^{-1}$  alongside the intervals expected for a Gaussian distribution. The simple toy configuration used to produce the log-likelihood function is used. A comparison between the coverage of  $\tau_{\mu\mu}$  and the Gaussian intervals shows that the coverage of the statistical uncertainties is very close to the expected values.

|           | $\tau_{\mu\mu}$    | $\Gamma_{\mu\mu} \equiv \frac{1}{t_{\mu\mu}}$ | Gaussian |
|-----------|--------------------|---|----------|
| $1\sigma$ | $68.50 \pm 0.08\%$ | $67.92 \pm 0.08\%$                            | 68.27%   |
| $2\sigma$ | $93.44 \pm 0.10\%$ | $95.91 \pm 0.10\%$                            | 95.45%   |
| $3\sigma$ | $98.06 \pm 0.10\%$ | $99.55 \pm 0.10\%$                            | 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.

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 6.12 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 6.12. Furthermore the statistical coverage of  $\Gamma_{\mu\mu}$  is closer to the expected Gaussian coverage than the coverage of  $\tau_{\mu\mu}$ .

Ideally, the fit strategy would be performed to extract the lifetime not the inverse lifetime, however for the moment the maximum likelihood fit for both  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  are used in the toy studies. The statistical coverage for both parameters is good and using



**Fig. 6.12** Pull distribution for  $\Gamma_{\mu\mu}$  using simplified toy studies for 4.4 (top left), 50 (top right) and 300 (bottom left)  $\text{fb}^{-1}$  and the likelihood profile as a function of  $\Gamma_{\mu\mu}$  (bottom right).

either is reasonable. The final decision was made based on the statistical coverage for the observed number of decays in the data set.

### 6.4.2 Optimisation results

The mass distribution of the expected number of  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates passing the effective lifetime selection is shown in Figure 6.13 alongside the corresponding decay time distribution for one pseudoexperiment for 4.4  $\text{fb}^{-1}$  of Run 1 and Run 2 data. 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  $B \rightarrow h^+ h'^-$ , 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  $\Gamma_{\mu\mu}$ .

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.

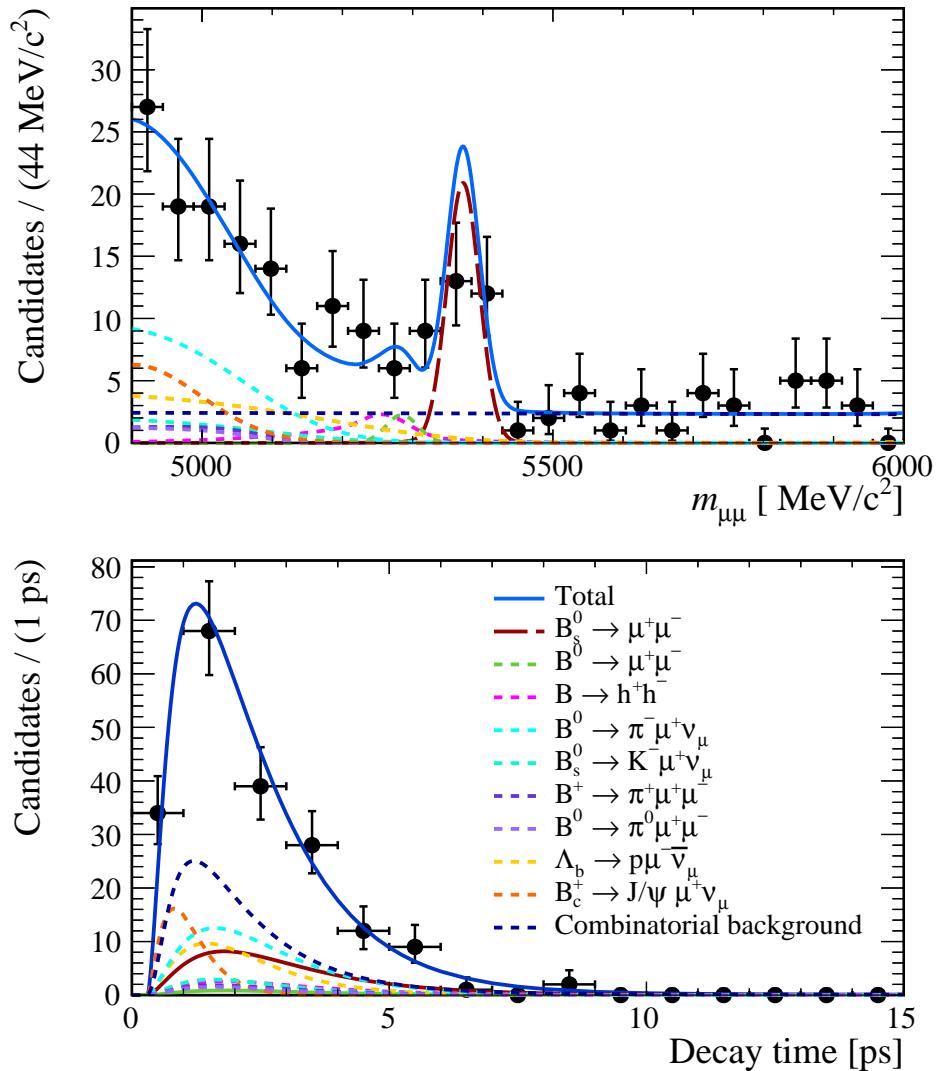
For each possible mass fit configuration the  $B_s^0 \rightarrow \mu^+\mu^-$ ,  $B^0 \rightarrow \mu^+\mu^-$  and combinatorial background yields are left free in the fit whereas the yields of any other backgrounds are constrained to their expected values. The mass shapes of all components are fixed 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,  $\tau_H$ , the average value calculated by the Particle Data Group is used to generated events for the pseudoexperiments [48]. 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  $\Gamma_{\mu\mu}$ , the  $B_s^0 \rightarrow \mu^+\mu^-$  yield and combinatorial background yield and slope as well as the median expected uncertainty 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.

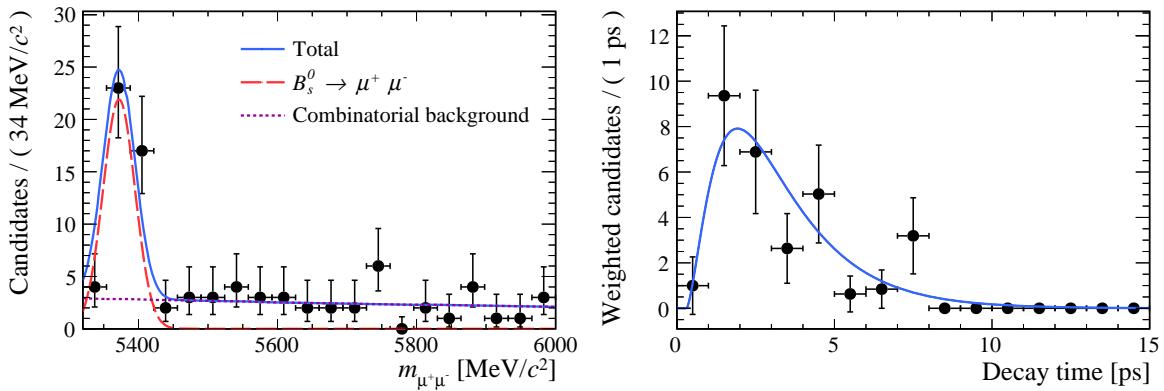
The expected statistical uncertainties for  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  are smallest for fit configuration 11, where the mass range is restricted to 5320 - 6000 MeV/ $c^2$  and only the  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background components are used in the total mass PDF. 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 expected uncertainties for  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  as well as clearly biased pull distributions.

Therefore, the fit configuration number 11 is chosen to measure the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime. Figure 6.14 gives an example of the mass and decay time maximum likelihood fits for the chosen configuration Figure 6.13 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 affect on the final result of not modelling these backgrounds is estimated in Section 7.2.

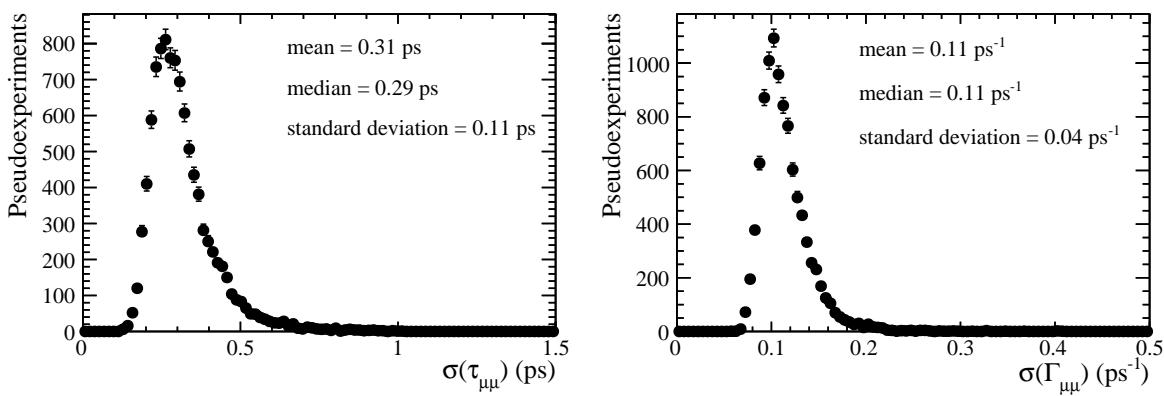
The expected uncertainties for the chosen fit configuration for  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  are  $\sigma(\tau_{\mu\mu}) = 0.28$  ps and  $\sigma(\Gamma_{\mu\mu}) = 0.11$  ps $^{-1}$ . However, due to the low expected number of decays there is a large spread in the expected uncertainties as shown in Figure 6.15. 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  $\Gamma_{\mu\mu}$ .



**Fig. 6.13** Mass and decay time distributions for the generated decays in the mass range 4900 - 6000 for on pseudoexperiment.



**Fig. 6.14** Example of the mass and decay time maximum likelihood fits for one pseudo-experiment 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. 6.15** Expected statistical uncertainties for  $\tau_{\mu\mu}$  (left) and  $\Gamma_{\mu\mu}$  (right) using fit configuration number 11.

| Fit no. | Mass Range  | Components included in the mass PDF<br>/MeV/ $c^2$  | Yields<br>free | Yields<br>fixed |
|---------|-------------|---|----------------|-----------------|
|         |             | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
| 1.      | 4900 - 6000 | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ , $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ ,<br>$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ , $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ ,<br>$B^+ \rightarrow \pi^+ \mu^+ \mu^-$ , $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ |                | ✓               |
| 2.      | 4900 - 6000 | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
|         |             | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ , $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ ,<br>$B_{(s)}^0 \rightarrow \pi^- (K^-) \mu^+ \nu_\mu$ , $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$  |                | ✓               |
| 3.      | 5150 - 6000 | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
|         |             | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ , $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ ,<br>$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ , $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ ,<br>$B^+ \rightarrow \pi^+ \mu^+ \mu^-$ , $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ |                | ✓               |
| 4.      | 5150 - 6000 | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
|         |             | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ , $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ ,<br>$B_{(s)}^0 \rightarrow \pi^- (K^-) \mu^+ \nu_\mu$ , $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$  |                | ✓               |
| 5.      | 5200 - 6000 | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
|         |             | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ , $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$   |                | ✓               |
| 6.      | 5200 - 6000 | $B_s^0 \rightarrow \mu^+ \mu^-$ , $B^0 \rightarrow \mu^+ \mu^-$ , comb. bkg.  | ✓              |                 |
|         |             | $B \rightarrow h^+ h'^-$ , $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$  |                | ✓               |
| 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 | $\sigma/\text{ps}$ | Mean   | $\sigma/\text{ps}$ | Width           | $\sigma/\text{ps}^{-1}$ | $\sigma/\text{ps}$ | Width             | $\sigma/\text{ps}^{-1}$ |  |
| 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.

## 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  $4.4 \text{ fb}^{-1}$  of Run 1 and Run 2 data are shown in Figure 6.16. The number of observed decays was  $22 \pm 6$   $B_s^0 \rightarrow \mu^+ \mu^-$  decays and  $20 \pm 6$  combinatorial background decays. The measured values of  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  are

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

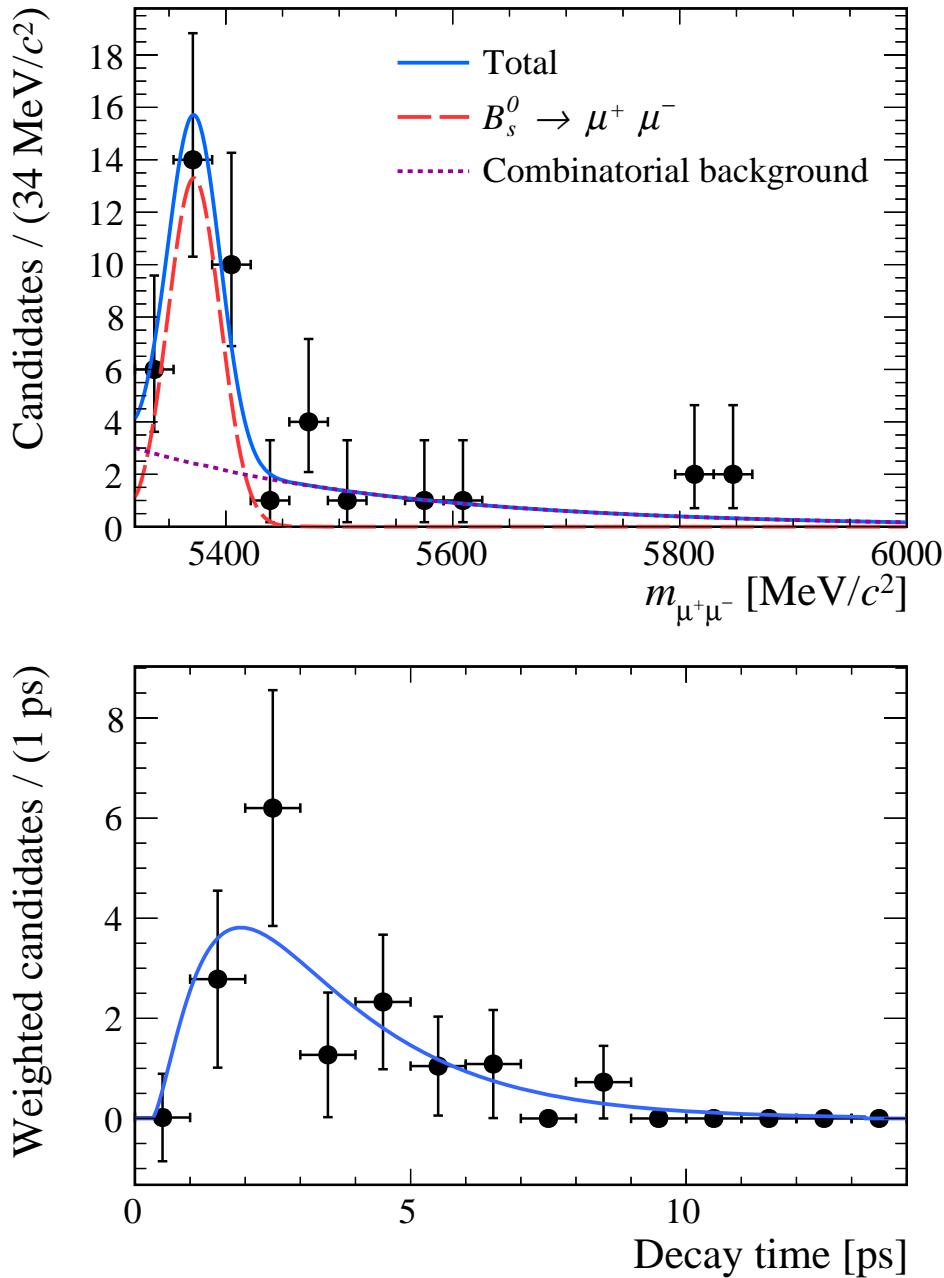
$$\Gamma_{\mu\mu} = 0.489 \pm 0.117 \text{ ps}^{-1} \quad (6.9)$$

where the uncertainties are only statistical. The results are consistent with the Standard Model prediction of  $\tau_{\mu\mu} = \tau_H$  within  $1\sigma$  and within  $1.5\sigma$  of  $\tau_{\mu\mu} = \tau_L$ .

The observed number of signal and background decays is lower than expected yields given in Table 6.5. Therefore, it is important to check whether the statistical coverage for  $\tau_{\mu\mu}$  uncertainties is still reasonable. The pseudoexperiments in Section 6.4.1 were 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.

|           | $\tau_{\mu\mu}$    | $\Gamma_{\mu\mu}$  | Gaussian |
|-----------|--------------------|--------------------|----------|
| $1\sigma$ | $68.83 \pm 0.08\%$ | $67.76 \pm 0.08\%$ | 68.27%   |
| $2\sigma$ | $93.11 \pm 0.10\%$ | $95.55 \pm 0.10\%$ | 95.45%   |
| $3\sigma$ | $97.92 \pm 0.10\%$ | $99.67 \pm 0.10\%$ | 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. 6.16** Maximum likelihood fit to the invariant mass distribution (top) and weighted decay time distribution (bottom) of  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates using an integrated luminosity of 4.4 fb<sup>-1</sup> of data collected by the LHCb experiment.  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates are described by the red peak in the mass plot and combinatorial background by the blue dashed line, the total PDF is given by the solid blue line.

# Chapter 7

## Systematic uncertainties and cross checks

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 different systematic biases 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 for measuring  $\tau_{\mu\mu}$  is given at the end of the chapter.

### 7.1 Fit Accuracy

The fit strategy used to measure the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime was presented in Chapter 6. The final fit configuration 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  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$ . The values for the figures of merit for a set of 10,000 pseudoexperiments using the final fit configuration and assuming the Standard Model  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions and effective lifetime 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. 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  $\tau_{\mu\mu}$ . The pull distributions for  $\tau_{\mu\mu}$  are biased, as discussed in Section 6.4.1, and are therefore not used to evaluate the fit performance and the pull distribution for  $\Gamma_{\mu\mu}$  can be used as a measure of the fit performance instead.

Furthermore, the uncertainty on  $\Gamma_{\mu\mu}$  is not longer needed to evaluate the fit performance since the final measured results are quoted in terms of  $\tau_{\mu\mu}$ .

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

The pull distribution for  $\Gamma_{\mu\mu}$  and the coverage of the statical uncertainties given in Section 6.4 show that the fit gives a good estimated 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 were performed for a range of generated  $B_s^0 \rightarrow \mu^+\mu^-$  lifetimes different to the Standard Model prediction. Only  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background decays were generated in the pseudoexperiments 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 all return accurate pull distributions for the fitted  $\Gamma_{\mu\mu}$  values with means and widths consistent with 0 and 1, respectively, and the expected uncertainties for  $\tau_{\mu\mu}$  are similar. Therefore, the fit returns an accurate measured value for a range of  $B_s^0 \rightarrow \mu^+\mu^-$  lifetimes independent of the lifetime chosen for the  $B_s^0$ .

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

**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 Standard Model predication ( $\tau_H$ ); the lifetime of the light  $B_s^0$  mass eigenstate ( $\tau_L$ ); and the mean lifetime of the  $B_s^0$  ( $\tau_{B_s^0}$ ). 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 median statistical uncertainty on  $\tau_{\mu\mu}$ . The uncertainties on the means are 0.010 and widths are 0.007 for both configurations.

### 7.1.2 $B_s^0 \rightarrow \mu^+ \mu^-$ yields

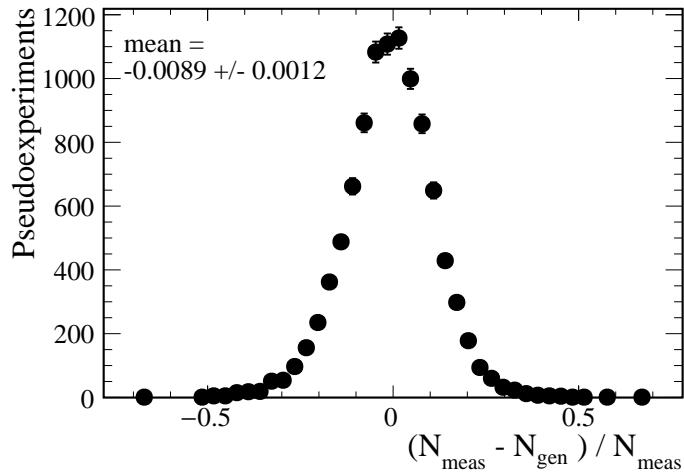
The pull distributions for  $B_s^0 \rightarrow \mu^+ \mu^-$  yields have slightly biased mean values of  $\sim 0.010$  ps as shown in Tables 7.1, implying that the mass fit does not accurately estimate the  $B_s^0 \rightarrow \mu^+ \mu^-$  yield. However, the pull distribution for  $\Gamma_{\mu\mu}$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

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 an accurate 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. Therefore, the uncertainty on the measured yields is proportional to  $\sqrt{N_{gen}}$  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 %.

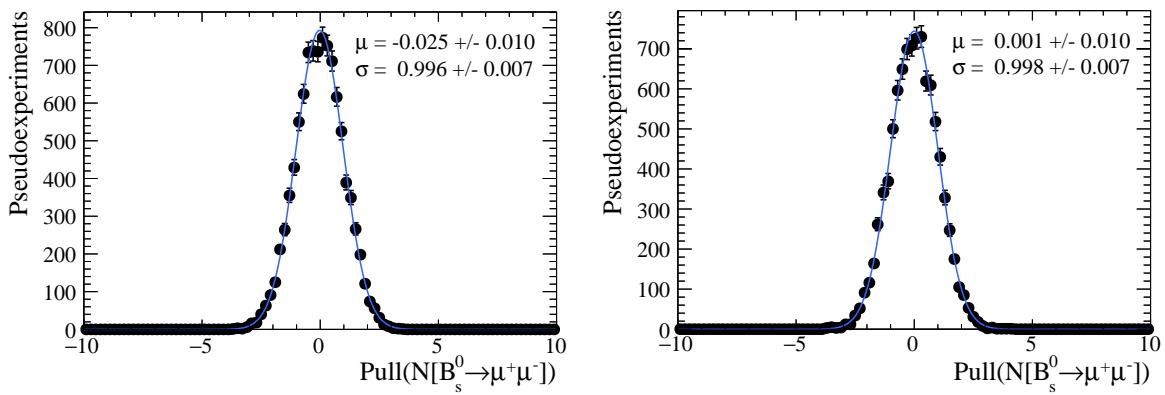
Furthermore, the pull distributions for  $B_s^0 \rightarrow \mu^+ \mu^-$  yields for pseudoexperiments with higher statistics produce means that tend towards 0 as the number of decays increases as shown in Figure 7.3 for the expected number of decays with 50 and 300 fb $^{-1}$ . Therefore the mass fit 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 bias mean value of  $\sim 0.006$  compared to the  $B_s^0 \rightarrow \mu^+ \mu^-$  yields.

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

The remaining area of the fit to investigate is any underlying bias in the fit on the measured value of  $\tau_{\mu\mu}$ . As discussed in Section 6.4.1 the pull distribution for the measured effective lifetime is biased for the expected number of statistics but the pull distribution for  $\Gamma_{\mu\mu}$  produces a mean and width consistent with 0 and 1, respectively. However, the coverage



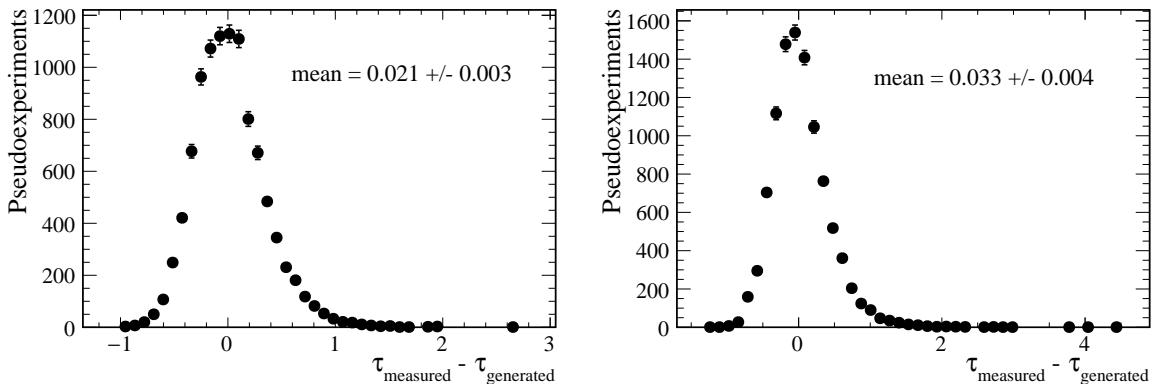
**Fig. 7.1** Fractional bias of the measured  $B_s^0 \rightarrow \mu^+\mu^-$  yield from pseudoexperiments for the expected number of decays with  $4.4 \text{ fb}^{-1}$ . Only  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background decays are included in the pseudoexperiments.



**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  $50$  and  $300 \text{ fb}^{-1}$ . Only  $B_s^0 \rightarrow \mu^+\mu^-$  and combinatorial background decays are included in the pseudoexperiments.

of the uncertainties of both  $\tau_{\mu\mu}$  and  $\Gamma_{\mu\mu}$  are reasonable and the biased  $\tau_{\mu\mu}$  pull arises from the likelihood function as discussed in Section 6.4.1.

The overall bias in the fit for measuring  $\tau_{\mu\mu}$  is evaluated from the difference between the measured and generated values 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 with the expected and also the observed number of  $B_s^0 \rightarrow \mu^+\mu^-$  decays. The fit bias is evaluated for the observed number of decays because there are fewer than expected. 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 cause by the fit on the final result for  $\tau_{\mu\mu}$ .



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

## 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 and candidates with a  $B_s^0$  mass between 5320 - 6000 MeV/ $c^2$ . Although the majority of background decays from mis-identified decays and  $B^0 \rightarrow \mu^+\mu^-$  decays fall outside this mass window, as shown in Figure 6.13, the tails of some backgrounds still need to be accounted for. The backgrounds of particular importance for the chosen mass range include  $B^0 \rightarrow \mu^+\mu^-$ ,  $B \rightarrow h^+h'^-$ ,  $\Lambda_b^0 \rightarrow p\mu^-\nu_\mu$ ,  $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ ,  $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ ,  $B^+ \rightarrow \pi^+\mu^+\mu^-$ ,

$B^0 \rightarrow \pi^0 \mu^+ \mu^-$ ,  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ , with  $B \rightarrow h^+ h'^-$ ,  $B^0 \rightarrow \mu^+ \mu^-$  and  $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$  being of particular importance for the chosen mass range.

The number of expected background decays and their mass PDFs in the range 4900 - 6000 MeV/ $c^2$  were 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 computed by integrating the mass PDFs. 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/ $c^2$ .

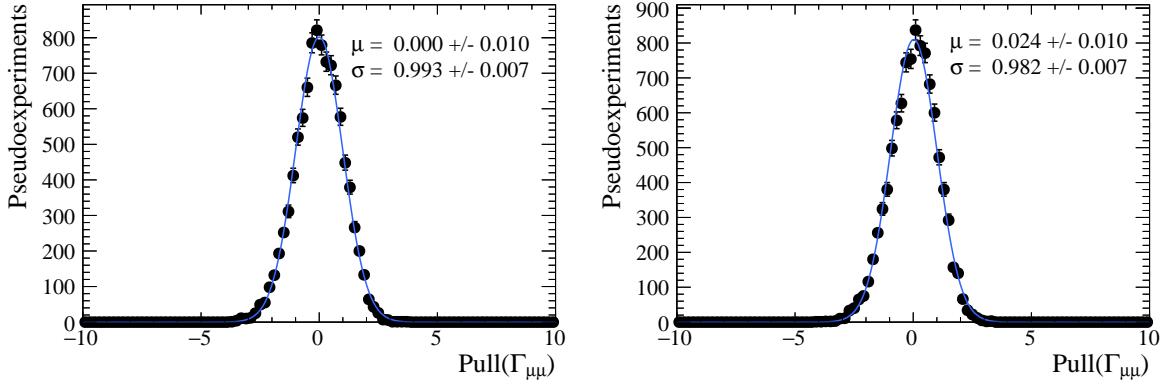
| Decay                                     | Expected yields in mass ranges |                        |
|---|--------------------------------|------------------------|
|   | 4900 - 6000 MeV/ $c^2$         | 5320 - 6000 MeV/ $c^2$ |
| $B_s^0 \rightarrow \mu^+ \mu^-$           | 30.9                           | 30.5                   |
| $B^0 \rightarrow \mu^+ \mu^-$             | 3.3                            | 0.2                    |
| $B \rightarrow h^+ h'^-$                  | 9.7                            | 0.9                    |
| $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$ | 13.3                           | 0.6                    |
| $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$     | 40.5                           | 0.1                    |
| $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$     | 9.1                            | 0.0                    |
| $B^+ \rightarrow \pi^+ \mu^+ \mu^-$       | 6.0                            | 0.0                    |
| $B^0 \rightarrow \pi^0 \mu^+ \mu^-$       | 4.9                            | 0.6                    |
| $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  | 9.8                            | 0.0                    |
| 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/ $c^2$ .

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 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 pseudoexperiment set up is 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.



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

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, excluding the results presented in Chapter 5. Using the measured values from the combined analysis of the Run 1 CMS and LHCb data [40], the expected number of  $B_s^0 \rightarrow \mu^+\mu^-$  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 world average branching fractions but the shift in the mean of the pull distribution was smaller, therefore the larger value from the SM predictions are used.

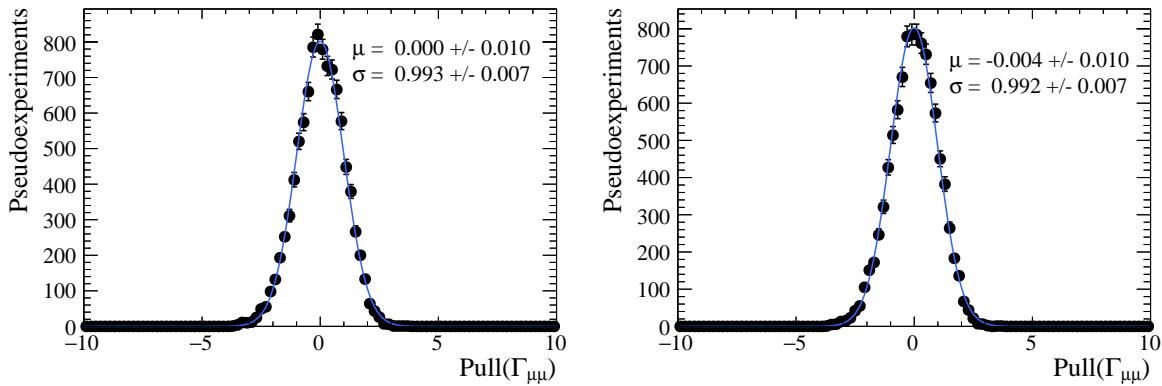
| Decay                          | Expected yield in 5320 - 6000 MeV/ $c^2$ |               |
|--------------------------------|--|---------------|
|                                | Standard Model                           | World average |
| $B_s^0 \rightarrow \mu^+\mu^-$ | 30.5                                     | 22.5          |
| $B^0 \rightarrow \mu^+\mu^-$   | 0.2                                      | 0.7           |

**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 Standard Model predictions and branching fraction measurements from the combined analysis of the Run 1 CMS and LHCb data [40].

### 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 in the mass window.  $B_s^0 \rightarrow \mu^+ \mu^-$  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 are 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 using Run 1 parameters (left) and Run 2 parameters (right).

### 7.4 Acceptance function accuracy

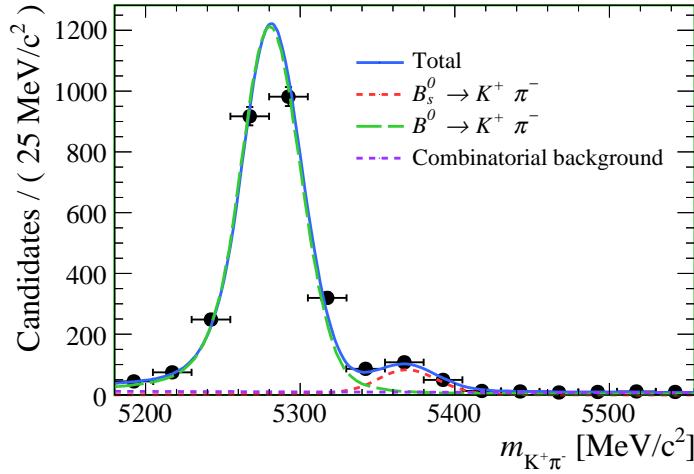
The decay time acceptance function is determined from weighted simulated decays, as described in Section 6.3.1. It relies on the assumption that weighted simulated

decays model the data reasonably well. 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 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 trigger lines that identify  $B_s^0 \rightarrow \mu^+\mu^-$  decays in data are relatively unbiased with respect to the  $B_s^0$  decay time distribution. The same is true for trigger lines that identify  $B \rightarrow h^+h'^-$  decays as TIS, whereas TOS triggers that identify  $B \rightarrow h^+h'^-$  decays create a large bias of the decay time distribution due to the dependence on the trigger lines on  $B_s^0$  IP, IP  $\chi^2$  and flight distance. The DLL <sub>$K\pi$</sub>  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 <sub>$K\pi$</sub>  value assigned the kaon mass hypothesis and the daughter particle with the lowest DLL <sub>$K\pi$</sub>  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 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 are 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 dependant 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



**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. Components for  $B^0 \rightarrow K^+\pi^-$ ,  $B_s^0 \rightarrow K^+\pi^-$  and combinatorial background decays are included.

each year. The weights used to combine simulated decays in different years are

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

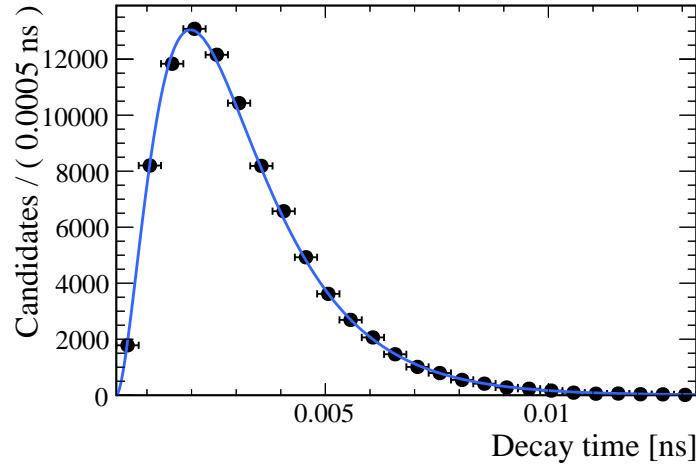
where  $Y_i^{data}$  are the  $B^0 \rightarrow K^+\pi^-$  yields in each year of data and  $N_i^{K\pi}$  the number of simulated decays available for each year. The acceptance function fit is shown in Figure 7.7 and for consistency the same simulation versions are used for  $B^0 \rightarrow K^+\pi^-$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

The measured  $B^0 \rightarrow K^+\pi^-$  lifetime is

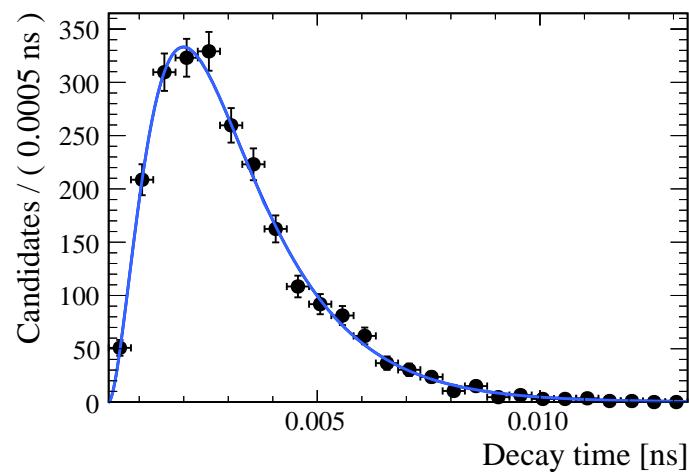
$$\tau_{K\pi} = 1.52 \pm 0.03 \text{ ps}, \quad (7.2)$$

where only the statistical uncertainty is given and the decay time fit is shown in Figure 7.8. The measured results are consistent with the PDF value of  $1.520 \pm 0.004$  ps [48] 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  $\tau_{\mu\mu}$ .

However, the determination of the  $B_s^0 \rightarrow \mu^+\mu^-$  acceptance function relies on weights taken from the number of tracks in an event for  $B^0 \rightarrow K^+\pi^-$  decays in data and



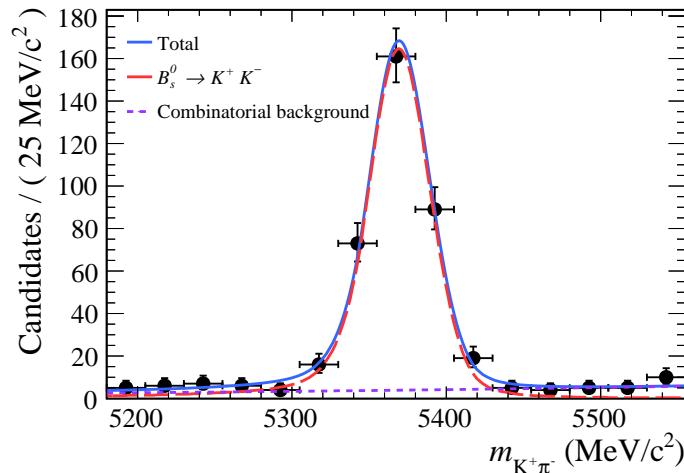
**Fig. 7.7** Decay time distribution in weighted 2011, 2012, 2015 and 2016 simulated decays and the maximum likelihood fit results to determine the acceptance function parameters.



**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.

simulation. Although the measurement of the  $B^0 \rightarrow K^+\pi^-$  lifetime shows the procedure and weighting method works for these decays, it does not show that the weights taken from  $B^0 \rightarrow K^+\pi^-$  decays can be applied to other decays. Therefore, as a cross check, the lifetime of  $B_s^0 \rightarrow K^+K^-$  decays is also measured.

The same measurement strategy is used for  $B_s^0 \rightarrow K^+K^-$  decays as  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow K^+\pi^-$  and candidates in 2012 and 2015 data are identified using the selection requirements in Chapter 4. Only 2012 and 2015 data are used due to the available simulation versions of simulated  $B_s^0 \rightarrow K^+K^-$  decays compared to  $B_s^0 \rightarrow \mu^+\mu^-$ . 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 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.

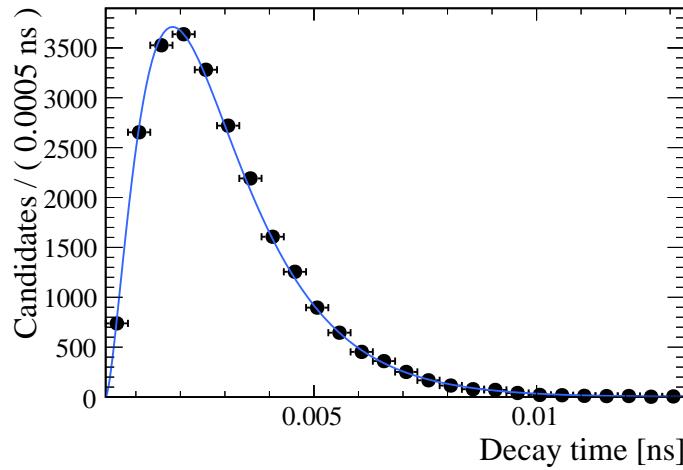


**Fig. 7.9** Maximum likelihood fit to the mass distribution of  $B_s^0 \rightarrow K^+K^-$  decays for data taken in 2012 and 2015. Components for  $B_s^0 \rightarrow K^+K^-$  and combinatorial background decays are included in the mass fit.

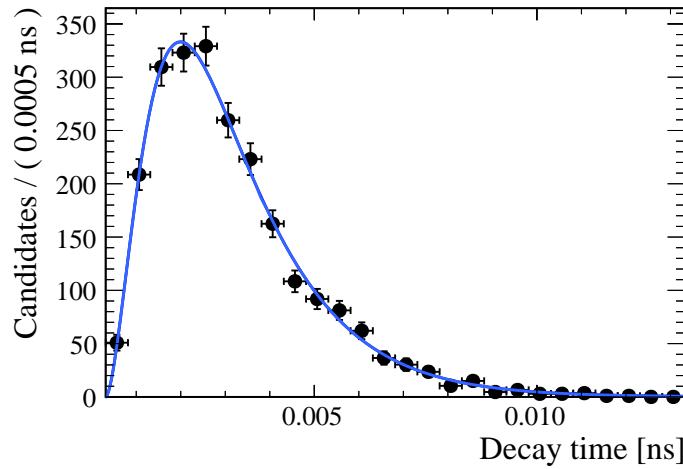
The  $B_s^0 \rightarrow K^+K^-$  acceptance is found using the same method as  $B_s^0 \rightarrow \mu^+\mu^-$  with  $B_s^0 \rightarrow J/\psi\phi$  decays used to determine the relative proportions of decays in each year of data. Figure 7.10 shows the acceptance fit and the results of the decay time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

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

where only the statistical uncertainty is given. The measured values are consistent with the predicted value of  $1.395 \pm 0.020$  ps [137] and shows that  $B^0 \rightarrow K^+ \pi^-$  weights can be used for other decays as well as  $B^0 \rightarrow K^+ \pi^-$ .



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



**Fig. 7.11** Maximum likelihood fit to the signal weighted decay time distribution of  $B_s^0 \rightarrow K^+ K^-$  decays for data taken in 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 the references [138, 139] that study the lifetimes of  $B \rightarrow J/\psi X$  decays at LHCb, a component is included into the decay time fit to model the number of incorrectly assigned primary vertices (PVs) as well as the resolution of the detector. The decay time fit consists of a PDF describing the decay time distribution 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 the references [138, 139] 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  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime is equal to the lifetime of the heavy  $B_s^0$  mass eigenstate of  $\tau_H = 1.610$  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.

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 with which decays were generated is computed for each decay that passes the full selection. The resulting distribution is fitted with a resolution function composed of the sum of three Gaussian functions. Each Gaussian function has the same mean value, which is left free in the fit, but different widths that are also free in the final fit. The fit parameters are shown in Table 7.4 and the results in Figure ???. The resulting distribution has a similar form to those used in references [138, 139], where the detector resolution is modelled with two

narrow Gaussian functions and the Gaussian function for incorrectly assigned PVs is broader and describes a small fraction of decays.

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

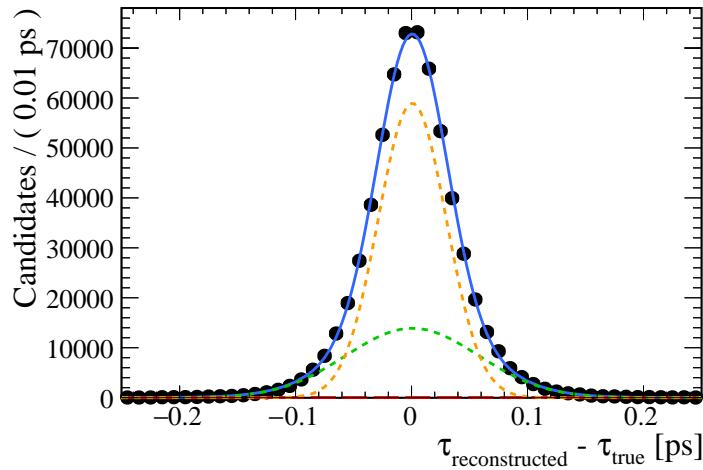
**Table 7.4** Parameters from the fit to the difference between the reconstructed decay time and the true decay time for simulated decays that pass the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime selection. The mean used for all Gaussian is  $\mu$  and  $\sigma_i$  are the widths of each Gaussian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

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 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.

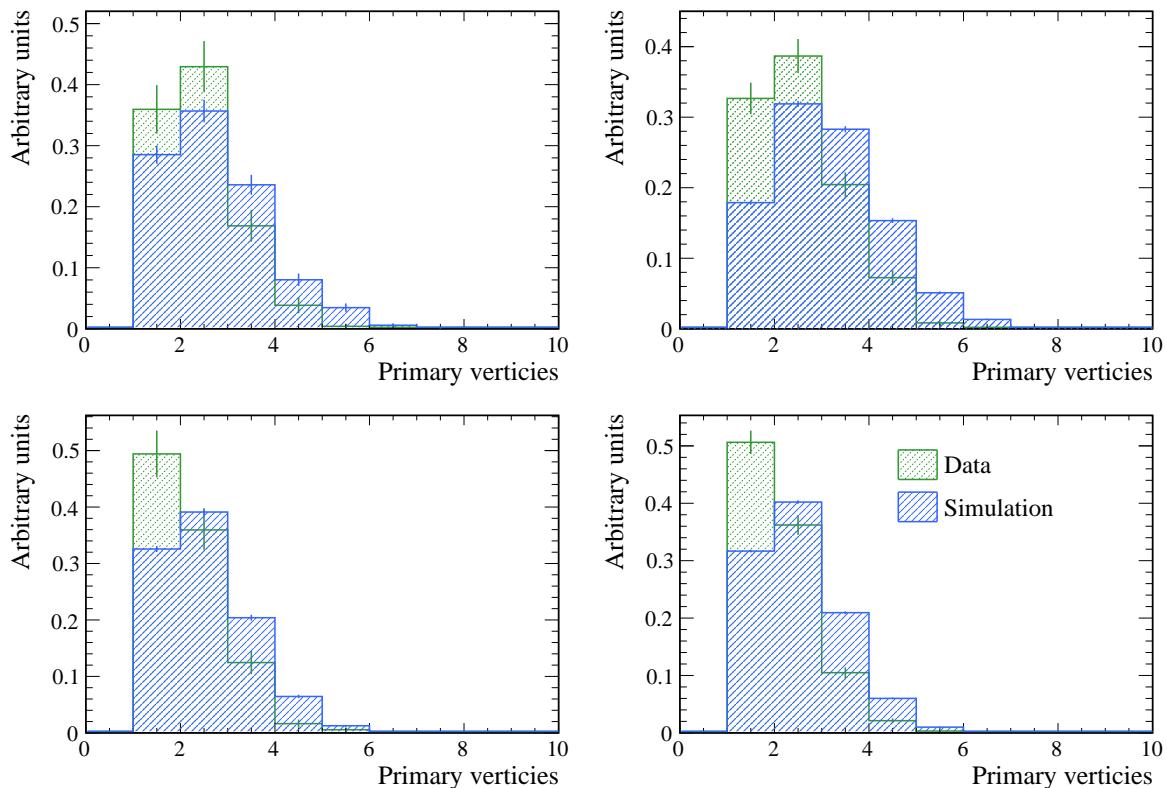
However, this check assumed that simulated decays provide a good estimate of the number of incorrectly assigned PVs. Figure 7.13 shows the number of  $B^0 \rightarrow K^+ \pi^-$  decays passing the selection for simulated  $B^0 \rightarrow K^+ \pi^-$  decays and sWeighted decays data for each year. On 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



**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.



**Fig. 7.13** The distributions for the number of primary vertices in an event  $B^0 \rightarrow K^+ \pi^-$  data and simulated decays for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right).

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 the background decay time components that are significantly longer lived than the signal lifetimes and 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 uncertainties in the 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:  $\tau_1$  and  $\tau_2$  are both increased by  $1\sigma$ ; and the fraction of decays with lifetime  $\tau_1$  is increased by  $1\sigma$ . 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<sup>-1</sup> 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

|         |                      | $B_s^0 \rightarrow \mu^+\mu^-$ |                      | $B \rightarrow h^+h'^-$ |  |
|---------|----------------------|--------------------------------|----------------------|-------------------------|--|
| BDT bin | mean decay time / ps | Number of candidates           | mean decay time / ps | Number of candidates    |  |
| 1       | $1.178 \pm 0.005$    | 50,695                         | $1.124 \pm 0.001$    | 964,502                 |  |
| 2       | $1.936 \pm 0.098$    | 244                            | $2.394 \pm 0.022$    | 8,838                   |  |
| 3       | $2.570 \pm 0.327$    | 46                             | $2.781 \pm 0.051$    | 2,373                   |  |
| 4       | $2.210 \pm 0.361$    | 17                             | $3.023 \pm 0.076$    | 1,125                   |  |
| 5       | $2.582 \pm 1.103$    | 4                              | $3.417 \pm 0.112$    | 655                     |  |
| 6       | $2.540 \pm 0.390$    | 3                              | $3.978 \pm 0.187$    | 313                     |  |
| 7       | $2.868 \pm 1.048$    | 2                              | $4.626 \pm 0.363$    | 109                     |  |
| 8       | -                    | 0                              | $5.706 \pm 0.683$    | 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.

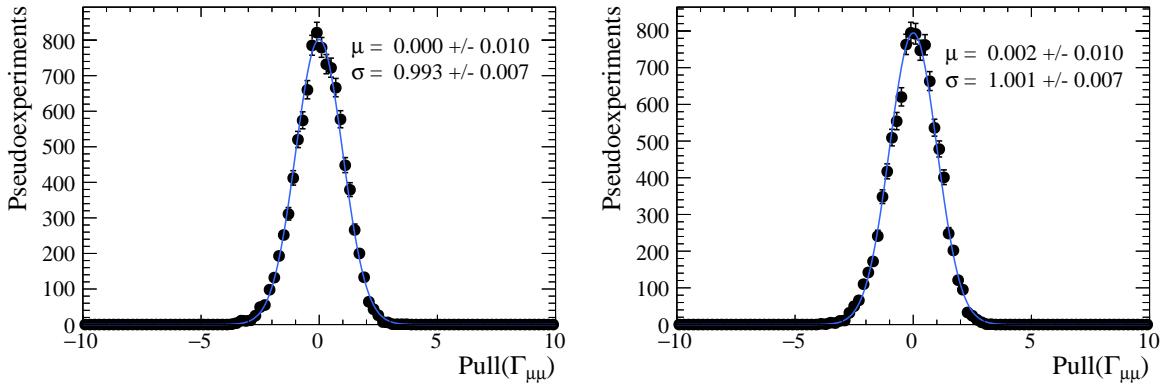
$\tau_{\mu\mu}$  a systematic uncertainty of 0.002 ps is assigned due to the combinatorial background decay time model.

## 7.7 Mix of $B_s^0$ mass eigenstates

In the Standard Model 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.

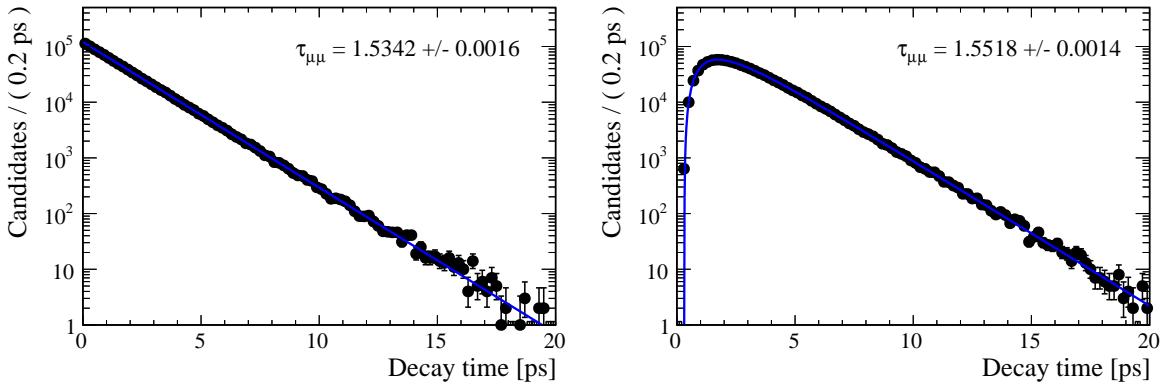
The size of this effect has been estimated using two simple pseudoexperiments. The first assumes that the selection has no bias on the decay time distribution and 1 million candidates are generated with equal contributions from the heavy and light  $B_s^0$  mass eigenstates. A second set of 1 million candidates are generated with the same mix of eigenstates but with a more realistic model including the  $B_s^0 \rightarrow \mu^+\mu^-$  acceptance function.



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

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.



**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 flat acceptance function (left) and the acceptance function used to describe  $B_s^0 \rightarrow \mu^+\mu^-$  decays in data (right).

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

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 ??, 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  $p\bar{p}$  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 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  $A_p = (1.09 \pm 2.61 \pm 0.66)\%$  [140]. The presence of the production asymmetry modifies the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay rate to

$$\begin{aligned} \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) &= \left( \frac{1+A_p}{2} \right) \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \\ &\quad + \left( \frac{1-A_p}{2} \right) \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-). \end{aligned} \quad (7.8)$$

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 2.32 and 2.33.

The total decay rate with the production asymmetry is

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

where  $C_\lambda = (1 - |\lambda_{\mu\mu}|^2)/(1 + |\lambda_{\mu\mu}|^2)$  and  $S_\lambda = 2\text{Im}/(1 + |\lambda_{\mu\mu}|^2)$  are  $\mathcal{CP}$  asymmetries and are related to  $A_{\Delta\Gamma}$  by  $|A_{\Delta\Gamma}|^2 + |C_\lambda|^2 + |S_\lambda|^2 = 1$  [42].

The production asymmetry introduces an additional oscillatory term 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  $\mathcal{N}' \equiv \frac{1}{4}\mathcal{N}|A_{\mu\mu}|^2(1 + |\lambda_{\mu\mu}|^2)$ . This decay rate can be 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 affect of the production asymmetry on the effective lifetime can now be calculated using the PDG 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  $\Gamma_H = 0.621 \text{ ps}^{-1}$  [48]. A value of  $A_p = 0.40$  is used, which is 1 standard deviation greater than the value measured by LHCb, and  $A_{\Delta\Gamma}^{\mu^+\mu^-} = 0.0$  is chosen. Since  $(A_{\Delta\Gamma}^{\mu^+\mu^-})^2 + (C_\lambda)^2 + (S_\lambda)^2 = 1$ , 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 leads to a total uncertainty of 0.05 ps for  $\tau_{\mu\mu}$ , which corresponds to 11 % of the observed statistical uncertainty. The small size of the total systematic uncertainties compared to the statistical uncertainty is expected given the observed number of decays.

| 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.

# Chapter 8

## Summary and Outlook

### 8.1 Summary

The LHCb experiment was built to test the predictions of the Standard Model and search for new physics 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, show no significant deviations from predictions and confirm the predictive power of the SM. 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 [141] as an indirect search for new physics. In 2011 LHCb joined 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  $2.1 \text{ fb}^{-1}$  of Run 1 data from  $pp$  collisions [38]. A combined analysis of the Run 1 datasets 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 [40]. The measured branching fractions of these decays are consistent with the SM predictions and place strong constraints on new physics models, although the precision of the measurements still leaves room for new physics effects to be revealed. With the observation of the  $B_s^0$  mode the search for  $B_s^0 \rightarrow \mu^+ \mu^-$  decays is complete and properties of these 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 fraction and the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime with  $4.4 \text{ fb}^{-1}$  of Run 1 and Run 2 data collected by the LHCb experiment

are presented in this dissertation. The measured branching fractions are

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

The  $B_s^0$  mode is observed with a statistical significance of  $7.8\sigma$ , making this result the first single experiment observation of this decay. The  $B^0$  mode is observed with a significance of  $1.6\sigma$ , 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 concisest with the SM predictions.

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

$$\tau_{\mu\mu} = 2.04 \pm 0.44 \pm 0.05 \text{ ps},\tag{8.2}$$

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

## 8.2 Outlook

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

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 flavour structure of the SM and new physics models. The current precision of the ratio of branching fractions is X%, the future runs of the LHC will enable the precision of the ratio of branching fractions to be reduced to 40% with  $50 \text{ fb}^{-1}$  of  $pp$  data and 20% with  $300 \text{ fb}^{-1}$  of data [142].

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 has been estimated using pseudoexperiments based on the observed numbers of decays with  $4.4 \text{ fb}^{-1}$  and the current measurement strategy. At the end of Run 2 the median uncertainty of the effective lifetime will be  $\sim 0.2 \text{ ps}$  which is reduced to  $\sim 0.08 \text{ ps}$  with  $50 \text{ fb}^{-1}$  and  $\sim 0.03 \text{ ps}$  with

300 fb<sup>-1</sup>. Therefore, with 300 fb<sup>-1</sup> the precision on the effective lifetime will be able to distinguish new physics effects. These expected uncertainties 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 is 0.05 ps which is too large to allow possible new physics 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 new physics 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 new physics models.



# Bibliography

- [1] LHCb, 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*, [arXiv:1703.0574](https://arxiv.org/abs/1703.0574).
- [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] R. Davis, D. S. Harmer, and K. C. Hoffman, *Search for neutrinos from the sun*, Phys. Rev. Lett. **20** (1968) 1205.
- [6] Super-Kamiokande, 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](https://arxiv.org/abs/hep-ex/9805021), [Erratum: Phys. Rev. Lett. 81, 4279(1998)].
- [7] 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.
- [8] 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.
- [9] Z. F., *Spectral displacement of extra galactic nebulae*, Helv. Phys. Acta **86** (1933).
- [10] R. M. S. and R. A. H., *Comparison of rotation curves of different galaxy types*, Astronomy and Astrophysics **26** (1973) 483.
- [11] WMAP, J. Dunkley *et al.*, *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Likelihoods and Parameters from the WMAP data*, Astophys. J. Suppl. **180** (2009) 306, [arXiv:0803.0586](https://arxiv.org/abs/0803.0586).
- [12] Planck, P. A. R. Ade *et al.*, *Planck 2015 results. XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13, [arXiv:1502.0158](https://arxiv.org/abs/1502.0158).
- [13] 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)].

- [14] 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](https://arxiv.org/abs/hep-ph/9312215).
- [15] D. M., *Supersymmetry and String Theory: Beyond the Standard Model*, Cambridge University Press, 2007.
- [16] 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](https://arxiv.org/abs/hep-ph/0211168).
- [17] 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](https://arxiv.org/abs/1202.1391).
- [18] 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](https://arxiv.org/abs/1207.7235).
- [19] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1, [arXiv:1207.7214](https://arxiv.org/abs/1207.7214).
- [20] Belle Collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of  $b \rightarrow K^* \ell^+ \ell^-$* , *Phys. Rev. Lett.* **118** (2017) 111801.
- [21] 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](https://arxiv.org/abs/1403.8044).
- [22] 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:1506.0861](https://arxiv.org/abs/1506.0861), [Erratum: *Phys. Rev. Lett.* 115,no.15,159901(2015)].
- [23] 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](https://arxiv.org/abs/1303.0571).
- [24] 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:1507.0323](https://arxiv.org/abs/1507.0323).
- [25] 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](https://arxiv.org/abs/1205.5442).
- [26] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  decay using  $3 \text{ fb}^{-1}$  of integrated luminosity*, *JHEP* **02** (2016) 104, [arXiv:1512.0444](https://arxiv.org/abs/1512.0444).
- [27] 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:1506.0877](https://arxiv.org/abs/1506.0877).
- [28] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using  $B^+ \rightarrow K^+ \ell^+ \ell^-$  decays*, *Phys. Rev. Lett.* **113** (2014) 151601.

- [29] LHCb collaboration, R. Aaij *et al.*, *Search for new physics with  $b \rightarrow s\ell^+\ell^-$  decays at lhcb*, CERN seminar 18th April 2017.
- [30] ATLAS collaboration, G. Aad *et al.*, *Search for the decay  $B_s^0 \rightarrow \mu\mu$  with the ATLAS detector*, *Phys. Lett. B* **713** (2012) 387, [arXiv:1204.0735](#).
- [31] 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. C* **76** (2016), no. 9 513, [arXiv:1604.0426](#).
- [32] CMS collaboration, S. Chatrchyan *et al.*, *Search for  $B(s)$  and  $B$  to dimuon decays in  $pp$  collisions at 7 TeV*, *Phys. Rev. Lett.* **107** (2011) 191802, [arXiv:1107.5834](#).
- [33] 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](#).
- [34] CMS collaboration, S. Chatrchyan *et al.*, *Measurement of the  $B(s)$  to mu+ mu- branching fraction and search for  $B0$  to mu+ mu- with the CMS Experiment*, *Phys. Rev. Lett.* **111** (2013) 101804, [arXiv:1307.5025](#).
- [35] 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. B* **699** (2011) 330, [arXiv:1103.2465](#).
- [36] 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. B* **708** (2012) 55, [arXiv:1112.1600](#).
- [37] 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](#).
- [38] 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](#).
- [39] 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](#).
- [40] 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](#).
- [41] 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:1606.0091](#).
- [42] 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](#).
- [43] I. Dunietz, R. Fleischer, and U. Nierste, *In pursuit of new physics with  $B_s$  decays*, *Phys. Rev. D* **63** (2001) 114015, [arXiv:hep-ph/0012219](#).
- [44] 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](#).

- [45] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.
- [46] M. Kobayashi and T. Maskawa, *Cp-violation in the renormalizable theory of weak interaction*, Progress of Theoretical Physics **49** (1973), no. 2 652, [arXiv:/oup/backfile/Content\\_public/Journal/ptp/49/2/10.1143/PTP.49.652/2/49-2-652.pdf](https://arxiv.org/abs/1011.4311).
- [47] L. Wolfenstein, *Parametrization of the kobayashi-maskawa matrix*, Phys. Rev. Lett. **51** (1983) 1945.
- [48] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.
- [49] K. De Bruyn *et al.*, *Branching Ratio Measurements of  $B_s$  Decays*, Phys. Rev. **D86** (2012) 014027, [arXiv:1204.1735](https://arxiv.org/abs/1204.1735).
- [50] 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.
- [51] K. G. Wilson, *Non-lagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.
- [52] W. Wilson, Kenneth G.and Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Communications in Mathematical Physics **24** (1972), no. 2 87.
- [53] 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](https://arxiv.org/abs/1204.1737).
- [54] 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](https://arxiv.org/abs/1303.3820).
- [55] Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and  $\tau$ -lepton properties as of summer 2016*, [arXiv:1612.0723](https://arxiv.org/abs/1612.0723).
- [56] 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](https://arxiv.org/abs/1109.5115).
- [57] 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](https://arxiv.org/abs/1311.0903).
- [58] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, Eur. Phys. J. **C77** (2017), no. 2 112, [arXiv:1607.0029](https://arxiv.org/abs/1607.0029).
- [59] R. Fleischer, R. Jaarsma, and G. Tetlalmatzi-Xolocotzi, *In Pursuit of New Physics with  $B_{s,d}^0 \rightarrow \ell^+ \ell^-$* , [arXiv:1703.1016](https://arxiv.org/abs/1703.1016).
- [60] R. J. Knegjens, *Strategies to Hunt for New Physics with Strange Beauty Mesons*, PhD thesis, Vrije U., Amsterdam, 2014.

- [61] 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](https://arxiv.org/abs/1411.3161).
- [62] 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](https://arxiv.org/abs/hep-ph/0207036).
- [63] S. P. Martin, *A Supersymmetry primer*, [arXiv:hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356), [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [64] W.-S. Hou, *Source of CP Violation for the Baryon Asymmetry of the Universe*, *Chin. J. Phys.* **47** (2009) 134, [arXiv:0803.1234](https://arxiv.org/abs/0803.1234).
- [65] 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](https://arxiv.org/abs/1012.3893).
- [66] LHCb collaboration, CMS collaboration, *Combination of results on the rare decays  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  from the CMS and LHCb experiments*, .
- [67] L. J. and M. B. Wise, *Flavor changing higgs boson couplings*, *Nuclear Physics B* **187** (1981), no. 3 397 .
- [68] E. Witten, *Dynamical Breaking of Supersymmetry*, *Nucl. Phys.* **B188** (1981) 513.
- [69] A. J. Buras, P. H. Chankowski, J. Rosiek, and L. Slawianowska, *Correlation between  $\Delta M_s$  and  $B_{s,d}^0 \rightarrow \mu^+ \mu^-$  in supersymmetry at large  $\tan \beta$* , *Phys. Lett.* **B546** (2002) 96, [arXiv:hep-ph/0207241](https://arxiv.org/abs/hep-ph/0207241).
- [70] 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](https://arxiv.org/abs/hep-ph/9909476).
- [71] G. Isidori and A. Retico, *Scalar flavor changing neutral currents in the large tan beta limit*, *JHEP* **11** (2001) 001, [arXiv:hep-ph/0110121](https://arxiv.org/abs/hep-ph/0110121).
- [72] 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:1611.0493](https://arxiv.org/abs/1611.0493).
- [73] 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:1608.0850](https://arxiv.org/abs/1608.0850).
- [74] 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).
- [75] 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:1511.0190](https://arxiv.org/abs/1511.0190).
- [76] S. Fajfer and N. Košnik, *Vector leptoquark resolution of  $R_K$  and  $R_{D^{(*)}}$  puzzles*, *Phys. Lett.* **B755** (2016) 270, [arXiv:1511.0602](https://arxiv.org/abs/1511.0602).
- [77] W. Altmannshofer, C. Niehoff, and D. M. Straub,  *$B_s \rightarrow \mu^+ \mu^-$  as current and future probe of new physics*, [arXiv:1702.0549](https://arxiv.org/abs/1702.0549).

- [78] CERN. <http://home.cern/about/member-states>.
- [79] L. collaboration. <http://lhcb-public.web.cern.ch/lhcb-public/>.
- [80] collaboration. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResultsOverview>.
- [81] A. collaboration. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> and <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [82] LHCb collaboration, S. Amato *et al.*, *LHCb technical proposal*, .
- [83] LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.
- [84] LHCb collaboration, *LHCb technical design report: Reoptimized detector design and performance*, .
- [85] LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, *Int. J. Mod. Phys. A* **30** (2015), no. 07 1530022, [arXiv:1412.6352](https://arxiv.org/abs/1412.6352).
- [86] 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).
- [87] R. E. Kalman, *A New Approach to Linear Filetering and Prediction Problems*, *Journal of Basic Engineering* **82** (1960) 35.
- [88] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444.
- [89] 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).
- [90] R. W. Forty and O. Schneider, *RICH pattern recognition*, Tech. Rep. LHCb-98-040, CERN, Geneva, Apr, 1998.
- [91] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, *Eur. Phys. J. C* **73** (2013) 2431, [arXiv:1211.6759](https://arxiv.org/abs/1211.6759).
- [92] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, *JINST* **8** (2013) P10020, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249).
- [93] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, *Nucl. Instrum. Meth. A* **559** (2006) 190.
- [94] 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).
- [95] 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).
- [96] 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).

- [97] 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.
- [98] P. Mato, *GAUDI-Architecture design document*, .
- [99] LHCb collaboration, R. Antunes-Nobrega *et al.*, *LHCb computing: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2005. Submitted on 11 May 2005.
- [100] F. Stagni *et al.*, *LHCbDirac: Distributed computing in LHCb*, *J. Phys. Conf. Ser.* **396** (2012) 032104.
- [101] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, *Nucl. Instrum. Meth.* **A389** (1997) 81.
- [102] 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.
- [103] LHCb collaboration, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, *J. Phys. Conf. Ser.* **331** (2011) 032023.
- [104] 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).
- [105] 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).
- [106] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.
- [107] 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).
- [108] GEANT4, S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.
- [109] J. Allison *et al.*, *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [110] I. Bird, *Computing for the Large Hadron Collider*, *Ann. Rev. Nucl. Part. Sci.* **61** (2011) 99.
- [111] W. L. C. Grid. <http://www.cern.ch/LHCgrid>.
- [112] S. Paterson and A. Tsaregorodtsev, *DIRAC Infrastructure for Distributed Analysis*, .
- [113] <http://ganga.web.cern.ch/ganga/>.
- [114] K. Harrison *et al.*, *GANGA: a user-grid interface for atlas and lhcb*, CoRR [cs.SE/0306085](https://arxiv.org/abs/cs/0306085) (2003).

- [115] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength  $|V_{ub}|$  using baryonic decays*, *Nature Phys.* **11** (2015) 743, [arXiv:1504.0156](#).
- [116] 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:1509.0041](#).
- [117] C. M. Bouchard *et al.*,  *$B_s \rightarrow K \ell \nu$  form factors from lattice QCD*, *Phys. Rev.* **D90** (2014) 054506, [arXiv:1406.2279](#).
- [118] RBC and UKQCD Collaborations, J. M. Flynn *et al.*,  *$b \rightarrow \pi \ell \nu$* , *Phys. Rev. D* **91** (2015) 074510.
- [119] 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](#).
- [120] 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. D* **90** (2014), no. 3 032009, [arXiv:1407.2126](#).
- [121] 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. D* **86** (2012) 114025, [arXiv:1207.0265](#).
- [122] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.
- [123] 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.
- [124] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, [arXiv:physics/0703039](#).
- [125] 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.
- [126] J. Stevens and M. Williams, *uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers*, *JINST* **8** (2013) P12013, [arXiv:1305.7248](#).
- [127] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, *Nucl. Instrum. Meth.* **A555** (2005) 356, [arXiv:physics/0402083](#).
- [128] 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.
- [129] 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**.
- [130] ARGUS, H. Albrecht *et al.*, *Search for hadronic  $b \rightarrow u$  decays*, *Phys. Lett.* **B241** (1990) 278.

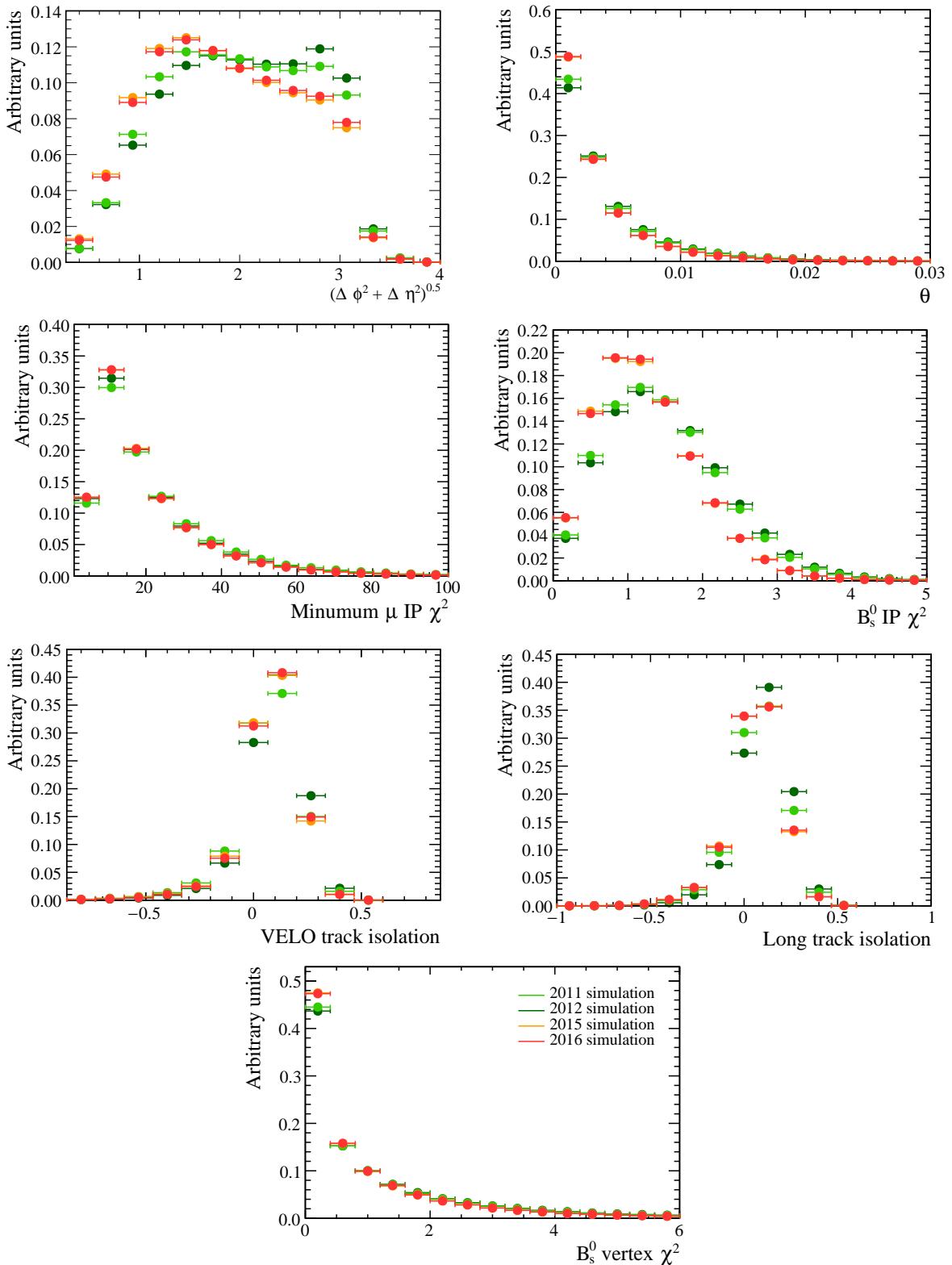
- [131] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, [arXiv:1312.5000](https://arxiv.org/abs/1312.5000).
- [132] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data-driven measurement of trigger efficiencies with the TisTos method*, Tech. Rep. LHCb-INT-2013-038. CERN-LHCb-INT-2013-038, CERN, Geneva, Jun, 2013.
- [133] LHCb collaboration, *Updated average  $f_s/f_d$  b-hadron production fraction ratio for 7 TeV pp collisions*, .
- [134] A. L. Read, *Presentation of search results: the cl s technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002), no. 10 2693.
- [135] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [arXiv:physics/0306116](https://arxiv.org/abs/physics/0306116), [,186(2003)].
- [136] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP violating phase  $\phi_s$  in  $\bar{B}_s^0 \rightarrow J/\psi f_0(980)$* , Phys. Lett. **B707** (2012) 497, [arXiv:1112.3056](https://arxiv.org/abs/1112.3056).
- [137] LHCb collaboration, R. Aaij *et al.*, *Effective lifetime measurements in the  $B_s^0 \beta K^+ K^-$  and  $B_s^0 \beta^+ K^-$  decays*, Phys. Lett. **B736** (2014) 446, [arXiv:1406.7204](https://arxiv.org/abs/1406.7204).
- [138] 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:1608.0485](https://arxiv.org/abs/1608.0485).
- [139] 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:1503.0708](https://arxiv.org/abs/1503.0708).
- [140] 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](https://arxiv.org/abs/1408.0275).
- [141] LHCb collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, [arXiv:0912.4179](https://arxiv.org/abs/0912.4179).
- [142] 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.
- [143] O. Augusto, *LHCb; LHCb Jet Reconstruction*, Dec, 2012.
- [144] 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.
- [145] D. J. Jackson, *A Topological vertex reconstruction algorithm for hadronic jets*, Nucl. Instrum. Meth. **A388** (1997) 247.
- [146] A. Mordà and G. Mancinelli, *Rare dileptonic B0(s) meson decays at LHCb*, PhD thesis, Aix-Marseille U., 2015, presented 28 Sep 2015.

- [147] 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](https://arxiv.org/abs/hep-ex/0508036), [Erratum: Phys. Rev. Lett.95,249905(2005)].

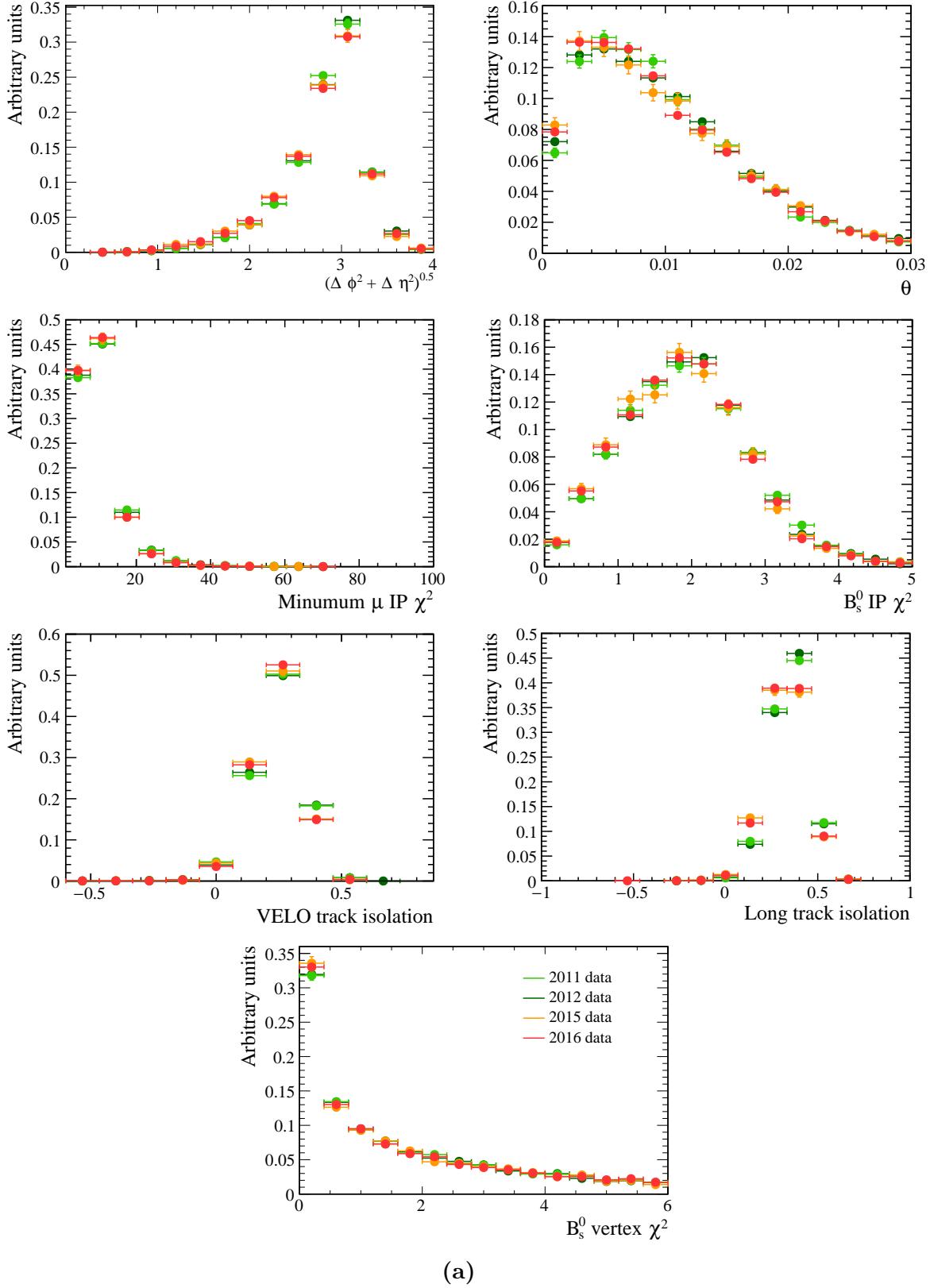
## Appendix A

# 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  $m_{\mu\mu} > 5447 \text{ MeV}/c^2$  and passing the selection cuts in Table 4.12.



**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  $m_{\mu\mu} > 5447 \text{ MeV}/c^2$  and 2012 simulated  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays.



# 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 in the 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 [143] 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 [144]. 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 constrain 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 IP  $\chi^2$  of the  $B_s^0$ ;
- vertex fit  $\chi^2/ndof$  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 IP  $\chi^2$  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 [125].<sup>1</sup>; and

---

<sup>1</sup>Replacing this isolation variable with the Long track and VELO track isolations does not significantly improve the overall performance of either BDT.

- ZVtop isolation variable which uses a topological vertex algorithm [145] and is defined in reference [146].

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;
- 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 [147]. 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,  $I_\mu$ , 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 ??;
- 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  $\mu^+$  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 force 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  $B_s^0$ .

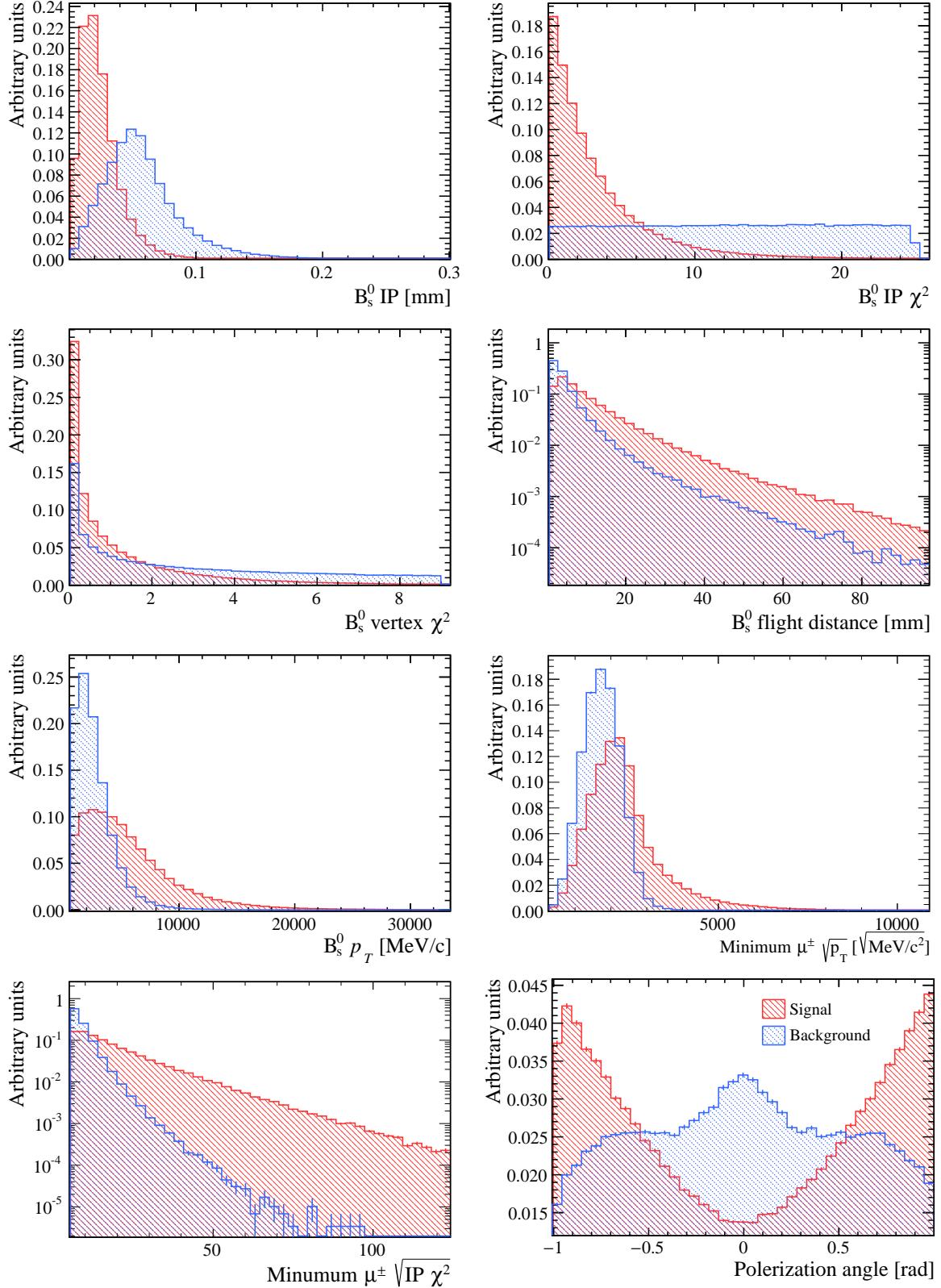
The distributions of the input variables for  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  2012 sim06 simulated decays are shown in Figure B.3.

## B.2 Training parameters

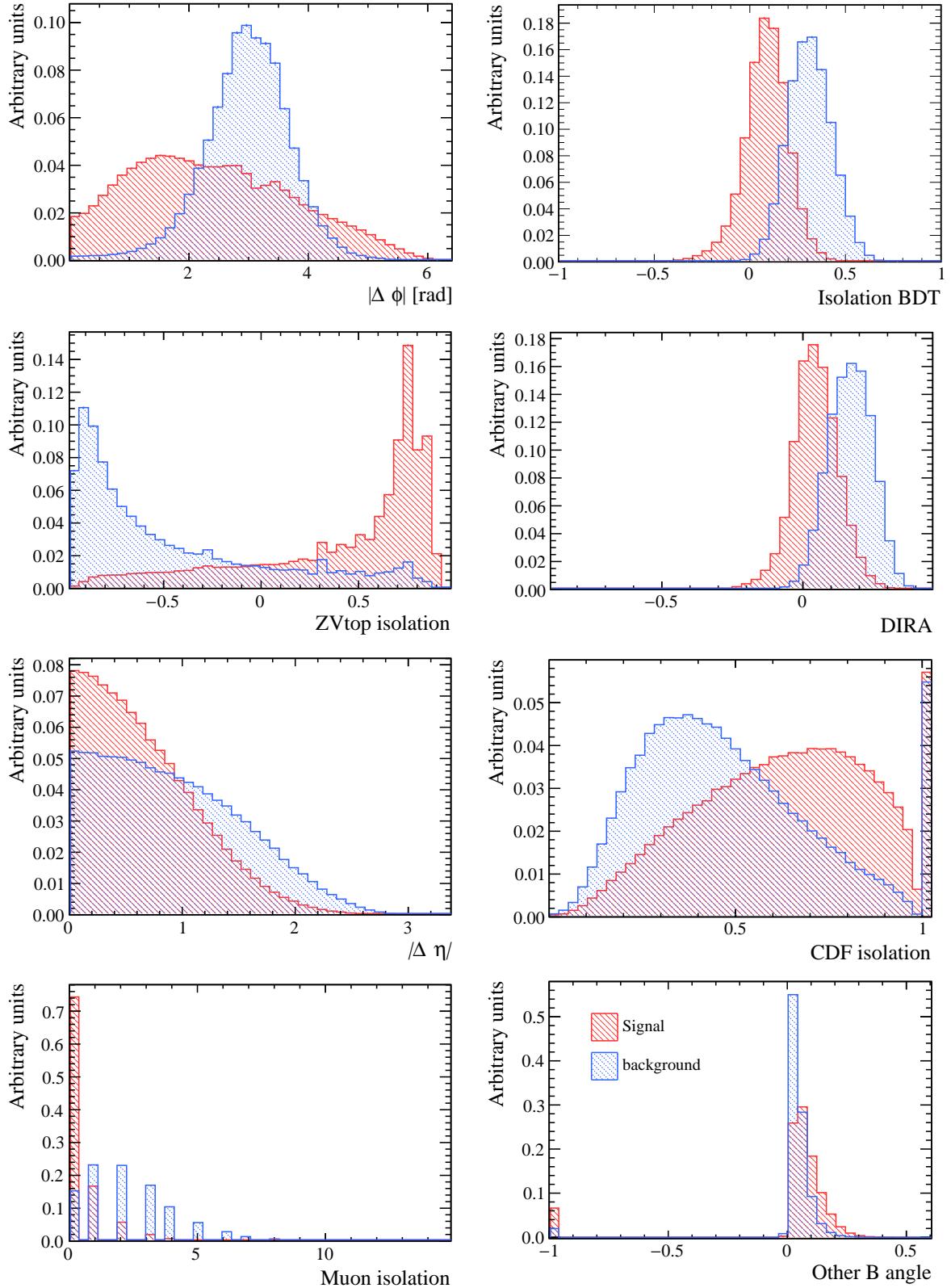
The training parameters discussed in Section 4.3.4.1 put constraints on how a BDT separates signal and background decays. The training parameters used in the adaptive boost BDT were optimised by iterating over different training parameter values and choosing the BDT that gave the best signal significance 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 [126] have been used where it was shown that different training parameters had a small impact of the overall BDT performance.

## B.3 Overtraining test

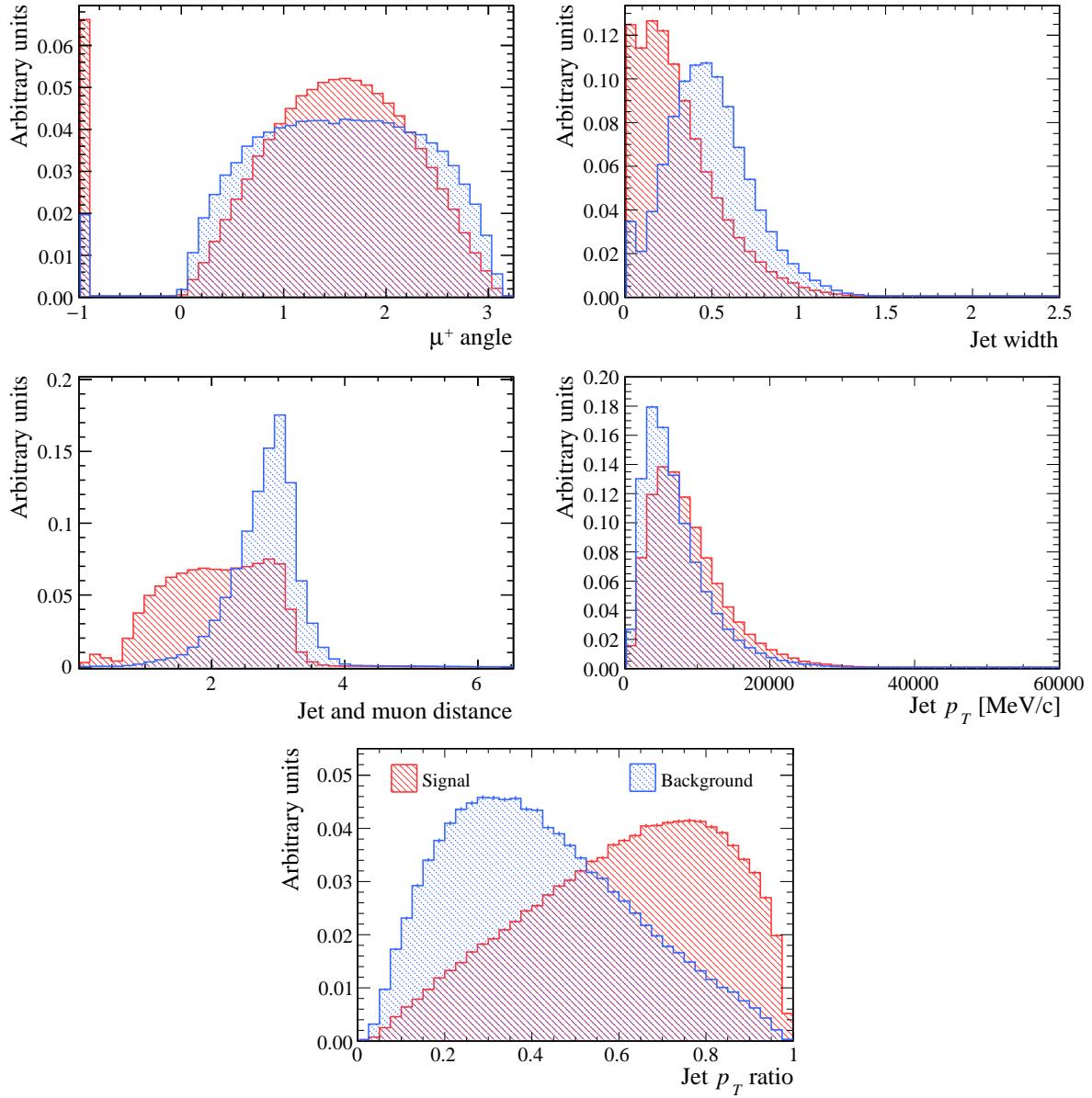
As discussed in Section 4.3.4.1, it is important that BDTs are not overtrained. To test this assumption the signal and background samples are both split in two to create a training set and a testing set. A BDT is trained using the training set, and the BDT is then applied to both the training and testing 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 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.



**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 sim06 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 sim06 simulated decays.

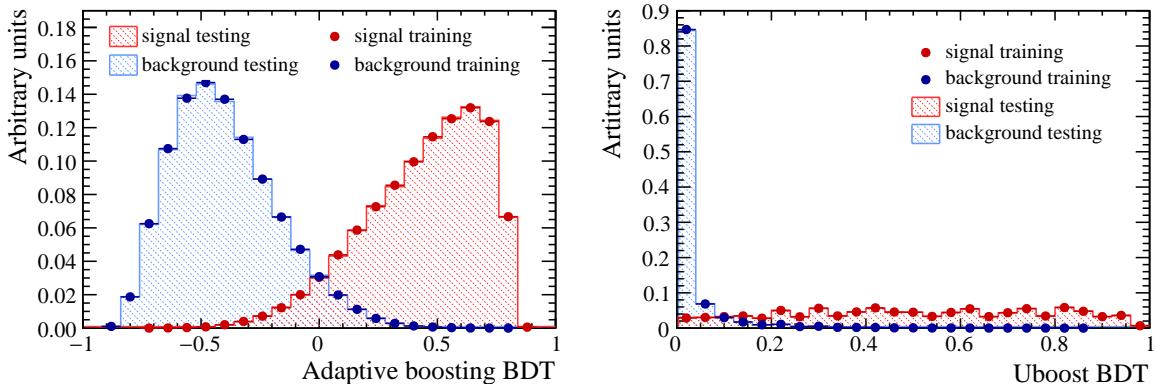


**Fig. B.3** The distributions of the input variables used in only the uBoost BDT for  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  2012 sim06 simulated decays.

| Adaptive Boost BDT |       | uBoost BDT |       |
|--------------------|-------|------------|-------|
| Parameter          | Value | Parameter  | Value |
| nTrees             | 1000  | nTrees     | 100   |
| MinNodeSize        | 5%    | nEventsMin | 100   |
| MaxDepth           | 3     | MaxDepth   | 4     |
| $\beta$            | 0.1   | $\beta$    | 1.0   |
| nCuts              | 30    | nCuts      | 200   |

**Table B.1** Training parameters used to specify the training of the adaptive boost and uBoost BDT.

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 the adaptive boost BDT (left) and the uBoost BDT (right).

## **Appendix C**

### **Details of pseudoexperiment configuration**

Toy study info. Fun!

